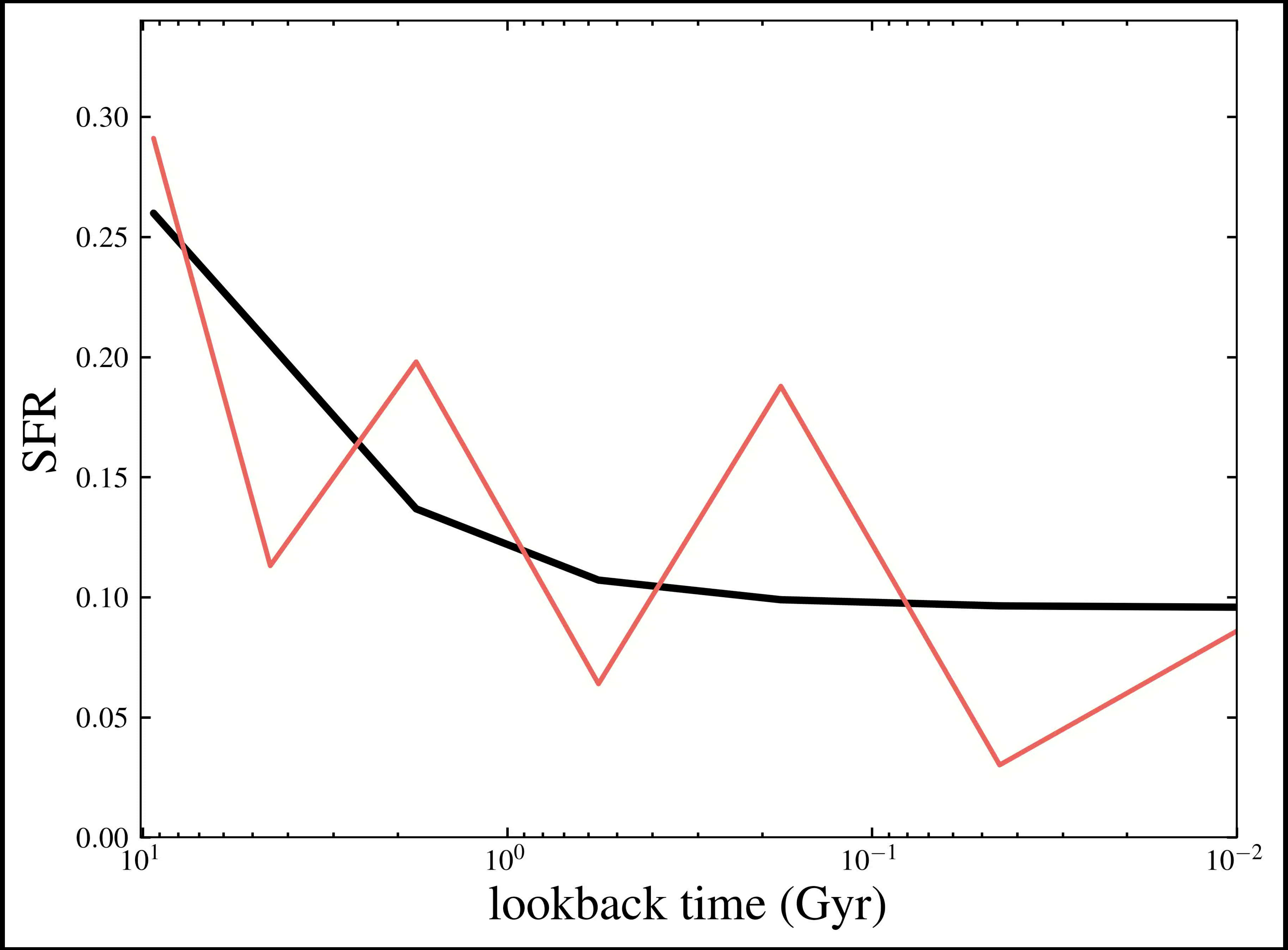
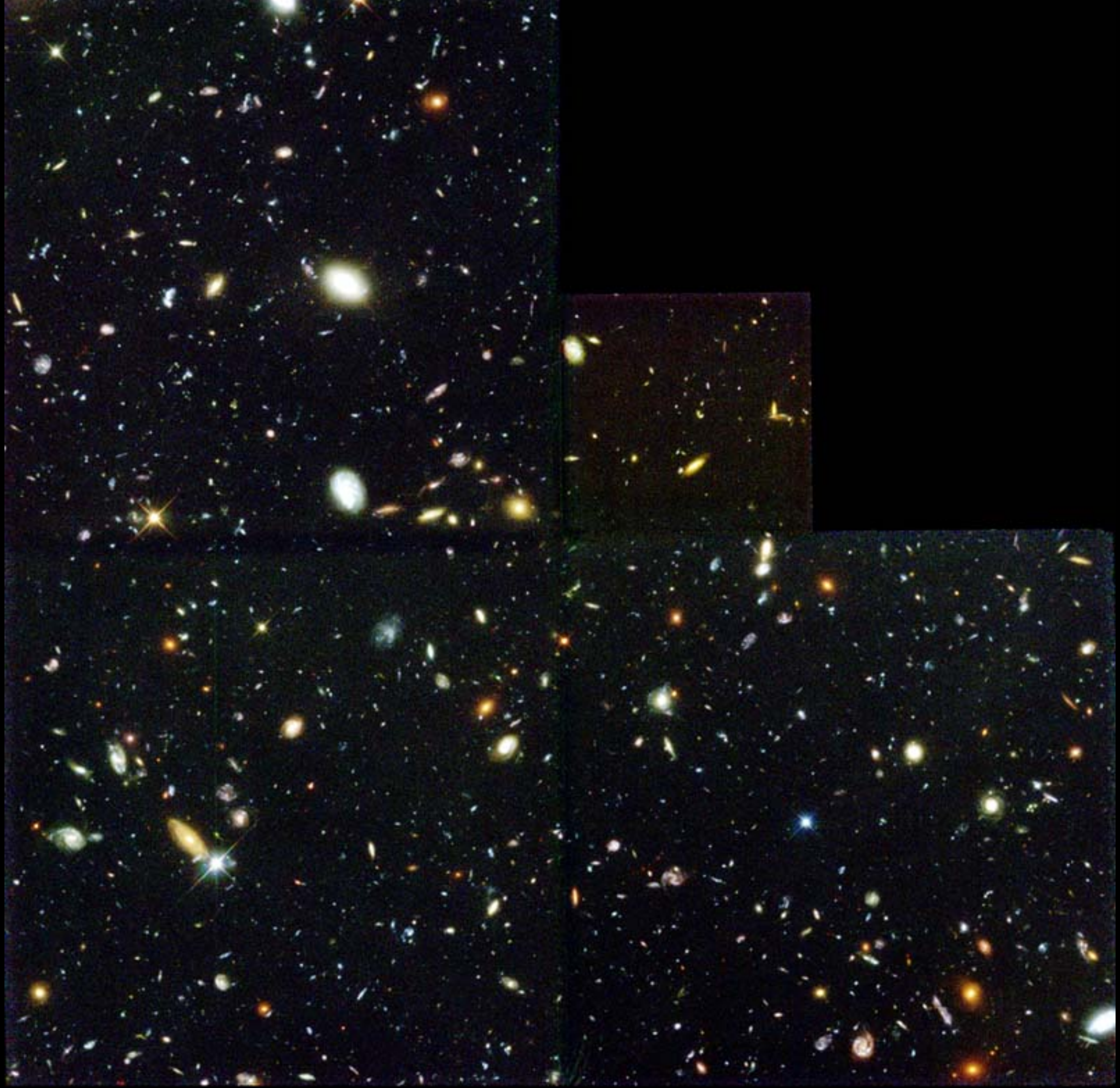
Rapid inference of galaxy properties in the age of **deep** and **large-scale** surveys of the universe

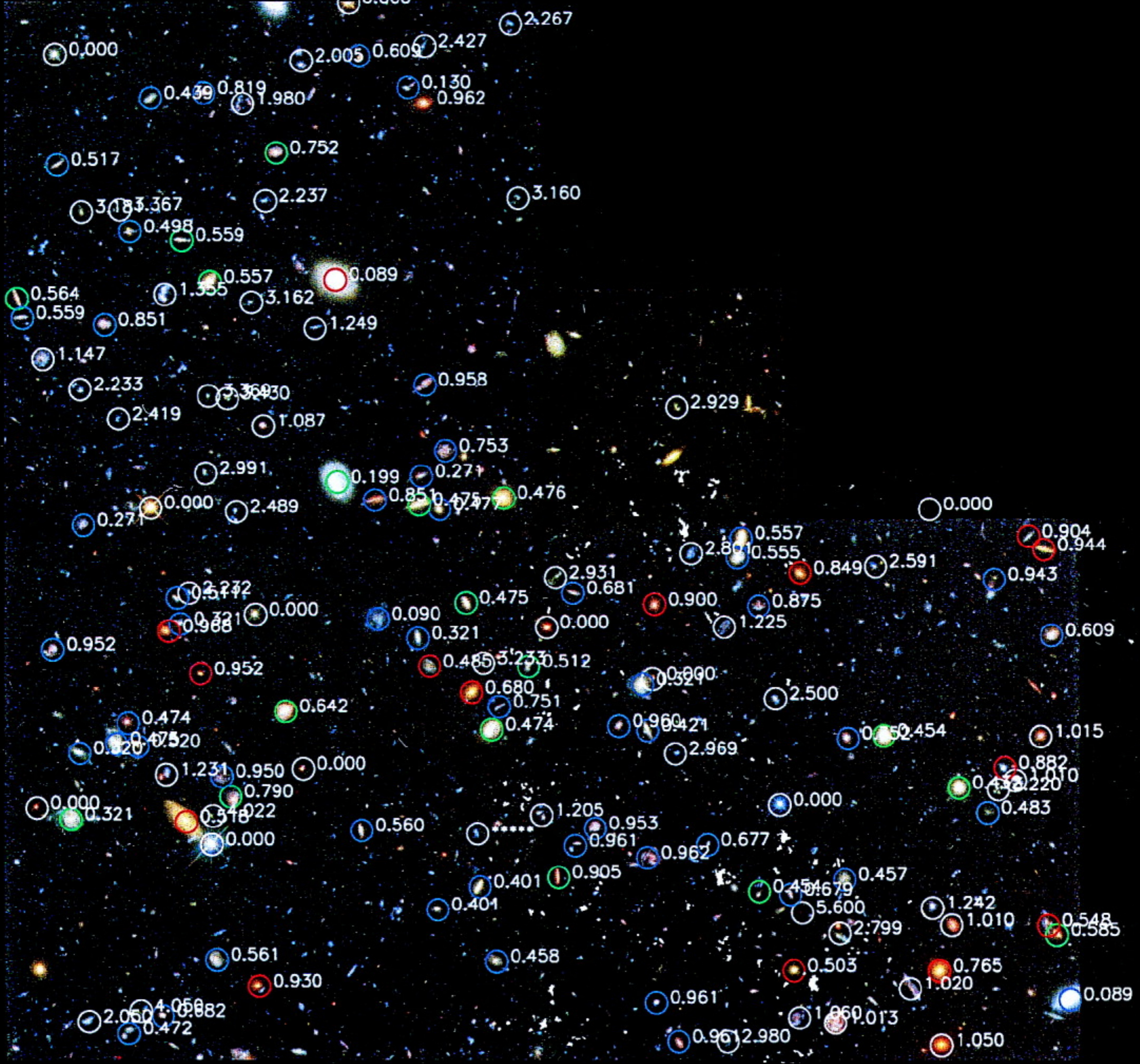


Joel Leja
Penn State Astronomy & Astrophysics
Dr. Keiko Miwa Ross Early Career Endowed Chair



Hubble Deep Field **HST WFPC2**

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

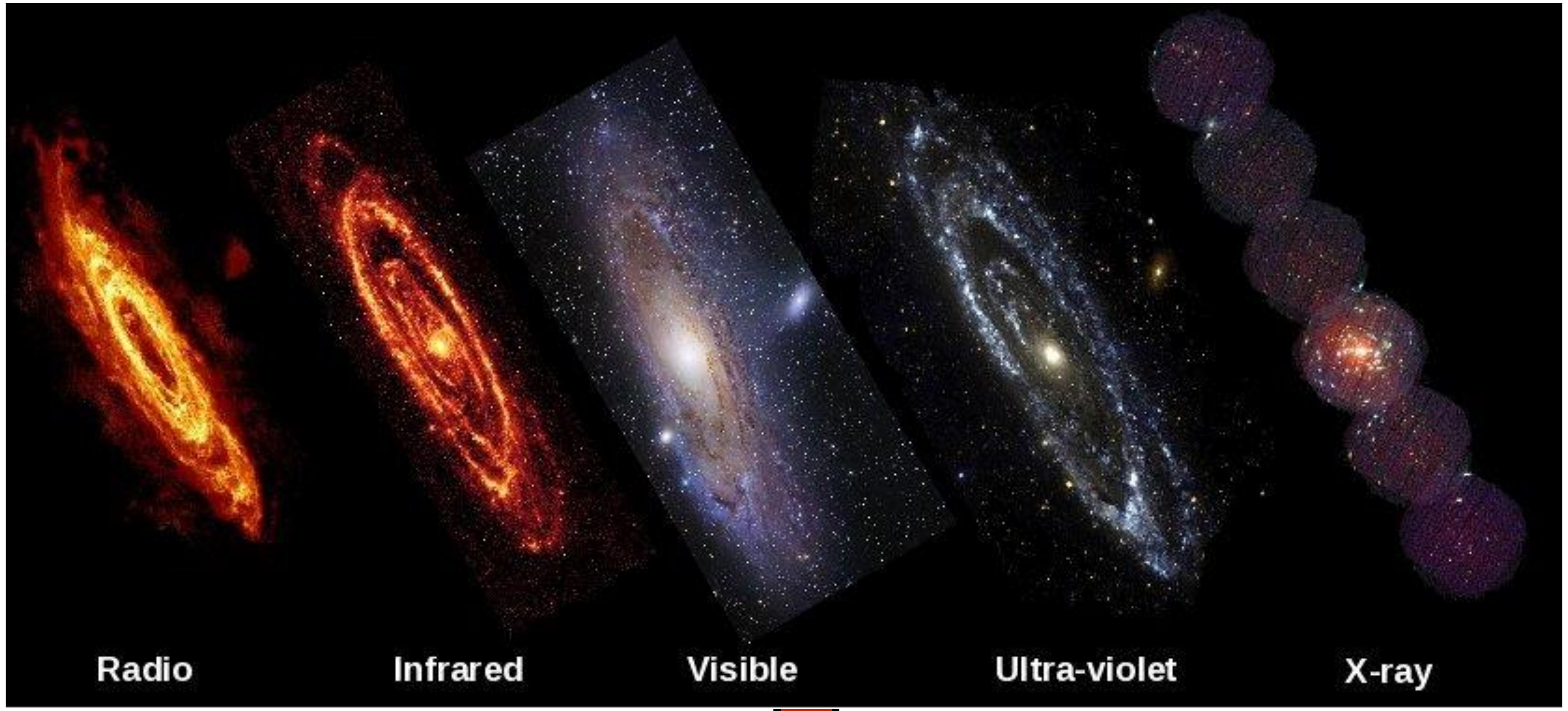


In addition to distance, **galaxies are cosmic ecosystems** that contain rich information about their growth:

- Formation histories
- Number of stars
- Production of heavy elements
- Dust obscuration

but also 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

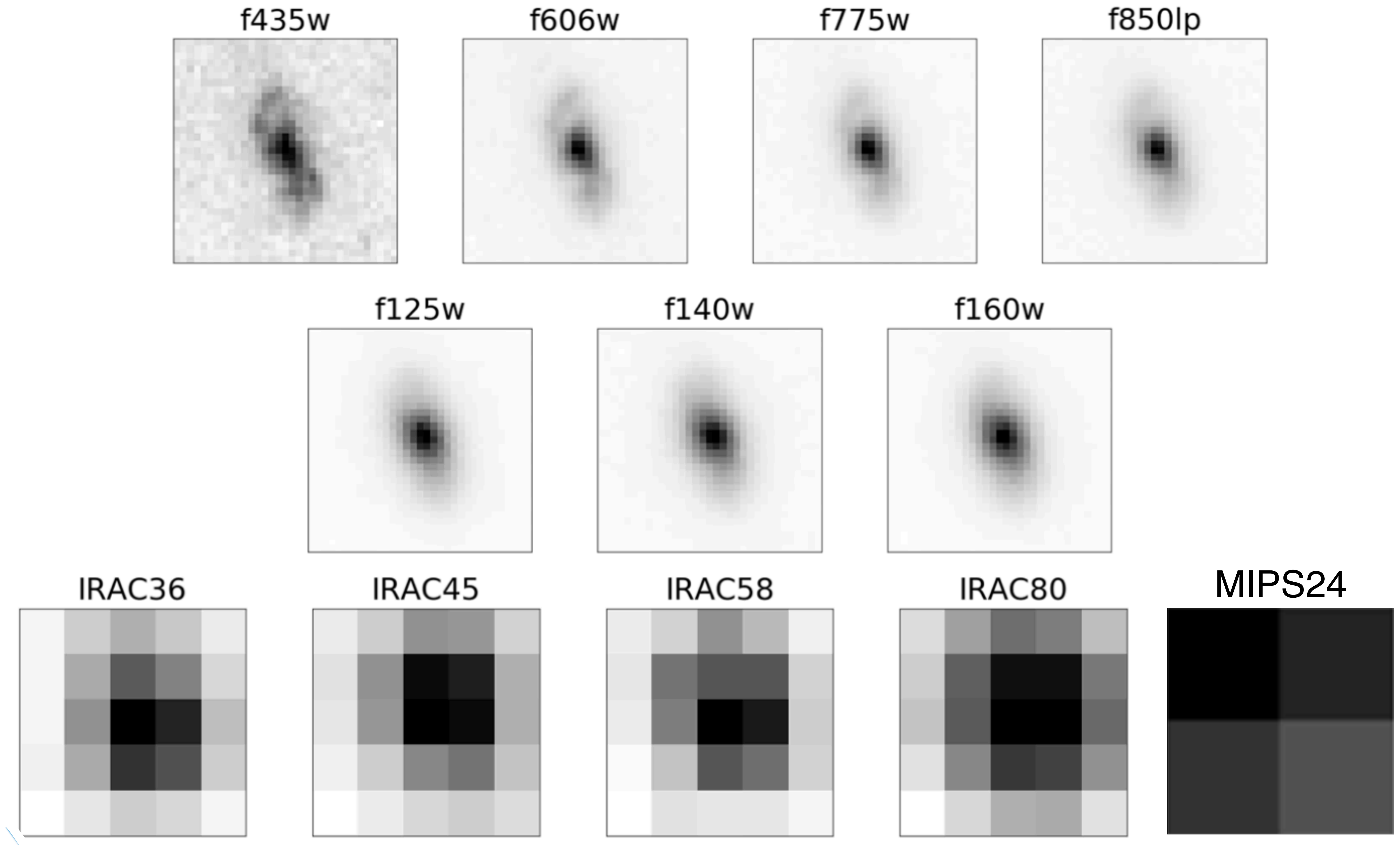
… and use models to turn them into *even more beautiful* inferred parameters.

stellar mass	dust content
star formation history	chemical abundances
nebular properties	active black holes

Key Idea: Stellar Populations in Distant Galaxies are (almost) always **Unresolved** (i.e. stars are blended)

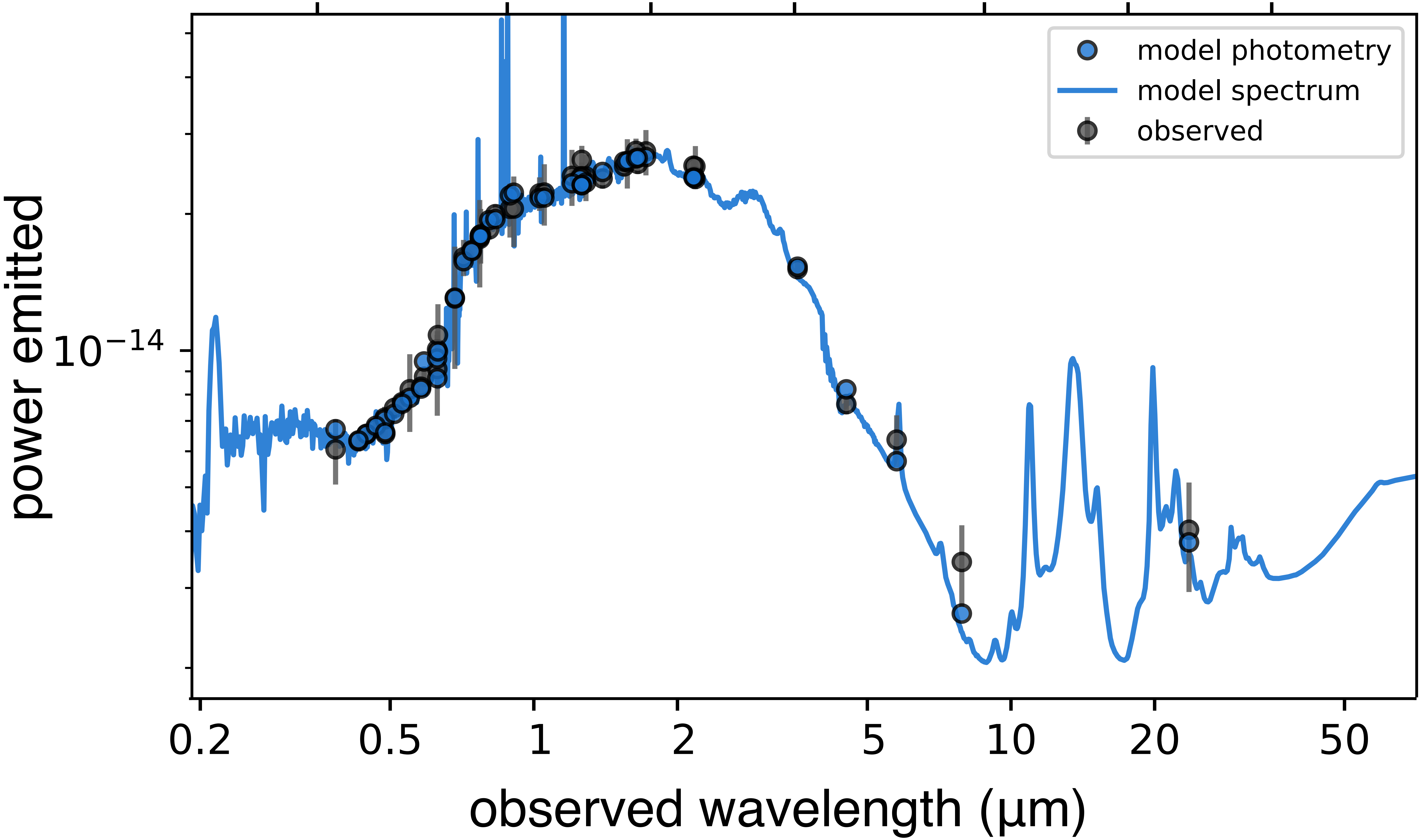
Andromeda

A typical distant galaxy

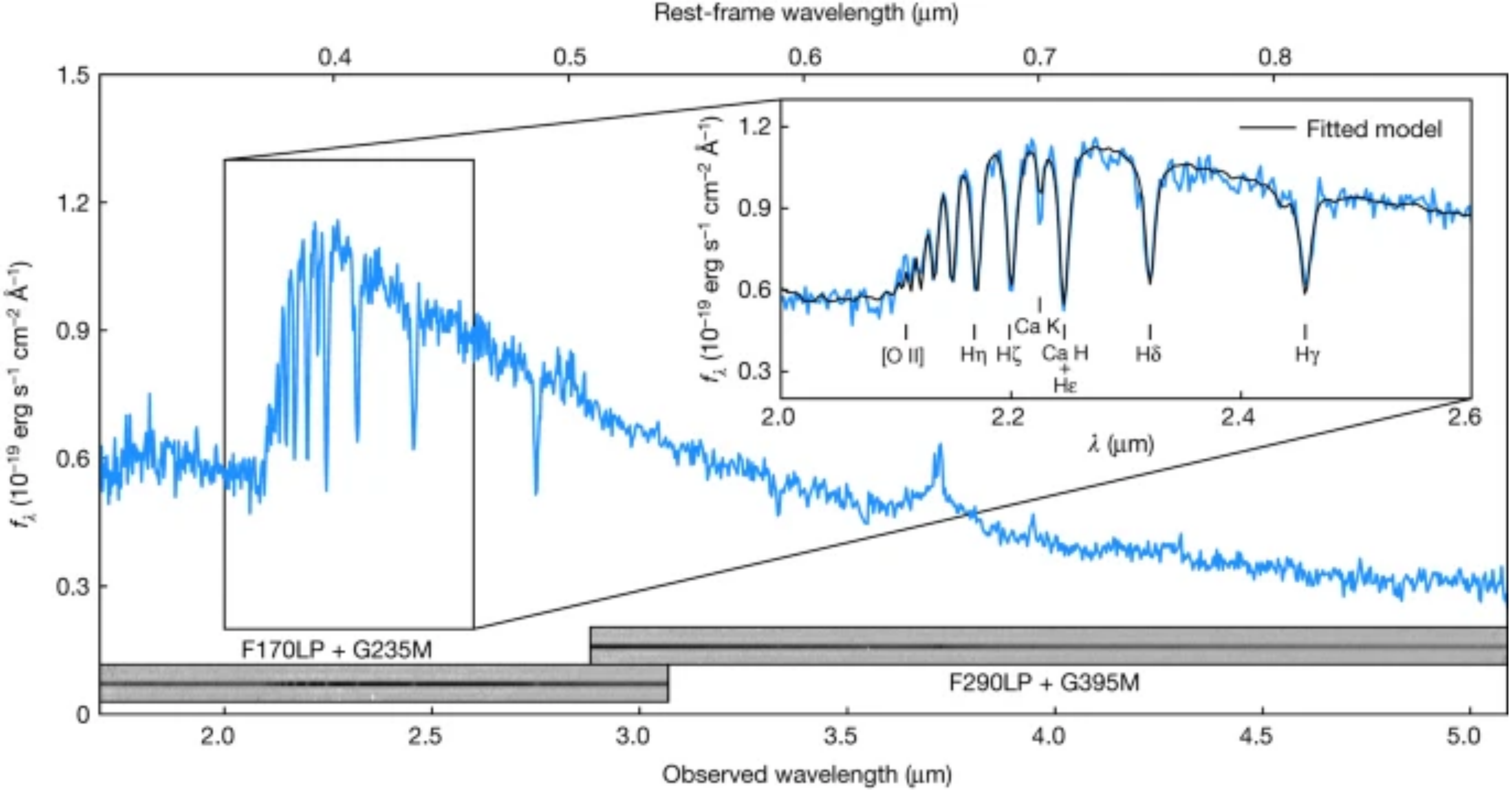


In most cases we model the **sum of the light** — hard to reconstruct dim populations (e.g. low-mass stars).

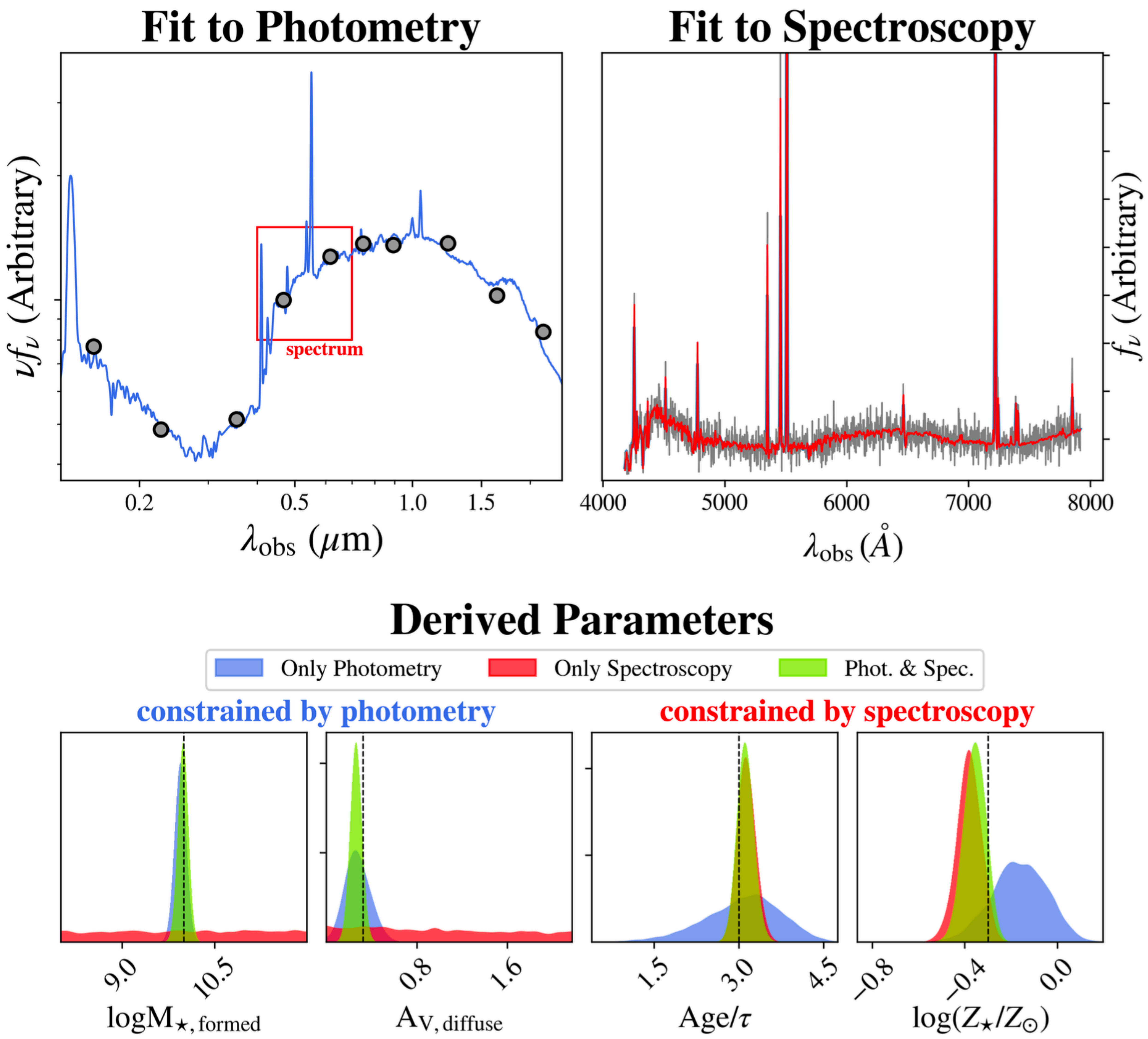
Thanks to decades of investment, most modern galaxy evolution surveys are multiwavelength, with tens of image types across the electromagnetic spectrum - usually spanning ultraviolet to infrared.



More rarely, spectroscopy is available. Spectroscopy is richly informative with many different absorption and emission features, and promises constraints on galaxy formation histories and physical conditions to **factors of 2-3**.



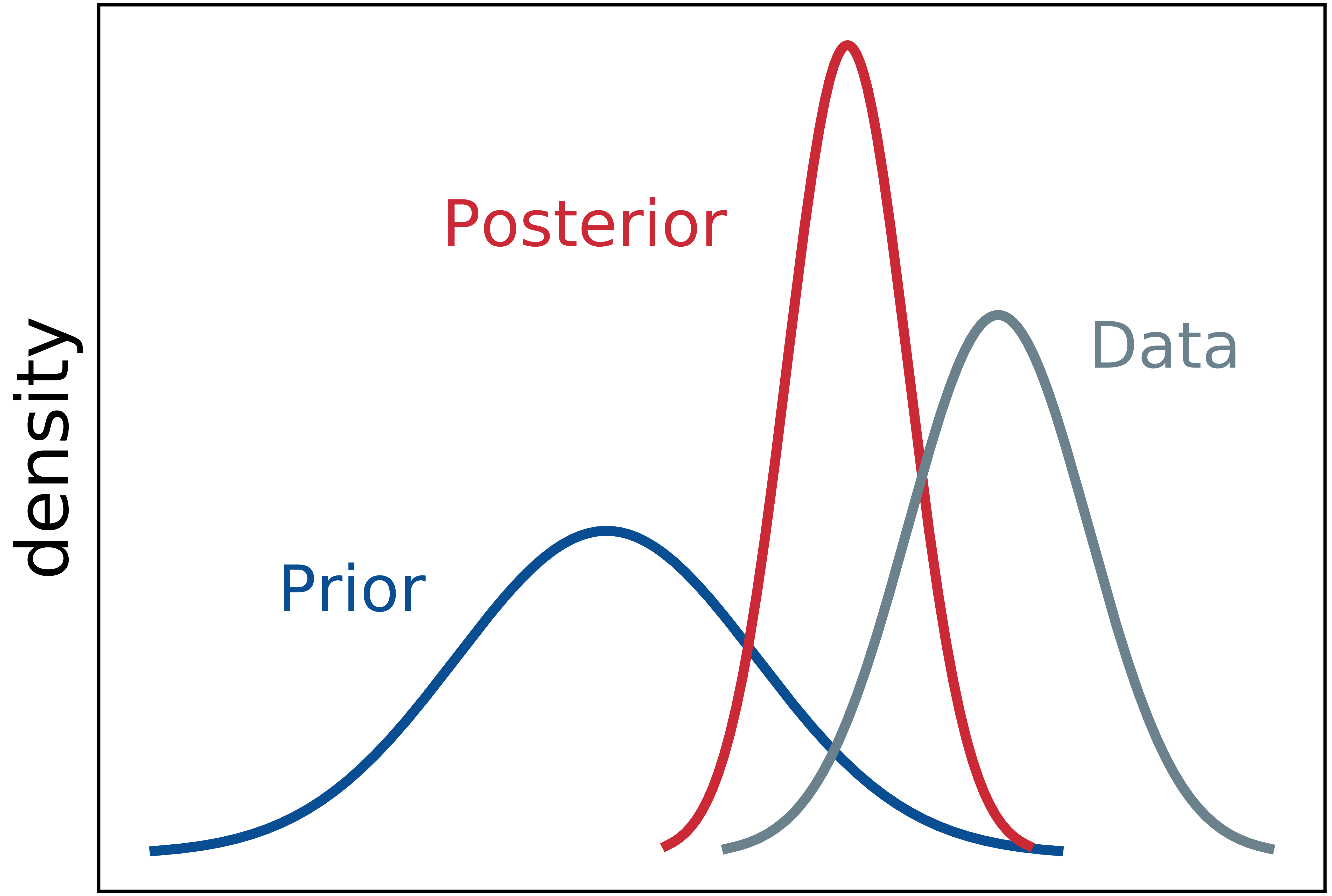
The best scenario is **both spectroscopy and imaging**. Imaging probes **dust and mass** constraints, while spectroscopy yields detailed **star formation history and heavy element abundance.**



Bayesian Thinking is Powerful for Interpreting Galaxy Data

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

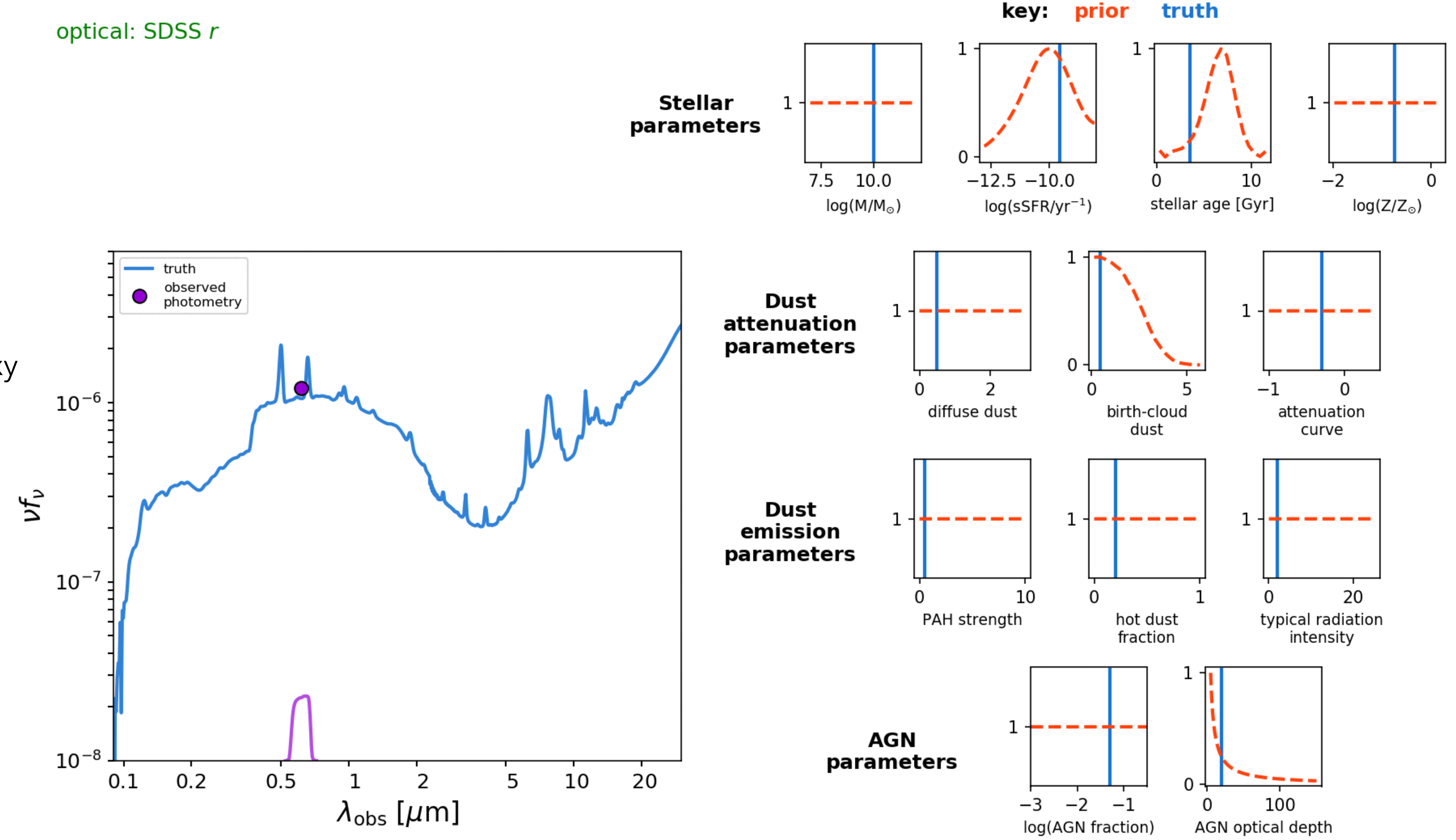
This is the 'sweet spot' for Bayesian statistics.



Bayesian Thinking is Powerful for Interpreting Galaxy Data

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?

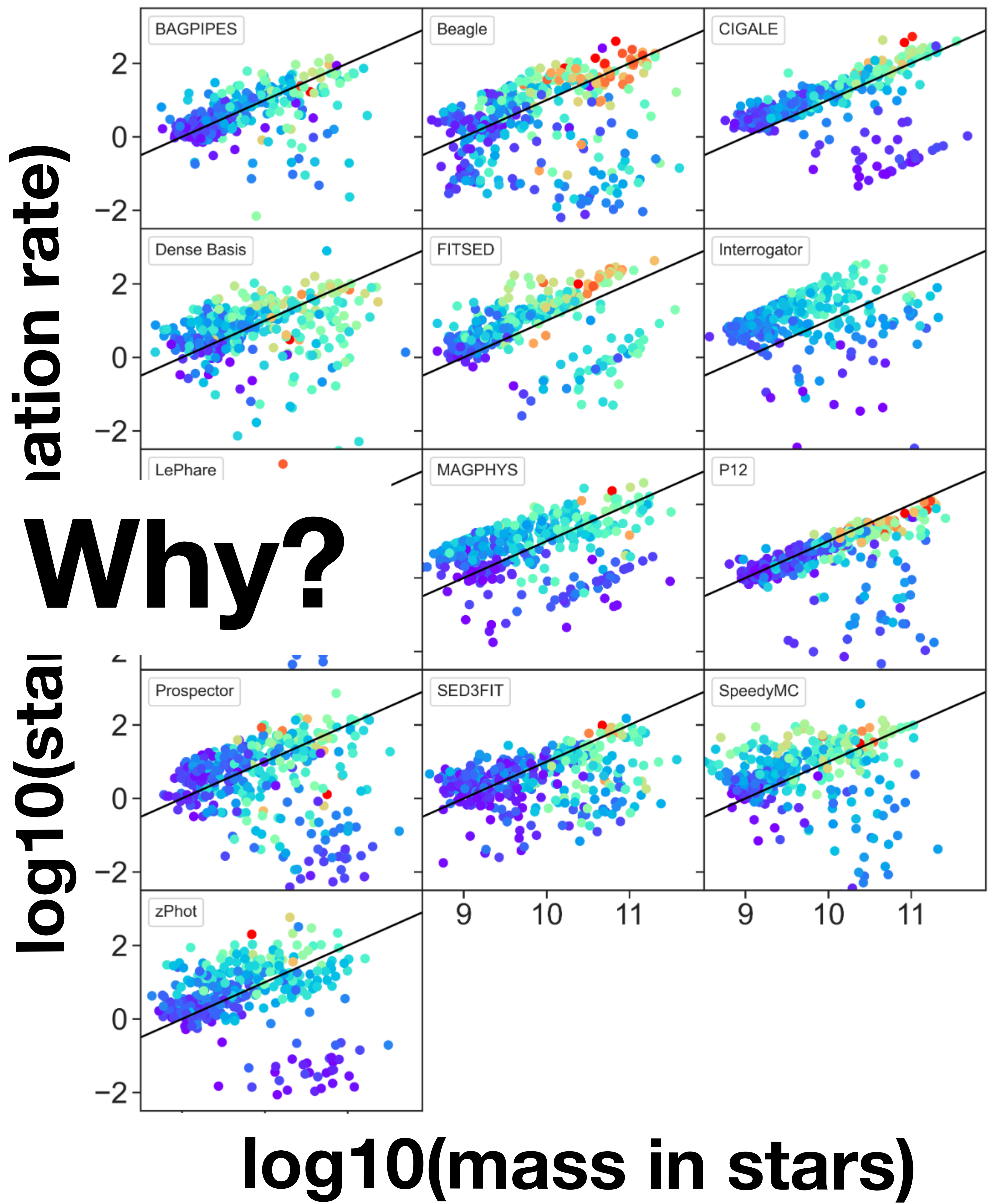
Systematic Uncertainties in Modeling Photometry of Distant Galaxies

A galaxy modeling experiment

Popular, different modeling codes applied to…

…**identical** high-quality space-based imaging…

… produce **qualitatively** and **quantitatively** different galaxy populations!



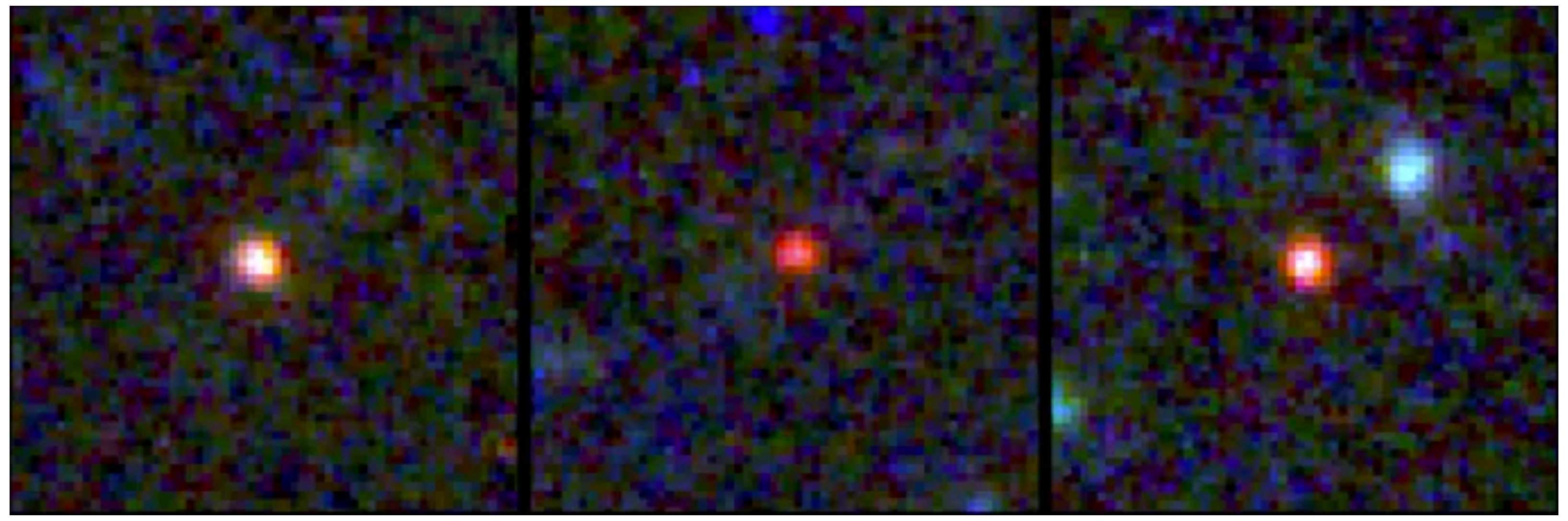
Pacifici et al. 2023 (+**Leja**)

Key Statistical & Modeling Challenges for Distant Galaxies

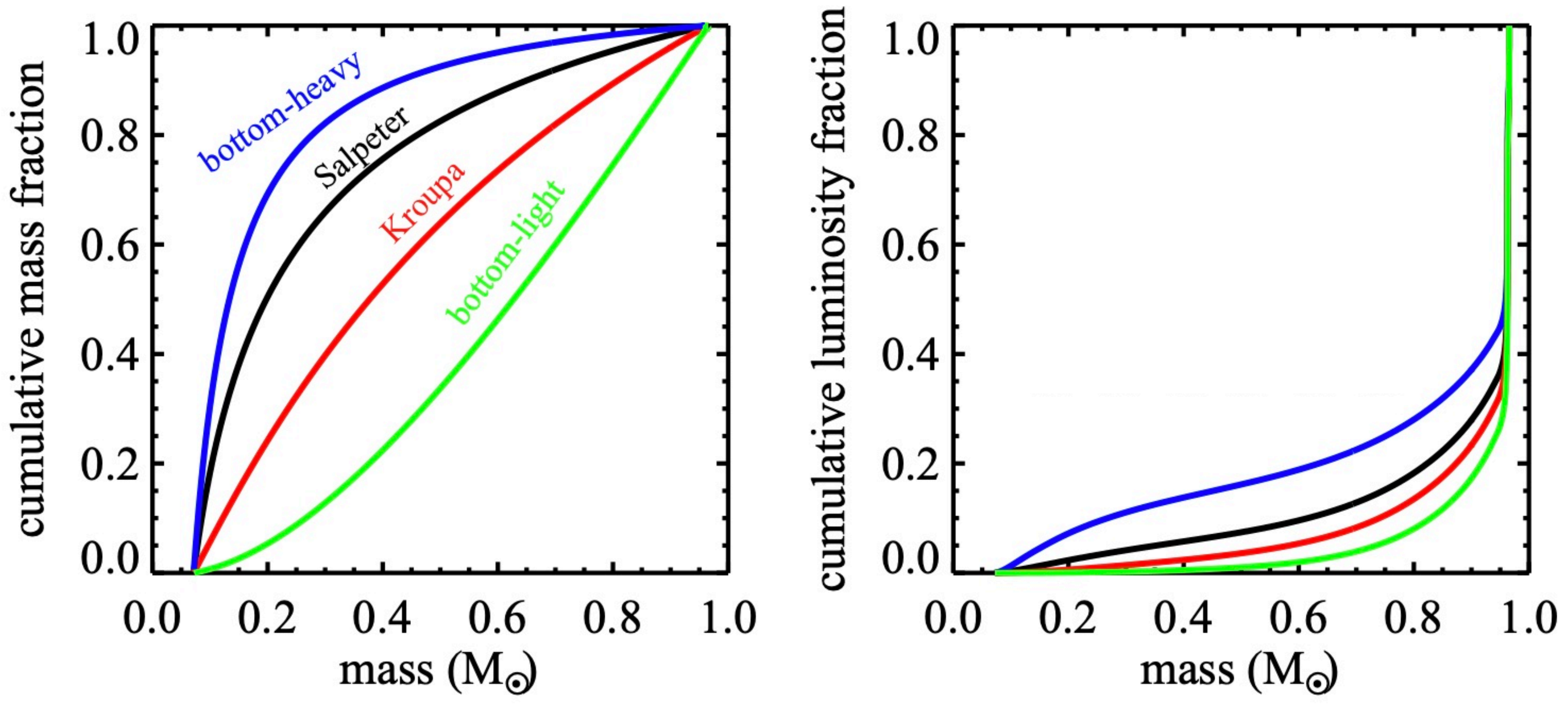
The Outshining Problem

- Distant galaxies are nearly point-sources, so light from every star, nebula, black hole, etc is combined linearly.

- Most of the light comes from a few **very bright objects** (e.g. O-stars, red giants); much of what we want to know is the great majority of **dim things** (e.g., older, low-mass stars)



Labbé+23



Conroy+13

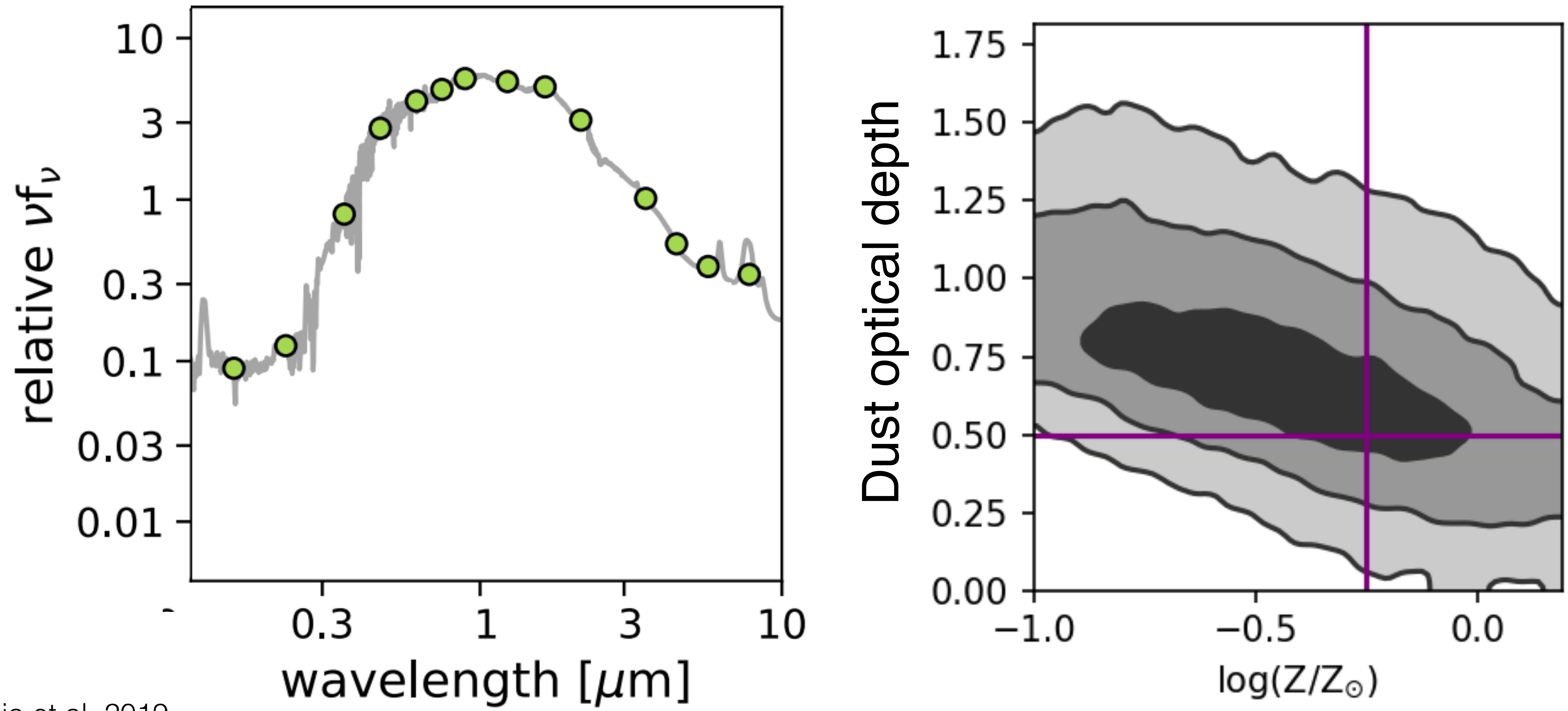
Inferring the properties of distant galaxies dominated by modeling choices:
what do bright things tell us about the rest of the system?

True size of Andromeda on sky, compared to full moon

Most Parameters are Degenerate, Especially With Only Imaging

Star formation history, metallicity, dust all have similar effects on imaging

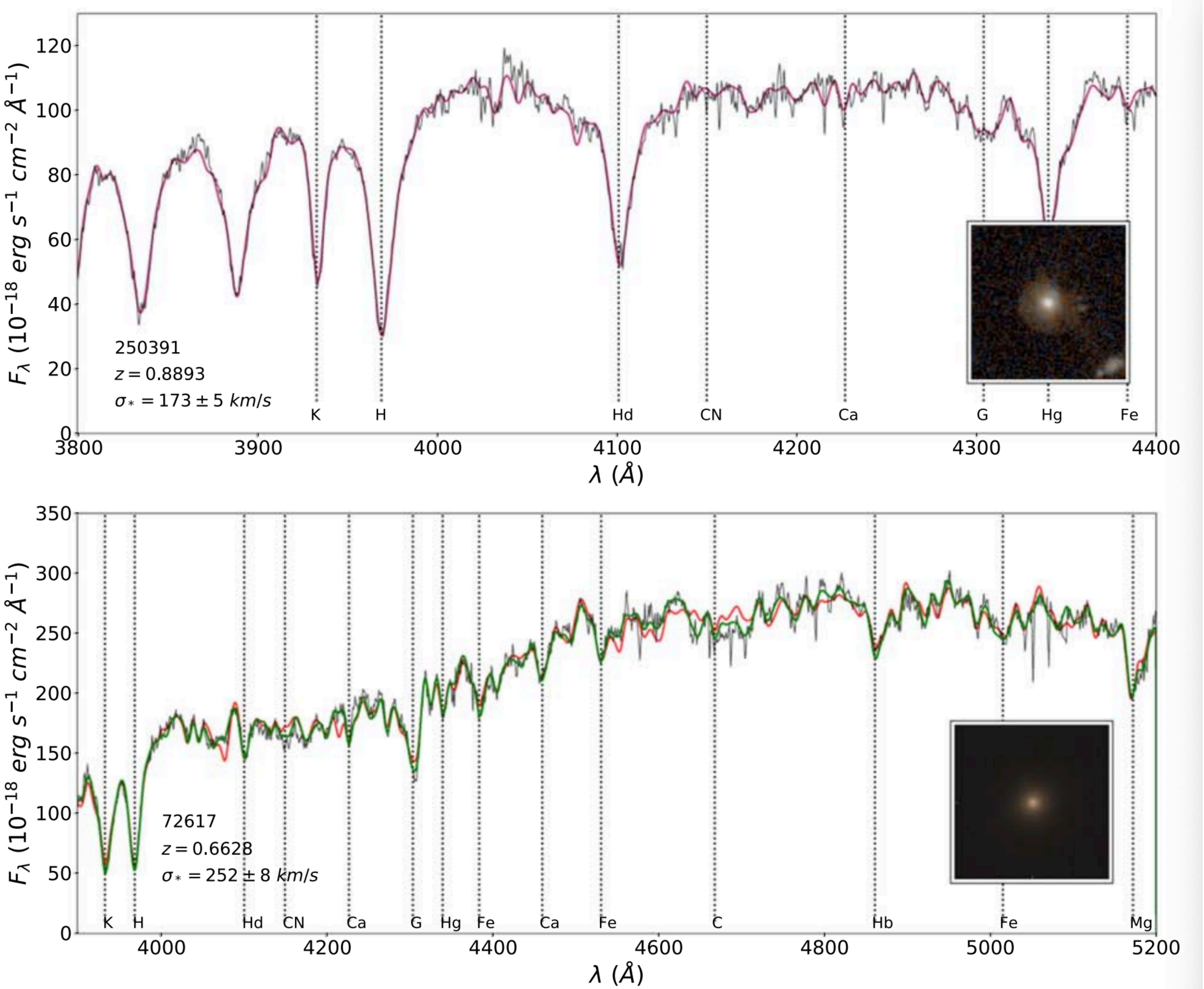
–> limited to mapping out inherent degeneracies or adopting informed priors from better data!



Leja et al. 2019

Imaging Isn't Enough. What About The "Best Distant Galaxy Spectra We've Ever Had"?

The LEGA-C survey: ~3000 galaxies at 0.6 < z < 1, with ~20 hour rest-optical spectra on an 8.2 meter telescope (VLT), yielding signal to noise~ 20-70 per wavelength element with excellent resolution



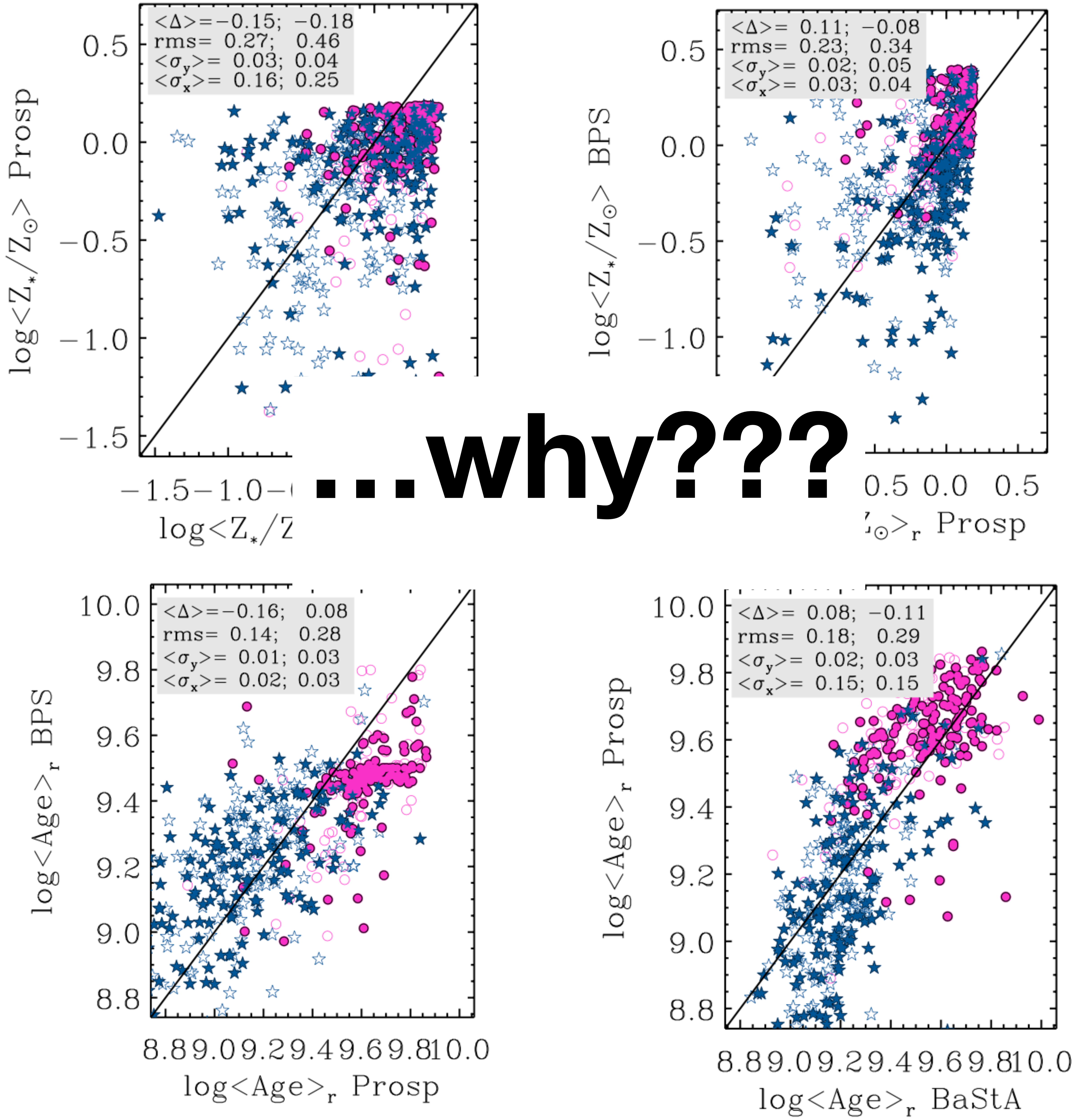
Van der Wel et al. 2021

Comparing Results of LEGA-C Spectra from Leading Analysis Codes

Stellar metallicity

"Light-weighted" age

(Mass-weighted not shown; much worse)



...why???

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 create systematic uncertainties.

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

SED Parameter	approximate effect on SED
stellar mass	orders of magnitude

(ad infinitum)

Complex Emergent Behavior Means Key Physics Are Uncertain

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

Below are the formation histories of 4-5 galaxies, under different simulation rules.

Iyer et al. 2020

Bayesian Inference: Powerful but Dangerous

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

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.*

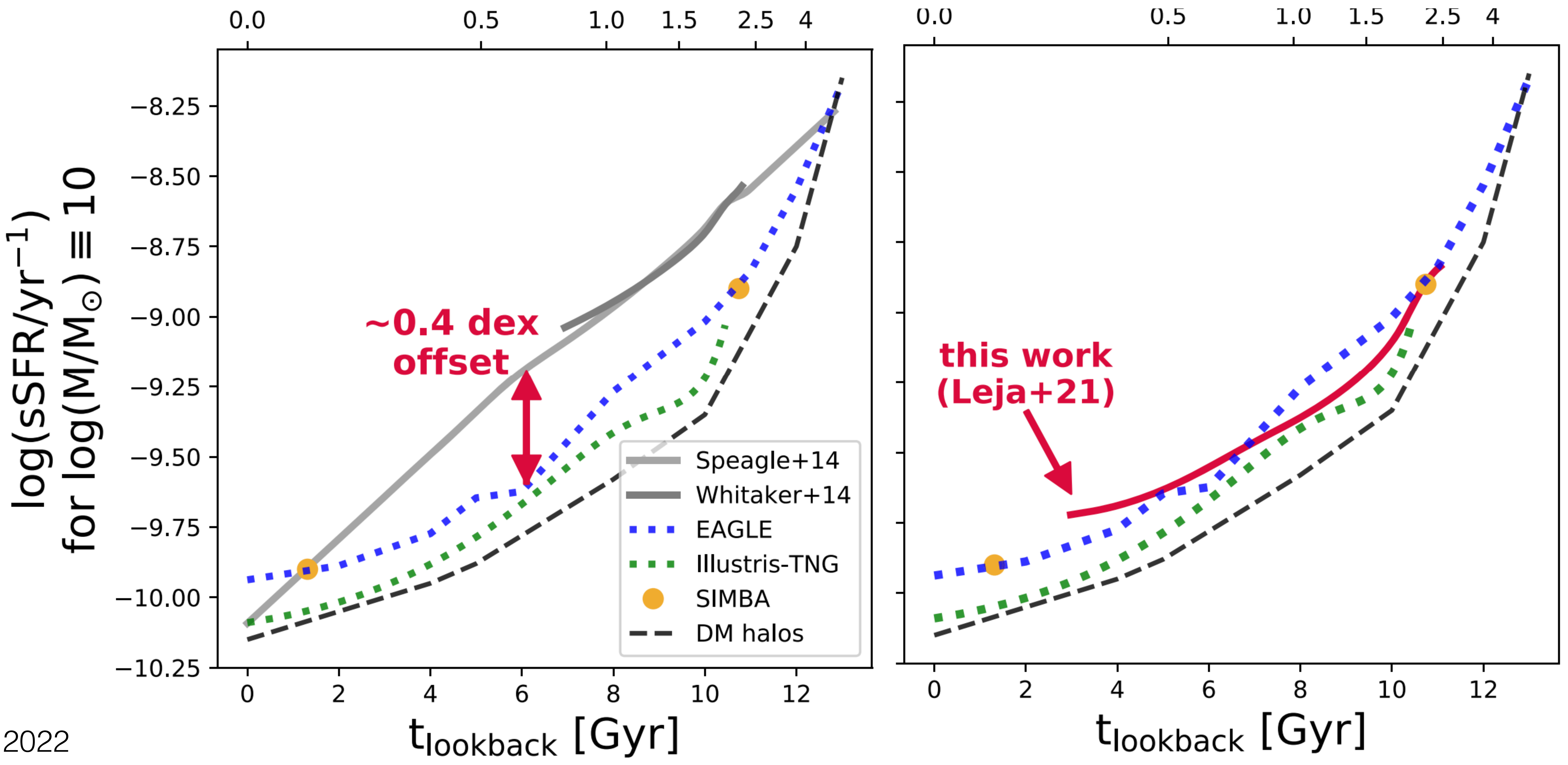
So, are our models accurate enough to be useful?

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

But - this was computationally expensive, and many puzzles remain



Leja et al. 2022

High Dimensionality Already Pushing Computational Limits

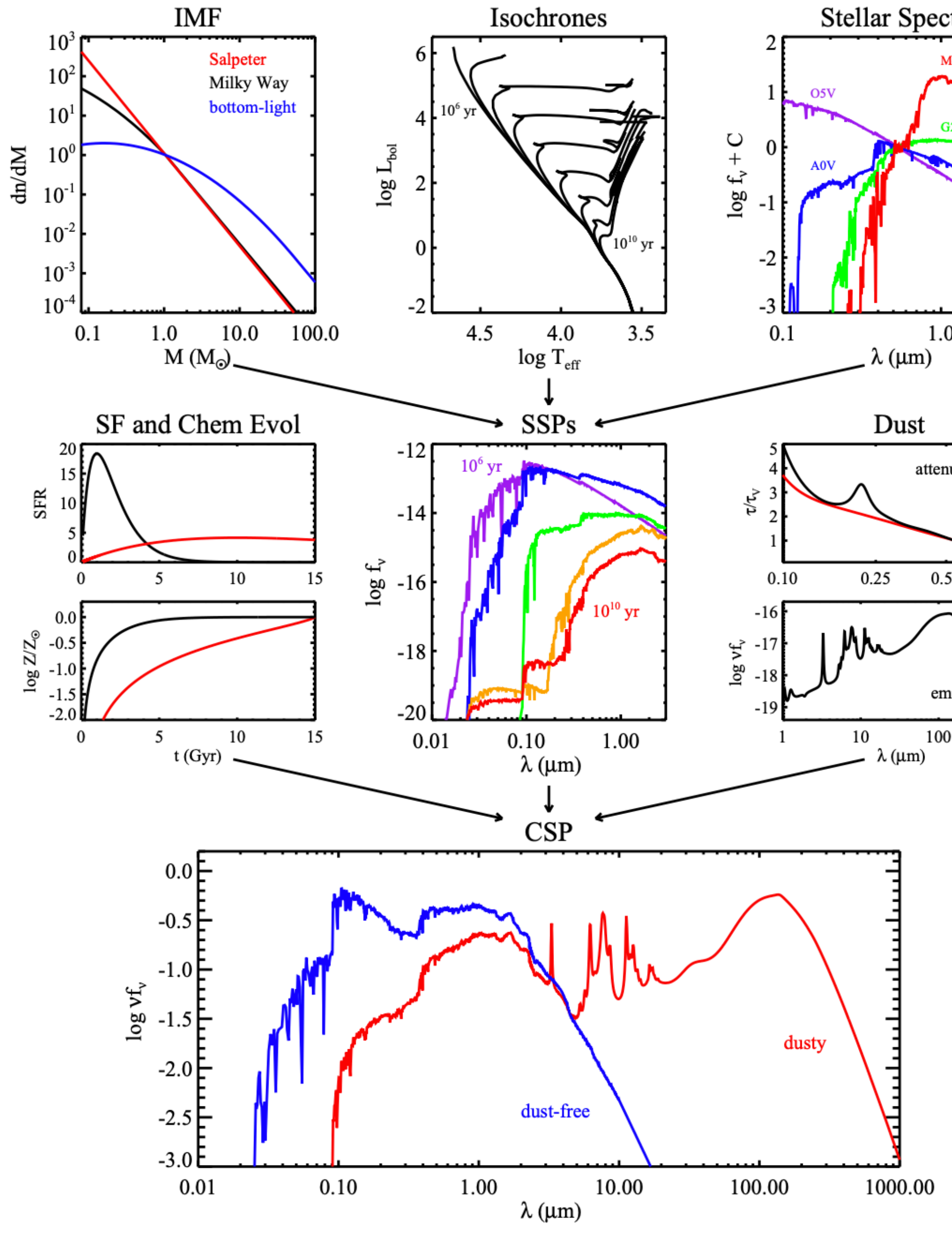
High-dimensional models cannot live on a grid (*curse of dimensionality*): this means each model must be generated **on-the-fly**, a compute-intensive task (~15 hours/fit!).

- **several million CPU-hours** to analyze ~10^5 objects in deep field

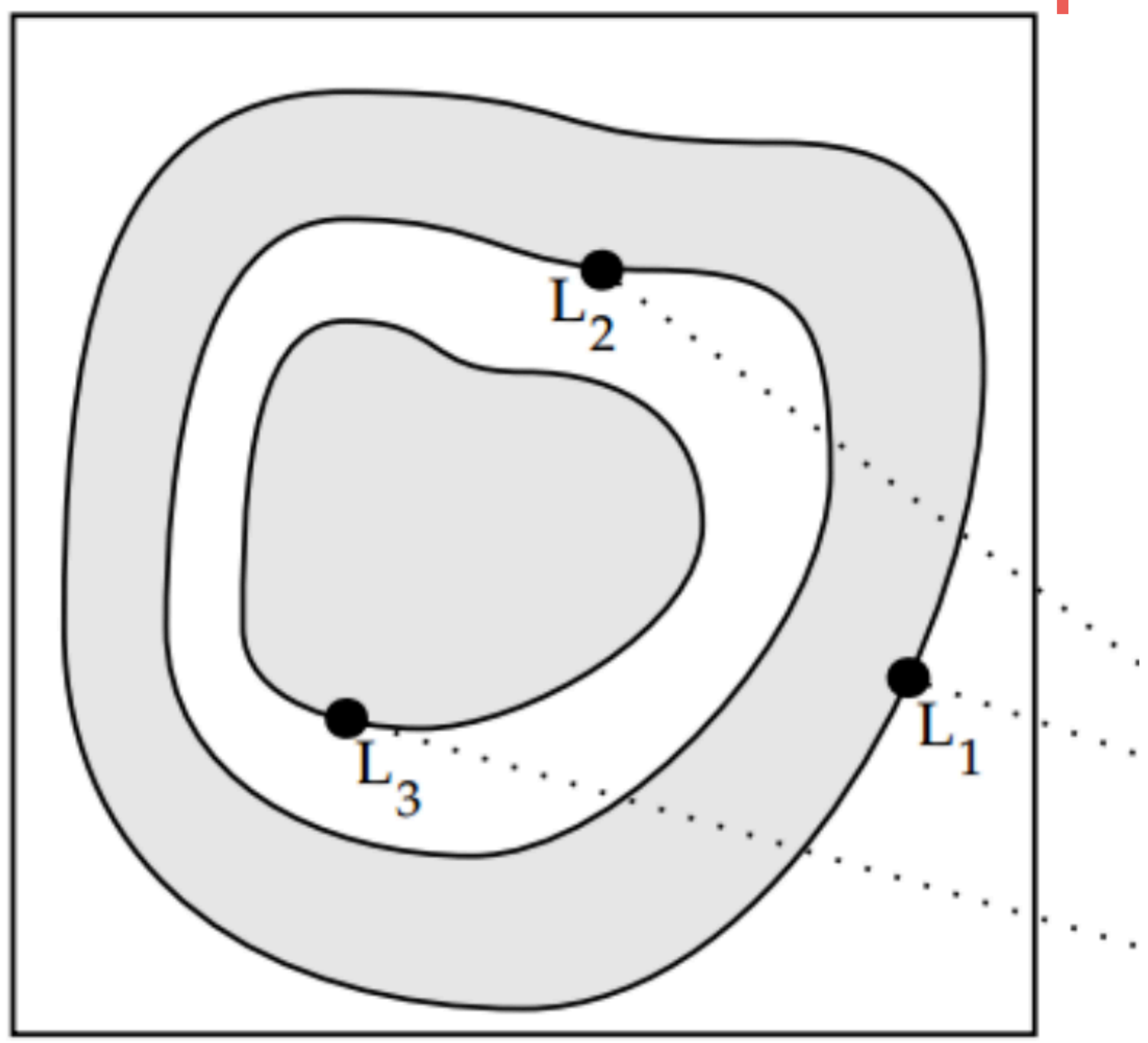
What is driving the computational requirements?

Galaxy observations

0.05s / model



~1 million models per fit



Parameter space

= 15 hours/ object

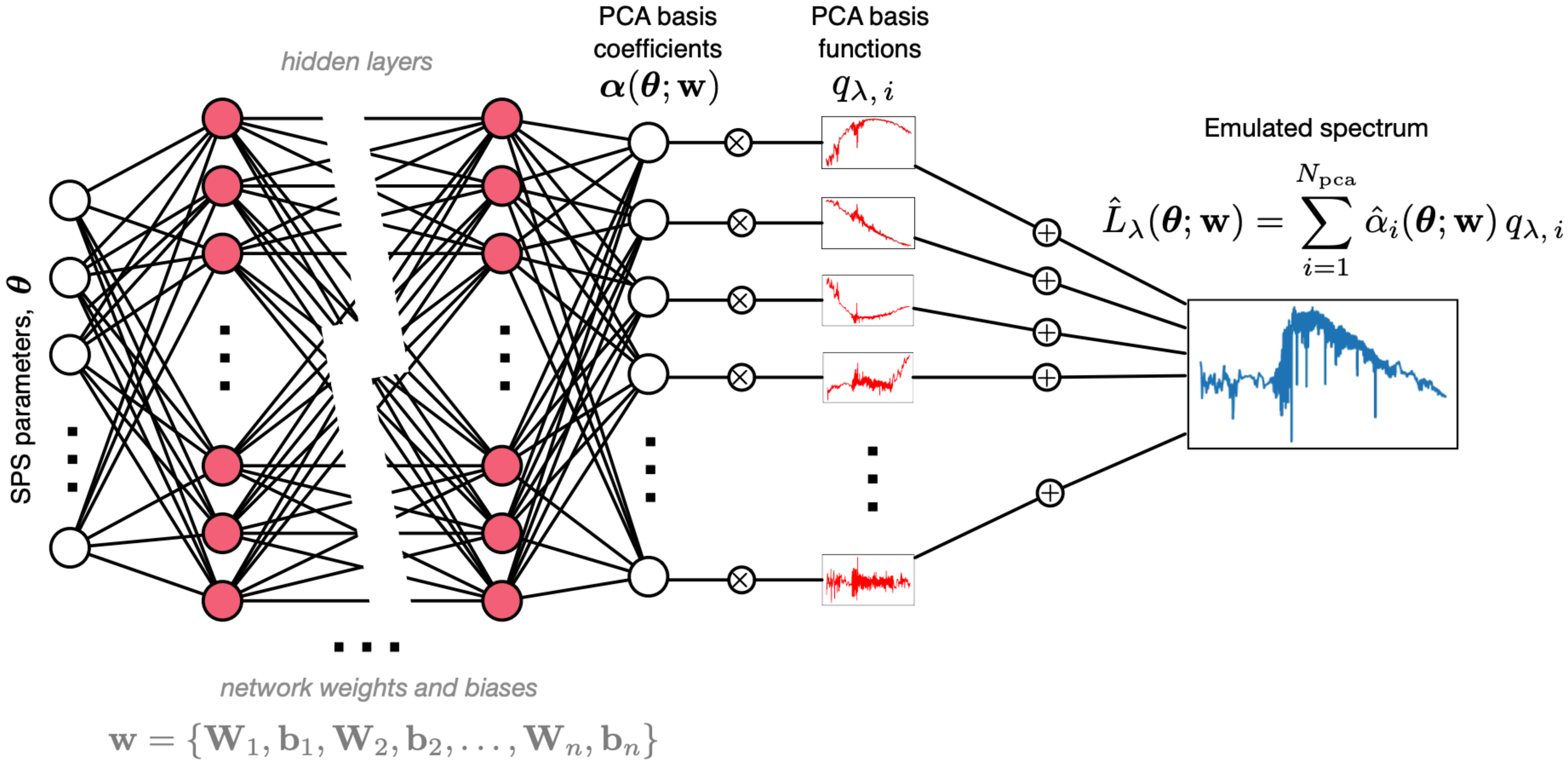
solutions

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 ~10^5 objects in deep field

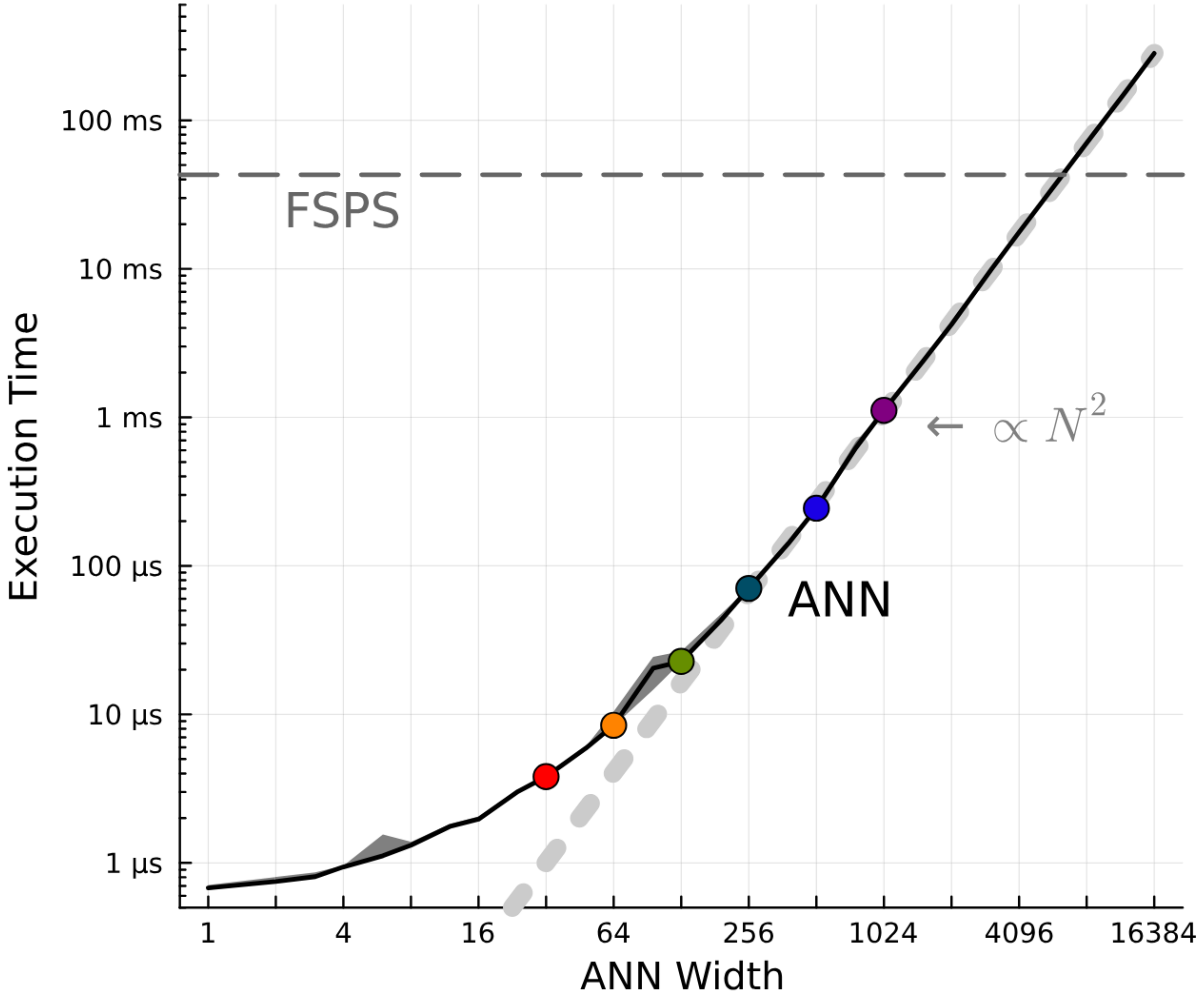
Neural net emulation of **photometric predictions** reduces model generation time by **~100-1000 (10^4 on a GPU)**



Alsing, Peiris, Leja et al. (2020)

Mathews, Leja et al. (2023)

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

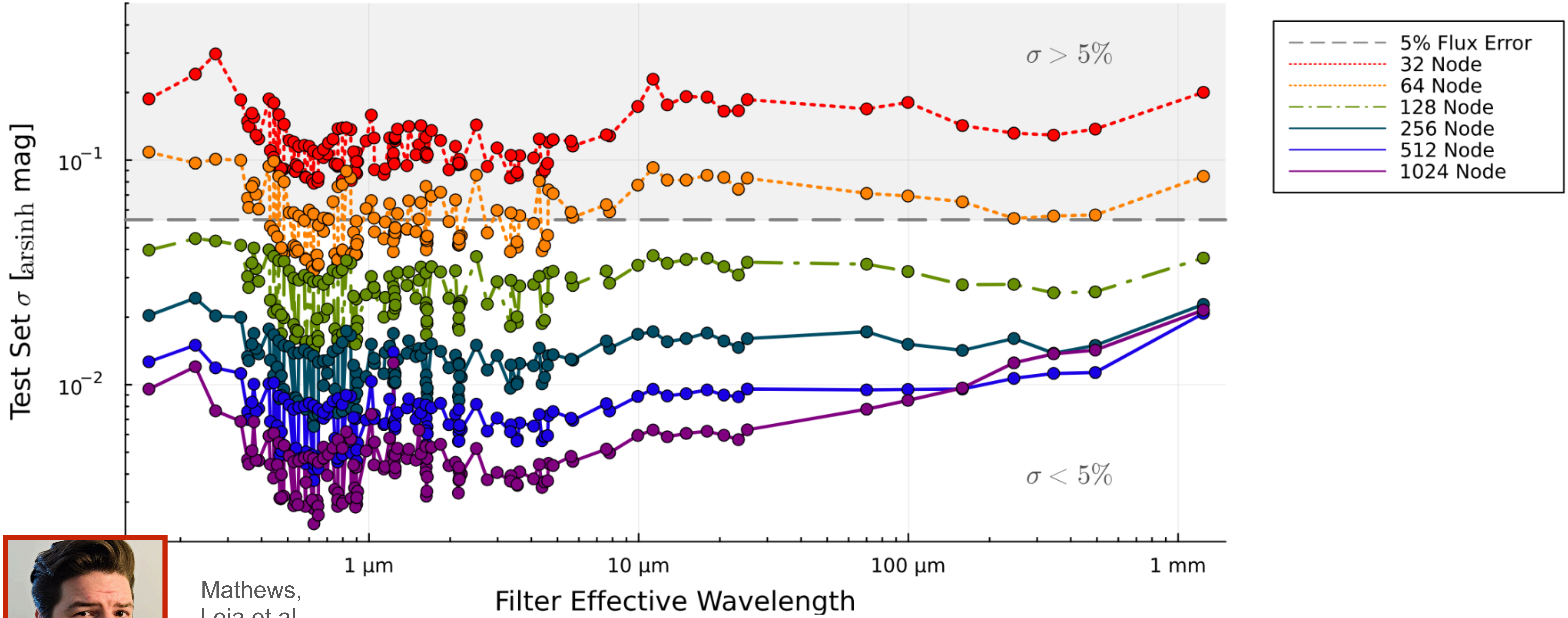


FSPS = classical stellar
population
synthesis (Conroy et al. 2009)

ANN = artificial neural network

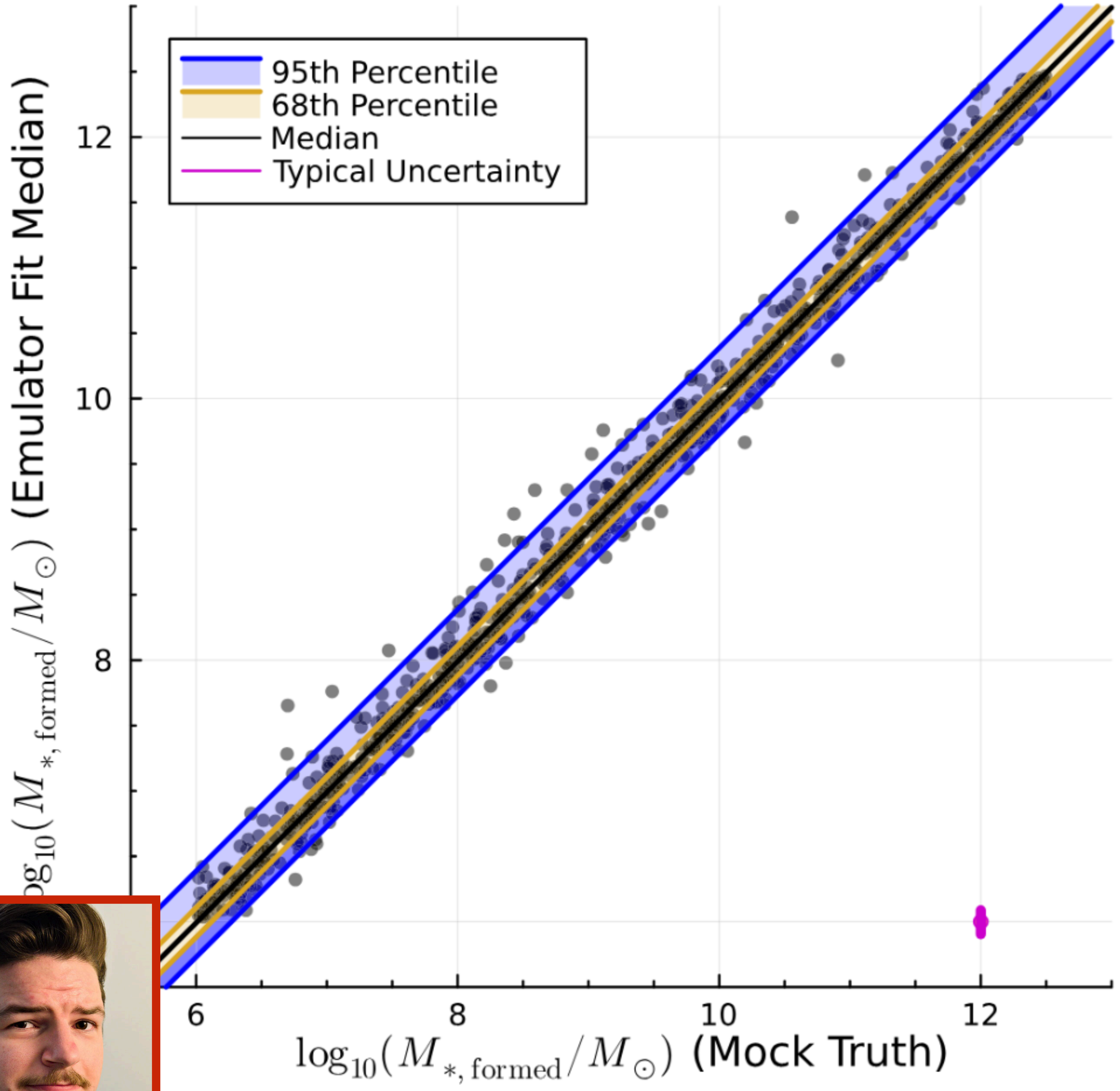
Mathews,
Leja et al.
(2023)

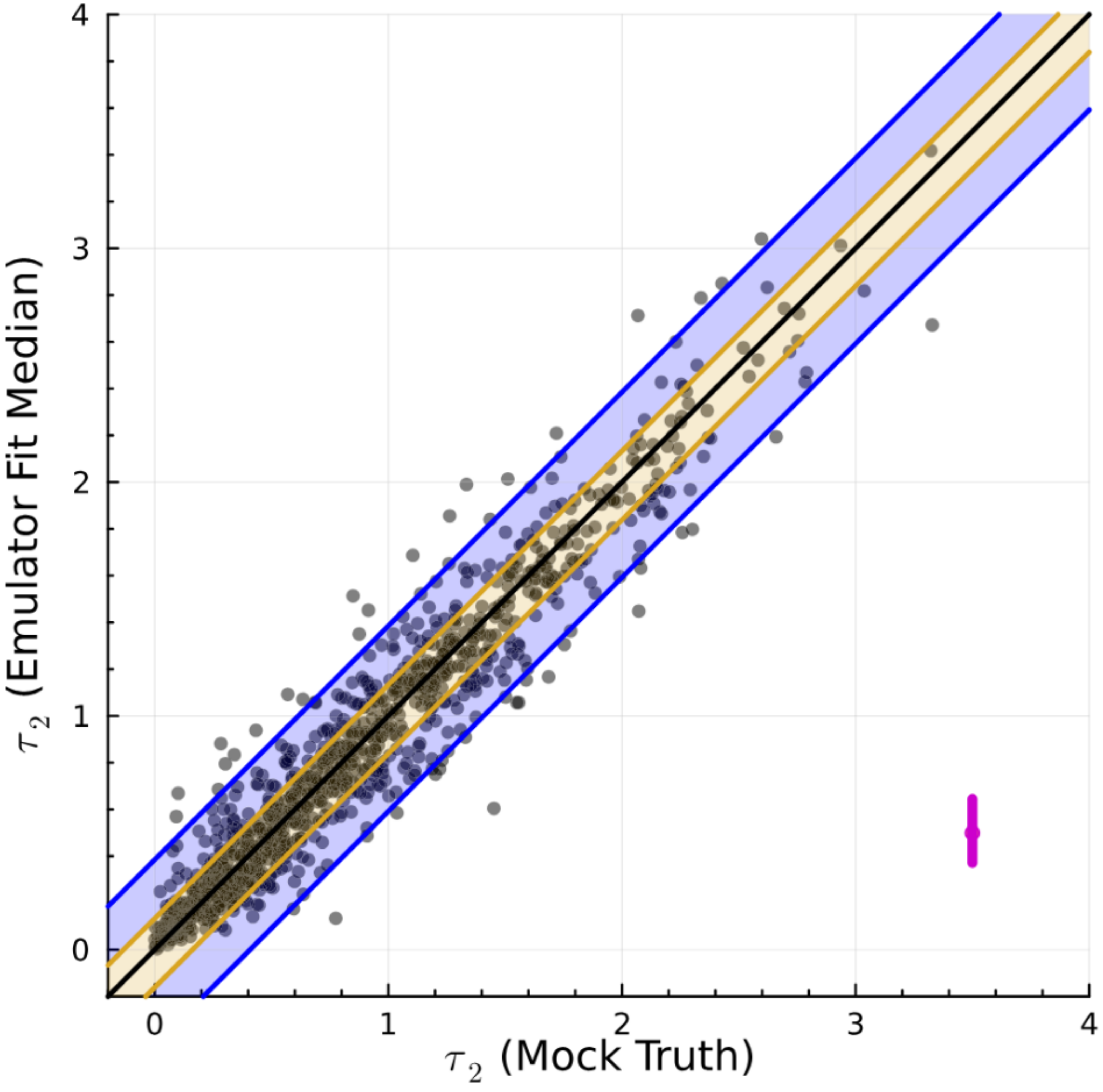
Neural networks are ***precise*** (if they're big enough; <[0.5 x obs error] works)



Mathews, Leja et al. (2023)

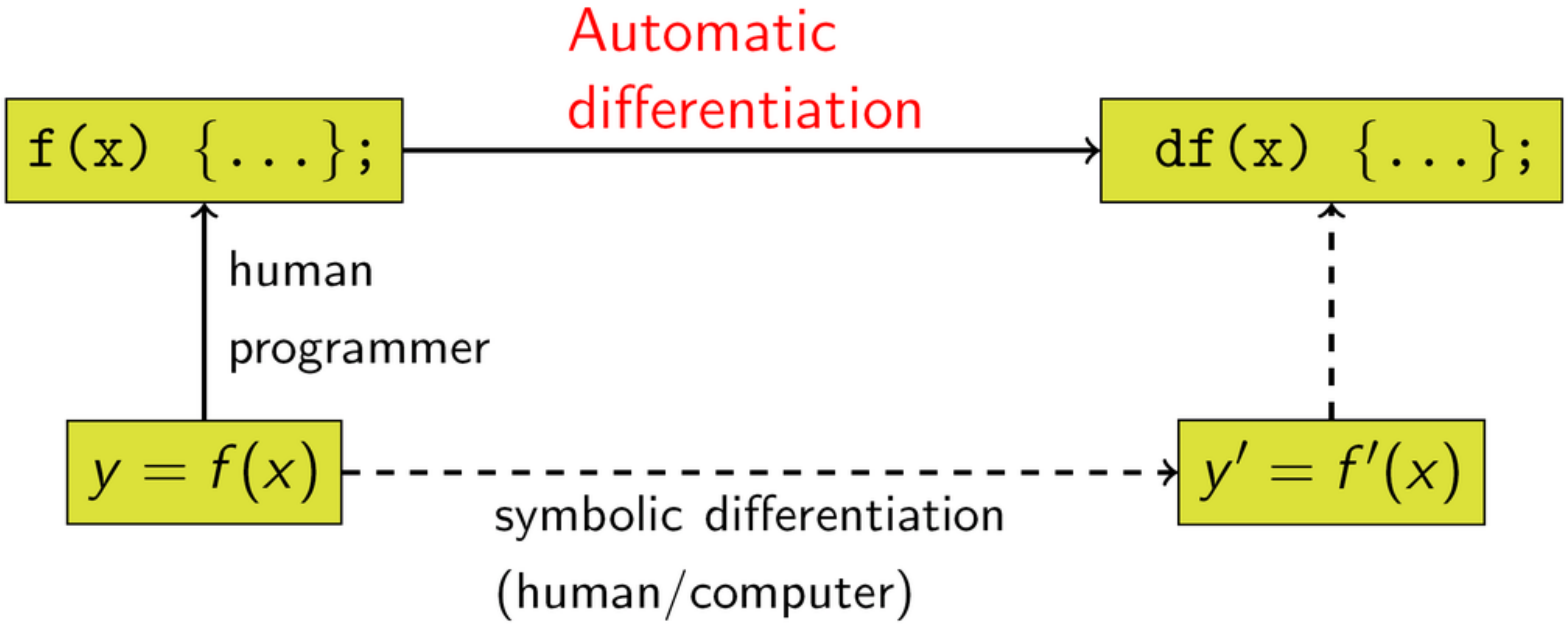
Neural networks ***give accurate posteriors***





Mathews, Leja et al. (2023)

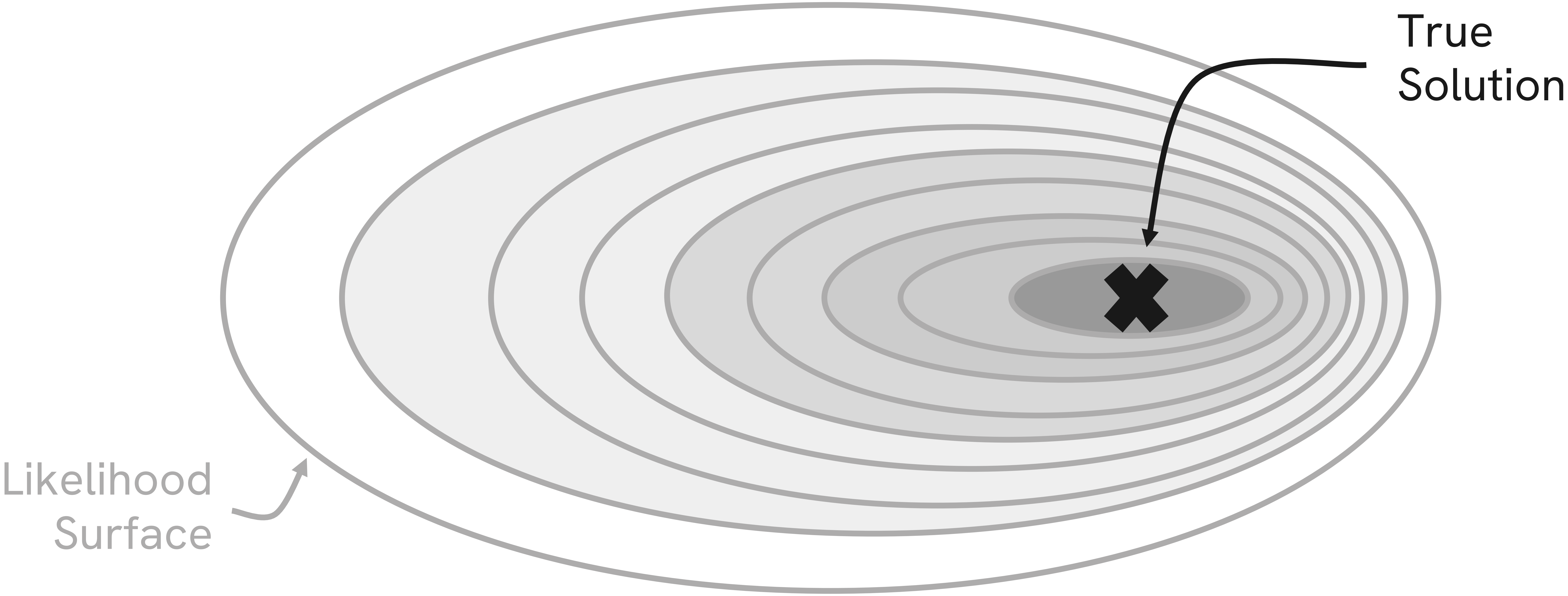
Neural Networks are ***differentiable***



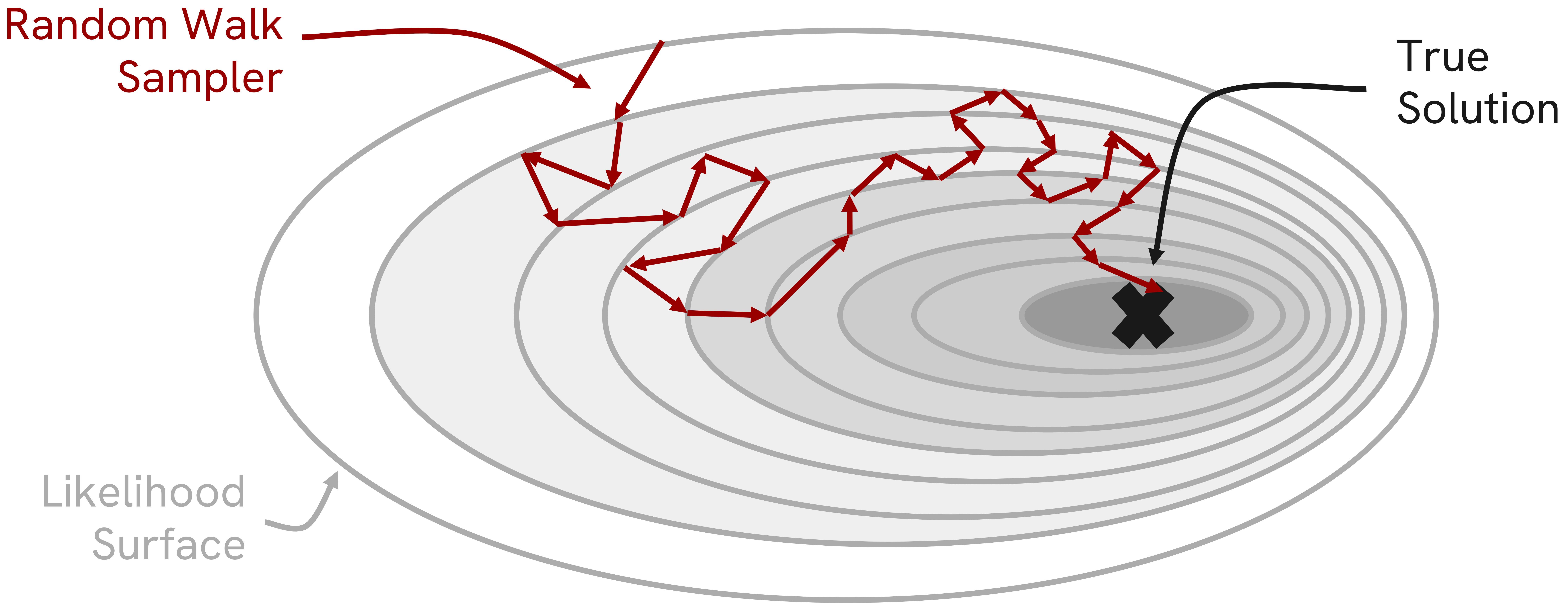
(from Wikipedia)

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

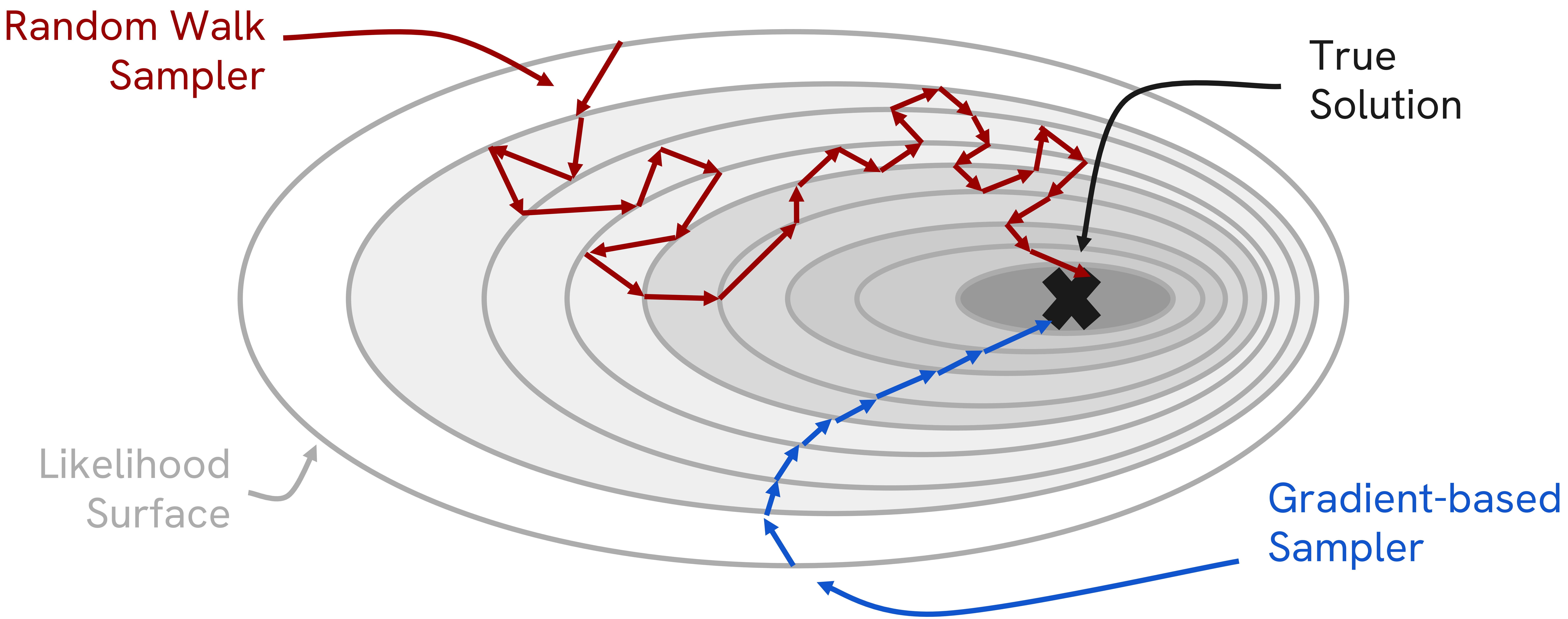
Differentiability Allows Highly Efficient Sampling



Differentiability Allows Highly Efficient Sampling

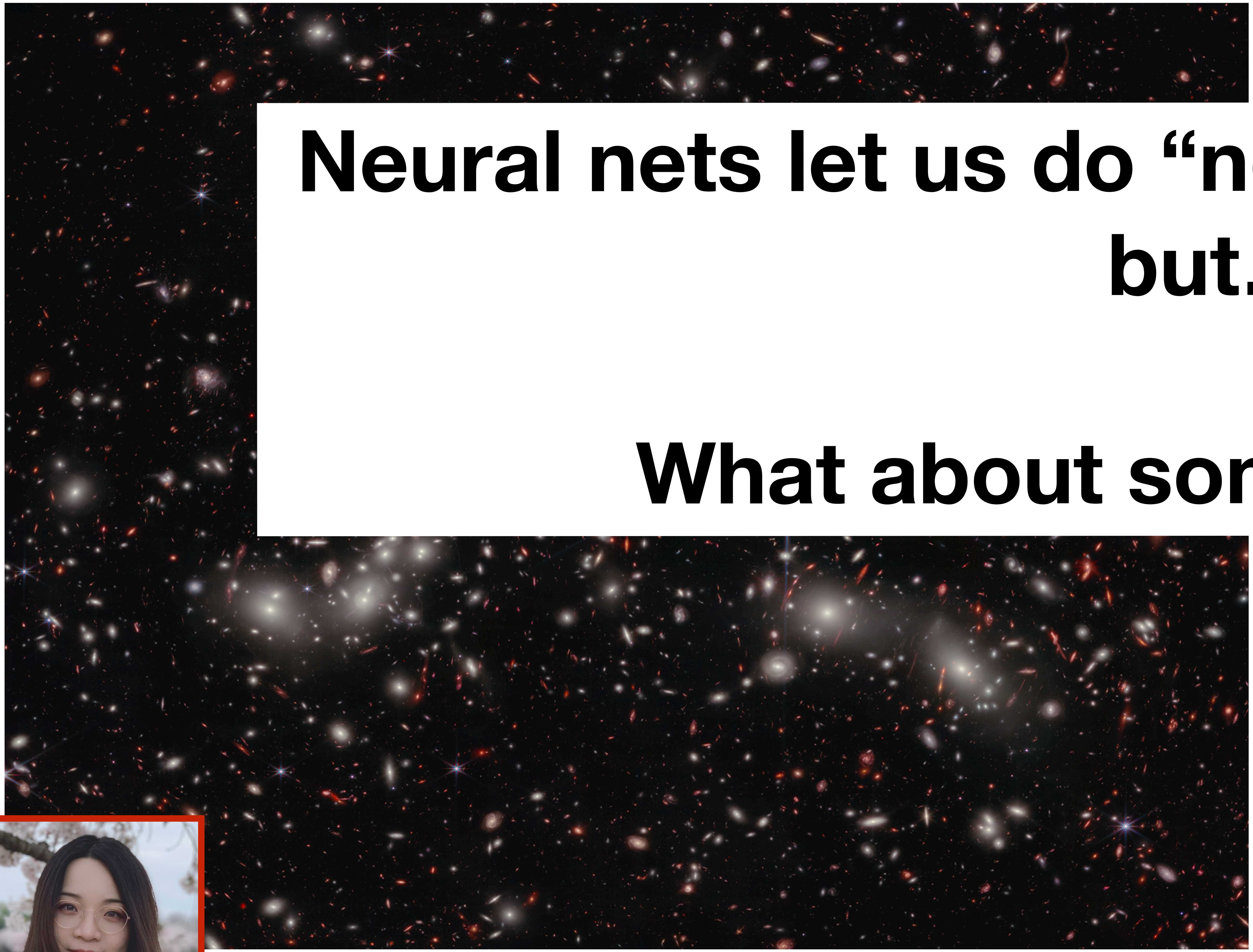


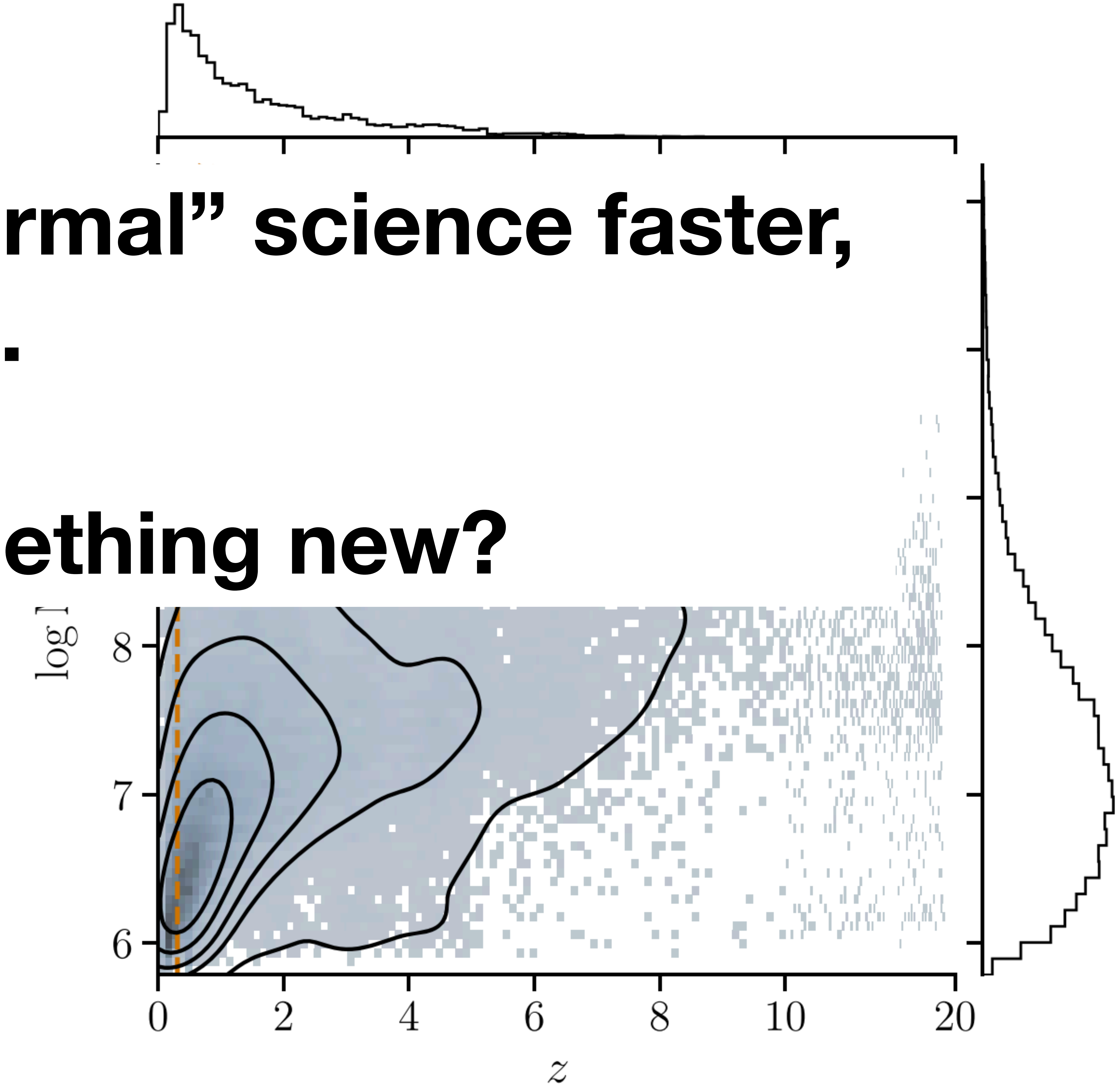
Differentiability Allows Highly Efficient Sampling



Gradient-based samplers are efficient at proposing good samples.

The result: ~70,000 galaxies in one of Webb's first deep fields (UNCOVER; PIs Labbé, Bezanson) can be analyzed in a couple of days on a modest compute allocation (Wang, Leja+ 2024).
Can answer a variety of questions about early galaxies, including discovery of second- and fourth-most distant galaxies (z=12.4, z=13.1; Wang et al. 2024, incl. Leja)





Neural nets let us do "normal" science faster, but…

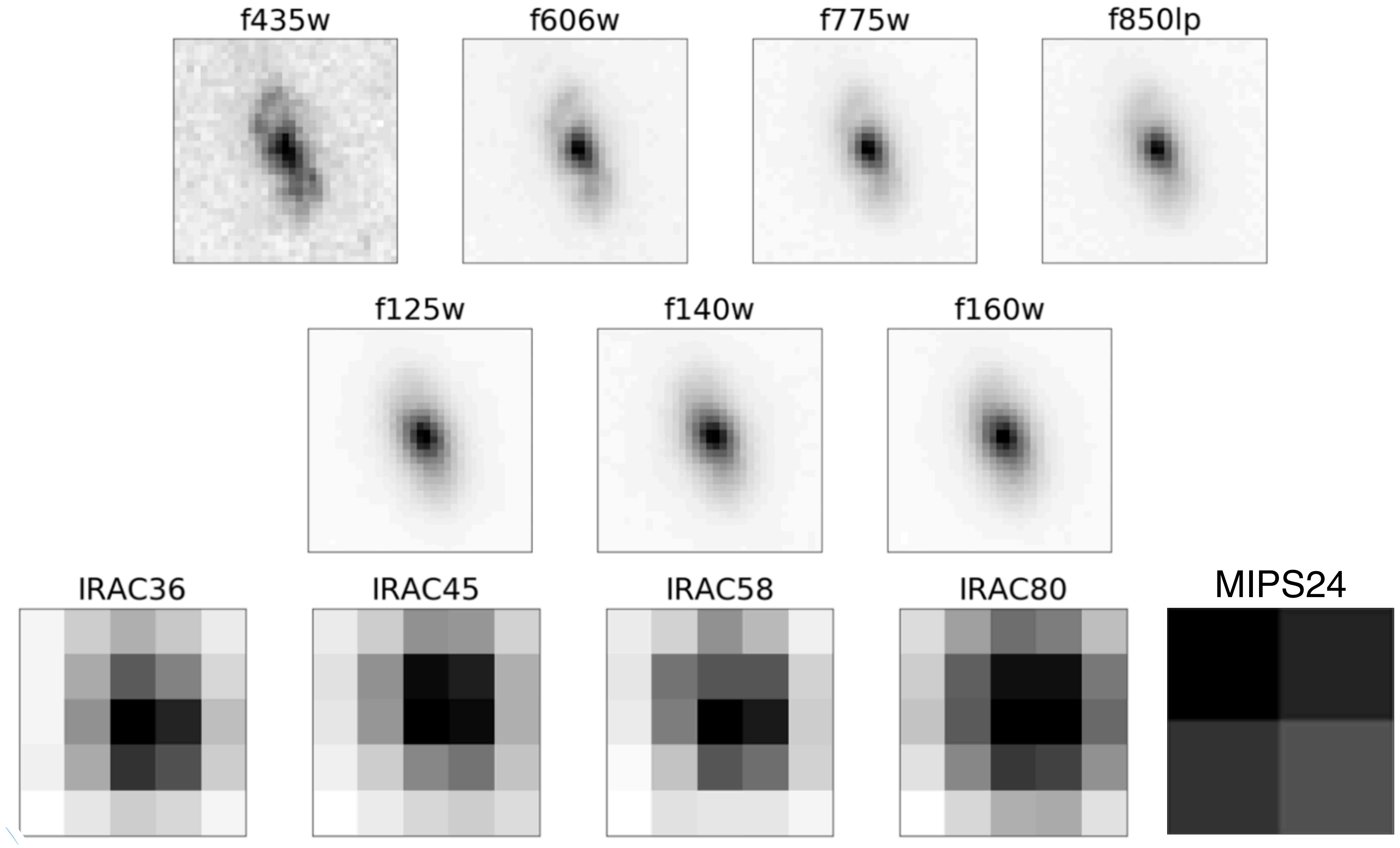
What about something new?

Recall: We treat *stellar populations* (not stars!) in distant galaxies as Unresolved

But - they aren't!

Andromeda

A typical distant galaxy

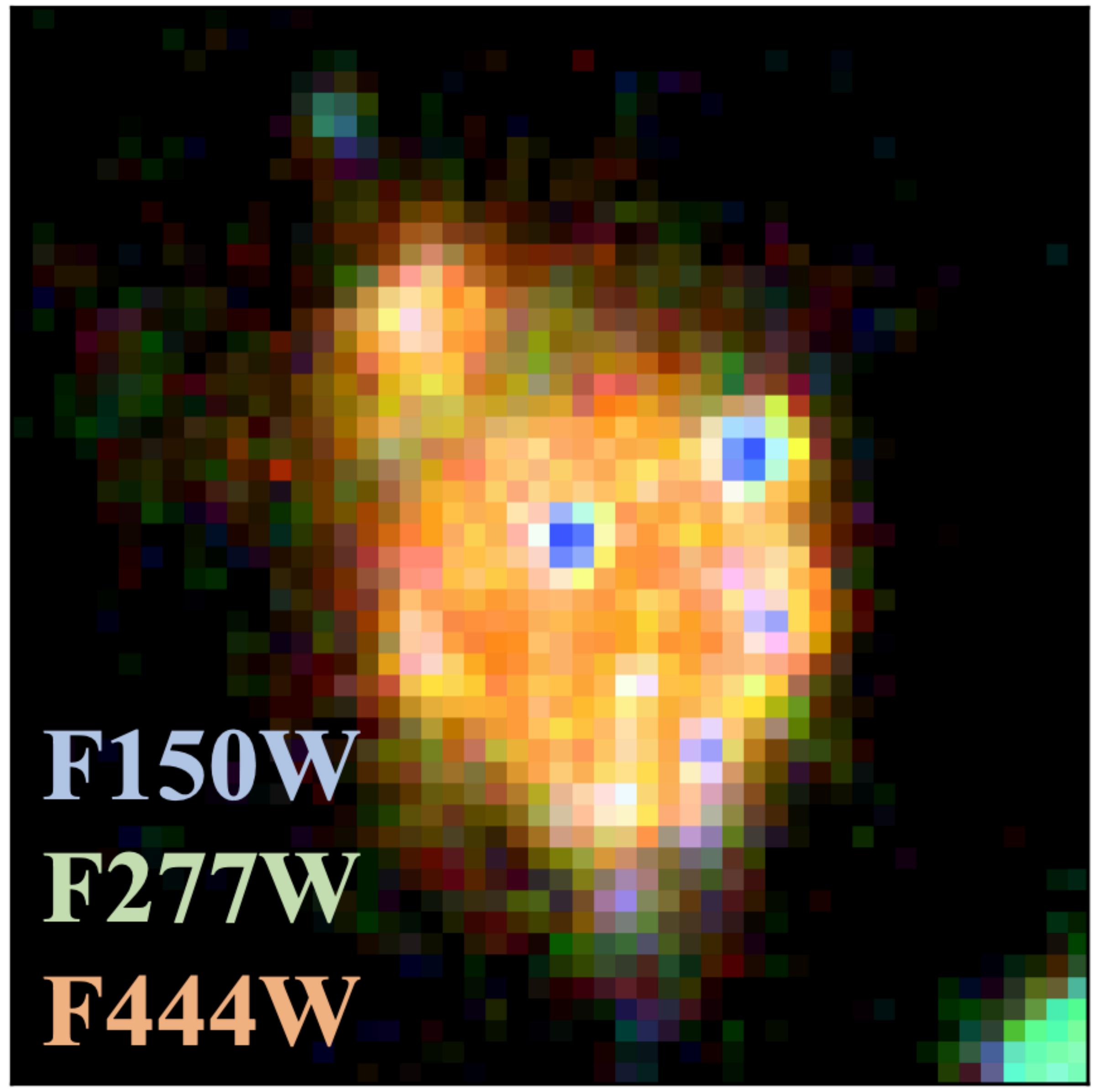


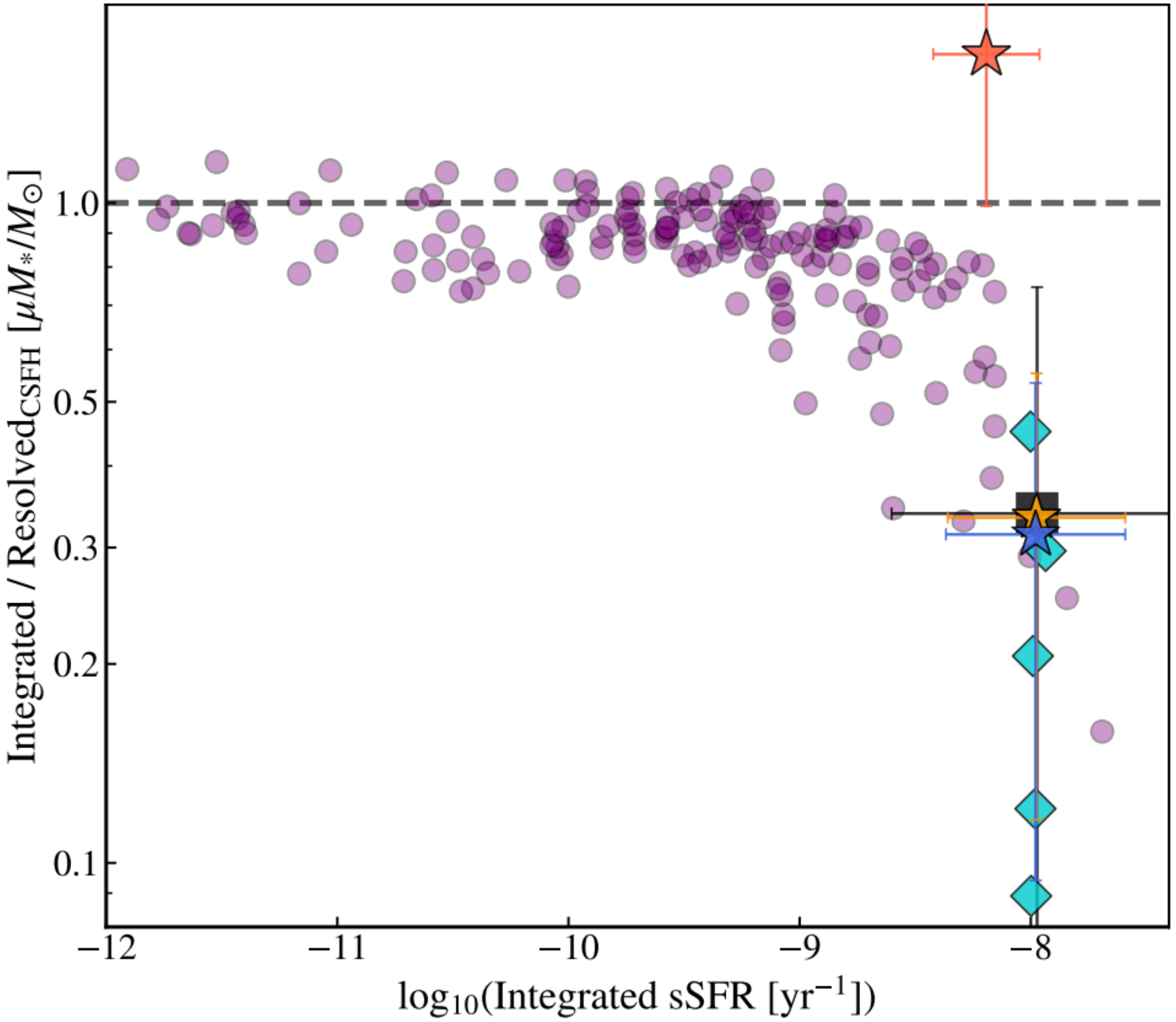
In principle we can fit every pixel - this has key advantages ! Break underlying degeneracies, learn fundamentally new things about distant galaxies!

Key challenge: pixels are not independent - correlated by limited resolution which spreads out light. So we must forward-model hundreds or thousands of pixels .. ***SIMULTANEOUSLY***

New Constraints: Bursty Star Formation And Outshining

JWST sees spatial complexity of distant galaxies, revealing older stars normally hidden by bright young stars. Resolved modeling boosts inferred stellar mass up to 5x for galaxies forming lots of stars.





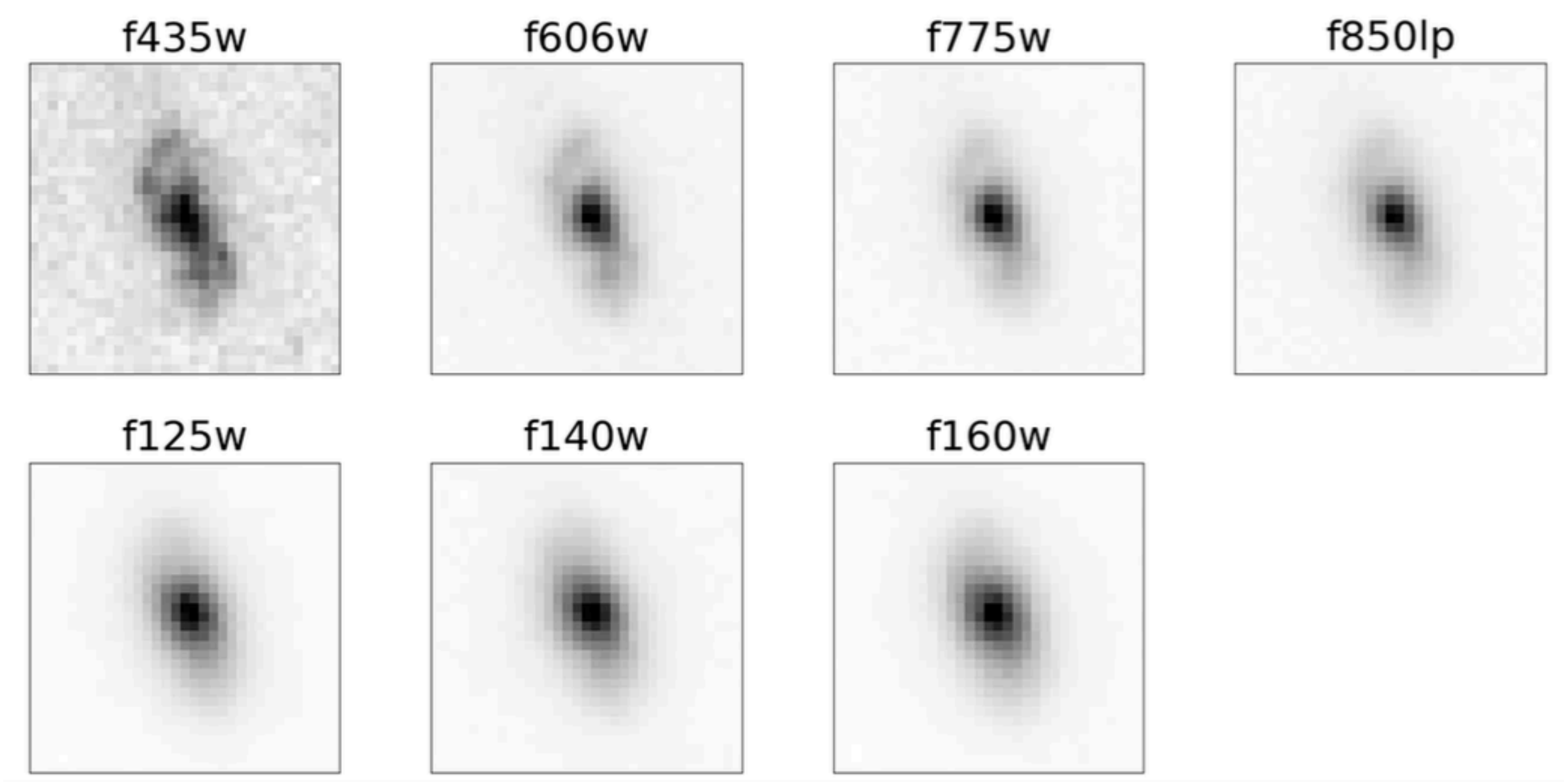
Giménez-Arteaga+24
see also Sorba & Sawicki 2018

Current techniques emerging treat every pixel like a little isolated galaxy, and smear all imaging to match lowest resolution. This throws (lots) of information and ignores correlations from convolution. With ML-acceleration we can do better!

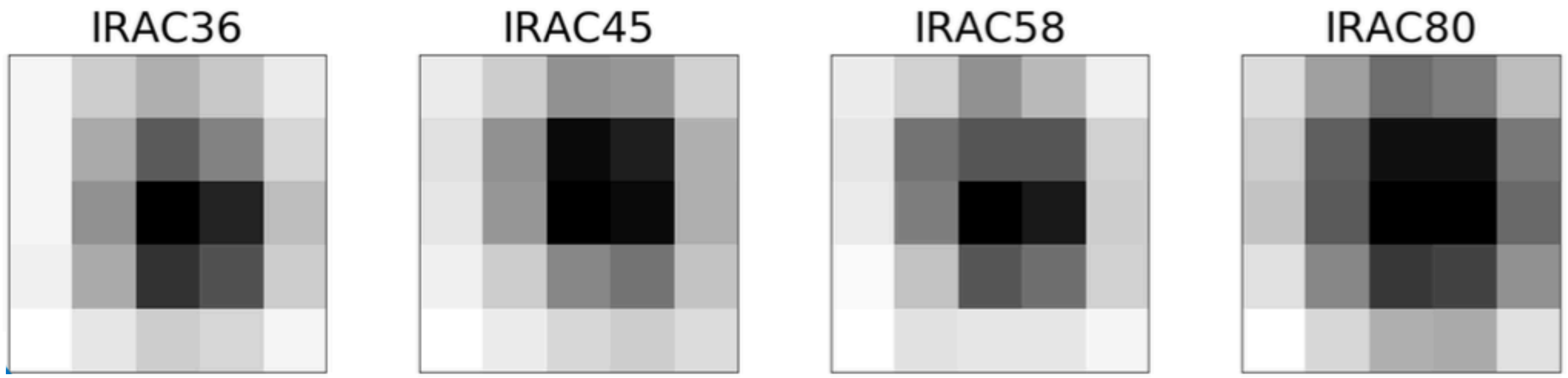
New Approach: Fit Every Pixel!

- New approach, "pirate", forward-models both stellar populations and light-smearing. Can simultaneously model images of any resolution.
- Give each pixel own stars, dust, black holes, …
- Neural nets give free derivatives - powers hyper-efficient Hamiltonian Monte Carlo sampler.
- Fitting a ~75x75 px image with 20 filters **finishes in about ~1 day**
  - To be clear, this is 75 * 75 * 15 = **84,375 free parameters**.
  - This would take about 112,000 core-hours with classic fitting - 5000x faster!
- So - does it work?

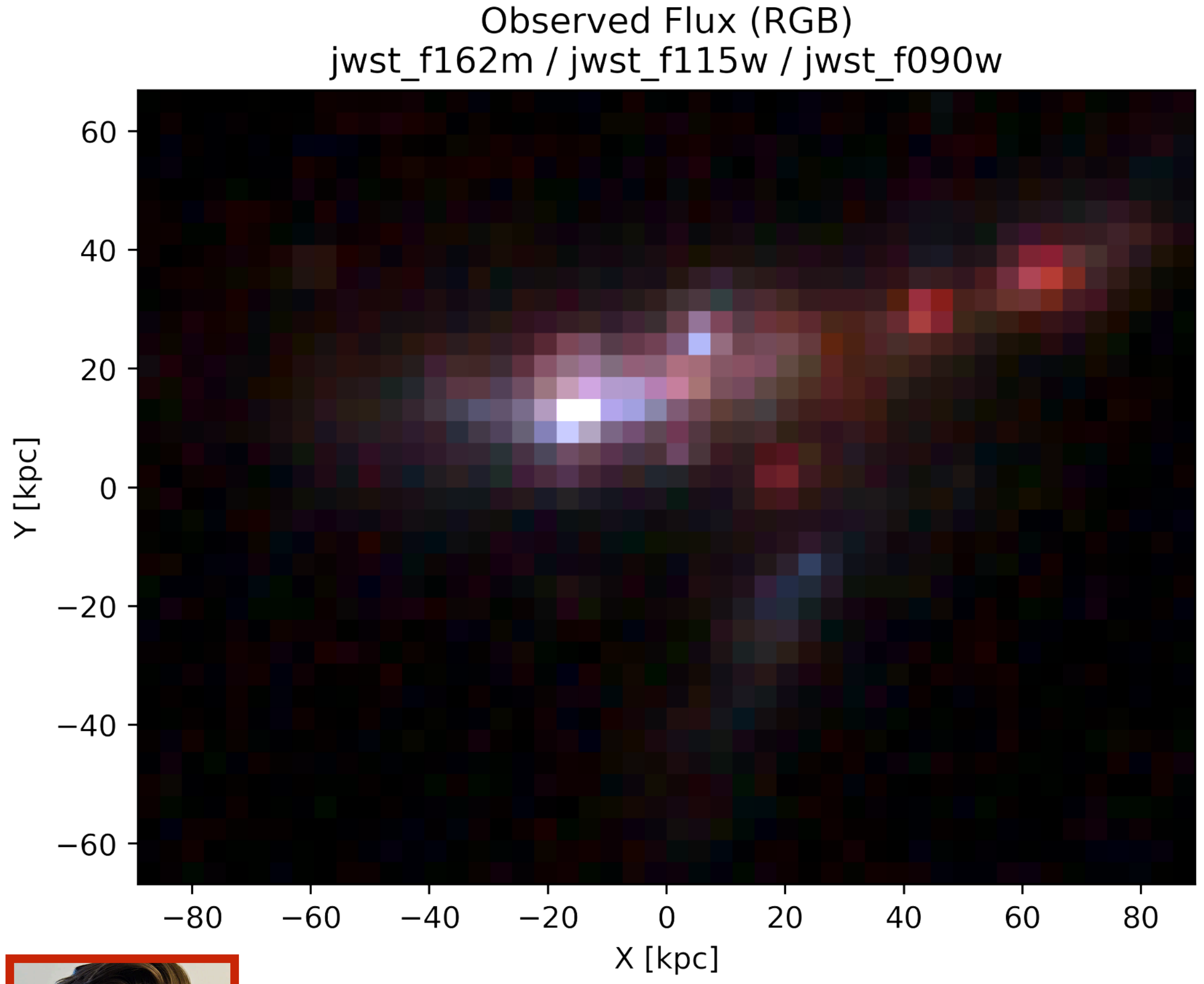
Hubble

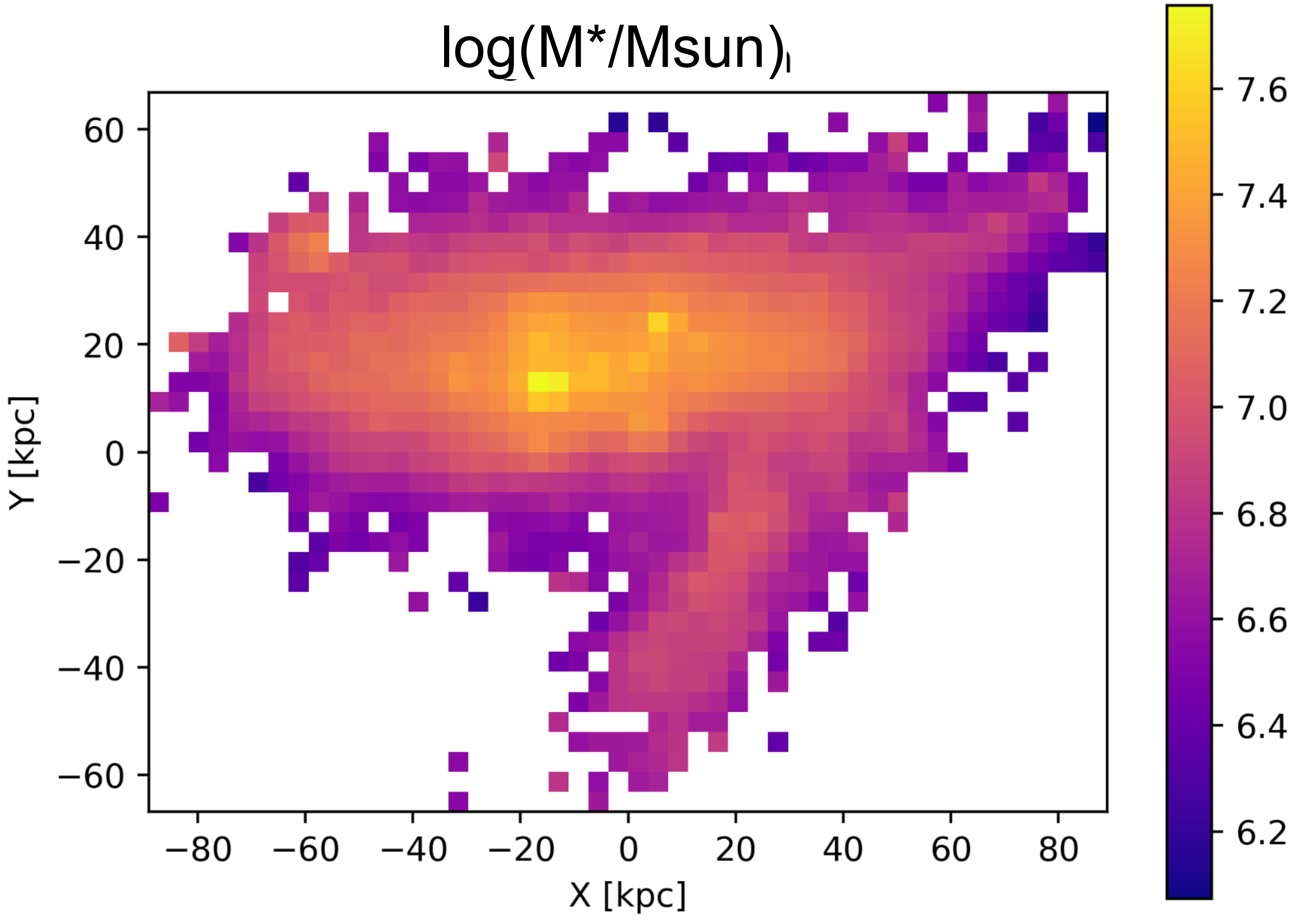


Spitzer (and Herschel, and Webb, and …)



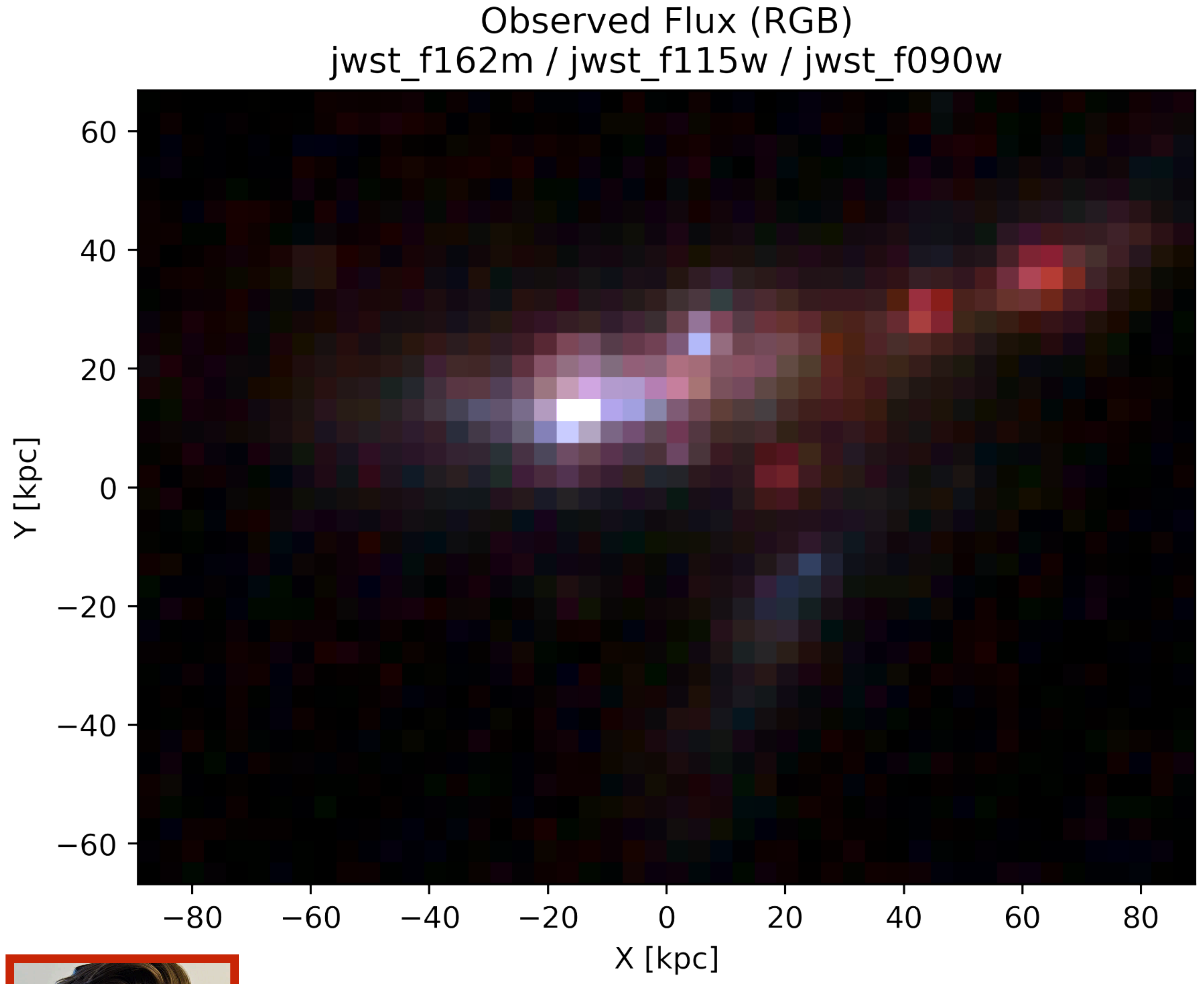
First fits in Webb deep fields reveal galaxies are **spatially complex**

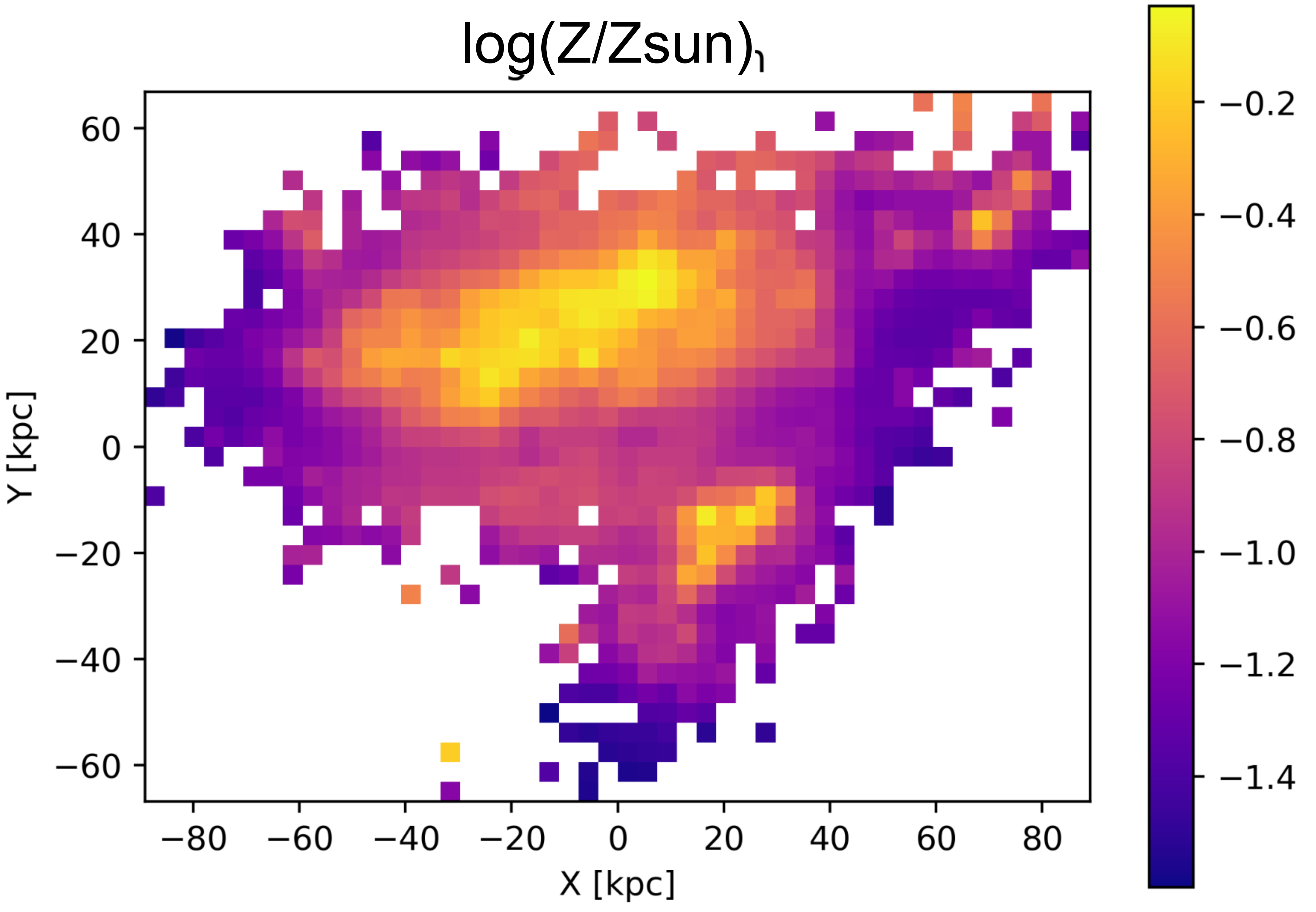




Mathews, Leja et al. in prep;
collaborators Nelson,
Speagle, Whitaker..

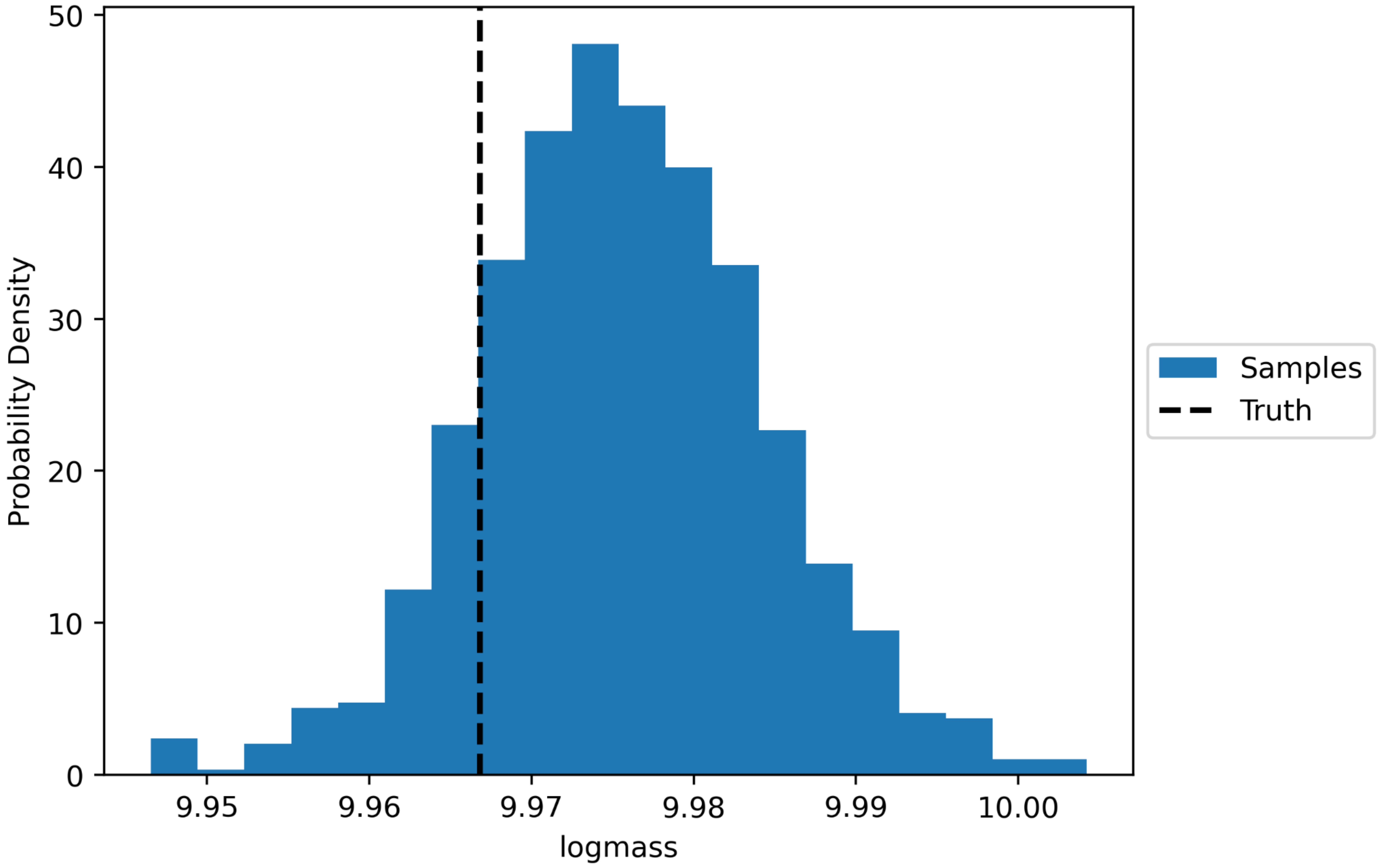
First fits in Webb deep fields reveal galaxies are **spatially complex**





Mathews, Leja et al. in prep;
collaborators Nelson,
Speagle, Whitaker..

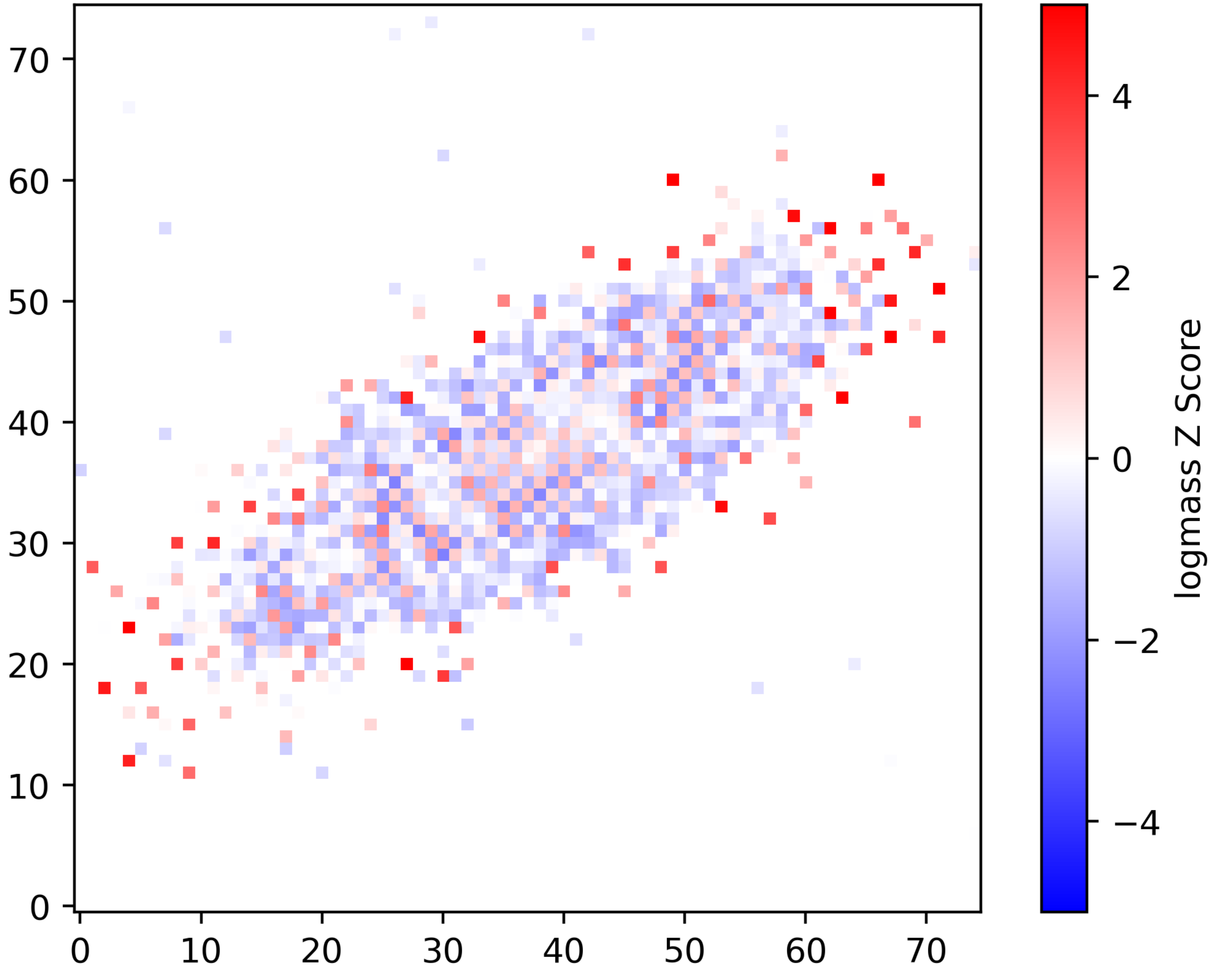
Better constraints: typical uncertainties for galaxy stellar masses in unresolved photometry are 20-50%…



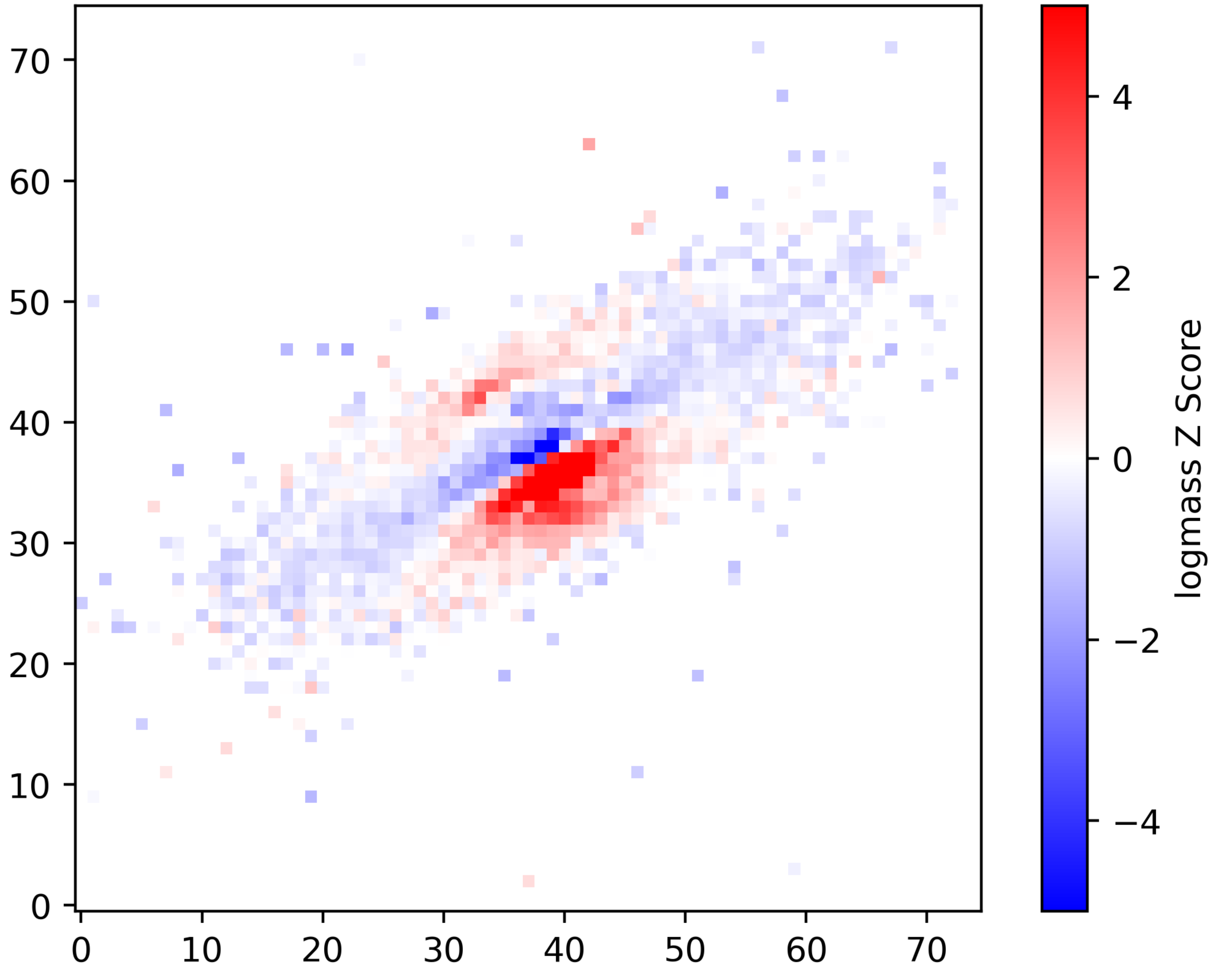
Mathews, Leja et al. in prep

Tests on spatially resolved mock galaxies suggest **well-calibrated uncertainties of ~5%!**

Simultaneous fits forward-modeling image-smearing are critical - “independent” pixel-by-pixel fits fail to give good answers



Accounting for PSF



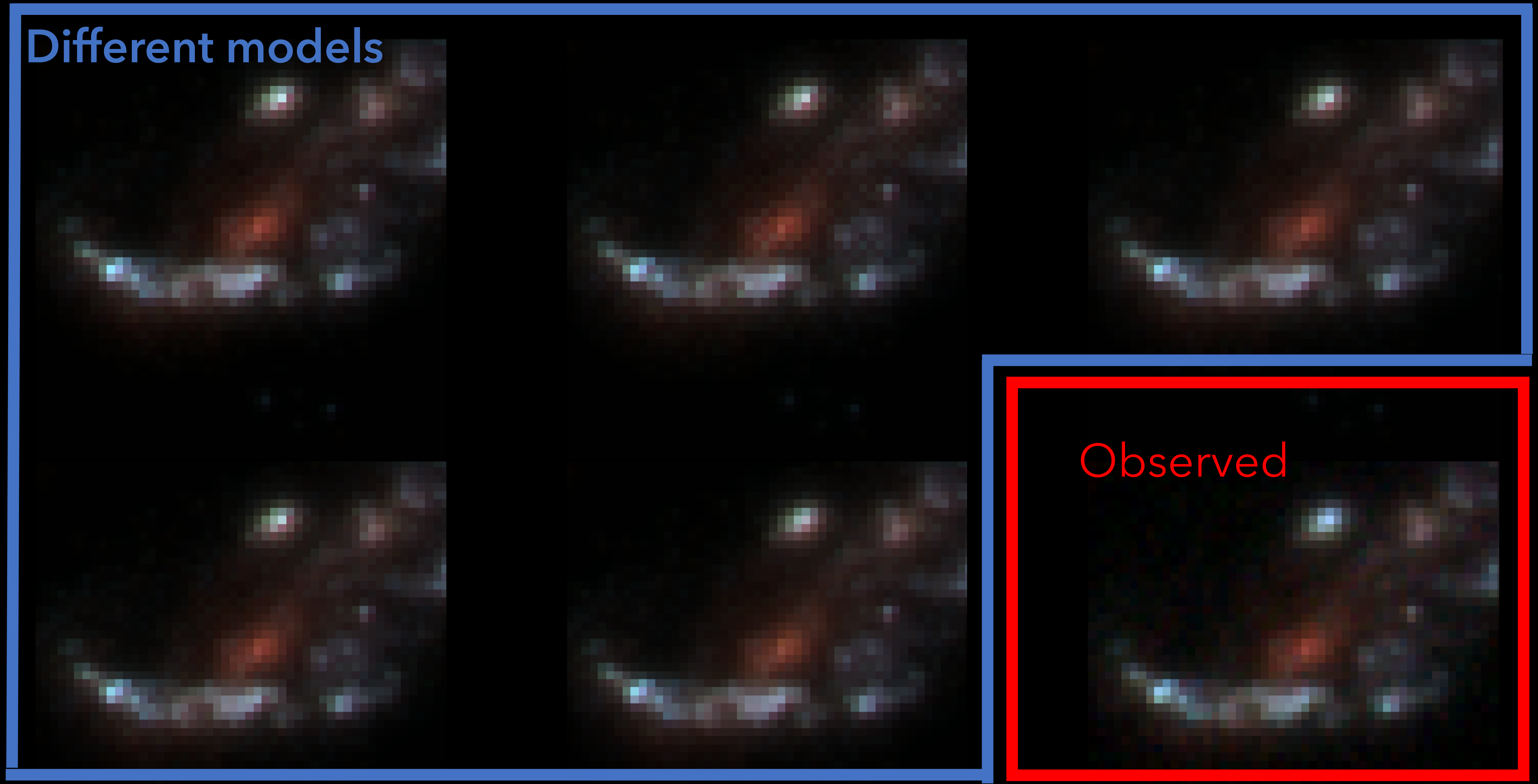
Ignoring PSF

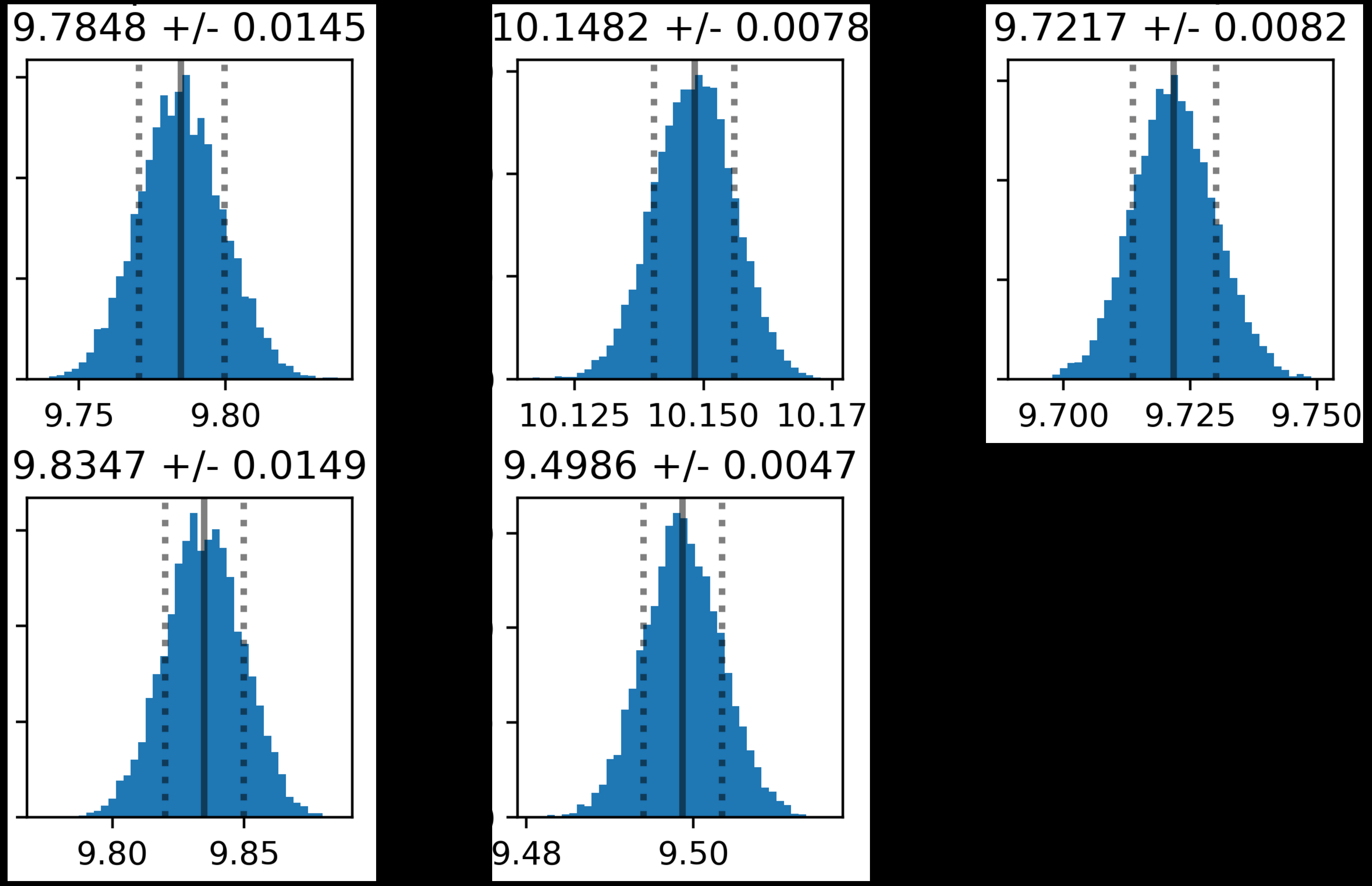
A Key Challenge Before Industrial-Scale Spatially Resolved Modeling..

More Pixels - More Problems!

Fit 20-band Webb photometry with different assumptions about small-scale variation of age/dust/metallicity - all provide beautiful match to light!

But small-scale variations very important - factor of 4-5 spread in recovered stellar masses!





Need to understand small-scale physics of dust and dynamical mixing of stellar populations of different age, heavy element composition. *Hard problem*!

Mathews, Leja et al. in prep

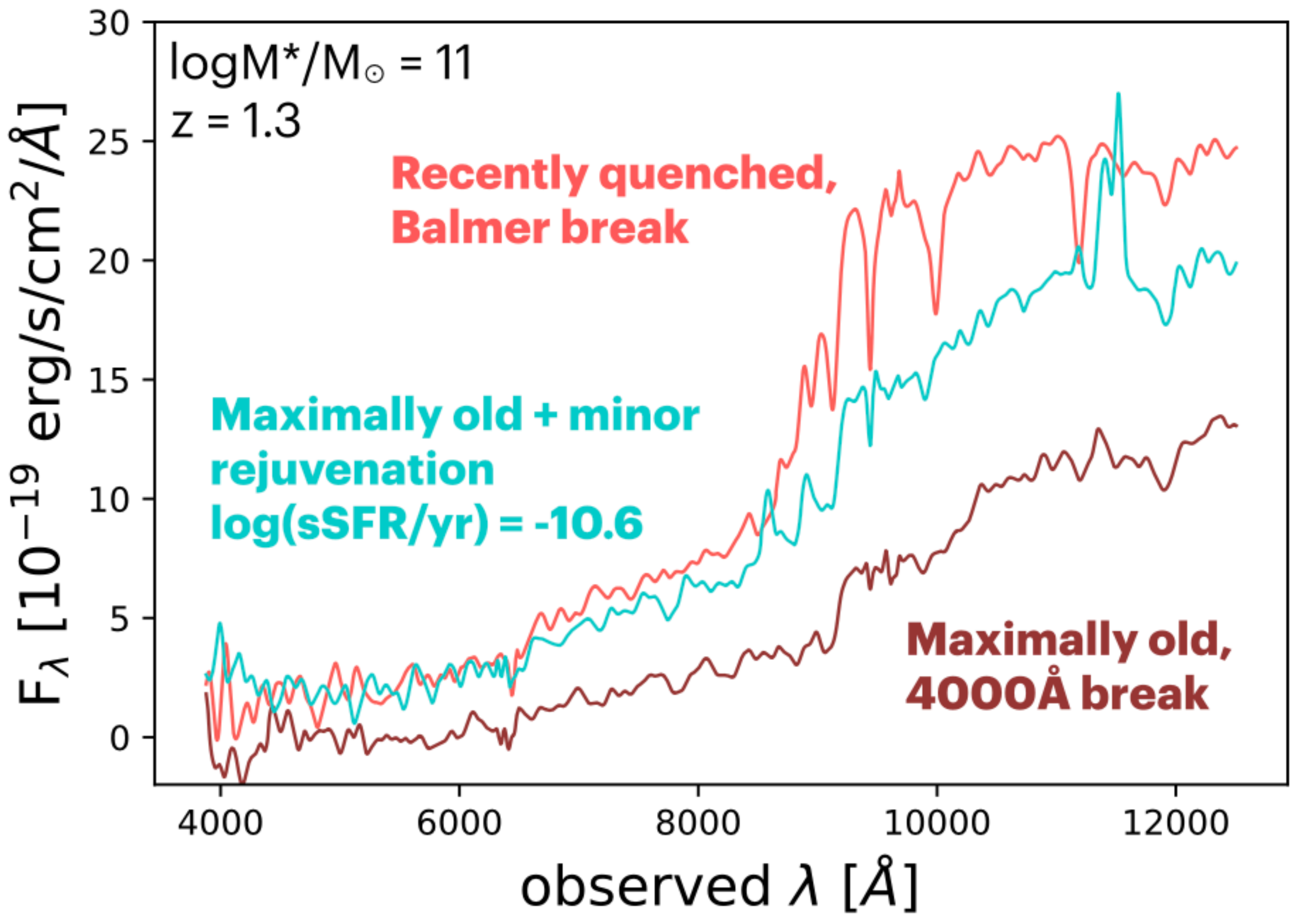
Mathews, Leja+23 (emulator)

Mathews, Leja+ in prep (GPU/spatially resolved)

What about Spectra?

Many large-scale spectral surveys of the observable universe starting, goal of understanding dark energy: DESI, PFS, Euclid, MOONS, … tens of millions of galaxies!

PFS 12-hour spectra



Advantages over imaging emulators

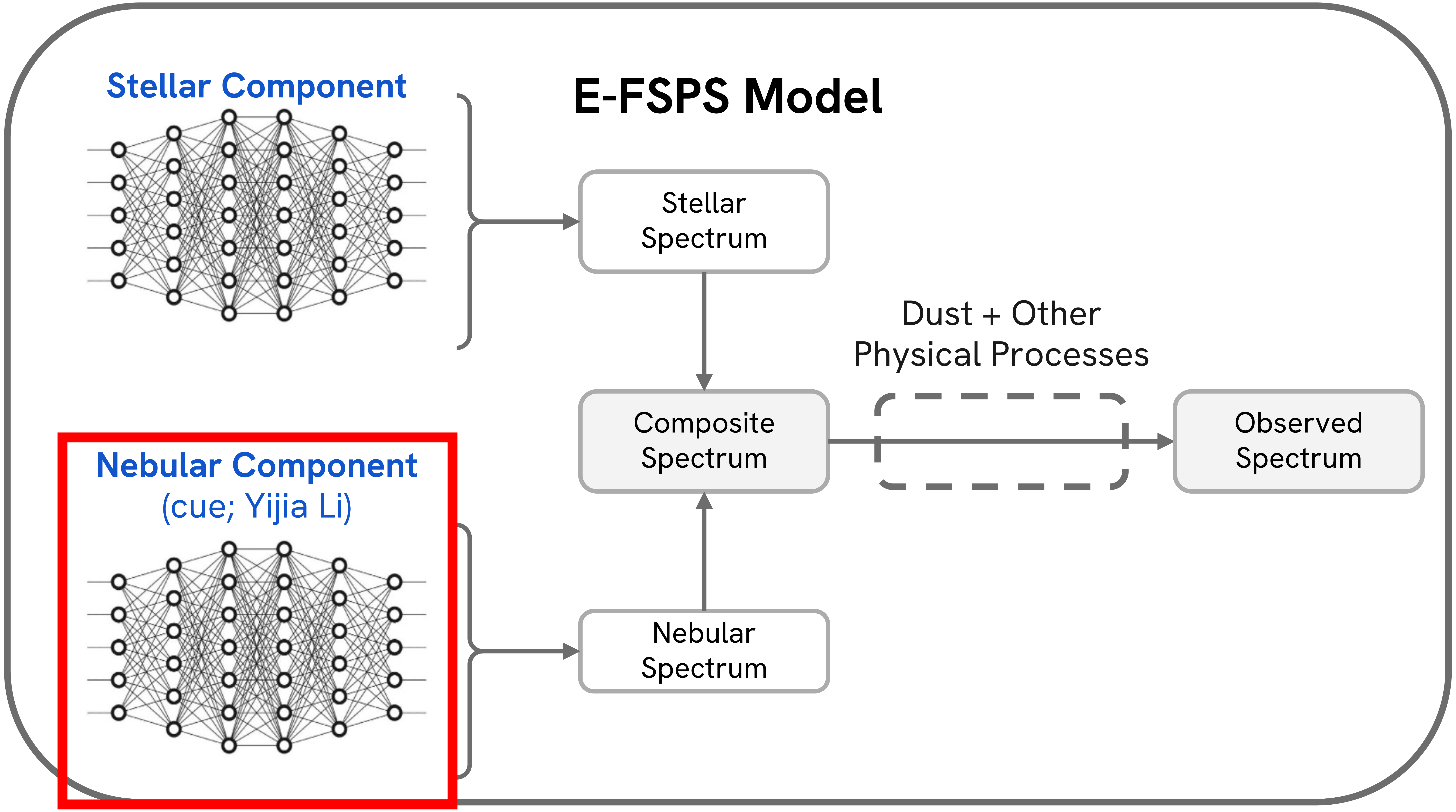
- Can emulate in rest-frame (redshift is biggest challenge!)
- Spectral pixels highly correlated, smooth

Challenges compared to imaging emulator

- Far more data (~10 filters —> 1000s of pixels)
- More detailed input physics (e.g. complex line emission physics)

Piecewise Emulation For Spectra (Not Single-Shot!)

Burnham et al. in prep

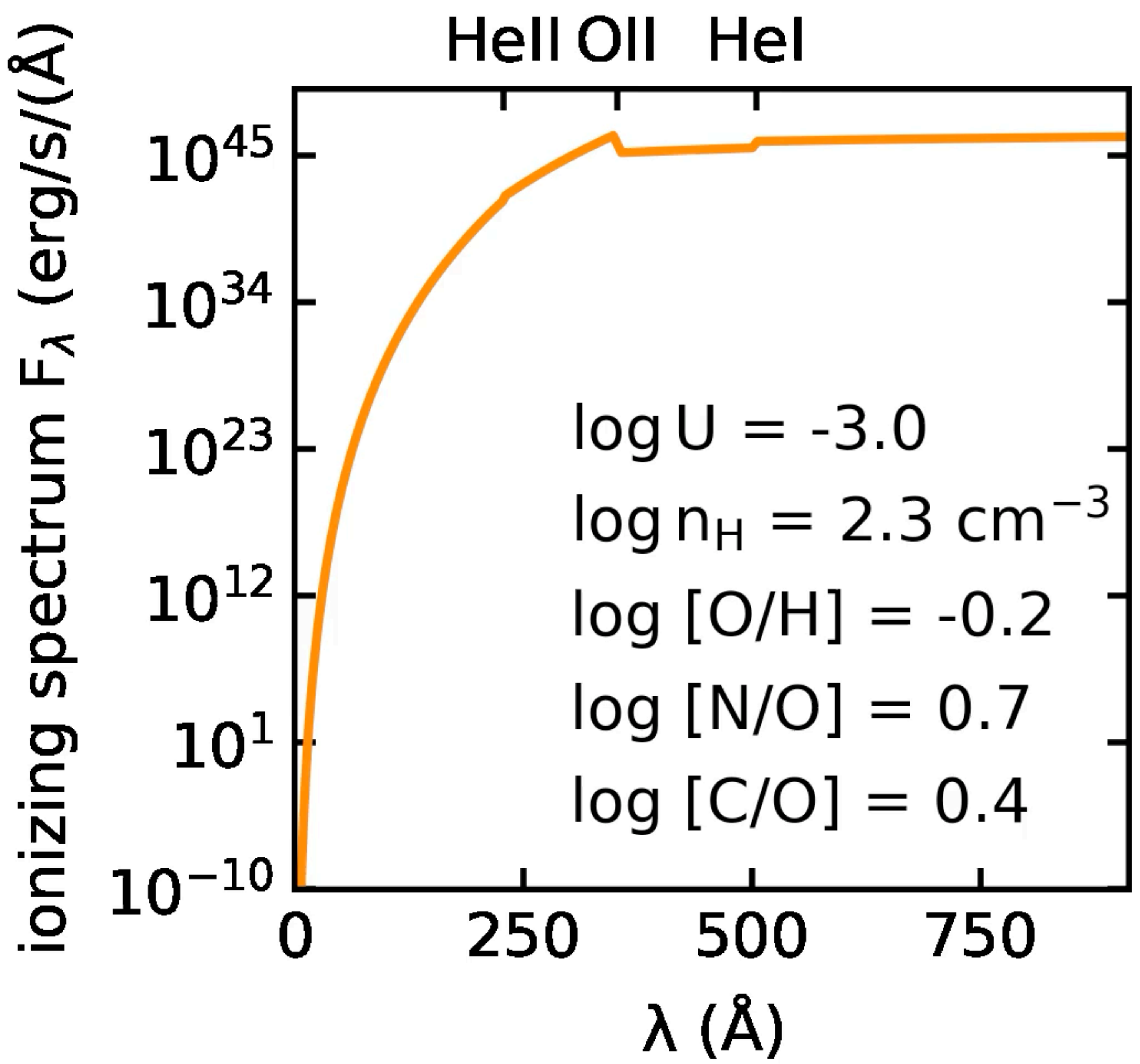


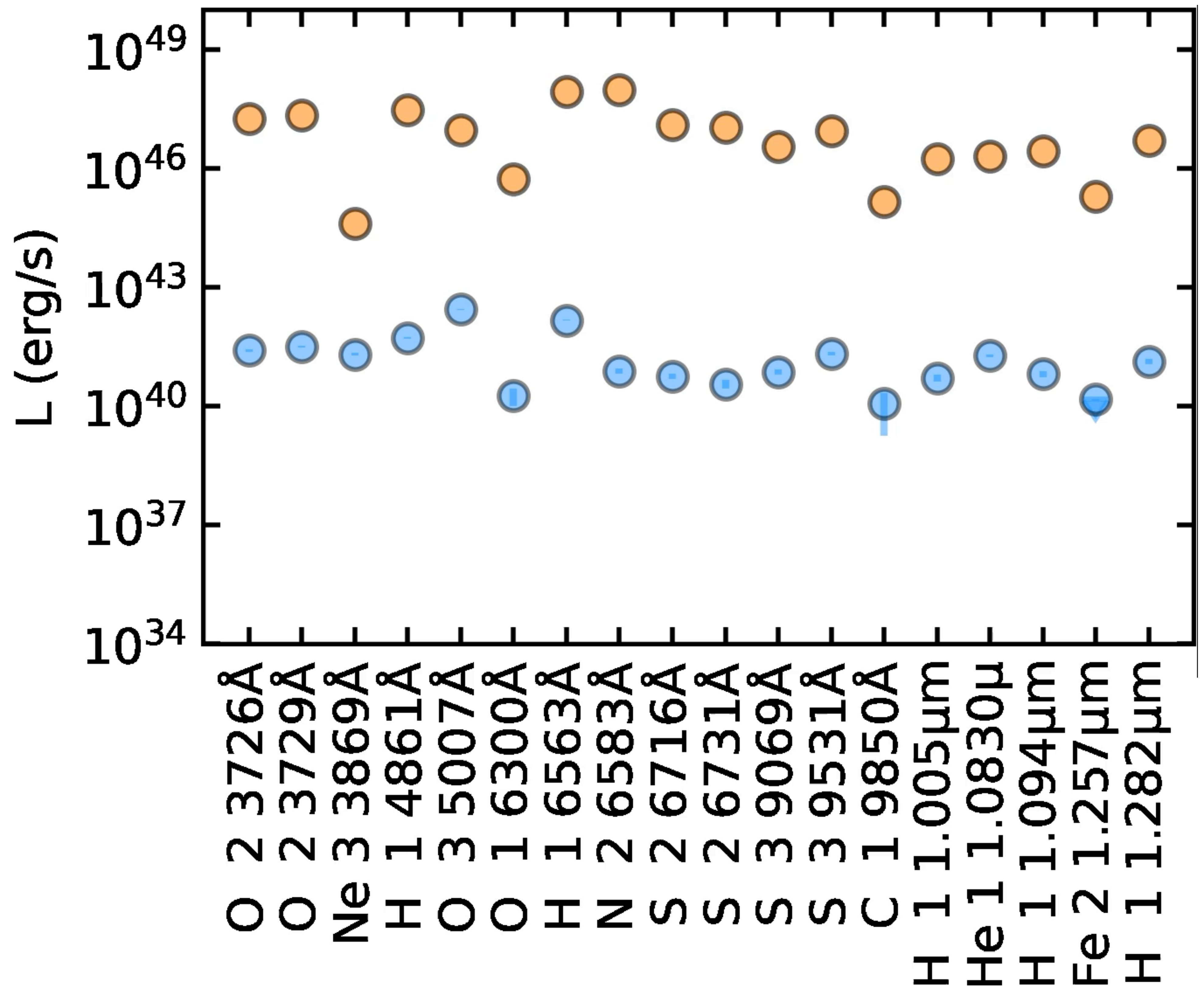
Emulated FSPS (E-FSPS) replaces gigabytes of stellar model grids with **deep neural networks**.

Cue: a fast and flexible neural net emulator for nebulae

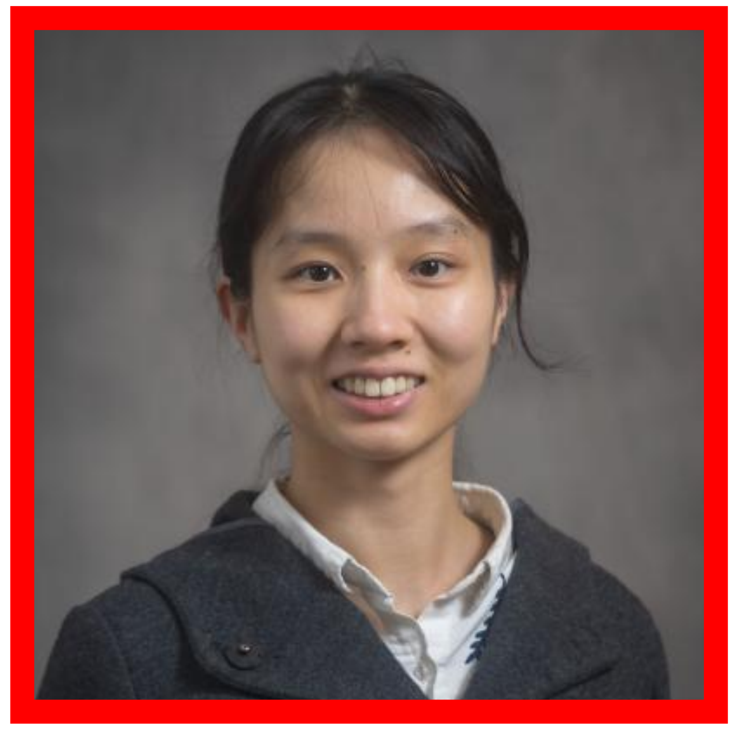
Inferred ionizing spectrum + nebular physics

Observed data





Li, Leja+24ab

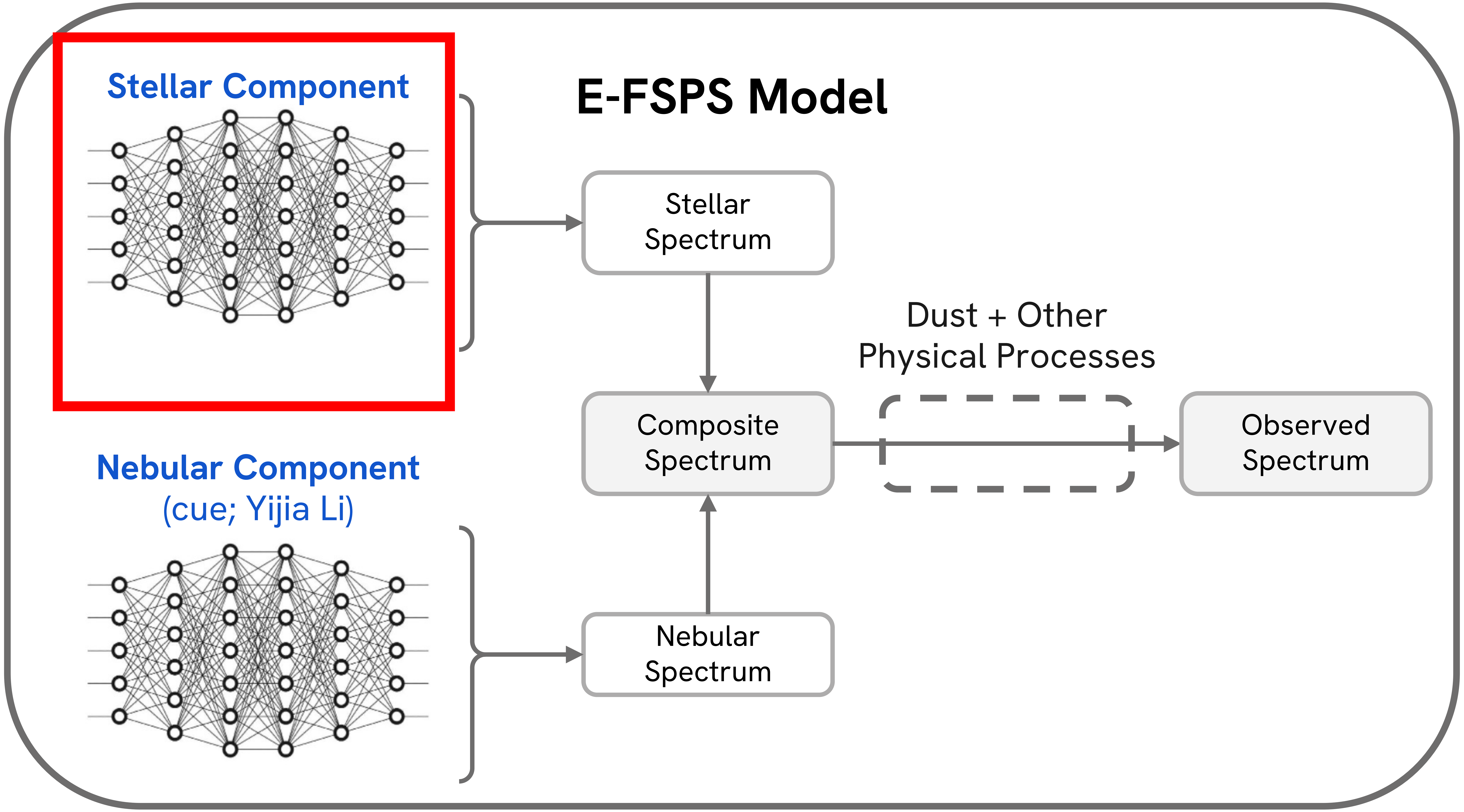


https://github.com/yi-jia-li/cue

- 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)

Piecewise Emulation For Spectra (Not Single-Shot!)

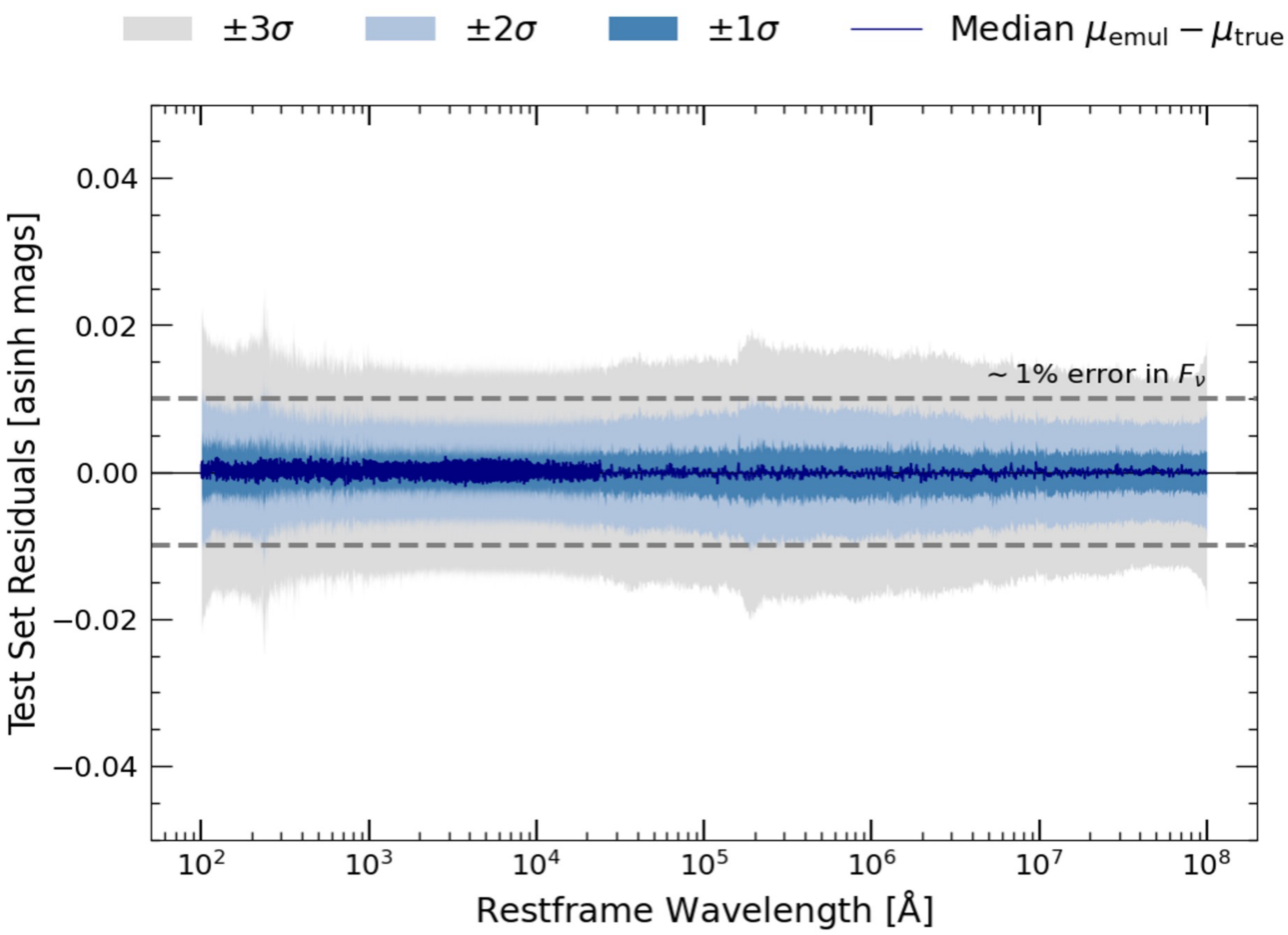
Burnham et al. in prep



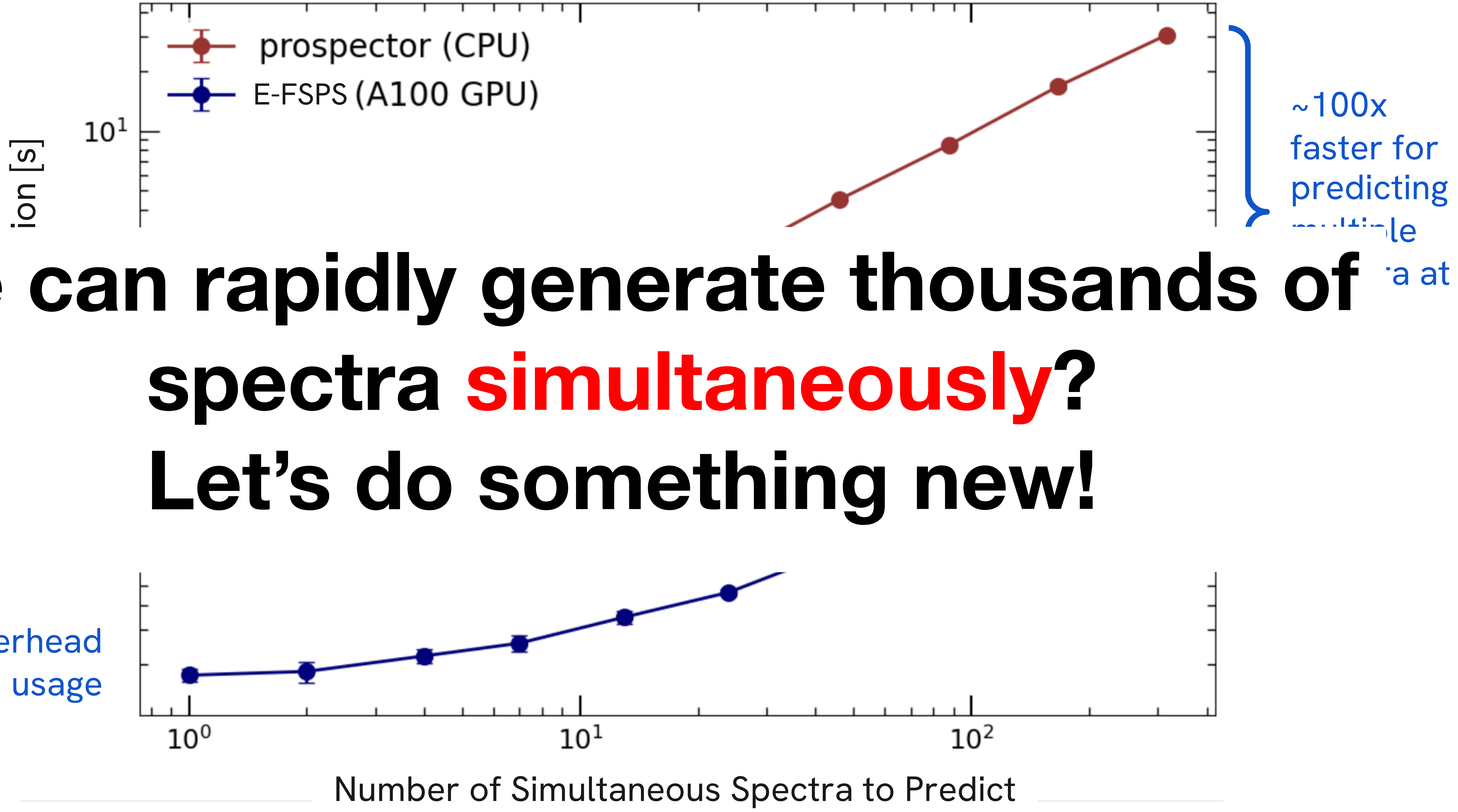
Emulated FSPS (E-FSPS) replaces gigabytes of stellar model grids with **deep neural networks**.

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!

+Additional speedups from gradient-enhanced sampling (TBD)

Annoyingly, Bursty Star Formation Is Nearly Impossible To Model In Single Objects

Recently-formed bright stars stars obscure nearly all history older than 100 Myr in 'summed' galaxy imaging.

A model-generated, high S/N Webb/

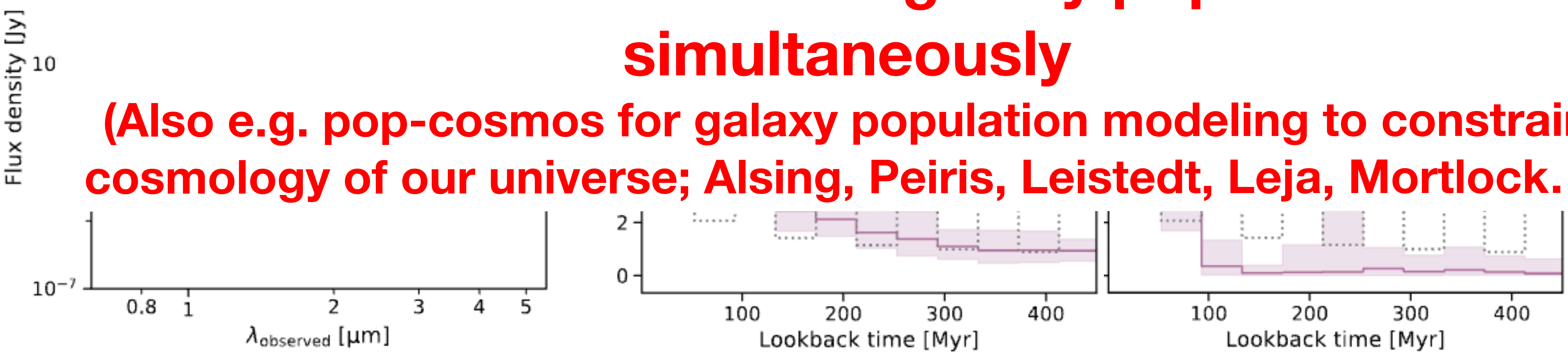
A "continuity" prior recovers *a*

A "bursty" prior recovers *only*

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…)



Wang, Leja et al., ApJ under review

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*.

The Next Step: Likelihood-Free Inference

With neural net emulators, we can generate **~2000 model spectra per second** (can you feel the wind in your hair??).

Yet not fast enough; need ~50k-100k model galaxies to propose a single mock galaxy population; need to generate 10^5-10^6 populations for a fit; 10k core-hours per fit?

New Workflow

galaxy observations

Simulation-based Inference using Normalizing Flows (~1-10s)

solutions

Simulation-Based Inference

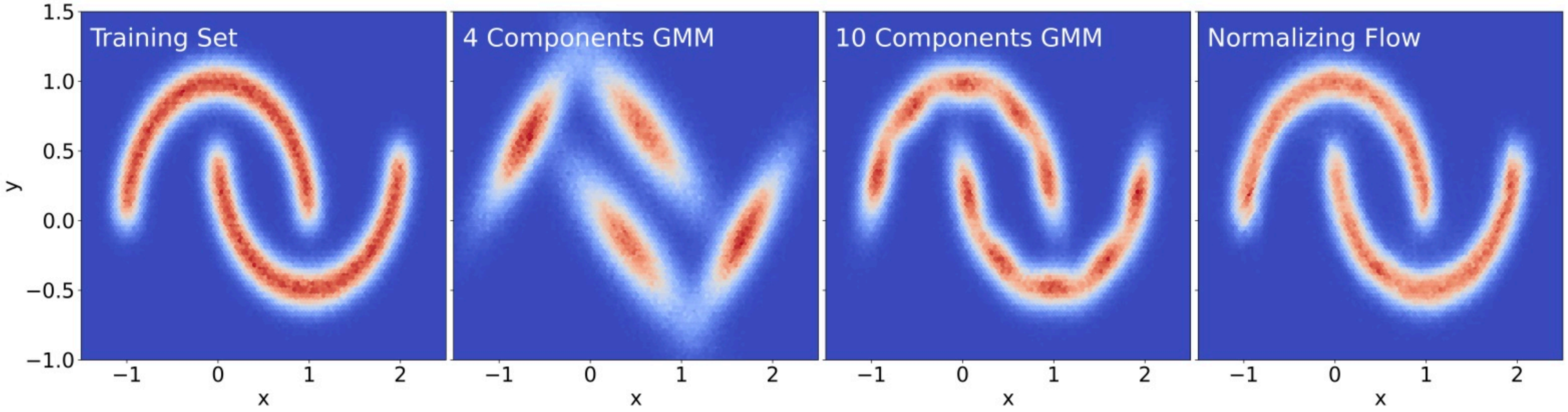
Wang, Leja et al., NeuRIPS 2022

Simulate your data, *plus noise,* many times, and **learn the direct transformation from noisy data to Bayesian posteriors.**

Use a "normalizing flow", an ML technique that **learns the transformation** from an N-dimensional Gaussian to an **arbitrary N-dimensional PDF**

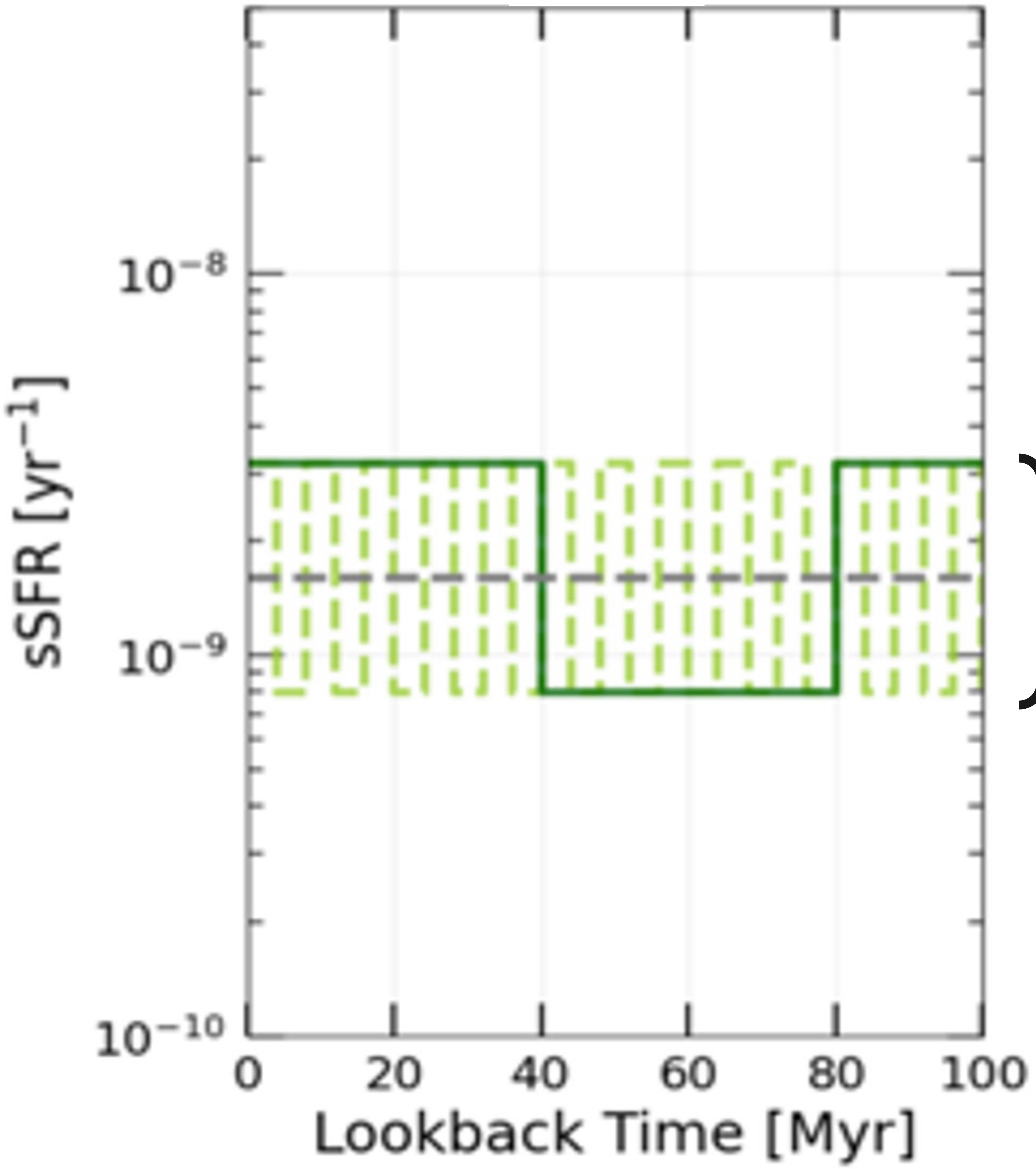
We use **SBI++** (Wang, Leja+22), which lets us Monte Carlo over missing data and uncertainties outside training set (crucial for astronomical data)

input: observed data
output: joint posteriors for your model



Ting & Weinberg 2021

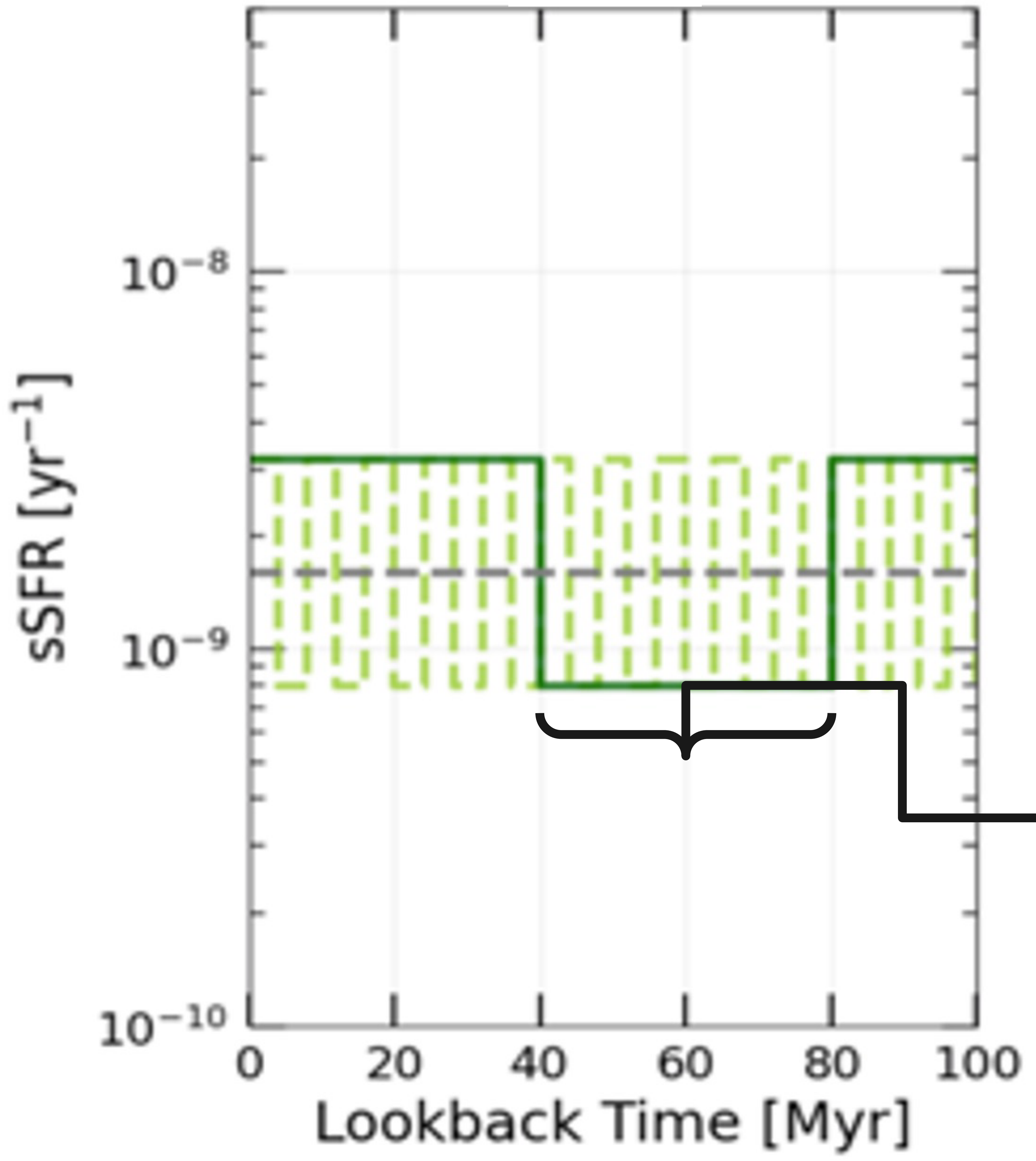
Generate Populations of Bursty Formation Histories



σ: Amplitude of up-and-down fluctuations

Burnham et al. in prep

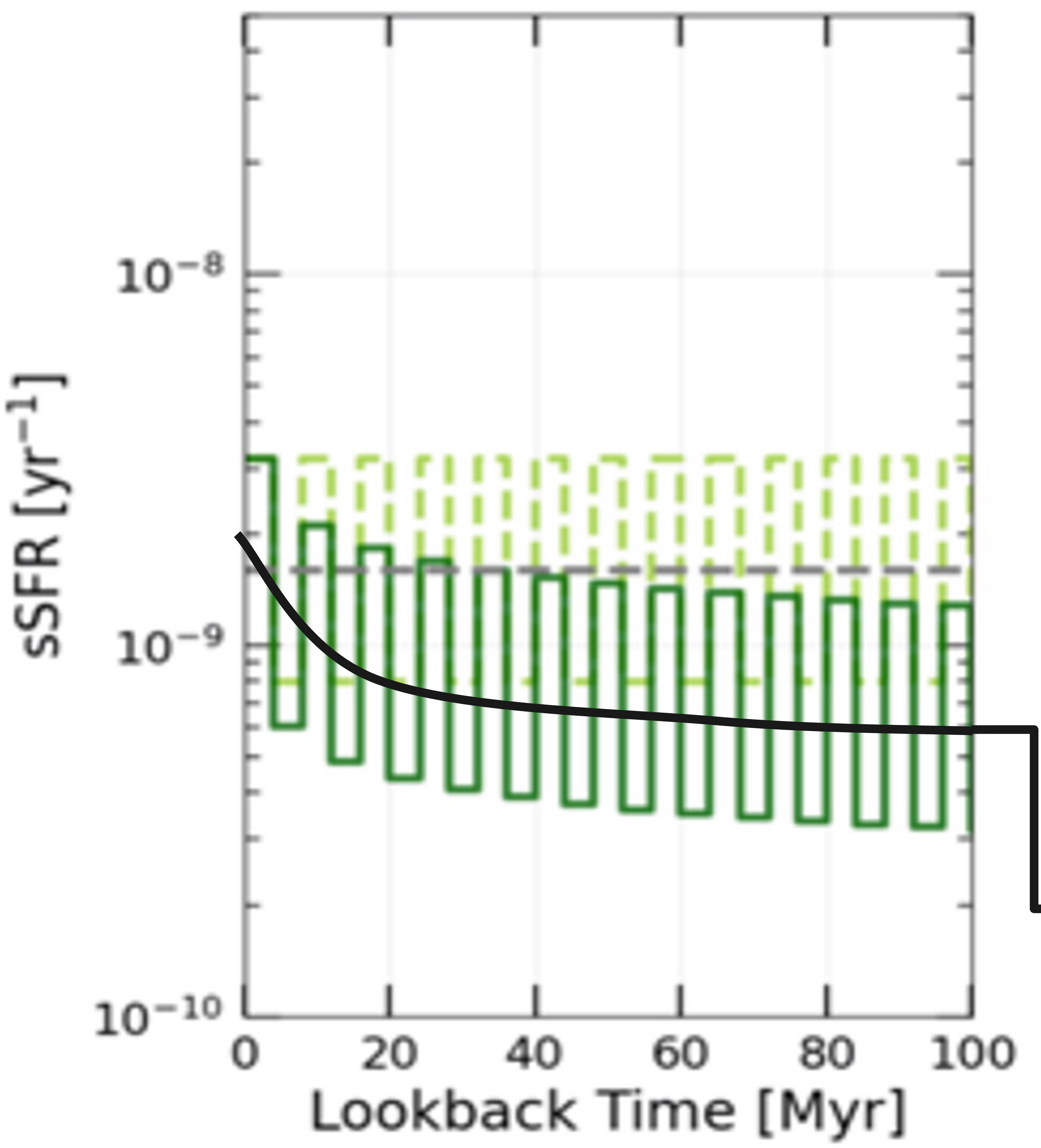
Generate Populations of Bursty Formation Histories



δt: Timescale for up-and-down fluctuations

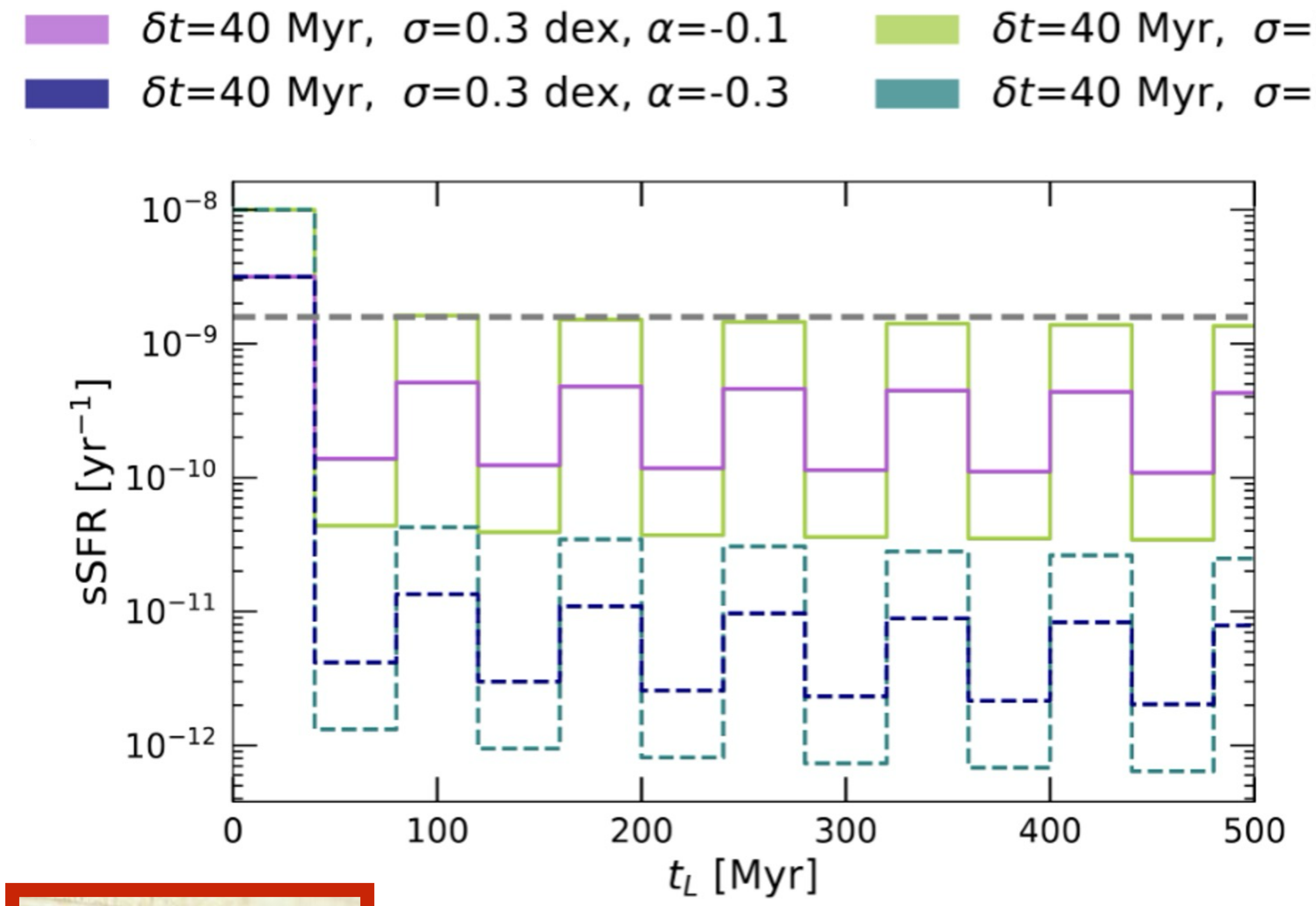
Burnham et al. in prep

Generate Populations of Bursty Formation Histories



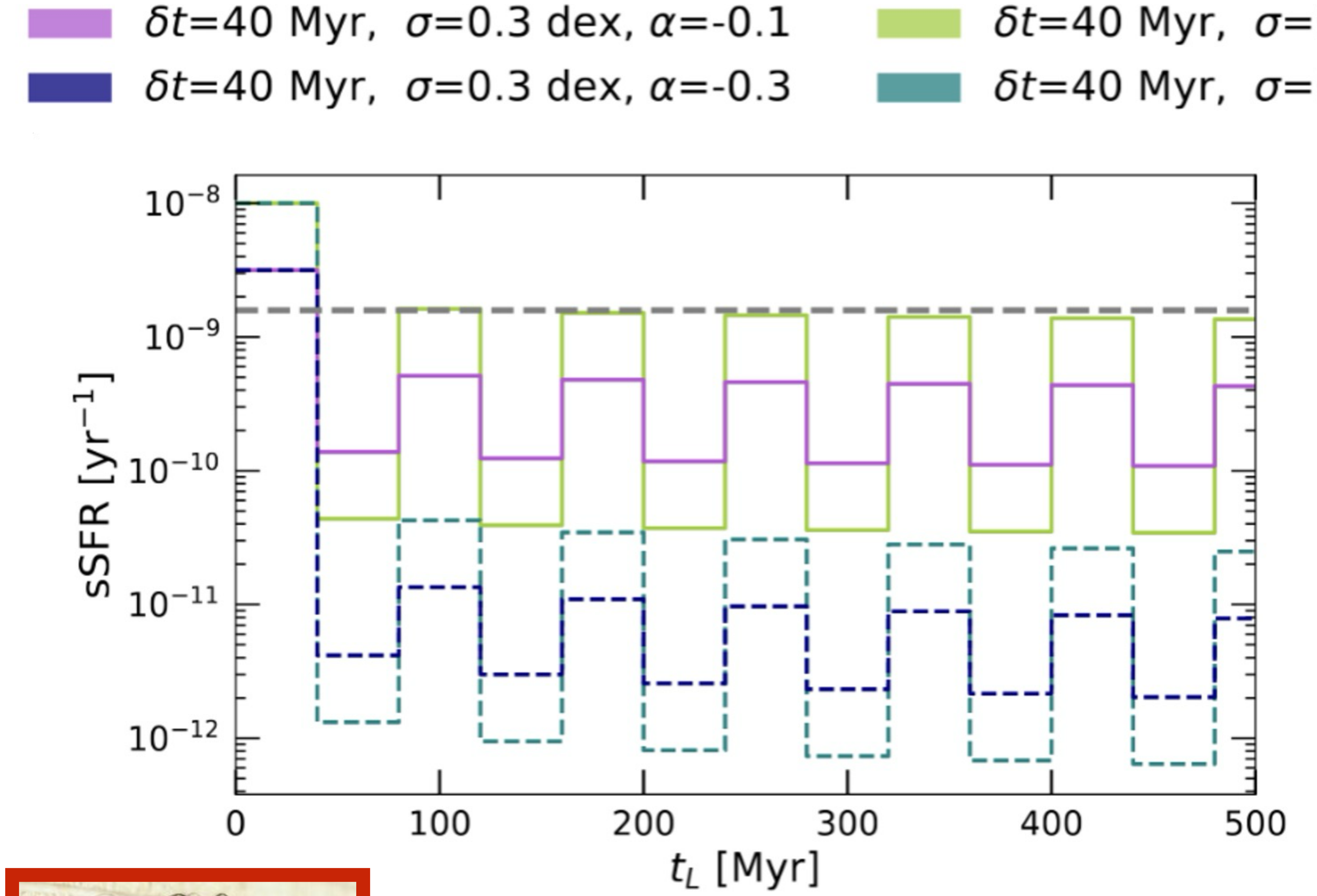
α: Slope
of recent star formation history (rising, falling, etc)

Burnham et al. in prep

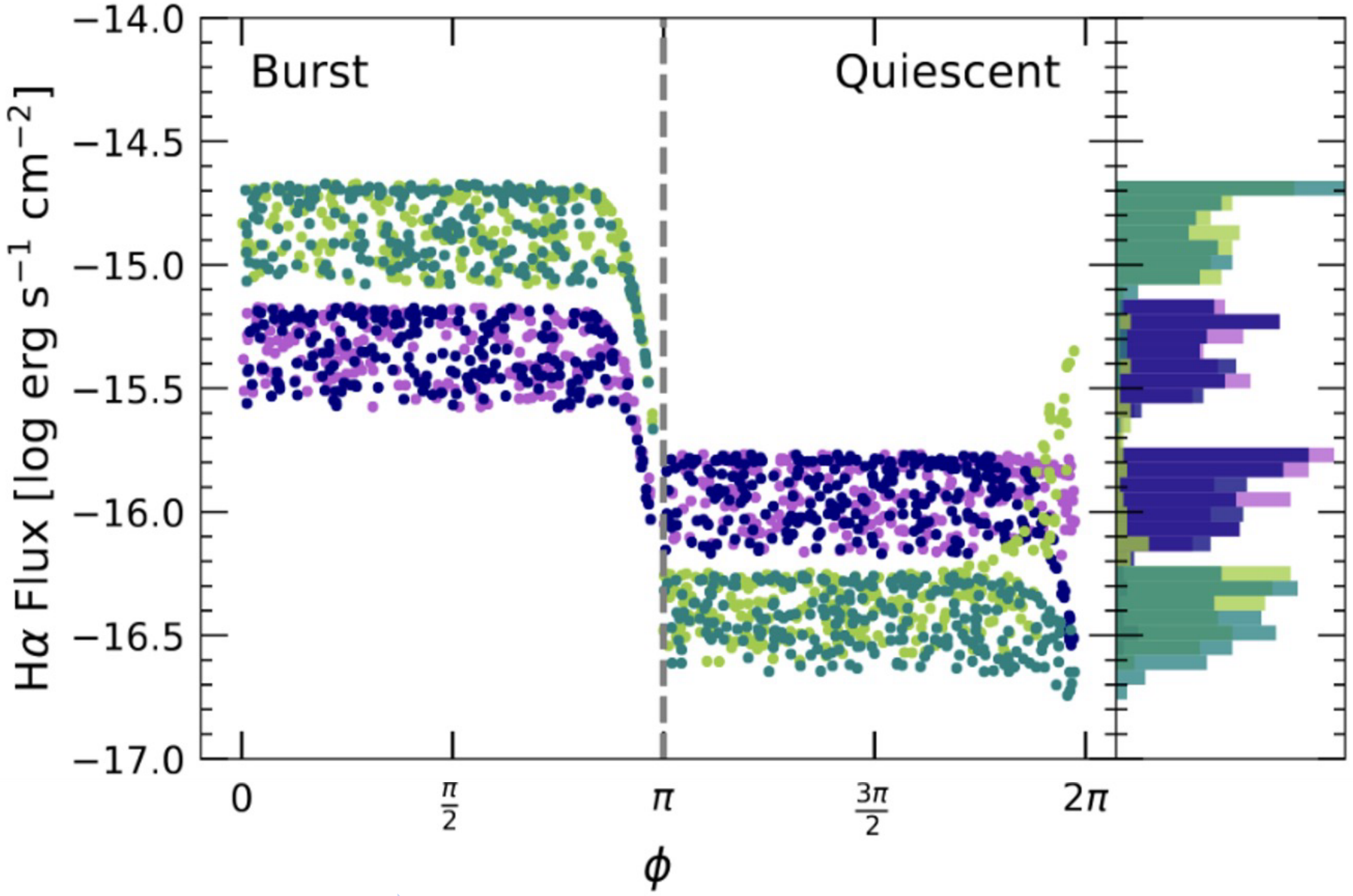


Burnham et al. in prep

Sample many galaxies with different dust, heavy element composition, etc from bursty population



(Sampling over SFH phase)

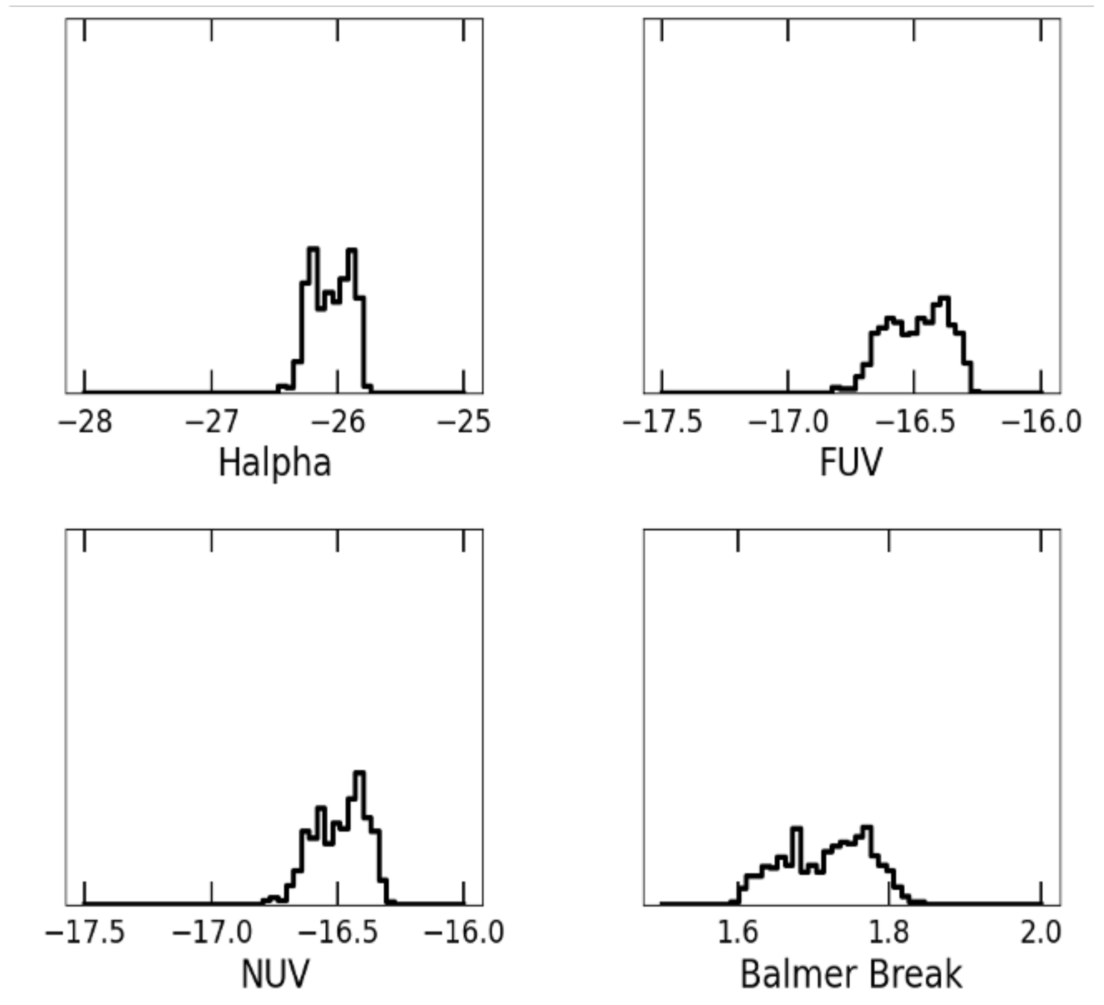


Burnham et al. in prep

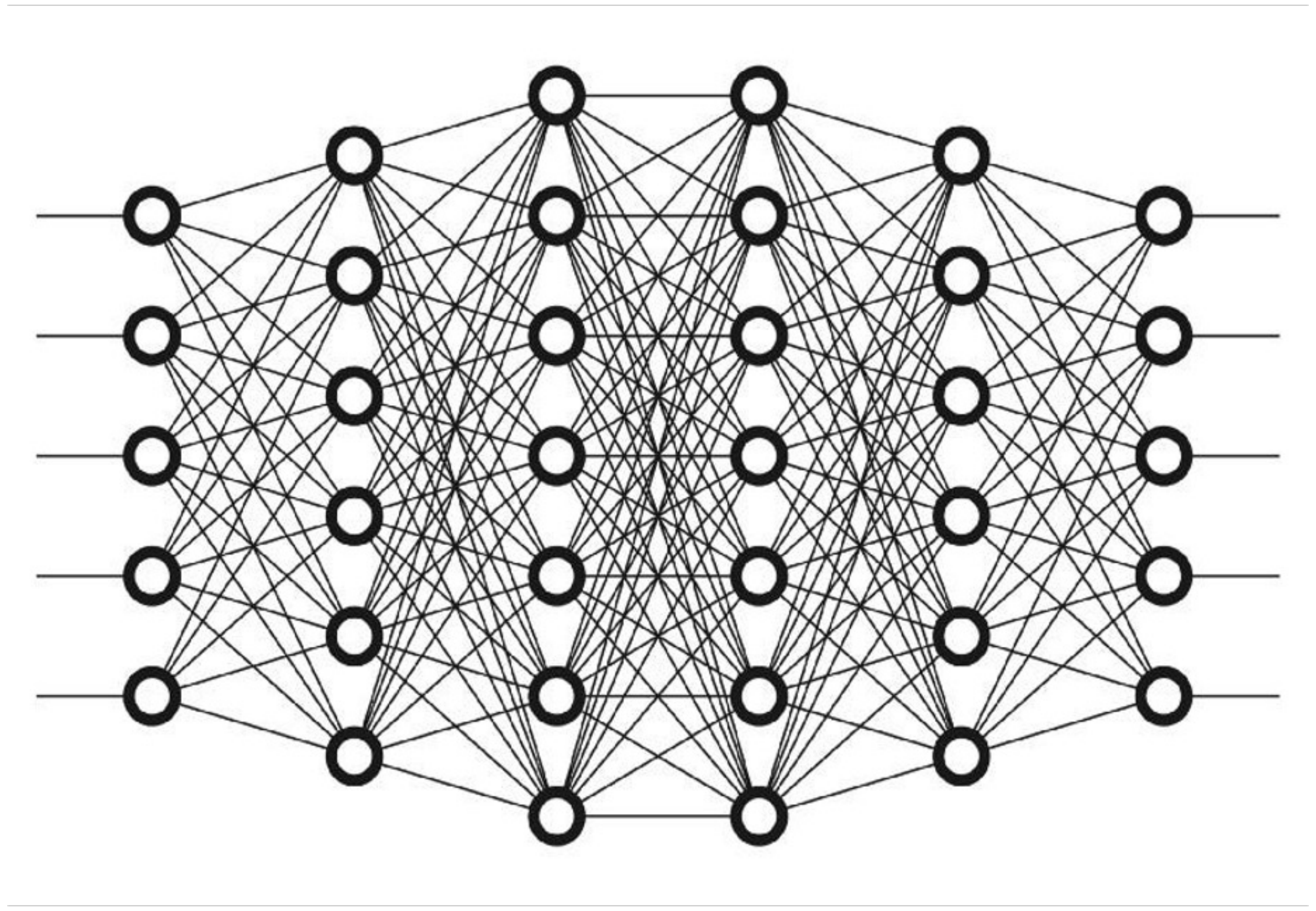
Different burstiness models produce different observed populations. Need hundreds of millions of model galaxies as a training set – now “easy”.

Simulation-Based Inference (SBI) For Galaxy Population Modeling

Observed Distributions

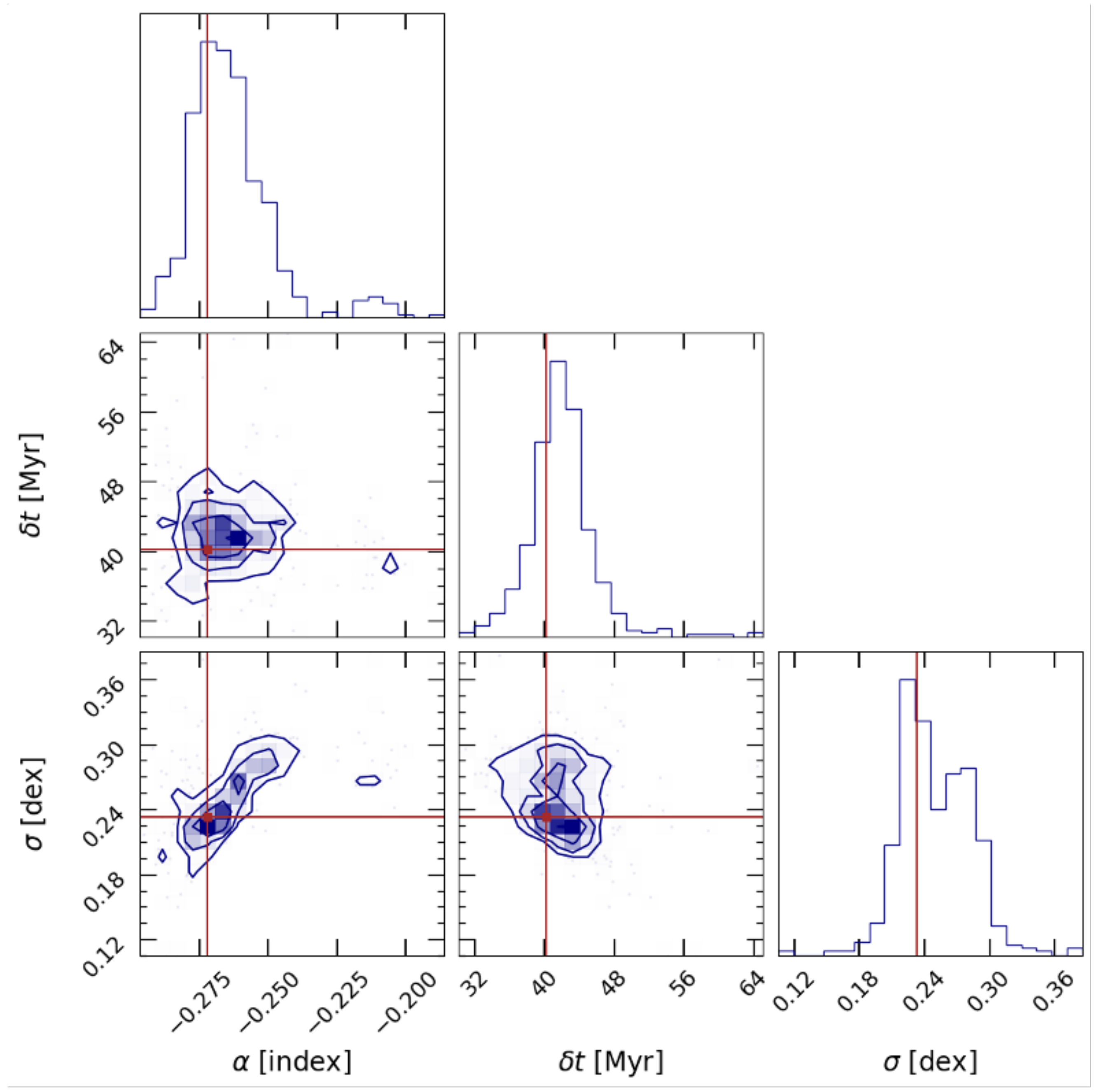


Simulation-Based Inference



Machine-learning inference framework pre-trained on simulated distributions

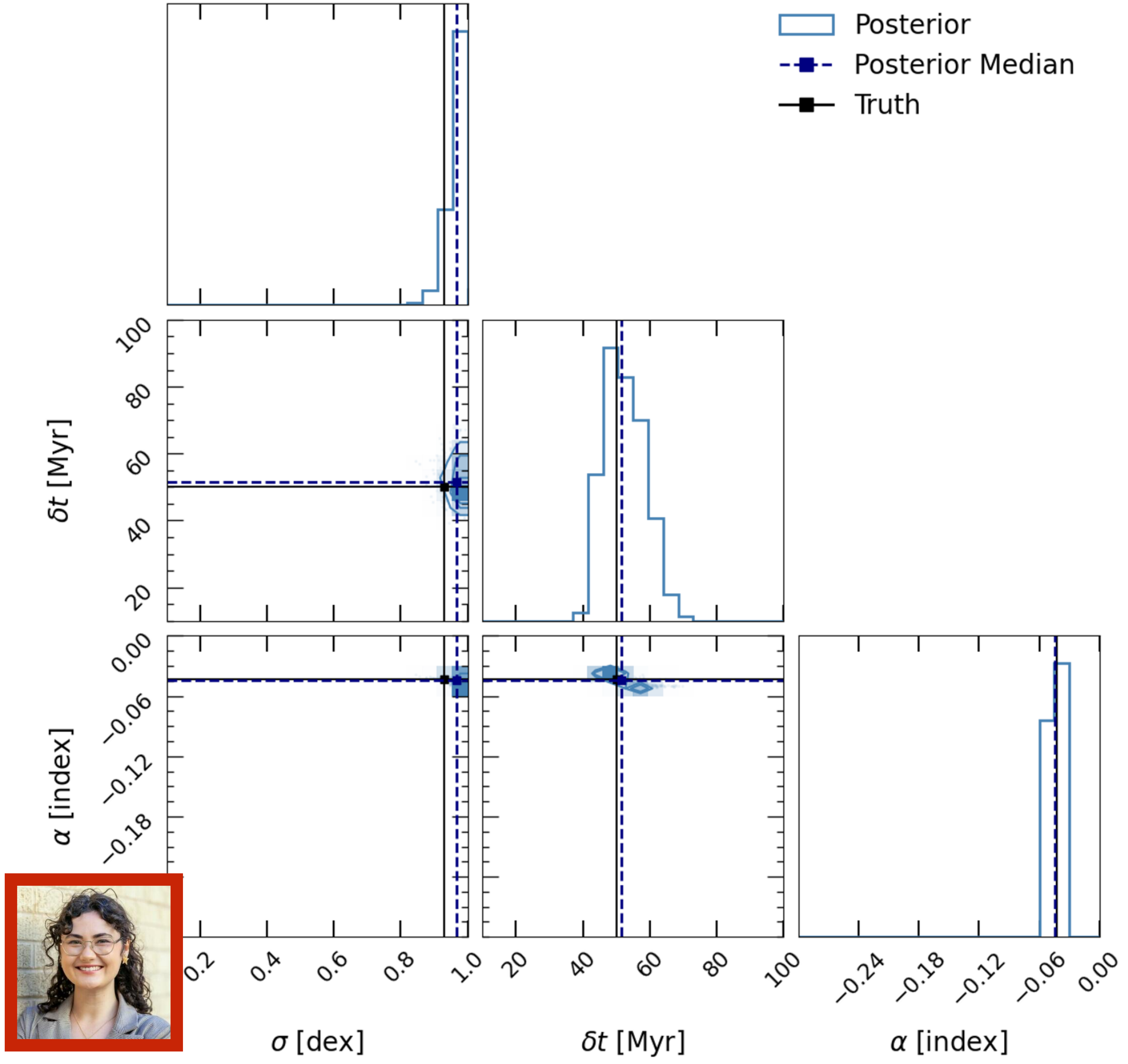
Population-level Parameters



SBI effectively learns the population posterior given an observation.

Burnham et al. in prep

Can We Recover the Simple Model Parameters?



Burnham et al. in prep

We can **accurately recover** timescales for realistic populations (N~500 galaxies)

slope: 3-5%
timescale: ~10%
amplitude: 1-2%

Can we 'solve' outshining by learning the right population prior, and applying to individual objects? **Unsolvable** with classic techniques -- now **feasible** with ML-enhanced approach.

Next up

- test epistemic uncertainties (e.g. stellar evolution models)
- use realistic formation histories
- apply to deep JWST populations!

A Quiet Computational Revolution

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

2017: 100 core-hours

Code optimization

2019: 20 core-hours

Neural net emulators (Alsing+20, Mathews+23, Kwon+23)

2021: 10 core-minutes

GPU acceleration + differentiability (Hearin+23, Alsing+23, Li+24)

2024: 1 core-minute

& simulation-based inference yields <10 core-seconds per object; (e.g. Hahn+21, Wang+22, Khullar+22…)

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 Galaxy Modeling Science with New Data-Intensive Techniques

- 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
- **NEW**: spatially complex modeling including light-smearing; hundreds of pixels, tens of thousands of parameters, more accurate galaxy inference (pirate, Mathews et al. 2023, +in prep)
- **NEW**: rapid interrogation of unknown or mixed ionizing sources (cue, Li et al. 2024ab)
- **NEW**: Bayesian population modeling with SBI; infer star formation rate fluctuations, mitigate old problems of outshining and unknown formation histories (E-FSPS, Burnham et al. in prep)