Auxiliary Dynamical Systems for Bayesian Inference
Second Order Langevin Markov Chain Monte Carlo

Thomas A. Catanach

California Institute of Technology

July 28, 2016

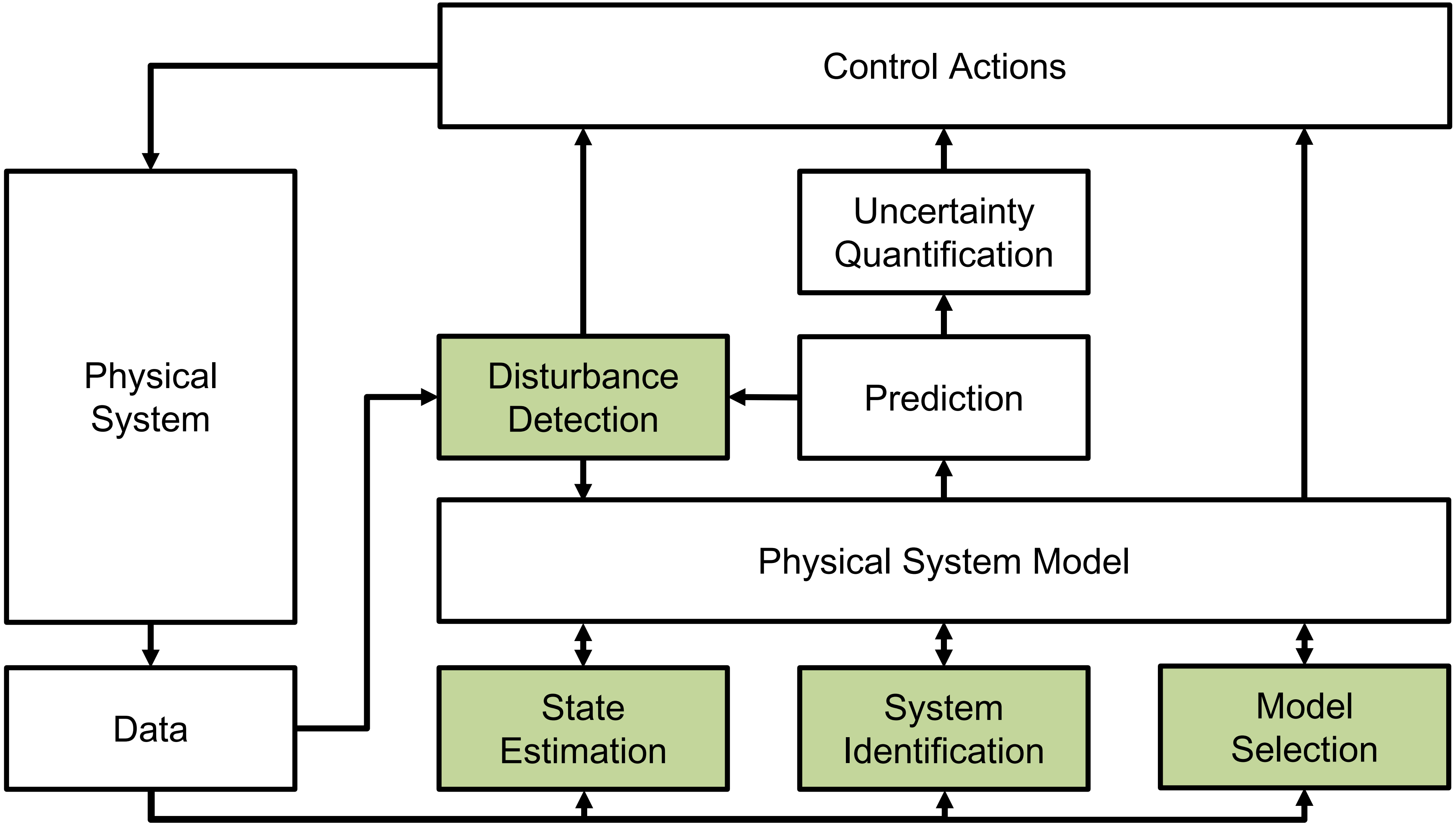
Motivation

- Bayesian methods for identification and estimation are critical to the robust system analysis
- The computational intensity of MCMC sampling methods is the main bottleneck for Bayesian inference

Goal:

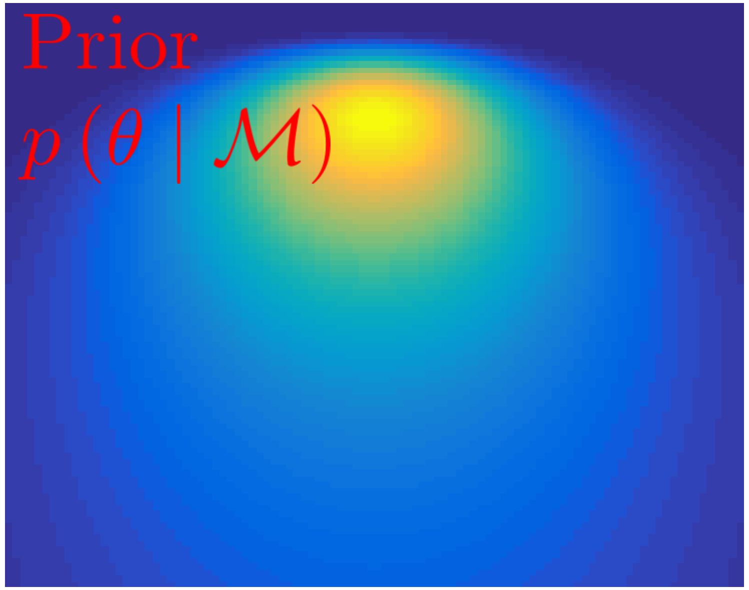
Develop new algorithms based upon dynamical systems to better sample probability distribution

Bayesian Inference for Physical Systems



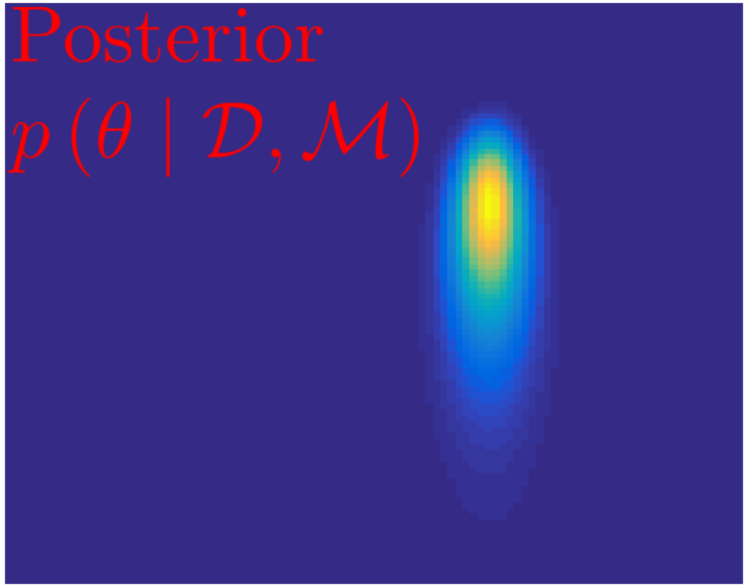
The Bayesian Inference Problem

Observations: $\mathcal{D}$



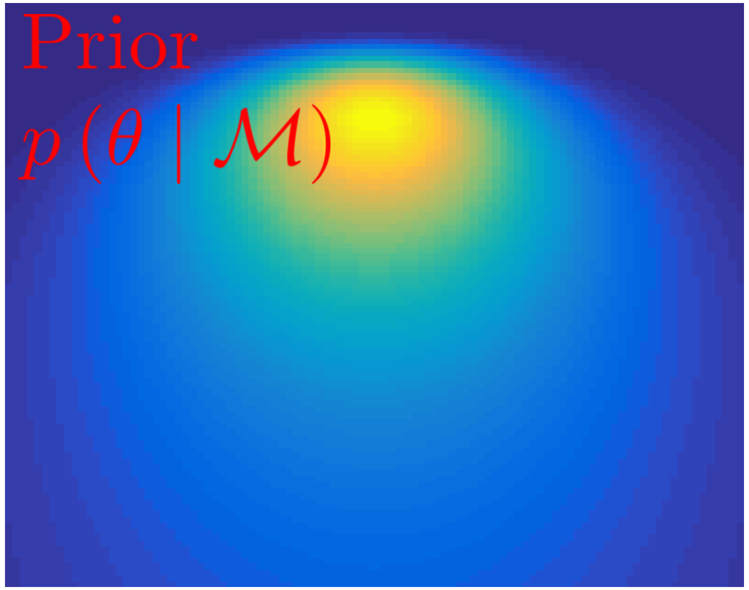
Bayes' Theorem

$$p(\theta \mid \mathcal{D}, \mathcal{M}) = \frac{p(\mathcal{D}|\theta,\mathcal{M})p(\theta|\mathcal{M})}{p(\mathcal{D}|\mathcal{M})}$$



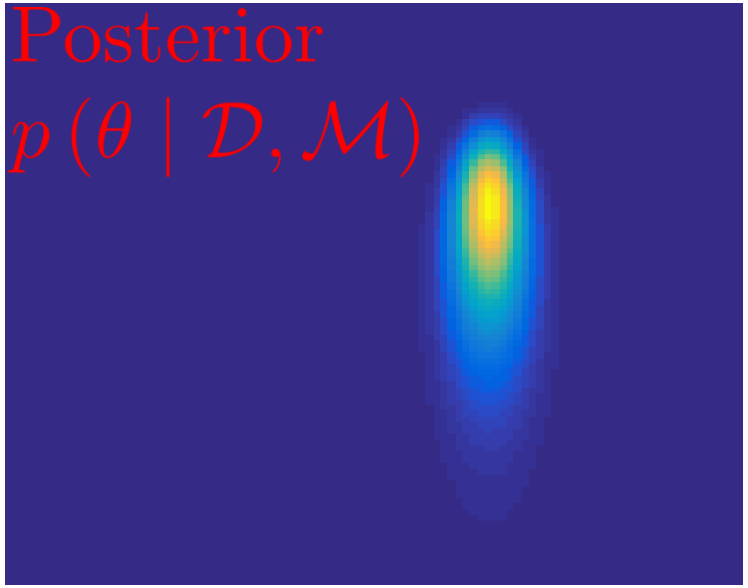
The Bayesian Inference Problem

Observations: $\mathcal{D}$



Bayes' Theorem

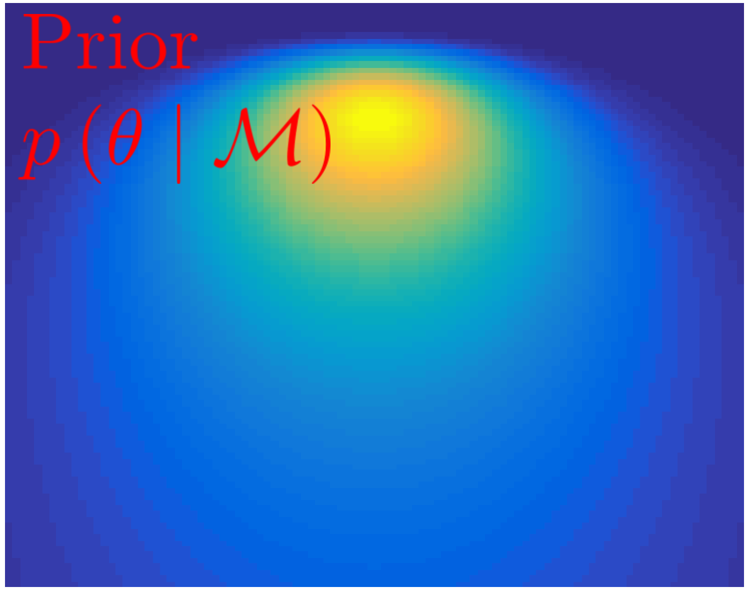
$$p\left(\theta \mid \mathcal{D}, \mathcal{M}\right) = \frac{p(\mathcal{D}|\theta,\mathcal{M})p(\theta|\mathcal{M})}{p(\mathcal{D}|\mathcal{M})}$$



$$p\left(\mathcal{D} \mid \mathcal{M}\right) = \underbrace{\int p\left(\mathcal{D} \mid \theta, \mathcal{M}\right) p\left(\theta \mid \mathcal{M}\right) d\theta}_{\text{Intractable}}$$

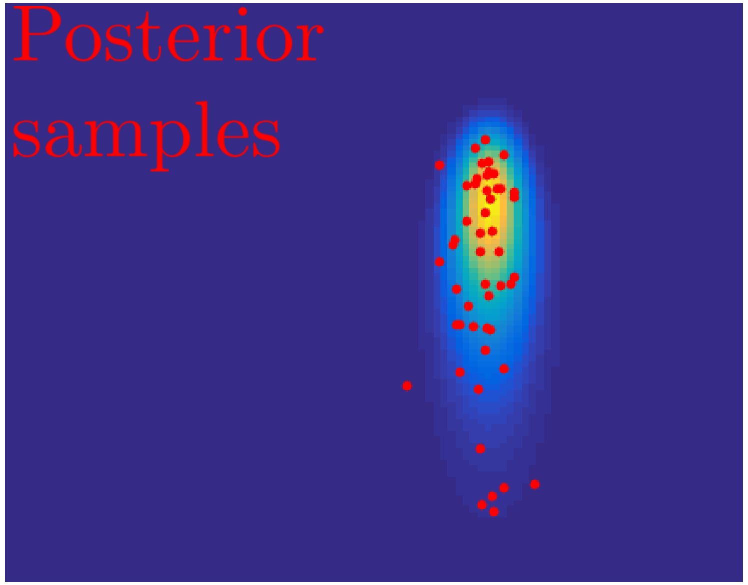
The Bayesian Inference Problem

Observations: $\mathcal{D}$



Bayes' Theorem

$$p\left(\theta \mid \mathcal{D}, \mathcal{M}\right) = \frac{p(\mathcal{D}|\theta,\mathcal{M})p(\theta|\mathcal{M})}{p(\mathcal{D}|\mathcal{M})}$$

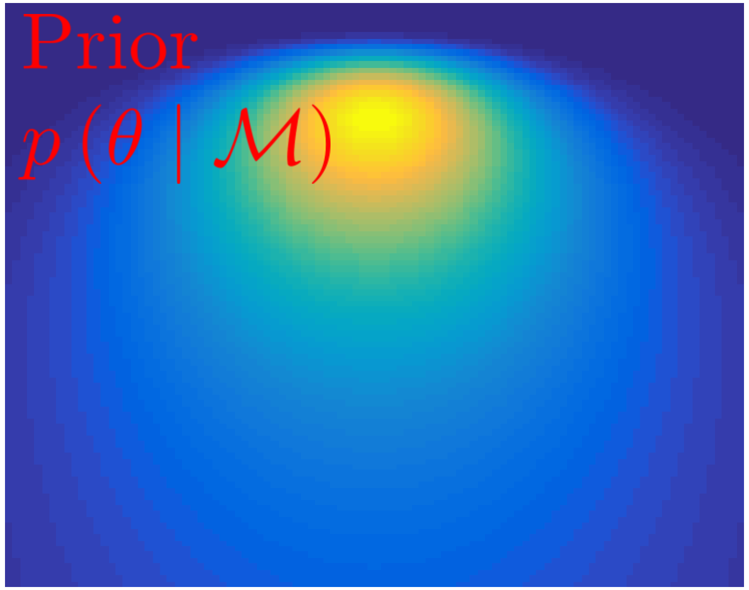


Posterior Estimation:

$$\mathbb{E}\left[g\left(\theta\right) \mid \mathcal{D}, \mathcal{M}\right] = \int g\left(\theta\right) p\left(\theta \mid \mathcal{D}, \mathcal{M}\right) d\theta \approx \frac{1}{N}\sum_{i=1}^{N} g\left(\theta_i\right)$$

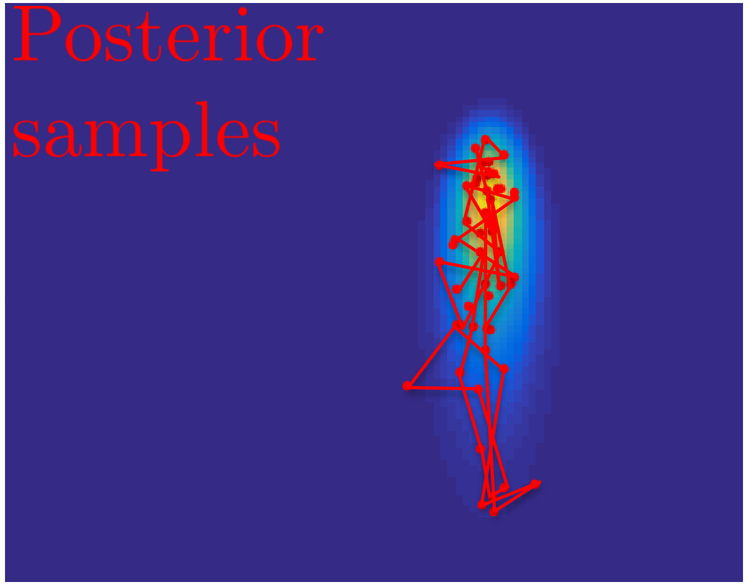
The Bayesian Inference Problem

Observations: $\mathcal{D}$



Bayes' Theorem

$$p\left(\theta \mid \mathcal{D}, \mathcal{M}\right) = \frac{p(\mathcal{D}|\theta,\mathcal{M})p(\theta|\mathcal{M})}{p(\mathcal{D}|\mathcal{M})}$$



Posterior Estimation: $$\mathbb{E}\left[g\left(\theta\right) \mid \mathcal{D}, \mathcal{M}\right] = \int g\left(\theta\right) p\left(\theta \mid \mathcal{D}, \mathcal{M}\right) d\theta \approx \frac{1}{N}\sum_{i=1}^{N} g\left(\theta_i\right)$$

Effective Number of Samples: $$ESS\left[g\left(\theta_{1:N}\right)\right] = \frac{N}{1 + 2\sum_{k=1}^{N} \rho_k\left(g\left(\theta_{1:N}\right)\right)}$$

Markov Chain Monte Carlo

- Design a Markov Chain kernel which:
  - Minimizes the convergence (burn-in) time to the stationary distribution
  - Minimizes the time correlation when sampling the stationary distribution
- Sufficient conditions for stationary distribution
  - Detailed Balance → Reversibility → Existence

$$\pi(\theta) K(\theta' \mid \theta) = \pi(\theta') K(\theta \mid \theta')$$

  - Ergodicity→ Uniqueness

Sampling using Auxiliary Dynamical Systems

- We can equate the joint probability and a Hamiltonian

Euclidean HMC: $H\left(\theta,p\right)=-\log\pi\left(\theta\right)+\frac{1}{2}p^{T}M^{-1}p$

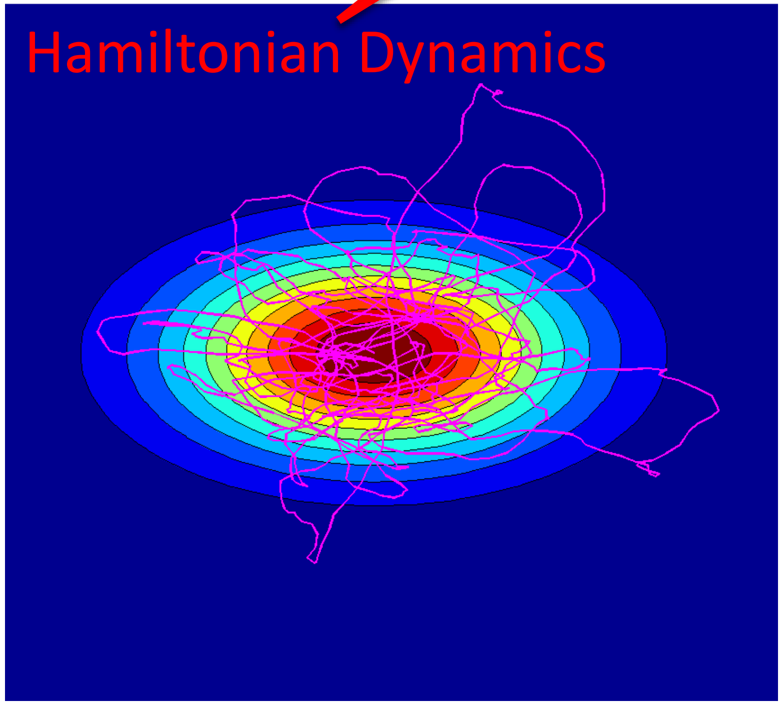
Riemannian HMC: $H\left(\theta,p\right)=-\log\pi\left(\theta\right)+\frac{1}{2}\log\left|\mathbf{G}\left(\theta\right)\right|+\frac{1}{2}p^{T}\mathbf{G}\left(\theta\right)^{-1}p$

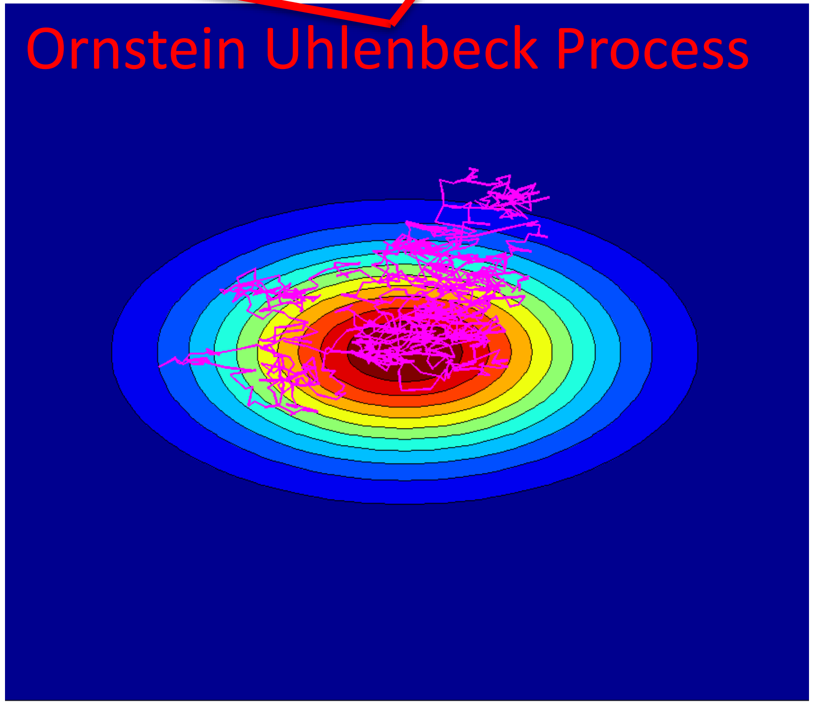
Posterior: $\Pi\left(\theta,p\right)\propto\exp\left(-H\left(\theta,p\right)\right)$

- Use the corresponding Hamiltonian dynamical system as an efficient proposal distribution

Second Order Langevin SDE

$$\begin{bmatrix} \dot{\theta} \\ \dot{p} \end{bmatrix} = \begin{bmatrix} 0 & \mathbf{I} \\ -\mathbf{I} & -\mathbf{C}(\theta) \end{bmatrix} \begin{bmatrix} \frac{\partial H(\theta,p)}{\partial \theta} \\ \frac{\partial H(\theta,p)}{\partial p} \end{bmatrix} + \begin{bmatrix} 0 \\ \sqrt{2\mathbf{C}(\theta)} \end{bmatrix} w$$





Numerical Implementation

Strang Splitting

$$(\theta_{k+1}, p_{k+1}) = \psi_{t_k+h, t_k+h/2} \circ \Theta_h \circ \psi_{t_k+h/2, t_k} (\theta_k, p_k)$$

$\psi_{\frac{h}{2}}$ Stochastic Integrator for Ornstein-Uhlenbeck Process

Θ_h Deterministic Integrator for Hamilton's Equations

Metropolis Step

$$(\theta_{n+1}, p_{n+1}) = \begin{cases} \left(\theta^*_{n+1}, p^*_{n+1}\right), \text{ if } \zeta_n < \alpha\left(\theta_n, p_n, \theta^*_{n+1}, p^*_{n+1}\right) \\ \left(\theta_n, -p_n\right), \text{ otherwise} \end{cases}$$

Flipping momentum for reversibility

SOL-MC Design

- Gaussian Posterior → Linear System

$$\begin{bmatrix} \dot{\theta} \\ \dot{p} \end{bmatrix} = \begin{bmatrix} 0 & \mathbf{M}^{-1} \\ -\mathbf{G} & -\mathbf{CM}^{-1} \end{bmatrix} \begin{bmatrix} \theta \\ p \end{bmatrix} + \begin{bmatrix} 0 \\ \sqrt{2\mathbf{C}} \end{bmatrix} w$$

  – We can optimize by choosing **C** to minimize the largest eigenvalue and by aligning **M** to **G**

- Non-Linear Problems
  – **C**(θ) can vary so we can locally linearize at θ to find the best C
  – Changing **M** changes the Hamiltonian so it is best to find using test runs to estimate the global structure of the posterior

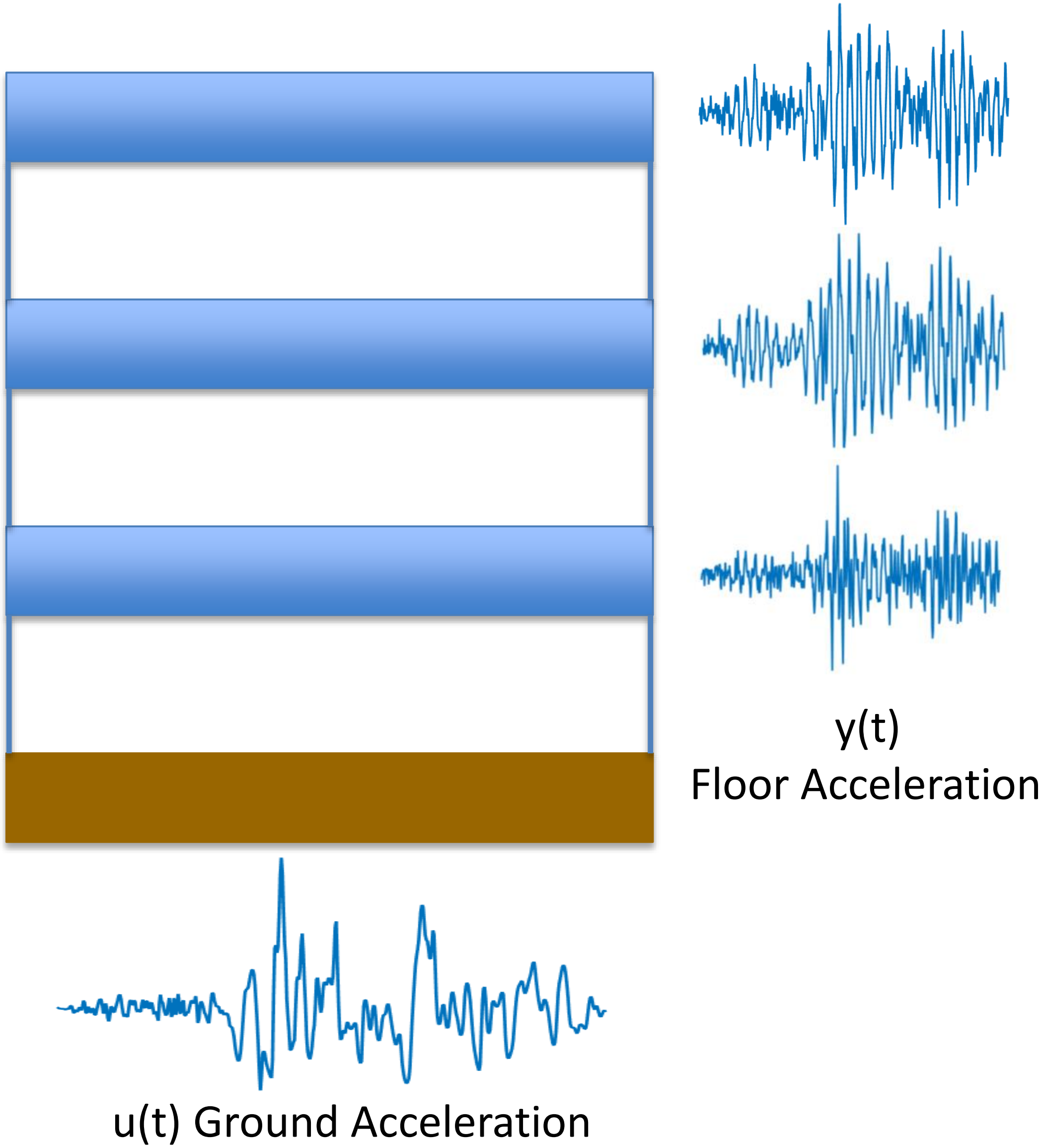
System Identification: Hysteretic Structure

Non-Linear Dynamical System

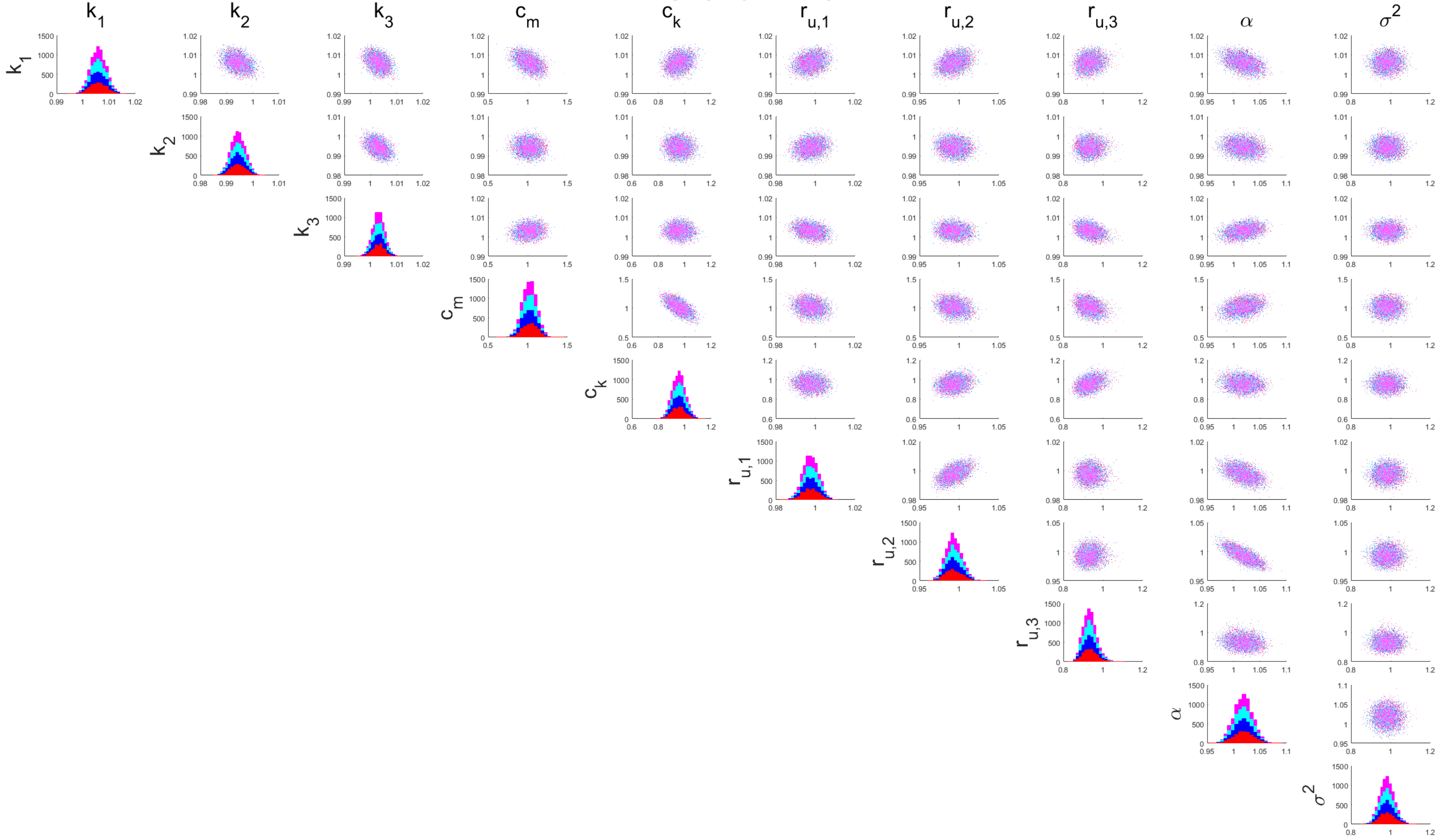
$$\dot{x}(t) = f(x(t), u(t), \theta)$$
$$y(t_i) = h(x(t_i), u(t_i), \nu(t_i, \theta), \theta)$$

Likelihood Function $p(\mathcal{D} \mid \theta, \sigma)$

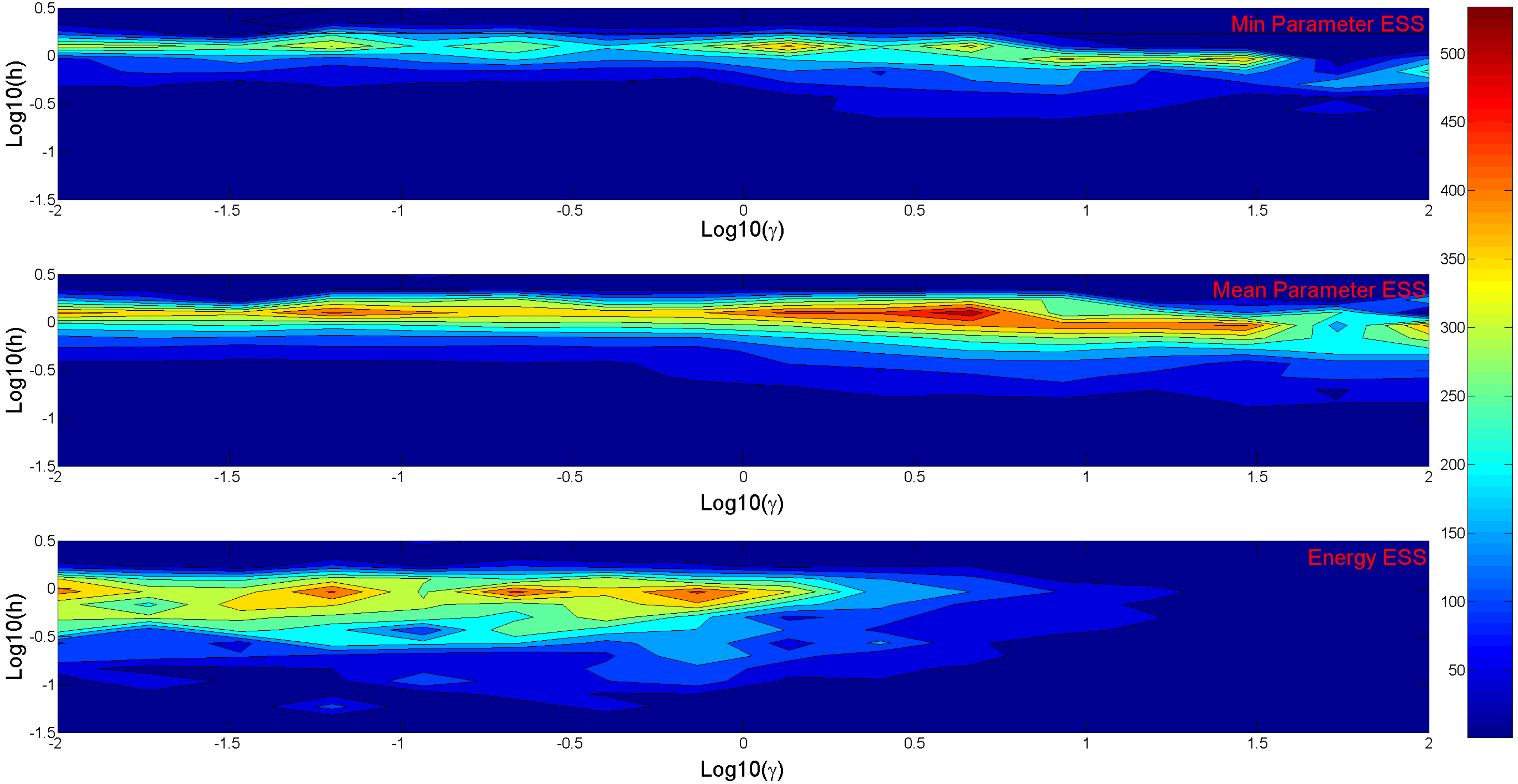
$$(2\pi\sigma^2)^{\frac{-N_d N_t}{2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N_d}\sum_{t=1}^{N_t}\left(y_t^{(i)}(\theta) - \hat{y}_t^{(i)}\right)^2\right]$$



Results



Performance Analysis



Future Directions

- Better Leverage Existing Dynamical Systems Methods
  - Adaptive Time steps
  - State Feedback Control
- Online Bayesian Inference
  - Parameter Estimation →Simulating Dynamical System→Filtering

Thank You

- Advisor: Jim Beck
- CMS/Caltech
- LANL Practicum
  - Russell Bent
  - Earl Lawrence
- The Krell Institute
- DOE