

Cosmological Emulation based on the Goku Simulations in a 10-Dimensional Parameter Space

PPC 2026

Yanhui Yang, University of California, Riverside

Yanhui
Yang



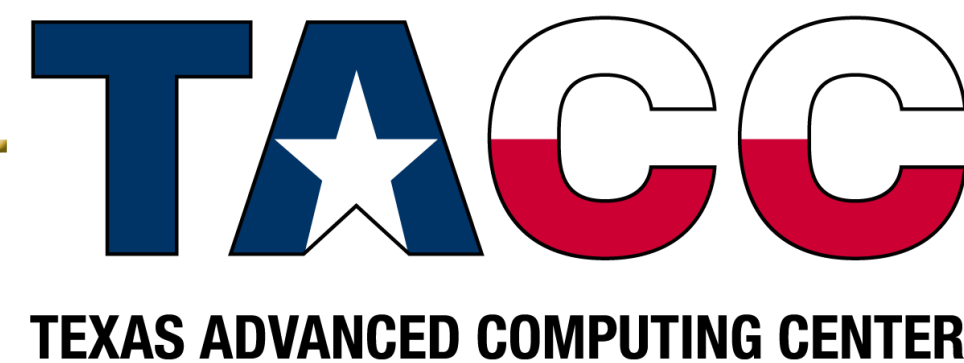
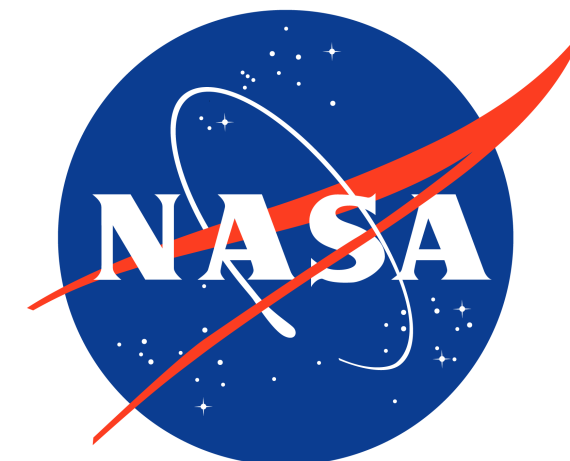
Simeon
Bird



Ming-Feng
Ho



Mahdi (Sum)
Qezlou

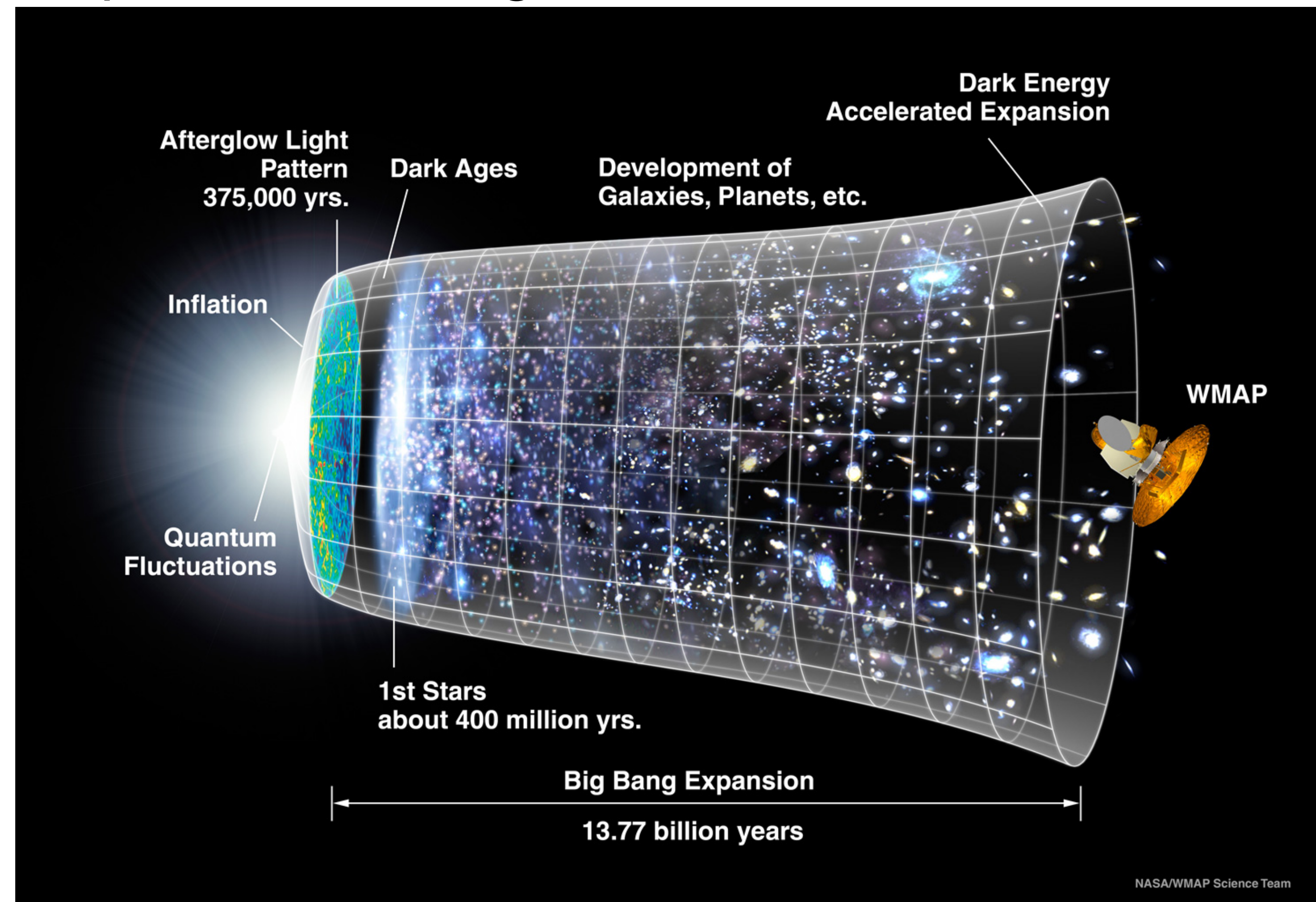


Contents

- **Background**
- **Cosmological Simulations: Goku** (Yang+2025, PRD, 2501.06296)
- **Multifidelity Emulation Framework: T2N-MusE** (Yang+2026a, PRD, 2507.07184)
- **Neural Network Emulator: GokuNEmu** (Yang+2026b, PRL, 2507.07177)
- **Future work**
- **Summary**

Background

- Big questions: How has the Universe evolved over cosmic time? What are the fundamental components driving this evolution?



NASA/WMAP

Background

- The standard cosmological model: Λ Cold Dark Matter (Λ CDM)
- Big Bang picture of cosmic evolution
- Parameters:
 - Ω_m : matter density parameter
 - Ω_b : baryon density parameter
 - h : expansion rate (Hubble parameter)
 - | | | |
|--|---|----------------------------------|
| <ul style="list-style-type: none">• A_s (amplitude)• n_s (tilt) | $\left. \vphantom{\begin{matrix} A_s \\ n_s \end{matrix}} \right\}$ | primordial density perturbations |
|--|---|----------------------------------|

Background

- LambdaCDM has been very successful, but is it complete?
- Open questions (extensions):
 - dark energy: E.o.S parameter is $w = -1$ or time-dependent (DESI 2025)? $\rightarrow w(a) = w_0 + w_a(1 - a)$
 - neutrinos have mass; what is the mass scale? $\rightarrow \Sigma m_\nu$
 - additional relativistic species (effective number of neutrinos)? $\rightarrow N_{\text{eff}}$
 - primordial perturbations: power spectrum exactly a power law? $\rightarrow \alpha_s = \frac{d n_s}{d \ln k}$
(tilt of tilt)

Background

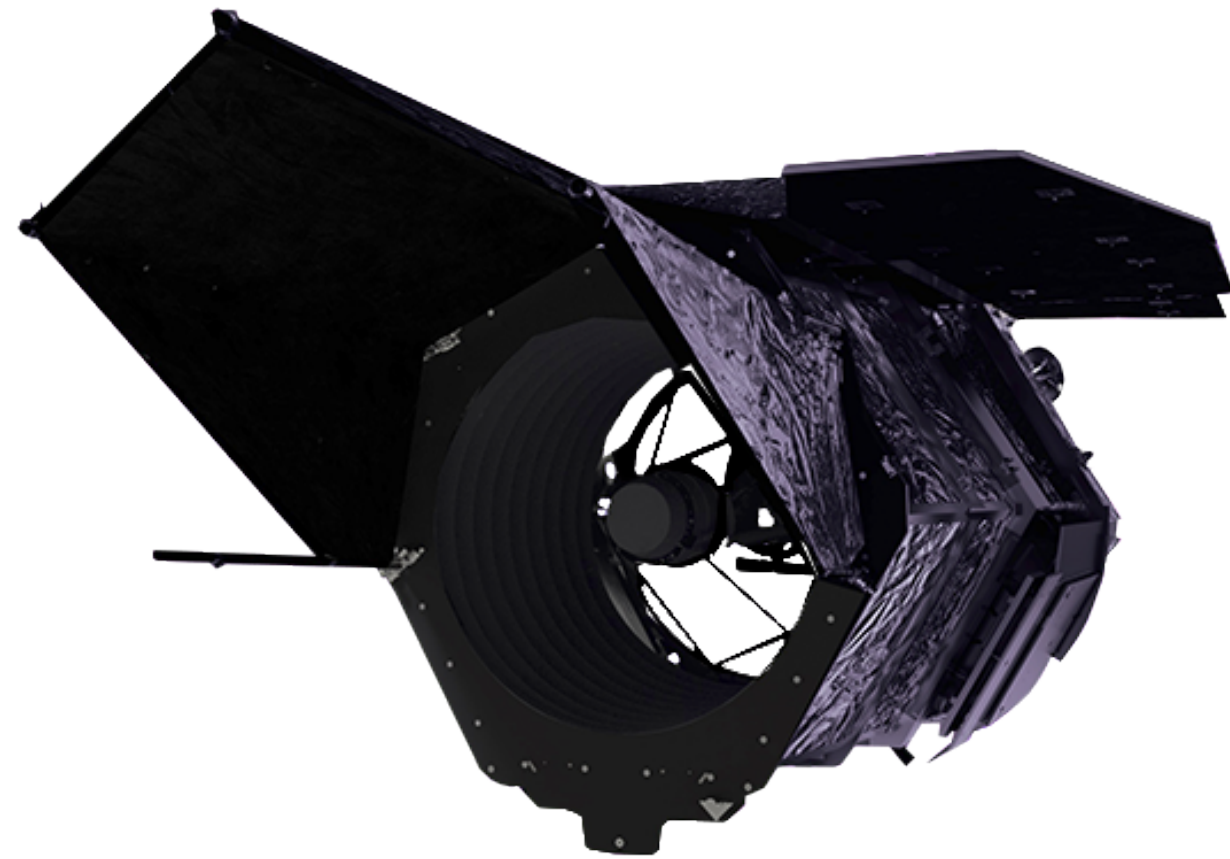
- LambdaCDM has been very successful, but is it complete?
- Open questions (extensions):
 - dark energy: E.o.S parameter is $w = -1$ or time-dependent (DESI 2025)? $\rightarrow w(a) = w_0 + w_a(1 - a)$
 - neutrinos have mass; what is the mass scale? $\rightarrow \Sigma m_\nu$
 - additional relativistic species (effective number of neutrinos)? $\rightarrow N_{\text{eff}}$
 - primordial perturbations: power spectrum exactly a power law? $\rightarrow \alpha_s = \frac{d n_s}{d \ln k}$
(tilt of tilt)



5 extra cosmological parameters: w_0 w_a Σm_ν N_{eff} α_s

Background

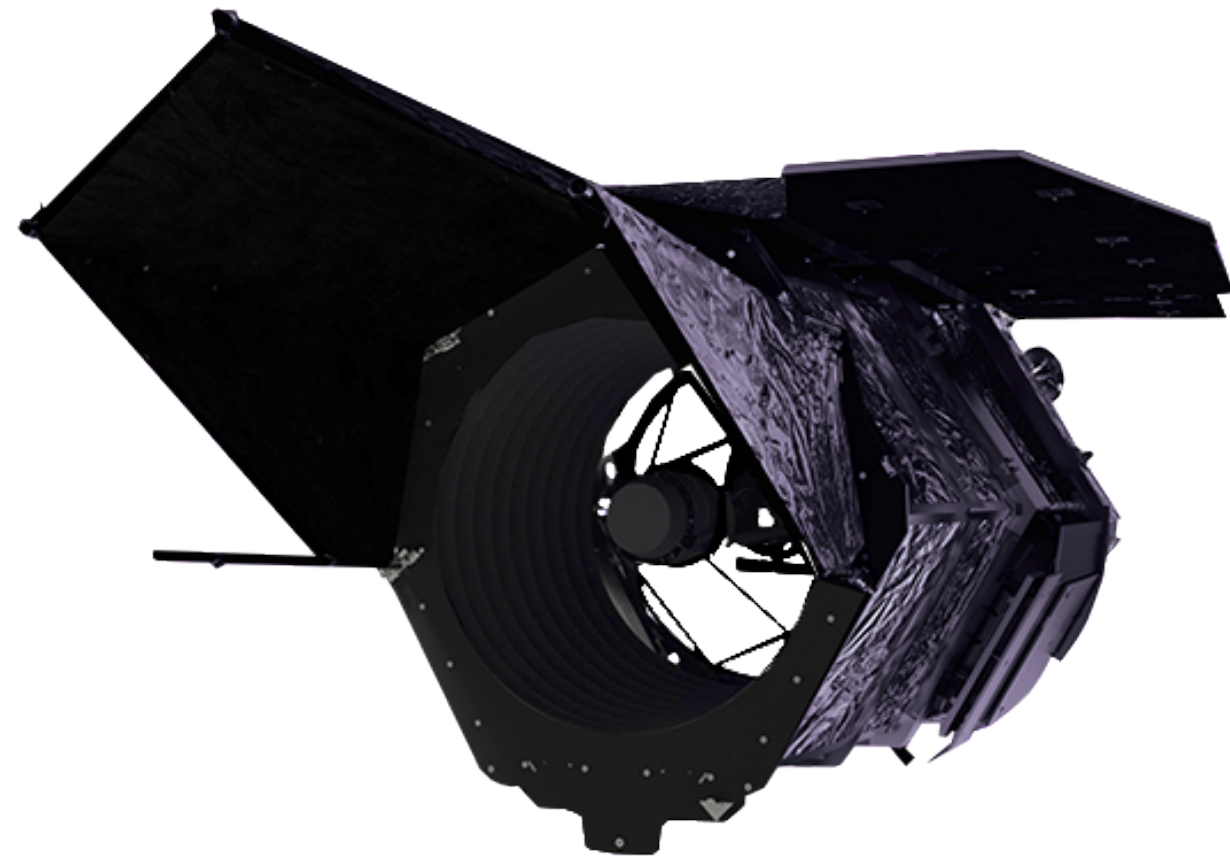
- Cosmological surveys: the Roman Space Telescope, Euclid, LSST, CSST, DESI, DES, HSC, etc.



Roman Space Telescope

Background

- Cosmological surveys: the Roman Space Telescope, Euclid, LSST, CSST, DESI, DES, HSC, etc.



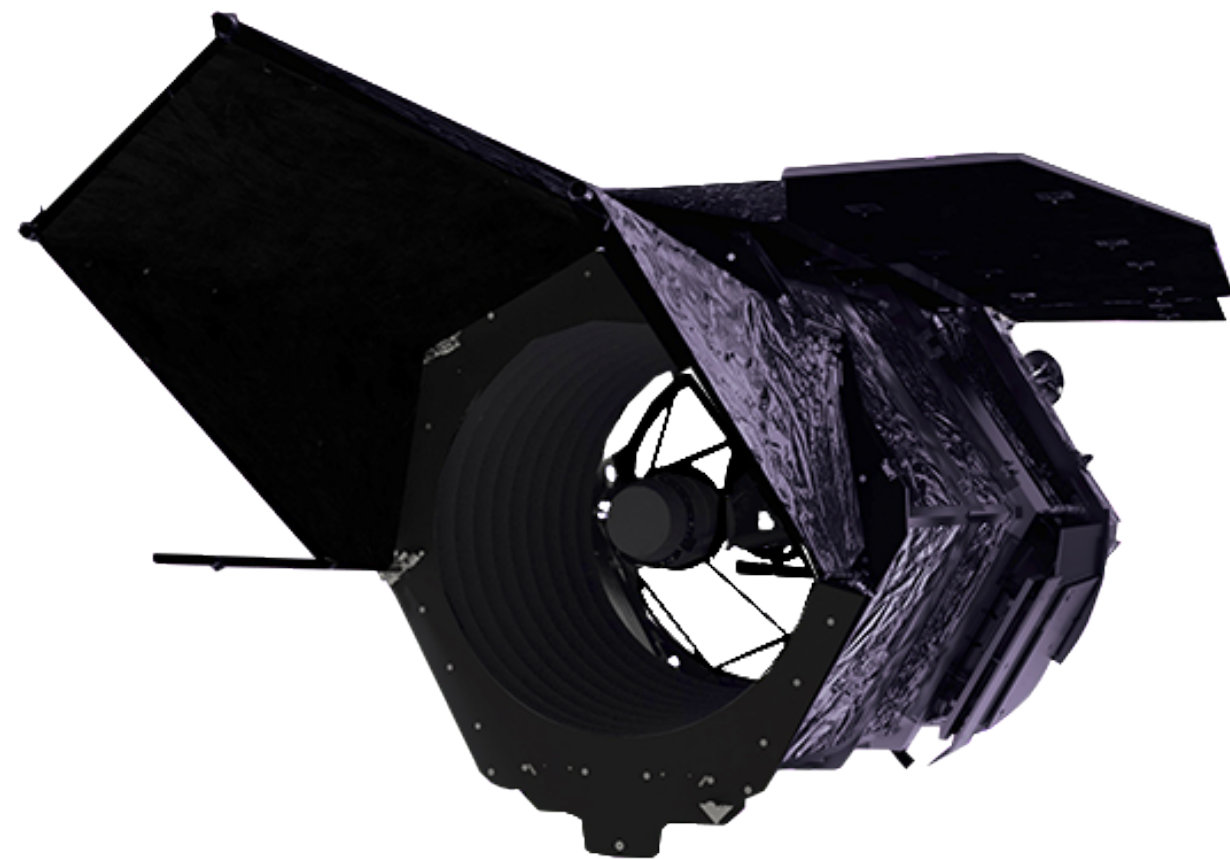
Roman Space Telescope



$\sim 10^9$ galaxies will be observed

Background

- Cosmological surveys: the Roman Space Telescope, Euclid, LSST, CSST, DESI, DES, HSC, etc.



Roman Space Telescope



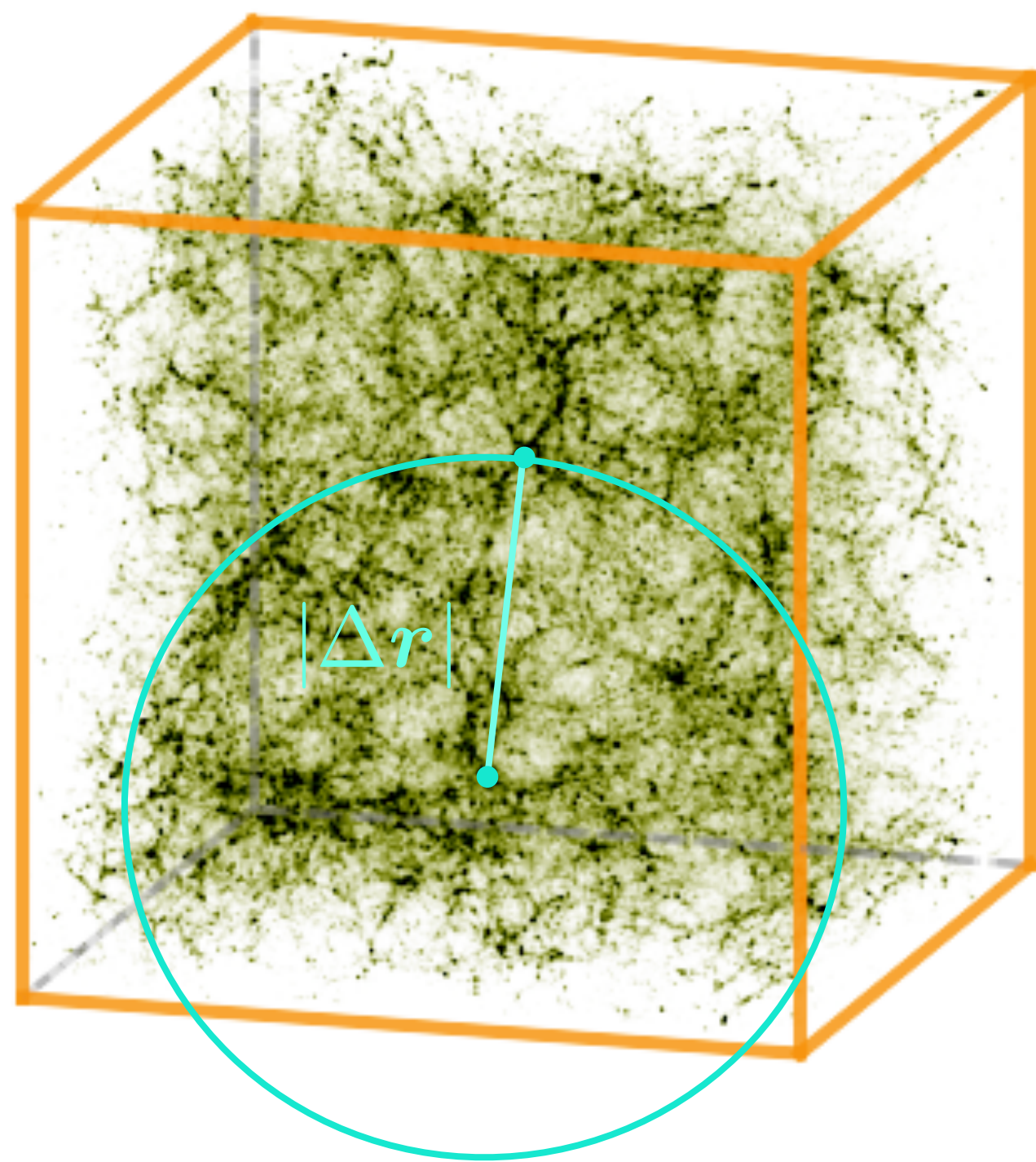
$\sim 10^9$ galaxies will be observed



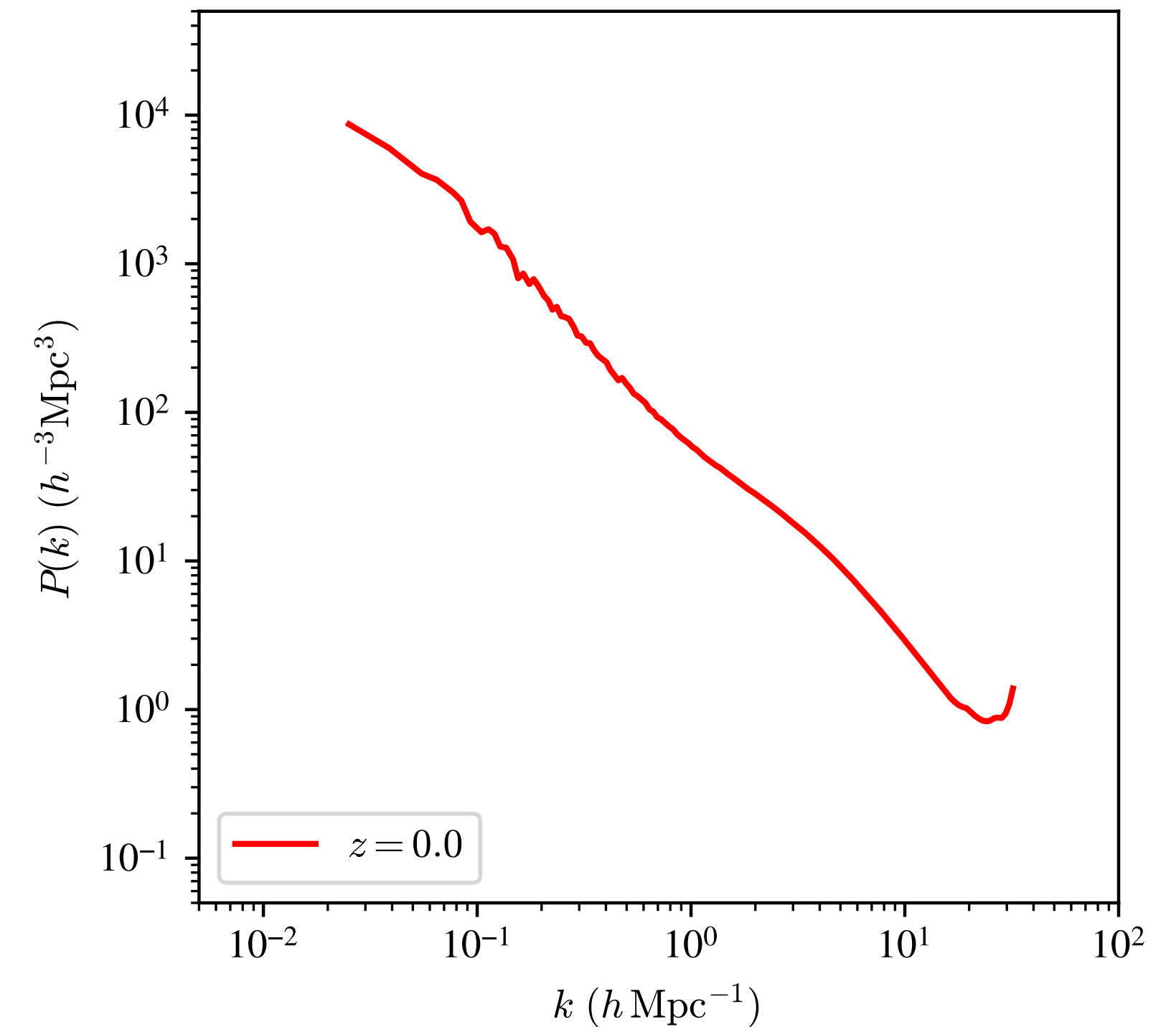
Summary statistics of large-scale matter distribution, e.g., the matter power spectrum

Background

- The matter power spectrum: how "clumpy" the Universe is on different scales (Fourier transform of the two-point correlation function of the overdensity field)



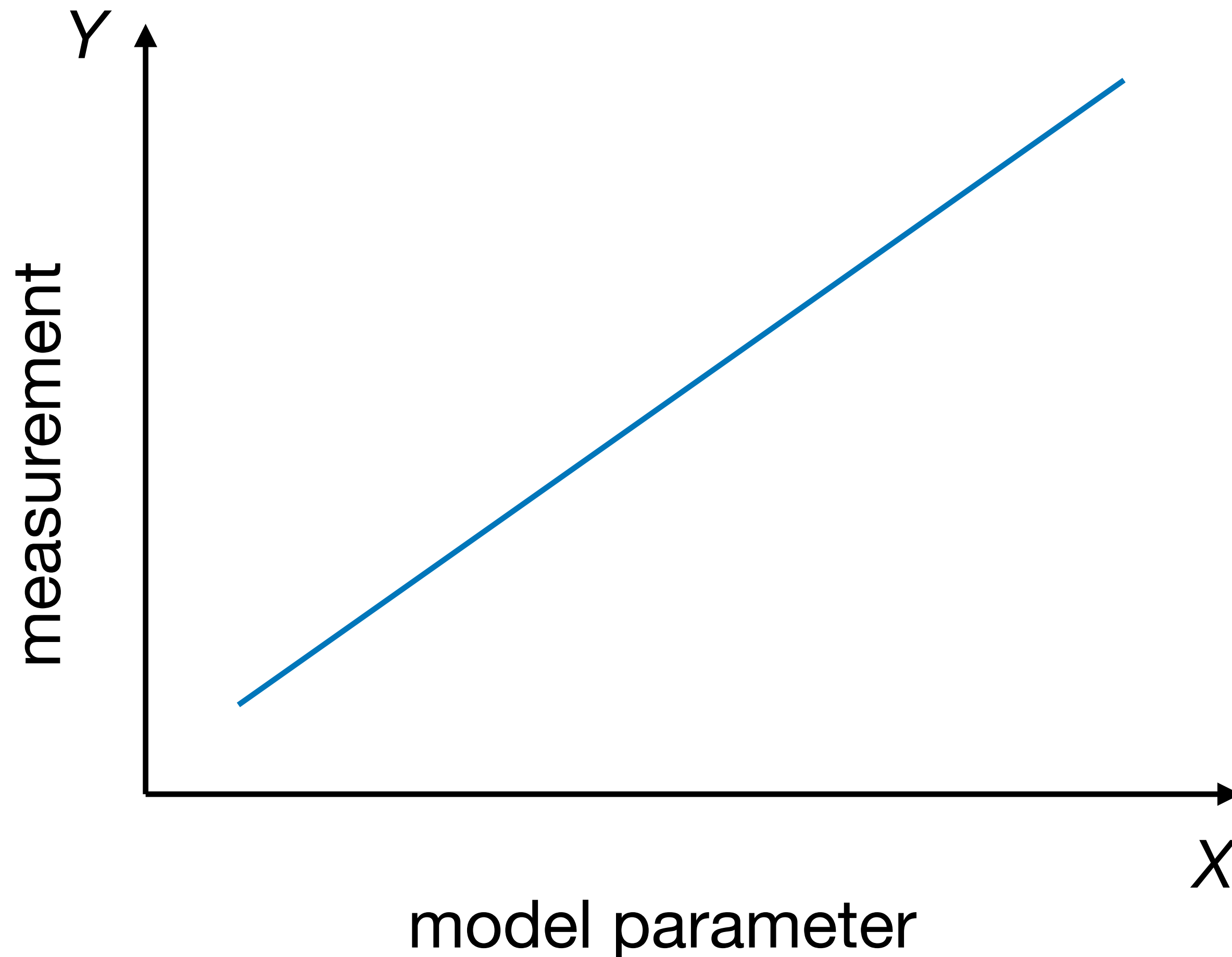
$$k = \frac{2\pi}{|\Delta \mathbf{r}|}$$



Background

- Cosmological inference: Bayesian methods $P(\text{model}|\text{data}) = \frac{P(\text{data}|\text{model})P(\text{model})}{P(\text{data})}$

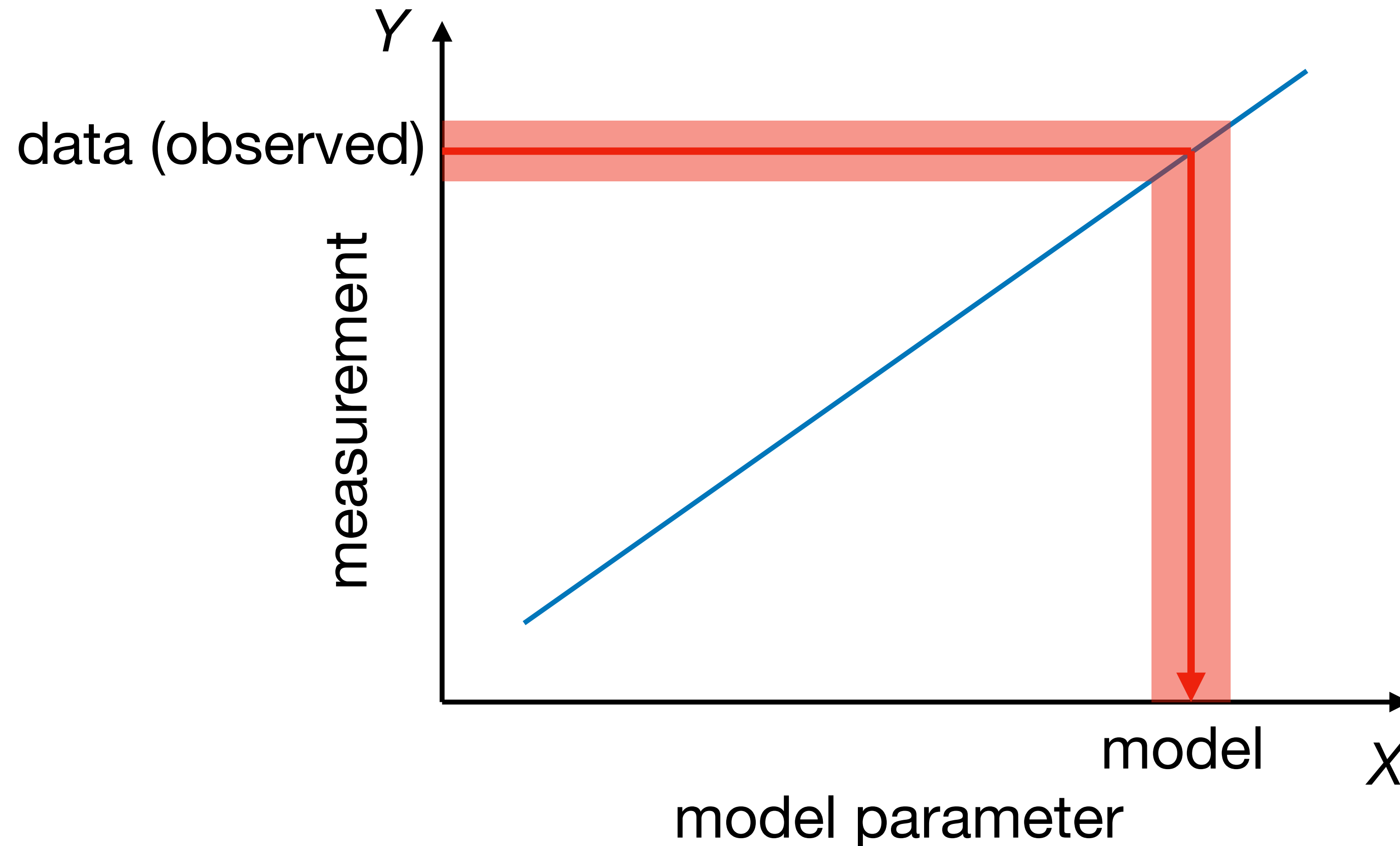
Basic idea: an analogy, suppose we know $Y=f(X)$



Background

- Cosmological inference: Bayesian methods $P(\text{model}|\text{data}) = \frac{P(\text{data}|\text{model})P(\text{model})}{P(\text{data})}$

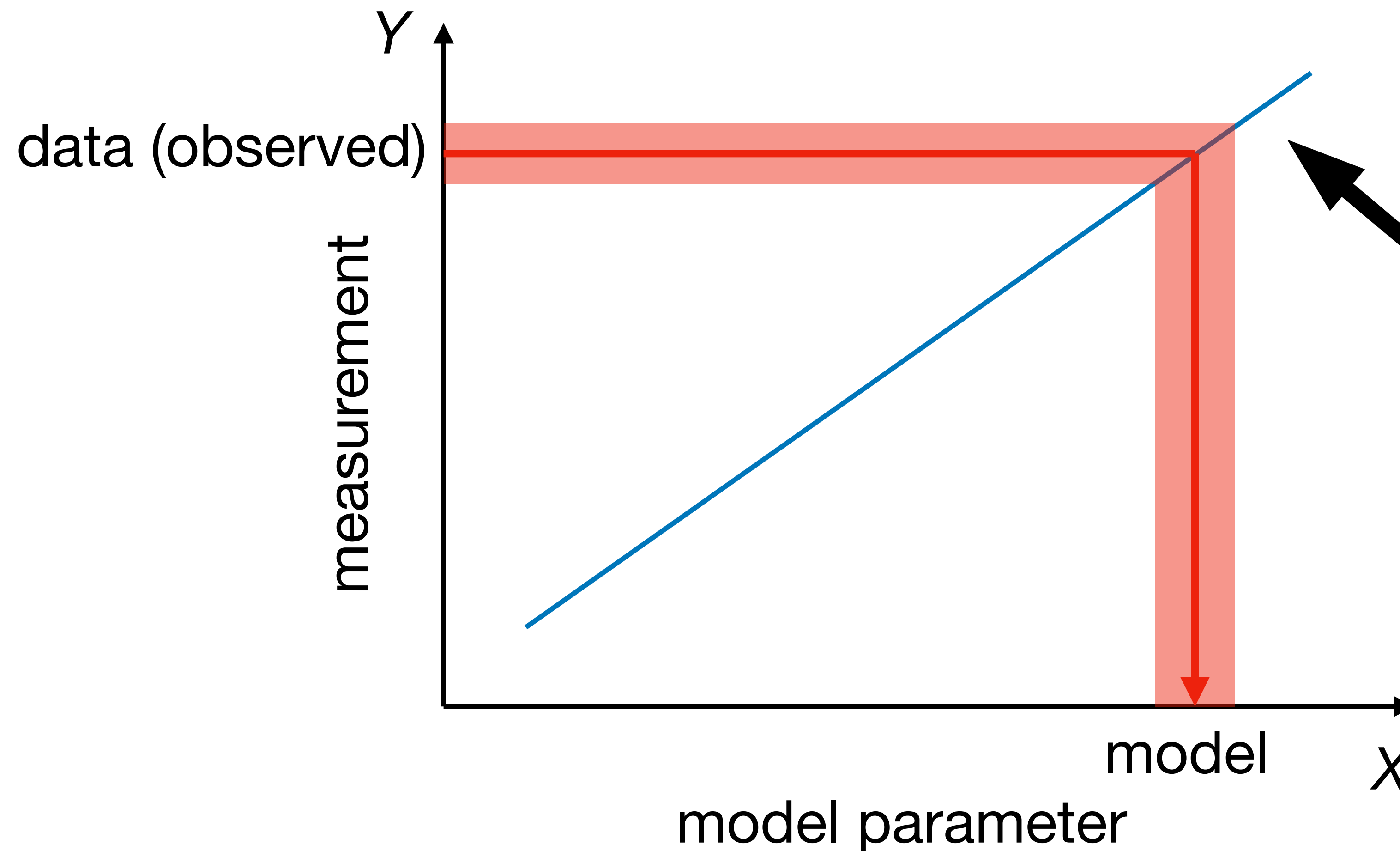
Basic idea: an analogy, suppose we know $Y=f(X)$



Background

- Cosmological inference: Bayesian methods $P(\text{model}|\text{data}) = \frac{P(\text{data}|\text{model})P(\text{model})}{P(\text{data})}$

Basic idea: an analogy, suppose we know $Y=f(X)$



In practice, X and Y can be vectors!

And f may be complex!

Focus

How do we get the dependence f ?

Background

by N-body simulations

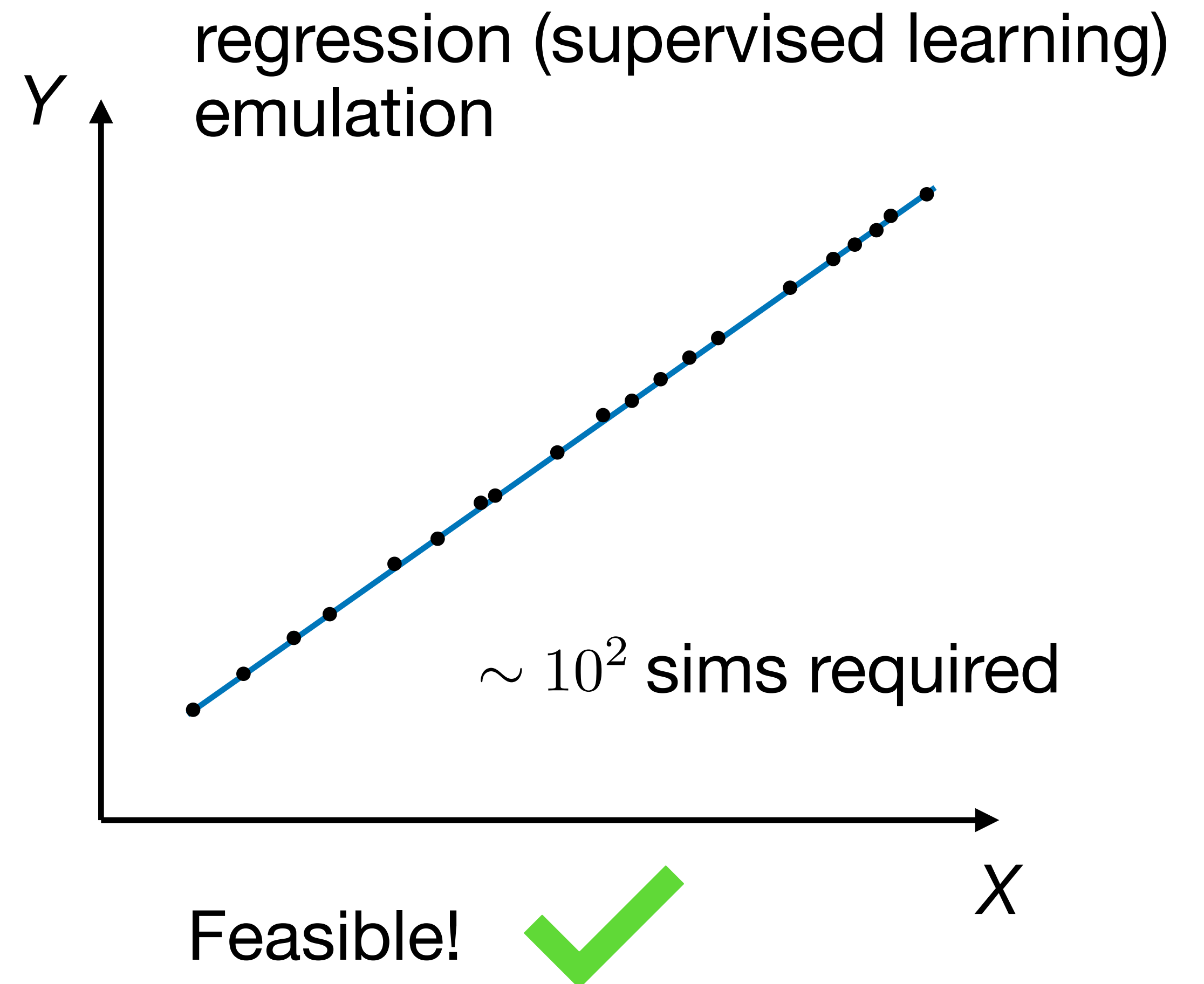
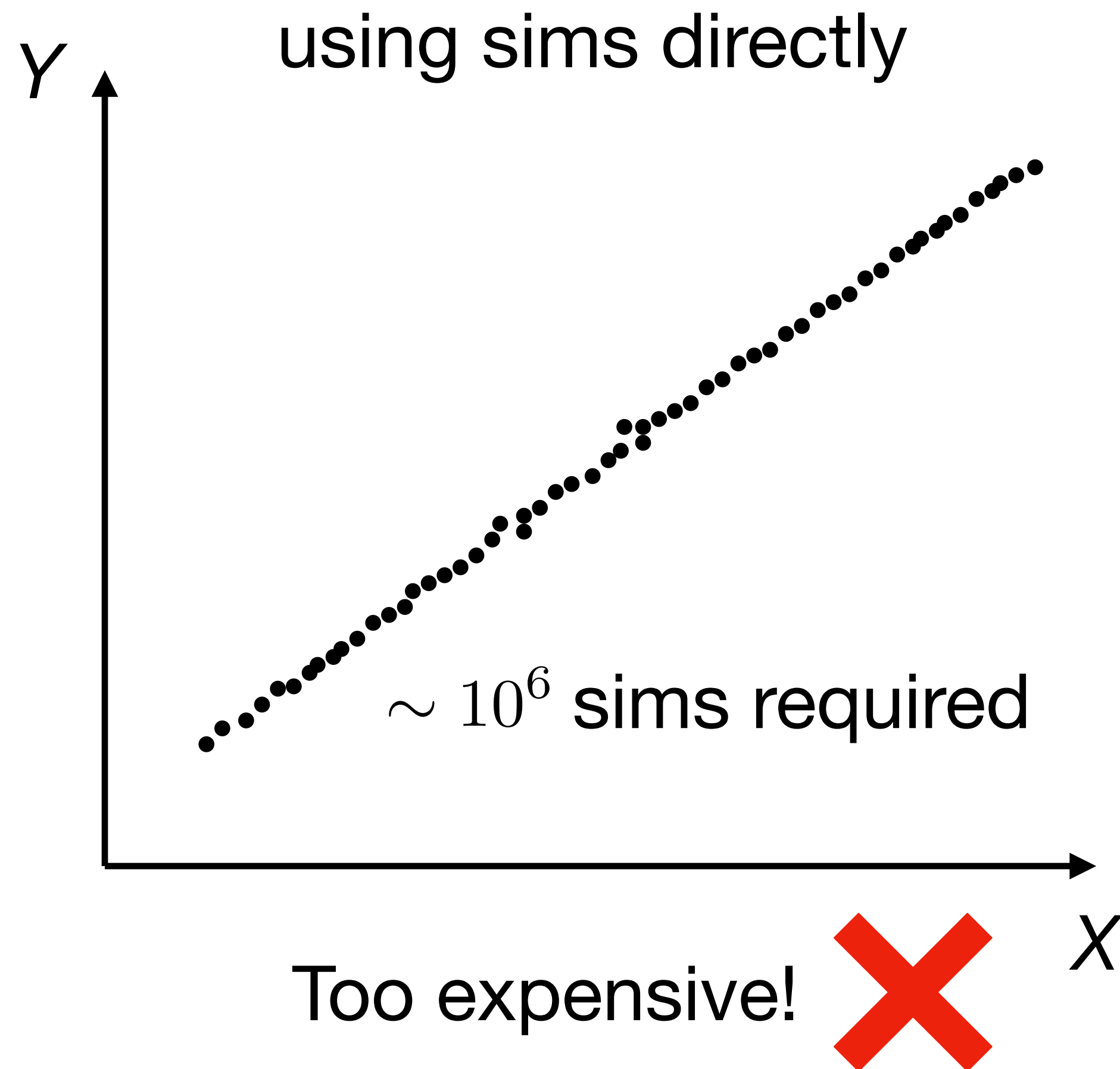
- Theoretical predictions: X (cosmological parameters) $\rightarrow$ Y (matter power)



Background

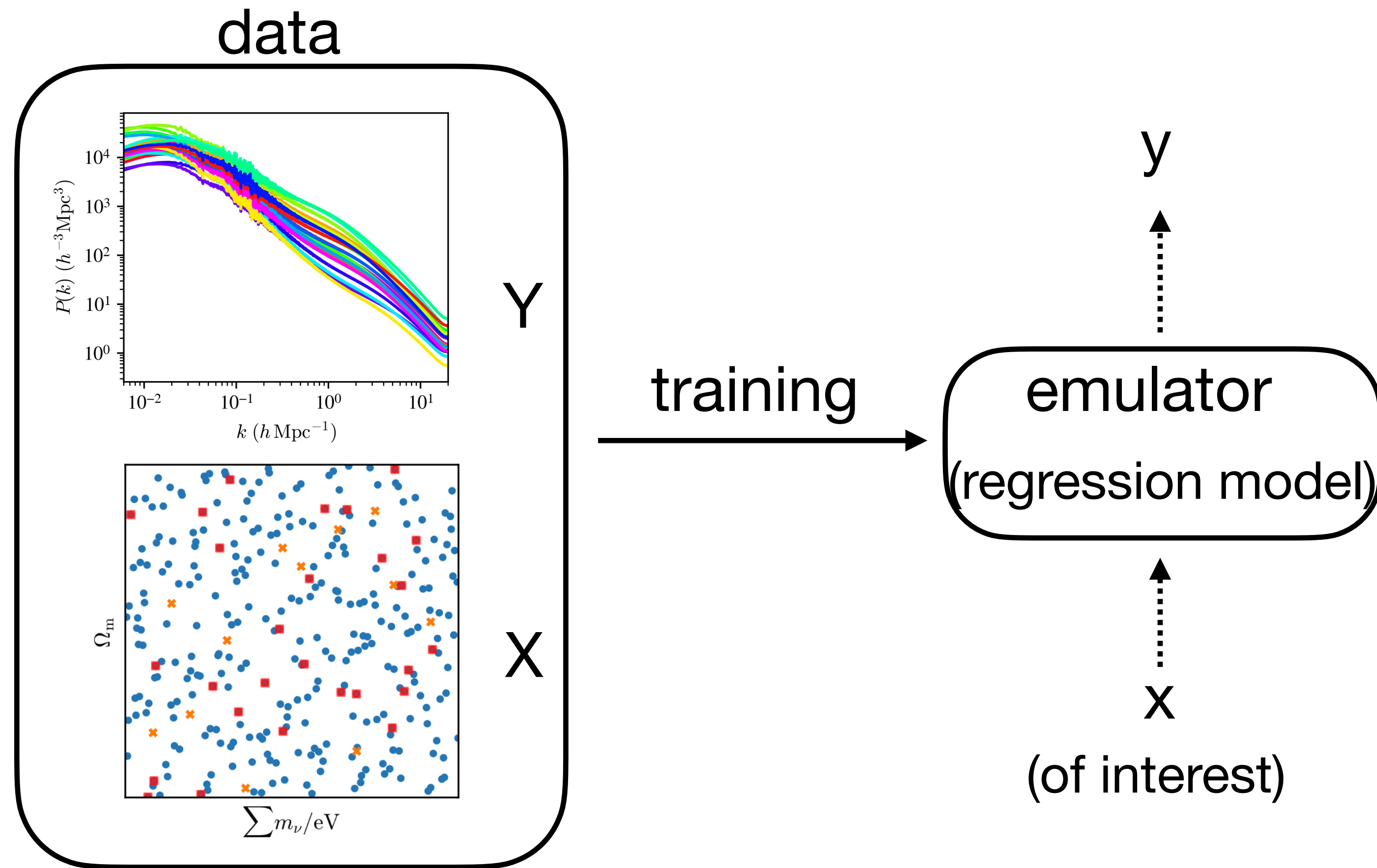
by N-body simulations

- Theoretical predictions: X (cosmological parameters) $\rightarrow$ Y (matter power)



Background

- Simulation-based emulation:

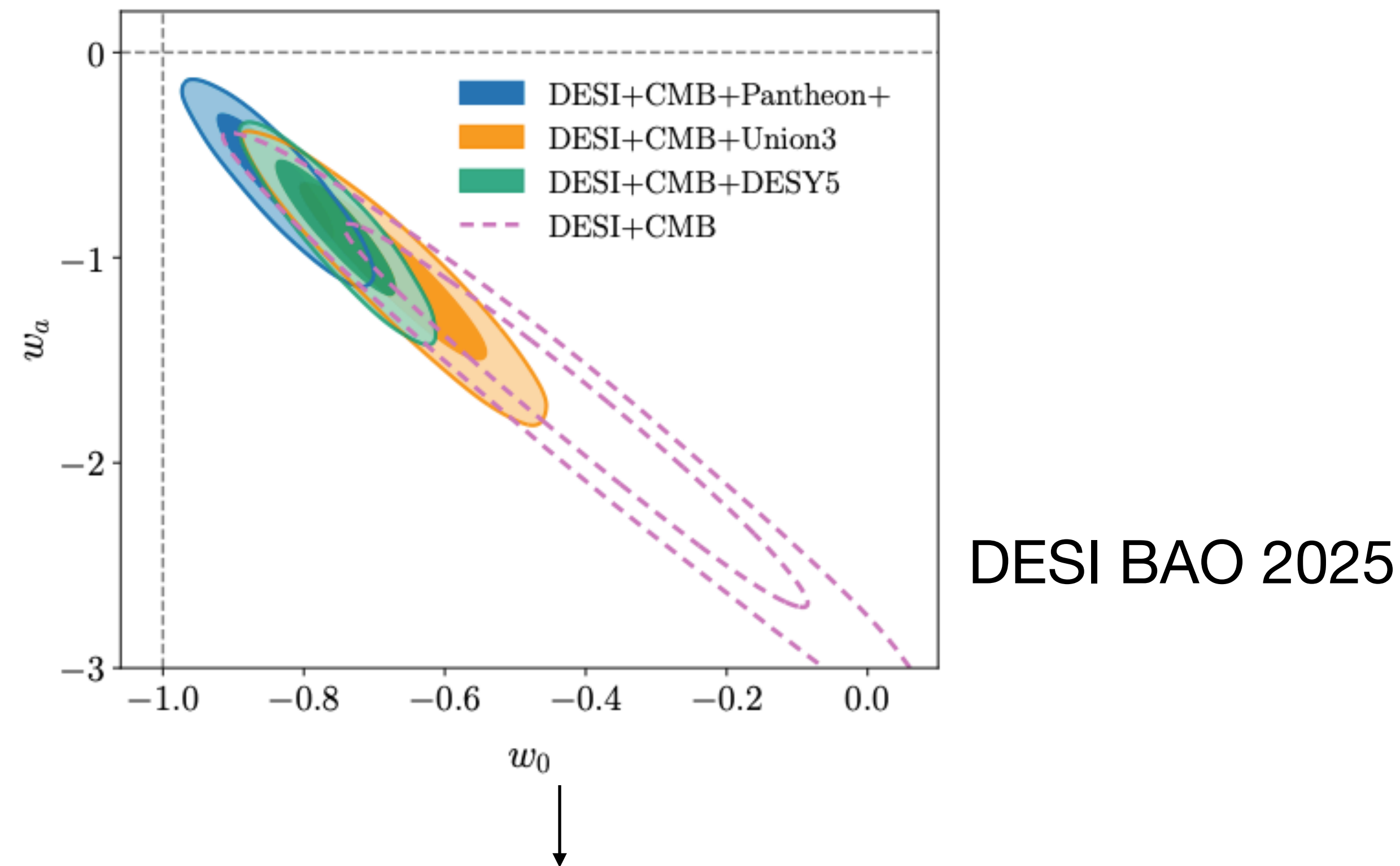


Background

- Emulators: percent-level accuracy, $\Lambda\text{CDM}+w_0w_a+\Sigma m_\nu$
 - EuclidEmulator2 (Euclid 2020): 8D, ~ 127 simulations with 3000^3 particles evolving in $1 (\text{Gpc}/h)^3$ volumes
 - CSST emulator (Chen+2025): 8D, 129 simulations, similar fidelity
- N_{eff} and α_s not included

Background

- Wider prior ranges for w_0 and w_a might be needed (considering constraints from DESI 2025)



We need new emulators (higher dimensions, broader priors)!

Background

- Our goal is difficult to achieve using traditional single-fidelity emulation techniques
- Traditional methods:
 - single-fidelity: simulations with the same accuracy (determined by box size and resolution)
 - computationally expensive for high-dimensional parameter space (a large number of simulations required)
 - e.g., EuclidEmulator2: 8D, ~ 127 simulations

Background

- Our goal is difficult to achieve using traditional single-fidelity emulation techniques
- Traditional methods:
 - single-fidelity: simulations with the same accuracy (determined by box size and resolution)
 - computationally expensive for high-dimensional parameter space (a large number of simulations required)
 - e.g., EuclidEmulator2: 8D, ~ 127 simulations
- We use MF-Box (Ho, Bird et al. 2023), a multifidelity emulation framework, to reduce the number of high-fidelity simulations

Cosmological Simulations: Goku (Yang, Bird et al. 2025)

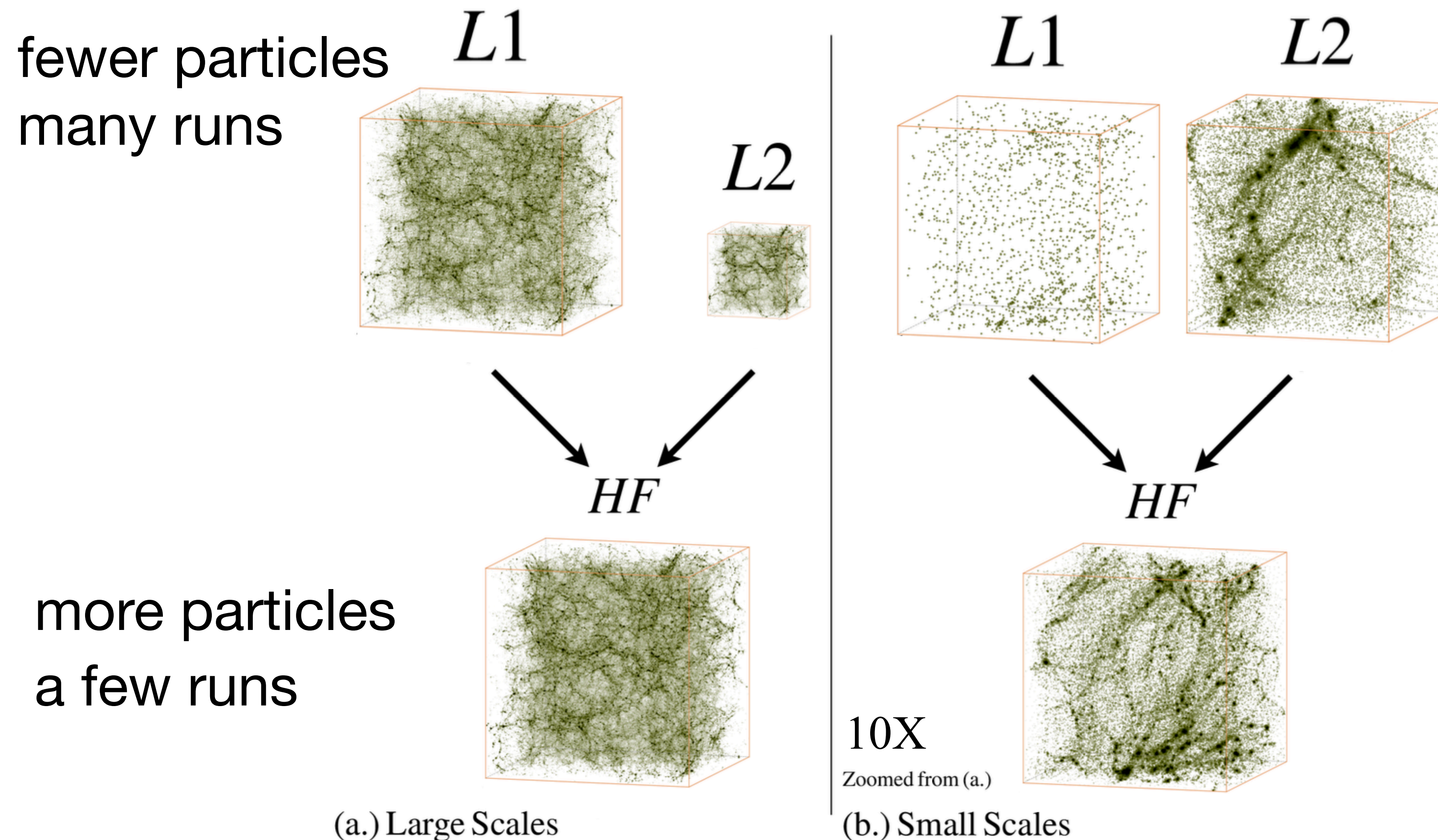
Methods

- Multifidelity emulation: combines high-fidelity data with low-fidelity data in training
 - a low-fidelity (fewer particles) simulation is much less expensive than a high-fidelity one, while still providing a fairly good prediction on certain scales
 - low-fidelity predictions can be corrected using a few high-fidelity samples

Cosmological Simulations: Goku (Yang, Bird et al. 2025)

Methods

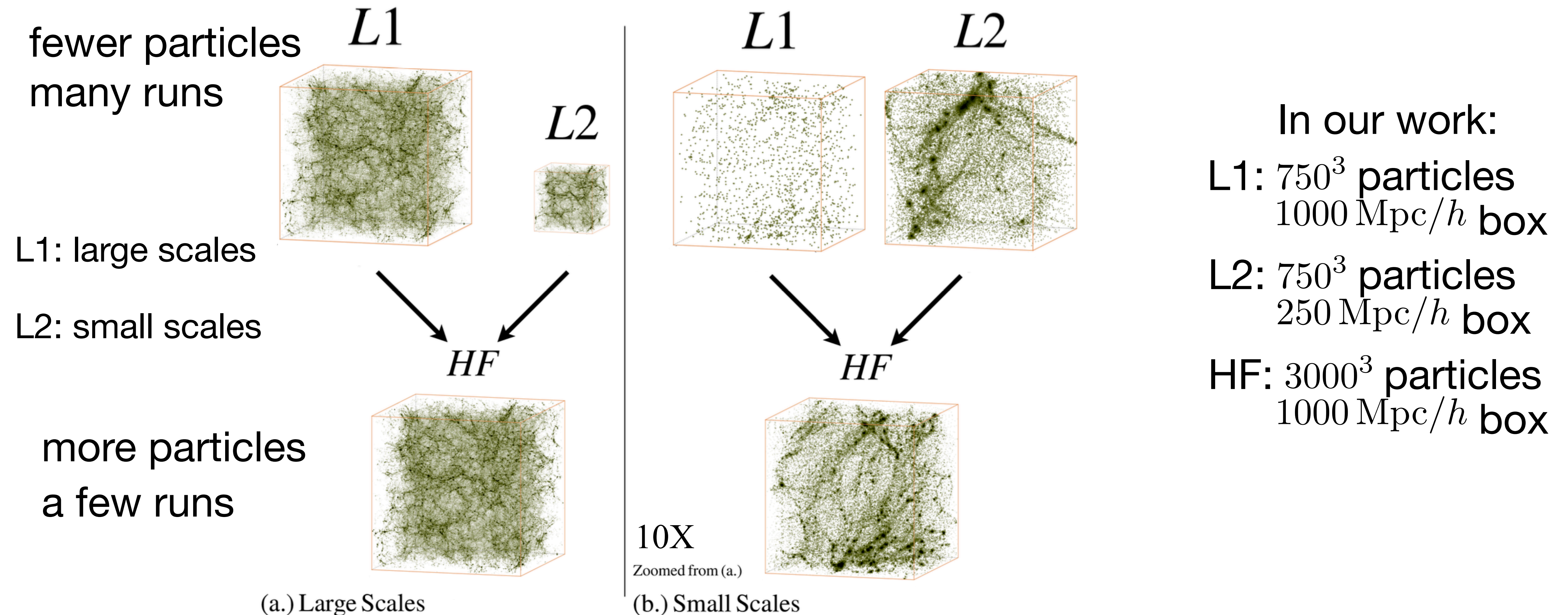
- MF-Box emulation: tree structure (Ho, Bird et al. 2023), two low-fidelity (LF) nodes (information sources) and one high-fidelity (HF) node



Cosmological Simulations: Goku (Yang, Bird et al. 2025)

Methods

- MF-Box emulation: tree structure (Ho, Bird et al. 2023), two low-fidelity (LF) nodes (information sources) and one high-fidelity (HF) node



Cosmological Simulations: Goku (Yang, Bird et al. 2025)

Methods

- Cosmological parameters: 10-dimensional parameter space (unprecedented)
- Sufficiently large prior ranges for the extensions:
 - Safely covers recent DESI constraints (only Goku does!)

Parameter	min(W)	max(W)
Ω_{m}	0.22	0.40
Ω_{b}	0.040	0.055
h	0.60	0.76
A_{s}	1.0×10^{-9}	3.0×10^{-9}
n_{s}	0.80	1.10
extensions {	w_0	-1.30
	w_a	-3.0
	$\sum m_{\nu}$	0.00
	N_{eff}	2.2
	α_{s}	-0.05
		0.60 eV

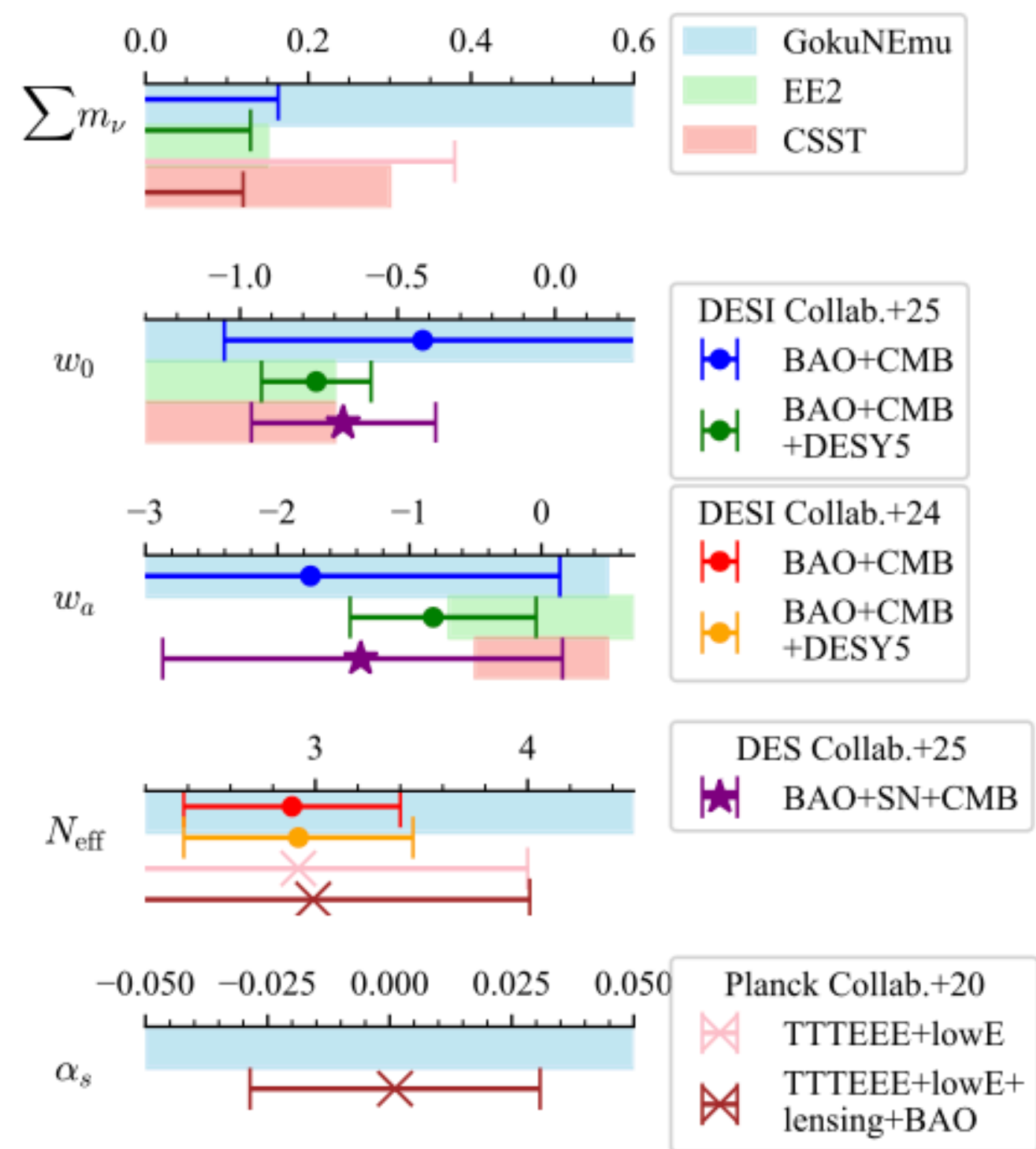
Cosmological Simulations: Goku (Yang, Bird et al. 2025)

Methods

- Cosmological parameters: 10-dimensional parameter space (unprecedented)
- Sufficiently large prior ranges for the extensions:
 - Safely covers recent DESI constraints (only Goku does!)

extensions {

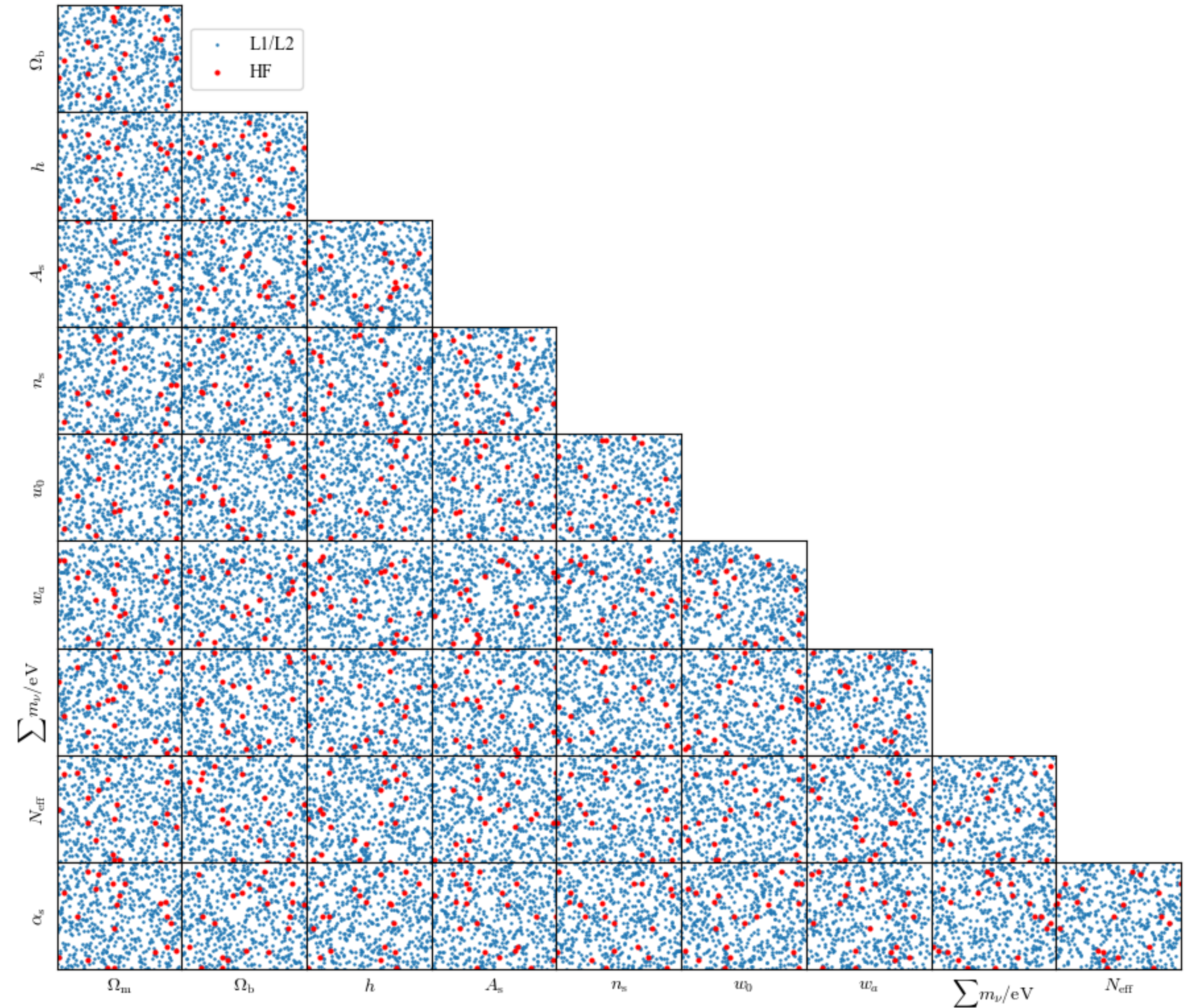
Parameter	min(W)	max(W)
Ω_{m}	0.22	0.40
Ω_{b}	0.040	0.055
h	0.60	0.76
A_{s}	1.0×10^{-9}	3.0×10^{-9}
n_{s}	0.80	1.10
w_0	-1.30	0.25
w_a	-3.0	0.5
$\sum m_{\nu}$	0.00	0.60 eV
N_{eff}	2.2	4.5
α_{s}	-0.05	0.05



Cosmological Simulations: Goku (Yang, Bird et al. 2025)

Methods

- Perform simulations with MP-Gadget (Feng, Bird et al. 2018)
- Goku: 1128 LF, 36 HF



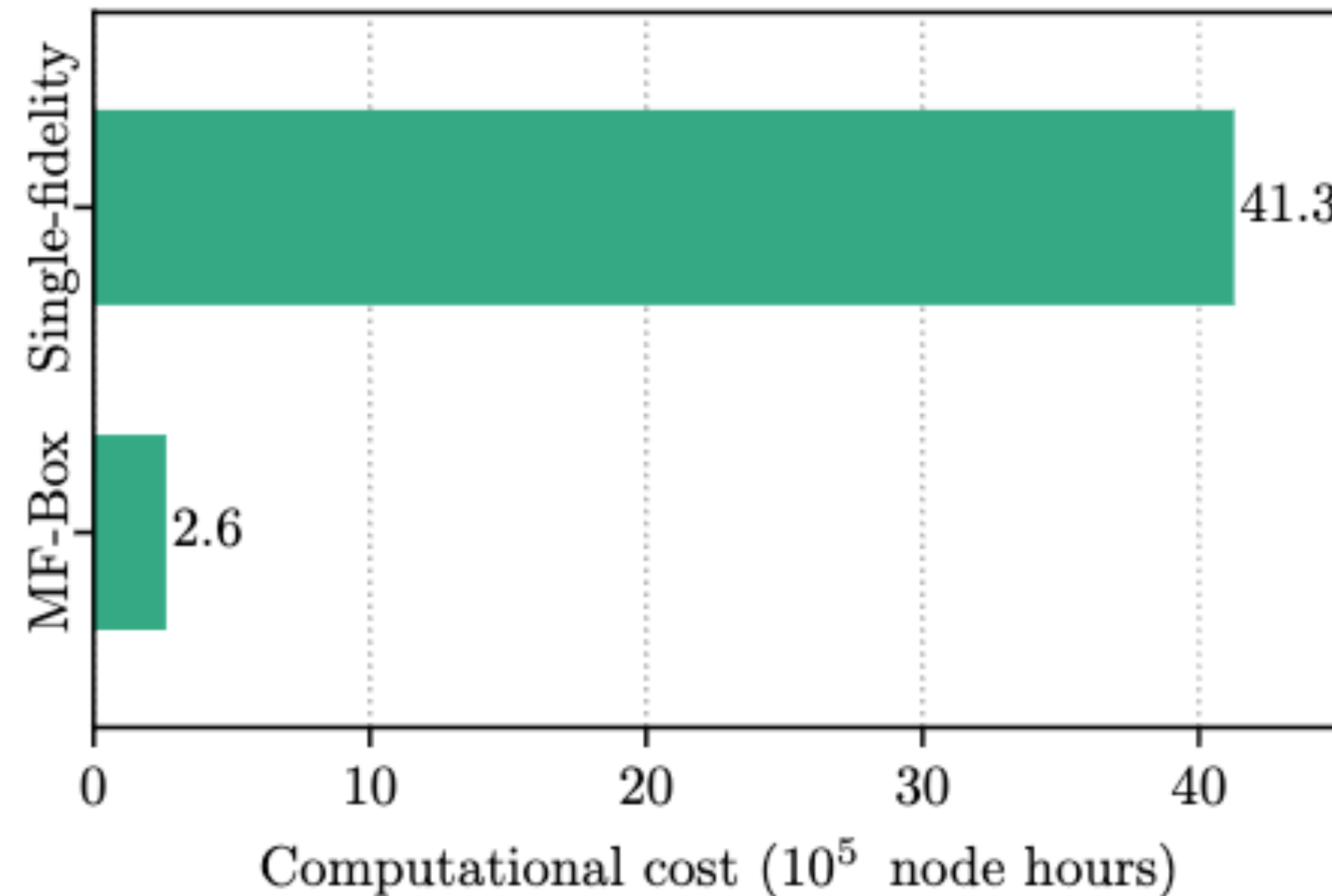
Cosmological Simulations: Goku (Yang, Bird et al. 2025)

Methods

- computational expense ~260,000 node-hours (Frontera, TACC)
- ~16x more efficient than a single-fidelity counterpart

(many years!)

a few months in practice



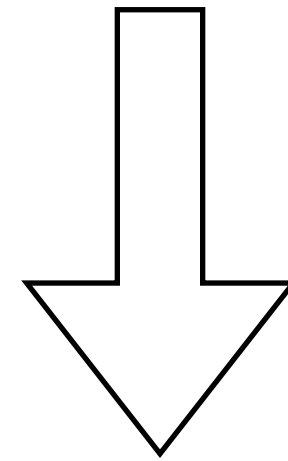
94% reduction!

Cosmological Simulations: Goku (Yang, Bird et al. 2025)

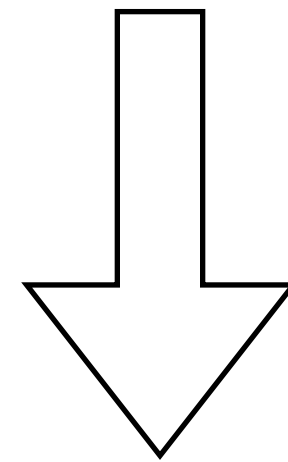
- Emulation using Gaussian process regression (MF-Box)?
- Computational complexity scales poorly with sample size:

Cosmological Simulations: Goku (Yang, Bird et al. 2025)

- Emulation using Gaussian process regression (MF-Box)?
- Computational complexity scales poorly with sample size:



slow in training and inference time!

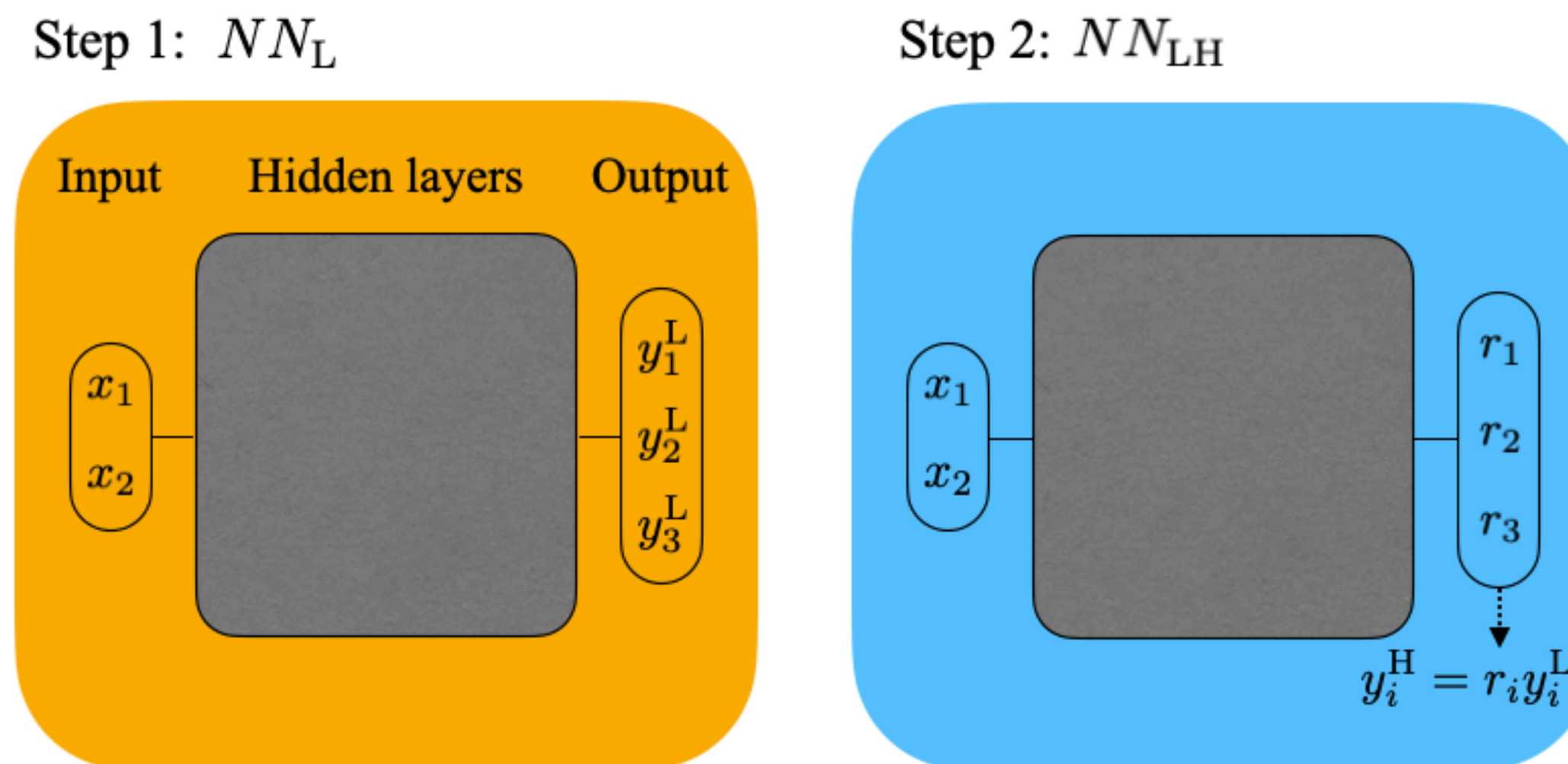


We next explore *neural-network (NN)* approaches

Multifidelity Emulation Framework: T2N-MusE (Yang, Bird et al. 2026a)

- We developed the "Triple 2" Neural network Multifidelity cosmological Emulation framework, with

- a 2-step multifidelity architecture



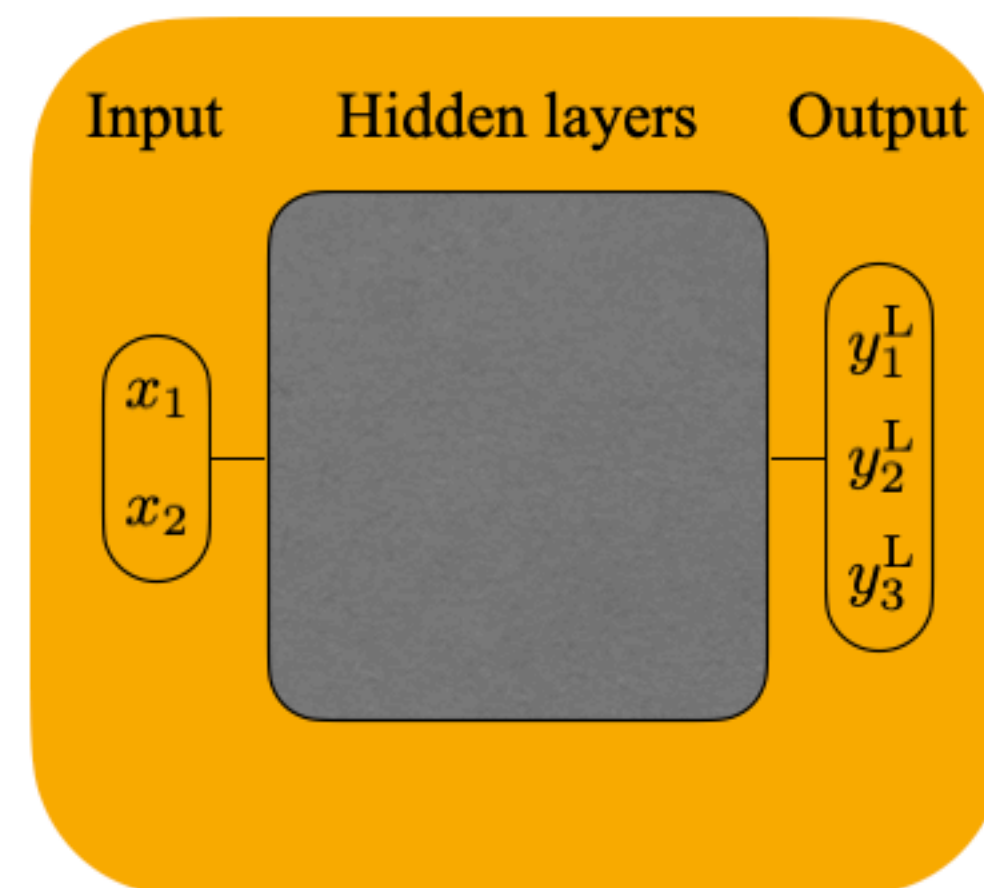
Multifidelity Emulation Framework: T2N-MusE (Yang, Bird et al. 2026a)

- We developed the "Triple 2" Neural network Multifidelity cosmological Emulation framework, with

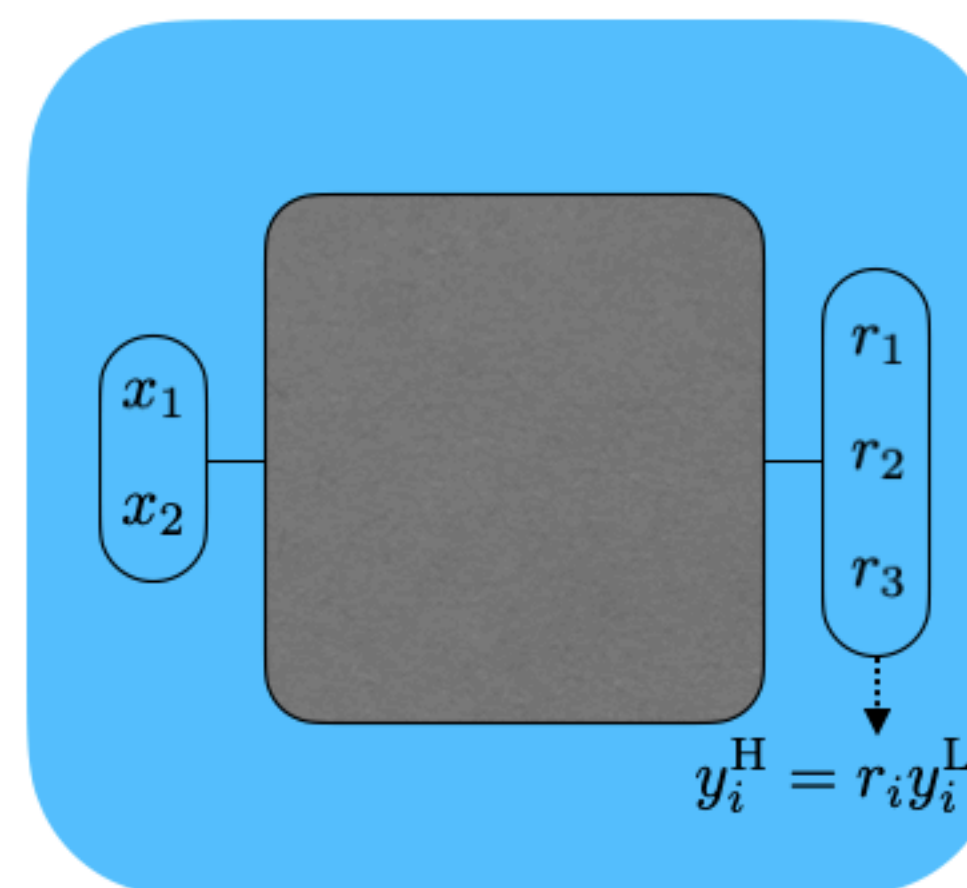
- a 2-step multifidelity architecture
- a 2-stage hyperparameter optimization strategy
- a 2-phase k-fold training and validation technique

technical details...

Step 1: NN_L

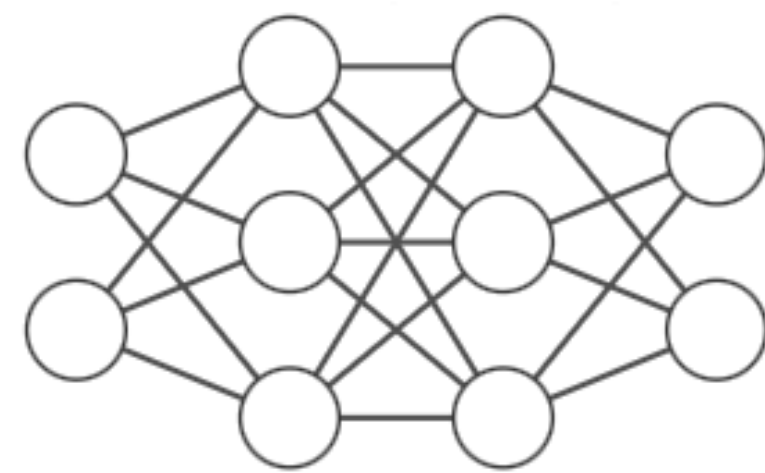
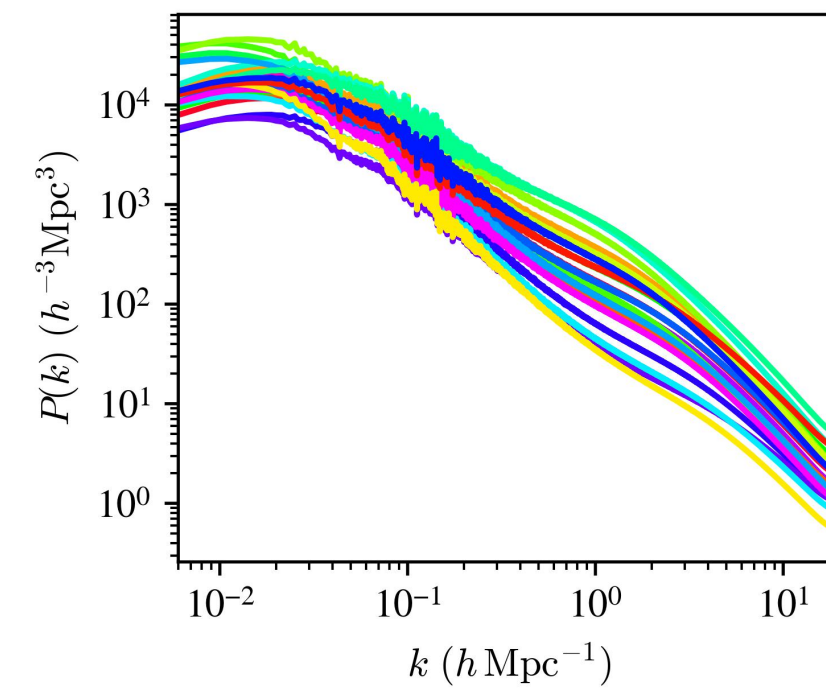
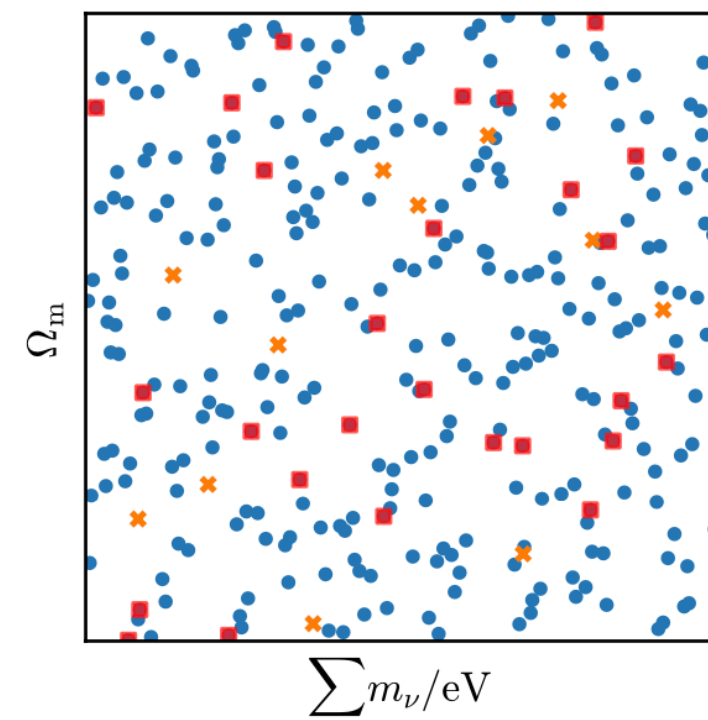


Step 2: NN_{LH}



Neural network emulator: GokuNEmu (Yang, Bird et al. 2026b)

- build the neural-network emulator



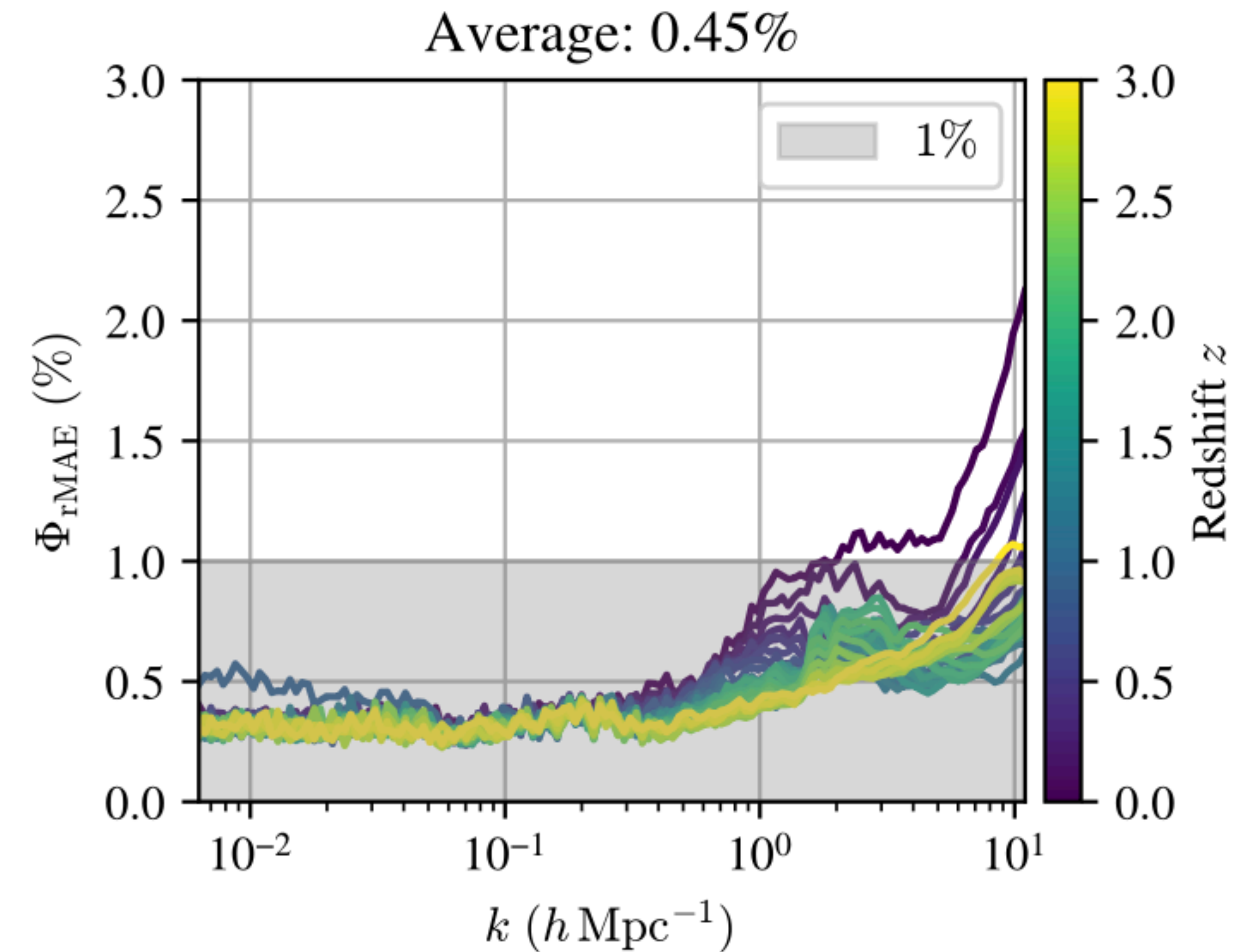
Goku

T2N-MuSE

GokuNEmu

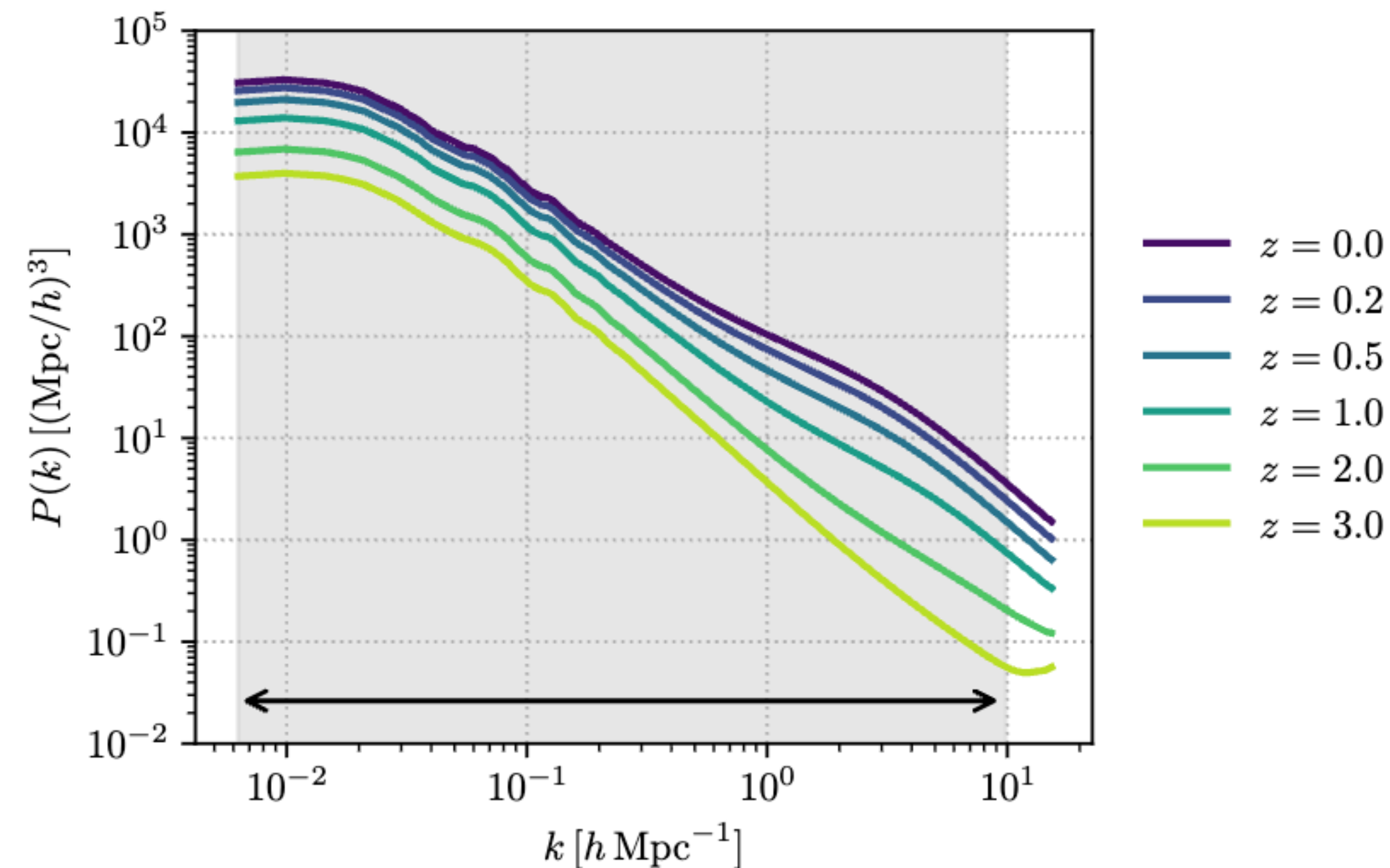
Neural network emulator: GokuNEmu (Yang, Bird et al. 2026b)

- Generalization accuracy
 - Leave-one-out cross-validation
- Percent-level error, sufficient for future analyses



Neural network emulator: GokuNEmu (Yang, Bird et al. 2026b)

- redshift and scale ranges:
 - $0 \leq z \leq 3$ (34 redshift bins; negligible interpolation error)
 - $0.006 < k/(h \text{ Mpc}^{-1}) < 10$



Neural network emulator: GokuNEmu

- Fastest evaluation: orders of magnitude faster in making predictions than some others
- **GokuNEmu**: a few **ms** (per evaluation)

Estimator	single redshift [s]	50 redshifts [s]	DE2 cost ratio
DARKEMULATOR2	0.0169	0.1006	1.00
DARKEMULATOR2 + direct CLASS	21.3455	21.5335	1263.05
EUCLIDEMULATOR2	0.9182	1.0332	54.33
BACCOEMU	0.0293	1.0536	1.73
MIRA-TITAN UNIV. IV	0.0117	0.2644	0.69
AEMULUSv	1.9274	94.2646	114.05
GOKUITEMU	0.0039	0.0042	0.23
ALETHEIA	1.0490	52.1157	62.07
CSSTEMULATOR	0.0603	0.0619	3.57
FRANKENEMU	0.0067	0.3390	0.40
PKANN	0.0257	1.2068	1.52
HALOFIT	7.2766	7.4387	430.57
HMCODE	8.0016	8.0575	473.47

Table from Tanaka+2026

Neural network emulator: GokuNEmu

- Fastest evaluation: orders of magnitude faster in making predictions than some others
- **GokuNEmu**: a few **ms** (per evaluation)

Estimator	single redshift [s]	50 redshifts [s]	DE2 cost ratio
<u>DARKEMULATOR2</u>	<u>0.0169</u>	0.1006	1.00
DARKEMULATOR2 + direct CLASS	21.3455	21.5335	1263.05
EUCLIDEMULATOR2	0.9182	1.0332	54.33
BACCOEMU	0.0293	1.0536	1.73
MIRA-TITAN UNIV. IV	0.0117	0.2644	0.69
AEMULUSv	1.9274	94.2646	114.05
<u>GOKUITEMU</u>	<u>0.0039</u>	0.0042	0.23
ALETHEIA	1.0490	52.1157	62.07
CSSTEMULATOR	0.0603	0.0619	3.57
FRANKENEMU	0.0067	0.3390	0.40
PKANN	0.0257	1.2068	1.52
HALOFIT	7.2766	7.4387	430.57
HMCODE	8.0016	8.0575	473.47

Table from Tanaka+2026

Future Work

- Applying GokuNEmu to weak-lensing analyses (cosmic shear, or 3x2pt)
- Goku simulations --> mock weak-lensing maps (including non-Gaussian info)
- Predicting the halo mass function and galaxy correlation function (Sum Qezlou at UT Austin)

Summary

GokuNEmu



- **Goku** simulation suite: flat Λ CDM+ $w_0w_a+\Sigma m_\nu+N_{\text{eff}}+\alpha_s$ (10D)
- **T2N-MusE**: a novel NN framework for MF cosmological emulation
- **GokuNEmu**: a uniquely *powerful* emulator for the nonlinear matter power spectrum (will be useful for Euclid, Roman etc.):
 - *unprecedentedly high-dimensional*
 - broad prior ranges for w_0 and w_a (the *only* emulator that allows tests on recent DESI constraints)
 - fastest in runtime ($\sim 2\text{--}3$ ms per evaluation)
- Publications: i. Yang+2025 (PRD), [arXiv:2501.06296](#); ii. Yang+2026a (PRD), [arXiv:2507.07184](#); iii. Yang+2026b (PRL), [arXiv:2507.07177](#).
- Future: data analysis, WL maps and other summary statistics