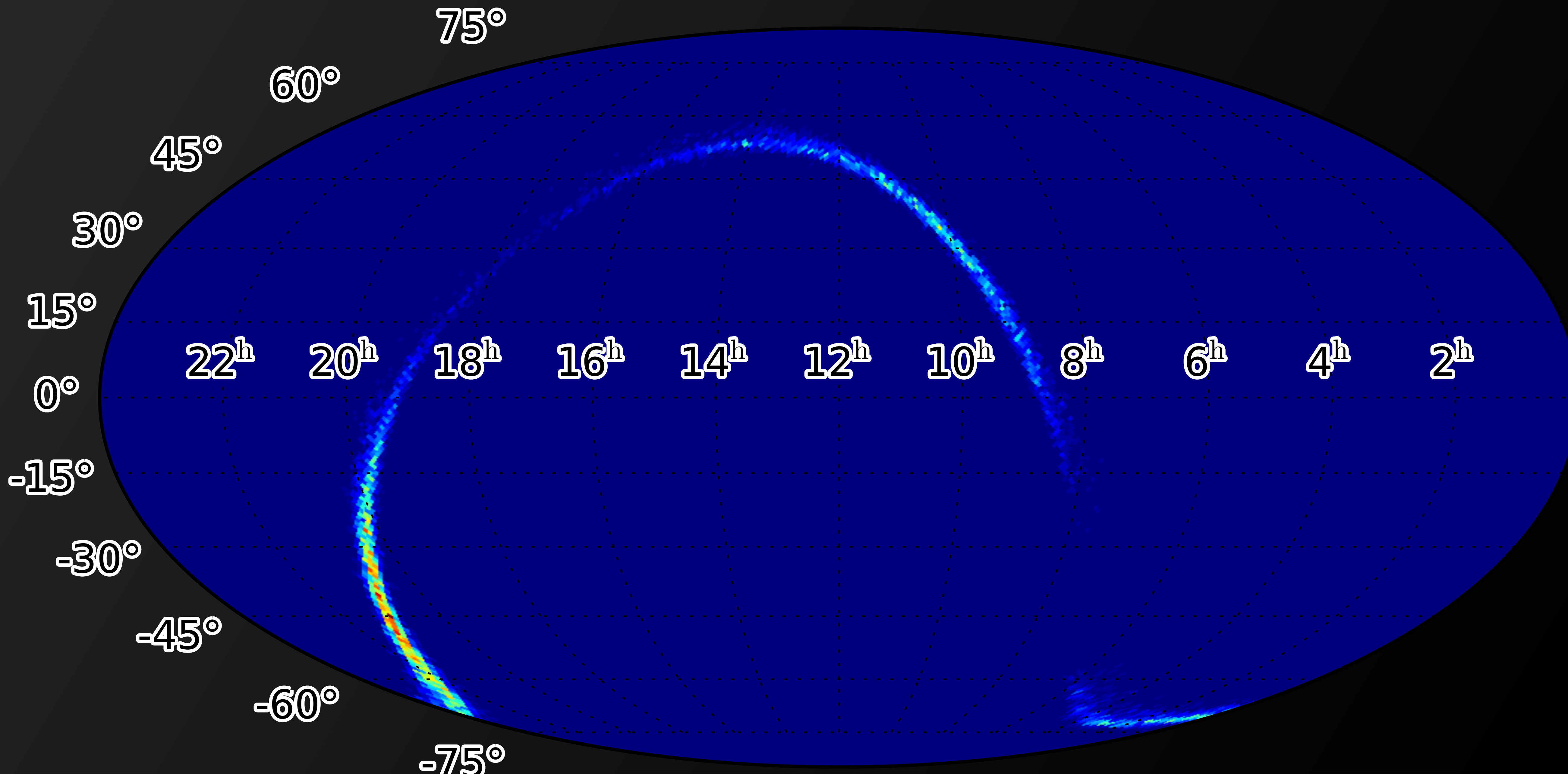
Parameter Estimation in the Advanced-Detector Era



Ben Farr

Cardiff University

LIGO-G1301239

Outline

- Intro to Parameter Estimation
- Intro Markov-chain Monte Carlo Methods
- Improving MCMC efficiency
- Improving noise modeling

Bayesian Inference

Posterior

Likelihood

Prior

$$p(\vec{\theta}|d) = \frac{p(d|\vec{\theta})p(\vec{\theta})}{p(d)}$$

Evidence

Single-detector Likelihood Function

$$p(d|\vec{\theta}) = \exp\left[-2\int_0^\infty \frac{\left|\tilde{d}(f) - \tilde{h}(f;\vec{\theta})\right|^2}{S_n(f)} df\right]$$

- Where *d* assumed to be Gaussian noise with some signal.
- *h* is the pN model signal.
- S_n is the noise power spectral density.

Priors

- Uninformative priors:
  - Flat in component mass space.
  - Volumetric in space.
  - Flat in time, phase, orbital orientation.

Posterior Sampling

1. Markov-Chain Monte Carlo:

   - Stochastically wonders about parameter space
   - Metropolis-Hastings algorithm guarantees samples distributed according to posterior

2. Nested sampling:

   - Samples the prior with 'mini-MCMCs'
   - weighs points along the way

Metropolis–Hastings

1. A chain is at current location $\vec{\theta}$.

2. A new location $\vec{\theta'}$ is proposed according to $Q(\vec{\theta} \to \vec{\theta})$.

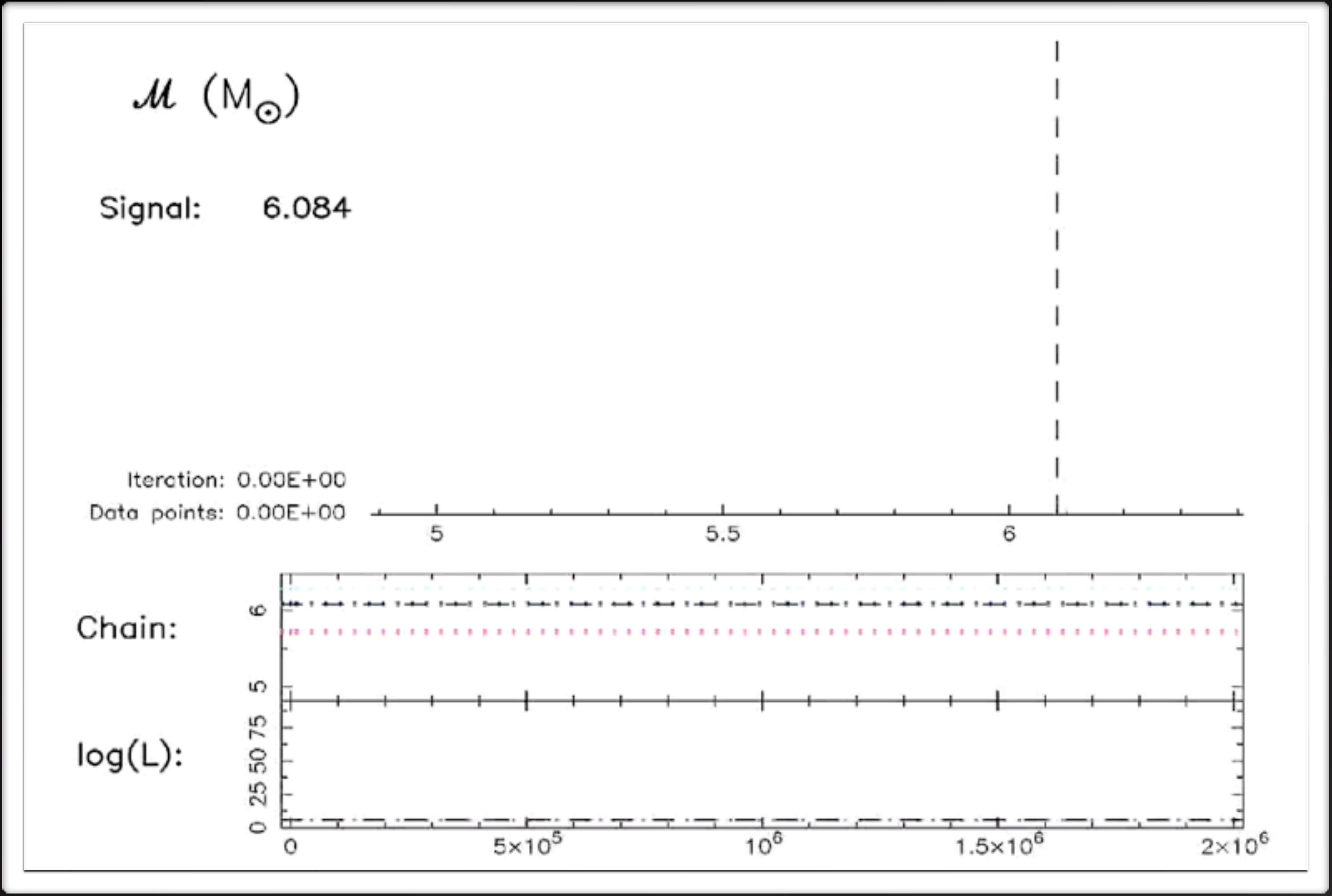
3. Proposal is accepted if:

$$\frac{p(\vec{\theta'}|d)Q(\vec{\theta};\theta')}{p(\vec{\theta}|d)Q(\vec{\theta'};\theta)} > r, \text{ where } \; r = U(0,1)$$

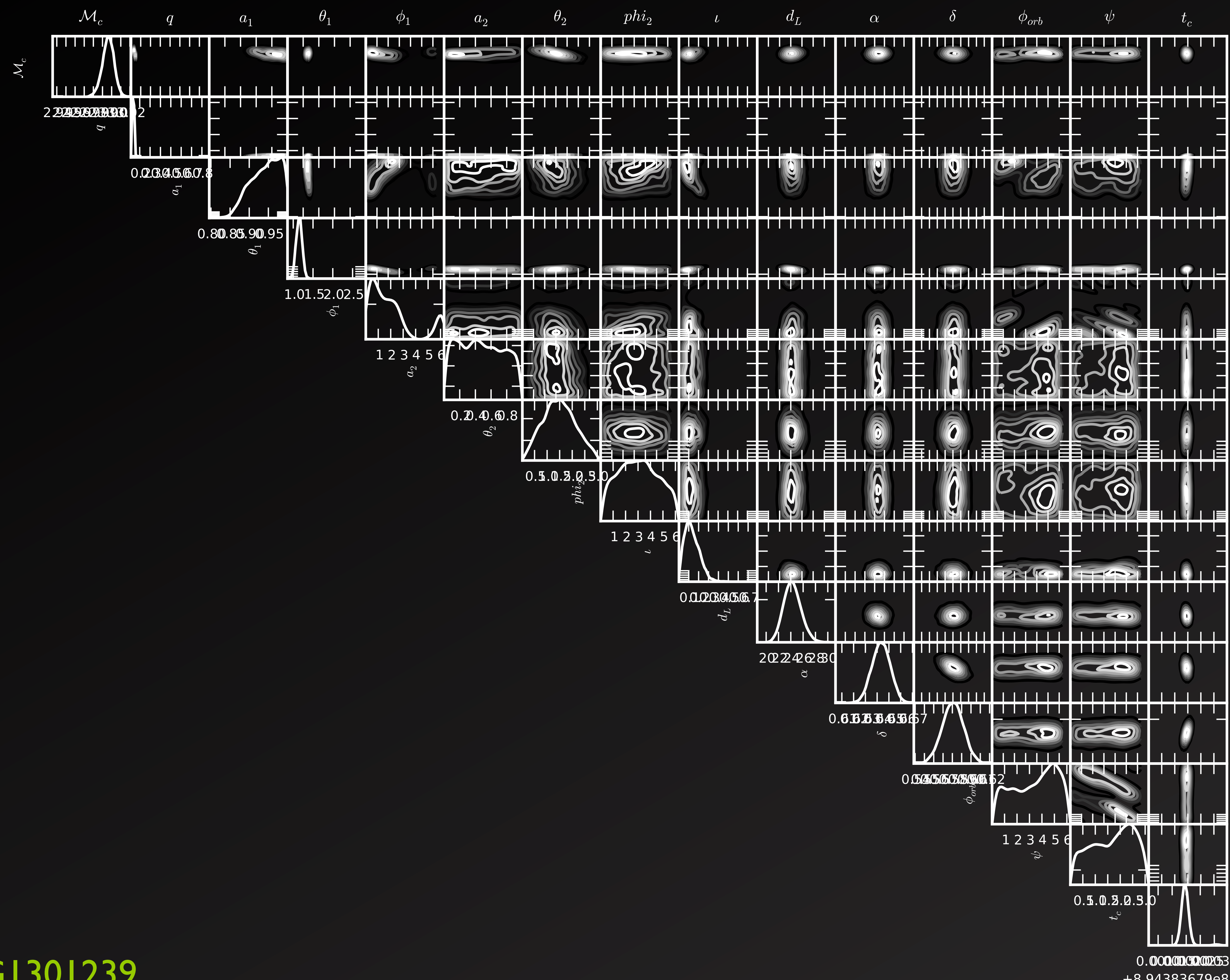
or else $\vec{\theta}$ is repeated.

4. Repeat. A lot.

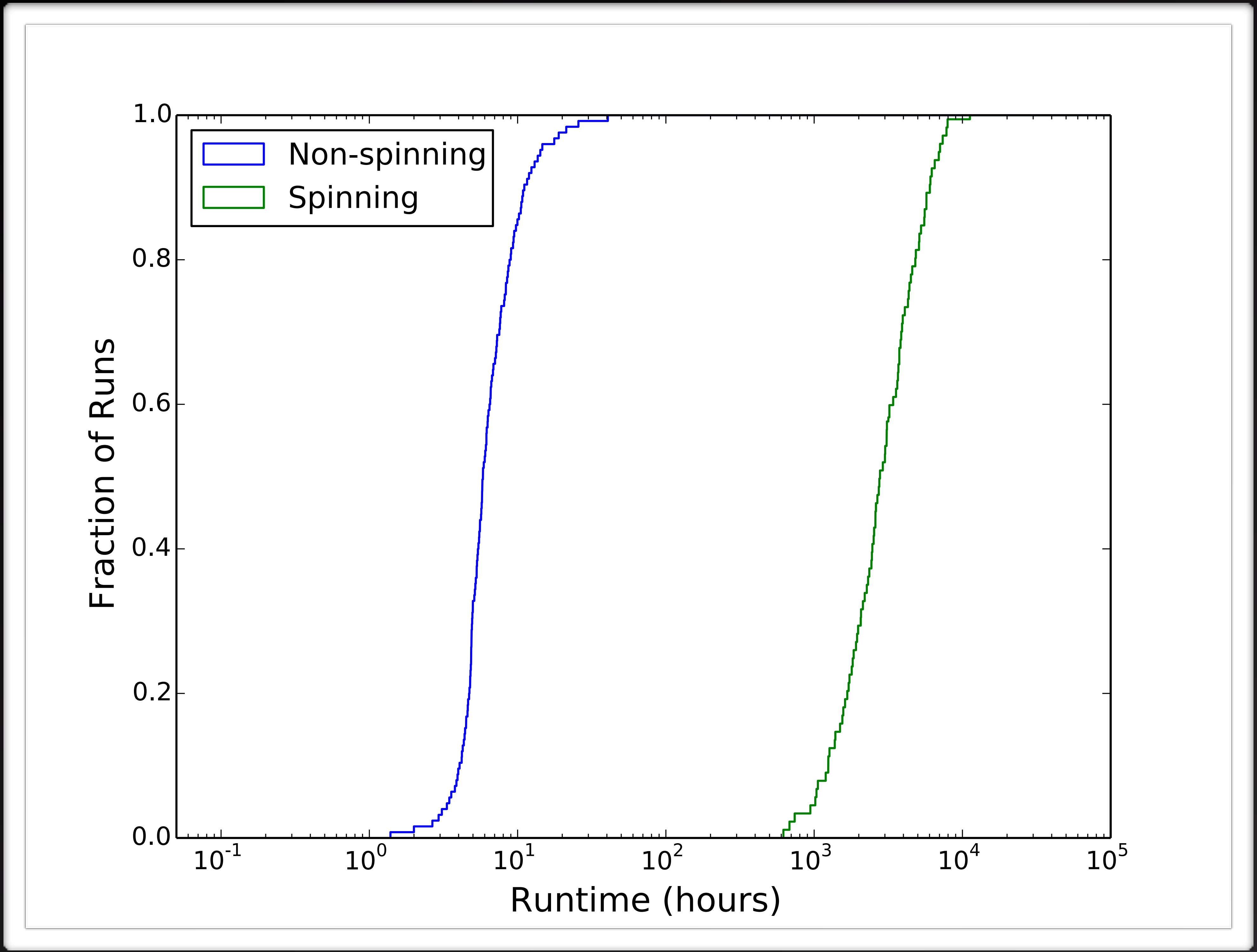
1-D Posterior Estimate



15-D Posterior Estimate



MCMC Efficiency



MCMC Efficiency

- Only independent samples used to estimate posterior.
- Chains are thinned by autocorrelation length.
- Target: Smaller ACLs through proposals.

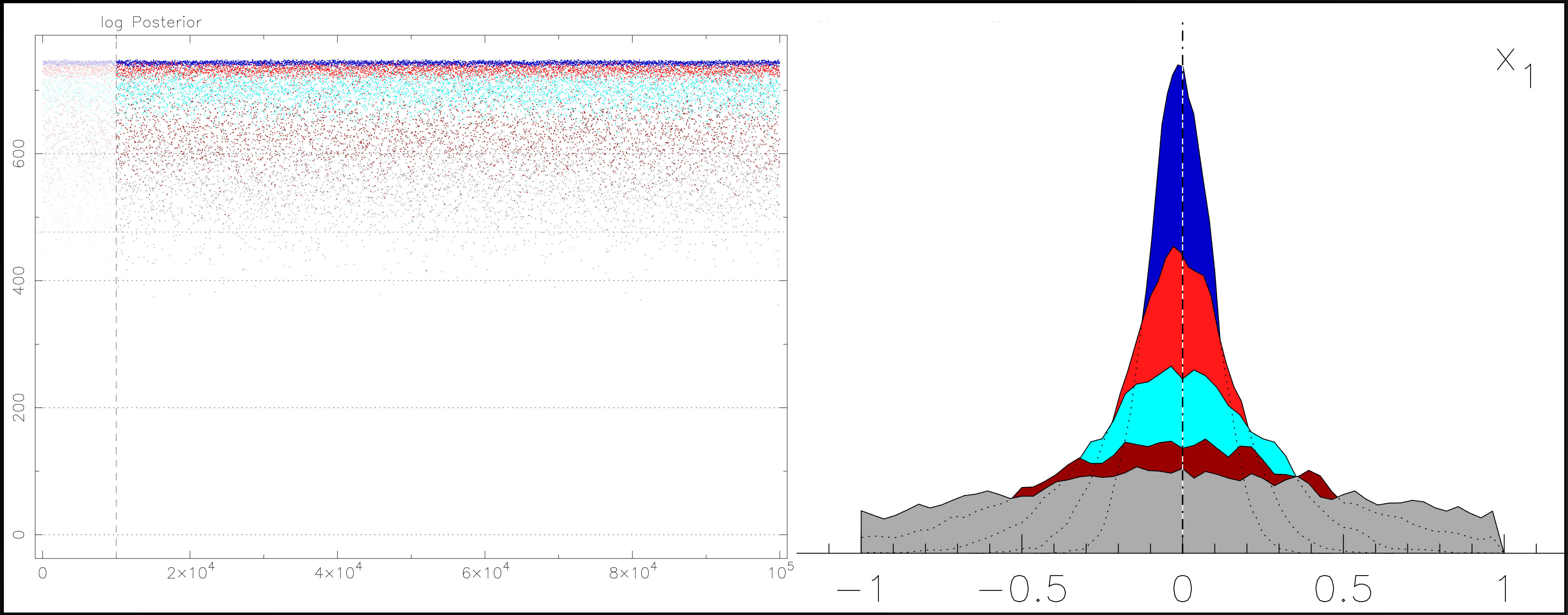
Parallel Tempering

- Modify the likelihood function.

$$p(\vec{\theta}|d) = \frac{p(d|\vec{\theta})p(\vec{\theta})}{p(d)} \quad \Rightarrow \quad p(\vec{\theta}|d) = \frac{p(d|\vec{\theta})^{1/T_i}p(\vec{\theta})}{p(d)}$$

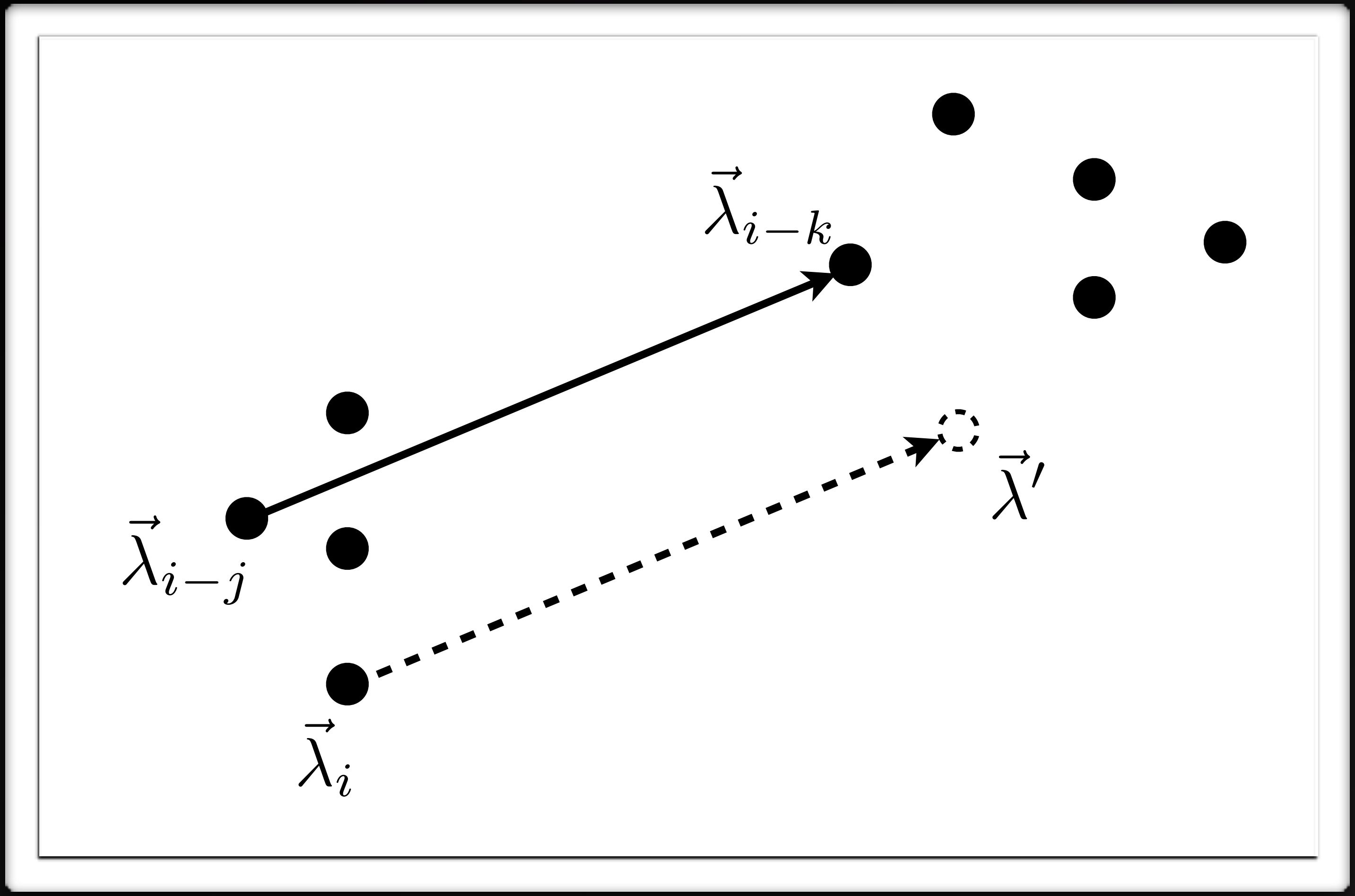
- Run many chains in parallel.
- Propose swaps between chain locations.

Parallel Tempering



- Higher temperatures reduce contrast of posterior.
- Effective for multimodal sampling, but wasteful.
- Most likelihood calculations discarded.

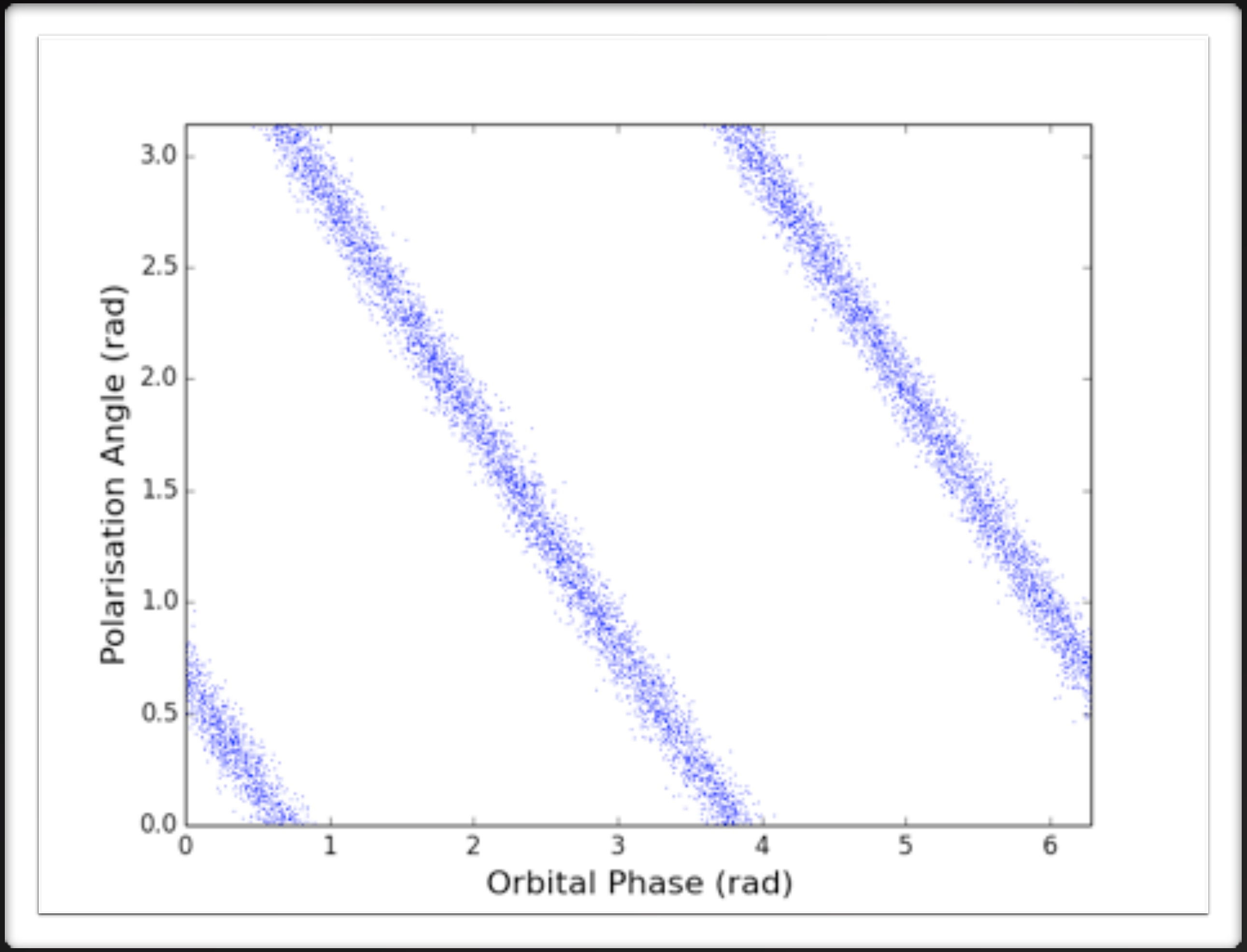
Differential Evolution



- Uses previous samples to generate proposals.
- Helps with linear correlations and multiple modes.

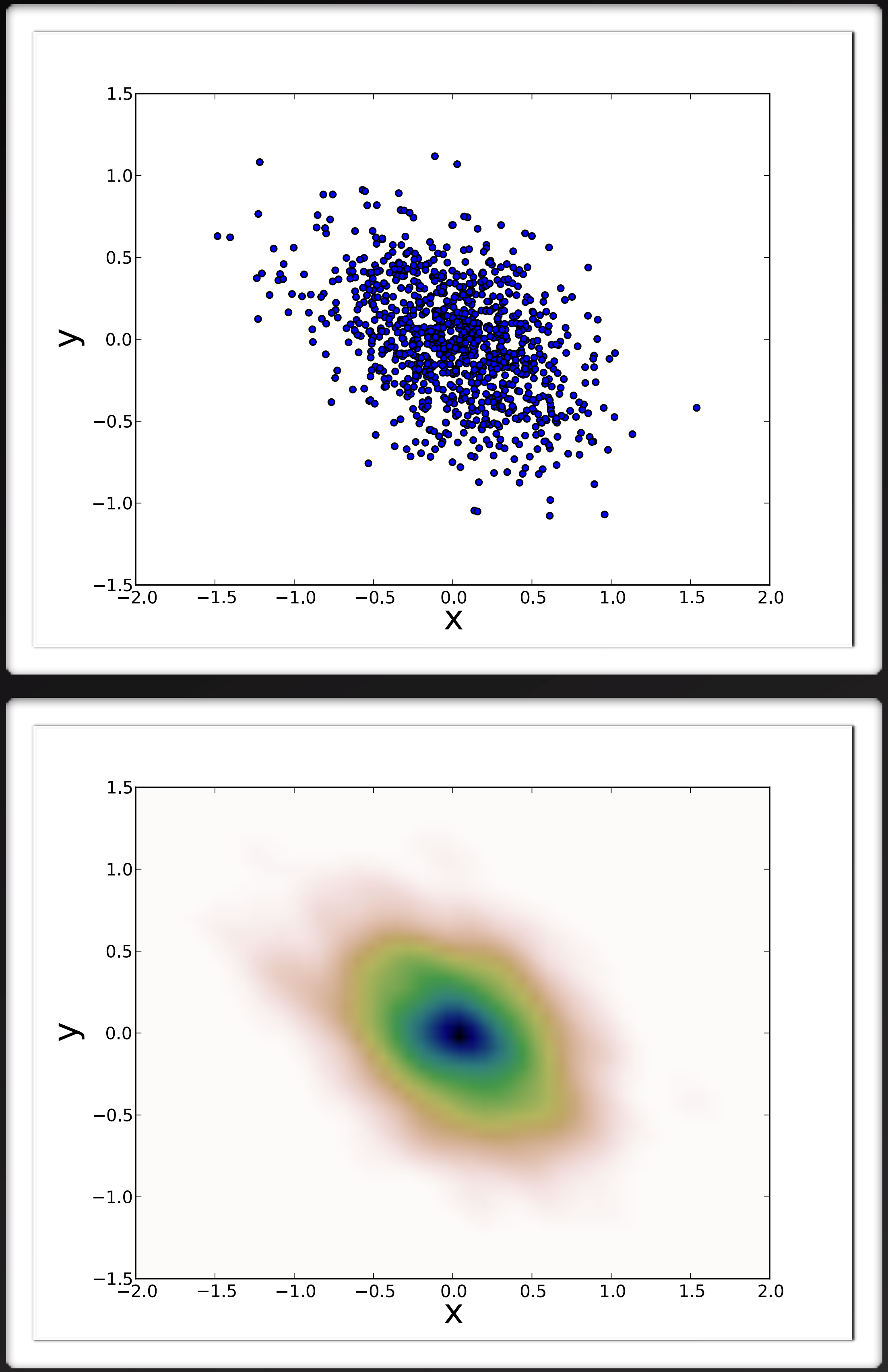
Jump Proposals

- Vanilla MCMC: local Gaussian
- Some account for known degeneracies.
  - e.g. polarisation and phase
- Some are problem-agnostic.



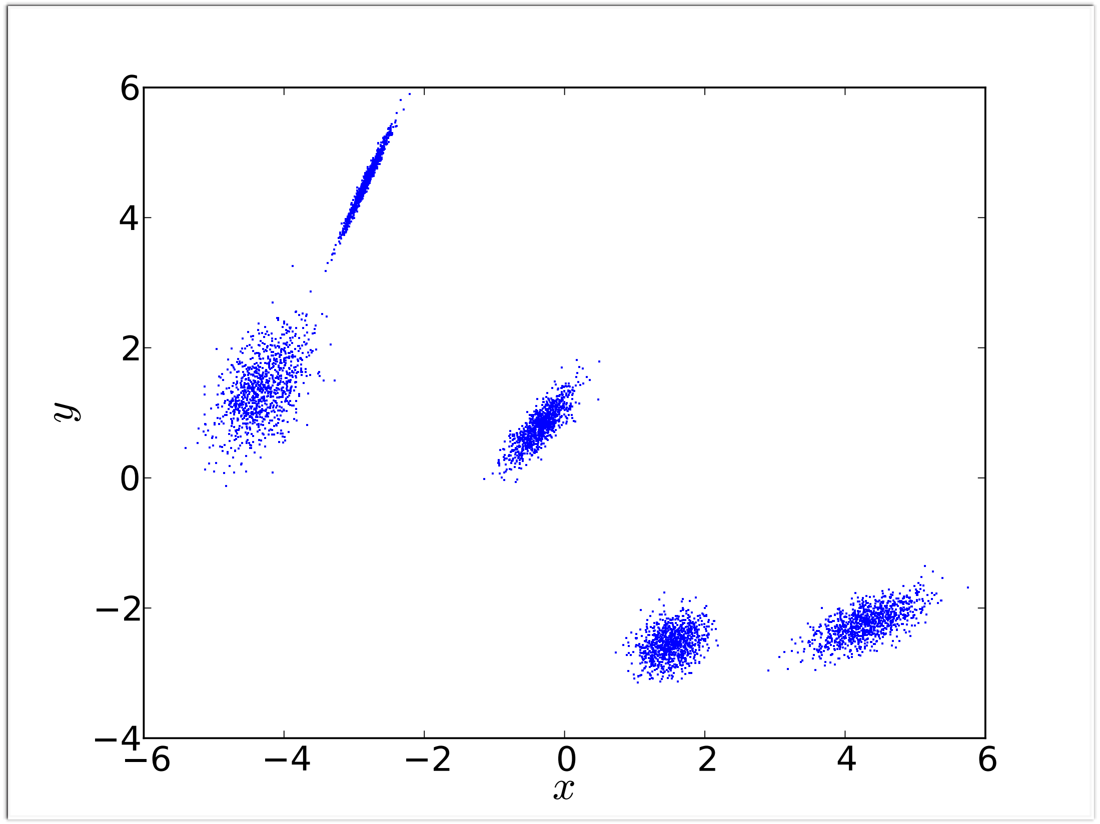
Educating an MCMC

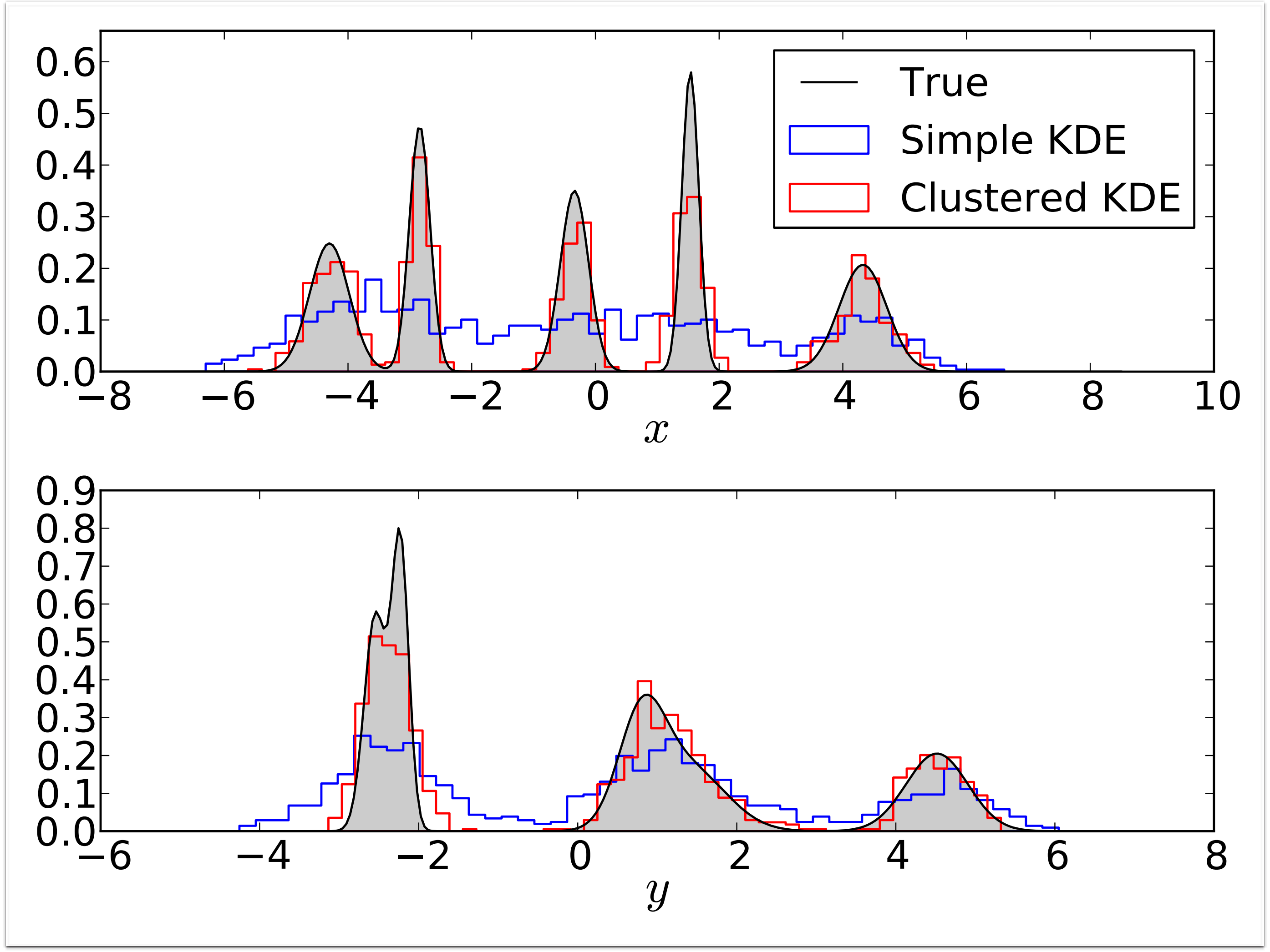
- Can be done through jump proposals.
- Kernel Density Estimators (KDEs):
  - \+ Estimates distribution from samples.
  - \+ Easily draw samples from result.
  - \+ Continuous (helps with detailed balance).
  - \- Over-smoothes multimodal distributions.



Educating an MCMC

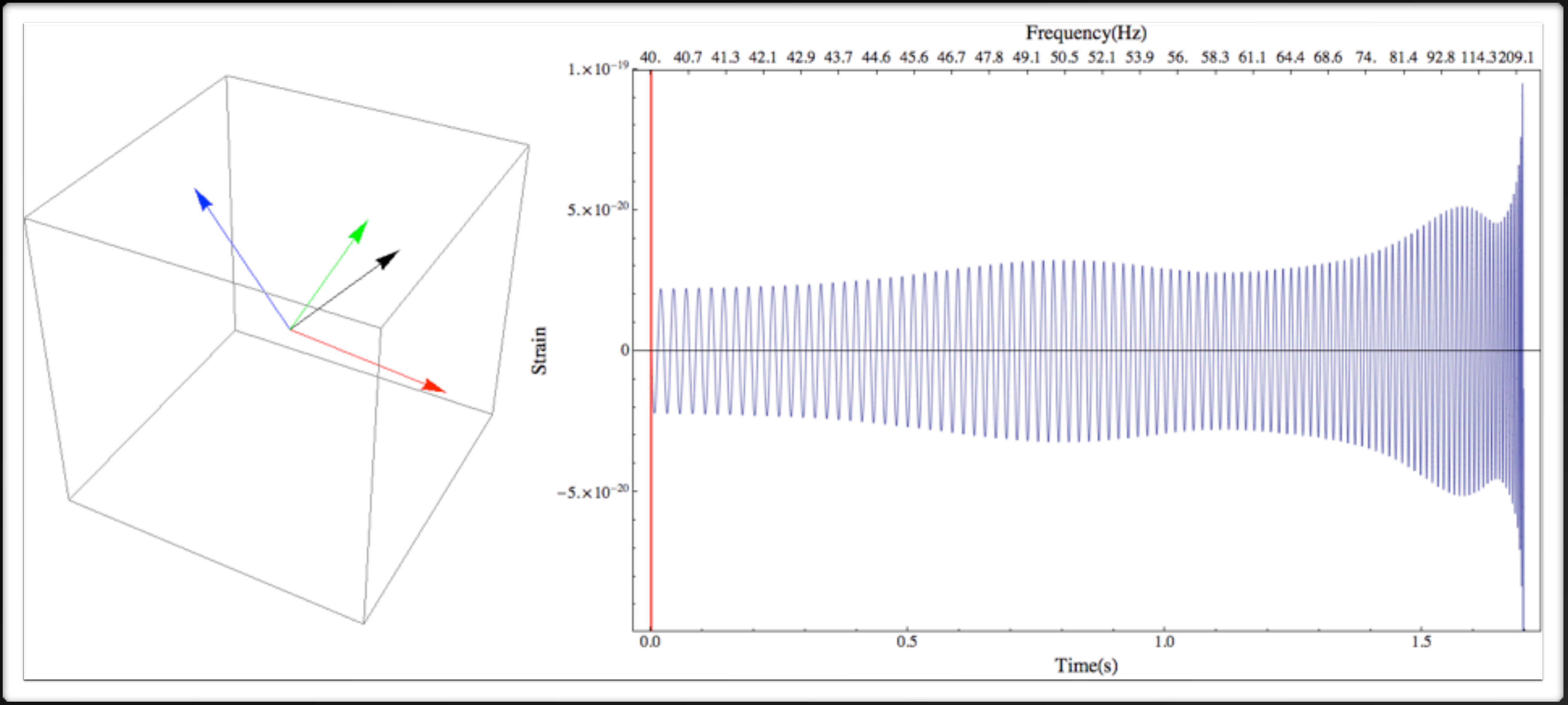
- Use a density-based clustering algorithm (OPTICS) to extract clusters.
- Estimate posterior in each mode with a KDE.
- Weight KDEs by cluster size.





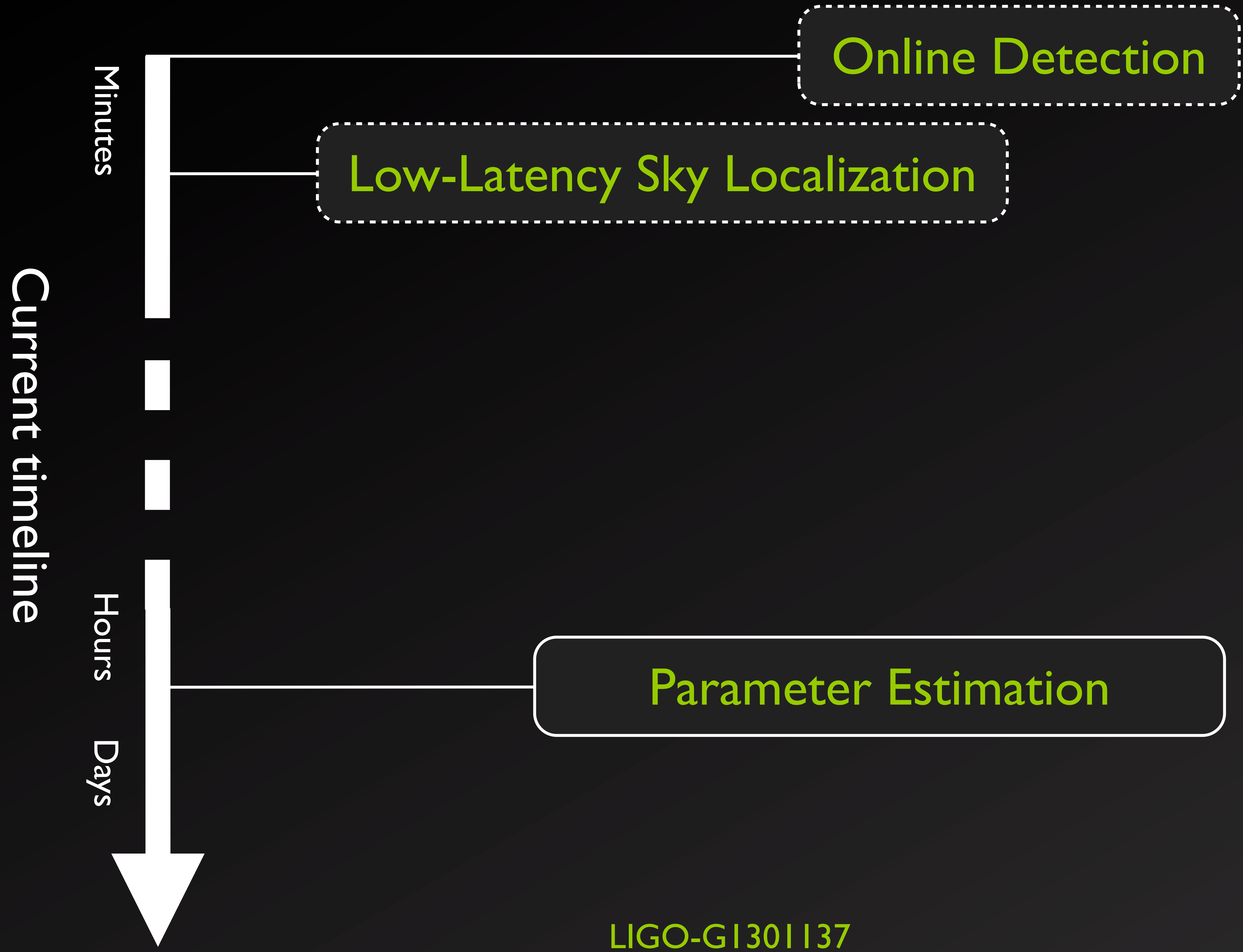
Reparameterisation

- Move away from the radiation frame.

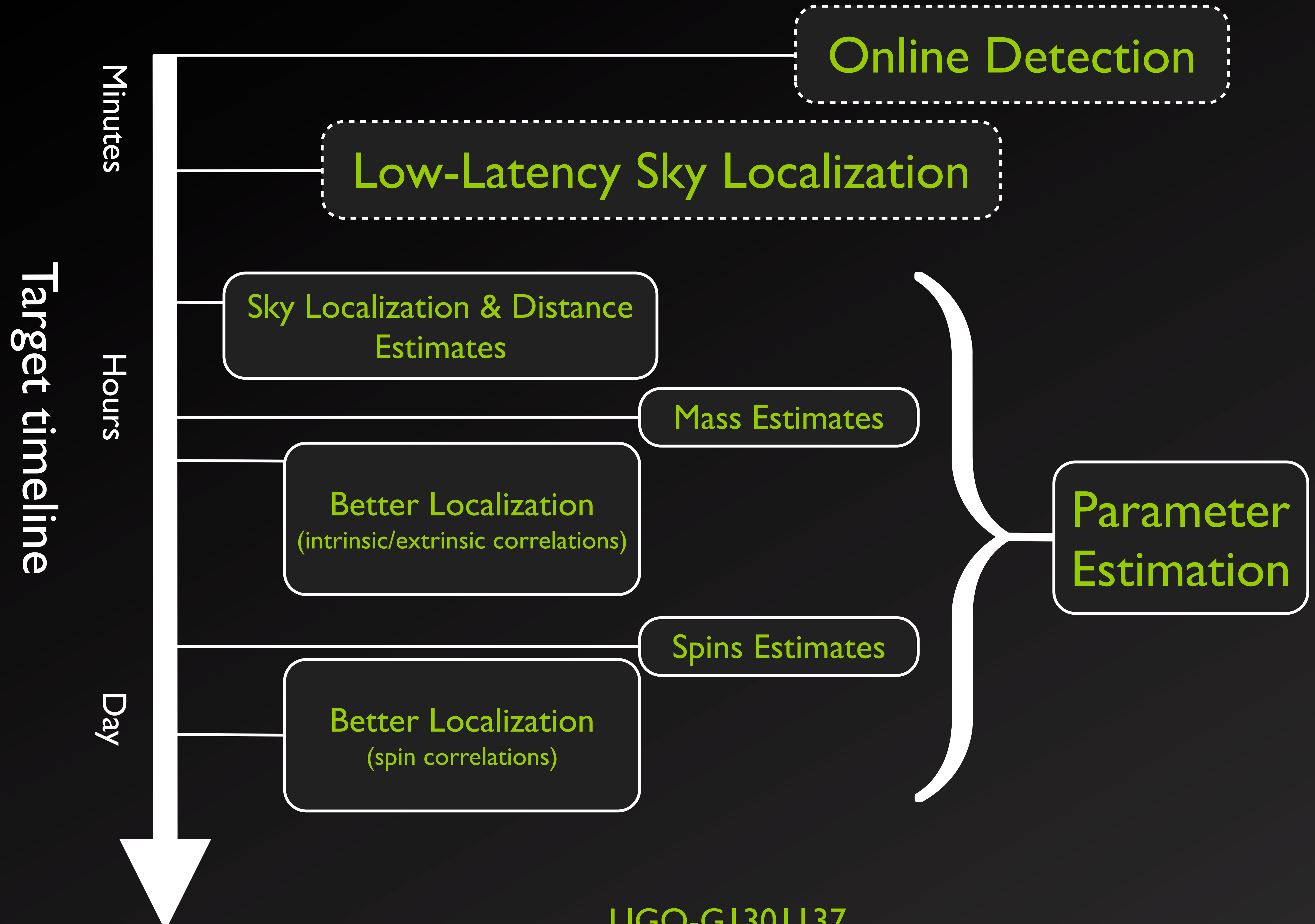


- Define evolving parameters where you are sensitive to them.

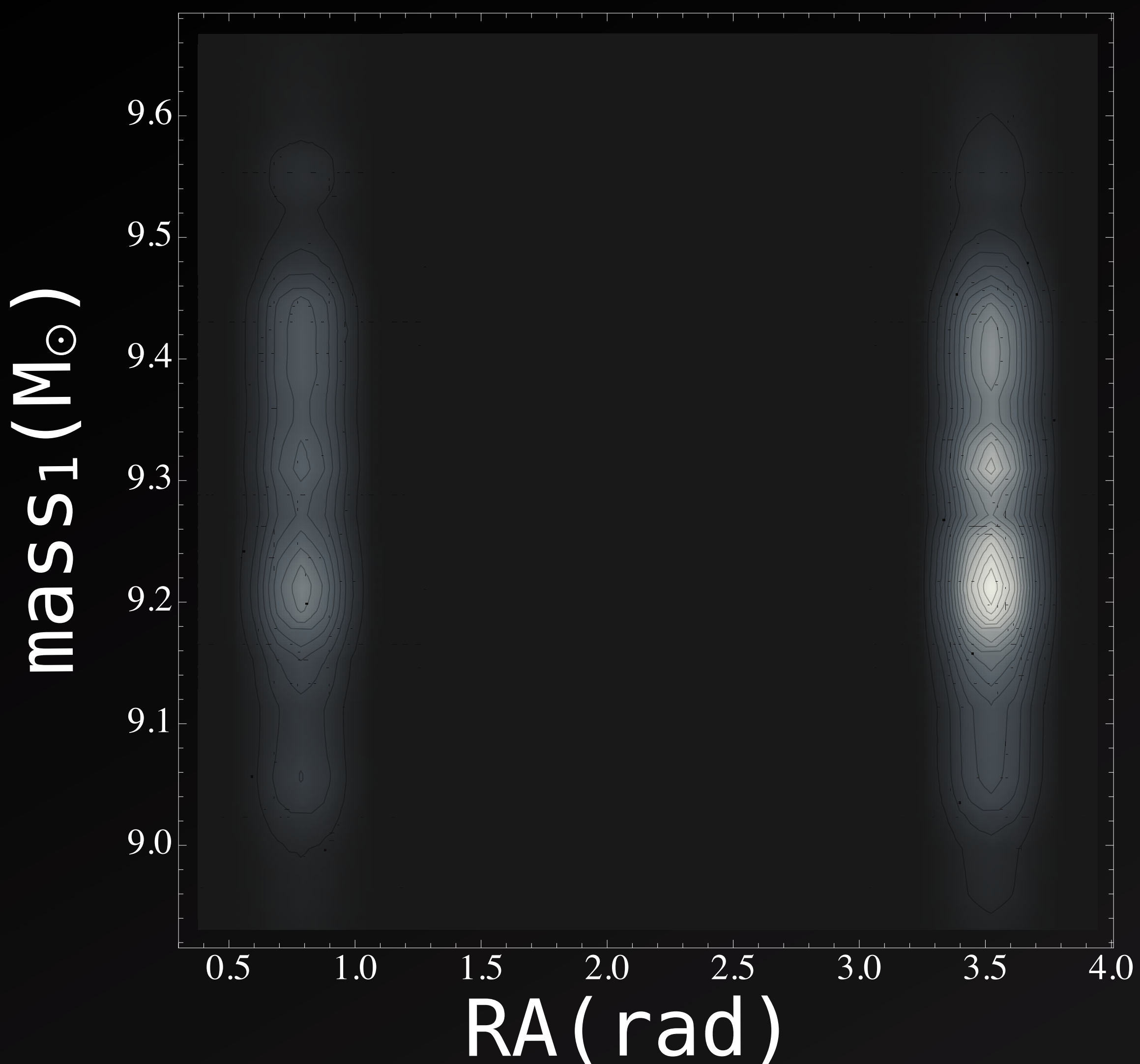
Hierarchical MCMC

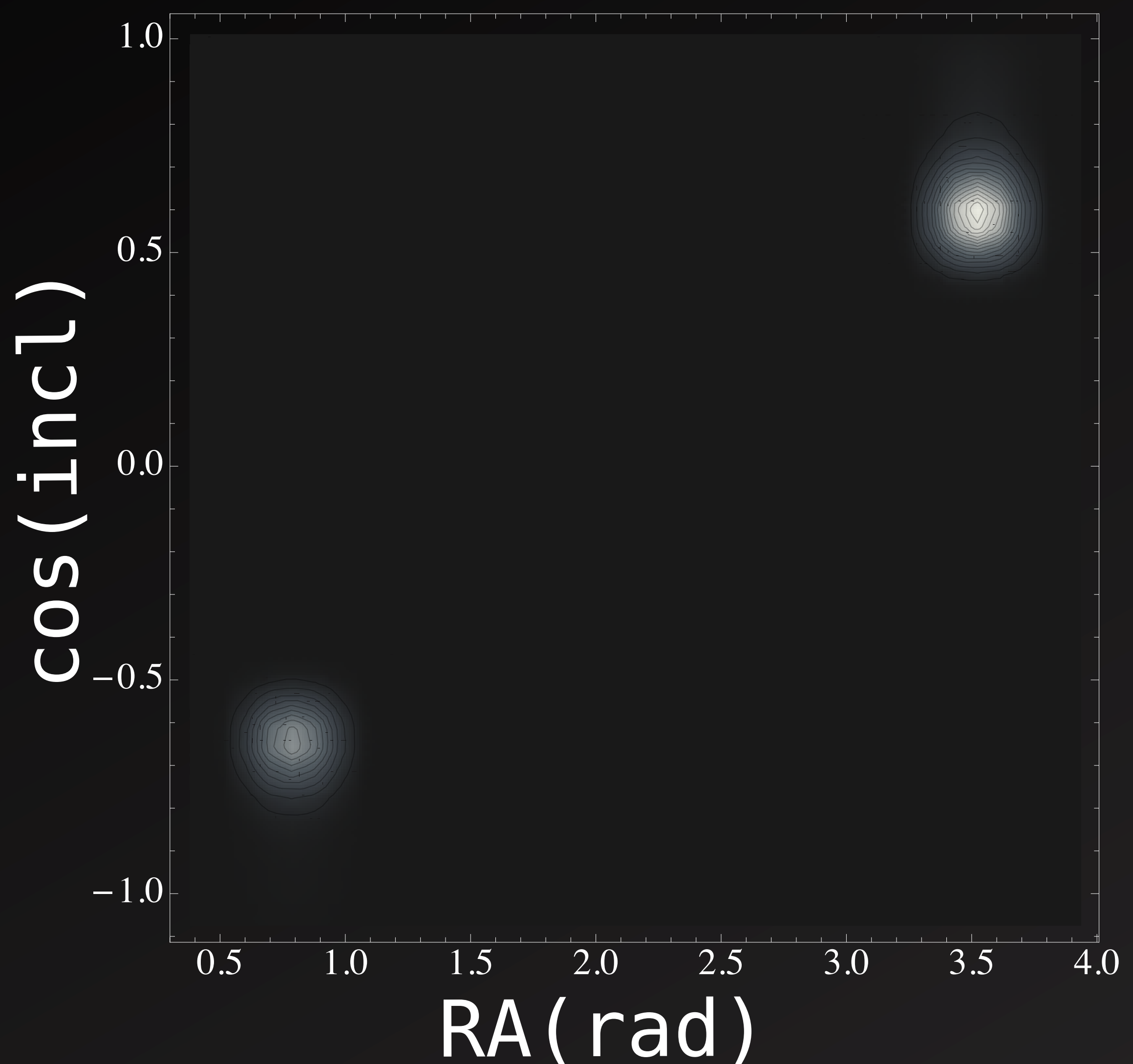


Hierarchical MCMC



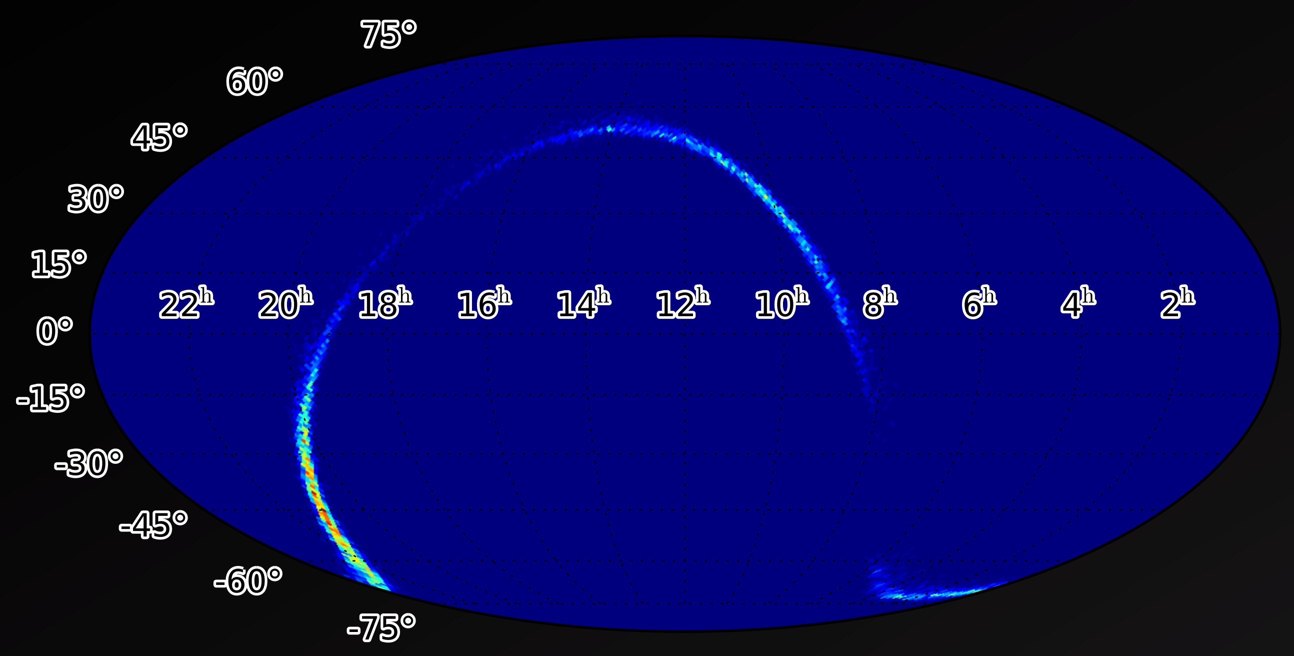
Hierarchical MCMC

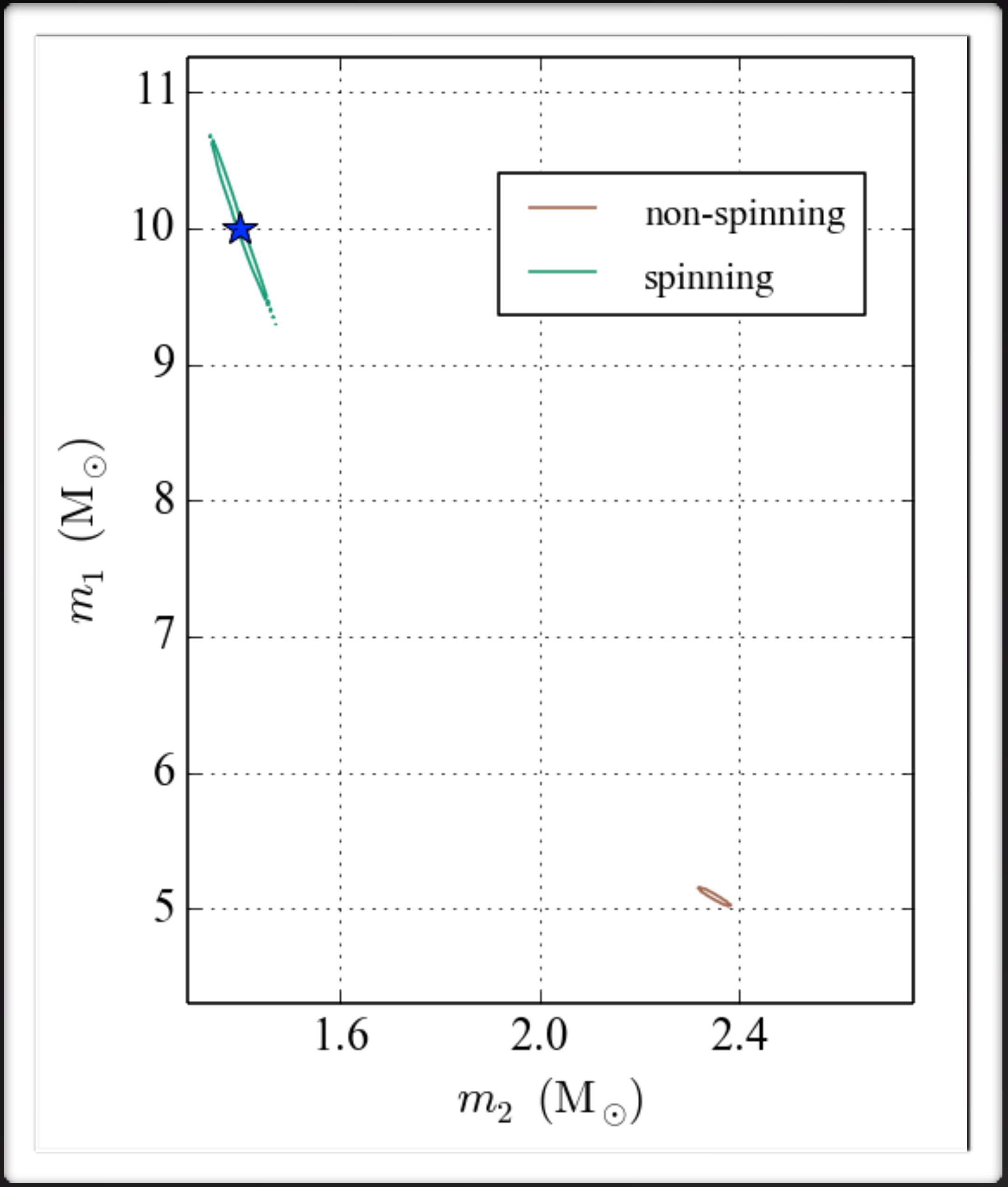




- Intrinsic and extrinsic parameters decouple.
- Full PE follow-up is the most accurate, but also slow.
- Exploit lack of correlation with mass.

Remaining Work





- Leverage low-latency results.
- Test on 2-detector data.
- Extend to the spinning parameter space.

Better Noise Models

- Goal: account for non-stationarity and non-Gaussianity.

- We have started including more sophisticated noise models in the MCMC.

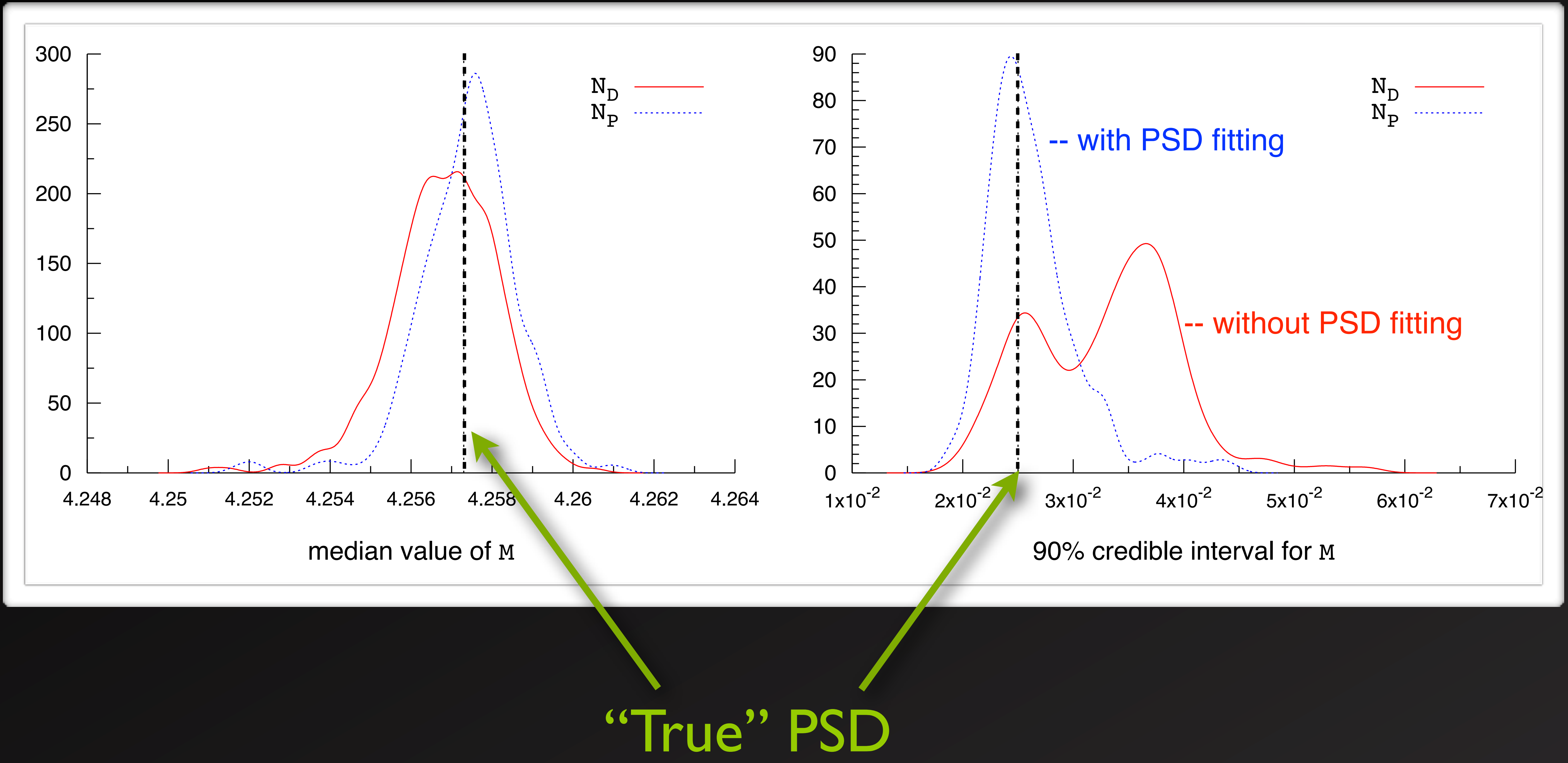
- Marginalize over uncertainties in the model.

[Littenberg, Coughlin, Farr, Farr, arXiv:1307.8195 (soon in PRD)]

PSD Bin Scaling

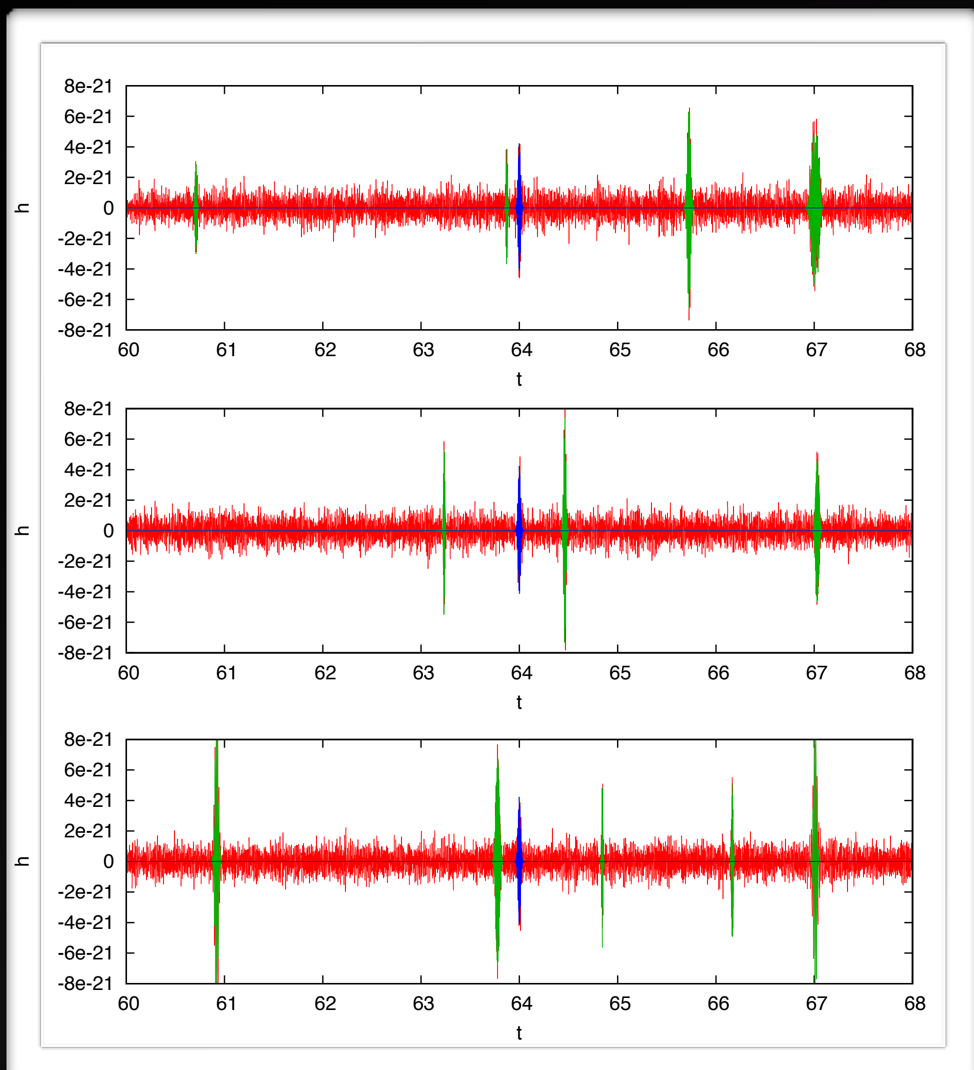
Distribution of medians from ~300 runs

Distribution of 90% CIs from ~300 runs



[Littenberg, Coughlin, Farr, Farr, arXiv:1307.8195 (soon in PRD)]

Glitch Modelling



Independent | Coherent

$$d^{\mathrm{H}} = n^{\mathrm{H}} + \sum_i^{N_g^{\mathrm{H}}} \psi(\vec{\gamma_i}^{\mathrm{H}}) + h^{\mathrm{H}}(\vec{\theta})$$

$$d^{\mathrm{L}} = n^{\mathrm{L}} + \sum_i^{N_g^{\mathrm{L}}} \psi(\vec{\gamma_i}^{\mathrm{L}}) + h^{\mathrm{L}}(\vec{\theta})$$

$$d^{\mathrm{V}} = n^{\mathrm{V}} + \sum_i^{N_g^{\mathrm{V}}} \psi(\vec{\gamma_i}^{\mathrm{V}}) + h^{\mathrm{V}}(\vec{\theta})$$

$$h^{\mathrm{IFO}} = \mathcal{R}^{\mathrm{IFO}}(\vec{\lambda}) \times h(\vec{\theta})$$

Extrinsic

$$\{D, \vec{\Omega}, \phi_{\mathrm{ref}}, \iota, \psi, t_{\mathrm{ref}}\}$$

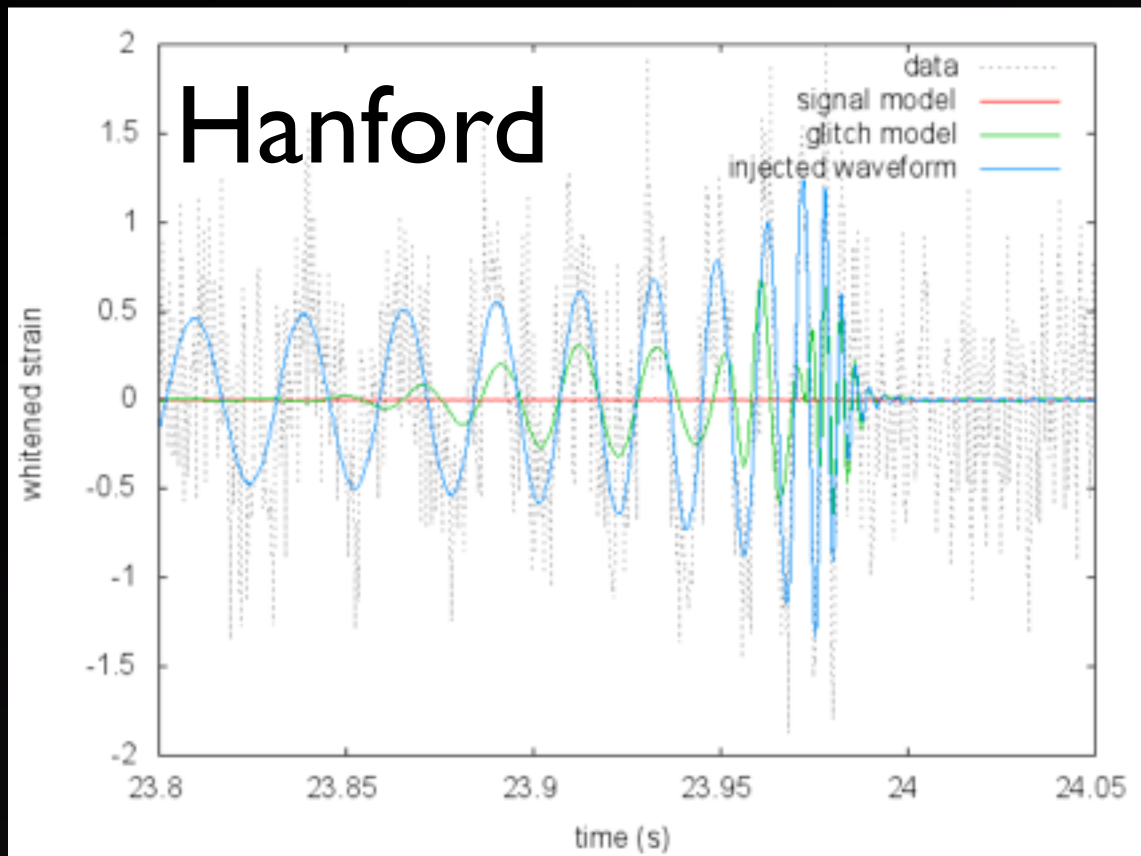
sky location, orientation, etc.

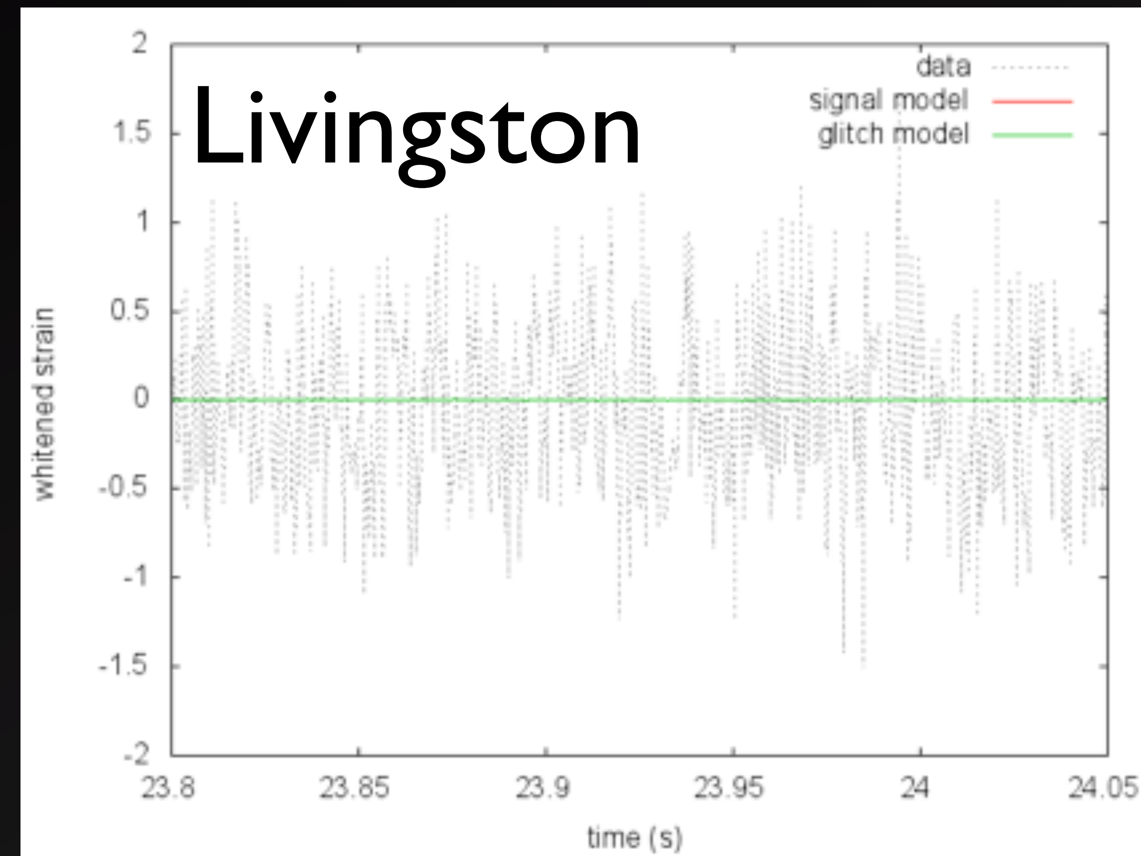
Intrinsic

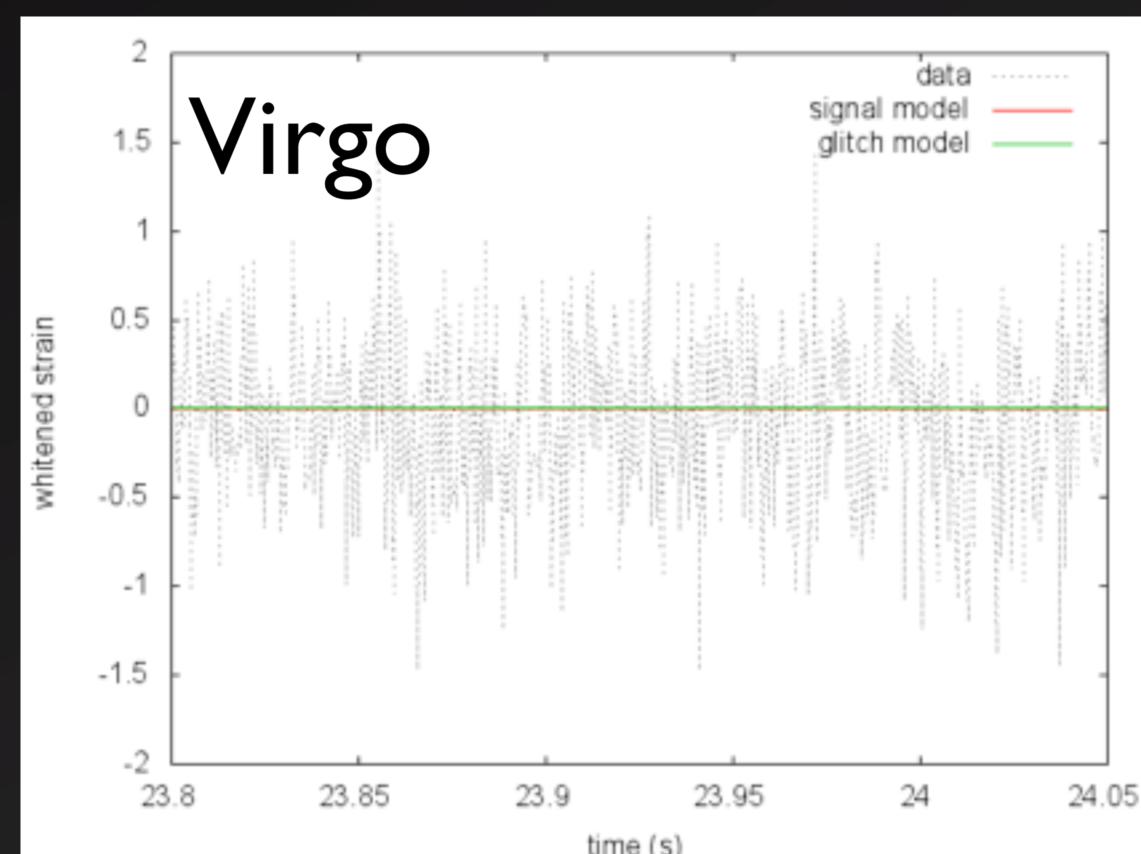
$$\{m_1, m_2, \vec{s_1}, \vec{s_2}\}$$

physical parameters

Glitch Modelling







Hanford SNR= 20

Bayes Factor$_{\mathrm{GW\ v.\ Glitch}}$ ~ $4\mathrm{x}10^{-3}$:1

Conclusions

- LVC parameter estimation has come a long way since big dog.
- We are ready for the start of the advanced detector era.
- Work still needs to be done to handle design sensitivity.
- Better noise models will soon allow for meaningful Bayes factors.