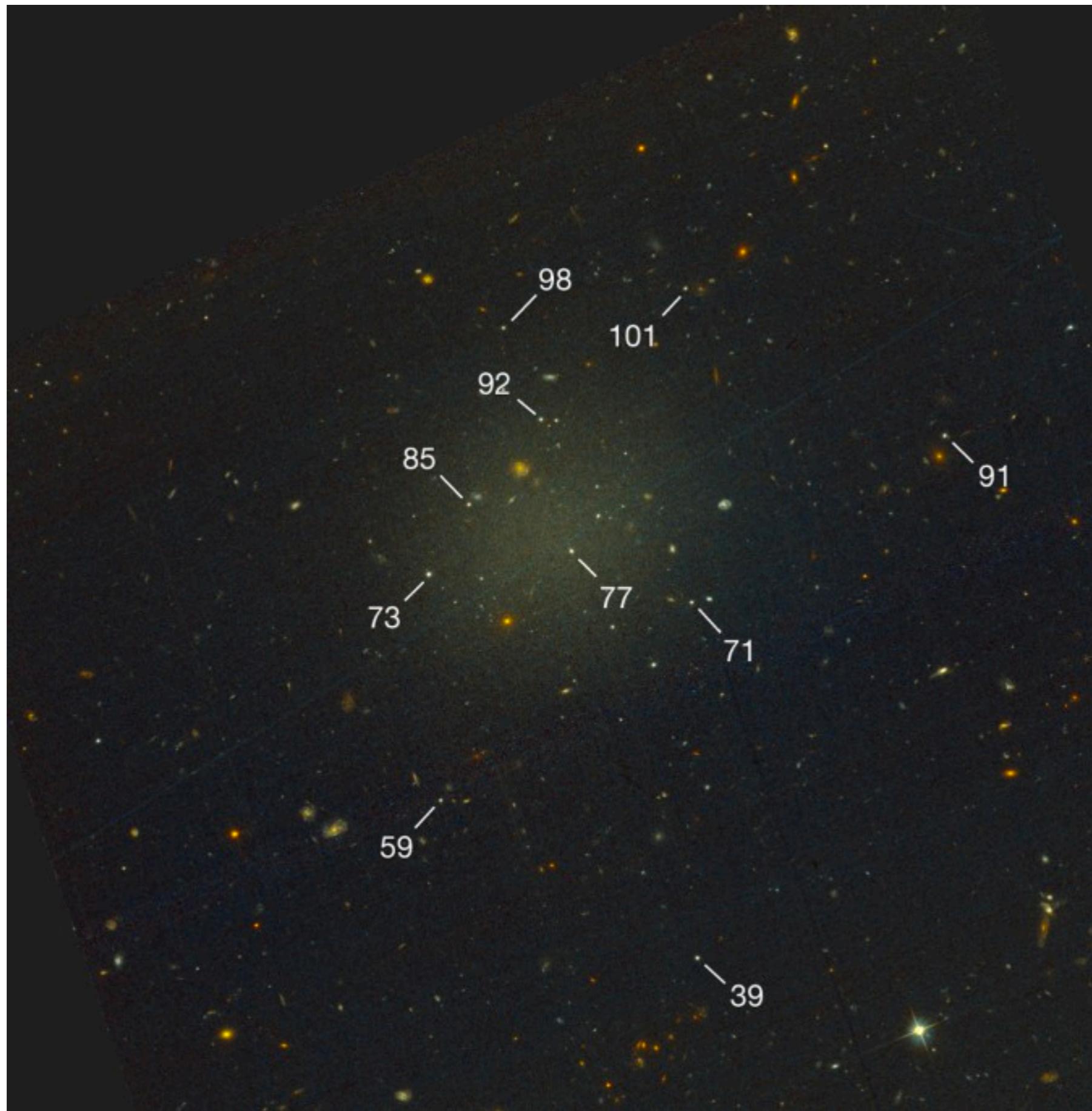


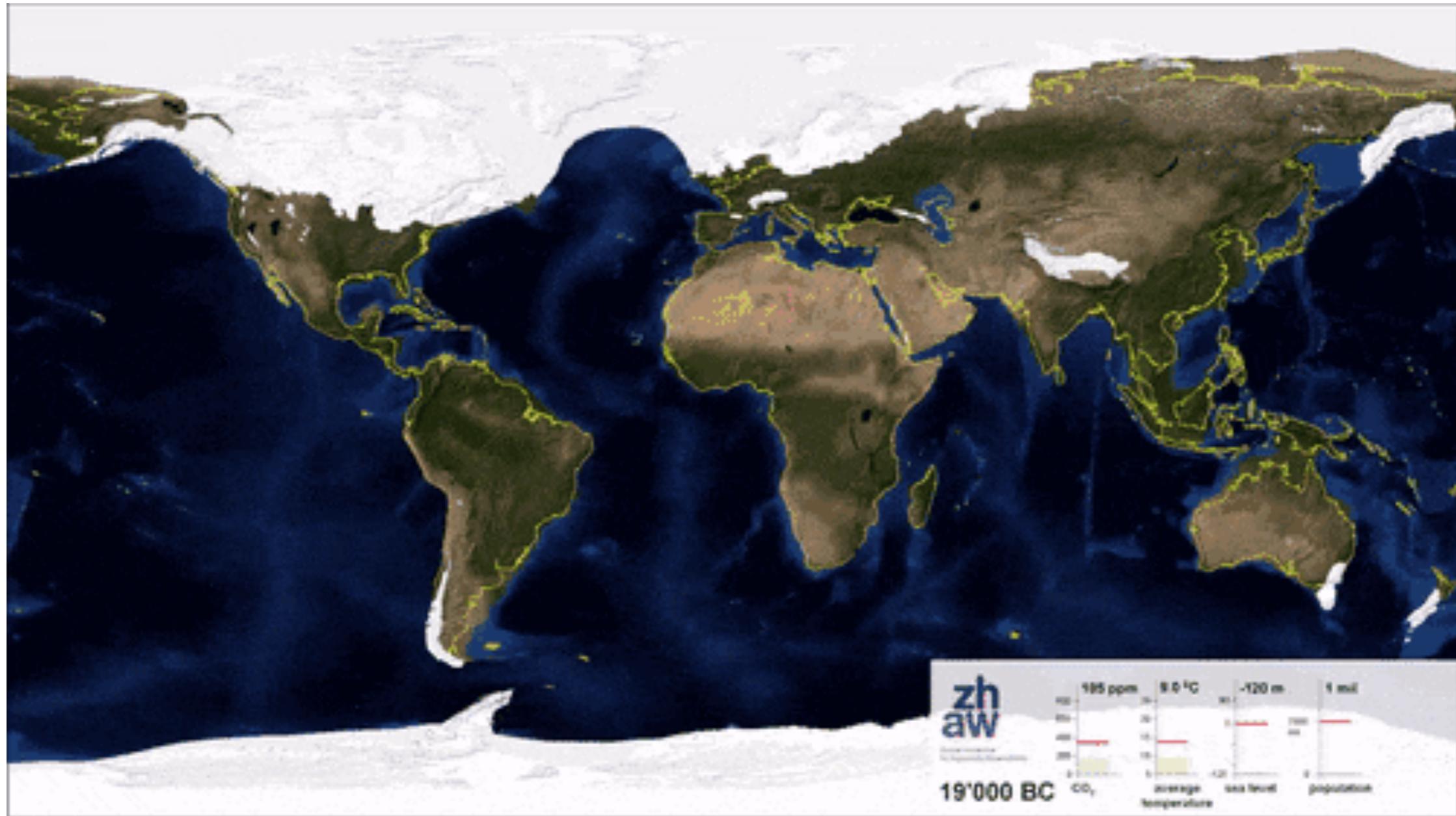


Lecture 22: Markov Chain Monte Carlo (update)

MCMC

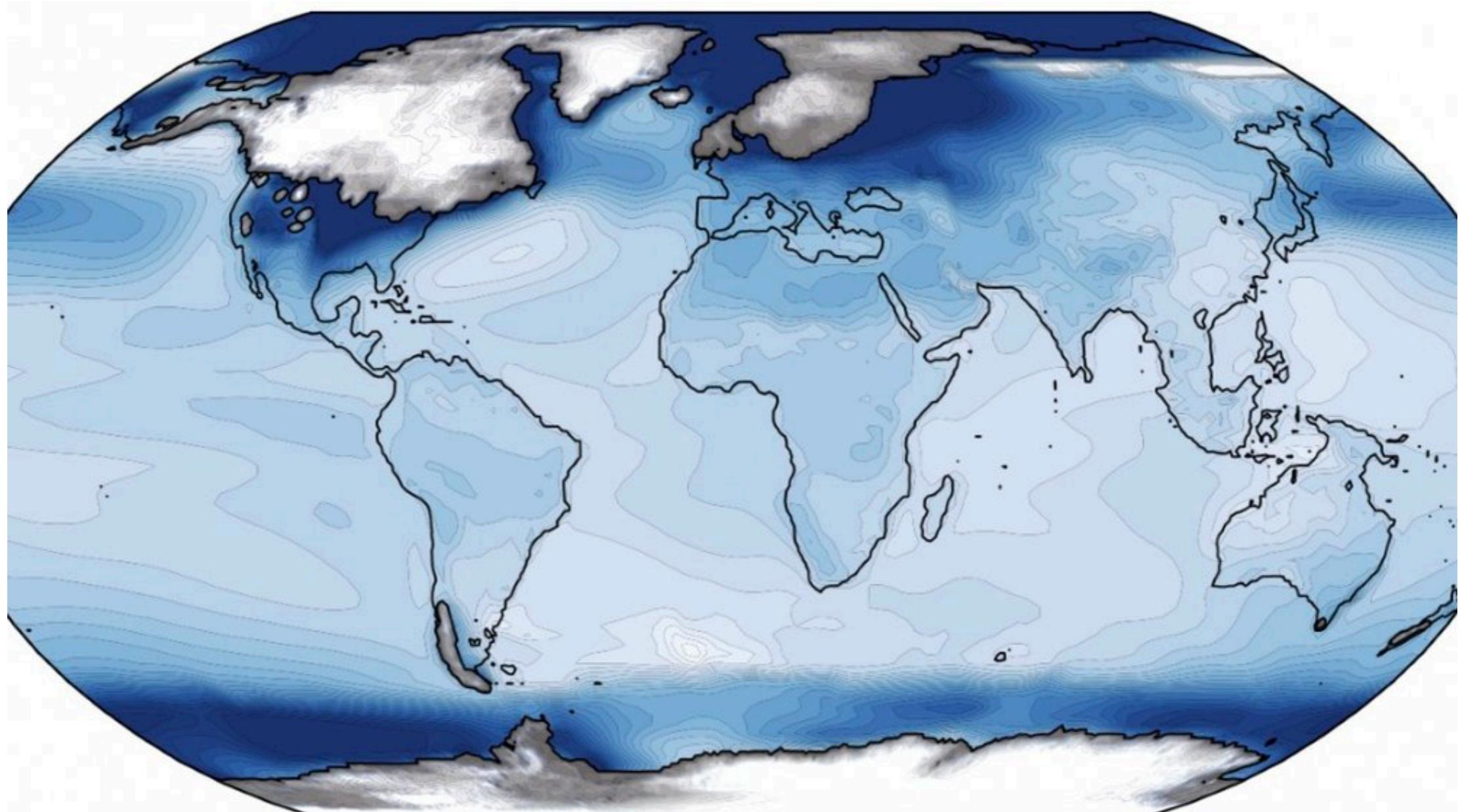


The Ice Age



- Ice age had a profound impact on the earth
- Crazy to think humans were alive at this time

The Ice Age



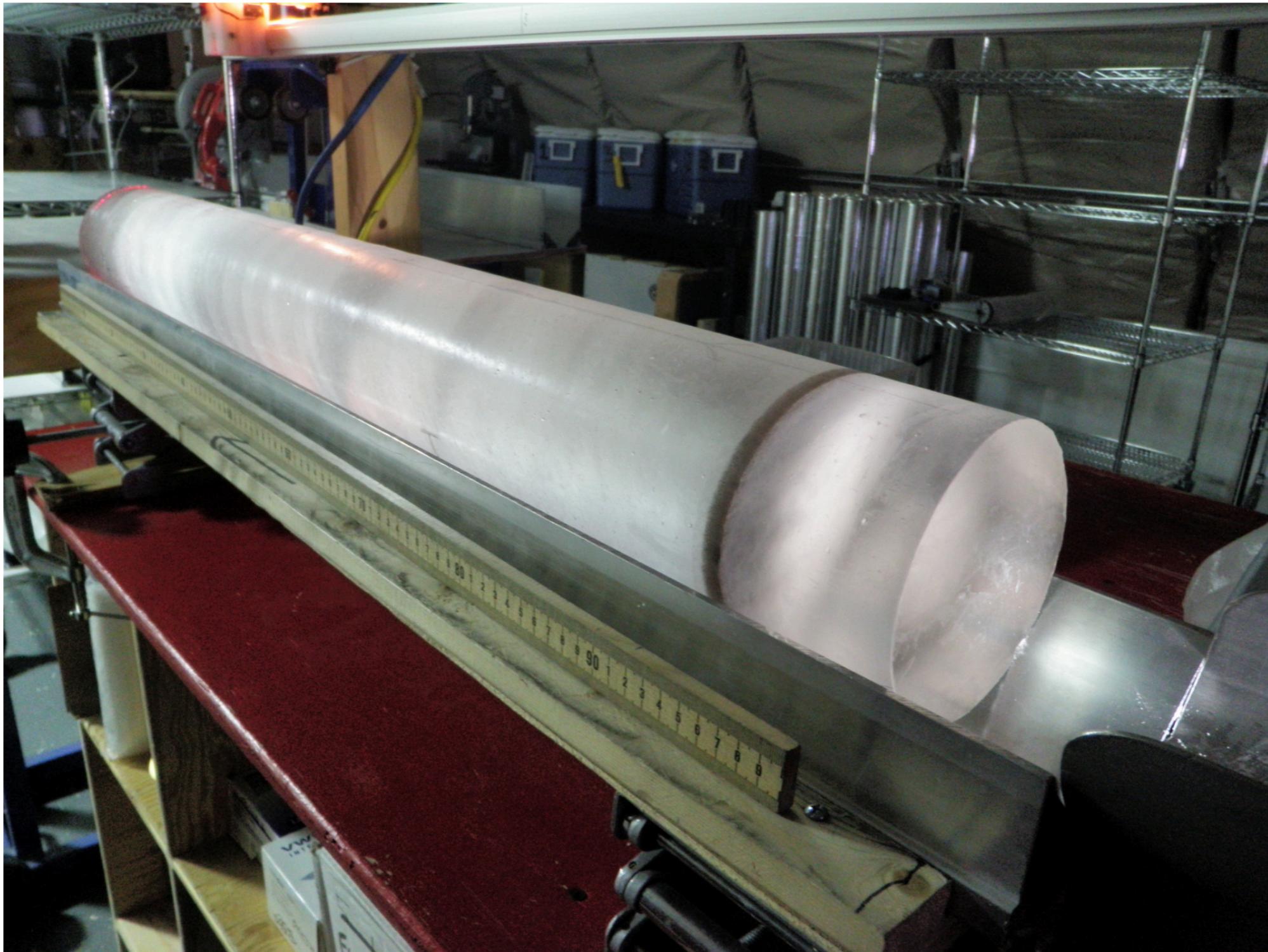
Last Glacial Maximum Surface Air Temperature

Difference from Preindustrial ($^{\circ}\text{C}$)



-14 -12 -10 -8 -6 -4 -2 0

Ice Core Temps



The dark band in this ice core from the West Antarctic Ice Sheet Divide (WAIS Divide) is a layer of volcanic ash that settled on the ice sheet approximately 21,000 years ago. — Credit: Heidi Roop, NSF

Ice Core

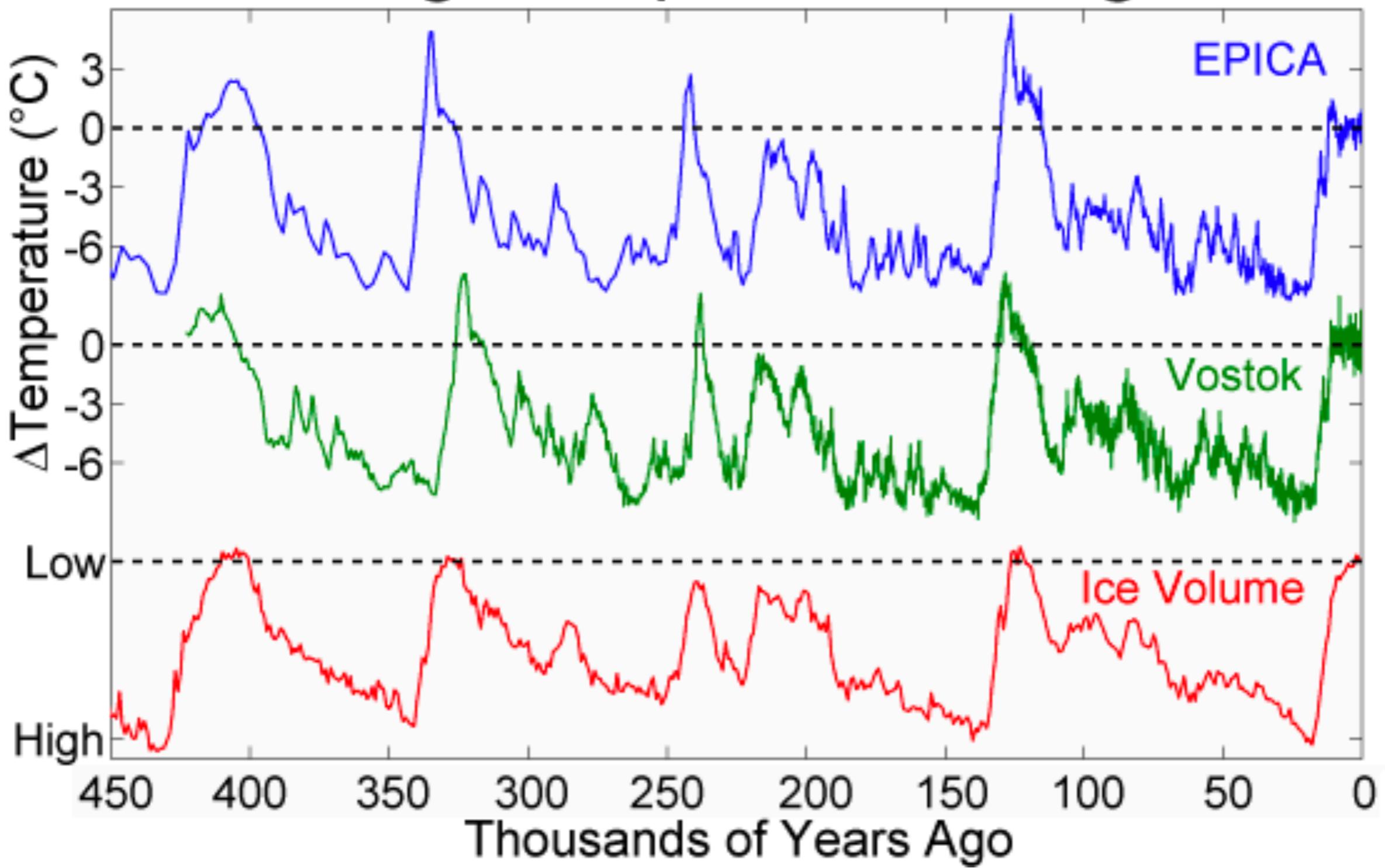


Ice Core recovery has been
A critical element In many
locations on earth

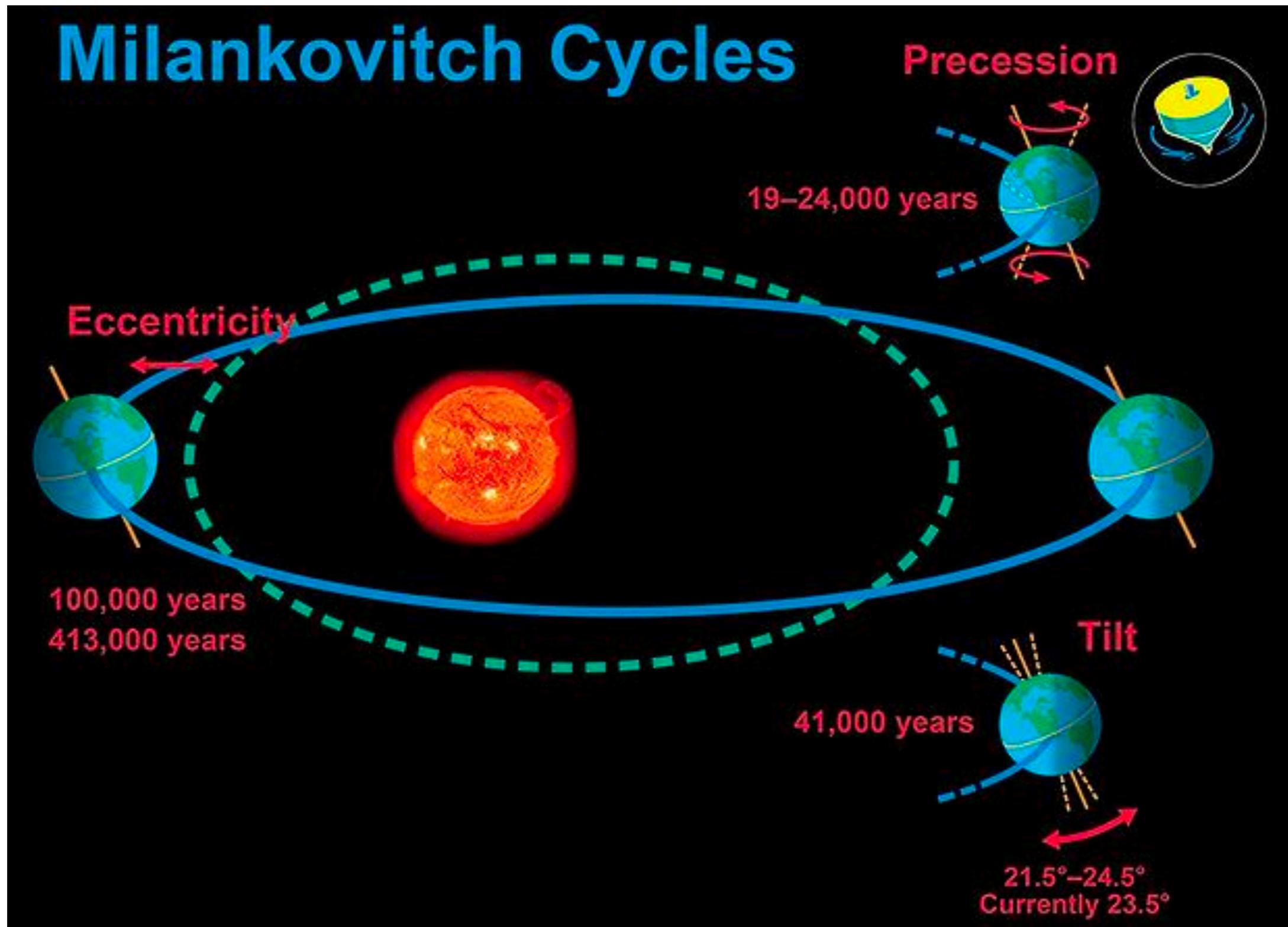


Ice Age over time

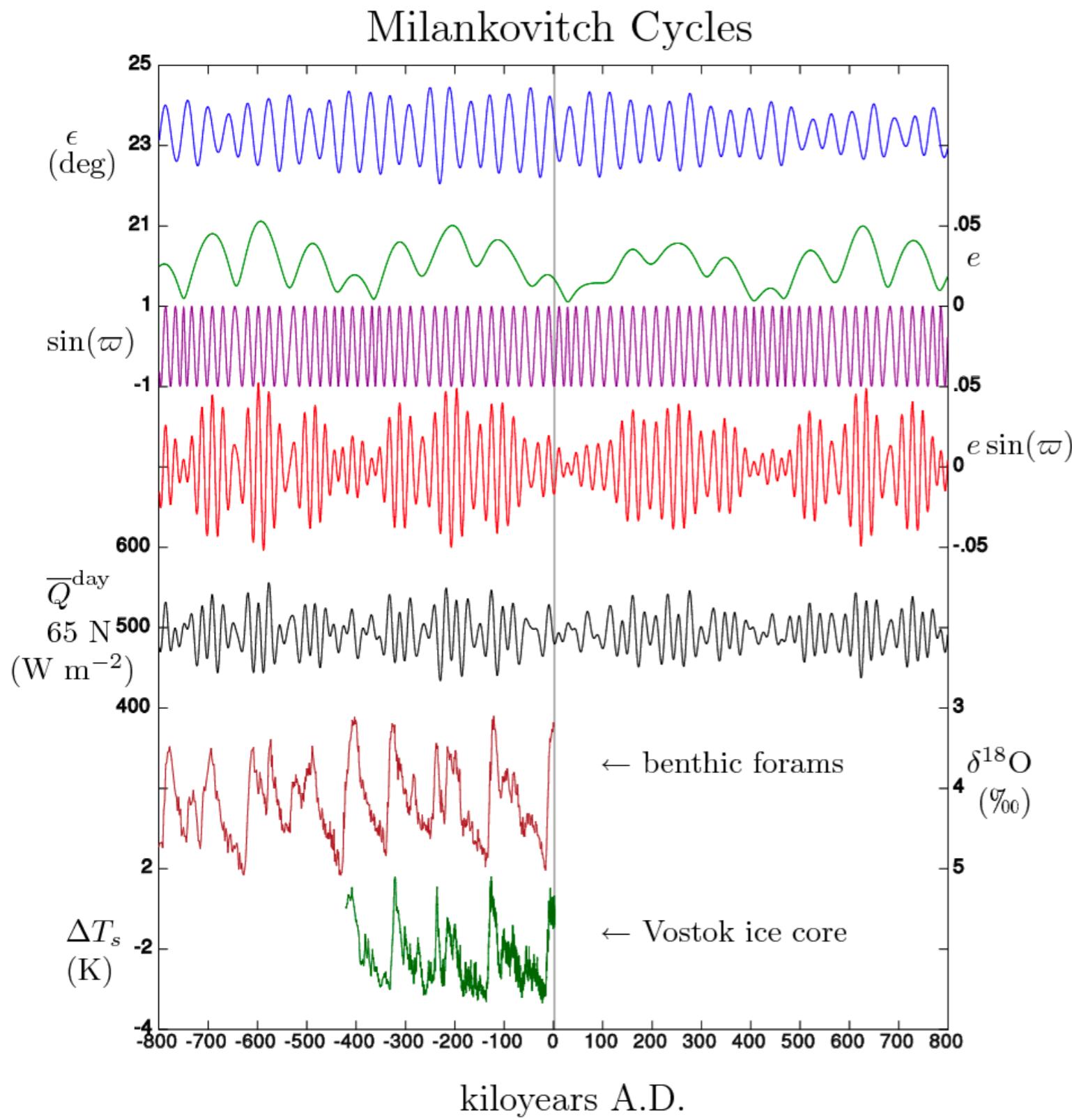
Ice Age Temperature Changes



Milhanovitch Cycles

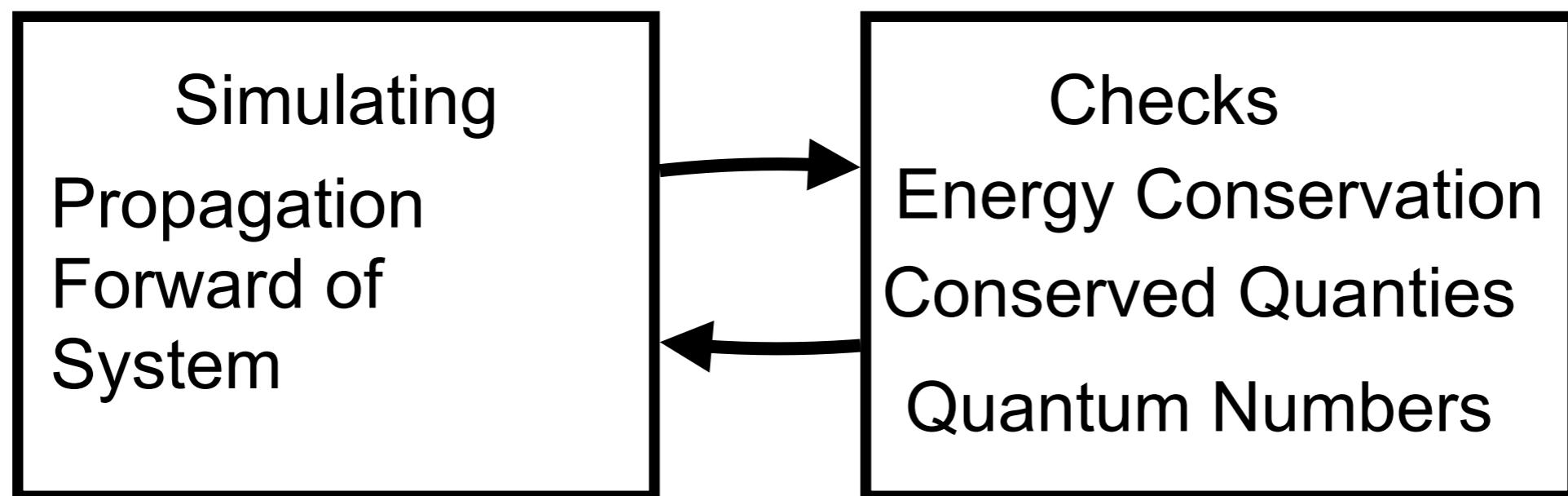


Scale of Eccentricity



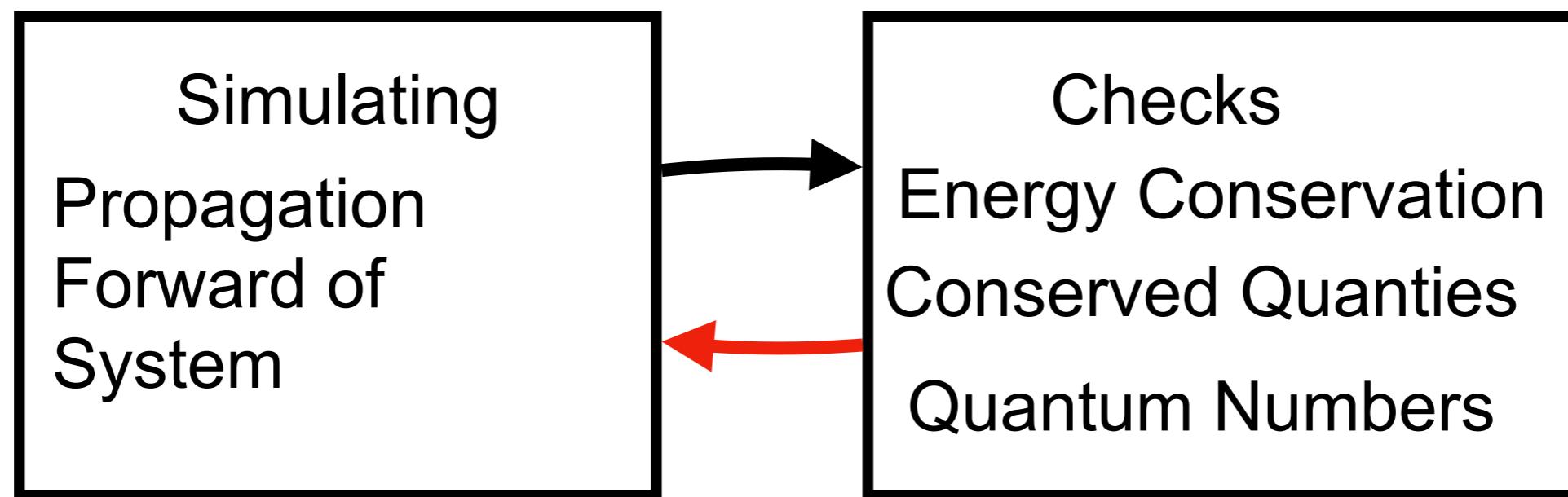
Making MC Better

- In this simulation part of this class
 - We have learned that simulations are not accurate
 - There are a few things we can do to make it better
 - Key is to have a notion of when we are going wrong



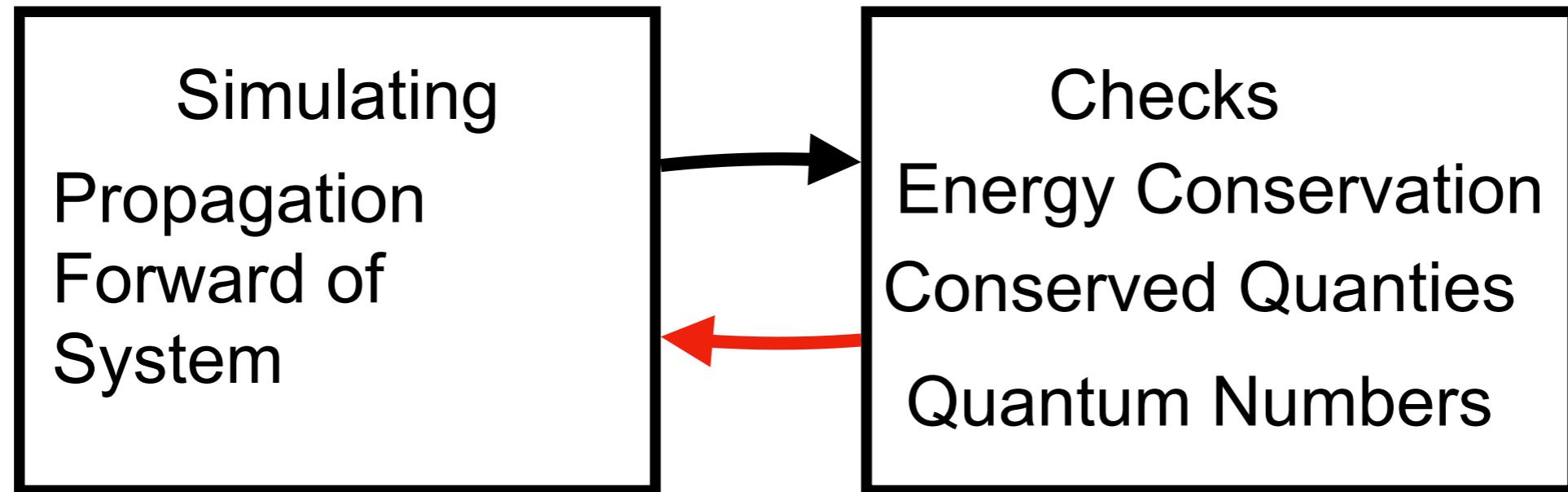
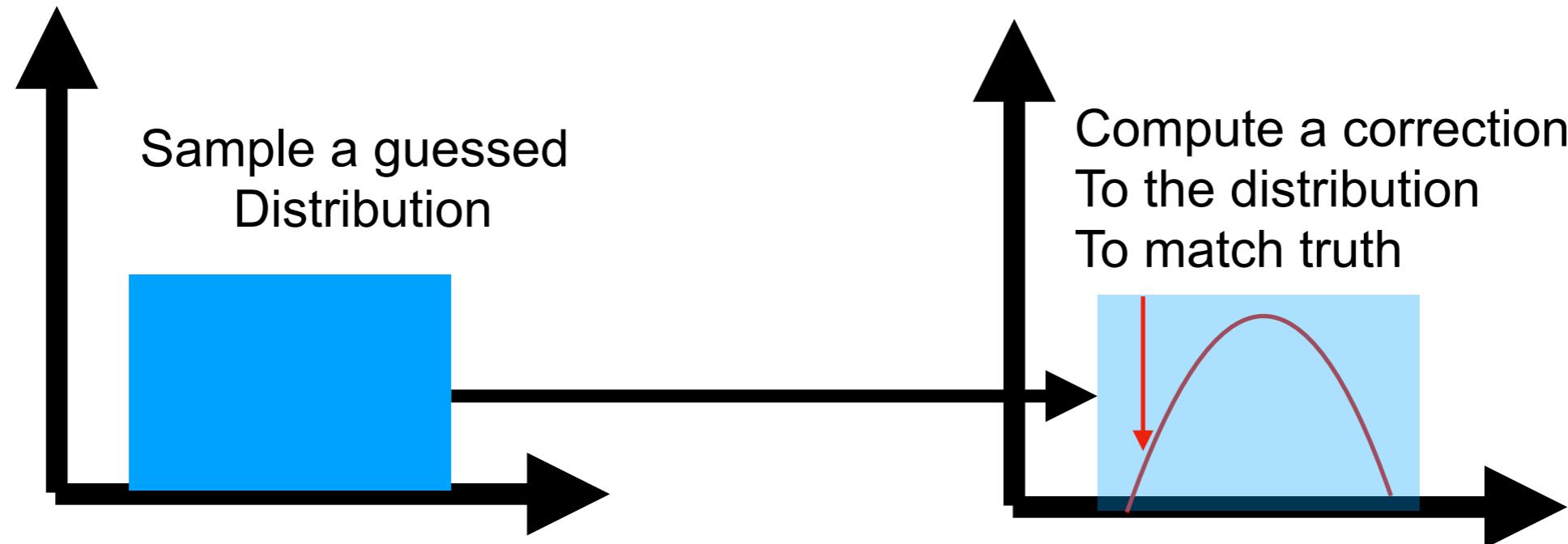
Correcting

- In this simulation part of this class
 - We have learned that simulations are not accurate
 - There are a few things we can do to make it better
 - Key is to have a notion of when we are going wrong



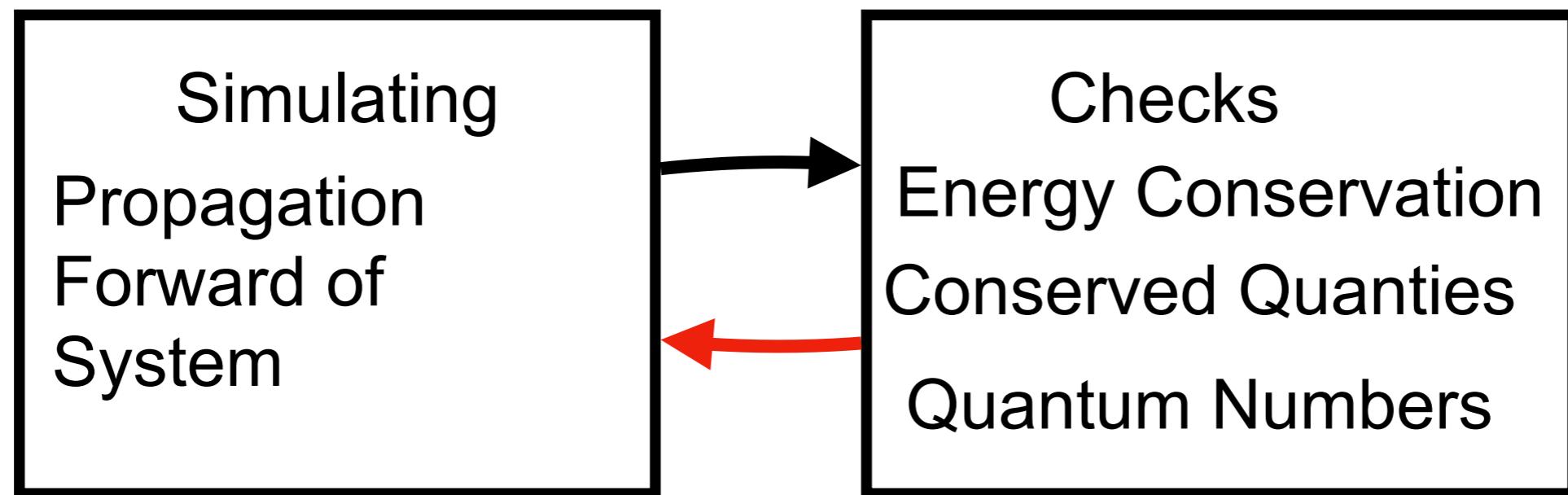
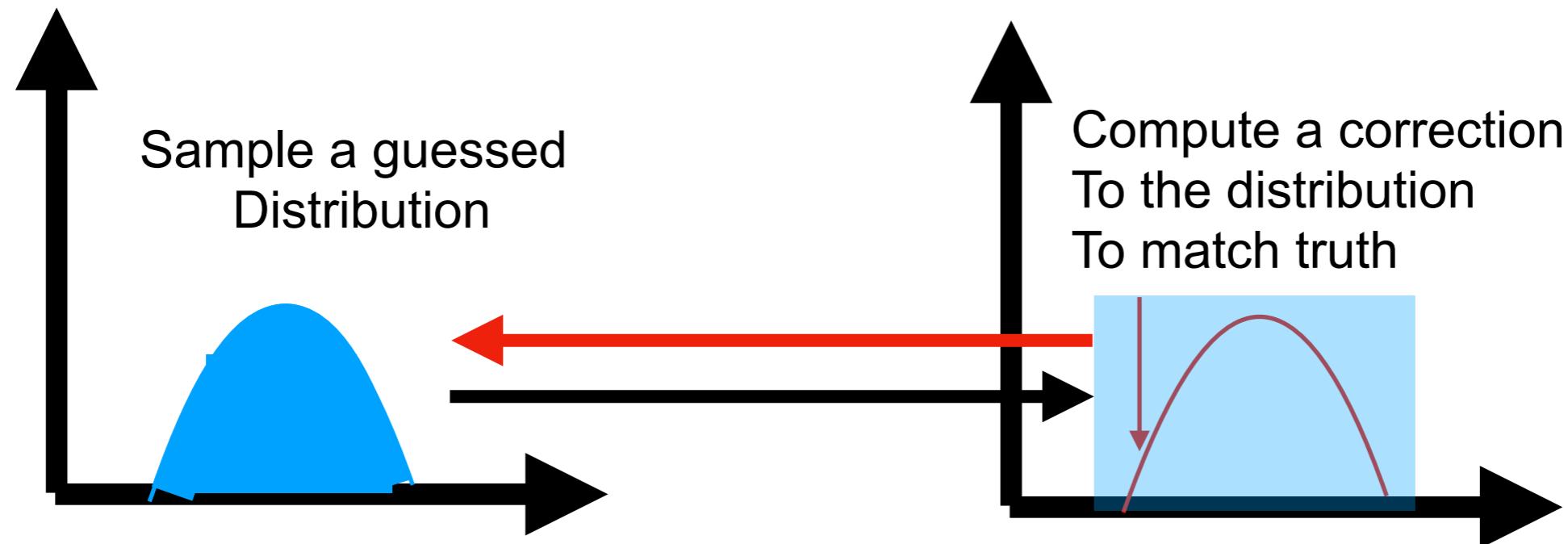
Correct Our Simulation Through a Probabilistic Rescaling

Markov Chain MC



Correct Our Simulation Through a Probabilistic Rescaling

Markov Chain MC



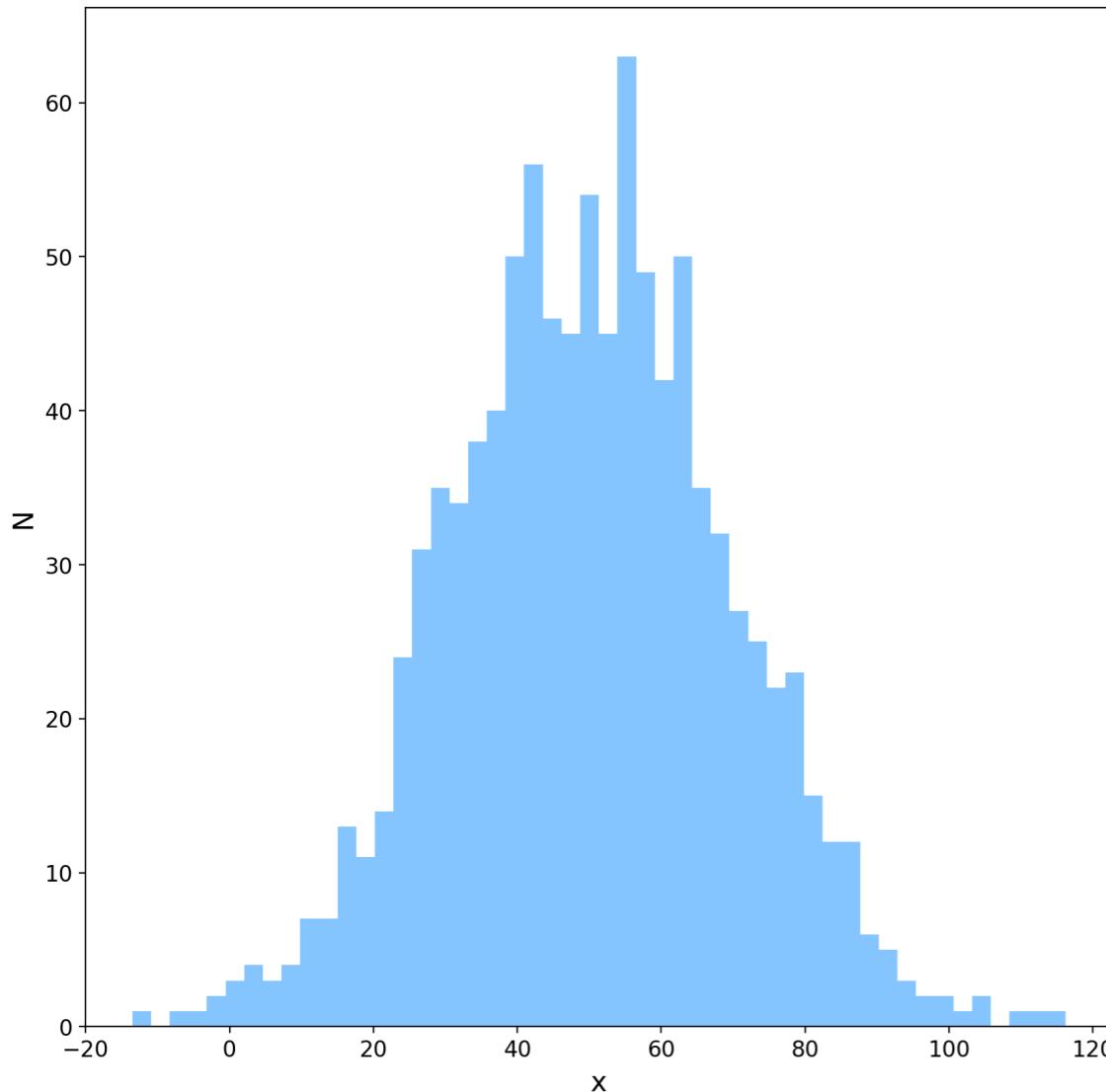
Correct Our Simulation Through a Probabilistic Rescaling

Metropolis-Hastings

- Step 0: Randomly sample a parameter \mathbf{x}_1
- Step 1: Sample a new parameter \mathbf{x}_2
 - Use a chosen “Proposal Function” for sampling
 - Compute the probability of stepping \mathbf{x}_2 to stepping \mathbf{x}_1
- Step 2: Sample a flat distribution from 0 to 1 (s_2)
 - Accept \mathbf{x}_2 if $s_2 < \frac{p(x_2)}{p(x_1)}$
- Step 3 : Go back to step 1

Fitting a Gaussian

- Strategy to randomly sample mean(μ) and sigma σ
 - Accept the values for μ, σ if probability is higher
 - Keep accepting/rejecting until we hit equilibrium

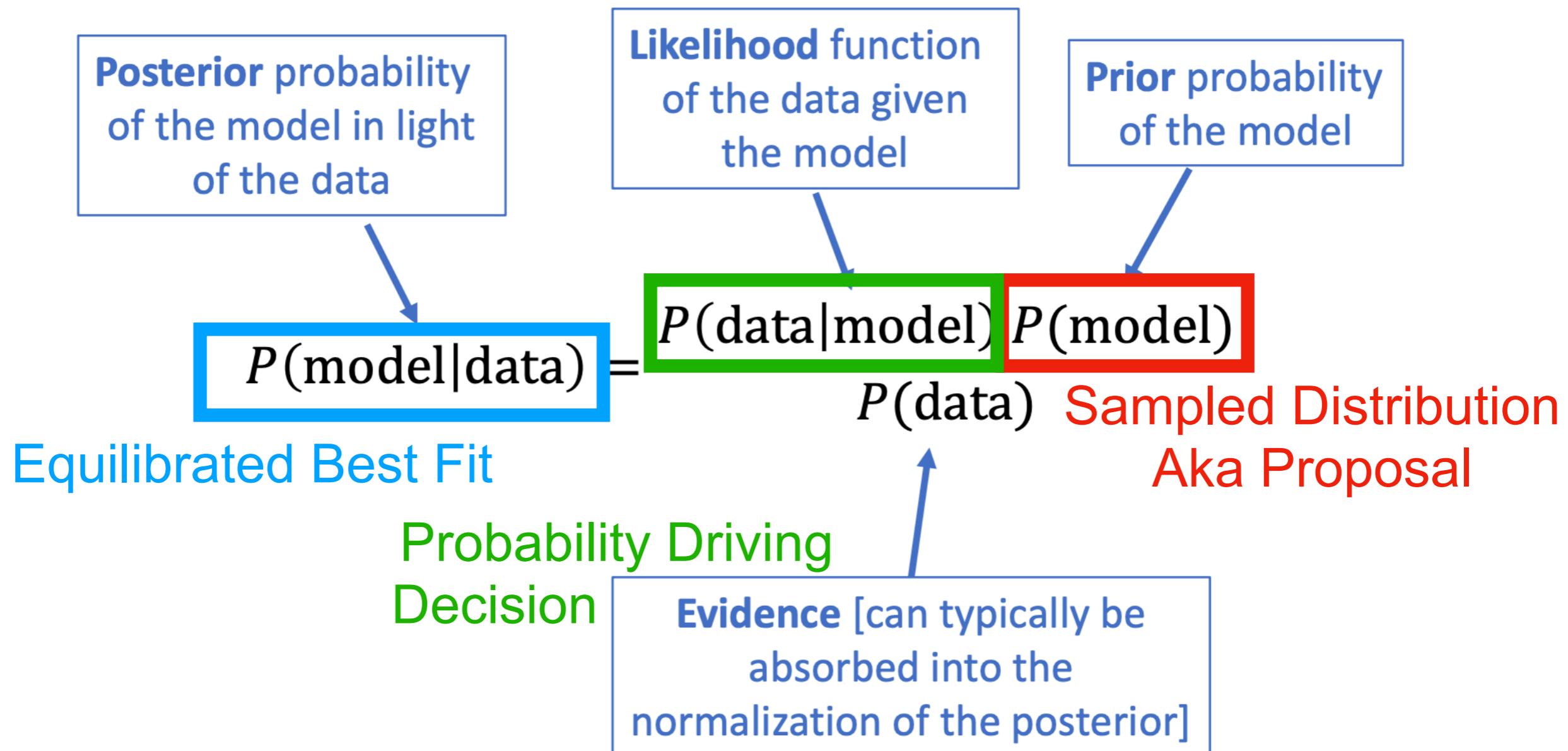


Log(Probability)

$$\begin{aligned} \log(\mathcal{L}) &= \sum_i \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\mu-x_i)^2}{\sigma^2}}\right) \\ &= \sum_i \left(\frac{x_i - \mu}{\sigma}\right)^2 - \frac{1}{2} \log(2\pi\sigma^2) \end{aligned}$$

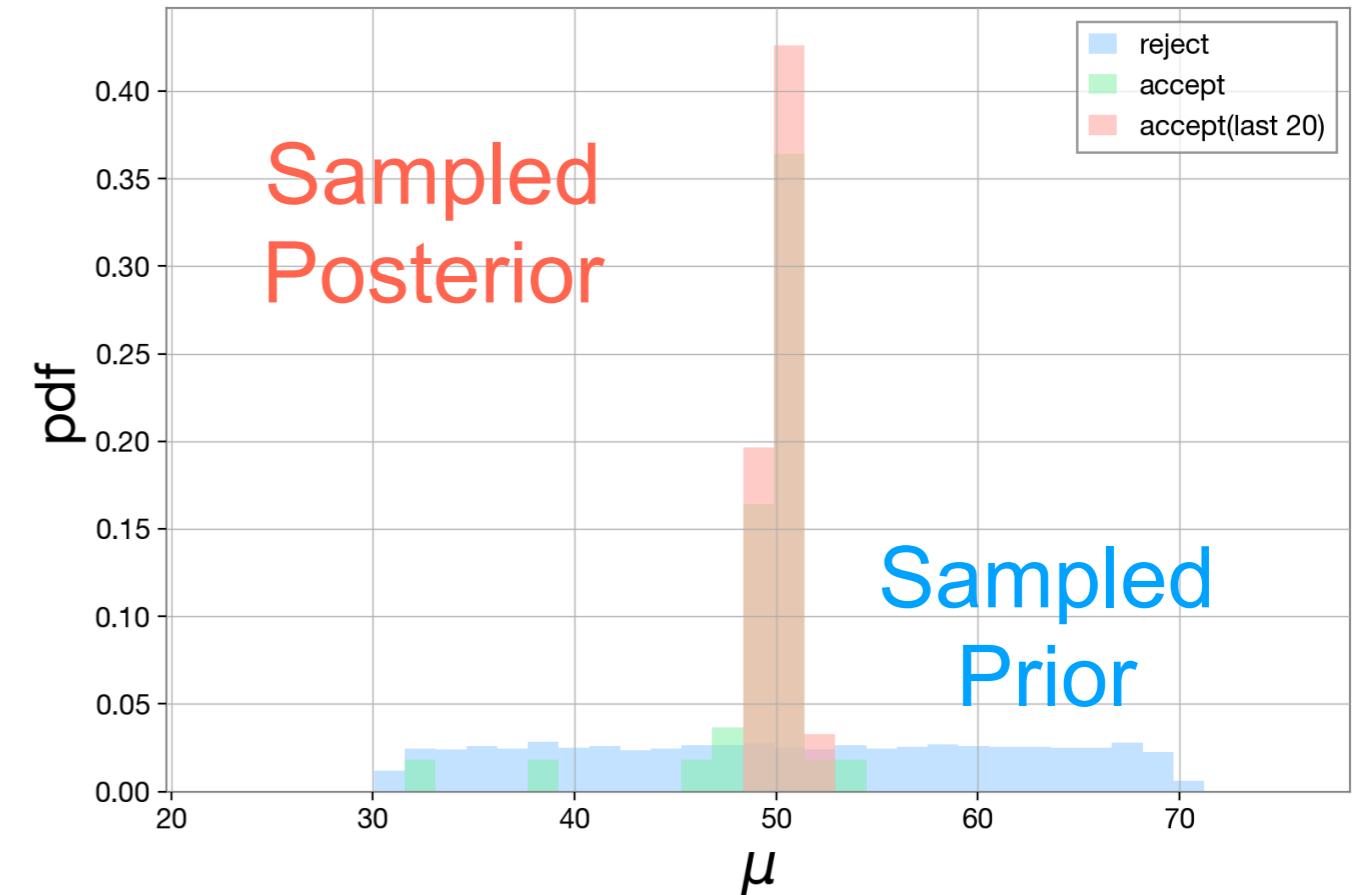
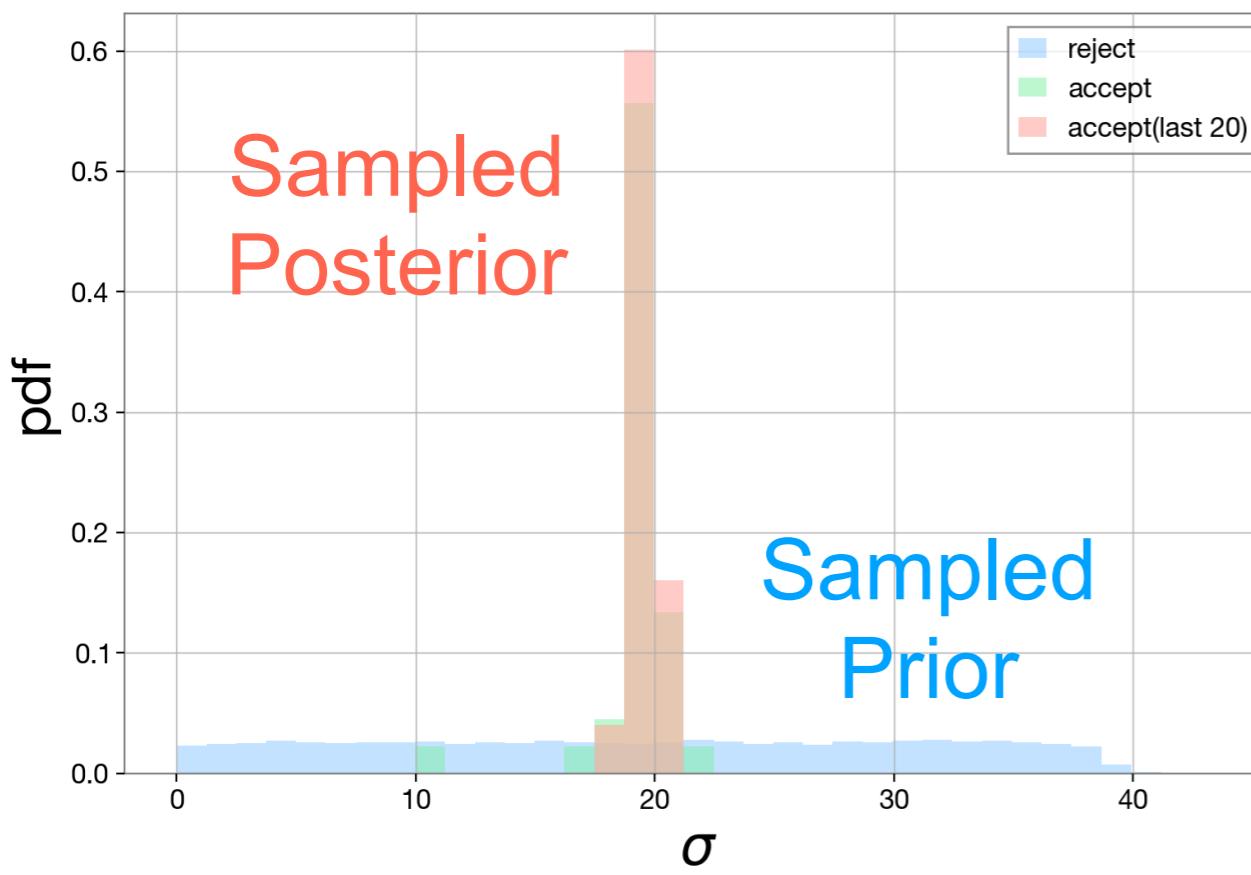
Accepted μ, σ yield the best fits

Visualizing in Bayes



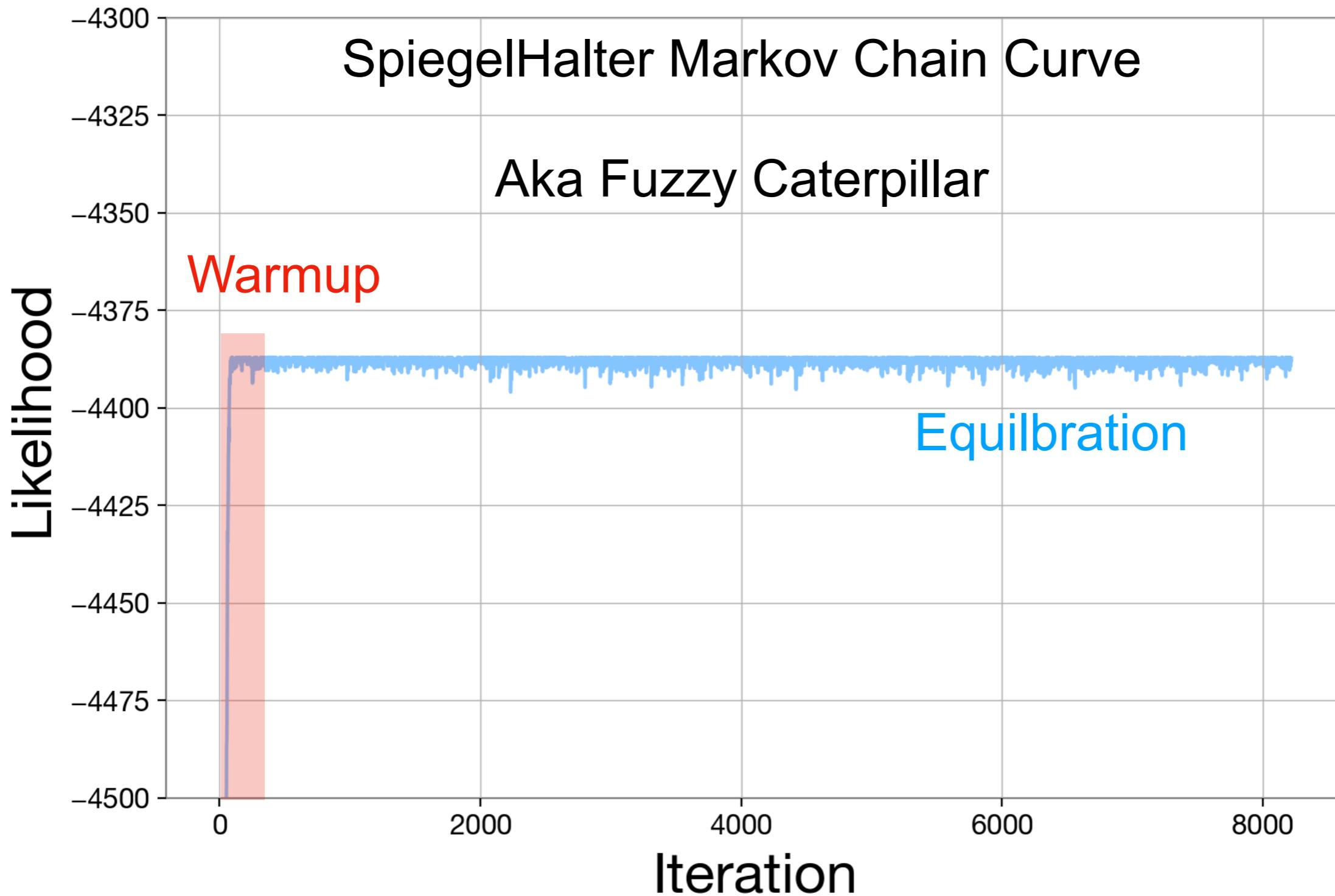
The Likelihood reweights the Prior to the Posterior

Best Fit Parameters



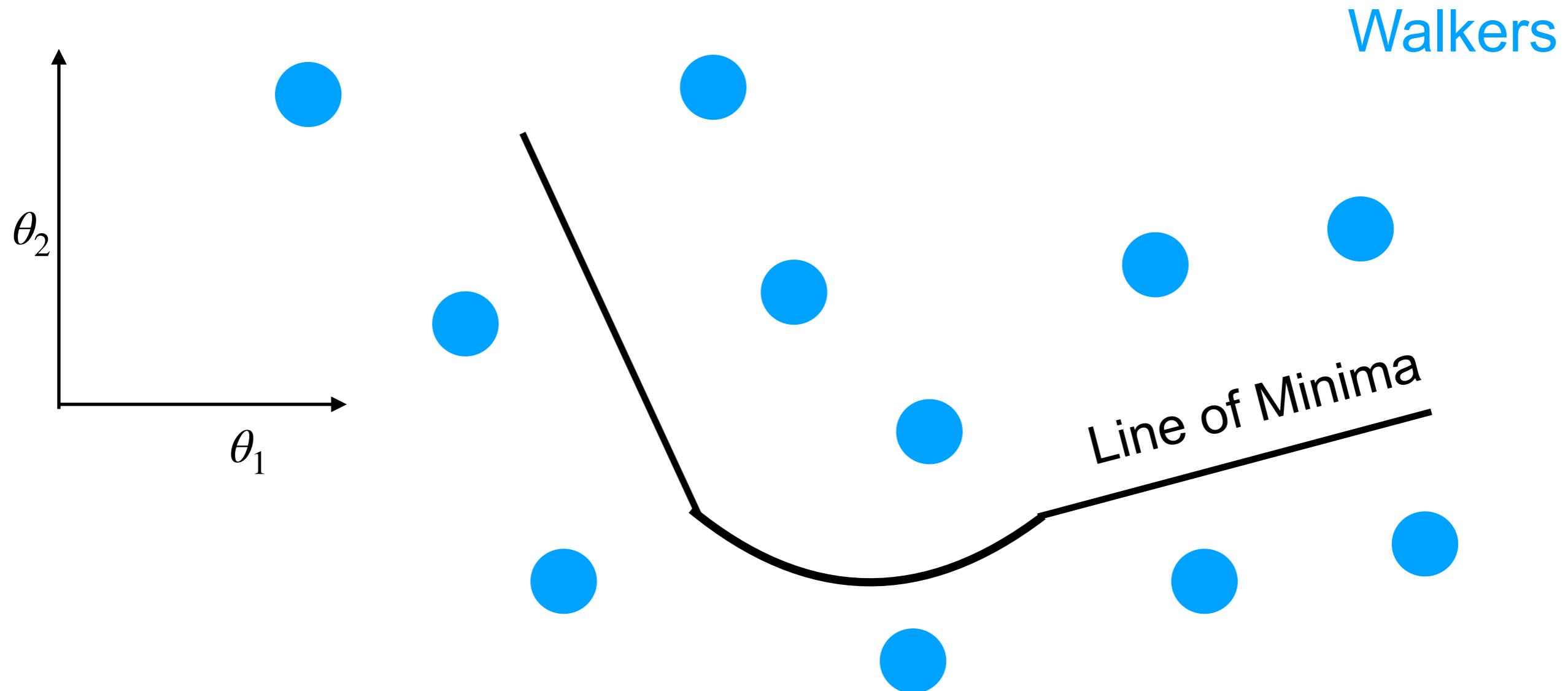
- This is where I start to hint that there are limitations here
- What we are doing is running a check
 - We are not taking a derivative (No gradients or Hessians)

Best Fit Parameters



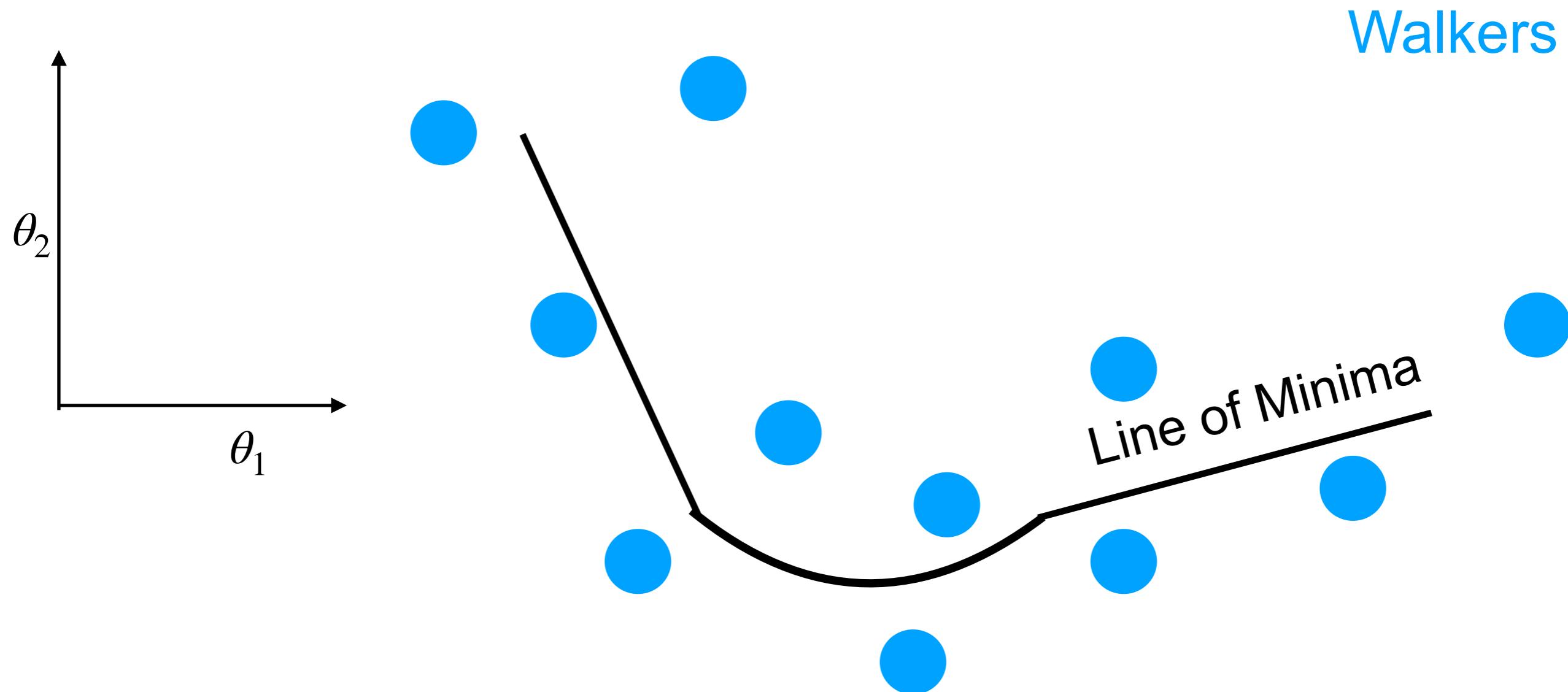
Speeding up MCMC

- We can consider having many walkers probe our space
 - Many walkers at the same time speed up convergence



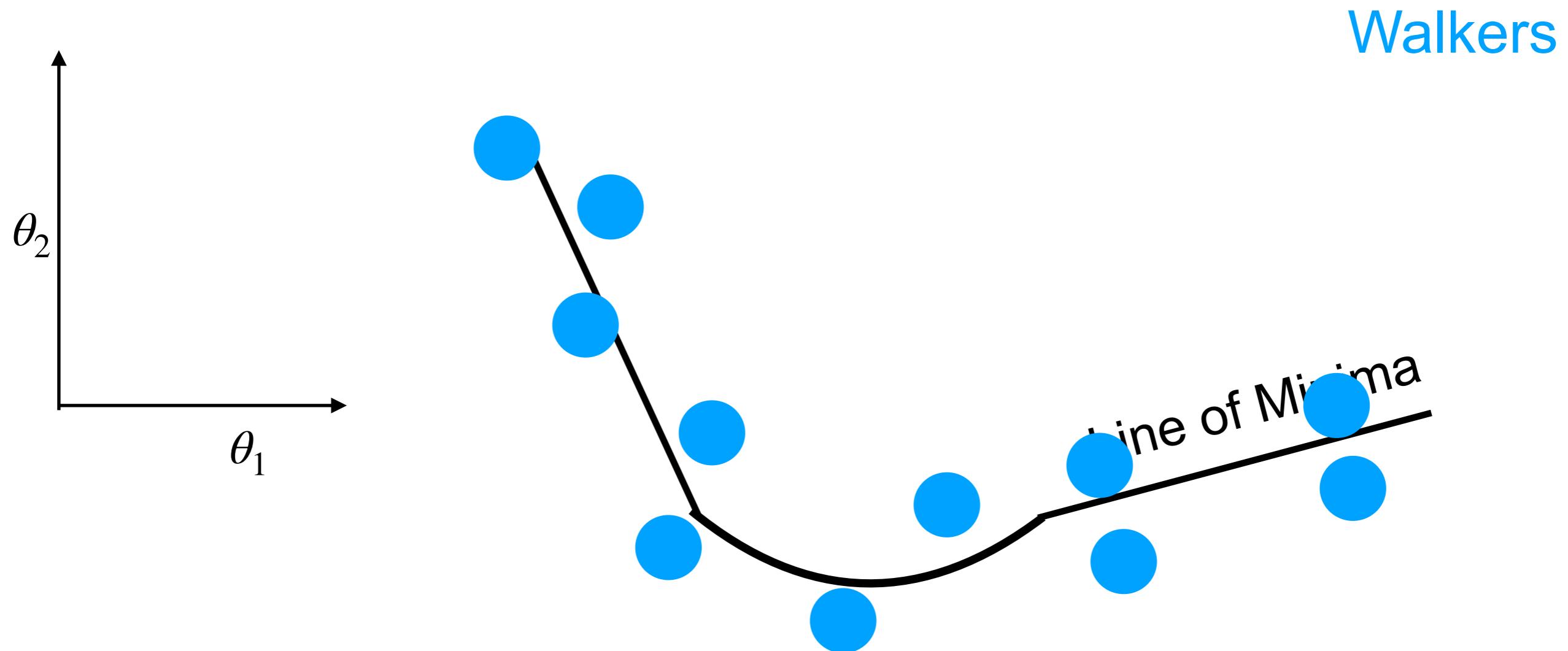
Speeding up MCMC

- We can consider having many walkers probe our space
 - Many walkers at the same time speed up convergence

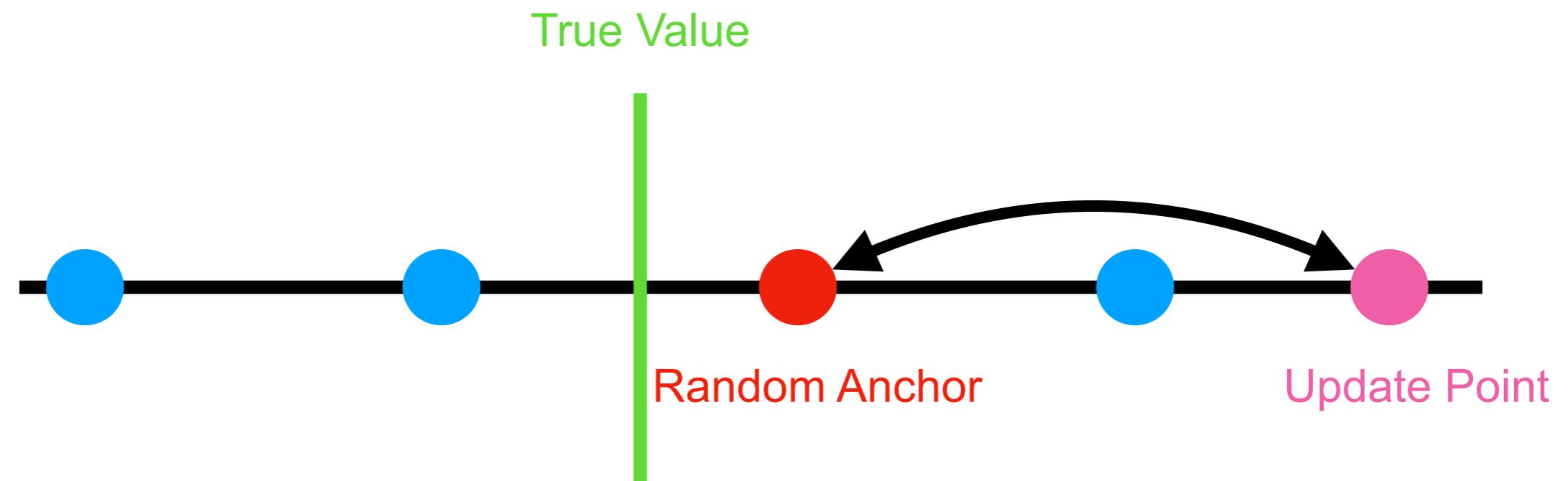


Speeding up MCMC

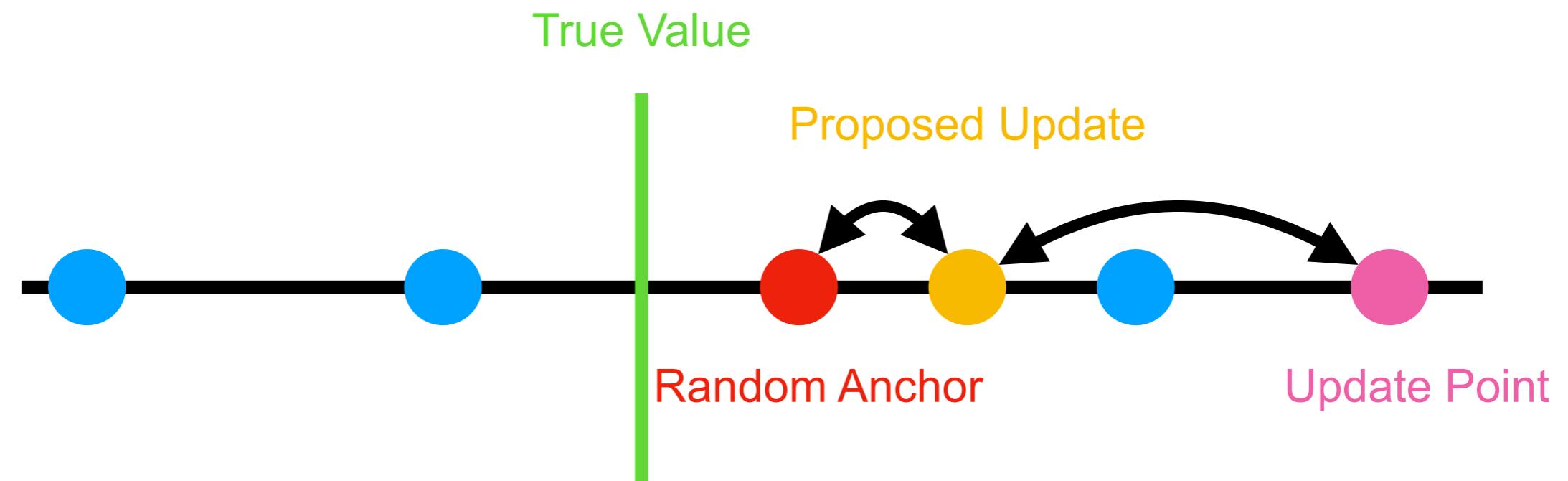
- We can consider having many walkers probe our space
 - Many walkers at the same time speed up convergence



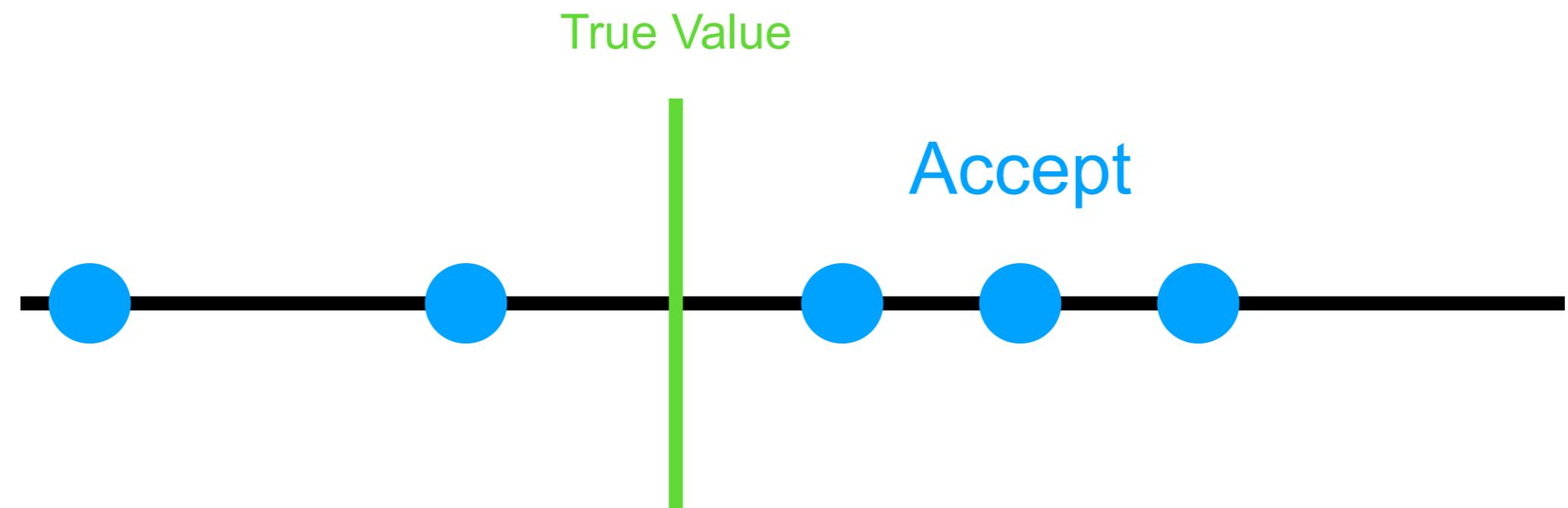
Example Fit



Example Fit

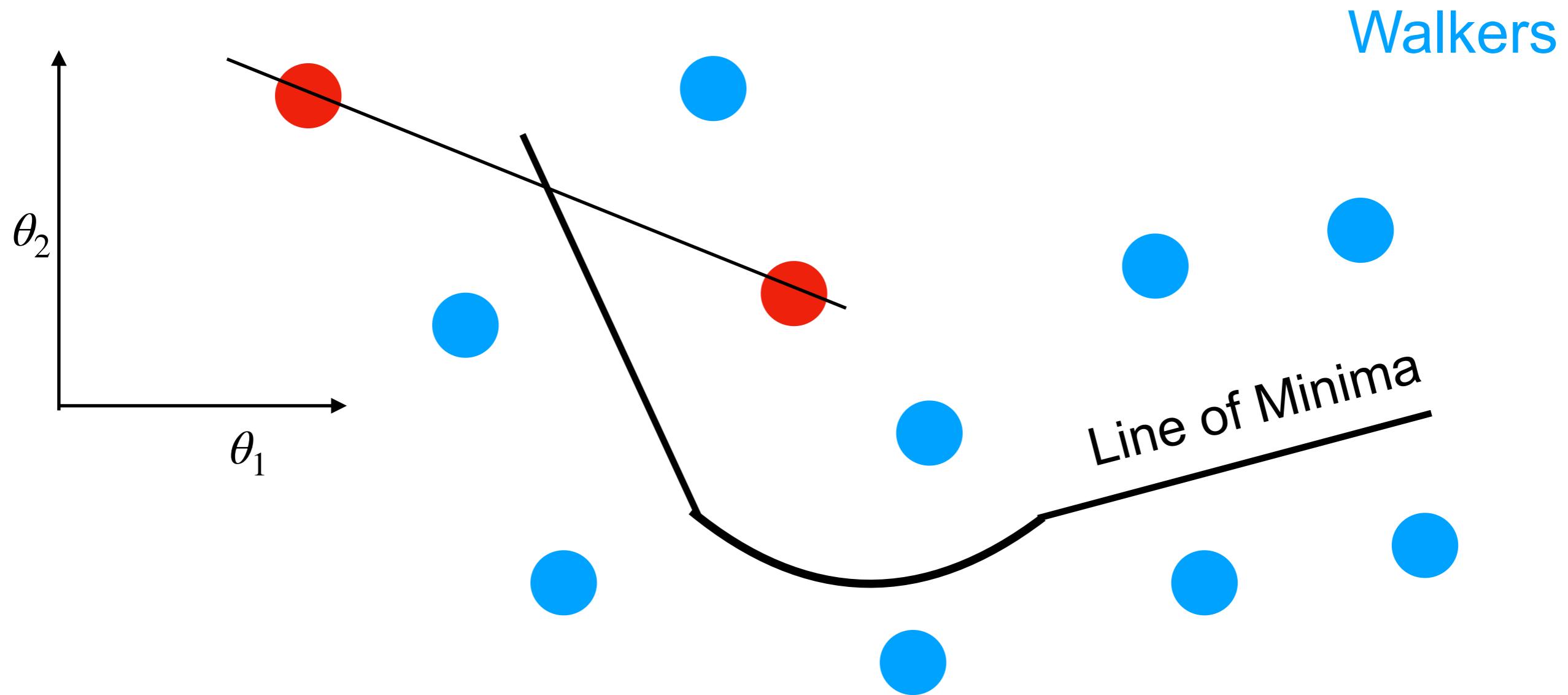


Example Fit



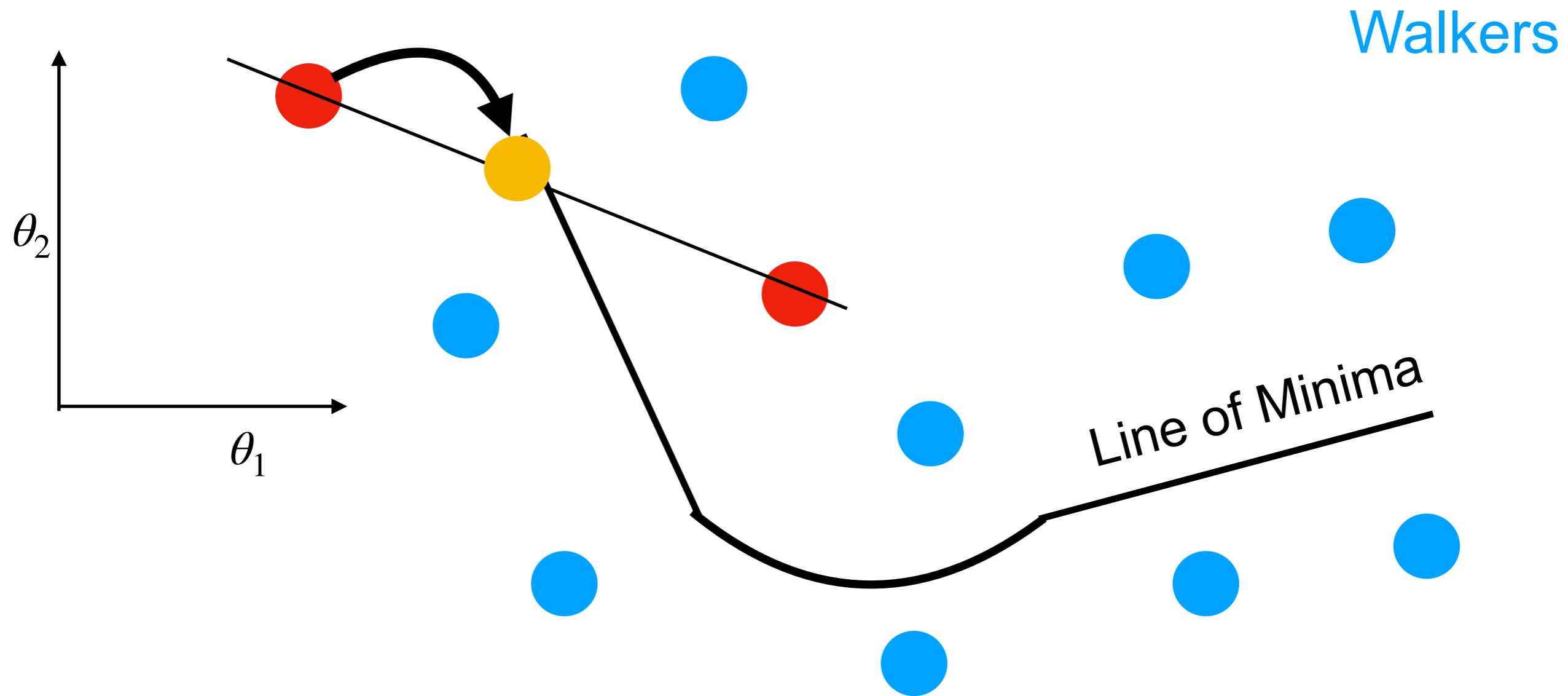
Updating w/Random Points

- We randomly choose a pair of points
 - Move one of the points along the line between them



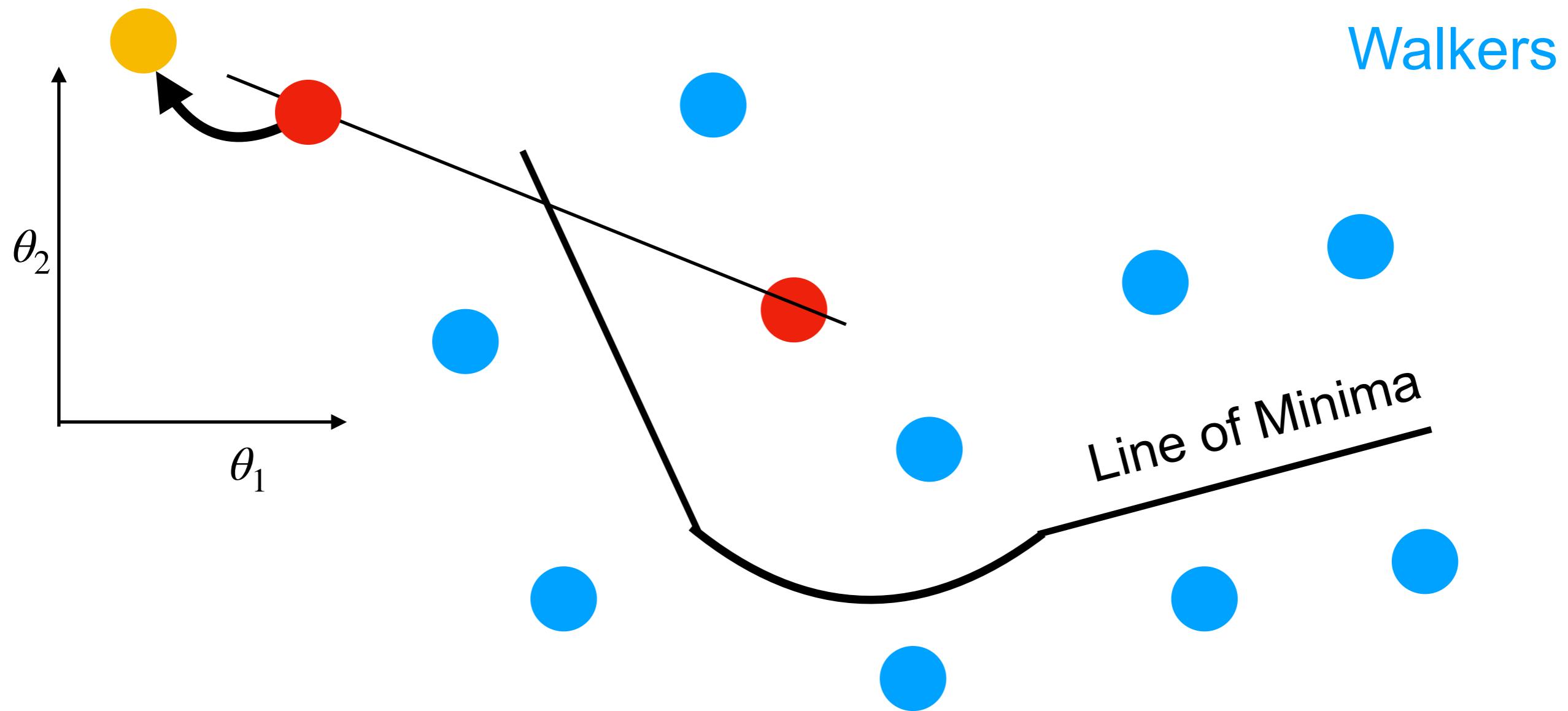
Updating w/Random Points

- We randomly choose a pair of points
 - Move one of the points along the line between them



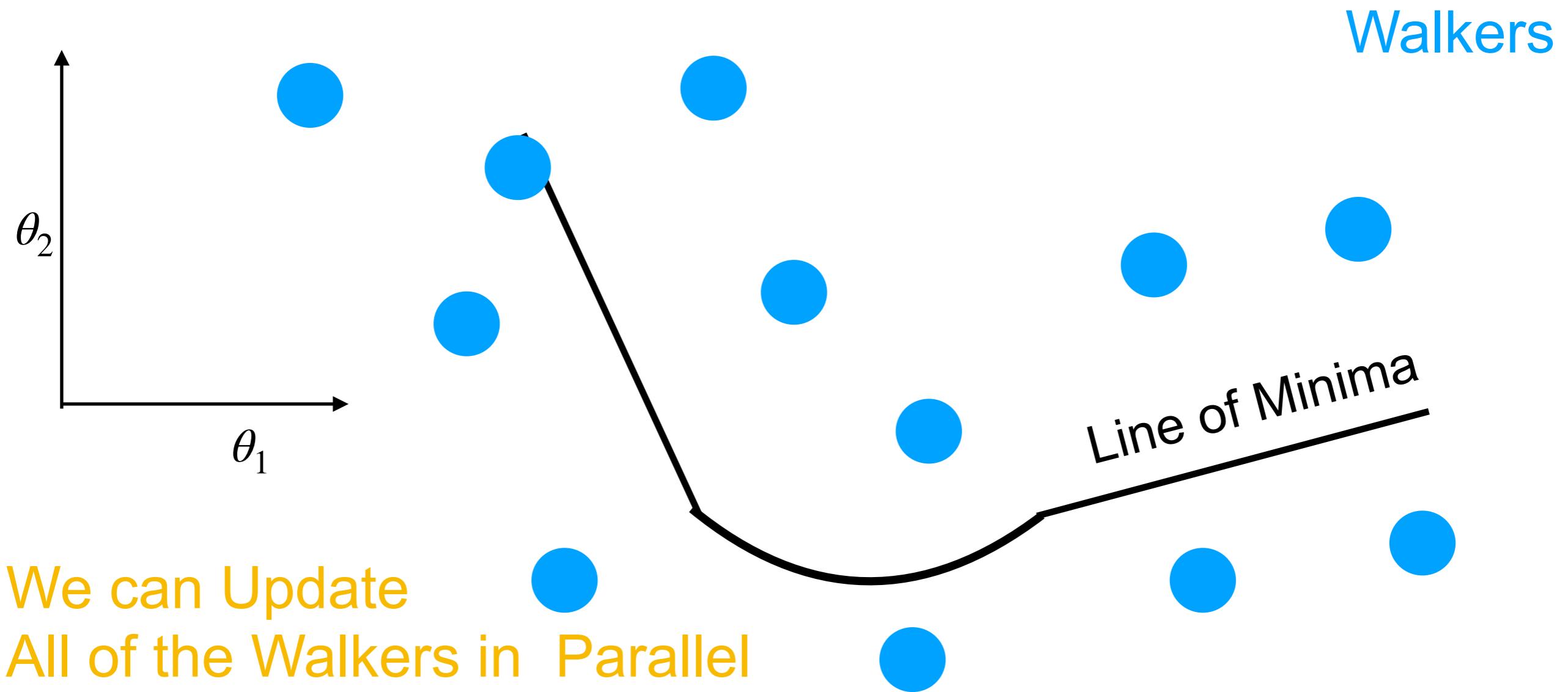
Updating w/Random Points

- We randomly choose a pair of points
 - Move one of the points along the line between them



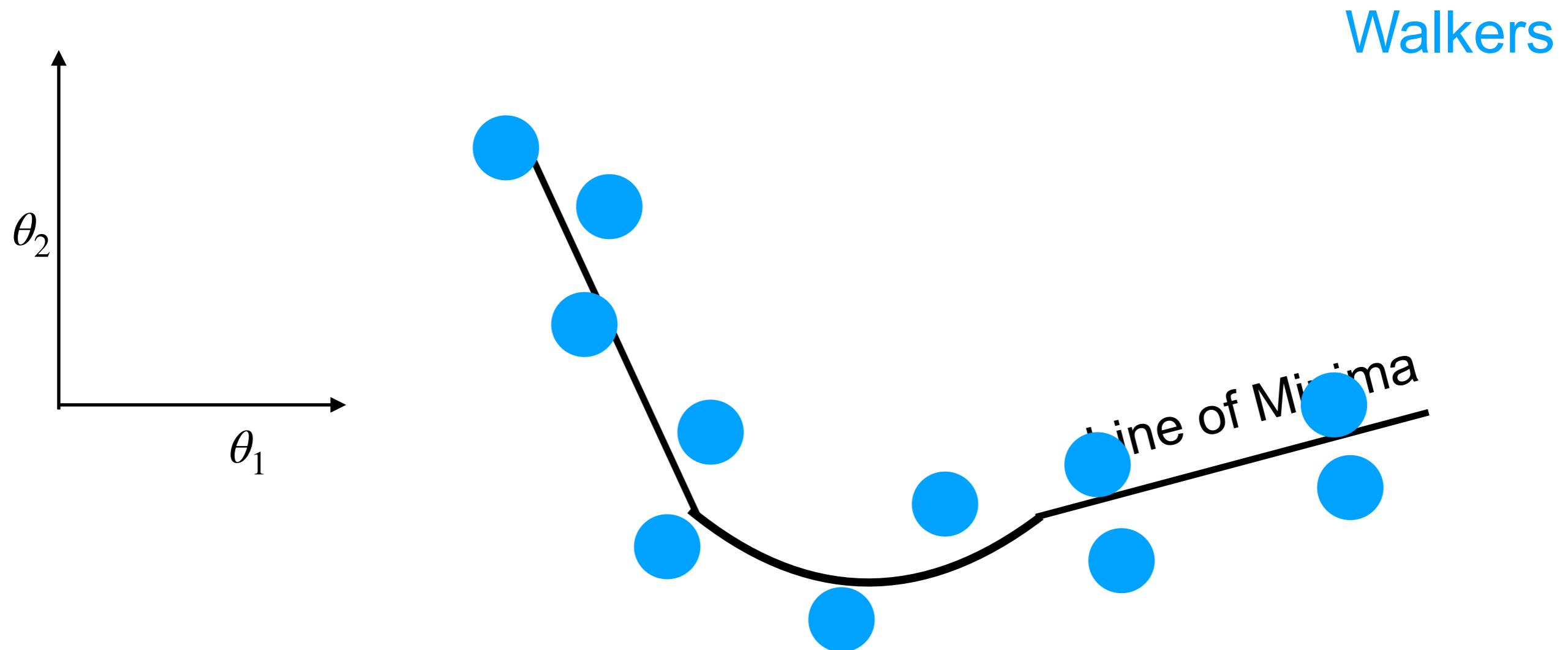
Updating w/Random Points

- We randomly choose a pair of points
 - Move one of the points along the line between them

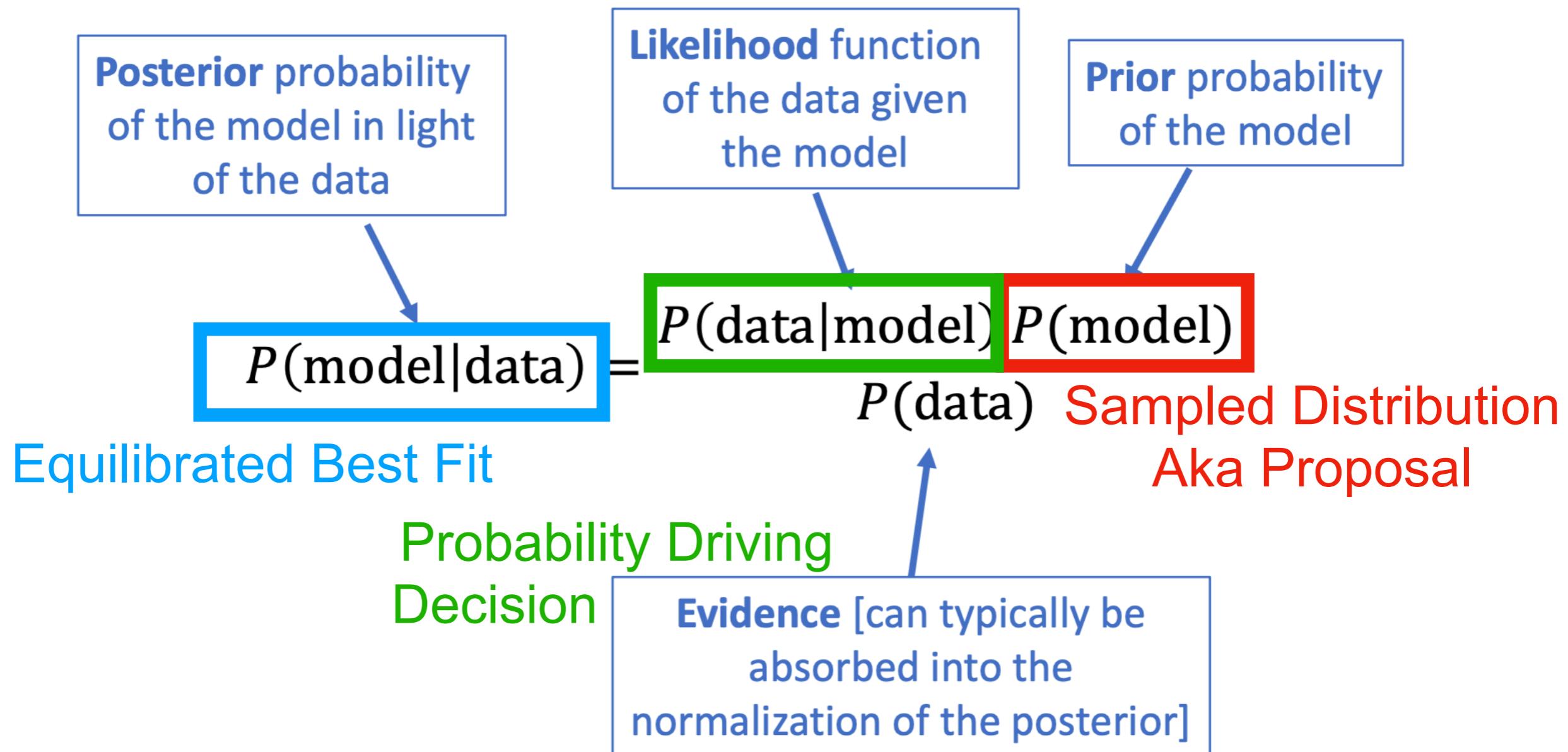


Speeding up MCMC

- We randomly choose a pair of points
 - Move one of the points along the line between them



Visualizing in Bayes



The Likelihood reweights the Prior to the Posterior

Quantum Monte Carlo

- Can use the same MCMC to populate a wave function
 - We can then scan parameters to solve Schrödinger's Eq

$$\psi(\vec{r} | \vec{\theta}) = Ae^{-r/\theta_0}$$

Guess a Form for the wavefunction

$$p(\vec{r} | \vec{\theta}) = \frac{\psi^*(\vec{r} | \vec{\theta})\psi(\vec{r} | \vec{\theta})}{\langle \psi | \psi \rangle}$$

We can define probability from wavefunction

$$w_{i+1} = \frac{p(\vec{r}_{i+1} | \vec{\theta})}{p(\vec{r}_i | \vec{\theta})} = \frac{\psi^*(\vec{r}_{i+1})\psi(\vec{r}_{i+1})}{\psi^*(\vec{r}_i)\psi(\vec{r}_i)}$$

Our proposal
Doesn't need integral
Aka $\langle \psi | \psi \rangle$

Multiple Walkers Populate

- The key is to MCMC evolve the wave function many times
 - We can use the aggregate Particles solve QM stuff

$$\sum_j \psi_j(\vec{r} | \vec{\theta}) = A e^{-r/\theta_0}$$

Guess a Form for
the wavefunction

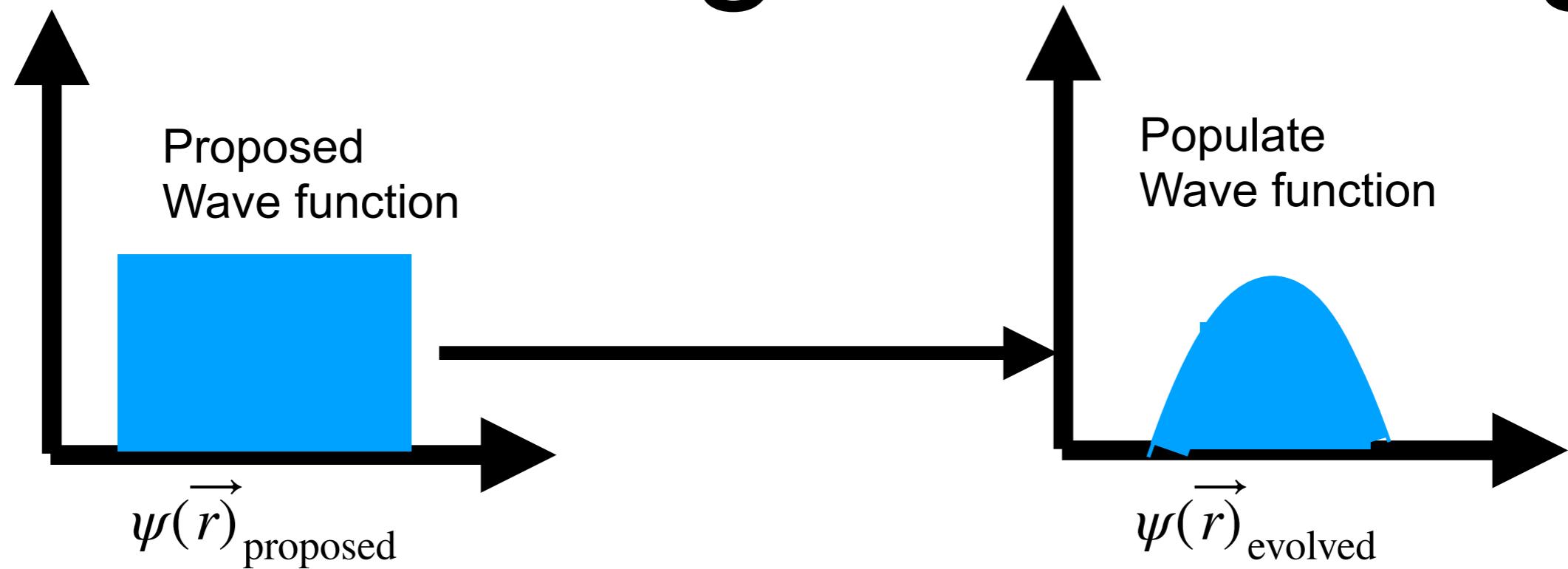
$$\sum_j p_j(\vec{r} | \vec{\theta}) = \frac{\psi_j^*(\vec{r} | \vec{\theta}) \psi_j(\vec{r} | \vec{\theta})}{\langle \psi | \psi \rangle}$$

We can define
probability from
wavefunction

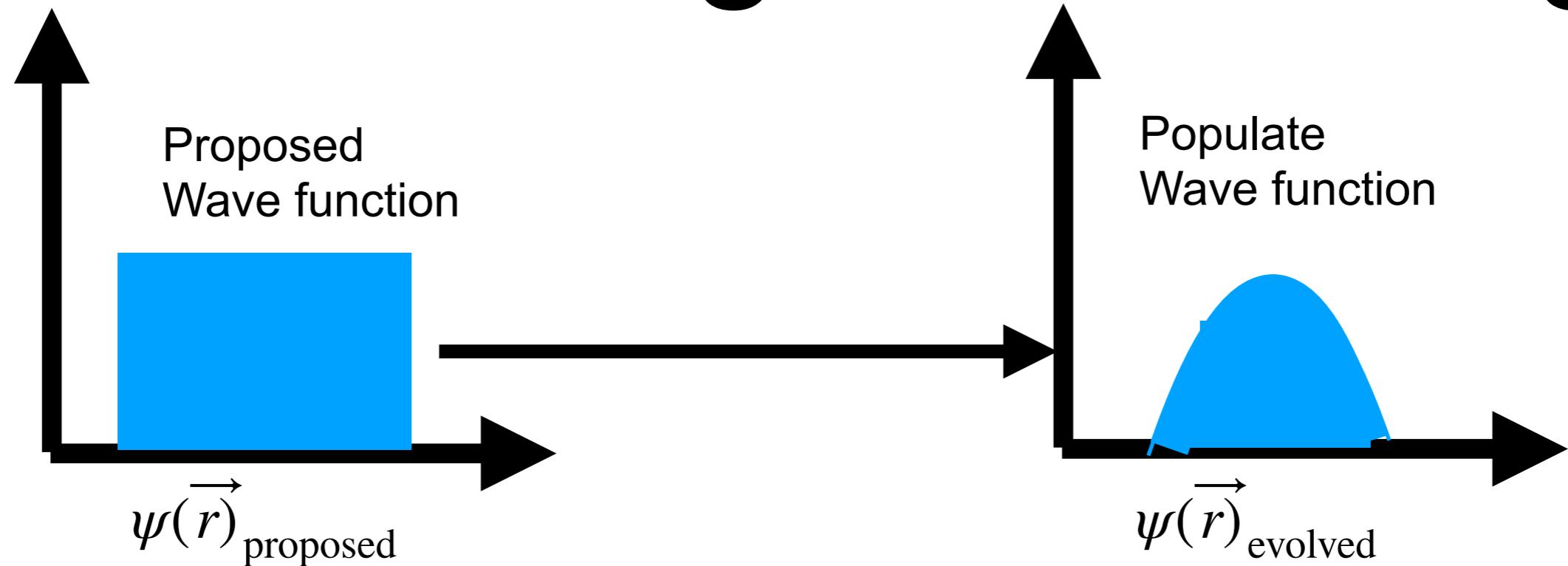
$$\sum_j w_{i+1}^j = \frac{p_j(\vec{r}_{i+1} | \vec{\theta})}{p_j(\vec{r}_i | \vec{\theta})}$$

Our proposal
Doesn't need
 $\langle \psi | \psi \rangle$

Solving Schroedinger



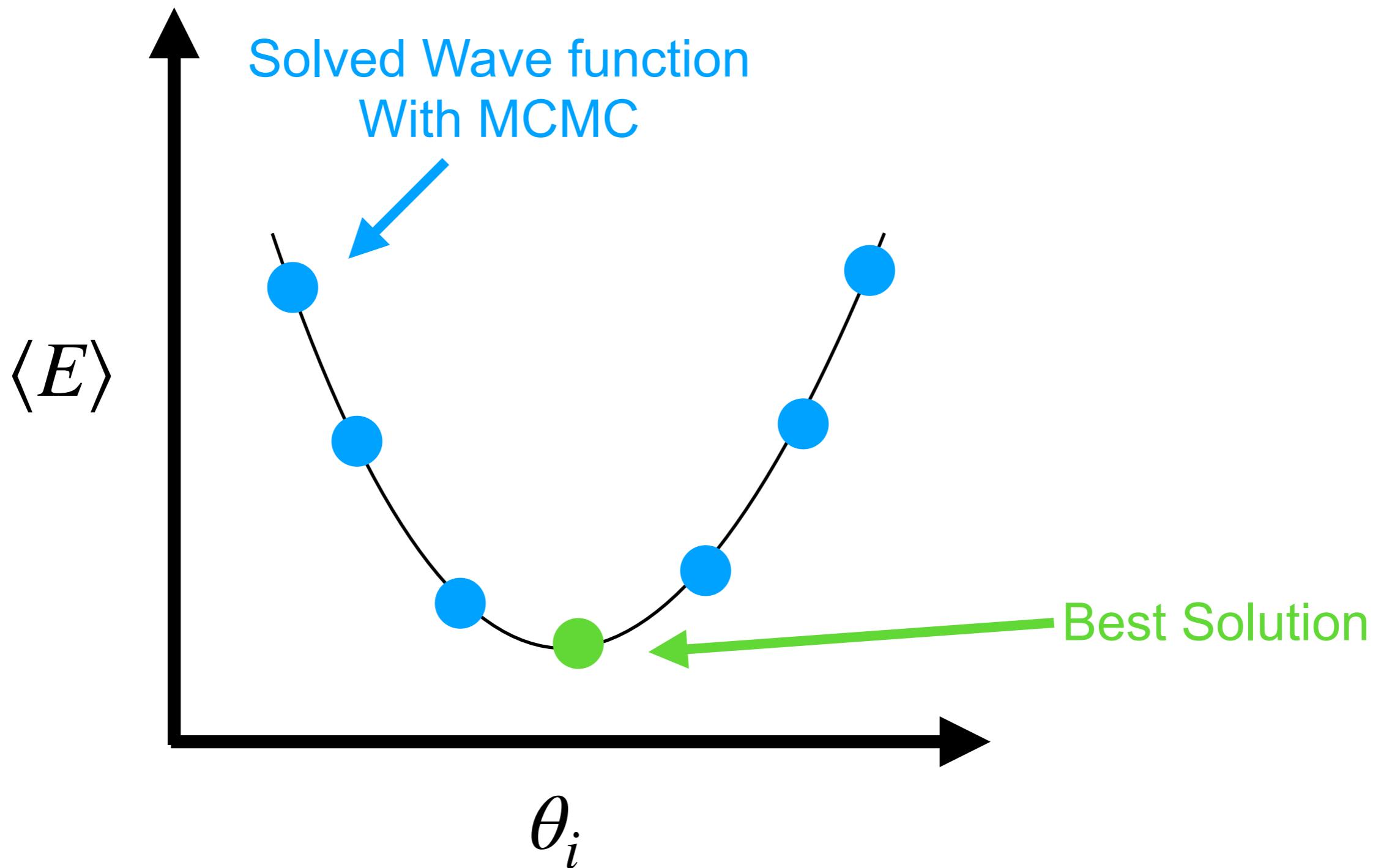
Solving Schroedinger



- Once we have the evolved wave function
 - We can compute expectations
 - No need to integrate (Really this is MC integration)

$$\langle E \rangle = \sum_j p_j(\vec{r} | \vec{\theta}) E_j(\vec{r} | \vec{\theta}) = \sum_j \psi_j^*(\vec{r} | \vec{\theta}) \psi_j(\vec{r} | \vec{\theta}) E_j(\vec{r} | \vec{\theta})$$

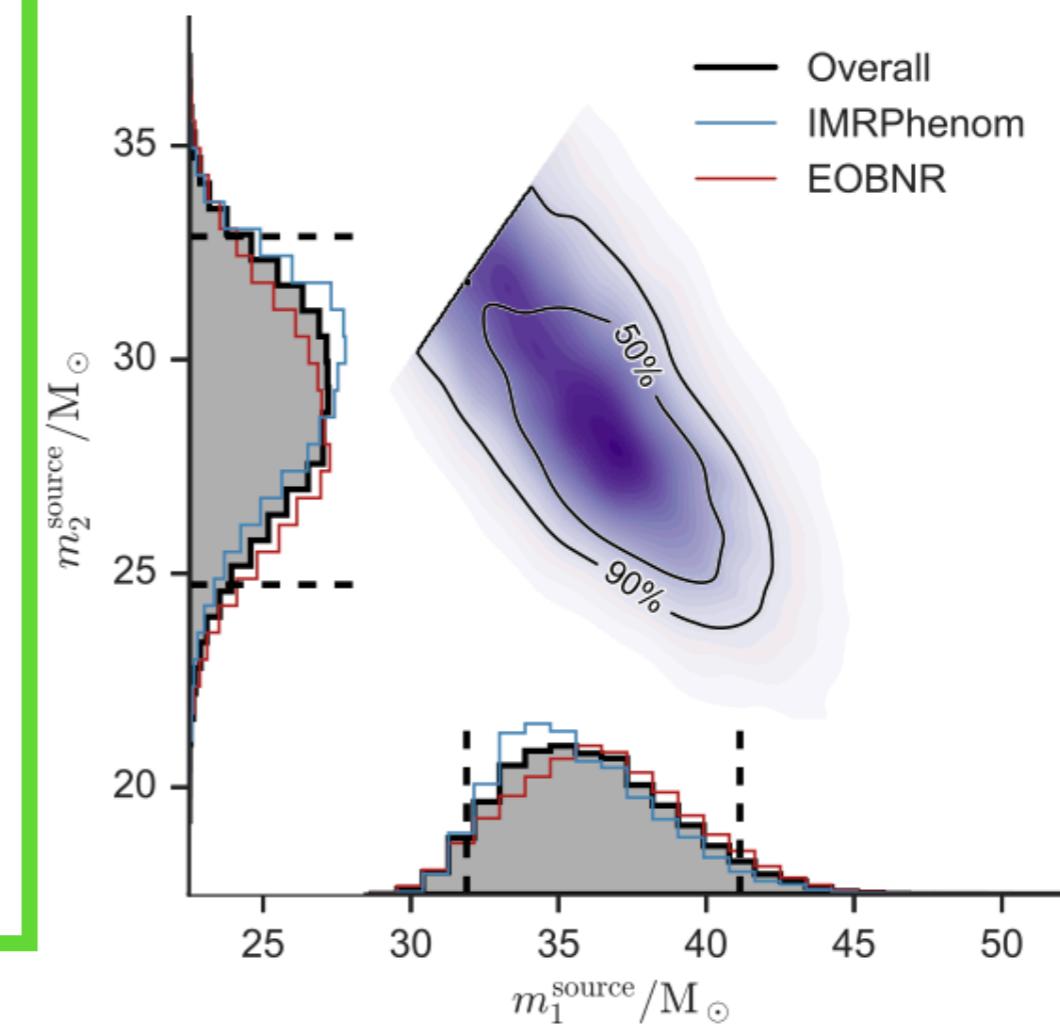
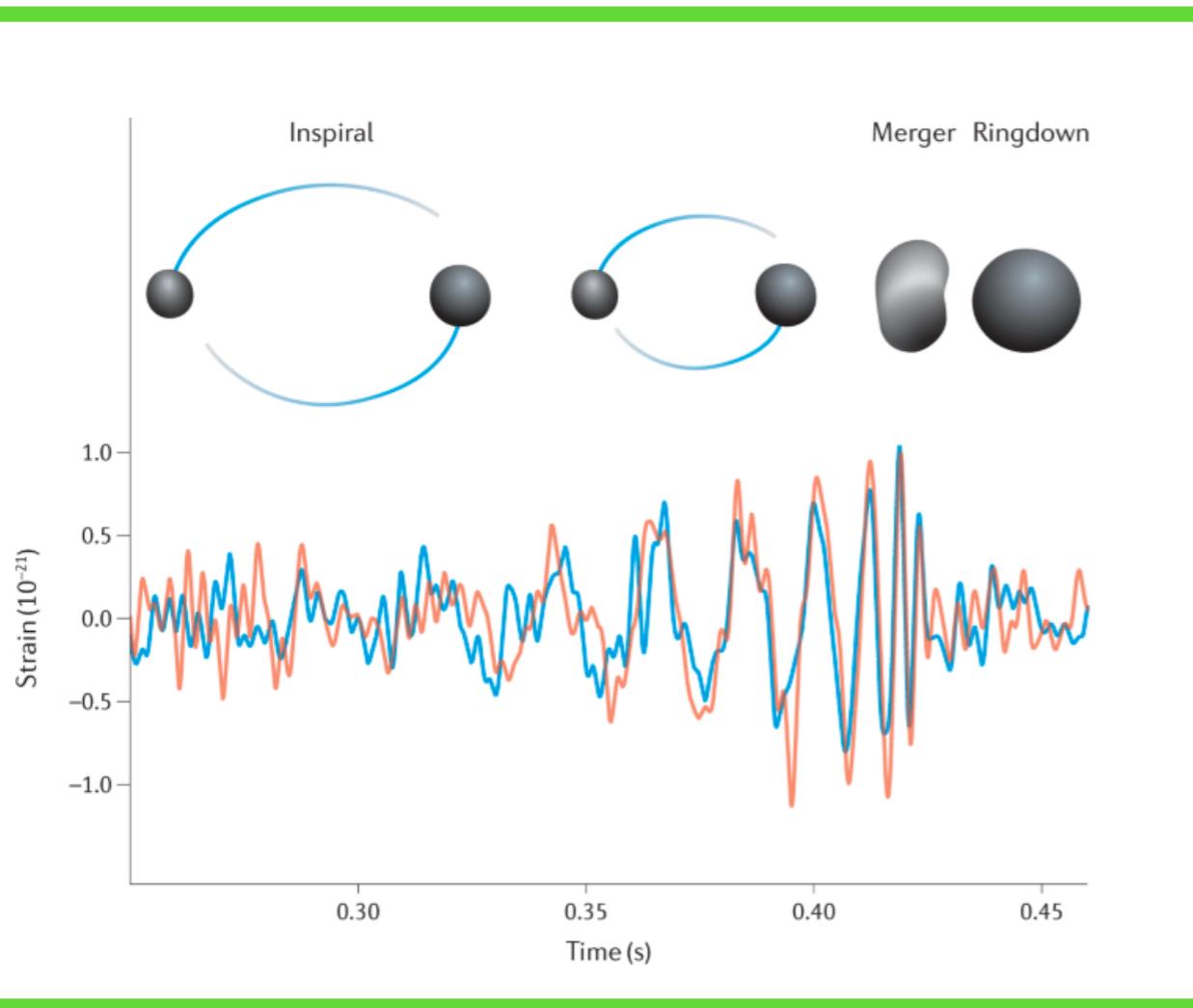
Solving Schroedinger



Our goal is to minimize the Energy given a wave functional form

Full Blow MCMC

- Ultimately the big gain from this are complex fits
 - Cases where normal gradient descent breaks down
 - What better case than to go back to LIGO



Full Waveform not differentiable (cannot fit it)

Observations

- There is some elegance to the MCMC approach
 - Builds directly to Bayesian fitting
 - MC allows us to explore parameters and correlations
- However, it is really slow
 - Migration towards differentiable loss is becoming popular
 - Training an NN to replace part of sampling helps with this
 - Project 3 starts to illustrate modern approach to this all

