Extracting the Full Cosmological Information of Roman

Changhoon Hahn

Princeton University

changhoon.hahn@princeton.edu

changhoonhahn.github.io

the *standard* ΛCDM model of cosmology

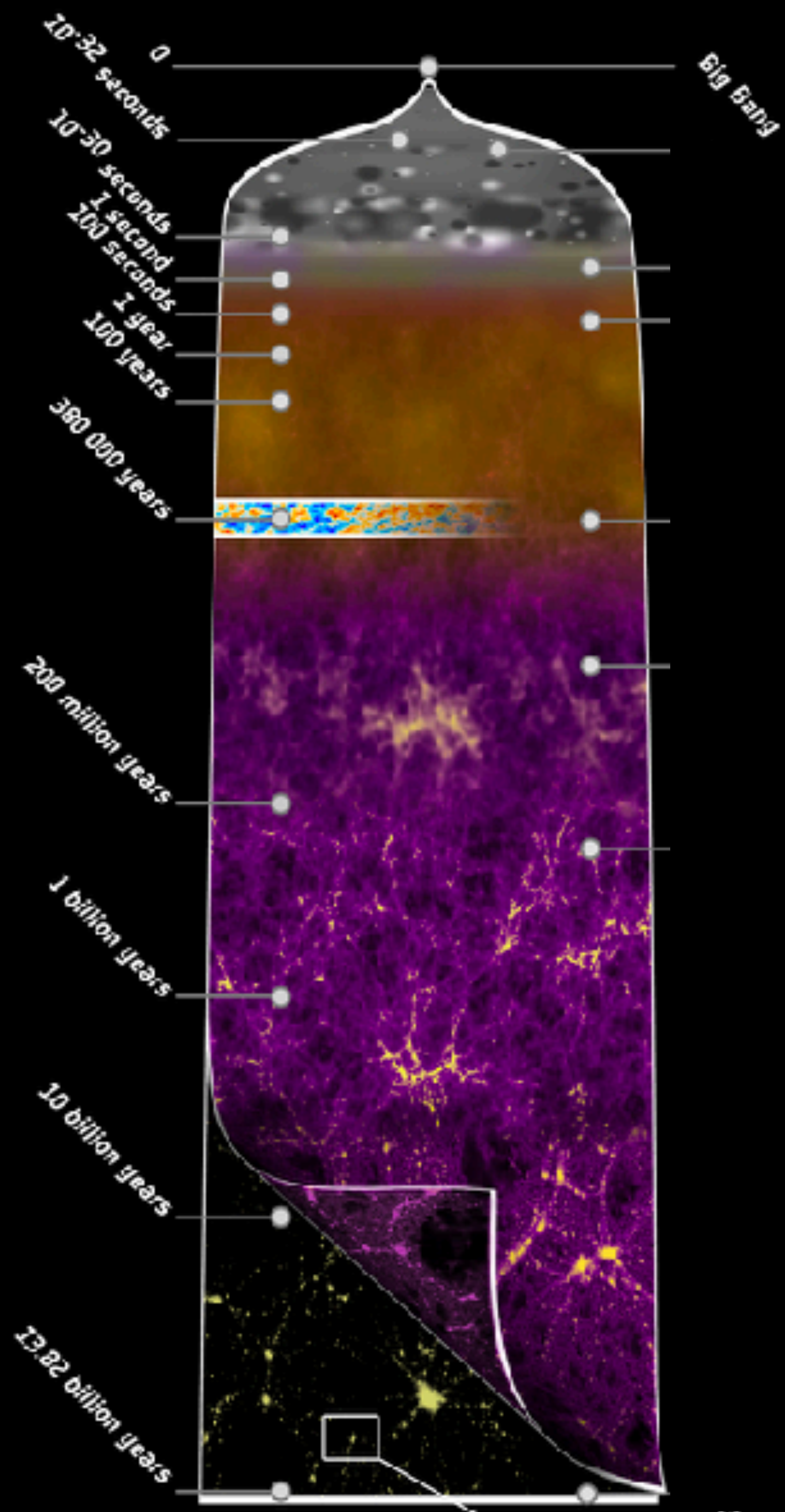
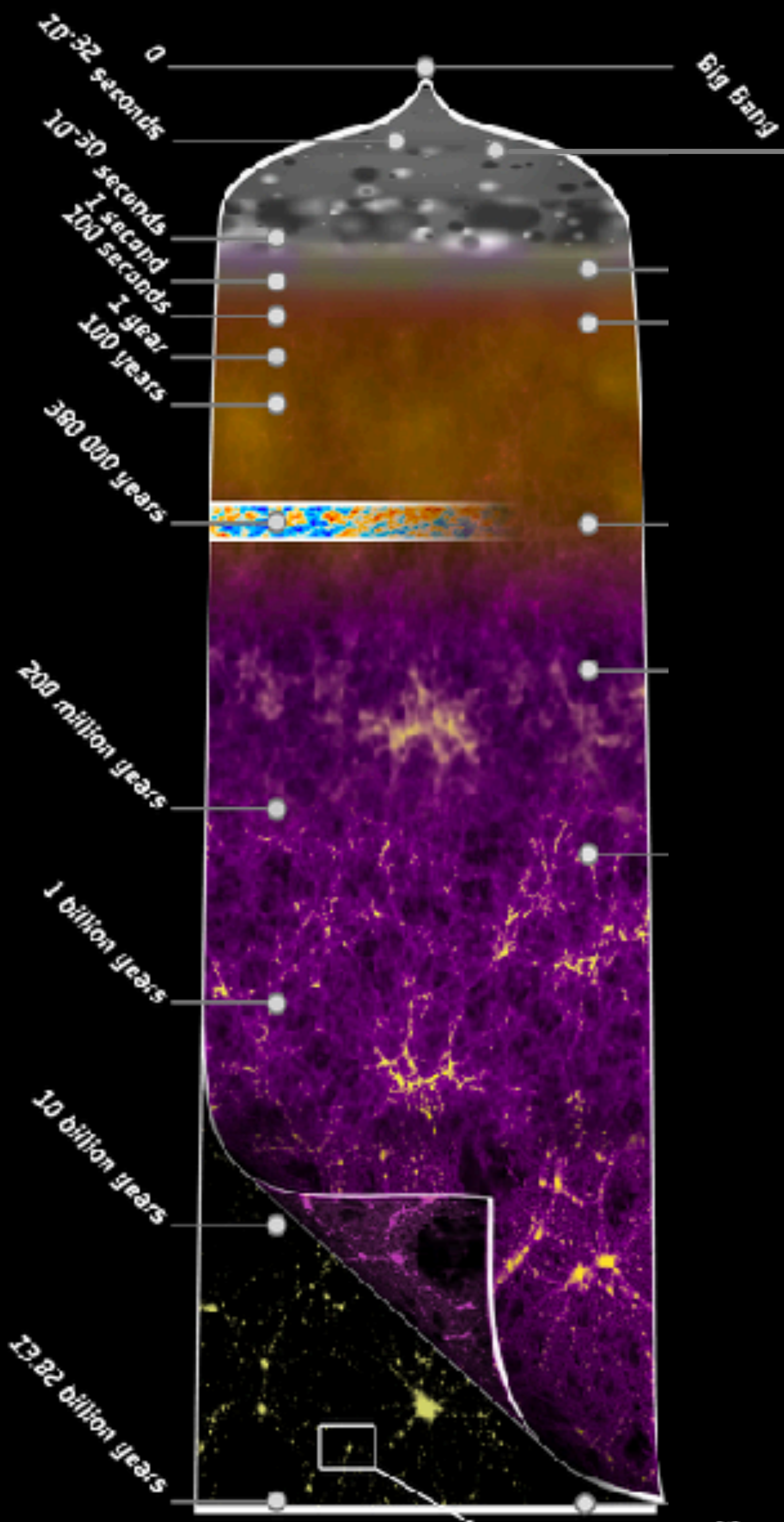


image: NASA

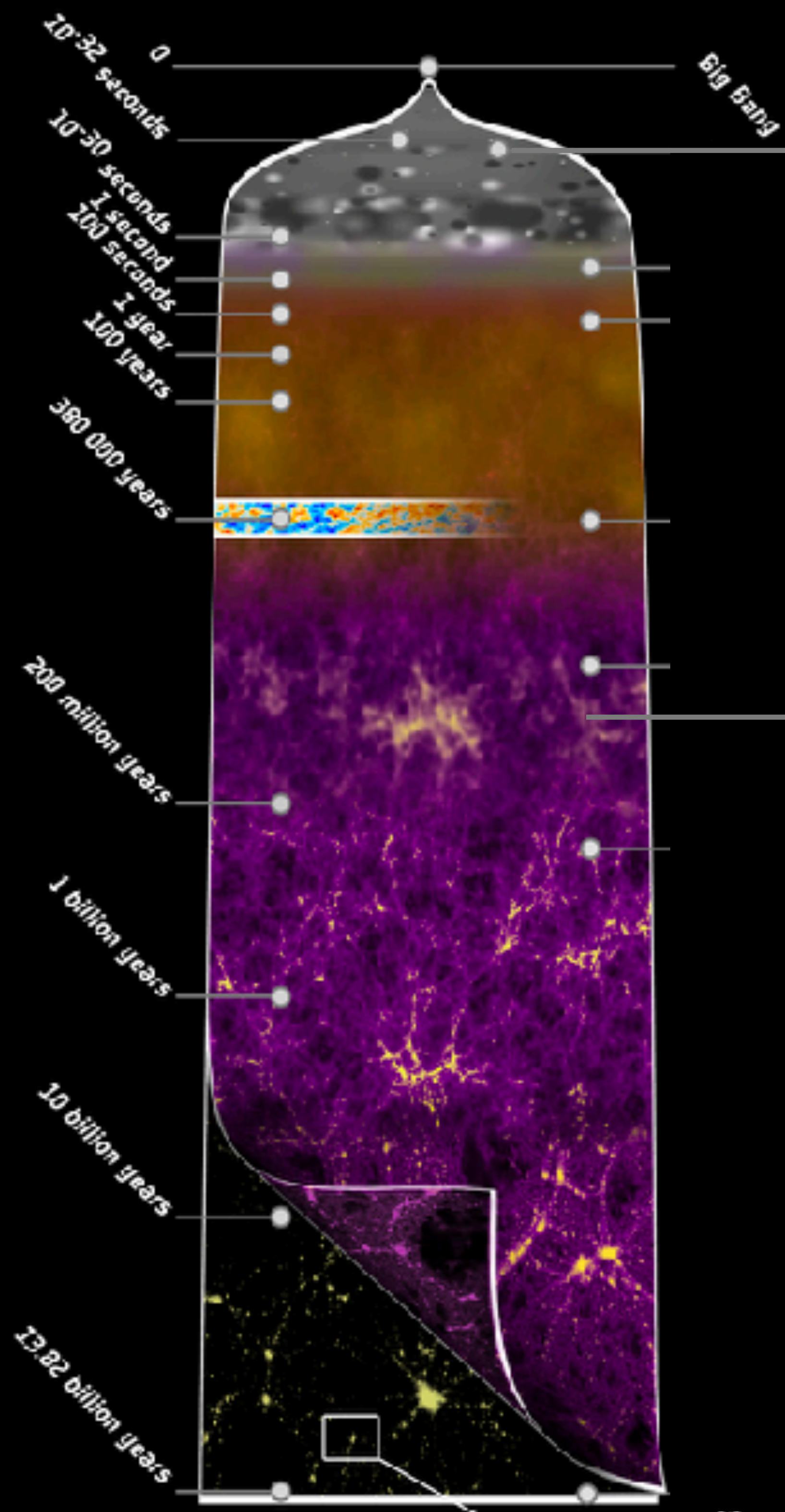
the *standard* ΛCDM model of cosmology



cosmic inflation — *origin of fluctuations*

image: NASA

the *standard* ΛCDM model of cosmology

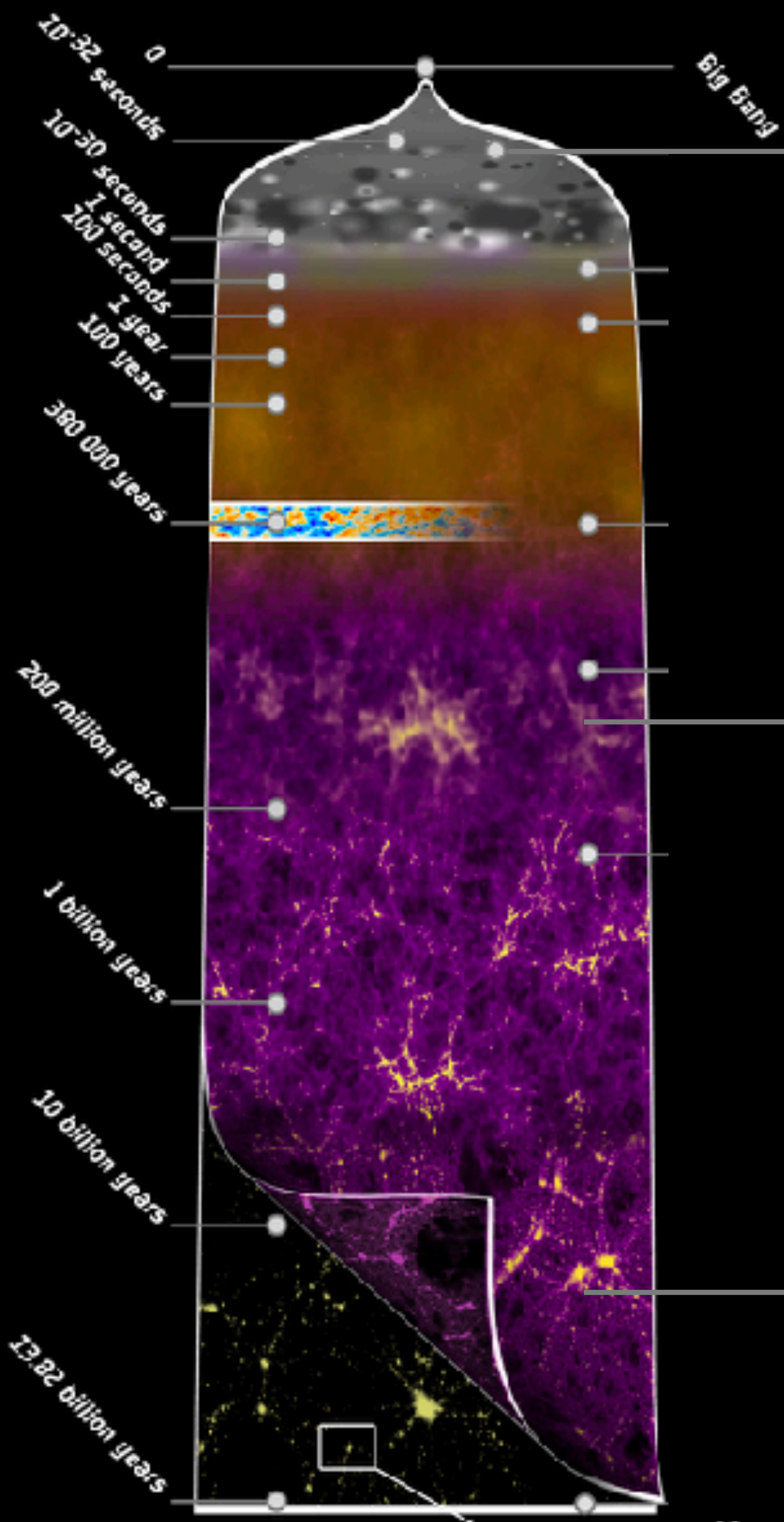


cosmic inflation — *origin of fluctuations*

dark matter — *~85% of matter*

image: NASA

the *standard* ΛCDM model of cosmology



cosmic inflation — *origin of fluctuations*

dark matter — *~85% of matter*

dark energy — *accelerated expansion*

the *standard* ΛCDM model of cosmology is *remarkably successful* at describing current observations

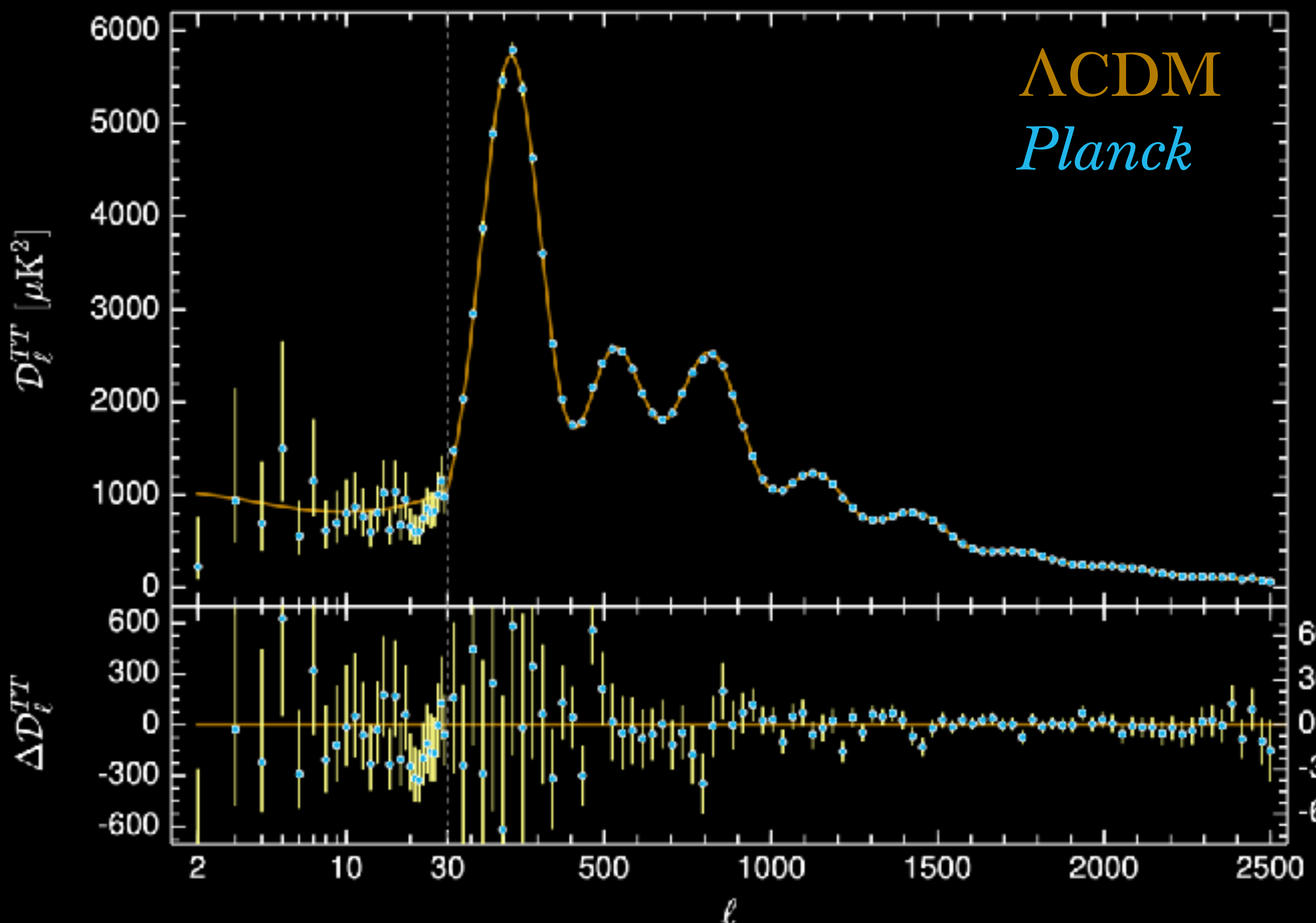
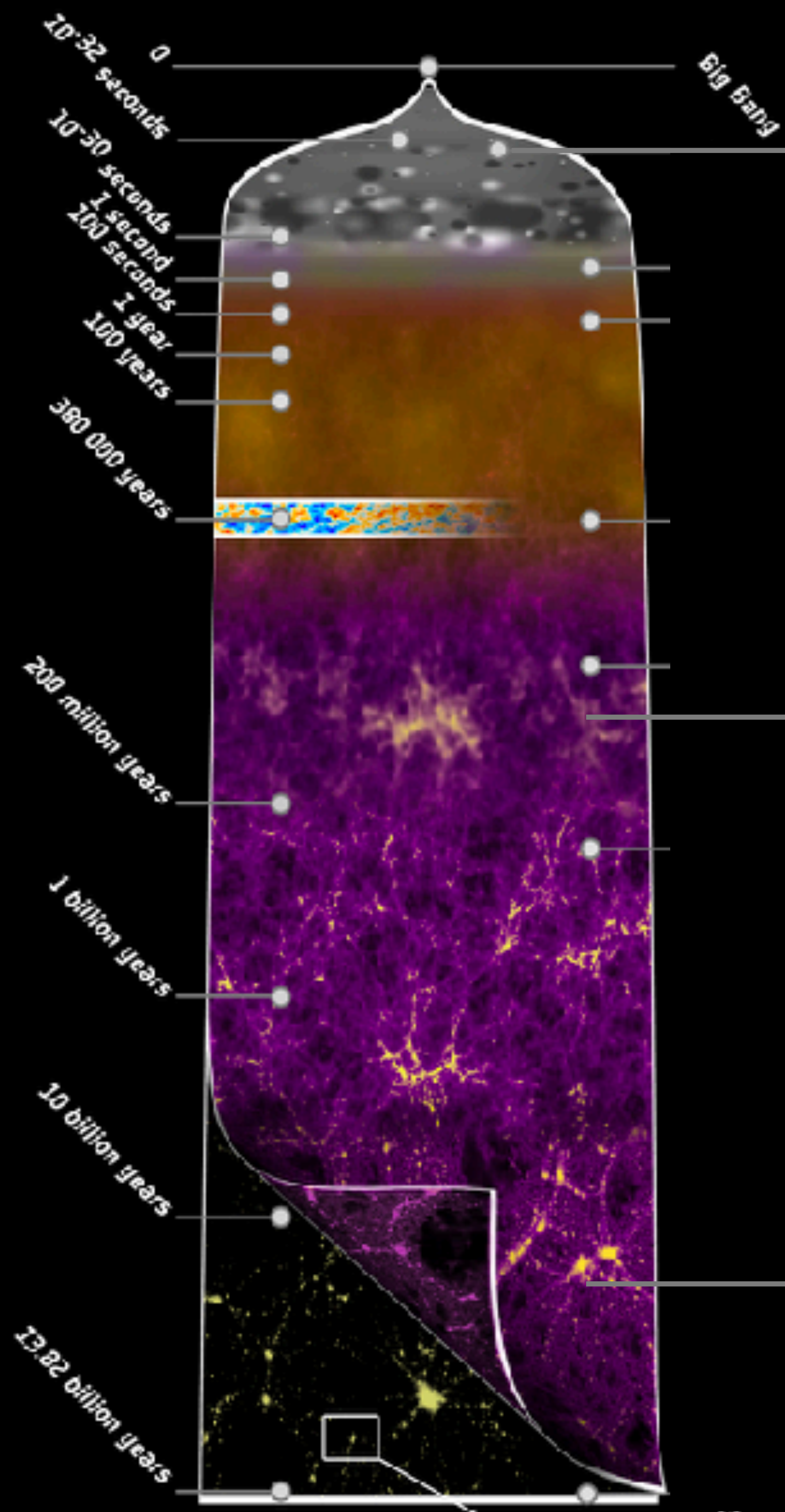


image: Planck

the *standard* ΛCDM model of cosmology



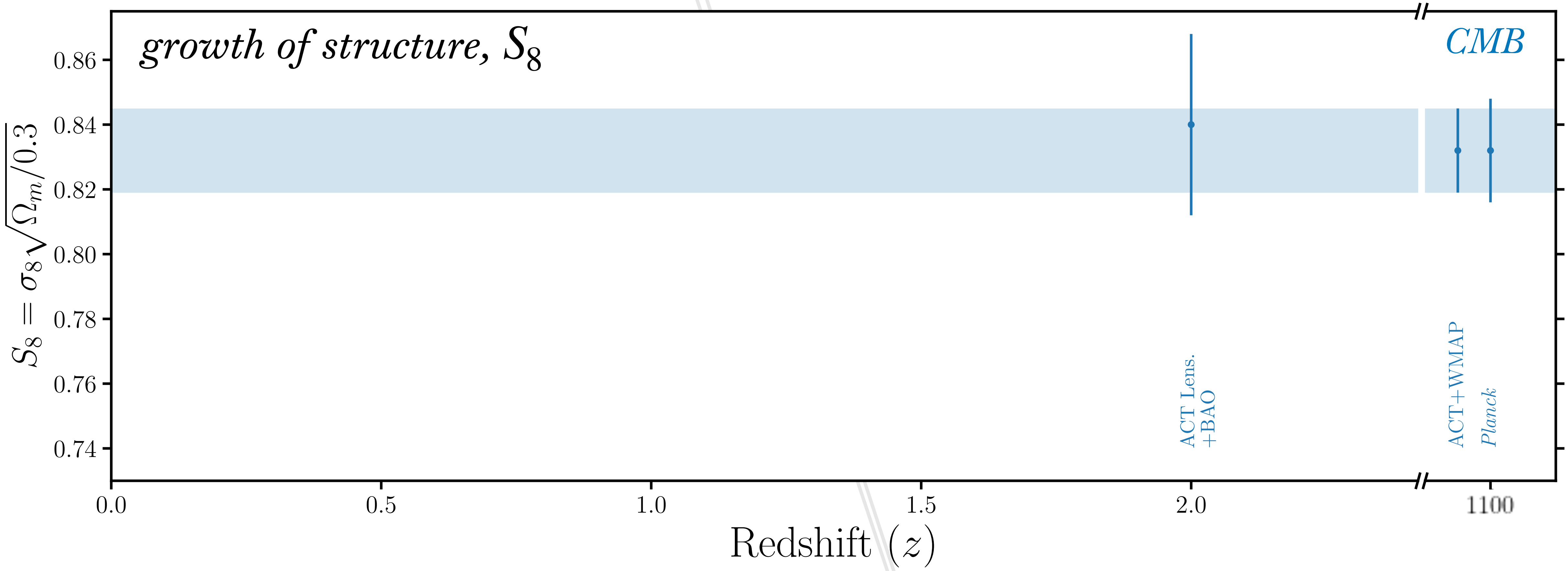
cosmic inflation — ***not understood***

dark matter — ***not understood***

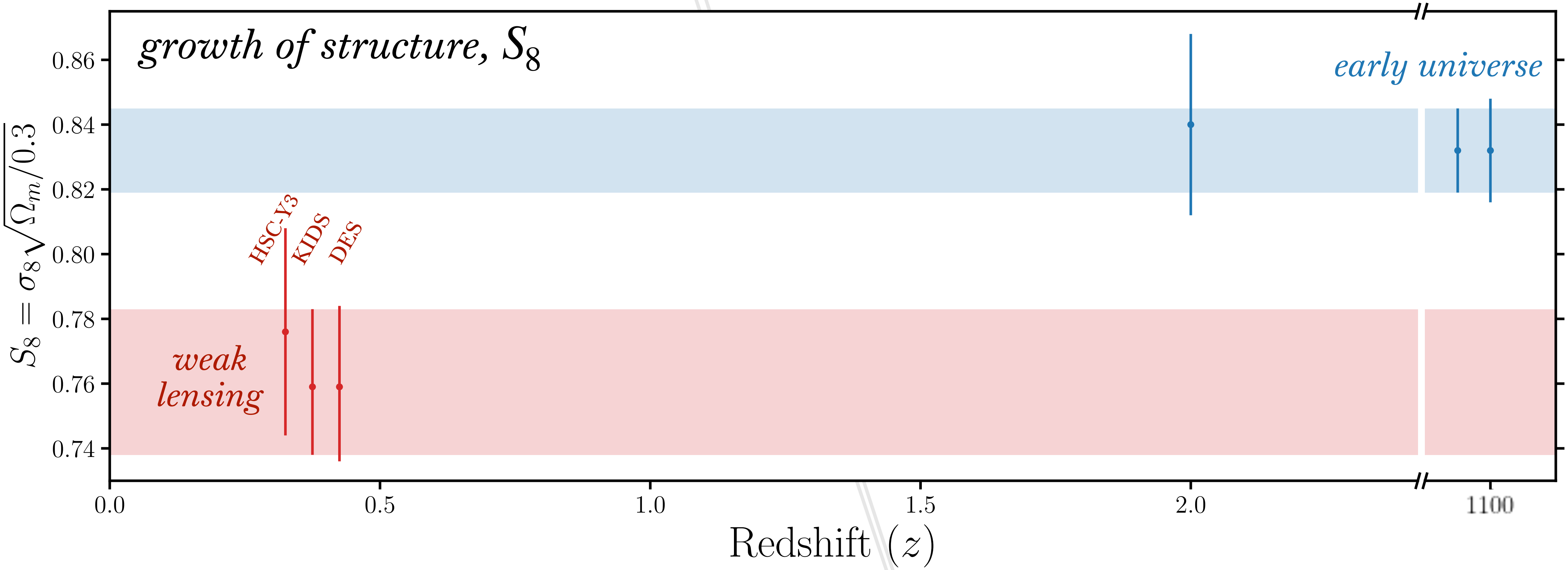
dark energy — ***not understood***

cosmic tensions

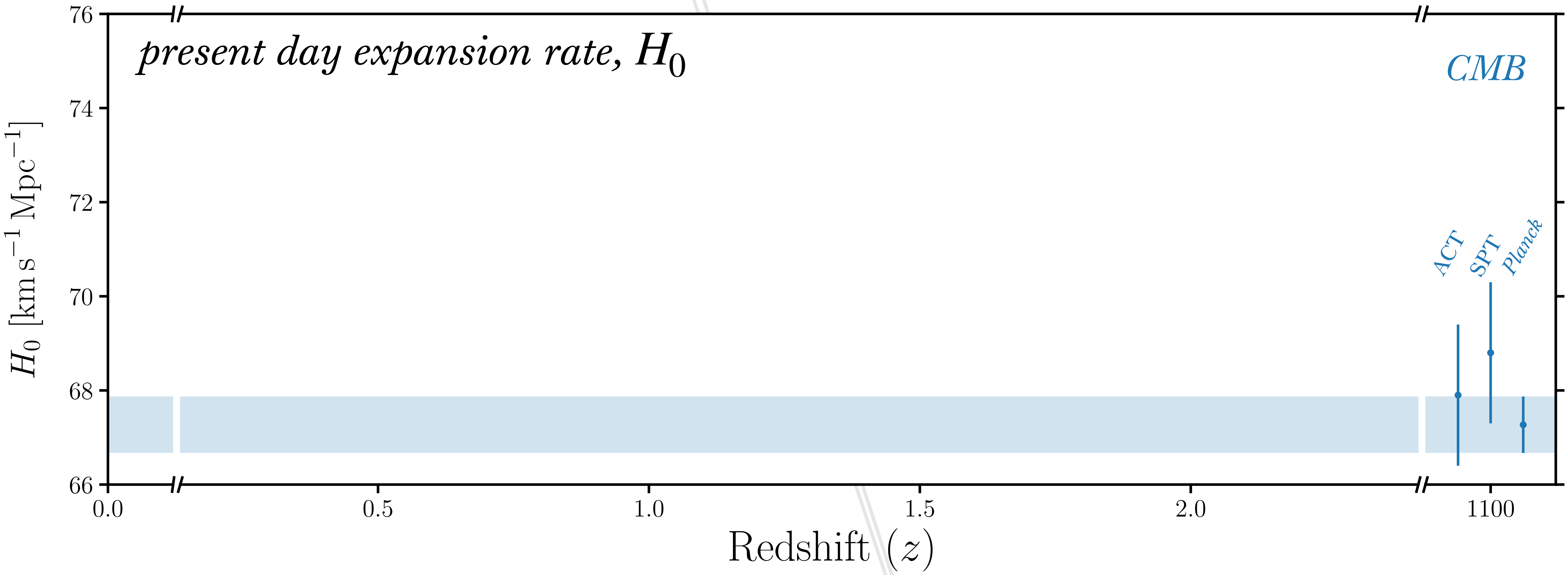
cosmic tensions between probes of the early and the late universe



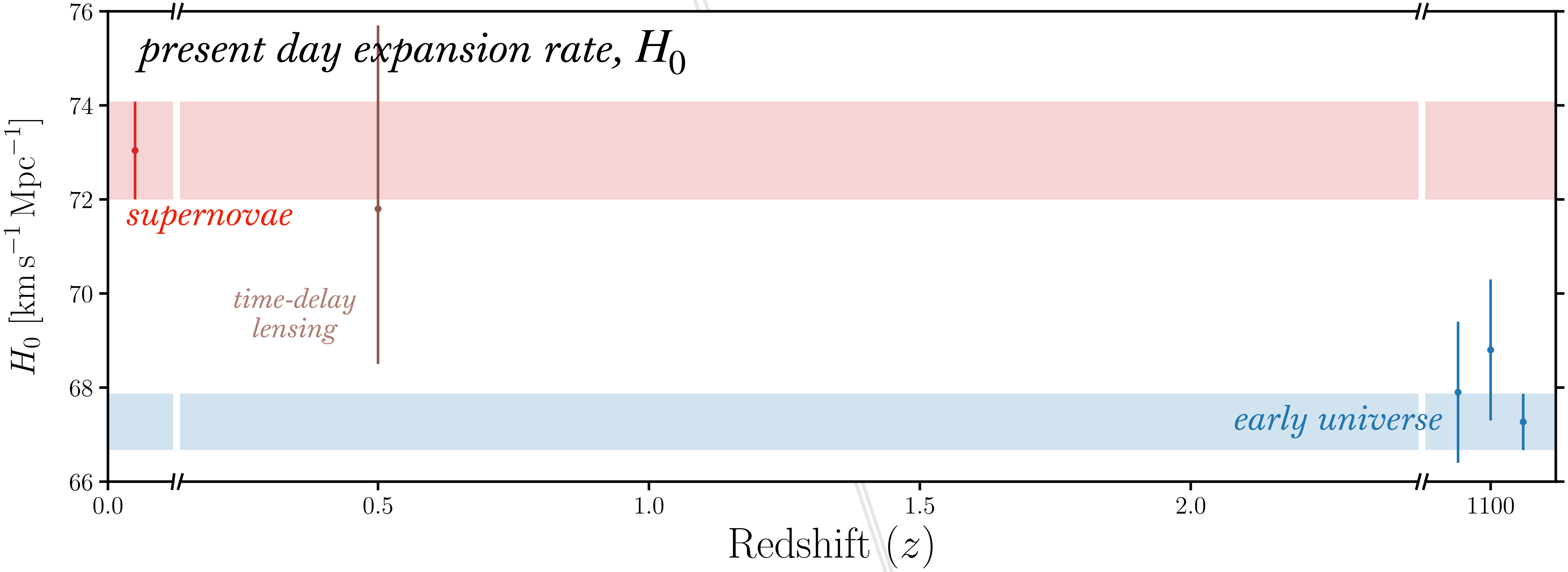
cosmic tensions between probes of the early and the late universe



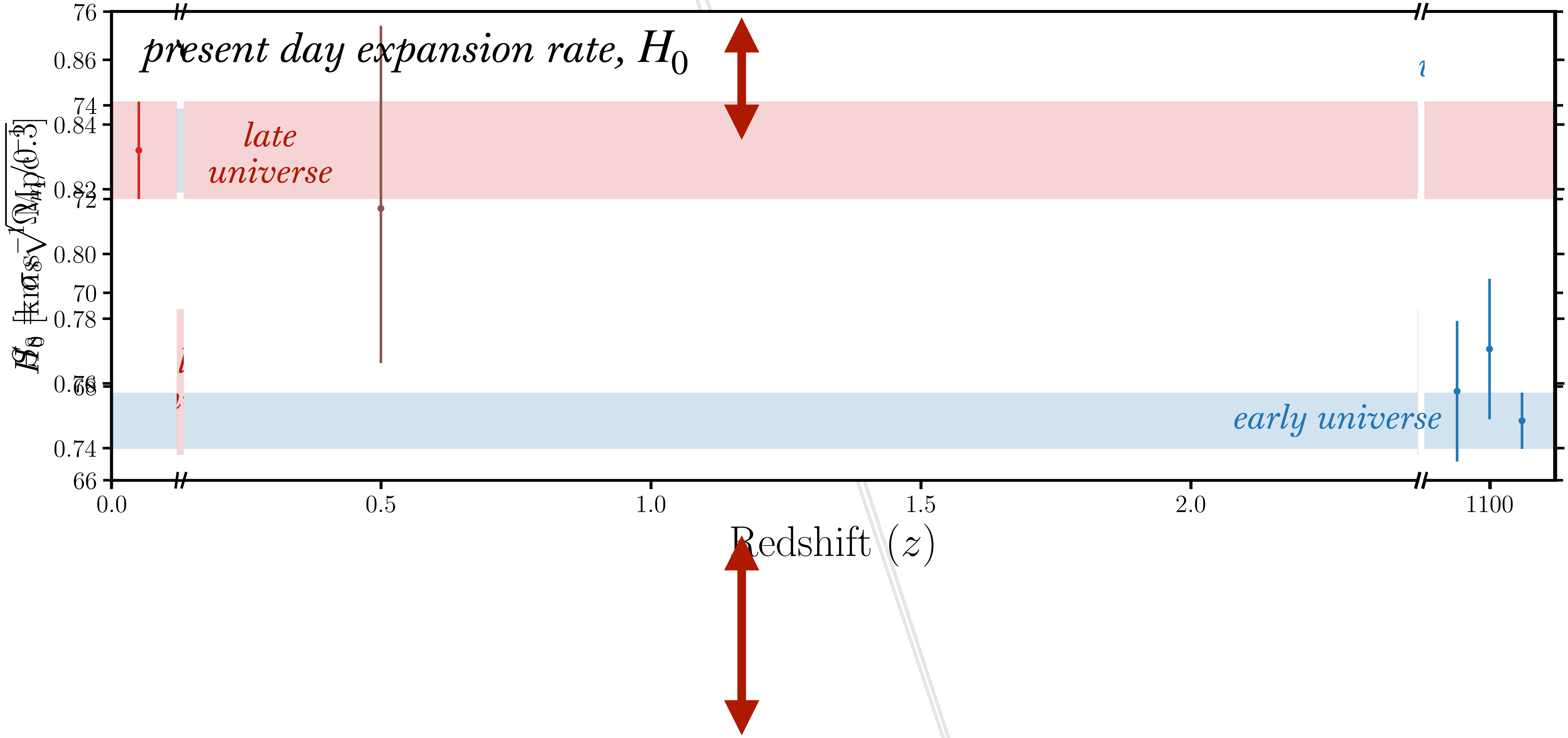
cosmic tensions between probes of the early and the late universe



cosmic tensions between probes of the early and the late universe



cosmic tensions between probes of the early and the late universe



spectroscopic galaxy surveys probe both the growth of structure and the expansion rate

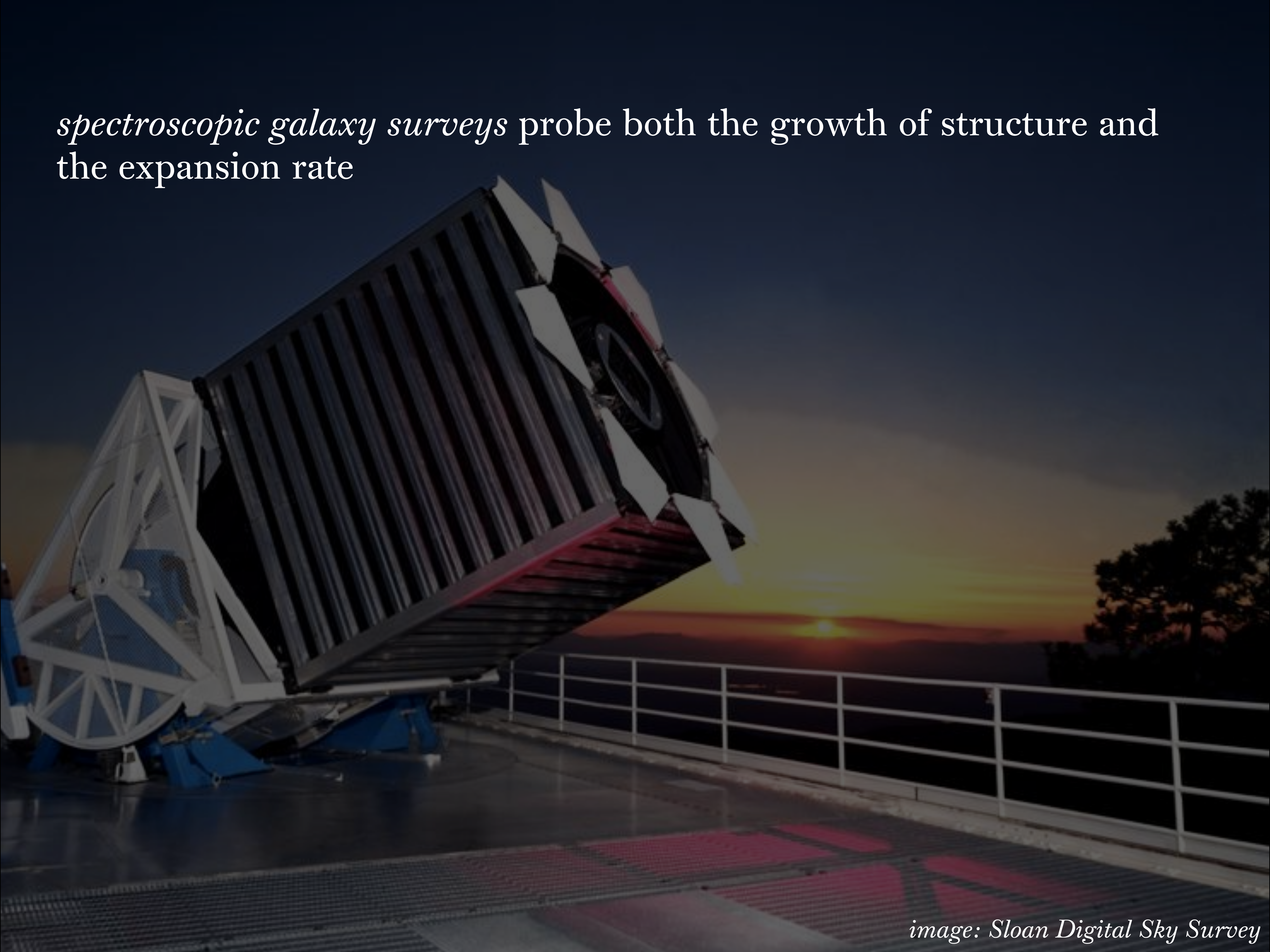
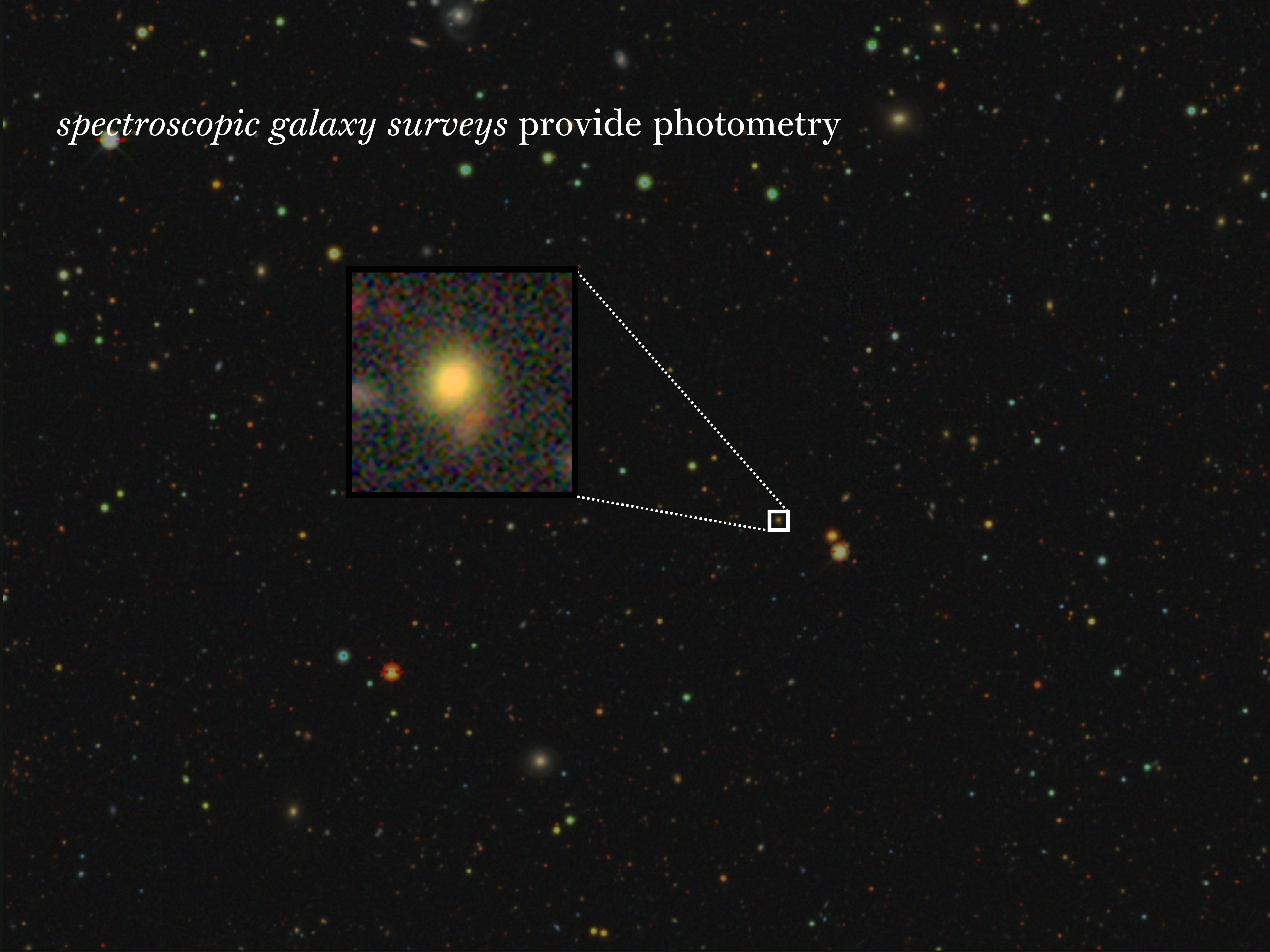
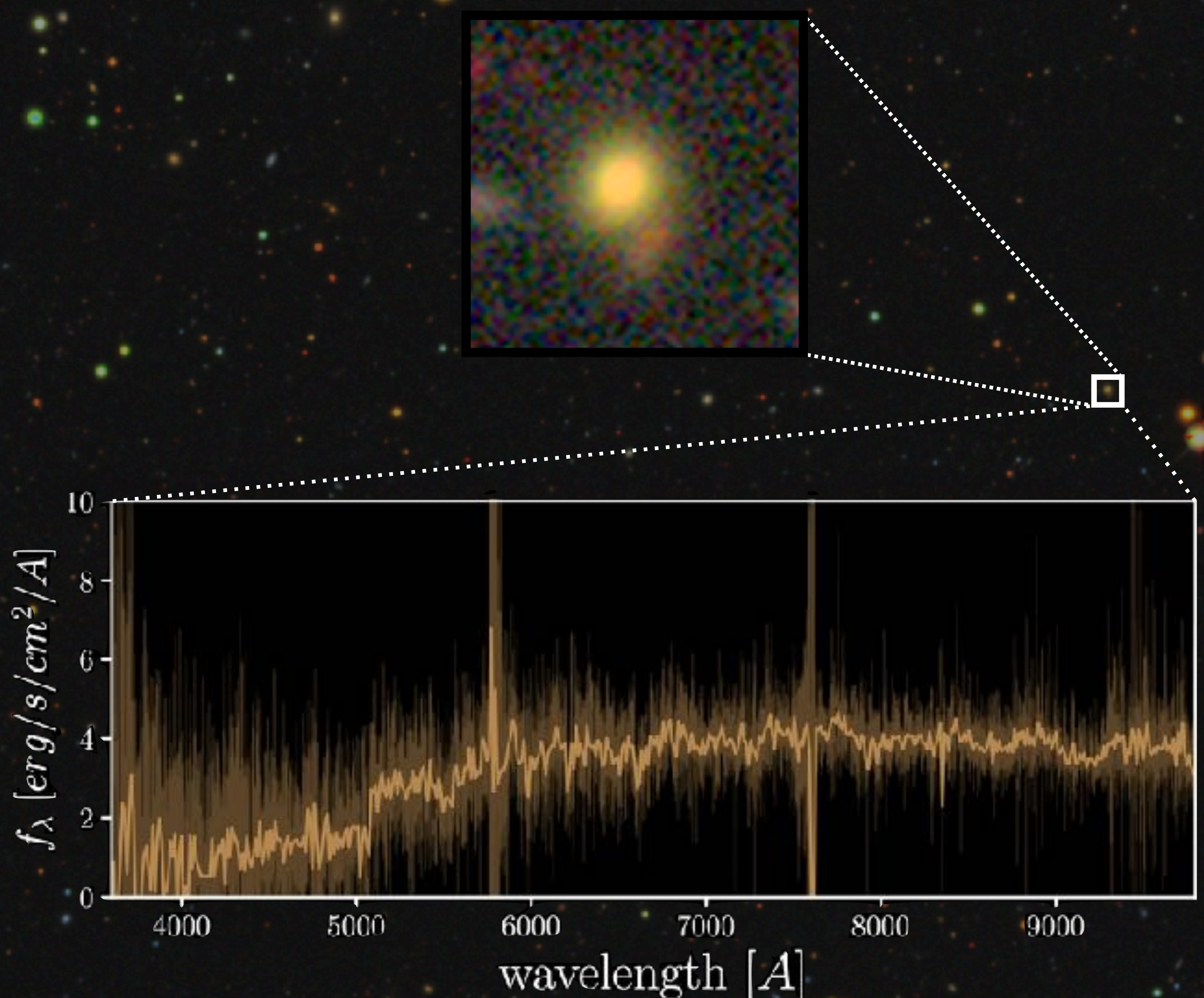


image: Sloan Digital Sky Survey


spectroscopic galaxy surveys provide photometry

spectroscopic galaxy surveys provide photometry and spectra of galaxies



spectroscopic galaxy surveys map the detailed three-dimensional spatial distribution of galaxies

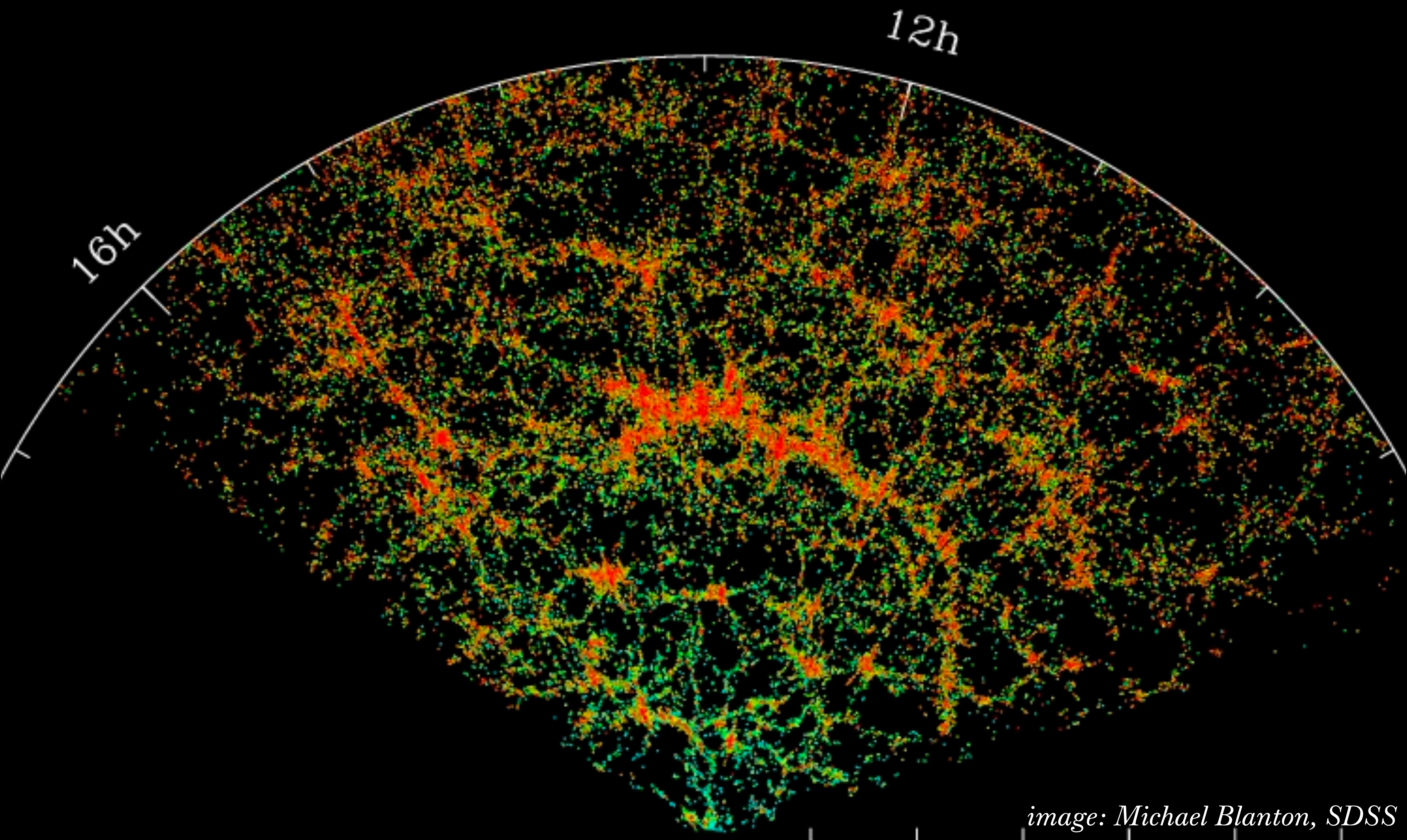


image: Michael Blanton, SDSS

3D distribution of galaxies encodes cosmological information on the **growth of structure** — *redshift-space distortions*

3D distribution of galaxies encodes cosmological information on the **growth of structure** — *redshift-space distortions*

$$z_{\rm obs} = z_{\rm cosmo} + \frac{v_{\rm pec}}{c}$$

cosmological expansion

3D distribution of galaxies encodes cosmological information on the **growth of structure** — *redshift-space distortions*

$$z_{\rm obs} = z_{\rm cosmo} + \frac{v_{\rm pec}}{c}$$

peculiar velocity

3D distribution of galaxies encodes cosmological information on the **growth of structure** — *redshift-space distortions*

$$z_{\rm obs} = z_{\rm cosmo} + \frac{v_{\rm pec}}{c}$$

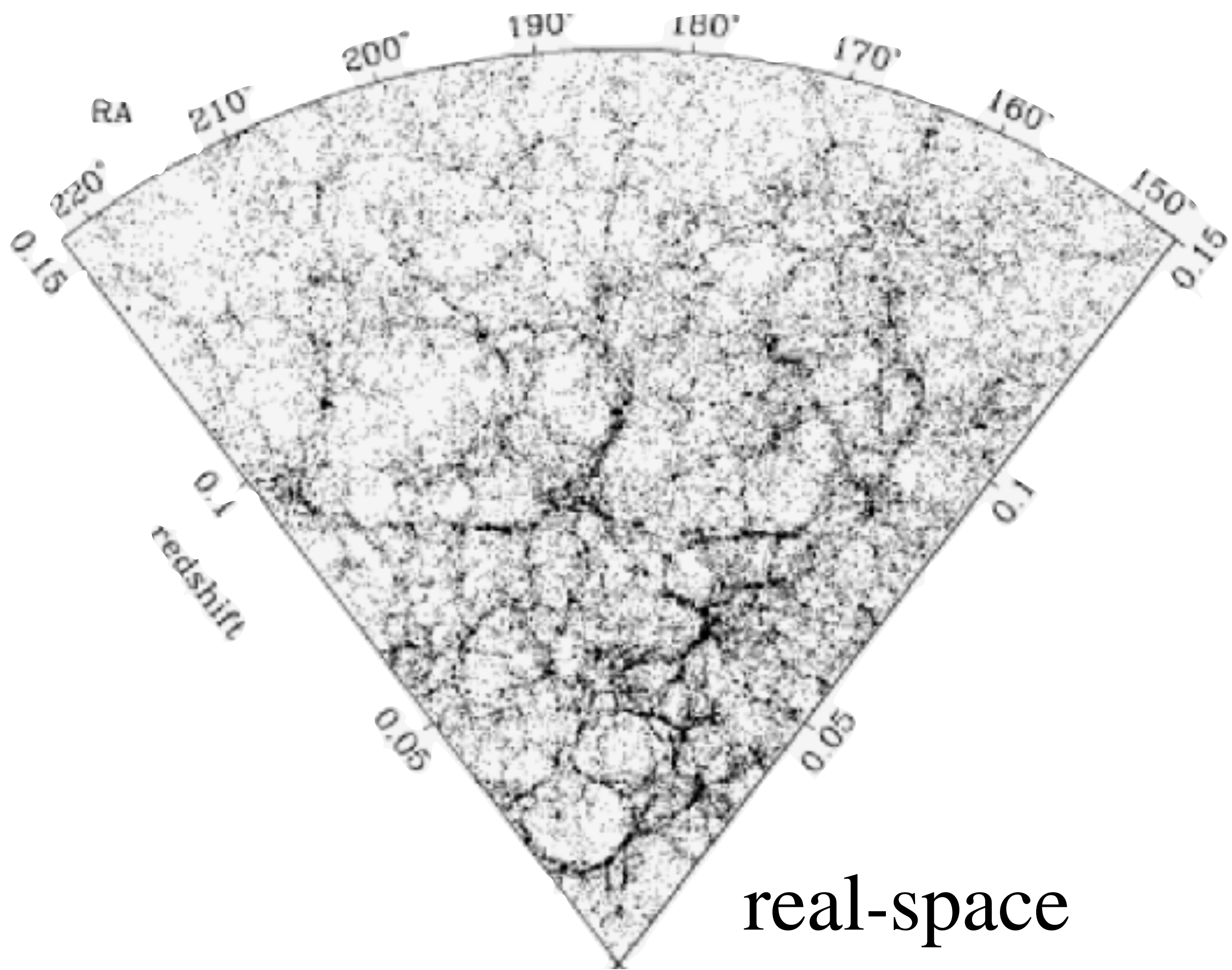
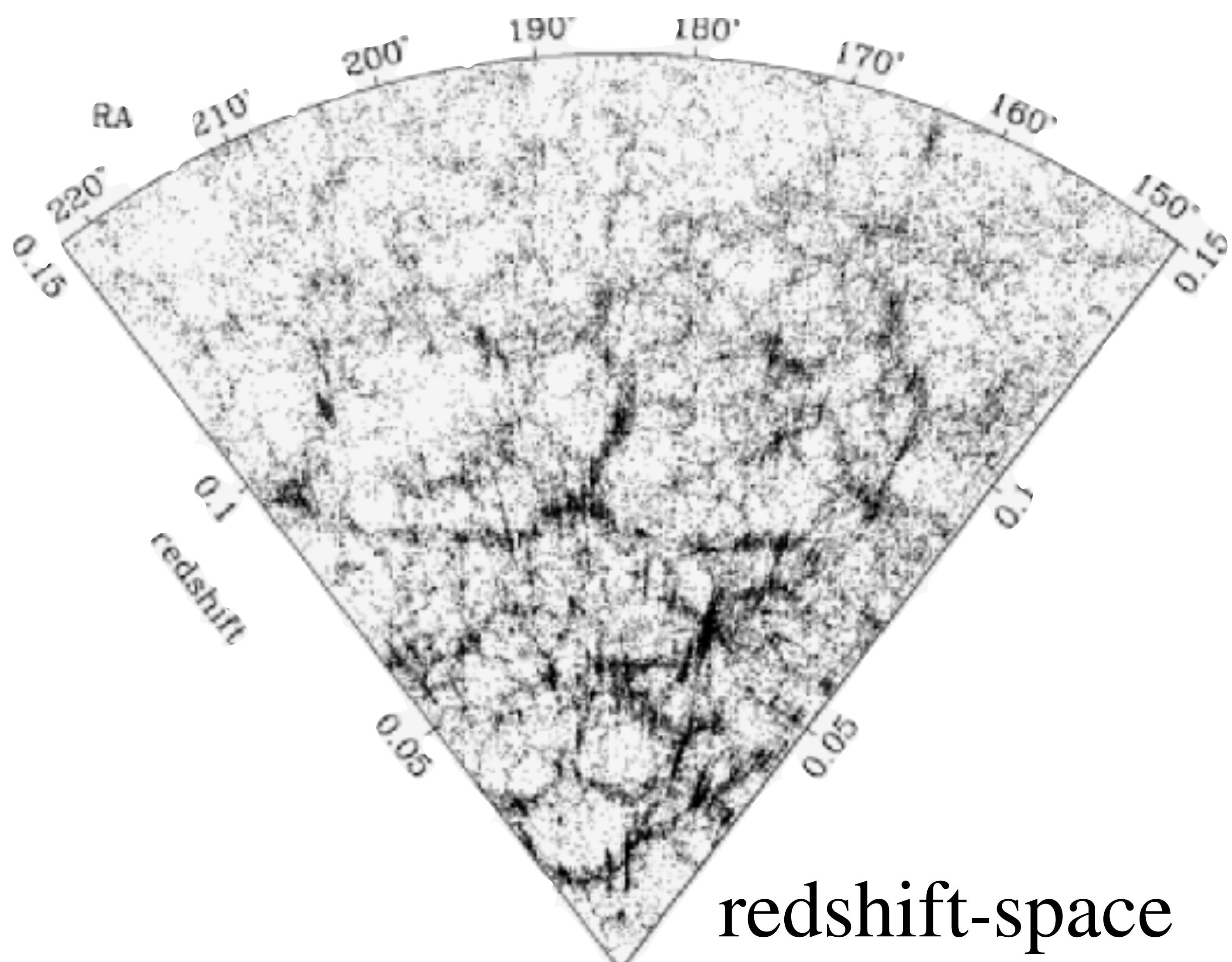


image: Eke et al. (2003)

3D distribution of galaxies encodes cosmological information on the **growth of structure** — *redshift-space distortions*

$$z_{\mathrm{obs}} = z_{\mathrm{cosmo}} + \frac{v_{\mathrm{pec}}}{c}$$

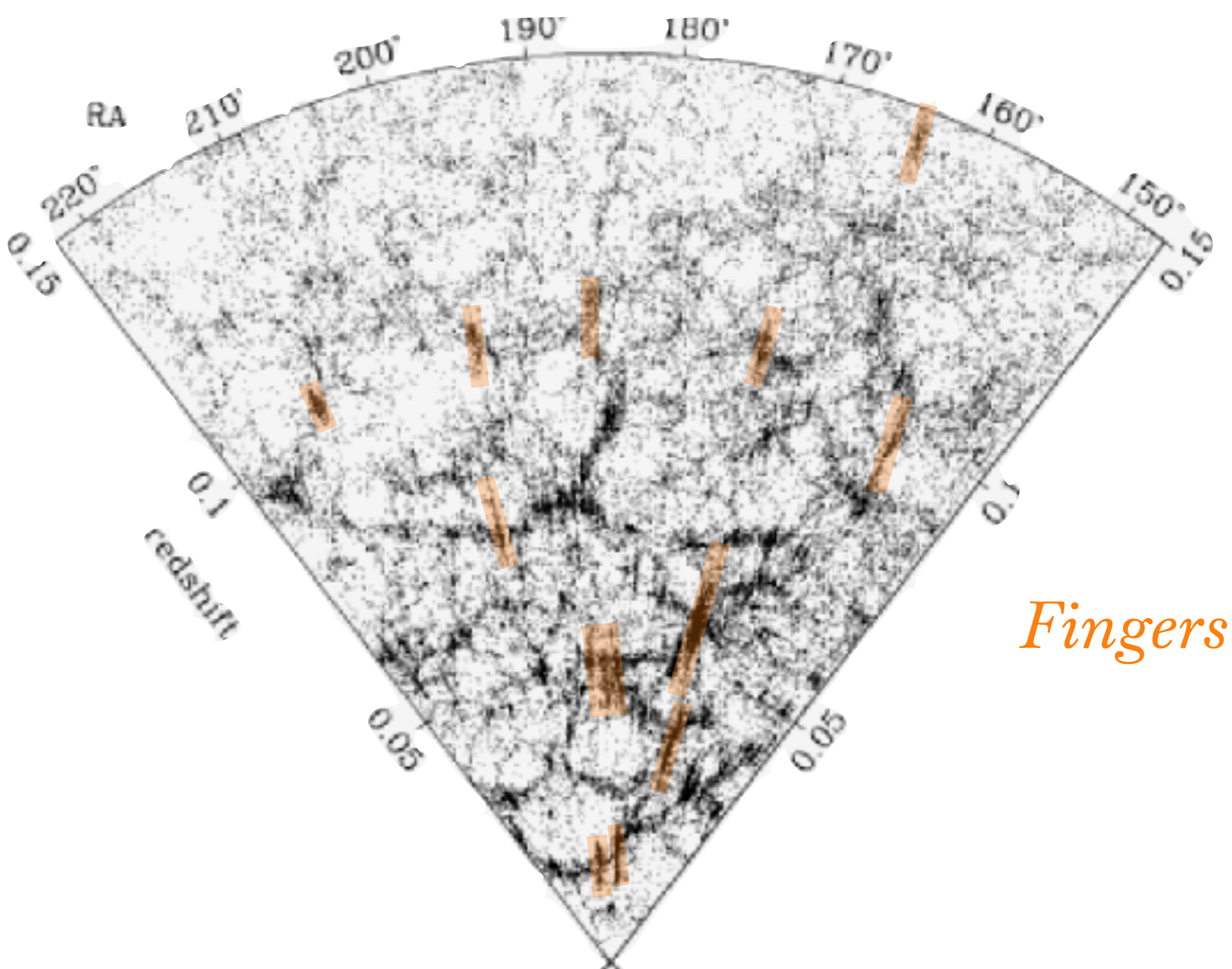


redshift-space

image: Eke et al. (2003)

3D distribution of galaxies encodes cosmological information on the **growth of structure** — *redshift-space distortions*

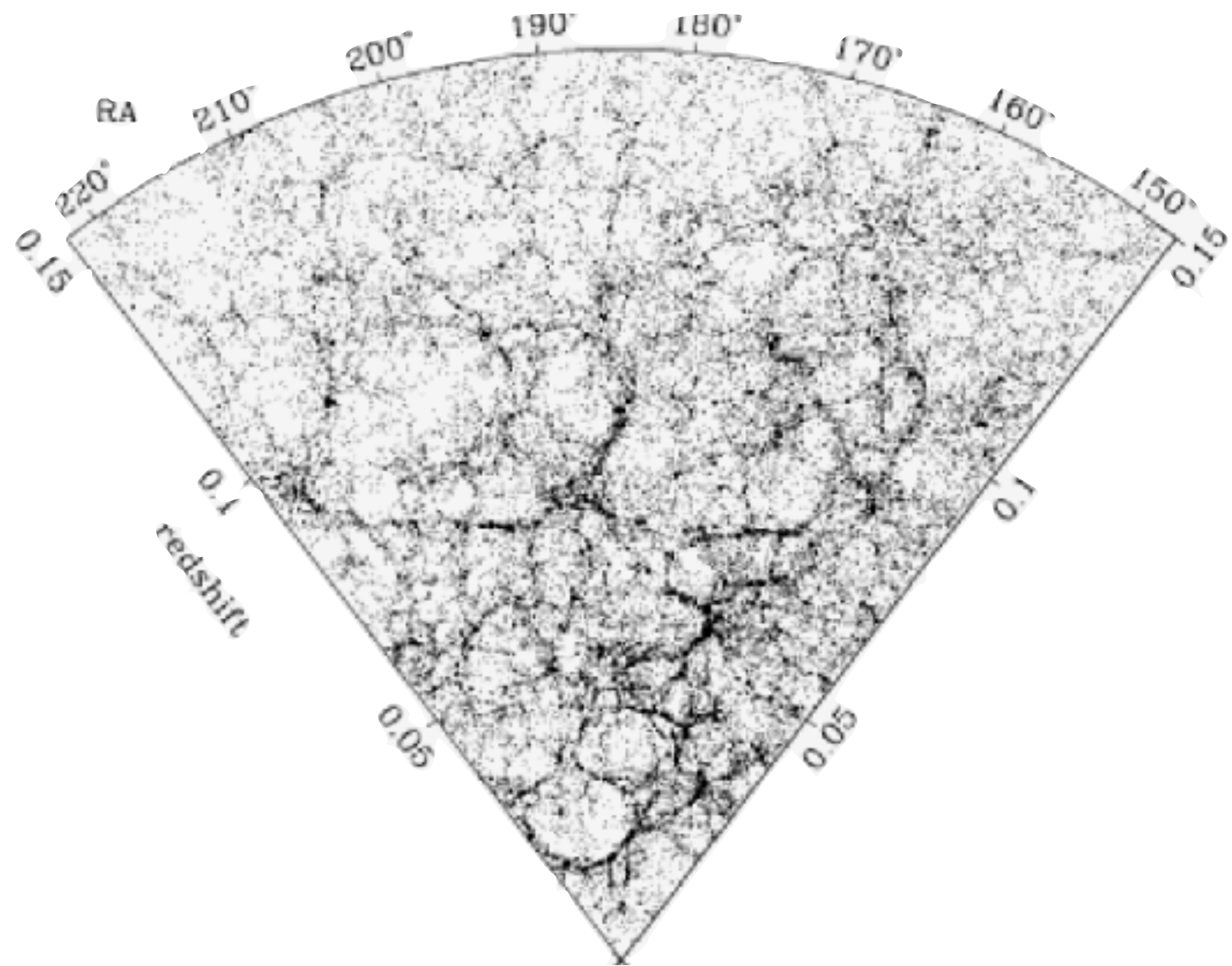
$$z_{\text{obs}} = z_{\text{cosmo}} + \frac{v_{\text{pec}}}{c}$$



Fingers-of-God

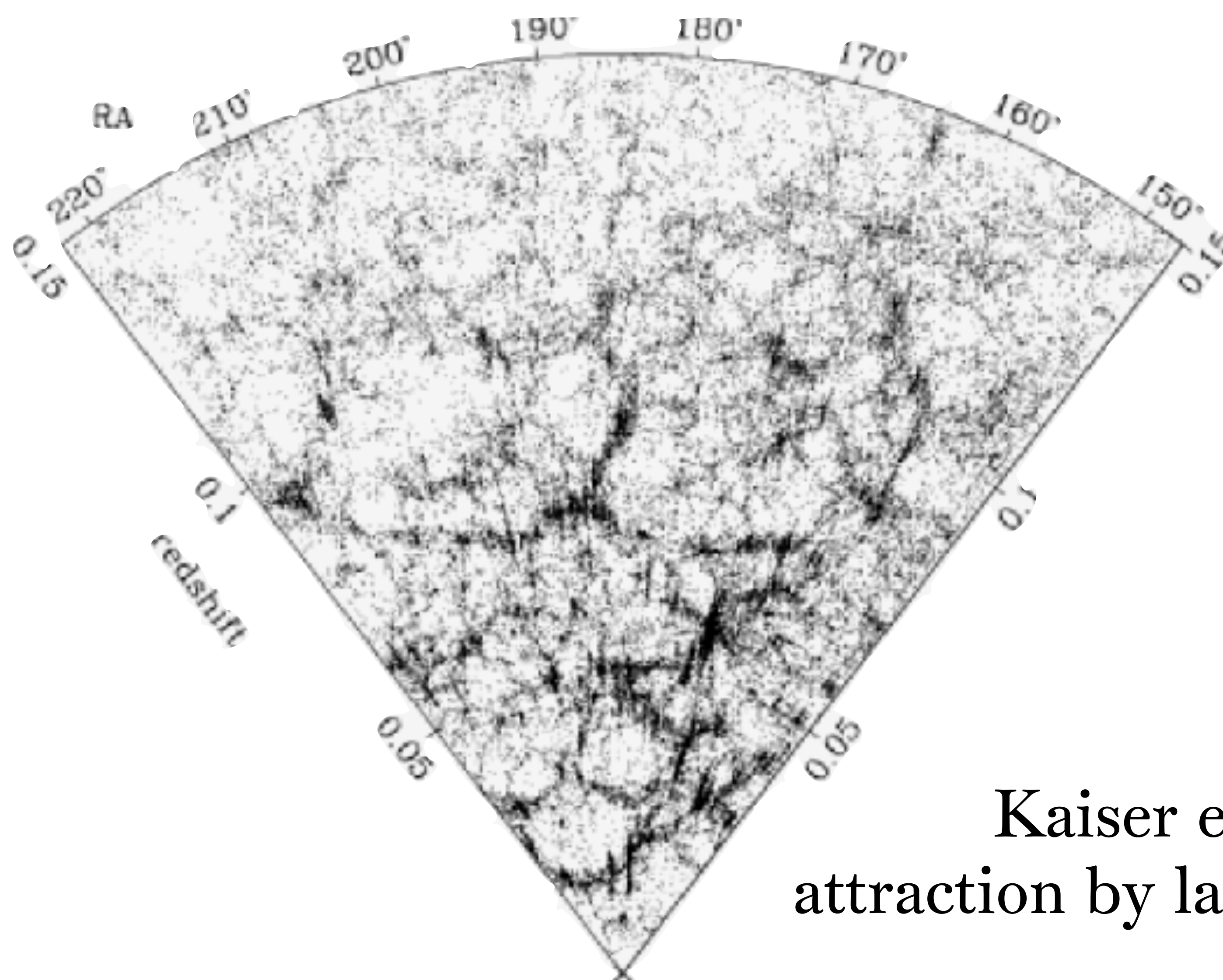
3D distribution of galaxies encodes cosmological information on the **growth of structure** — *redshift-space distortions*

$$z_{\rm obs} = z_{\rm cosmo} + \frac{v_{\rm pec}}{c}$$



3D distribution of galaxies encodes cosmological information on the **growth of structure** — *redshift-space distortions*

$$z_{\mathrm{obs}} = z_{\mathrm{cosmo}} + \frac{v_{\mathrm{pec}}}{c}$$



Kaiser effect: gravitational attraction by large-scale structure

3D distribution of galaxies encodes cosmological information on the **growth of structure** — *redshift-space distortions*

galaxy overdensity in redshift-space

$$\delta_g^{(s)}(k)$$

**linear theory*

3D distribution of galaxies encodes cosmological information on the **growth of structure** — *redshift-space distortions*

$$\delta_g^{(s)}(k) = \delta_g(k)$$

galaxy overdensity in real-space

*linear theory

3D distribution of galaxies encodes cosmological information on the **growth of structure** — *redshift-space distortions*

$$\delta_g^{(s)}(k) = \delta_g(k) + f\mu^2\,\delta_m(k)$$

growth rate of structure *matter overdensity*

**linear theory*

3D distribution of galaxies encodes cosmological information on the **growth of structure** — *redshift-space distortions*

$$\delta_g^{(s)}(k) = \delta_g(k) + f\mu^2\,\delta_m(k)$$

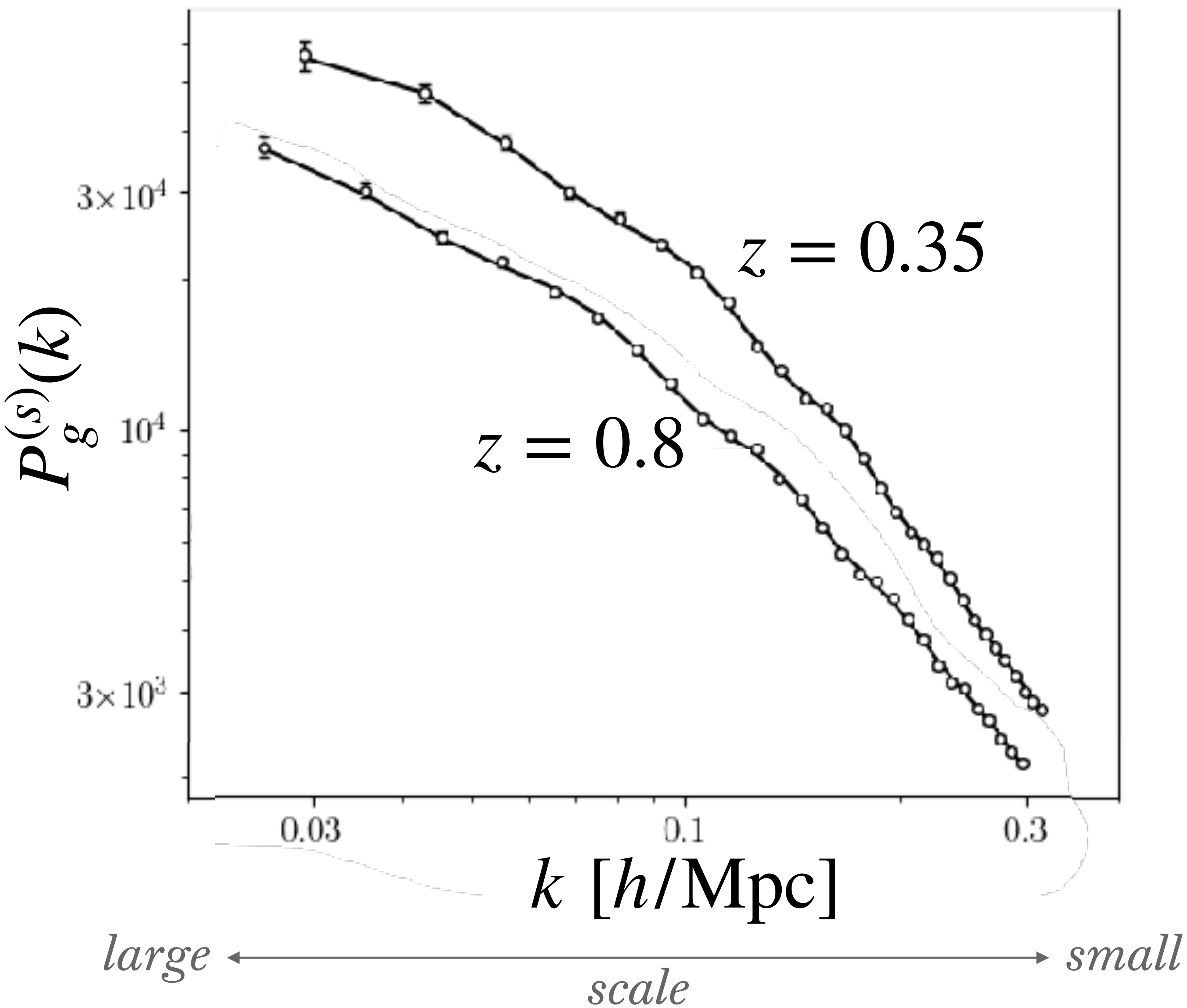
galaxy power spectrum

$$P_g^{(s)}(k) = \langle\, \delta_g^{(s)}(k)\,\delta_g^{(s)}(k')\,\rangle$$

$$= \left(b + f\mu^2\right)^2 P_m(k)$$

**linear theory*

3D distribution of galaxies encodes cosmological information on the **growth of structure** — *redshift-space distortions*

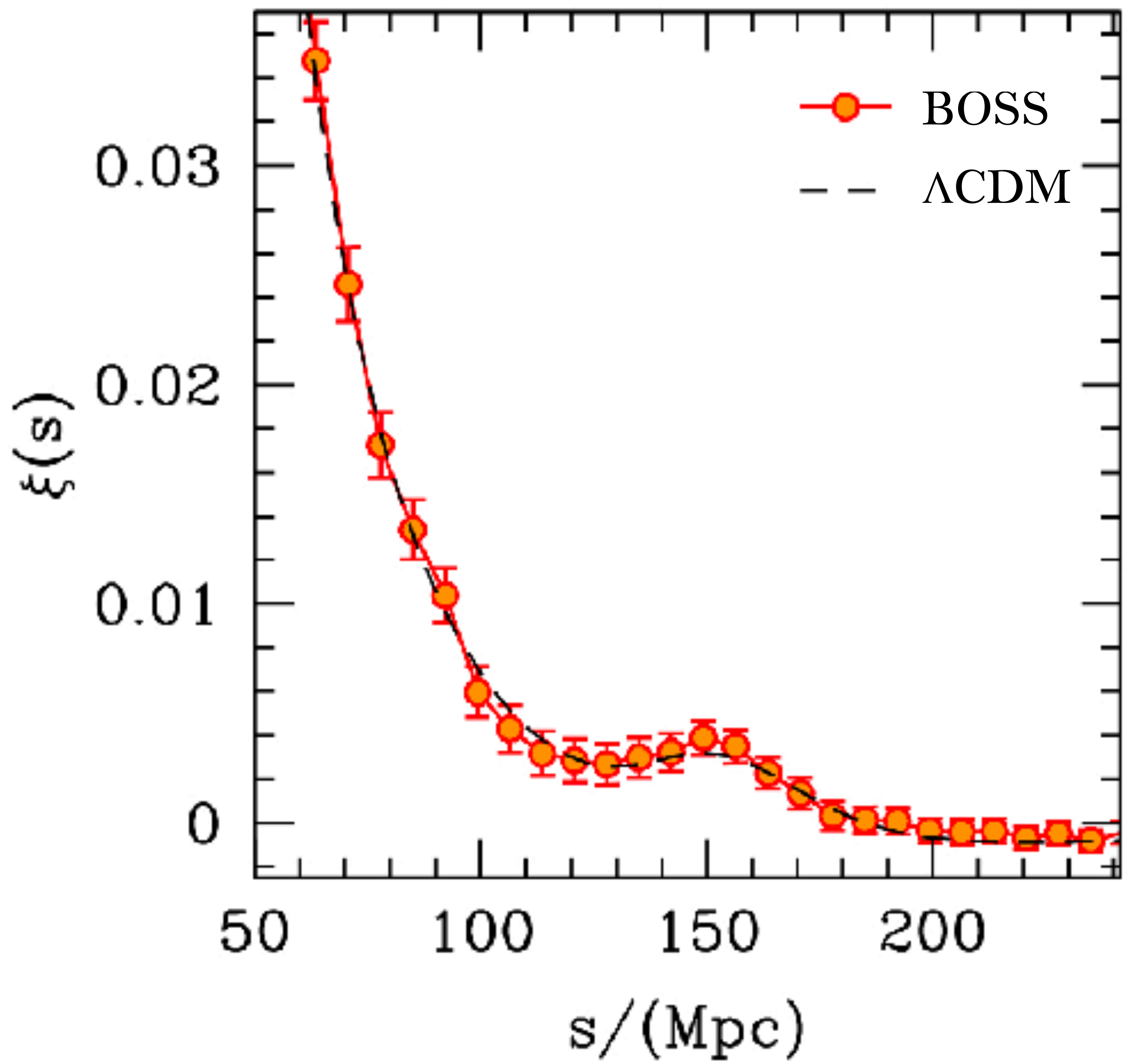


SDSS — Beutler et al.(2017), Gil-Marín et al.(2020)

3D distribution of galaxies encodes cosmological information on the **expansion history** from Baryon Acoustic Oscillations (BAO)

video: Scott Wiessinger, Ashley Balzer

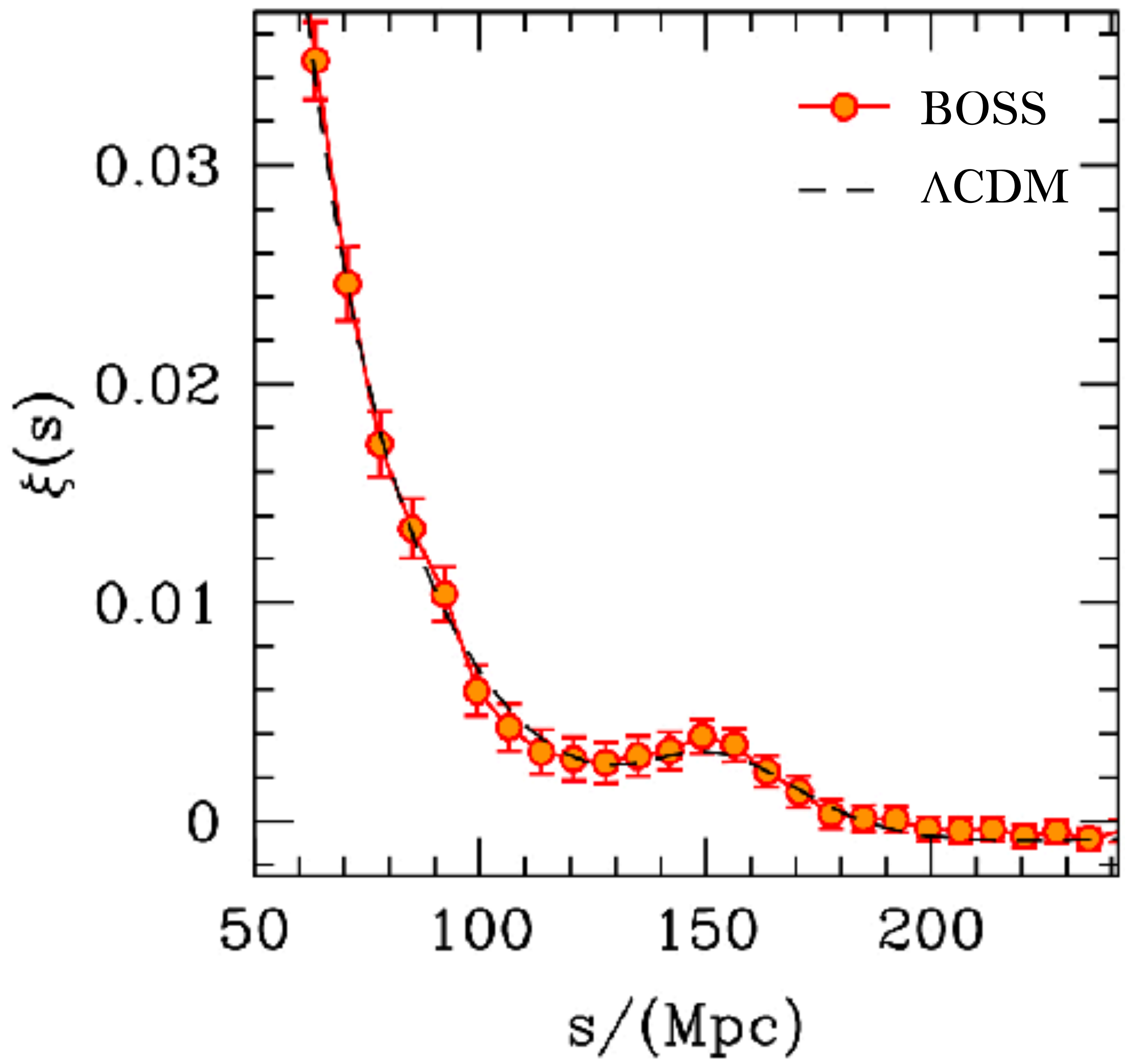
3D distribution of galaxies encodes cosmological information on the **expansion history** from Baryon Acoustic Oscillations (BAO)



2pt correlation function $\xi \equiv$ Fourier transform of P_g

BOSS — Sanchez et al. (2013)

3D distribution of galaxies encodes cosmological information on the **expansion history** from Baryon Acoustic Oscillations (BAO)



standard ruler

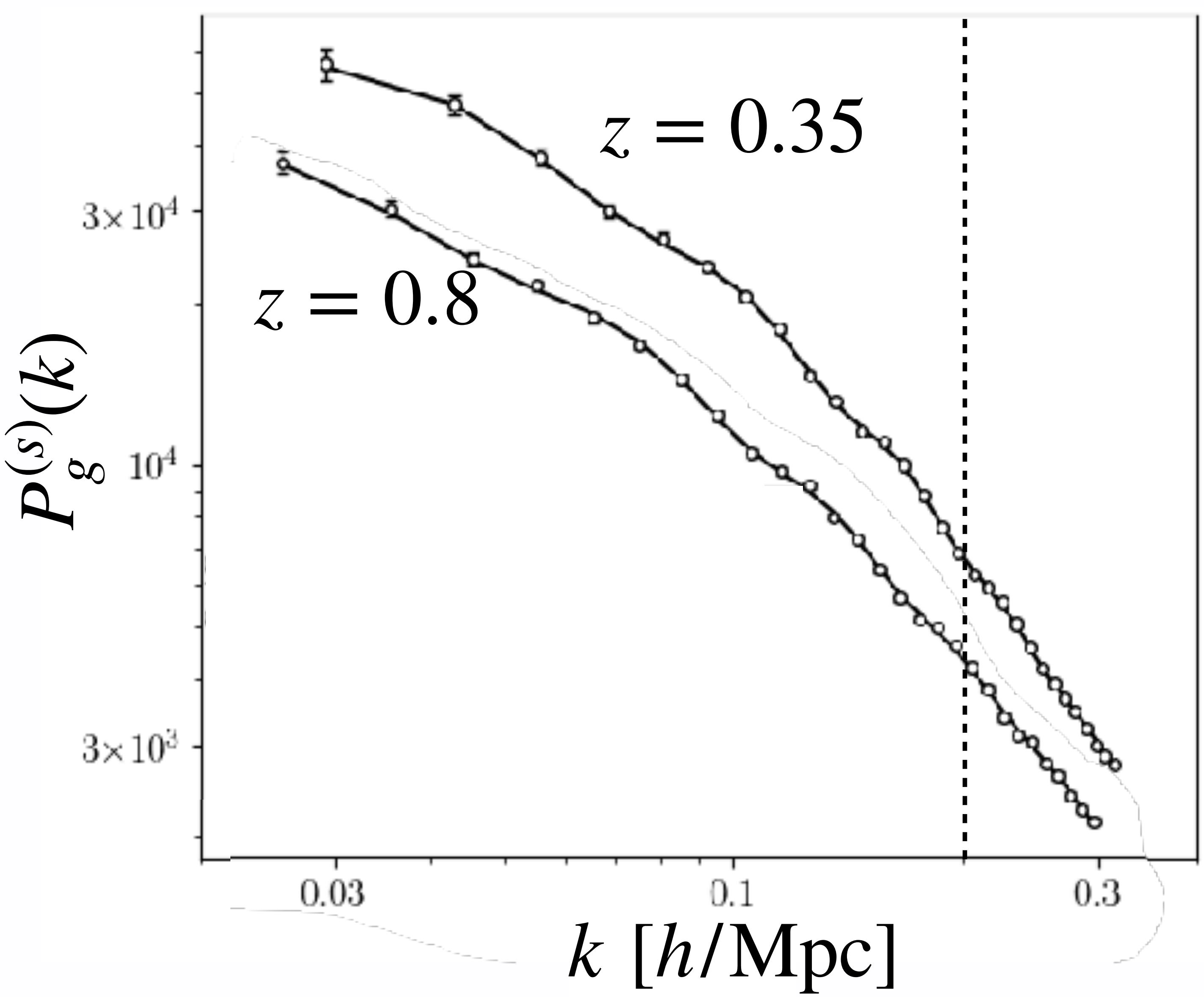
$$\theta = \frac{r_s}{d_A}$$

BOSS — Sanchez et al. (2013)

3D galaxy distribution encodes cosmological information on the **growth** and **expansion history** of the Universe

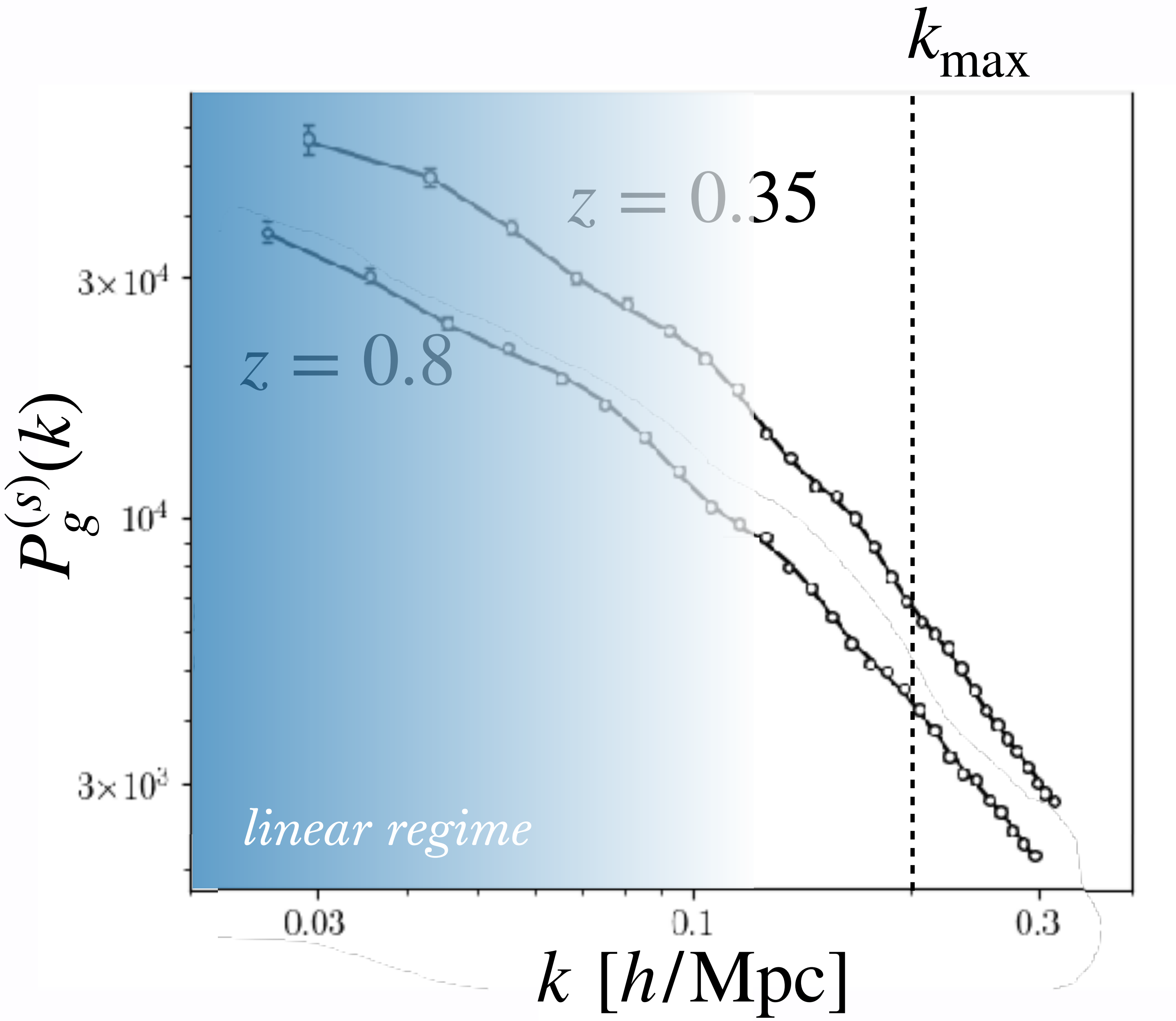

current analyses
galaxy surveys

current galaxy clustering analyses only use the **power spectrum** on large **linear scales**

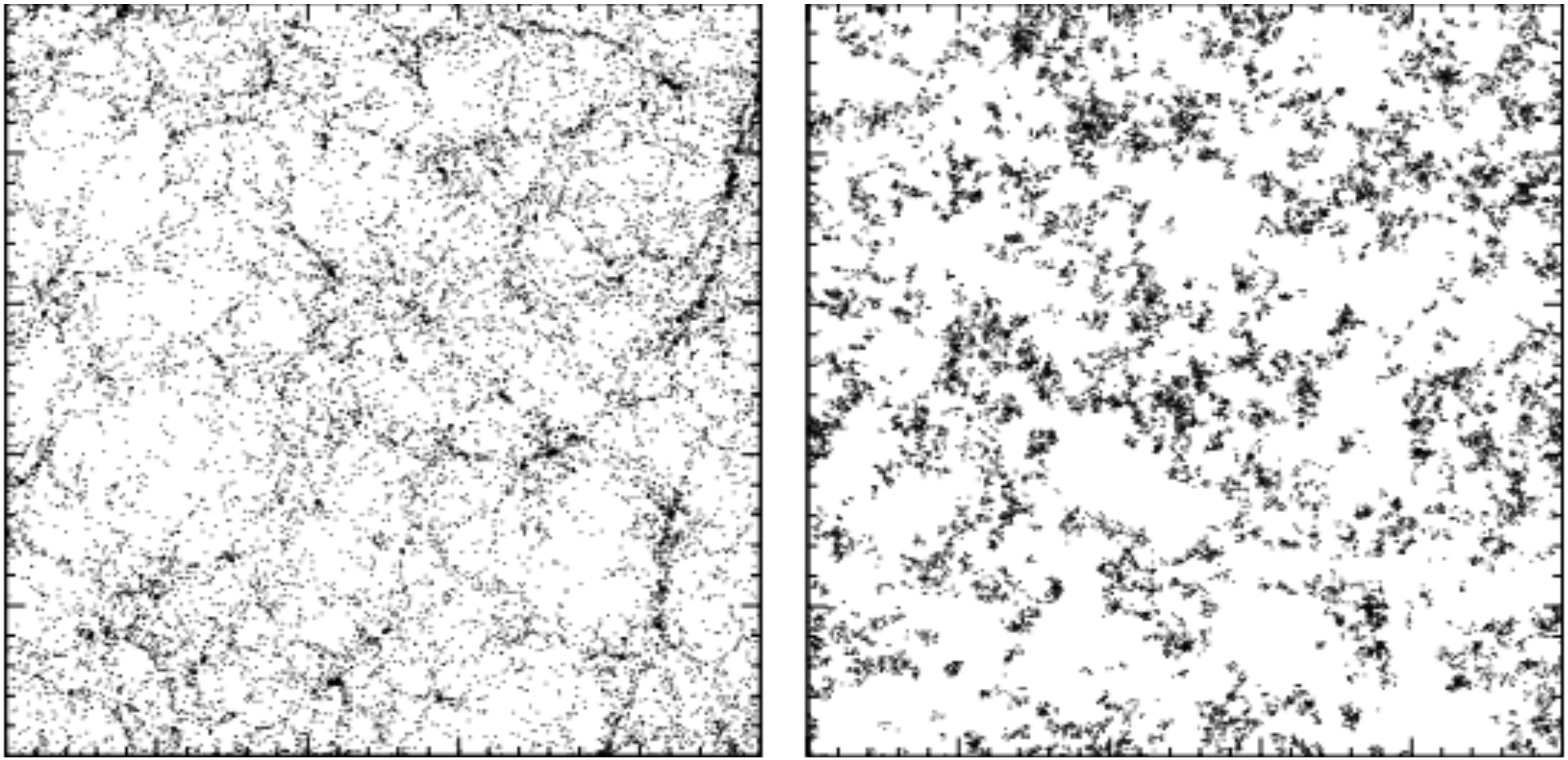


SDSS — Beutler et al.(2017), Gil-Marín et al.(2020)

current galaxy clustering analyses only use the **power spectrum** on large **linear scales**

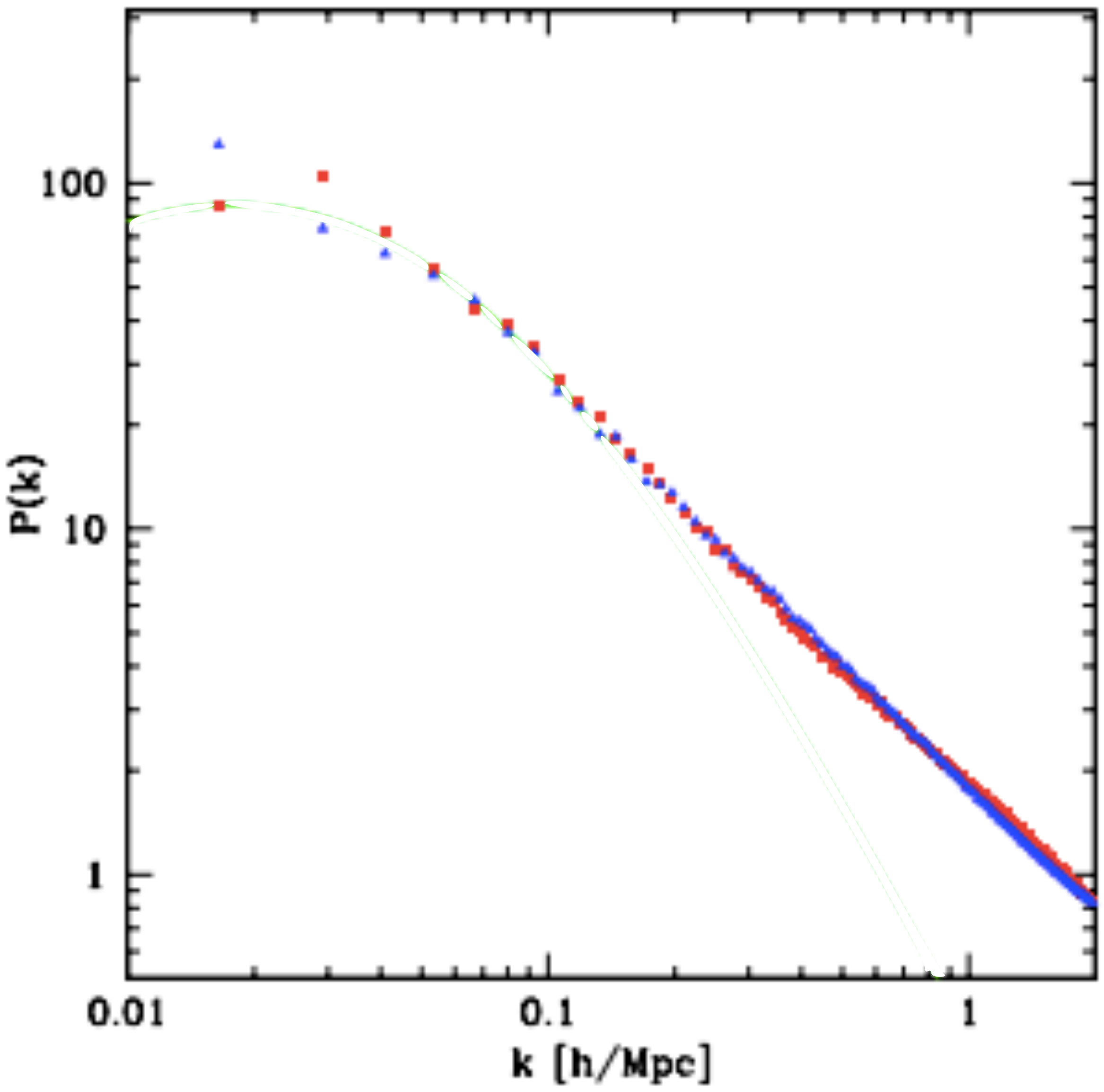


SDSS — Beutler et al.(2017), Gil-Marín et al.(2020)



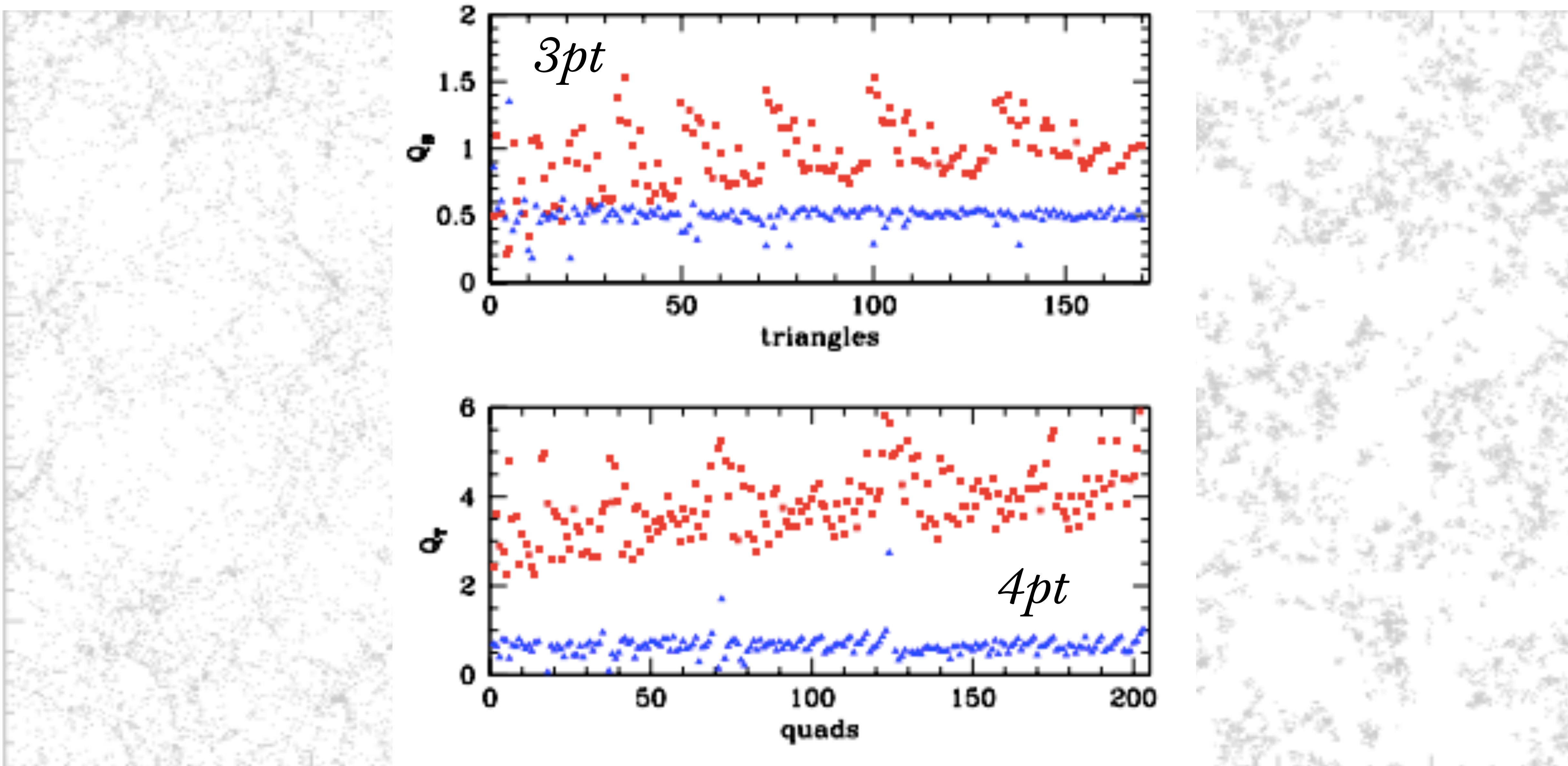
Sefusatti et al.(2005)

these two distributions have the *same* power spectrum



Sefusatti et al.(2005)

these two distributions have the *same power spectrum* but very different *higher-order clustering*



Sefusatti et al.(2005)

how much cosmological information is available *beyond the power spectrum*?

Quijote

82,000 full N-body simulations

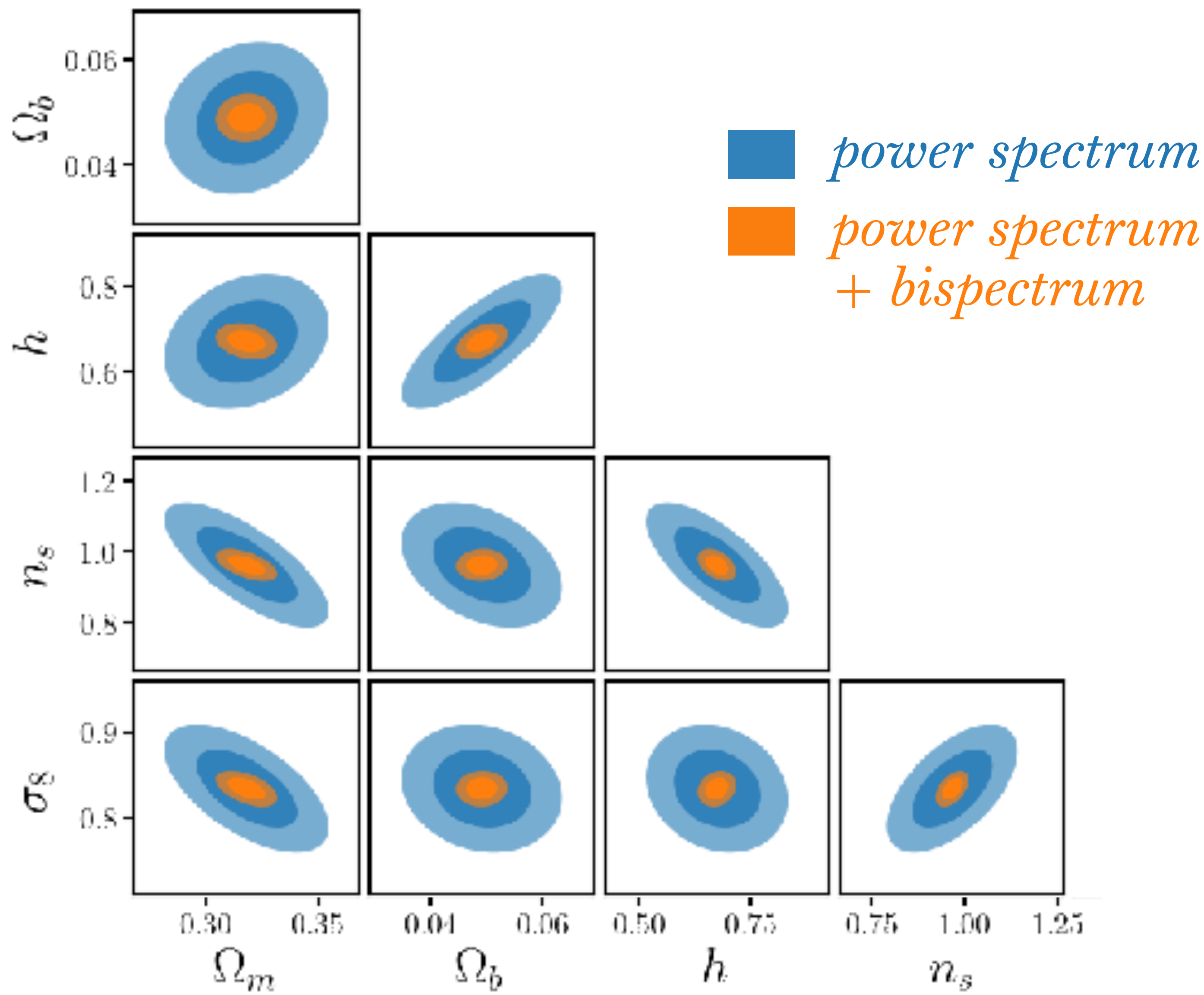
*Villaescusa-Navarro, **Hahn** et al. (2019)*

---

molino

75,000 simulated galaxy catalogs

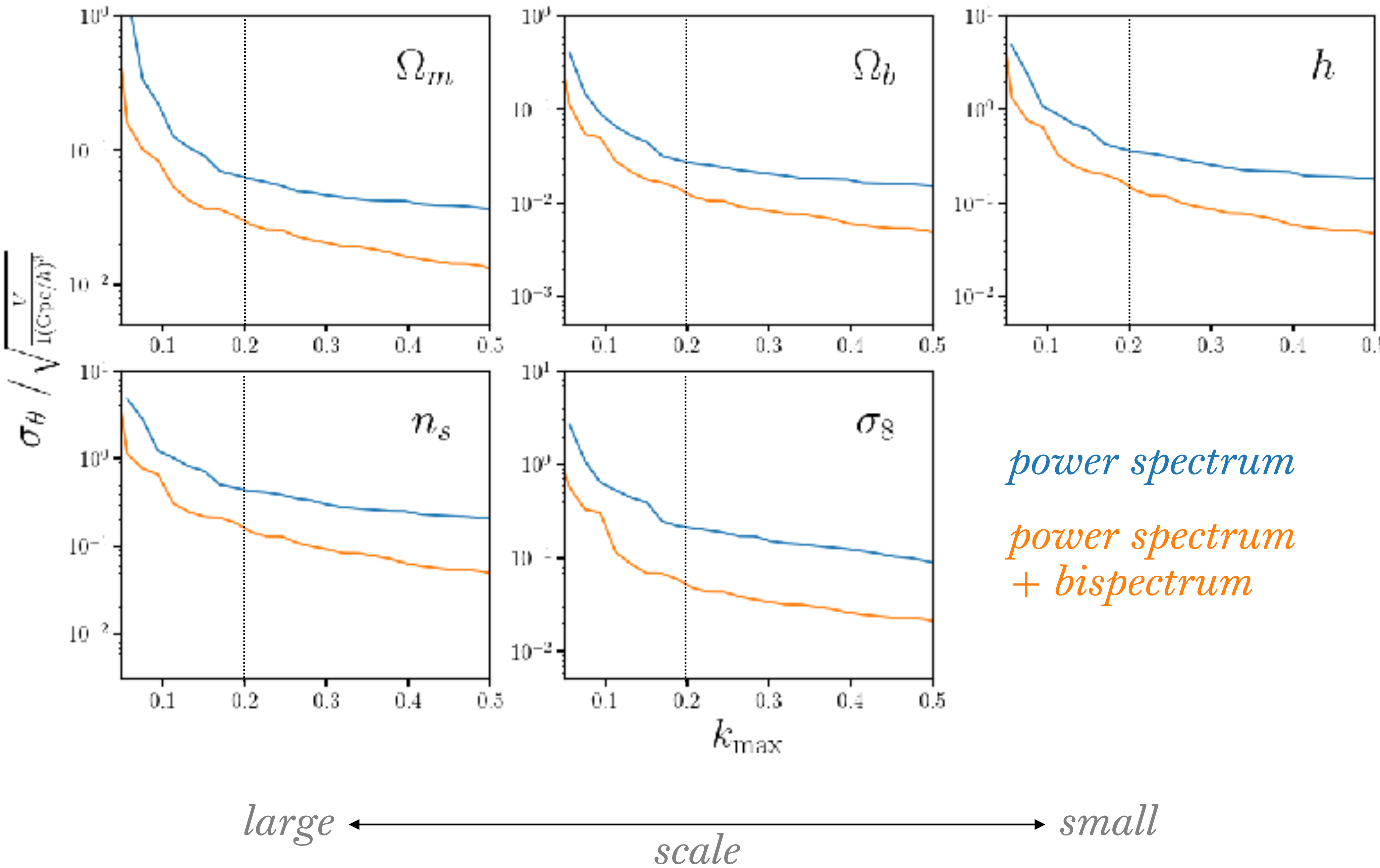
***Hahn** & Villaescusa-Navarro (2021)*



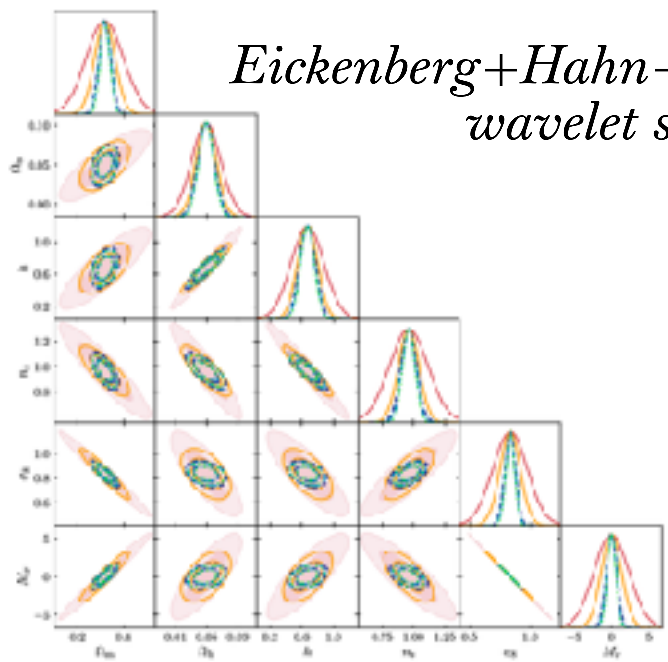
with the bispectrum we can constrain the ΛCDM parameters $(\Omega_m, \Omega_b, h, n_s, \sigma_8)$ $\gtrsim$ **3 × tighter** than the power spectrum alone

Hahn et al. (2020)
Hahn & Villaescusa-Navarro (2021)

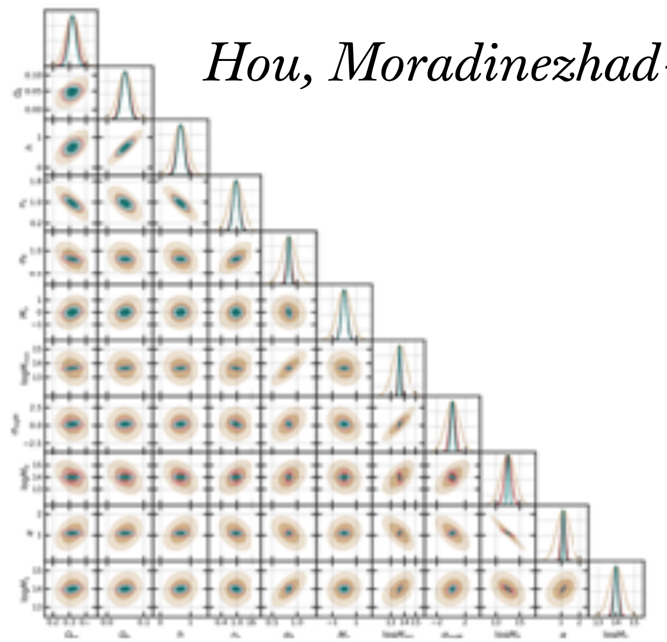
significant cosmological information on *non-linear scales*



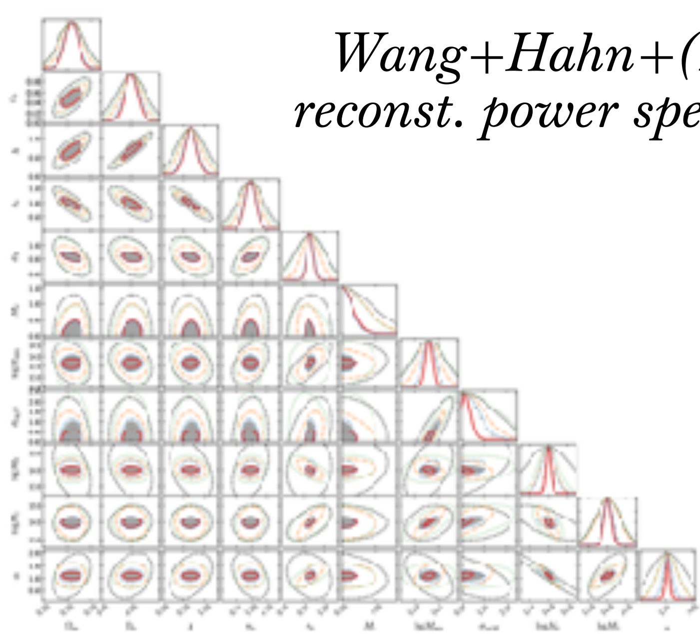
many promising clustering statistics beyond the power spectrum — *e.g.*



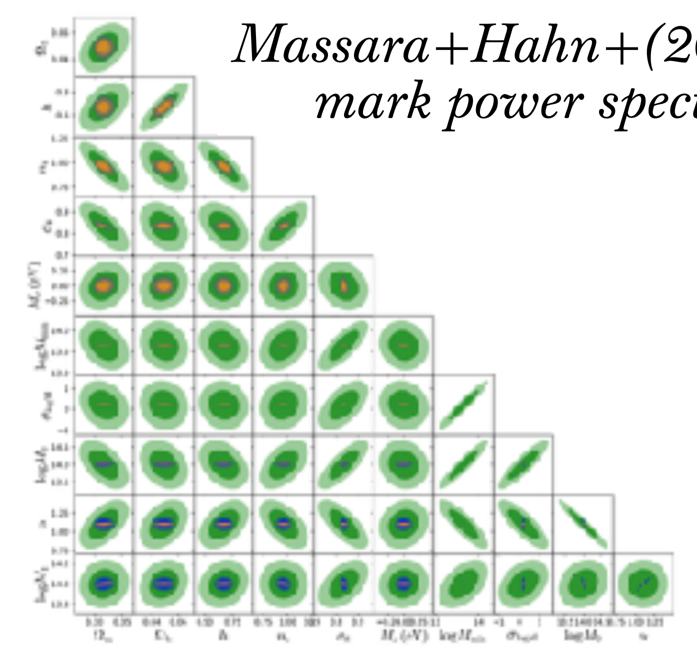
Eickenberg+Hahn+(2022)
wavelet statistics



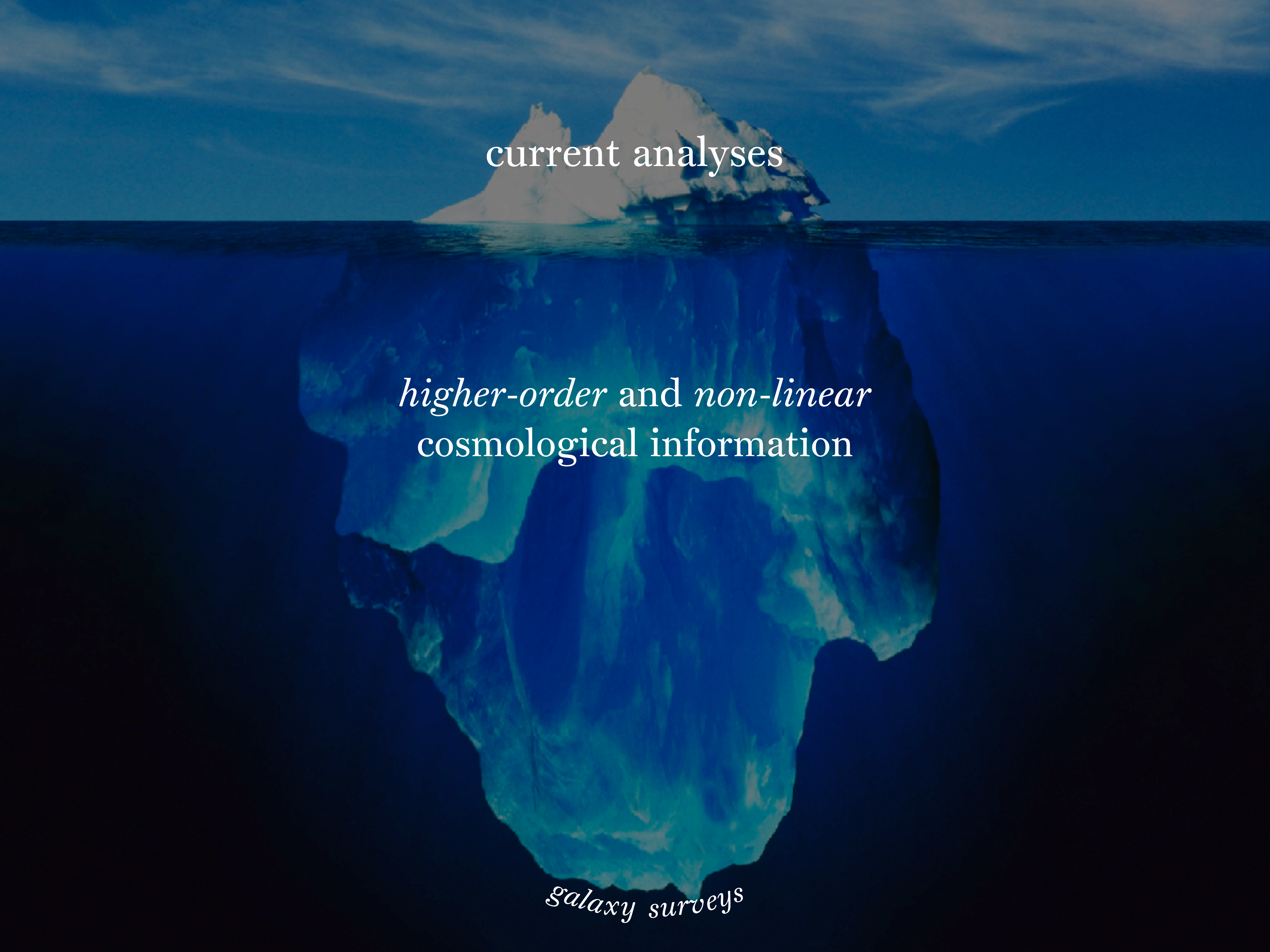
Hou, Moradinezhad+Hahn+(2022)
skew spectra



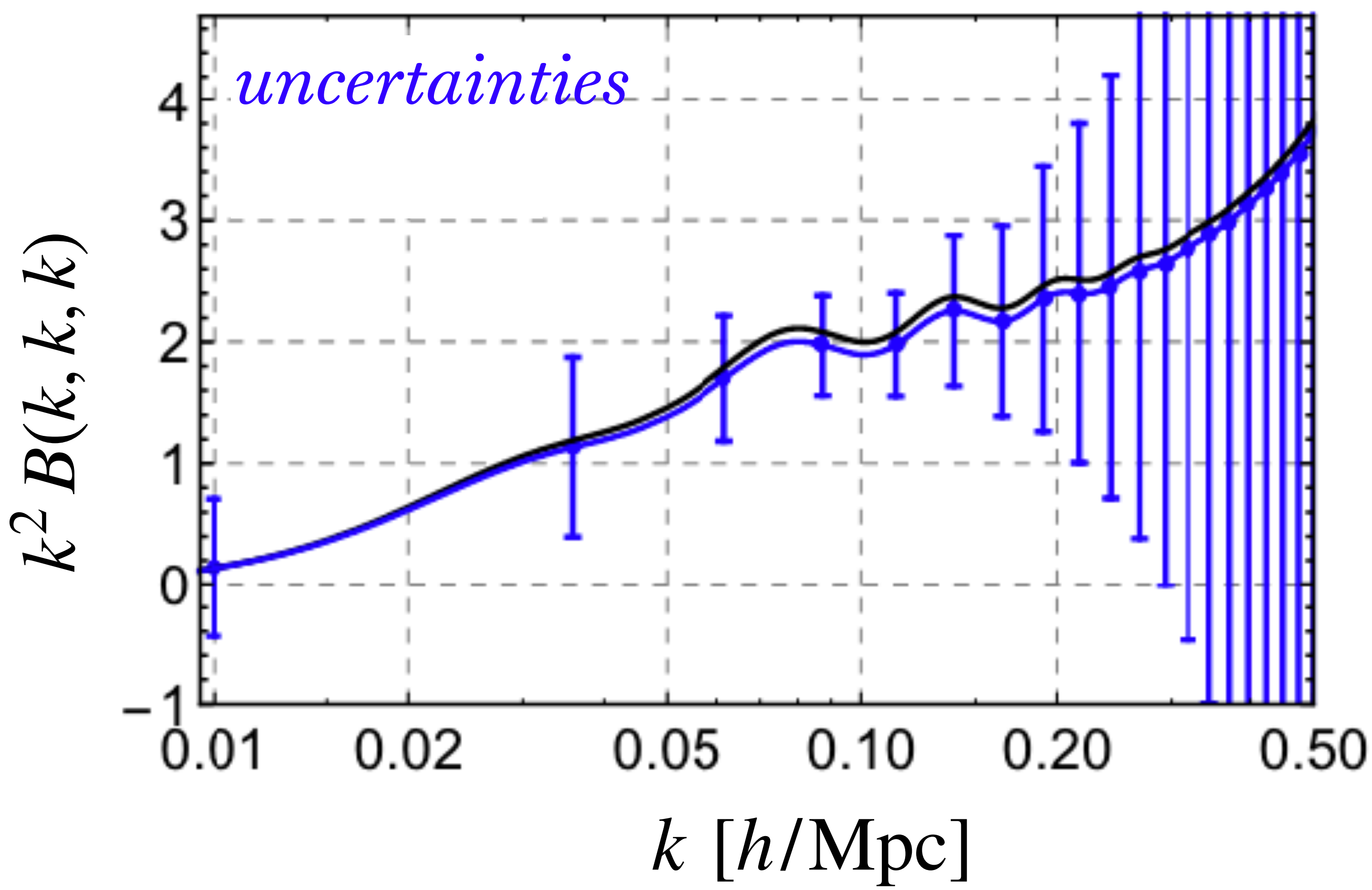
Wang+Hahn+(2024)
reconst. power spectrum



Massara+Hahn+(2022)
mark power spectrum

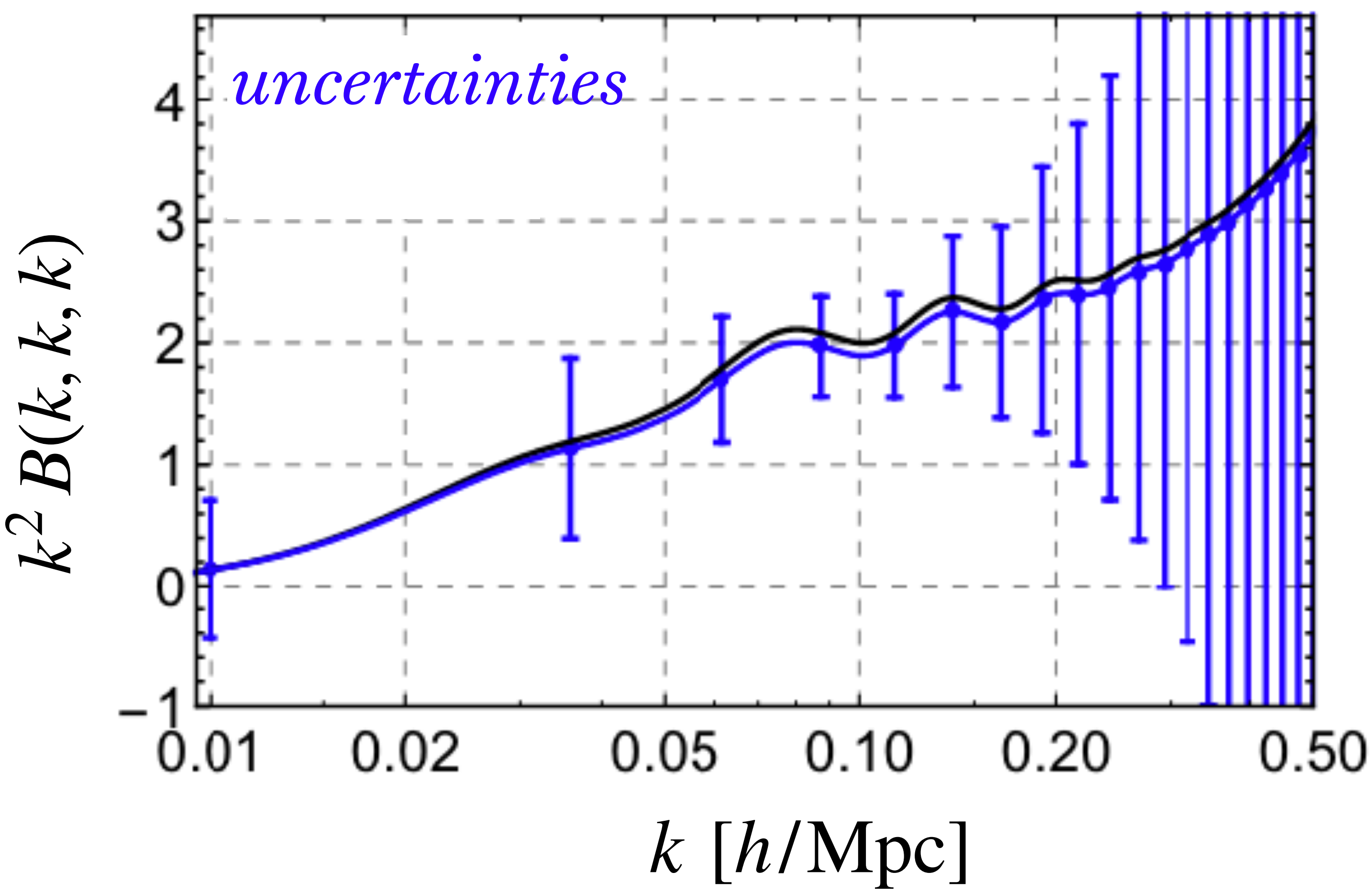

current analyses
higher-order and *non-linear*
cosmological information
galaxy surveys

challenges: analytic models of galaxy clustering are *inaccurate* on non-linear scales $k_{\text{max}} \gtrsim 0.2\, h/\text{Mpc}$



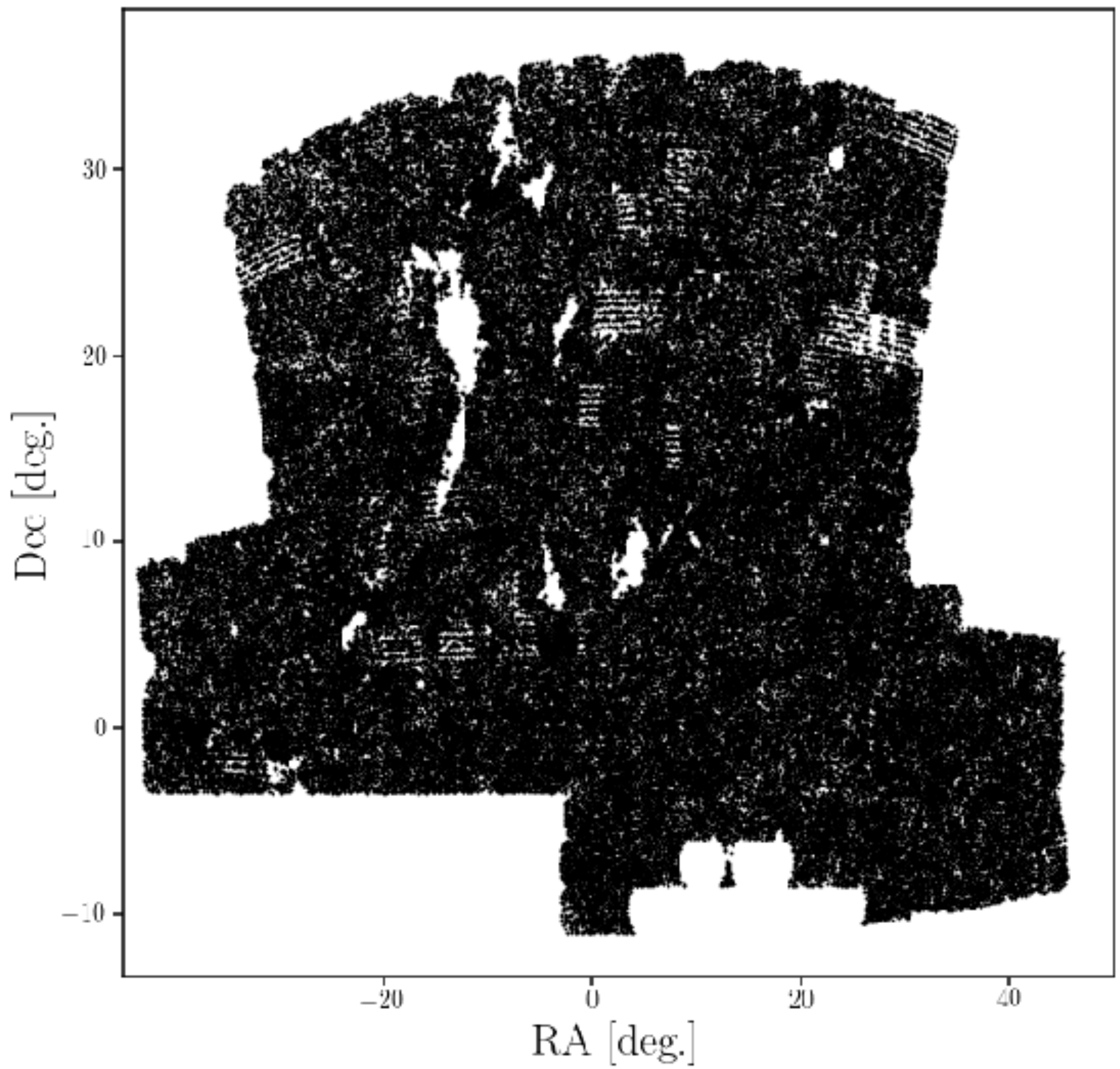
Chudaykin & Ivanov (2019)

challenges: analytic models of galaxy clustering are *inaccurate* on non-linear scales $k_{\mathrm{max}} \gtrsim 0.2\,h/\mathrm{Mpc}$



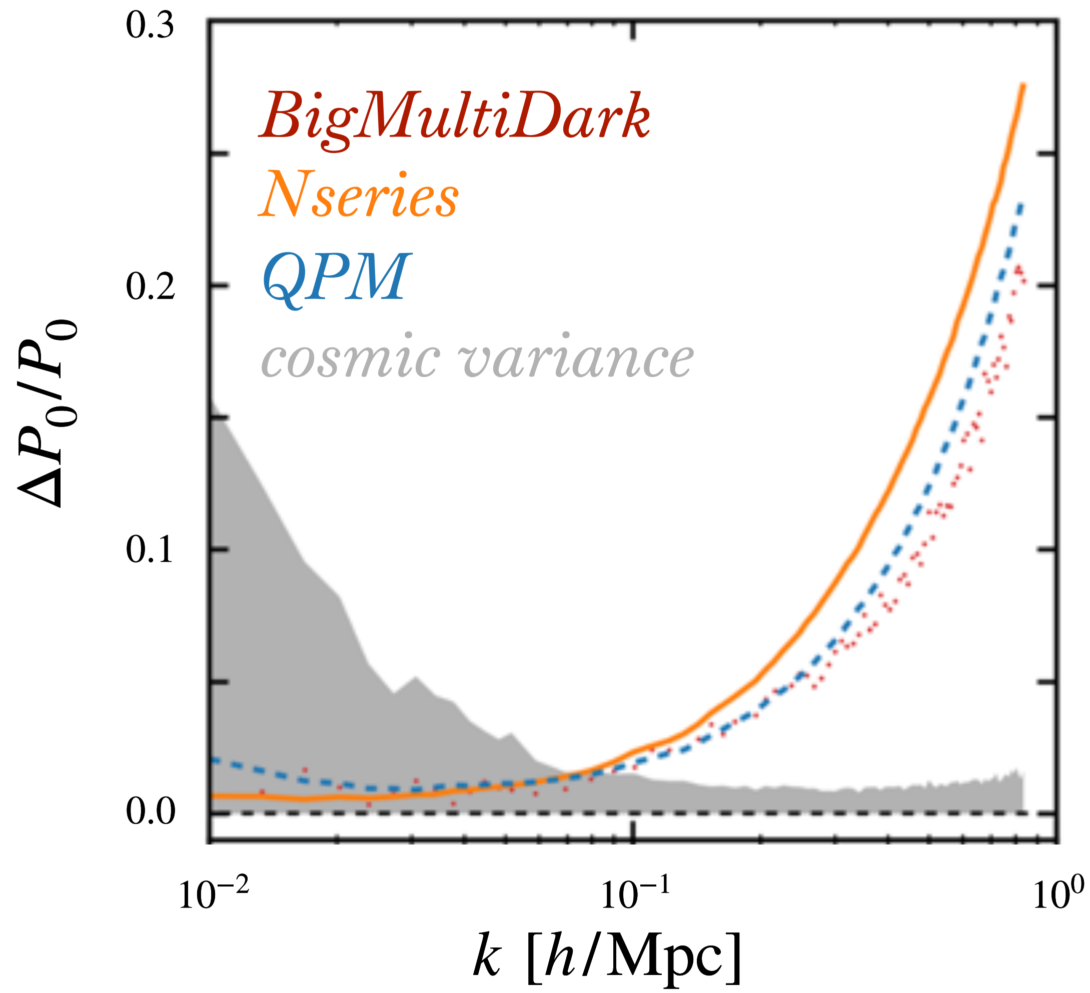
no analytic models available for — *e.g. wavelet statistics, k^{th}-nearest neighbor, minimum spanning tree…*

challenges: observations are *messy*



SDSS-III: BOSS CMASS Southern Galactic Cap

challenges: observations are *messy* — *e.g.* fiber collisions strongly affect small scale clustering



no correction scheme currently available for higher-order statistics

current challenges for clustering using higher-order statistics on non-linear scales

1. modeling *non-linear* scales

2. modeling clustering statistics *beyond* P_ℓ

3. observational *systematics*

current challenges can be addressed with a **simulation-based approach**

1. modeling *non-linear* scales

N-body simulations can accurately model small scales

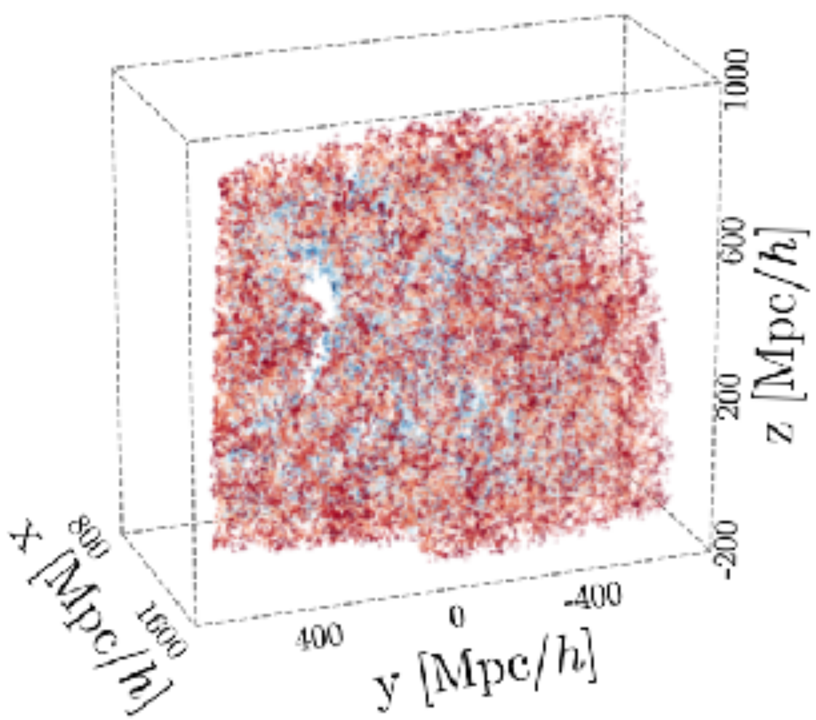
2. modeling clustering statistics *beyond* P_ℓ

we can use any statistic that can be measured in observations

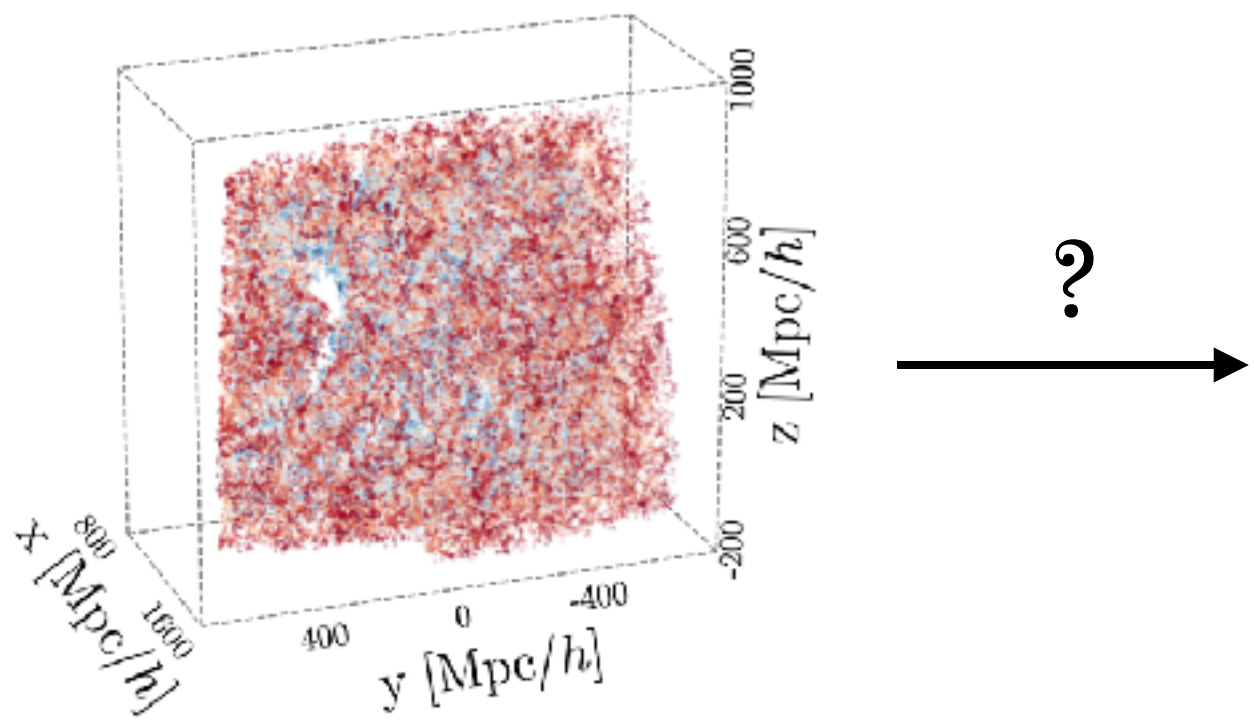
3. observational *systematics*

we already have forward models of geometry, fiber collisions, etc

image: Quijote

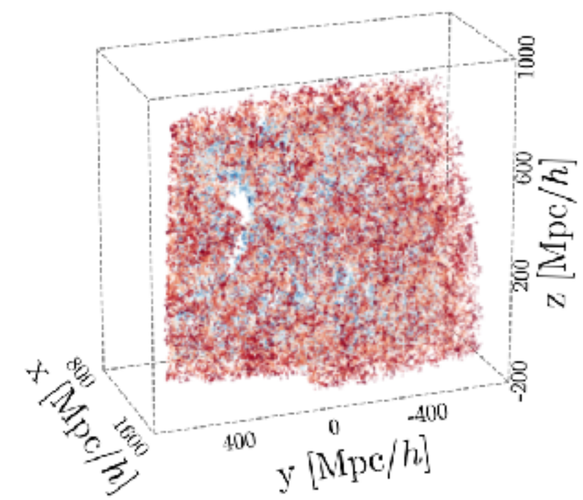


1000
600
200
-200
z [Mpc/h]
400
0
-400
y [Mpc/h]
800
1600
x [Mpc/h]



$$\theta = \left\{ \begin{matrix} \Omega_m, \Omega_b, h \\ n_s, \sigma_8 \end{matrix} \right\}$$

$$p\left(\begin{matrix} \Omega_m, \Omega_b, h \\ n_s, \sigma_8 \end{matrix} \;\middle|\; \right.$$

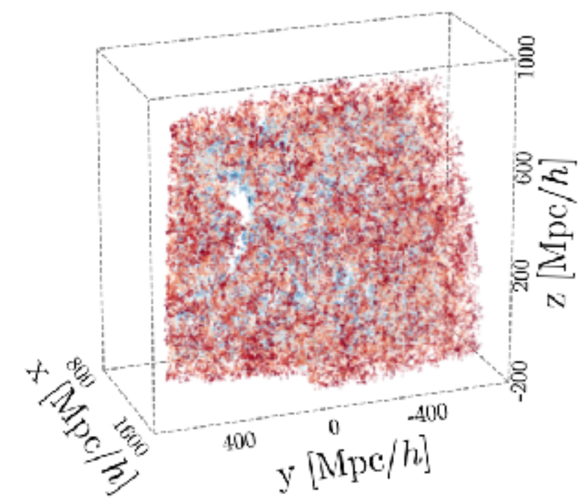


$$\left. \right)$$

ΛCDM *parameters*

galaxy distribution

$$p\left(\begin{matrix} \Omega_m, \Omega_b, h \\ n_s, \sigma_8 \end{matrix} \;\middle|\; \right.$$



$$)$$

using **simulation-based inference**

what is **simulation-based inference**?

what is **simulation-based inference**?

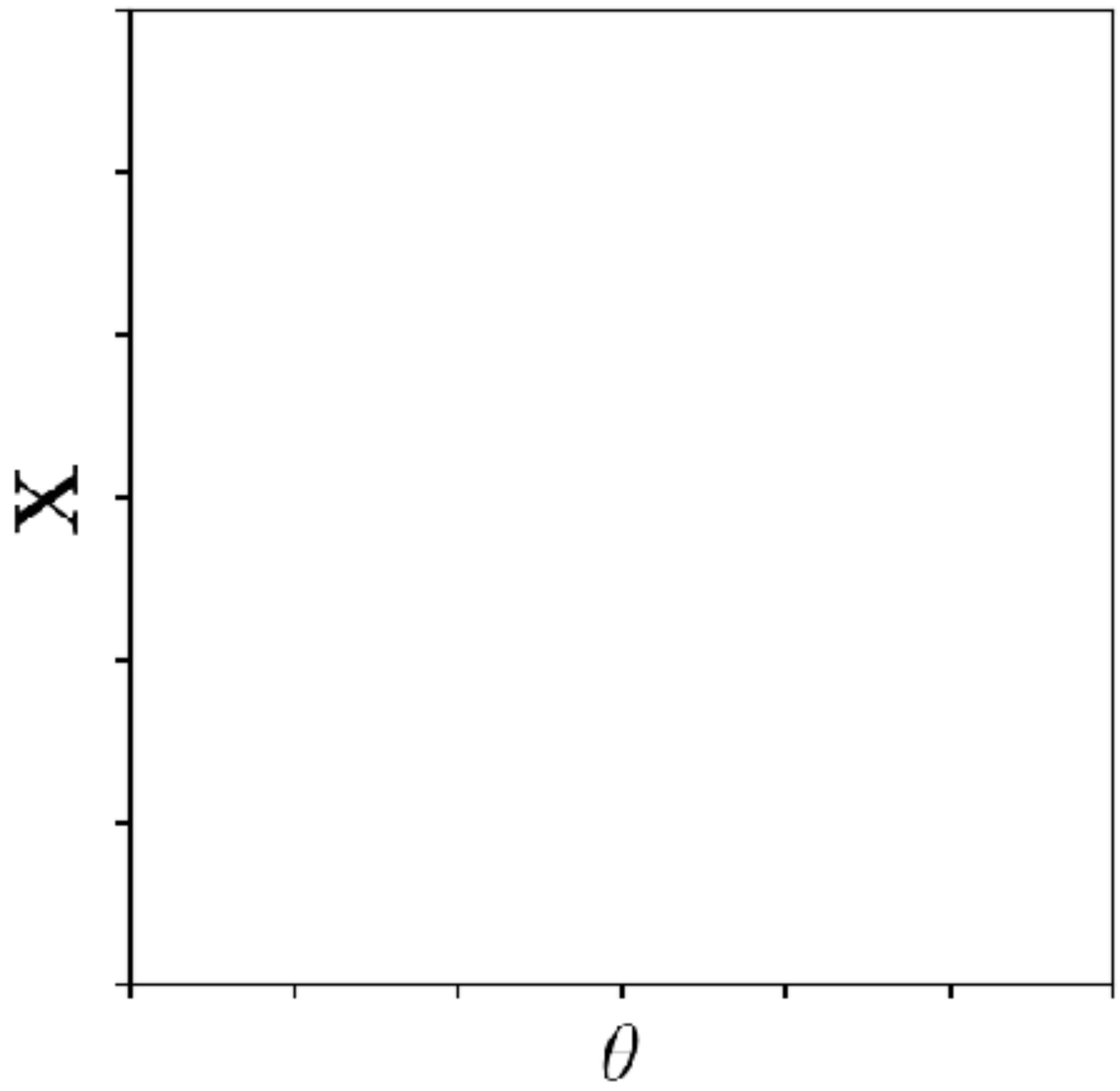
some stochastic forward model/simulator

$$\mathbf{X}' \sim F(\theta')$$

what is **simulation-based inference**?

some stochastic forward model/simulator

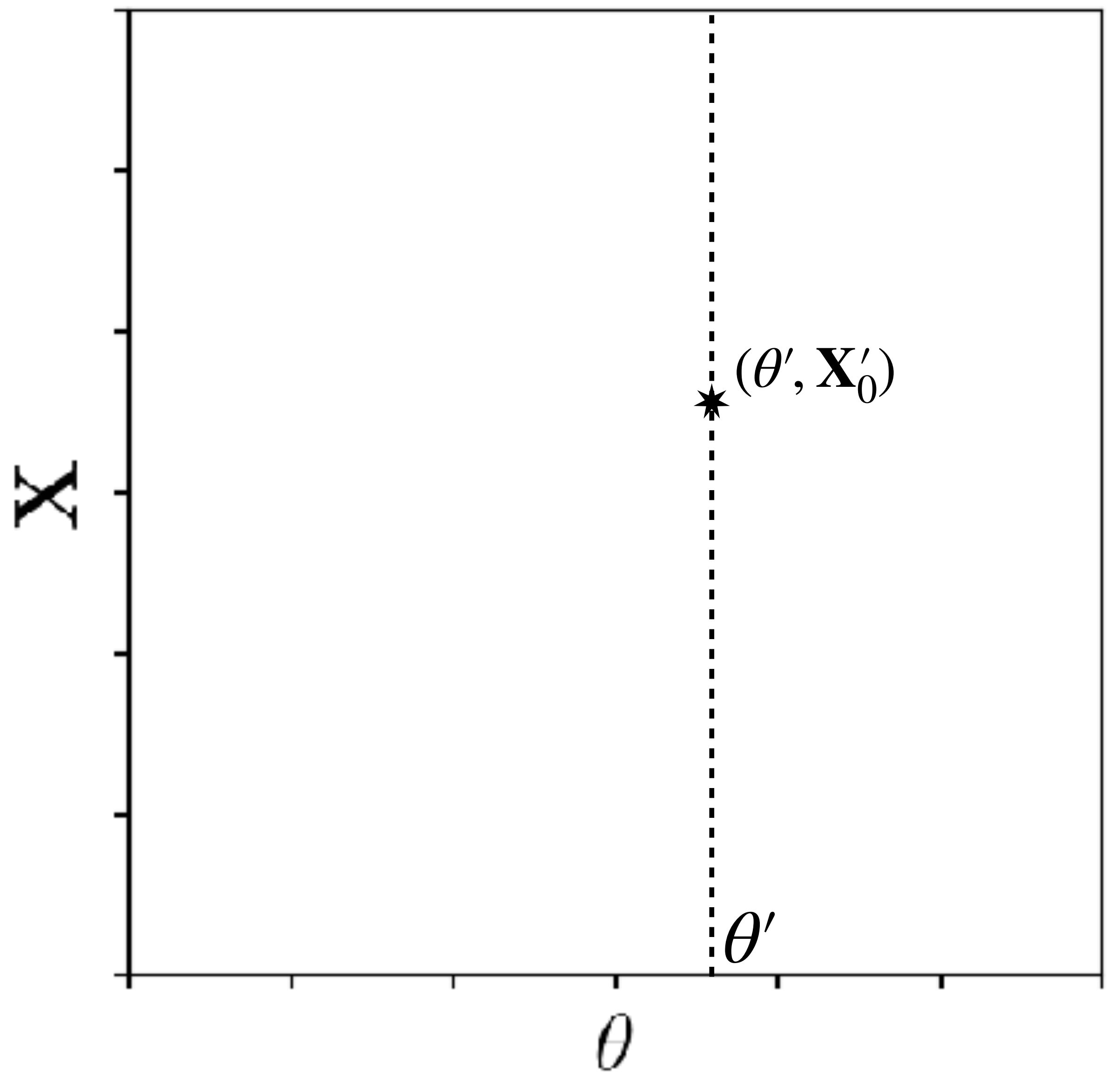
$$\mathbf{X}' \sim F(\theta')$$



what is **simulation-based inference**?

some stochastic forward model/simulator

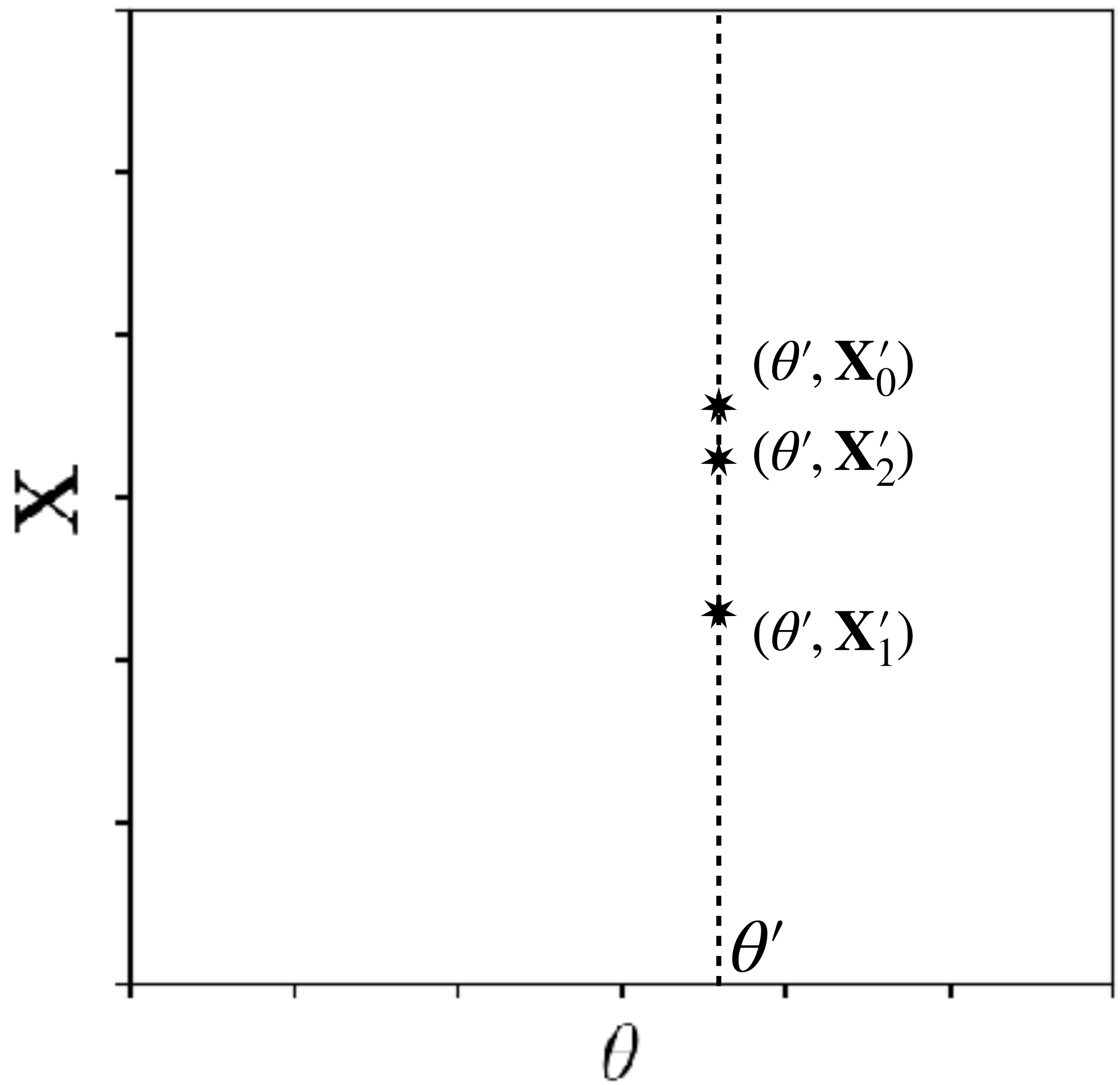
$$\mathbf{X}' \sim F(\theta')$$



what is **simulation-based inference**?

some stochastic forward model/simulator

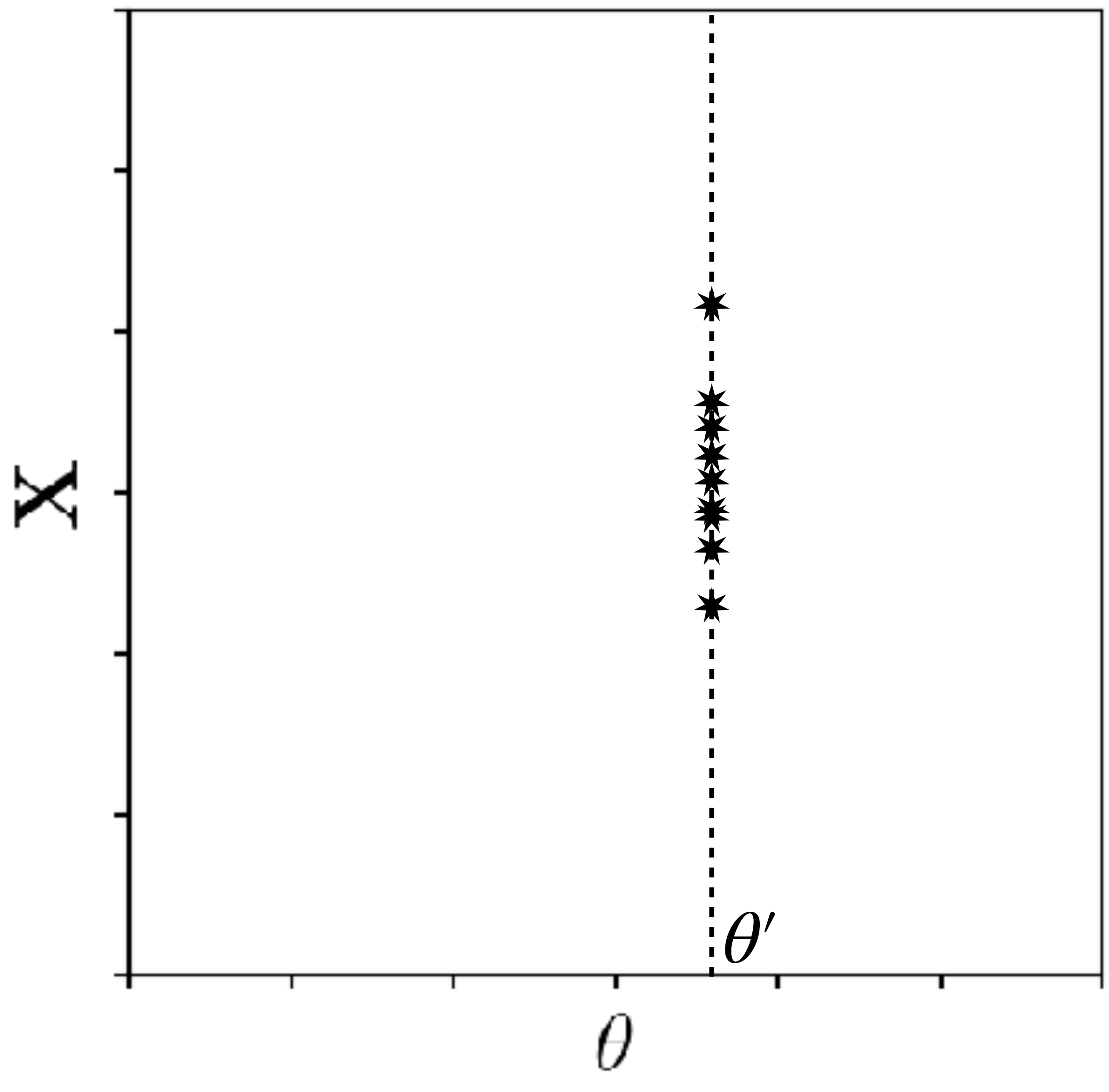
$$\mathbf{X}' \sim F(\theta')$$



what is **simulation-based inference**?

some stochastic forward model/simulator

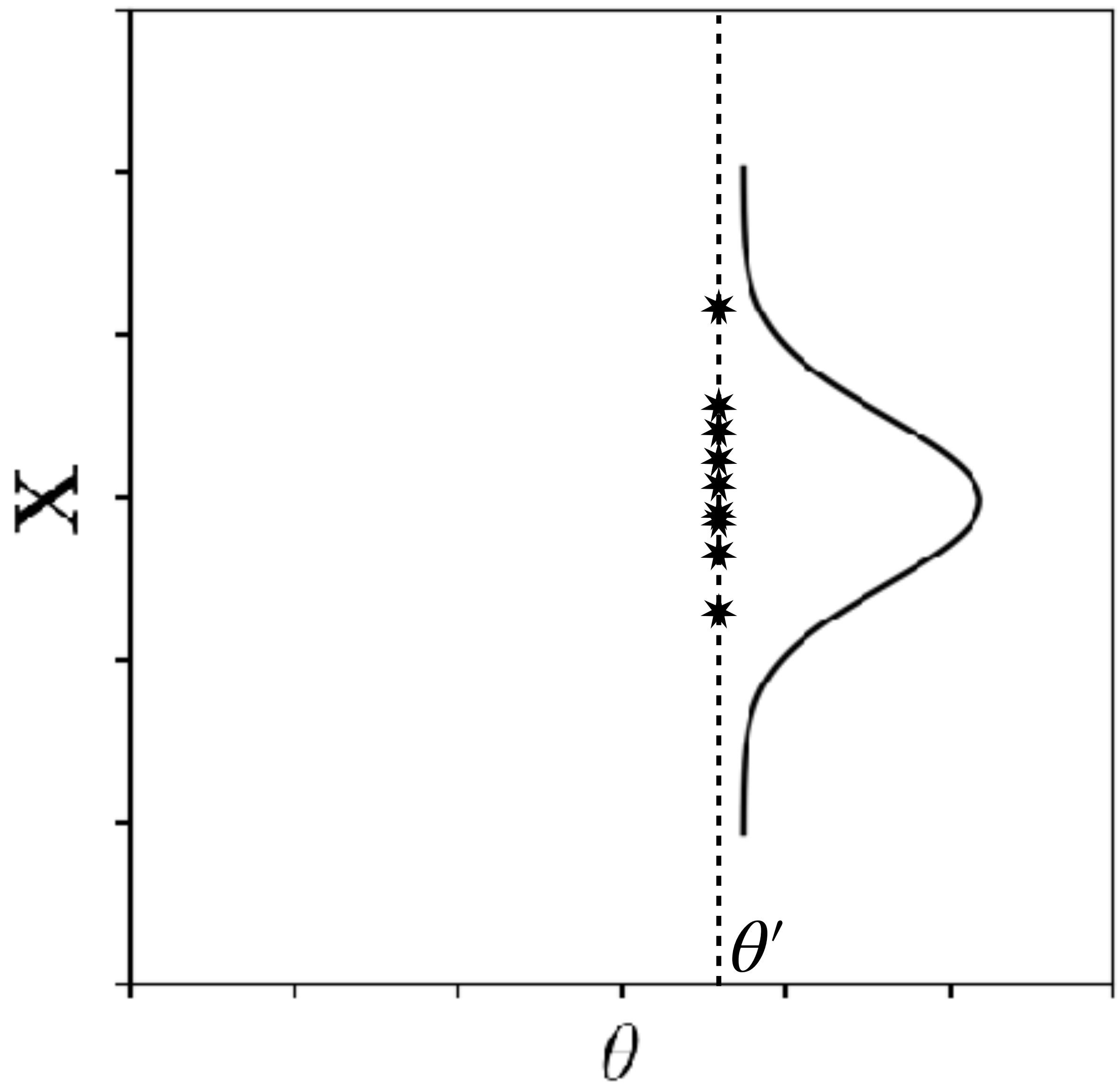
$$\mathbf{X}' \sim F(\theta')$$



what is **simulation-based inference**?

the forward model/simulator implicitly defines our likelihood

$$\mathbf{X}' \sim F(\theta') \quad \equiv \quad \mathbf{X}' \sim p(\mathbf{X} \,|\, \theta')$$



what is **simulation-based inference**?

the forward model/simulator implicitly defines our likelihood

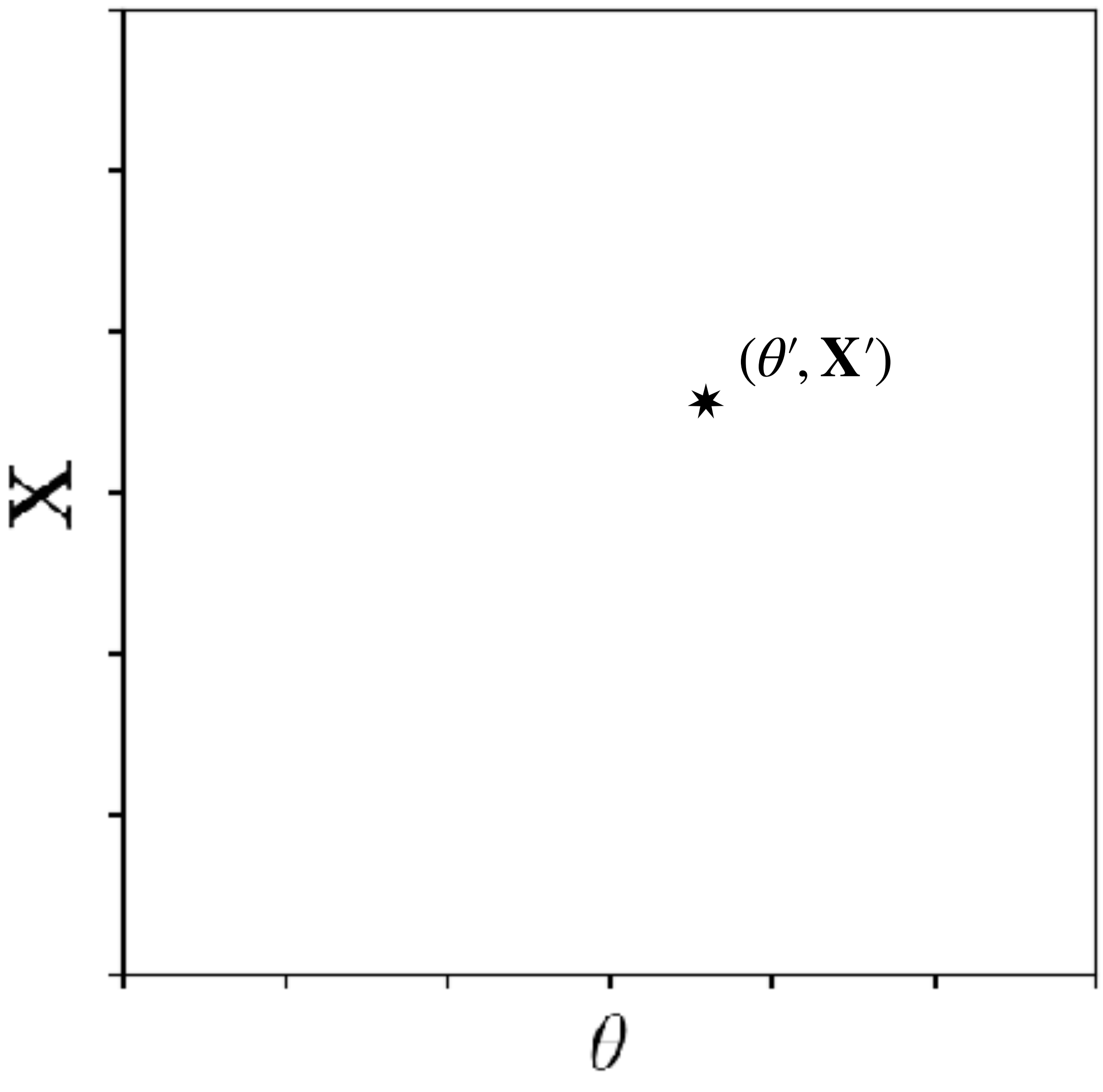
$$\mathbf{X}' \sim F(\theta') \quad \equiv \quad \mathbf{X}' \sim p(\mathbf{X} \mid \theta')$$

1. sample the prior

$$\theta' \sim p(\theta)$$

2. run simulator

$$\mathbf{X}' \sim F(\theta')$$



what is **simulation-based inference**?

the forward model/simulator implicitly defines our likelihood

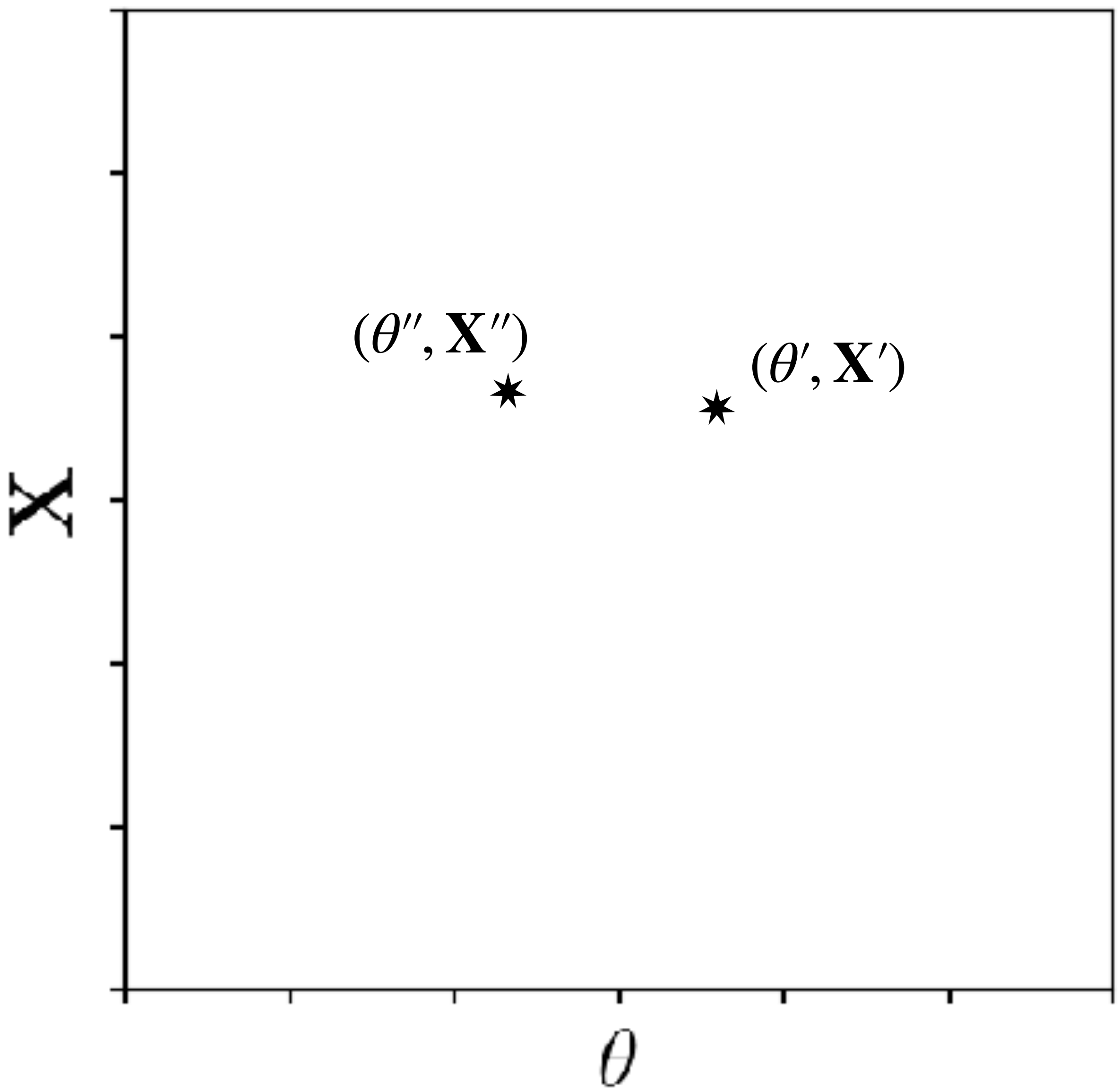
$$\mathbf{X}' \sim F(\theta') \quad \equiv \quad \mathbf{X}' \sim p(\mathbf{X} \,|\, \theta')$$

1. sample the prior

$$\theta' \sim p(\theta)$$

2. run simulator

$$\mathbf{X}' \sim F(\theta')$$



what is **simulation-based inference**?

the forward model/simulator implicitly defines our likelihood

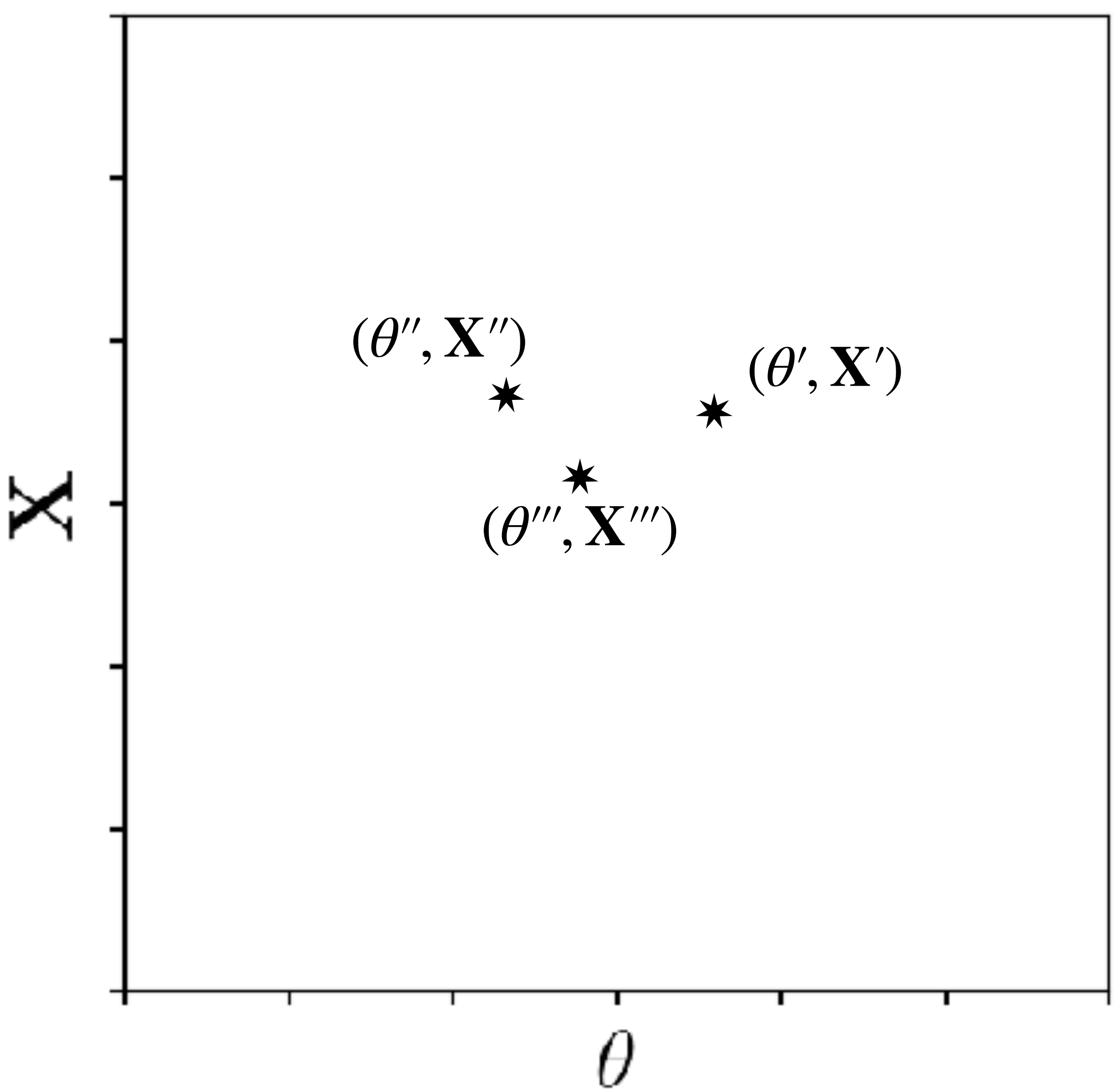
$$\mathbf{X}' \sim F(\theta') \quad \equiv \quad \mathbf{X}' \sim p(\mathbf{X} \,|\, \theta')$$

1. sample the prior

$$\theta' \sim p(\theta)$$

2. run simulator

$$\mathbf{X}' \sim F(\theta')$$



what is **simulation-based inference**?

the forward model/simulator implicitly defines our likelihood

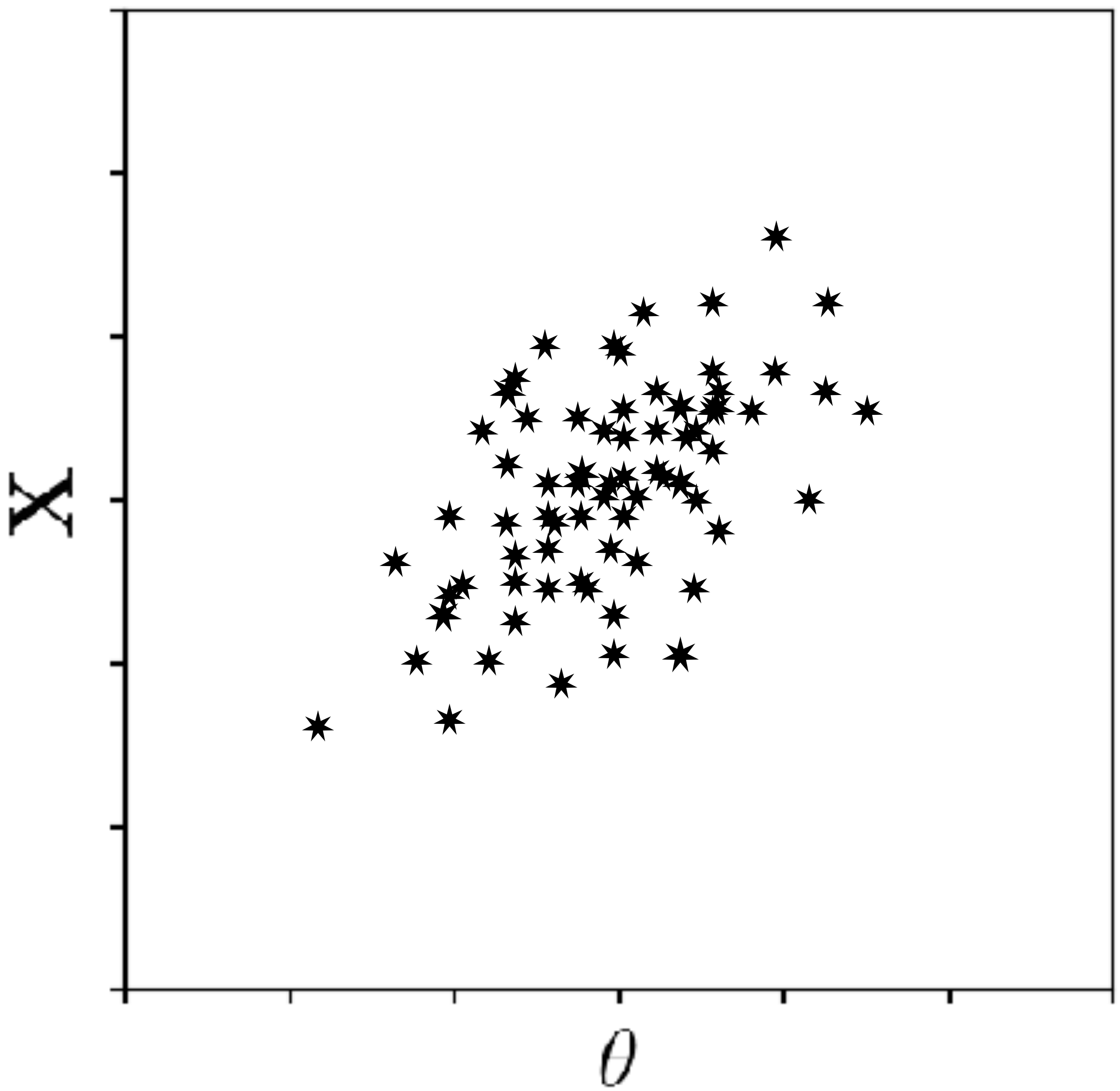
$$\mathbf{X}' \sim F(\theta') \quad \equiv \quad \mathbf{X}' \sim p(\mathbf{X} \mid \theta')$$

1. sample the prior

$$\theta' \sim p(\theta)$$

2. run simulator

$$\mathbf{X}' \sim F(\theta')$$



what is **simulation-based inference**?

the forward model/simulator implicitly defines our likelihood

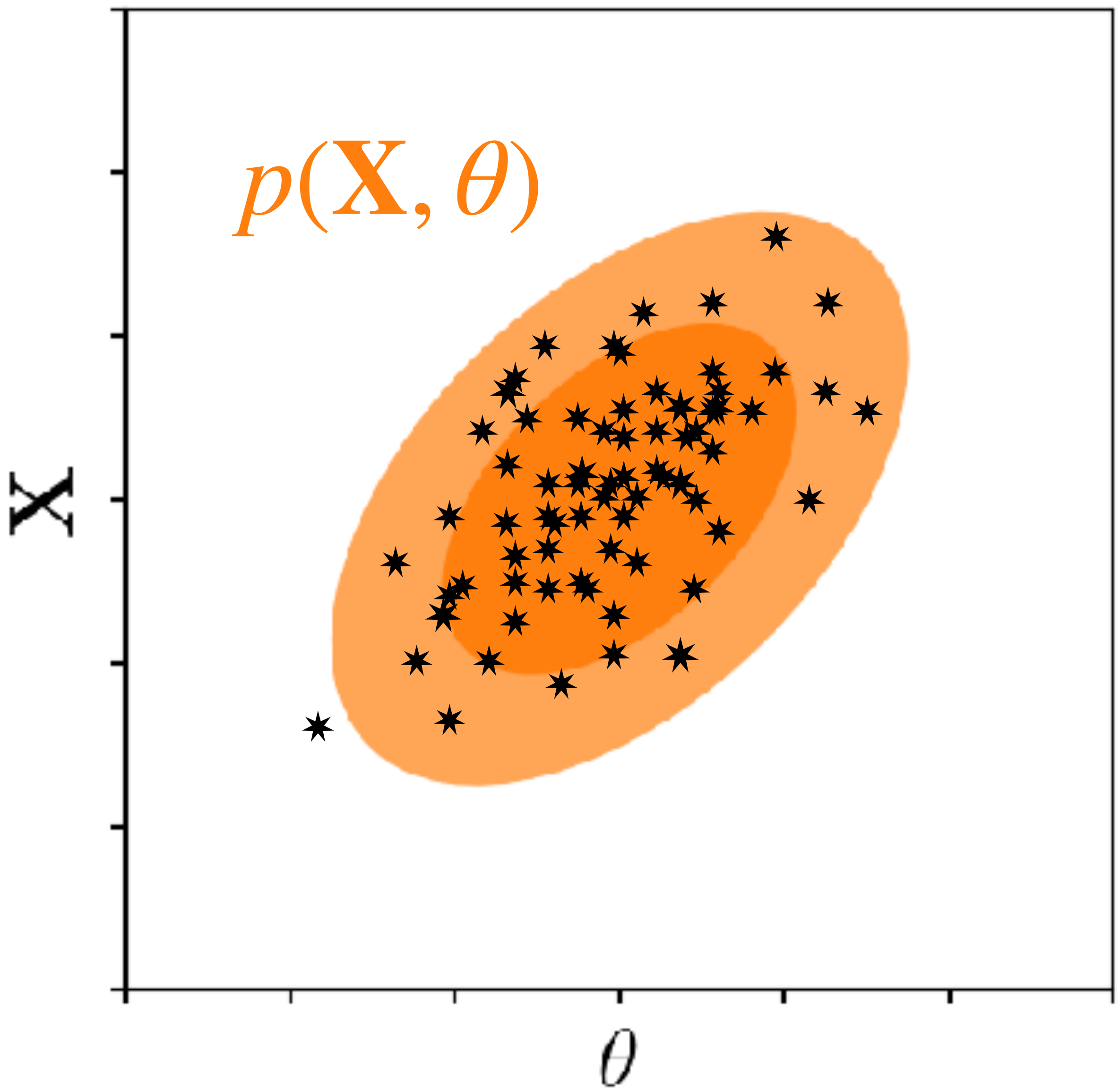
$$\mathbf{X}' \sim F(\theta') \quad \equiv \quad \mathbf{X}' \sim p(\mathbf{X}\,|\,\theta')$$

1. sample the prior

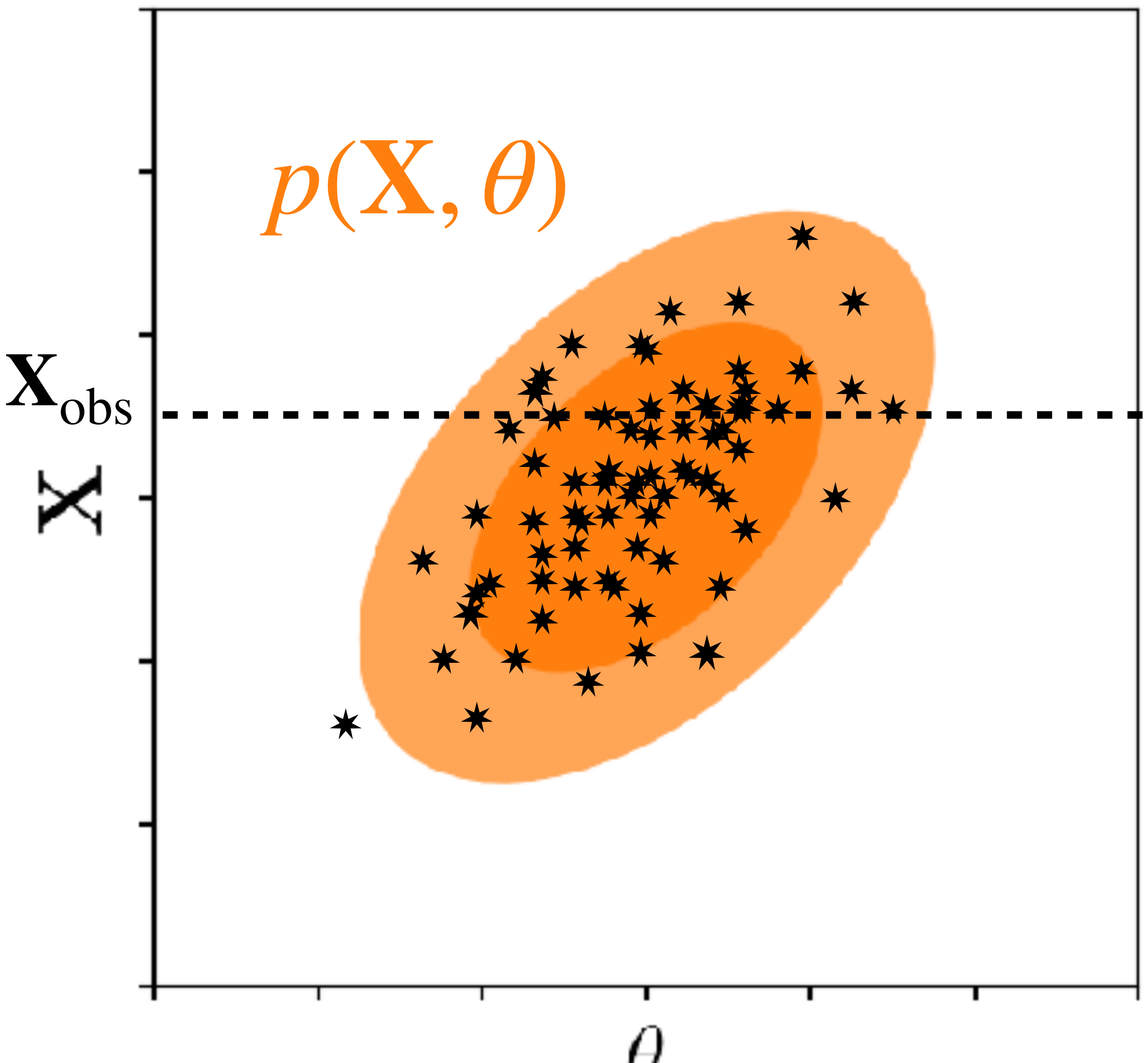
$$\theta' \sim p(\theta)$$

2. run simulator

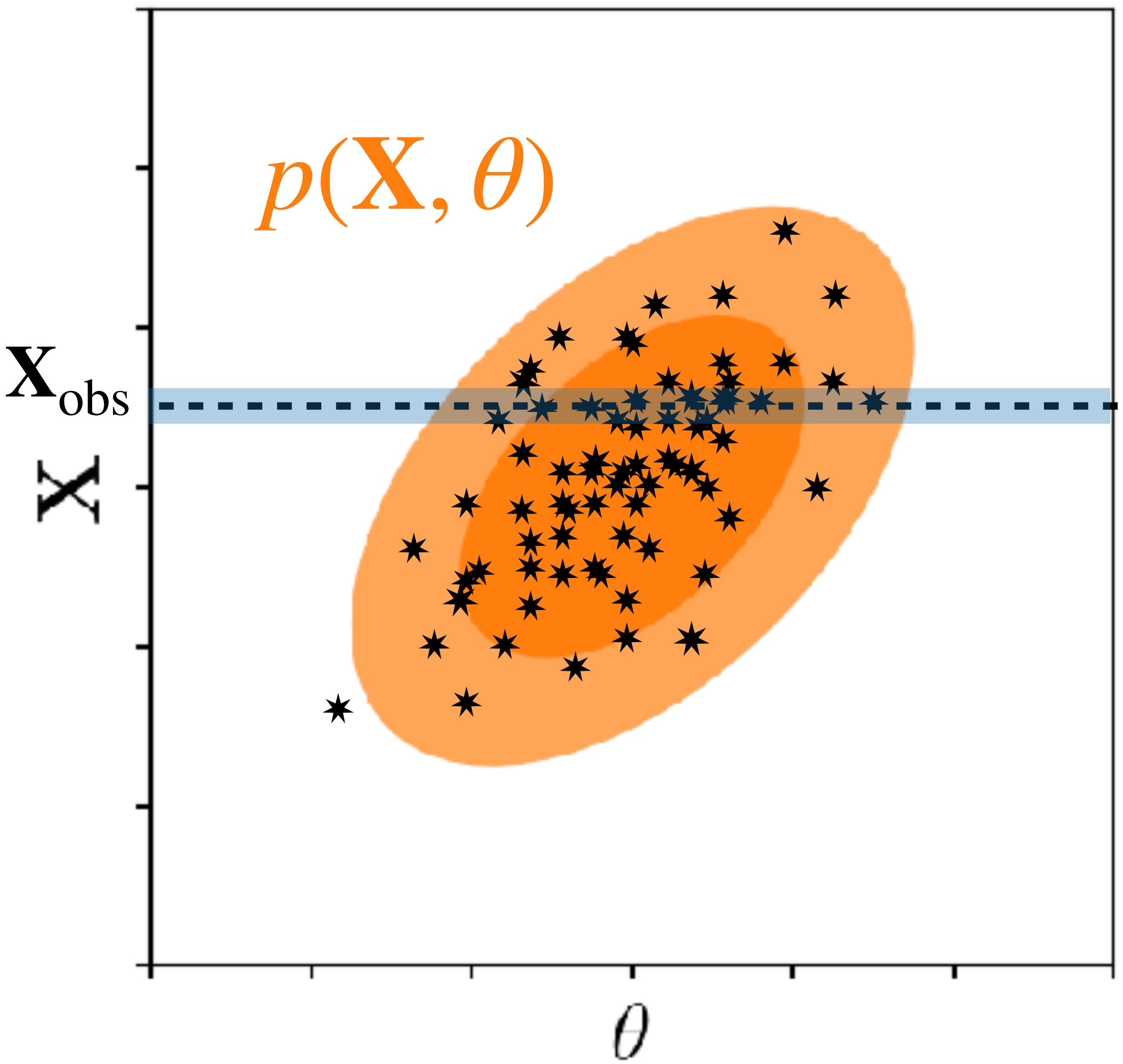
$$\mathbf{X}' \sim F(\theta')$$



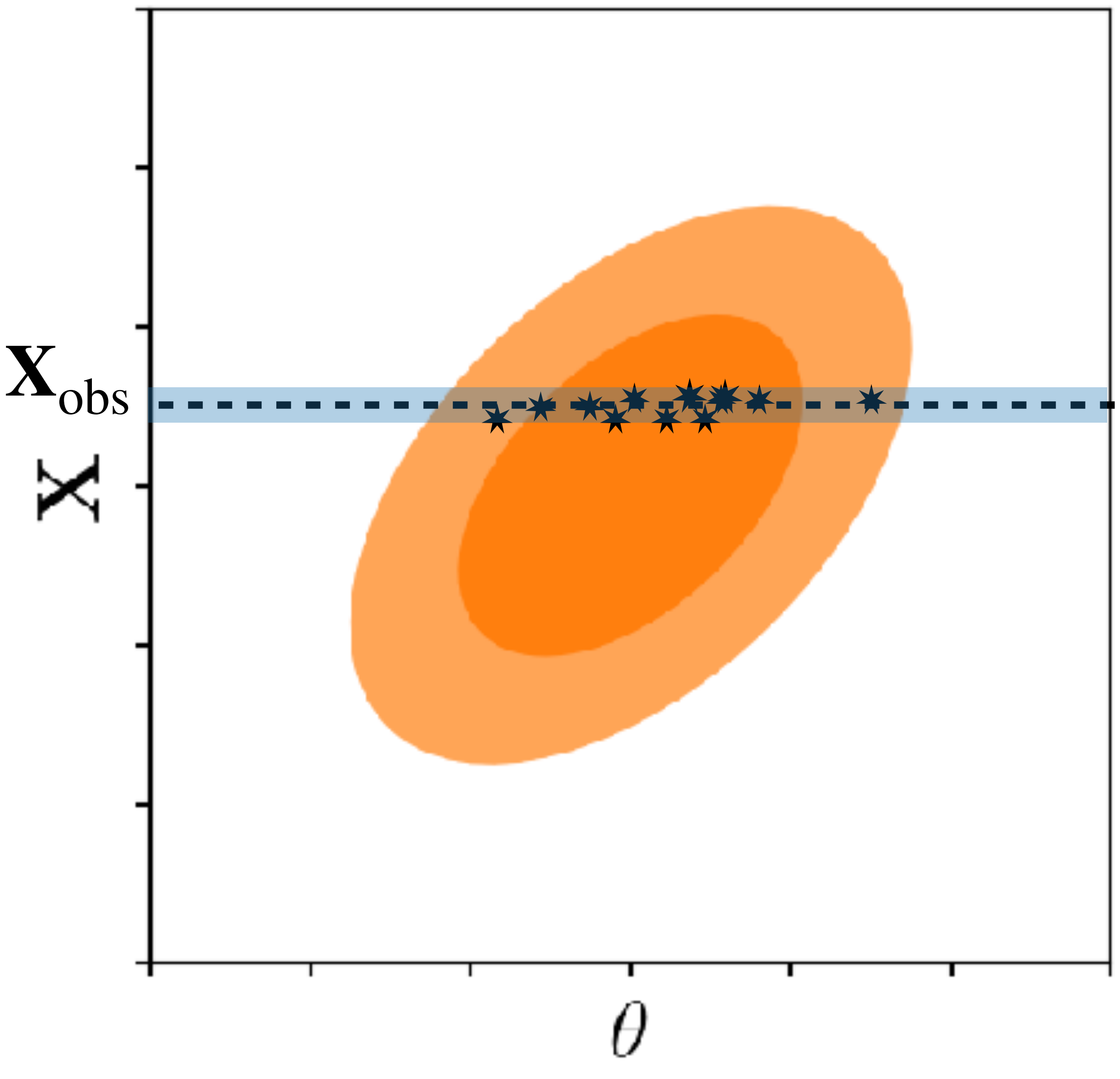
what is **simulation-based inference**?



what is **simulation-based inference**?

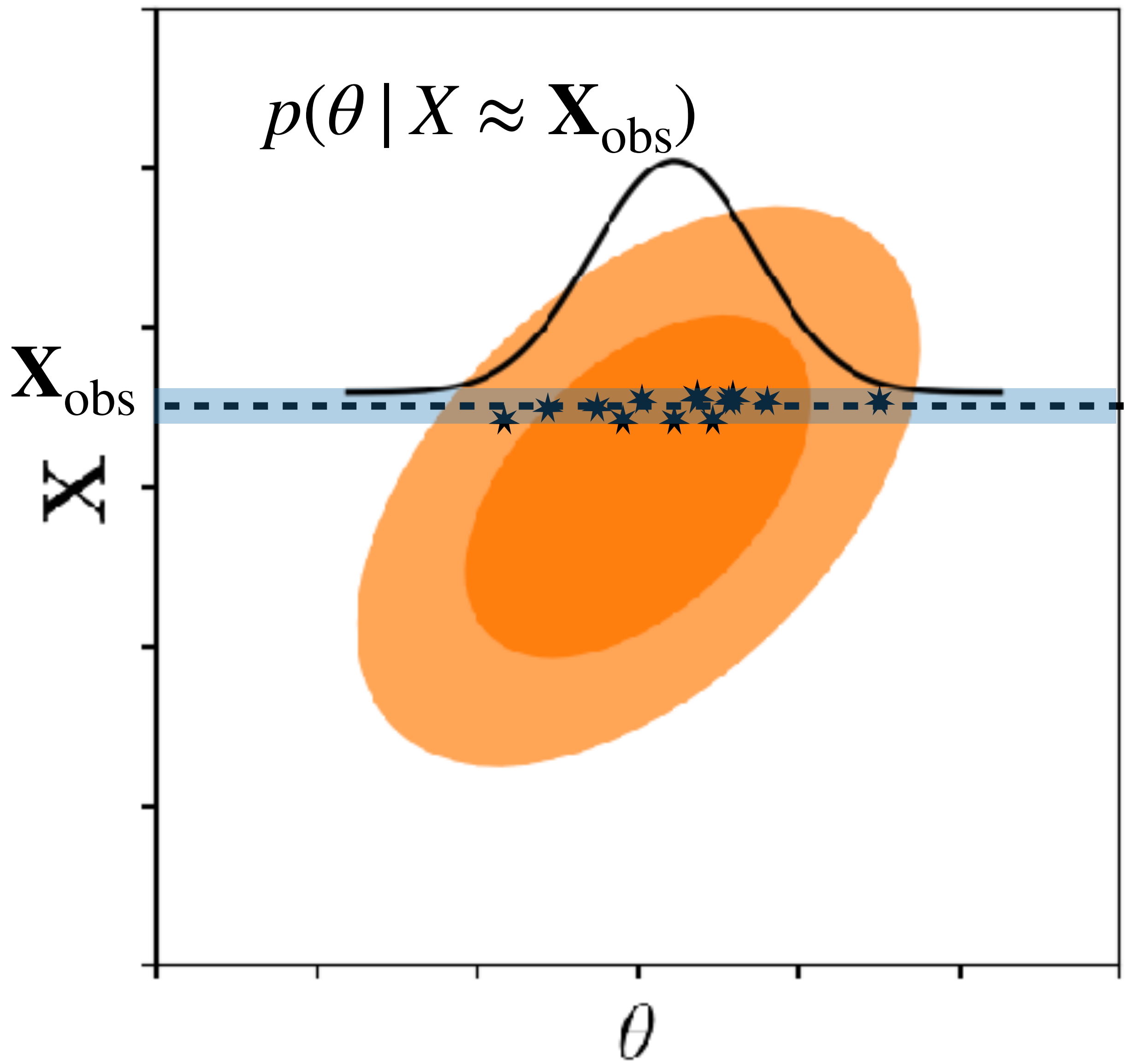


what is **simulation-based inference**?



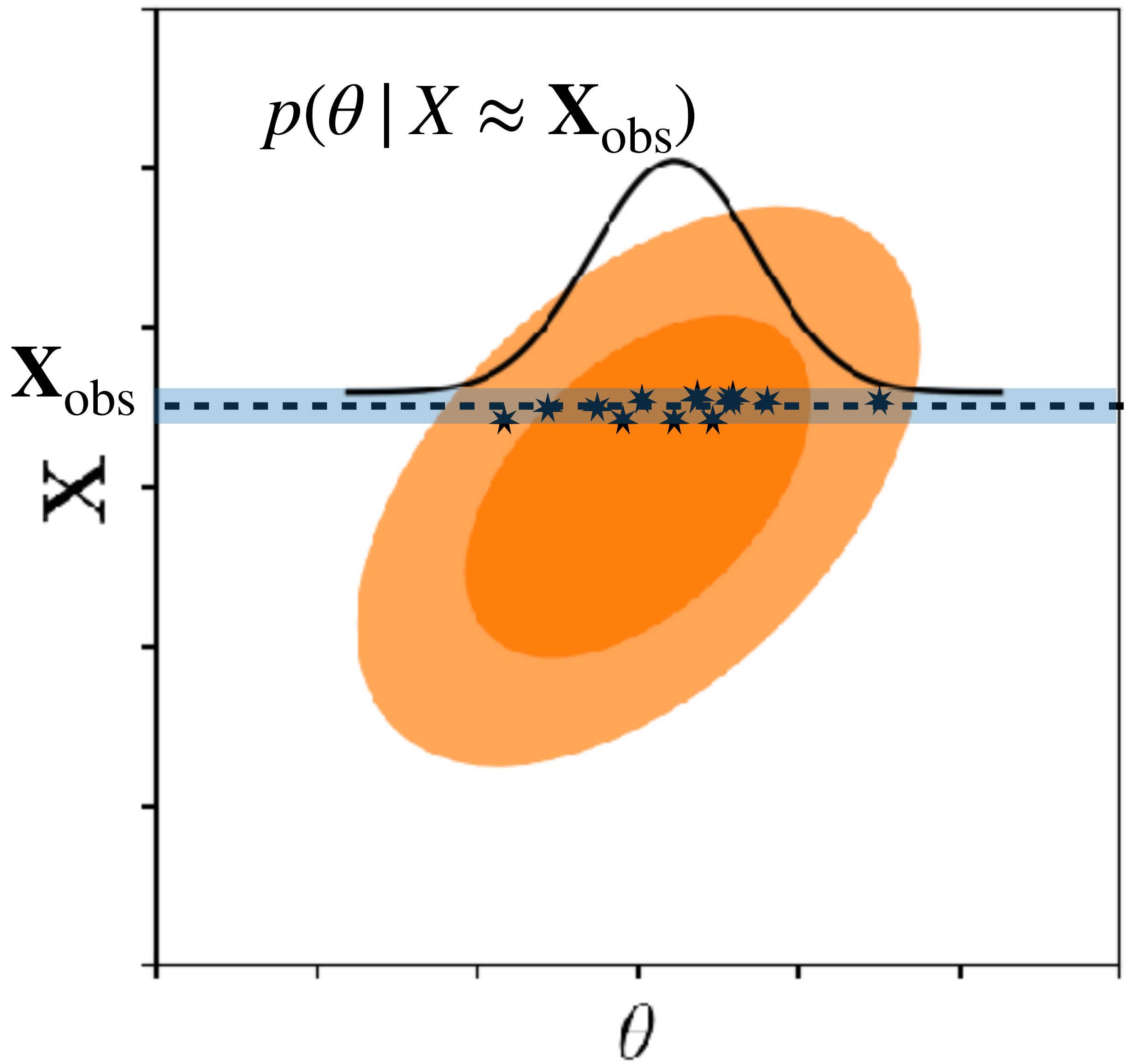
what is **simulation-based inference**?

$$p(\theta \mid \mathbf{X}_{\text{obs}}) \approx p(\theta \mid X \approx \mathbf{X}_{\text{obs}})$$

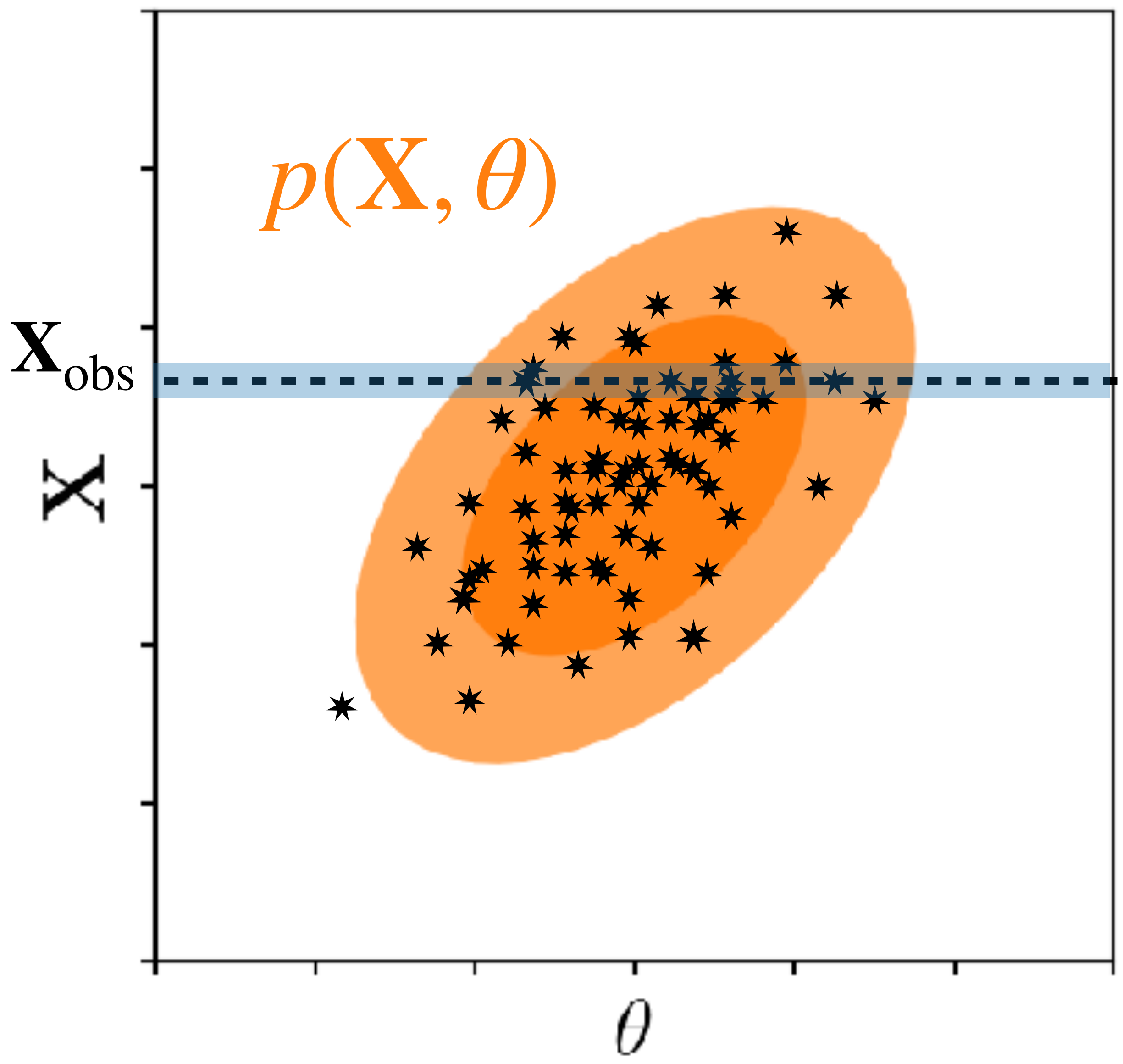


what is **simulation-based inference**?

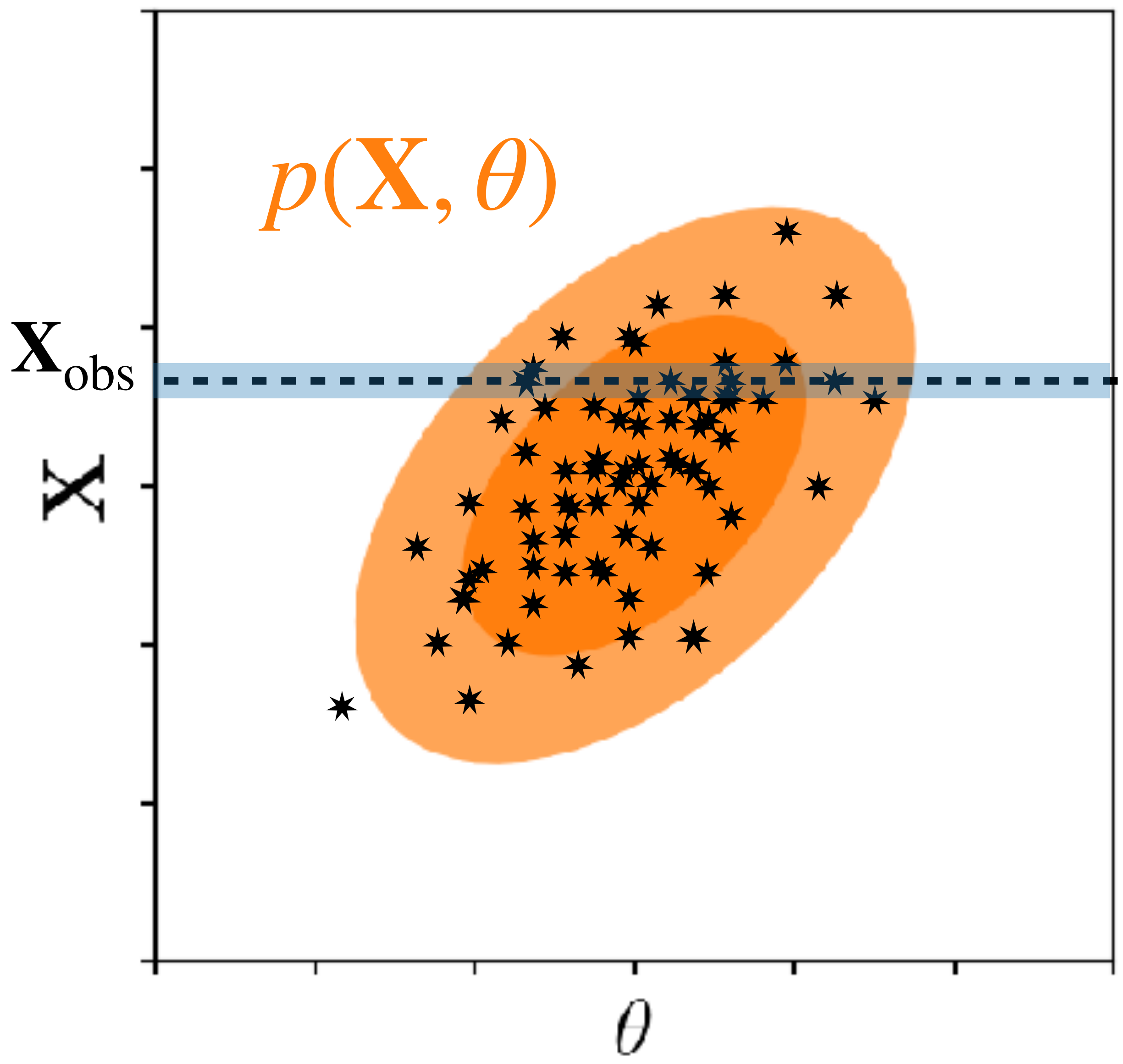
$$p(\theta \,|\, \mathbf{X}_{\text{obs}}) \approx p(\theta \,|\, X \approx \mathbf{X}_{\text{obs}})$$



simulation-based inference *in practice*

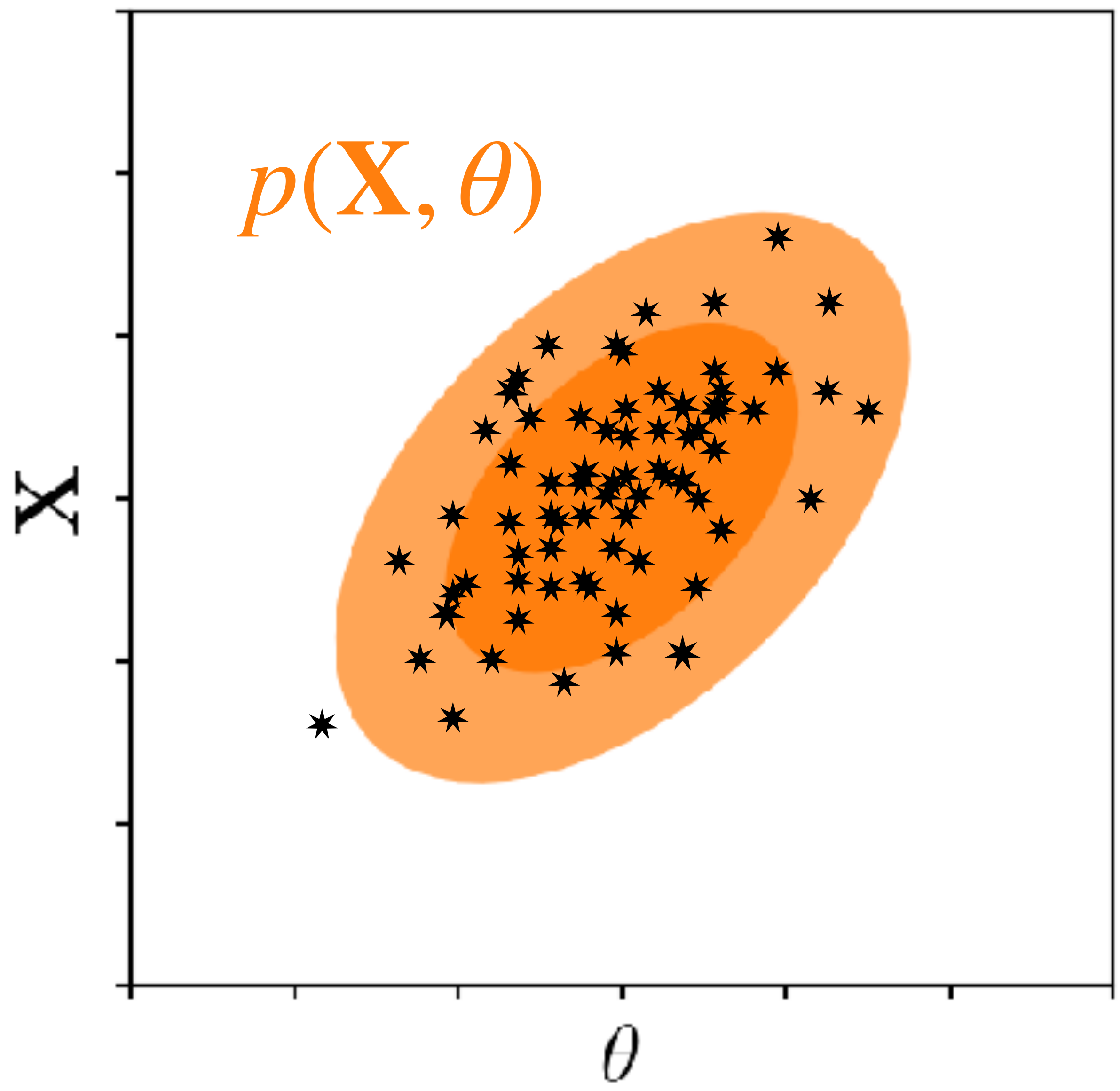


simulation-based inference *in practice*



approximate bayesian computation is often *infeasible*

simulation-based inference *in practice* — density estimation



can we estimate $p(\theta \,|\, \mathbf{X})$ from $\mathbf{X}' \sim F(\theta)$?

$\sim p(\mathbf{X} \,|\, \theta)$

estimate $p(\theta \mid \mathbf{X}) \approx q_\phi(\theta \mid \mathbf{X})$ from $\{(\theta', \mathbf{X}')\} \sim p(\mathbf{X}, \theta)$?

estimate $p(\theta \mid \mathbf{X}) \approx q_\phi(\theta \mid \mathbf{X})$ from $\{(\theta', \mathbf{X}')\} \sim p(\mathbf{X}, \theta)$?

some model q with free parameters ϕ

**Gaussian Mixture Models,*
Independent Component Analysis, neural density estimators…

$$\text{estimate } p(\theta \mid \mathbf{X}) \approx q_\phi(\theta \mid \mathbf{X}) \qquad \text{from } \{(\theta', \mathbf{X}')\} \sim p(\mathbf{X}, \theta) \text{ ?}$$

we can determine ϕ by

$$\min_\phi \; D_{\mathrm{KL}}(\, p(\theta \mid \mathbf{X})\, p(\mathbf{X}) \;\|\; q_\phi(\theta \mid \mathbf{X})\, p(\mathbf{X})\,)$$

Kullback-Leibler divergence (a.k.a. relative entropy) estimate $p(\theta \,|\, \mathbf{X}) \approx q_\phi(\theta \,|\, \mathbf{X})$ from $\{(\theta', \mathbf{X}')\} \sim p(\mathbf{X}, \theta)$?

we can determine ϕ *by*

$$\min_\phi \; D_{\mathrm{KL}}(\, p(\theta \,|\, \mathbf{X})\, p(\mathbf{X}) \;\|\; q_\phi(\theta \,|\, \mathbf{X})\, p(\mathbf{X})\,)$$

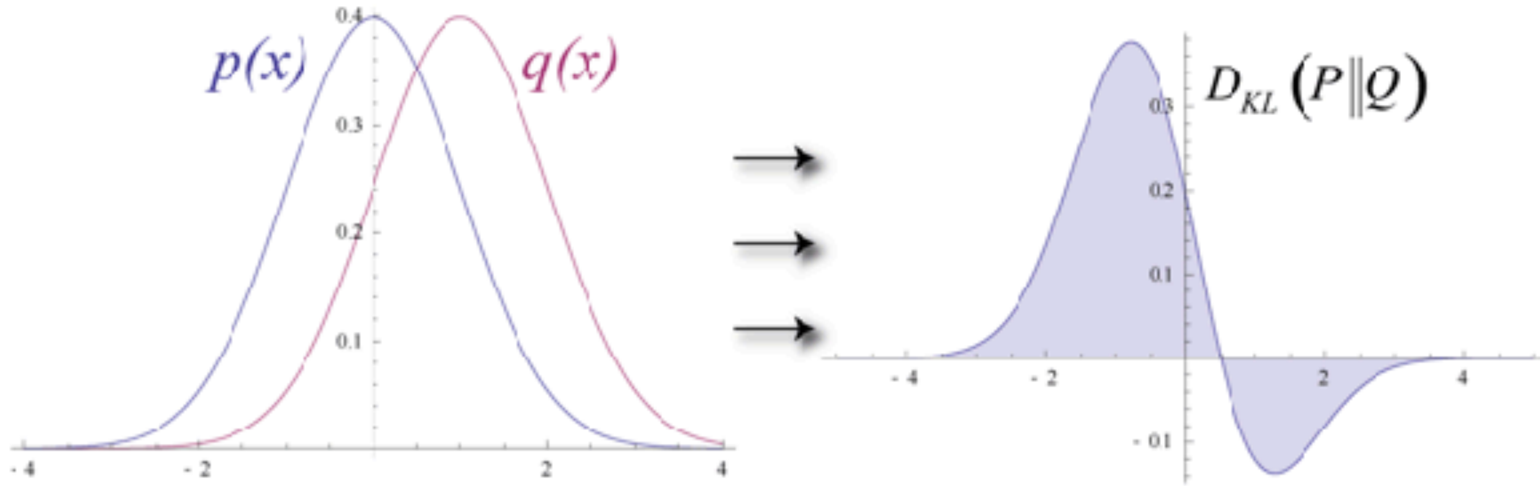


image: wikipedia

estimate $p(\theta \,|\, \mathbf{X}) \approx q_\phi(\theta \,|\, \mathbf{X})$ from $\{(\theta', \mathbf{X}')\} \sim p(\mathbf{X}, \theta)$?

we can determine ϕ by

$$\min_\phi \; D_{\mathrm{KL}}(\, p(\theta \,|\, \mathbf{X})\, p(\mathbf{X}) \;\|\; q_\phi(\theta \,|\, \mathbf{X})\, p(\mathbf{X})\,)$$

$$= \min_\phi \; \int p(\theta \,|\, \mathbf{X})\, p(\mathbf{X}) \;\log \frac{p(\theta \,|\, \mathbf{X})\, p(\mathbf{X})}{q_\phi(\theta \,|\, \mathbf{X})\, p(\mathbf{X})}$$

Kullback-Leibler divergence (a.k.a. relative entropy)

estimate $p(\theta \mid \mathbf{X}) \approx q_\phi(\theta \mid \mathbf{X})$ from $\{(\theta', \mathbf{X}')\} \sim p(\mathbf{X}, \theta)$?

we can determine ϕ *by*

$$\min_\phi \; D_{\mathrm{KL}}(\, p(\theta \mid \mathbf{X})\, p(\mathbf{X}) \;\|\; q_\phi(\theta \mid \mathbf{X})\, p(\mathbf{X})\,)$$

$$= \min_\phi \int \underbrace{p(\theta \mid \mathbf{X})\, p(\mathbf{X})}^{p(\mathbf{X}, \theta)} \log \frac{p(\theta \mid \mathbf{X})\, p(\mathbf{X})}{q_\phi(\theta \mid \mathbf{X})\, p(\mathbf{X})}$$

$$\text{estimate } p(\theta \,|\, \mathbf{X}) \approx q_\phi(\theta \,|\, \mathbf{X}) \qquad \text{from } \{(\theta', \mathbf{X}')\} \sim p(\mathbf{X}, \theta) \text{ ?}$$

we can determine ϕ *by*

$$\min_\phi \; D_{\mathrm{KL}}(\, p(\theta \,|\, \mathbf{X})\, p(\mathbf{X}) \;\|\; q_\phi(\theta \,|\, \mathbf{X})\, p(\mathbf{X})\,)$$

$$= \min_\phi \; \int p(\theta \,|\, \mathbf{X})\, p(\mathbf{X}) \; \log \frac{p(\theta \,|\, \mathbf{X})\, p(\mathbf{X})}{q_\phi(\theta \,|\, \mathbf{X})\, p(\mathbf{X})}$$

$$\approx \min_\phi \sum_{(\mathbf{X}', \theta') \sim p(\mathbf{X}, \theta)} \log \frac{p(\theta' \,|\, \mathbf{X}')}{q_\phi(\theta' \,|\, \mathbf{X}')}$$

estimate $p(\theta \mid \mathbf{X}) \approx q_\phi(\theta \mid \mathbf{X})$ from $\{(\theta', \mathbf{X}')\} \sim p(\mathbf{X}, \theta)$?

we can determine ϕ by

$$\min_\phi \; D_{\mathrm{KL}}(\, p(\theta \mid \mathbf{X})\, p(\mathbf{X}) \;\|\; q_\phi(\theta \mid \mathbf{X})\, p(\mathbf{X})\,)$$

$$= \min_\phi \int p(\theta \mid \mathbf{X})\, p(\mathbf{X}) \; \log \frac{p(\theta \mid \mathbf{X})\, p(\mathbf{X})}{q_\phi(\theta \mid \mathbf{X})\, p(\mathbf{X})}$$

$$\approx \min_\phi \sum_{(\mathbf{X}', \theta') \sim p(\mathbf{X}, \theta)} \log \frac{p(\theta' \mid \mathbf{X}')}{q_\phi(\theta' \mid \mathbf{X}')}$$

estimate $p(\theta \mid \mathbf{X}) \approx q_\phi(\theta \mid \mathbf{X})$ from $\{(\theta', \mathbf{X}')\} \sim p(\mathbf{X}, \theta)$?

we can determine ϕ *by*

$$\min_\phi \; D_{\mathrm{KL}}(\, p(\theta \mid \mathbf{X})\, p(\mathbf{X}) \;\|\; q_\phi(\theta \mid \mathbf{X})\, p(\mathbf{X})\,)$$

$$= \min_\phi \int p(\theta \mid \mathbf{X})\, p(\mathbf{X}) \; \log \frac{p(\theta \mid \mathbf{X})\, p(\mathbf{X})}{q_\phi(\theta \mid \mathbf{X})\, p(\mathbf{X})}$$

$$\approx \min_\phi \sum_{(\mathbf{X}',\theta') \sim p(\mathbf{X},\theta)} \log \frac{p(\theta' \mid \mathbf{X}')}{q_\phi(\theta' \mid \mathbf{X}')}$$

$$= \min_\phi \sum_{(\mathbf{X}',\theta') \sim p(\mathbf{X},\theta)} - \log \; q_\phi(\theta' \mid \mathbf{X}')$$

$$\text{estimate } p(\theta \mid \mathbf{X}) \approx q_\phi(\theta \mid \mathbf{X}) \qquad \text{from } \{(\theta', \mathbf{X}')\} \sim p(\mathbf{X}, \theta) \text{ ?}$$

we can determine ϕ *by*

$$\min_\phi \; D_{\mathrm{KL}}(\, p(\theta \mid \mathbf{X})\, p(\mathbf{X}) \;\|\; q_\phi(\theta \mid \mathbf{X})\, p(\mathbf{X})\,)$$

$$= \min_\phi \; \int p(\theta \mid \mathbf{X})\, p(\mathbf{X}) \; \log \frac{p(\theta \mid \mathbf{X})\, p(\mathbf{X})}{q_\phi(\theta \mid \mathbf{X})\, p(\mathbf{X})}$$

$$\approx \min_\phi \sum_{(\mathbf{X}', \theta') \sim p(\mathbf{X}, \theta)} \log \frac{p(\theta' \mid \mathbf{X}')}{q_\phi(\theta' \mid \mathbf{X}')}$$

$$= \max_\phi \; \sum_{(\mathbf{X}', \theta') \sim p(\mathbf{X}, \theta)} \log \; q_\phi(\theta' \mid \mathbf{X}')$$

estimate $p(\theta \mid \mathbf{X}) \approx q_\phi(\theta \mid \mathbf{X})$ from $\{(\theta', \mathbf{X}')\} \sim p(\mathbf{X}, \theta)$?

we can determine ϕ by

$q_\phi(\theta \mid \mathbf{X})$ is guaranteed to converge to $p(\theta \mid \mathbf{X})$ if

q_ϕ is flexibly expressive

$N \to \infty$ samples from $p(\mathbf{X}, \theta)$

successful optimization

$$= \max_\phi \sum_{(\mathbf{X}', \theta') \sim p(\mathbf{X}, \theta)} \log q_\phi(\theta' \mid \mathbf{X}')$$

normalizing flows: generative models that are easy to evaluate and flexibly expressive

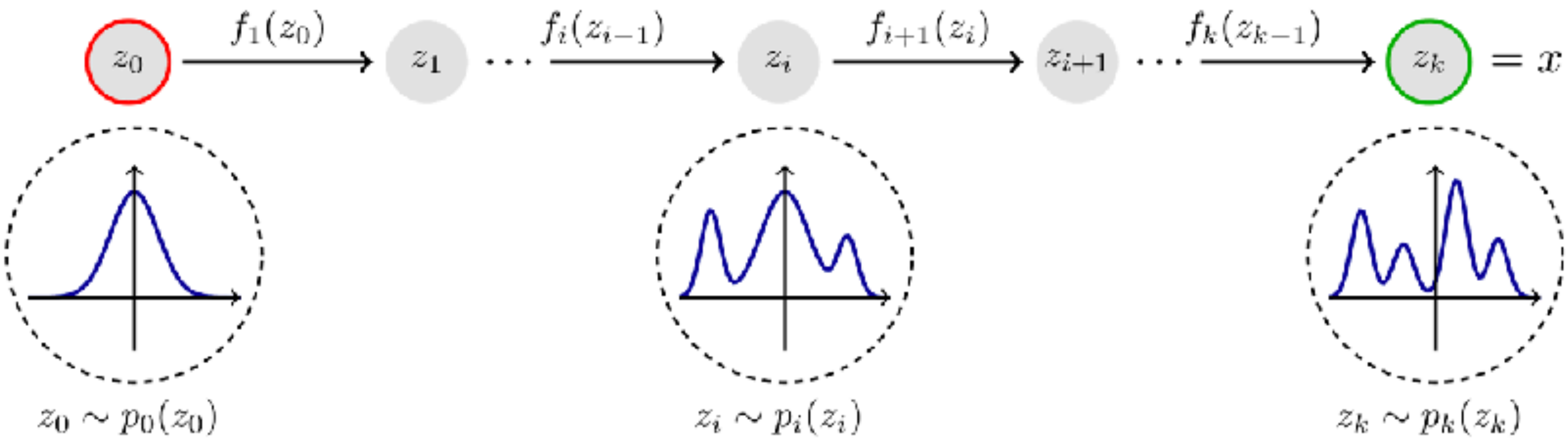
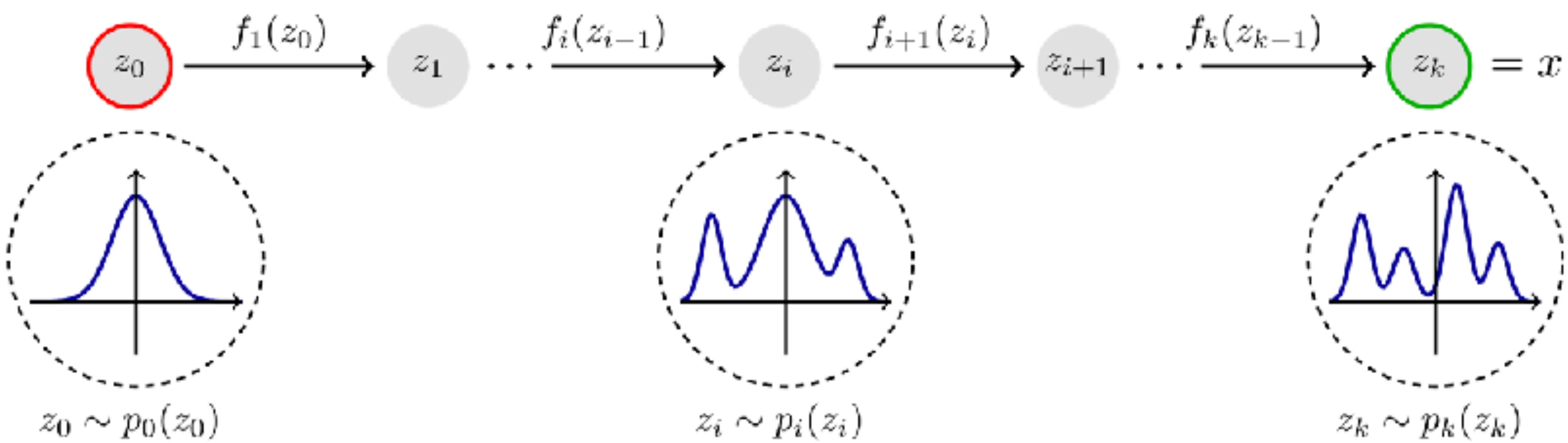


image: Janosh Riebesell

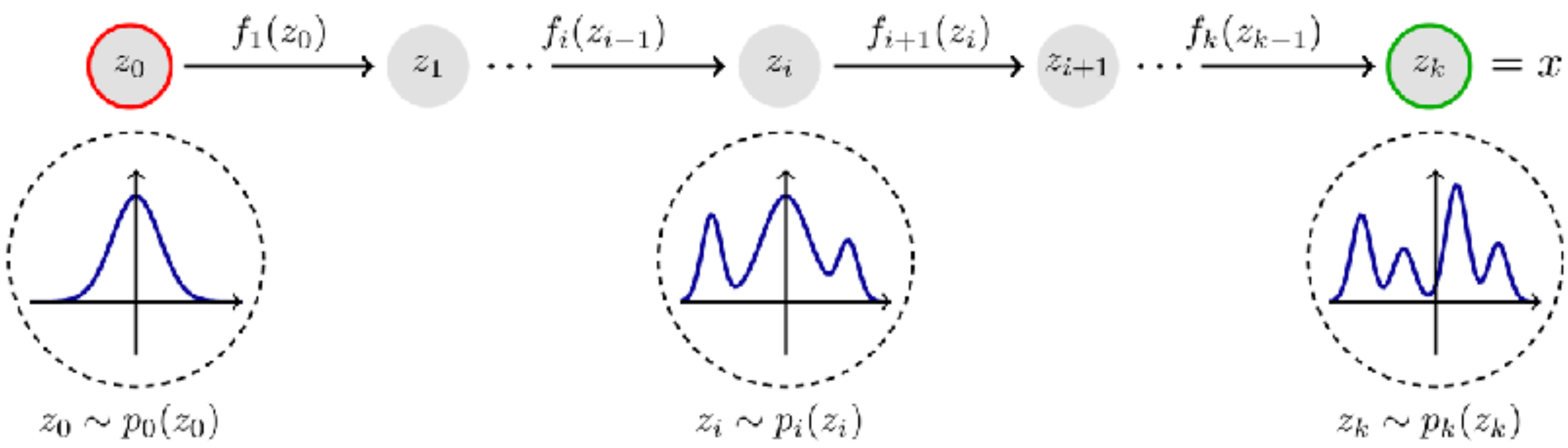
normalizing flows: generative models that are easy to evaluate and flexibly expressive



$z_i = f_i(z_{i-1})$ are invertible and differentiable transformations

$$p(z_i) = p(z_{i-1}) \left| \det \left(\frac{\partial f_i^{-1}}{\partial z_i} \right) \right|$$

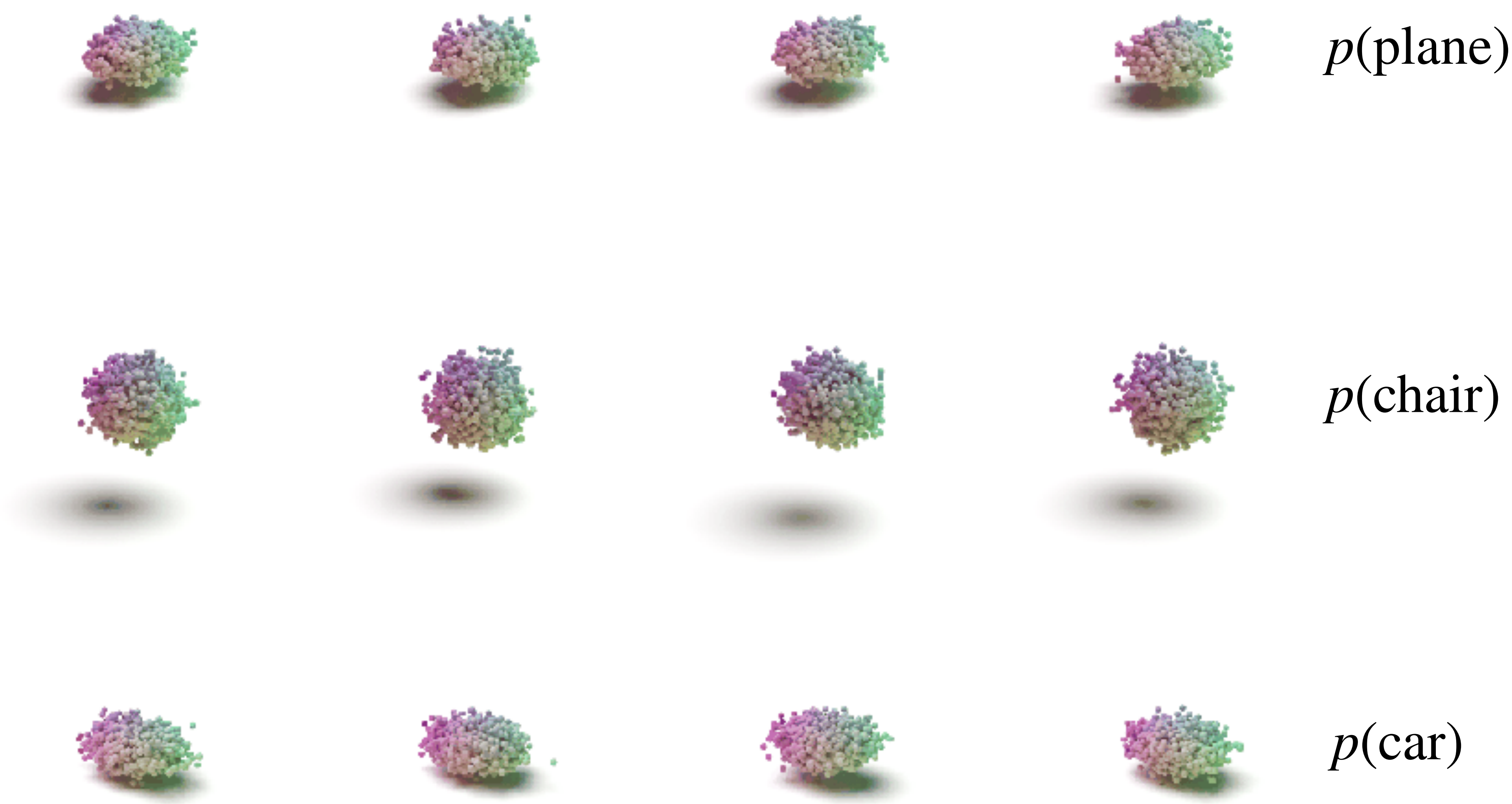
normalizing flows: generative models that are easy to evaluate and flexibly expressive



$z_i = f_i(z_{i-1})$ are invertible and differentiable transformations

$f = f_1 \circ f_2 \ldots \circ f_{k-1} \circ f_k$ is also invertible and differentiable

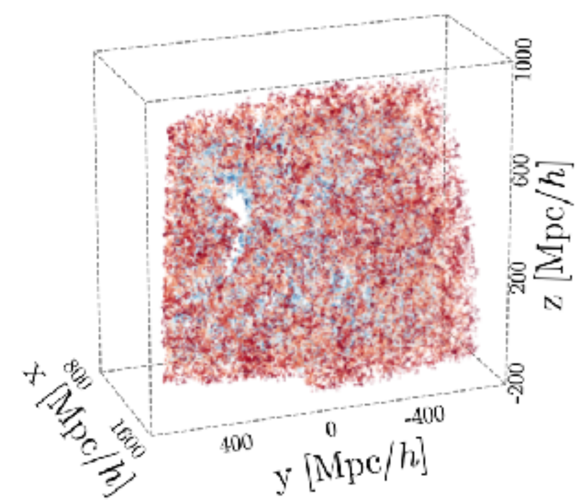
normalizing flows: generative models that are easy to evaluate and flexibly expressive



Yang et al.(2019) PointFlow

normalizing flows: generative models that are easy to evaluate and flexibly expressive

$$p\left(\begin{matrix}\Omega_m, \Omega_b, h \\ n_s, \sigma_8\end{matrix} \;\middle|\; \right.$$



$$\left. \vphantom{\begin{matrix}a\\b\end{matrix}} \right)$$

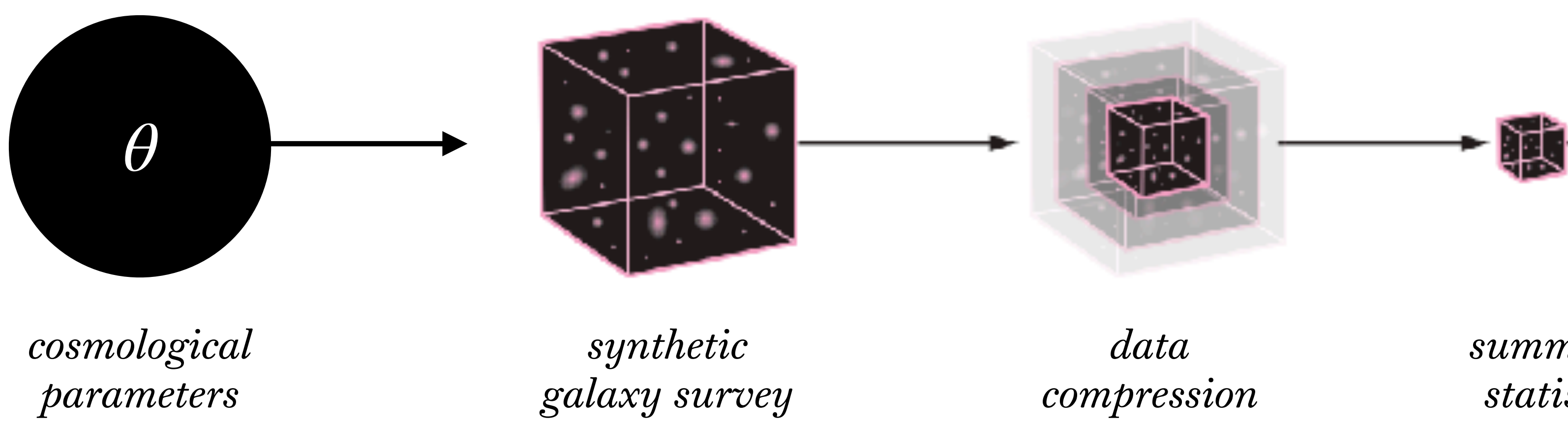
ΛCDM *parameters* — *observed galaxy distribution*

SimBIG

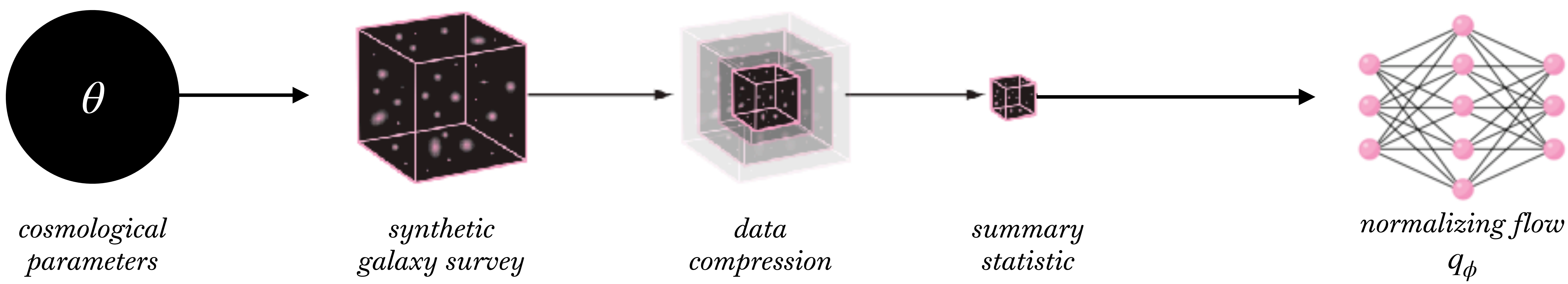
Simulation-**B**ased **I**nference of **G**alaxies

https://changhoonhahn.github.io/simbig/

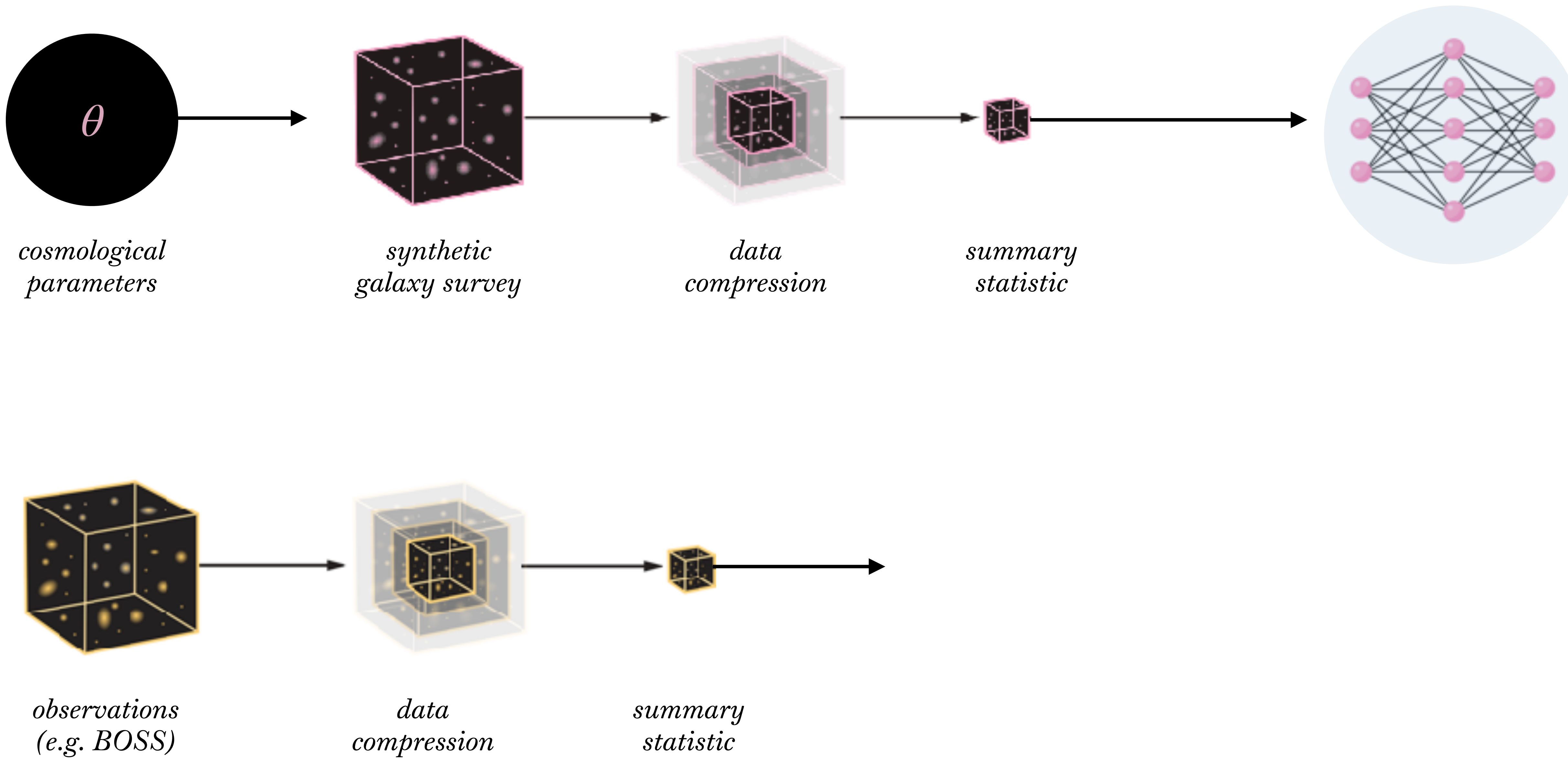
SIMBIG — 1. *generating training data of synthetic observations*



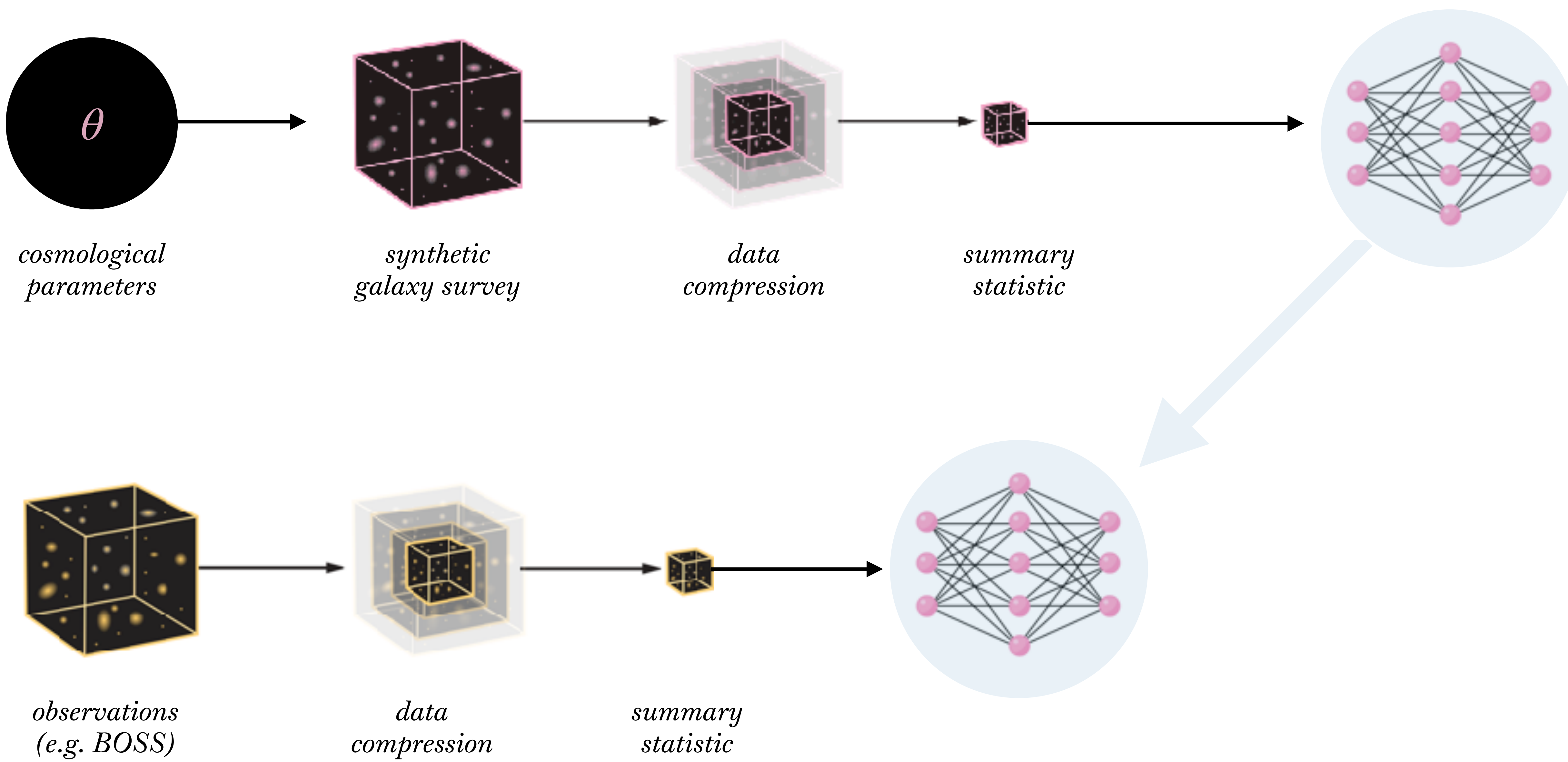
SimBIG — 2. *training the normalizing flow*



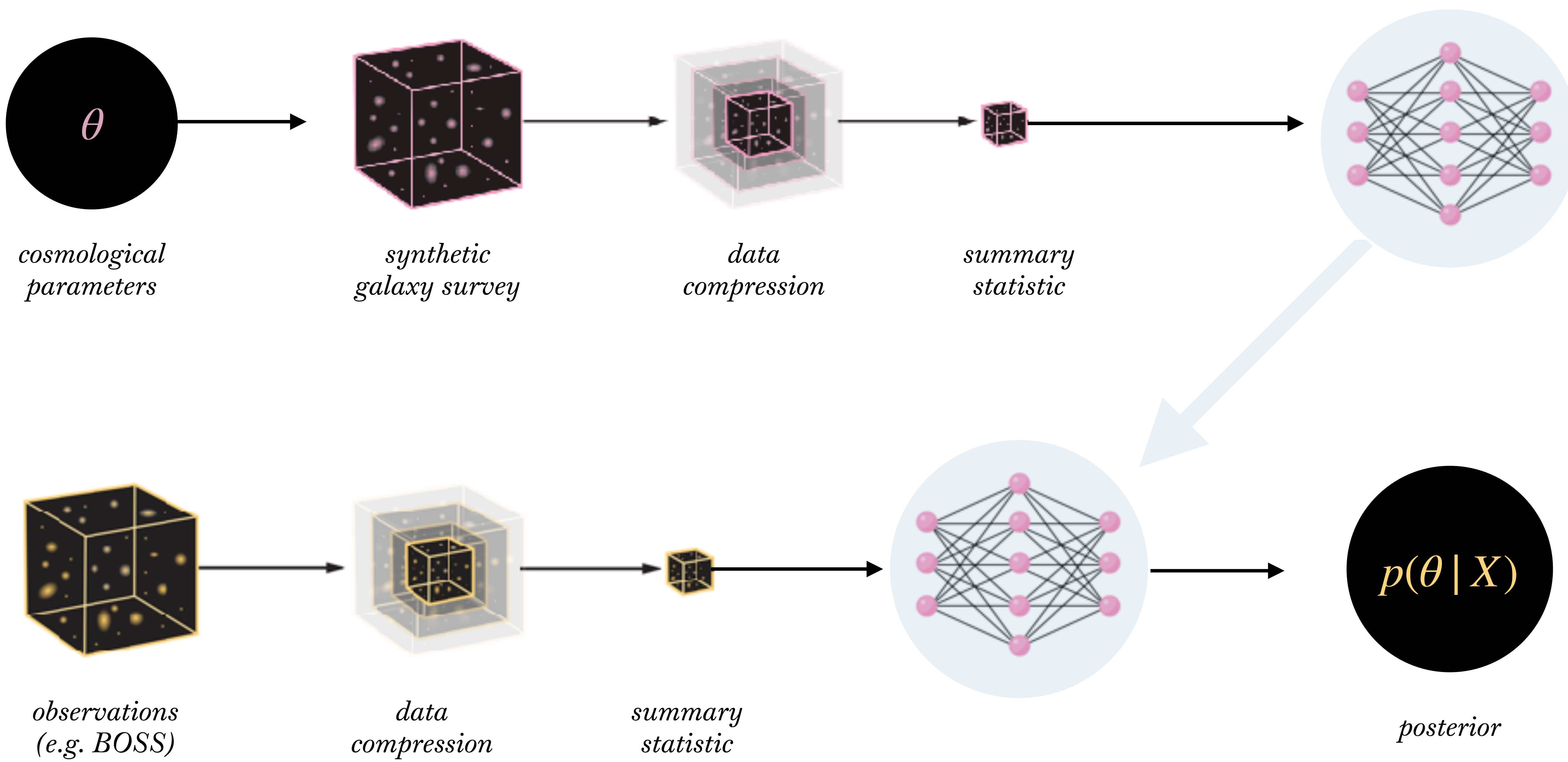
SIMBIG — 3. inference using real observations



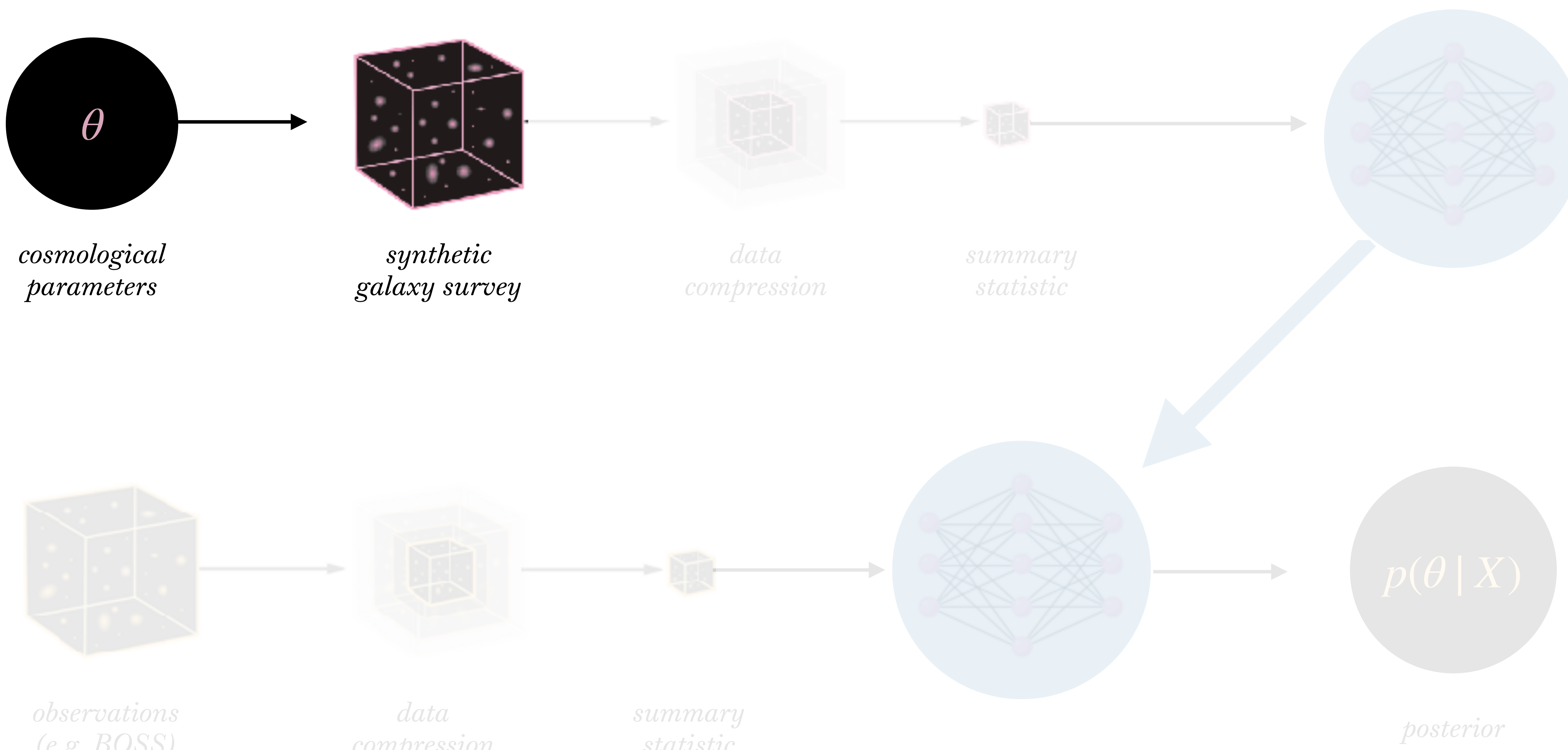
SIMBIG — 3. inference using real observations



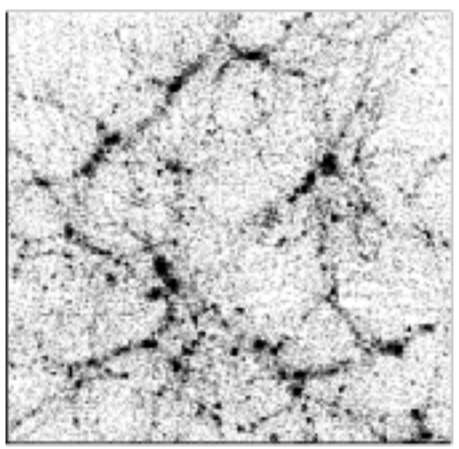
SimBIG — 3. inference using real observations



SimBIG — 3. inference using real observations

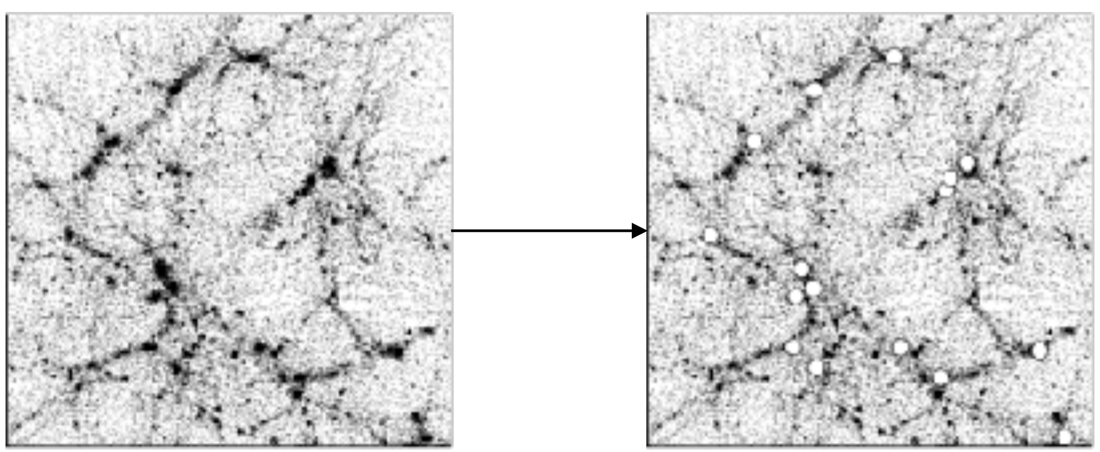


SimBIG **forward model** — SDSS-III: BOSS observations



Quijote high-res
N-body simulations

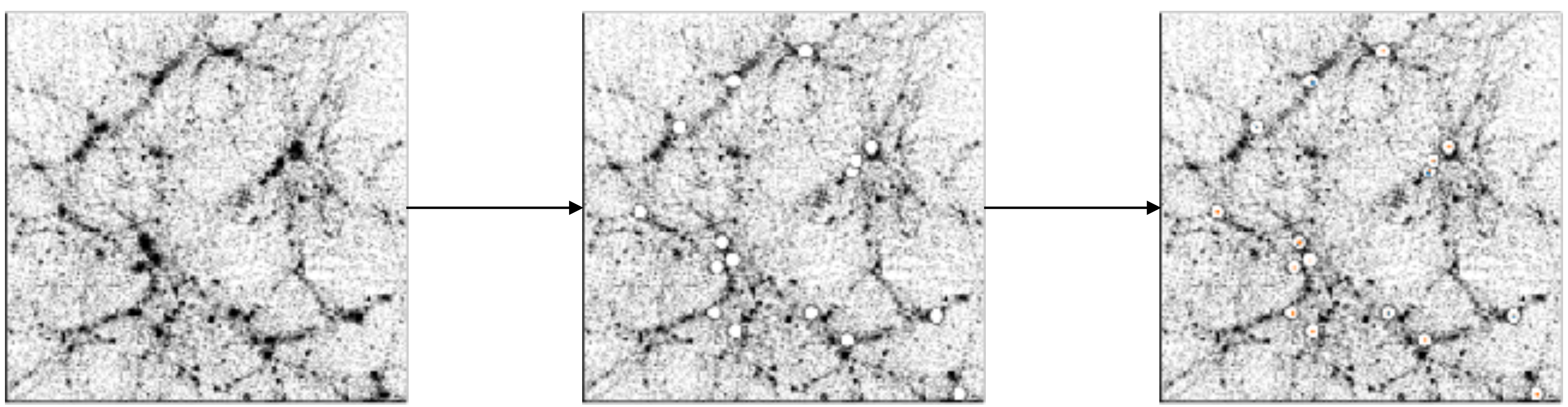
SimBIG **forward model** — SDSS-III: BOSS observations



Quijote high-res
N-body simulations

Rockstar phase-space
halo finder

SIMBIG **forward model** — SDSS-III: BOSS observations

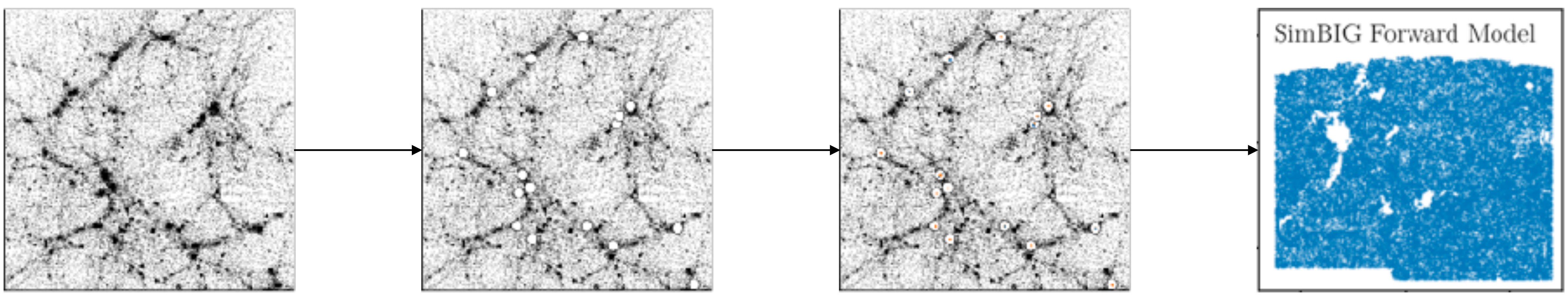


Quijote high-res
N-body simulations

Rockstar phase-space
halo finder

HOD model with *assembly,*
velocity, concentration biases

SimBIG **forward model** — SDSS-III: BOSS observations



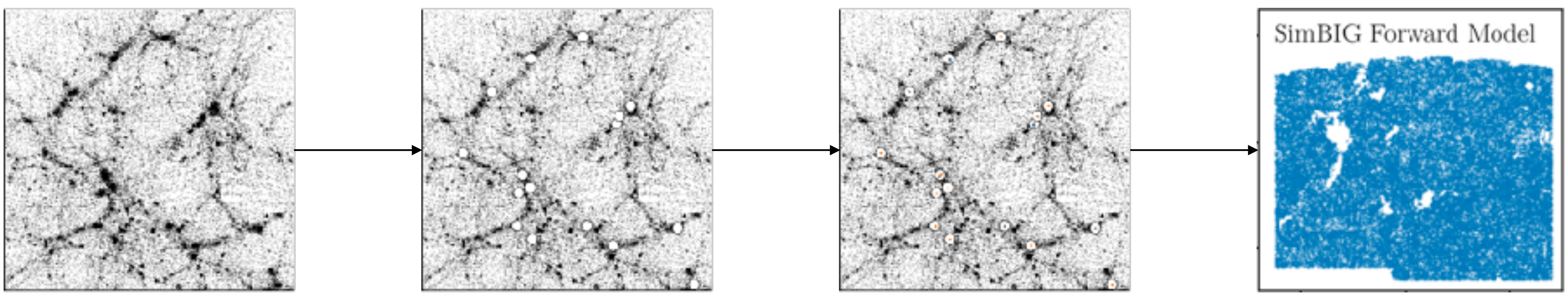
Quijote high-res N-body simulations

Rockstar phase-space halo finder

HOD model with *assembly, velocity, concentration biases*

survey realism: *redshift-space, geometry, mask, fiber collisions*

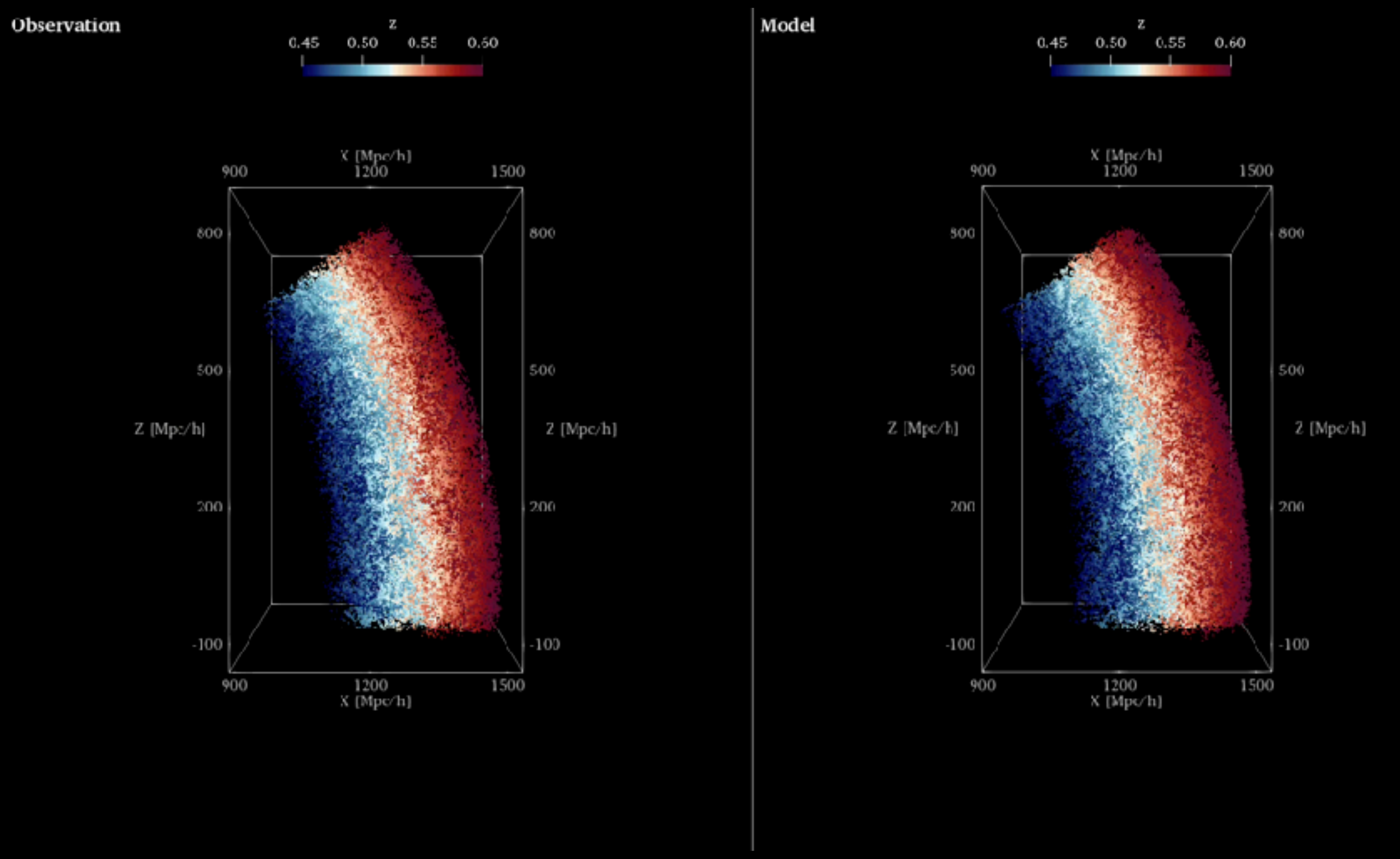
SimBIG **forward model** — SDSS-III: BOSS observations



Quijote high-res N-body simulations | Rockstar phase-space halo finder | HOD model with *assembly, velocity, concentration biases* | survey realism: *redshift-space, geometry, mask, fiber collisions*

20,000 training simulations spanning broad range of cosmologies and HOD parameters

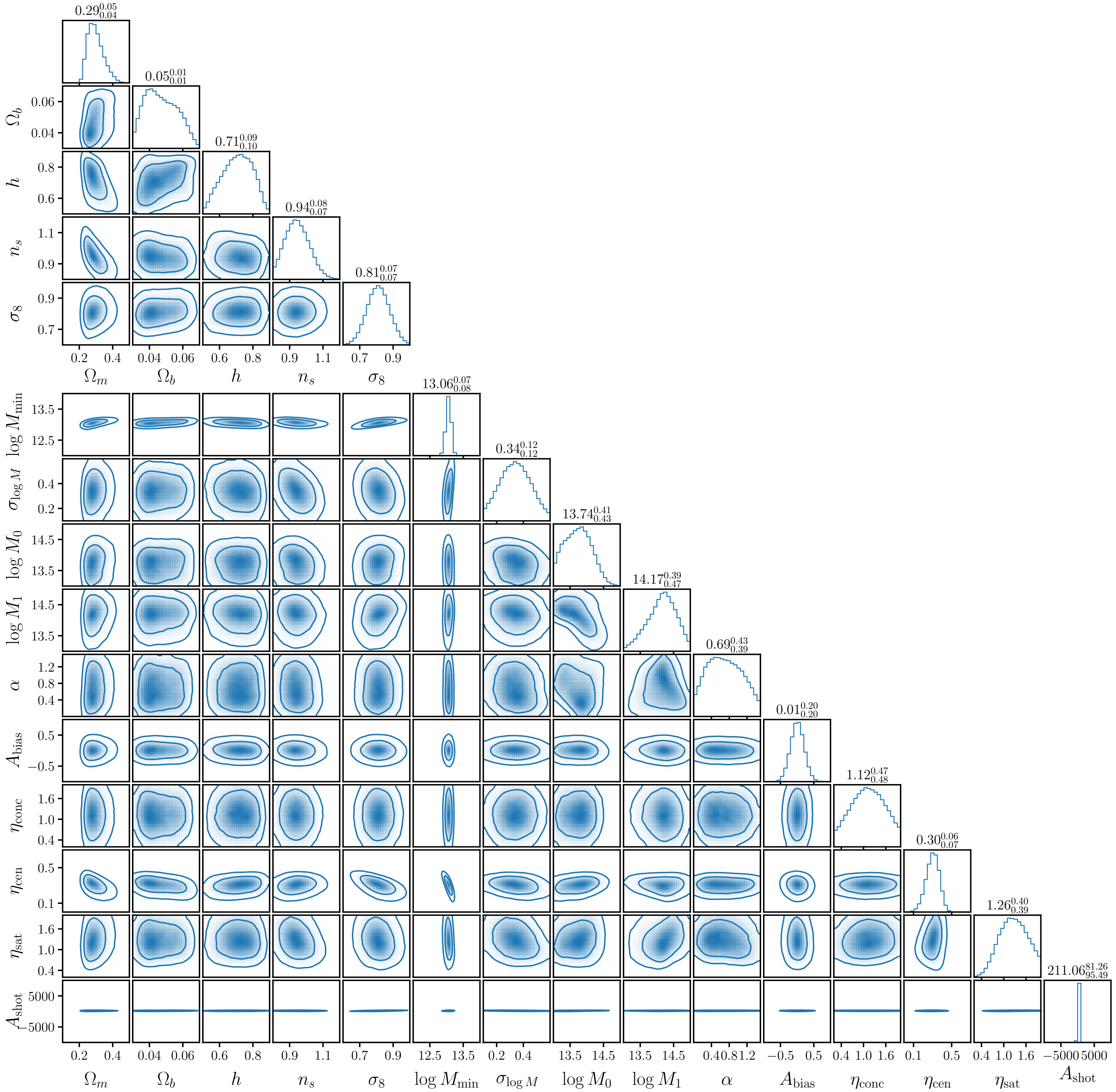
Hahn *et al.(2023c, 2023d)*



Hahn *et al.(2023c, 2023d)*
video: Bruno Régaldo-Saint Blancard

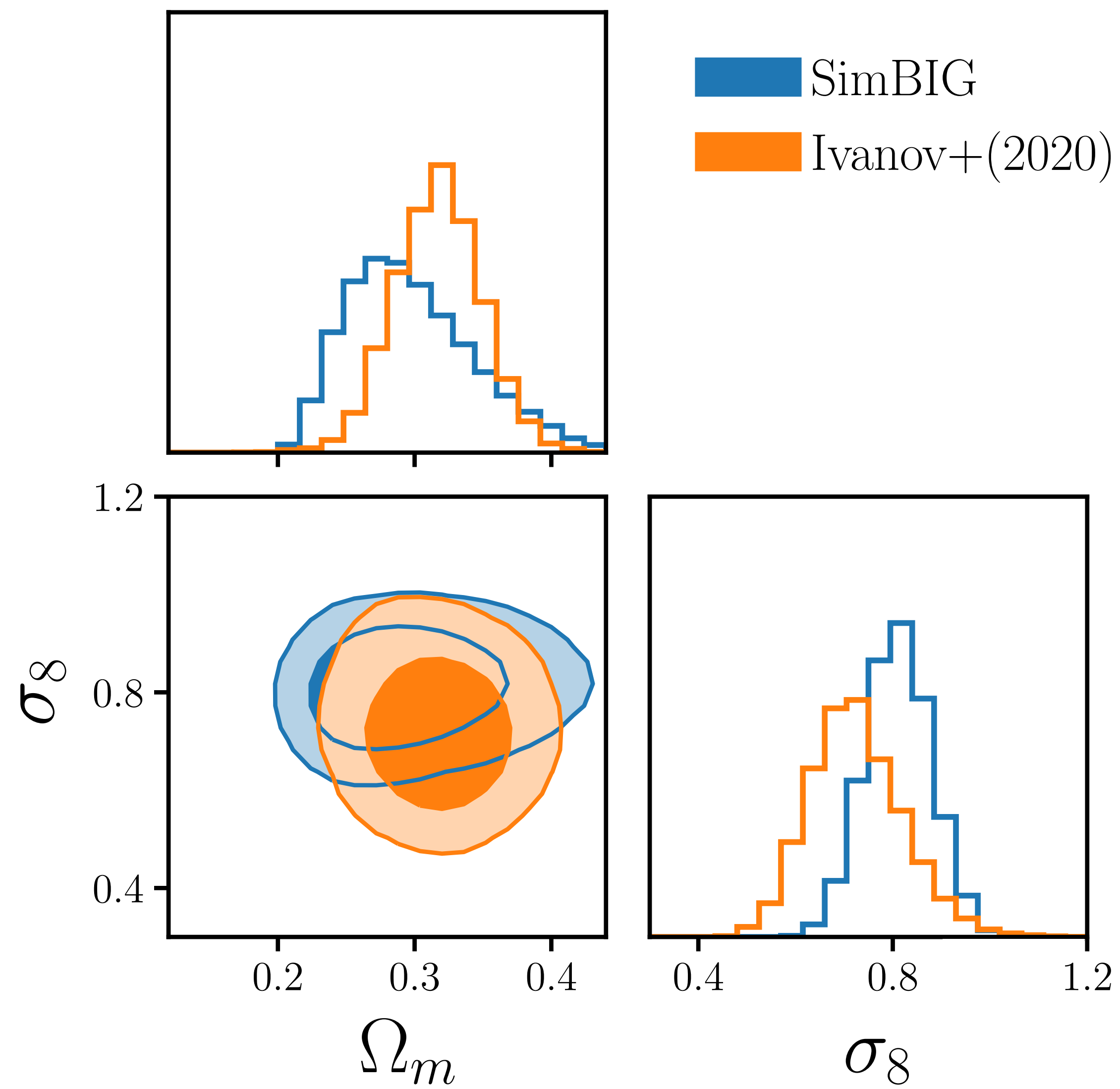
SimBIG: non-linear galaxy power spectrum $P_\ell(k < 0.5\,h/\mathrm{Mpc})$

SimBIG: non-linear galaxy power spectrum $P_\ell(k < 0.5\,h/\mathrm{Mpc})$



Hahn *et al.(2023c);* ***Hahn*** *et al.(2023d) PNAS*

SimBIG: non-linear galaxy power spectrum $P_\ell(k < 0.5\, h/\mathrm{Mpc})$

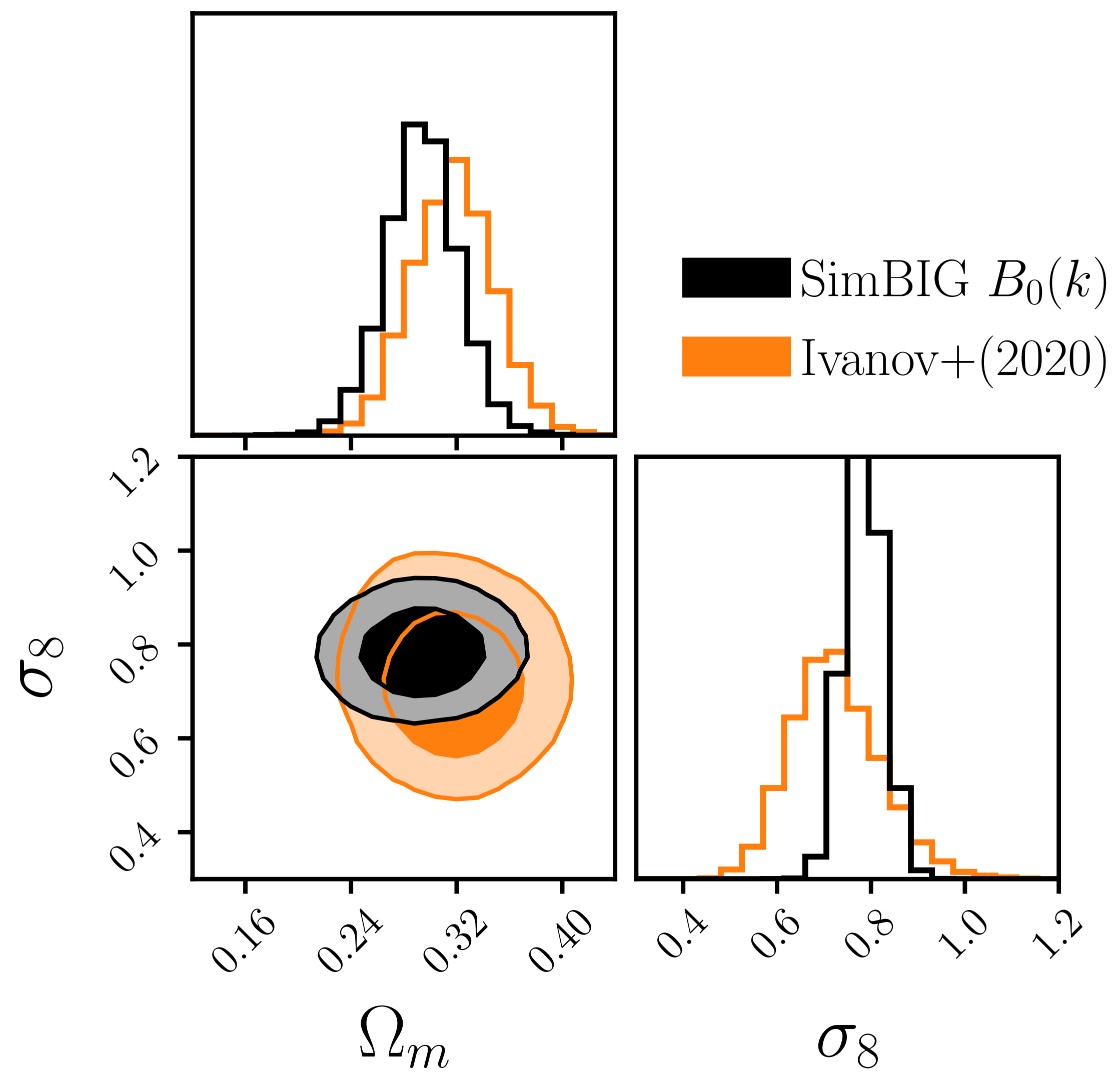


$1.4\times$ *tighter* σ_8 from non-linear scales

Hahn *et al.(2023c);* ***Hahn*** *et al.(2023d) PNAS*

SIMBIG: non-linear galaxy bispectrum $B_0(k_1, k_2, k_3 < 0.5\,h/\mathrm{Mpc})$

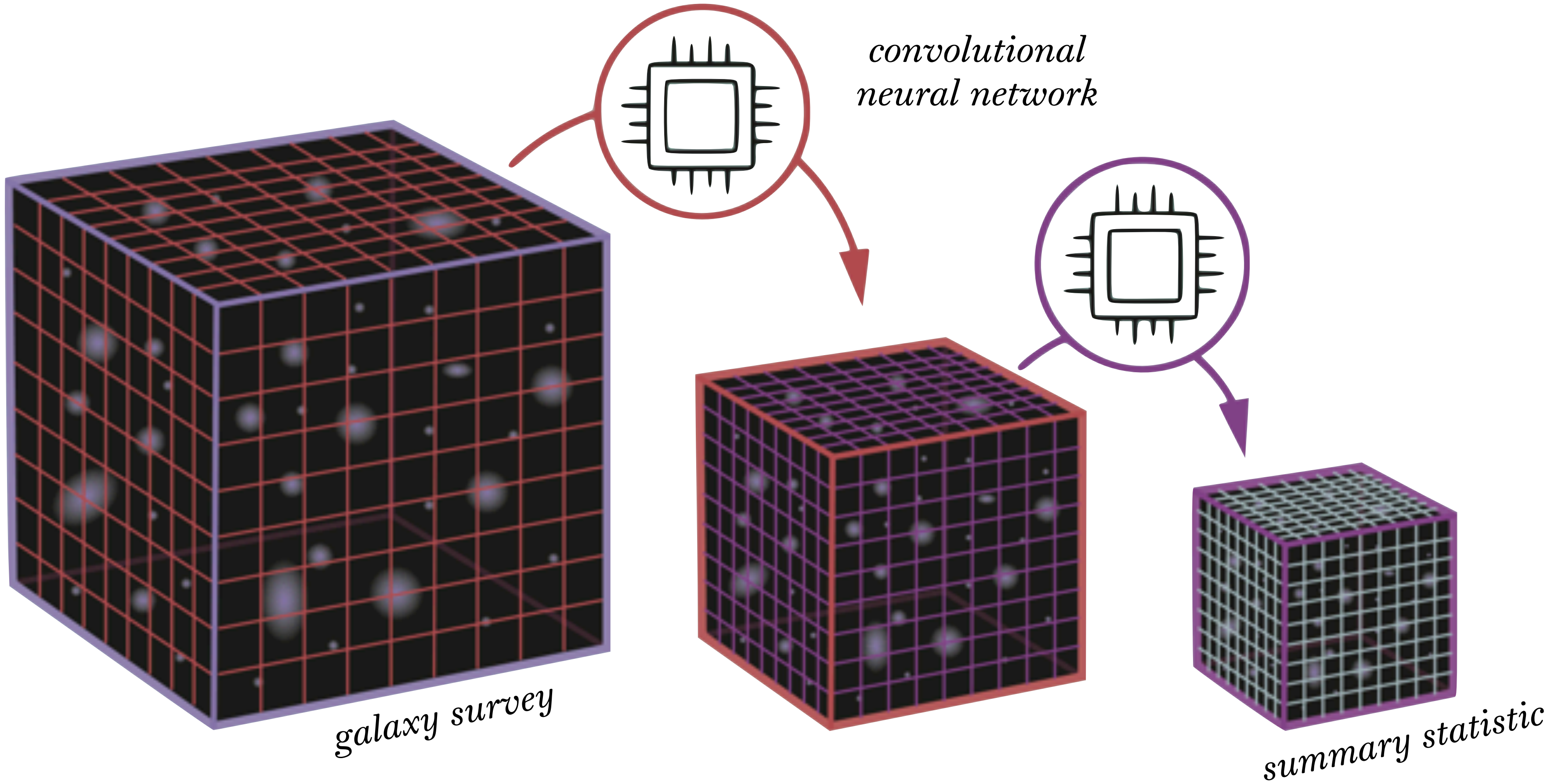
SimBIG: non-linear galaxy bispectrum $B_0(k_1, k_2, k_3 < 0.5\,h/\mathrm{Mpc})$



1.2 and 2.4 $\times$ *tighter* Ω_m and σ_8 from **non-linear** + **higher-order** clustering

Hahn *et al. (2023h)*

SIMBIG: convolutional neural network field-level summary



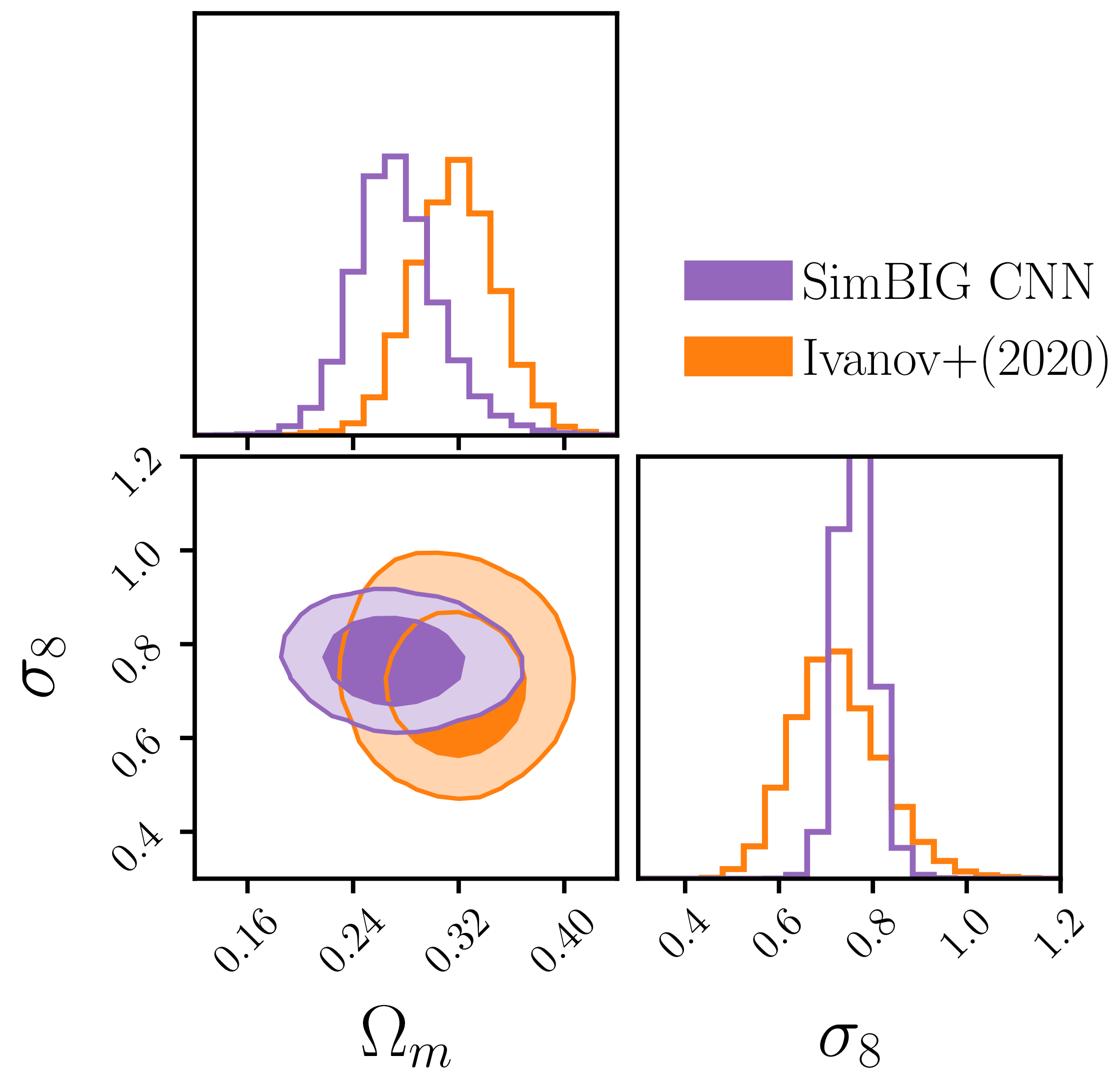
extracting *all* relevant cosmological information in N-pt functions

*Lemos, Parker, **Hahn** et al. (2023)*

SimBIG: convolutional neural network field-level summary

Liam Parker
Princeton Univ.

Pablo Lemos
MILA



extracting *all* relevant cosmological information in N-pt functions

*Parker, Lemos, **Hahn** et al. (2023)*

SIMBIG

wavelet scattering transforms

*Régaldo-Saint Blancard, **Hahn** et al. (2023)*

Bruno Régaldo-Saint Blancard
CCM Flatiron

skew spectra

*Hou, Moradinezhad Dizgah, **Hahn** et al. (2024)*

Jiamin Hou
Univ. of Florida

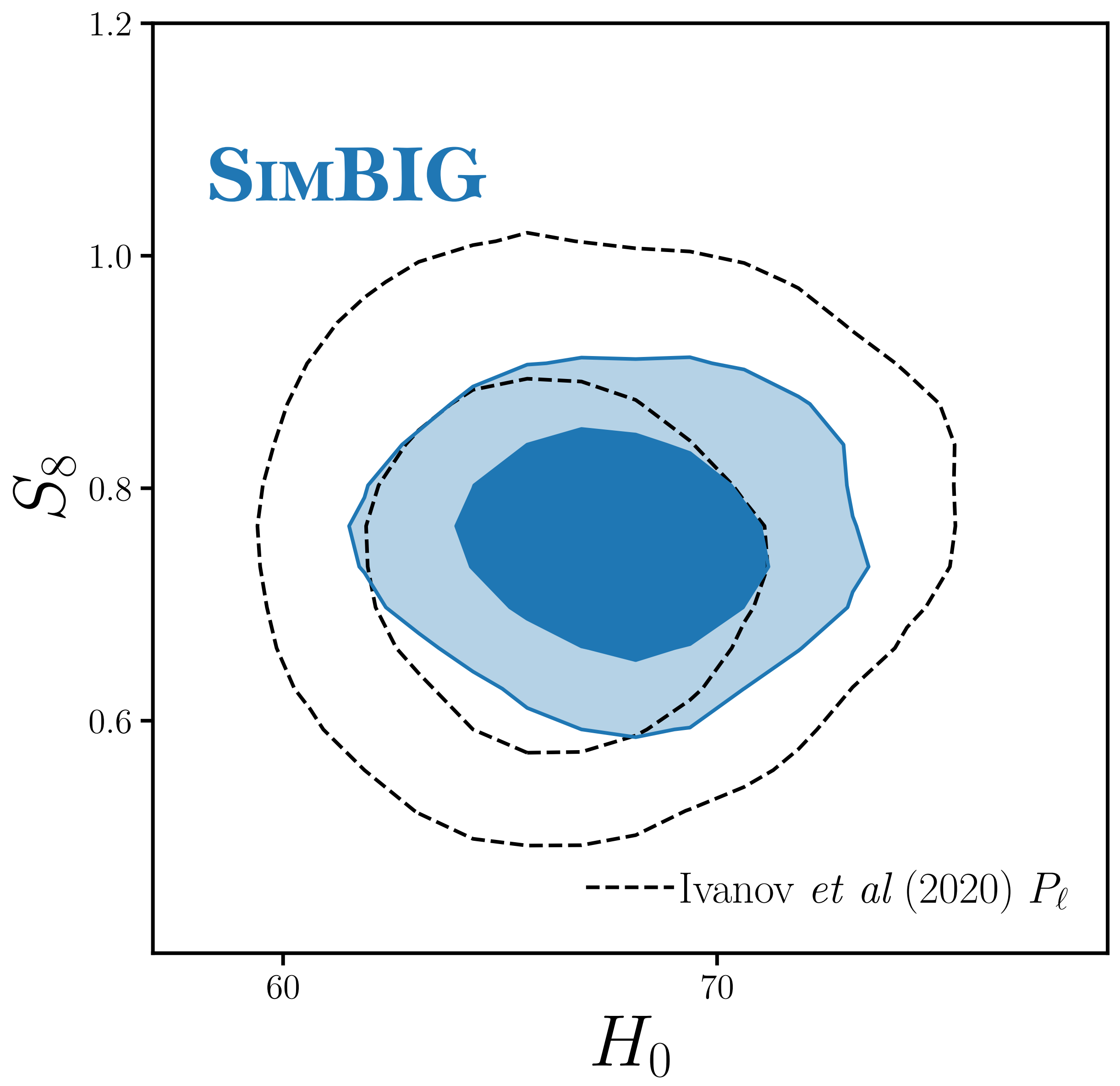
marked powerspectrum

*Massara, **Hahn** et al. (2024)*

Elena Massara
UWaterloo

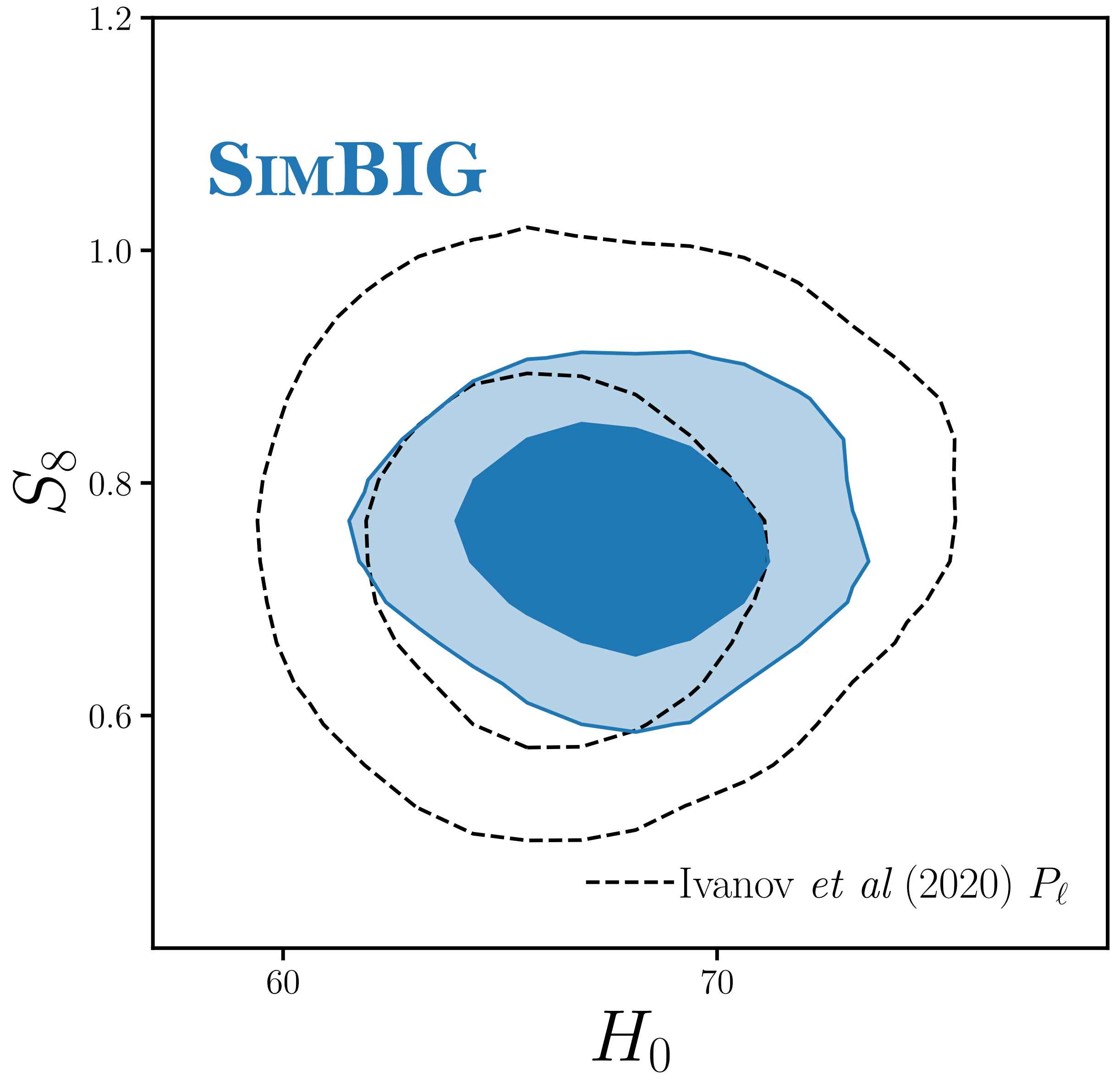
voids, graph neural network, combined … *coming soon*

SimBIG: **~1.9 and 1.5× tighter** S_8 and H_0



Hahn *et al. (2023i)*

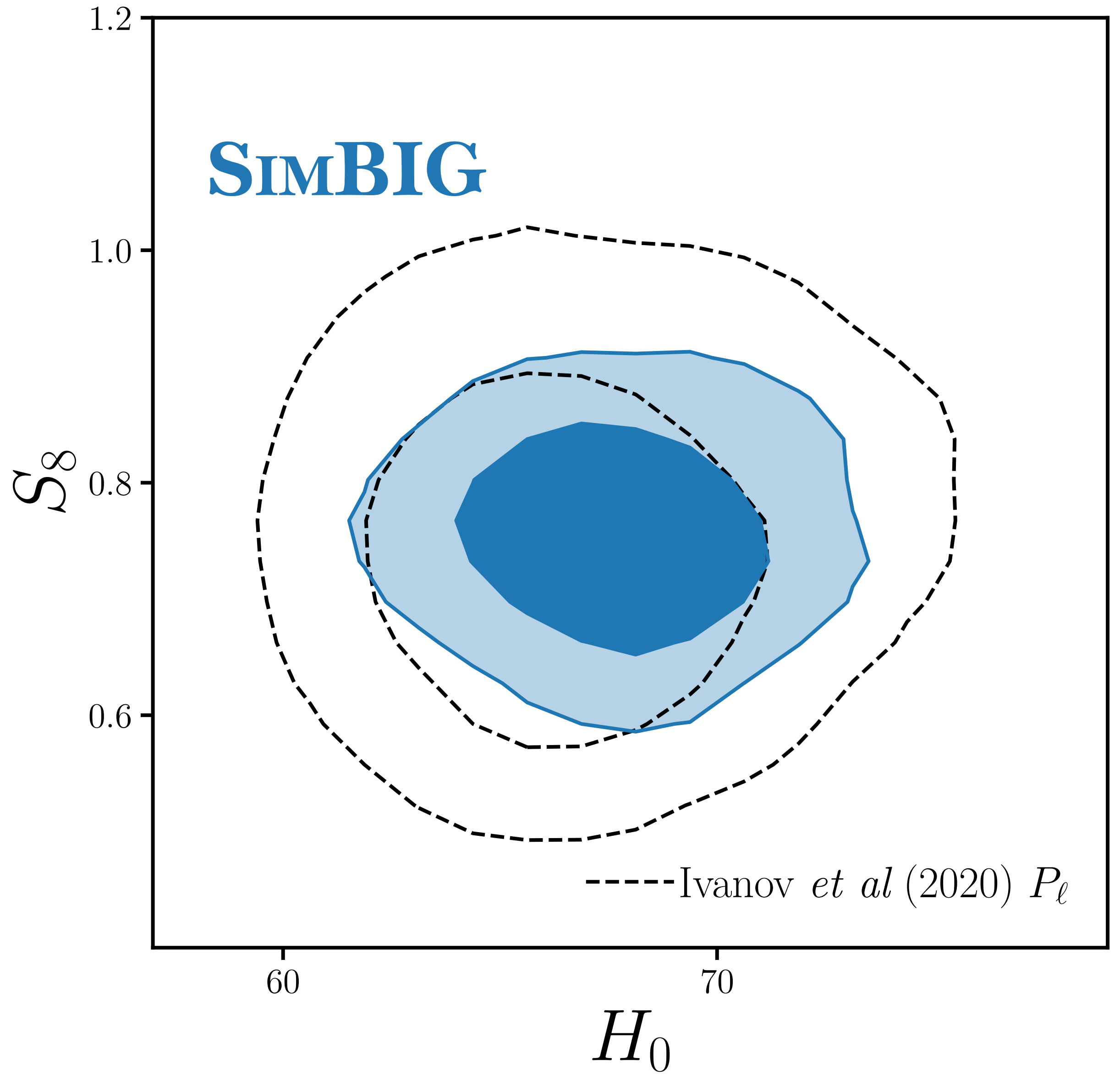
SIMBIG: **~1.9 and 1.5× tighter** S_8 and H_0



production level cosmological constraints — *not a proof-of-concept!*

***Hahn** et al. (2023i)*

SimBIG: **~1.9 and 1.5× tighter** S_8 and H_0



S_8 improvement is equivalent to analyzing a *survey of ~4× larger volume*

Hahn *et al. (2023i)*


~100,000 galaxies at z ~ 0.5
galaxy surveys

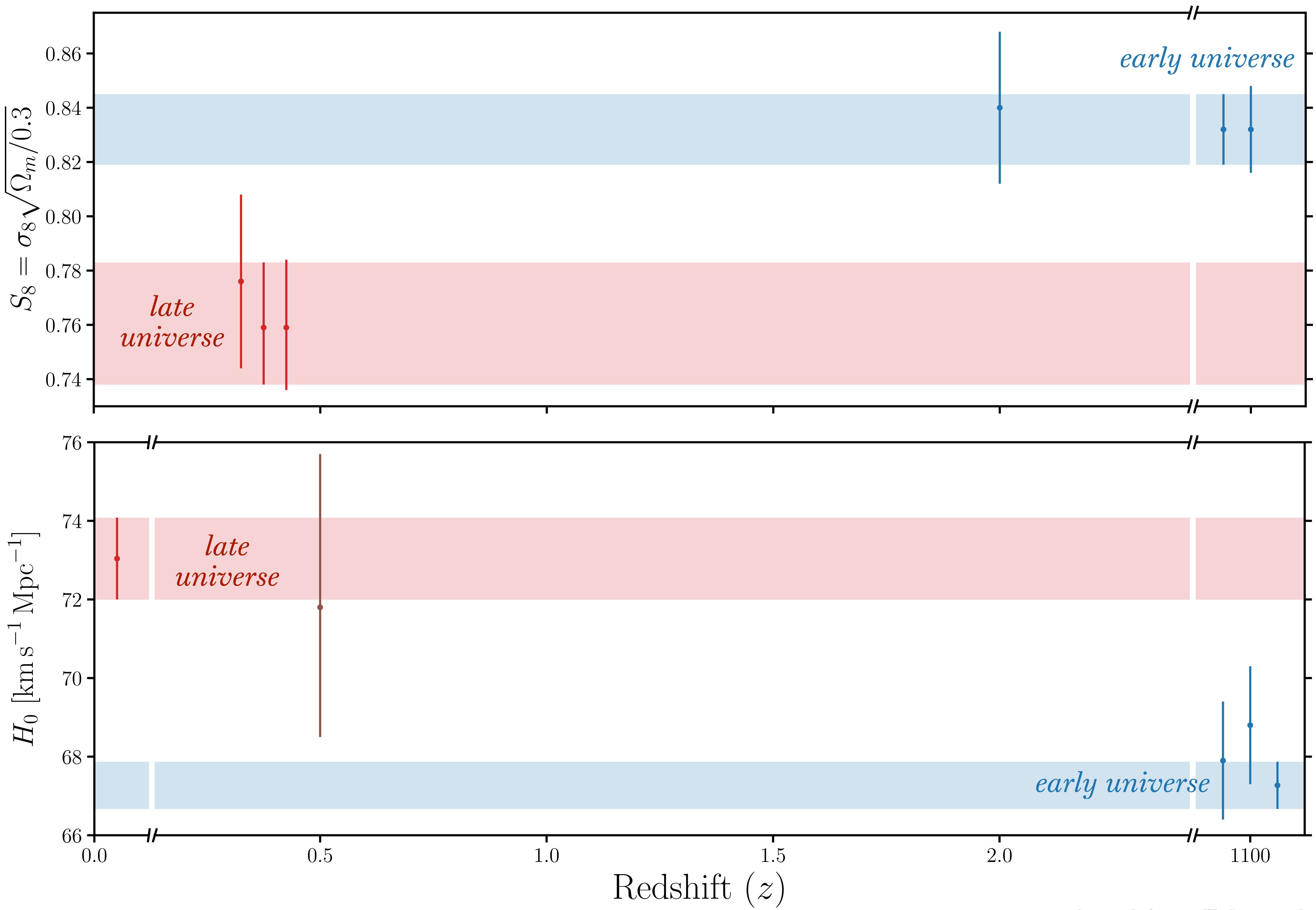

~100,000 *galaxies* at $z \sim 0.5$
galaxy surveys

Roman *High Latitude Spectroscopic Survey*

~10 million Hα Emission-Line Galaxies $1 < z < 2$

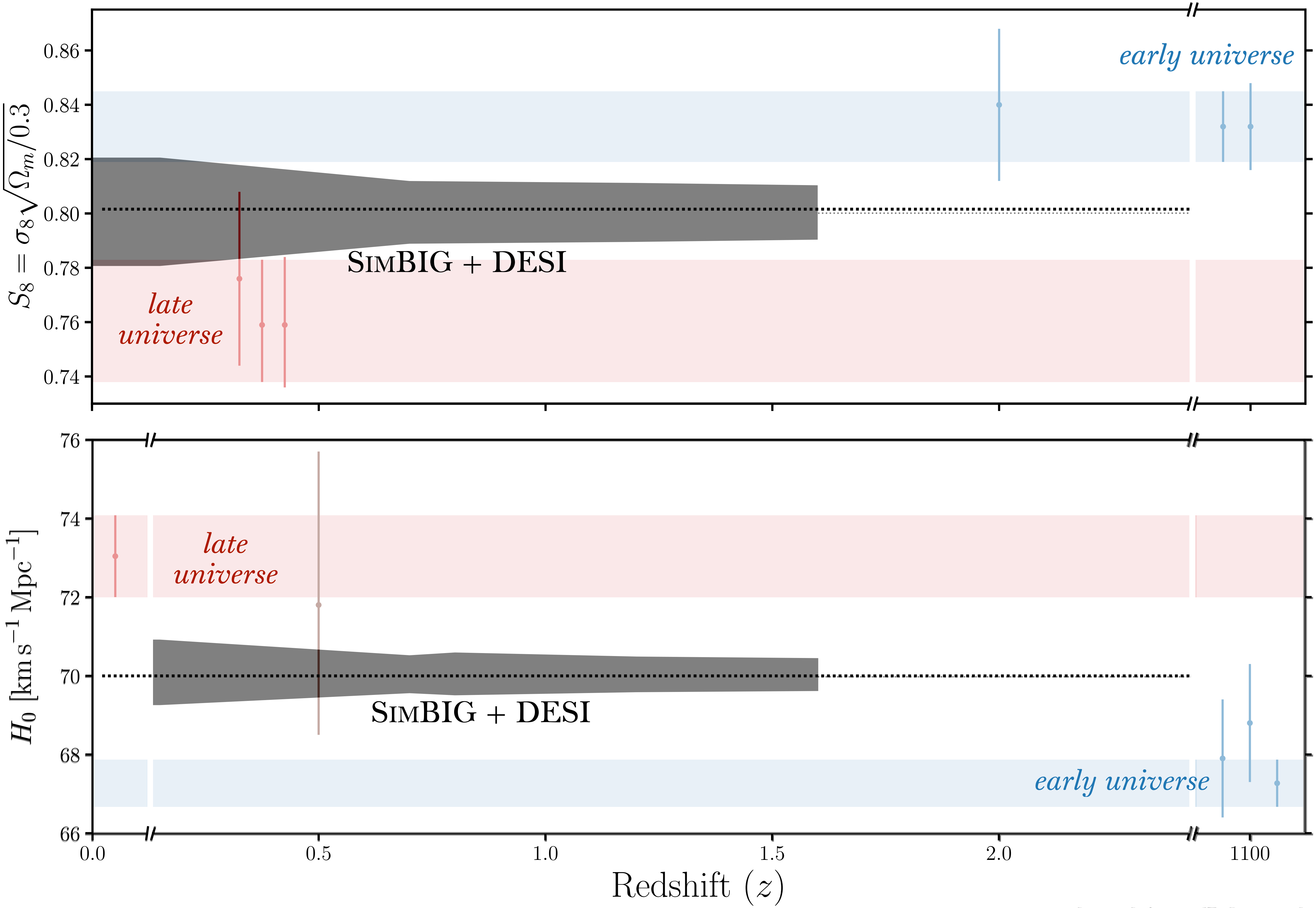
~2 million OIII Emission-Line Galaxies $2 < z < 3$

Wang et al. (2022)



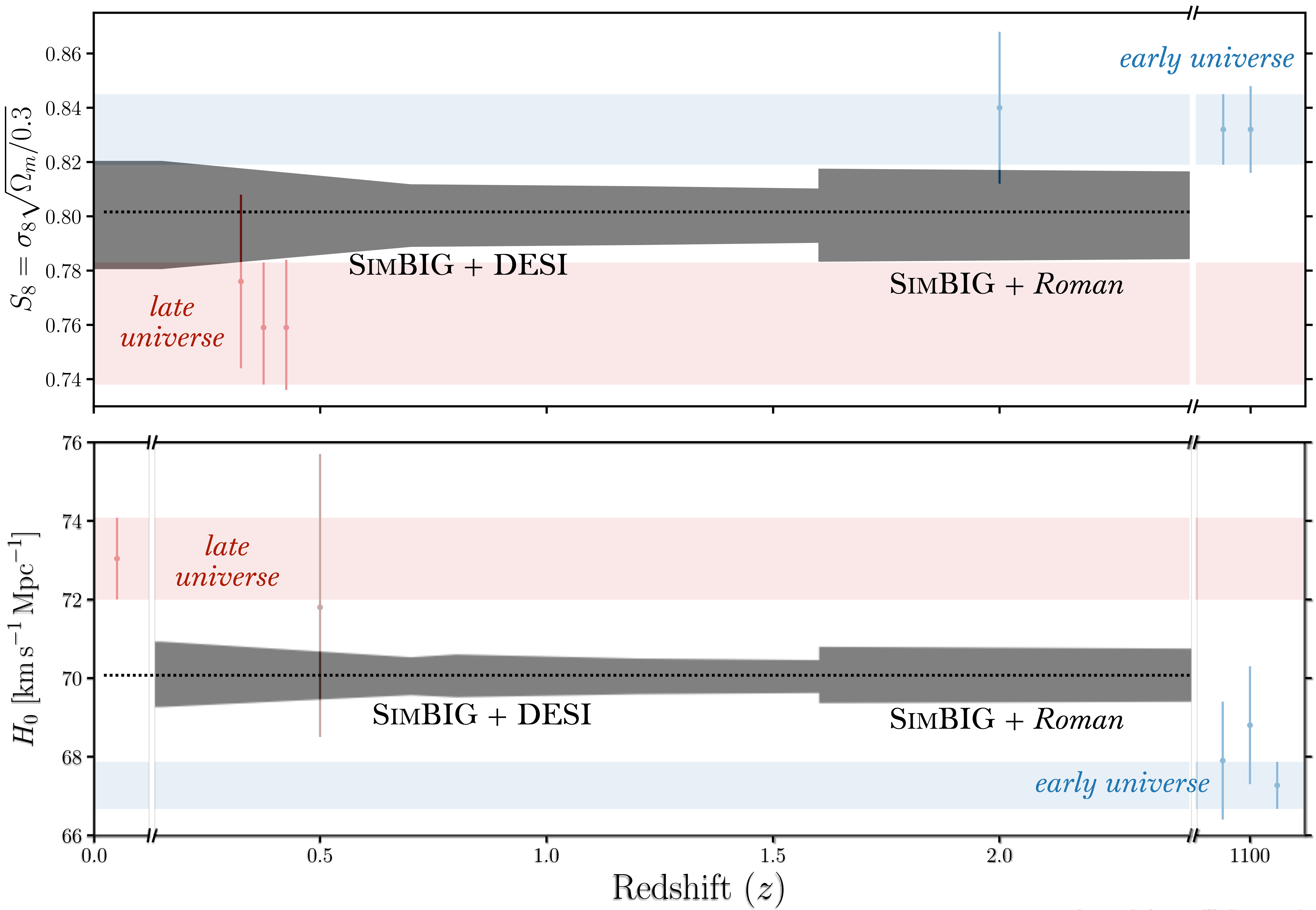
adapted from ***Hahn*** *et al. (2023i)*

SimBIG + DESI and *Roman* will probe *new regimes*



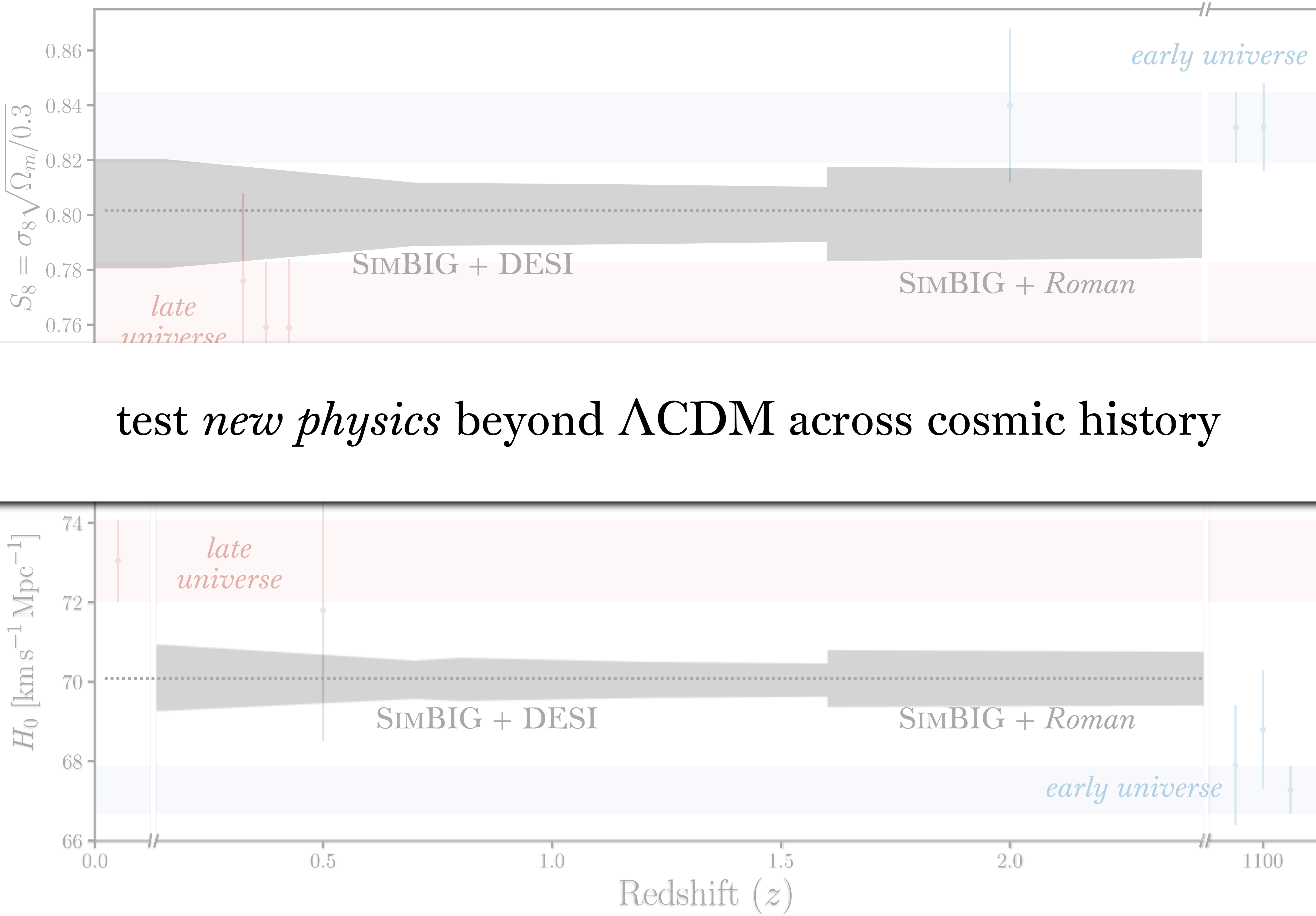
adapted from ***Hahn*** *et al. (2023i)*

SimBIG + DESI and *Roman* will probe *new regimes*



adapted from ***Hahn*** *et al. (2023i)*

SimBIG + DESI and *Roman* will probe *new regimes*



test *new physics* beyond ΛCDM across cosmic history

*adapted from **Hahn** et al. (2023i)*

galaxy surveys encode the *growth* and *expansion histories* of the Universe

ML×Cosmo: SimBIG analyses leverage *non-linear* and *higher-order* galaxy clustering to **double** the cosmological impact of galaxy surveys

Roman with SimBIG will *settle* cosmic tensions and probe *new physics*

ChangHoon Hahn
changhoon.hahn@princeton.edu
changhoonhahn.github.io