Cosmology and Statistics Days - 01/02/2024

📍CEA Paris-Saclay, France

A talk on Titan

"Morphology and spatial distribution of high-redshift dust emission using component separation and deep learning"

Arnab Lahiry

Ph.D candidate, FORTH

Supervised by
Dr. Jean-luc Starck (CEA Paris-Saclay)
Dr. Tanio Diaz-Santos (IA - FORTH)

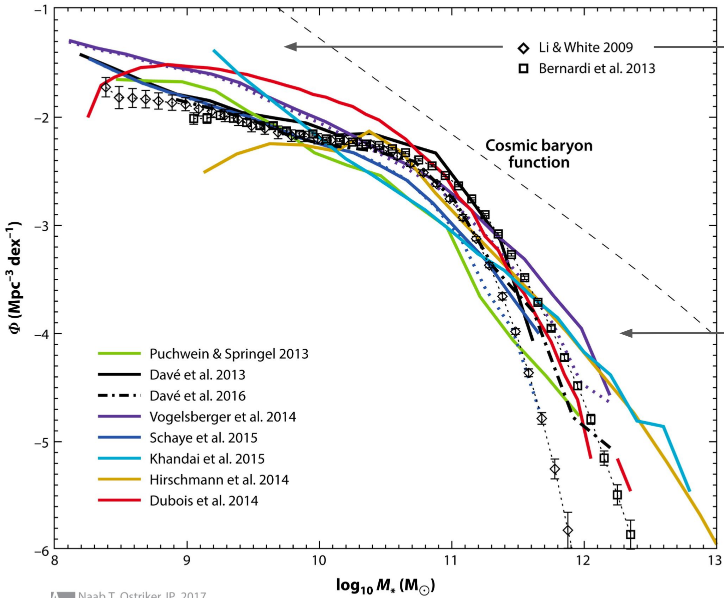
TITAN ARTIFICIAL INTELLIGENCE IN ASTROPHYSICS · FORTH · Signal Processing Laboratory · IA · CosmoStat · cea

A disparity between ΛCDM predictions and observations:

Observations show that there is a lack of low mass and high mass galaxies with respect to number of DM halos

Comparison of galaxy stellar mass functions from recent large-scale cosmological simulations of representative volumes of the Universe. The simulations include stellar and AGN feedback with the exception of Davé et al. (2013), who use an empirical heating model in massive halos. The different groups typically adjust the key parameters in the varying sub-resolution models to match observations of galaxy mass functions

The dashed line for Vogelsberger et al. (2014) and Schaye et al. (2015) indicate different mass estimates. **The dashed line shows the hypothetical galaxy mass function assuming the cosmic baryon fraction**.



Naab T, Ostriker JP. 2017.
Annu. Rev. Astron. Astrophys. 55:59–109

Disparity in the number of observed **low mass** galaxies

Disparity in the number of observed **high mass** galaxies

A disparity between ΛCDM predictions and observations:

Observations show that there is a lack of low mass and high mass galaxies with respect to number of DM halos

In the **local universe**, these phenomena are understood

- **For low mass galaxies:**
  Baryonic processes such as feedback from supernovae, and stellar winds in small galaxies can expel cold gas necessary for star formation, hench quenching them, leading to less observations.

- **For high mass galaxies:**
  Energetic quasar feedback causes the quenching of star formation in galaxies, leading to the observed rarity of ultra-massive galaxies

But what about for earlier epochs?

The mechanisms of quasar and SMBH accretion activities are **still unclear** at higher redshifts - beyond $z > 3$, as the gas and dust had much different physical conditions

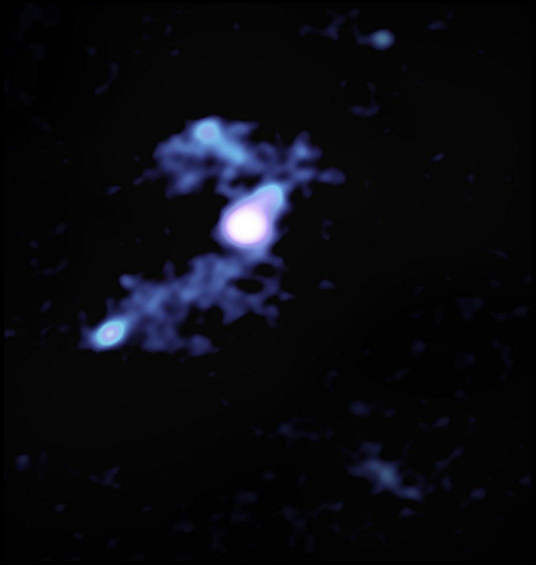
Broad goal

To perform a morpho-kinematic analysis of high-redshift quasar sources and their role in the overall galaxy evolution timeline

What are hot dust-obscured galaxies (Hot DOGs)?

The end goal is to focus on data from a **high redshift, hot dust-obscured galaxy population**, undergoing multiple merger events

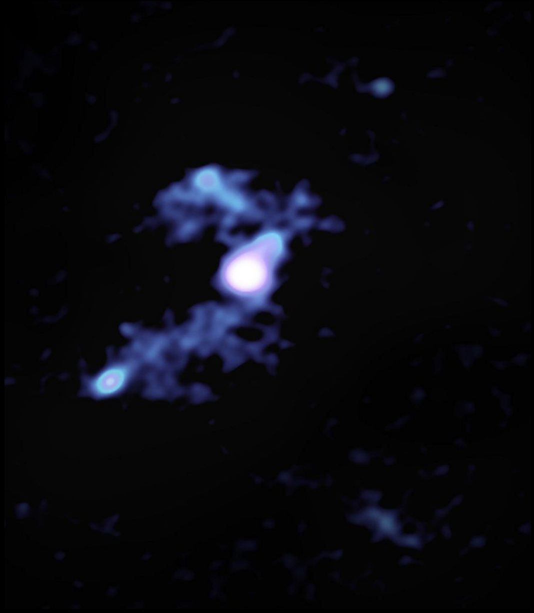
- Population of **hyper-luminous quasars** initially identified by WISE
- **Very high** bolometric luminosity ($L_{bol} > 10^{14} L_{\odot}$)
- **Heavy dust obscuration** leading to the re-emission of stellar and AGN radiation in the infrared - Intense mid-IR emission (~ 24 μm)
- This population gets its name from the high temperature of the surrounding dust particles
- Will be focusing on the multiple merger system around W2246-0526, also known as **W2246** (image on the right)
- Thought to represent a stage in galaxy evolution where there is simultaneous occurrence of intense star-formation and AGN activity



Composite image of **W2246** generated from ALMA data, showing the central galaxy undergoing mergers with > 2 neighbouring sources. Also depicting morphological features like tidal tails and bridges

An artist's illustration of the multiple-merger system around **W2246** showing the different morphological components

Source: (NRAO/AUI/a) S. Dagnello)



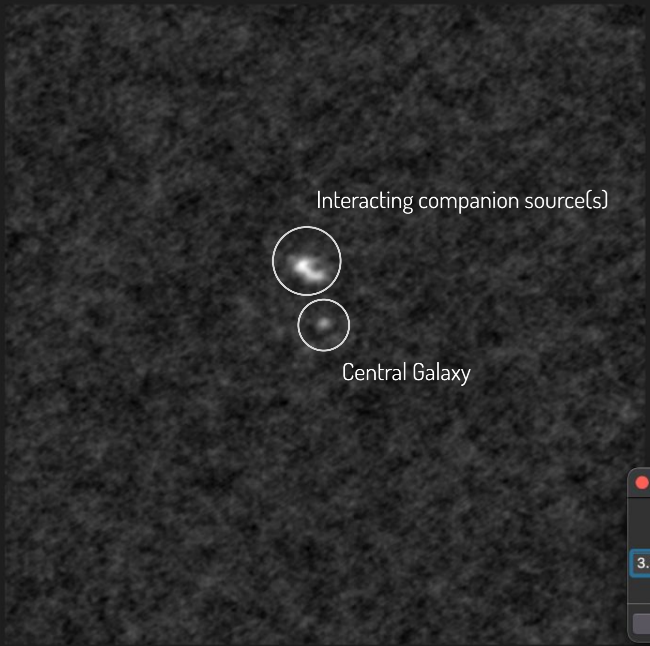
Composite image of **W2246** generated from ALMA data, showing the central galaxy undergoing mergers with > 2 neighbouring sources. Also depicting morphological features like tidal tails and bridges

An artist's illustration of the multiple-merger system around **W2246** showing the different morphological components

Source: (NRAO/AUI/a) S. Dagnello)

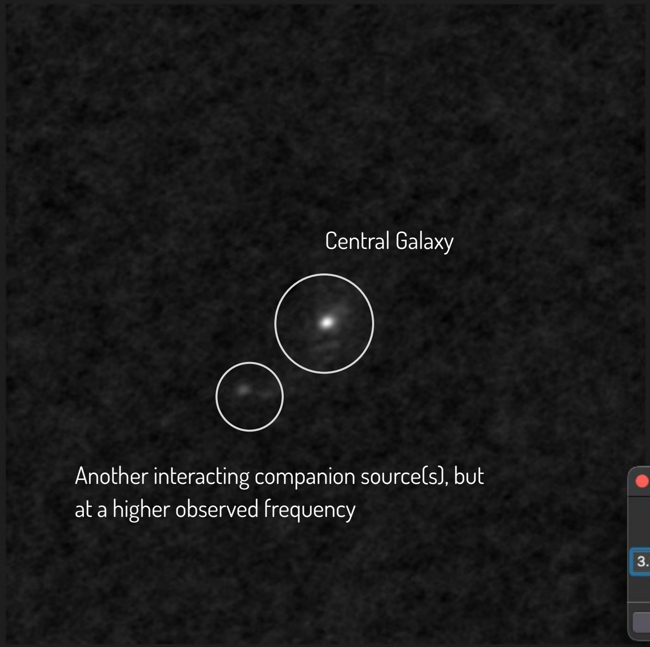
Another artist's rendition of the central galaxy **W2246** showing the bright and dust-obscured centre

Source: (JPL/NASA)



Preliminary visualisation of one slice of the (continuum subtracted) CII emission line ALMA datacube.

Dimensions :
600 X 600 (spatial) **X 125** (spectral)



Preliminary visualisation of one slice of the (continuum subtracted) CII emission line ALMA datacube.

Dimensions :

600 X 600 (spatial) **X 125** (spectral)

Tentative Goals

Different sources can be (and have been) qualitatively identified (visually), and as evident in the previous 2 slides, some interacting sources are at different velocities.

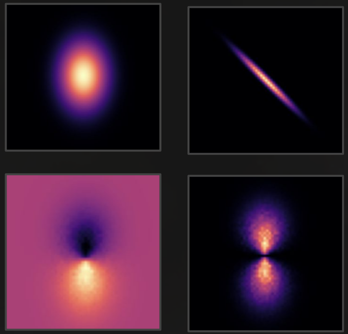
Next steps in this direction:

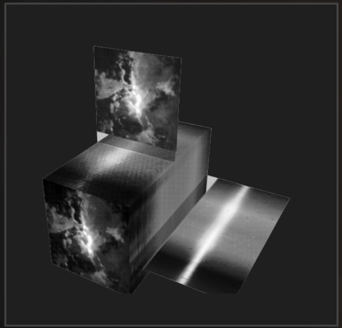
Employing **novel blind source separation algorithms** on the ALMA datacubes

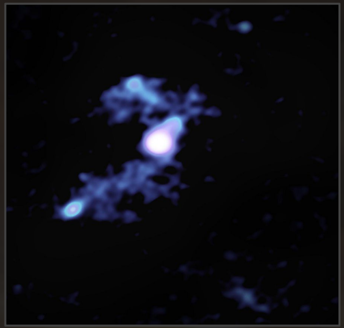
- More quantitative identification of different sources, allowing for the morphological and kinematic analysis of the same
- Better understanding of the dynamics of the interacting system at high redshift

Working with accurate zoom-in **simulations** to **identify the galaxy evolutionary stage**, the observed high-redshift quasars belong to, by employing **machine learning algorithms**.

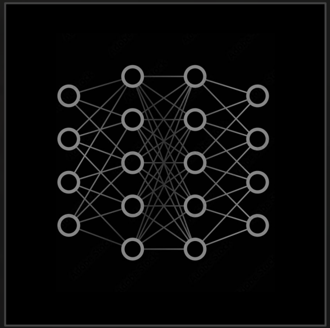
Methodology & project plan







Blind Source Separation



1. Generating Simple simulations made from scratch with various controllable parameters

2. Using simulated spectral cubes generated from state-of-the-art simulations, with same parameters as observational datacubes from ALMA

3. Using spectral Observational datacubes in the infrared regime obtained from the ALMA telescope

4. Performing principal component analysis, and blind source separation methods on data from (1), (2) and (3) to separate out components

5. Developing machine/deep learning algorithms to identify the stage of the source in the overall galaxy evolution timeline

Tentative deep learning flowchart

Simulation data
(producing datacubes with same attributes as ALMA cubes)
FIRE: Feedback In Realistic Environments

Creating a **Training dataset** to train machine learning model

Building a **deep learning model** to analyse the identified morpho-kinematic features and predict the galaxy evolutionary stage

Testing the model results with simulated datacubes

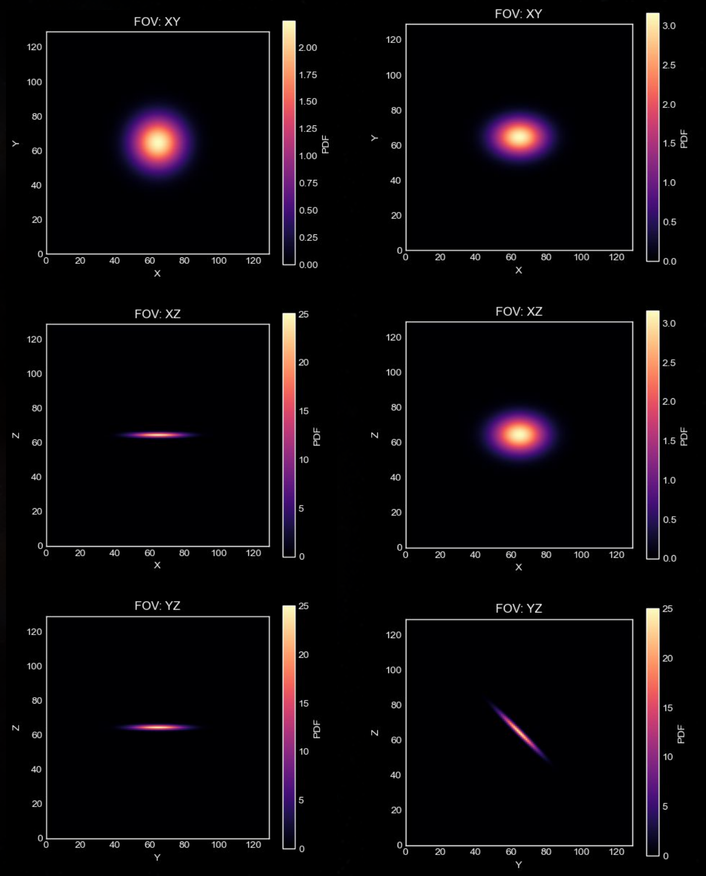
Using this trained model to identify the stage of the observed sources in the overall galaxy evolution timeline

(As simulations are theoretically generated, they would allow us to know the evolution of a source through cosmic time, providing a simulated picture of the galaxy evolution timeline, allowing the possibility for supervised learning for predictions)

Illustration

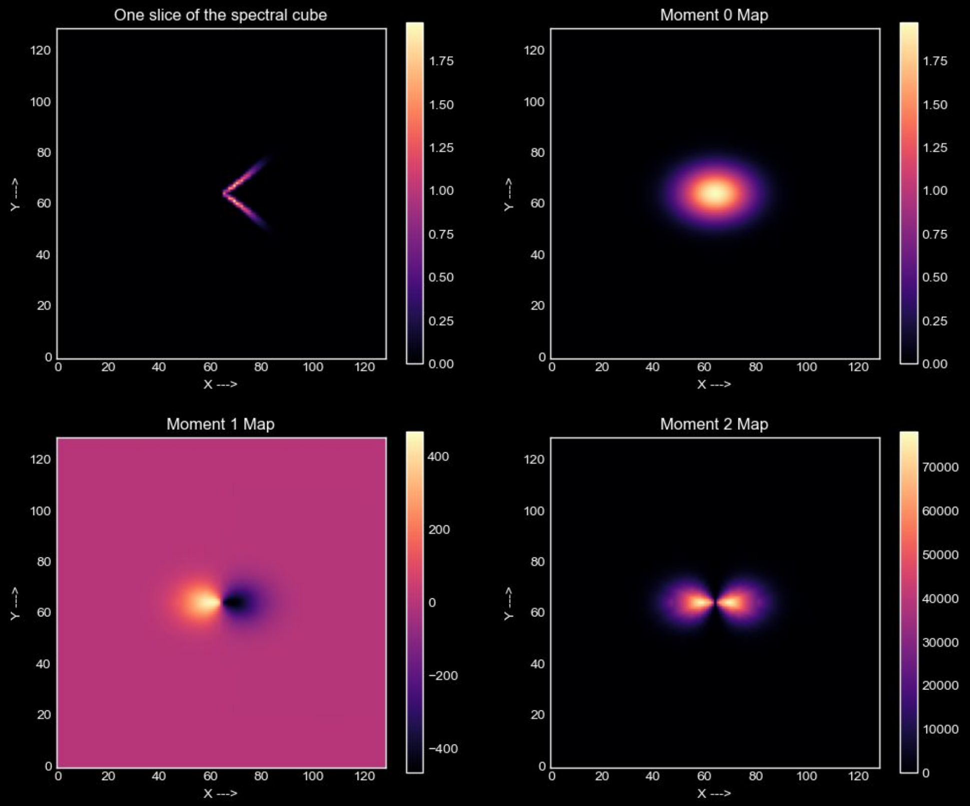
Ongoing work

- (Lots and lots of reading)
- Creating a simple datacube simulation from scratch
  - Defining a flat disk in 3D space using a Gaussian
  - Calculating and assigning the rotational velocity for each particle in the disk
  - Rotating the entire system along different axes of rotation
  - Forming a spectral cube with 2 spatial axes and one spectral (velocity) axis, where each *slice* contains information about the points within that particular velocity bin, integrated along a chosen line of sight
- Creating a composite datacube (no. of disks >1)
- Applying principal component analysis to express this data in terms of eigenvalues and eigenvectors (principal components)



Current progress

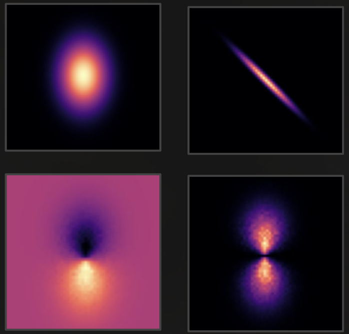
Rotated single spectral cube and moment maps

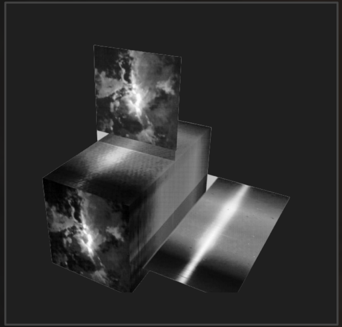


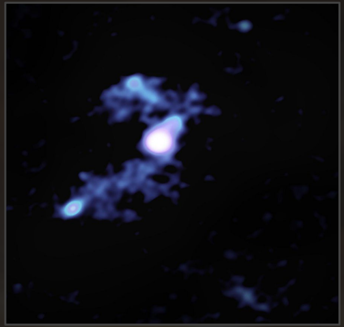
Next steps

- Combining >1 disks
- Applying PCA
- Making the simulations better (bulges, exponential disks)
- Working with more accurate state-of-the-art simulations

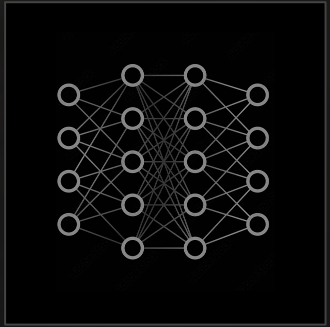
Methodology & project plan (again)(summary)







Blind
Source
Separation



1 Generating Simple simulations made from scratch with various controllable parameters

2 Using simulated spectral cubes generated from state-of-the-art simulations, with same parameters as observational datacubes from ALMA

3 Using spectral Observational datacubes in the infrared regime obtained from the ALMA telescope

4 Performing principal component analysis, and blind source separation methods on data from (1), (2) and (3) to separate out components

5 Developing machine/deep learning algorithms to identify the stage of the source in the overall galaxy evolution timeline

Thank you