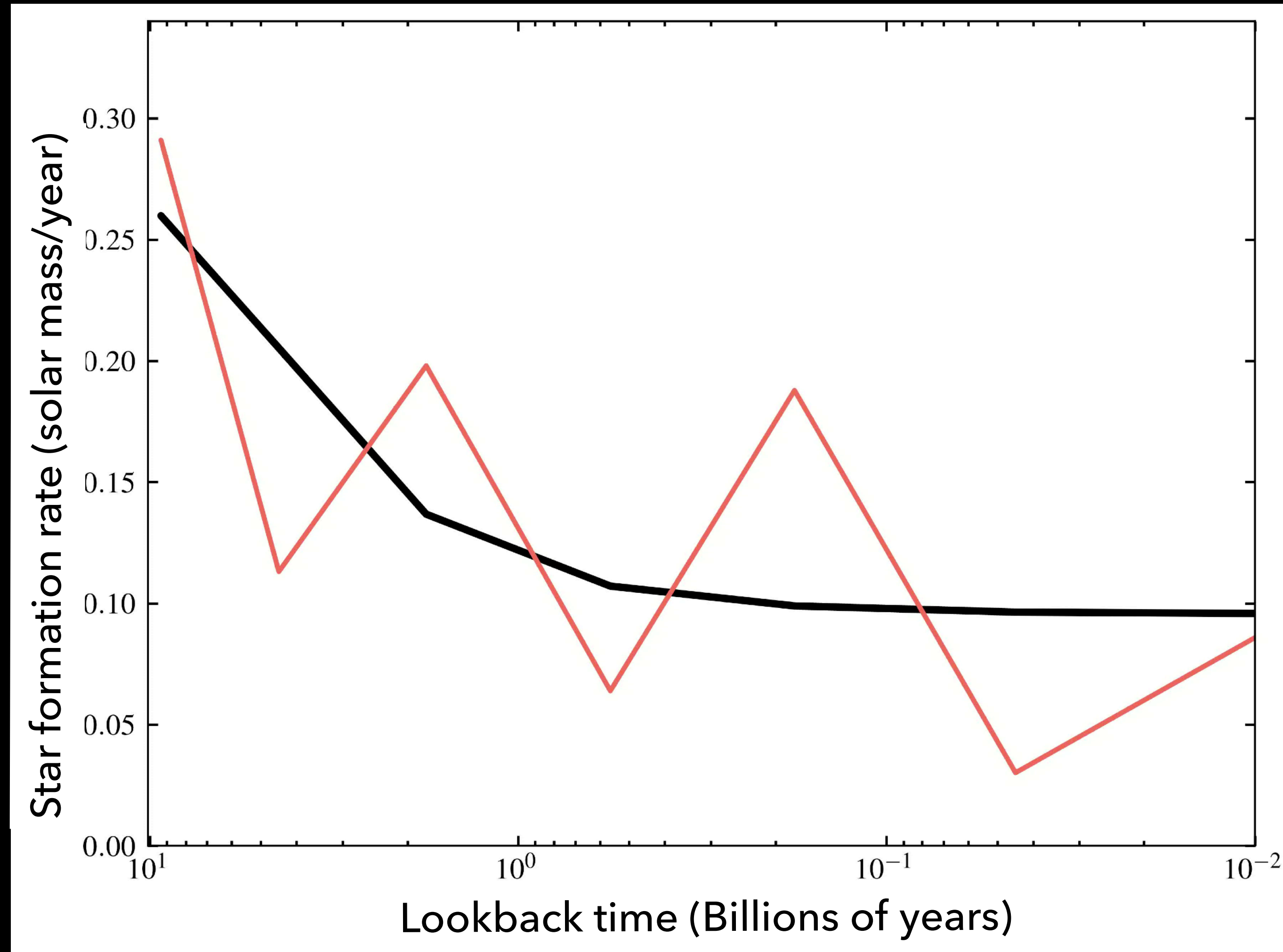


Rapid inference of galaxy properties in the age of **deep** and **large-scale** surveys of the universe



Joel Leja

Associate Professor in Astronomy & Astrophysics
Institute for Computational & Data Sciences @ Penn State
Dr. Keiko Miwa Ross Mid-Career Endowed Chair

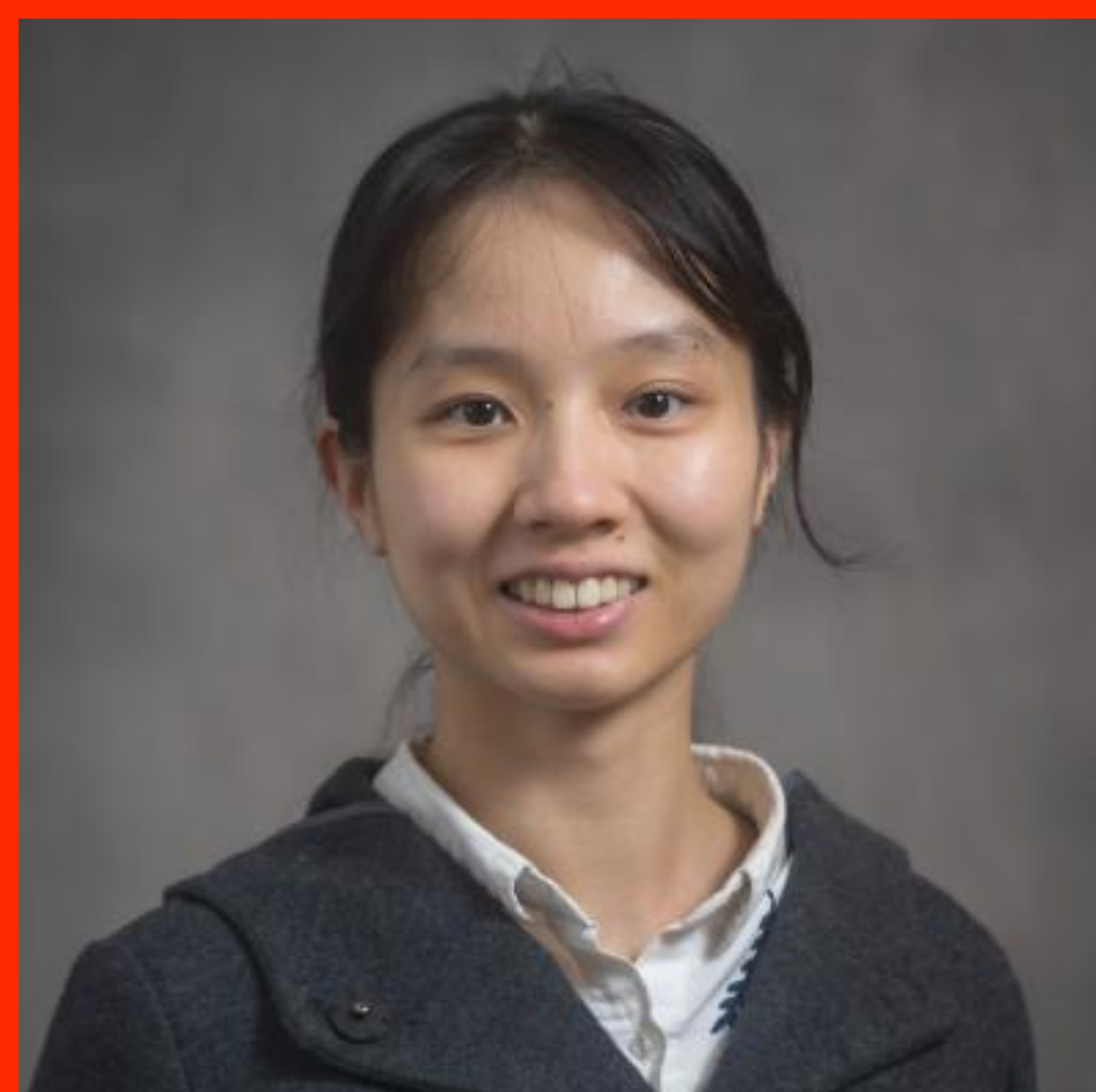
Emilie Burnham
PSU grad
Was here - see her poster!



Dr. Bingjie Wang (former PSU)
NASA Hubble Postdoctoral
Fellow @ Princeton
Here - next talk!



Dr. Yijia Li (former PSU)
Postdoc @ Northwestern



Elijah Mathews
PSU grad



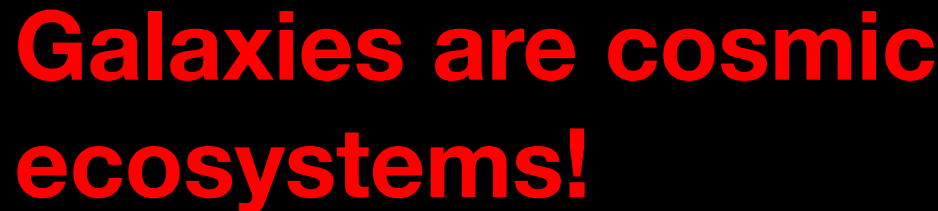
Perhaps the most important astronomical
picture of all time.
But all I see are problems.



Hubble Deep Field

HST WFPC2

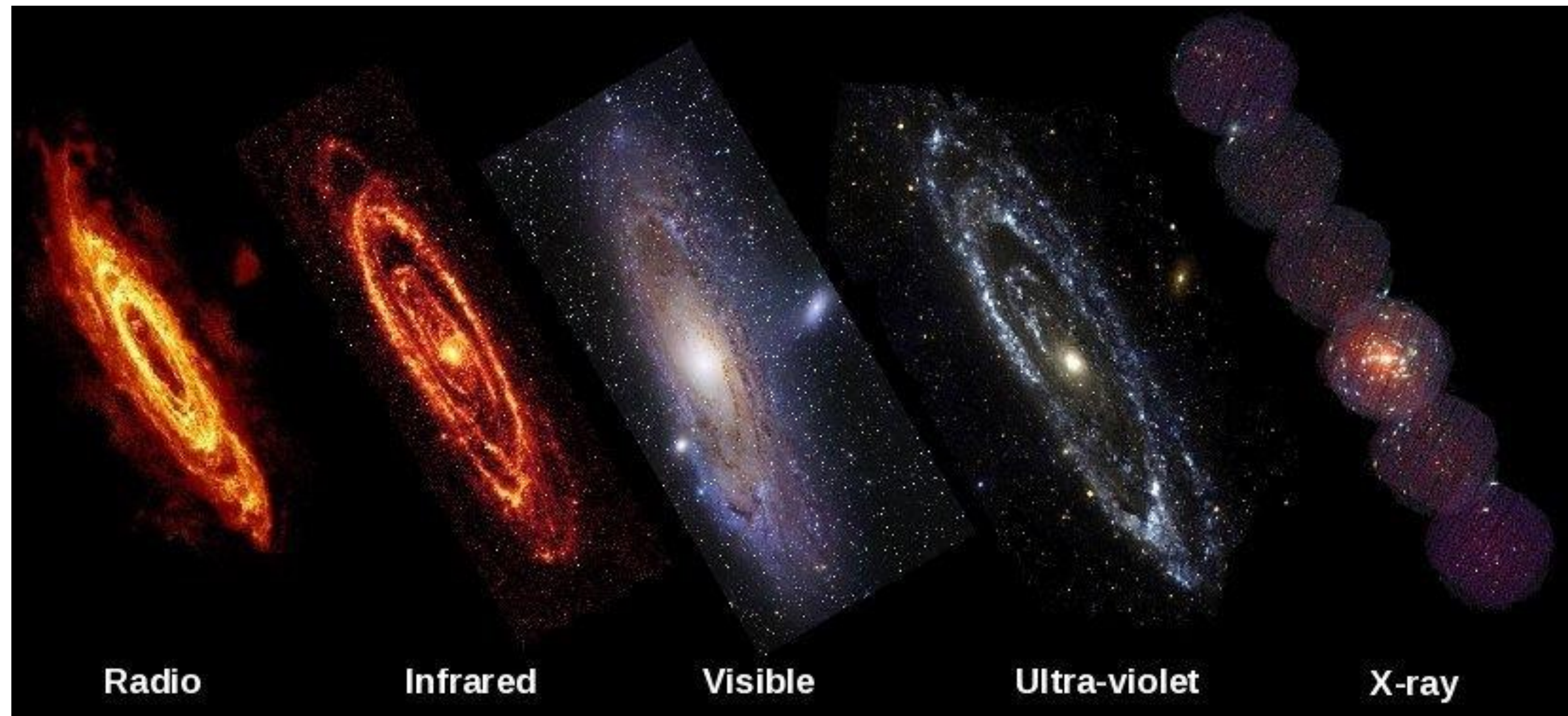
ST ScI OPO January 15, 1996 R. Williams and the HDF Team (ST ScI) and NASA



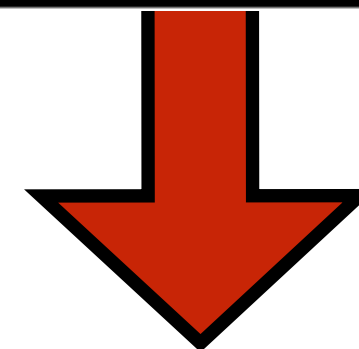
Hidden inside: rich information about the astrophysics of the universe!

- Galaxy formation histories
- Production rate of heavy elements, cosmic dust
- supernovae progenitors
- cosmology,
- stellar evolution,
-!

Galaxy properties are inferred by fitting observed data with models.
Take beautiful galaxy data:



The Andromeda Galaxy
Planck / NASA / ESA



... and use models to turn them into *more beautiful*
inferred parameters.

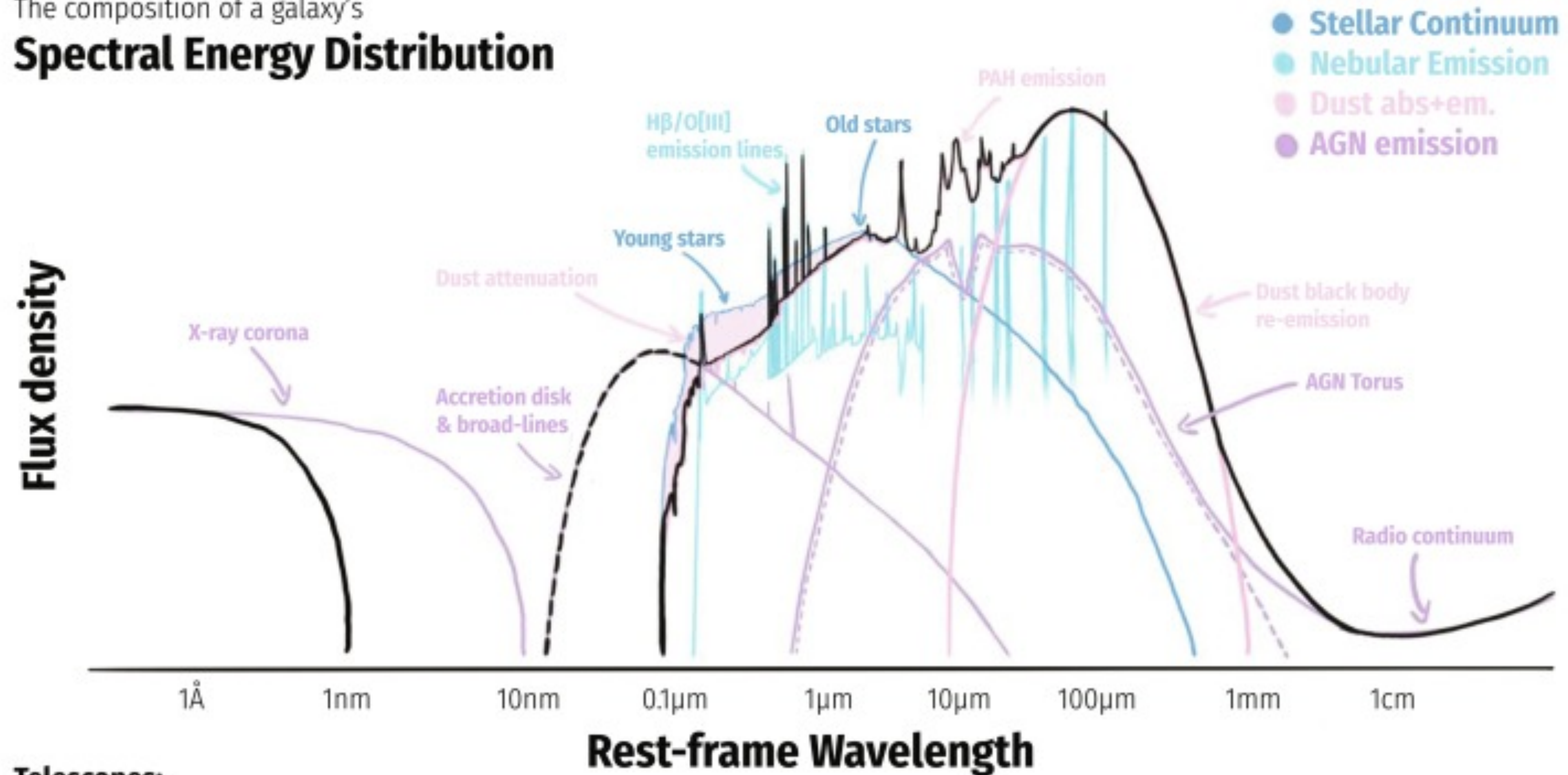
stellar mass
star formation history
nebular properties

dust content
chemical abundances
active black holes

Thanks to decades of investment, modern galaxy evolution surveys are **multiwavelength**, usually spanning **ultraviolet** to **infrared** (sometimes more!)

The composition of a galaxy's

Spectral Energy Distribution



Iyer+25

Telescopes:

Fermi

XMM-Newton
Chandra

ROSAT

GALEX
HST

Gemini
Subaru
Keck

WISE
JWST
Spitzer

Herschel
SOFIA

ALMA
NOEMA
JCMT

VLA

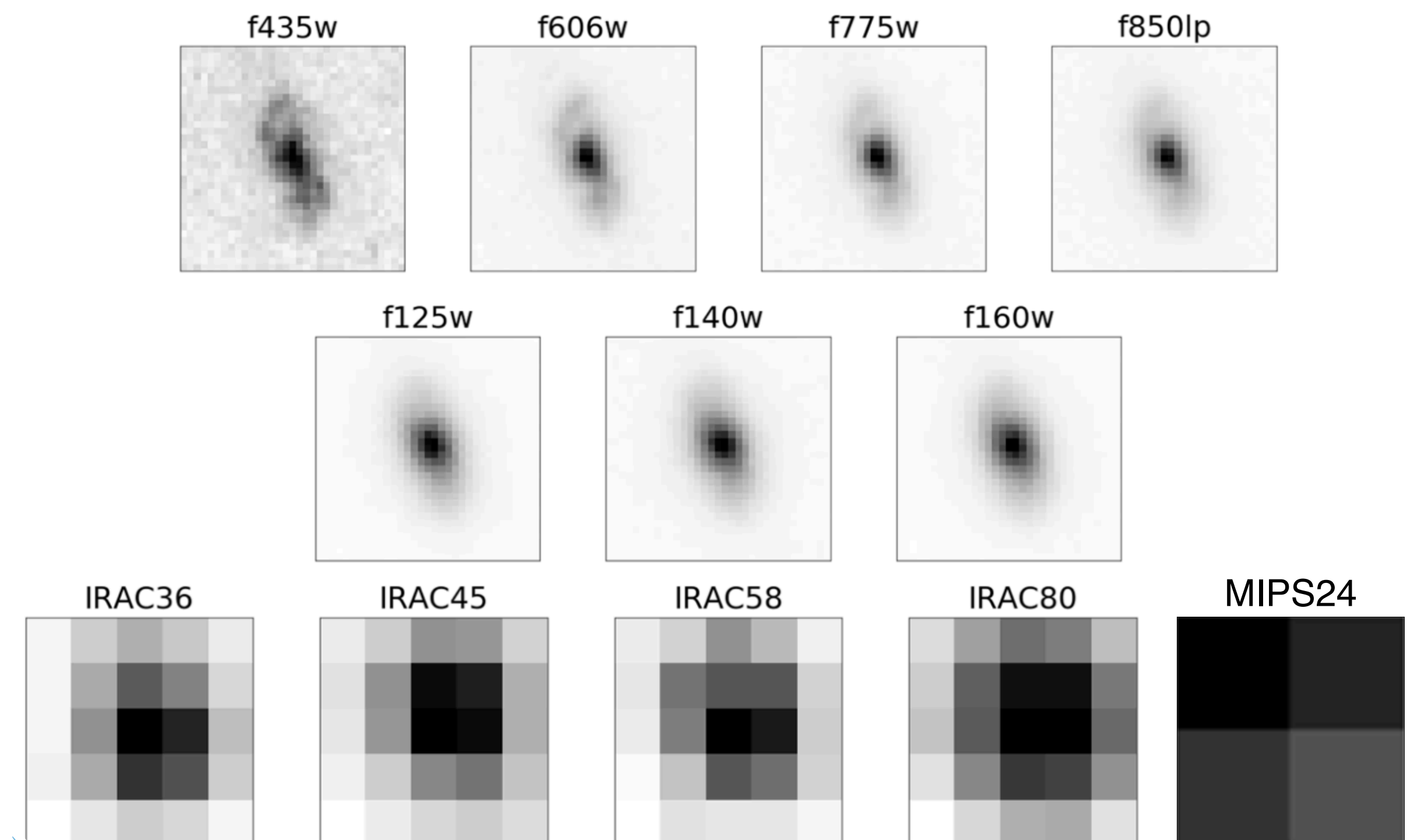
- Distant galaxies are nearly point-sources, so light from every star, nebula, black hole, etc is combined linearly.
- “**Outshining**” - Most of the light comes from a few **very bright objects** (e.g. O-stars, red giants); much of what we want to know comes from **dim things** (e.g., older, low-mass stars)

Andromeda

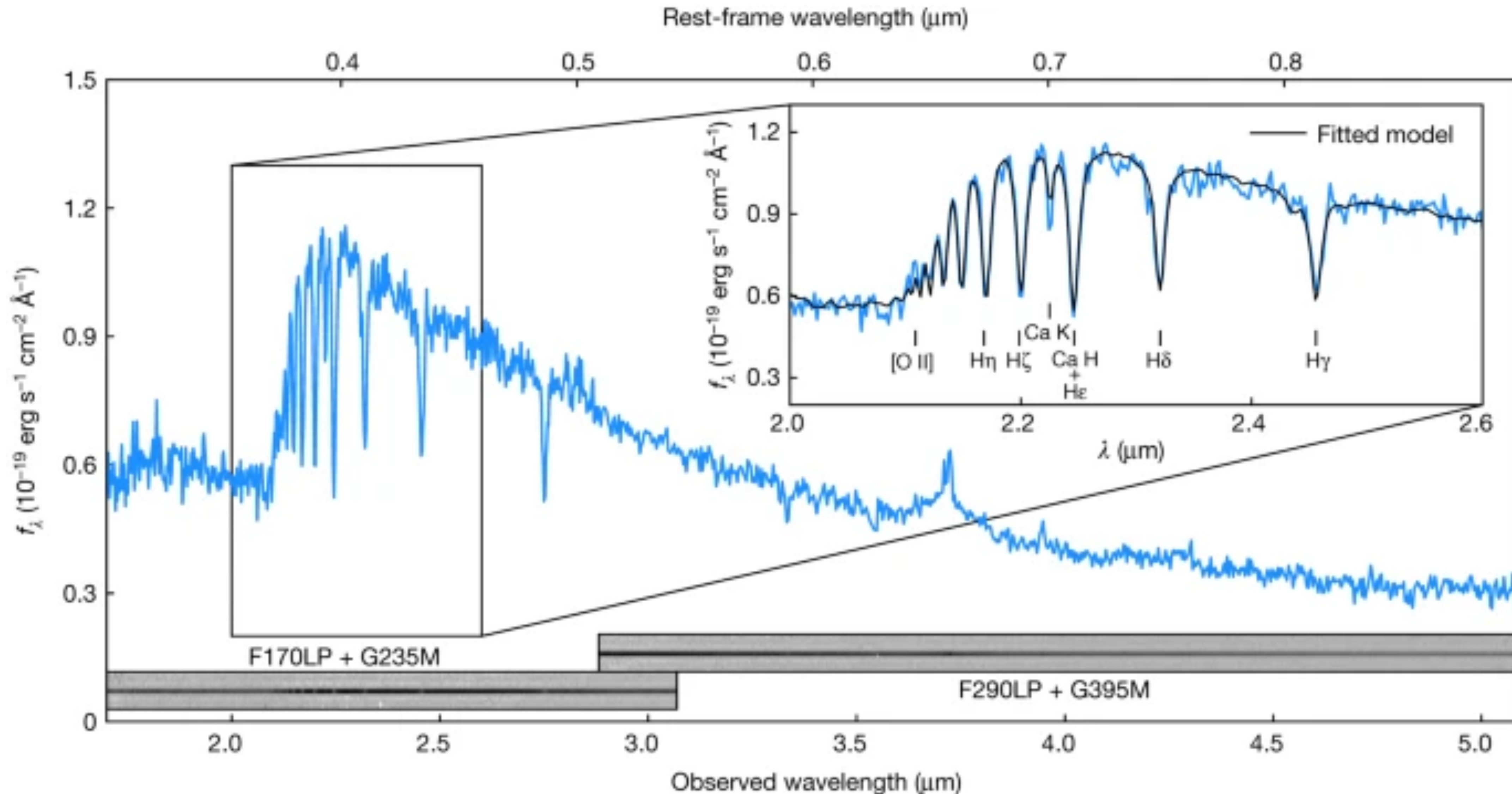


Inferring the properties of distant galaxies
dominated by modeling choices:
what do bright things tell us about the rest of the system?

A typical distant galaxy



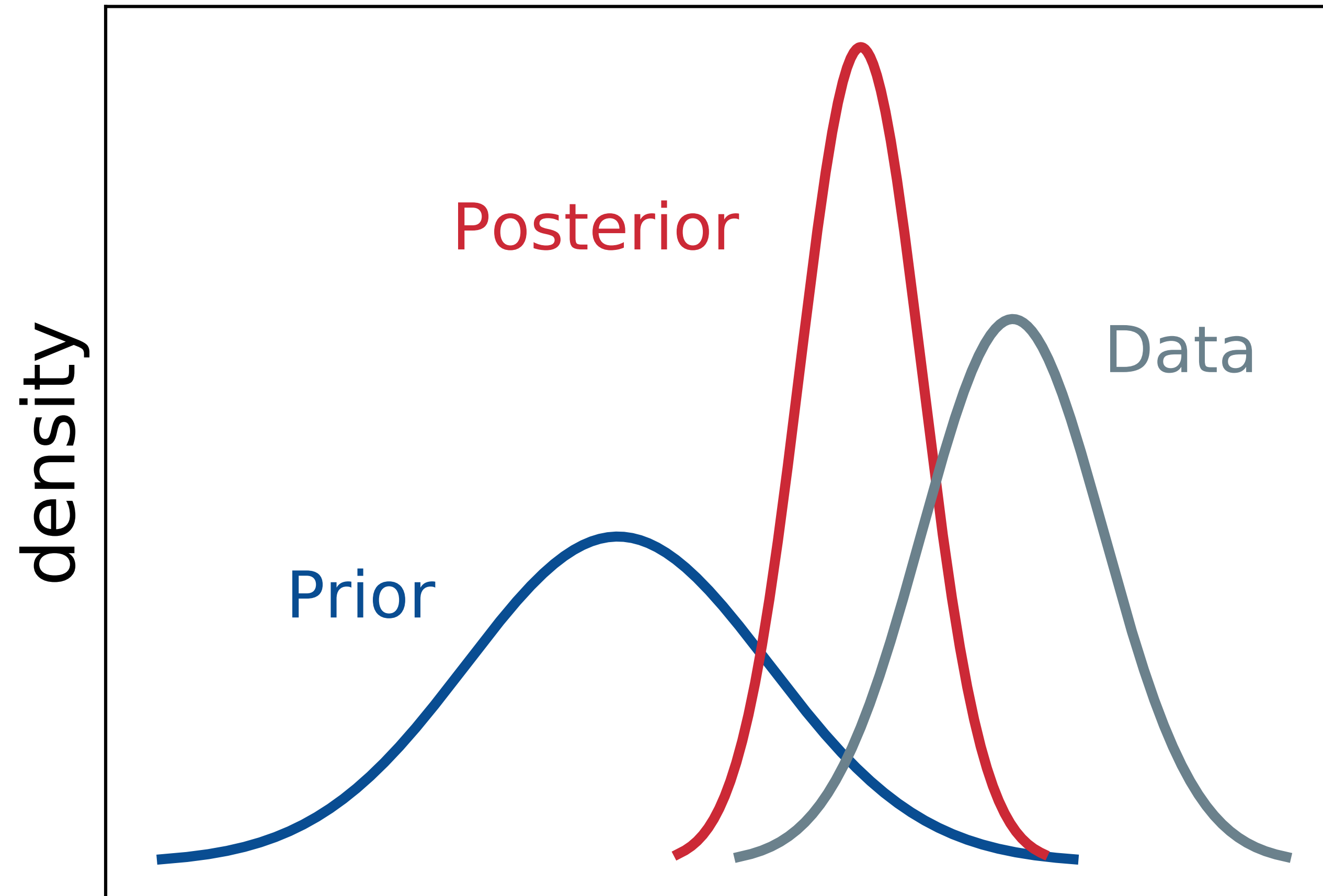
More rarely, spectroscopy is available. Spectroscopy is **richly informative** with many different absorption and emission features; in theory, can constrain galaxy formation histories and physical conditions to **factors of 2-3**.



Bayesian Thinking is **Powerful** for Interpreting Galaxy Data

The combination of **tens/hundreds of parameters** and **weakly constraining data** for distant galaxies mean the prior and the data are ~equally important.

This is the 'sweet spot' for Bayesian statistics.



Complex Emergent Behavior Means Key Physics Are Uncertain

Huge scales in space and time (20+ orders of magnitude in both) **limit accuracy of galaxy formation simulations**

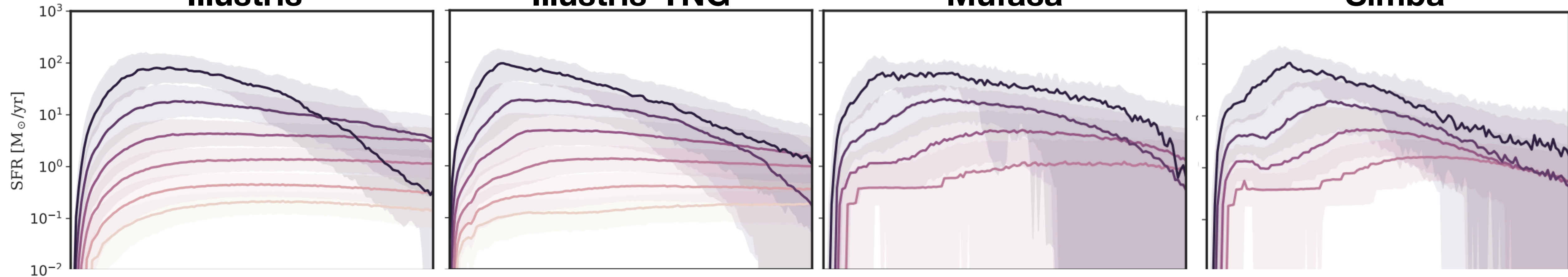
Below are the formation histories of same galaxies, **under different simulation rules.**

Illustris

Illustris-TNG

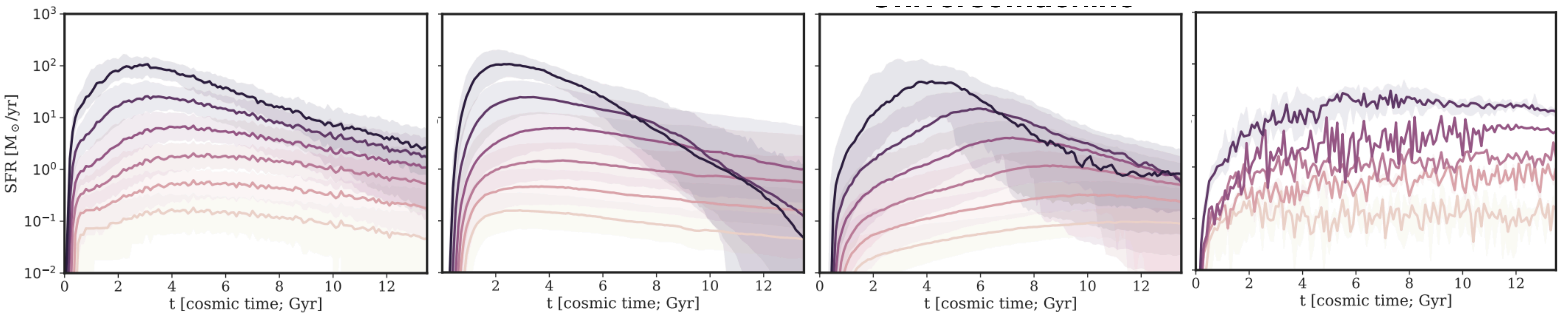
Mufasa

Simba



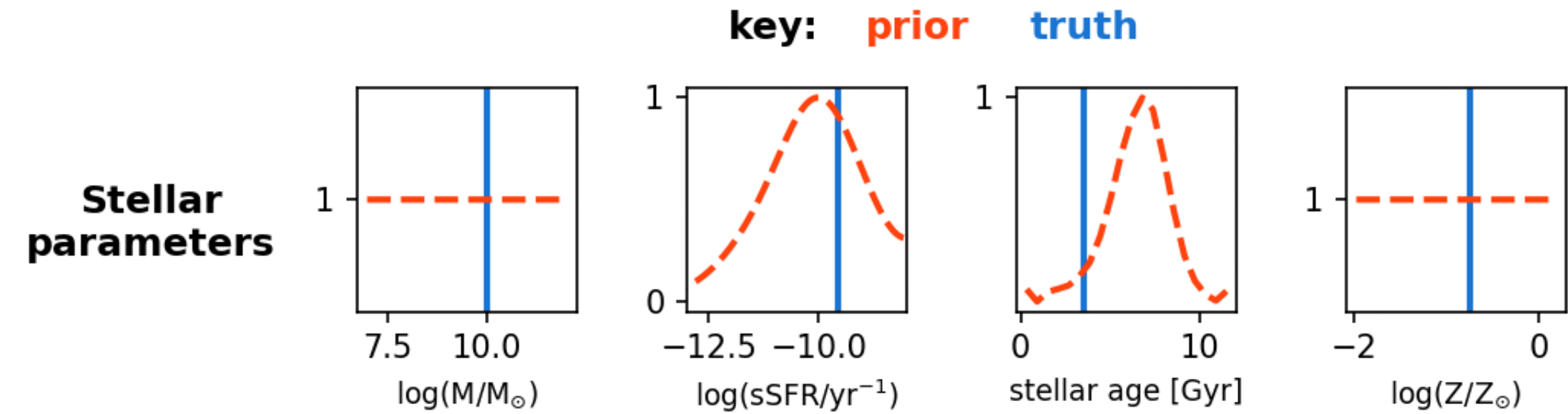
Do real galaxies behave like one of these? None of these?

ly Simulations do not show enough agreement, and don't predict enough important 'microphysics', to use as basis
We combine models of stars, gas, dust, black holes, with physically-motivated priors (guesses!) to make fake galaxies



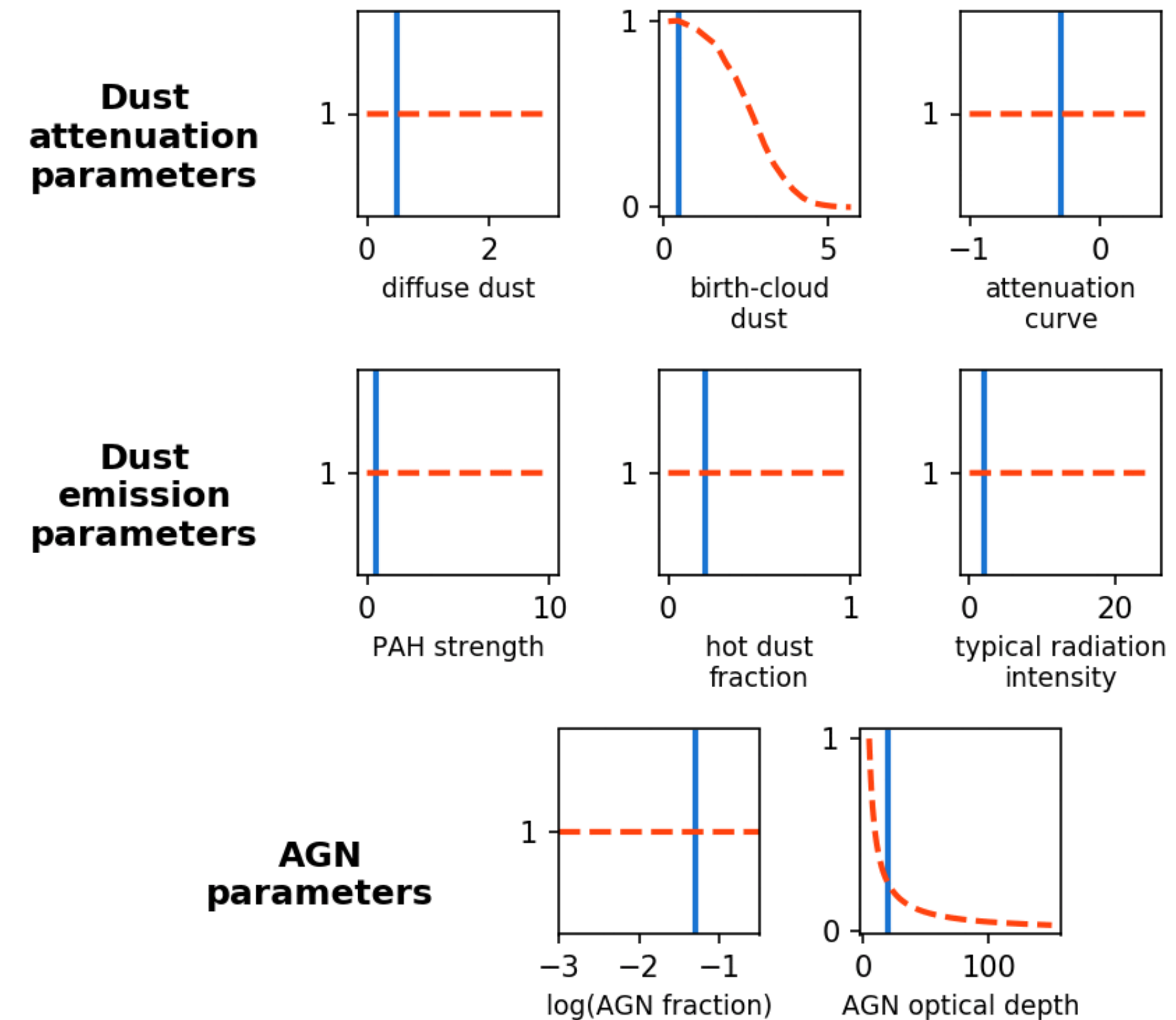
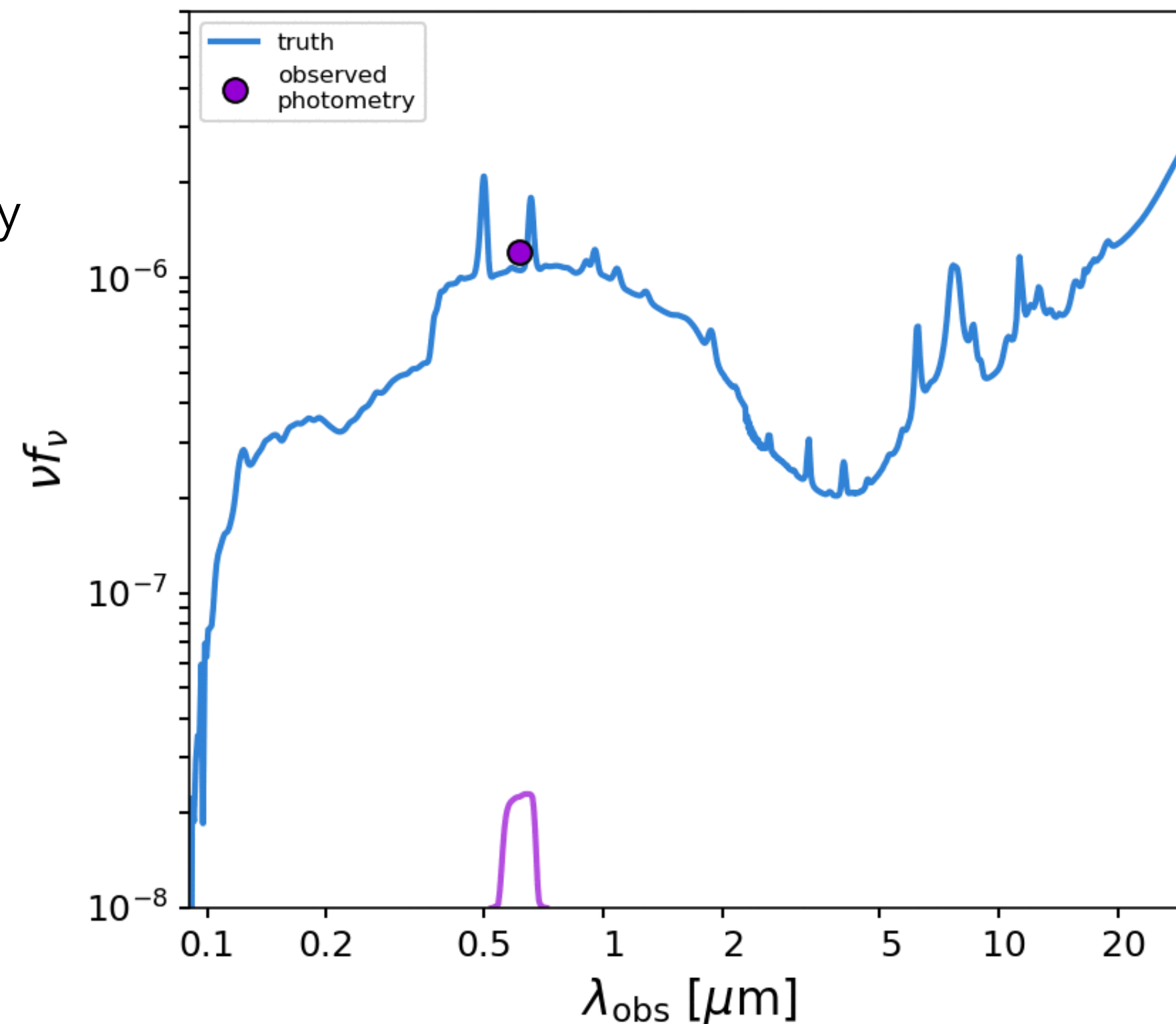
Bayesian Thinking is **Powerful** for Interpreting Galaxy Data

optical: SDSS *r*



Prospector
Open-source galaxy
SED-fitting code

Johnson, Leja+21



The Universe can be **arbitrarily complex** - yet as long as we can simulate the correct physics, we can marginalize over those physics.

Can we do that?

Model Misspecification is Key Problem for Distant Galaxies

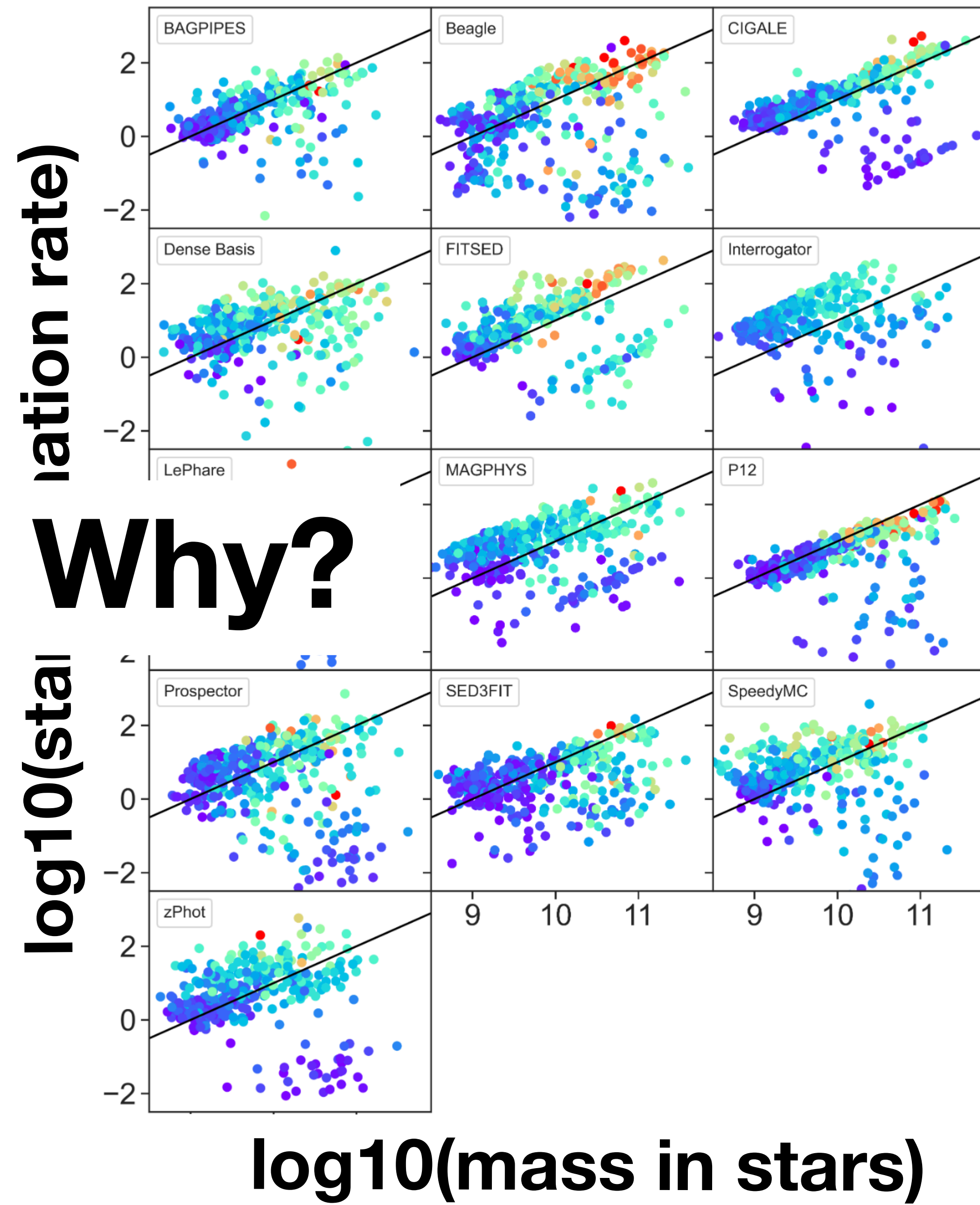
Galaxies have no
'ground truths'; not like
e.g. weather prediction!

Popular, different
modeling codes
applied to...

...**identical** high-
quality space-
based images...

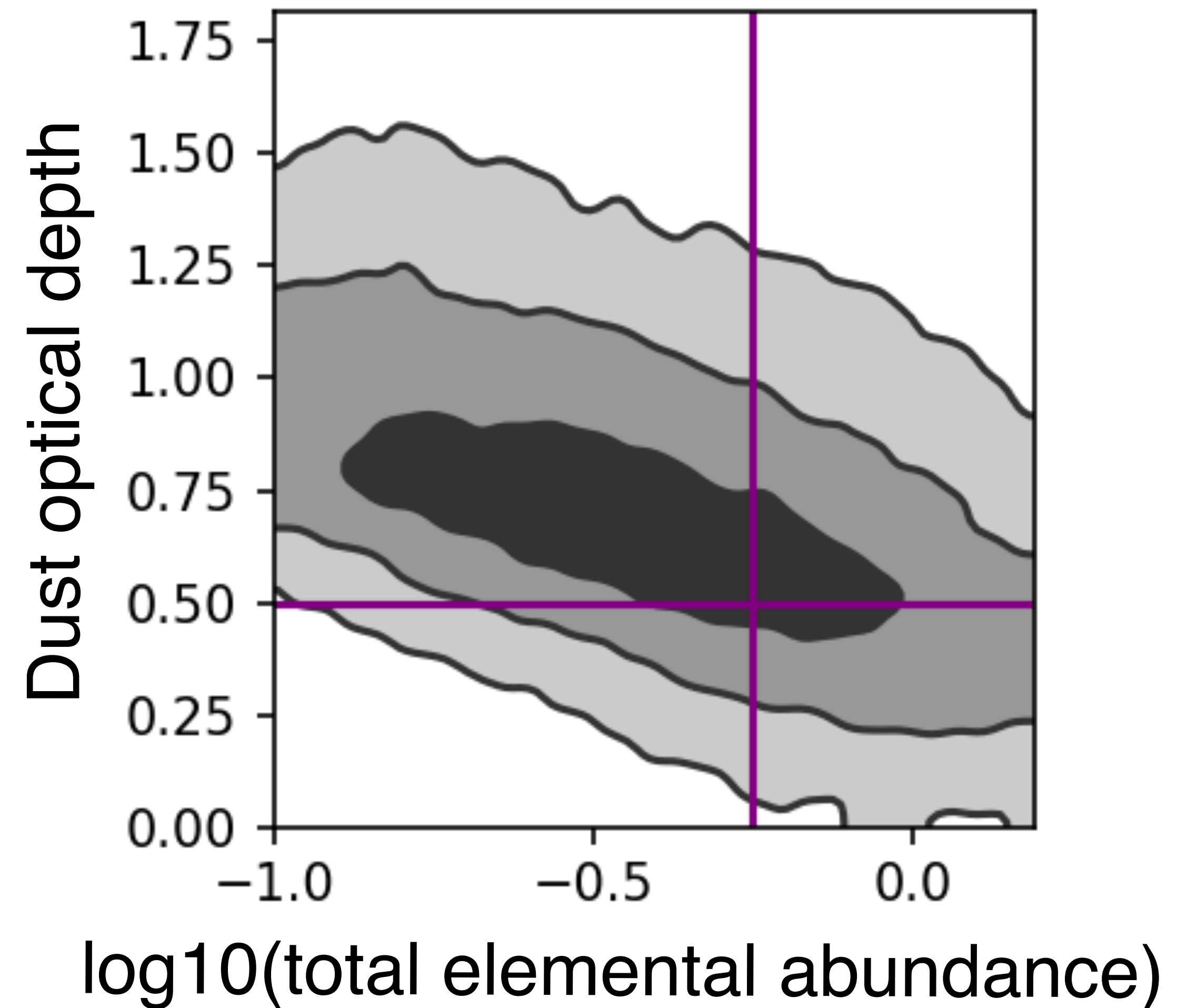
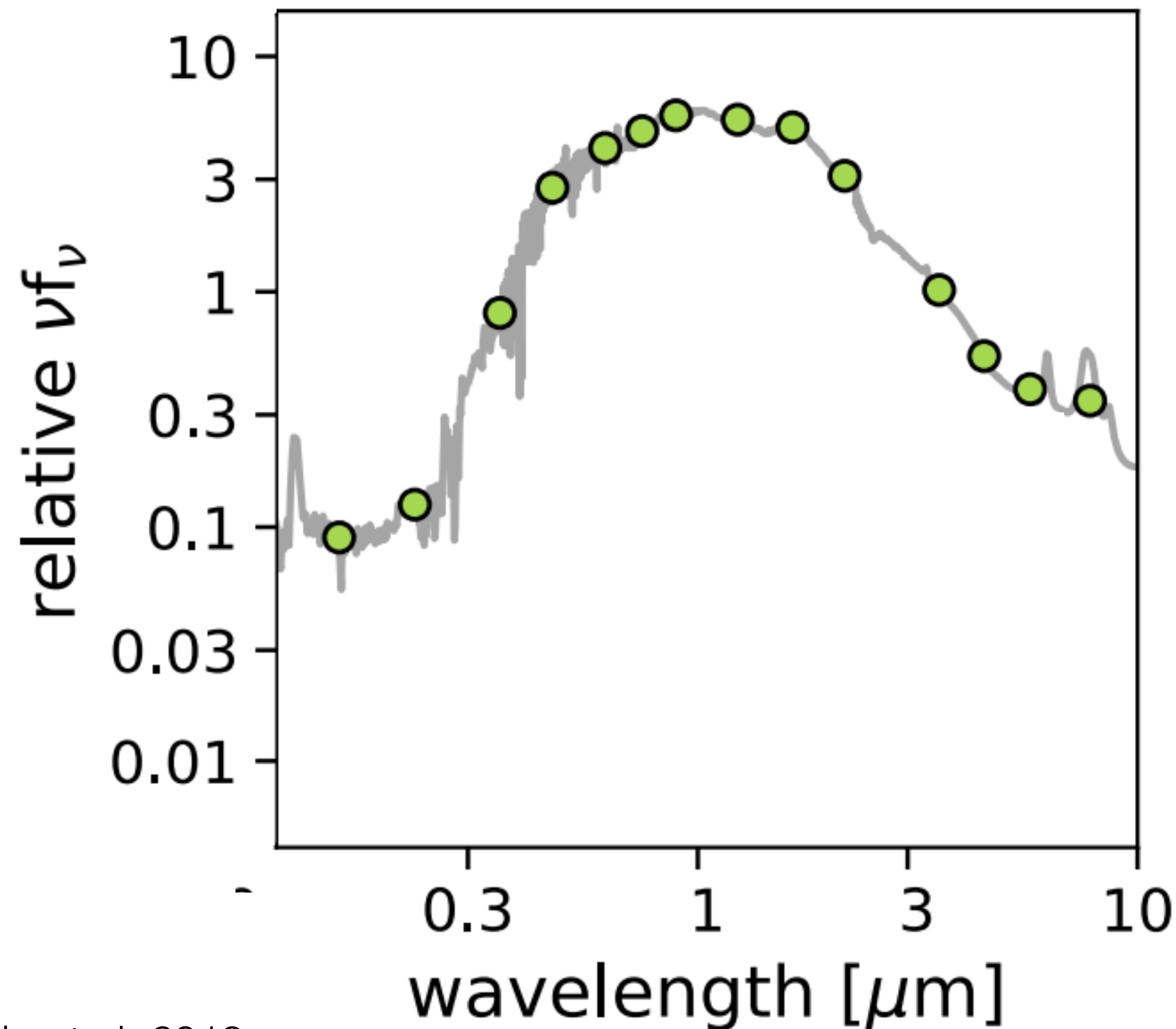
... produce
qualitatively and
quantitatively
different galaxy
populations!

Pacifici et al.
2023 (+Leja)



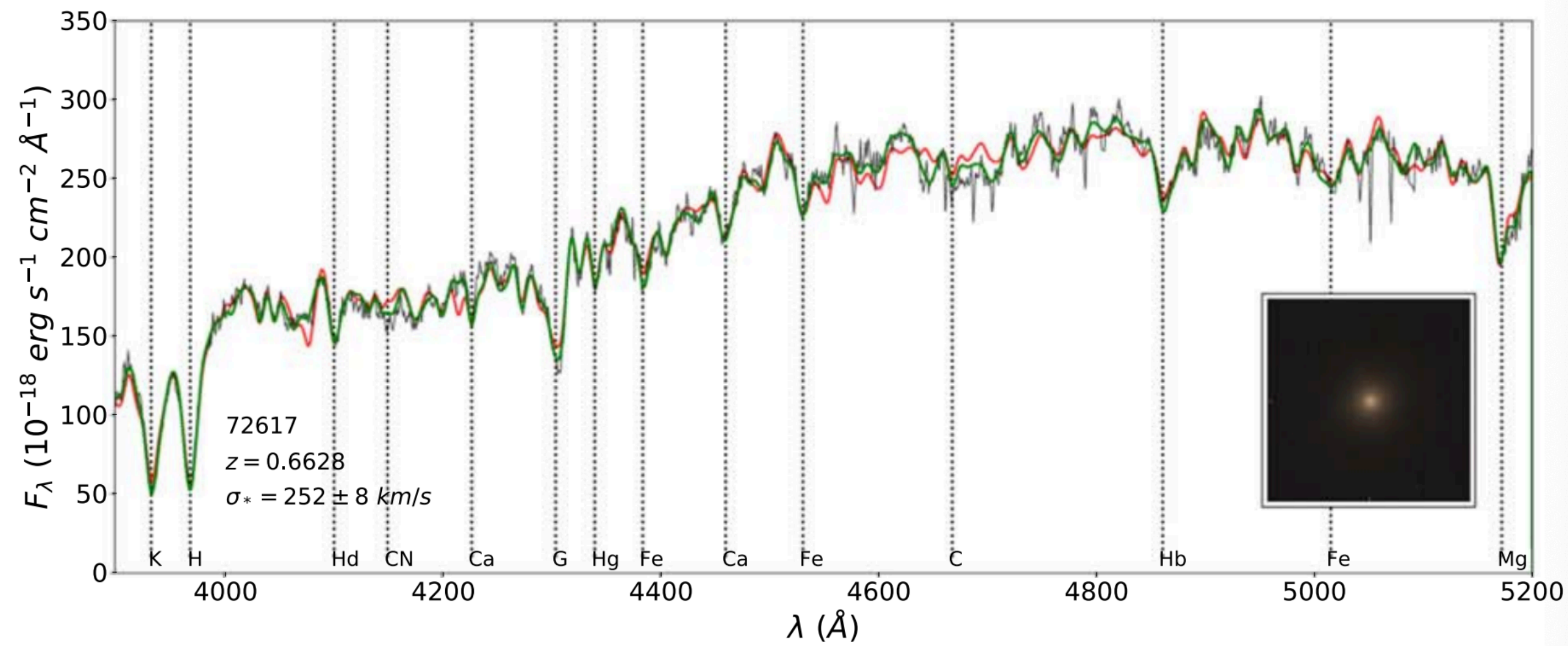
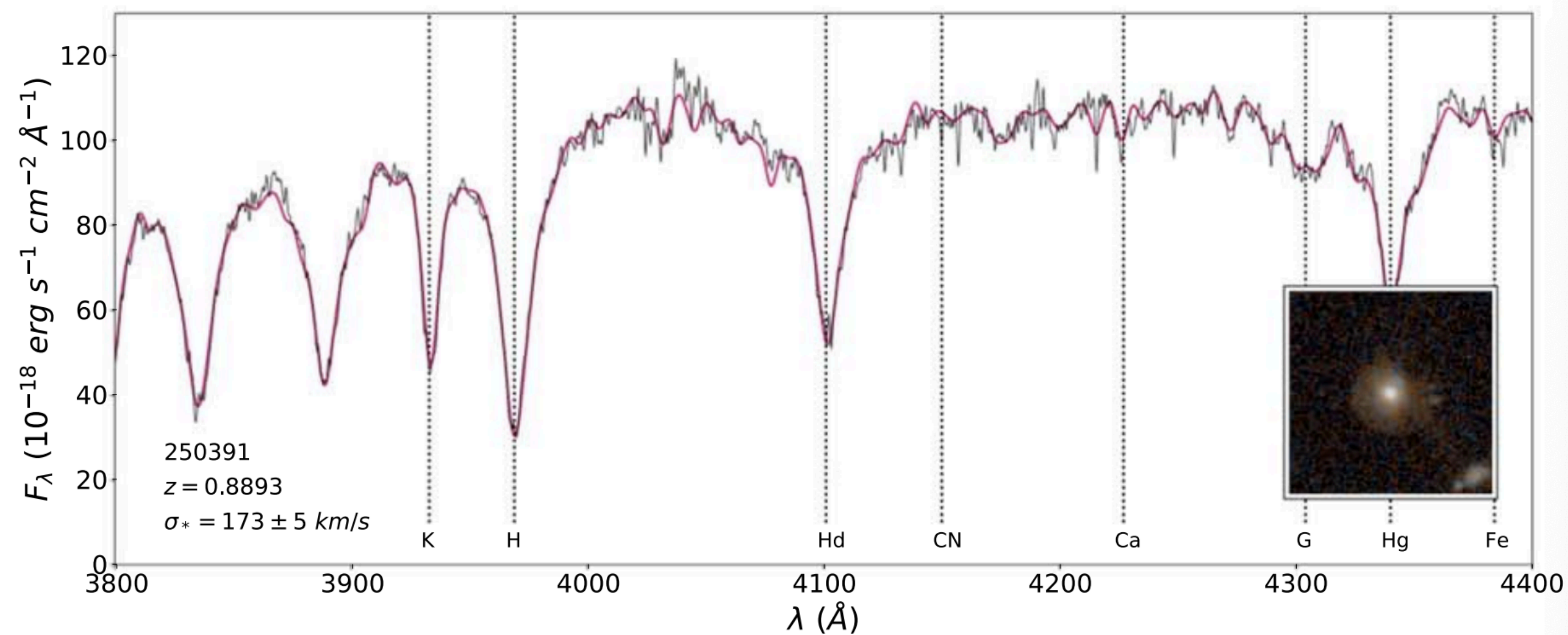
Most Parameters are Degenerate, Especially With Only Imaging

Star formation history, chemical abundances ('metallicity'), dust have **similar effects on imaging**
→ limited to **mapping out inherent degeneracies** or adopting **informed priors from better data!**



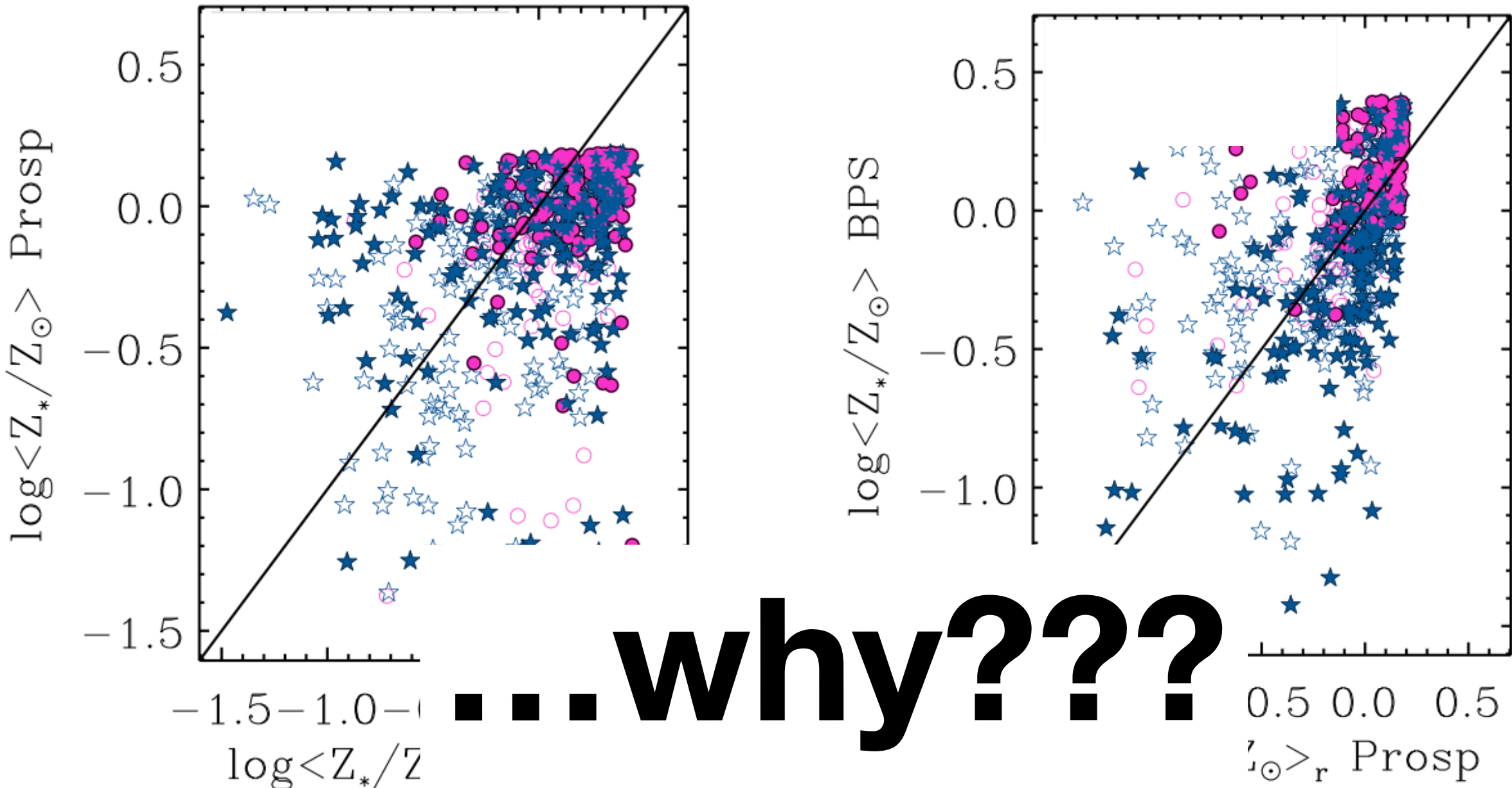
Imaging Isn't Enough. What About The “Best Distant Galaxy Spectra We Have”?

The LEGA-C survey: ~3500 galaxies existing 6-8 billion years ago, each with **~20 hour spectrum** from an **8 meter telescope (VLT)**, yielding **signal to noise ~20-70** per Angstrom (1/10th of a nanometer)



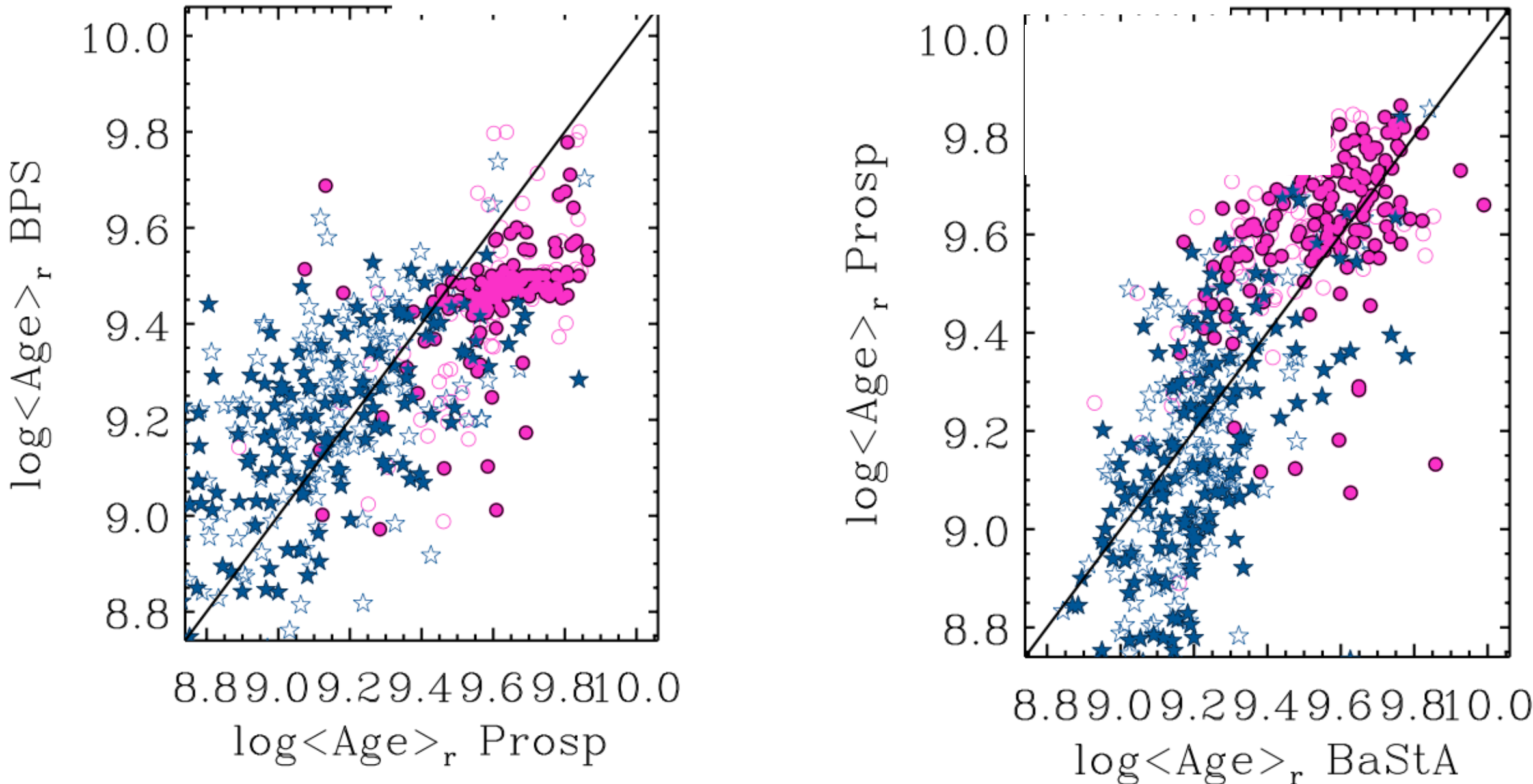
Comparing Inferences Leading Inference Codes Applied To High-Quality Spectra

**Abundance
of Elements
in Stars**



...why???

**“Light-weighted”
age**
(Mass-weighted not
shown; much worse)



Gallazzi et al.
submitted (incl
Leja), 2025

Codes: Bagpipes
(Carnall+), Prospector
(Johnson,Leja+), BaStA
(Zibetti & Gallazzi)

We Do Not Include Enough Physics To Capture the Complexities of Galaxies

Many different physics in play, forcing approximations/assumptions and creating model misspecification.

Physical Parameter	approximate effect on data
stellar mass (ad infinitum)	orders of magnitude

Bayesian Inference: Powerful but Dangerous

"All models are wrong - but some are useful."

(Kyle Cranmer beat me to it)

What does "useful" mean? We create new knowledge!

But - must **understand data** (measurement uncertainties, instrumental effects, ...)

Must have **reasonably accurate generative models** for what we're fitting (stars, gas, black holes, ...)

If **not true** - careful! *Modeling can do more harm than good.*

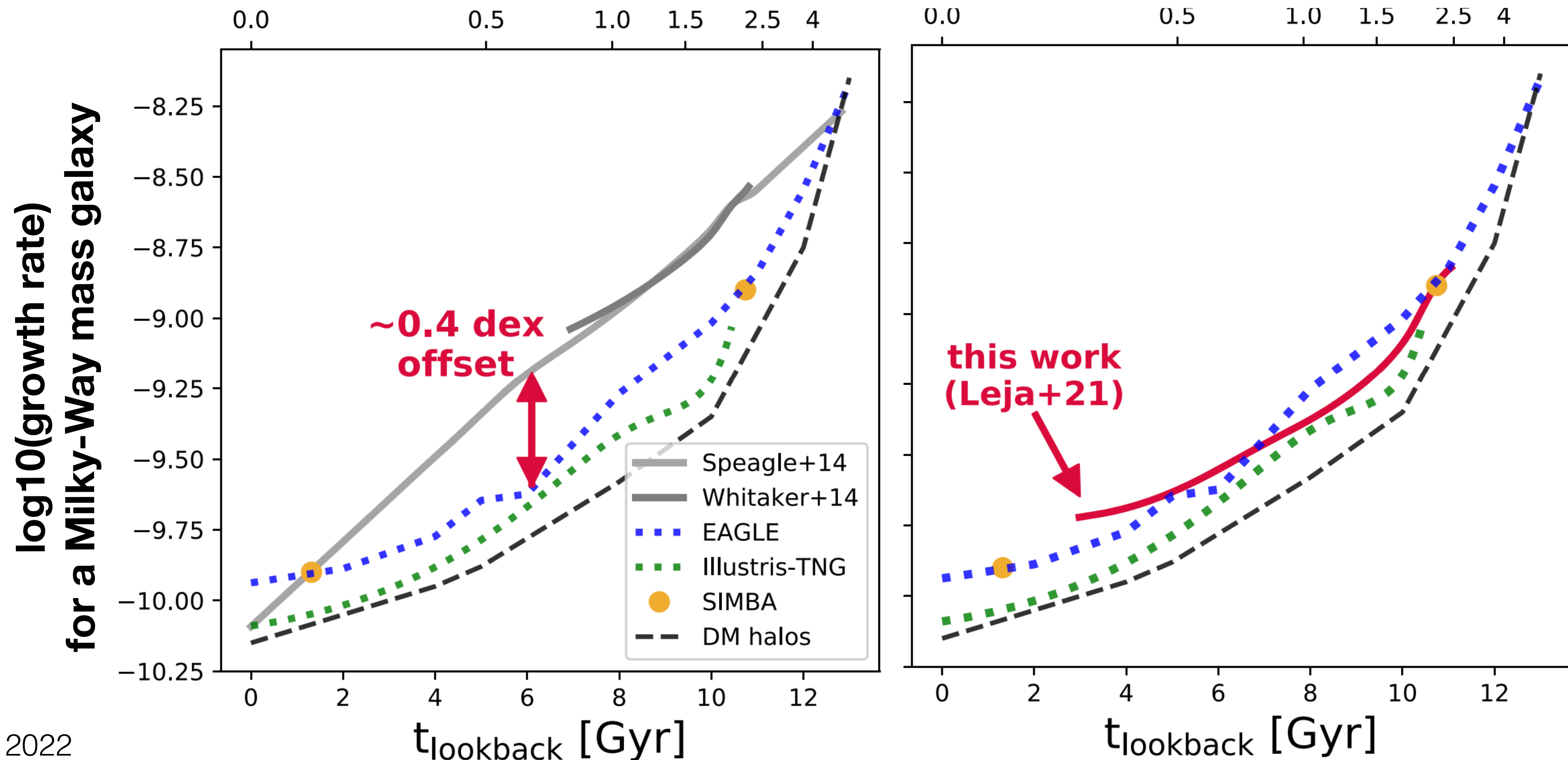
So, are our models accurate enough to be useful?

Adding More Physics Can Solve Big Problems!

Straightforward interpretations of ultraviolet, infrared images left a long-standing mystery: galaxies form too many stars at early times

More physics: 15-parameter forward-model solves, cosmic star formation rate now agrees with simulations!

But - this was computationally expensive, and many puzzles remain



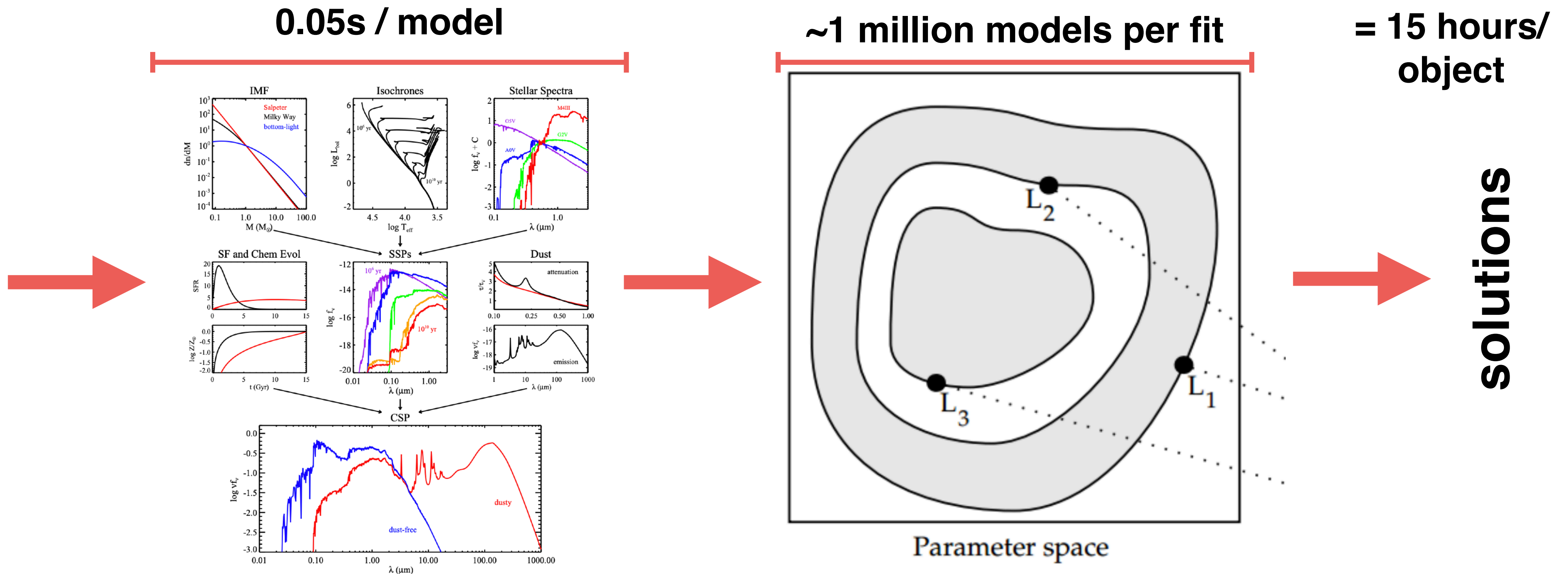
High Dimensionality Already Pushing Computational Limits

High-dimensional models cannot live on a grid (*curse of dimensionality*): this means each model must be generated **on-the-fly**, a compute-intensive task (~15 hours/fit!).

- **several million CPU-hours** to analyze $\sim 10^5$ objects in deep field

What drives computational requirements?

Galaxy observations

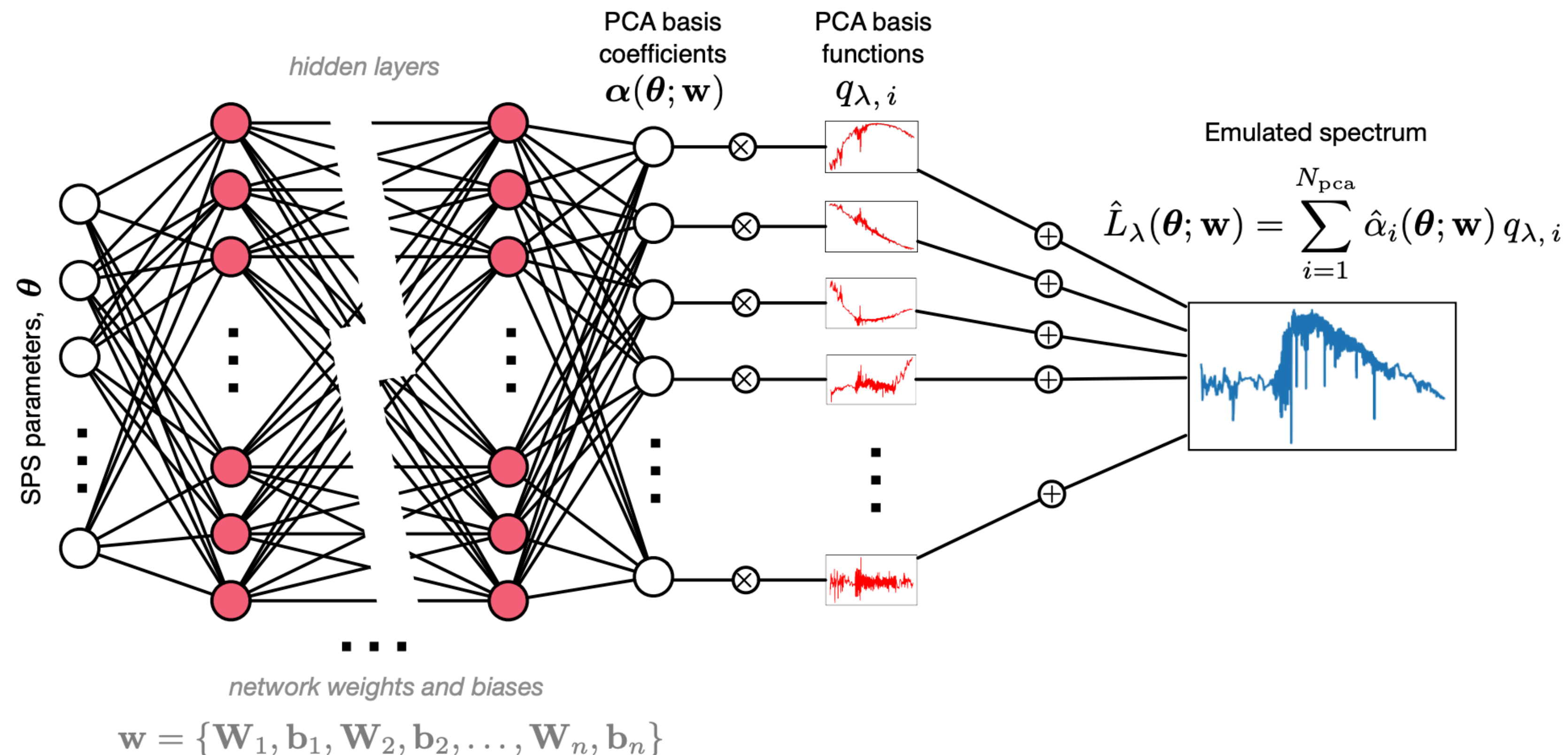


High Dimensionality Already Pushing Computational Limits

High-dimensional models cannot live on a grid (*curse of dimensionality*): this means each model must be generated **on-the-fly**, a compute-intensive task (~15 hours/fit!).

- **several million CPU-hours** to analyze $\sim 10^5$ objects in deep field

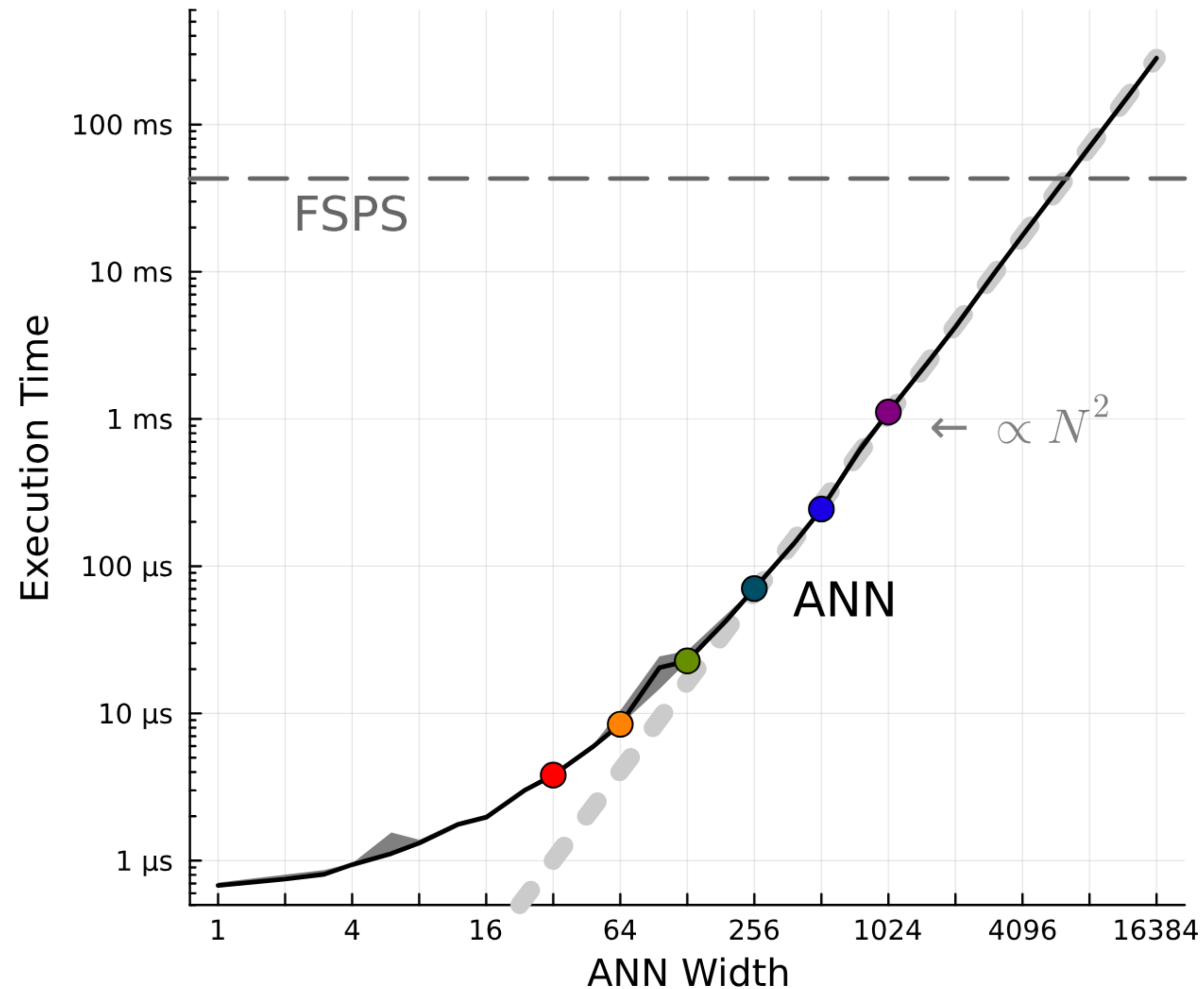
Neural net emulation of **imaging predictions** reduces model generation time by **~ 100 - 1000 (10^4 on a GPU)**



Alsing, Peiris,
Leja et al.
(2020)

Mathews, Leja
et al. (2023)

Neural networks are *fast* (by trading off precision)



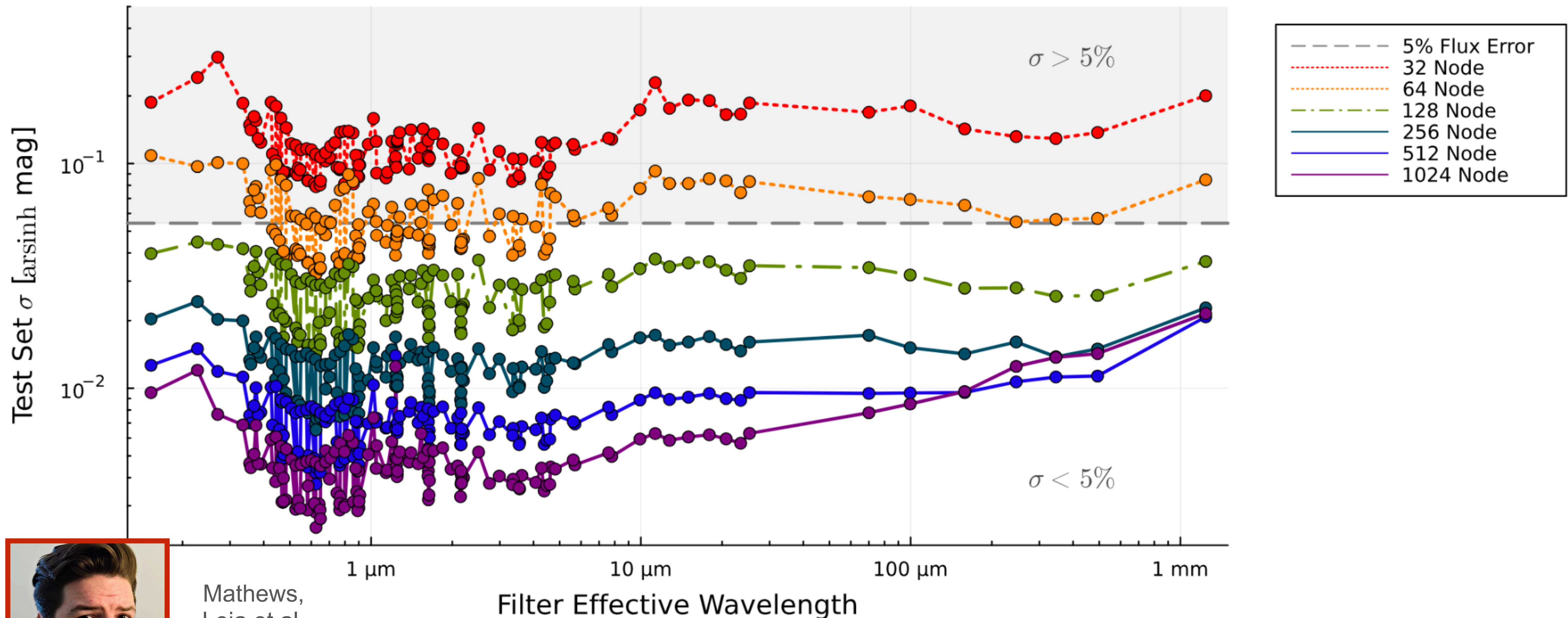
FSPS = classical stellar
population
synthesis (Conroy et al. 2009)

ANN = artificial neural network



Mathews,
Leja et al.
(2023)

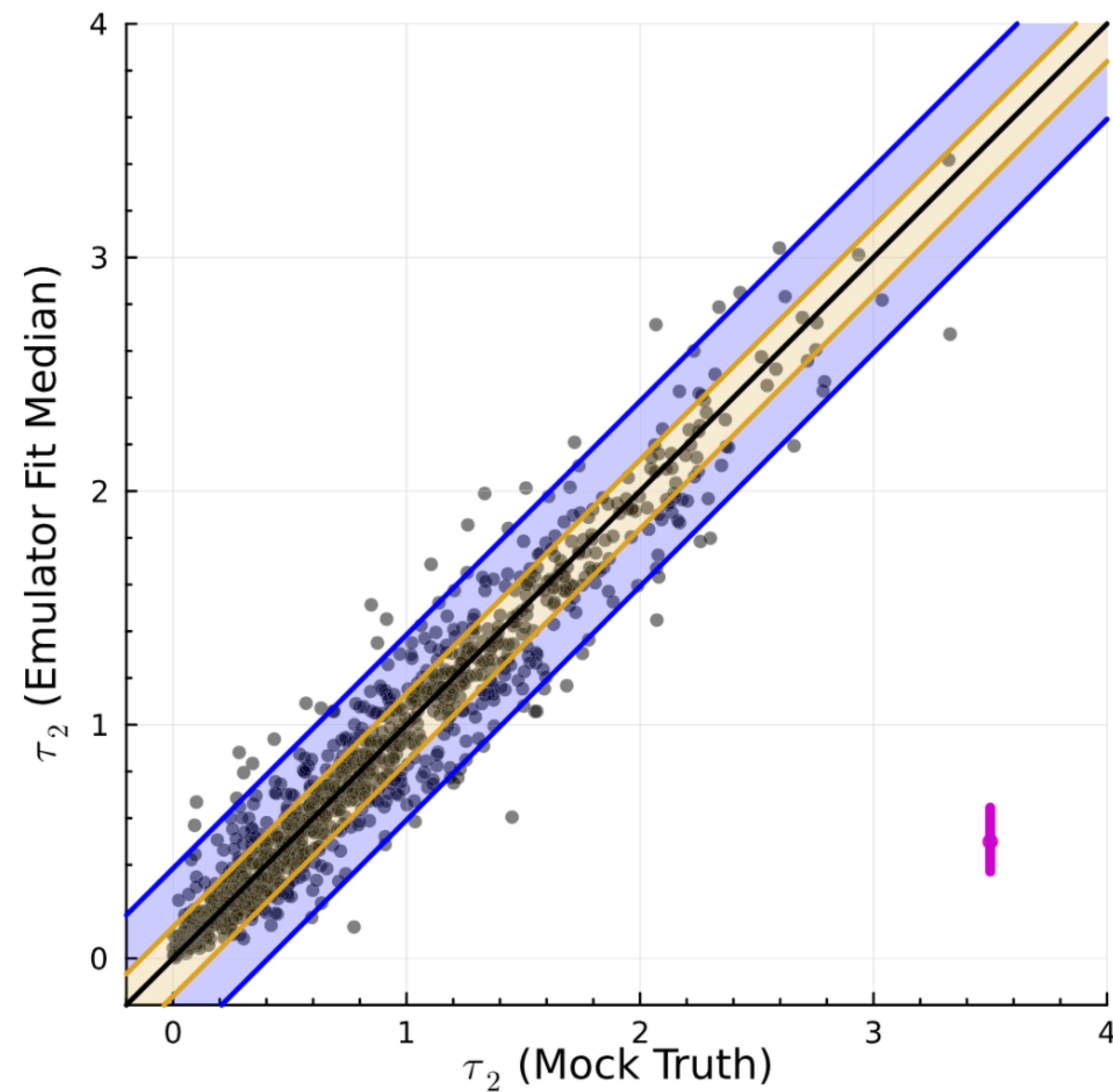
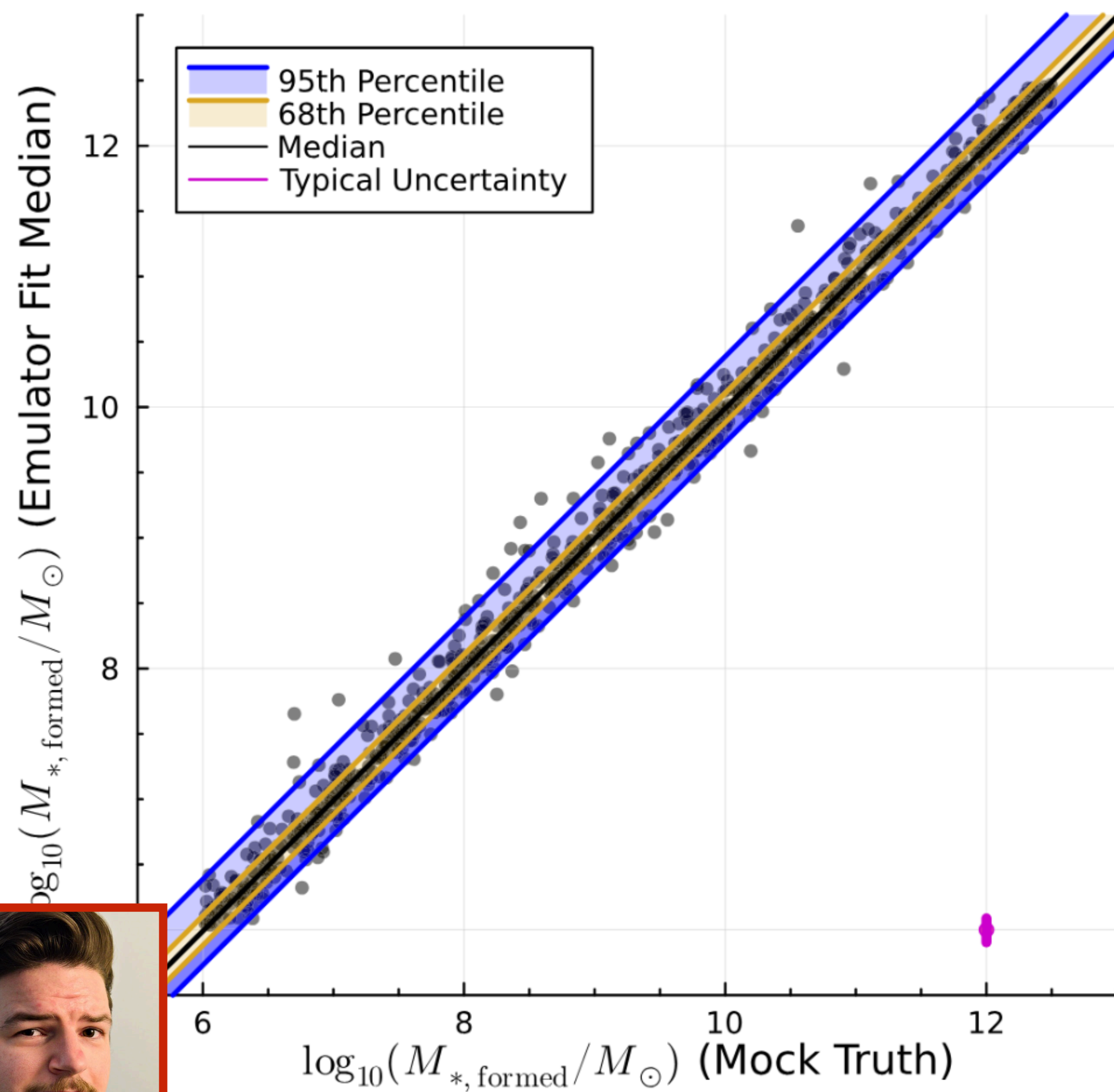
Neural networks are ***precise*** (if they're big enough; $<[0.5 \times \text{obs error}]$ works)



Mathews,
Leja et al.
(2023)



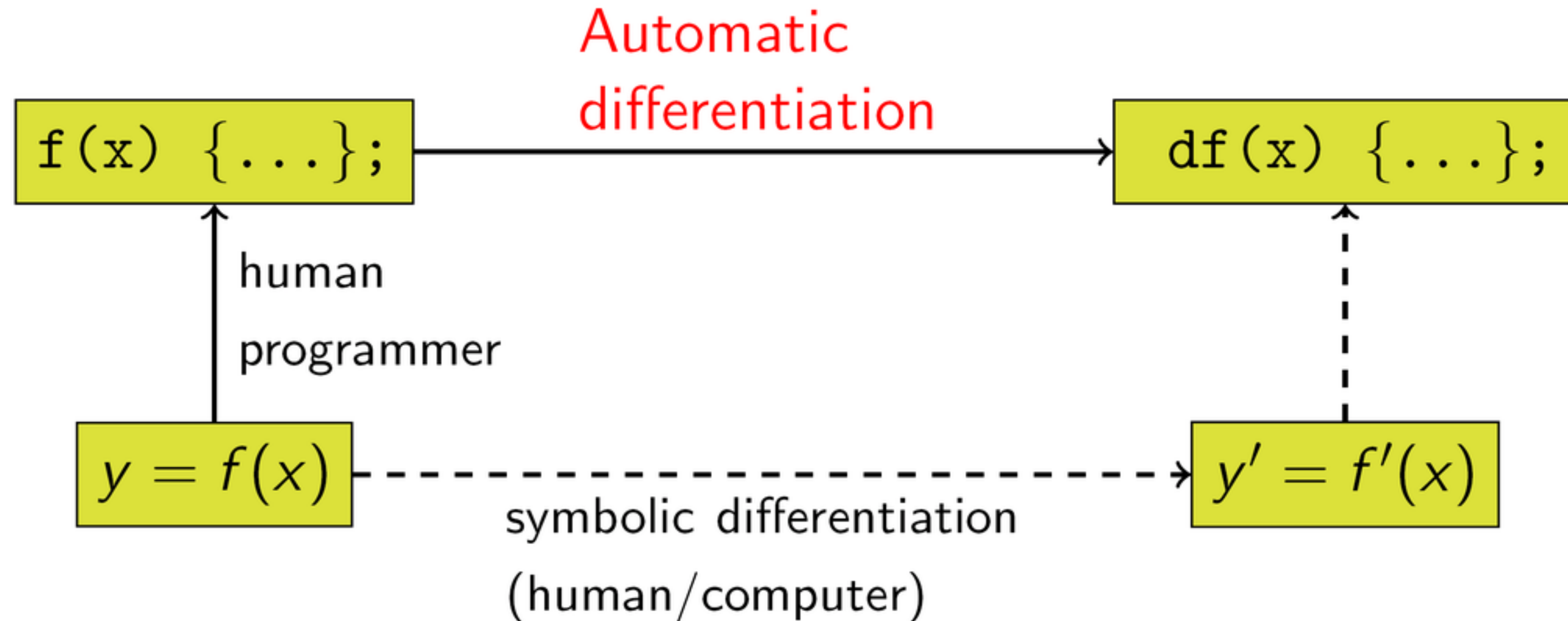
Neural networks *give accurate posteriors*



Mathews, Leja et al. (2023)



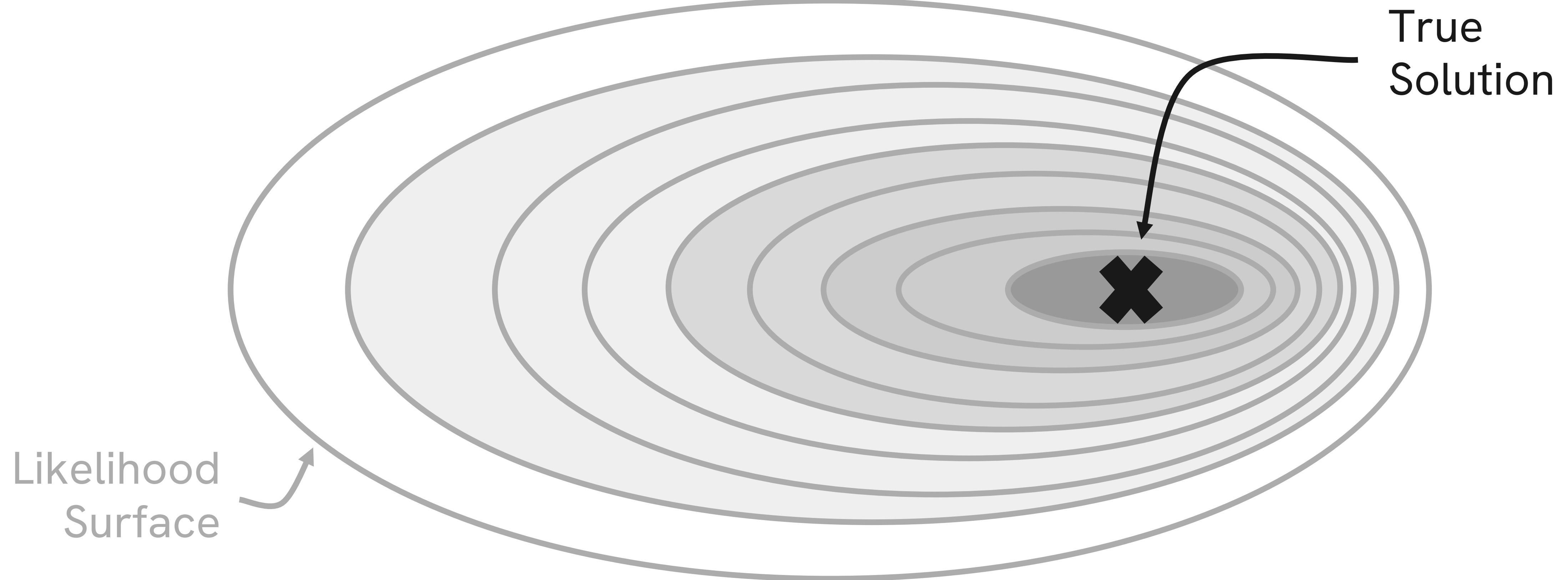
Neural Networks are *differentiable*



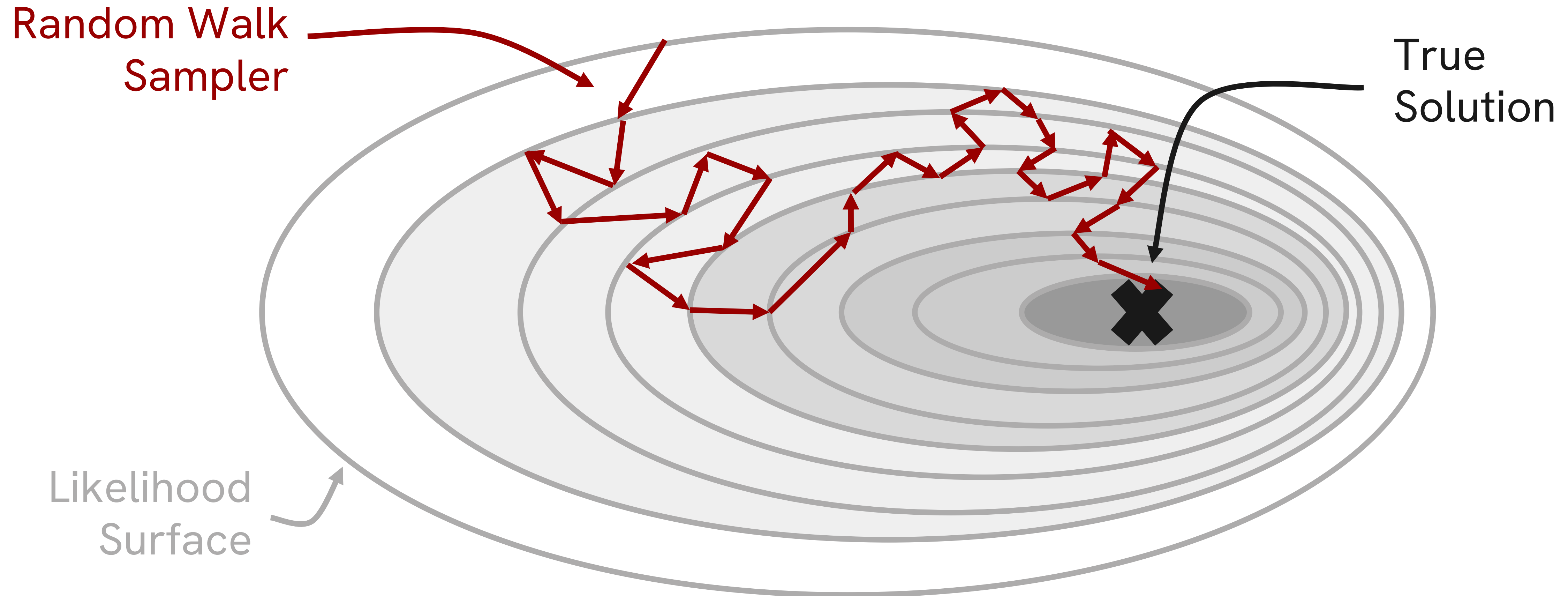
(from Wikipedia)

This opens the door for using **gradient-based samplers** such as Hamiltonian Monte Carlo (HMC) or the No-U-Turn Sampler (NUTS)

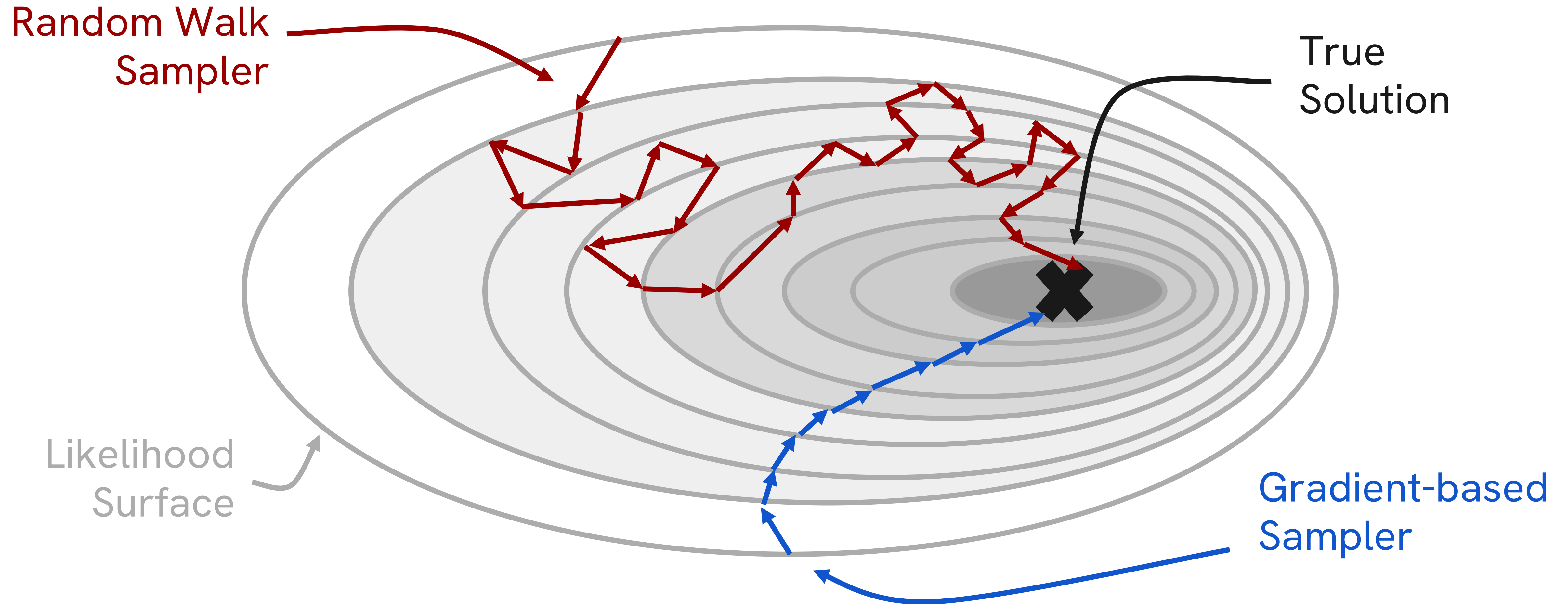
Differentiability Allows Highly Efficient Sampling



Differentiability Allows Highly Efficient Sampling



Differentiability Allows Highly Efficient Sampling



Gradient-based samplers are efficient at proposing good samples.

Smart analysis of deep fields - sort through tens of thousands of sources to find most distant galaxies!

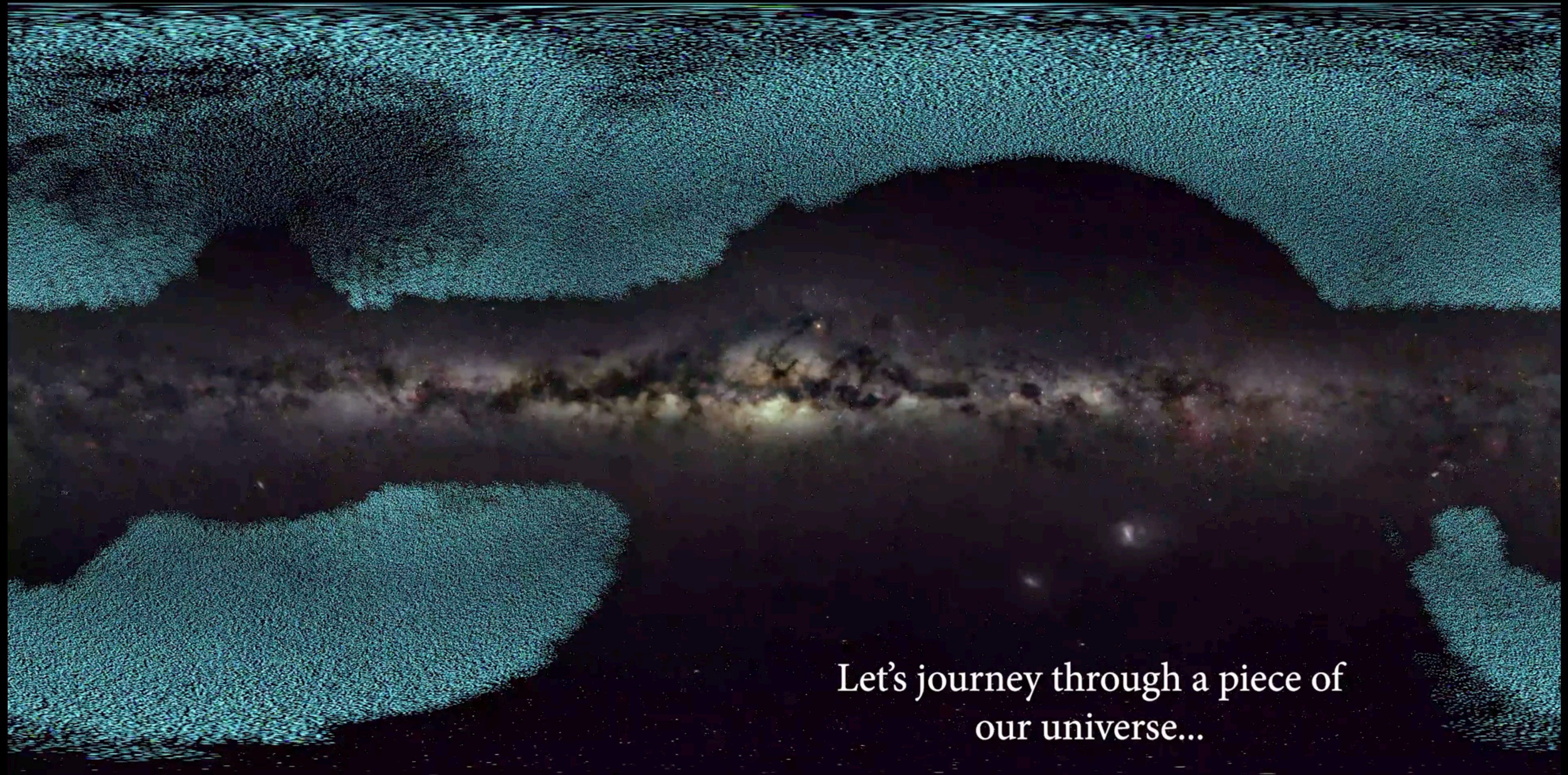
$z = 13.20$

Neural net emulators let us do “normal” science faster!

But also - what about new directions?



This decade, there are multiple large-scale spectroscopic campaigns to **map large fractions of the observable universe** with spectroscopy of distant galaxies, with the goal of understanding **dark matter**, **dark energy**, and **galaxy formation**.



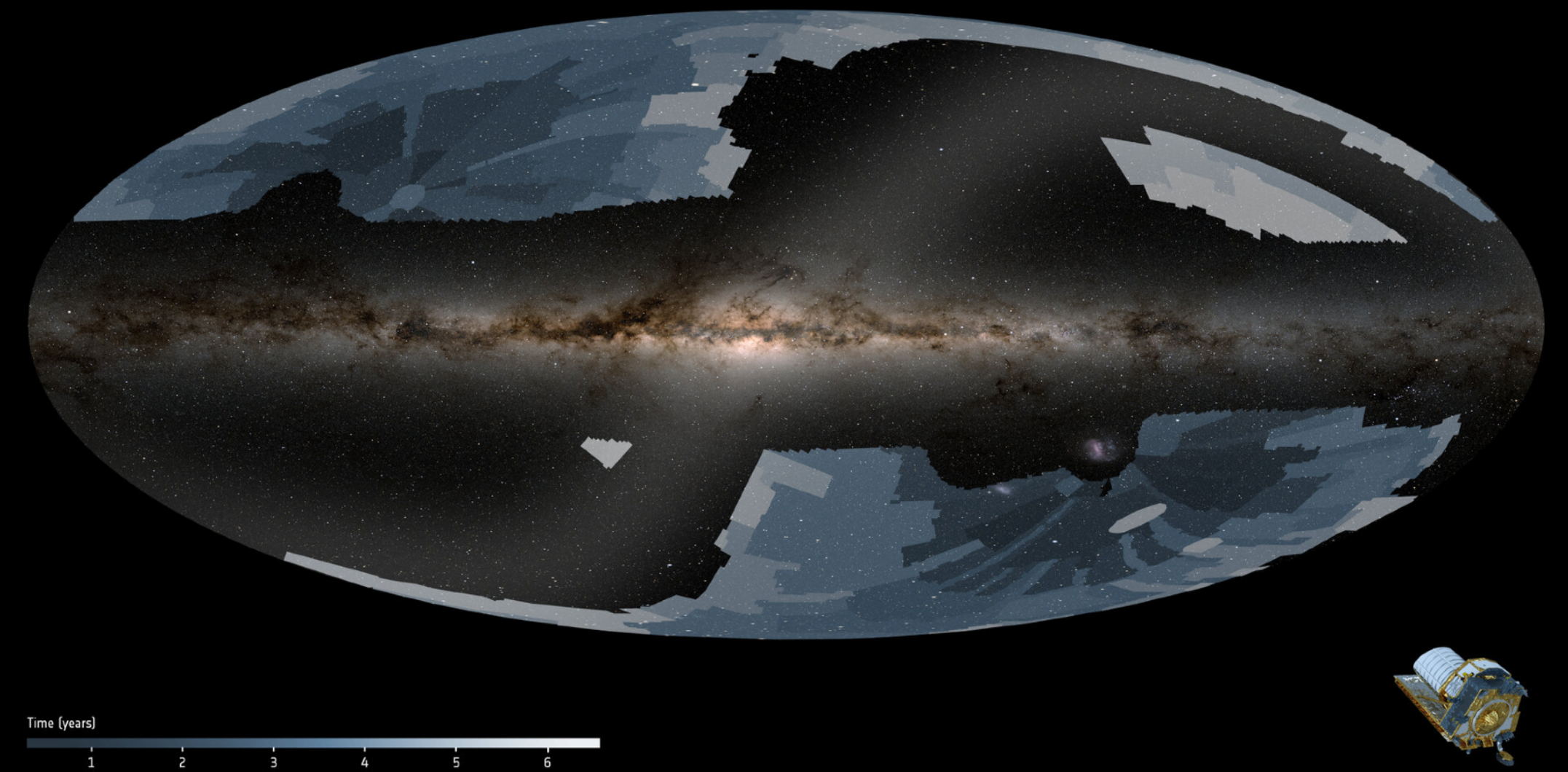
Let's journey through a piece of
our universe...



DESI
~50M galaxies
2021-2026

Euclid
~35M galaxies
2025-2031

esa



PFS
~5M galaxies
2025-2030

Roman
>100M galaxies
2027-2032

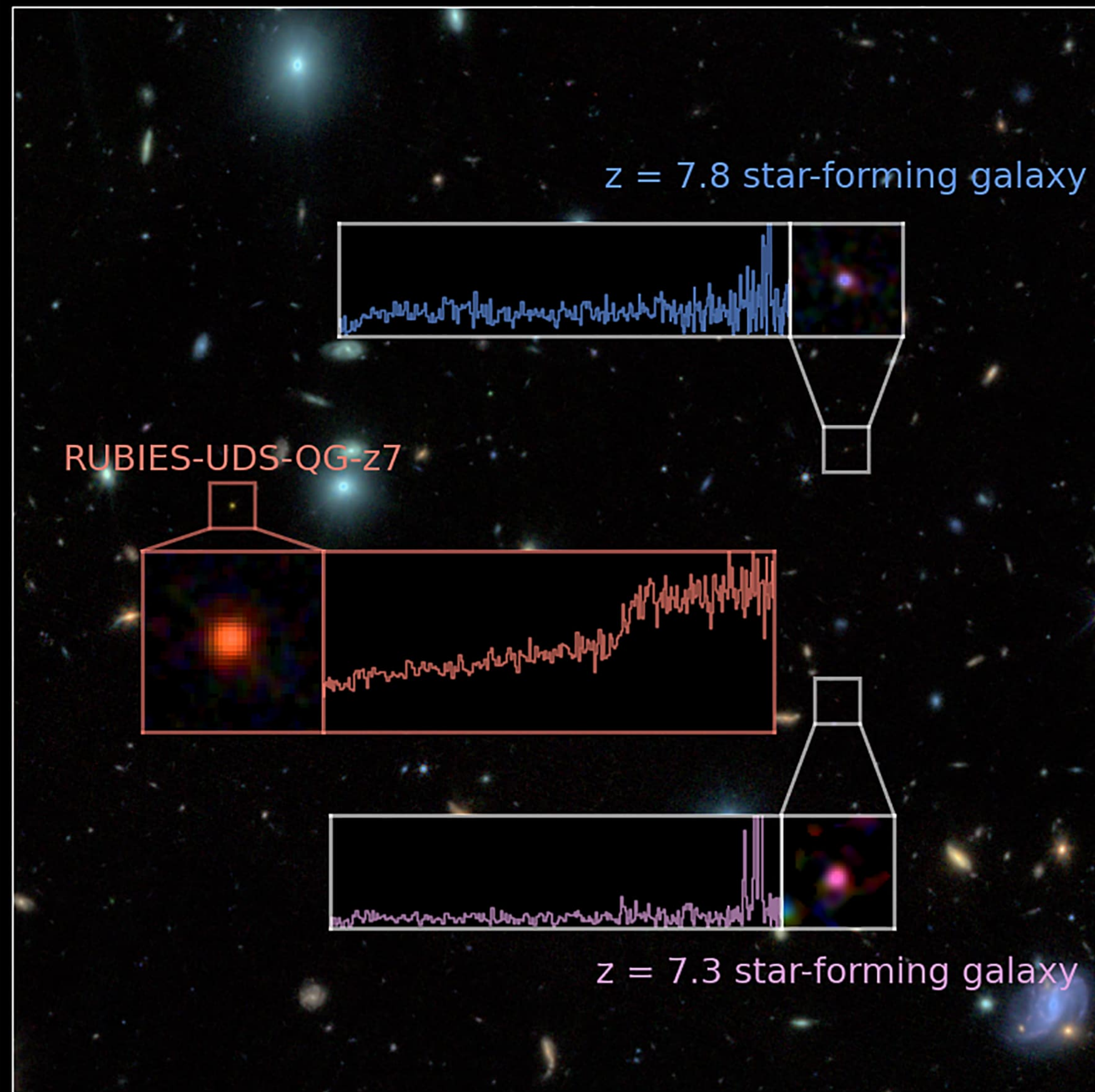


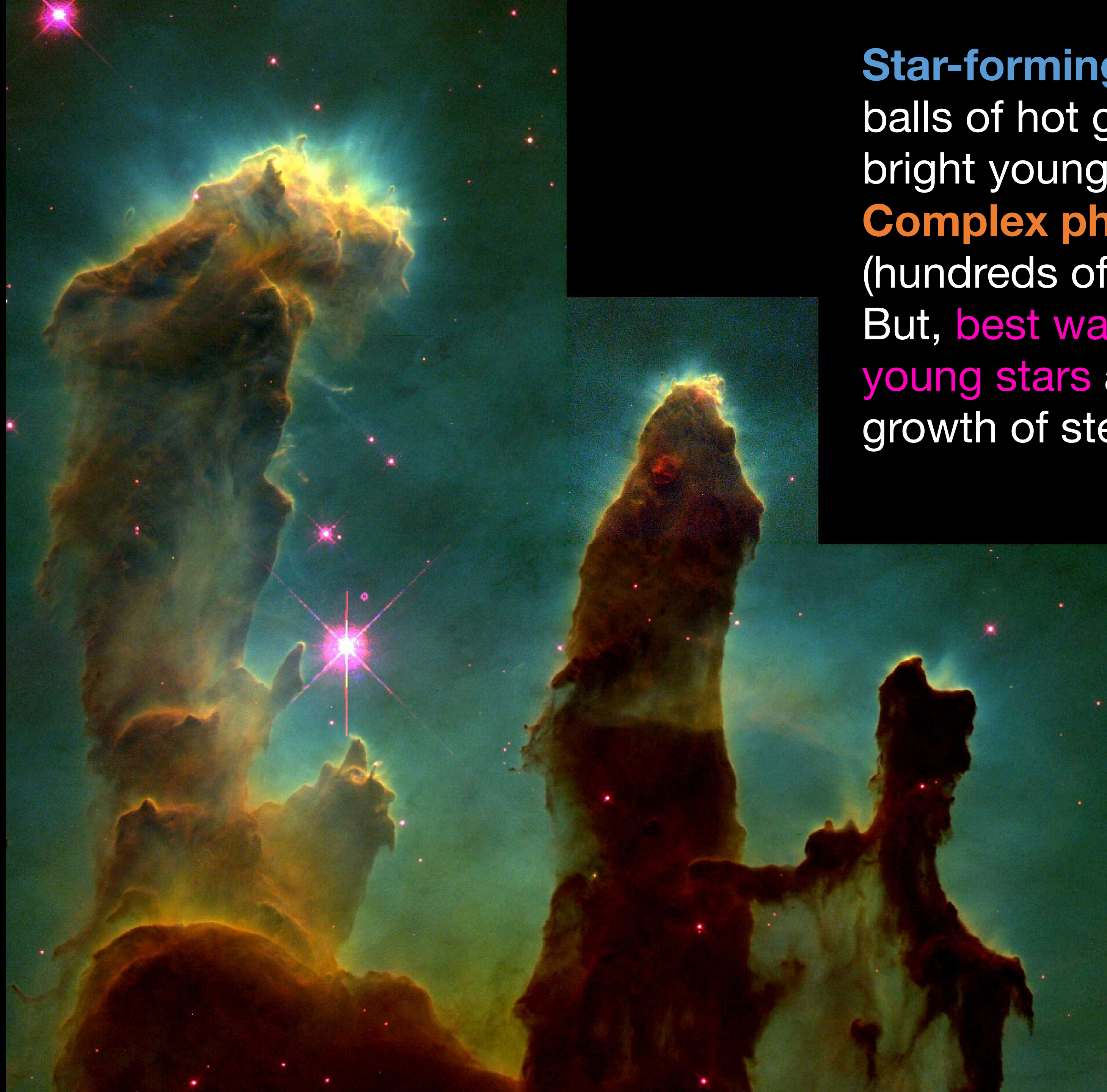
What about Spectra?

Much richer information! But:

- Far **more data** (~ 10 images $\rightarrow$ 1000s of pixels);
- **more detailed input physics** (e.g. complex gas physics; ~ 10 parameters $\rightarrow$ 30-40 parameters);
- need many more **nuisance parameters** (flux calibration, broadening, outlier modeling, correlated noise, ...)

“Classic” nested sampling fits take $\sim$ a few days to finish!





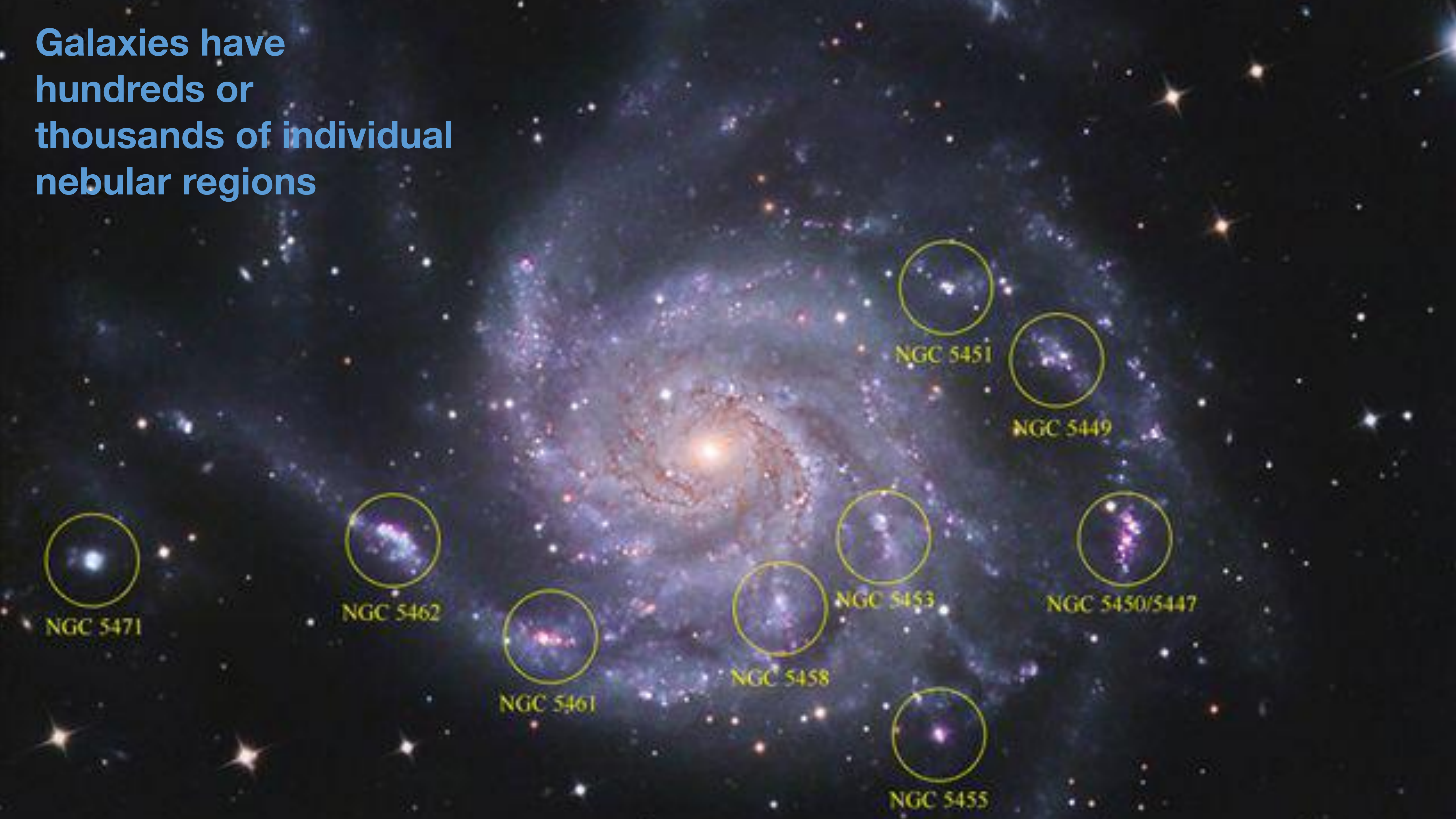
Star-forming regions: huge balls of hot gas, ionized by bright young stars.

Complex physics

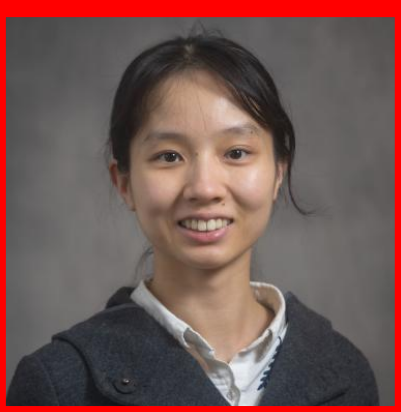
(hundreds of parameters!).

But, **best way to “count” young stars** and understand growth of stellar systems.

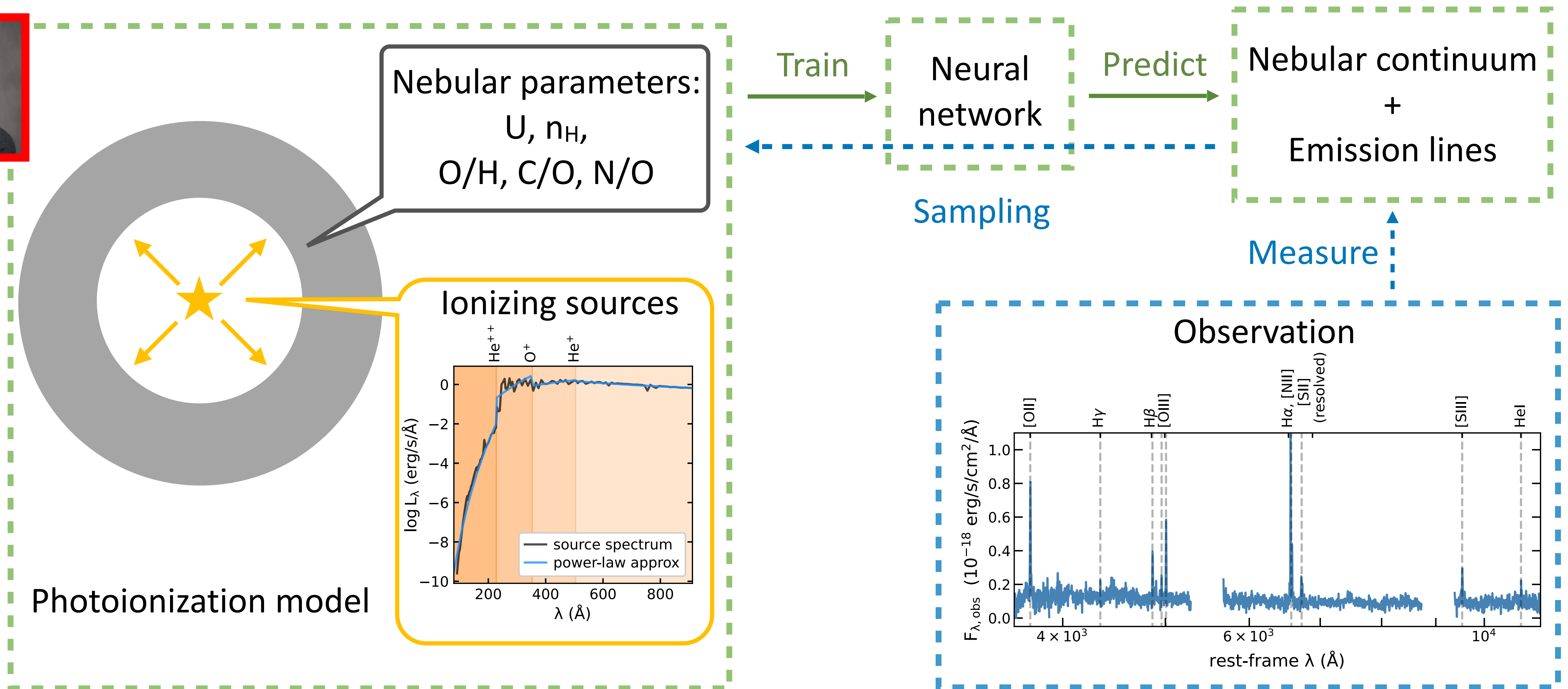
**Galaxies have
hundreds or
thousands of individual
nebular regions**



Cue: fast, flexible neural net emulator for nebulae



Li, Leja+
24ab



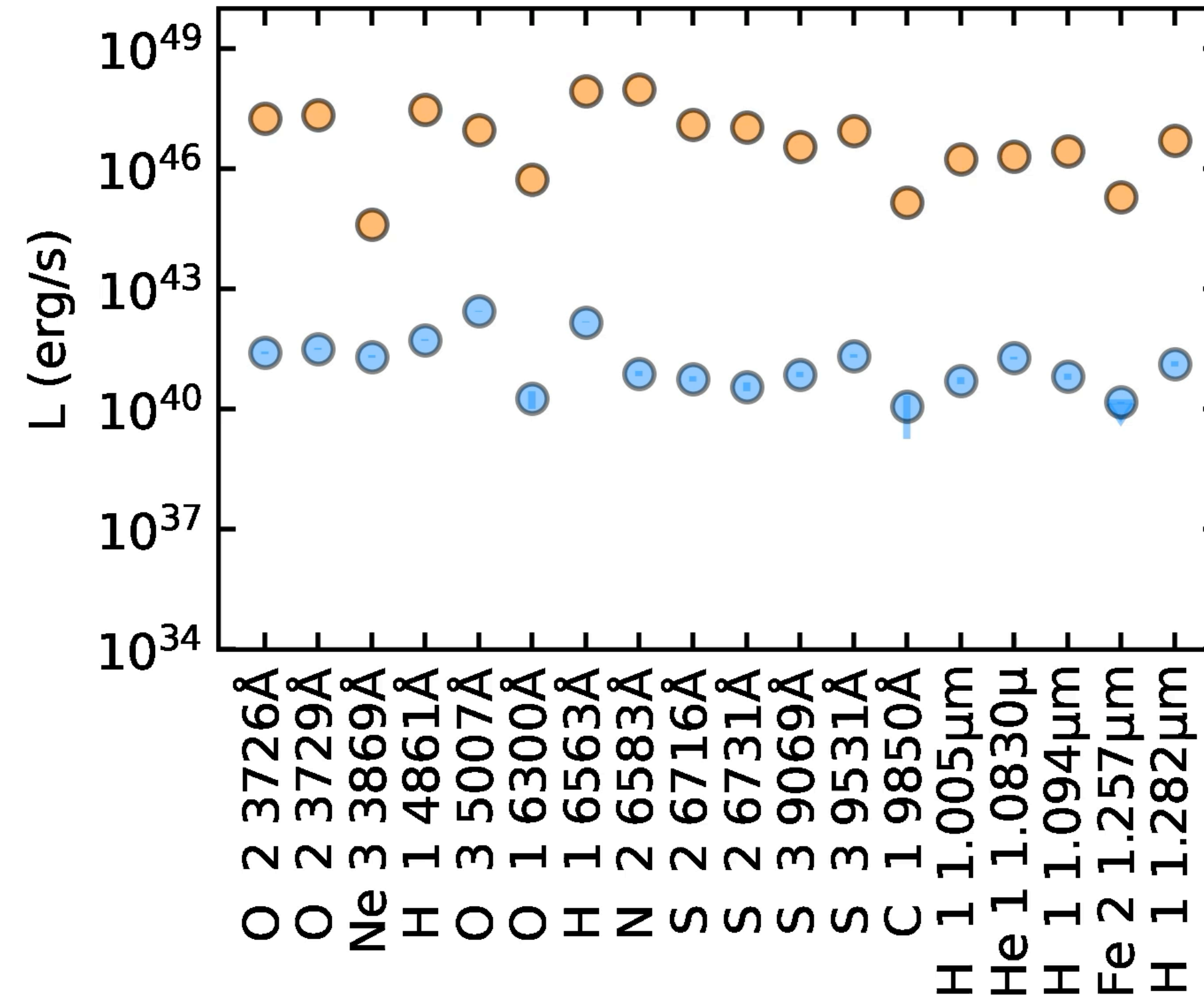
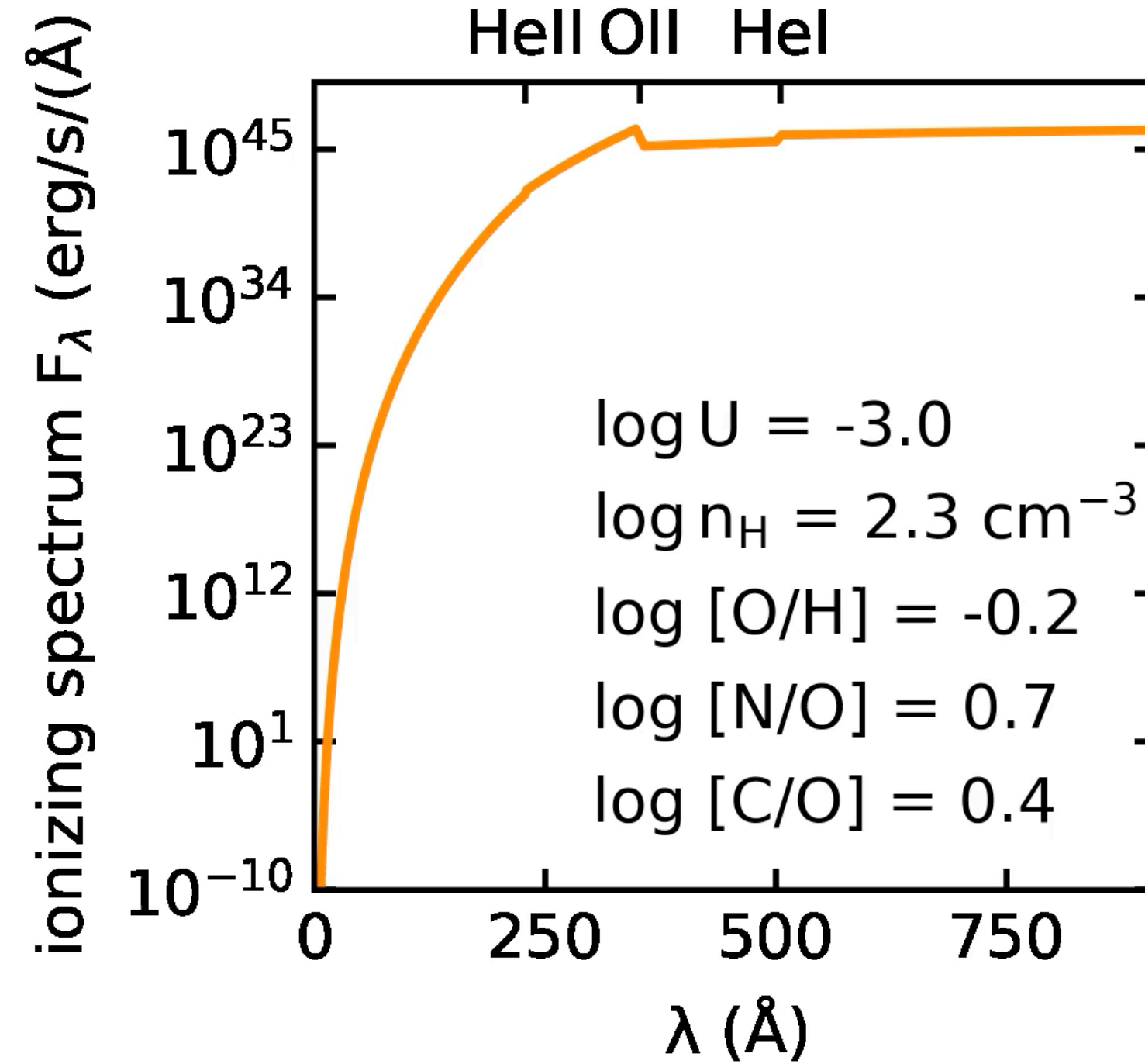
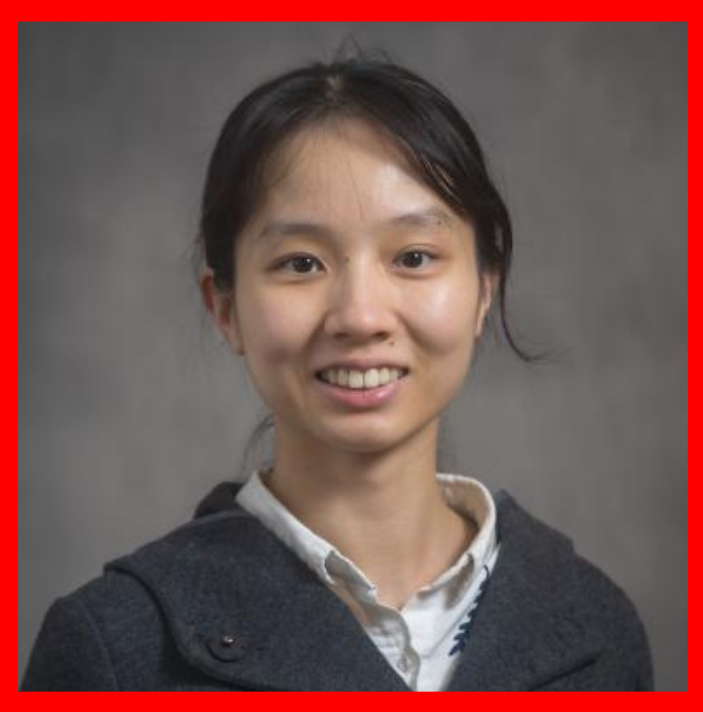
- **Source-agnostic:** ionizing spectrum by a 4-piece-wise continuous power law (test for stars, black holes,...)
- **Flexible nebular conditions:** ionizing strength, gas density, $[O/H]$, $[N/O]$, $[C/O]$...

Cue: fast, flexible neural net emulator for nebulae

Inferred 'invisible' ionizing spectrum + nebular physics

Observed data

Li, Leja+24ab

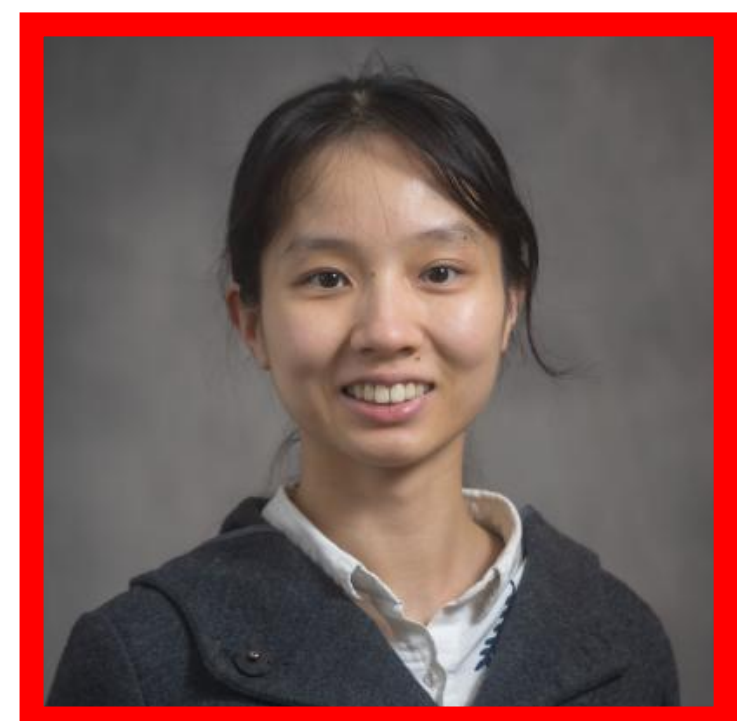
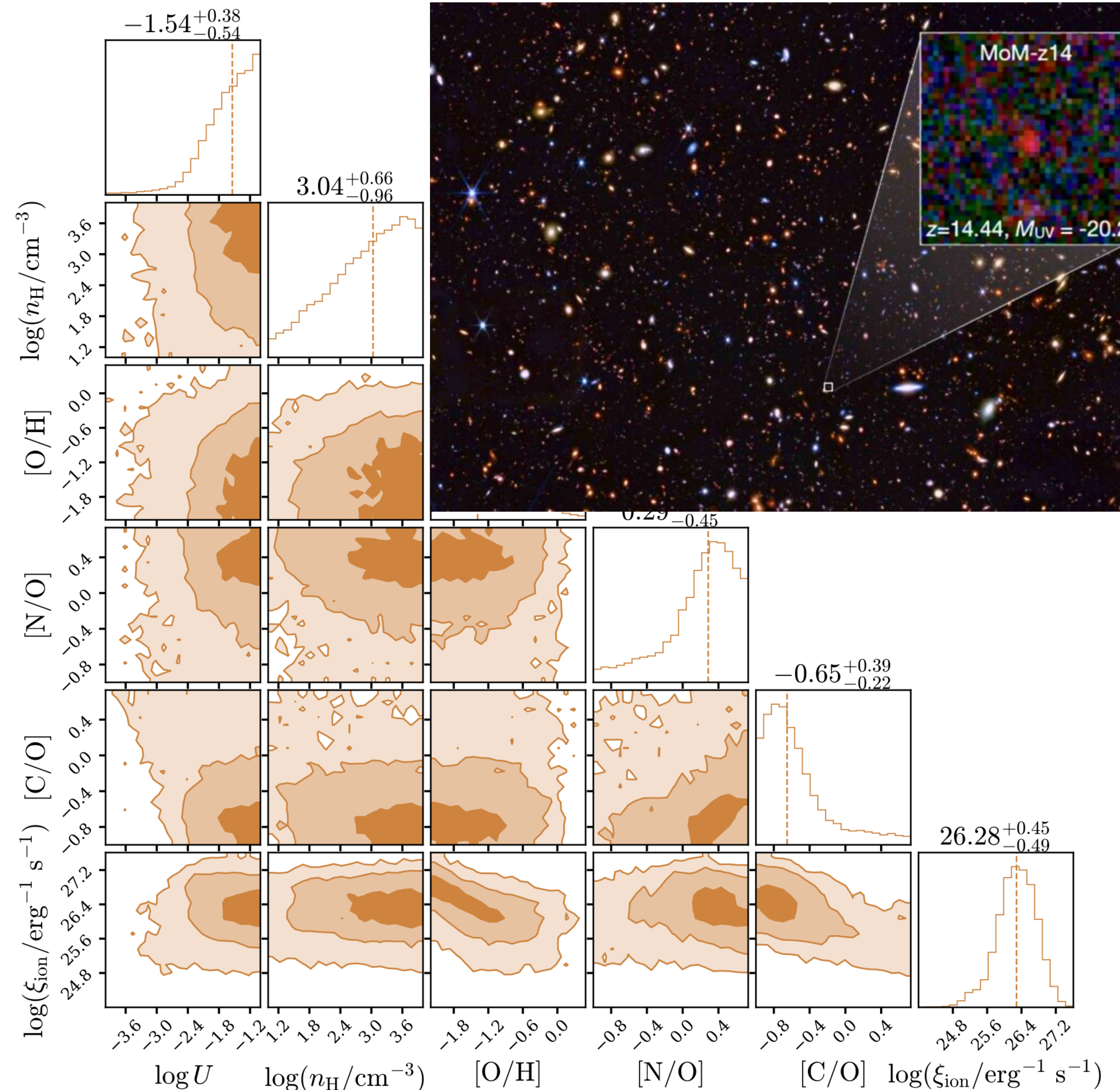


- Predict **brightest 130 emission lines**+**nebular continuum** with **< 5% uncertainty**.
- Each prediction takes 5ms → **10^4 faster** than full nebular models
- Can *investigate mysterious ionizing sources to calibrate models* (good data), or *marginalize over uncertainty in ionizing sources* (most data)

<https://github.com/yi-jia-li/cue>

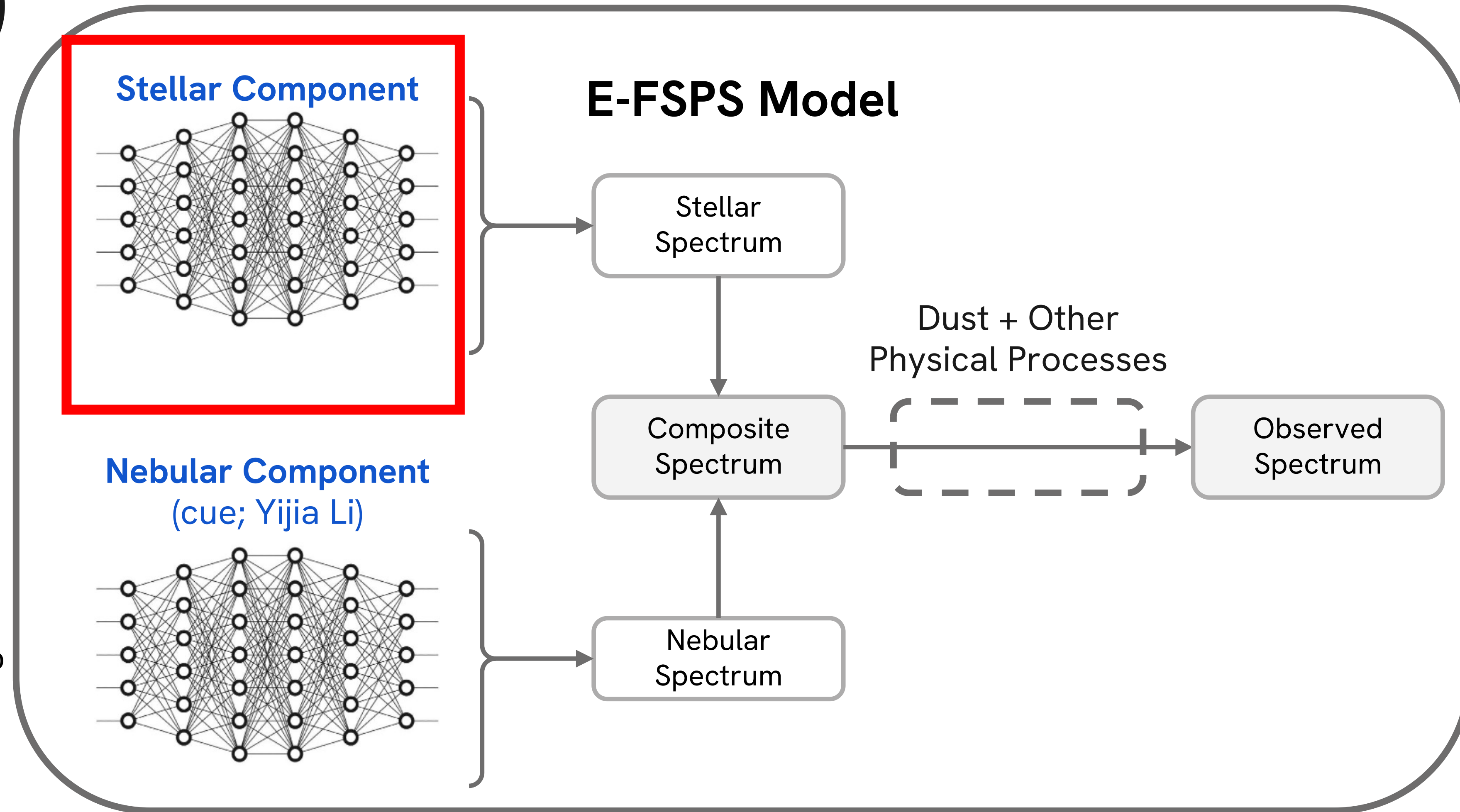
Example: used in discovery of most distant galaxy, a few months ago! 290 Myr after Big Bang. **Cue** reveals

- *Something* pumping out hot photons faster than normal stars (not normal!)
- Way too much nitrogen (exotic element production?)
- Surprisingly dense gas (early universe conditions)



Naidu et al. 2025,
incl Li, Leja

Piecewise Emulation For Spectra (Not Single-Shot!)



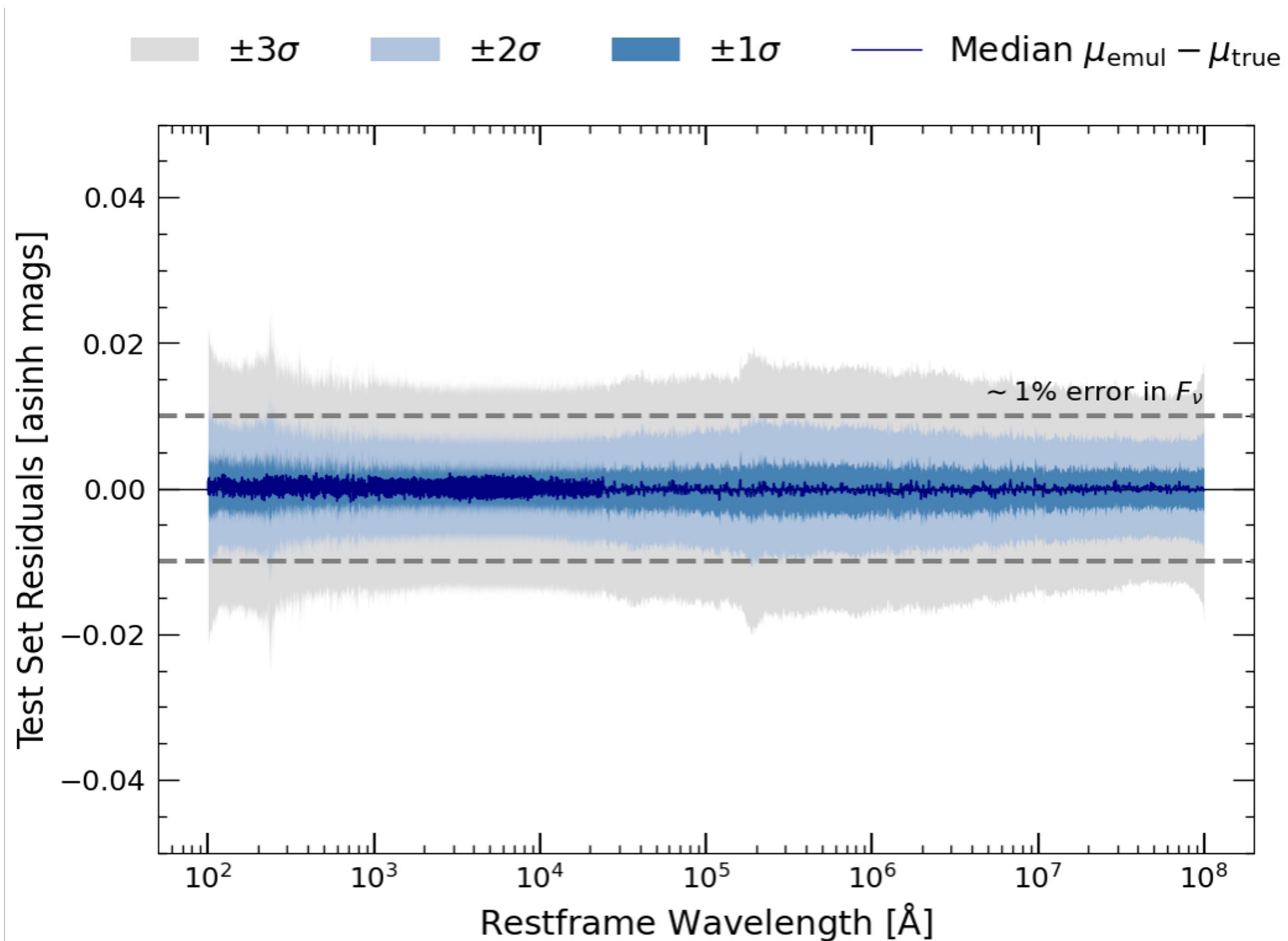
Burnham et al. in prep

Emulated FSPS (E-FSPS) replaces gigabytes of stellar model grids with **deep neural networks**.

Emulation Error for Stellar Spectrum



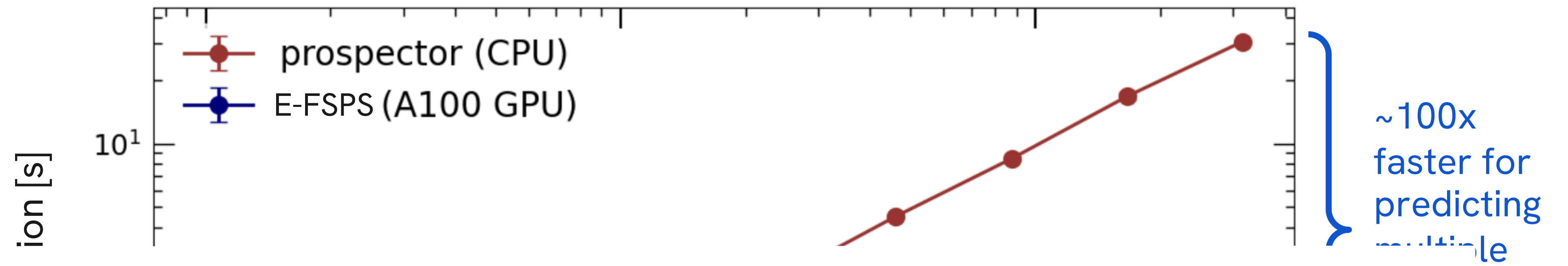
Burnham et al. in prep



How does E-FSPS compare to Prospector?

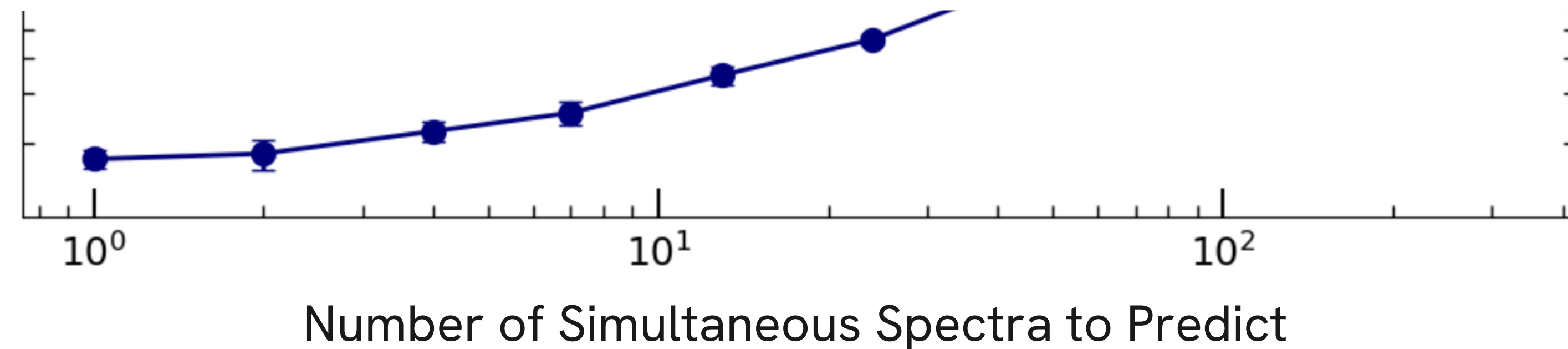


Burnham et al. in p



We can rapidly generate thousands of spectra **simultaneously**?
Let's do something new!

Initial overhead
for GPU usage



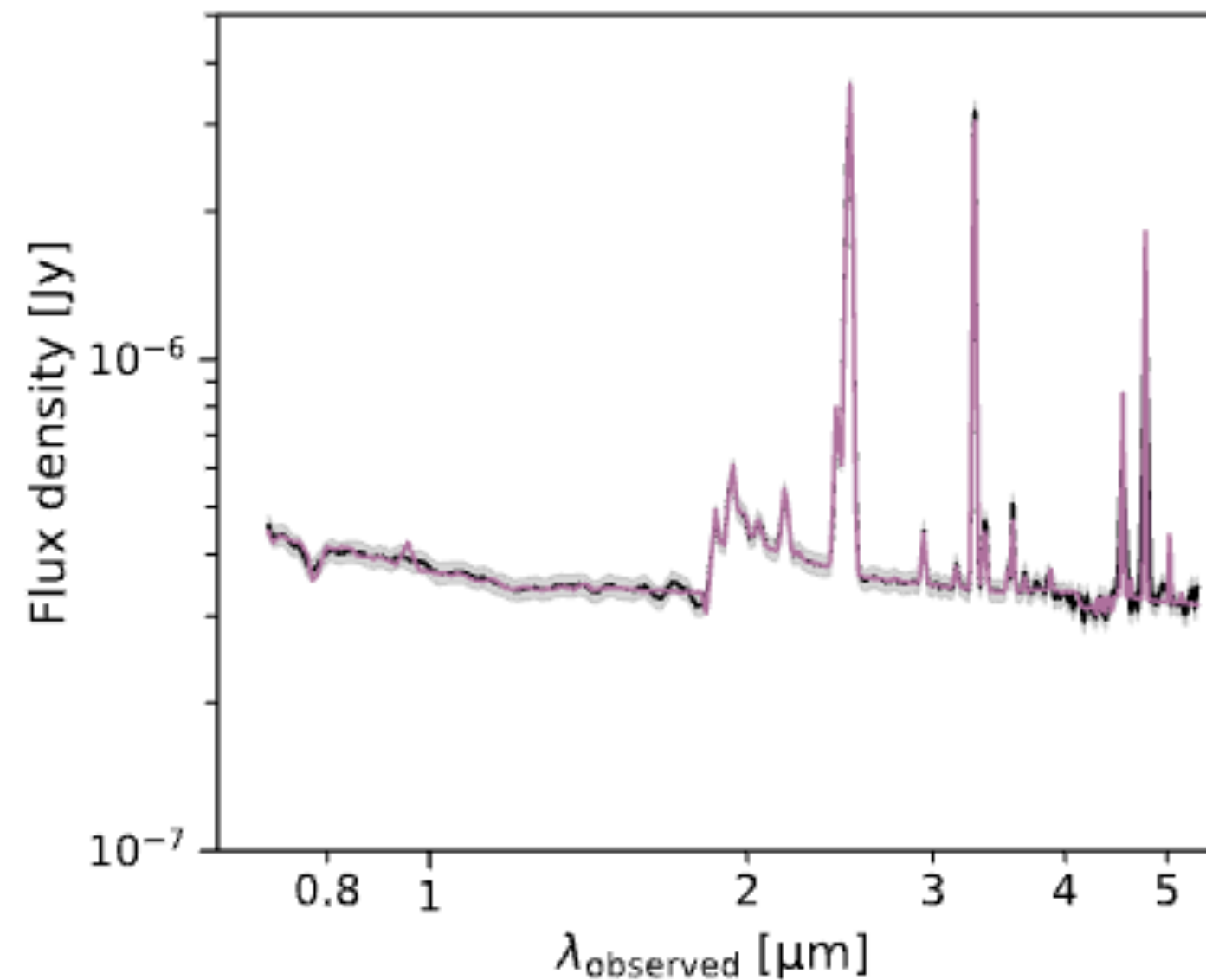
+Additional speedups from gradient-enhanced sampling (TBD)

Key challenge: highly multimodal likelihood space; how to cleverly initialize?

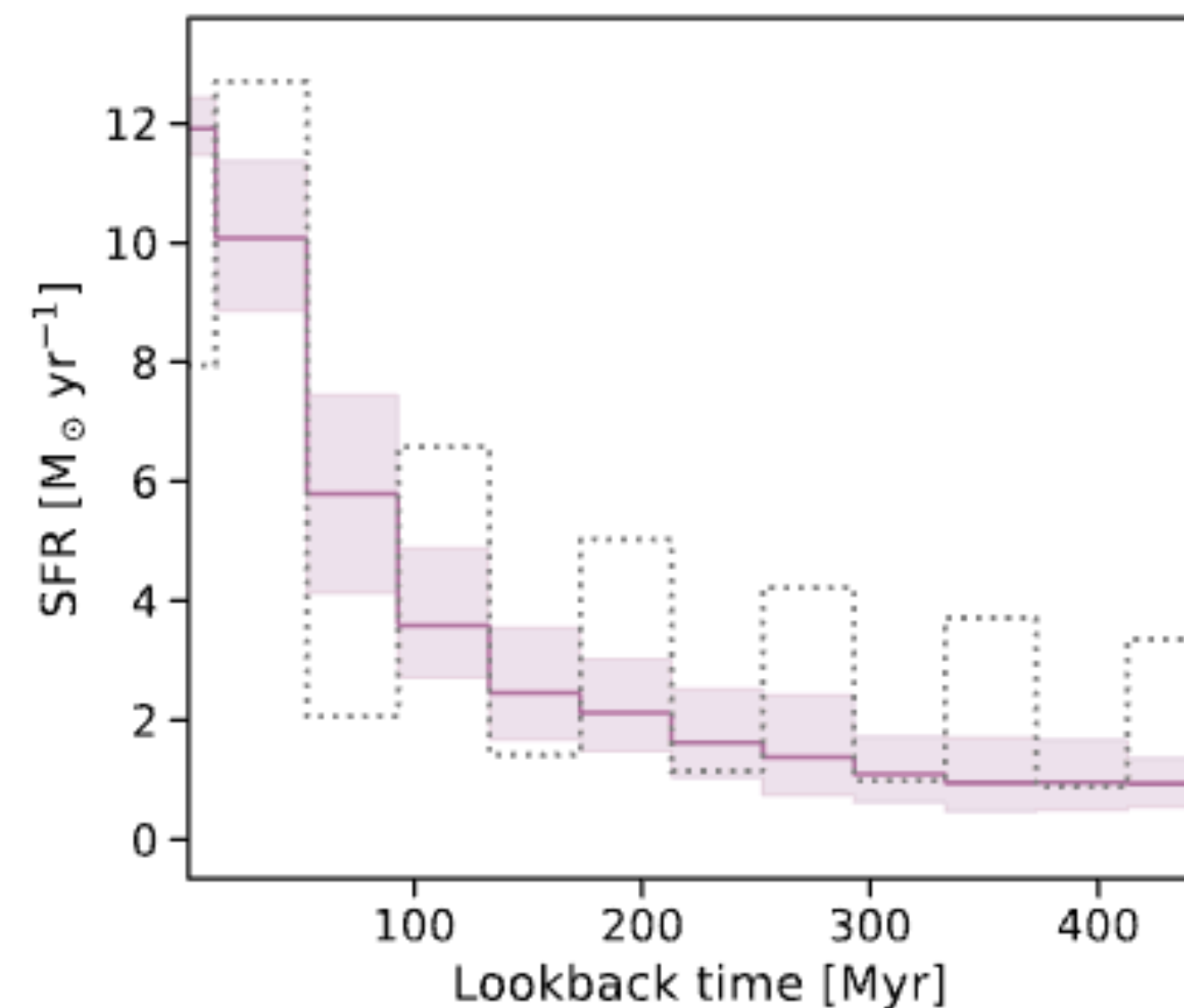
"Bursty" Star Formation Common But Nearly Impossible To Model In Single Galaxies

Recently-formed bright stars **obscure nearly all stellar history older than 100 Myr** in images of distant galaxies

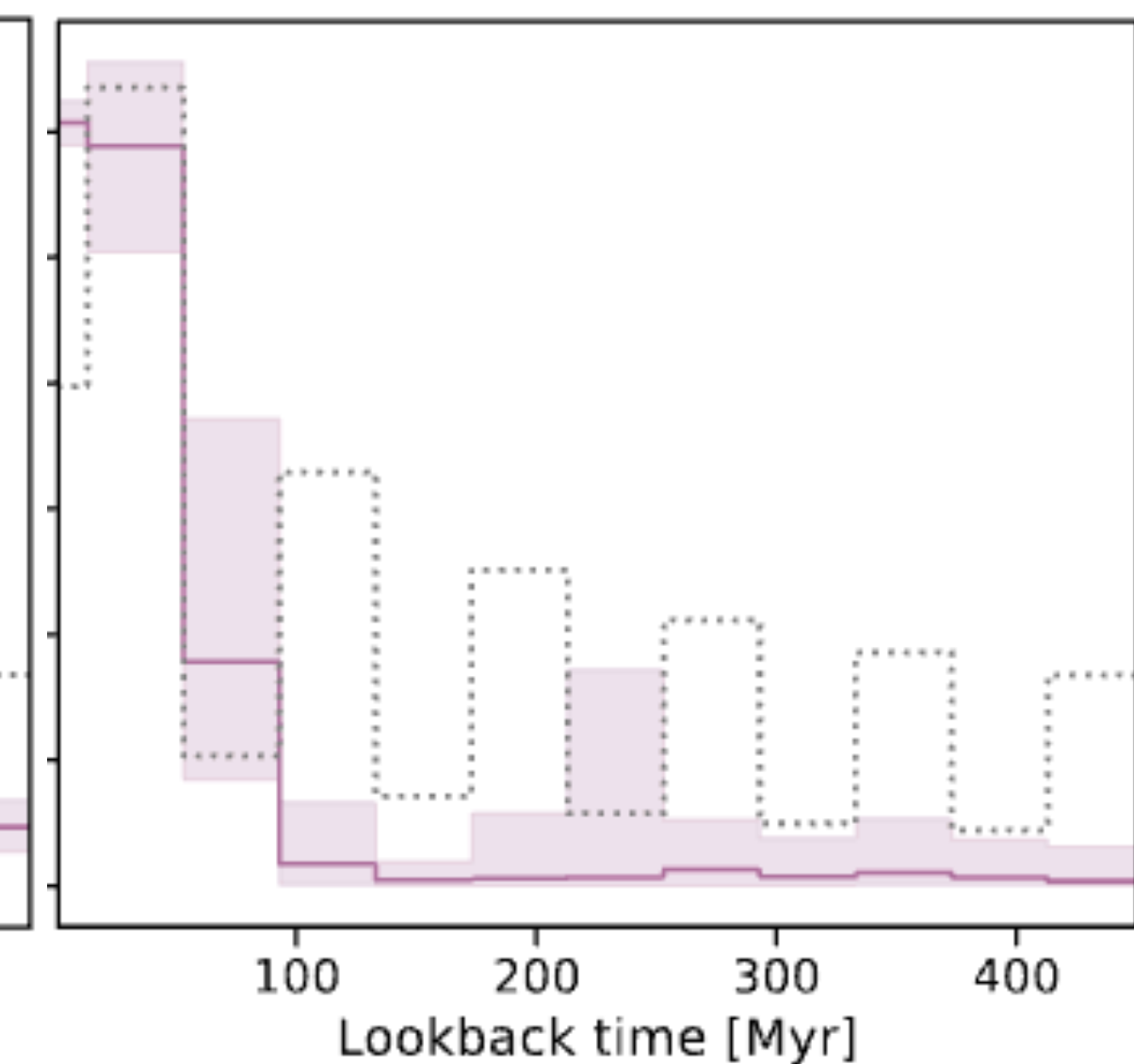
A model-generated, high S/N spectrum from the Webb telescope, **perfectly fitted**



A "smooth" prior recovers a *smoothly rising history but no bursts*



A "bursty" prior recovers *only bursts (no old stars)*



Wang, Leja et al.,
ApJ, 2025

see also Narayanan+23, Haskell+24, Wan+24

In addition to fluctuations being unrecoverable, bursty SFHs **significantly degrade accuracy** of masses, ages, star formation rates (>factor of two) - *even if we know it's bursty.*

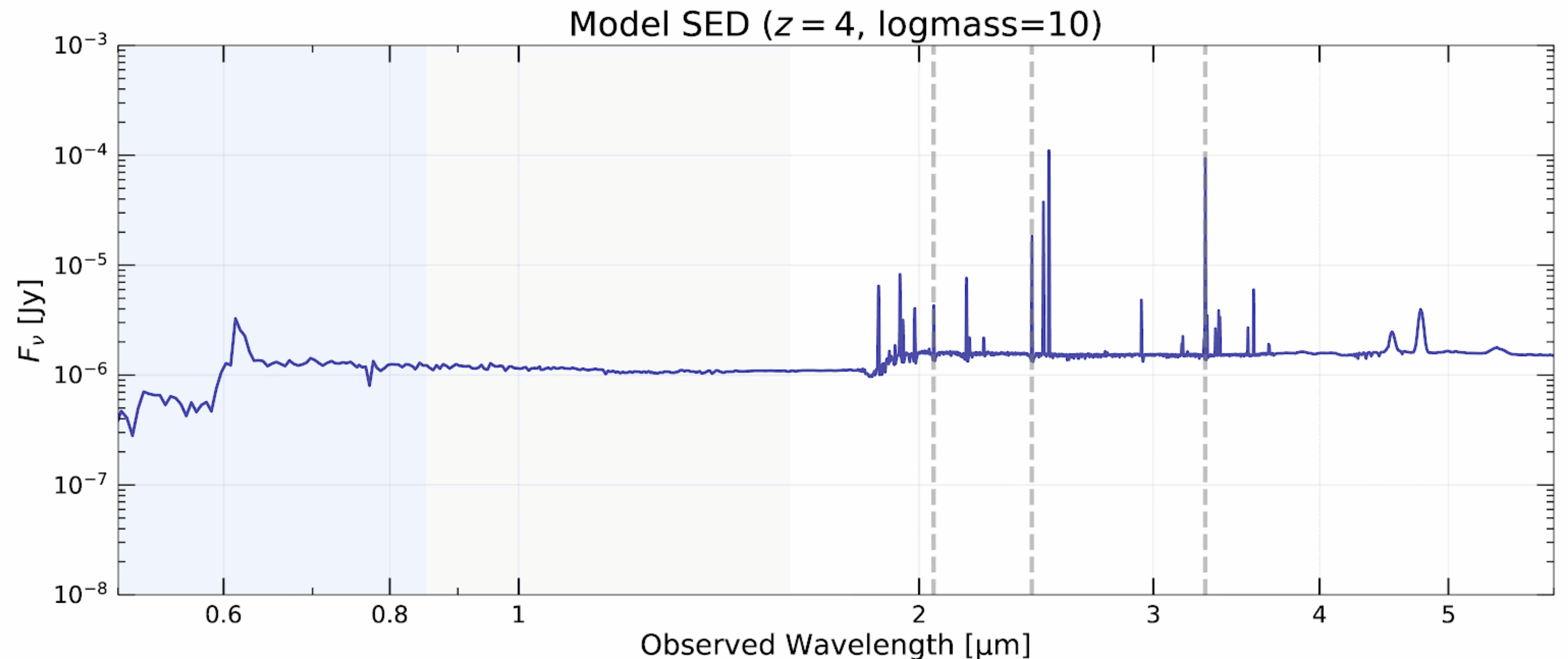
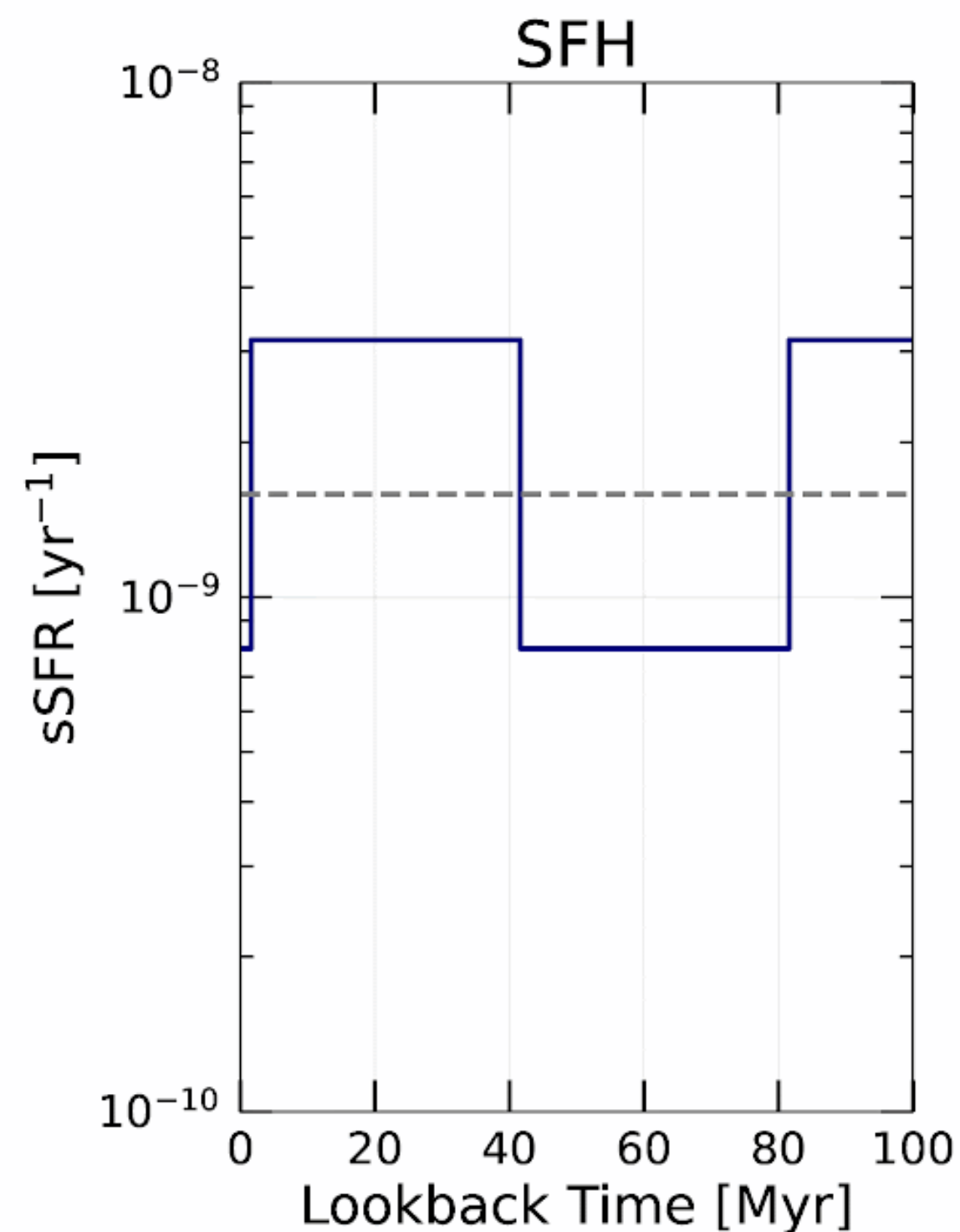
Single objects degenerate?

**Solution: model entire galaxy populations
simultaneously**

(Also e.g. **pop-cosmos** for galaxy population modeling to constrain cosmology of our universe; Alsing, Peiris, Leistedt, Leja, Mortlock...)

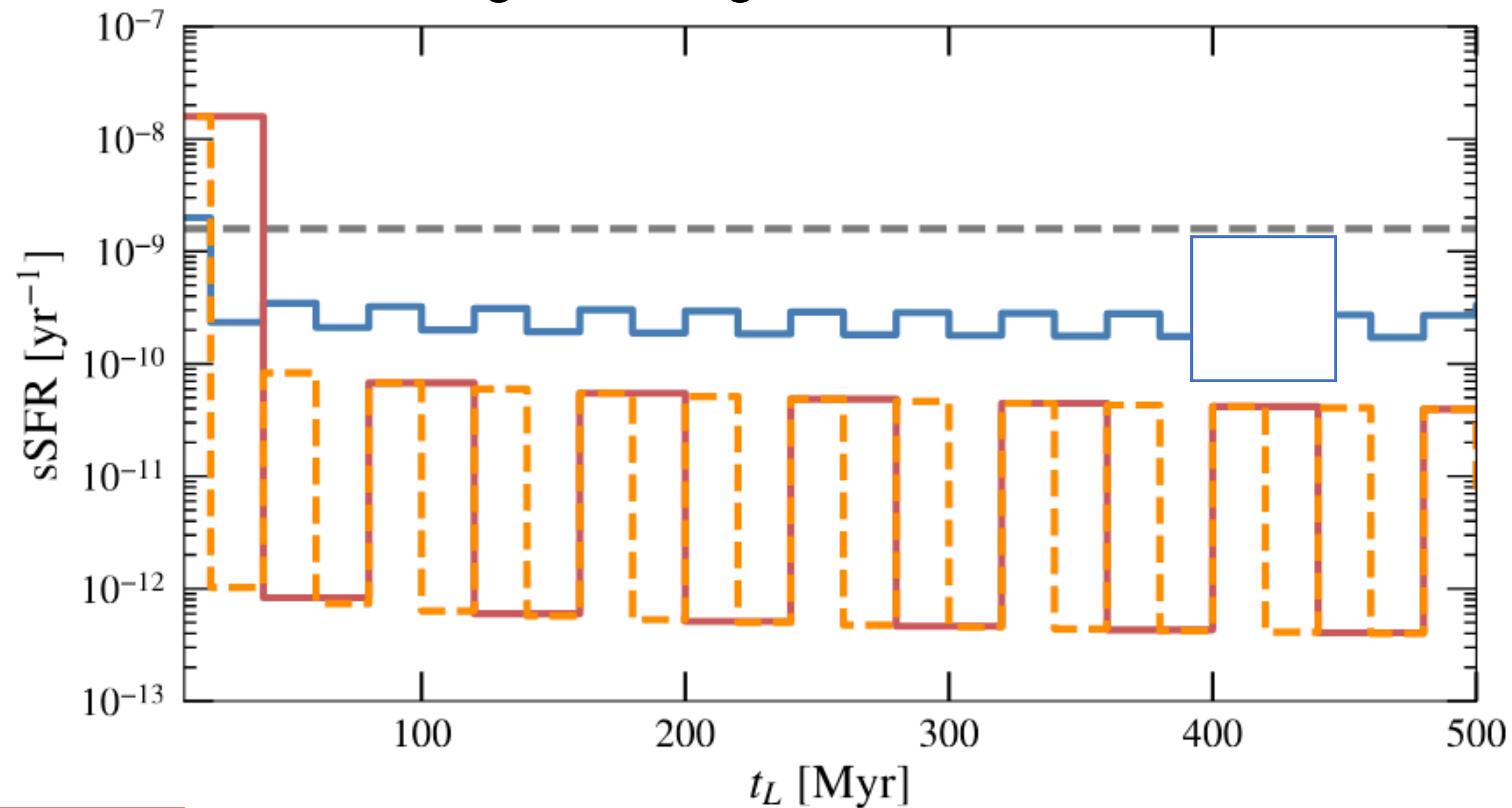


Burnham, Wang, Leja et
al. in prep



Impossible to Measure Fluctuation Timescales, Frequencies in One Object - But Solvable For Populations!

different fluctuation timescales, amplitudes, and
“rising” or “falling” formation histories



Burnham, Wang, Leja et
al. in prep

Simple amplitude, timescale, slope **not sufficient to describe complex feedback physics in galaxies.**

Use full, complete model: ‘stochastic’ formation histories, described by **power at different frequencies.**
(Caplar & Tacchella 2019; Iyer+20; Iyer, Speagle+24;...)

Burnham,
Wang, Leja et
al. in prep



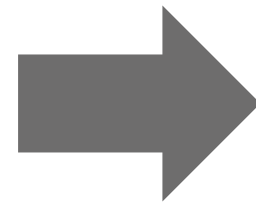
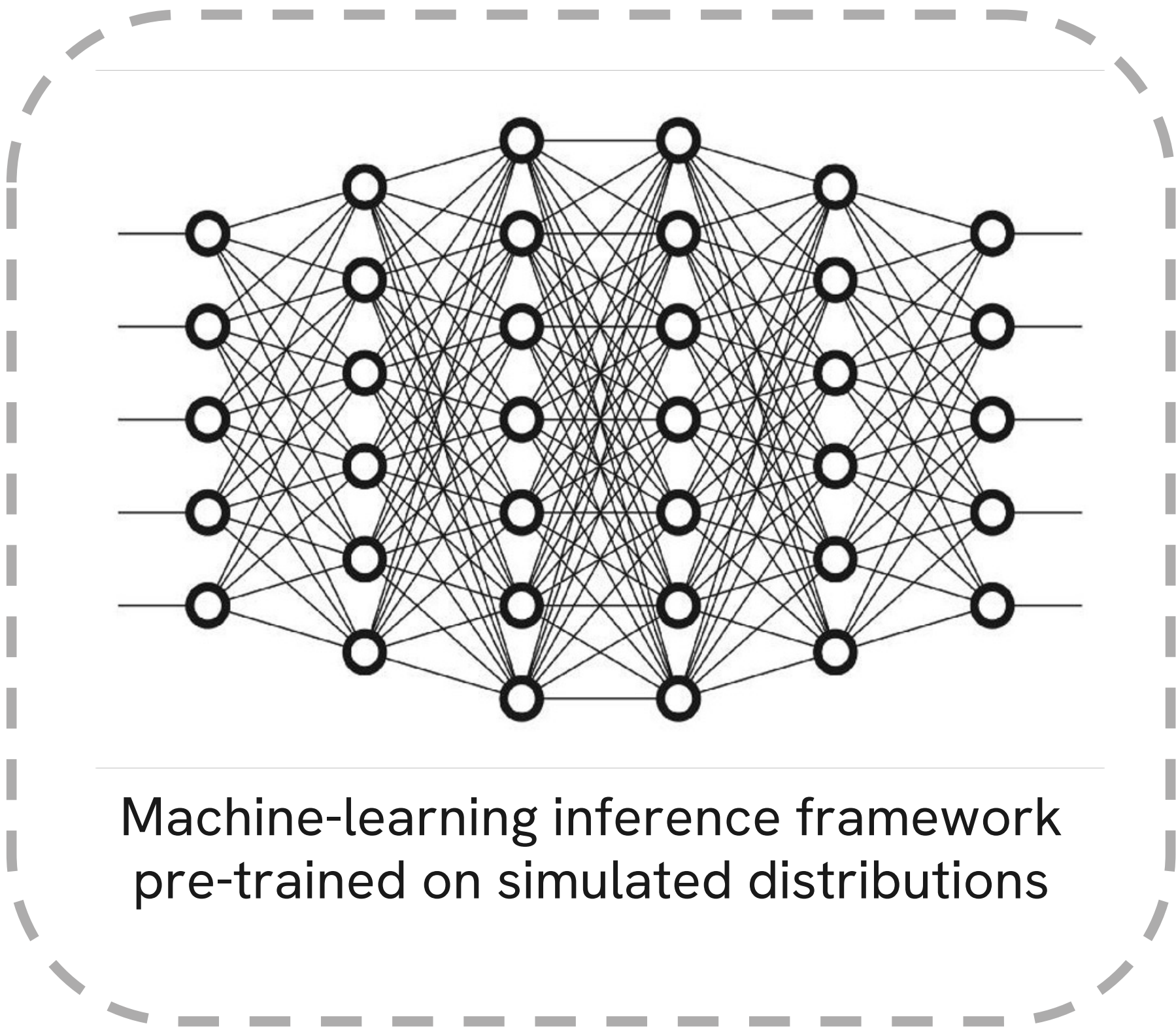
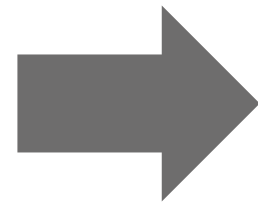
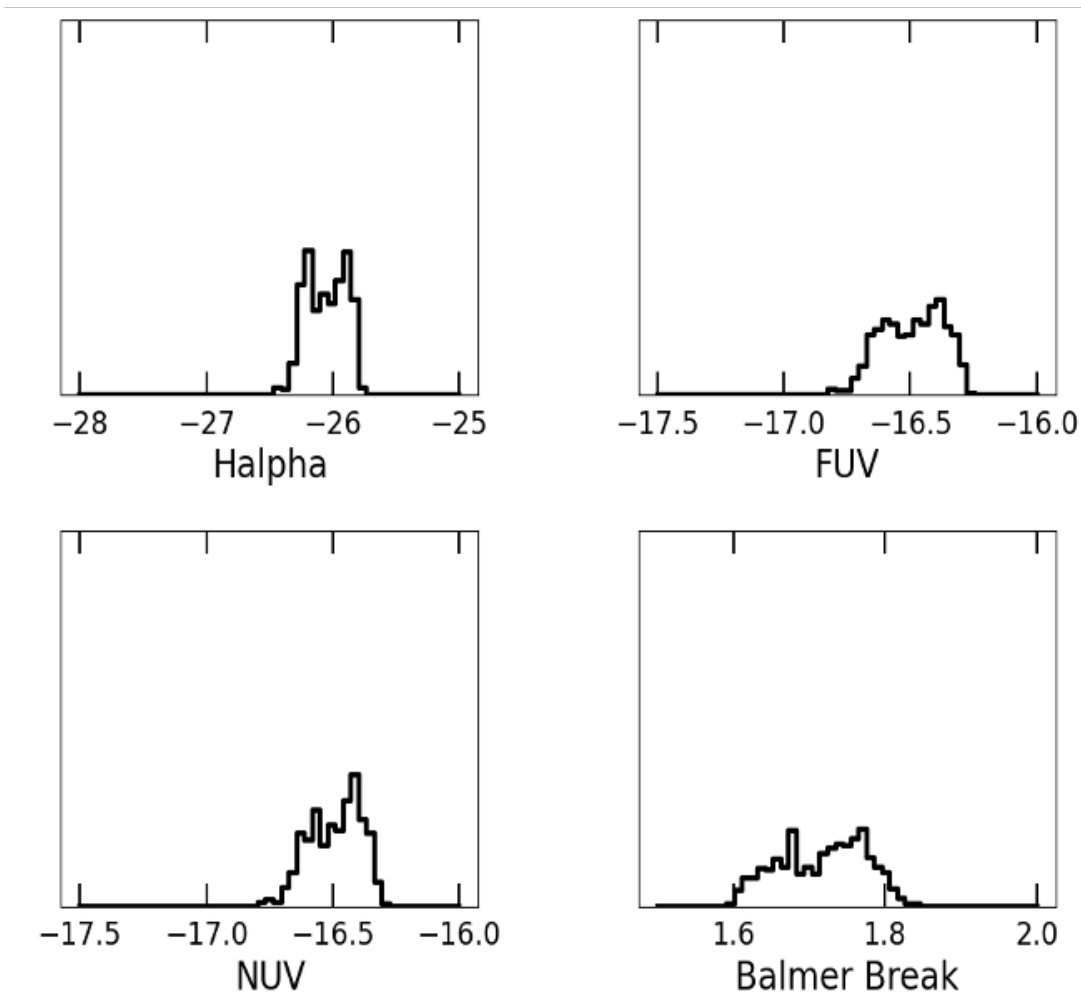
But - how to forward-model populations; need hundreds of spectra per likelihood call, millions of likelihood calls —> **10-20k core-hours per fit ?**

Use **simulation-based inference (SBI)**!

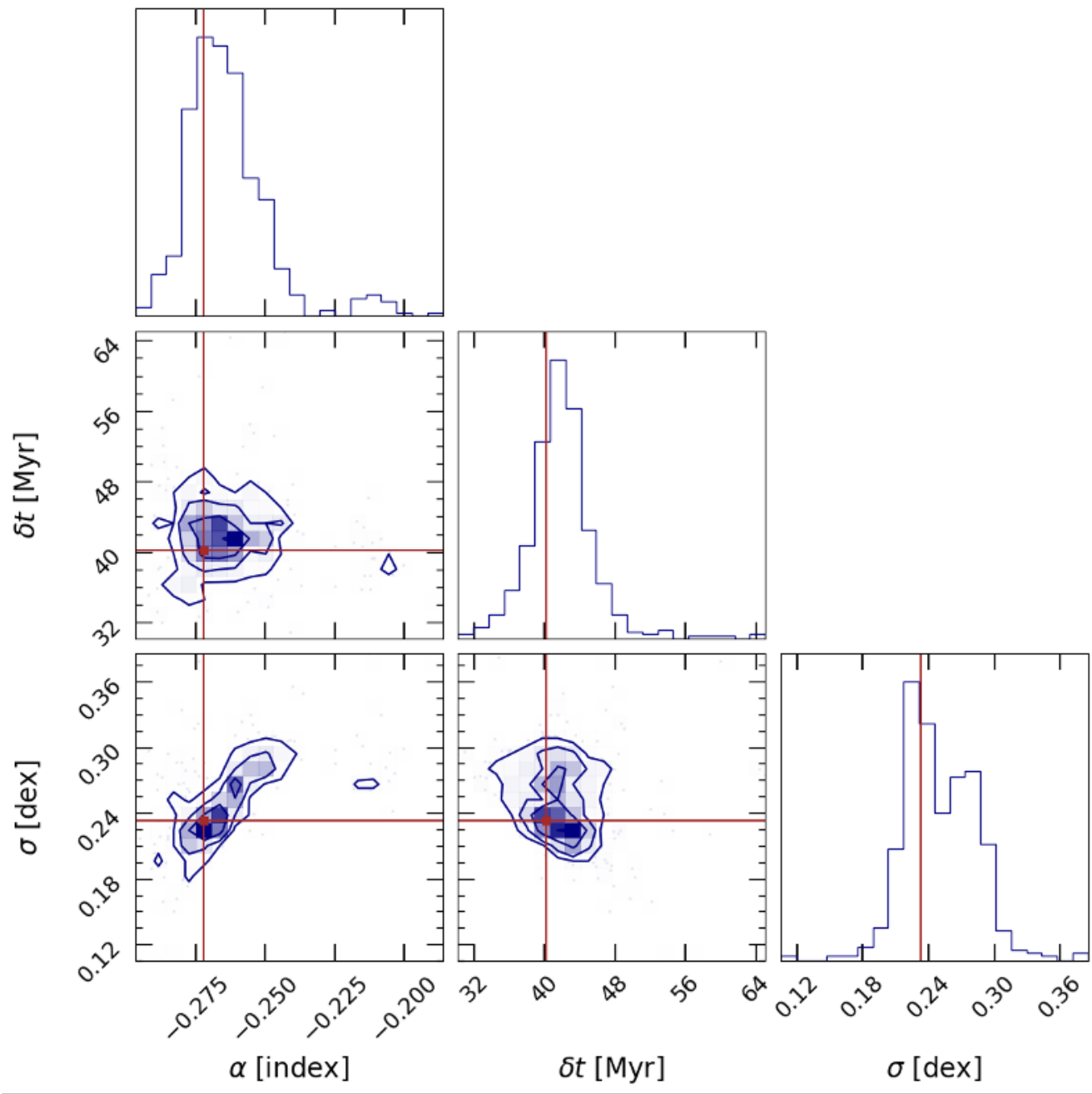
Marginalize over nuisance parameters (e.g. distribution of stellar abundance patterns, dust...); include measurement and Poisson (counting) noise for small-N surveys of a population; train on cluster, run in seconds on laptop.

Simulation-Based Inference

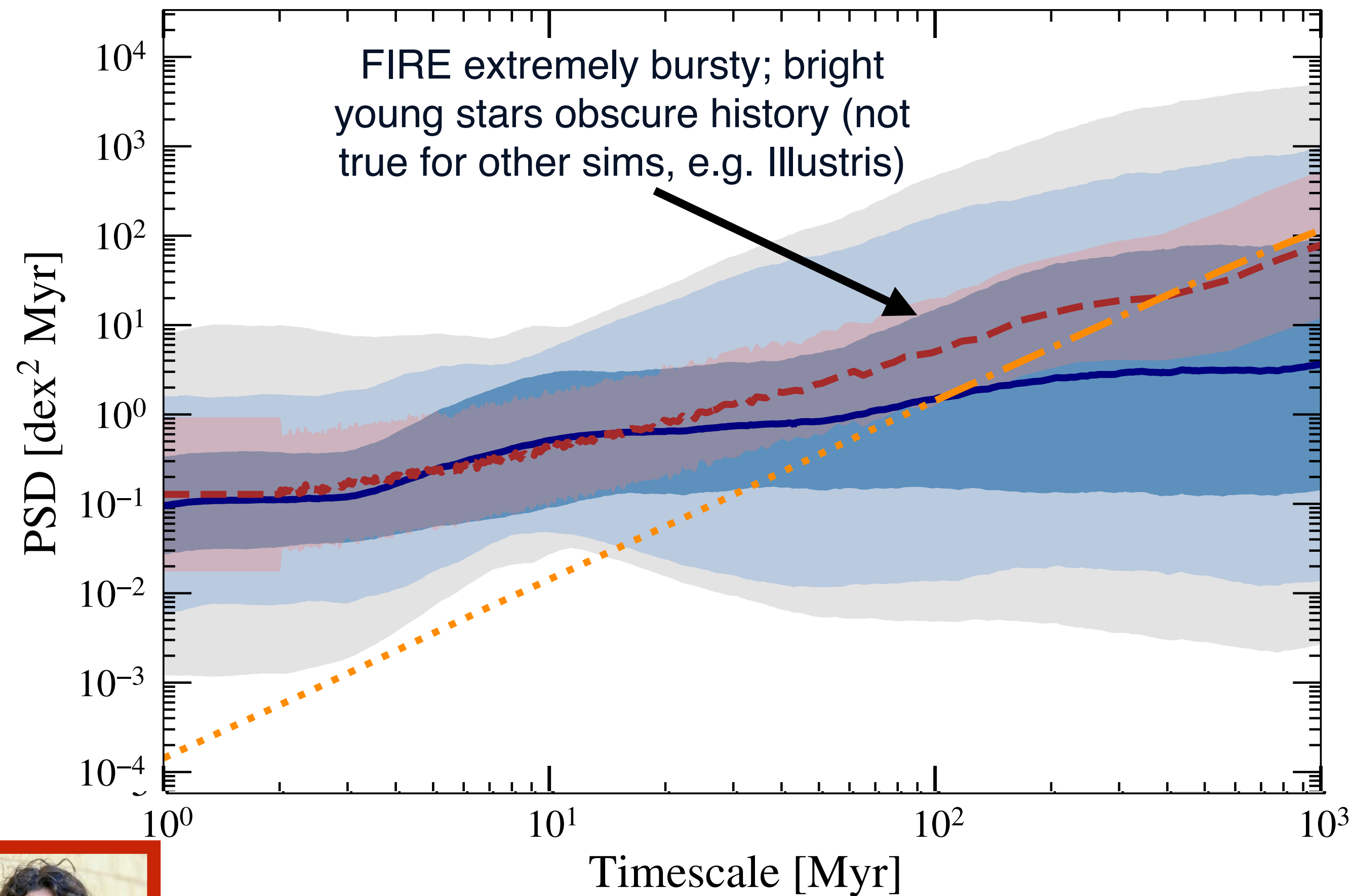
Distributions of (Noisy) Observables



Population-level Parameters



Burnham, Wang, Leja et al. in prep



can **accurately recover** short-term fluctuations for realistic populations ($N \sim 500$ galaxies with observational noise equivalent to 5 hours on James Webb).

Short-term fluctuations **very different in galaxy formation sims** (FIRE, Illustris, FLAMINGO, ...); Iyer+20. Can finally **falsify feedback models!**

Can **avoid outshining** by **learning the true prior**, and **applying to individual objects**. **Unsolvable** with classic techniques -- now **feasible** with SBI-powered approach.

Next up

- test epistemic uncertainties (e.g. stellar libraries)
- apply to galaxies in deep fields



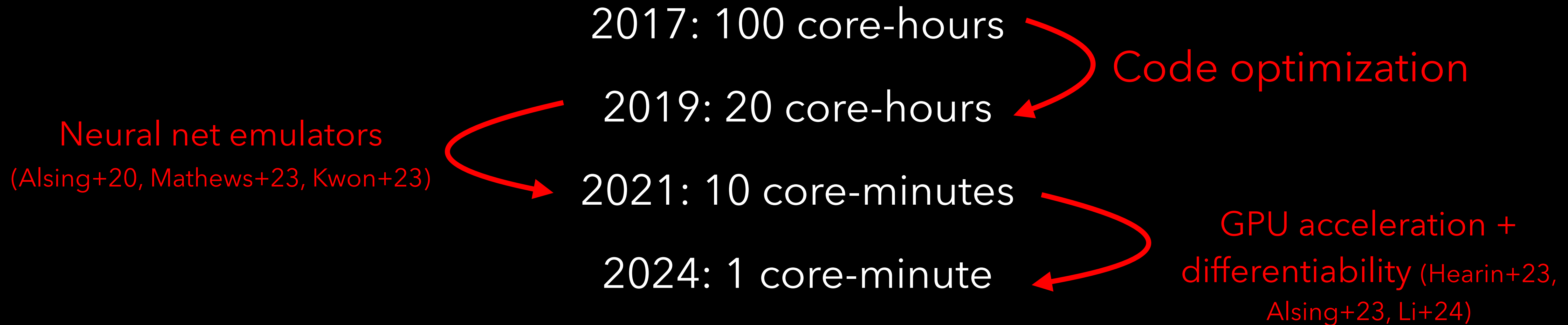
Burnham, Wang, Leja et al. in prep

- “Burstiness” population model used a few well-chosen features from spectra.
- Lots more to learn; lots more pixels to use; far more data coming. **Can we adapt SBI to fit entire spectra?**
- Challenges:
 - Our physical priors are fuzzy; update while running?
 - Rife with model misspecification (astrophysics is hard)
 - Balance coverage and highly informative priors (FreB?)
 - Selection effects very strong (only see bright things)



A Quiet Computational Revolution

Time to fit galaxy imaging using *on-the-fly* model generation:



& ***simulation-based inference*** yields <10 core-seconds per object

10^5 speedup in 5-6 years - hold on to your hats, folks! Now can efficiently model complex systems with hundreds, thousands, or *tens of thousands* of parameters.

... just need to **parameterize the physics!**

(Already **hundreds of parameters** in existing models of black holes, chemical evolution, stellar evolution, photoionization...)

New Quests in Galaxy Formation Enabled By Rapid High-Dimensional Sampling

- Neural net emulators yield speed increases of **100-1000x for galaxy data**, + gradient-enhanced sampling and GPU-acceleration
- Solves 'curse of dimensionality'; permits more physics and/or faster inference (**e.g. deep Webb fields; Wang+24**), AND qualitatively new science questions
- **NEW**: rapid interrogation of unknown or mixed ionizing sources (**cue, Li et al. 2024ab**)
- **NEW**: Bayesian population modeling with SBI; infer short-term timescales, mitigate old problems of outshining and unknown formation histories (**E-FSPS, Burnham et al. in prep**)
- **FUTURE**: Can we meaningfully ingest hundreds of millions of galaxy spectra, each with thousands of pixels mapping to tens or hundreds of parameters, using SBI?