COSMO21

20-24 MAY, 2024
CHANIA, GREECE

CONFERENCE BOOKLET

FOUNDATION FOR RESEARCH AND TECHNOLOGY - HELLAS

Imperial College London

BE GREEN KEEP IT ON THE SCREEN

CONTENTS

COSMO21

Statistical Challenges in 21st Century Cosmology

All times are Greece time (EEST / GMT+3)

PREFACE

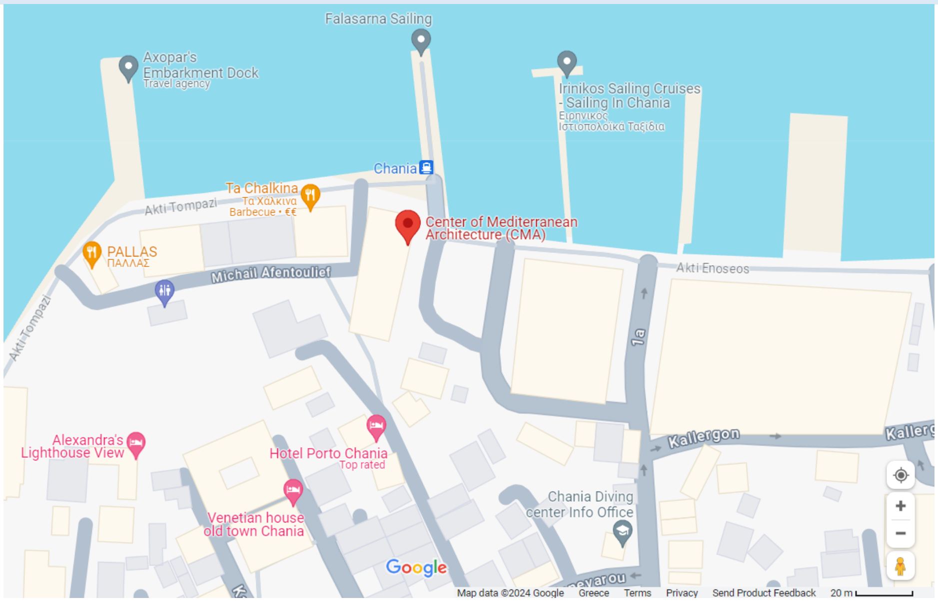
Rationale

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.

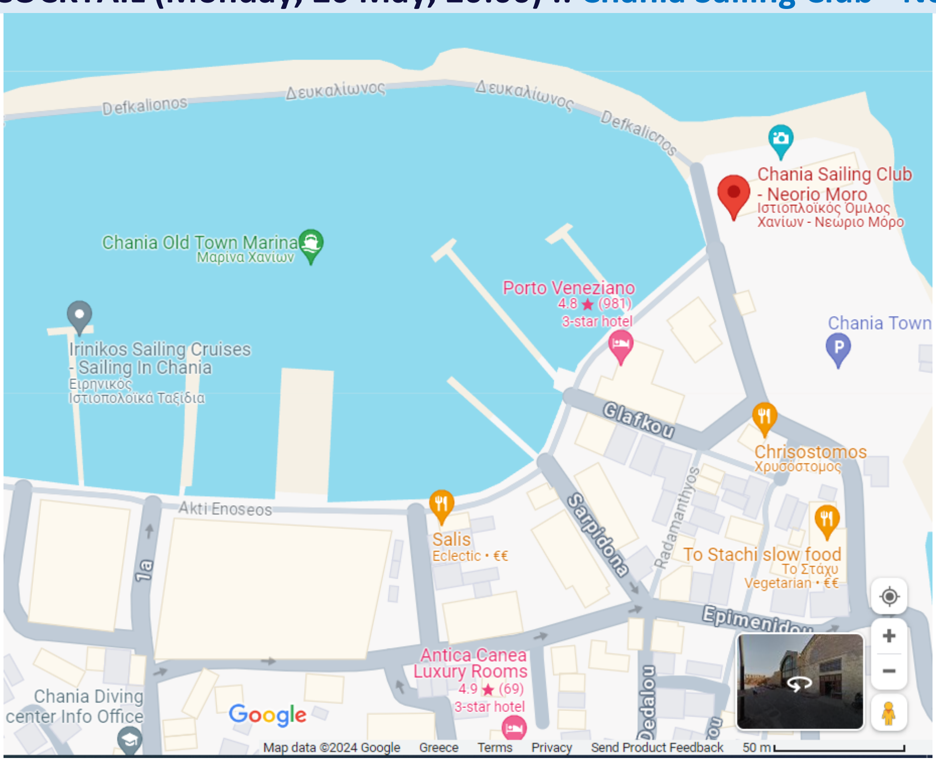
The 4th edition of COSMO21 will be hosted in Chania, Greece and will gather leading experts from around the world in statistical methods and cosmology to discuss the state-of-the-art in data analysis and interpretation. Highlights will include topics such as Bayesian techniques, machine learning, likelihood-free inference, radio data, strong and weak gravitational lensing, joint probes, and gravitational waves. The organizers also hope to attract participants to discuss other novel developments in the field.

VENUE & SOCIAL EVENTS

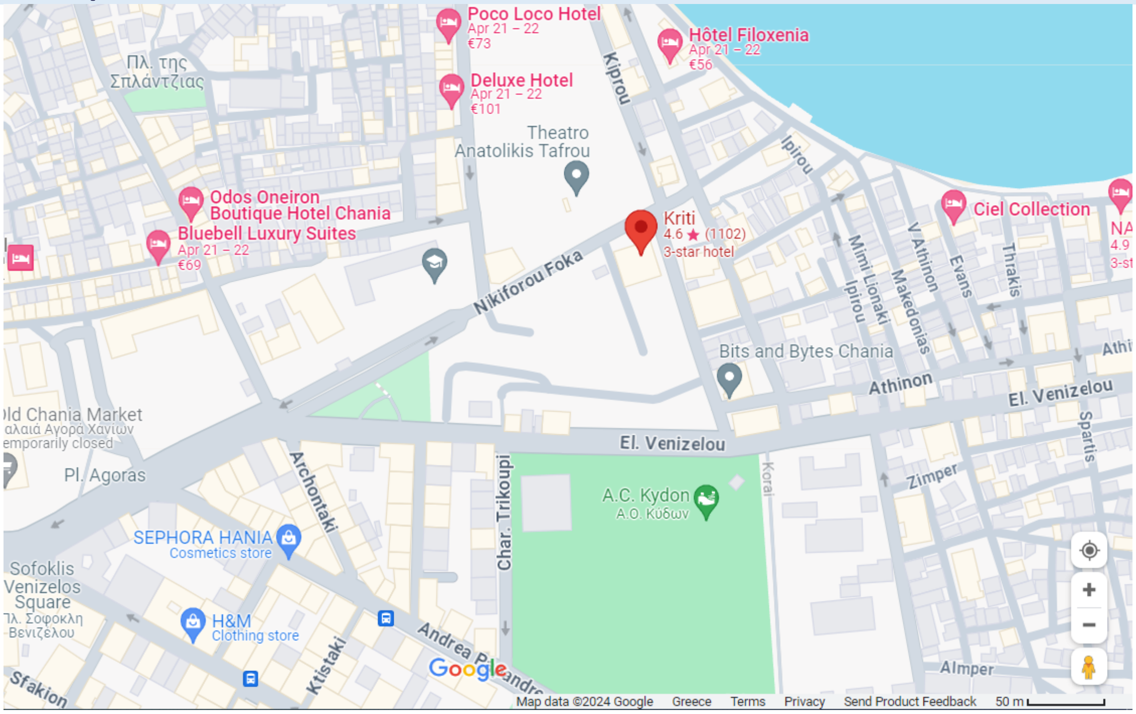
CONFERENCE VENUE :: Grand Arsenal – Conference Area



WELCOME COCKTAIL (Monday, 20 May, 20:00) :: Chania Sailing Club - Neorio Moro



GALA DINNER (Wednesday, 22 May, 20:00) :: Tavern Douliana
Departure point: KRITI Hotel, at 19:00



ORGANIZING COMMITTEES

COSMO21 – Scientific Organizing Committee

Susmita Adhikari	Indian Institute of Science Education and Research (IISER)
Alan Heavens	ICIC, Imperial College London
Shirley Ho	Simons Foundation/Flatiron Institute, New York University & Princeton University
Arthur Loureiro	Oskar Klein Centre & ICIC
Vaso Pavlidou	University of Crete & Institute of Astrophysics/FORTH
Elena Sellentin	MI & STRW Leiden University
Huanyuan Shan	Shanghai Astronomical Observatory, CAS
Jean-Luc Starck	CEA Paris-Saclay
Roberto Trotta	SISSA and Imperial College London
George Tzagkarakis	Institute of Computer Science/FORTH
Licia Verde	ICREA & ICC-UB
Yanxia Zhang	National Astronomical Observatories, CAS

COSMO21 – Local Organizing Committee

George Tzagkarakis	Institute of Computer Science/FORTH
Samuel Farrens	CEA Paris-Saclay
Greg Tsagkatakis	University of Crete & Institute of Computer Science/FORTH
Panagiotis Tsakalides	University of Crete & Institute of Computer Science/FORTH
Andreas Tersenov	University of Crete & Institute of Computer Science/FORTH

Administrative Support

Maria Prevelianaki	Institute of Computer Science/FORTH
Marian Papadaki	Institute of Computer Science/FORTH

SPONSORS

Financial support for COSMO21 has been provided by the Institute of Computer Science – FORTH, CEA Paris-Saclay, Imperial College London, as well as the European Commission projects **TITAN** and **ARGOS**, and the ANR/PCR project **TOSCA**.

LIST OF PARTICIPANTS

Last name	First name	Institution	Country
Abramo	Raul	University of São Paulo	Brazil
Acharya	Anshuman	Max Planck Institute for Astrophysics	Germany
Adhikari	Susmita	IISER Pune	India
Akhaury	Utsav	EPFL	Switzerland
Aldridge	Henry	University College London	UK
Alsing	Justin	Oskar Klein Centre, Stockholm University	Sweden
Armijo	Joaquin	Kavli IPMU	Japan
Bayer	Adrian	Princeton University/Simons Foundation	USA
Berger	Sabrina	University of Melbourne	Australia
Bhuvanendran	Karthika	Newcastle University	UK
Bonjean	Victor	FORTH-ICS	Greece
Chang	Chihway	University of Chicago/KICP	USA
Charmandaris	Vassilis	FORTH-IA	Greece
Cheng	Sihao	Institute for Advanced Study	USA
Courbin	Frederic	EPFL & University of Barcelona	Switzerland
Craigie	Matt	University of Queensland	Australia
Desmond	Harry	ICG Portsmouth	UK
Doeser	Ludvig	Stockholm University	Sweden
Dupac	Xavier	ESA	Spain
Esteban Gutierrez	Ana	University of Valparaiso	Chile
Farrens	Samuel	CosmoStat, CEA Paris-Saclay	France
Gkogkou	Athanasia	FORTH-IA	Greece
Gorce	Adélie	IAS, Université Paris-Saclay	France
Goh	Lisa	CEA Paris-Saclay	France
Gorbatchev	Pauline	University of Crete & FORTH-IA	Greece
Grandón	Daniela	Leiden University	Netherlands
Grayling	Matt	University of Cambridge	UK
Harnois-Deraps	Joachim	Newcastle University	UK
Heavens	Alan	Imperial College London	UK
Ibitoye	Ayodeji	NAOC, CAS, Beijing	China
Jain	Bhuvnesh	U Penn	USA
Jeffrey	Niall	University College London	UK
Khorashadizadeh	AmirEhsan	University of Basel/Univ. College London	Switzerland
Kist	Timo	Leiden Observatory	Netherlands
Korkidis	George	FORTH-IA	Greece
Kovačić	Inja	Ghent University	Belgium
Lahiry	Arnab	FORTH-ICS, University of Crete	Greece

Leclercq	Florent	Institut d'Astrophysique de Paris	France
Li	Xing	Imperial College London	UK
Li	Nan	National Astronomical Observatories	China
Li	Changhua	National Astronomical Observatories	China
Liaudat	Tobías	CEA Paris-Saclay	France
Lin	Kiyam	University College London	UK
Liu	Jia	Kavli IPMU	Japan
Loureiro	Arthur	Oskar Klein Centre, Stockholm University	Sweden
Lucie-Smith	Luisa	Max Planck for Astrophysics	Germany
Makinen	Lucas	Imperial College London	UK
Mastrogiovanni	Simone	Italian Institute for Nuclear Physics	Italy
Mazde	Kratika	Institut d'Astrophysique de Paris, Sorbonne Université	France
McEwen	Jason	University College London	UK
Mittal	Shree Hari	University of Hertfordshire	UK
Moretti	Chiara	SISSA	Italy
Moskowitz	Irene	Rutgers University	USA
Mulder	Kevin	University College London	UK
Nazli	Kutay	Leiden University	Netherlands
Oehl	Veronika	ETH Zurich	Switzerland
Panayidou	Klea	European University Cyprus	Cyprus
Pawlosky	Annalisa	Google	Switzerland
Perez	Lucia	Flatiron Institute	USA
Petrakis	Manolis	Technical University of Crete	Greece
Piras	Davide	University of Geneva	Switzerland
Polanska	Alicja	University College London	UK
Porqueres	Natalia	University of Oxford	UK
Pourtsidou	Alkistis	University of Edinburgh	UK
Prat Marti	Judit	Nordita, Stockholm University	Sweden
Price	Matthew	University College London	UK
R. S.	Pranjal	Univeristy of Arizona	USA
Robertson	Naomi	University of Edinburgh	UK
Rogers	Keir	University of Toronto	Canada
Sarma Boruah	Supranta	University of Pennsylvania	USA
Seher Gandhi	Suroor	New York University	USA
Sellentin	Elena	University of Leiden	Netherlands
Shan	Huanyuan	SHAO, CAS	China
Simon Onfroy	Hugo	CEA Paris-Saclay	France
Spinelli	Marta	OCA	France

Starck	Jean-Luc	CosmoStat, CEA Paris-Saclay	France
Su	Chen	SHAO, CAS	China
Tersenov	Andreas	University of Crete & FORTH-ICS	Greece
Thummel	Linus	University of Edinburgh	UK
Tinnaneri Sreekanth	Vilasini	CEA Paris-Saclay	France
Trotta	Roberto	SISSA & Imperial College London	Italy
Truttero	Ottavia	University of Edinburgh	UK
Tsagkatakis	Gregory	University of Crete & FORTH-ICS	Greece
Tsaprazi	Eleni	Imperial College London	UK
Tsedrik	Maria	University of Edinburgh	UK
Tsouros	Alexandros	University of Crete & FORTH-IA	Greece
Tzagkarakis	George	FORTH-ICS	Greece
Valogiannis	Georgios	University of Chicago	USA
Vitorelli	André	NASA/JPL	USA
von Wietersheim-Kramsta	Maximilian	University College London	UK
Wang	Bonny	Carnegie Mellon University	USA
Whitney	Jessica	University College London	UK
Williamson	Joshua	University College London	UK
Xie	Yushan	SHAO, CAS	China
Zhang	Yanxia	National Astronomical Observatories	China

Monday - May 20th, 2024

9:00 AM → 12:15 PM **Radio Cosmology: Chair Elena Sellentin**

9:00 AM	**Registration**	45m
9:45 AM	**Introduction by Alan Heavens, George Tzagkarakis and Jean-Luc Starck**	15m
10:00 AM	**Cosmology with the SKA Observatory** **Speaker:** Marta Spinelli	45m
10:45 AM	**Short break**	10m
10:55 AM	**Synergies between radio intensity mapping and optical galaxy surveys** **Speaker:** Alkistis Pourtsidou	20m
11:15 AM	**Machine Learning for recovering the EoR 21-cm signal power spectrum at z~9.1 from LOFAR data** **Speaker:** Anshuman Acharya	20m
11:35 AM	**Learning Reionization History with Quasar IGM Damping Wings** **Speaker:** Timo Kist	20m
11:55 AM	**Using gradient descent-like methods to reconstruct the matter density field during reionization** **Speaker:** Sabrina Berger	20m

3:00 PM → 4:05 PM **Bayesian Methodology: Chair Marta Spinelli**

3:00 PM	**Our Universe in Simulation** **Speaker:** Jia Liu	45m
3:45 PM	**Towards robust Bayesian inference using physics informed priors from cosmological simulations** **Speaker:** Ludvig Doeser	20m

4:05 PM → 4:25 PM **Coffee break** — 20m

4:25 PM → 5:55 PM **Poster Session** — 1h 30m

5:55 PM → 6:55 PM **Gravitational Waves: Chair Alan Heavens**

5:55 PM	**Gravitational-wave cosmology: current results and statistical challenges ahead** **Speaker:** Simone Mastrogiovanni	45m

8:00 PM → 10:00 PM **Cocktail** — 2h

Tuesday - May 21st, 2024

9:00 AM → 10:25 AM **Weak Lensing IV: I: Chair Jia Liu**

9:00 AM **Higher Order Statistics vs. Deep learning for lensing cosmology** 45m
Speaker: Bhuvnesh Jain

9:45 AM **Cosmic shear analysis of KiDS and DES with higher-order statistics** 20m
Speaker: Joachim Harnois-Deraps

10:05 AM **Baryons and non-Gaussian statistics in weak lensing cosmology** 20m
Speaker: Daniela Grandon Silva

10:25 AM → 10:55 AM **Coffee Break** 30m

10:55 AM → 12:15 PM **Weak Lensing IV: II: Chair Justin Alsing**

10:55 AM **Dark Energy Survey Year 3 results: likelihood-free, simulation-based wCDM inference with neural compression of weak-lensing map statistics** 20m
Speaker: Niall Jeffrey

11:15 AM **Simulation based inference with contrastive learning and persistent homology** 20m
Speaker: Judit Prat

11:35 AM **KiDS-SBI: Simulation-Based Inference Analysis of KiDS-1000 Cosmic Shear** 20m
Speaker: Maximilian von Wietersheim-Kramsta

11:55 AM **Lossless Neural Compression for Weak Lensing Implicit Inference** 20m
Speaker: T. Lucas Makinen

3:00 PM → 4:25 PM **Artificial Intelligence I: Chair Georgios Tzagkarakis**

3:00 PM **Good Data Analysis Practices for AI** 45m
Speaker: Annalisa Pawlosky

3:45 PM **Scientific machine learning in cosmology** 20m
Speaker: Jason McEwen

4:05 PM **Don't trust neural networks? Me neither, but here's how I use them anyway** 20m
Speaker: Florent Leclercq

4:25 PM → 4:55 PM **Break** 30m

4:55 PM → 5:55 PM **Artificial Intelligence II: Chair Sam Farrens**

4:55 PM **Neural Networks as Classifiers of Cosmological Models** 20m
Speaker: Lisa Goh

5:15 PM **Exhaustive Symbolic Regression: Learning Astrophysics directly from Data** 20m
Speaker: Harry Desmond

5:35 PM **One Model to Handle Them All: A versatile framework for galaxy vision tasks by Implanting Human-in-the-loop upon a Large Vision Model** 20m
Speaker: Nan Li

Wednesday - May 22nd, 2024

9:00 AM → 10:25 AM **Galaxies & Redshifts: Chair Vassilis Charmandaris**

9:00 AM **A comprehensive picture of the galaxy population with simulation-based inference** 45m
Speaker: Justin Alsing

9:45 AM **Machine Learning Pipeline for speeding up SED fitting with Radiative Transfer Models** 20m
Speaker: Klea Panayidou

10:05 AM **A Comparative Analysis of Regression Methods for Spectroscopic Redshift Estimation** 20m
Speaker: Greg Tsagkatakis

10:25 AM → 10:55 AM **Coffee break** 30m

10:55 AM → 1:00 PM **Field Level Inference: Chair Susmita Adhikari**

10:55 AM **Field-level inferences of galaxy clustering** 20m
Speaker: Eleni Tsaprazi

11:15 AM **Almanac: Generic Field Level Inference for Full-Sky Cosmological Fields and Angular Power Spectra** 20m
Speaker: Arthur Loureiro

11:35 AM **Weak lensing field -level inference with KaRMMa** 20m
Speaker: Supranta Sarma Boruah

11:55 AM **Extracting optimal and robust information from cosmological surveys with field-level inference and joint analyses** 20m
Speaker: Adrian Bayer

3:00 PM → 4:25 PM **Weak Lensing III: Chair Arthur Loureiro**

3:00 PM **How to quantify textures and random fields in cosmology?** 45m
Speaker: Sihao Cheng

3:45 PM **Pseudaria: Map-Level Inference of Unknown Systematics in Galaxy Number Counts** 20m
Speaker: André Zamorano Vitorelli

4:05 PM **Mass-Mapping with Conditional Generative Adversarial Networks** 20m
Speaker: Jessica Whitney

4:25 PM → 4:55 PM **Coffee break** 30m

4:55 PM → 5:55 PM **Bayesian Methodology: : Chair Jason Mc Ewen**

- 4:55 PM **Scalable Bayesian uncertainty quantification with learned convex regularisers** 20m
  Speaker: Tobias Liaudat
- 5:15 PM **GPU-accelerated Hierarchical Bayesian Inference of Intrinsic and Extrinsic Properties of Type Ia Supernovae** 20m
  Speaker: Matthew Grayling
- 5:35 PM **Learned harmonic mean with normalizing flows for Bayesian model comparison in cosmology** 20m
  Speaker: Alicia Polanska

8:00 PM → 11:00 PM **Conference Diner** 3h

Thursday - May 23rd, 2024

9:30 AM → 10:55 AM **Machine Learning II: Chair Klea Panayidou**

- 9:30 AM — **Machine learning cosmological structure formation** — 45m
  Speaker: Luisa Lucie-Smith
- 10:15 AM — **CAMELS-SAM: untangling the galaxy-halo connection with machine learning in observational space** — 20m
  Speaker: Lucia Perez
- 10:35 AM — **Unveiling the connections between halos and galaxies with Machine Learning** — 20m
  Speaker: Raul Abramo

10:55 AM → 11:25 AM **Coffee break** — 30m

11:25 AM → 12:25 PM **Machine Learning II: Chair Yanxia Zhang**

- 11:25 AM — **Using graph neural networks to detect dark matter in stellar streams** — 20m
  Speaker: Keir Rogers
- 11:45 AM — **A representation learning approach to probe for dynamical dark energy in matter power spectra** — 20m
  Speaker: Davide Piras
- 12:05 PM — **Unsupervised Searches for Parity-Violation** — 20m
  Speaker: Matt Craigie

3:00 PM → 4:25 PM **Cosmic Fields: Chair Frederic Courbin**

- 3:00 PM — **Cosmic fields beyond 2pt: practical challenges opportunities** — 45m
  Speaker: Chihway Chang
- 3:45 PM — **Theoretical wavelet l1-norm from one-point PDF prediction** — 20m
  Speaker: Vilasini Tinnaneri Sreekanth
- 4:05 PM — **An exact likelihood for the weak lensing correlation function** — 20m
  Speaker: Veronika Oehl

4:25 PM → 4:55 PM **Coffee Break** — 30m

4:55 PM → 5:55 PM **Weak Lensing IV: Chair Chihway Chang**

- 4:55 PM — **The interplay between analysis choices in cosmic shear** — 20m
  Speaker: Naomi Robertson
- 5:15 PM — **DeepLensing Flow: Score based and flow models for weak lensing statistics** — 20m
  Speaker: Joaquin Armijo
- 5:35 PM — **Mitigating 3x2pt Systematics: Augmented Photo-z Training Samples and Optimal Tomographic Binning** — 20m
  Speaker: Irene Moskowitz

Friday - May 24th, 2024

9:00 AM → 11:05 AM **Cosmology: Chair Yanxia Zhang**

9:00 AM	**The strong lensing contribution to the H0 tension** **Speaker**: Frederic Courbin	45m
9:45 AM	**Abundance of Primordial Black Holes from a Bayesian inference analysis between quasar microlensing observations and simulated magnification maps** **Speaker**: Ana Esteban Gutiérrez	20m
10:05 AM	**Understanding the distribution of dark matter in halos from cluster to dwarf scales** **Speaker**: Susmita Adhikari	20m
10:25 AM	**MLPatch: Learned local patch-based models for high-dimensional cosmological inverse problems** **Speaker**: Amir Ehsan Khorashadizadeh	20m
10:45 AM	**Joint Multi-channel Deconvolution for Euclid and LSST Images** **Speaker**: Utsav Akhaury	20m

11:05 AM → 11:35 AM **Coffee Break** 30m

11:35 AM → 12:15 PM **Panel Discussion: Moderator Elena Sellentin** 40m

12:15 PM → 12:35 PM **Conclusions** 20m

A comprehensive picture of the galaxy population with simulation-based inference

Justin Alsing

Characterizing the evolution of galaxy demographics over cosmic history is one of the key challenges in extragalactic astronomy, being of central importance for both understanding galaxy evolution and for cosmological analyses of galaxy surveys (for example, for accurate photometric redshift calibration). However, accurately and comprehensive characterization of the galaxy population from photometric and/or spectroscopic survey data has remained an intractable task for a number of reasons: the stellar population synthesis (SPS) models connecting galaxy properties to their observables (SEDs) have been too slow to run at scale; the distribution of galaxy characteristics is too complex to be captured by parametric modeling approaches; selection effects and complex photometric noise and calibration effects make the associated (hierarchical) inference task challenging or intractable. I present pop-cosmos — a new framework for constraining the galaxy population from large surveys, applied to COSMOS data, to obtain the most comprehensive picture of the galaxy population out to redshift 4 ever obtained. Leveraging advances in simulation-based inference (SBI), generative ML, and ML emulation techniques, we overcome the previous challenges to define a new gold-standard in large-scale galaxy survey data analysis.

Cosmic fields beyond 2pt: Practical challenges opportunities

Chihway Chang

In this talk I will first give an overview of how information about cosmology and astrophysics is imprinted in the cosmic fields sourced by the large-scale structure, and a schematic way to think about how we use different statistics to extracting that information. I will then focus on the weak lensing field and present a couple examples where we implement these statistics and describe the practical challenges as well as opportunities.

How to quantify textures and random fields in cosmology?

Sihao Cheng

Extracting information from stochastic fields is a ubiquitous task in science. However, from cosmology to biology, it tends to be done either through a power spectrum analysis, which is often too limited, or through the use of convolutional neural networks, which require large training sets and lack interpretability. I will present a powerful statistical tool called the “scattering transform” and its recent updates, which stand nicely between the two extremes. These techniques combine the idea of wavelet scale separation with simple nonlinear transform. I will use various examples in cosmology and beyond, including its application to HSC weak lensing data, to demonstrate its efficiency, interpretability, and advantage over traditional statistics.

Higher Order Statistics vs Deep learning for lensing cosmology

Bhuvnesh Jain

Weak lensing mass maps are well suited for the application of a variety of non-Gaussian/Higher Order Statistics (HOS). Deep learning via CNNs and ViTs has also been applied. Systematic uncertainties of various kinds can mess up any of these approaches! How can we meaningfully compare their constraining power, given the range of systematics that affect different surveys in different ways? I will address this question with some worked examples from the Dark Energy Survey and simulations. I will also discuss some of the currently available methods for ‘interpretability’ for cosmological maps.

Our Universe in Simulation

Jia Liu

We are expecting high-precision observations from upcoming CMB surveys, such as the Simons Observatory, CMB-S4, and LiteBIRD, as well as from surveys of the large-scale structure, such as Rubin LSST, Euclid, DESI, PSF, SPHEREx, and Roman. Most of the observables from these independent surveys will be correlated due to their sky and redshift overlaps. Joint analysis of these surveys will be key to transformative discoveries.

Machine learning cosmological structure formation

Luisa Lucie-Smith

I will present an explainable deep learning framework for extracting new knowledge about the underlying physics of cosmological structure formation. I will focus on an application to dark matter halos, which form the building blocks of the cosmic large-scale structure and wherein galaxy formation takes place. The goal is to use an interpretable neural network to generate a compressed, “latent” representation of the data, which encodes all the relevant information about the final output of interest; the latent representation can then be interpreted using mutual information. I will show how such networks can be used to model final emergent properties of dark matter halos, such as their density profiles, and connect them to the physics that determines those properties. The results illustrate the potential for machine-assisted scientific discovery in cosmological structure formation and beyond.

Gravitational-wave cosmology: current results and statistical challenges ahead

Simone Mastrogiovanni

For over 20 years, measurements of the Universe expansion rate from close-by and far sources are in tension hinting at the presence of new physics. Gravitational Waves (GWs) from compact binary coalescences (CBCs) are emergent cosmological probes, potentially observable from close to far scales. GWs are Standard Sirens as they are the only source for which it is possible to measure the distance. In this talk, I will discuss how populations of dark sirens, namely GW sources observed without electromagnetic counterparts, can be exploited to measure cosmic expansion. I will mainly discuss two methodologies that can assign a redshift, using hierarchical Bayesian inference, to GW events either thanks to galaxy surveys and the source mass spectrum of binary black holes. I will focus on possible pitfalls of the two methodologies that could potentially introduce a systematic bias. Finally, I will present the latest results on the measurement of the cosmic expansion rate from the dark sirens reported in the latest Gravitational Wave Transient catalogue and the computational and statistical challenges expected for the future.

Good data analysis practices for AI

Annalisa Pawlosky

The potential number of problems, and wrong turns to be made, seems to scale with the size of datasets. Large data sets offer numerous ways of looking at data, opening up pathways for including AI. While AI can offer new approaches to evaluate datasets and generate predictions, it is essential that good data analysis approaches are used to produce credible insights. Good data analysis approaches for large datasets will be discussed, alongside mistakes made and lessons learned. Additionally, strategies for model evaluation will be covered. Finally, a brief discussion on considerations for Quantum Computing will help lay down the foundation for how to approach the evaluation of datasets beyond classical capabilities.

Cosmology with the SKA Observatory

Marta Spinelli

The SKA Observatory is an international project to build a next-generation radio facility comprising two telescopes: a dish array (SKA-Mid) in South Africa, and an array of dipole antennas (SKA-Low) in Western Australia. In this talk, I will briefly review the status of the science cases of interest for the cosmology community. I will describe in particular the measurement of the distribution of cosmic neutral hydrogen via its 21cm line using Intensity Mapping. I will discuss how such a technique could provide a measurement of the underlying large-scale structure of the Universe and its evolution, and potentially unveil the nature of dark matter and dark energy. I will present the current data-taking campaign of the SKAO precursor MeerKAT and discuss the challenges that we need to overcome to fully exploit this new incredible window into the evolution of the Universe.

Unveiling the connections between halos and galaxies with Machine Learning

Raul Abramo

Large-scale structures can be successfully described in terms of analytical or semi-analytical models only down to mildly non-linear scales. Structure formation in the non-linear regime, including the emergence of observable tracers of those structures such as galaxies, voids, and quasars, are quite outside the scope of numerical solvers. Together with hydrodynamical simulations, Machine Learning offers non-parametric tools to mimic, and therefore predict, the intricate relationships between the cosmic web and some of those tracers. I will show how it is possible to reproduce with high accuracy the mapping between the multivariate distributions of halo properties and the distribution of galaxy properties in those simulations. This allows us to predict which types of galaxies can be found in different types of halos, paving the way to high-fidelity galaxy mocks based on dark-matter-only N-body simulations. I will also discuss the role of stochasticity in the emergence of the tracers of large-scale structure. Results based on the papers: arXiv:2301.06398, arXiv:2201.06054, and Rodrigues et al. 2024 (to appear).

Detecting the epoch of reionization 21-cm signal: A machine learning upgrade to GPR

Anshuman Acharya

The LOFAR Epoch of Reionisation (EoR) working group has been using Gaussian Process Regression (GPR) for foreground subtraction from data, to constrain the EoR 21-cm signal power spectrum. However, recent work has noted that there can be significant amounts of signal loss if the EoR covariance is misestimated. To have better covariance models, it has been proposed to use Machine Learning trained models, by using training sets from a large variety of N-body + 1D radiative transfer simulations (GRIZZLY). I will thus first talk about the results from Acharya et al. (2024) which show the limits of improvement provided by ML by testing against a variety of mock signals injected in simulated data. Secondly, I will talk about the results of applying this model to 10 nights of observed data at redshift z ~ 9.1, and compare results with upper limits obtained on the same data in Mertens et al. (2020). Lastly, I will talk about future plans including the building of a redshift agnostic model for the 21-cm signal.

Understanding the substructure within halos

Susmita Adhikari

Dark Matter halos are collapsed, self-gravitating objects. Their dynamical structure contains information about the growth history of the halo. I will describe how using machine learning and clustering techniques we can separate the different dynamical regions of the halo and understand substructure within it.

DeepLensing Flow: Score based and flow models for weak lensing statistics

Joaquin Armijo

In the past few years, deep learning models have been used to model the statistics of the matter density field at the level of the power spectrum. More recently, generative methods can replicate the features of the Large-scale structure, including the non-linear scales. Normalizing flow and diffusion models are powerful tools for this task, as they generate new information from random Gaussian noise keeping the original dimensionality of the original data. I will show results of these models for generating weak lensing convergence maps and the results of their non-Gaussian statistics. We focus on how well the mean values of power spectrum, PDF and Minkowski functions are recovered, but also at what level the covariance of these statistics can be reproduced.

Extracting optimal and robust information from cosmological surveys with field-level inference and joint analyses

Adrian Bayer

Extracting maximal information from upcoming cosmological surveys is a pressing task on the journey to understanding phenomena such as neutrino mass, dark energy, and inflation. This can be achieved by both making advancements in methodology and by carefully combining multiple cosmological datasets. In the first part of my talk, I will discuss methodological advancements to obtain optimal constraints, focusing on field-level inference with differentiable forward modeling. I will first motivate this approach to both reconstruct the initial conditions of the Universe and to obtain cosmological constraints. I will then tackle one of the bottlenecks of this approach -- sampling a high-dimensional parameter space -- by presenting a novel method, Microcanonical Langevin Monte Carlo. This method is orders of magnitude more efficient than the traditional Hamiltonian Monte Carlo and will enable scaling field-level inference to the regime of upcoming surveys.

I will then discuss combining multiple cosmological datasets to break parameter degeneracies and calibrate systematics. In particular, I will present the HalfDome cosmological simulations, a set of large-volume simulations designed specifically to model the Universe from CMB to LSS for the joint analysis of Stage IV surveys. I will show how these simulations are being used to mitigate systematics, obtain tighter constraints on cosmological parameters, and as a playground for machine learning applications. Time permitting, I will also present ongoing work on agnostically mitigating systematics using machine learning.

The talk will be based on 2307.09504, 2210.15649, and ongoing work. The HalfDome website can be found at https://halfdomesims.github.io/

Using gradient descent-like methods to reconstruct the matter density field during reionization

Sabrina Berger

As the first stars, quasars, and galaxies formed, their light ignited the universe by ionizing neutral hydrogen in the intergalactic medium during a period called reionization. A multitude of intensity mapping experiments are currently underway to detect 21-cm neutral hydrogen emission during reionization. This will enable an unparalleled understanding of high redshift galaxy formation and evolution through the only direct probe of matter in the early universe. Although no detection has been made, the tools necessary to map the evolving density field of the universe from a brightness temperature measurement must be in place. We propose a new Bayesian method to convert a measured brightness temperature map to its corresponding density field. Using gradient descent-like methods, we determine the maximum a posteriori (MAP) solution for this conversion near the midpoint of reionization. Our model successfully allows us to map the density field even in the ionized bubbles where the brightness temperature measurement is zero. This will allow the next generation of precision 21-cm intensity mapping experiments such as SKA-low to gain an immediate glimpse into the distribution of matter at high redshift.

Strong lensing as a way to address the Hubble tension: stellar dynamic le lenses vs. hierarchical Bayesian analysis

Frederic Courbin

I will show how strong lensing of quasars by galaxies can be used to measure the Hubble constant and what are the limitations of the method which is a single-step (no ladder) method fully independent of any other method in use. I will also show the limitation of the method and the mitigation strategies to minimize systematics. Strong lensing supports the value of the Hubble constant inferred from the distance ladder.

Is the Universe the same as its mirror image? Unsupervised searches for cosmological parity violation

Matt Craigie

Recent measurements of parity-odd modes in the spectroscopic 4-point correlation function and cosmic bifringence in the CMB provide tantalizing evidence for parity violations in cosmological fields. This could point to physics beyond the standard model. Traditional searches for parity-violations in large scale structure are limited to low-order N-point functions and theoretical uncertainties in the covariance of the 4-point function. In this talk I present an unsupervised approach which is immune to these difficulties. Conceptually this is achieved by subdividing the observed cosmological field into sub-volumes and feeding siamese pairs of the original and parity-flipped sub-volumes to a neural network which learns to classify the images as real or parity-flipped. If the network can reliably make an accurate classification, parity-violation is present in the data.

Exhaustive Symbolic Regression: Learning astrophysics directly from data

Harry Desmond

Machine learning is the new frontier in data-driven science. A key challenge is to make machine-assisted discovery interpretable, enabling it not only to uncover correlations but enhance our physical understanding of the world. A nascent branch of machine learning -- Symbolic Regression (SR) - aims to discover the optimal functional representations of datasets, producing perfectly interpretable outputs (equations) by construction. While SR is traditionally done using a “genetic algorithm” which stochastically generates trial functions through an analogue of natural selection, I will describe a more ambitious approach that exhaustively searches and evaluates function space.

Coupled to an information-theoretic model selection principle based on minimum description length, Exhaustive Symbolic Regression is guaranteed to find the simple functions that optimally balance simplicity with accuracy on any dataset, giving it broad application across science. I will describe the method and use it to quantify the extent to which state-of-the-art astrophysical theories - FLRW cosmology, General Relativity and Inflation - are implied by the current data.

Towards robust Bayesian inference using physics informed priors from cosmological simulations

Ludvig Doeser

Next-generation surveys such as DESI, Euclid, LSST, SPHEREx, and SKA will push cosmology into a data-driven era with their unprecedented survey volume and instrument sensitivity. The effectiveness of how much we can leverage the informational wealth and statistical power contained in the observational data will greatly depend on the accuracy of our data models. To this end, using cosmological simulations is established as a principal method to test and validate a given inference pipeline. Furthermore, simulations are also leveraged to distil insights into the properties of the expected survey data. However, these physical intuitions gained from simulations are often empirically quantified and not directly connected to the final model parameters of interest. Therefore, more conservative prior choices are still necessary. In this talk, we will introduce a fully Bayesian formalism on how to distil implicit priors for a given model by leveraging existing suites of cosmological simulations. Our method is model-agnostic and can improve parameter constraints without introducing biases.

Abundance of PBHs from a Bayesian inference analysis of quasar microlensing effect in lens galaxies

Ana Esteban Gutierrez

The quasar microlensing effect allows us to estimate the mass and abundance of any kind of compact object in the lens galaxies, being sensitive to both, baryonic and dark matter (DM) distributions. Recent results of black hole (BH) mergers coming from the Gravitational Wave (GW) experiments LIGO/Virgo raised again the interest in studying other astrophysical candidates to be part of the elusive DM nature (specifically, the intermediate-mass BHs). In particular, the favorite candidate among the community is the so-called Primordial Black Holes (PBHs), which are extensively studied by various groups to explain the observed amount of DM in the Universe. Thus, constraining the abundance of this kind of compact object in galaxies could give us clues about the DM content or even about the possibility of PBHs as seeds for the formation of SMBHs. In this talk, I present the results of using optical and X-ray quasar microlensing observations of differential magnitude measurements of image pairs for each strong lensed system and applying the corresponding microlensing magnification statistics with a Bayesian analysis to simulations of a mixed population of Primordial Black Holes (PBHs) and stars (typically the ones taken into account for the microlensing effect).

For comparison, we also performed different tests including the change in the source size involved in the convolution of the maps and a limited bootstrapping analysis to obtain an average standard deviation of the method applied to the selected parameters of each lens system. Finally, the mass range explored in this work, ranging from planetary to intermediate-mass PBHs, provides new upper limits for the estimated fraction of the DM in form of PBHs and sets the strongest bounds in this mass range according to other microlensing studies.

Neural networks as classifiers of cosmological models

Lisa Goh

We build a neural network (NN) that is capable of differentiating between a LCDM and a tomographic Coupled Dark Energy model, based on datasets of fs8. We simulate mock data of the product of the growth factor and sigma8, fs8, in both a LCDM and a Coupled Dark Energy model, adding noise to mimic a Stage IV-like survey. We see that our NN can well distinguish between the two models, regardless of what redshift coupling is activated.

Baryons and non-Gaussian statistics in weak lensing cosmology

Daniela Grandón

The non-Gaussian (NG) statistics comprise a set of statistical tools that are designed to capture the non-Gaussianities encoded in the cosmic density field. In weak lensing cosmology, they are implemented to unveil complementary information to the two-point statistics, and to calibrate systematic effects such as baryons, photo-z, etc. In this talk, I will introduce weak lensing non-Gaussian statistics, their modeling and the challenge of training accurate emulators for Stage-III and Stage-IV galaxy surveys. My talk will focus on the impact of baryonic feedback on these various summary statistics, and how to combine them to obtain tighter and unbiased cosmological constraints. I will also show preliminary results on a Bayesian method for unveiling the impact of baryons on weak lensing NG statistics, that takes into account the uncertainty we have on the modeling of baryons, and averages over the various baryonic feedback models.

Scalable hierarchical Bayesian inference of intrinsic and extrinsic properties of type Ia supernovae

Matt Grayling

Type Ia supernovae have played a key role within cosmology, used to infer both the properties of dark energy and the local expansion rate of the universe; there exists a tension between measurements of this latter parameter as inferred from supernovae and from the Cosmic Microwave Background. However, much remains unknown about the exact physical nature of these events. A key open question is the cause of the 'mass-step', whereby SNe Ia in massive galaxies are brighter than those in lower mass hosts post-standardization. Understanding the cause of this effect will further our knowledge of the progenitor systems of these events and reduce the scatter of supernova distances used to probe cosmological parameters. Recent discussion has centered around whether the mass-step is caused by extrinsic effects such as host galaxy dust or intrinsic, environmentally dependent differences across the SN Ia population. We will discuss what we can infer about the relative contributions of these two effects with hierarchical Bayesian modelling, using a new, GPU-accelerated implementation of the SN Ia SED model BayeSN. Hierarchical modelling enables BayeSN to model a physically-motivated distribution of dust laws and intrinsic variation as separate effects at the population level, providing better constraint of the dust properties of SN Ia hosts. We will present a combined analysis of SNe Ia in Foundation, DES3YR and PS1MD, and comment on the how the results obtained can be influenced by modelling choices.

Cosmic shear analysis of KiDS and DES with higher-order statistics

Joachim Harnois-Deraps

Cosmic shear galaxy surveys are now achieving high precision in their measurements of cosmological parameters that describe the dark sector of our Universe, competing with the precision of CMB experiments such as Planck. In particular, novel lensing statistics have been shown to be significantly more robust to an array of systematic uncertainties, offering compelling alternatives to the mainstream methods based on Gaussian statistics. In this talk, I will present a joint analysis of the Kilo Degree Survey and the Dark Energy Survey based on peak count statistics, highlighting some of the key interplay between baryonic feedback, intrinsic alignments and parameters of the dark sector. I will finally discuss some of the potential and challenges we will face when analyzing the upcoming generation of data from LSST and Euclid.

Dark Energy Survey Year 3 results: likelihood-free, simulation-based wCDM inference with neural compression of weak-lensing map statistics

Niall Jeffrey

I will present the simulation-based cosmological wCDM inference using Dark Energy Survey Year 3 weak-lensing maps, via neural data compression of weak-lensing map summary statistics: power spectra, peak counts, and direct map-level compression/inference with convolutional neural networks (CNN). Using simulation-based inference, we use forward-modelled mock data to estimate posterior probability distributions of unknown parameters. This approach allows all statistical assumptions and uncertainties to be propagated through the forward-modelled mock data; these include: sky masks, non-Gaussian shape noise, shape measurement bias, source galaxy clustering, photometric redshift uncertainty, intrinsic galaxy alignments, non-Gaussian density fields, neutrinos, and non-linear summary statistics. We include a series of tests to validate our inference results – this is a key feature of this analysis. For wCDM inference our most constraining result uses power spectra combined with map-level (CNN) inference, leading to more than a factor 2 improvement in our inference of dark energy parameters.

MLPatch: Learned local patch-based models for high-dimensional cosmological inverse problems

AmirEhsan Khorashadizadeh

Convolutional neural networks (CNNs) have become a standard approach for solving inverse problems in cosmology and beyond. Despite the satisfactory results on `in-distribution' test data similar to the training data, these methods can overfit the 'global' structure of the input image leading to poor generalization on out-of-distribution data. This issue is particularly pertinent in cosmology where we often do not have a great deal of training data. A wide range of imaging inverse problems, however, have a localized relationship between clean and degraded images like image denoising and even scientific imaging problems like mass mapping. In this work, we propose a `local' processing network termed MLPatch which recovers the image intensity at a given pixel by processing a small `neighborhood' of the pixel in the degraded image using a multi-layer perception (MLP). With comparable performance with CNNs on in-distribution test data, MLPatch can significantly outperform strong CNNs like U-Net on out-of-distribution data while maintaining a small memory usage. Finally, we generalize this local processing model for solving general inverse problems by employing MLPatch as an efficient image denoiser in the plug-and-play framework. Our experiments on a variety of imaging inverse problems, including radio interferometric imaging and mass mapping, show that MLPatch can reconstruct images with high accuracy at high resolution and with efficient memory usage.

Inferring reionization history from quasar IGM damping wings

Timo Kist

Constraining the Epoch of Reionization remains one of the pivotal tasks of modern cosmology, and next-generation telescopes are opening up the path to the first precision constraints on the timing of reionization derived from the Ly-alpha damping wing signature imprinted on the spectra of high-redshift quasars by the foreground neutral intergalactic medium (IGM). In the coming years, EUCLID will detect a number of high-redshift quasars never seen before, whose exquisite spectra collected by JWST are calling for powerful statistical methods to infer precision constraints on the IGM neutral hydrogen fraction as a function of redshift. We developed a state-of-the-art Bayesian inference pipeline that allows us to disentangle the IGM damping wing from a quasar's unknown intrinsic spectrum and infer its lifetime as well as the neutral hydrogen (HI) column density in front it, directly translating into a constraint on the global IGM neutral fraction. We account for covariances across the full spectral range caused by IGM transmission fluctuations, quasar continuum reconstruction, and spectral noise. We show how simulation-based inference enables us to overcome non-Gaussianities in the IGM transmission likelihood by training a normalizing flow as neural likelihood estimator as well as a binary classifier as likelihood ratio estimator. These developments are facilitated by our fully differentiable JAX-based inference pipeline, exploiting the latest ML infrastructure. After marginalizing out nuisance parameters associated with the quasar continuum, we find that we can constrain the HI column density within the first 15 pMpc of each individual quasar to 0.2 - 0.9 dex and its lifetime to 0.2 - 1.0 dex. By applying our procedure to a set of mock observational spectra resembling the distribution of EUCLID quasars with realistic spectral noise, we show that our method applied to upcoming observational data can robustly constrain the evolution of the IGM neutral fraction at the $< 5\%$ level at all redshifts between $6 < z < 11$.

COmoving Computer Acceleration: Don't trust neural networks? Me neither, but here’s how I use them anyway

Florent Leclercq

Interpretability and accuracy are pivotal challenges in the application of machine learning to cosmology. If machines find something humans don't understand, how can we check (and trust) the results? In this presentation, I contend that addressing this concern is not always obligatory, whether machine learning is used to build a posterior approximator or an emulator of an expensive model. I will elucidate this argument through two case studies where the use of neural networks is safe *by construction*. My first example will involve information maximizing neural networks to autonomously define statistical summaries for implicit likelihood inference. My second example will involve neural networks to emulate a frame of reference (rather than the simulation output) in a COLA-like framework for N-body simulations of dark matter particles.

One Model to Handle Them All: A versatile framework for Astronomical Vision Tasks by Implanting Human-in-the-loop on a Large Vision Model

Nan Li

Large-scale astronomical surveys bring exponentially increased datasets with considerable complexity, leading to a challenging but must-solve problem: How to mine scientific information from such enormous datasets efficiently. Astronomers adopt deep learning approaches rapidly to extract physical parameters from photometric and spectroscopic data. However, the specific applications can only deal with particular problems defined by the training set. To avoid meaningless duplicate workloads, we build a framework for general analysis of galaxy images based upon a large vision model for downstream tasks, including classification, image restoration, object detection, parameter extraction, etc. Given the low signal-to-noise ratio in astronomical images and the highly imbalanced distribution of different celestial objects, we incorporate a Human-in-the-loop module onto the aforementioned large vision model. This involves human knowledge priors to enhance the reliability in processing astronomical images through interactive improvement. On the galaxy images from DESI legacy Surveys, the proposed scheme exhibits notable few-shot learning capabilities and versatile adaptability to various tasks. In conclusion, our framework can successfully process galaxy images for multiple downstream tasks, requiring few training sets thanks to the human-in-the-loop mechanism. In addition to images, multimodal data can also be integrated in a similar manner, which holds promise for conducting joint analyses with datasets spanning diverse domains in the era of multi-message astronomy.

Scalable Bayesian uncertainty quantification with learned convex regularisers for radio interferometric imaging

Tobías Liaudat

The last decade brought us substantial progress in computational imaging techniques for current and next-generation interferometric telescopes, such as the SKA. Imaging methods have exploited sparsity and more recent deep learning architectures with remarkable results. Despite good reconstruction quality, obtaining reliable uncertainty quantification (UQ) remains a common pitfall of most imaging methods. The UQ problem can be addressed by reformulating the inverse problem in the Bayesian framework. The posterior probability density function provides a comprehensive understanding of the uncertainties. However, computing the posterior in high-dimensional settings is an extremely challenging task. Posterior probabilities are often computed with sampling techniques, but these cannot yet cope with the high-dimensional settings from radio imaging.

This work proposes a method to address uncertainty quantification in radio-interferometric imaging with data-driven (learned) priors for very high-dimensional settings. Our model uses an analytic physically motivated model for the likelihood and exploits a data-driven prior learned from data. The proposed prior can encode complex information learned implicitly from training data and improves results from handcrafted priors (e.g., wavelet-based sparsity-promoting priors). We exploit recent advances in neural-network-based convex regularisers for the prior that allow us to ensure the log-concavity of the posterior while still being expressive. We leverage probability concentration phenomena of log-concave posterior functions that let us obtain information about the posterior avoiding the use of sampling techniques. Our method only requires the maximum-a-posteriori (MAP) estimation and evaluations of the likelihood and prior potentials. We rely on convex optimization methods to compute the MAP estimation, which are known to be much faster and better scale with dimension than sampling strategies. The proposed method allows us to compute local credible intervals, i.e., Bayesian error bars, and perform hypothesis testing of structure on the reconstructed image. We demonstrate our method by reconstructing simulated radio-interferometric images and carrying out fast and scalable uncertainty quantification.
Based on arXiv:2312.00125.

Almanac: Generic field level inference for full-sky cosmological fields and angular power spectra

Arthur Loureiro

With the advent of Euclid, LSST, Simons Observatory and other upcoming Stage-IV Surveys, we will soon map cosmic structure at an unprecedented precision over a large portion of the sky, such that sky curvature becomes influential. In this talk, I will present Almanac: a generic Bayesian solution for inferring the full-sky underlying cosmological fields and their power spectra from noisy partial sky observations. The crux to inferring these science-ready data products is to develop a Monte Carlo sampler that can handle the high resolution of upcoming data maps. A further challenge is that cosmological structures often have power spectra spanning many orders of magnitude. Thus, Almanac handles strongly scale-dependent signal-to-noise cases for spin-0 and spin-2 cosmological fields. This talk will show different applications of Almanac, focusing on the challenging applications to upcoming Weak Lensing data. Using a Euclid-like survey as a test study, we jointly infer all-sky E-mode and B-mode tomographic auto- and cross-power spectra from the masked sky and potentially parity-violating EB-mode power spectra. In this test, we probe scales up to a maximum multipole of 2048 – a Hamiltonian Monte Carlo with a total of ~16.8 million parameters. The main output and natural outcome is the set of samples of the posterior, which does not suffer from leakage of power from E to B unless reduced to point estimates.

Lossless neural compression for weak lensing implicit inference

Lucas Makinen

We present optimal neural compression techniques for weak lensing data. We show that massive catalogues and cosmological fields can be compressed to asymptotically lossless summary statistics, which can be used in implicit inference without having to assume their statistical properties. Training networks over simulations automatically generates Fisher matrices for cosmological parameter constraint forecasts at a fiducial model for the implicit likelihood, and the learned compression can then be used to construct posterior distributions with simulation-based inference. We show that these techniques are less computationally expensive than field-level Bayesian sampler and n-point correlation function inference and can learn nonlinear combinations of data features whose distributions are not known, but can be simulated. Moreover, we quantify how much more information neural techniques can extract beyond traditional summary statistics.

Scientific machine learning in cosmology

Jason McEwen

Machine learning (ML) is having a transformative impact on cosmology. The field is starting to mature, where we are moving beyond the naive application of off-the-shelf, black-box ML models towards approaches where ML is an integral component in a larger, principled analysis methodology. Furthermore, not only are cosmological analyses benefiting from the use of ML, but ML models themselves can be greatly enhanced by integrating knowledge of relevant physics. I will briefly review three maturing areas where ML and cosmology have already demonstrated some success, while still providing many further opportunities and challenges. (1) Physics-enhanced learning integrates knowledge of relevant physics into ML models, either through augmentation, encoding symmetries and invariances, encoding dynamics, or directly through physical models that are integrated into the ML model. (2) In statistical learning, ML and statistical models are tightly coupled to provide probabilistic frameworks, often in a Bayesian setting, that offer uncertainty quantification, generative models, accelerated inference, and data-driven priors. (3) For scientific analyses in particular, it is important that ML models are not opaque, black-boxes but are intelligible, ensuring truthfulness, explainability and interpretability. Throughout I will provide numerous examples of cosmology studies where such approaches have or are being developed and applied, in the context of upcoming observations from the Euclid satellite, the Rubin Observatory Legacy Survey of Space and Time (LSST), and the Square Kilometre Array (SKA).

Mitigating 3x2pt Systematics: Augmented photo-z training samples and optimal tomographic binning

Irene Moskowitz

Large imaging surveys use the 3x2pt method combining weak lensing and galaxy clustering to measure cosmological parameters. These surveys, including the Vera C. Rubin Observatory's Legacy Survey of Space and Time (LSST) rely on photometric redshifts, but spectroscopic training samples are biased towards brighter, redder, lower redshift galaxies. These non-representative training samples can lead to photo-z outlier fractions nearly 4 times larger than expected for a representative training sample. In Moskowitz et al. (in internal review), we show that augmenting the training sample with simulated galaxies possessing otherwise unrepresented features, we can reduce the outlier fraction of the photo-z estimates by nearly 50%, and the scatter by 56%. With these improved photo-z estimates, we then sort the galaxies into tomographic redshift bins that maximize the 3x2pt signal determined using the generalized binning parameterization introduced in Moskowitz et al. (2023, ApJ 950, 49). We use a neural network classifier to identify galaxies that are highly likely to be sorted into the correct redshift bin, which can improve the dark energy figure of merit by ~13%, equivalent to a 28% increase in data volume. With the improved photo-z estimates and binning choices, we investigate their effect on bias in the final cosmological parameter estimates using a 3x2pt analysis.

An exact likelihood for the weak lensing correlation function

Veronika Oehl

The two-point correlation function is a widely used summary statistic for cosmological fields. Within Bayesian parameter inference, the likelihood for this observable, particularly in weak lensing, is mostly assumed to be Gaussian. However, this is an approximation at best, as the distribution of functions that are quadratic in the Gaussian random variables is manifestly non-Gaussian. Since the exact shape of the likelihood determines the obtained posterior distribution of the model parameters, it is crucial to be able to quantify the possible influence of such simplifying assumptions on the final parameter constraints. This is particularly relevant in view of the on-going discrepancies in the amplitude of clustering inferred from high- and low-redshift probes: the S8 tension. On small enough scales, the likelihood becomes well-approximated by a multivariate Gaussian, due to the central limit theorem.

On large scales, where the assumption of the cosmological field to be Gaussian is most valid, this approximation, as well as the flat-sky approximation, are not valid anymore. The increased precision and angular scales accessed by stage-IV surveys will challenge these approximations, as even current stage-III surveys are seeing indications of non-Gaussianity in their correlation functions at large scales. We present a framework to calculate the exact likelihood for auto- and cross-correlation functions of Gaussian random fields, even in the presence of masks. We show that the resulting likelihoods contain significant levels of skewness such that the mode of the distribution is shifted towards lower values. The framework can be applied to spin-2 fields on the sphere, to in principle calculate n-dimensional likelihoods and is compared to simulated distributions as well as the standard Gaussian likelihoods of noisy weak lensing maps.

Machine learning pipeline for speeding up SED fitting with radiative transfer models

Klea Panayidou

The Spectral Energy Distribution (SED) of galaxies from the ultraviolet to the millimetre contains important information about the physics of the galaxies. The SEDs of galaxies are usually decomposed into a number of components that correspond to star formation regions in the galaxy, an active galactic nucleus (AGN) or the host galaxy. Currently the most popular codes for SED fitting are energy balance codes such as CIGALE and MAGPHYS mainly due to their ease of use. Additionally, more physically motivated methods like CYGNUS (CYprus models for Galaxies and their NUclear Spectra) which employ radiative transfer models have already yielded useful results on physical quantities such as stellar mass, star formation rate, starburst timescale, AGN fraction, dust masses etc. Computing the radiative transfer equations 'on the fly' during the fitting is very computing intensive and for this reason, SED fitting codes employ pre-computed libraries of radiative transfer models. Drawing upon an existing computational pipeline (used for exoplanet research) we couple a machine learning model with our MCMC SED fitting framework. We demonstrate that the aforementioned pipeline considerably improves the speed of the calculation of the radiative transfer models and therefore the SED fitting and retrieval of key parameters of galaxies.

CAMELS-SAM and beyond: Untangling the galaxy-halo connection with machine learning in observational space

Lucia Perez

While many advanced methods for the statistical analysis of cosmological data have been developed to handle the looming breadth of new astronomical observations, a common constraint is the availability of training cosmological simulation data, especially that which includes realistic galaxy formation physics and the volume and/or resolution necessary to match observations. CAMELS-SAM and its newest updates offer crucial and unique training data sets of realistic galaxies across an enormous range of cosmologies and galaxy physics formulations. As the larger-volume 'hump' of the Cosmology and Astrophysics with MachinE Learning Simulations (CAMELS) project, CAMELS-SAM uses semi-analytic models (SAMs) of galaxy formation to flexibly and quickly generate galaxies over 1000 dark-matter only simulations of L=100 h^-1 cMpc and N=640^3. We present: the updated galaxy catalogs using the Santa Cruz SAM with 1) up to 9 varied astrophysical parameters, with 2) complete photometry generated for all galaxies; and, 3) new galaxy catalogs with the L-Galaxies SAM. We show initial exploration of how well neural networks are able constrain OmegaM and sigma8 while marginalizing over both varied parameters within a particular SAM, as well as across distinct SAMs; and the many possibilities that the CAMELS-SAM datasets offer for the next generation of simulation-based inference on observed galaxies.

Less is enough: Extending ΛCDM with representation learning

Davide Piras

We present DE-VAE, a variational autoencoder (VAE) architecture to search for a compressed representation of dynamical dark energy (DE) models in observational studies of the cosmic large-scale structure. DE-VAE is trained on matter power spectra boosts generated at wavenumbers $k \in (0.01-2.5)$ h/Mpc and at four redshift values $z \in (0.1,0.48,0.78,1.5)$ for the most typical dynamical DE parametrization with two extra parameters describing an evolving DE equation of state. The boosts are compressed to a lower-dimensional representation, which is concatenated with standard cold dark matter (CDM) parameters and then mapped back to reconstructed boosts; both the compression and the reconstruction components are parametrized as neural networks. Remarkably, we find that a single latent parameter is sufficient to predict 95% (99%) of DE power spectra generated over a broad range of cosmological parameters within 1σ (2σ) of a Gaussian error which includes cosmic variance, shot noise and systematic effects for a Stage IV-like survey.

This single parameter shows a high mutual information with the two DE parameters, and these three variables can be linked together with an explicit equation through symbolic regression. Considering a model with two latent variables only marginally improves the accuracy of the predictions, and adding a third latent variable has no significant impact on the model's performance. We discuss how the DE-VAE architecture can be extended from a proof of concept to a general framework to be employed in the search for a common lower-dimensional parametrization of a wide range of beyond-ΛCDM models and for different cosmological datasets. Such a framework could then both inform the development of cosmological surveys by targeting optimal probes, and provide theoretical insight into the common phenomenological aspects of beyond-ΛCDM models.

Learned harmonic mean with normalizing flows for Bayesian model comparison in cosmology

Alicja Polanska

Computing the Bayesian evidence is an important task in Bayesian model selection, providing a principled quantitative way to compare models. In this work, we introduce normalizing flows to improve the learned harmonic mean estimator of the Bayesian evidence. This recently presented estimator leverages machine learning to address the exploding variance problem associated with the original harmonic mean. The improved method provides an accurate, robust and scalable estimator of the Bayesian evidence. Moreover, it is agnostic to the sampling strategy, meaning it can be combined with various efficient MCMC sampling techniques or variational inference approaches. We present numerical experiments demonstrating the effectiveness of the use of normalizing flows for the learned harmonic mean. We also apply the method to practical cosmological examples, including a cosmic shear analysis using CosmoPower-JAX, a JAX-based implementation of the CosmoPower framework that accelerates cosmological inference by building differentiable neural emulators of cosmological power spectra. We leverage the flexibility of the learned harmonic mean estimator to couple it with the differentiable CosmoPower-JAX emulator and the efficient No U-Turn Sampler (NUTS) implemented in NumPyro, and perform model comparison for cosmic shear in 37 dimensions. Our method takes around 1 day of compute time on 3 GPUs offering a significant speed-up relative to the conventional method using nested sampling, which takes 4 months on 48 CPU cores. We also successfully perform a cosmic shear analysis in a 157-dimensional setting, where using conventional methods is not feasible. This shows that the scalability CosmoPower-JAX and the learned harmonic mean estimator offer could allow for the comparison between models of unprecedented complexity, thus unlocking the full potential of Bayesian analysis even in high-dimensional settings. We make the harmonic code implementing the learned harmonic mean estimator publicly available (https://github.com/astro-informatics/harmonic).

Cosmology with Euclid

Alkistis Pourtsidou

Ongoing optical galaxy surveys like DESI and Euclid are promising to deliver cosmological measurements of unprecedented statistical precision. At the same time, the 21cm intensity mapping technique is being tested with instruments like MeerKAT. Combining the two methods (and CMB observations) can alleviate systematics and provide galaxy evolution and cosmological constraints. I will discuss a suite of possible synergies and their scientific promise.

Persistent homology applied to DES Y3 lensing maps: Using SBI to extract cosmological information

Judit Prat Marti

In this talk I will provide an overview of our approach to extracting cosmological information from weak lensing mass maps using state-of-the-art simulation-based inference techniques applied to data from the first three years of observations of the Dark Energy Survey. I will focus specifically on two higher-order summary statistics: Persistent Homology and Convolutional Neural Networks trained through a contrastive learning approach.

Quantifying the impact of and interplay between different analysis choices for LSST-Y1 Cosmic Shear

Naomi Robertson

Achieving robust cosmological constraints from cosmic shear involves several stages and many different analysis choices. Recent galaxy weak lensing analyses (DES & KiDS 2023) have shown that small shifts in parameter constraints are exacerbated by some combinations of analysis choices. As constraining power from cosmic shear improves, more complex modelling and accounting of systematics is required. Galaxy lensing is affected by baryonic feedback, primarily through active galactic nuclei redistributing matter at the centre of galaxies, when measuring small scales. We have performed a full mock-Rubin cosmic shear analysis and have found that removing data that is strongly affected by baryonic feedback would mean excluding almost 80% of the data, however marginalizing over baryonic feedback parameters without any external information leads to minimal improvements in constraining power.

This is because data from the Rubin Observatory is so precise that even a small effect is detectable. For baryon systematics we are currently dependent on cosmological hydrodynamical simulations, however cross-correlations with external probes such as X-ray or CMB measurements provide a more direct way to constrain baryons. We will present results on the potential impact of baryons on cosmic shear with Rubin Observatory and the interplay between baryons and other cosmic shear systematics.

Using graph neural networks to detect dark matter in stellar streams

Keir Rogers

The fundamental nature of dark matter so far eludes direct detection experiments, but it has left its imprint in the population of low-mass sub-halos in the Milky Way. The Rubin Observatory is forecast to probe the lightest dark halos ever detectable (~ 10^5 M_Sol) through their gravitational perturbations with thin tidal streams of stars in our Galaxy. But the data will be highly complex with observational selection functions, contamination from background galaxies and incomplete velocity information from complementary spectroscopic surveys like the Dark Energy Spectroscopic Instrument (DESI). I will present a new approach to robust inference of fundamental physics from stellar stream data using graph neural networks to compress the full 6D phase space and normalizing flows to infer posterior distributions. This new pipeline is intended for next-generation surveys like Rubin and DESI and will supersede sub-optimal density power spectrum analyses.

Field-level inference for weak lensing – DES-Y3 mass map reconstruction with lognormal priors

Supranta Sarma Boruah

In this talk, I will present new results from our field-level inference code, KaRMMa. Assuming a lognormal prior on the weak-lensing convergence field, KaRMMa performs tomographic, curved-sky field-level inference with weak-lensing data. With simplifying assumptions, we have previously shown that field-level inference with lognormal prior can substantially improve cosmological constraints for LSST-Y1 data (2204.13216, 2307.00070). After improving many of these assumptions, we have now used KaRMMa to produce the first Bayesian weak lensing mass maps with Stage-III survey data — namely DES-Y3, HSC-Y1 and KiDS-1000. Along with these results, I will also talk about the first steps towards including systematic effects in our pipeline.

Theoretical prediction for wavelet l1-norm

Vilasini Tinnaneri Sreekanth

The phenomenon of light deflection due to the presence of massive objects is called gravitational lensing, which leads to the distortion of the observed images of these distant galaxies. These distortions are usually very small and can be detected only by averaging over a huge number of galaxies. This regime is what we call weak lensing. Weak gravitational lensing serves as a major tool in unraveling the universe's large-scale structure. One of the key focuses of upcoming surveys is quantifying non-Gaussianities. Traditional two-point statistics fall short in capturing these non-Gaussian features, necessitating the adoption of higher-order statistics. However, a missing piece of the puzzle is a robust theoretical framework, because of which we currently rely on simulations, which need a huge amount of resources and time. One of the higher-order statistics methods that enables us to extract the non-Gaussian information from cosmic shear surveys is by using the one-point probability density functions. In a significant step forward, recent work by [Barthelemy et al. 2021] introduced a theoretical prescription to obtain the one-point probability density function based on the large deviation principle. Building upon this foundation, our study extends the theoretical framework to provide, for the first time, predictions for the ℓ1-norm. Previous work by [Ajani et al. 2021] has shown that ℓ1-norm outperforms the power spectrum by a considerable margin. With this work now have a theoretical prediction for the ℓ1-norm. We also explore the cosmological dependence of this statistic and validate our findings using simulations. Our results demonstrate that the theoretical predictions of the aperture mass ℓ1-norm align remarkably well with existing simulations, accurately capturing non-Gaussian information. Furthermore, our work sheds light on the cosmological implications of these finding.

A comparative analysis of regression methods for spectroscopic redshift estimation

Gregory Tsagkatakis

Machine learning, and particularly deep learning, is revolutionizing the field of astronomical data analysis, extending its influence to areas such as galaxy classification and stellar categorization. However, a significant subset of problems in this domain involves continuous variables, exemplified by tasks like photometric and spectroscopic redshift estimation. These problems are commonly approached through regression models which, while providing accurate point estimates, often fail to adequately convey uncertainty information. Alternatively, probabilistic priors can be introduced by targeting the estimation of specific parameters, e.g., the mean (μ) and standard deviation (σ) of a Gaussian distribution, under a maximum a-posteriori approach.

Although such approaches provide uncertainty insights, they typically assume a specific distribution for the predictions. An alternative approach involves formulating the problem as that of ordinal regression, which diverges from traditional regression by discretizing the continuous ranges and employing statistical metrics like cross-entropy instead of mean squared error. Our study focuses on the work presented in "Convolutional Neural Networks for Spectroscopic Redshift Estimation on Euclid Data' by Stivaktakis, R et al., published in IEEE Transactions on Big Data (2020). We aim to present both published and new results related to the comparative analysis between traditional regression and ordinal regression, utilizing metrics like accuracy, robustness to noise, and uncertainty quantification. The anticipated insights from this comparison will not only shed light on the strengths and limitations of each method but are also expected to inform similar research endeavors in other astronomical data analysis tasks.

Cosmological field-level inference from supernovae

Eleni Tsaprazi

Synthetic catalogues of supernova observations are used to forecast the constraining power of surveys and explore astrophysical effects. In this talk, I will be presenting a low-redshift supernova catalogue which follows the clustering of real galaxies in the large-scale structure and is equipped with realistic emulated galaxy evolution properties and peculiar velocities. I will discuss applications on the interface between galaxy evolution and Cosmology, highlighting results which suggest that the simulation, which reproduces the physics of supernova occurrence in galaxies, in accordance with a multitude of observations, provides a pristine ground to study the intersection between astrophysical and cosmological effects.

Pseudaria: Discovering unknown systematics in galaxy clustering surveys

André Vitorelli

Upcoming and ongoing large surveys like J-PAS, SPHEREx, DESI, Euclid will measure galaxy clustering at large (>10deg) scales. This offers an opportunity to constrain primordial non-gaussianities - a key component in understanding the dynamics of the early universe. However, clustering is extremely sensitive to systematic errors in survey completeness - which can induce large scale correlations and bias results, possibly in unpredictable ways. We have developed a new method that, by using pseudo-Gibbs sampling of constrained realizations of systematic maps from two overlapping surveys, is able to quantify unknown completeness errors. In this talk, I will present tests and results applicable to the analysis of the DESI Legacy Survey angular clustering as an example case, and discuss future applications of the technique.

KiDS-SBI: Simulation-based inference analysis of KiDS-1000 cosmic shear

Maximilian von Wietersheim-Kramsta

Cosmic shear, the weak gravitational lensing effect on distant galaxies due to matter in the foreground, is a powerful tool to study the distribution of matter, to probe its large-scale structure, and infer the cosmological model of the Universe. Standard analyses are typically based on the assumption of a Gaussian likelihood with a parameter-independent covariance, but these assumptions may not hold for all observables, scales and/or all systematics. Simulation-based inference (SBI) addresses this by evaluating an effective likelihood from forward-simulations which map parameters to data vectors. To this end, I will present a novel application of SBI to a cosmic shear analysis of the Kilo-Degree Survey's KiDS-1000 data release. The forward model is based on lognormal random fields which take into consideration systematics which are typically not modelled in standard inference, such as variable depth, point-spread function variations, shear biases, etc. I will also describe how the simulated galaxy catalogues are compressed to shear-shear angular two-point statistics which are further compressed using score compression. I will show how we train a 12-dimensional neural likelihood estimation to obtain a converged and unbiased posterior of the cosmological parameters within LambdaCDM. We achieve this with only 10,000 model evaluations which run in a time comparable to a standard MCMC. We find that our constraints on the weak lensing parameter, S8, are similar to constraints from previous analyses of KiDS-1000. We note a non-negligible parameter-dependence in the learnt likelihood which is consistent with cosmic variance. At the same time, we find that systematics such as variable depth can have significant impacts on the posterior estimates. Lastly, I will highlight how these findings and SBI will help address the modelling/inference challenges facing upcoming stage IV galaxy surveys.

Mass-mapping with conditional generative adversarial networks

Jessica Whitney

Reconstruction of the cosmological matter distribution from lensing observables requires solving an ill-posed inverse problem known as ‘mass-mapping’. Because of the ill-posed nature of this problem, having uncertainties to pair with reconstructions can be highly valuable. Upcoming surveys such as Euclid and the Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST) will create an unprecedented amount of new data for analysis, meaning reconstruction techniques must also be fast and efficient to be feasible. Machine learning models have shown great promise for problems involving imaging and large datasets. Conditional generative adversarial networks (GANs) are a particularly good model choice as they are known for their high-fidelity samples, which they can produce rapidly.

Our approach employs a conditional GAN to approximate samples from a Bayesian posterior distribution, meaning they can be interpreted in a statistically robust manner. Unlike typical GANs, our approach stands out for its resilience against mode collapse. We avoid mode collapse through the incorporation of regularization techniques that actively promote sample diversity. We present our methods, which uses data-driven priors to generate high-fidelity mass maps, which can be generated in a matter of seconds alongside associated uncertainties. To demonstrate the effectiveness of our approach, we train our model on mock COSMOS-style data, made using Colombia Lensing’s kappaTNG mock weak lensing suite. Subsequently, we apply our trained model to the real COSMOS data, and we will showcase the corresponding reconstruction results.

Joint multi-channel deconvolution for Euclid and LSST images

Utsav Akhaury

High spatial resolution and high signal-to-noise observations are prerequisites to most observational astrophysical problems. However, expecting the two conditions to happen simultaneously is challenging: space telescopes have limited collecting power, while larger telescopes necessarily need to be ground-based and are therefore affected by atmospheric turbulence. Exploiting the best of both worlds is possible, provided reliable deconvolution techniques are developed to remove blurring by the atmosphere and instrument Point Spread Function (PSF) while minimizing noise in the solution. While deconvolution is primarily used to reconstruct galaxy images at high spatial resolution in each photometric band independently, there are scenarios, particularly at low signal-to-noise ratios, where joint multi-channel deconvolution can enhance the detection and characterization of systems. One such application is the joint deconvolution of Euclid-LSST images, where the Euclid VIS-band overlaps with three of the LSST filters: r, i, z. We introduce an iterative ML-based algorithm for jointly deconvolving lower-resolution images from the r, i, z bands of LSST using the higher-resolution Euclid VIS-band image as a prior. Our method demonstrates effectiveness in terms of resolution recovery, flux preservation, and generalization across different noise levels. Through our joint deconvolution approach, we achieve resolution recovery in LSST simulated images close to that of HST (0.05”), an accomplishment nearly impossible with independent deconvolution of each photometric band.

Proximal nested sampling with data-driven priors

Henry Aldridge

Bayesian model selection provides a powerful framework for objectively comparing models directly from observed data, without reference to ground truth data. However, Bayesian model selection requires the computation of the marginal likelihood (model evidence), which is computationally challenging, prohibiting its use in many high-dimensional Bayesian inverse problems. With Bayesian imaging applications in mind, we introduce the proximal nested sampling methodology to objectively compare alternative Bayesian imaging models for applications that use images to inform decisions under uncertainty. The methodology is based on nested sampling, a Monte Carlo approach specialized for model comparison, and exploits proximal Markov chain Monte Carlo techniques to scale efficiently to large problems and to tackle models that are log-concave and not necessarily smooth (e.g., involving l1 or total-variation priors). Taking one step further, we show how proximal nested sampling can be extended using Tweedie's formula to support data-driven priors, such as deep neural networks learned from training data. We demonstrate our method by carrying out Bayesian model comparison between data-driven and hand-crafted priors in imaging applications like radio-interferometric image reconstruction.

Cosmic shear

Karthika Bhuvanendran

Analyzing the data from future cosmological surveys is going to be a big challenge. These surveys promise measurements with an accuracy of a few percent, and to make sense of these measurements, our theoretical predictions need to be just as accurate, if not more so. For large scale structure probes, which explore the nonlinear regime of structure formation, this level of accuracy can only be achieved with costly high-performance simulations. While the state-of-the-art two-point statistics serve as clean summary statistics, they can encapsulate all cosmological information only if the underlying matter distribution is Gaussian. However, gravity introduces non-linear features on small scales, rendering the distribution non-Gaussian. Non-Gaussian statistics become necessary to extract complete cosmological information from cosmic shear data as it can double, sometimes triple, the precision on some parameters given the same data. Obtaining additional information from the same lensing data necessitates alternative estimators. Unlike in CMB analysis where one can generate a large number (tens to hundred thousand) of power spectra with fast codes like CAMB relatively easily this is impossible for large scale structure simulations. Running simulations of large-scale structure formation is computationally expensive and time-consuming. To address this challenge, researchers often develop emulators, which are fast prediction tools based on a relatively small number of high-precision simulations. These emulators can replace the need for running a vast number of simulations in the analysis, providing a more efficient and practical approach to match the accuracy required for the measurements from future cosmological surveys. In this project we use a Gaussian Process Regression emulator to train on the analytical predictions for shear 2-point correlation functions and thus predict weak lensing statistics for wCDM cosmologies. Where our parameters of interests are $\pi * = [\Omega m, S8, w0, wa]$. Cosmic shear is maximally sensitive to Ωm, S8, with sample range carefully chosen as $\Omega m = [0.1429, 0.5735]$, $S8 = [0.55, 0.95]$ and parameters in the dark energy equation of state as $w0 = [-3.0, -0.33]$ and $wa = [-3.0, 3.0]$. These accuracy requirements will inform next-generation N-body simulations, which will be used to train the other non-Gaussian statistics.

Self-supervised component separation for the extragalactic submillimeter sky

Victor Bonjean

We use a new approach based on self-supervised deep learning networks originally applied to transparency separation in order to simultaneously extract the components of the extragalactic submillimeter sky, namely the cosmic microwave background (CMB), the cosmic infrared background (CIB), and the Sunyaev-Zel'dovich (SZ) effect. In this proof-of-concept study, we test our approach on the WebSky extragalactic simulation maps in a range of frequencies from 93 to 545 GHz, and compare with one of the state-of-the-art traditional methods, MILCA, for the case of SZ. We first visually compare the images, and then statistically analyse the full-sky reconstructed high resolution maps with power spectra. We study the contamination from other components with cross spectra, and particularly emphasise the correlation between the CIB and the SZ effect and compute SZ fluxes around positions of galaxy clusters. The independent networks learn how to reconstruct the different components with less contamination than MILCA. Although this is tested here in an ideal case (without noise, beams, or foregrounds), this method shows significant potential for application in future experiments such as the Simons Observatory (SO) in combination with the Planck satellite.

Bayesian inference of initial conditions from non-linear cosmic structures using field-level emulators

Ludvig Doeser

Unlocking the full potential of next-generation cosmological data requires navigating the balance between sophisticated physics models and computational demands. We propose a solution by introducing a machine learning-based field-level emulator within the HMC-based Bayesian Origin Reconstruction from Galaxies (BORG) inference algorithm. The emulator, an extension of the first-order Lagrangian Perturbation Theory (LPT), achieves remarkable accuracy compared to N-body simulations while significantly reducing evaluation time. Leveraging its differentiable neural network architecture, the emulator enables efficient sampling of the high-dimensional space of initial conditions. To demonstrate its efficacy, we use the inferred posterior samples of initial conditions to run constrained N-body simulations, yielding highly accurate present-day nonlinear dark matter fields compared to the underlying truth used during inference.

The Euclid Survey

Xavier Dupac

Euclid was launched on the 1st of July 2023, and with the first data came new challenges for the survey planning. In particular, the VIS stray light levels obliged us to completely rethink the survey strategy.

Higher order statistics for neutral hydrogen intensity mapping

Pauline Gorbatchev

Mapping the integrated emission of the 21-cm line of neutral hydrogen (HI) from all the galaxies at a given redshift, a technique known as HI intensity mapping is the new frontier to probe the cosmic large-scale structure. Indeed, it consists of treating the 21 cm signal as a diffuse background instead of measuring the 21 cm emission of each galaxy separately. Thus, it is less costly, less time consuming and it allows to treat a large cosmological volume. It has a great potential for cosmology, as thanks to its exquisite redshift resolution it allows us to track the evolution of cosmic structures across time in a way unparalleled by traditional galaxy redshift surveys. Furthermore, the study of individual galaxies is not required for the study of the large-scale structures of the Universe.

Constraining the large-scale diffused gas using the cross-correlation between the tSZ and the ISW effect

Ayodeji Ibitoye

We present a joint cosmological analysis of the power spectra measurement of the Planck Compton parameter and the integrated Sachs–Wolfe (ISW) maps. We detect the statistical correlation between the Planck Thermal Sunyaev\textendash Zeldovich (tSZ) map and ISW data with a significance of a 3.6σ confidence level~(CL), with the autocorrelation of the Planck tSZ data being measured.

Predicting the physical parameters in nearby galaxies with machine learning

Inja Kovačić

During its mission, Euclid will image thousands of nearby galaxies in four bands at a high angular resolution. We investigate the degree with which we can extract the distribution of physical parameters (stellar mass, mass-averaged stellar metallicity and age) from synthetic Euclid images of nearby galaxies. We generated these synthetic Euclid maps for a set of simulated galaxies, which were extracted from the TNG50 cosmological simulation, with the radiative transfer code SKIRT. We use a machine learning scheme to map the surface brightness and colors to physical parameters on a sub-kpc level.

Dissecting the information content in the large-scale structures of the Universe using machine learning

Arnab Lahiry

Prior research has shown that convolutional neural networks (CNNs) can be used as an estimator of cosmological information from 2D baryonically contaminated maps as well as density field maps from simulations belonging to the CAMELS project. This research has been a major step forward in extracting cosmological information, overcoming the hurdle of the non-Gaussianity of data, on a purely theoretical basis. One of the major questions after the results obtained in that work was “From where in the cosmos does this information arise?”. To attempt to answer these questions, we investigate the scales and morphological features of the matter density cosmic web which give rise to the cosmological inferences gained by the neural network. We identify the bins in the frequency domain of the cosmic web maps responsible for the information content. We also investigate the interpretability of the CNN with respect to cosmological inference as well as feature dependence on the results to identify the morphological features responsible for this information content, followed by augmentations on the maps to obtain a more robust inference about the origins of cosmological information in the large-scale structures of the Universe.

Photometric redshift estimation of galaxies by machine learning

Changhua Li

The accurate estimation of photometric redshifts plays a crucial role in accomplishing science objectives of the large survey projects. We used many machine learning methods, such as CatBoost, XGBoost, Multi-Layer Perceptron, and Random Forest to train redshift prediction models in the DESI Legacy Imaging Surveys and obtained good performance in many metrics. Finally, we released a photometric redshift catalog that includes all galaxies of DESI Legacy Imaging Surveys DR9.

Simulation-based inference of cosmic shear with the Kilo-Degree Survey

Kiyam Lin

The standard approach to inference from cosmic large-scale structure data employs summary statistics that are compared to analytic models in a Gaussian likelihood with pre-computed covariance. To overcome the idealising assumptions about the form of the likelihood and the complexity of the data inherent to the standard approach, we employed simulation-based inference (SBI), which learns the sampling distribution as a probability density parameterised by a neural network. We first validated that SBI is a viable methodology by testing its application towards a suite of exactly Gaussian-distributed data vectors for the most recent Kilo-Degree Survey (KiDS) weak gravitational lensing analysis (Lin et al., 2023). New simulations were then constructed and used to train our neural density estimators. We compare the differences between our version of the fiducial KiDS analysis done using SBI vs. our new analysis that include previously not modelled systematics such as variable depth and also compare the effects of different levels of Gaussianity imposed on the inference. We present the newest KiDS-1000 results analysed using SBI (von Wietersheim-Kramsta in prep.)

Trajectory approach to Quantum Cosmology

Kratika Mazde

In quantum cosmology, the problem of time is one of the many issues that require imminent attention. Over the past few decades, multiple attempts have been made to solve this problem vis-à-vis Wheeler-De Witt (WdW) equation. WdW equation happens to be timeless and that implies that all wave functions are constant, nothing is being occurred. Majorly, it implies that time is frozen which is strange because classically evolution is known to occur. This further implies that the notions of 'time' in General Relativity and Quantum Theory clash. When attempting to replace these two disciplines of physics with a single framework in circumstances when both criteria apply, such as in black holes or in the very early times, this presents a dilemma. One of the many ideas put out is to use a perfect fluid, whose Hamiltonian naturally converts the Wheeler-De Witt equation into the Schrödinger form when the momentum is quantized. In the quantum mechanical trajectory approach, this kind of solution further rids us of the cosmological singularities. Assigning an internal degree of freedom to define time could also be immensely helpful in this regard.

CMB cross correlations

Shree Hari Mittal

The second data release of LOFAR Two-meter sky survey covers ~5.600 deg^2 and more than 4 million sources however, lacks redshift information of its radio sources. Using the optical counterparts to the radio sources with photometric redshifts, we analyze the harmonic space cross-correlation of galaxy over-density with CMB lensing map from Planck, as well as auto-correlation of such radio galaxies. We aim to better constrain the galaxy bias of radio galaxies and the amplitude of matter perturbations on a large scale, characterized by σ_8.

Accelerating discrete-continuous convolutions for scalable and equivariant spherical CNNs

Kevin Mulder

A rotationally equivariant and scalable spherical convolutional neural network (CNN) framework (arXiv:2209.13603) applying the hybrid discrete-continuous group convolution (DISCO) has previously been employed to resolve the tension between continuous and discrete approaches, which were either equivariant or scalable but not both. Recent advances in machine learning and high-performance numerical computing are leveraged to simultaneously streamline, accelerate and optimise the DISCO framework. Through implementation within the JAX computational framework a number of key features become accessible that are essential for usage in large-scale machine learning tasks. Due the rapid integration of machine learning techniques coupled with the increasing data requirements of upcoming cosmological surveys, e.g., the large field observations of Euclid and the Legacy Survey or Space and Time over the celestial sphere, cosmological data analyses are rapidly becoming part of the aforementioned large-scale machine learning tasks. JAX allows for the composability of both natively supported automatic differentiation and XLA (accelerated linear algebra) compilation which enables deployment on hardware accelerators such as GPUs and TPUs. Additionally, differentiability of sparse matrices, through the sparsify transform, enables the direct usage of sparse gradients in the DISCO convolution. The central components of the DISCO convolution in the JAX accelerated DISCO framework are primarily expressed as pure functions. This enables the usage of JAX transformations and yields effortless access directly to the constituent functions. These functions can in turn be composed to form the full DISCO convolution in various neural network libraries. The JAX accelerated DISCO framework includes a fully composed DISCO spherical convolutional layer in the format of a custom Flax layer, and will be made publicly available.

Normalizing flows for cosmological parameter inference

Kutay Nazli

Inferring signals from cosmological maps requires the extraction of many hundred thousand data points efficiently and accurately. In Euclid's science case, this challenge is further exacerbated by the non-Gaussianity of the signal's posterior. It is possible to fully model and evaluate non-Gaussian probability distributions at arbitrary points using normalizing flow-based machine learning approaches. However, the key challenge is scaling this architecture to the high-dimensions needed for the Euclid science case. I present the high dimensional flow-based architecture trained on the posterior samples generated by Almanac, which allows for inference of cosmological parameters in a fully Bayesian way. Access to the full posterior distributions instead of the best estimates or summary statistics of the signal allows for optimal removal of all biases the current standard approaches cannot remove.

Differentiable spherical harmonic transforms and generative modeling for cosmology

Matthew Price

Many areas of science and engineering encounter data defined on spherical manifolds. In cosmology in particular, fields are observed over the celestial sphere. Modelling and analysis of spherical data often necessitates spherical harmonic transforms, at high degrees, and increasingly requires efficient computation of gradients for machine learning or other differentiable programming tasks. We develop novel algorithmic structures for accelerated and differentiable computation of generalized Fourier transforms on the sphere and rotation group, i.e., spherical harmonic and Wigner transforms, respectively. We present a recursive algorithm for the calculation of Wigner d-functions that is both stable to high harmonic degrees and extremely parallelizable. By tightly coupling this with separable spherical transforms, we obtain algorithms that exhibit an extremely parallelizable structure that is well-suited for the high throughput computing of modern hardware accelerators (e.g., GPUs). Our algorithms are implemented within the JAX differentiable programming framework in the S2FFT software code (https://pypi.org/project/s2fft/). With such transforms to hand we have developed a new generation of scattering covariances natively on the sphere. Leveraging automatic differentiation, we demonstrate that this representation can be used as a robust generative model for complex cosmological signals over the full celestial sphere, e.g., cosmic string signatures and large-scale structure, which exhibit highly non-Gaussian filamentary structure. We demonstrate that such generative models are an effective alternative to numerical simulation, in the case of cosmic strings reducing compute time for high-resolution fields from approximately one million CPU hours to a matter of hours on a GPU.

Cosmic shear without shape noise

Pranjal R. S.

The unknown intrinsic shape of source galaxies is one of the largest uncertainties of weak gravitational lensing (WL). It results in the so-called shape noise at the level of $\sigma_\varepsilon^{WL} \approx 0.26$, whereas the shear effect of interest is of order percent. Kinematic lensing (KL) is a new technique that combines photometric shape measurements with resolved spectroscopic observations to infer the intrinsic galaxy shape and directly estimate the gravitational shear. This work presents a KL inference pipeline that jointly forward-models galaxy imaging and slit spectroscopy to extract the shear signal. We build a set of realistic mock observations and show that the KL inference pipeline can robustly recover the input shear. To quantify the shear measurement uncertainty for KL, we average the shape noise over a population of randomly oriented disc galaxies and estimate it to be σ_ε^{KL} 0.022-0.038 depending on emission line signal-to-noise. This order of magnitude improvement over traditional WL makes a KL observational program feasible with existing spectroscopic instruments. To this end, we characterize the dependence of KL shape noise on observational factors and discuss implications for the survey strategy of future KL observations. In particular, we find that prioritizing quality spectra of low inclination galaxies is more advantageous than maximizing the overall number density.

Improved and novel statistical methods for dark matter-baryon scattering in linear cosmology

Suroor Seher Gandhi

Linear cosmological observations, particularly those involving the cosmic microwave background (CMB), offer a unique avenue to investigate the elusive properties of dark matter (DM) through its interactions with standard model (SM) particles. The precision of contemporary experiments has reached sub-percent-level accuracy (<1%), necessitating a theoretical formalism of definitive accuracy to extract meaningful constraints on DM-SM interactions.

Standardized benchmark for field-level cosmological inference from galaxy surveys

Hugo Simon Onfroy

The advent of new-generation galaxy surveys such as DESI and Euclid poses significant challenges in data analysis. Standard analyses relying on analytical models for the 2-pt statistics of the galaxy density field are limited in scale and do not exploit the non-Gaussian information of the field.

Void Lensing x SBI: a new tool for constraining cosmology

Chen Su

Void are the large underdensity regions in the cosmic web. As a comprehensive probe to the galaxies, it is known that void can provide additional information in cosmological constraints. Meanwhile, thanks to its underdensity feature, voids are less affected by the non-linear effects. However, accurately modeling void statistics is still a problem, meaning that using traditional likelihood inference is difficult when using void statistics to constrain cosmology. In this work we apply the Simulation-Based Inference (SBI), a likelihood-free inference method, to avoid modeling the void statistics directly. Using void lensing as an example, we will show the constraining power of void statistics.

The impact of mass-mapping methods on cosmological inference

Andreas Tersenov

Weak lensing mass-mapping provides a way to reconstruct the projected matter density field from the observed galaxy ellipticities, and is commonly utilized in cosmological inference from weak lensing observations. One of its main advantages is that it allows for analyzing the data in alternative ways, beyond the standard two-point cosmic shear, giving access to the full distribution of dark matter in the sky. However, the reconstruction of mass maps is an ill-posed inverse problem, due to missing data and noise in the observations, and there exist several different methods to solve it. In this study, we investigate the impact of the mass-mapping method on cosmological parameter estimation in a systematic way, utilizing a likelihood-based approach and advanced higher-order summary statistics. Such an investigation is especially important for large-scale weak lensing surveys, as understanding the impact of these algorithms on cosmology inference is crucial for selecting the most suitable ones for those surveys. Our analysis is conducted on the CosmoSLICS suite of simulations, a comprehensive dataset tailored for weak lensing studies beyond the standard two-point statistics. We employ different state-of-the-art mass mapping techniques: the Kaiser-Squires algorithm, Wiener filter, sparsity-based reconstruction, and deep learning models, to transform shear measurements into mass maps. The cosmological inference is derived from these mass maps using three higher-order statistics: peak counts, starlet multiscale peak counts, and the starlet L1-norm. These statistics are chosen for their sensitivity to the underlying cosmology, providing a robust framework for assessing the impact of mass mapping methods on cosmological parameter estimation.

Detecting modified gravity with machine learning

Linus Thummel

Forthcoming stage-IV cosmological surveys will perform measurements with unprecedented accuracies up to non-linear scales. We have developed a novel machine learning approach to detect beyond-standard-model physics in the data using Bayesian Neural Networks. Based on the halo model reaction framework, we create non-linear dark matter power spectra for a variety of modified gravity and dark energy theories. The modeling includes baryonic effects and massive neutrinos to increase the accuracy on non-linear scales and enhance our ability to detect deviations from ΛCDM.

Constraining extended cosmologies with weak lensing data in the non-linear regime

Maria Tsedrik

Cosmological surveys, such as Euclid and LSST, will provide us with unprecedented measurements of the large-scale structure. This also means that our knowledge of the Universe is now constrained by the accuracy of our modeling rather than our measurements. Several frameworks have been proposed to generalize a broad class of extended cosmologies, aiming to avoid testing models individually in the vast realm of modified gravities and dark energy models. However, these frameworks only allow us to model on linear scales, resulting in the exclusion of valuable information from nonlinear scales, thereby neglecting the primary advantage of the Stage-IV mission.

Galactic magnetic field reconstruction using the optical polarization of starlight

Alexandros Tsouros

Ultra-high energy cosmic rays (UHECRs) are highly energetic charged particles with energies exceeding 10^18 eV. Identifying their sources and production mechanism can provide insight into many open questions in astrophysics and high energy physics. Crucial for understanding their origins and propagation mechanisms is the 3-D structure of the Galactic magnetic field, which influences their trajectories. However, this structure is not well understood as the GMF is mainly probed through integral measurements. Instead, observations of a large number of stars at known distances in optical polarization, tracing the Galactic dust, can map the magnetic field between them with the use of novel statistical methods. The Polar Areas Stellar Imaging in Polarisation High Accuracy Experiment (PASIPHAE) will deliver such a map utilizing novel-technology wide-field-optimized optical polarimeters (WALOPs). This endeavor will not only update our current understanding of UHECR arrival directions but also significantly contribute to a broad spectrum of astrophysical research, including the study of the dust polarization foreground for the hunt for B-modes in the cosmic microwave background polarization, the interstellar medium, high-energy astrophysics, and galactic evolution. This presentation will delve into the objectives of PASIPHAE and explore the latest advancements in statistical methods designed to tackle the inverse problems presented by the influx of data that the survey will generate.

Precise cosmological constraints from galaxy clustering using the wavelet scattering transform

Georgios Valogiannis

Optimal extraction of the non-Gaussian information encoded in the Large-Scale Structure (LSS) of the universe lies at the forefront of modern precision cosmology. In this poster, I summarize recent efforts to achieve this task using the Wavelet Scattering Transform (WST), which subjects an input field to a layer of non-linear transformations that are sensitive to non-Gaussianity through a generated set of WST coefficients. In order to assess its applicability in the context of LSS surveys, I discuss the first WST application to actual galaxy observations, through a WST re-analysis of the BOSS DR12 CMASS dataset. After laying out the procedure on how to capture all necessary layers of realism for an application to data obtained from a spectroscopic survey, I show results for the marginalized posterior probability distributions of multiple cosmological parameters obtained from a likelihood analysis of the CMASS data. A joint WST+ 2-point correlation function (2pcf) analysis is found to deliver a substantial improvement in the values of the predicted 1σ errors compared to the regular 2pcf-only analysis, highlighting the exciting prospect of employing higher-order statistics in order to fully exploit the potential of upcoming Stage-IV spectroscopic observations from DESI and Euclid.

Machine learning cosmology from void properties

Bonny Wang

Cosmic voids are the largest and most underdense structures in the Universe. Their properties have been shown to encode precious information about the laws and constituents of the Universe. We show that machine learning techniques can unlock the information in void features for cosmological parameter inference. We rely on thousands of void catalogs from the GIGANTES dataset, where every catalog contains an average of 11,000 voids. We focus on three properties of cosmic voids: ellipticity, density contrast, and radius. We train 1) fully connected neural networks on histograms from individual void properties and 2) deep sets from void catalogs, to perform likelihood-free inference on the value of cosmological parameters. We find that our best models are able to constrain the value of Ωm, σ8, and ns with mean relative errors of 10%, 4%, and 3%, respectively, without using any spatial information from the void catalogs. Our results provide an illustration for the use of machine learning to constrain cosmology with voids.

Field-level galaxy biasing from DES Y3 weak-lensing mass & galaxy clustering

Joshua Williamson

We present simulation-based inference results on galaxy biasing using Dark Energy Survey Year 3 galaxy clustering and weak-lensing mass maps. We use neural compression of power spectra as a summary statistic and field-level convolutional neural networks (CNN) direct from the field-level to constrain both galaxy biasing models and parameters. In our simulation-based analysis, we forward-model from the Gower Street N-body simulation suite to realistic mock weak-lensing and galaxy clustering maps. We model all relevant systematics and uncertainties in our mock data. We analyze both the Dark Energy Survey Year 3 MagLim and redMaGiC clustering catalogs for consistency.

Cluster strong lens modeling based on pixelated image distribution

Yushan Xie

Cluster-scale strong lensing is a powerful tool for exploring the properties of dark matter and constraining cosmological models. However, due to the complex parameter space, current strong lens modeling in galaxy clusters is using an approximation of point-like multiple images with Gaussian positional errors, which potentially introduces systematic bias. In this talk, I will introduce a novel pixel-level modeling method. Our method demonstrates that the bias in the recovered magnification maps, cosmological parameters can be substantially reduced. Moreover, considering the unacceptable computational costs with traditional algorithms, our modeling method is accelerated, and is applicable to real observations with the help of JAX.

COSMO21

20-24 MAY, 2024
CHANIA, GREECE

FOUNDATION FOR RESEARCH AND TECHNOLOGY - HELLAS

BE GREEN
KEEP IT ON
THE SCREEN