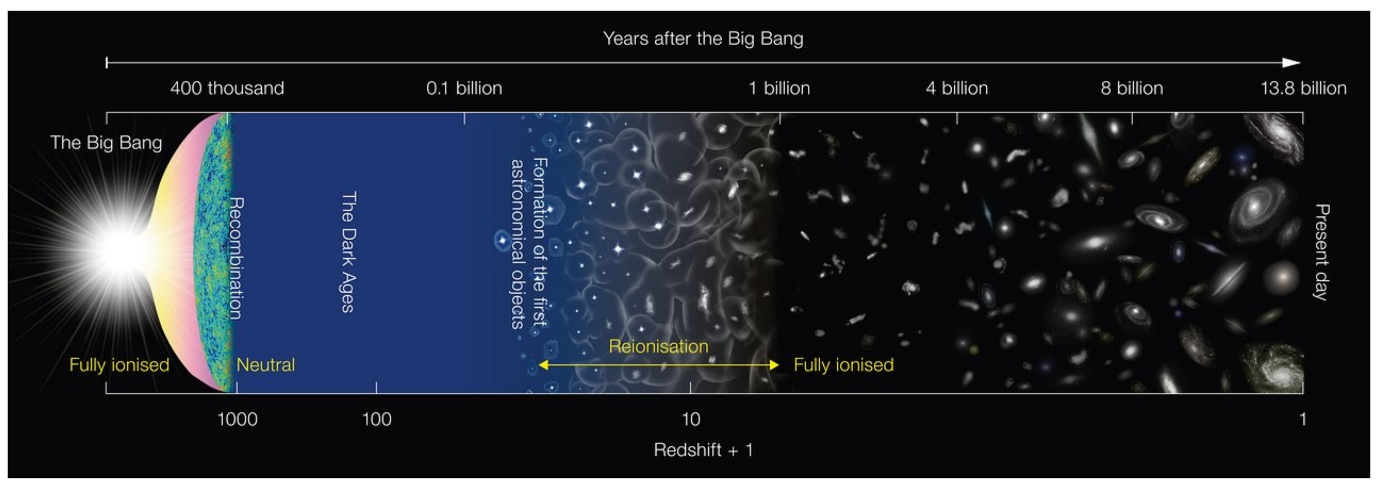
Diffusion Models to Infer the Density Fields from SKA-Low Maps

Pietro Albanese Guidi[1], Adélie Gorce[1], David Cornu[2], Benoit Semelin[2]
[1] Université Paris-Saclay, CNRS, Institut d'Astrophysique Spatiale, 91405 Orsay, France
[2] Observatoire de Paris, PSL Research University, Sorbonne Université, CNRS, LUX, 75014 Paris, France

Co-funded by the European Union

1 Scientific Context

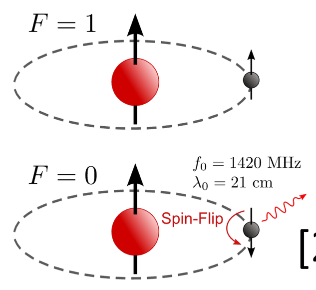
The Epoch of Reionization (EoR) is the last major phase change of the universe. It is expected to have occurred between z = 14 and z = 6.
During this period, galaxies and stars emit photons which can ionize the InterGalactic Medium (IGM) and create bubbles of ionized gas around them. As time passes, the size of the bubbles increases and this process continues until the entire IGM is ionized.



[1]

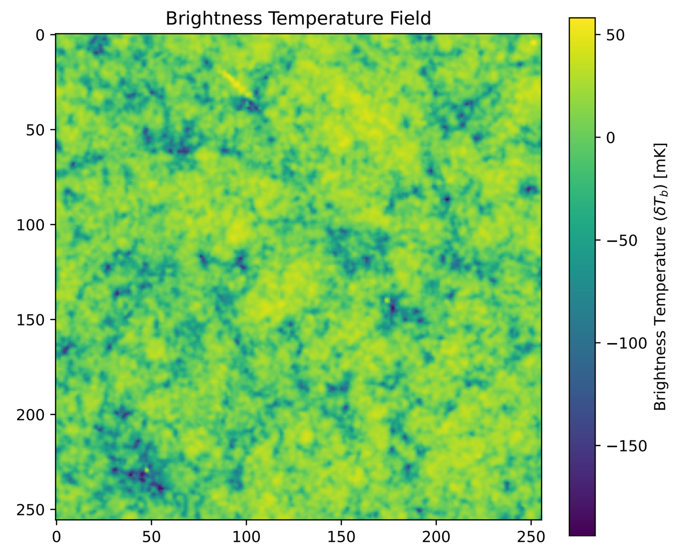
2 21 cm Line & Simulations

The EoR can be observed by detecting the 21 cm line, which is emitted by a spin flip in the hyperfine structure of neutral hydrogen (HI). The brightness temperature of the 21 cm line is proportional to the baryon density and the neutral hydrogen fraction.



[2]

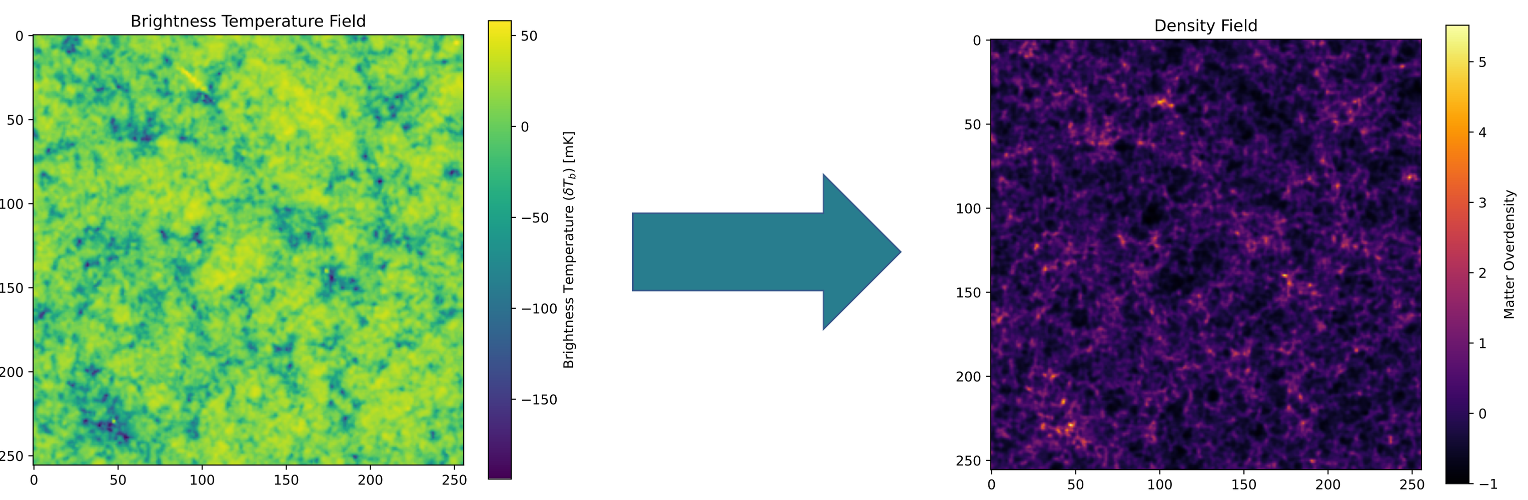
The 21 cm maps and matter density fields used in this project are part of the Loreli dataset and are generated with a numerical simulation. This dataset contains more than 9000 simulations.
The maps are cubes with 256^3 pixel, with a side length of 300 Mpc and at redshifts between z = 53.6 and z = 4.97 [3].
In this project, we use simulations at z = 10.
SKA-Low is expected to measure the 21 cm line and provide maps of its brightness temperature.



3 Objective: Field-Level Inference

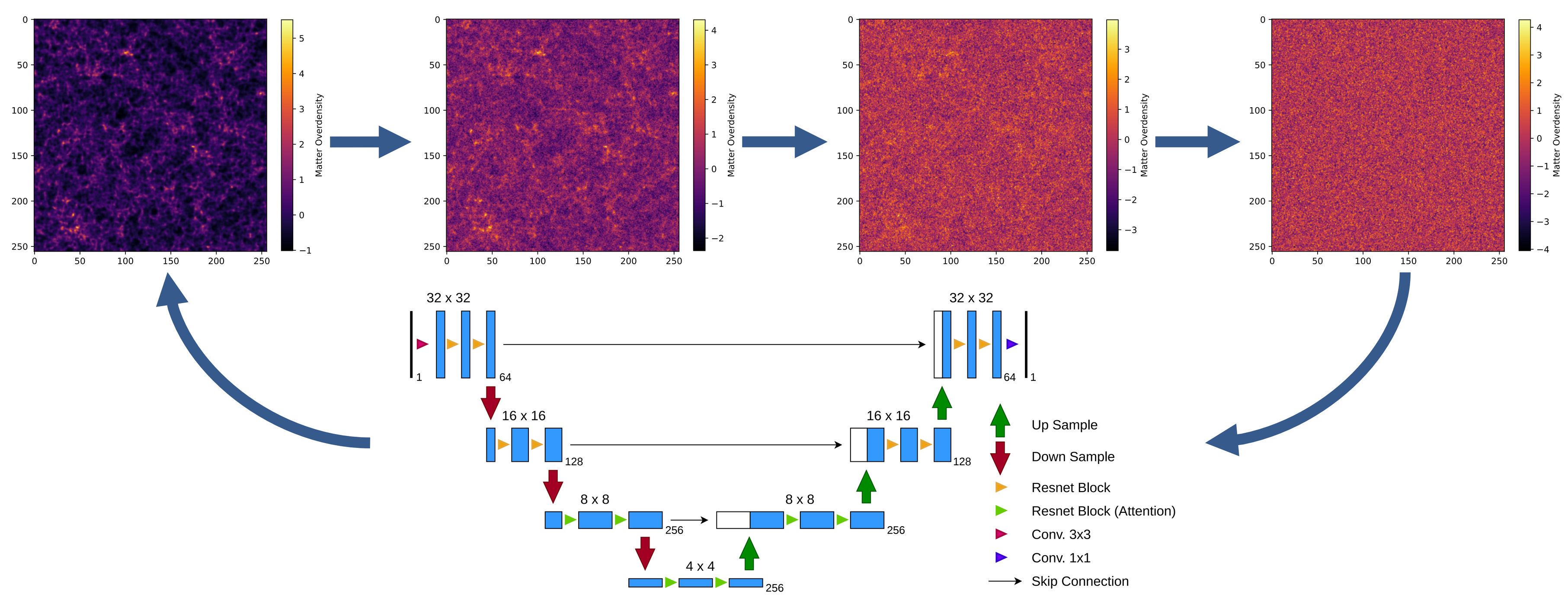
Field-Level Inference is a method used to infer cosmological parameters and/or physical fields from observed data or simulations without first compressing them with summary statistics [4].

Objective: Reconstruct the matter density field from the simulated 21 cm maps.



4 Diffusion Models

Diffusion Models are a generative AI model which generates data by denoising Gaussian noise. It is trained on images with different amounts of Gaussian noise with the goal of predicting the original image and the noise.

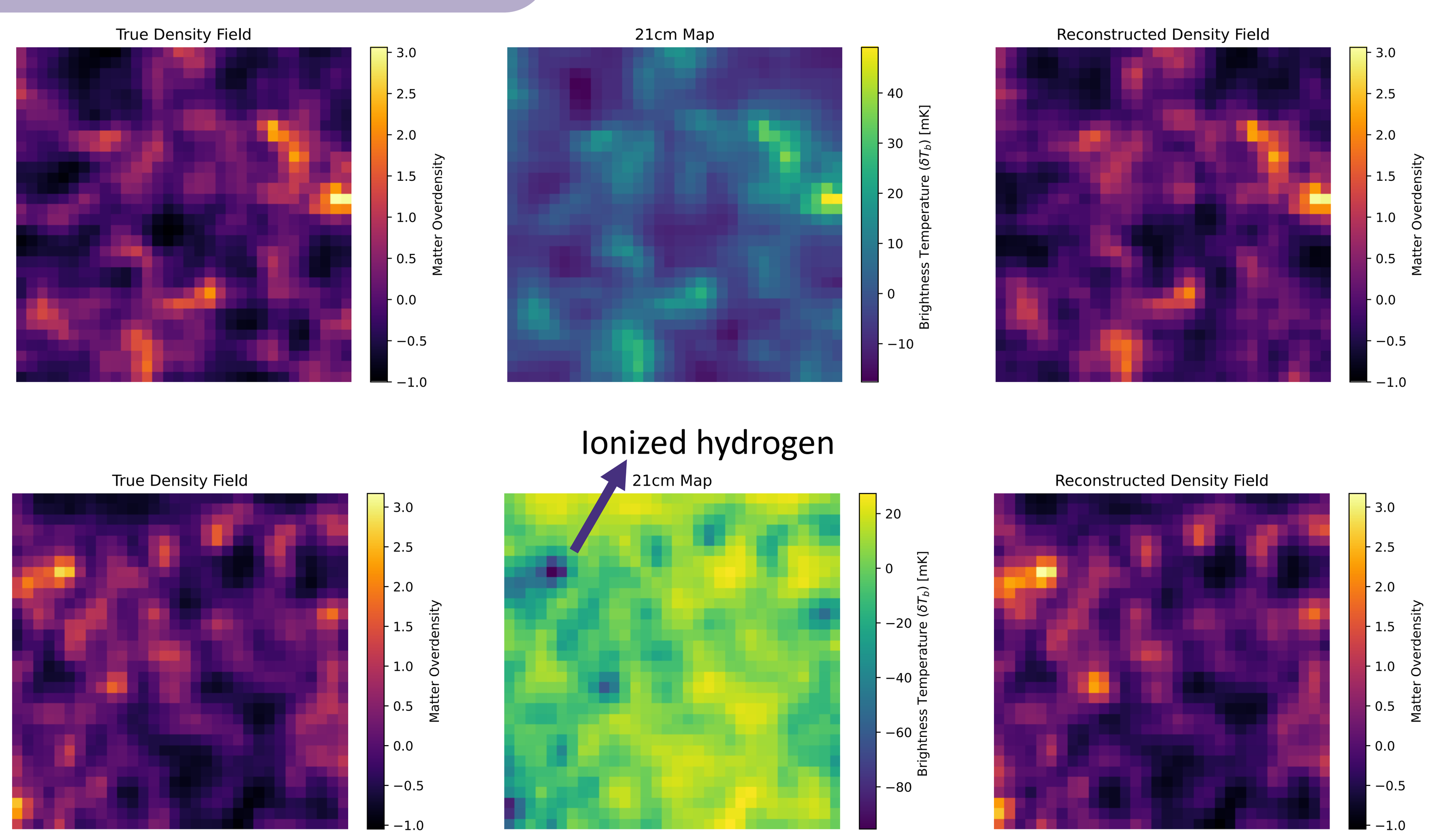


5 Results

1) Generated Samples:

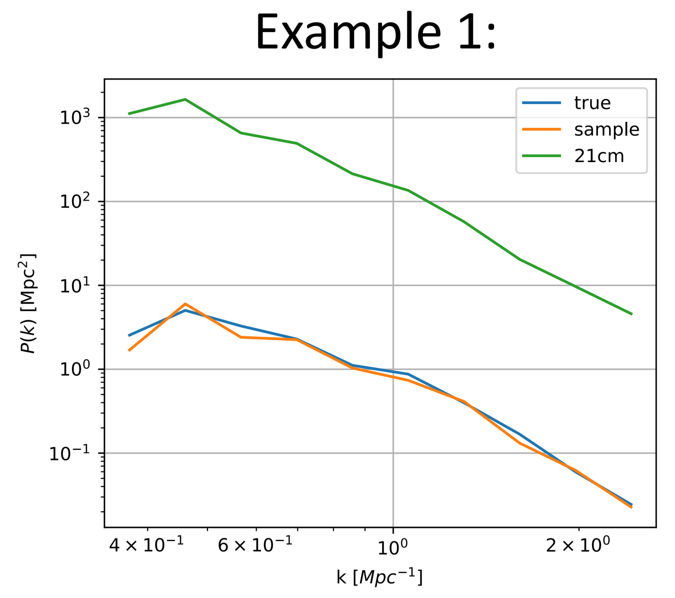
Example 1:

Example 2:

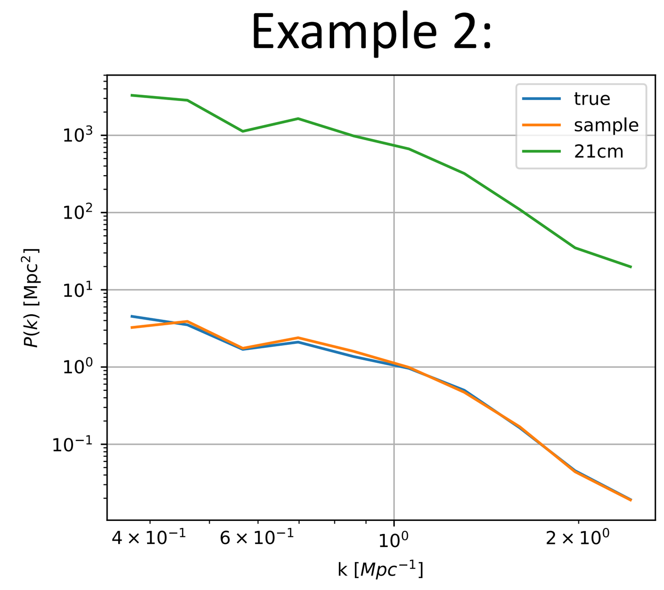


2) Power Spectrum:

Example 1:

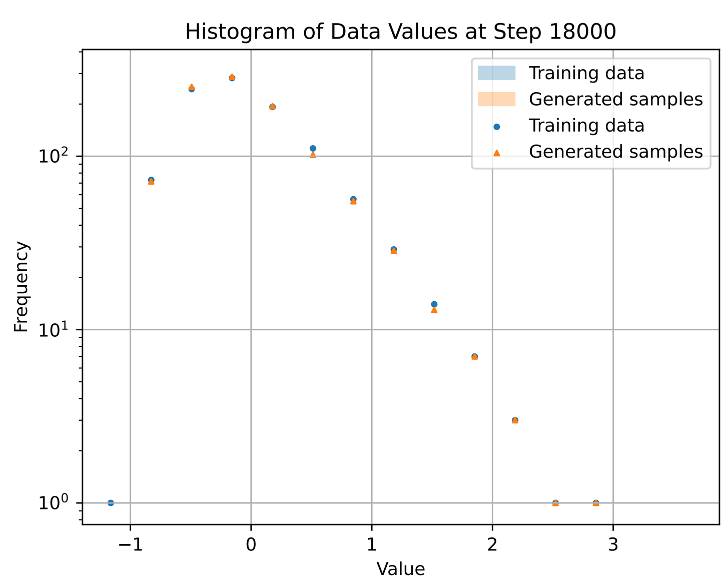


Example 2:

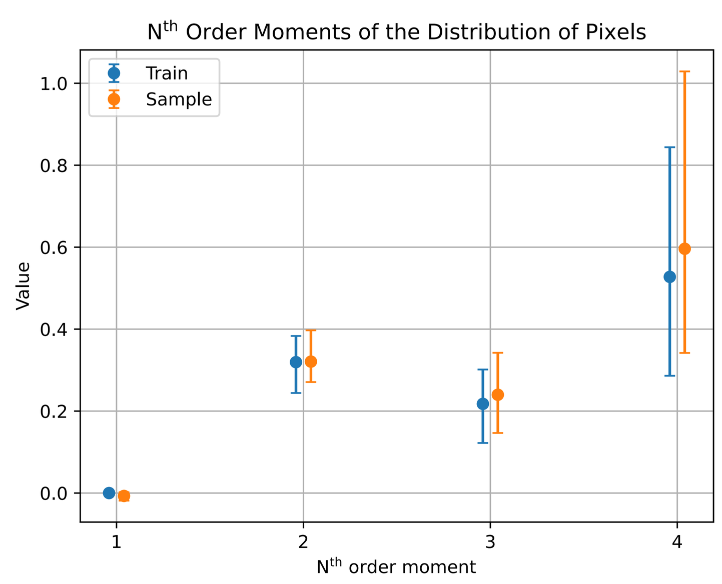


The power spectrum of the generated density fields matches the true power spectrum.

3.1) Distribution of Pixels:



3.2) N^{th} Order Moments:



The distribution of pixels of the generated density fields matches the one of the training set. This can also be verified by looking at the first four order moments.

6 Conclusion & Next Steps

Conclusion:

- The diffusion model can reconstruct the matter density field even in the presence of reionization, although not perfectly.
- The power spectrum of the reconstructed density field closely matches the one of the true density field.
- The distribution of pixels is also well recovered.

Next Steps:

- The reconstruction is still not perfect and some changes to the model architecture are probably necessary.
- Increasing the number of pixels from 32x32 to 256x256.
- Using metrics such as precision and recall to evaluate the performance of the model.
- Training diffusion models on 21 cm maps located at lower redshifts where there is more reionization.

References:
[1] Credits: NAOJ
[2] https://commons.wikimedia.org/wiki/File:Hydrogen-SpinFlip.svg
[3] R. Meriot, B. Semelin, *The LoReLi database: 21 cm signal inference with 3D radiative hydrodynamics simulations* (2023)
[4] F. Leclercq, *Field-level inference in cosmology* (2025)