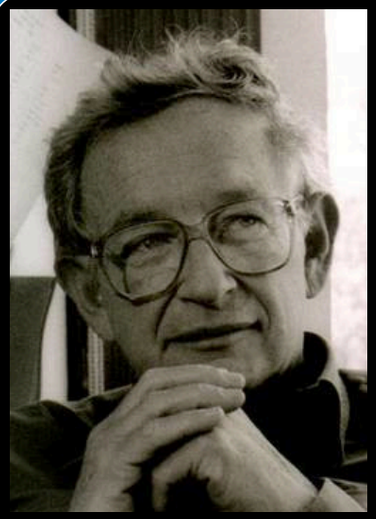
Sloppy Models, Differential geometry, and Why Science Works



More is Different

Unreasonable Effectiveness of Mathematics

Why is Nature comprehensible?

Emergence of laws from complex underpinnings.

Today: a different *kind* of emergence

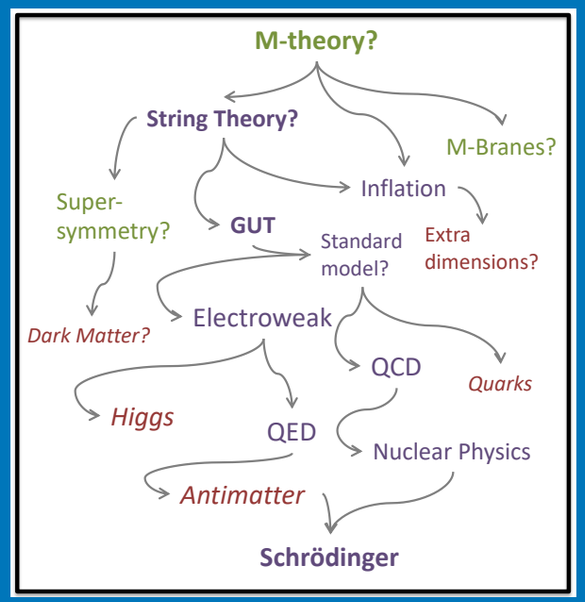
JPS, Han Kheng Teoh, Katherine Quinn, Colin Clement, Lorien Hayden, Alex Alemi, Archishman Raju, Mark Transtrum, Ben Machta, Heather Wilber, Alex Townsend, Will Bergen, Cameron Duncan, Danilo Liarte, Ricky Chachra, Kevin Brown, Ryan Gutenkunst, Josh Waterfall, Fergal Casey, Chris Myers, …

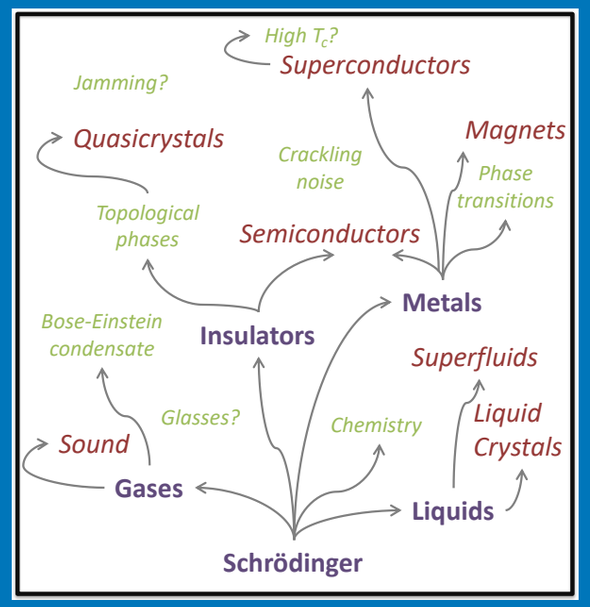
We are gathered here to explain why nature is comprehensible. As Eugene Wigner asked, why does mathematics describe the natural world? Phil Anderson asserted that there are emergent laws describing the collective behavior of complex systems — that the laws governing our economy or ecosystem are every bit as ‘fundamental’ as those of high-energy physics.

Today I want to introduce a different kind of emergence. We shall use information geometry and a kind of interpolation theory to show that complex systems with many microscopic parameters will quite generally exhibit comprehensible collective behaviors.

Systematic Emergence

Low dimensional behavior of complex systems





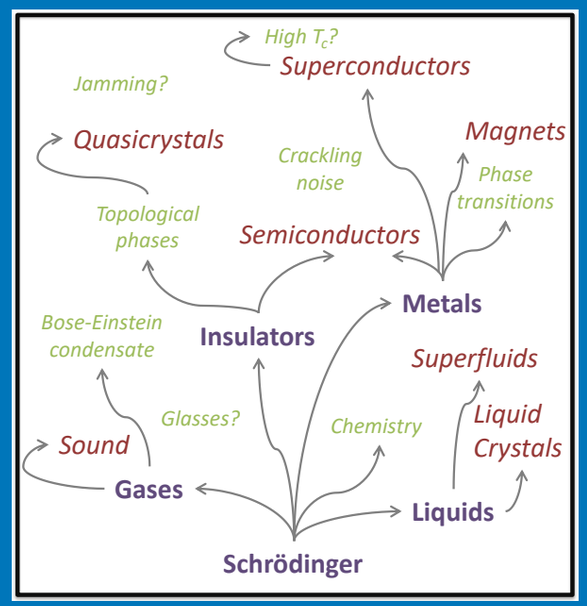
Physics: Controlled approximations, small parameters

High-energy physicists, in their quest to understand nature, try climbing up the energy tree to search for simplicity and elegance. Condensed matter physicists view the ‘fundamental’ Schrodinger equation as horribly complex, and climb a length-scale tree in the search for elegant, emergent models.

Systematic Emergence

Low dimensional behavior of complex systems

- Continuum Limits
- Renormalization
- Separation of Time Scales
- Asymptotic analysis
- Symmetry
- Center manifold theorem
- …



Physics: Controlled approximations, small parameters

High energy theorists and we share many of the same tools — symmetry, renormalization under coarse graining, asymptotic analysis and instantons — in deriving our emergent models. In dynamical systems theory, time-scale separation and the center manifold theorem allow for systematic derivation of emergent, simpler behavior.

Systematic Emergence

Low dimensional behavior of complex systems

- Continuum Limits
- Renormalization
- Separation of Time Scales
- Asymptotic analysis
- Symmetry
- Center manifold theorem
- …

Diffusion on a square lattice a:
Gradient expansion

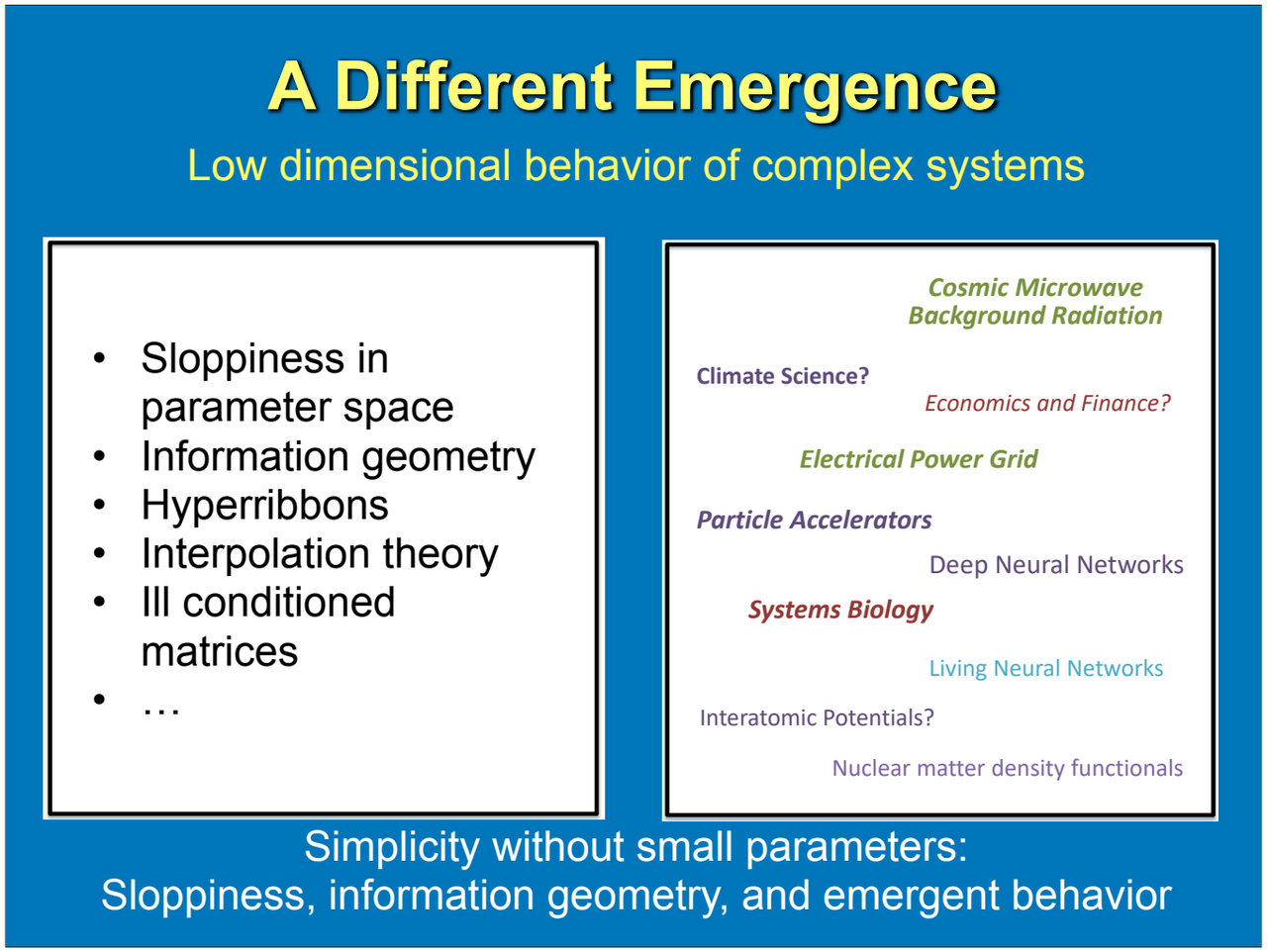
$$\frac{\partial \rho}{\partial t} = -\nabla \cdot \mathbf{J}$$

$$J_i = -D_{ij}\partial_j \rho + \cancel{E_{ijk}\partial_j\partial_k \rho} + F_{ijkl}\partial_j\partial_k\partial_l\rho$$

$$\frac{\partial \rho}{\partial t} = D\partial_i^2\rho + \cancel{F_{\text{sym}}(\partial_i^2)^2\rho} + \cancel{F_{\text{square}}(\partial_x^4 + \partial_y^4)\rho}$$

Continuum limit: wave vector $q << 1/a$. $\quad F/D \sim a^2; q^2a^2 \ll 1$

So, to take a continuum limit, we use the fact that we are large, slow creatures compared to atoms, and expand in low frequencies and small gradients. The size of the higher gradient terms for diffusion on a square lattice can be estimated by dimensional analysis. They are smaller by factors of the microscopic length (the lattice constant a) times the macroscopic gradient in density (the wave vector q). Physics here is a well-controlled expansion in powers of the small parameter (q a). Diffusion on a square lattice will have two ‘least sloppy’ corrections ~q^2 a^2: one isotropic, and one with square symmetry. Further corrections will be smaller and smaller by higher powers of q a, giving us a sloppy spectrum.

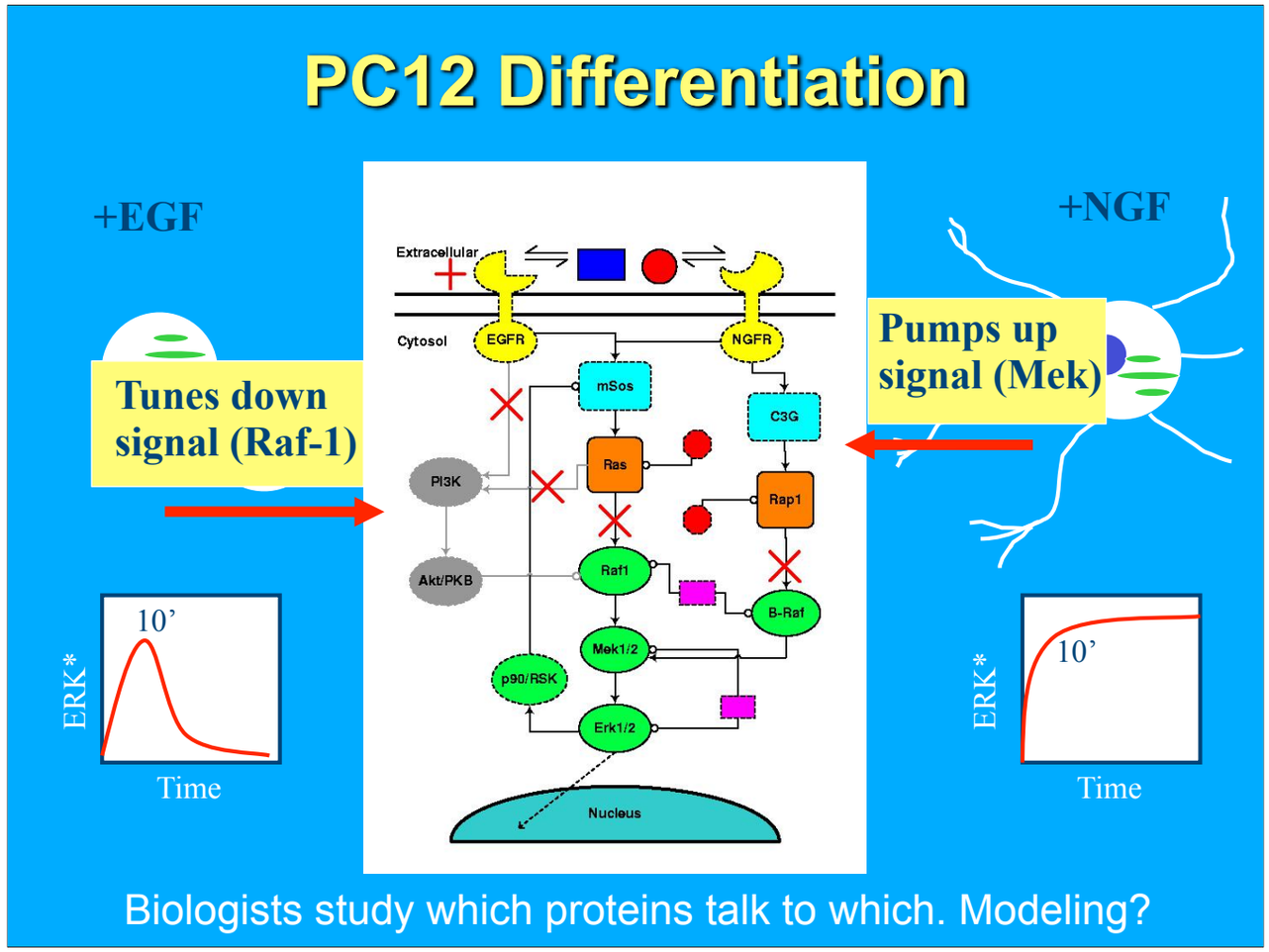


Today we'll introduce a different mechanism, based on interpolation theory and information geometry, that also provides a fundamental expectation that complex systems with multiple microscopic parameters should exhibit comprehensible, low-dimensional behavior. We have applied our methods to models in systems biology, complex instruments like particle accelerators, the electrical power grid, and the cosmic microwave background radiation.

We'll begin by arguing that many multiparameter models are sloppy — their behavior depends strongly on only a few parameter combinations. We will then use methods of information geometry to show that the model manifold of possible behaviors is an effectively low-dimensional hyperribbon. Finally, we will see how these hyperribbons leads to emergent models, and visualize them for the cosmic microwave background radiation and a deep neural network.

Systems Biology

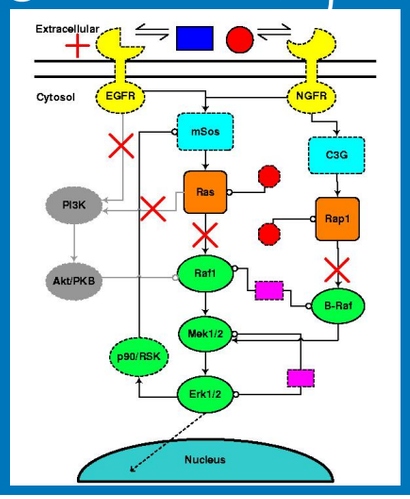
Two members of my group two decades ago got us together with Rick Cerione, a cancer biologist here. We started working on a model of one signaling pathway in a cell — doing what is now called ‘systems biology’. The model behaved in quite surprising ways.



Cerione wrote down this diagram of the interactions between proteins that carry a signal from outside the membrane of a cell into the nucleus. If the cell is exposed to EGF, it triggers a cascade of reactions that pump a pulse of the protein ERK into the nucleus, after about ten minutes. The nucleus then causes the cell to divide. If it is exposed to NGF, the ERK signal builds for ten minutes but stays high, causing the cell to develop neuronal branches.

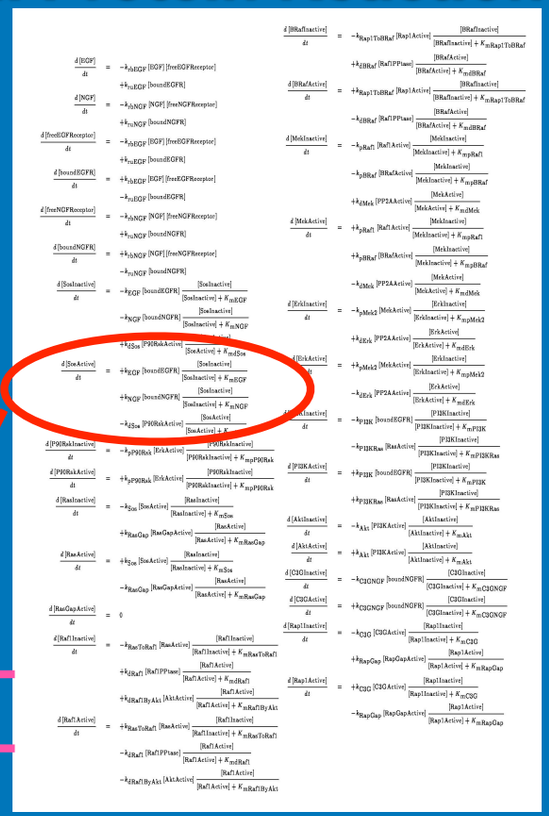
Systems Biology: Cell Protein Reactions

Kevin Brown



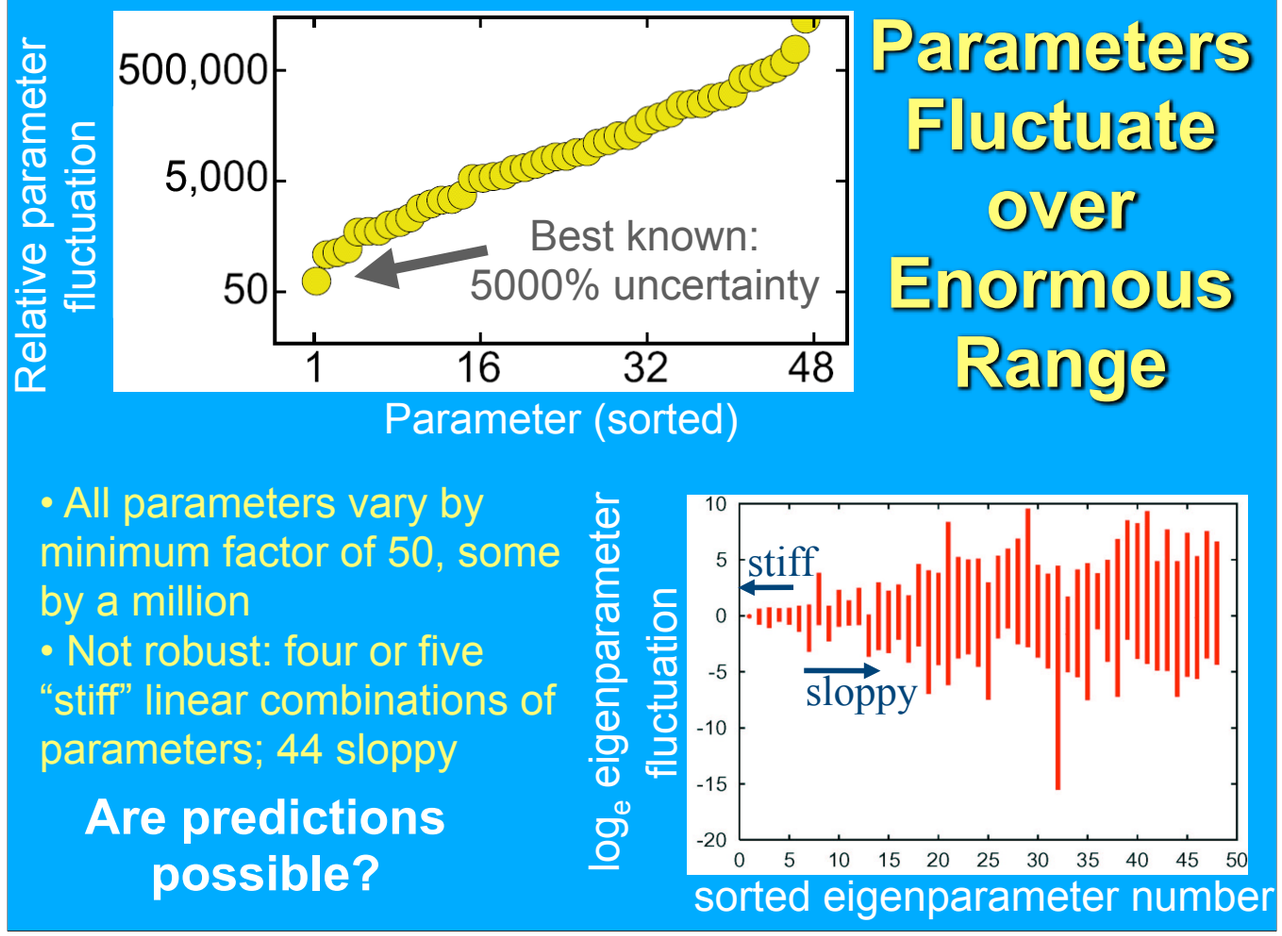
Membrane → Nucleus

$$\frac{d[\text{SosActive}]}{dt} = +k_{\text{EGF}}\,[\text{boundEGFR}]\,\frac{[\text{SosInactive}]}{[\text{SosInactive}]+K_{\text{mEGF}}} + k_{\text{NGF}}\,[\text{boundNGFR}]\,\frac{[\text{SosInactive}]}{[\text{SosInactive}]+K_{\text{mNGF}}} - k_{\text{dSos}}\,[\text{P90RskActive}]\,\frac{[\text{SosActive}]}{[\text{SosActive}]+K_{\text{mdSos}}}$$

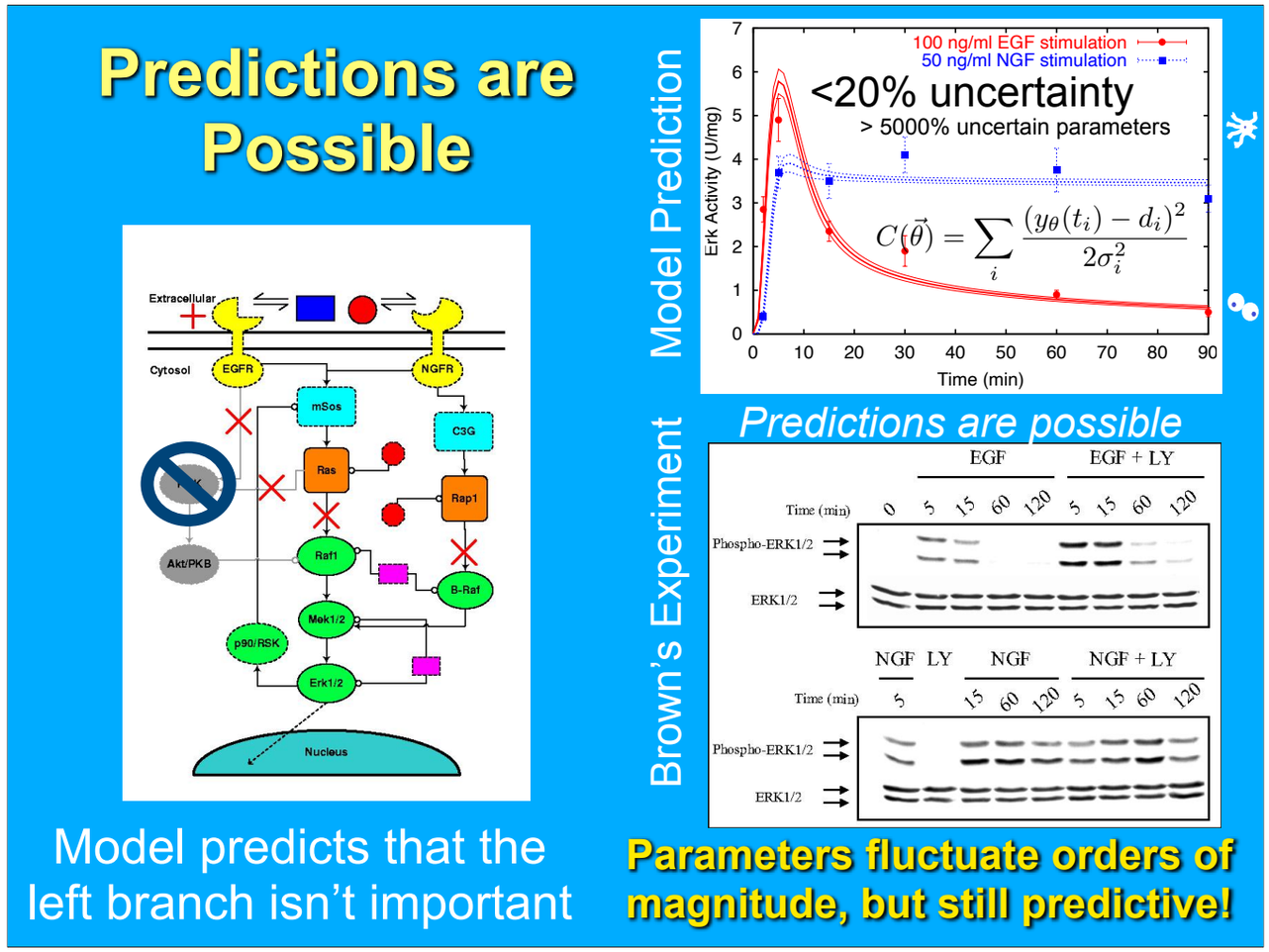


48 Parameter Fit!

Here is a multiparameter model, describing how a crazy collection of proteins carries information from the cell membrane to the nucleus. The 29 nonlinear differential equations that govern this systems biology model depend on 48 reaction rates and saturation constants. There were 63 experimental data points for this well-understood biological system. Cerione told us that the reactions were pretty well known, but the reaction constants were not. Can we extract the 48 rates from 63 measurements? Can we make predictions for new experiments?



I was skeptical. With 48 parameters, surely we could not estimate all their values with only 63 data points. And if we made a prediction, surely if we were wrong we could then fit 64 data points with the same 48 parameters. I was half right. When we typed in Cerione's model (and modified and prodded it to fit the data, challenges we shall describe later), we found an enormous range of parameters could fit the behavior well. The best known parameter has 5000% error bars!

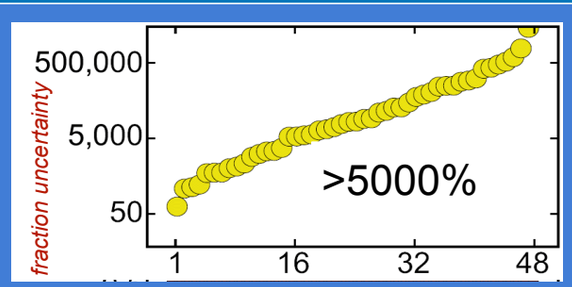


Cerione explained that, until the model got so complicated, he could work out what it would do without simulations. He was confident it would make good predictions. Sure enough, we tested our ensemble of fits by making a prediction for an experiment that used a drug LY to turn off one of the reactions (the EGF receptor triggering PI3D). Cerione thought that this would cause EGF stimulated ERK to stop being a pulse, and continue higher for longer. Our model disagreed. Kevin Brown, a theory grad student, did the experiment — and our model was correct! The left branch of the model, thought to suppress ERK production at late times, apparently did not.

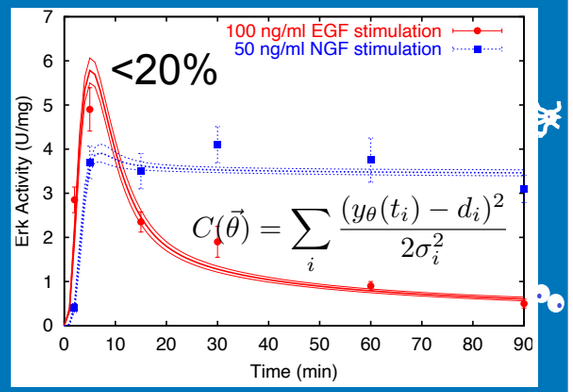
The amazing thing was not that our one prediction was correct (or three other post-dictions were correct). Nor was it that we made a prediction that was better than our expert (although Cerione is awesome). It was that we could make a prediction at all! With 5000% uncertain parameters, we made predictions with 20% errors.

48 Parameter Fit to Data

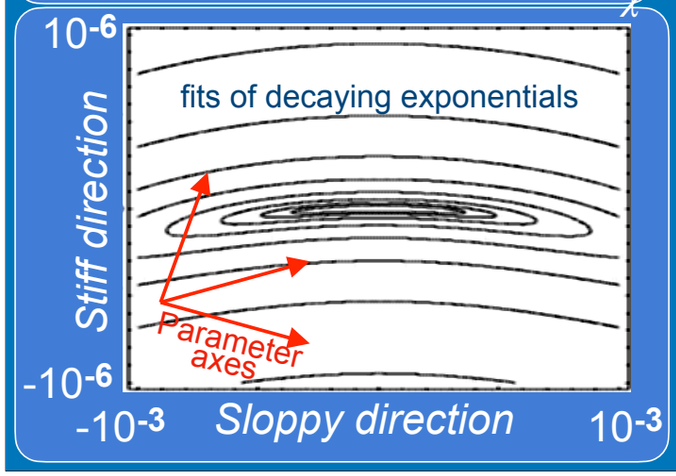
Kevin Brown, Josh Waterfall



Parameters are unknown



Predictions are possible



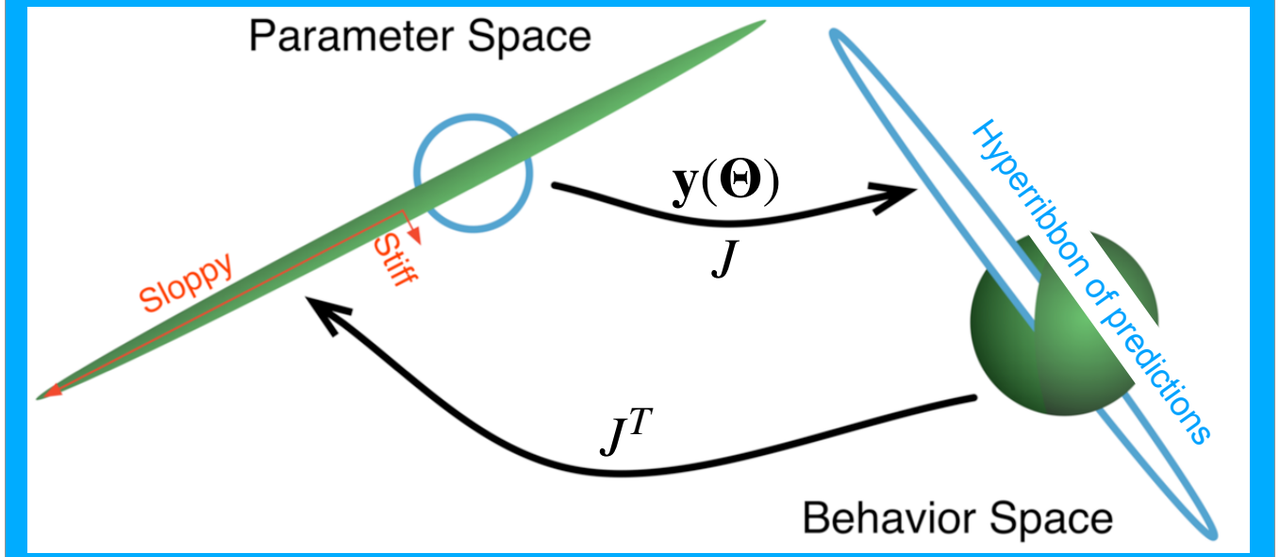
The behavior is *ill-conditioned*: it depends only on a few stiff parameter combinations.
Eigenvalues of cost Hessian

$$g_{\mu\nu} = \partial_\mu \partial_\nu C(\vec{\theta})$$

span enormous range

At lower left is our basic explanation, illustrated by a contour plot of the 'cost' C (squared error, aka chi squared). The axes are chosen along two of the 48 eigenvalues of the cost Hessian g_mn (which will become the metric tensor later). Vertical is a stiff direction — small changes cross many contours; horizontal is a sloppy direction, for which large changes in that combination of parameters hardly matter. A few of the eigenvalues for our model are stiff; the other forty-some are sloppy. Later we shall use this to find emergent descriptions with fewer parameters.
Two more things.
(1) The original 'bare' parameters are tilted with respect to the eigenvectors — each of which varies a lot as we traverse the sloppy direction.
(2) The vertical and horizontal scales differ by a factor of 1000. The real figure stretches about a kilometer to the right.

Models: Predictions about Data



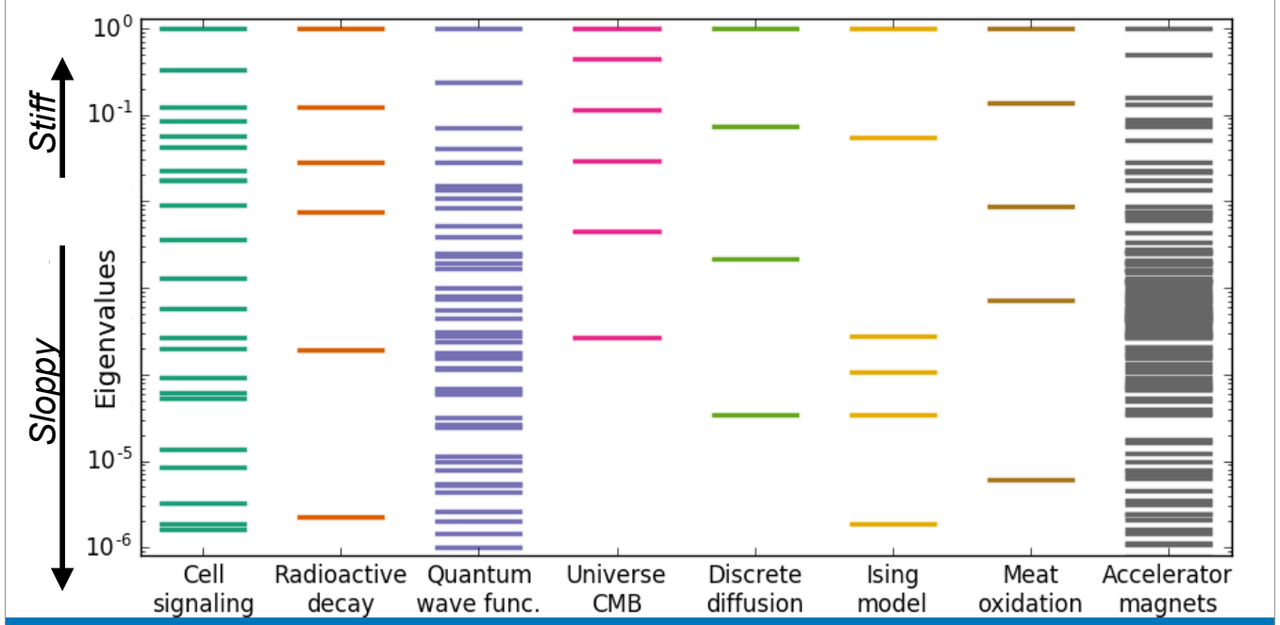
Scientific model: Predictions about behavior depend on physical constants (parameters) in the model.
Sloppiness: the behavior only depends on a few stiff parameter combinations.
Hyperribbons: The model predictions are confined to a flat, narrow region with a hierarchy of widths.

In these lectures, we will explore the sloppiness in parameter space, and the resulting manifold of predictions in behavior space. We shall understand sloppiness by showing that the Jacobian giving the linear approximation of y is extremely skewed. We shall make the case that science is possible — that multiparameter models give comprehensible results — because this skewness leads to a low-dimensional ‘hyperribbon’ of predictions in parameter space.

Other applications

But first, let us discuss models we have studied from a wide variety of fields. We did so
(1) to test whether sloppiness is found generally for multiparameter nonlinear models,
(2) to use ‘information geometry’ ideas from differential geometry and approximation theory to understand why they are sloppy, and
(3) to develop new algorithms for finding best fits and controlling multiparameter systems
(4) to find reduced, simpler models, allowing humans to understand what’s going on (and judge whether predictions can be trusted), and
(5) to understand why science works when there are no small parameters.

Sloppy Universality

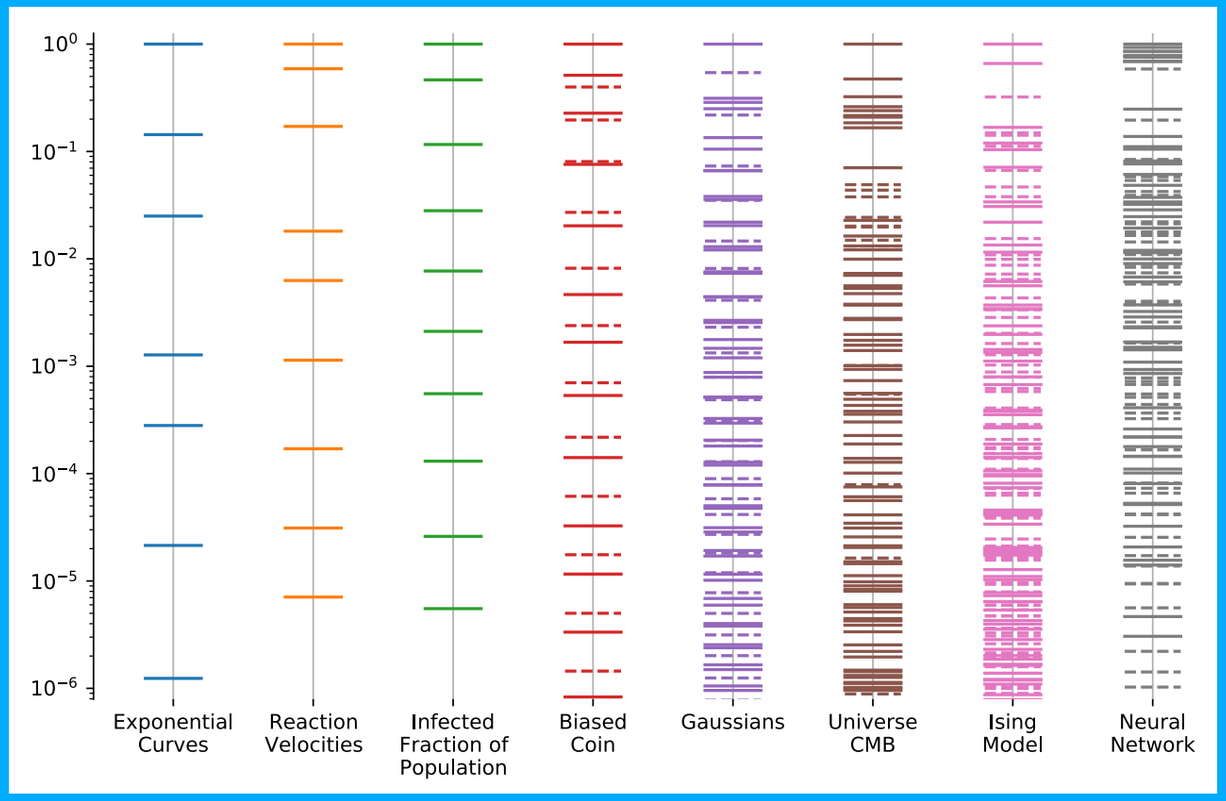


Enormous range of eigenvalues; Roughly equal density in log;
Observed in broad range of systems.

How can we use this to develop emergent laws?

The symptom of sloppiness is a broad range of eigenvalues, roughly equally spaced in log. On the left is our cell signaling model; each parameter difference is roughly a factor of three (and 3^48 is a very big number). Sums of exponentials, quantum wavefunctions used in high-precision chemistry, our model for the cosmic microwave background radiation, and particle accelerators are all sloppy. (Particle accelerators have hundreds of tunable magnets, so they have more eigenvalues — but the same enormous range). We also see that physics models are sloppy (diffusion and Ising), here for reasons we understand.

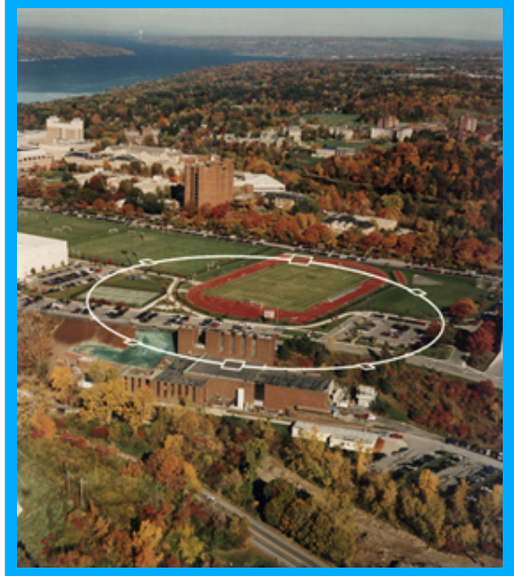
Hyperribbon widths in behavior



The widths of the model manifold in behavior space also show a hierarchy. There is a longest direction, and shorter and shorter axes with geometrically smaller widths. Ignoring the thin directions allows one to understand the overall behavior without encompassing all the details at once — an emergent simplicity.

Accelerator Optimization

CESR (Cornell Electron Storage Ring) as a sloppy model



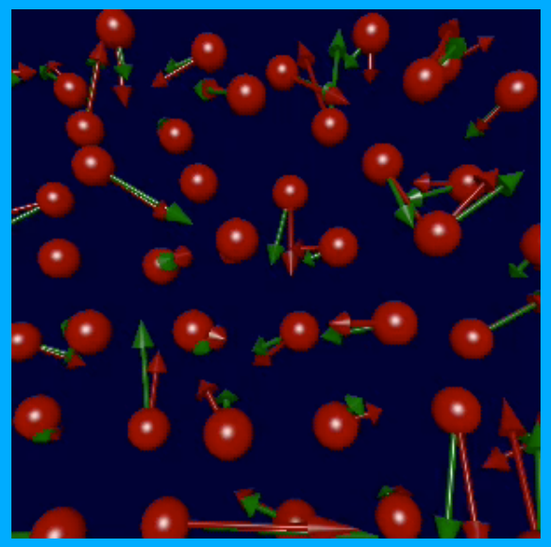
- 81 parameter space of magnet tuning parameters (‘knobs’) that need to be varied to optimize beam brightness.
- Used sloppy model analysis to reduce the search space to eight ‘stiff’ dimensions.
- Developed the method using simulations
- Tested the method on the (many million dollar) machine itself

Will Bergen, Cameron Duncan, Ivan Bazarov

Online storage ring optimization using dimension-reduction and genetic algorithms , William F. Bergan, Ivan V. Bazarov, Cameron J. R. Duncan, Danilo B. Liarte, David L. Rubin, and James P. Sethna, Phys. Rev. Accel. Beams 22, 054601 (2019), https://sethna.lassp.cornell.edu/pubPDF/SloppyAccelerator.pdf

Interatomic Potentials

Søren Frederiksen, Karsten Jacobsen, Kevin Brown, JPS



- Need atomic forces and energies
- Guess functional form for potential (17 parameters for MEAM, 5 for Finnis-Sinclair, …)
- Least-squares fit to DFT calculations of energy, forces for a variety of "important" atomic configurations
- Large errors for some predictions, small for others. Used sloppy ensembles to make error estimates.

Density functional theory's exchange-correlation potential (see Arias lectures) is sloppy too. Ensembles can predict accuracy…

"Bayesian Error Estimation in Density Functional Theory", J. J. Mortensen, K. Kaasbjerg, S. L. Frederiksen, J. K. Norskov, James P. Sethna, K. W. Jacobsen, Phys. Rev. Letters 95, 216401 (2005), https://sethna.lassp.cornell.edu/pubPDF/SloppyMo.pdf.

Variational Wavefunctions for Quantum Monte Carlo

Cyrus Umrigar, Josh Waterfall

The most accurate method for solving Schrödinger's equation for molecules rests on a variational wavefunction (followed by diffusion Monte Carlo):

$$\Psi(R) = \sum_m a_m \exp(-\sum_{i<j} u(r_{ij})) \begin{vmatrix} \phi_1(r_1) & \phi_1(r_2) & \dots & \phi_1(r_N) \\ \phi_2(r_1) & \phi_2(r_2) & \dots & \phi_2(r_N) \\ \dots & \dots & \dots & \dots \\ \phi_N(r_1) & \phi_N(r_2) & \dots & \phi_N(r_N) \end{vmatrix}_{m\uparrow} \begin{vmatrix} \phi_1(r_1) & \phi_1(r_2) & \dots & \phi_1(r_N) \\ \phi_2(r_1) & \phi_2(r_2) & \dots & \phi_2(r_N) \\ \dots & \dots & \dots & \dots \\ \phi_N(r_1) & \phi_N(r_2) & \dots & \phi_N(r_N) \end{vmatrix}_{m\downarrow} \dots$$

Often parameters θ are varied to minimize energy (find ground states)
Umrigar developed a method which fits the local energies at randomly chosen R_j

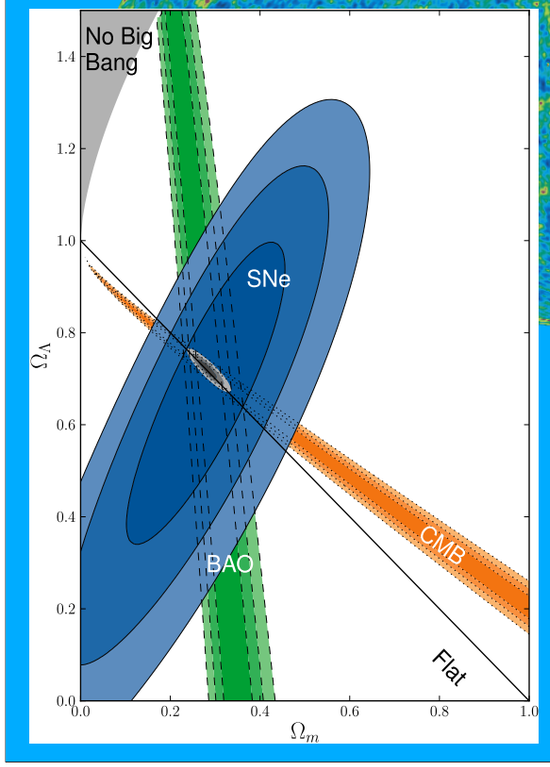
$$\text{Data} = H\ \Psi_\theta(R_j) / |\Psi_\theta(R_j)| = E(R_j)$$

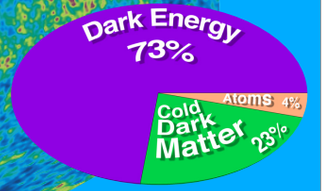
then varies parameters to minimize fluctuations in local E's away from eigen-energy.

The most accurate solutions to molecular energies of that time were based on a sloppy variational wave function!

The Universe

ΛCDM fit for cosmic microwave background radiation





Universe is flat, mostly unknown dark stuff

Six parameter ΛCDM model is sloppy fit to CMB. (Good parameter estimates from amazing data.)

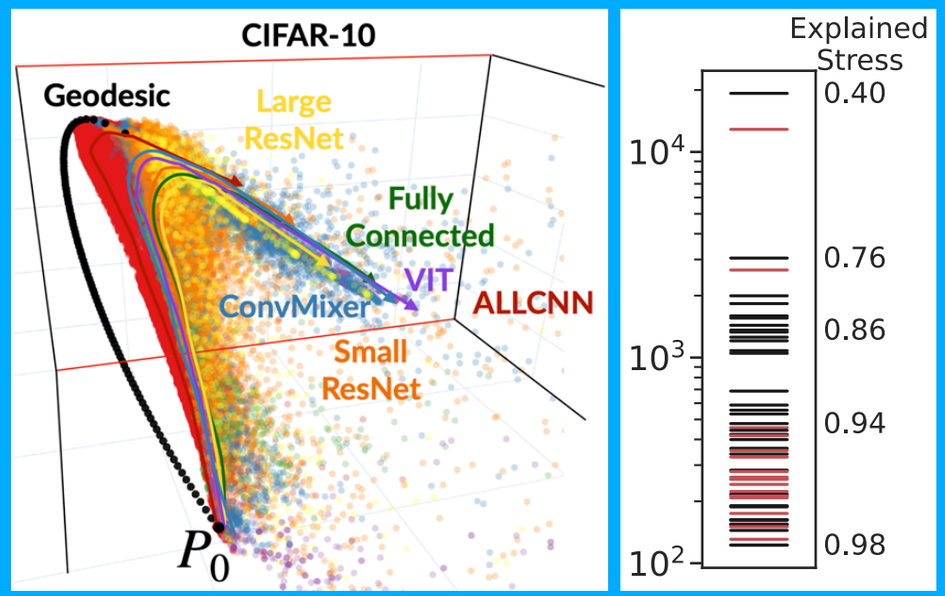
Katherine Quinn
Michael Niemack

Visualizing probabilistic models with Intensive Principal Component Analysis, Katherine N. Quinn, Colin B. Clement, Francesco De Bernardis, Michael D. Niemack, and James P. Sethna, Proceedings of the National Academy of Sciences 116, 13762-13767 (2019); https://sethna.lassp.cornell.edu/pubPDF/InPCA.pdf.

See also Data visualization could reveal nature of the universe (Cornell Chronicle article by Melanie Lefkowitz), and Algorithm to map universe, solve mysteries: Study (Times of India, June 26, 2019).

Deep neural network training

Mark Transtrum; Jialin Mao, Pratik Chaudhari, Itay Griniasty,
Han Kheng Teoh…

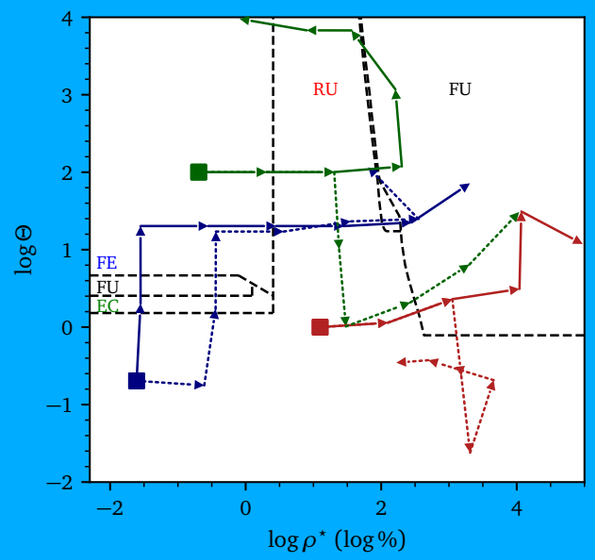


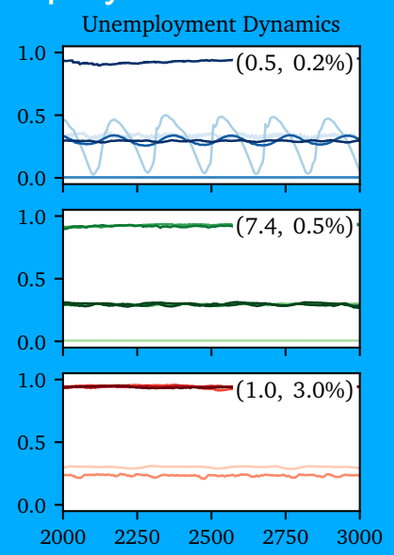
- Millions of parameters (weights); prediction space half-million dimensions
- Learning trajectories of all DNN architectures mostly contained in 3D!

The Training Process of Many Deep Networks Explores the Same Low-Dimensional Manifold, Jialin Mao, Itay Griniasty, Han Kheng Teoh, Rahul Ramesh, Rubing Yang, Mark K. Transtrum, James P. Sethna, and Pratik Chaudhari
https://arxiv.org/abs/2305.01604

Macroeconomics

Agent based models guiding response to pandemic and subsequent inflation / unemployment





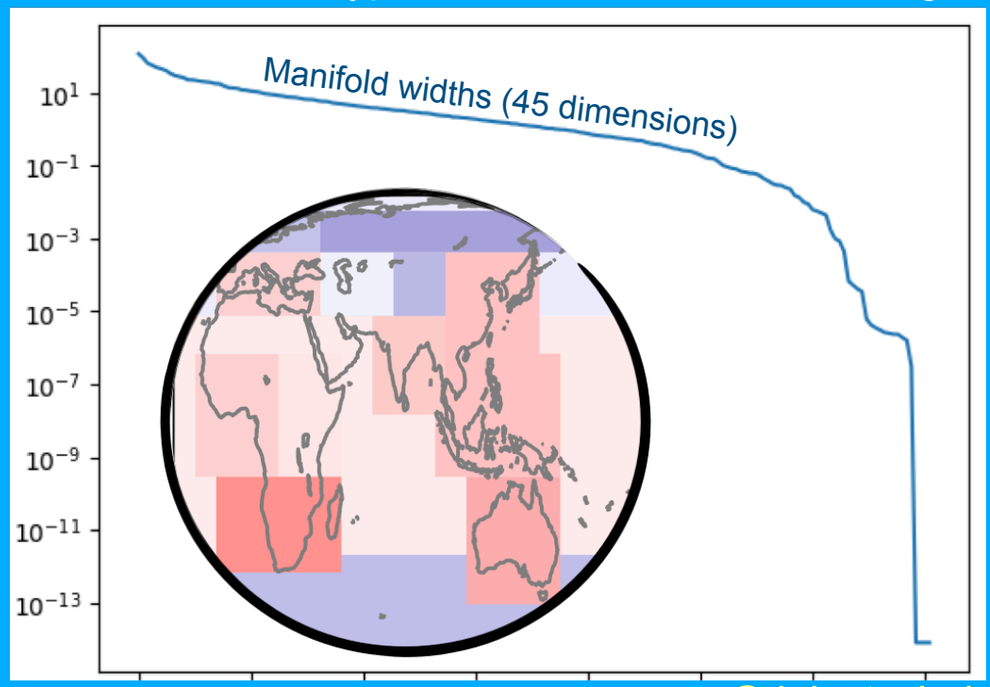
Stiff parameter directions in Jacobian guide search for phase boundaries

Naumann-Woleske, Knicker, Benzaquen, Zamponi, Bouchaud

Work in progress

Climate Change

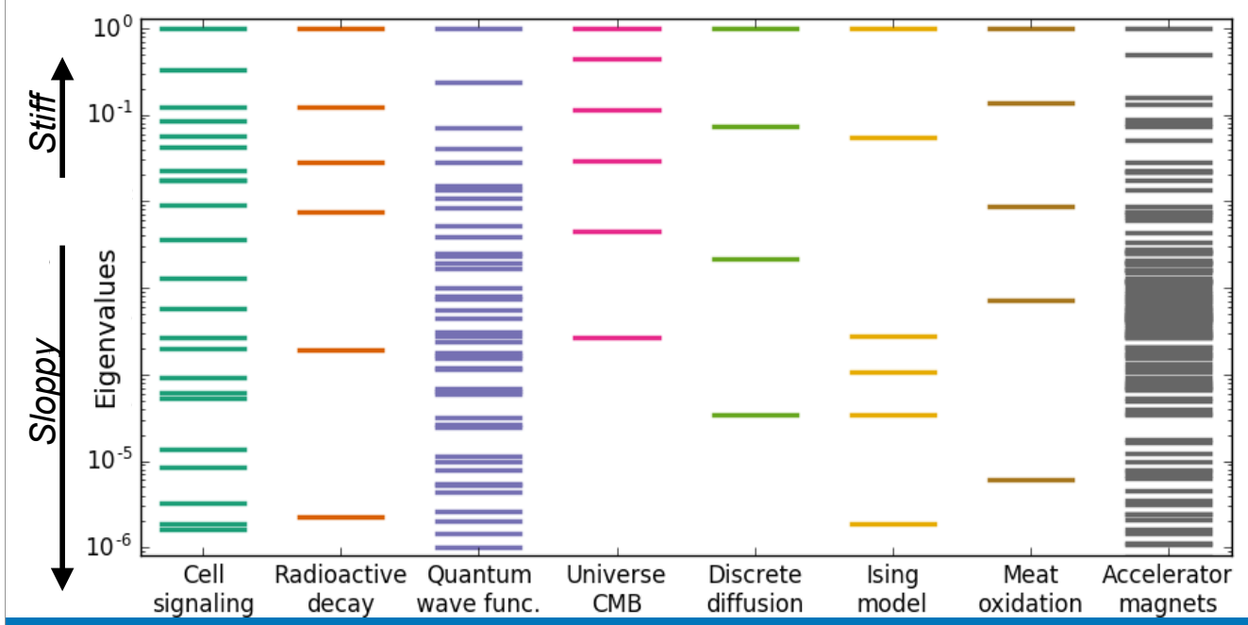
Initial work on hyperribbons in climate modeling



175 climate outputs, 45 parameters

Griniasty, Lehner, Pendergrass

Sloppy Universality

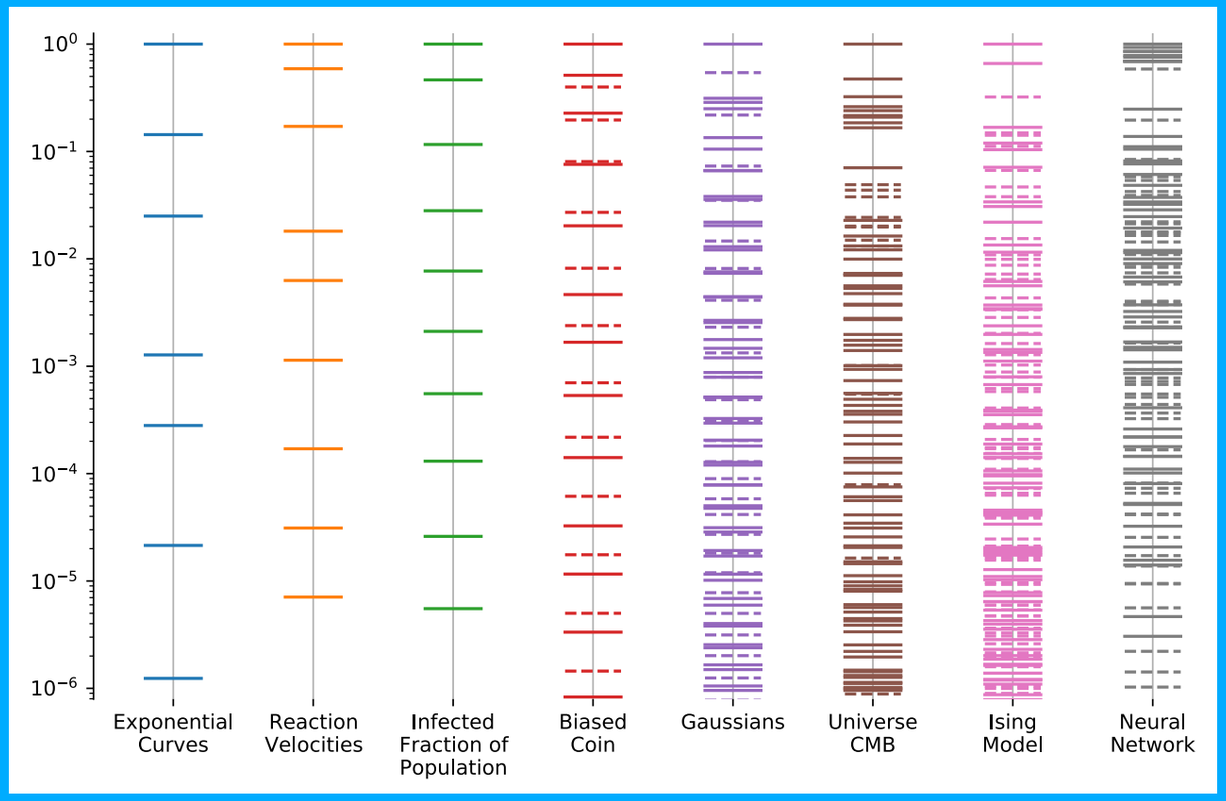


Enormous range of eigenvalues; Roughly equal density in log;
Observed in broad range of systems.

How can we use this to develop emergent laws?

We won't discuss meat oxidation here, and won't discuss quantum wave functions or accelerators again (but note that they are sloppy!) We shall return at length to discuss cell signaling and model reduction, the CMB and how to visualize its hyperribbon, the Ising model and renormalization group sloppiness. And those of you who do the homework will spend several exercises looking at our simple model of radioactive decay.

Hyperribbon widths in behavior



And towards the end of the module, we shall explore rigorous bounds — not describing sloppiness, but describing the hyperribbon of model predictions. And we shall discover new ways of visualizing probabilistic models.