Using machine learning to unlock Gaia's full potential to determine the dark matter halo

Bryan Ostdiek

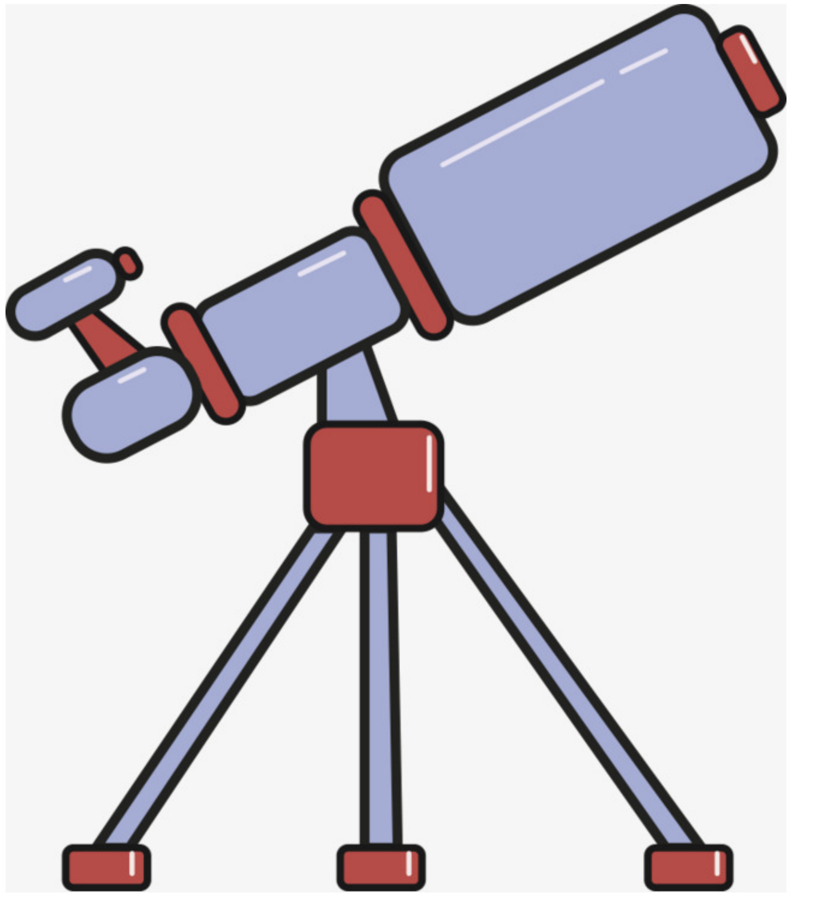
LBNL, September 26, 2018

with Timothy Cohen, Marat Freytsis, Phillip Hopkins, Mariangela Lisanti, Lina Necib, and Andrew Wetzel

First Gaia skymap in color [https://www.cosmos.esa.int/web/gaia/gaiadr2_gaiaskyincolour]

Introduction

Goal: Use data from the Gaia satellite to make measurements about the halo of the Milky Way



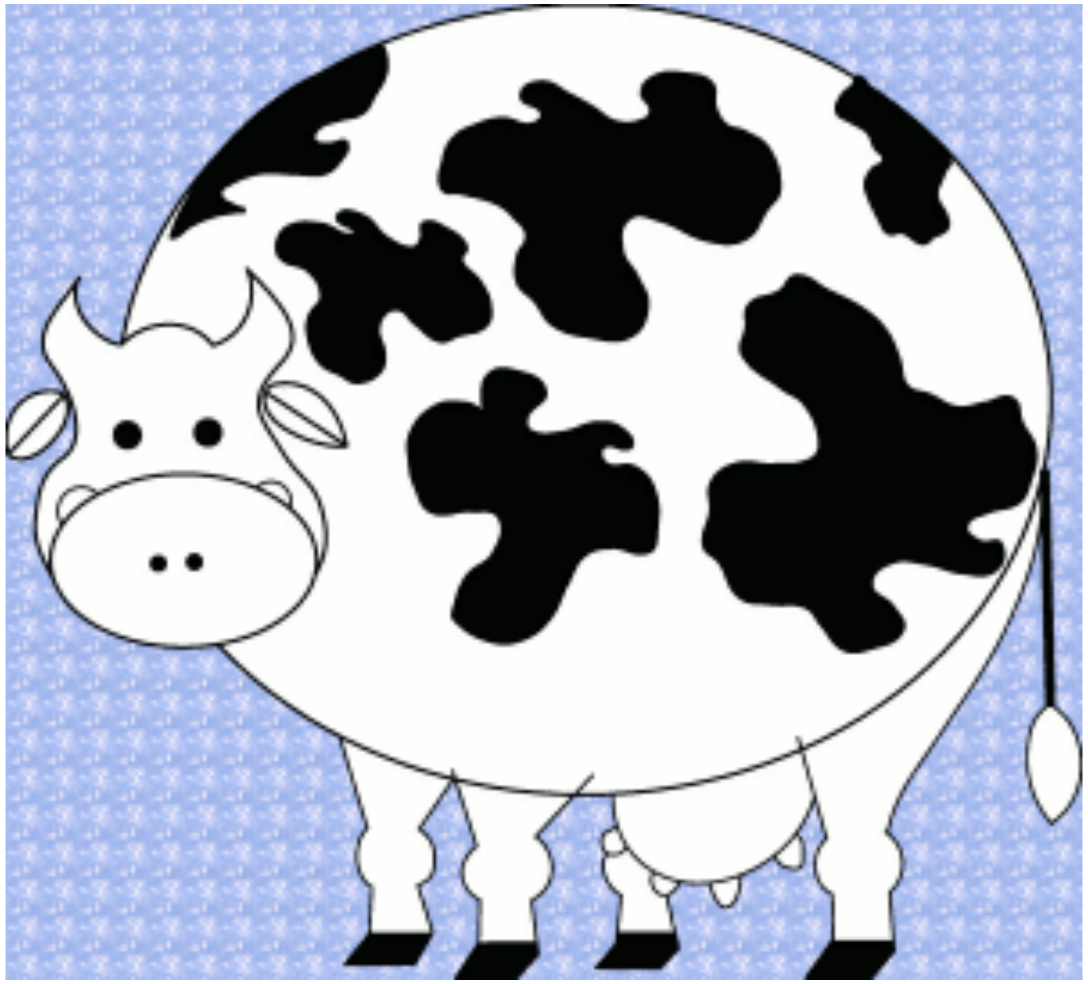
Why (Astronomers): How galaxies form

Why (Particle physicists): Dark matter makes up the halo

How: Old stars act as tracers for dark matter

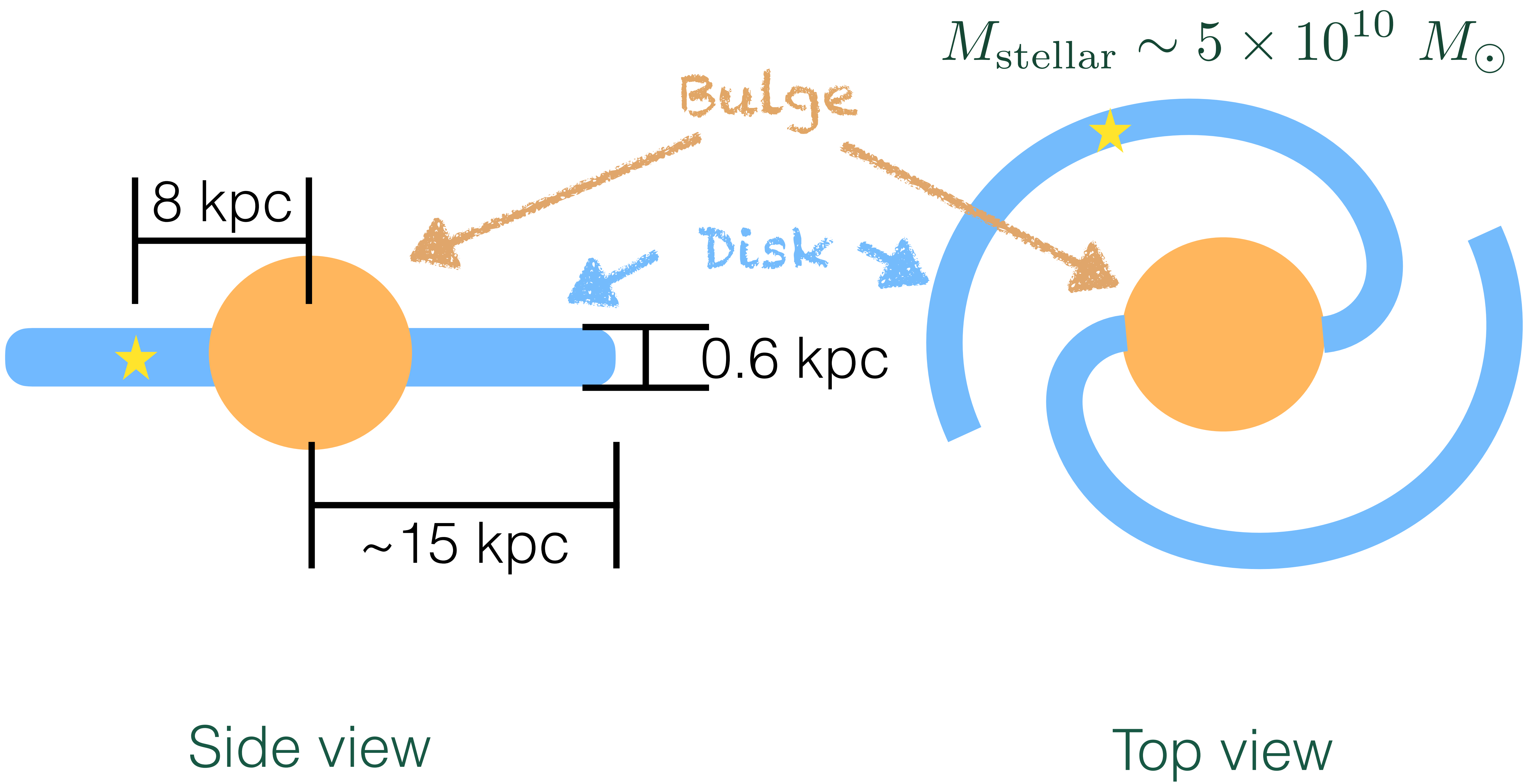
Challenges: Identifying old stars with limited information

Spherical cow model of a galaxy

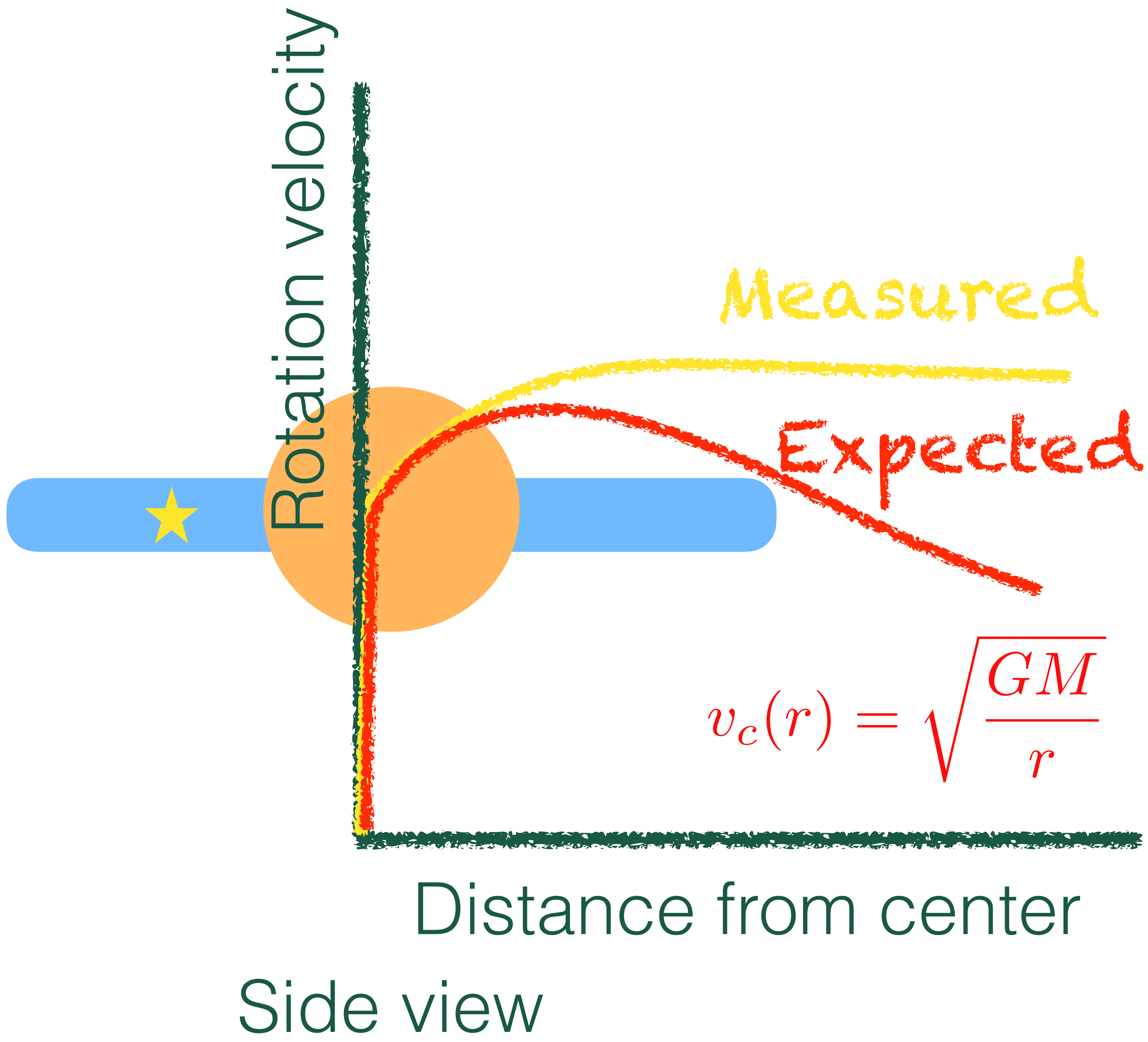


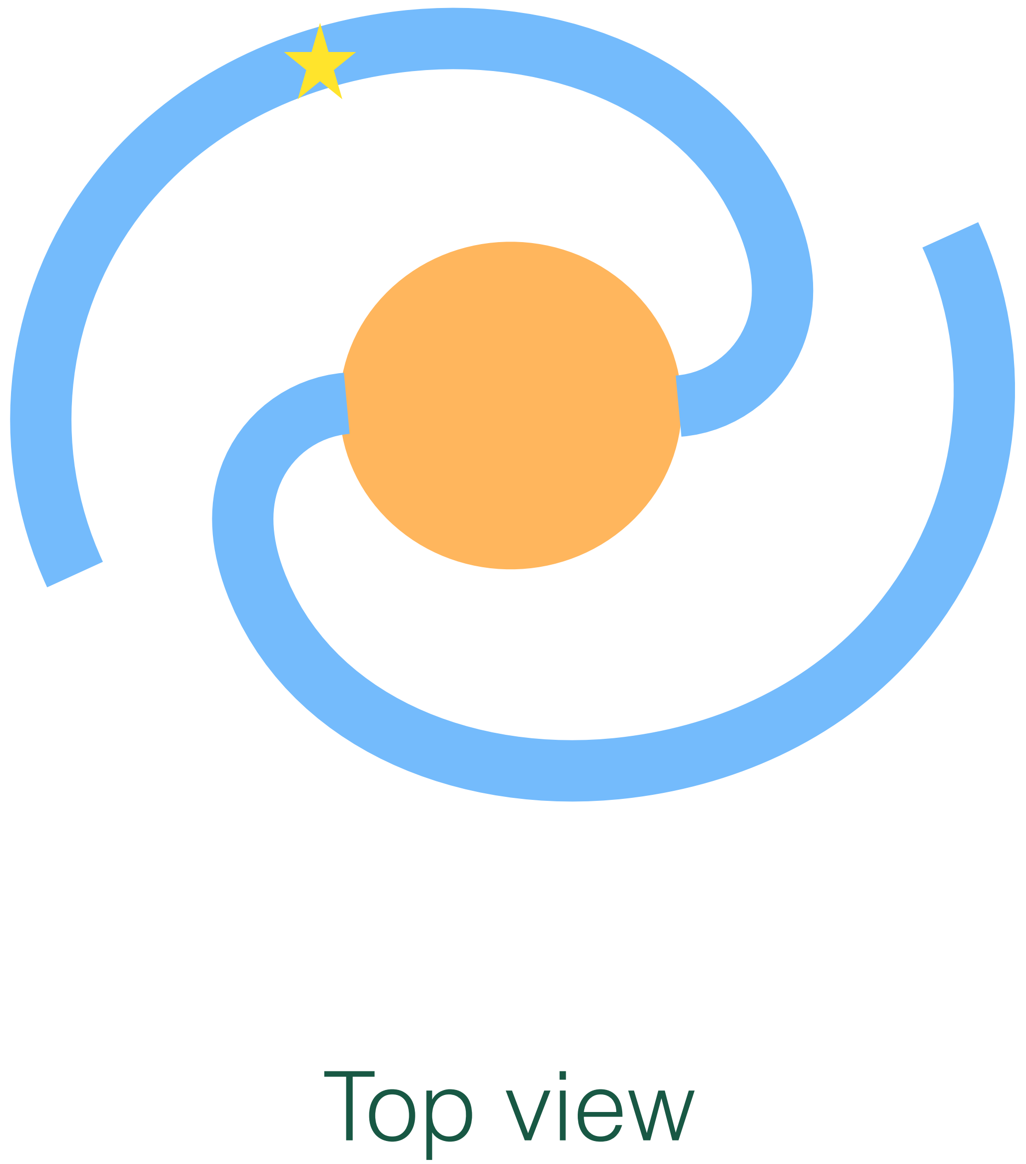
http://www.physics.csbsju.edu/stats/WAPP2_cow.html

Toy model of spiral galaxy

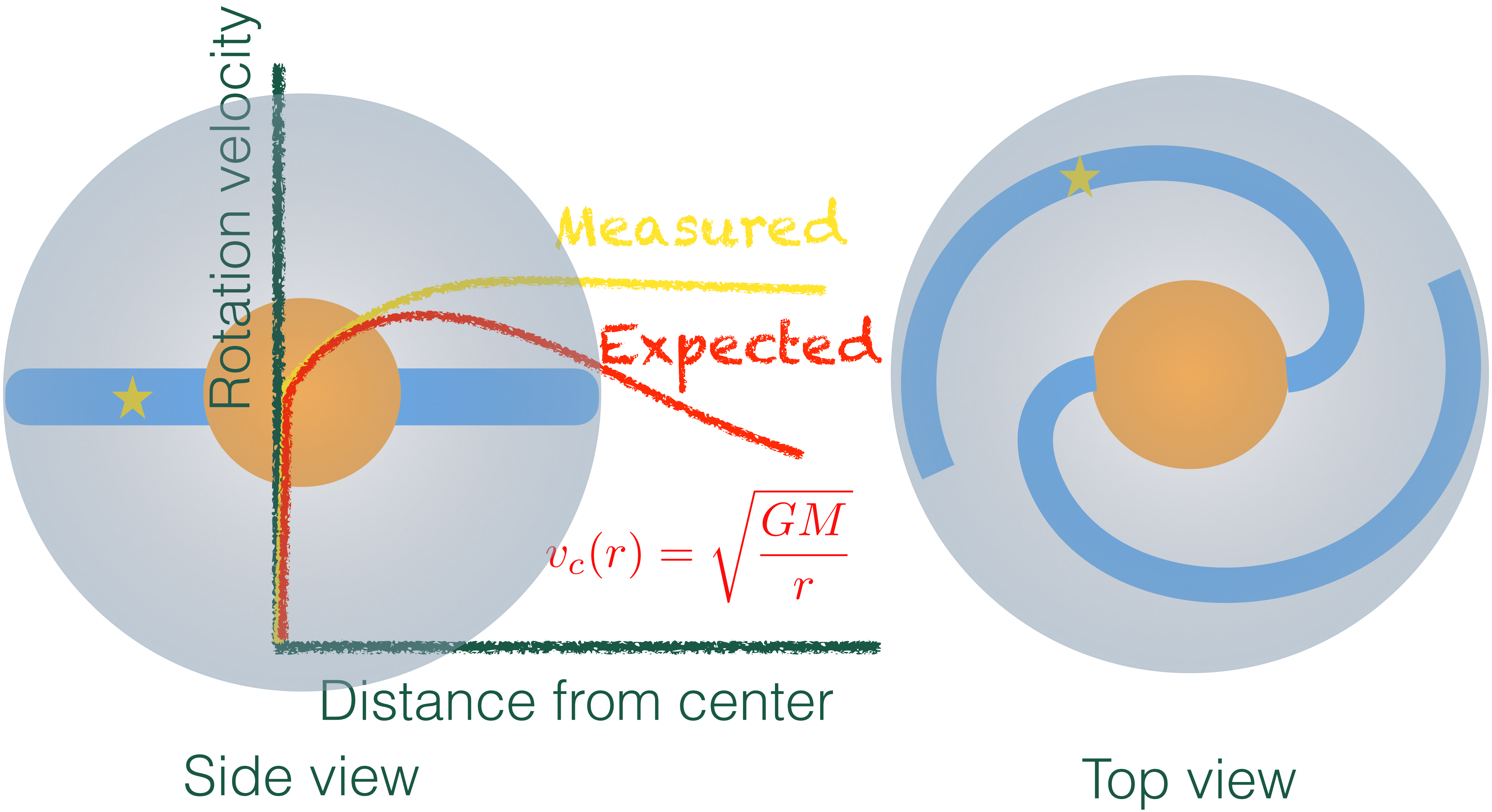


Toy model of spiral galaxy

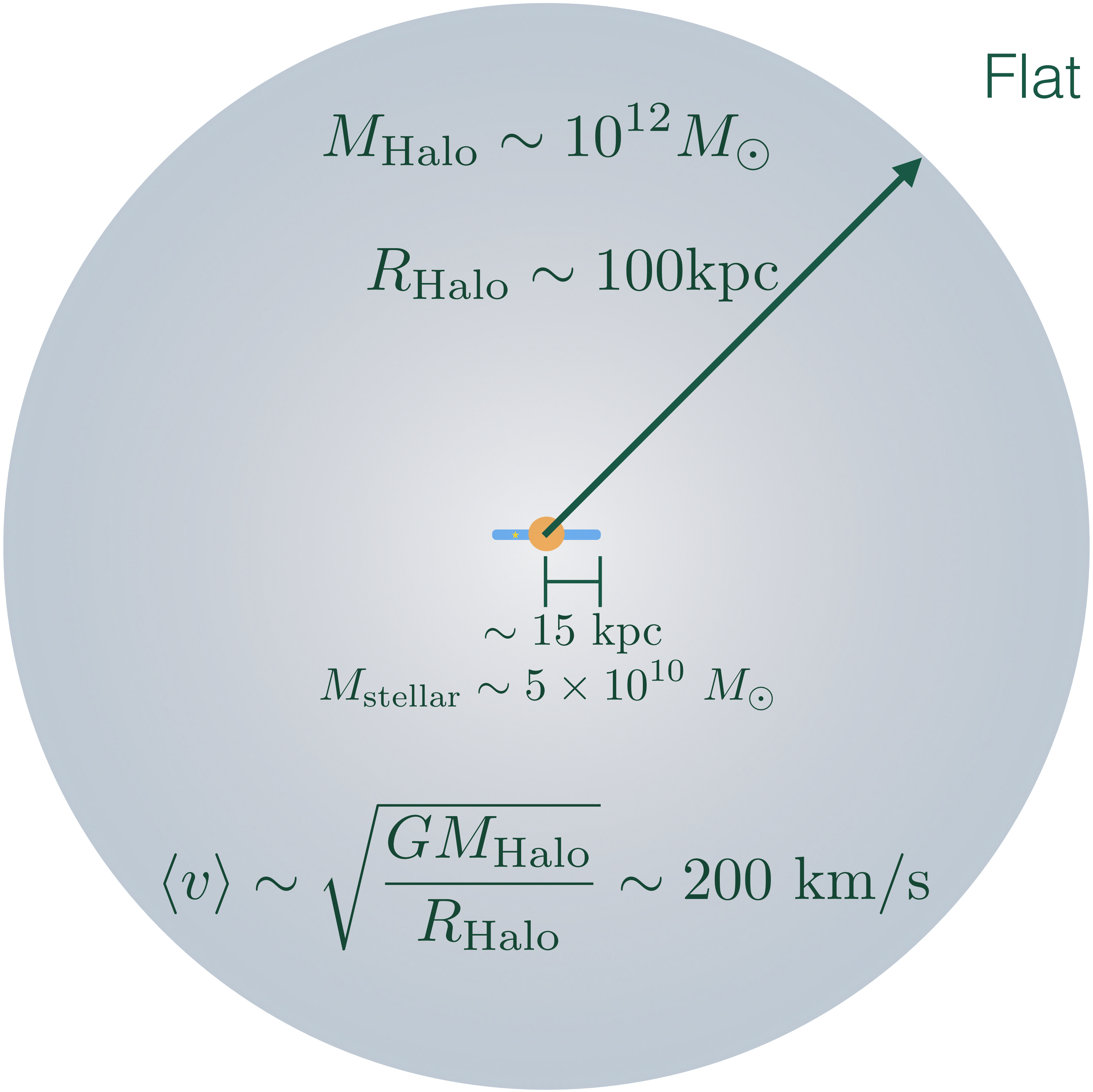




Toy model of spiral galaxy



Toy model of spiral galaxy



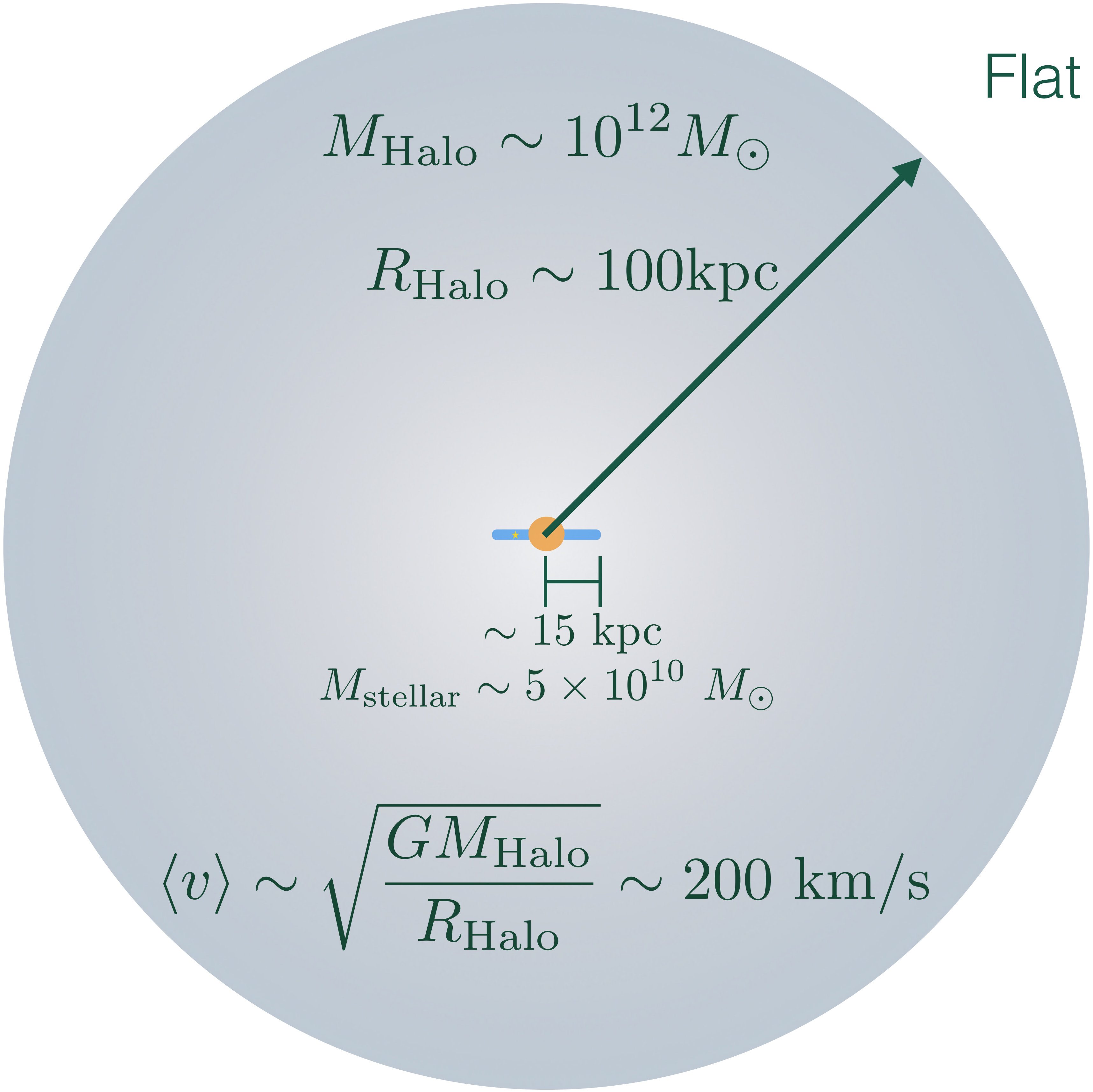
Flat rotation curve implies

$$M(r) \propto r$$

Assuming spherical symmetry

$$\rho(r) \propto \frac{1}{r^2}$$

Toy model of spiral galaxy



Flat rotation curve implies

$$M(r) \propto r$$

Assuming spherical symmetry

$$\rho(r) \propto \frac{1}{r^2}$$

- Collisionless
- Nonrelativistic
- Self-gravitating
- Isotropic
- Isothermal gas

Hierarchical Merger Model

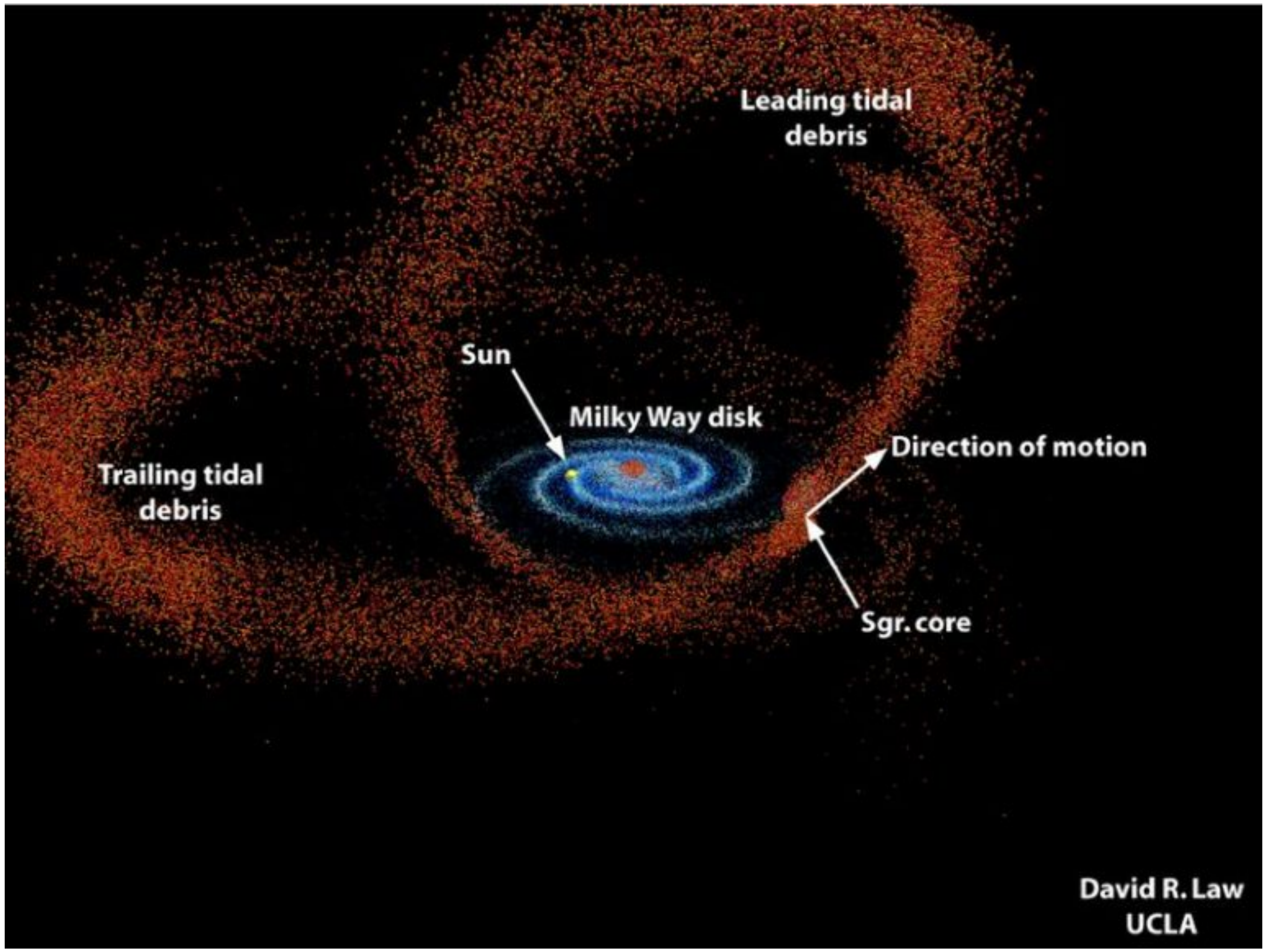
1) Density fluctuations after big bang lead to proto-galactic fragments of order $10^6 - 10^8 M_\odot$
2) Fragments evolve in isolation creating stars / globular clusters
3) Collisions and tidal disruptions lead to distribution of halo (stars and DM)
4) Gas in the mergers interact and collapse to disk
5) Young and metal rich stars produced in the disk

Last major merger ~10 Gyr ago	Minor mergers still happening

Hierarchical Merger Model

Minor mergers still happening

Stars and DM in the proto-galactic fragments only interact via gravity



http://www.stsci.edu/~dlaw/Sgr/TimeEvol.html

To find dark matter distribution, find stars from early mergers

Dark Matter Tracers

To find dark matter distribution, find stars from early mergers

Early merger ➞ Old star ➞ Low metallicity

$$[\mathrm{Fe/H}] = \log_{10}\left(\frac{\mathrm{N_{Fe}}}{\mathrm{N_H}}\right) - \log_{10}\left(\frac{\mathrm{N_{Fe}}}{\mathrm{N_H}}\right)_{\odot}$$

Dark Matter Tracers

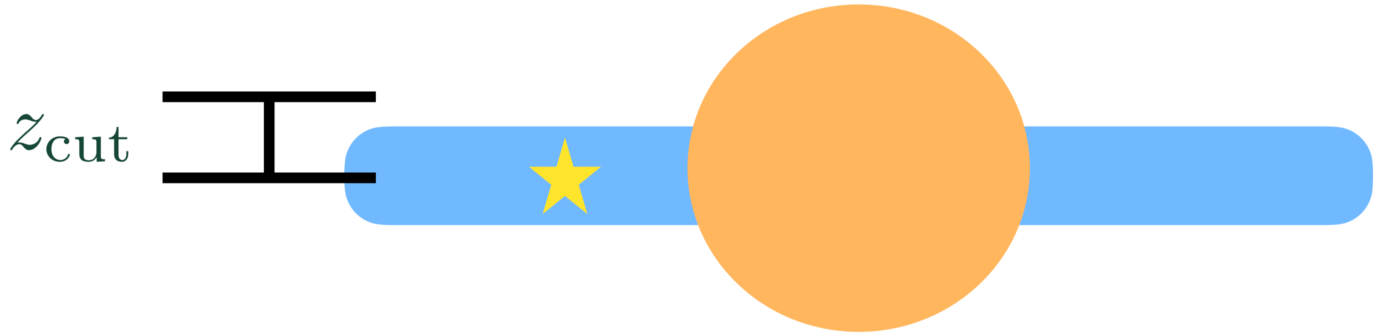
To find dark matter distribution, find stars from early mergers

Early merger → Old star → Low metallicity

$$[\mathrm{Fe/H}] = \log_{10}\left(\frac{\mathrm{N_{Fe}}}{\mathrm{N_H}}\right) - \log_{10}\left(\frac{\mathrm{N_{Fe}}}{\mathrm{N_H}}\right)_{\odot}$$

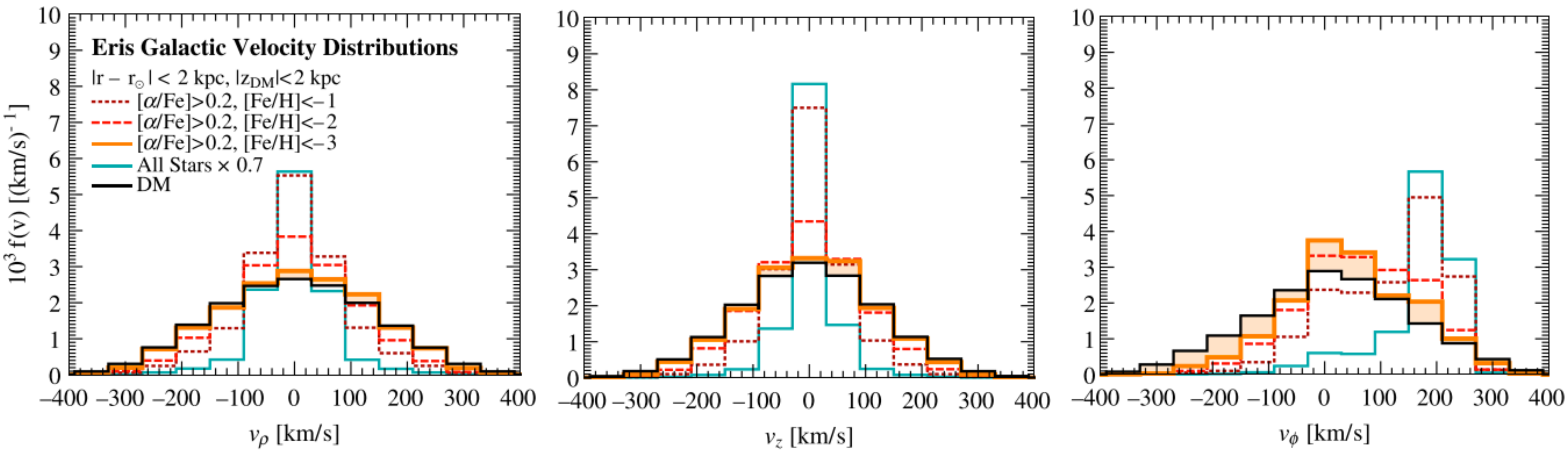
Also helps to not look directly in the disk

$$|z| > z_{\mathrm{cut}}$$



Dark Matter Tracers

To find dark matter distribution, find stars from early mergers

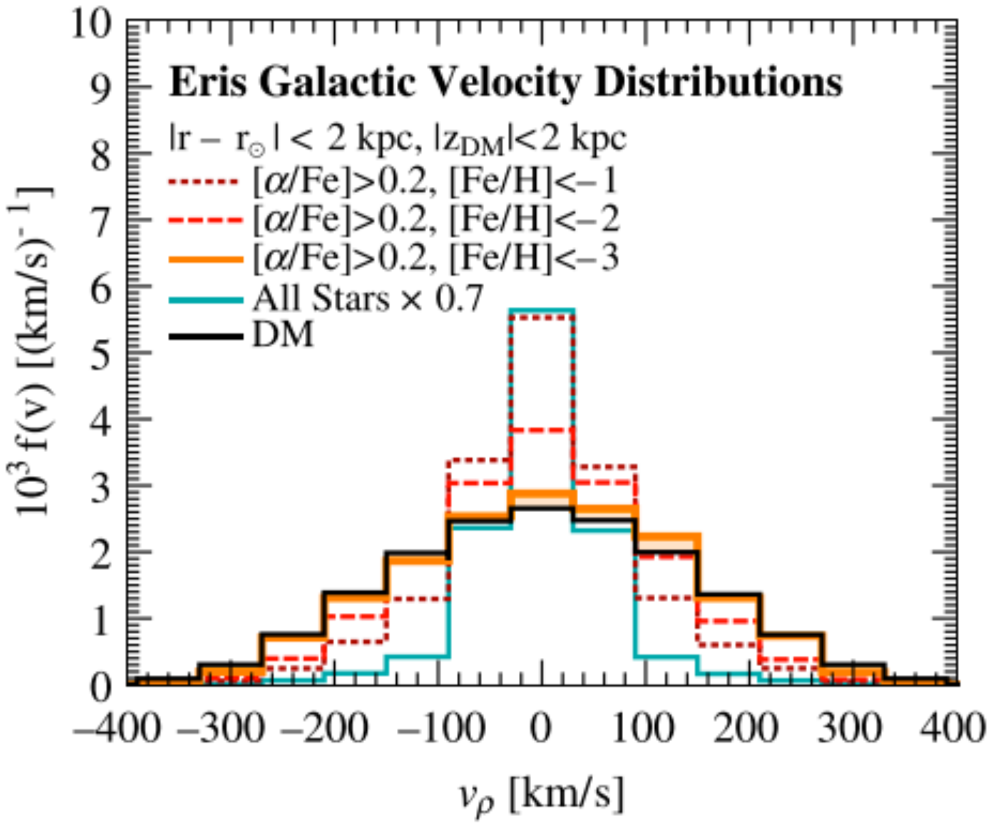


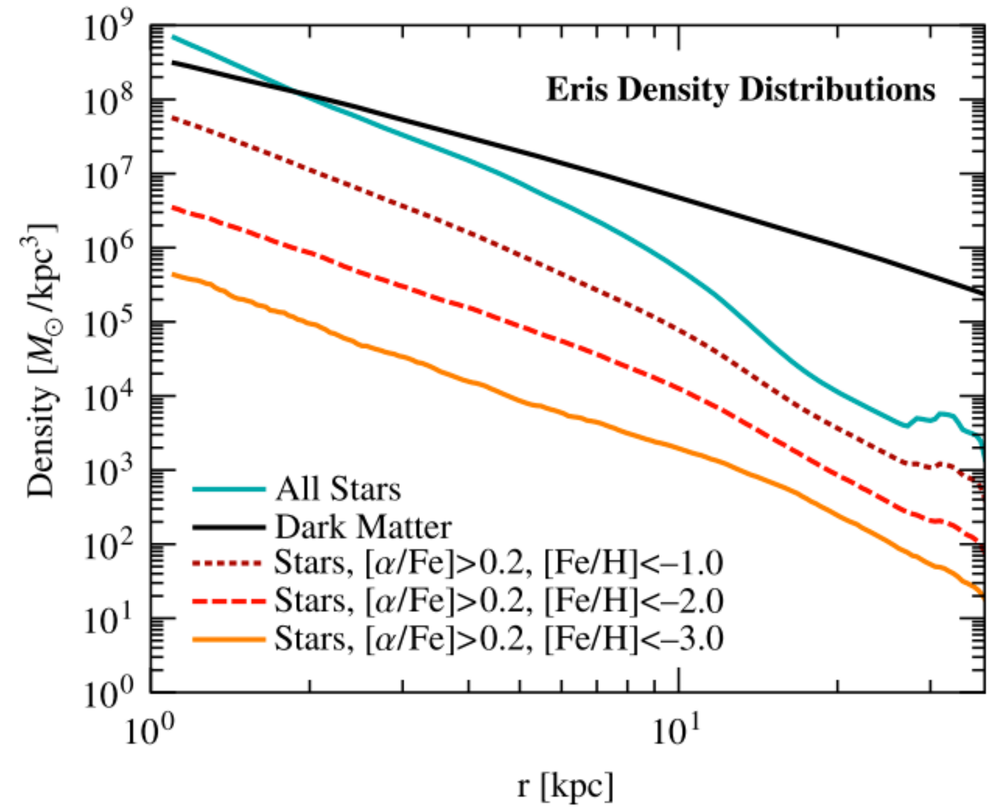
[arXiv:1704.04499]

Old (low [Fe/H]) stars and dark matter share the same **velocity** distributions!

Dark Matter Tracers

To find dark matter distribution, find stars from early mergers



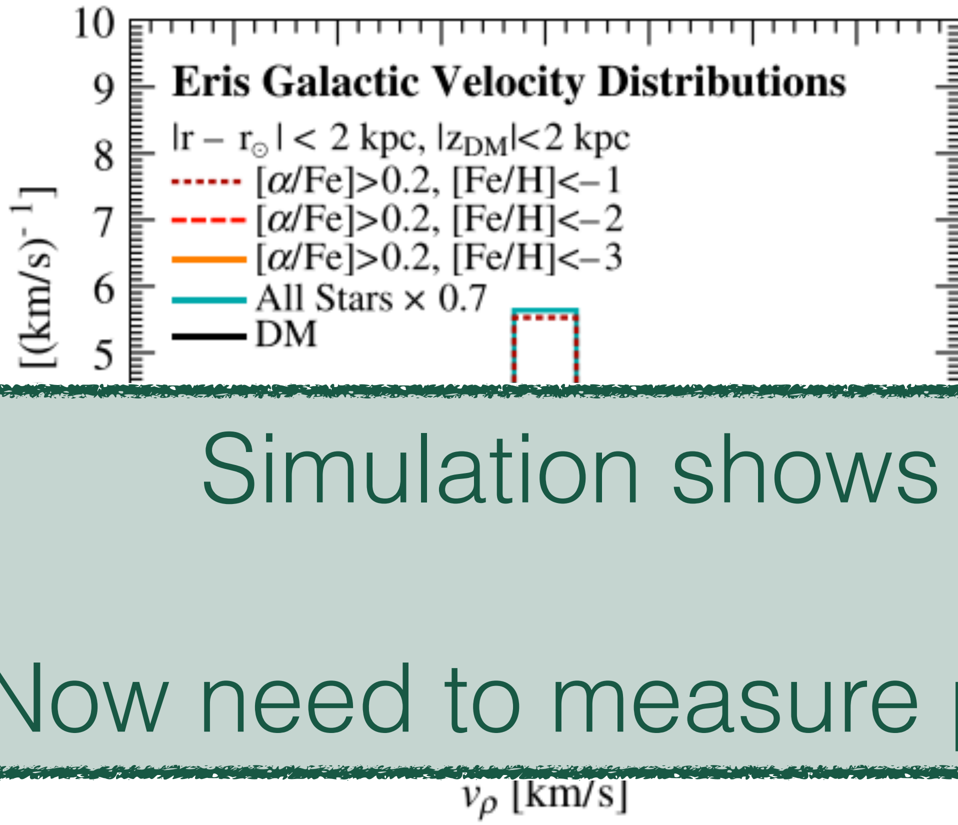


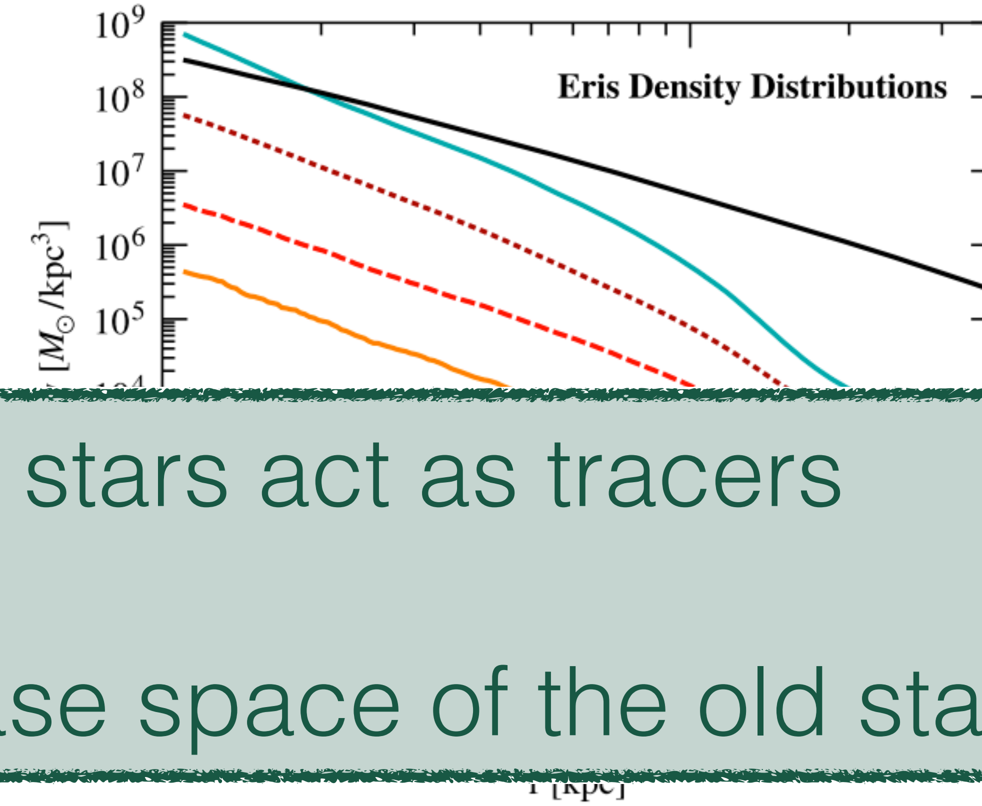
[arXiv:1704.04499]

Old (low [Fe/H]) stars and dark matter share same **density** profile!

Dark Matter Tracers

To find dark matter distribution, find stars from early mergers





Simulation shows old stars act as tracers

Now need to measure phase space of the old stars

[arXiv:1704.04499]

Old (low [Fe/H]) stars and dark matter share same **density** profile!

Catalogs of real data

Phase Space (5-d)

- Gaia DR1 (2-d) location for 1 billion stars
  - ★ Cross matched with Tycho-2 catalog of Hipparcos → 2 million stars
- Gaia DR2 (5-d) information for 1 billion stars

Spectroscopy

- Radial VElocity Experiment
- Sloan Digital Sky Survey

Catalogs of real data

Phase Space (5-d)	Spectroscopy
• Gaia DR1 (2-d) location for 1 billion stars ★ Cross matched with Tycho-2 catalog of Hipparcos → 2 million stars • Gaia DR2 (5-d) information for 1 billion stars	• Radial VElocity Experiment • Sloan Digital Sky Survey

RAVE-TGAS (255,922 stars)

Catalogs of real data

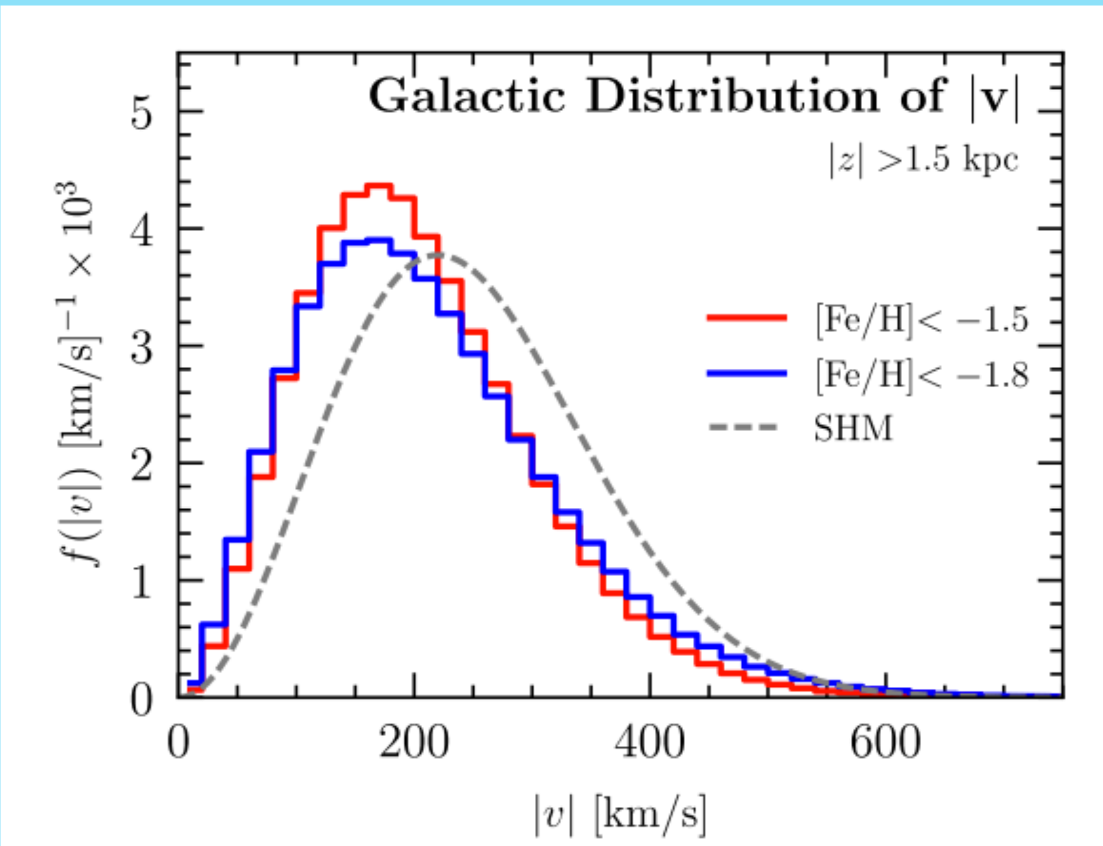
Phase Space (5-d)	Spectroscopy
• Gaia DR1 (2-d) location for 1 billion stars ★ Cross matched with Tycho-2 catalog of Hipparcos → 2 million stars • Gaia DR2 (5-d) information for 1 billion stars	• Radial VElocity Experiment • Sloan Digital Sky Survey

RAVE-TGAS (255,922 stars)

Gaia-SDSS (193,162 stars)

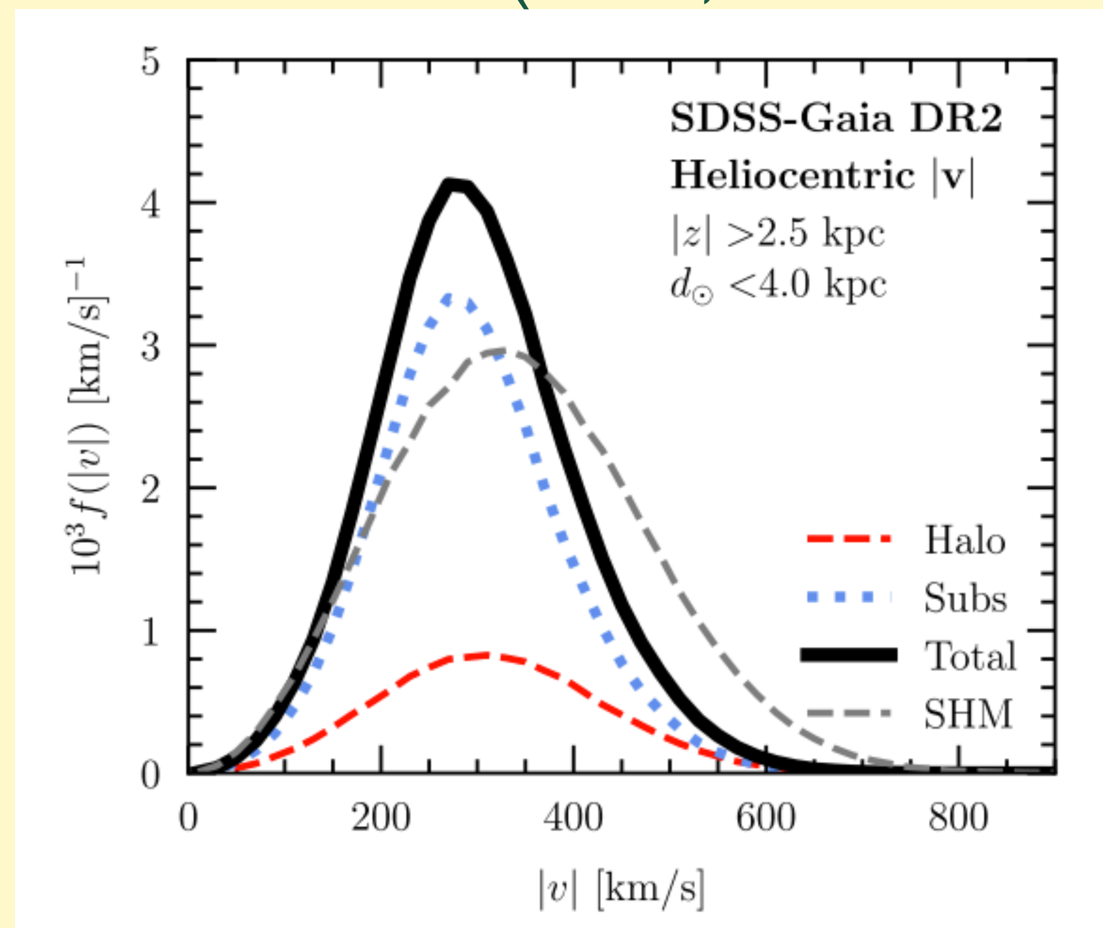
Catalogs of real data

RAVE-TGAS (255,922 stars)



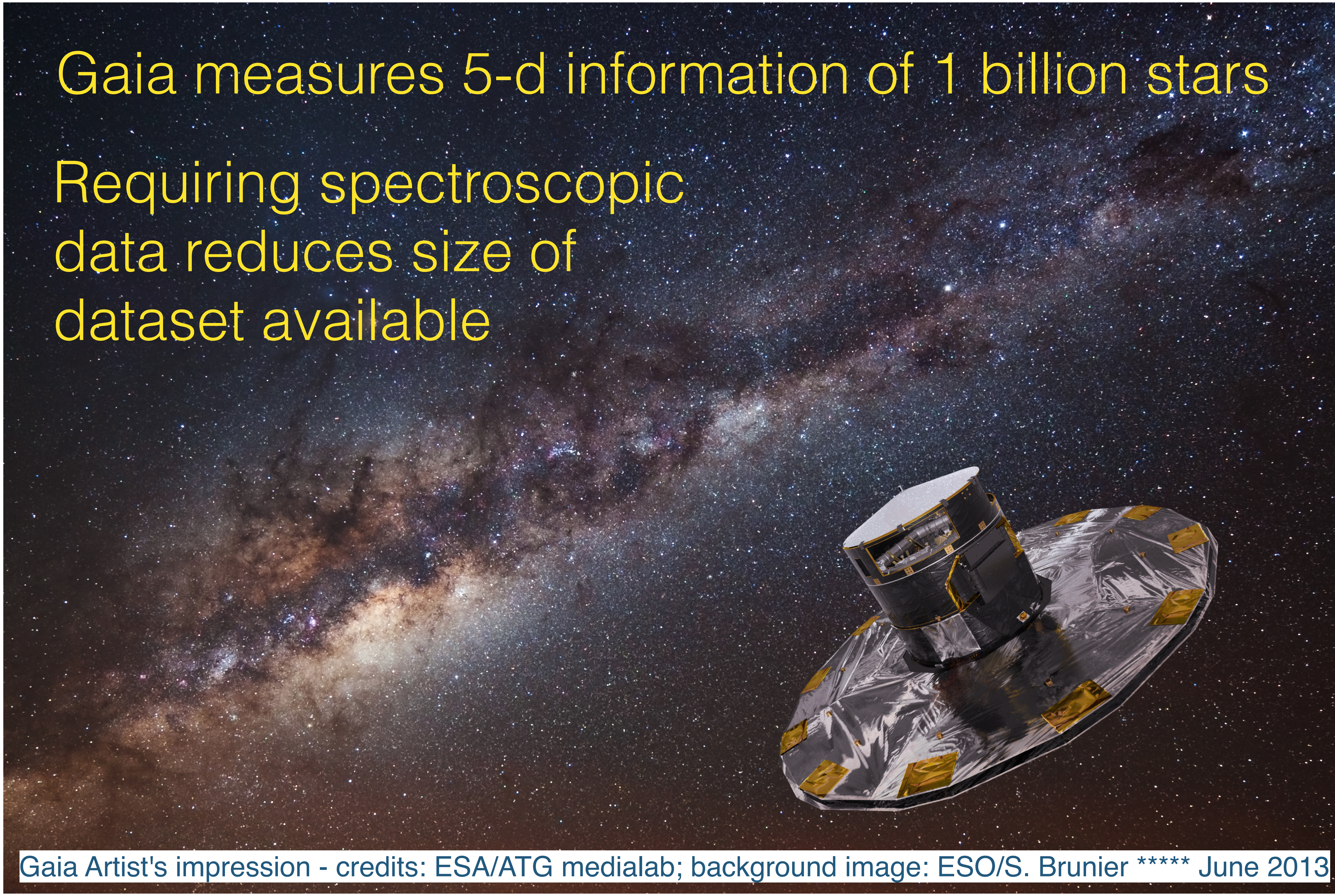
J. Herzog-Arbeitman, M. Lisanti and L. Necib, JCAP **1804**, no. 04, 052 (2018) doi:10.1088/1475-7516/2018/04/052 [arXiv:1708.03635 [astro-ph.GA]].

Gaia-SDSS (193,162 stars)



L. Necib, M. Lisanti and V. Belokurov, arXiv:1807.02519 [astro-ph.GA].

Empirical determination of halo velocity distribution smaller than standard model halo → direct detection interpretation



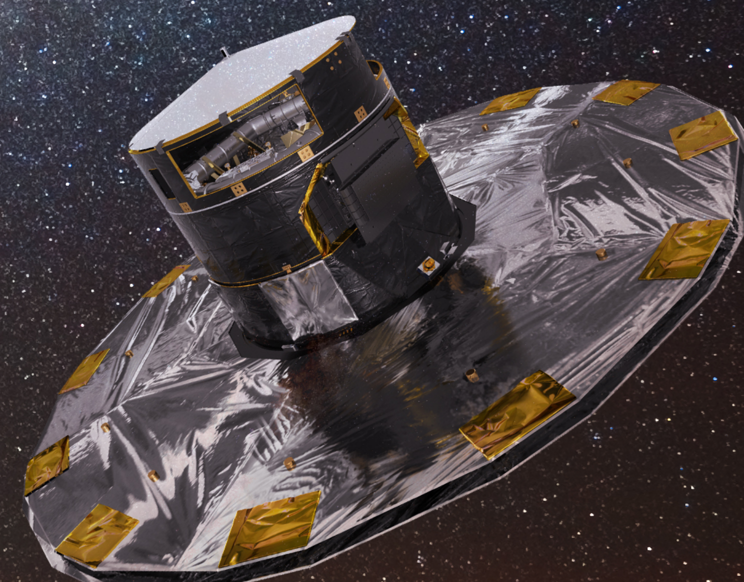
Gaia Artist's impression - credits: ESA/ATG medialab; background image: ESO/S. Brunier ***** June 2013

Is it possible to classify halo stars using only 5-d information?

Gaia measures 5-d information of 1 billion stars

Requiring spectroscopic data reduces size of dataset available

Use deep neural network as generic distribution fitter



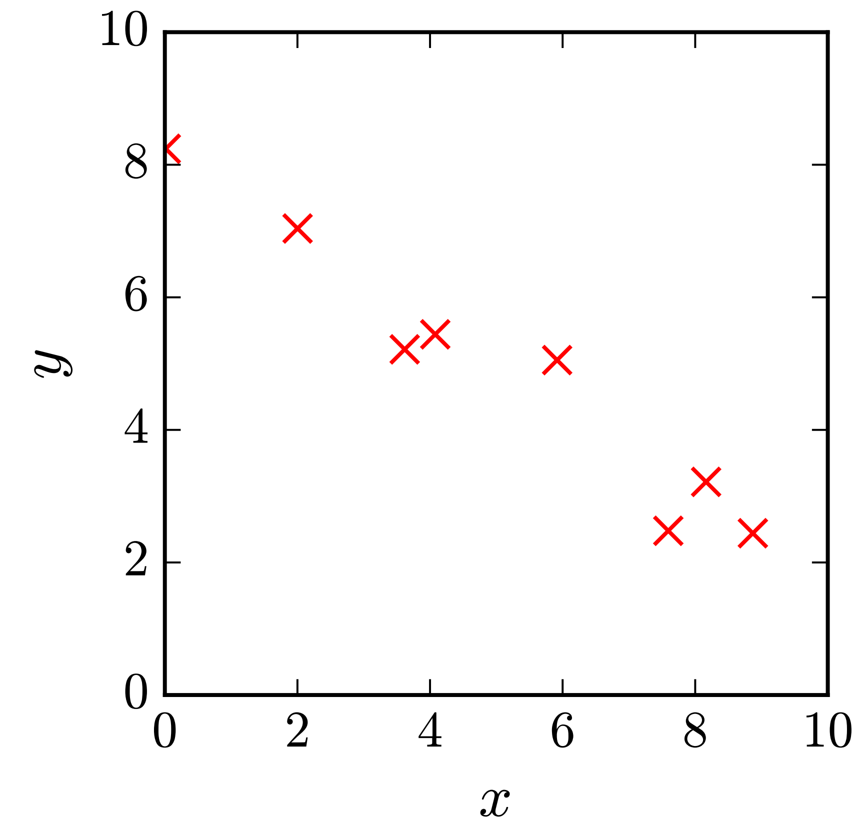
Gaia Artist's impression - credits: ESA/ATG medialab; background image: ESO/S. Brunier ***** June 2013

Is it possible to classify halo stars using only 5-d information?

Brief aside on Machine Learning

https://www.techemergence.com/what-is-machine-learning/

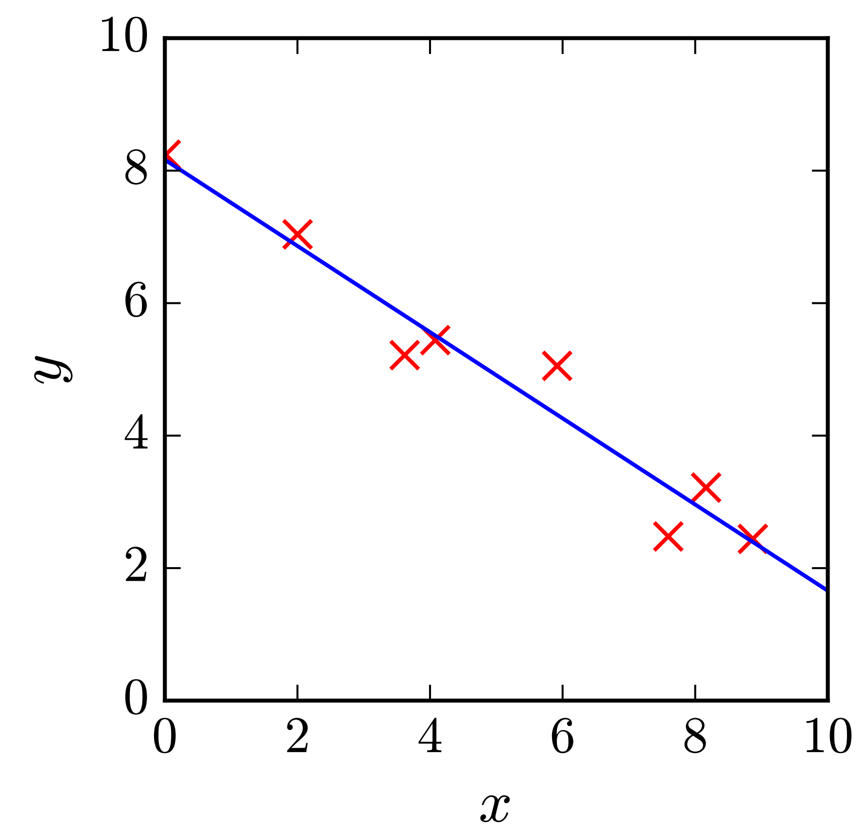
Review: Linear Regression



<u>How to fit data</u>

1. Plot the data
2. Define the function
   - $f(x, \vec{a}) = a_0 + a_1 x$
3. Choose how to know what fits best
   - a.k.a. *Loss Function*
   - MSE: $L(x, y, \vec{a}) = \frac{1}{N} \sum_{i=1}^{N} (f(x_i, \vec{a}) - y_i)^2$
5. Find the minimum error (loss) (cost)
   - $a_{\text{best}} = a \text{ when } \left(\frac{\partial L(x, y, \vec{a})}{\partial \vec{a}} \Big|_{x,y} = 0 \right)$

Review: Linear Regression



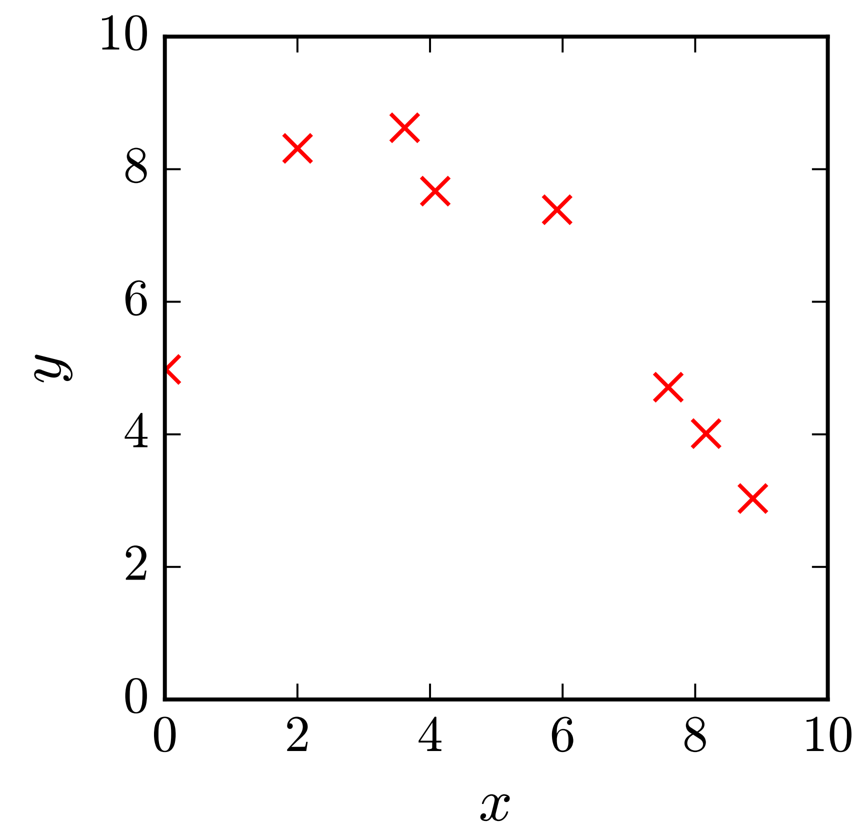
How to fit data

1. Plot the data
2. Define the function
   - $f(x, \vec{a}) = a_0 + a_1 x$
3. Choose how to know what fits best
   - a.k.a. *Loss Function*
   - MSE: $L(x, y, \vec{a}) = \frac{1}{N}\sum_{i=1}^{N}\left(f(x_i, \vec{a}) - y_i\right)^2$
5. Find the minimum error (loss) (cost)
   - $a_{\text{best}} = a \text{ when } \left(\frac{\partial L(x, y, \vec{a})}{\partial \vec{a}}\Big|_{x,y} = 0 \right)$

Review: Linear Regression

<u>How to fit data</u>

1. Plot the data
2. Define the function
   - $f(x, \vec{a}) = a_0 + a_1 x + a_2 x^2$
3. Choose how to know what fits best
   - a.k.a. *Loss Function*
   - MSE: $L(x, y, \vec{a}) = \frac{1}{N}\sum_{i=1}^{N}(f(x_i, \vec{a}) - y_i)^2$
5. Find the minimum error (loss) (cost)
   - $a_{\text{best}} = a \text{ when } \left(\frac{\partial L(x, y, \vec{a})}{\partial \vec{a}}\bigg|_{x,y} = 0 \right)$

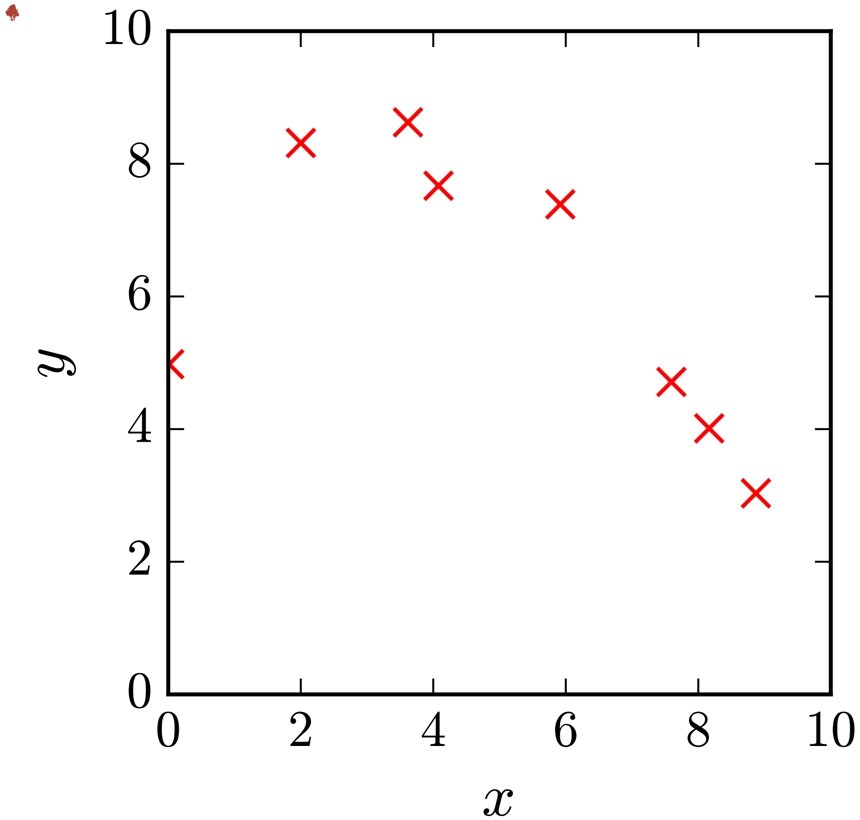


Review: ~~Linear~~ Regression

Quadratic?

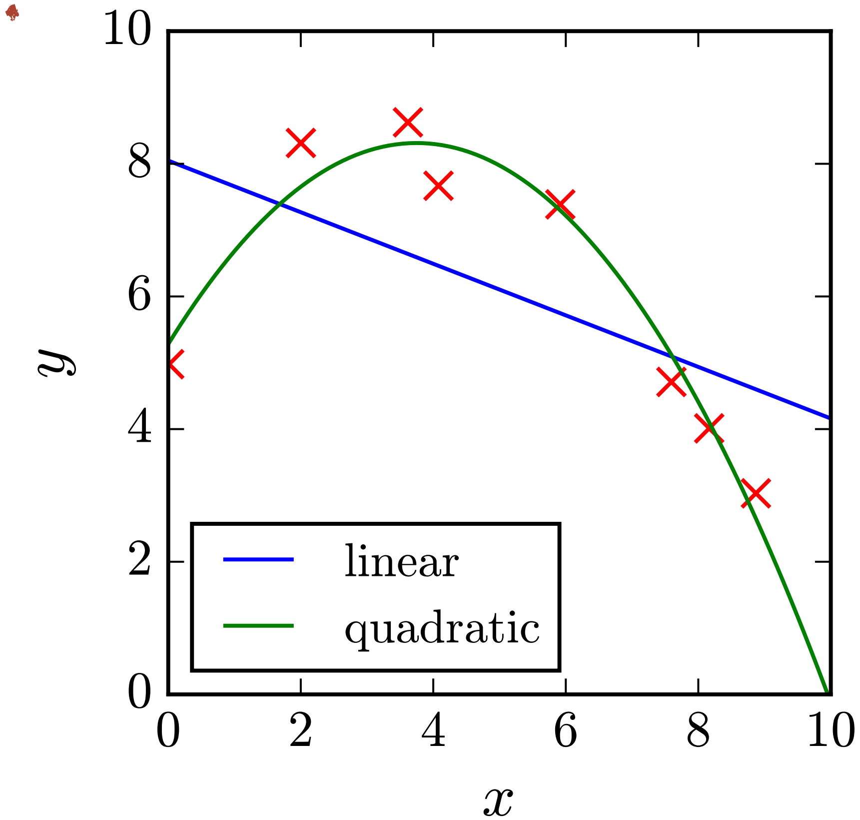
How to fit data

1. Plot the data
2. Define the function
   - $f(x, \vec{a}) = a_0 + a_1 x + a_2 x^2$
3. Choose how to know what fits best
   - a.k.a. *Loss Function*
   - MSE: $L(x, y, \vec{a}) = \frac{1}{N}\sum_{i=1}^{N}\left(f(x_i, \vec{a}) - y_i\right)^2$
5. Find the minimum error (loss) (cost)
   - $a_{\text{best}} = a \text{ when } \left(\left.\frac{\partial L(x, y, \vec{a})}{\partial \vec{a}}\right|_{x,y} = 0\right)$



Review: ~~Linear~~ Regression

Quadratic?

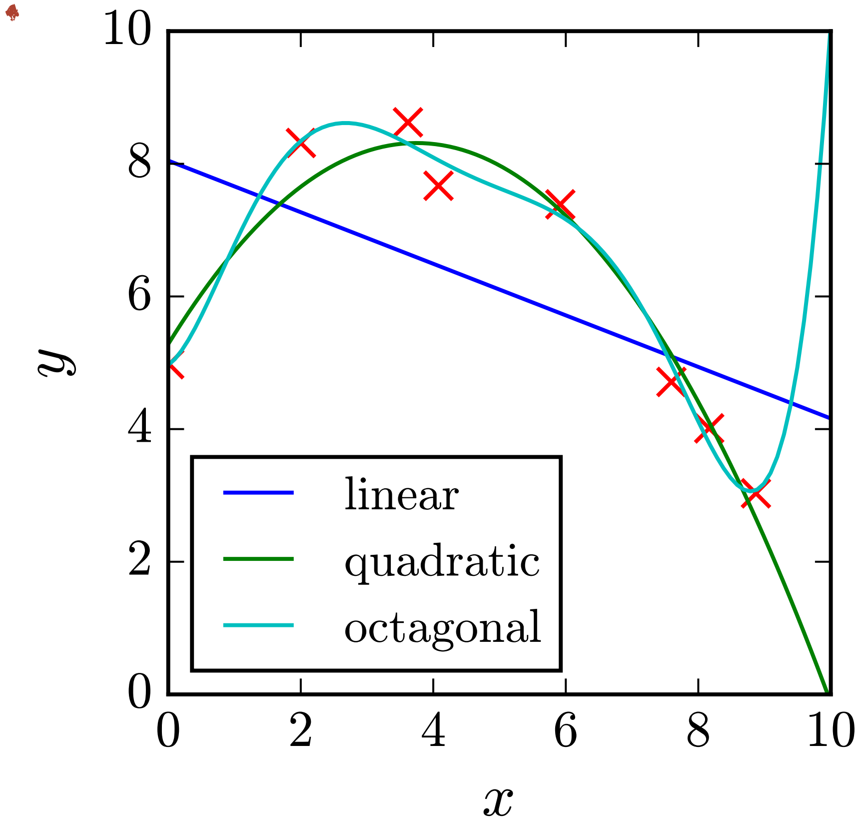


How to fit data

1. Plot the data
2. Define the function
   - $f(x, \vec{a}) = a_0 + a_1 x + a_2 x^2$
3. Choose how to know what fits best
   - a.k.a. *Loss Function*
   - MSE: $L(x, y, \vec{a}) = \frac{1}{N} \sum_{i=1}^{N} (f(x_i, \vec{a}) - y_i)^2$
5. Find the minimum error (loss) (cost)
   - $a_{\text{best}} = a \text{ when } \left(\frac{\partial L(x, y, \vec{a})}{\partial \vec{a}} \bigg|_{x,y} = 0 \right)$

Review: Linear Regression

Quadratic?



How to fit data

1. Plot the data
2. Define the function
   - $f(x, \vec{a}) = a_0 + a_1 x + a_2 x^2$
3. Choose how to know what fits best
   - a.k.a. *Loss Function*
   - MSE: $L(x, y, \vec{a}) = \frac{1}{N} \sum_{i=1}^{N} (f(x_i, \vec{a}) - y_i)^2$
5. Find the minimum error (loss) (cost)
   - $a_{\text{best}} = a \text{ when } \left(\frac{\partial L(x, y, \vec{a})}{\partial \vec{a}} \Big|_{x,y} = 0 \right)$

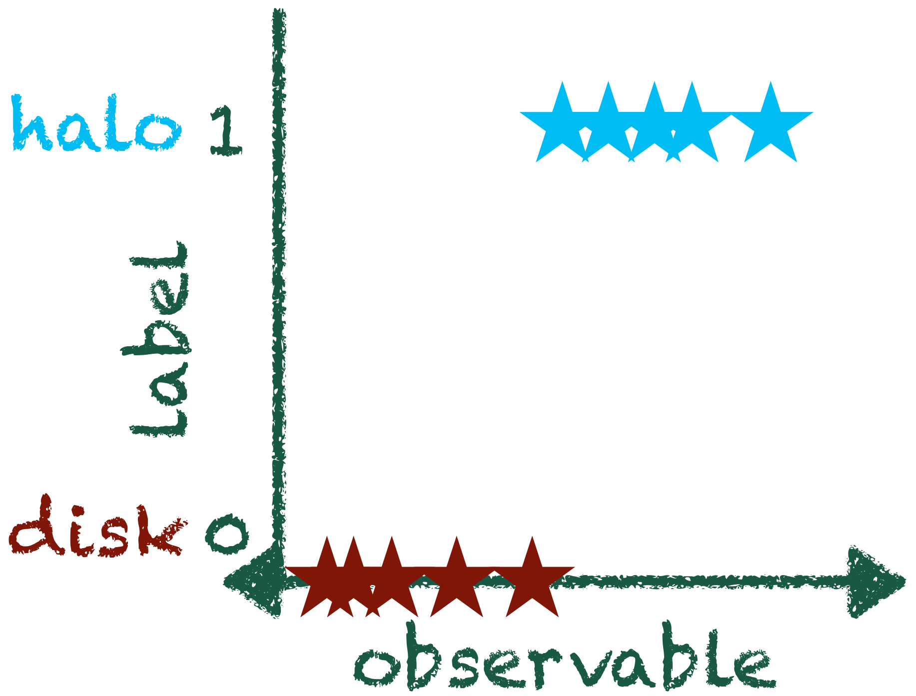
Is that good enough? How many parameters can we add?

Logistic Regression

What if we are trying to predict a class, not a number?

Logistic Regression

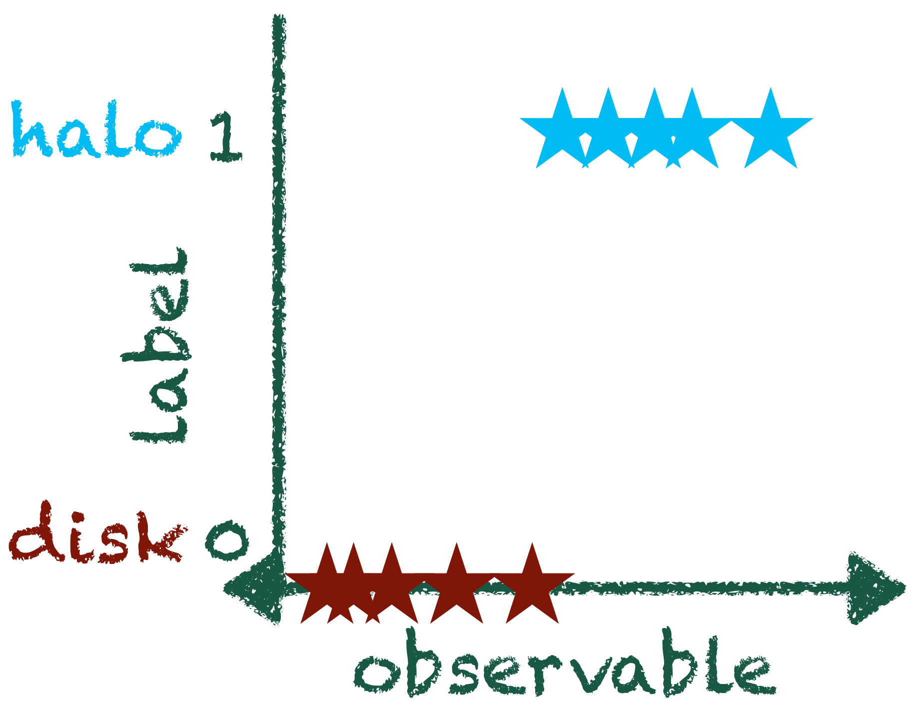
What if we are trying to predict a class, not a number?



Logistic Regression

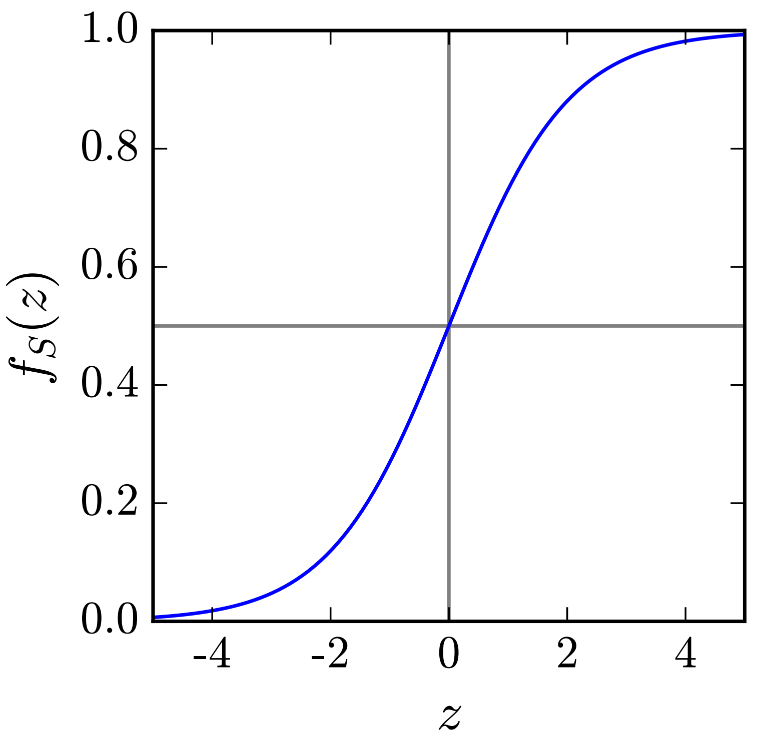
What if we are trying to predict a class, not a number?

- Change the shape of function: Logistic/Sigmoid function



$$f_S(z) = \frac{1}{1 + e^{-z}}$$

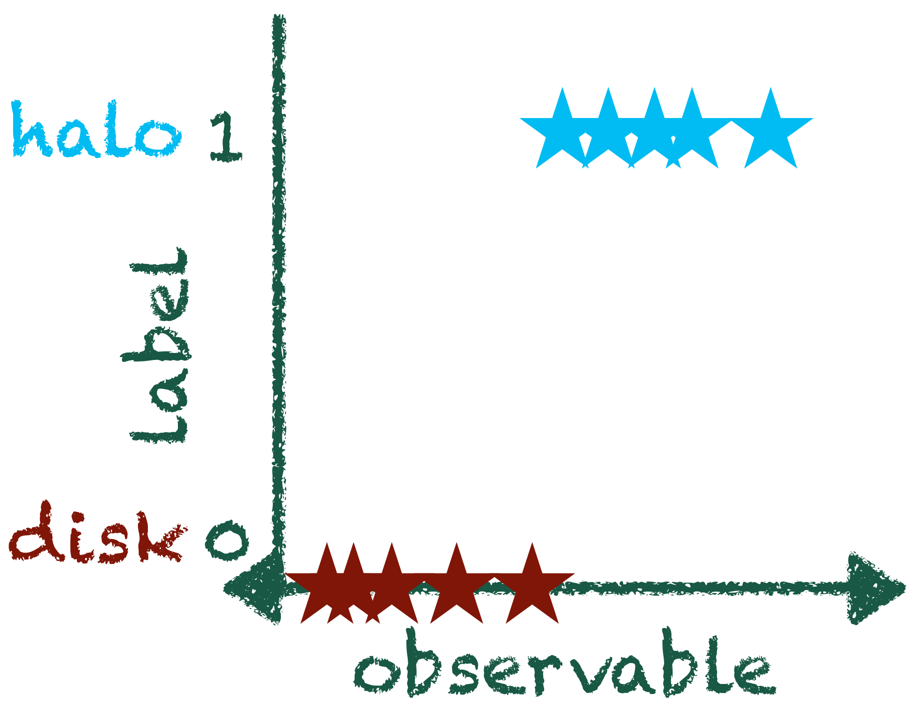
Does not add parameters



Logistic Regression

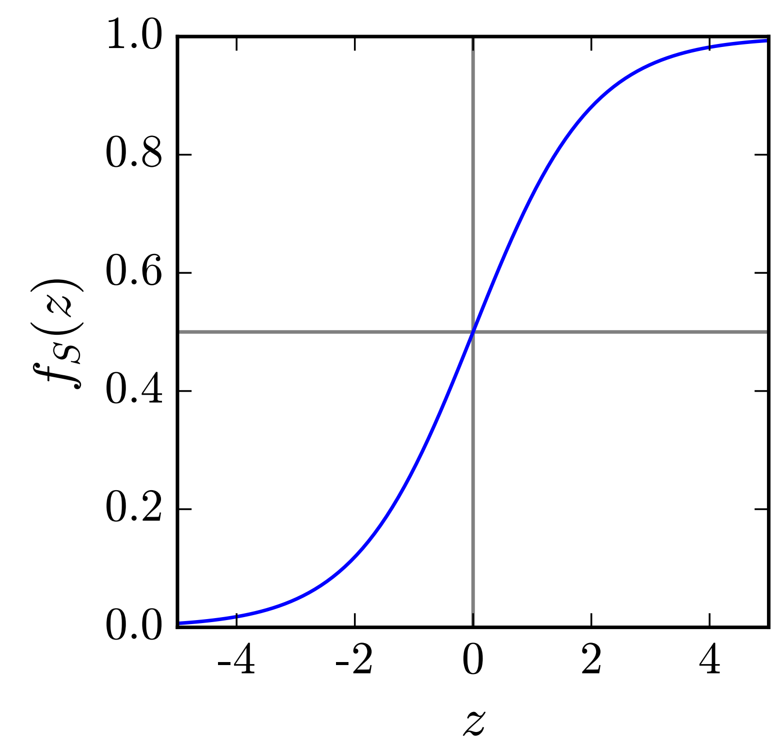
What if we are trying to predict a class, not a number?

- Change the shape of function: Logistic/Sigmoid function



$$f_S(z) = \frac{1}{1 + e^{-z}}$$

Does not add parameters

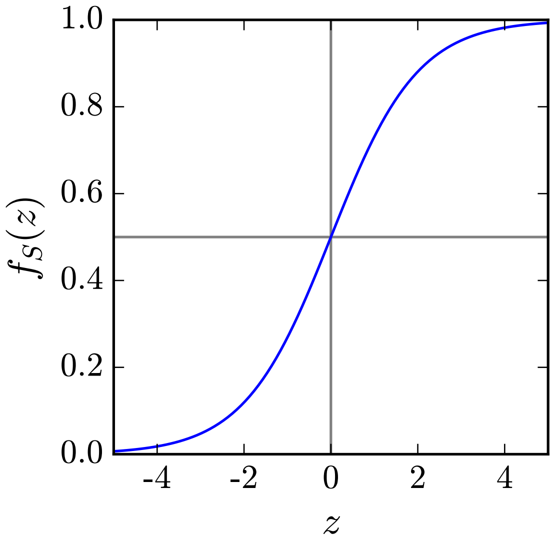


- Change the loss function: BCE

$$L(\vec{x}, \vec{y}, \vec{a}) = -\frac{1}{N}\sum_{i=1}^{N}\left(y_i \log\Big(f_S(p(x,a))\Big) + (1 - y_i)\log\Big(1 - f_S(p(x,a))\Big)\right)$$

Logistic Regression

What if we are trying to predict a class, not a number?

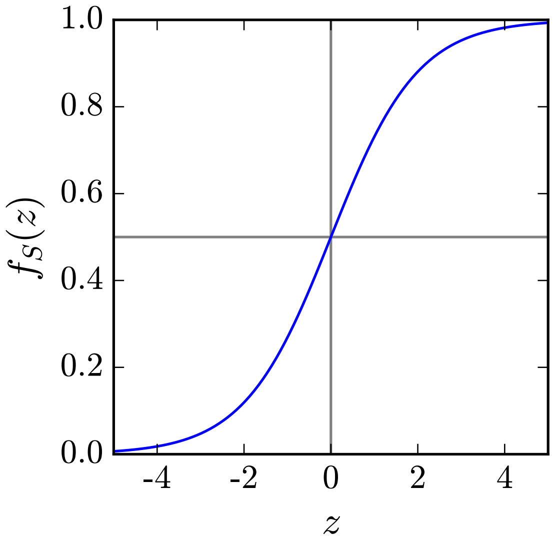


$$L(\vec{x}, \vec{y}, \vec{a}) = -\frac{1}{N}\sum_{i=1}^{N}\left(y_i \log\Big(f_S(p(x,a))\Big) + (1-y_i)\log\Big(1 - f_S(p(x,a))\Big)\right)$$

$$f_S(z) = \frac{1}{1+\mathrm{e}^{-z}} \qquad z = p(x,a)$$

Logistic Regression

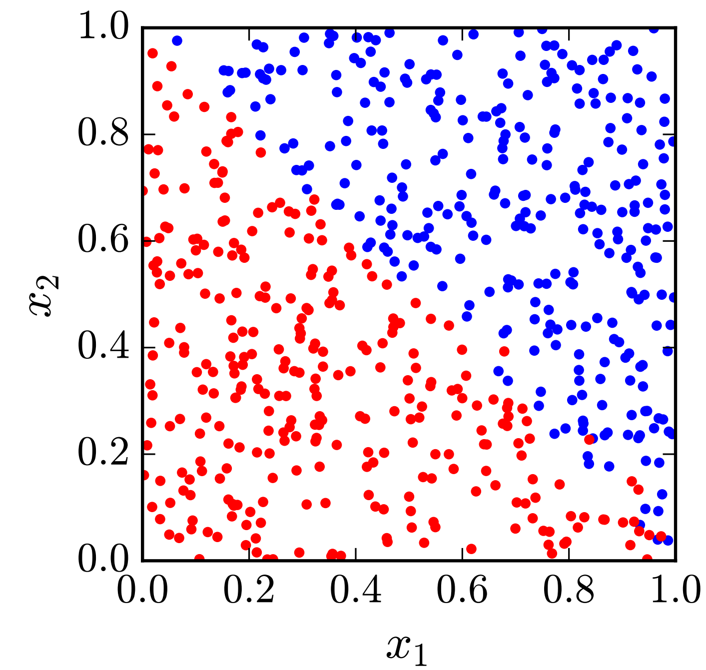
What if we are trying to predict a class, not a number?



$$L(\vec{x},\vec{y},\vec{a}) = -\frac{1}{N}\sum_{i=1}^{N}\left(y_i \log\Big(f_S(p(x,a))\Big) + (1-y_i)\log\Big(1-f_S(p(x,a))\Big)\right)$$

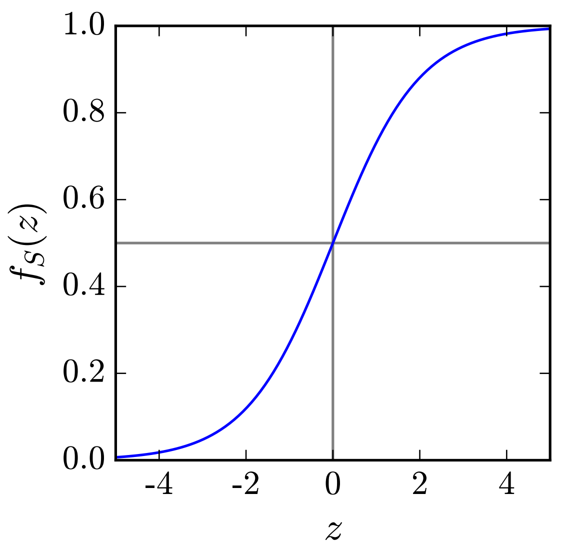
$$f_S(z) = \frac{1}{1+\mathrm{e}^{-z}} \qquad z = p(x,a)$$

What is $p(x,a)$?



Logistic Regression

What if we are trying to predict a class, not a number?

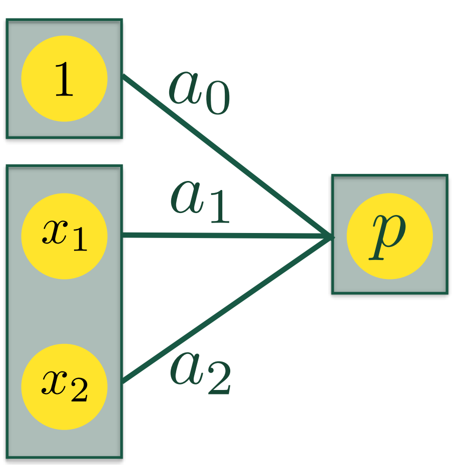


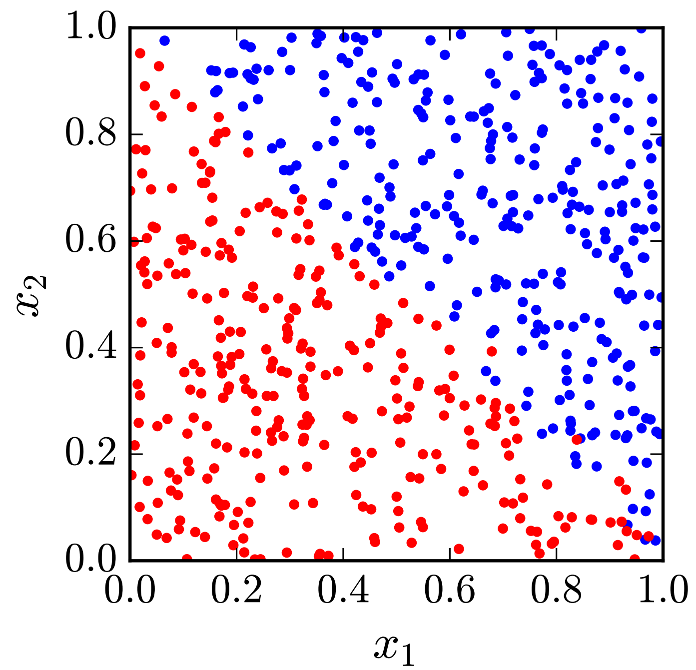
$$L(\vec{x},\vec{y},\vec{a}) = -\frac{1}{N}\sum_{i=1}^{N}\left(y_i \log\left(f_S(p(x,a))\right) + (1-y_i)\log\left(1 - f_S(p(x,a))\right)\right)$$

$$f_S(z) = \frac{1}{1+\mathrm{e}^{-z}} \qquad z = p(x,a)$$

What is $p(x,a)$?

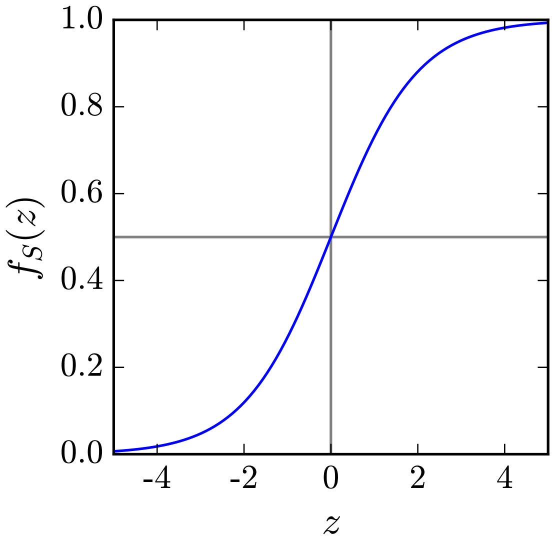
$$p(x,a) = a_0 + x_1 a_1 + x_2 a_2$$





Logistic Regression

What if we are trying to predict a class, not a number?

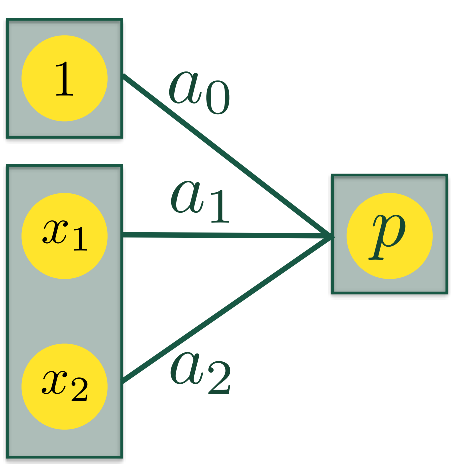


$$L(\vec{x},\vec{y},\vec{a}) = -\frac{1}{N}\sum_{i=1}^{N}\Big(y_i \log\big(f_S(p(x,a))\big) + (1-y_i)\log\big(1-f_S(p(x,a))\big)\Big)$$

$$f_S(z) = \frac{1}{1+\mathrm{e}^{-z}} \qquad z = p(x,a)$$

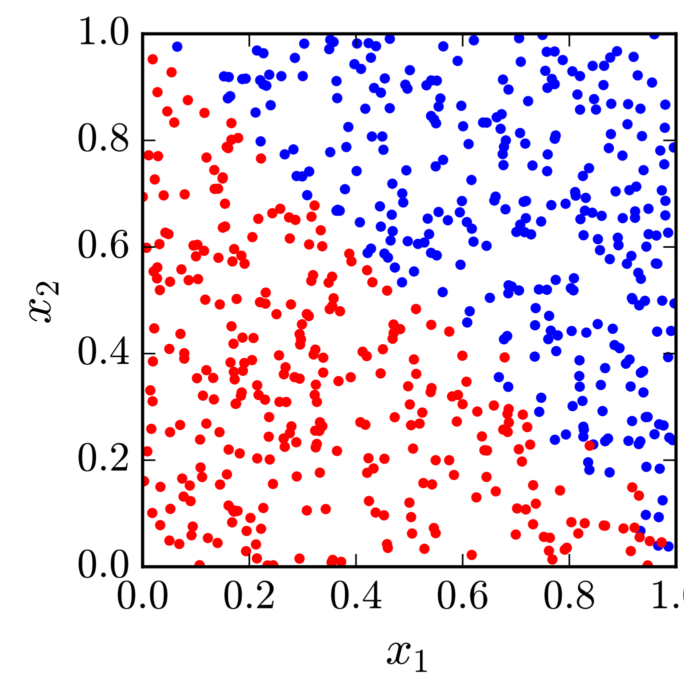
What is $p(x,a)$?

$$p(x,a) = a_0 + x_1 a_1 + x_2 a_2$$



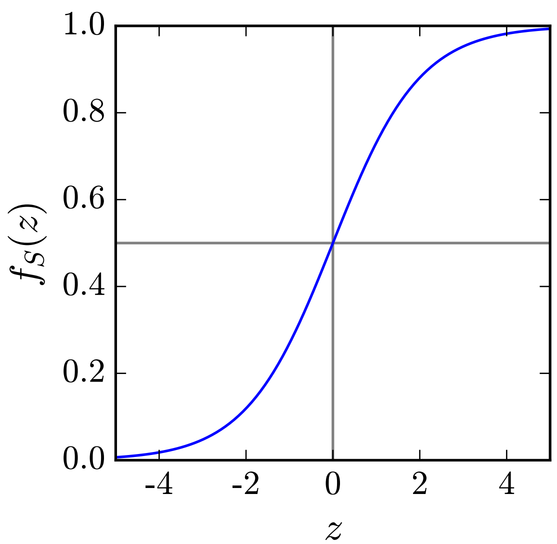
Minimize the loss with respect to $\vec{a}$

Boundary at $p(x,a) = 0$



Logistic Regression

What if we are trying to predict a class, not a number?

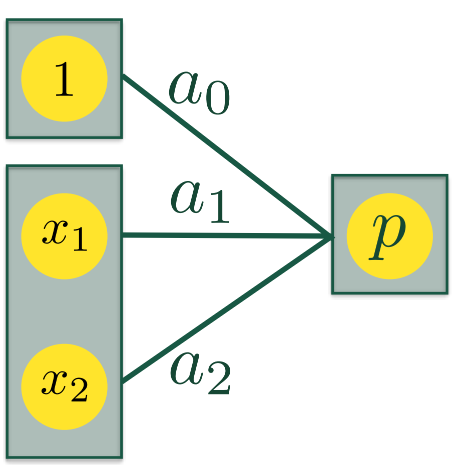


$$L(\vec{x},\vec{y},\vec{a}) = -\frac{1}{N}\sum_{i=1}^{N}\left(y_i \log\left(f_S(p(x,a))\right) + (1-y_i)\log\left(1-f_S(p(x,a))\right)\right)$$

$$f_S(z) = \frac{1}{1+\mathrm{e}^{-z}} \qquad z = p(x,a)$$

What is $p(x,a)$?

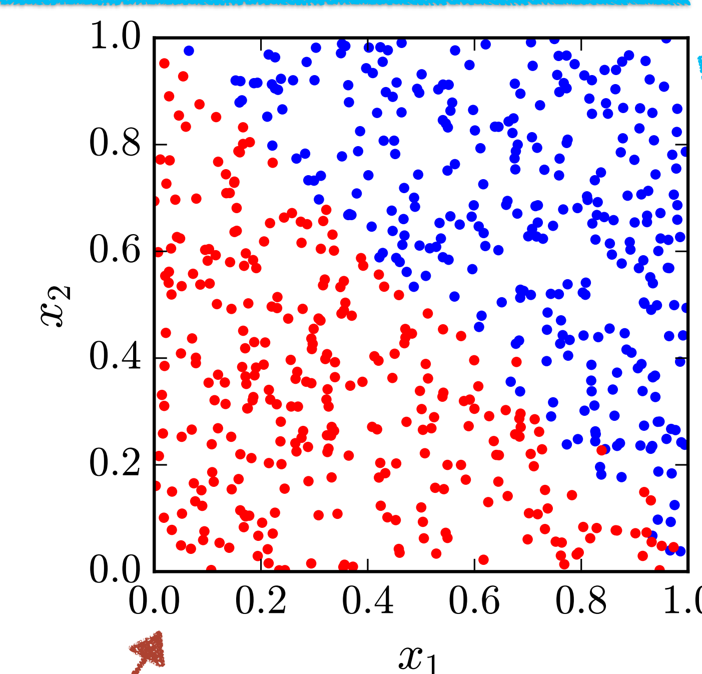
$$p(x,a) = a_0 + x_1 a_1 + x_2 a_2$$



Minimize the loss with respect to $\vec{a}$

Boundary at $p(x,a) = 0$

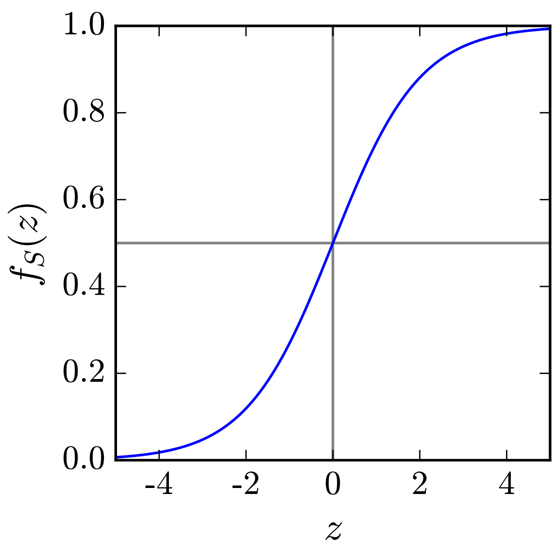
Large + values of p



Large - values of p

Logistic Regression

What if we are trying to predict a class, not a number?

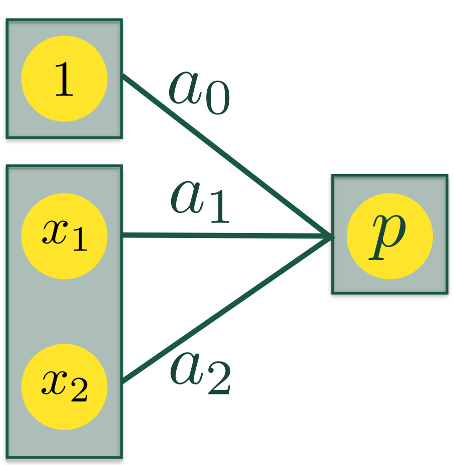


$$L(\vec{x}, \vec{y}, \vec{a}) = -\frac{1}{N} \sum_{i=1}^{N} \Big(y_i \log\big(f_S(p(x,a))\big) + (1 - y_i) \log\big(1 - f_S(p(x,a))\big) \Big)$$

$$f_S(z) = \frac{1}{1 + \mathrm{e}^{-z}} \qquad z = p(x,a)$$

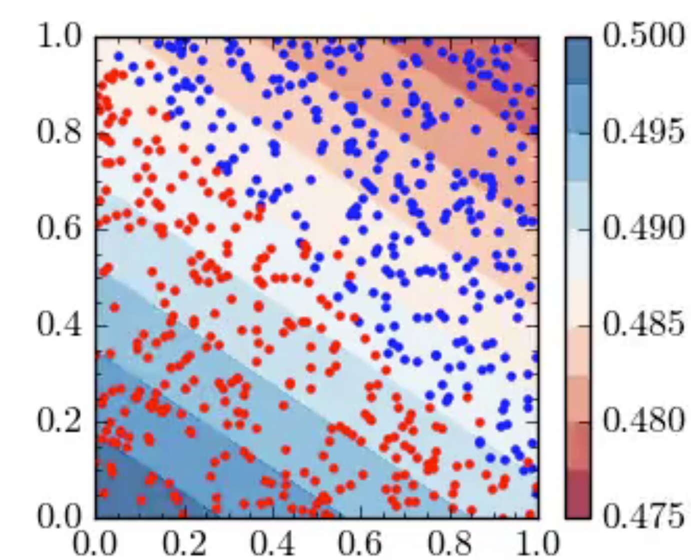
What is $p(x,a)$?

$$p(x,a) = a_0 + x_1 a_1 + x_2 a_2$$



Minimize the loss with respect to $\vec{a}$

Boundary at $p(x,a) = 0$

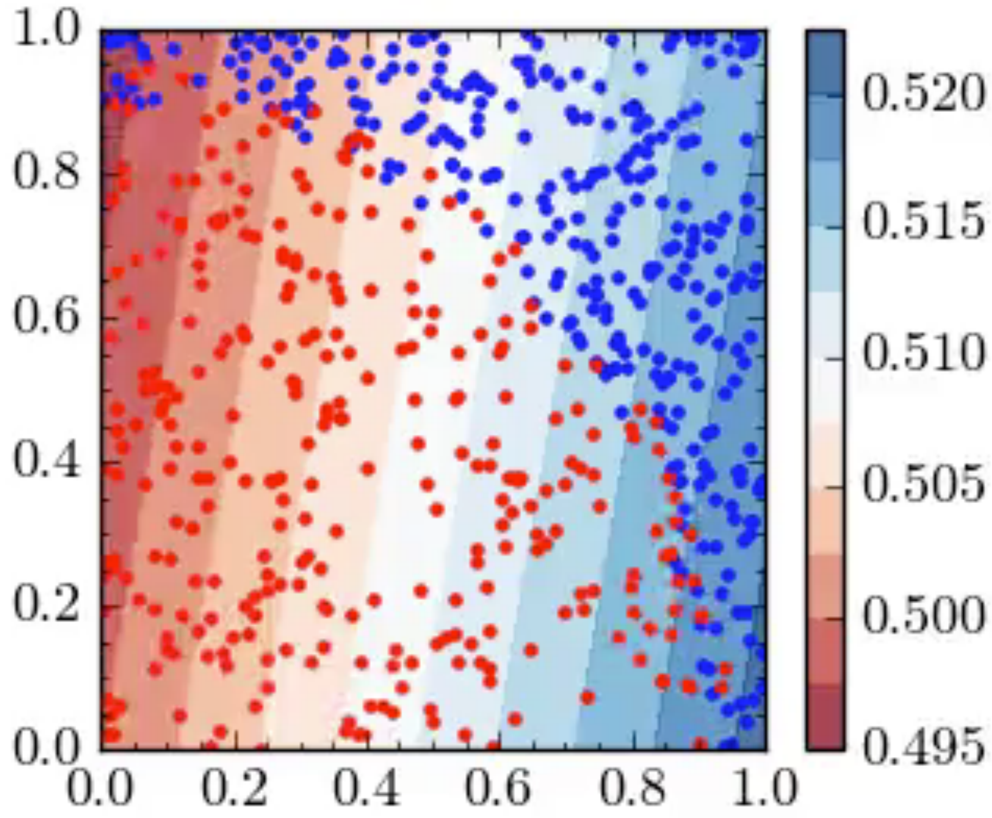


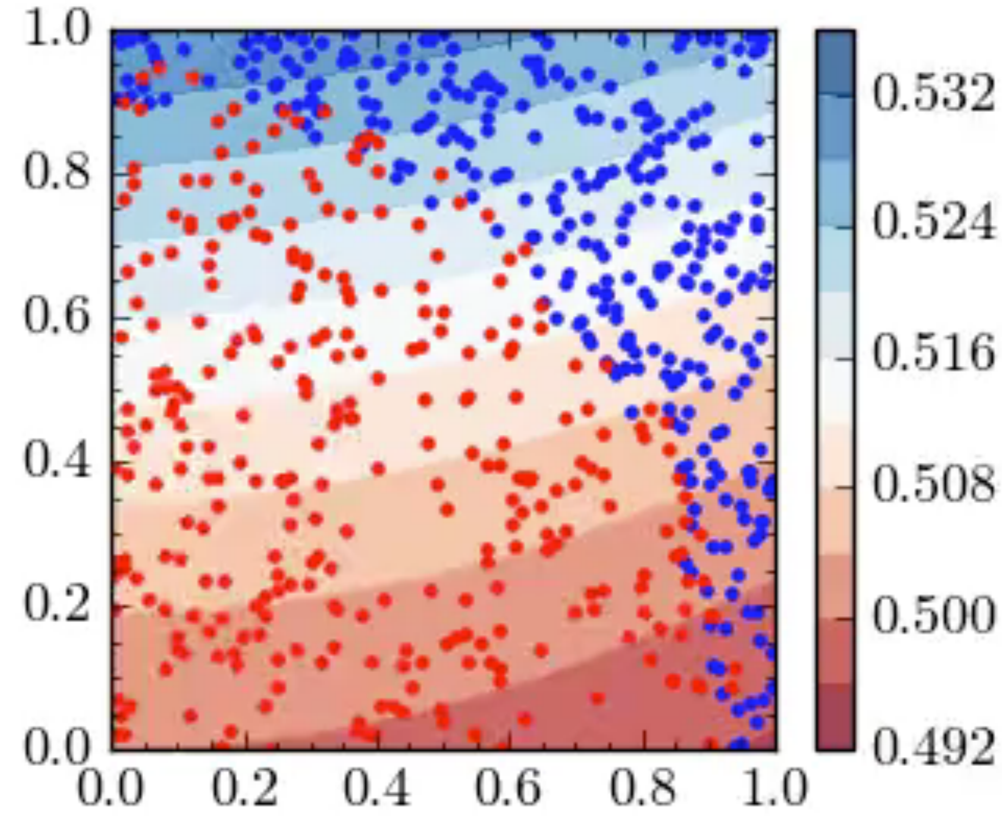
Logistic Regression

What if there is a shape in the data?

$$p(x,a) = a_0 + x_1 a_1 + x_2 a_2$$

$$p(x,a) = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_1^2 + a_4 x_2^2 + a_5 x_1 x_2$$



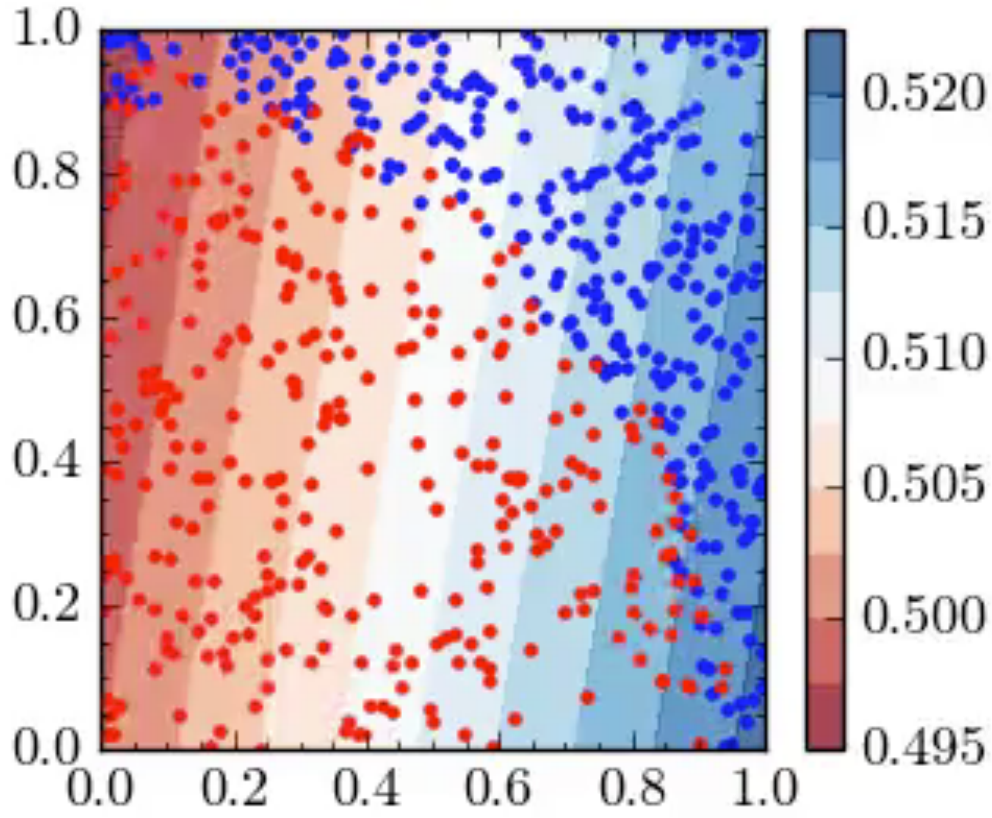


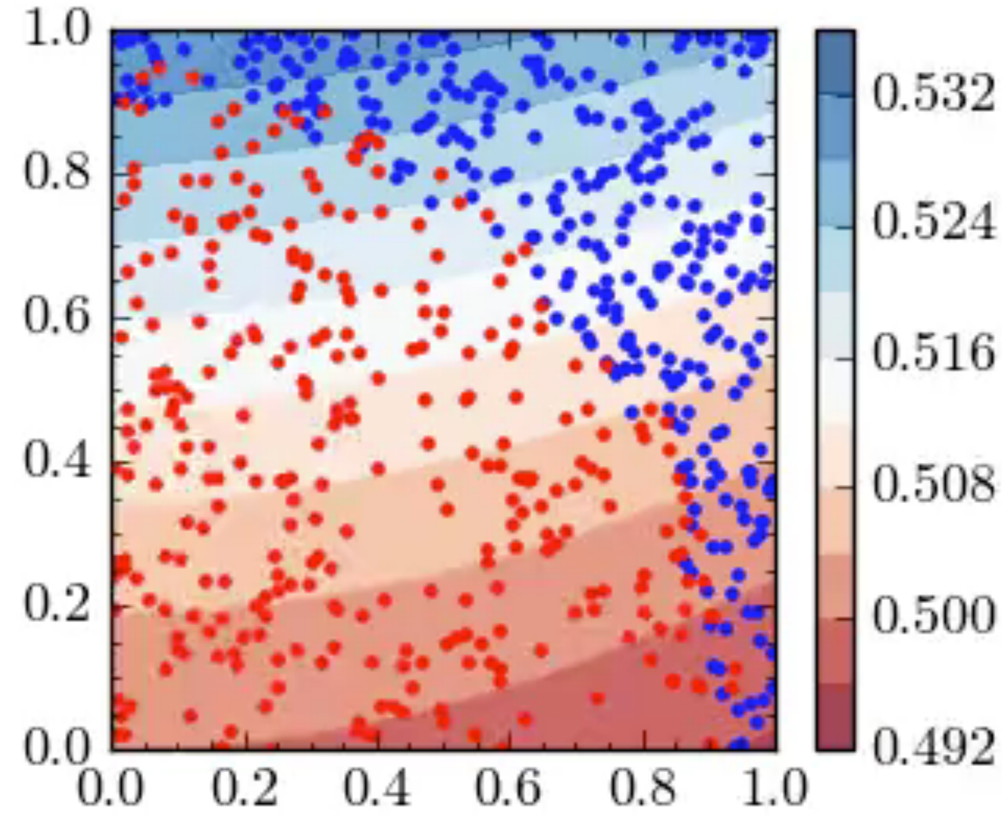
Logistic Regression

What if there is a shape in the data?

$$p(x, a) = a_0 + x_1 a_1 + x_2 a_2$$

$$p(x, a) = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_1^2 + a_4 x_2^2 + a_5 x_1 x_2$$





Regression Review

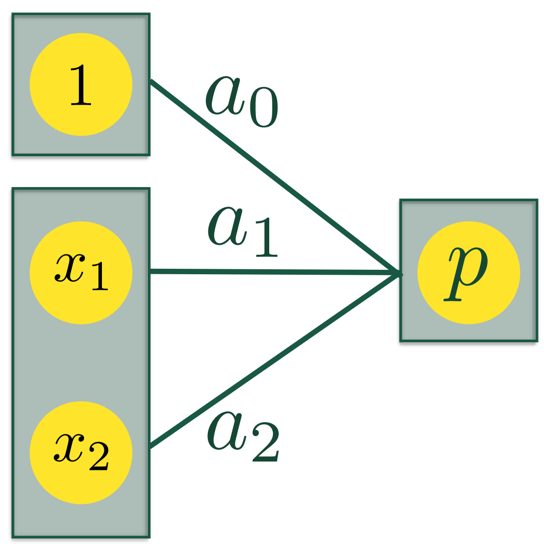
1. Can use nearly the same process for fitting a curve (predicting a number) or classification

2. Minimize a defined cost function

3. Easy to add parameters if shape is unknown — worry about over-fitting

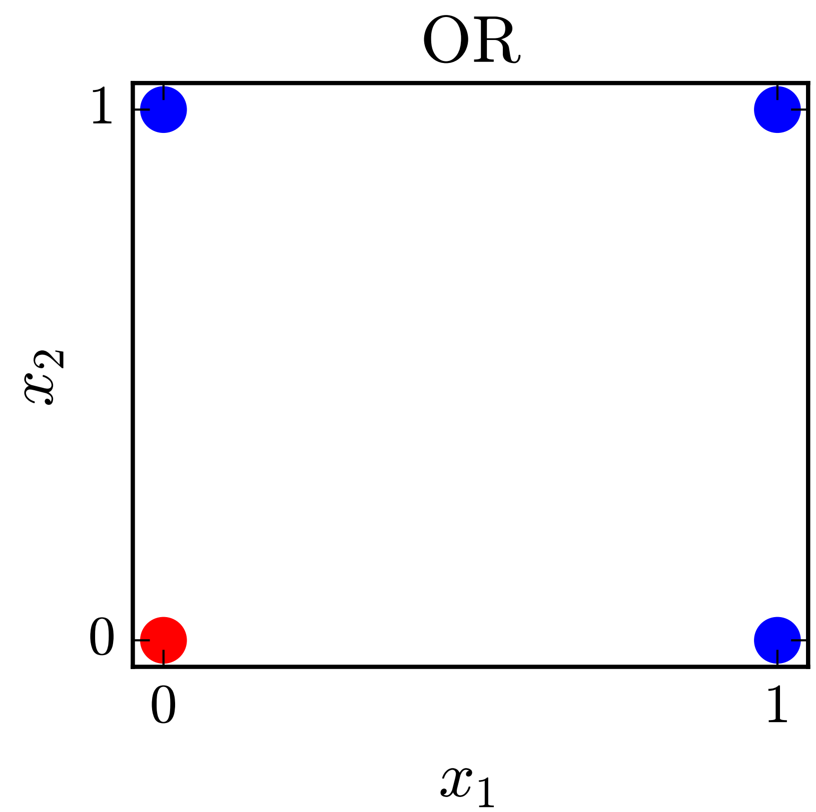
4. If many inputs and complicated shapes, number of parameters necessary grows very quickly

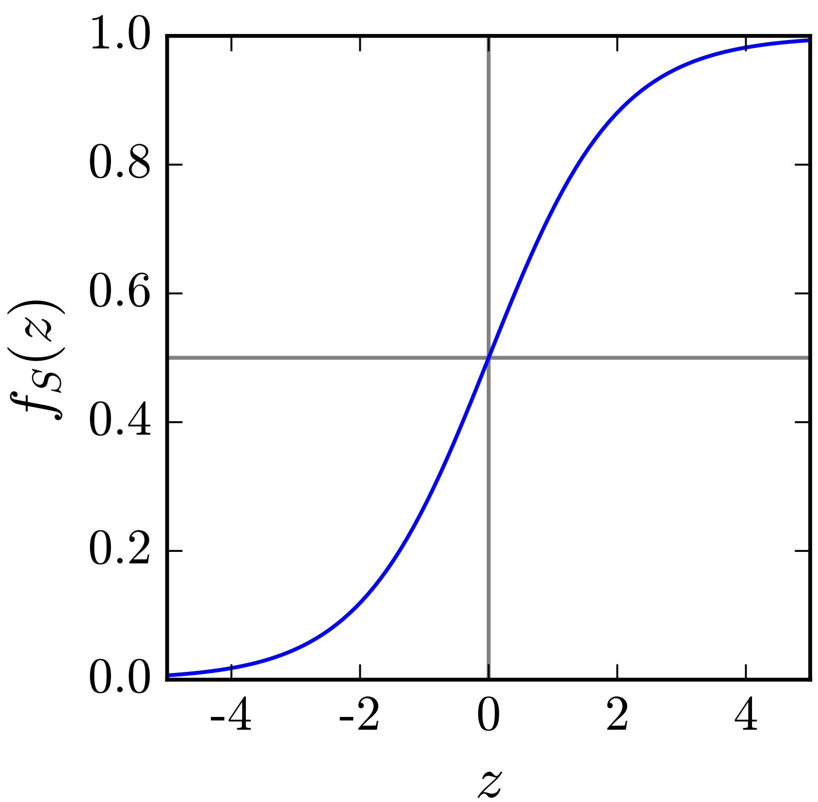
Neural Networks



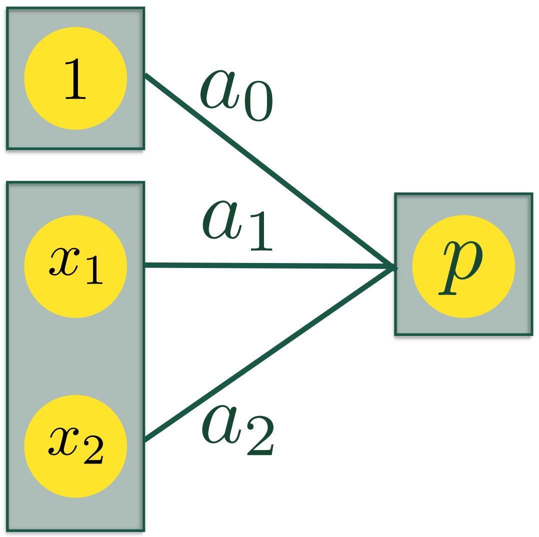
OR

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	1



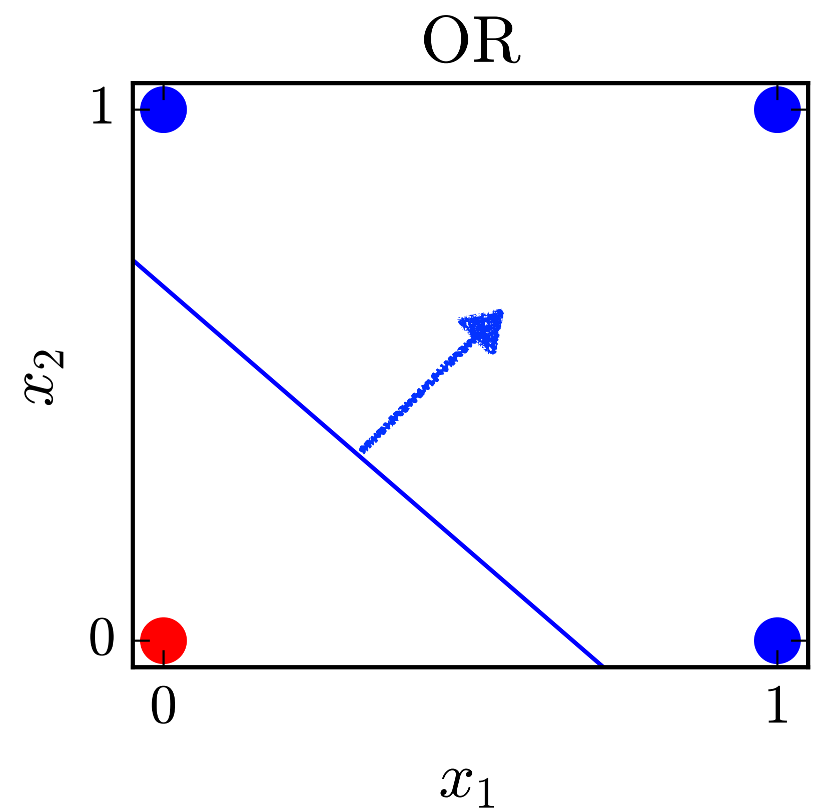


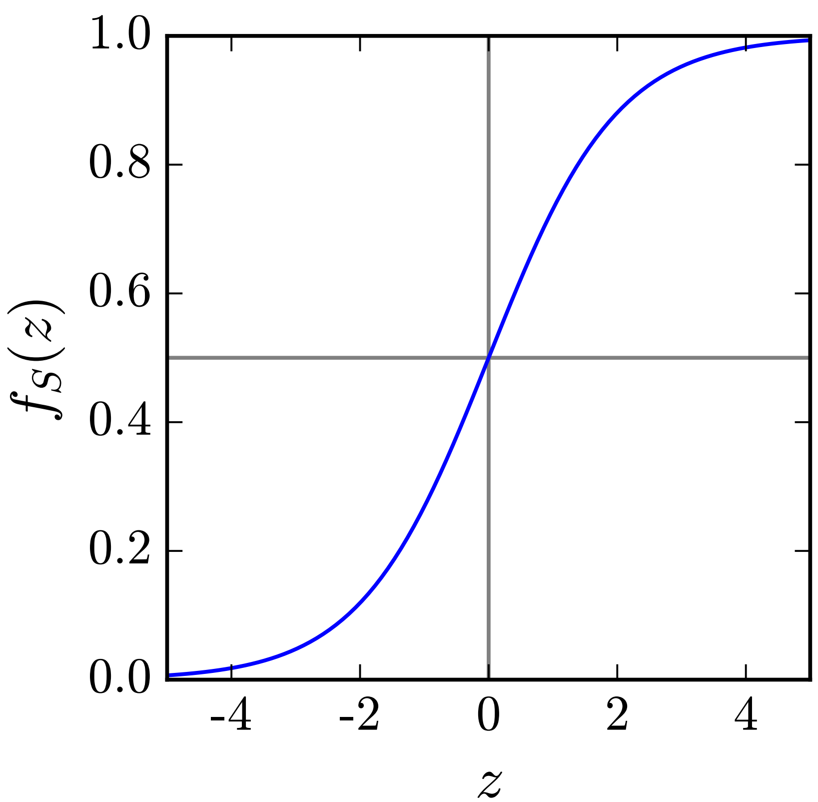
Neural Networks



OR

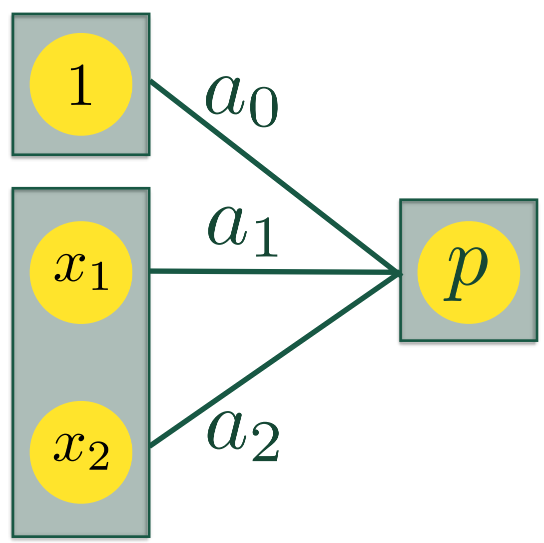
x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	1

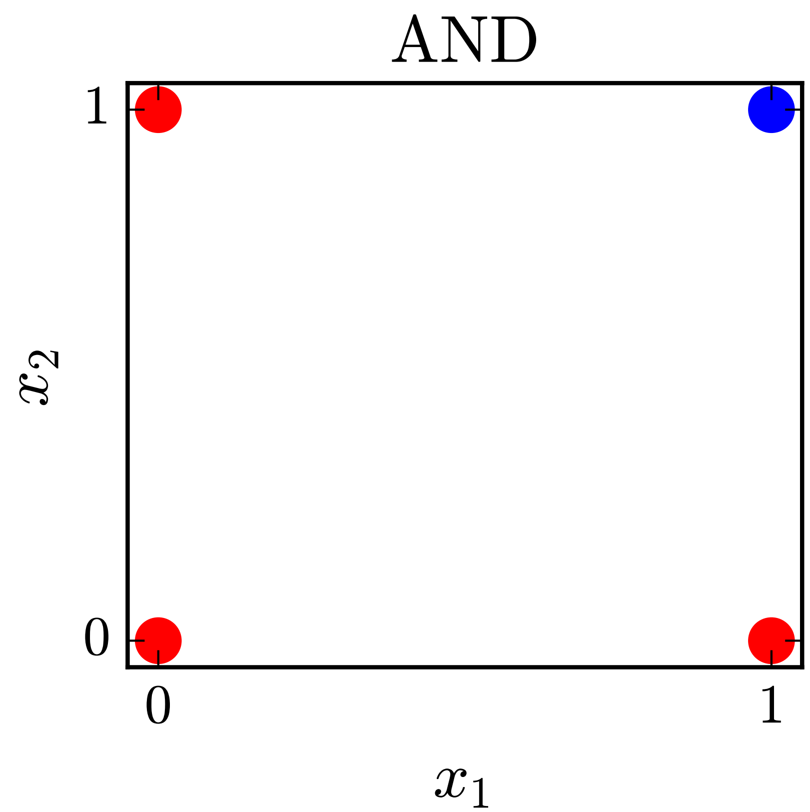




$$a_0 = -10, \quad a_1 = 15, \quad a_2 = 15$$

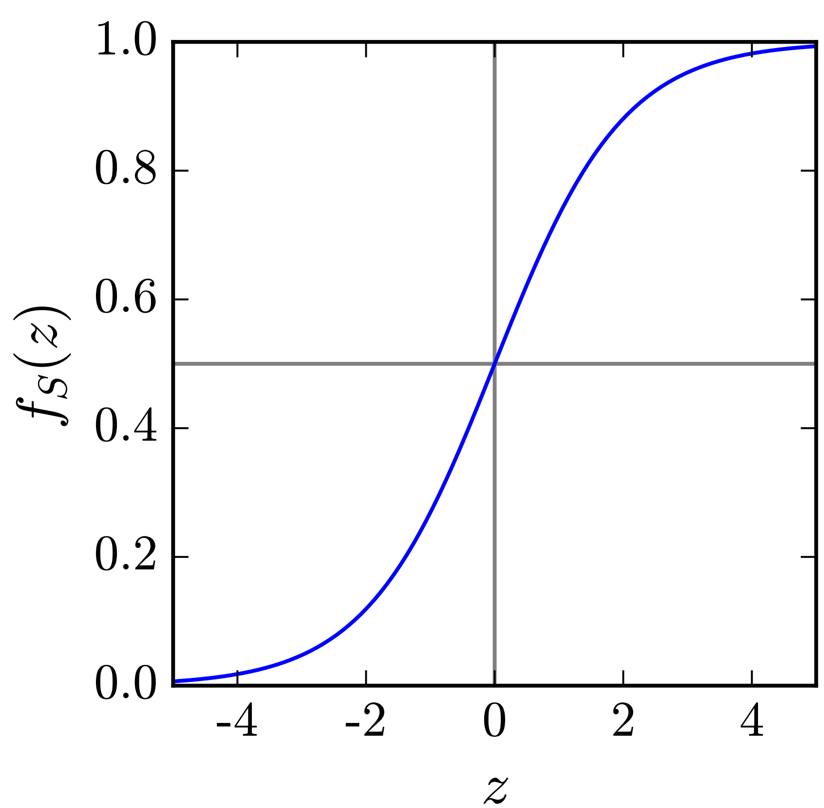
Neural Networks



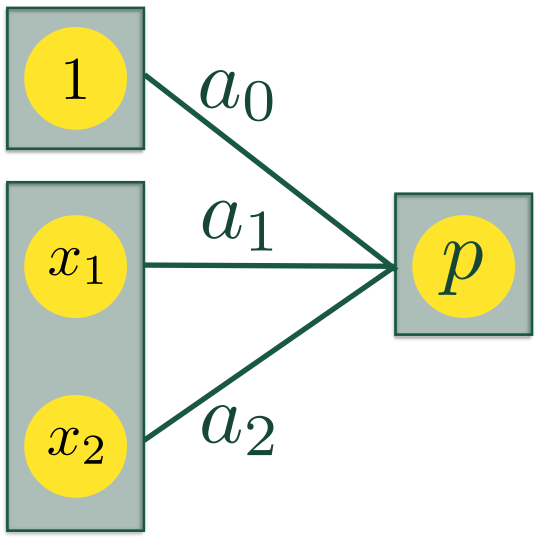


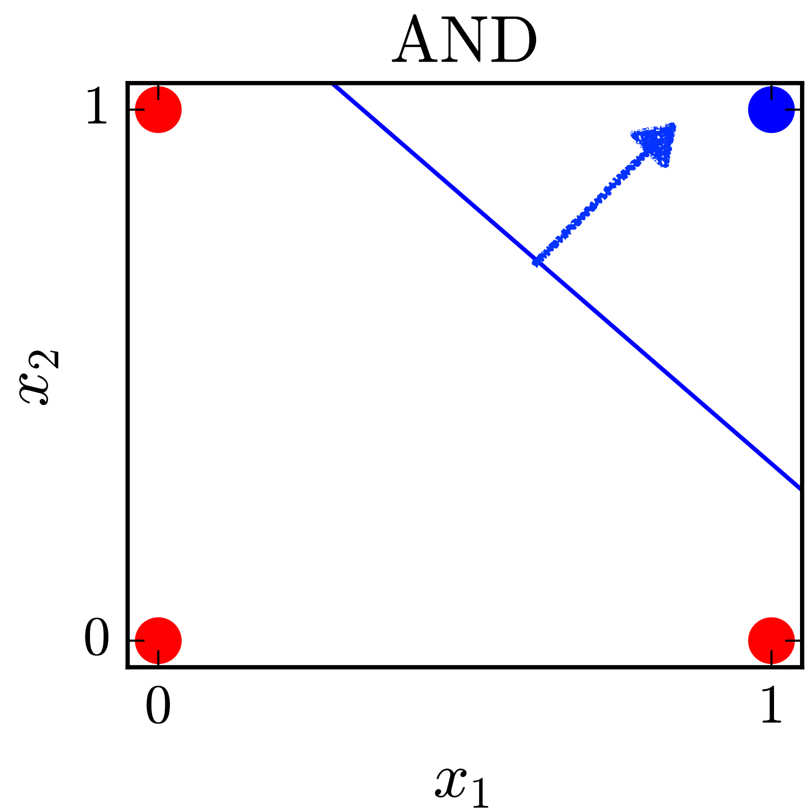
AND

x_1	x_2	y
0	0	0
0	1	0
1	0	0
1	1	1



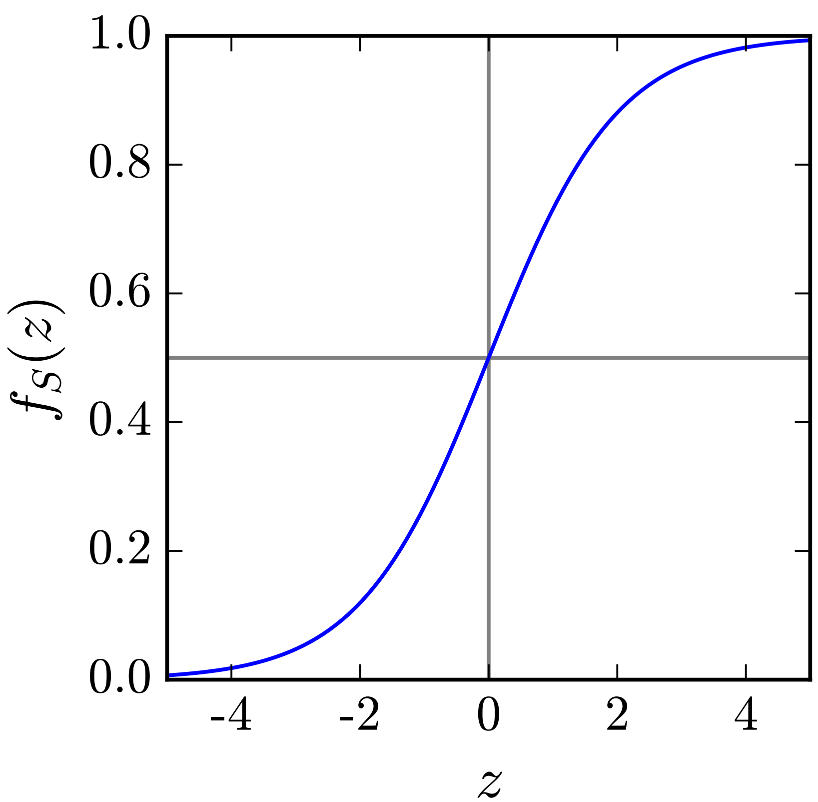
Neural Networks





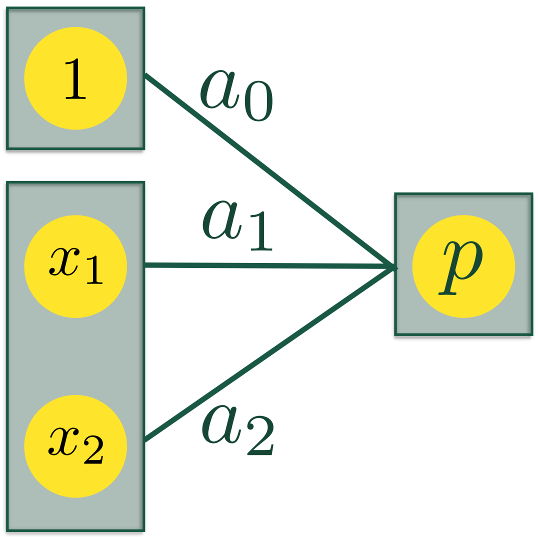
AND

x_1	x_2	y
0	0	0
0	1	0
1	0	0
1	1	1



$$a_0 = -20, \quad a_1 = 15, \quad a_2 = 15$$

Neural Networks



OR

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	1

$$a_0 = -10, \quad a_1 = 15, \quad a_2 = 15$$

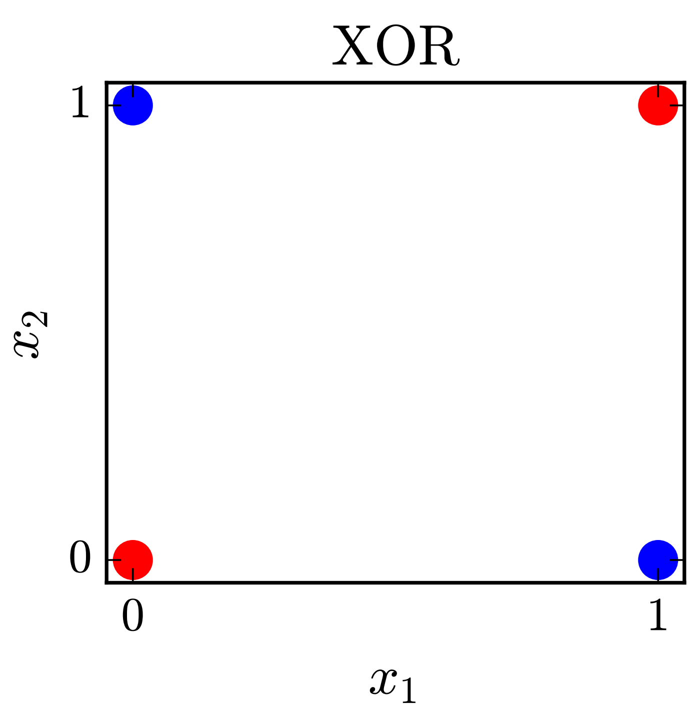
AND

x_1	x_2	y
0	0	0
0	1	0
1	0	0
1	1	1

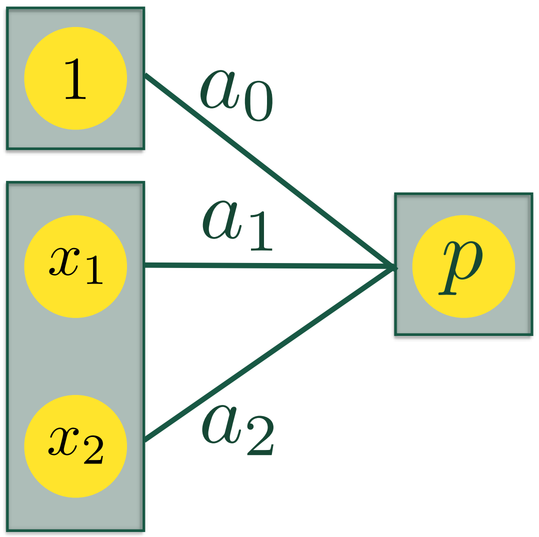
$$a_0 = -20, \quad a_1 = 15, \quad a_2 = 15$$

XOR

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0



Neural Networks



OR

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	1

$$a_0 = -10, \quad a_1 = 15, \quad a_2 = 15$$

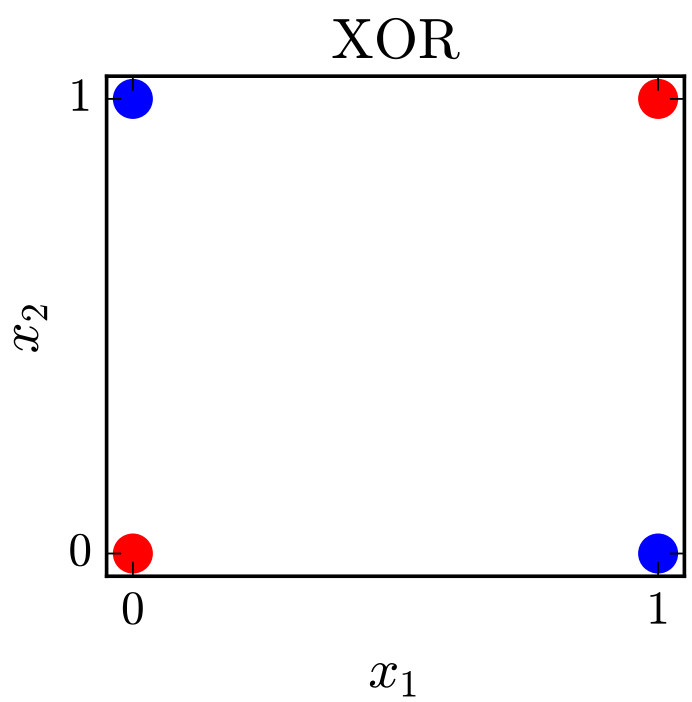
AND

x_1	x_2	y
0	0	0
0	1	0
1	0	0
1	1	1

$$a_0 = -20, \quad a_1 = 15, \quad a_2 = 15$$

XOR

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0



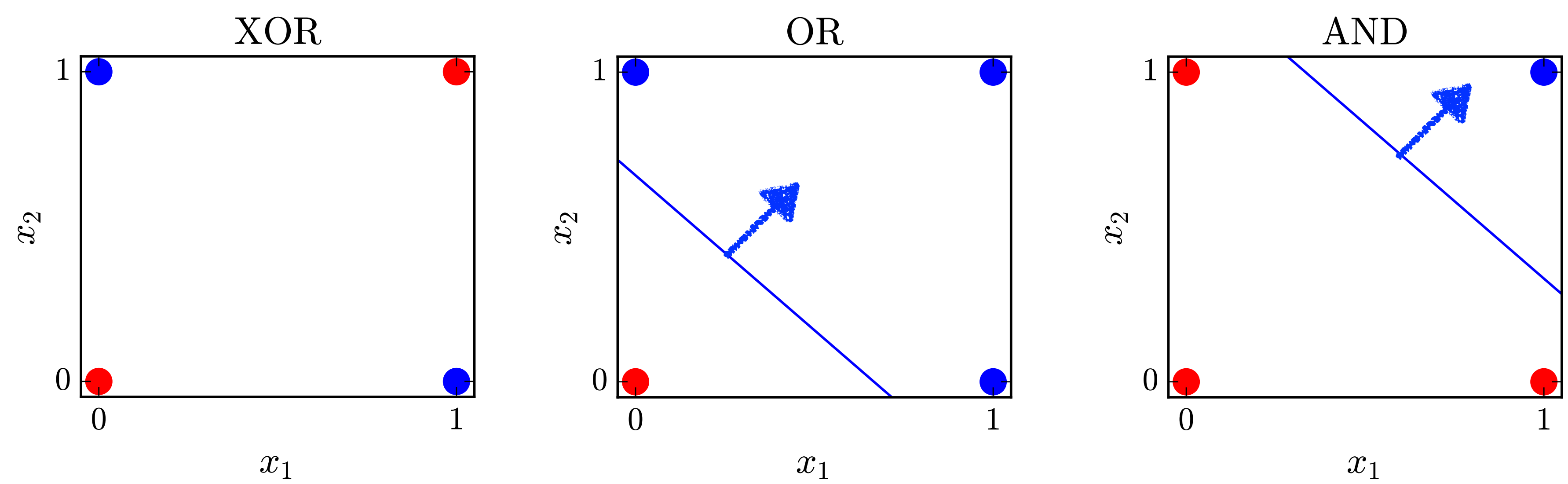
This system cannot produce XOR

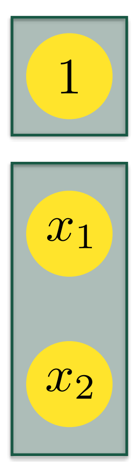
(cannot make a two sided cut)

Neural Networks

XOR

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0

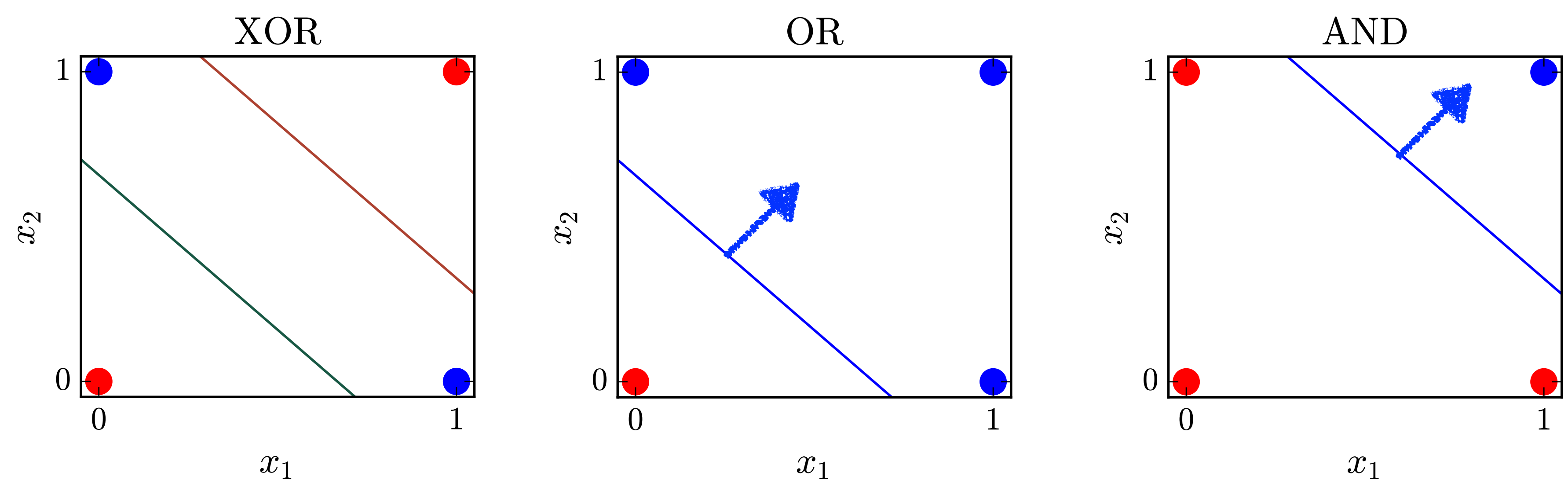


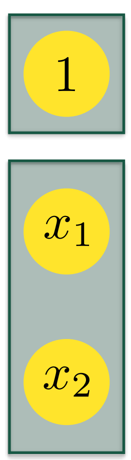


Neural Networks

XOR

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0

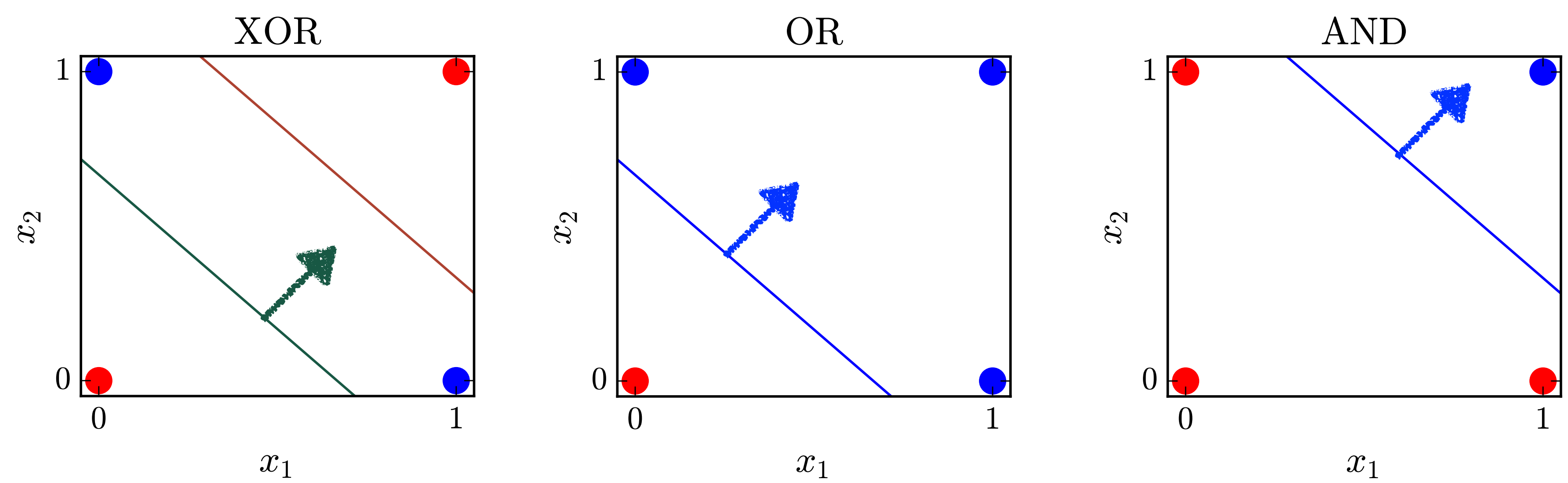


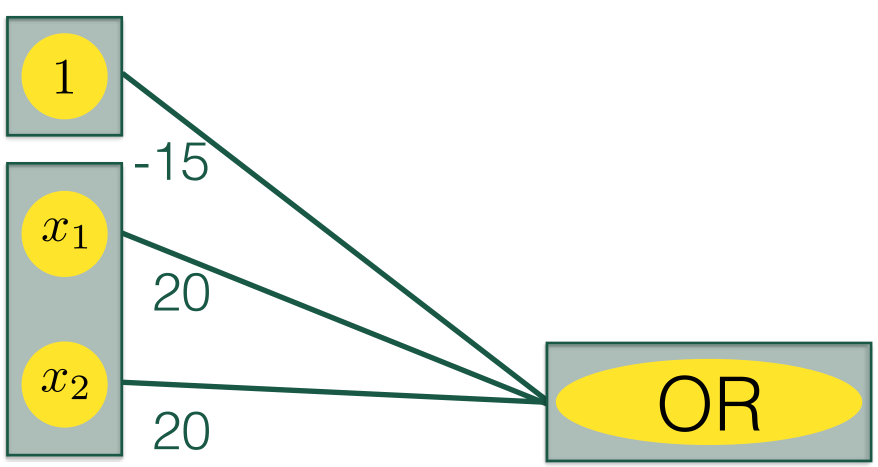


Neural Networks

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0

XOR

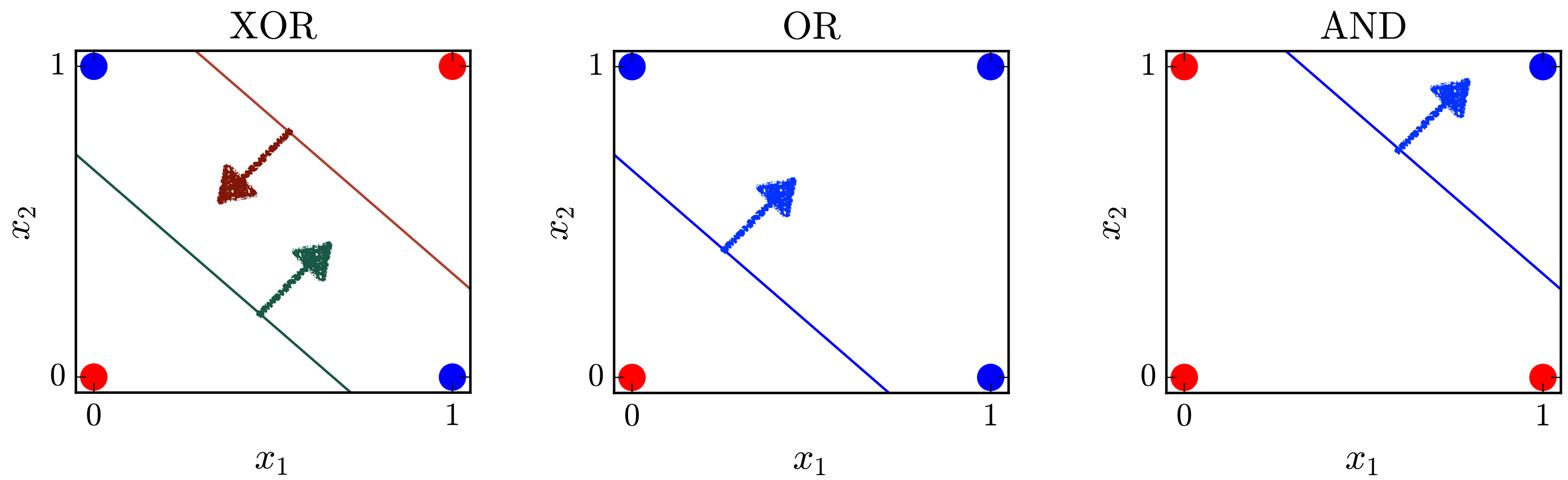


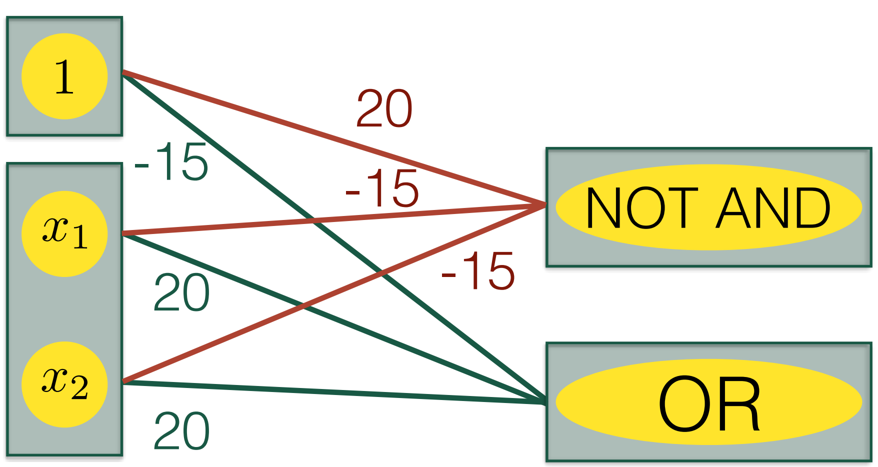


Neural Networks

XOR

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0

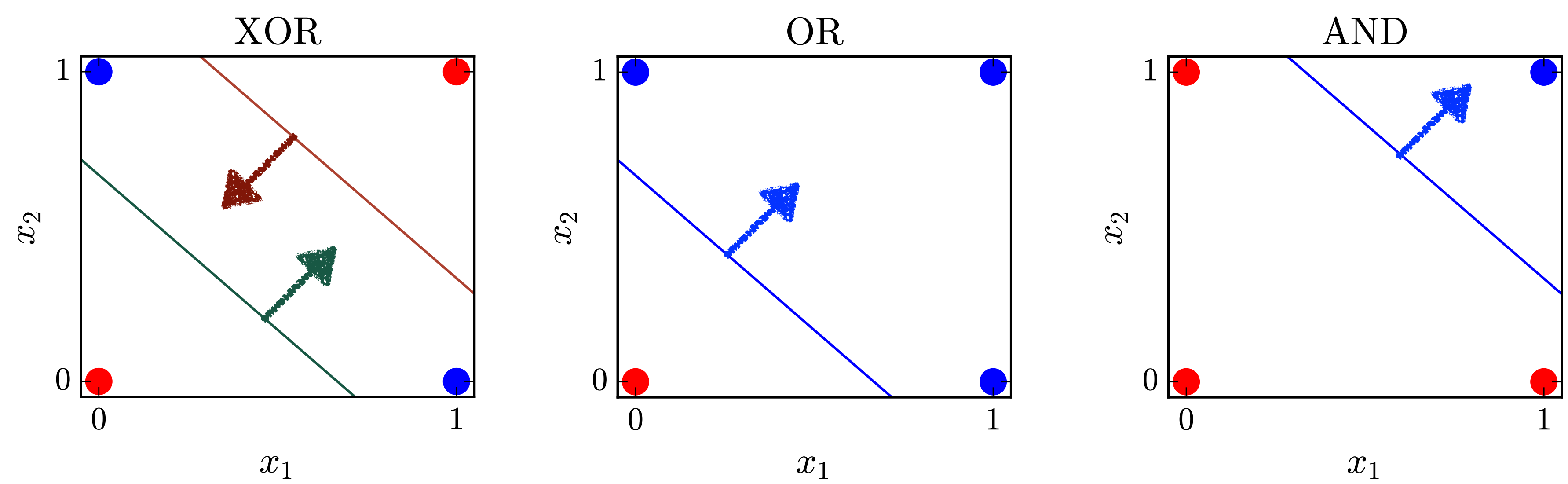


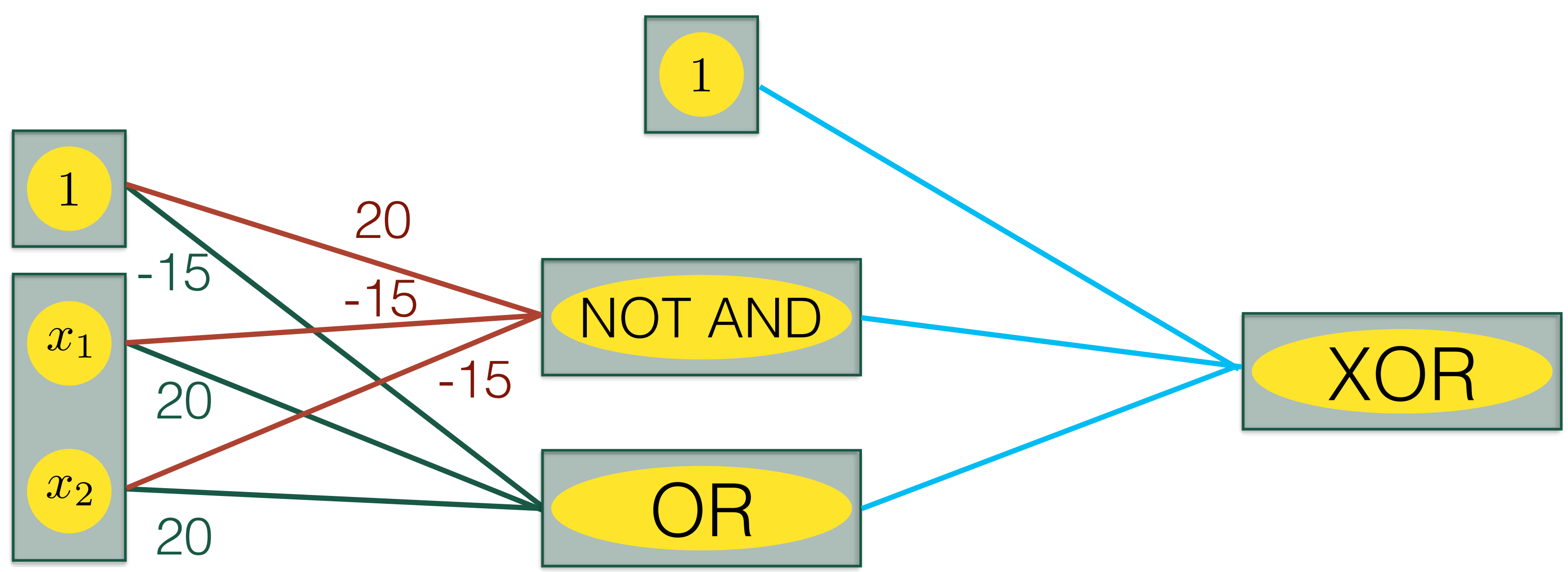


Neural Networks

XOR

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0

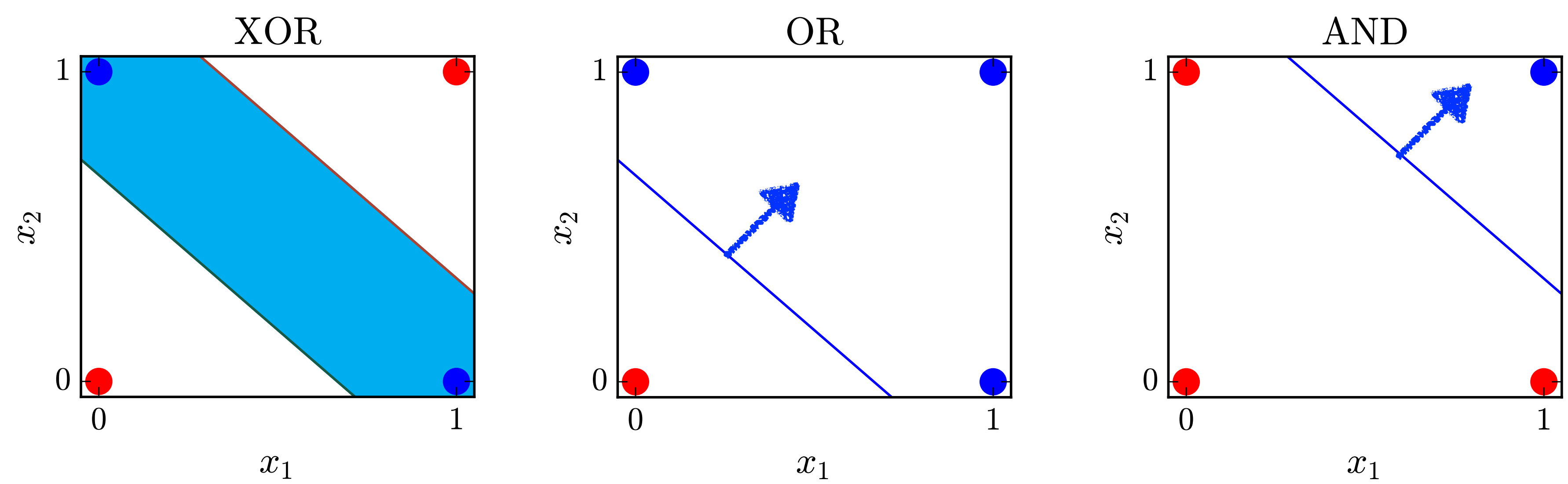


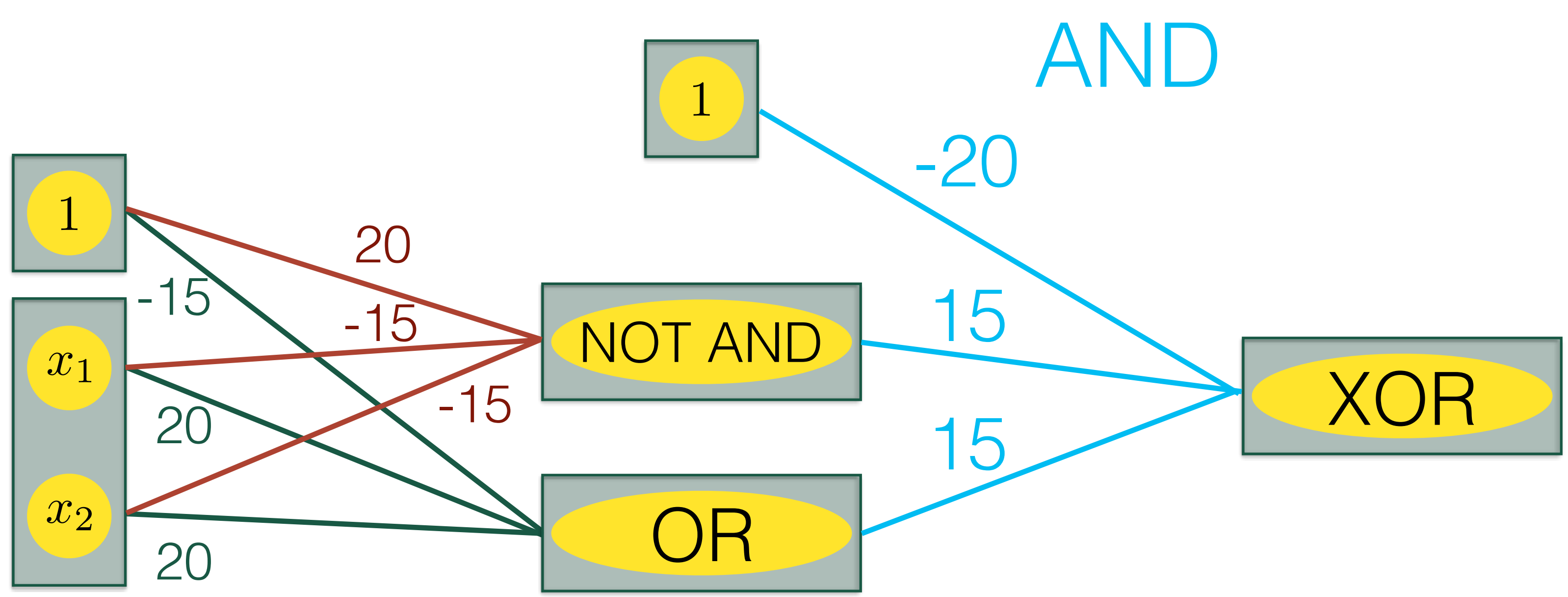


Neural Networks

XOR

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0

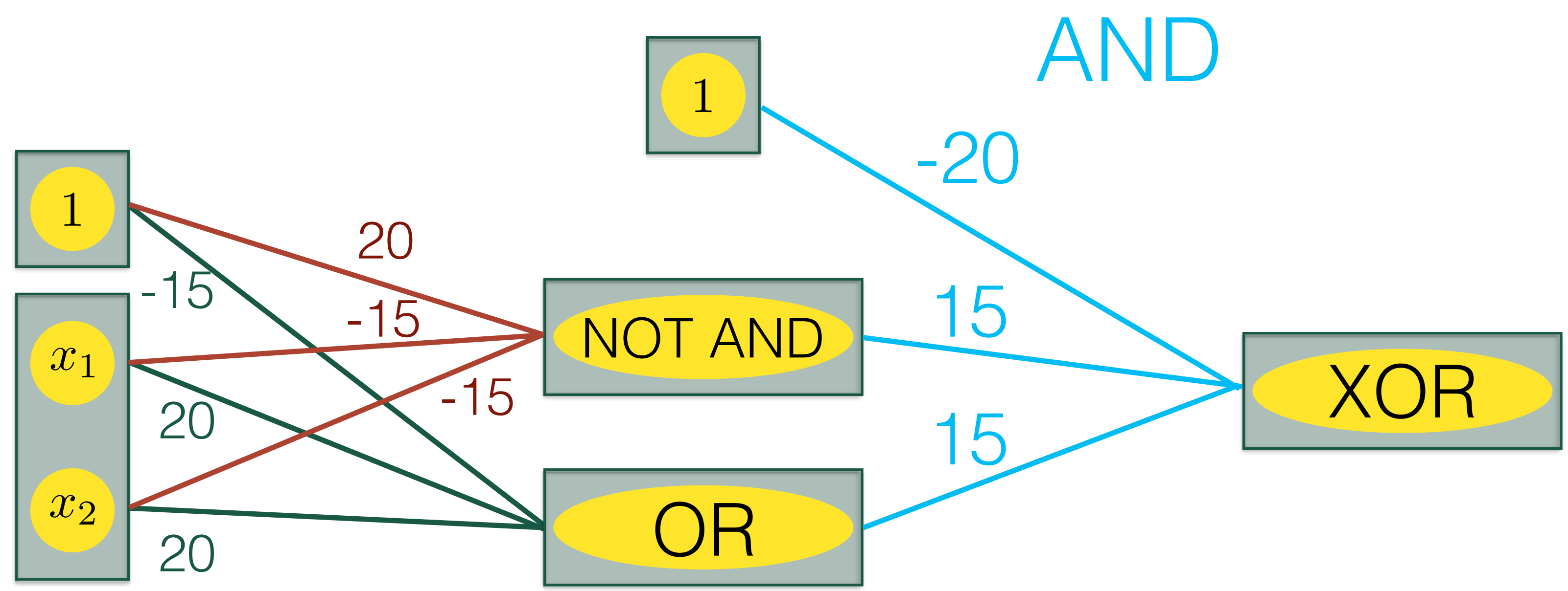




Neural Networks

XOR

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0

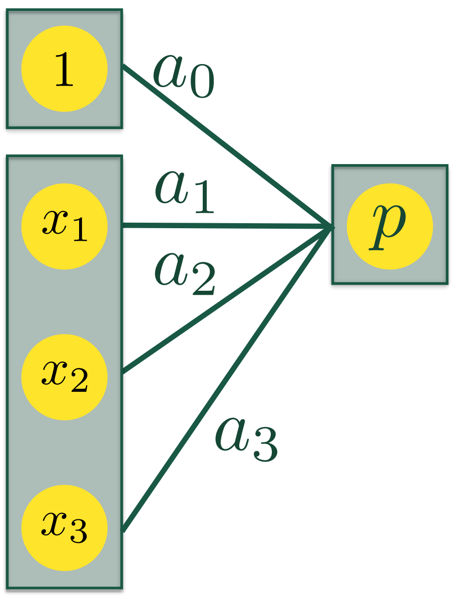


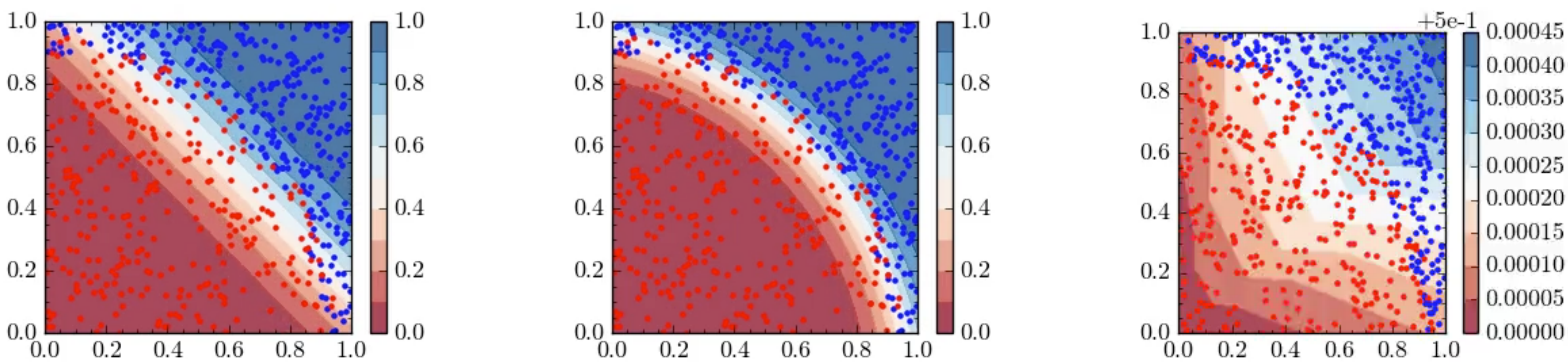
Simple example showing that neural network can access 'high-level' functions

To learn weights, need large training set and CPU time

Neural Networks

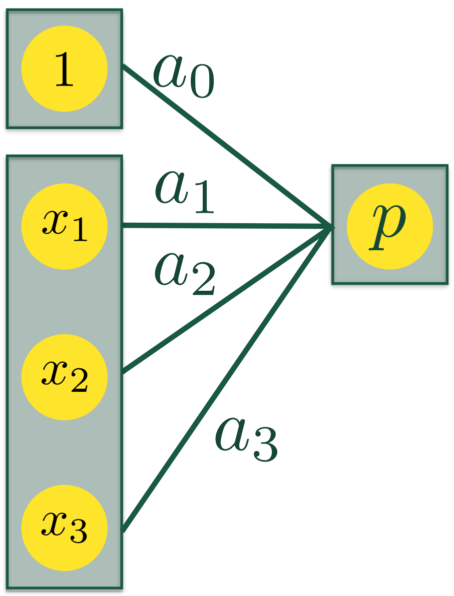
- Don't add more inputs, let machine find own shape
- Ability to learn 'any' function
- More nodes/hidden layers allows for more complex features

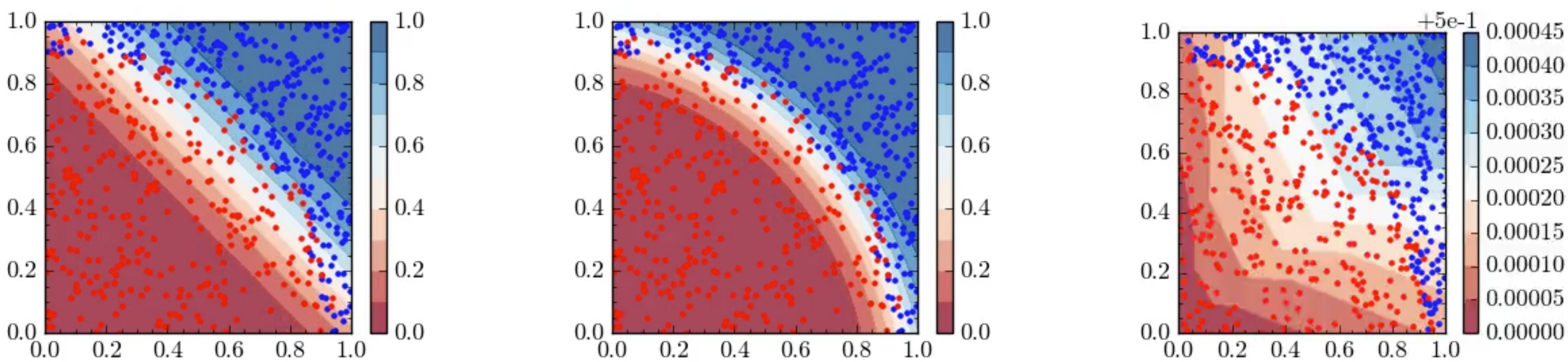




Neural Networks

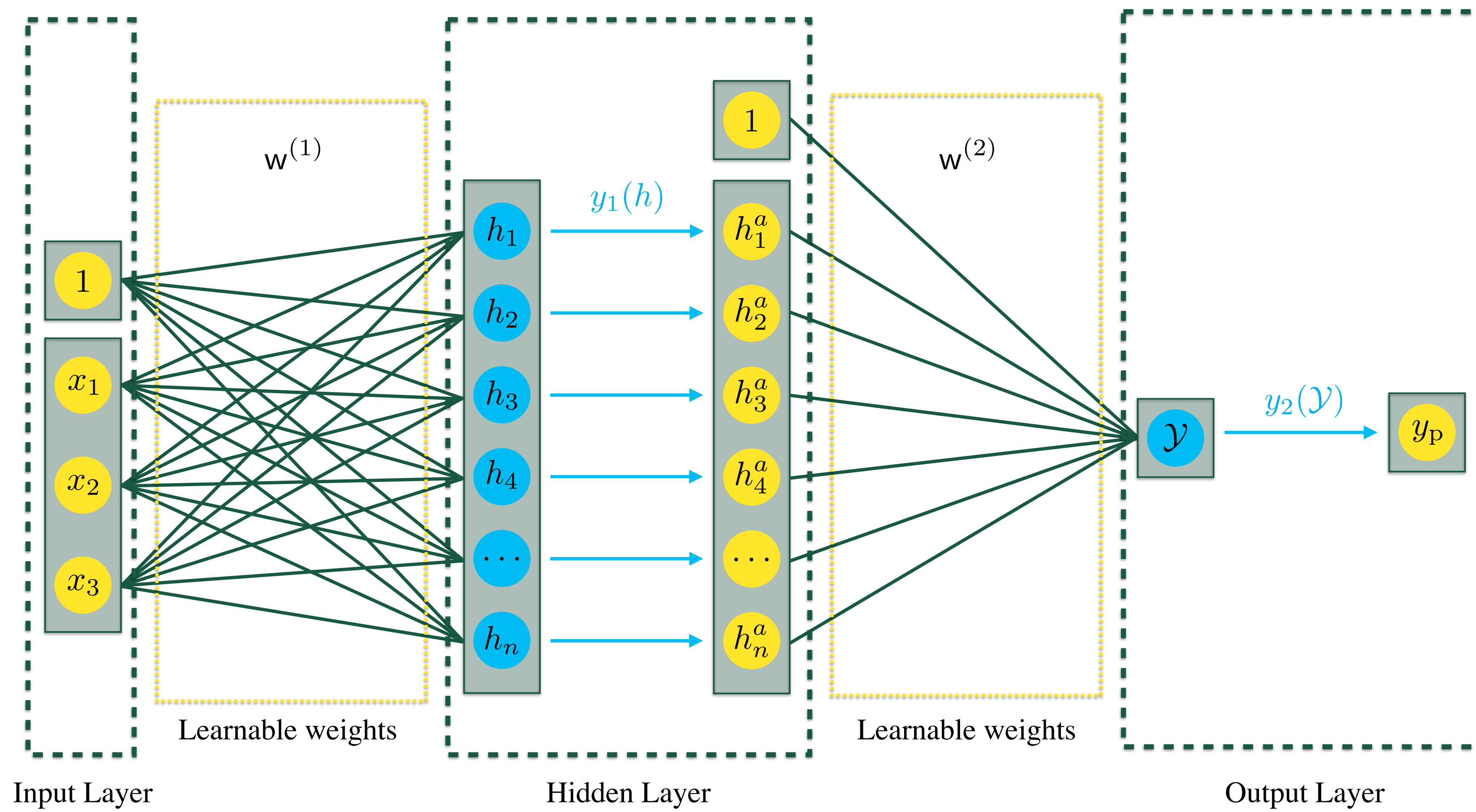
- Don't add more inputs, let machine find own shape
- Ability to learn 'any' function
- More nodes/hidden layers allows for more complex features

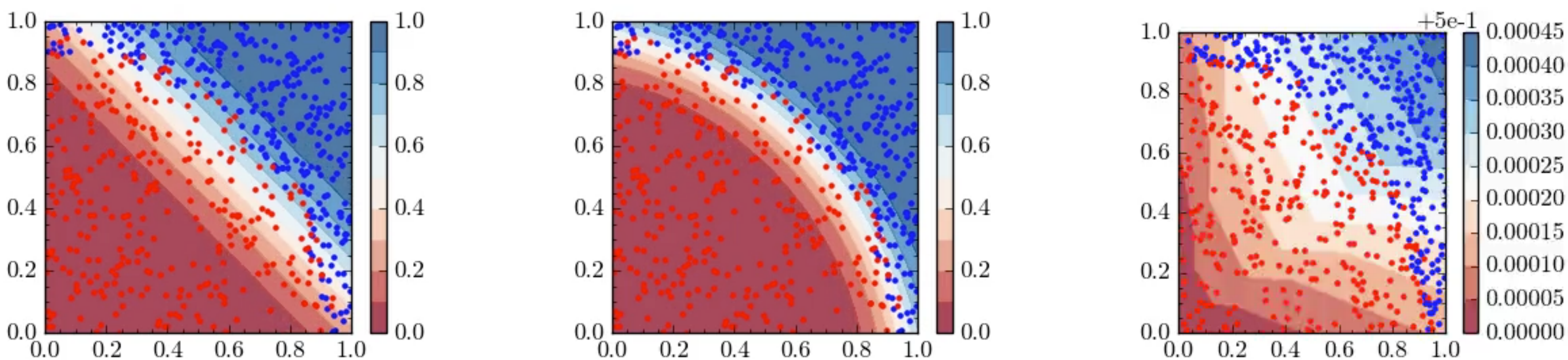




Neural Networks

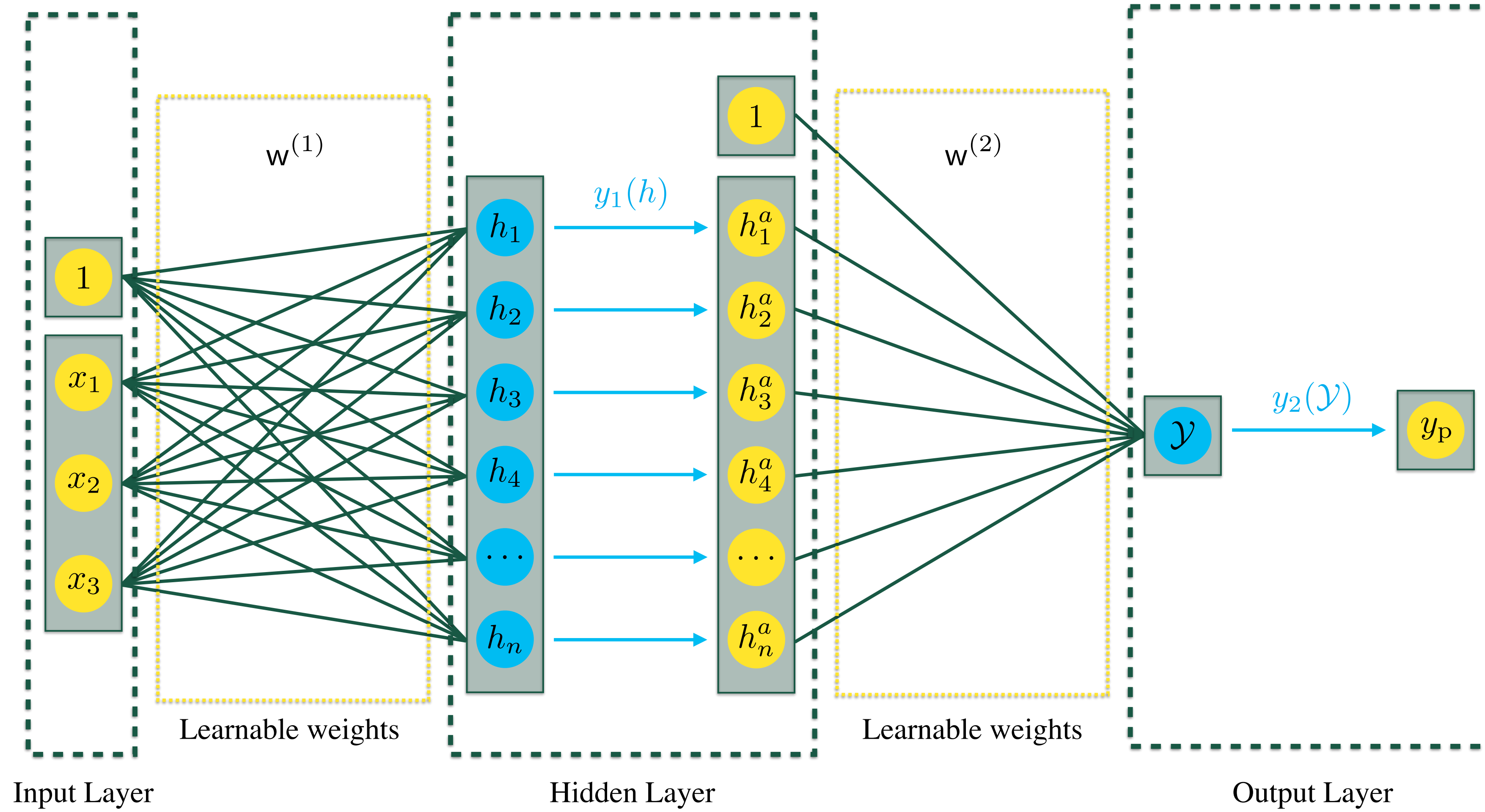
- Don't add more inputs, let machine find own shape
- Ability to learn 'any' function
- More nodes/hidden layers allows for more complex features

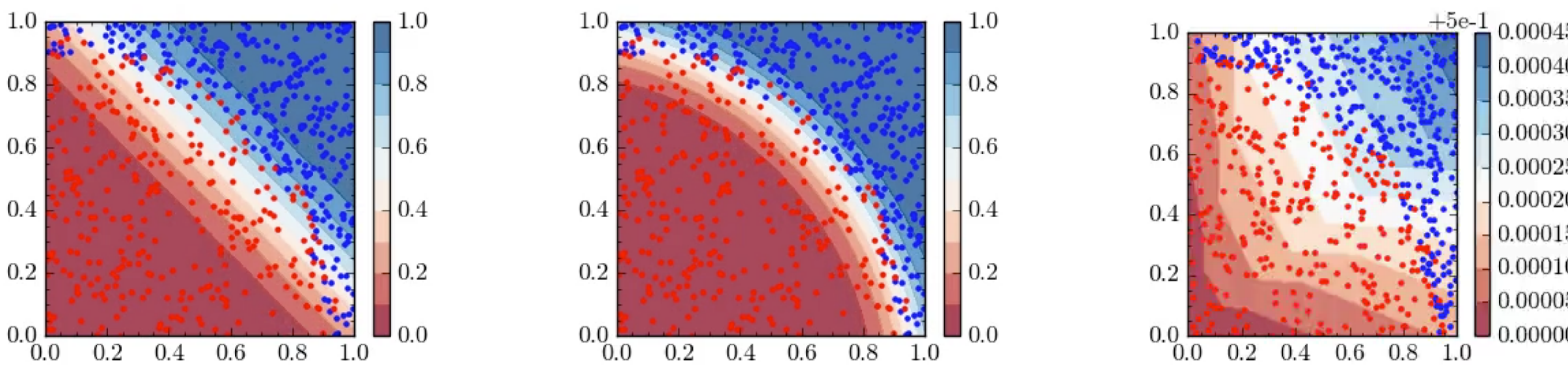




Neural Networks

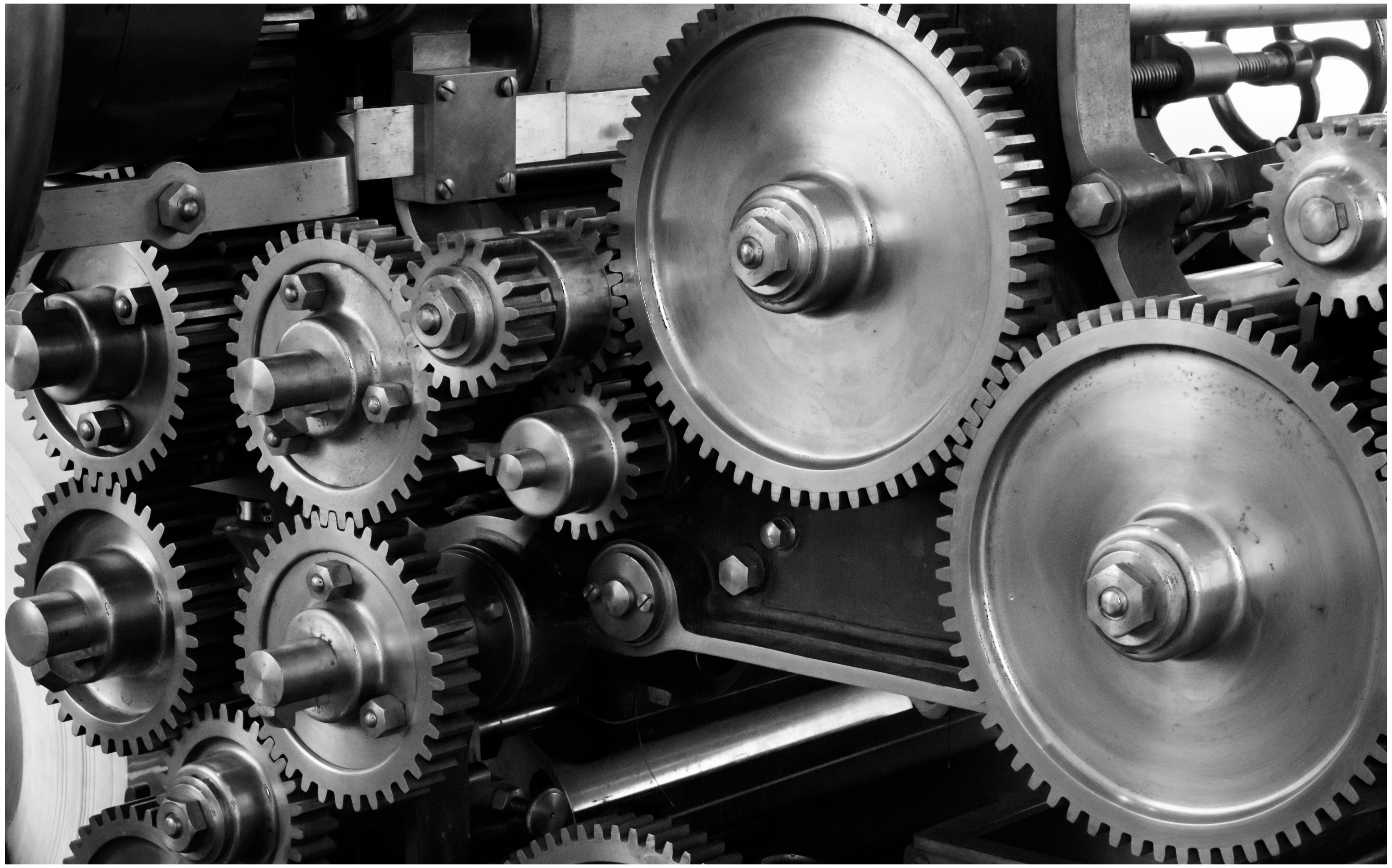
- Don't add more inputs, let machine find own shape
- Ability to learn 'any' function
- More nodes/hidden layers allows for more complex features





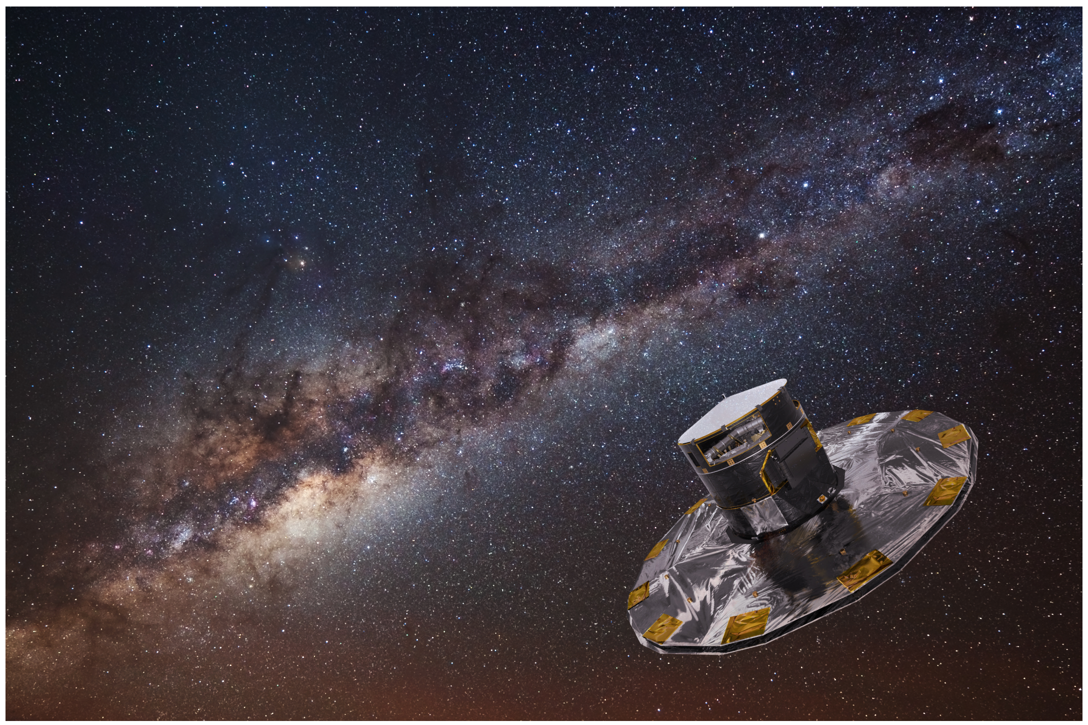
Neural Network Review

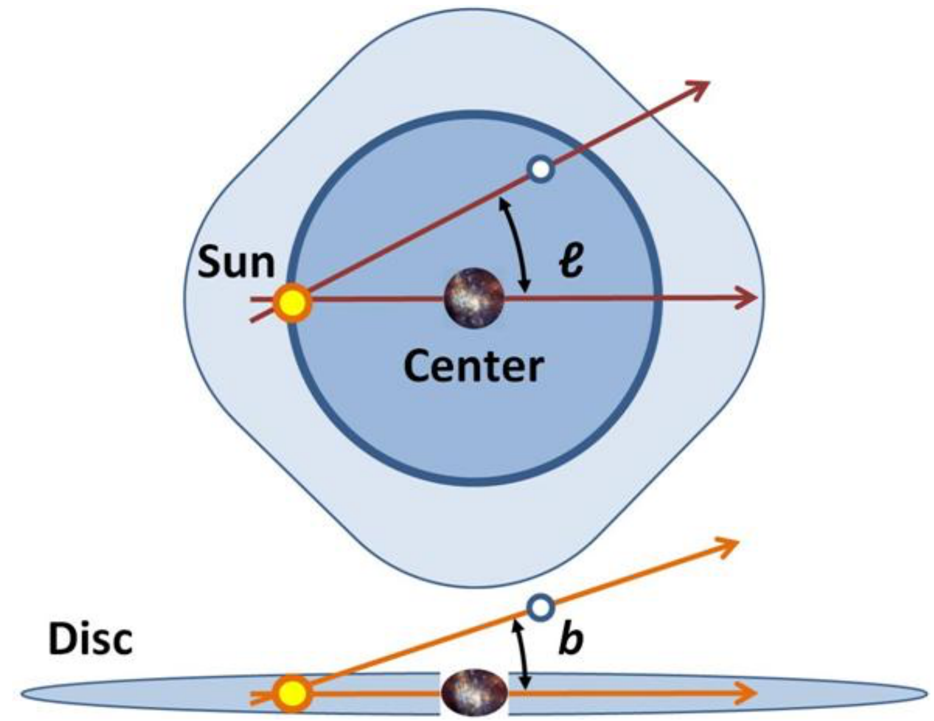
- Neural networks act as universal function fitter
- Deep networks (many hidden layers) allow the network to pick its own features



Is it possible to classify halo stars without spectroscopy?

Is it possible to classify halo stars using only 5-d information?





Stellar information from Gaia:

- Galactic longitude (*l*)
- Galactic latitude (b)
- Proper motion (ascension)
- Proper motion (declination)
- Parallax (distance = 1 / parallax)

rate of change of these, transferred to different coordinate system

Is it possible to classify halo stars using only 5-d information?

How to train the network if we don't know labels for the stars?

Sampling

- Draw stars from model distributions
- Defined labels
- Fast data generation

Simulation

- Distributions from interaction
- Labels from merger history
- Can't generate ourselves

Learning the halo

Sampling

THE ASTROPHYSICAL JOURNAL, 730:3 (20pp), 2011 March 20 doi:10.1088/0004-637X/730/1/3

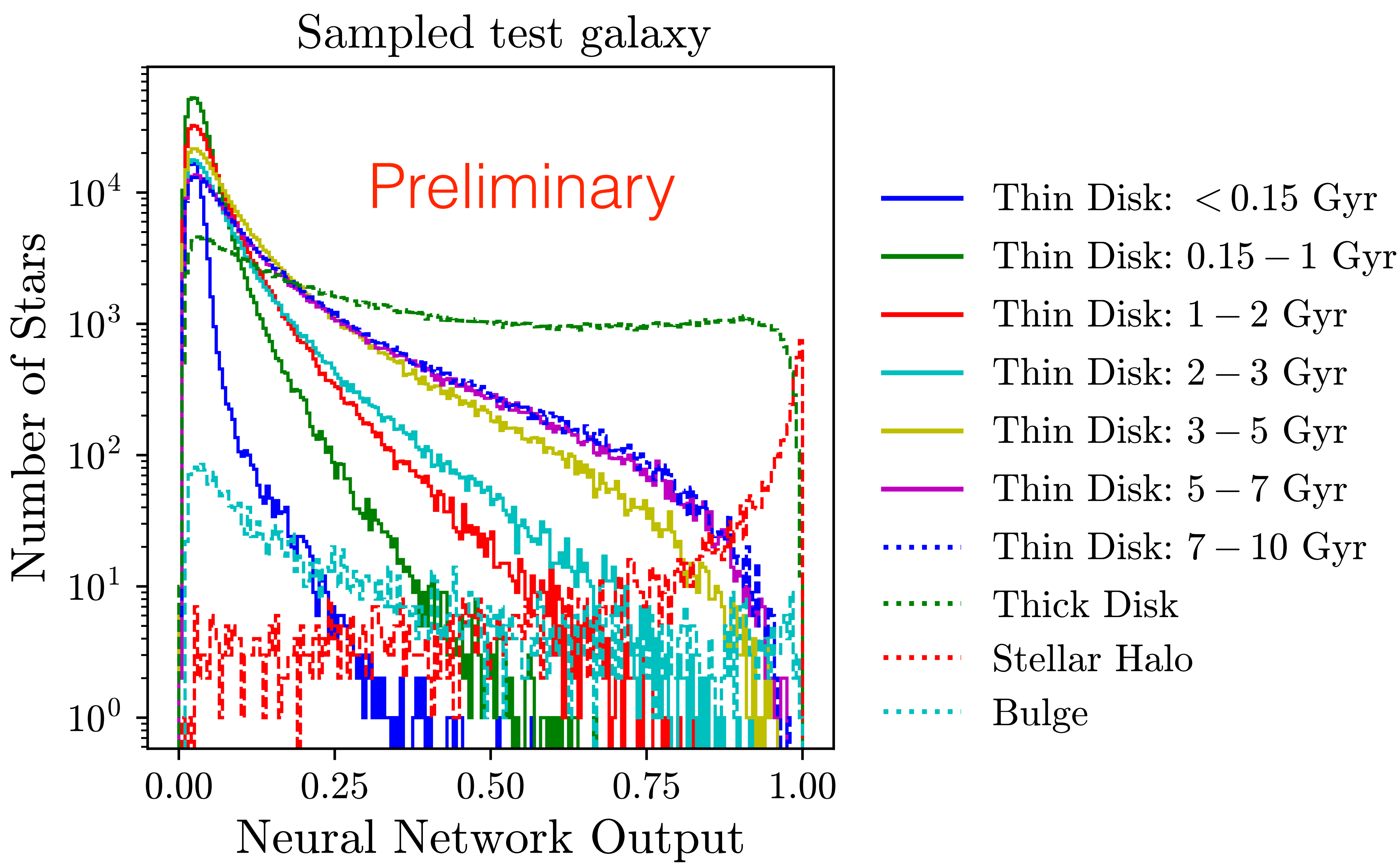
GALAXIA: A CODE TO GENERATE A SYNTHETIC SURVEY OF THE MILKY WAY

SANJIB SHARMA[1], JOSS BLAND-HAWTHORN[1,4], KATHRYN V. JOHNSTON[2], AND JAMES BINNEY[3]
[1] Sydney Institute for Astronomy, School of Physics, University of Sydney, NSW 2006, Australia
[2] Department of Astronomy, Columbia University, New York, NY 10027, USA
[3] Rudolf Peierls Centre for Theoretical Physics, 1 Keble Rd, Oxford OX1 3NP, UK
Received 2010 September 16; accepted 2011 January 12; published 2011 February 23

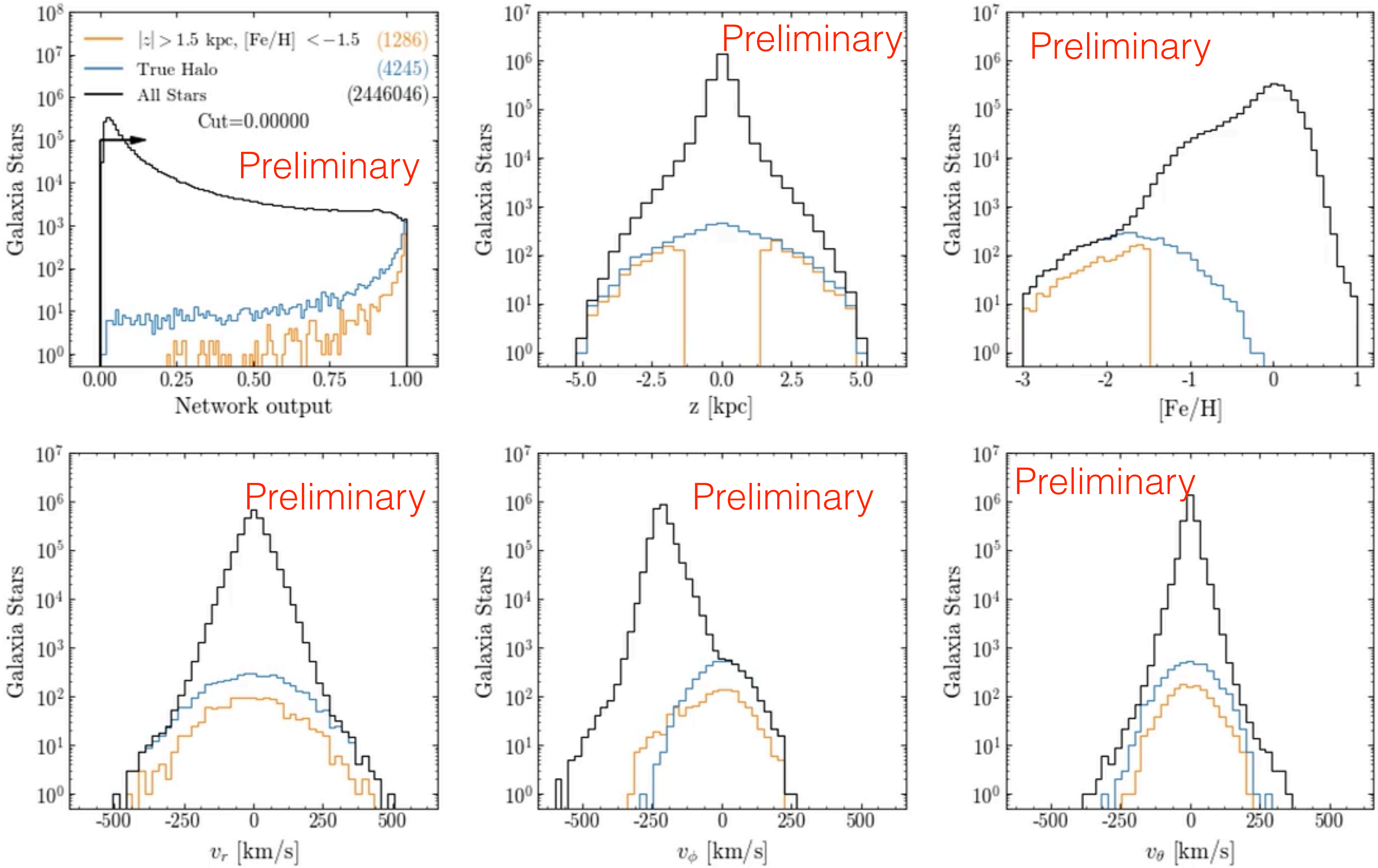
ABSTRACT

We present here a fast code for creating a synthetic survey of the Milky Way. Given one or more color–magnitude bounds, a survey size, and geometry, the code returns a catalog of stars in accordance with a given model of the Milky Way. The model can be specified by a set of density distributions or as an *N*-body realization. We provide fast and efficient algorithms for sampling both types of models. As compared to earlier sampling schemes which generate stars at specified locations along a line of sight, our scheme can generate a continuous and smooth distribution of stars over any given volume. The code is quite general and flexible and can accept input in the form of a star formation rate, age–metallicity relation, age–velocity-dispersion relation, and analytic density distribution functions. Theoretical isochrones are then used to generate a catalog of stars, and support is available for a wide range of photometric bands. As a concrete example, we implement the Besançon Milky Way model for the disk. For the stellar halo we employ the simulated stellar halo *N*-body models of Bullock & Johnston. In order to sample *N*-body models, we present a scheme that disperses the stars spawned by an *N*-body particle, in such a way that the phase-space density of the spawned stars is consistent with that of the *N*-body particles. The code is ideally suited to generating synthetic data sets that mimic near future wide area surveys such as *GAIA*, LSST, and HERMES. As an application we study the prospect of identifying structures in the stellar halo with a simulated *GAIA* survey. We plan to make the code publicly available.

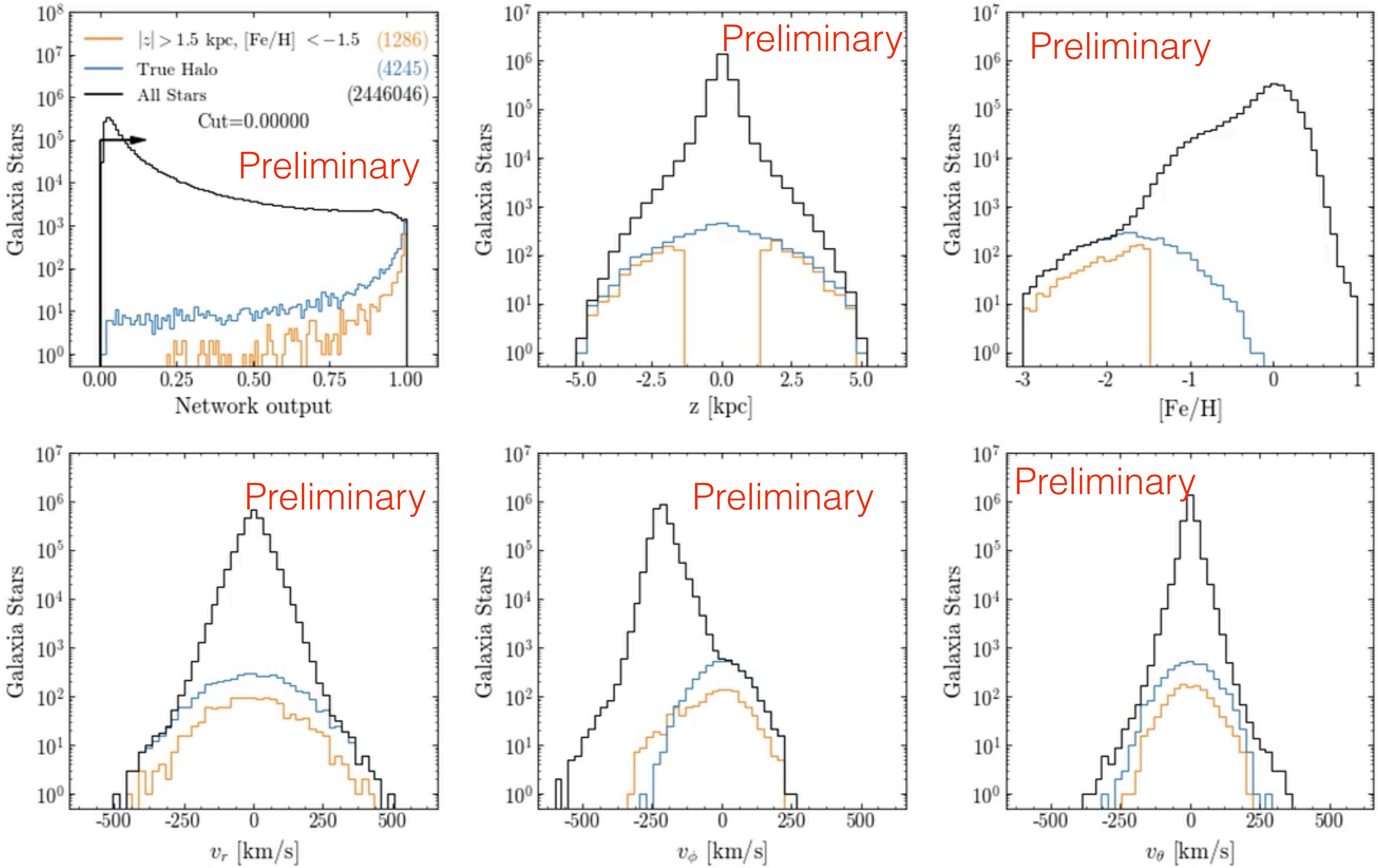
Learning the halo



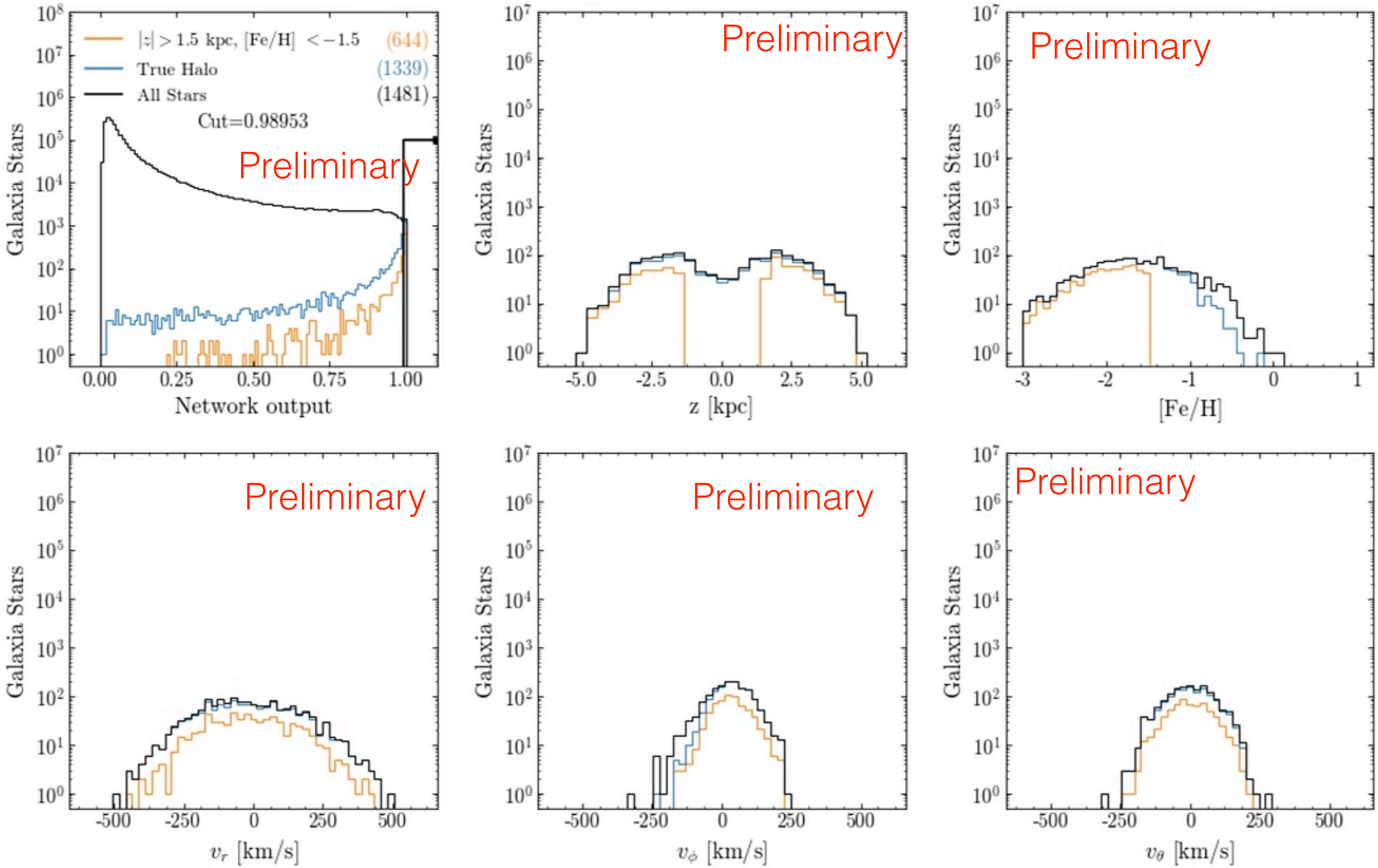
Learning the halo



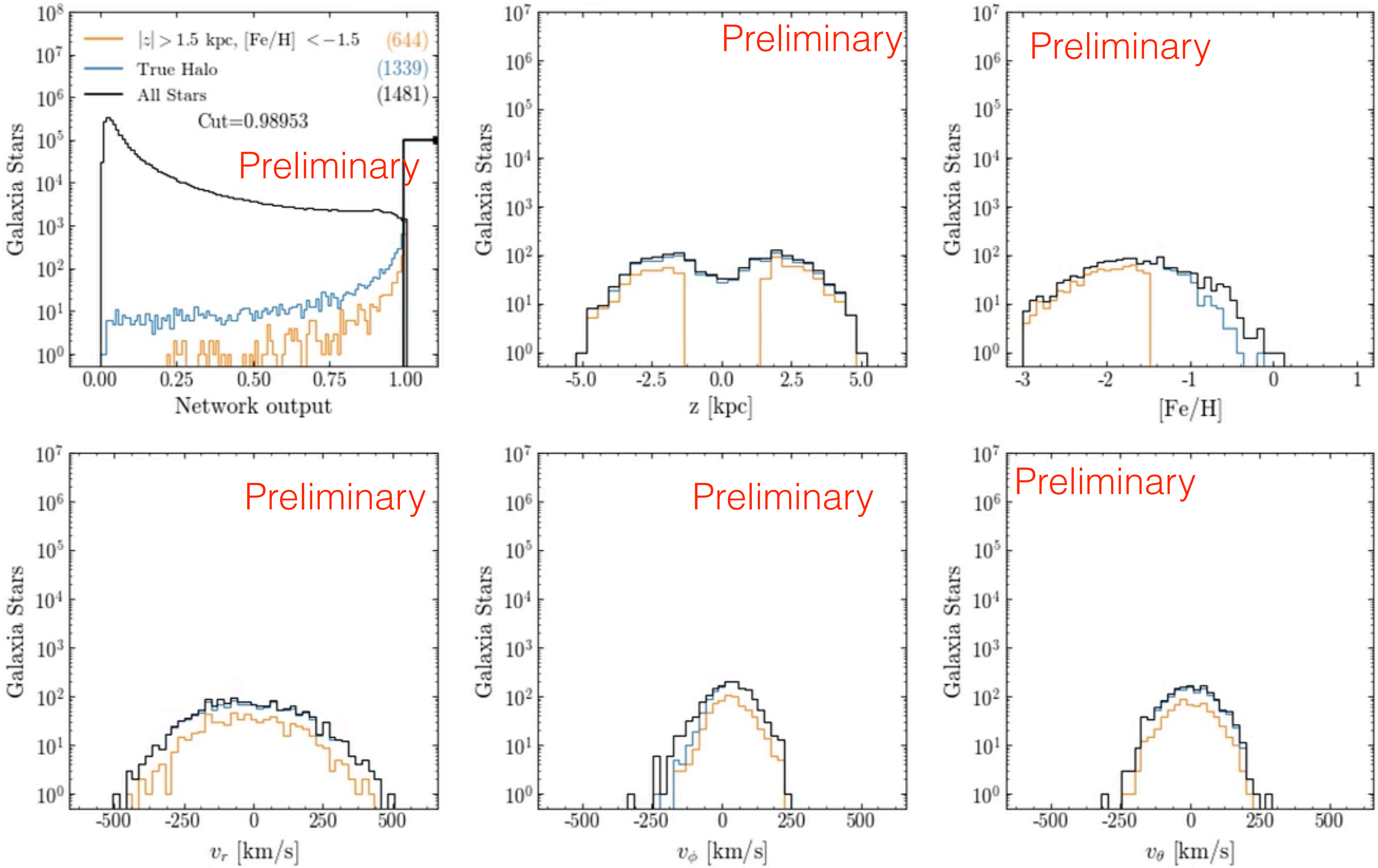
Learning the halo



Learning the halo



Learning the halo



Compare with other methods

ZM: $|z| > 1.5$ kpc an [Fe/H] < -1.5

Kinematic selection defines any star which has $|\mathbf{v} - \mathbf{v}_{LSR}| > \mathbf{v}_{LSR}$ as halo, where $\mathbf{v} = (v_x, v_y, v_z)$ and $\mathbf{v}_{LSR} = (0, 232, 0)$ km/s.

Metallicity selection use gaussian mixture model on 3D velocities. One group should have a peak consistent with the disk (either in v_y or v_ϕ). The halo stars are then defined as the stars which have [Fe/H]< -1 and are not part of the group with velocities consistent with the disk.

	Halo	Non-halo	FPR	TPR	Purity
Galaxia test set	4245	2441801	-	-	-
$NN > 0.98929$	1359	151	$6.26{\times}10^{-5}$	0.320	90%
ZM	1093	193	$7.90{\times}10^{-5}$	0.257	85%
Kinmatic	3139	1763	$7.22{\times}10^{-4}$	0.739	64%
Metallicity	3880	44404	0.0182	0.914	8.0%

Learning the halo

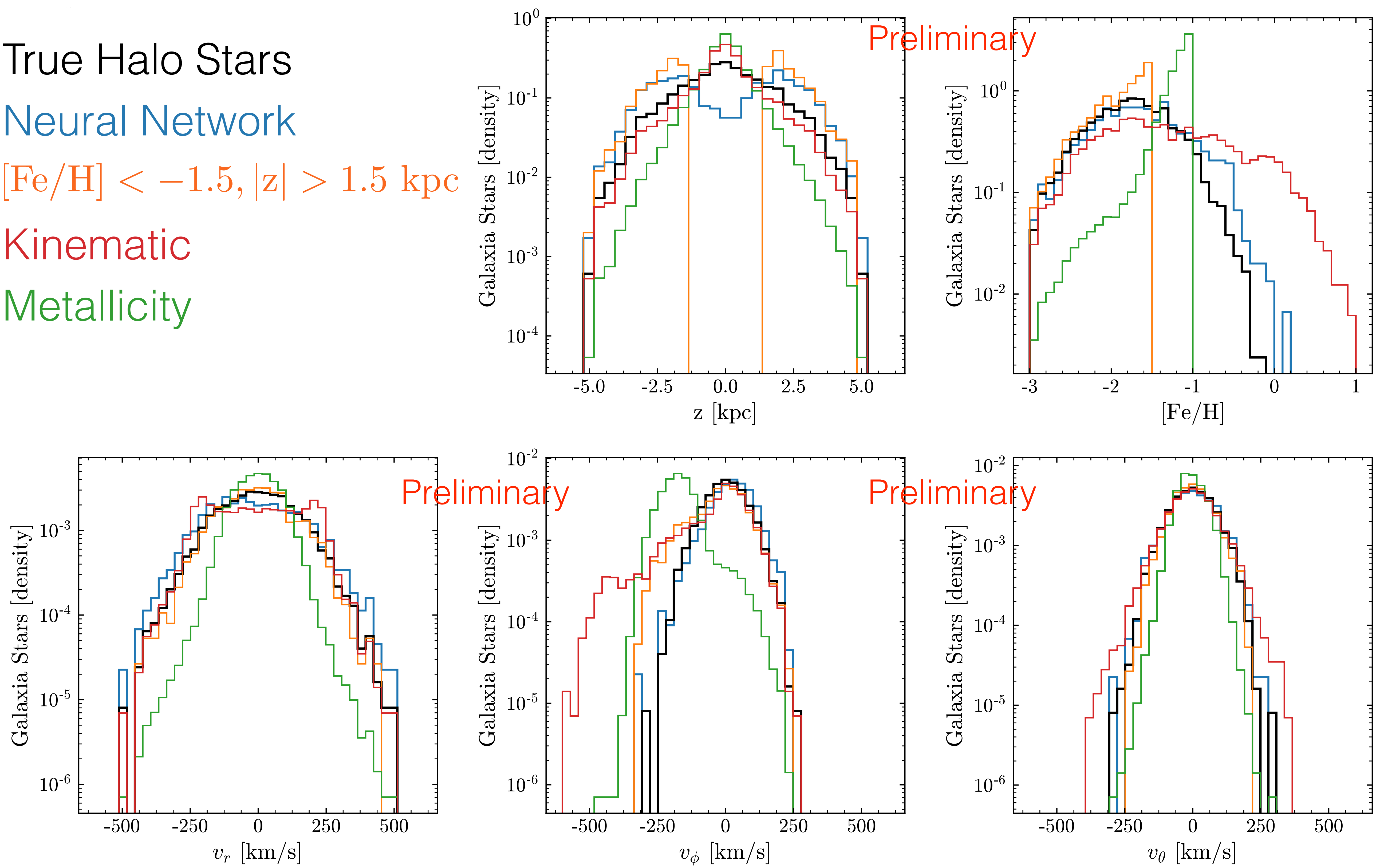
True Halo Stars

Neural Network

$[Fe/H] < -1.5, |z| > 1.5$ kpc

Kinematic

Metallicity



Learning the halo

Is it possible to classify halo stars using only 5-d information?

Sampling

- Draw stars from model distributions
- Defined labels
- Fast data generation

- Classification is possible!
- Can perform better than traditional methods
- High purity still preserves underlying distributions

Learning the halo

Is it possible to classify halo stars using only 5-d information?

Sampling

- Draw stars from model distributions
- Defined labels
- Fast data generation

Simulation

- Distributions from interaction
- Labels from merger history
- Can't generate ourselves

Learning the halo

Simulation

arXiv.org > astro-ph > arXiv:1806.10564

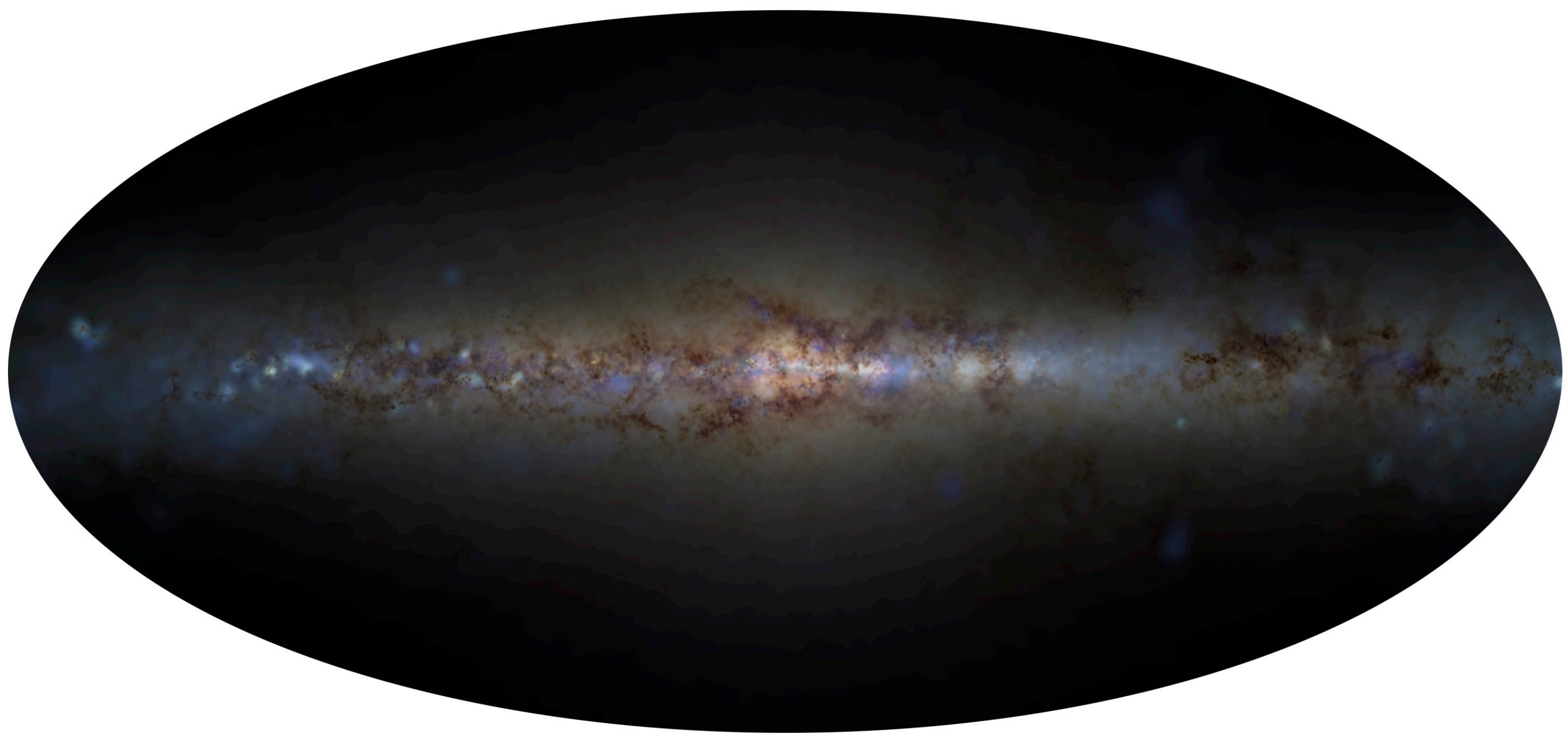
Search or Article

(Help | Advanced sear

Astrophysics > Astrophysics of Galaxies

Synthetic Gaia surveys from the FIRE cosmological simulations of Milky-Way-mass galaxies

Robyn E. Sanderson (1), Andrew Wetzel (2), Sarah Loebman (2), Sanjib Sharma (3), Philip F. Hopkins (1), Shea Garrison-Kimmel (1), Claude-André Faucher-Giguère (4), Dušan Kereš (5), Eliot Quataert (6) ((1) California Institute of Technology, (2) University of California at Davis, (3) University of Sydney, (4) Northwestern University, (5) University of California at San Diego, (6), University of California Berkeley)

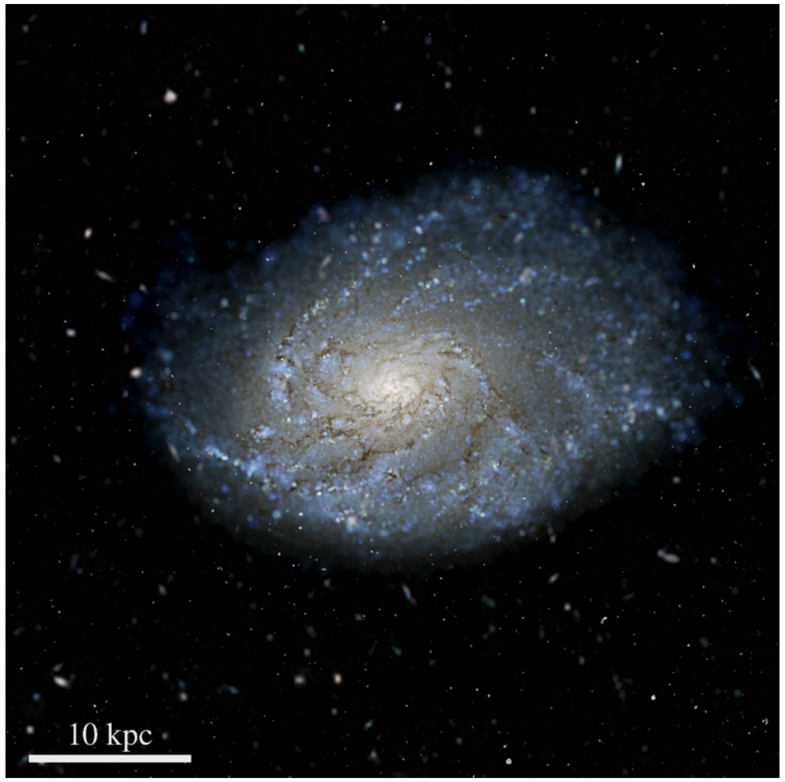


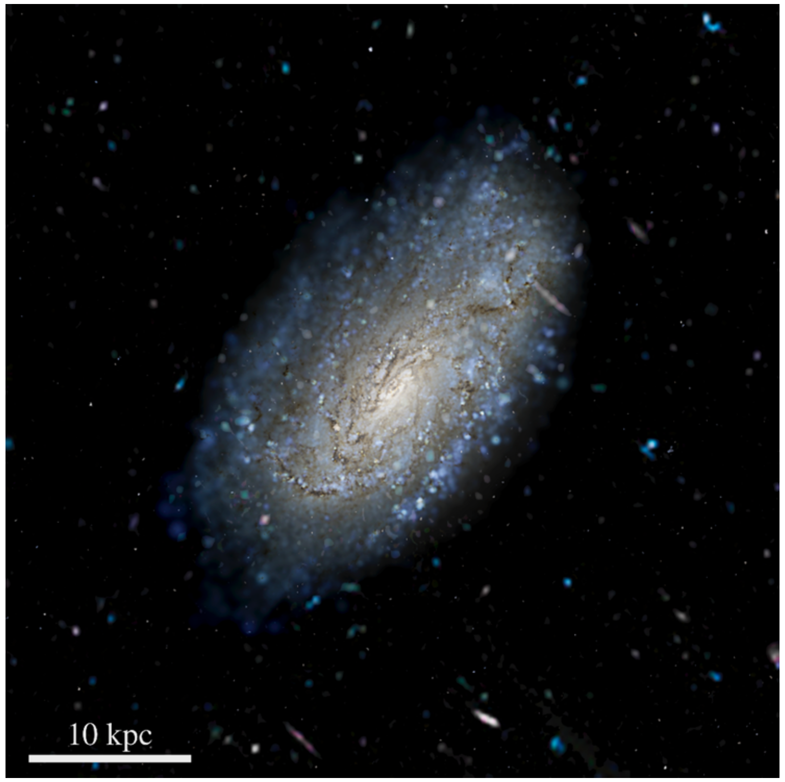
The Latte suite of FIRE-2 cosmological zoom-in baryonic simulations of Milky Way-mass galaxies (Wetzel et al 2016), part of the Feedback In Realistic Environments (FIRE) simulation project, were run using the Gizmo gravity plus hydrodynamics code in meshless finite-mass (MFM) mode (Hopkins 2015) and the FIRE-2 physics model (Hopkins et al 2018).

Synthetic Gaia DR2-like surveys of the Latte suite of FIRE-2 simulations were created via the Ananke framework (Sanderson et al 2018).

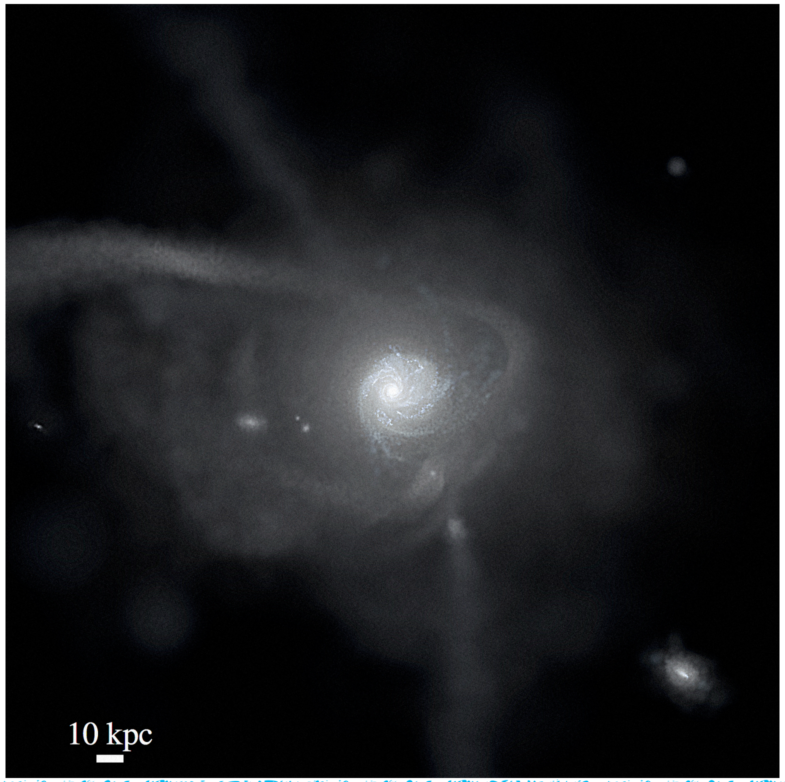
Learning the halo

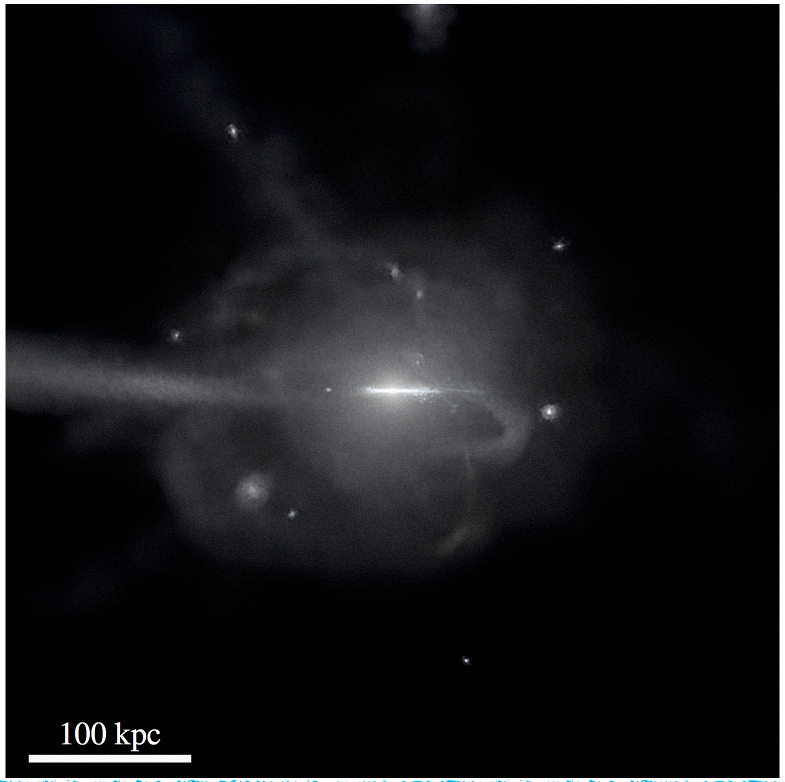
Simulation





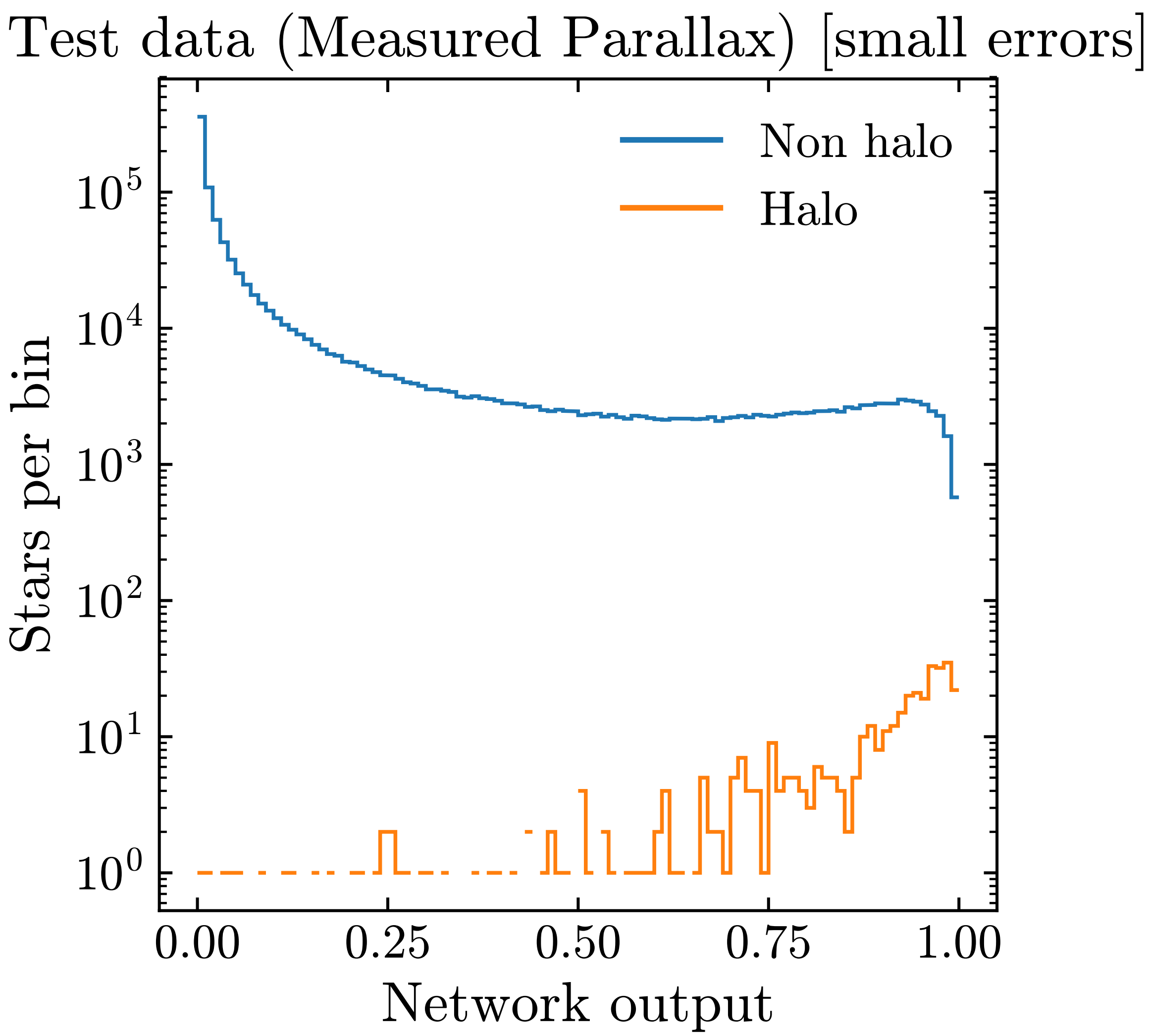
- Not smooth distributions
- Very large dataset
- Expect this to be more challenging
- How do deal with "measurement" errors

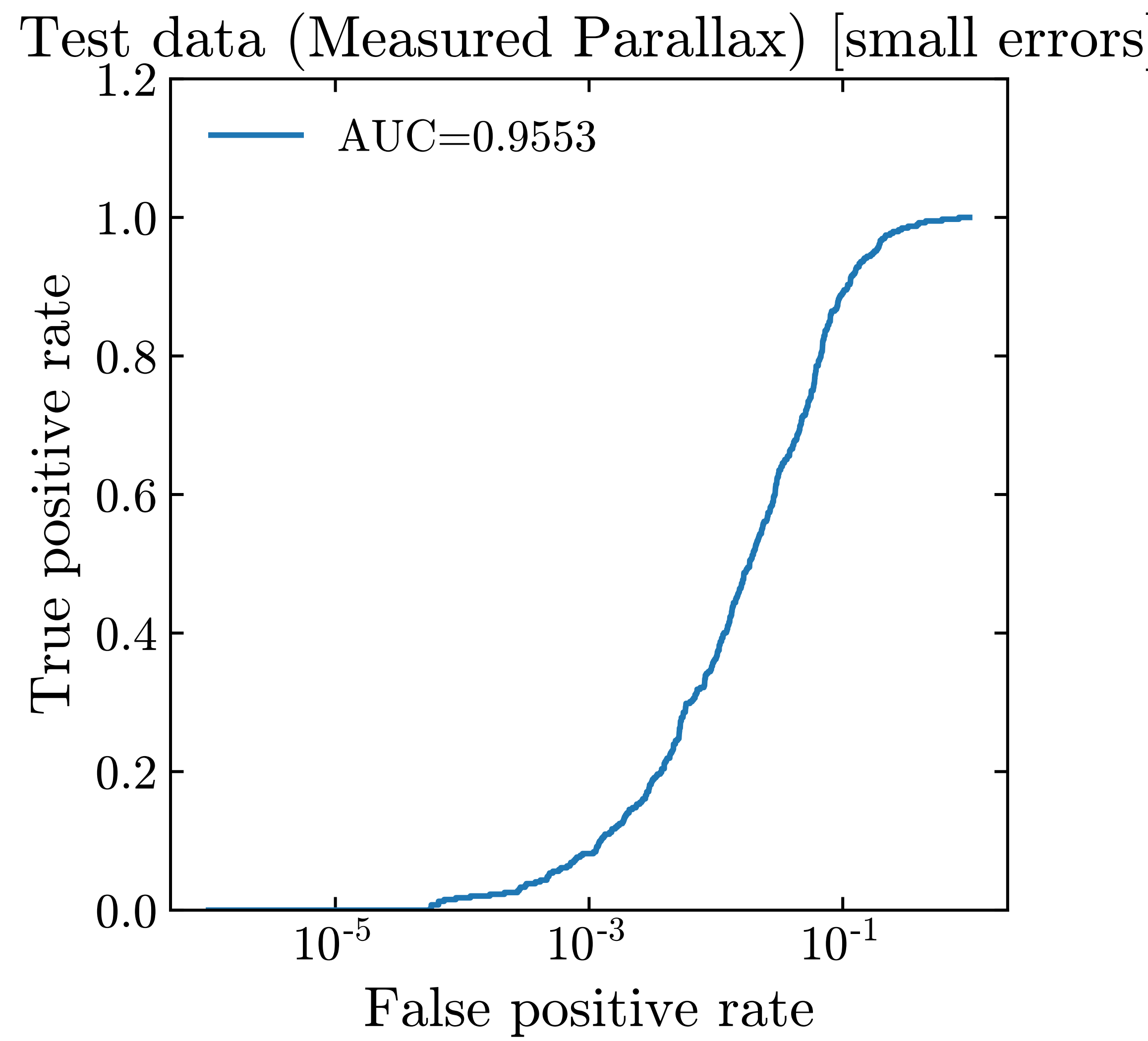




Learning the halo

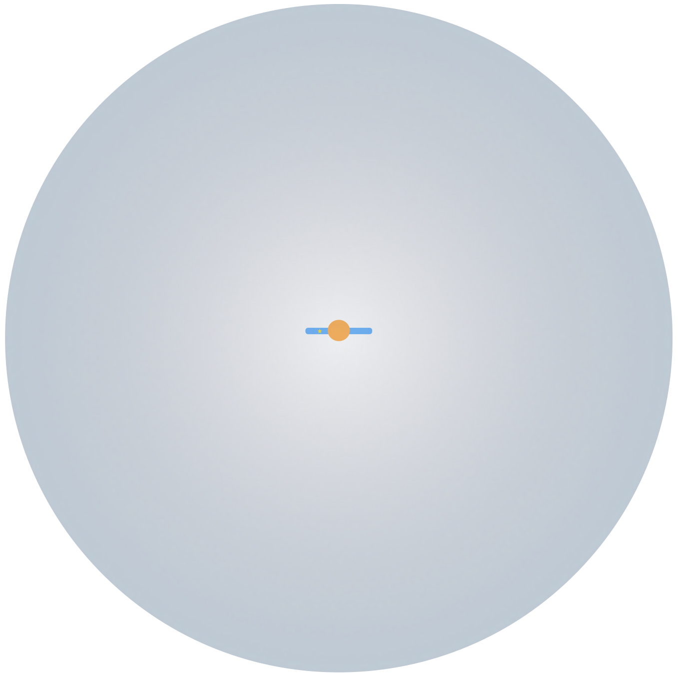
Simulation





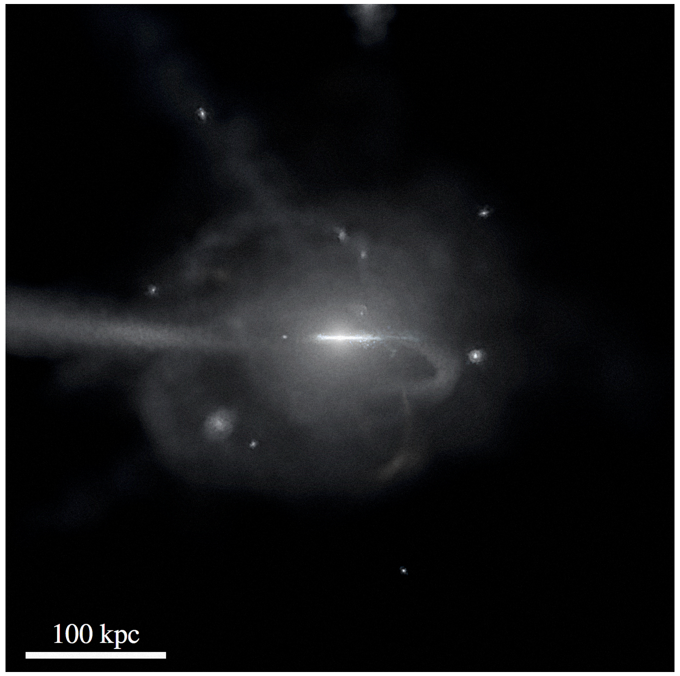
Looks promising, still have issues to deal with

Conclusion

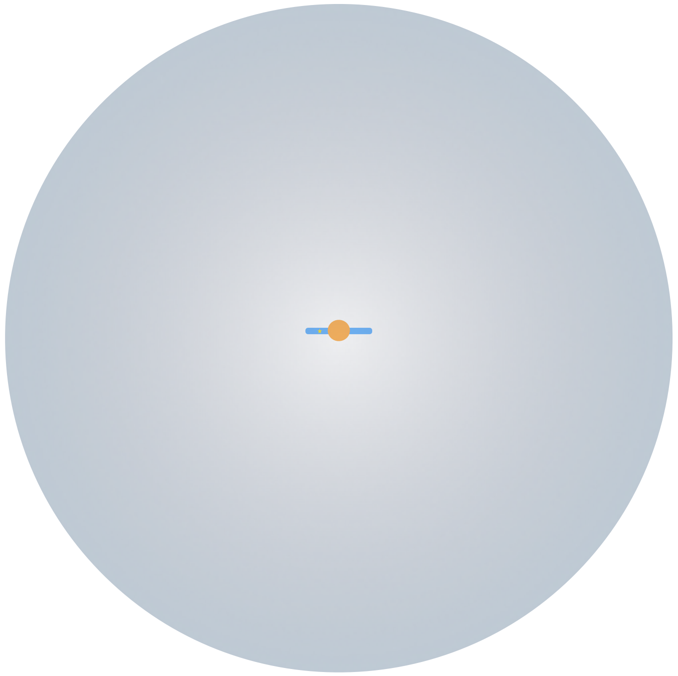


Hierarchical
Merger Model

DM tracers

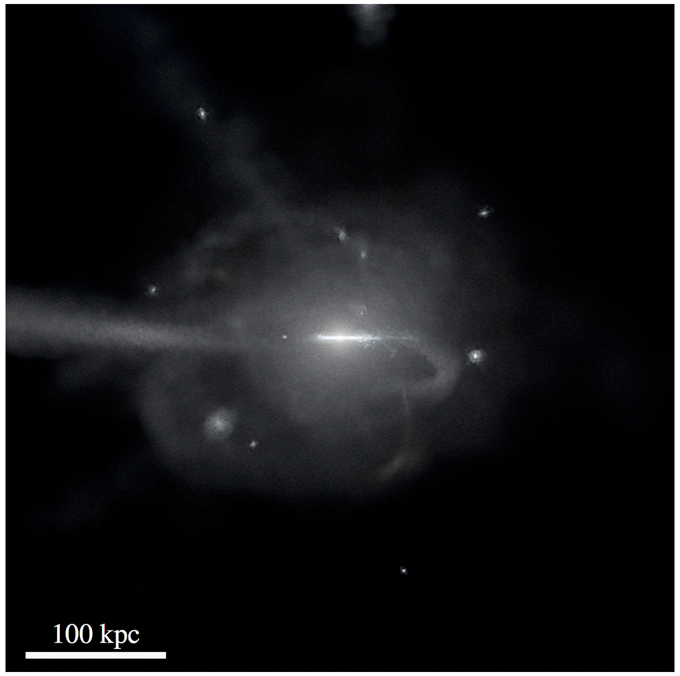


Conclusion

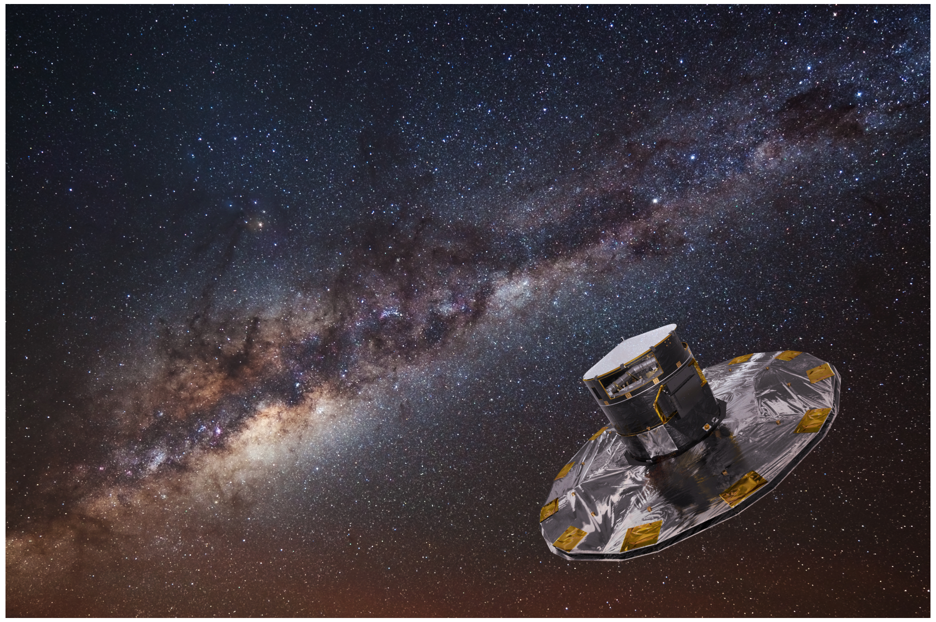


Hierarchical Merger Model

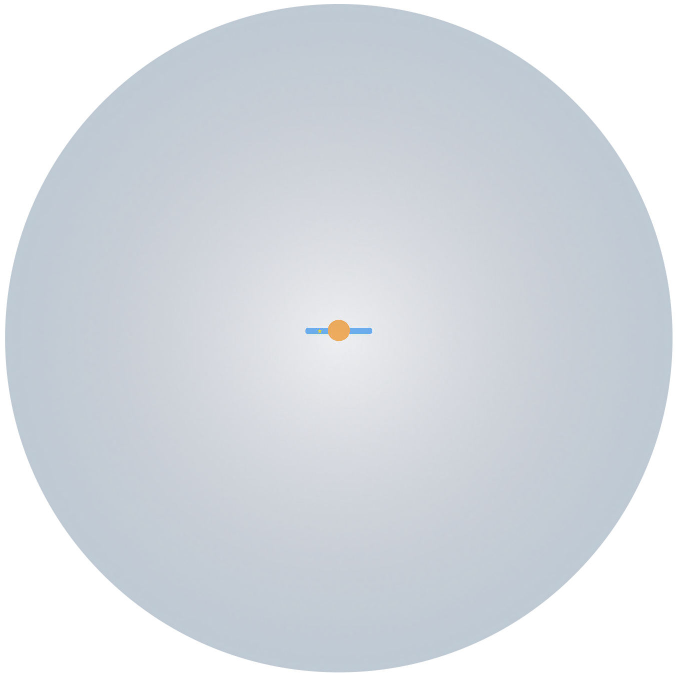
DM tracers



Gaia+Spectroscopic

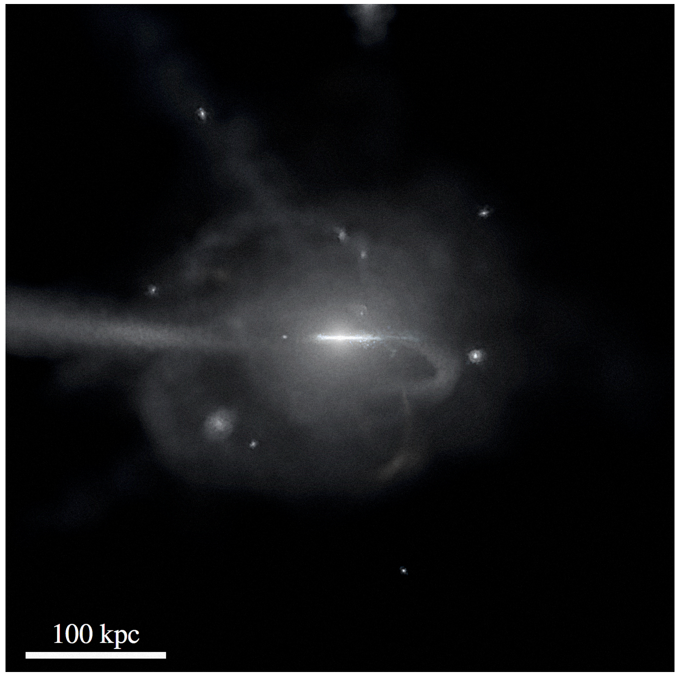


Conclusion

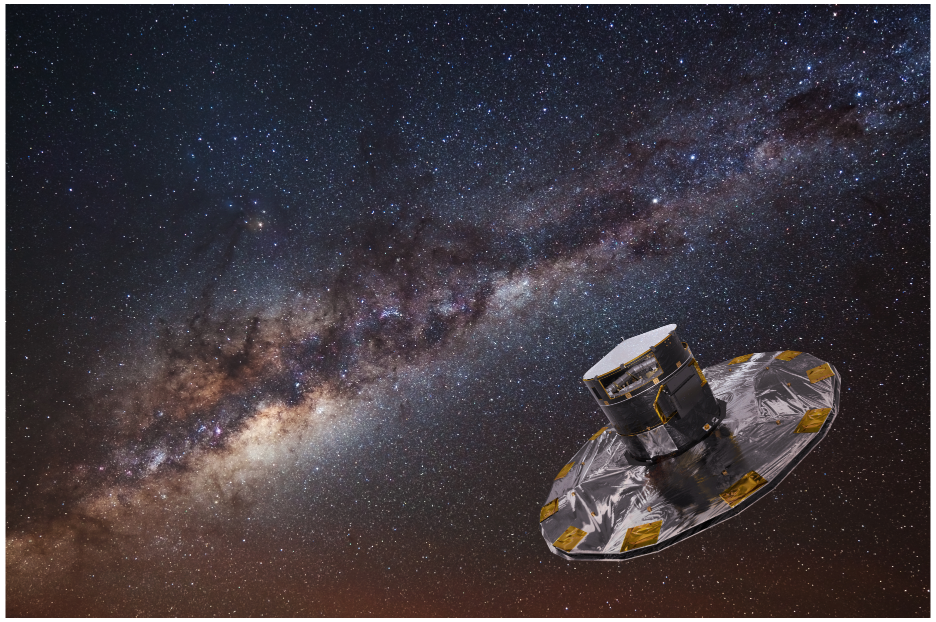


Hierarchical
Merger Model

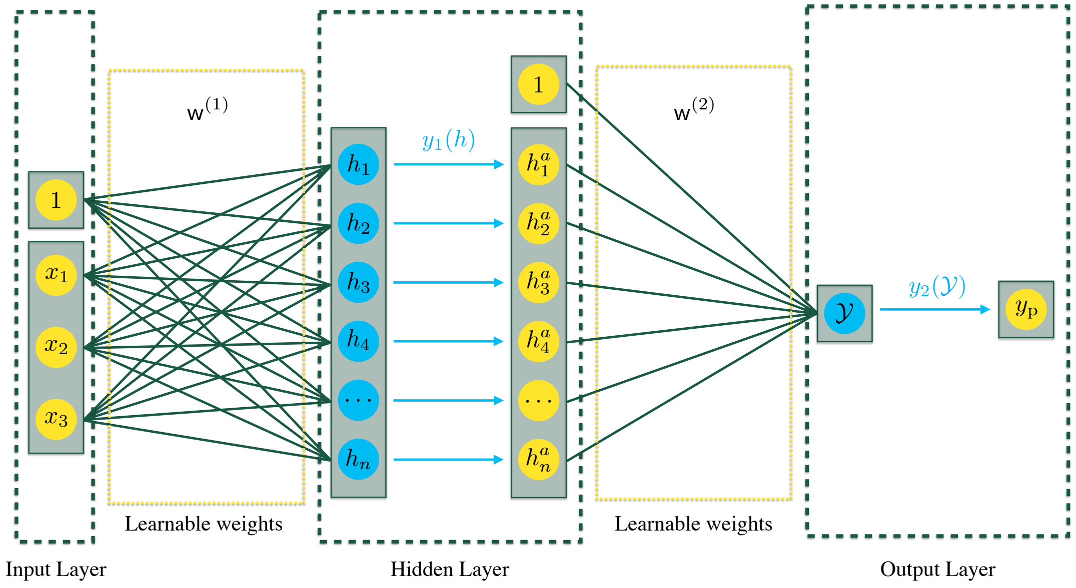
DM tracers



~~Gaia+Spectroscopic~~



Use NN to unlock full
potential of Gaia



Backup

Toy model of spiral galaxy

Examine the phase space distribution of the halo: $f(\mathbf{x}, \mathbf{v})$

$$\int f(\mathbf{x}, \mathbf{v})\, d^3\mathbf{x}\, d^3\mathbf{v} = 1$$ (probability)

Conservation of probability

$$\frac{\partial f}{\partial t} + \dot{\mathbf{x}}\frac{\partial f}{\partial \mathbf{x}} + \dot{\mathbf{v}}\frac{\partial f}{\partial \mathbf{v}} = 0$$

Boltzmann equation

- Non-relativistic
- Collisionless

Toy model of spiral galaxy

Examine the phase space distribution of the halo: $f(\mathbf{x}, \mathbf{v})$

$$\int f(\mathbf{x}, \mathbf{v})\, d^3\mathbf{x}\, d^3\mathbf{v} = 1$$ (probability)

Conservation of probability

$$\frac{\partial f}{\partial t} + \dot{\mathbf{x}} \frac{\partial f}{\partial \mathbf{x}} + \dot{\mathbf{v}} \frac{\partial f}{\partial \mathbf{v}} = 0$$

($\frac{\partial f}{\partial t} \to 0$)

Boltzmann equation
- Non-relativistic
- Collisionless

Steady state / Virialized / Equilibrium

$$f(\mathbf{x}, \mathbf{v}) = f(\mathcal{E})$$

$$\mathcal{E} = \phi - \frac{1}{2} v^2$$

must be function of integrals of motion

$$f(\mathbf{x}, \mathbf{v}) \propto e^{\frac{\phi - v^2/2}{\sigma^2}}$$

Toy model of spiral galaxy

Use the phase space density to get derive the particle density

$$\rho(\mathbf{x}) = \int d^3v\ f(\mathbf{x}, \mathbf{v}) = \int 4\pi v^2 f(\mathbf{x}, \mathbf{v})\ dv \propto e^{\phi/\sigma^2}$$

$$\phi \propto \sigma^2 \log \rho(\mathbf{x})$$

Toy model of spiral galaxy

Use the phase space density to get derive the particle density

$$\rho(\mathbf{x}) = \int d^3v \; f(\mathbf{x}, \mathbf{v}) = \int 4\pi v^2 f(\mathbf{x}, \mathbf{v}) \; dv \propto e^{\phi/\sigma^2}$$

$$\phi \propto \sigma^2 \log \rho(\mathbf{x})$$

<u>Use Gauss' Law</u>:

$$\nabla^2 \phi = -4\pi G \rho(\mathbf{x}) \propto \nabla^2 \left(\sigma^2 \log \rho(\mathbf{x})\right)$$

Toy model of spiral galaxy

Use the phase space density to get derive the particle density

$$\rho(\mathbf{x}) = \int d^3v\ f(\mathbf{x},\mathbf{v}) = \int 4\pi v^2 f(\mathbf{x},\mathbf{v})\ dv \propto e^{\phi/\sigma^2}$$

$$\phi \propto \sigma^2 \log \rho(\mathbf{x})$$

Use Gauss' Law:

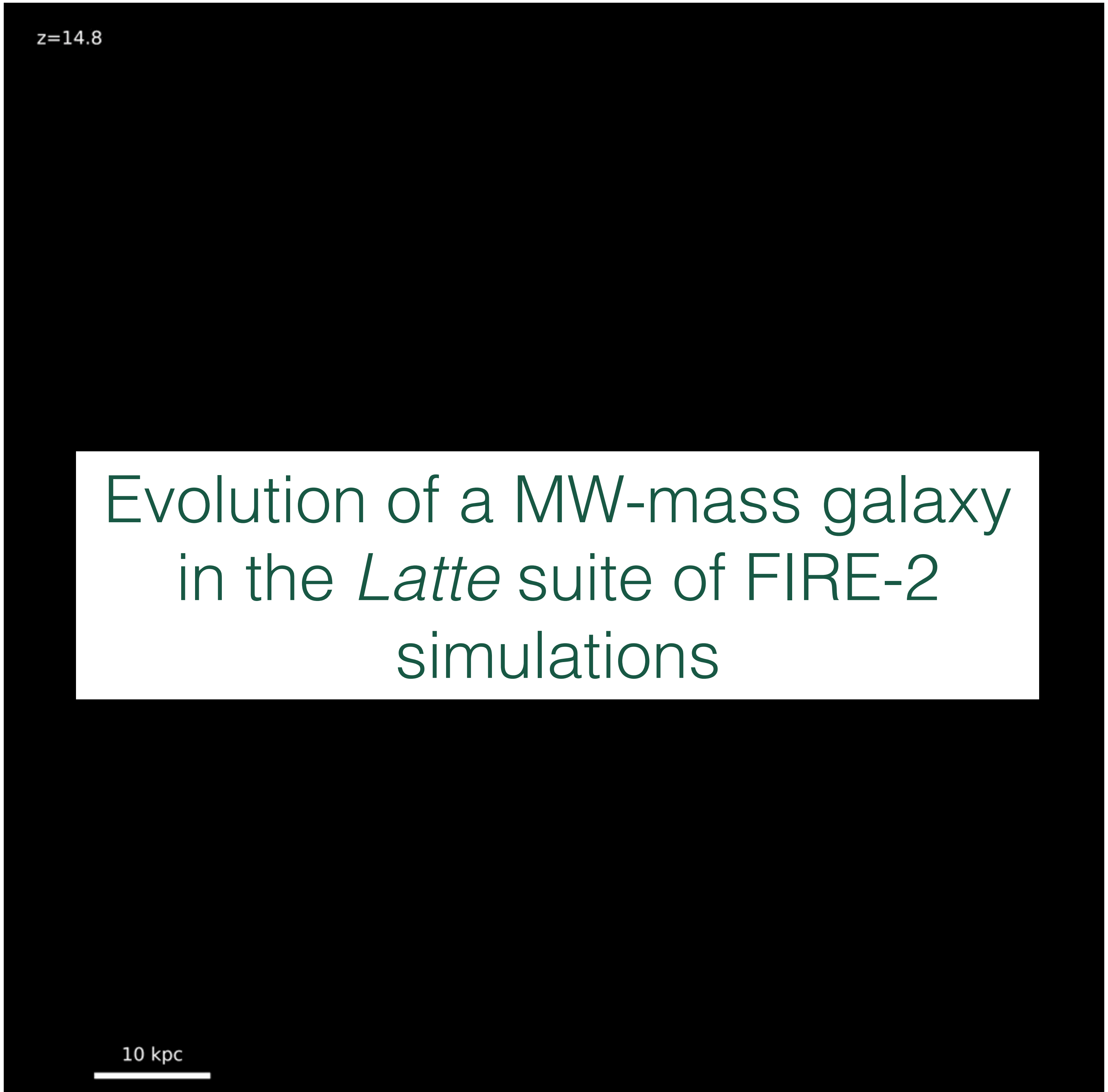
$$\nabla^2\phi = -4\pi G\rho(\mathbf{x}) \propto \nabla^2\left(\sigma^2 \log \rho(\mathbf{x})\right)$$

- collisionless
- self-gravitating
- isotropic
- isothermal gas

$$\rho(\mathbf{x}) \propto \frac{\sigma^2}{2\pi G r^2}$$

$$f(\mathbf{v}) \propto e^{\frac{-v^2}{2\sigma^2}}$$

NOT a toy model of spiral galaxy



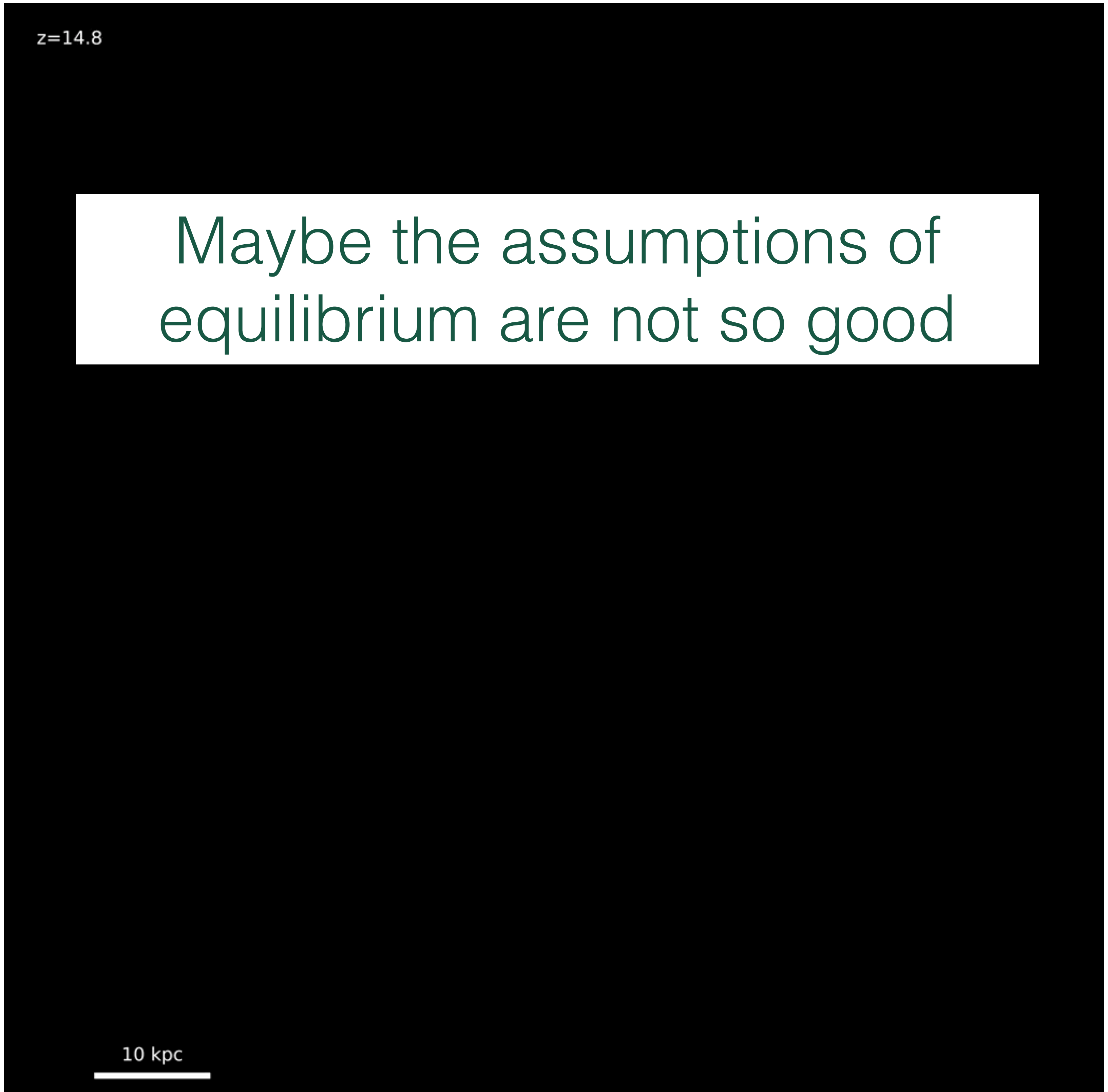
Evolution of a MW-mass galaxy in the *Latte* suite of FIRE-2 simulations

http://www.tapir.caltech.edu/~sheagk/starvids.html

NOT a toy model of spiral galaxy

http://www.tapir.caltech.edu/~sheagk/starvids.html

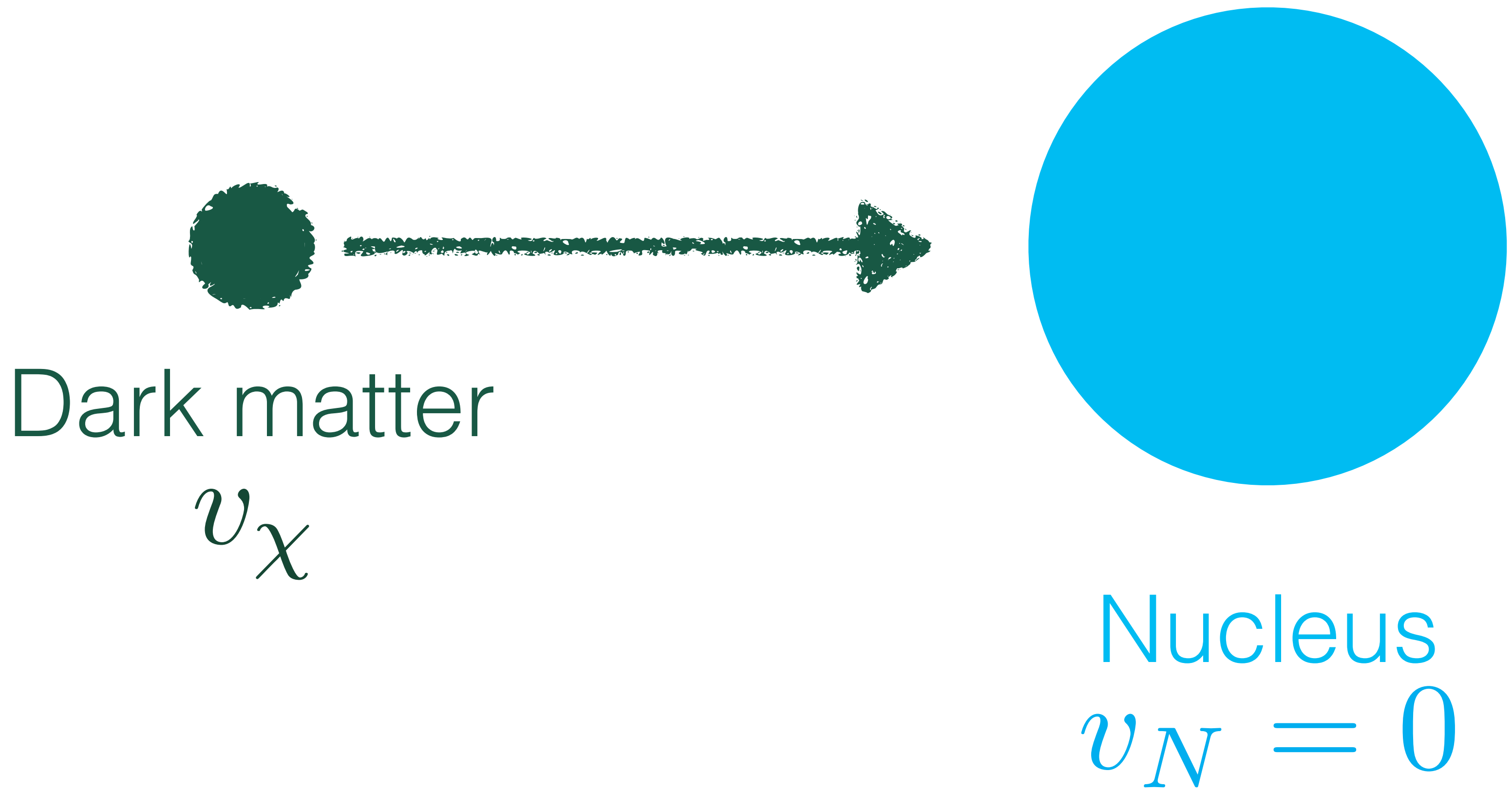
NOT a toy model of spiral galaxy



Maybe the assumptions of equilibrium are not so good

http://www.tapir.caltech.edu/~sheagk/starvids.html

Dark matter direct detection



Dark matter direct detection

Dark matter

Nucleus

Kinetic energy of recoiling nucleus: $E_{R,\mathrm{max}} = \dfrac{2\mu^2 v_\chi^2}{M_N}$

Only detect recoils above threshold energy

$$v_{\chi,\mathrm{min}} = \sqrt{\frac{M_N E_{R,\mathrm{threshold}}}{2\mu^2}}$$

Small mass needs more velocity

Dark matter direct detection

The rate of nuclear recoils:

$$\frac{dR}{dE_R} = \frac{1}{m_N}\frac{\rho_\chi}{m_\chi}\int_{v_{\rm min}}^{v_{\rm max}} d^3v\ v\tilde{f}(\mathbf{v}) \quad \frac{d\sigma(v)}{dE_R}$$

Dark matter direct detection

The rate of nuclear recoils:

$$\frac{dR}{dE_R} = \frac{1}{m_N} \frac{\rho_\chi}{m_\chi} \int_{v_{\min}}^{v_{\max}} d^3v \; v \tilde{f}(\mathbf{v}) \; \frac{d\sigma(v)}{dE_R}$$

$\frac{\rho_\chi}{m_\chi} = n_\chi$: how many dark matter particles around

Dark matter direct detection

The rate of nuclear recoils:

$$\frac{dR}{dE_R} = \frac{1}{m_N} \frac{\rho_\chi}{m_\chi} \int_{v_{\min}}^{v_{\max}} d^3v \; v \tilde{f}(\mathbf{v}) \frac{d\sigma(v)}{dE_R}$$

$\frac{\rho_\chi}{m_\chi} = n_\chi$: how many dark matter particles around

$\frac{d\sigma(v)}{dE_R}$: how likely to scatter given the velocity

Dark matter direct detection

The rate of nuclear recoils:

$$\frac{dR}{dE_R} = \frac{1}{m_N} \frac{\rho_\chi}{m_\chi} \int_{v_{\min}}^{v_{\max}} d^3v \; v \tilde{f}(\mathbf{v}) \frac{d\sigma(v)}{dE_R}$$

$\frac{\rho_\chi}{m_\chi} = n_\chi$: how many dark matter particles around

$\frac{d\sigma(v)}{dE_R}$: how likely to scatter given the velocity

$\tilde{f}(\mathbf{v})$: the probability to have certain velocity

Dark matter direct detection

The rate of nuclear recoils:

$$\frac{dR}{dE_R} = \frac{1}{m_N} \frac{\rho_\chi}{m_\chi} \int_{v_{\min}}^{v_{\max}} d^3v \; v \tilde{f}(\mathbf{v}) \frac{d\sigma(v)}{dE_R}$$

galaxy escape velocity

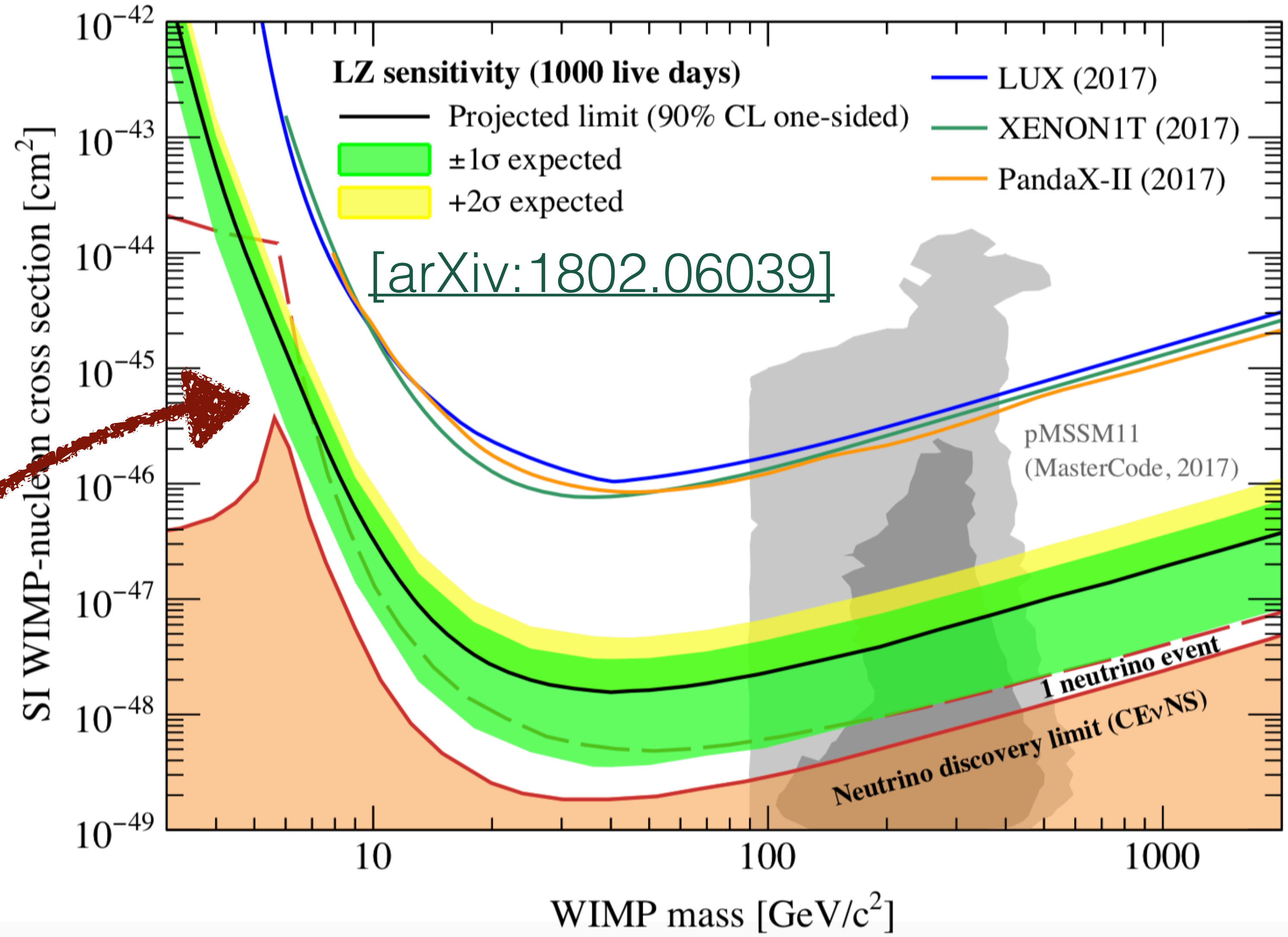
minimum velocity to achieve E_R

$\frac{\rho_\chi}{m_\chi} = n_\chi$: how many dark matter particles around

$\frac{d\sigma(v)}{dE_R}$: how likely to scatter given the velocity

$\tilde{f}(\mathbf{v})$: the probability to have certain velocity

Dark matter direct detection



[arXiv:1802.06039]

E_{R}, threshold

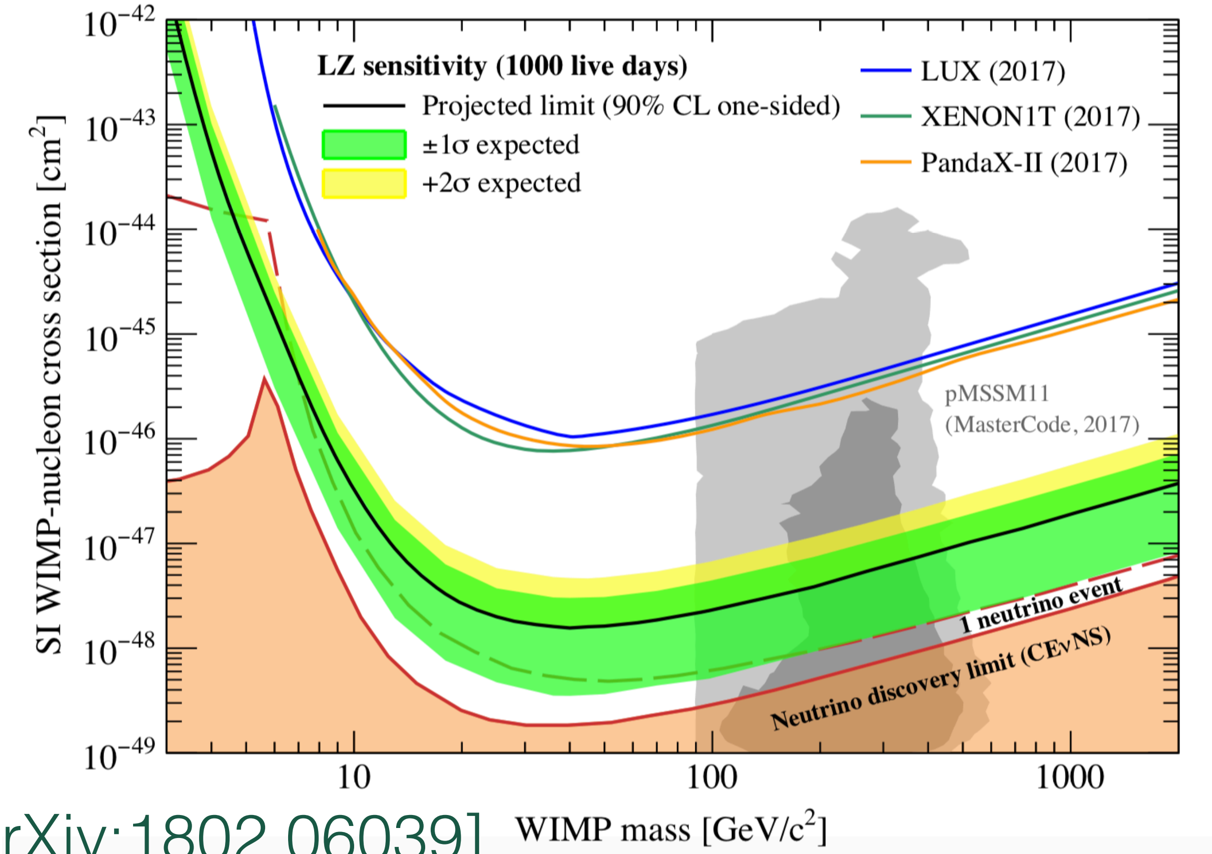
sensitivity falls off as $1/m_\chi$ for fixed mass density.

$\tilde{f}(\mathbf{v})$: the probability to have certain velocity

	# sources in Gaia DR2	# sources in Gaia DR1
Total number of sources	**1,692,919,135**	**1,142,679,769**
Number of 5-parameter sources	1,331,909,727	2,057,050
Number of 2-parameter sources	361,009,408	1,140,622,719
Sources with mean G magnitude	1,692,919,135	1,142,679,769
Sources with mean G_{BP}-band photometry	1,381,964,755	-
Sources with mean G_{RP}-band photometry	1,383,551,713	-
Sources with radial velocities	7,224,631	-
Variable sources	550,737	3,194
Known asteroids with epoch data	14,099	-
Gaia-CRF sources	556,869	2,191
Effective temperatures (T_{eff})	161,497,595	-
Extinction (A_G) and reddening ($E(G_{BP}$-$G_{RP})$)	87,733,672	-
Sources with radius and luminosity	76,956,778	-

https://www.cosmos.esa.int/web/gaia/dr2

Toy model of spiral galaxy



[arXiv:1802.06039]

$$f(\mathbf{v}) = \begin{cases} \frac{1}{N}\left(e^{-v^2/v_0^2} - e^{-v_{esc}^2/v_0^2}\right) & v < v_{esc} \\ 0 & v > v_{esc} \end{cases}$$