

CONFRONTING THE THEORY OF GALAXY FORMATION WITH OBSERVATIONS USING GENERATIVE MODELS

M. Huertas-Company

Instituto de Astrofísica de Canarias
Observatoire de Paris - Université de Paris

Collaborators:

Primack, Dekel,
Lanusse, Charnock, Zanisi, Dubois, Margalef-Bentabol,
Dominguez-Sanchez, Bernardi ...



Ringberg Castle - December 10 2019

DOES A NEURAL NETWORK KNOW ABOUT HORSES IF IT HAS ONLY SEEN CATS AND DOGS?

M. Huertas-Company

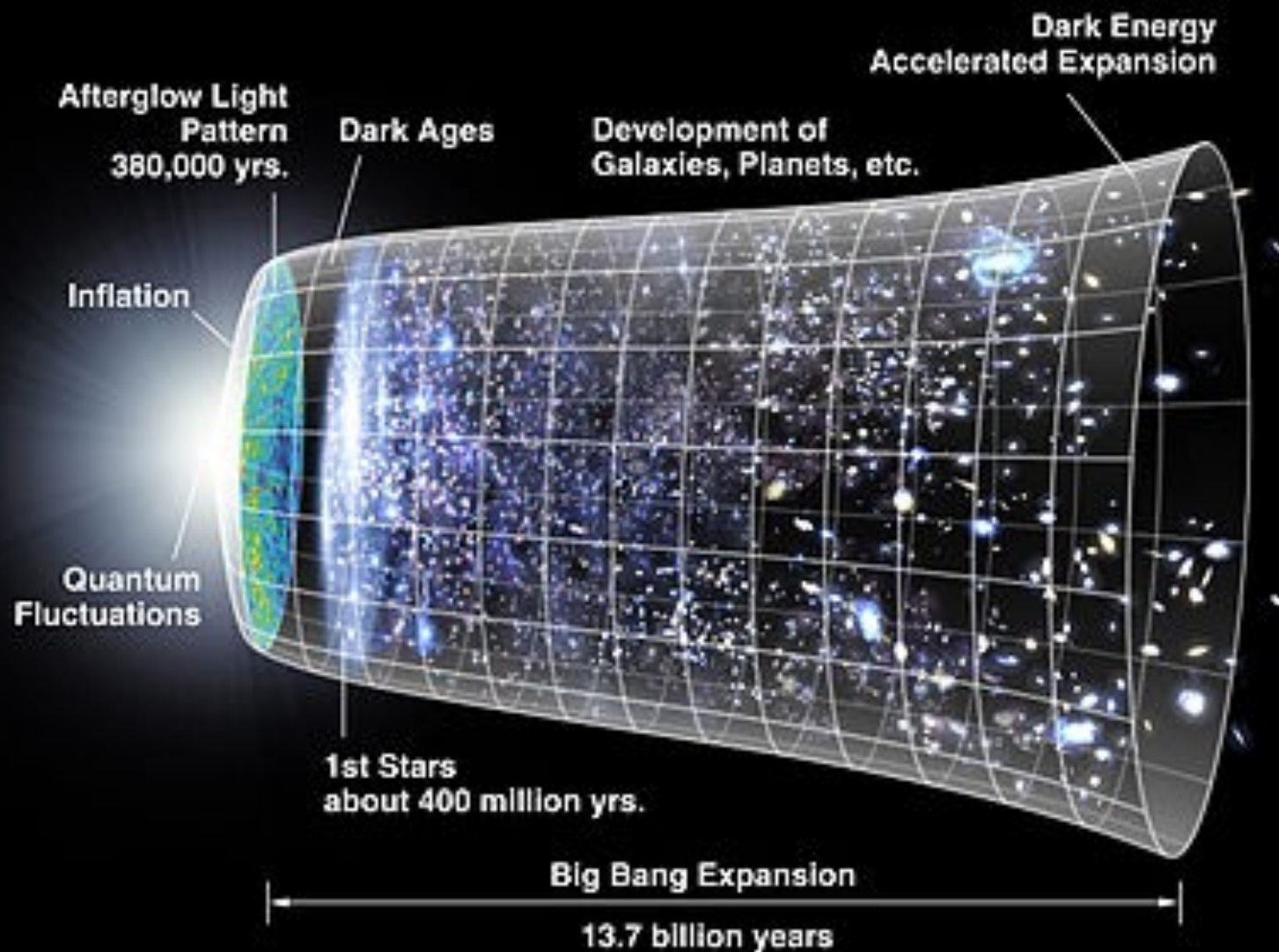
Instituto de Astrofísica de Canarias
Observatoire de Paris - Université de Paris

Collaborators:

Primack, Dekel,
Lanusse, Charnock, Zanisi, Dubois, Margalef-Bentabol,
Dominguez-Sanchez, Bernardi ...

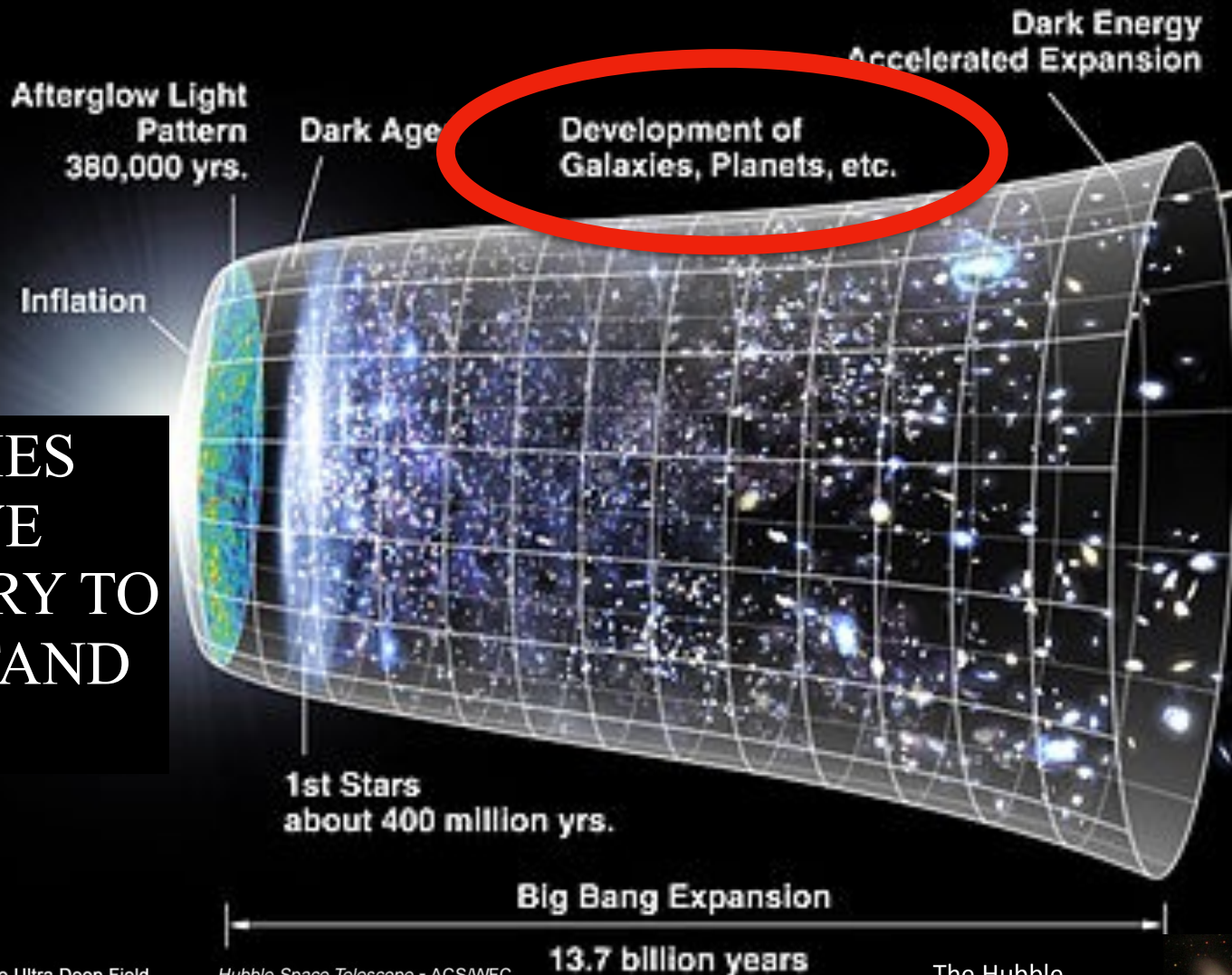


Ringberg Castle - December 10 2019



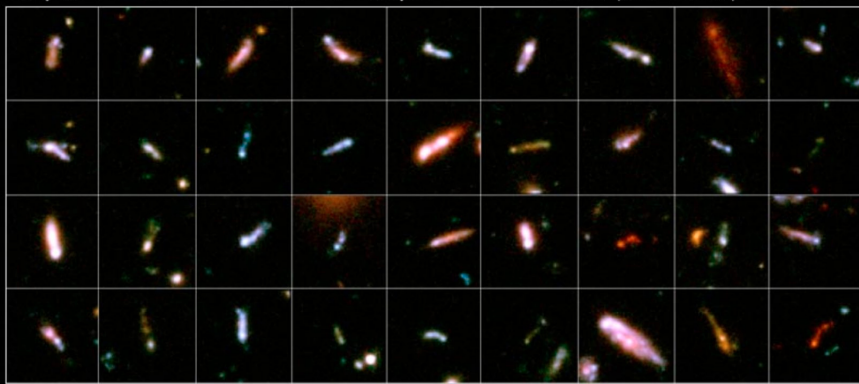
THIS IS OUR
CURRENT
UNDERSTANDING OF
THE HISTORY
OF THE UNIVERSE

GALAXIES EVOLVE AND WE TRY TO UNDERSTAND HOW



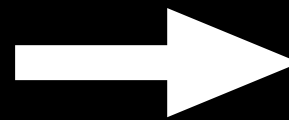
"Tadpole" Galaxies in the Hubble Ultra Deep Field

Hubble Space Telescope • ACS/WFC



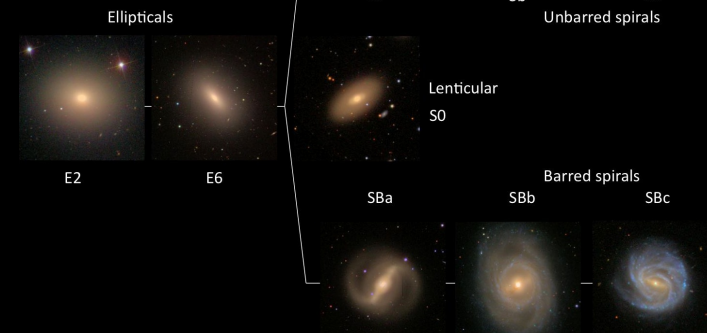
NASA, ESA, A. Straughn, S. Cohen and R. Windhorst (Arizona State University), and the HUDF team (STScI)

STScI-PRC06-04

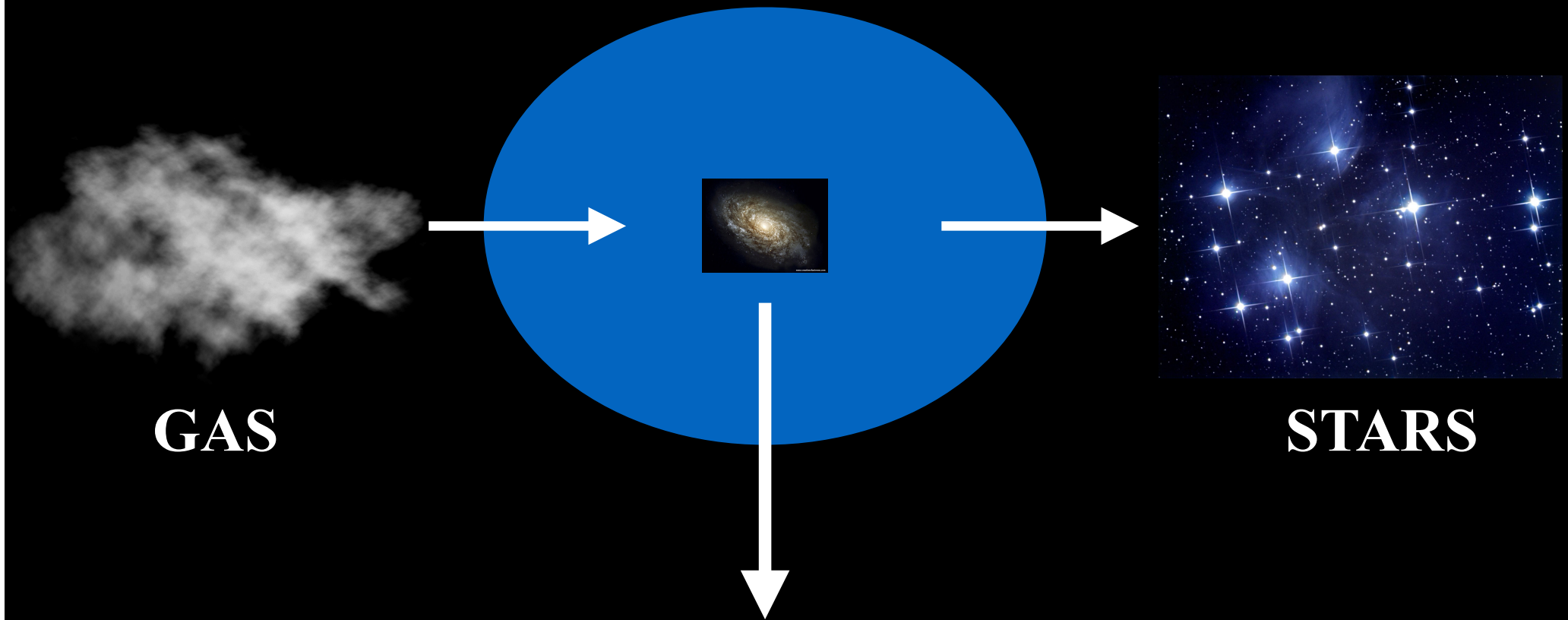


?

The Hubble Tuning Fork



GALAXIES ARE “EFFICIENT” MACHINES TO MAKE STARS

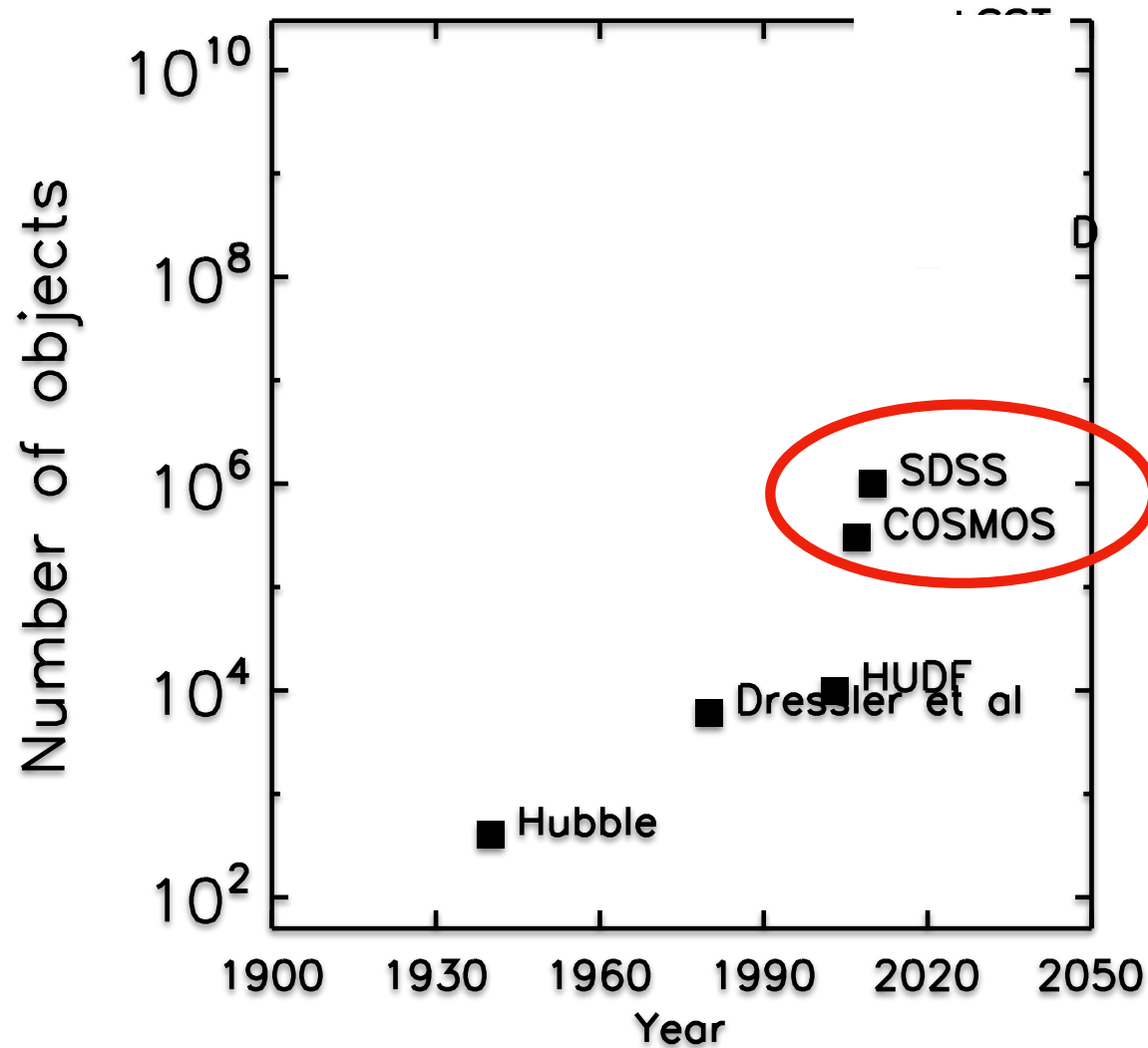


GAS

STARS

COMPLEX, NON-LINEAR PHYSICS

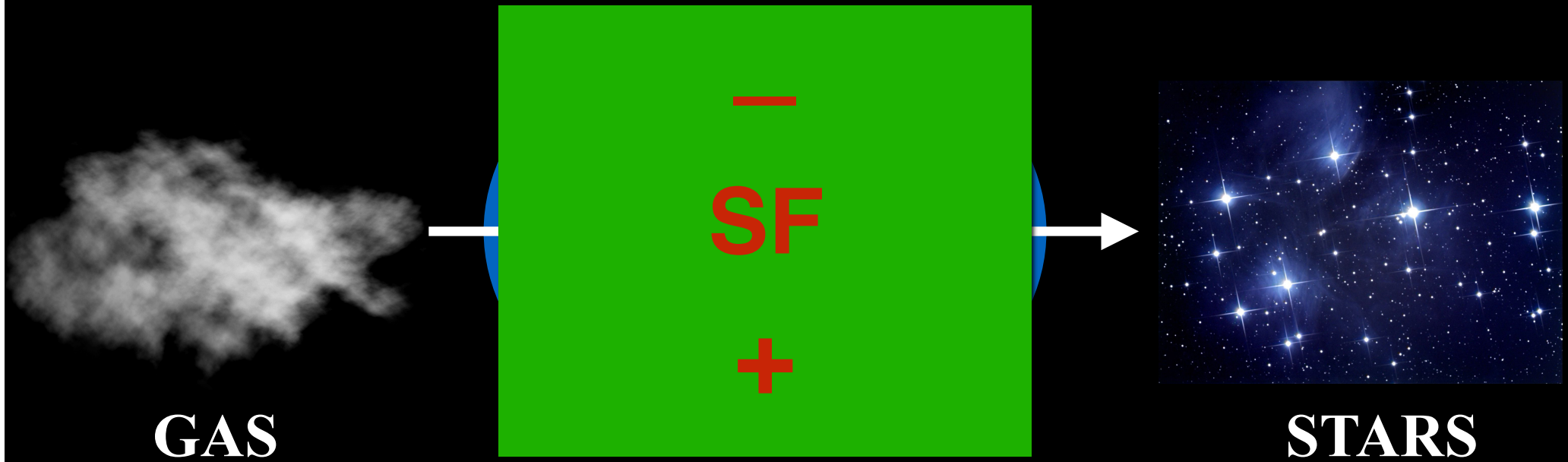
THE ERA OF STATISTICS



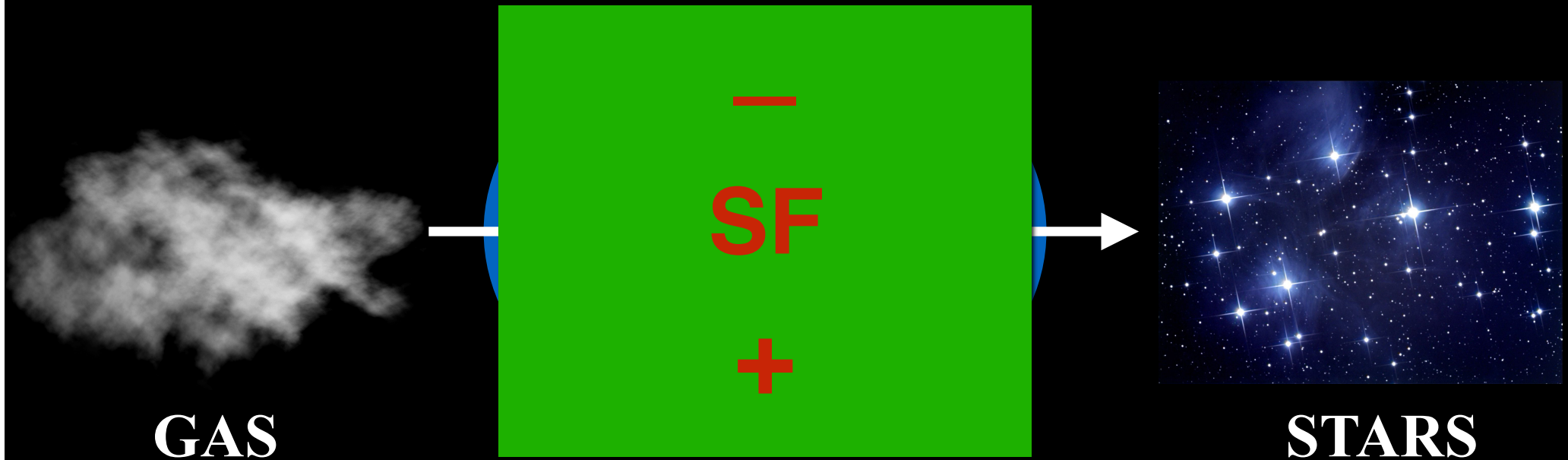
IN THE LAST
~20 YEARS

**THE FIELD OF
OBSERVATIONAL
GALAXY EVOLUTION
HAS EVOLVED INTO A
STATISTICAL SCIENCE**

GALAXIES ARE “EFFICIENT” MACHINES TO MAKE STARS

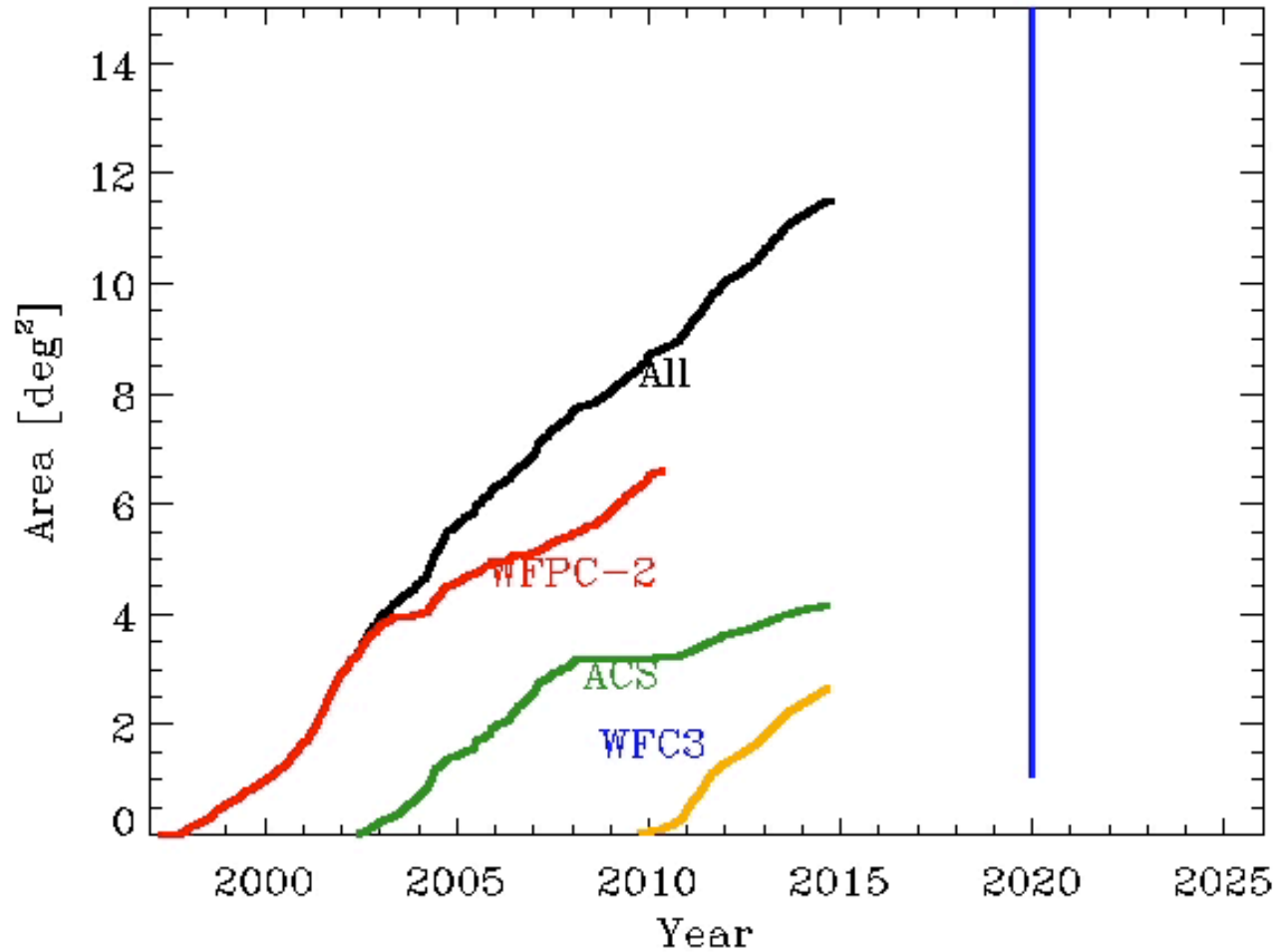


GALAXIES ARE “EFFICIENT” MACHINES TO MAKE STARS

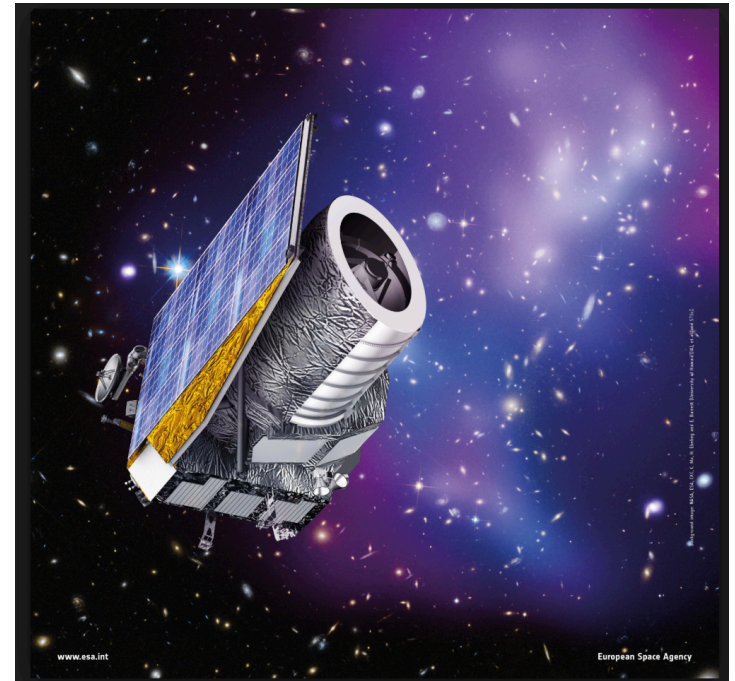


**GAS INFLOWS
OUTFLOWS
INTERACTIONS
BLACK HOLES
SUPERNOVAE
DISK INSTABILITIES
ETC..**

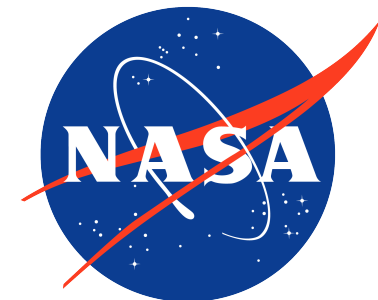
THE BIG-DATA ERA



(Thanks to J. Brinchmann)

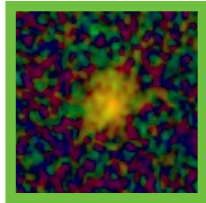


EUCLID space telescope
(2021)



WE "MIGHT KNOW" THE INGREDIENTS, WE DON'T KNOW THE RECIPE

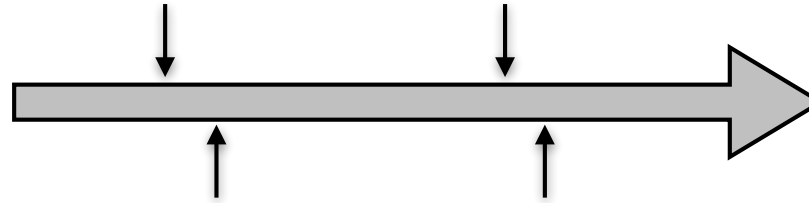
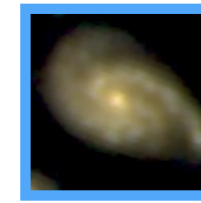
11 Gyrs ago



(Toomre+77)
mergers?

(Silk+98)
feedback?

today



clumps?
(Bournaud+07)

accretion?
(Dekel+09)

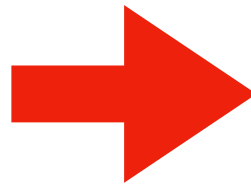
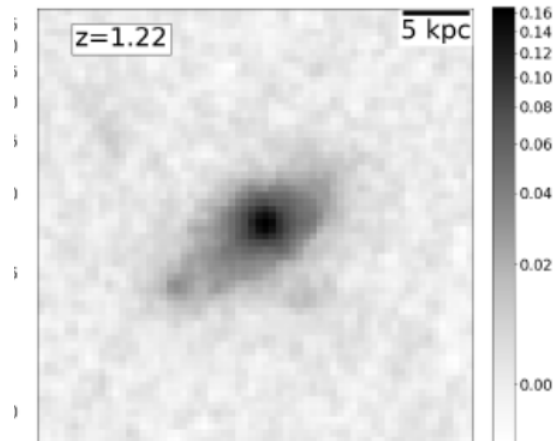
irregular (MHC+15,16), turbulent (Genzel+13), rotating (Wisnioski+15), compact (Trujillo,+06 Shibuya+15), high star-formation efficiency (Daddi+07), gas-rich (Genzel+15)

hubble sequence bulges/disks (Hubble+36), bimodal SF/passive (Kauffmann+03)

WHAT IS THE HISTORY OF MASS ASSEMBLY ?

WHEN / HOW IS FEEDBACK TRIGGERED?

AN OBSERVATION IS A SINGLE “SNAPSHOT” IN TIME



WAS THERE A MERGER? WHEN?

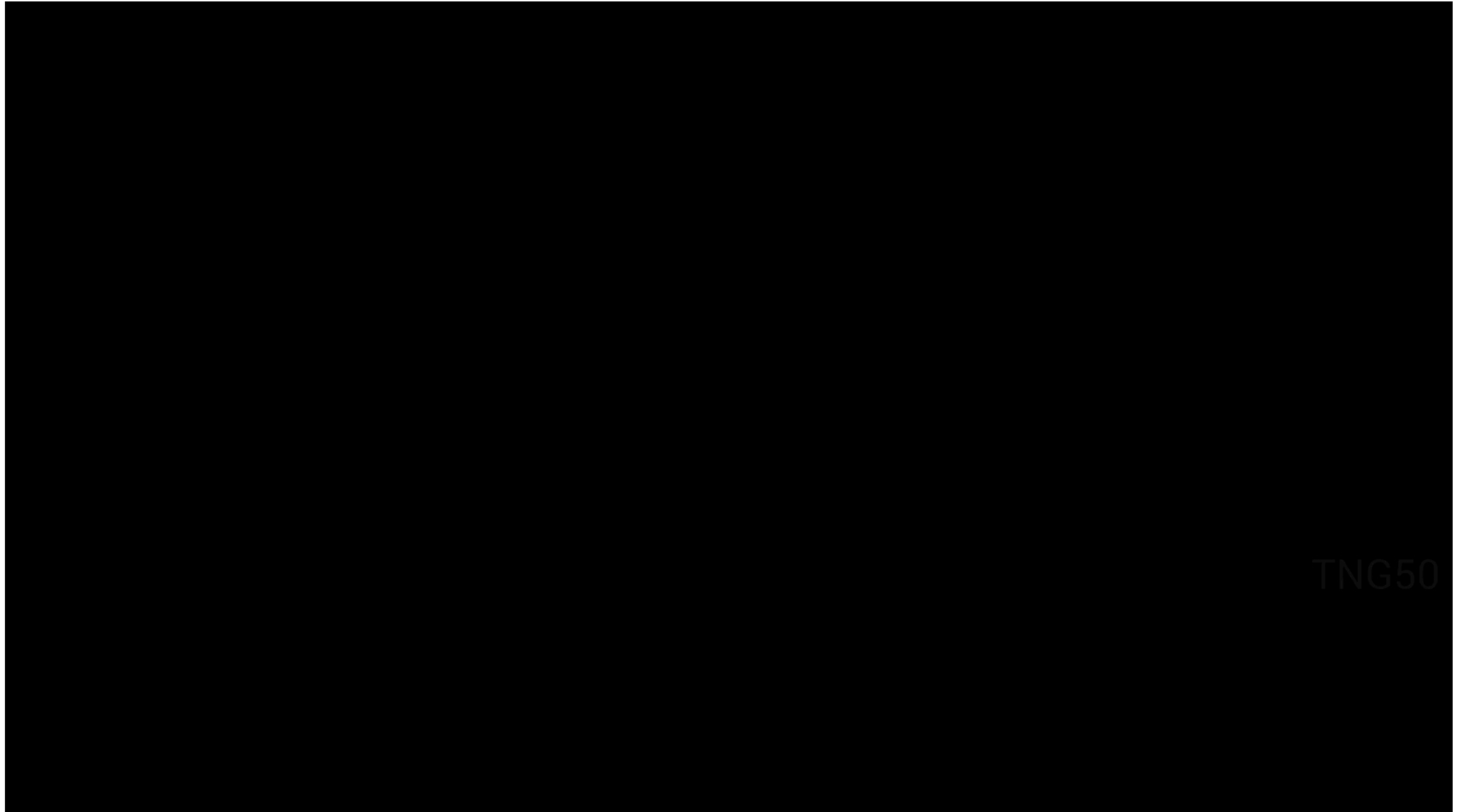
DID IT HAVE AN ACTIVE BLACK HOLE?

WHEN/HOW WAS THE BULGE FORMED?

GIVEN AN OBSERVATION...

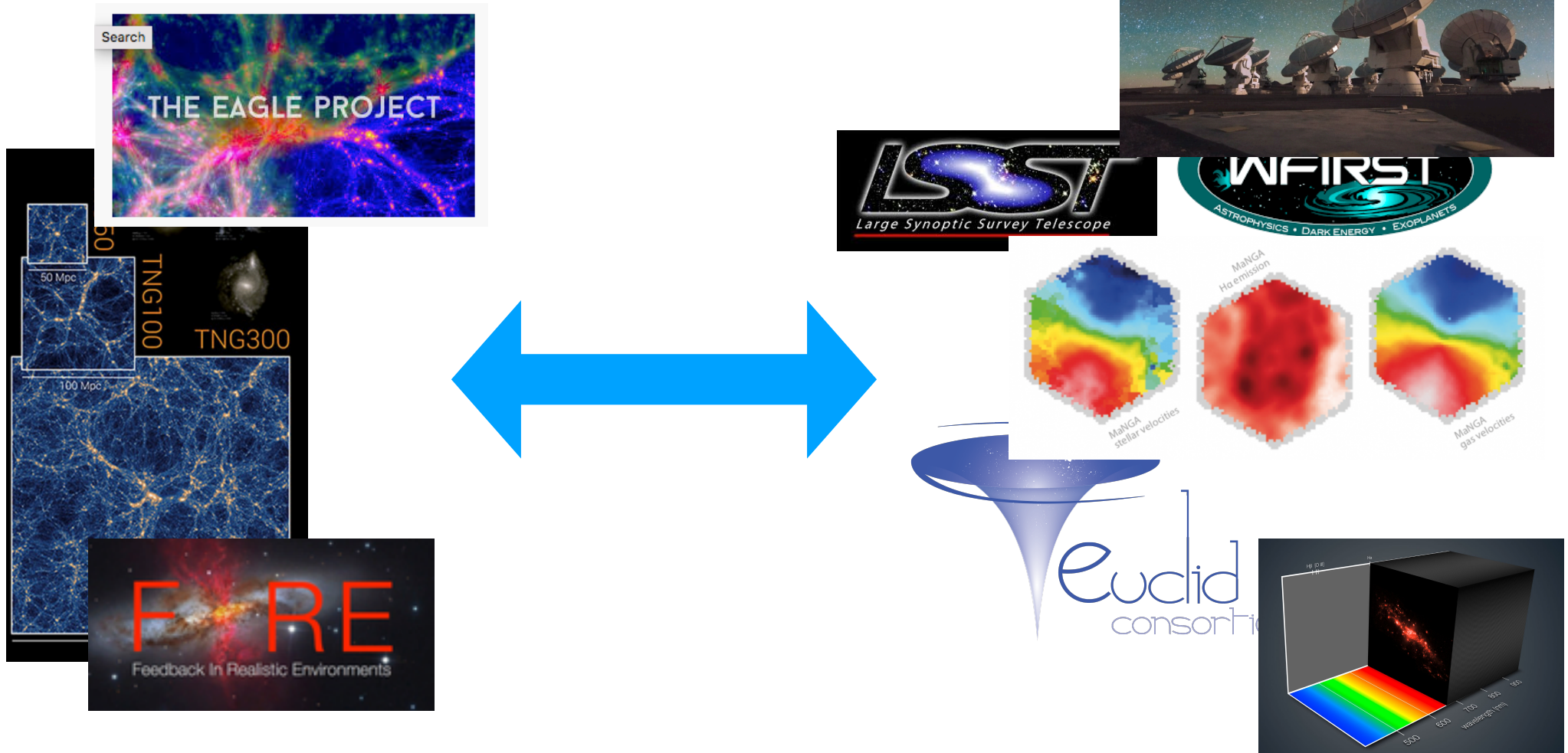
CAN WE INFER ITS
ASSEMBLY HISTORY?

GALAXIES EVOLVE AND GO THROUGH DIFFERENT PHASES DURING THEIR LIFETIME. THIS IS CAPTURED IN NUMERICAL SIMULATIONS



ILLUSTRIS-TNG SIMULATION: <http://www.tng-project.org>

AS SIMULATIONS AND OBSERVATIONS BECOME LARGER AND MORE COMPLEX....



...THERE IS A NEED OF MEANINGFUL WAYS OF LINKING THE
TWO

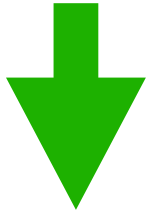
**THEORY /
SIMULATIONS**



**Illustris, EAGLE,
Horizon-AGN ...**

ML TO LINK THEORY AND OBSERVATION IN THE DATA SPACE

[FULL 3D EVOLUTION HISTORY]



**PROJECT HYDRO
SIMS IN
THE
“OBSERVATIONAL
PLANE”**

**ASSUMPTIONS OF MASS
TO LIGHT CONVERSION**

**+ DUST
+ PSF
+ NOISE**

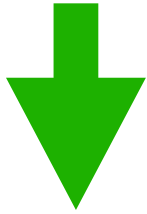
**THEORY /
SIMULATIONS**



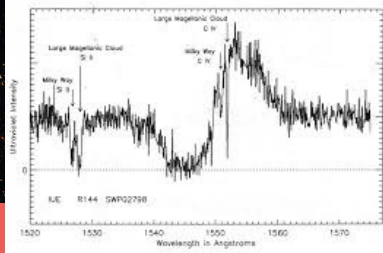
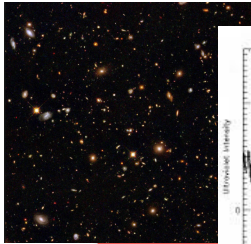
**Illustris, EAGLE,
Horizon-AGN ...**

ML TO LINK THEORY AND OBSERVATION IN THE DATA SPACE

[FULL 3D EVOLUTION HISTORY]



**PROJECT HYDRO
SIMS IN
THE
“OBSERVATIONAL
PLANE”**



**MOCK
OBSERVATIONS**



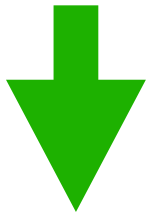
**THEORY /
SIMULATIONS**



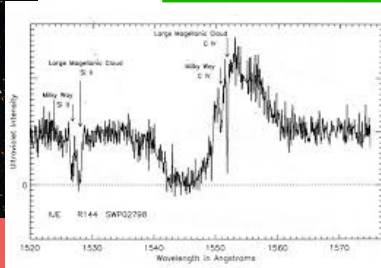
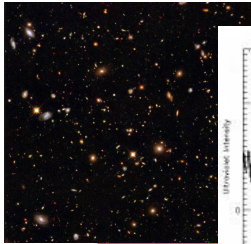
**Illustris, EAGLE,
Horizon-AGN ...**

ML TO LINK THEORY AND OBSERVATION IN THE DATA SPACE

[FULL 3D EVOLUTION HISTORY]

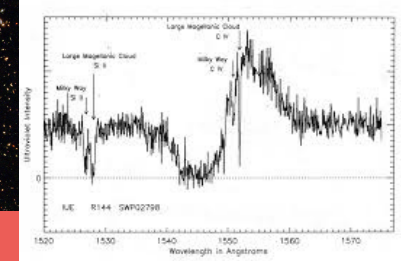
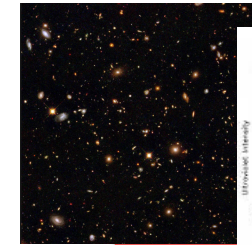


**PROJECT HYDRO
SIMS IN
THE
“OBSERVATIONAL
PLANE”**



TRAIN

TEST



**MOCK
OBSERVATIONS**

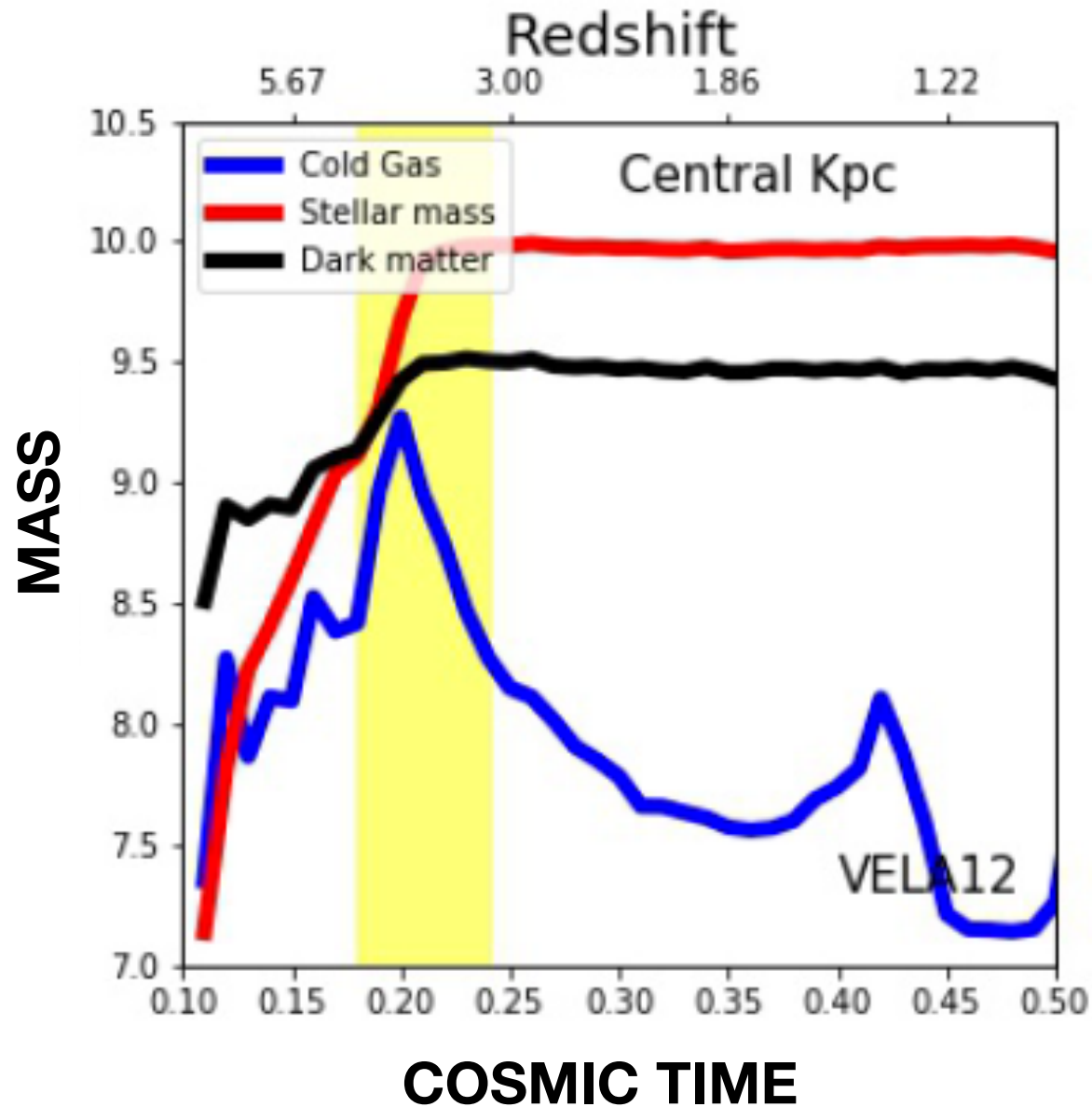


OBSERVATIONS



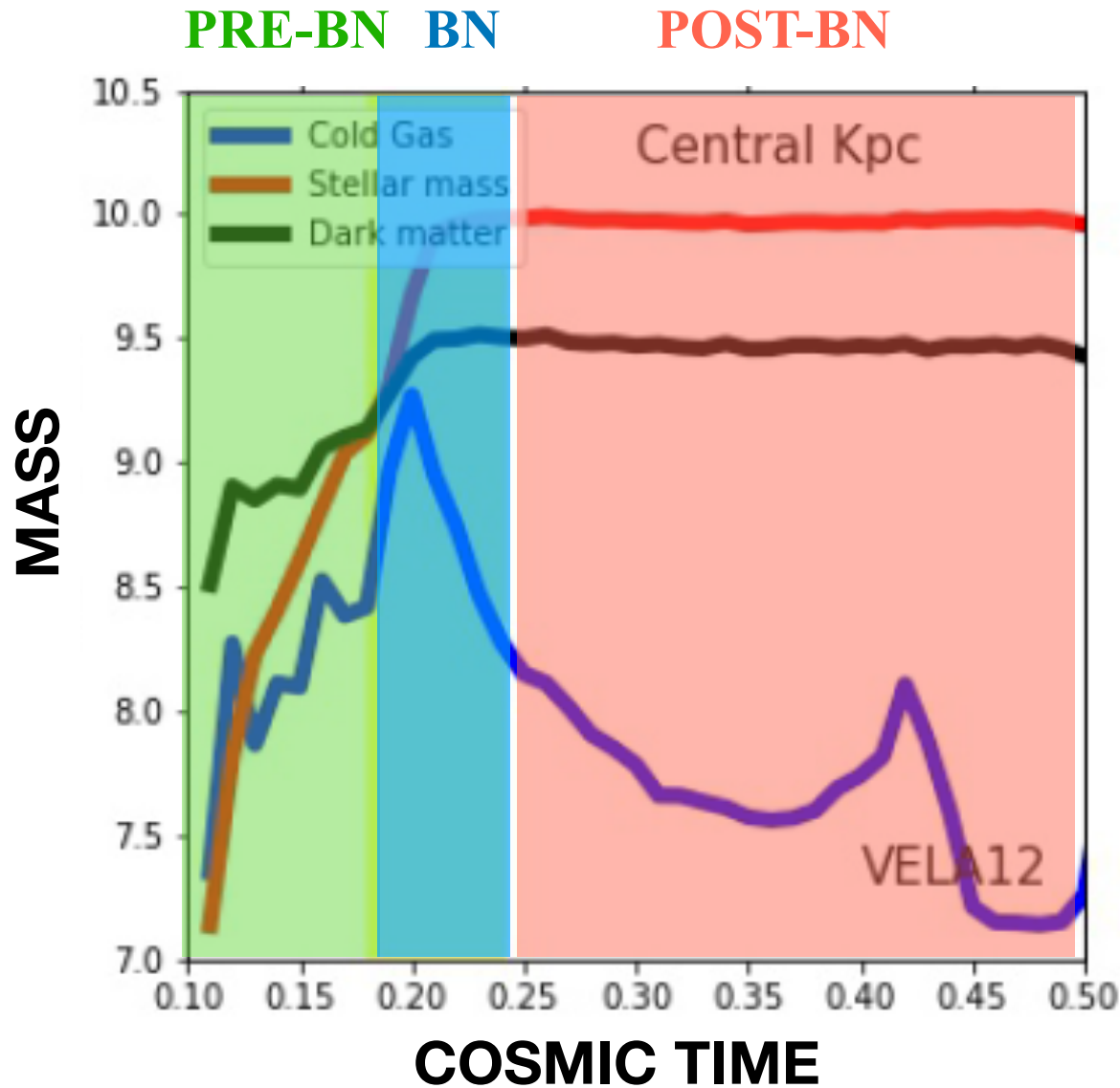
**MACHINE (DEEP)
LEARNING**

GALAXIES EVOLVE AND GO THROUGH DIFFERENT PHASES DURING THEIR LIFETIME. THIS IS CAPTURED IN NUMERICAL SIMULATIONS.



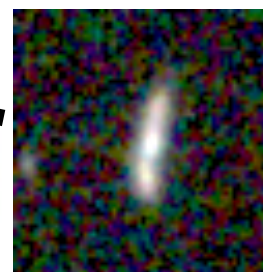
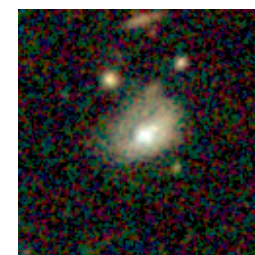
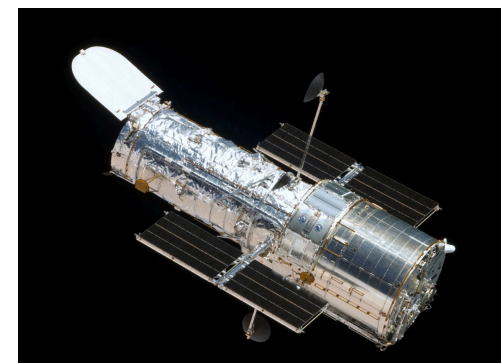
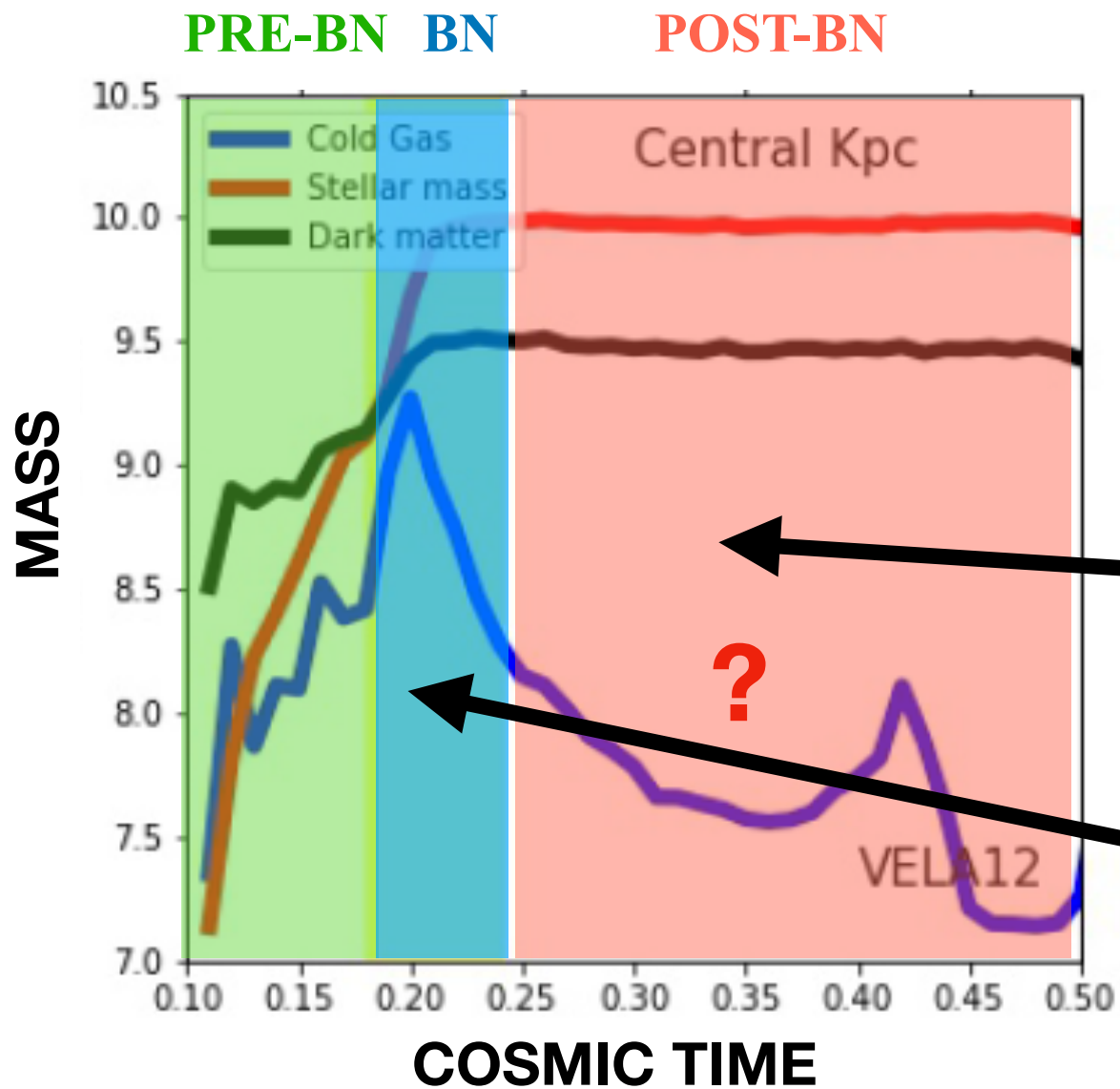
**VELA suite of hydro simulations
[Ceverino, Dekel+]**

GALAXIES EVOLVE AND GO THROUGH DIFFERENT PHASES DURING THEIR LIFETIME. THIS IS CAPTURED IN NUMERICAL SIMULATIONS.



**LABELLING
EXCLUSIVELY
BASED ON
THE FULL HISTORY
OF THE GALAXY
FROM ADVANCED
NUMERICAL
SIMULATIONS
BASED ON PHYSICAL
PRINCIPLES**

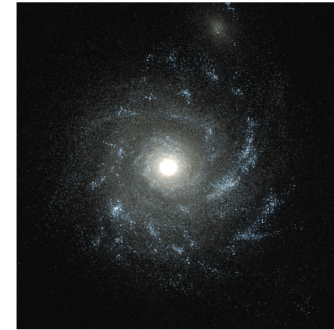
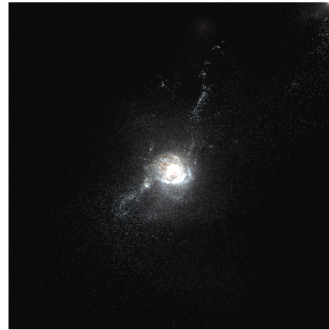
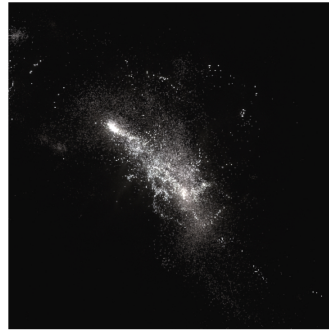
HOW CAN WE ESTIMATE THE PHASE FROM A UNIQUE IMAGE?



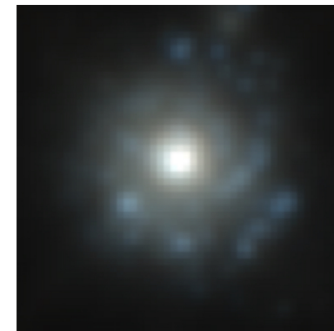
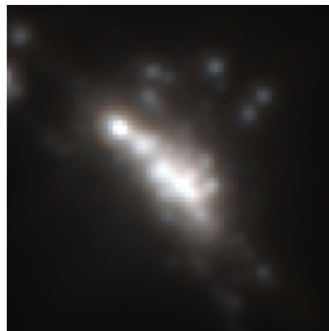
Pre-Blue-Nugget-Stage

Blue-Nugget-Stage

Post-Blue-Nugget-Stage



Images derived from computer simulations of one distant young Galaxy going through the 3 phases at a maximum resolution



Same images derived from computer simulations of a distant young Galaxy as it would have been observed by the Hubble Space Telescope



Hubble Space Telescope images of distant young Galaxies classified with a deep learning algorithm

DEEP LEARNING LINKS THEORY AND OBSERVATIONS



16.000 light years
[1 light year = 5.6 trillion miles]

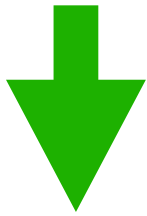
**THEORY /
SIMULATIONS**



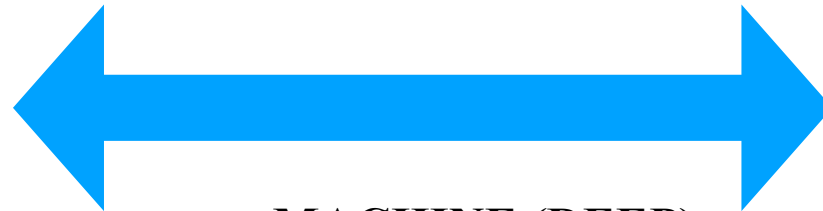
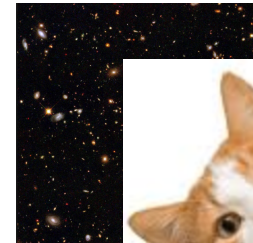
**Illustris, EAGLE,
Horizon-AGN ...**

WHAT IF OBSERVATIONS AND SIMULATIONS DO NOT COME FROM THE SAME DISTRIBUTIONS?

[FULL 3D EVOLUTION HISTORY]

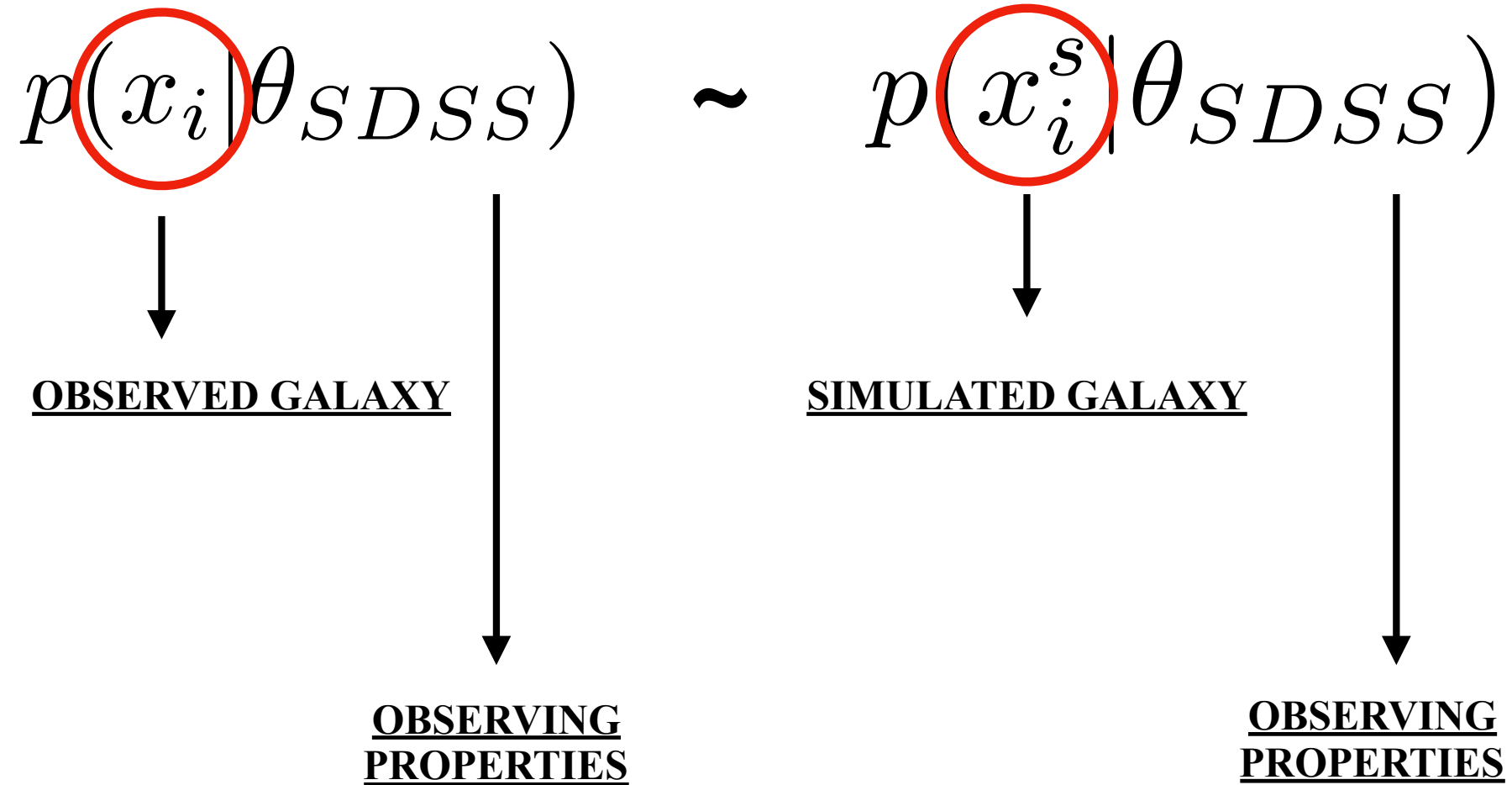


**FORWARD MODEL
HYDRO SIMS IN
THE
“OBSERVATIONAL
PLANE”**



**MACHINE (DEEP)
LEARNING**

WHAT DOES IT MEAN THAT SIMULATIONS “MATCH” OBSERVATIONS?



**WHAT DOES IT MEAN THAT SIMULATIONS “MATCH”
OBSERVATIONS?**

$$p(x_i | \theta_{SDSS}) \sim p(x_i^s | \theta_{SDSS})$$

\simeq

$$p(R_e | \theta_{SDSS})$$

$$p(SFR | \theta_{SDSS})$$

$$p(SFR, M_* | \theta_{SDSS})$$

$$p(R_e, M_* | \theta_{SDSS})$$

\simeq

$$p(R_e^s | \theta_{SDSS})$$

$$p(SFR^s | \theta_{SDSS})$$

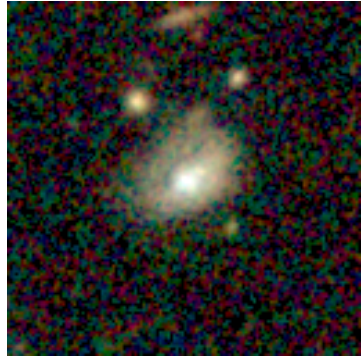
$$p(SFR^s, M_*^s | \theta_{SDSS})$$

$$p(R_e^s, M_*^s | \theta_{SDSS})$$

USING GENERATIVE MODELS TO ESTIMATE $P(X)$



x

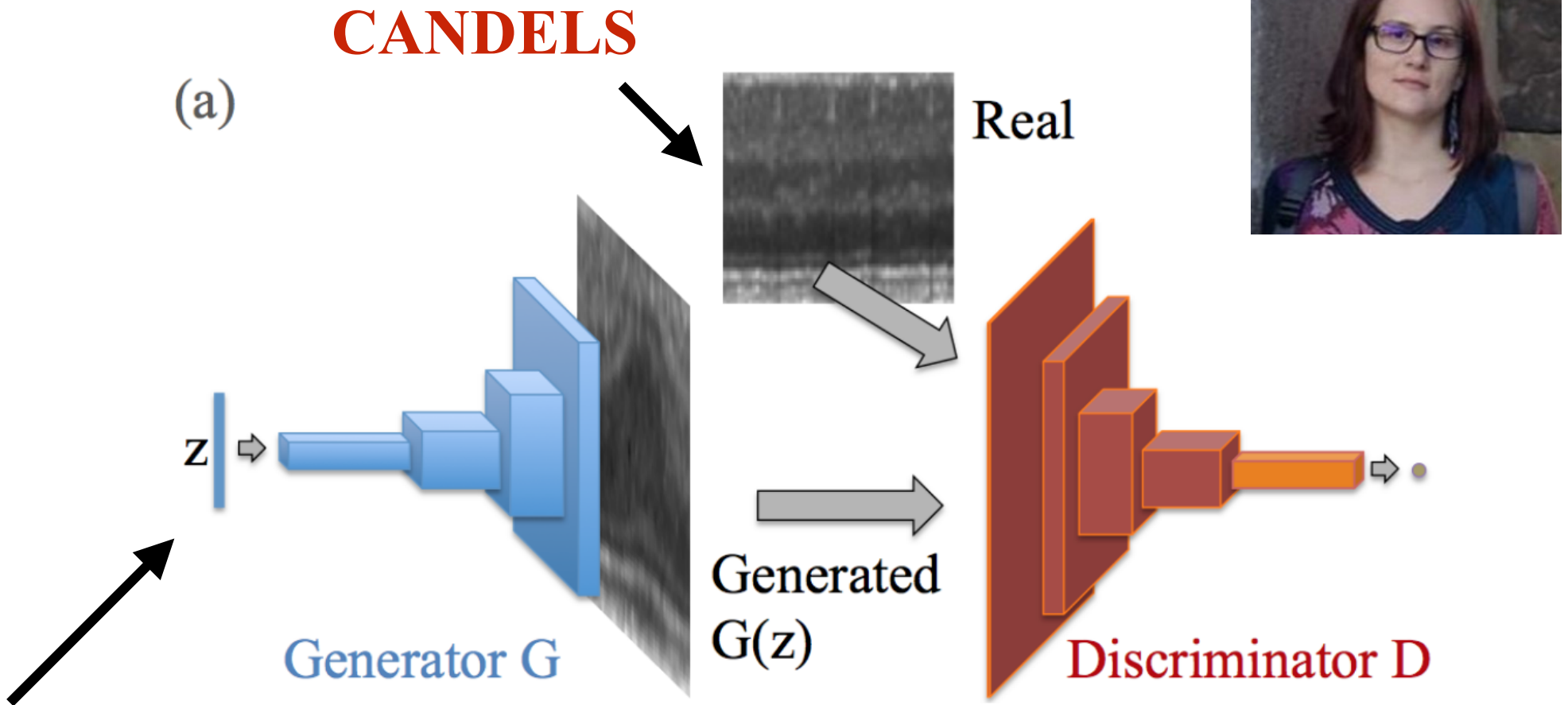


x'



$p(x)$

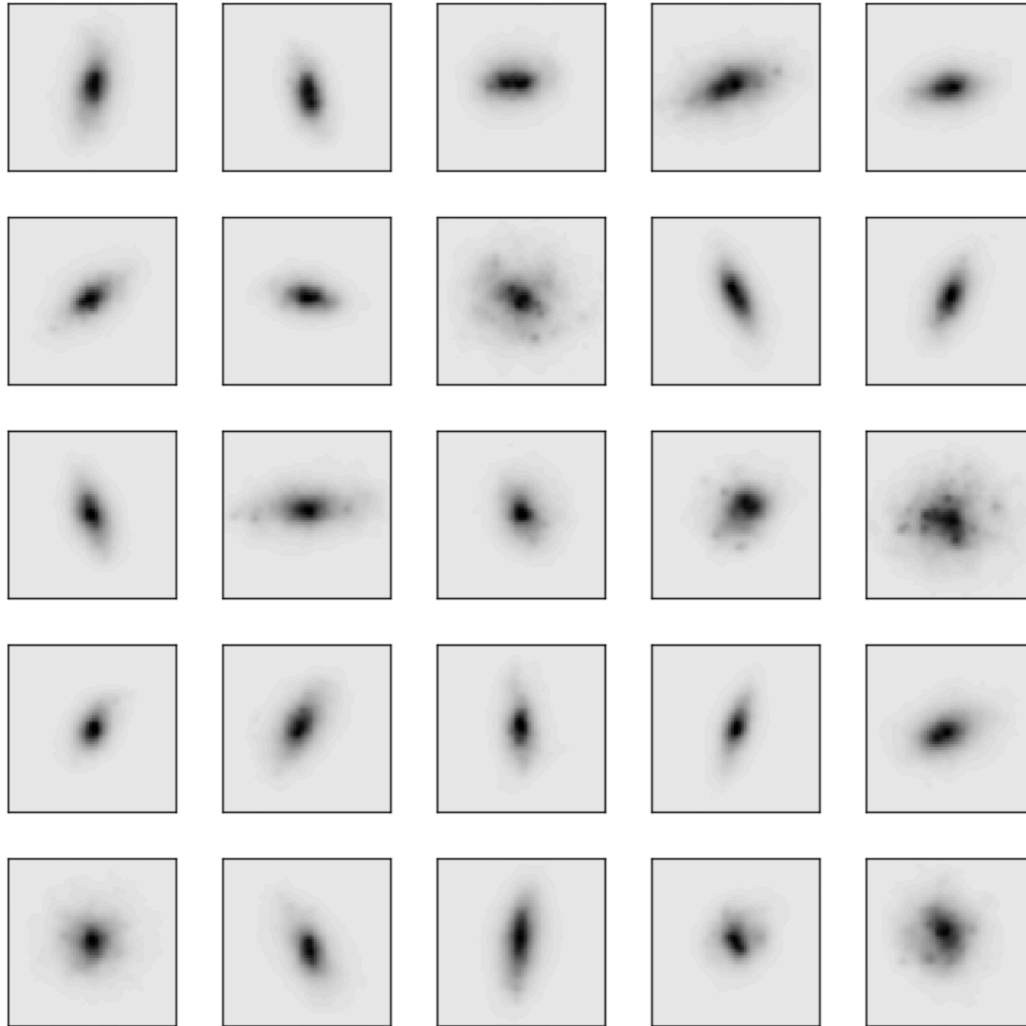
astroGANomaly: ANOMALY DETECTION WITH GANs



SIMULATIONS [HORIZON-AGN]

The Horizon Simulation

DUBOIS+14

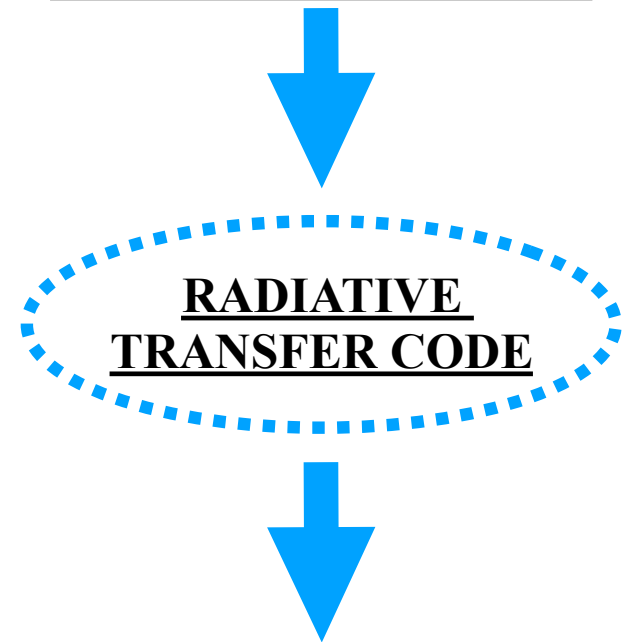


100 Mpc box, 1kpc res, RAMSES code

“NORMAL DATA”

ALL GALAXIES WITH:

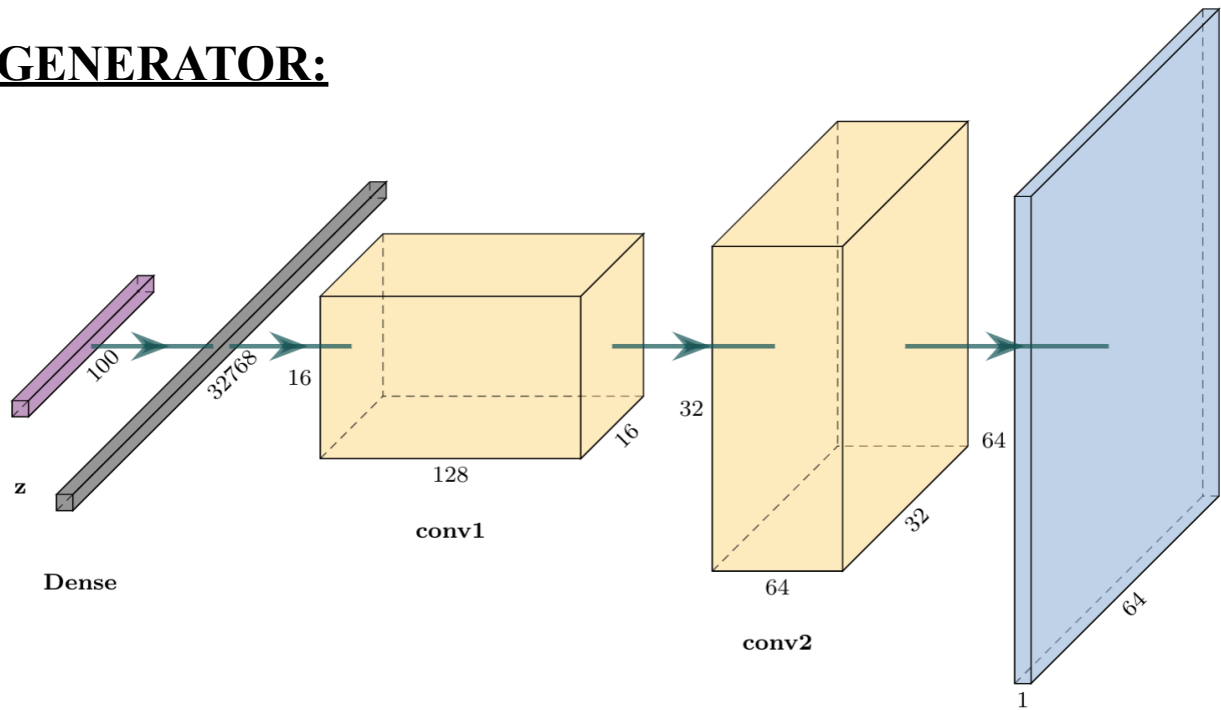
$0.5 < Z < 3$
 $\text{Log}(M^*) > 10$
~20 Myrs time steps



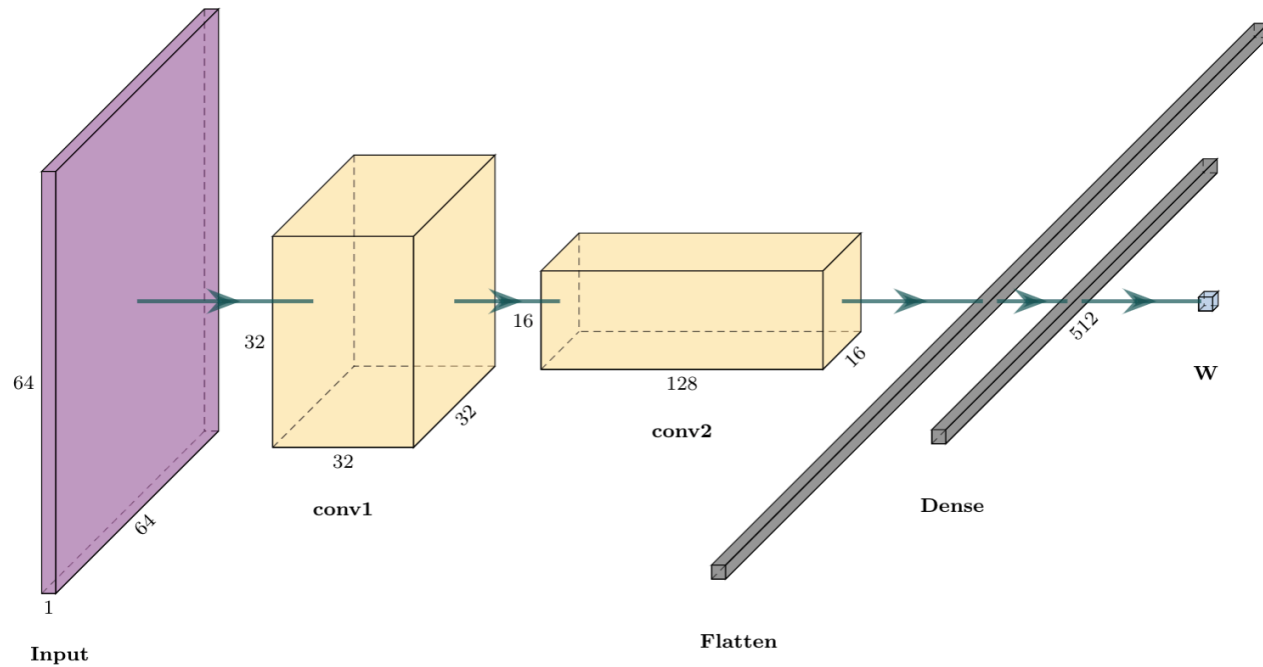
~1.5 million images

(HST RESOLUTION, H BAND)

GENERATOR:



CRITIC:



COMPUTE ANOMALY SCORE

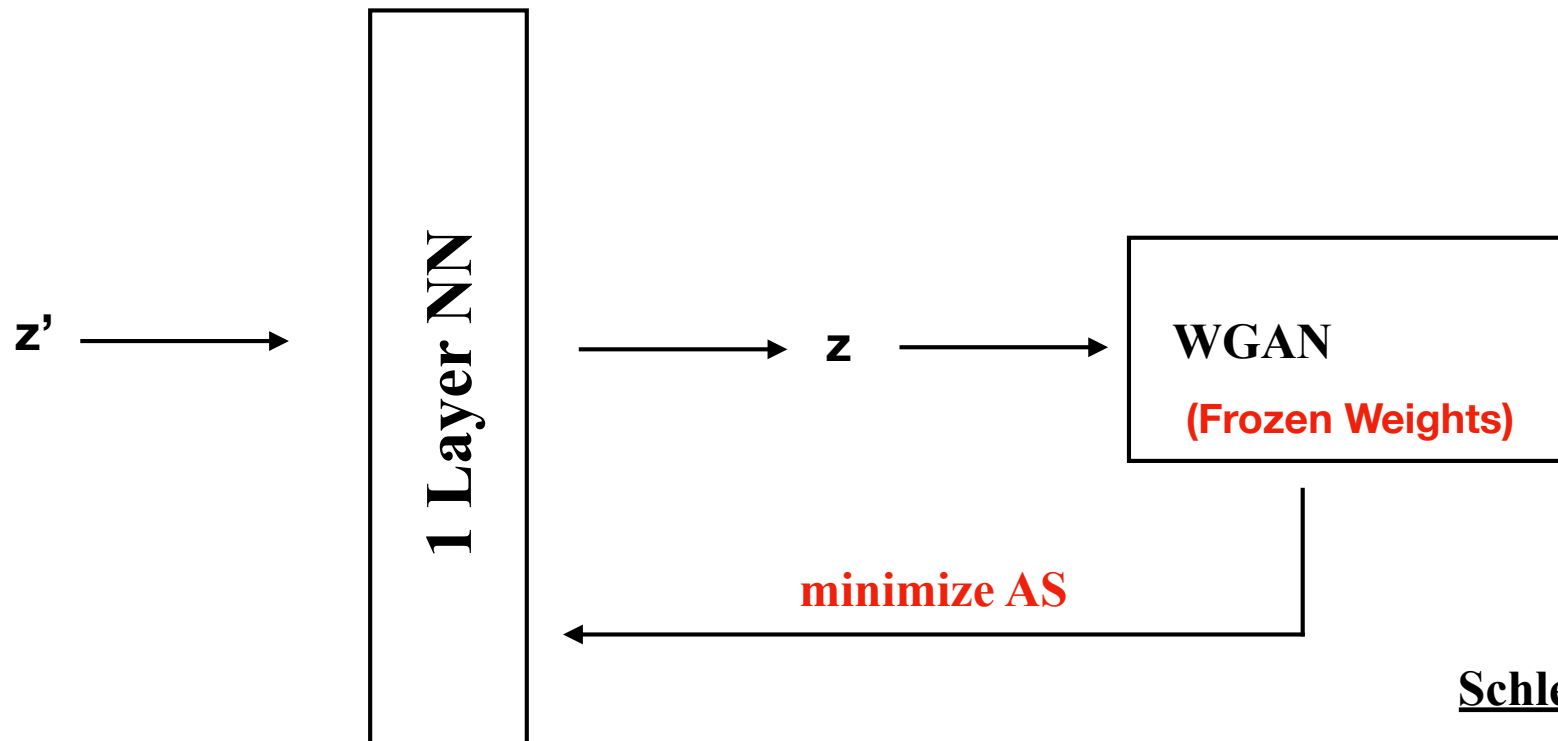
ANOMALY SCORE:

$$AS = \lambda G + (1 - \lambda)C$$

Hyperparameter

RMS btw input
and generated

RMS btw critic features
from input and generated



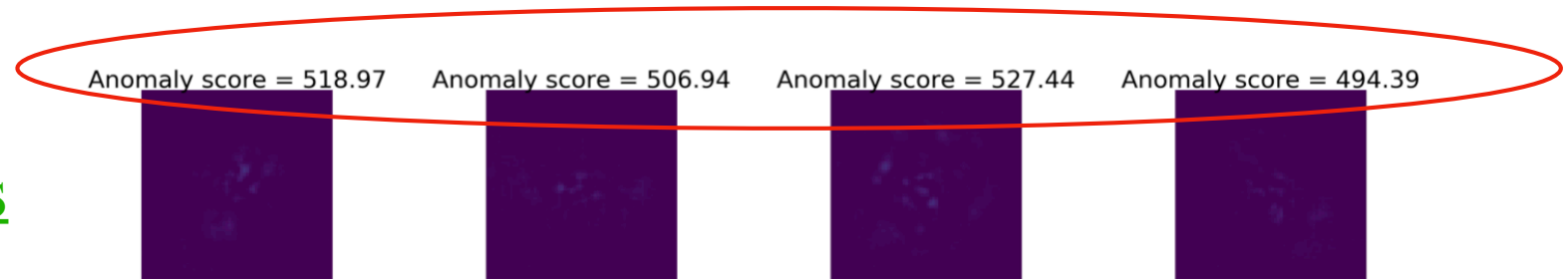
“NORMAL” GALAXIES



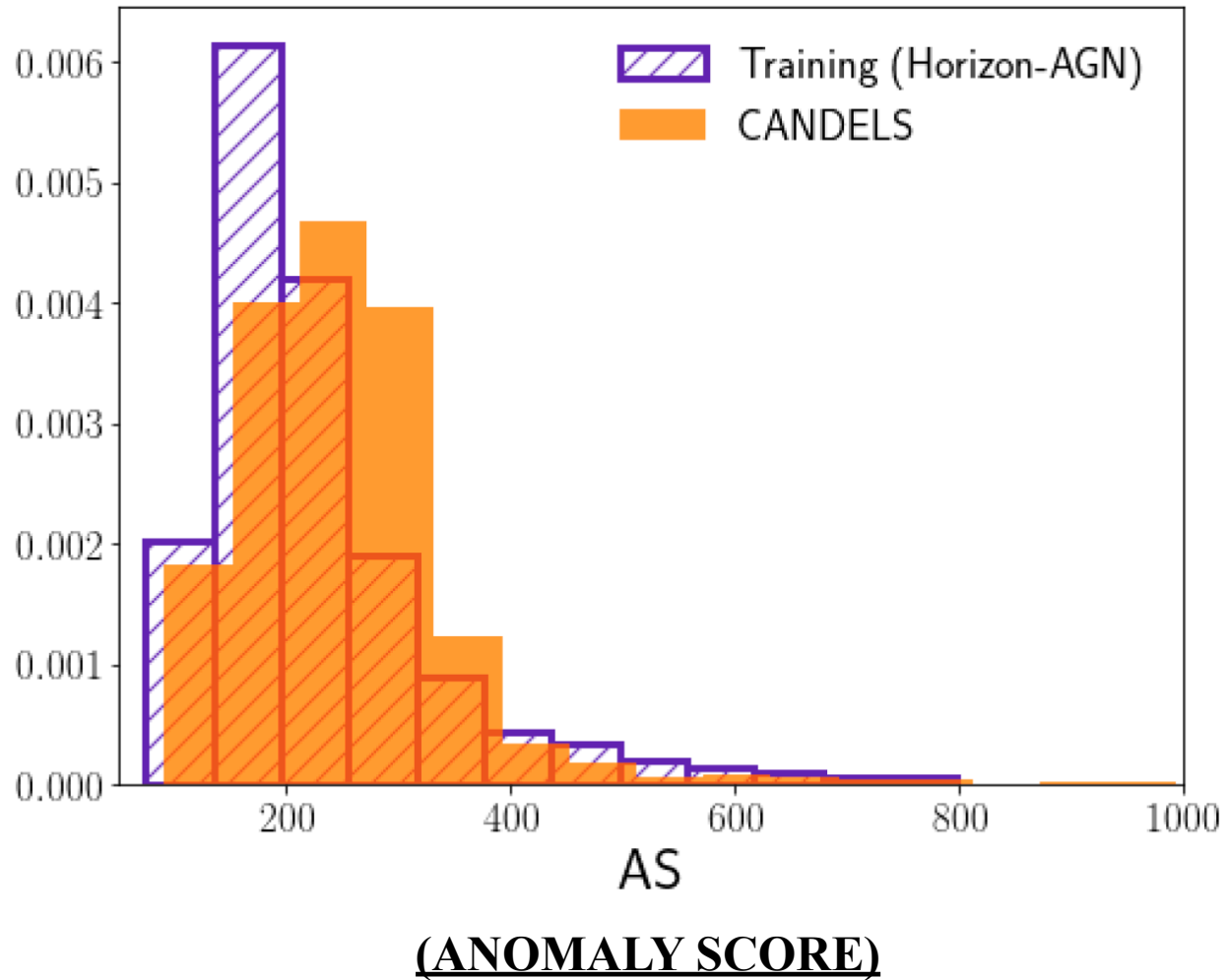
GAN BEST GUESS FOR
NORMAL GALAXIES

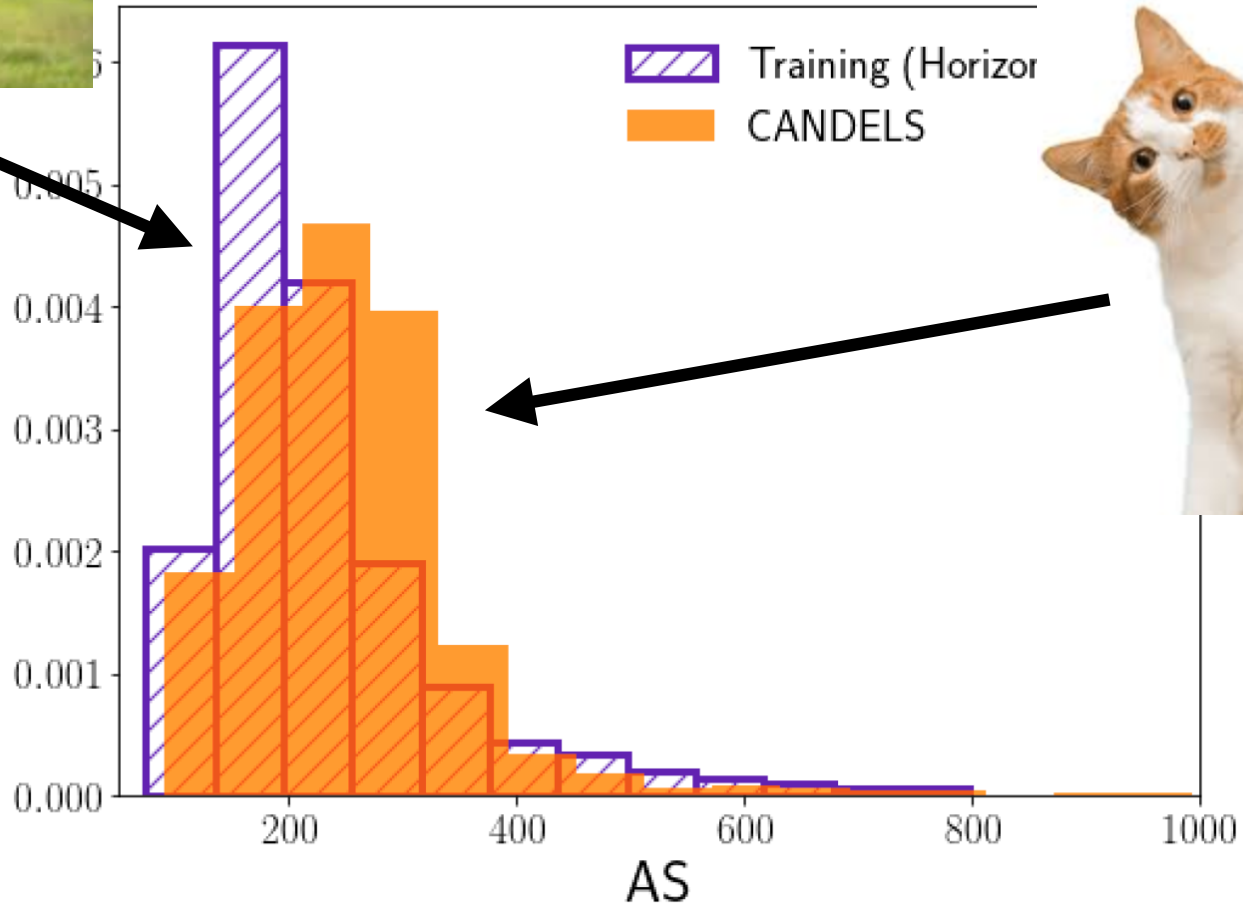


RESIDUAL MAPS



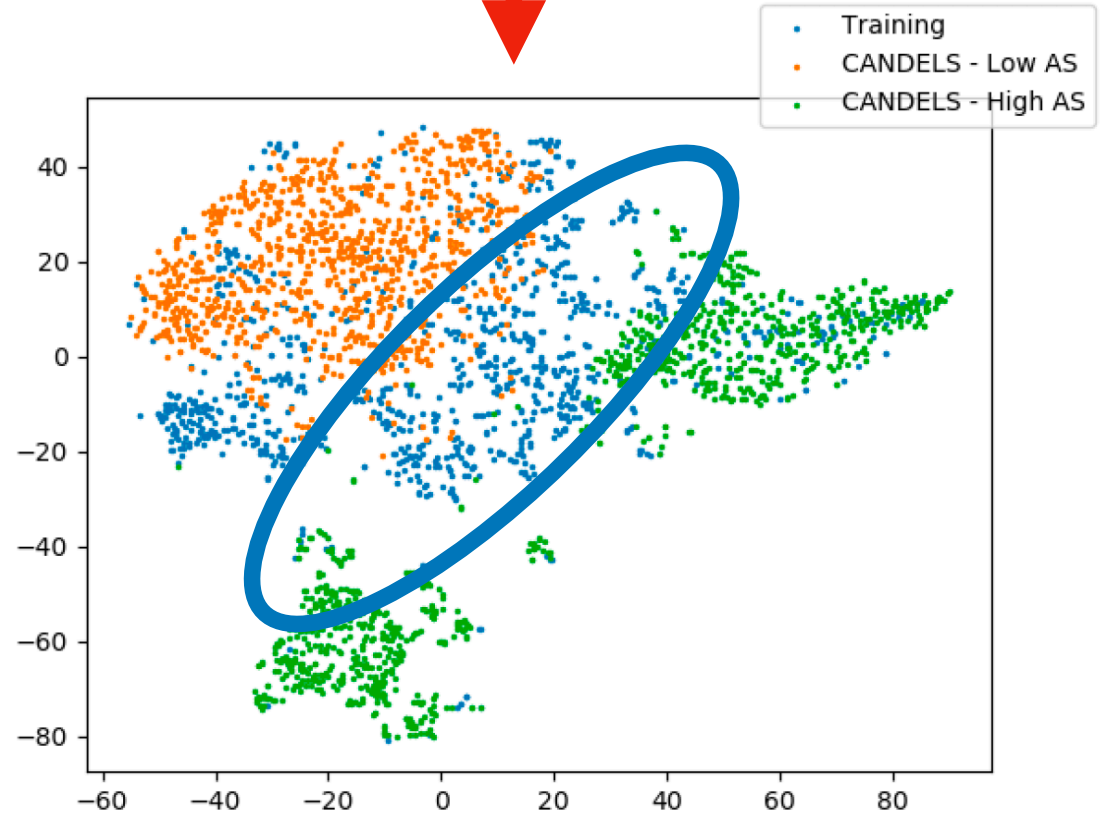
ANOMALY SCORE DISTRIBUTION OF OBSERVATIONS AND SIMULATIONS [TRAINED ON SIMS]



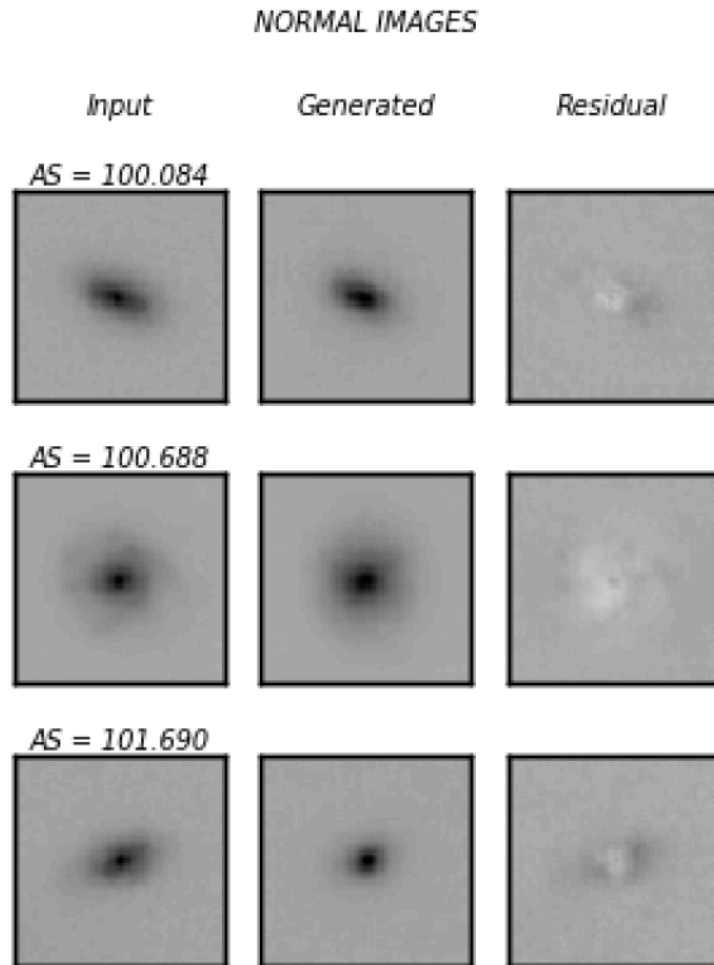


(ANOMALY SCORE)

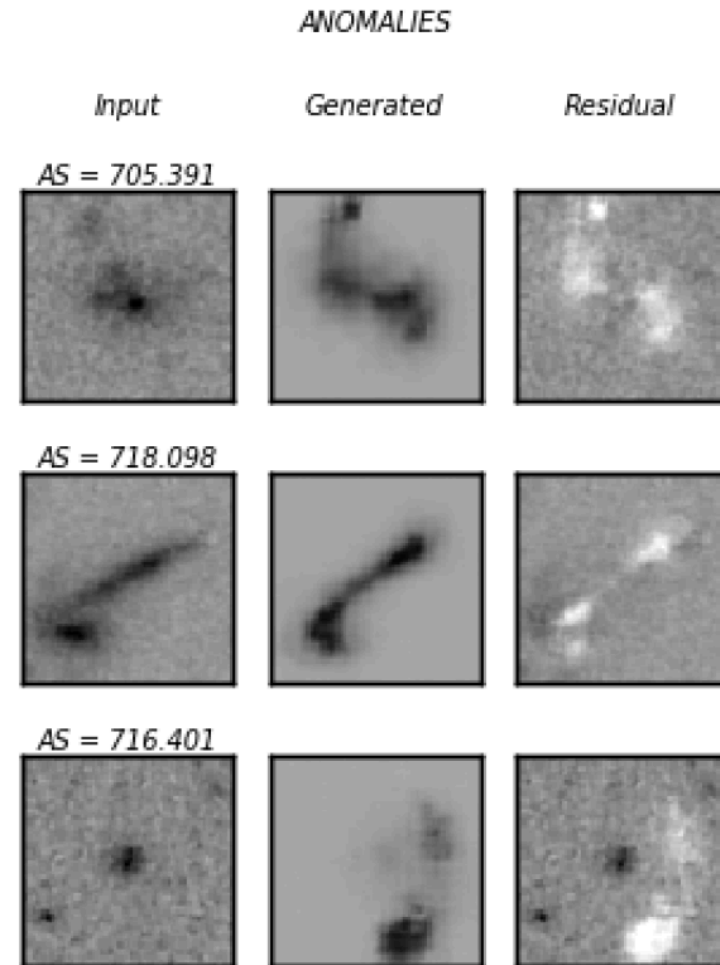
T-SNE REPRESENTATION OF LEARNED FEATURES



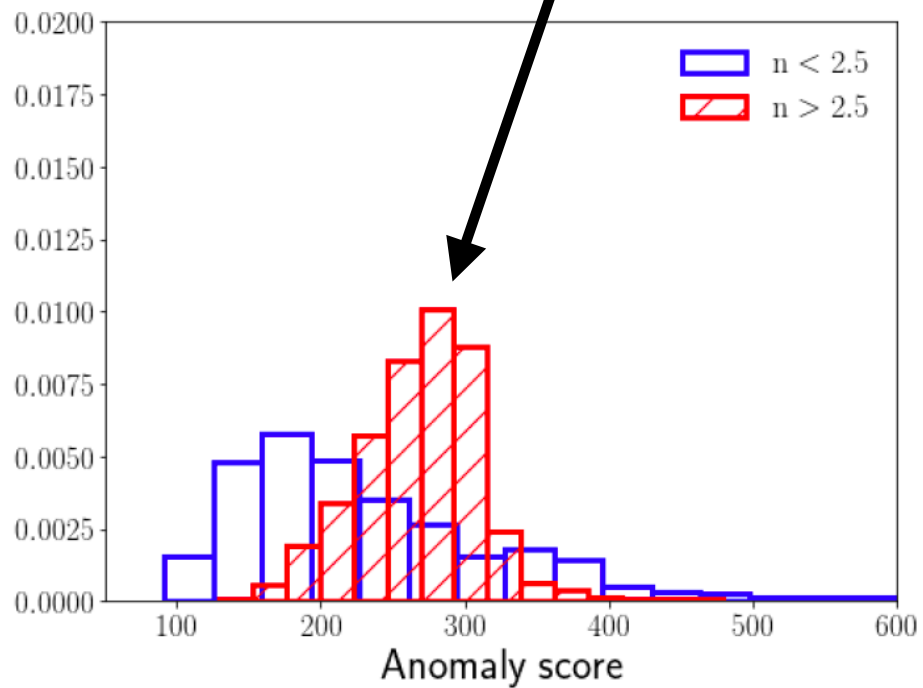
CANDELS GALAXIES WELL CAPTURED BY HORIZON-AGN SIM



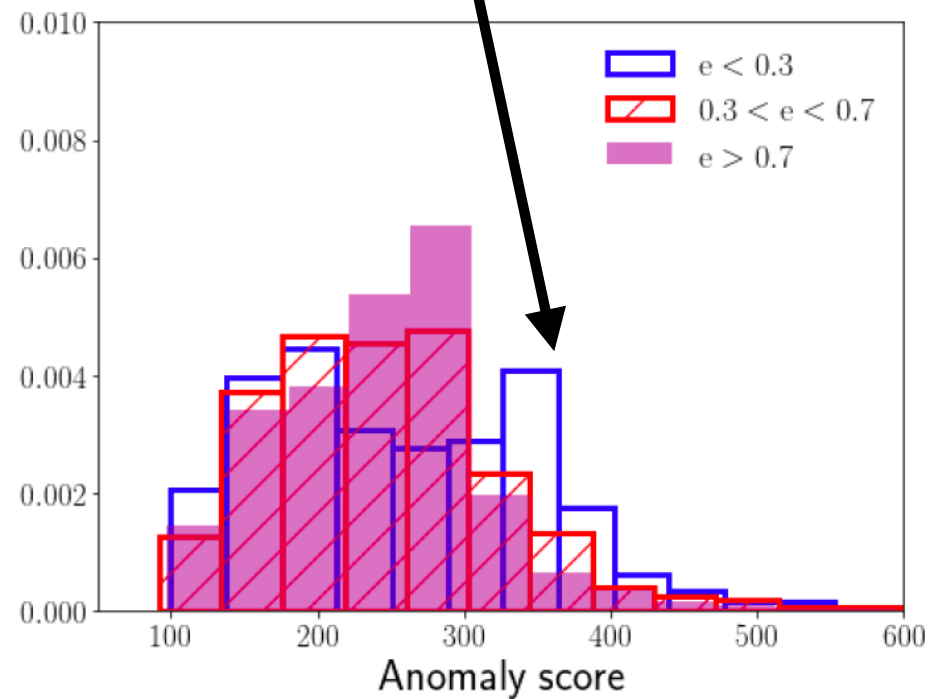
CANDELS GALAXIES **NOT** WELL CAPTURED BY HORIZON-AGN SIM



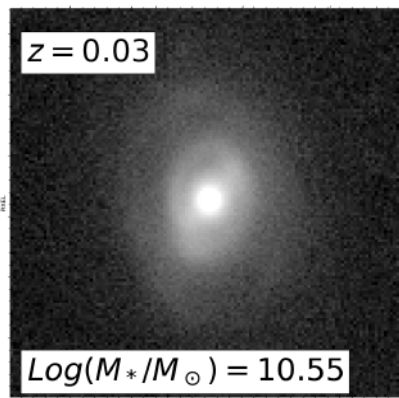
SPHEROIDS



ELONGATED GALAXIES



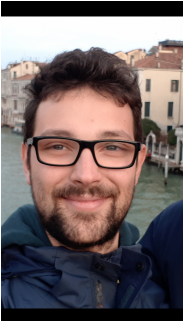
NOT WELL CAPTURED BY HORIZON-AGN



**SDSS
GALAXY**

**AUTOREGRESSIVE IMAGE
GENERATION:**
pixelCNN

[van der Oord+16, Salimans+17]



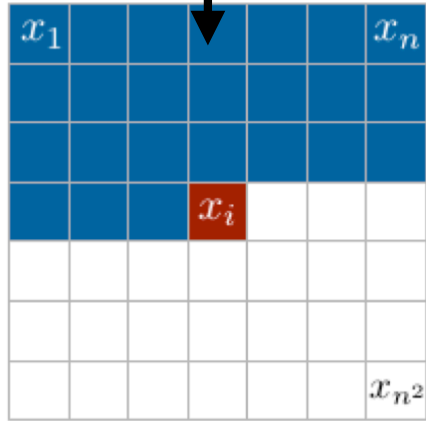
$z = 0.03$

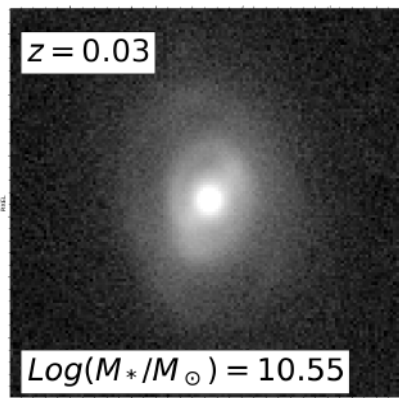
$\text{Log}(M_*/M_\odot) = 10.55$

**SDSS
GALAXY**

**AUTOREGRESSIVE IMAGE
GENERATION:**
pixelCNN

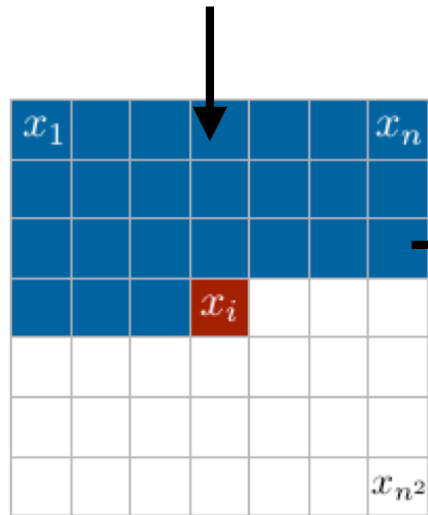
[van der Oord+16, Salimans+17]



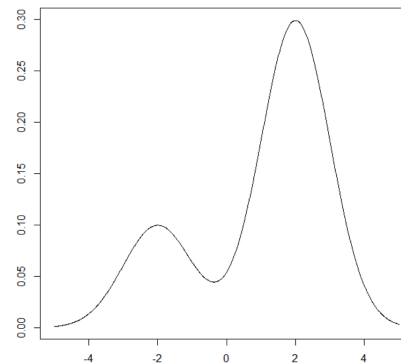


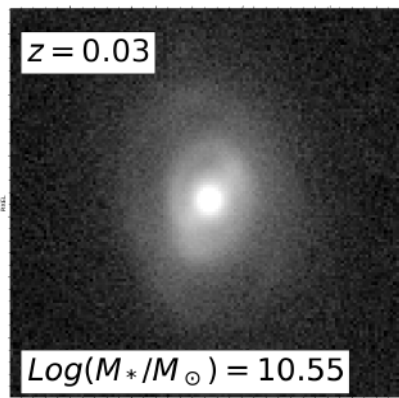
AUTOREGRESSIVE IMAGE GENERATION: pixelCNN

[van der Oord+16, Salimans+17]



PDF FOR ONE PIXEL

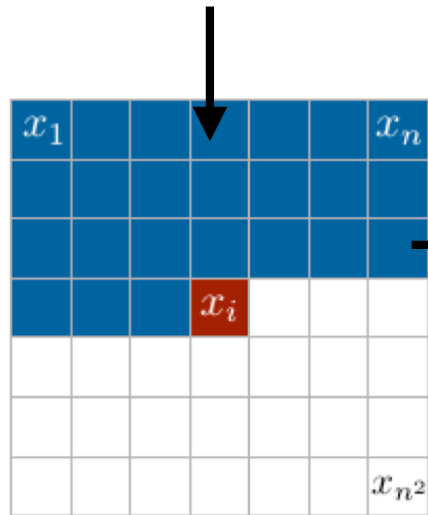




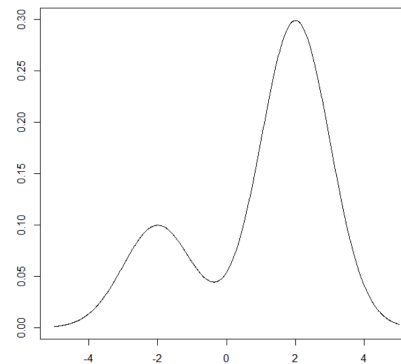
**SDSS
GALAXY**

AUTOREGRESSIVE IMAGE GENERATION: pixelCNN

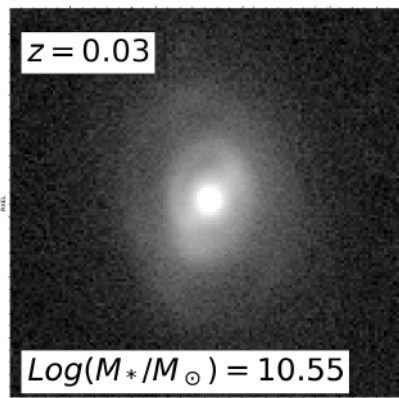
[van der Oord+16, Salimans+17]



PDF FOR ONE PIXEL



$$p(x) = p(x_0, x_1, \dots, x_{n^2} | \theta_{SDSS})$$

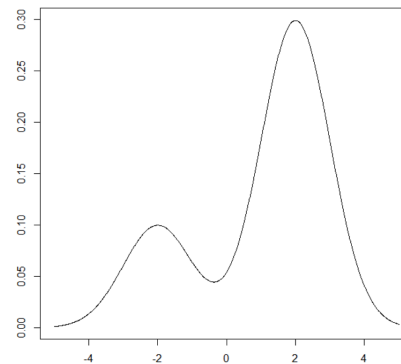
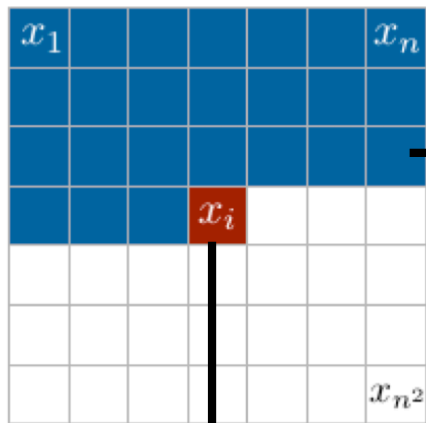


**SDSS
GALAXY**

AUTOREGRESSIVE IMAGE GENERATION: pixelCNN

[van der Oord+16, Salimans+17]

PDF FOR ONE PIXEL



$$p(x) = p(x_0, x_1, \dots, x_{n^2} | \theta_{SDSS})$$

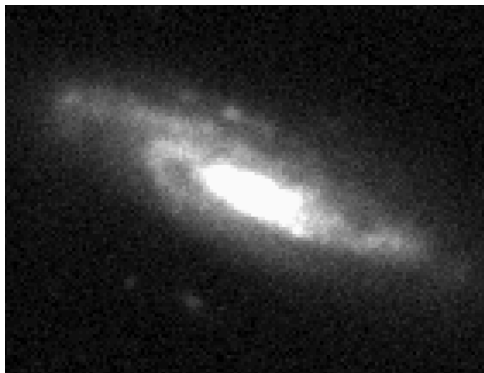
$$p(x_i | x_1, \dots, x_{i-1})$$

$$p(\mathbf{x}) = \prod_{i=1}^{n^2} p(x_i | x_1, \dots, x_{i-1})$$

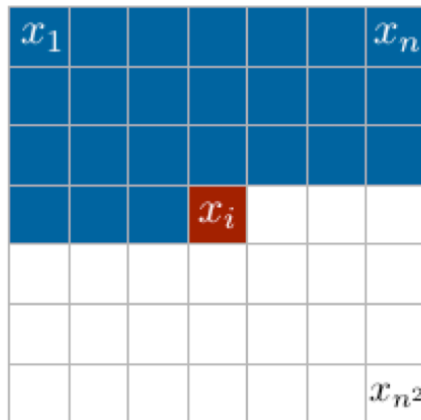
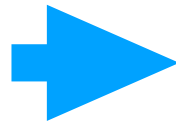
SDSS DR7 DATASET

$\text{Log}(M^*) > 10$
 $0.02 < Z < 0.08$

~ 100,000 galaxies



r-band images



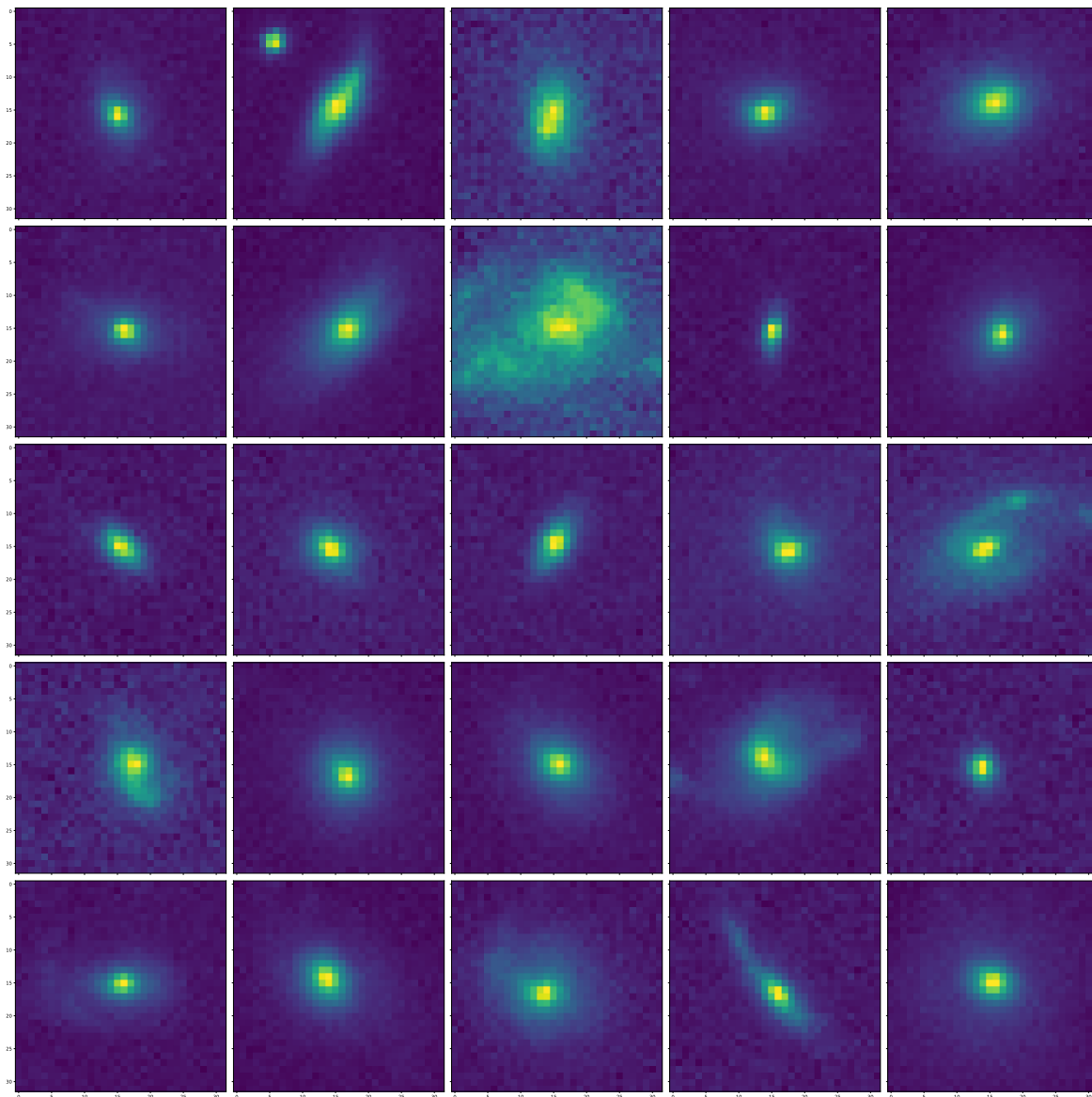
pixelCNN



$$p(x_0, x_1, \dots, x_{n^2} | \theta_{SDSS}^r)$$

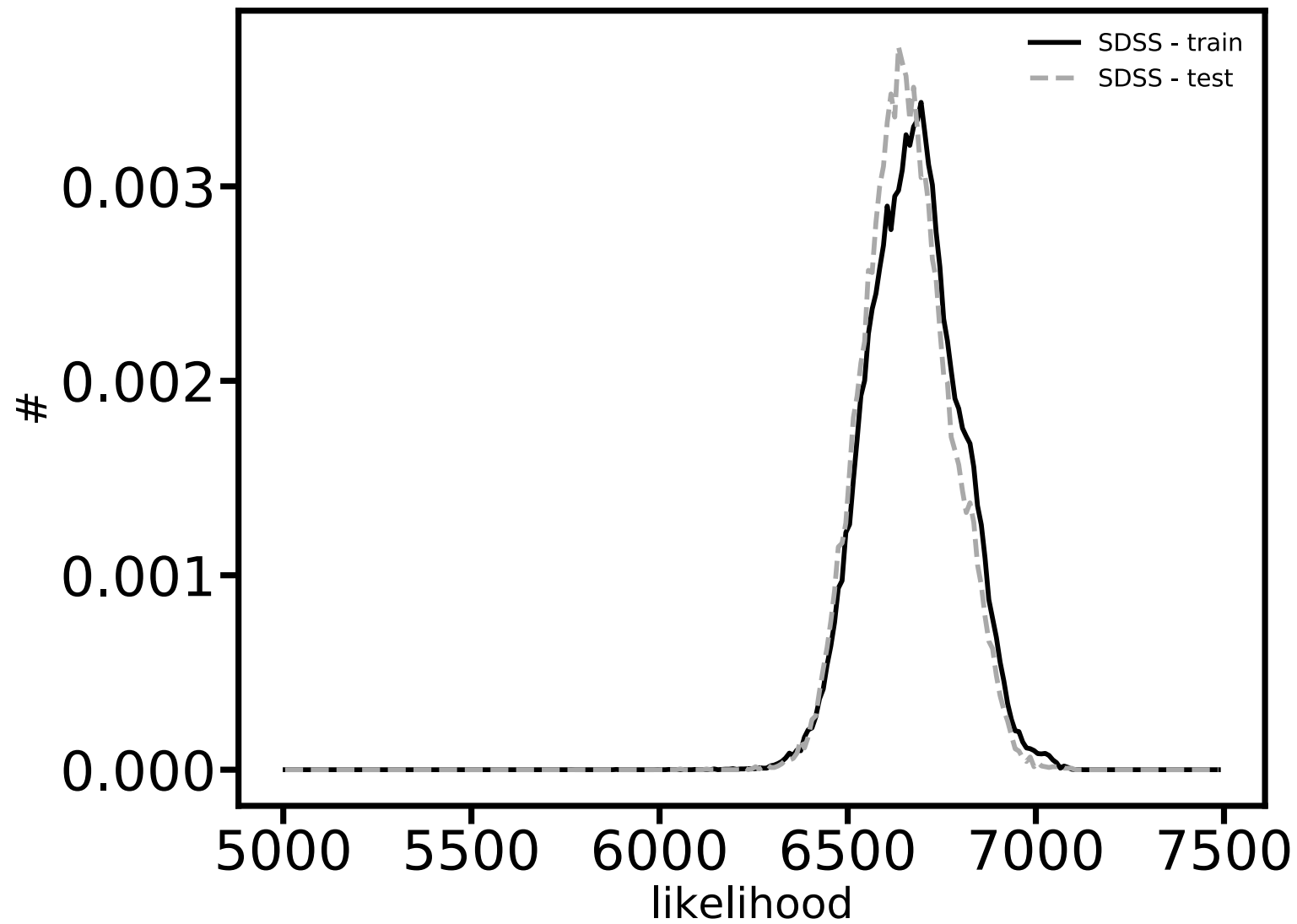
RANDOM SAMPLES FROM THE MODEL

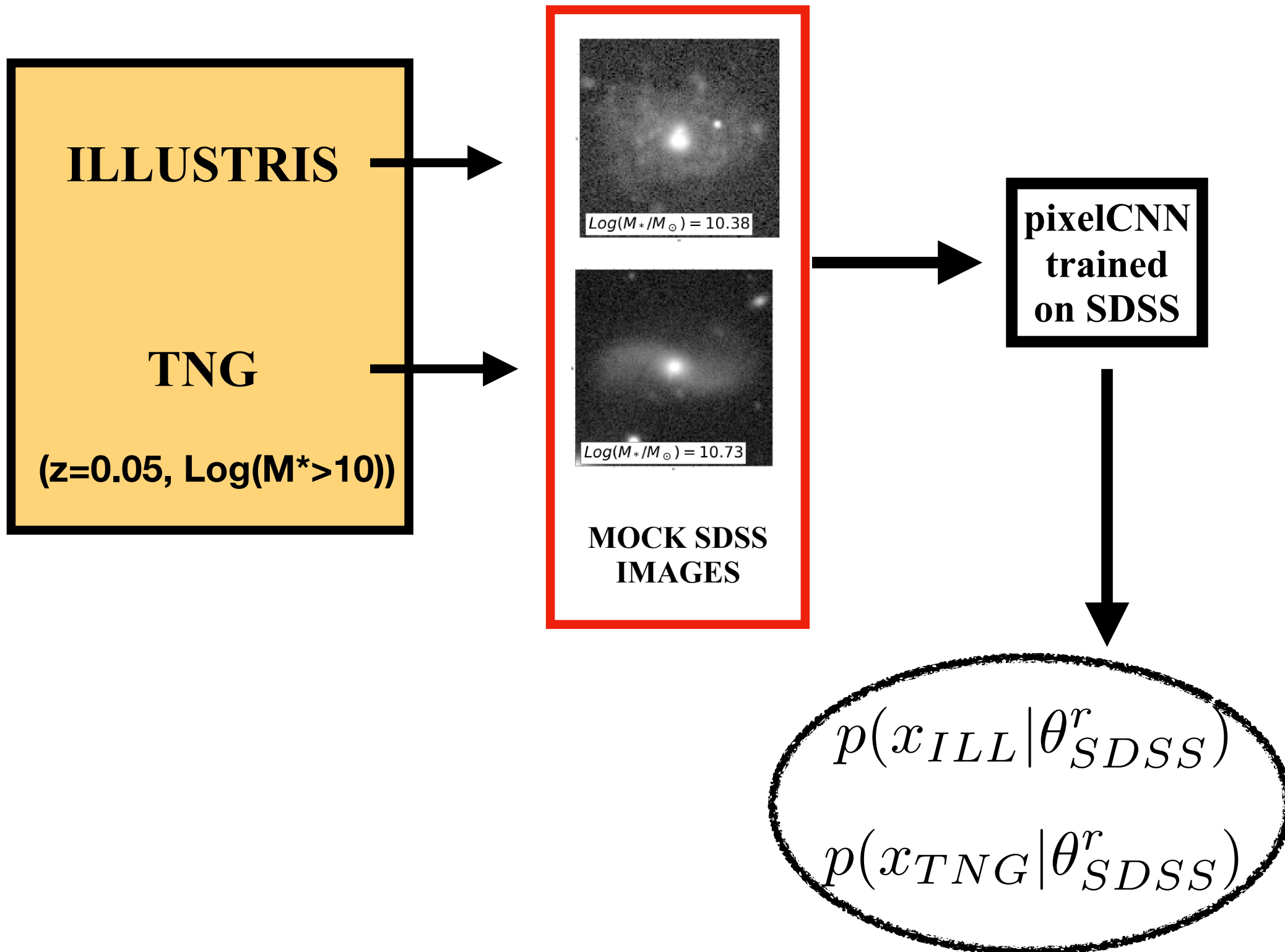
$$\theta_{SDSS}^r$$

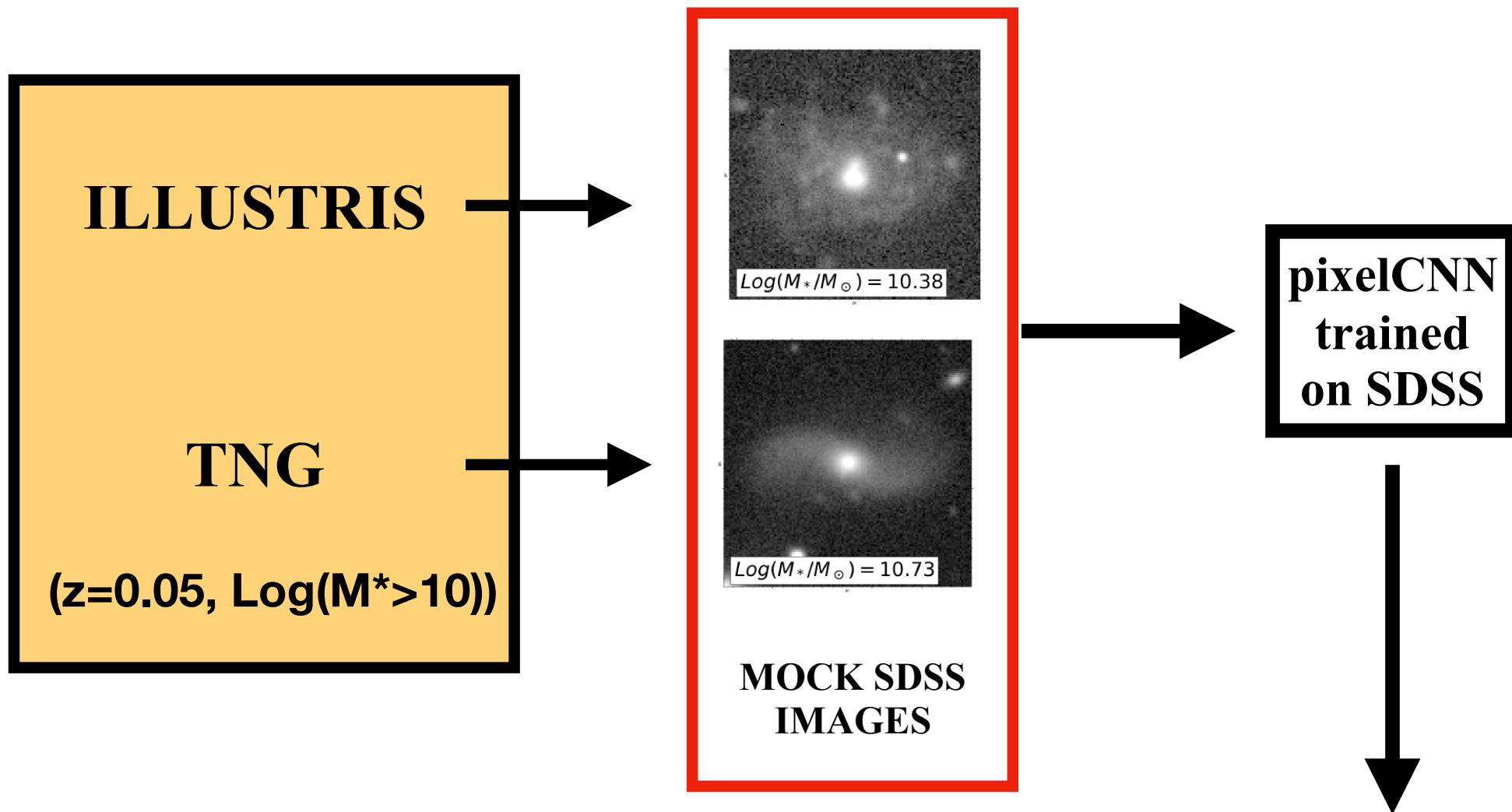


**RANDOM “FAKE” OF
SDSS GALAXIES
OBTAINED THROUGH
SAMPLING OF THE
PDFs**

DISTRIBUTION OF $p(x)$ for SDSS GALAXIES







$$p(x_{ILL} | \theta_{SDSS}^r, SKIRT)$$

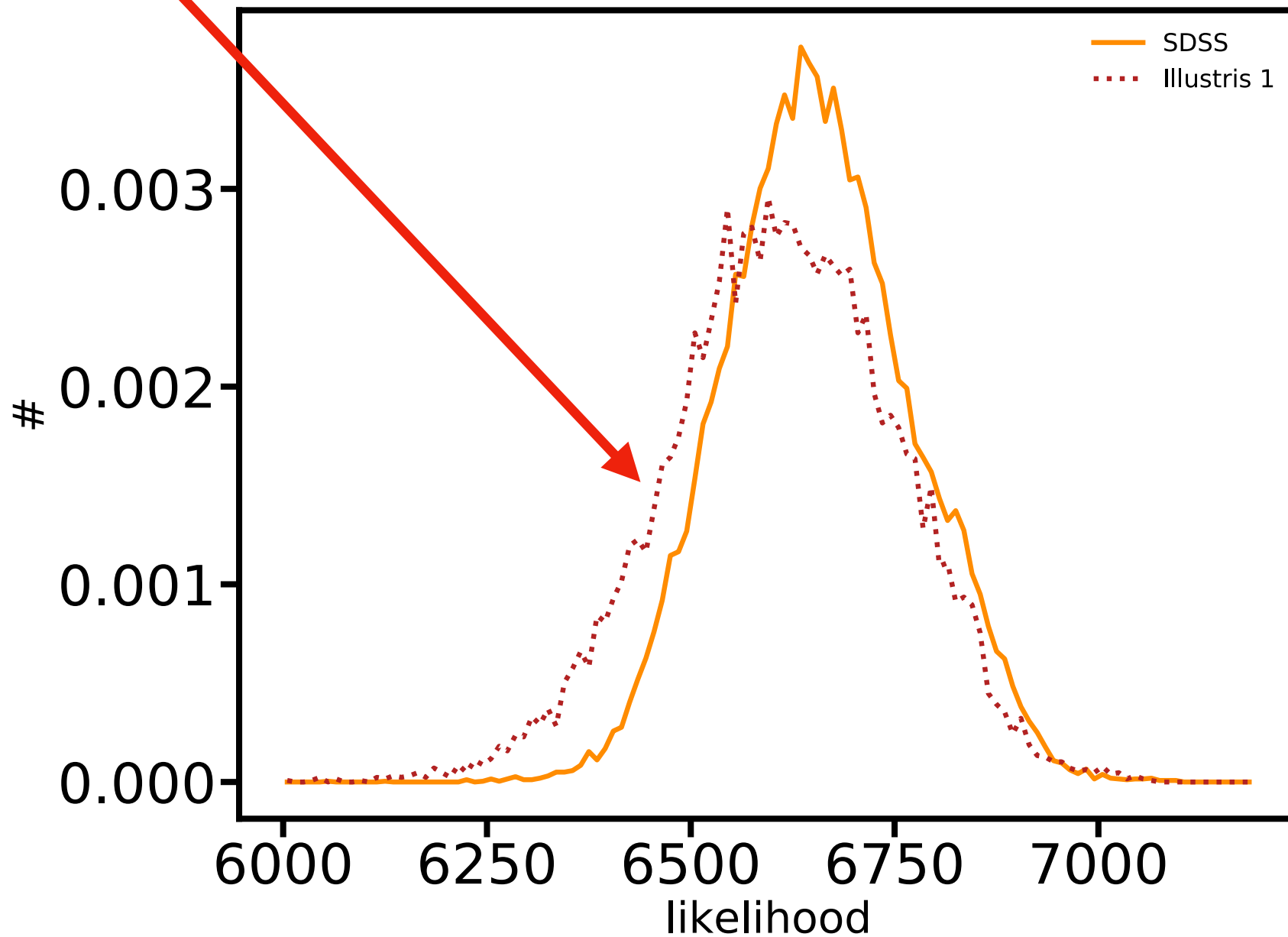
$$\approx$$

$$p(x_{TNG} | \theta_{SDSS}^r, SKIRT)$$

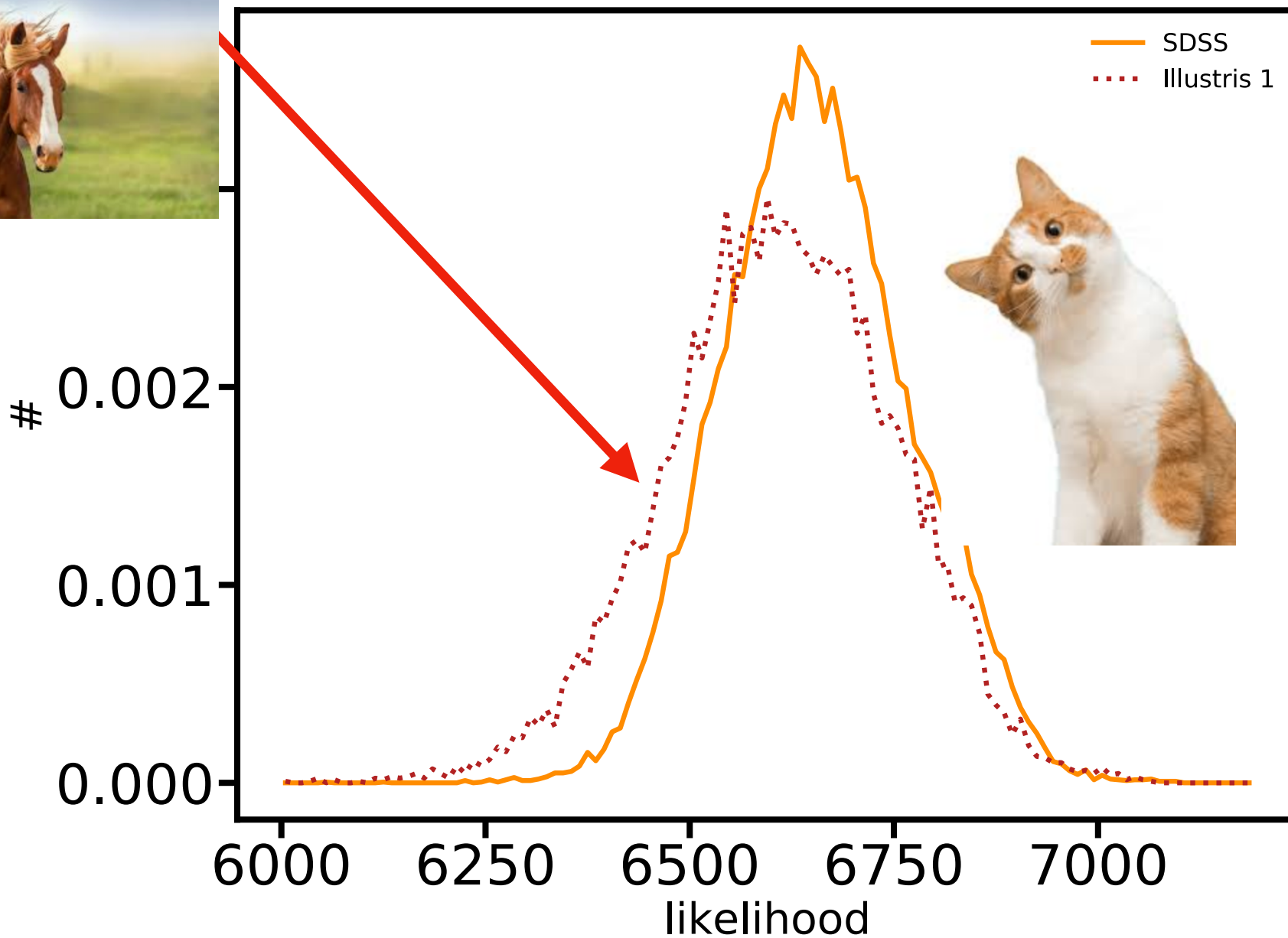
$$p(x_{ILL} | \theta_{SDSS}^r)$$

$$p(x_{TNG} | \theta_{SDSS}^r)$$

ILLUSTRIS

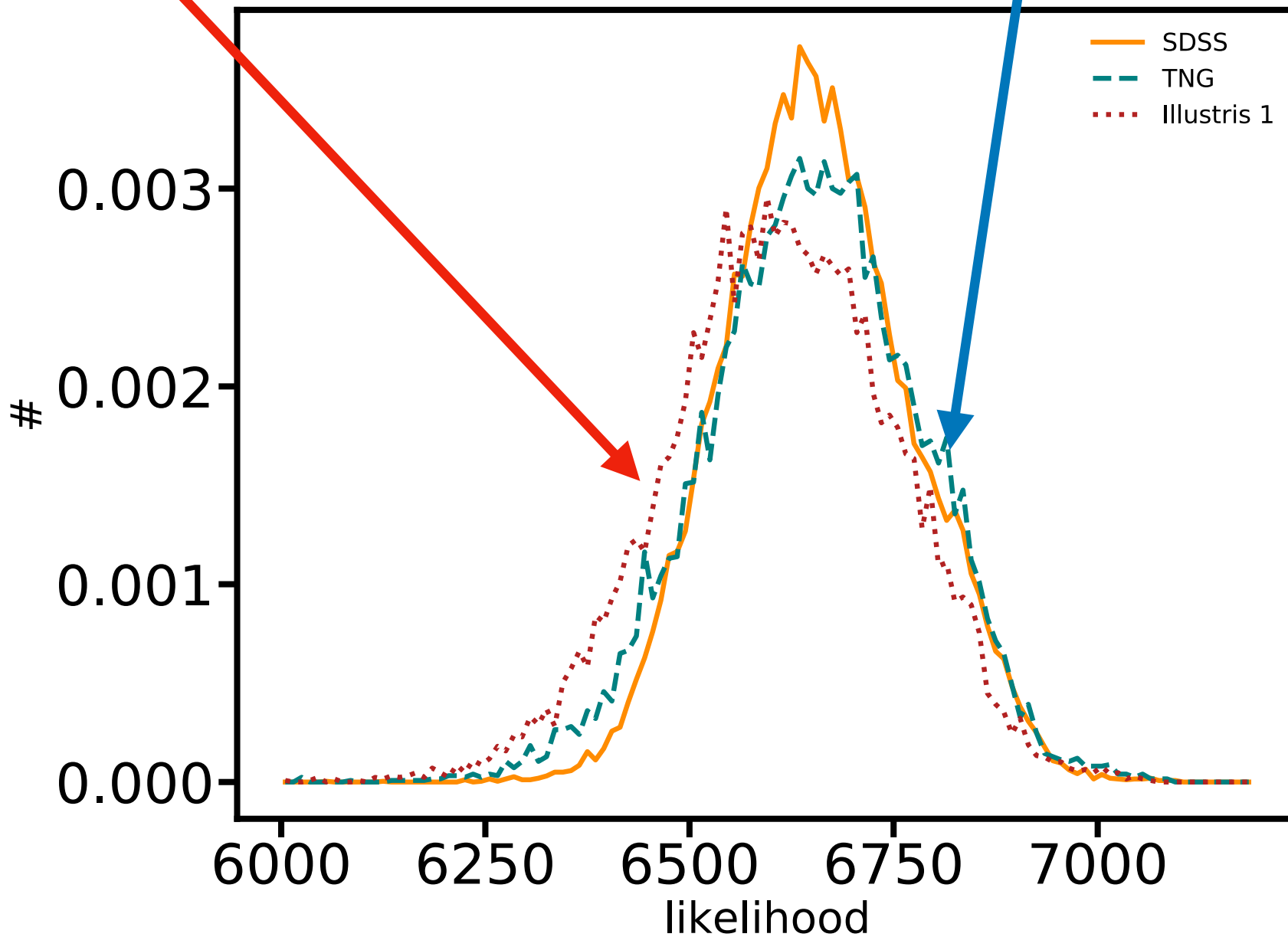


ILLUSTRIS



ILLUSTRIS

TNG



***DOES A NEURAL NETWORK
KNOW ABOUT HORSES IF IT HAS
ONLY SEEN CATS AND DOGS?***

(MAYBE) YES!

