MACHINE LEARNING FOR ASTROPHYSICS
2ND EDITION CATANIA, 8-12 JULY, 2024

11/07/2024

Catania, Italy

Morphology and spatial distribution of high-redshift galaxy gas and dust emission using source identification and deep learning

ARNAB LAHIRY

PhD Student
IA / ICS -FORTH

Supervised by:

Dr. Jean-Luc Starck
CEA Paris-Saclay / IA-FORTH

Dr. Tanio Diaz-Santos
IA - FORTH

TITAN ARTIFICIAL INTELLIGENCE IN ASTROPHYSICS

FORTH INSTITUTE OF COMPUTER SCIENCE

Signal Processing Laboratory

CosmoStat

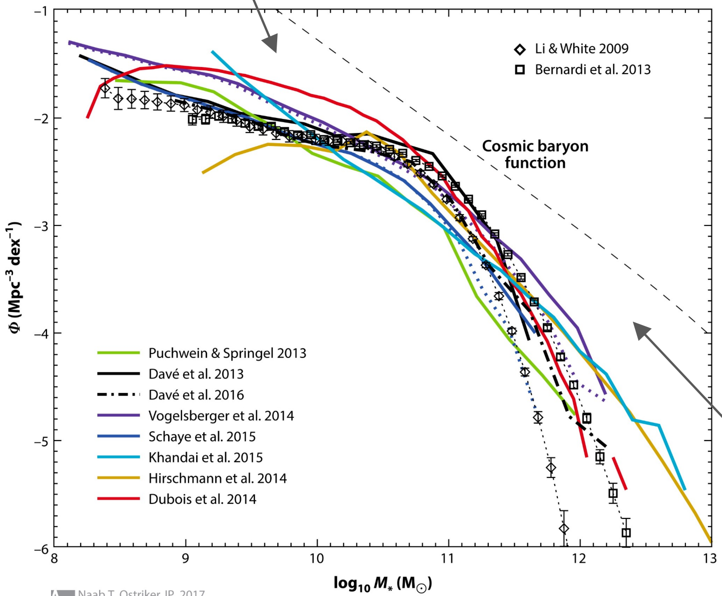
cea

Funded by the European Union

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

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



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

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

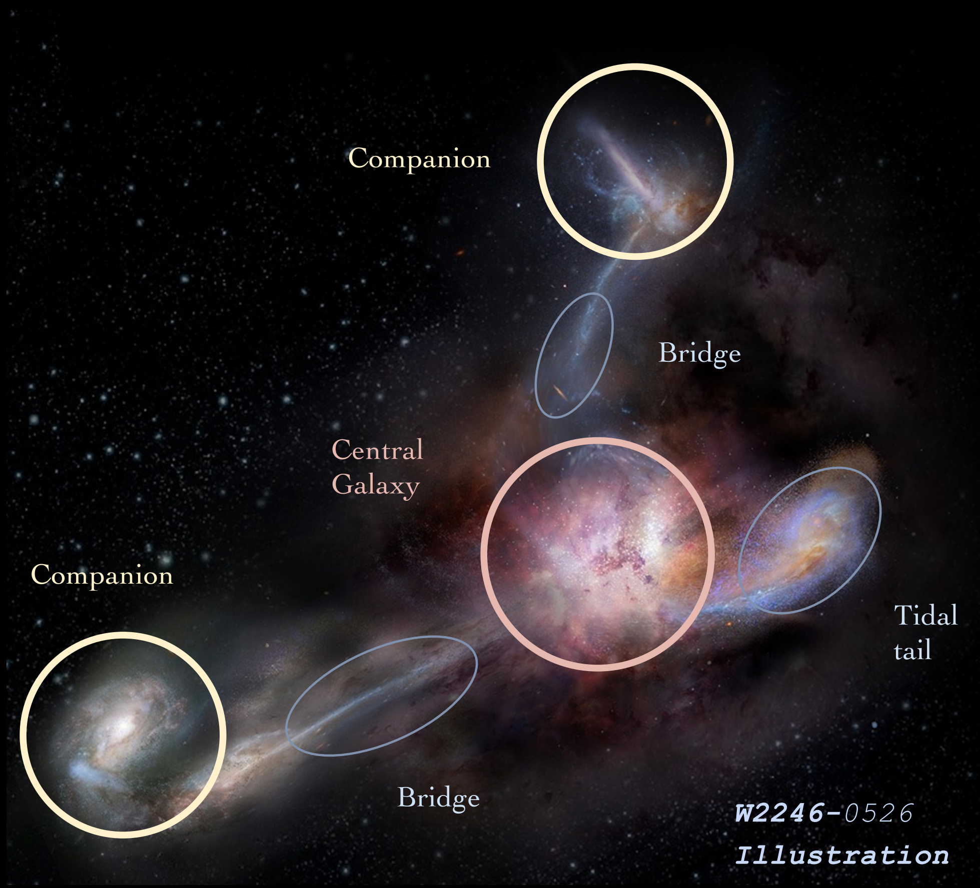
In the **local universe**, these phenomena are understood

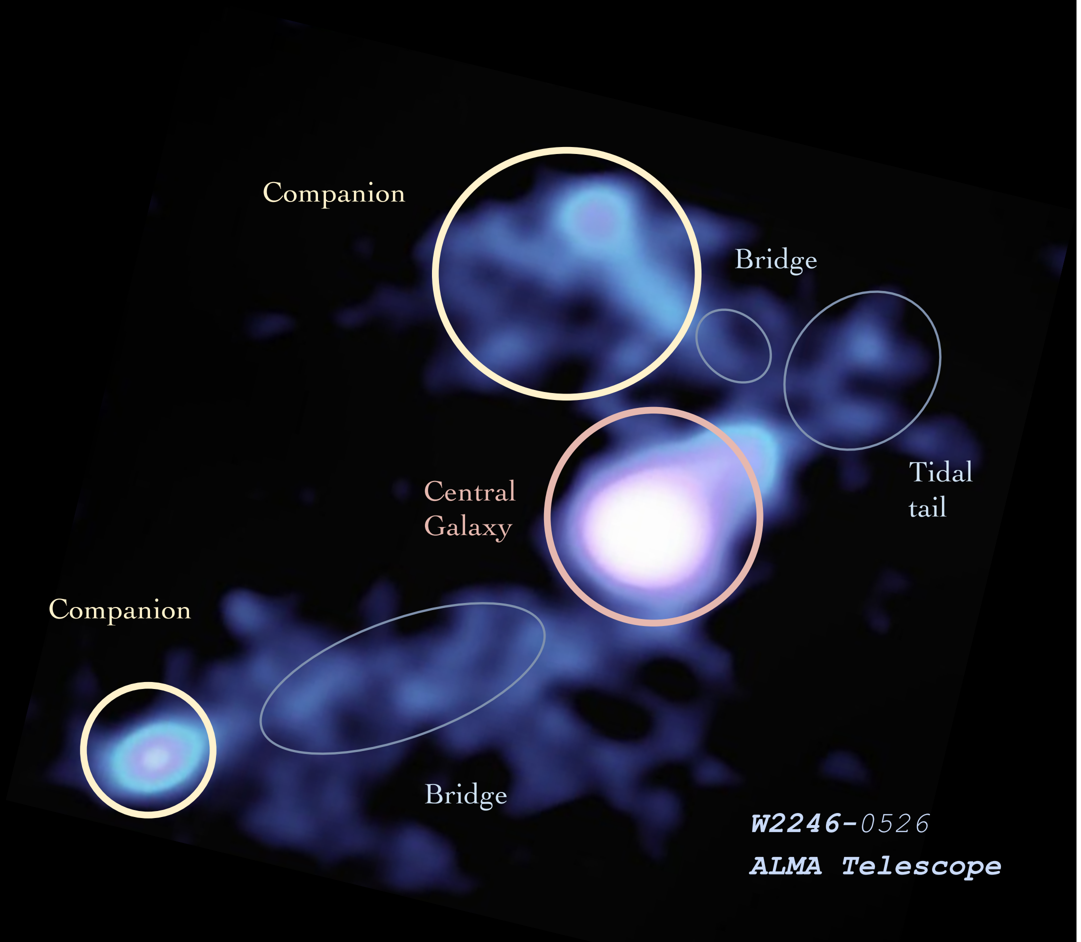
- **For low mass galaxies:**
  Baryonic processes can expel cold gas necessary for star formation, hence quenching them, leading to less observations.
- **For high mass galaxies:**
  Energetic quasar feedback causes the quenching of star formation in galaxies, leading to the observed rarity of ultra-massive galaxies

But what about for earlier epochs?

The mechanisms of quasar and SMBH accretion activities are still unclear at higher redshifts - as the gas and dust had much different physical conditions

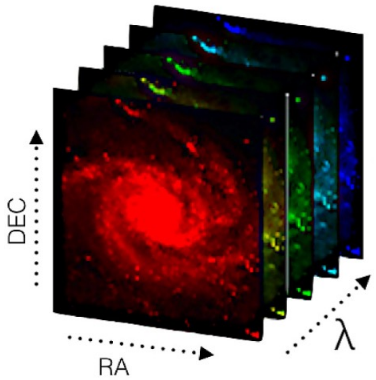
Analysis of a high-redshift (z = 4.6) Hot DOG system





TITAN ARTIFICIAL INTELLIGENCE IN ASTROPHYSICS
FORTH INSTITUTE OF COMPUTER SCIENCE
Signal Processing Laboratory
CosmoStat
cea
Funded by the European Union

Image source: H. Farias, C. Nuñez, M. Solar, *TensorFit a tool to analyse spectral cubes in a tensor mode*



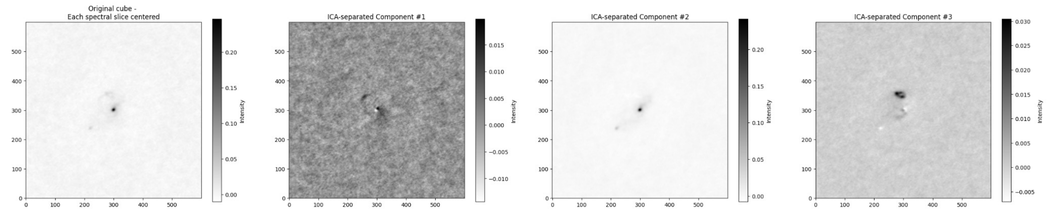
Spectral cube

Each spectral 'slice' corresponds to a spectral 'observation'

Therefore, in data matrix Y, each column corresponds to one spectral observation, and each row has information from the flattened 2D slice

Attempting to separate out sources using ***blind source separation***

FastICA using `scikit-learn`



Original Input data is noisy

One of the sources is mostly noise

The central galaxy and the companion with similar spectra are considered together in one source

ICA is able to identify the other companion with separate spectra

$$\mathbf{Y} = \mathbf{A} \cdot \mathbf{S} + \mathbf{N}$$

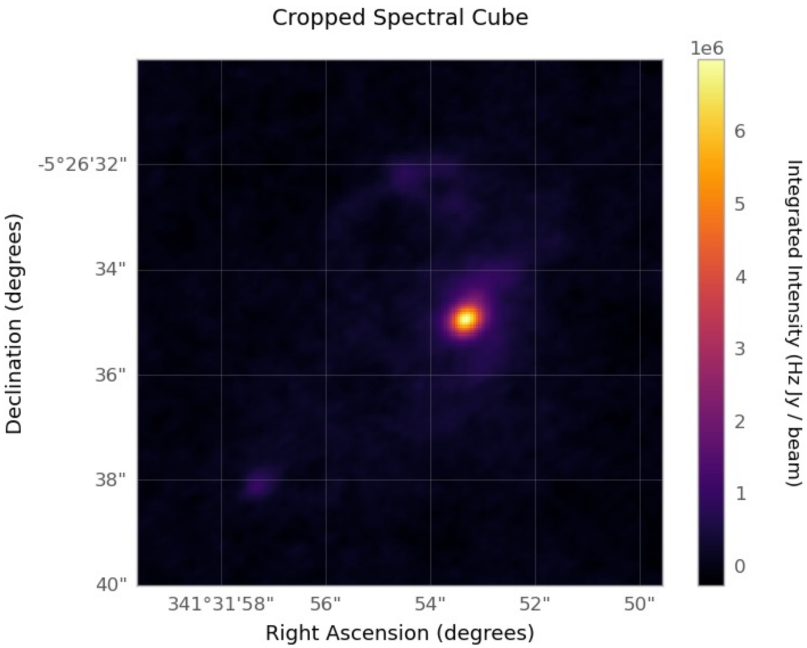
Y	A	S	N
Mixed data which is a mixture of sources	Mixing Matrix	Source Matrix	Noise
Each column is an observed signal	**[UNKNOWN]**	**[UNKNOWN]**	

Blind Source Separation is **unable to separate out sources with similar spectral signatures.**

Furthermore there are sources which appear in different spatial locations at different spectral observations, suggesting kinematically distinct sources which cannot be separated by BSS

TITAN Artificial Intelligence in Astrophysics
FORTH Institute of Computer Science
Signal Processing Laboratory
CosmoStat
cea
Funded by the European Union

Starlet transform based source identification



Moment 0 map of the CII (continuum subtracted) emission line map from ALMA

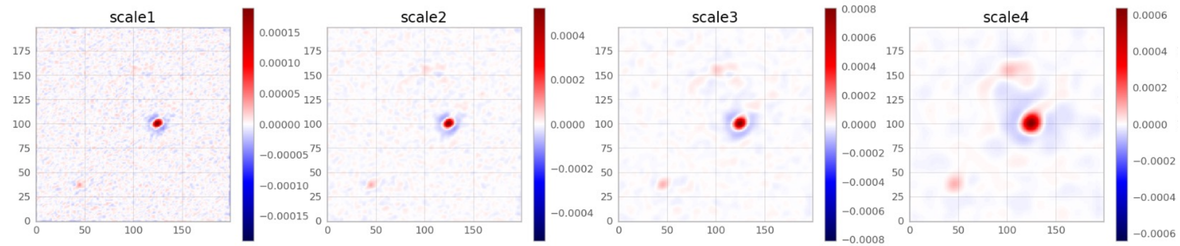


Fig A: Different wavelet scales for different detail levels

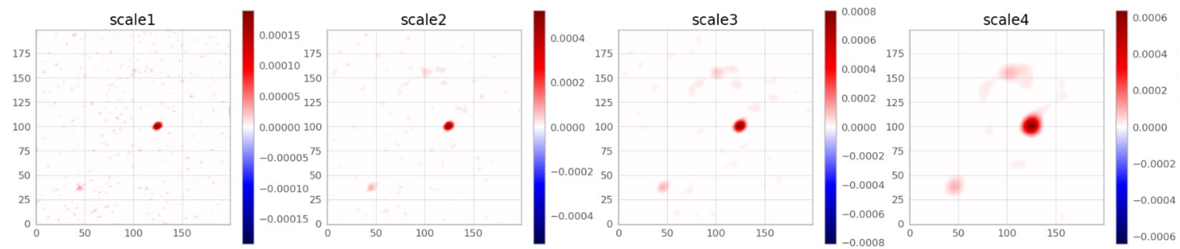
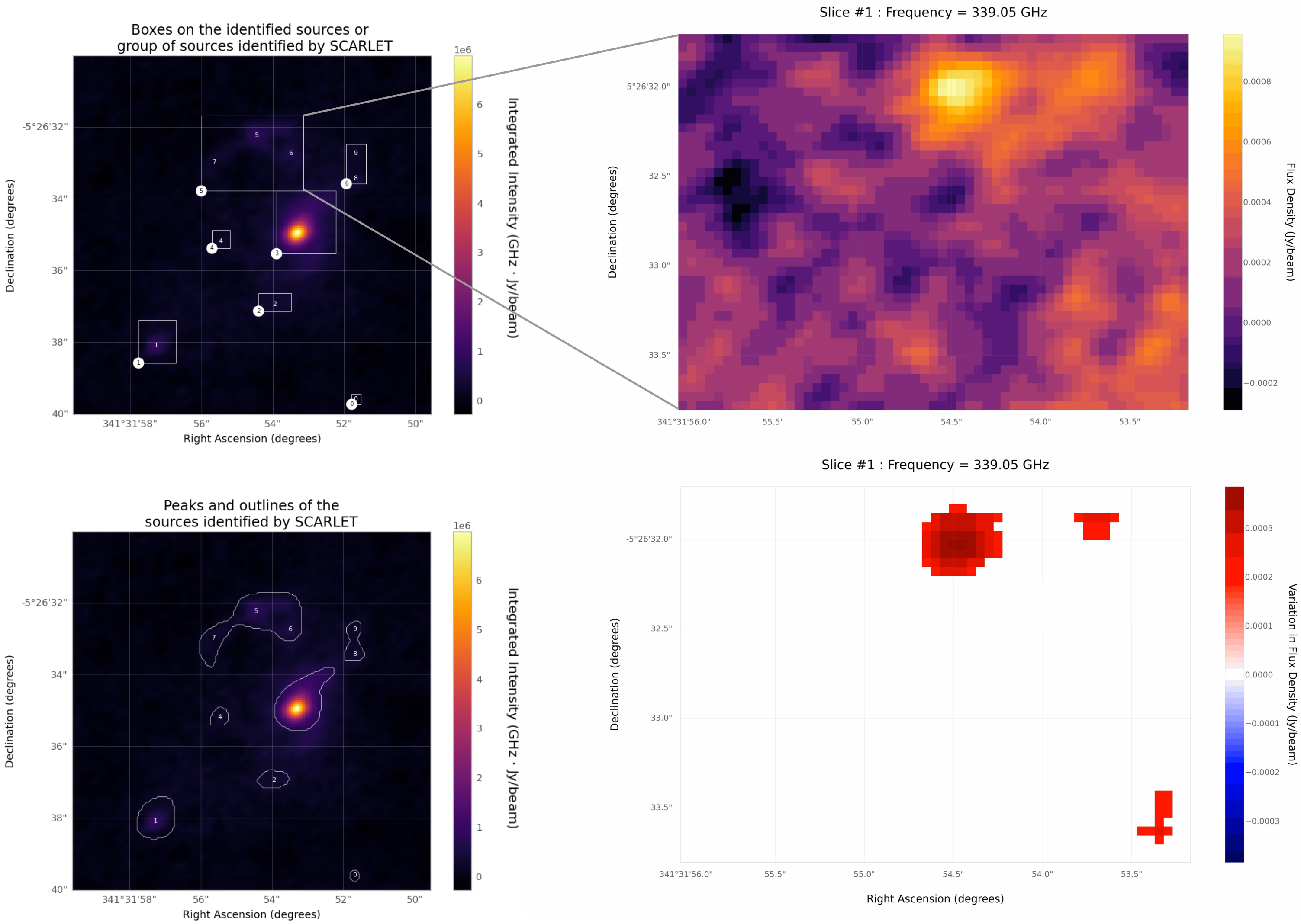


Fig B: "Detections" upon applying a non-negativity constraint and thresholding on the wavelet scales

Methodology performed using SCARLET software (P. Melchior et al.)



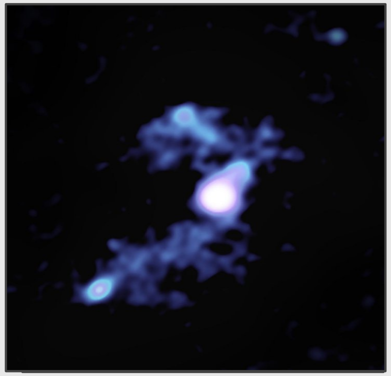
Observations:

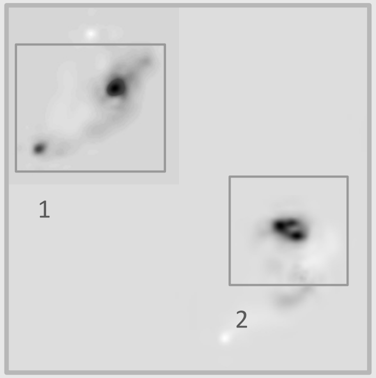
- In box #5, SCARLET has detected 3 peaks from the mean map of the cube, which may suggest the presence of multiple sources
- Upon slice-wise analysis of box #5, we observe that the the spatial positions of 'sources' change upon different frequency observations.
- This suggests that the point of interest may be a single source which is kinematically active, leading to different spatial locations at different frequencies.
- SCARLET-based detections were performed on each slice to better visualise the flux-density peaks

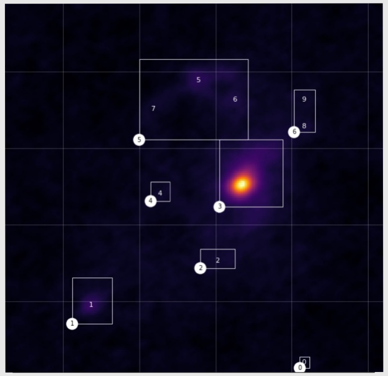
Ongoing work:

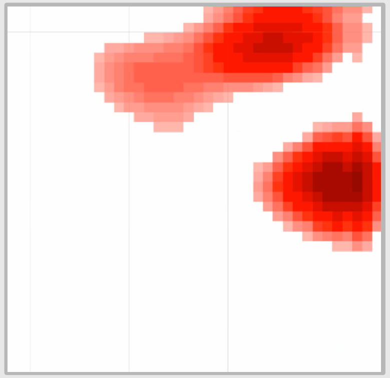
- Classifying whether detections with multiple peaks are a single or multiple sources quantifiably (constraints on the spatial movement per velocity, etc.)
- Implementing method on simulated spectral cubes (FIRE)
- Creating software to automate and make the process more efficient.

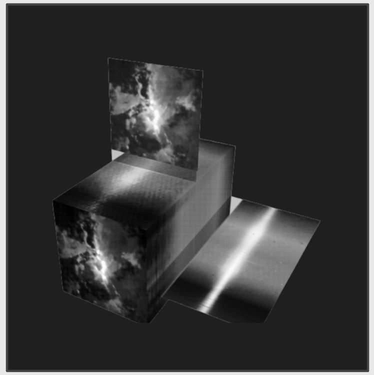
Summary (and thank you for listening :))











1 Preprocessing the observational spectral cube and applying preliminary component identification methods like PCA

2 Source separation methods - ICA and GMCA to separate out sources with different spectra

3 Using SCARLET and wavelet based peak identifications to detect possible individual sources and regions of interests

4 Analysing the kinematics of the detected peaks in regions of interest for each spectral slice to understand the physical dynamics of sources

5 Implementing this method with simulated spectral cubes from the FIRE simulations, and using deep learning to improve results

TITAN ARTIFICIAL INTELLIGENCE IN ASTROPHYSICS
FORTH INSTITUTE OF COMPUTER SCIENCE
Signal Processing Laboratory
CosmoStat
cea
Funded by the European Union