Data handling in Cosmology

Soumak Maitra
DTP, TIFR

ML4HEP Pre-School, 25th May 2026

The questions Cosmology tries to answer

- **What is the Universe made of?**

  Ordinary matter, dark matter, radiation, neutrinos, dark energy all contribute to the cosmic energy budget.

- **How did the Universe begin and evolve?**

  Cosmology studies the history from Big Bang to our current Universe.

- **How did the structures in the Universe form?**

  The growth of structure from primordial density fluctuations to Galaxies, clusters, filaments, voids.

- **What is the fate of the Universe?**

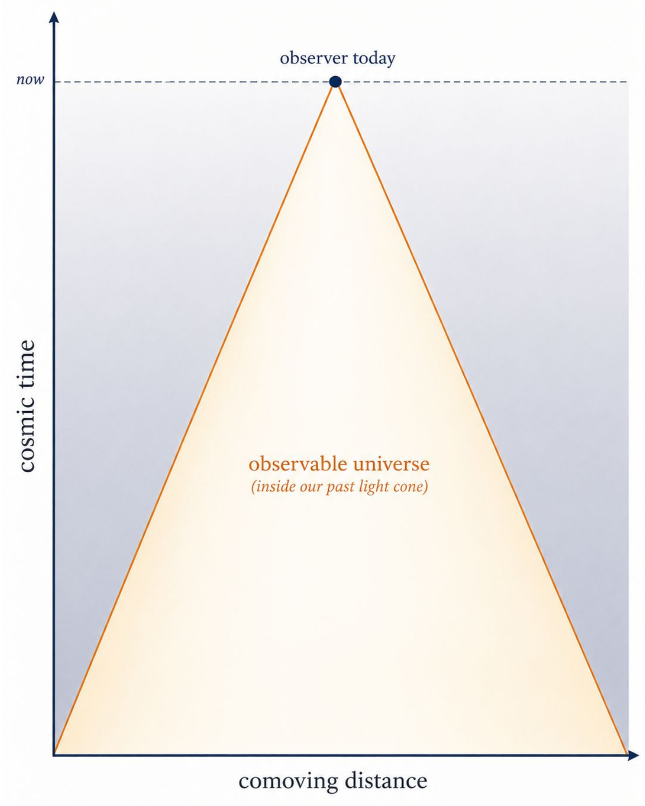
  Depends on the expansion history, spatial curvature and nature of dark energy.

In practice, many of these physical questions are encoded in a set of cosmological parameters that describe the contents, expansion history, and growth of structure in the Universe.

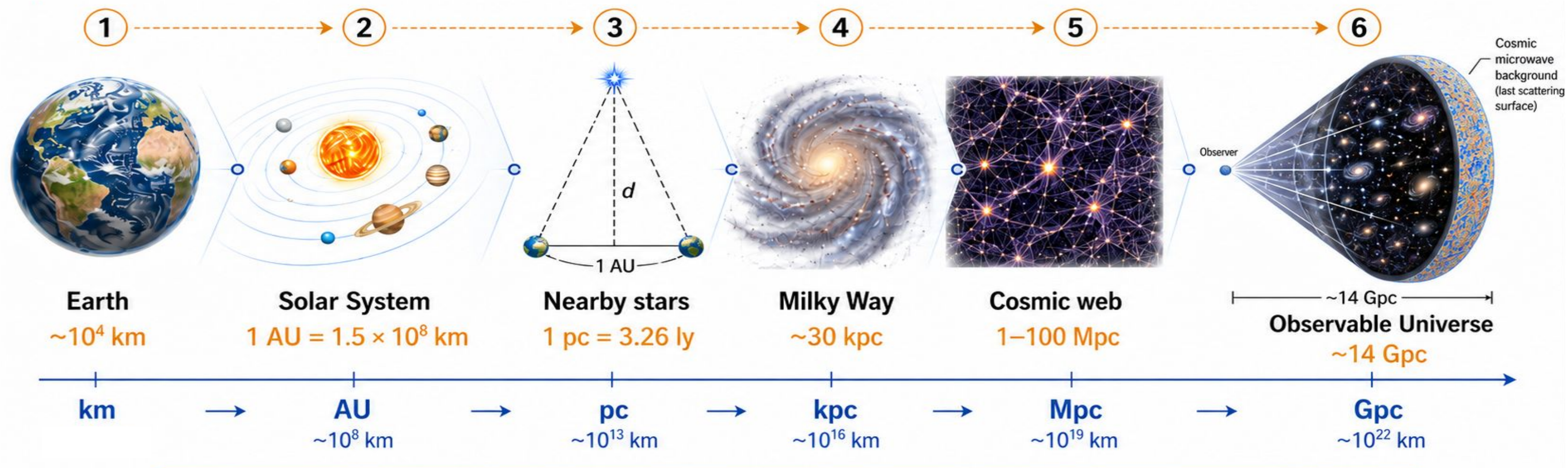
The Observable Universe:
What Can We Actually see?

In this lecture, the **Universe** refers to the **Observable Universe**.

- **Observable universe** is the region from which light has had time to reach us. Limited by finite light speed and age of Universe.
- **Whole universe** may be larger than what we can observe directly.
- **Looking farther away also means looking further back in cosmic time.** Distance and **lookback time** are therefore entangled in cosmological data. The collection of all light reaching us today forms our **past light cone**.



Cosmological Distance scales



Why is cosmology different from a laboratory experiment?

Cosmology deals with extremely large scales. Distance, time and evolution are tied together.

Cosmology studies the Universe as a single physical system that can be measured through observations and described with mathematical models.

- Unlike laboratory experiments, we cannot rerun the Universe under different conditions. Instead, we infer its history by comparing observations with physical models.

- The aim is not simply to catalogue astronomical objects, but to understand how different observations constrain the origin, composition, evolution, and structure of the Universe.

We observe limited pieces of the Universe and use them to infer the larger physical story that connects them.

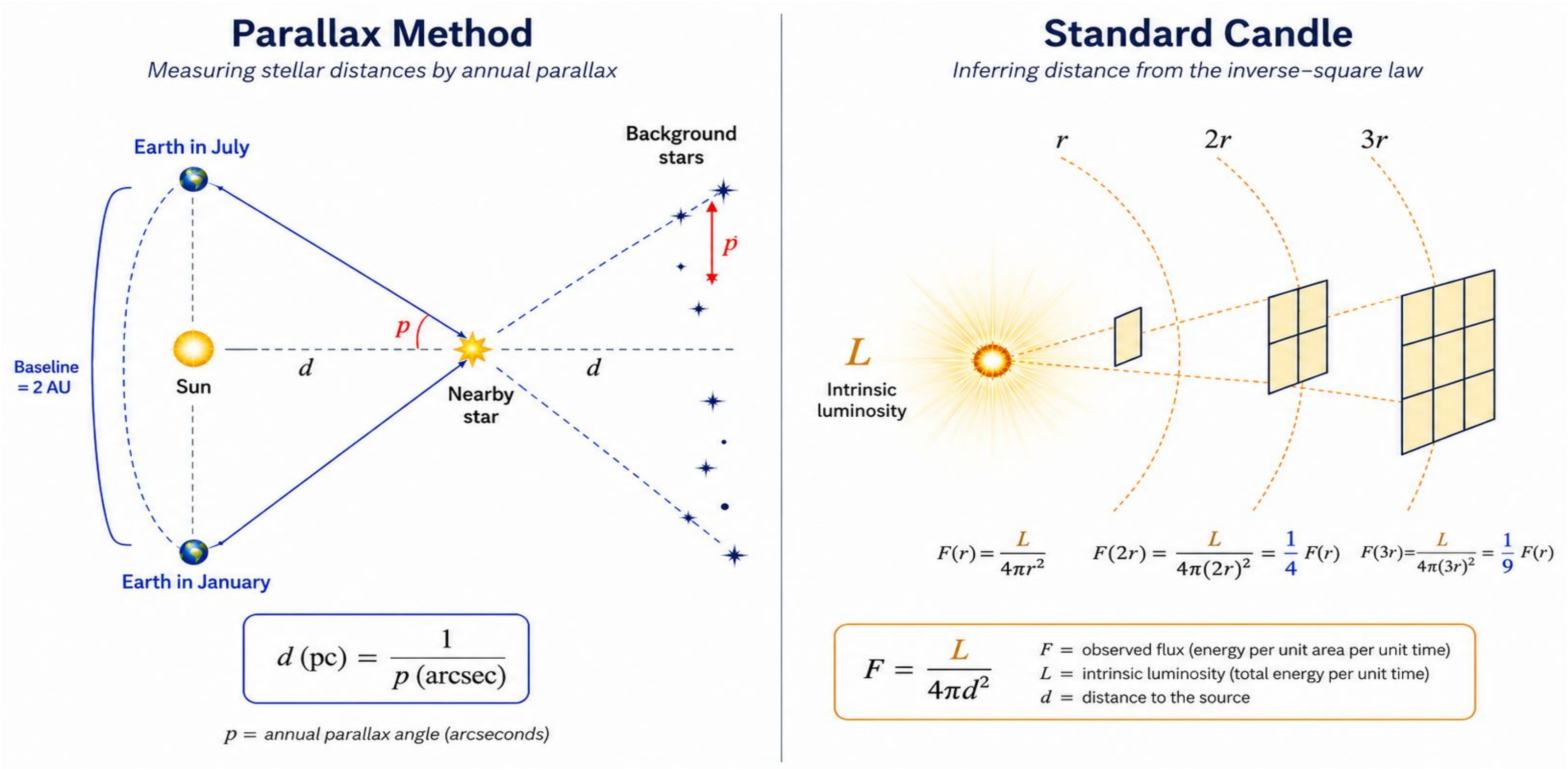
Cosmology is a historical science; we reconstruct the Universe's past using light, matter tracers, and statistical patterns observed today.

Cosmological data: What does a telescope measure?

Cosmological data do not give us the Universe directly. They give us light, positions, spectra, fluxes, images, and counts, which we interpret through physical models.

- Each photon carries information about where, when, and how it was emitted or absorbed.
- A telescope measures photons arriving here and now.
- Cosmological observables begin as the following properties of photons, not as parameters of model.
    - ❖ **Direction** on the sky: angular position
    - ❖ **Wavelength/frequency**: spectral information and redshift
    - ❖ **Flux/counts**: brightness after calibration
    - ❖ **Image shape/polarization/time variation**: probe-dependent information
    - ❖ **Noise, masks, resolution, selection**: part of the measurement

How do astronomers measure distances?



Hubble's Law: The First Evidence for an Expanding Universe

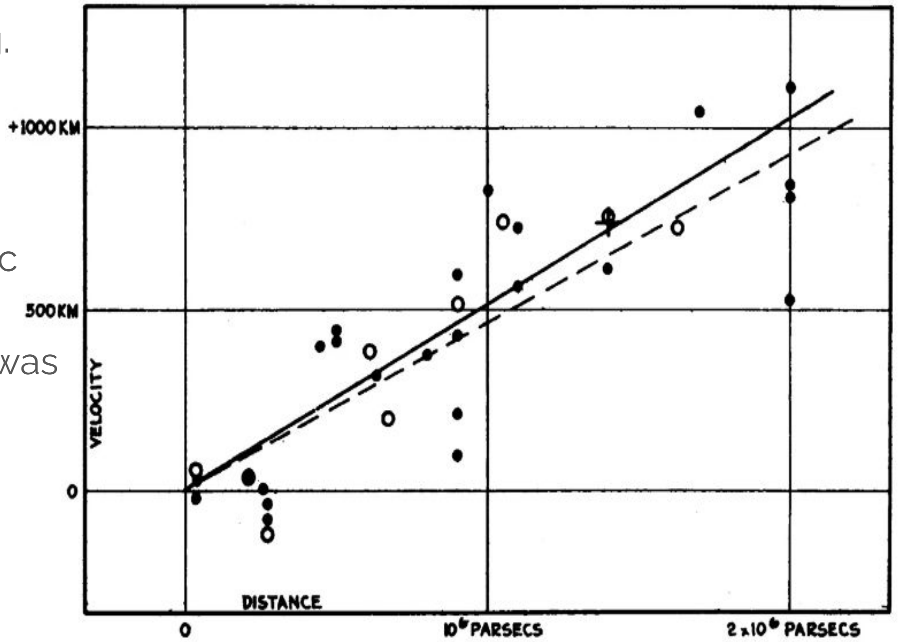
- The distance to the standard candles (eg. Cepheid Variables) were estimated.

$$d = \sqrt{\frac{L}{4\pi F}}$$

- Galaxy spectra contain identifiable atomic lines. Many were observed at longer wavelengths than in the laboratory. This was interpreted as a recession speed:

$$\frac{\Delta\lambda}{\lambda_{\rm lab}} \simeq \frac{v}{c}$$

- Hubble found a linear trend $v \simeq H_0 d$ (**Hubble's Law**)



Hubble 1929

The Universe is not static. On large scales, distances between galaxies grow with time.
The Universe is expanding.

Expansion means the cosmic distance scale changes

On large scales, the physical distance between objects changes because the scale of the Universe changes with time.

$$r(t) = a(t)\chi$$

where:

- χ : **comoving** distance
- *a* (t) : scale factor
- *r* (t) : **physical** distance

If *a* (t) grows with time, distances between objects grows. This naturally gives:

$$v = \dot{r} = \frac{\dot{a}}{a} r = H(t) r.$$

At present time, $v \simeq H_0 d$

From expansion to redshift, an observational coordinate

Light travelling through expanding space is stretched. Therefore, a photon emitted with wavelength $\boldsymbol{\lambda}_{\mathbf{em}}$ is observed with a longer wavelength $\boldsymbol{\lambda}_{\mathbf{obs}}$.

$$\lambda_{\rm obs} > \lambda_{\rm em}$$

This is called **redshift**

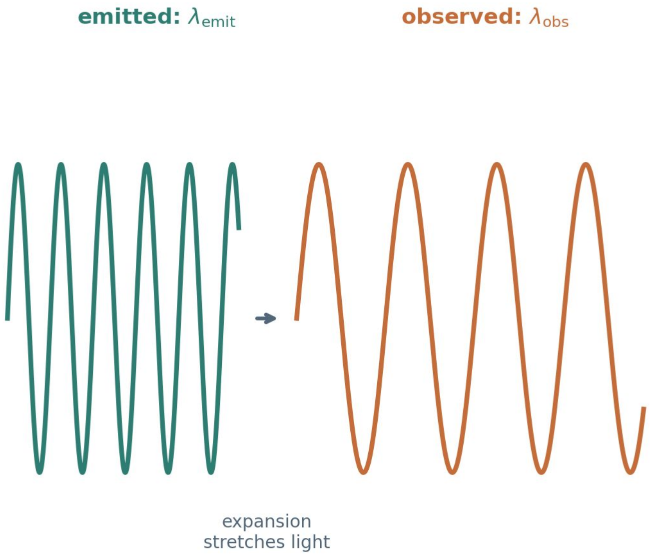
$$1 + z = \frac{\lambda_{\rm obs}}{\lambda_{\rm em}}.$$

In an expanding Universe

$$1 + z = \frac{a(t_0)}{a(t_{\rm em})}.$$

With $a(t_o) = 1$,

$$a(t_{\rm em}) = \frac{1}{1 + z}.$$



Redshift tells us how much the Universe has expanded since the light was emitted.

From expansion to redshift, an observational coordinate

- A telescope directly measures angular position on the sky and spectral information.

- For a galaxy survey, the basic observed coordinates are **(θ , ϕ , *z*)**

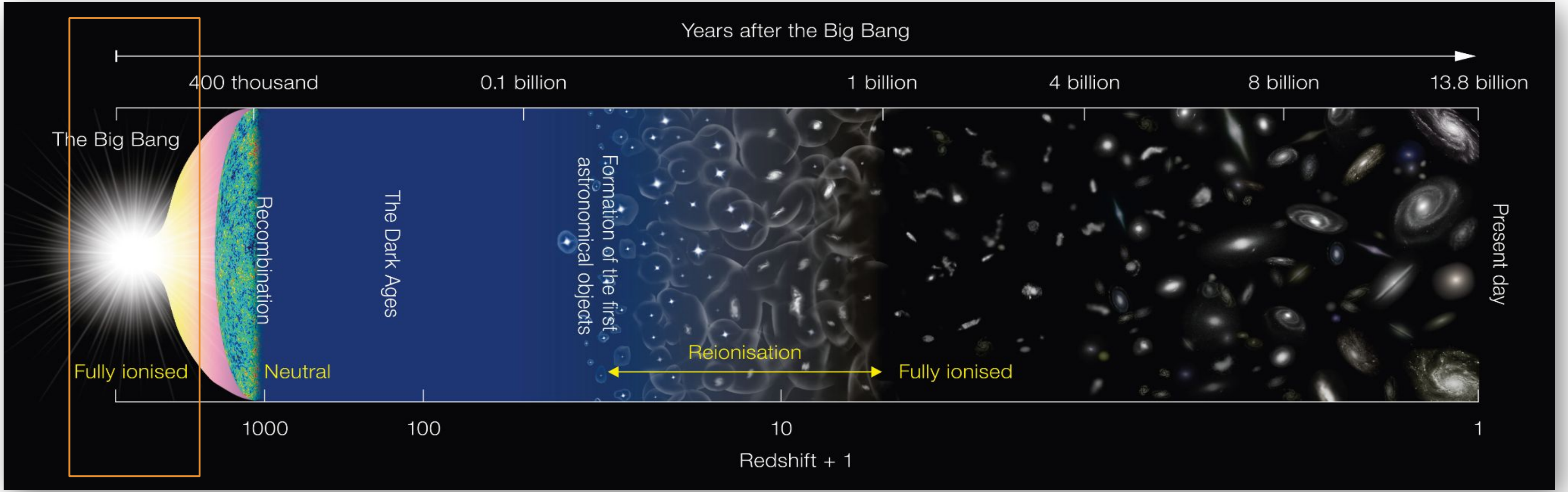
- The angles **(θ , ϕ)** locate the object on the sky.
  The redshift **z** tells us where it lies along the line of sight.

- With a cosmological model, redshift can be converted into distance.

$$\chi(z) = c \int_0^z \frac{dz'}{H(z')}$$

Redshift acts both as a coordinate along the cosmic light cone and as a marker of cosmic time. Higher redshift means we are observing an earlier epoch of the Universe.

A brief history of Cosmic Timeline



Big Bang (t=0) :

- Hot dense Universe
- Smooth Initial state
- Beginning of Cosmic History

Inflation ($t \sim 10^{-36}s - 10^{-32}s$) :

- Extremely rapid expansion
- Quantum fluctuations
- Seeds of structure

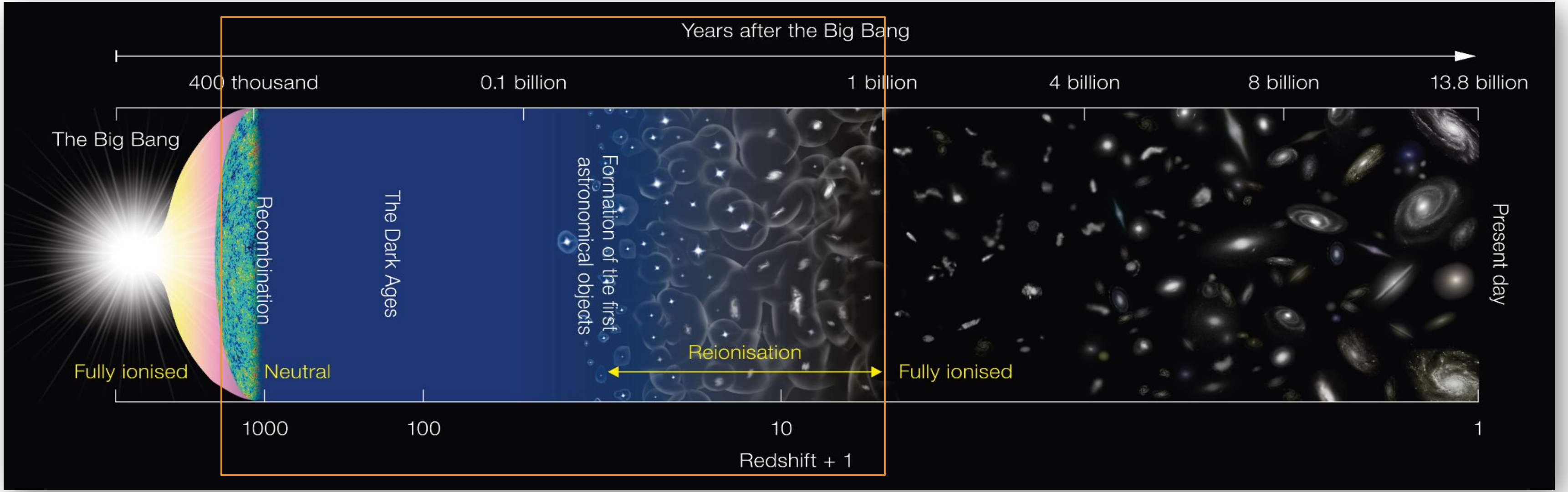
Particle Plasma (t~1s) :

- Hot particle soup
- Radiation domination
- Cooling with expansion

Big Bang Nucleosynthesis (t~3-20min) :

- First nuclei, Hydrogen, Helium, trace Lithium.
- No neutral atoms yet.

A brief history of Cosmic Timeline



Recombination (t=380,000 yr) :

- Neutral atoms,
- Photon decoupling
- Transparent Universe
- Cosmic Microwave Background

Dark Ages (t=0.4-100 Myr) :

- Neutral hydrogen.
- No luminous source
- Gravitational collapse, cosmic structure growth.

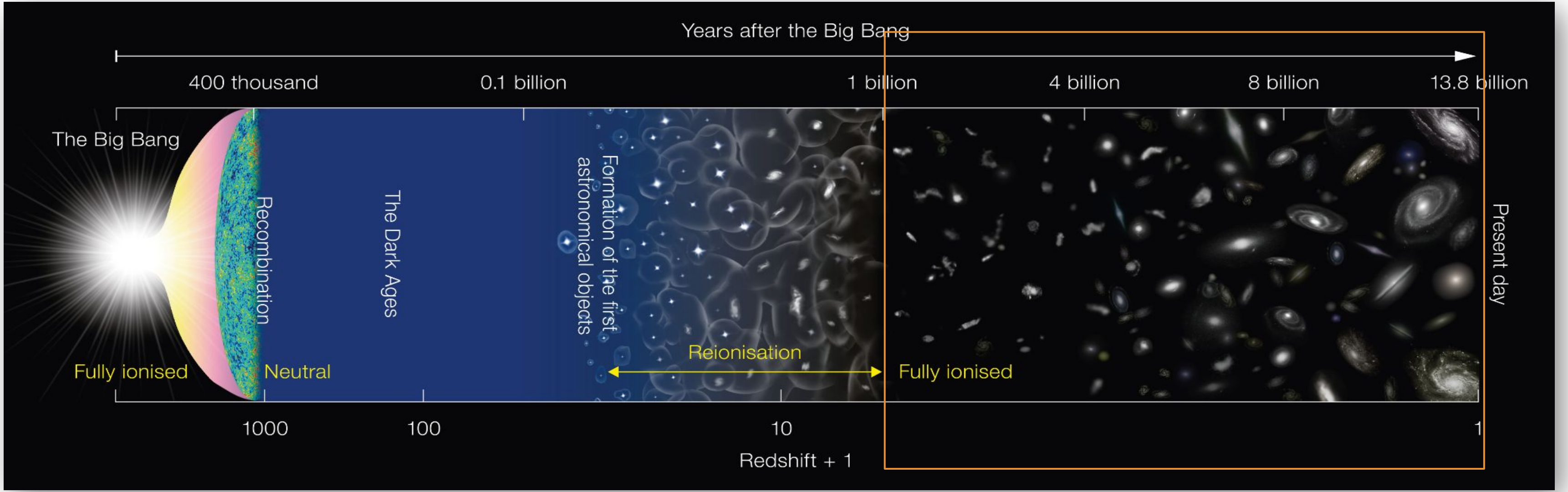
Cosmic Dawn (t~100 Myr) :

- Dark matter halos, gas cooling.
- First stars and galaxies.
- UV radiation begins

Reionization (t~0.1-1 Gyr) :

- UV photons from galaxies ionizes the Universe.
- Formation of ionized bubbles in IGM.

A brief history of Cosmic Timeline



Cosmic Web (t> 1Gyr) :

- Large scale structures, Gravity-driven growth.
- Formation of voids, filaments, sheets, nodes.

Cosmic Noon (t=2-4 Gyr) :

- Peak star formation
- Rapid galaxy growth
- Active black holes
- Bright quasars
- Metal enrichment

Accelerated Expansion (z<0.7) :

- Dark energy domination.
- Accelerating expansion.
- Slower structure growth
- Expanding cosmic distances
- Late-time cosmology

Universe is homogeneous and isotropic at large scales

- At sufficiently large scales (>100Mpc), the Universe is:
  - Homogenous (no preferred position).
  - Isotropic (no preferred direction)
- Cosmic evolution is determined by scale factor a(t).
- Local structures exist, but are smoothed over large scales.
- Under these symmetry conditions, Friedmann–Lemaître–Robertson–Walker (**FLRW**) metric describes the expanding Universe.

$$ds^2 = -c^2dt^2 + a^2(t)\left[\frac{dr^2}{1-kr^2} + r^2 d\Omega^2\right]$$

- $a(t)$: scale factor
- $k = 0, +1, -1$: spatial curvature
- t: cosmic time
- r, θ, ϕ: comoving coordinates

What controls Universe's expansion?

- The Universe contains different components:

$$\rho_{\rm tot} = \rho_m + \rho_r + \rho_\Lambda + \rho_k$$

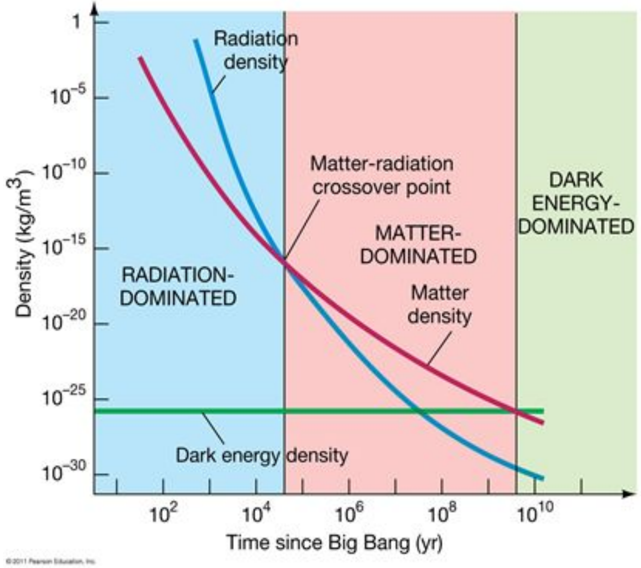
- **Matter**: **Baryonic Matter (~15%), Dark matter (~85%)**
- **Radiation**: photons, relativistic species
- **Dark energy**: accelerated expansion
- **Curvature**: spatial geometry

- Different components evolve differently with time.

$$\rho_m \propto a^{-3}$$

$$\rho_r \propto a^{-4}$$

$$\rho_\Lambda = \text{constant}$$



The Friedmann equation

$$H^2(t) = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{kc^2}{a^2} + \frac{\Lambda c^2}{3}$$

$$H^2(a) = H_0^2\left[\Omega_r a^{-4} + \Omega_m a^{-3} + \Omega_k a^{-2} + \Omega_\Lambda\right]$$

$$H^2(z) = H_0^2\left[\Omega_r (1+z)^4 + \Omega_m (1+z)^3 + \Omega_k (1+z)^2 + \Omega_\Lambda\right]$$

Main cosmological parameters:

- H_0: expansion rate today
- Ω_m: matter density
- Ω_b: baryon density
- Ω_Λ: dark energy density
- Ω_k: curvature

Structure formation in the background smooth Universe

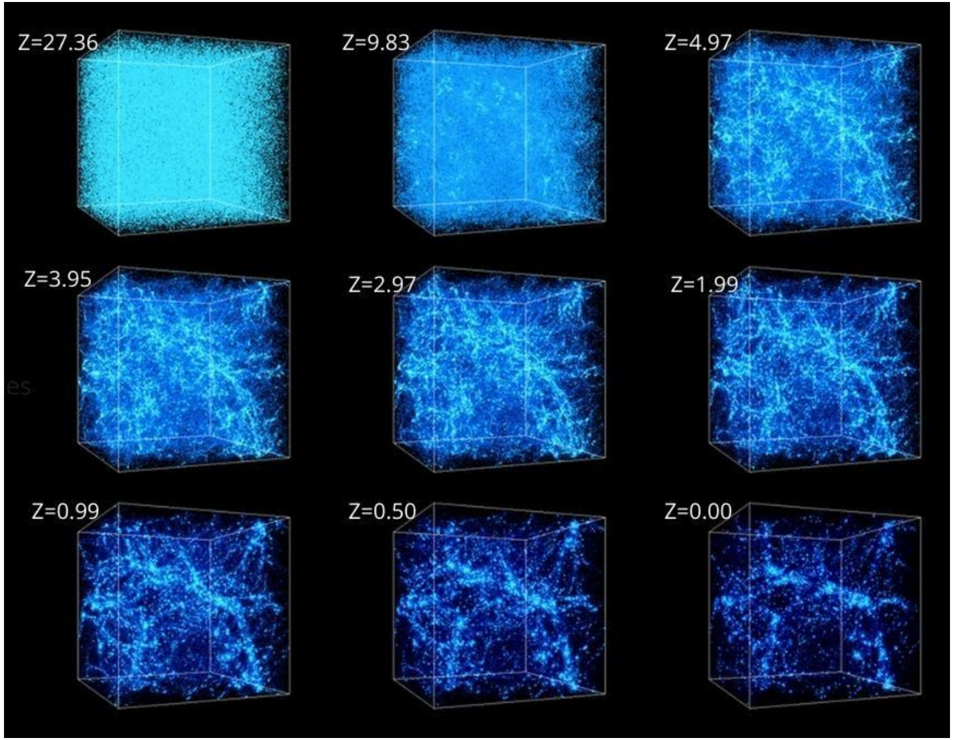
- FLRW describes the homogeneous and isotropic expanding Universe.
- Real universe: small departures from the mean density.
- These perturbations grow under gravity.

$$\rho_m(\mathbf{x},t) = \bar{\rho}_m(t)\,[1+\delta_m(\mathbf{x},t)]$$

$$\delta_m(\mathbf{x},t) = \frac{\rho_m(\mathbf{x},t)-\bar{\rho}_m(t)}{\bar{\rho}_m(t)}$$

- $\bar{\rho}_m(t)$: homogeneous background density.
- $\delta_m > 0$: overdense regions.
- $\delta_m < 0$: underdense regions.
- Structure formation is the evolution of $\delta_m(\mathbf{x},t)$.

The background tells us how the universe expands; δ_m tells us how matter clusters.



Artist's impression of the cosmic web, showing diffuse intergalactic matter distributed along filaments between galaxies. Image credit: Mark Garlick / Science Photo Library / Getty Images.

Structure formation in the background smooth Universe

- Linear perturbation equation:

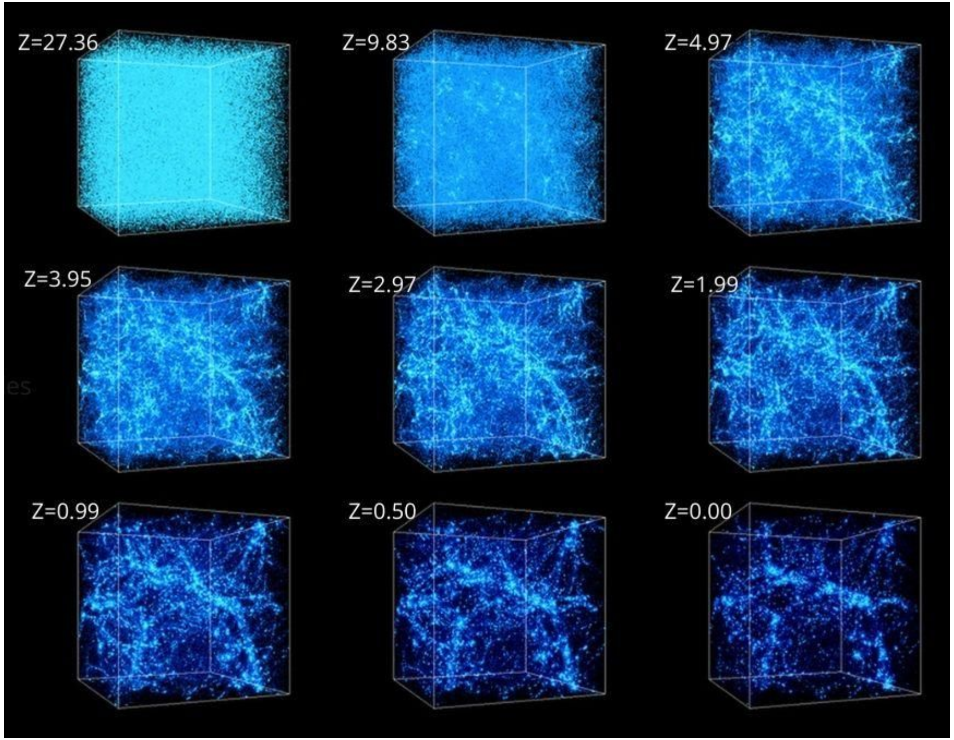
$$\ddot{\delta} + 2H\dot{\delta} - 4\pi G\bar{\rho}_m\delta = 0$$

$$\delta(\mathbf{x}, a) = D(a)\,\delta(\mathbf{x}, a_{\text{ini}})$$

where $D(a)$ is the **linear growth factor**.

- Gravity makes overdensities grow. Dark matter dominates the contribution.
- Expansion acts as a friction term through $2H\dot{\delta}$.
- The growth history depends on cosmological parameters like Ω_m, Ω_Λ, σ_8, and initial conditions.

The background tells us how the universe expands; δ_m tells us how matter clusters.



Artist's impression of the cosmic web, showing diffuse intergalactic matter distributed along filaments between galaxies. Image credit: Mark Garlick / Science Photo Library / Getty Images.

Matter power spectrum: Structure as a function of scale

Fourier transform:

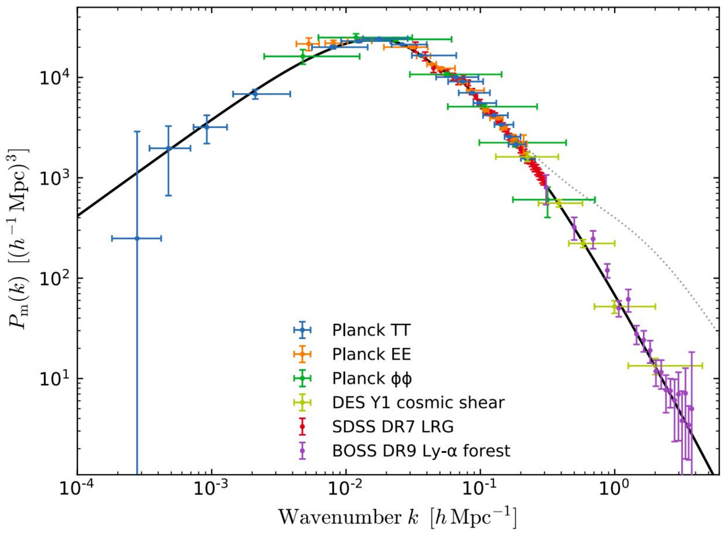
$$\delta_m(\mathbf{x}) \rightarrow \delta_m(\mathbf{k})$$

Power spectrum:

$$\langle \delta_m(\mathbf{k})\delta_m^*(\mathbf{k}')\rangle = (2\pi)^3\delta_D(\mathbf{k}-\mathbf{k}')P_{mm}(k)$$

Dimensionless power:

$$\Delta^2(k) = \frac{k^3 P(k)}{2\pi^2}$$

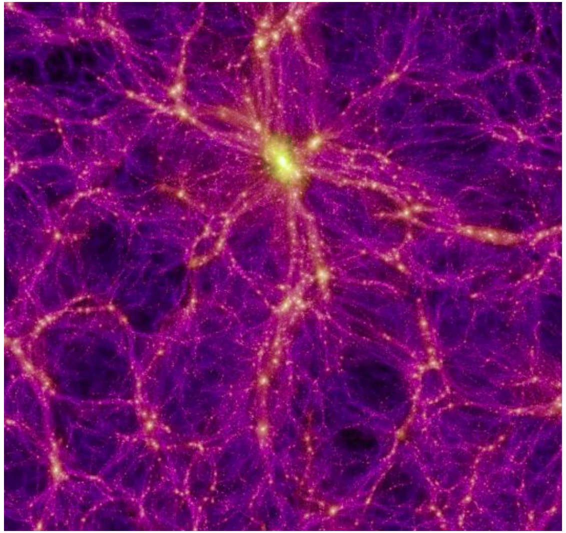


Planck Collaboration 2020, A&A 641, A1

- Small k: large physical scales.
- Large k: small physical scales.
- P(k) measures variance per Fourier mode.
- Shape of P(k) contains crucial cosmological information.
- Cosmology usually predicts statistics, not the exact position of every galaxy.

The power spectrum compresses the density field into fluctuation amplitude versus scale.

Tracers: What we actually observe



Millennium Simulation

Credit: ESA/Webb, NASA & CSA, G. Gozaliasl, A. Koekemoer, M. Franco, and the COSMOS-Web team

We want the matter field, but we observe tracers of it.

Tracers: What we actually observe

Underlying field:

$$\delta_m(\mathbf{x})$$

Observed tracer fields:

$$\delta_g(\mathbf{x}), \quad \delta_q(\mathbf{x}), \quad \delta_{\mathrm{Ly}\alpha}(\mathbf{x}), \quad \kappa(\mathbf{x}), \quad T_{21}(\mathbf{x}), \quad T_{\mathrm{CMB}}(\hat{\mathbf{n}}), \quad \mu_{\mathrm{SN}}(z)$$

Examples:

Tracer	What it roughly traces	Data product
Galaxies	biased matter peaks	catalog
Quasars	rare massive environments	sparse catalog
Lyα forest	neutral hydrogen / IGM density	spectra
Weak lensing	projected matter field	shear map
21 cm	neutral hydrogen field	intensity map/cube
Clusters	massive halo population	catalog
CMB	primordial fluctuations / projected early-Universe field	temperature / polarization maps
Supernovae	expansion history / distance indicator	light-curve / distance catalog

Tracers: What we actually observe

Galaxies: Luminous tracers of cosmic structures

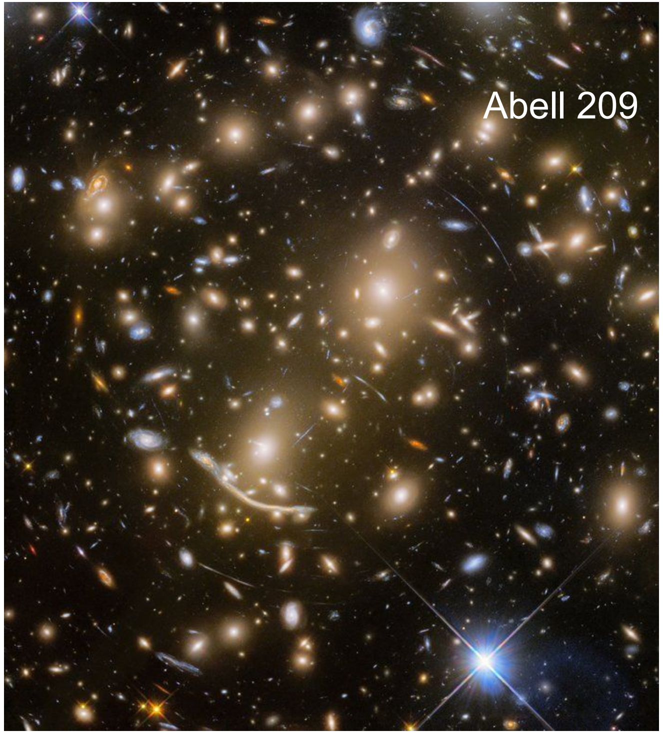
- Massive systems of stars, gas, and dust held together by gravity.
- Observed as positions, redshifts, luminosities, colours.
- Discrete tracers of high-density peaks.
- Data product: galaxy catalogue

Credit: ESA/Webb, NASA & CSA, G. Gozaliasl, A. Koekemoer, M. Franco, and the COSMOS-Web team

Tracers: What we actually observe

Galaxy clusters: Rare high-mass peaks

- Largest gravitationally bound structures.
- Trace rare peaks of δ_m.
- Data product: cluster catalogue

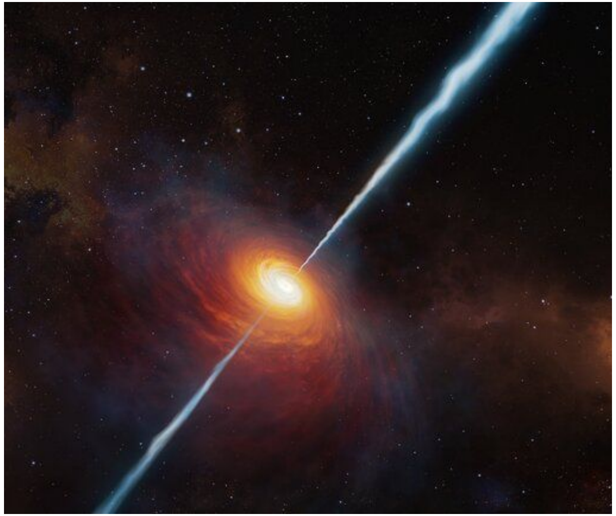


Credit: NASA, ESA/Hubble, HST Frontier Fields

Tracers: What we actually observe

Quasars: Rare, bright beacons in massive environments

- Galaxies powered by accretion onto supermassive black holes
- Very luminous, visible to high redshifts.
- Rare objects → sparse catalogue.
- Strongly biased environments.
- Also act as background lamps to spectroscopically probe the foreground intergalactic gas.

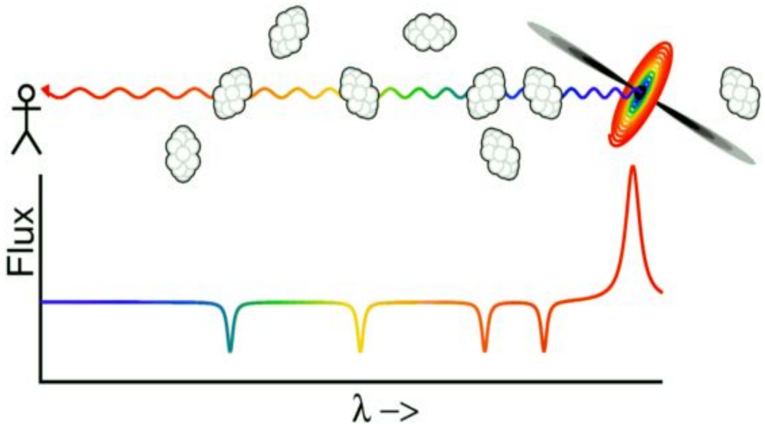


Artist’s impression of quasar P172+18

Tracers: What we actually observe

Lyα forest: 1D sightlines through the cosmic web

- Background quasar provides the light.
- Intervening neutral hydrogen (H I) absorbs at Lyα.
- Each spectrum gives a 1D line-of-sight skewer.
- Sensitive to cosmic mean densities.
- Data product: quasar spectra / transmitted-flux field.

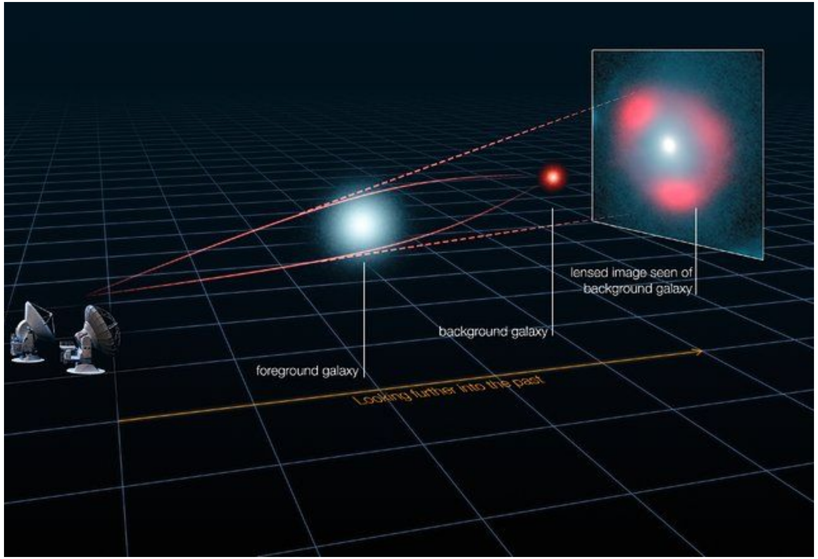


https://www.astro.ucla.edu/~wright/Lyman-alpha-forest.html

Tracers: What we actually observe

Weak lensing: Seeing matter through distorted light

- Gravity distorts background galaxy images.
- Sensitive to total matter, not just luminous matter.
- Measures projected mass along the line of sight.
- Data product: shear catalogue / convergence map.

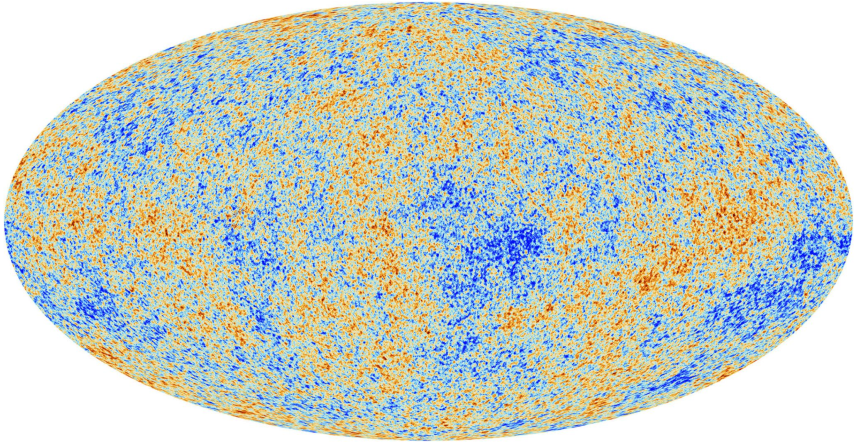


Credit: ALMA (ESO/NRAO/NAOJ), L. Calçada (ESO), Y. Hezaveh et al.

Tracers: What we actually observe

Cosmic Microwave Background: The oldest light and initial condition map

- Relic radiation from the early Universe.
- Emitted when the Universe became neutral and transparent.
- Maps tiny temperature anisotropies on the sky.
- Traces primordial density perturbations.
- Data product: full-sky temperature / polarization maps

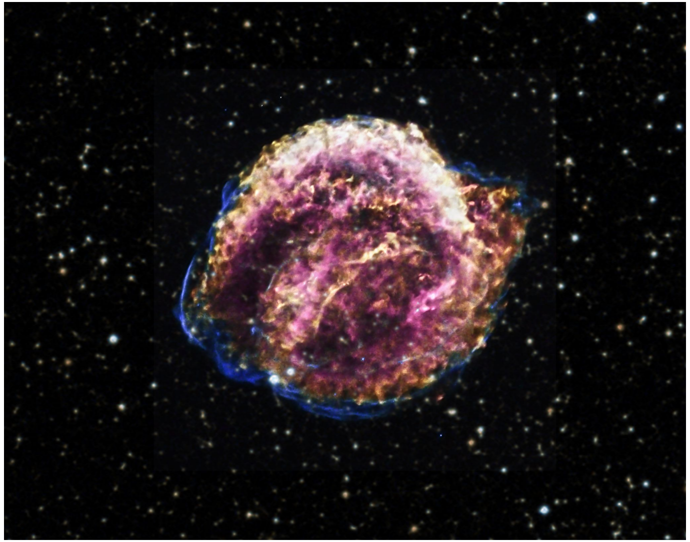


Credit: ESA and the Planck Collaboration

Tracers: What we actually observe

Supernovae: Distance tracers of cosmic expansion

- Standardized candles: Type Ia supernovae have standardizable luminosities
- Observed brightness gives luminosity distance.
- Redshift gives position on cosmic timeline.
- Distance–redshift relation probes expansion history.
- Data product: supernova light curves + redshifts

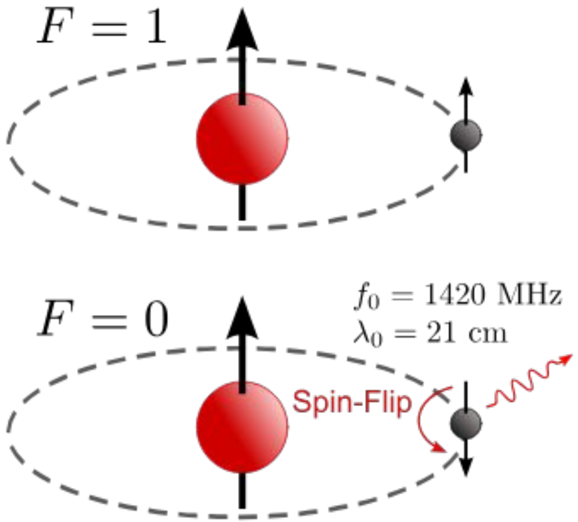


Kepler's supernova remnant, shown here in a combination of X-ray and optical wavelengths
[X-ray: NASA/CXC/SAO/D.Patnaude, Optical: DSS]

Tracers: What we actually observe

21 cm: Mapping neutral hydrogen directly

- Hyperfine transition of neutral hydrogen.
- Traces H I rather than starlight.
- During reionization: sensitive to ionized bubbles.
- Post-reionization: H I intensity mapping traces large-scale structure.
- Data product: radio intensity map / cube.



Cosmological Tracers and Major surveys/telescopes

Tracer (What we observe)	Major Surveys / Telescopes (examples)				
1 **Galaxies** (Positions, spectra, shapes)	SDSS (I–V)	DESI (Dark Energy Spectroscopic Instrument)	Euclid (ESA)	LSST (Vera C. Rubin Observatory)	HSC (Subaru)
2 **Weak Gravitational Lensing** (Galaxy shapes)	KiDS-1000 KiDS (VLT)	DESY (Dark Energy Survey)	HSC (Subaru)	LSST (Vera C. Rubin)	Euclid (ESA)
3 **Cosmic Microwave Background** (CMB temperature & polarization)	Planck (ESA)	ACT (Atacama)	SPT-3G (South Pole)	Simons Observatory	CMB-S4 (Future)
4 **Type Ia Supernovae** (Standard candles)	Pantheon+	DES (SN program)	ZTF (Transient Factory)	LSST (Vera C. Rubin)	Roman (ESA)
5 **Quasars & Lyα Forest** (Spectra)	eBOSS (SDSS)	BOSS (SDSS)	DESI (DESI)	KODIAQ (Keck)	SQUAD (VLT)
6 **Galaxy Clusters** (Positions, masses, X-ray/SZ)	eROSITA (All-sky X-ray)	ACT (SZ clusters)	SPT (SZ clusters)	DES (Optical clusters)	LSST (Future clusters)

Other / Emerging Tracers

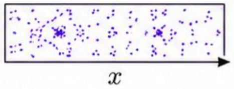
- **Strong Lensing**
  Mass maps, substructure, cosmography
  Surveys: HSC, LSST, Euclid
- **21-cm Intensity Mapping**
  HI distribution over cosmic time
  Surveys: HERA, CHIME, IMPACT, SKA (future)
- **Gravitational Waves** (standard sirens)
  Independent distance measurements
  Surveys: LIGO–Virgo–KAGRA, Einstein Telescope (future)
- **Gamma-ray Bursts**
  High-z standardizable candles
  Surveys: Fermi, Swift, SVOM (future)

Different tracers probe different physics and epoch of the Universe. Together, they can break degeneracies and provide a consistent picture of the Universe.

Tracer statistics: Auto, Cross and Higher-order

1D statistics

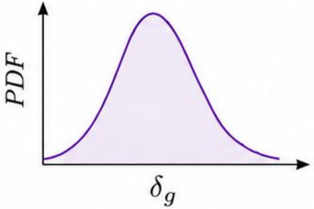
1D statistics compress the field into one-point information and summarize the distribution of tracer values.



Example: PDF of galaxy overdensity / transmitted flux

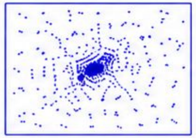
$P(\delta_g)$ or $\mathrm{PDF}(\delta_g)$

- mean: $\langle \delta_g \rangle$



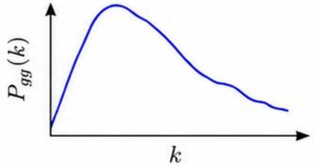
Auto-correlation / auto-power

$$P_{gg}(k) = \langle \delta_g(\mathbf{k})\, \delta_g^*(\mathbf{k}) \rangle$$



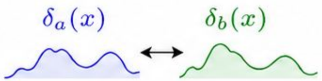
Measures how strongly the tracer clusters with itself.

Example: galaxy × galaxy



Cross-correlation / cross-power

$$P_{ab}(k) = \langle \delta_a(\mathbf{k})\, \delta_b^*(\mathbf{k}) \rangle$$



Examples:

$$r_{ab}(k) = \frac{P_{ab}(k)}{\sqrt{P_{aa}(k) P_{bb}(k)}}$$

- galaxy × weak lensing
- Lyα forest × quasars
- LAEs × 21 cm

$r = 1$: perfectly correlated fields.

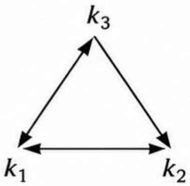
$r < 1$: stochasticity, noise, nonlinear physics, or different astrophysical response.

Higher-order tracer statistics

Two-point statistics measure variance, but nonlinear structure contains more information.

$$\langle \delta(\mathbf{k}_1)\, \delta(\mathbf{k}_2)\, \delta(\mathbf{k}_3) \rangle = (2\pi)^3 \delta_D(k_1 + k_2 + k_3)\, B(k_1, k_2, k_3)$$

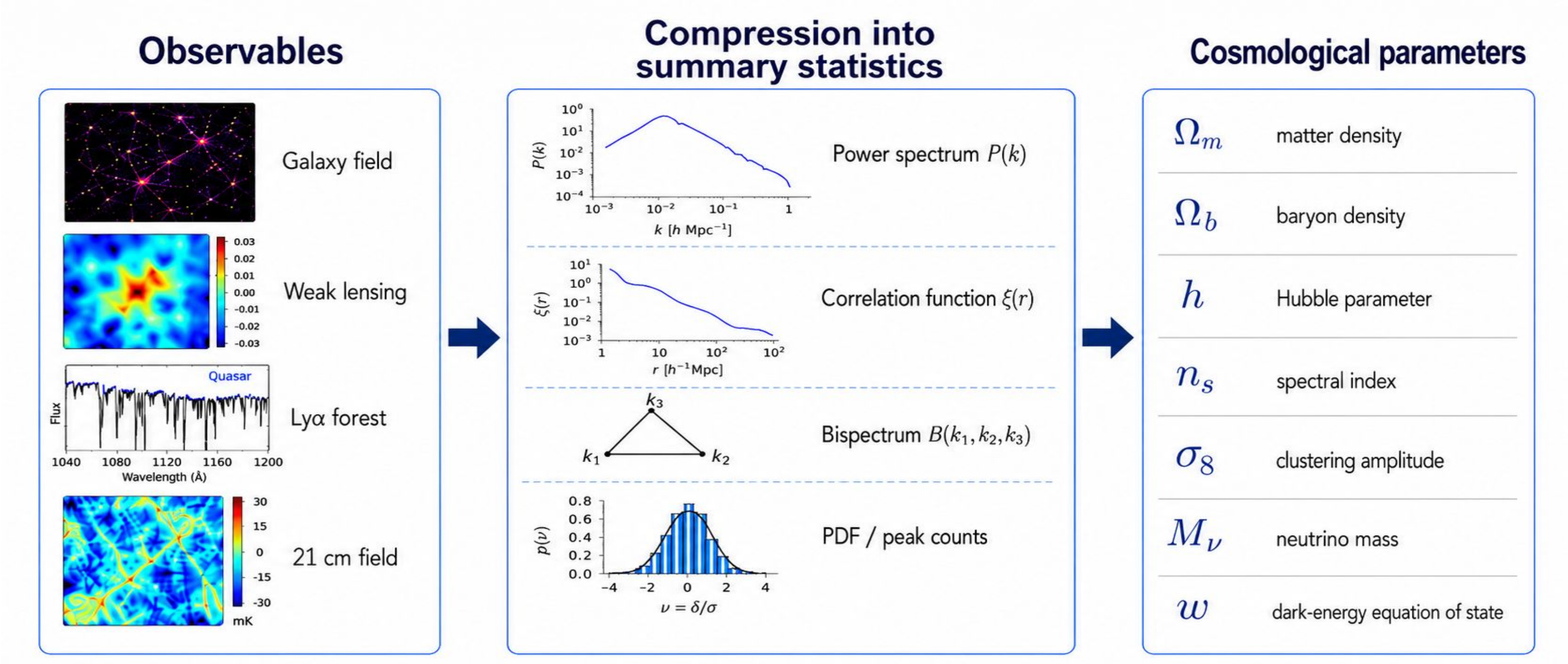
$$k_1 + k_2 + k_3 = 0$$



Higher-order statistics capture:

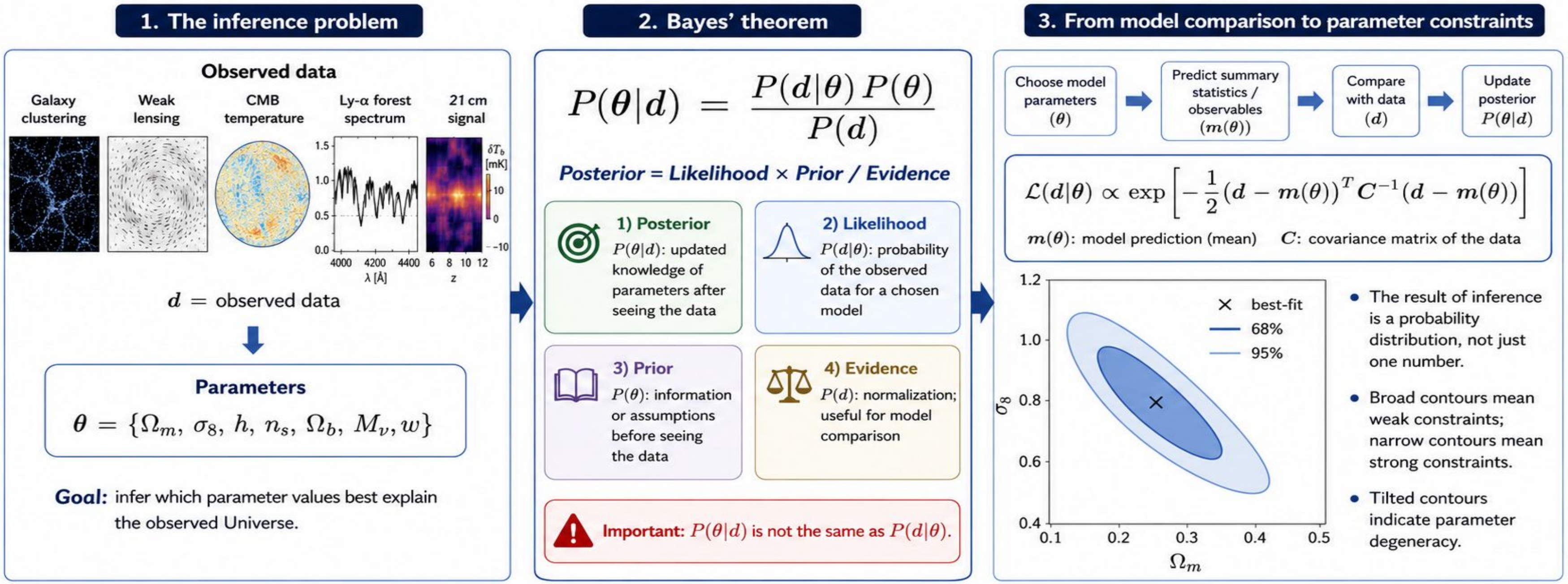
- nonlinear gravitational evolution
- nonlinear bias
- phase correlations
- filamentarity
- bubble morphology
- reionization topology
- feedback-induced non-Gaussianity

Inferring cosmological parameters from tracers



We compress the tracer data into summary statistics and ask which cosmological model most likely produced them.

Bayesian Parameter Inference in Cosmology



Bayesian inference turns cosmological observations into probabilistic constraints on physical models.

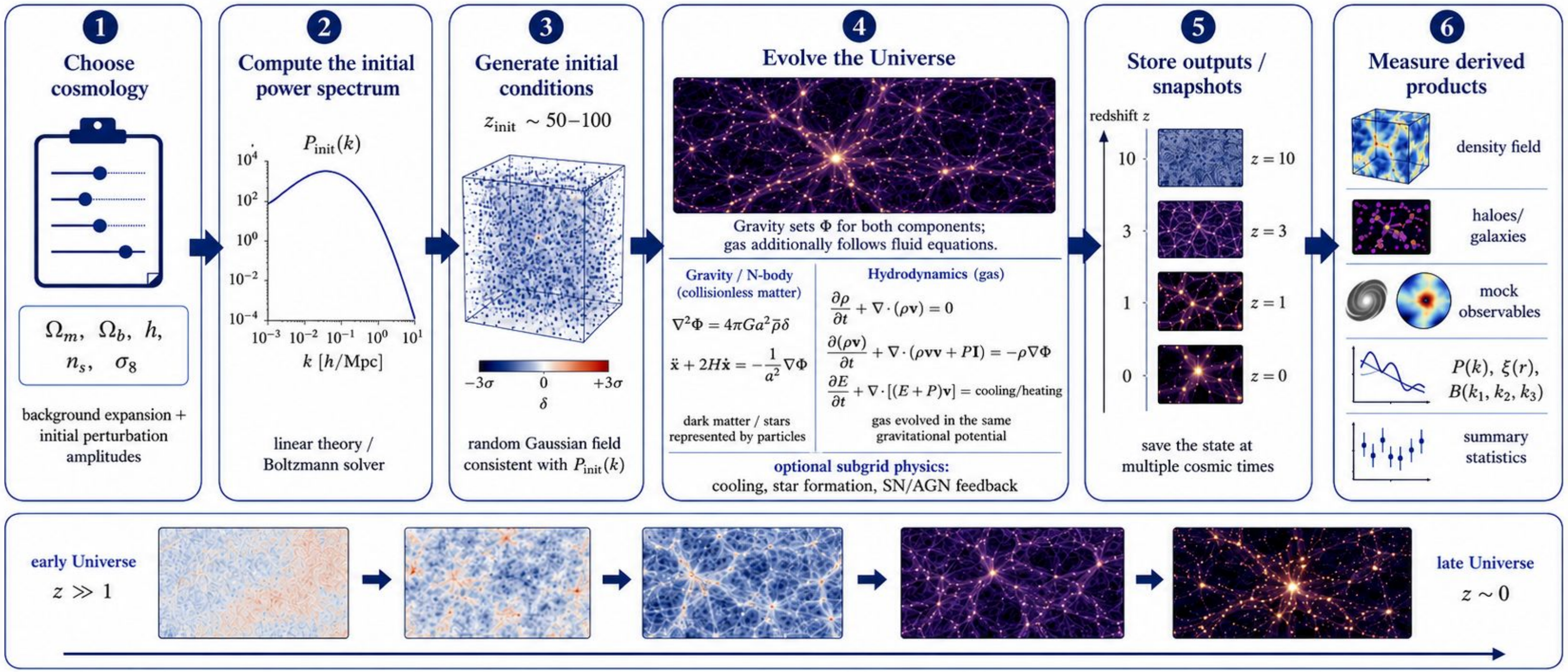
Inference requires a family of models

- To infer parameters, we need predictions for many possible choices of cosmology and astrophysics. The observed Universe is then compared against this ensemble.
- In cosmology, the models usually come from simulations.
- A simulation takes model parameters and forward models a mock instance of the Universe.

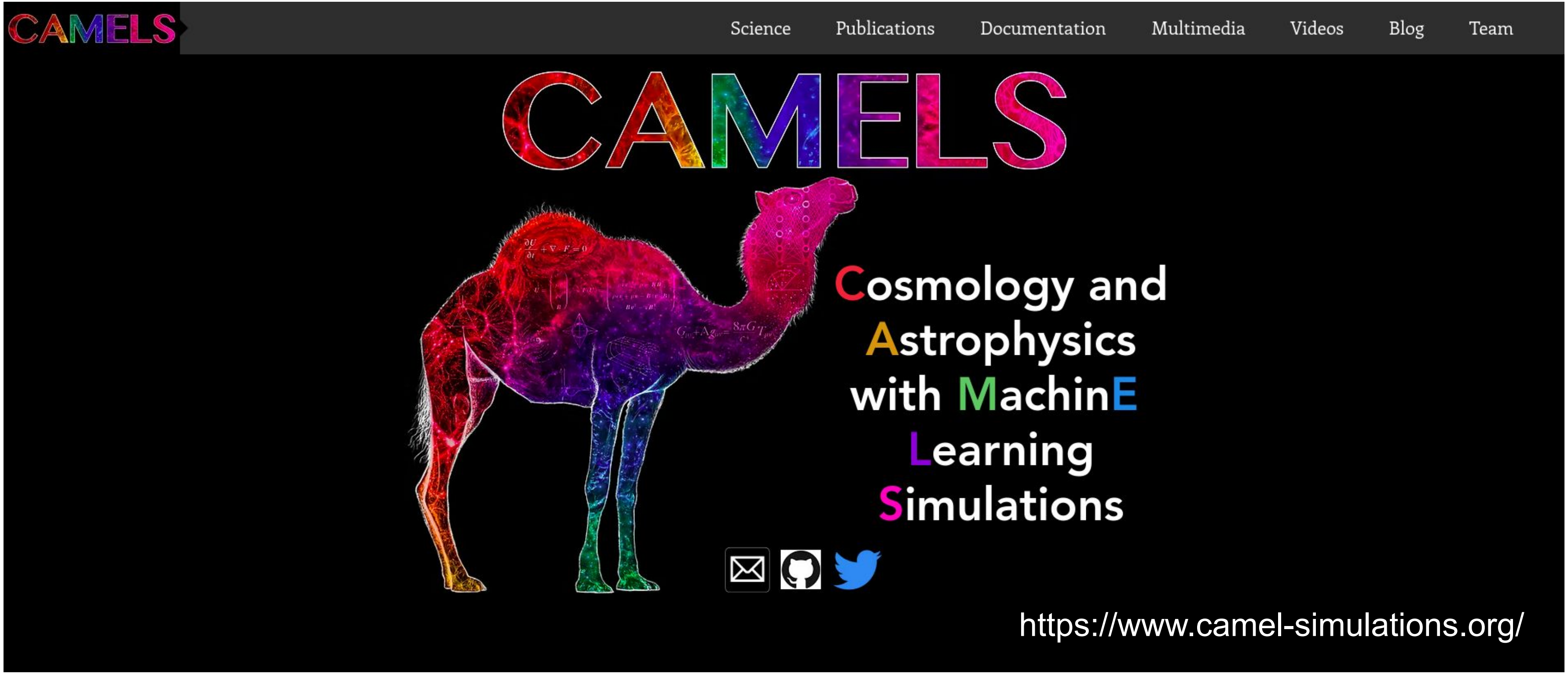
$$\boxed{\theta \rightarrow \text{simulation} \rightarrow \text{mock Universe} \rightarrow \text{summary statistics}}$$

Simulation type	What it models	Used for
N-body	collisionless dark matter	matter clustering, halo formation
hydrodynamical	dark matter + gas + stars + black holes	galaxies, baryons, feedback
radiative transfer	propagation of radiation	reionization, Lyα, 21 cm
mock survey pipelines	observational selection/noise	realistic comparison to data

How cosmological simulations are run



CAMELS: A public Simulation Suite for inference



CAMELS: A public Simulation Suite for inference

General description

As of March 2025, CAMELS host more than 2 petabytes of data from 16,960 cosmological simulations: 7,208 N-body and 9,752 hydrodynamic simulations. This project is the result of a large, collaborative effort and represents the most extensive suite of cosmological hydrodynamic simulations ever conducted.

The simulations are run with different codes (suites), cover different volumes and number of particles (generations), and sample different points in parameter space (sets), defined as the space formed from cosmological and astrophysical parameters and different initial conditions.

All CAMELS simulations share a few characteristics:

- All simulations start at redshift $z = 127$ and finish at $z = 0$.
- All simulations have the same mass and spatial resolution.
- For each simulation we store multiple snapshots, halo/subhalo catalogs, and summary statistics (e.g. power spectra) see Data organization.
- The N-body simulations only model the spatial phase-space distribution of total matter under the influence of gravity.
- The hydrodynamic simulations account for gravity + (magento-)hydrodynamics + astrophysical processes (e.g. supernova and AGN feedback).

Organization

The simulations can be classified into:

- **Suites**: depending on the code and subgrid physics model used to run them. CAMELS has 9 suites:

Suite	Code	Subgrid Physics Model
IllustrisTNG	AREPO	IllustrisTNG
SIMBA	GIZMO	SIMBA
Astrid	MP-Gadget	Astrid
Magneticum	OpenGadget	Magneticum-like
Swift-EAGLE	Swift	EAGLE
Ramses	Ramses	–
Enzo	Enzo	–
CROCODILE	Gadget4-Osaka	CROCODILE
Obsidian	GIZMO	Obsidian

CAMELS: A public Simulation Suite for inference

Cosmological parameters

All simulations share the value of these cosmological parameters:

w	M_ν	Ω_k
-1	0.0 eV	0.0

For the other cosmological parameters, the different sets vary them differently:

	$\Omega_{\rm m}$	σ_8	$\Omega_{\rm b}$	h	n_s
CV	0.3	0.8	0.049	0.6711	0.9624
BE	0.3	0.8	0.049	0.6711	0.9624
EX	0.3	0.8	0.049	0.6711	0.9624
LH	0.1 - 0.5	0.6 - 1.0	0.049	0.6711	0.9624
1P	0.1 - 0.5	0.6 - 1.0	0.029 - 0.069	0.4711 - 0.8711	0.7624 - 1.1624
SB	0.1 - 0.5	0.6 - 1.0	0.029 - 0.069	0.4711 - 0.8711	0.7624 - 1.1624

Astrophysical parameters

We emphasize that every subgrid physics model is different, and the parameters of one model does not mean anything in another one. Thus, we will describe these parameters for each suite and what is varied.

IllustrisTNG

The IllustrisTNG suite contains all sets: 1P, CV, LH, EX, BE, and SB. This table shows which parameters are varied in each set:

	Astrophysical parameters	
	First generation	**Second generation**
CV	*fiducial* IllustrisTNG	
BE	*fiducial* IllustrisTNG	
LH	*standard* 4 astrophysical parameters	—
EX	*standard* 4 astrophysical parameters	
1P	23 astrophysical parameters	30 astrophysical parameters
SB	SB28: 23 astrophysical parameters	SB35: 30 astrophysical parameters

The meaning and range of variation of the 4 *standard* IllustrisTNG parameters are these:

- $A_{\rm SN1}$: it represents the energy per unit SFR of the galactic winds. It can vary from 0.25 to 4. Fiducial value is 1.
- $A_{\rm SN2}$: it represents the wind speed of the galactic winds. It can vary from 0.5 to 2. Fiducial value is 1.
- $A_{\rm AGN1}$: it represents the energy per unit blach-hole accretion rate. It can vary from 0.25 to 4. Fiducial value is 1.
- $A_{\rm AGN2}$: it represents the ejection speed/burstiness of the kinetic mode of the black-hole feedback. It can vary from 0.5 to 2. Fiducial value is 1.

CAMELS: A public Simulation Suite for inference

Data Access

Binder	Globus	URL	FlatHUB	Rusty
Binder link	Globus link	URL link	FlatHUB link	`/mnt/ceph/users/camels/PUBLIC_RELEASE`

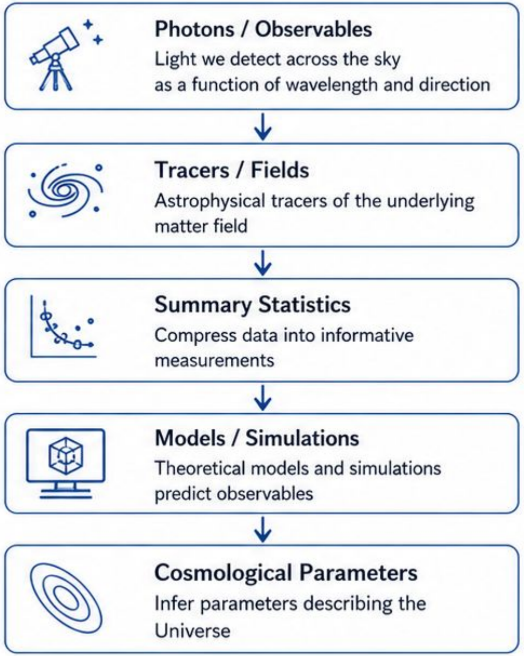
This table shows the availability of the different simulations:

	First generation	Second generation	Third generation
IllustrisTNG	Public	Private	
SIMBA	Public		
Astrid	Public	Private (running)	
Swift-EAGLE	Public		
Magneticum	Private		
Ramses	Private	Private (running)	
CROCODILE	Private	Private (running)	
Obsidian	Private		
Enzo	Private		
N-body	Public	Public	

Take-home messages: From light to cosmology

Cosmology as data-driven inference

1. We observe the Universe on our past light cone, not as a single snapshot in time.
2. Redshift is both an observational coordinate and a marker of cosmic time.
3. The expanding background Universe is described by $a(t)$, $H(z)$, and parameters such as $\Omega_m, \Omega_\Lambda, H_0, \sigma_8, n_s$.
4. Structure formation is described by the growth of density perturbations: $\delta_m(\mathbf{x}, t) = [\rho_m(\mathbf{x}, t) - \bar{\rho}_m(t)] / \bar{\rho}_m(t)$.
5. We do not observe the matter field directly; we observe tracers: galaxies, quasars, Lyα forest, weak lensing, CMB, supernovae, and 21 cm.
6. Cosmological information is extracted using summary statistics such as $P(k)$, $\xi(r)$, $B(k_1, k_2, k_3)$, PDFs, and cross-correlations.



Cosmological data handling turns observed photons into quantitative statements about the composition, expansion history, and structure of the Universe.

References:

- **T. Padmanabhan** — *Structure Formation in the Universe*
- **Scott Dodelson & Fabian Schmidt** — *Modern Cosmology*
- **P. J. E. Peebles —** *Principles of Physical Cosmology* or *The Large-Scale Structure of the Universe*

Lyman-α Forest as a 1D Probe of the Cosmic Structures

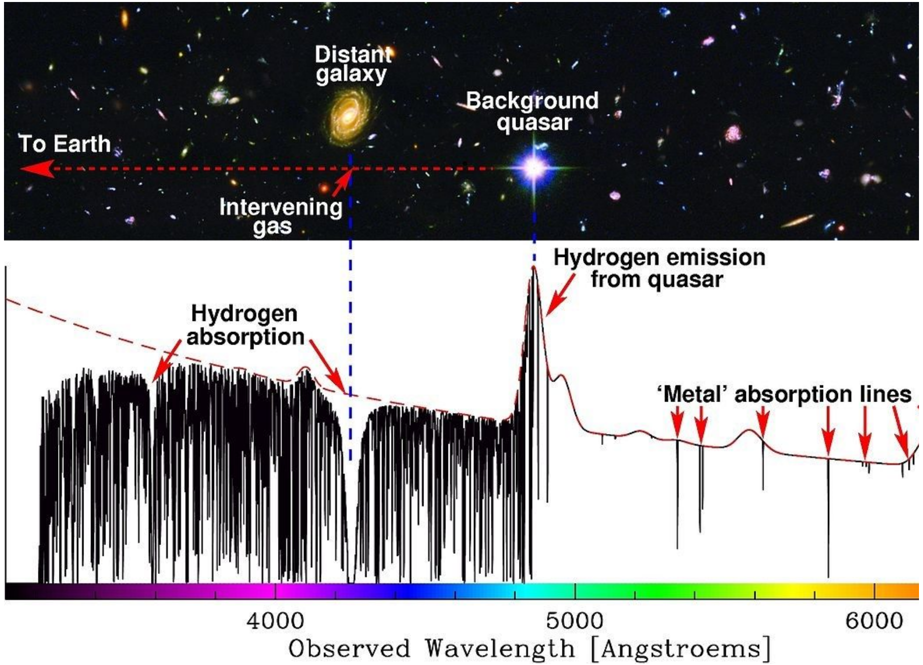
- Lyα forest is a series of absorption produced by neutral hydrogen (HI) due to the 1s→2p transition in the Quasar spectra.
- These absorptions get redshifted due to the expansion of the Universe.

$$E_\alpha = 10.2\,\mathrm{eV},\ \lambda_\alpha = 1215.67\,\mathrm{\AA}$$

$$\lambda_{\mathrm{obs}} = \lambda_\alpha(1 + z_{\mathrm{abs}})$$

- This makes them a 1D map of the intervening HI gas present in the sightline of the distant Quasar.
- Lyman-α forest probes matter in a quasi-linear regime.
- $F_{obs} = F_{cont}\, e^{-\tau_{HI}}$ [Transmitted Flux: F_{obs}/F_{cont}]

where F_{obs} is the observed Flux and F_{cont} is the continuum flux emitted by the quasar before absorption.

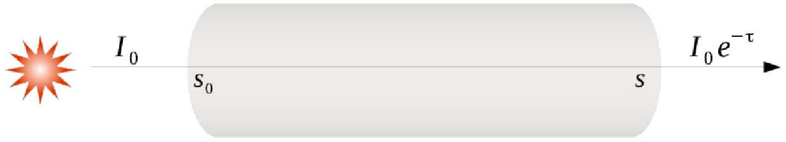


Lyman-α Radiative transfer

- The basic radiative transfer equation for absorption is $I_\nu = I_{\nu,0} e^{-\tau_\nu}$.

- For a particular frequency ν, the optical depth over a distance s_o to s is given by

$$\tau_\nu = \int_{s_0}^{s} ds' n(s', t') \sigma_{\nu'}$$



where, n : Number density of the absorbers

σ_ν : Frequency dependent absorption cross-section.

- σ_ν for the absorbers at rest is given by a Lorentzian profile as

$$\sigma_\nu = \left(\frac{\pi e^2}{m_e c}\right) \left(\frac{f_{lu}}{4\pi\epsilon_0}\right) \frac{\Gamma_{ul}/4\pi^2}{(\nu - \nu_{lu})^2 + (\Gamma_{ul}/4\pi)^2}.$$

- Here, f_{lu} is the upward oscillator strength, Γ_{ul} is the damping width of the upper level and ν_{lu} is the resonance line frequency.

- For hydrogen Lyman-α, f_{lu}=0.4162, λ_{lu} = 1215.67 A , and Γ_{ul} = $6.262 \times 108 s^{-1}$.

Lyman-α Radiative transfer

- The absorbers are rarely at rest.
  Ignoring non-thermal velocities, the actual absorption cross-section is obtained by convolving the Lorentzian with a Maxwellian profile.

$$\sigma_\nu = \left(\frac{\pi e^2}{m_e c}\right)\left(\frac{f_{lu}}{4\pi\epsilon_0}\right)\boxed{\varphi_V(\alpha,\nu)}\ \textbf{Voigt Profile}$$

$$\varphi_V(a,x) = \frac{1}{\pi^{1/2}\Delta\nu_D}V(a,x) = \frac{a}{\pi^{3/2}\Delta\nu_D}\int_{-\infty}^{\infty}\frac{e^{-y^2}}{(x-y)^2+a^2}$$

  where, $\Delta v_D = v_{lu} b/c$, with b being the doppler parameter ($b = \sqrt{(2k_B T/m)}$), $a = \Gamma_{ul}/(4\pi\Delta v_D)$ and $x = (v - v_{lu})/\Delta v_D$.

- The Lyman-α absorption optical depth at redshift z_0 will then be given as

$$\tau(z_0) = \frac{cI_\alpha}{\pi^{1/2}}\int ds\left(\frac{n_{HI}[s,z(s)]}{b[s,z(s)][1+z(s)]}\right)V\left[a, \frac{cz(s)-z_0}{b[s,z(s)][1+z_0]} + \frac{v[s,z(s)]}{b[s,z(s)]}\right]$$

  where n_{HI} is the neutral hydrogen density and $I_\alpha = 4.45 \times 10^{-18} cm^2$.

Lyman-α Radiative transfer

- The absorbers are rarely at rest.
  Ignoring non-thermal velocities, the actual absorption cross-section is obtained by convolving the Lorentzian with a Maxwellian profile.

$$\sigma_\nu = \left(\frac{\pi e^2}{m_e c}\right)\left(\frac{f_{lu}}{4\pi\epsilon_0}\right)\boxed{\varphi_V(\alpha,\nu)}$$ **Voigt Profile**

$$\varphi_V(a,x) = \frac{1}{\pi^{1/2}\Delta\nu_D}V(a,x) = \frac{a}{\pi^{3/2}\Delta\nu_D}\int_{-\infty}^{\infty}\frac{e^{-y^2}}{(x-y)^2+a^2}$$

  where, $\Delta v_D = v_{lu} b/c$, with b being the doppler parameter ($b = \sqrt{(2k_B T/m)}$), $a = \Gamma_{ul}/(4\pi\Delta v_D)$ and $x = (v - v_{lu})/\Delta v_D$.

- The Lyman-α absorption optical depth at redshift z_0 will then be given as

$$\tau(z_0) = \frac{cI_\alpha}{\pi^{1/2}}\int ds\left(\frac{\boxed{n_{HI}}[s,z(s)]}{\boxed{b}[s,z(s)][1+z(s)]}\right)V\left[\boxed{a}, \frac{cz(s)-z_0}{\boxed{b}[s,z(s)][1+z_0]}+\frac{\boxed{v}[s,z(s)]}{\boxed{b}[s,z(s)]}\right]$$

  where n_{HI} is the neutral hydrogen density and $I_\alpha = 4.45 \times 10^{-18} cm^2$.

Understanding the Voigt profile

Important parameters:

- nHI: Neutral hydrogen density
- T: IGM temperature
- v: peculiar velocity, i.e, velocity on top of the Hubble expansion.

Under Photoionization equilibrium assumption, $n_{\mathrm{HI}}\Gamma_{\mathrm{HI}} = n_e n_p \alpha_A(T)$

Ionization Recombination

For highly ionized hydrogen ($n_e \approx n_p \approx n_H$), $n_{\mathrm{HI}} \simeq \dfrac{\alpha_A(T) n_H^2}{\Gamma_{\mathrm{HI}}}$

Recombination coefficient roughly scales as $\alpha_A(T) \propto T^{-0.7}$.

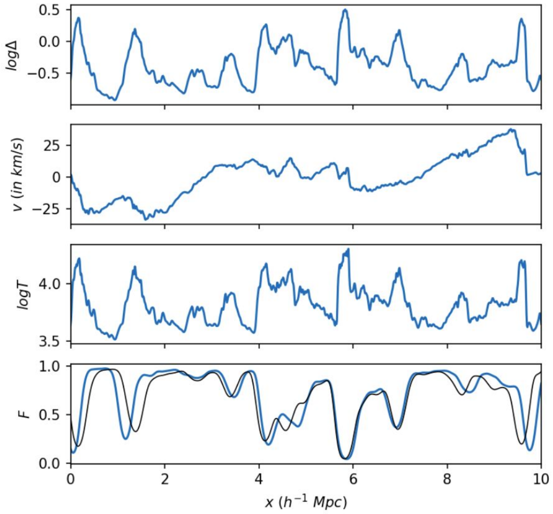
Hence, $n_{\mathrm{HI}} \propto \dfrac{n_H^2 T^{-0.7}}{\Gamma_{\mathrm{HI}}}$, where nH is the hydrogen density that is proportional to the matter density $\Delta = \varrho / \langle\varrho\rangle$ and

Γ_{HI} is the hydrogen photoionization rate.

So the final important parameters governing Lyman-alpha forest Voigt profile are Δ (Density), Γ_{HI} (photoionization rate), T (IGM temperature) and v (peculiar velocity)

Understanding the Voigt profile

- Effect of Δ (Density): Increases n_{HI}, stronger absorption.
- Effect of Γ_{HI} (photoionization rate): Higher Γ_{HI} decreases n_{HI}, weaker absorption.
- Effect of T (IGM temperature):
  - Higher temperature decreases recombination rate, lower n_{HI}.
  - Higher temperature increase b-parameter, wider absorption lines (**Thermal broadening**)
- Effect of peculiar velocity:



Fluctuating Gunn Peterson Approximation:
Approximate dependence on cosmological and astrophysical parameters

Main approximations:

- Highly ionized IGM: $n_e \simeq n_p \simeq n_H$
- Photoionization equilibrium, temperature dependent recombination (previous slide)
- Power law temperature-density relation: $T = T_0(1+\delta_b)^{\gamma-1}$
- Local optical depth approximation: $\tau(\mathbf{x}) \propto n_{\rm HI}(\mathbf{x})$

$$\tau(\Delta, z) \propto \frac{(\Omega_b h^2)^2 (1+z)^{4.5}}{h\,\Omega_m^{1/2}\,\Gamma_{\rm HI}} T_0^{-0.7} \Delta^{2-0.7(\gamma-1)}$$

$$F(\Delta, z) = \exp[-\tau(\Delta, z)]$$

Fluctuating Gunn Peterson Approximation:
Approximate dependence on cosmological and astrophysical parameters

Main approximations:

- Highly ionized IGM: $n_e \approx n_p \approx n_H$
- Photoionization equilibrium, temperature dependent recombination (previous slide)
- Power law temperature-density relation: $T = T_0(1+\delta_b)^{\gamma-1}$
- Local optical depth approximation: $\tau(\mathbf{x}) \propto n_{\mathrm{HI}}(\mathbf{x})$

$$\tau(\Delta, z) \propto \frac{(\Omega_b h^2)^2 (1+z)^{4.5}}{h\,\Omega_m^{1/2}\,\Gamma_{\mathrm{HI}}} T_0^{-0.7} \Delta^{2-0.7(\gamma-1)}$$

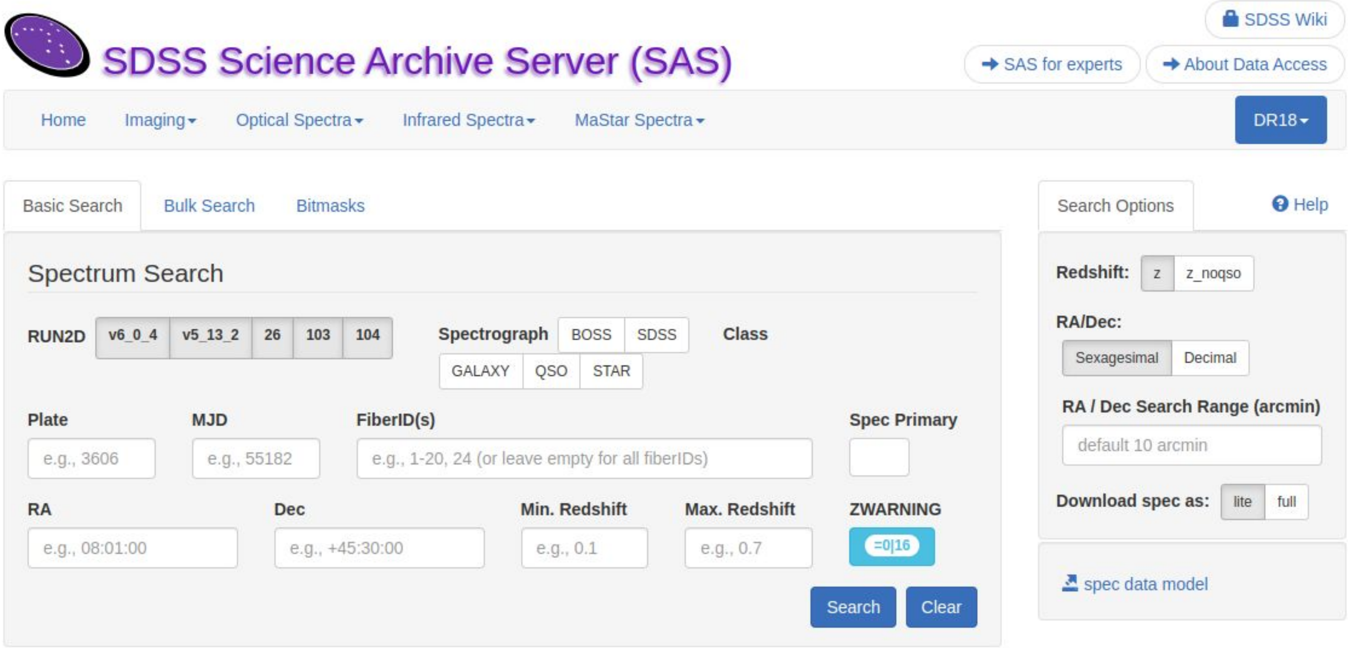
$$F(\Delta, z) = \exp[-\tau(\Delta, z)]$$

References:

- **Avery A. Meiksin**, “The physics of the intergalactic medium,” *Reviews of Modern Physics*, 81, 1405–1469, 2009.
- **Matthew McQuinn**, “The Evolution of the Intergalactic Medium,” *Annual Review of Astronomy and Astrophysics*, 54, 313–362, 2016.

Now let’s look at some realistic Lyman-α forest spectra

SDSS: Low resolution spectra (R~2000), large in number



https://dr18.sdss.org/optical/spectrum/search

Now let’s look at some realistic Lyman-α forest spectra

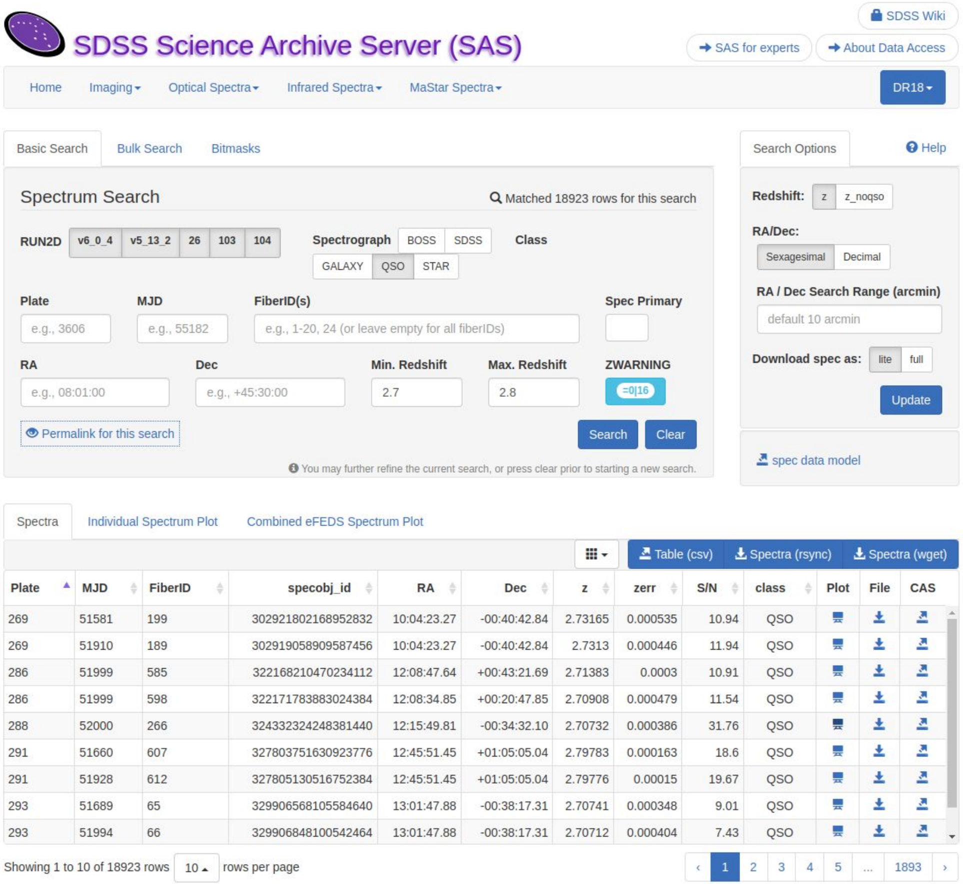
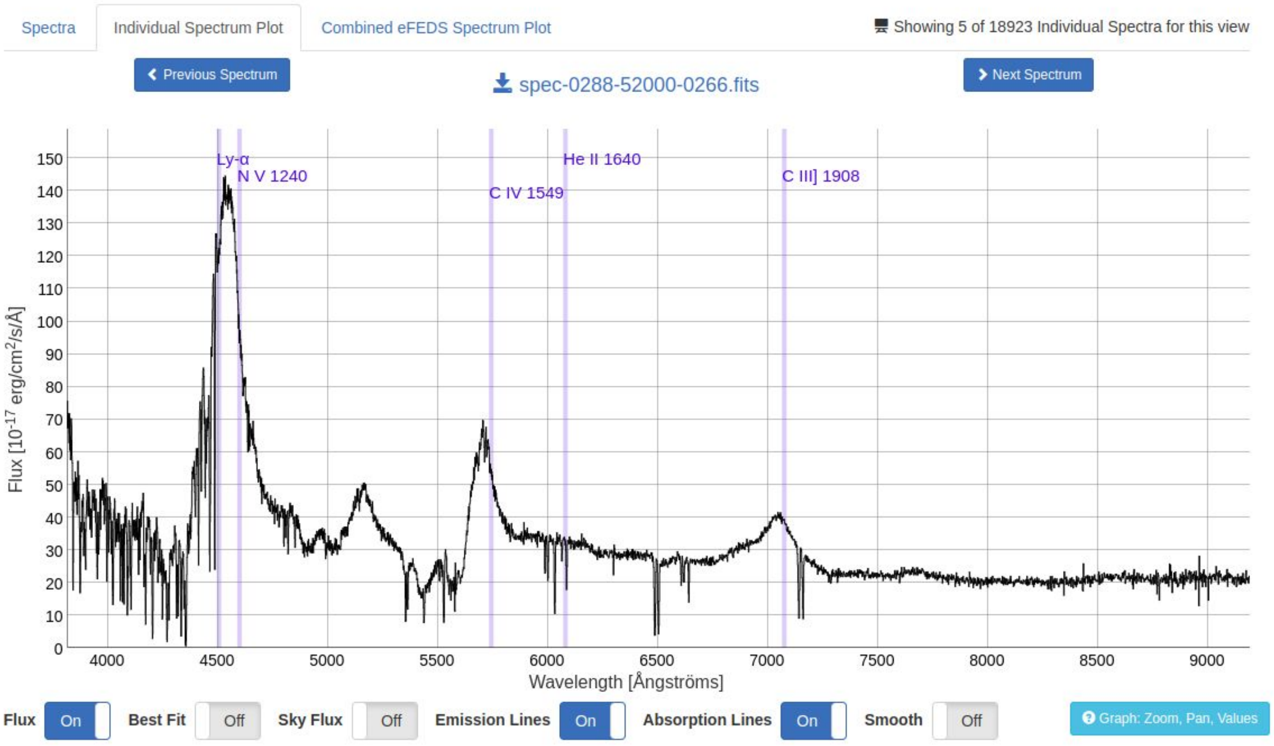
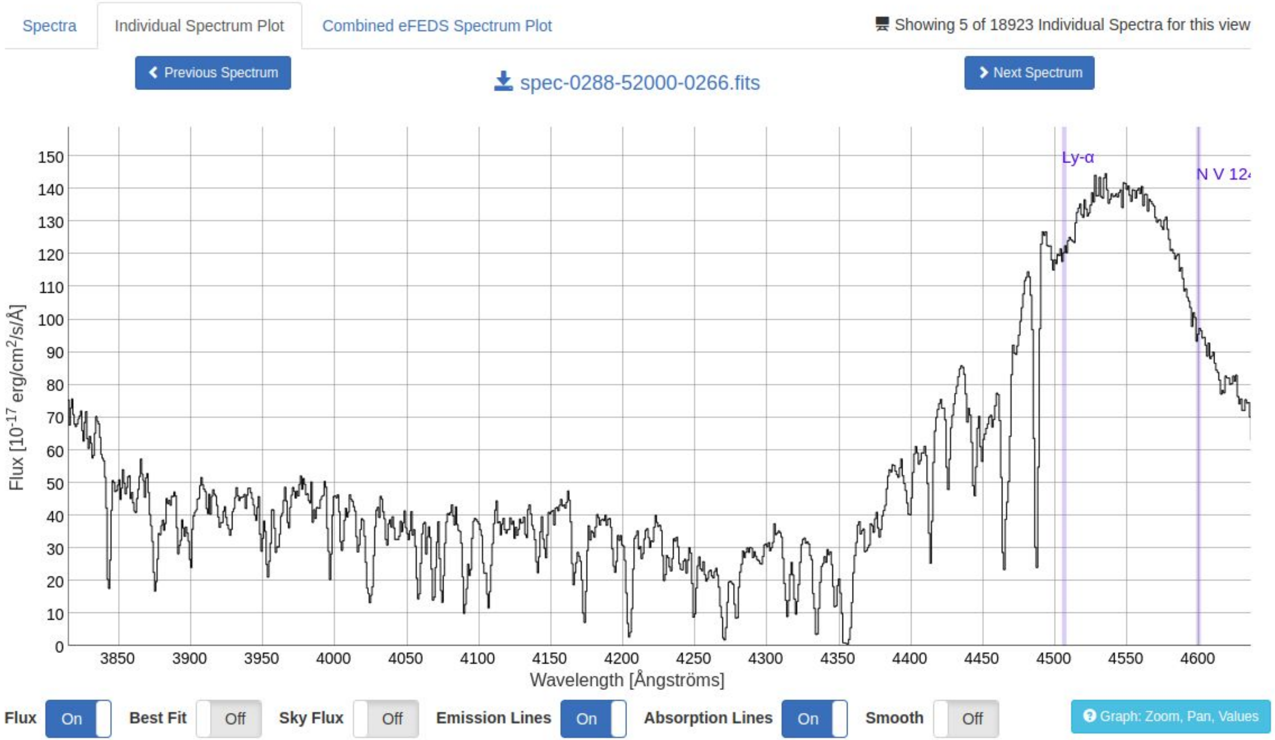


Plate	MJD	FiberID	specobj_id	RA	Dec	z	zerr	S/N	class	Plot	File	CAS
269	51581	199	302921802168952832	10:04:23.27	-00:40:42.84	2.73165	0.000535	10.94	QSO			
269	51910	189	302919058909587456	10:04:23.27	-00:40:42.84	2.7313	0.000446	11.94	QSO			
286	51999	585	322168210470234112	12:08:47.64	+00:43:21.69	2.71383	0.0003	10.91	QSO			
286	51999	598	322171783883024384	12:08:34.85	+00:20:47.85	2.70908	0.000479	11.54	QSO			
288	52000	266	324332324248381440	12:15:49.81	-00:34:32.10	2.70732	0.000386	31.76	QSO			
291	51660	607	327803751630923776	12:45:51.45	+01:05:05.04	2.79783	0.000163	18.6	QSO			
291	51928	612	327805130516752384	12:45:51.45	+01:05:05.04	2.79776	0.00015	19.67	QSO			
293	51689	65	329906568105584640	13:01:47.88	-00:38:17.31	2.70741	0.000348	9.01	QSO			
293	51994	66	329906848100542464	13:01:47.88	-00:38:17.31	2.70712	0.000404	7.43	QSO			

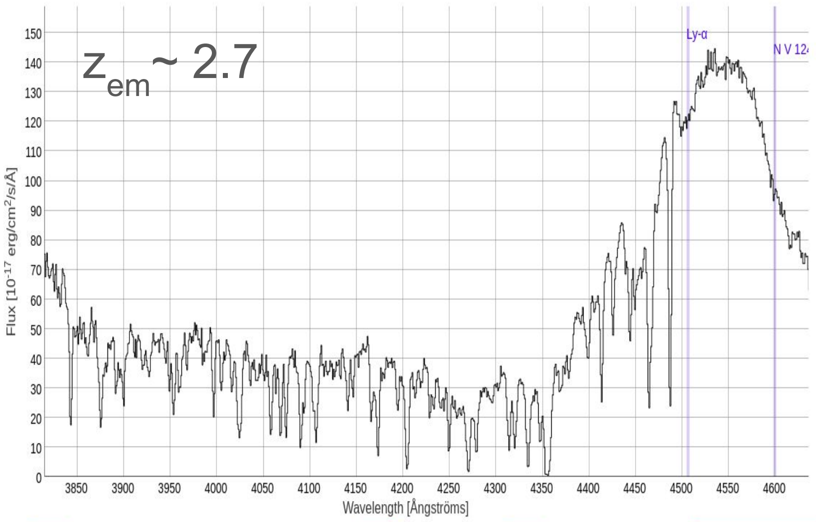
Now let's look at some realistic Lyman-α forest spectra

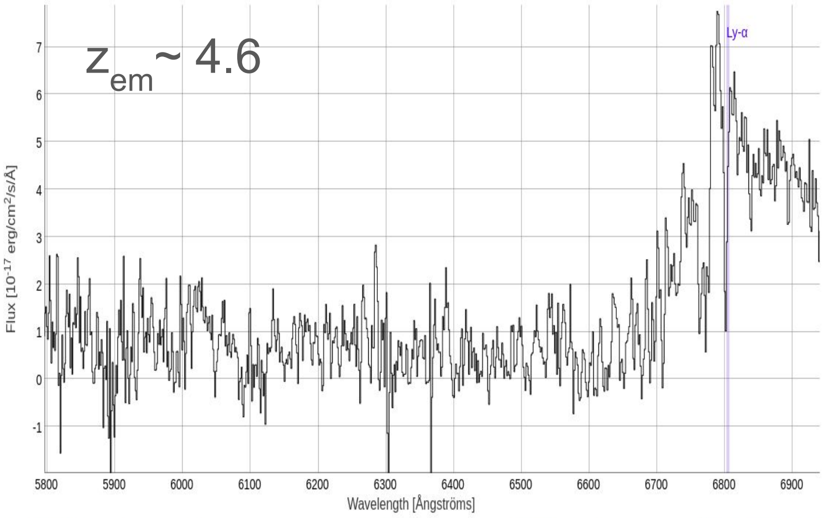


Now let’s look at some realistic Lyman-α forest spectra



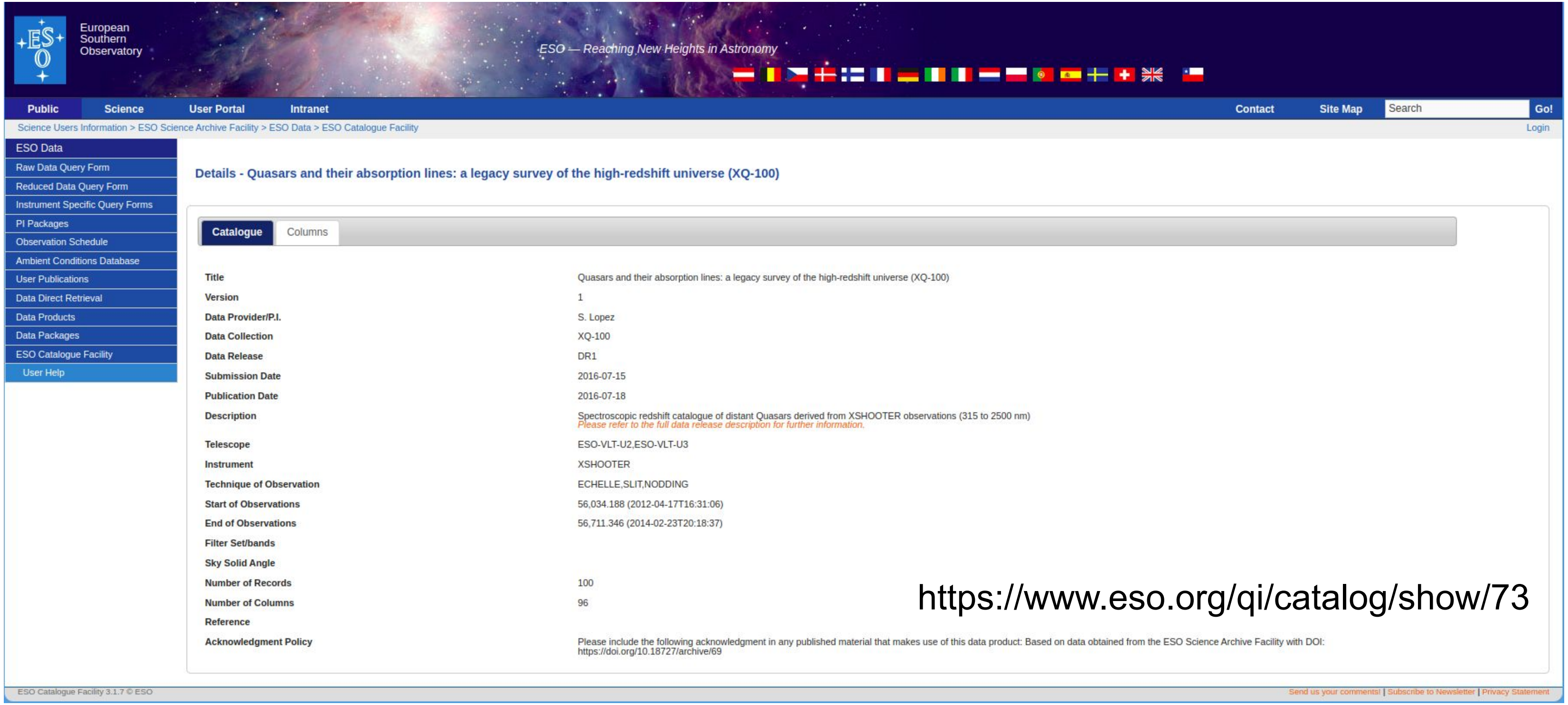
Now let's look at some realistic Lyman-α forest spectra





Now let's look at some realistic Lyman-α forest spectra

XQ100: Medium resolution (R=6000-9000), VLT/X-shooter spectrograph


European Southern Observatory

ESO — Reaching New Heights in Astronomy

Public | Science | User Portal | Intranet | Contact | Site Map | Search | Go!

Science Users Information > ESO Science Archive Facility > ESO Data > ESO Catalogue Facility | Login

- ESO Data
- Raw Data Query Form
- Reduced Data Query Form
- Instrument Specific Query Forms
- PI Packages
- Observation Schedule
- Ambient Conditions Database
- User Publications
- Data Direct Retrieval
- Data Products
- Data Packages
- ESO Catalogue Facility
- User Help

Details - Quasars and their absorption lines: a legacy survey of the high-redshift universe (XQ-100)

Catalogue | Columns

Title	Quasars and their absorption lines: a legacy survey of the high-redshift universe (XQ-100)
Version	1
Data Provider/P.I.	S. Lopez
Data Collection	XQ-100
Data Release	DR1
Submission Date	2016-07-15
Publication Date	2016-07-18
Description	Spectroscopic redshift catalogue of distant Quasars derived from XSHOOTER observations (315 to 2500 nm) *Please refer to the full data release description for further information.*
Telescope	ESO-VLT-U2,ESO-VLT-U3
Instrument	XSHOOTER
Technique of Observation	ECHELLE,SLIT,NODDING
Start of Observations	56,034.188 (2012-04-17T16:31:06)
End of Observations	56,711.346 (2014-02-23T20:18:37)
Filter Set/bands	
Sky Solid Angle	
Number of Records	100
Number of Columns	96
Reference	
Acknowledgment Policy	Please include the following acknowledgment in any published material that makes use of this data product: Based on data obtained from the ESO Science Archive Facility with DOI: https://doi.org/10.18727/archive/69

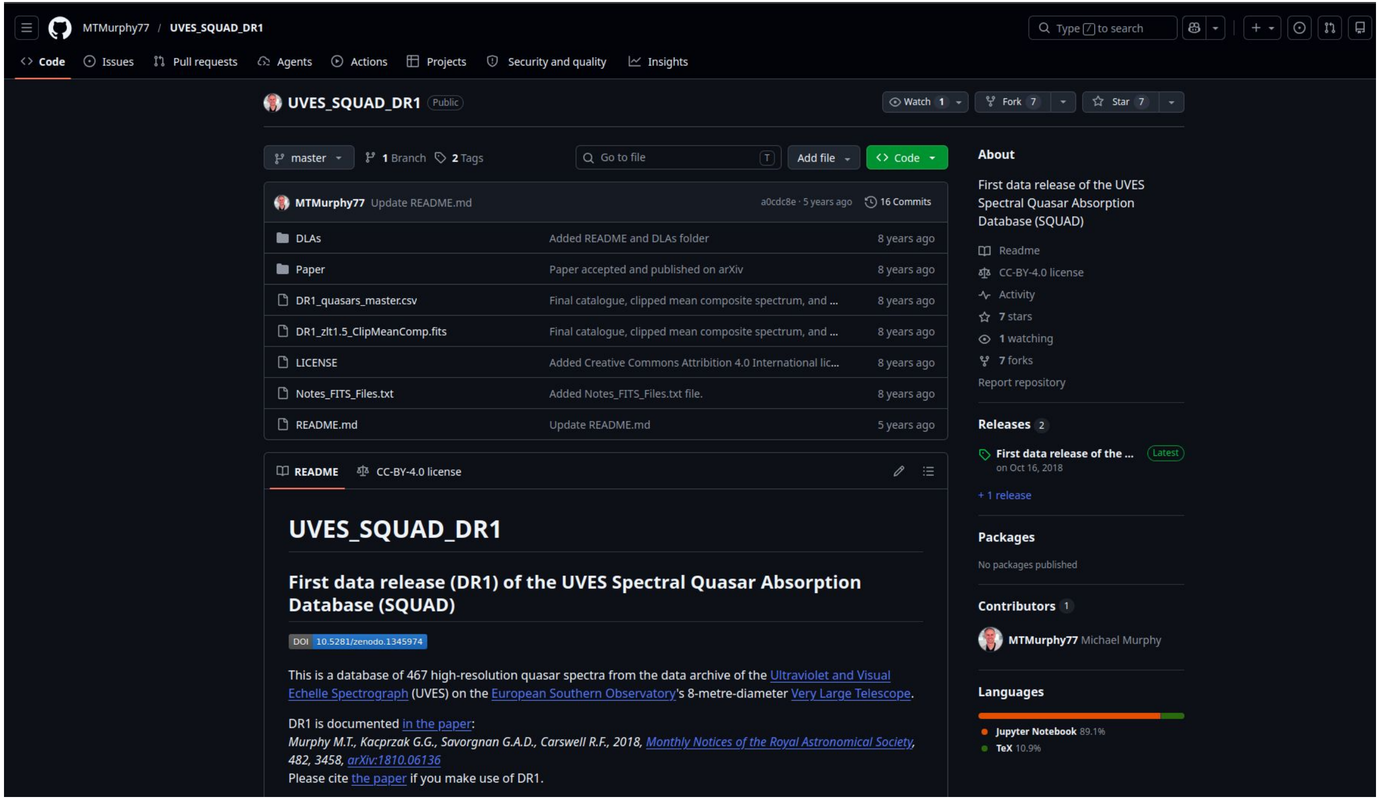
ESO Catalogue Facility 3.1.7 © ESO

Send us your comments! | Subscribe to Newsletter | Privacy Statement

https://www.eso.org/qi/catalog/show/73

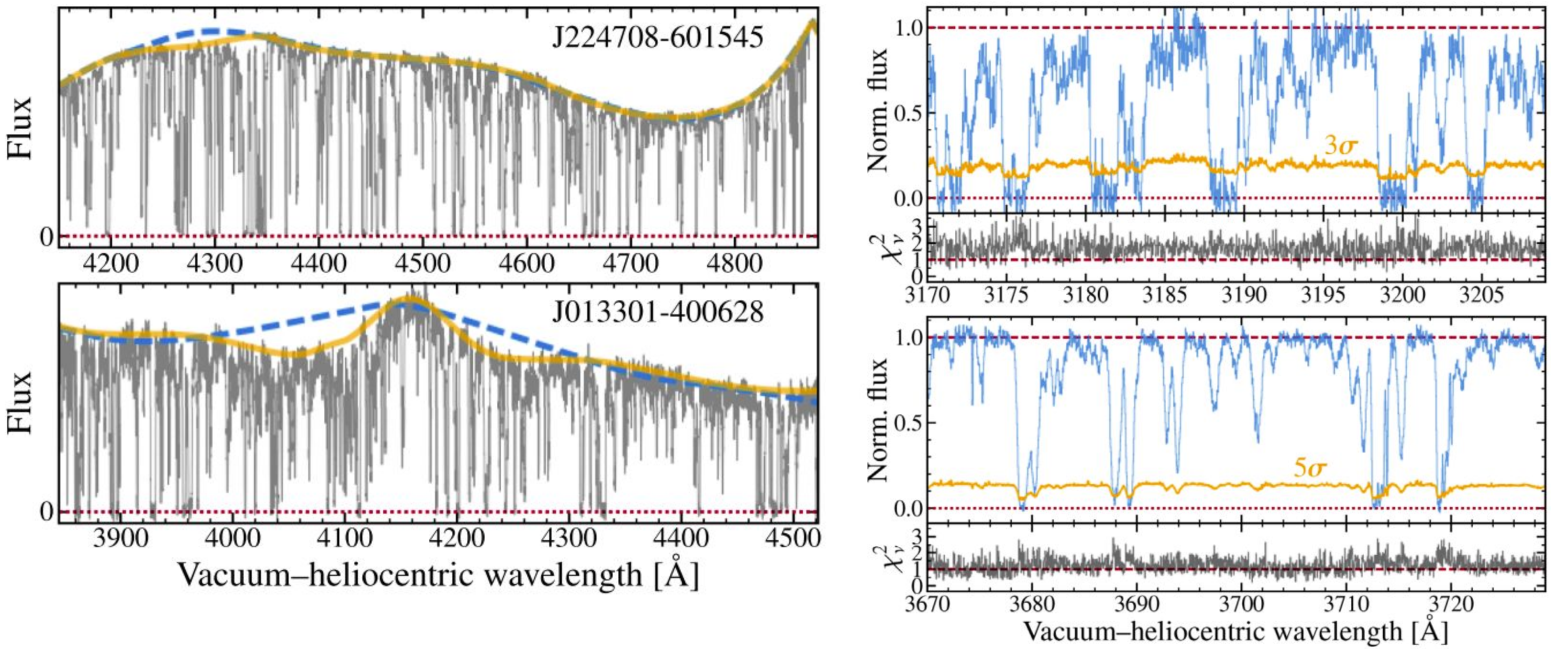
Now let's look at some realistic Lyman-α forest spectra

SQUAD: High resolution (R~45,000), VLT/UVES spectrograph



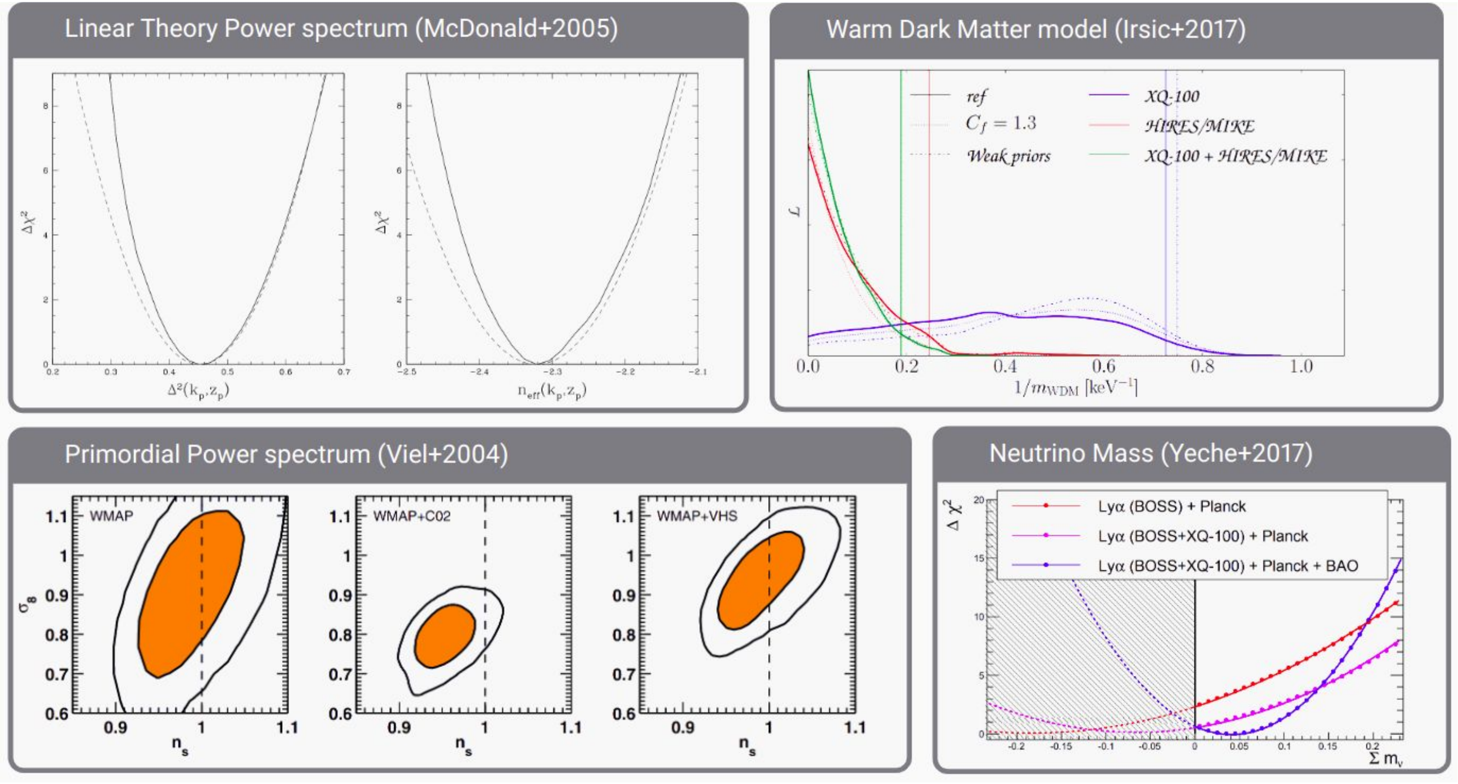
Now let's look at some realistic Lyman-α forest spectra

SQUAD: High resolution (R~45,000), VLT/UVES spectrograph

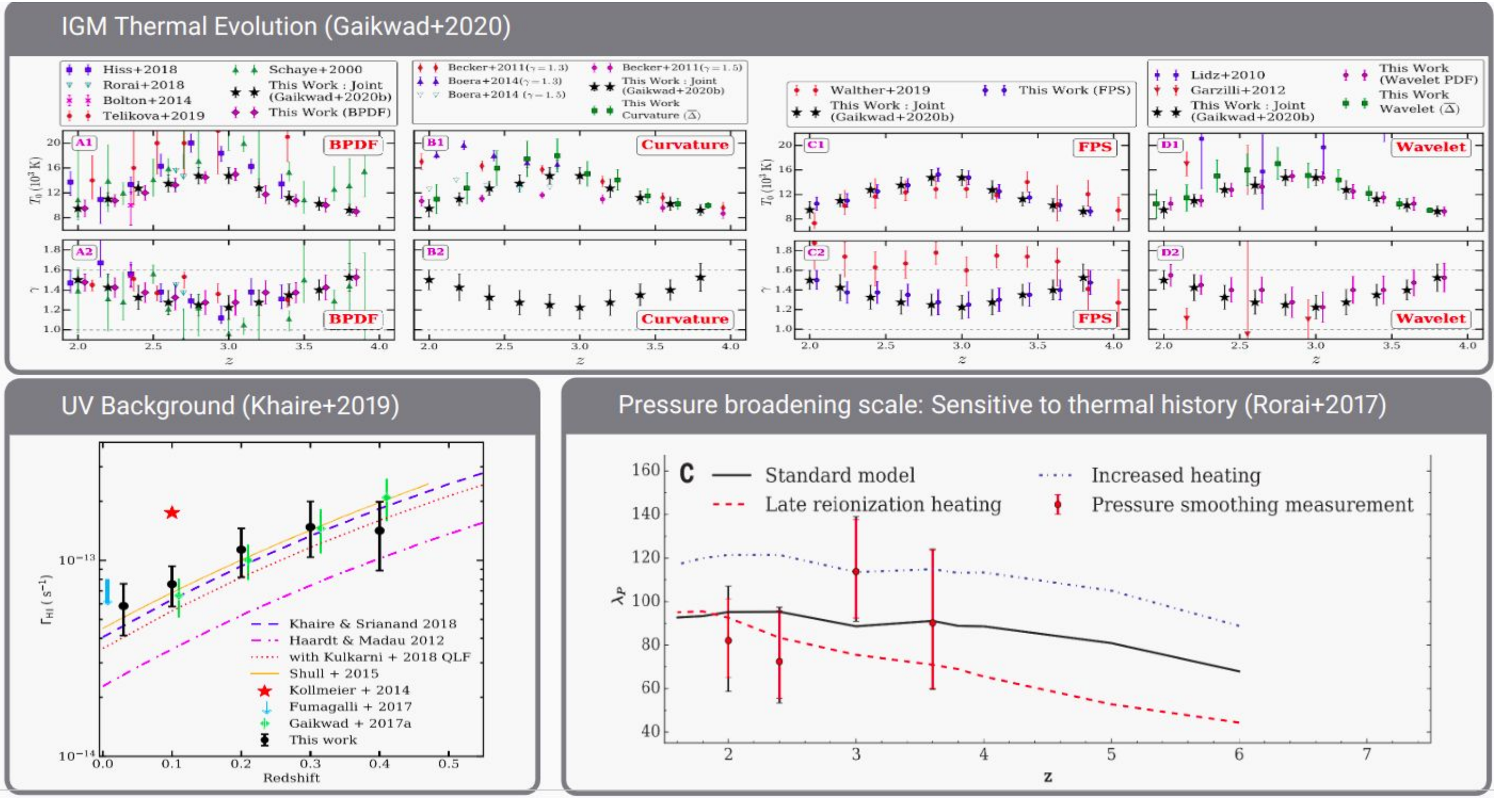


Murphy et al. 2018

Lyman-α forest: Cosmological Utility



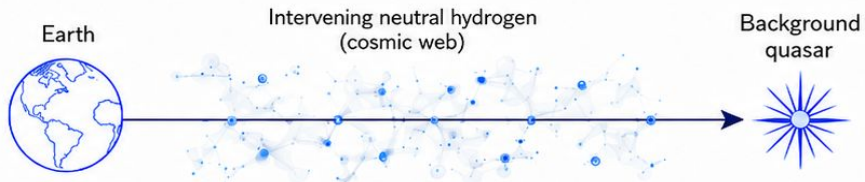
Lyman-α forest: Astrophysical Utility

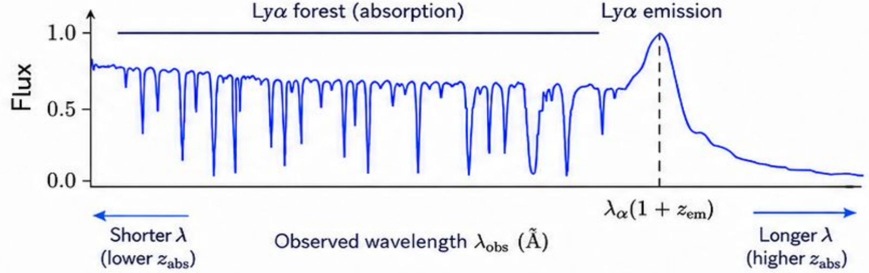


Take-home messages: Lyα forest in cosmology

A one–dimensional probe of the cosmic web

1. A background quasar provides bright light that passes through intervening neutral hydrogen.
2. Absorption at the Lyα transition ($\lambda = 1215.67\,\tilde{\text{Å}}$) imprints a series of features blueward of the quasar's Lyα emission line.
3. Observed wavelength maps to absorber redshift: $\lambda_{\rm obs} = \lambda_\alpha(1 + z_{\rm abs})$.
4. The transmitted flux is $F = F_{\rm obs}/F_{\rm cont} = \exp(-\tau_{\rm HI})$.
5. The optical depth depends on density Δ, temperature T, photoionization rate $\Gamma_{\rm HI}$, and peculiar velocity $v_{\rm pec}$.
6. Many quasar sightlines together provide flux statistics and constraints on the IGM, the cosmic web, and cosmology.





The Lyα forest is a sparse but powerful map of the cosmic web: each quasar spectrum is a 1D skewer through the intergalactic medium, and many such skewers become a statistical probe of structure formation.