Intuition for MCMC Sampling

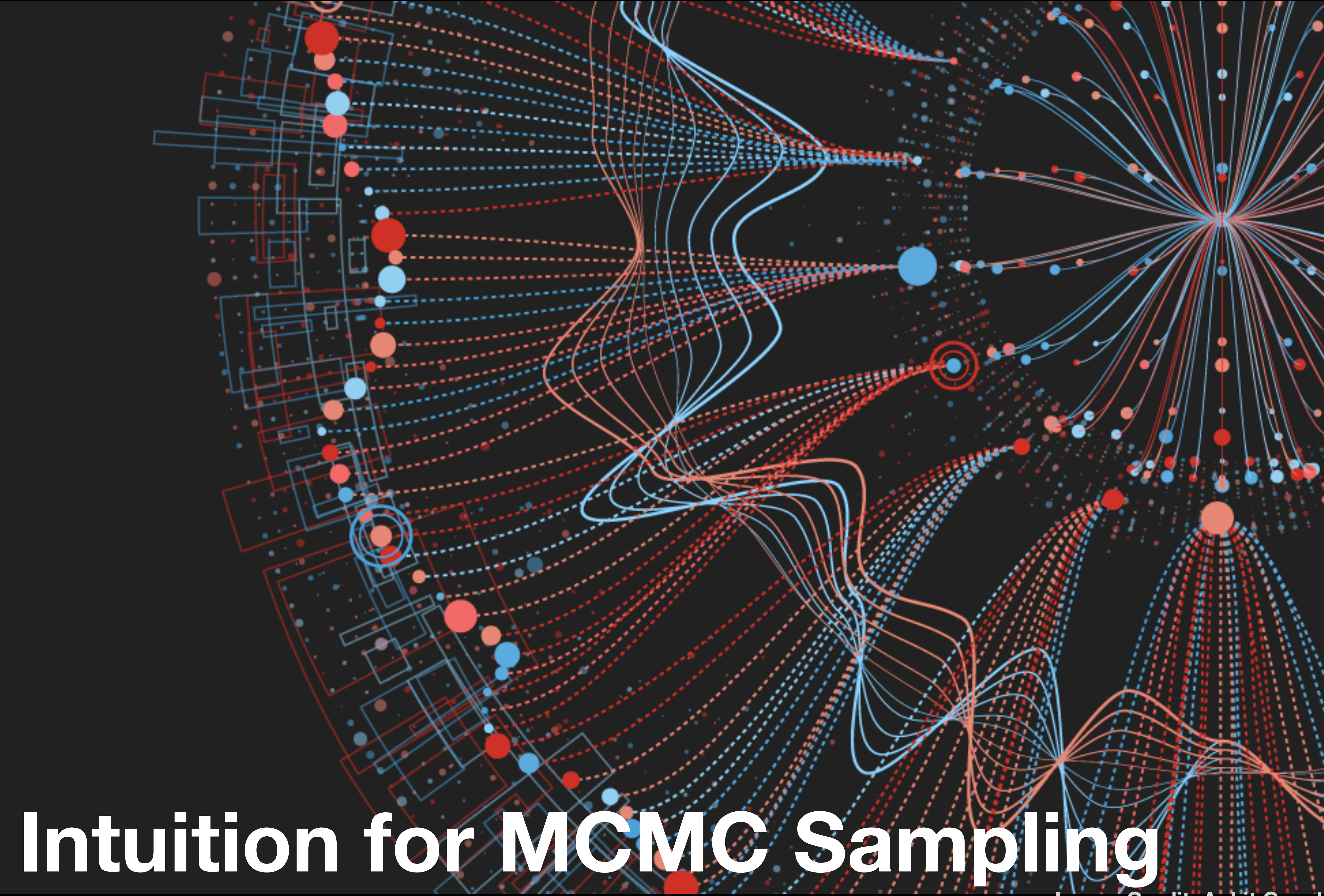
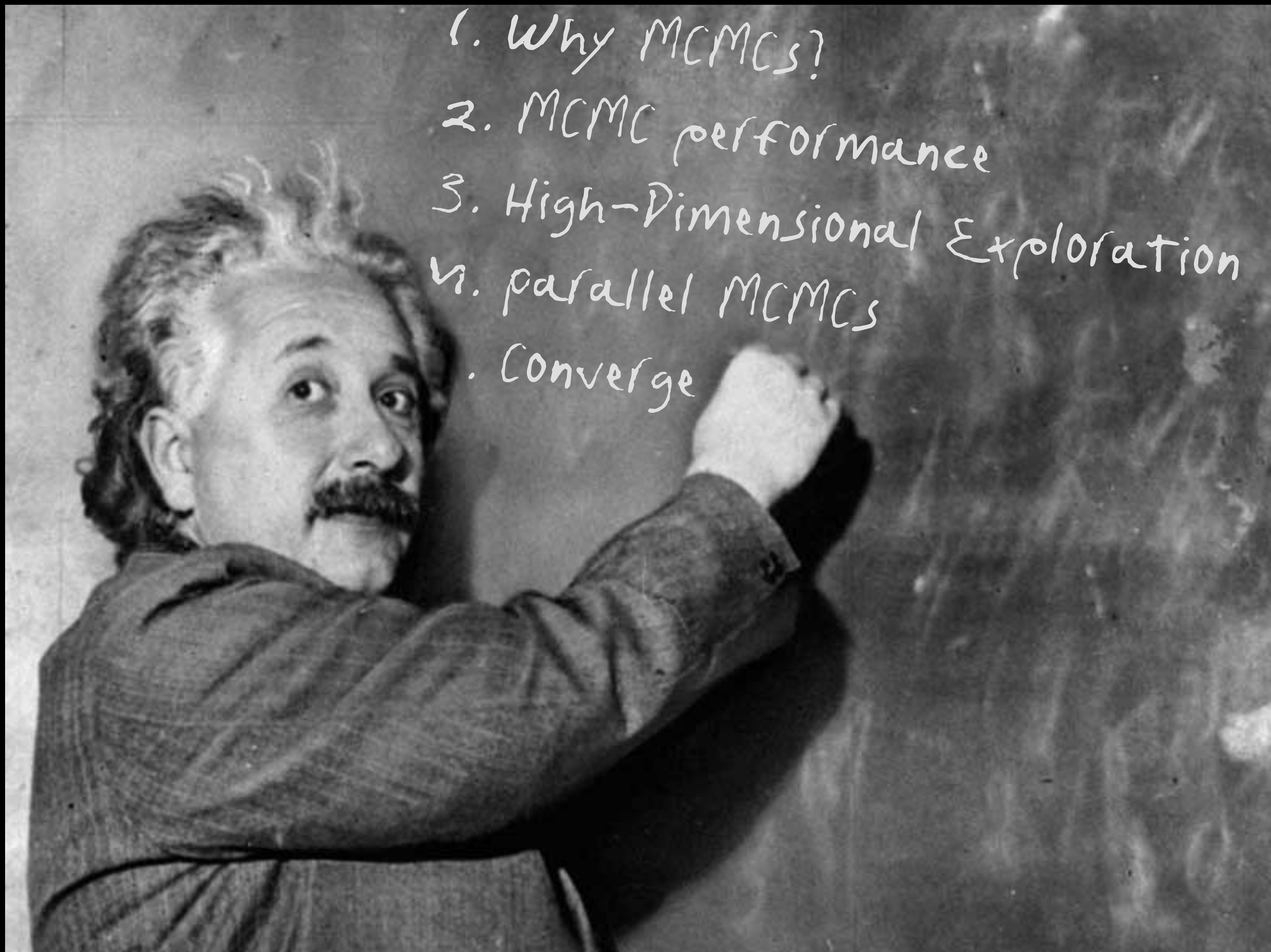


Image Credit: Ankit Goyal

Peter Behroozi

Outline


1. Why MCMCs?
2. MCMC performance
3. High-Dimensional Exploration
4. parallel MCMCs
. Converge

Why?

Observations

Why?

Observations

$$\mathcal{L}(x) = \exp(-\chi^2/2)$$

Models

MCMC

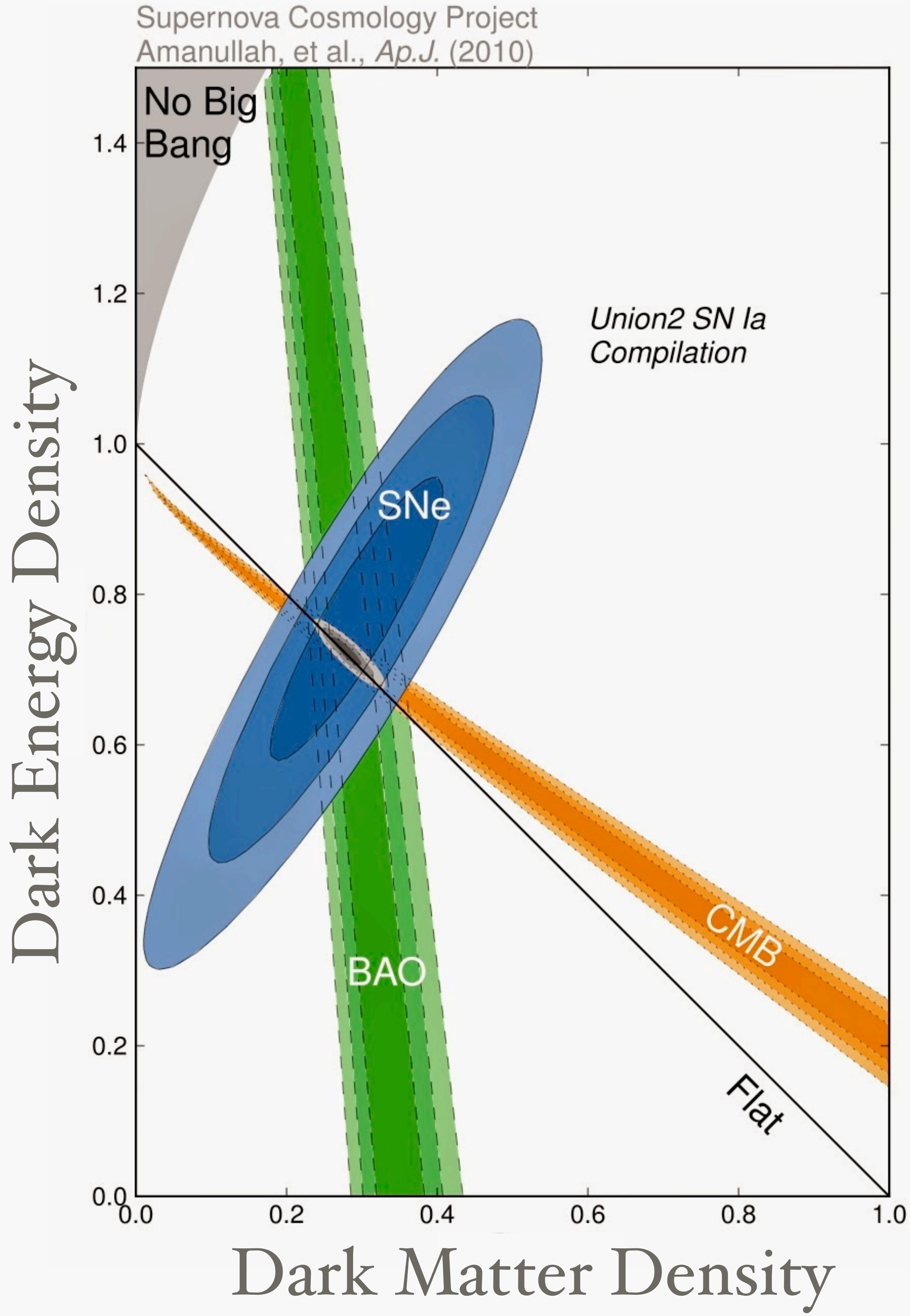
Why?

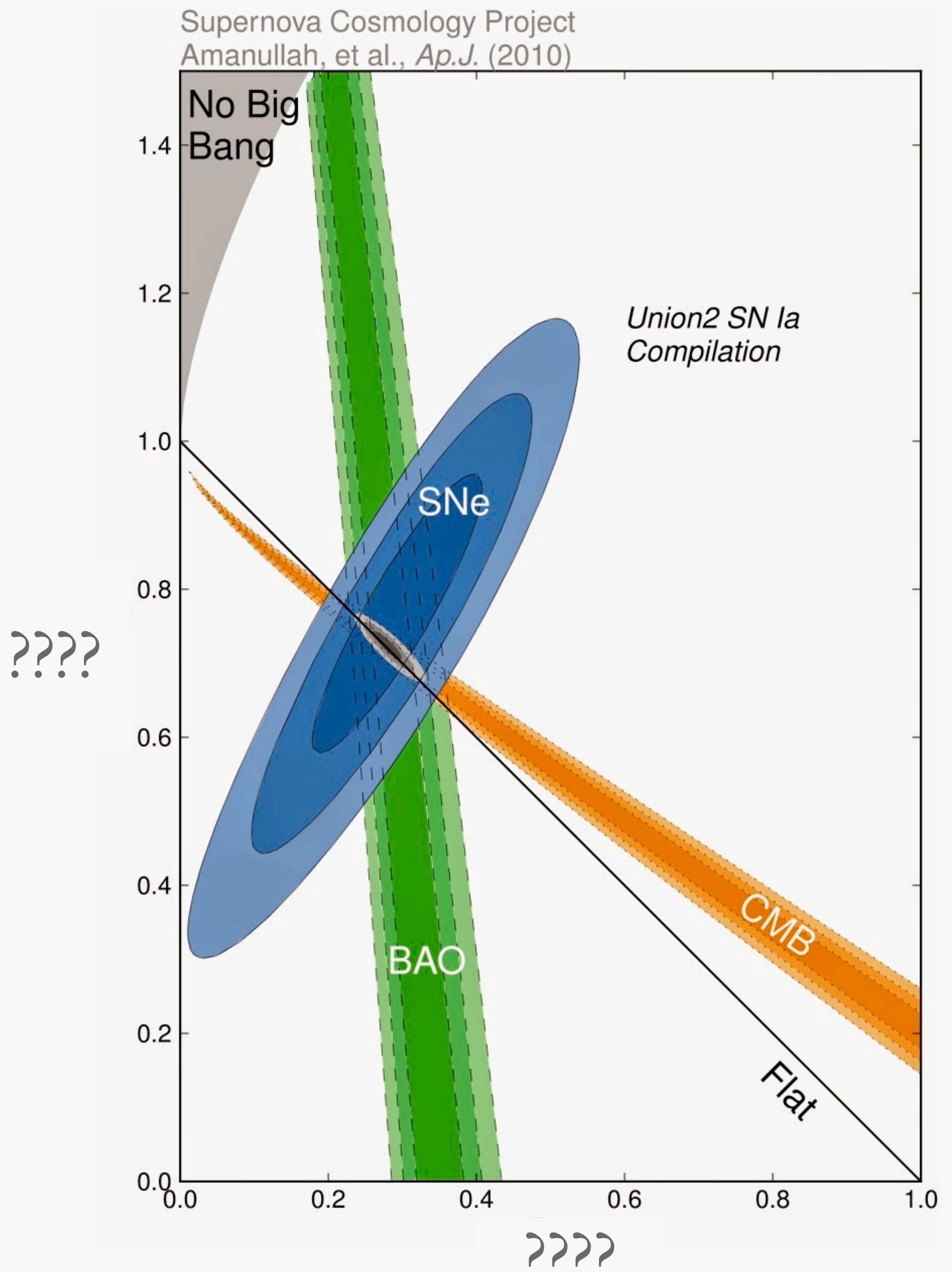
Observations

$$\mathcal{L}(x) = \exp(-\chi^2/2)$$

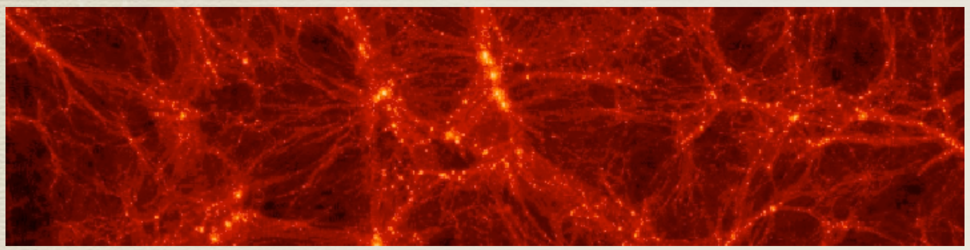
Models → **MCMC**

Range of Reasonable Physical Models


Supernova Cosmology Project
Amanullah, et al., Ap.J. (2010)
No Big Bang
Union2 SN Ia Compilation
SNe
BAO
CMB
Flat
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Dark Energy Density
Dark Matter Density


Supernova Cosmology Project
Amanullah, et al., Ap.J. (2010)
No Big Bang
Union2 SN Ia Compilation
SNe
BAO
CMB
Flat
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
????
????

UniverseMachine



DM Simulation

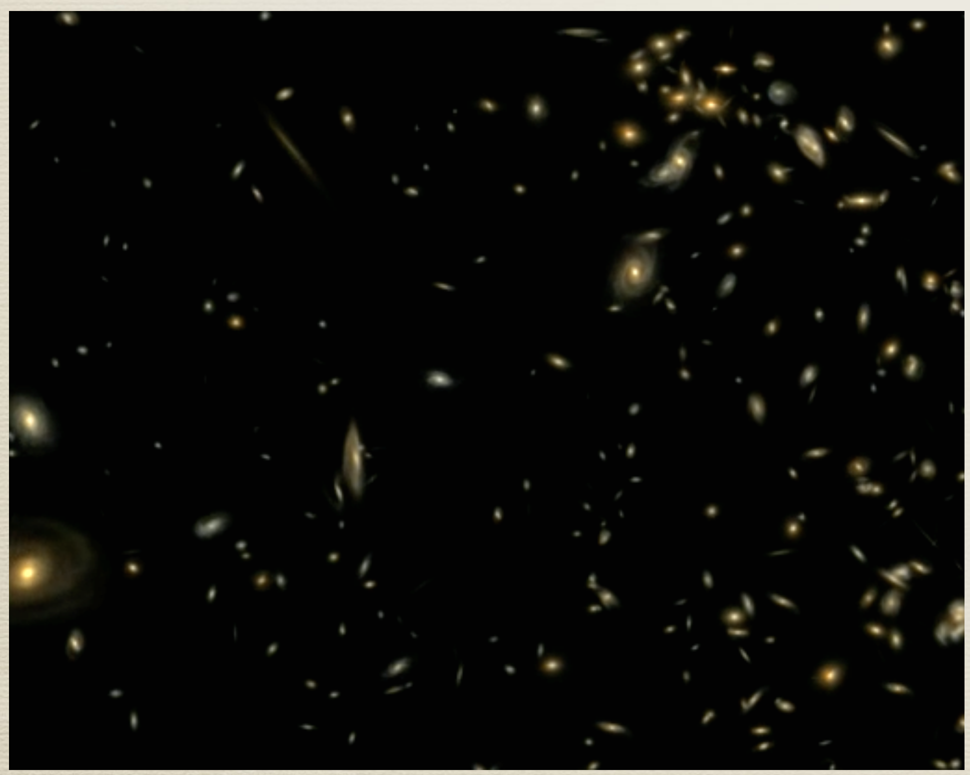
+

$$\mathrm{SFR} = f(M_h, \dot{M}_h, z)$$

=

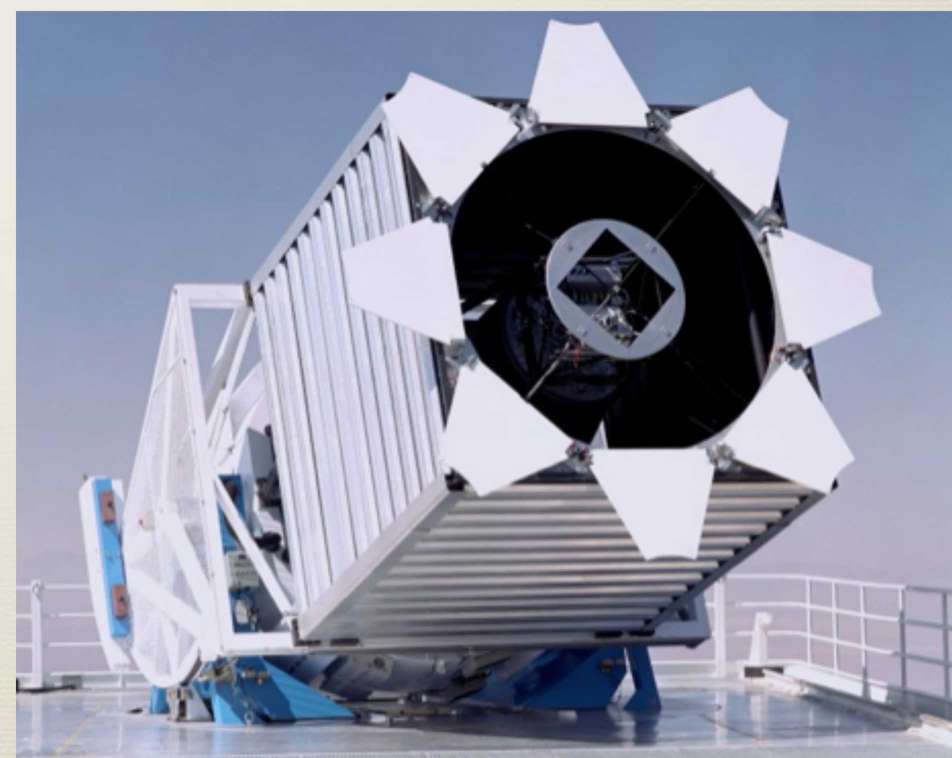
Minimal Galaxy Model

MCMC



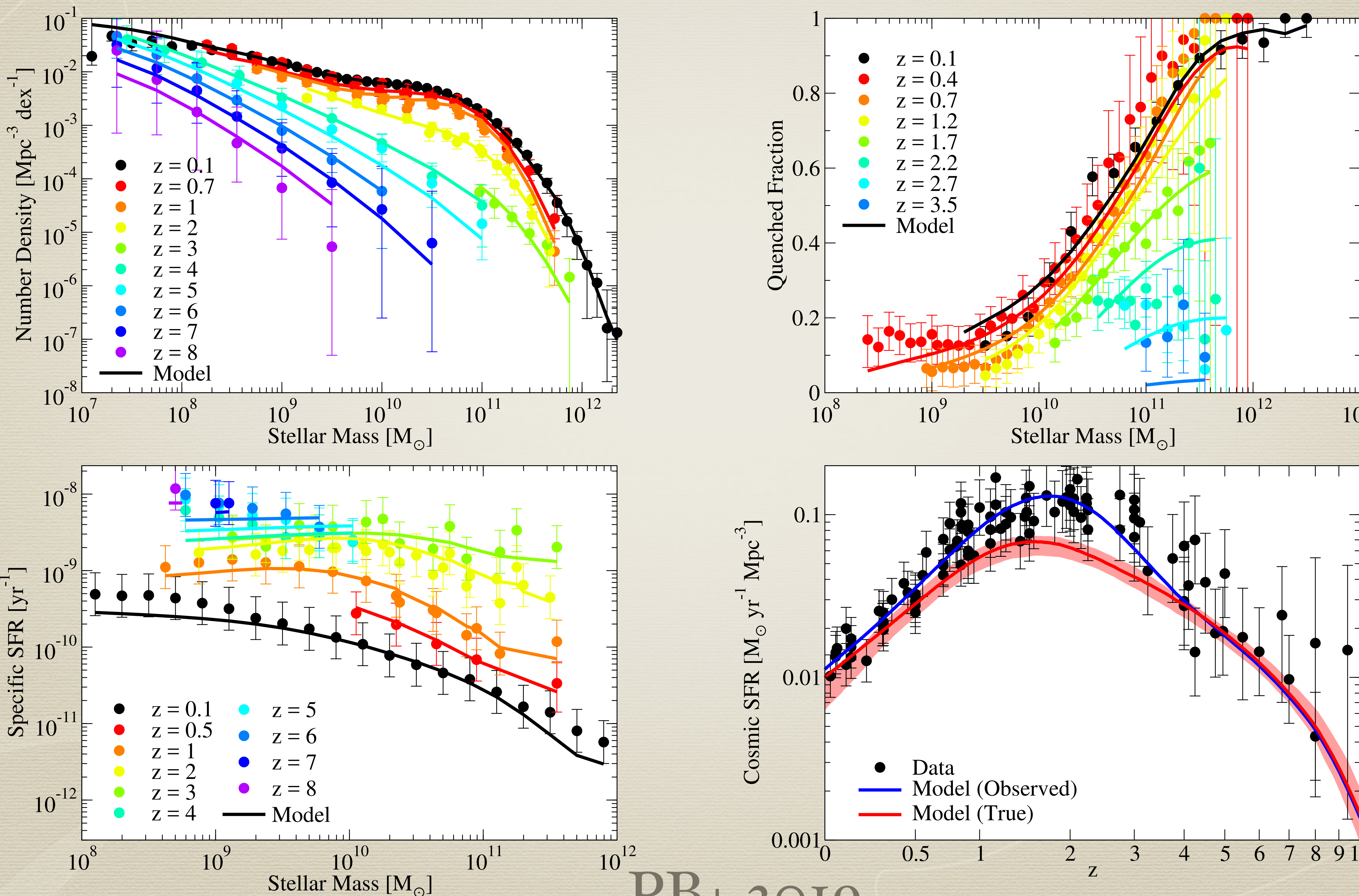
Mock Universe

$\stackrel{?}{=}$



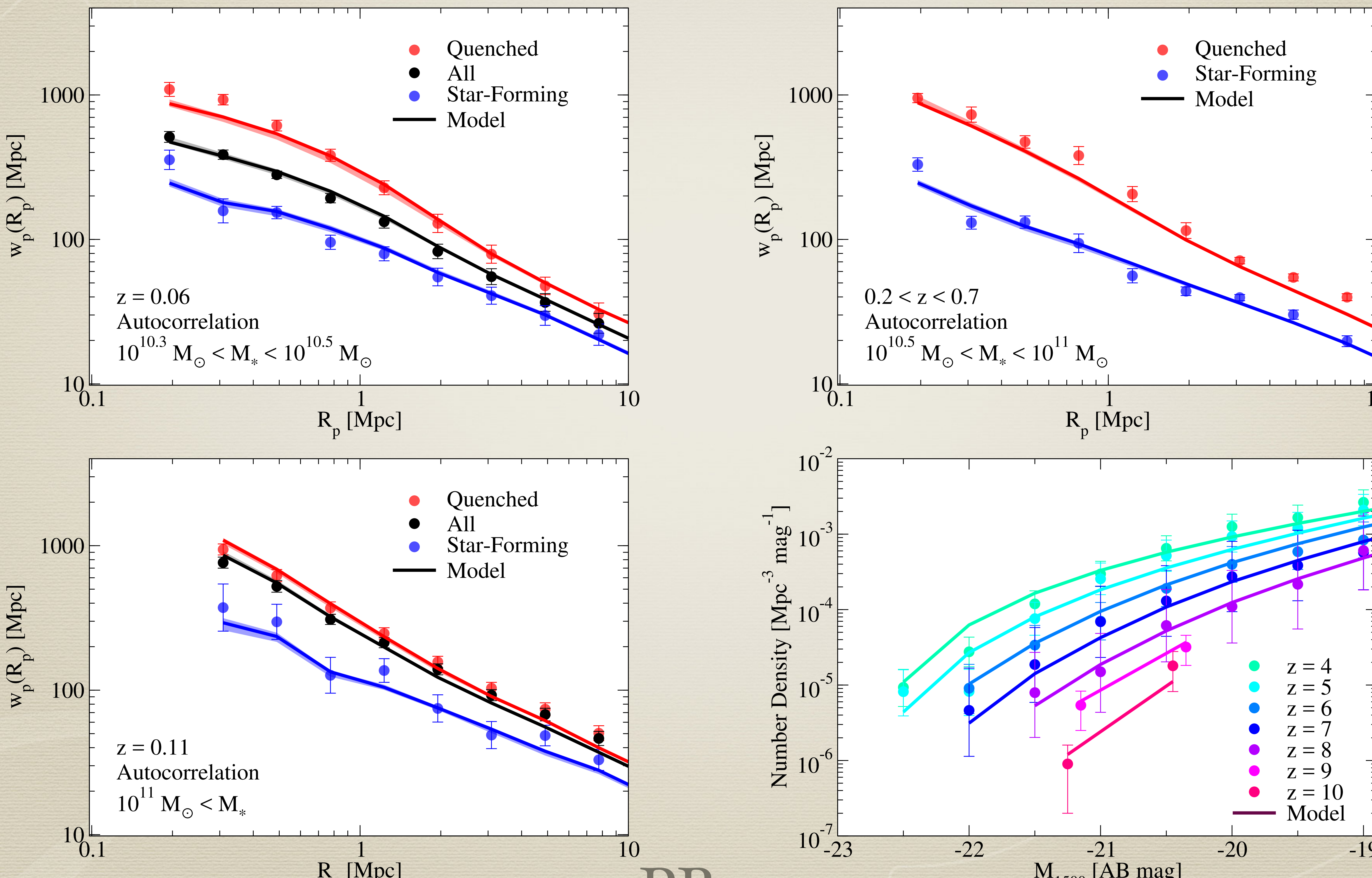
PB+ 2019

Observational Constraints



PB+ 2019

Observational Constraints



PB+ 2019

Why?

Observations

$$\mathcal{L}(x) = \exp(-\chi^2/2)$$

Models → **MCMC**

Range of Reasonable Physical Models

Why?

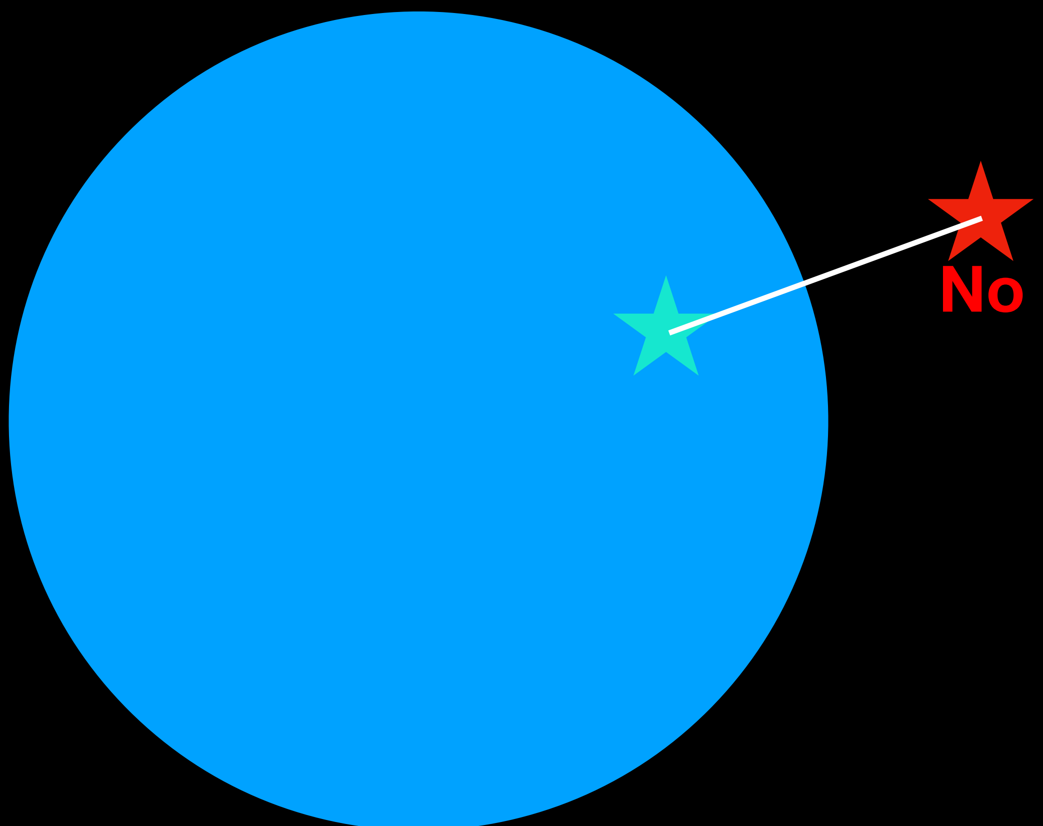
Observations

$$\mathcal{L}(x) = \exp(-\chi^2/2)$$

Models **MCMC**

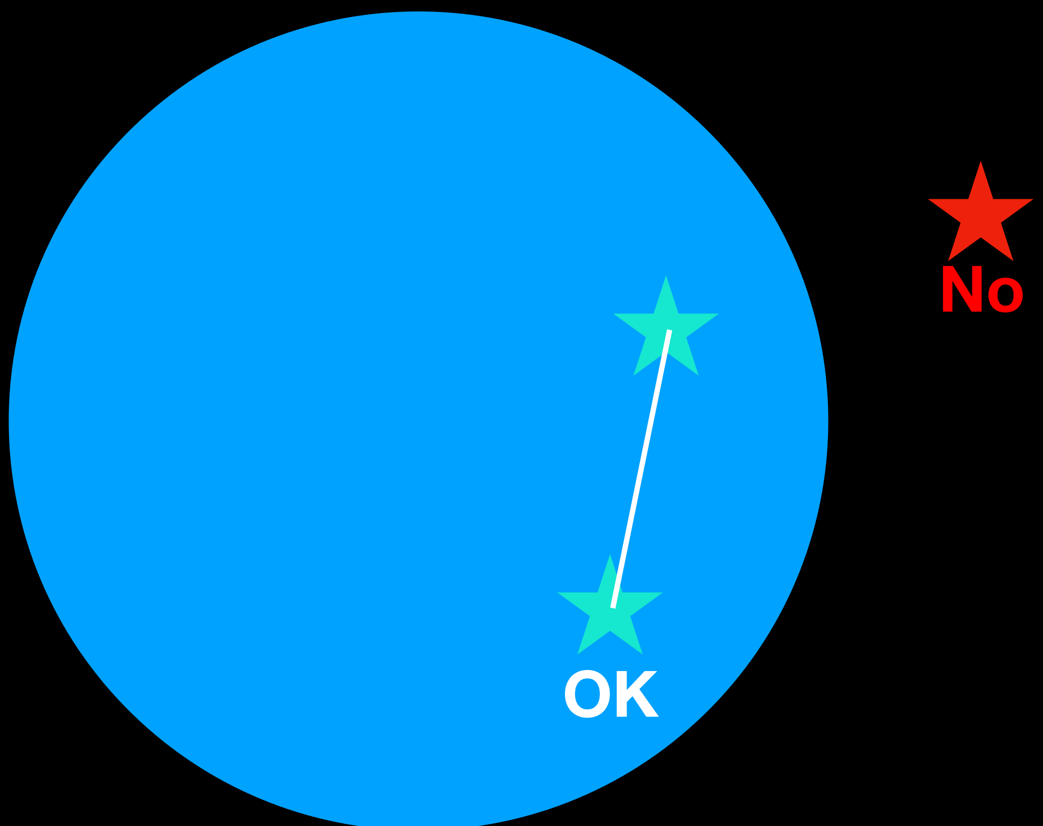
Career in Astronomy

MCMCs



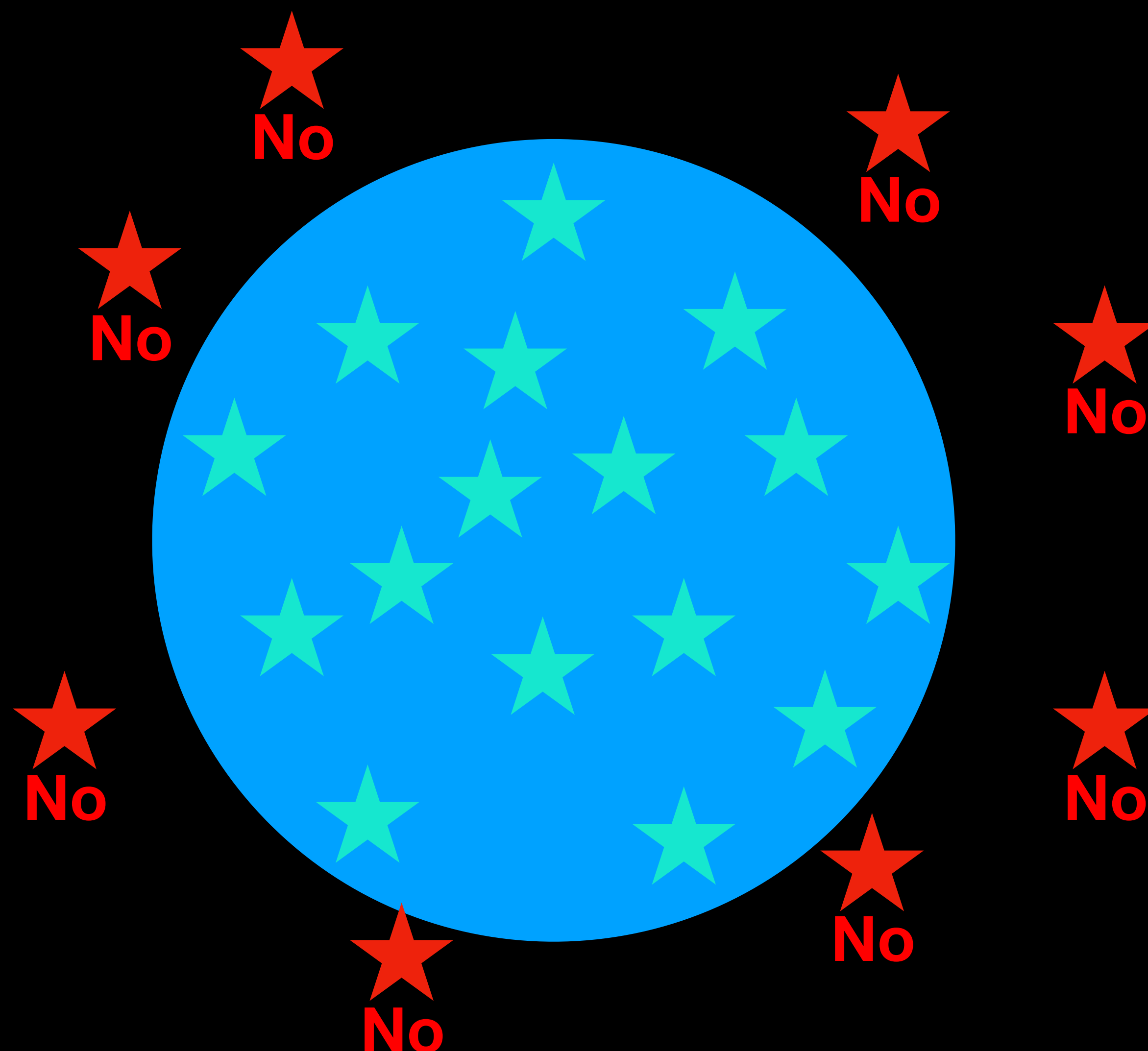
Random walk w/ rejections

MCMCs



Random walk w/ rejections

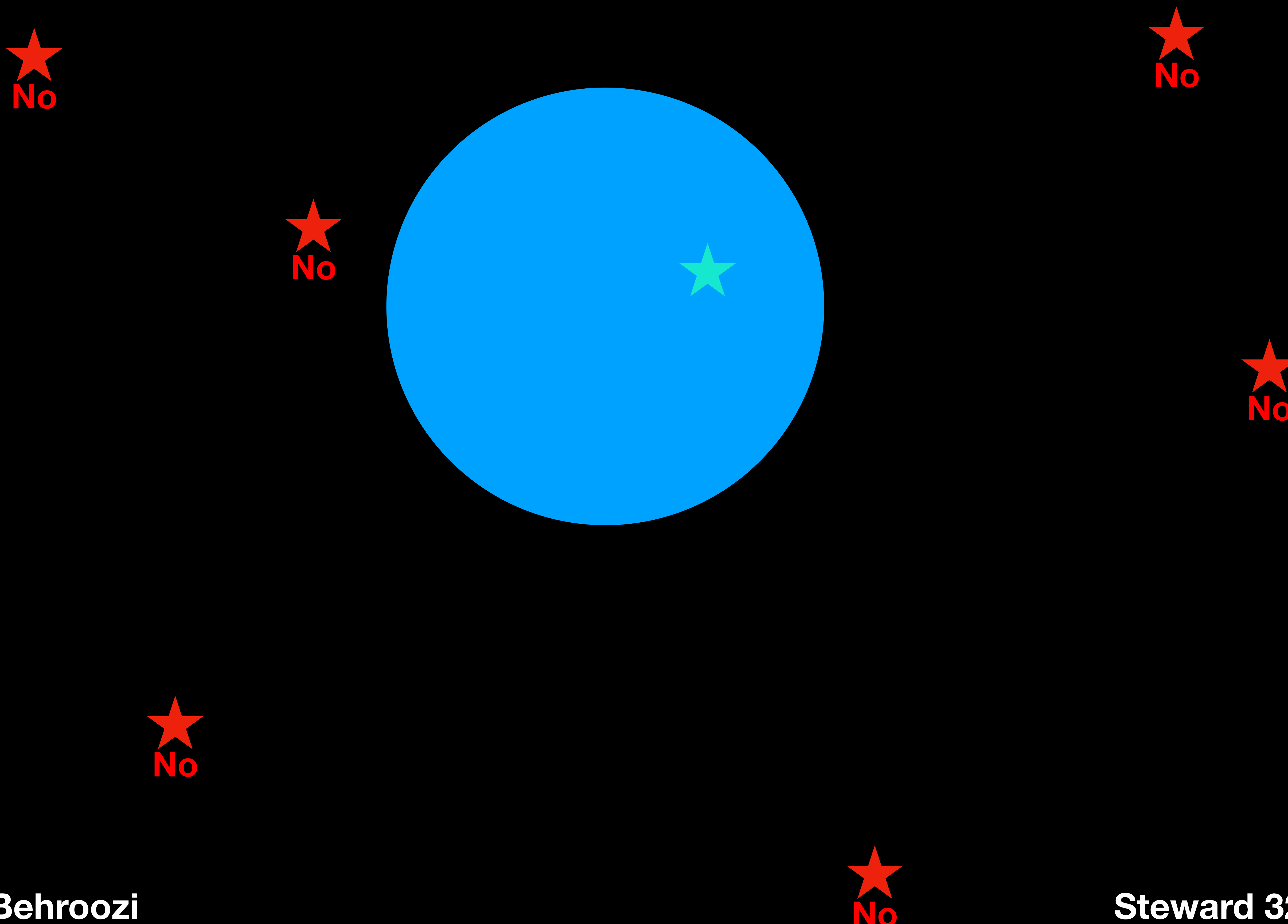
MCMCs



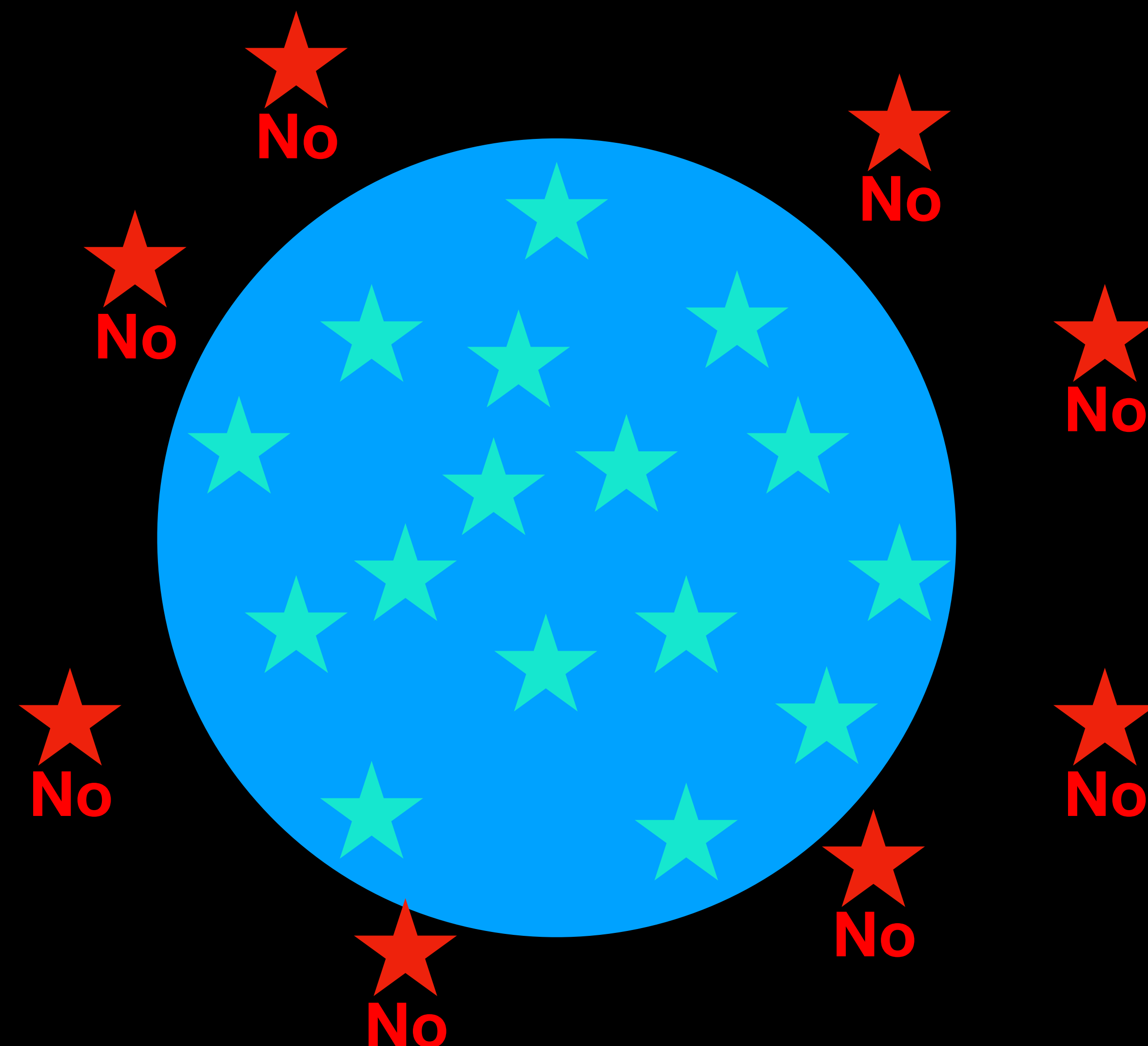
Random walk w/ rejections

MCMC Performance

MCMC Performance

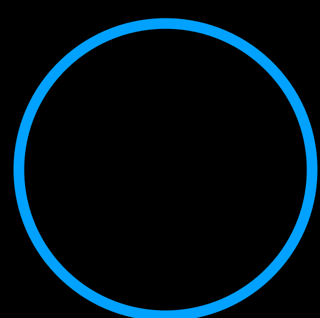


MCMC Performance

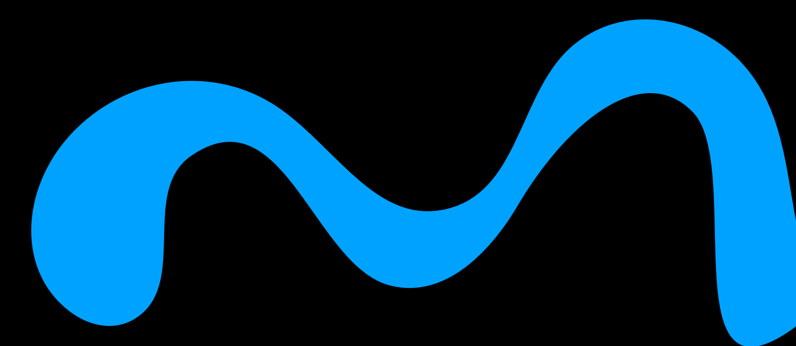


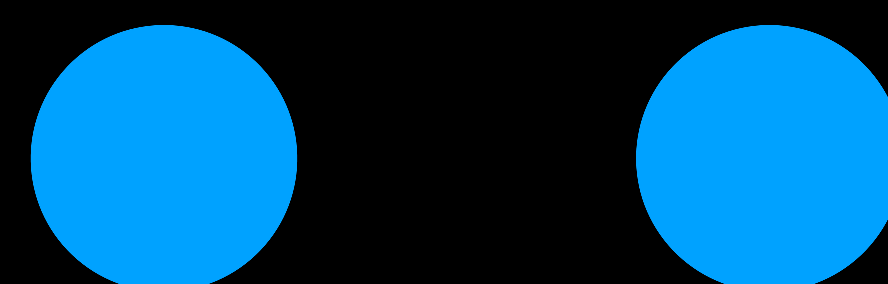
Rapidly degrades when step size is not ~ distribution size.

Which distribution will be hardest to explore with a uniform random walk?

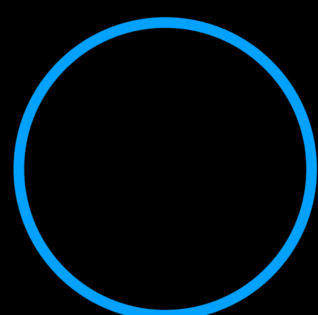
1. 

2. 

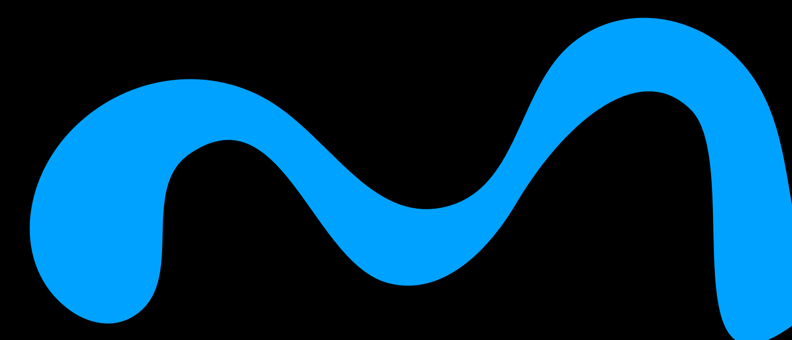
3. 

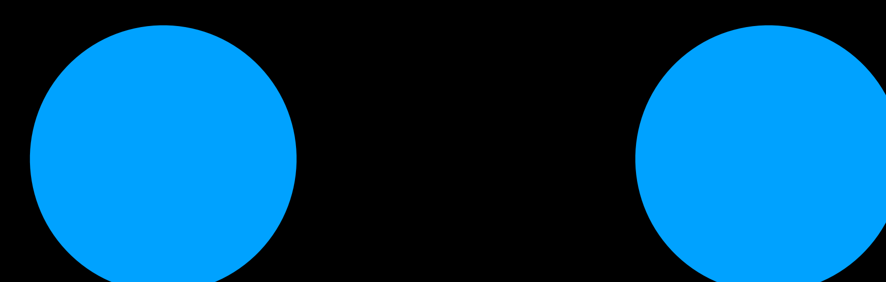
4. 

Which distribution will be second hardest to explore with a uniform random walk?

1. 

2. 

3. 

4. 

Takeaway point:

MCMC algorithms that adapt to the target distribution are vastly more efficient.

Takeaway point:

emcee (stretch-step)
Adaptive Metropolis
NUTS (Hamiltonian)

Hamiltonian Sampling Demo

Hamiltonian Sampling Demo

$$U(\vec{x}) = \frac{1}{2}\chi^2(\vec{x}) \qquad\qquad T(\vec{v}) = \frac{1}{2}v^2$$

Hamiltonian Sampling Demo

$$U(\vec{x}) = \frac{1}{2}\chi^2(\vec{x}) \qquad T(\vec{v}) = \frac{1}{2}v^2$$

$$\mathcal{H}(\vec{x},\vec{v}) = U(\vec{x}) + T(\vec{v})$$

Hamiltonian Sampling Demo

$$U(\vec{x}) = \frac{1}{2}\chi^2(\vec{x}) \qquad\qquad T(\vec{v}) = \frac{1}{2}v^2$$

$$\mathcal{H}(\vec{x}, \vec{v}) = U(\vec{x}) + T(\vec{v})$$

$$\mathcal{L}(\vec{x}, \vec{v}) = \exp(-\mathcal{H}(\vec{x}, \vec{v}))$$

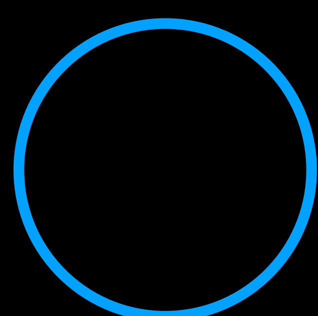
Hamiltonian Sampling Demo

$$U(\vec{x}) = \frac{1}{2}\chi^2(\vec{x}) \qquad T(\vec{v}) = \frac{1}{2}v^2$$

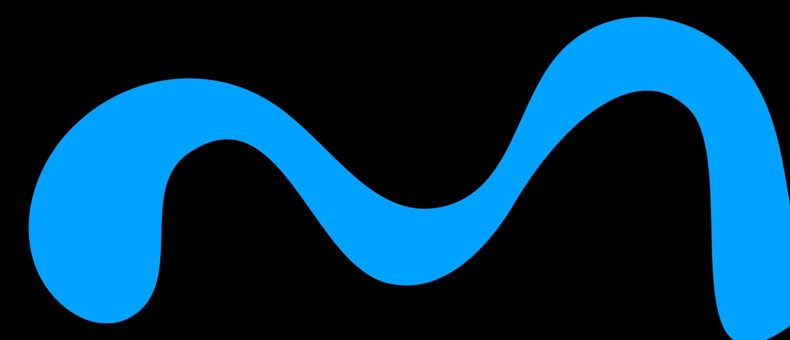
$$\mathcal{H}(\vec{x},\vec{v}) = U(\vec{x}) + T(\vec{v})$$
$$\mathcal{L}(\vec{x},\vec{v}) = \exp(-\mathcal{H}(\vec{x},\vec{v}))$$

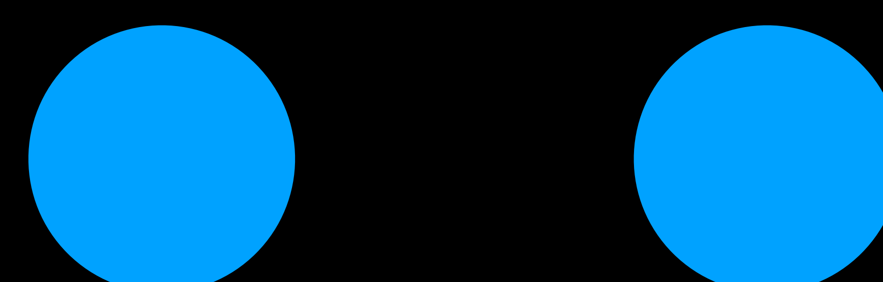
$$\frac{\partial\vec{x}}{\partial t} = \nabla T = \vec{v} \qquad \frac{\partial\vec{v}}{\partial t} = -\nabla U = -\frac{1}{2}\nabla\chi^2$$

Which distribution will be second hardest to explore with a uniform random walk?

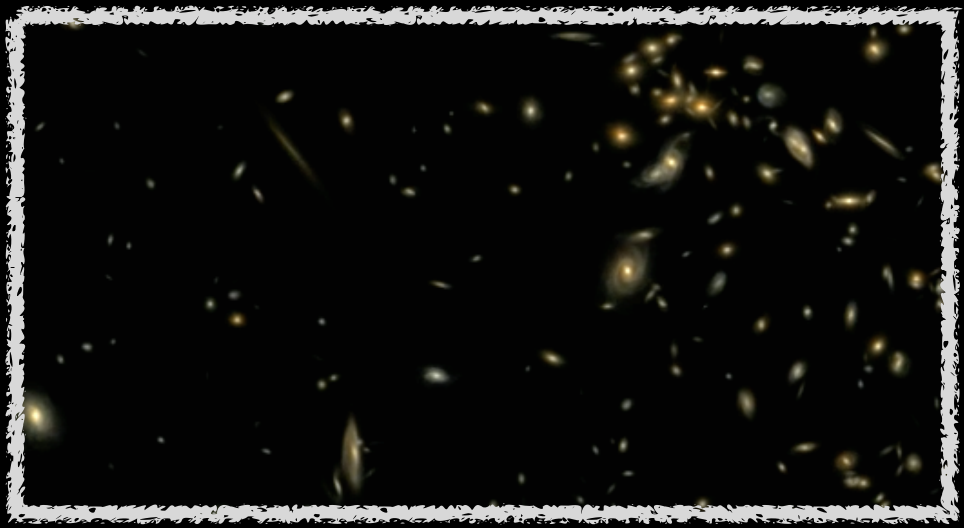
1. 

2. 

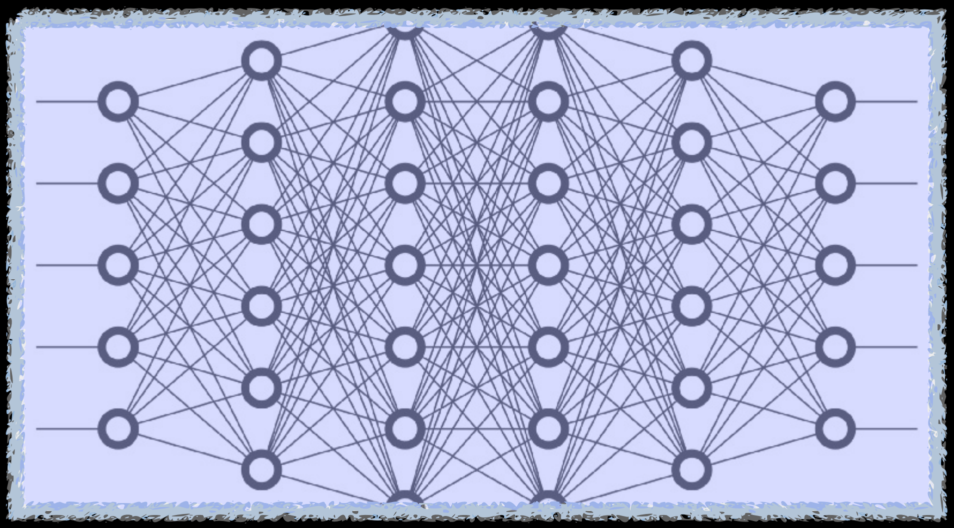
3. 

4. 

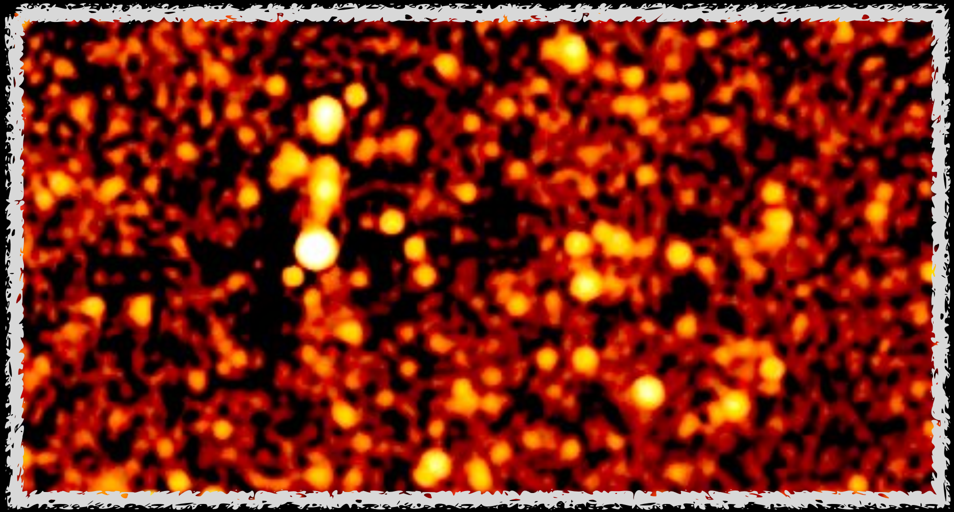
High-Dimensional Problems



Empirical Galaxy Modeling
40-70 dimensions



Deep Neural Networks
50-10,000 dimensions

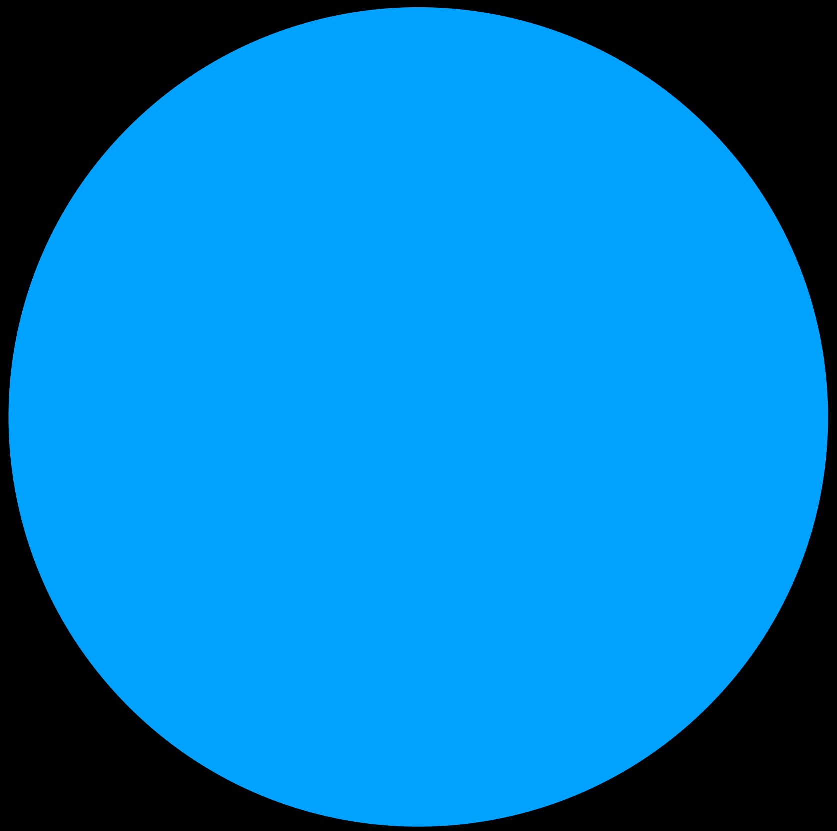


Confused Source Recovery
100 - 10^9 dimensions

Constrained Realizations
> 10^9 dimensions

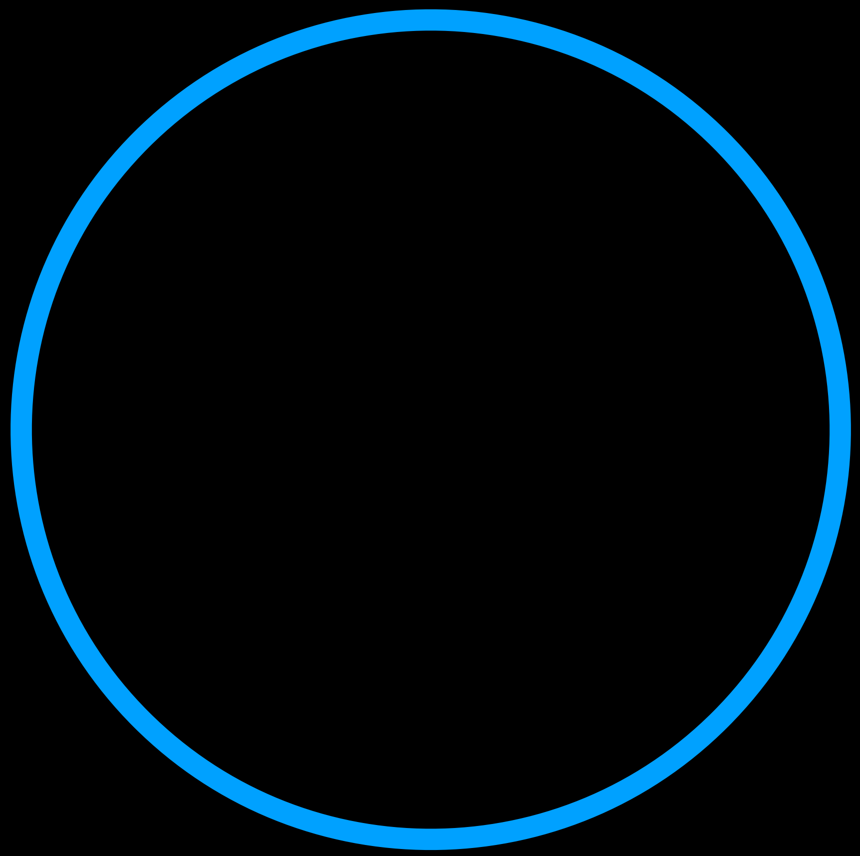
Image Credits: SDSS, Radu Raicea, Jim Condon, CLUES

High-Dimensional Problems



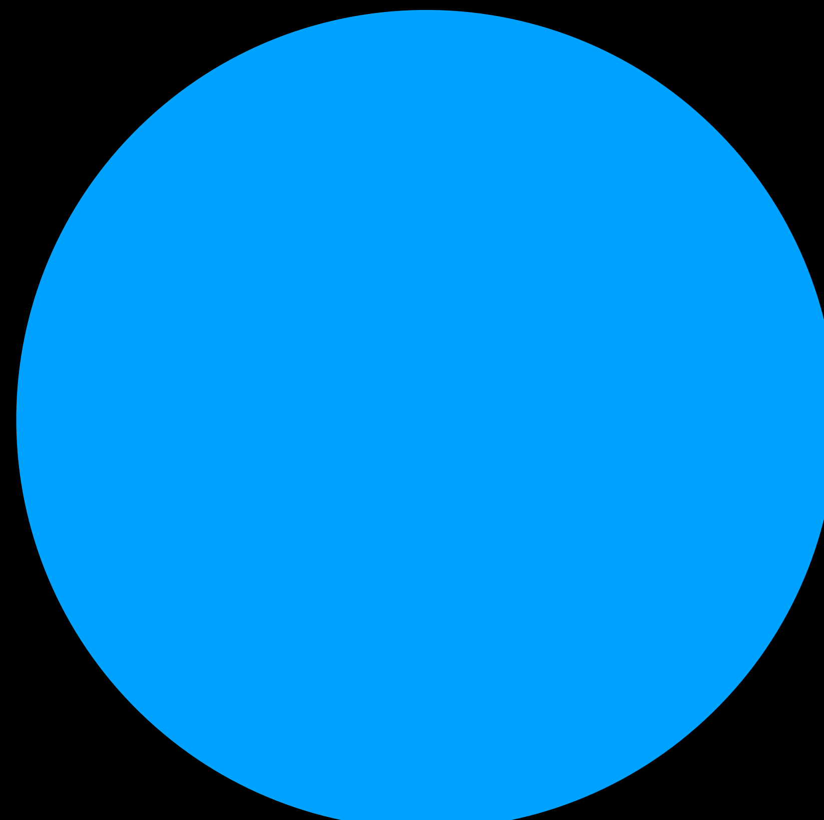
In 3D: 5% change in radius = 15% change in volume.

High-Dimensional Problems



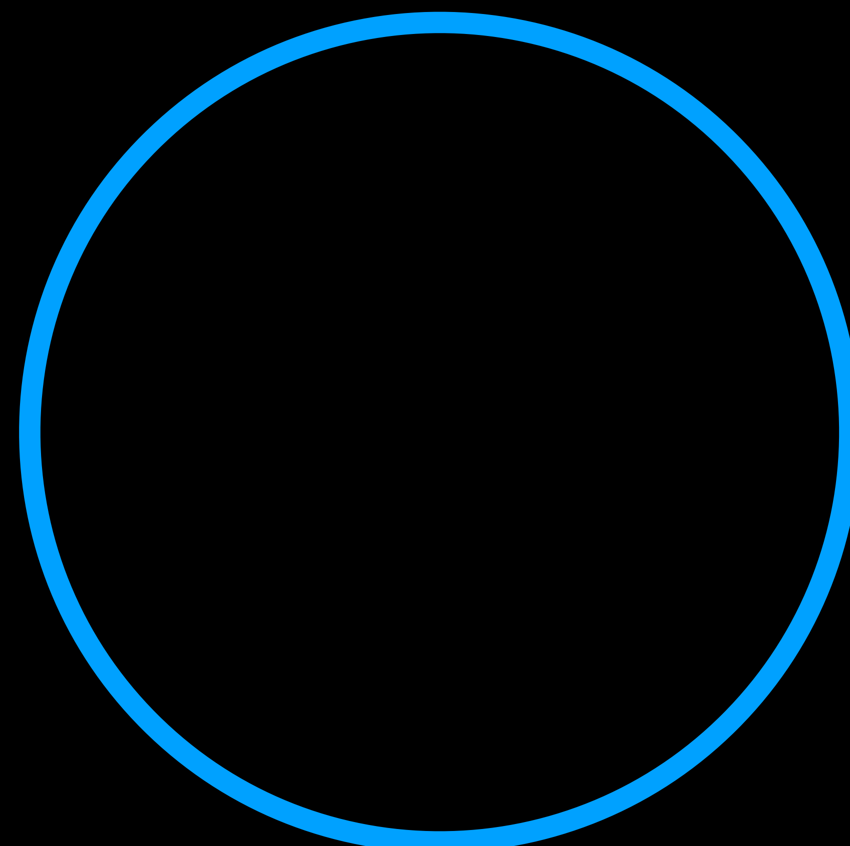
In 20D: 5% change in radius = 165% change in volume.

High-Dimensional Problems



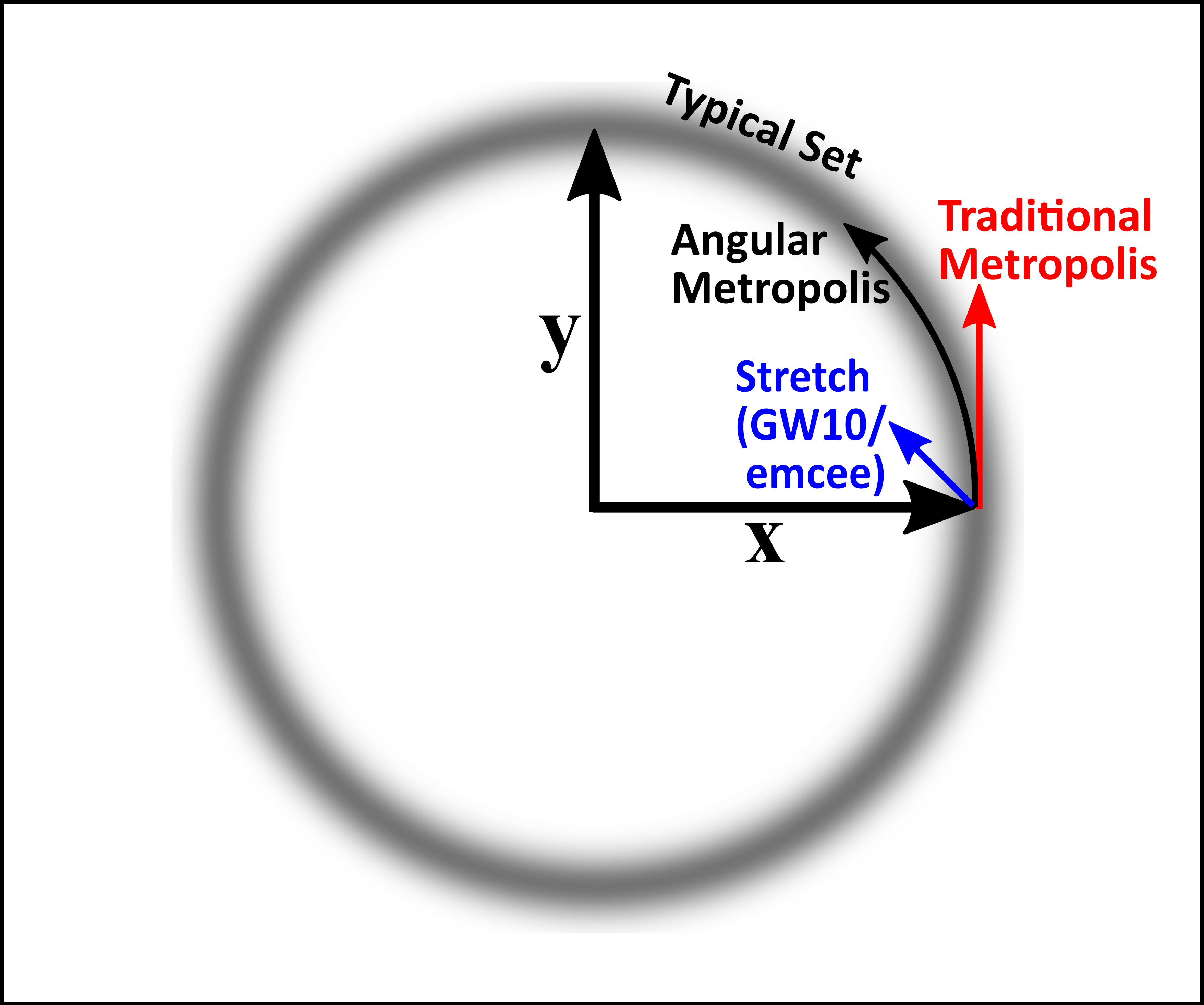
In 3D: covariance matrix has 6 parameters.

High-Dimensional Problems

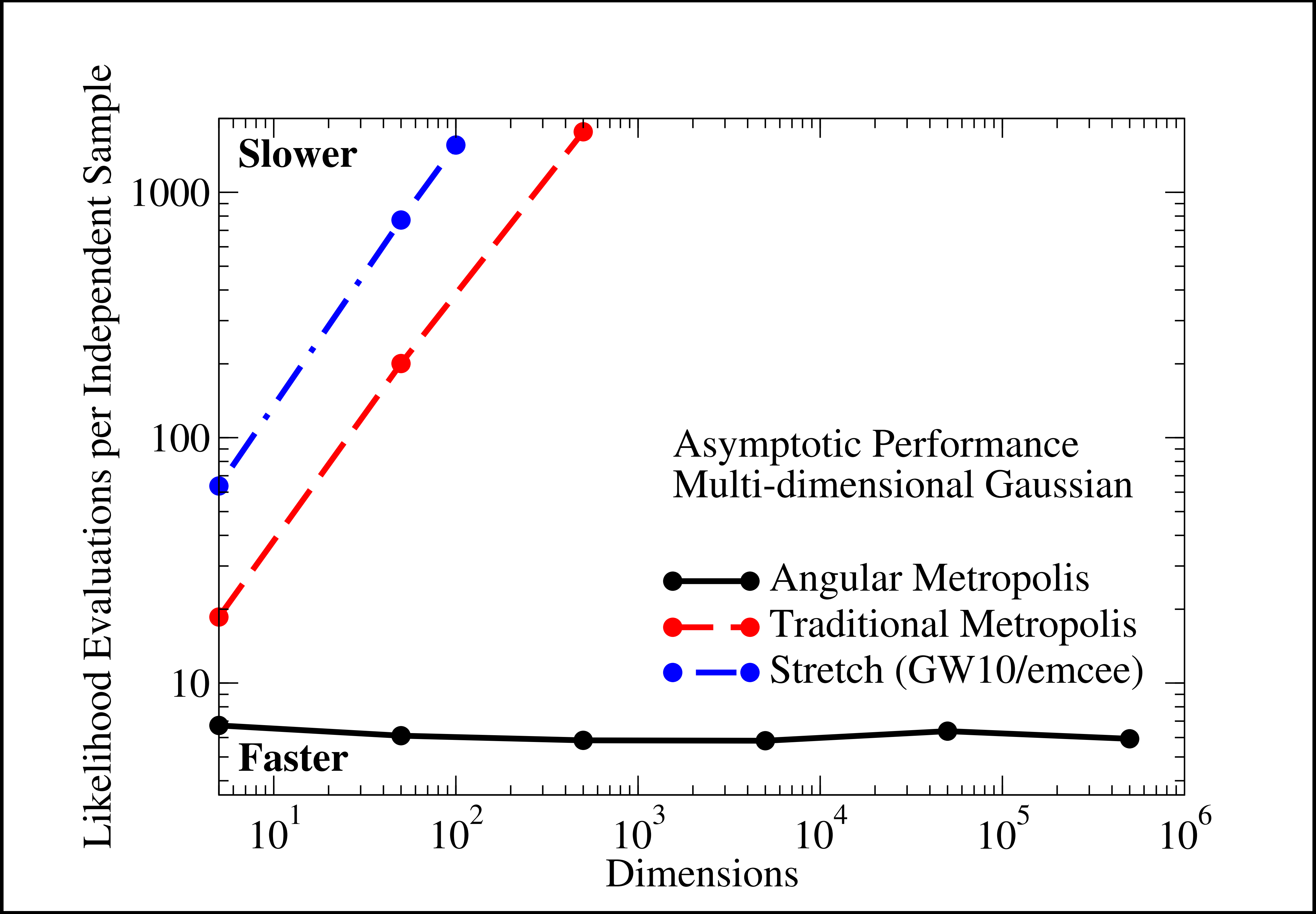


In 20D: covariance matrix has 210 parameters (D*(D+1)/2).

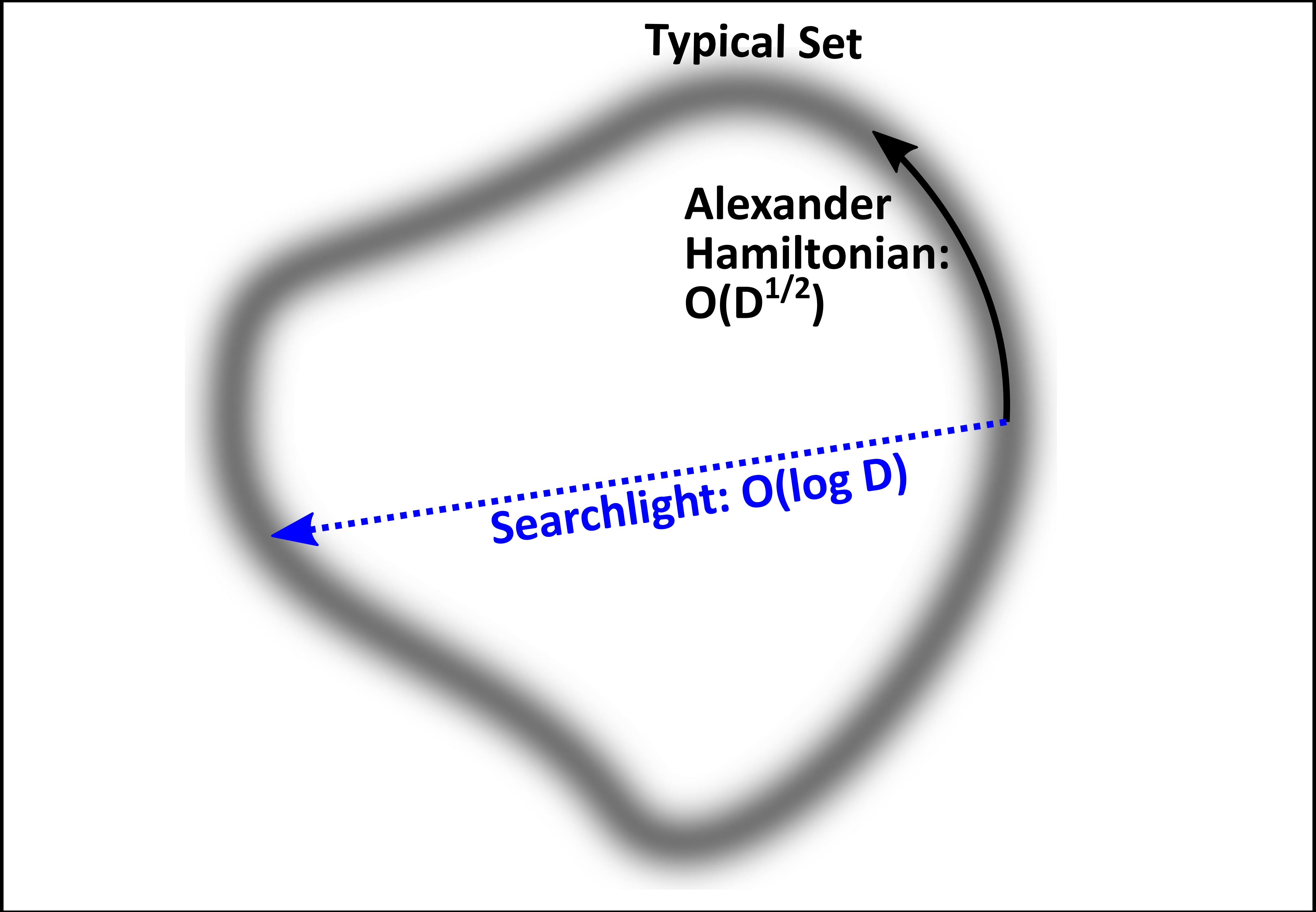
(Future) Solutions



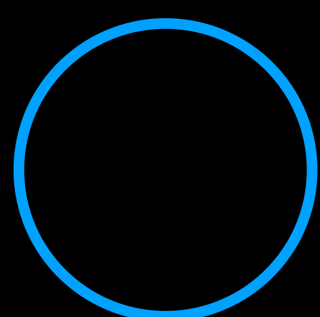
(Future) Solutions



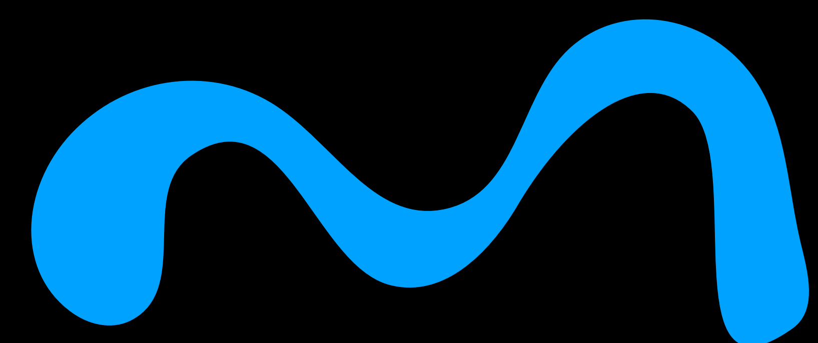
(Future) Solutions

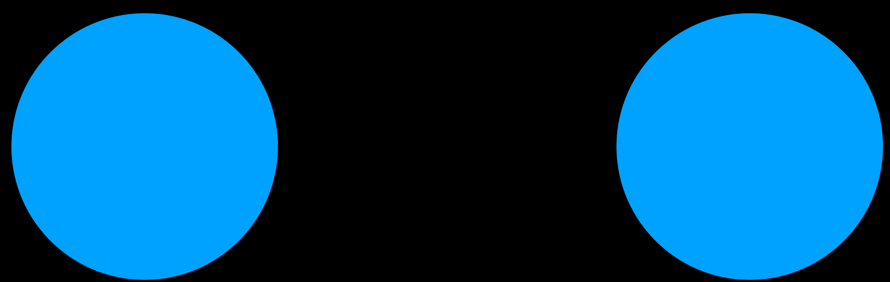


Which distribution will be hardest to explore with a uniform random walk?

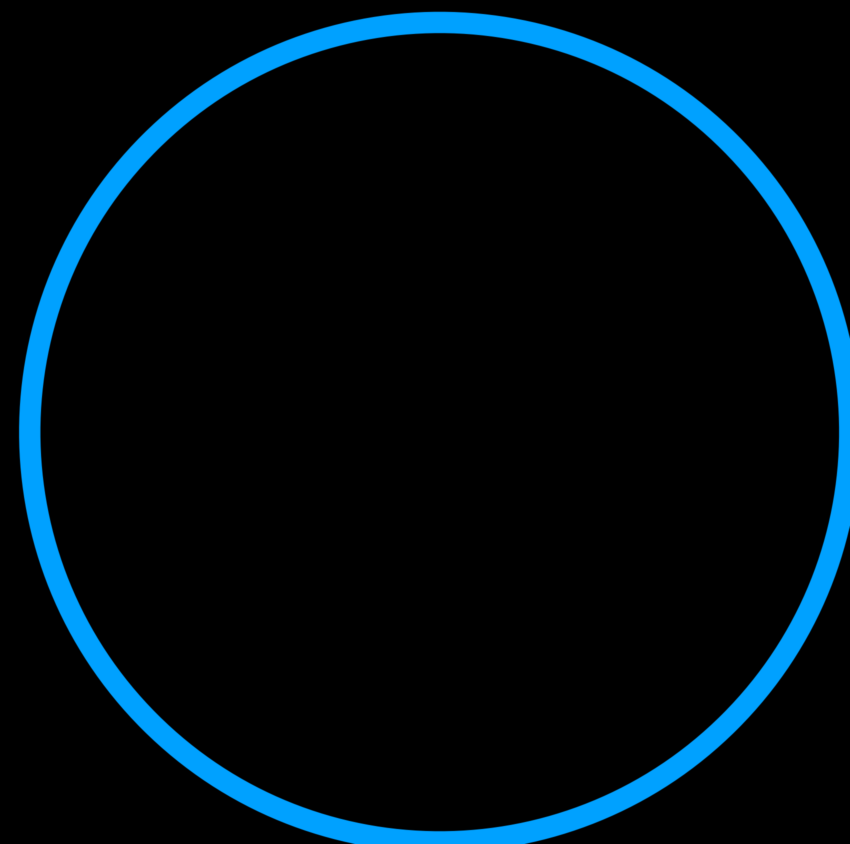
1. 

2. 

3. 

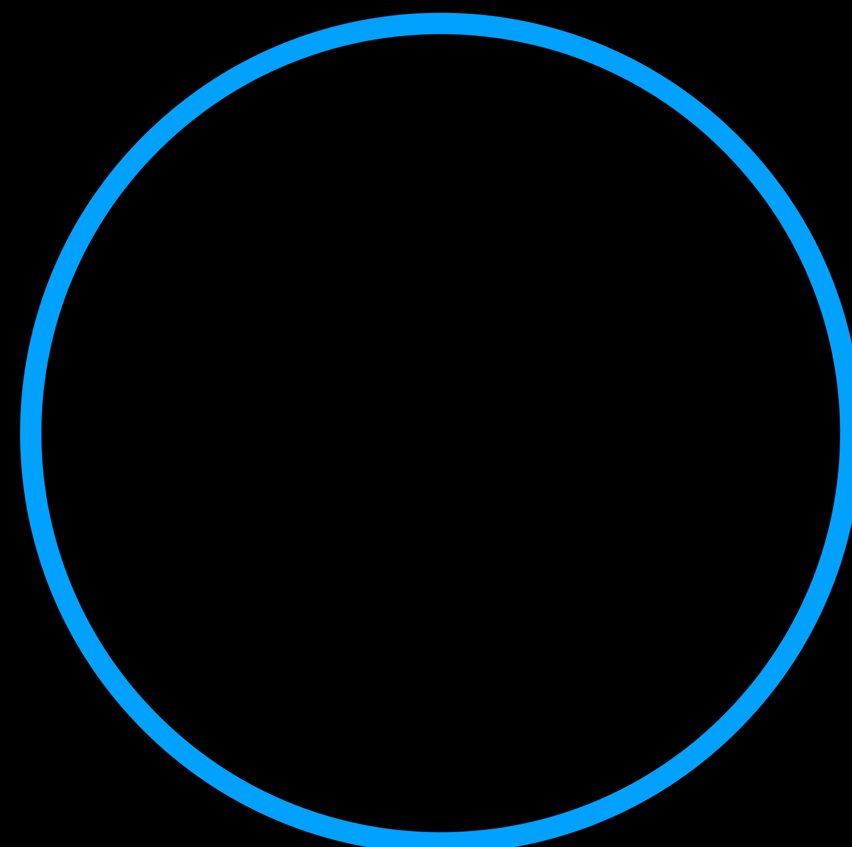
4. 

High-Dimensional Problems



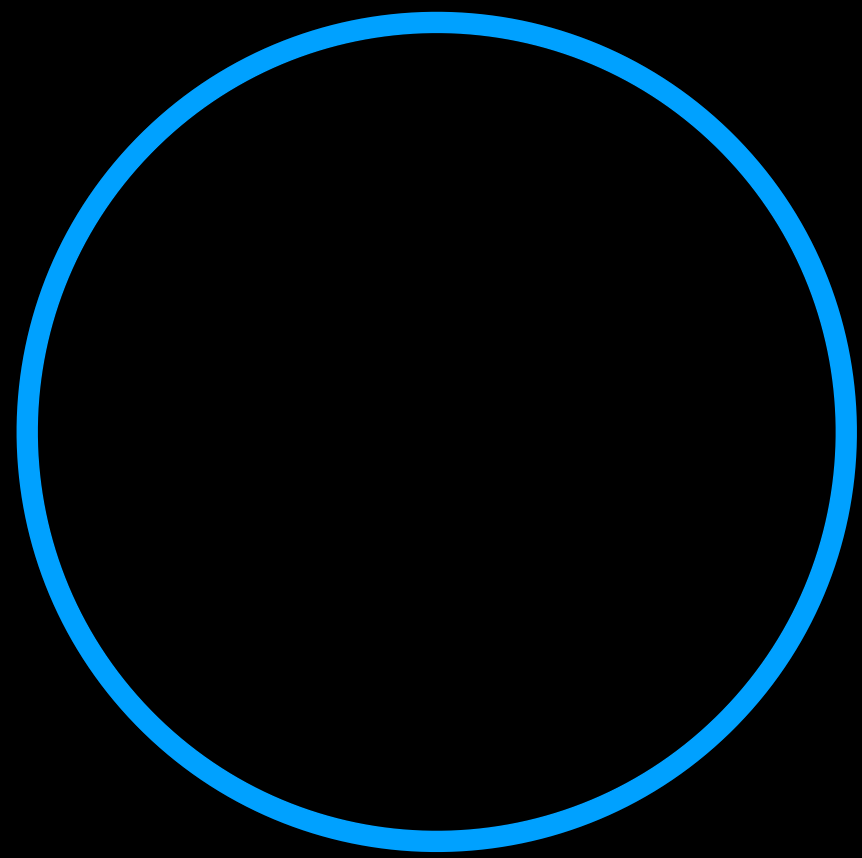
Use another algorithm (e.g. from scipy.optimize) to find minimum first.

High-Dimensional Problems



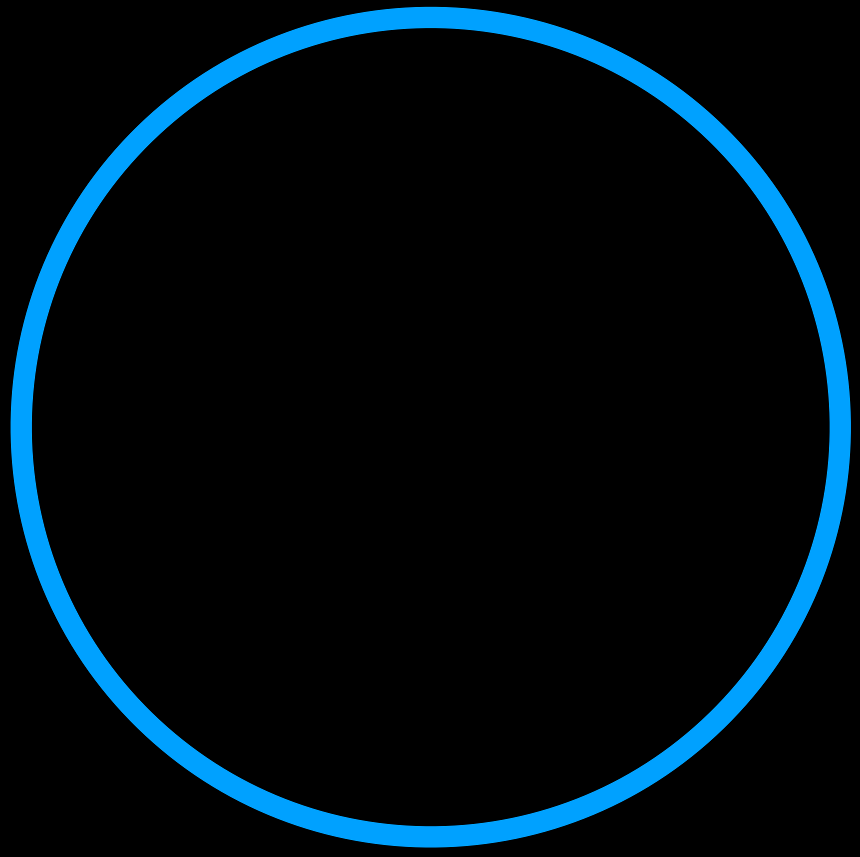
Estimate covariance matrix around minimum using the Fisher Information Matrix.

High-Dimensional Problems



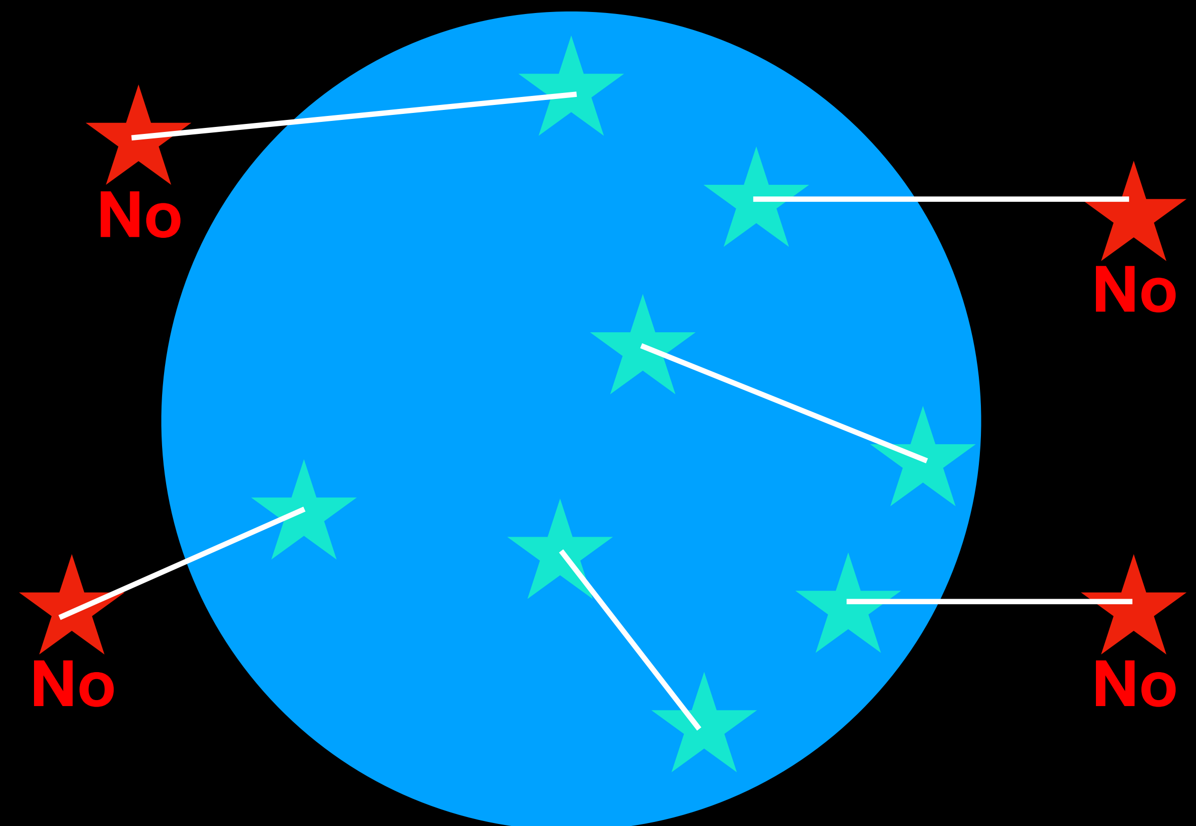
Use a Gaussian distribution with the estimated covariance matrix as starting points.

High-Dimensional Problems



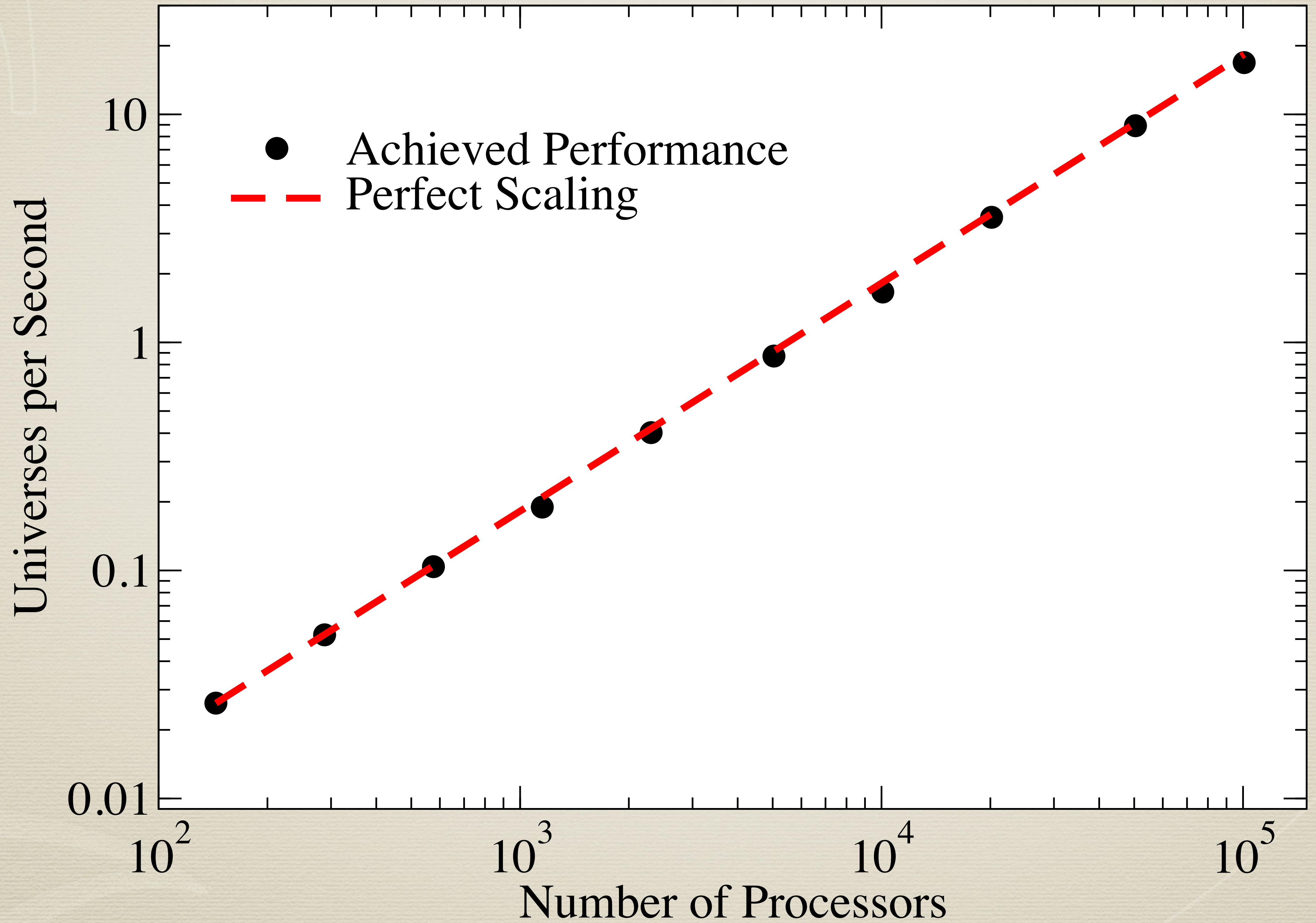
Use parallel sampling.

Parallel Sampling



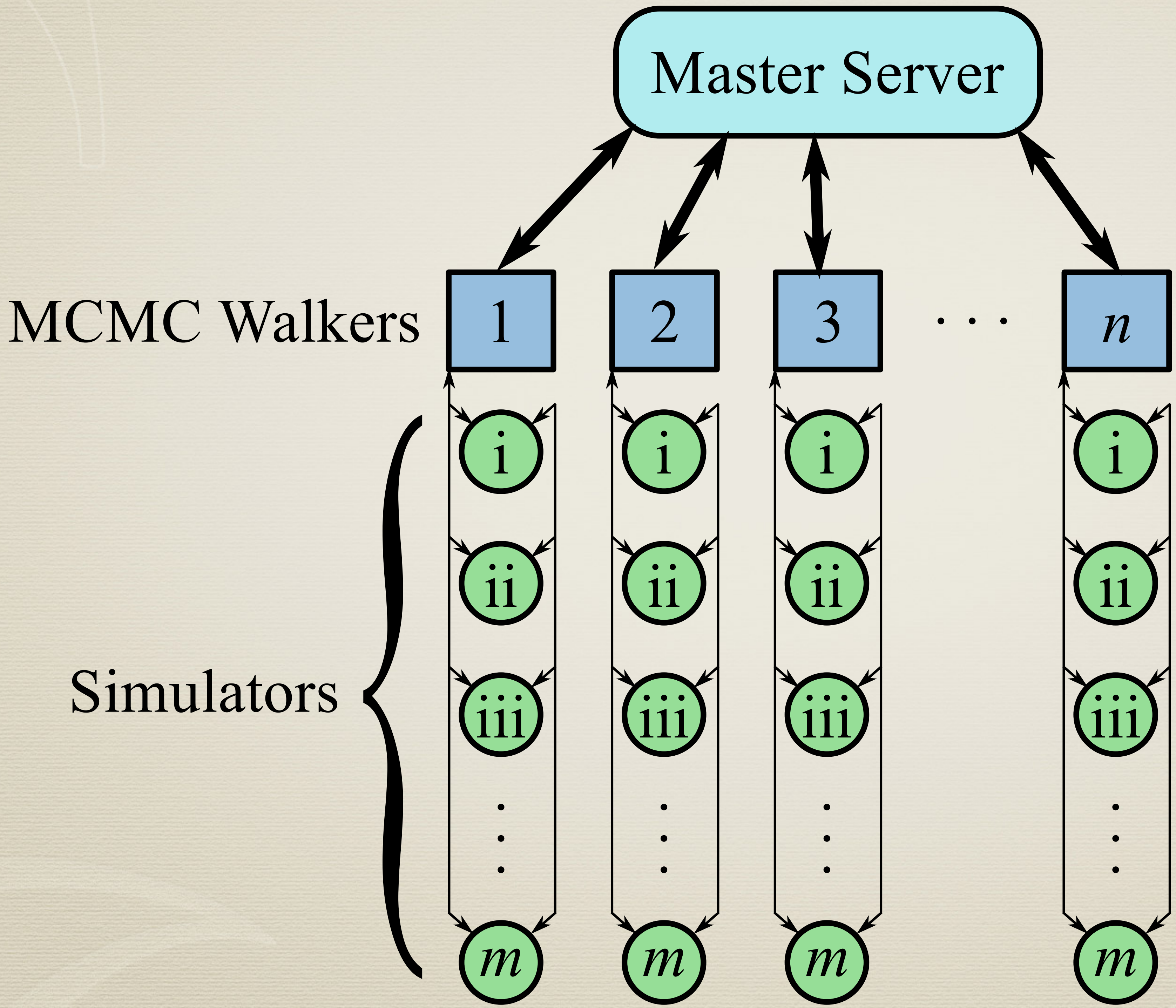
Multiple walkers generating multiple MCMC chains.

Universe Machine



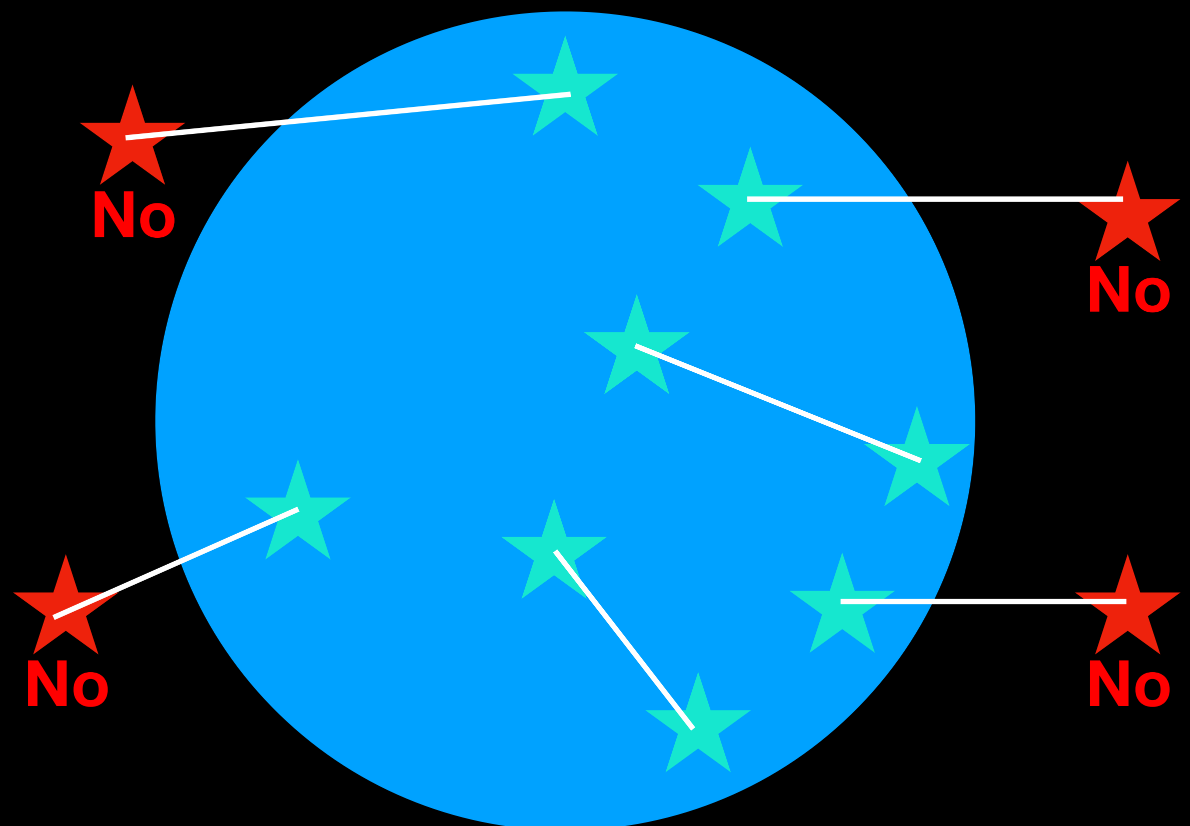
PB+ 2019

Universe Machine

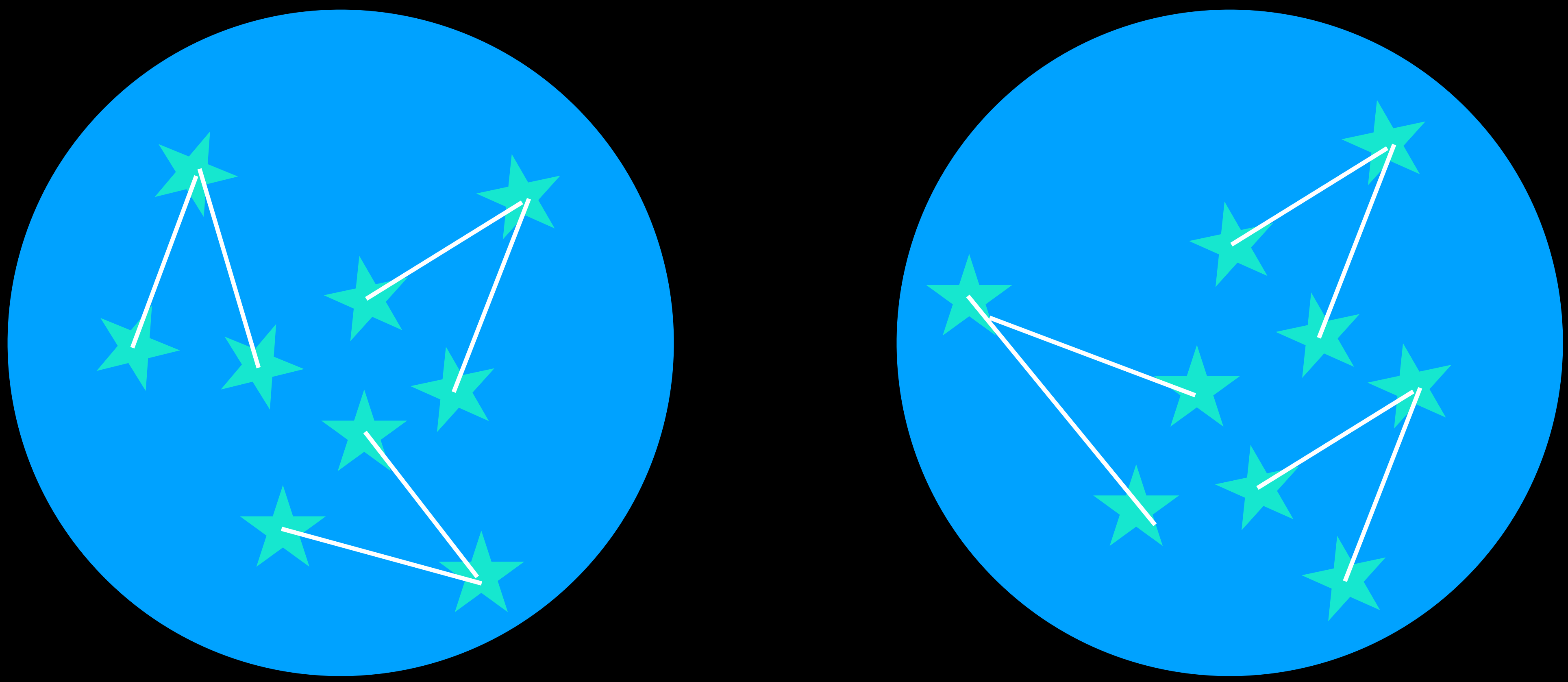


PB+ 2019

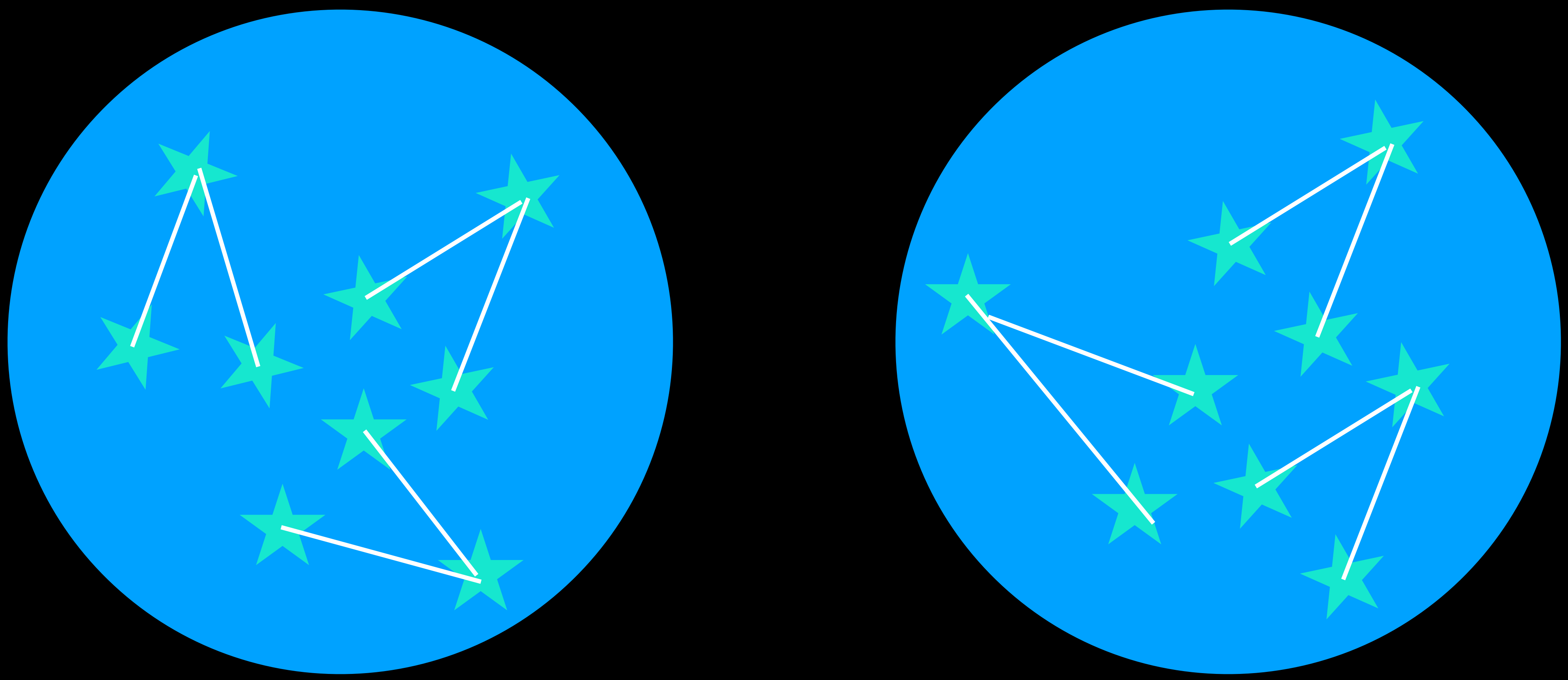
Convergence



Convergence



Convergence



Gelman-Rubin ~ $\frac{\text{Total SD}}{\text{Within-chain SD}}$

Summary

MCMCs help convert obs. to physics.

Use adaptive algorithms

Good algorithms exist in <5D;
not so much (currently) for >10D

Testing for convergence is generally hard, but easier if you parallel sample.

come talk to me if you have MCMC questions

(behroozi@email.arizona.edu)

Peter Behroozi

Steward 326