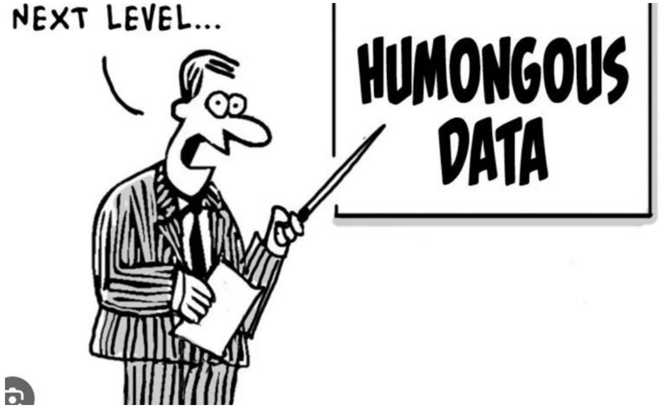
AI, Astrostat and Big Data Challenges for the ALMA2040 Era

Fabrizia Guglielmetti (ESO)

From ALMA WSU to ALMA 2040:
Telescope capability increases

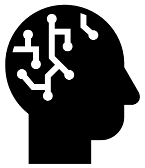


Data Volume

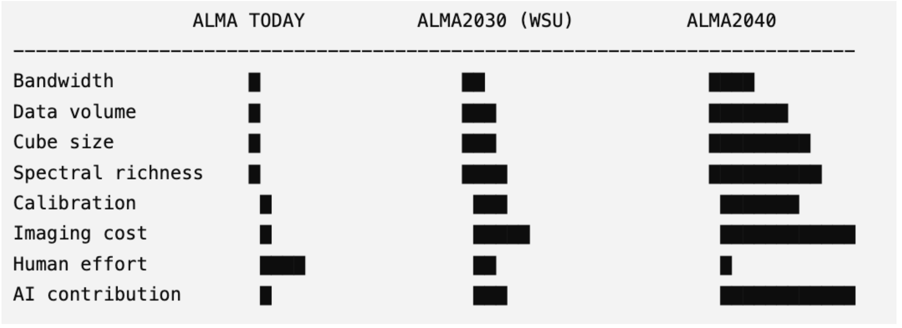
Scientific and Computing complexity

More users

Current reduction paradigm : PL and/or manual QA + CLEAN + inspection



Bayesian inference + AI : BRAIN framework



Is Big Data a 21st-century problem?

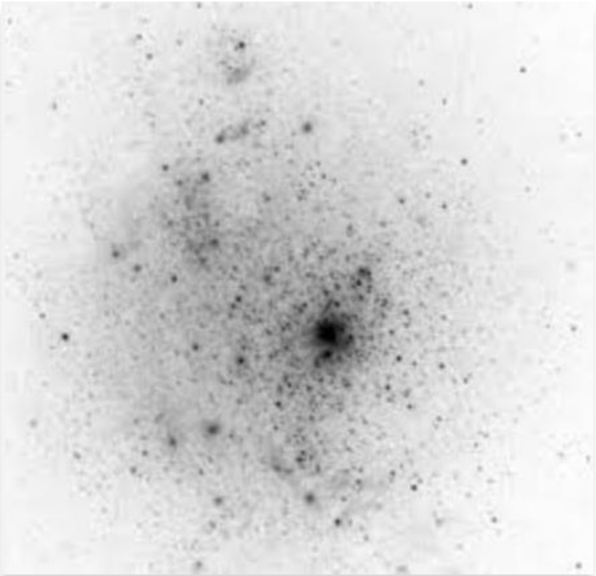
Big Data challenges triggered modern data analysis

A brief astronomical computing history

STScI fundation planning work :

- Newsletters and technical memos from late 1970s
- Early introduction of **AI, Machine Learning, and Bayesian methods** centers around a massive problem: **How to guide the Hubble Space Telescope.**
  - No digital map of the sky dense enough to guide HST
  - Digitizing thousands of photographic plates (20 x10^6 sources)
    - **Massive catalogue (10^9 objects) was produced**

      (Barry M. Lasker, **FG**, *et al,* 2008, *AJ,* **136** 735)



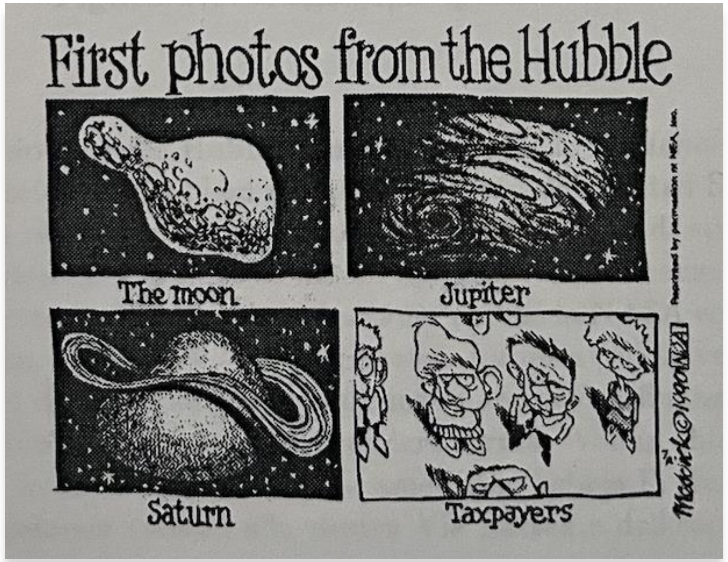
Schmidt plate, Coma cluster

STScI, AI and Bayesian Connections

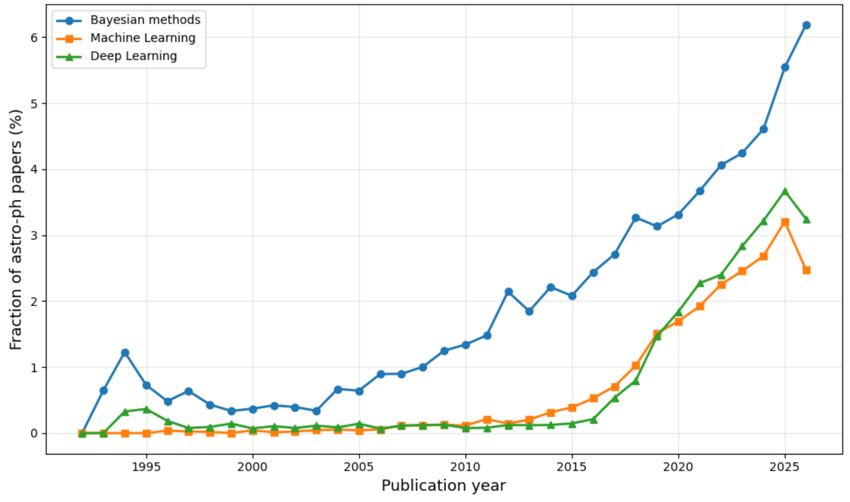
A brief astronomical computing history

Dealing with a data problem decades ahead of its time, its early newsletters and technical memos became a hotbed for discussing advanced computational techniques:

- **Automated Classification (The Roots of ML)**
  - **Bayesian approach to pattern recognition and classification complexities**
  - **First ML applications** "*Automated Star-Galaxy Discrimination with Neural Networks*" Odewahn, S. C., et al. 1992, **AJ**, 103, 318
- **Artificial Intelligence for Scheduling (SPIKE)**
  - Based on NN (1980s)
- **The Bayesian Image Restoration Shock (1990)**
  - Lucy-Richardson, MEM



Bayesian Inference in astronomy



Data taken from arXiv: astro-ph – statistics based on titles and abstracts

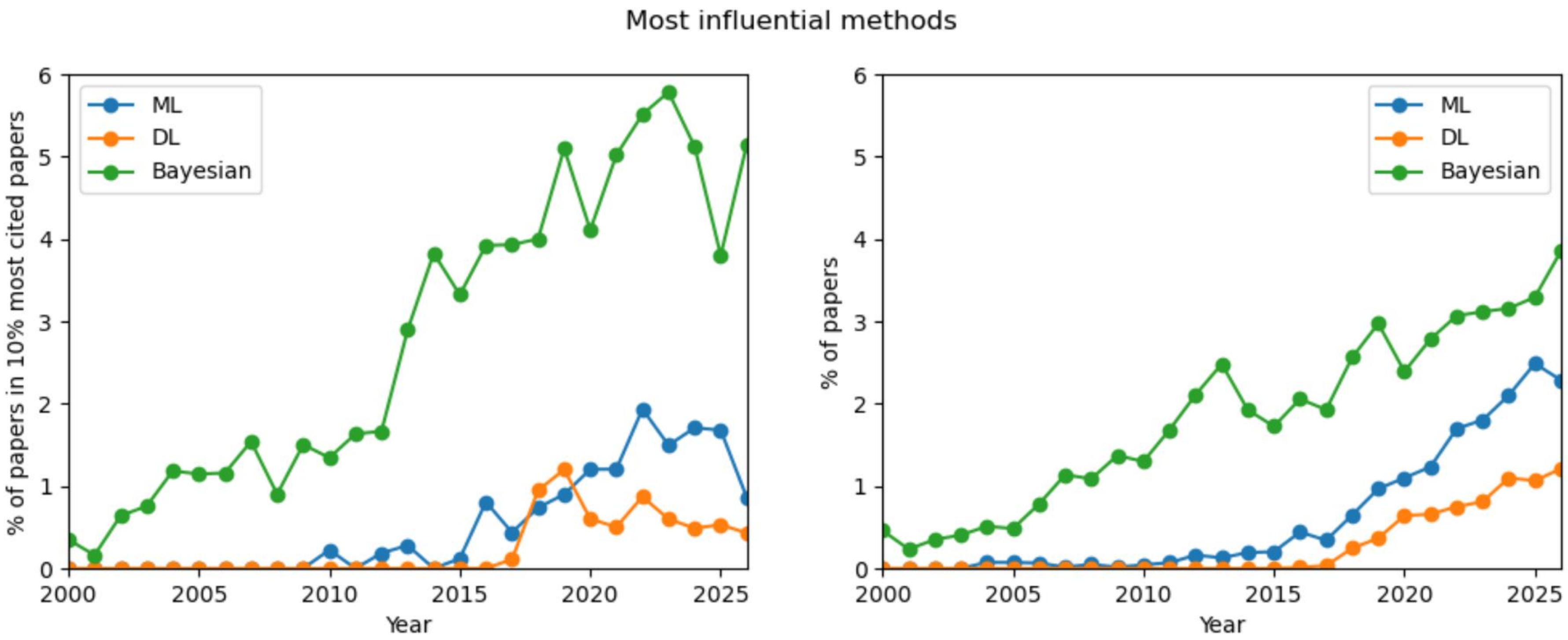
Gold standard for interpreting the cosmos

- Earliest roots (1960s-1970s)
  - The 1968 Foundation (Ed Jaynes)
  - The Lucy-Richardson Deconvolution (1974)
- Modern Revolution: MCMC (1990s)

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$

  - *"From Laplace to supernova SN 1987A: Bayesian inference in astrophysics" , Tom Loredo (1990)*
- ***Today :* treated as a core branch of probabilistic machine learning**
  - Exoplanet Detection, Cosmological Parameters, Gravitational Wave Analysis

Realizing the potential of astrostatistics and astroinformatics



(papers in A&A, ApJ, AJ & MNRAS – statistics based on titles & abstracts only!)

Credit Jarle Brinkmann (ESO)
AI Days @ ESO March 2, 2026

Bayesian techniques allow astronomers to bake the laws of physics directly into the model as "Priors” (pdf), meaning they can extract highly accurate, mathematically sound discoveries from even the faintest, most isolated signals in the sky

What is modern AI?

The emergence of modern AI

Major paradigm shift at the Garching Hub: Intersection of Deep Learning, Bayesian Inference, and Physical Laws

- From AI treated as a "black-box" (simply looking at patterns) to **Physics-Informed and Risk-Aware AI**
  - Systems that understand **uncertainty** and the **laws of physics**

```
      THE MODERN AI ECOSYSTEM (ODSL PERSPECTIVE)

 [ Traditional AI / ML ]          [ Pure Bayesian Inference ]
  (Deep Learning, CNNs,            (MCMC, Nested Sampling,
  unconstrained black-boxes)         Information Field Theory)
             \                              /
              \                            /
               v                          v
              [ PROBABILISTIC MACHINE LEARNING ]
              * Quantifiable Error Bars
              * Physics-Informed Neural Networks
              * Simulation-Based Inference (SBI)
```

The emergence of modern AI

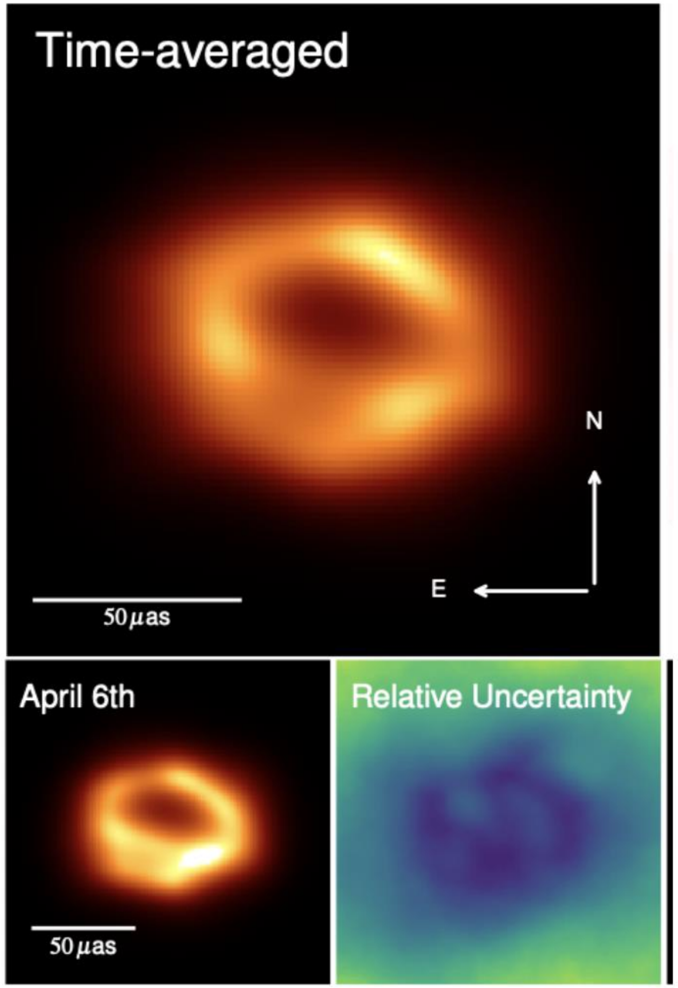
Major paradigm shift at the Garching Hub: Intersection of Deep Learning, Bayesian Inference, and Physical Laws

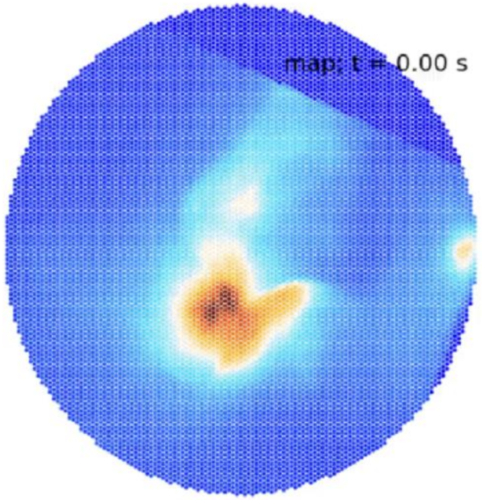
- **Origins Data Science Lab (ODSL)**
  - A group of interdisciplinary minded people interested in Statistics, AI/ML, Data Science, Open Data / Open Science, Distributed Computing
  - Bridging the two historically competing philosophies of data science:
    - **The Frequentist/Deep Learning camp:** "Give me massive data, and the neural network will figure out the rules."
    - **The Bayesian camp:** "Give me a physical model, and I will compute the exact probabilities."
- Demonstrating to the broader AI community how to build **trustworthy AI** for the natural sciences.
- Mapping out the future blueprint of how AI operates under strict physical constraints.

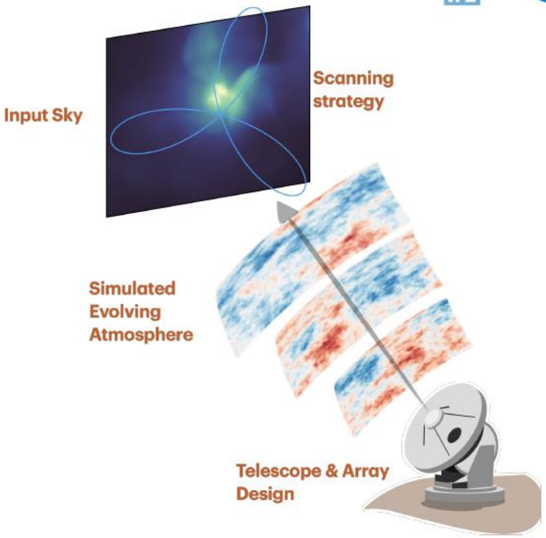
Credit Lukas Heinrich (TUM/MPP)
ODSL Days July 8, 2026

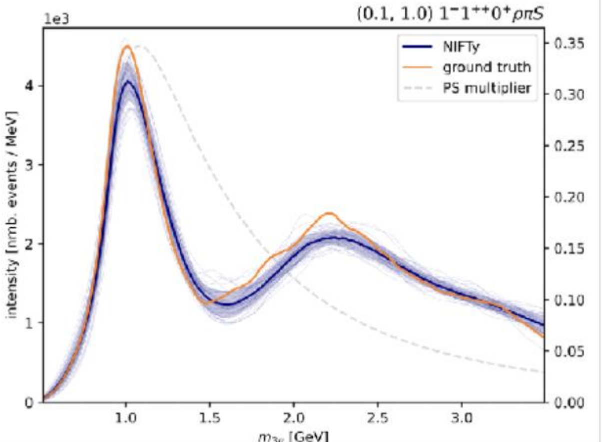
Bayesian Inverse Problems

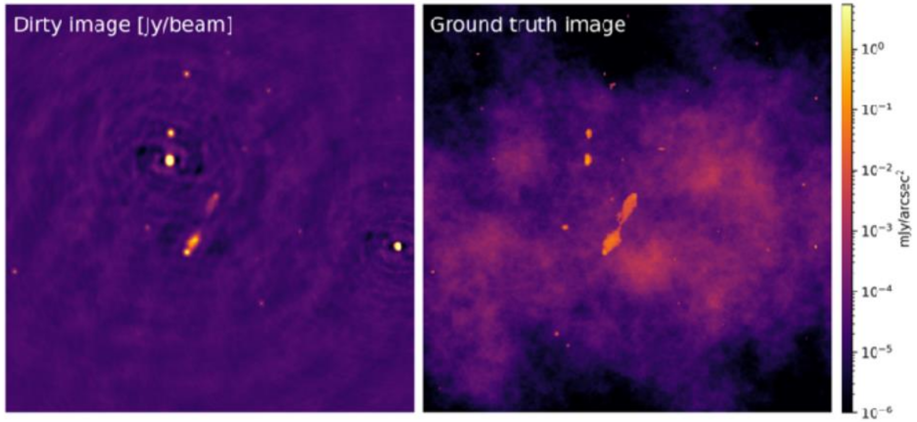
Transfer over "multiple generations" from Astro to Particle Physics







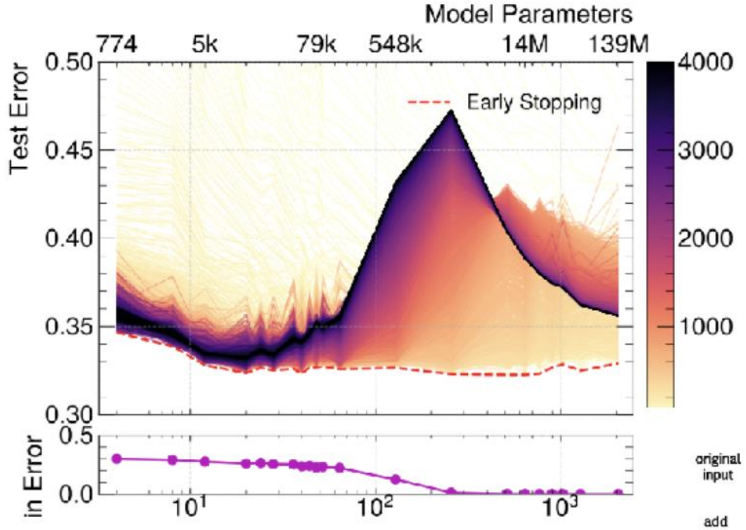


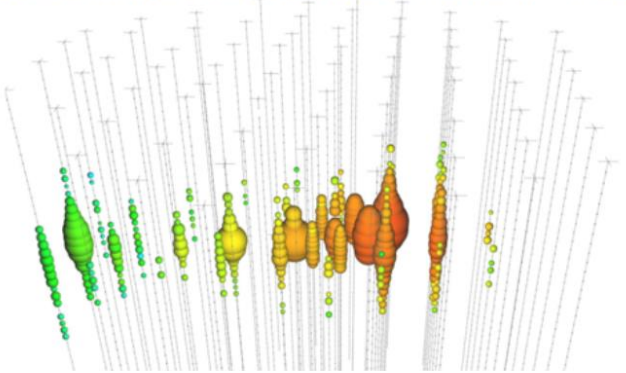


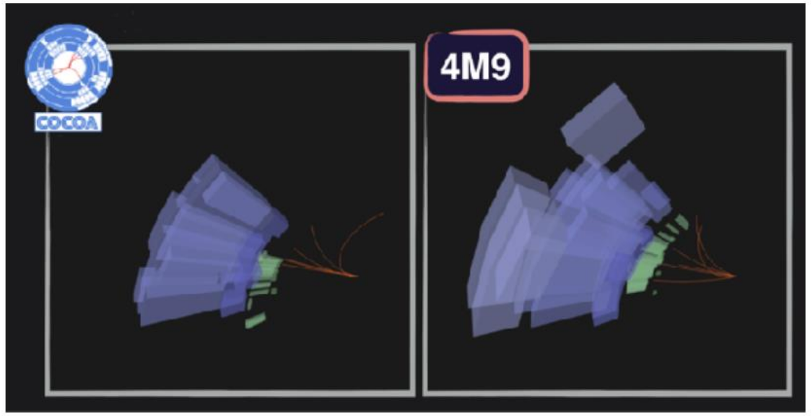
Credit Lukas Heinrich (TUM/MPP)
ODSL Days July 8, 2026

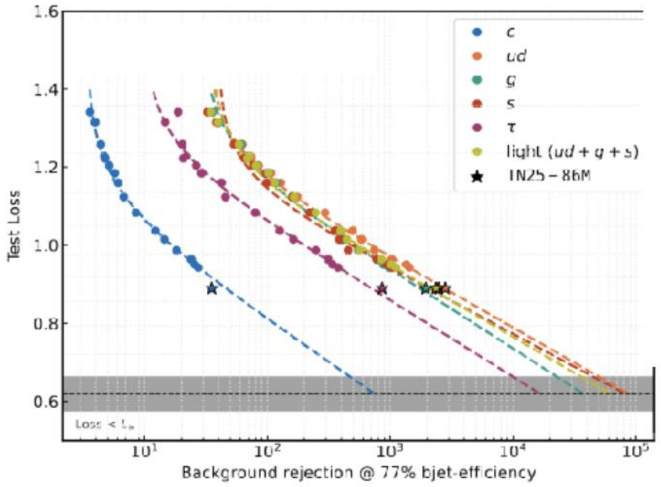
Foundation Models:

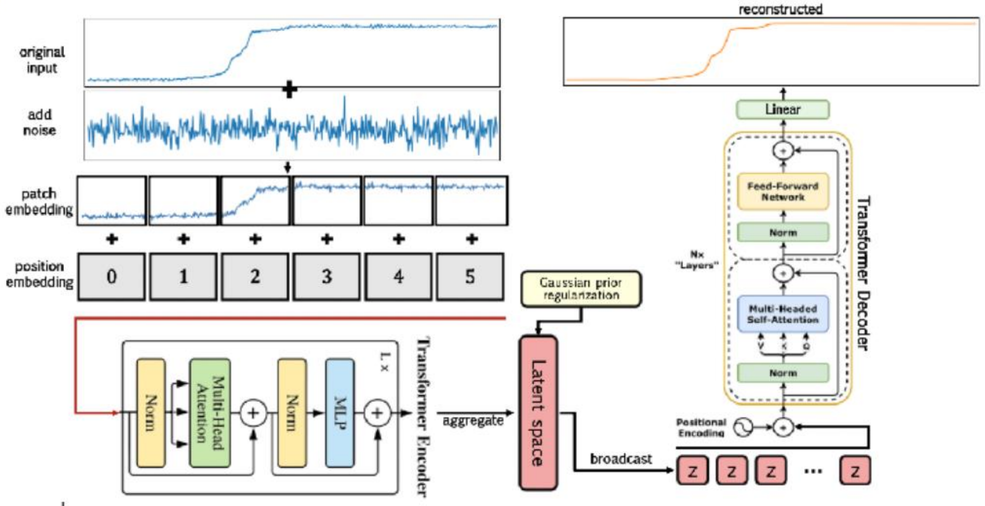
Big Trend: from specialists to generalist AI models. We can train O(1-10 B) models!

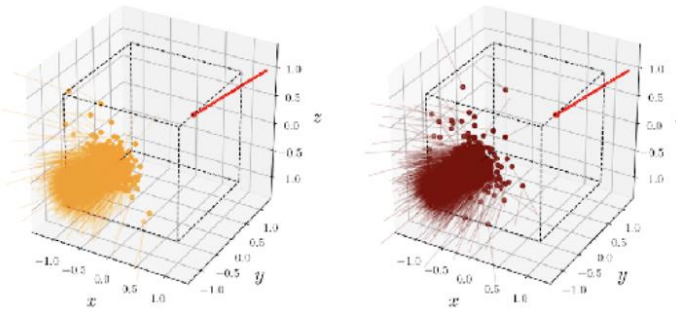








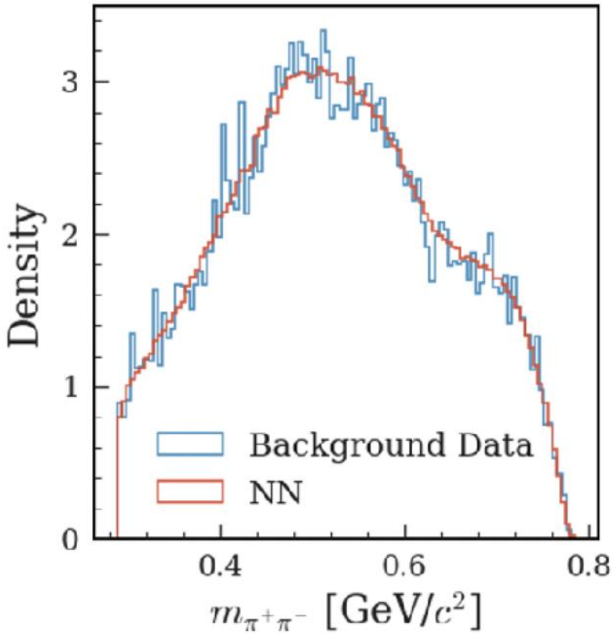


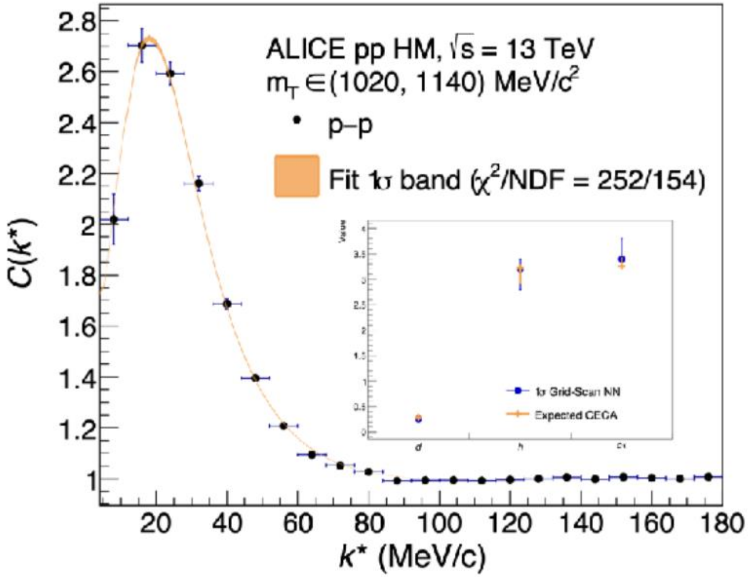


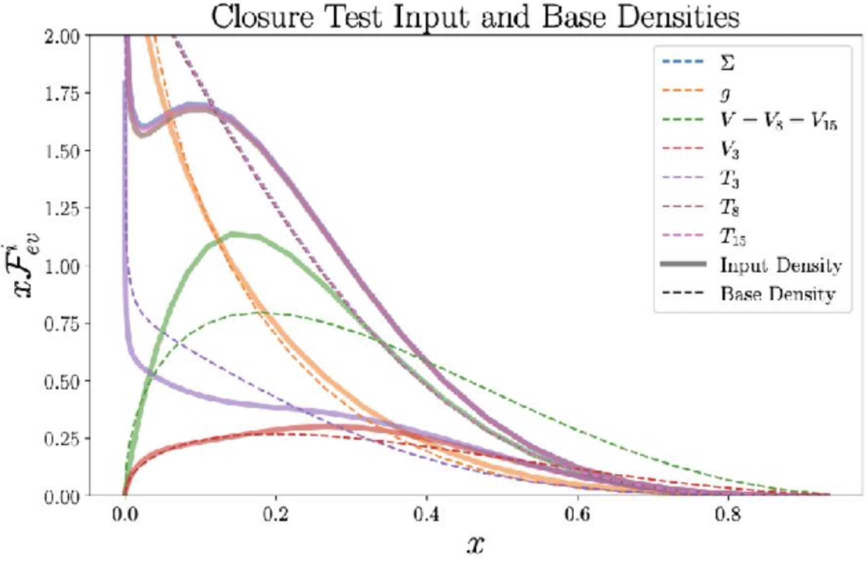
Credit Lukas Heinrich (TUM/MPP)
ODSL Days July 8, 2026

Neural Density Modelling

AI methods are excellent as a non-parametric density modelling tool. Multiple applications across ORIGINS sciences e.g. for background modelling



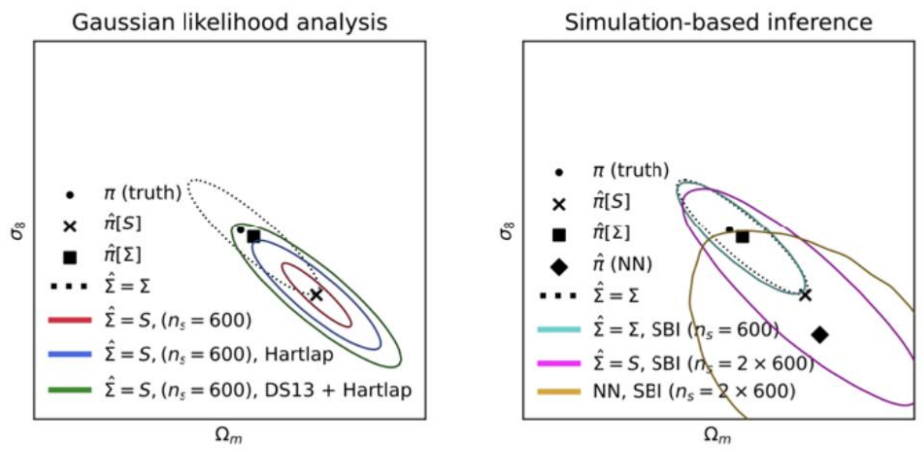


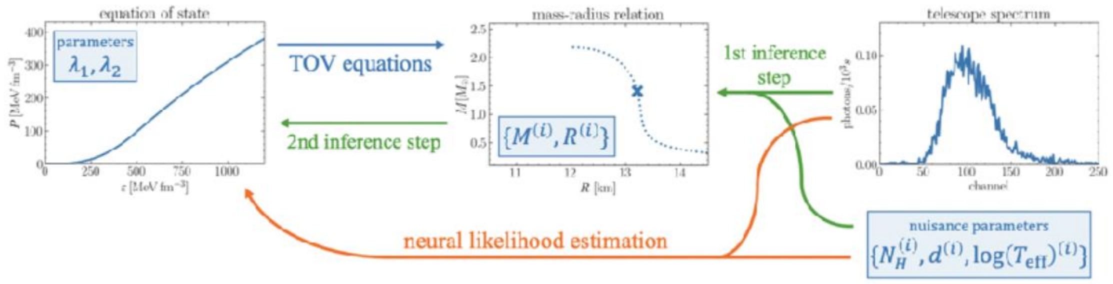


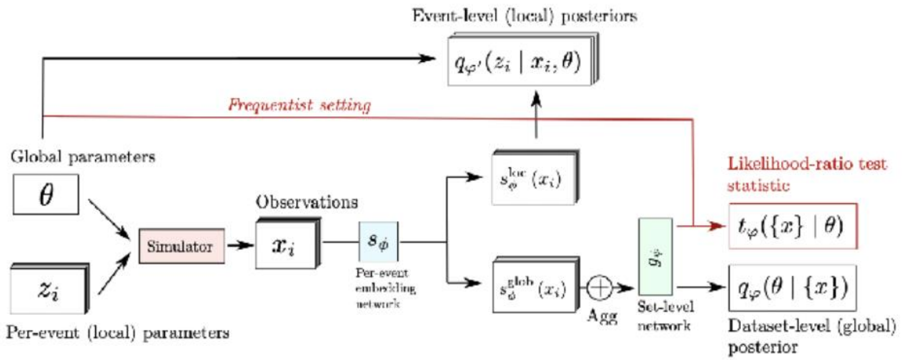
Credit Lukas Heinrich (TUM/MPP)
ODSL Days July 8, 2026

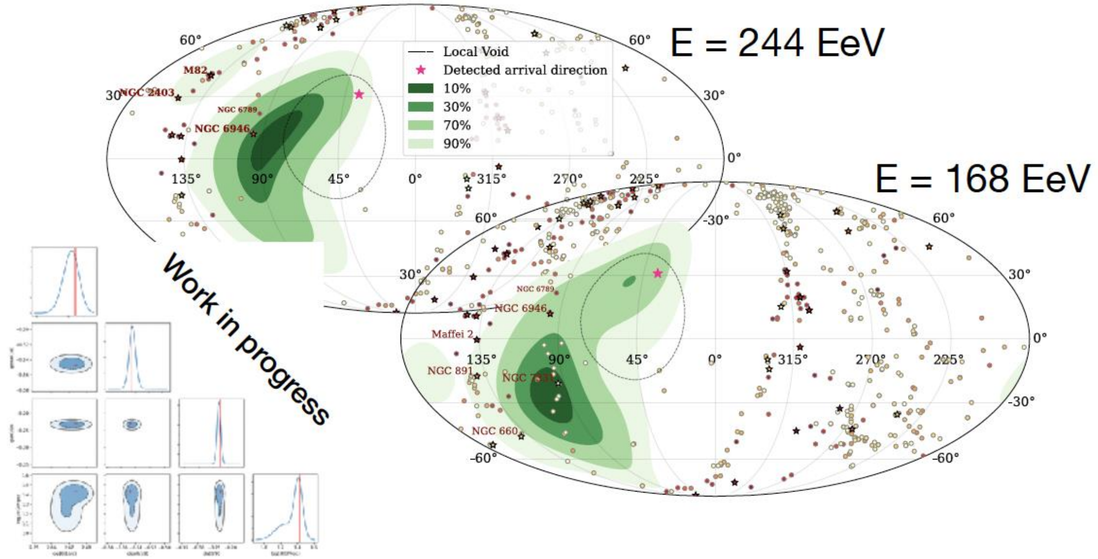
Simulation-based Inference

A updated, cross-disciplinary approach on statistical inference for big science



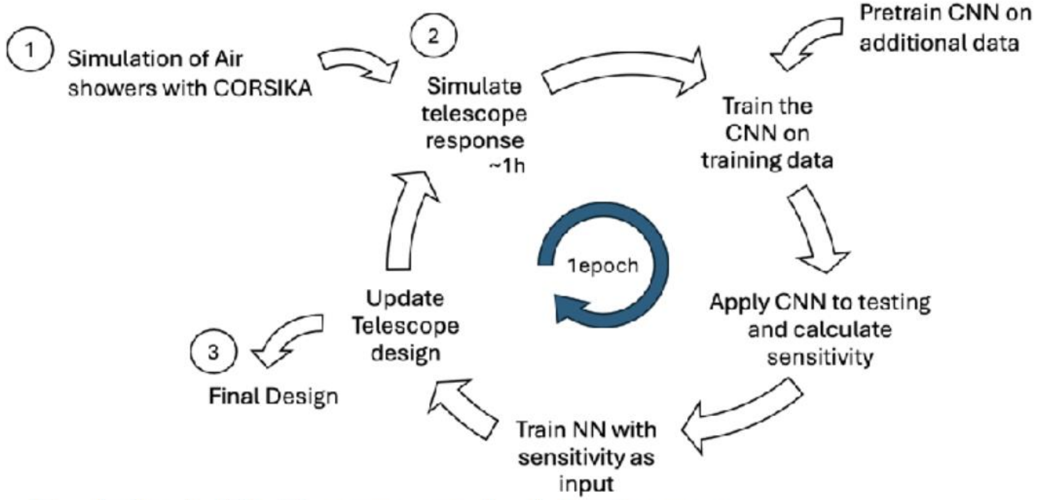




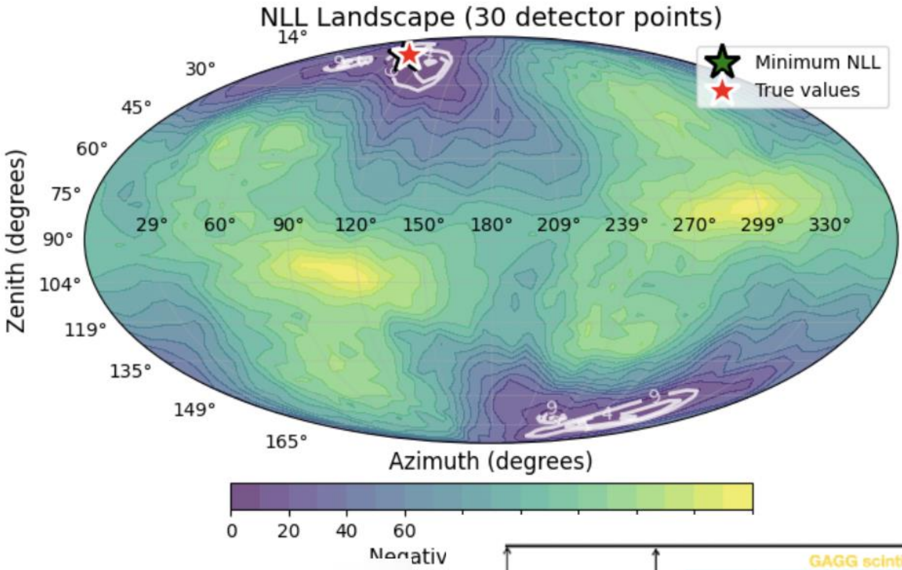


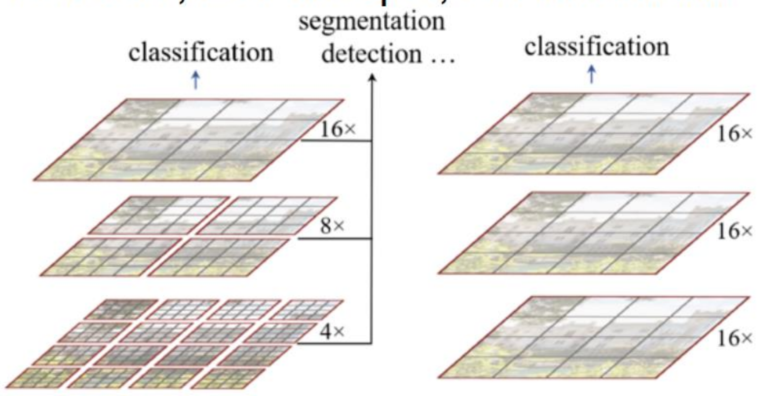
AI-driven Design

The Framework



E. Jobst, D. Paneque, L. Hackmann

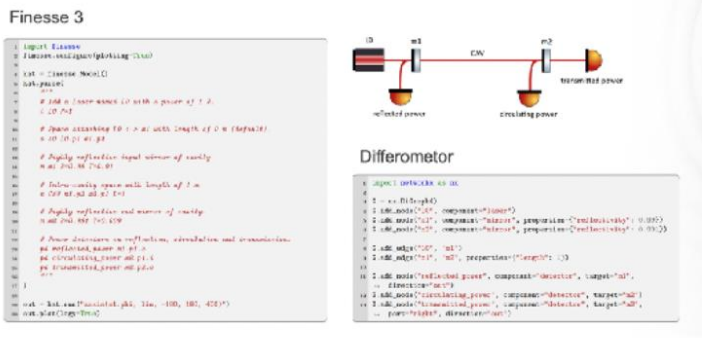


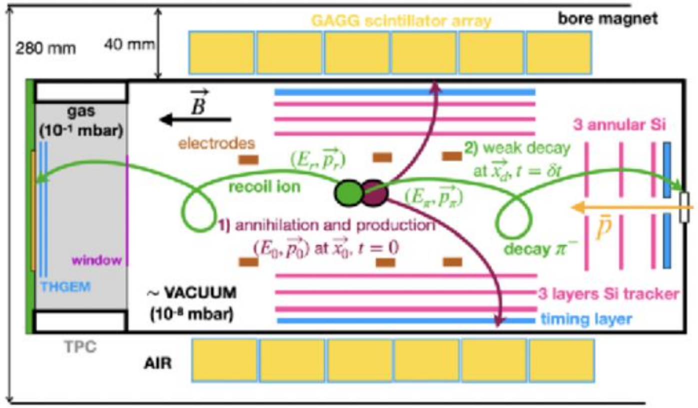


(a) Swin Transformer (ours)

(b) ViT

differometer - A Differentiable Interferometer Simulator





Credit Lukas Heinrich (TUM/MPP)
ODSL Days July 8, 2026

Data driven approaches

The ESO internal ALMA Development Study

Bayesian Reconstruction through Adaptive Image Notion (2019-2024), *PI Fabrizia Guglielmetti*

ALMA Development Studies: Motivation and Context

Why ALMA Development studies?

- Enable community-led innovation beyond standard operations
- Prepare for future scientific and technical challenges
- Test high-risk/high-gain ideas before operational adoption

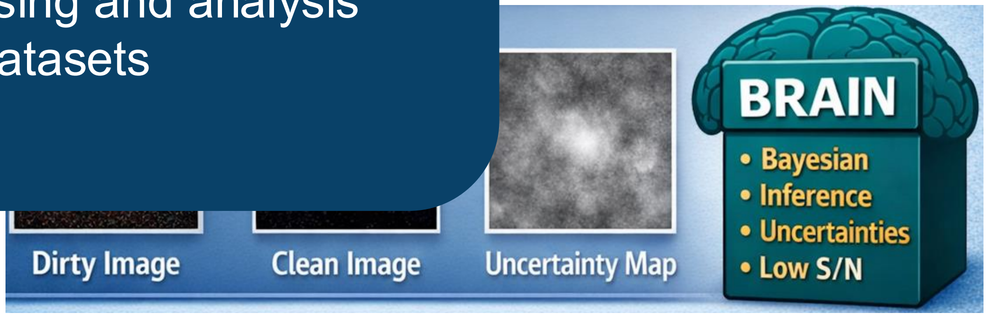
Challenges motivating new approaches

- Increasing sensitivity exposes faint, extended emission
- Complex source morphologies are difficult to image robustly
- Standard techniques may rely on manual tuning
- Limited propagation of uncertainties to science products

The BRAIN Study

- Enhance image analysis using data-driven algorithms
  - ➢ Prioritize objective insights over subjective interpretations
  - ➢ Achieve more precise and efficient image processing and analysis
- Improve computational efficiency for handling large datasets
- Automate data processing to elevate image quality

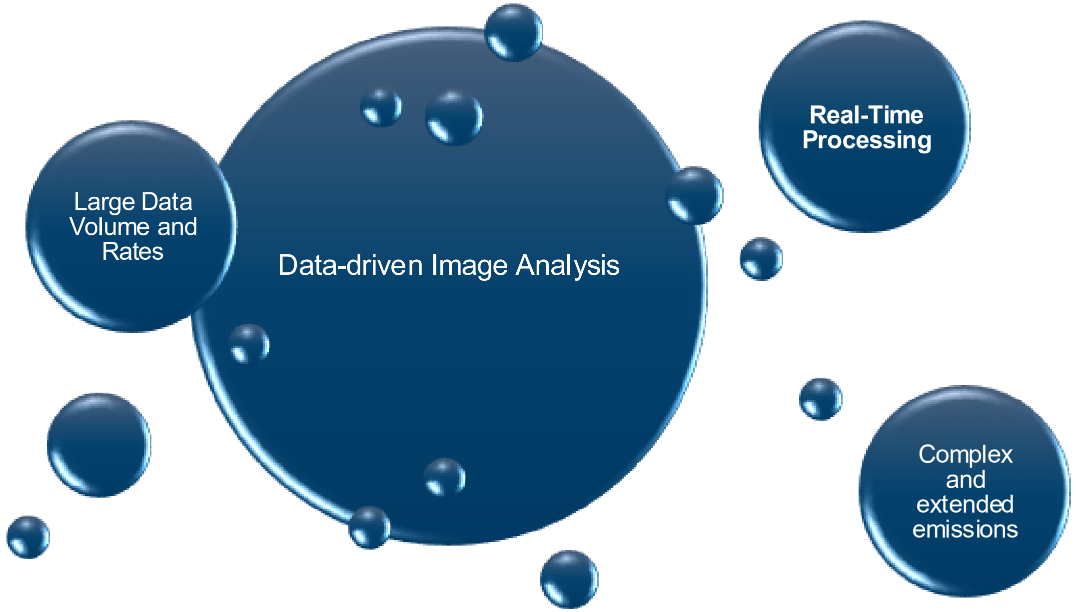
<u>**Guglielmetti+2019**</u>, <u>**Guglielmetti+2022**</u>, <u>**Guglielmetti+2023**</u>



Enhancing Data Processing in Astrophysics: The Brain Study

ALMA2030 Operational Challenges

(Gonzalez, A. et al., 13102, SPIE, 2024; Carpenter, J. et al., ALMA memo 621, 2022)



Solution Pathways:

Cutting-edge Resources and AI Innovation:

- **Expertise in AI and Astronomy:** Engage teams with dual expertise to create advanced algorithms tailored for astronomical data.
- **Enhanced Source Detection Accuracy**: AI-based approaches can refine source detection, leveraging knowledge embedded in Point Spread Functions (PSF), noise patterns, source distribution…

ALMA Simulation Tools:

- **Purpose**: Develop an ALMA simulator designed to generate datasets for AI training, testing and validation.
- **Applications**: This simulator is supported by empirical noise studies and used to validate AI models, ensuring alignment with real-world conditions in the ALMA2030 environment.

Resolve ⊂ NIFTy ⊂ UBIK

RESOLVE:

- **Methodology:**
  - ✓ Bayesian approach within IFT
  - ✓ Synthesizing radio apertures
  - ✓ Enabling precise image reconstruction
- **Key Capabilities**:
  - ✓ Low signal-to-noise diffuse emissions detection
  - ✓ Data combination
  - ✓ Self-calibration, Full Polarization imaging
- **Results**:
  - ✓ High accuracy in identifying and preserving weak celestial signals, even in noisy observational environments

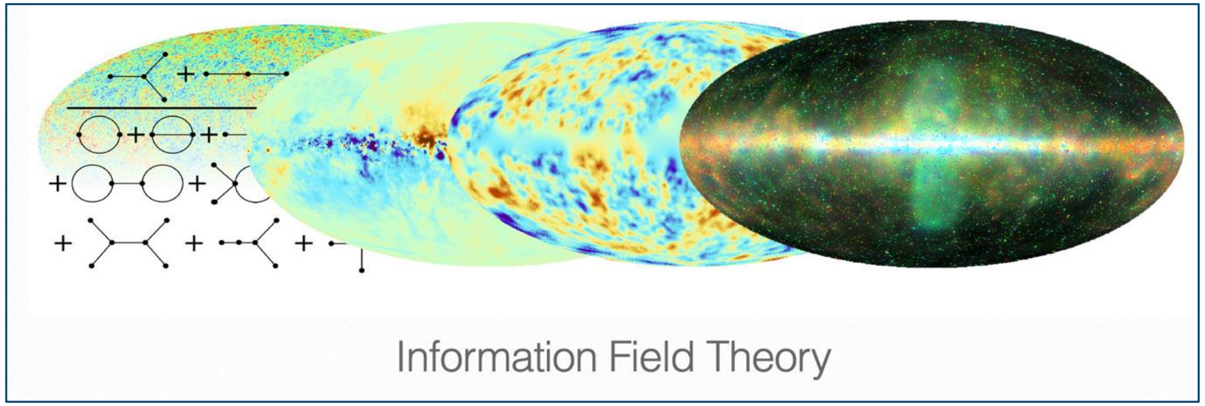
https://github.com/information-field-theory/nifty
https://ift.pages.mpcdf.de/nifty/index.html

https://github.com/NIFTy-PPL/J-UBIK
https://ift.pages.mpcdf.de/j-ubik/

https://gitlab.mpcdf.mpg.de/ift/resolve
https://ift.pages.mpcdf.de/resolve/

Resolve user guide

This guide is an overview and explains the main idea behind resolve. More details are found in the API reference.



Research, Applications, Software and Publications of the IFT group

RESOLVE:

https://github.com/information-field-theory/nifty
https://ift.pages.mpcdf.de/nifty/index.html

https://github.com/NIFTy-PPL/J-UBIK
https://ift.pages.mpcdf.de/j-ubik/

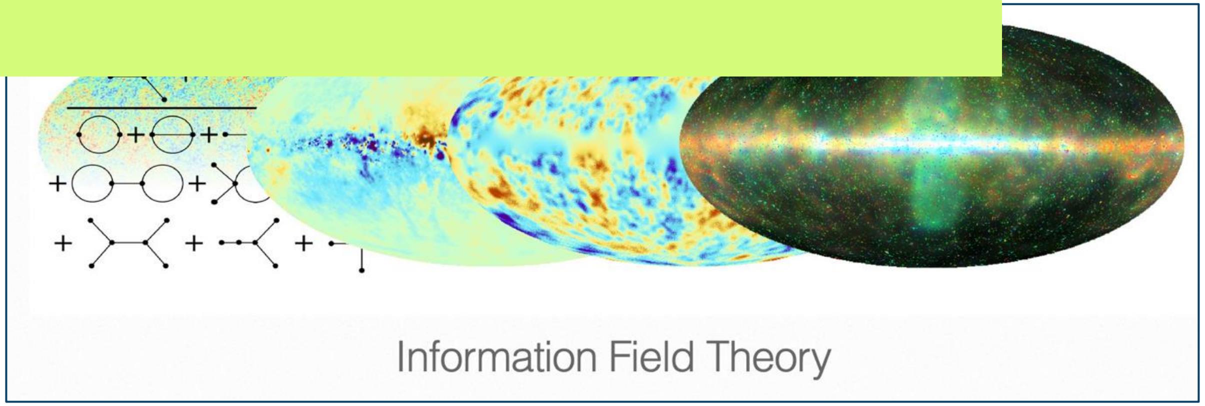
- **Methodology:**
  - ✓ Bayesian approach within IFT
  - ✓ Synthesizing radio apertures

https://gitlab.mpcdf.mpg.de/ift/resolve

Enßlin+2009, Junklewitz+2016, Tychoniec+2022, Roth+2024, Johnson+2026, Popp+2026, Rüstig+2026

- **Ke**

und in the



- **Results**:
  - ✓ High accuracy in identifying and preserving weak celestial signals, even in noisy observational environments

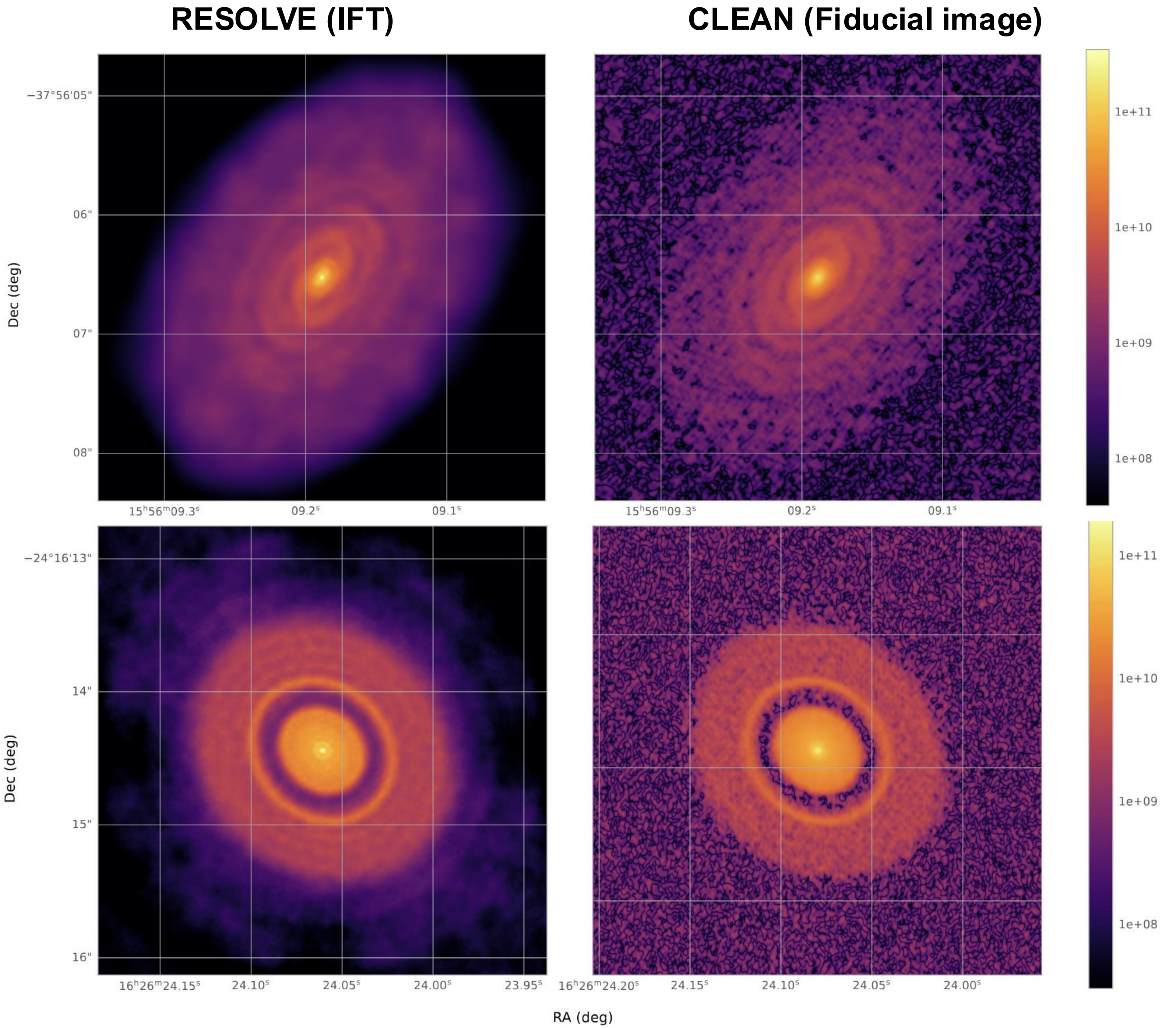
Research, Applications, Software and Publications of the IFT group

Outcome from the BRAIN Study
PI: F. Guglielmetti (ESO)

Re-imaging ALMA DSHARP Large Program (2016.1.00484.L) with resolve

Johnson et al, in prep

(left) *R*econstructed images of **IM Lup** (top) and **Elias 24** (bottom) with **resolve** (IFT algorithm), **revealing extended emission and new features, all consistent with the LAS and not previously detected;**
(right) *Corresponding* **fiducial images** from the DSHARP team, obtained employing the traditional **CLEAN algorithm**

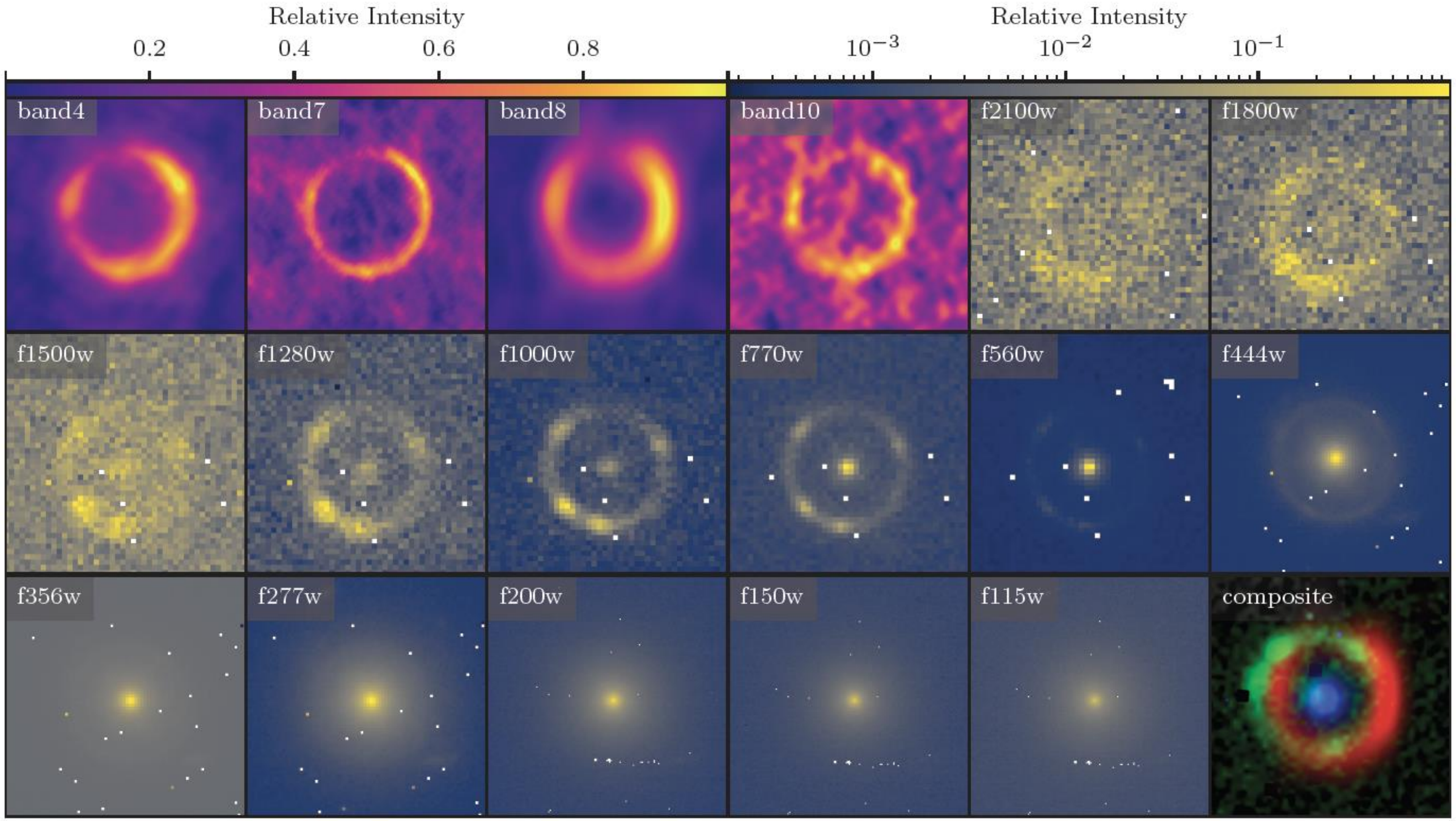


Combining ALMA and JWST data for source reconstruction (z~4.2) with Information Field Theory

ALMA

JWST/MIRI

JWST/NIRCam

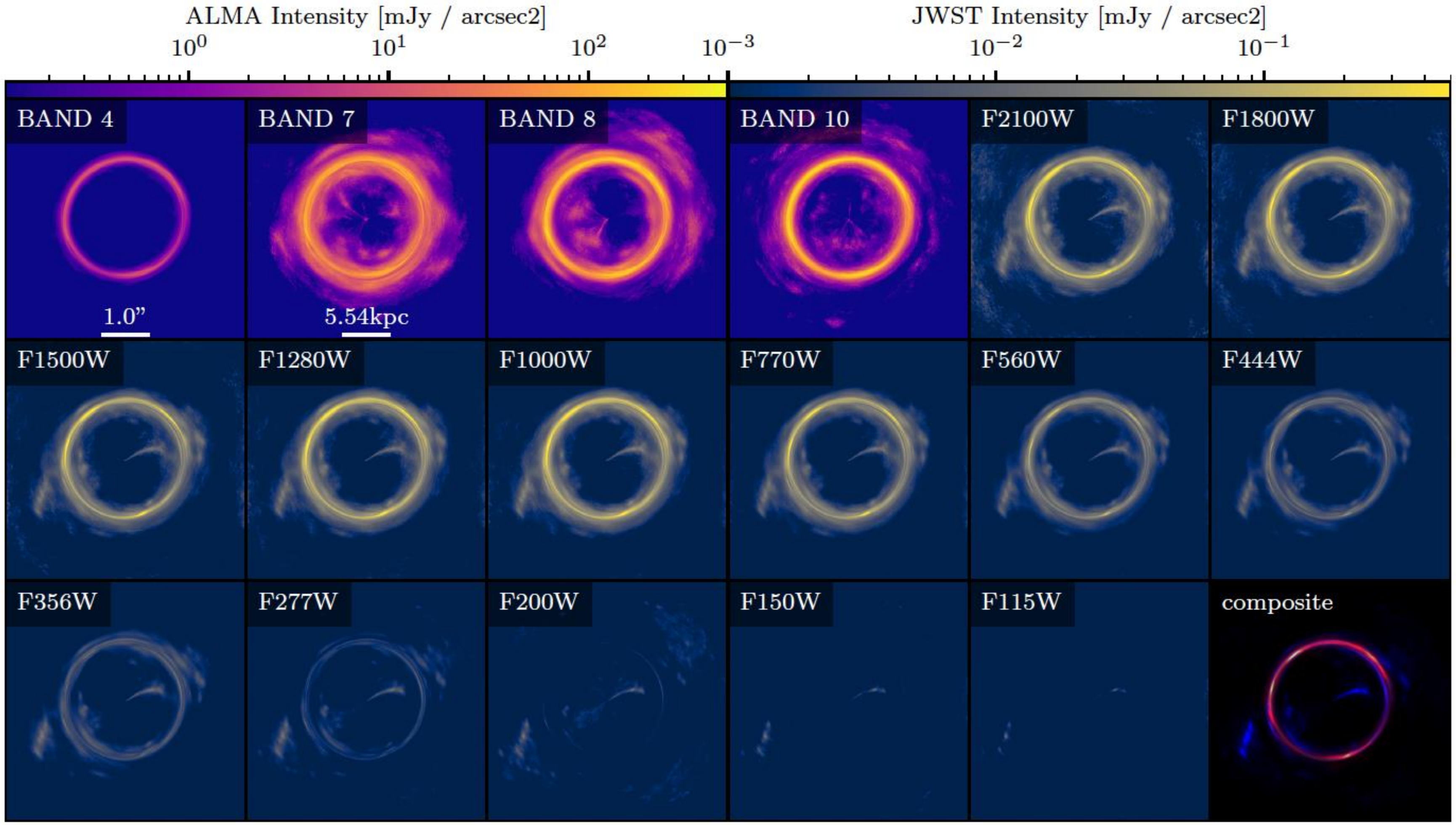


Ruestig et al, submitted **Outcome from the Early-Career Scientific Visitor Programme, 2025**

IFT is capable to perform analyses on ALMA multi-bands and JWST multi-filters analyses and to provide a composite image.

Combining ALMA and JWST data for source reconstruction (z~4.2) with Information Field Theory

ALMA



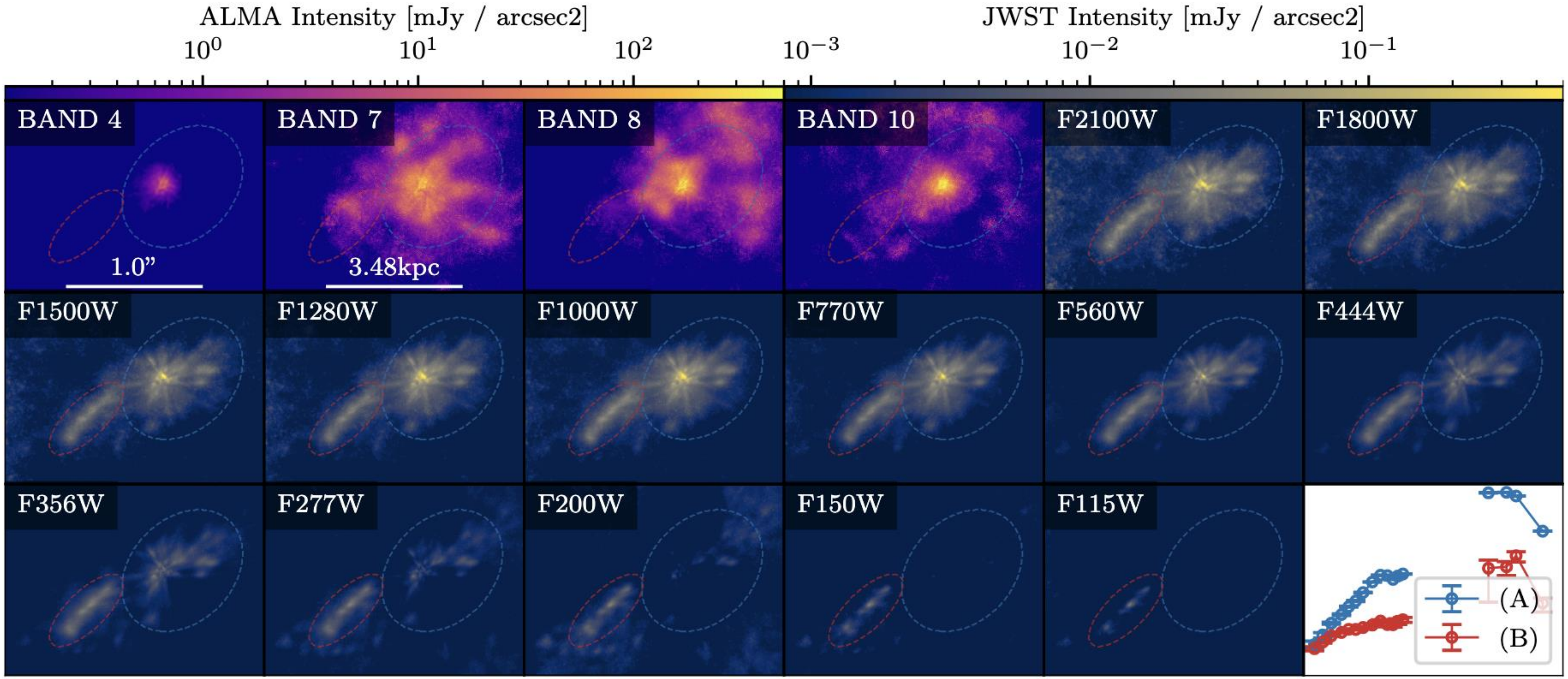
JWST/MIRI

JWST/NIRCam

Ruestig et al, submitted **Outcome from the Early-Career Scientific Visitor Programme, 2025**

IFT is capable to provide Posterior-mean lensed-light reconstruction across all 17 bands.

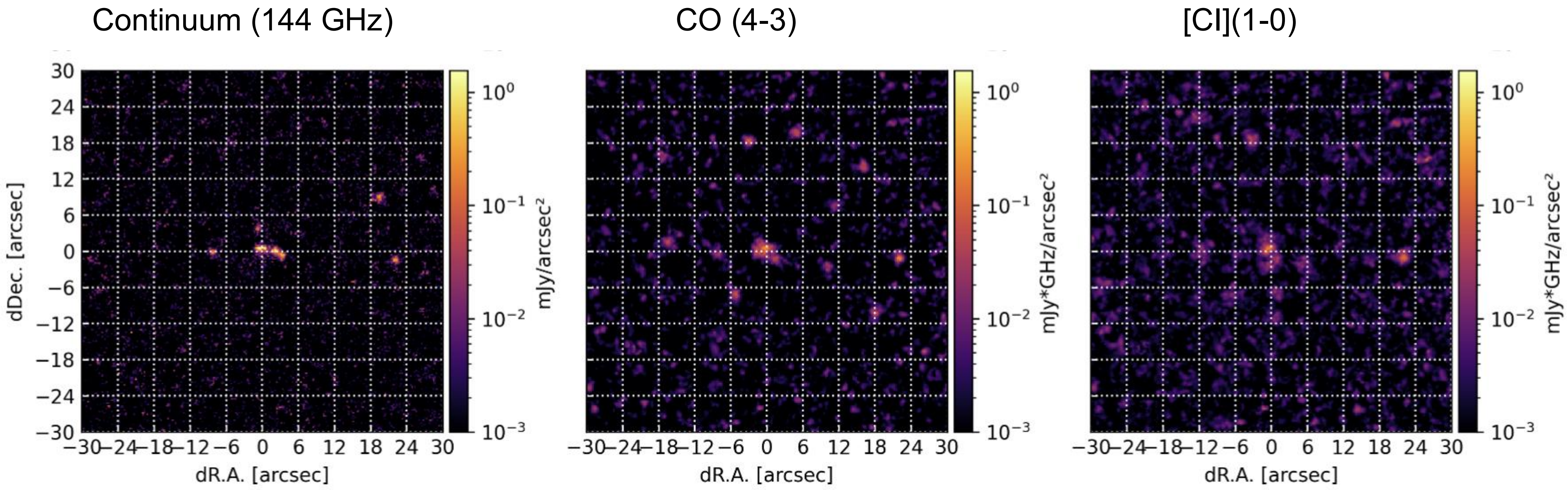
Combining ALMA and JWST data for source reconstruction (z~4.2) with Information Field Theory



Ruestig et al, submitted **Outcome from the Early-Career Scientific Visitor Programme, 2025**

IFT is capable to perform joint analyses and to provide the information of 2 merging galaxies.

Targeting the Spiderweb (z~2.16) with ALMA in Band 4 with resolve



Popp et al, in prep **Outcome from the SSDF project, 2025**

Scale independent detection allows us to reveal more sources than previous analyses and to detect faint galaxies connected to the spiderweb.

- Employing 3 ALMA projects

Credit T. Ensslin (MPA/DZA)
ESO/DZA partnership

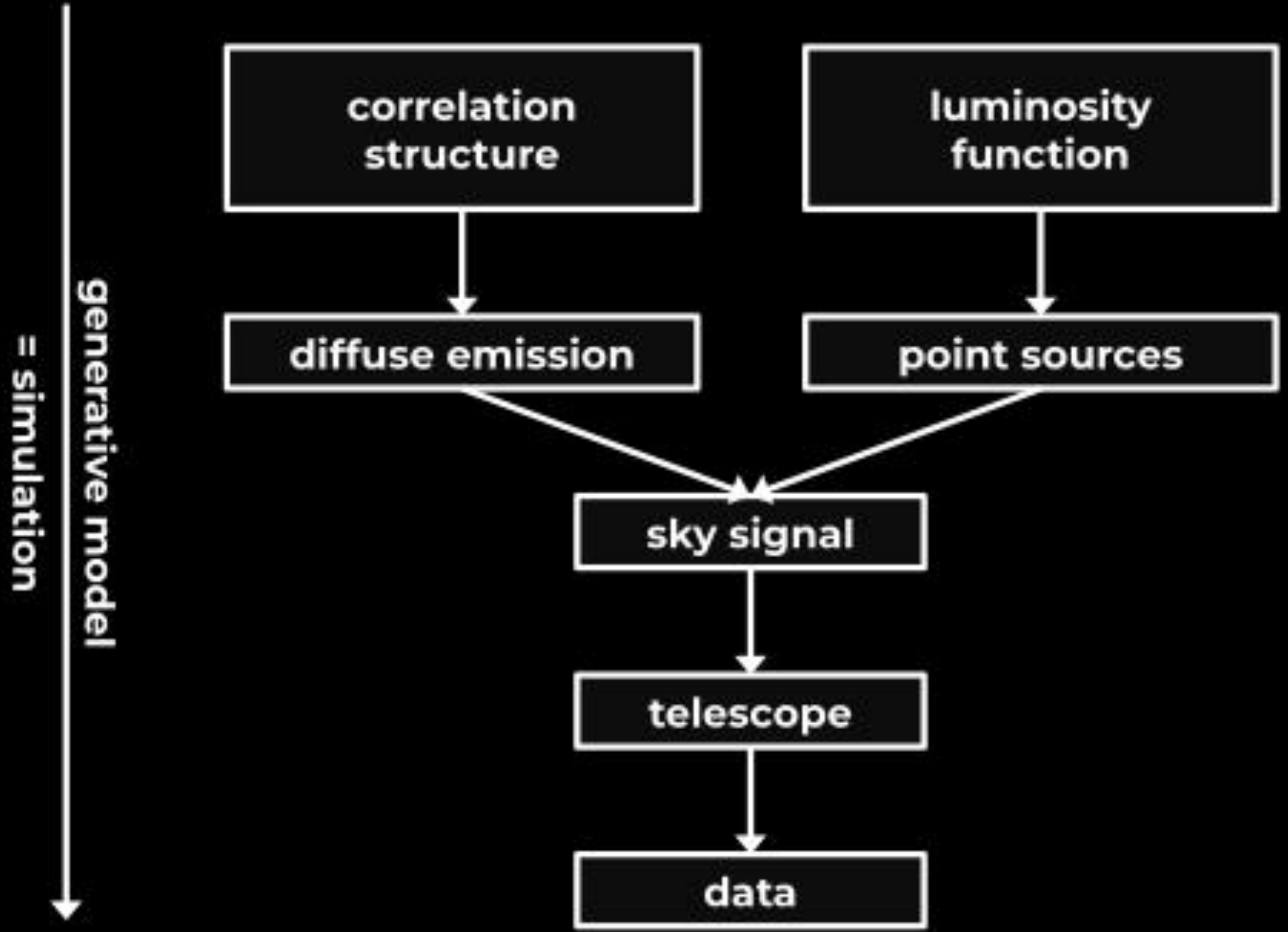
Universal
Bayesian
Imaging
Kit

UBIK

- unifies imaging for a growing number of instruments
- is multi-frequency, -instrument, & -component data analysis
- provides Bayesian uncertainty quantification
- is artificial intelligence without training
- is based on information field theory, NIFTy, & JAX
- is open source

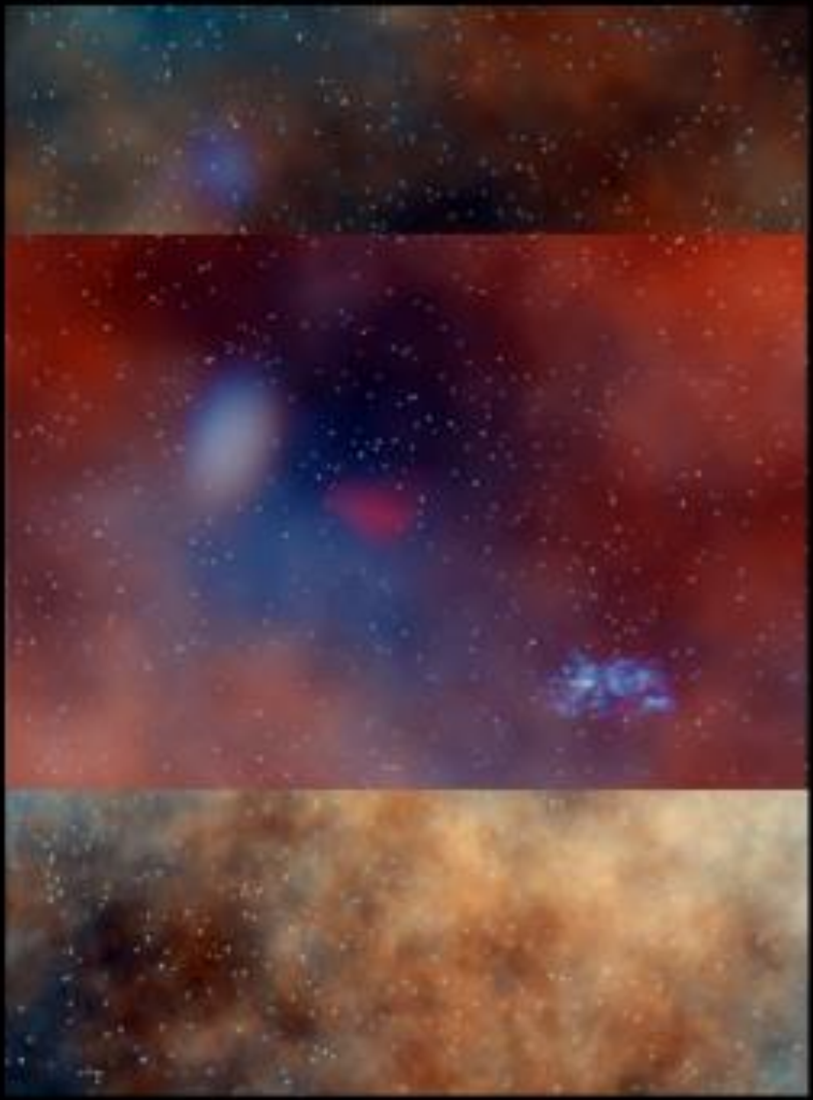
Credit T. Ensslin (MPA/DZA)

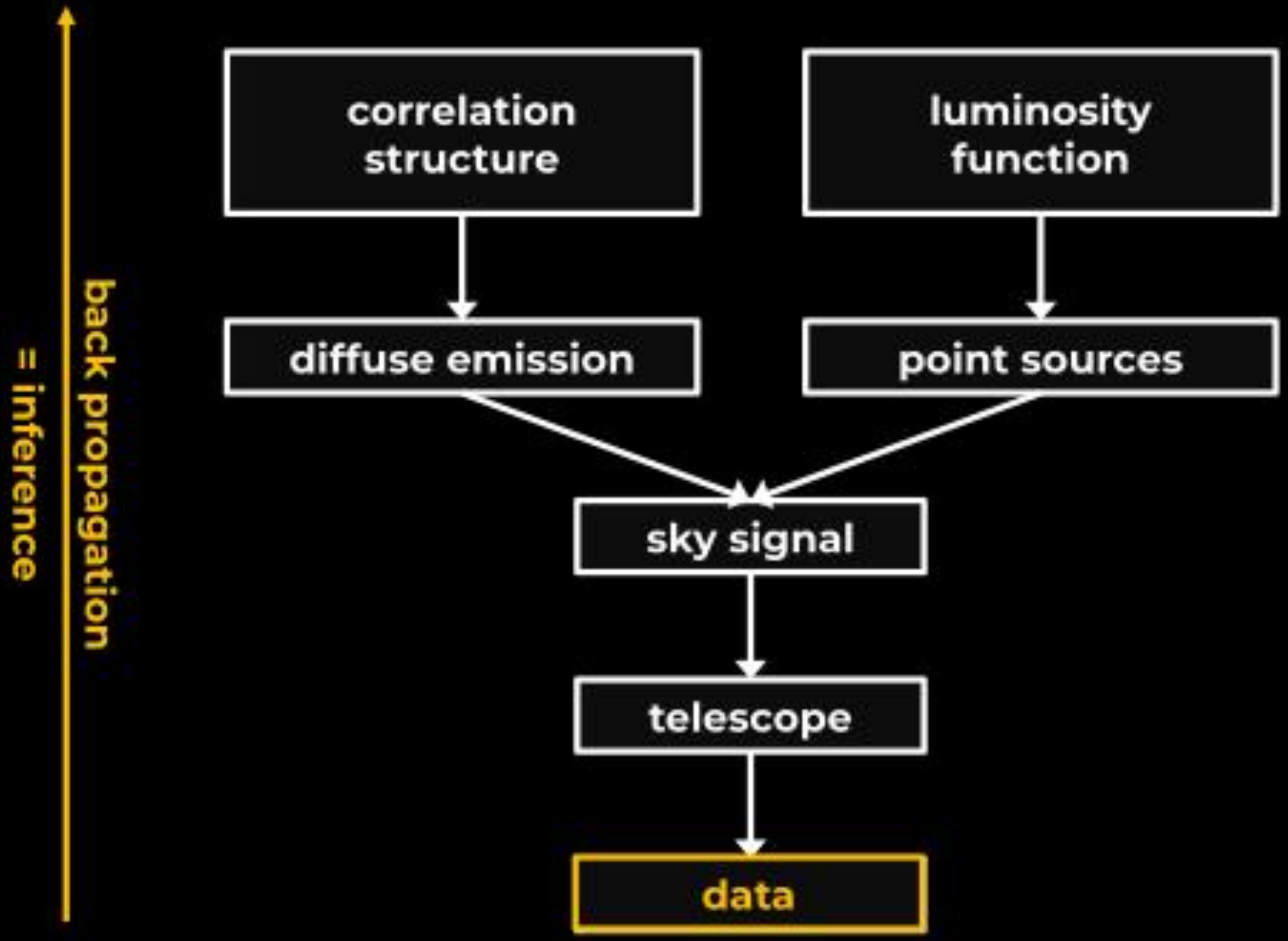
Digital Perception of Reality

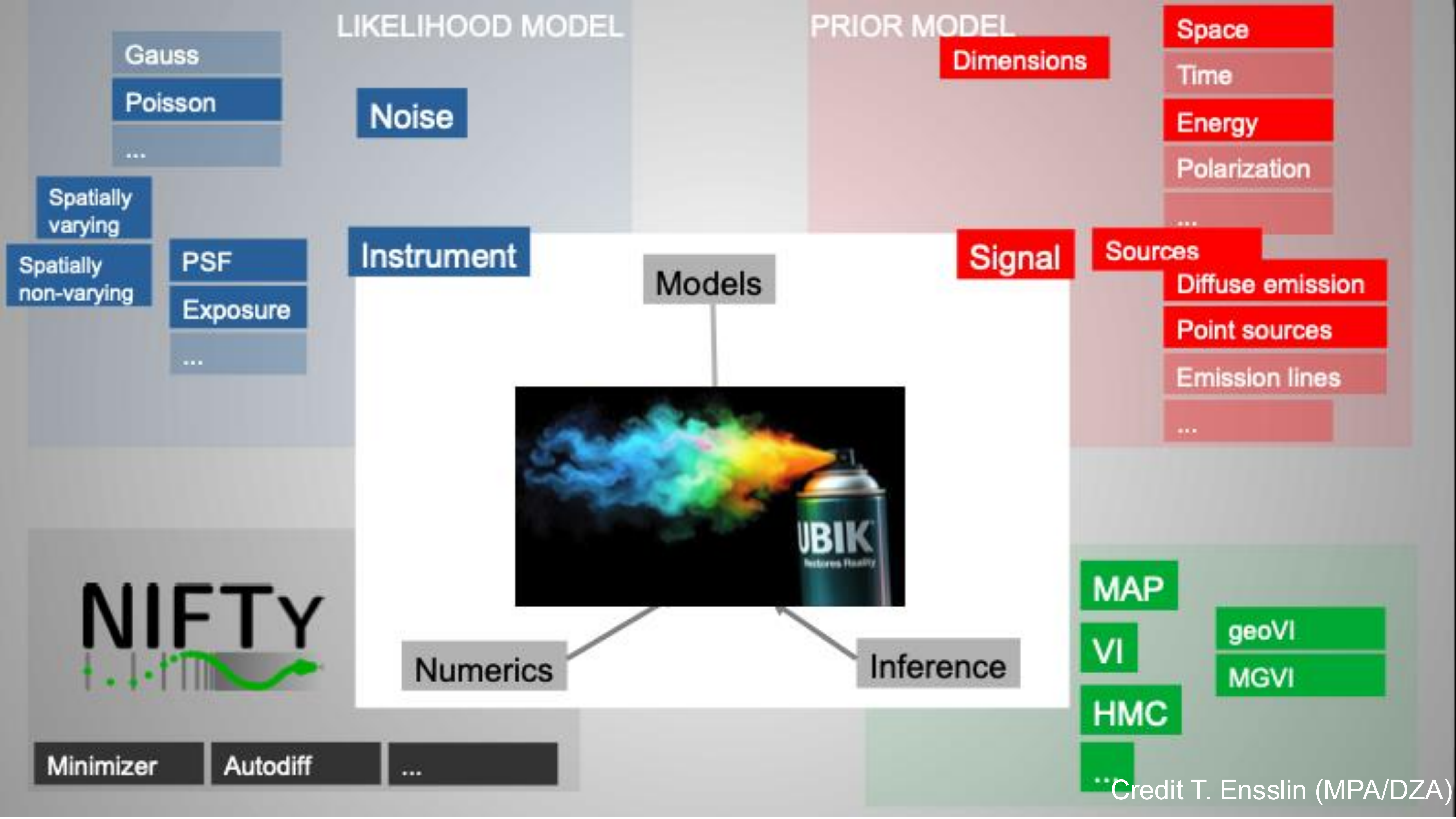


Credit T. Ensslin (MPA/DZA)

Digital Perception of Reality







Credit T. Ensslin (MPA/DZA)

Deep Learning

DeepFocus:

- **Overview**:
  - ✓ Deep Learning architecture
  - ✓ Meta-learner: Taxonomy with Bayesian Surrogates
  - ✓ Designed to analyze and process large ALMA data cubes
- **Enhanced by Simulation**:
  - ✓ Utilizes ALMASIM for realistic data generation and NOISEMPIRE for accurate noise representation, ensuring fidelity to real ALMA observations
- **Outcomes**:
  - ✓ Real-time image analysis
  - ✓ Catalogues of extracted sources

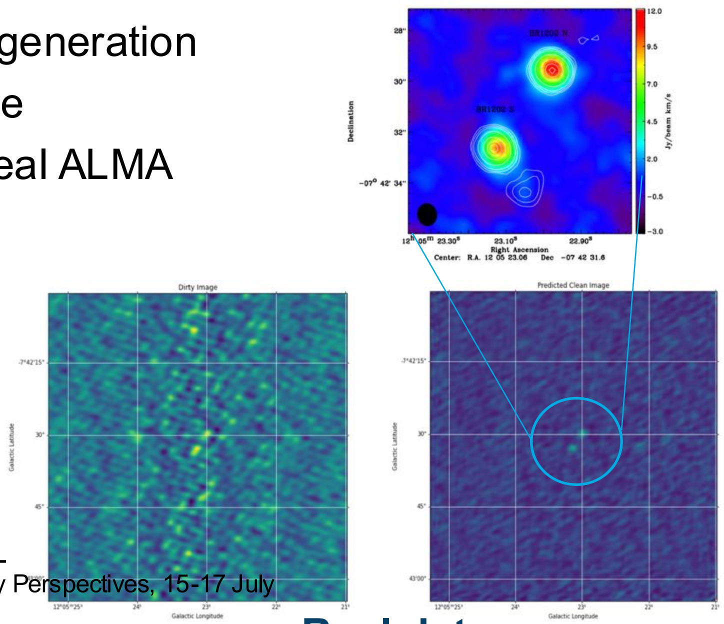
GitHub

https://github.com/MicheleDelliVeneri/DeepFocus Delli Veneri +2023

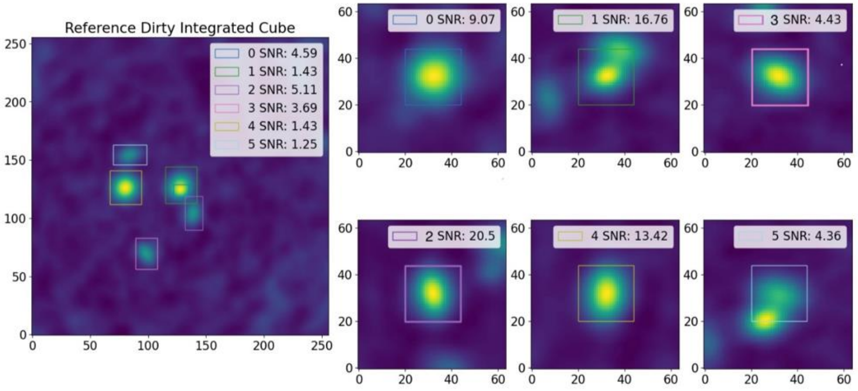
https://github.com/MicheleDelliVeneri/ALMASim

https://github.com/lbaronch/NOISEMPIRE. Baronchelli, I.

Carniani et al.+2013



Real data



Simulated data

INPUT DIRTY CUBE

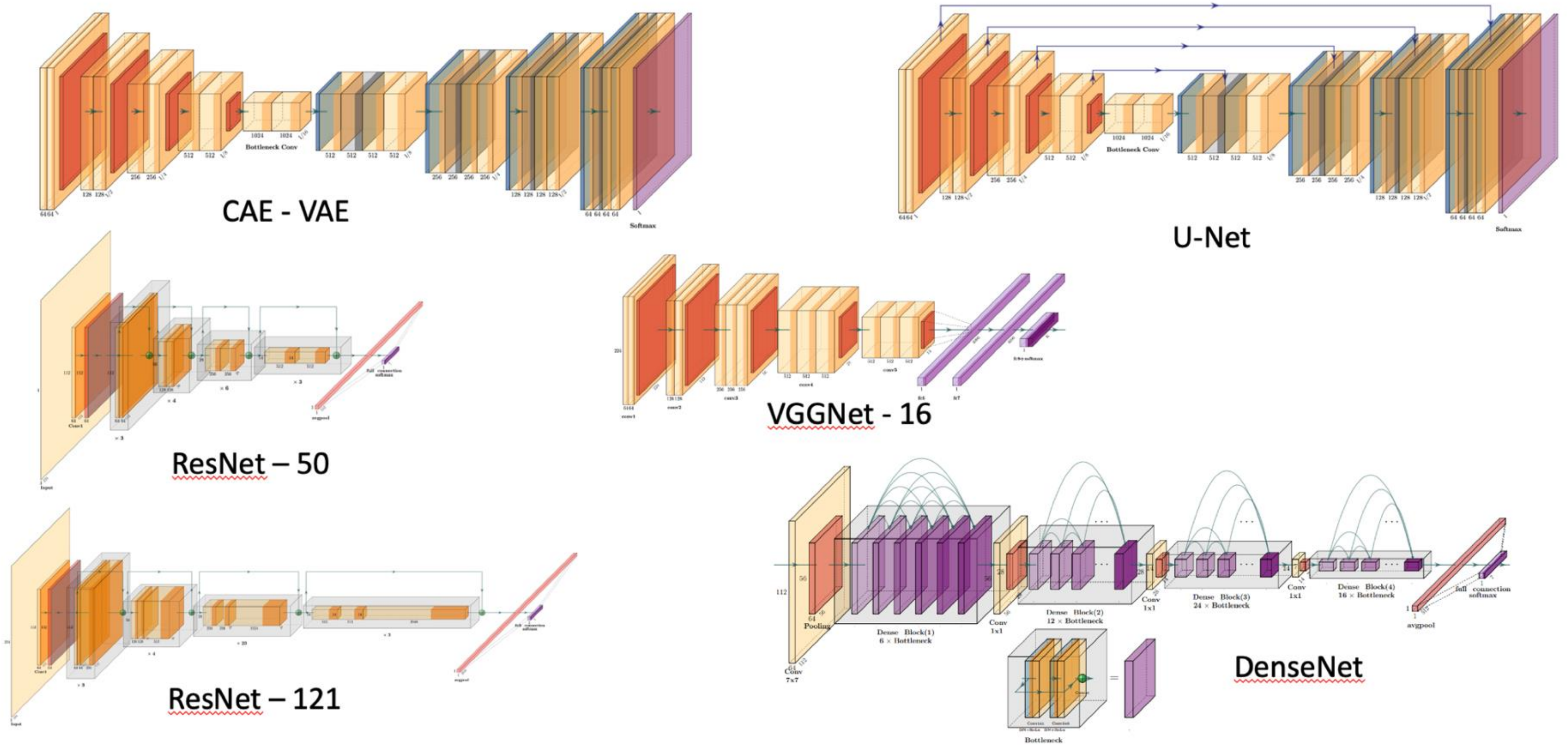
DEEP FOCUS

TAXONOMY

DECONVOLUTION AND DENOISING

FOCUSING

REGRESSION AND CLASSIFICATION



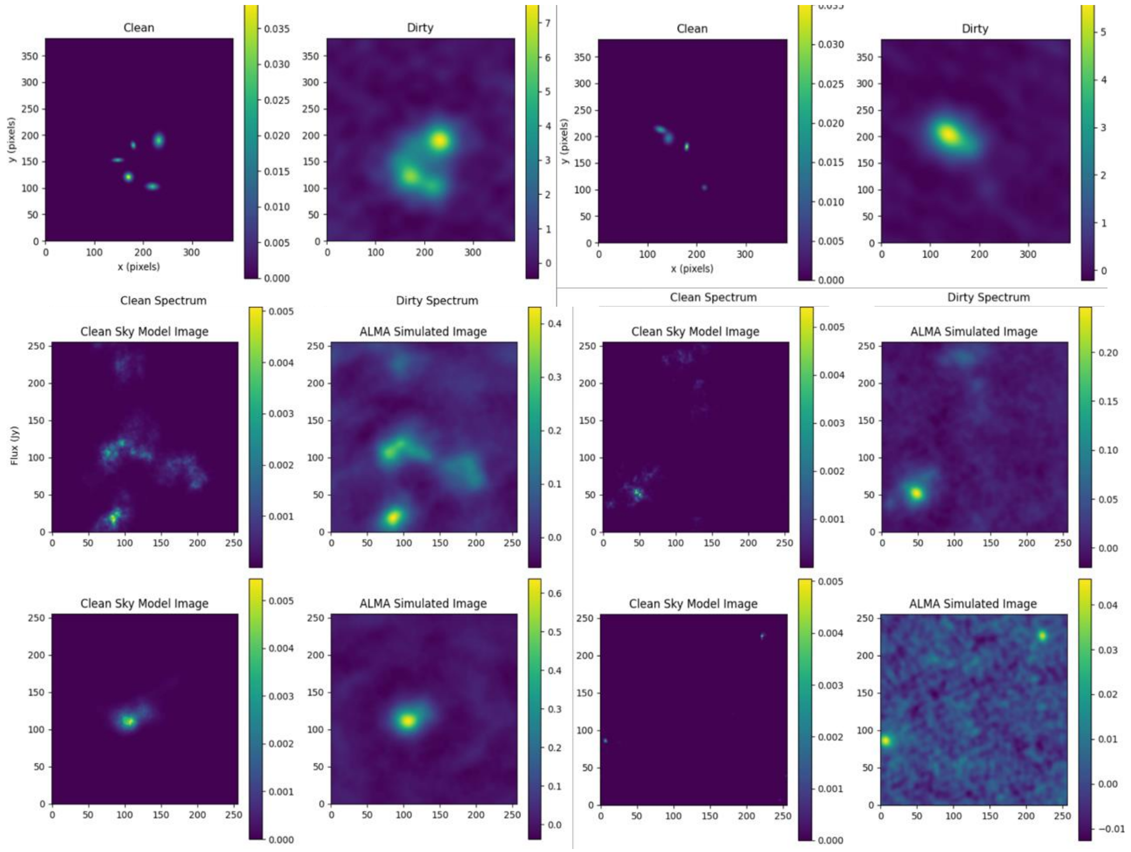
https://github.com/lbaronch/NOISEMPIRE

https://github.com/MicheleDelliVeneri/ALMASim.git

https://github.com/MicheleDelliVeneri/DeepFocus

ALMASim (Simulator)

- **Metadata retrieval** : query the ALMA Archive
- **Source reasoning** :
  - Extract fitting lines for the observation window
  - Calculate continuum and line fluxes
- **Sky model generation** :
  - Serendipitous sources
  - Storing source properties and observations on disk
- **Mock ALMA Observations** :
  - Accounts for atmospheric noise and gain errors
  - Save data products to disk



ALMASim: the ALMA Simulator for AI Developments

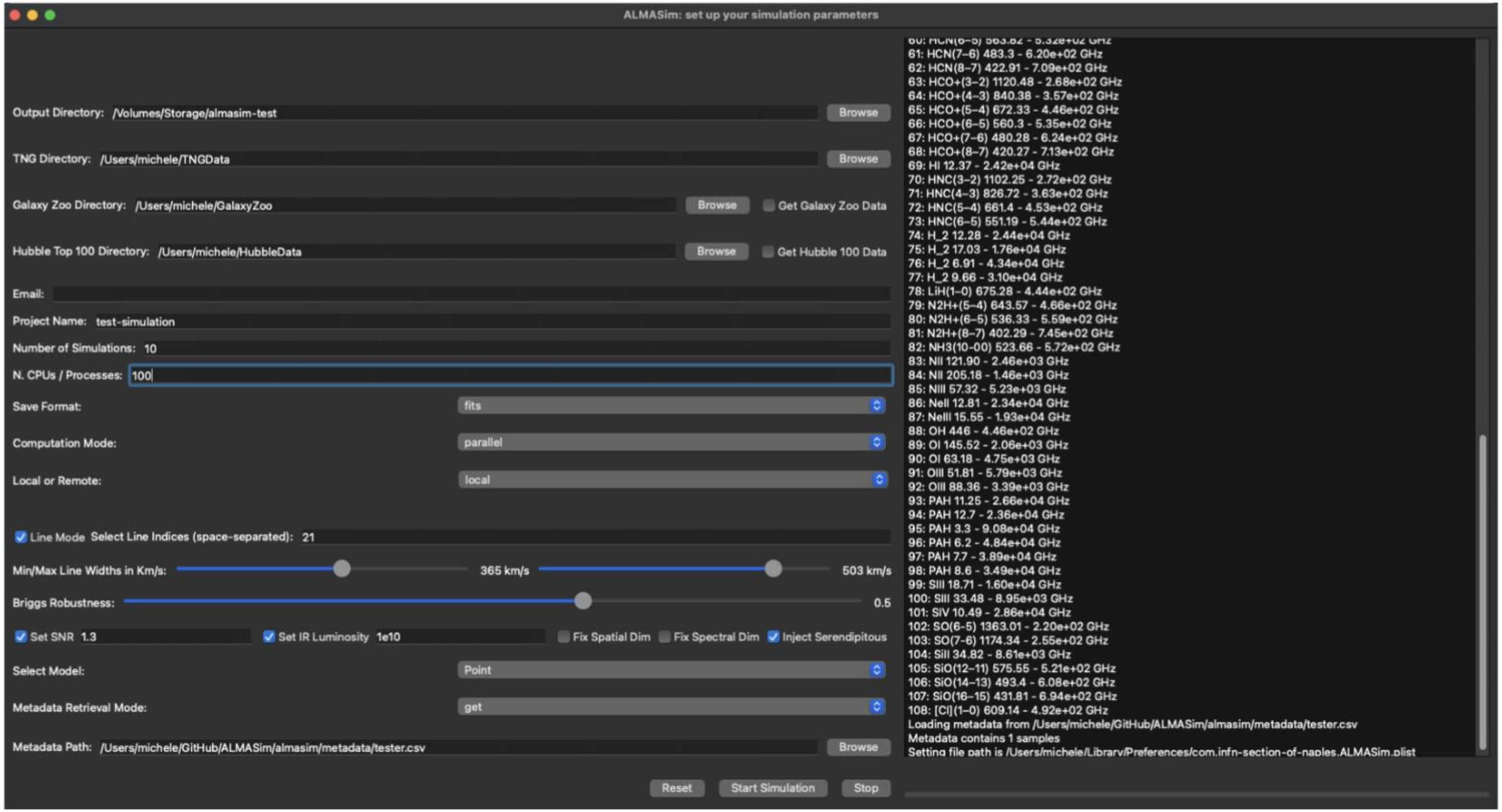


Figure 33: The ALMASim Graphical User Interface.

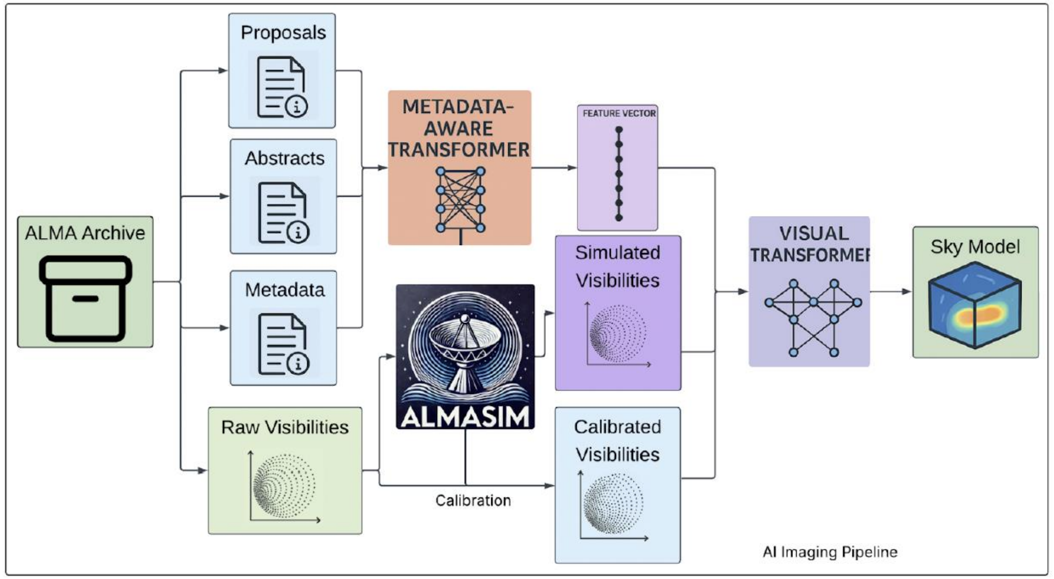
Deep Learning Models: Future extensions

MADIV: a Metadata Aware transformer for Direct Imaging from Visibilities *(P.I. Michele Delli Veneri)*

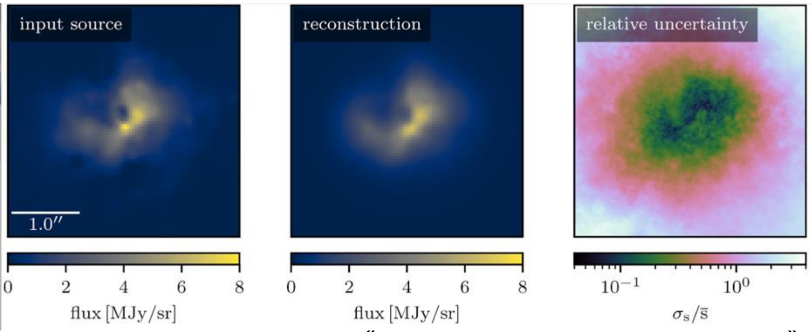
Approved ESO external ALMA Development Study

ESO/SKA partnership

- *Visual Transformer architecture as deconvolver/imager operating on visibility domain*
- *Metadata-Aware Transformer as prior knowledge on signal distribution*
- *ALMASim will learn from ALMA Archival data*
- *Diffusion models will be used to create simulated data*
- **Applications to ALMA Archival data**



Potential AI applications for ALMA



Rüstig+2024

RESOLVE
IFT + UQ
Improves
accuracy in
ALMA imaging

GitLab

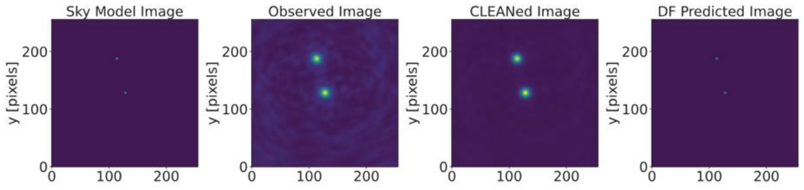


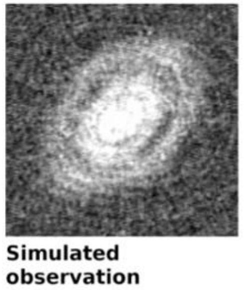
Fig. 1. – From left to right: Sky Model Image, Observed Image, CLEAN image, DF Prediction of the Sky Model.

Delli Veneri+2024

DeepFocus
Meta Learner
Enables
Real-time
ALMA imaging
+ cataloguing

GitHub

Baronchelli+2024



NOISEMPIRE 1.0.3 Noce

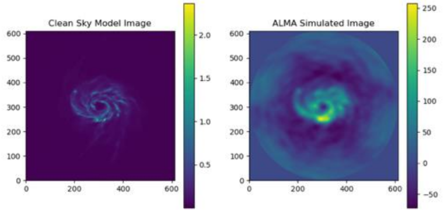
Empirical noise simulator

Simulates pure noise 2D images given a real ALMA cube or image in input. A real sky image (same size as reference image) can also be added.

GitHub

Real-time ALARM
Systems for
Observatory's
health?

ALMASim
Encodes +100
Molecular
transitions at
several z, HPC
enabled



Delli Veneri, ALMASim documentation

GitHub

Conclusion on Imaging beyond CLEAN

AI = Modern shift in Data Analysis	✓Objectivity above subjectivity ✓Model fidelity and statistical rigor ✓Data-driven approaches to overcome common limitations
RESOLVE	✓Improved the detection of faint signals, expanding the frontiers of astronomical discovery ✓Enhanced uncertainty quantification for more reliable results
DeepFocus	✓Real-time image analysis for immediate insights ✓Automated identification of interesting features, cataloguing
ALMASim	✓Enabled supervised ML approaches for more accurate modeling ✓Generated more realistic simulations for training AI models
Future	❑Improve AI techniques to image analysis, popularize to community, intensify data mining

Thank you!

GitHub GitLab

the BRAIN Study Team

fgugliel@eso.org

https://gitlab.mpcdf.mpg.de/ift/resolve

https://github.com/NIFTy-PPL

https://github.com/MicheleDelliVeneri/DeepFocus

https://pypi.org/project/ALMASim

https://github.com/lbaronch/NOISEMPIRE

@ESOAstronomy

@esoastronomy

@ESO

european-southern-observatory

@ESOobservatory