Euclid: probing the high-redshift universe at the epoch of re-ionization

Name of the laboratory:	Laboratoire d'Astrophysique de Marseille (LAM)
Thesis advisor:	Jean-Gabriel Cuby
Email:	jean-gabriel.cuby@lam.fr

Name of the laboratory:	Institut de Mathématiques de Marseille (I2M)
Thesis advisor:	Pierre Pudlo
Email:	pierre.pudlo@univ-amu.fr

Name of the laboratory:	Laboratoire d'Informatique Fondamentale (LIF)
Thesis advisor:	François-Xavier Dupé
Email:	francois-xavier.dupe@univ-amu.fr

Context

Galaxies and quasars at high redshift (z>7) are probes of the distant universe that enable studies of the formation and evolution of the first objects in the Universe and of the re-ionization epoch that occurred at z = 7.7 ± 0.7 (mid-point, Planck collaboration VI, 2018). Observing the Primordial Universe in this redshift range is a high priority objective of most space and ground-based projects of the next decades. EUCLID (launch in 2022) and JWST (launch in 2021) will be the first missions to unveil large samples of these high-redshift populations, followed by Roman, SKA precursors, ATHENA, LISA, etc. Complementary spectroscopic observations of the Euclid-detected sources with JWST, VLT, ALMA and the ELT will enable unprecedented detailed studies of the epoch of reionization and of the formation of the first galaxies and quasars. Euclid will have a considerable and long lasting impact and legacy value, enabling for the first time cross-correlation studies with 21-cm data.

EUCLID will detect hundreds (resp. tens) of quasars at z > 7 (resp. z > 8) from its 15,000 sq. degree wide survey, and thousands of galaxies at z > 7 from its 40 sq. degree deep survey. However, brown dwarfs and early type galaxies at intermediate redshifts have colours similar to those of high-redshift objects and act as contaminants. Sophisticated statistical tools facilitating the selection of the high-z objects are therefore required. Bayesian selection methods have been developed and used for the selection of high-redshift quasars in wide field infrared surveys (see e.g. Euclid Collaboration: Barnett et al., 2019 ; Pipien et al. 2018). These methods use detailed models of the high-redshift object and contaminant populations. The sensitivity of Euclid will probe domains of the parameter space not covered by these models (in magnitudes and/or in space), and more sophisticated models will need to be elaborated. Conversely, simplified models may be used to train machine or deep learning algorithms.

Methodology and Work Program

In a first phase, the work will consist in consolidating the models of the high-redshift object and contaminant populations and the Bayesian selection method, based on the Bayes Factor or full Bayesian model selection. The models will use existing spectral and morphological data and consider i) the luminosity functions of the populations of high-redshift objects, ii) early-type galaxies at intermediate redshifts and iii) the distribution of Brown Dwarfs (including binaries) in the Milky Way. These models will then be used to run an end-to-end simulation by generating mock catalogues fed into Euclid-simulated images and recovered with the Euclid pipeline. From a statistical point of view, the main challenge is to provide an algorithm to compute the evidence of the models that scales up to the number of objects to be screened. The performance of various selection algorithms will then be tested and compared. Ancillary and publicly available data from wide field surveys (e.g. PanStarrs, CFIS, MzLS, WISE, etc.) will be considered to further enhance the efficiency of the selection methods. The Bayesian selection method will be compared to state-of-the-art machine learning classifiers based on bagging, boosting, or deep neural networks trained on long databases. An important aspect of the work will be to test the performance of the selection methods on existing infrared datasets (e.g. HST, UKIDSS, VISTA, CFHQSIR, WISE, etc.) and to recover high-redshift galaxies and quasars so far detected, and possibly new ones. The model and tools will be applied to the initial Euclid datasets.

This work will be developed in the context of the Euclid Primeval Universe Working Group and in relation with the teams developing the ground segment of the mission. This work will lead to several publications. Upon completion, the successful laureate will have first-hand experience of a high-profile space mission and will be ideally positioned to continue working on the scientific exploitation of Euclid or of other large-scale astrophysical projects. He or she will also be positioned as an early career researcher in the fields of computational statistics and machine learning.

Applications

The ideal candidate will have a strong background in physics, space science or data science. Experience in Astronomy and Astrophysics is desirable. Applications must include a cover letter, a CV and contact information for reference letters.

Bibliography

- A Luminous Quasar at Redshift 7.642, Wang et al., 2021, ApJL, 907(1),
- Euclid preparation. V. Predicted yield of redshift 7 < z < 9 quasars from the wide survey, Euclid Collaboration: Barnett et al., 2019, A&A, 631
- An 800-million-solar-mass black hole in a significantly neutral Universe at a redshift of 7.5, Bañados et al., 2018, Nature, Volume 553, Issue 7689
- High-redshift quasar selection from the CFHQSIR survey, Pipien et al., 2018, A&A, 617, id.A127
- On the nature of the luminous Lya emitter CR7 and its restframe UV components, Sobral et al., arXiv:1710.08422
- A Remarkably Luminous Galaxy at z=11.1 Measured with Hubble Space Telescope Grism Spectroscopy, Oesch et al., The Astrophysical Journal, Volume 819, Issue 2, article id. 129, 11 pp. (2016)
- A luminous quasar at a redshift of z = 7.085, Mortlock et al., Nature, Volume 474, Issue 7353, pp. 616-619 (2011)
- Euclid Definition Study Report (Red book). arXiv:1110.3193 (pages 29-30)