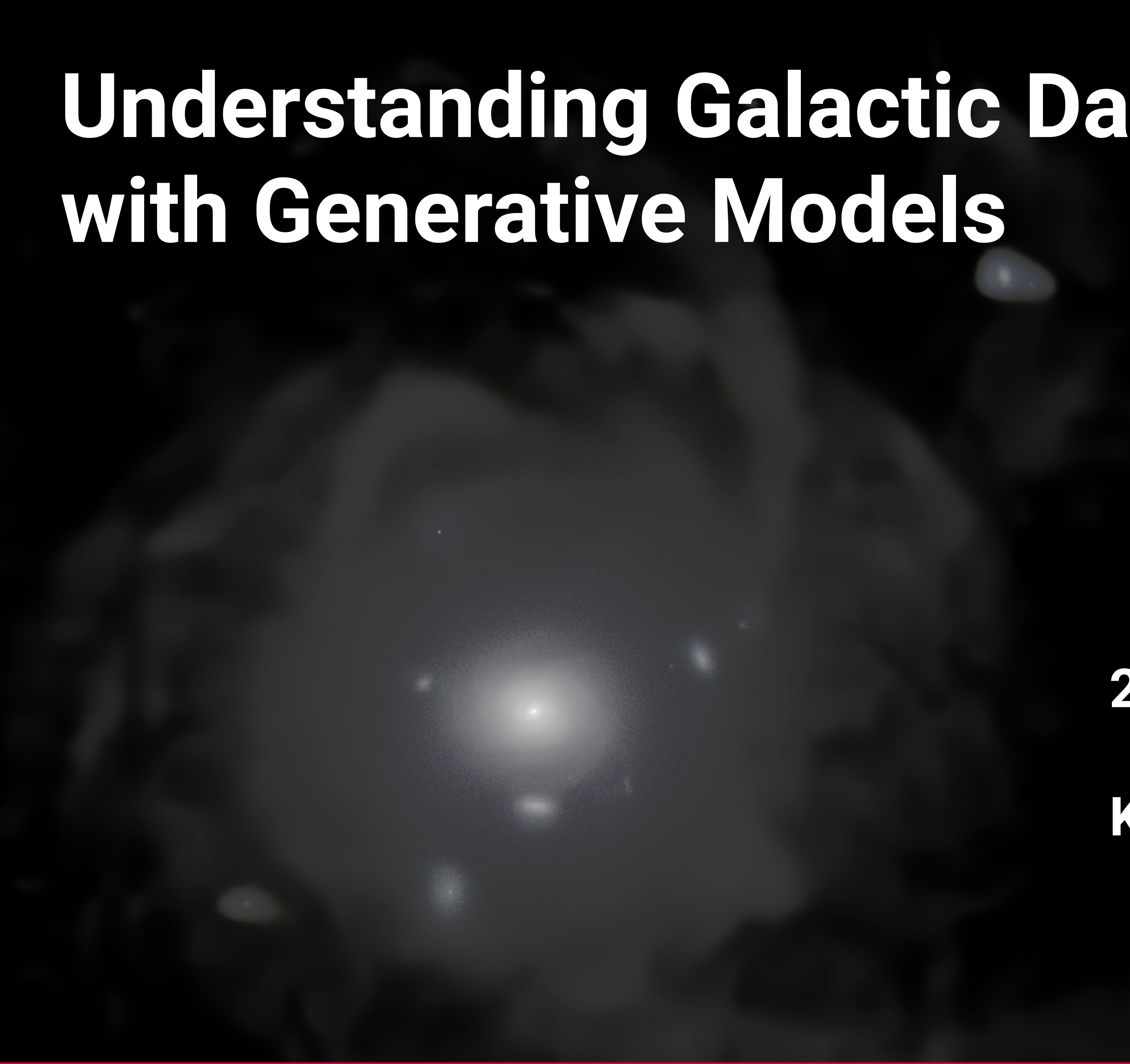
Understanding Galactic Dark Matter with Generative Models



2nd AI+HEP in East Asia

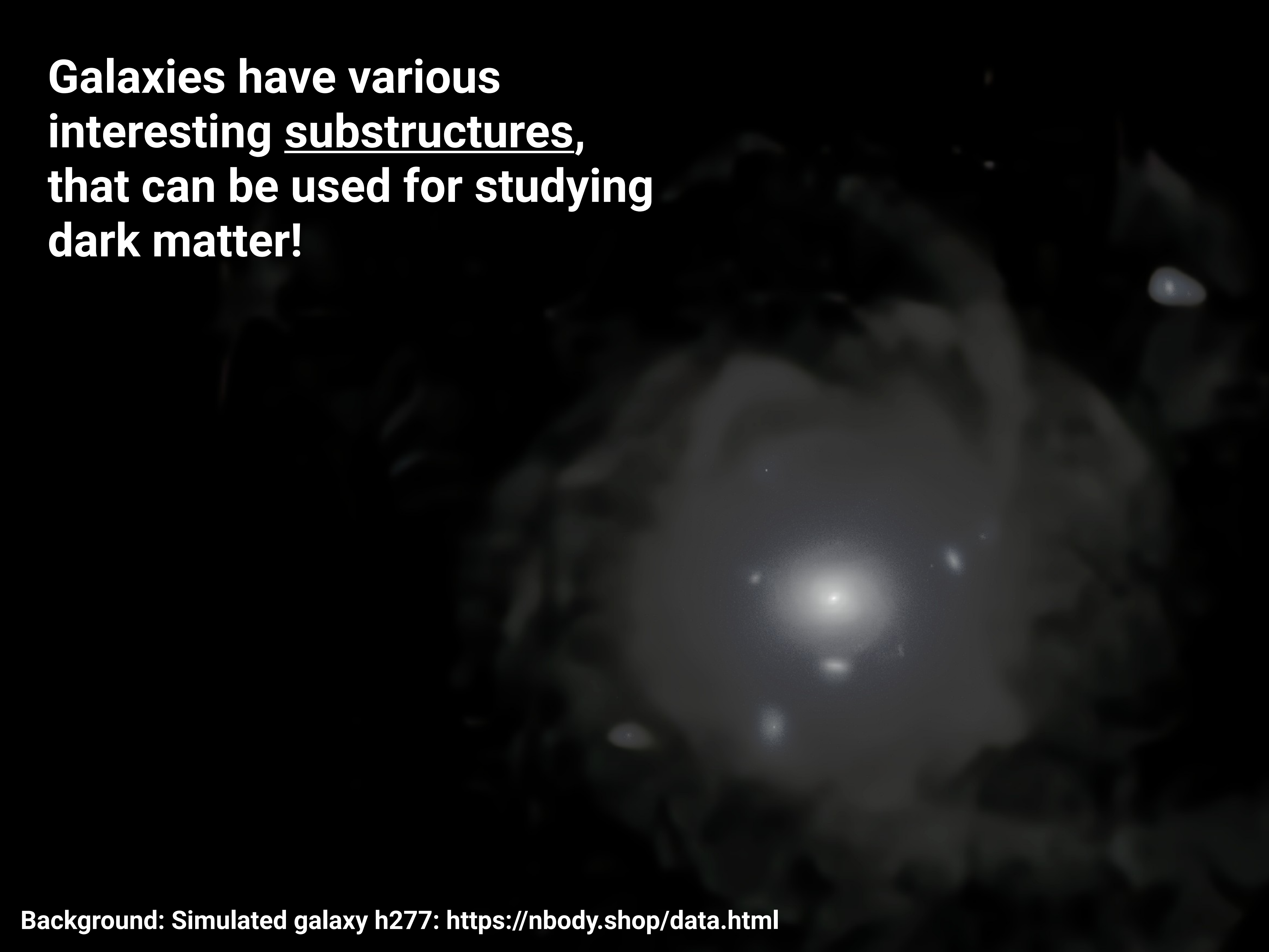
KEK, Tsukuba, Japan

Jan. 2025

Sung Hak Lim
CTPU-PTC, IBS

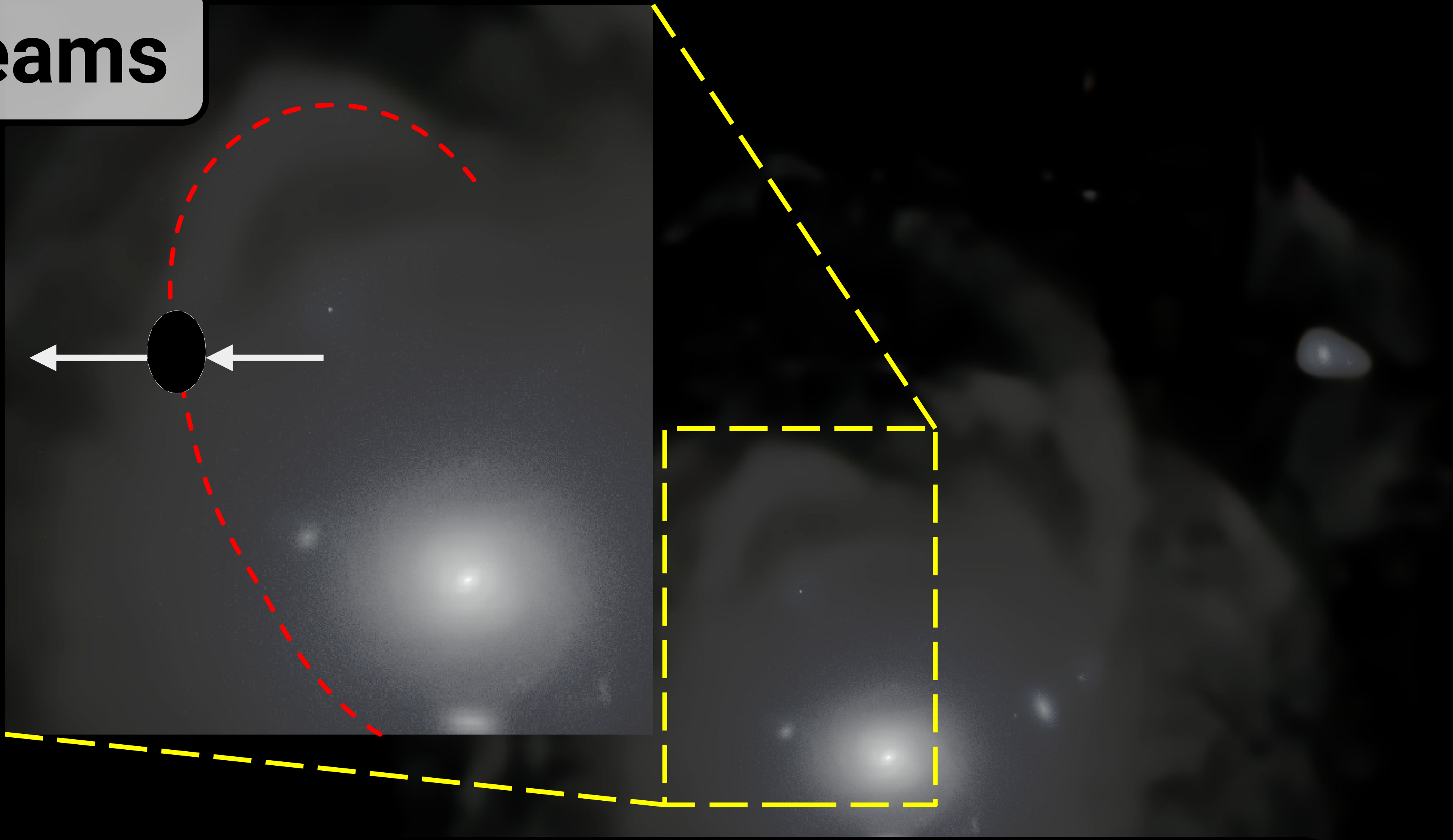
Institute for Basic Science

Galaxies have various interesting substructures, that can be used for studying dark matter!



Background: Simulated galaxy h277: https://nbody.shop/data.html

Streams

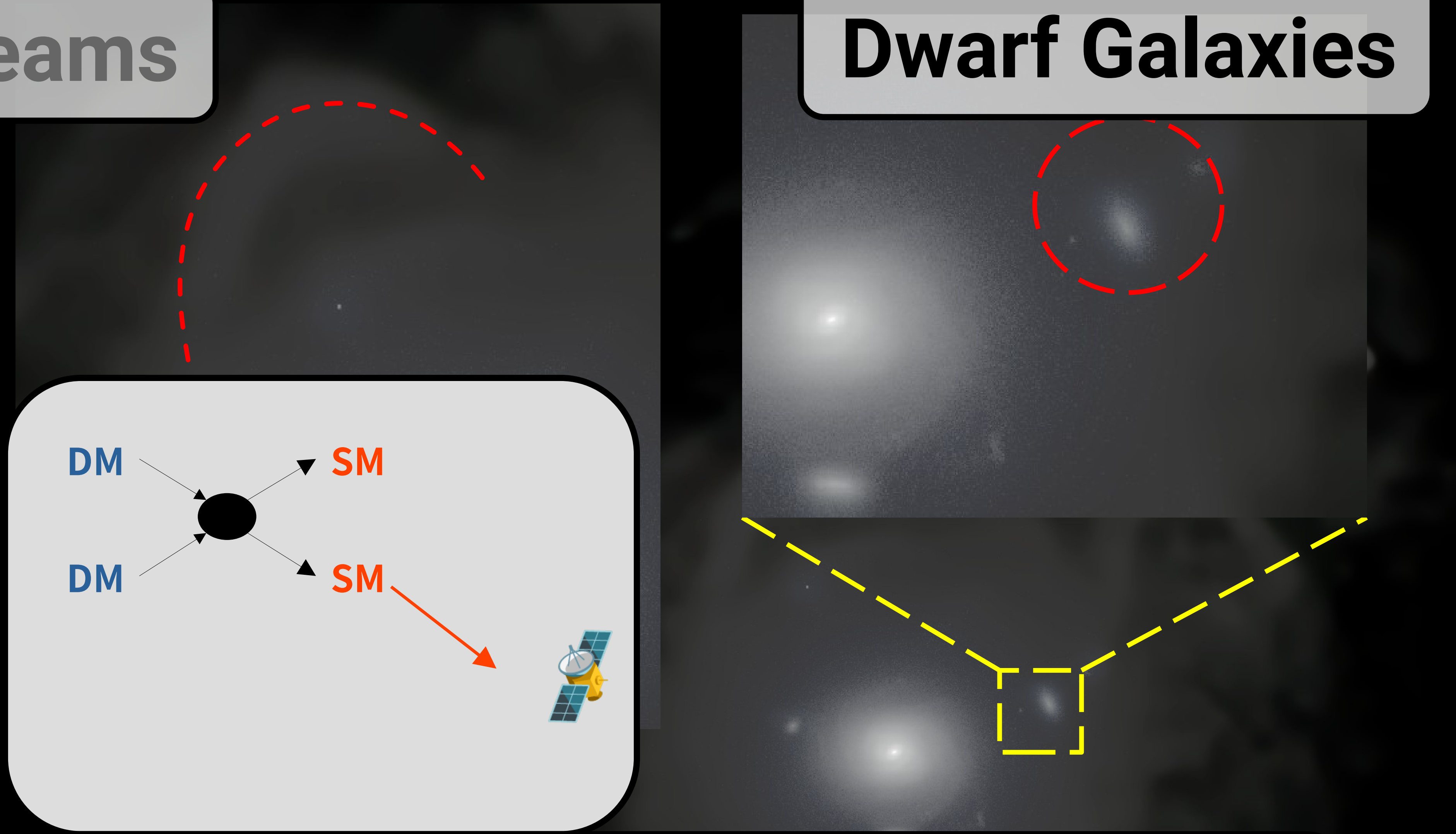


Stars on the same orbit, originating from globular clusters or dwarf galaxies tidally stripped

- Orbit is visible → measuring gravity
- Any gaps in streams may be a sign of collision with dark matter subhalo!

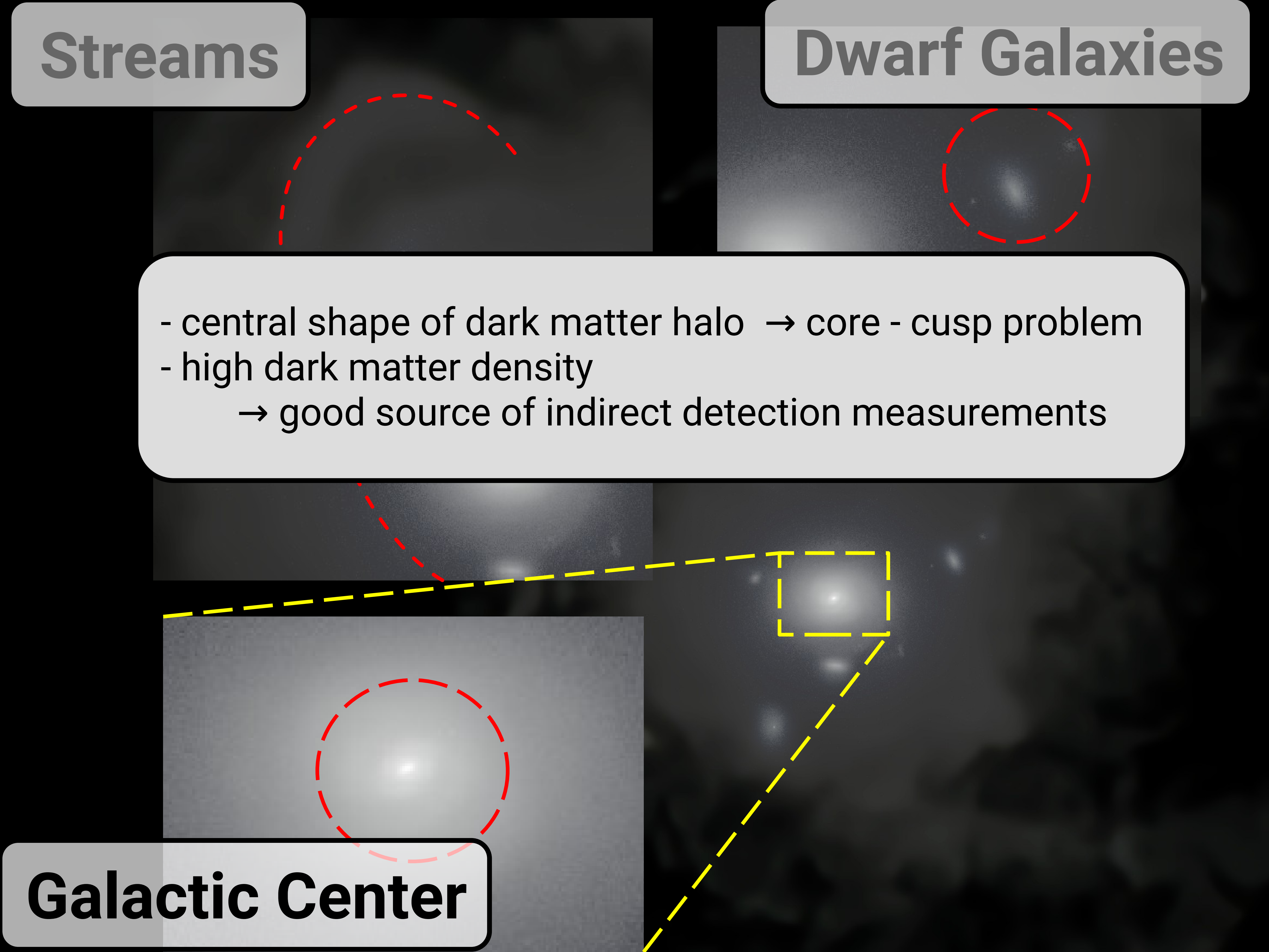
Streams

Dwarf Galaxies



A round and faint satellite galaxy orbiting the galaxy.

- whole structure is clearly visible → core - cusp problem
- less baryonic activity
    → good source of indirect detection measurements

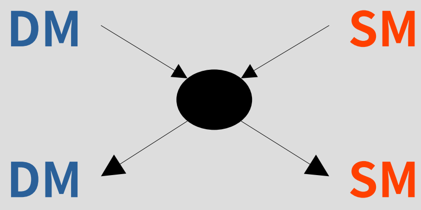


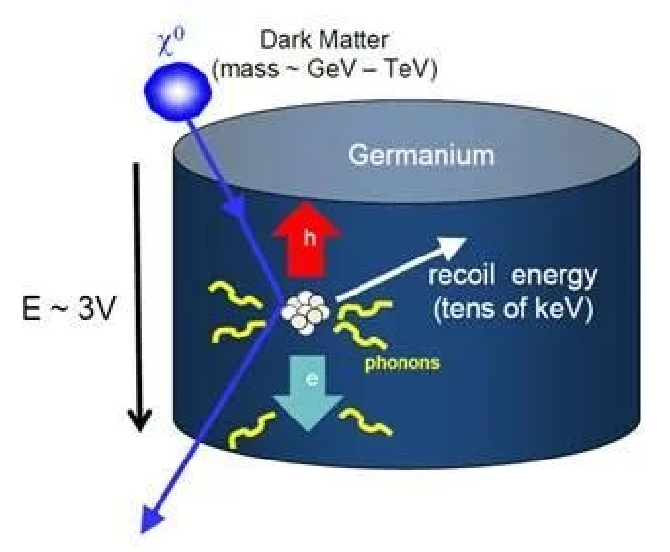
- central shape of dark matter halo → core - cusp problem
- high dark matter density
  → good source of indirect detection measurements

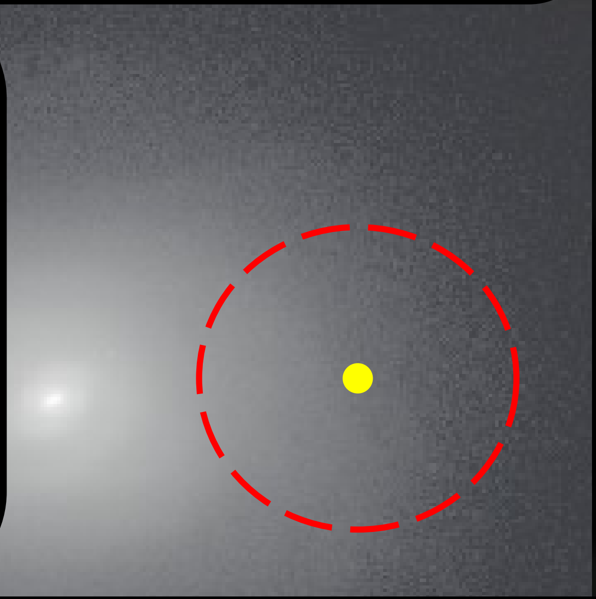
Streams

Dwarf Galaxies

- nearby stars → highly precise and complete measurement
  → high quality measurement of local dark matter density
- measuring dark matter density around solar system
  → relevant for direct detection of dark matter

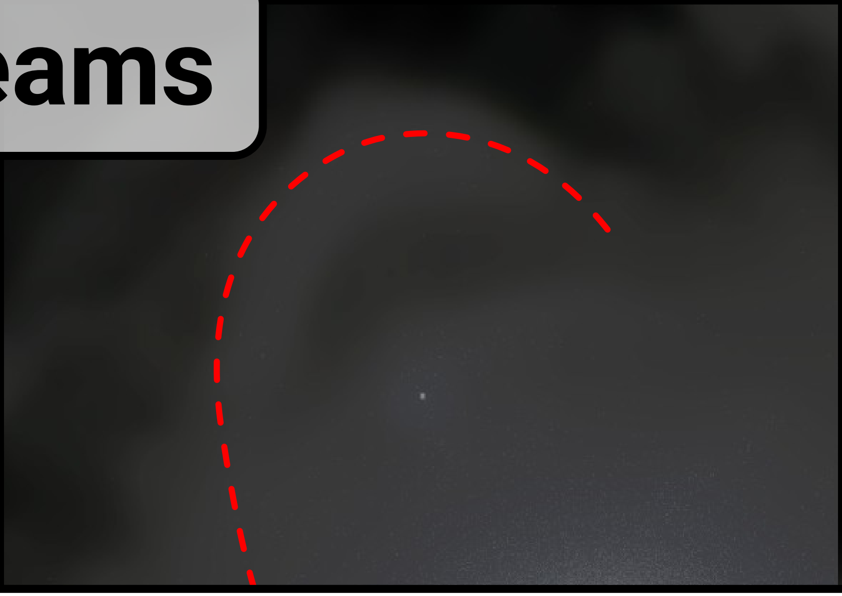






Galactic Center

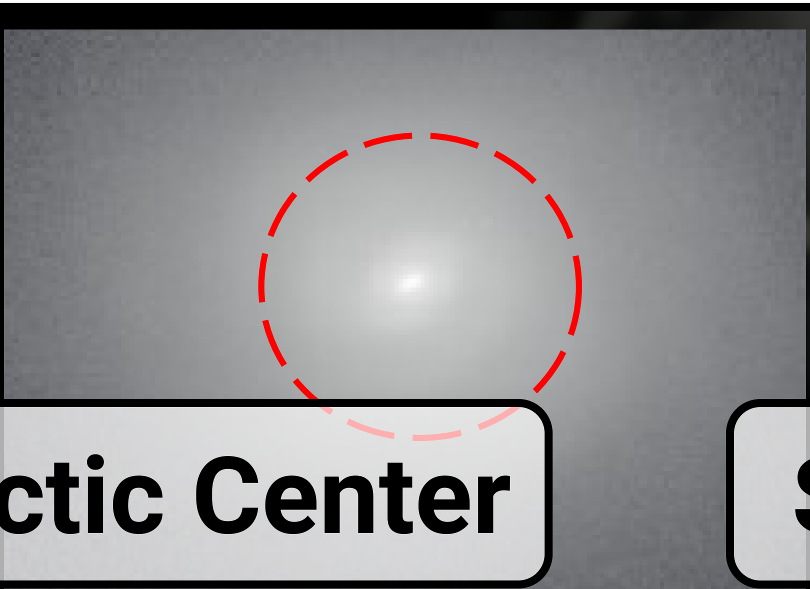
Solar Neighborhood



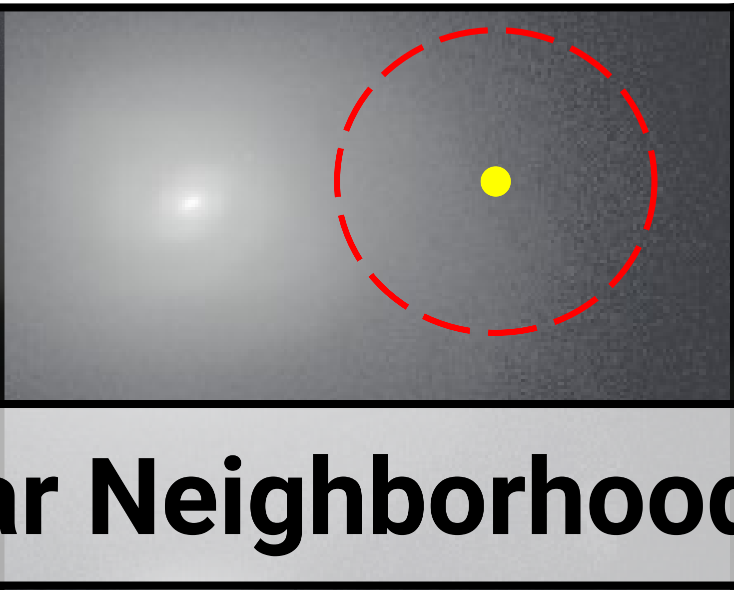
Streams

Dwarf Galaxies

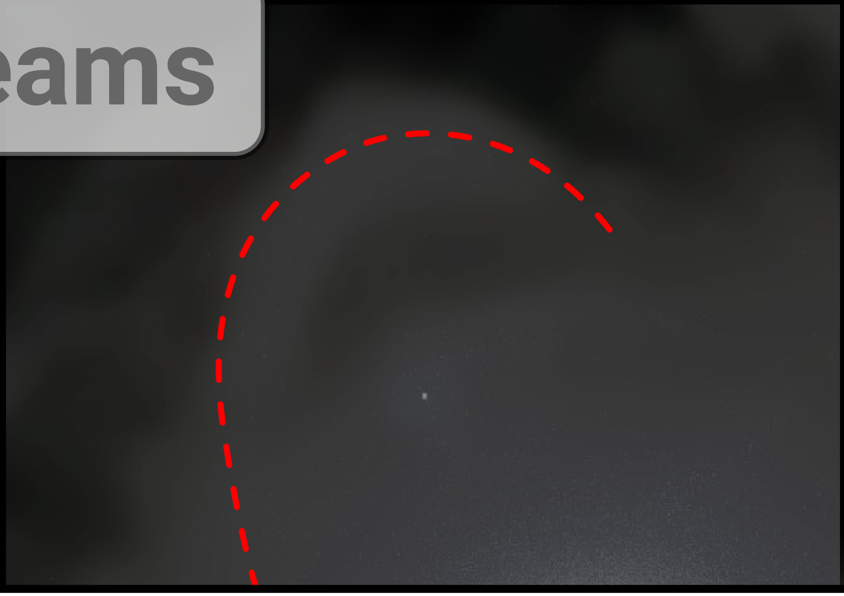
We can study the details of galactic dark matter by analyzing these substructures!



Galactic Center



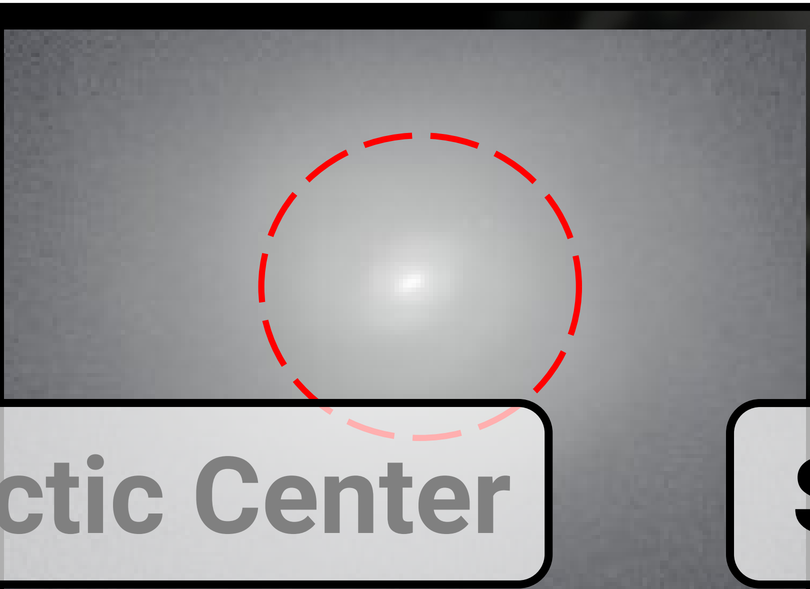
Solar Neighborhood



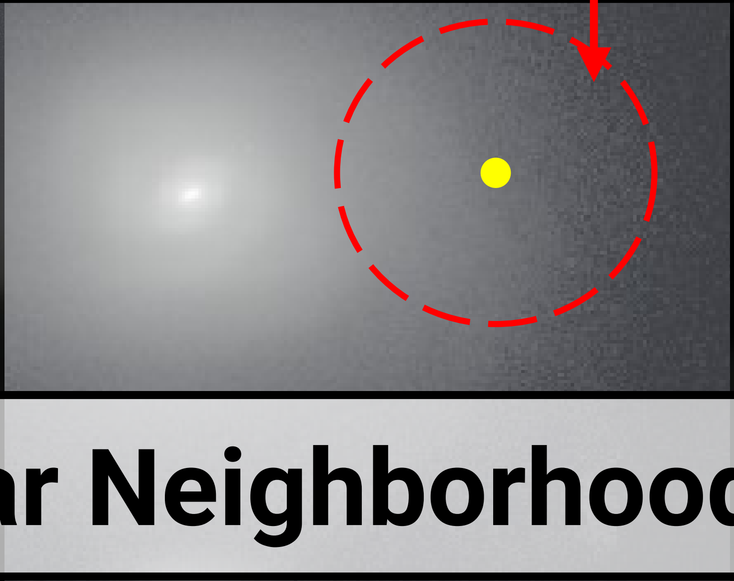
Streams

Dwarf Galaxies

My current research focuses on these two.

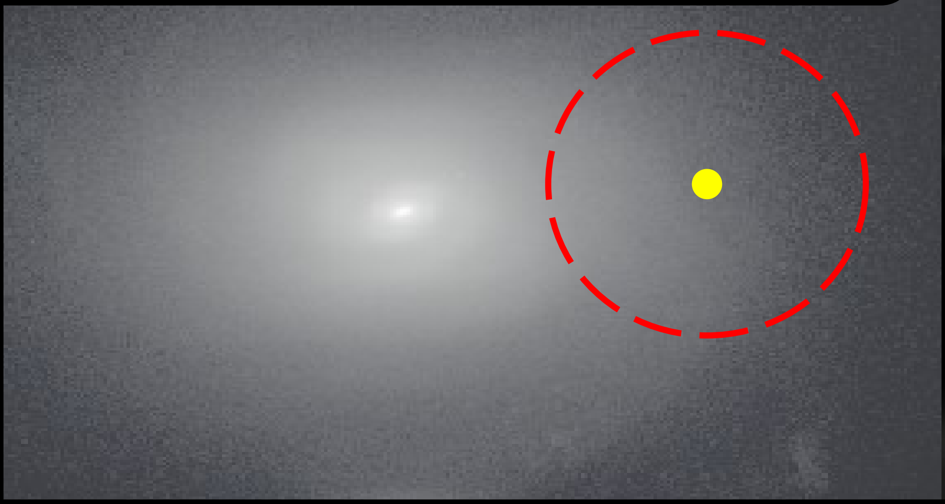


Galactic Center



Solar Neighborhood

Solar Neighborhood



Dwarf Galaxies

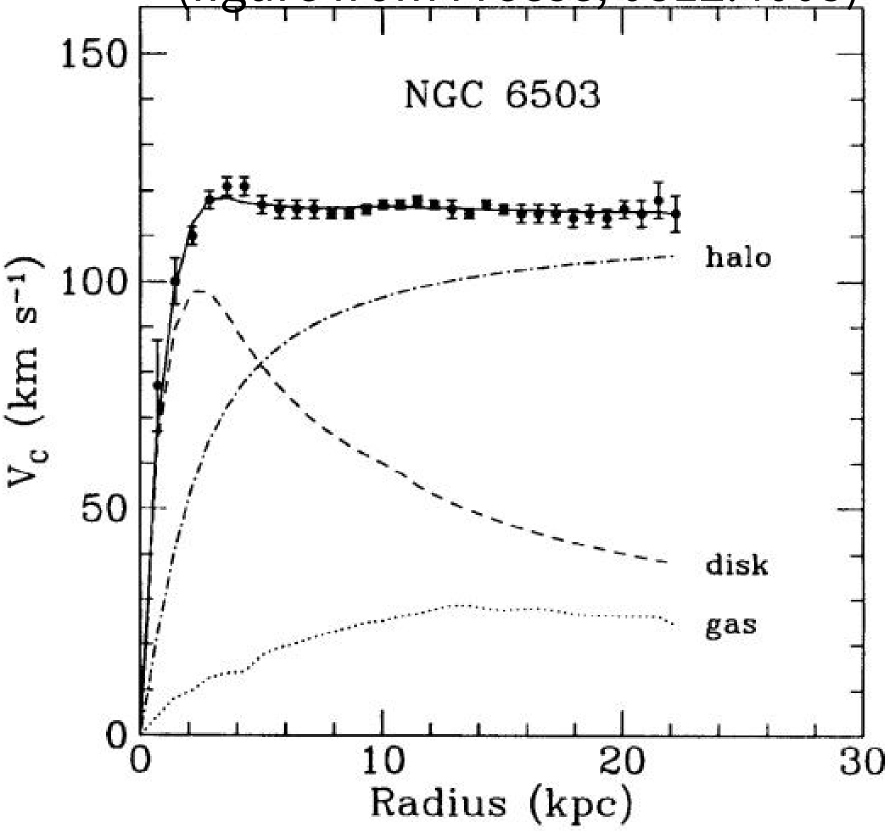
Talk title:

Understanding Galactic Dark Matter with Generative Models

But, why do we need to use machine learning and generative models to analyze these galactic structures?

Return to old school example: Galaxy rotation curve and dark matter

(figure from Freese, 0812.4005)



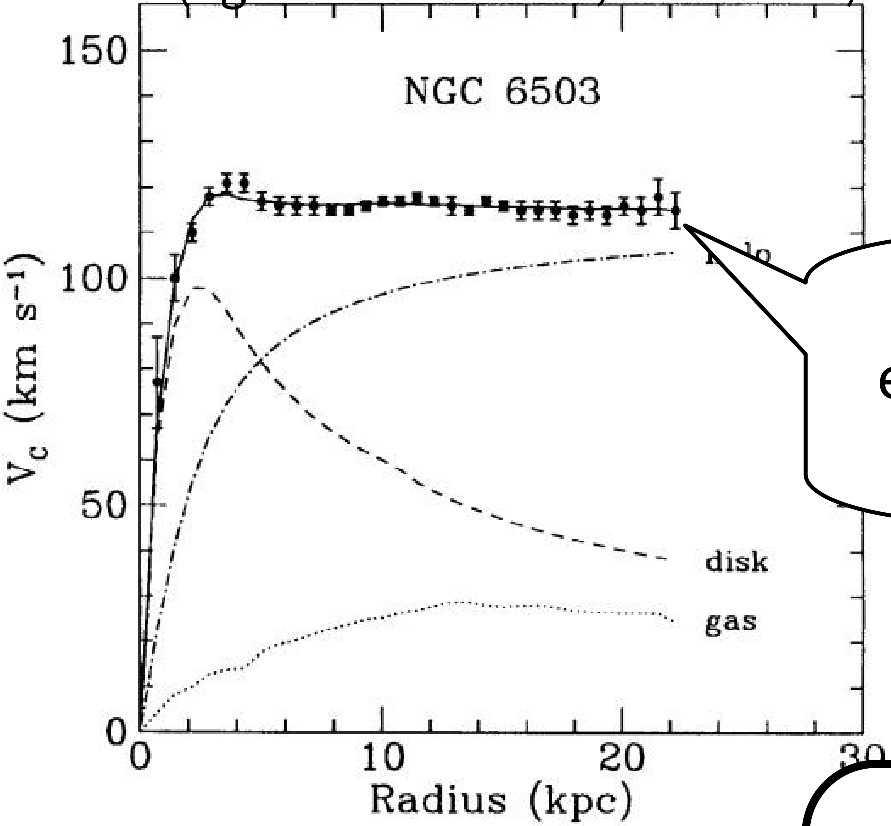
NGC 6503 from NASA Hubble telescope

$$v_{\rm circ}(R) = \sqrt{\frac{GM(R)}{R}}$$

Obtain mass density from enclosed mass M(R)

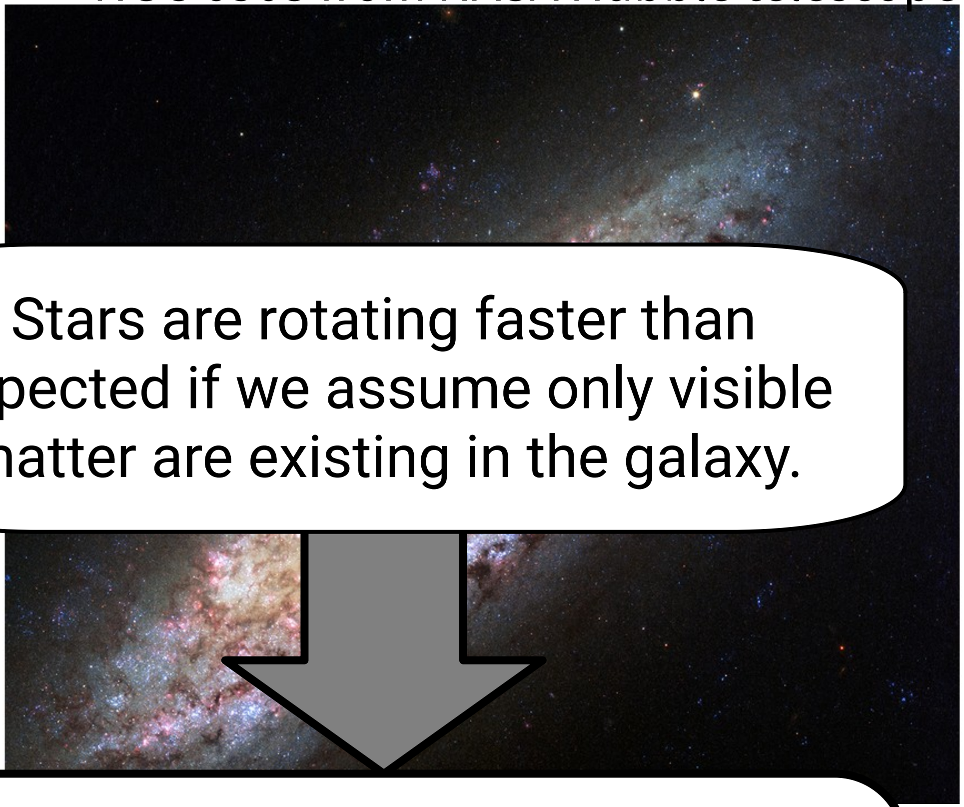
Return to old school example: Galaxy rotation curve and dark matter

(figure from Freese, 0812.4005)



$$v_{\mathrm{circ}}(R) = \sqrt{\frac{GM(R)}{R}}$$

NGC 6503 from NASA Hubble telescope

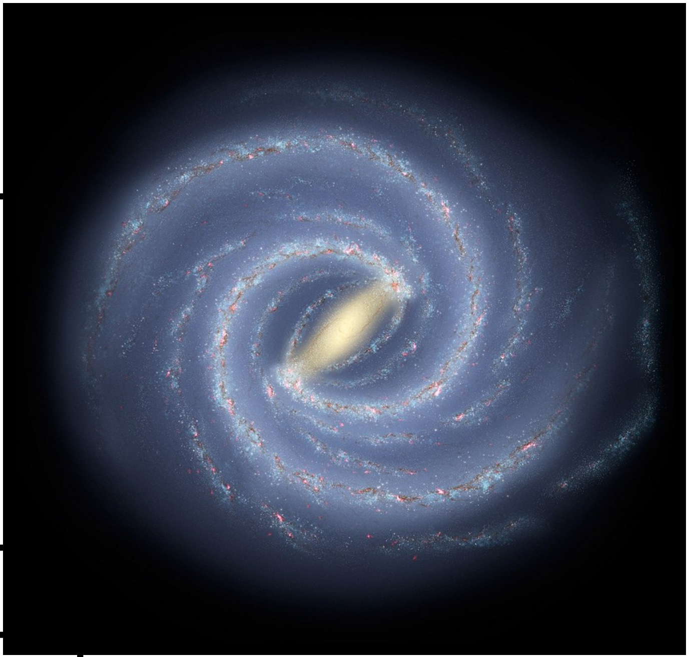


Stars are rotating faster than expected if we assume only visible matter are existing in the galaxy.

DARK MATTER!

General Classic Approach for Dark Matter Density Estimation

Galactic Dynamics



Equation of motion: Boltzmann Equation

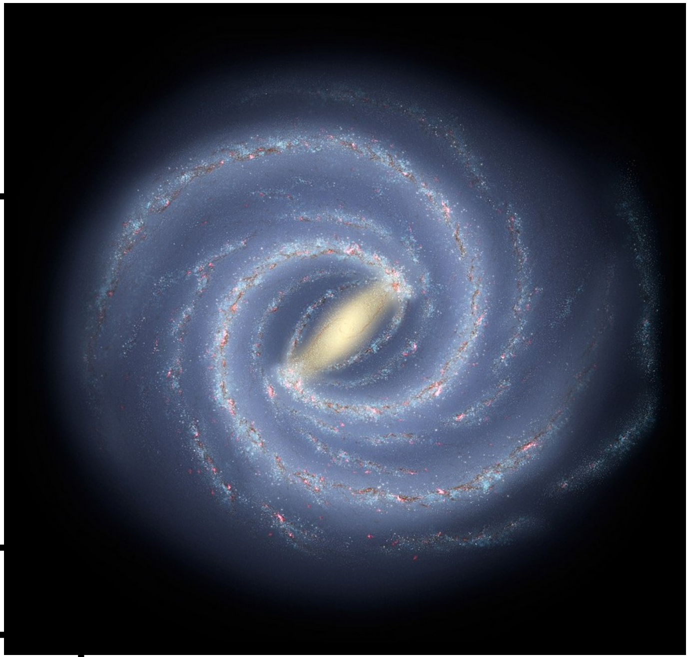
$$\left[\frac{\partial}{\partial t} + \vec{v} \cdot \frac{\partial}{\partial \vec{x}} + \vec{a} \cdot \frac{\partial}{\partial \vec{v}}\right] f(\vec{x}, \vec{v}) = 0$$

Classic inference method:

Introduce **a simple parametric model** of galaxy to fit data and infer dark matter halo shape.

Why do we need to use machine learning and neural networks to analyze these galactic structures?

Galactic Dynamics



Equation of motion: Boltzmann Equation

$$\left[\frac{\partial}{\partial t} + \vec{v}\cdot\frac{\partial}{\partial \vec{x}} + \vec{a}\cdot\frac{\partial}{\partial \vec{v}}\right] f(\vec{x},\vec{v}) = 0$$

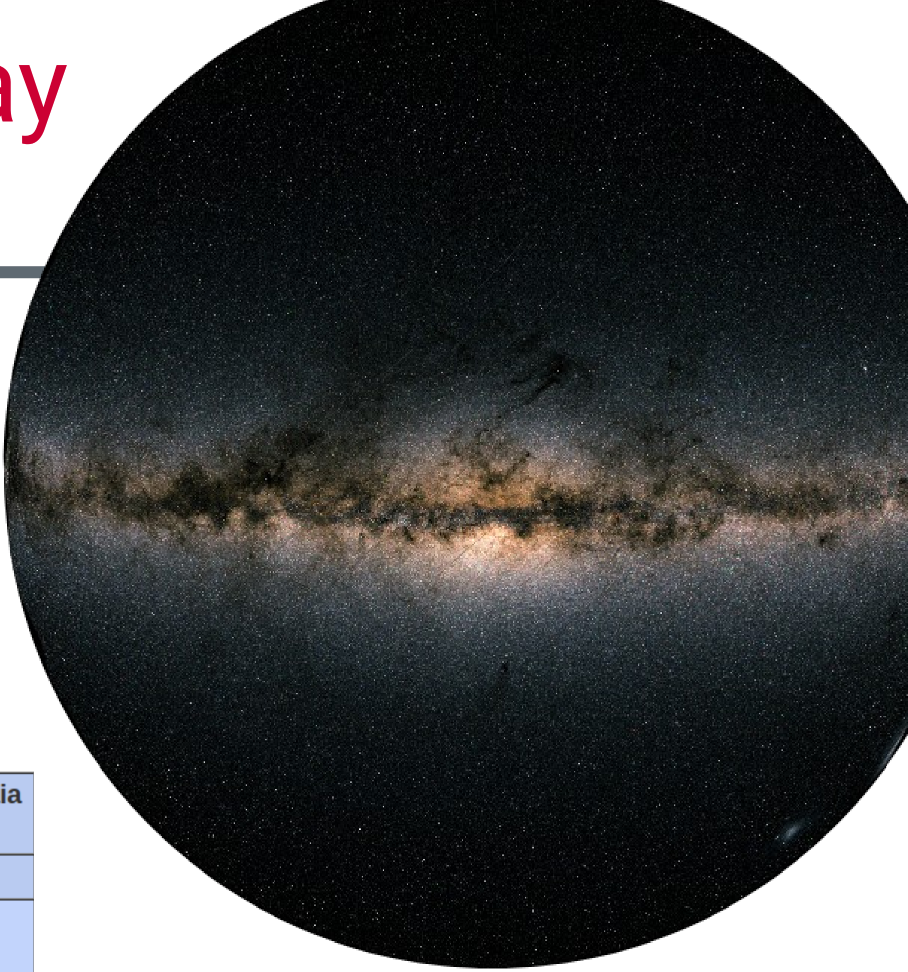
Classic inference method:

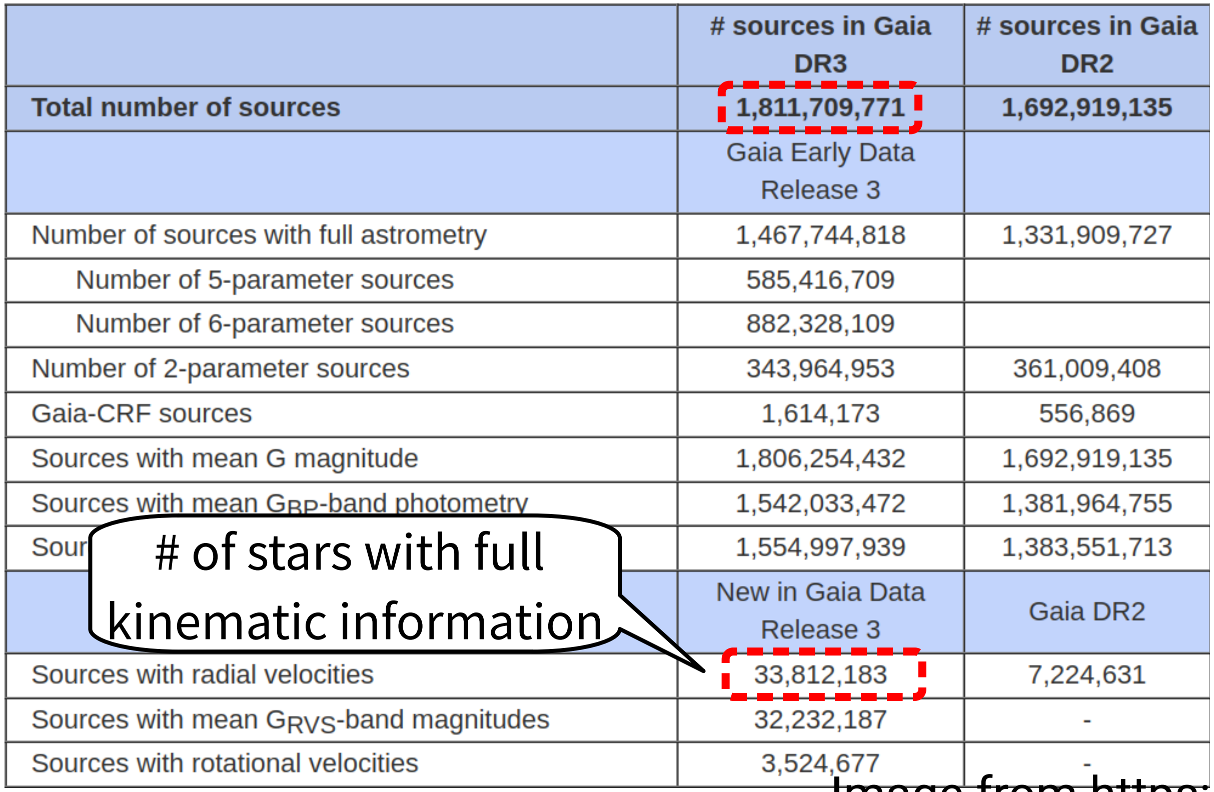
Introduce **a simple parametric model** of galaxy to fit data and infer dark matter halo shape.

- Not a **"local"** dark matter density measurement!
- This strategy **may not be the best strategy** in the current data-driven era of astrophysics!

A Snapshot of Milky Way from Gaia

Recently, Gaia mission from European Space Agency (ESA) released a new catalog containing very detailed measurement of stars in the Milky Way that can be used for various physics analysis.





	# sources in Gaia DR3	# sources in Gaia DR2
Total number of sources	**1,811,709,771**	**1,692,919,135**
	Gaia Early Data Release 3	
Number of sources with full astrometry	1,467,744,818	1,331,909,727
Number of 5-parameter sources	585,416,709	
Number of 6-parameter sources	882,328,109	
Number of 2-parameter sources	343,964,953	361,009,408
Gaia-CRF sources	1,614,173	556,869
Sources with mean G magnitude	1,806,254,432	1,692,919,135
Sources with mean G_{BP}-band photometry	1,542,033,472	1,381,964,755
Sour	1,554,997,939	1,383,551,713
	New in Gaia Data Release 3	Gaia DR2
Sources with radial velocities	33,812,183	7,224,631
Sources with mean G_{RVS}-band magnitudes	32,232,187	-
Sources with rotational velocities	3,524,677	-

# of stars with full kinematic information

We could use this dataset to understand structure of the **Milky Way** very precisely.

Image from https://gea.esac.esa.int/archive/visualization/

Why do we need to use machine learning and neural networks to analyze these galactic structures?

Galactic Dynamics

Neural Network

Equation of motion: Boltzmann Equation

$$\left[\frac{\partial}{\partial t} + \vec{v}\cdot\frac{\partial}{\partial \vec{x}} + \vec{a}\cdot\frac{\partial}{\partial \vec{v}}\right] f(\vec{x},\vec{v}) = 0$$

Machine-learning based inference method:

Neural Networks for
arbitrary **density estimation**:

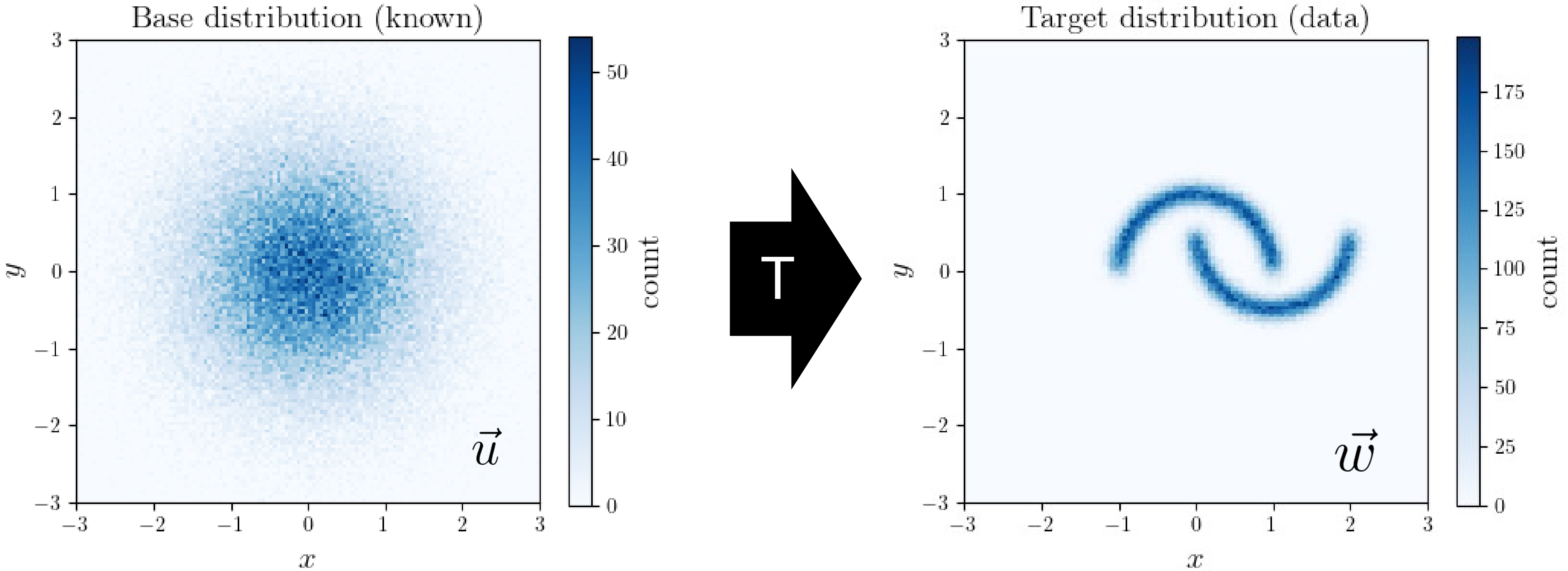
- Normalizing Flows

$$\vec{u}_0 \rightarrow \vec{u}_1 \rightarrow \cdots \rightarrow \vec{u}_n = (\vec{x},\vec{v})$$

This is a fully
model-independent
strategy!

Normalizing Flows: Neural Network learning a Transformation

Normalizing Flows (NFs) is an artificial neural network that learns a transformation of random variables.



Main idea: if we could find out such transformation, we can use the transformation formula for the density estimation:

$$p_W(\vec{w}) = p_U(\vec{u}) \cdot \left| \frac{d\vec{u}}{d\vec{w}} \right|$$

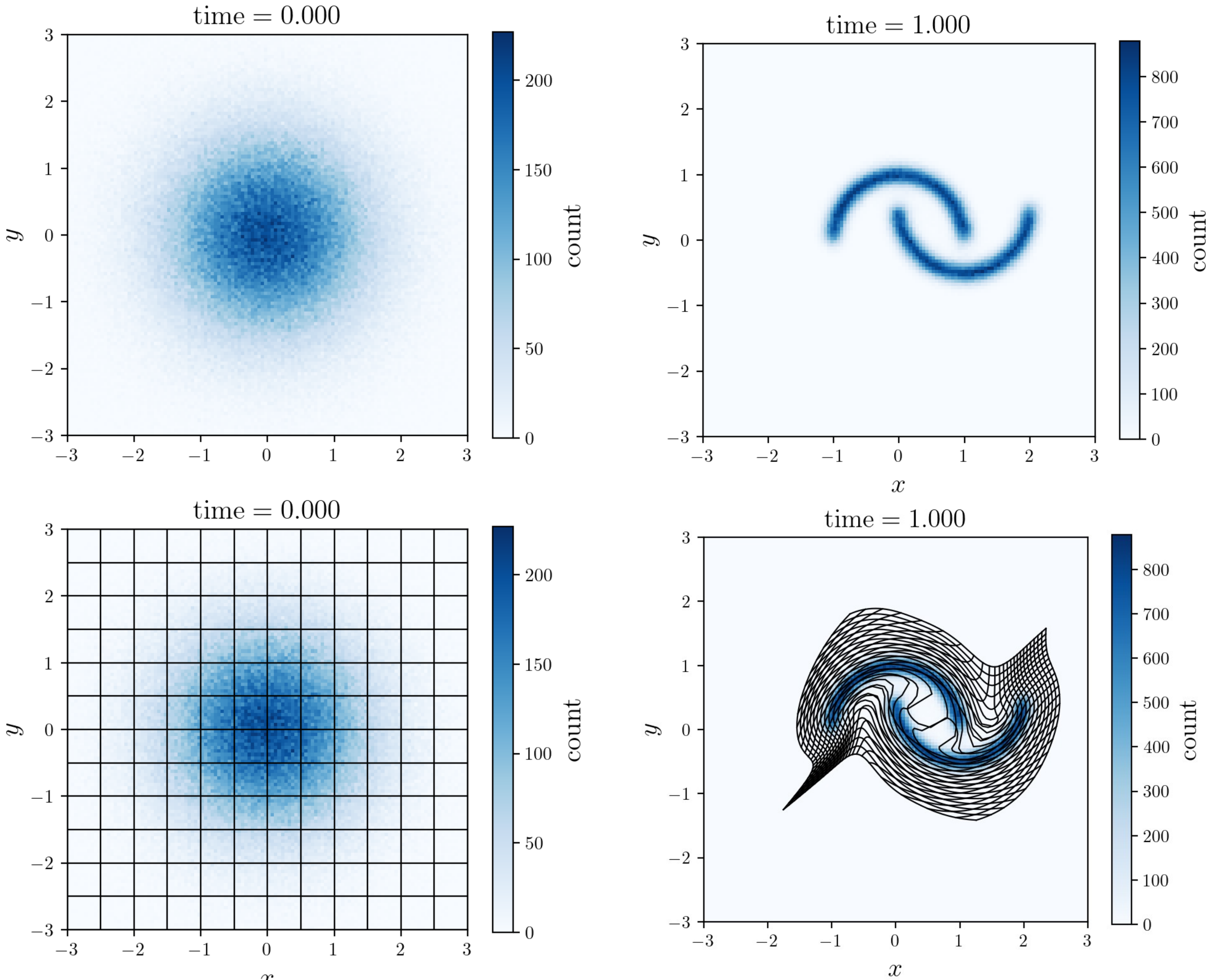
Neural Network

$$\vec{w} = T(\vec{u})$$

This formula can be used for training normalizing flows, too:

Maximum likelihood estimation

Learned transformation

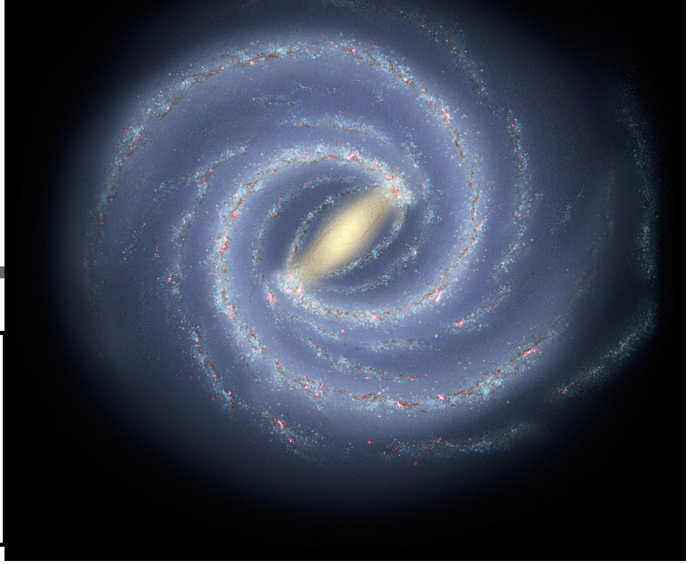


Outline of Strategy

Star catalog

$\{(\vec{x}, \vec{v})\}$

Gaia DR3: star catalog of the Milky Way



Phase space density

$f(\vec{x}, \vec{v})$

Neural Networks for Density Estimation:
Normalizing Flows

$$\vec{u}_0 \to \vec{u}_1 \to \cdots \to \vec{u}_n = (\vec{x}, \vec{v})$$

Gravitational accel.

$\vec{a}(\vec{x})$

Solving EOM (Boltzmann Equation)

$$\left[\frac{\partial}{\partial t} + \vec{v} \cdot \frac{\partial}{\partial \vec{x}} + \vec{a} \cdot \frac{\partial}{\partial \vec{v}}\right] f(\vec{x}, \vec{v}) = 0$$

Mass density

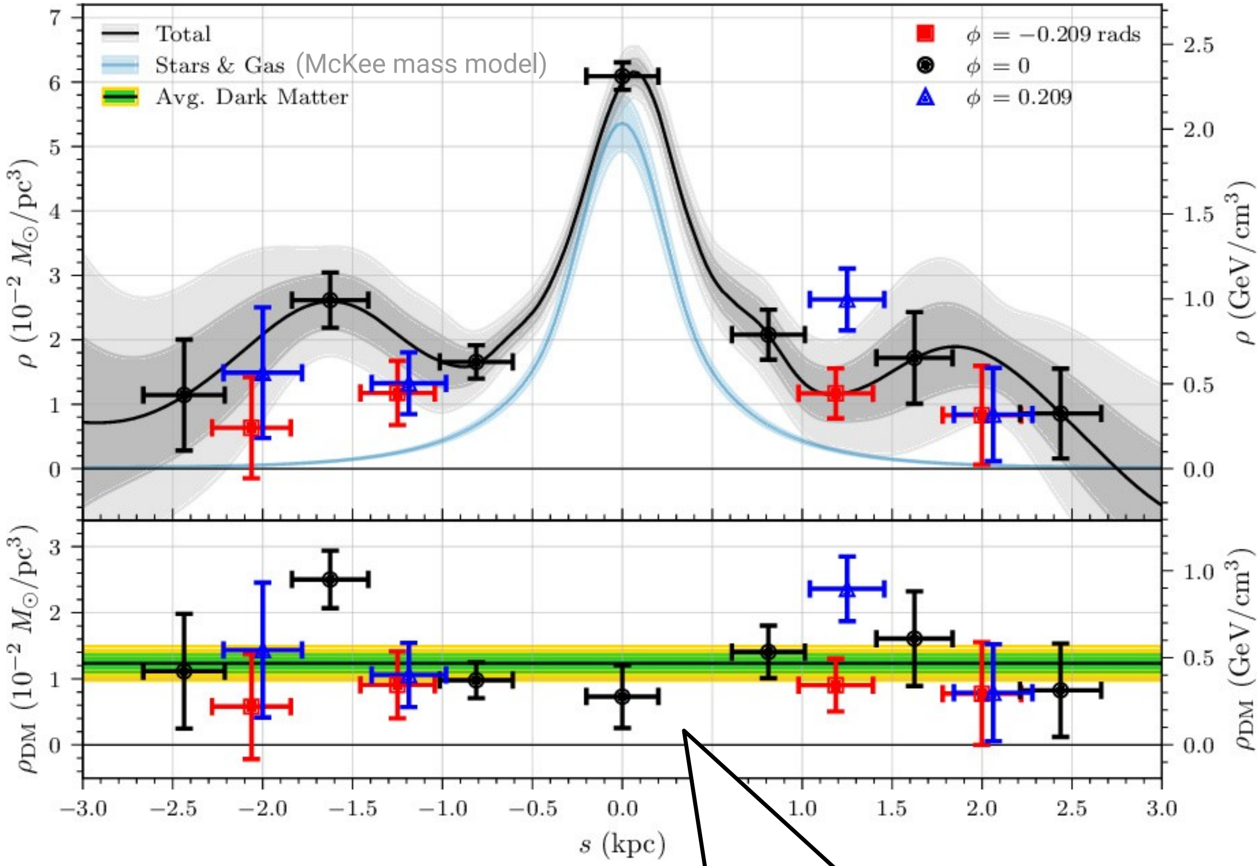
$\rho(\vec{x})$

Solving Gauss's Equation

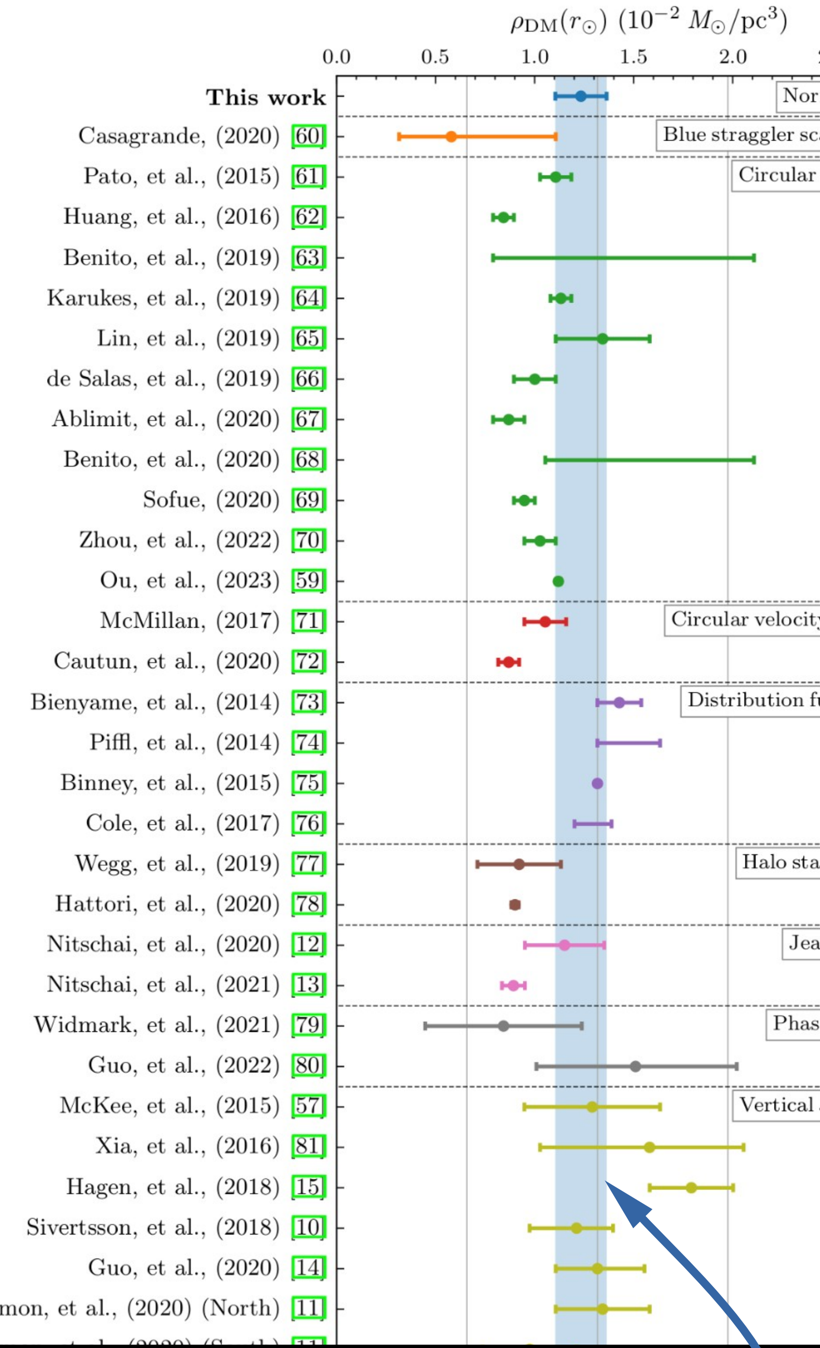
$$-4\pi G\rho = \nabla \cdot \vec{a}$$

arXiv: 2205.01129, 2305.13358
See also Green et. al. arXiv:2011.04673, arXiv:2205.02244

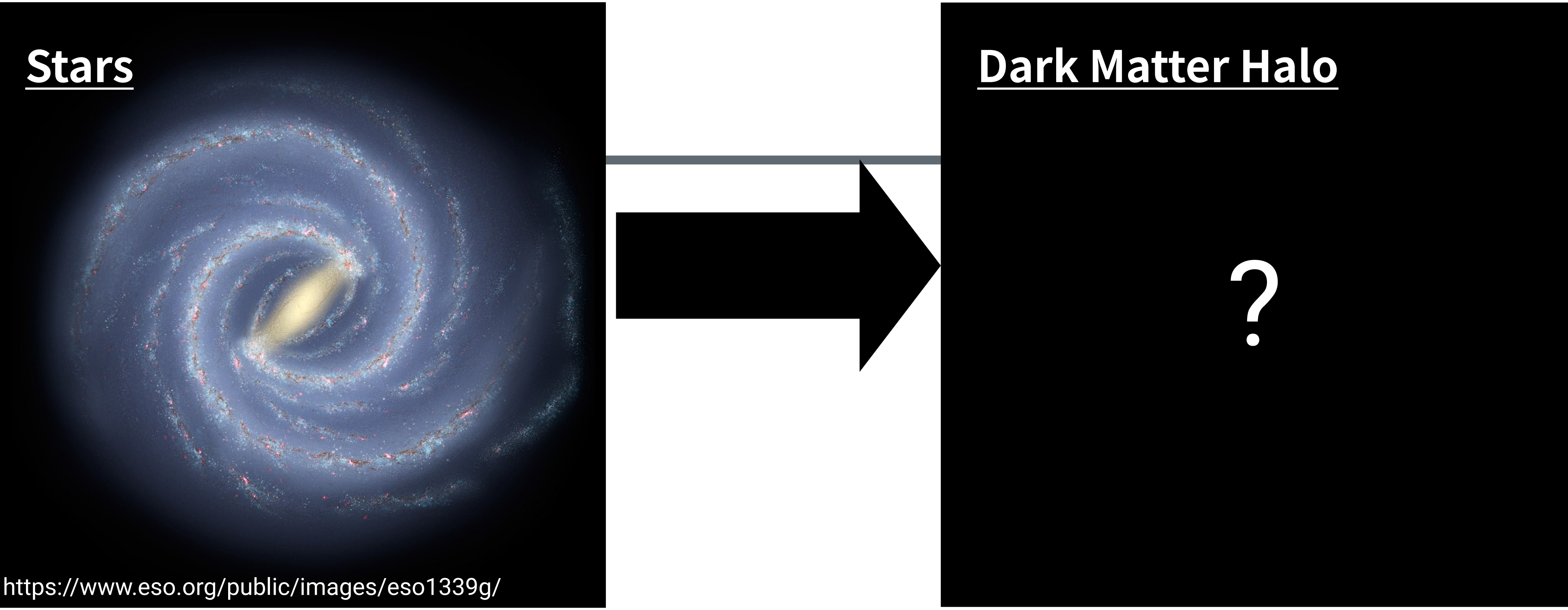
Local DM Mass Density of the Milky Way



0.32±0.18 GeV/cm³



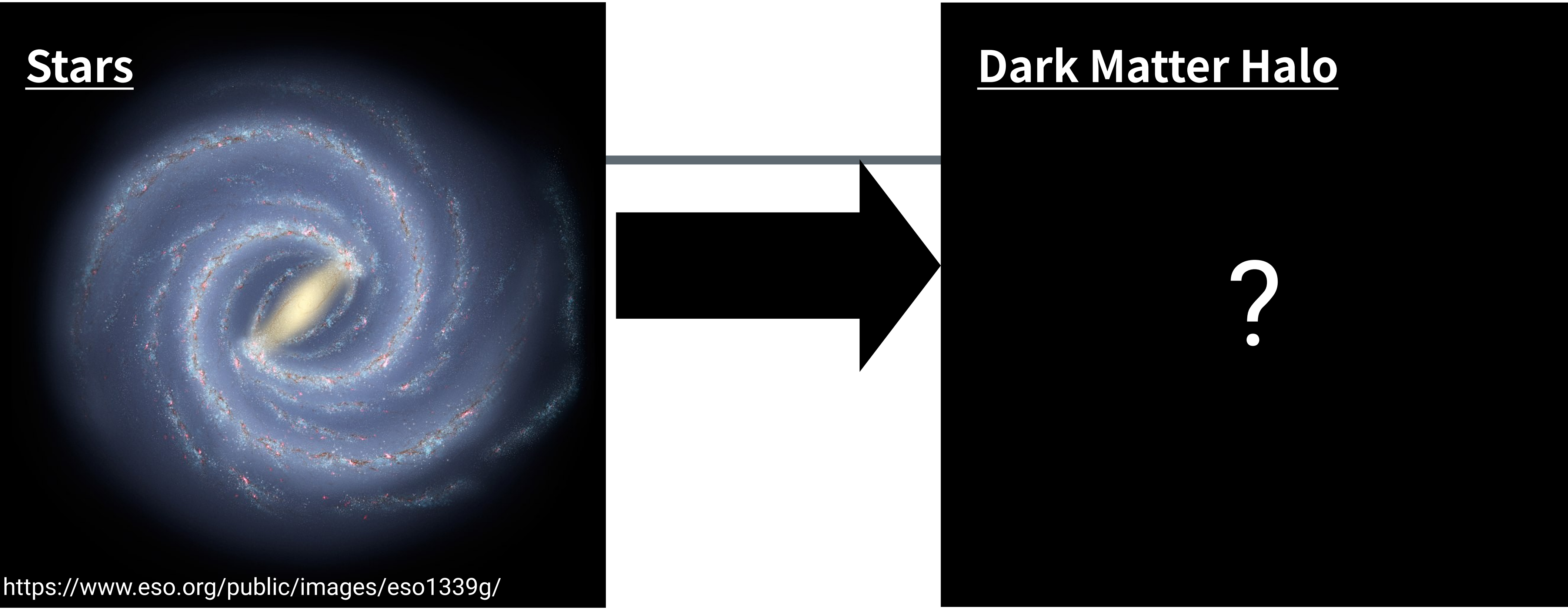
Taking the average of the DM mass density at the Solar radius, we find a local dark matter density: **0.47±0.05 GeV/cm³**



https://www.eso.org/public/images/eso1339g/

We have an unsupervised ML method to estimate **dark matter density** given **stellar distribution of a galaxy**.

THE END?

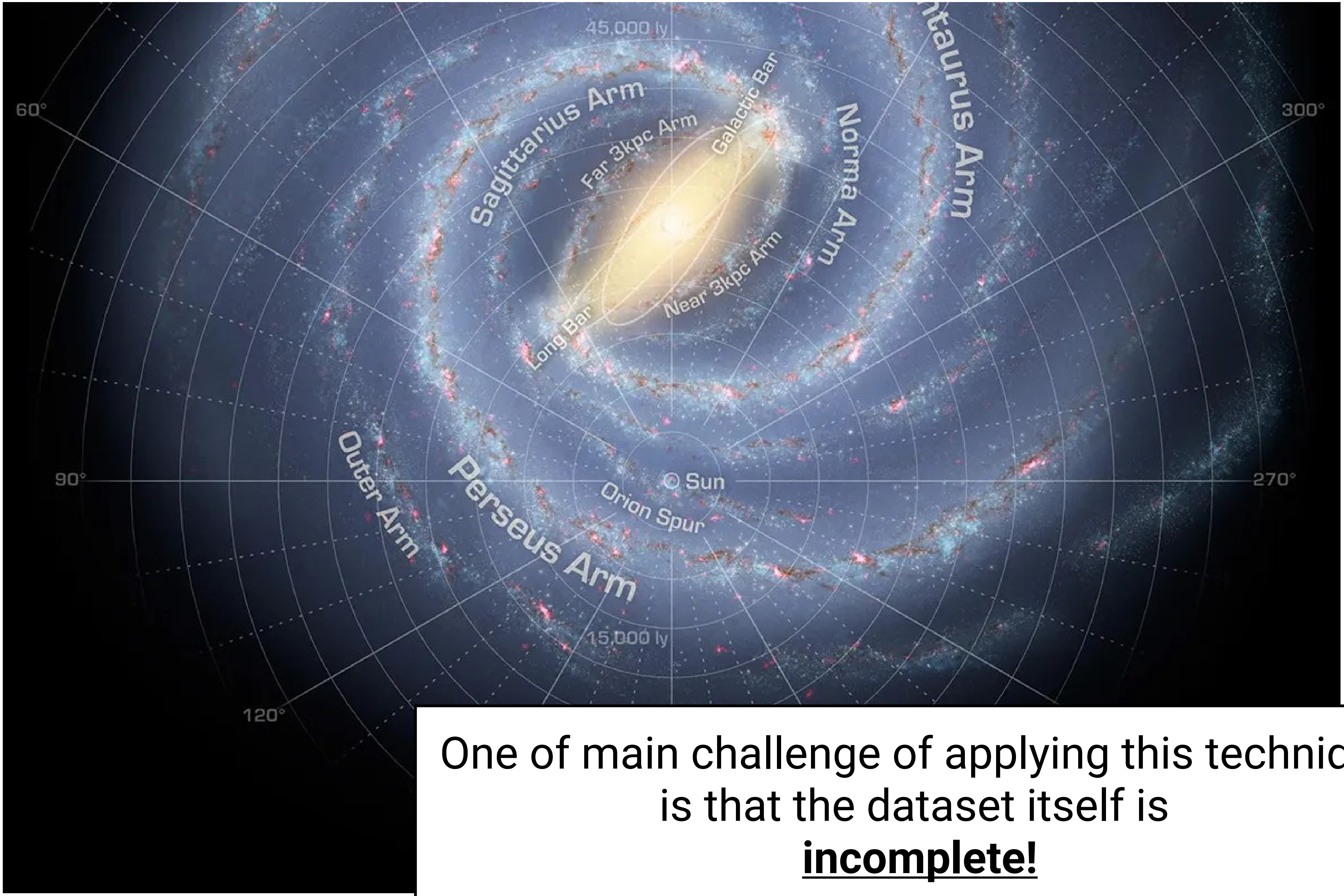


https://www.eso.org/public/images/eso1339g/

We have an unsupervised ML method to estimate **dark matter density** given **stellar distribution of a galaxy**.

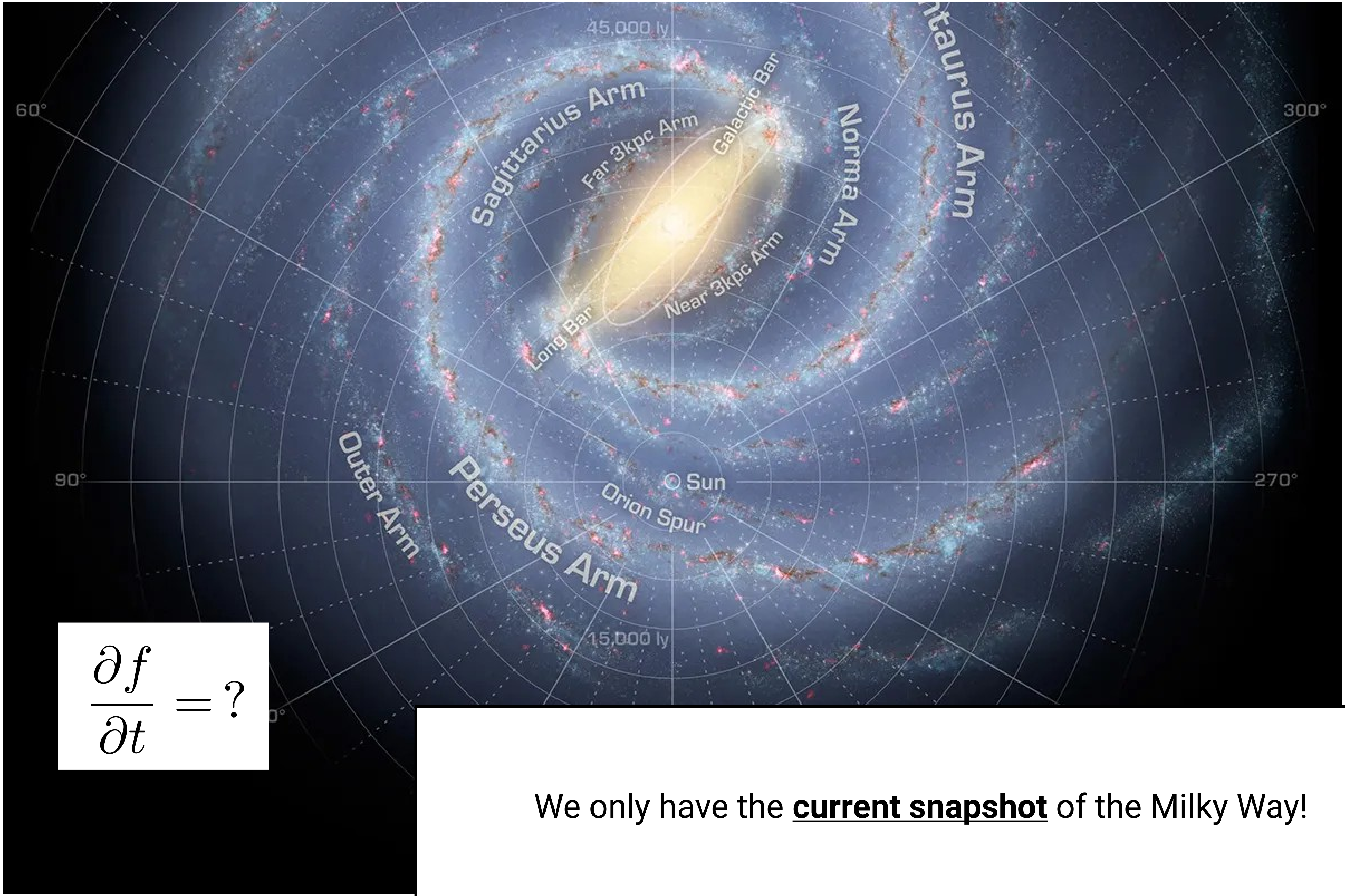
THE END? → Of course not!

Galactic Dynamics and Incomplete Datasets



One of main challenge of applying this technique is that the dataset itself is **<u>incomplete!</u>**

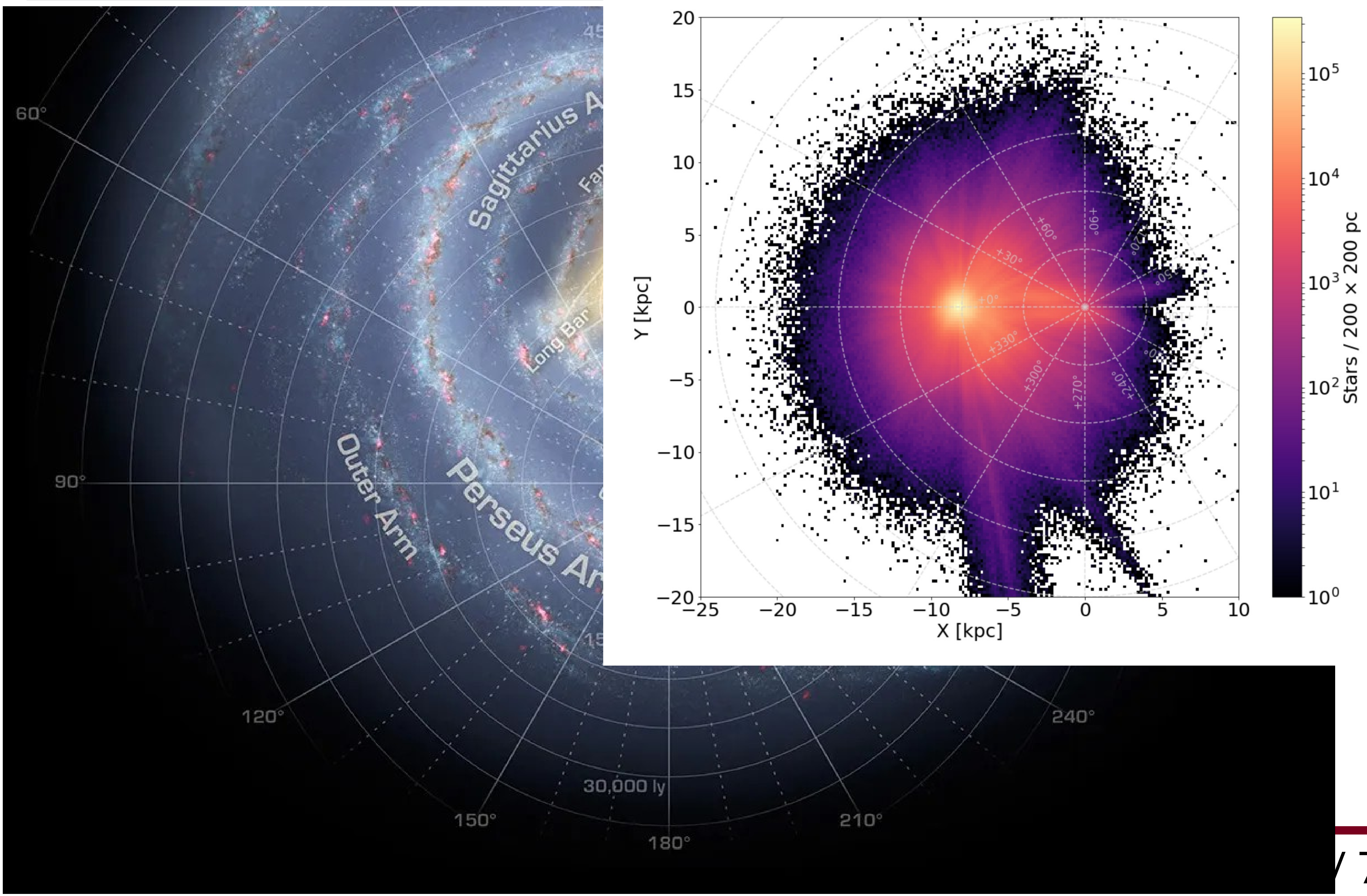
No time derivative information



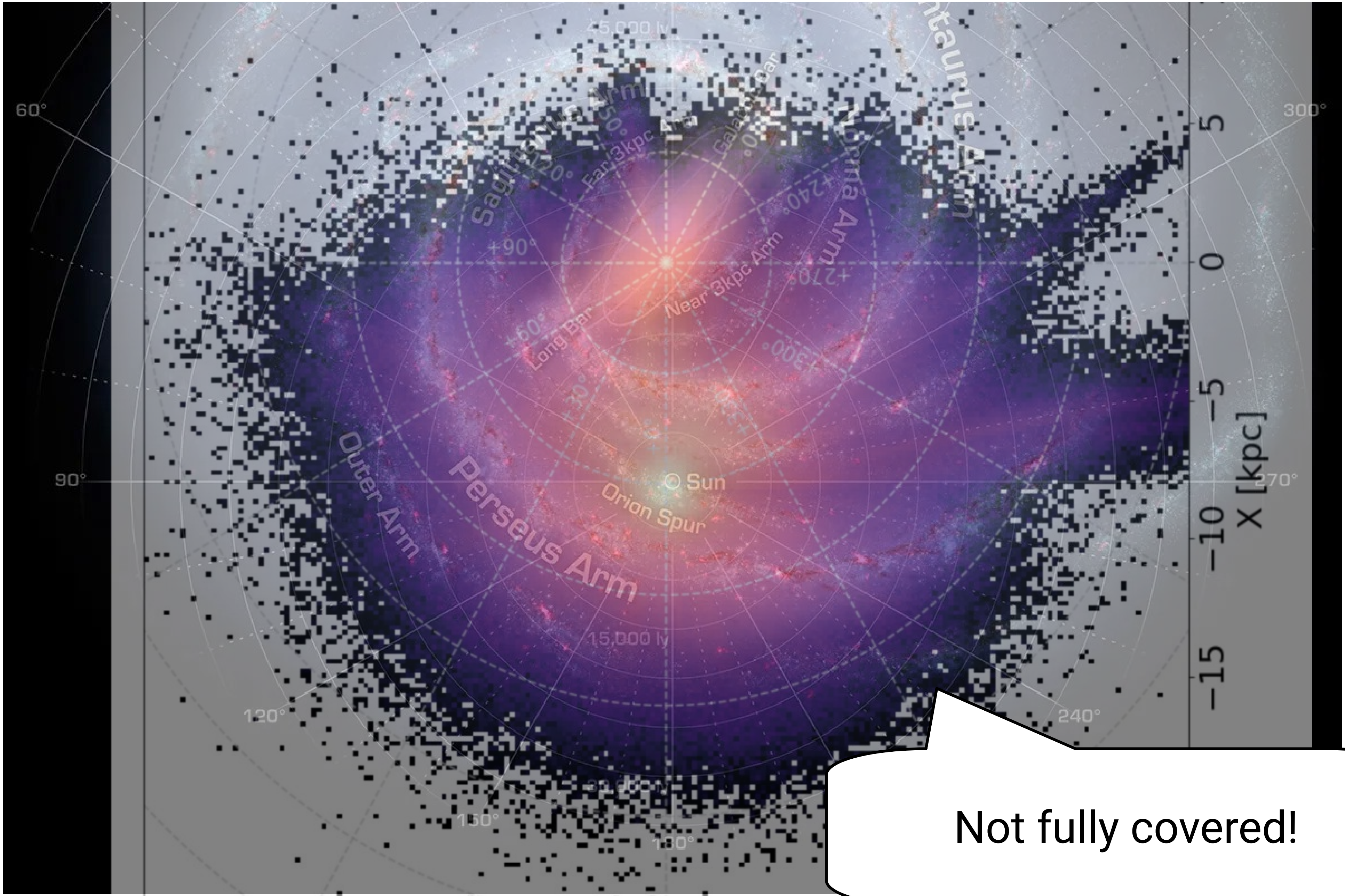
$$\frac{\partial f}{\partial t} = ?$$

We only have the **current snapshot** of the Milky Way!

Radial Velocity Distribution of Gaia DR3

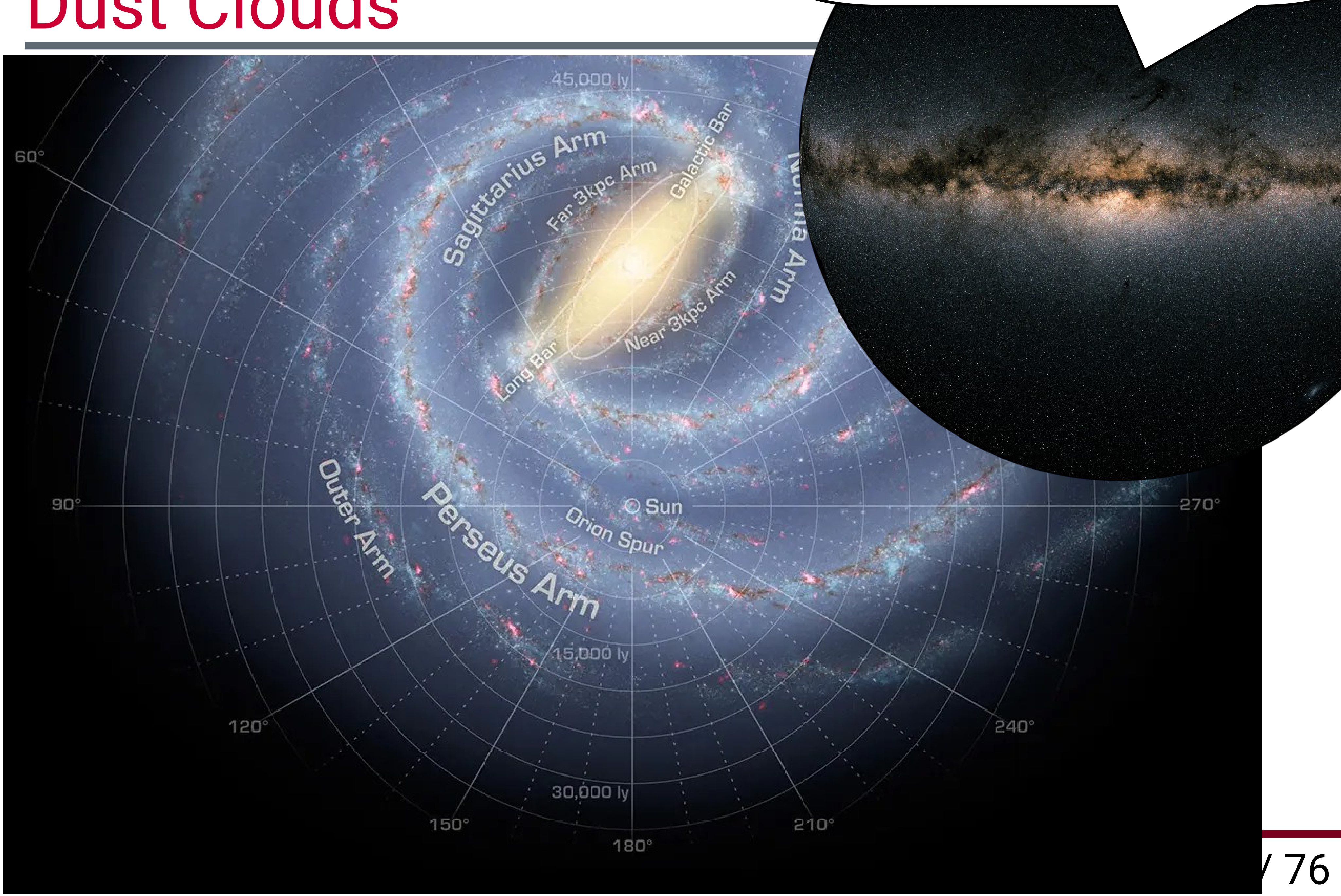


Incompleteness in Spacial Coverage



Not fully covered!

Dust Clouds



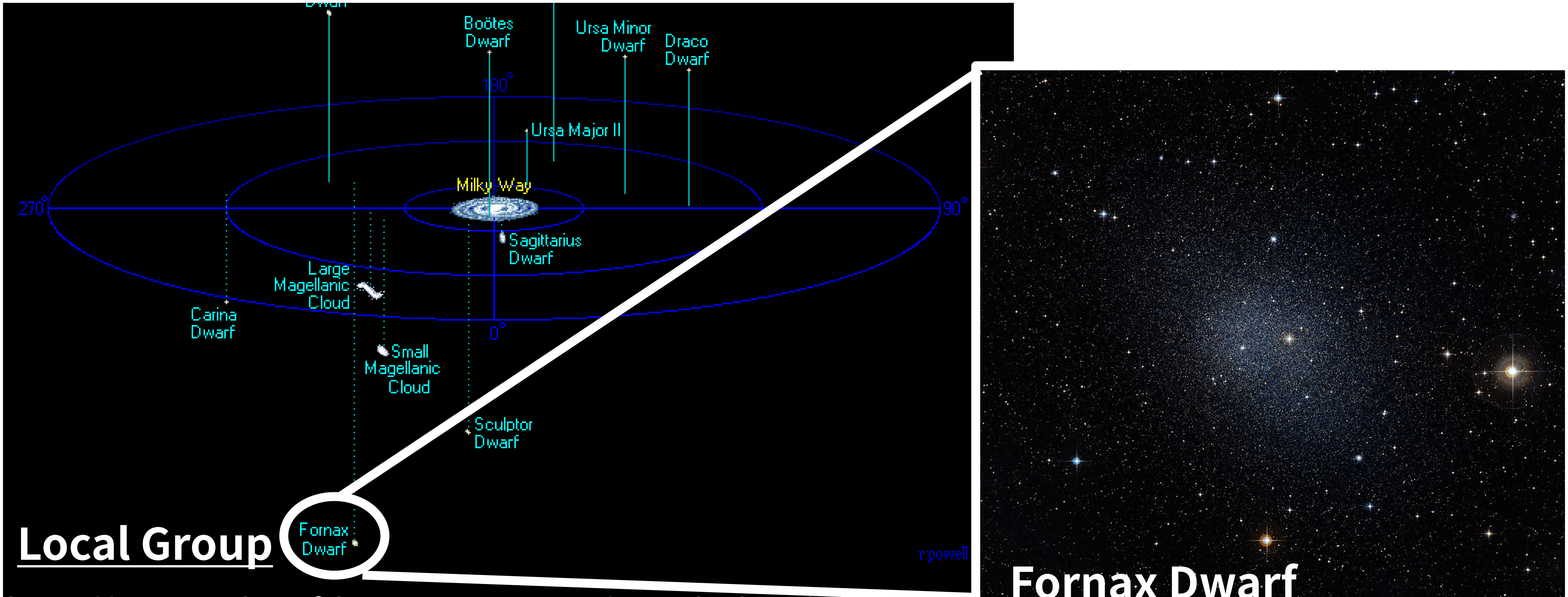
Intergalactic dust cloud obscuring light from stars!

Dust Obscuring Stars

Intergalactic dust cloud
obscuring light from stars!

Distant Galaxy – Only partial information is available

- Available kinematic information is **limited**!
  - Position of stars on the sky (x, y) (phot.)
  - ~~Distance to the stars (z)~~
  - ~~Proper motion of stars on the sky (v_x, v_y)~~
  - Radial velocity (v_z) (spec.)



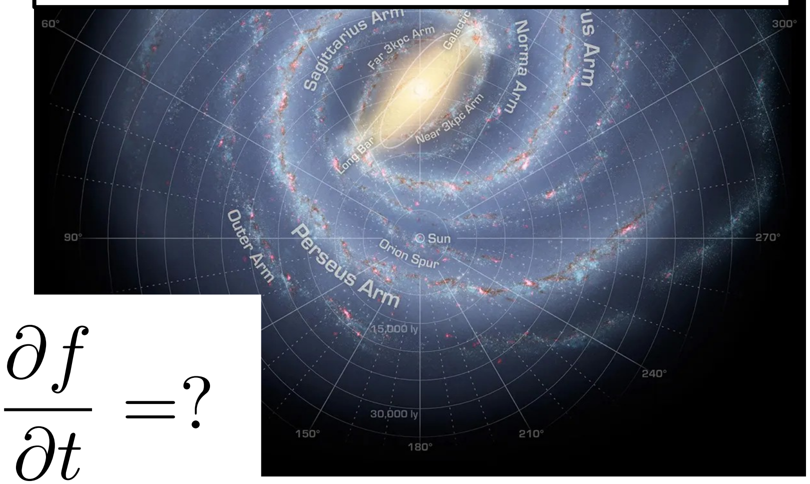
Local Group

Fornax Dwarf

http://www.atlasoftheuniverse.com/sattelit.html

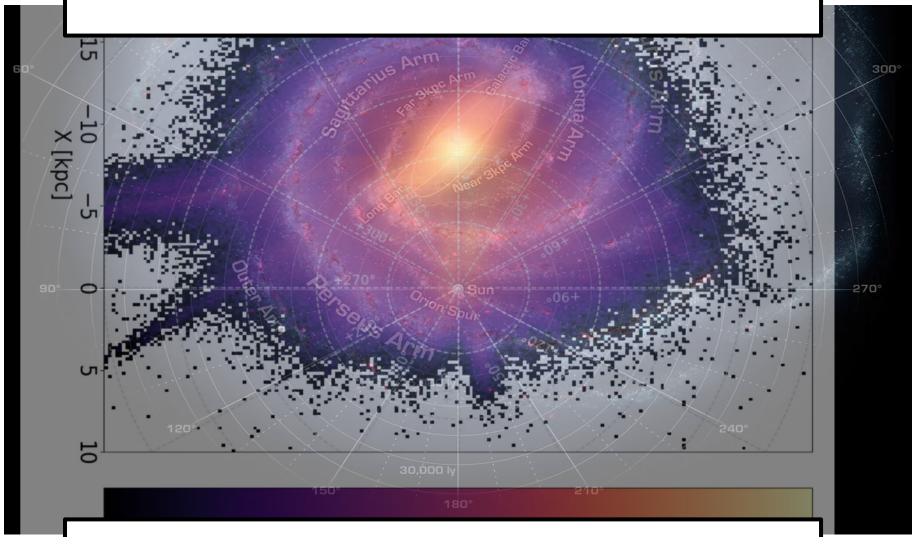
Various challenging incompleteness!

Disequilibrium

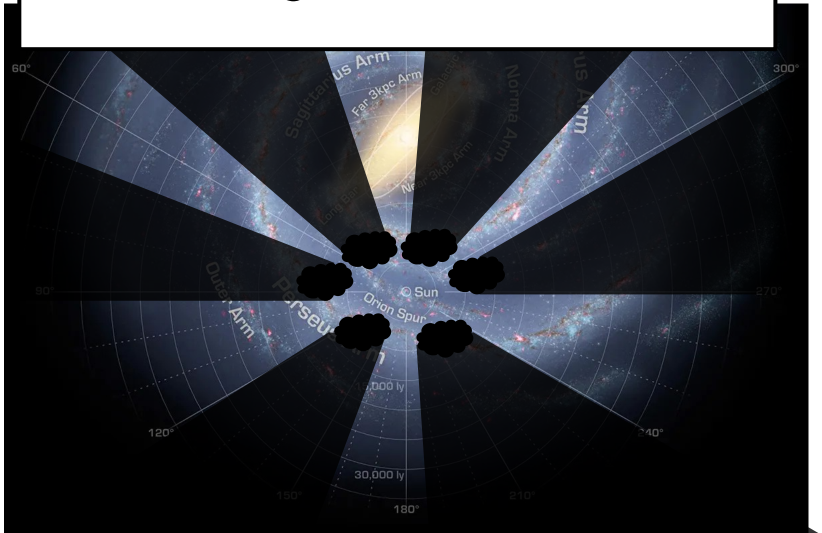


$$\frac{\partial f}{\partial t} =?$$

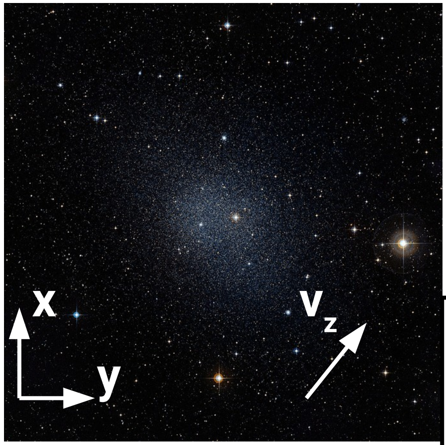
Spatial Incompleteness



Intergalactic Dust



Lack of information



Only 3D info. available, not the full 6D PS info.

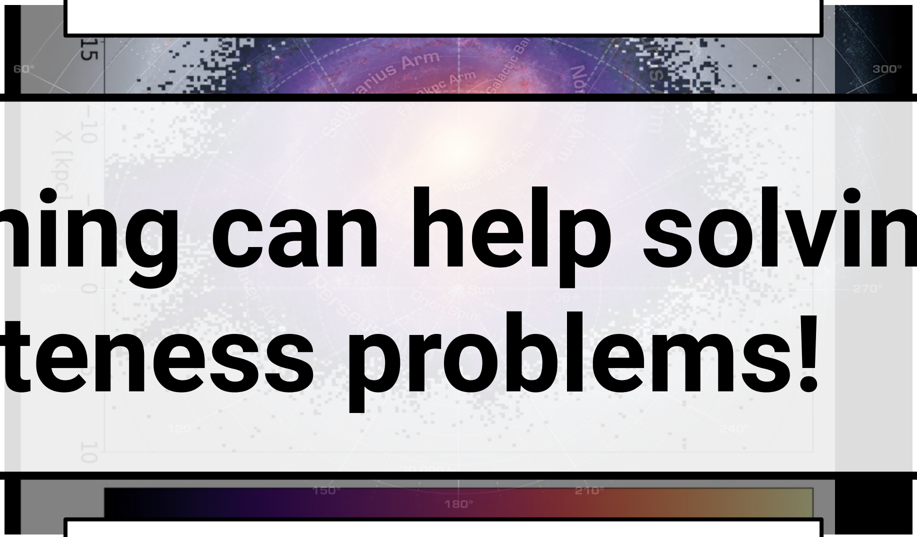
More challenges are waiting!

Various challenging incompleteness!

Disequilibrium

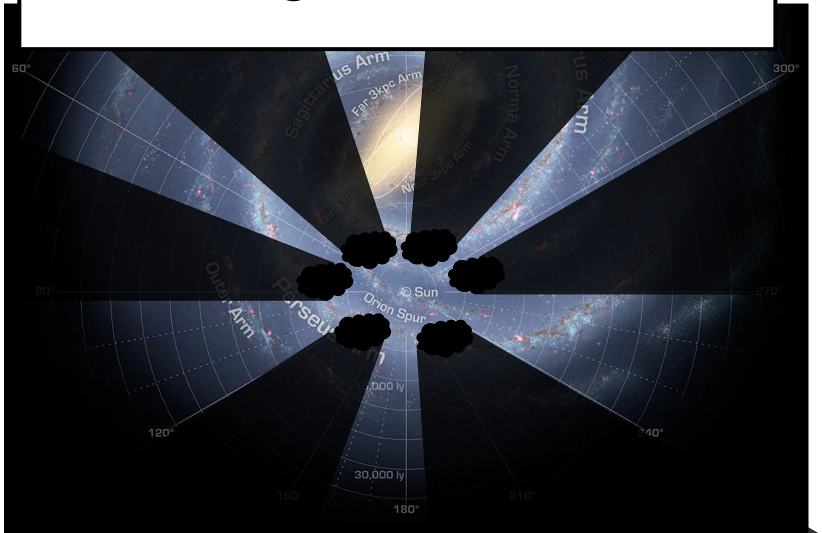
$\frac{\partial f}{\partial t} = ?$

Spatial Incompleteness

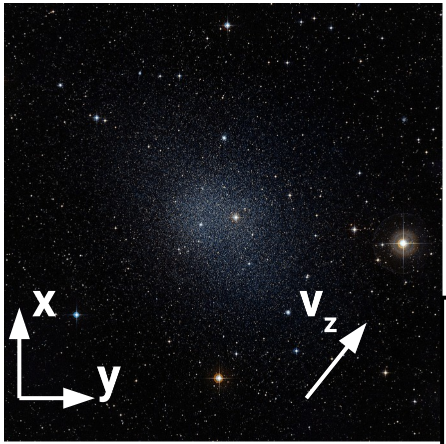


Again, machine learning can help solving these data incompleteness problems!

Intergalactic Dust



Lack of information



Only 3D info. available, not the full 6D PS info.

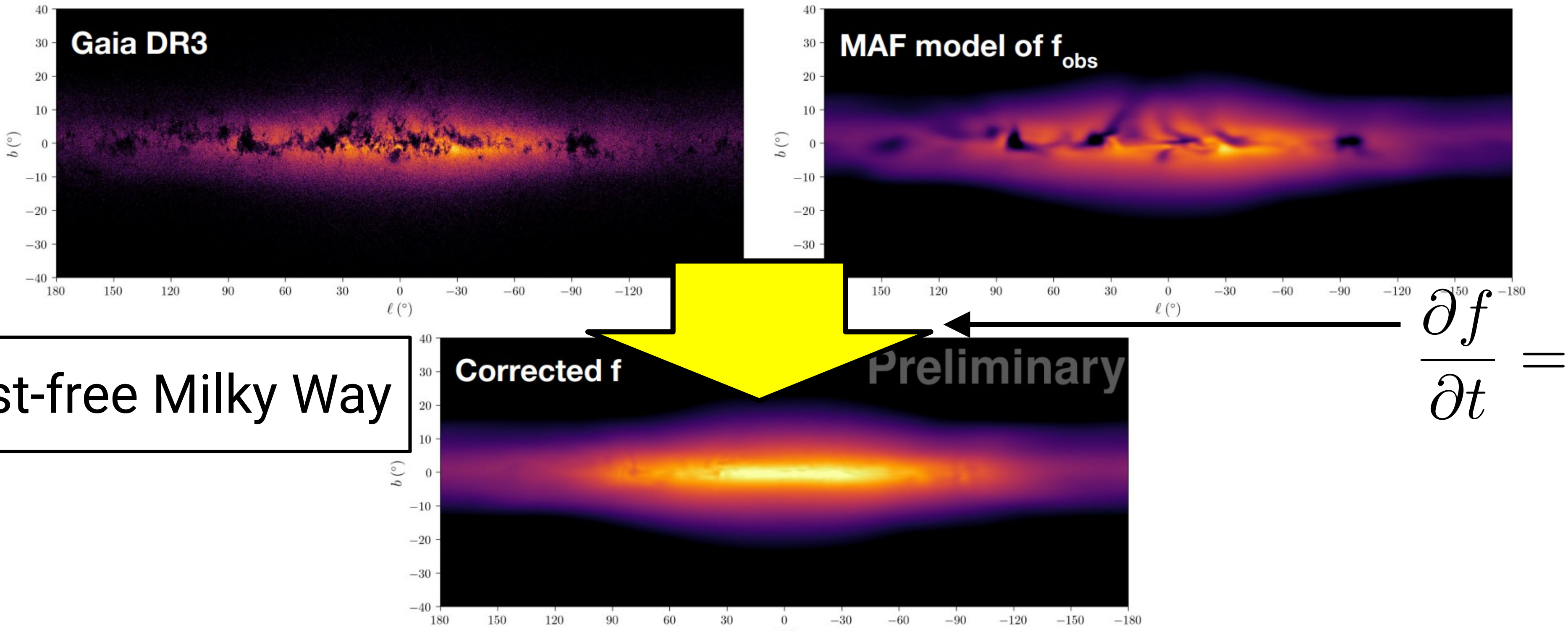
More challenges are waiting!

Erasing Dust using Neural Network and Equilibrium Assmptions

Dusty Milky Way

E. Putney, D. Shih, **SHL**, and M. R. Buckley, arXiv:2412.14236

Unbiased Phase Space Density



Dust-free Milky Way

$$\frac{\partial f}{\partial t} = 0$$

Equilibrium assumption can be used to interpolate dust-obscured regions!

Dust Correction

Boltzmann equation also provides us an alternative way of measuring intergalactic dust clouds.

Observed density
modeled by
Normalizing flows

Selection Efficiency
of measuring
obscured stars
due to **dust clouds**

$$f_{\mathrm{obs}}(\vec{x},\vec{v}) = f_{\mathrm{true}}(\vec{x},\vec{v}) \times \epsilon(\vec{x})$$

**True phase-space
density of stars**

$$\left[\vec{v}\cdot\frac{\partial}{\partial\vec{x}} + \vec{a}\cdot\frac{\partial}{\partial\vec{v}}\right] f_{\mathrm{true}}(\vec{x},\vec{v}) = 0, \quad \vec{a} = -\frac{d\Phi(\vec{x})}{d\vec{x}}$$

Dust Correction

Boltzmann equation also provides us an alternative way of measuring intergalactic dust clouds.

Observed density modeled by Normalizing flows

Selection Efficiency of measuring obscured stars due to **dust clouds**

$$f_{\text{obs}}(\vec{x}, \vec{v}) = f_{\text{true}}(\vec{x}, \vec{v}) \times \epsilon(\vec{x})$$

True phase-space density of stars

Linear equation in **acceleration** and **efficiency**

$$(\vec{a}(x), \nabla \log \epsilon(x))$$

$$= (-\nabla\phi(x), \nabla \log \epsilon(x))$$

↓

Solvable by MSE minimization!

$$\left[\vec{v} \cdot \frac{\partial}{\partial \vec{x}} + \color{red}{\vec{a}} \cdot \frac{\partial}{\partial \vec{v}}\right] \log f_{\text{obs}}(\vec{x}, \vec{v}) - \vec{v} \cdot \color{red}{\frac{\partial}{\partial \vec{x}} \log \epsilon(\vec{x})} = 0$$

Dust Correction

Boltzmann equation also provides us an alternative way of measuring intergalactic dust clouds.

Observed density modeled by Normalizing flows

Selection Efficiency of measuring obscured stars due to **dust clouds**

$$f_{\text{obs}}(\vec{x}, \vec{v}) = f_{\text{true}}(\vec{x}, \vec{v}) \times \epsilon(\vec{x})$$

True phase-space density of stars

Linear equation in **acceleration** and **efficiency**

$$(\vec{a}(x), \nabla \log \epsilon(x))$$

$$= (-\nabla\phi(x), \nabla \log \epsilon(x))$$

$$\downarrow$$

Solvable by MSE minimization!

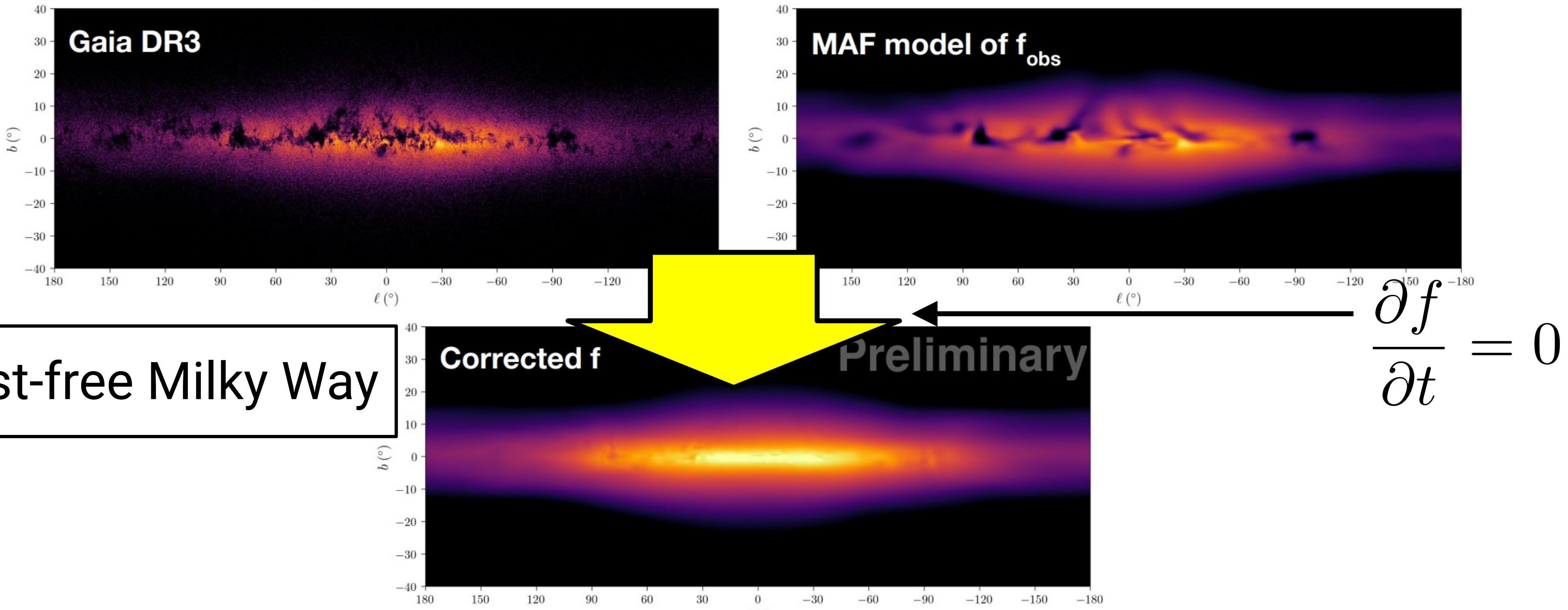
$$\mathcal{L}(\phi, \log \epsilon) = \mathrm{E}_{\vec{x},\vec{v}} \left| \left[\vec{v} \cdot \frac{\partial}{\partial \vec{x}} + \vec{a} \cdot \frac{\partial}{\partial \vec{v}} \right] \log f_{\text{obs}}(\vec{x}, \vec{v}) - \vec{v} \cdot \frac{\partial}{\partial \vec{x}} \log \epsilon(\vec{x}) \right|^2$$

Erasing Dust using Neural Network and Equilibrium Assmptions

Dusty Milky Way

E. Putney, D. Shih, **SHL**, and M. R. Buckley, arXiv:2412.14236

Unbiased Phase Space Density

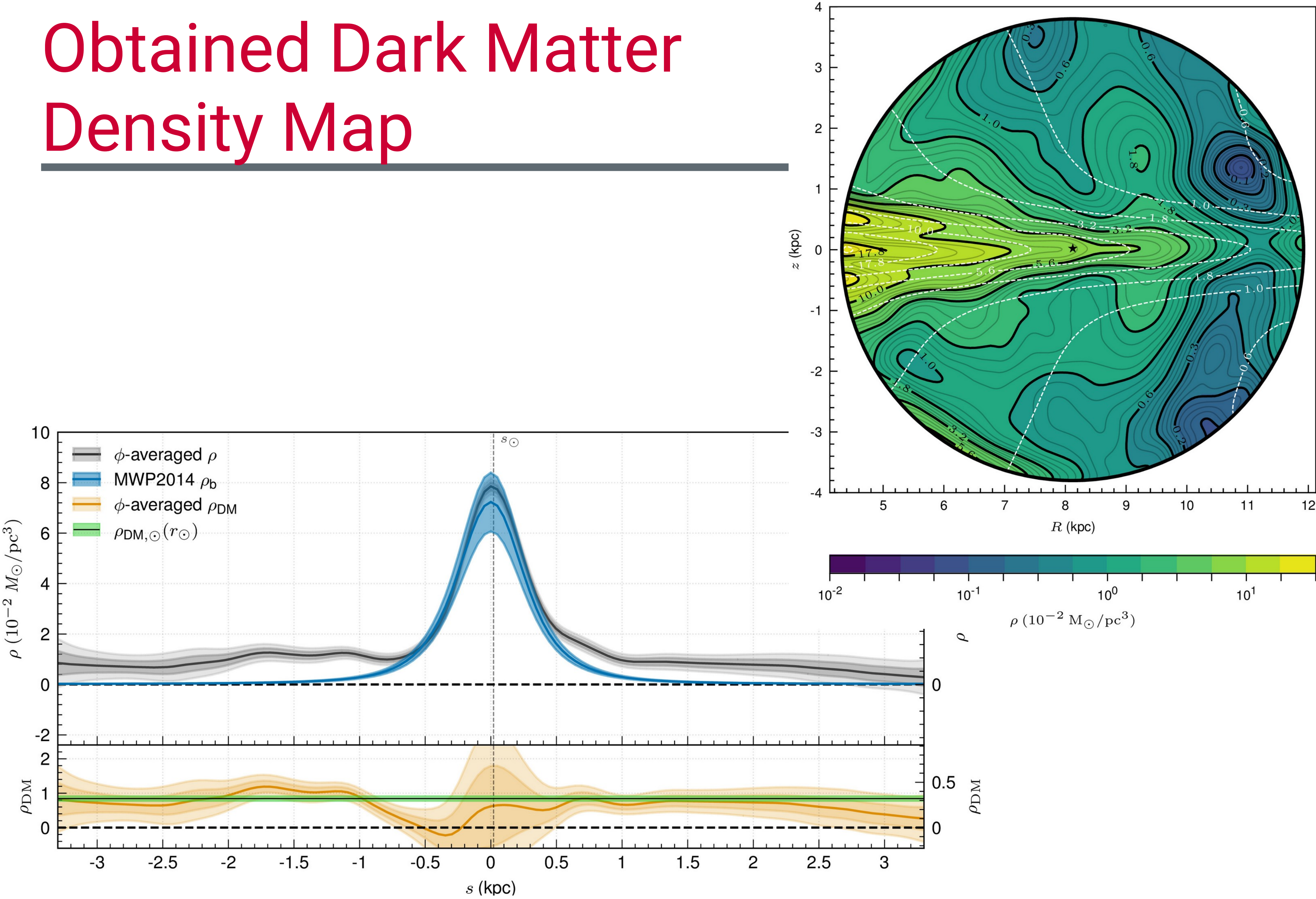


$$\frac{\partial f}{\partial t} = 0$$

Dust-free Milky Way

We could measure dark matter density using dust-obscured populations (disk stars)!

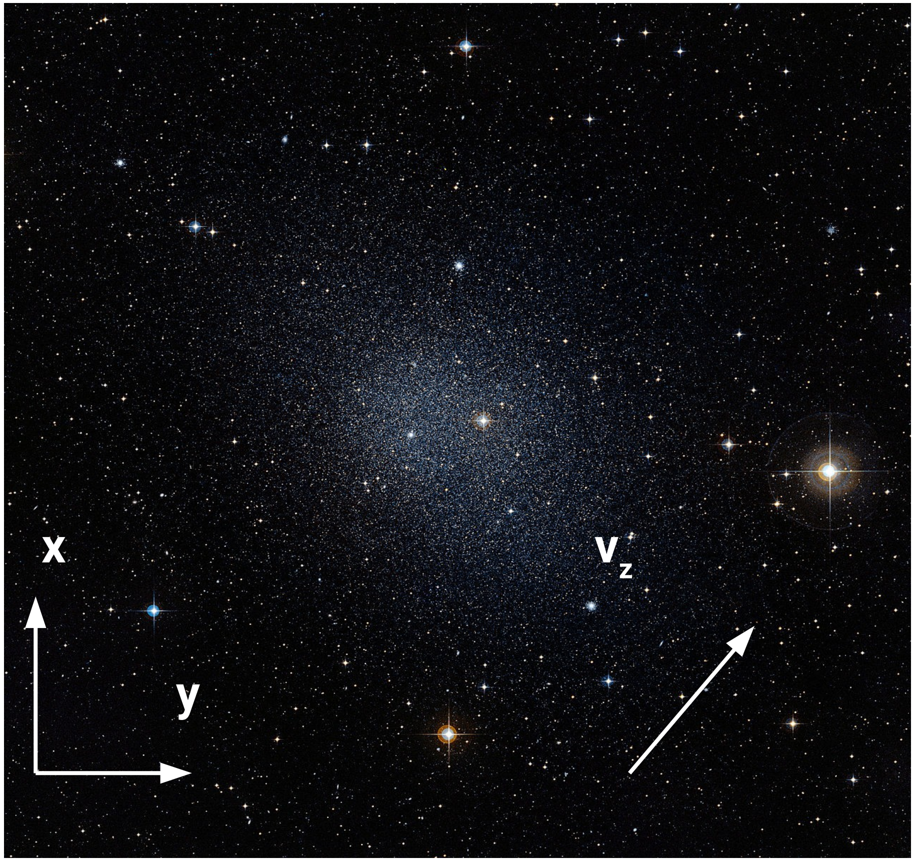
Obtained Dark Matter Density Map



E. Putney, D. Shih, **SHL**, and M. R. Buckley, arXiv:2512.09989
workshop paper in NeurIPS 2025 ML4PS workshop

How about other incompleteness in data?

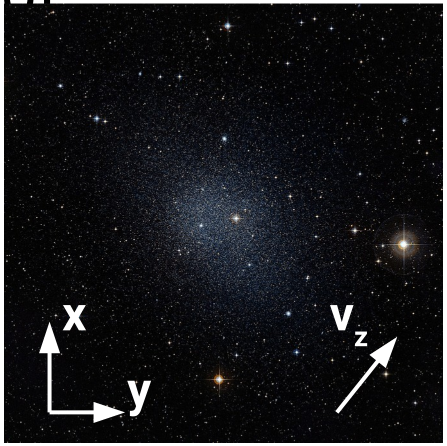
Lack of information



Only 3D info. available, not the full 6D PS info.

Challenges in Analyzing dSphs - Lack of information: how to recover?

- Faint galaxy
  → less number of observed stars O[100] ~ O[1000]
- Available kinematic information is **limited**!
  - Position of stars on the sky (x, y) (phot.)
  - ~~Distance to the stars (z)~~
  - ~~Proper motion of stars on the sky (v_x, v_y)~~
  - Radial velocity (v_z) (spec.)



- Phase space density of stars are not accessible, and hence we cannot solve the equation of motion yet.. (Jeans equation)

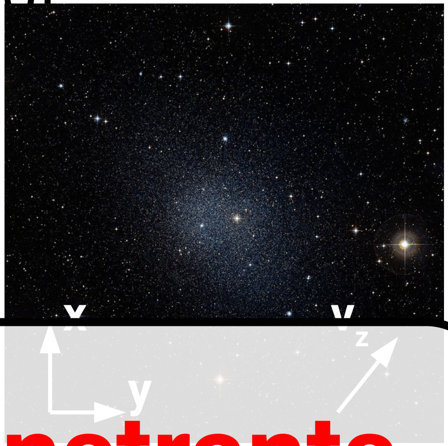
$$\frac{\partial n\langle v_j\rangle}{\partial t} + n\frac{\partial \Phi}{\partial x_j} + n\frac{\partial n\langle v_i v_j\rangle}{\partial x_i} = 0$$

Can we recover the full 6D information somehow?

Classical solution: assume **spherical symmetry**

Challenges in Analyzing dSphs - Lack of information: how to recover?

- Faint galaxy
  → less number of observed stars O[100] ~ O[1000]
- Available kinematic information is **limited**!
  - Position of stars on the sky (x, y) (phot.)
  - ~~Distance to the stars (z)~~
  - ~~Proper motion of stars on the sky (v_x, v_y)~~
  - Radial velocity (v_z)   (spec.)



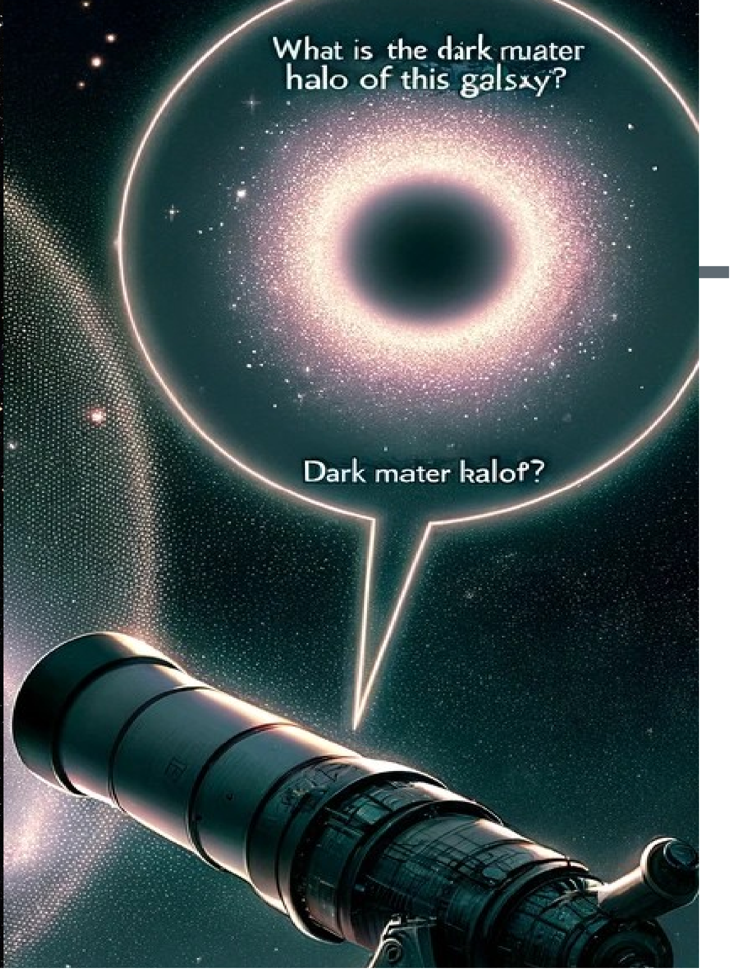
- Phase space density of stars are not accessible, and hence we cannot solve the equation of motion yet.. (Jeans equation)

We can impose such physical constrants to neural networks!

$$\frac{\partial n\langle v_j\rangle}{\partial t} + n\frac{\partial \Phi}{\partial x_j} + n\frac{\partial n\langle v_i v_j\rangle}{\partial x_i} = 0$$

Can we recover the full 6D information somehow?

Classical solution: assume **spherical symmetry**

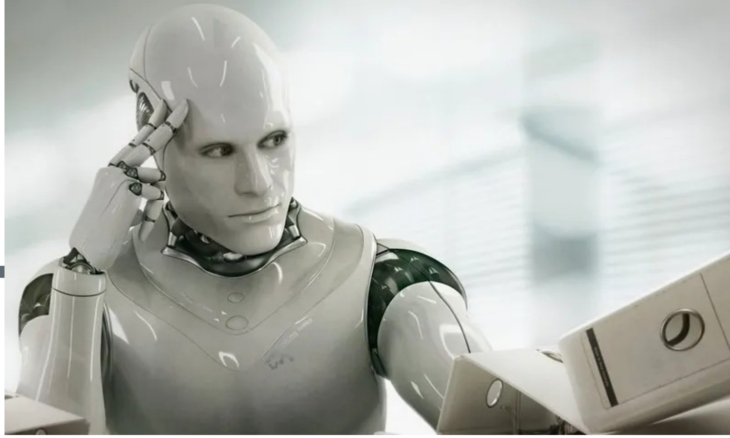


JFlow: Model-Independent Spherical Jeans Analysis using Equivariant Continuous Normalizing Flows

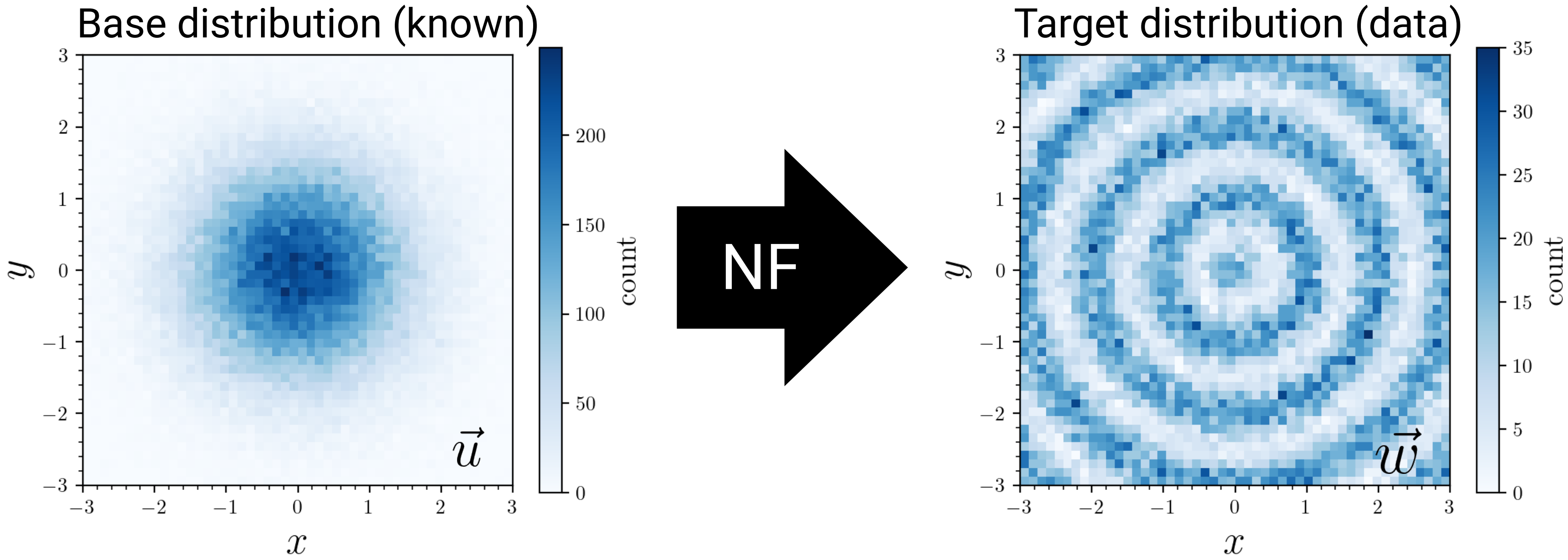
Collaboration with
K. Hayashi (NIT, Sendai College), S. Horigome (Tohoku),
S. Matsumoto (IPMU), M. M. Nojiri (KEK),

arXiv:2505.00763

Normalizing Flows: Neural Density Estimator



<u>Normalizing Flows</u> (NFs) is an artificial neural network that learns a transformation of random variables.



Base distribution (known)

Target distribution (data)

Main idea: if we could find out such transformation, we can use the transformation formula for the density estimation:

$$p_W(\vec{w}) = p_U(\vec{u}) \cdot \left| \frac{d\vec{u}}{d\vec{w}} \right|$$

We will use this model for estimating the phase space density f(x,v) from the data.

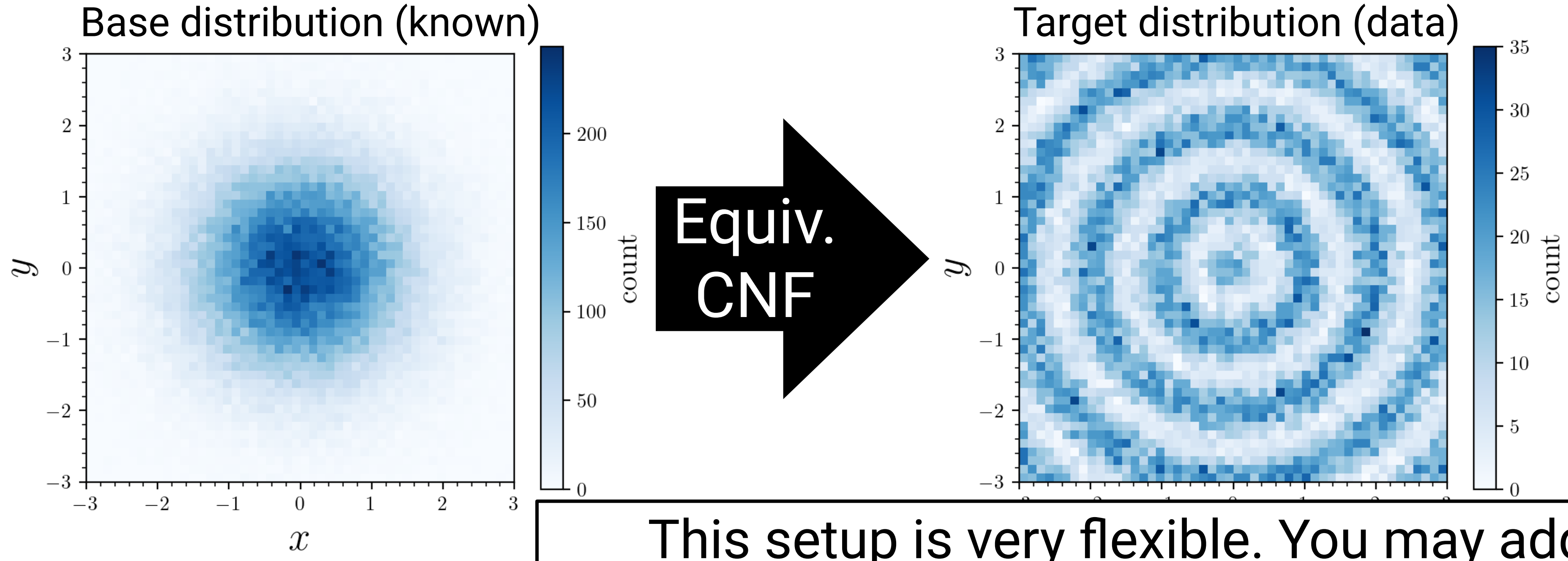
$\boxed{n(r)}$

Equivariant Continous Normalizing Flows

How to model spherically symmetric density using normalizing flows?
→ Use Equivariant Continuous Normalizing Flows!

$$\frac{d\vec{x}}{dt} = \vec{F}(\vec{x}, t) \longrightarrow \frac{d\vec{x}}{dt} = \hat{r} f(\vec{x}, t)$$

- Invariant (Gaussian) base distribution
- Equivariant vector field


Base distribution (known)

Target distribution (data)

This setup is very flexible. You may add **physics constraints** to **neural networks**, too!

Cored Spherical Density Model

In dSph analysis, we may further constrain the density model as conventional analysis often only consider the following type of densities.

- **Cored** density (constant density at r << 0)
- **Cuspy** density

ex) plummer sphere:

$$p(r) = \left(1 + \frac{r^2}{r_0^2}\right)^{-5/2}$$

Equivariant CNF for modeling **cored** density profile

$$\frac{d\vec{x}}{dt} = \hat{r} f(\vec{x}, t) \longrightarrow \frac{d\vec{x}}{dt} = \hat{r}\tanh\left(\frac{|\vec{x}|}{r_0}\right) f(\vec{x}, t)$$

Transformation at the origin is suppressed, remaining as Gaussian-shape. → cored density

$$\boxed{n(r)}$$

Cuspy Spherical Density Model

In dSph analysis, we may further constrain the density model as conventional analysis often only consider the following type of densities.

- **Cored** density (constant density at r << 1)
- **Cuspy** density

Equivariant CNF for modeling **cuspy** density profile

ex) NFW profile:

$$p(r) = \left(\frac{r}{r_0}\right)^{-1} \left(1 + \frac{r}{r_0}\right)^{-2} \to \frac{1}{r}$$

Apply power-law transform to radial component

$$|r| \to |r|^{c+1} \qquad \text{Jacobian} \propto r^{-\frac{3c}{1+c}}$$

to **cored** spherical symmetric density model

Here is a 6D density model, but...

Now we have a full 6D phase-space density model ready for solving spherical Jeans equation.

$p(\vec{r}) = n(r;\theta)$ modeled by equivariant CNF for cuspy halos

$$p(\vec{v}|\vec{r}) = \mathrm{GaussPDF}(\vec{v};\mu = 0, \Sigma(r;\theta))$$

$$f(\vec{r},\vec{v}) = p(\vec{r}) \times p(\vec{v}|\vec{r})$$

Wait, we only have x, y, vz.
How can we train this network by MLE?
We cannot use a conventional loss function.

Loss Function for Modeling Dwarf Spheroidal Galaxy

- In order to train the normalizing flow with spherical symmetry using limited kinematic information, we minimize the following entropy:

$$\mathcal{L}(\theta) = \int d\vec{w}_\perp \; p * K_h(\vec{w}_\perp) \; \log \hat{p} * K_h(\vec{w}_\perp; \theta)$$

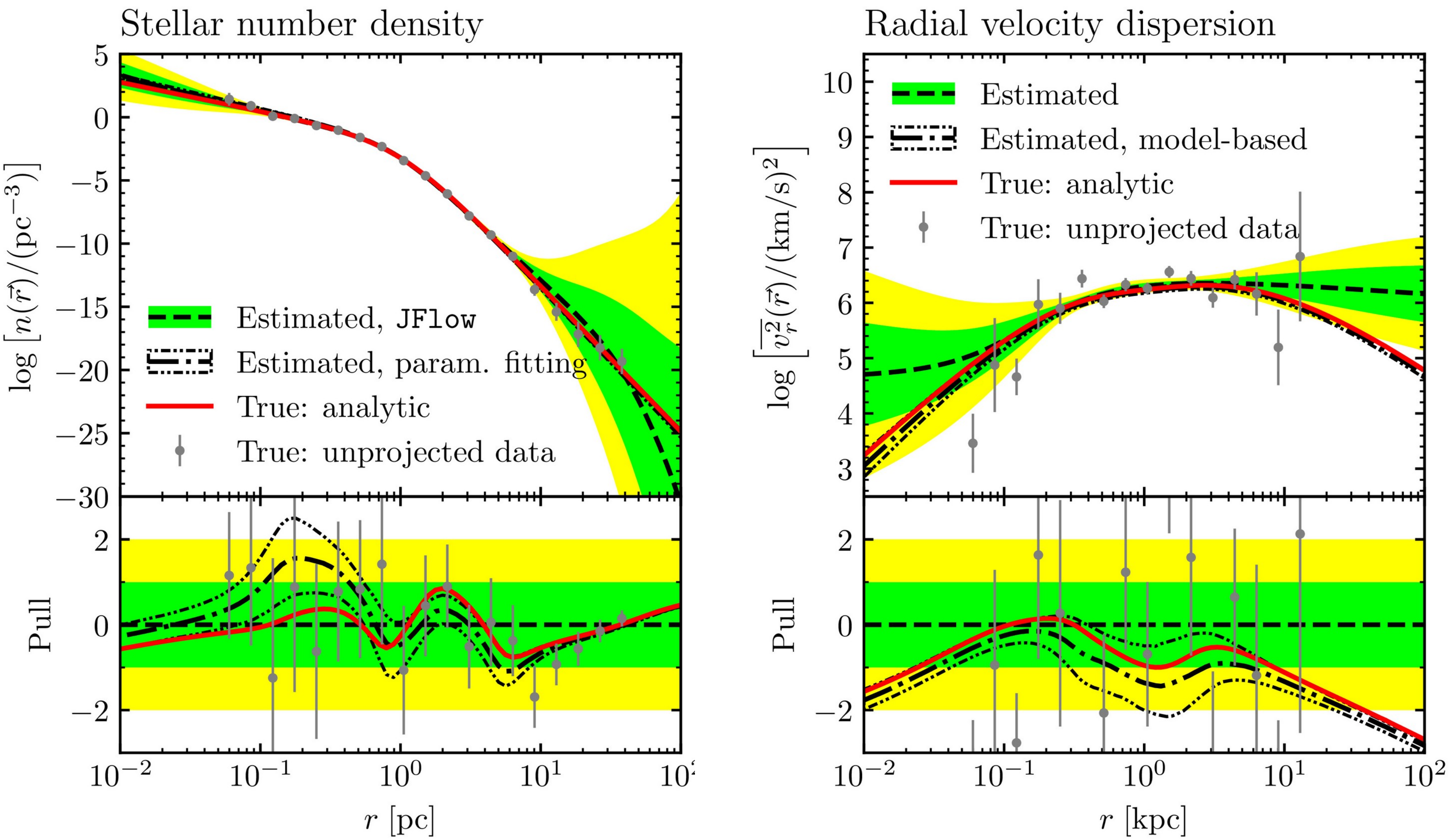
- Importance sampling: N_T training sample (stars) ~ p, N_K noise samples ~ K_h

$$\mathcal{L}(\theta) = \frac{1}{N N_K} \sum_{a=1}^{N} \sum_{b=1}^{N_K} \log \hat{p} * K_h(\vec{w}_\perp^{(a)} + \vec{\epsilon}^{(b)}; \theta)$$

- KDE for the smeared likelihood model:
  N_G generated stars from the normalizing flows~ \hat{p}

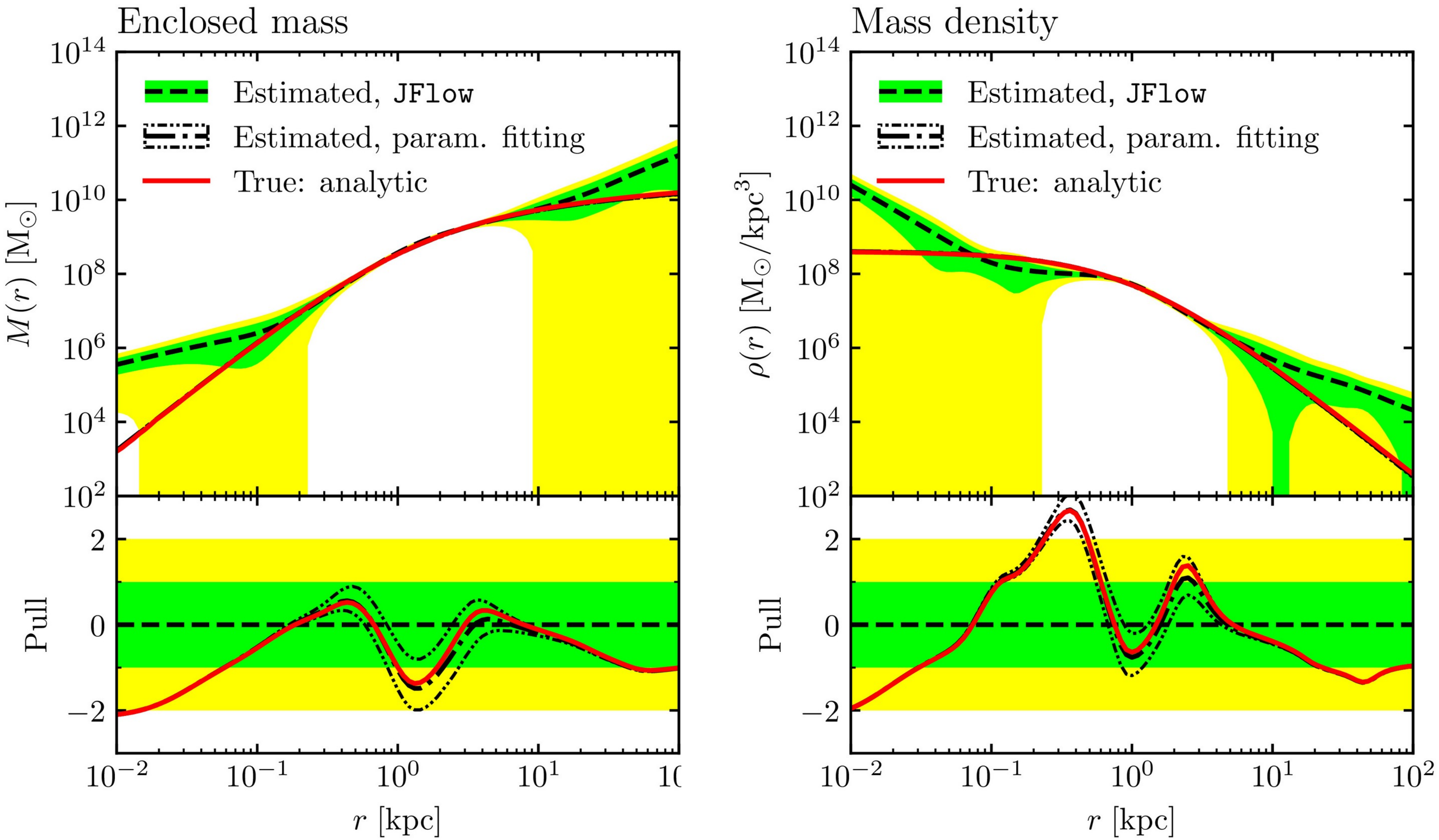
$$\mathcal{L}(\theta) = \frac{1}{N N_K} \sum_{a=1}^{N} \sum_{b=1}^{N_K} \log \frac{1}{N_G} \sum_{c=1}^{N_G} K_h \left[\vec{w}_\perp^{(a)} + \vec{\epsilon}^{(b)} - \vec{T}(\vec{z}^{(c)}; \theta) \right]$$

Results: stellar number density & radial velocity dispersion



Dataset: simulated dwarf spherodal galaxy from Gaia Challenge Dataset
https://astrowiki.surrey.ac.uk/doku.php?id=tests:sphtri

Results: dark matter mass density



Dataset: simulated dwarf spherodal galaxy from Gaia Challenge Dataset
https://astrowiki.surrey.ac.uk/doku.php?id=tests:sphtri

This approach is a bit slow… Can we improve somehow by removing redundant component?

Star catalog

$$\{(\vec{x}, \vec{v})\}$$

Gaia DR3:
star catalog of
the Milky Way

Phase space density

$$f(\vec{x}, \vec{v})$$

Neural Networks for Density Estimation:
Normalizing Flows

$$\vec{u}_0 \to \vec{u}_1 \to \cdots \to \vec{u}_n = (\vec{x}, \vec{v})$$

Gravitational accel.

$$\vec{a}(\vec{x})$$

Solving EOM (Boltzmann Equation)

$$\left[\frac{\partial}{\partial t} + \vec{v} \cdot \frac{\partial}{\partial \vec{x}} + \vec{a} \cdot \frac{\partial}{\partial \vec{v}}\right] f(\vec{x}, \vec{v}) = 0$$

Mass density

$$\rho(\vec{x})$$

Solving Gauss's Equation

$$-4\pi G\rho = \nabla \cdot \vec{a}$$

arXiv: 2205.01129, 2305.13358
See also Green et. al. arXiv:2011.04673, arXiv:2205.02244

Return to outline of strategy...

Star catalog
$\{(\vec{x}, \vec{v})\}$

Gaia DR3:
star catalog of
the Milky Way

Phase space density
$f(\vec{x}, \vec{v})$

Neural Networks for Density Estimation:
Normalizing Flows

$$\vec{u}_0 \to \vec{u}_1 \to \cdots \to \vec{u}_n = (\vec{x}, \vec{v})$$

Gravitational accel.
$\vec{a}(\vec{x})$

Solving EOM (Boltzmann Equation)

$$\left[\frac{\partial}{\partial t} + \vec{v} \cdot \frac{\partial}{\partial \vec{x}} + \vec{a} \cdot \frac{\partial}{\partial \vec{v}}\right] f(\vec{x}, \vec{v}) = 0$$

Mass density
$\rho(\vec{x})$

Solving Gauss's Equation

$$-4\pi G\rho = \nabla \cdot \vec{a}$$

arXiv: 2205.01129, 2305.13358
See also Green et. al. arXiv:2011.04673, arXiv:2205.02244

We need just density derivative!

Gravitational accel.

$$\vec{a}(\vec{x})$$

Solving EOM (Boltzmann Equation)

$$\left[\frac{\partial}{\partial t} + \vec{v} \cdot \frac{\partial}{\partial \vec{x}} + \vec{a} \cdot \frac{\partial}{\partial \vec{v}}\right] f(\vec{x}, \vec{v}) = 0$$

Equilibriuum Collisionless Boltzmann Equation:

$$\left[\vec{v} \cdot \frac{\partial}{\partial \vec{x}} + \vec{a} \cdot \frac{\partial}{\partial \vec{v}}\right] f(\vec{x}, \vec{v}) = 0$$

In order to estimate acceleration, knowing the following density derivatives are **sufficient**!

Density Derivatives

$$\frac{\partial f}{\partial \vec{x}} \qquad \frac{\partial f}{\partial \vec{v}}$$

→

Gravitational accel.

$$\vec{a}(\vec{x})$$

Machine Learning Method for Density Derivative Estimation?

Score Matching (Hyvariene 2005)

Score function = log-density derivative

Score functions

$$\frac{\partial}{\partial \vec{x}} \log f \qquad \frac{\partial}{\partial \vec{v}} \log f$$

We want to make **neural network** regressing **score functions!**

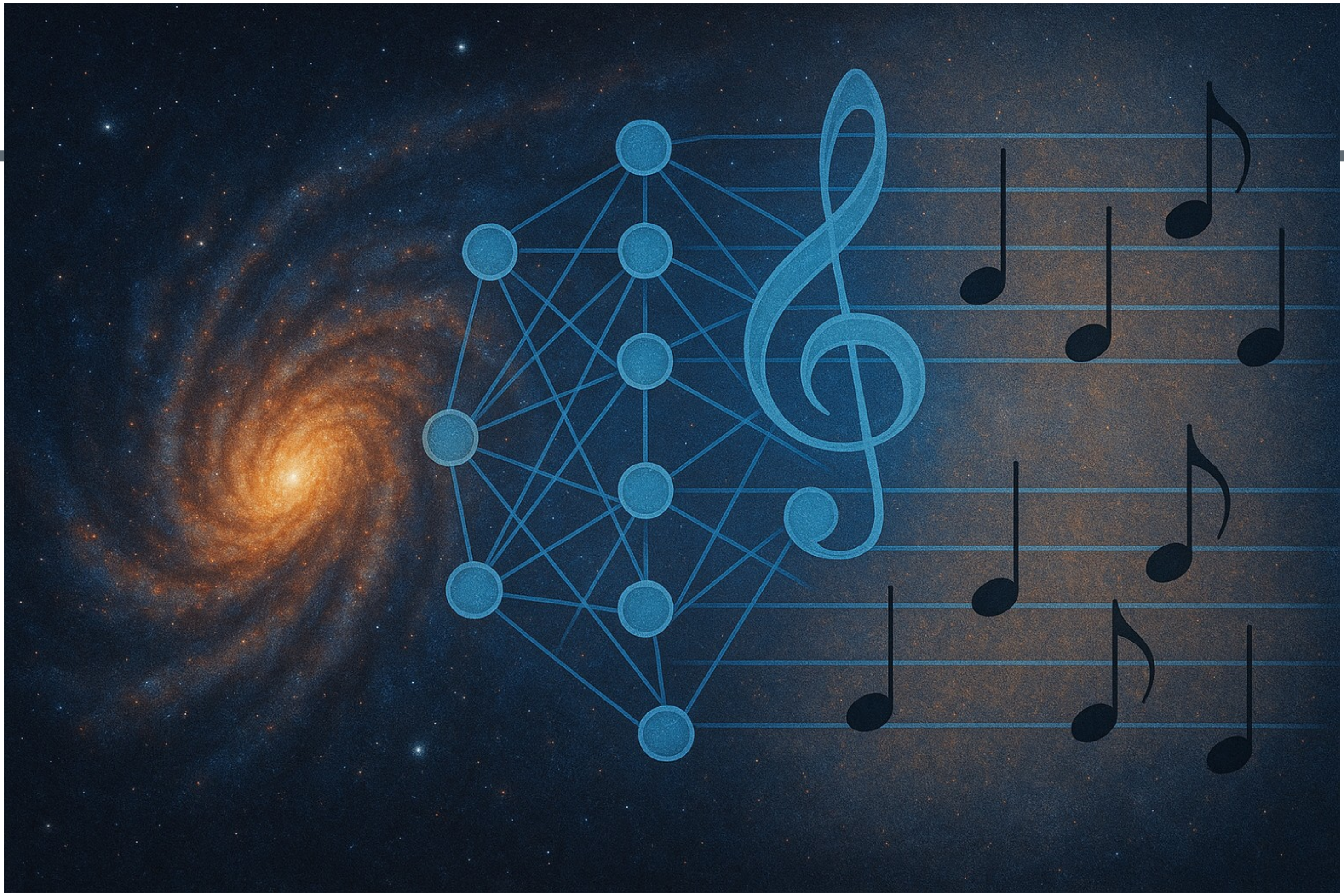
Explicit Score Matching Loss = (MSE Loss)

$$\mathcal{L} = E_x \left| s_i(\vec{x}; \theta) + \frac{\partial}{\partial \vec{x}} \log p(\vec{x}) \right|^2$$

Inplicit Score Matching Loss

$$\mathcal{L} = E_x \left[\frac{\partial}{\partial x_i} s_i(\vec{x}; \theta) + \frac{1}{2} |s_i(\vec{x}; \theta)|^2 \right]$$

Equivalent loss functions!



GalaxyScore: Score Matching for Galactic Dark Matter Density Estimation

Collaboration with M. R. Buckley, E. Putney, D. Shih (Rutgers)
proceeding accepted in NeurIPS 2025 ML4PS workshop

New faster outline of strategy!

Star catalog

$\{(\vec{x}, \vec{v})\}$

Gaia DR3:
star catalog of
the Milky Way

Score functions

$$\frac{\partial}{\partial \vec{x}} \log f \qquad \frac{\partial}{\partial \vec{v}} \log f$$

Score Matching (Hyvariene 2005)

$$\mathcal{L} = E_x \left[\frac{\partial}{\partial x_i} s_i(\vec{x};\theta) + \frac{1}{2} |s_i(\vec{x};\theta)|^2 \right]$$

Gravitational accel.

$\vec{a}(\vec{x})$

Solving EOM (Boltzmann Equation)

$$\left[\frac{\partial}{\partial t} + \vec{v} \cdot \frac{\partial}{\partial \vec{x}} + \vec{a} \cdot \frac{\partial}{\partial \vec{v}} \right] \log f(\vec{x}, \vec{v}) = 0$$

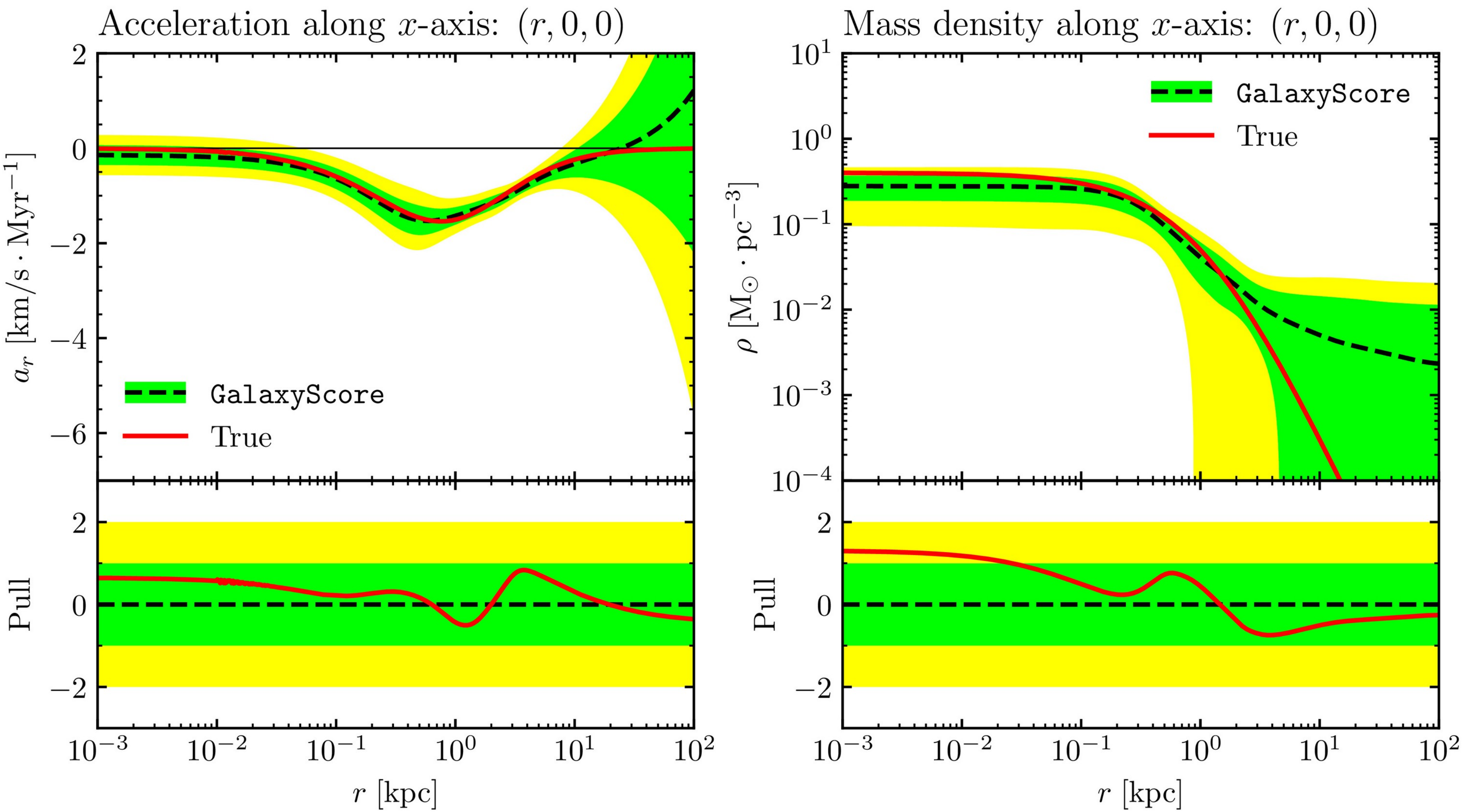
Mass density

$\rho(\vec{x})$

Solving Gauss's Equation

$$-4\pi G \rho = \nabla \cdot \vec{a}$$

Proof-of-Concept Results



Dataset: simulated dwarf spherodal galaxy from Gaia Challenge Dataset
https://astrowiki.surrey.ac.uk/doku.php?id=tests:sphtri

Awesome collaborators of my projects:

Prof. David Shih
(Rutgers)

Prof. Matthew Buckley
(Rutgers)

Prof. Mihoko Nojiri
(KEK)

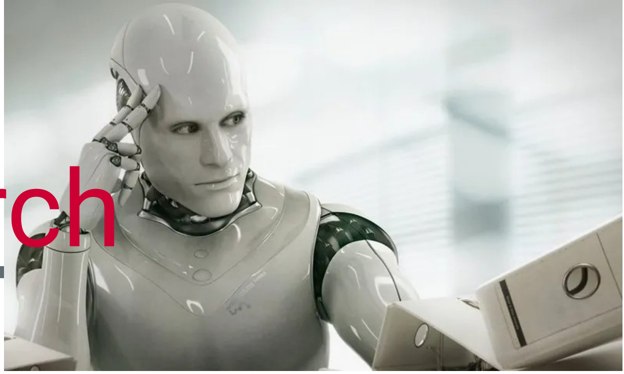
Prof. Kohei Hayashi
(NIT, Sendai College)

Eric Putney
(Ph.D. student at Rutgers,
on **job market this year!**)

Prof. Shigeki
Matsumoto
(IPMU)

Dr. Shunichi
Horigome
(Tohoku)

Future Timeline for Galactic Dynamics Research

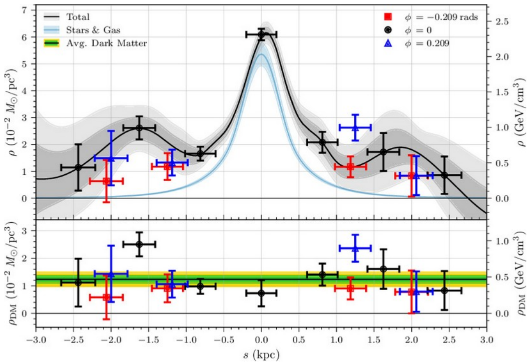


2025

Gaia DR3

2026 ... 2030

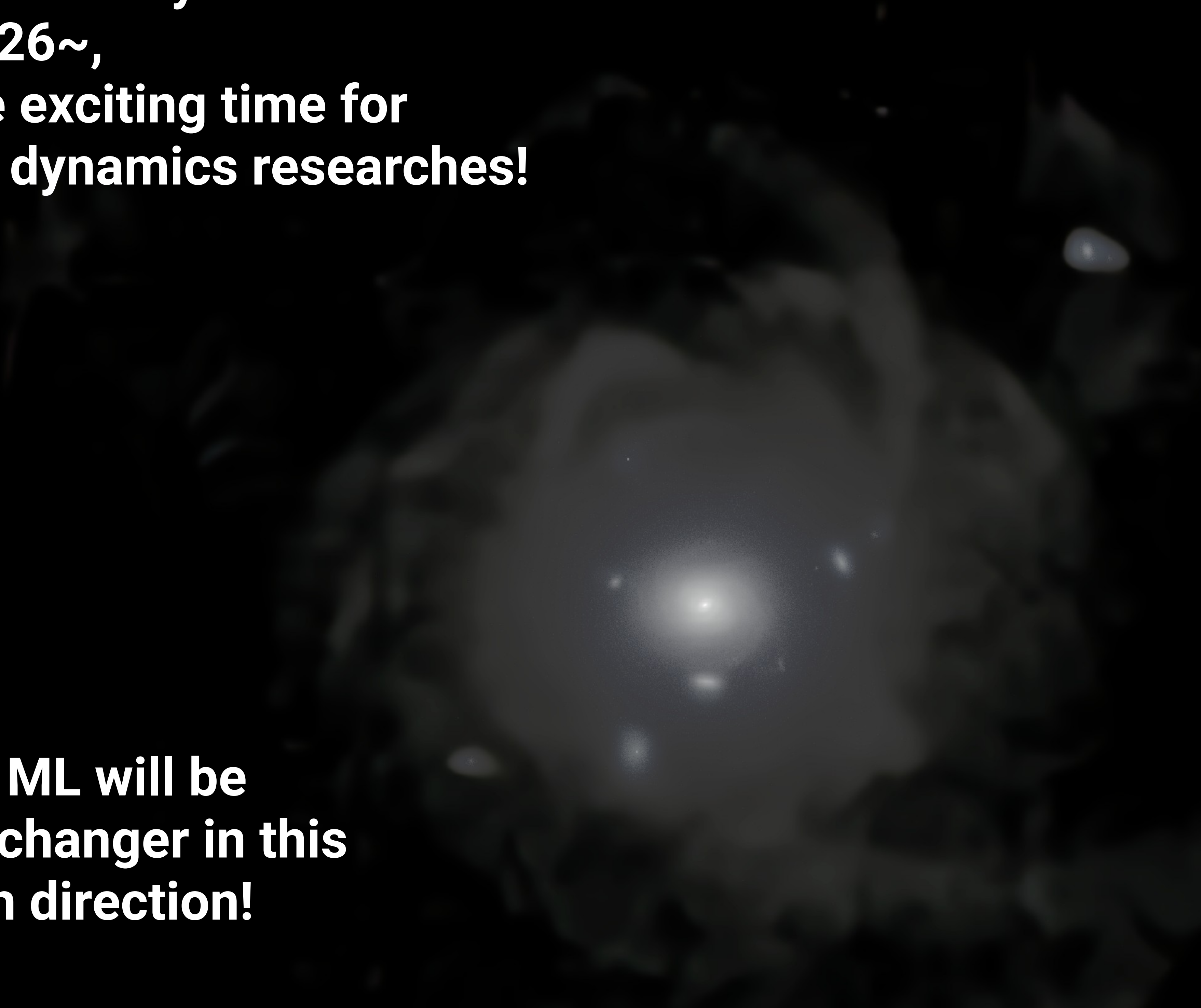
Gaia DR4, SuperPFS, and so on!



Understanding DM in solar neighborhood

New methods for analyzing distant satellite galaxies

More precision and new opportunities for studying galactic substructures and dark matter!

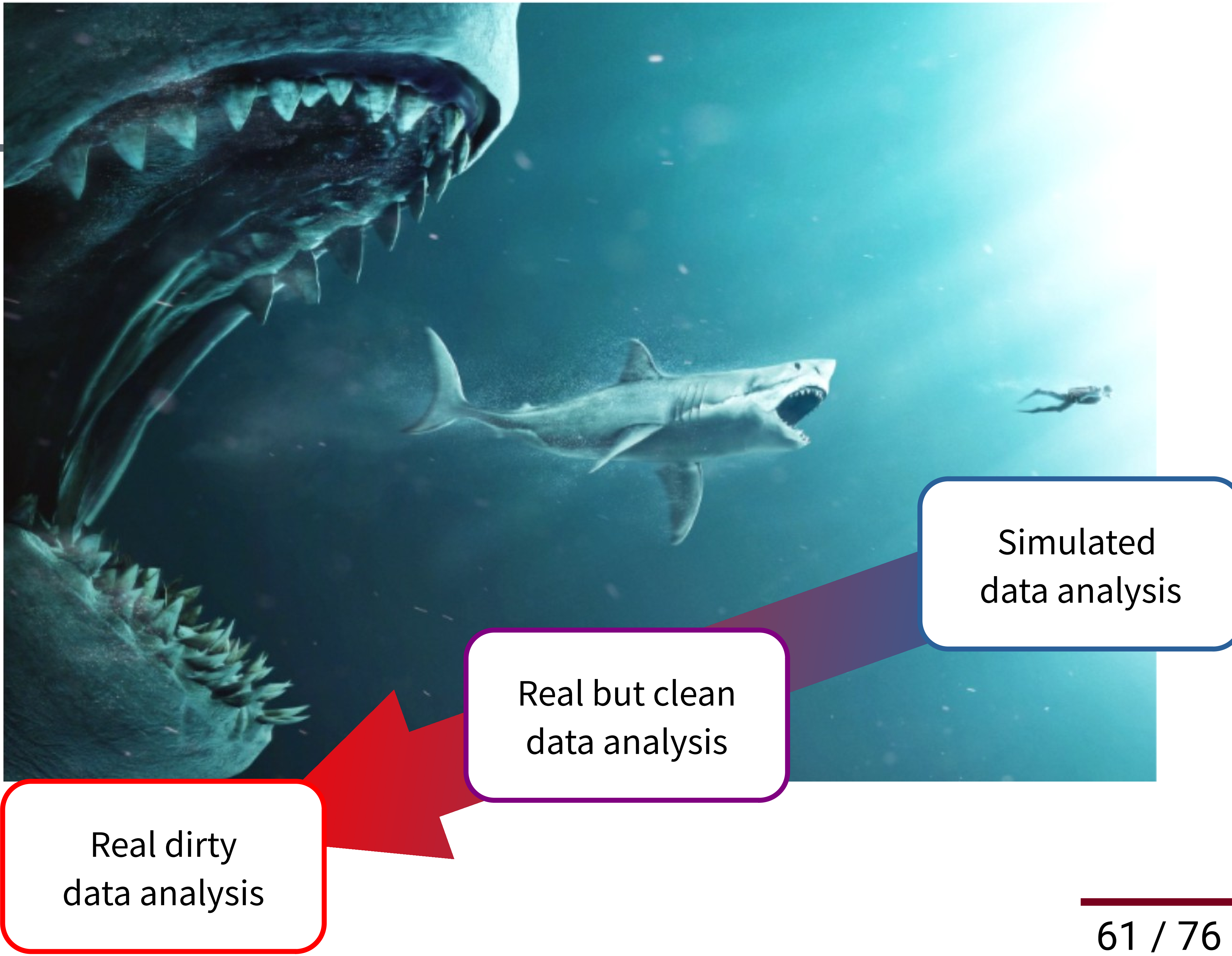


With new survey data delivered from 2026~, it will be exciting time for galactic dynamics researches!

and AI / ML will be a game changer in this research direction!

Thank you
for listening!

Backups



Simulated data analysis

Real but clean data analysis

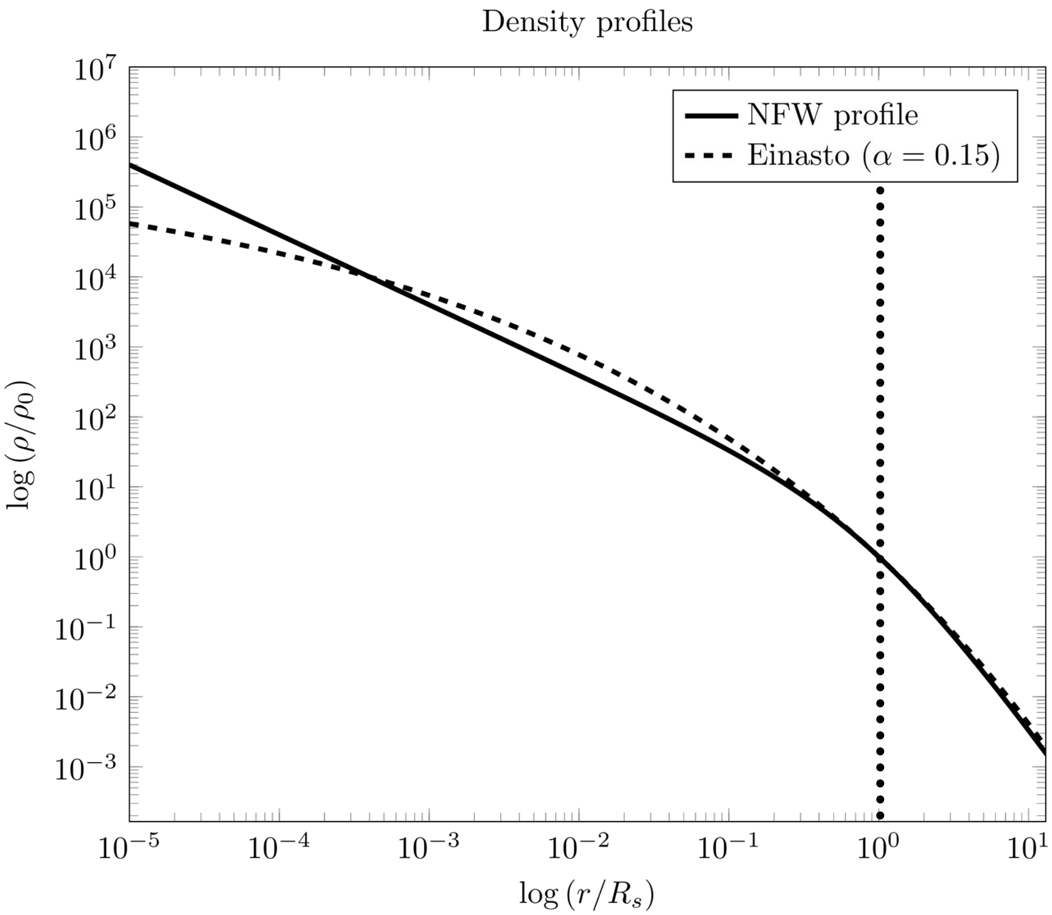
Real dirty data analysis

Navarro–Frenk–White (NFW) profile

A commonly used dark matter halo model empirically identified in N-body simulations

$$\rho(r) = \frac{\rho_0}{\frac{r}{R_s}\left(1 + \frac{r}{R_s}\right)^2}$$

If dark matter exhibits non-trivial interactions, the **halo shape may vary**.



Self-interacting dark matter, wave dark matter ….

https://en.wikipedia.org/wiki/Navarro%E2%80%93Frenk%E2%80%93White_profile

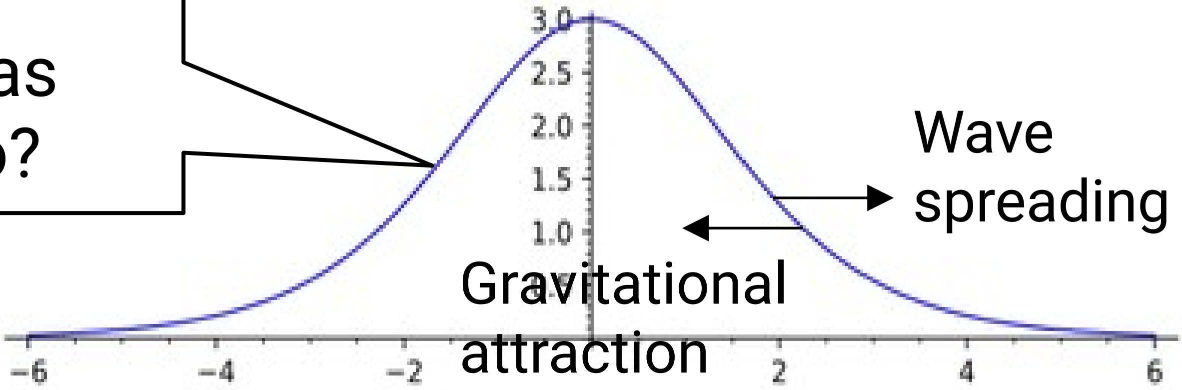
Example: Wave Dark Matter

If DM mass is so light (e.g. very light axions) so that

inter-particle spacing << de Broglie wavelength

DM exhibits **wave-like** behavior.

Nontrivial stable solution:
Soliton
→ Soliton solution as
a dark matter halo?



Smoking gun signatures

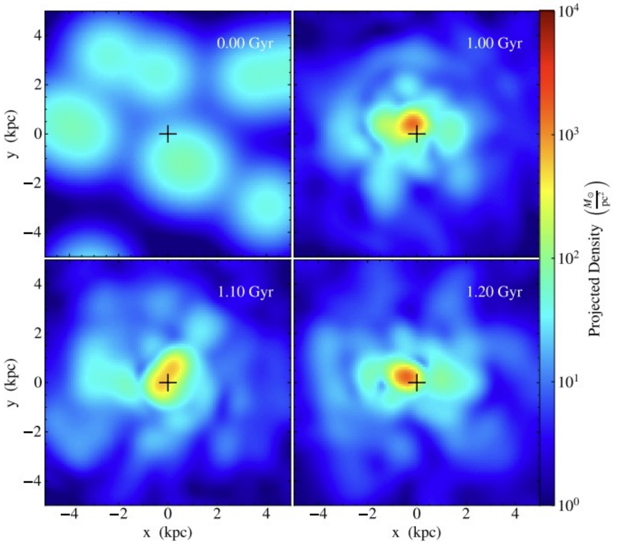
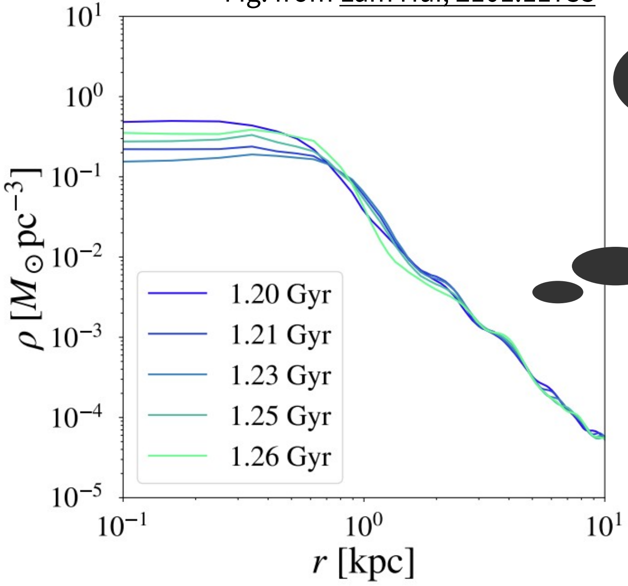


Fig. from Lam Hui, 2101.11735



Soliton Oscillations

Soliton Core

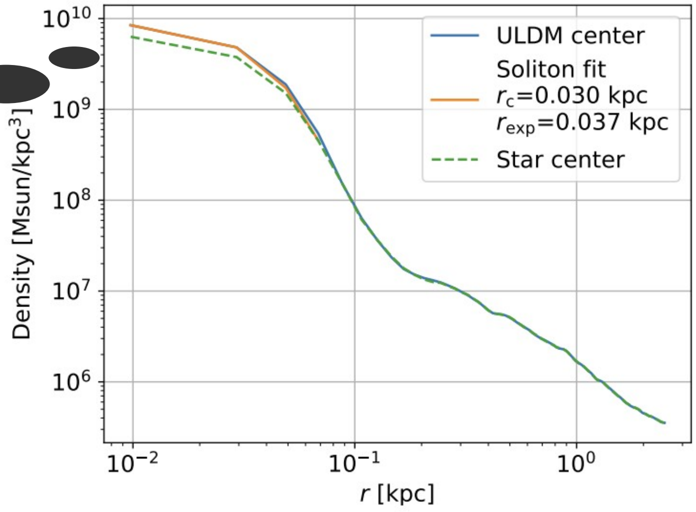
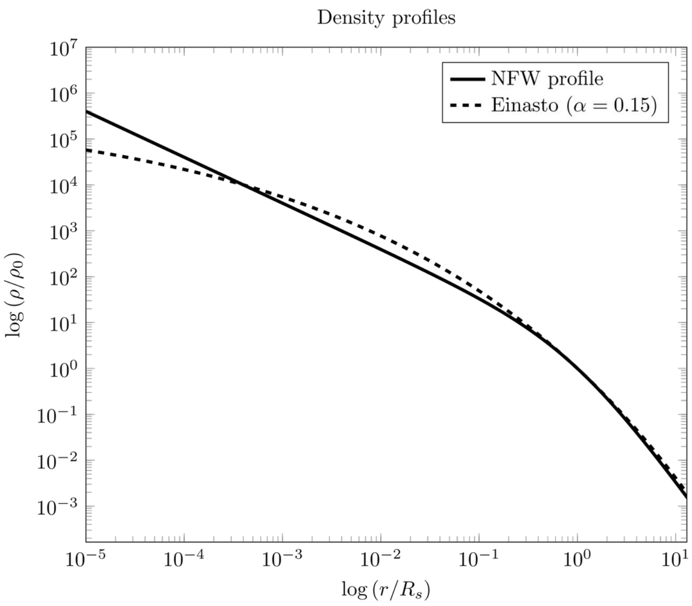
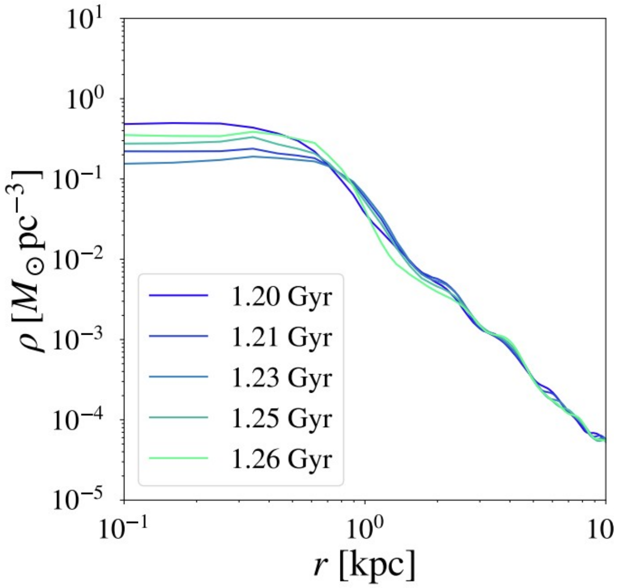
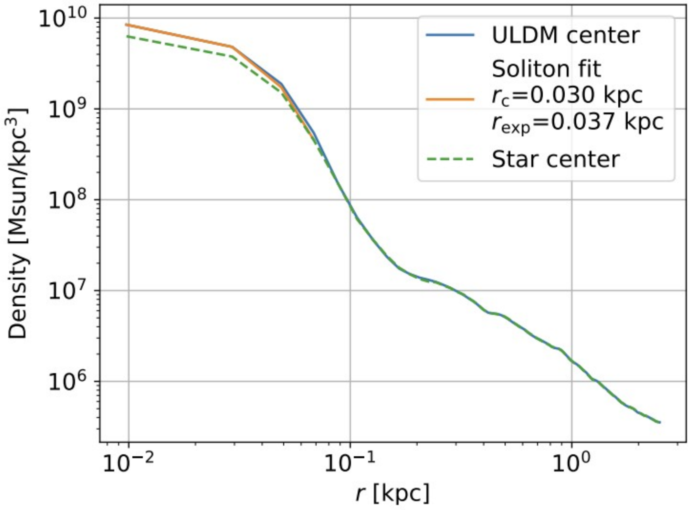


Fig. from talk by Teodori Luca, IBS Let there be light (particles) Workshop

Need for model-independent analysis





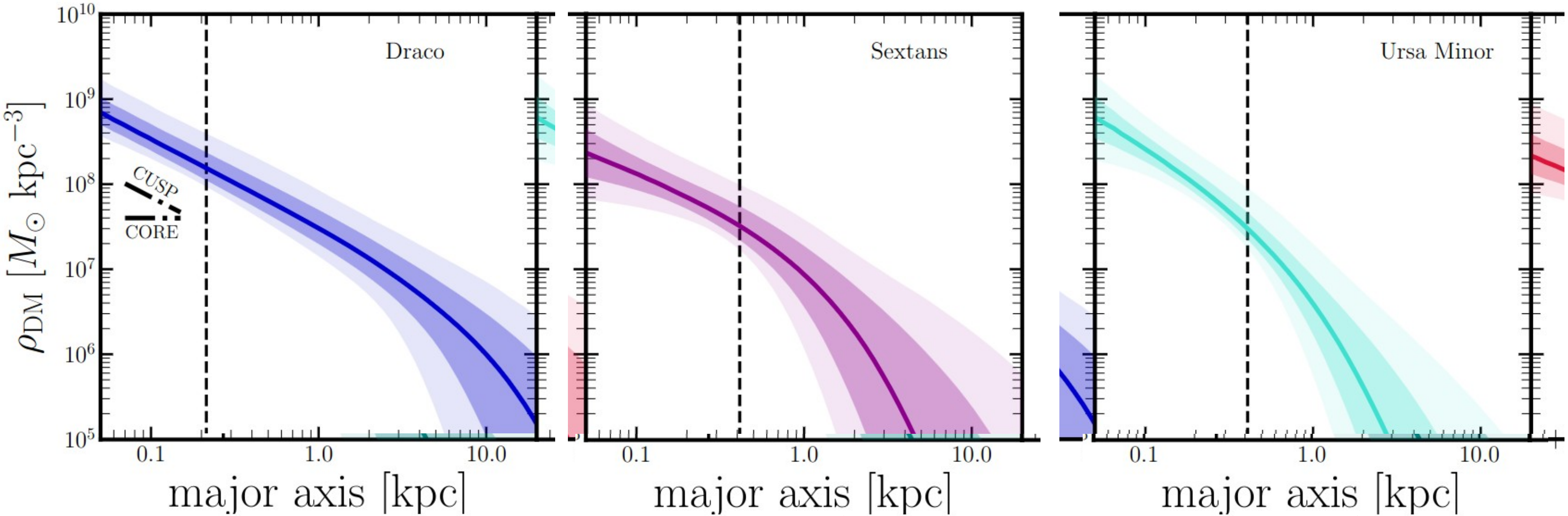


As many non-trivial DM halos are considered nowadays, we need a **free-form DM density estimation** in order to do a **model-indepdent** DM halo analysis.

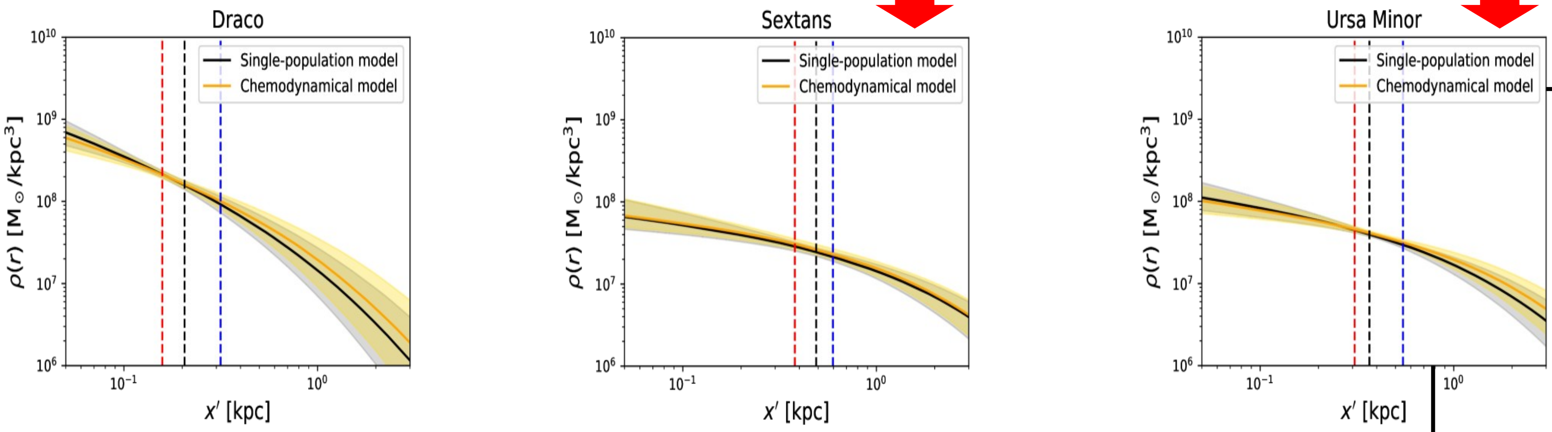
Again, unsupervised **machine learning** can help solving this type of problem!

Need for model-independent analysis

Hayashi et. al., 2007.13780



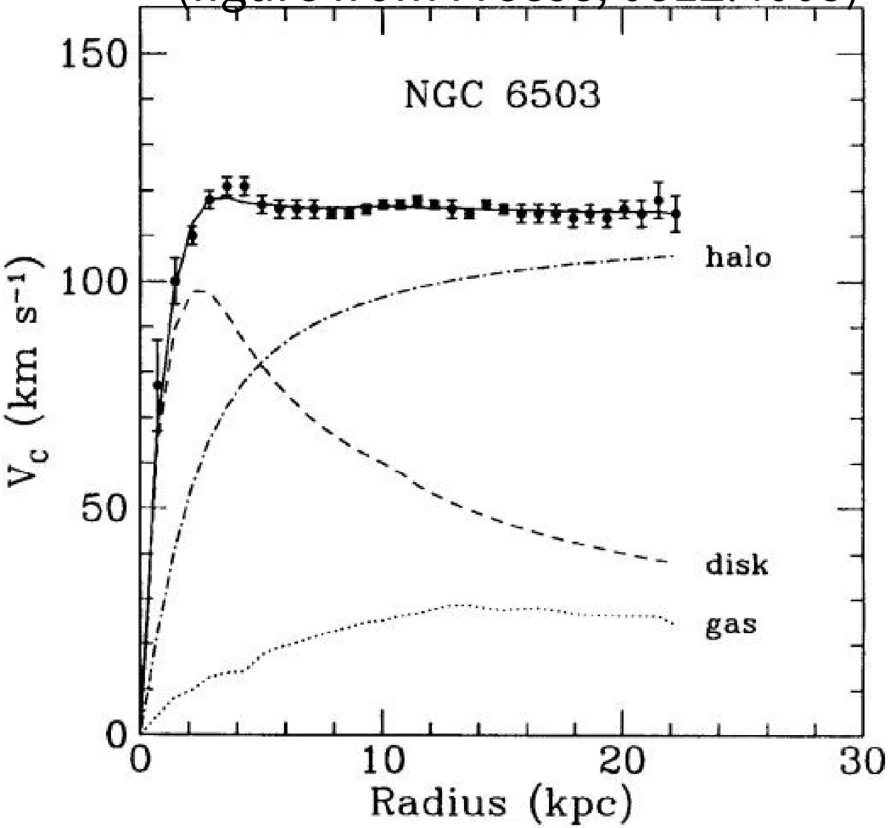
DESI Milky Way Survey, 2507.02284



Tension between two measurements..

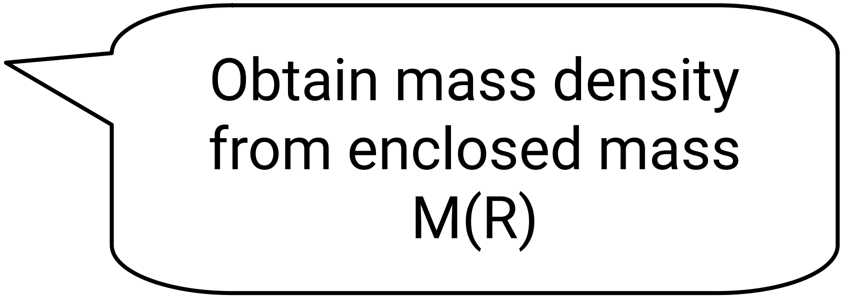
Galaxy rotation curve and dark matter NGC 6503

(figure from Freese, 0812.4005)



NGC 6503 from NASA Hubble telescope

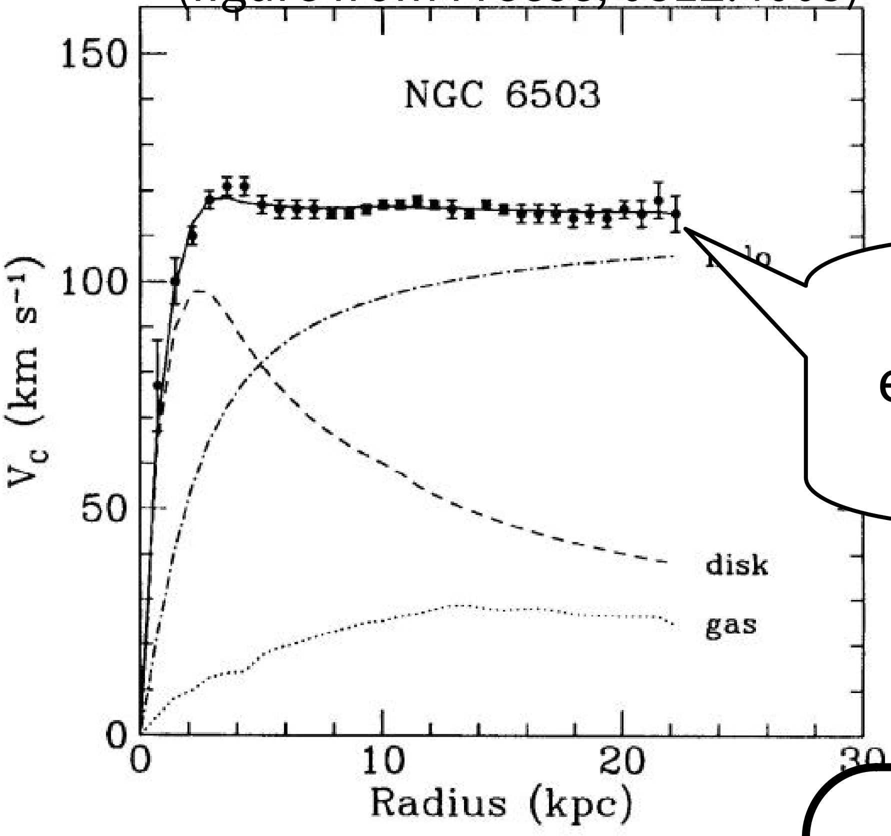
$$v_{\rm circ}(R) = \sqrt{\frac{GM(R)}{R}}$$



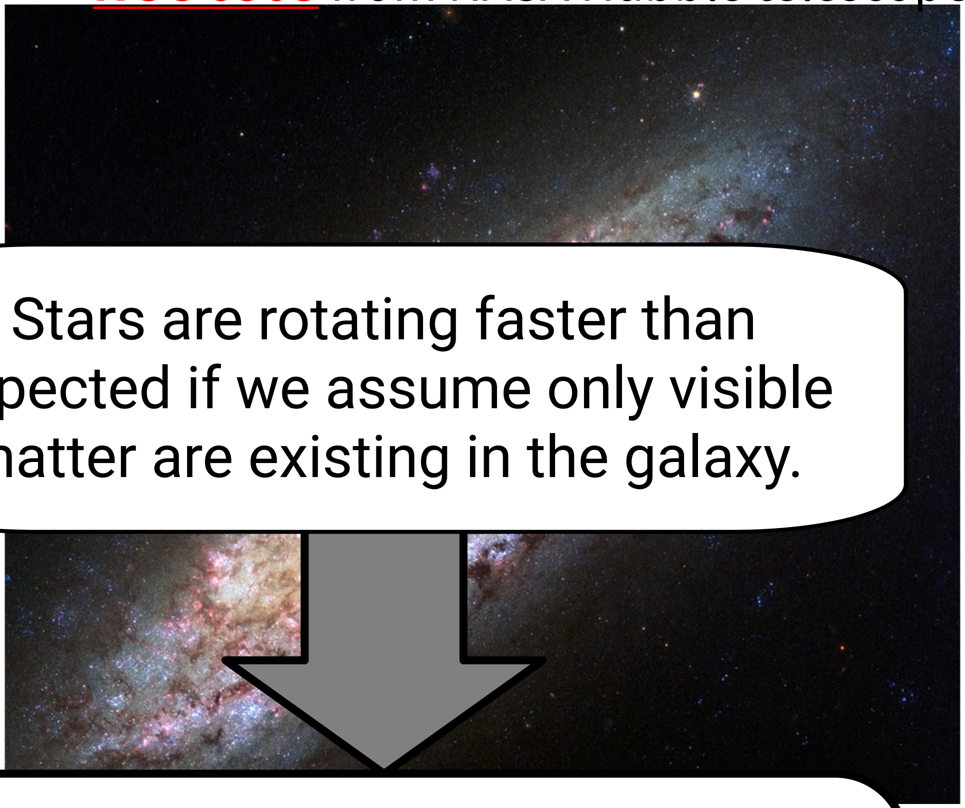
Obtain mass density from enclosed mass M(R)

Galaxy rotation curve and dark matter NGC 6503

(figure from Freese, 0812.4005)



NGC 6503 from NASA Hubble telescope



Stars are rotating faster than expected if we assume only visible matter are existing in the galaxy.

$$v_{\rm circ}(R) = \sqrt{\frac{GM(R)}{R}}$$

DARK MATTER!

Galaxy rotation curve of Milky Way

Ou, et. al., 2303.12838

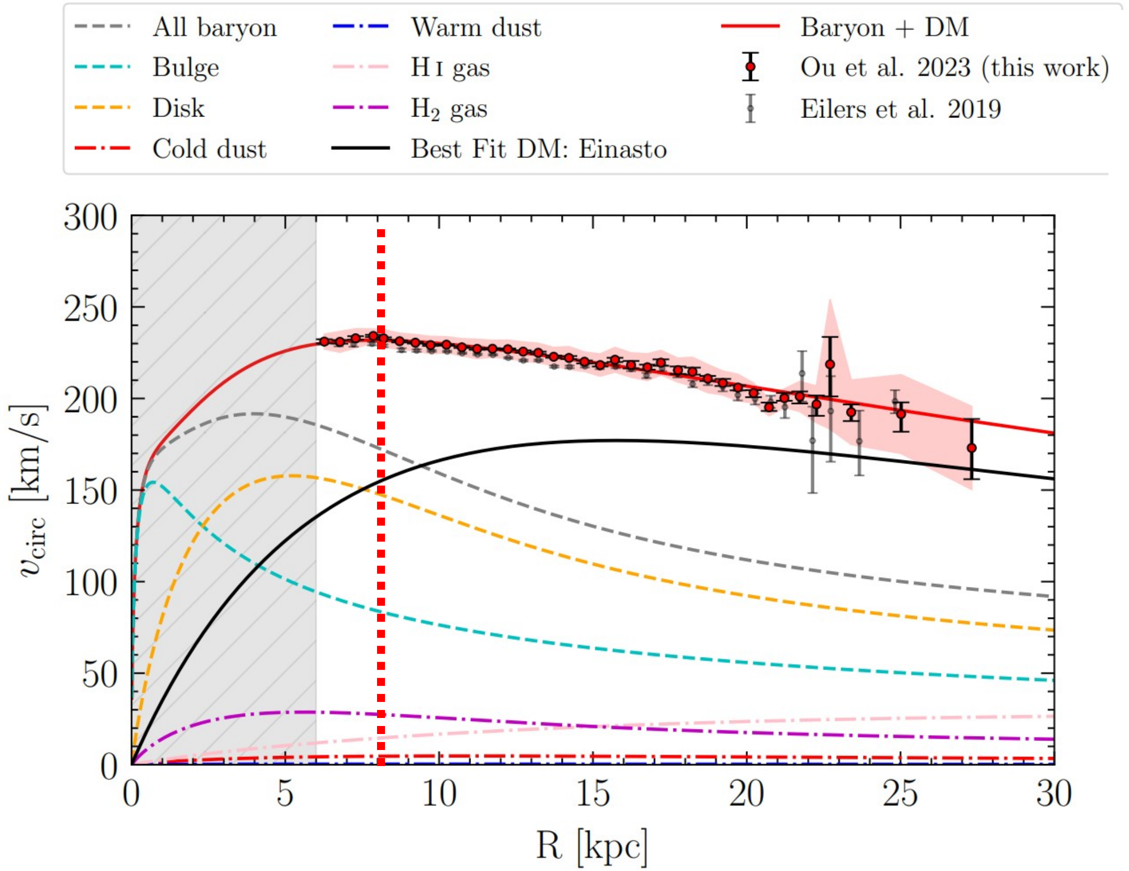
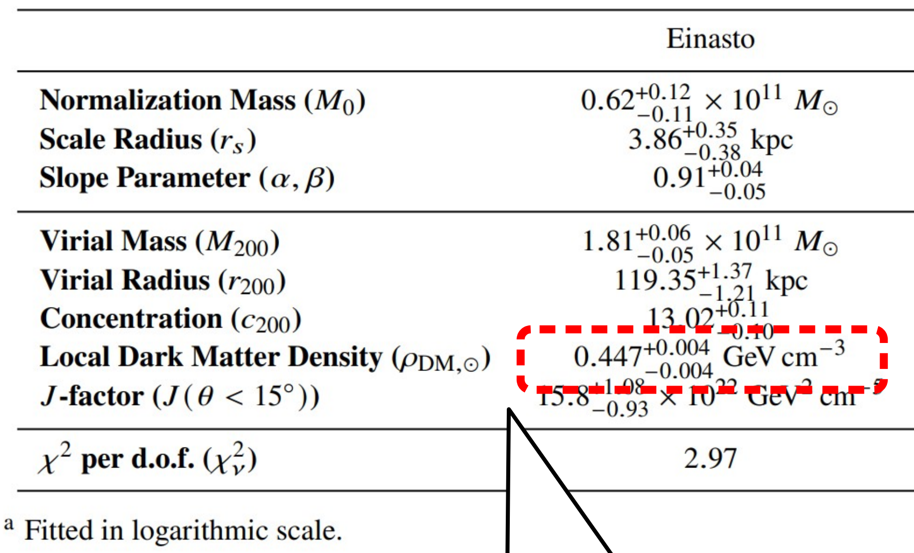


Figure 4. Comparison between the circular velocity curve measured from Eilers et al. (2019) (black) and this work (red). The best-fit Einasto DM profile, with the baryonic model from de Salas et al. (2019), is also shown here. The grey shaded region represents the bulge region, which we do not model due to the non-axisymmetric potential near the galactic bar. The red shaded region represents the total uncertainty estimate from the dominating systematic sources, as shown in Figure 5.

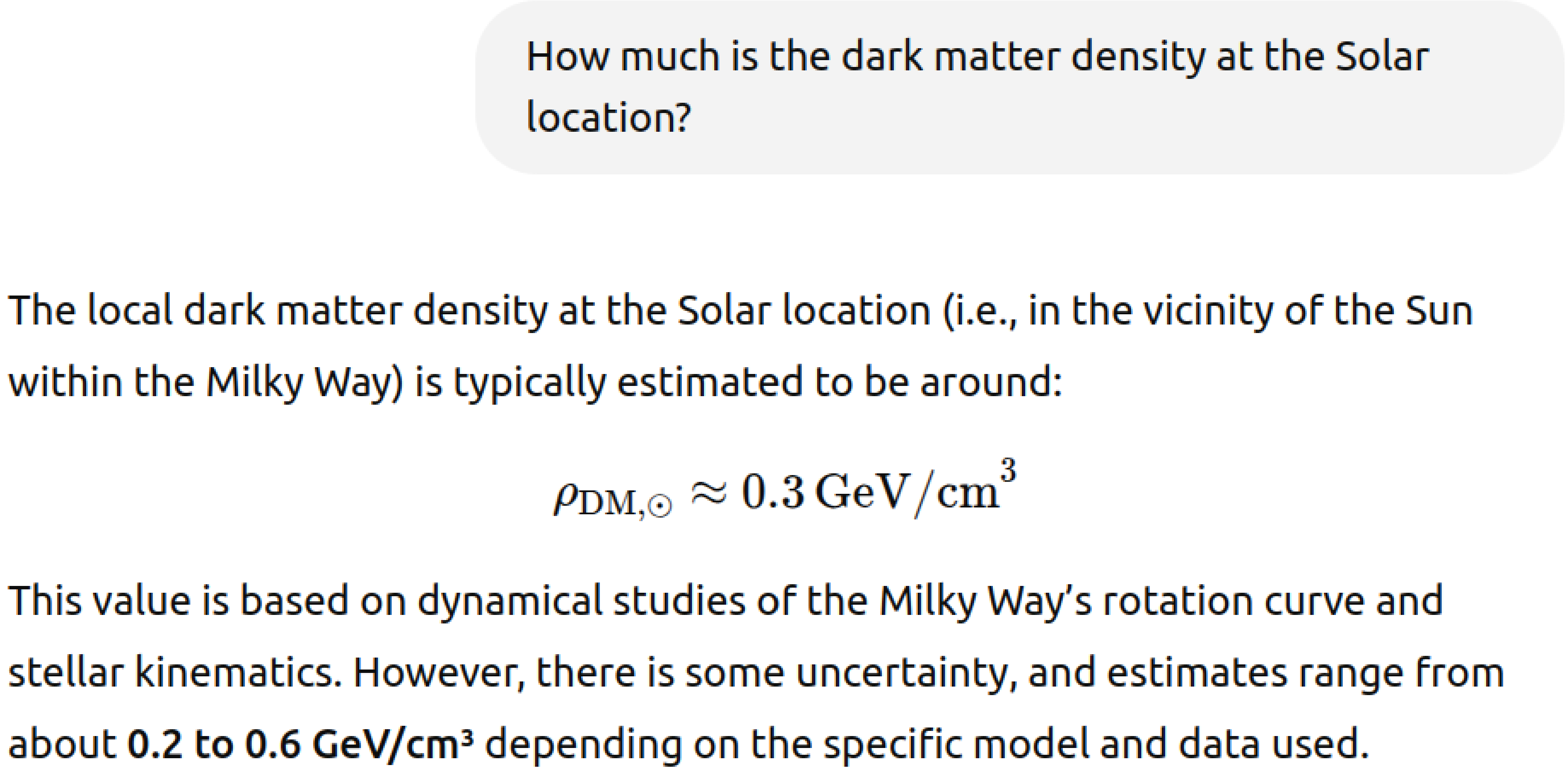


	Einasto
Normalization Mass (M_0)	$0.62^{+0.12}_{-0.11} \times 10^{11}\ M_\odot$
Scale Radius (r_s)	$3.86^{+0.35}_{-0.38}$ kpc
Slope Parameter (α, β)	$0.91^{+0.04}_{-0.05}$
Virial Mass (M_{200})	$1.81^{+0.06}_{-0.05} \times 10^{11}\ M_\odot$
Virial Radius (r_{200})	$119.35^{+1.37}_{-1.21}$ kpc
Concentration (c_{200})	$13.02^{+0.11}_{-0.10}$
Local Dark Matter Density ($\rho_{\rm DM,\odot}$)	$0.447^{+0.004}_{-0.004}$ GeV cm^{-3}
J**-factor** ($J(\theta < 15°)$)	$15.8^{+1.08}_{-0.93} \times 10^{22}$ GeV^2 cm^{-5}
χ^2 **per d.o.f.** (χ^2_ν)	2.97

[a] Fitted in logarithmic scale.

Very precise DM density measurement!

Probably, you heard about the dark matter density on Earth...



How much is the dark matter density at the Solar location?

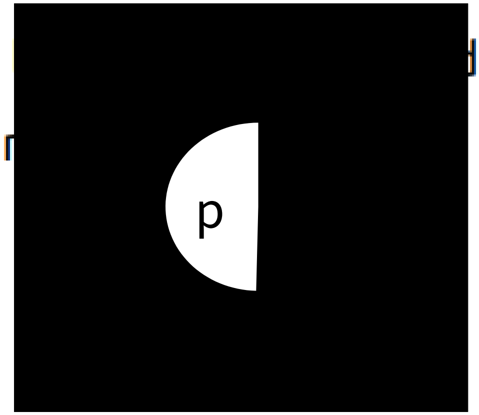
The local dark matter density at the Solar location (i.e., in the vicinity of the Sun within the Milky Way) is typically estimated to be around:

$$\rho_{\mathrm{DM},\odot} \approx 0.3\,\mathrm{GeV}/\mathrm{cm}^3$$

This value is based on dynamical studies of the Milky Way's rotation curve and stellar kinematics. However, there is some uncertainty, and estimates range from about **0.2 to 0.6 GeV/cm³** depending on the specific model and data used.

Would you

observatio



About 1/3-proton
in a cubic centimeter box!

Galaxy rotation curve of Milky Way

Ou, et. al., 2303.12838

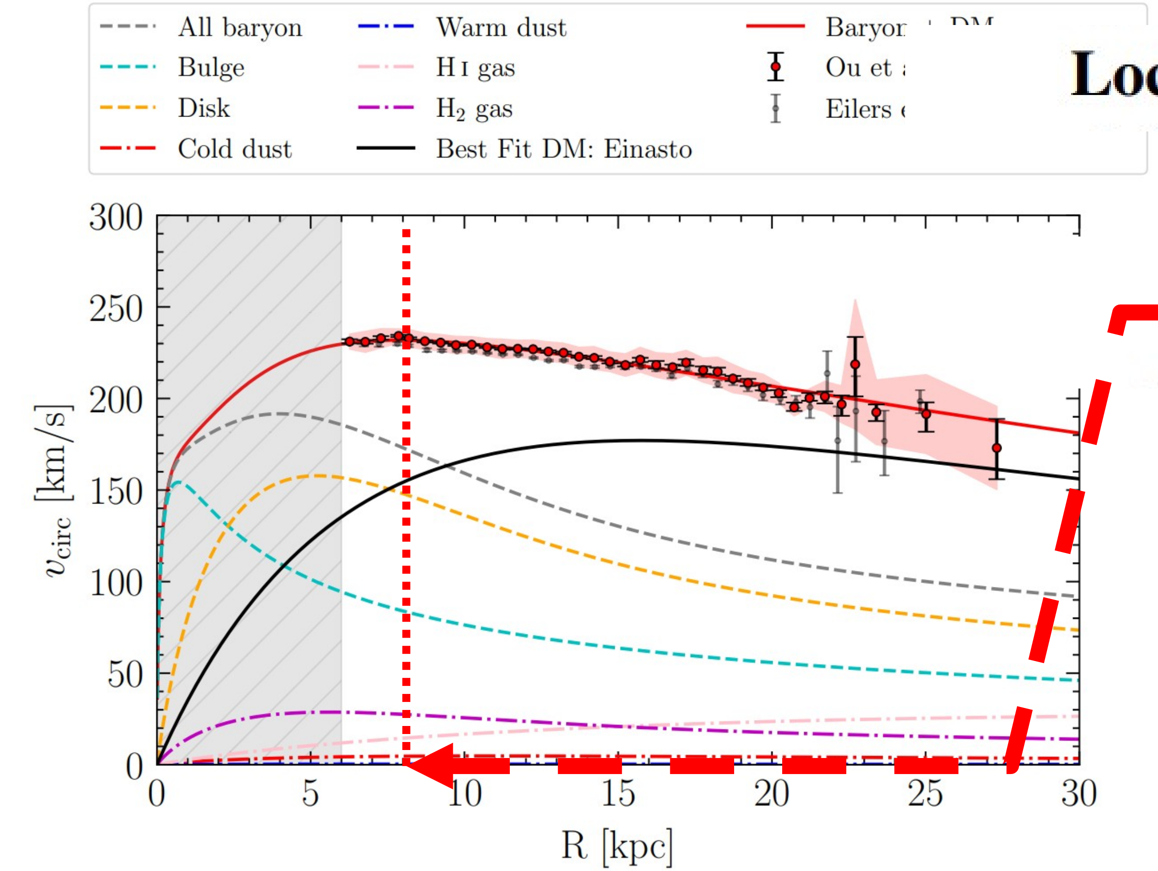
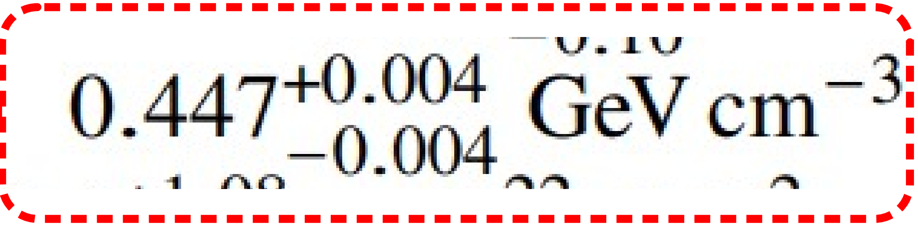


Figure 4. Comparison between the circular velocity curve measured from Eilers et al. (2019) (black) and this work (red). The best-fit Einasto DM profile, with the baryonic model from de Salas et al. (2019), is also shown here. The grey shaded region represents the bulge region, which we do not model due to the non-axisymmetric potential near the galactic bar. The red shaded region represents the total uncertainty estimate from the dominating systematic sources, as shown in Figure 5.

Local Dark Matter Density ($\rho_{\rm DM,\odot}$)

at Sun's location

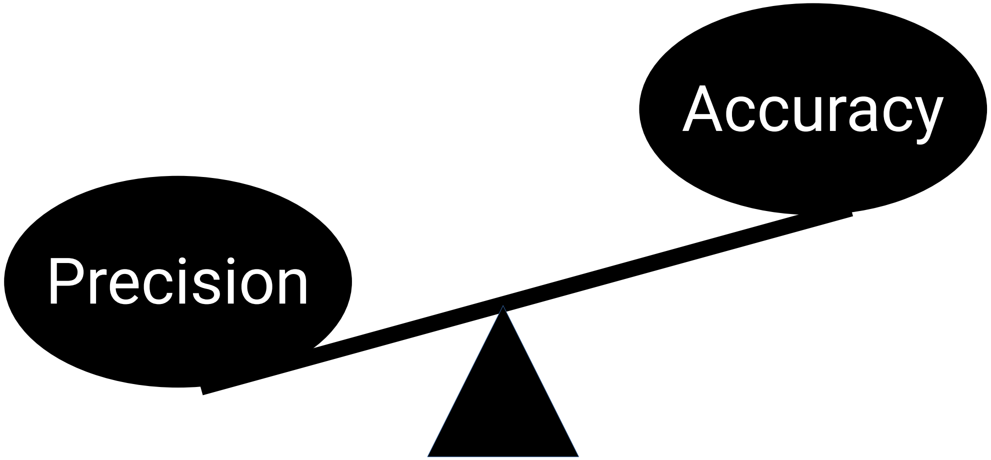


$0.447^{+0.004}_{-0.004}$ GeV cm^{-3}

Very precise
DM density
measurement!

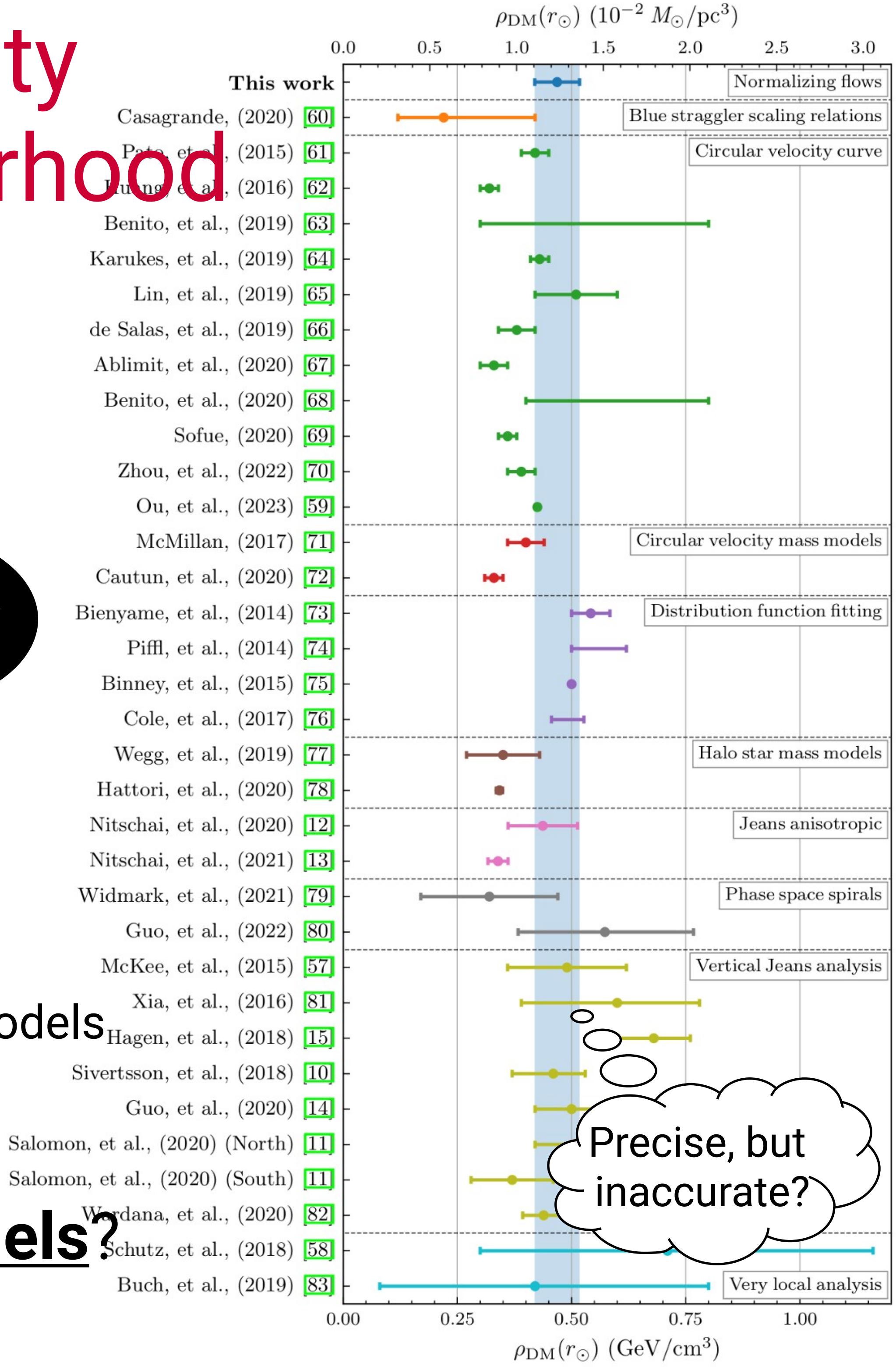
Measuring DM density in the Solar Neighborhood

Thanks to recent progress in observing stars in the Milky Way, we can measure **the dark matter density in the Solar neighborhood** in very high precision using model-based analyses.



When **sufficient number of data are available**, using over-constrained models may results in inaccurate results.

Need of analysis without assumed **symmetries** and **models**?



Here is a 6D density model, but...

Now we have a full 6D phase-space density model ready for solving spherical Jeans equation.

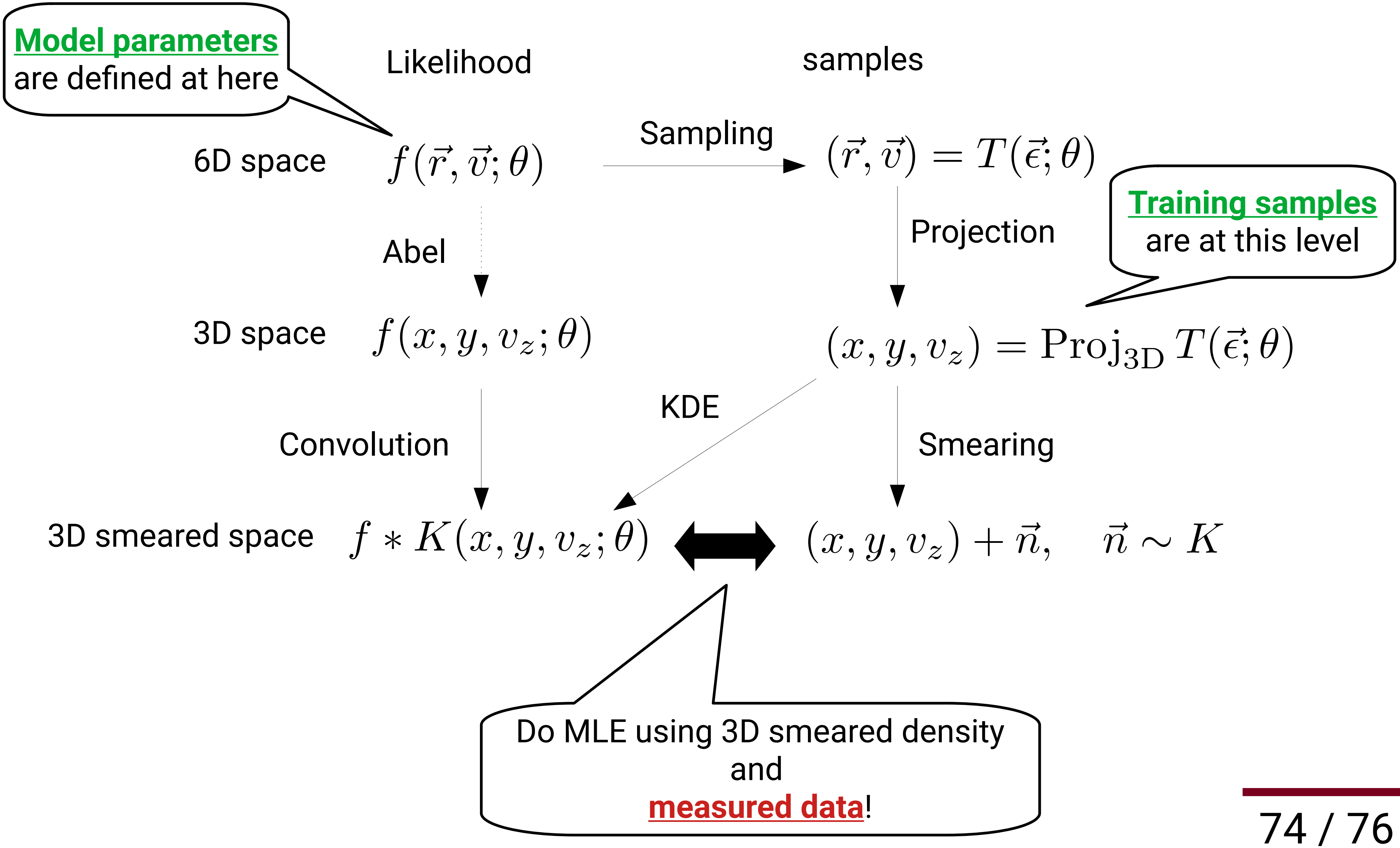
$$p(\vec{r}) = n(r;\theta)$$ modeled by equivariant CNF for cuspy halos

$$p(\vec{v}|\vec{r}) = \mathrm{GaussPDF}(\vec{v};\mu = 0, \Sigma(r;\theta))$$

$$f(\vec{r},\vec{v}) = p(\vec{r}) \times p(\vec{v}|\vec{r})$$

Wait, we only have x, y, vz.
How can we train this network by MLE?
We cannot use a conventional loss function.

How to train this model?



Model parameters are defined at here

$$\begin{tikzcd}
 & \text{Likelihood} & & \text{samples} \\
\text{6D space} & f(\vec{r},\vec{v};\theta) \arrow[rr, "\text{Sampling}"] \arrow[d, dotted, "\text{Abel}"'] & & (\vec{r},\vec{v}) = T(\vec{\epsilon};\theta) \arrow[d, "\text{Projection}"] \\
\text{3D space} & f(x,y,v_z;\theta) \arrow[d, "\text{Convolution}"'] & & (x,y,v_z) = \mathrm{Proj}_{3\mathrm{D}}\, T(\vec{\epsilon};\theta) \arrow[d, "\text{Smearing}"] \arrow[dll, "\text{KDE}"'] \\
\text{3D smeared space} & f * K(x,y,v_z;\theta) \arrow[rr, leftrightarrow, line width=2pt] & & (x,y,v_z) + \vec{n}, \quad \vec{n} \sim K
\end{tikzcd}$$

Training samples are at this level

Do MLE using 3D smeared density
and
measured data!

Loss Function for Modeling Dwarf Spheroidal Galaxy

- In order to train the normalizing flow with spherical symmetry using limited kinematic information, we minimize the following entropy:

$$\mathcal{L}(\theta) = \int d\vec{w}_\perp \; p * K_h(\vec{w}_\perp) \; \log \hat{p} * K_h(\vec{w}_\perp; \theta)$$

- Importance sampling: N_T training sample (stars) ~ p, N_K noise samples ~ K_h

$$\mathcal{L}(\theta) = \frac{1}{N N_K} \sum_{a=1}^{N} \sum_{b=1}^{N_K} \log \hat{p} * K_h(\vec{w}_\perp^{(a)} + \vec{\epsilon}^{(b)}; \theta)$$

- KDE for the smeared likelihood model:
  N_G generated stars from the normalizing flows~ \hat{p}

$$\mathcal{L}(\theta) = \frac{1}{N N_K} \sum_{a=1}^{N} \sum_{b=1}^{N_K} \log \frac{1}{N_G} \sum_{c=1}^{N_G} K_h \left[\vec{w}_\perp^{(a)} + \vec{\epsilon}^{(b)} - \vec{T}(\vec{z}^{(c)}; \theta) \right]$$