Focus Meeting 7: New Horizons at the interface between Computational Astrophysics and Big Data

https://doi.org/10.1017/S174392132400334X Published online by Cambridge University Press

Astronomy in Focus
XXXIInd IAU General Assembly, August 2024
Diana M. Worrall, ed.
doi:10.1017/S174392132400334X

Exploring the High-Redshift Universe Using SKA and Other Upcoming Facilities

Tirthankar Roy Choudhury

National Centre for Radio Astrophysics, Tata Institute of Fundamental Research, Ganeshkhind, Pune 411007, India
email: tirth@ncra.tifr.res.in

Abstract. The formation of the first stars in the Universe remains a challenge in cosmology, necessitating a multi-wavelength approach with various observational facilities. While instruments like the James Webb Space Telescope (JWST) can directly observe early galaxies, upcoming radio telescopes such as the Square Kilometre Array (SKA) will provide insights into the hydrogen signal from the early Universe. Interpreting the extensive and complex datasets generated by these observations requires sophisticated theoretical models that balance detail with computational efficiency. These models must be coupled with advanced statistical techniques to ensure reliable interpretation of the underlying physical processes. This proceeding highlights recent developments in modelling the high-redshift Universe, alongside the evolution of machine learning-driven statistical techniques. These innovations are crucial for bridging the gap between theory and observation, allowing for a more refined understanding of cosmic evolution in light of forthcoming observational capabilities.

Keywords. galaxies: high-redshift, reionization, cosmology, radio astronomy, machine learning

1. Introduction

Understanding the formation of the first stars and galaxies in the high-redshift universe is crucial to uncovering the history of cosmic evolution. The study of reionization, driven by radiation from the first light sources, provides insights into the epoch when neutral hydrogen in the intergalactic medium (IGM) was ionized (Choudhury 2022). Recent advances in observational capabilities, such as the James Webb Space Telescope (JWST), have opened new opportunities for probing these distant epochs. With several upcoming probes, particularly the Square Kilometre Array (SKA), our understanding of the first stars is expected to deepen significantly in the coming decade.

This article summarizes the current status of observational probes and recent advances in theoretical modelling of reionization. We also highlight the challenges faced in theoretical modelling, which can complicate the interpretation of observational data.

2. Observational Probes of Reionization

There are two primary approaches to studying the first stars at high redshifts. One approach is to directly observe the galaxies that host these stars using ultraviolet (UV), optical, and near-infrared (NIR) telescopes such as the Hubble Space Telescope (HST) and JWST. Recent JWST data have already allowed researchers to study star-forming galaxies at redshifts as high as $z \sim 14$ and fainter magnitudes than ever before (Adamo et al. 2024). However, these observations are inherently limited by luminosity bias, i.e., the preferential detection of more luminous sources.

© The Author(s), 2026. Published by Cambridge University Press on behalf of The International Astronomical Union.

https://doi.org/10.1017/S174392132400334X Published online by Cambridge University Press

A complementary approach, which avoids this luminosity bias, is to study the hydrogen signal from the IGM, either in its neutral or ionized form. Lyman-α absorption from neutral hydrogen at $z \sim 6$ (Bosman et al. 2022) and secondary Cosmic Microwave Background (CMB) signals from ionized electrons (Planck Collaboration et al. 2020; Reichardt et al. 2021) have already provided valuable insights. Perhaps the most promising future probe of neutral hydrogen is the redshifted 21 cm signal, detectable using low-frequency radio telescopes. Current facilities, such as Experiment to Detect the Global EoR Signature (EDGES, Bowman et al. 2018) and Shaped Antenna measurement of the background Radio Spectrum 3 (SARAS 3, Singh et al. 2022), are probing the globally averaged signal, while LOw Frequency ARray (LOFAR, Mertens et al. 2020), Murchison Widefield Array (MWA, Trott et al. 2020) and other interferometric arrays are providing constraints on fluctuations. Future surveys with SKA-Low will significantly enhance both sensitivity and resolution, further advancing our understanding.

3. Modelling Reionization

Interpreting these observations requires comparison to sophisticated theoretical models that predict the evolution of the IGM and the sources of ionizing radiation. These models must be capable of calculating the relevant observables. Key components of these models include:

(*a*) **Large-scale structure:** Models must capture the distribution of dark matter haloes that host galaxies, along with the large-scale density and velocity fields. This can be achieved through N-body simulations or analytical prescriptions.
(*b*) **Star formation:** In reionization models, simple prescriptions can be used to estimate star formation in dark matter haloes, as it is impossible to resolve the relevant scales in cosmological simulations. These prescriptions, often described by free parameters or functions, include the fraction of baryons converted into stars, the timescale of star formation, and feedback processes. UV luminosity from galaxies thus computed can be compared with observed luminosity functions to constrain high-redshift star formation.
(*c*) **Ionizing radiation:** This involves relating the star-formation rate or UV luminosity to the number of ionizing photons produced. These can be estimated using stellar population synthesis models.
(*d*) **Escape of photons:** The fraction of ionizing photons that escape from galaxies into the IGM is a critical ingredient in modelling reionization. As with star formation, this too is described by simple parametrizations.
(*e*) **Ionization and thermal history:** The final component of the models is to track the evolution of ionization fronts and the thermal history of the IGM.

A schematic diagram showing the different steps involved in modelling the epoch of reionization is shown in Figure 1.

In order to account for all the relevant physical processes, many of which are extremely complex and non-linear, one needs to use numerical simulations that include gravity, gas dynamics, and radiative transfer. However, these simulations face several challenges, which can be broadly classified into three categories:

(*a*) **High dynamic range:** Reionization simulations must cover a wide range of scales, from large simulation boxes of ~ 200 Mpc in comoving units (representing masses of $\sim 10^{17} \mathrm{M}_{\odot}$), corresponding to volumes of typical surveys with upcoming telescopes, down to small dark matter haloes (or minihaloes) of masses $\sim 10^{6} \mathrm{M}_{\odot}$ hosting the

https://doi.org/10.1017/S174392132400334X Published online by Cambridge University Press

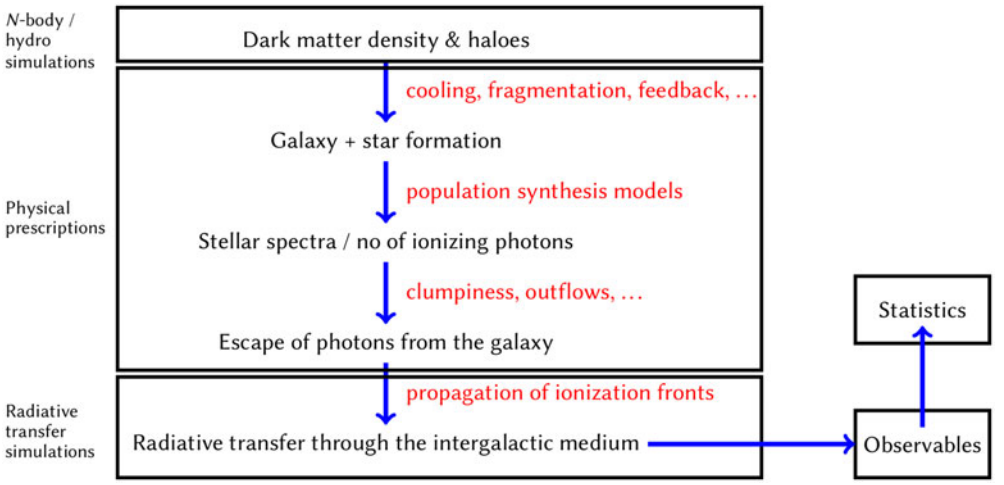


Figure 1. Steps involved in modelling the epoch of reionization, represented by black text. The three rectangular blocks represent the distinct techniques used within these models: large-scale structure modelled through N-body or hydrodynamical simulations; star formation, ionizing radiation, and photon escape represented by simplified parametrizations; and the ionization and thermal history modelled via radiative transfer simulations. Blue arrows indicate the information flow between these stages, while red statements denote the specific physical components essential to modelling each step.

first stars. This vast range requires computationally expensive simulations with $\sim 10^{11}$ particles or fluid elements.

This challenge is often overcome by using sub-grid prescriptions where one simulates a low-resolution large volume, and the small mass haloes are populated using analytical prescriptions for the conditional halo mass function (Barkana & Loeb 2004). Such methods allow one to simulate large volumes with relatively less computing power. An alternative approach is to construct a hybrid simulation consisting of low-resolution large boxes and high-resolution small boxes, each set containing lesser particles but effectively covering the entire dynamic range (Barsode & Choudhury 2024).

(b) **Radiative transfer:** Modelling the transfer of ionizing photons through the IGM is computationally challenging, as the radiative transfer equation is a seven-dimensional partial differential equation. To reduce computational costs, simplified methods such as the method of characteristics or moment-based approaches are often employed in the radiative transfer simulations (Gnedin & Madau 2022). Alternatively, photon-counting algorithms combined with sub-grid models for thermal evolution can offer a more efficient solution.

(c) **Parameter inference:** These numerical or semi-numerical models cannot probe the very small scales (e.g., scales relevant for star-formation in a galaxy) in a cosmological volume. These processes are hence captured through unknown parameters or functions (e.g., efficiency of star formation, escape of photons, small-scale clumping of the IGM which act as sinks of ionizing radiation). The parameter inference using simulations can be quite expensive, particularly using conventional Bayesian techniques where one may need numerous runs of the simulation. Recent advancements in machine learning have shown promise in addressing some of these challenges.

https://doi.org/10.1017/S174392132400334X Published online by Cambridge University Press

4. A Semi-Numerical Model: SCRIPT

The most widely used technique for semi-numerical models of reionization is based on the "excursion set" approach, where one determines if a region can remain self-ionized by evaluating it across different scales until it meets a specific ionization criterion. Independently, our group has developed a semi-numerical model, Semi-numerical Code for ReIonization with PhoTon-conservation (SCRIPT), to simulate reionization†. This model uses a photon-conserving algorithm to distribute photons around sources under the assumption of spherical symmetry, effectively addressing certain challenges encountered in excursion set-based models (Paranjape et al. 2016; Choudhury & Paranjape 2018). Its computational efficiency allows for the exploration of a wide parameter space, making it an effective tool for interpreting various observations, such as Lyman-α absorption (Choudhury et al. 2021b), CMB anisotropies (Mukherjee et al. 2019; Paul et al. 2021; Choudhury et al. 2021a; Jain et al. 2023, 2024b,a), UV luminosity functions, and measurements of the IGM temperature (Maity & Choudhury 2022a,b).

SCRIPT has also been used for forecasting future missions like CMB experiments (SO, LiteBIRD, CMB-S4, PICO) (Jain et al. 2023, 2024b,a) and 21 cm experiments (SKA) (Maity & Choudhury 2023; Choudhury et al. 2024). Recently, we employed Machine Learning (ML) techniques to explore the parameter space more efficiently using SCRIPT by integrating Gaussian Process Regression (GPR) with Bayesian Markov Chain Monte Carlo (MCMC) analysis. This hybrid method narrows the prior range of unknown parameters using a fast, low-resolution MCMC analysis and employs a GPR-based emulator to interpolate the likelihood function. This ML-based approach is more than an order of magnitude faster than conventional MCMC methods (Maity et al. 2023; Choudhury et al. 2024).

5. Summary

The epoch of reionization, when the first stars formed and ionized the hydrogen in the IGM, offers a unique window into the early universe. Current data, such as CMB anisotropies, Lyman-α absorption, and galaxy luminosity functions, provide valuable insights into this epoch. Upcoming facilities like SKA and the Thirty Meter Telescope (TMT) will enhance our ability to observe both neutral hydrogen and early galaxies, shedding further light on the reionization process.

Interpreting these observations requires efficient and accurate models capable of simulating the complex physical processes involved in reionization. Given the large dynamic range and the computational demands of these models, semi-numerical approaches offer a practical compromise between accuracy and efficiency. The uncertainties in these models, coupled with the vast parameter space, make advanced statistical techniques essential for parameter inference. Machine learning algorithms present a promising solution to these challenges, enabling us to bridge the gap between theory and observation.

The semi-numerical model SCRIPT, based on a photon-conserving algorithm, provides an effective framework for studying reionization. It has been successfully compared with a wide range of observational data and used to make forecasts for upcoming missions. This model serves as a valuable tool for exploring the parameter space and gaining deeper insights into the processes driving cosmic reionization.

References

Adamo, A., Atek, H., Bagley, M. B., et al. 2024, arXiv e-prints, arXiv:2405.21054, doi: 10.48550/arXiv.2405.21054

Barkana, R., & Loeb, A. 2004, ApJ, 609, 474, doi: 10.1086/421079

† https://bitbucket.org/rctirthankar/script

https://doi.org/10.1017/S174392132400334X Published online by Cambridge University Press

Barsode, A., & Choudhury, T. R. 2024, arXiv e-prints, arXiv:2407.10585, doi: 10.48550/arXiv.2407.10585
Bosman, S. E. I., Davies, F. B., Becker, G. D., et al. 2022, MNRAS, 514, 55, doi: 10.1093/mnras/stac1046
Bowman, J. D., Rogers, A. E. E., Monsalve, R. A., Mozdzen, T. J., & Mahesh, N. 2018, Nature, 555, 67, doi: 10.1038/nature25792
Choudhury, T. R. 2022, General Relativity and Gravitation, 54, 102, doi: 10.1007/s10714-022-02987-4
Choudhury, T. R., Mukherjee, S., & Paul, S. 2021a, MNRAS, 501, L7, doi: 10.1093/mnrasl/slaa185
Choudhury, T. R., & Paranjape, A. 2018, MNRAS, 481, 3821, doi: 10.1093/mnras/sty2551
Choudhury, T. R., Paranjape, A., & Bosman, S. E. I. 2021b, MNRAS, 501, 5782, doi: 10.1093/mnras/stab045
Choudhury, T. R., Paranjape, A., & Maity, B. 2024, J. Cosmology Astropart. Phys., 2024, 027, doi: 10.1088/1475-7516/2024/03/027
Gnedin, N. Y., & Madau, P. 2022, Living Reviews in Computational Astrophysics, 8, 3, doi: 10.1007/s41115-022-00015-5
Jain, D., Choudhury, T. R., Mukherjee, S., & Paul, S. 2023, MNRAS, 522, 2901, doi: 10.1093/mnras/stad1149
Jain, D., Choudhury, T. R., Raghunathan, S., & Mukherjee, S. 2024a, MNRAS, 530, 35, doi: 10.1093/mnras/stae748
Jain, D., Mukherjee, S., & Choudhury, T. R. 2024b, MNRAS, 527, 2560, doi: 10.1093/mnras/stad3277
Maity, B., & Choudhury, T. R. 2022a, MNRAS, 511, 2239, doi: 10.1093/mnras/stac182
—. 2022b, MNRAS, 515, 617, doi: 10.1093/mnras/stac1847
—. 2023, MNRAS, 521, 4140, doi: 10.1093/mnras/stad791
Maity, B., Paranjape, A., & Choudhury, T. R. 2023, MNRAS, 526, 3920, doi: 10.1093/mnras/stad2984
Mertens, F. G., Mevius, M., Koopmans, L. V. E., et al. 2020, MNRAS, 493, 1662, doi: 10.1093/mnras/staa327
Mukherjee, S., Paul, S., & Choudhury, T. R. 2019, MNRAS, 486, 2042, doi: 10.1093/mnras/stz1002
Paranjape, A., Choudhury, T. R., & Padmanabhan, H. 2016, MNRAS, 460, 1801, doi: 10.1093/mnras/stw1060
Paul, S., Mukherjee, S., & Choudhury, T. R. 2021, MNRAS, 500, 232, doi: 10.1093/mnras/staa3221
Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, A&A, 641, A6, doi: 10.1051/0004-6361/201833910
Reichardt, C. L., Patil, S., Ade, P. A. R., et al. 2021, ApJ, 908, 199, doi: 10.3847/1538-4357/abd407
Singh, S., Jishnu, N. T., Subrahmanyan, R., et al. 2022, Nature Astronomy, 6, 607, doi: 10.1038/s41550-022-01610-5
Trott, C. M., Jordan, C. H., Midgley, S., et al. 2020, MNRAS, 493, 4711, doi: 10.1093/mnras/staa414

https://doi.org/10.1017/S174392132400334X Published online by Cambridge University Press