Computational Challenges in Computational Cosmology - Simulators, Differentiable Models, and Surrogates

CSGF Program Review 2022

Jamie Sullivan
w/ Uros Seljak, Sukhdeep Singh, JD Emberson, Matt Becker, Salman Habib, Zack Li, Marius Millea

Berkeley UNIVERSITY OF CALIFORNIA

DOE CSGF

BERKELEY CENTER *for* COSMOLOGICAL PHYSICS

Overview

1. The Cosmological Inverse Problem

2. Fast Differentiable Models

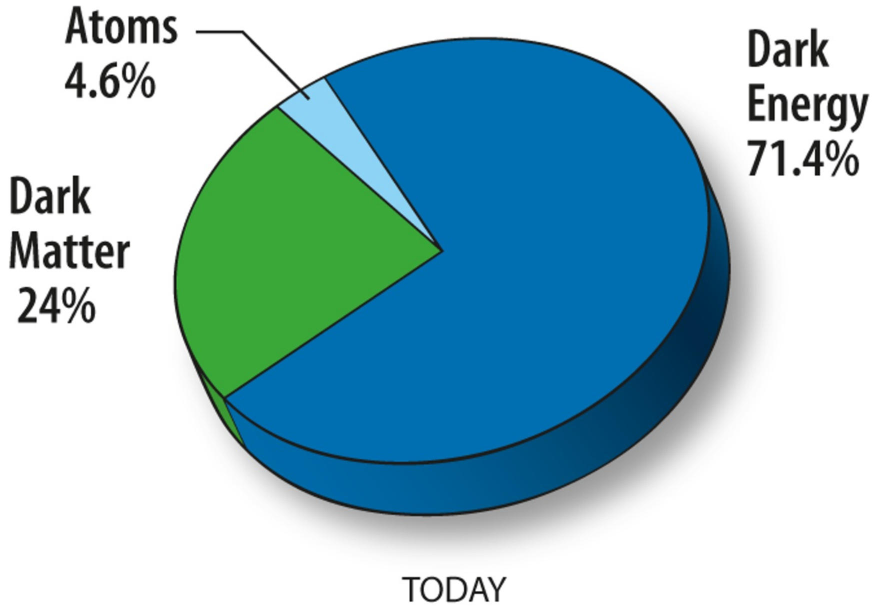
3. Efficient Posterior Inference

The Cosmological Inverse Problem

An abundance of ignorance

What is:

- dark matter?
- dark energy?
- quantum gravity?



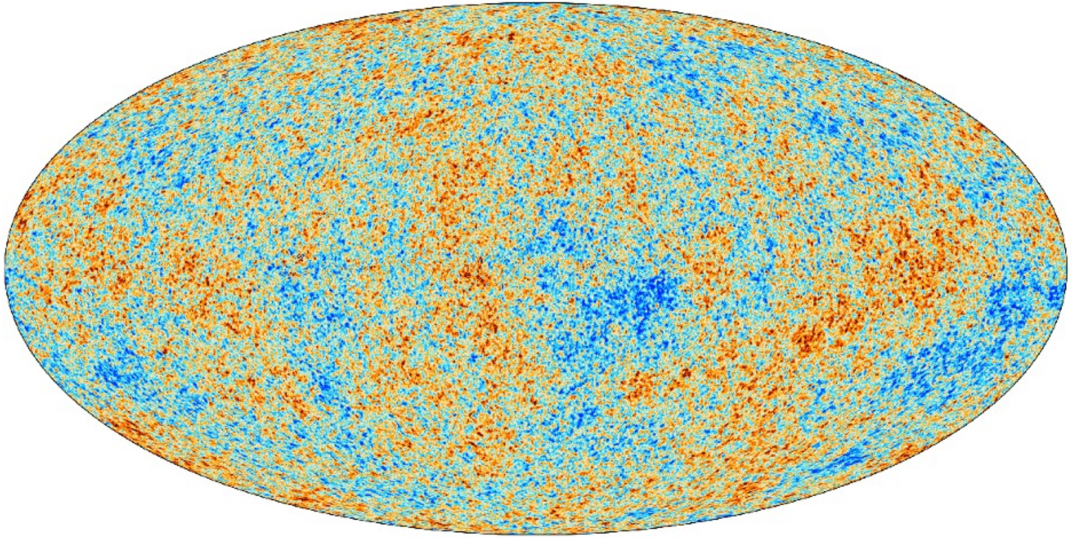
WMAP/NASA GSFC

An abundance of ignorance

What is:

- dark matter?
- dark energy?
- quantum gravity?

Cosmology brings **lots of data** to these questions!



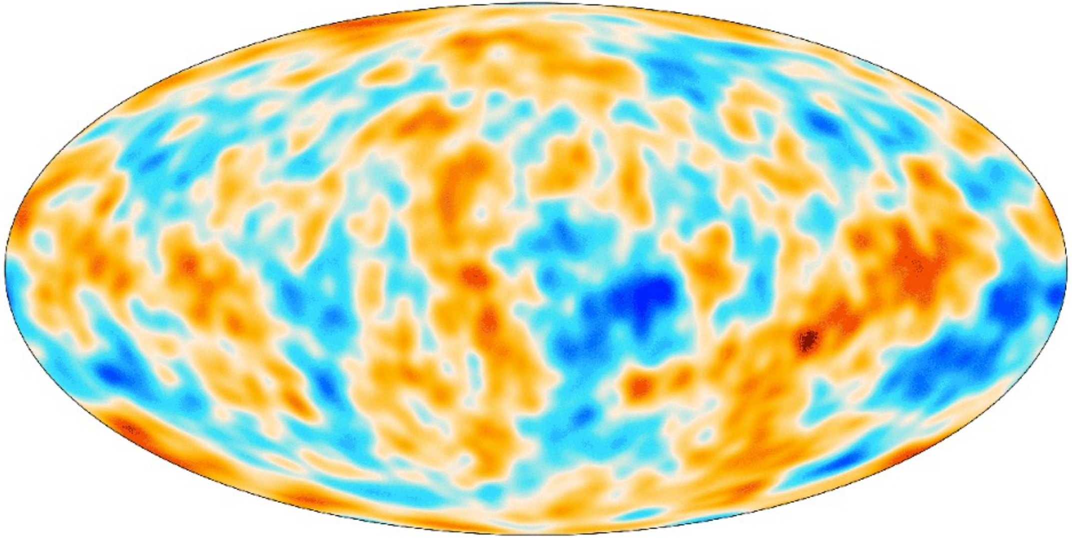
ESA/Planck

An abundance of ignorance

What is:

- dark matter?
- dark energy?
- quantum gravity?

Cosmology brings **lots of data** to these questions!



ESA/Planck

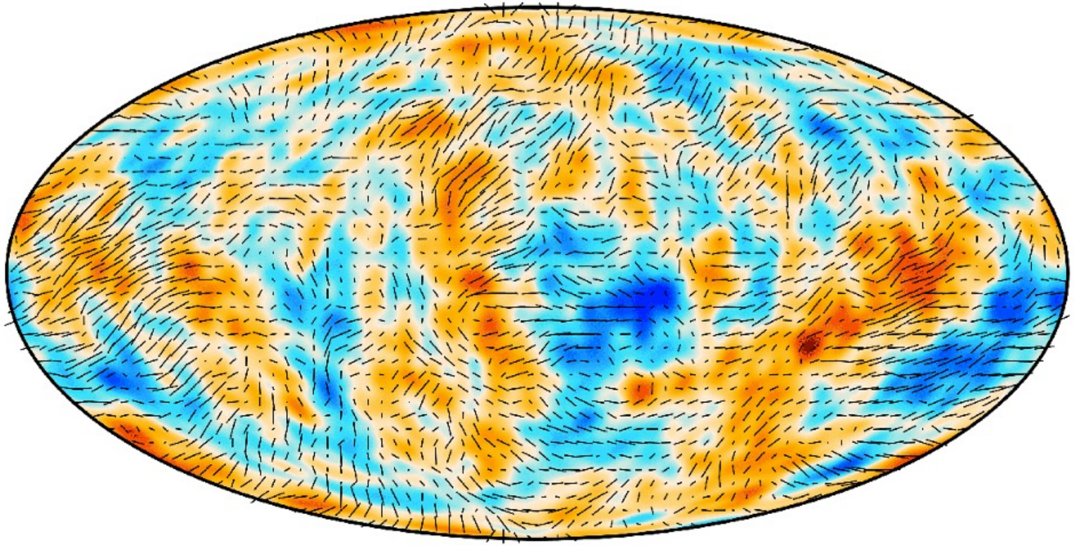
An abundance of ignorance

What is:

- dark matter?
- dark energy?
- quantum gravity?

Cosmology brings **lots of data** to these questions!

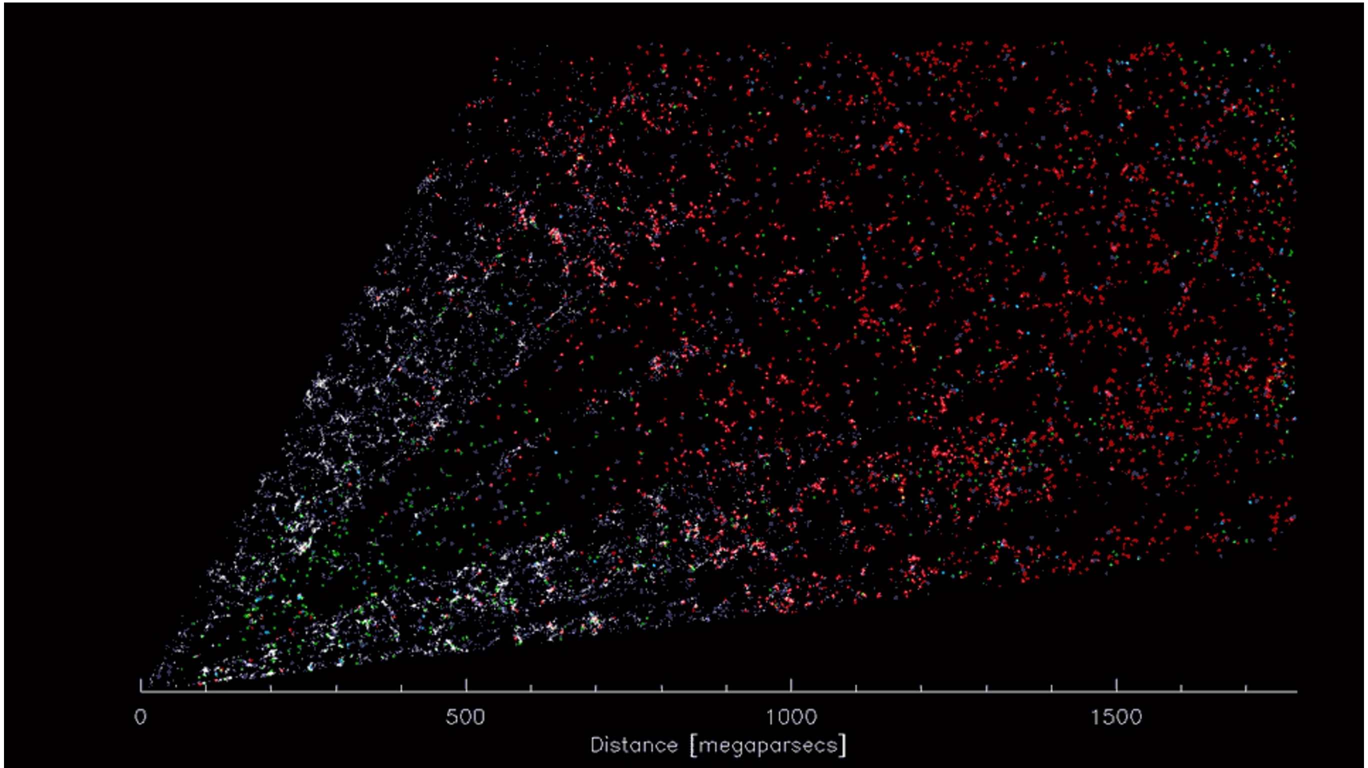
How do we use it?



ESA/Planck

Large-scale Structure Cosmology

Focus only on *spatial statistics* - how clustered are galaxies?



Large-scale Structure Cosmology

Focus only on *spatial statistics* - how clustered are galaxies?

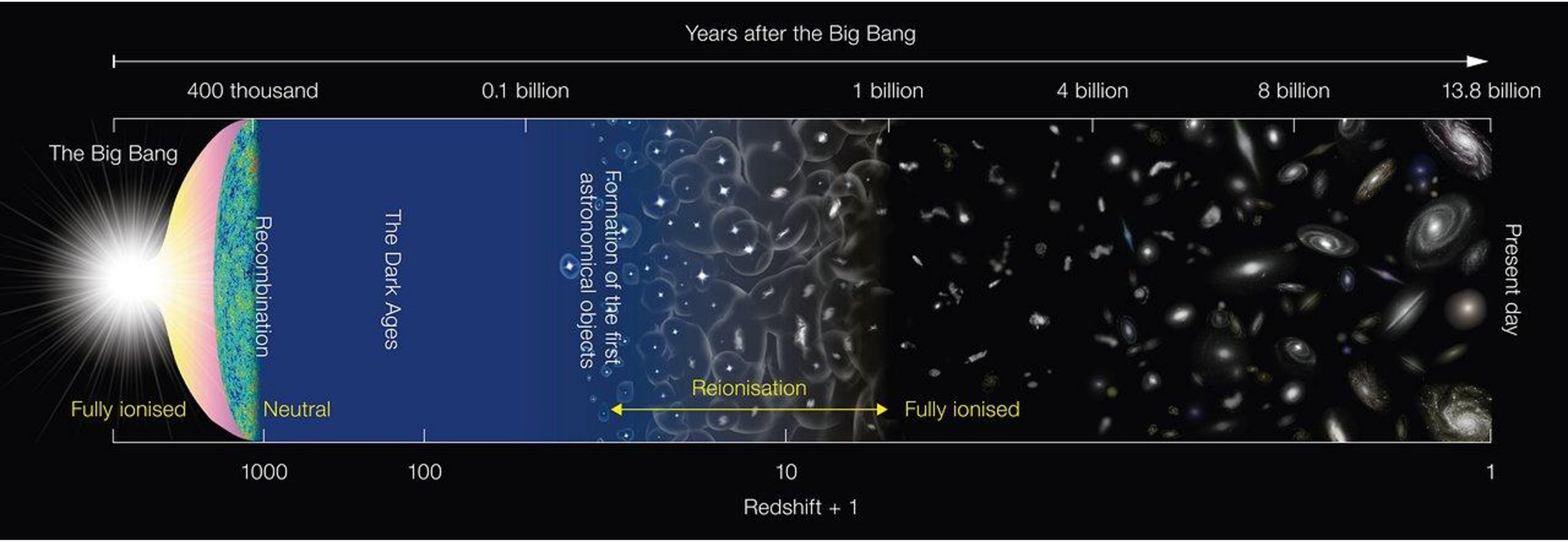
Answer in the form of Bayesian inverse problem:

1. Choose parameters $\boldsymbol{\theta}$
2. Pass through forward model $f(\boldsymbol{\theta})$ for spatial statistic
3. Compare $f(\boldsymbol{\theta})$ to data via the posterior
4. Optimize+sample the posterior

What is the forward model? How expensive is it?

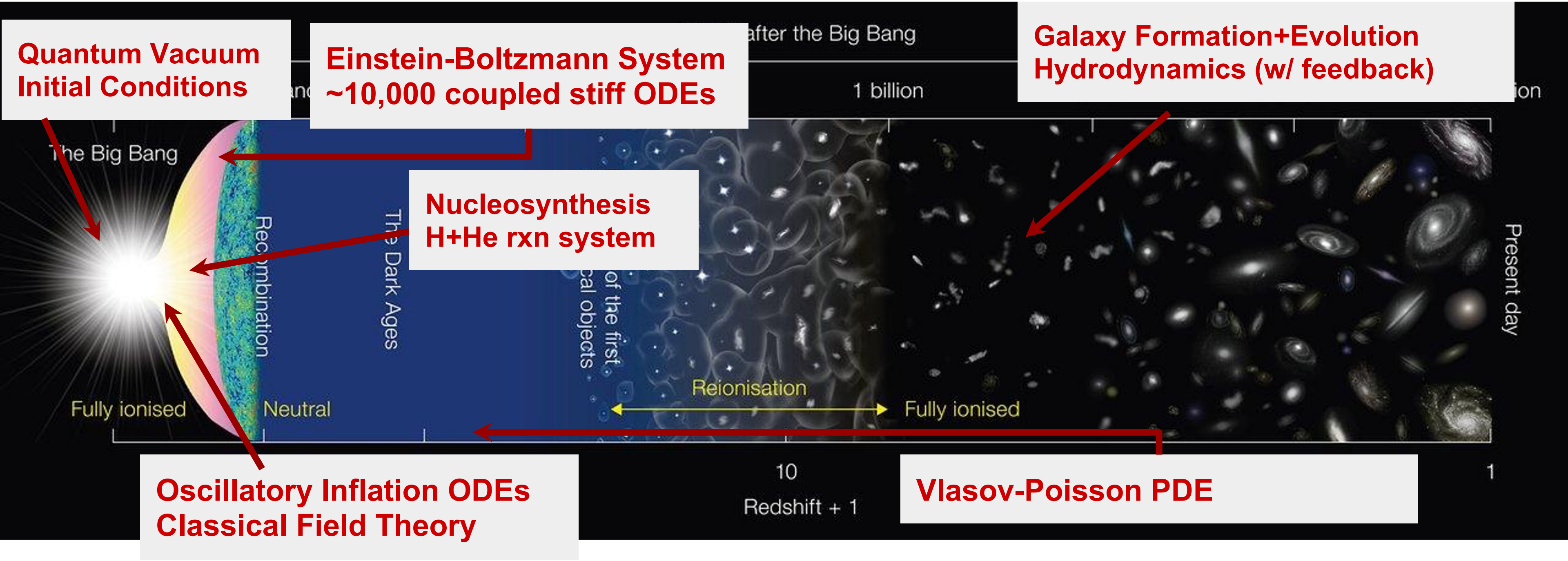
Timeline of the Universe - in equations

How do we get to galaxies in LSS surveys? **Solve equations!**



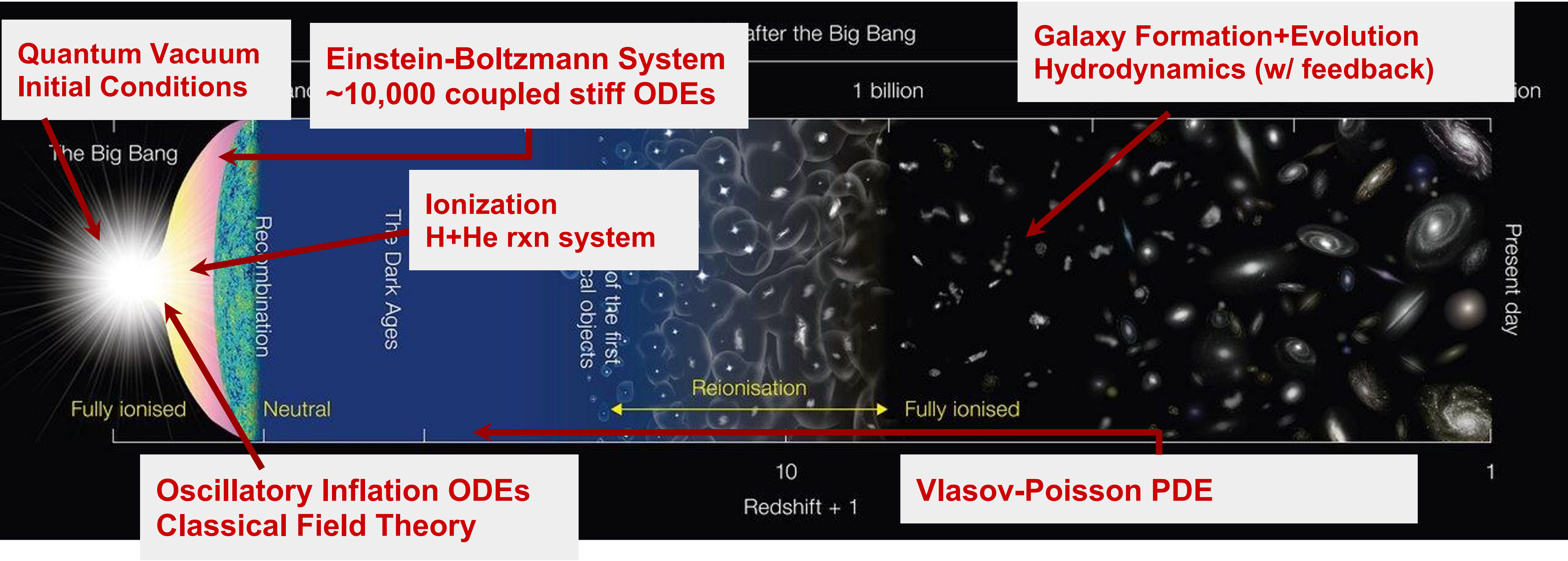
Timeline of the Universe - in equations

How do we get to galaxies in LSS surveys? **Solve equations!**



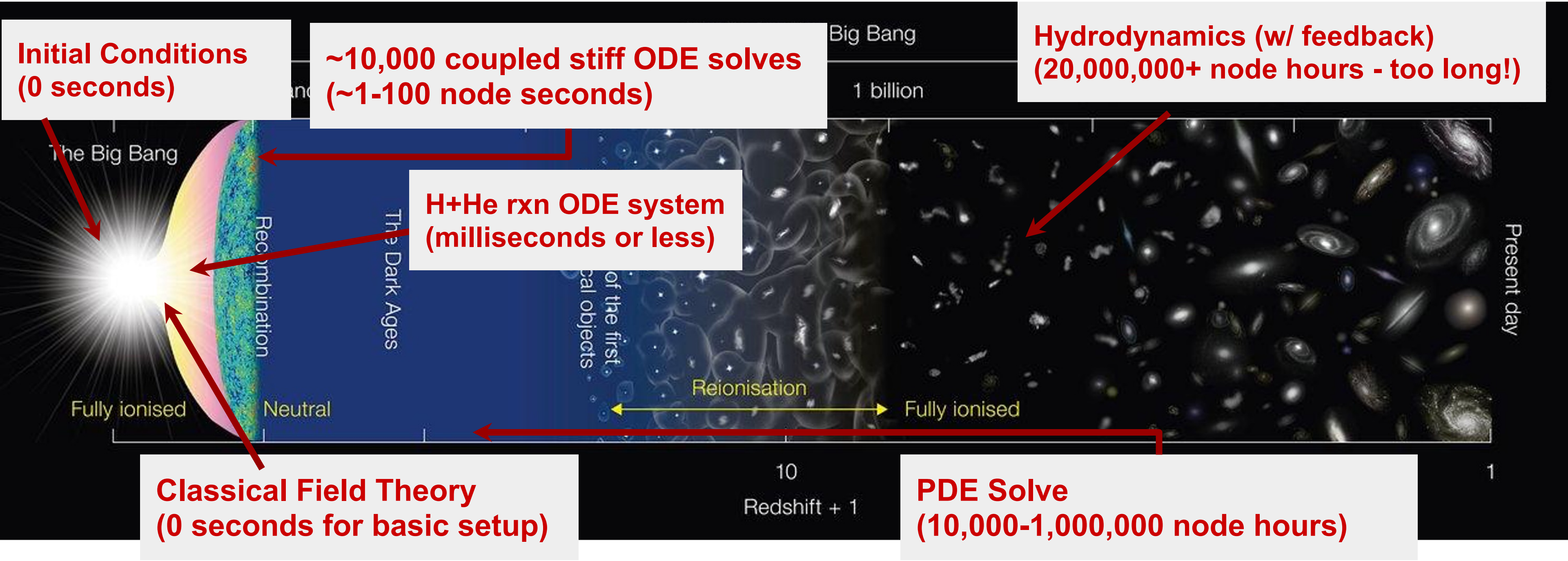
Timeline of the Universe - ~~in equations~~

How do we get to galaxies in LSS surveys? **Solve equations!**



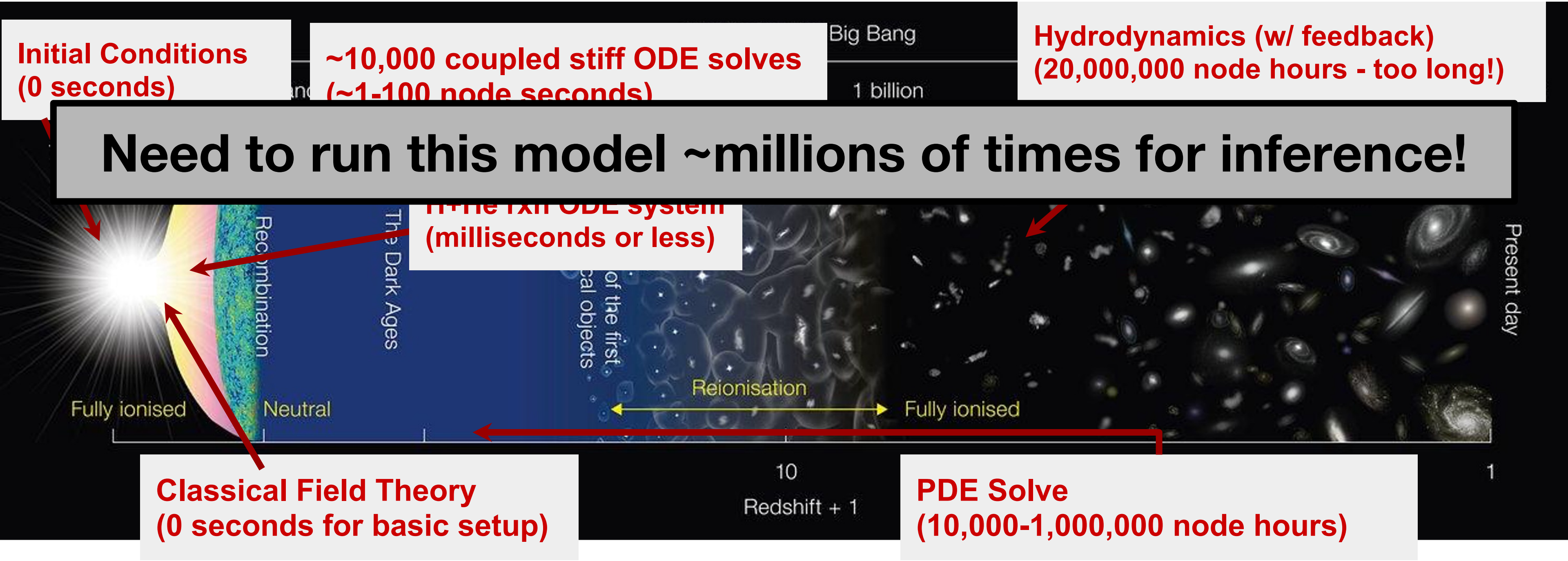
Timeline of the Universe - in *compute time*

How do we get to galaxies in LSS surveys? **Solve equations!**



Timeline of the Universe - in *compute time*

How do we get to galaxies in LSS surveys? **Solve equations!**



Making the Solvable Soluble

Two bottleneck solutions:

- Faster or differentiable solvers
- Smart posterior evaluation for inference

Improved computational capability can enable previously infeasible methods!

Gradients of the Einstein-Boltzmann System

A Differentiable Simulator for the CMB and LSS

w/ Zack Li, Marius Millea

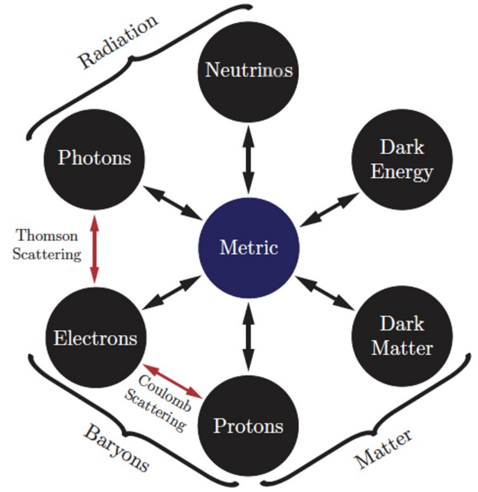
The Cornerstone of Cosmology

LSS and the CMB models evolve perturbations described by GR and the Boltzmann eqn. (Einstein-Boltzmann system)

Accurate modeling requires solving this (stiff) ODE system

Current bottleneck in many applications

Gradient-based methods make this tractable - need model gradient

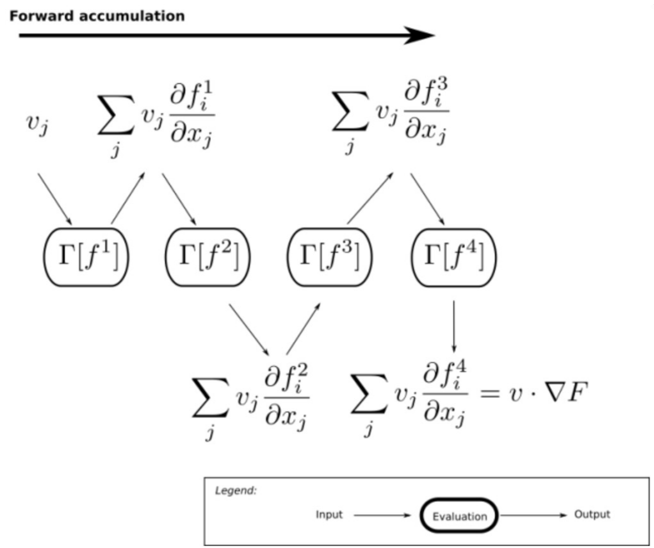


Ma&Bertschinger95, Baumann

Automatic Differentiation

Chain rule for code!

```
julia> derivative(x -> x^2, 1)
2
```



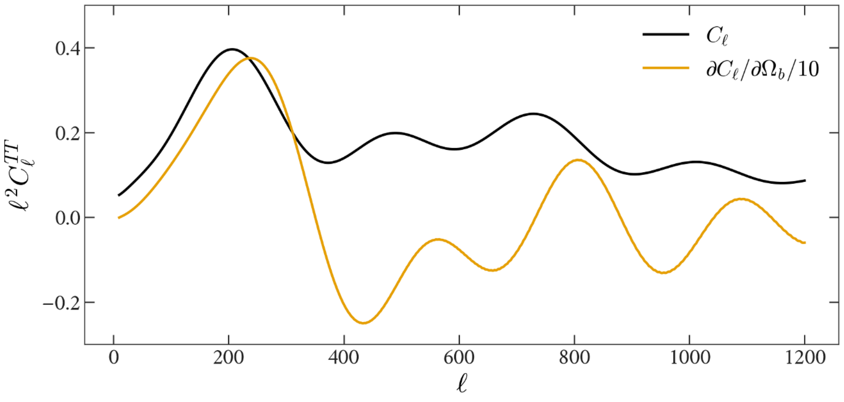
Forward mode AD with dual numbers

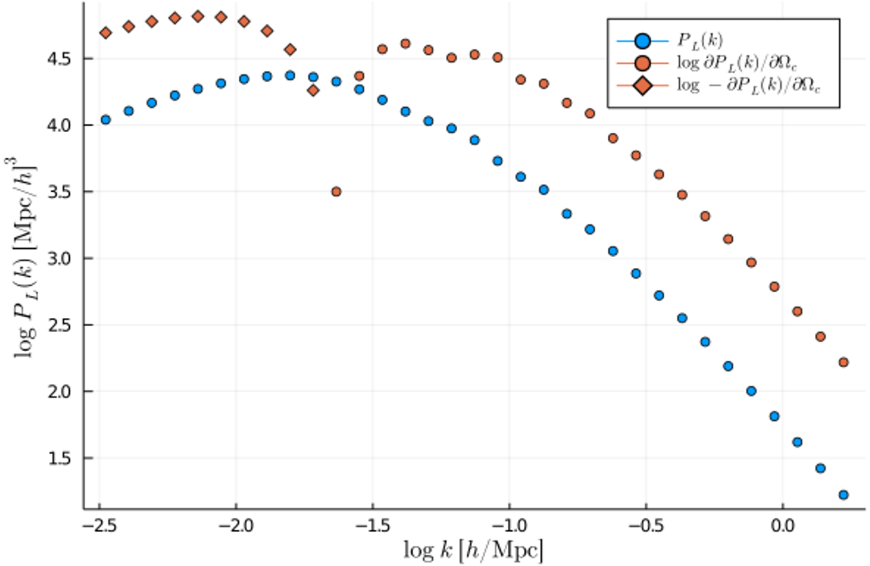
$$f\left(a + \sum_{i=1}^{N} b_i \epsilon_i\right) = f(a) + f'(a) \sum_{i=1}^{N} b_i \epsilon_i$$

AD through stiff ODE solver

Gradients for the CMB and Galaxy Surveys

Exact gradients of the two main observables: CMB anisotropy, linear matter power spectra





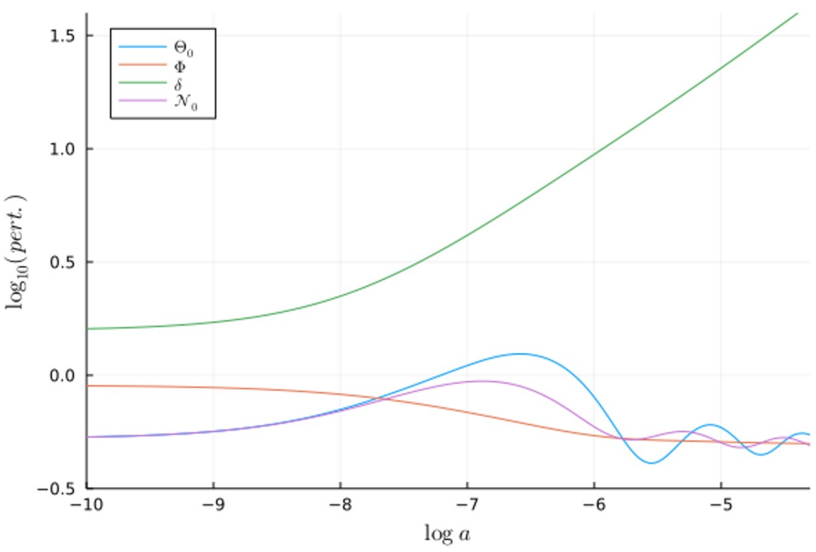
Bolt.jl runs on GPU!

First Boltzmann code that can run on GPU

In the right setting, can lead to enormous performance gains

Minimal code rewriting for GPU with CUDA.jl, Adapt.jl

These perturbations came out of a GPU!



A Practicum Aside! (Another efficient model)

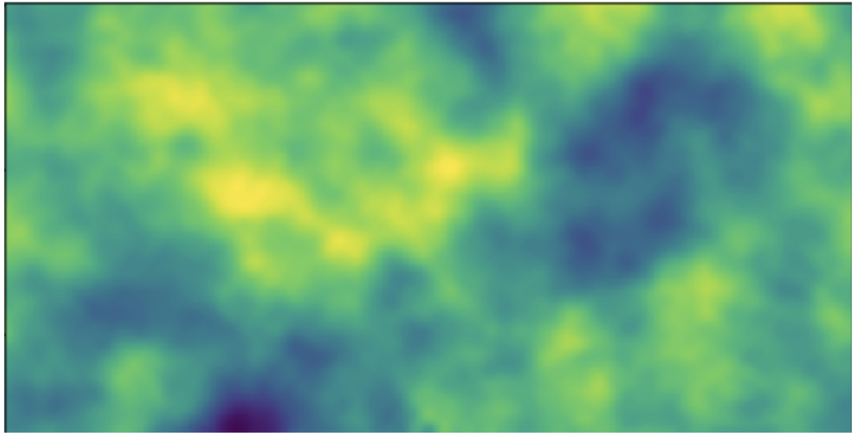
LSS sensitive to neutrino mass

Neutrinos are a pain to simulate

Implemented:

- symmetric momenta
- iterative scheme for ICs
- numerical artifact removal

Argonne NATIONAL LABORATORY



Deterministic Langevin Optimization

for Efficient Posterior Inference

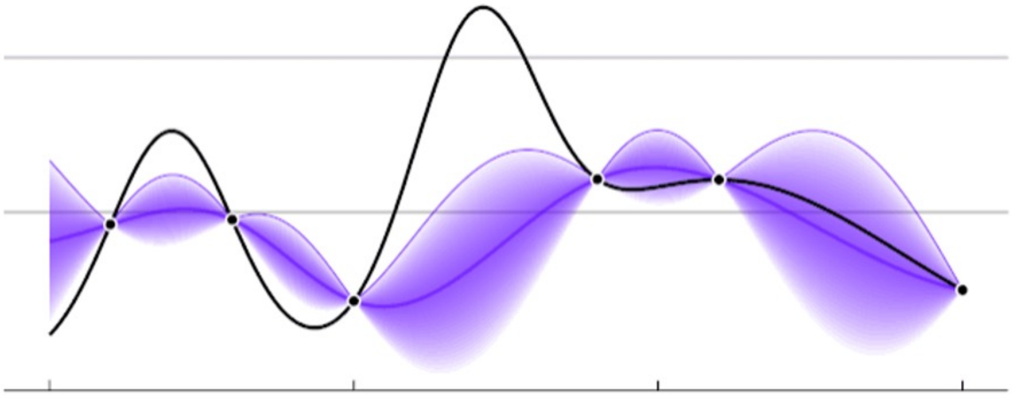
w/ Uroš Seljak

Bayesian Optimization

Strategy for expensive functions:

1. Think very carefully about choosing a domain point
2. Evaluate the function at the top candidate point
3. Build upon success or learn from mistakes

Iterate!

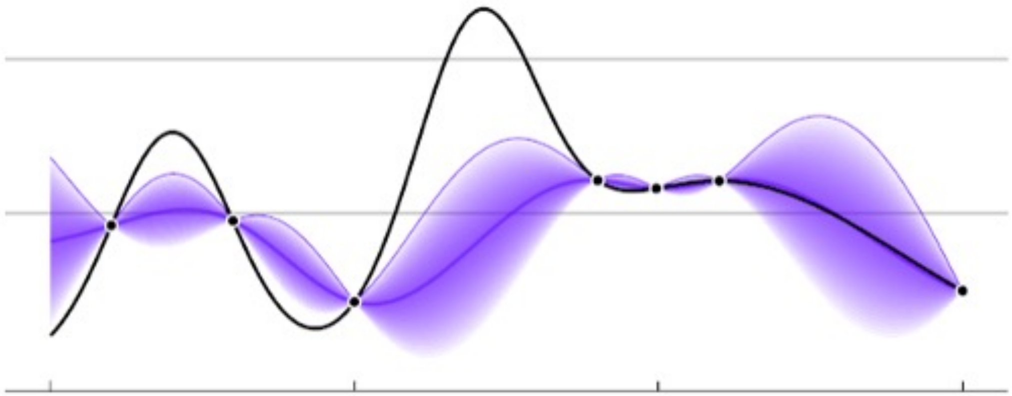


Bayesian Optimization

Strategy for expensive functions:

1. Think very carefully about choosing a domain point
2. Evaluate the function at the top candidate point
3. Build upon success or learn from mistakes

Iterate!

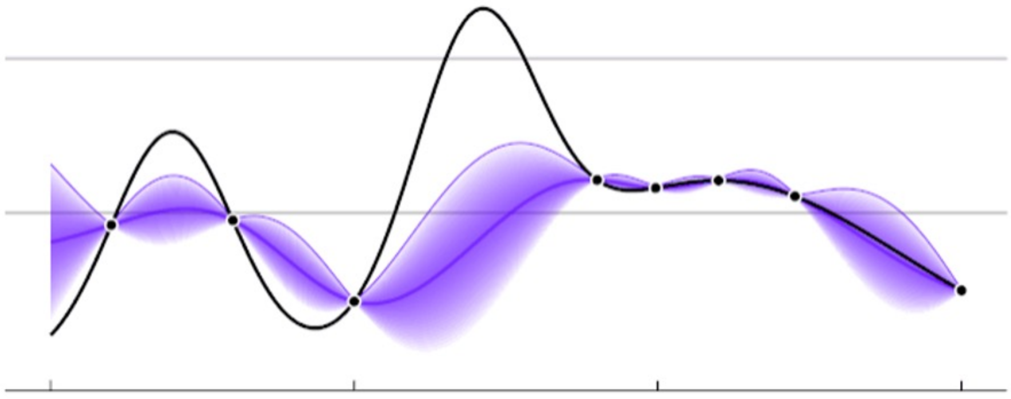


Bayesian Optimization

Strategy for expensive functions:

1. Think very carefully about choosing a domain point
2. Evaluate the function at the top candidate point
3. Build upon success or learn from mistakes

Iterate!

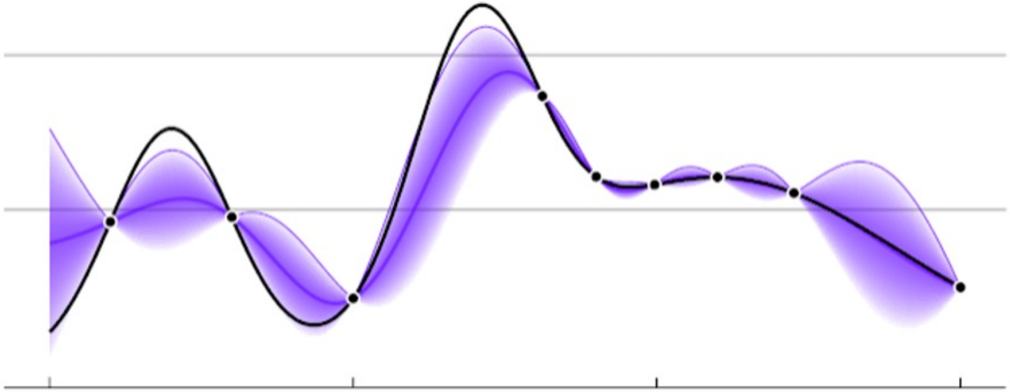


Bayesian Optimization

Strategy for expensive functions:

1. Think very carefully about choosing a domain point
2. Evaluate the function at the top candidate point
3. Build upon success or learn from mistakes

Iterate!

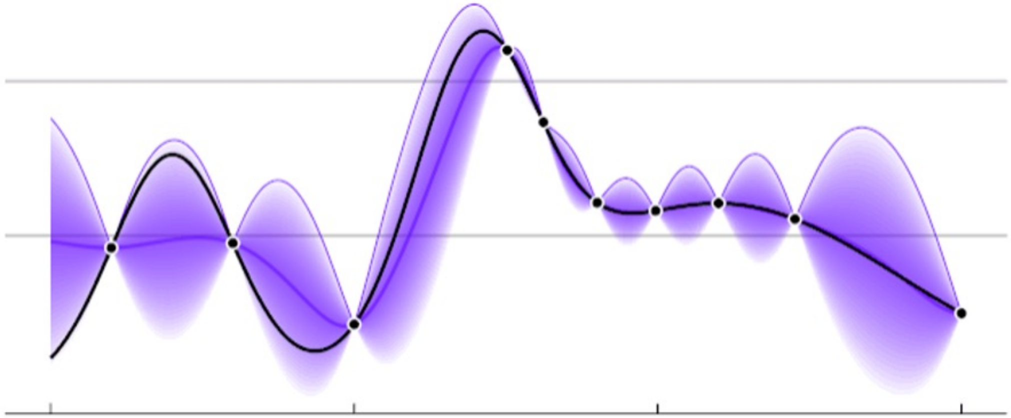


Bayesian Optimization

Strategy for expensive functions:

1. Think very carefully about choosing a domain point
2. Evaluate the function at the top candidate point
3. Build upon success or learn from mistakes

Iterate!

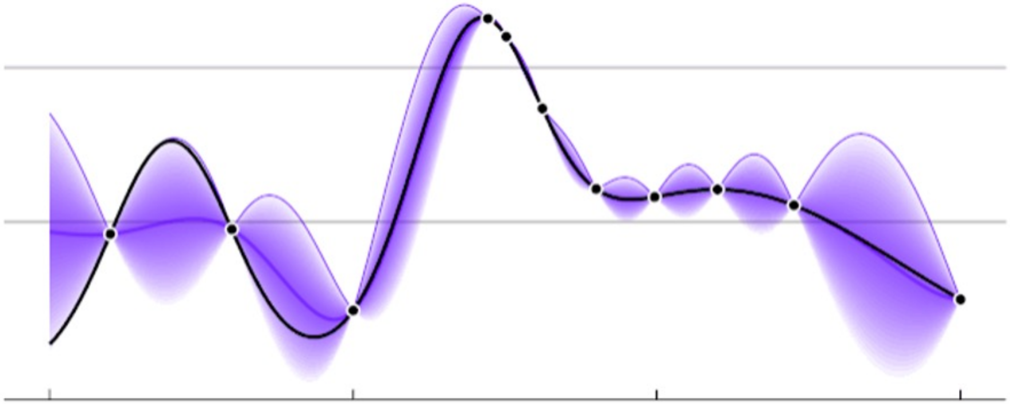


Bayesian Optimization

Strategy for expensive functions:

1. Think very carefully about choosing a domain point
2. Evaluate the function at the top candidate point
3. Build upon success or learn from mistakes

Iterate!

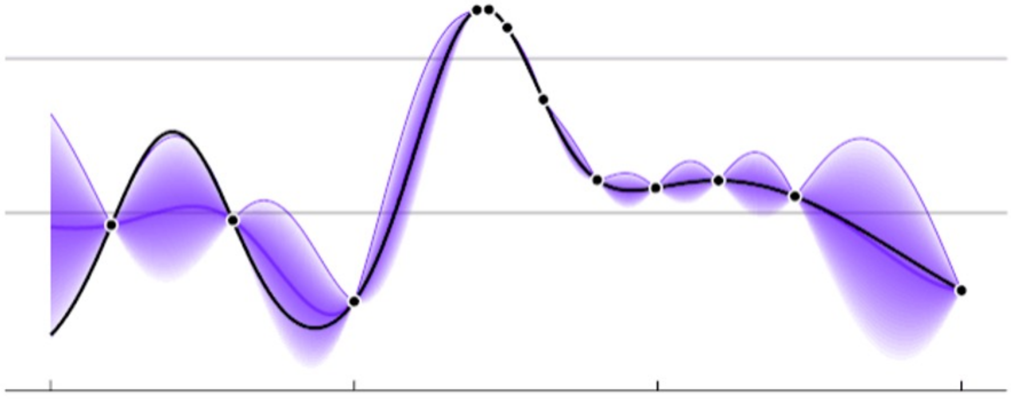


Bayesian Optimization

Strategy for expensive functions:

1. Think very carefully about choosing a domain point
2. Evaluate the function at the top candidate point
3. Build upon success or learn from mistakes

Iterate!

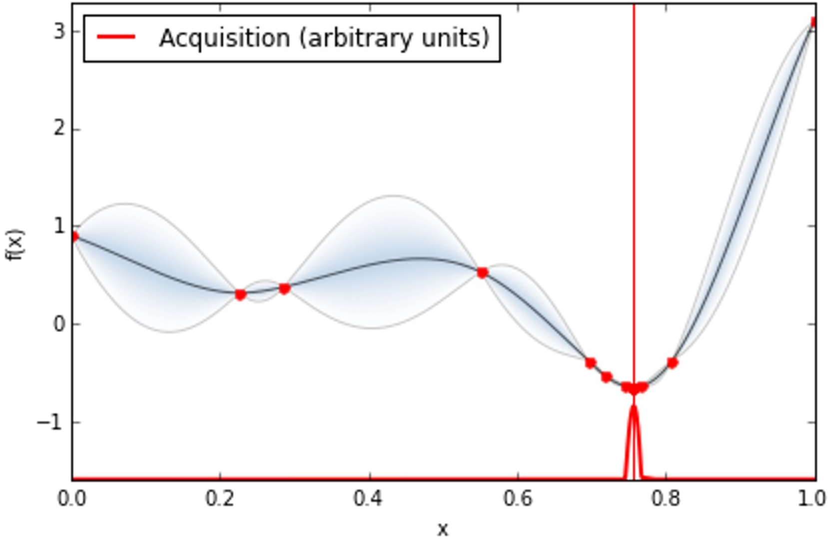


BO with Gaussian Processes

1. Surrogate model: *s(θ)* - approximates the function
2. Acquisition function (AF): weights exploration vs exploitation to select future points

Gaussian Processes (GPs) are non-parametric interpolators with uncertainty attached

Typically *s(θ)* is a Gaussian Process mean, and the *AF* uses GP uncertainty



DLO - Acquisition Function

GP comes with a “natural” uncertainty for the AF

We instead use a density estimate for uncertainty inspired by the deterministic Langevin equation:

$$\theta_{t+1} = \theta_t + v\epsilon = \theta_t + \frac{d}{d\theta}[\beta f(\theta_t) + V_t(\theta_t)(t)]\epsilon$$

DLO - Acquisition Function

GP comes with a "natural" uncertainty for the AF

We instead use a density estimate for uncertainty inspired by the deterministic Langevin equation:

stochastic update

$$\theta_{t+1} = \theta_t + \boxed{v\epsilon} = \theta_t + \boxed{\frac{d}{d\theta}[\beta f(\theta_t) + V_t(\theta_t)(t)]}\epsilon$$

parameter vector
"particle motion"

Model the deterministic piece *v* as the gradient of a scalar potential with two pieces

DLO - Acquisition Function

GP comes with a “natural” uncertainty for the AF

We instead use a density estimate for uncertainty inspired by the deterministic Langevin equation:

$$\theta_{t+1} = \theta_t + v\epsilon = \theta_t + \frac{d}{d\theta}[\beta f(\theta_t) + V_t(\theta_t)(t)]\epsilon$$

$$\theta_{t+1} = \arg\max_\theta[\beta f(\theta) + V_t(\theta)] = \arg\max_\theta \ln \frac{\exp(\beta f(\theta))}{q_t(\theta)}$$

Formulate as optimization problem!

(q = exp(-V))

DLO - Acquisition Function

GP comes with a “natural” uncertainty for the AF

We instead use a density estimate for uncertainty inspired by the deterministic Langevin equation:

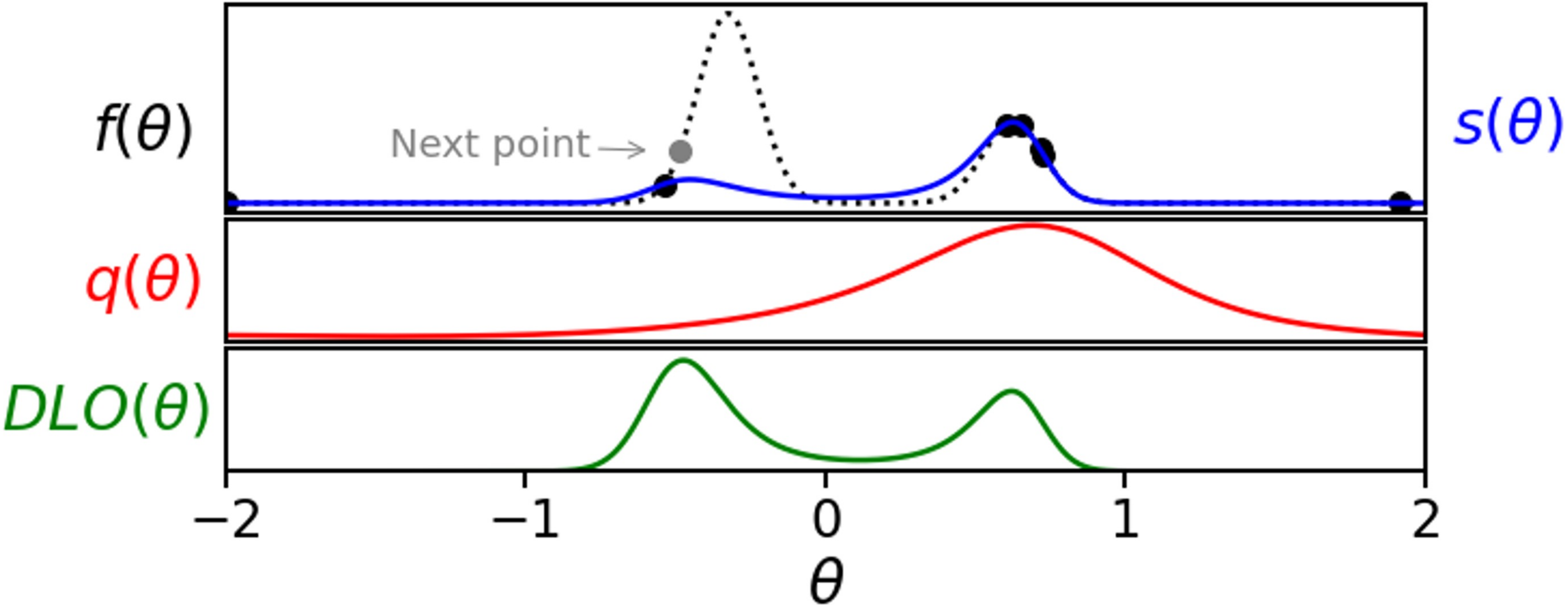
$$\theta_{t+1} = \theta_t + v\epsilon = \theta_t + \frac{d}{d\theta}[\beta f(\theta_t) + V_t(\theta_t)(t)]\epsilon$$

$$\theta_{t+1} = \arg\max_{\theta}[\beta \boxed{f(\theta)} + \boxed{V_t(\theta)}] = \arg\max_{\theta} \ln \frac{\exp(\beta \boxed{f(\theta)})}{\boxed{q_t(\theta)}}$$

target

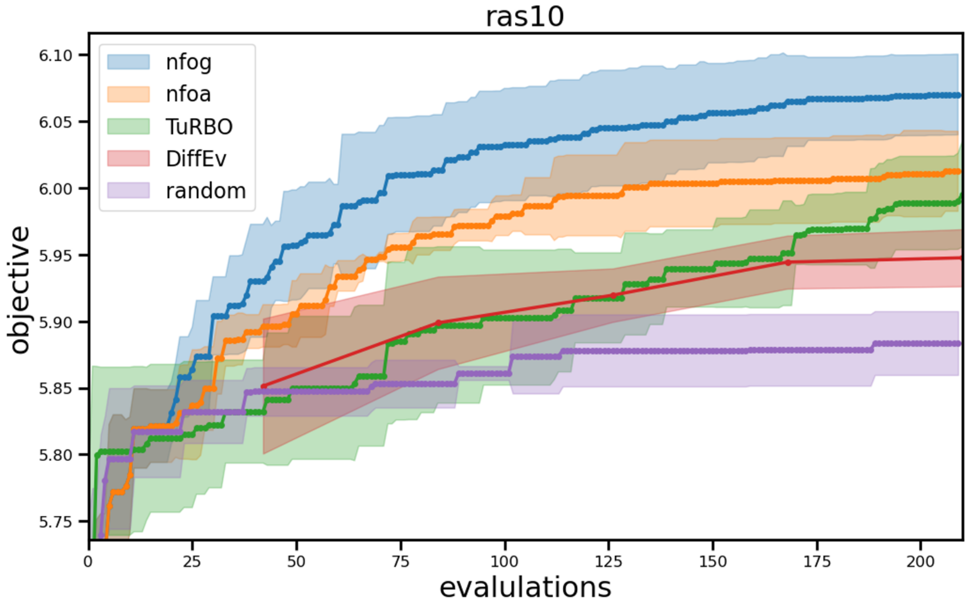
density estimate

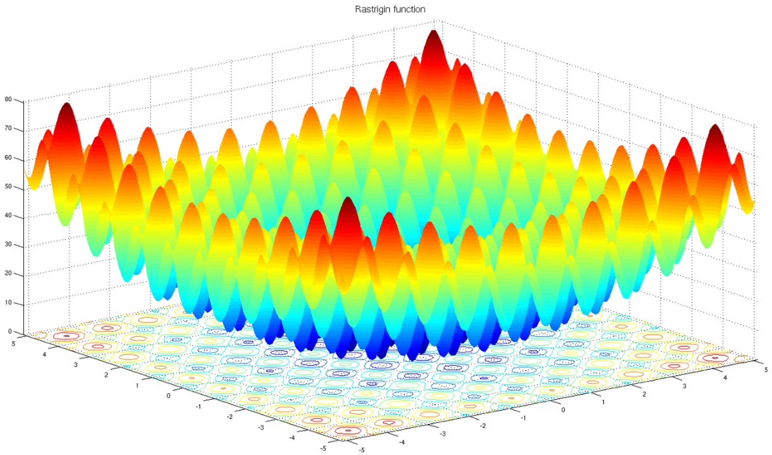
DLO - almost Bayesian Optimization



DLO Results: Test Functions

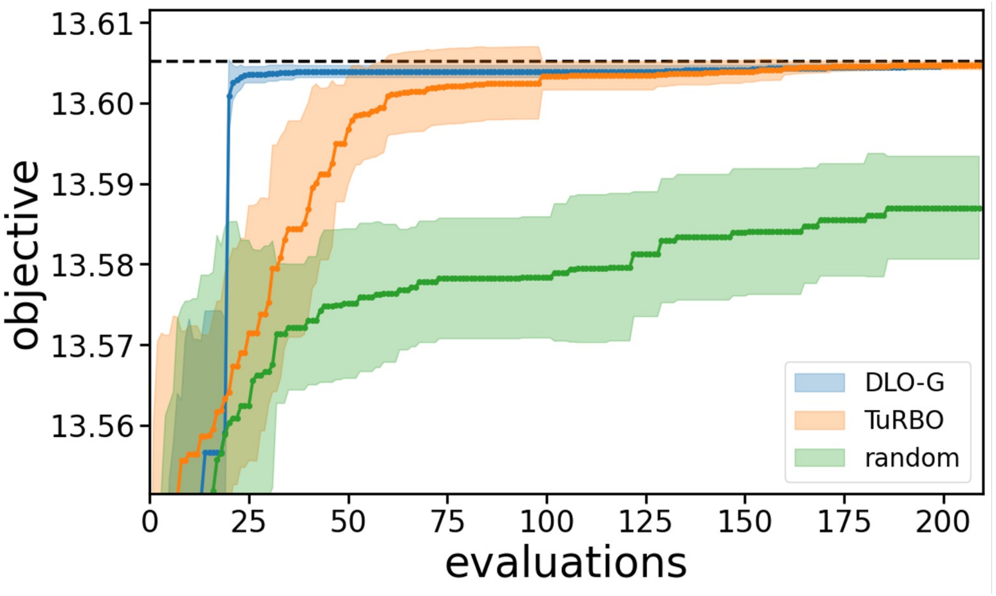
Synthetic (**Rastrigin**) & posterior (Rosenbrock “banana”) objectives in 10d

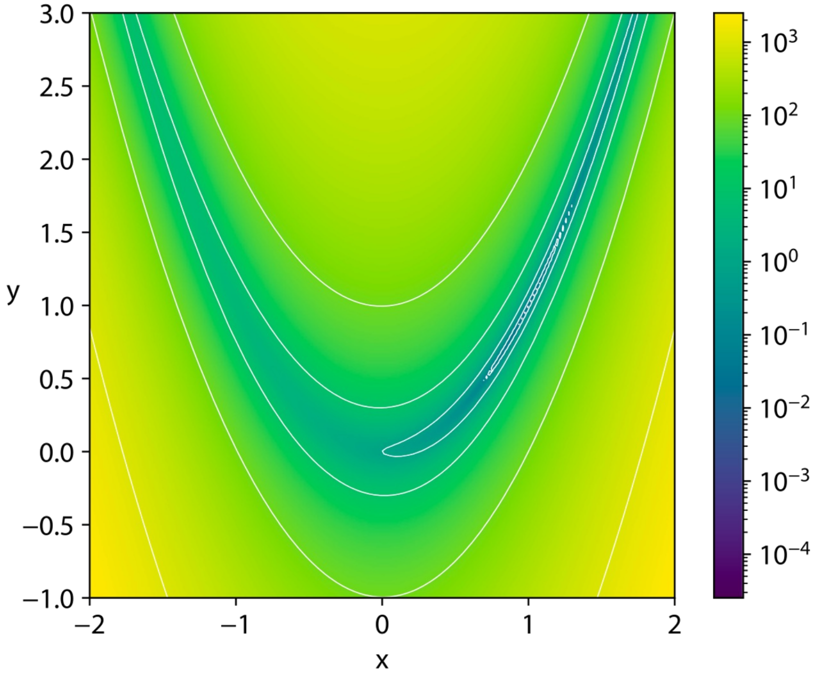




DLO Results: Test Functions

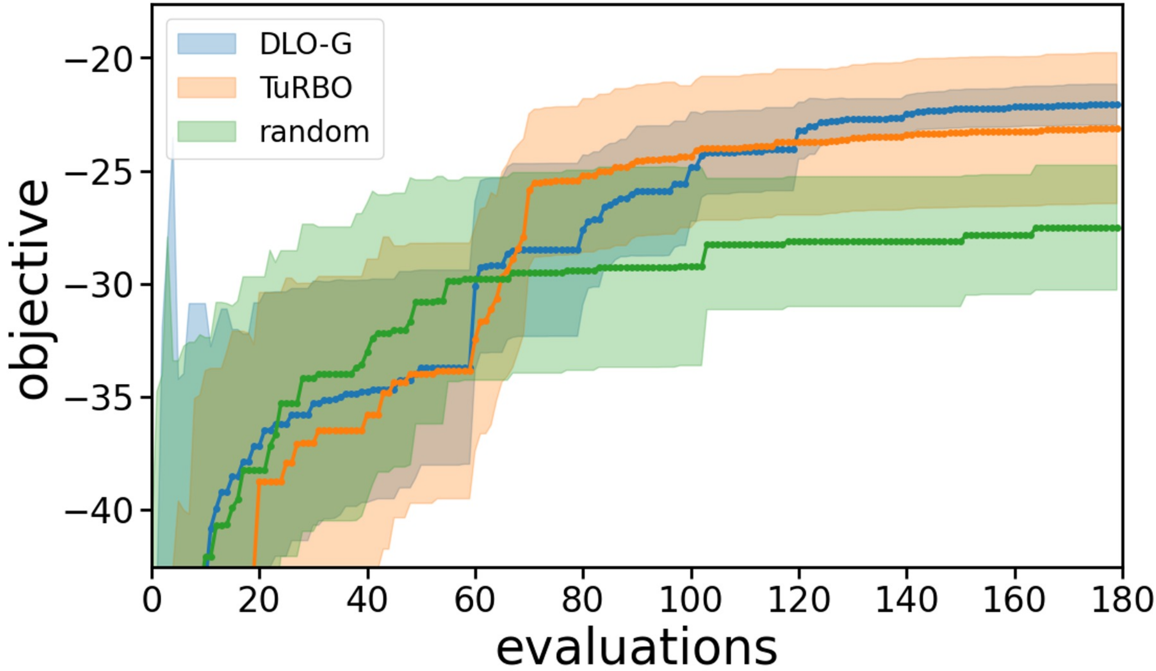
Synthetic (Rastrigin) & posterior (**Rosenbrock** “banana”) objectives in 10d





DLO Results: Applied example

Cosmology application: Luminous Red Galaxy clustering



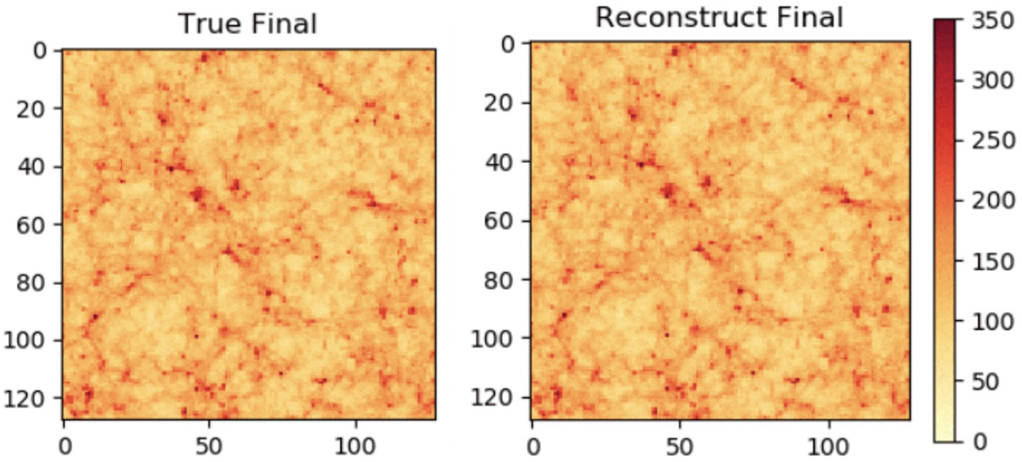
Also competitive for
ML hyperparameter optimization

Possible progress at the field-level

Field has all the information - no loss in compression to a summary statistic

Potentially easier than using summaries, but requires running simulations to obtain initial condition phases

May help probe details of inflation (WIP)!



Modi++20

Conclusions

Cosmological inference is expensive

Can make more efficient use of HPC resources by:

- Using faster/differentiable models
- Efficient choice of posterior evaluations

Progress on both fronts!