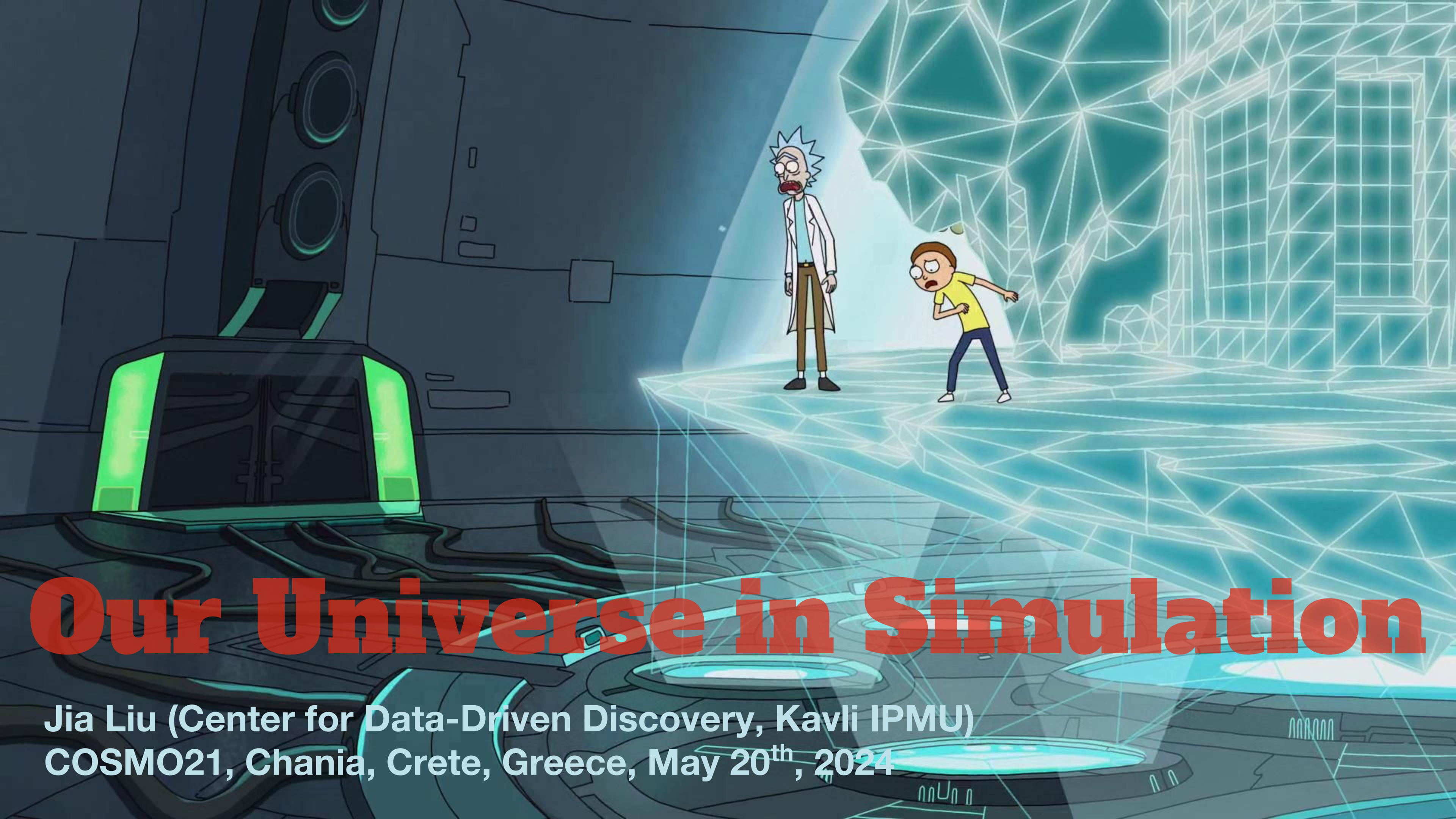


Our Universe in Simulation

Jia Liu (Center for Data-Driven Discovery, Kavli IPMU)

COSMO21, Chania, Crete, Greece, May 20th, 2024

Center for Data-Driven Discovery

CD3

Est. April 2023 cd3.ipmu.jp/

Theoretical Physics

Mathematics

Experimental (Astro)physics

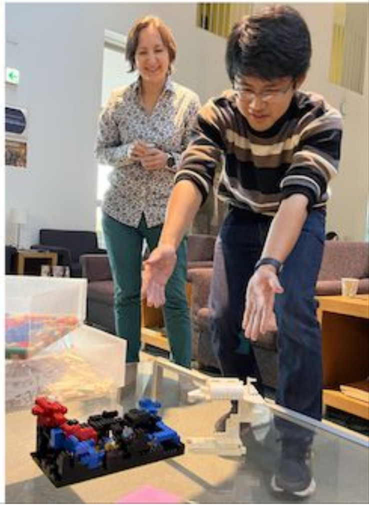
+ **Data Science & AI/ML**

Located 39 min. outside of **Tokyo**



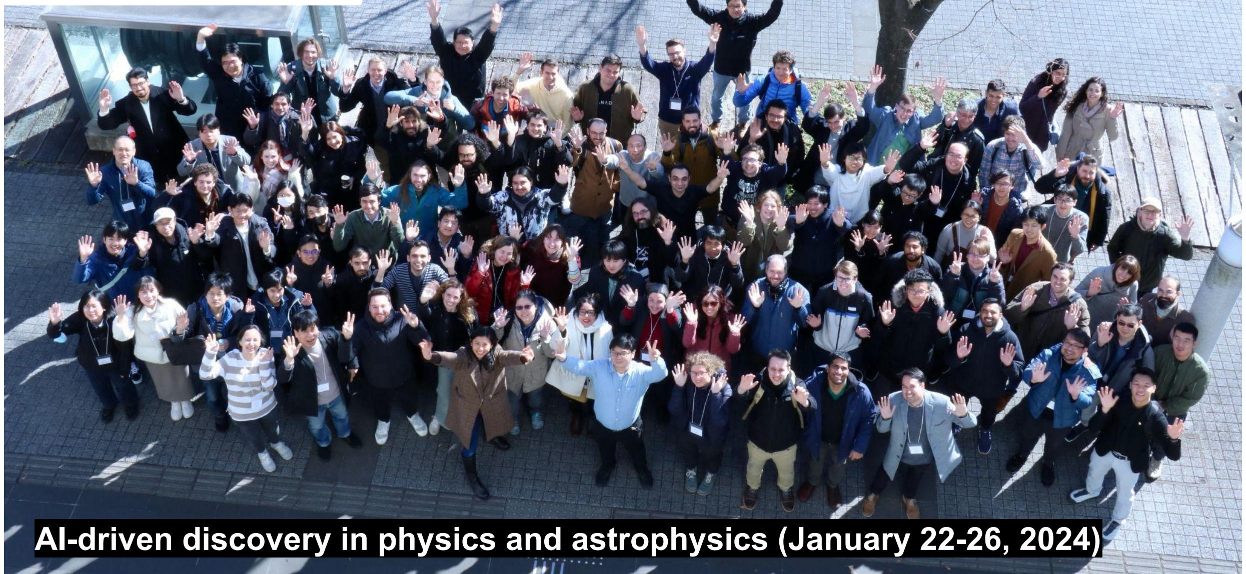


CD3 x DLX
LEGO Hackathon
November 24, Friday 13:30
Kavli IPMU 3F

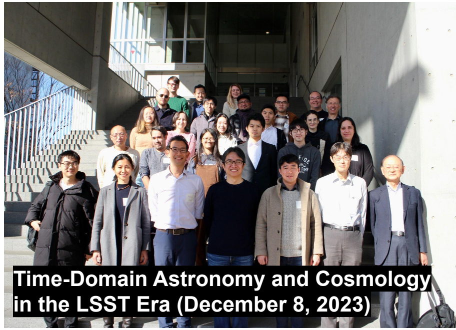
Future Science with CMB x LSS (April 10-14, 2023, YITP, Kyoto)

Astro AI with Fugaku (Sept. 11-12, 2023, U Tsukuba)



AI-driven discovery in physics and astrophysics (January 22-26, 2024)

CD3 Opening Symposium (April 19-20, 2023)



Time-Domain Astronomy and Cosmology in the LSST Era (December 8, 2023)

Baryons in the Universe 2024 (April 8-12, 2024)

Future Science with CMB x LSS (April 10-14, 2023, YITP, Kyoto)

Baryons in the Universe 2024 (April 8-12, 2024)

AI-driven discovery in physics and astrophysics (January 22-26, 2024)

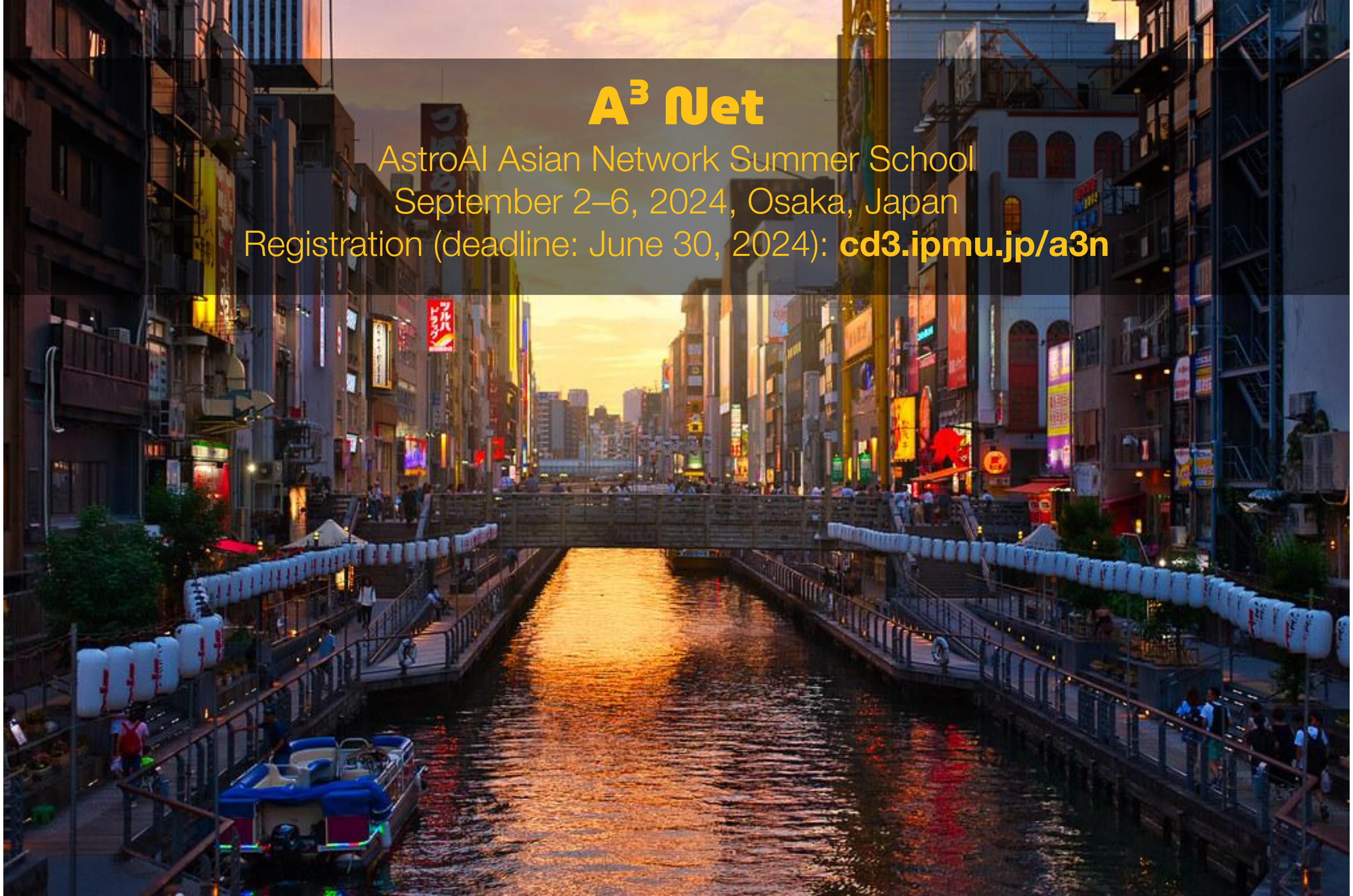
CD3 x CMB x Astro Seminar by Jo Dunkley (Feb. 5, 2024)

A³ Net

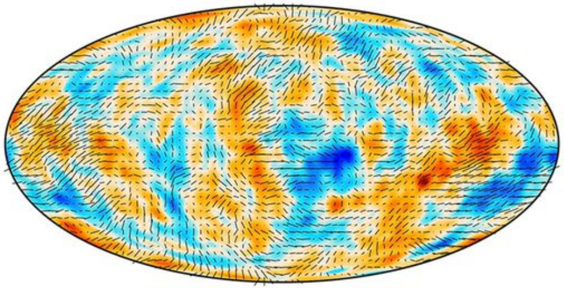
AstroAI Asian Network Summer School

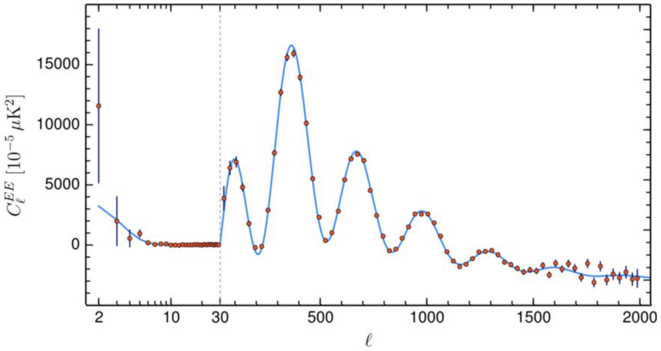
September 2–6, 2024, Osaka, Japan

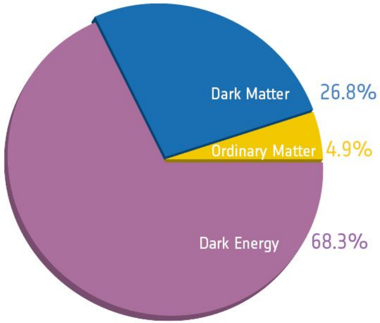
Registration (deadline: June 30, 2024): **cd3.ipmu.jp/a3n**



Cosmological analysis: the usual practice







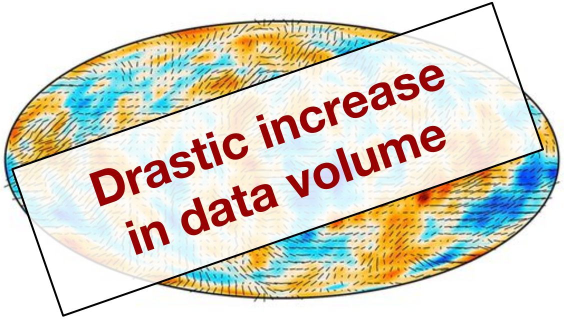
Observables
10^7 pixels/galaxies

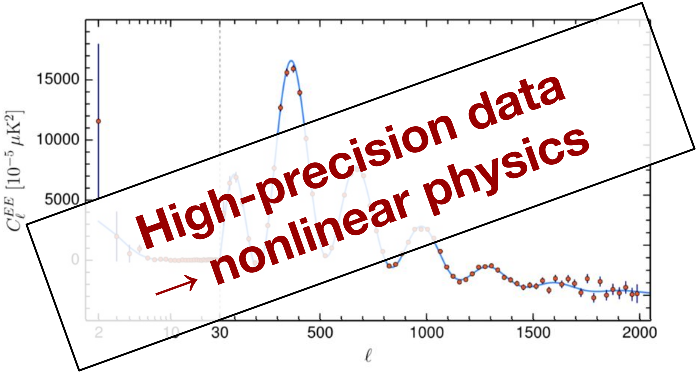
Summary Statistics
10^2–10^4 numbers

Physical Parameters
O(10)

Left: CMB polarization map; middle: polarization power spectrum; right: inferred cosmological parameters; Credit: ESA and Planck Collaboration

Cosmological analysis: the usual practice





Observables
10^7 → **10^{10}** pixels/galaxies

Summary Statistics
10^2–10^4 numbers

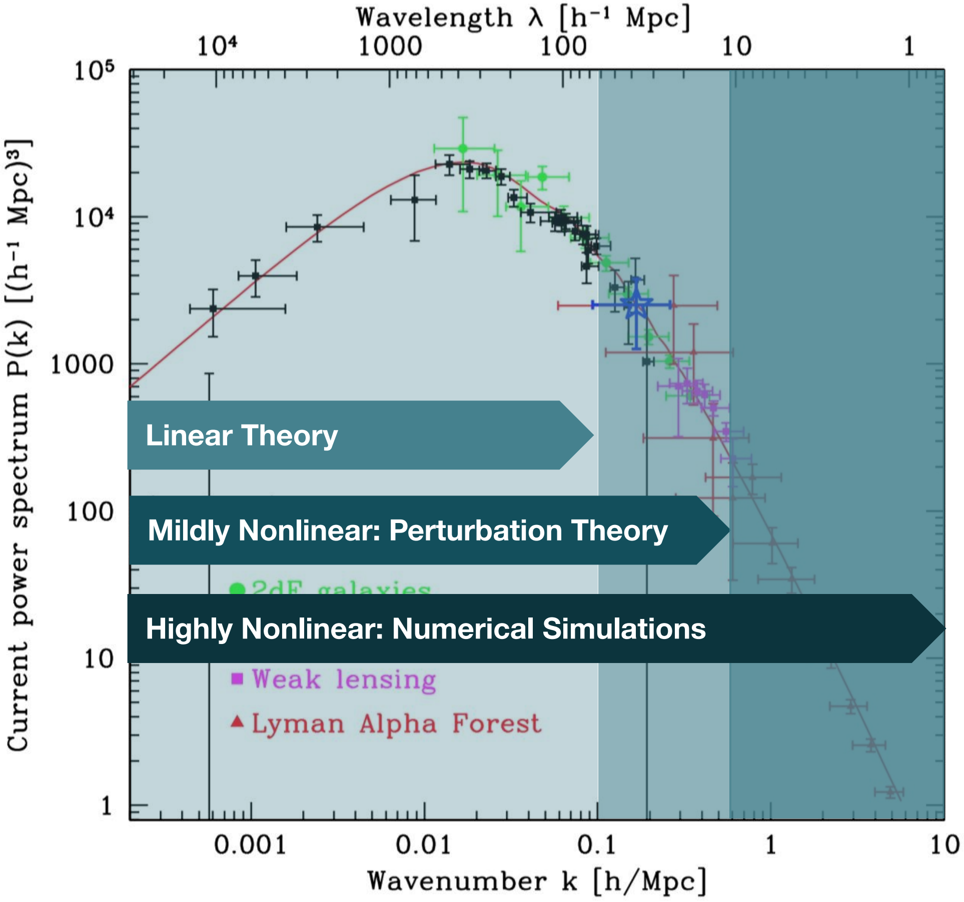
Physical Parameters
O(10)

Left: CMB polarization map; middle: polarization power spectrum; right: inferred cosmological parameters; Credit: ESA and Planck Collaboration

Rationale

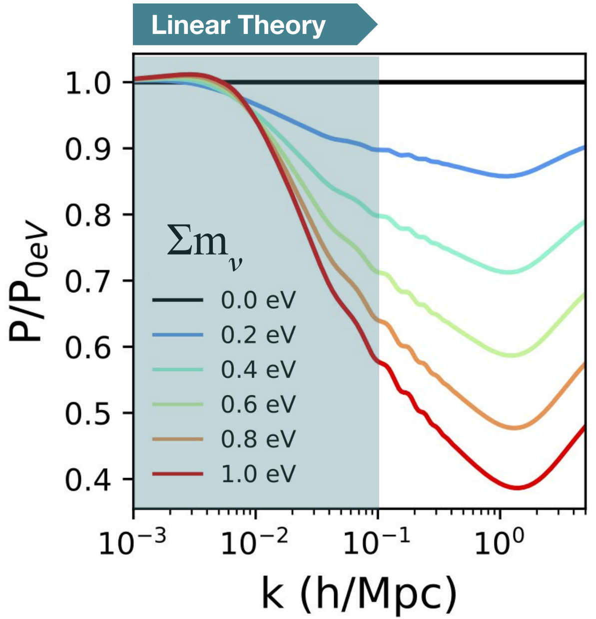
The COSMO21 series began in Lisbon in 2014, as an IAU Symposium, and was last held in Valencia in 2018. Since then, much has changed in the statistical analysis of cosmological data, with a large rise in the use of machine learning techniques in particular. Bayesian hierarchical modelling, using field-level inference, has been increasingly widespread, and simulation-based (or likelihood-free) inference is growing rapidly in the field, to tackle the complexity of low-redshift data whose (non-Gaussian) statistical properties may be very poorly known. Allied to that has been an increased interest in extreme data compression such as MOPED and more general score compression, to reduce the dimensionality for SBI. Furthermore, "information-maximizing" neural network techniques for finding highly informative summary statistics is emerging as an effective technique for extreme data compression for data whose statistical properties are unknown. Machine learning techniques are also coming to the fore in characterizing the complex posterior distributions, such as using neural ratio estimation (as one of a number of options) and often using variational inference techniques such as normalizing flows. With the advent of automatic differentiation, differentiable forward models are also emerging as powerful tools for Bayesian inference. And finally, interpretable machine learning is one that is beginning to challenge researchers in our field as elsewhere. This methodological development is accompanied by an upcoming explosion of data, expected from Euclid and the Rubin Observatory, and in due course the Roman Space Telescope and the Square Kilometre Array, to add to the current and recent survey data from KiDS, DES, HSC, and DESI, for example.

Most (if not all) these methods require simulations

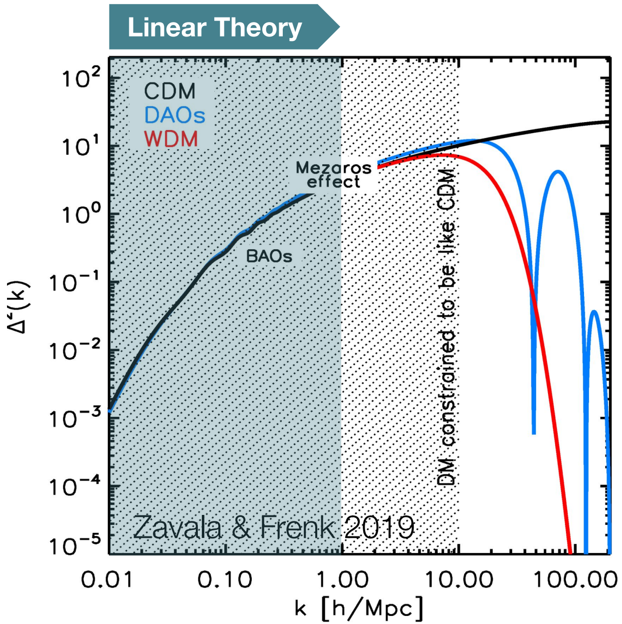


Tegmark 2004

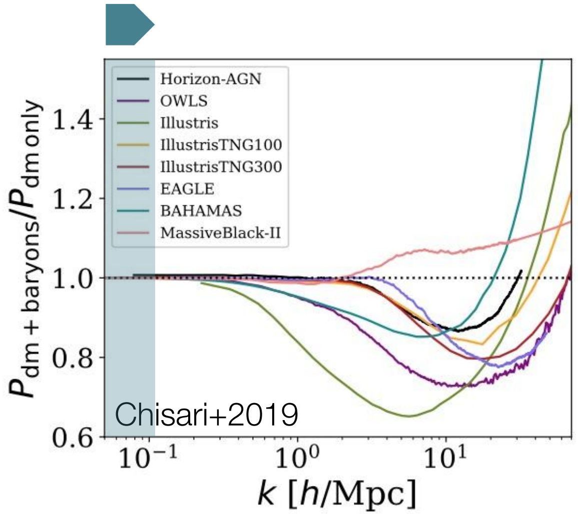
Neutrino Mass



Dark Matter



Baryons



See Daniela Grandon's talk

	Variable parameters	Number of simulations for variable parameters	Number of simulations for the fiducial cosmology	Box size [Mpc/h]	Number of particles
CosmoGridV1	Ω_m, σ_8, H_0, w_0, n_s, Ω_b	2500×7	200	900	832^3
cosmo-SLICS	Ω_m, σ_8, H_0, w_0	25×2	800	505	1536^3
DarkGridV1	Ω_m, σ_8	58×5	50	900	768^3
MassiveNuS	Ω_m, σ_8, M_ν	101×1	n/a	512	1024^3
DH10	Ω_m, σ_8	158×1	n/a	140	256^3

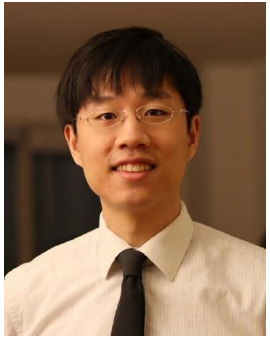
Table 1. List of simulation sets used in map-level simulation-based inference of cosmological parameters from LSS probes, either for forecasts or measurements. The cosmo-SLICS simulations description is taken from Harnois-Déraps et al. [100], the MassiveNus from Liu et al. [98], the DarkGridV1, a precursor to CosmoGridV1, from Zürcher et al. [46], and the DH10 from Dietrich & Hartlap [36]. This list includes only simulations that have lightcone shells/snapshots output dense enough to enable map-level inference with projected probe maps from LSS surveys. This list is not exhaustive; other notable datasets were used by Fluri et al. [17], Liu et al. [42], Zorrilla Matilla et al. [106].

Also the **Gower Street Sims** (Jeffrey+2024) and for 3D fields: the **Quijote** simulations (Villaescusa-Navarro+2019) and the **CAMELS** Simulations (Villaescusa-Navarro+2020)

Kacprzak+2023

Gabriela Marques
Fermilab
(Peaks & Minima)

Leander Thiele
Kavli IPMU
(PDF)



Sihao Cheng
IAS
(Scattering Transform)

Daniela Grandon
Leiden
(Baryonic Effects)

Jess Cowell
Kavli IPMU/Oxford
(Marked PS)

Joaquin Armijo
Kavli IPMU
(Minkowski Functionals)

HSC Weak Lensing
Higher-Order Statistics

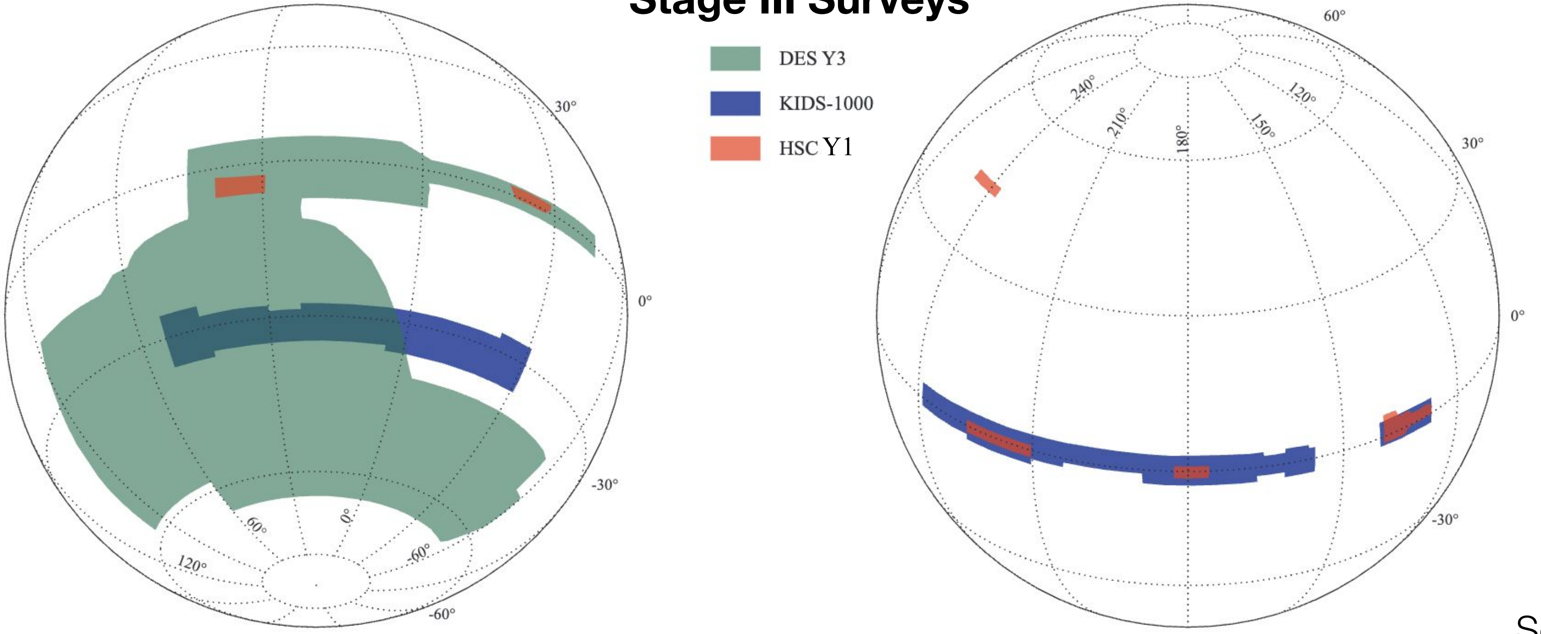
Camila Noaves
INPE
(Implicit Likelihood)

Will Coulton
Cambridge
(Bispectrum)

Masato Shirasaki
NAOJ
(DMO Simulation)

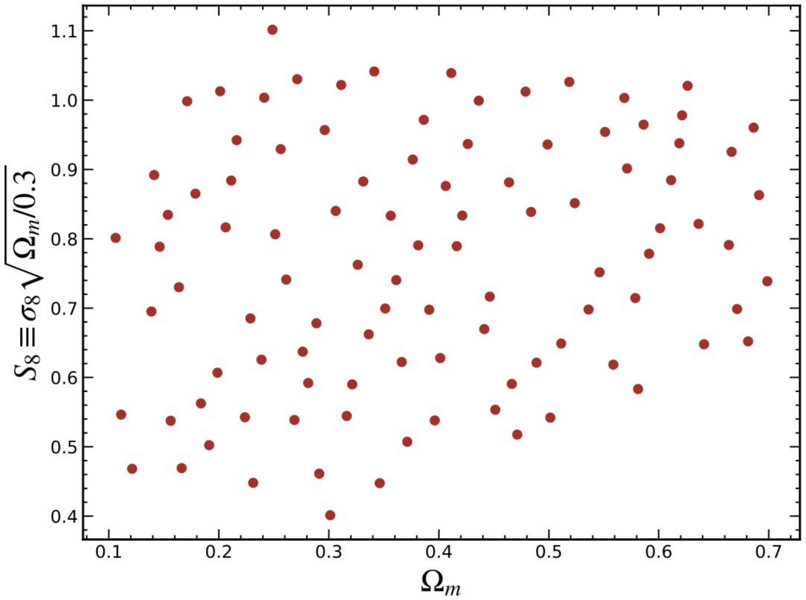
Ken Osato
Chiba U
(Baryonic Simulation)

Stage III Surveys



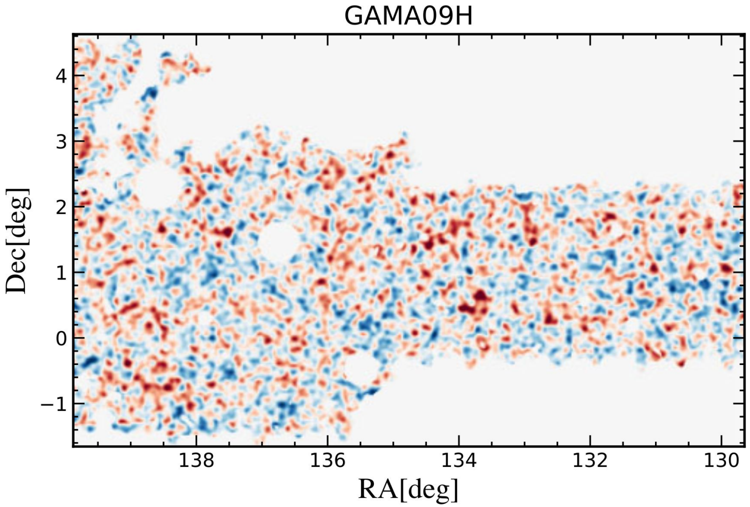
Secco+2022

	HSC	DES	KIDS
FoV (deg2)	1.8	3	1
Area (deg2)	1100	5000	1350
N_{gal} (arcmin $^{-2}$)	**22**	7	9



Name	# of realizations
Fiducial	2268
Photo-z run 1	100
Photo-z run 2	100
multiplicative-bias run 1	100
multiplicative-bias run 2	100
Cosmology-varied run	50×100

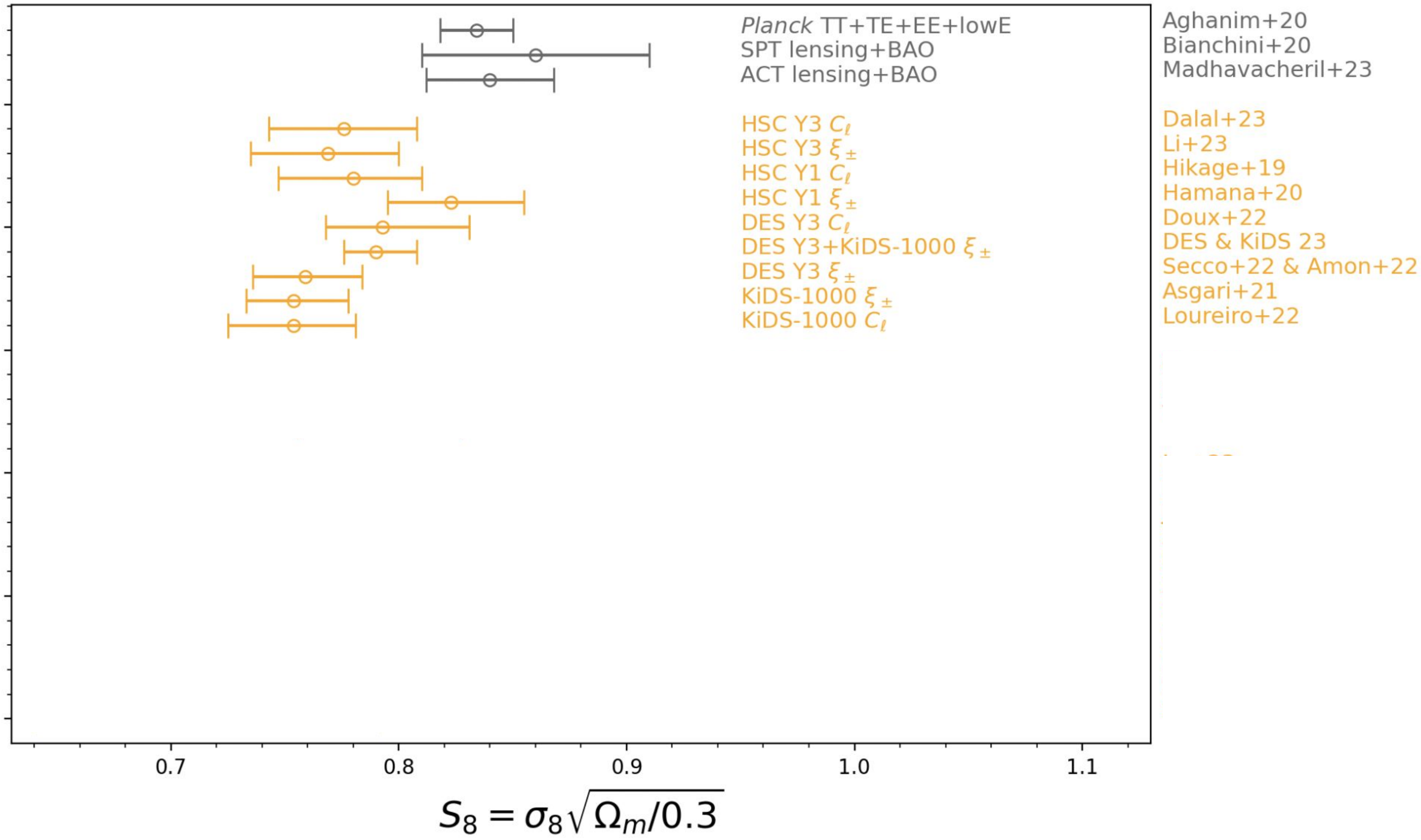
Shirasaki+2021
Marques+2024



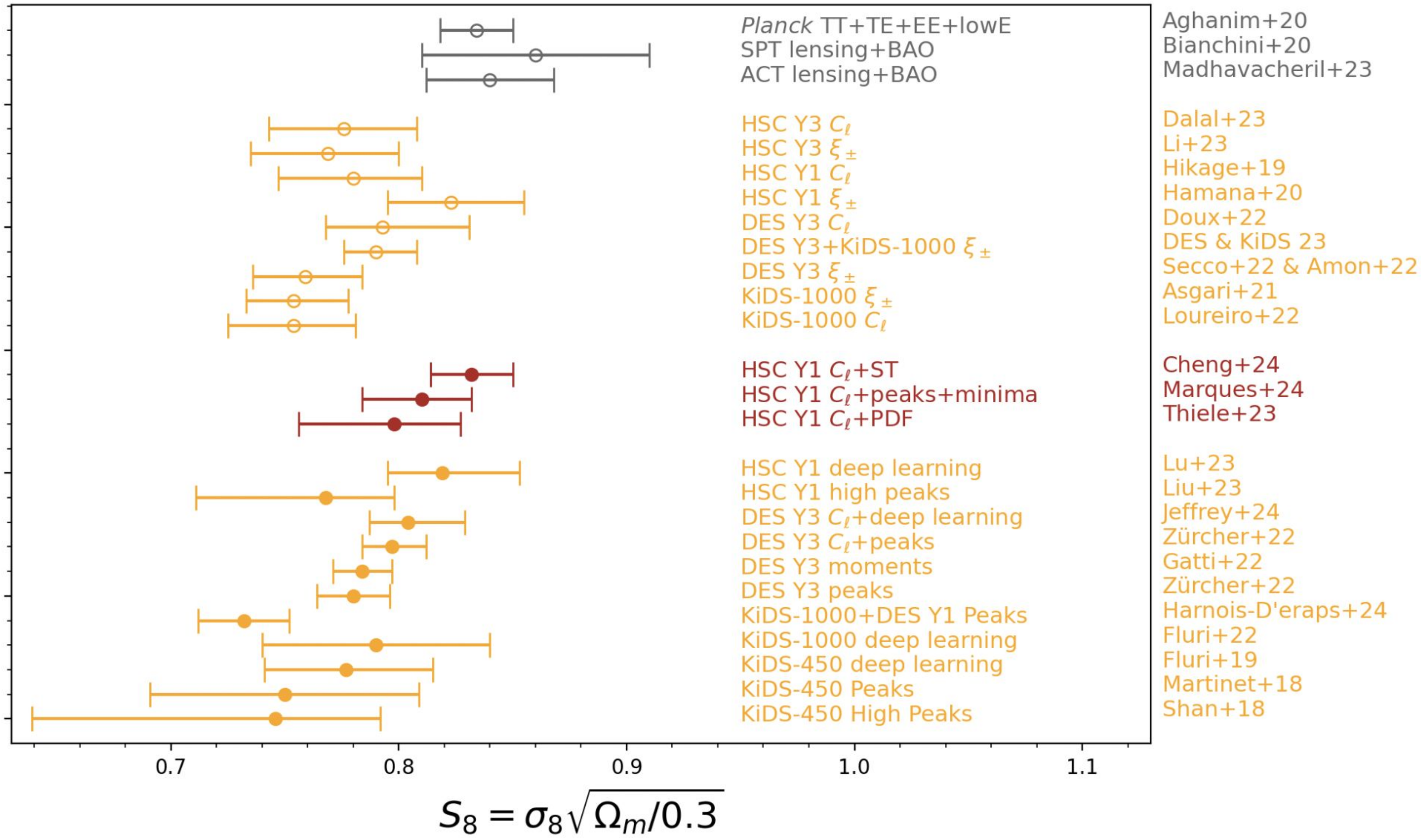
Three level blinded analysis (honor system):

- Built pipeline with only simulations and define data cuts to mitigate systematics
- Unblind B-mode and randoms
- Unblind the power spectrum
- Unblind the full data

Marques+2024



Thiele+2023
Marques+2024
Grandon+2024
Cheng+2024



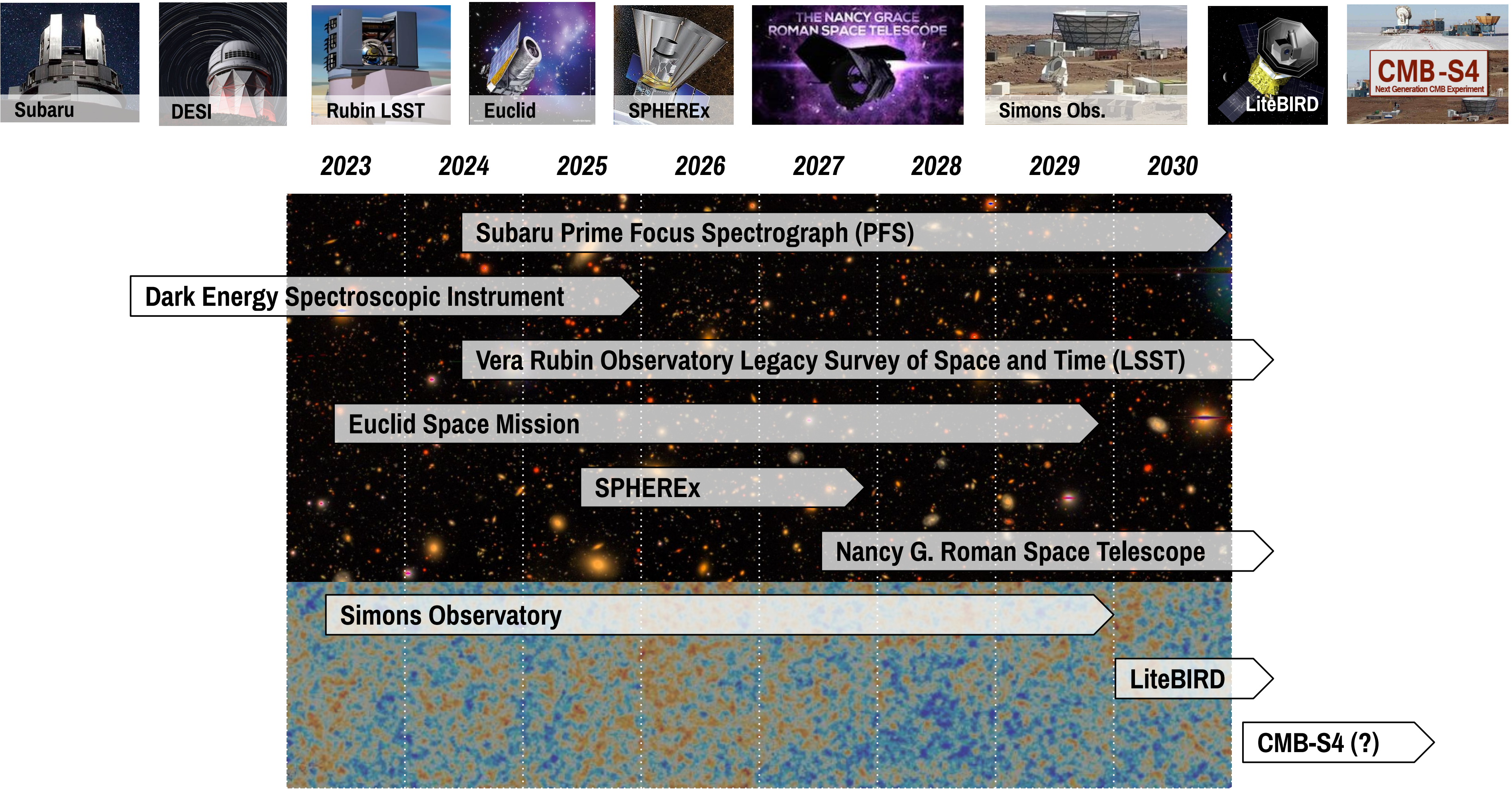
Thiele+2023
Marques+2024
Grandon+2024
Cheng+2024

Large-scale structure surveys: **Stage III → Stage IV**

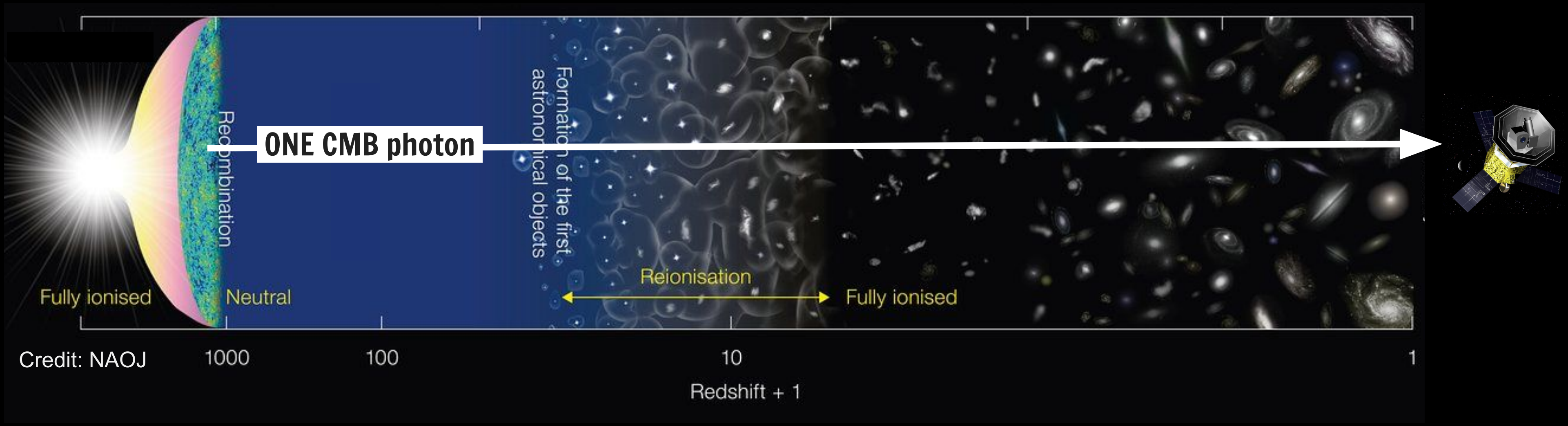
	HSC	DES	KIDS	EUCLID	LSST	CSST	ROMAN
FoV (deg2)	1.8	3	1	0.5	3.5	1.1	0.28
Area (deg2)	1100	5000	1350	15000	18000	15000	2000
N_{gal} (arcmin^{-2})	22	7	9	30	30	28	50

Higher-order stats topical team, co-lead Joachim Harnois-Deraps

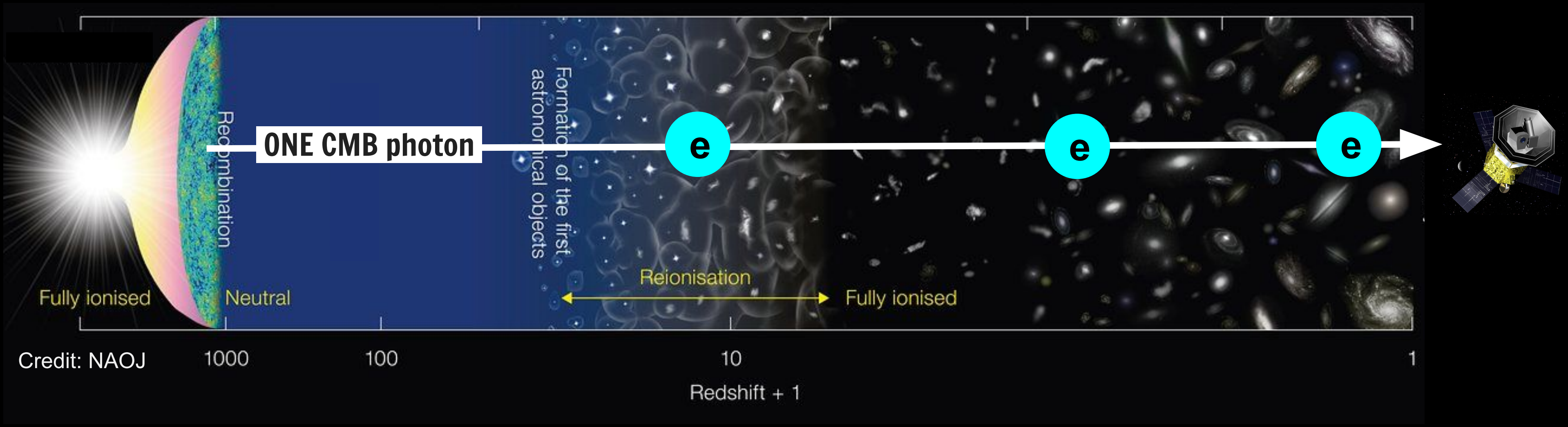
* These are approximate numbers. Please refer to official documents for up-to-date numbers.



Subaru
DESI
Rubin LSST
Euclid
SPHEREx
THE NANCY GRACE ROMAN SPACE TELESCOPE
Simons Obs.
LiteBIRD
CMB-S4
Next Generation CMB Experiment
2023
2024
2025
2026
2027
2028
2029
2030
Subaru Prime Focus Spectrograph (PFS)
Dark Energy Spectroscopic Instrument
Vera Rubin Observatory Legacy Survey of Space and Time (LSST)
Euclid Space Mission
SPHEREx
Nancy G. Roman Space Telescope
Simons Observatory
LiteBIRD
CMB-S4 (?)

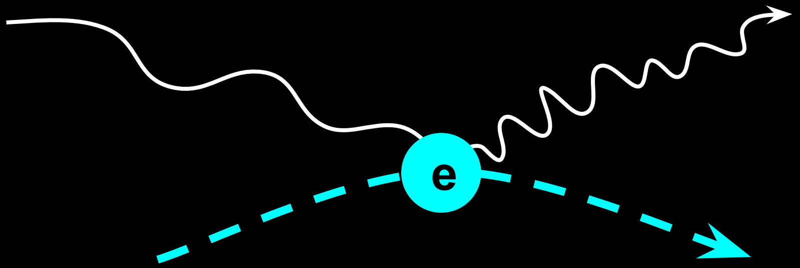

ONE CMB photon
Recombination
Formation of the first astronomical objects
Reionisation
Fully ionised
Neutral
Fully ionised
1000
100
10
1
Redshift + 1

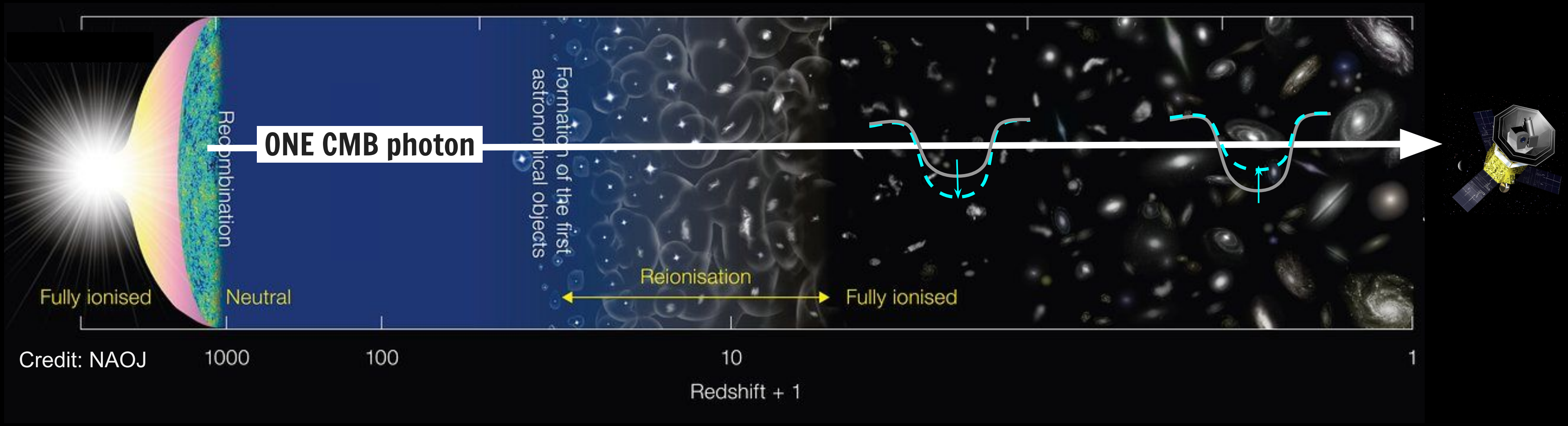
Credit: NAOJ

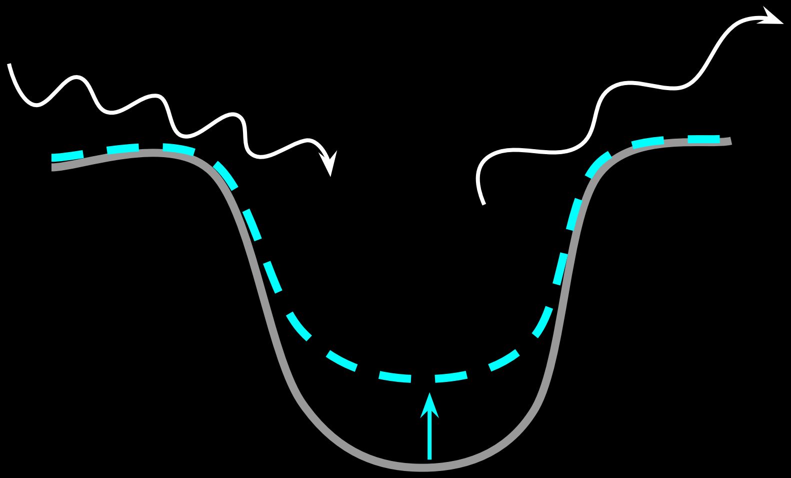


kinetic SZ
thermal SZ

Interaction with electrons



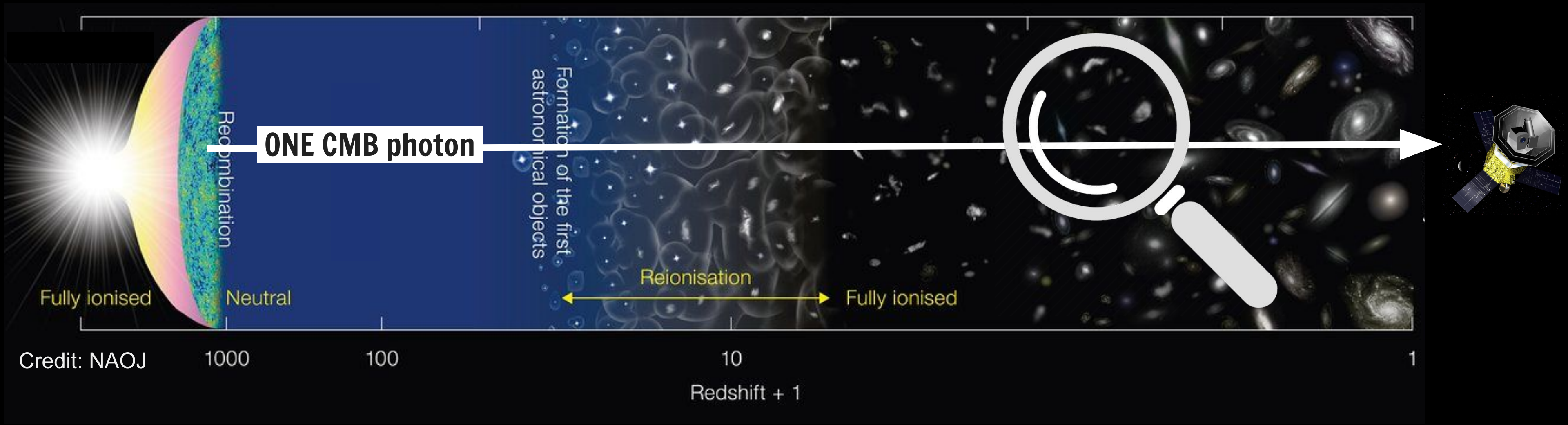




Integrated SW

(also: Rees-Sciama, Moving lens)

Evolving gravitational potential

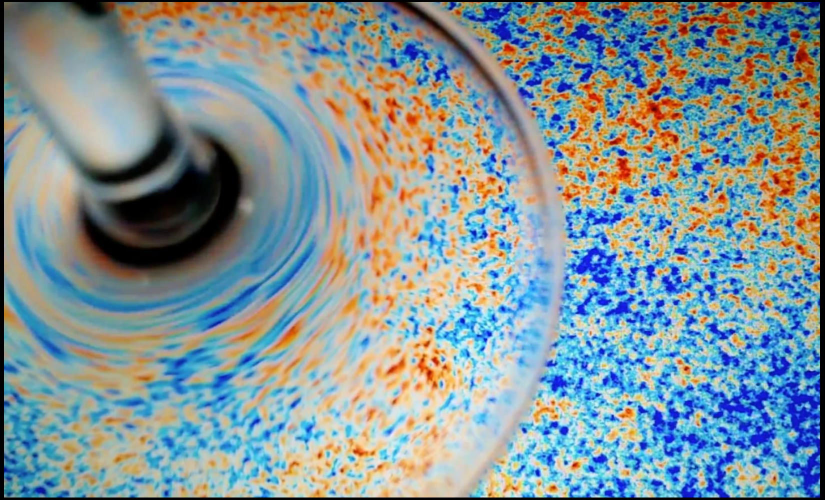


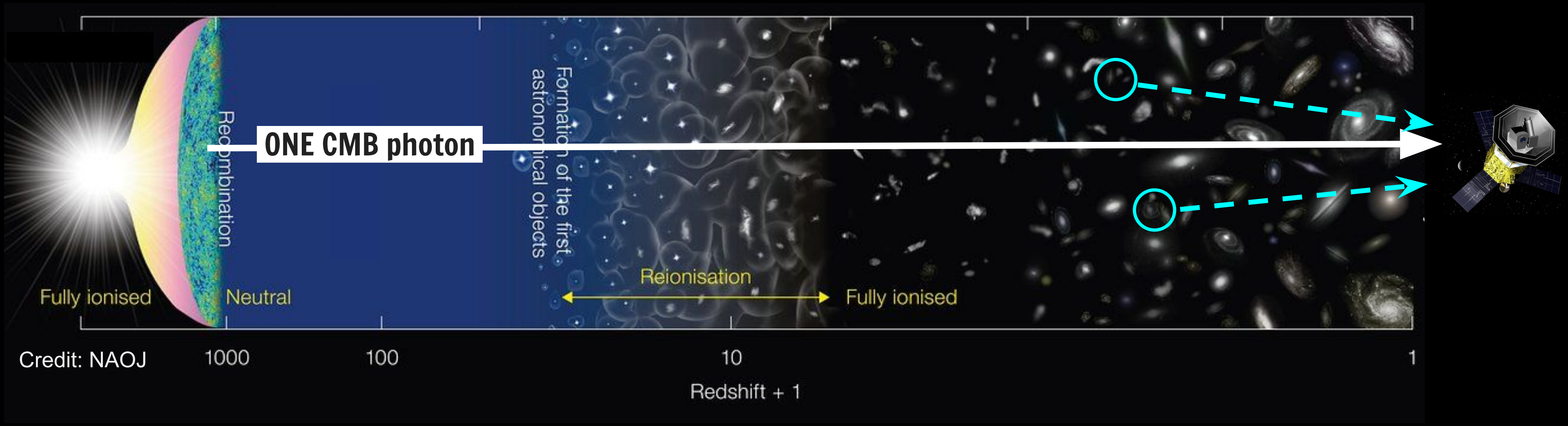
Credit: NAOJ

CMB Lensing

Photon’s path bent by curved spacetime

Credit: Emmanuel Schaan



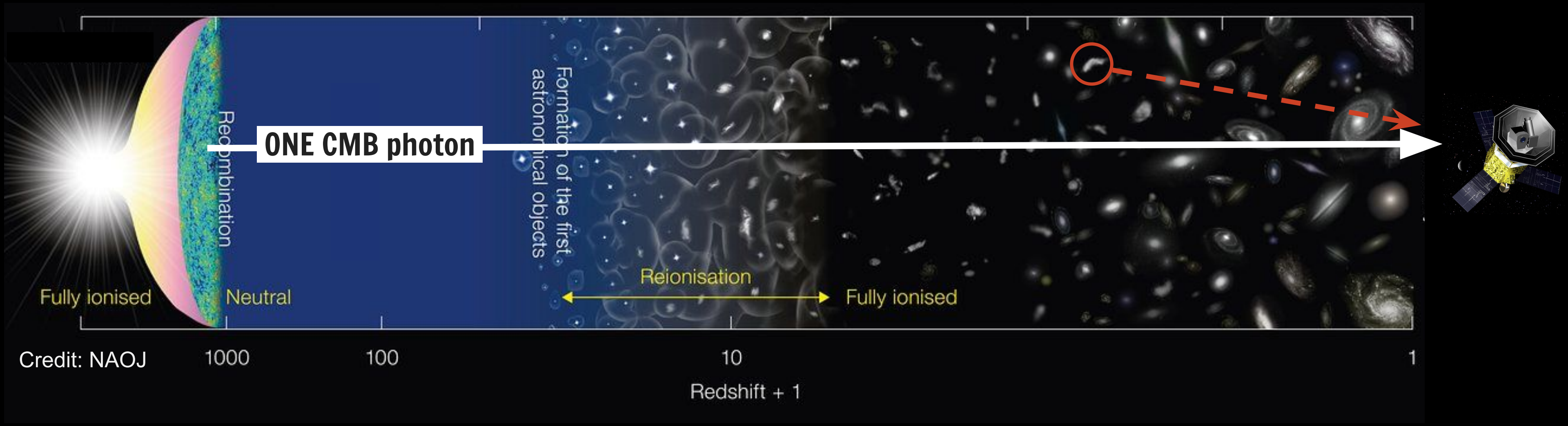


Credit: NAOJ

Cosmic Infrared Background

Emission from dusty galaxies

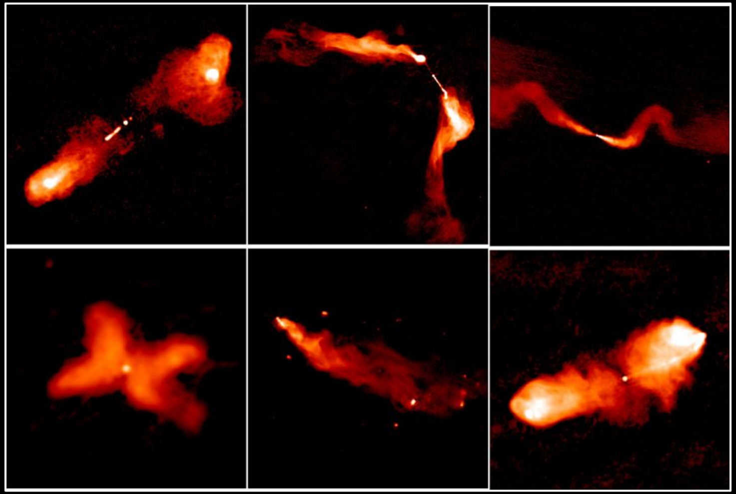

Credit: HST

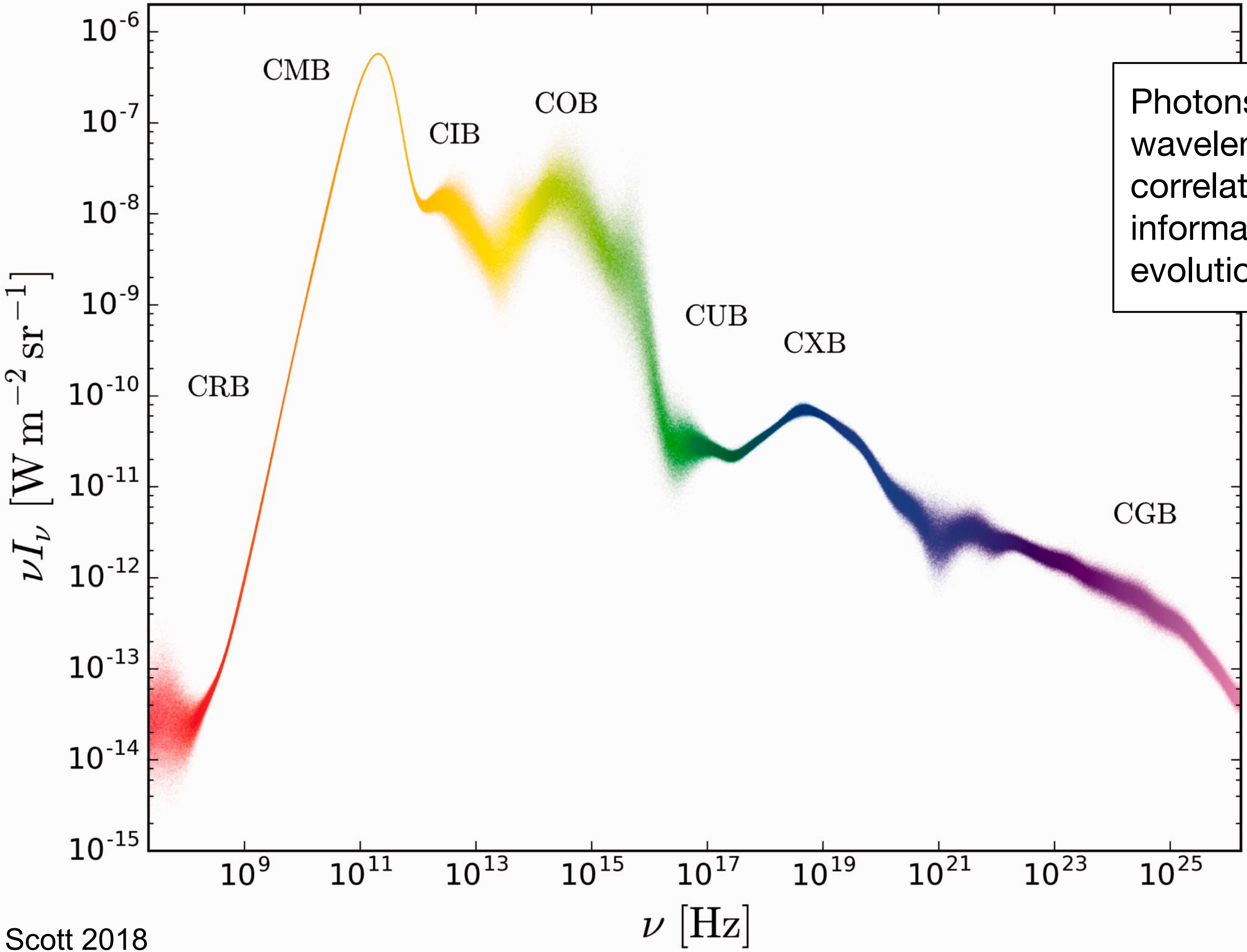


Radio Galaxies

Synchrotron emission from AGNs

Credit: Hardcastle & Crostonb

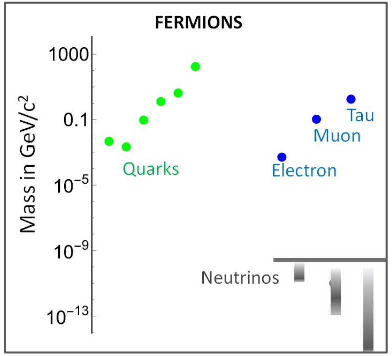




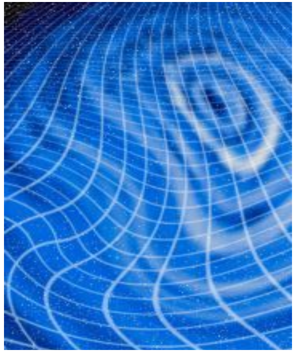
Photons at multiple wavelengths are correlated and contain information of the evolution of our universe

Hill, Masui & Scott 2018

Future discoveries rely on CMB & LSS

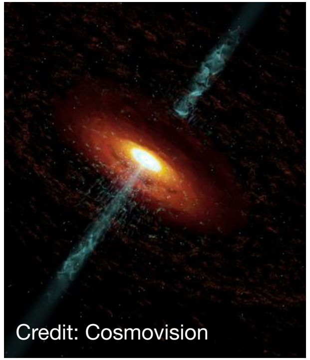


Neutrino Mass



Inflationary grav. waves

Primordial non-Gaussianity



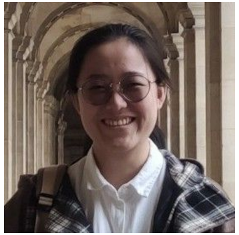
Astrophysics, and more!

HalfDome simulations for Stage IV Surveys

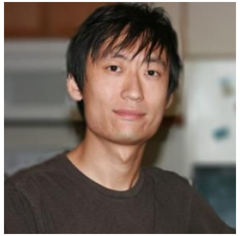
Adrian Bayer
Princeton/CCA

Zack Li
Berkeley/LBL

Joe DeRose
Berkeley/LBL



Yici Zhong
U Tokyo



Yu Feng
Google

Linda Blot
Kavli IPMU

Marcelo Alvarez
Stanford

Junjie Xia
Kavli IPMU

Hideki Tanimura
Kavli IPMU

Will Coulton
Cambridge

Giuseppe Puglisi
U Catania

Alex Laguë
U Penn

Mat Madhavacheril
U Penn

HalfDome simulations for Stage IV Surveys

	mmDL Sehgal+2010 Han+2021	**Websky** Stein+2020 Li+2022	**Agora** Omori 2022
N-body box $N_{particles}$	1 Gpc/h 1024^3	7.7 Gpc 6144^3	1 Gpc/h 3840^3
Min. M_{halo}	10^{13} $M_{\odot}$	1.2×10^{13} $M_{\odot}$	1.5×10^{9} $M_{\odot}/h$
LSS observables	None	None	κ, clusters, LIM
No. realizations	**1**	**1**	**1**

* Inputs from SO, CMB-S4, LSST, DESI, PFS, SPHEREx, Roman collaborators

HalfDome simulations for Stage IV Surveys

	mmDL Sehgal+2010 Han+2021	**Websky** Stein+2020 Li+2022	**Agora** Omori 2022	**Stage IV requirements***
N-body box $N_{particles}$	1 Gpc/h 1024^3	7.7 Gpc 6144^3	1 Gpc/h 3840^3	a few Gpc
Min. M_{halo}	$10^{13}\ M_\odot$	$1.2 \times 10^{13}\ M_\odot$	$1.5 \times 10^9\ M_\odot/h$	$10^{12}\ M_\odot$
LSS observables	None	None	κ, clusters, LIM	κ, galaxies,clusters
No. realizations	**1**	**1**	**1**	**10–100**

* Inputs from SO, CMB-S4, LSST, DESI, PFS, SPHEREx, Roman collaborators

HalfDome simulations for Stage IV Surveys

	mmDL Sehgal+2010 Han+2021	**Websky** Stein+2020 Li+2022	**Agora** Omori 2022	**Stage IV requirements***
N-body box $N_{particles}$	1 Gpc/h 1024^3	7.7 Gpc 6144^3	1 Gpc/h 3840^3	a few Gpc
Min. M_{halo}	10^{13} $M_\odot$	1.2 x 10^{13} $M_\odot$	1.5 x 10^9 $M_\odot/h$	10^{12} $M_\odot$
LSS observables	None	None	κ, clusters, LIM	κ, galaxies,clusters
No. realizations	**1**	**1**	**1**	**10–100**

* Inputs from SO, CMB-S4, LSST, DESI, PFS, SPHEREx, Roman collaborators

HalfDome simulations for Stage IV Surveys

	mmDL Sehgal+2010 Han+2021	**Websky** Stein+2020 Li+2022	**Agora** Omori 2022	**Stage IV requirements***	**Half Dome** 2024-
N-body box $N_{particles}$	1 Gpc/h 1024^3	7.7 Gpc 6144^3	1 Gpc/h 3840^3	a few Gpc	3.5 Gpc/h, 6144^3
Min. M_{halo}	10^{13} $M_\odot$	1.2 x 10^{13} $M_\odot$	1.5 x 10^9 $M_\odot$/h	10^{12} $M_\odot$	10^{12} $M_\odot$
LSS observables	None	None	κ, clusters, LIM	κ, galaxies,clusters	κ, galaxies, clusters, +more
No. realizations	**1**	**1**	**1**	**10–100**	11+1 f_{NL} (more to come)

See Adrian Bayer's talk

* Inputs from SO, CMB-S4, LSST, DESI, PFS, SPHEREx, Roman collaborators

Run	$N_{\rm sim}$	N_c	L [Mpc/h]	$N_{\rm side}$	$N_{\rm node}$	Hour	IC [%]	PM [%]	FOF [%]	IO [%]	Halos	Particles	Sheets
		Simulation setup			**Run time**						**Data volume**		
full res	11 (+1)	6144^3	3750	8192	2048	4.38	2.89	72.40	14.48	10.24	362G	17T	1.4T
1/2 res	1	4096^3	5000	8192	512	1.95	3.26	74.17	10.84	11.73	29G	7.7T	824G
1/4 res	1	2048^3	5000	8192	64	2.42	3.72	81.61	11.92	2.76	1.4G	1.2T	376G
1/8 res	1	1024^3	5000	2048	8	2.32	3.08	77.22	18.81	0.89	21M	149G	33G
1/16 res	1	512^3	5000	2048	1	3.01	1.73	65.86	32.21	0.21	176K	19G	8.9G

TABLE II. Summary of the HalfDome simulation N-body runs, including the number of simulations $N_{\rm sim}$, particle number N_c, box size L, the resolution of the HEALPix maps $N_{\rm side}$, run time, and data products and their corresponding volumes per simulation. Except for 1/2 resolution, which runs on NERSC Cori, all simulations are performed on Stampede2. The cosmic variance is turned off for 1/4 resolution. At the full resolution, in addition to the 11 default runs, we provide an additional run with primordial non-Gaussianity $f_{\rm NL}$=20.

Your contribution to the next generation simulations!

- New sciences / methods that can be enabled by simulations
- What aspects of simulations you'd like to see improving (new observables, resolution, number of sims., parameter sampling…)
- Please share your thoughts in: https://t.ly/JvHdl

A³ Net [AstroAI Asian Network]

First Summer school

September 2-6, 2024, Osaka, Japan

https://cd3.ipmu.jp/a3n

Registration open this week!

Photo: boutiquejapan.com