Uncertainty-Aware Visibility Gridding with VisCube

With applications to Bayesian inference and amortized learning

Michael (Misha) J. Yantovski-Barth
PhD Student at University of Montreal
Visiting Balzan Fellow at Oxford (Fall 2026)

Uncertain e Visibility Griddin isCube

With applications to ence and amortized

Michael (Misha) J. Yantovski-Barth

PhD Student at University of Montreal
Visiting Balzan Fellow at Oxford (Fall 2026)

Université de Montréal Mila Ciela Institute mitacs University of Oxford

Radio interferometers and Fourier-space measurements

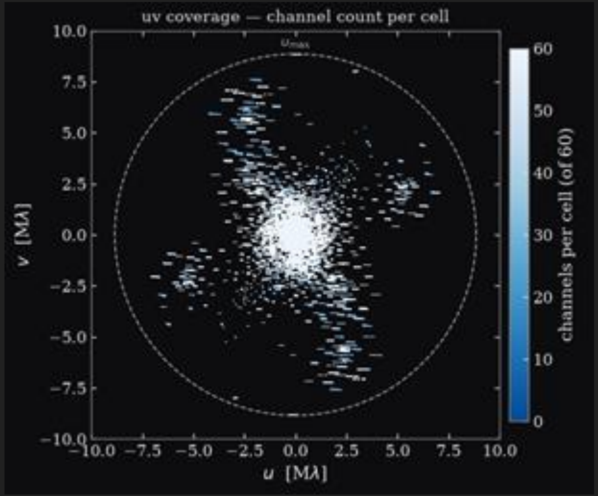
- Radio interferometers do not measure images
  - They irregularly sample the Fourier transform of the sky
    - Complex *visibilities*
  - Thermal noise
    - ~Gaussian and uncorrelated
- Visibilities ➢➢➢ images via iterative deconvolution
  - E.g., CLEAN
  - Introduces non-trivial noise correlations
  - Results depend on weighting, masking, and stopping criteria
- Alternative: fit a physical model directly to the visibilities
  - Can resolve signal below the CLEAN beam

Atacama Large Millimeter Array (ALMA)

Image credit: ALMA (ESO/NAOJ/NRAO)/W. Garnier

Problem statement: Visibilities are heterogeneous

- Radio interferometers: configurations vary between observations
  - Variable-length, continuous data
  - Sparse, high-dimensional, oversampled



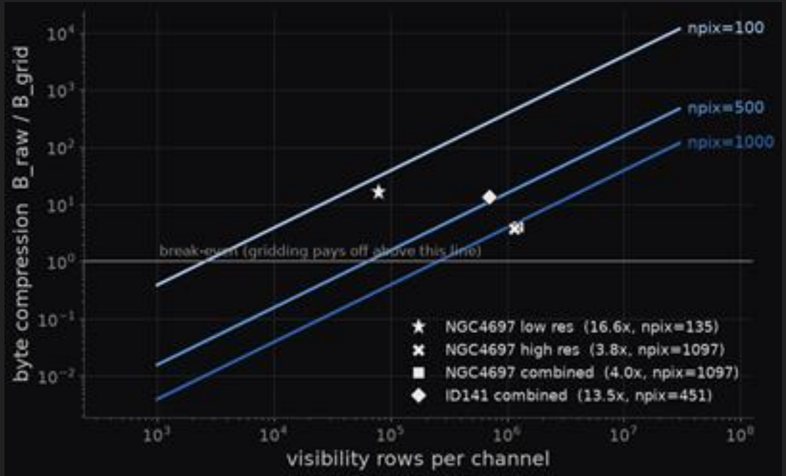
- FFT algorithm operates on regular grids
  - Non-uniform FFT adds interpolation cost

- Many neural architectures optimized for gridded data
  - CNNs, U-Nets, Transformers
    - Best for fixed, regular grids or sequences

Introducing VisCube

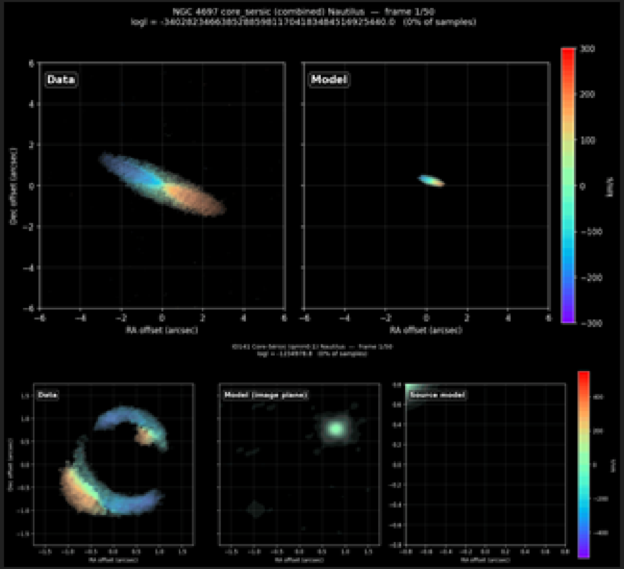
- Places visibilities onto a regular Fourier-space grid
  - Kaiser-Bessel window function weighted mean
- Calculates uncertainty for each grid cell
  - Weighted standard deviation
  - Uses per-scan neighbors for sparse cells
- Works with both continuum and spectral cubes

- Data compression ratios of ~3-20x
  - Depends on:
    - Field of view
    - Resolution
    - Number of visibilities

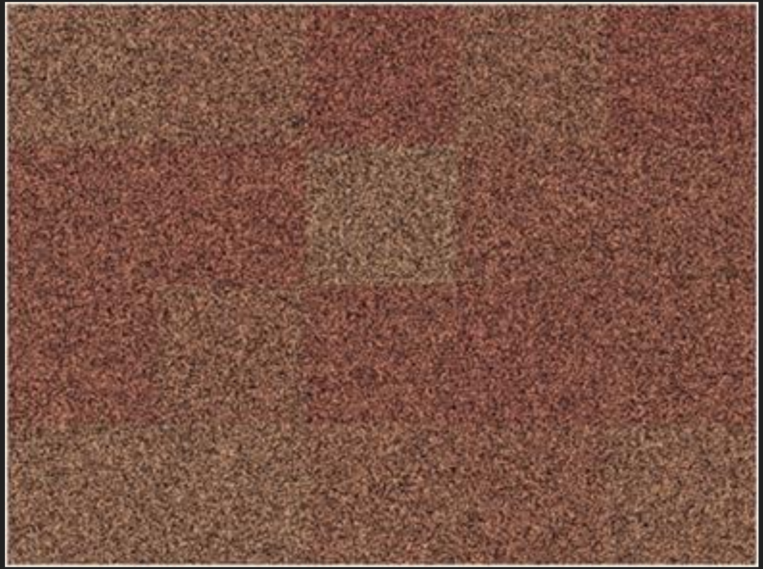


Applications of VisCube

Part 1:
Dynamical modeling of molecular gas in galaxies

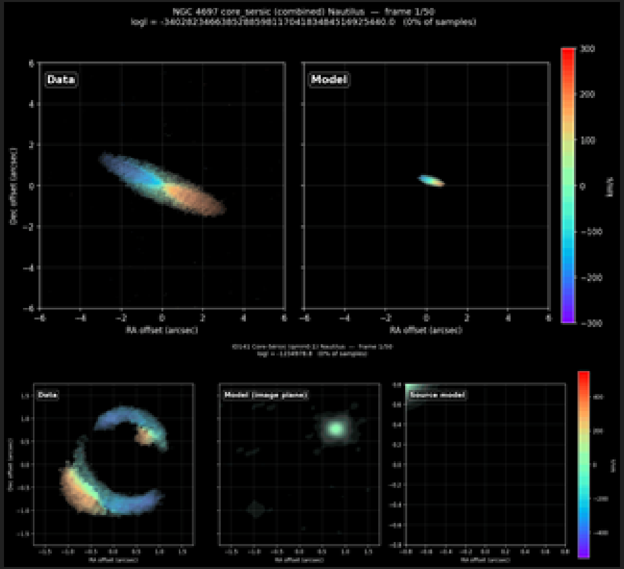


Part 2:
Neural imaging pipelines for interferometry data

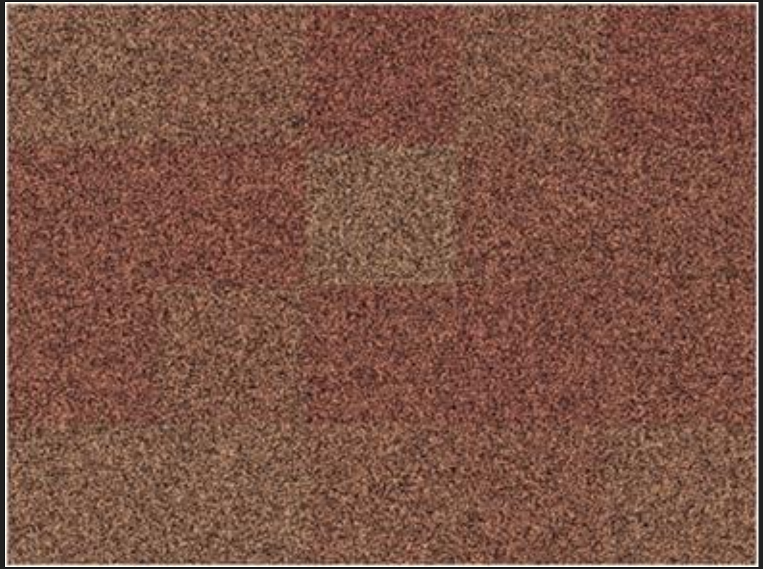


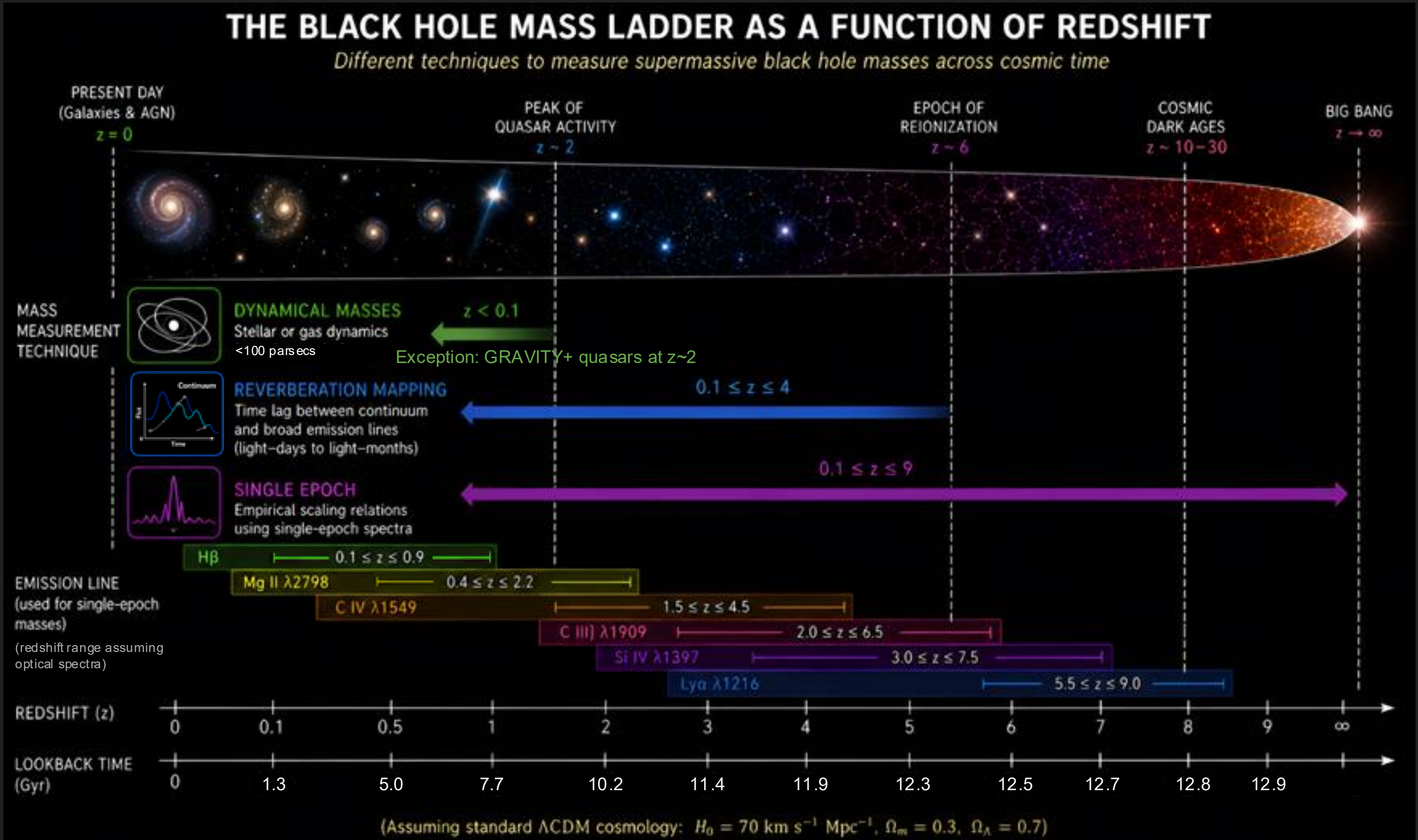
Applications of VisCube

Part 1:
Dynamical modeling of molecular gas in galaxies



Part 2:
Neural imaging pipelines for interferometry data

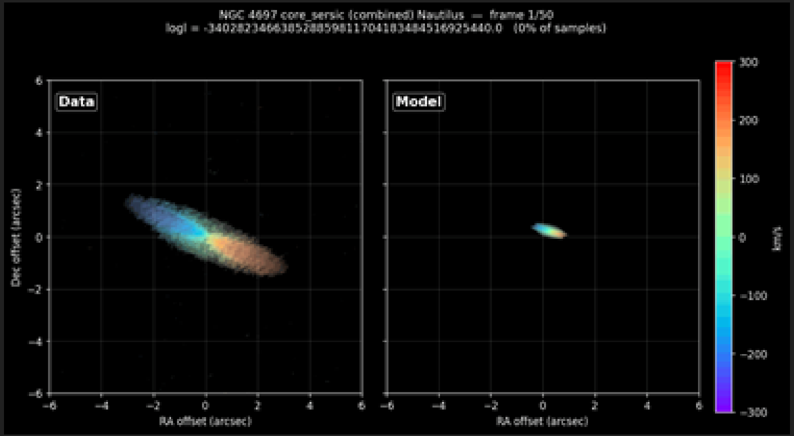


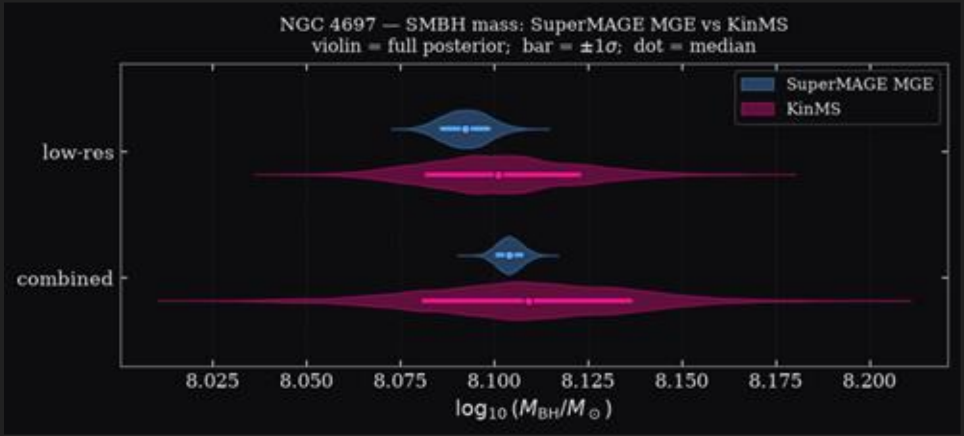


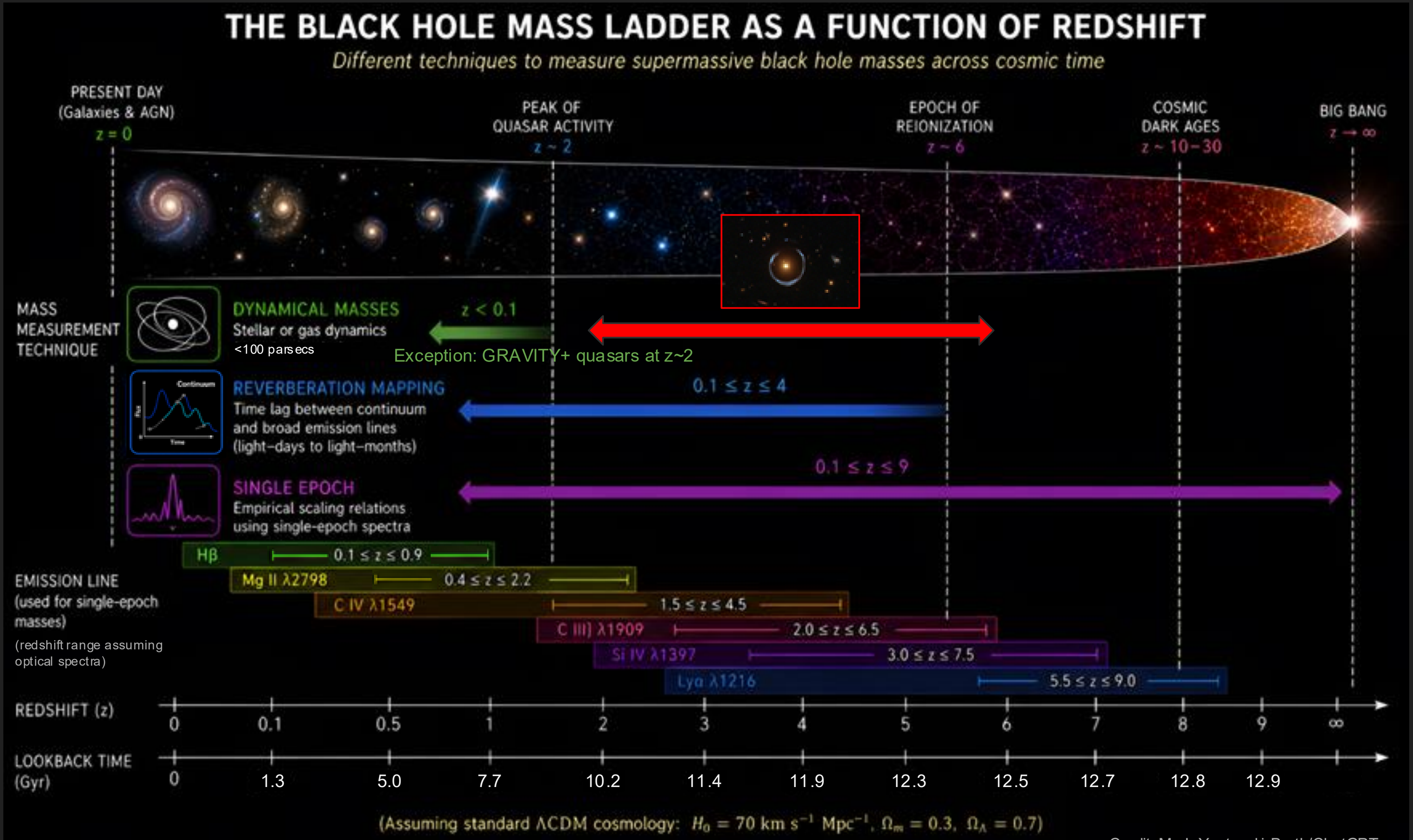
Credit: M. J. Yantovski-Barth/ChatGPT

Dynamical modeling of ALMA data: VisCube+SuperMAGE

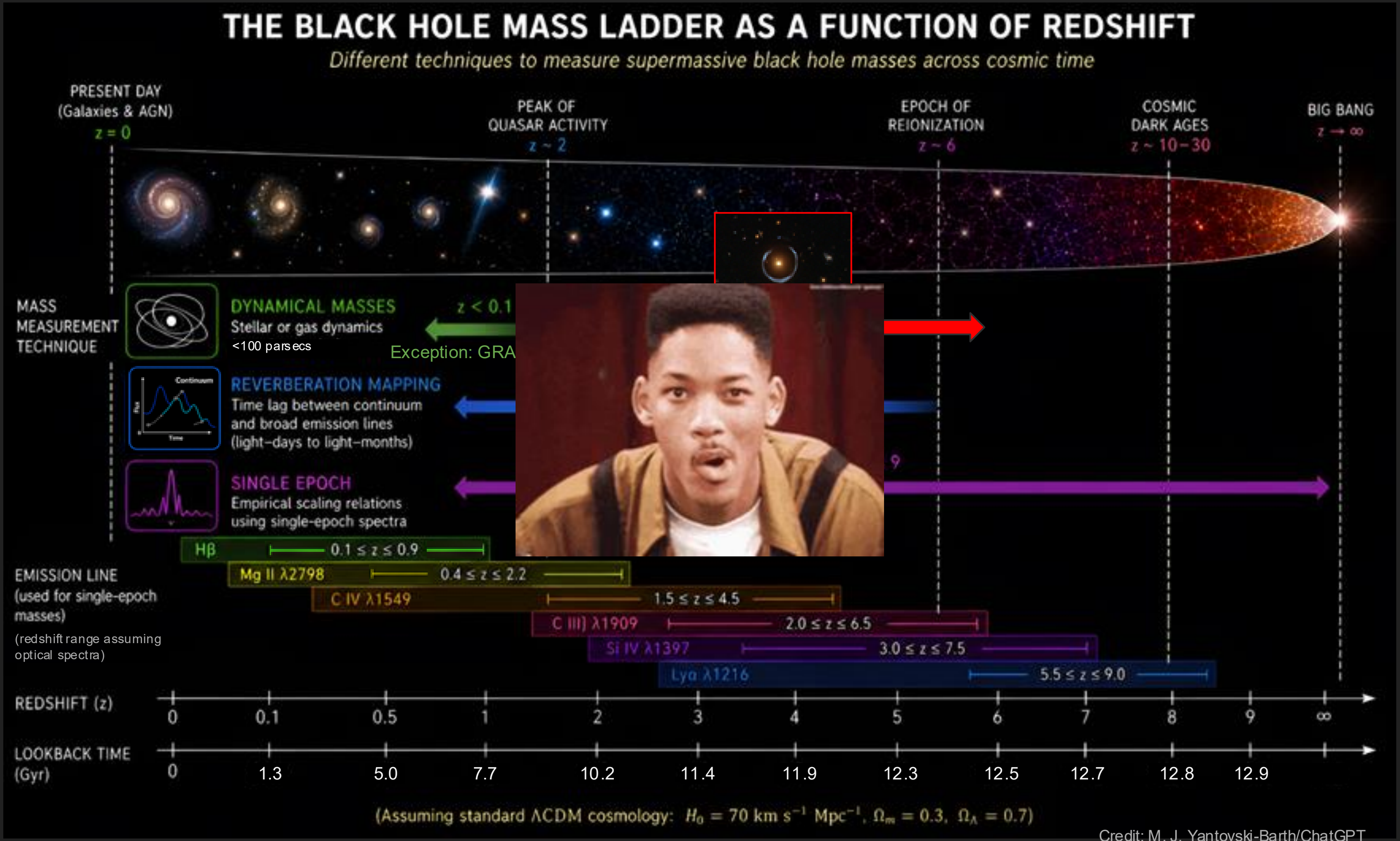
- Molecular gas dynamics via SuperMAGE
  - GPU-accelerated forward model (Yantovski-Barth et al., in prep)
  - Multiple-component fitting (supermassive BH, stars, gas, dark matter)
- Posterior inference via Nautilus (Lange 2023)
  - Nested sampling with neural network-driven exploration
  - Vectorized within and across GPUs





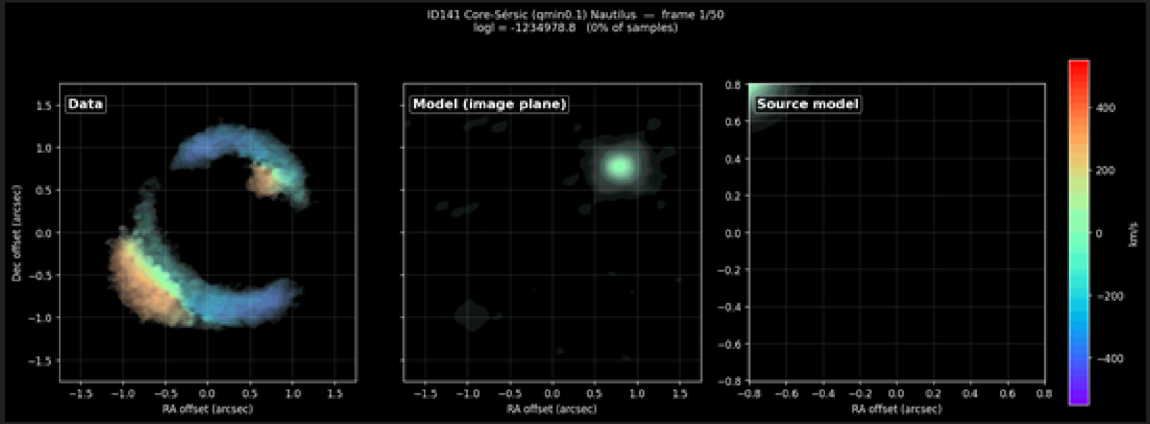


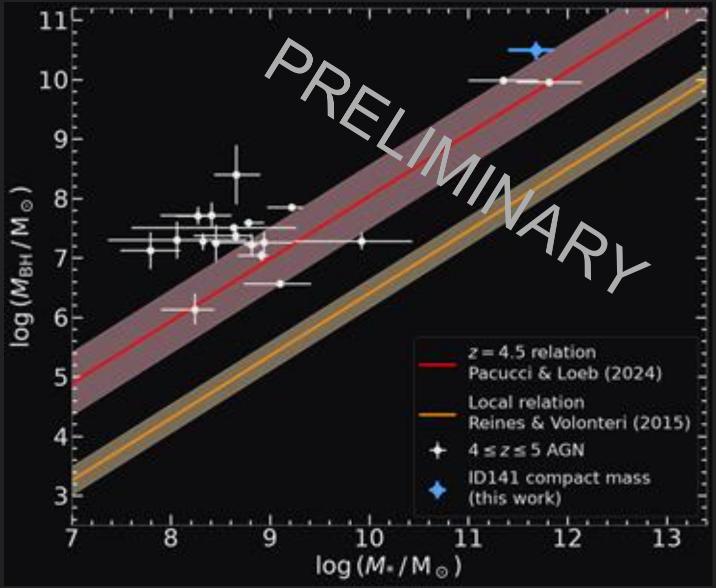
Credit: M. J. Yantovski-Barth/ChatGPT


THE BLACK HOLE MASS LADDER AS A FUNCTION OF REDSHIFT
Different techniques to measure supermassive black hole masses across cosmic time
PRESENT DAY (Galaxies & AGN) z = 0
PEAK OF QUASAR ACTIVITY z ~ 2
EPOCH OF REIONIZATION z ~ 6
COSMIC DARK AGES z ~ 10–30
BIG BANG z → ∞
MASS MEASUREMENT TECHNIQUE
DYNAMICAL MASSES
Stellar or gas dynamics
<100 parsecs
z < 0.1
Exception: GRA
REVERBERATION MAPPING
Time lag between continuum and broad emission lines (light-days to light-months)
SINGLE EPOCH
Empirical scaling relations using single-epoch spectra
EMISSION LINE (used for single-epoch masses)
(redshift range assuming optical spectra)
Hβ 0.1 ≤ z ≤ 0.9
Mg II λ2798 0.4 ≤ z ≤ 2.2
C IV λ1549 1.5 ≤ z ≤ 4.5
C III] λ1909 2.0 ≤ z ≤ 6.5
Si IV λ1397 3.0 ≤ z ≤ 7.5
Lyα λ1216 5.5 ≤ z ≤ 9.0
REDSHIFT (z) 0 0.1 0.5 1 2 3 4 5 6 7 8 9 ∞
LOOKBACK TIME (Gyr) 0 1.3 5.0 7.7 10.2 11.4 11.9 12.3 12.5 12.7 12.8 12.9
(Assuming standard ΛCDM cosmology: H_0 = 70 km s^{-1} Mpc^{-1}, Ω_m = 0.3, Ω_Λ = 0.7)
Credit: M. J. Yantovski-Barth/ChatGPT

VisCube+SuperMAGE: SMBH masses in lensed galaxies

- Strong lensing models via Caustics (Stone et al. 2024)
    - Integrated into SuperMAGE via Caskade
- <u>Dynamical mass constraint on a *z=4.2* supermassive black hole</u>
    - SNR and resolution frontier of ALMA

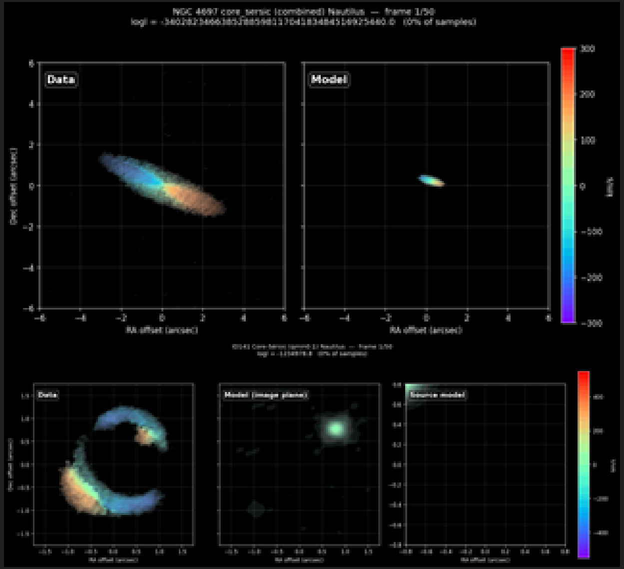




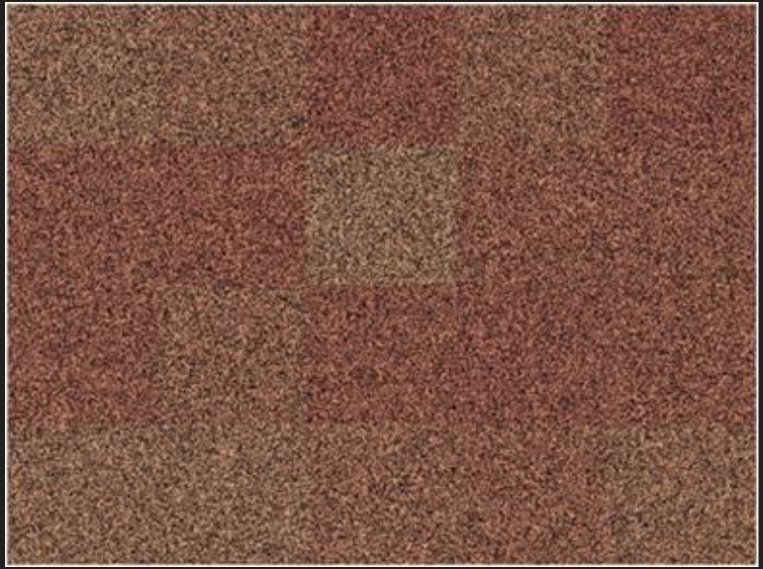
ALMA PI: Hengyue Zhang
(Oxford >>> MPIA)

Applications of VisCube

Part 1:
Dynamical modeling of molecular gas in galaxies



Part 2:
Neural imaging pipelines for interferometry data

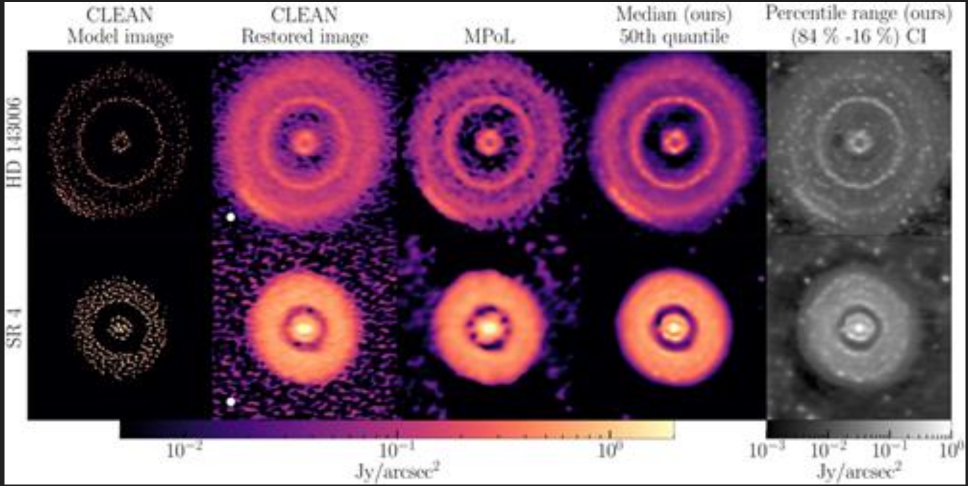


Neural imaging: Protoplanetary disks via diffusion models

- Every pixel is a parameter
  - Bayesian inference in high-dimensional space
  - Linear forward model (FFT+mask)
- IRIS: Diffusion models as high-dimensional priors
  - Sample posterior distribution of images corresponding to visibilities

Noé Dia

Alexandre Adam



Neural imaging of visibilities: other projects using VisCube

Anabel Q. Tan
MILA – Quebec AI Institute,
University of Montreal

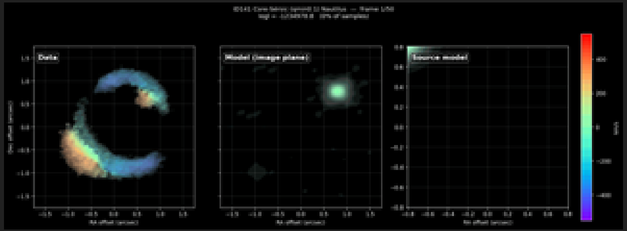
- TransformerRIM: High dynamic range interferometric imaging with Swin Transformers
  - Recurrent training + inference to recover faint structures while retaining information
  - VisCube provides uncertainties as input to the Transformer for added robustness

- Simulation-Based Inference (SBI) for detecting structure in protostellar disks
  - Neural network trained on VisCube-gridded simobserve output

Sidenote: 2018 NRAO REU (Laura Fissel)

Thank you for listening!

- VisCube: Visibility-space gridder with uncertainty quantification
  - Pros: Data compression, reliable noise statistics
- Applications: Bayesian posterior inference, uncertainty-aware learning
  - Bayesian dynamical modeling of molecular gas
    - Precision supermassive black hole masses
    - Efficient nested sampling
  - Bayesian neural imaging pipelines
    - Diffusion model prior for high-dimensional inference
    - Posterior samples of images
  - External projects: Iterative neural imaging pipelines and Simulation-Based Inference



Galaxy Dynamics with SuperMAGE

SuperMAGE

Superb **MA**sses from **G**as kin**E**matics

GPU-accelerated, auto-differentiable

Forward Modeling Pipeline:

VisCube

Gridding

Gas Distribution

Light

Stellar model

MGE (+BH)

Velocity Profile

Spectral image cube

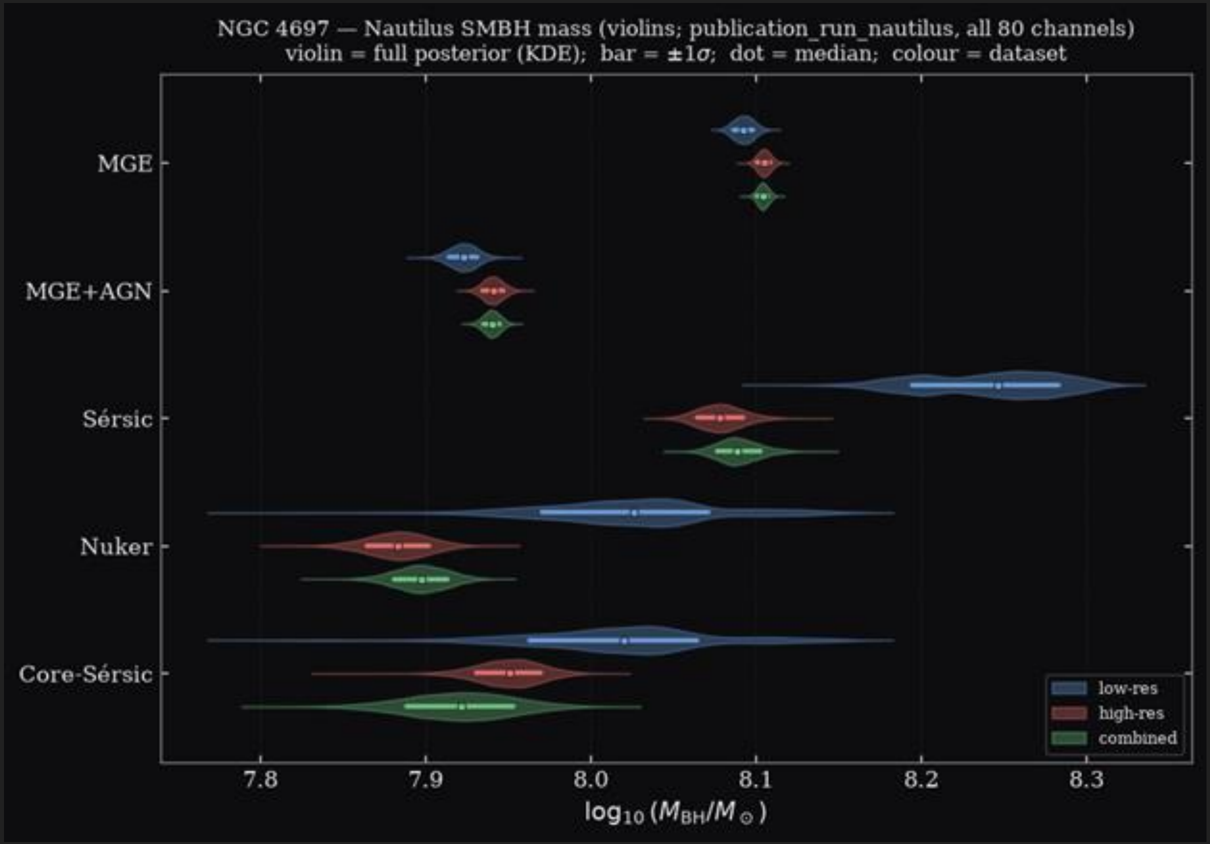
FFT

Visibility space cube

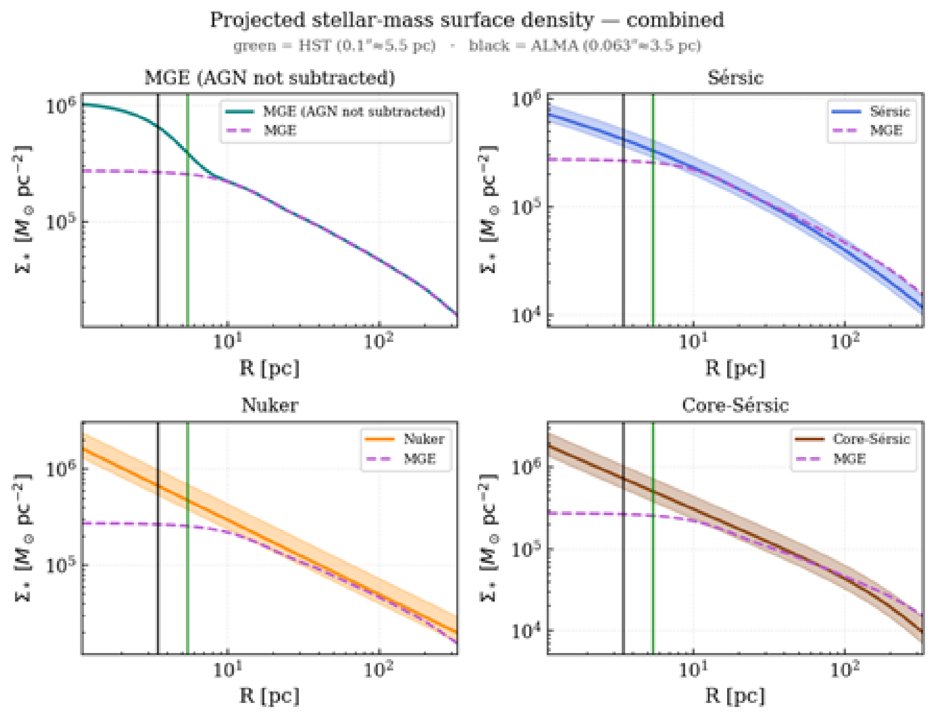
ALMA Data

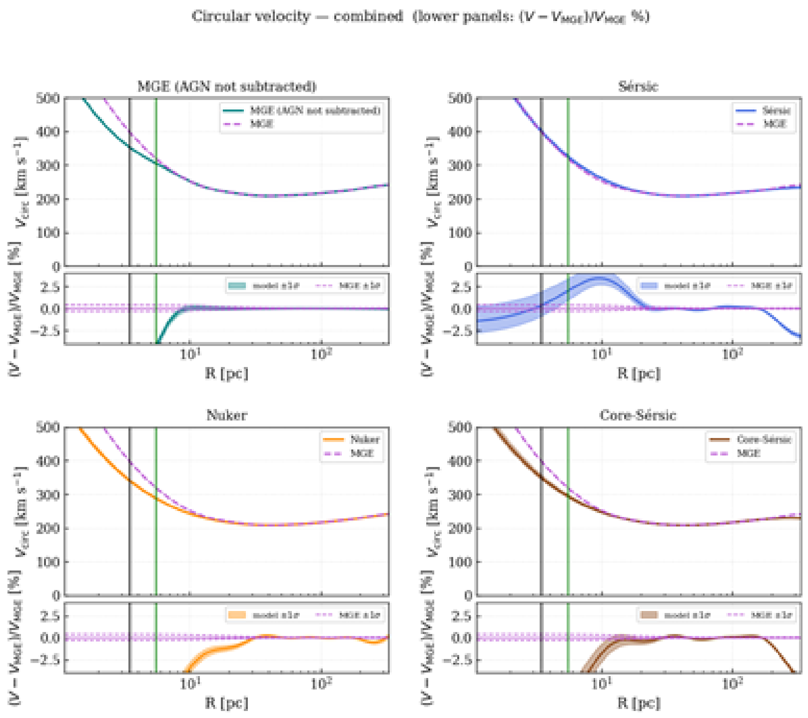
Key innovation: change of basis between 2D models (on a radial grid) and MGE via ridge regression/Tikhonov regularization

SuperMAGE stellar mass models + datasets



Stellar Mass and Velocity Profiles with SuperMAGE

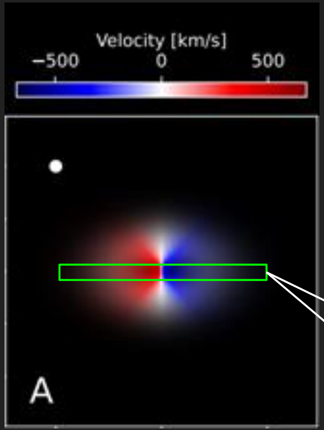


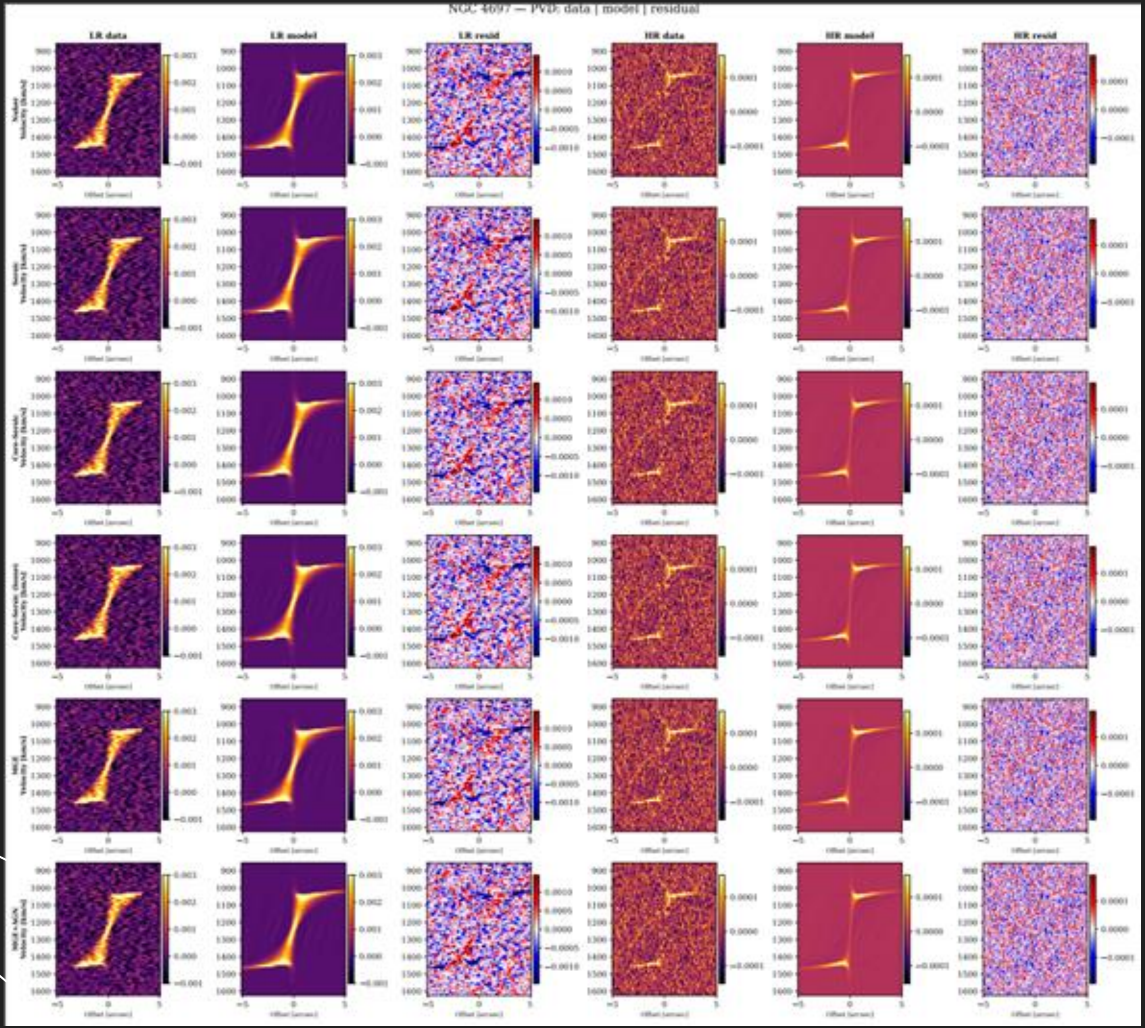


Residuals: NGC4697

What is a PVD?

Position-Velocity Diagram:





How does Caustics fit inside SuperMAGE? Caskade!

- Caskade: Differentiable physics simulator infrastructure
  - Stackable, modular simulators
    - Handles parameter unpacking
    - Facilitates user customization
- Examples of simulators built with Caskade
  - SuperMAGE: Galactic dynamical modeling
  - Caustics: gravitational lensing
  - More to come!

CASKADE

Ciela Institute

CAUSTICS

SuperMAGE

The Bayes Theorem: Score version

- Bayes in terms of scores:

$$\nabla log(p(\theta|V)) = \nabla log(p(V|\theta)) + \nabla log(p(\theta))$$

- No normalization needed (no “evidence” term)
- Posterior is sum of likelihood and prior

Diffusion models as “plug-and-play” priors

- Bayes in terms of scores:

$$\nabla log(p(\theta|V)) = \nabla log(p(V|\theta)) + \boxed{\nabla log(p(\theta))}$$

- Replace prior term (right) with the output of the diffusion model
- Sample the score of the posterior
    - Annealed Langevin sampling

TransformerRIM

Ground Truth
(PROBES)

WSClean
(1000 ite)

TransRIM
(3 FA x 100 epochs)