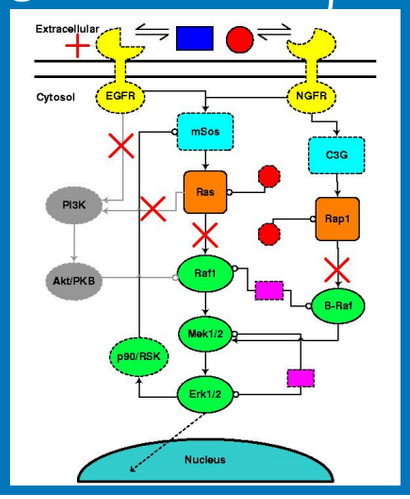
**Sloppy Models: Parameters are hard to find, but predictions are possible.**

**Implications for science**

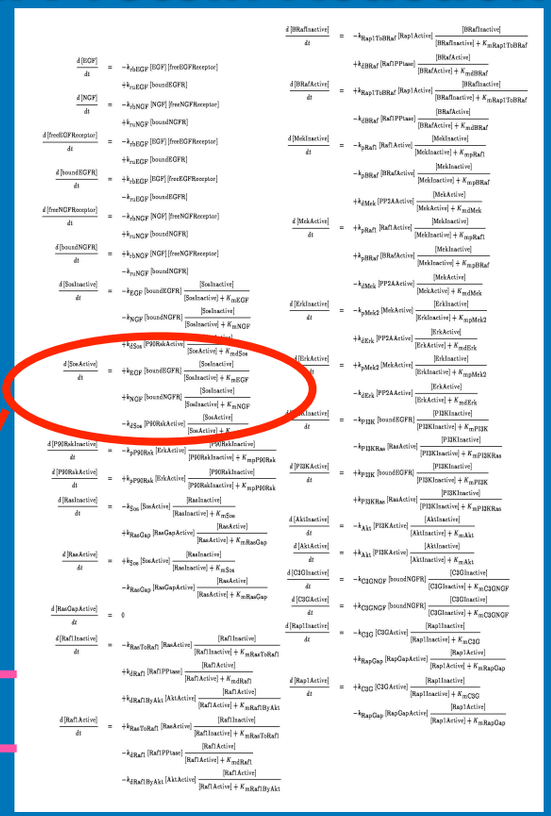
# Systems Biology: Cell Protein Reactions

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!

Remember our systems biology model from the first lecture?

# Numbers and notation

48 parameters $\theta_\alpha$ ($k_{EGF}$, $K_{mEGF}$…): $\Theta \in \mathbb{R}^{48}$

29 nonlinear differential equations

63 data points $d_i$

63 predictions $y_i(\Theta)$: behavior $\mathbf{y} \in \mathbb{R}^{63}$

Several *experimental control variables*

* time $t$ since stimulation by by EGF, NGF,
* strength of stimulation $u$,
* drug interventions $v$ like [LY] [ red Xs],

so $y_i = \mathbf{y}_\Theta(t_i, u_i, v_i)$ for experiment $i$.

Predictions of collective behavior

(It's not sloppy if one does separate experiments for each parameter. Also, such experiments are usually useless to predict collective behavior.)

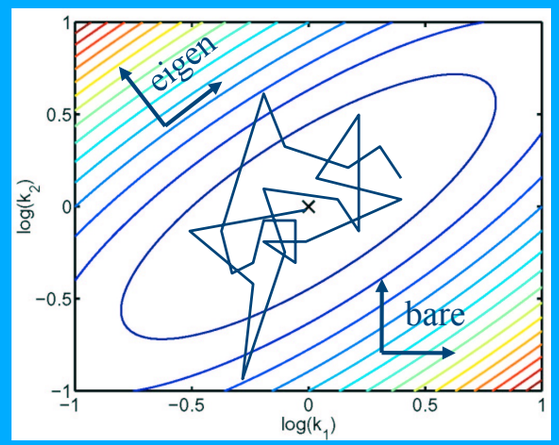
Let's set up some notation. We have parameters in a 48 dimensional real vector space, and behaviors in a 63 dimensional space of experiments. We view y as a map from parameter space to behavior space.

There are two other crucial ingredients. Our experiments have control variables, so the vector y_i should be viewed as y(t_i, u_i, v_i) for conditions t, u, and v.  And our experiments and predictions should be of collective behavior — no fair measuring each parameter independently! (Also

# Ensemble of Models

We want to consider not just minimum cost fits, but all parameter sets consistent with the available data. New level of abstraction: *statistical mechanics in model space*.

*Don't trust predictions that vary*




$$H_{ij} = \partial^2 C / \partial\theta_i \partial\theta_j$$

Cost is least-squares fit

$$C(\vec{\theta}) = \frac{1}{2}\sum_{i=1}^{N_D} \frac{(y(\vec{\theta}) - y_i)^2}{\sigma_i^2}$$

Boltzmann weights exp(-C/T)

$$\langle O \rangle = \frac{1}{N_E}\sum_{i=1}^{N_E} O(\vec{\theta}_i)$$

$$\sigma_O^2 = \left\langle O^2(\vec{\theta}) \right\rangle - \left\langle O(\vec{\theta}) \right\rangle^2$$

O is chemical concentration $y(t_i)$, or rate constant $\theta_n$…

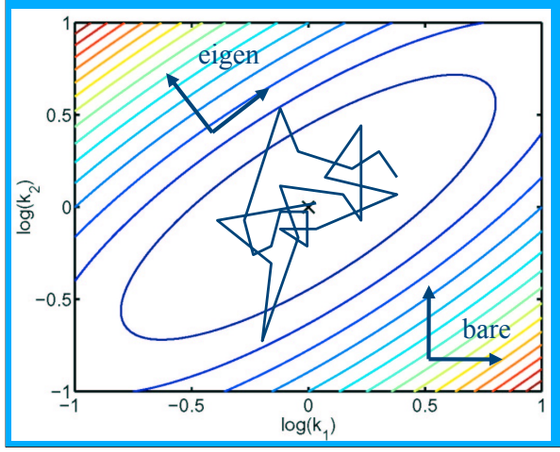
We do a Monte Carlo sampling of the Bayesian likelihood of different parameter sets, given the experimental data. We calculate error bars from our predictions y(theta) from the error bars sigma_i of the experimental data, by weighting the model predictions by exp(-C(theta)). We do a Monte-Carlo sampling of this nonlinear sum of squares.

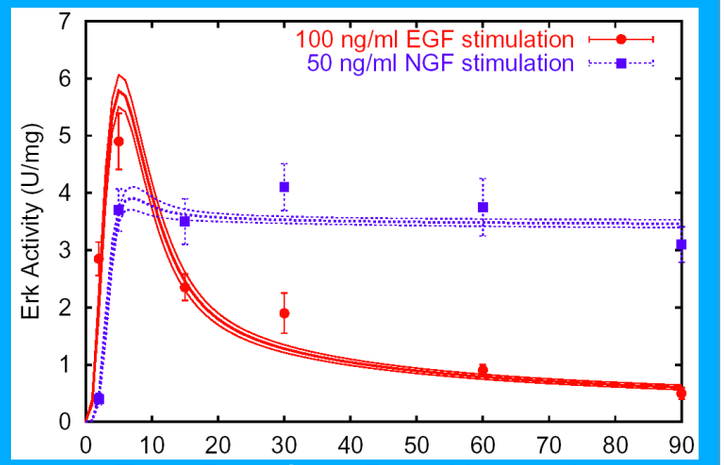
# 48 Parameter "Fit" to Data

Cost is "Energy"

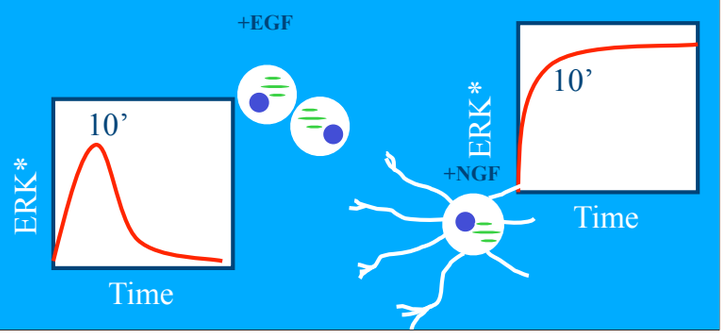
$$C(\Theta) = \frac{1}{2}\sum_{i=1}^{N}\frac{(y(\Theta) - d_i)^2}{\sigma_i^2}$$

Ensemble of Fits
Gives Error Bars



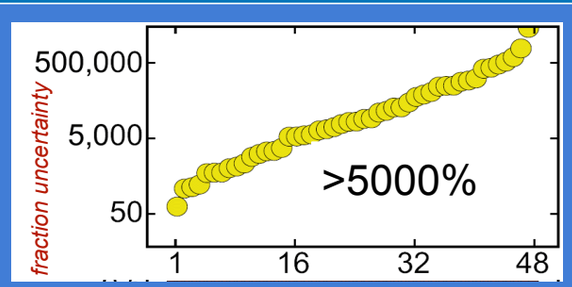

Error Bars from Data Uncertainty



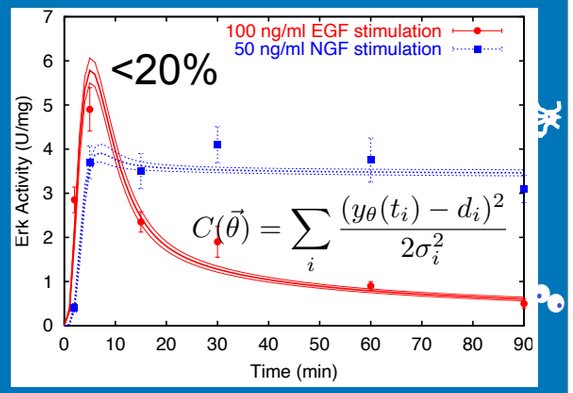

And we find that we can indeed make quite useful predictions.

# 48 Parameter Fit to Data

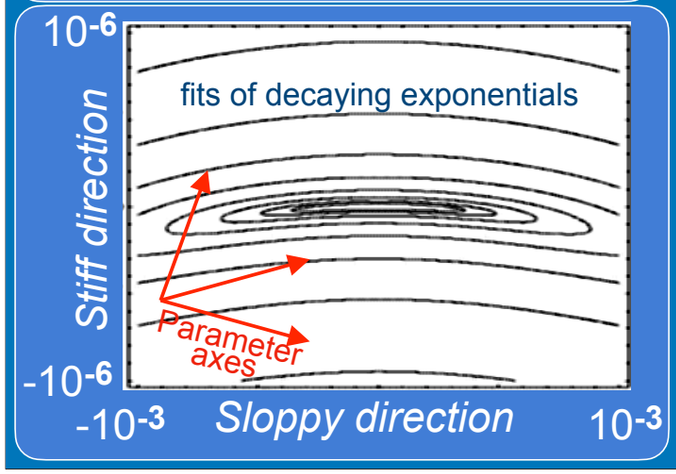
Kevin Brown, Josh Waterfall




*Parameters are unknown*




*Predictions are possible*



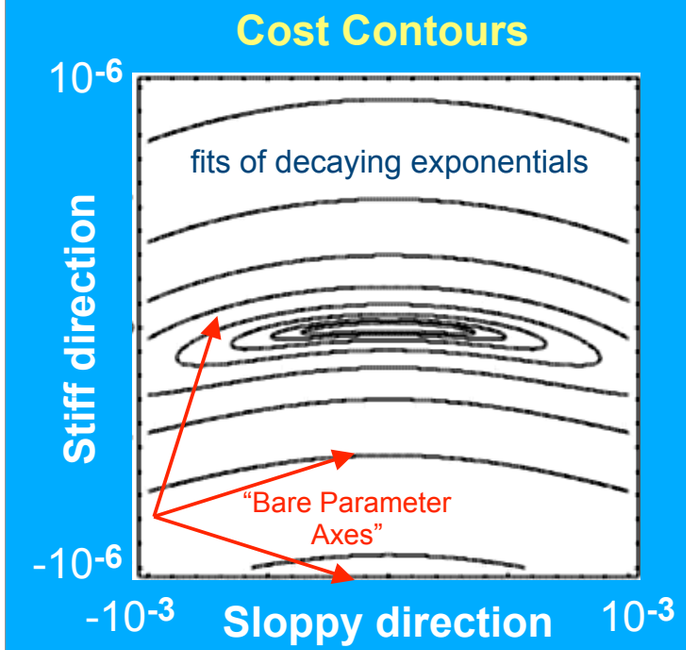

**The behavior is *ill-conditioned*: it depends only on a few stiff parameter combinations.**
**Eigenvalues of cost Hessian**

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

**span enormous range**

We do a Monte Carlo sampling of the Bayesian likelihood of different parameter sets, given the experimental data.
We cannot find 48 parameters from 63 data points. At upper left, the parameters are ordered according to their fractional uncertainty: note that the best known parameter has a 5000% error, and note the log scale.
We can make predictions. At right we show a prediction of the model, with errors ~20% given from the same ensemble. What's amazing is not that our predictions fit new experiments — it is that we can make predictions at all, given the enormous uncertainty in the parameters!
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.

# Parameter Indeterminacy and Sloppiness




**Note: Horizontal scale shrunk by 1000 times Aspect ratio = Human hair**

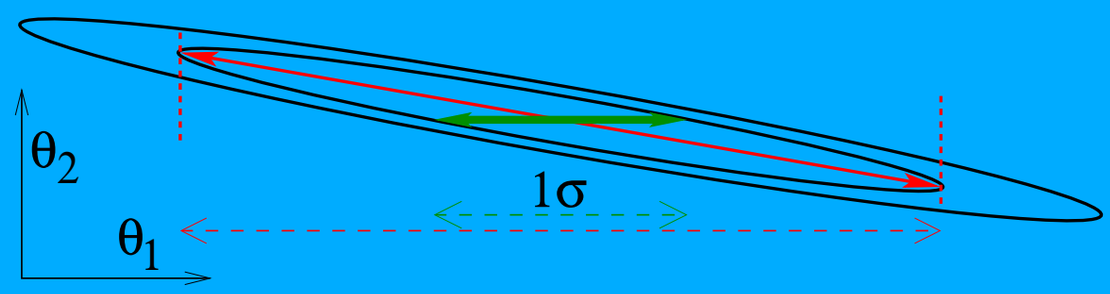
**48 parameter fits are sloppy: Many parameter sets give almost equally good fits**

**A few 'stiff' constrained directions allow model to remain predictive**

$$\mathcal{H}_{\alpha\beta} = \partial^2 C / \partial\theta_\alpha \partial\theta_\beta$$

So, the predictions as a function of parameters near the best fit form a really thin multidimensional hyper-ellipsoid. (This isn't actually true — many of the sloppy directions of the Hessian are associated with directions in which the parameter combination could reach out to infinity! More on this later.) Only a few parameter combinations control the behavior.

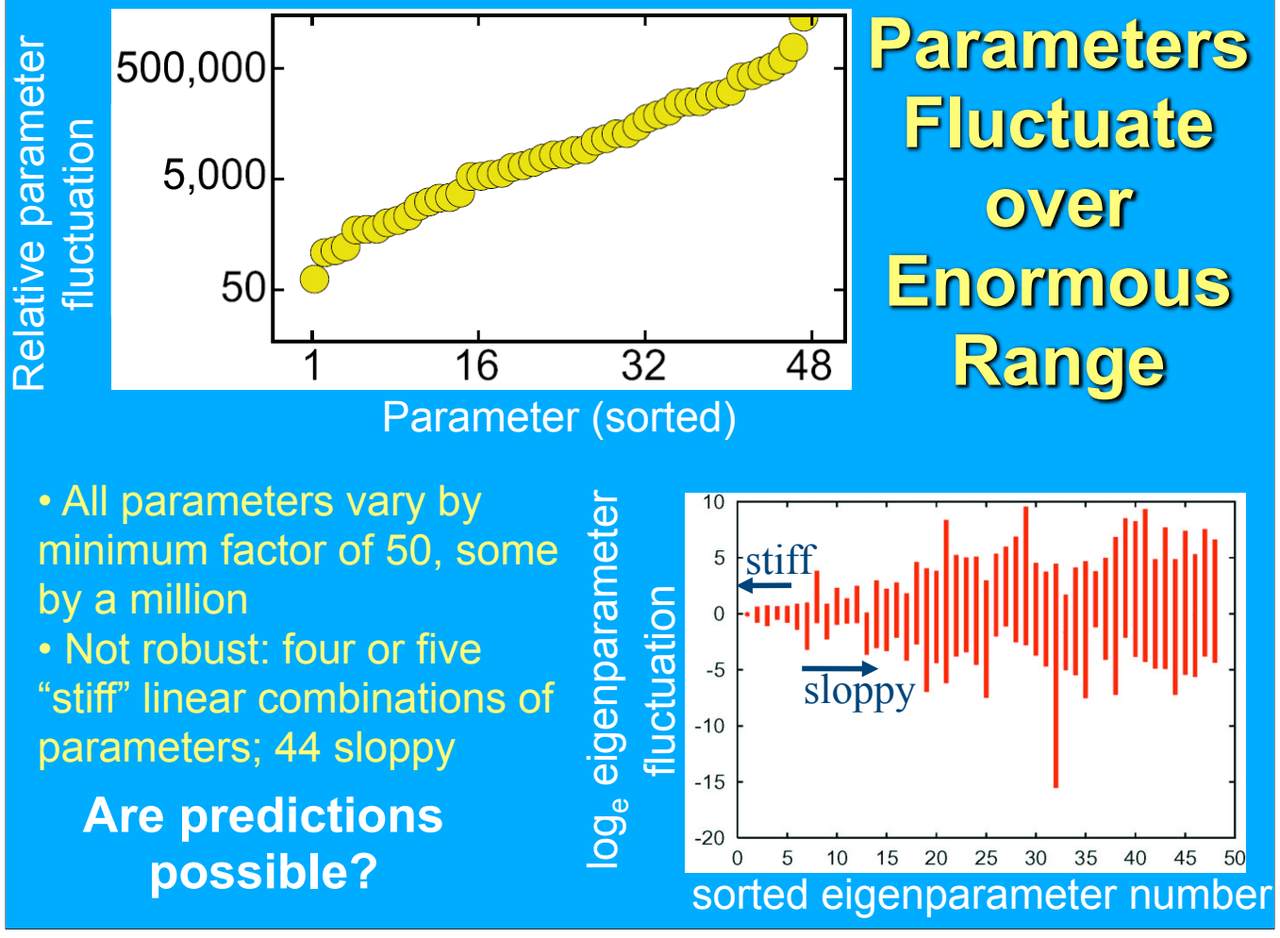
# Predictions are Possible



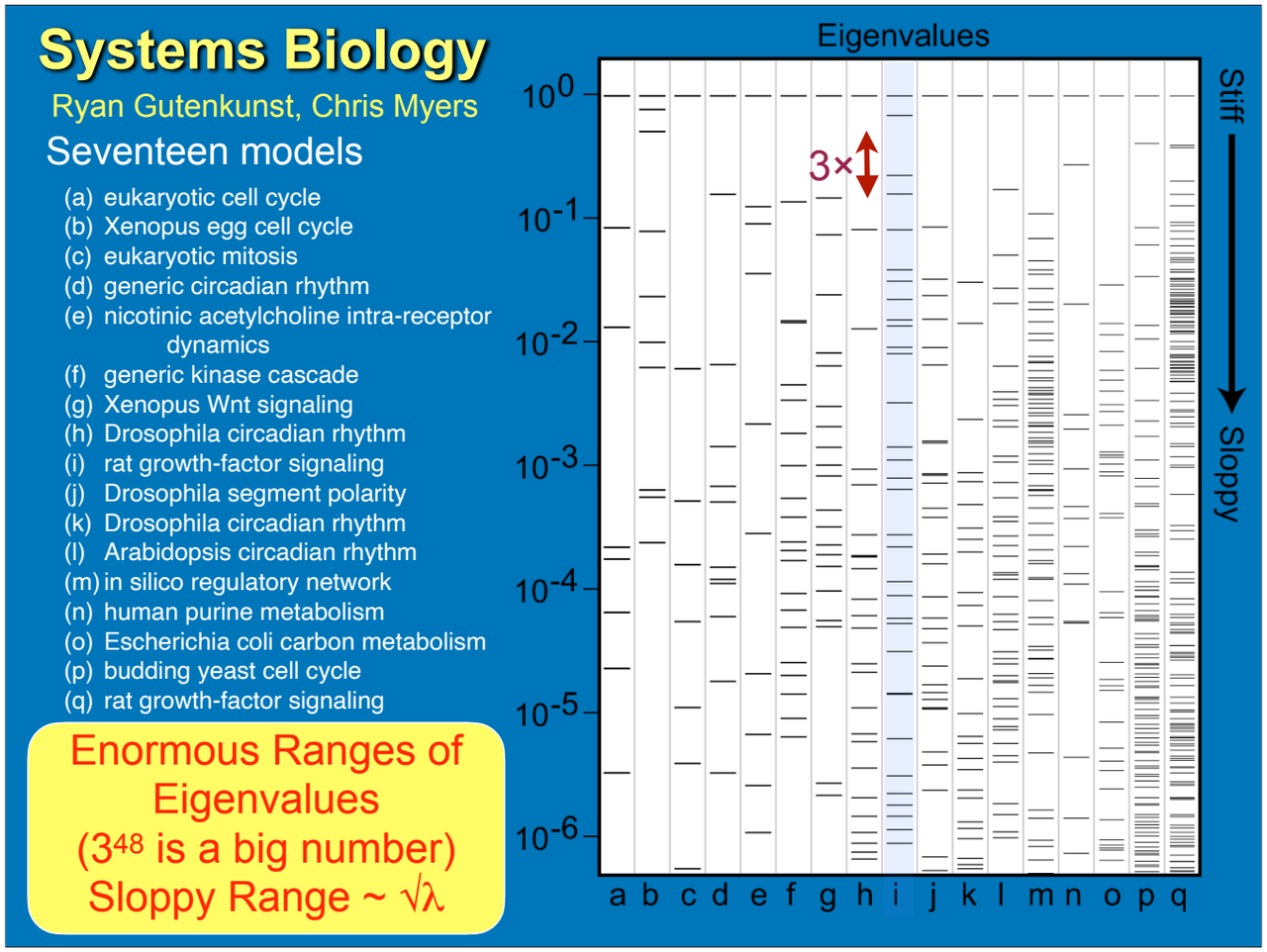

How can we make predictions when all parameters are undetermined?
— Almost all the parameters matter — shifting them by a factor of two ruins the fit. This is the range one can shift parameters without allowing others to vary: it is given by the diagonal elements of the cost Hessian, $(\sigma_\alpha^2)_{\text{fixed}} = 1/\mathscr{H}_{\alpha\alpha}$.
— All parameters are badly determined. The range that parameters can vary allowing others to compensate is given by the diagonal elements of the inverse Hessian: $(\sigma_\alpha^2)_{\text{collective}} = (\mathscr{H}^{-1})_{\alpha\alpha}$.

Linear algebra tells us that one standard deviation of behavior depends on whether the other parameters are known and fixed, or whether they can adjust to compensate.

Since our ellipsoids are extremely thin, and no parameters are parallel to the stiff axes, all our parameters are poorly determined.




But a few of the parameter combinations are well determined (lower left), because we choose them to move along the thin, stiff axes of the Hessian contours.



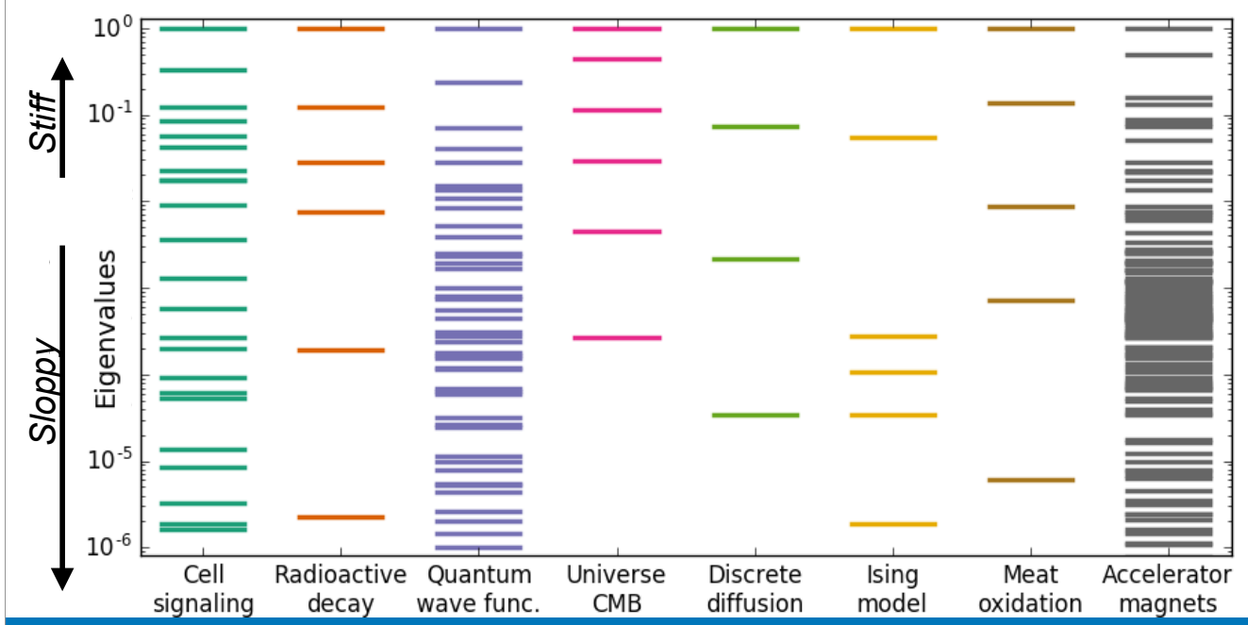

Ryan Gutenkunst, working with Chris Myers, wrote a wonderful Python software environment to interpret SBML files. (That's 'Systems Biology Markup Language', which was all the rage for a short time.) All kinds of biological systems turned out to be sloppy. We wrote some papers saying that biological 'robustness' (which then was a big exciting topic) was usually really sloppiness.

* "Universally Sloppy Parameter Sensitivities in Systems Biology", Ryan N. Gutenkunst, Joshua J. Waterfall, Fergal P. Casey, Kevin S. Brown, Christopher R. Myers, James P. Sethna, PLoS Comput Biol 3(10) e189 (2007). (PLoS, doi:10.1371/journal.pcbi.0030189), https://sethna.lassp.cornell.edu/pubPDF/SloppyEverywhere.pdf

* "Sloppiness, robustness, and evolvability in systems biology", Bryan C. Daniels, Yan-Jiun Chen, James P. Sethna, Ryan N. Gutenkunst, and Christopher R. Myers, Curr Opin Biotechnol 19, 389-395 (2008), doi:10.1016/j.copbio.2008.06.008, https://sethna.lassp.cornell.edu/pubPDF/SloppyRobust.pdf

We'll talk more about that later.

# Sloppy Universality




Enormous range of eigenvalues of $\mathcal{H}_{\alpha\beta}$; Roughly equal density in log;
Observed in broad range of systems.

***How can we use this to develop emergent laws?***

Many ask whether this is a biology-specific phenomenon — maybe we evolve to be sloppy? (Robustness was justified by those ideas.) Not so!

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. Only a few emergent parameters explain the behavior of systems at long length and time scales.

# Hessian for perfect data is given by the Jacobian

$$\mathscr{H}_{\alpha\beta} \approx \frac{\partial^2 C}{\partial\theta_\alpha \partial\theta_\beta} = \frac{\partial^2}{\partial\theta_\alpha \partial\theta_\beta}\left(\sum_i (y_i - d_i)^2/2\sigma^2\right)$$

$$= (1/\sigma^2)\frac{\partial}{\partial\theta_\alpha}\left(\sum_i (y_i - d_i)\frac{\partial y_i}{\partial\theta_\beta}\right)$$

$$= (1/\sigma^2)\left(\sum_i \frac{\partial y_i}{\partial\theta_\alpha}\frac{\partial y_i}{\partial\theta_\beta} + (y_i - d_i)\frac{\partial^2 y_i}{\partial\theta_\alpha \partial\theta_\beta}\right)$$

$$\approx (1/\sigma^2)(J^T J)_{\alpha\beta},$$

where

$$J_{i\alpha} = \frac{\partial y_i}{\partial\theta_\alpha}$$

- A good fit has y-d small.
- Second derivatives are expensive.
- Approximate Hessian is positive definite. (Sloppy minima often not perfect.)
- Quadratic approximation of distance between behaviors y as parameters are varied.
- Will be metric tensor on the model manifold.

Now, the Hessian for a 48 parameter model is a mess. But for a least-squares model that does a good job of fitting the data, the second derivative Hessian can be approximately written in terms of the first derivative J of the predictions! This Jacobian is the amount y twists and stretches parameter space into behavior space, and will be a focus of our attention. Also, if we want to measure the distances between nearby points on the model manifold, the approximate Hessian is exact, and J^T J becomes the metric on the model manifold.

# Fisher Information is the Metric for probabilistic models

**Fisher Information Matrix (FIM) measures distance**

$$d(\mathbf{\Theta}, \mathbf{\Theta} + \mathbf{\Delta}) \approx g_{\alpha\beta}\Delta_\alpha\Delta_\beta$$

$$g_{\alpha\beta}(\mathbf{\Theta}) = -\left\langle \frac{\partial^2 \log P(\mathbf{x})}{\partial\theta_\alpha \partial\theta_\beta} \right\rangle_{\mathbf{x}}$$

- Natural distance (metric tensor) between nearby probability distributions, parameterized by $\theta$ (see exercise Hellinger)
- For least-squares models (Gaussian $P(\mathbf{y}(\mathbf{\Theta}) - \mathbf{d})$, FIM distance equals $|\mathbf{y}(\mathbf{\Theta}_1) - \mathbf{y}(\mathbf{\Theta}_2)|/\sigma$
- Cramér–Rao bound: uncertainties in $\mathbf{\Theta}$ bounded below by inverse of Fisher Information matrix
- Machta (not me): Entropy cost for thermodynamic control is FIM path distance

What is the metric for more general, probabilistic models, like an N-spin Ising model (which predicts the probability of 2^N spin snapshots), or the Lambda-CDM model for the Big Bang (which predicts the probability of seeing a particular fluctuating temperature map of the microwave background radiation for our particular Universe)? In Exercise Hellinger, we'll motivate this Fisher Information Metric, which we'll use next to study the Ising model and the sloppy renormalization group.

# FIM for Stat Mech

## Deep links in local distances

Info Geom: $\mathrm{FIM} = g_{\alpha\beta}(\boldsymbol{\Theta}) = -\langle \partial_\alpha \partial_\beta \log \rho \rangle$ measures local dist

Stat Mech: $\log \rho = \beta \mathcal{F}$. Second derivatives of $\mathcal{F}$ are physical!

particle density $\rho = N/V$

isothermal compressibility $\kappa = -(1/V)(\partial V/\partial P)\Big|_T$

thermal expansion coefficient $\alpha = (1/V)(\partial V/\partial T)\Big|_P$

Specific heat at constant press $c_P = (T/N)(\partial S/\partial T)\Big|_P$

$$\mathrm{FIM} = g_{\alpha\beta}(\boldsymbol{\Theta}) = N \begin{pmatrix} -\kappa/\rho T & -\alpha/\rho T \\ -\alpha/\rho T & c_P/T^2 \end{pmatrix}$$

See exercise S2.3-5 (or text 6.20-22).

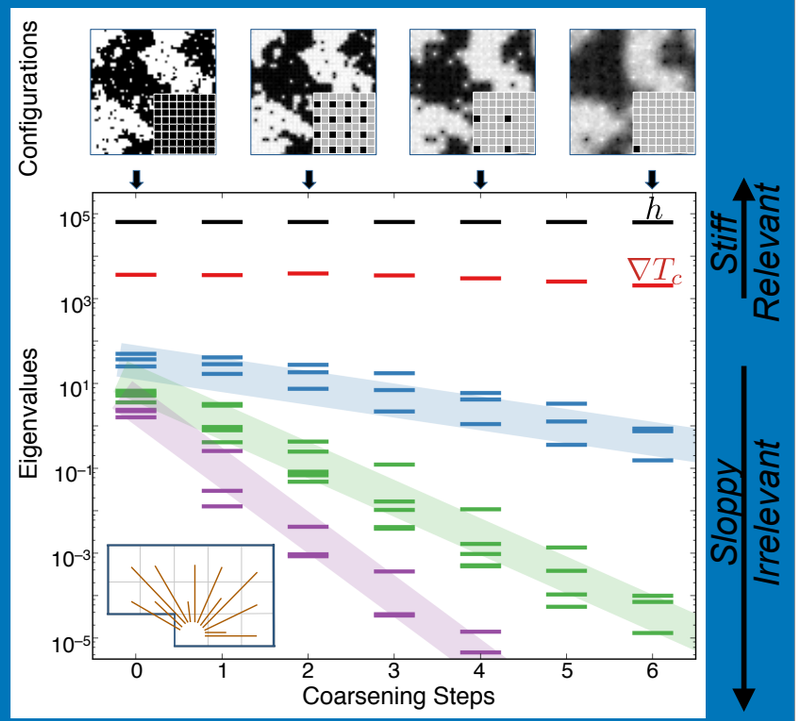
The natural metric in least-squares fits is the distance between the predicted data points and the measured ones, as measured in units of the error bars. In more general problems, the predictions are probability distributions P(x), and the natural metric is the Fisher information metric. This metric measures distances on the manifold of model predictions (to be discussed next).

# Physics: Sloppiness and Emergence

Ben Machta, Ricky Chachra, Mark Transtrum

Ising: long bonds
Diffusion: long hops
Details irrelevant on macroscale

***Emergence when only a few things matter***



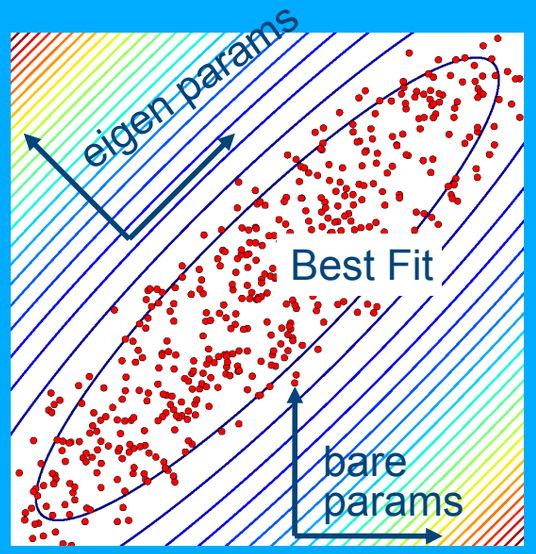

We studied Ising models using our sloppy model eigenvalues (here for the Fisher Information Matrix as $g_{\mu\nu}$). The Ising model predicts the probability of each spin configuration, and is not particularly sloppy (first column) as one varies the strength of long-range bonds (lower left). But if we only care about long wavelengths — with data on a coarse-grained grid — the relevant and irrelevant directions separate, and the model becomes sloppy. Our analysis thus captures key predictions of the renormalization group: it discovers the emergent relevant parameter combinations.

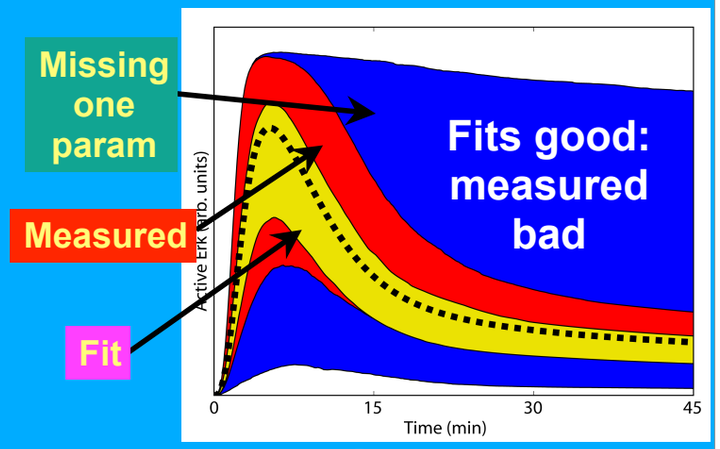
# *Implications of Sloppiness*

A variety of applications…

# Are rate constants useful?

## Fits vs. measurements




Monte Carlo (anharmonic)

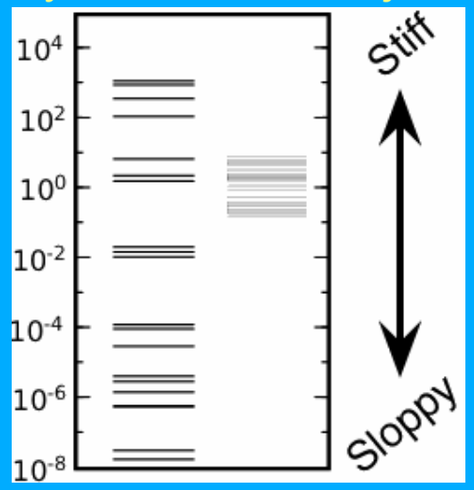
- Easy to Fit (14 expts); Measuring huge job (48 params, 25%)
- One missing parameter measurement = No predictivity
- Sloppy Directions = Enormous Fluctuations in Parameters
- Sloppy Directions often do not impinge on predictivity

"Universally Sloppy Parameter Sensitivities in Systems Biology", Ryan N. Gutenkunst, Joshua J. Waterfall, Fergal P. Casey, Kevin S. Brown, Christopher R. Myers, James P. Sethna, PLoS Comput Biol 3(10) e189 (2007). (PLoS, doi:10.1371/journal.pcbi.0030189), https://sethna.lassp.cornell.edu/pubPDF/SloppyEverywhere.pdf

# Parameter robustness and sloppiness

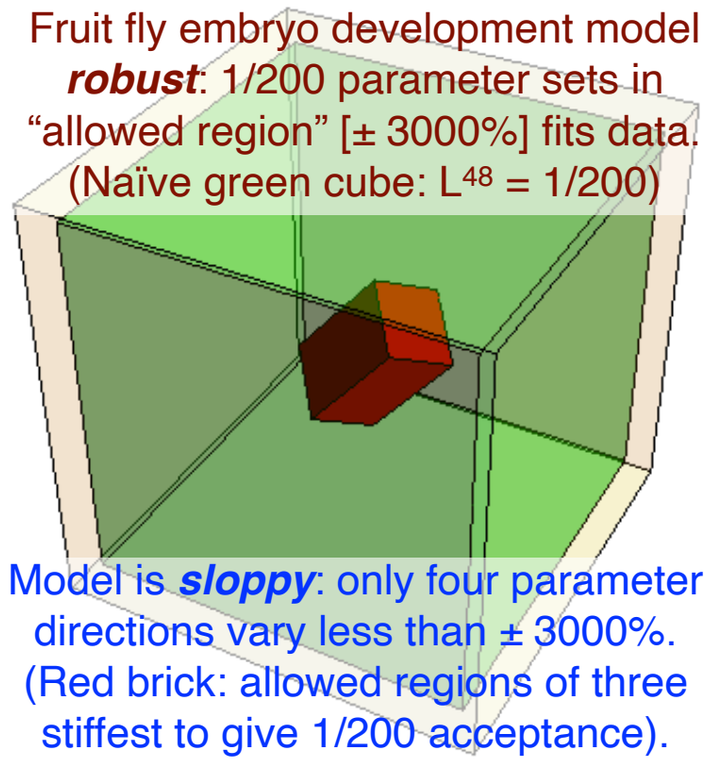
Do parameters matter at all?

Bryan Daniels, Yanjiun Chen, Ryan Gutenkunst, Chris Myers




Segment polarity model is sloppy (eigenvalues, left) and robust (PCA, phenotype-preserving in cube, right).

Fruit fly embryo development model ***robust***: 1/200 parameter sets in "allowed region" [± 3000%] fits data. (Naïve green cube: $L^{48}$ = 1/200)

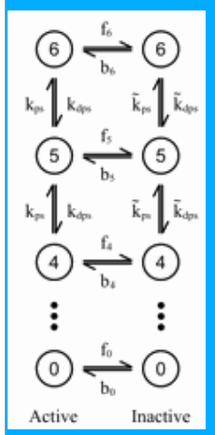


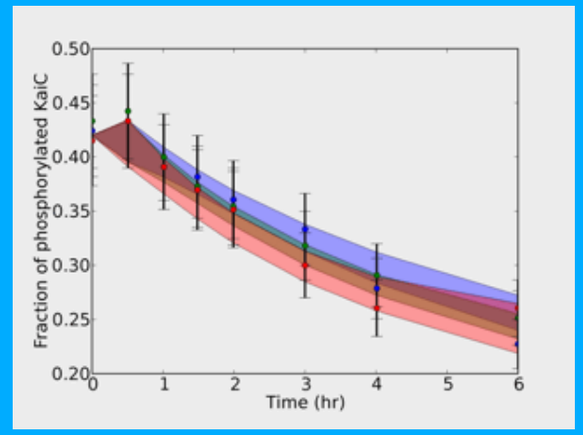
Model is ***sloppy***: only four parameter directions vary less than ± 3000%. (Red brick: allowed regions of three stiffest to give 1/200 acceptance).

* "Sloppiness, robustness, and evolvability in systems biology", Bryan C. Daniels, Yan-Jiun Chen, James P. Sethna, Ryan N. Gutenkunst, and Christopher R. Myers, Curr Opin Biotechnol 19, 389-395 (2008), doi:10.1016/j.copbio.2008.06.008, https://sethna.lassp.cornell.edu/pubPDF/SloppyRobust.pdf

# Environmental robustness and sloppiness

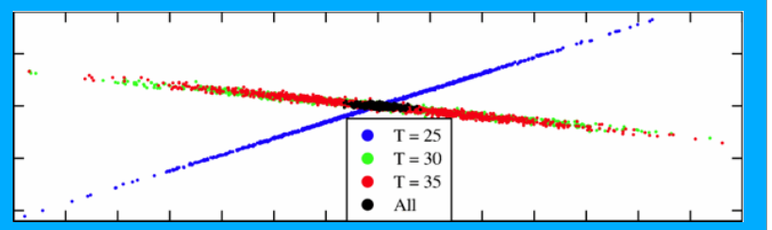
Circadian Rhythms




Bacteria know the time of day! How do they keep their clocks on time in the cold? All reaction rates exponentially dependent on temperature! Delicate cancellation?

Sloppiness facilitates finding parameter values for robust response to environmental change.

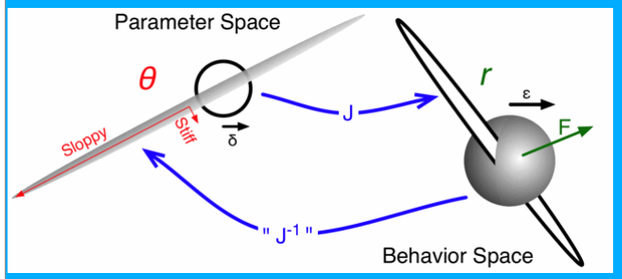



Three sloppy directions, 18 rates exponentially dependent on temperature. Three stiff directions at each temperature? 18 – 3 (25°C) – 3 (30°C) – 3 (35°C) = 9 dimensional robust parameter space

* "Sloppiness, robustness, and evolvability in systems biology", Bryan C. Daniels, Yan-Jiun Chen, James P. Sethna, Ryan N. Gutenkunst, and Christopher R. Myers, Curr Opin Biotechnol 19, 389-395 (2008), doi:10.1016/j.copbio.2008.06.008, https://sethna.lassp.cornell.edu/pubPDF/SloppyRobust.pdf

# Evolvability, robustness, and sloppiness

If it's robust, can it evolve?




How can evolution proceed if mutating the parameters doesn't matter? More robust, less evolvable!

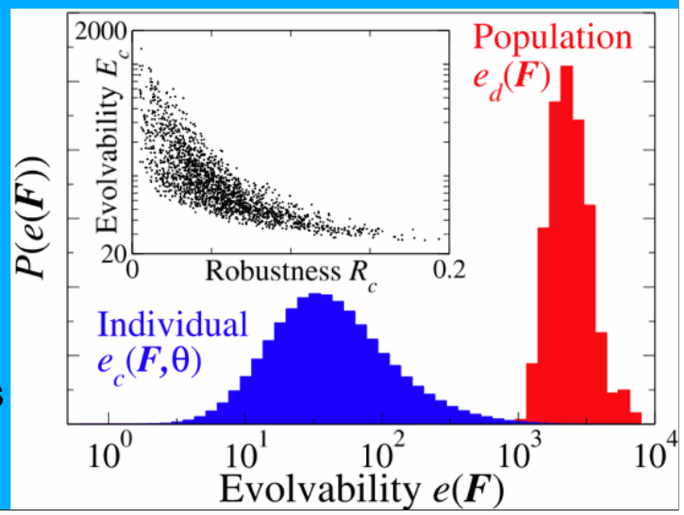
Evolutionary force $F$

*Sloppy signal transduction model*

Individual evolvabilities $e_c(F)$ within phenotype

Population evolvability $e_d$ within phenotype

[Sloppy neutral spaces allow species to explore large ranges of parameters]

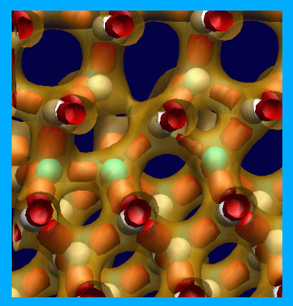



Read up on this!

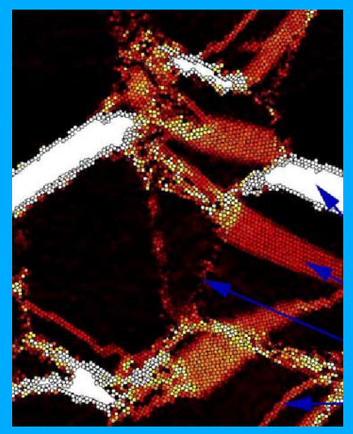
# E. Bayesian Errors for Atoms

'Sloppy Model' Approach to Error Estimation of Interatomic Potentials

Søren Frederiksen, Karsten W. Jacobsen, Kevin Brown, JPS



Quantum Electronic Structure (Si) 90 atoms (Mo) (Arias)



Atomistic potential 820,000 Mo atoms (Jacobsen, Schiøtz)

**Interatomic Potentials** $V(\mathbf{r}_1,\mathbf{r}_2,\ldots)$

- Fast to compute
- Limit $m_e/M \rightarrow 0$ justified
- Guess functional form
  Pair potential $\sum V(\mathbf{r}_i-\mathbf{r}_j)$ poor
  Bond angle dependence
  Coordination dependence
- Fit to experiment (old)
- Fit to forces from electronic structure calculations (new)

**17 Parameter Fit**

"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.

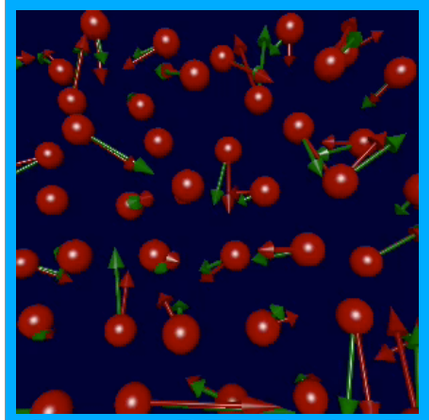
# E. Interatomic Potential Error Bars

## Ensemble of Acceptable Fits to Data

Not ***transferable***

Unknown errors

- 3% elastic constant
- 10% forces
- 100% fcc-bcc, dislocation core



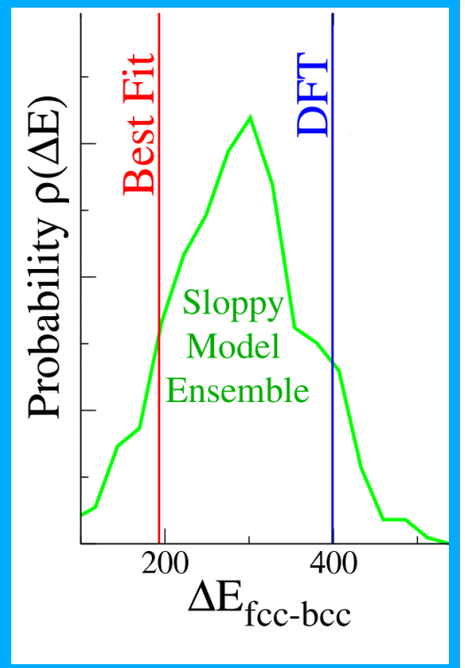
Green = DFT, Red = Fits

Best fit is ***sloppy***: ensemble of fits that aren't much worse than best fit.

**Ensemble in Model Space!**

$T_0$ set by equipartition energy = best cost

Error Bars from quality of best fit



# Sloppy Molybdenum: Does it Work?

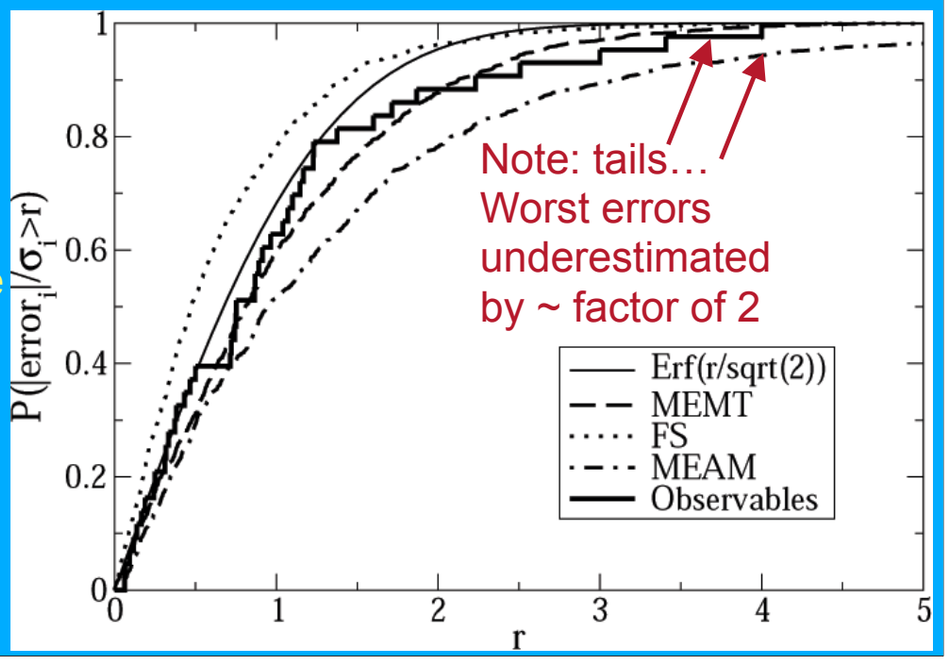
## Estimating *Systematic* Errors

Bayesian error $\sigma_i$ gives total error if ratio $r = \text{error}_i/\sigma_i$ distributed as a Gaussian: cumulative distribution $P(r)=Erf(r/\sqrt{2})$

Three potentials
- Force errors
- Elastic moduli
- Surfaces
- Structural
- Dislocation core
- $7\% < \sigma_i < 200\%$

"Sloppy model" systematic error most of total

~2 << 200%/7%

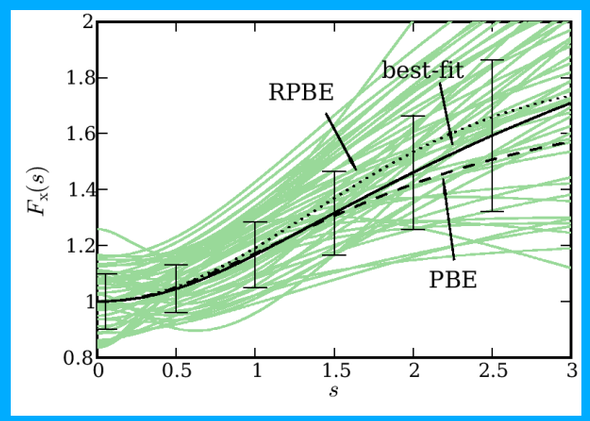



"Bayesian Error Estimation in Density Functional Theory", J. J. Mortensen, K. Kaasbjerg, S. L. Frederiksen, J. K. Norskov, James P. Sethna, K. W. Jacobsen, Phys. Rev. Letters 95, 216401 (2005), https://sethna.lassp.cornell.edu/pubPDF/SloppyMo.pdf.

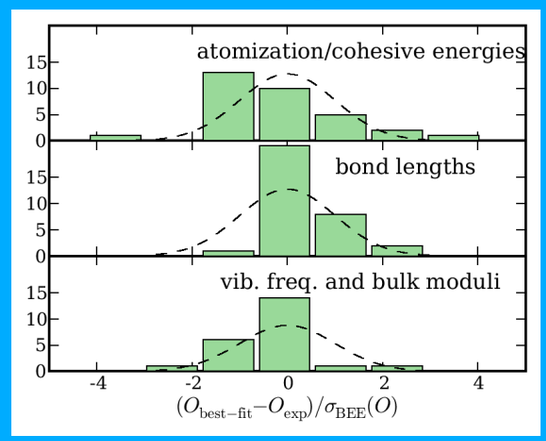
# Systematic Error Estimates for DFT

## GGA-DFT as Multiparameter Fit?

J. J. Mortensen, K. Kaasbjerg, S. L. Frederiksen,
J. K. Nørskov, JPS, K. W. Jacobsen,
(Anja Tuftelund, Vivien Petzold, Thomas Bligaard)




Enhancement factor $F_x(s)$
in the exchange energy $E_x$
**Large fluctuations**




Actual error / predicted error
**Deviation from experiment well described by ensemble!**

"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/SloppyDFT.pdf