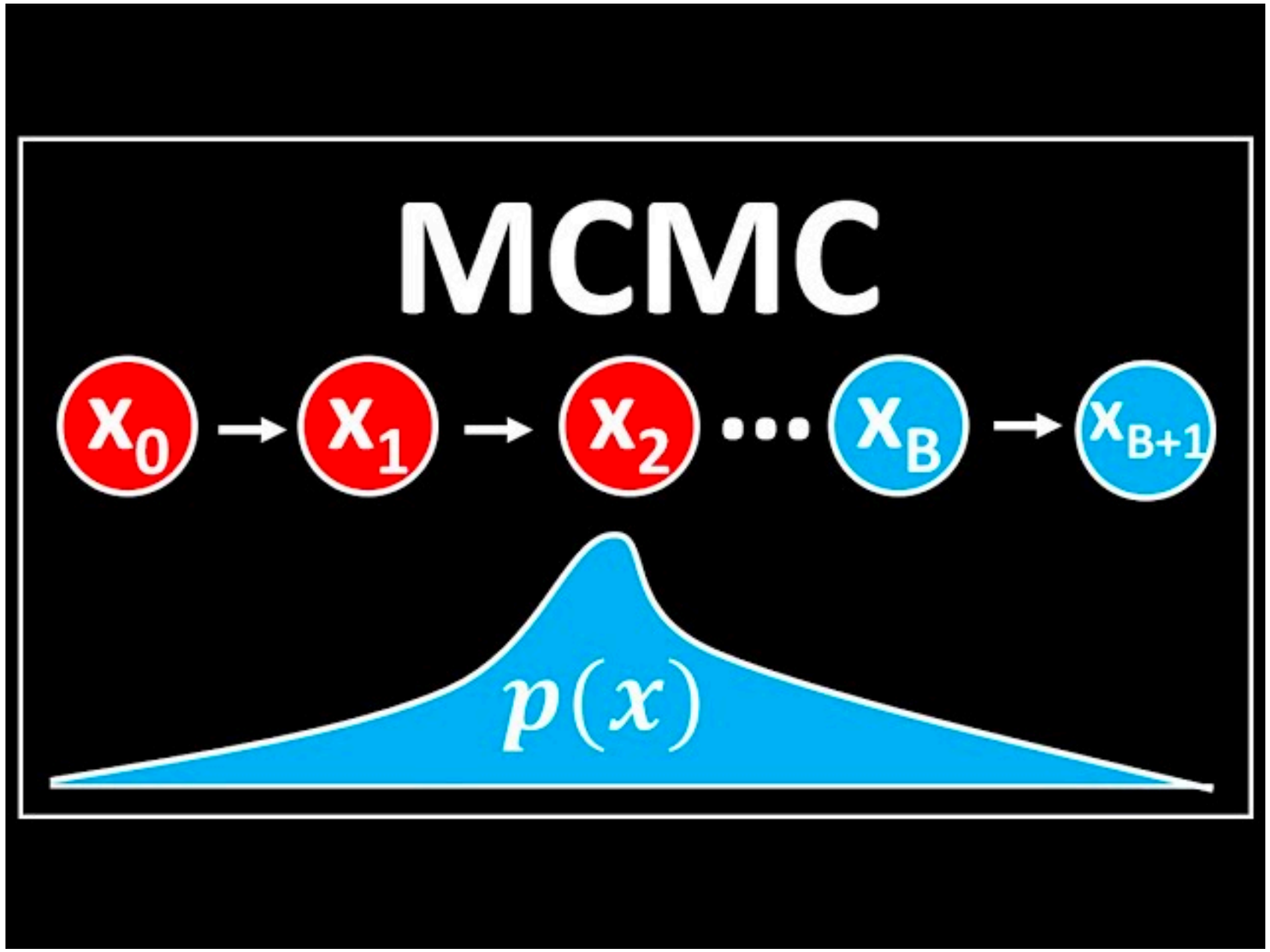
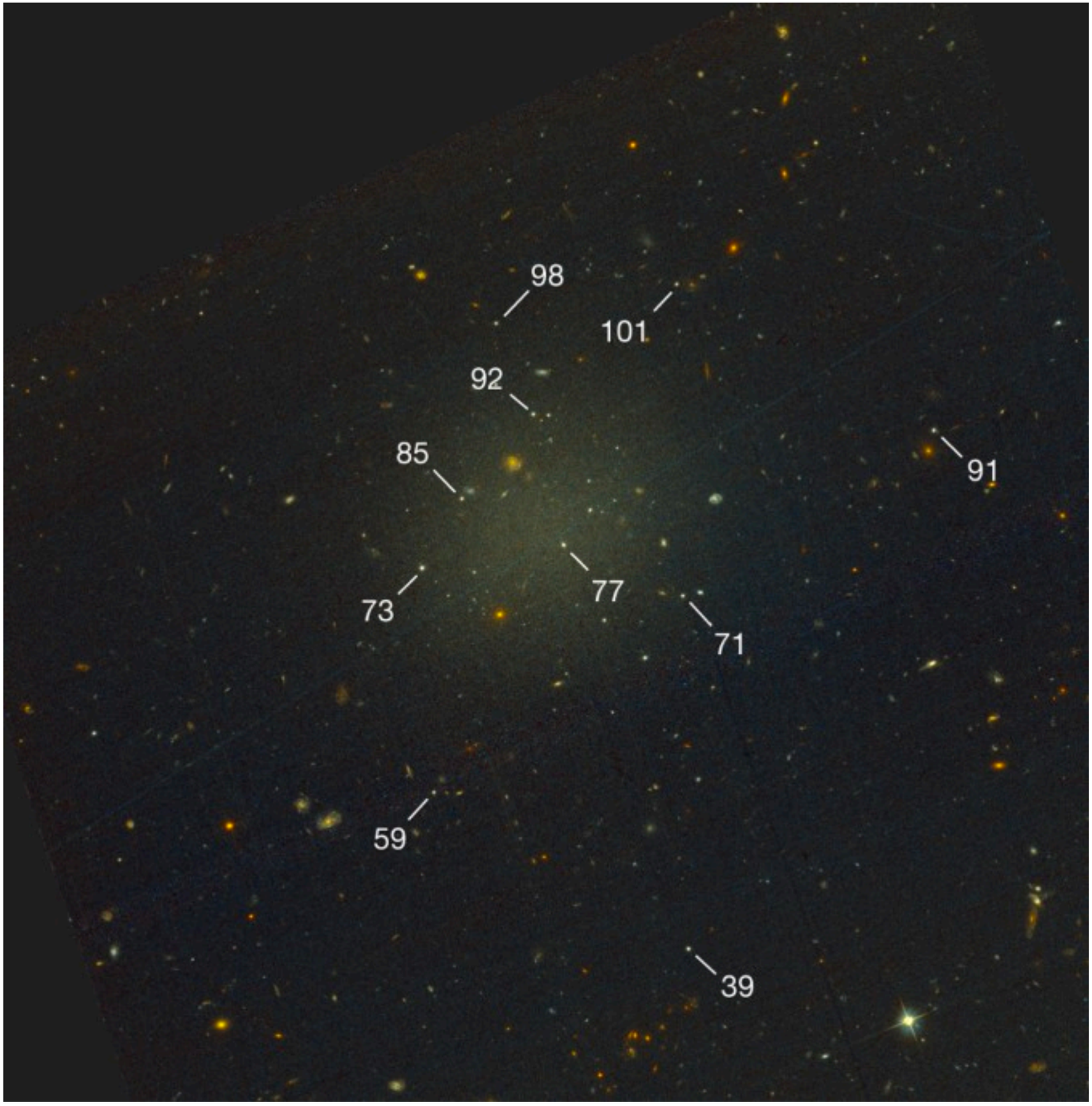
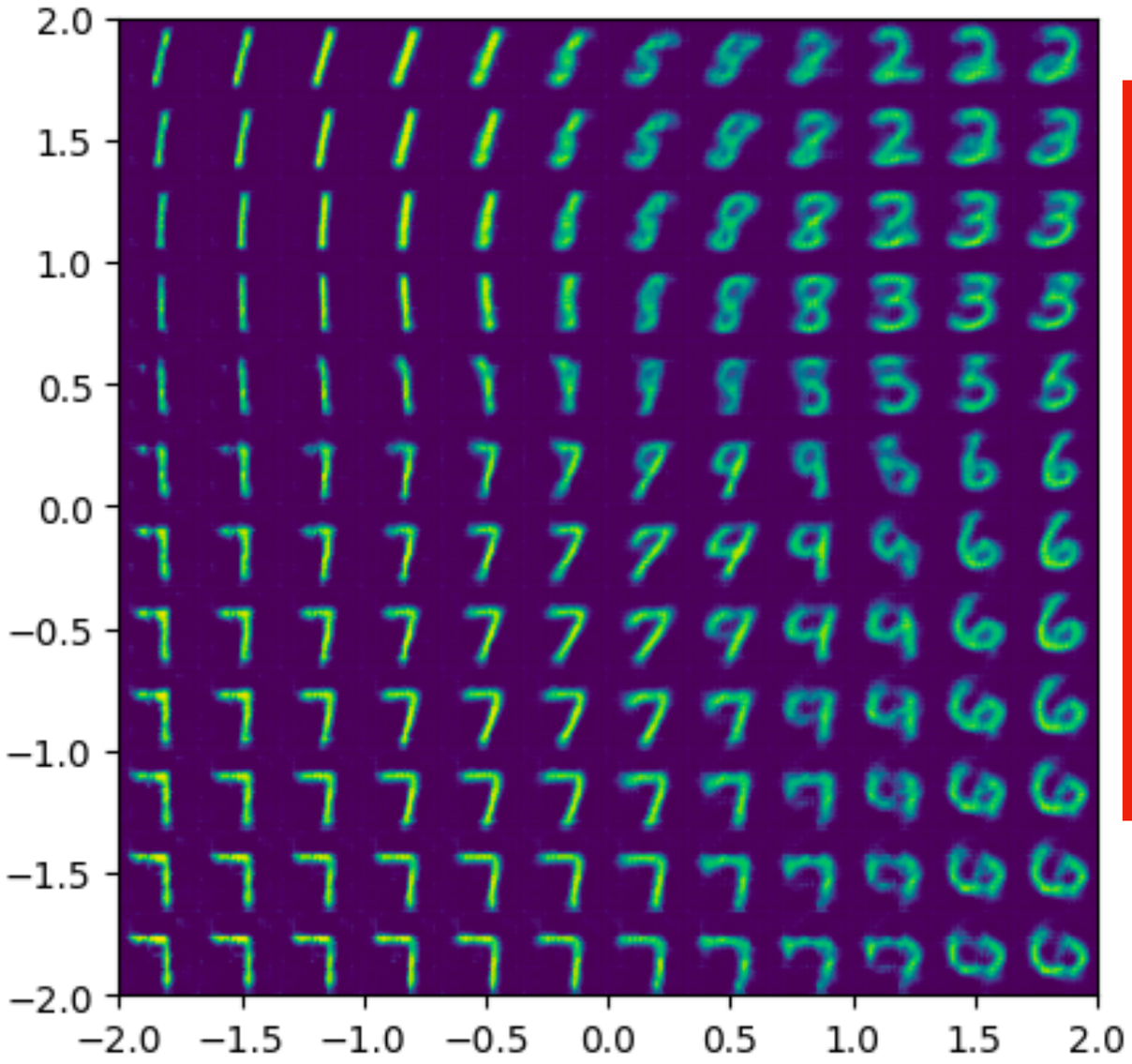


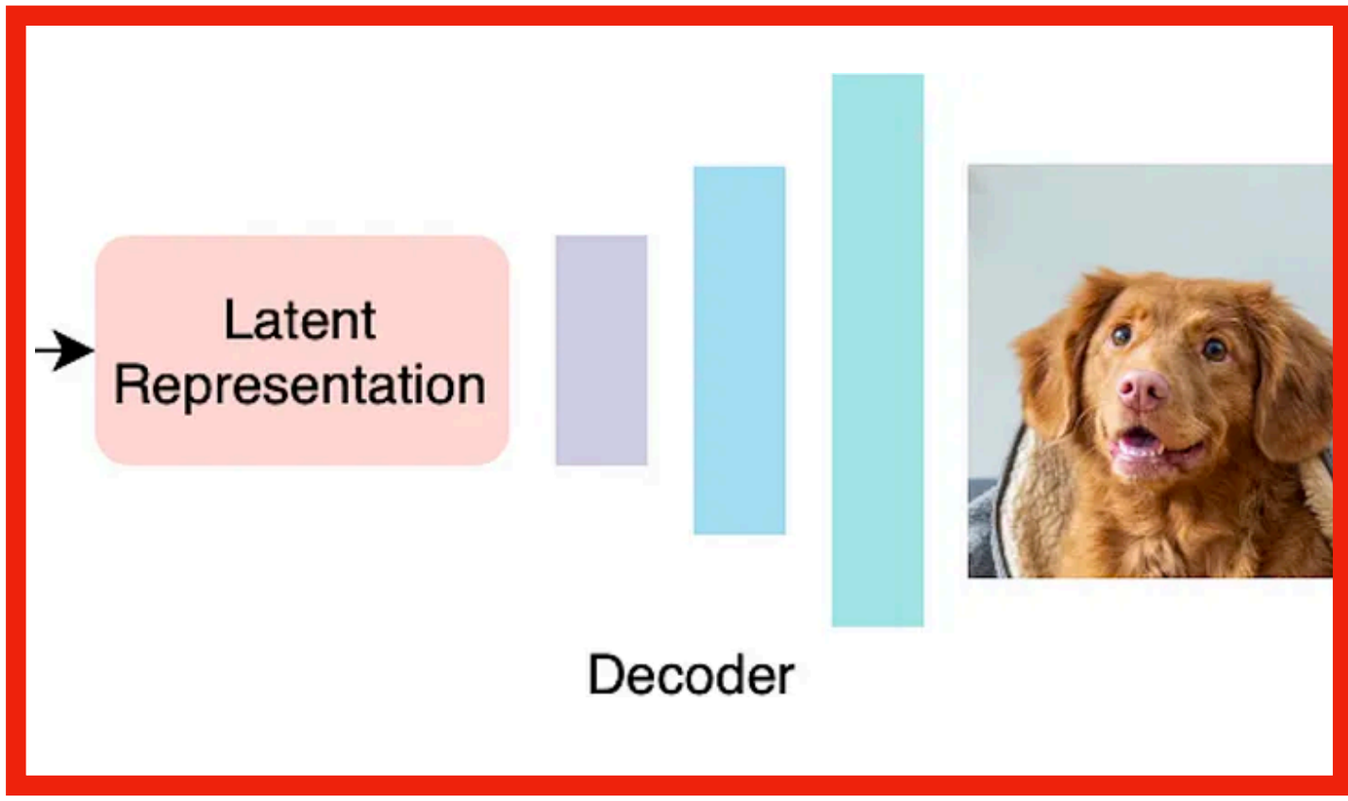
Lecture 20: Markov Chain MC

MCMC



Exploring the latent space?

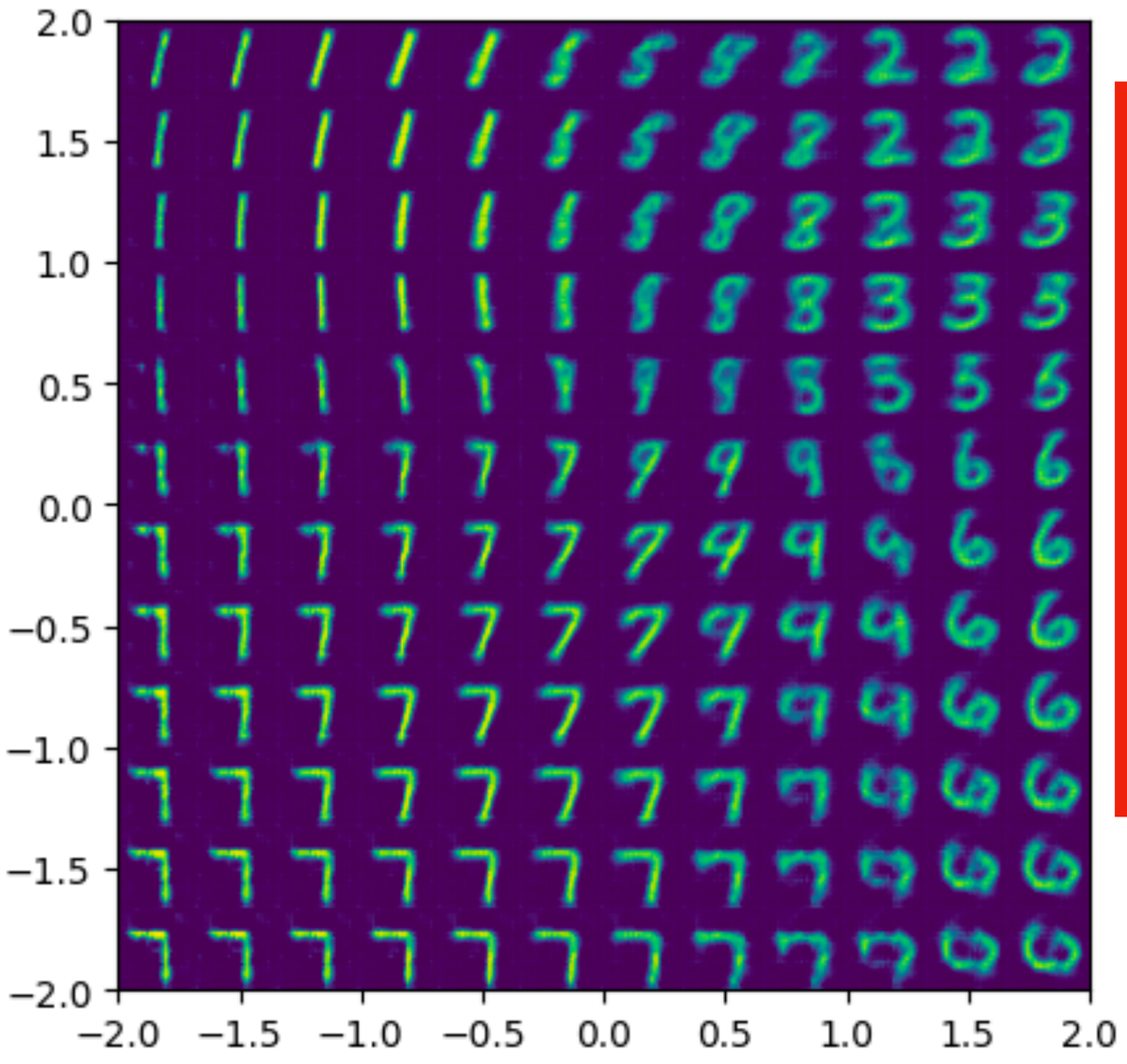


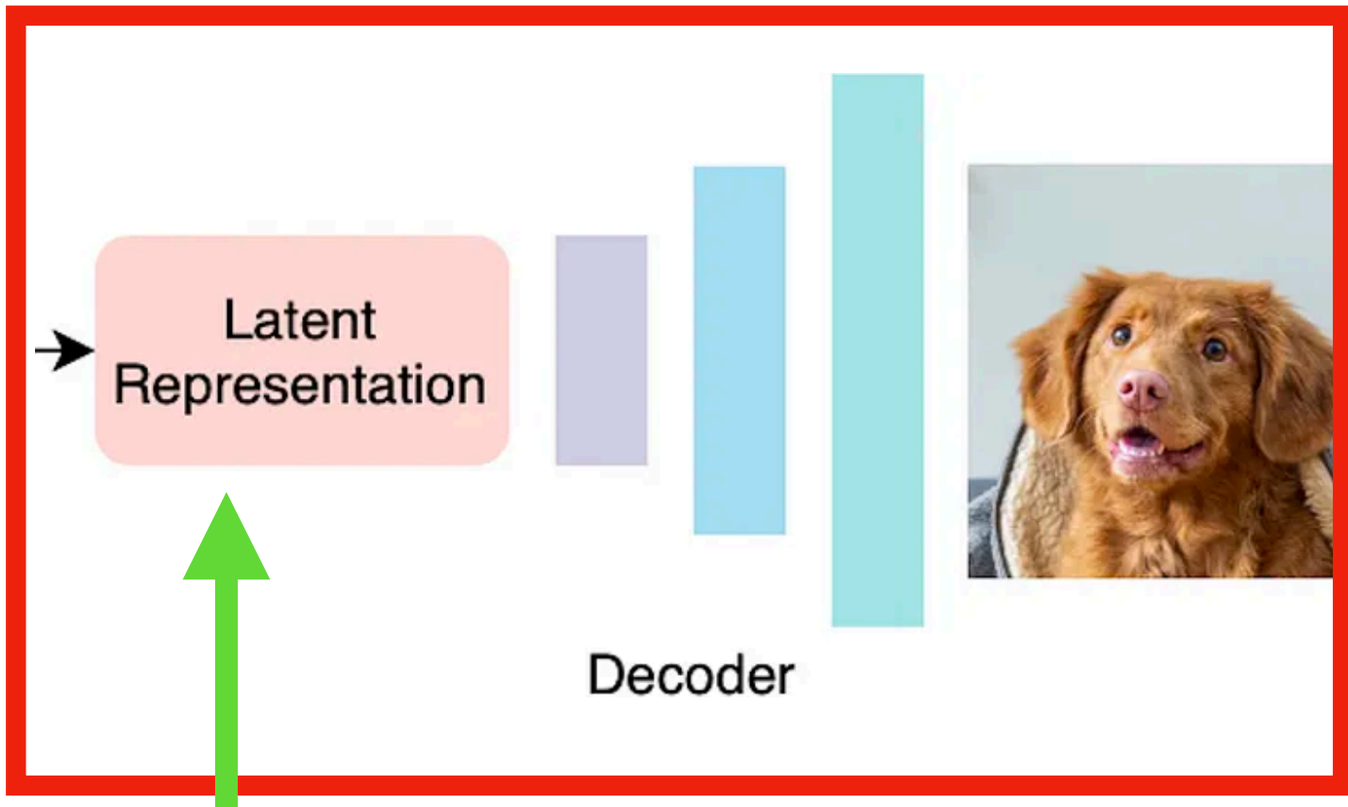


Generator

- We can sample the latent space as a generator

Conditional VAE





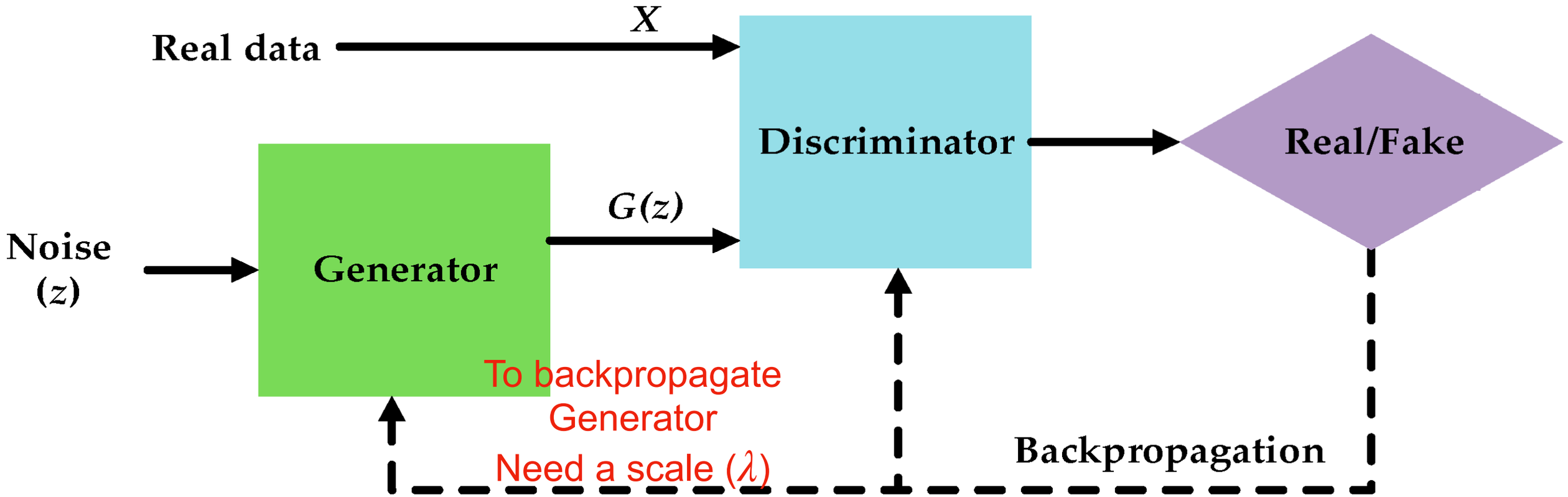
Generator

Variables that we would like to condition on

- Force known inputs into the VAE
  - That way our latent space has explicit knowledge of what is going on

GAN

- The other AI way of generating events uses
  - Generative Adversarial Networks (GANs)
  - I personally do not like these networks
    - They need too much tuning, **don't recommend them**

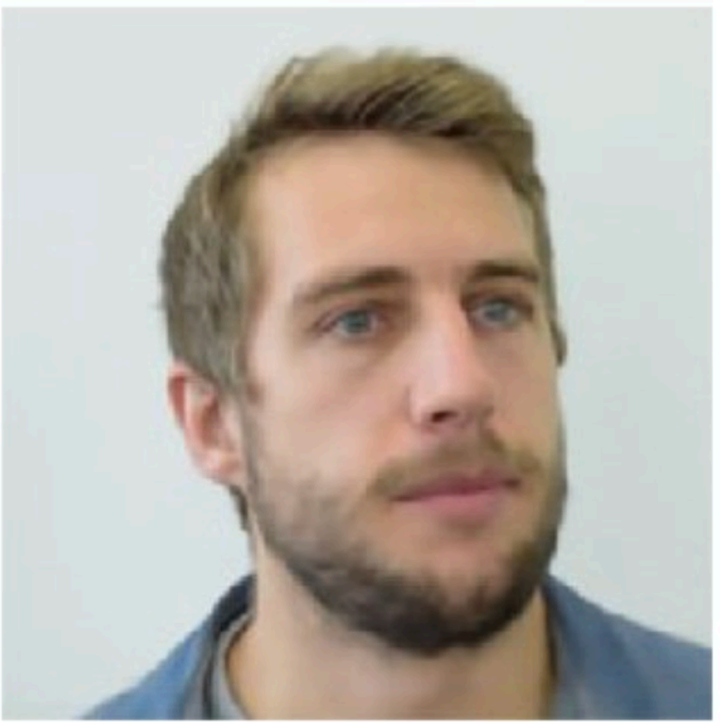


Power of GANs

- GANs became popular for generating fake human faces
  - Not as a hard a problem as typical physics problems
  - Our proton problem is probably already too hard for it

GAN

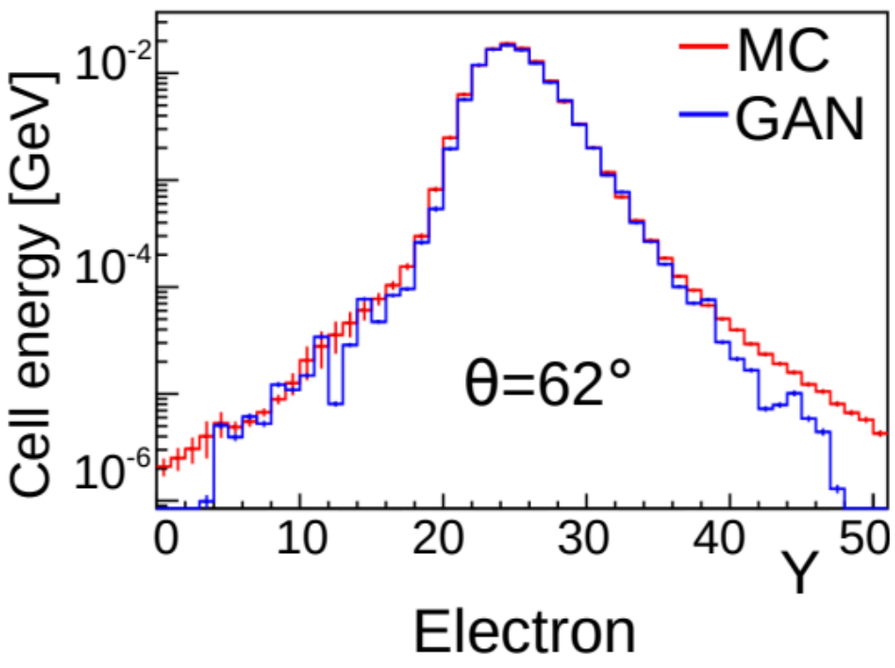
Autoencoder

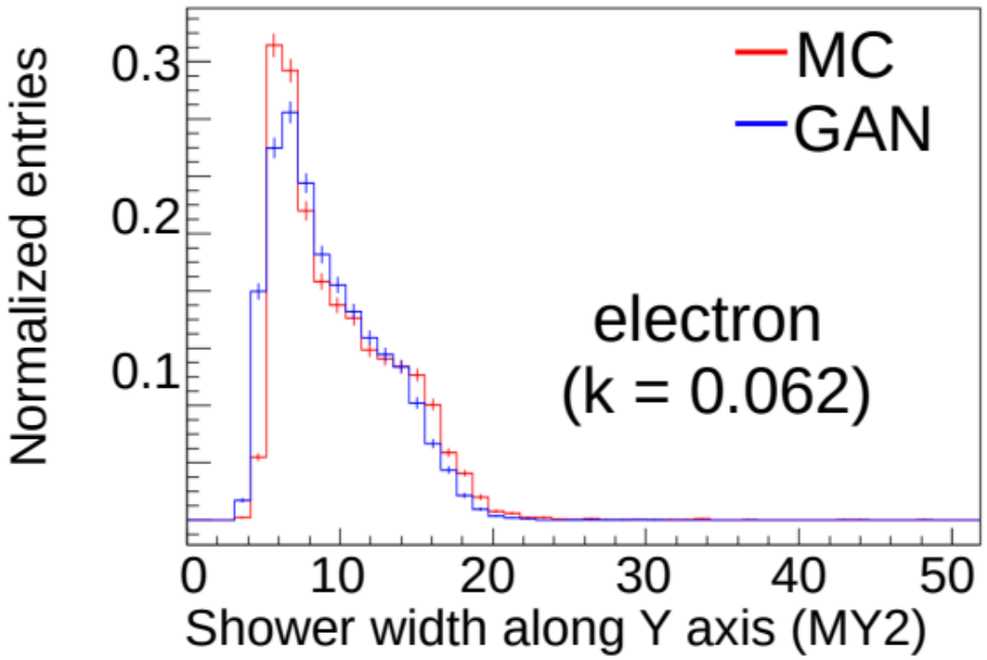


NeRF

Weakness of GANs

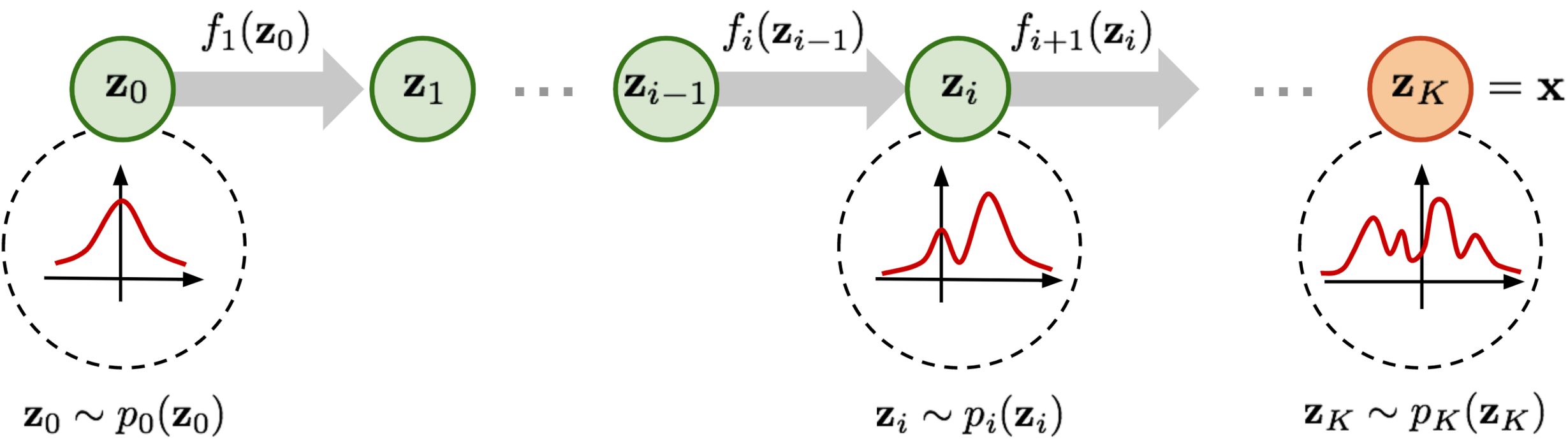
- GANs have not been great for physics
  - While they get the bulk of simulation ok
  - Struggle with the many orders of magnitude needed
  - Also they are pain to tune





Normalizing Flows

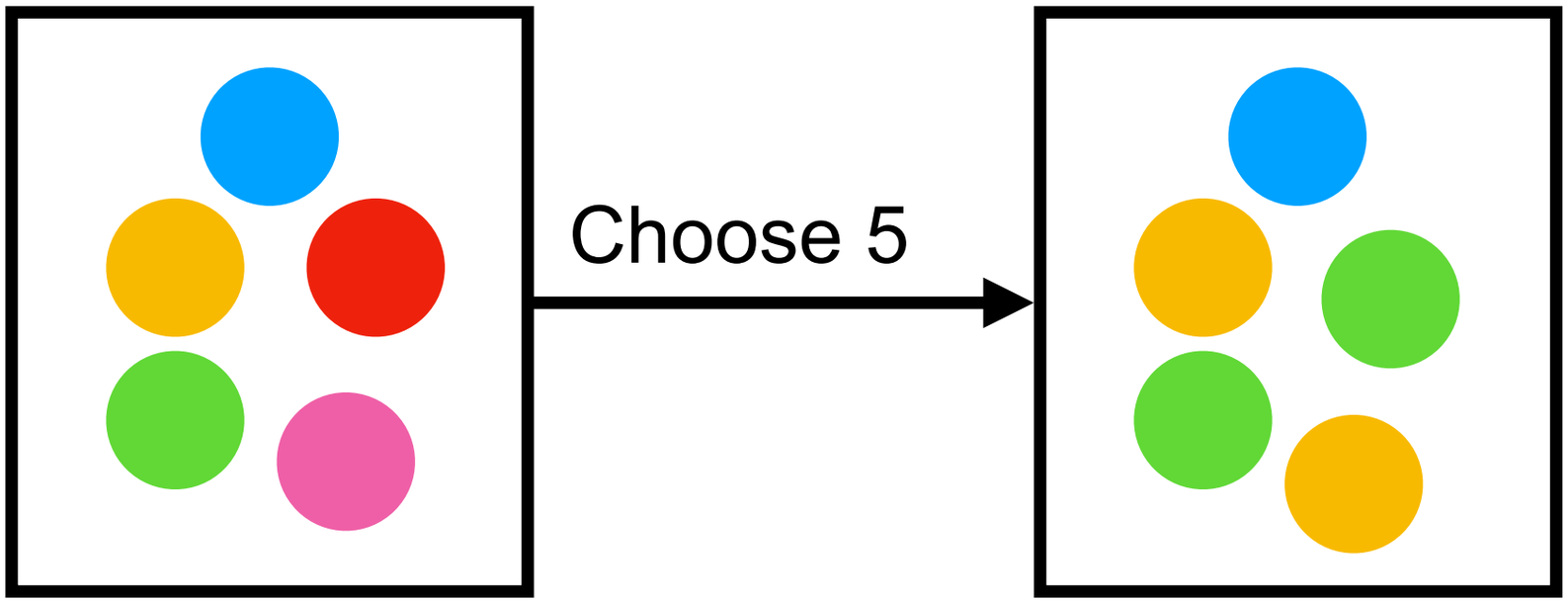
- While I don’t show it here
  - The normalizing flow is next evolution of VAE
  - Allows for a more expressive latent space
- Normalizing flow transforms flows to other spaces



Bootstrapping

- Lets say you have a fixed dataset ($\vec{x}$)
  - You cannot generate more
  - You compute something very complicated on it
    - For example $\vec{y} = \mathrm{NN}(\vec{x})$
- You want to know the uncertainty on $\vec{y}$ given by $\sigma_{\vec{y}}$
  - However your function $\frac{d\mathrm{NN}(\vec{x})}{dx} \notin$ exists

Bootstrapping: Strategy



You will
Get Duplicates

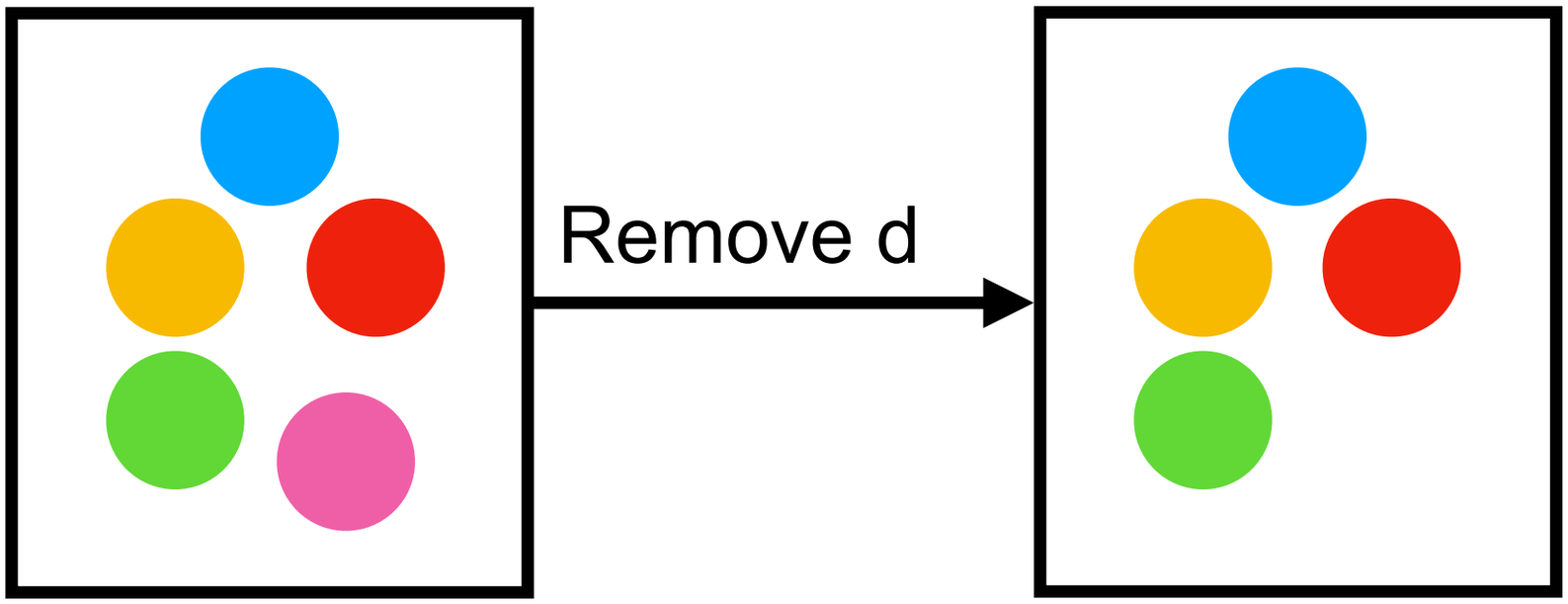
However if
things
Are truly
poission its ok

- Randomly sample the pool of events with the same size
  - Now use this sample to compute all our observables
- Variance over our random samples will be the uncertianty

$$\sigma_{\text{boot}} = \frac{1}{N_{\text{samples}} - 1} \sum_{i}^{N_{\text{samples}}} \left(\mathcal{O} - \bar{\mathcal{O}} \right)^2$$

We can do this
with any
observable

Delete-D Jackknife



- Like what we did with Bootstrapping
  - However, doesn't duplicate events

d=removed samples

$$\sigma_{\text{Jackknife}} = \frac{N_s - d}{\left({}_{N_s}C_d\right) d} \sum_{i}^{{}_{N_s}C_d} \left(\mathcal{O} - \bar{\mathcal{O}}\right)^2$$

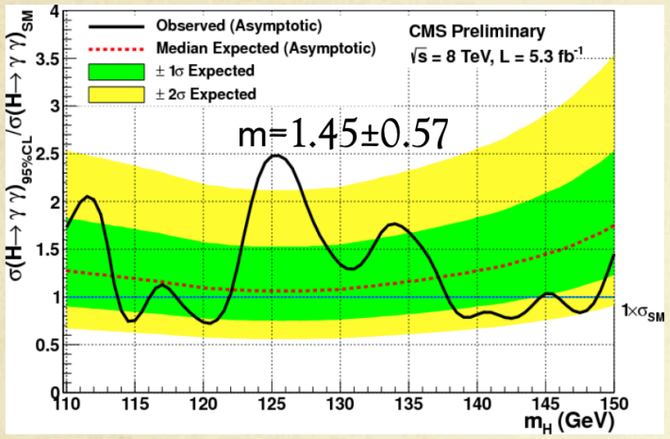
Total Unc
From Jackknife

Fun Story

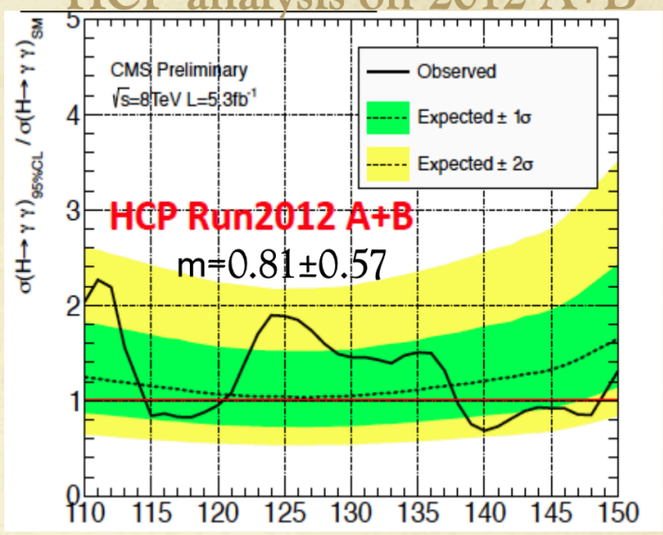
Higgs to diphoton result changed dramatically in 2012
On exactly the same data

The Observed Difference I

ICHEP analysis on 2012 A+B



HCP analysis on 2012 A+B

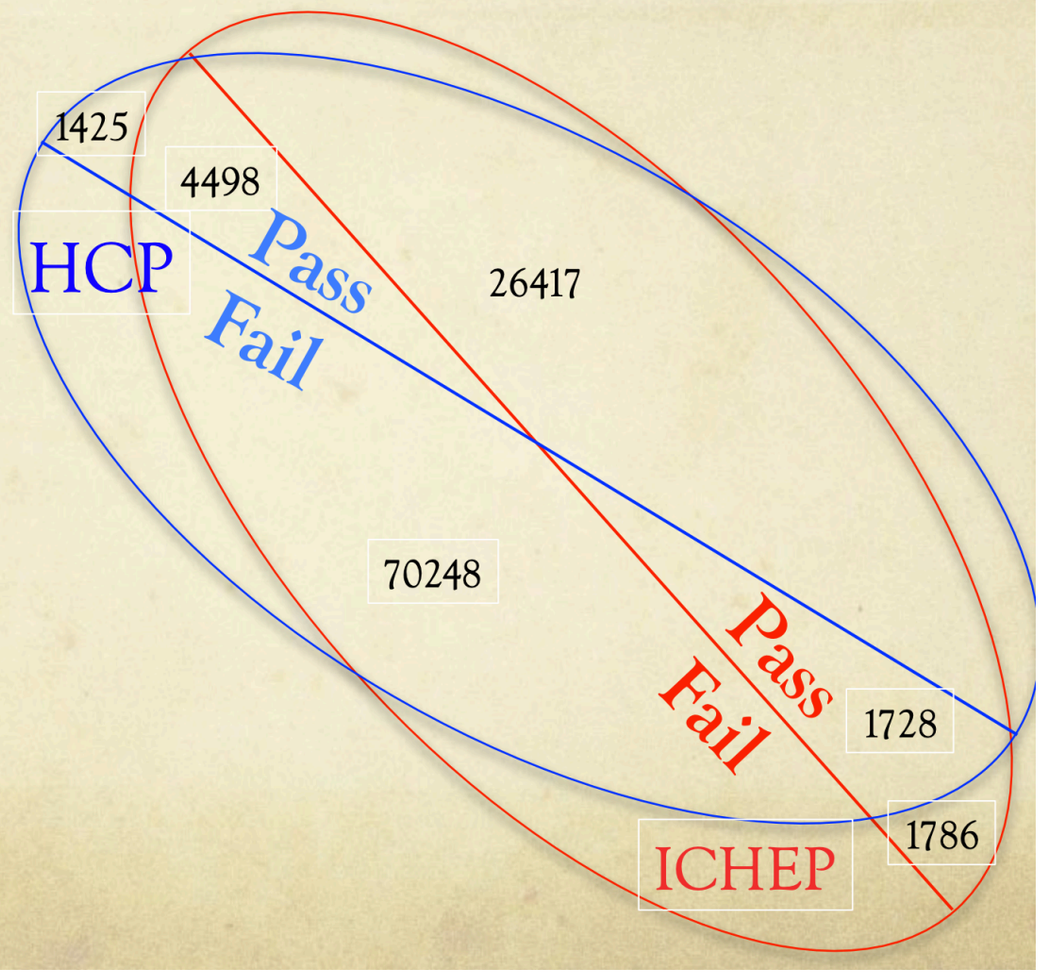


Difference = 0.79 σ_{diff}, but analysis and events are correlated
All checks concentrate on the difference between ICHEP and HCP Analyses on 2012 A+B.

Fun Story

Higgs to diphoton result changed dramatically in 2012
On exactly the same data

- Use Globe HCP analysis in 5.3
- Switch in steps back to ICHEP using event dump of ICHEP analysis
- 1786 Missing from HCP cannot be compared

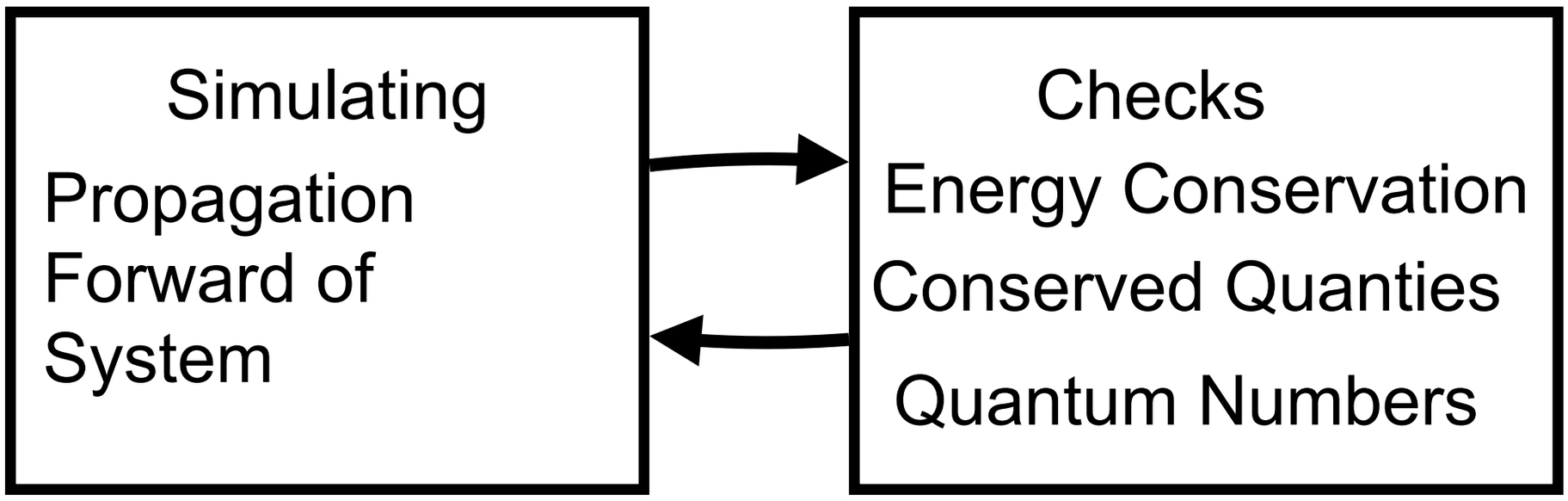


CMS UCSD

Delete-d Jackkinfe was the way we were able to analyze the unc.

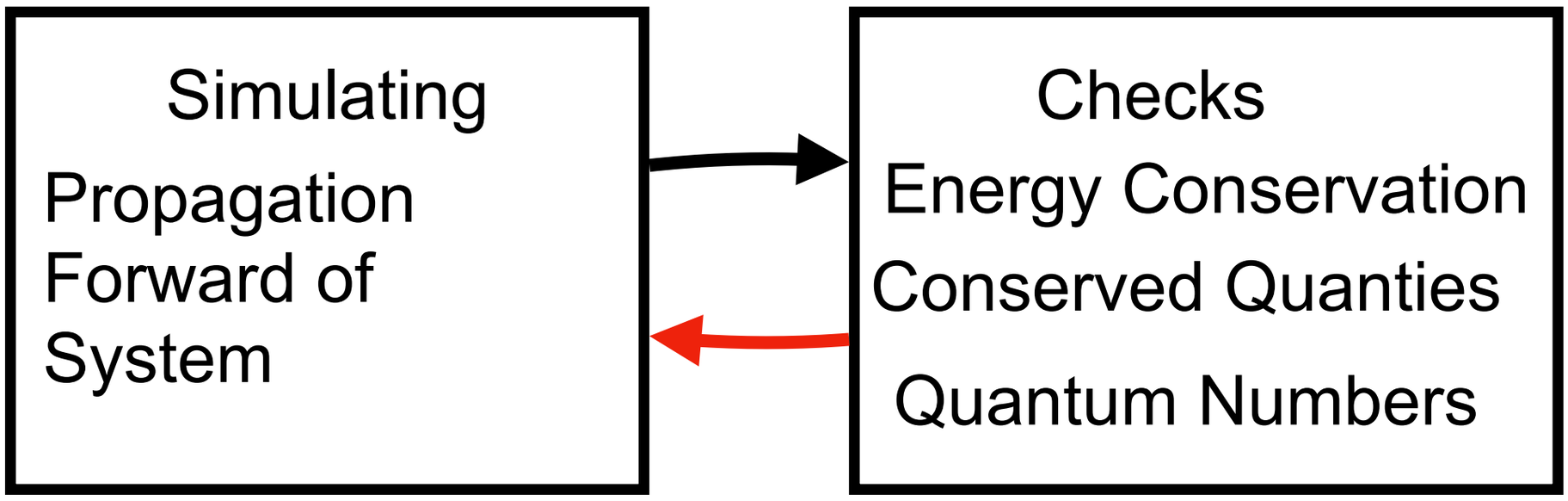
Making MC Better

- 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



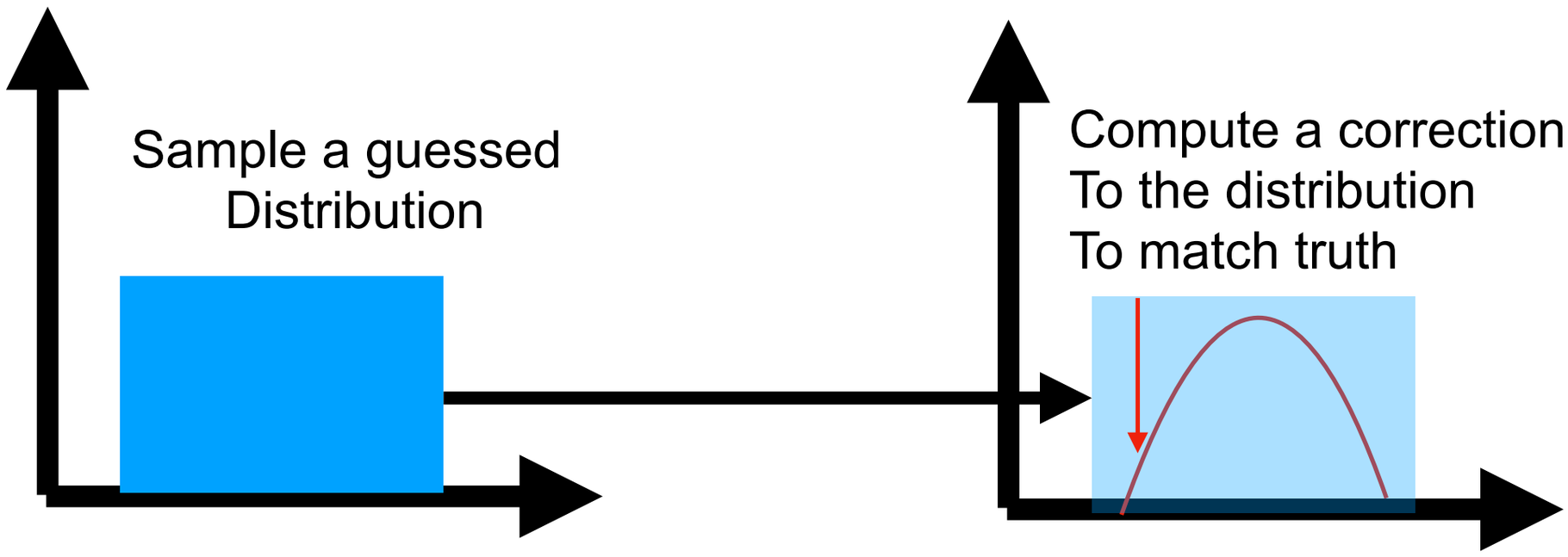
Correcting

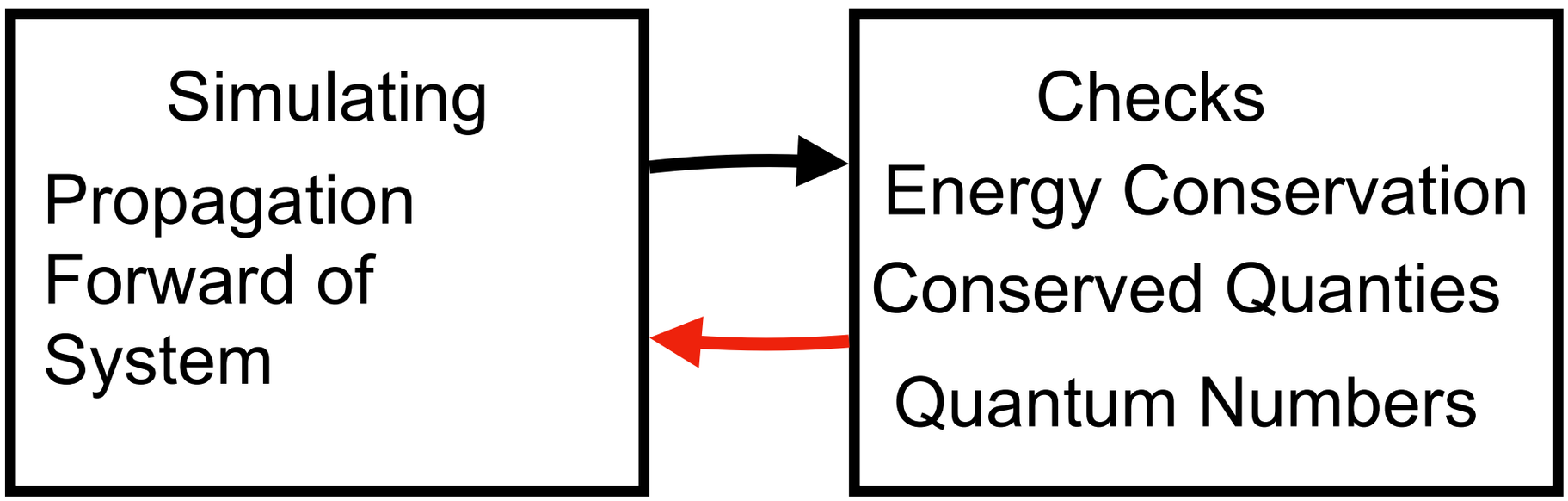
- In this simulation part of this class
  - We have learned that simulations are not accurate
  - There are a few things we can do to make it better
    - Key is to have a notion of when we are going wrong



Correct Our Simulation Through a Probabilistic Rescaling

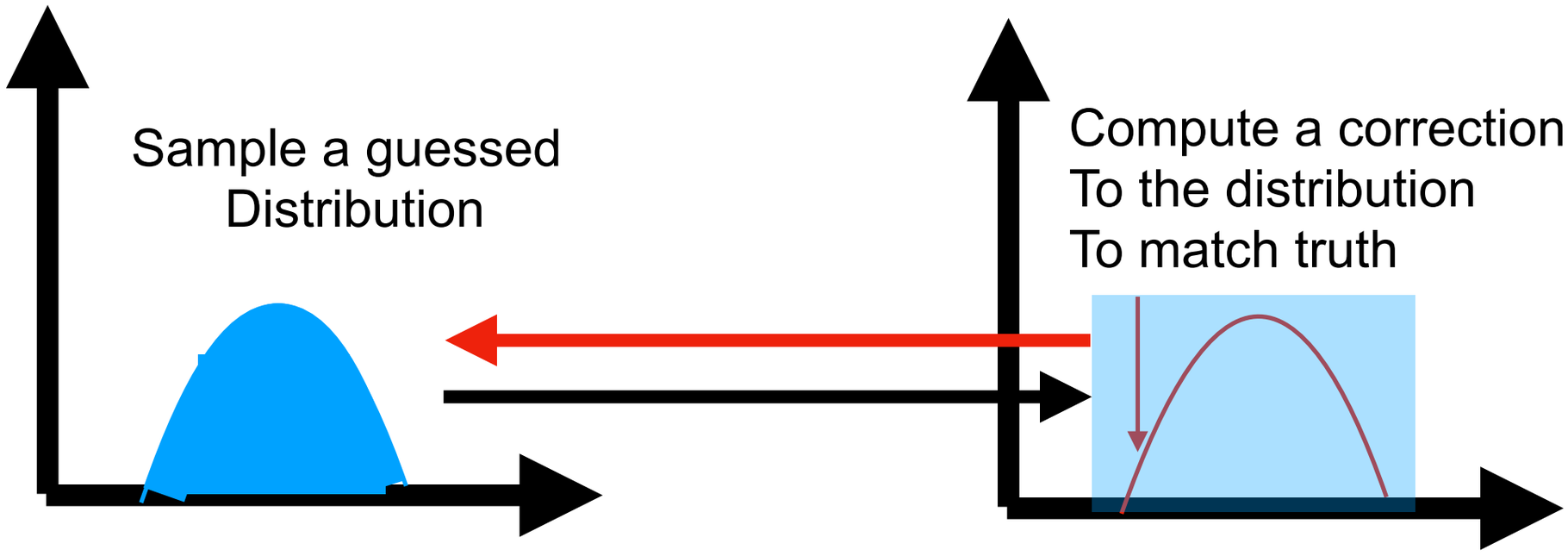
Markov Chain MC

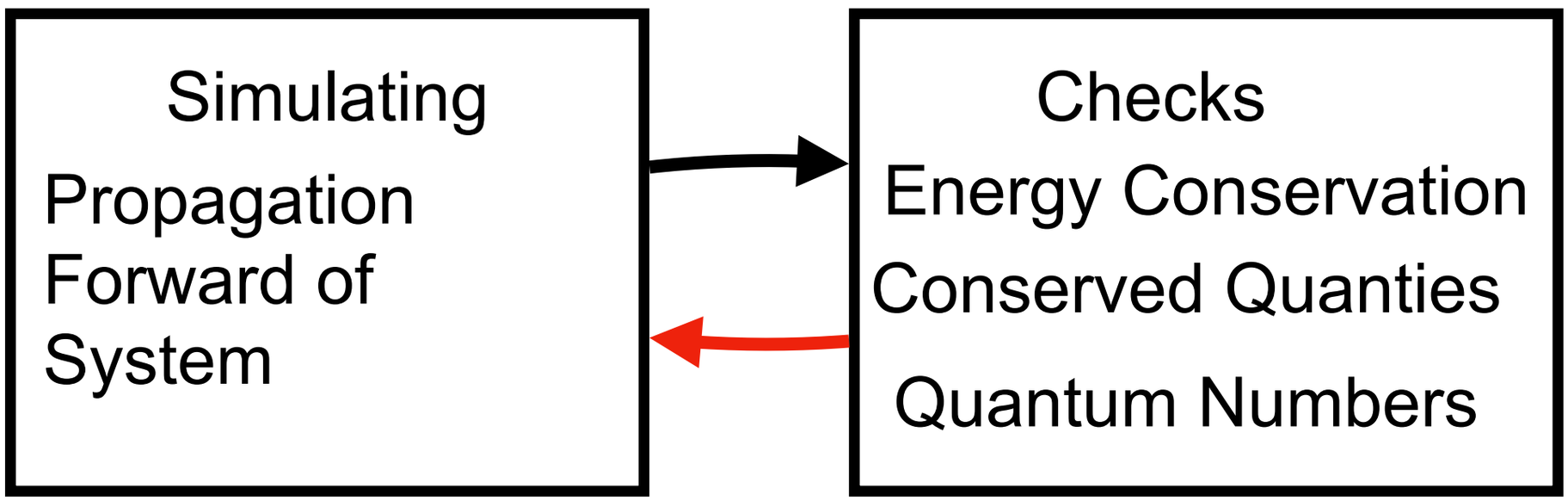




Correct Our Simulation Through a Probabilistic Rescaling

Markov Chain MC





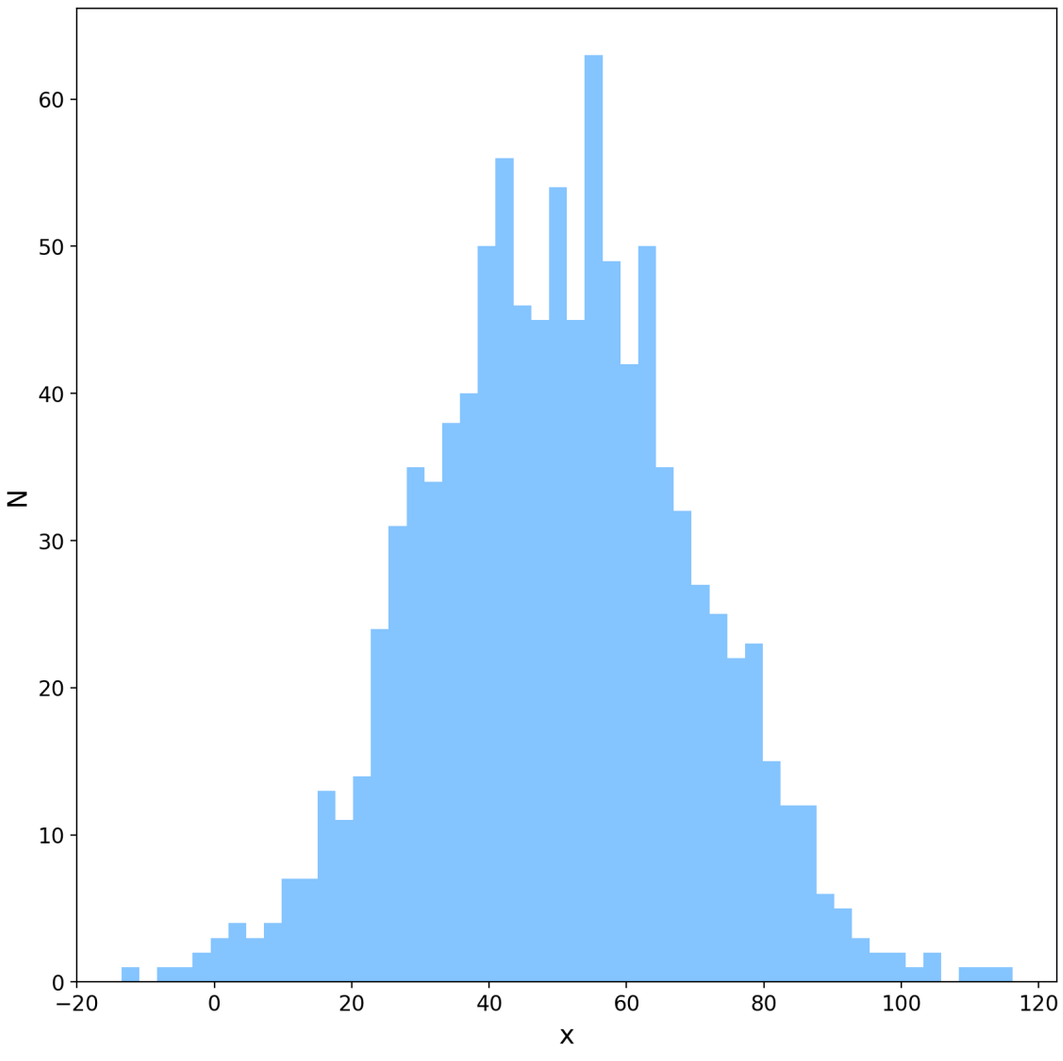
Correct Our Simulation Through a Probabilistic Rescaling

Metropolis-Hastings

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

Fitting a Guassian

- 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



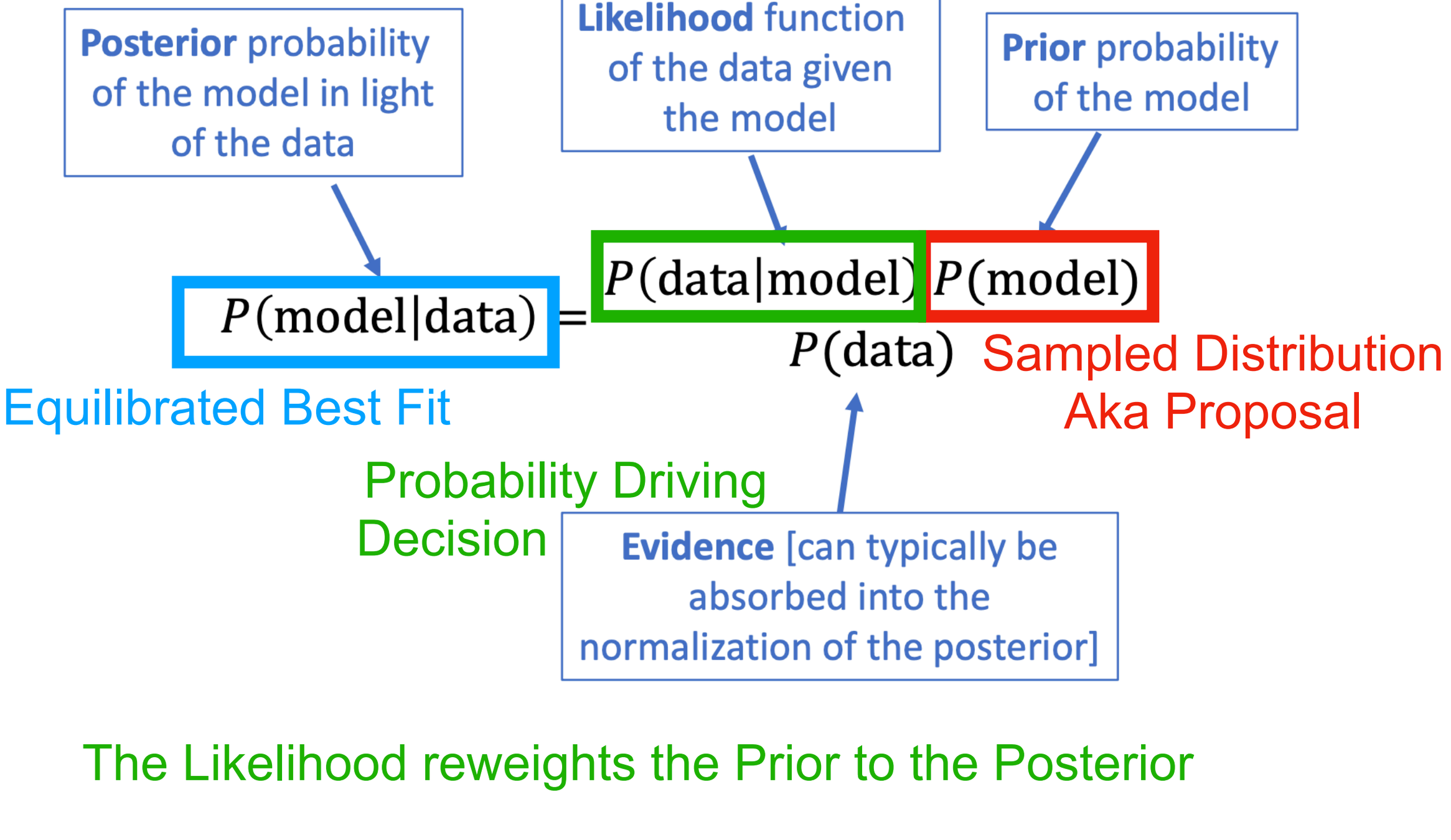
Log(Probability)

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

$$= \sum_i \left(\frac{x_i - \mu}{\sigma}\right)^2 - \frac{1}{2}\log\left(2\pi\sigma^2\right)$$

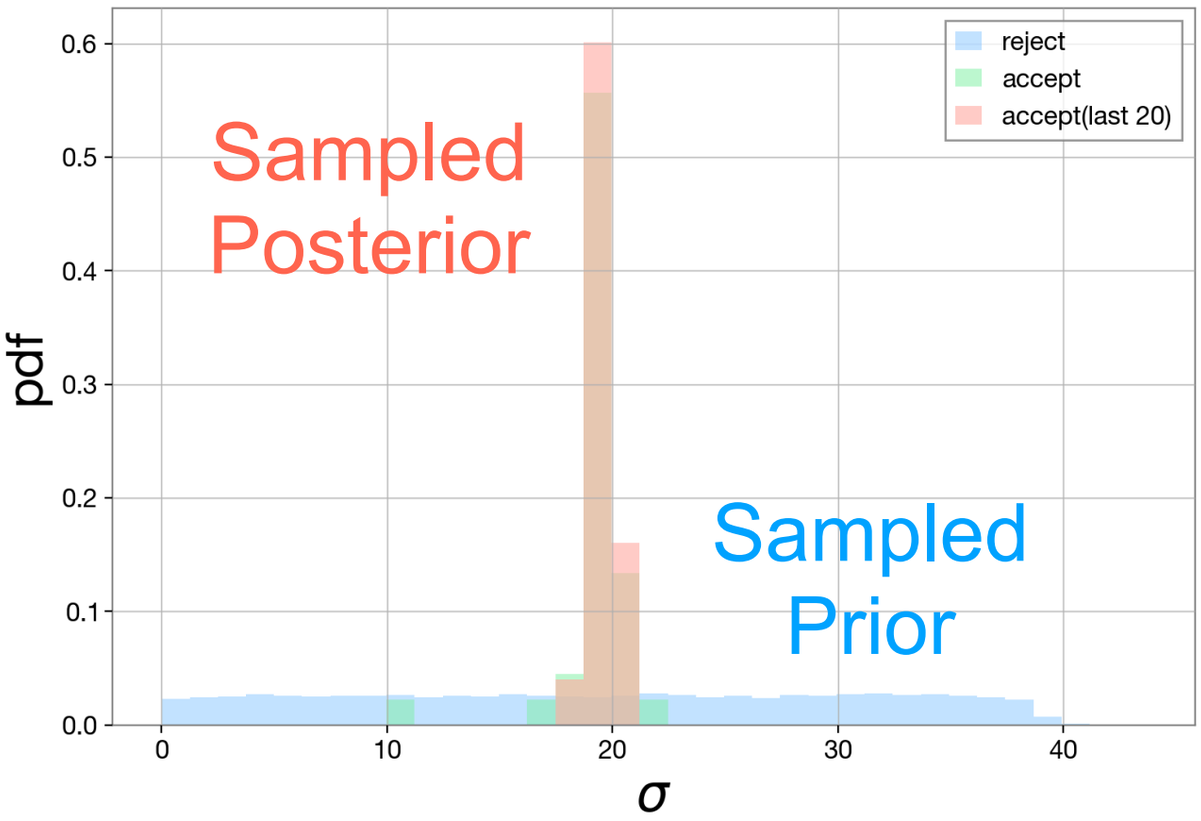
Accepted μ, σ yield the best fits

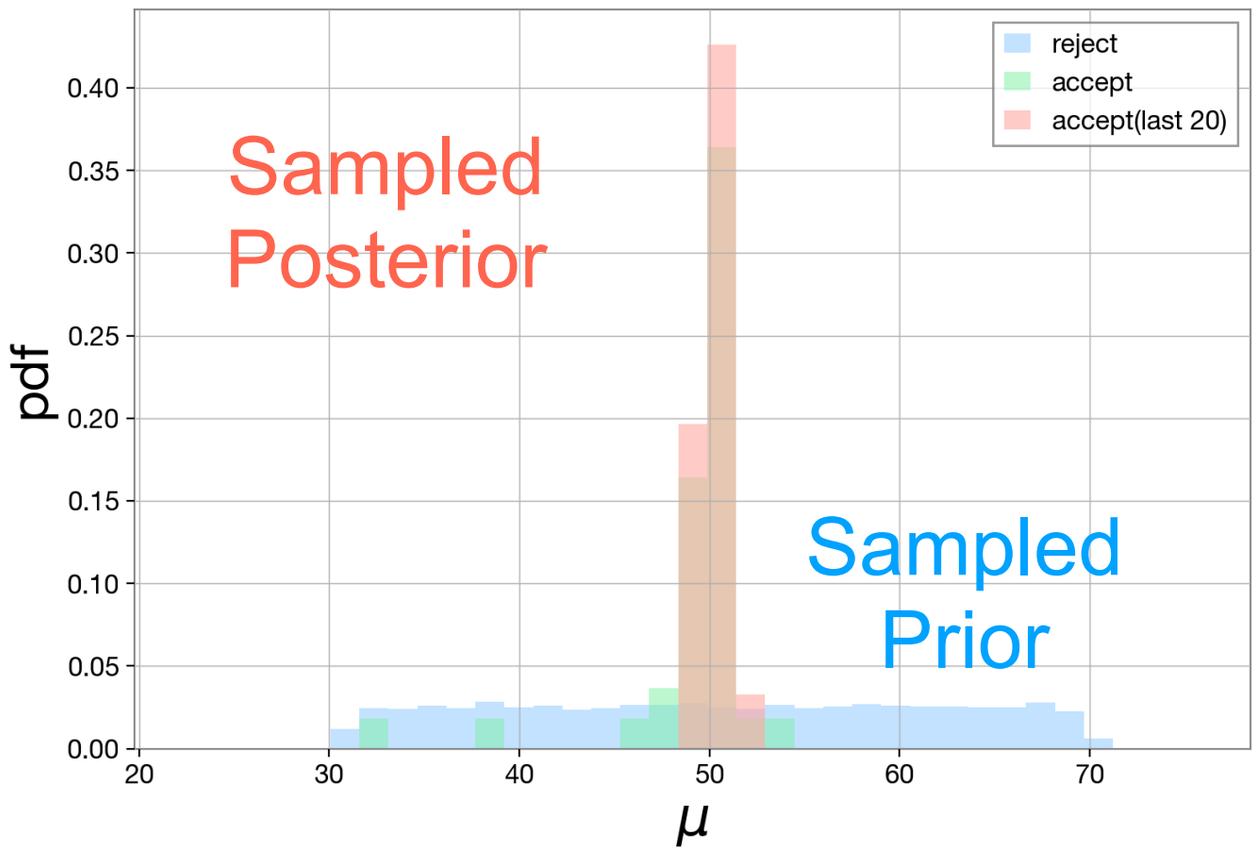
Visualizing in Bayes



The Likelihood reweights the Prior to the Posterior

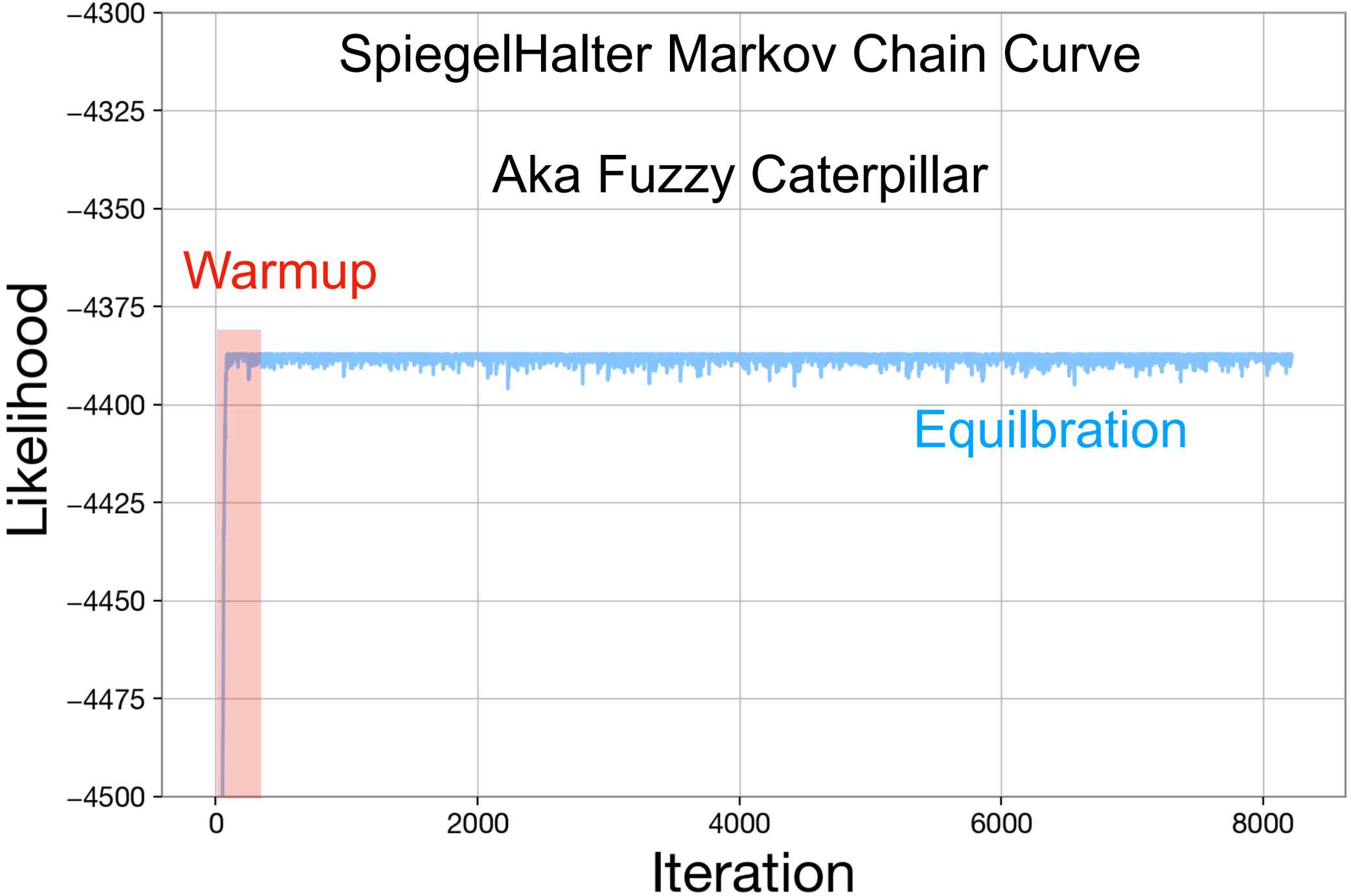
Best Fit Paramters





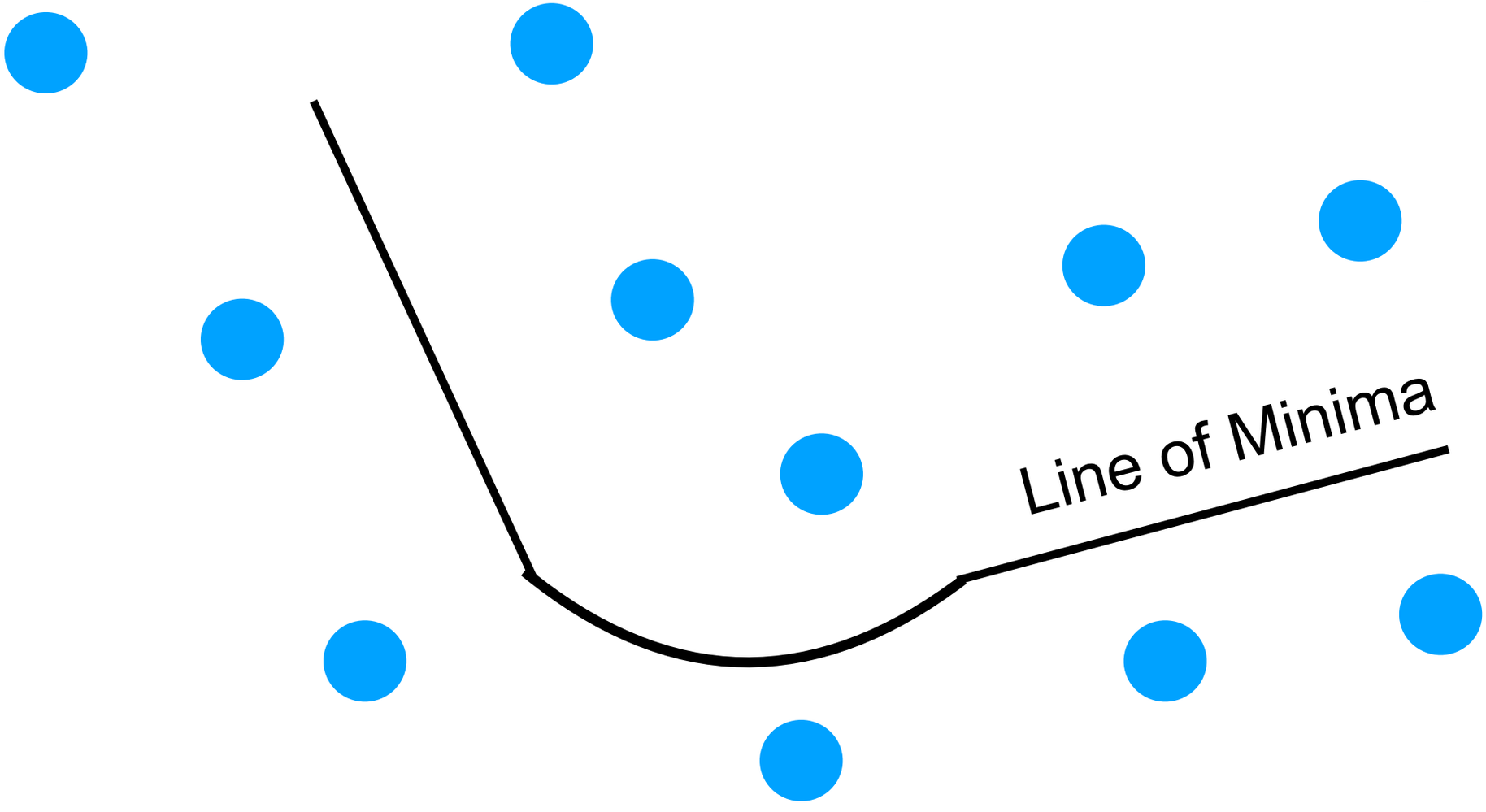
- This is where I start to hint that there are limitations here
- What we are doing is running a check
  - We are not taking a derivative (No gradients or Hessians)

Best Fit Paramters



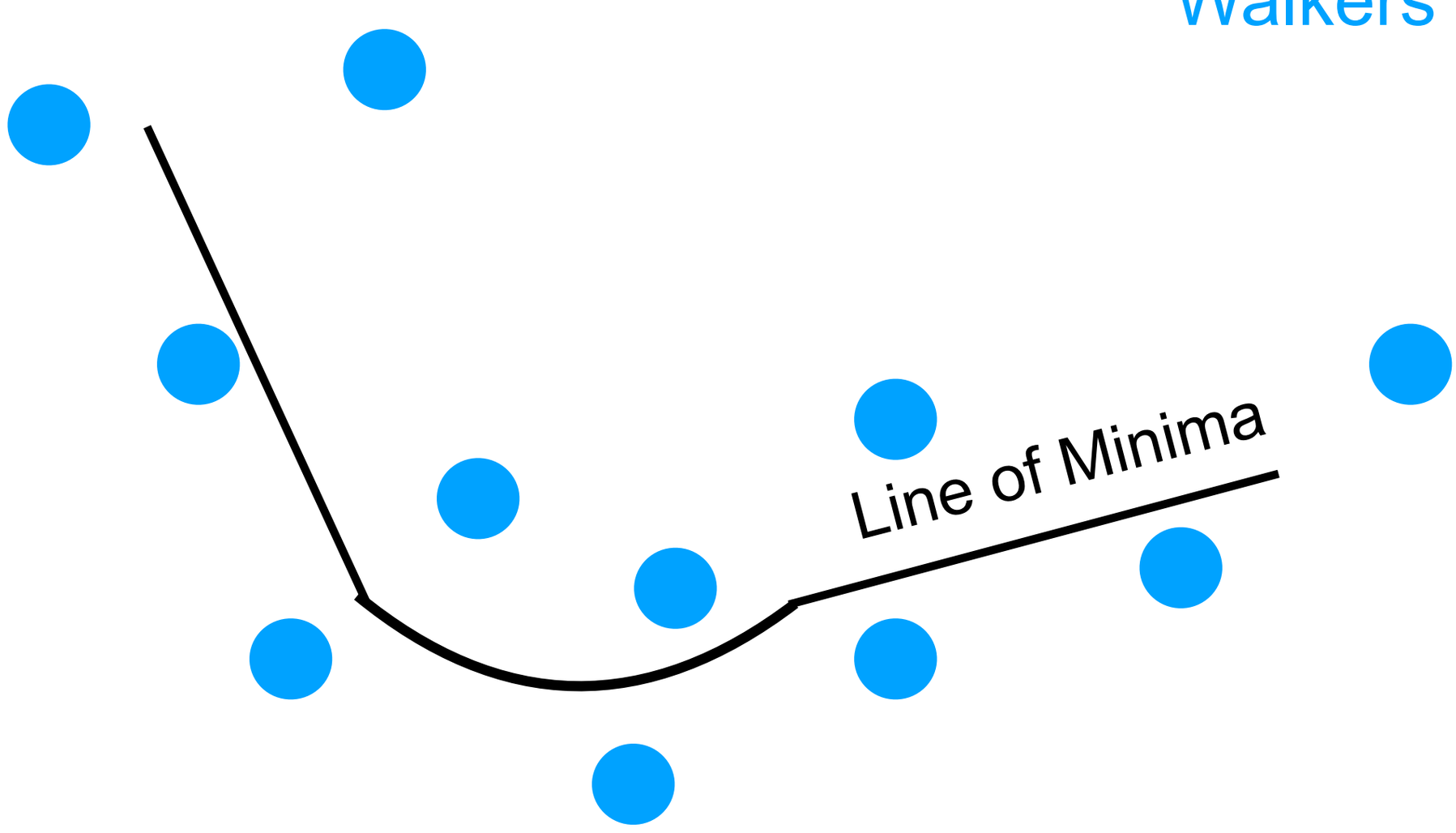
Speeding up MCMC

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



Speeding up MCMC

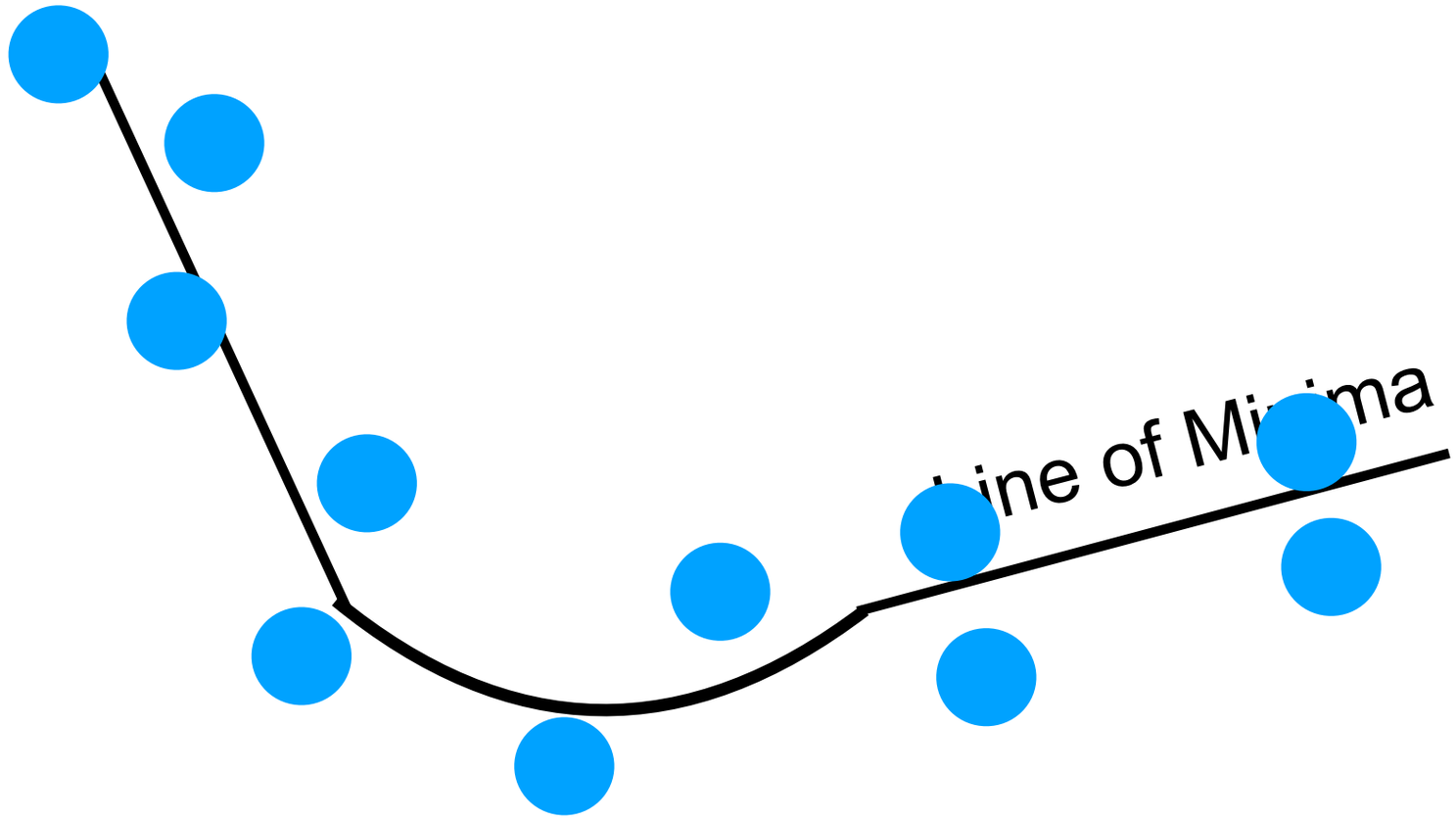
- We can consider having many walkers probe our space
  - Many walkers at the same time speed up convergence



Speeding up MCMC

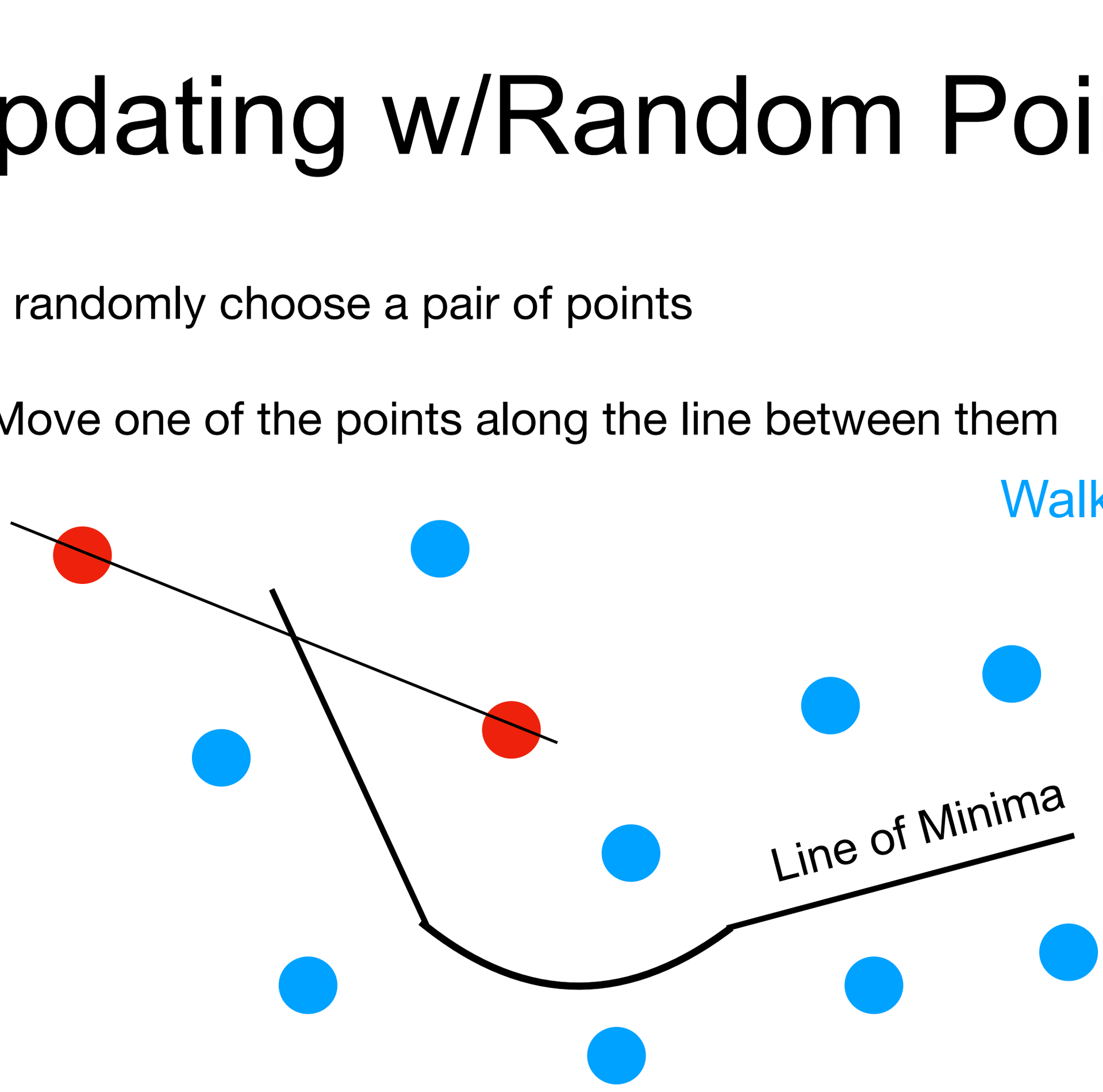
- We can consider having many walkers probe our space
  - Many walkers at the same time speed up convergence

Walkers



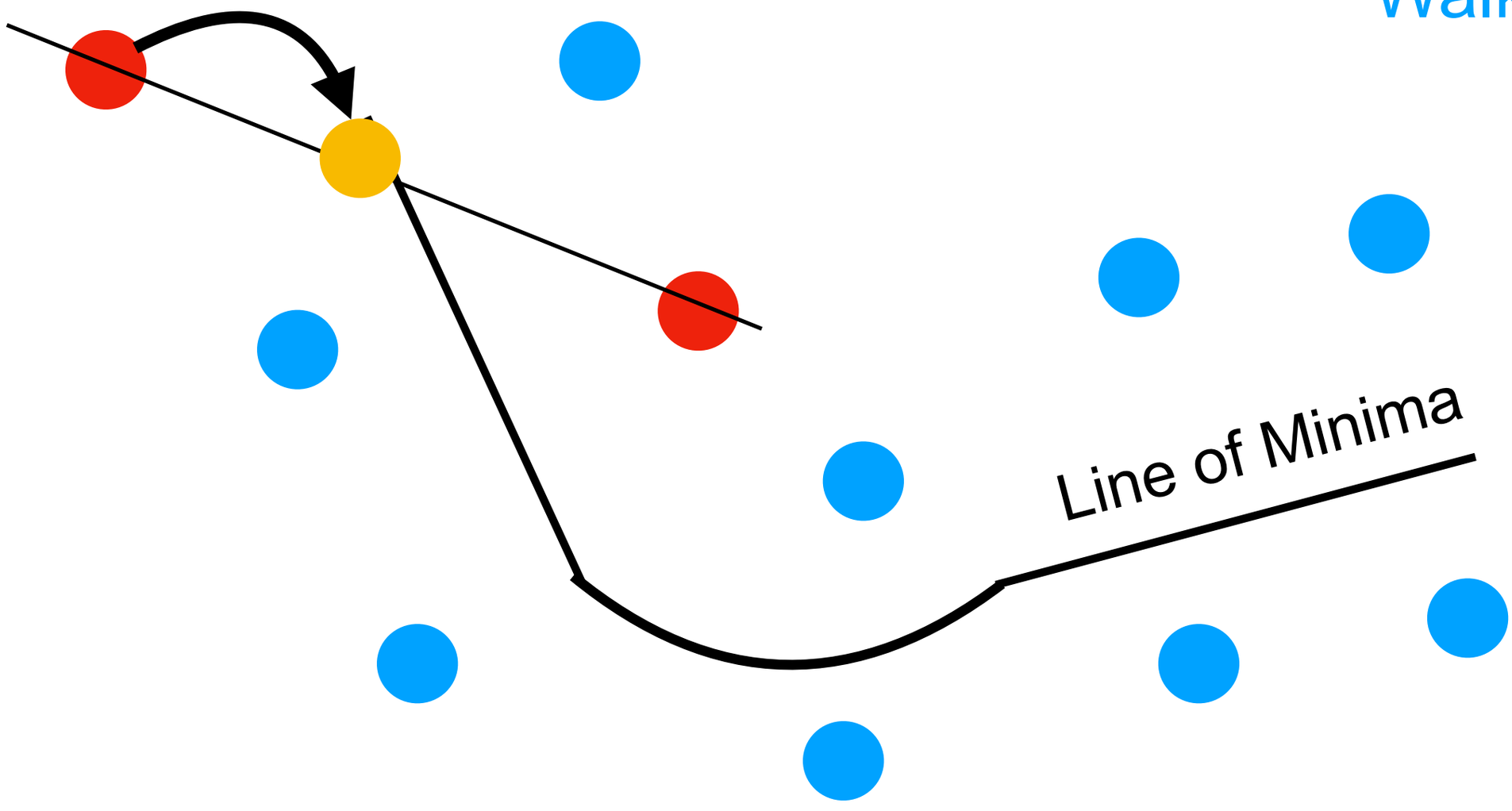
Updating w/Random Points

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



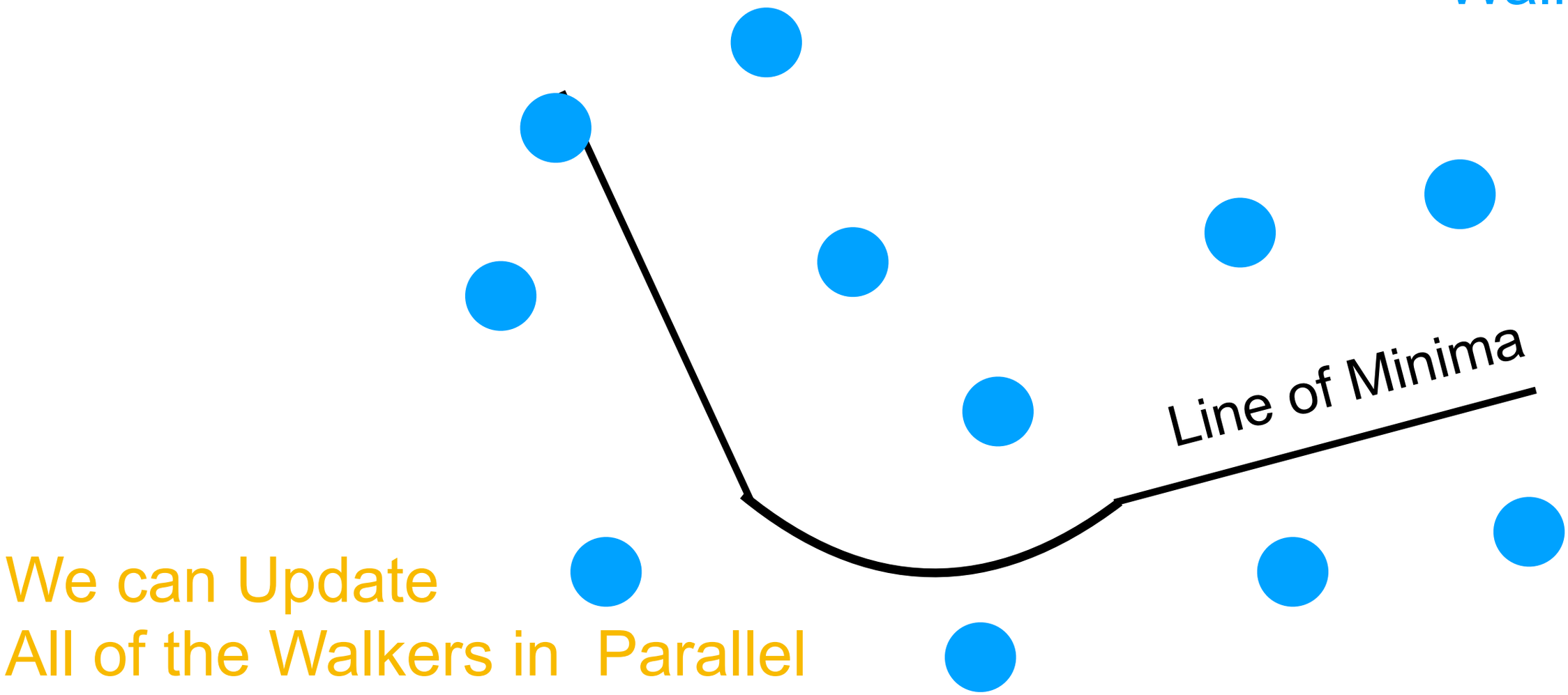
Updating w/Random Points

- 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



We can Update
All of the Walkers in Parallel

Quantum Monte Carlo

- Can use the same MCMC to populate a wave function
  - We can then scan paramaters to solve Shroedinger'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

$$w_{i+1} = \frac{p(\overrightarrow{r_{i+1}}\,|\,\vec{\theta})}{p(\overrightarrow{r_i}\,|\,\vec{\theta})}$$

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

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

Guess a Form for the wavefunction

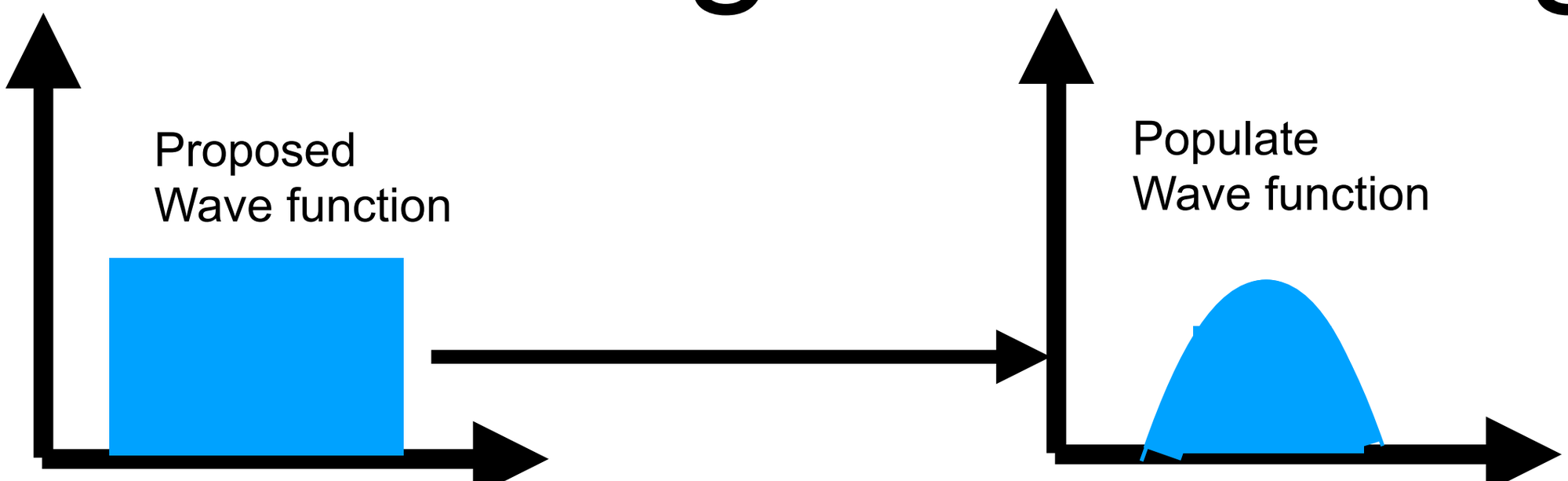
$$\sum_j p_j(\vec{r}\,|\,\vec{\theta}) = \frac{\psi_j^*(\vec{r}\,|\,\vec{\theta})\psi_j(\vec{r}\,|\,\vec{\theta})}{\langle\psi\,|\,\psi\rangle}$$

We can define probability from wavefunction

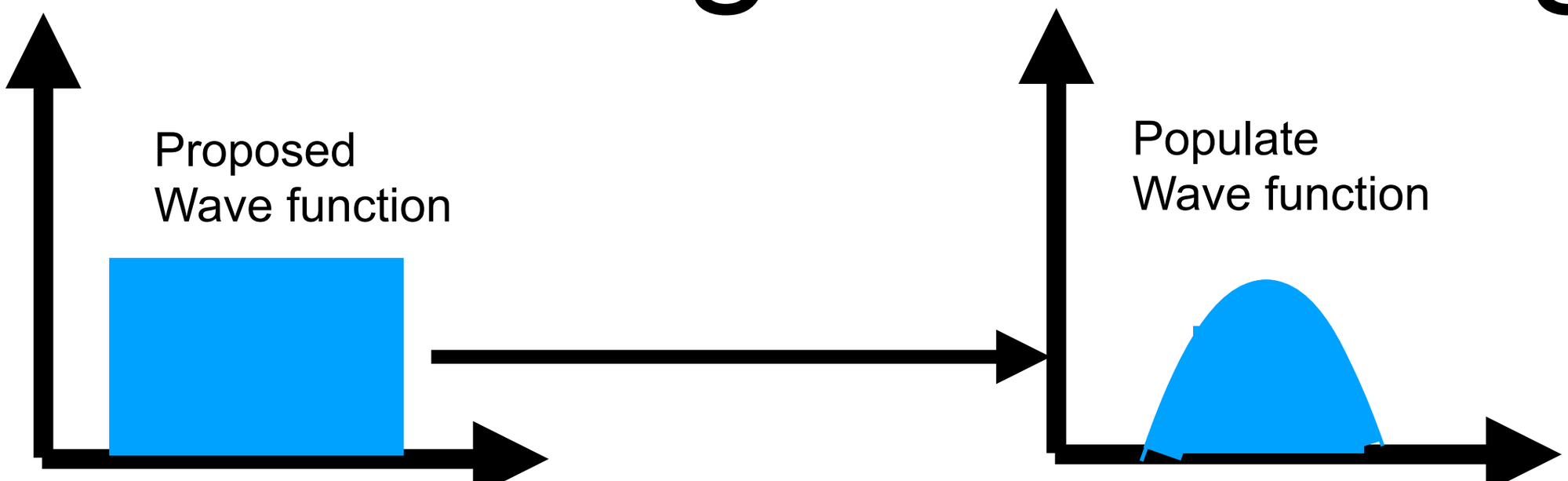
$$\sum_j w_{i+1}^j = \frac{p_j(\overrightarrow{r_{i+1}}\,|\,\vec{\theta})}{p_j(\overrightarrow{r_i}\,|\,\vec{\theta})}$$

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

Solving Schroedinger



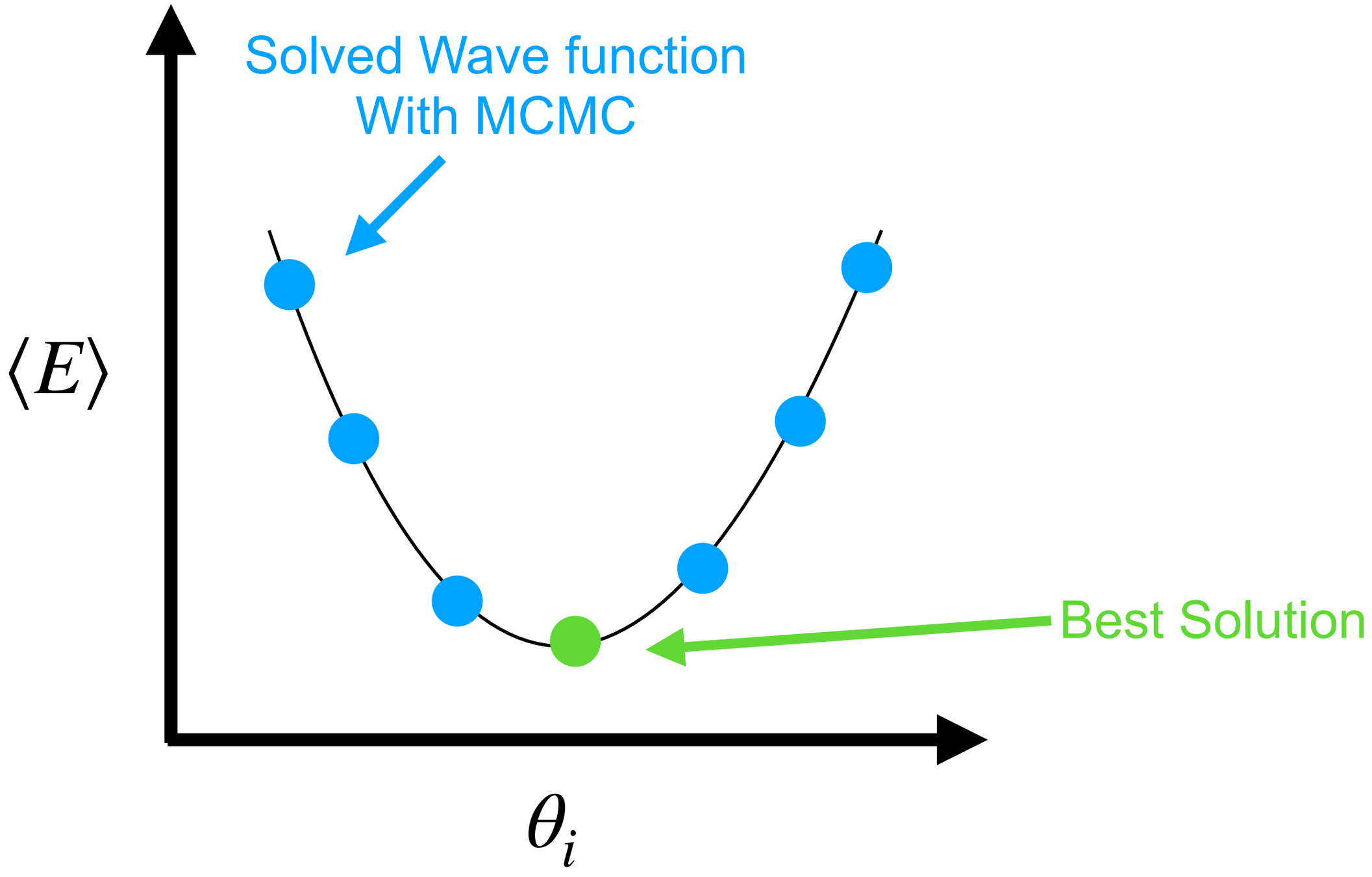
Solving Schroedinger



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

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

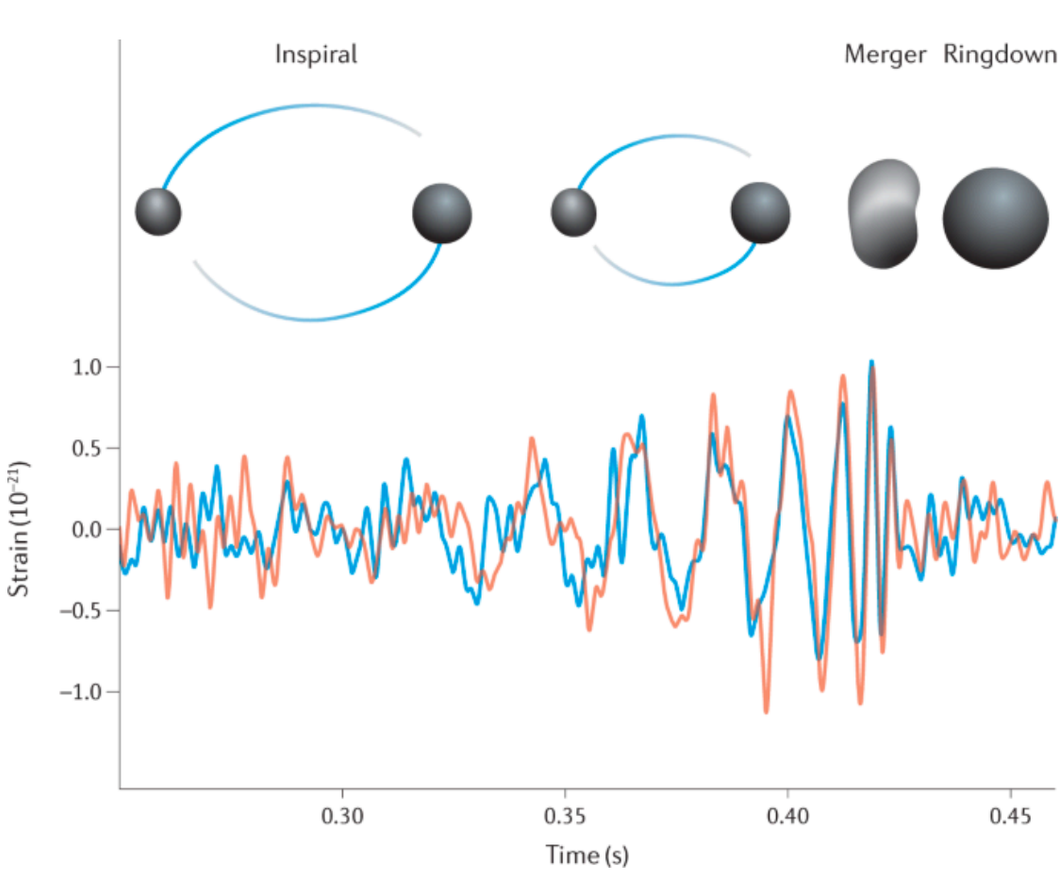
Solving Schroedinger



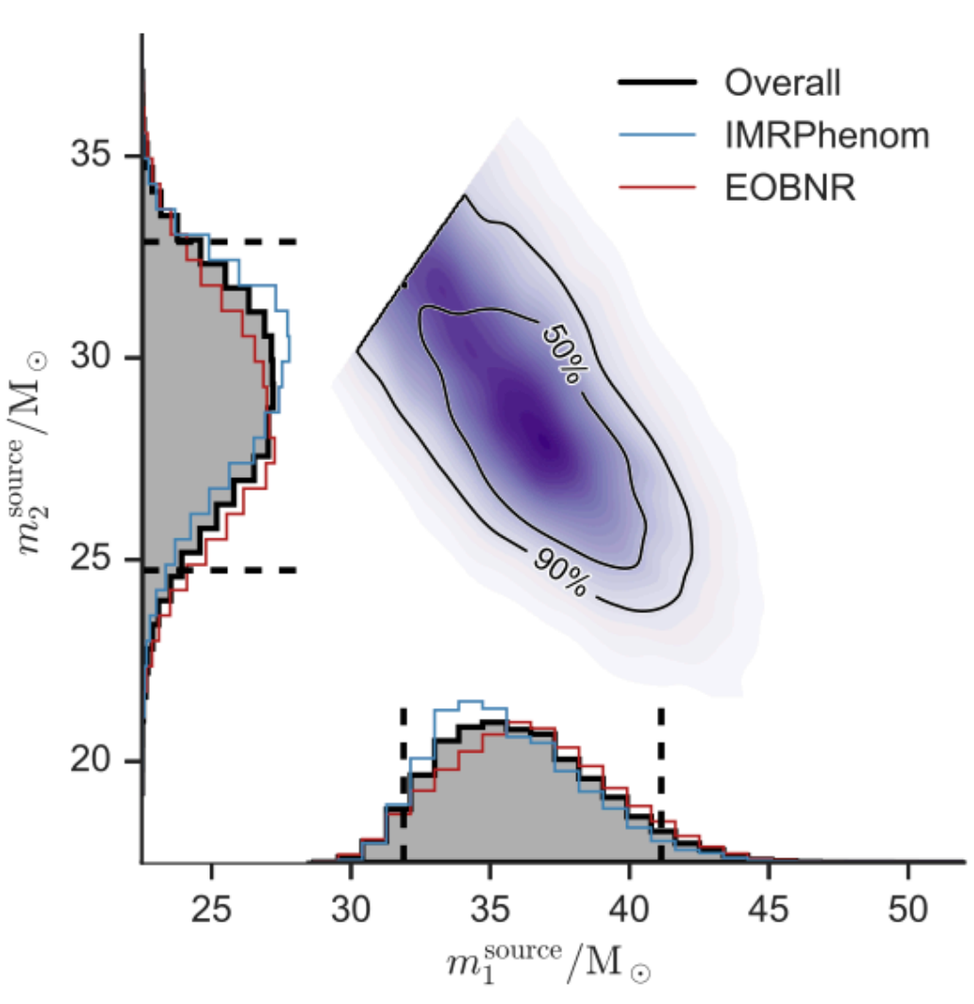
Our goal is to minimize the Energy given a wave functional form

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

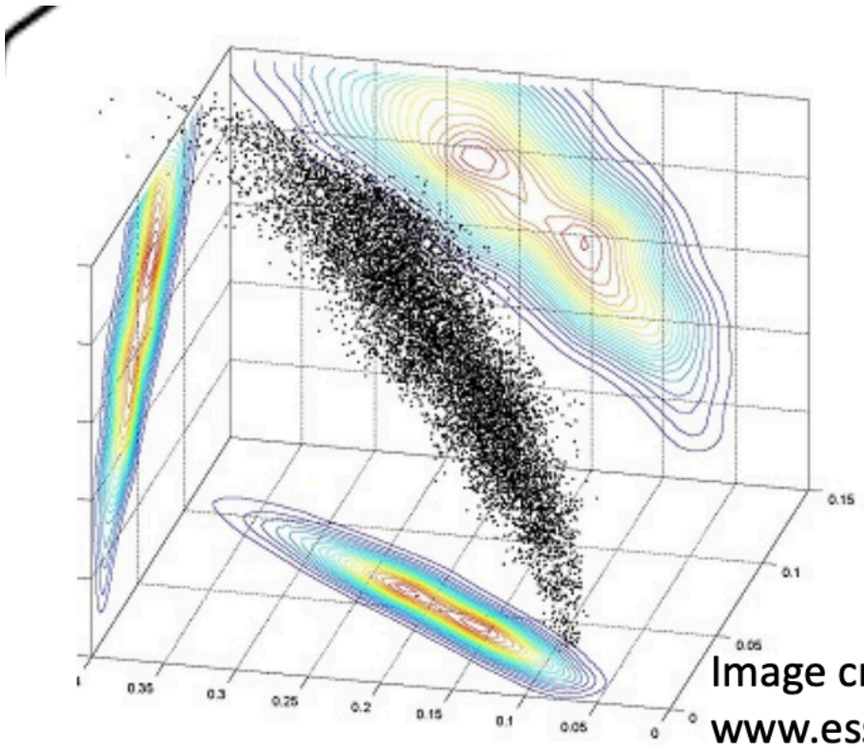


Image credit:
www.essenceps.com

- 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