TARTU ÜLIKOOL
Tartu observatoorium

Leveraging Machine Learning to Detect Dark Matter Subhalos in the Milky Way

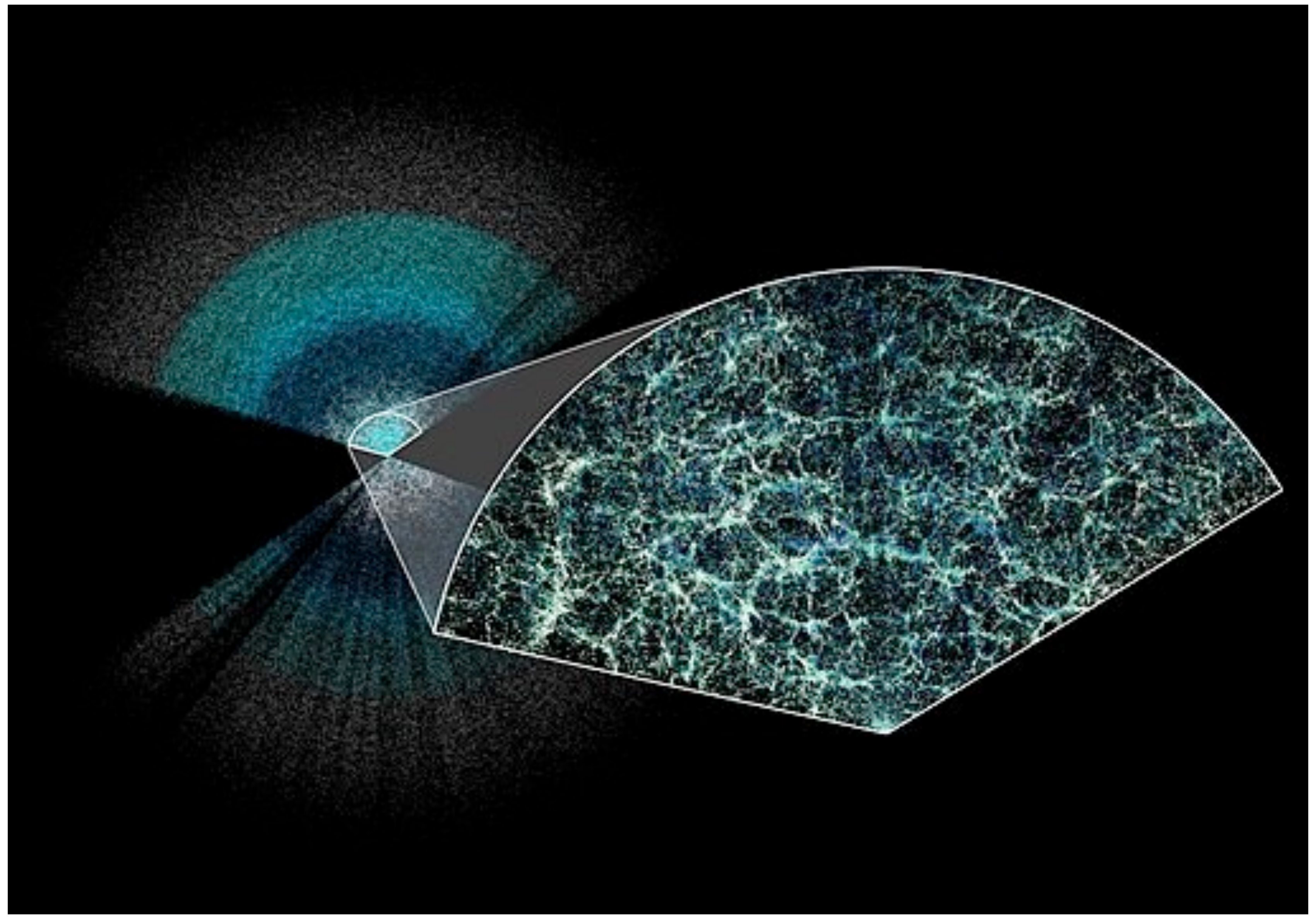
María Benito

Joosep Pata

Sven Põder

1st SMASHING WORKSHOP, 10/10/‘24

DESI collaboration



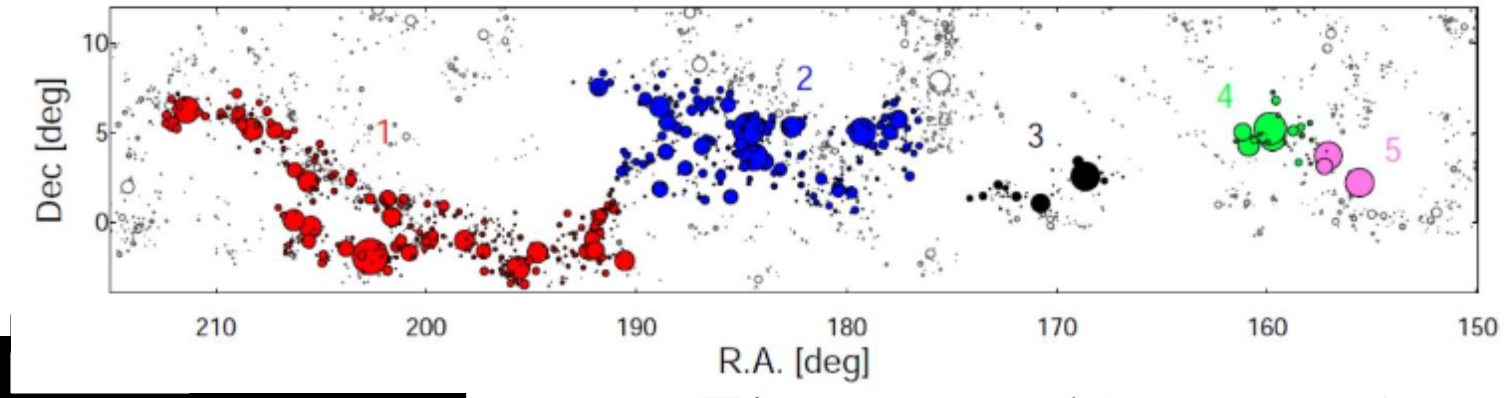
Cosmic Web (supercluster - void network)
Superclusters: 1% of the volume and ~ 15 % of mass of the Universe. **Voids:** 70 – 90% of the volume and 15 % of mass, ~ 10 – 30%: supercluster outskirts

"Large-scale structure of the Universe"
1977 Tallinn symposium

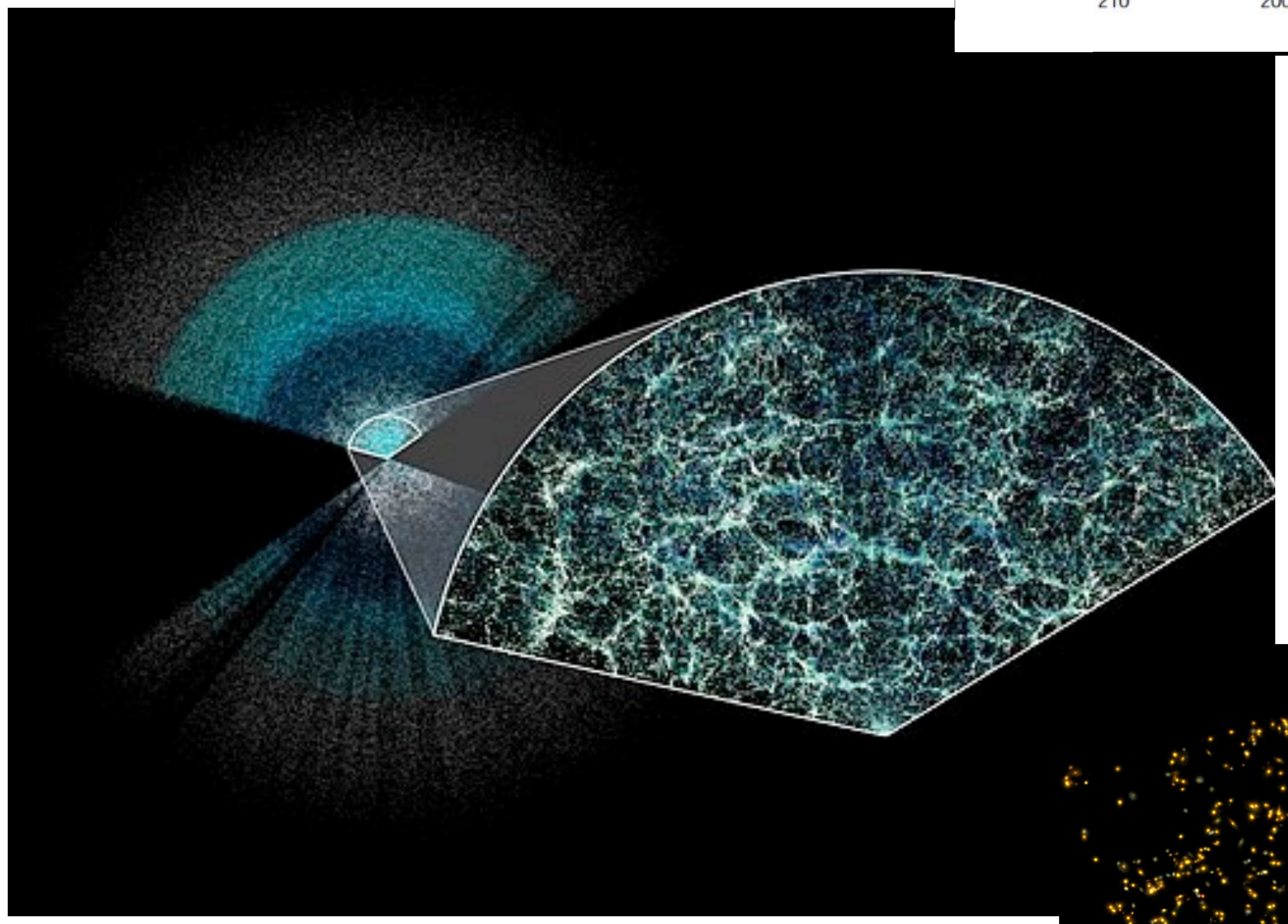
Einasto et al. '22

¿ Supercluster complexes, shells/ planes & regularity ??

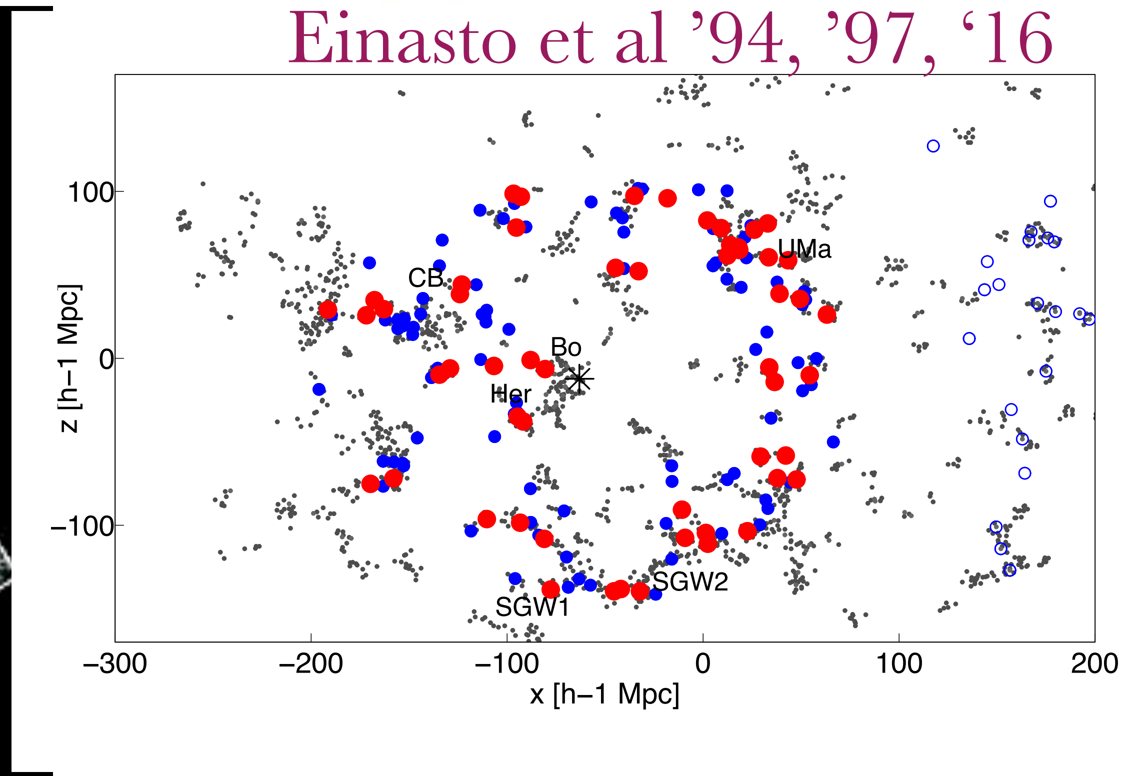
Liivamägi, Tempel & Saar ‘12

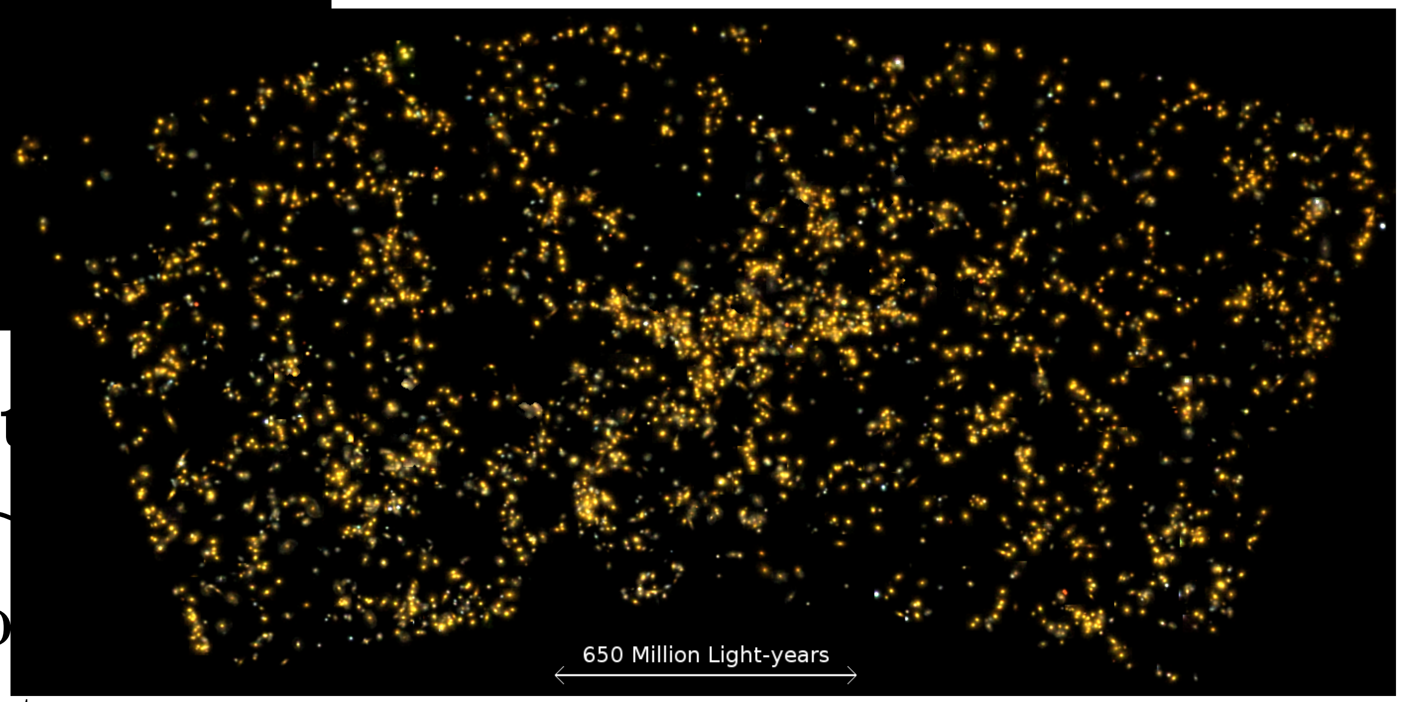


DESI collaboration



Einasto et al ’94, ’97, ‘16





Bagchi et al. ’17, Sankhyayan et al. ‘23

Cosmic Web (supercluster - void net
Superclusters: 1% of the volume and ~
mass of the Universe. **Voids:** 70 – 90% o
and 15 % of mass, ~ 10 – 30%: supercluster outskirts

“Large-scale structure of the Universe”
1977 Tallinn symposium

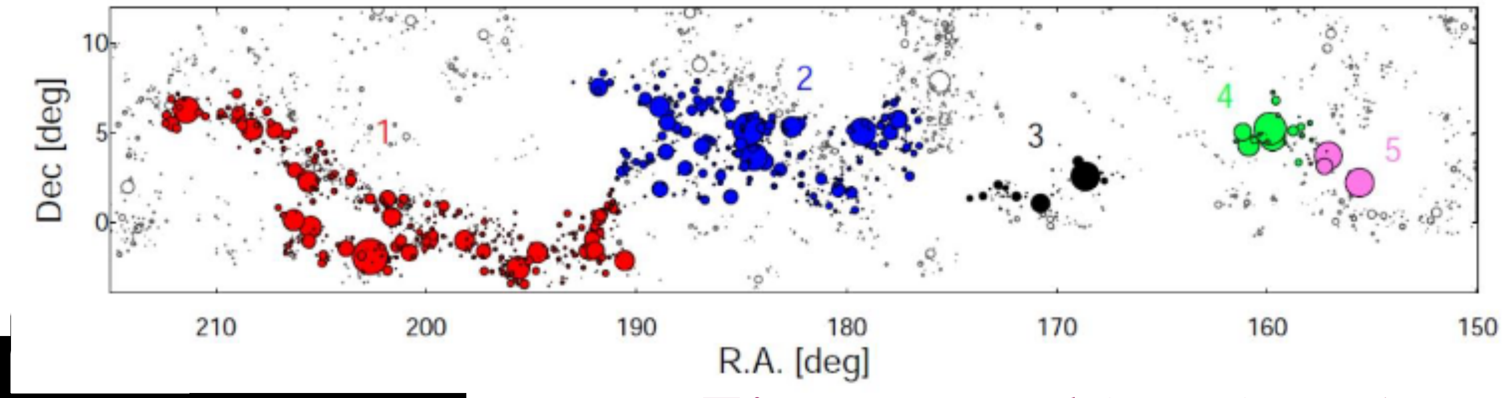
Einasto et al. ‘22

¿Ho’oleilana = individual BAO?

Tully et al. ‘23

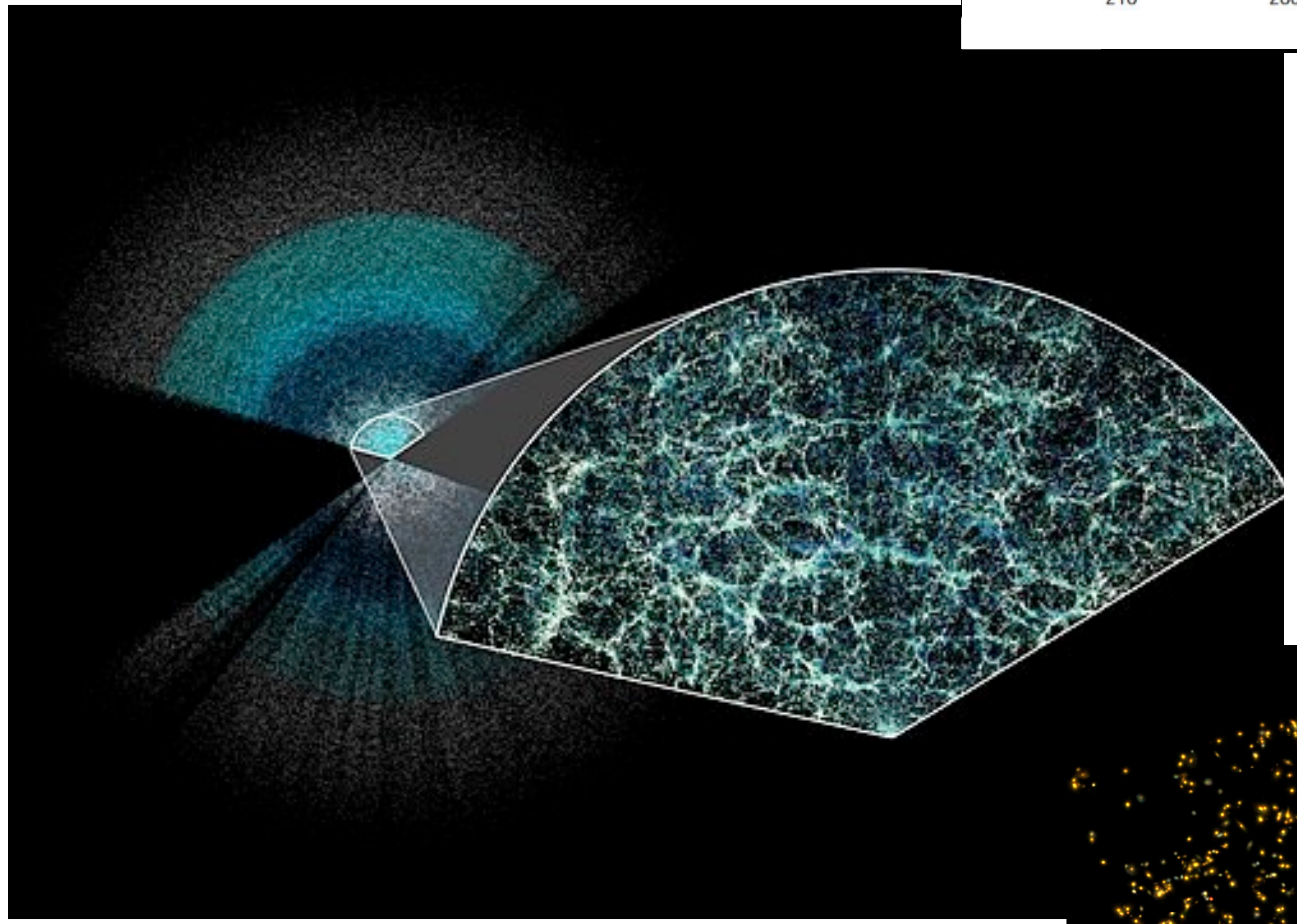
¿ **Supercluster complexes, shells/ planes & regularity** ??

Liivamägi, Tempel & Saar ‘12

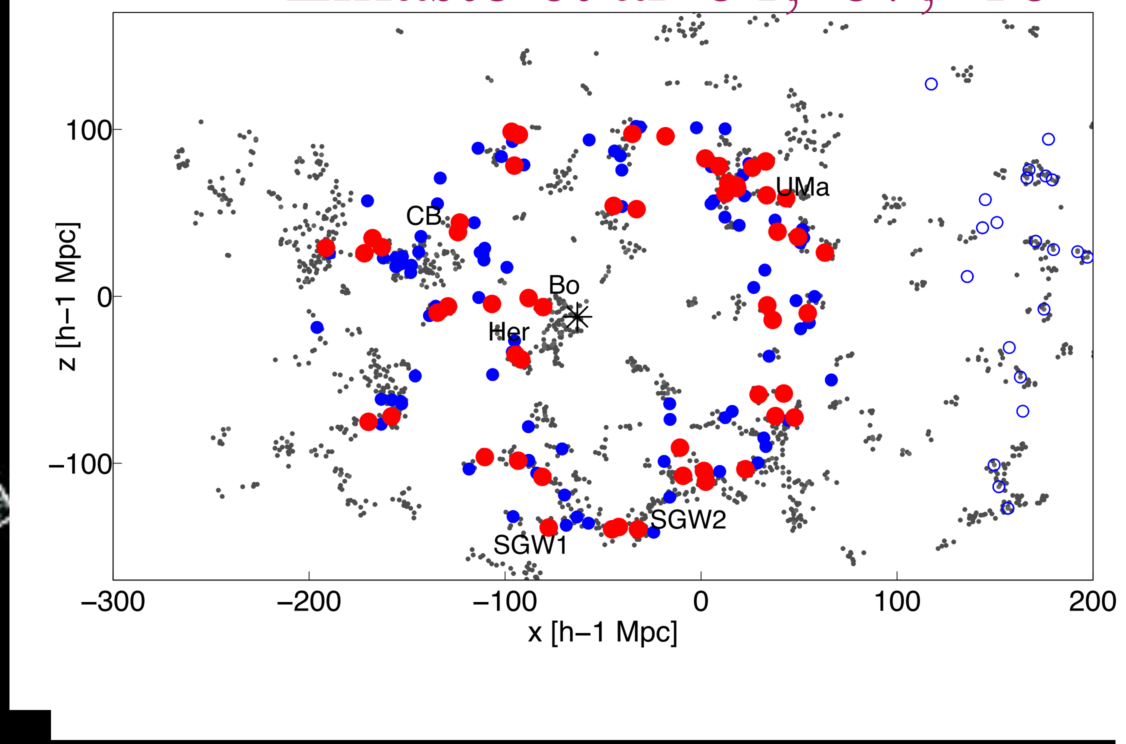


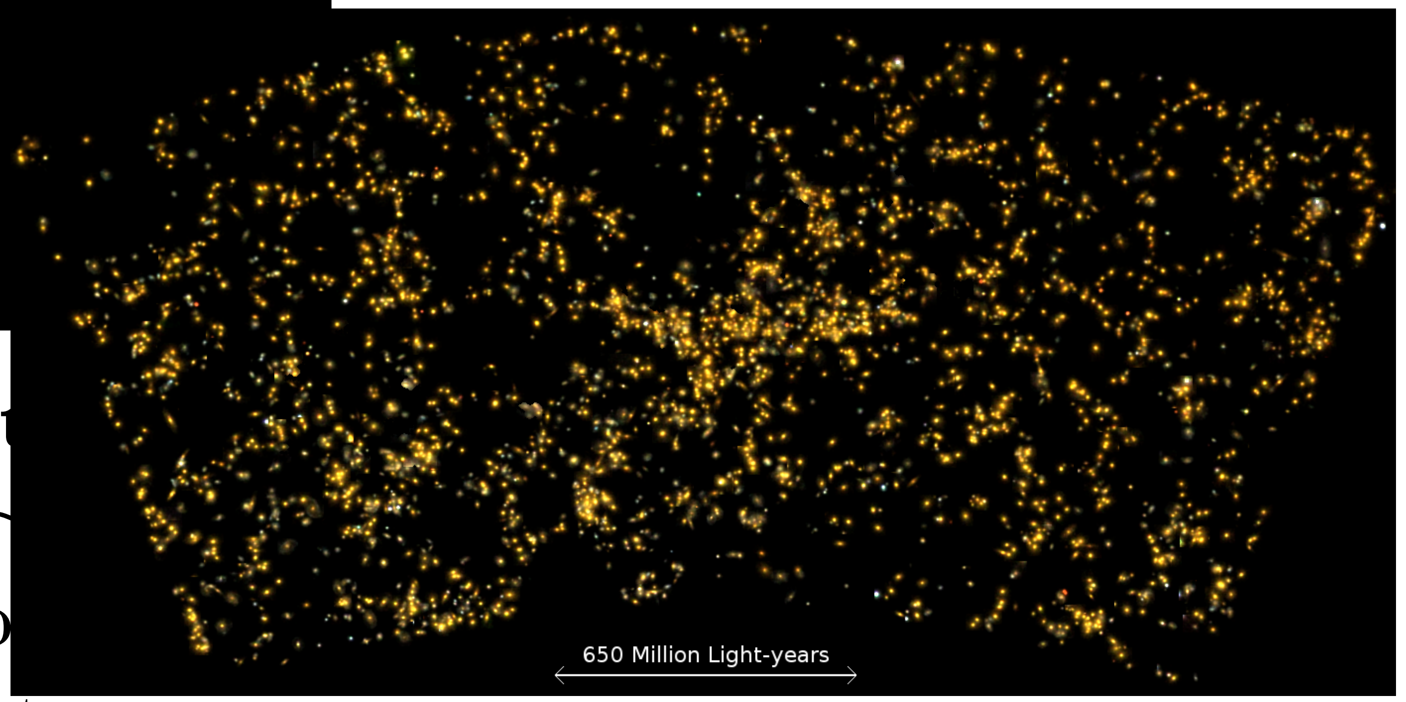
Observed visible & dark Milky Way @ 2024

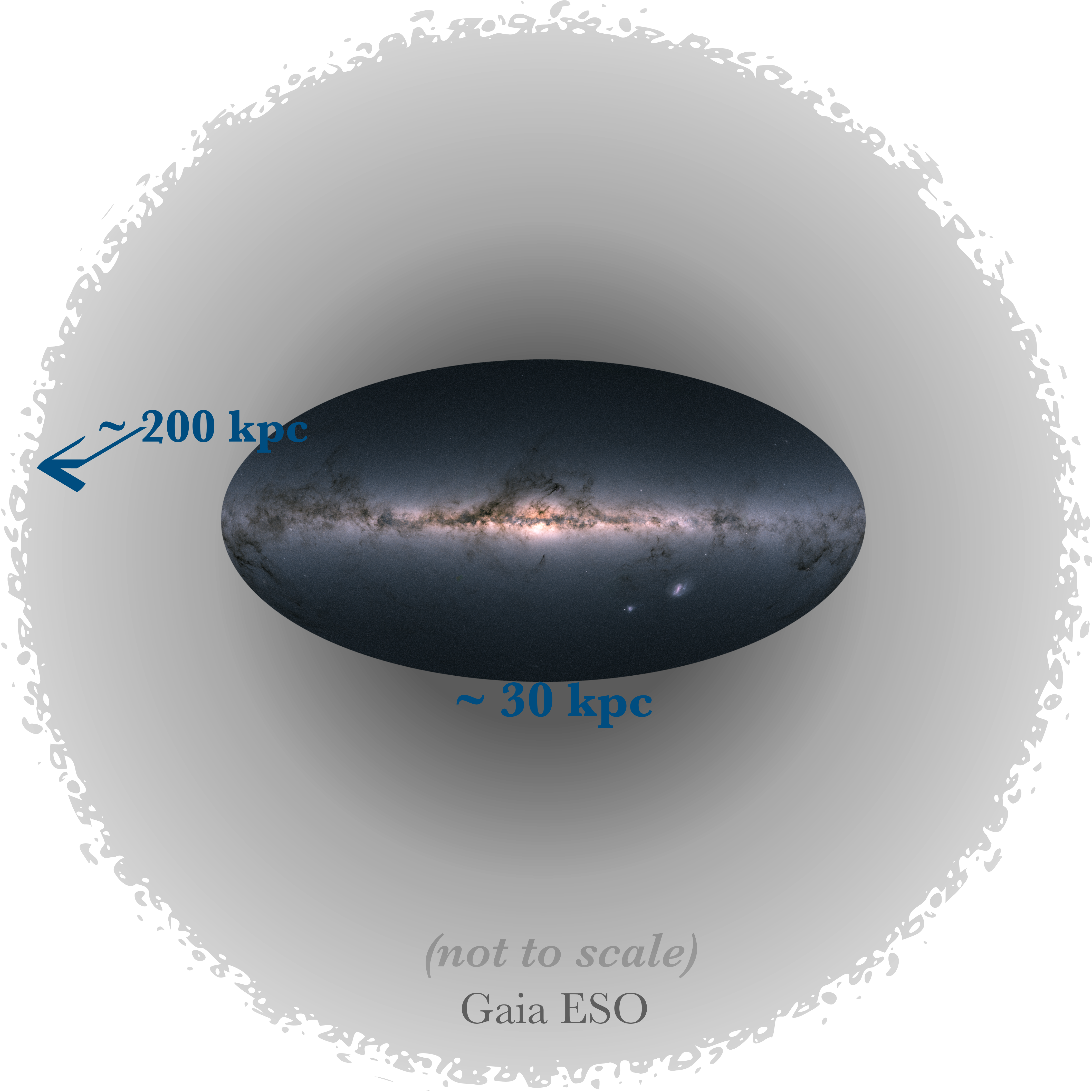
DESI collaboration



Einasto et al ’94, ’97, ‘16







Cosmic Web (supercluster - void net
Superclusters: 1% of the volume and ~
mass of the Universe. **Voids:** 70 – 90% o
and 15 % of mass, ~ 10 – 30%: supercluster
outskirts

Bagchi et al. ’17, Sankhyayan et al. ‘23

“Large-scale structure of the Universe”
1977 Tallinn symposium

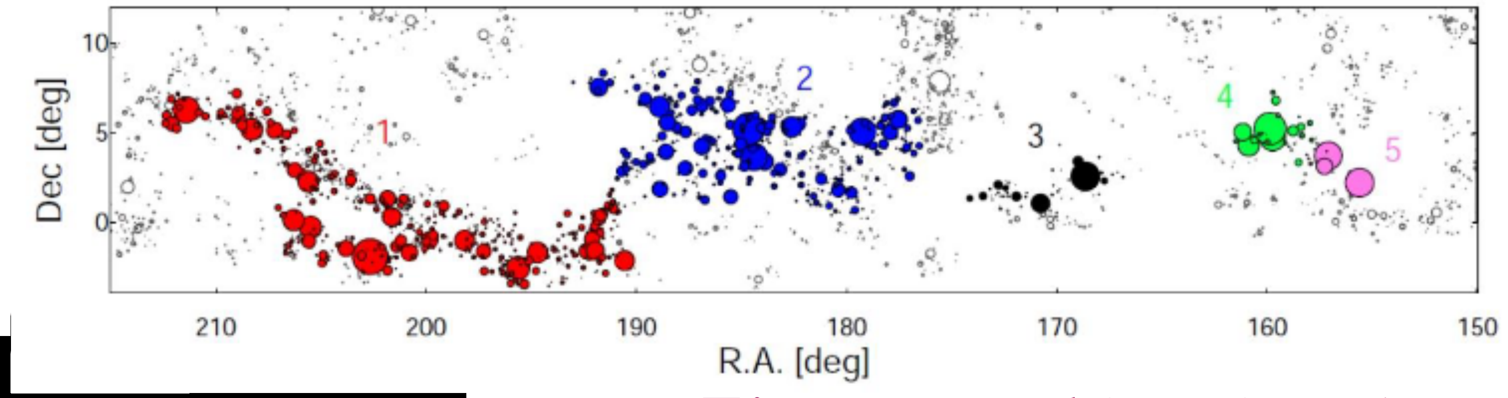
Einasto et al. ‘22

¿Ho’oleilana = individual BAO?

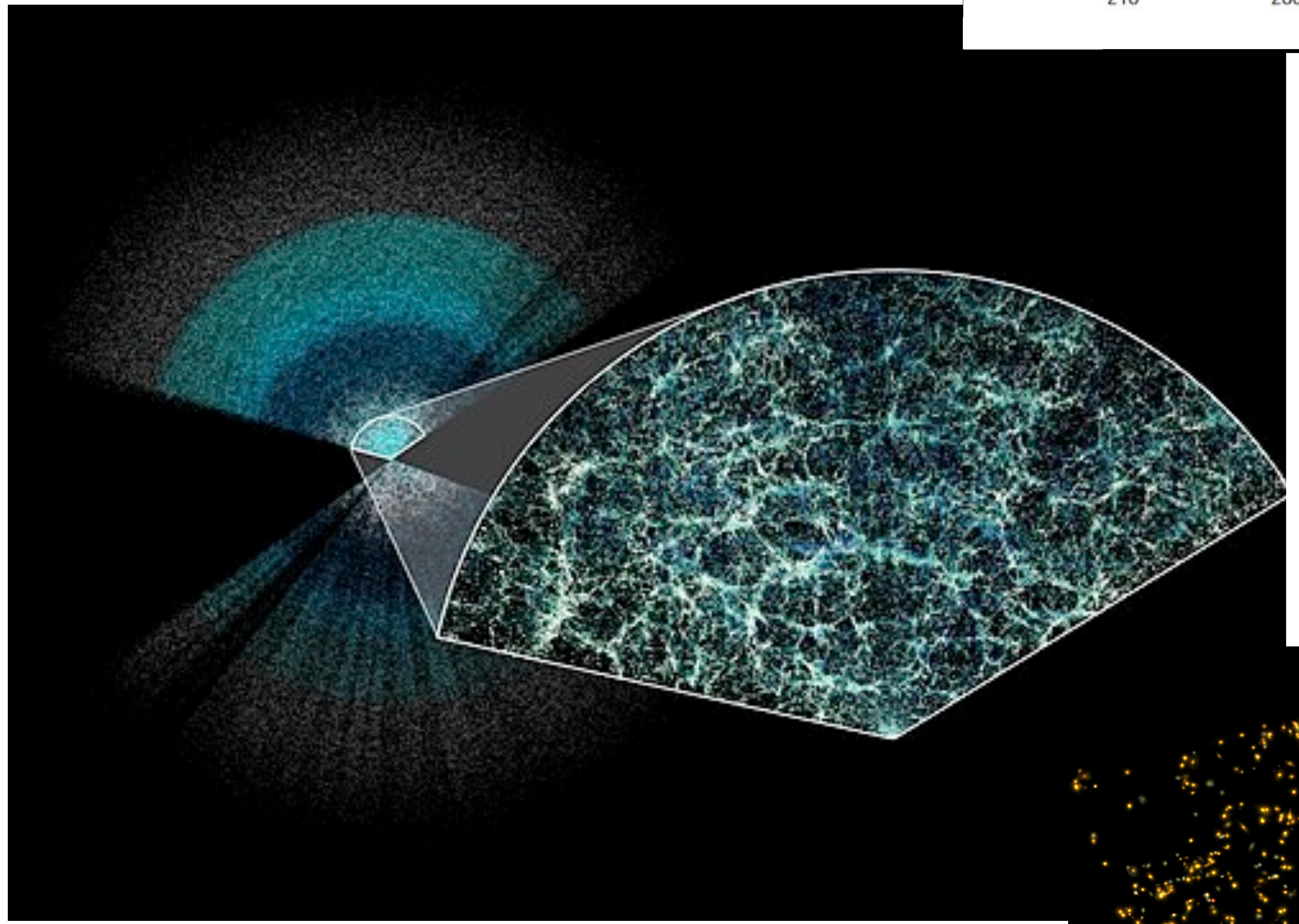
Tully et al. ‘23

¿ Supercluster complexes, shells/ planes & regularity ??

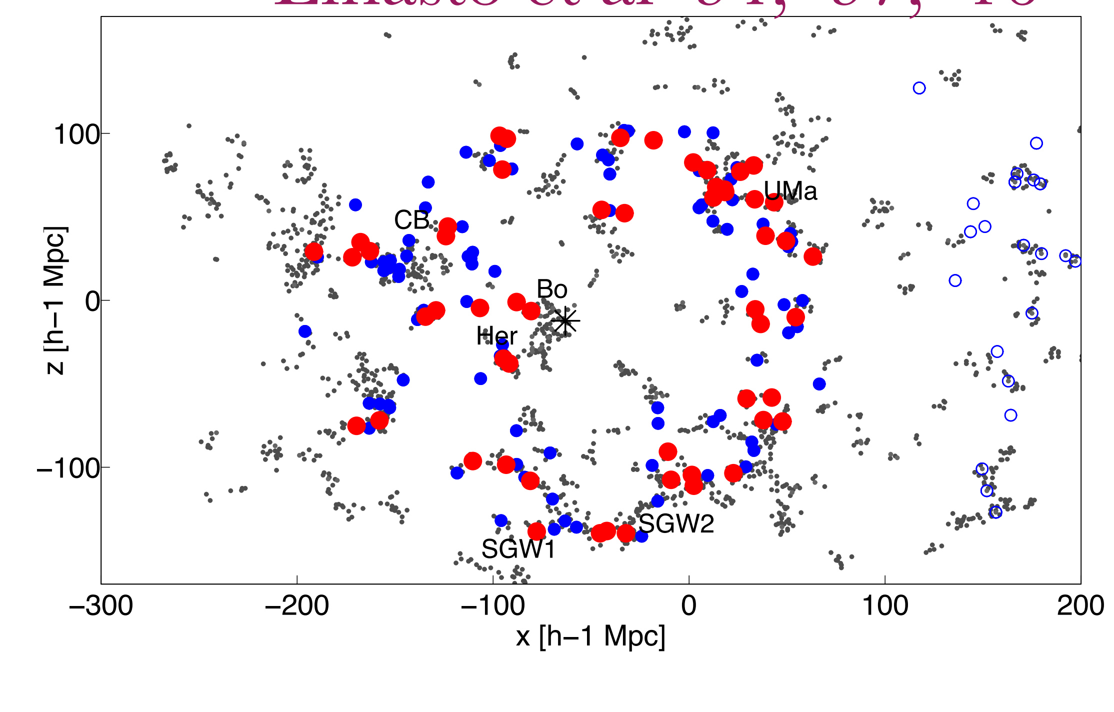
Liivamägi, Tempel & Saar ‘12

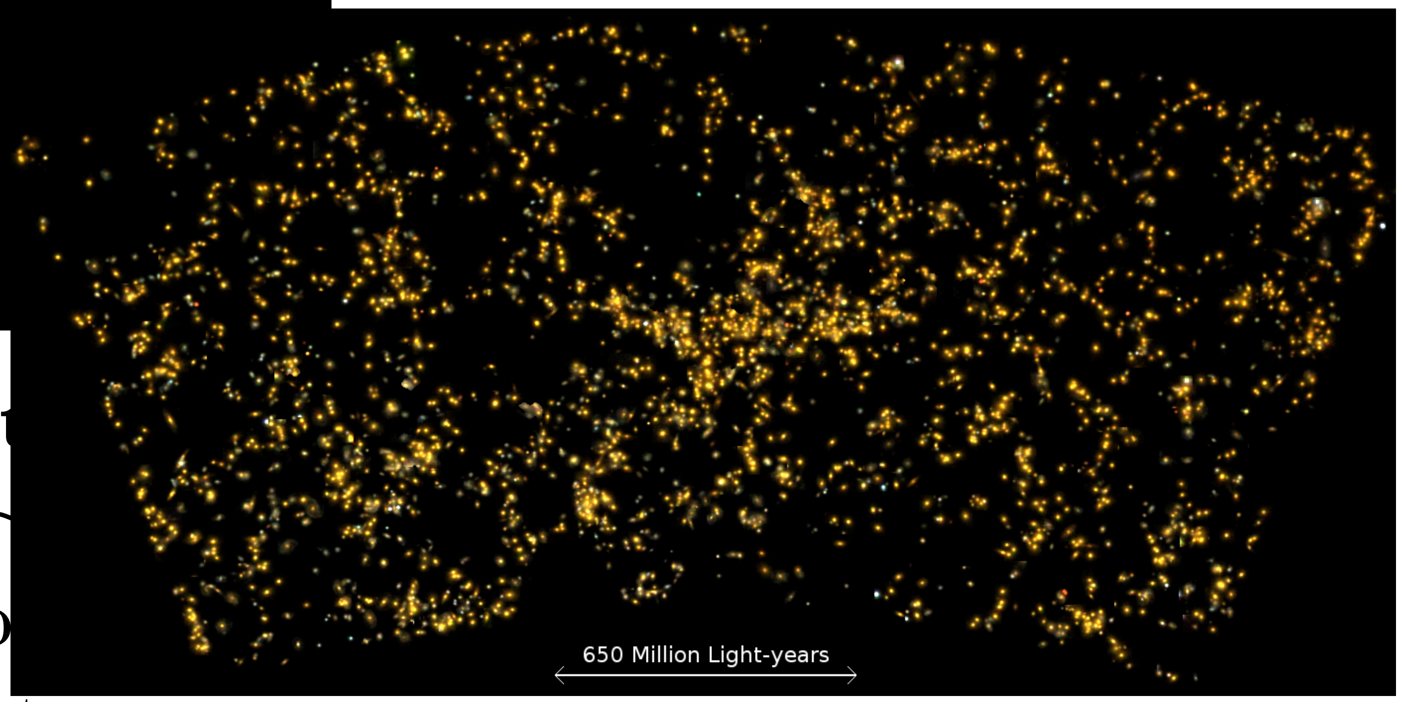


DESI collaboration



Einasto et al ’94, ’97, ‘16





Cosmic Web (supercluster - void net
Superclusters: 1% of the volume and ~
mass of the Universe. **Voids:** 70 – 90% o
and 15 % of mass, ~ 10 – 30%: supercluster outskirts

Bagchi et al. ’17, Sankhyayan et al. ‘23

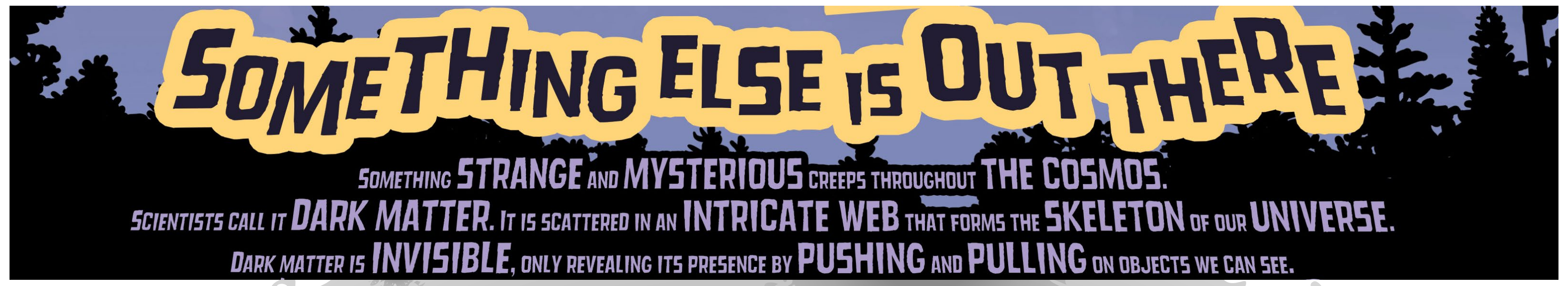
“Large-scale structure of the Universe”
1977 Tallinn symposium

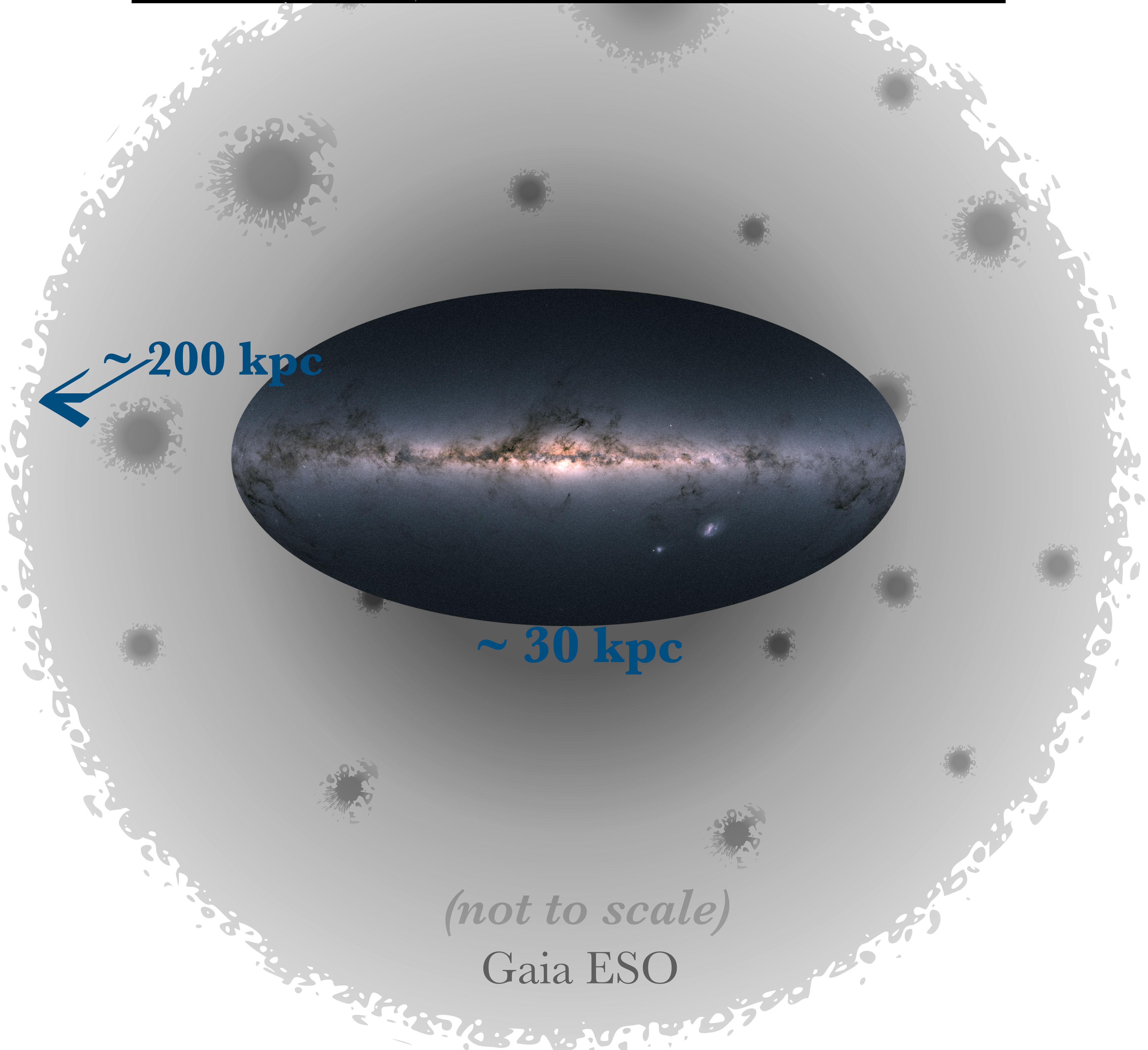
Einasto et al. ‘22

¿Ho’oleilana = individual BAO?

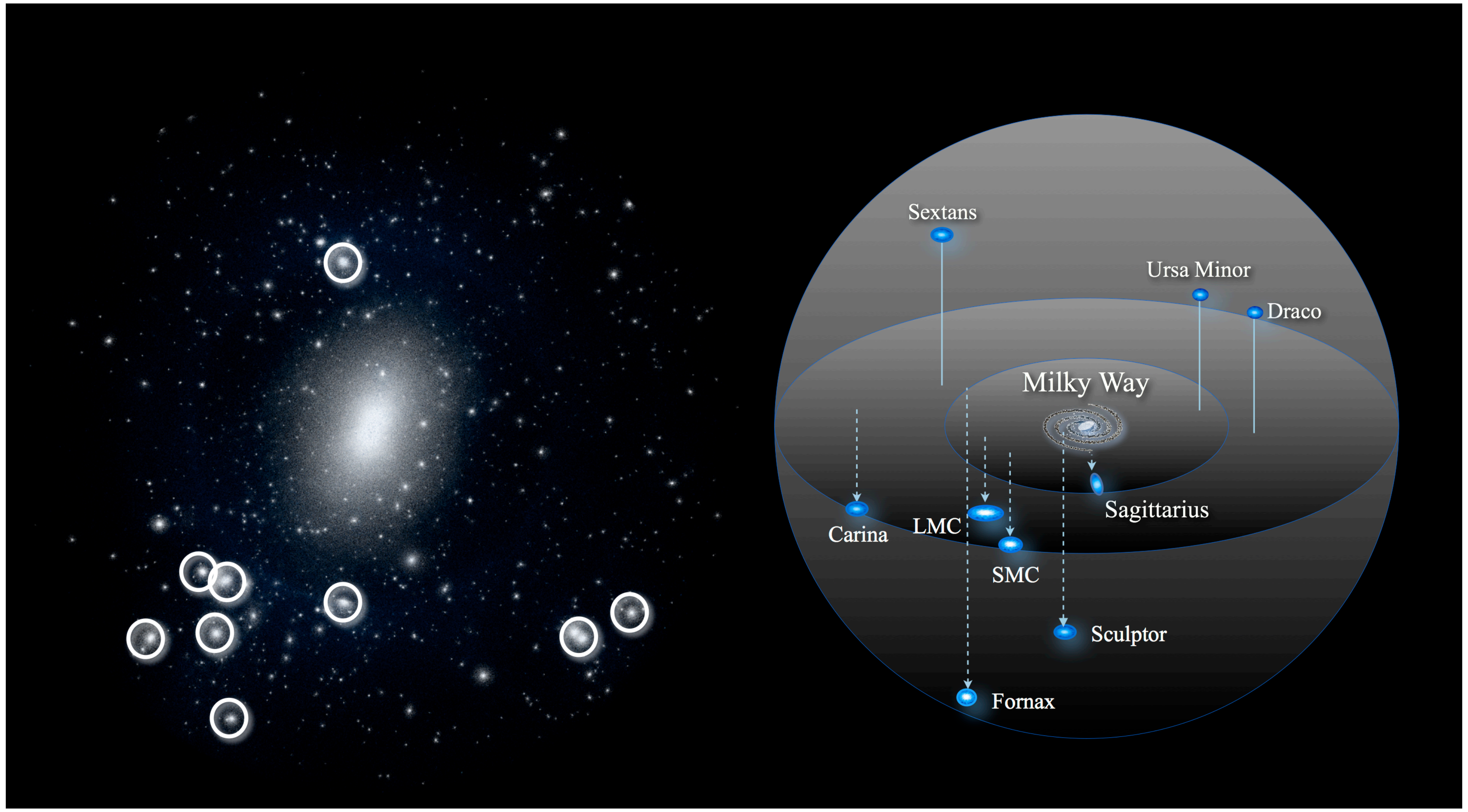
Tully et al. ‘23

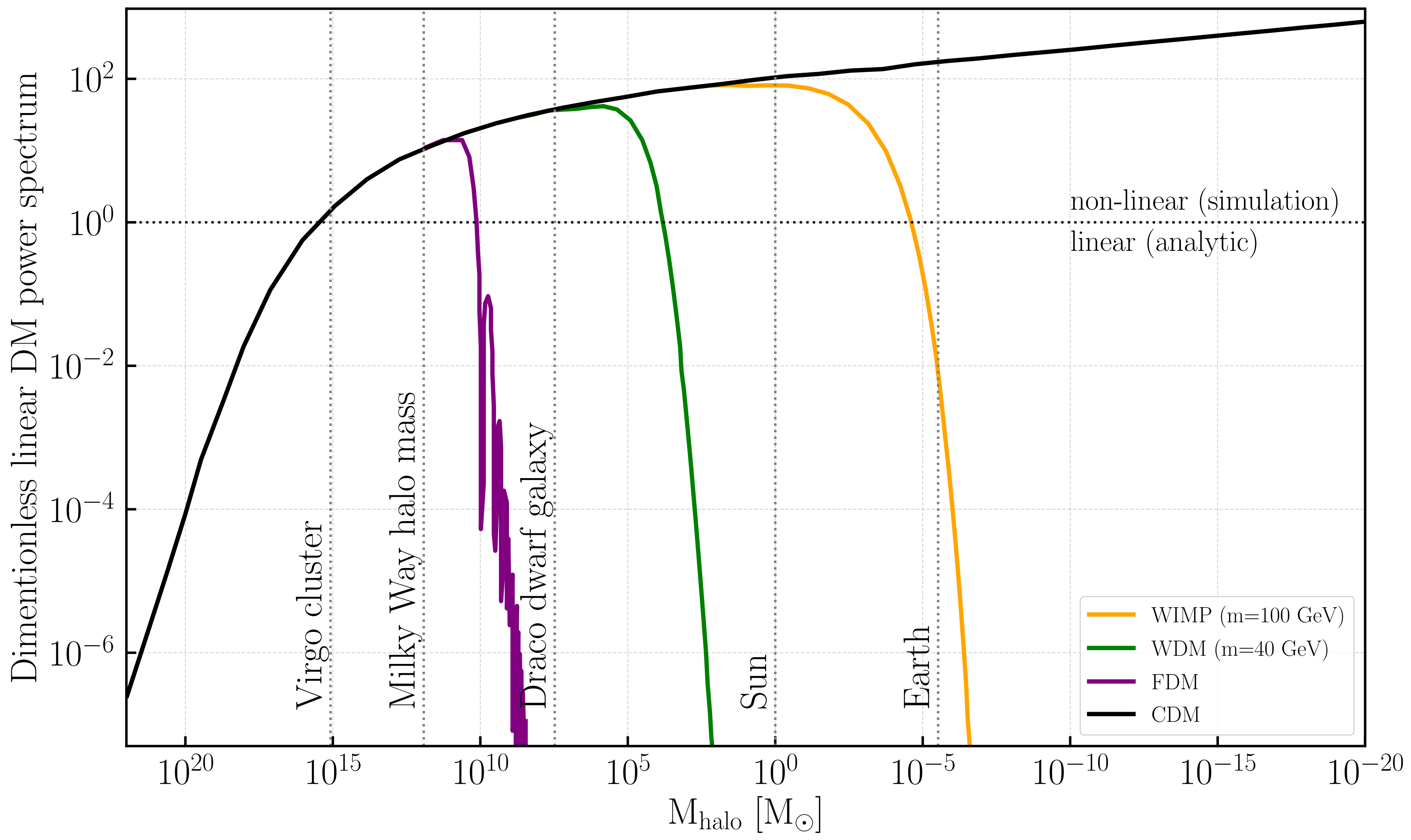
Observed visible & dark Milky Way @ 2024





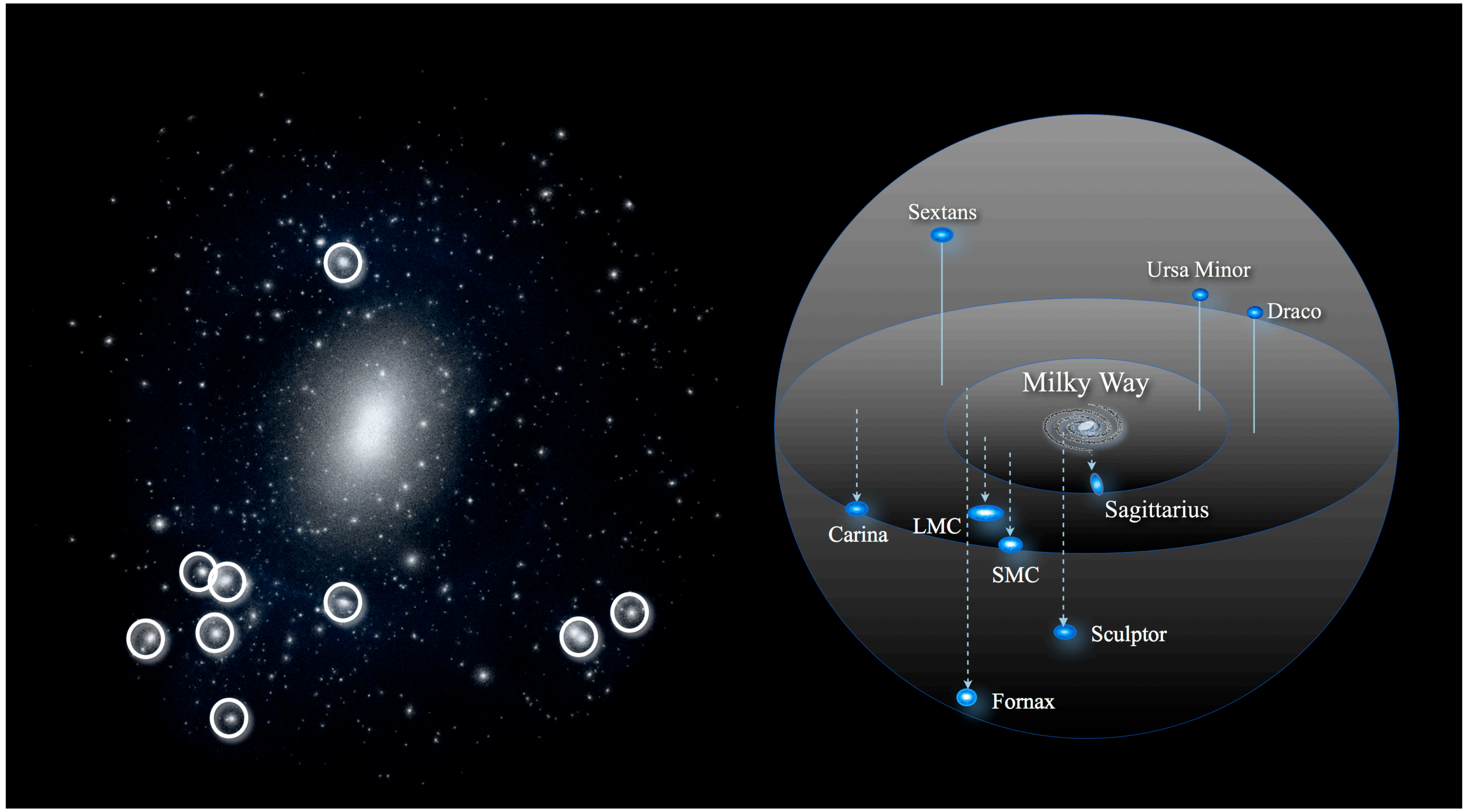
Gaia ESO





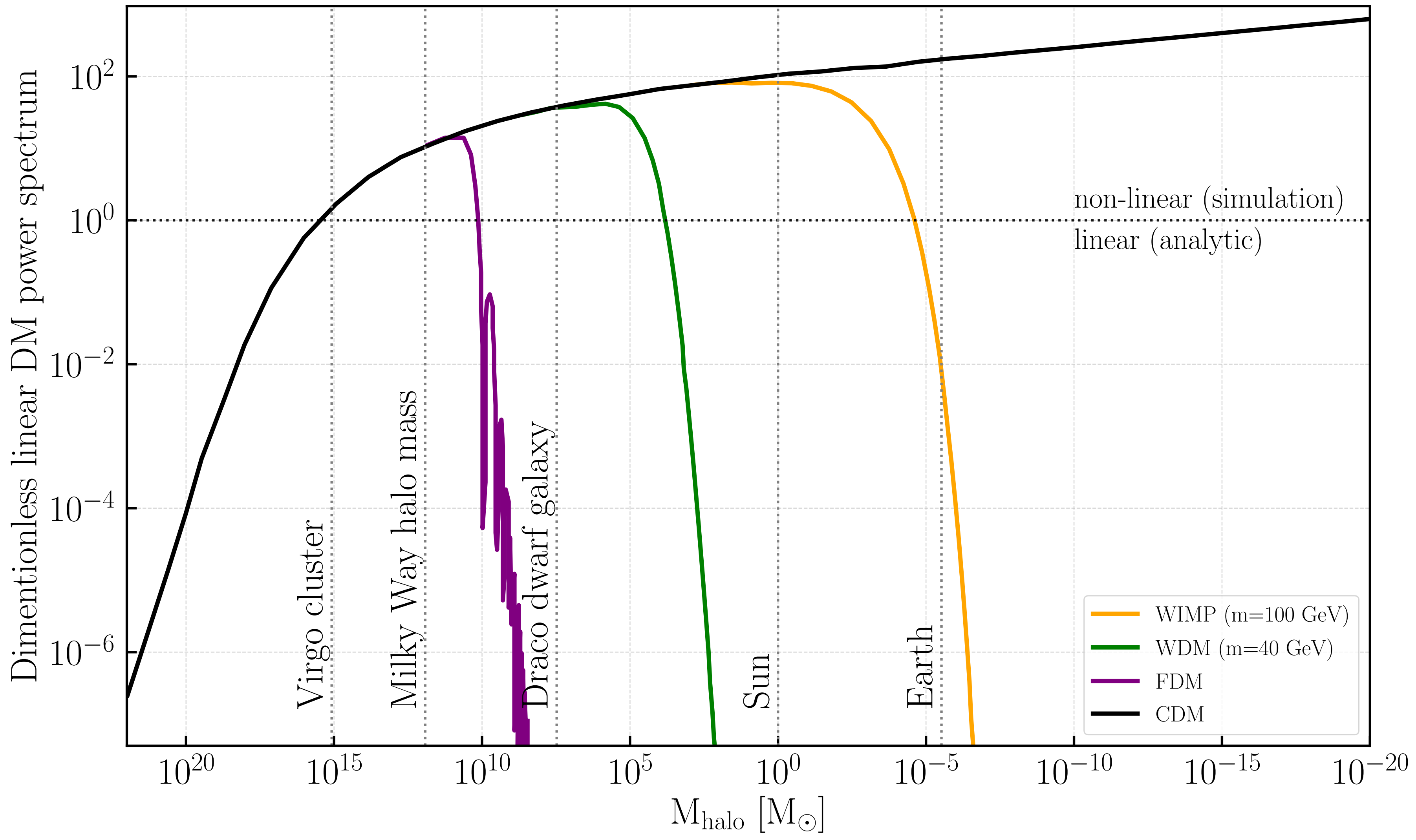
Goal

Search for dark subhalos (subhalos w/ masses smaller than ~10^9 M_{sun}) orbiting the Milky Way using stellar phase-space perturbations



Goal

Search for dark subhalos (subhalos w/ masses smaller than $\sim 10^9$ M_{sun}) orbiting the Milky Way using stellar phase-space perturbations



Why?

- Evidence of DM
- Assess the physics governing DM @ microscopic scales

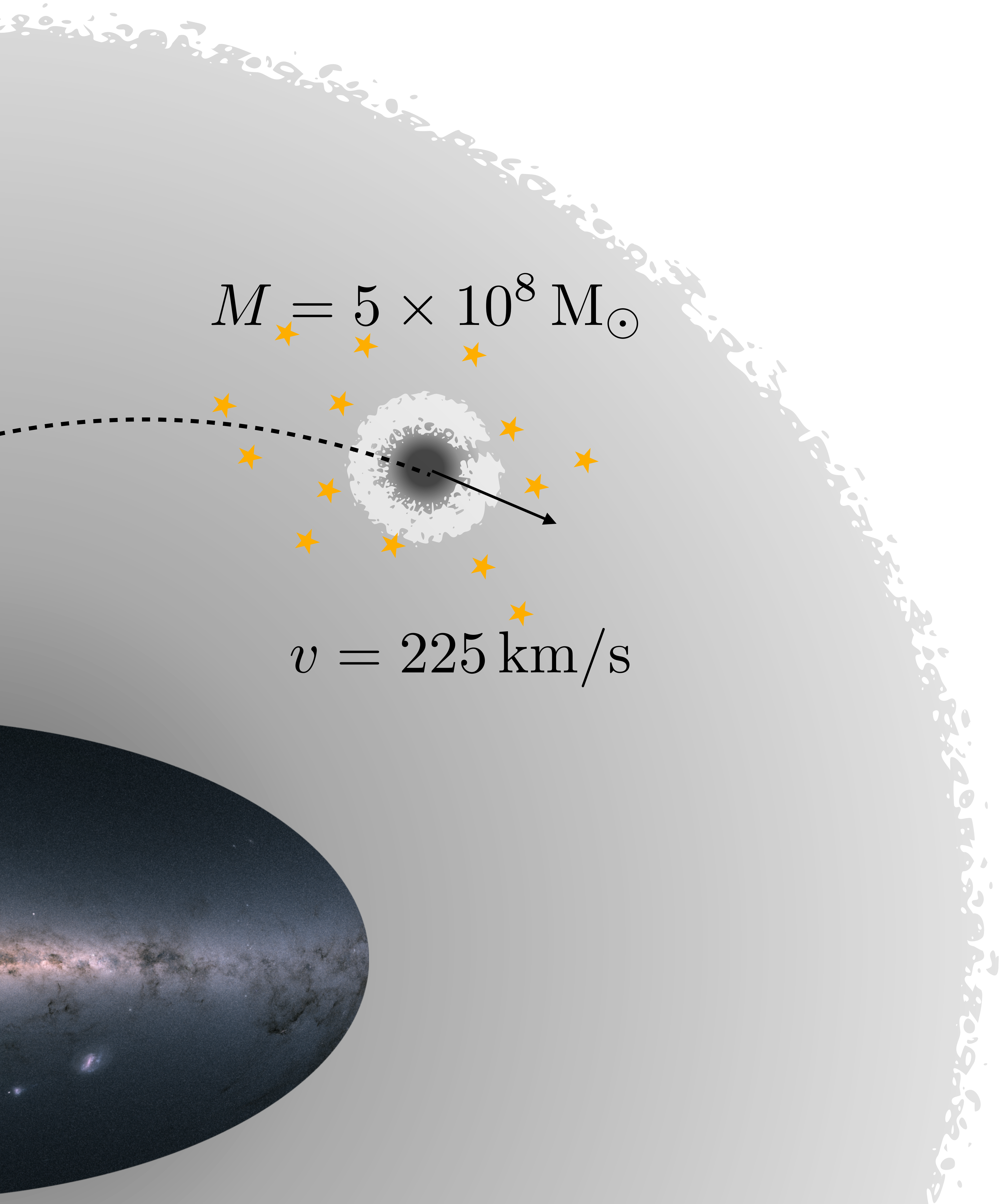
Searching for dark subhaloes in the Milky Way using …

- γ-ray instruments (may detect DM(WIMP) signals emitted therein)
- Stellar streams
- Pulsar timing arrays
- Stellar phase-space signatures:
    - Apparent perturbations due to (weak) lensing by subhalos
    - Real perturbations due to passing subhalos

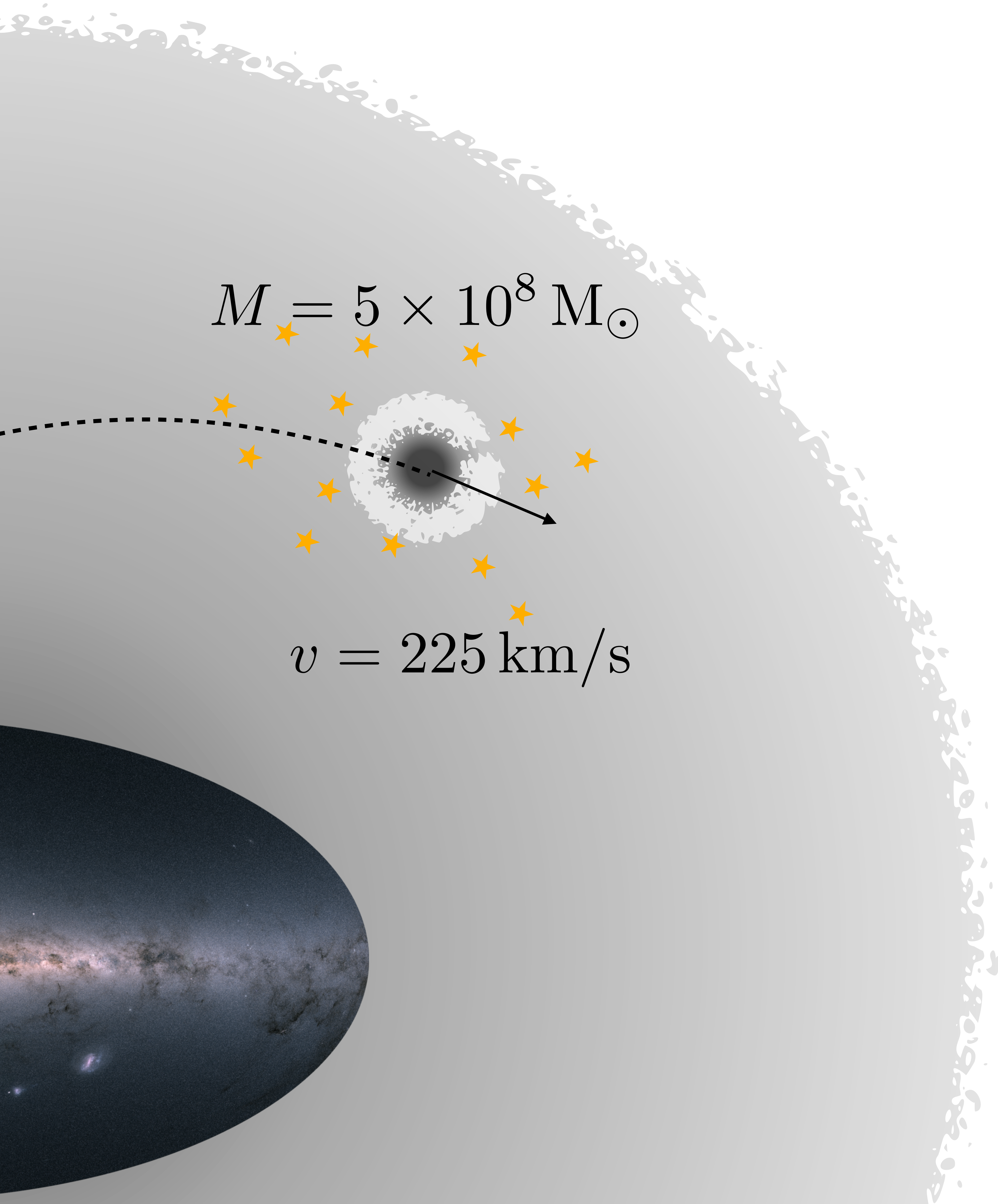
Searching for dark subhaloes in the Milky Way using …

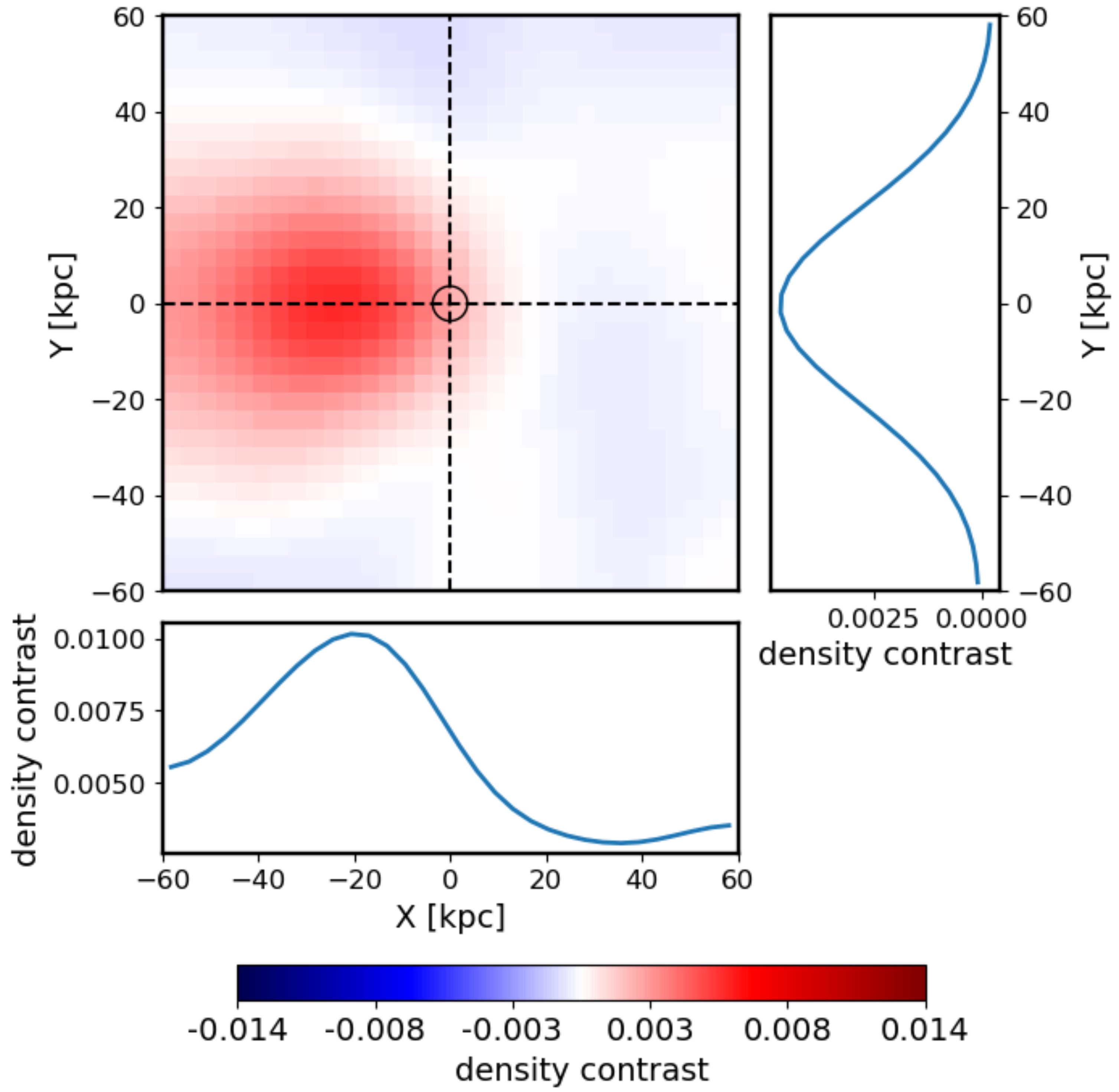
- γ-ray instruments (may detect DM(WIMP) signals emitted therein)
- Stellar streams
- Pulsar timing arrays
- Stellar phase-space signatures:
    - Apparent perturbations due to (weak) lensing by subhalos
    - **Real perturbations due to passing subhalos**

What's the signal?



What's the signal?

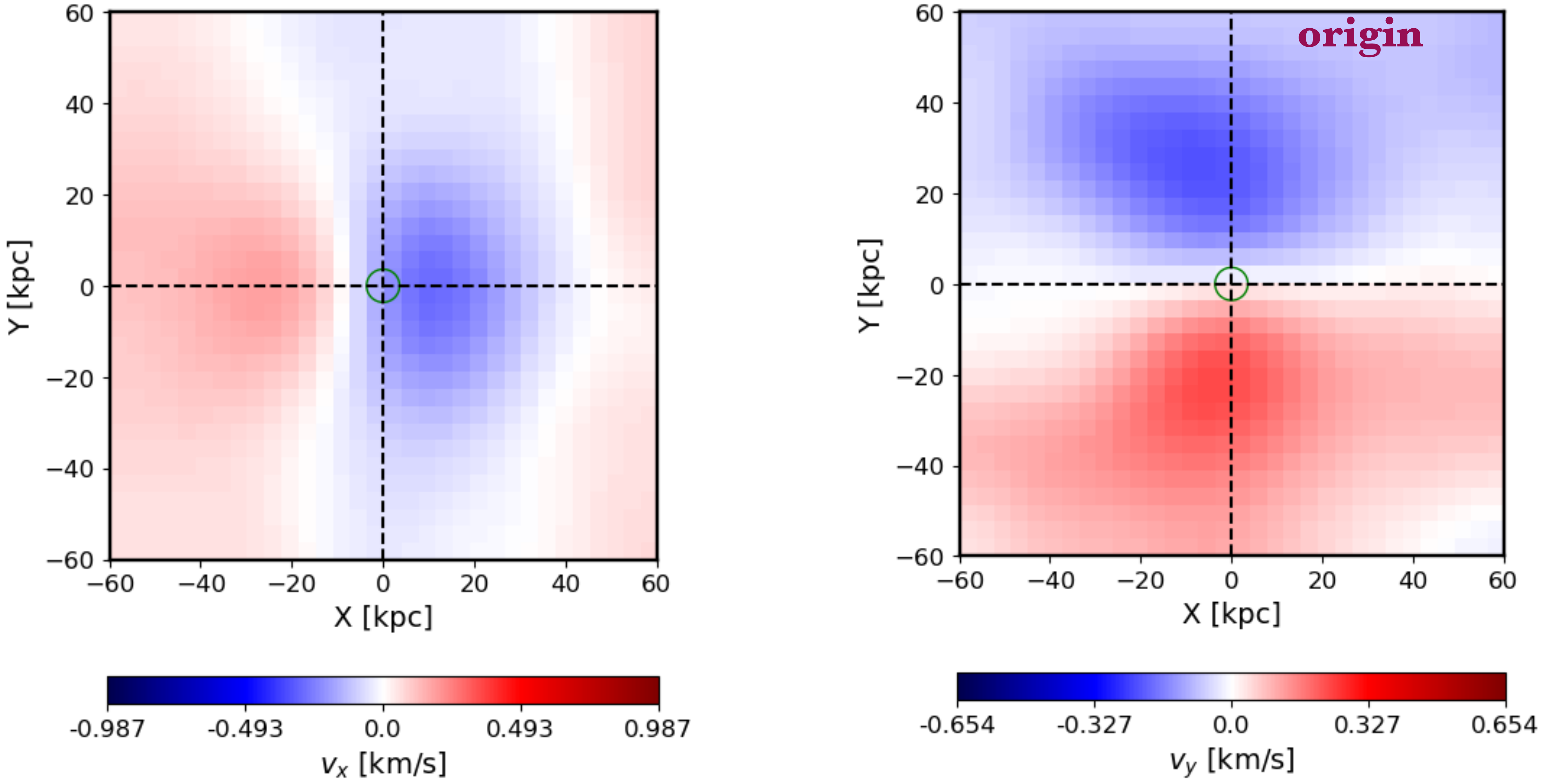




Density contrast ~ 1%

Velocity changes ~ 1 km/s

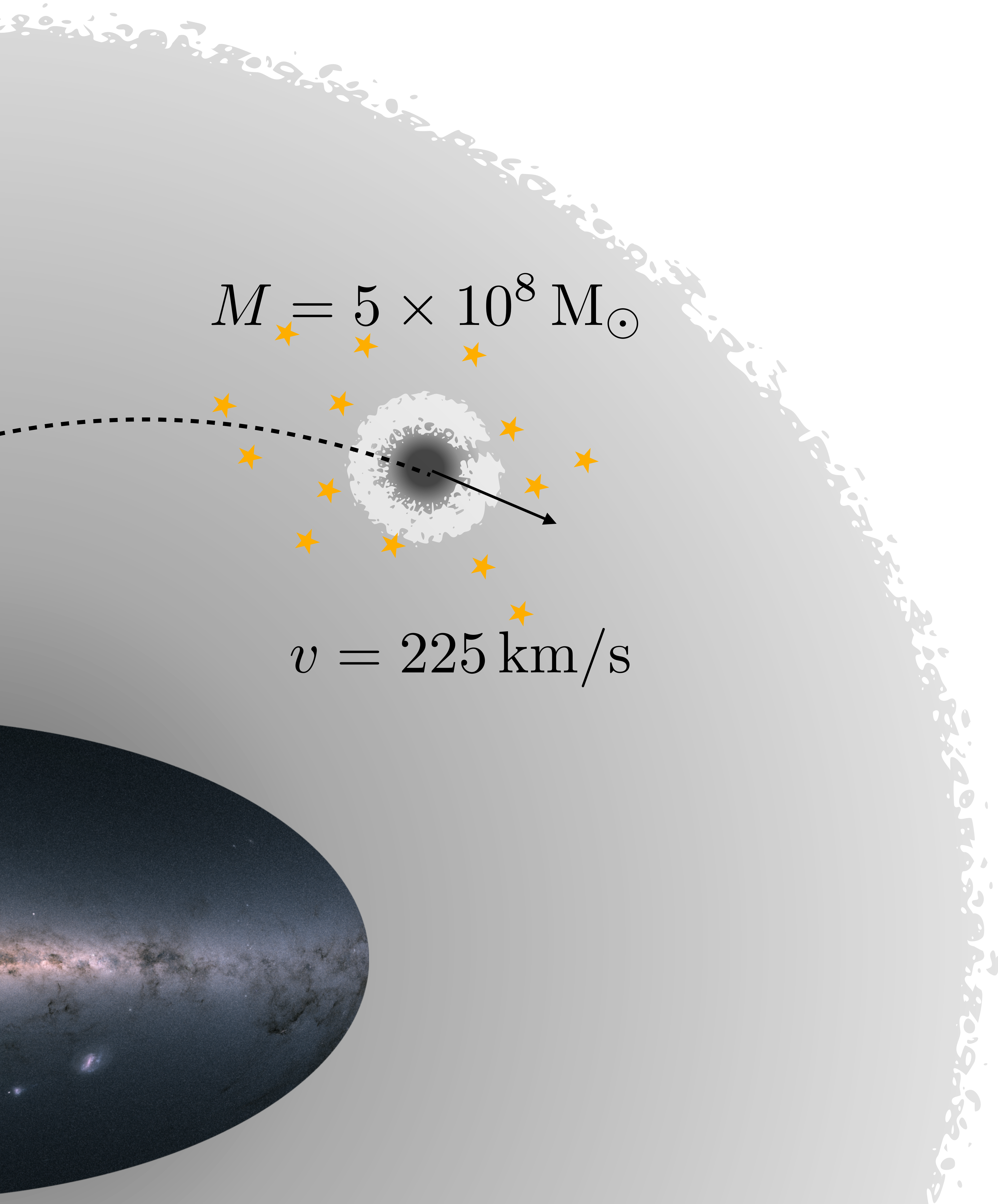
Subhalo located at origin

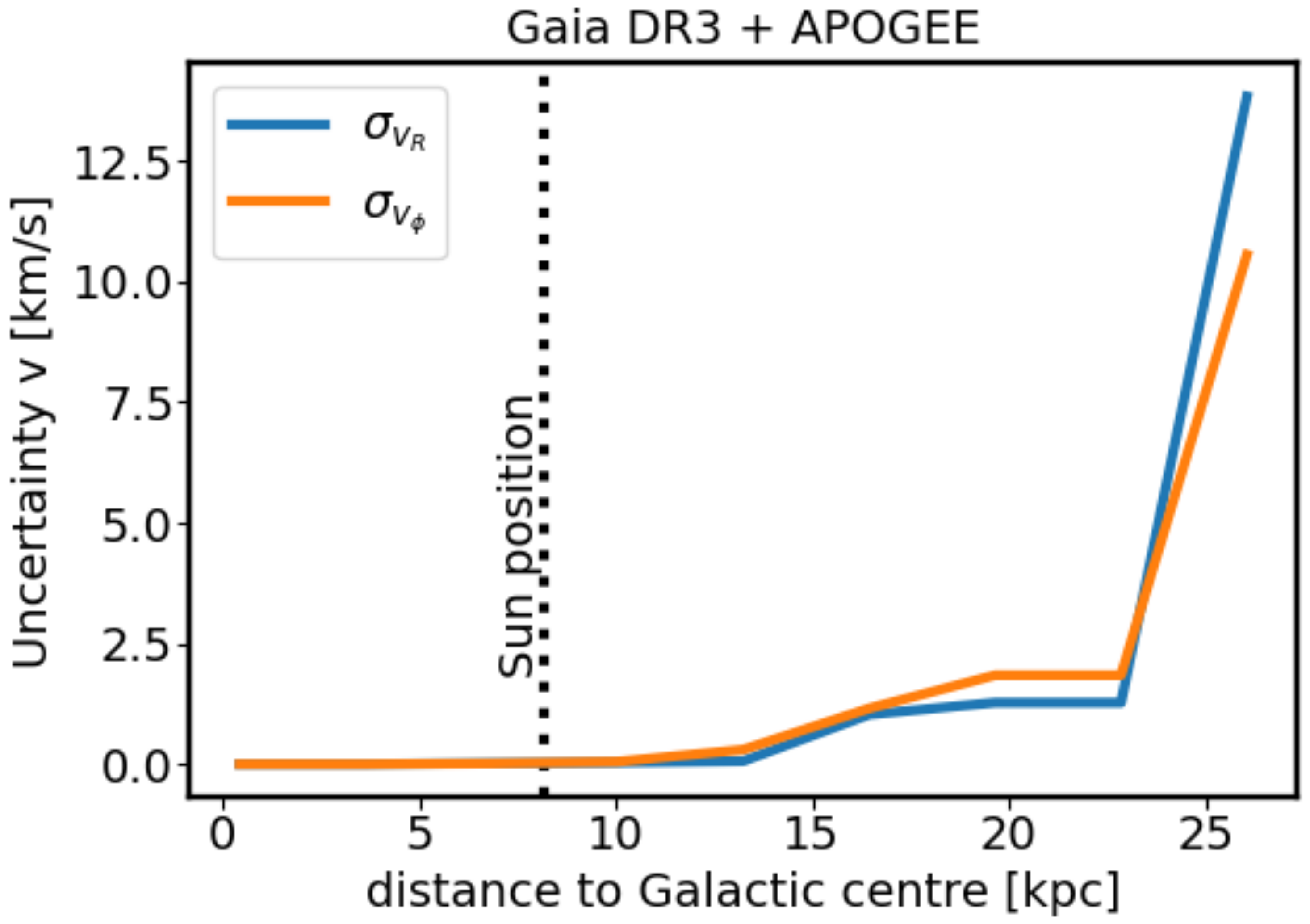


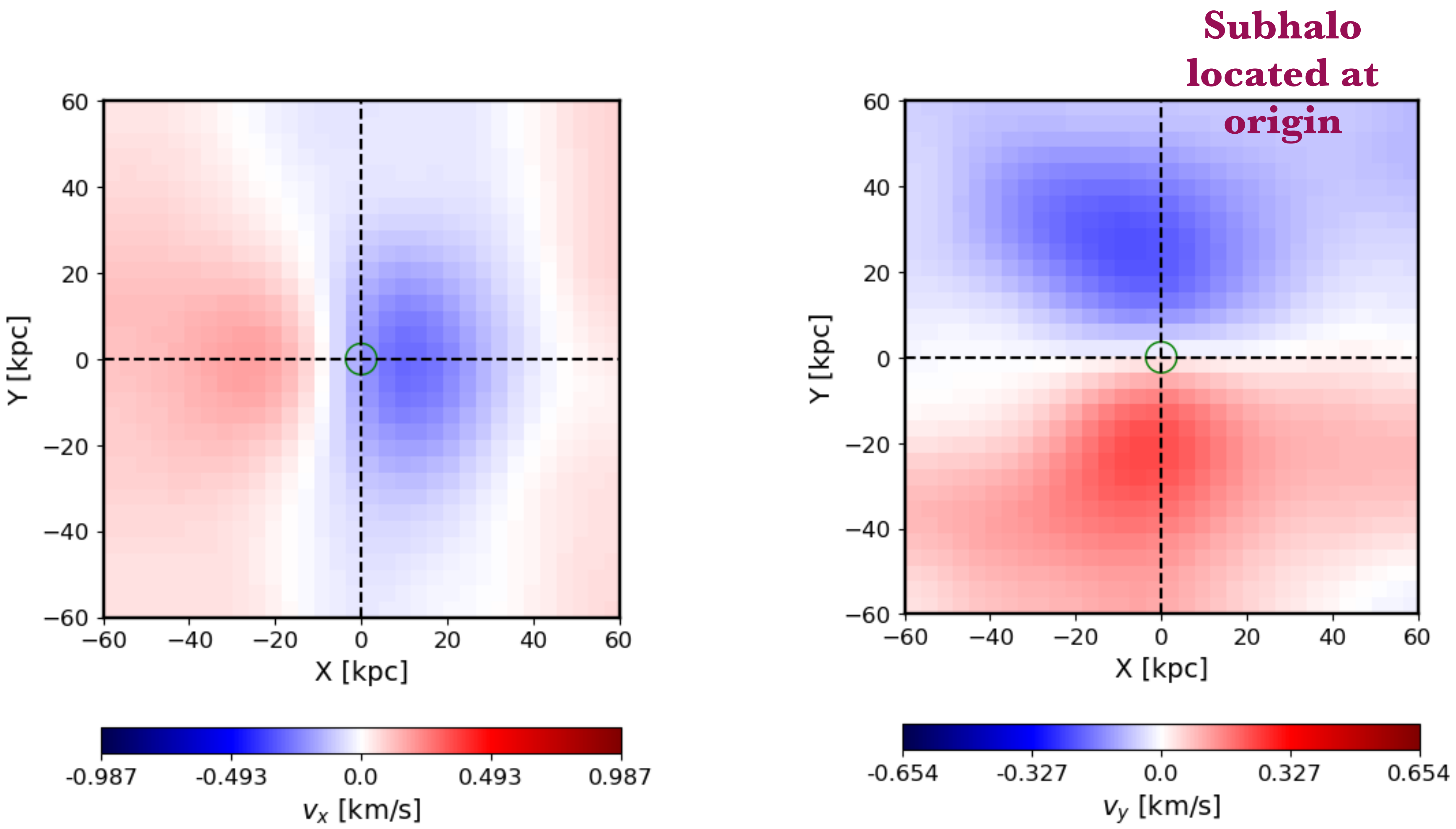
What's the signal?

Density contrast ~ 1%

Velocity changes ~ 1 km/s





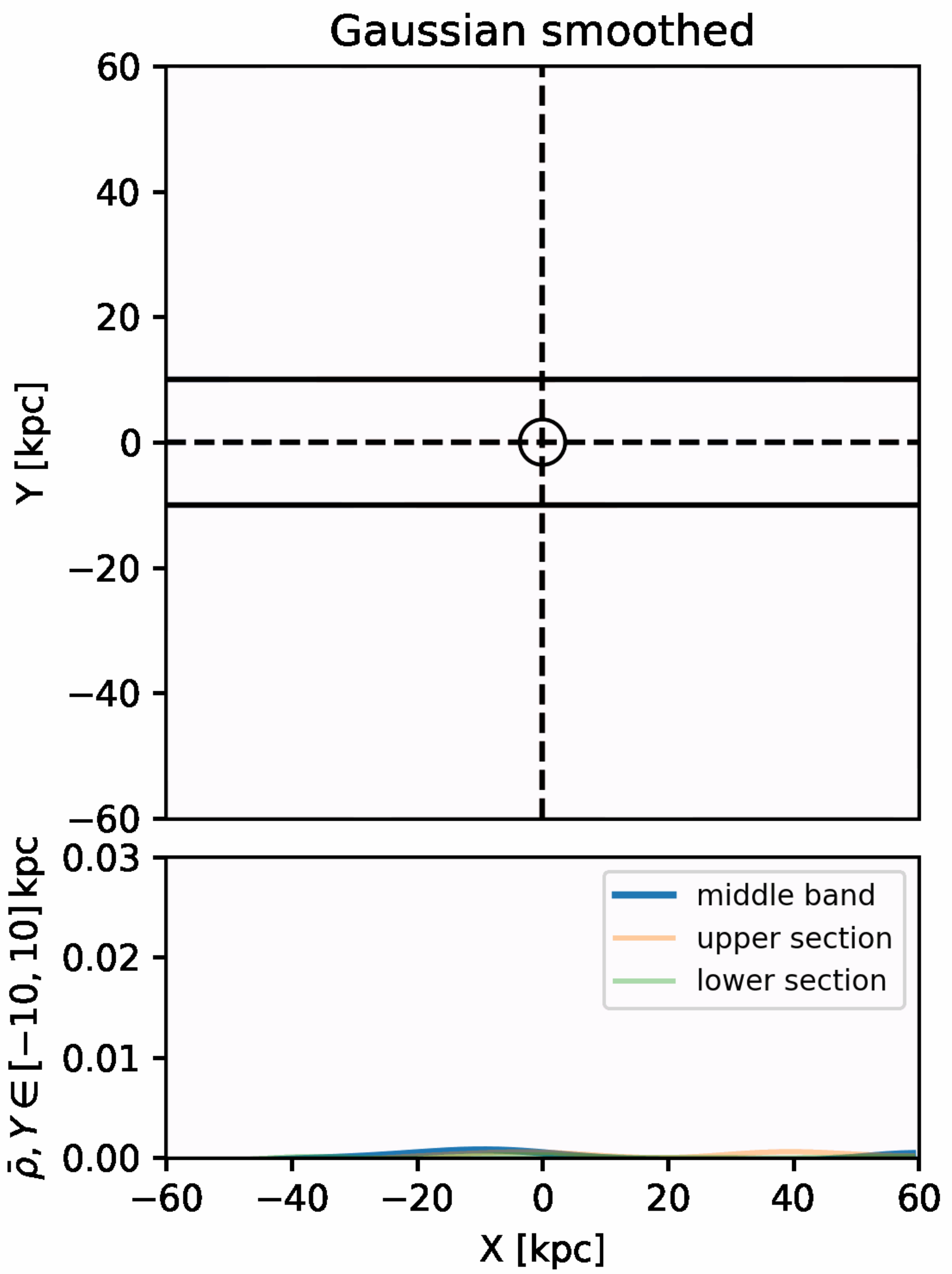


ML in windtunnel simulations:
Data generation

Sven Põder

Idealised simulation

Reference frame centred @ subhalo



In line with Foote et al ‘ 23

PKDGRAV3, 2x512^3

DM & stars have uniform density/ Maxwell velocity distribution with values as expected @ 30 kpc

1 snapshot = 20 GB

Plummer sphere

$$M = 5 \times 10^8, 10^8, 5 \times 10^7, 0\,\mathrm{M}_\odot$$

$$v = 225\,\mathrm{km/s}$$

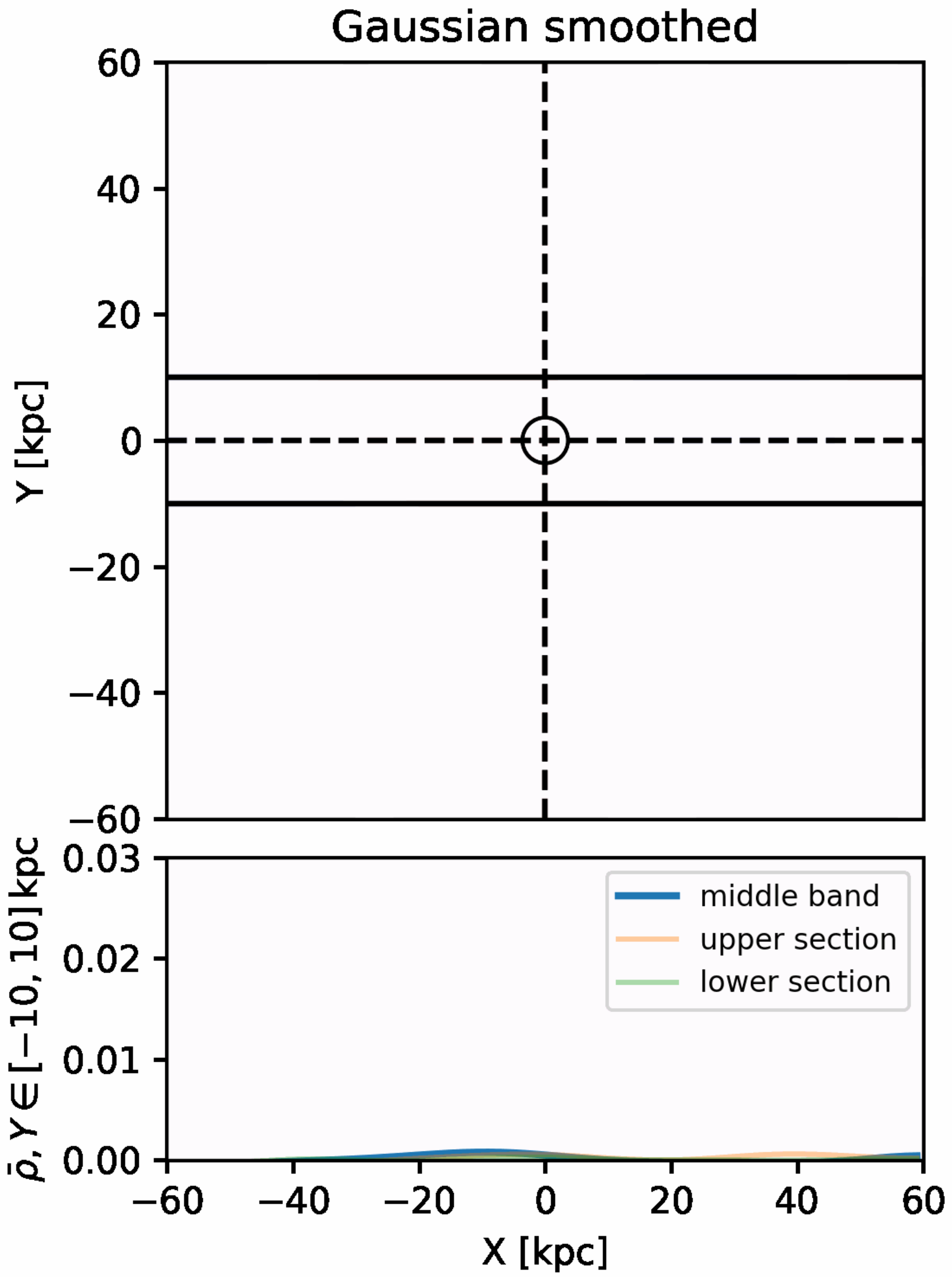
ML in windtunnel simulations: Data generation

On the detection of stellar wakes in the Milky Way: a deep learning approach

Sven Põder[1,2], Joosep Pata[1], María Benito[3], Isaac Alonso Asensio[4,5], and Claudio Dalla Vecchia[4,5]

Idealised simulation

Reference frame centred @ subhalo



In line with Foote et al ' 23

PKDGRAV3, 2x512^3

DM & stars have uniform density/ Maxwell velocity distribution with values as expected @ 30 kpc

1 snapshot = 20 GB

Plummer sphere

$M = 5 \times 10^8, 10^8, 5 \times 10^7, 0\,\mathrm{M}_\odot$

$v = 225\,\mathrm{km/s}$

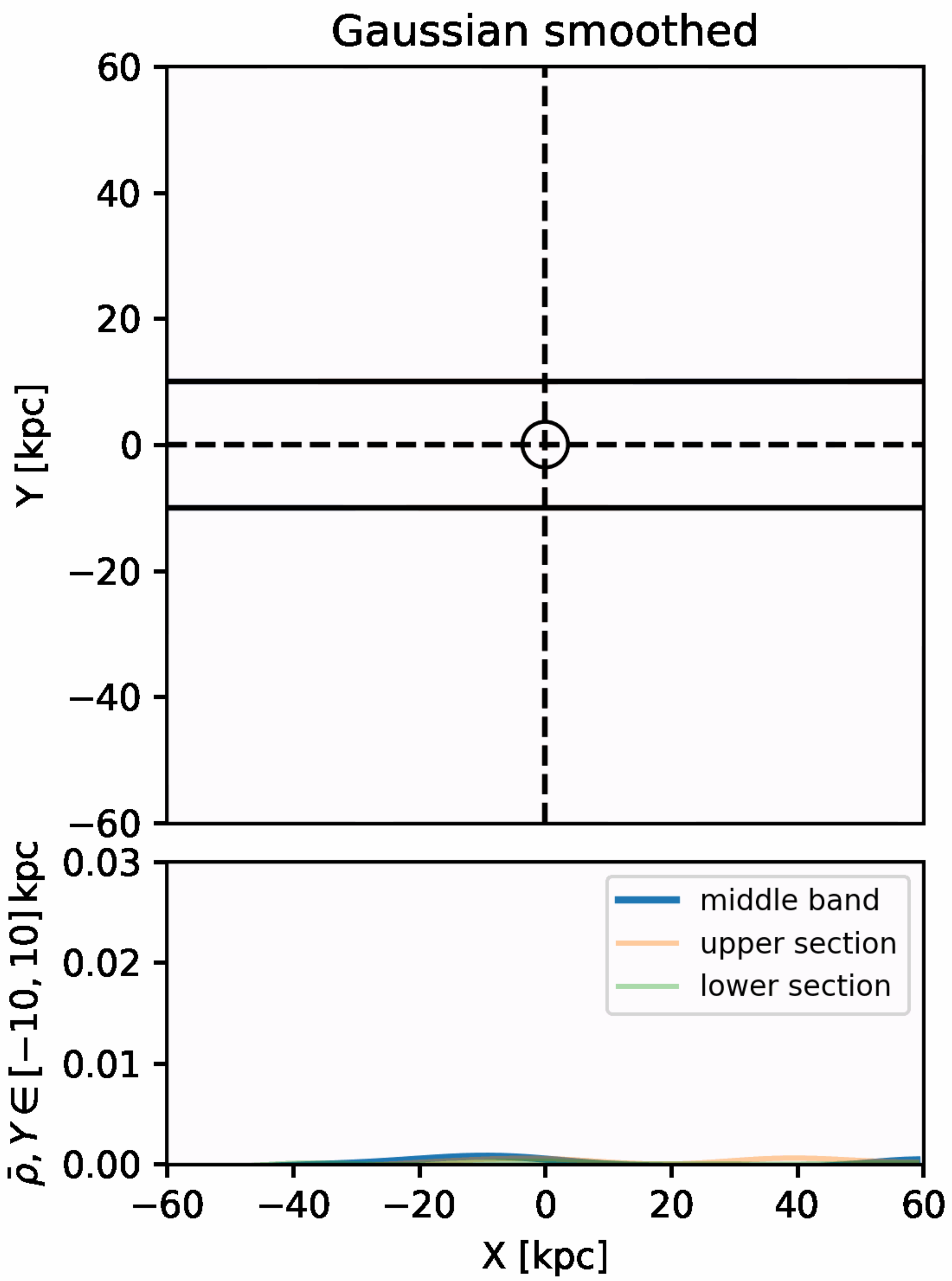
ML in windtunnel simulations:
Data generation

On the detection of stellar wakes in the Milky Way: a deep learning approach

Sven Põder[12], Joosep Pata[1], María Benito[3], Isaac Alonso Asensio[45], and Claudio Dalla Vecchia[45]

Idealised simulation

Reference frame centred @ subhalo



In line with Foote et al ‘ 23

PKDGRAV3, 2x512³

DM & stars have uniform density/ Maxwell velocity distribution with values as expected @ 30 kpc

1 snapshot = 20 GB

Plummer sphere

$$M = 5\times10^8, 10^8, 5\times10^7, 0\,\mathrm{M}_\odot$$

$$v = 225\,\mathrm{km/s}$$

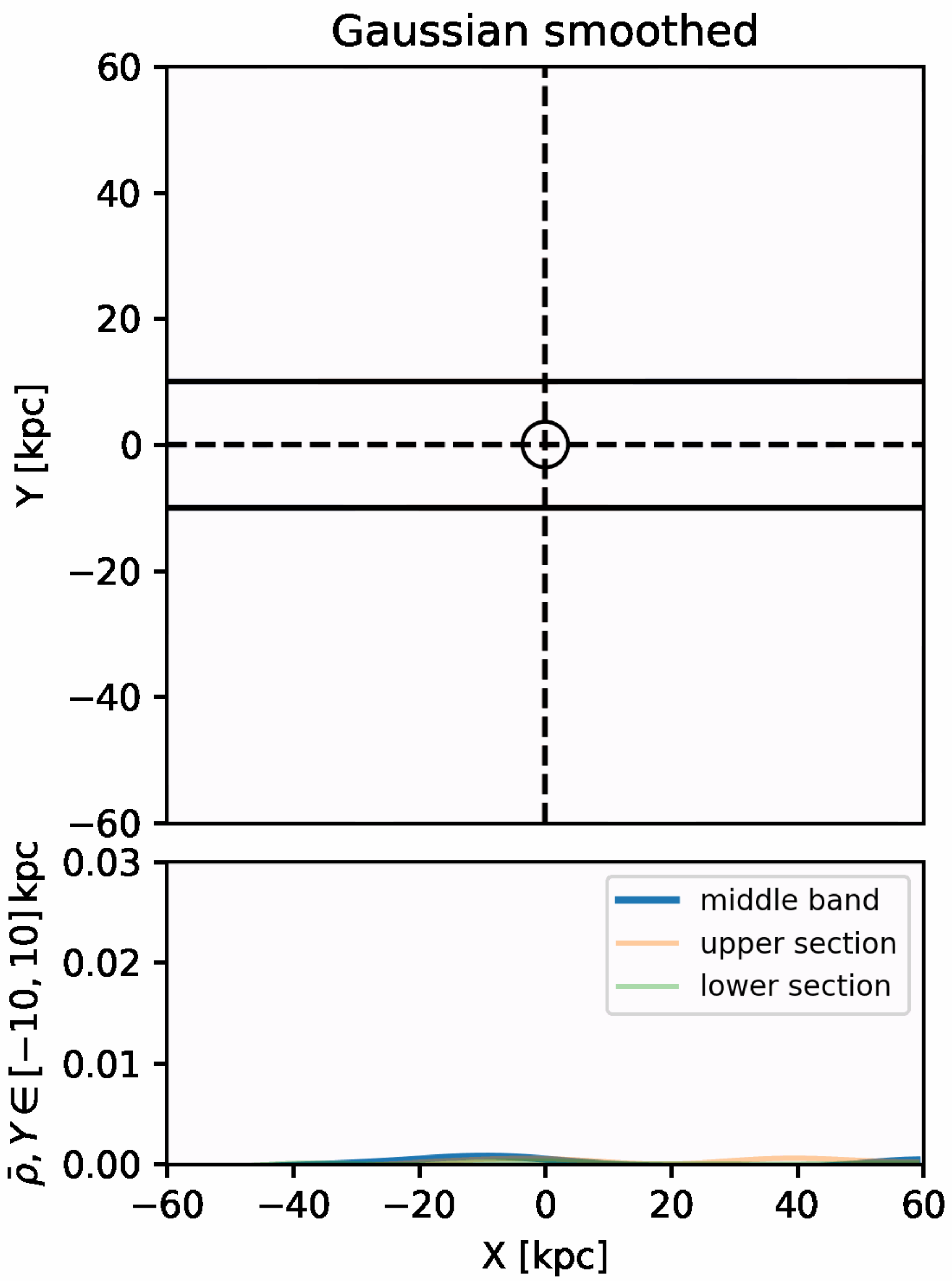
ML in windtunnel simulations: Data generation

On the detection of stellar wakes in the Milky Way: a deep learning approach

Sven Põder[1,2], Joosep Pata[1], María Benito[3], Isaac Alonso Asensio[4,5], and Claudio Dalla Vecchia[4,5]

Idealised simulation → “Observation” (selection) effect → **Sample**

Reference frame centred @ subhalo



PKDGRAV3, 2x512³

DM & stars have uniform density/ Maxwell velocity distribution with values as expected @ 30 kpc

1 snapshot = 20 GB

Plummer sphere

$M = 5 \times 10^8, 10^8, 5 \times 10^7, 0\,\mathrm{M}_\odot$

$v = 225\,\mathrm{km/s}$

1% of the particles

1.3×10^6 particles

$\times 6$ features

$\sim 8 \times 10^6$ raw values

In line with Foote et al ‘ 23

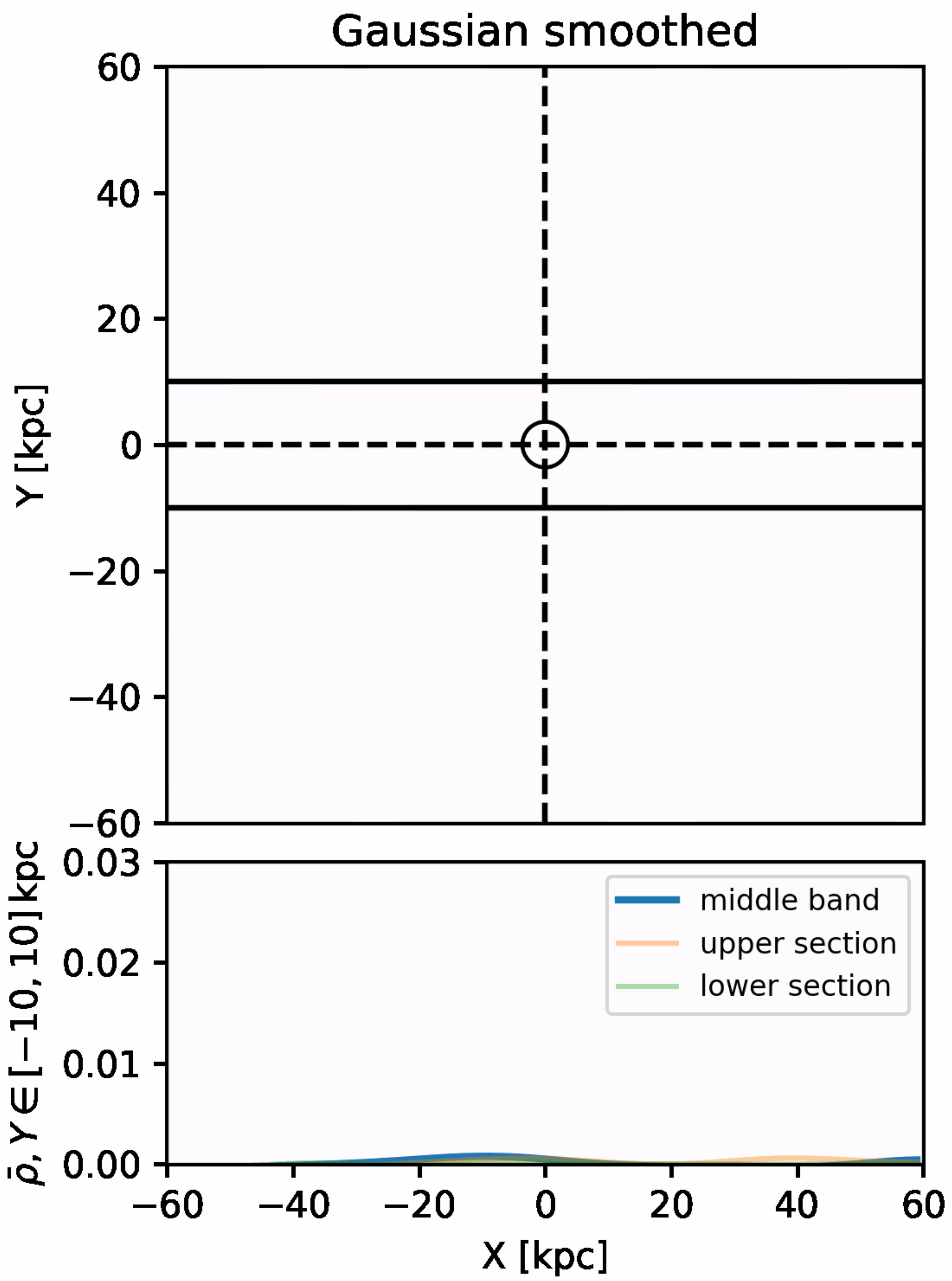
ML in windtunnel simulations:
Data generation

On the detection of stellar wakes in the Milky Way: a deep learning approach

Sven Põder[12], Joosep Pata[1], María Benito[3], Isaac Alonso Asensio[45], and Claudio Dalla Vecchia[45]

Idealised simulation → "Observation" (selection) effect → **Sample** → Effective observables → **Image**

Reference frame centred @ subhalo



In line with Foote et al ' 23

PKDGRAV3, 2x512³

DM & stars have uniform density/ Maxwell velocity distribution with values as expected @ 30 kpc

1 snapshot = 20 GB

Plummer sphere

$$M = 5\times10^8, 10^8, 5\times10^7, 0\,\mathrm{M}_\odot$$

$$v = 225\,\mathrm{km/s}$$

1% of the particles

1.3×10^6 particles

$\times 6$ features

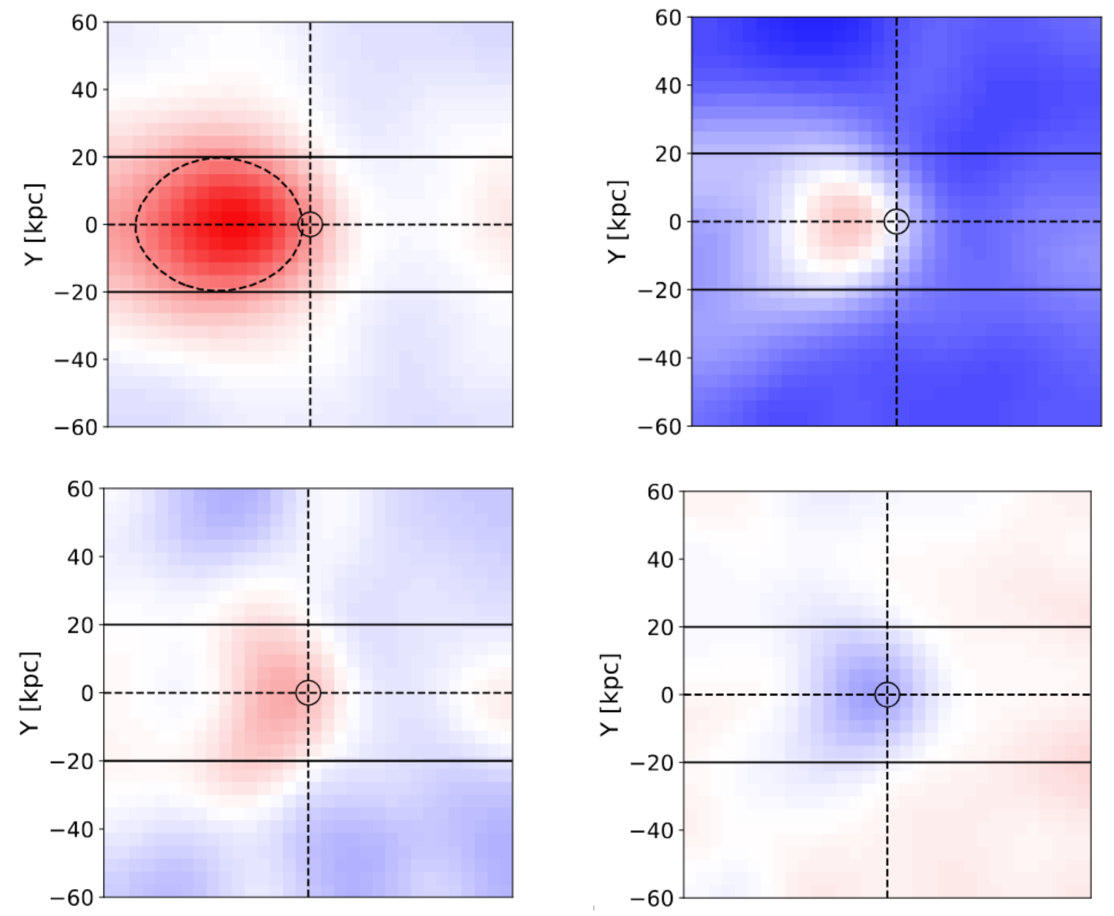
$\sim 8 \times 10^6$ raw values

32 x 32 x 3 binning

x

4 features

$\sim 1.2 \times 10^4$ effective observables



100 images = 9.4 MB

Training	Validation	Testing
50 %	33%	17%
2400	1600	800

ML in windtunnel simulations:
Data generation

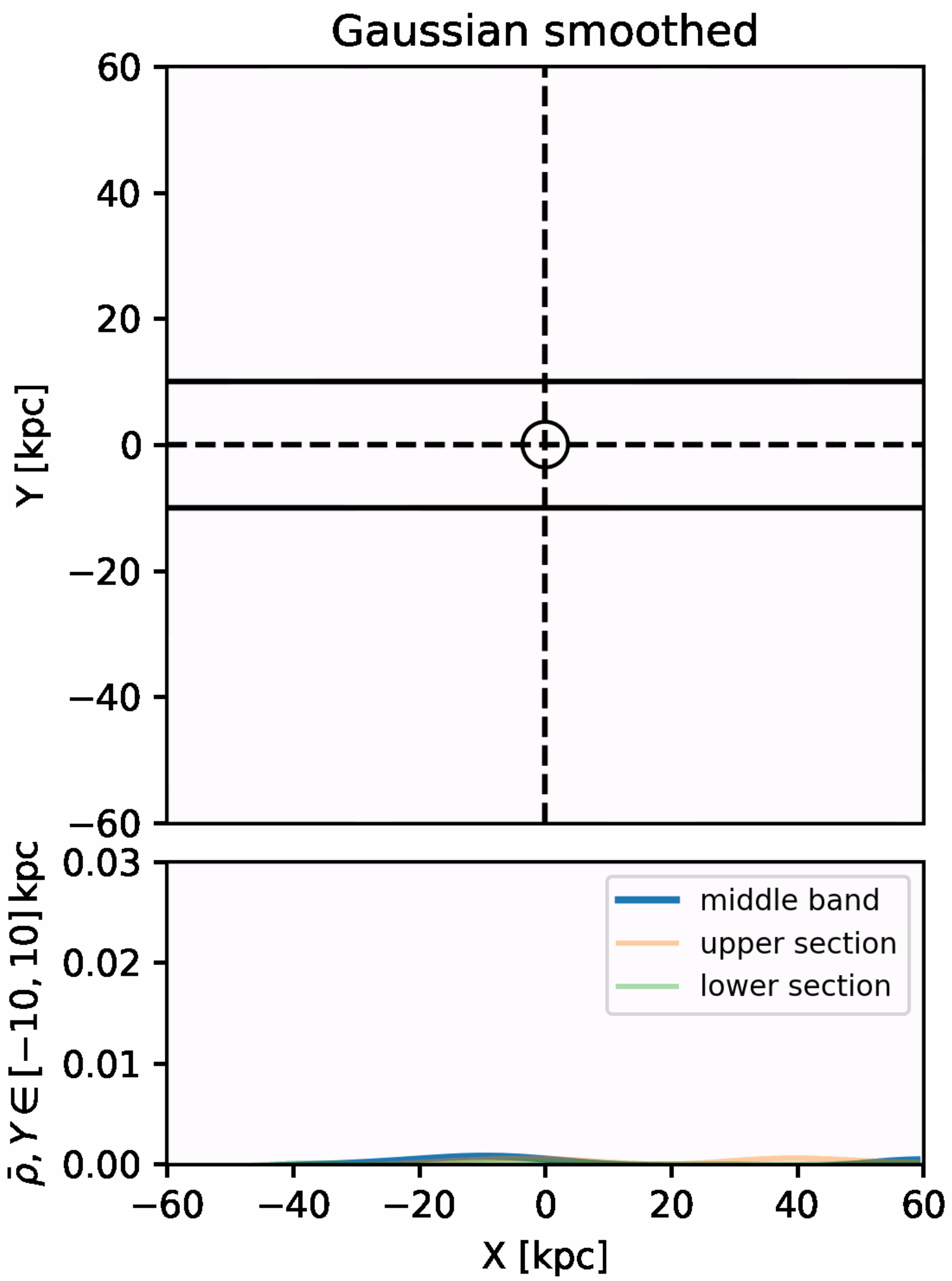
On the detection of stellar wakes in the Milky Way: a deep learning approach

Sven Põder[12], Joosep Pata[1], María Benito[3], Isaac Alonso Asensio[45], and Claudio Dall

ML catalogue

Idealised simulation → “Observation” (selection) effect → **Sample** → Effective observables → **Image**

Reference frame centred @ subhalo



In line with Foote et al ‘ 23

PKDGRAV3, 2x512³

DM & stars have uniform density/ Maxwell velocity distribution with values as expected @ 30 kpc

1 snapshot = 20 GB

Plummer sphere

$M = 5 \times 10^8, 10^8, 5 \times 10^7, 0\,\mathrm{M}_\odot$

$v = 225\,\mathrm{km/s}$

1% of the particles

1.3×10^6 particles

$\times 6$ features

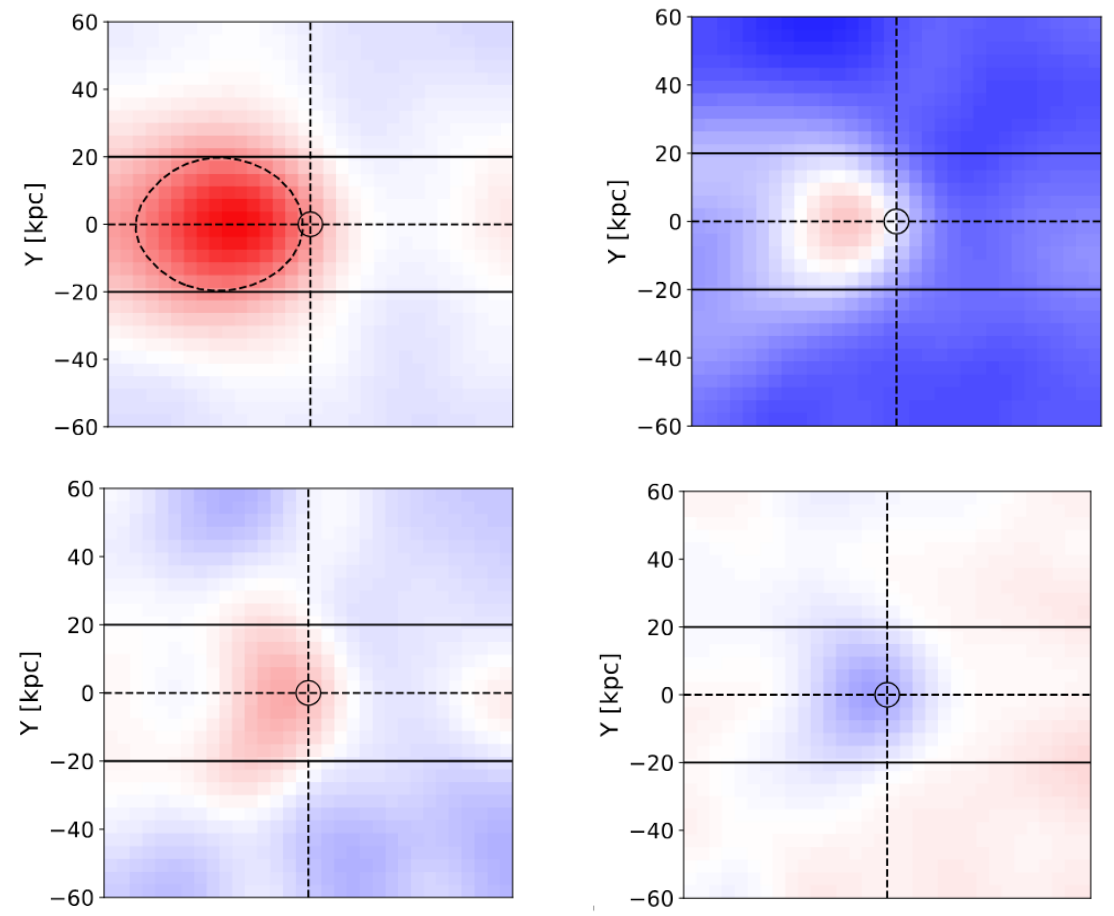
$\sim 8 \times 10^6$ raw values

32 x 32 x 3 binning

x

4 features

$\sim 1.2 \times 10^4$ effective observables



100 images = 9.4 MB

Training	Validation	Testing
50 %	33%	17%
2400	1600	800

ML in windtunnel simulations: Results

Binary classifier

Harmonic Network

Adam optimiser + binary focal cross entropy loss function

Hyperparameter	Range	Final Value
number of z-slices	[1, 2, 3]	3
filters	[4, 128]	32
learning rate	[1e-8, 1e-2]	1.9602e-06
dropout	[0, 0.6]	0.49259
activation	[relu, selu]	relu
kernel of 1st layer	[3, 10]	9
kernel of 2nd layer	[1, 3]	2
extra layers	[0, 3]	1
filter expansion	[1, 16]	2

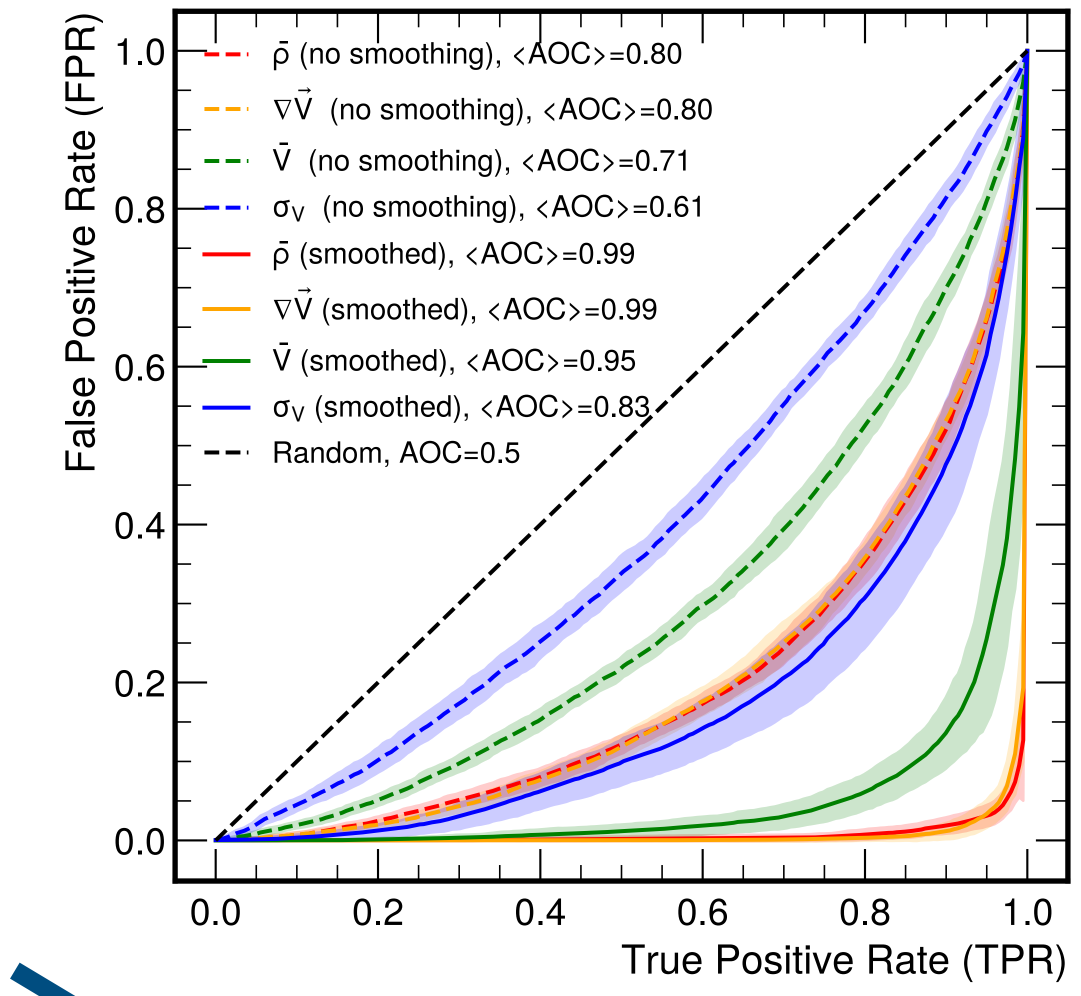
ML in windtunnel simulations: Results

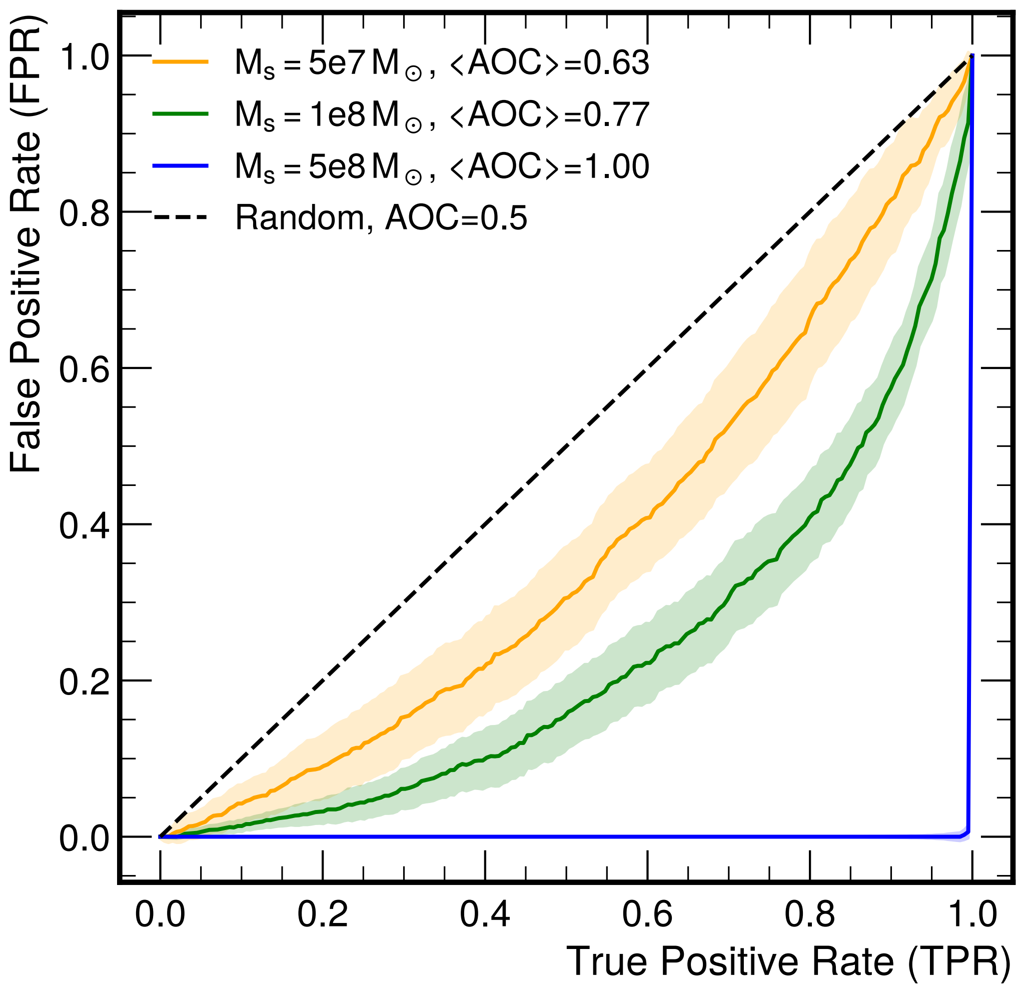
Binary classifier

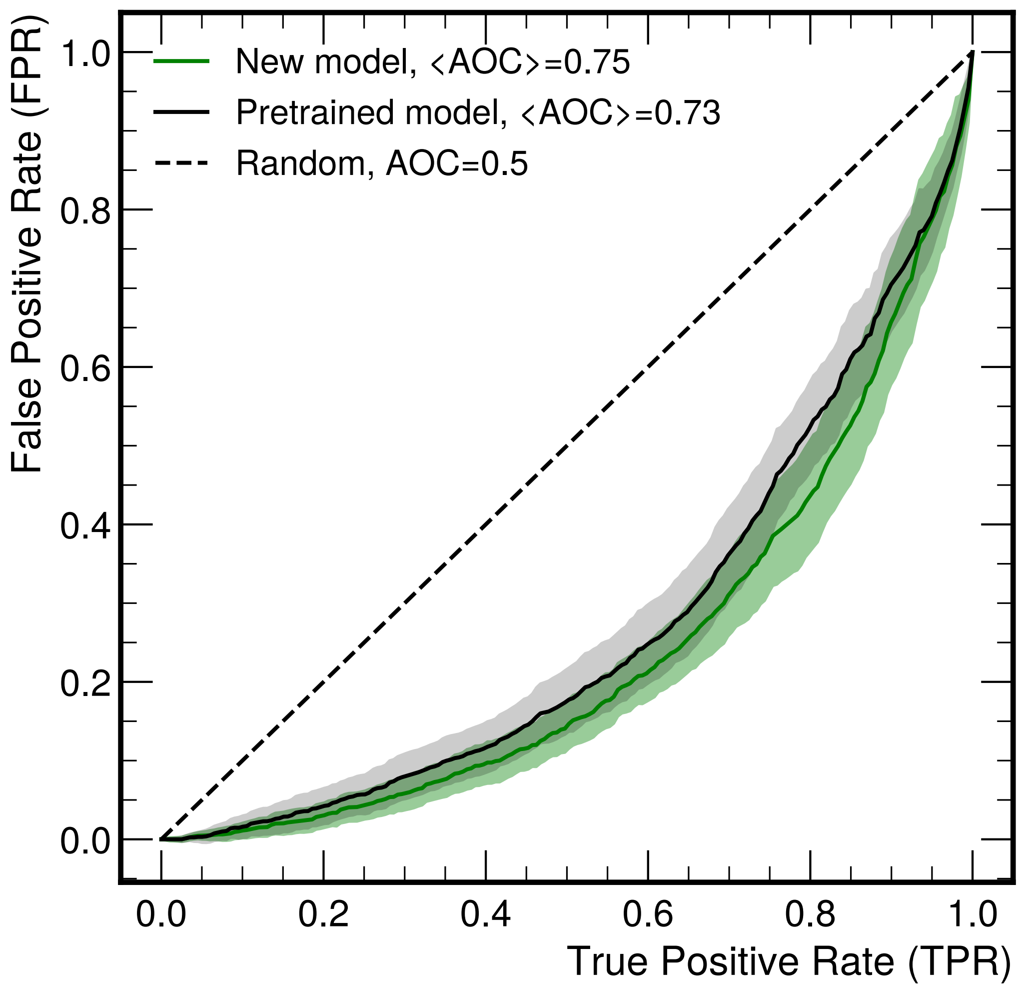
Harmonic Network

Adam optimiser + binary focal cross entropy loss function

Hyperparameter	**Range**	**Final Value**
number of z-slices	[1, 2, 3]	3
filters	[4, 128]	32
learning rate	[1e-8, 1e-2]	1.9602e-06
dropout	[0, 0.6]	0.49259
activation	[relu, selu]	relu
kernel of 1st layer	[3, 10]	9
kernel of 2nd layer	[1, 3]	2
extra layers	[0, 3]	1
filter expansion	[1, 16]	2







Smoothing plays a crucial role

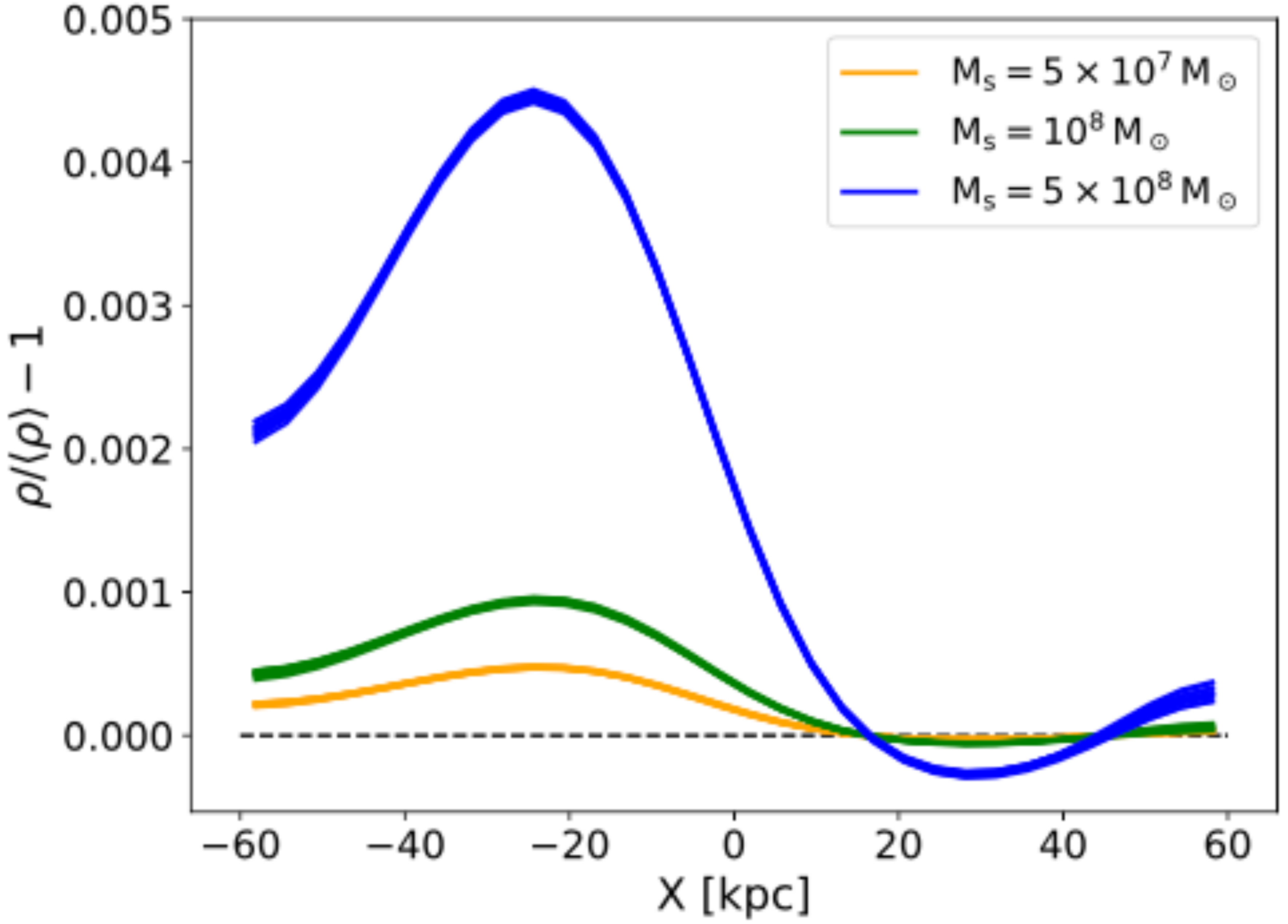
Overdensity/Velocity divergence yield maximal performance

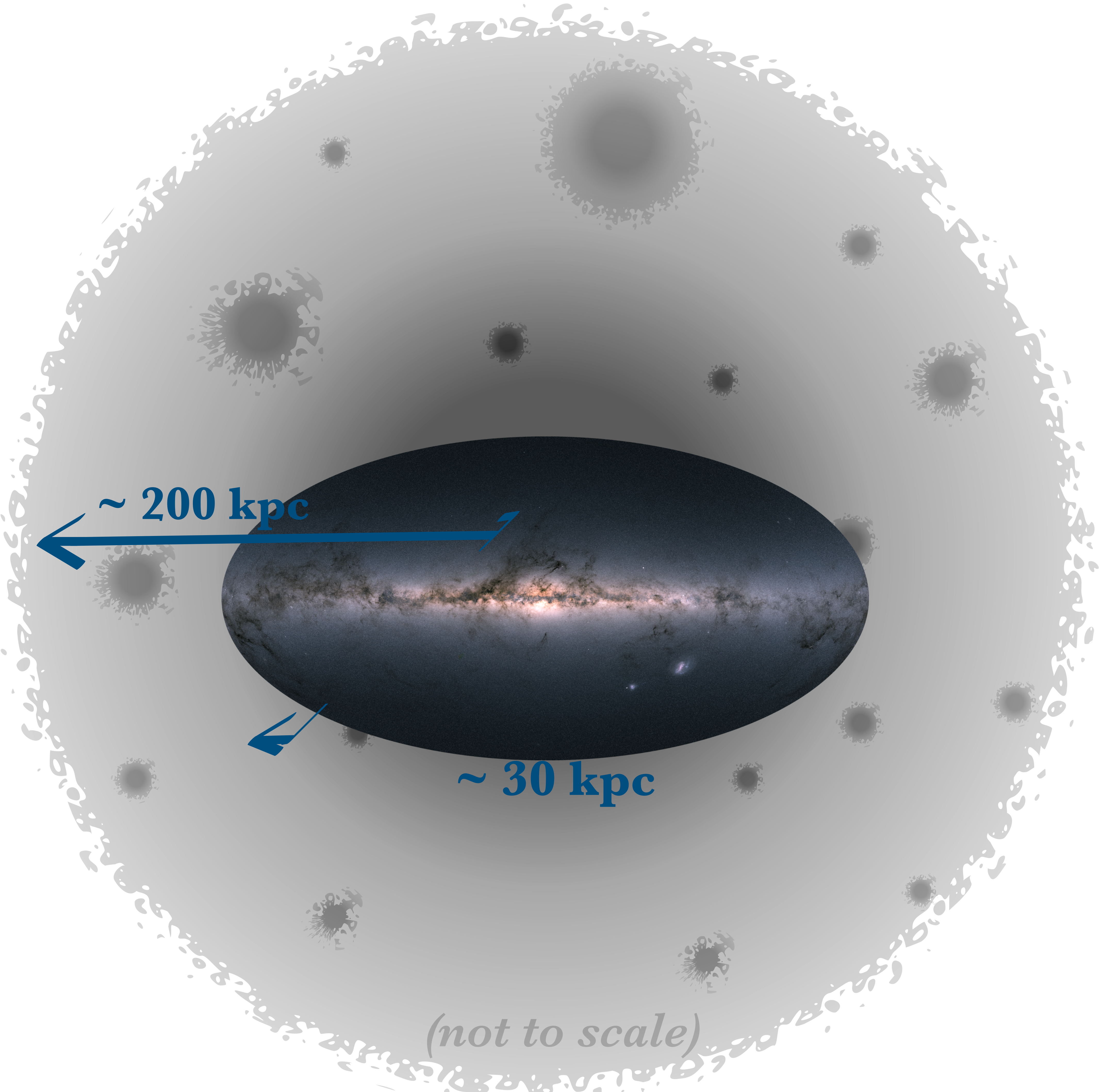
5e8 Msun is perfectly identified, but for smaller masses, amount of training data (4800 images) is the culprit of the drop in performance

The model is generalisable to other physical conditions

Elephant in the room

Wakes are too spatially extended





Gaia ESO

+ stellar halo of MW is a messy place made up of a smooth (virialised) component + non-virialised part

MW-like galaxies

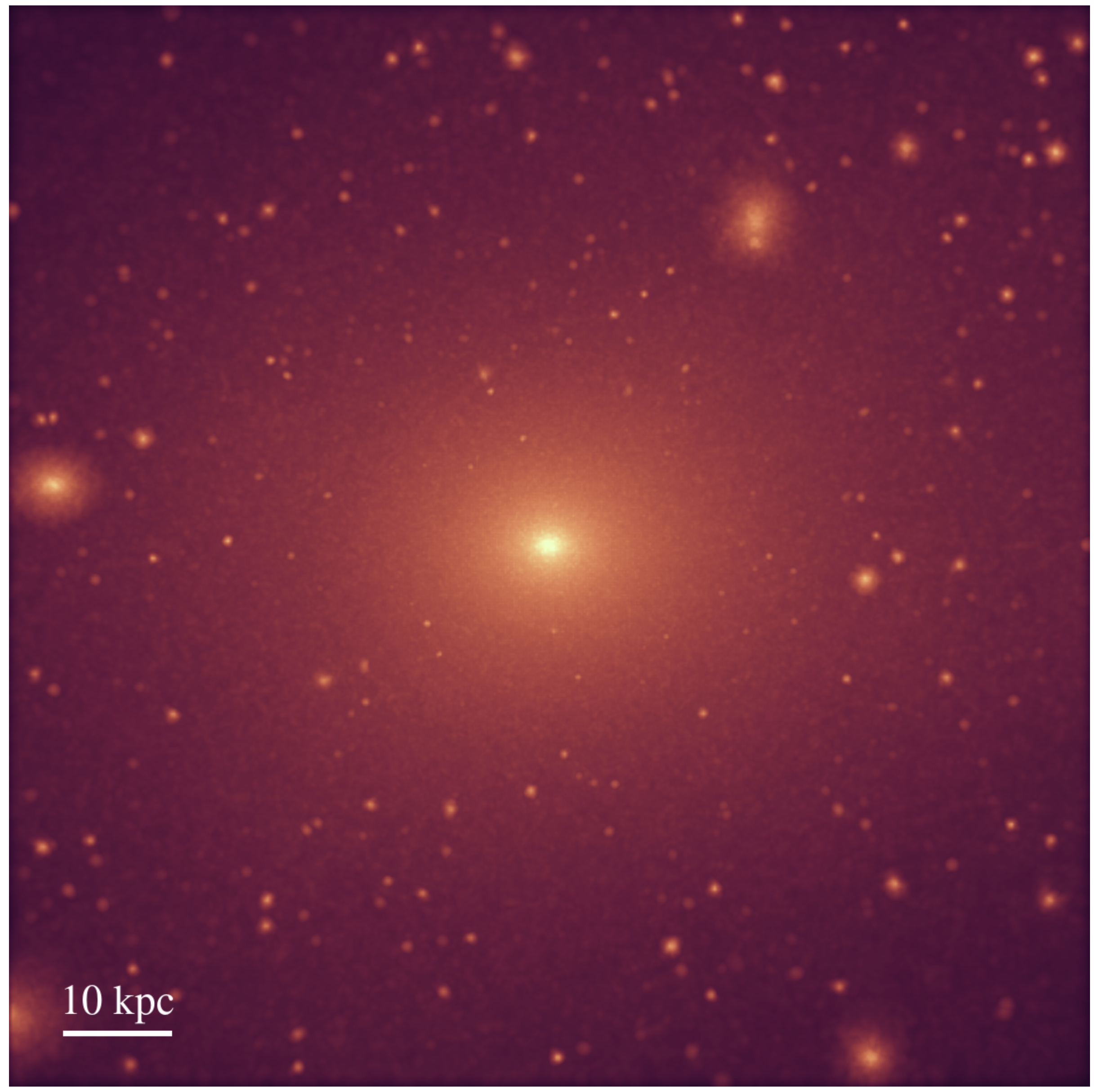
From Latte suite of FIRE-2 simulations

Sensitivity Estimation for Dark Matter Subhalos in Synthetic Gaia DR2 using Deep Learning

A. Bazarov[a], M. Benito[a,b], G. Hütsi[a], R. Kipper[b], J. Pata[a], S. Põder[a,*]

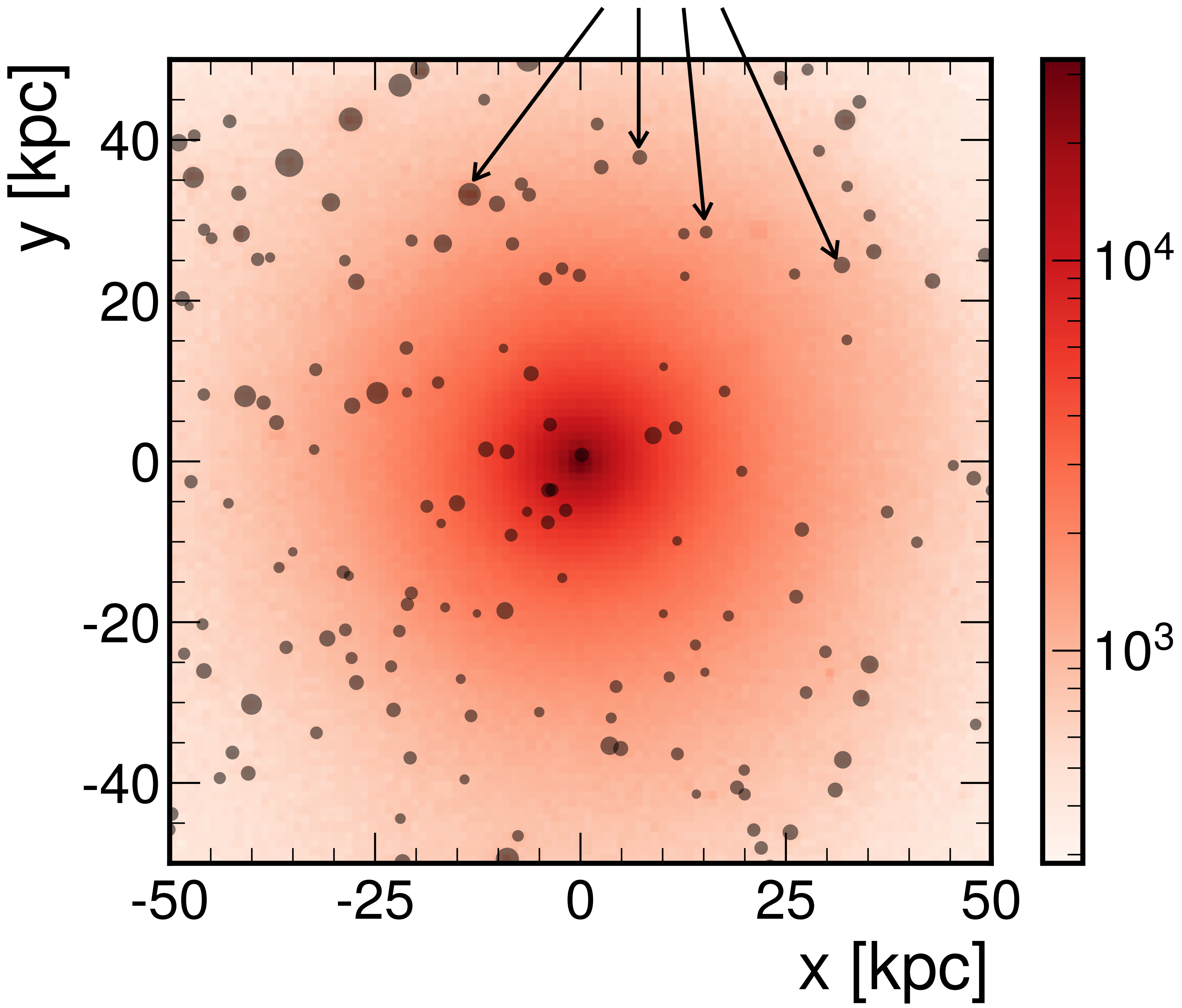
[a] *NICPB, Rävala 10, Tallinn 10143, Estonia*
[b] *Tartu Observatory, University of Tartu, Observatooriumi 1, Tõravere 61602, Estonia*

m12i galaxy



Garrison-Kimmel + [1701.03792]

AHF subhalos



Signal stars are those whose distance to central position of DM subhalo is less than 1 kpc

MW-like galaxies

Anomaly Detection algorithm

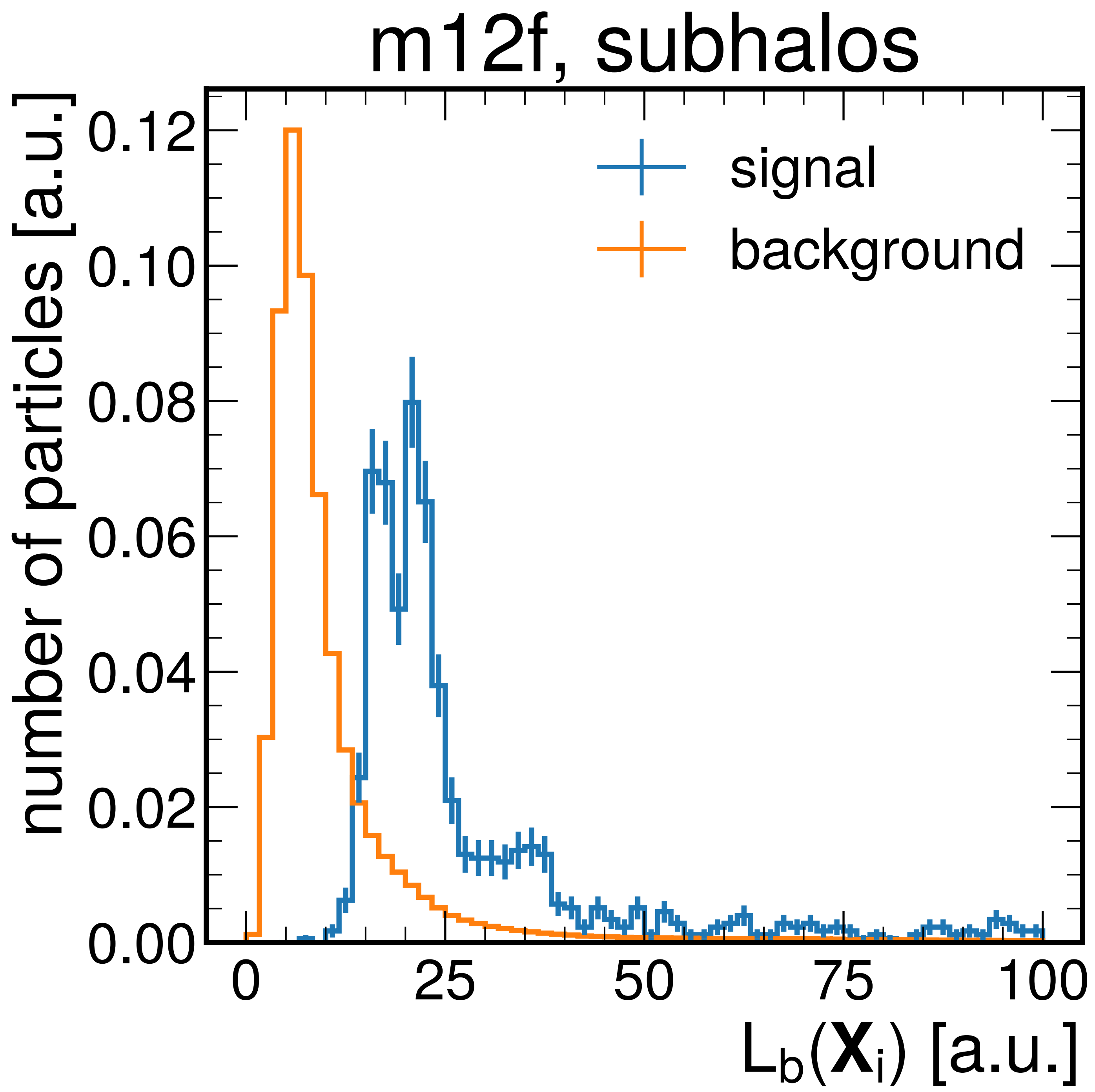
$$\mathbf{D}, \mathbf{E} = \arg\min_{D,E} \sum_{i \in \mathrm{bkg}} \|\mathbf{X}_i - D(E(\mathbf{X}_i))\|$$

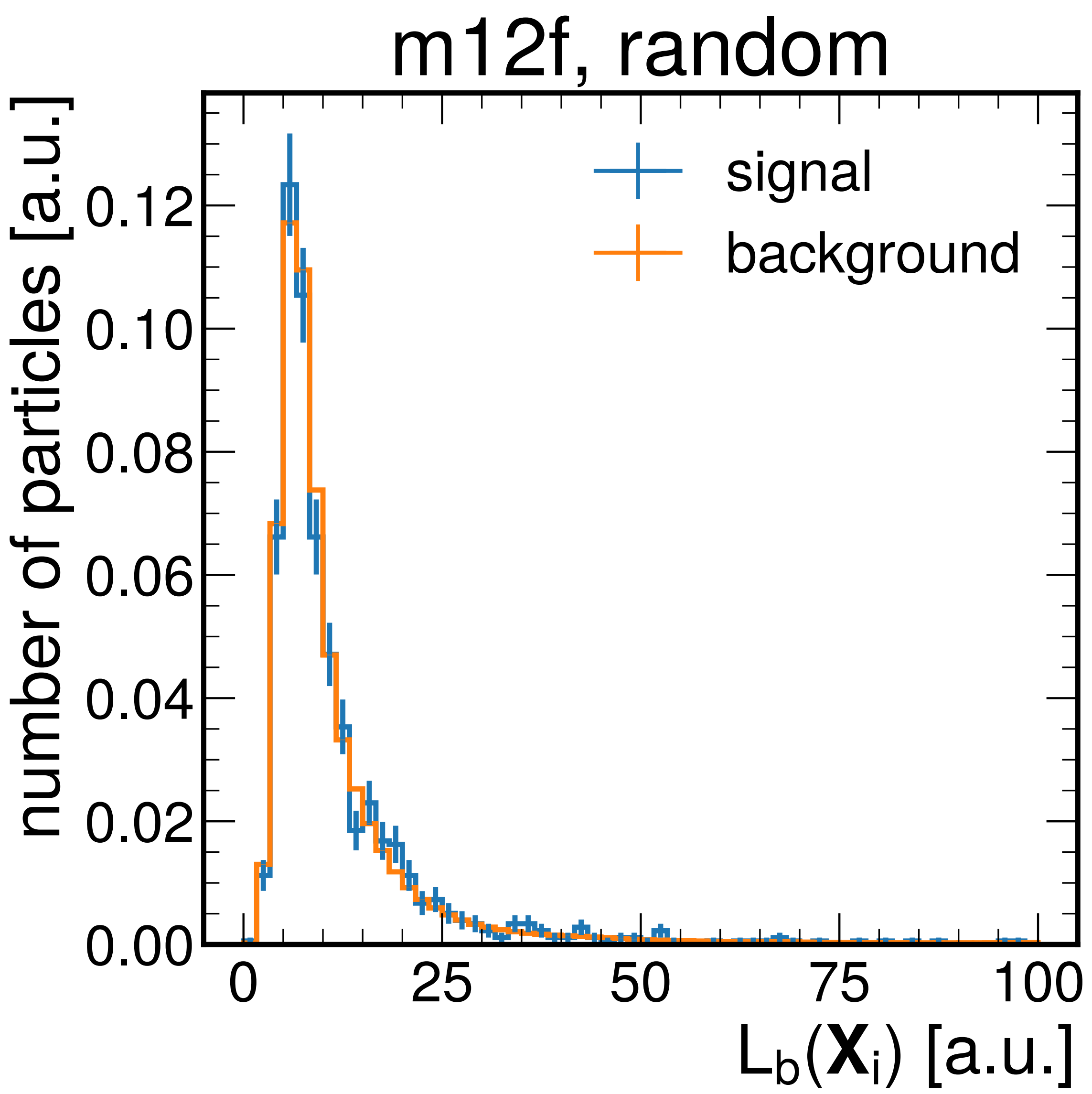
Test statistics to discriminate between signal & bckg stars:

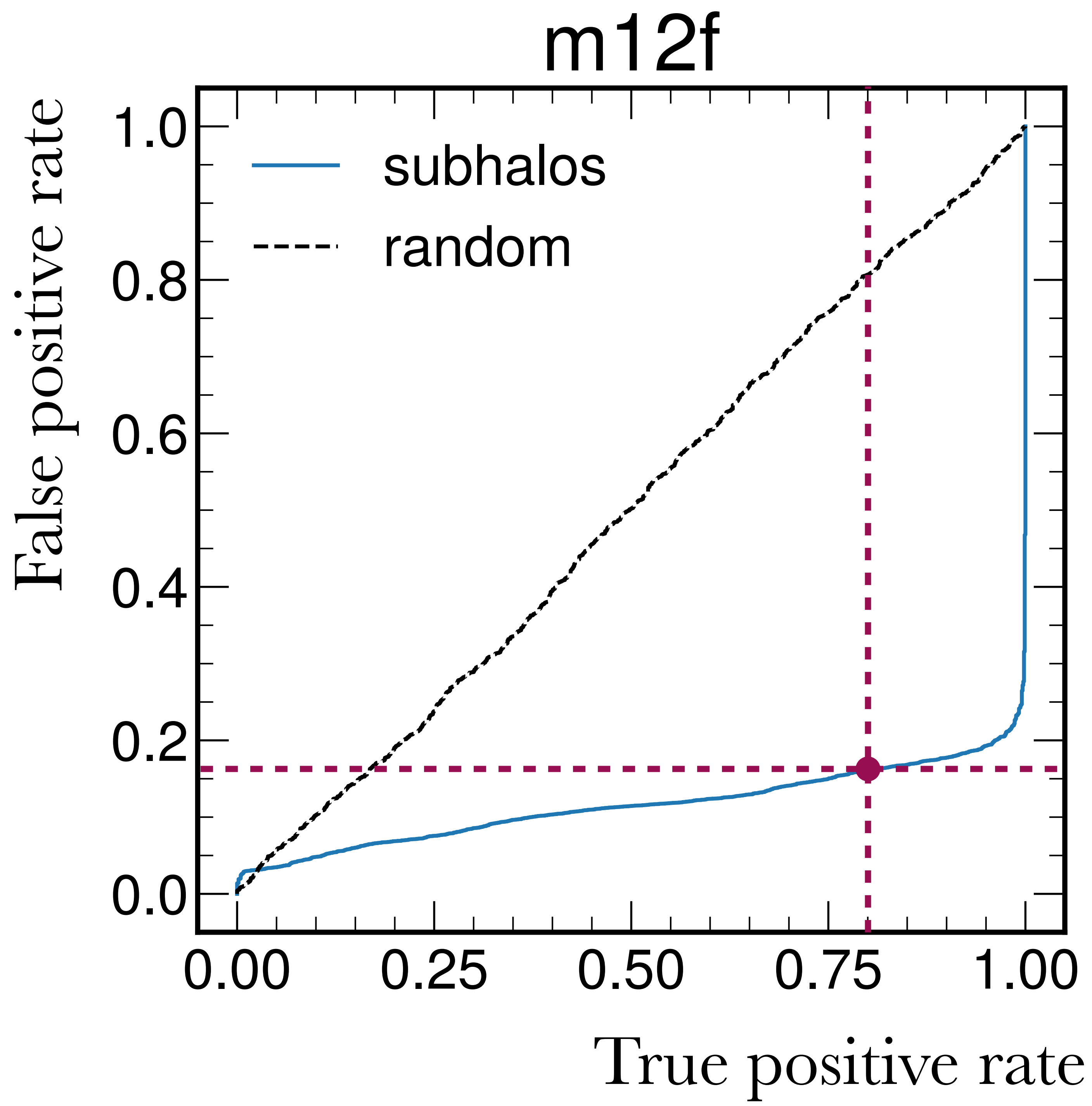
$$L_b(\mathbf{X}) = \|\mathbf{X} - D(E(\mathbf{X}))\|$$

MW-like galaxies

Results







$$L_b(\mathbf{X}) = \|\mathbf{X} - D(E(\mathbf{X}))\|$$

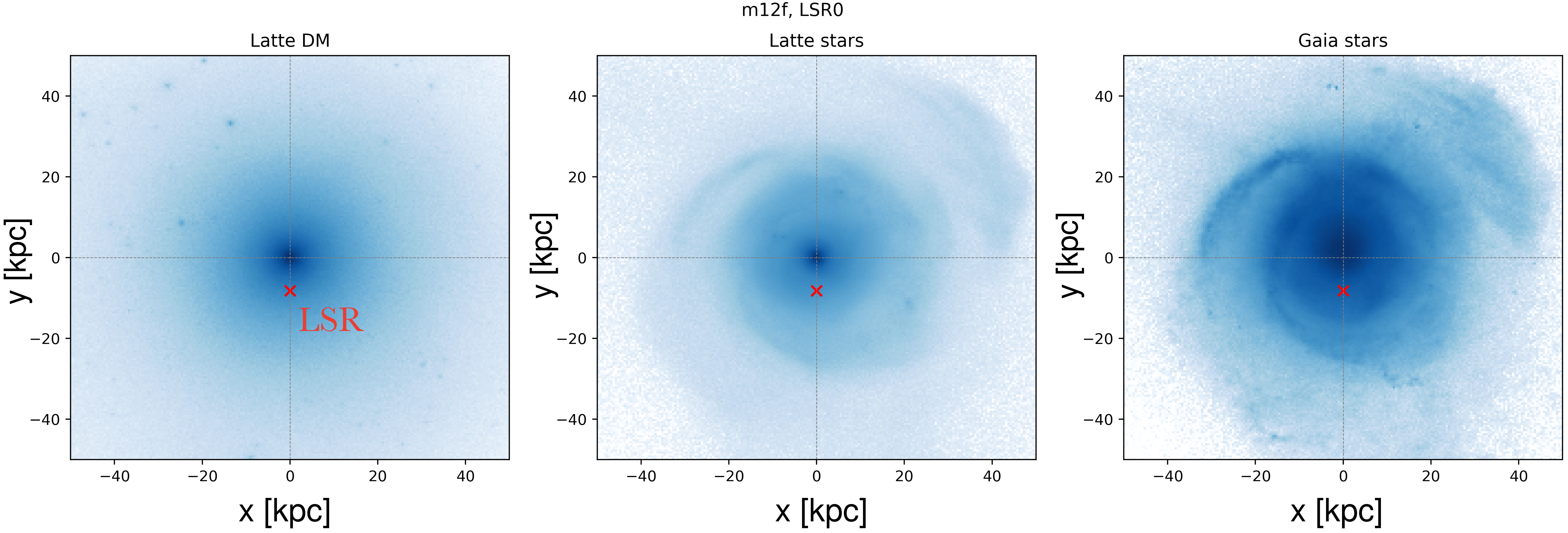
True 3D position & 3D velocity

Reconstructed 3D position & 3D velocity

80% of signal stars are correctly identified while we misclassify ~15% of the background stars as signal

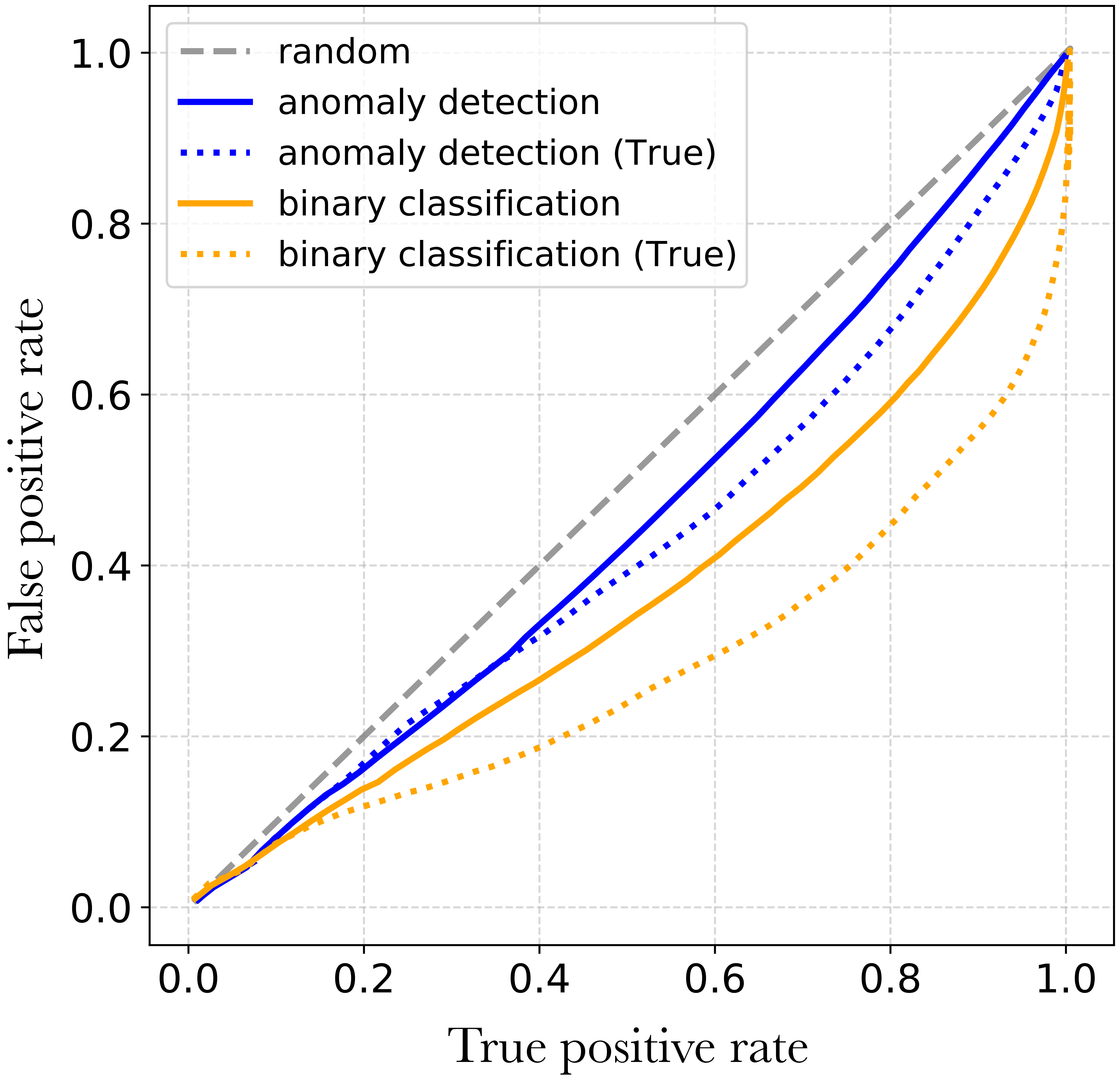
Gaia-like DR2 catalogs

Sanderson + [1806.10564]



Gaia-like DR2 catalogs

Results



Binary classification distinguishes between the halo-associated and background stars at a non-negligible level: FPR of ~35% at a TPR of ~50%

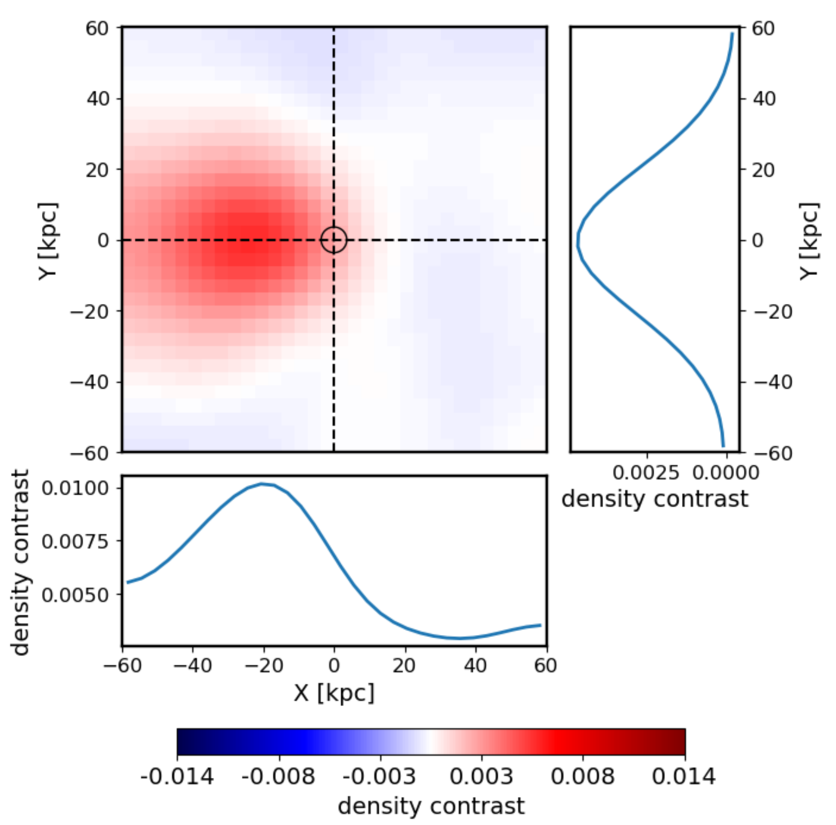
Anomaly detection does not differ significantly from purely random selection

Conclusions: väljakutse (a challenge)

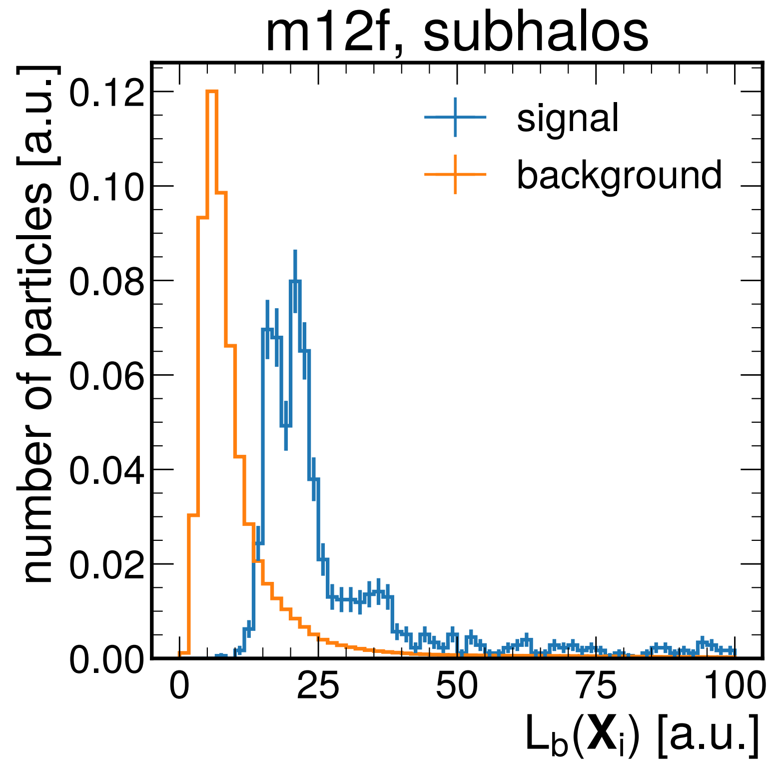
Stellar wakes program (not only relevant for DM science) is starting

—> we have characterised the signal

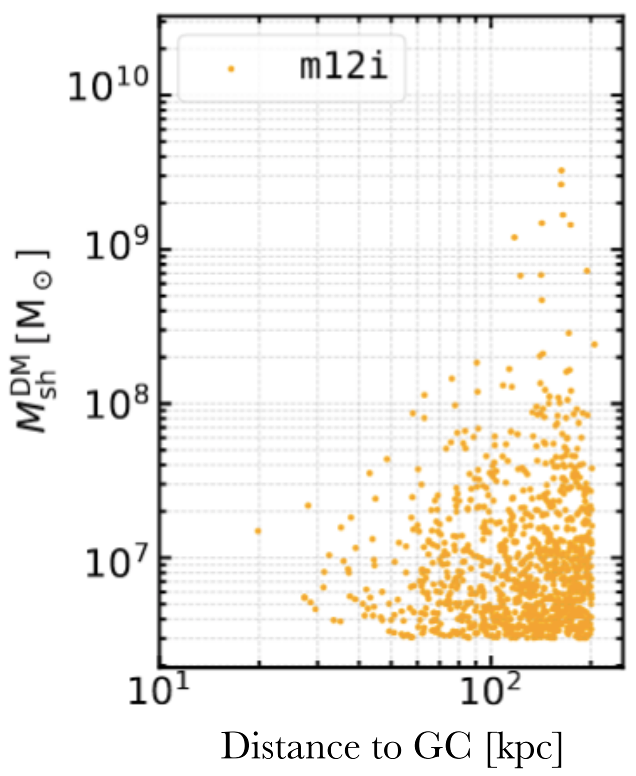
—> we have investigated the viability & performance of detecting individual wakes



Plenty of work to be done



Theoretical challenges



Sensitivity Estimation for Dark Matter Subhalos in Synthetic Gaia DR2 using Deep Learning

A. Bazarov[a], M. Benito[a,b], G. Hütsi[a], R. Kipper[b], J. Pata[a], S. Põder[a,*]

[a] NICPB, Rävala 10, Tallinn 10143, Estonia
[b] Tartu Observatory, University of Tartu, Observatooriumi 1, Tõravere 61602, Estonia

ML catalogue

On the detection of stellar wakes in the Milky Way: a deep learning approach

Sven Põder[1,2], Joosep Pata[1], María Benito[3], Isaac Alonso Asensio[4,5], and Claudio Dalla Vecchia[4,5]