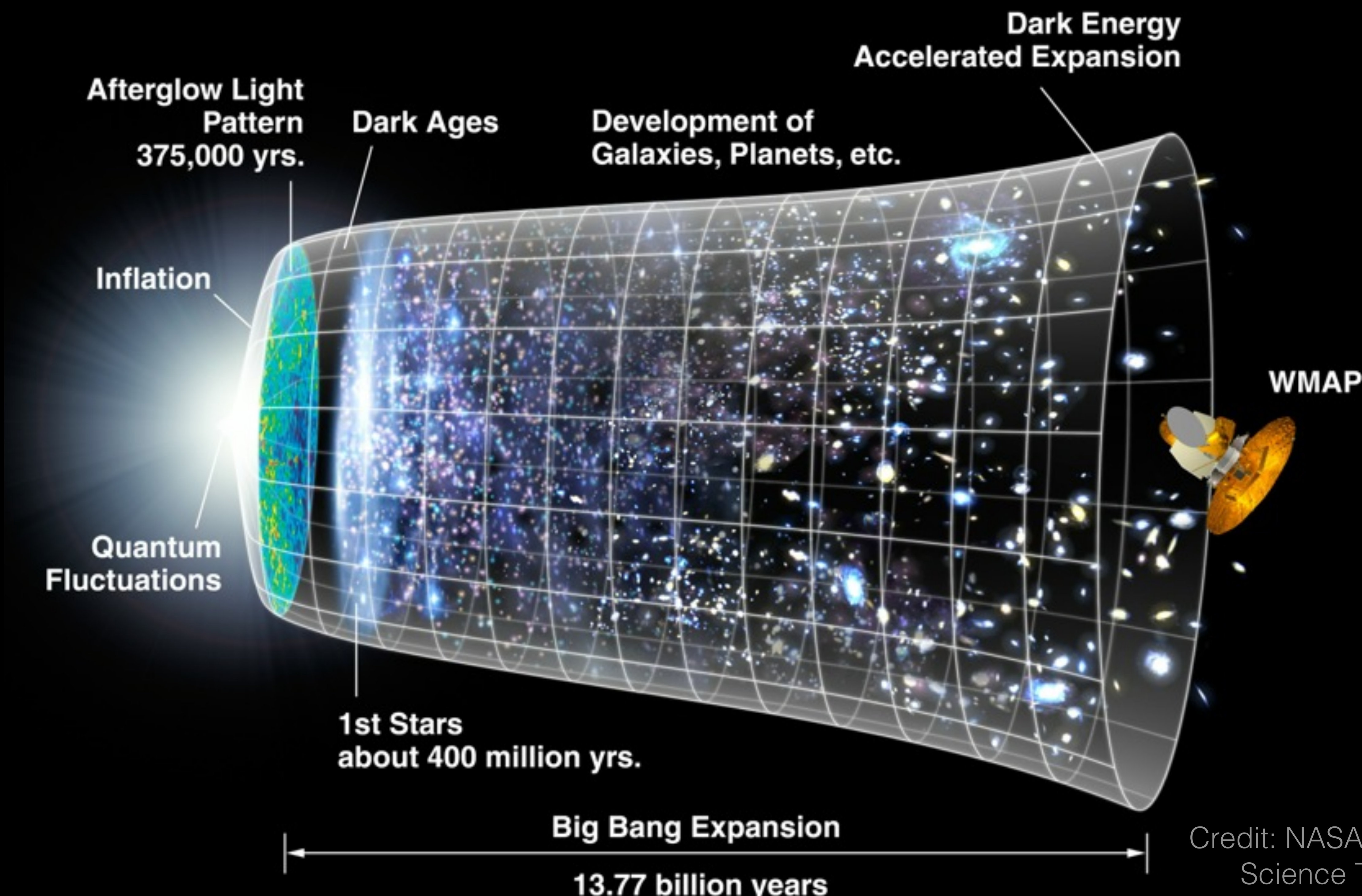
Computational Challenges in Cosmological Parameter Estimation

Sebastian Seehars and Joël Akeret

Joint work with Adam Amara, Alexandre Refregier, and André Csillaghy.
(In collaboration with University of Applied Sciences Northwestern Switzerland)

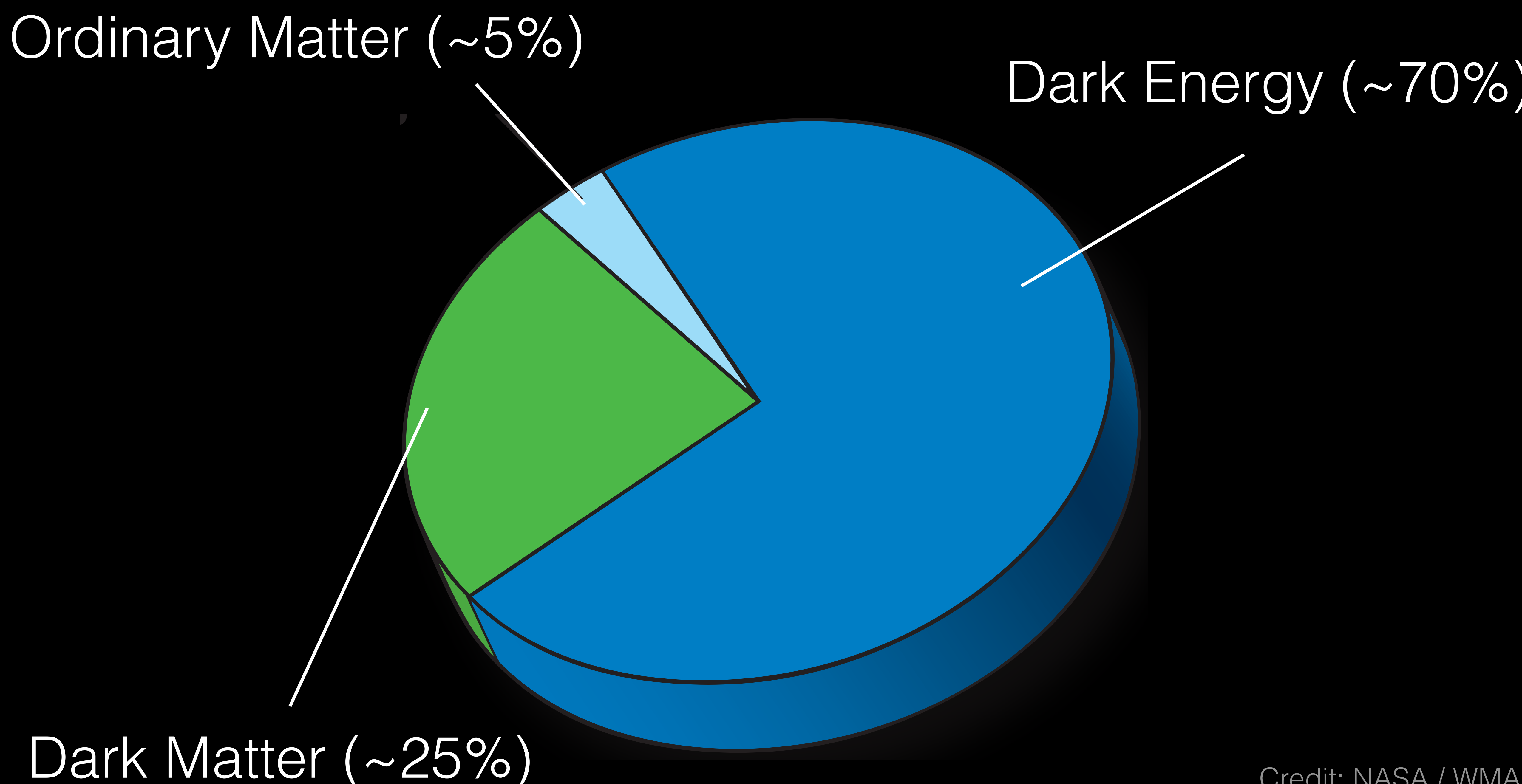
21.02.2014

Cosmology



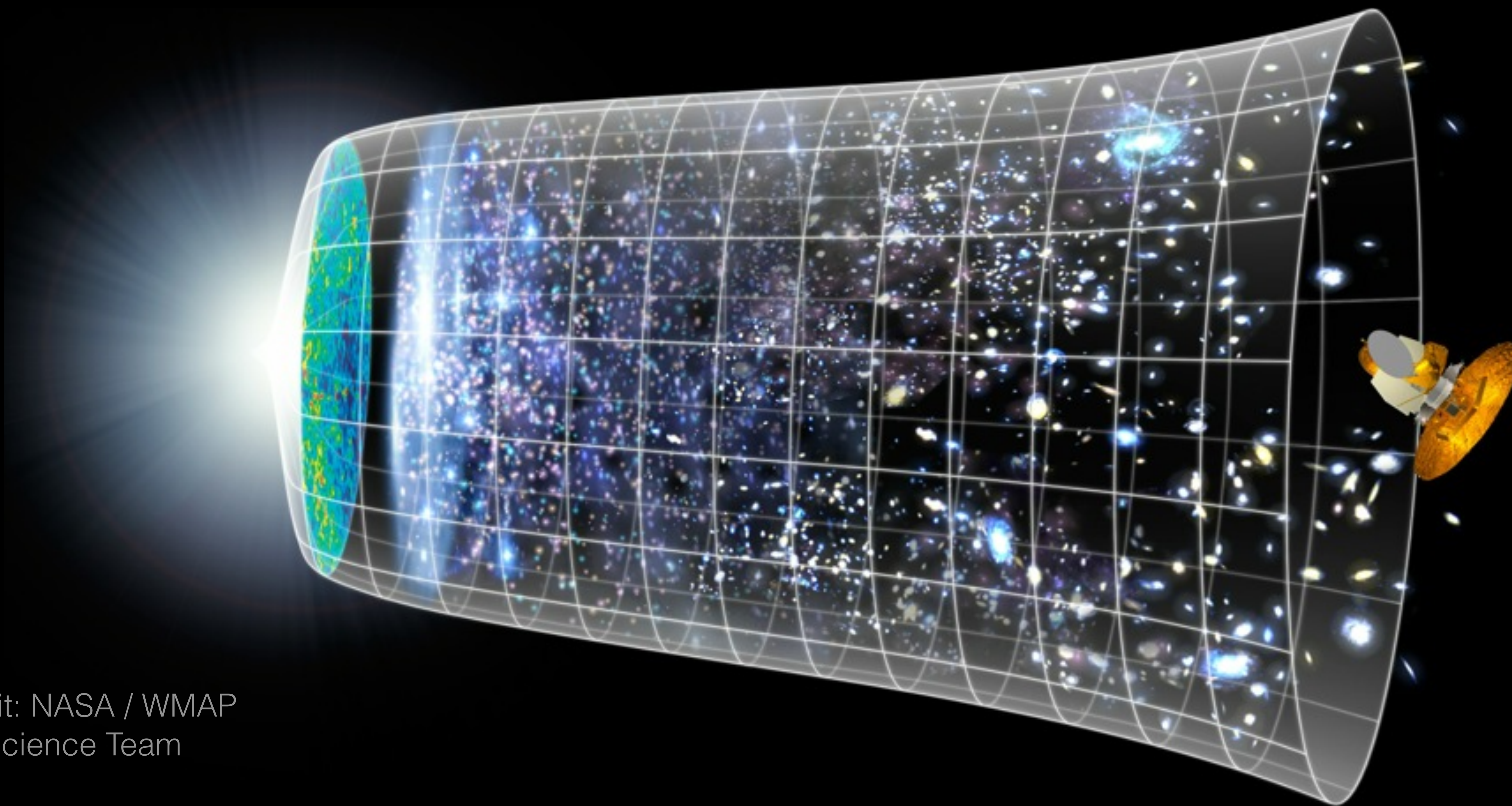
Credit: NASA / WMAP
Science Team

Dark Components



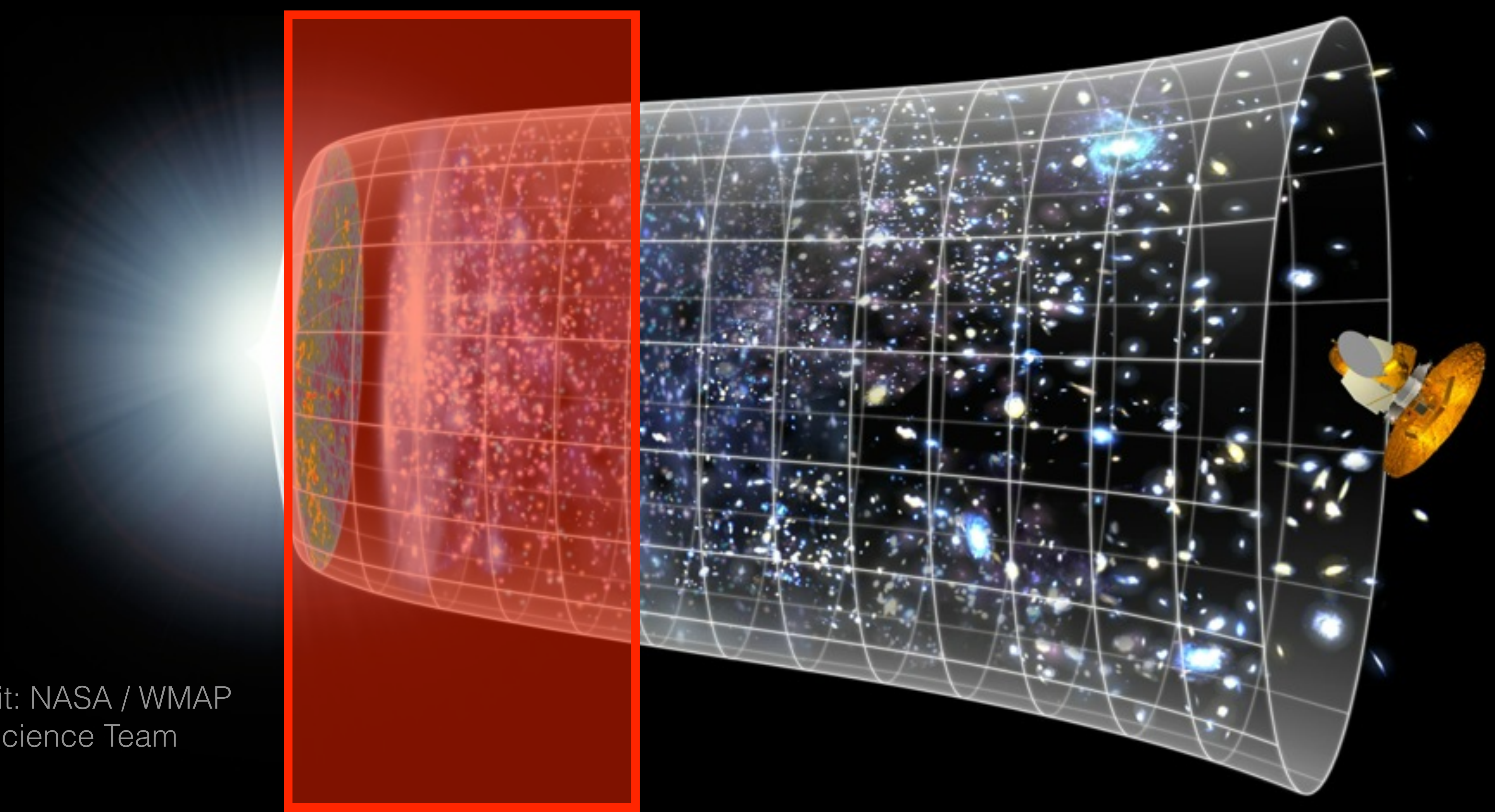
Credit: NASA / WMAP Science Team

Computational Cosmology



Credit: NASA / WMAP
Science Team

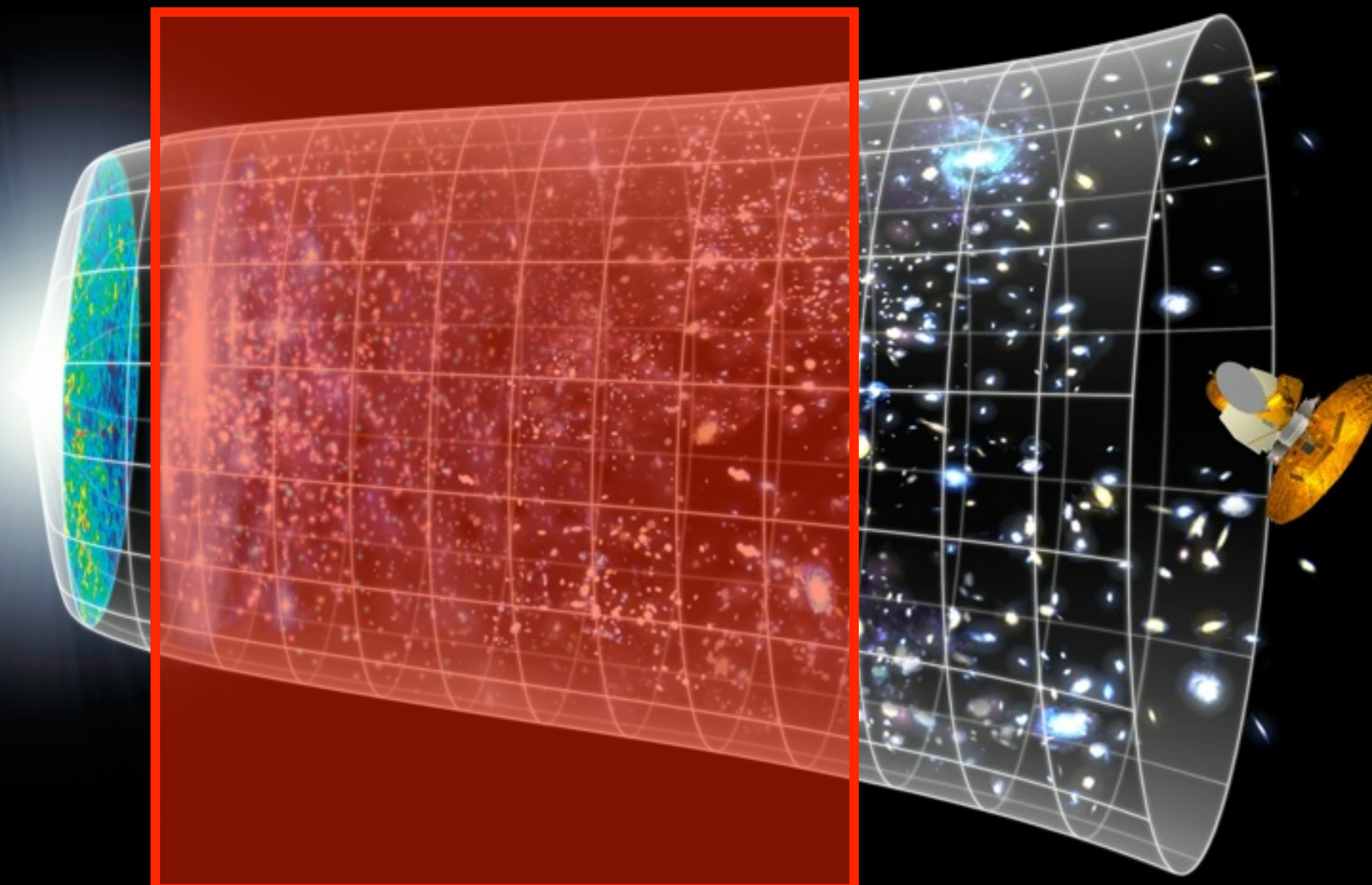
Computational Cosmology



Credit: NASA / WMAP
Science Team

Linear perturbation theory:
Solving Boltzmann equations numerically

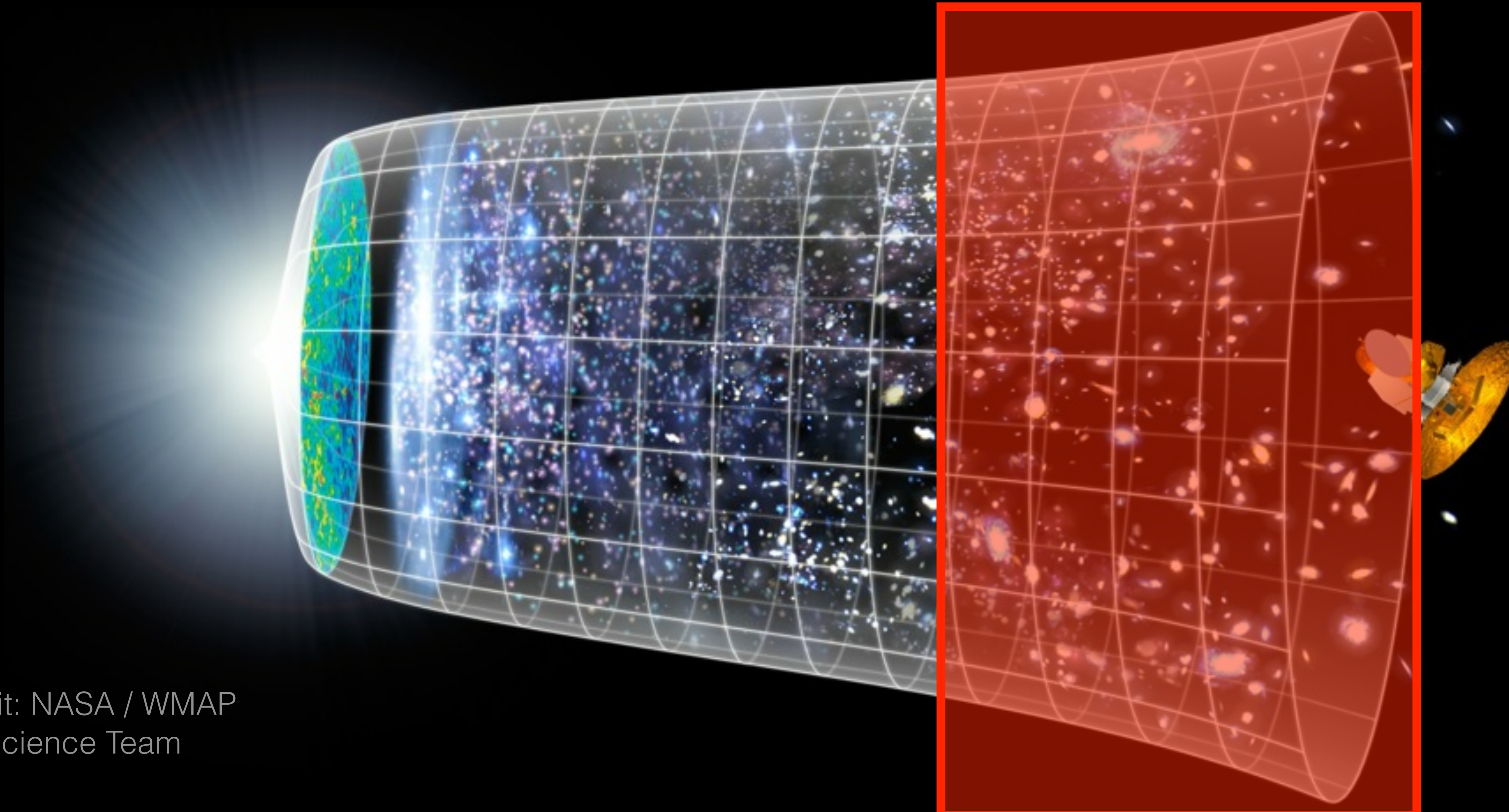
Computational Cosmology



Credit: NASA / WMAP Science Team

Large scale structure formation:
N-body simulations of dark matter dynamics

Computational Cosmology



Credit: NASA / WMAP Science Team

Galaxy formation:
Hydrodynamic simulations

Data Analysis

- Processing and analysis of large datasets
- Forward modelling
    - Astronomical foregrounds
    - Instrument effects
- Statistical inference

Incomplete History of Cosmic Microwave Background (CMB) Surveys

COBE
(1989)

→

BOOMERanG
(1997)

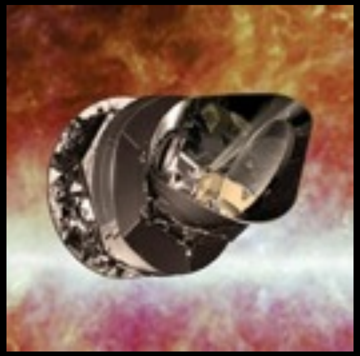
→

WMAP
(2001)

→

SPT
(2006)

→



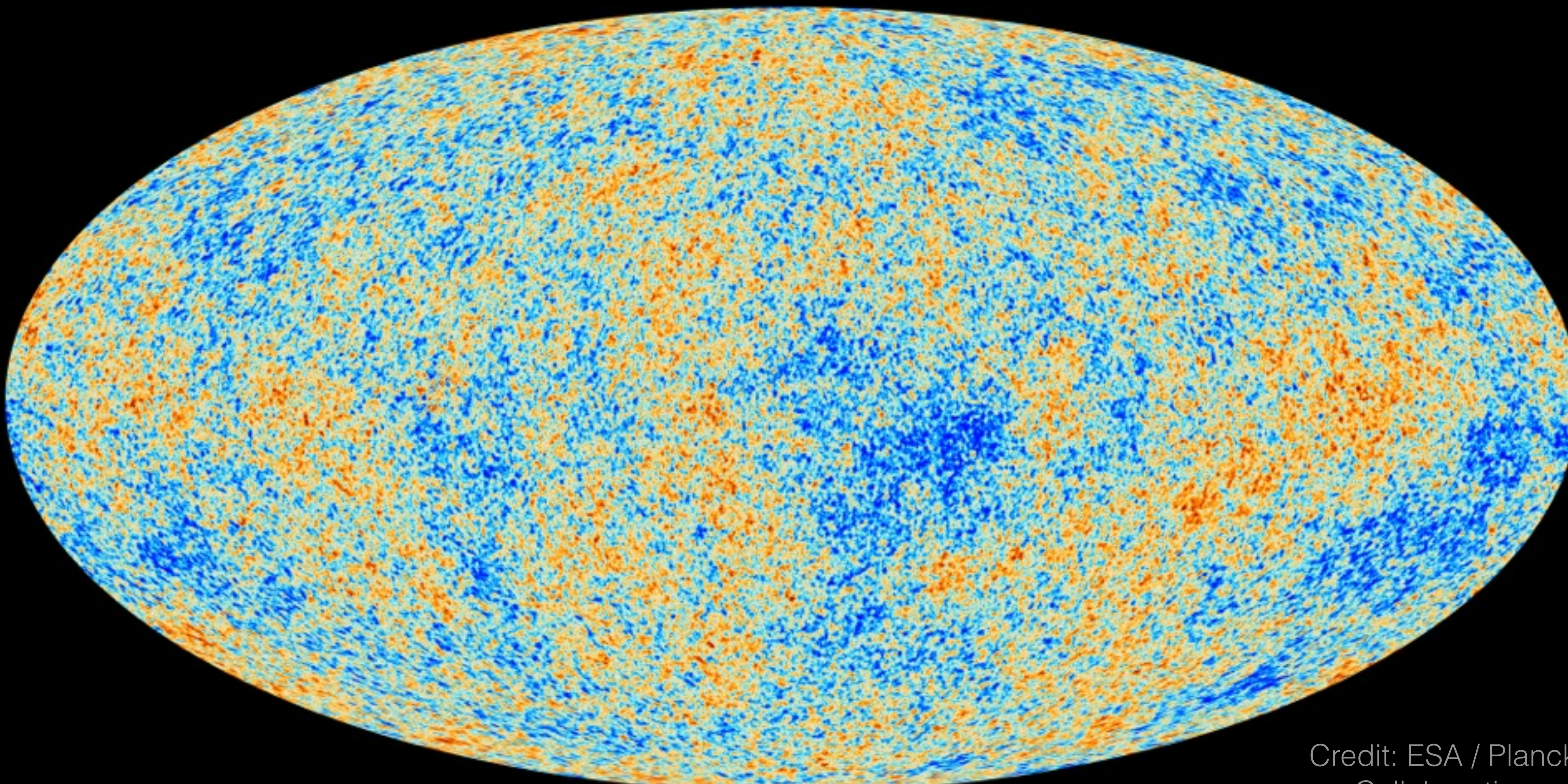
Planck
(2009)

→

PolarBear
Spider
…

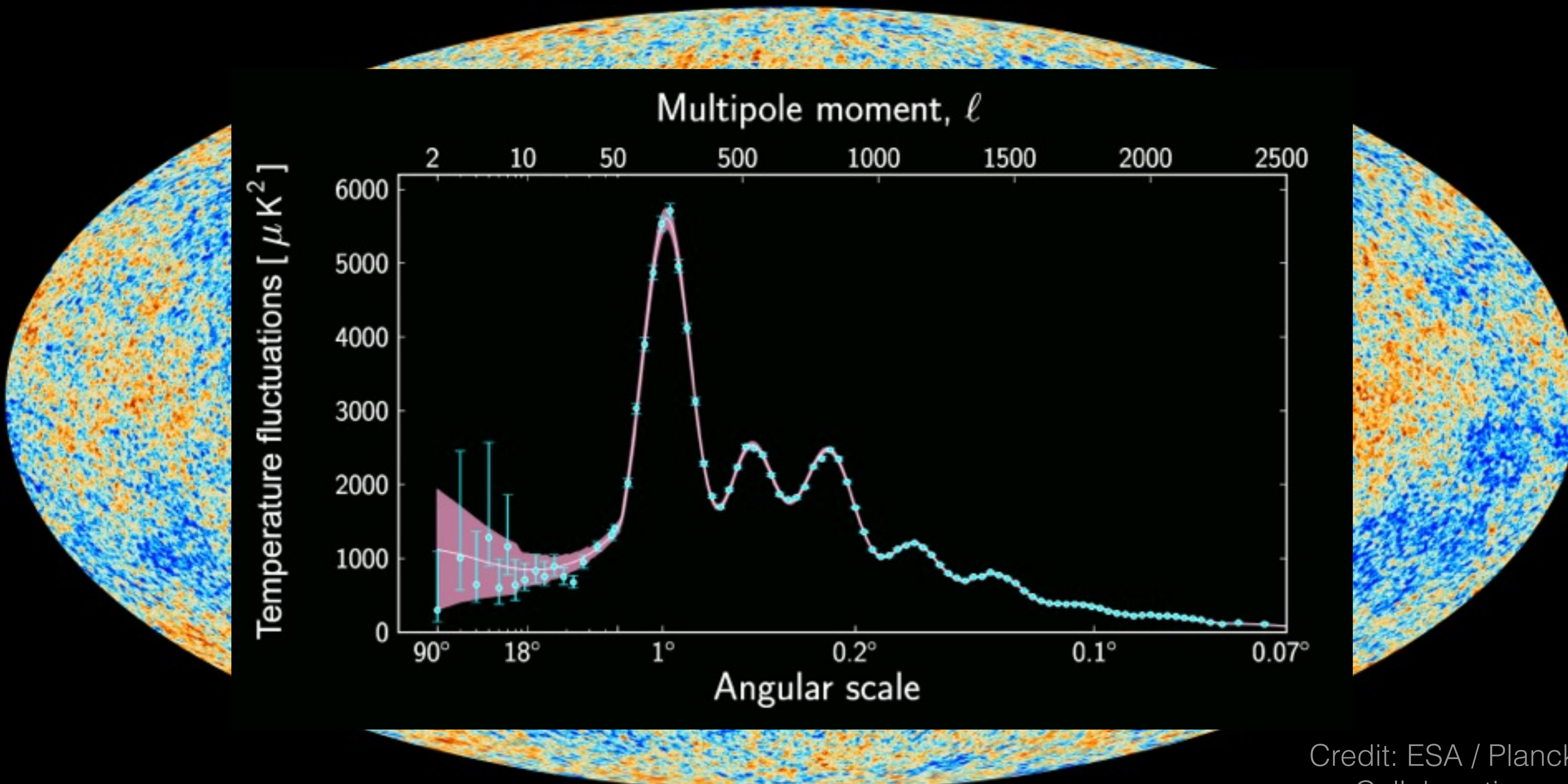
Future

Map of the CMB



Credit: ESA / Planck Collaboration

Map of the CMB



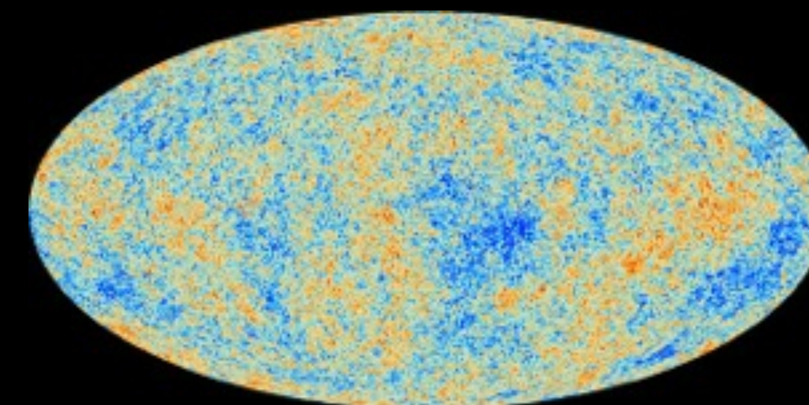
Credit: ESA / Planck Collaboration

Parameter Estimation

A_s

$\Omega_b h^2$ H_0 $\Omega_c h^2$

n_s τ



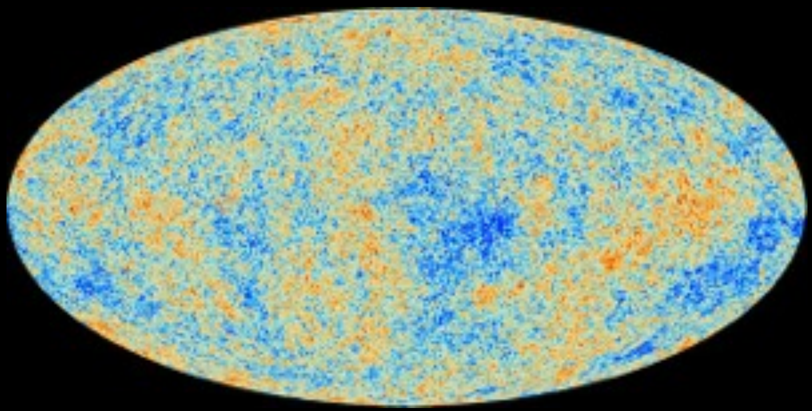
Credit: ESA / Planck Collaboration

Parameter Estimation

A_s

$\Omega_b h^2$ H_0 $\Omega_c h^2$

n_s τ



↓ Predictions from Theory

↓ Data and Errors

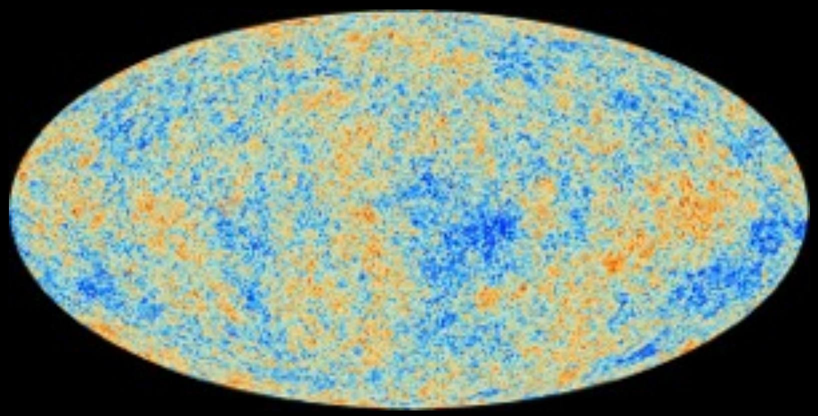
Credit: ESA / Planck Collaboration

Parameter Estimation

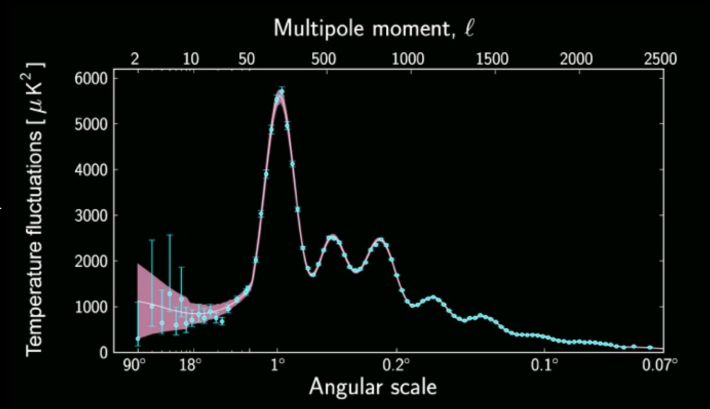
A_s

$\Omega_b h^2$ H_0 $\Omega_c h^2$

n_s τ



Predictions from Theory



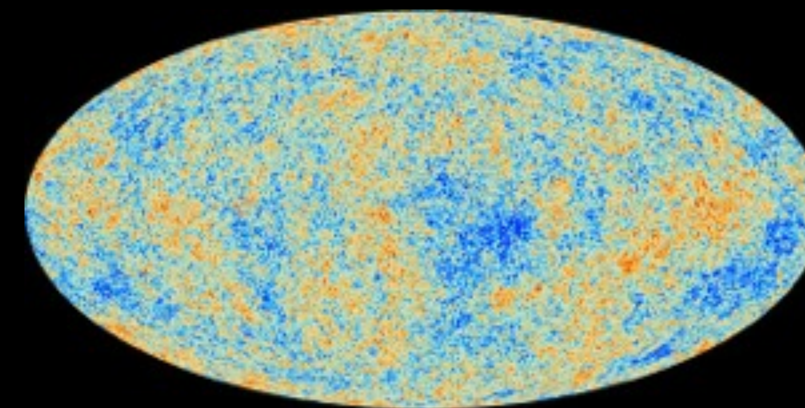
Data and Errors

Credit: ESA / Planck Collaboration

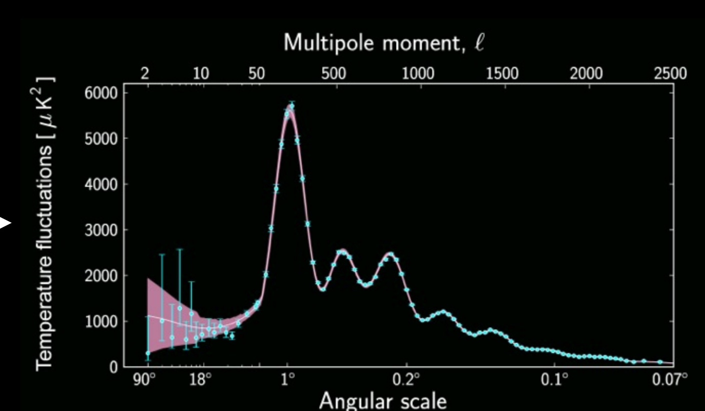
Parameter Estimation

$\Omega_b h^2$ A_s H_0 $\Omega_c h^2$ n_s τ

Parameter Estimation



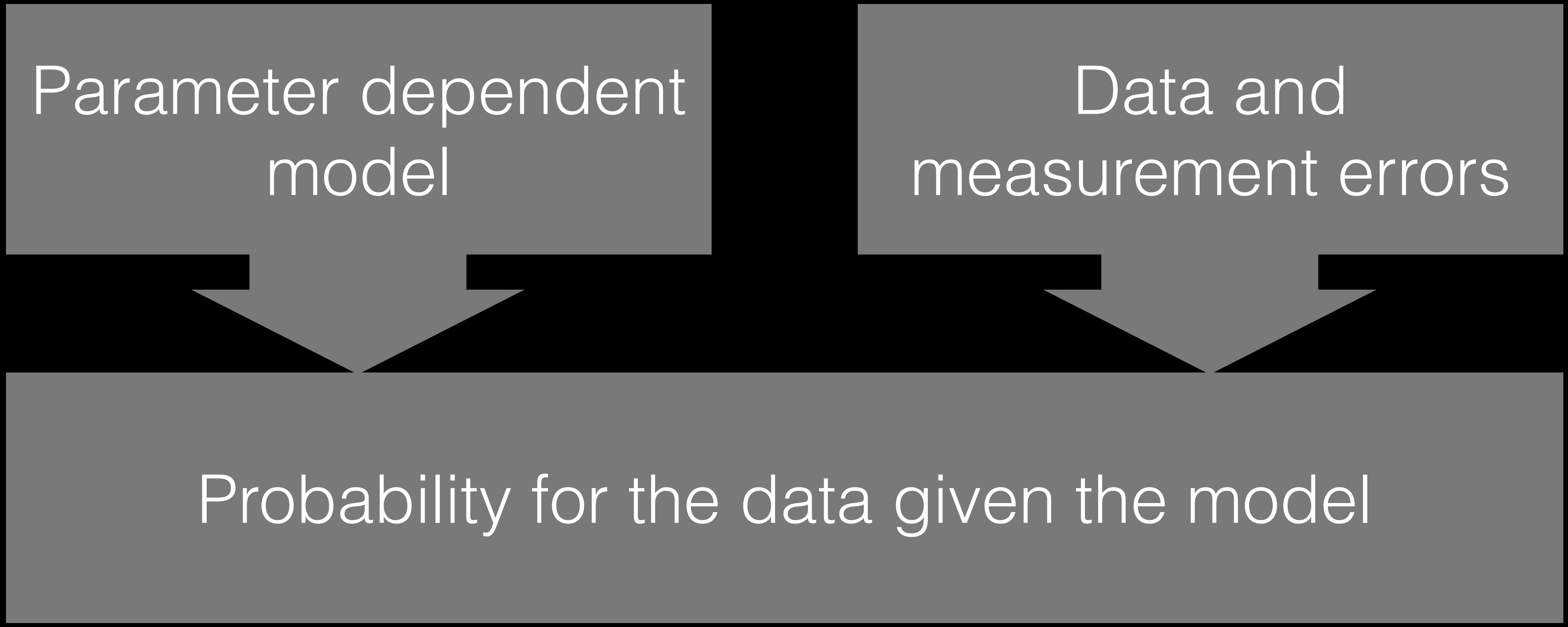
Predictions from Theory



Data and Errors

Credit: ESA / Planck Collaboration

Likelihood


Parameter dependent model
Data and measurement errors
Probability for the data given the model

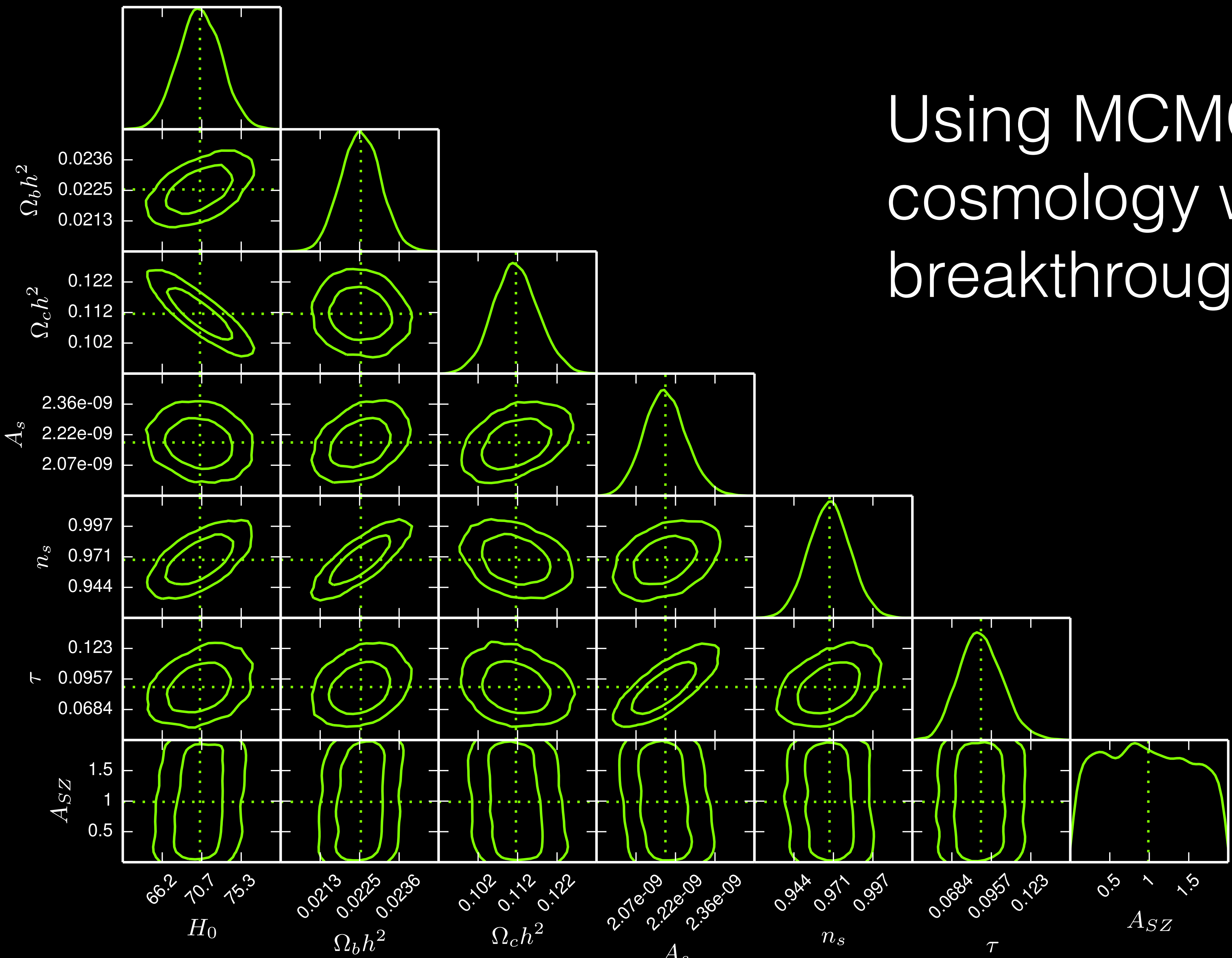
Bayesian Inference

$$p_{new}(\Theta) = \frac{\overbrace{p(D_{new}|\Theta)}^{\text{Likelihood}}\,\overbrace{p(\Theta)}^{\text{Prior}}}{\int d\Theta\, p(D_{new}|\Theta)p(\Theta)}$$

Bayesian Inference in Cosmology

- Parameter space is greater or equal to six
- Model predictions have to be evaluated numerically
- The likelihood is computationally intensive

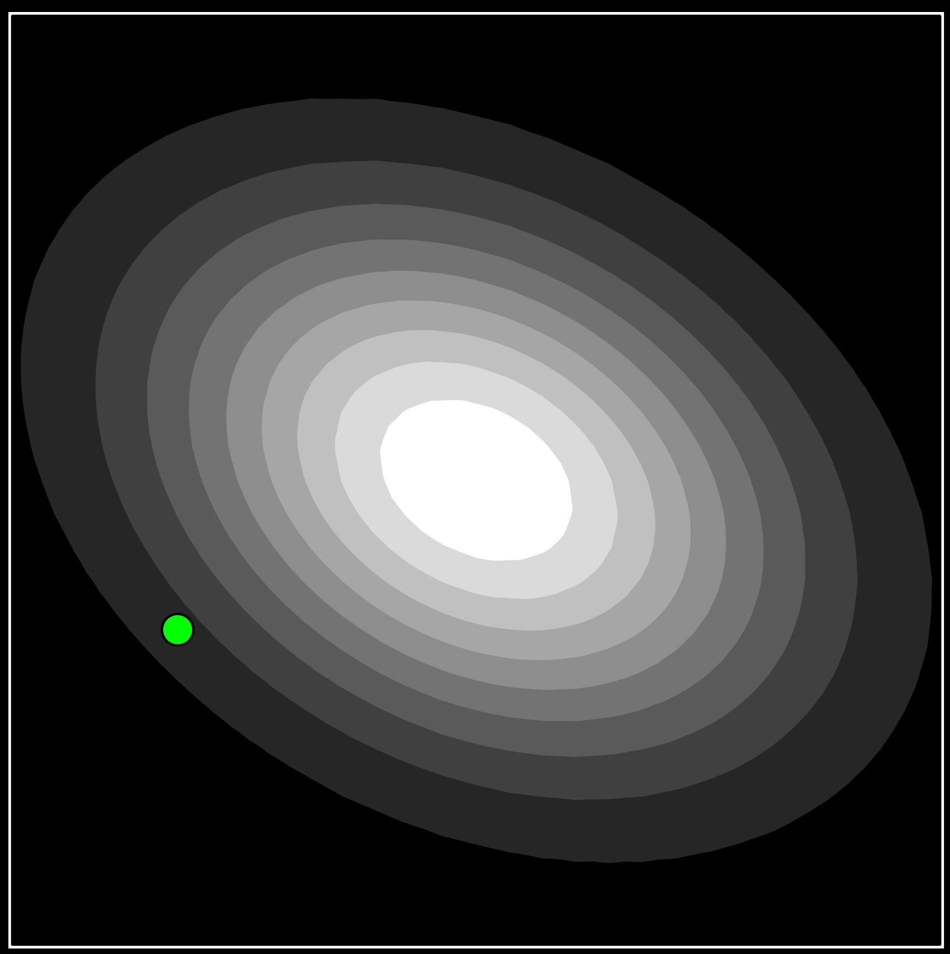
MCMC in Cosmology



Using MCMC in cosmology was a major breakthrough in 2001

Metropolis-Hasting

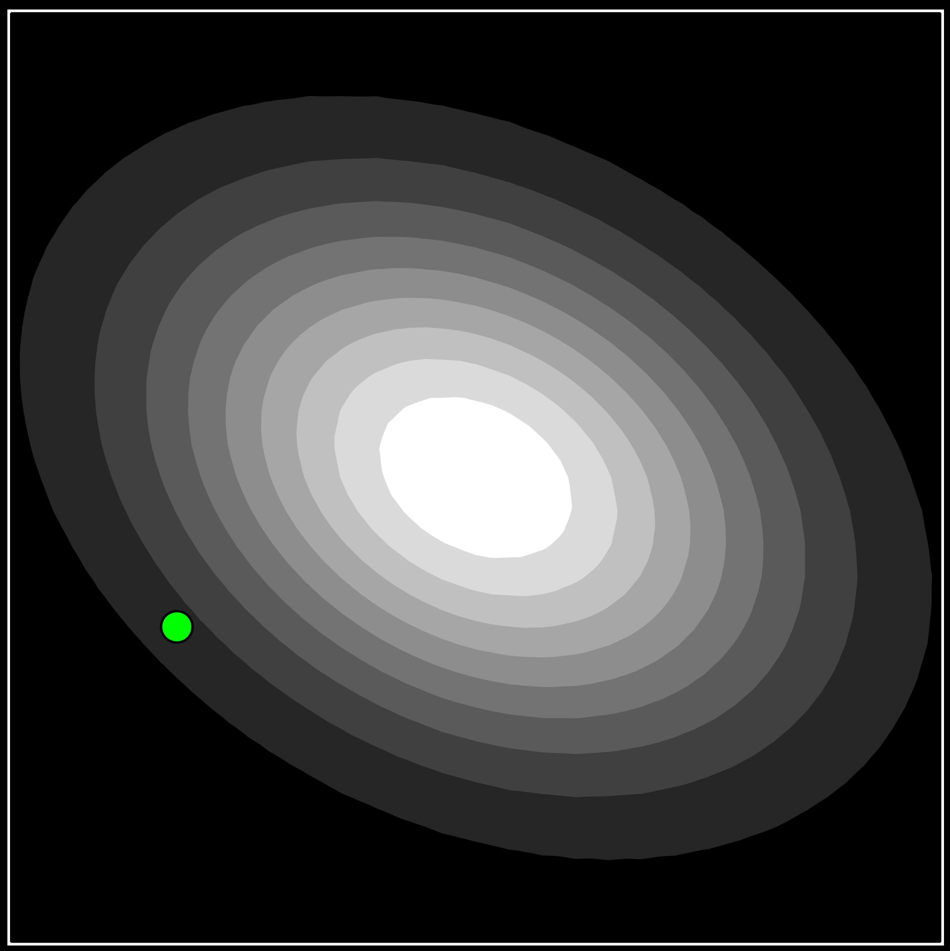
Metropolis+ 1953
Hastings 1970



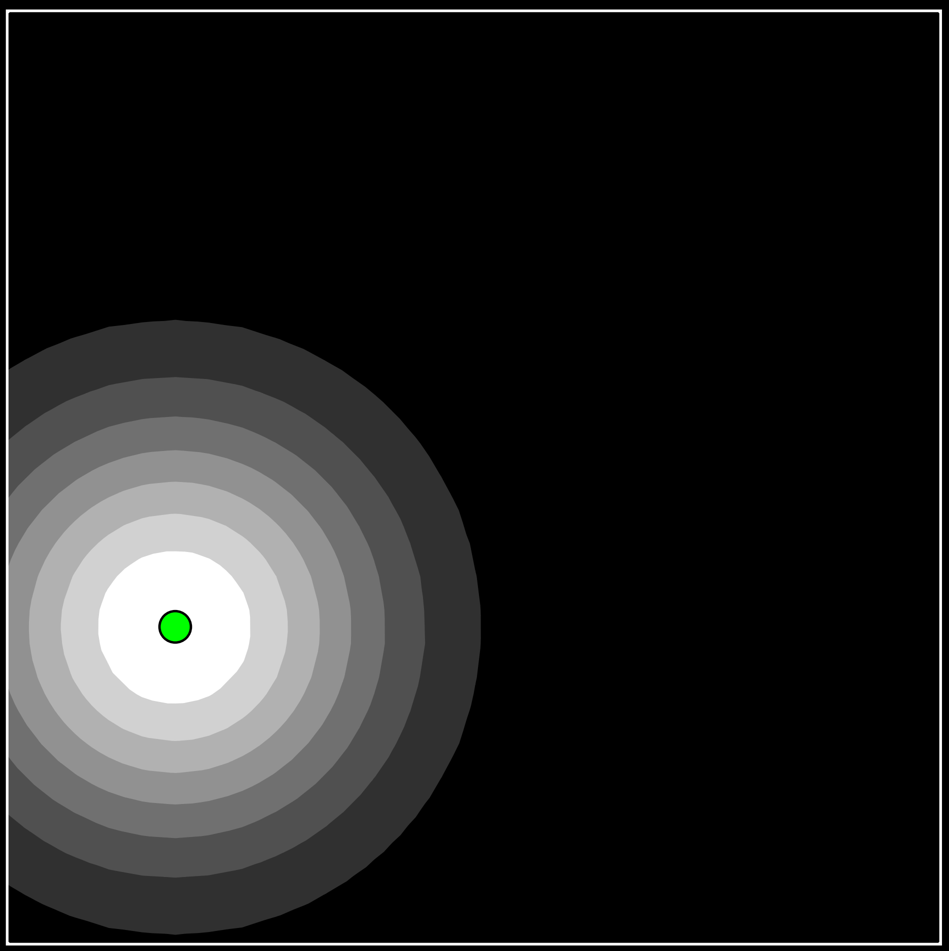
target distribution $P(\theta)$

Metropolis-Hasting

Metropolis+ 1953
Hastings 1970



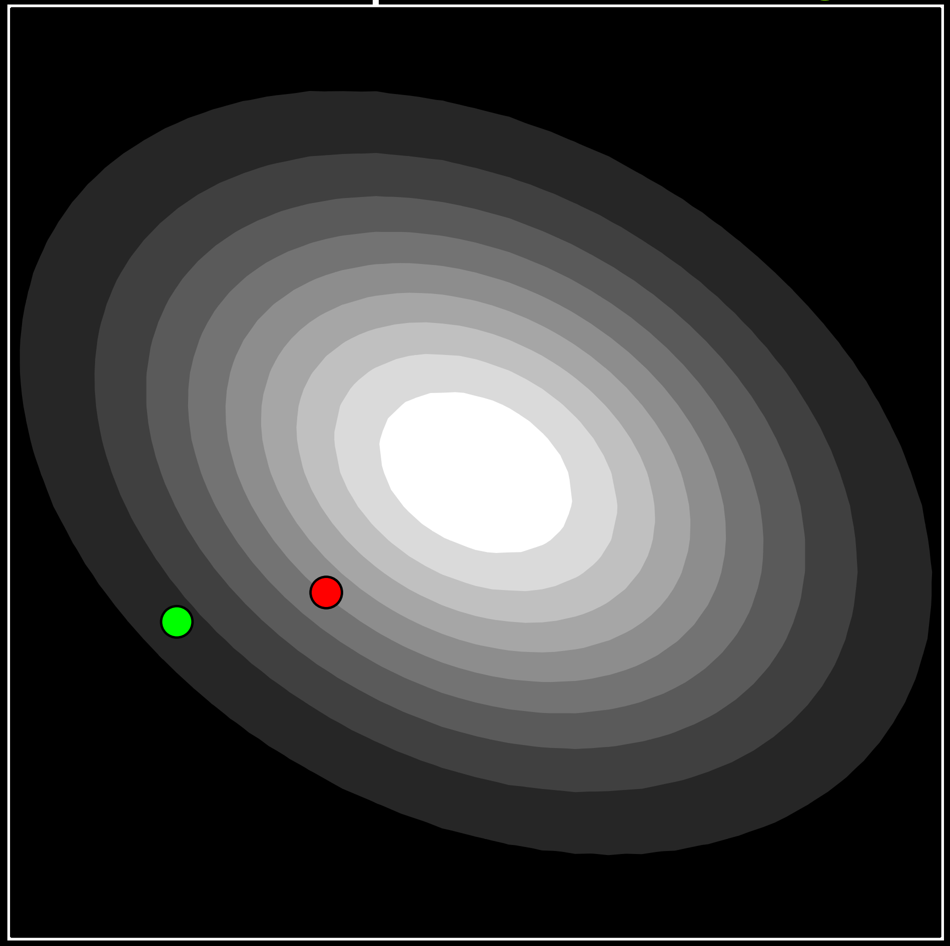
target distribution P(θ)



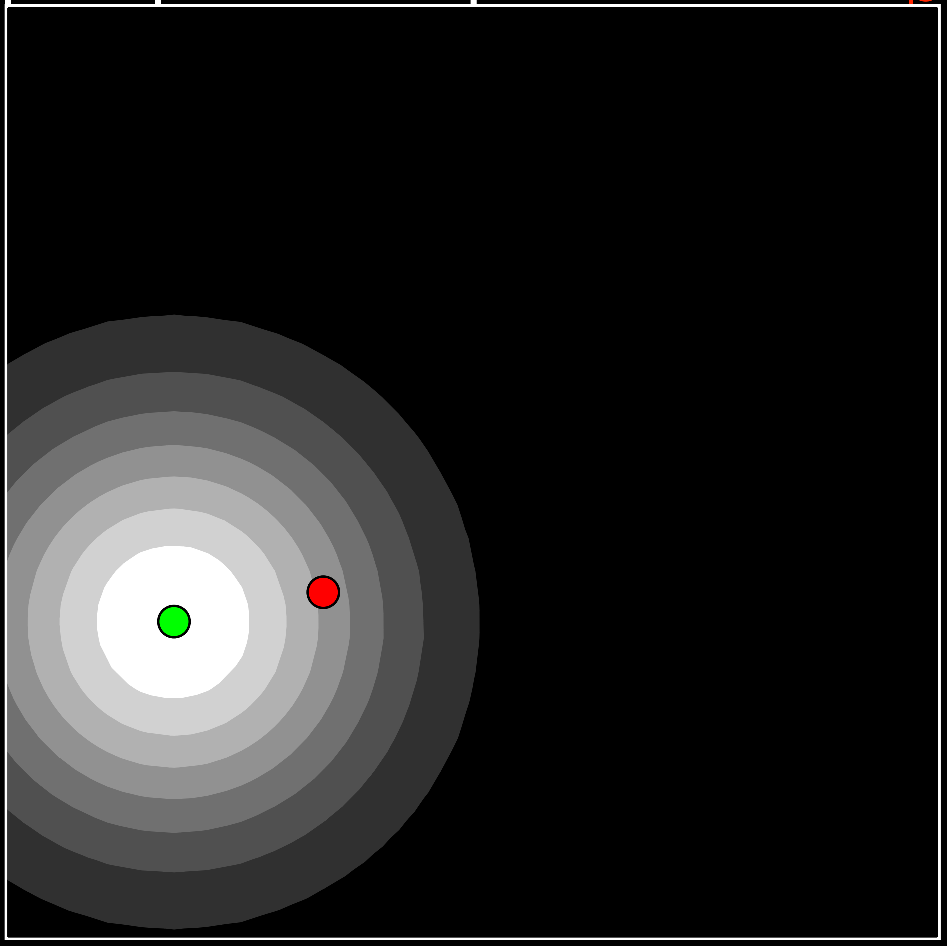
proposal distribution Q(θ)

Metropolis-Hasting

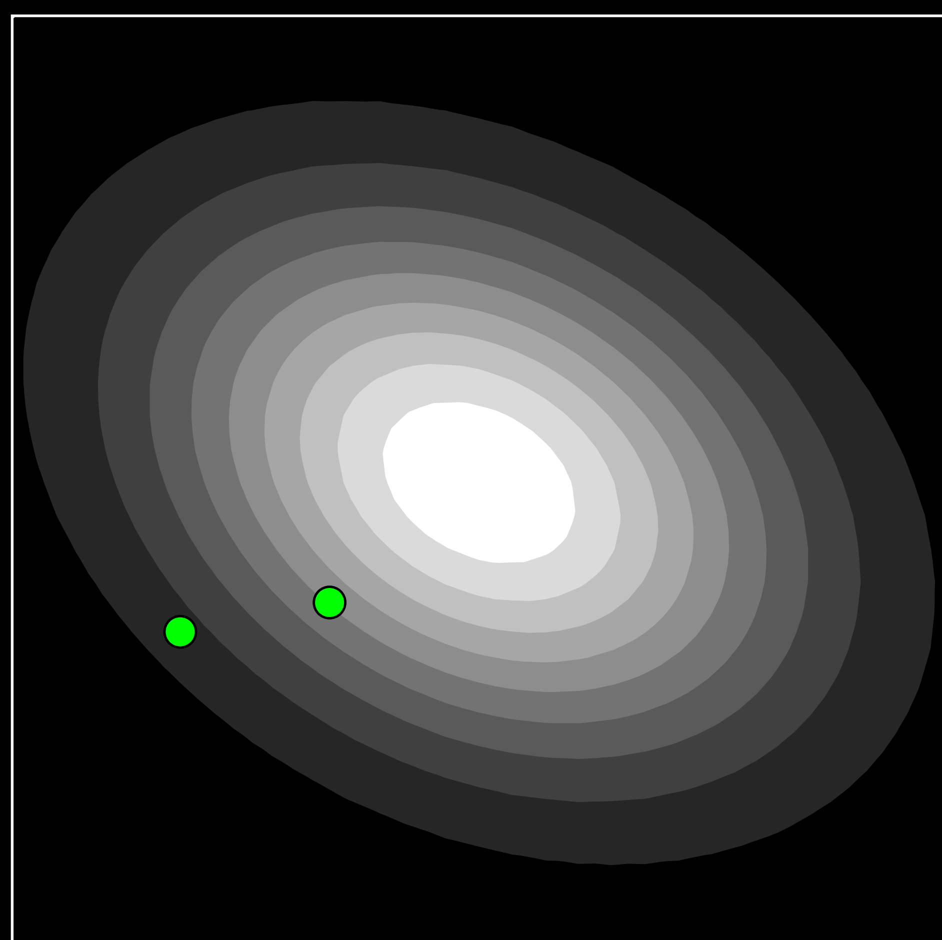
initial position θ_0



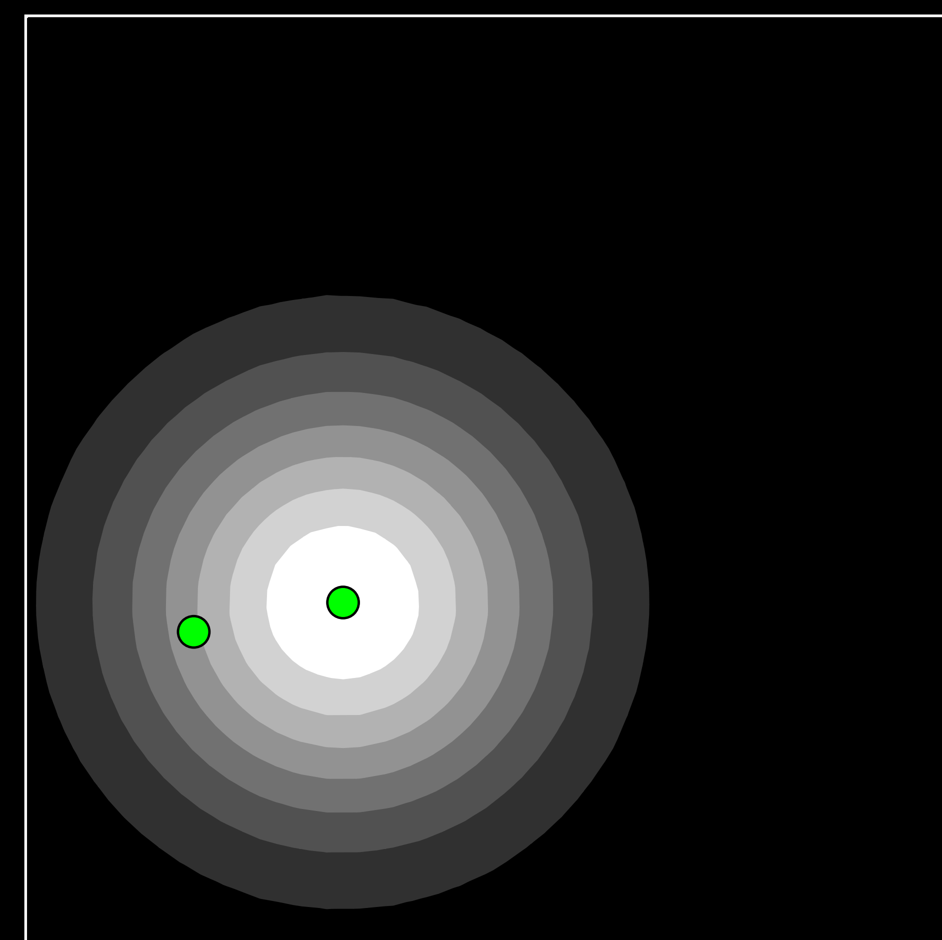
proposed position θ_p



Metropolis-Hasting

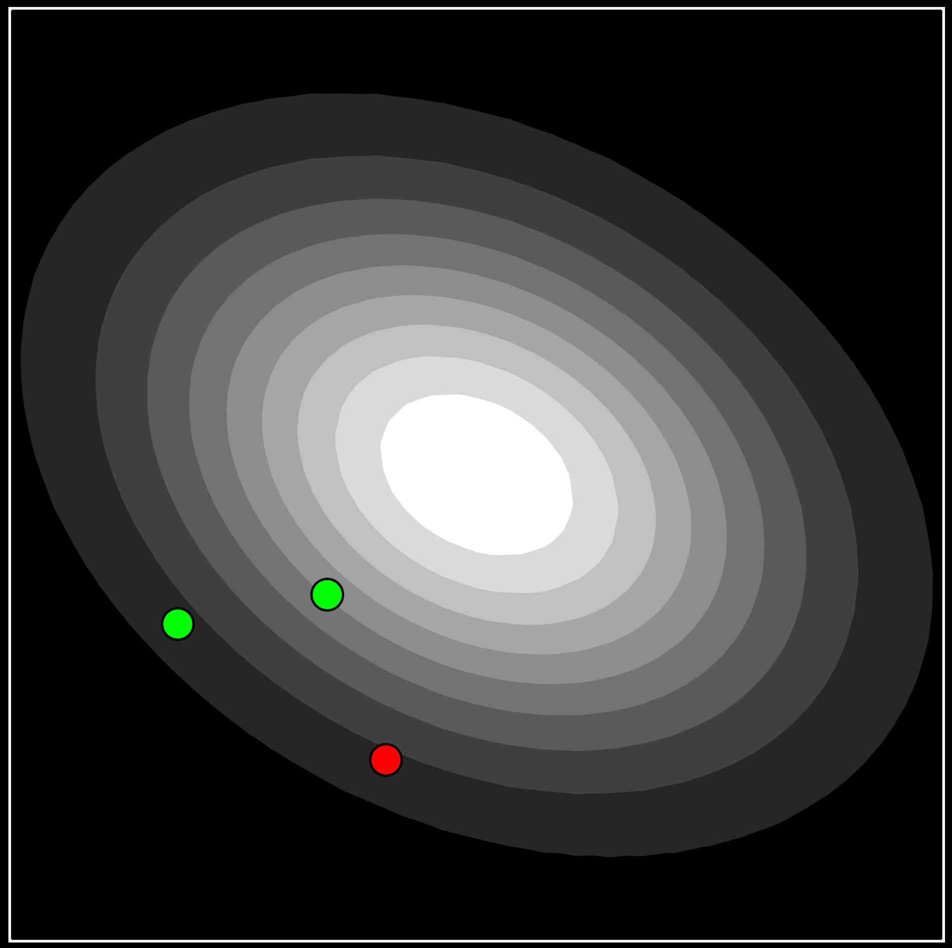


target distribution $P(\theta)$

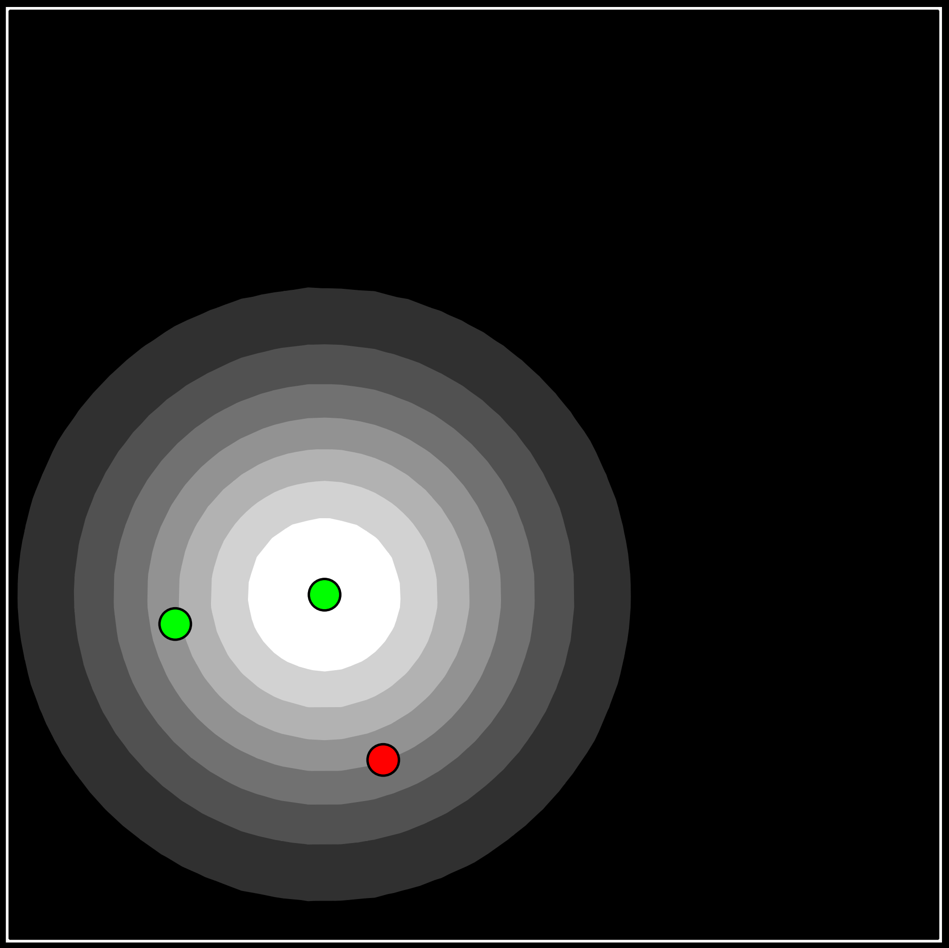


proposal distribution $Q(\theta)$

Metropolis-Hasting

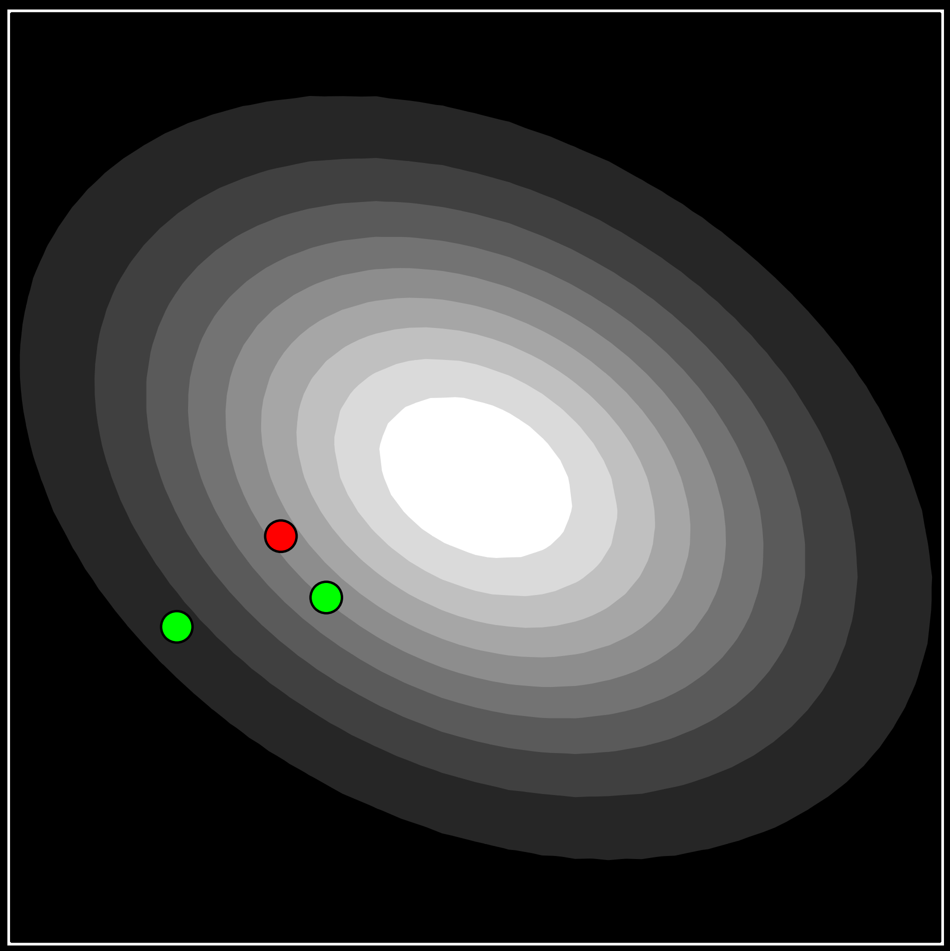


target distribution P(θ)

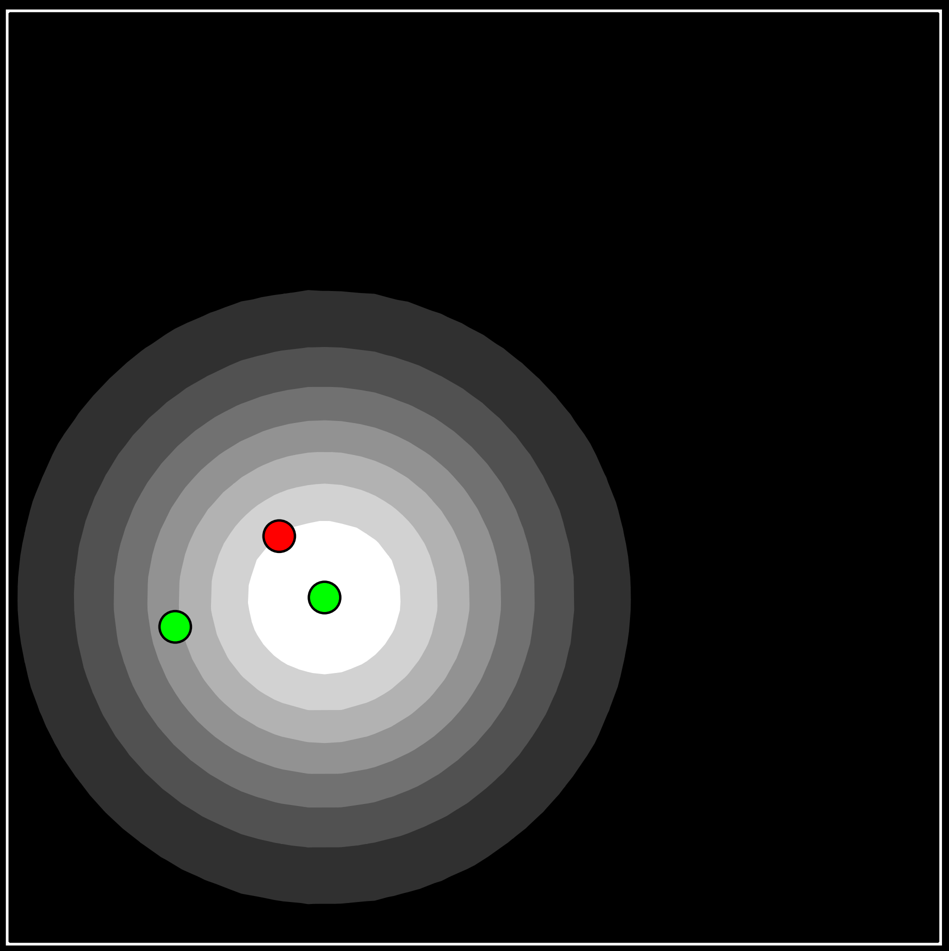


proposal distribution Q(θ)

Metropolis-Hasting

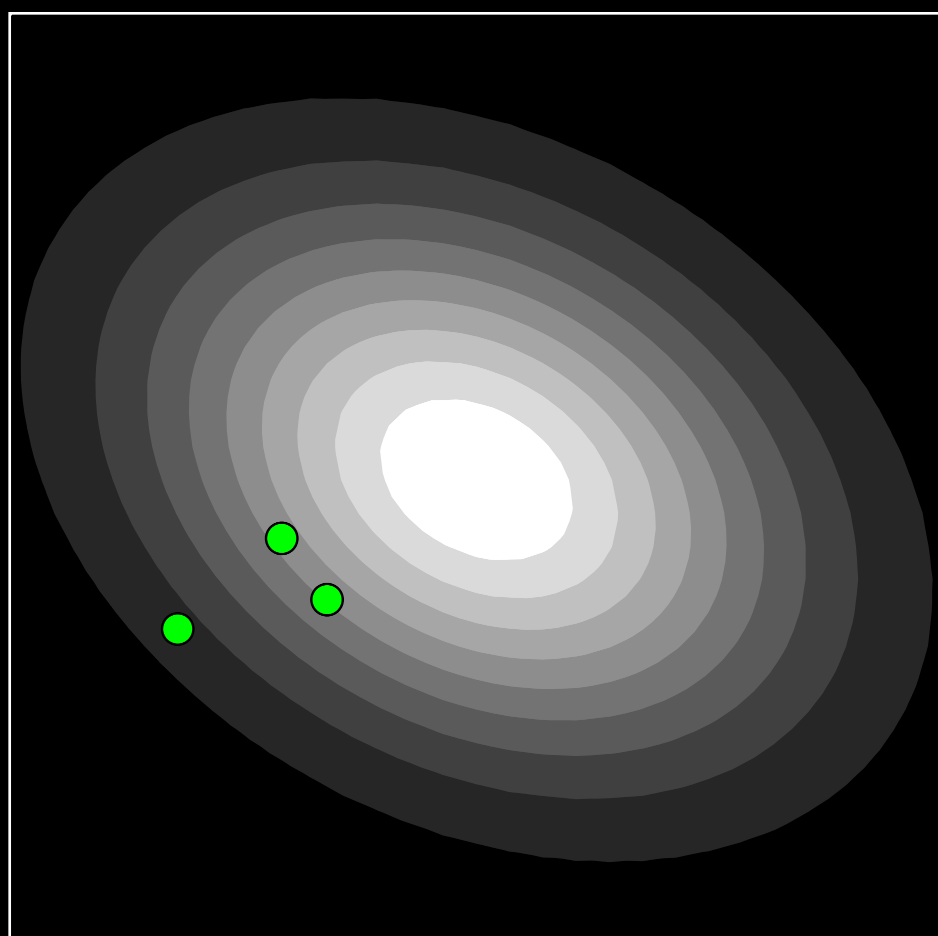


target distribution P(θ)

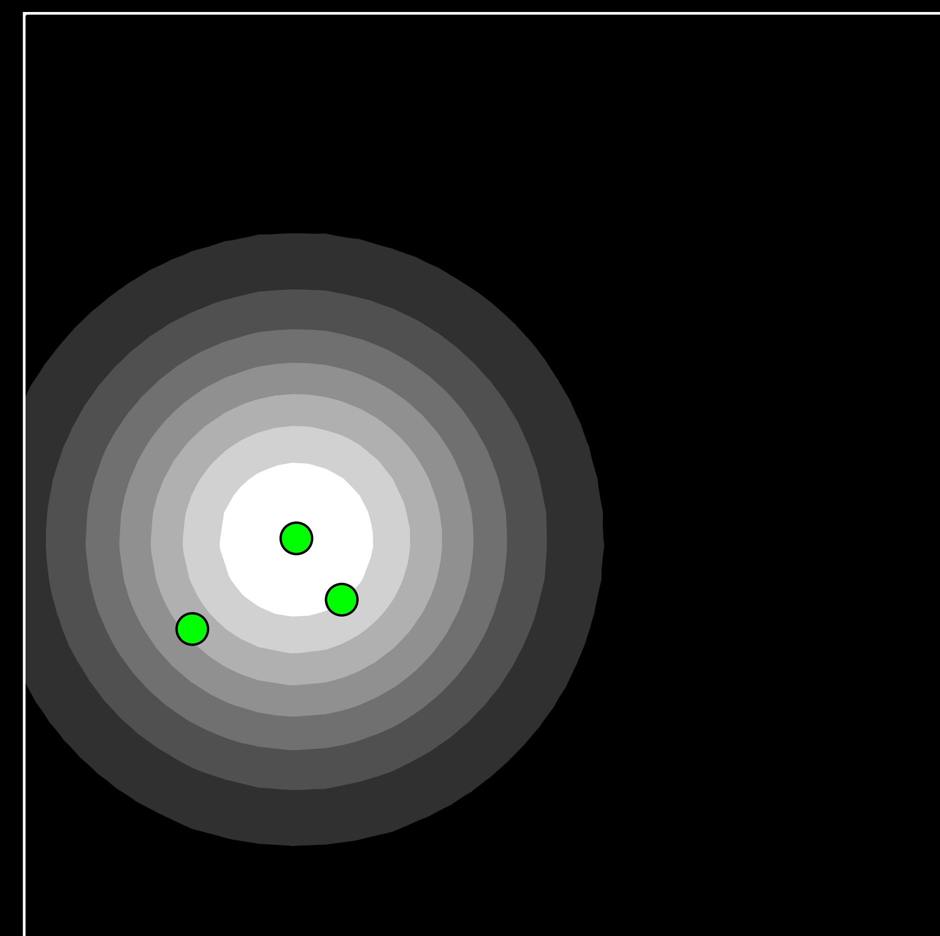


proposal distribution Q(θ)

Metropolis-Hasting

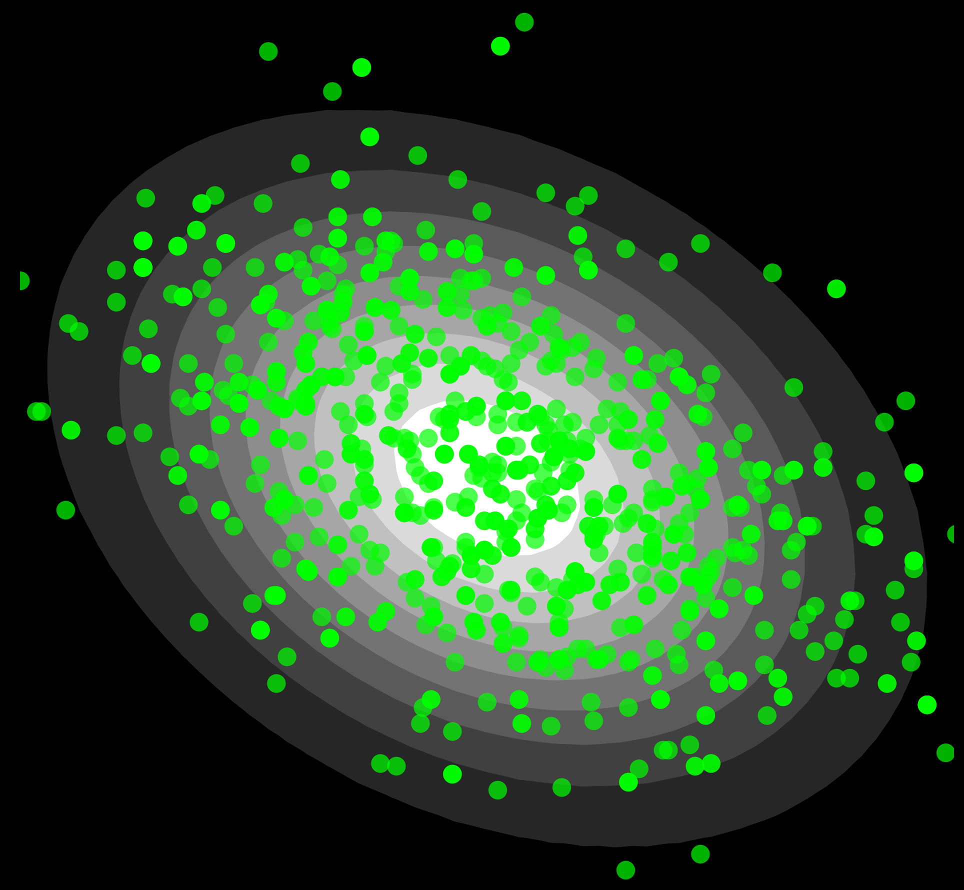


target distribution P(θ)



proposal distribution Q(θ)

Metropolis-Hasting



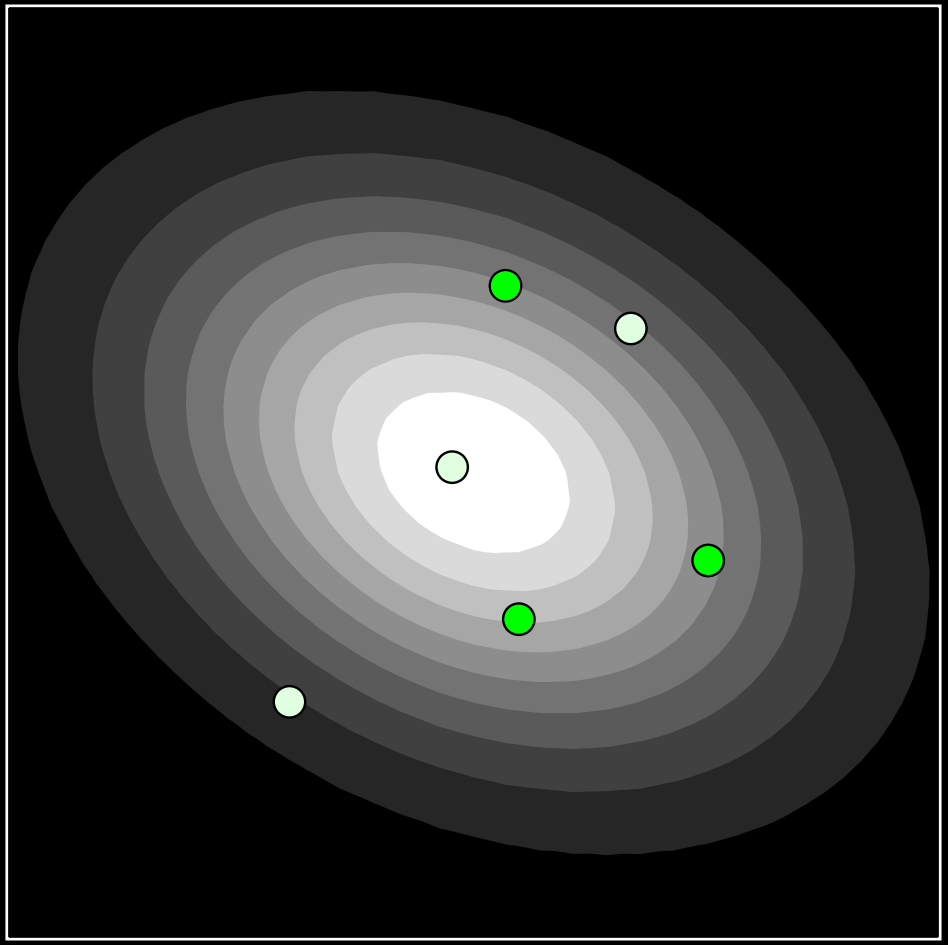
after 1,000 iterations

MCMC in Cosmology

- Estimation of cosmological parameters with MCMC is time consuming
- Multiple runs of the same likelihood
  - Learn about different models and configurations

emcee

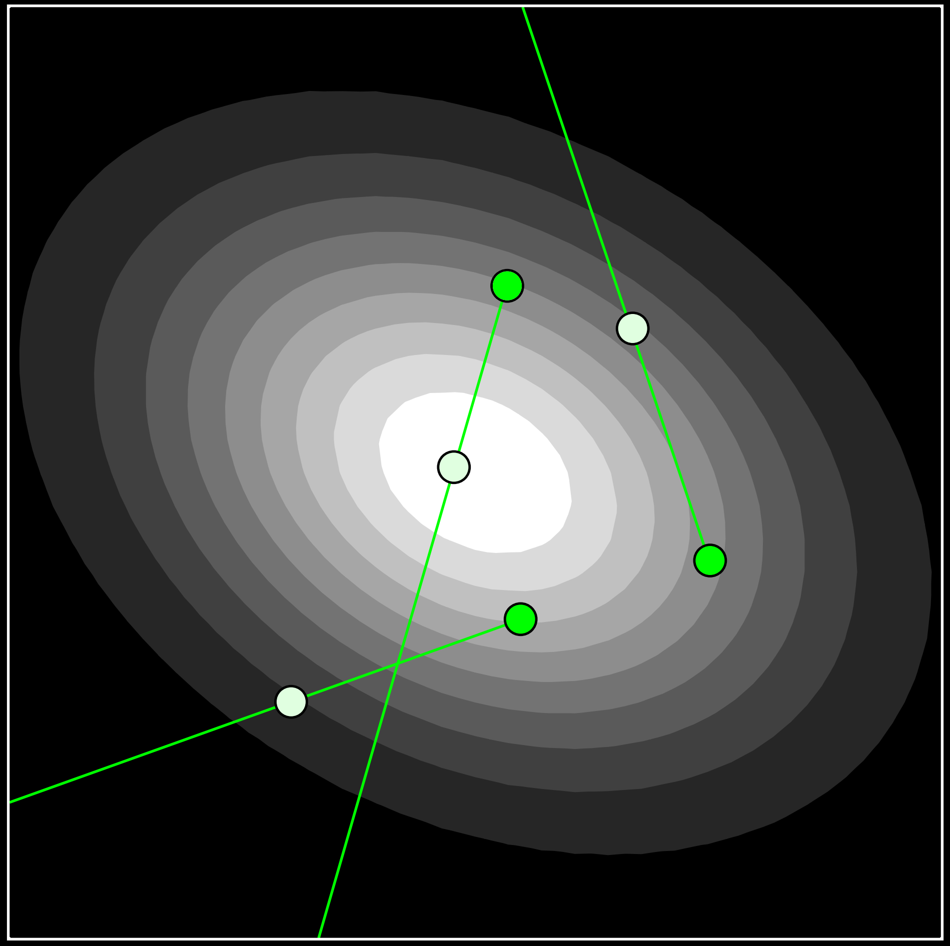
Foreman-Mackey+ 2012
Goodman+ 2010



initial positions $\theta_0^1, \cdots, \theta_0^m$

emcee

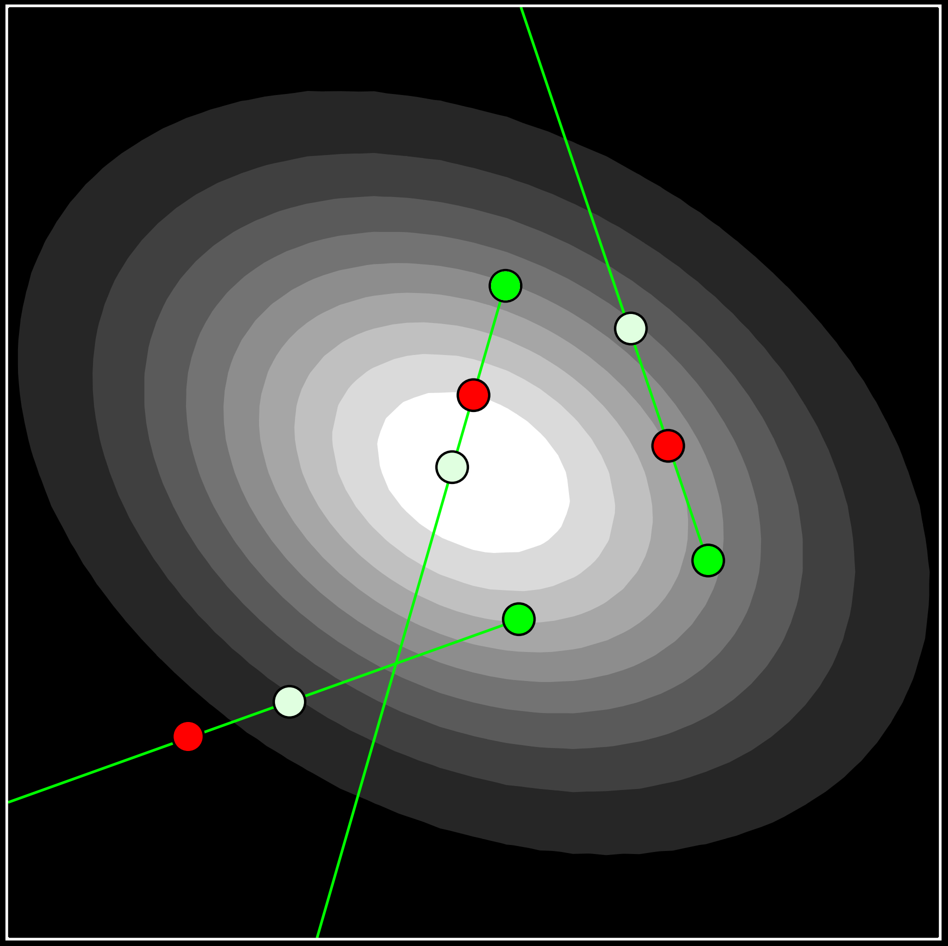
Foreman-Mackey+
2012
Goodman+ 2010



pick a partner at random

emcee

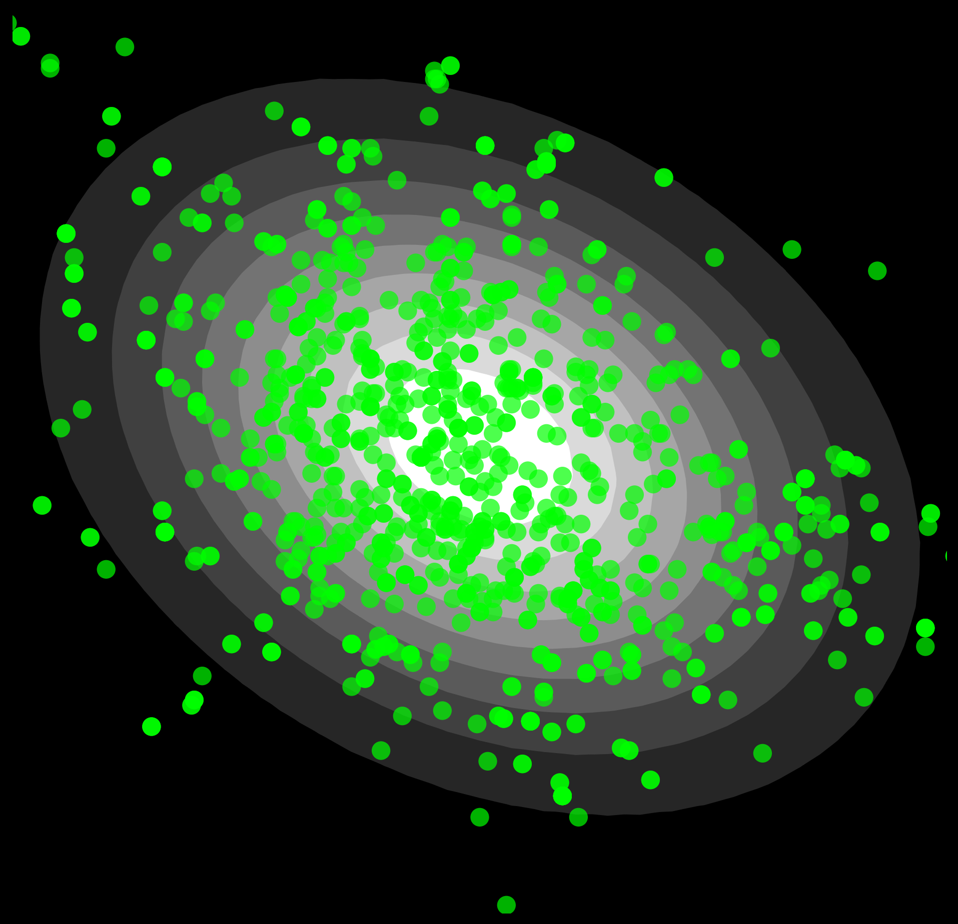
Foreman-Mackey+
2012
Goodman+ 2010



propose positions on connecting rays

emcee

Foreman-Mackey+ 2012
Goodman+ 2010

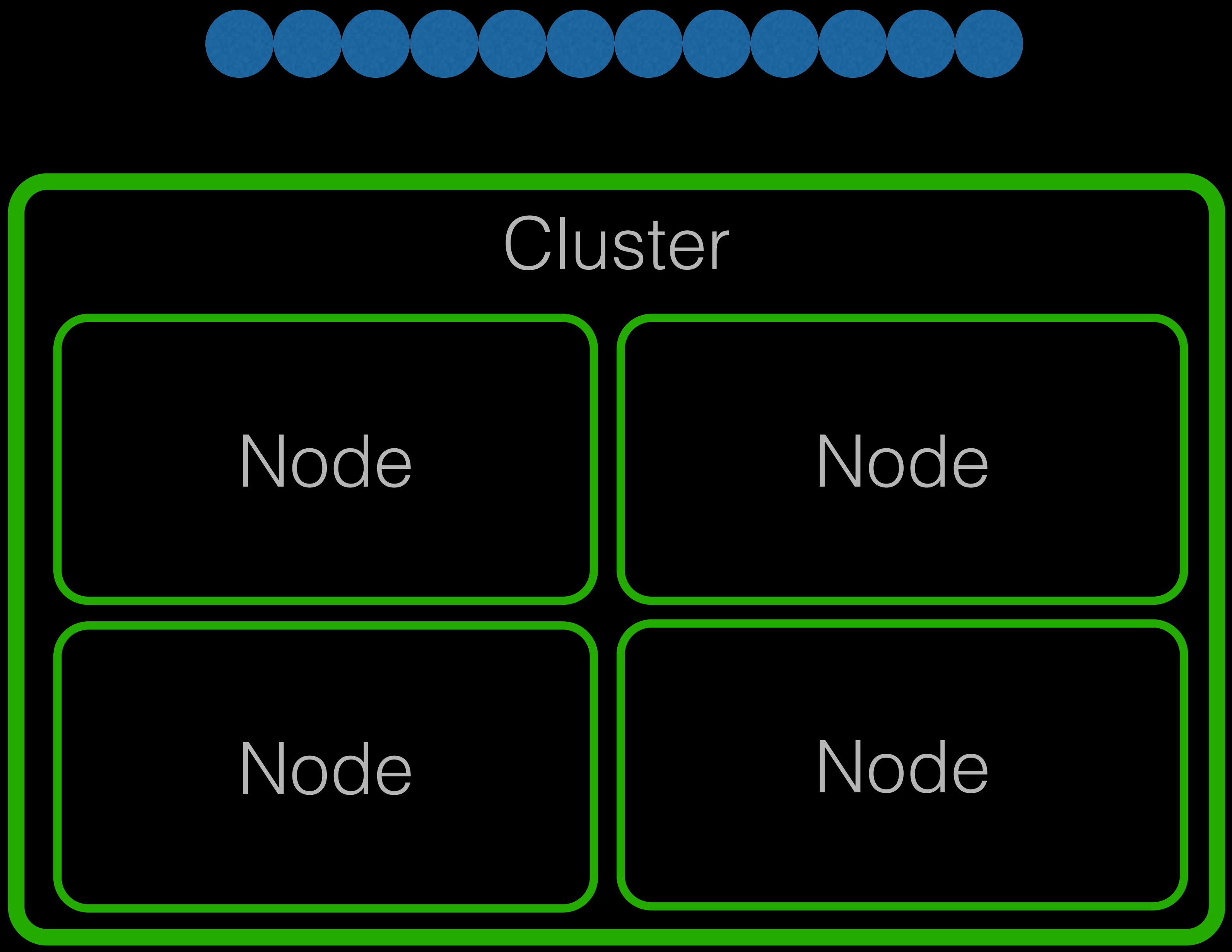


after 1,000 iterations

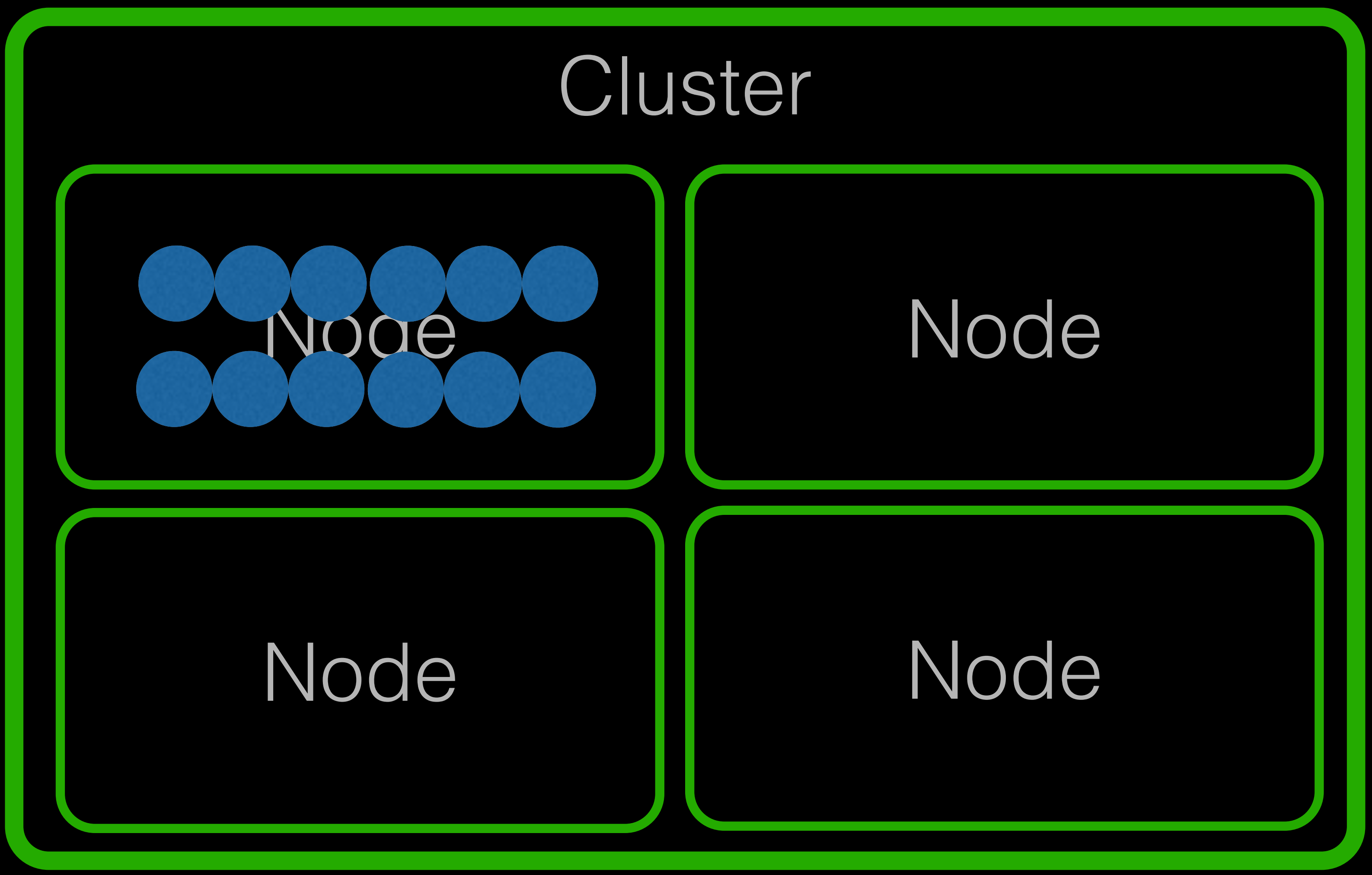
CosmoHammer

- Python framework for parallelised MCMC sampling with emcee
  - Flexible architecture to sample various models and data sets
- Distributes the workload over multiple nodes in a compute cluster
  - Built on MPI and Python multiprocessing

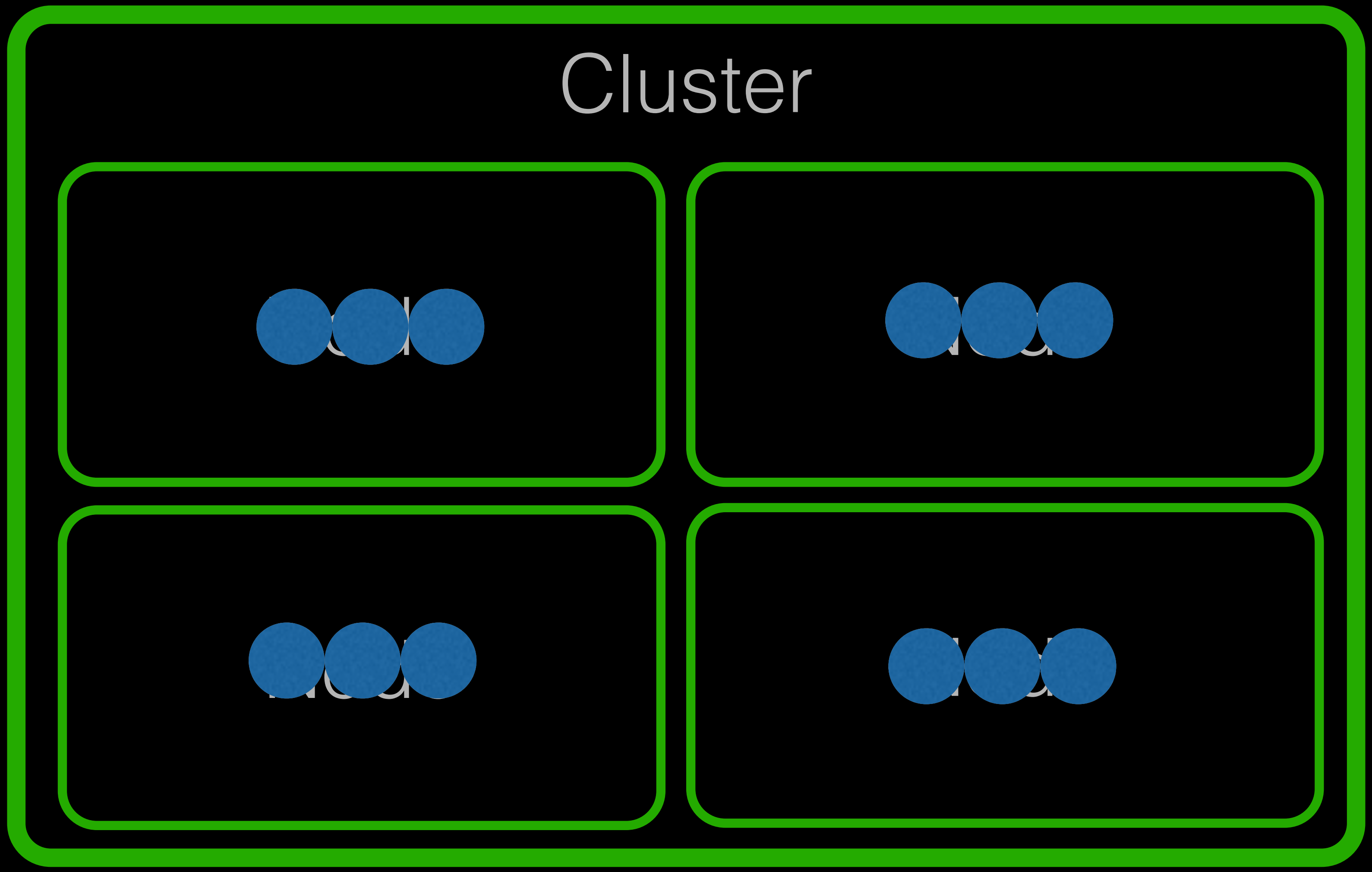
CosmoHammer



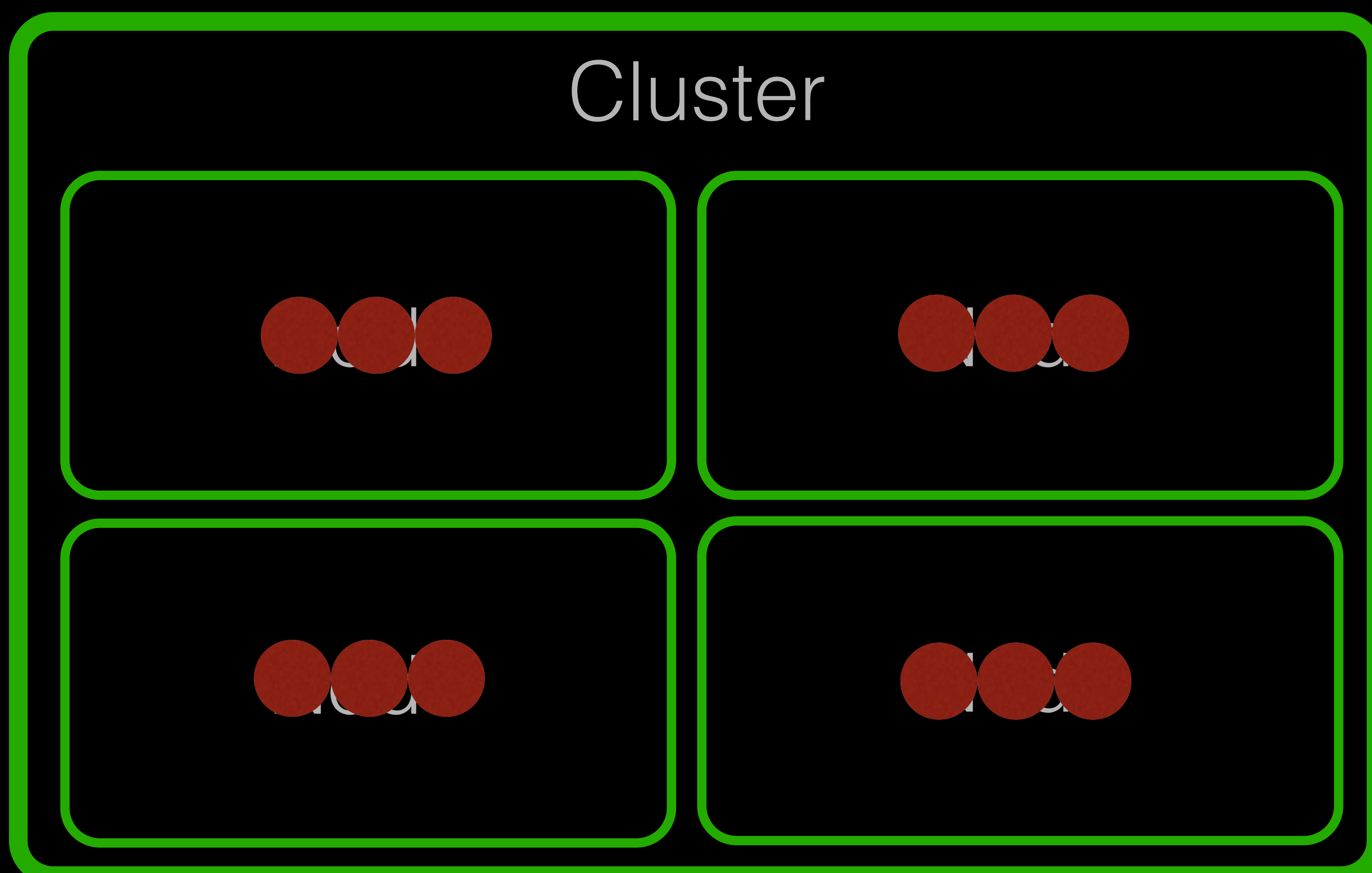
CosmoHammer



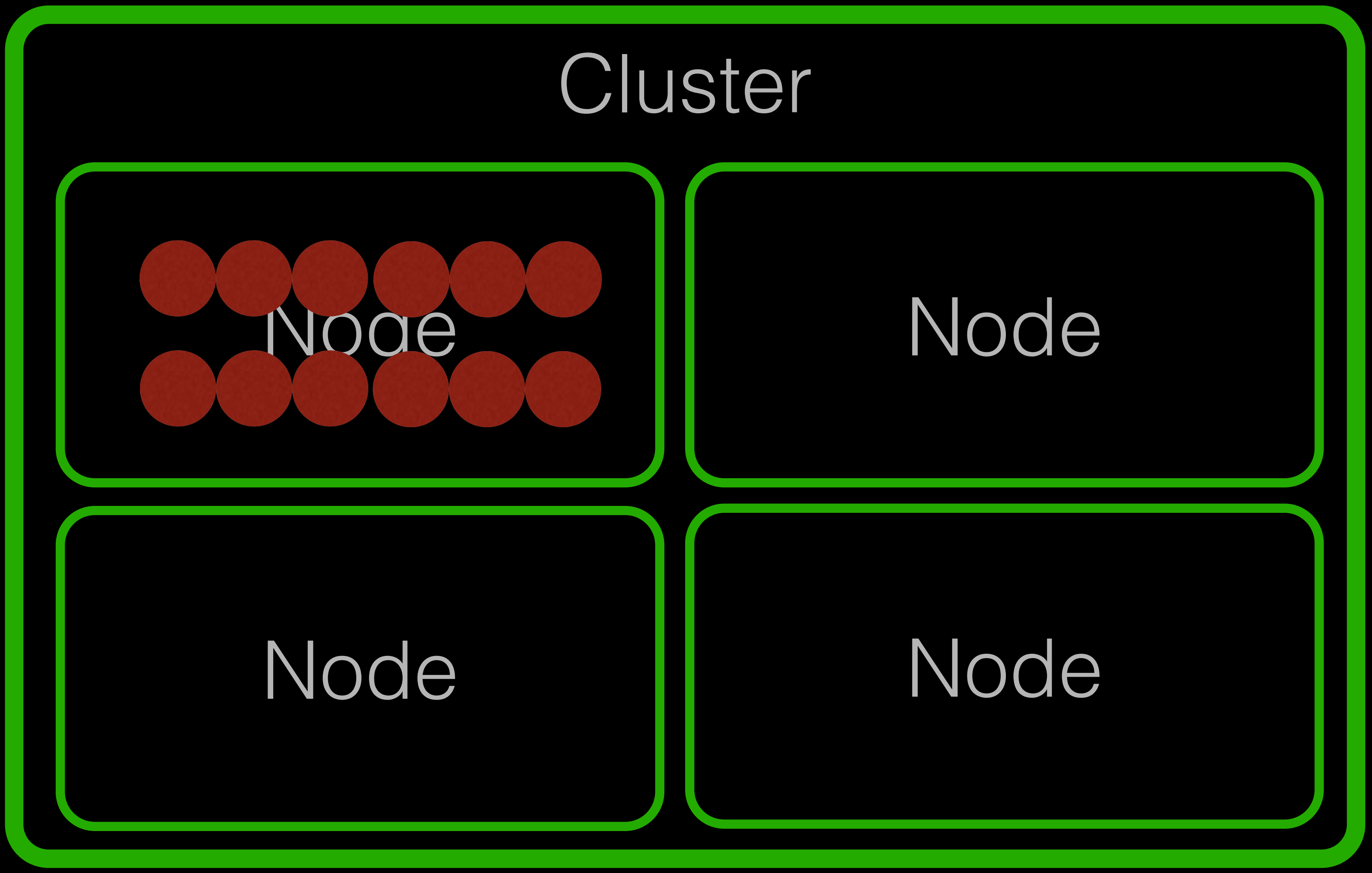
CosmoHammer



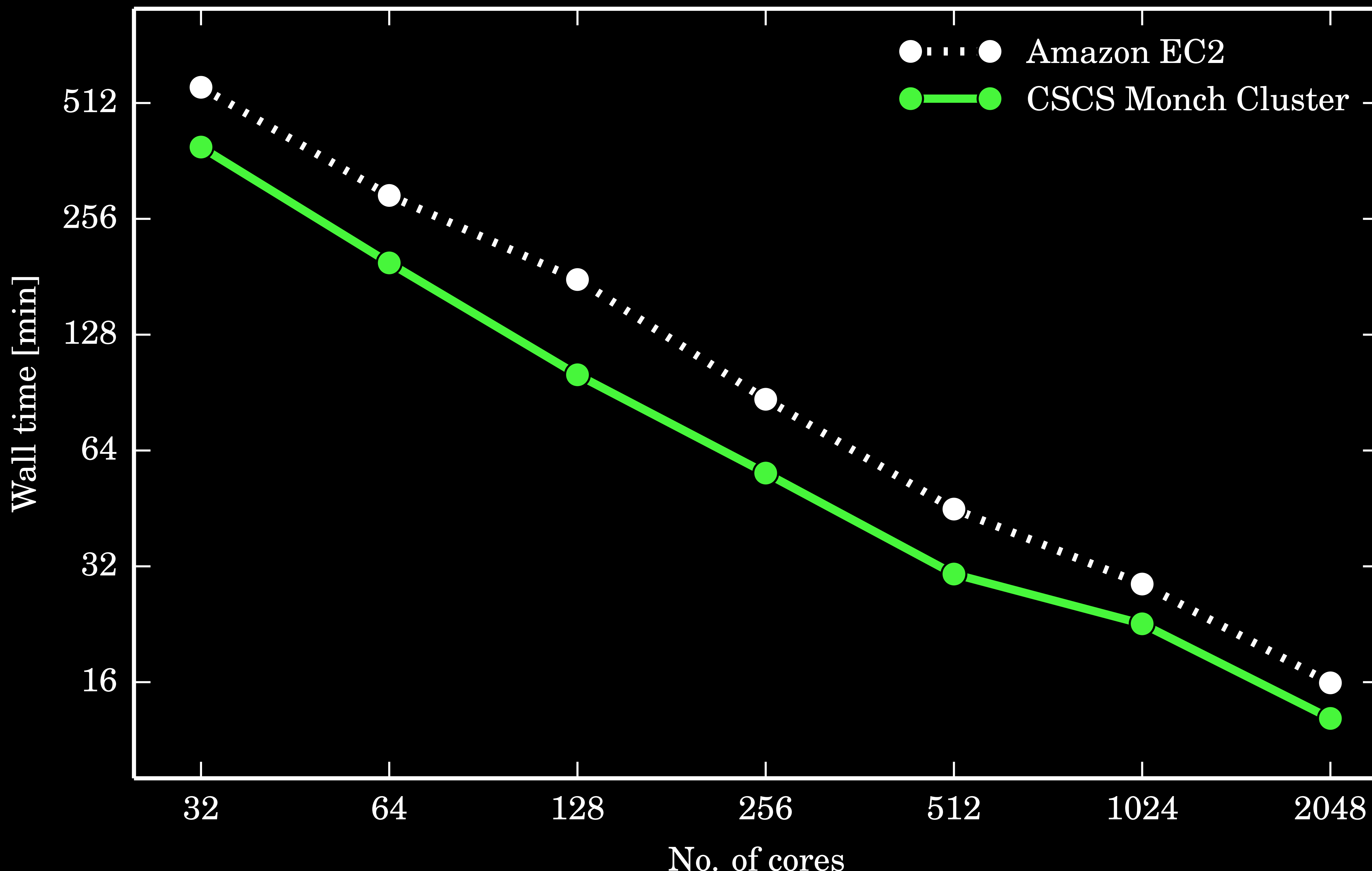
CosmoHammer



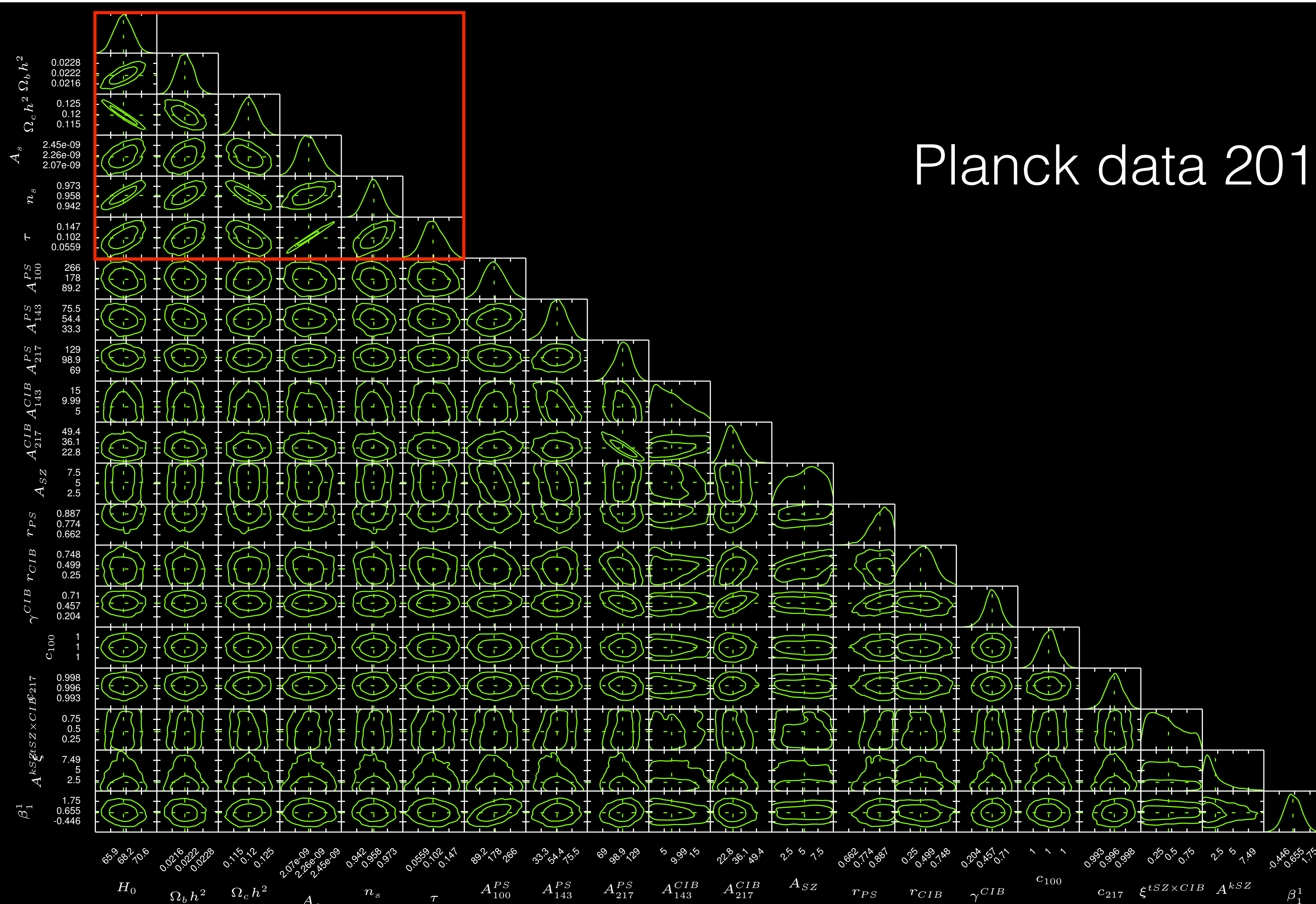
CosmoHammer



CosmoHammer



Current challenges



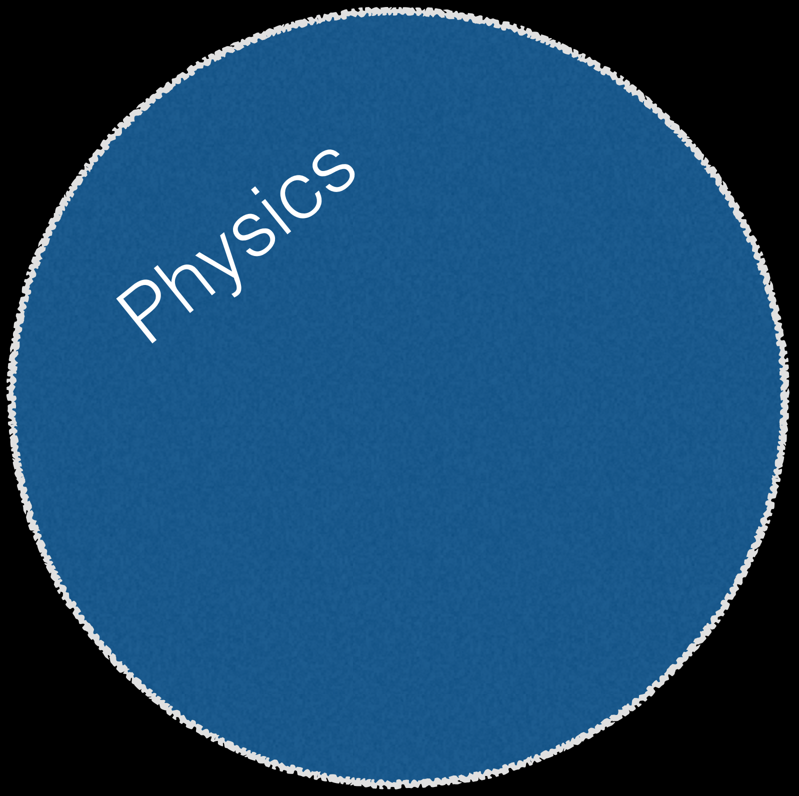
Planck data 2013

Current challenges

- Fast sampling in high dimensions
- Diagnostics & Convergence criteria
- Handling of systematic effects

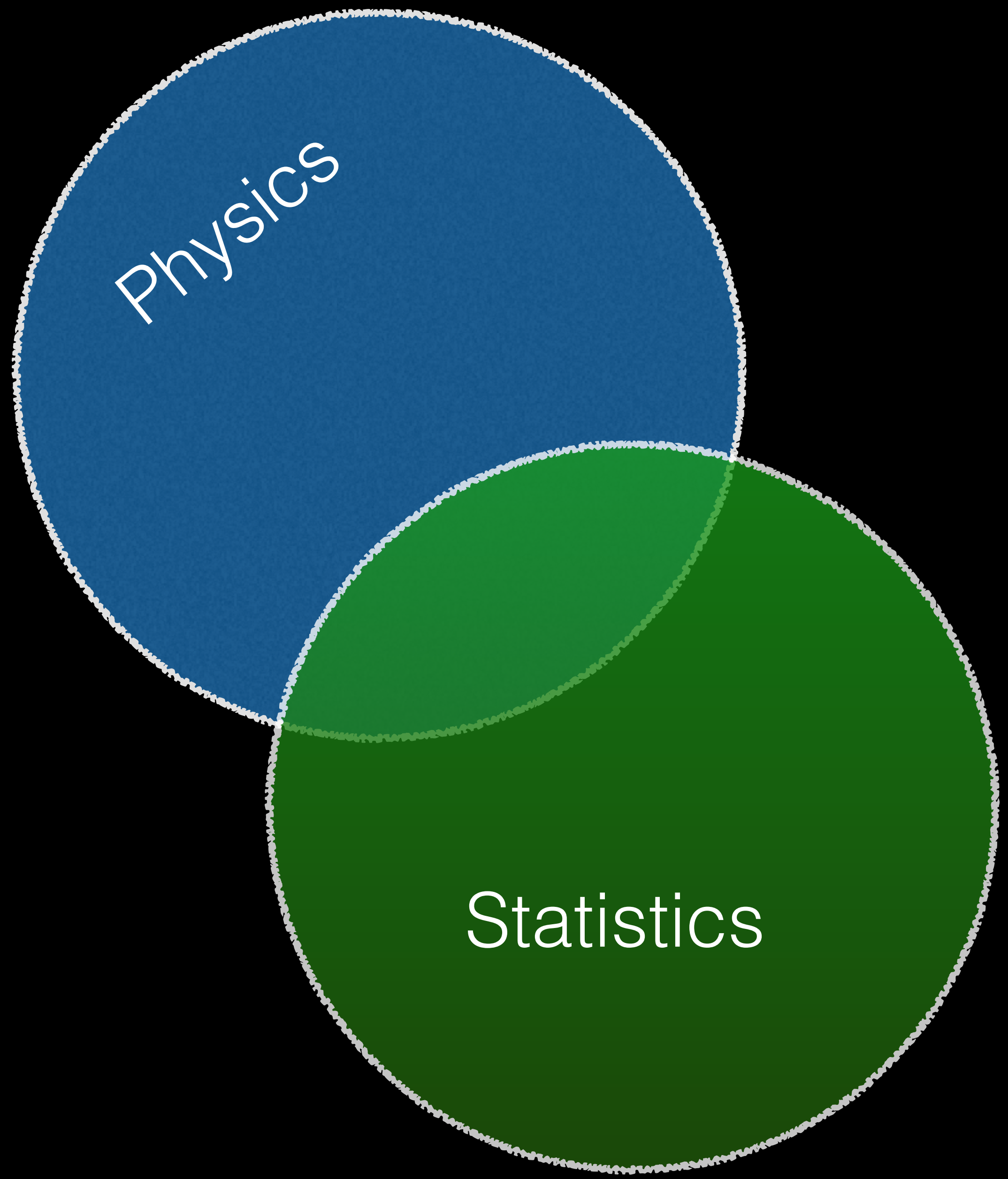
Current challenges

- Fast sampling in high dimensions
- Diagnostics & Convergence criteria
- Handling of systematic effects



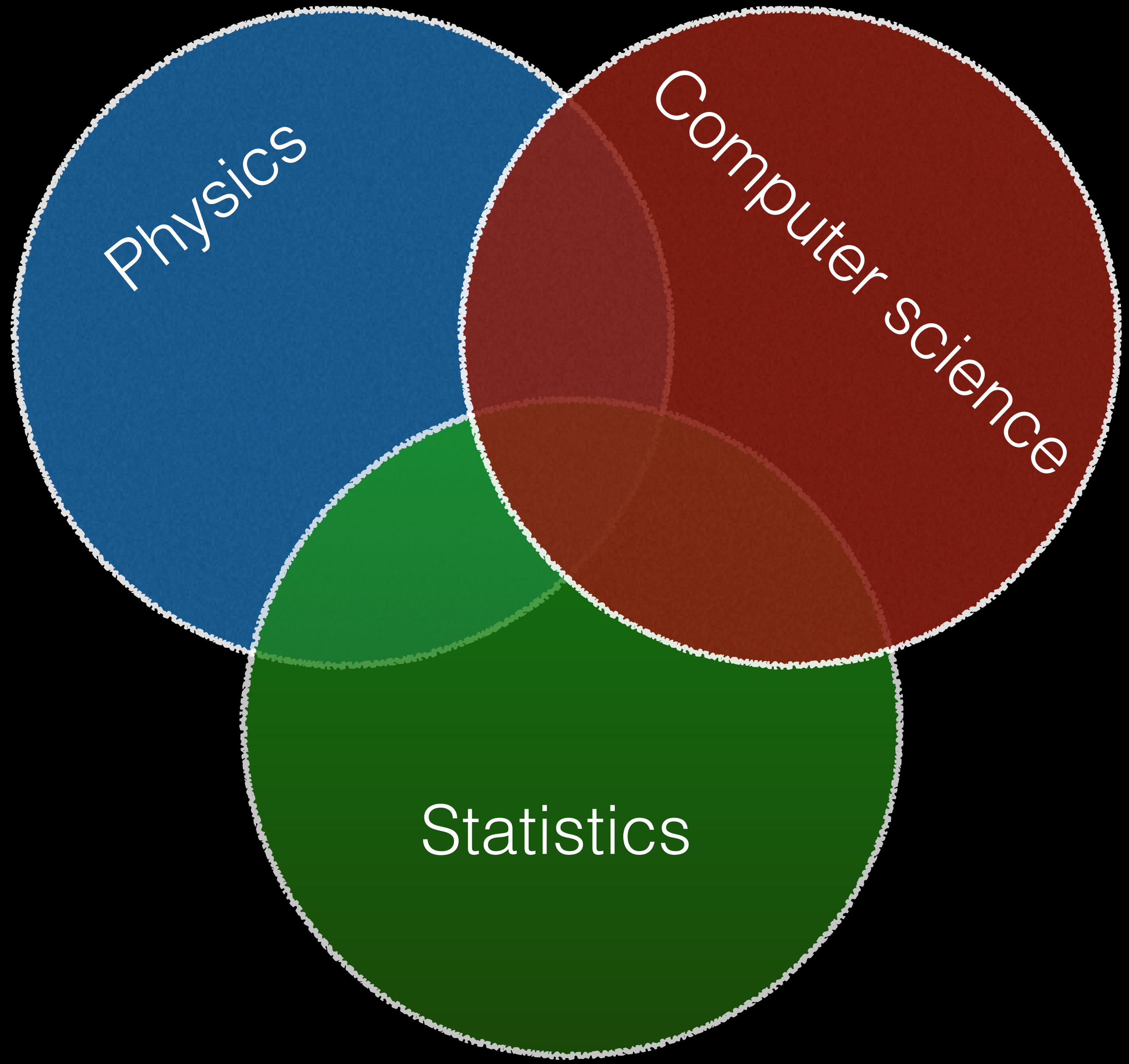
Current challenges

- Fast sampling in high dimensions
- Diagnostics & Convergence criteria
- Handling of systematic effects



Current challenges

- Fast sampling in high dimensions
- Diagnostics & Convergence criteria
- Handling of systematic effects



CosmoHammer References

- Akeret J., Seehars S., Amara A., Refregier A., and Csillaghy A. (2013). Astronomy and Computing, Volume 2, Pages 27-39
  - http://dx.doi.org/10.1016/j.ascom.2013.06.003
- Available at http://www.astro.ethz.ch/refregier/research/Software/cosmohammer