Understanding Galactic Dark Matter with Neural Networks

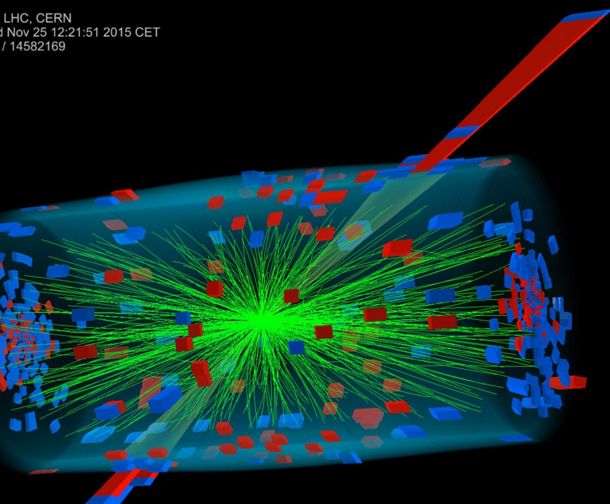
2025년
한국고에너지물리학회
및 핵-입자-천체분과
연합 워크샵
Chonnam National University, Gwangju, Korea

Nov. 2025

Sung Hak Lim
CTPU-PTC, IBS

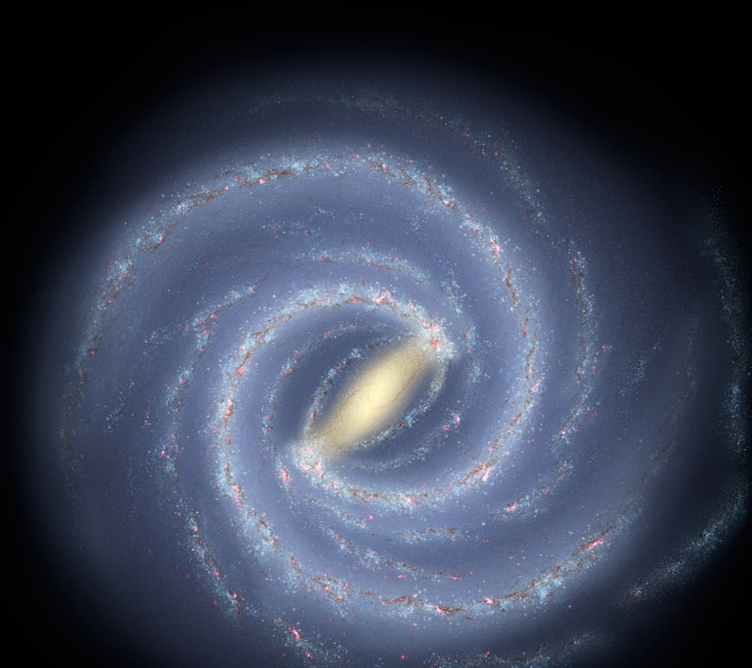
Institute for Basic Science

Phenomenology of Dark Matter appears in Various Length Scales!



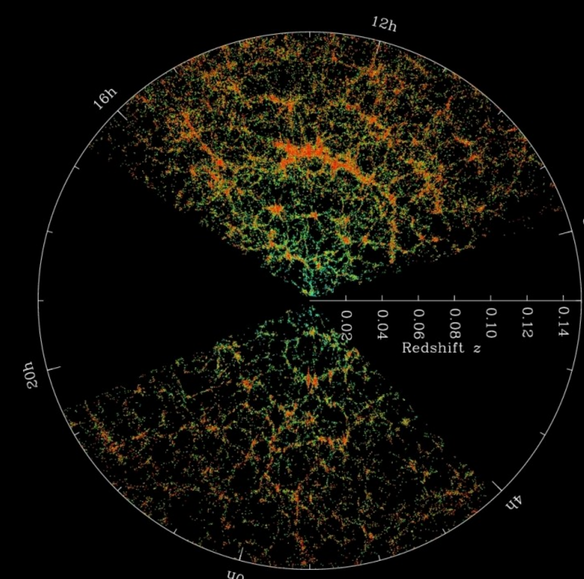
Collider Physics

(small length scale)



Galactic Astrophysics

(medium length scale?)



Cosmology

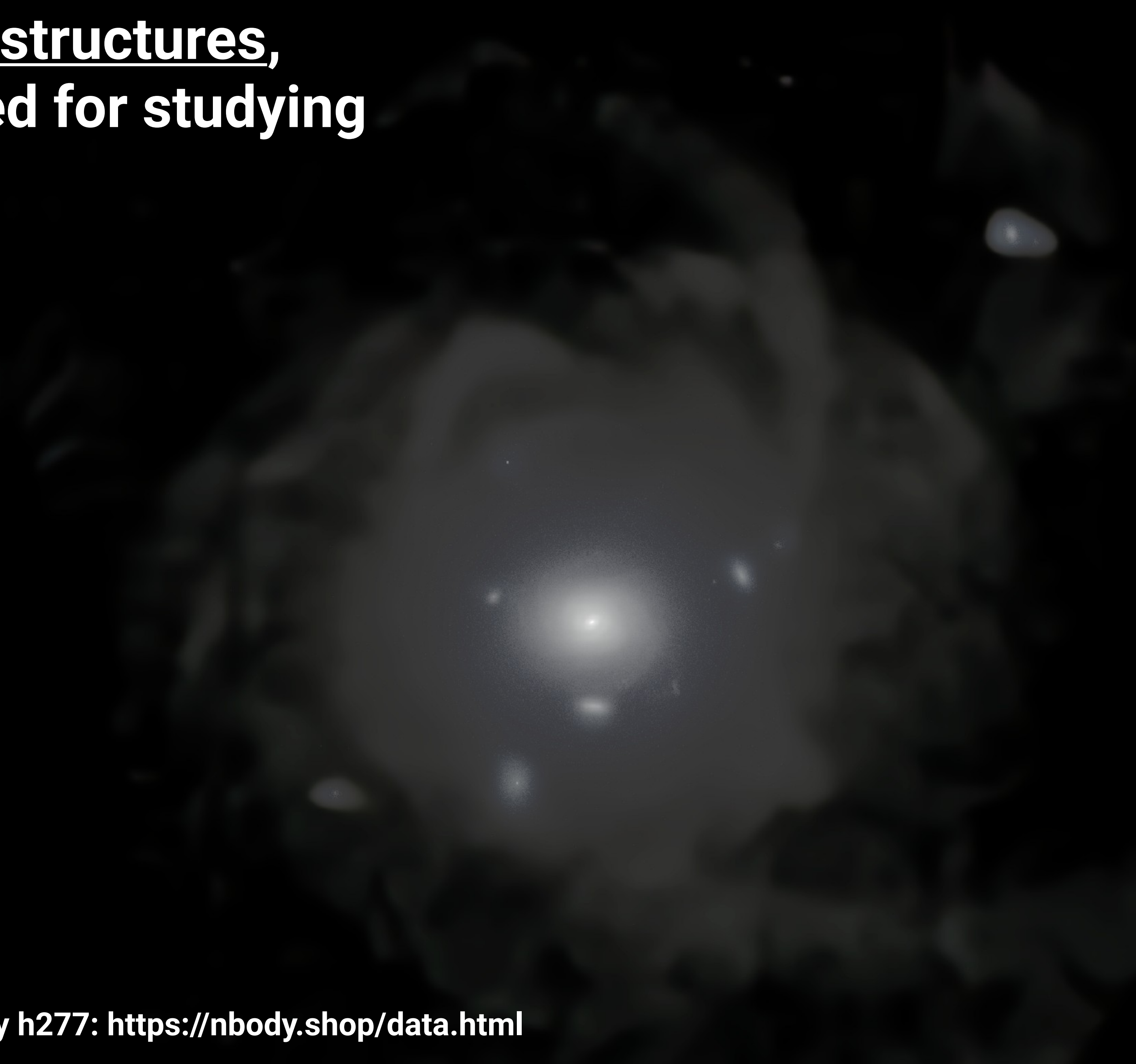
(large length scale)

But why <u>galactic astrophysics</u> is one of the astrophysics field may draw attention to <u>high energy particle physicists?</u>

But why galactic astrophysics is one of the astrophysics field may draw attention to high energy particle physicists?

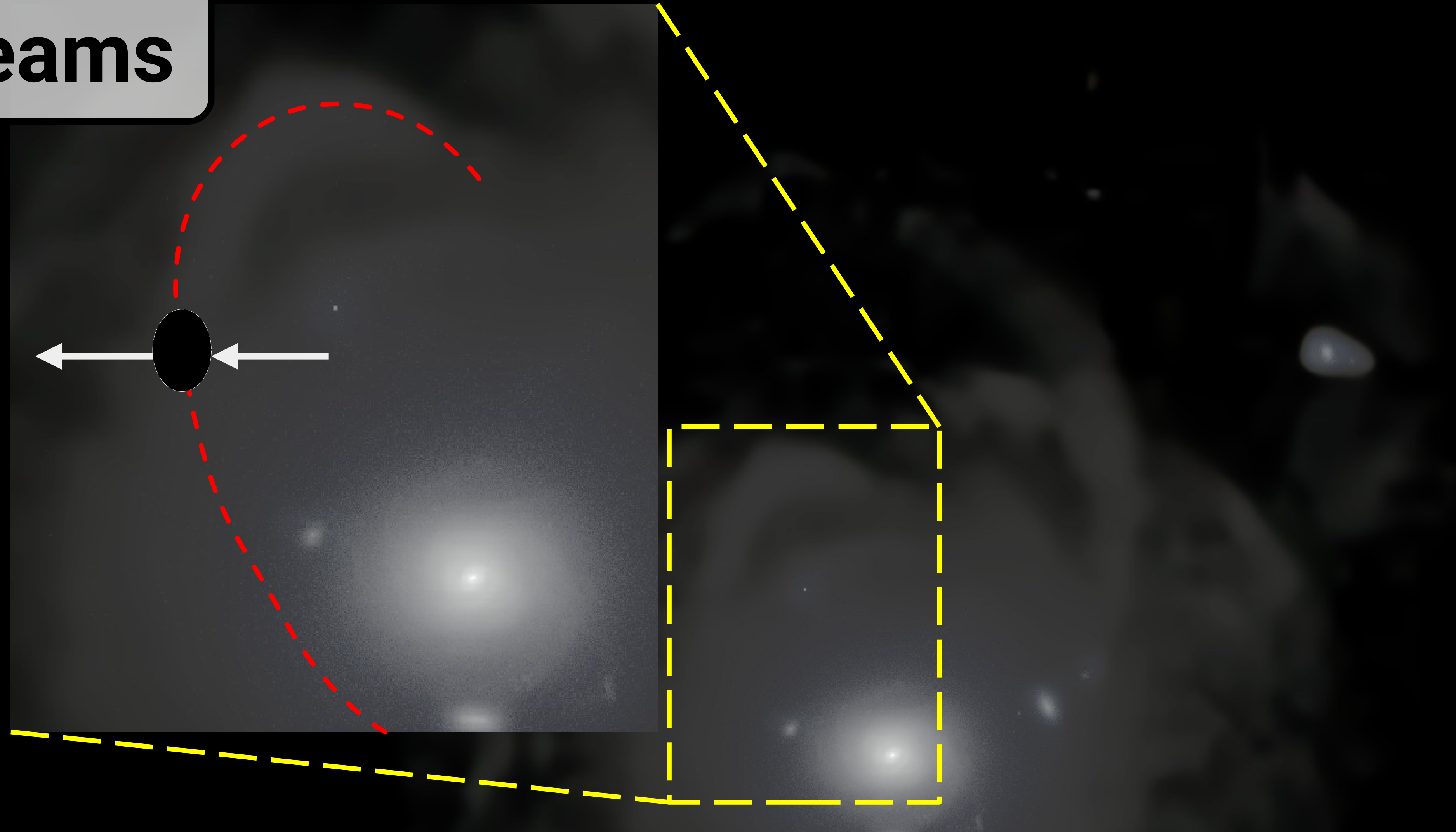
Because it provides us measurement and insight about dark matter.

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



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

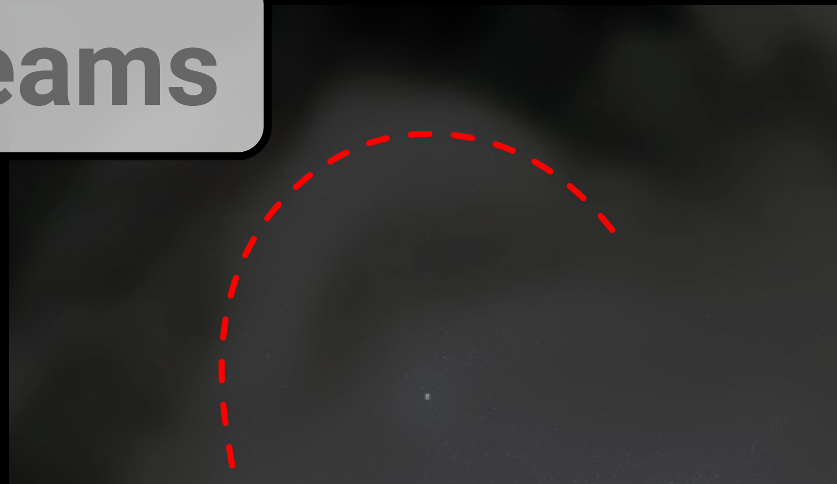
Streams

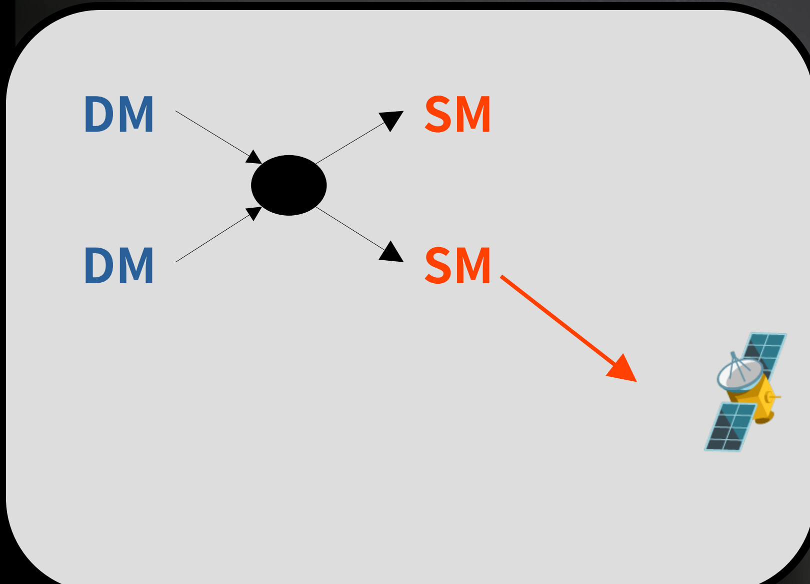


Stars on the same orbit, originating from globular clusters or dwarf galaxies tidally stripped

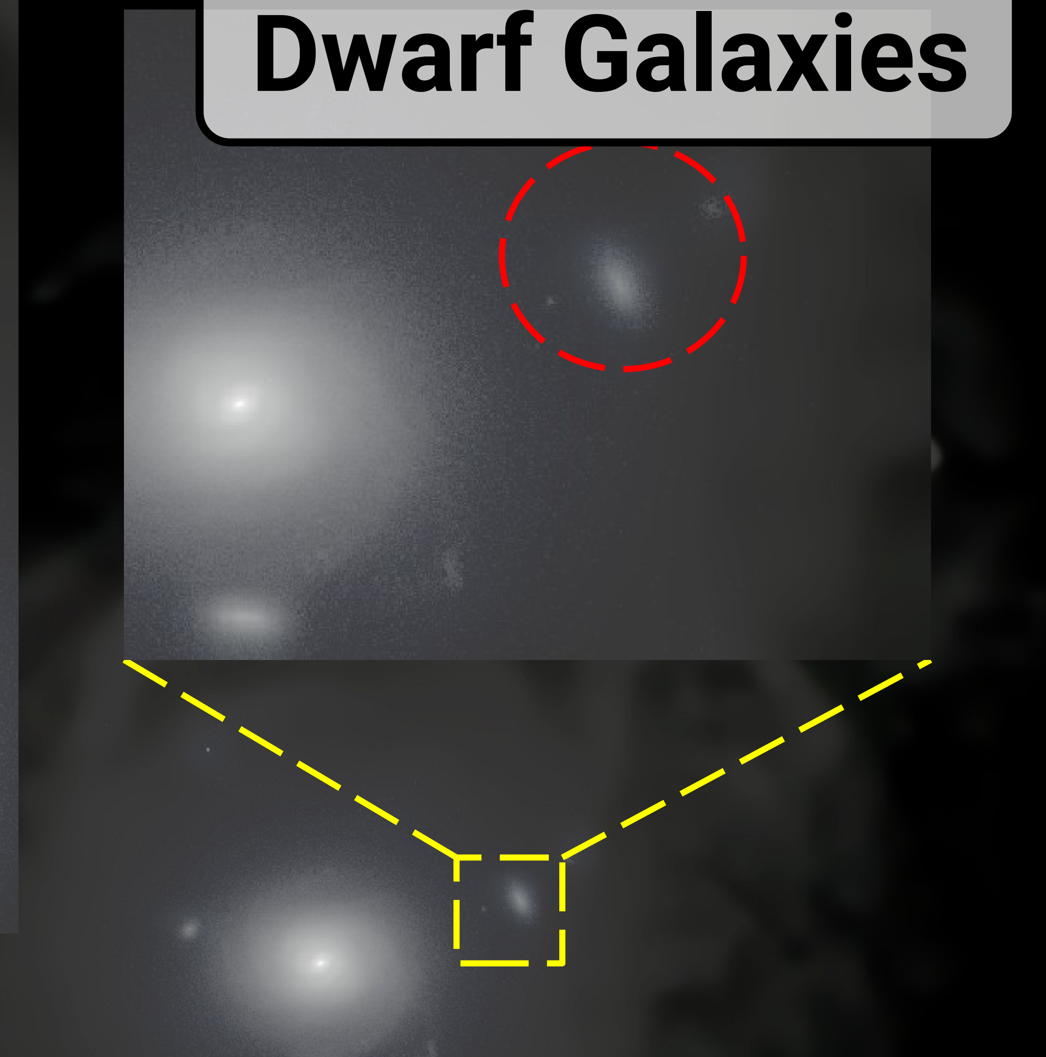
- Orbit is visible → measuring gravity
- Any gaps in streams may be a sign of collision with dark matter subhalo!

Streams

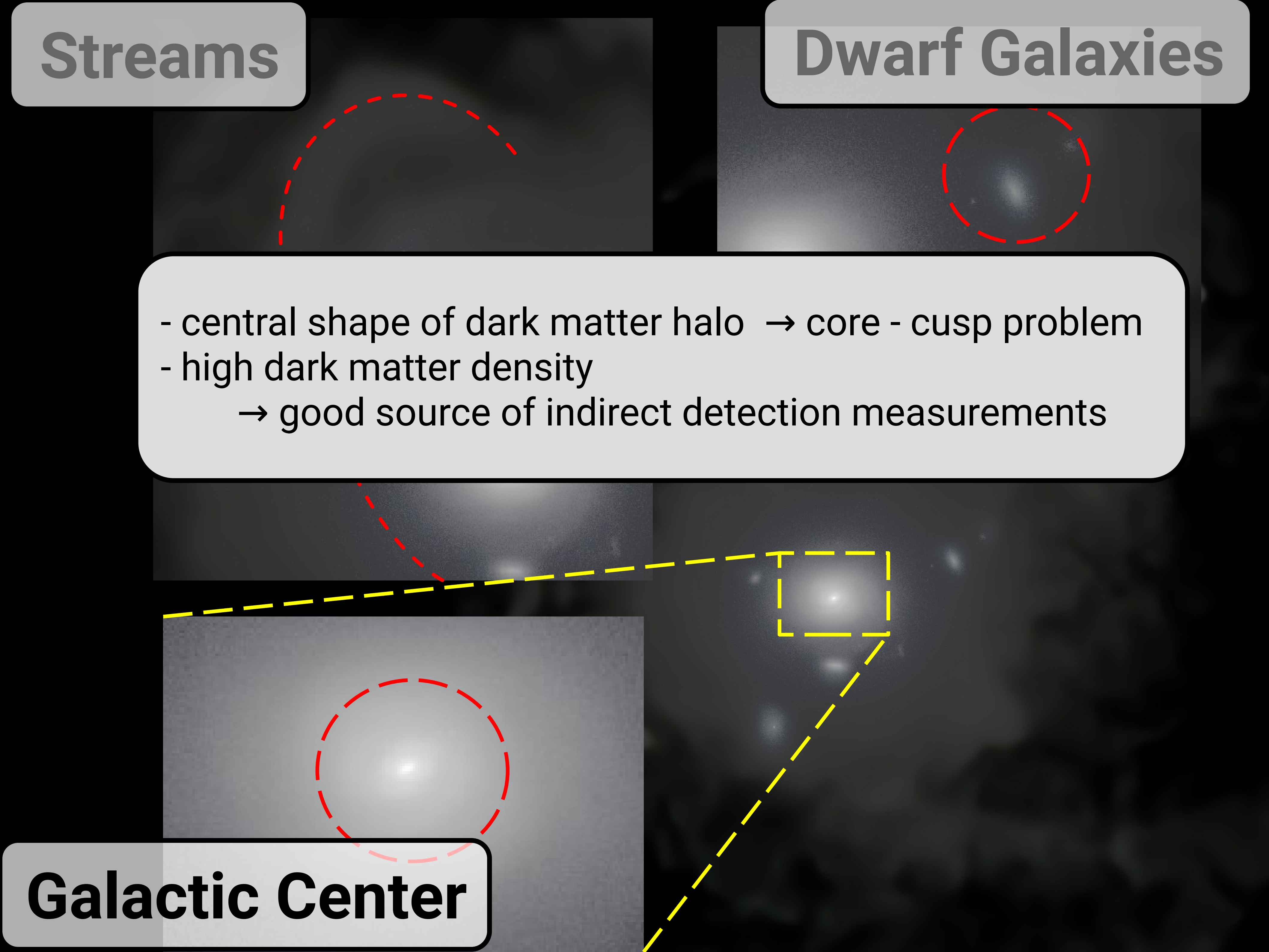




Dwarf Galaxies



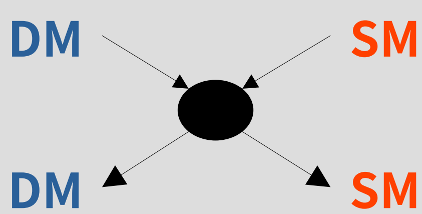
A round and faint satellite galaxy orbiting the galaxy.
- whole structure is clearly visible → core - cusp problem
- less baryonic activity
    → good source of indirect detection measurements

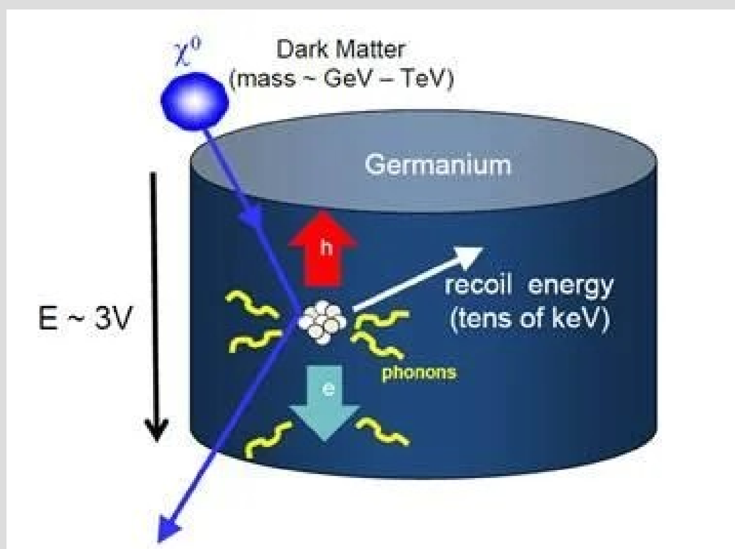

Streams
Dwarf Galaxies
- central shape of dark matter halo → core - cusp problem
- high dark matter density
→ good source of indirect detection measurements
Galactic Center

Streams

Dwarf Galaxies

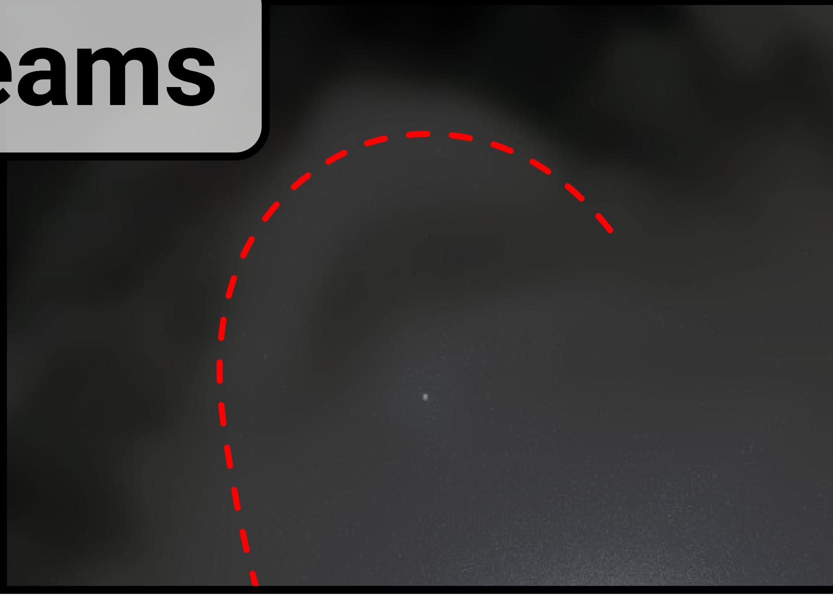
- nearby stars → highly precise and complete measurement
  → high quality measurement of local dark matter density
- measruing dark matter density around solar system
  → relevant for direct detection of dark matter



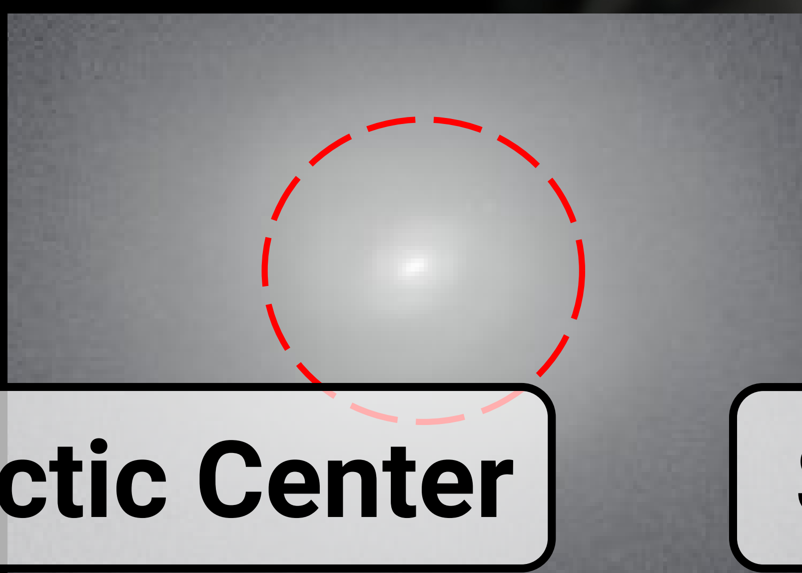


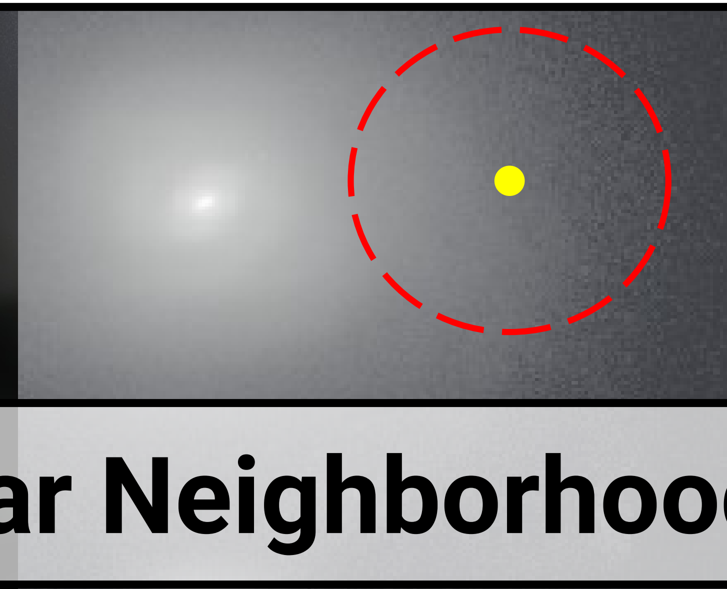
Galactic Center

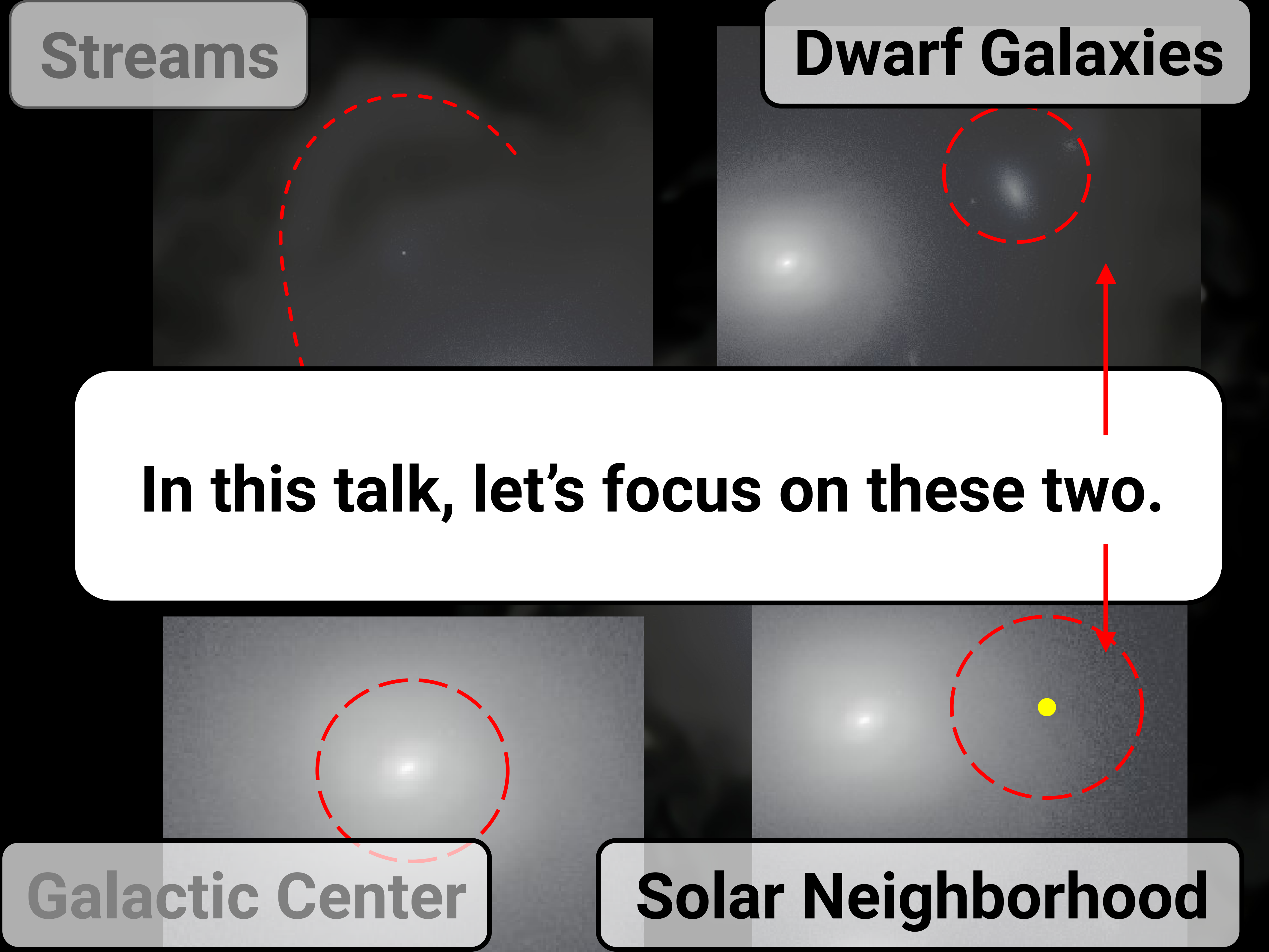
Solar Neighborhood

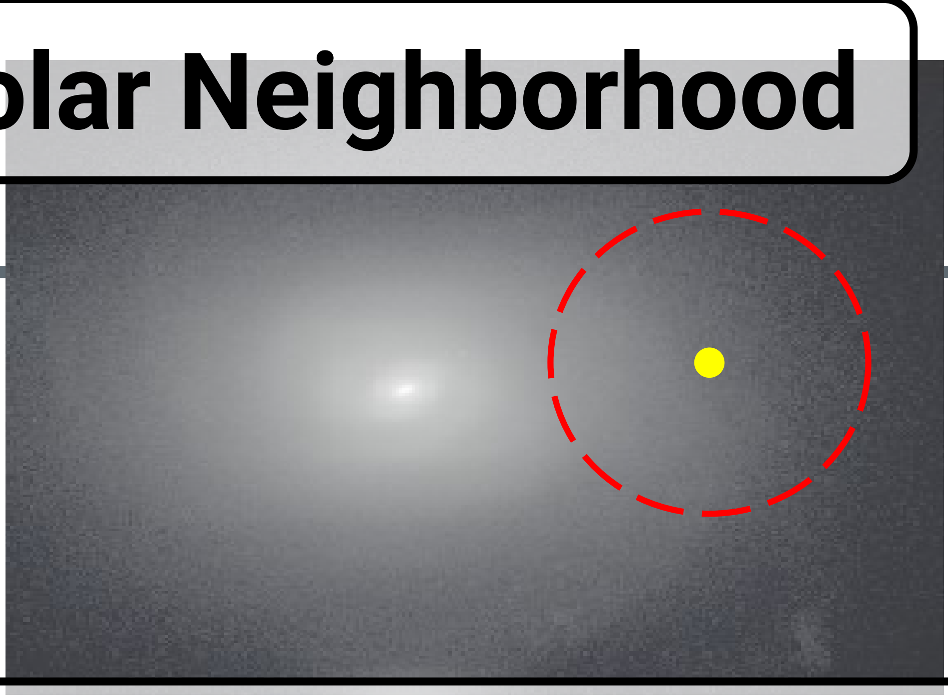


We can study the details of galactic dark matter by analyzing these galactic substructures!






Streams
Dwarf Galaxies
In this talk, let's focus on these two.
Galactic Center
Solar Neighborhood



Talk title:

Understanding Galactic Dark Matter with Neural Networks

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

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

Galactic Dynamics

Equation of motion: Boltzmann Equation

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

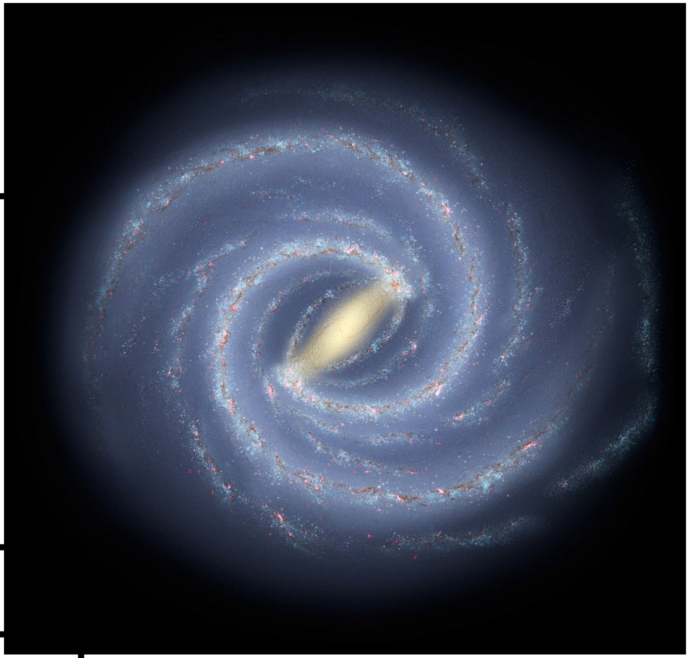
If we consider a galaxy as a hydrodynamic system $N \to \infty$ consisting of stars, phase-space density of a star (probability of finding a star with given position and velocity) describes the system.

If **phase-space density** is determined..

We can estimate the **gravitational acceleration field**!

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!

AMERICAN Scientist

Astrophysics Data Goes Big

BY JOEL R. PRIMACK

ASTRONOMY

Astrophysics is becoming a data science!

SPACEREF

All News | Earth | New Space & Tech | Space Commerce | Space Stations | Science & Exploration | Press Releases

SCIENCE AND EXPLORATION

Talks at Google: Data-Driven Discovery: Astronomy in the Era of Large Surveys

By Marc Boucher | Status Report | February 15, 2015
Filed under Astronomy, Big Data, Google, John Bochanski

Large, digital surveys of the night sky have revolutionized how astronomy is done. Astronomers are no longer tethered to their telescopes, but instead have access to terabytes of data on hundreds of millions of stars and galaxies through the internet.

John Bochanski, an assistant professor of Physics at Rider University, has used survey data over the last decade to study our home Galaxy, the Milky Way, and millions of its stellar constituents.

AMERICAN Scientist

Astrophysics Data Goes Big

BY JOEL R. PRIMACK

ASTRONOMY

Large Surveys
(**Gaia**, LSST, JWST, DESI …)
(N-body simulations…)

Public datasets

Machine Learning:
robust tool for
data driven discovery!

You can discover new things
(for example, dark matter)
by yourself!

SPACEREF★

All News | Earth | New Space & Tech | Space Commerce | Space Stations | Science & Exploration | Press Releases

SCIENCE AND EXPLORATION

Talks at Google: Data-Driven Discovery: Astronomy in the Era of Large Surveys

By Marc Boucher | Status Report | February 15, 2015
Filed under Astronomy, Big Data, Google, John Bochanski

Large, digital surveys of the night sky have revolutionized how astronomy is done. Astronomers are no longer tethered to their telescopes, but instead have access to terabytes of data on hundreds of millions of stars and galaxies through the internet.

John Bochanski, an assistant professor of Physics at Rider University, has used survey data over the last decade to study our home Galaxy, the Milky Way, and millions of its stellar constituents.

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

Galactic Dynamics

<u>Neural Network</u>

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:

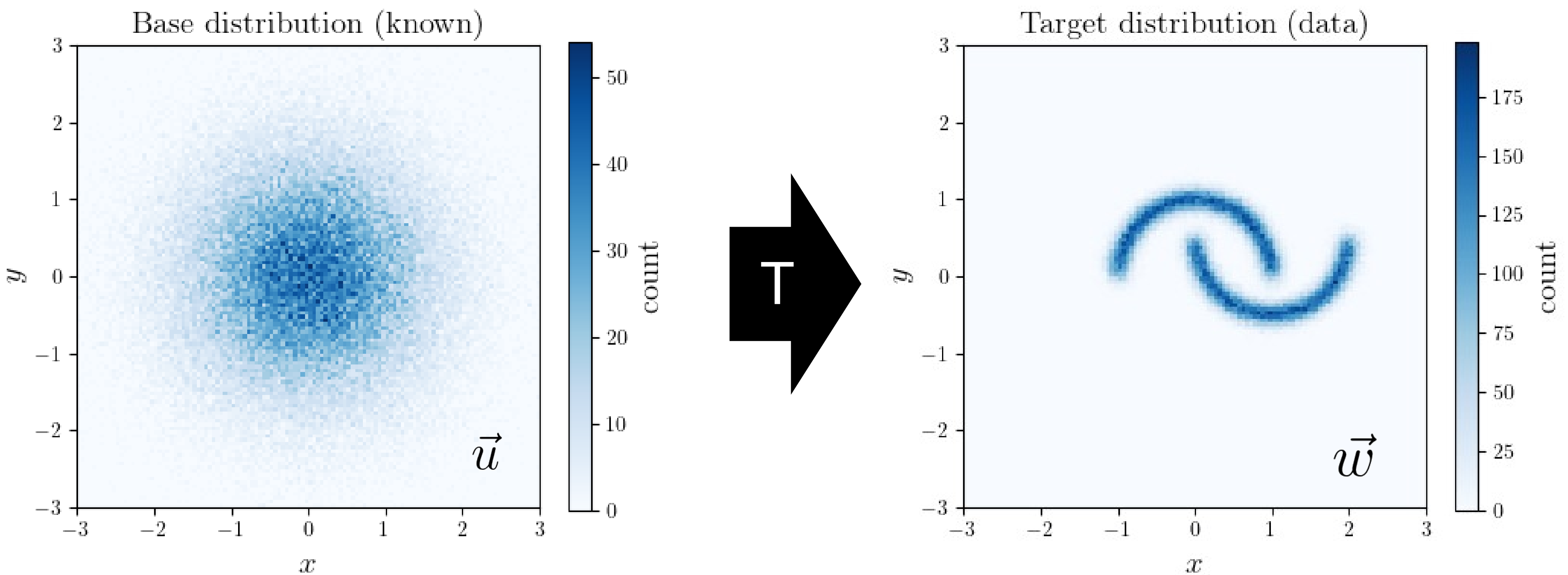
Neural Networks for
<u>**arbitrary**</u> <u>**density estimation**</u>:
- Normalizing Flows

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

This is a fully
<u>**model-independent**</u>
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:

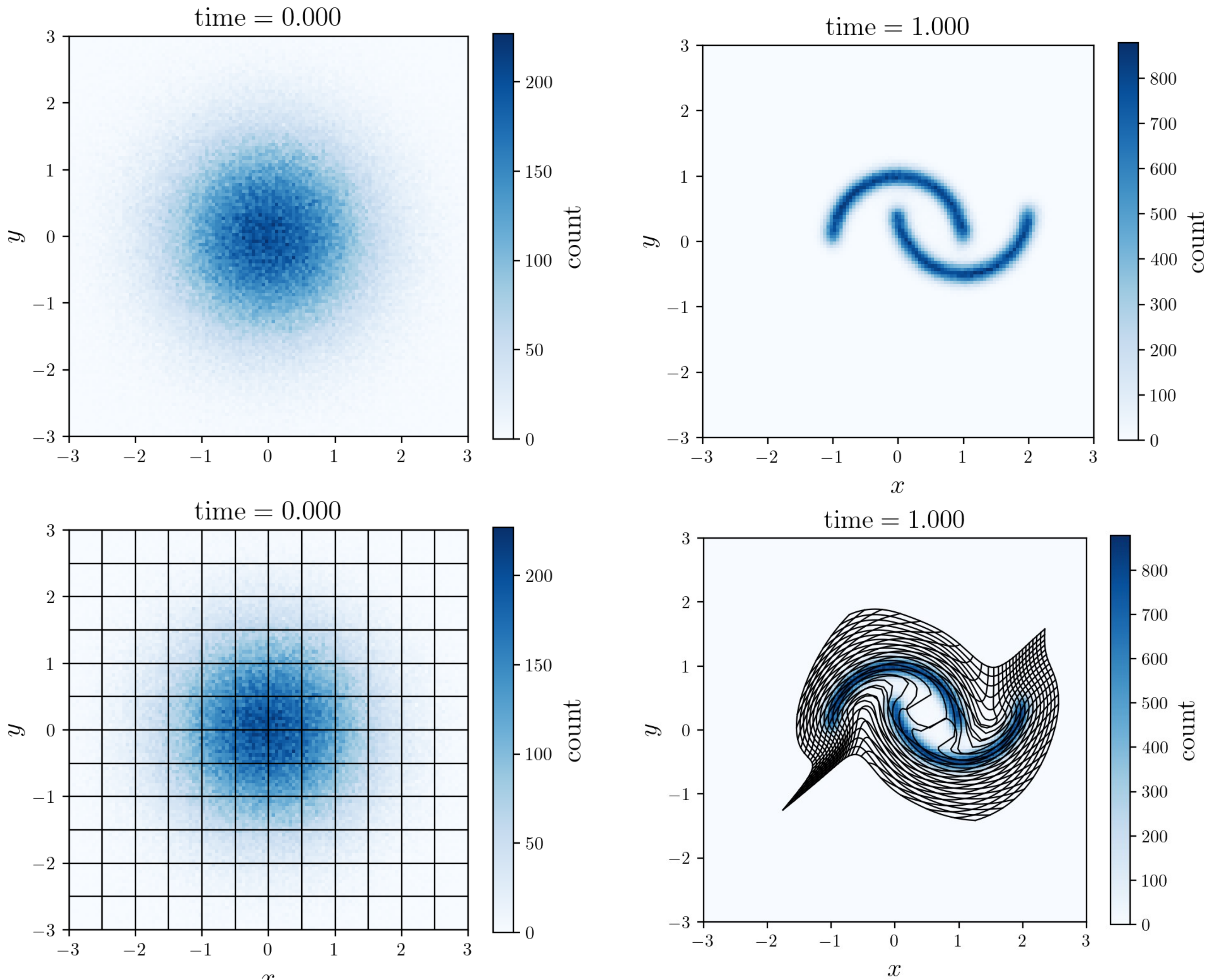
$$p_W(\vec{w}) = p_U(\vec{u}) \cdot \left| \frac{d\vec{u}}{d\vec{w}} \right| \qquad \vec{w} = T(\vec{u})$$

Neural Network

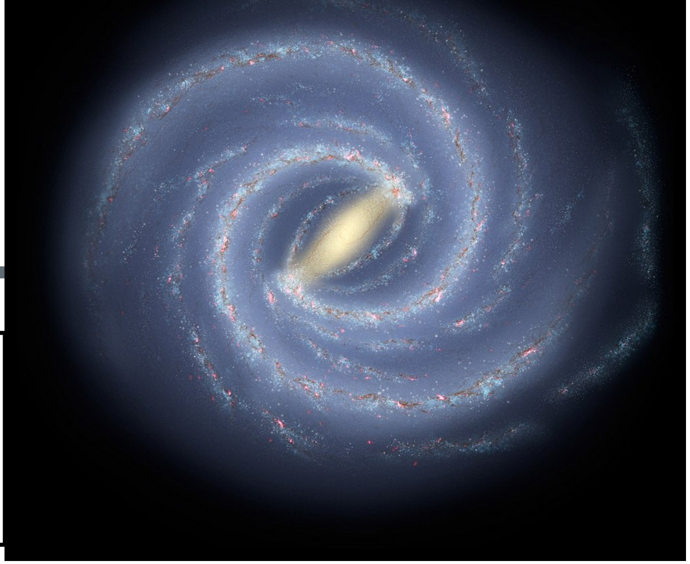
This formula can be used for training normalizing flows, too:

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
$\rho(\vec{x})$

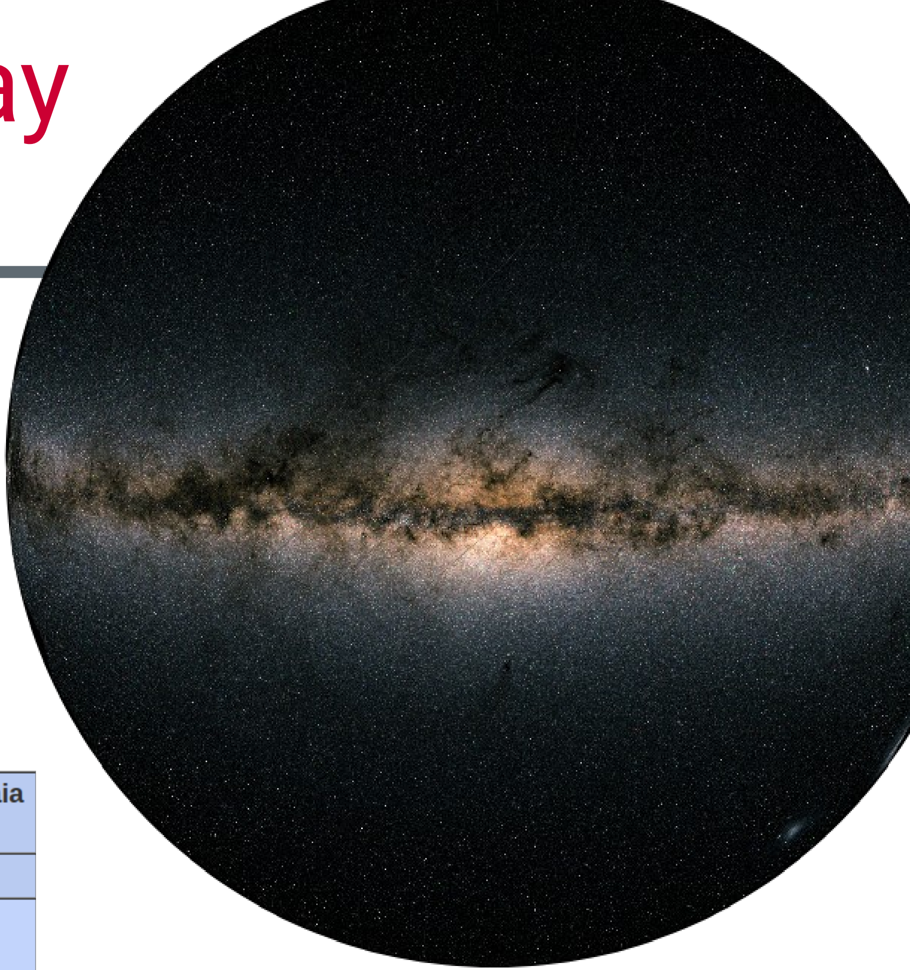
Solving Gauss's Equation

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

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

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.



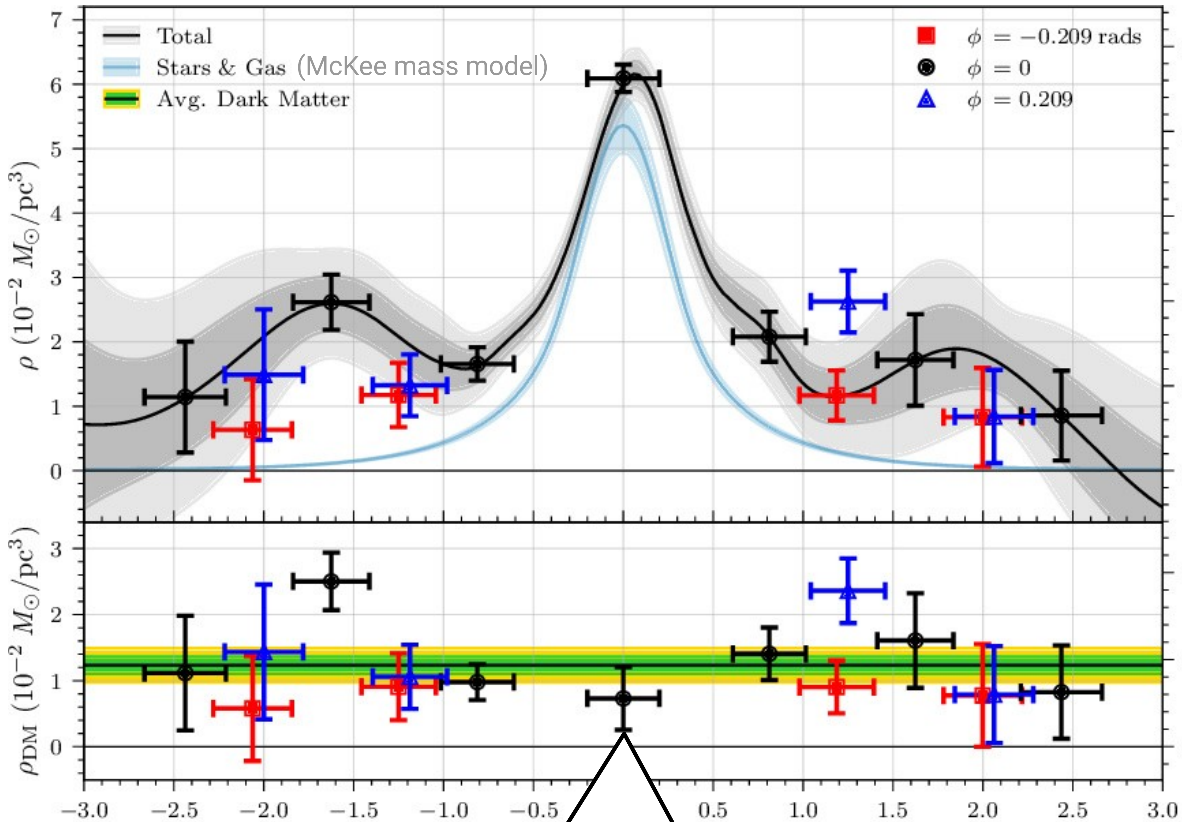
	# sources in Gaia DR3	# sources in Gaia DR2
Total number of sources	**1,811,709,771**	**1,692,919,135**
	Gaia Early Data Release 3	
Number of sources with full astrometry	1,467,744,818	1,331,909,727
Number of 5-parameter sources	585,416,709	
Number of 6-parameter sources	882,328,109	
Number of 2-parameter 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

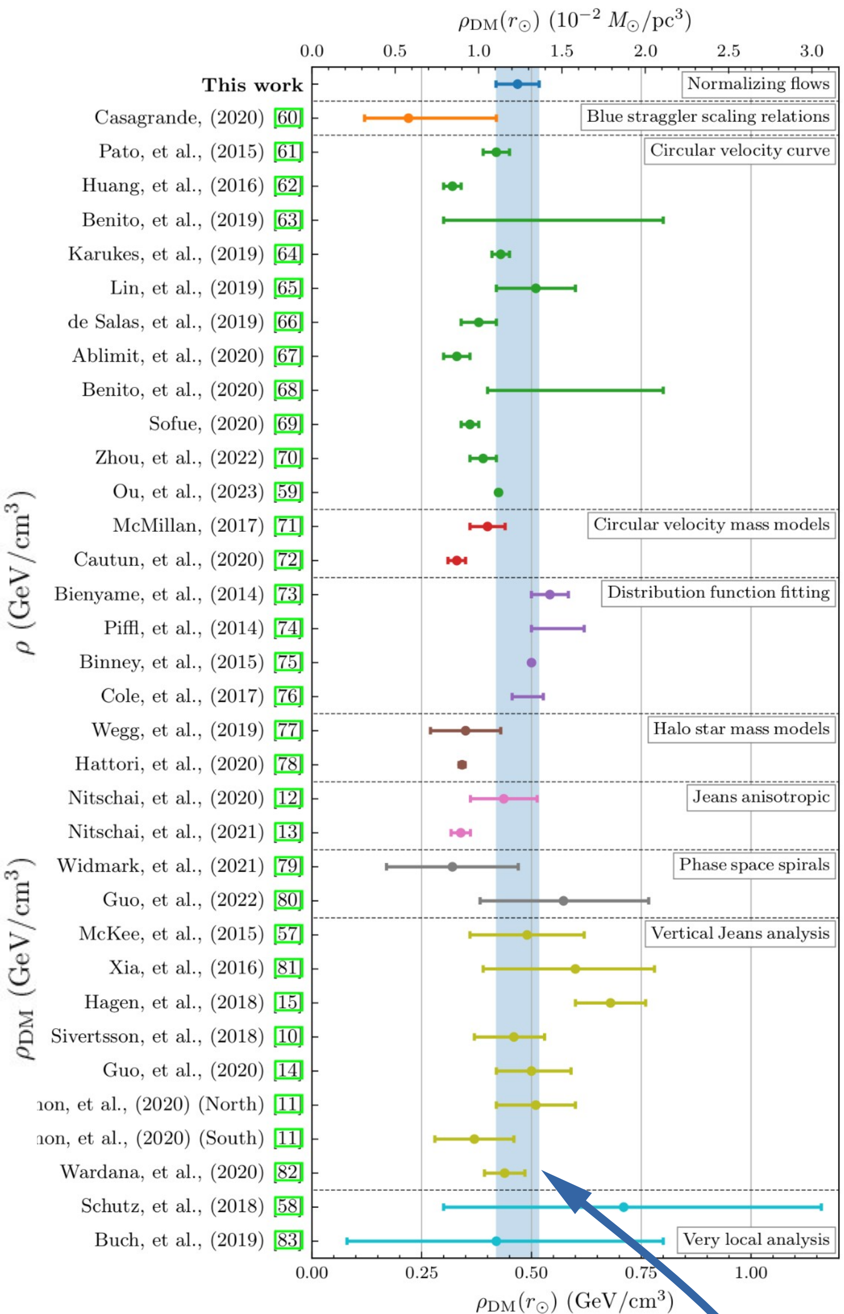
We could use this dataset to understand structure of the **Milky Way** very precisely.

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

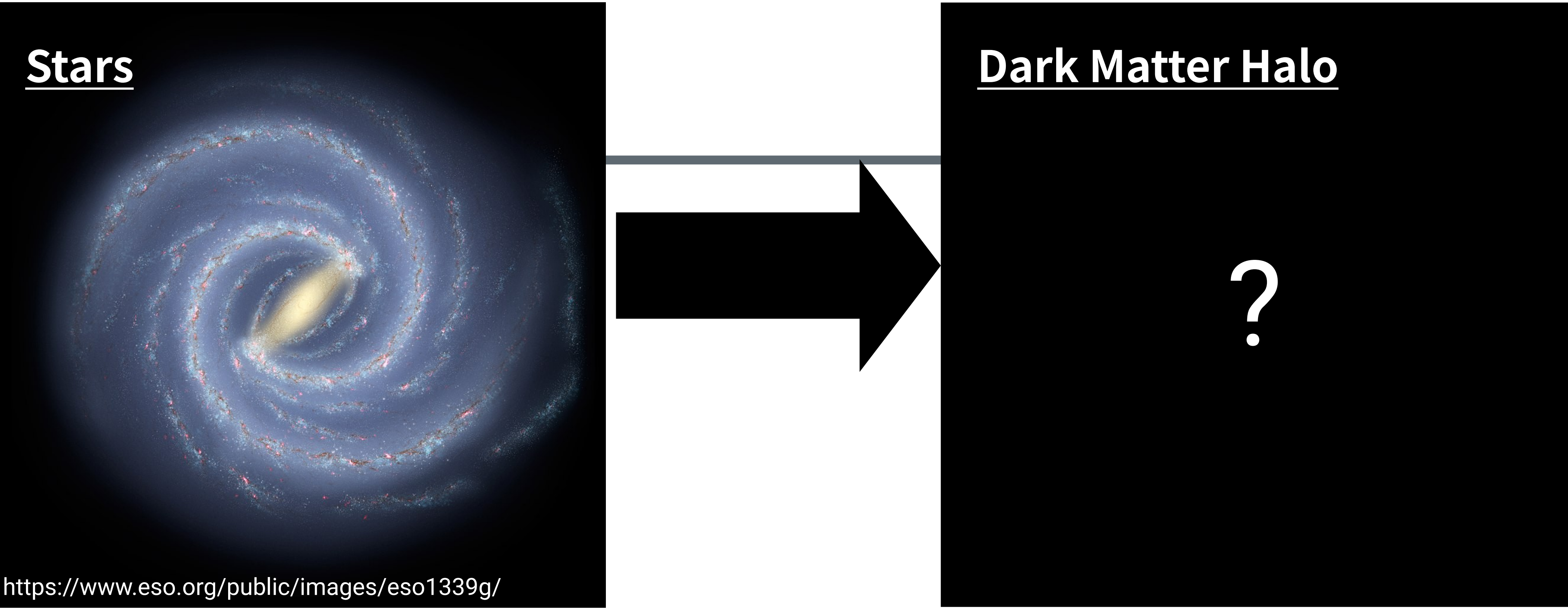
Local DM Mass Density of the Milky Way



0.32±0.18 GeV/cm³

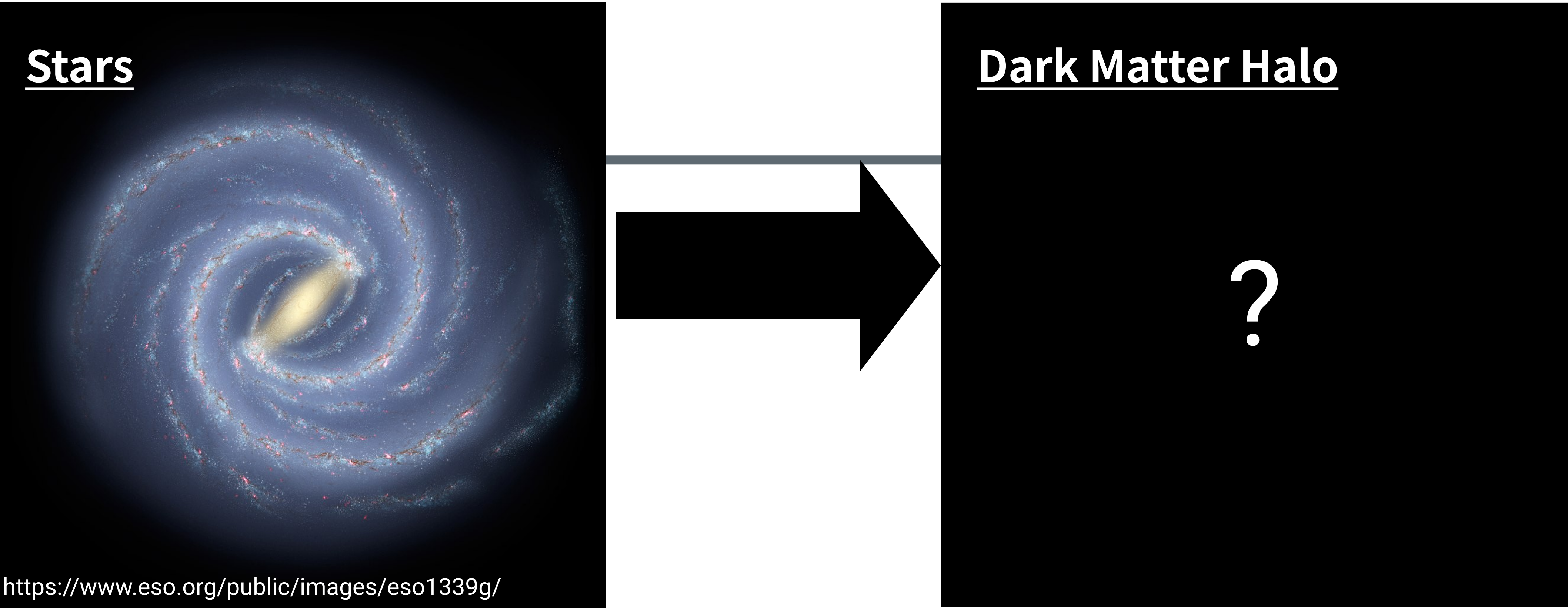


Taking the average of the DM mass density at the Solar radius, we find a local dark matter density: **0.47±0.05 GeV/cm³**



We have a model-independent (unsupervised) ML method to estimate **dark matter density** given **stellar distribution of a galaxy**.

THE END?



We have a model-independent (unsupervised) ML method to estimate **dark matter density** given **stellar distribution of a galaxy**.

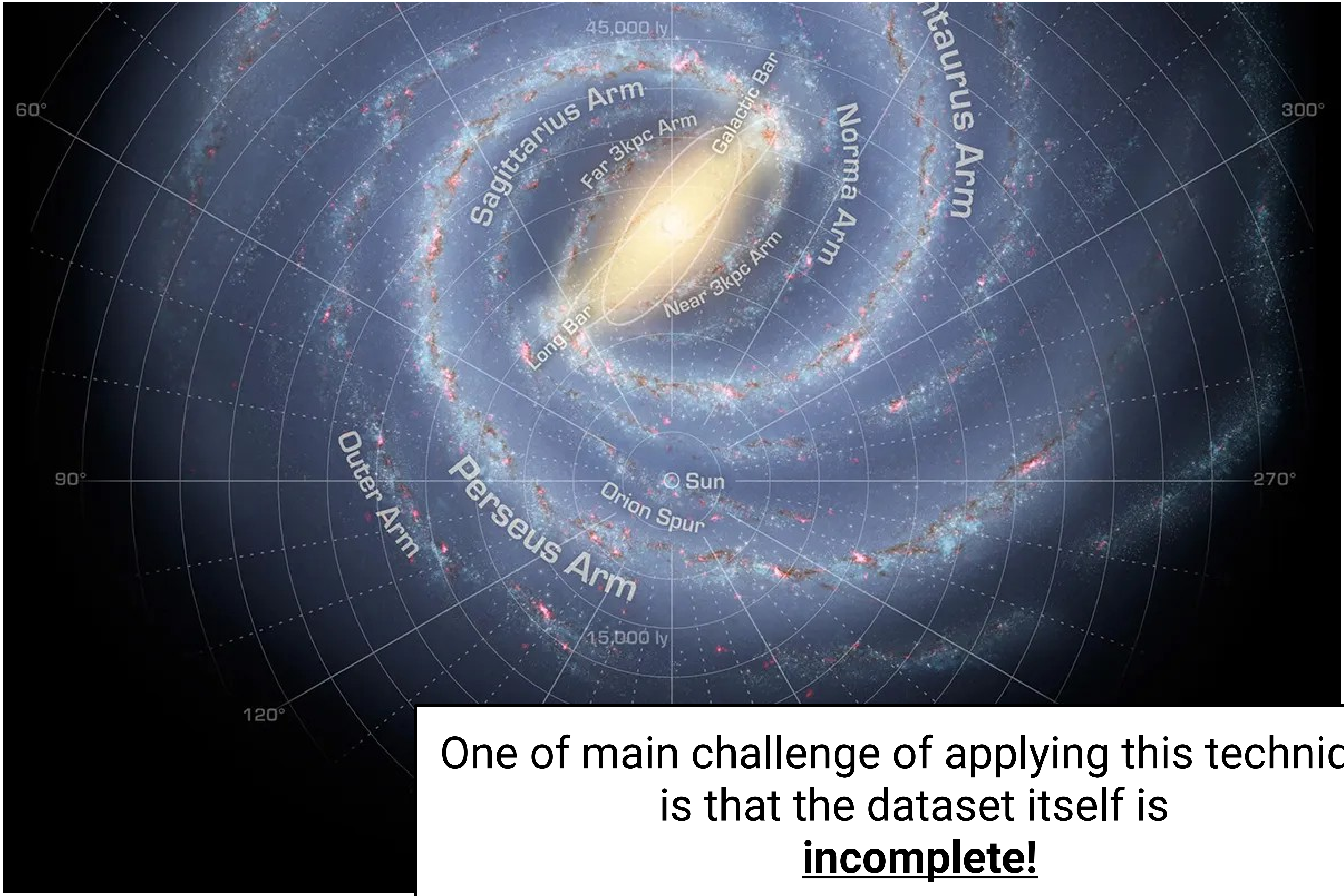
THE END? → Of course not!

Simulated data analysis

Real but clean data analysis

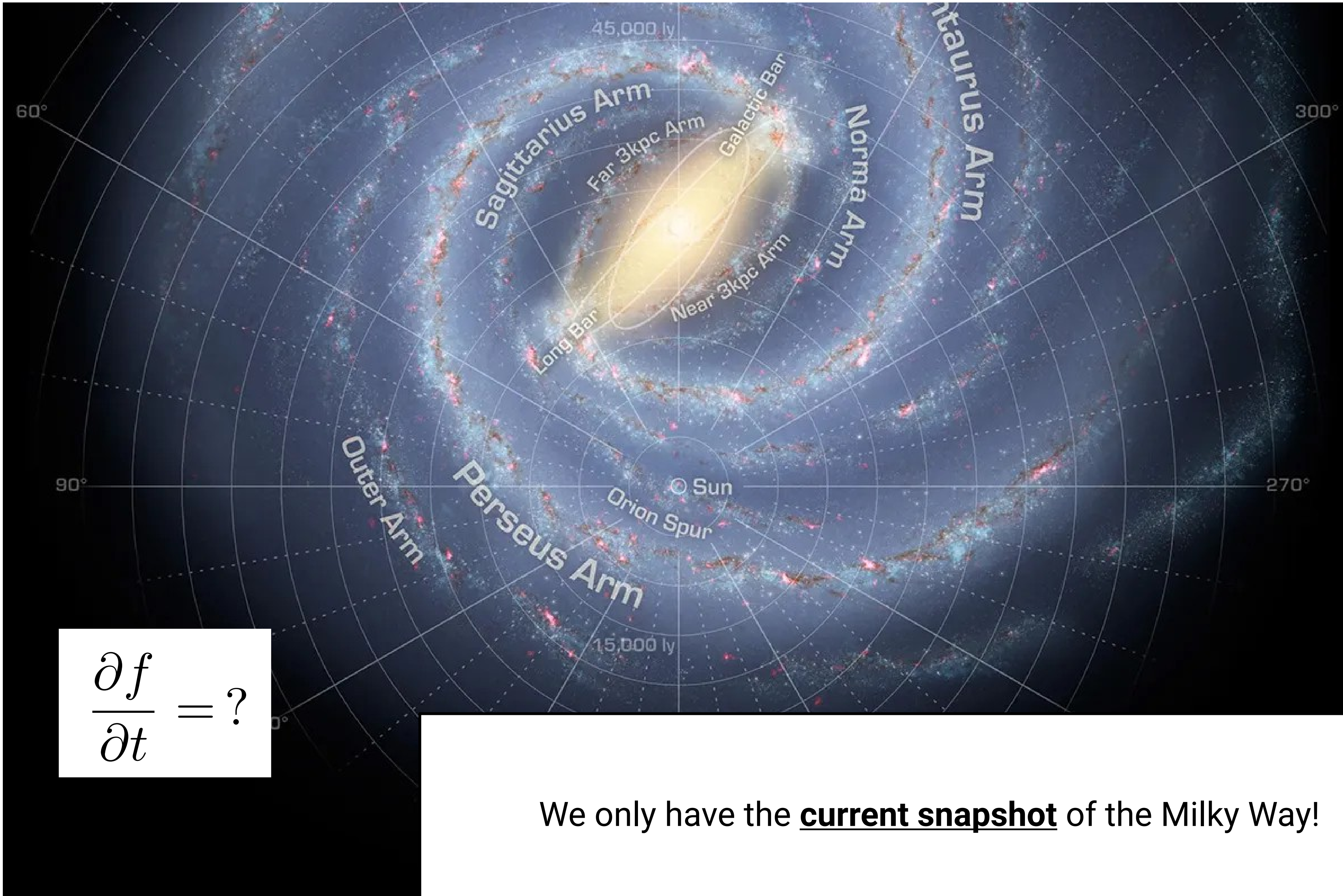
Real dirty data analysis

Galactic Dynamics and Incomplete Datasets



One of main challenge of applying this technique is that the dataset itself is **incomplete!**

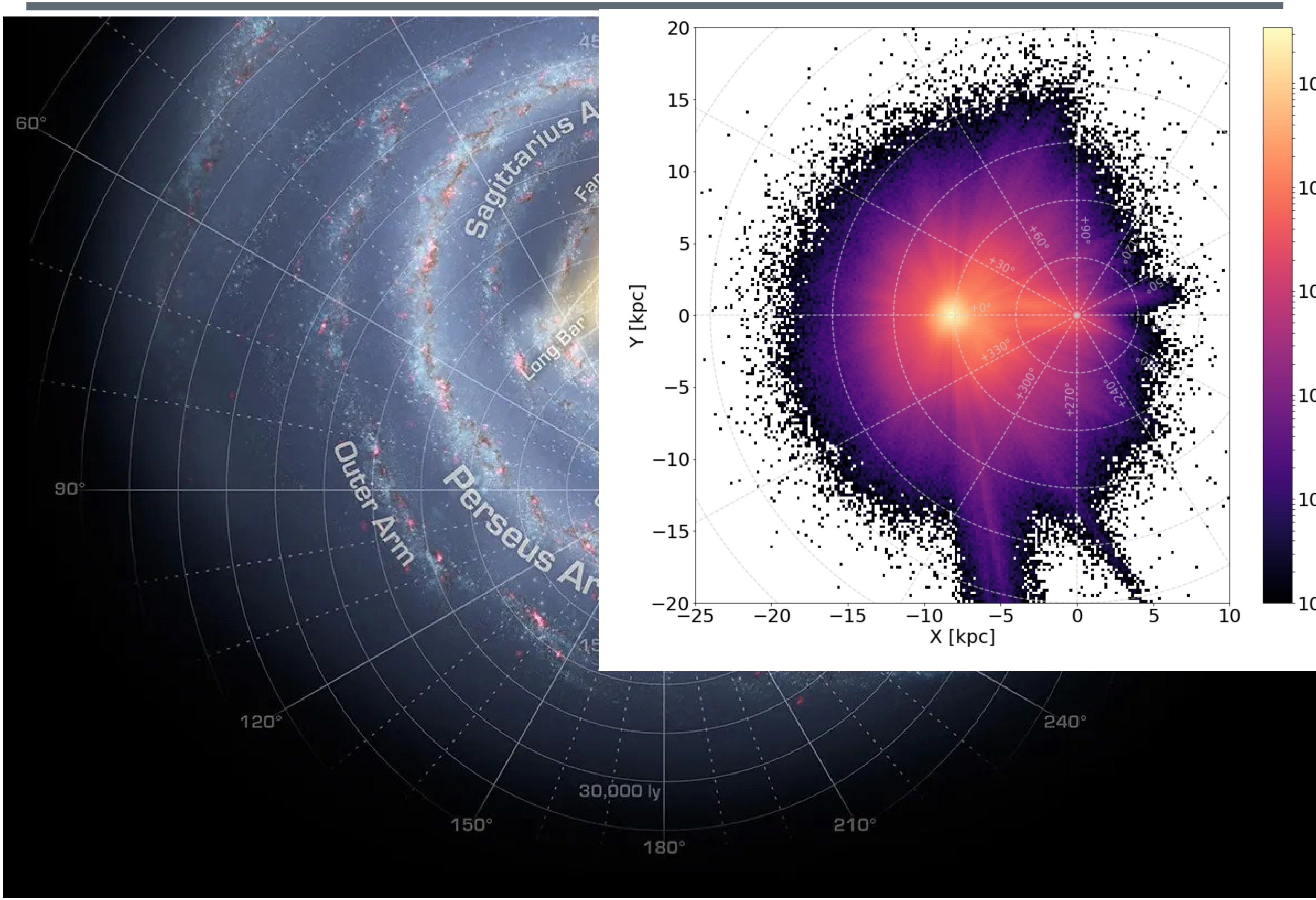
No time derivative information



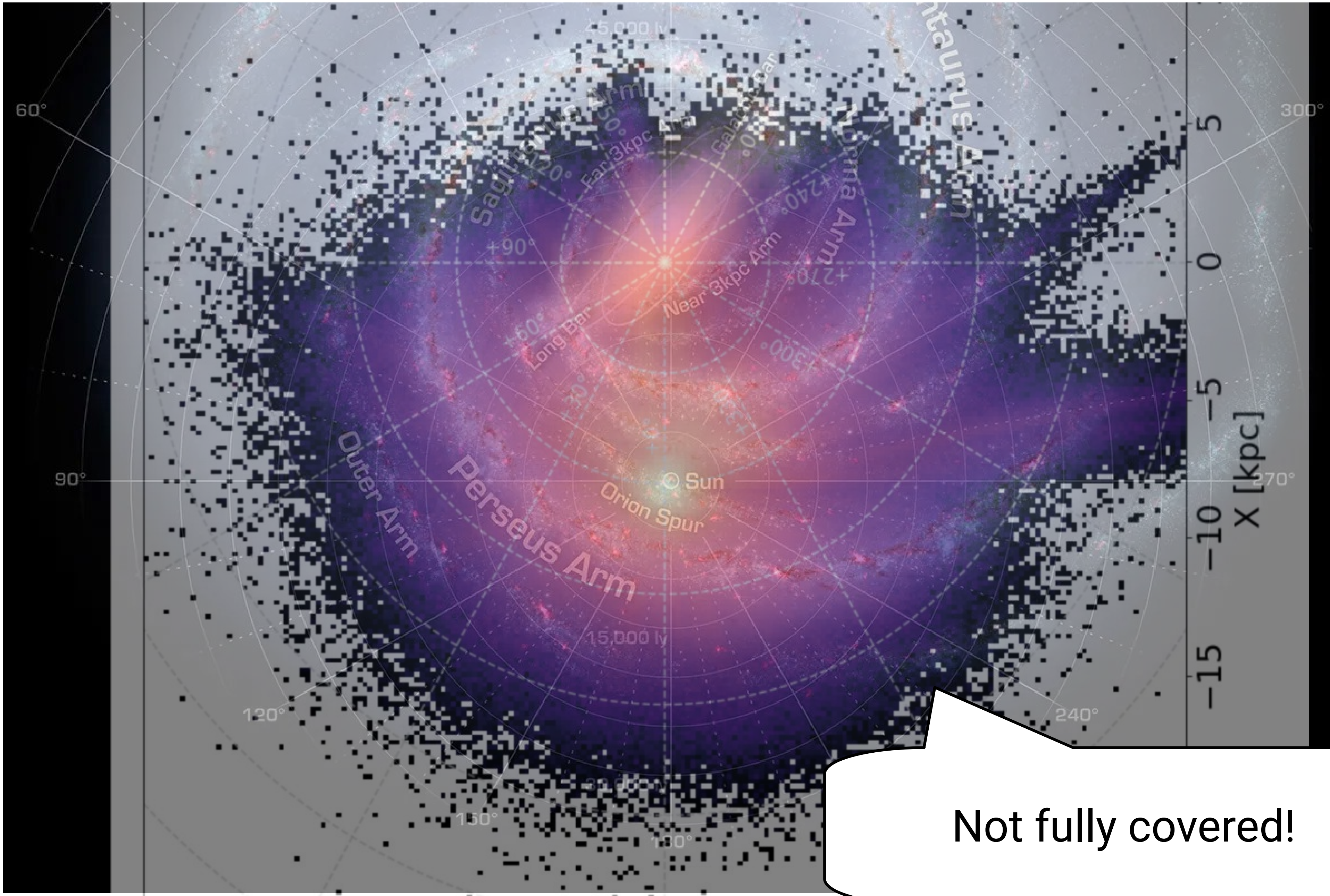
$$\frac{\partial f}{\partial t} = ?$$

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

Radial Velocity Distribution of Gaia DR3

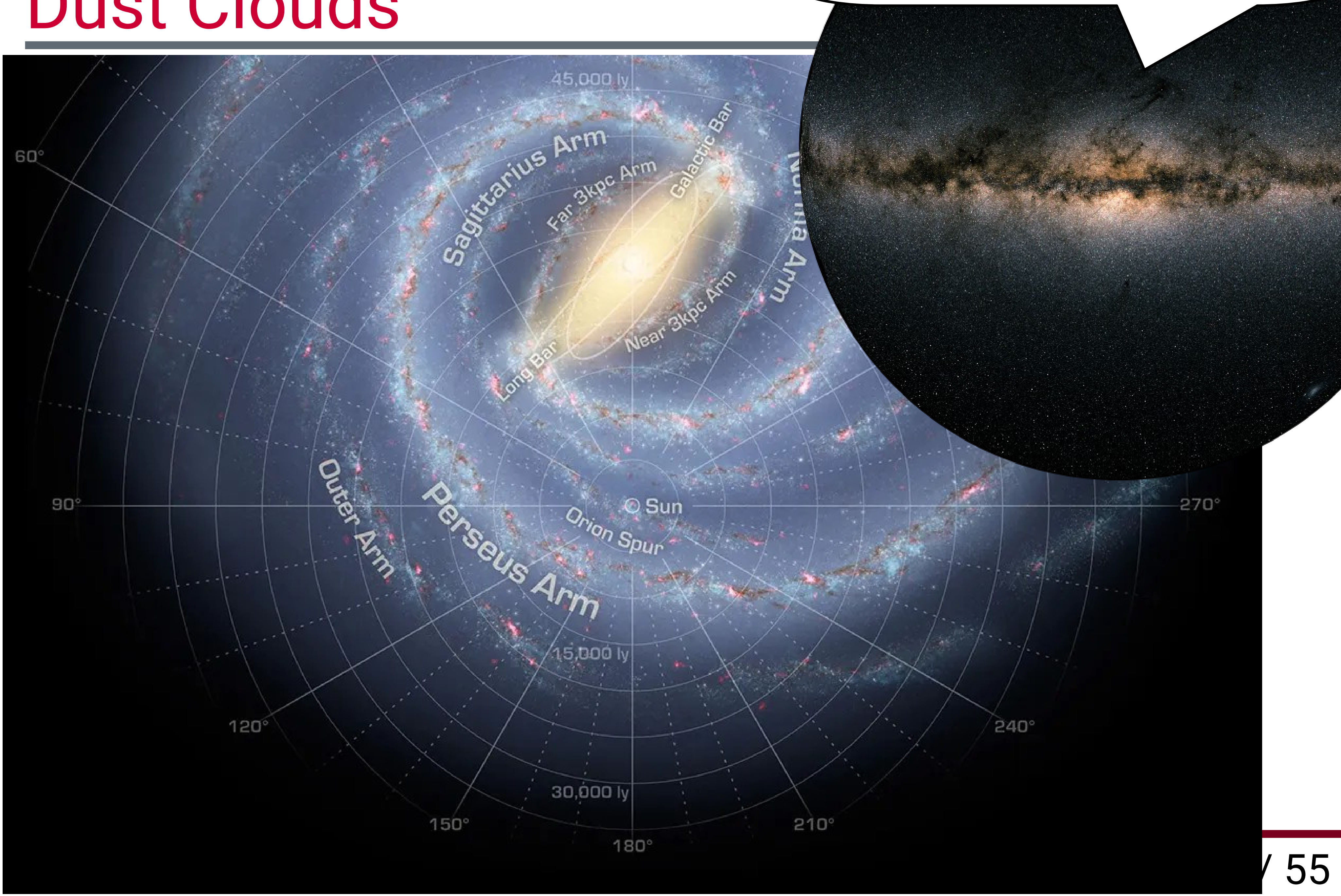


Incompleteness in Spacial Coverage



Not fully covered!

Dust Clouds



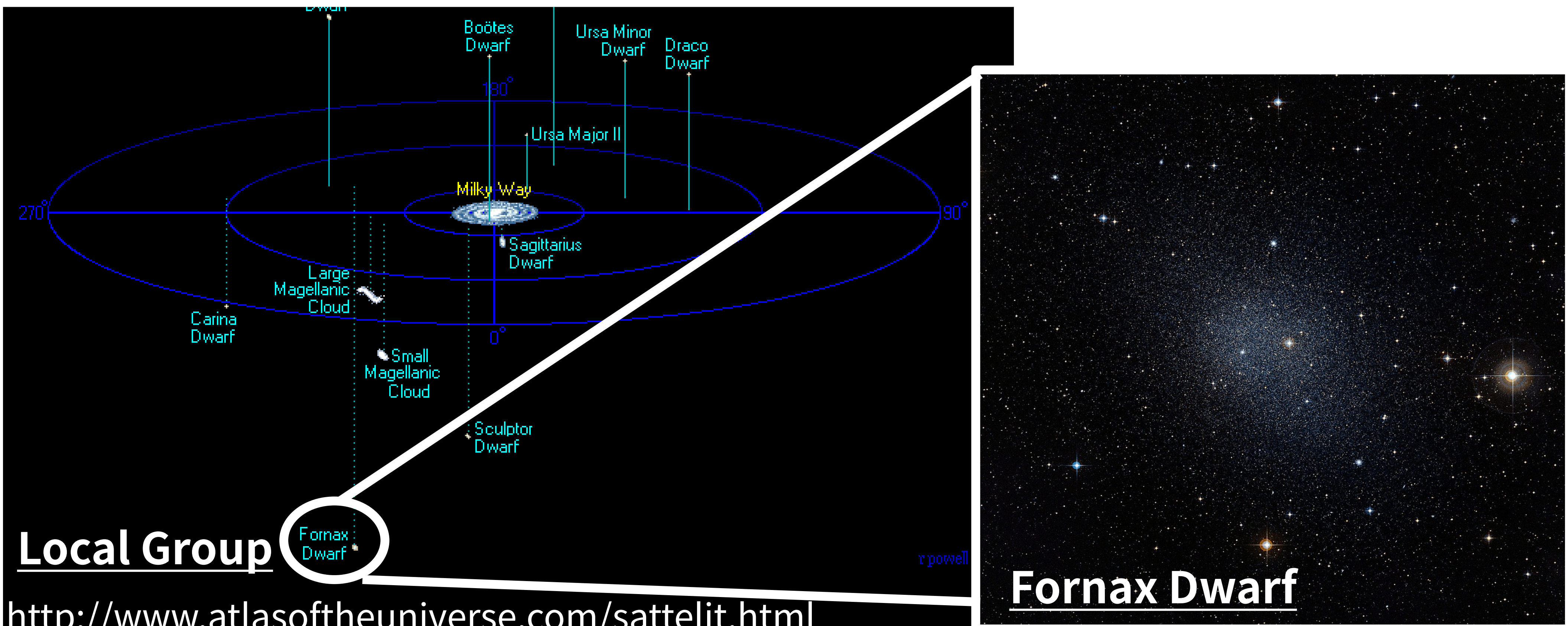
Intergalactic dust cloud
obscuring light from stars!

Dust Obscuring Stars

Intergalactic dust cloud obscuring light from stars!

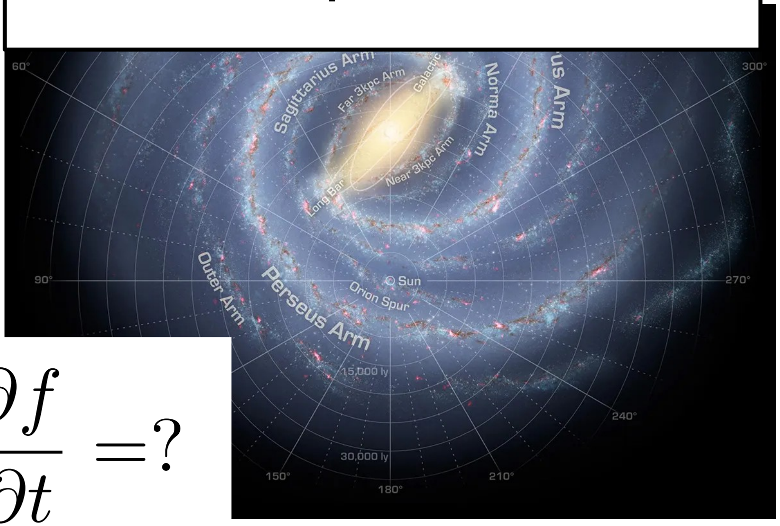
Distant Galaxy – Only partial information is available

- 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.)



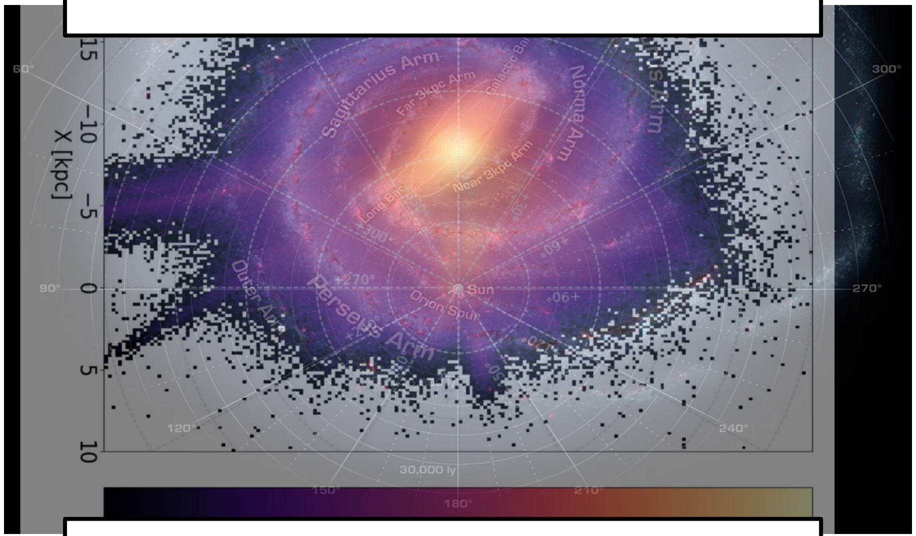
Various challenging incompleteness!

Disequilibrium

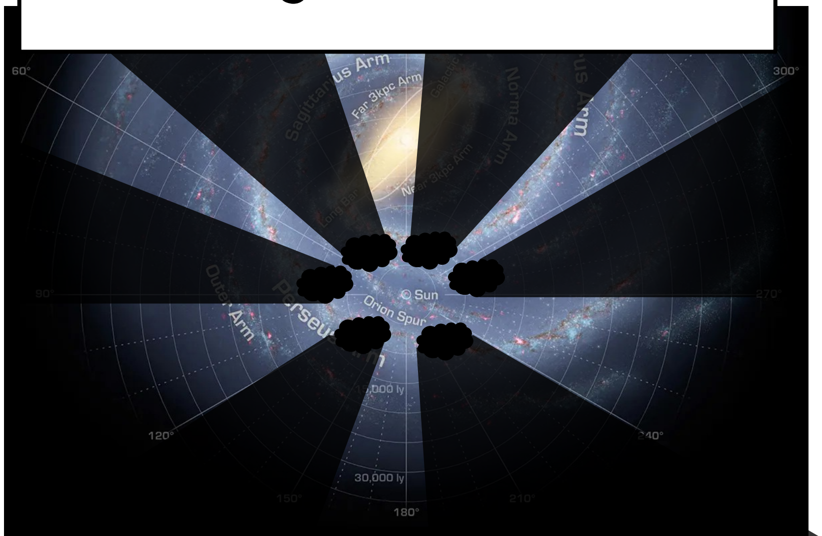


$$\frac{\partial f}{\partial t} = ?$$

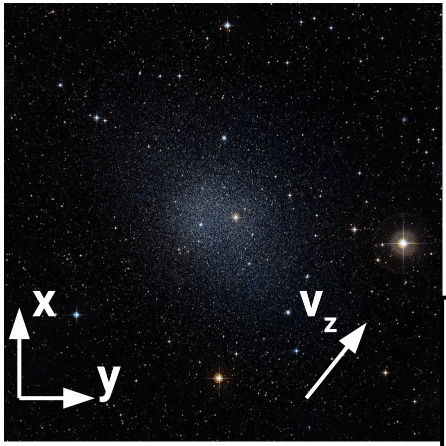
Spatial Incompleteness



Intergalactic Dust



Lack of information



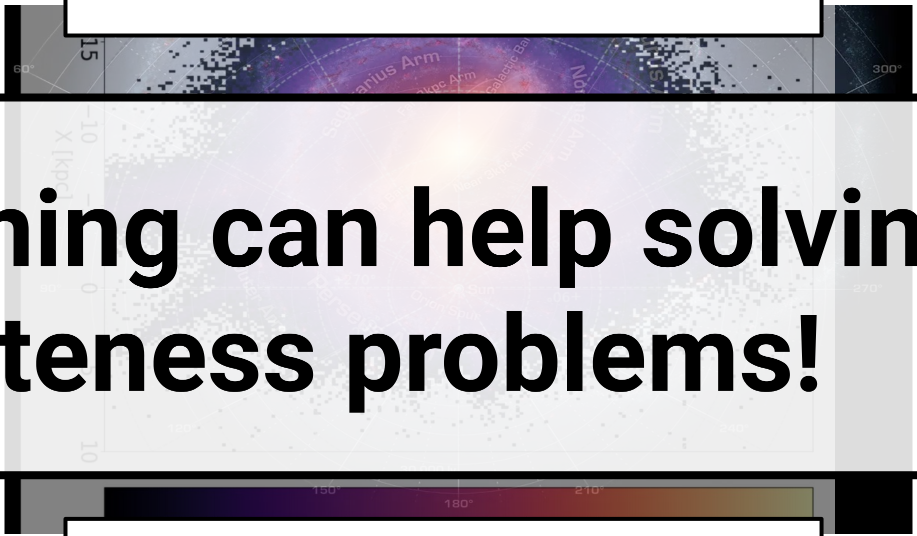
Only 3D info. available, not the full 6D PS info.

More challenges are waiting!

Various challenging incompleteness!

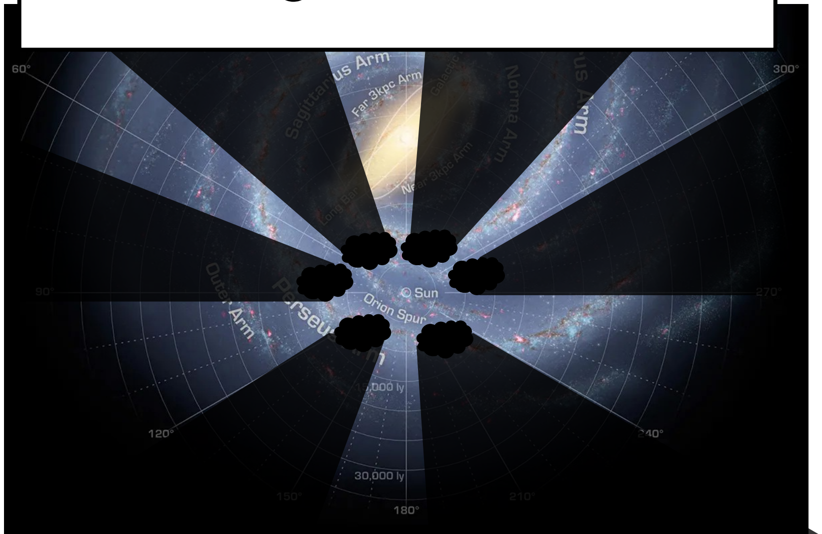
Disequilibrium

Spatial Incompleteness

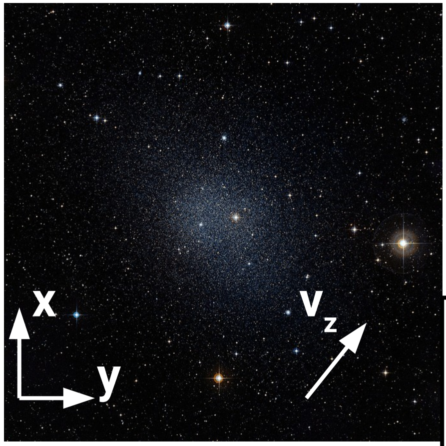


Again, machine learning can help solving these data incompleteness problems!

Intergalactic Dust



Lack of information



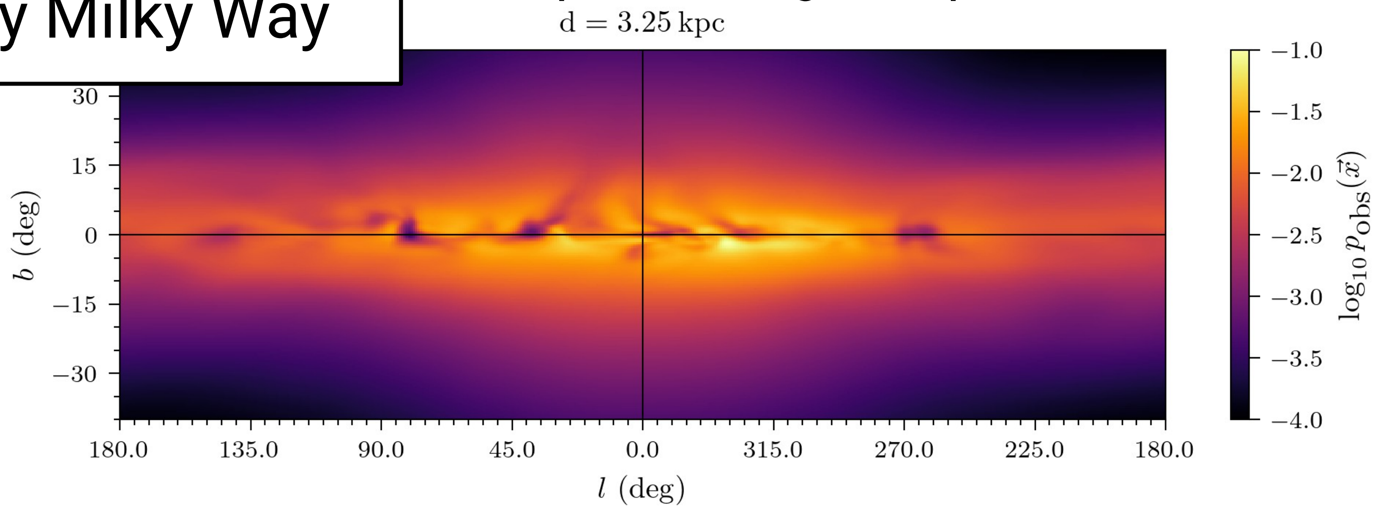
Only 3D info. available, not the full 6D PS info.

More challenges are waiting!

Erasing Dust using Neural Network and Equilibrium Assmptions

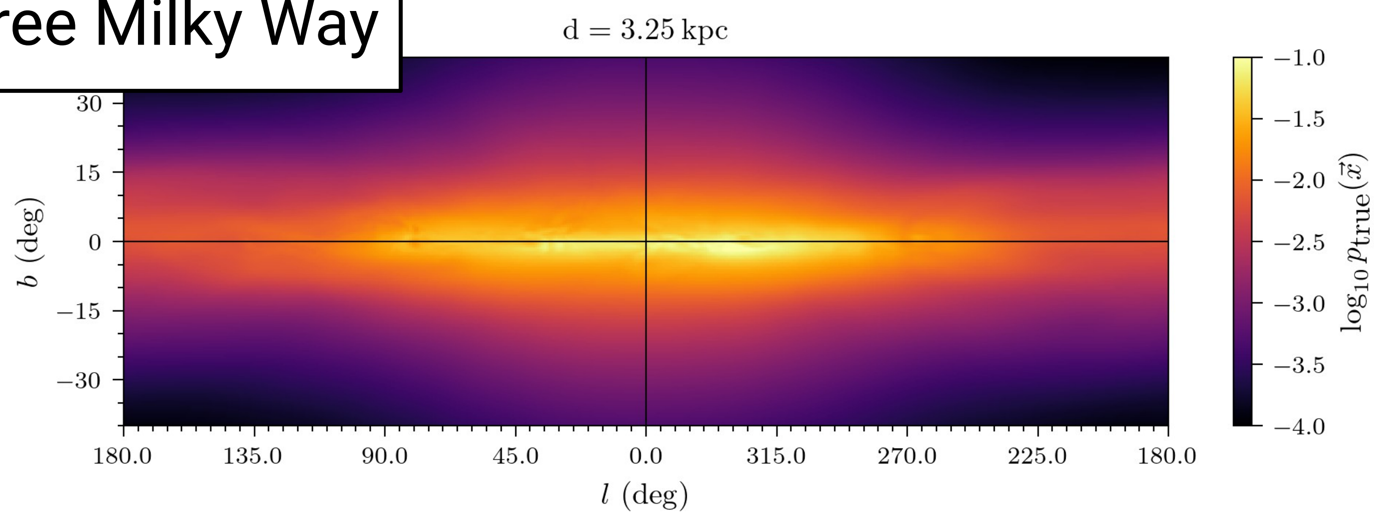
E. Putney, D. Shih, **SHL**, and M. R. Buckley, arXiv:2412.14236
proceeding accepted in NeurIPS 2025 ML4PS workshop

Dusty Milky Way



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

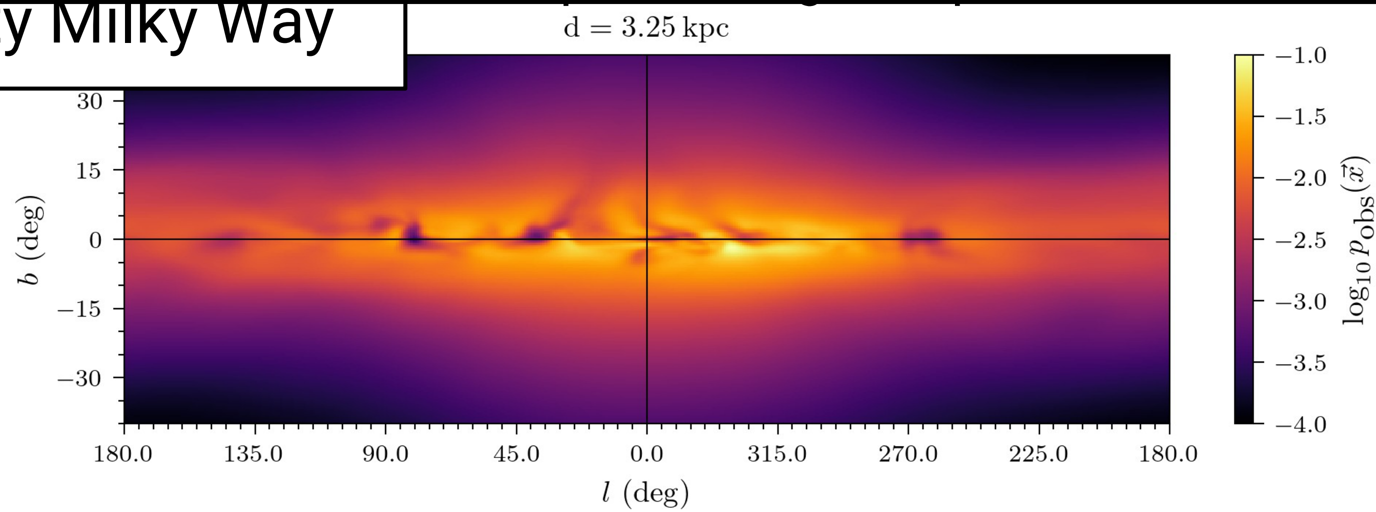
Dust-free Milky Way



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

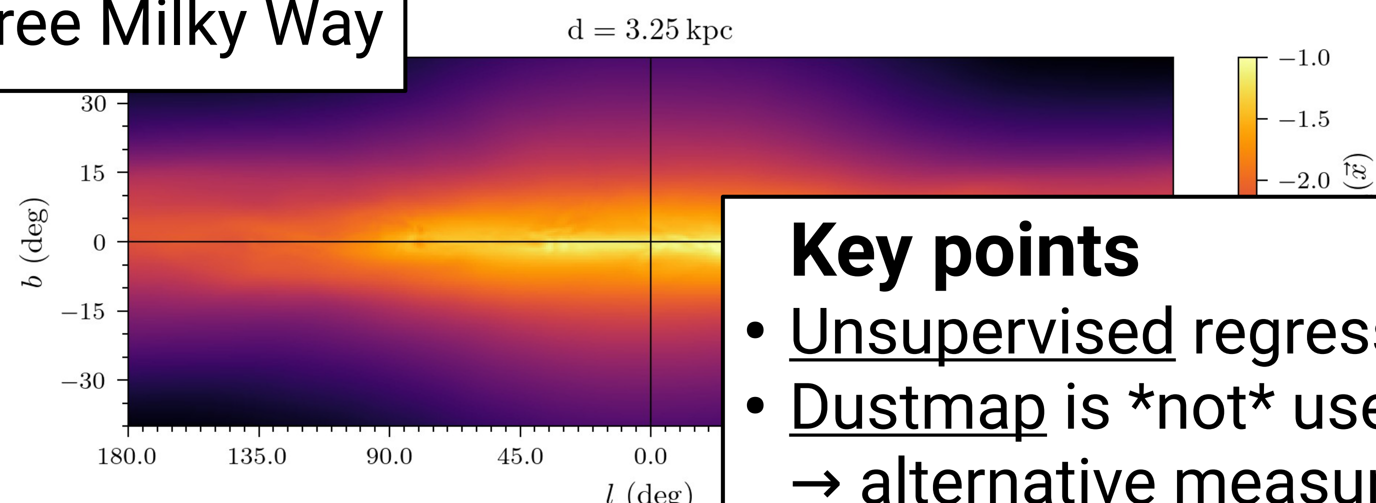
E. Putney, D. Shih, SHL, and M. R. Buckley, arXiv:2412.14236

Dusty Milky Way



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

Dust-free Milky Way

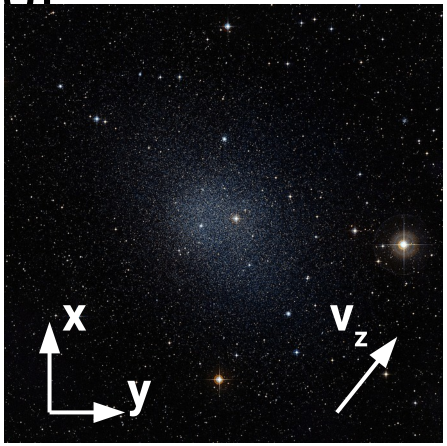


Key points

- Unsupervised regression
- Dustmap is *not* used
  → alternative measurement of dust!

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 **<u>limited</u>**!
  - 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.)

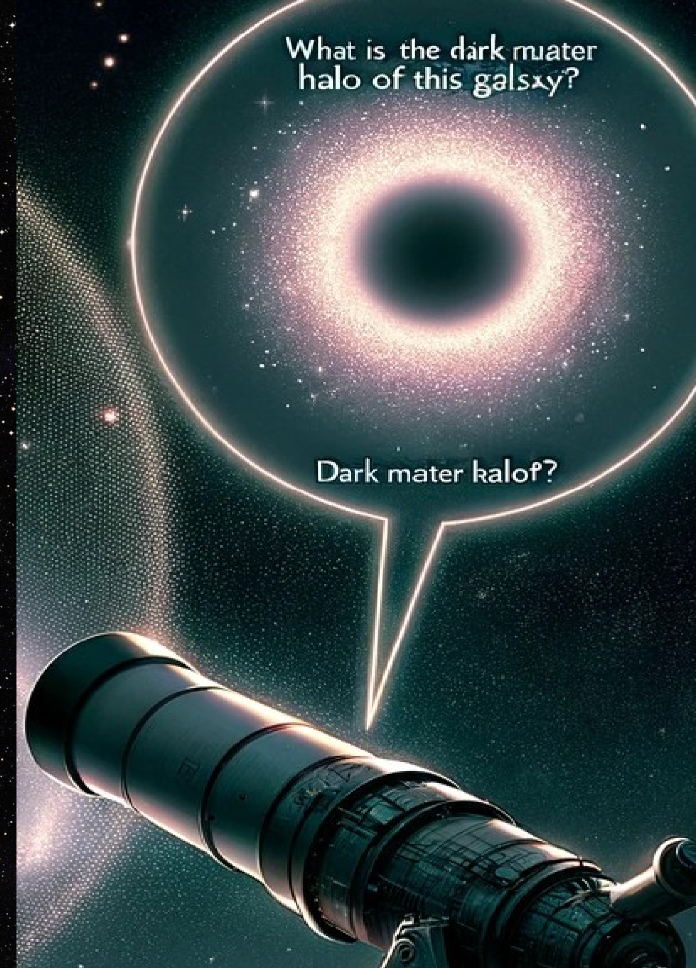
We can impose such <u>physical constrants</u> to neural networks!

- 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 **<u>spherical symmetry</u>**

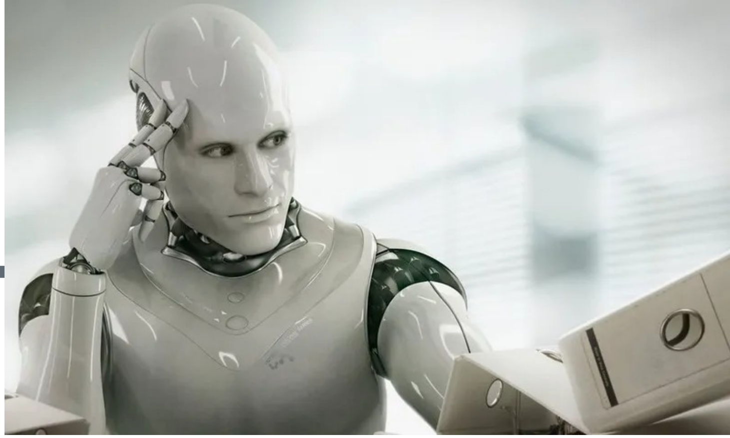


JFlow: Model-Independent Spherical Jeans Analysis using Equivariant Continuous Normalizing Flows

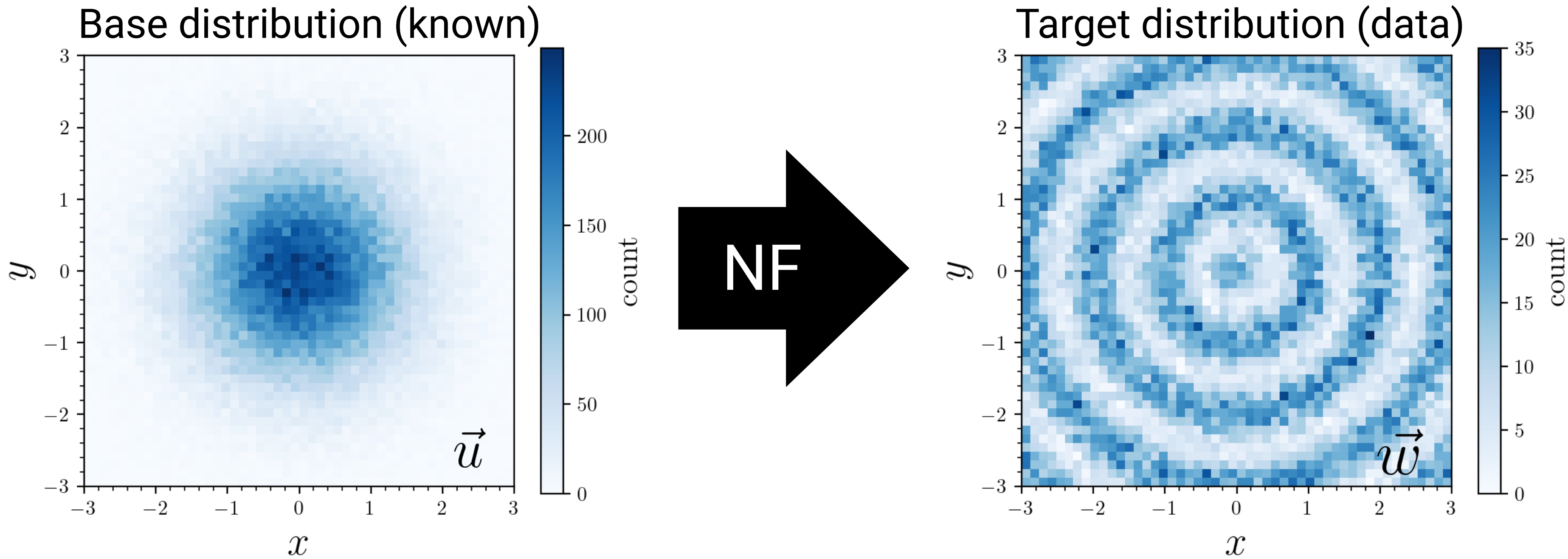
Collaboration with
K. Hayashi (NIT, Sendai College), S. Horigome (Tohoku),
S. Matsumoto (IPMU), M. M. Nojiri (KEK),

arXiv:2505.00763

Normalizing Flows: Neural Density Estimator



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:

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

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

Equivariant
Continous Normalizing Flows

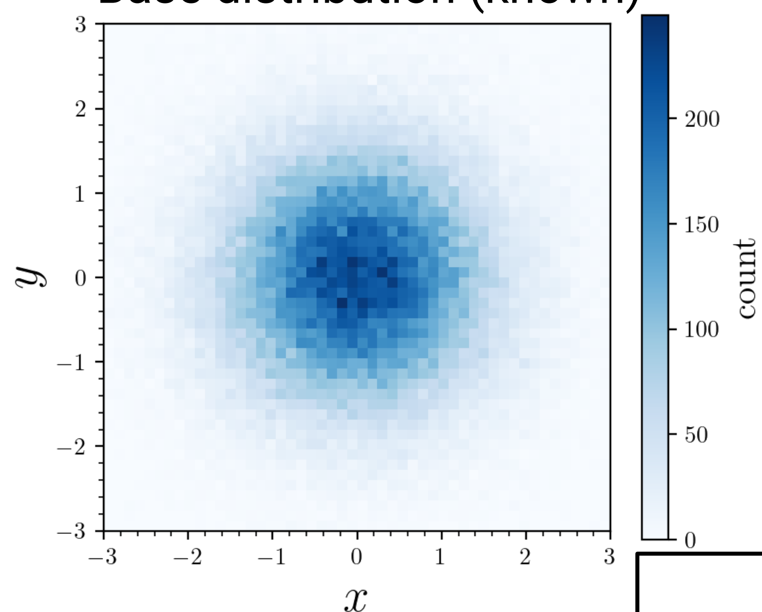
$n(r)$

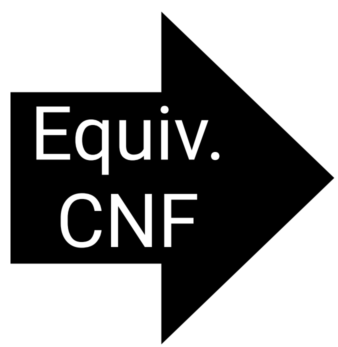
How to model spherically symmetric density using normalizing flows?
→ Use Equivariant Continuous Normalizing Flows!

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

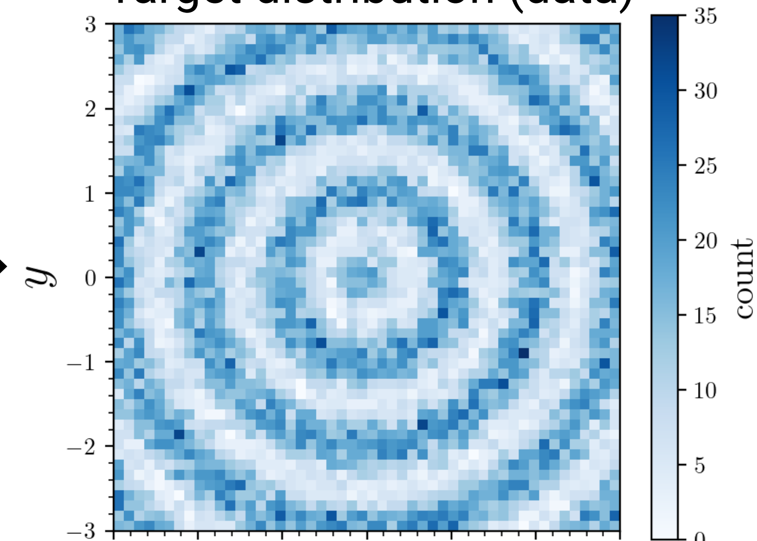
- Invariant (Gaussian) base distribution
- Equivariant vector field

Base distribution (known)





Target distribution (data)

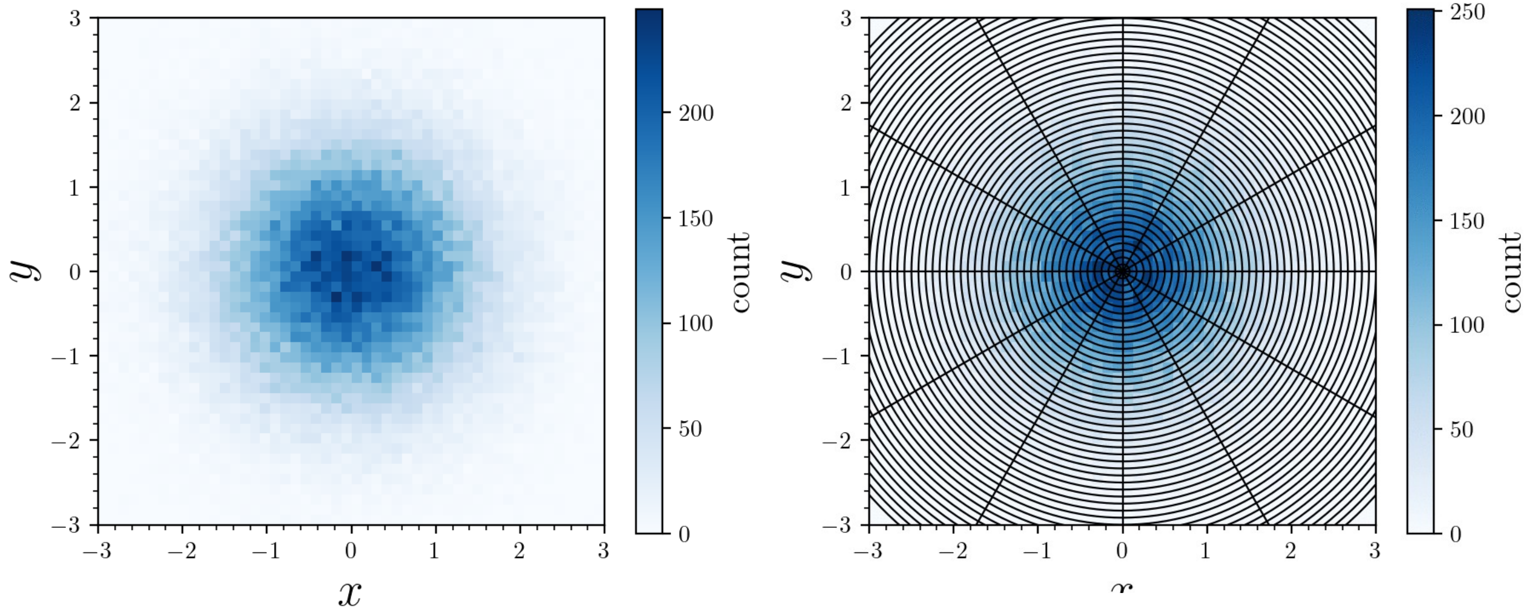


This setup is very flexible. You may add **physics constraints** to **neural networks**, too!

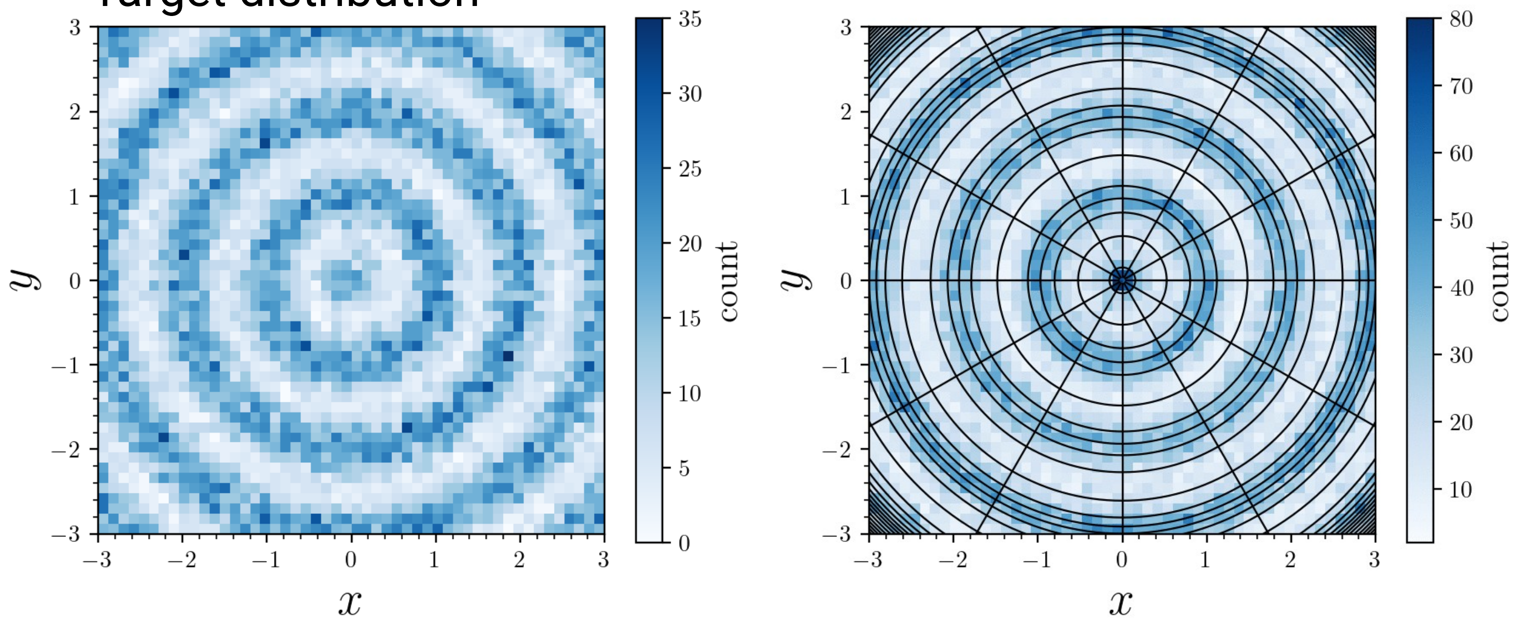
Normalizing Flows: How it works?

$n(r)$

Base distribution



Target distribution



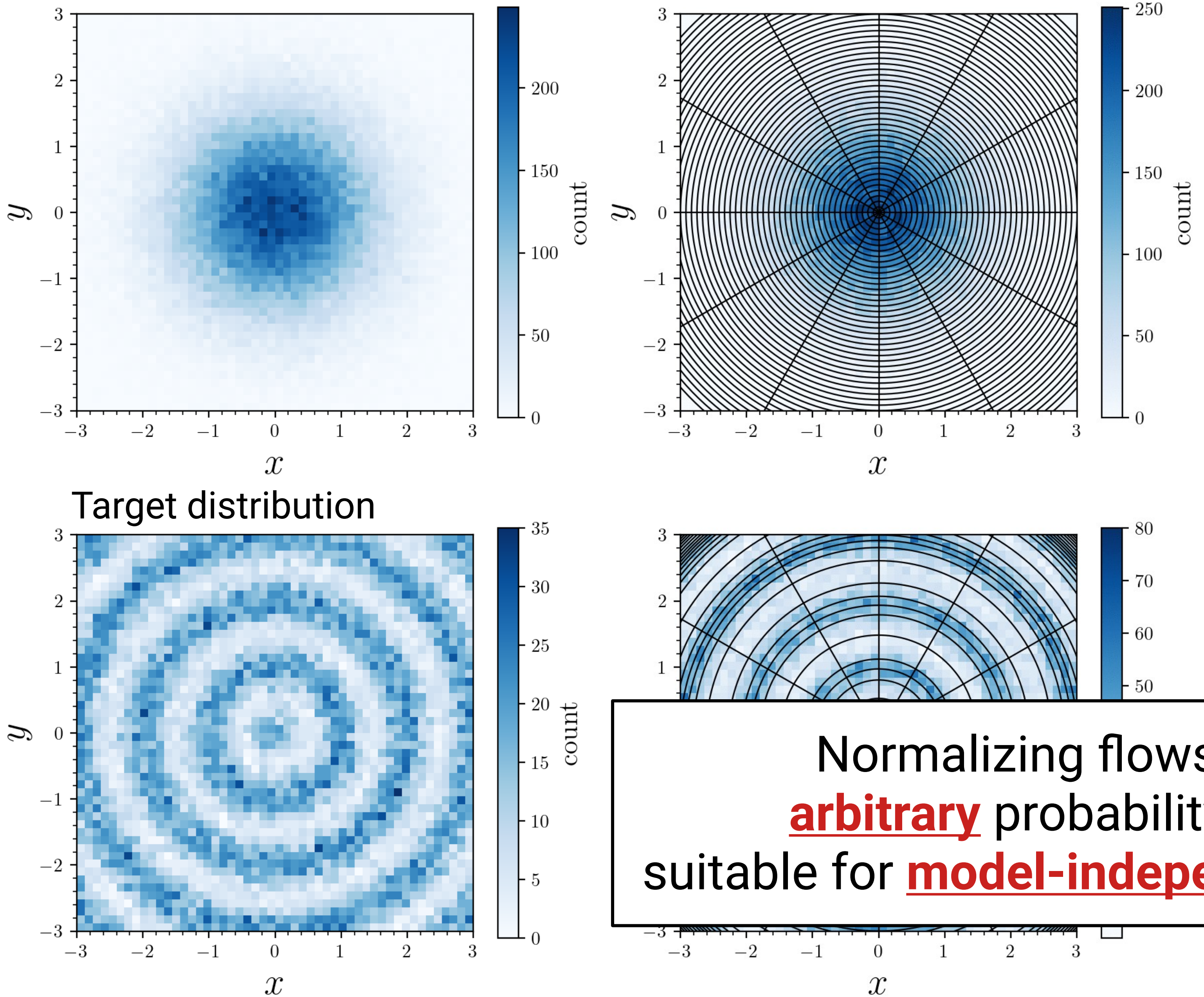
* result of a continuous normalizing flow learning infinitesimal transformations

Normalizing Flows: How it works?

$n(r)$

Base distribution

Target distribution



Normalizing flows can fit **arbitrary** probability density, suitable for **model-independent** analysis!

* result of a continuous normalizing flow learning infinitesimal transformations

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

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:

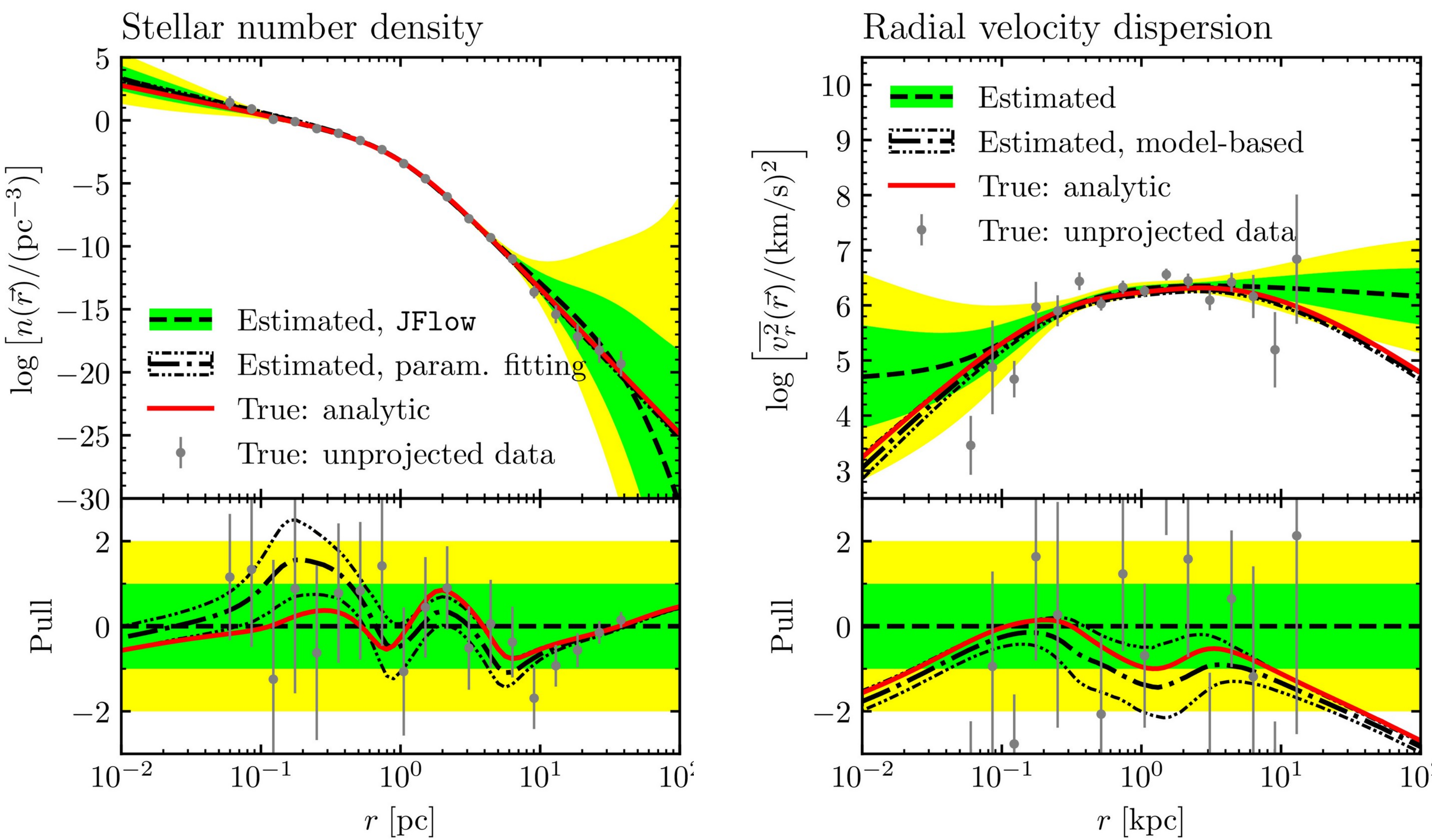
$$p(r) = \left(\frac{r}{r_0}\right)^{-1} \left(1 + \frac{r}{r_0}\right)^{-2} \rightarrow \frac{1}{r}$$

Apply power-law transform to radial component

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

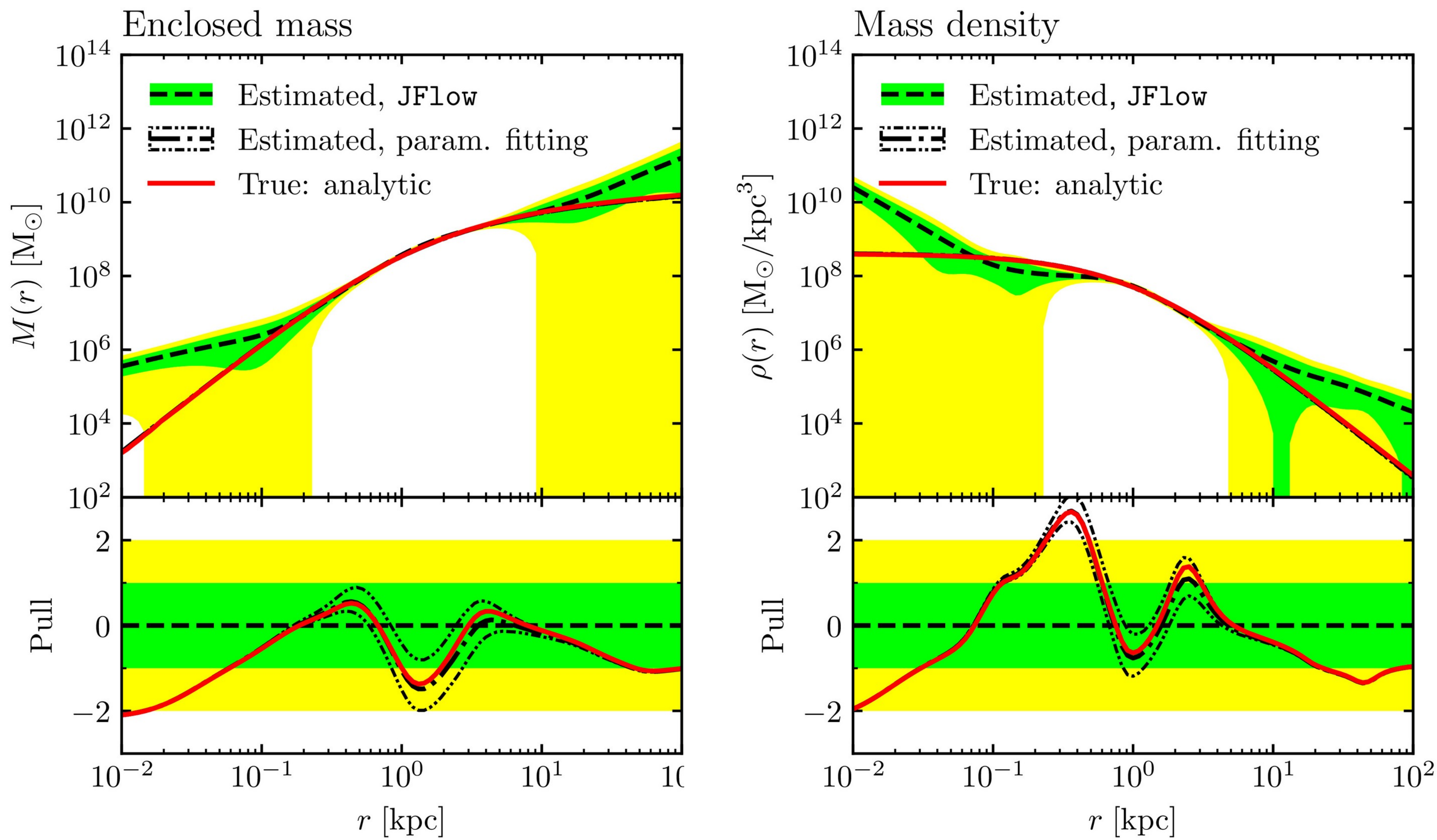
to **cored** spherical symmetric density model

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

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

$\rightarrow$

Gravitational accel.

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

Machine Learning Method for Density Derivative Estimation?

<u>Score Matching (Hyvariene 2005)</u>

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 **<u>neural network</u>** regressing **<u>score functions!</u>**

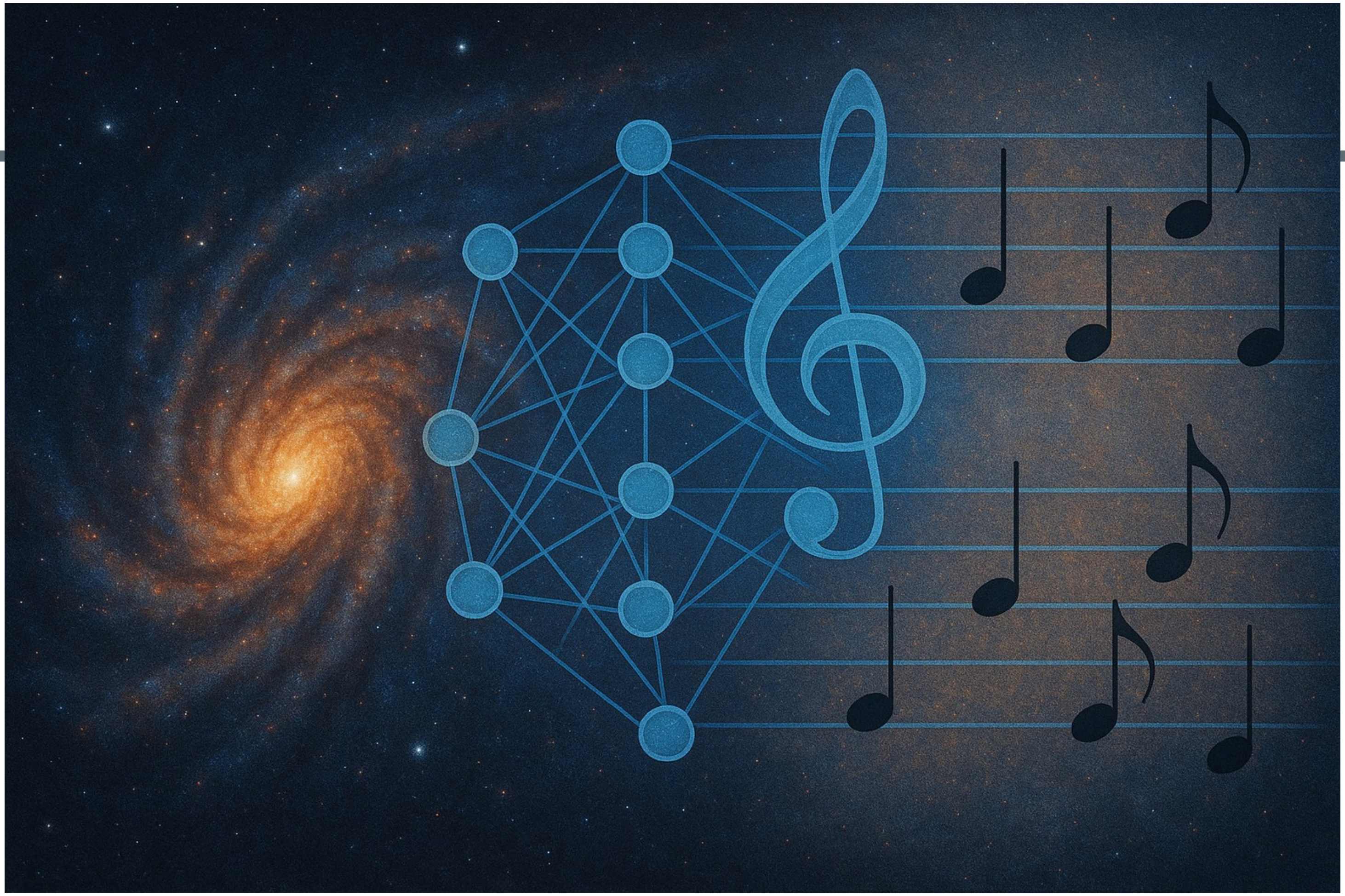
<u>Explicit Score Matching Loss = (MSE Loss)</u>

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

<u>Inplicit Score Matching Loss</u>

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

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

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

Score functions

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

Gravitational accel.

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

Mass density

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

Gaia DR3: star catalog of the Milky Way

Score Matching (Hyvariene 2005)

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

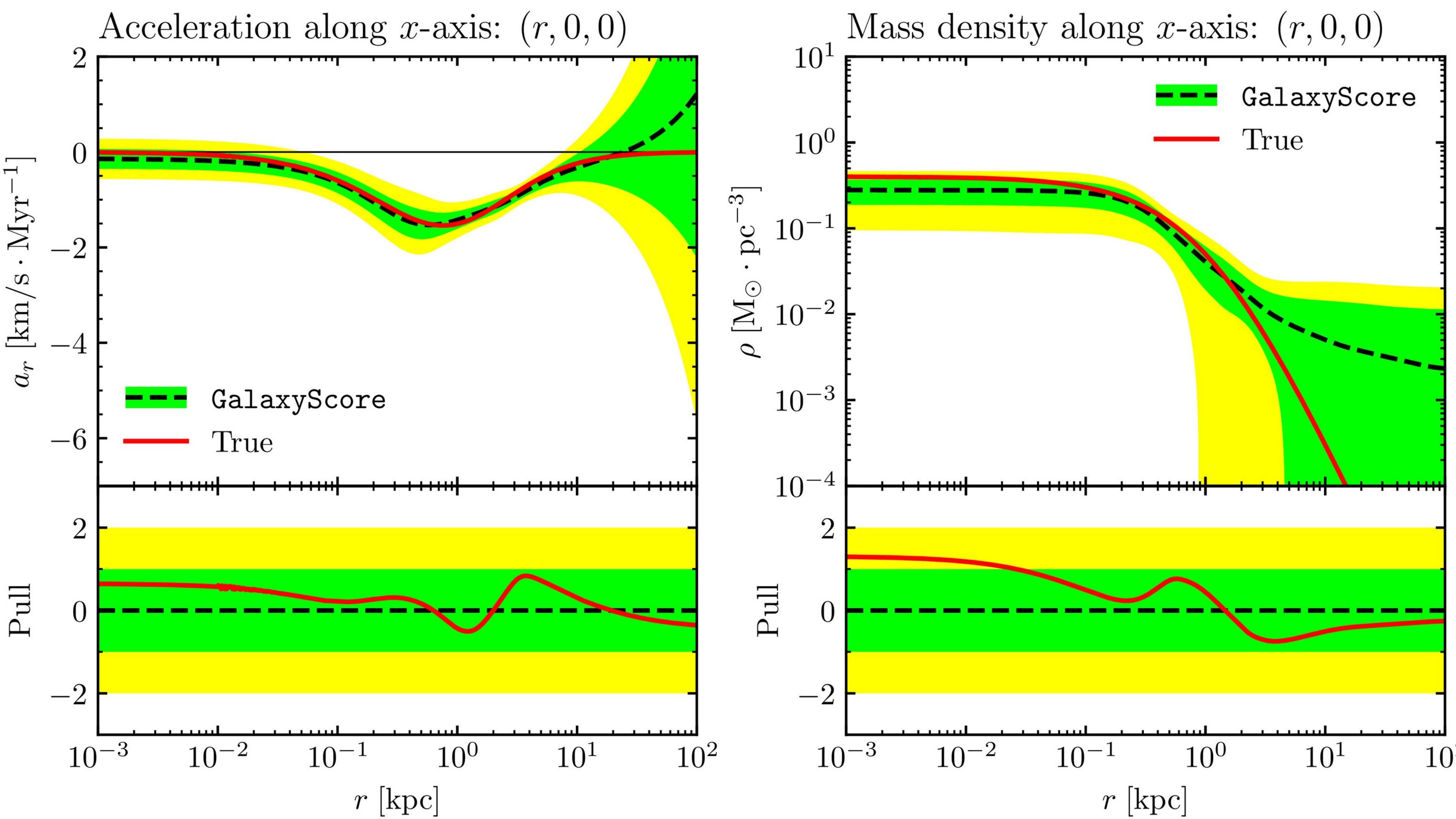
Solving EOM (Boltzmann Equation)

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

Solving Gauss's Equation

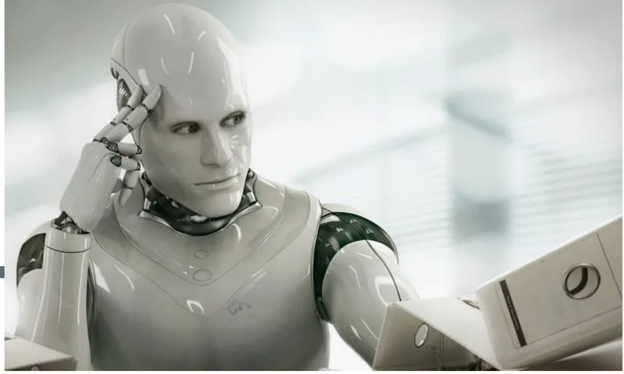
$$-4\pi G \rho = \nabla \cdot \vec{a}$$

Proof-of-Concept Results



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

Future Timeline

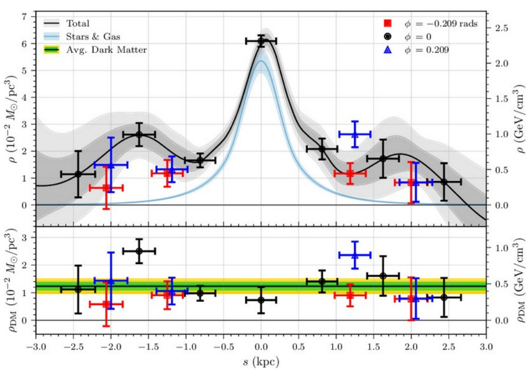


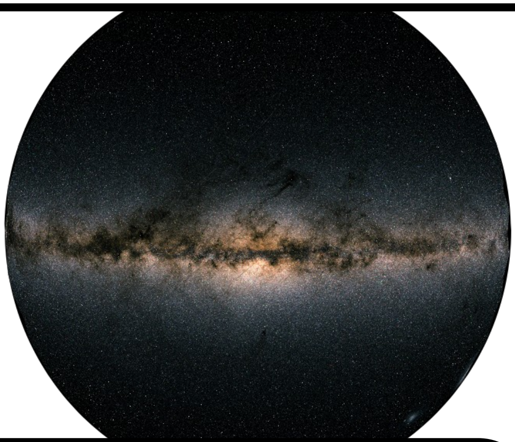
2025

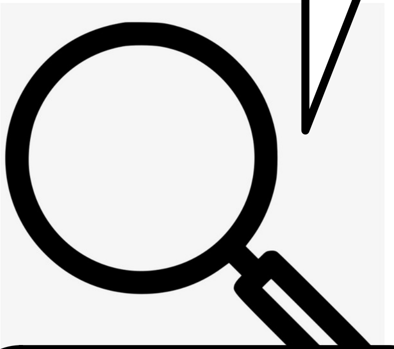
2026

Gaia DR3

Gaia DR4, SuperPFS, and so on!







Understanding DM in dust-free region (halo, solar neighborhood)

Understanding DM in dusty region (disk)

New methods for analyzing distant satellite galaxies

More precision and new opportunities!

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)

Thank you for listening!