Galaxy Cluster Mass Estimation Using Deep Learning

Matthew Ho
Department of Physics, Carnegie Mellon University
mho1@andrew.cmu.edu

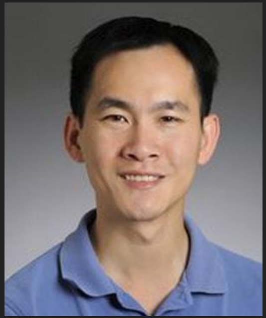
LBNL INPA 2021

MACS J0416.1-2403
Image Credit: Hubble

Outline

- Introduction
  - Dynamical Probes of Dark Matter
- Dynamical Masses from Deep Learning
  - Brief guide to CNNs
  - Application to cluster dynamics
- Uncertainty on Deep Learning Mass Estimates
  - Applications of Approximate Bayesian NNs
- Empirical Verification
  - Application to Coma Cluster
- Future Prospects
  - Multi-wavelength cluster measurements

Collaborators



Hy Trac (CMU)

Michelle Ntampaka (STScI)

Markus Michael Rau (Argonne)

Arya Farahi (UT Austin)

Gus Evrard (Michigan)

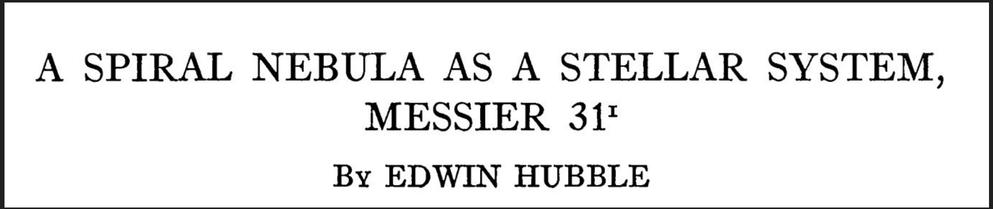
Daisuke Nagai (Yale)

Barnabás Póczos (CMU)

Minghan Chen (UCSB, CMU), **Alexa Lansberry** (CMU), **Faith Ruehle** (CMU), **Bryant Dean** (Morehouse) … and many more!

Introduction: Dynamical Probes of Dark Matter

Astronomy in the 1930s



A SPIRAL NEBULA AS A STELLAR SYSTEM,
MESSIER 31[1]

BY EDWIN HUBBLE

1924 - Edwin Hubble observes proof that `faint nebulae' were, in fact, other galaxies.

1925 - Cecilia Payne-Gaposchkin discovers the abundance of hydrogen in stellar spectra.

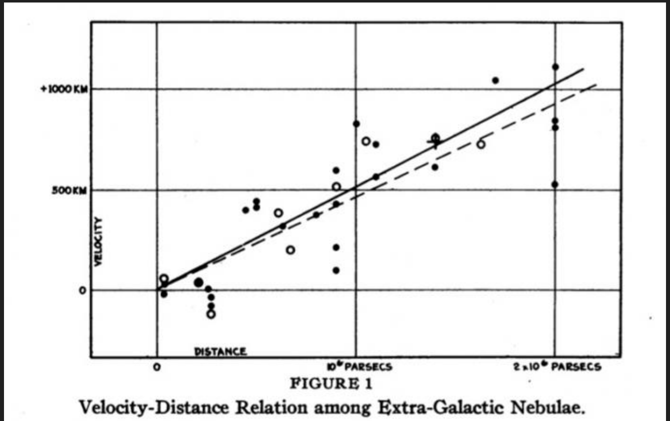
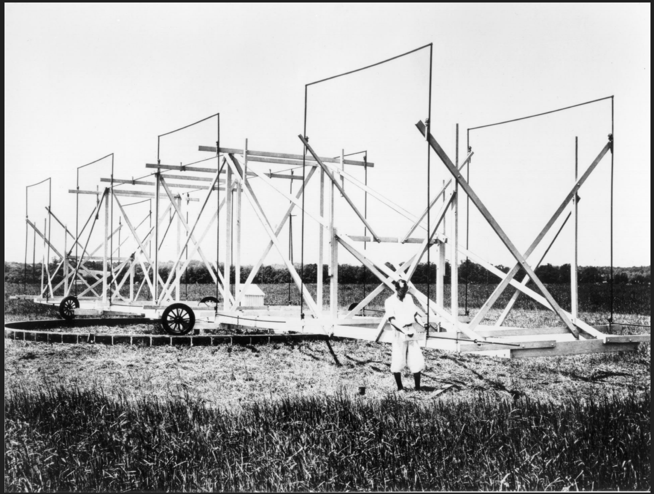


FIGURE 1
Velocity-Distance Relation among Extra-Galactic Nebulae.

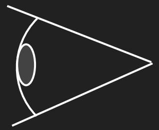
1929 - Edwin Hubble publishes studies on the distance-redshift relation.



1932 - Karl Jansky builds the first radio telescope

Fritz Zwicky and the Coma Cluster

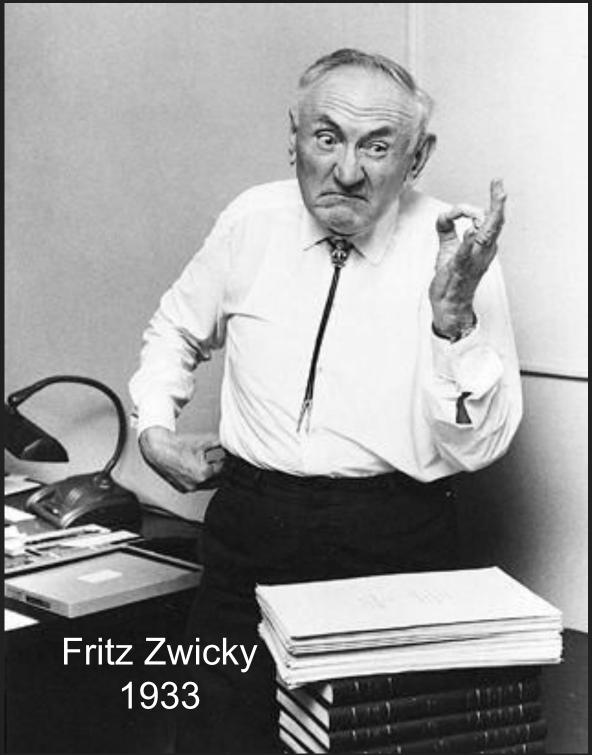

Coma Cluster,
Schulman Telescope



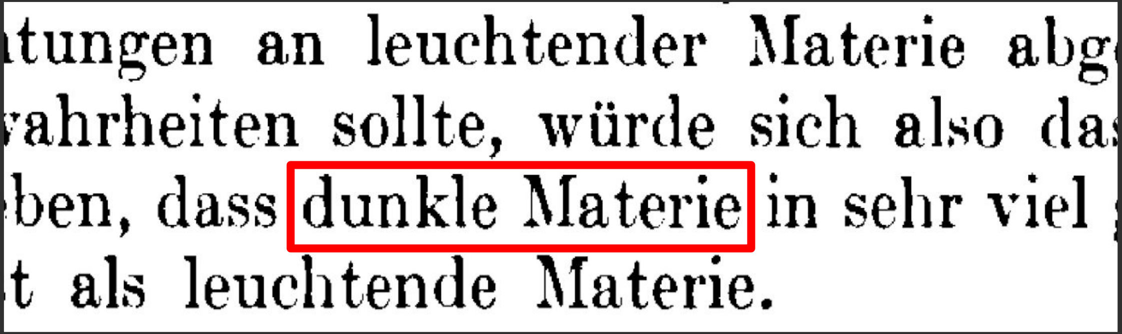
$$2\langle K\rangle = \langle V\rangle$$

Scheinbare Geschwindigkeiten im Comahaufen.

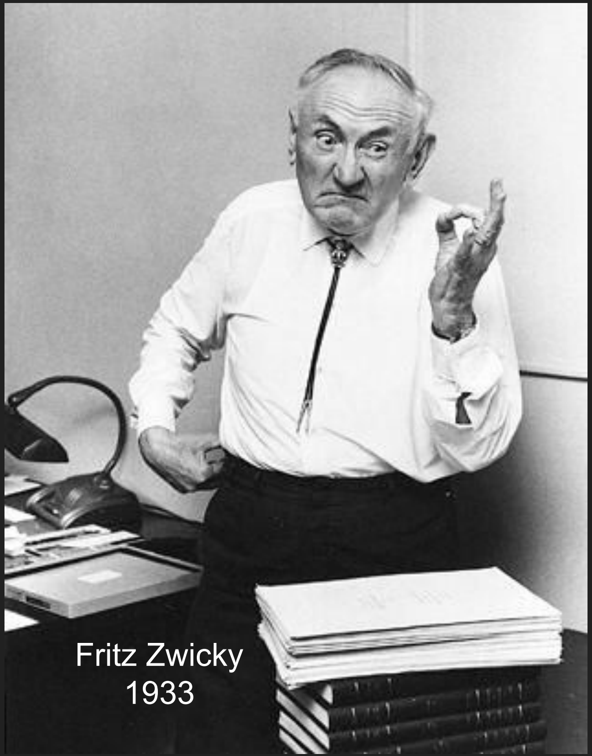
v = 8500 km/sek	6900 km/sek
7900	6700
7600	6600
7000	5100 (?)


Fritz Zwicky
1933

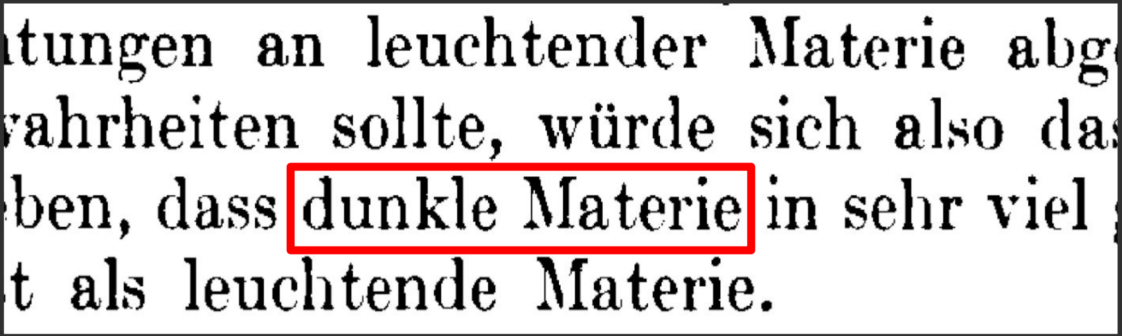
Fritz Zwicky and the Coma Cluster

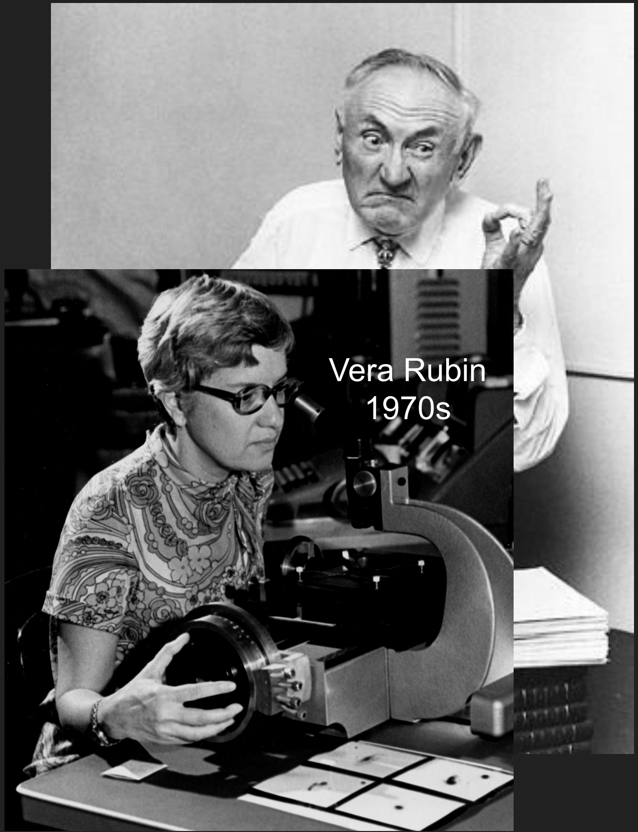

tungen an leuchtender Materie abg
ahrheiten sollte, würde sich also das
ben, dass dunkle Materie in sehr viel
t als leuchtende Materie.

Zwicky’s observation and analysis of the Coma cluster mass is often considered to be the first inference of dark matter!*†


Fritz Zwicky
1933

Fritz Zwicky and the Coma Cluster


tungen an leuchtender Materie abg
vahrheiten sollte, würde sich also das
ben, dass dunkle Materie in sehr viel
t als leuchtende Materie.



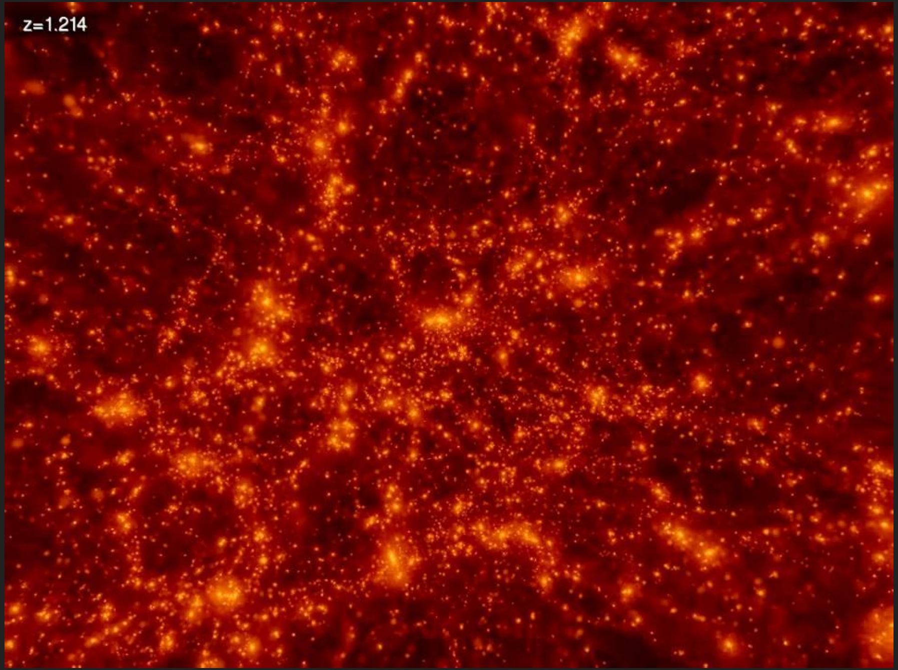
Zwicky's observation and analysis of the Coma cluster mass is often considered to be the first inference of dark matter!*†

* Kelvin (1884), Kapteyn (1922), Oort (1932)
† Not widely accepted until Rubin, Ford, and Freeman in the 1970s

90 Years Later…

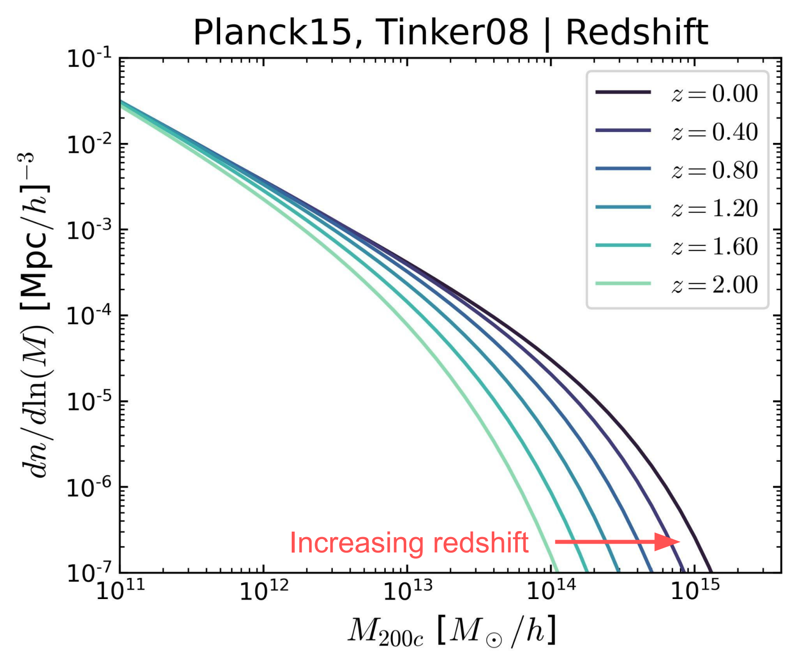
Millennium Simulations

The Halo Mass Function (HMF)

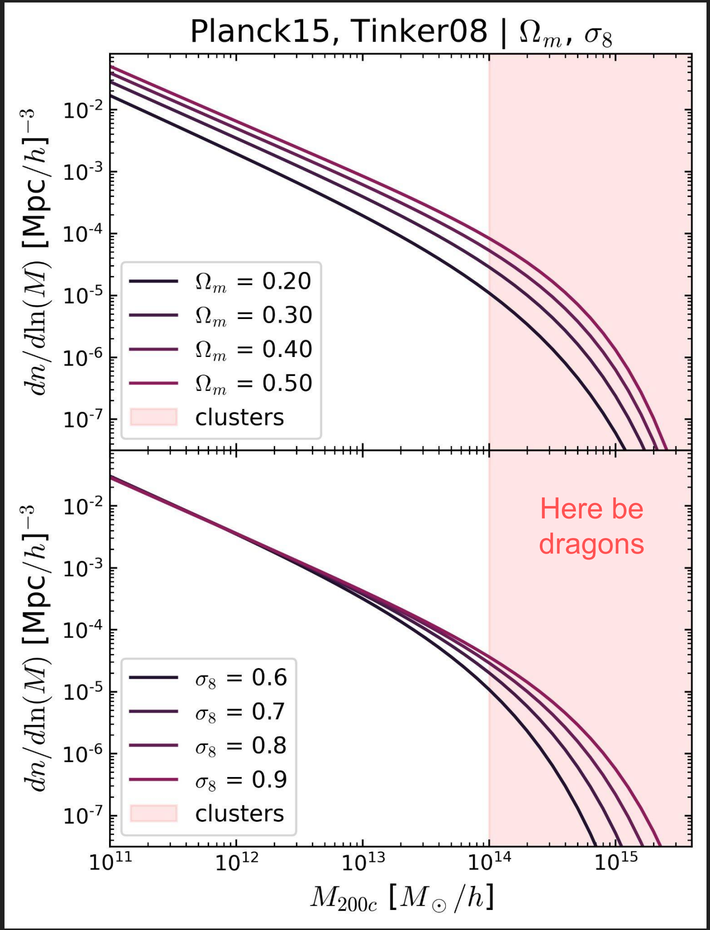


MultiDark Resimulations (Kristin Riebe)

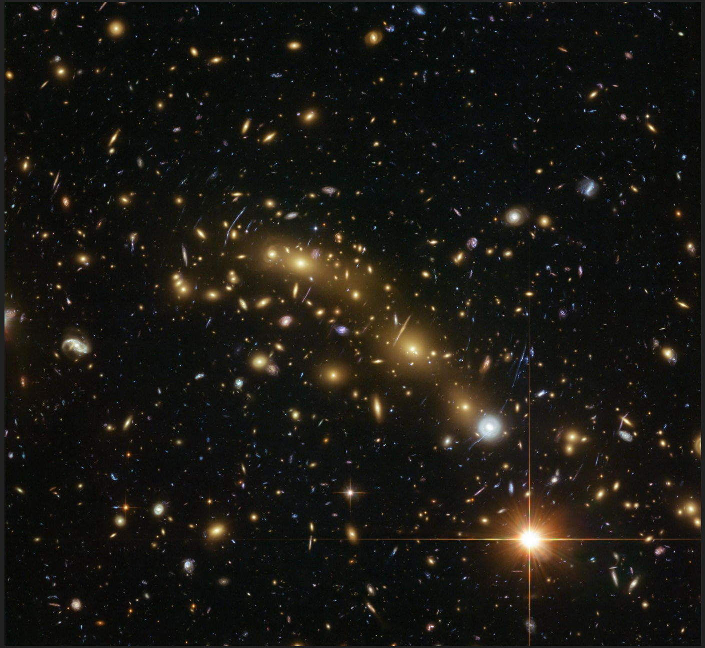
Halo abundance



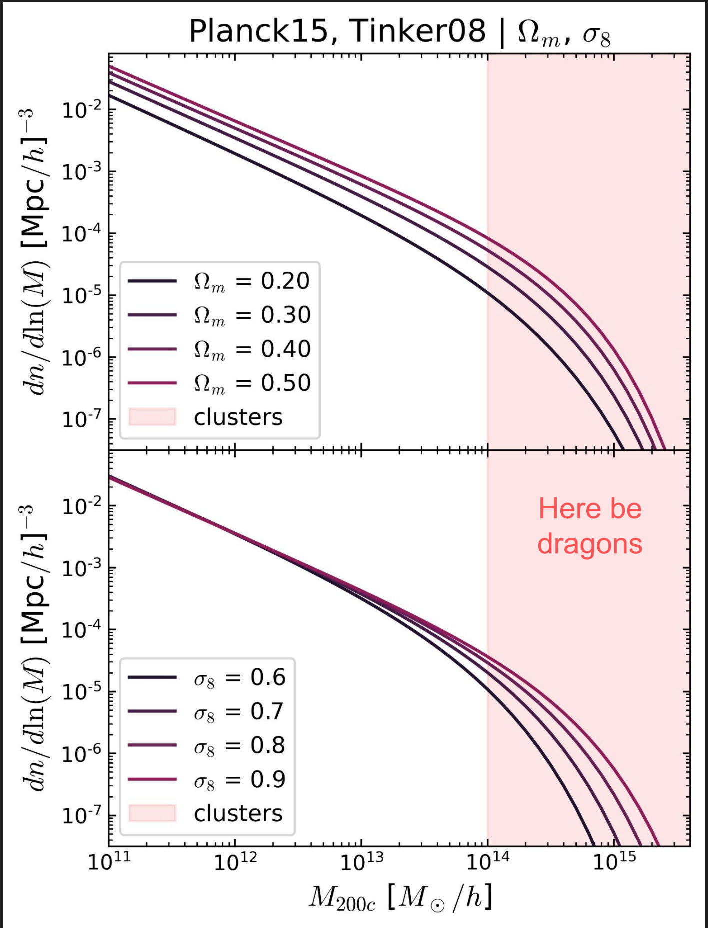
Halo mass

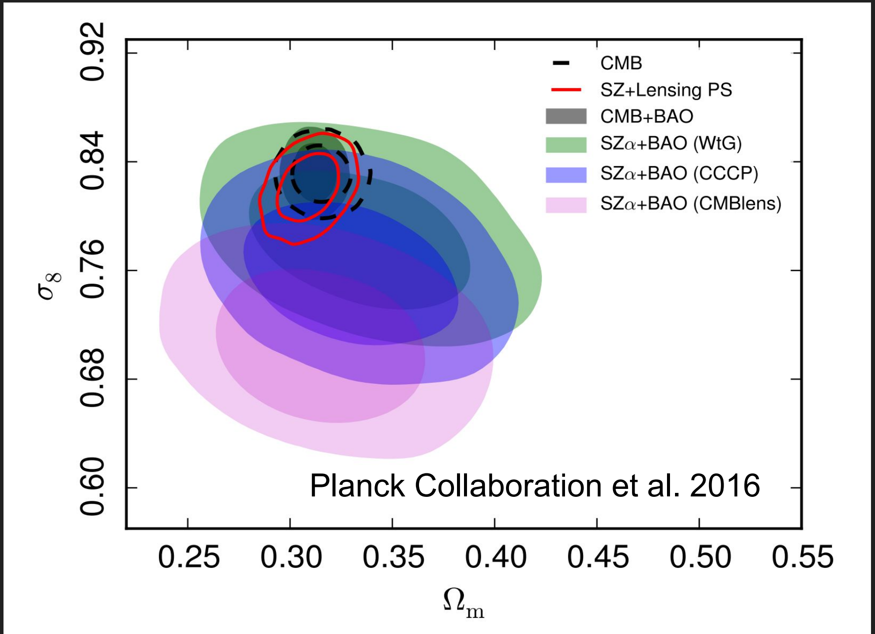


"Most massive bound structures in the universe"



MACS J0416.1-2403
Image Credit: Hubble

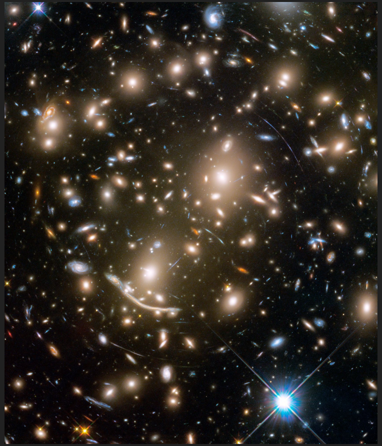
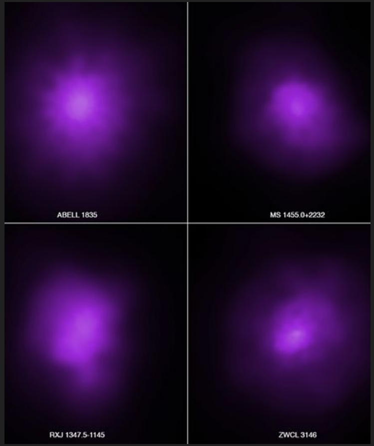
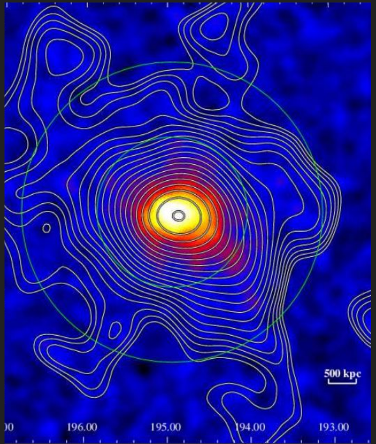
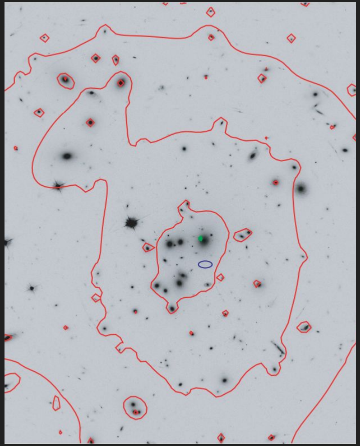
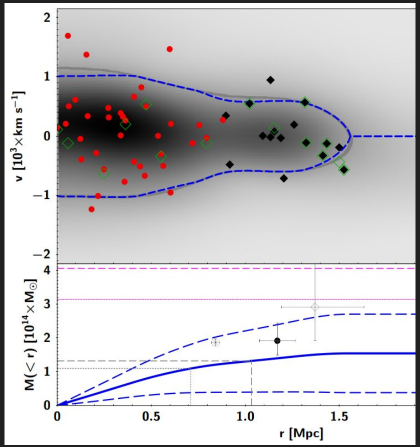




<u>To constrain the HMF:</u>

- Large, well-defined cluster sample
- Robust mass-measurement methods
  - Efficient and automated

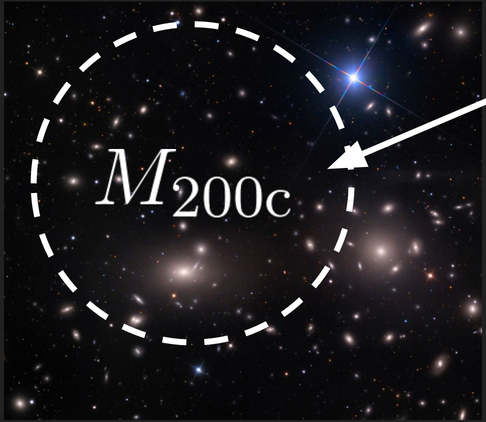
Mass Measurements of Galaxy Clusters

Optical	X-ray	Microwave	Multi-band	Spectroscopic
				
Number of galaxies above fixed luminosity	Hot, photon-emitting intracluster gas	Scattering of CMB photons on intracluster plasma	Lensing of background light	Galaxy dynamics

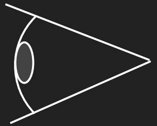
Dynamical Masses and The M-σ

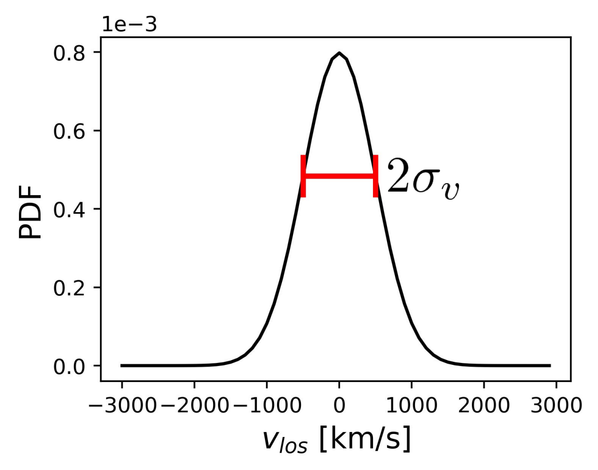
Assuming: spherical symmetry, gravitational equilibrium, identical galaxies, perfect selection

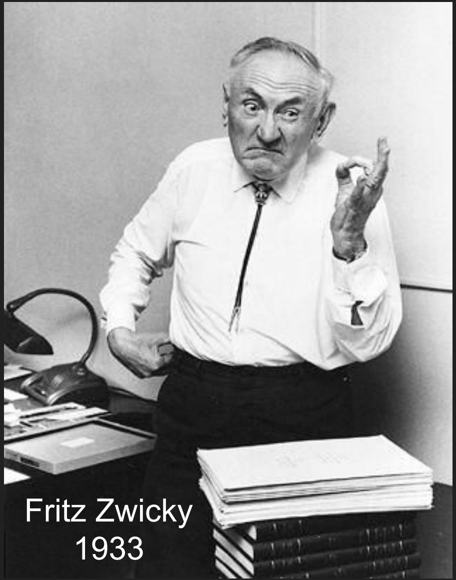
$$\boxed{\sigma_v \propto [M_{200c}]^{\alpha} \qquad \alpha \approx 1/3}$$



$\rho = 200\rho_c$





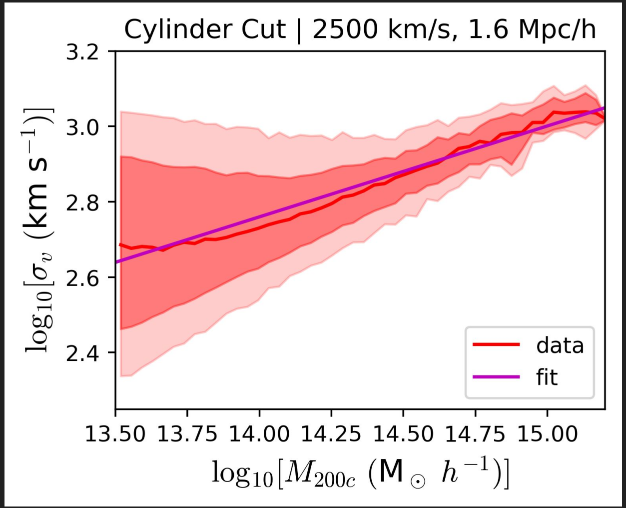


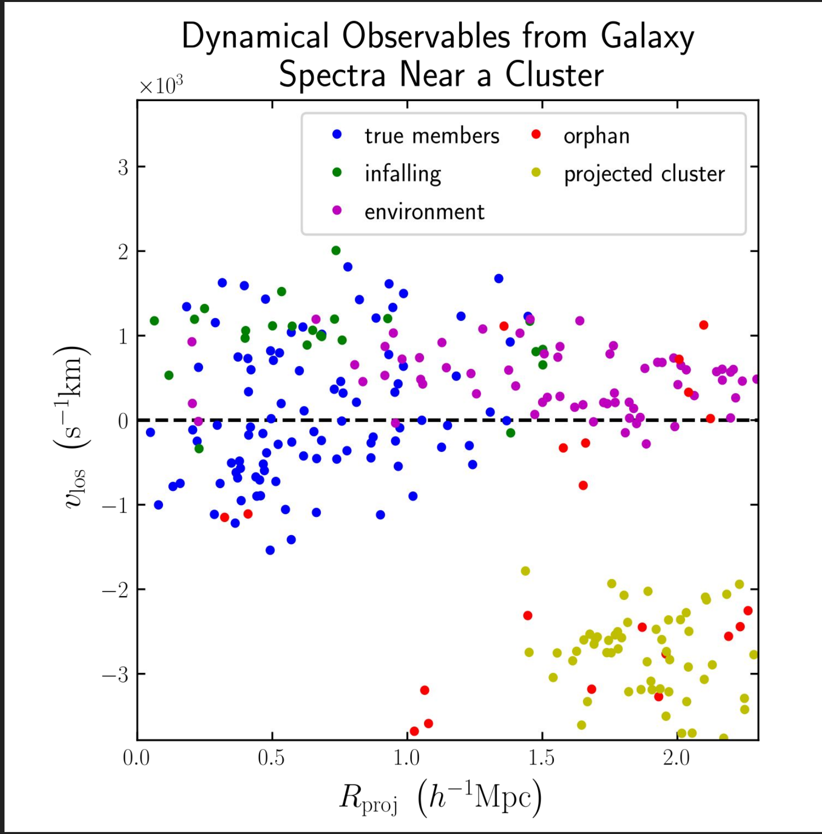
Fritz Zwicky
1933

First-order stats are not sufficient to capture galaxy dynamics!

Previous work has investigated impacts of:

- Dynamical substructure (Old et al. 2018)
- Halo environment (White et al. 2010)
- Triaxiality (Svensmark et al. 2015)
- Mergers (Evrard et al. 2008)
- Sample Contamination (Wojtak et al. 2018)





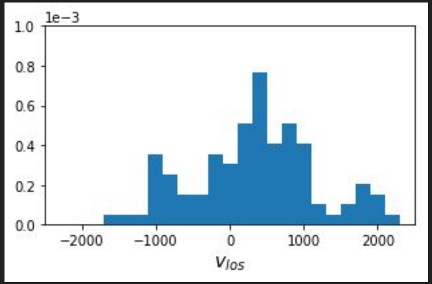
Dynamical Masses from Deep Learning

Dynamical Masses from Deep Learning (Ho et al. 2019)

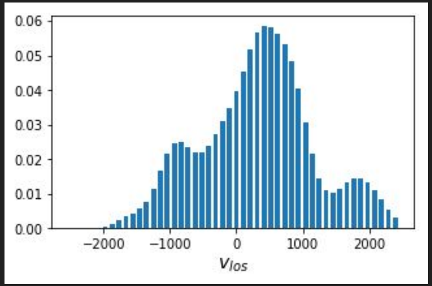
Our model should…

- Learn to identify features representative of cluster substructure and interlopers
- Relate these features to mass predictions in a complex manner

CNN_{1D}



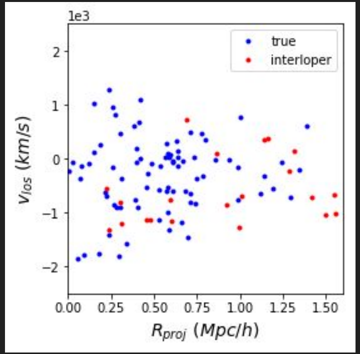
KDE



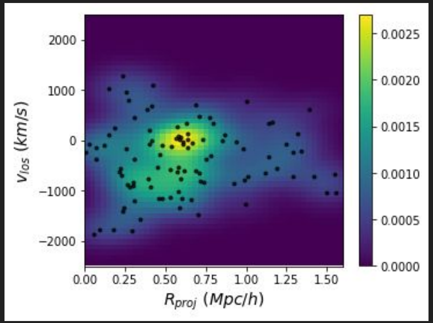
Convolutional Neural Network

M_{200c}

CNN_{2D}



KDE



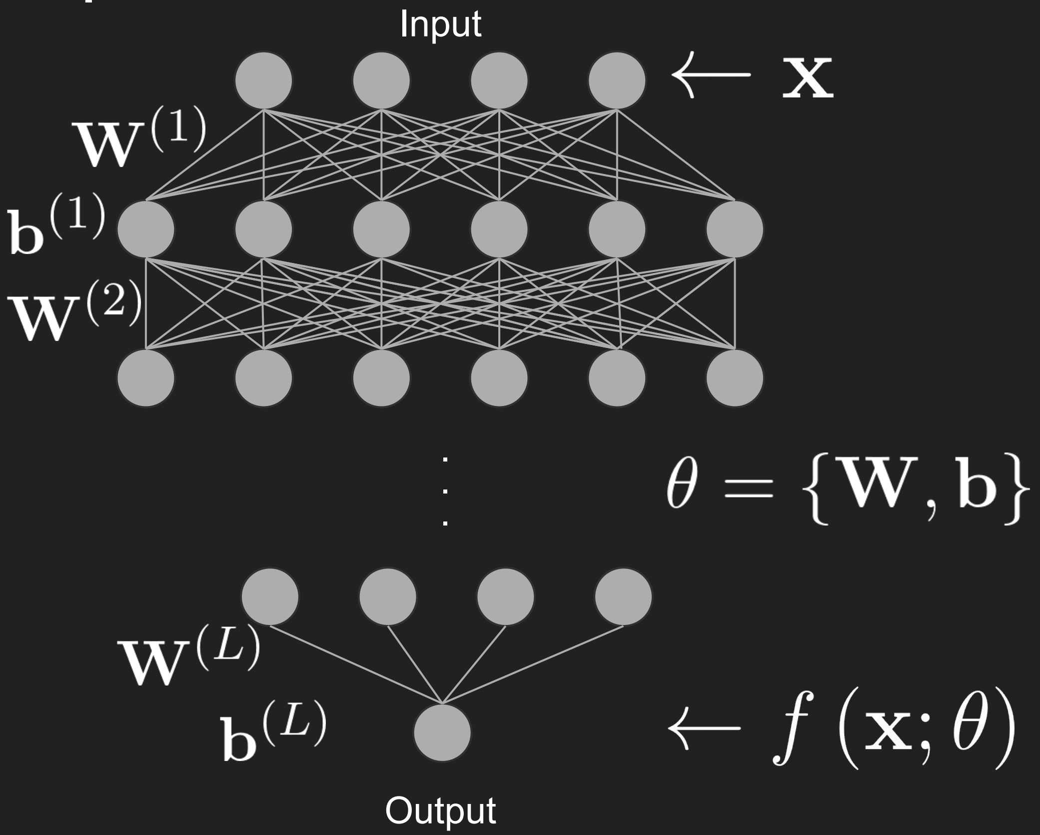
Convolutional Neural Network

M_{200c}

A Very Brief Guide to Deep Neural Networks

Deep Neural Networks

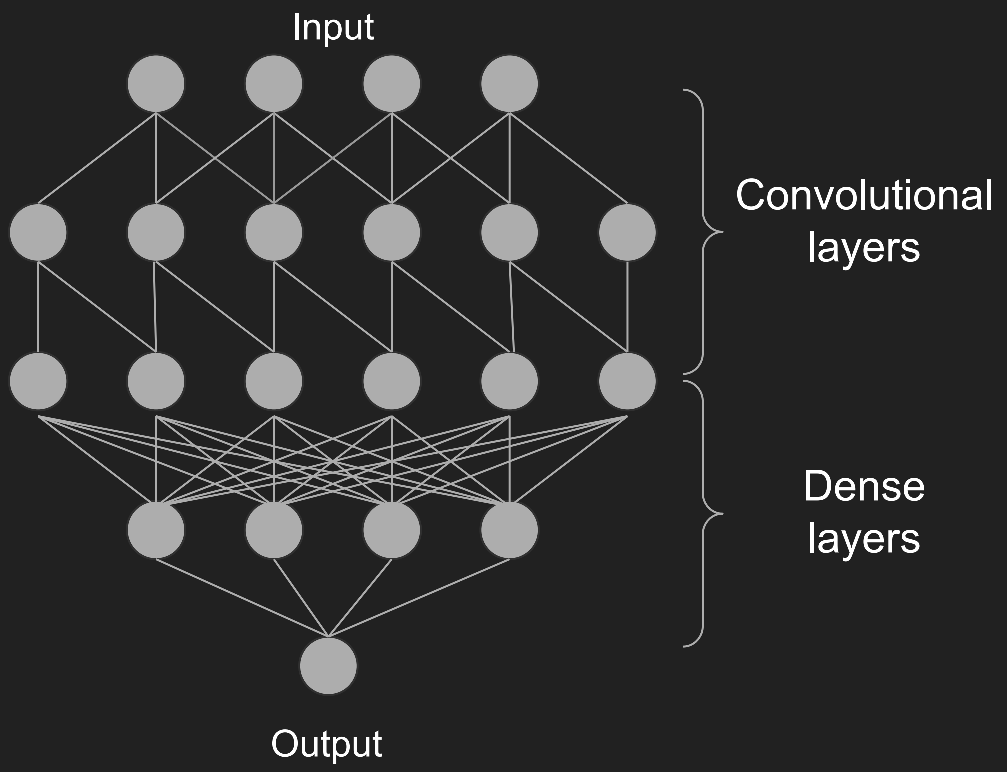
- Highly non-linear functions with nice gradients
- Very overparameterized (100,000+ parameters)



A Very Brief Guide to Convolutional Neural Networks

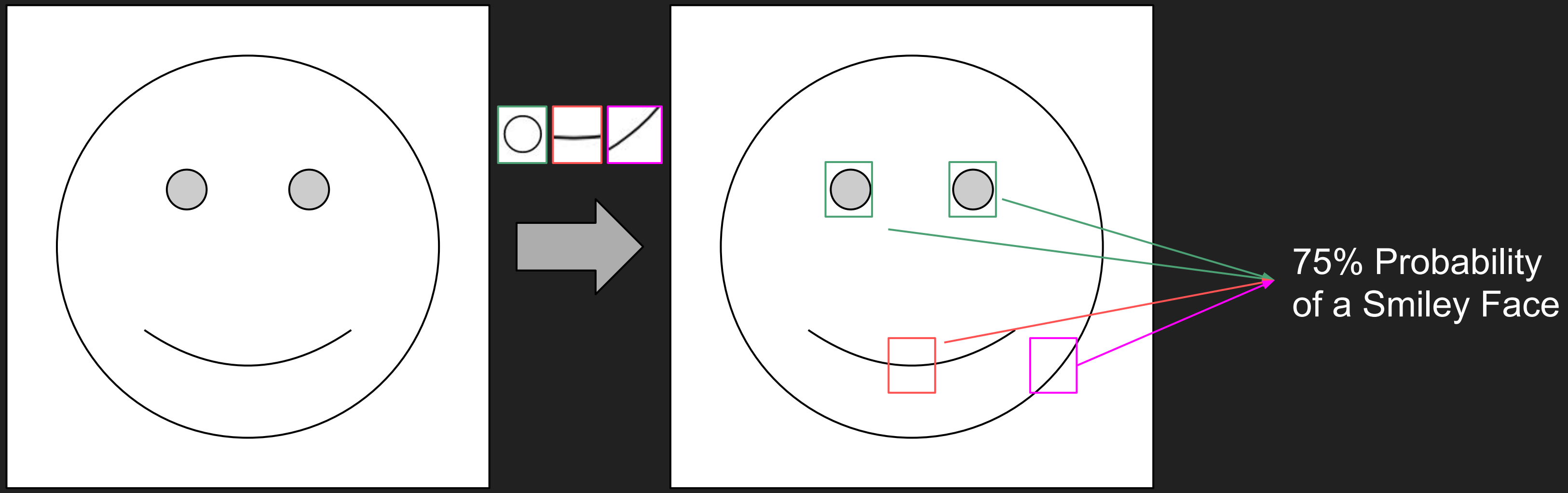
Convolutional Neural Networks (CNNs)

- Gold standard for image-recognition tasks
- Utilize shared feature filters in first layers
- Find localized patterns in subinputs



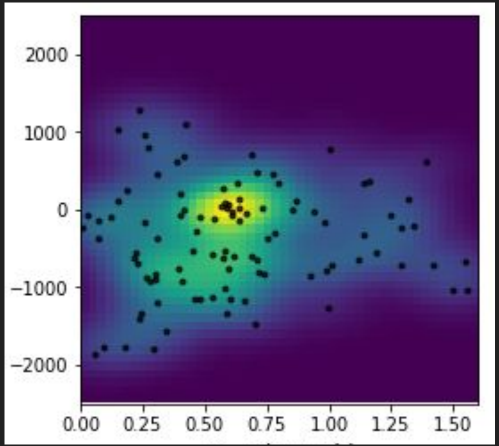
A Very Brief Guide to Convolutional Neural Networks

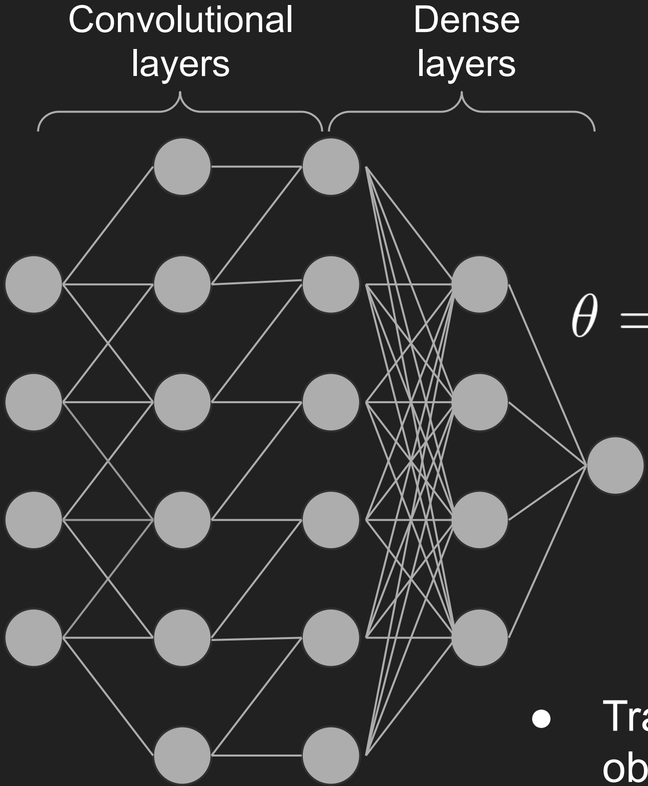
- Training feature filters:



Convolutional Networks for Cluster Mass Estimation

Galaxy positions and velocities





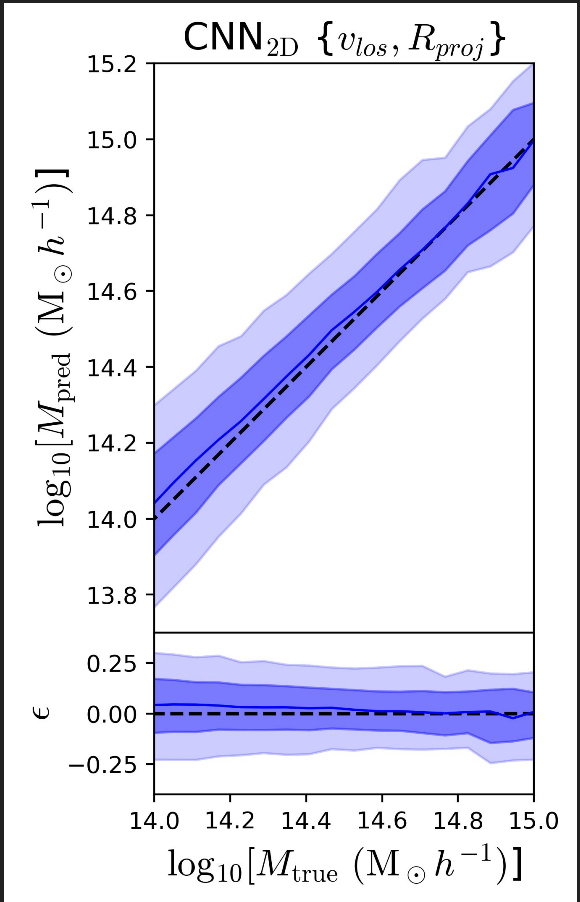
$$\theta = \{\mathbf{W}, \mathbf{b}\}$$

$$y = f(x;\theta) \rightarrow \hat{M}_{200c}$$

*simplified architecture

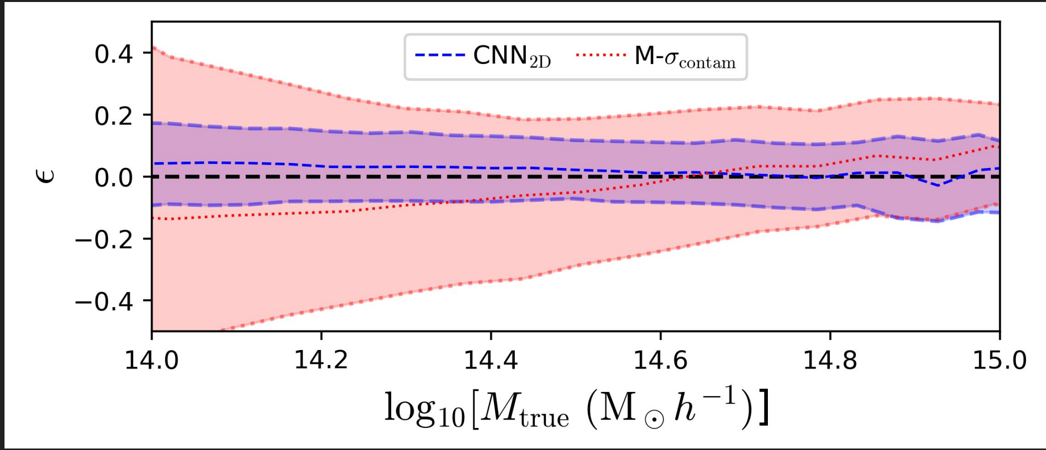
- Trained on realistic mock observations using MDPL2 halos and UniverseMachine galaxies

Dynamical Masses from Deep Learning (Ho et al. 2019)



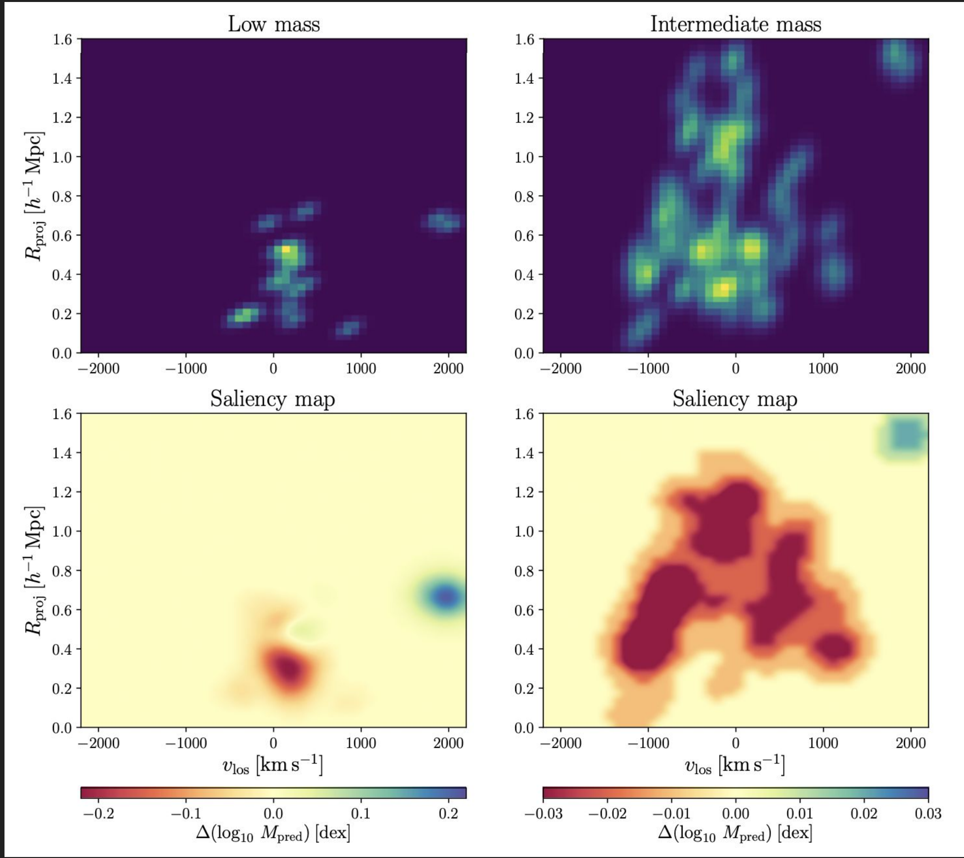
Low bias and Gaussian scatter

- Models reduce scatter of simple M-σ measurements by a factor of ~2.5.
- Models improve prediction scatter relative to ‘ideal’ M-σ measurements (i.e. no selection effects) by 30%.



Interpreting Dynamical Deep Learning

Deep learning models learn to downvote interlopers and emphasize substructure.



Doogesh Kodi Ramanah (DARK)

arXiv:2003.05951
arXiv:2009.03340

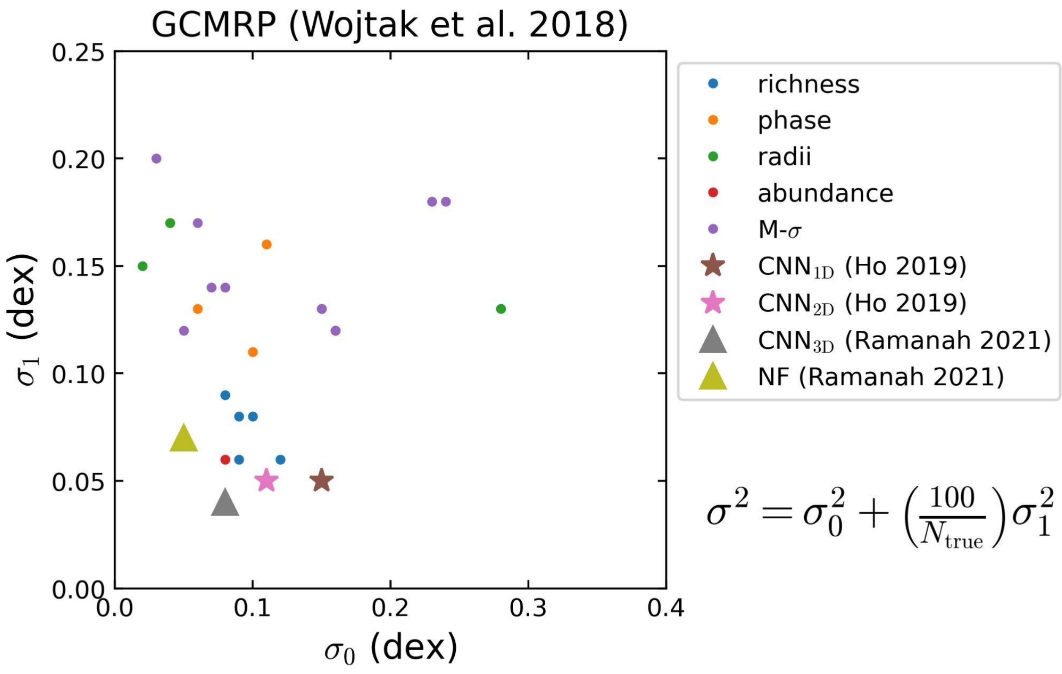
Dynamical Masses from Deep Learning (Ho et al. 2019)

Robustness

- Reduced sensitivity to cluster richness

Computational Efficiency

- Reduced training+inference time by 30x when compared to other ML approaches (SDM; Ntampaka et al. 2015, 2016)



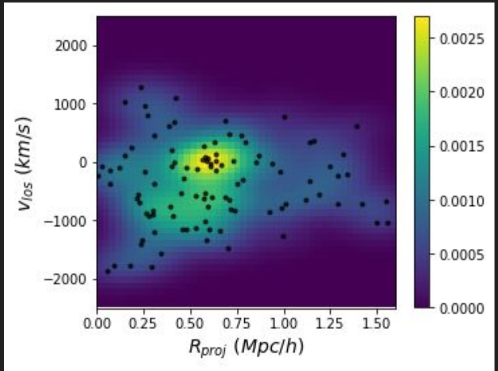
$$\sigma^2 = \sigma_0^2 + \left(\frac{100}{N_{\mathrm{true}}}\right)\sigma_1^2$$

Uncertainty on Deep Learning Mass Estimates

Uncertainty in Deep Learning

Methods to recover deep learning uncertainties:

- Approximate Bayesian Neural Networks (Ho et al. 2020)
- Forward modeling
- Normalizing Flows (Ramanah et al. 2020)
- Simulation-based inference (Ramanah et al. 2020)



→ Convolutional Neural Network → $p\left(M_{200c}|\mathbf{x}\right)$

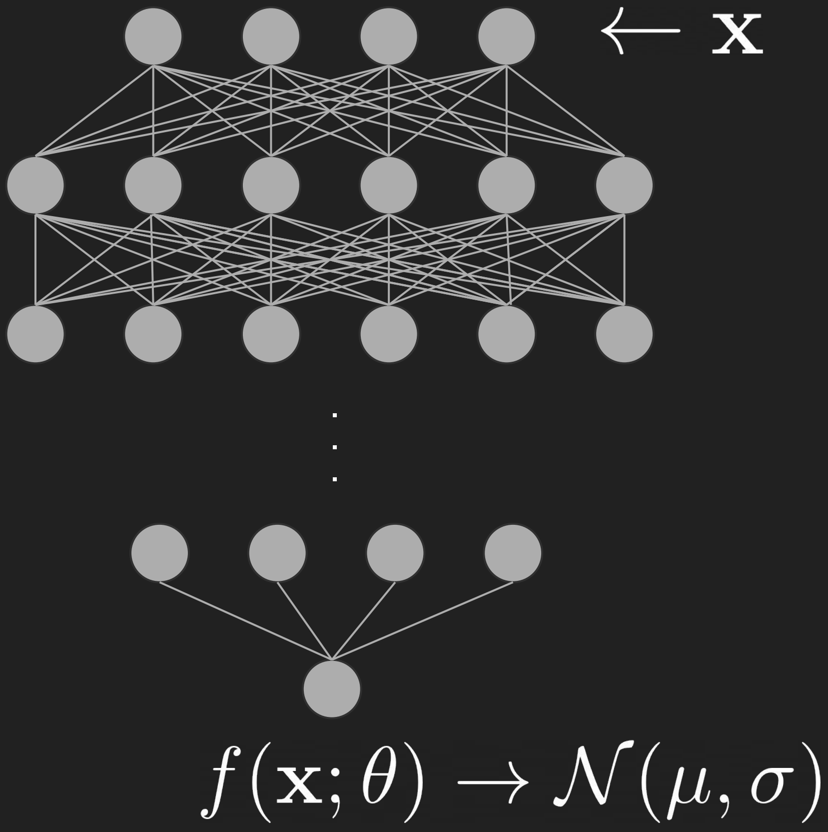
A Very Brief Guide to Approximate BNNs

<u>Aleatoric uncertainty</u> - Intrinsic scatter in input-output relationships

$$p\left(m \mid \mathbf{x}, \theta, \eta\right)$$

We can choose this distribution can be:

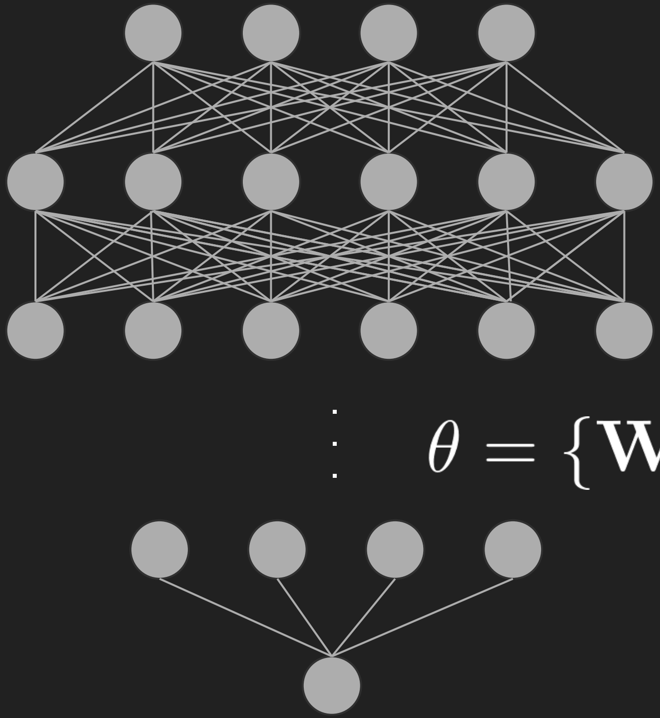
- Normal or log-normal
- Categorical
- Poisson
- Etc.



A Very Brief Guide to Approximate BNNs

Epistemic uncertainty - Uncertainty in parameter settings achieved during model training

$$p(\theta \mid \eta, \mathcal{D})$$



$$\theta = \{\mathbf{W}, \mathbf{b}\}$$

Sources of epistemic uncertainty:

- Insufficient training data
- Limited training time
- Inflexible model architectures

A Very Brief Guide to Approximate BNNs

$$\underbrace{p(\mathbf{m} \mid \mathbf{x}, \boldsymbol{\eta}, \mathcal{D})}_{\text{Mass posterior}} = \int \underbrace{p(\mathbf{m} \mid \mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\eta})}_{\text{Output of NN}} \underbrace{p(\boldsymbol{\theta} \mid \boldsymbol{\eta}, \mathcal{D})}_{\text{Uncertainty of weights}} d\boldsymbol{\theta}$$

Intractable for deep neural networks!

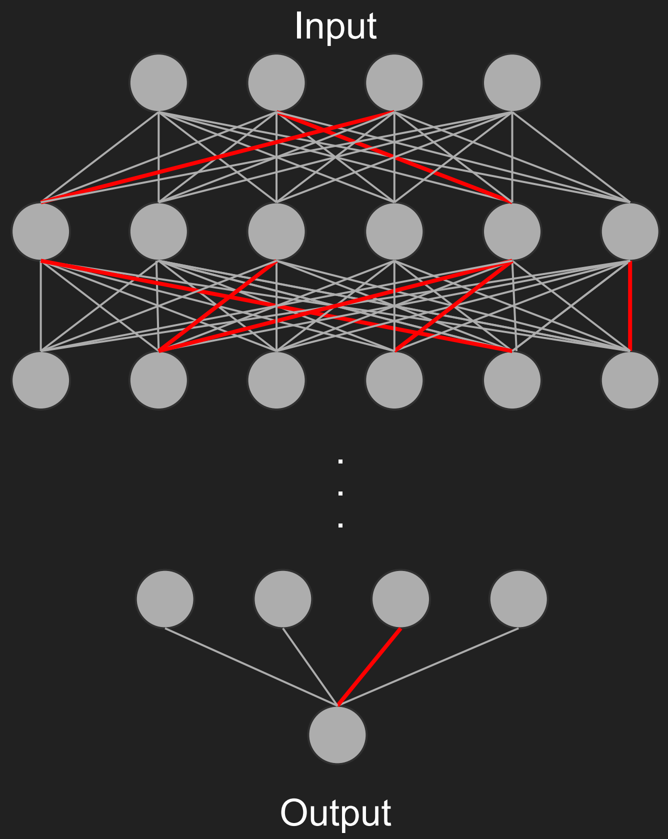
$$\mathbf{m} := \text{Cluster mass}$$
$$\mathbf{x} := \text{Input observables}$$
$$\boldsymbol{\eta} := \text{Model architecture}$$
$$\mathcal{D} := \text{Training data}$$
$$\boldsymbol{\theta} := \text{Model weights}$$

A Very Brief Guide to Approximate BNNs

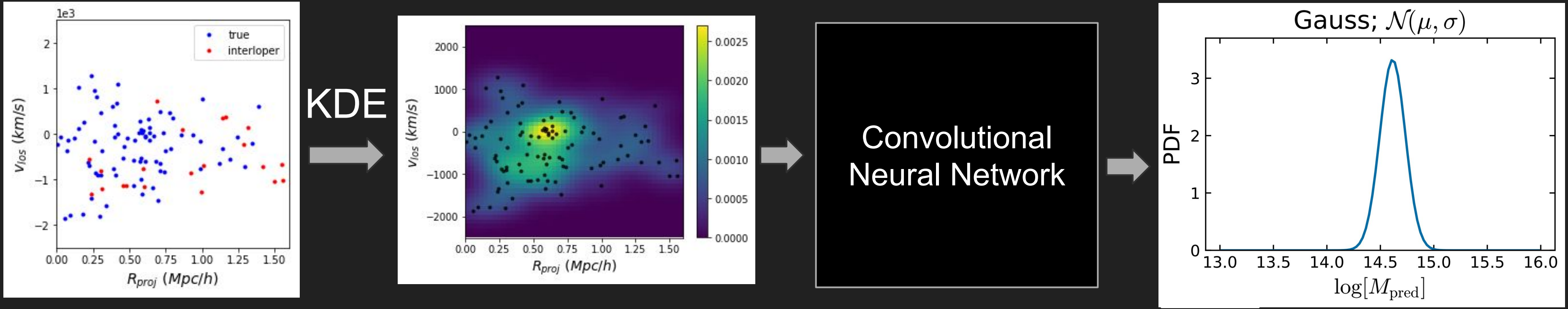
$$\theta \sim \theta_0 \cdot \mathrm{Bernoulli}(p)$$

Dropout Variational Inference (Gal & Ghahramani 2016)

- Randomly set some fraction *p* of weights to 0 during both training and inference
- Evaluate many random realizations, then average their outputs



Approximate Bayesian Uncertainties on Deep Learning Mass Estimates (Ho et al. 2020)



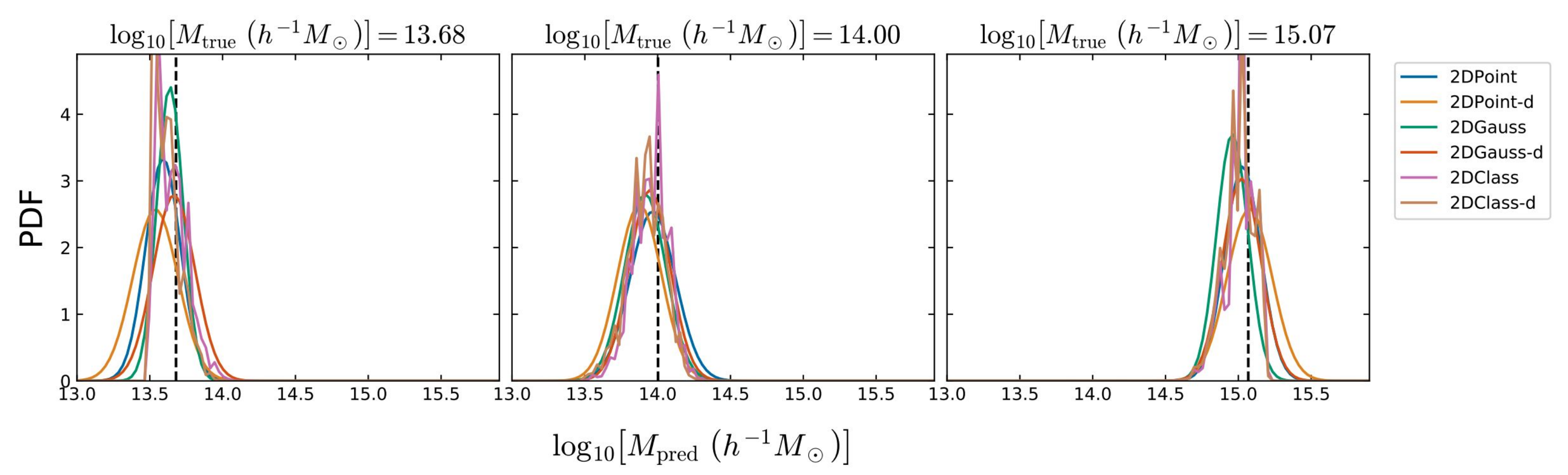
$$p\left(m \mid \mathbf{x}, \theta, \eta\right) \longrightarrow \text{Normal, Categorical}$$

$$p(\theta \mid \eta, \mathcal{D}) \longrightarrow \theta \sim \theta_0 \cdot \mathrm{Bernoulli}(p)$$

Same catalog and train/test procedures as original paper

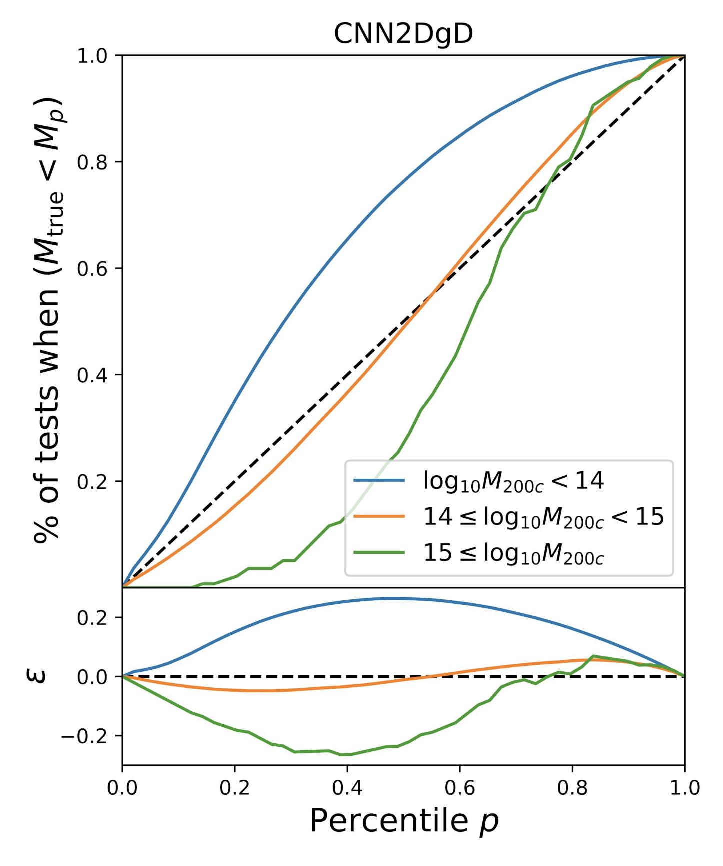
Approximate Bayesian Uncertainties on Deep Learning Mass Estimates (Ho et al. 2020)

- Model posteriors are Gaussian, even when given high flexibility. They are consistent with true cluster masses, with low predictive scatter and bias for median predictions.



Approximate Bayesian Uncertainties on Deep Learning Mass Estimates (Ho et al. 2020)

- Model posteriors are well calibrated for mid-range mass clusters. The best performing models can recover within +/- 1% of 64 and 90 percentile confidence intervals.
- Slight biases exist for very high/low mass clusters at the edges of our training set
- Epistemic uncertainties don't necessarily improve our posterior calibration



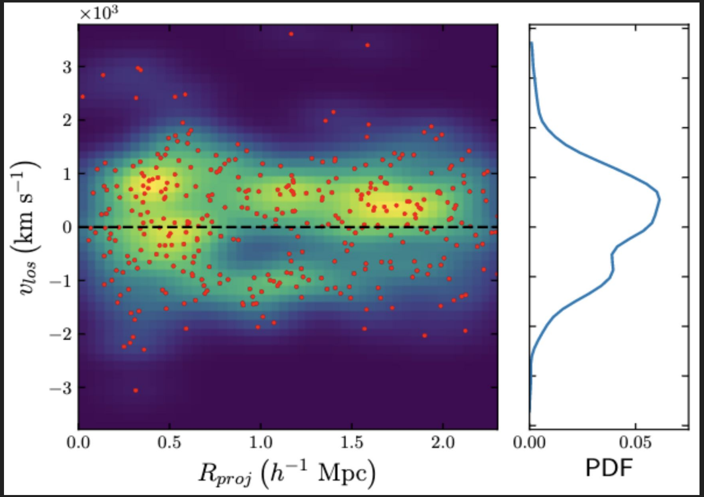
Empirical Verification

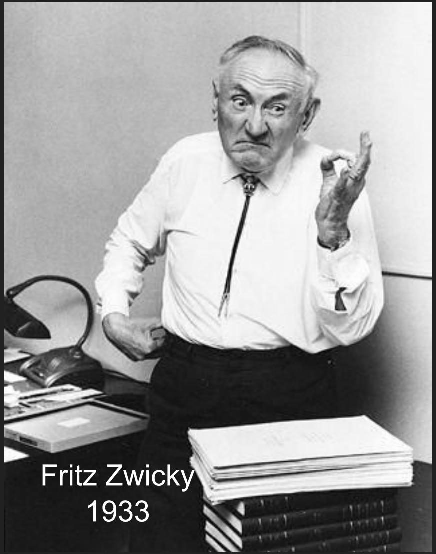
Application to Observation - Coma Cluster

Validate our model prediction on well-studied systems

~100 galaxy spectra from SDSS above $M_{\text{stellar}} \geq 10^{10.5}\ h^{-1} M_{\odot}$

Coma Cluster,
Schulman Telescope



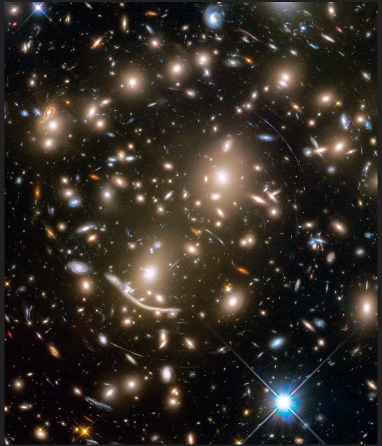


Fritz Zwicky
1933

Multiwavelength Probes of Dark Matter

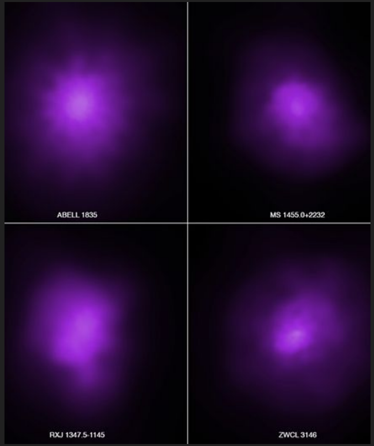
Multiwavelength Measurements

Optical



Number of galaxies above fixed luminosity

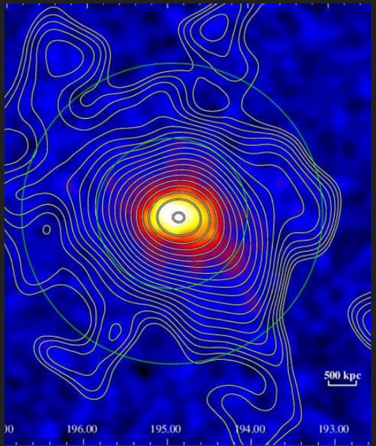
X-ray



Hot, photon-emitting intracluster gas

(Ntampaka et al. 2019)

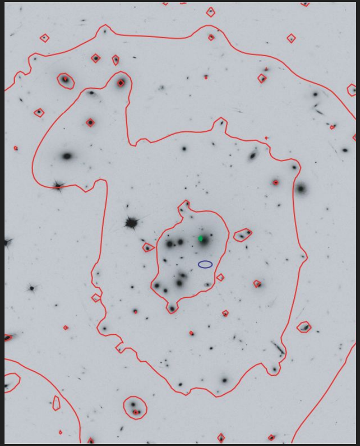
Microwave



Scattering of CMB photons on intracluster plasma

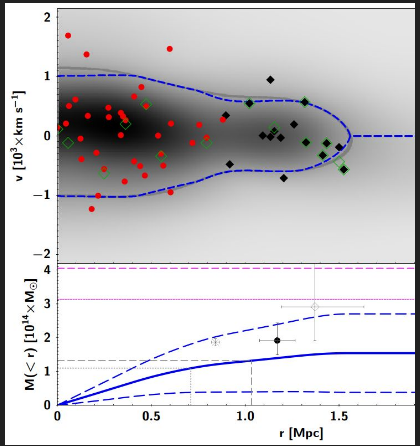
(Gupta et al. 2020a, 2020b)

Multi-band



Lensing of background light

Spectroscopic



Galaxy dynamics

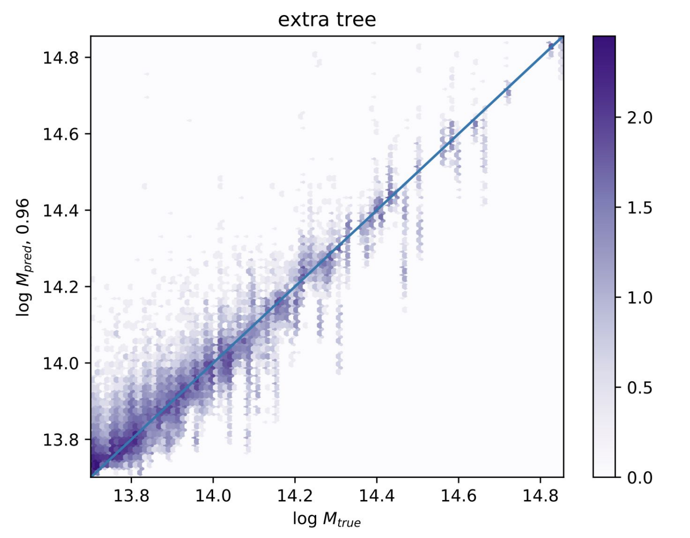
(Ho et al. 2019, 2020, 2021)
(Ramanah et al. 2020a, 2020b)

Multiwavelength Feature Analysis

Combining mass proxies using machine learning tools (LR, RFs, DTs, k-NN):

Feature set	Included Features
PHOT:	$R_{\text{mean, std, skew, kurt}}$, $m_{\text{mean, std, skew, kurt}}$, N_{gal}, e
SPEC:	PHOT + $v_{\text{mean, std, skew, kurt}}$
X:	T_{500c}, $L_{X,500c}$
SZ:	$Y_{SZ,5r_{500c}}$

M_{FOF} scatter of ~0.039 dex (Cohn). M_{200c} scatter of ~0.03 dex (Armitage)

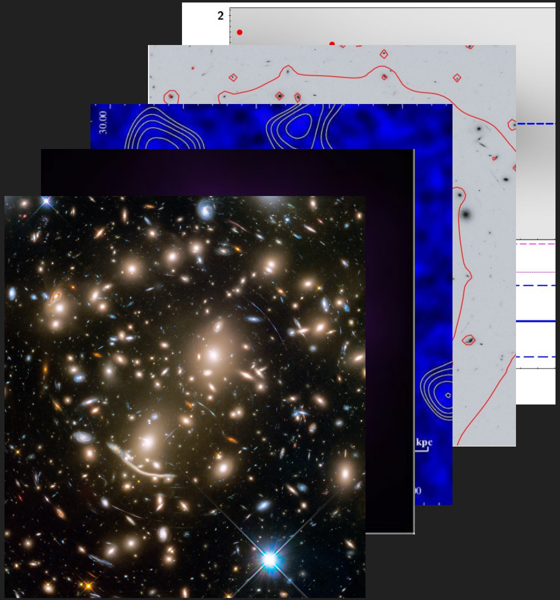


Joanne Cohn (LBNL)
arXiv:1905.09920

Thomas J. Armitage (Manchester)
arXiv:1810.08430

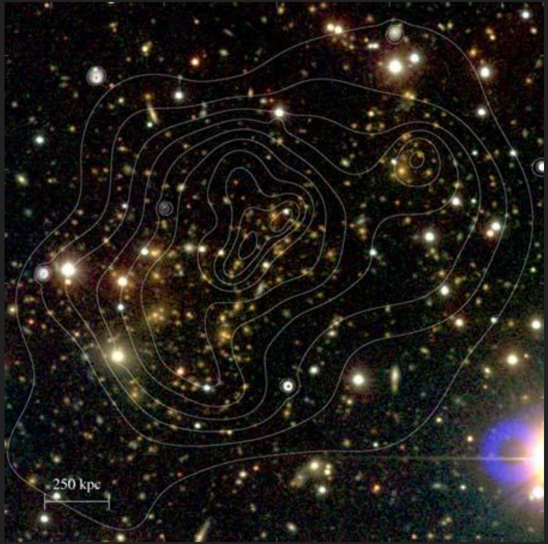
Multi-Wavelength Cluster Measurements

Multi-wavelength observations

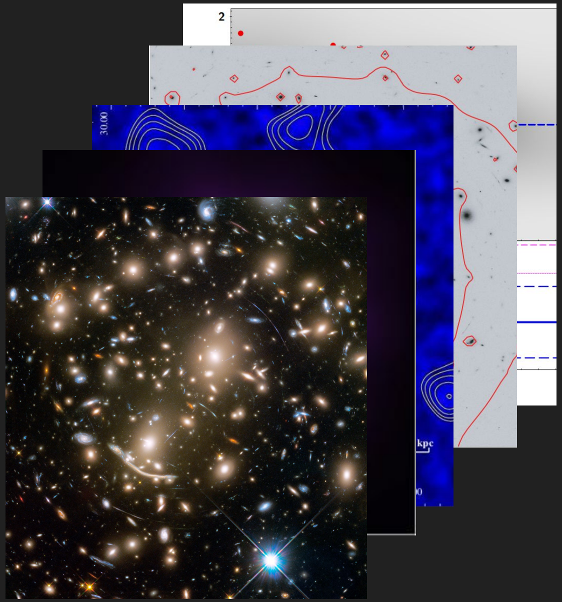


Deep Learning

Cluster mass distribution



Multi-Wavelength Cluster Measurements

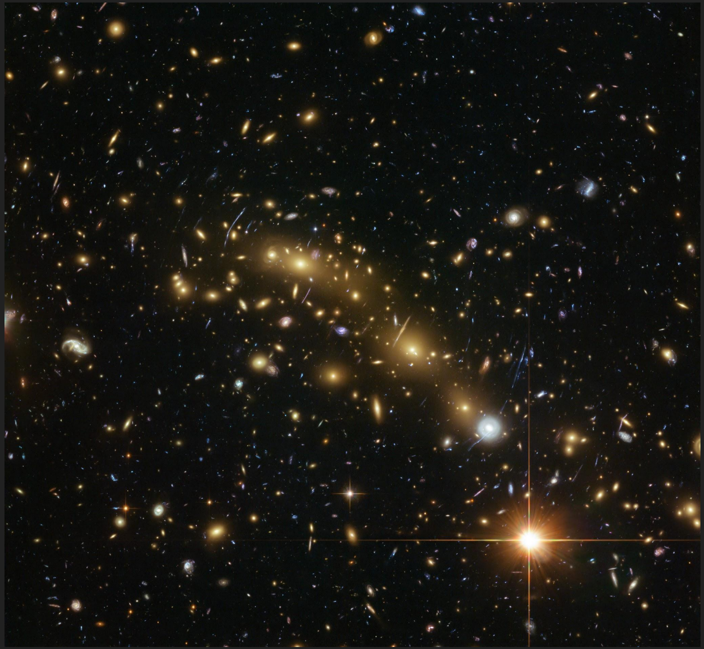


Deep Learning requires **Big Data**

- We are entering an era of large, high-resolution simulations
- Fast, large-volume hydrosims
  - HYPER (He et al. 2021)
- Advanced baryon pasting methods
  - Generative modeling
  - Subgrid pasting

Conclusion

- We introduced an image-recognition based model for calculating cluster masses from galaxy dynamics (Ho et al. 2019).
- Discussed methods for measuring uncertainties from deep learning (Ho et al. 2020).
- Described current applications on real systems such as the Coma, CLASH, and HeCS clusters.
- Detailed attempts toward fully-informed multiwavelength mass estimators.



MACS J0416.1-2403
Image Credit: Hubble

Matthew Ho ~ mho1@andrew.cmu.edu