A Strong Lens Is Worth a Thousand Dark Matter Halos: Inference on Small-Scale Structure Using Sequential Methods

EuCAIFCon 2024

Sebastian Wagner-Carena, Jaehoon Lee, Jeffrey Pennington, Jelle Aalbers, Simon Birrer, Risa H. Wechsler

arXiv: [2404.14487](2404.14487)

github: paltax

FLATIRON INSTITUTE

NYU

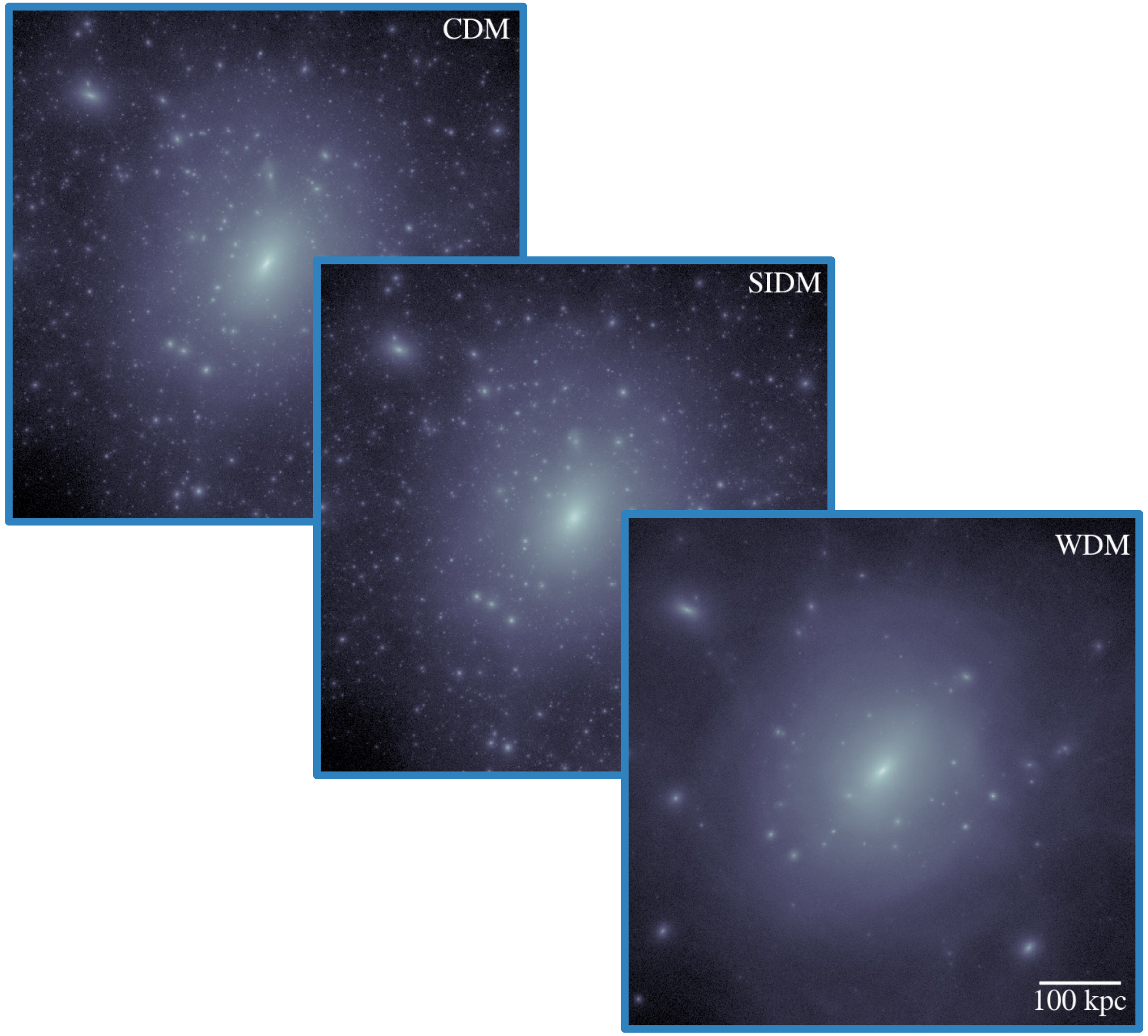
CDM Open Question

Bullock & Boylan-Kolchin, 2017

- CDM is a description of the **behavior** of dark matter, not a fundamental model

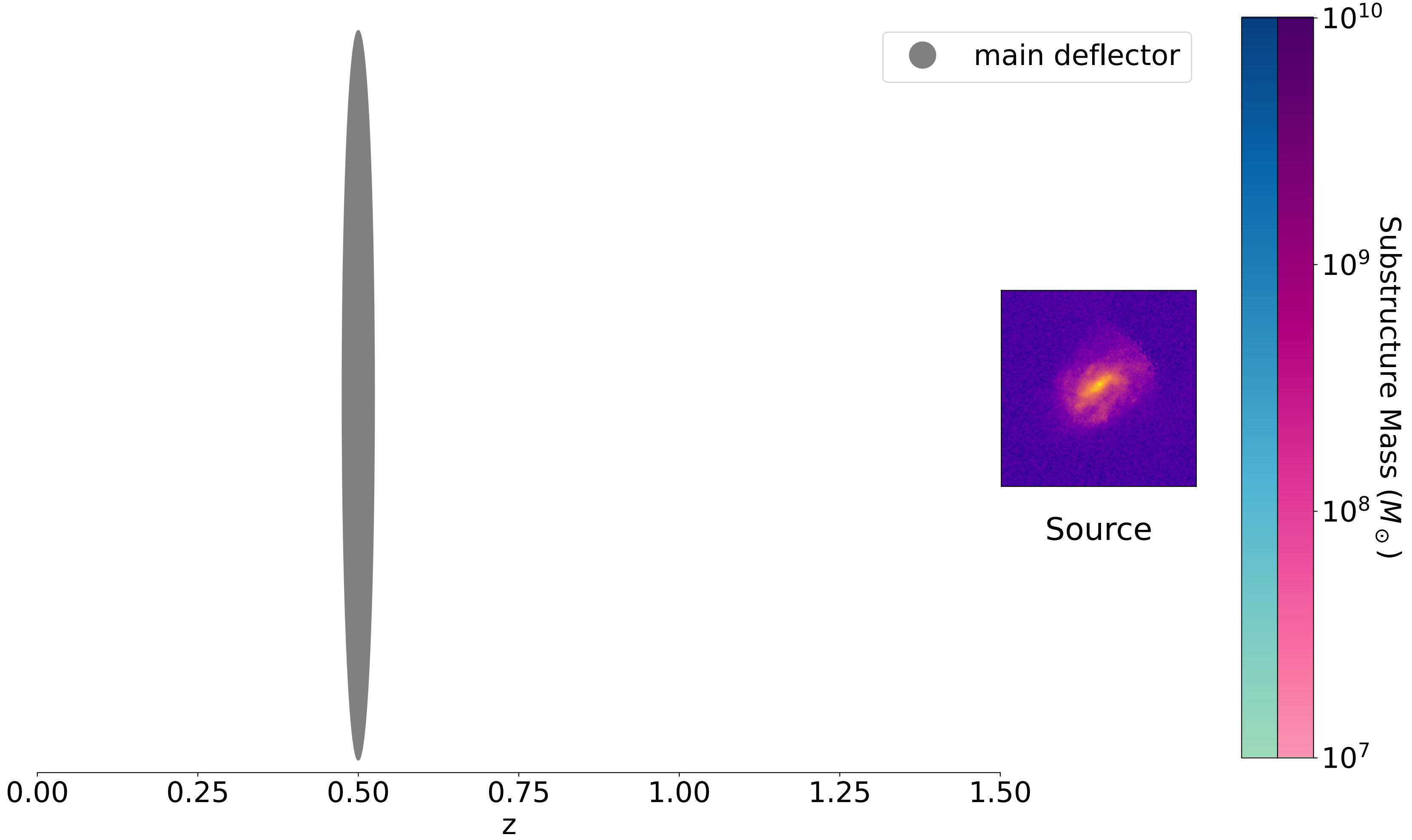
What is the fundamental nature of dark matter?

- Many compelling theories for dark matter **violate** the CDM paradigm
- **Low-mass** halos → **dark matter physics**



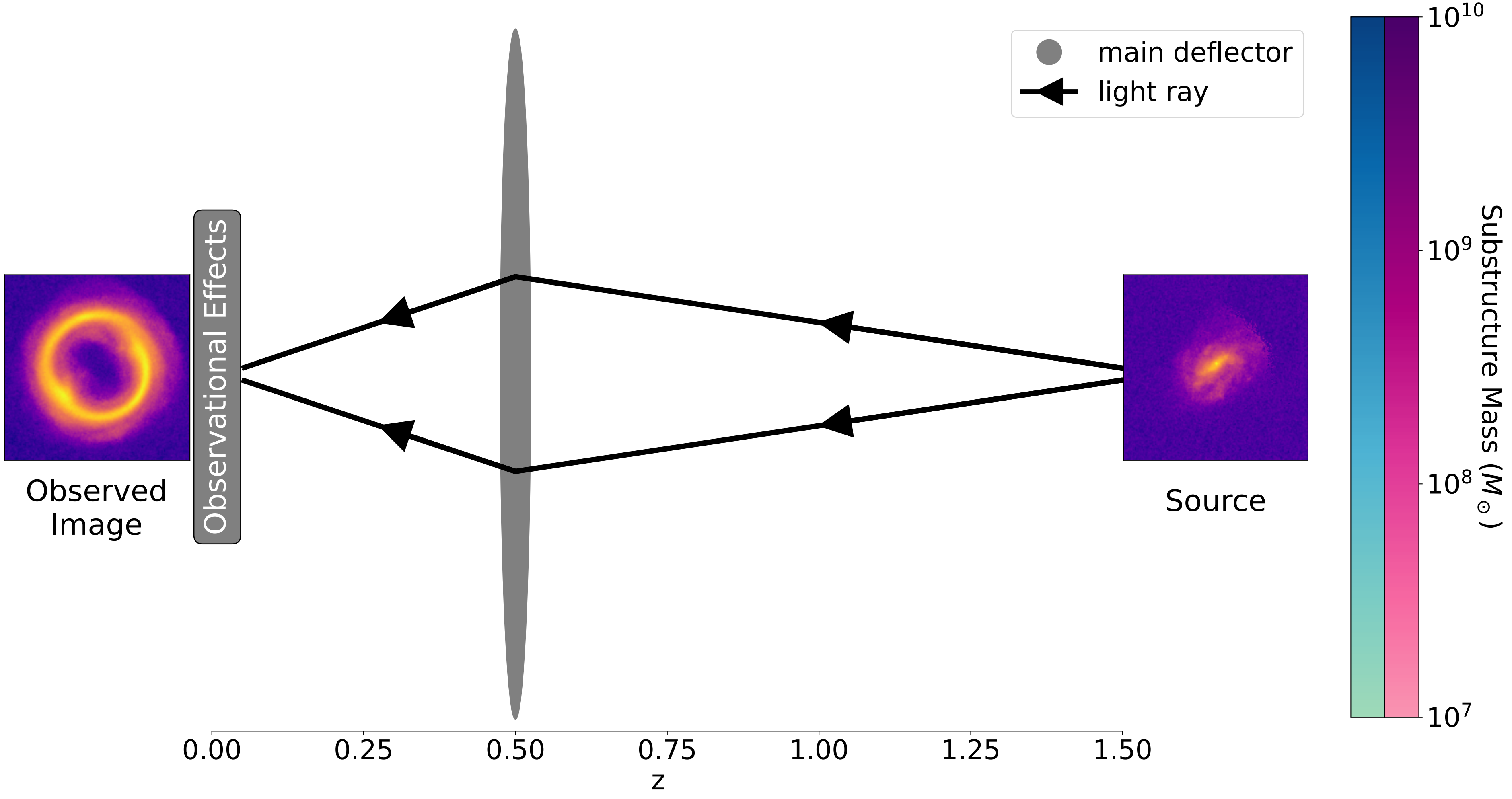
Wagner-Carena et al., 2022

Measuring Halos: Strong Lensing



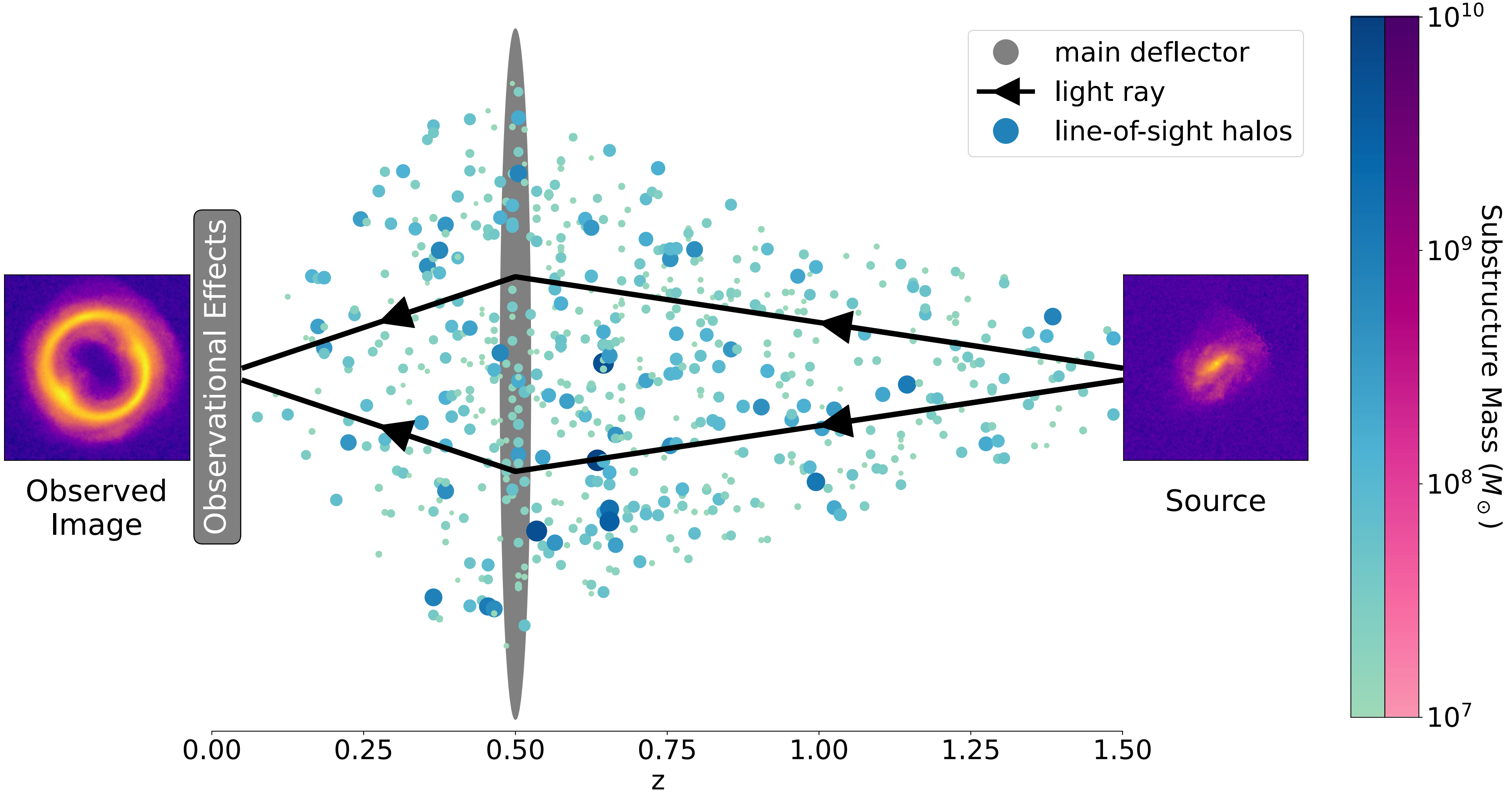
Wagner-Carena et al., 2022

Measuring Halos: Strong Lensing



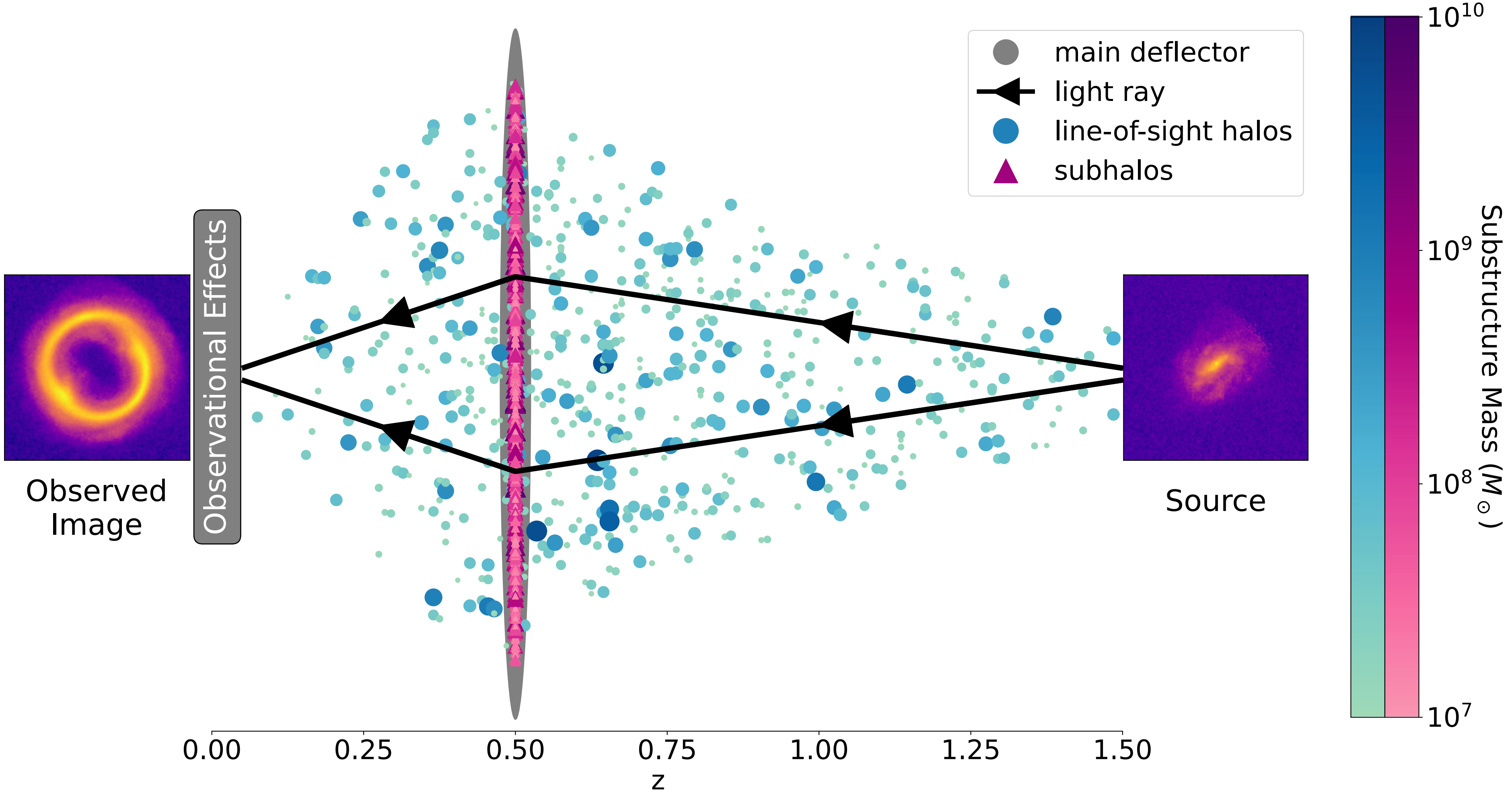
[Wagner-Carena et al., 2022]

Measuring Halos: Strong Lensing



Wagner-Carena et al., 2022

Measuring Halos: Strong Lensing



Wagner-Carena et al., 2022

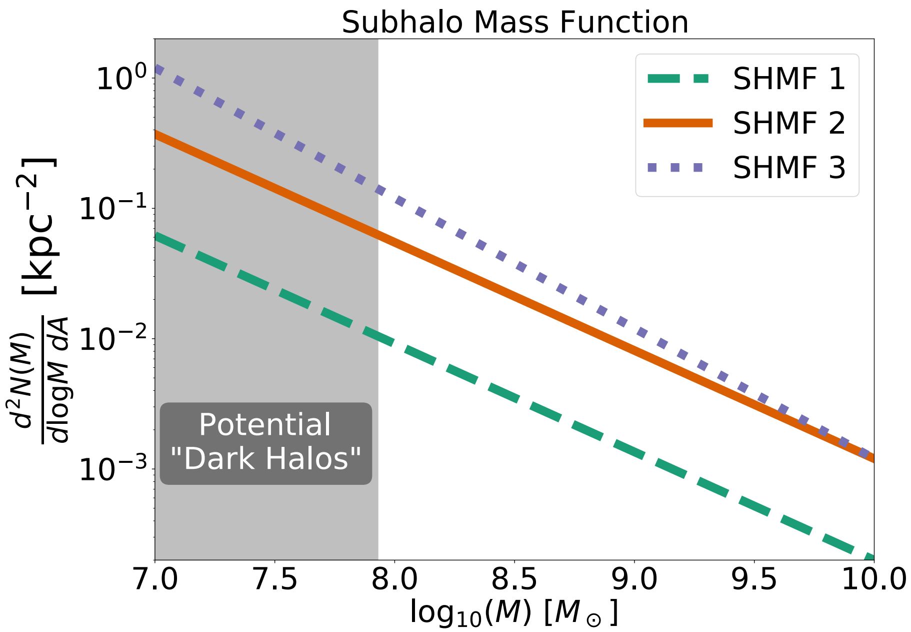
"Statistical" Detection

$$\frac{d^2 N_{\mathrm{sub}}}{dA\ dm_{\mathrm{sub}}} = \boxed{\Sigma_{\mathrm{sub}}} \frac{m_{\mathrm{sub}}^{\gamma_{\mathrm{sub}}}}{m_{\mathrm{pivot,sub}}^{\gamma_{\mathrm{sub}}+1}}$$

Wagner-Carena et al., 2022

"Statistical" Detection

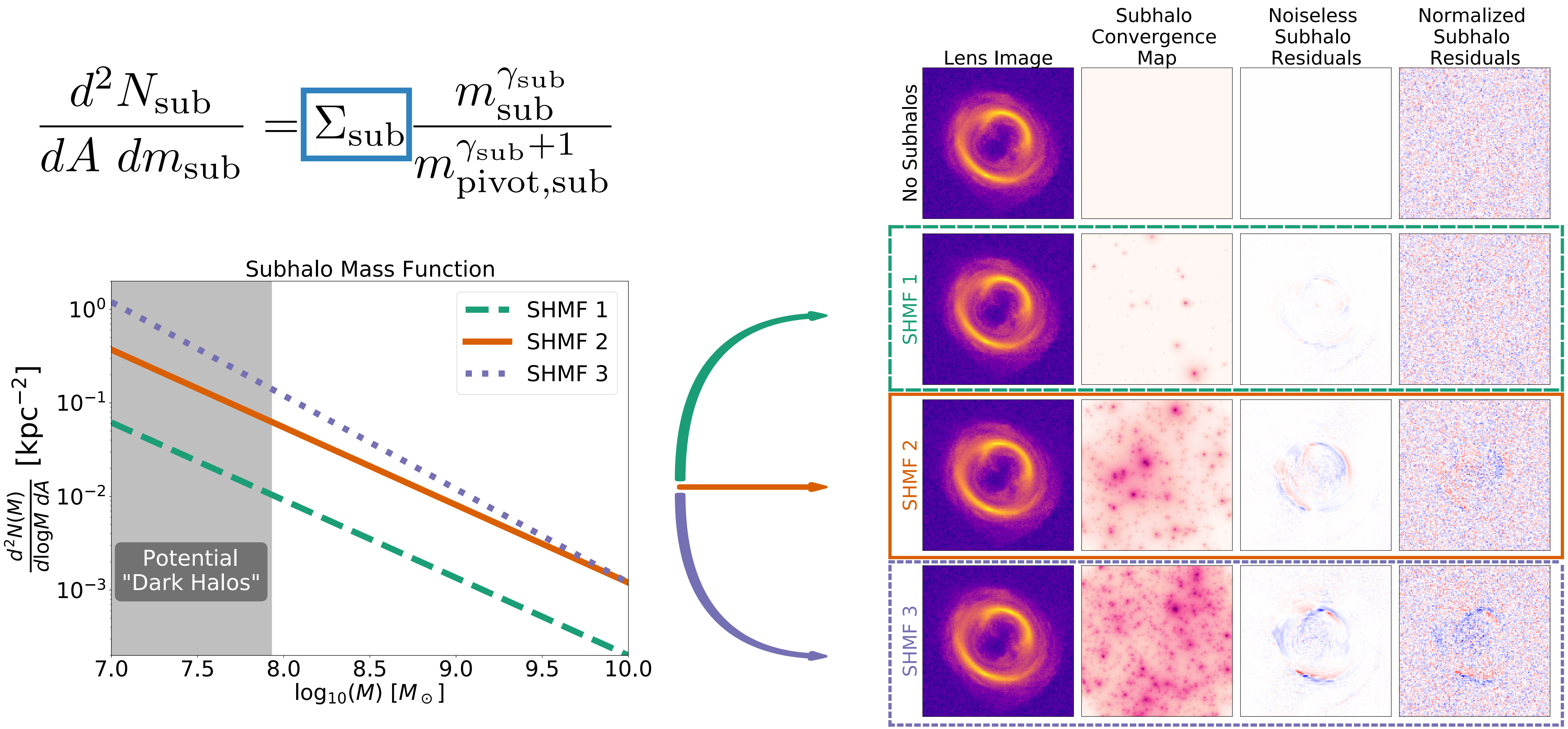
$$\frac{d^2 N_{\rm sub}}{dA\ dm_{\rm sub}} = \boxed{\Sigma_{\rm sub}} \frac{m_{\rm sub}^{\gamma_{\rm sub}}}{m_{\rm pivot,sub}^{\gamma_{\rm sub}+1}}$$



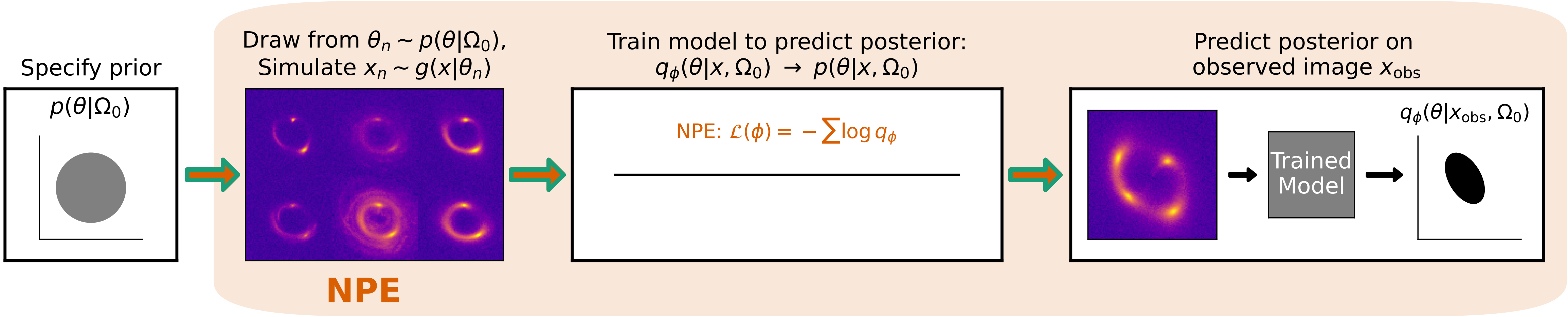
[Wagner-Carena et al., 2022](#)

“Statistical” Detection

$$\frac{d^2 N_{\rm sub}}{dA\ dm_{\rm sub}} = \boxed{\Sigma_{\rm sub}} \frac{m_{\rm sub}^{\gamma_{\rm sub}}}{m_{\rm pivot,sub}^{\gamma_{\rm sub}+1}}$$

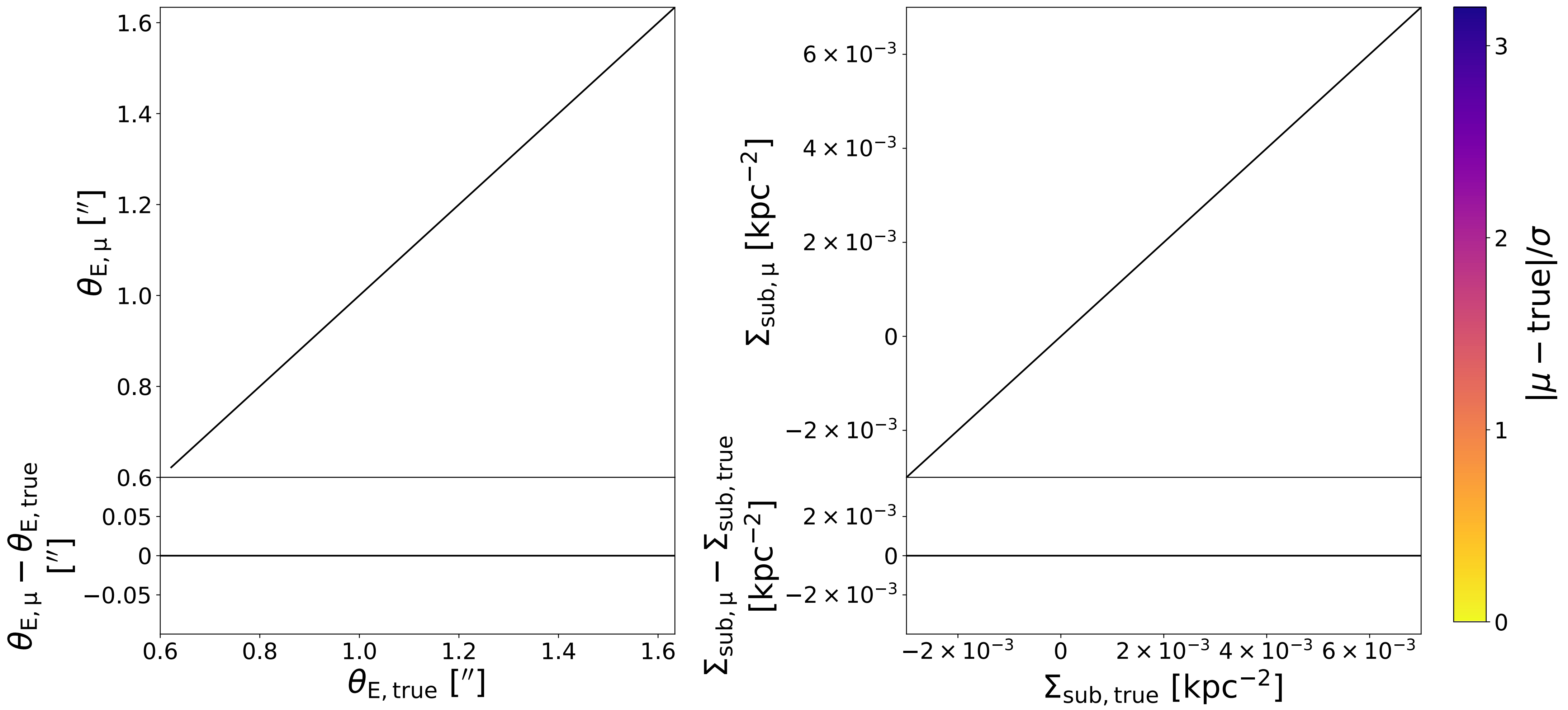


The Framework: Neural Posterior Estimation



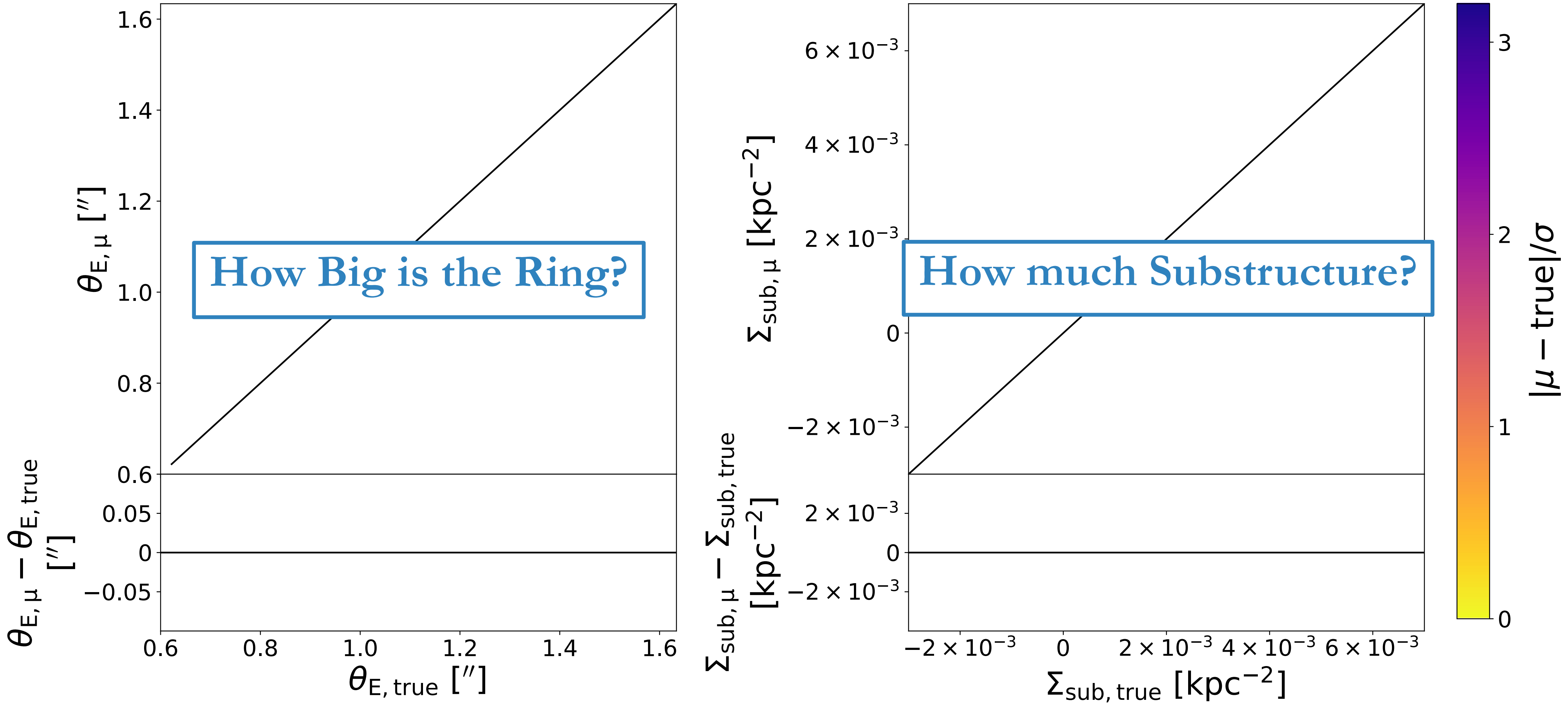
Wagner-Carena et al., 2022

Individual Constraints



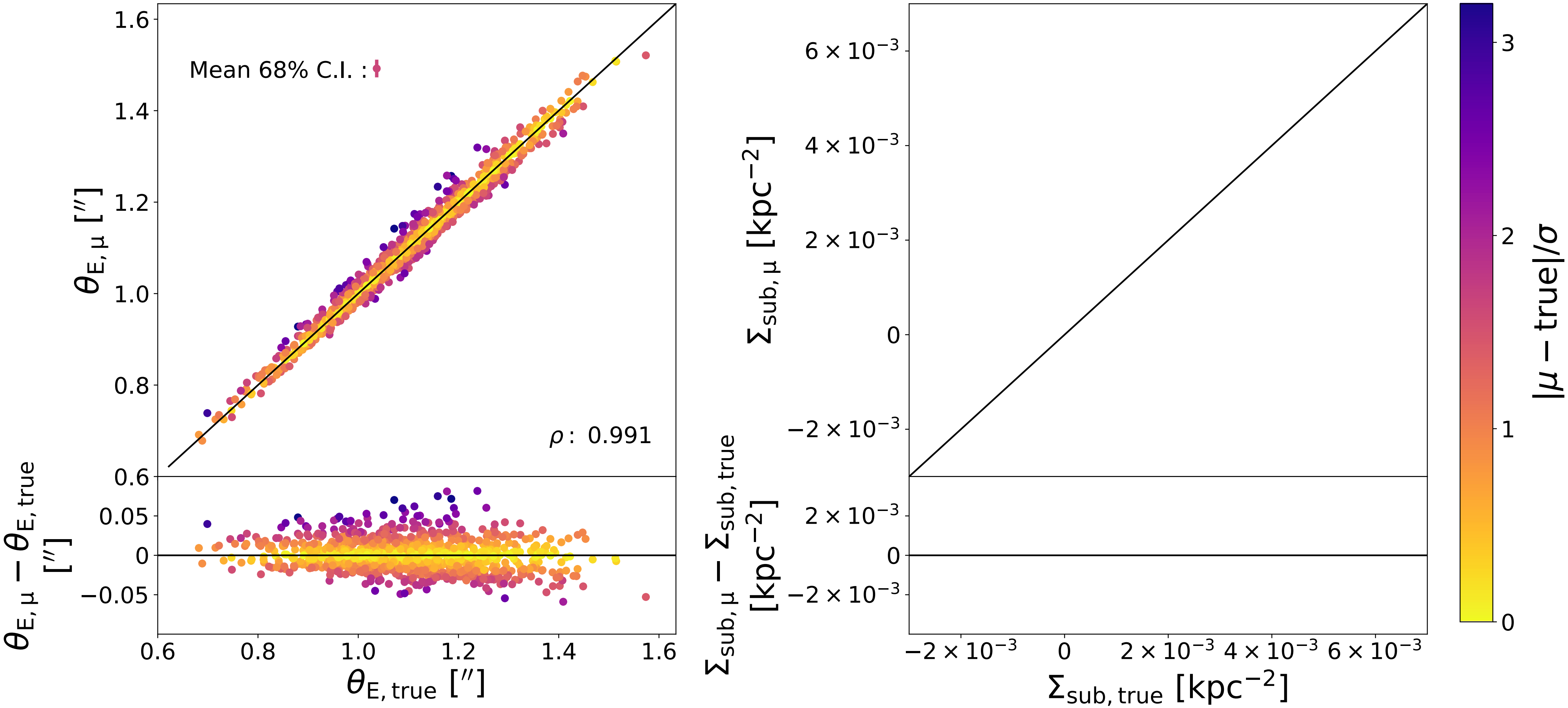
[Wagner-Carena et al., 2022]

Individual Constraints



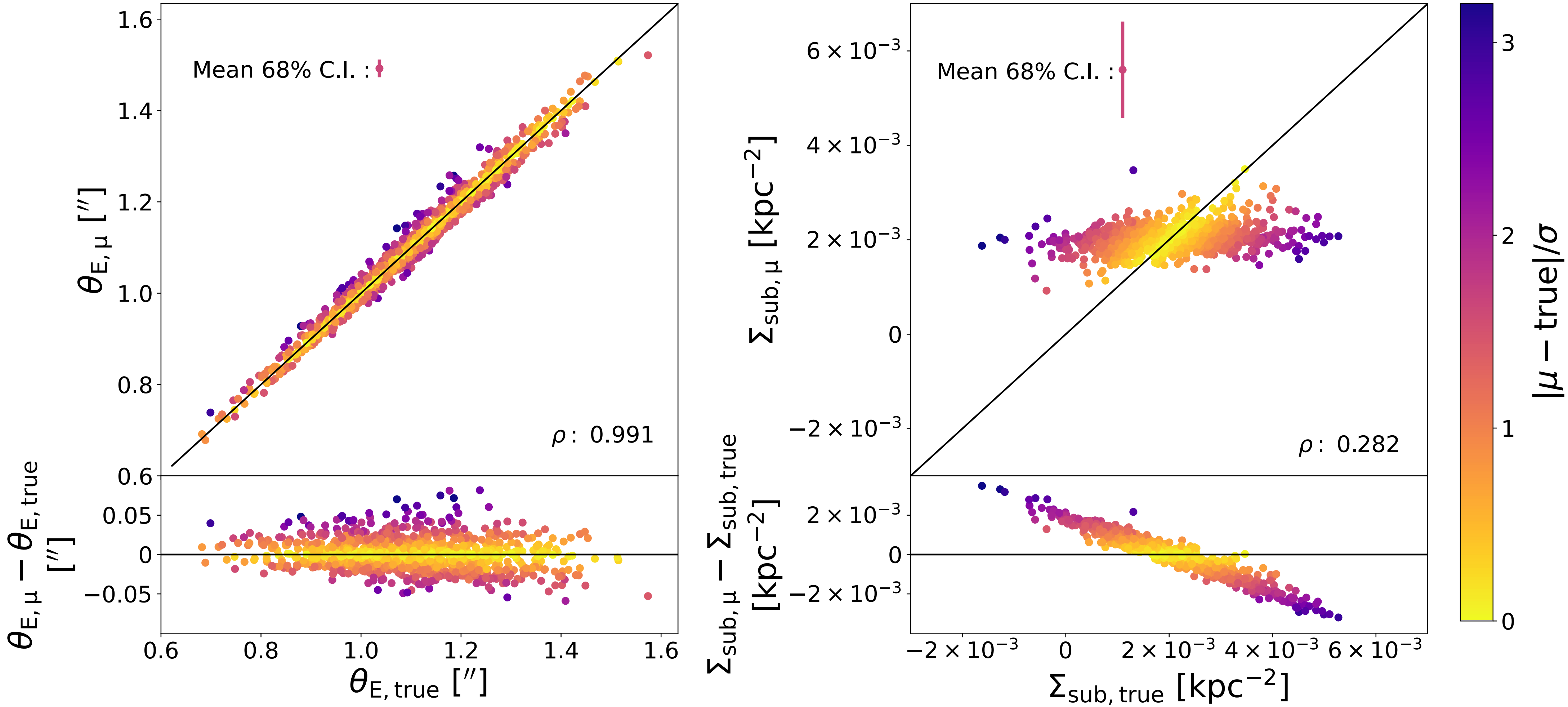
[Wagner-Carena et al., 2022]

Individual Constraints



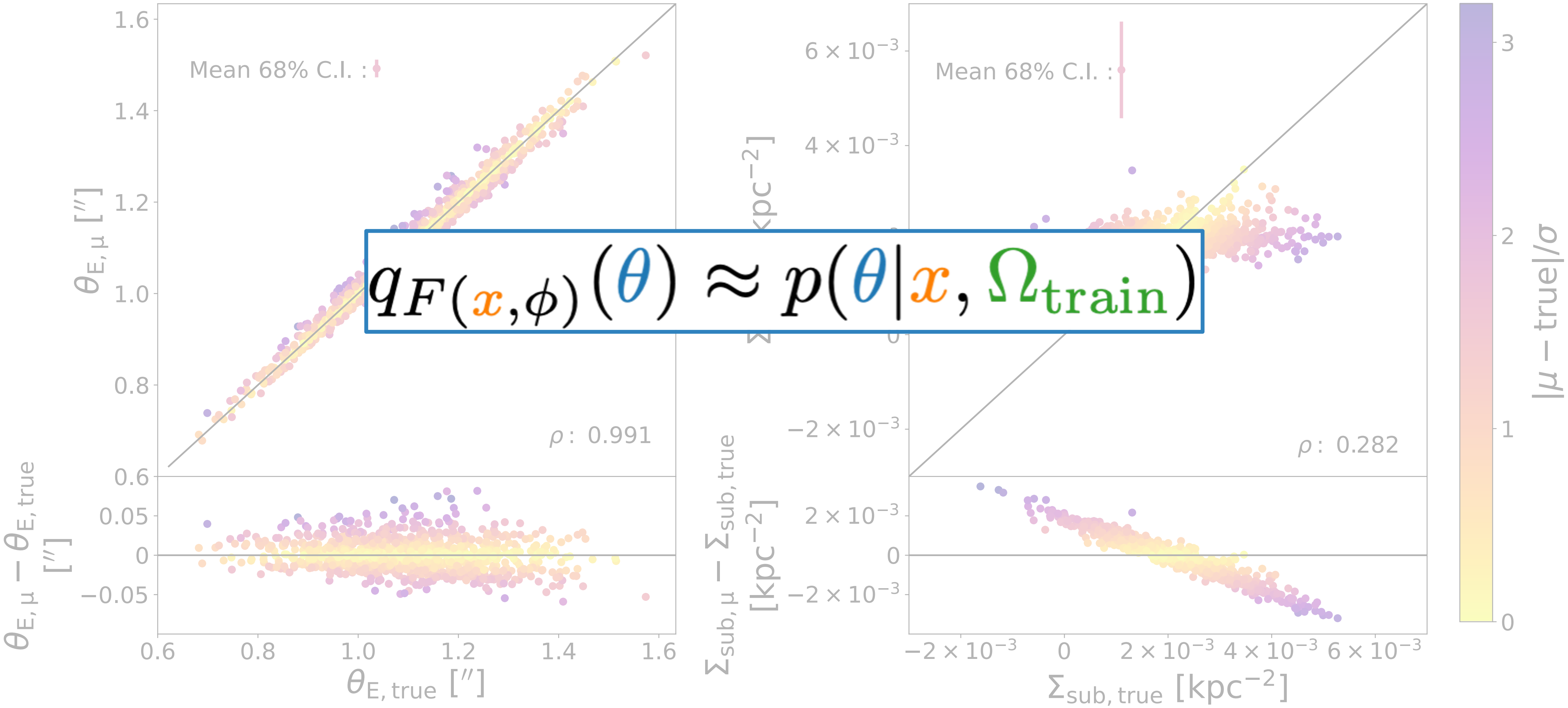
Wagner-Carena et al., 2022

Individual Constraints



[Wagner-Carena et al., 2022]

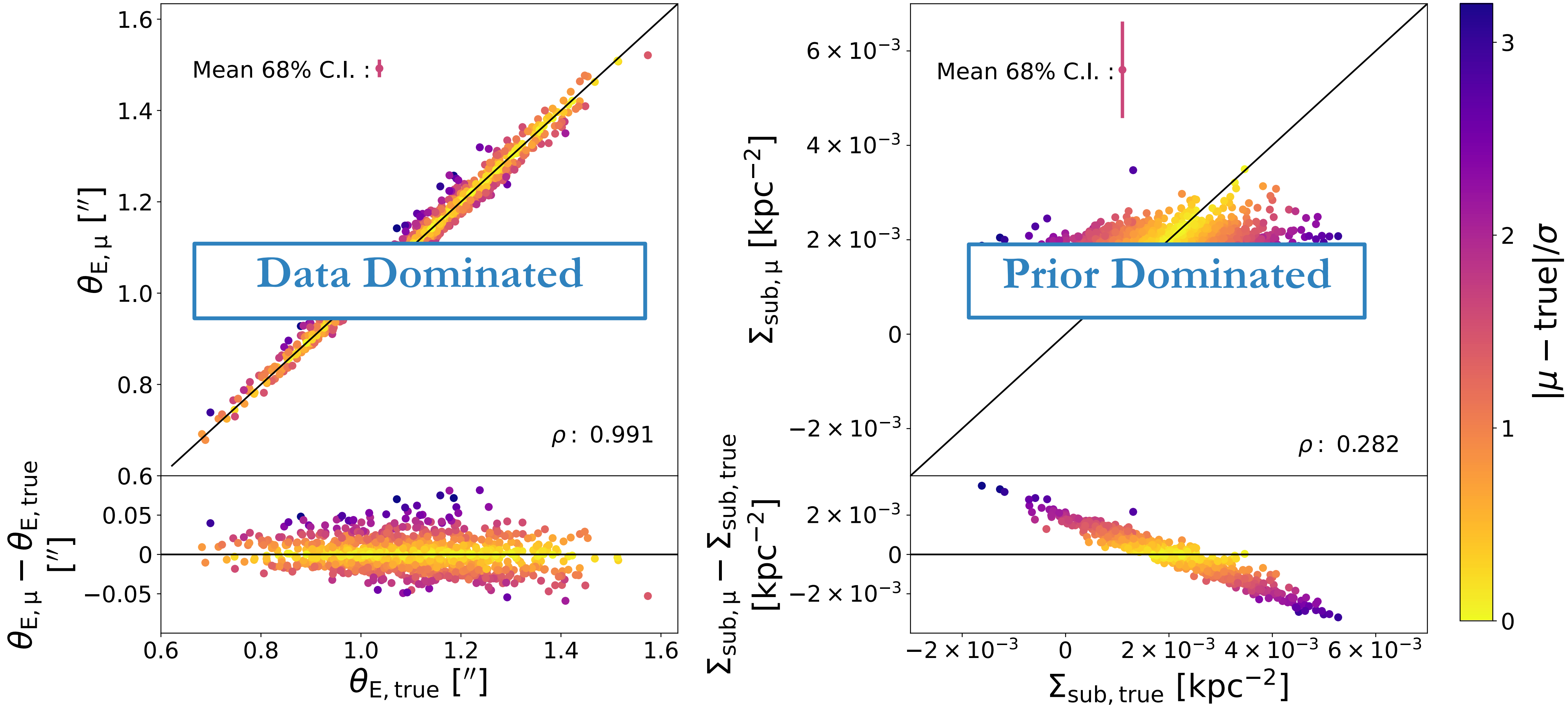
Individual Constraints



$$q_{F(x,\phi)}(\theta) \approx p(\theta | x, \Omega_{\text{train}})$$

Individual Constraints

[Wagner-Carena et al., 2022]

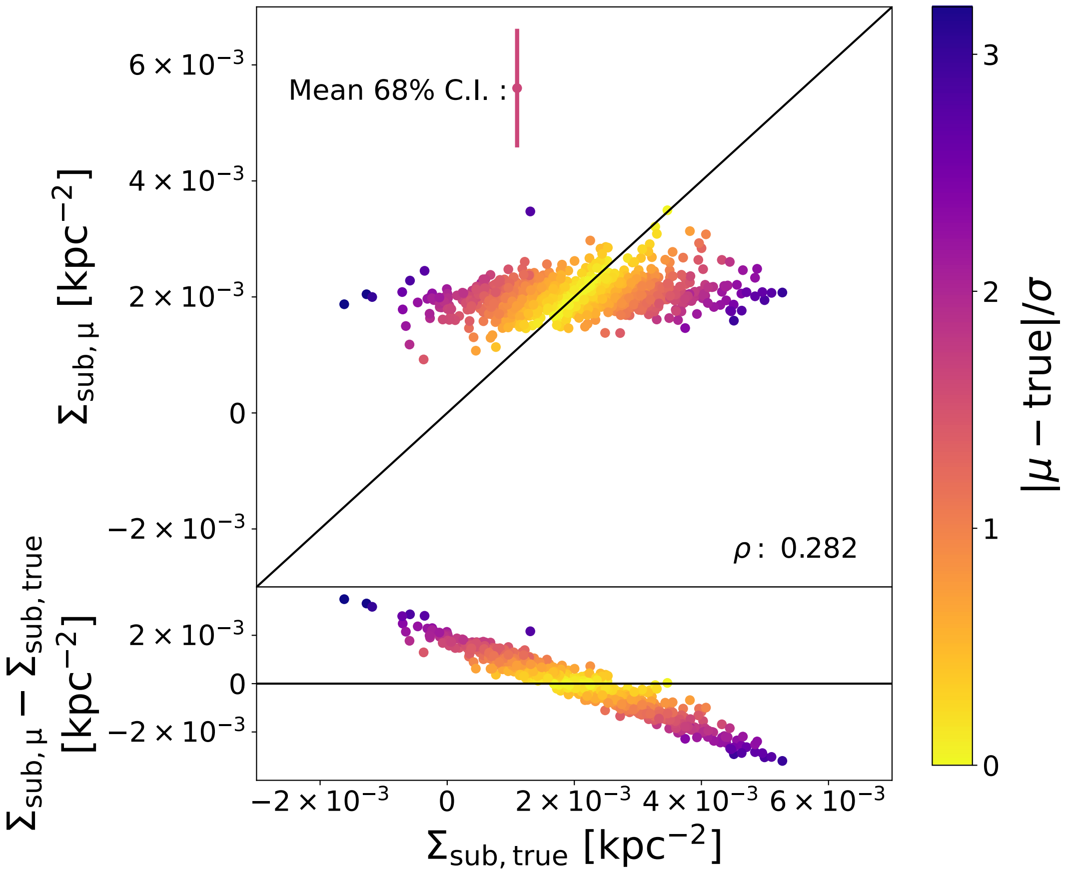


[Wagner-Carena et al., 2024]

Our Current Limitation?

Have neural-density based SBI methods already reached the ***information limit of the data****, or are there* ***methodology choices*** *that are imposing artificial bottlenecks?*

- NPE has theoretical guarantees, but only in the limits of **infinite data**, a sufficiently **expressive model**, and **perfect optimization**



[Wagner-Carena et al., 2024]

Infinite Training Data

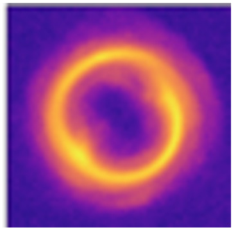
Ω : population hyperparameters

paltas

θ_i : physical parameters z_i : latent variables

LENSTRONOMY

x_i : data



6-9 seconds

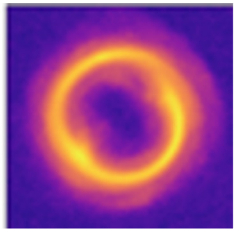
Ω : population hyperparameters

θ_i : physical parameters

paltas + JAX +

LENSTRONOMY

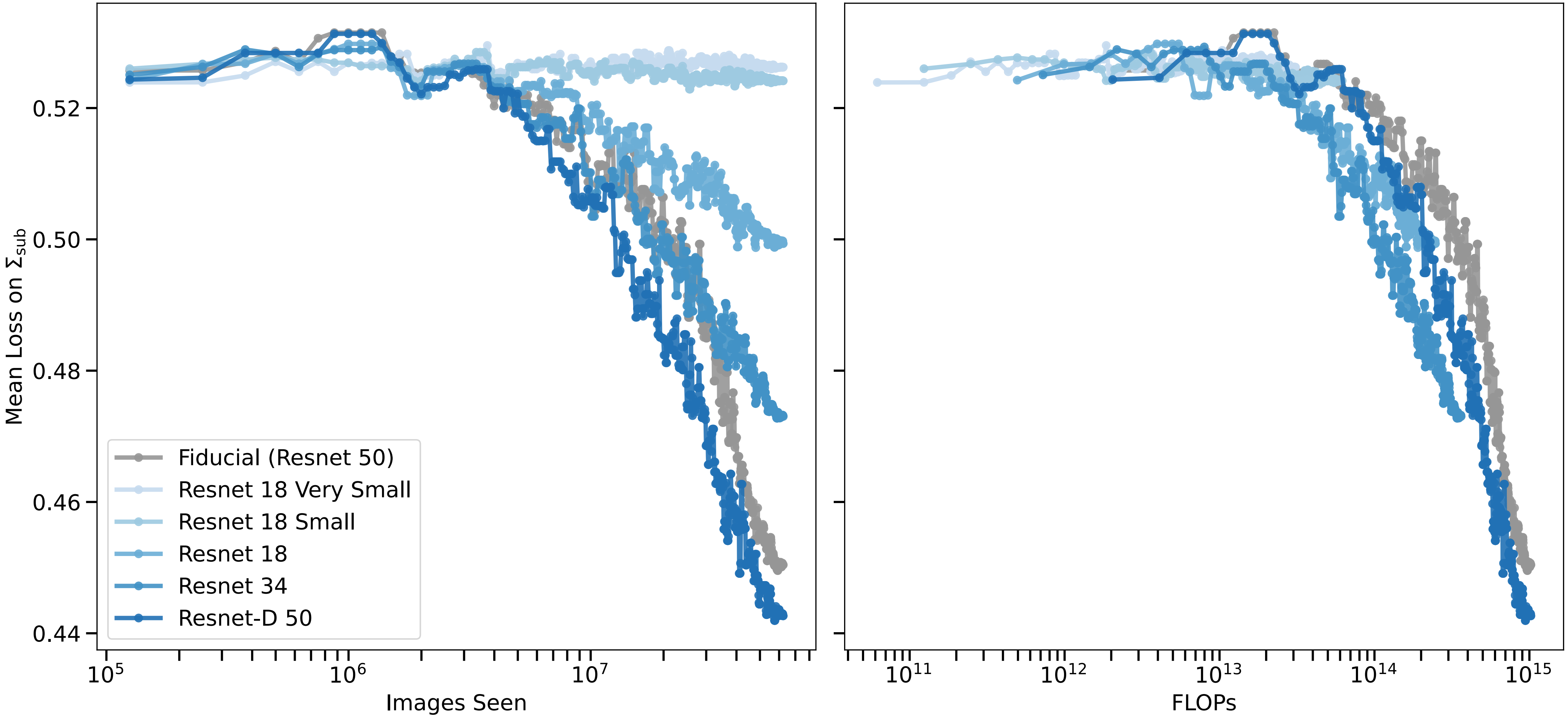
x_i : data



9 ms
differentiable,
parallelizable

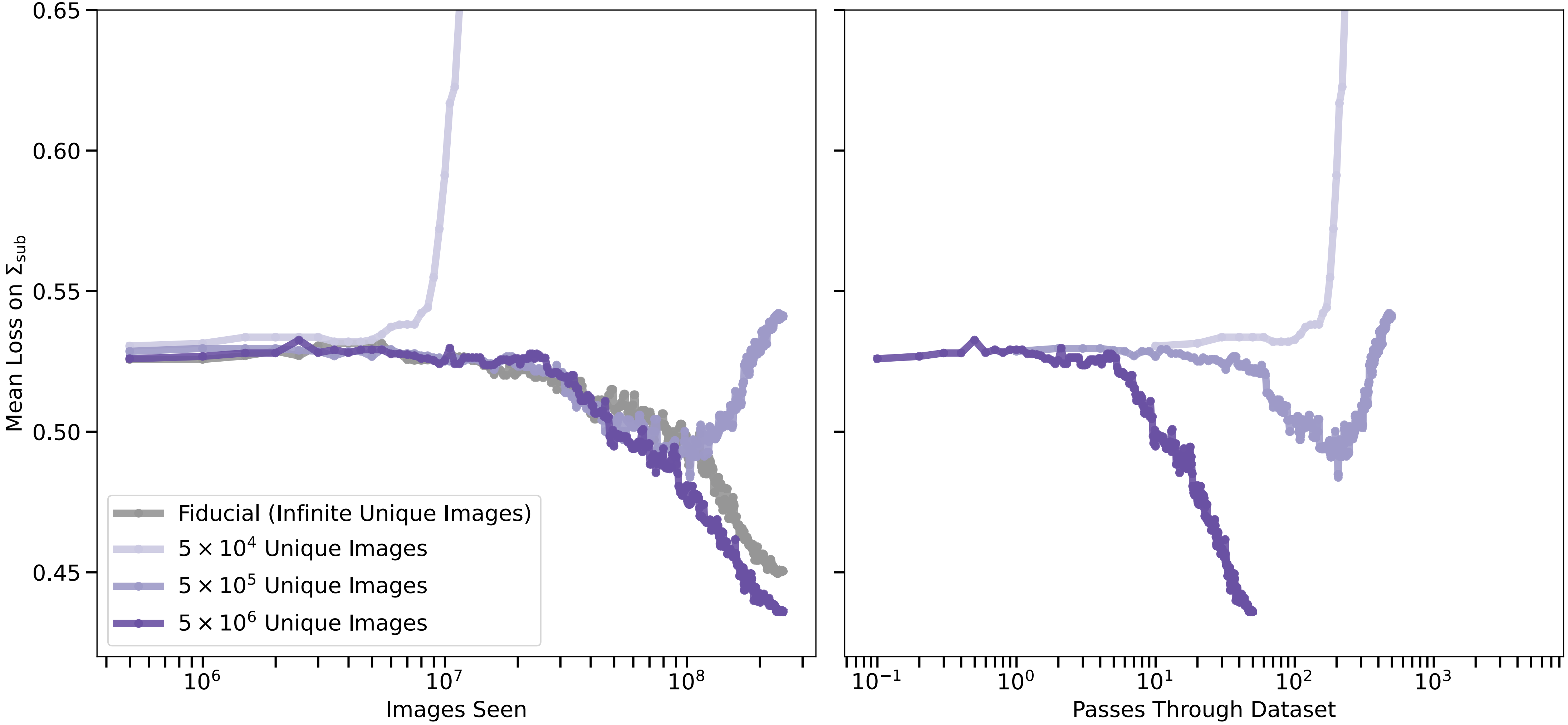
Wagner-Carena et al., 2024

Limits: Model



Wagner-Carena et al., 2024

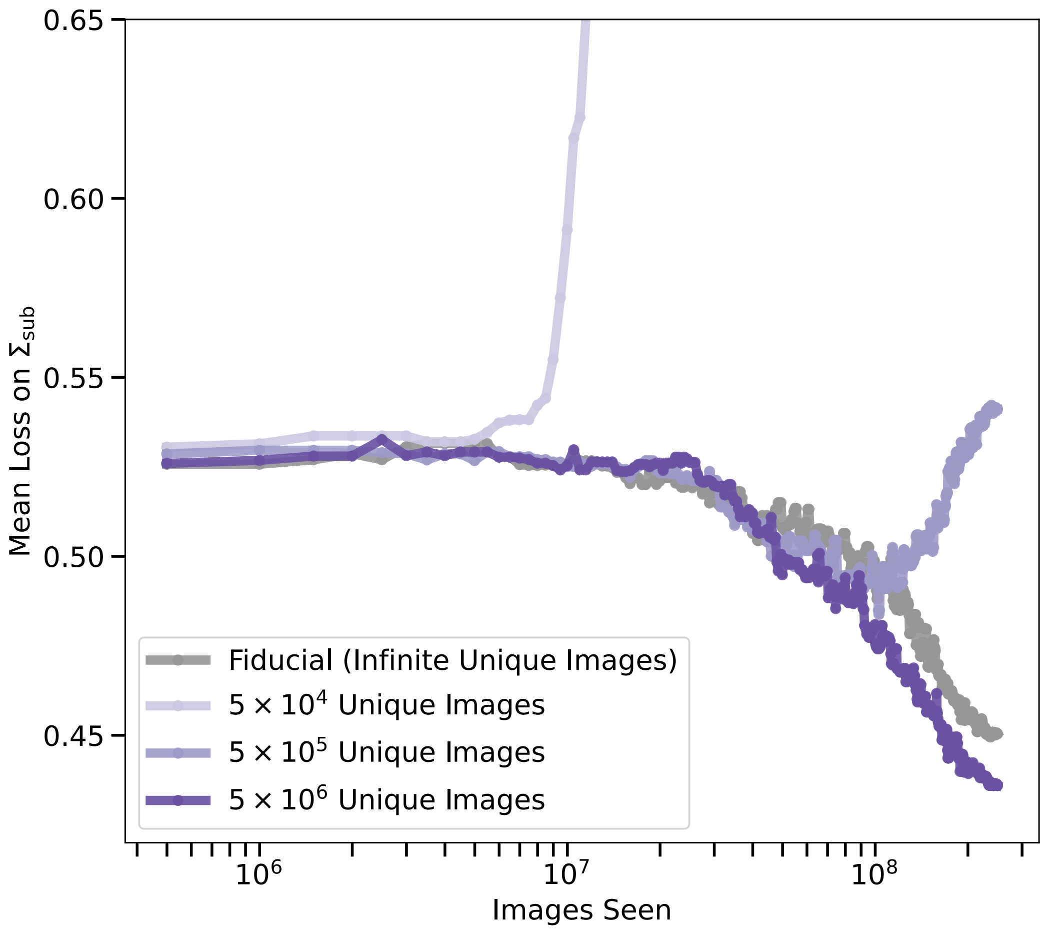
Limits: Training Set Size



Our Current Limitation?

Have neural-density based SBI methods already reached the ***information limit of the data****, or are there* ***methodology choices*** *that are imposing artificial bottlenecks?* - ***No***

- Pushing farther limited by **power-law scaling** in images seen

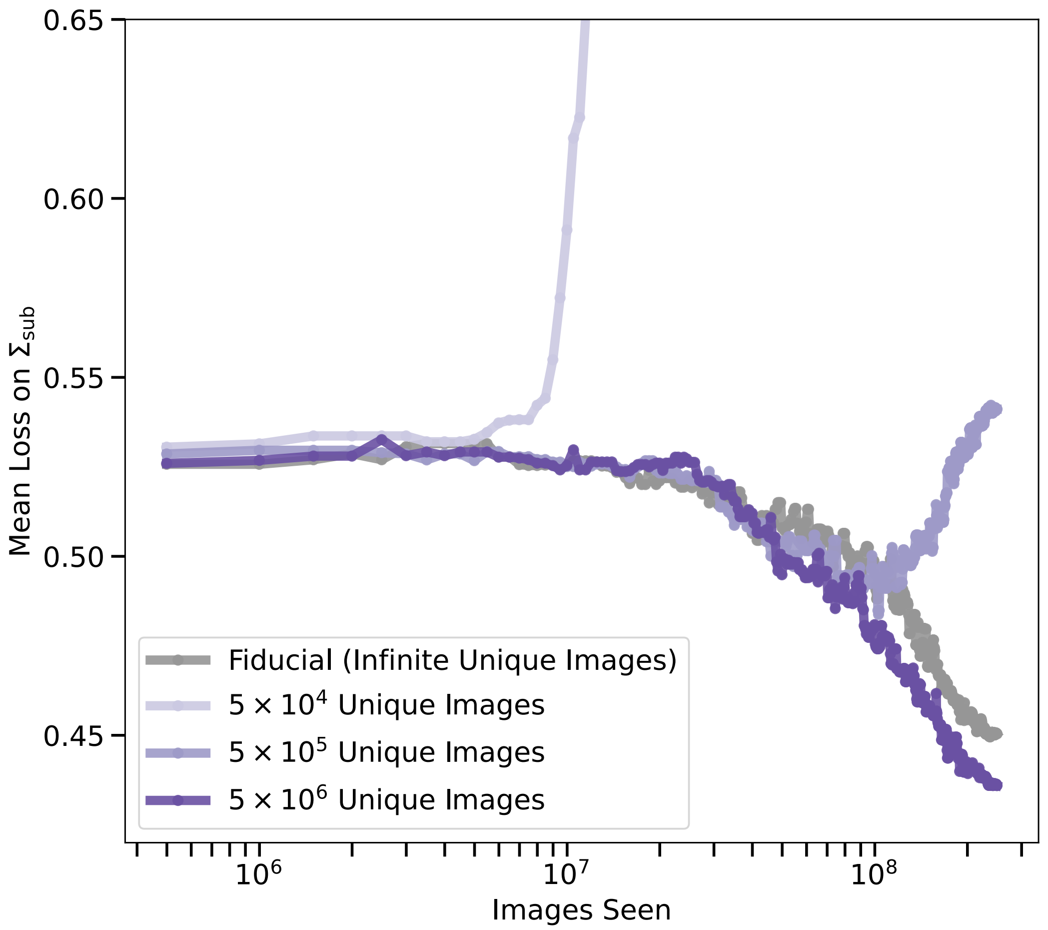


[Wagner-Carena et al., 2024]

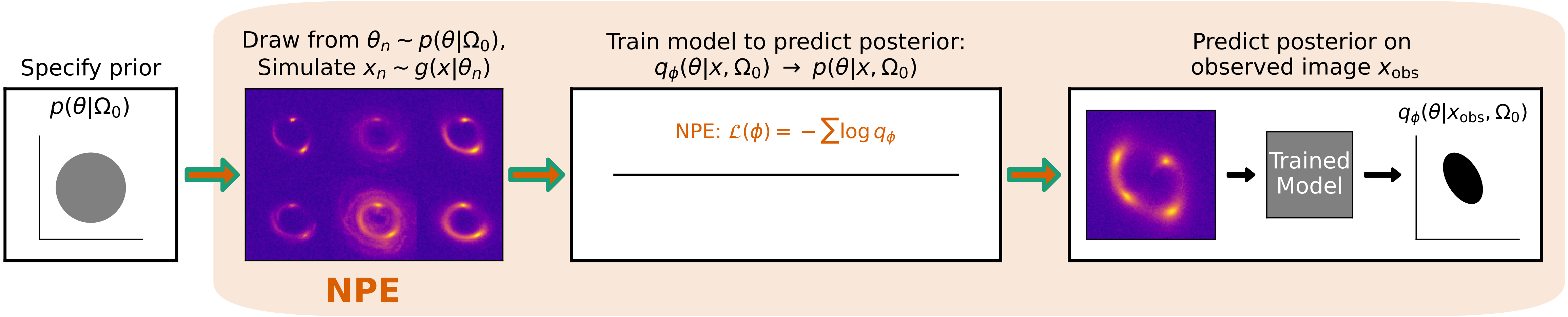
Our Current Limitation?

Have neural-density based SBI methods already reached the ***information limit of the data****, or are there* ***methodology choices*** *that are imposing artificial bottlenecks?* - ***No***

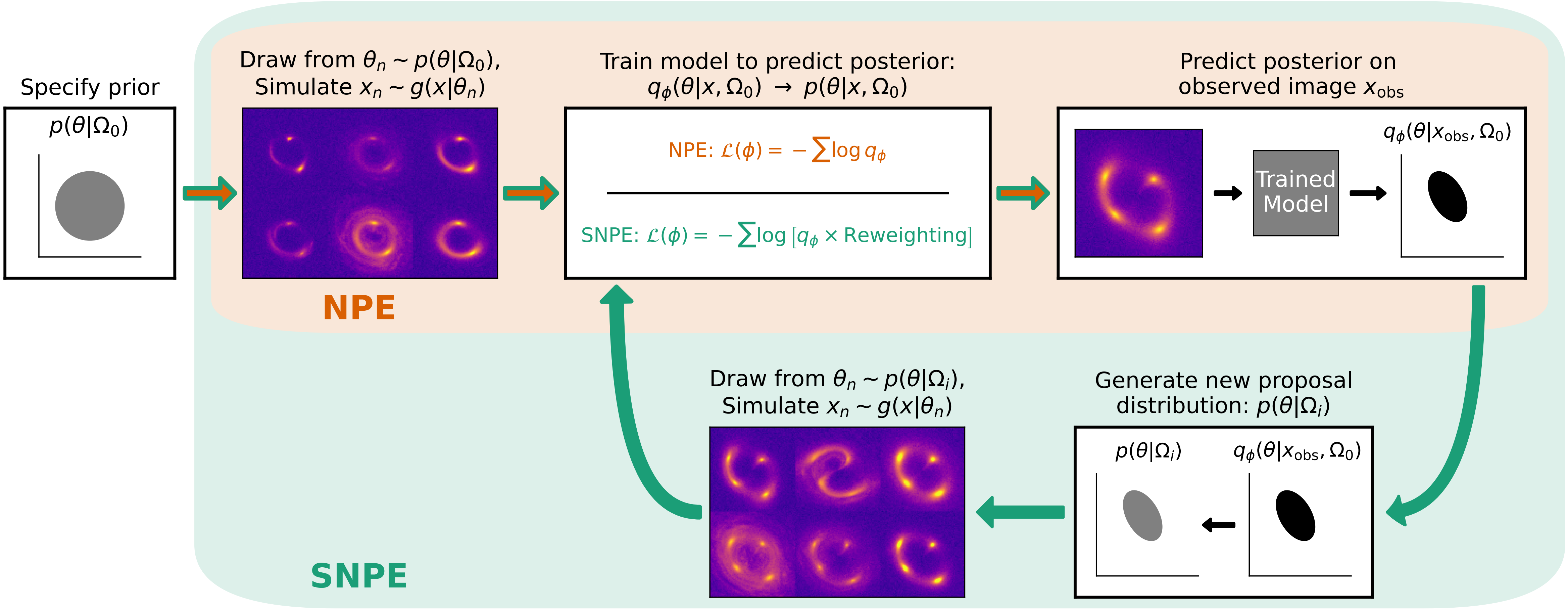
- Pushing farther limited by **power-law scaling** in images seen
- If only we had a **more efficient** training set…
  - … but **I made the training set**



Sequential Neural Posterior Estimator



Sequential Neural Posterior Estimator



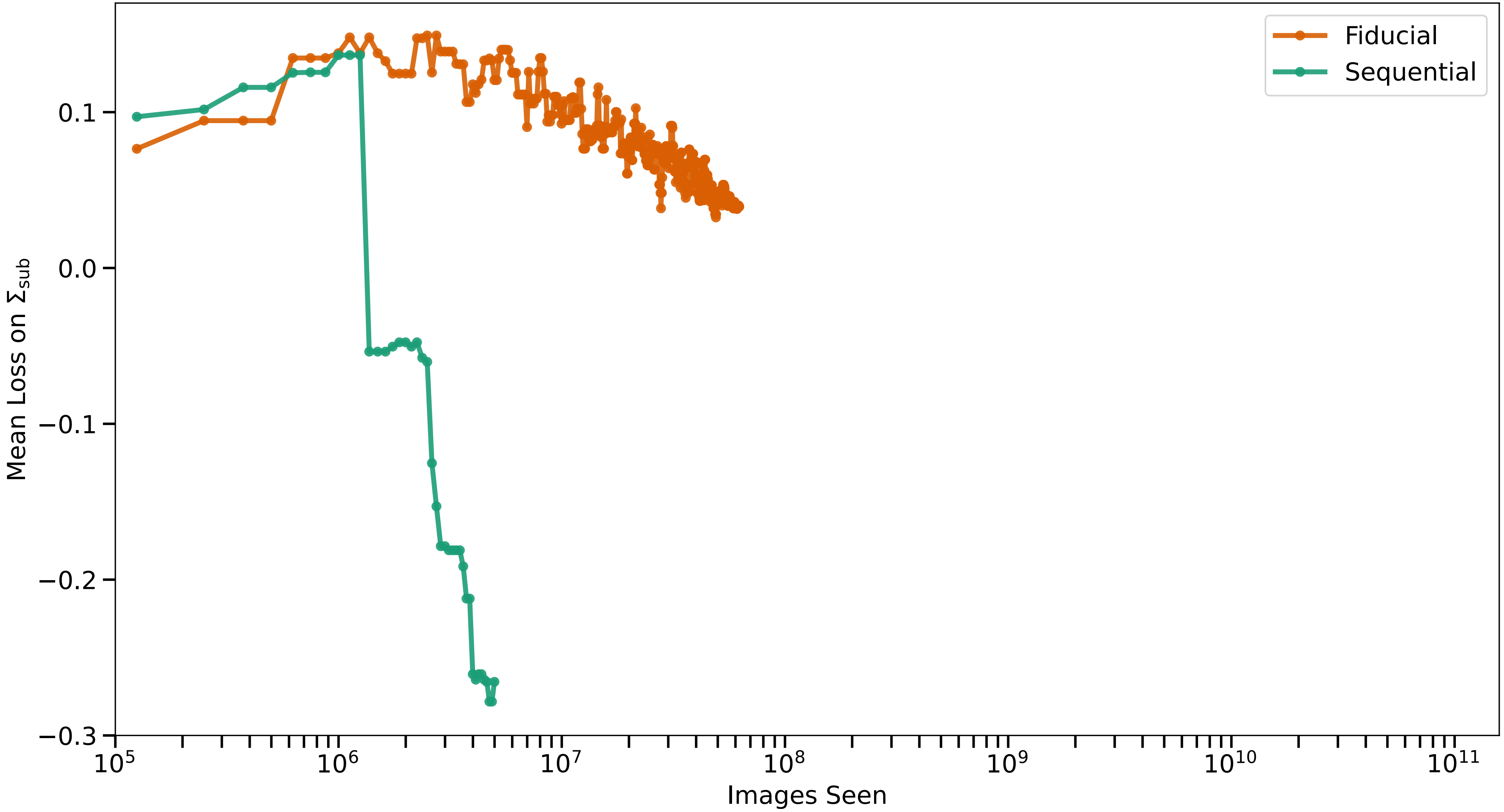
Wagner-Carena et al., 2024

SNPE Comparison

- Generate a set of **30 'true' observations** with the mean and scatter in the SHMF normalization **from N-body simulations**.
- Run sequential inference to answer:
  - *Are we still limited by the same power-law scaling, or do we accelerate learning?*
  - *Are we more data-efficient at the population level?*

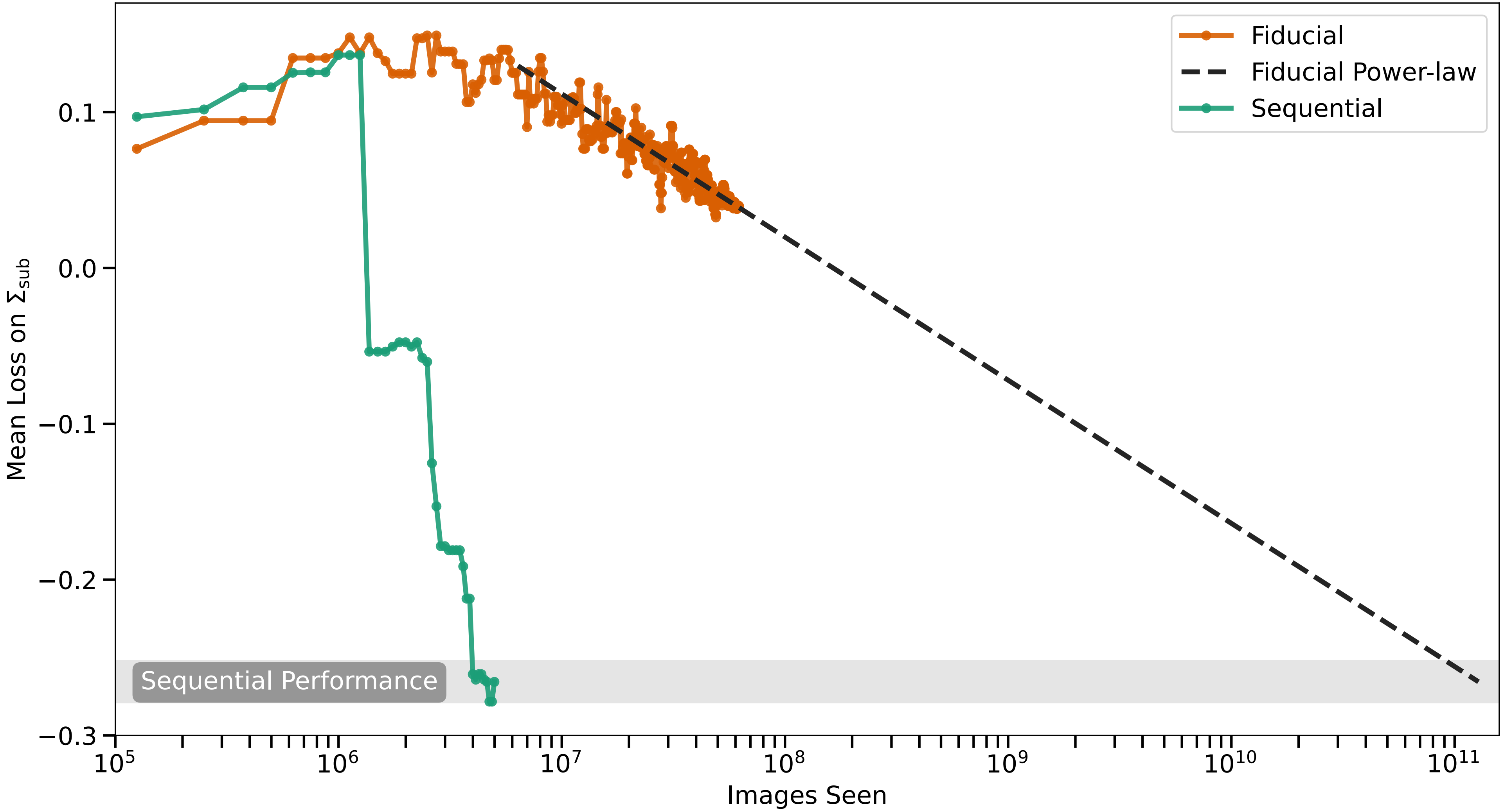
Wagner-Carena et al., 2024

Loss on Σ_{sub} Comparison



Loss on Σ_{sub} Comparison

Wagner-Carena et al., 2024



Wagner-Carena et al., 2024

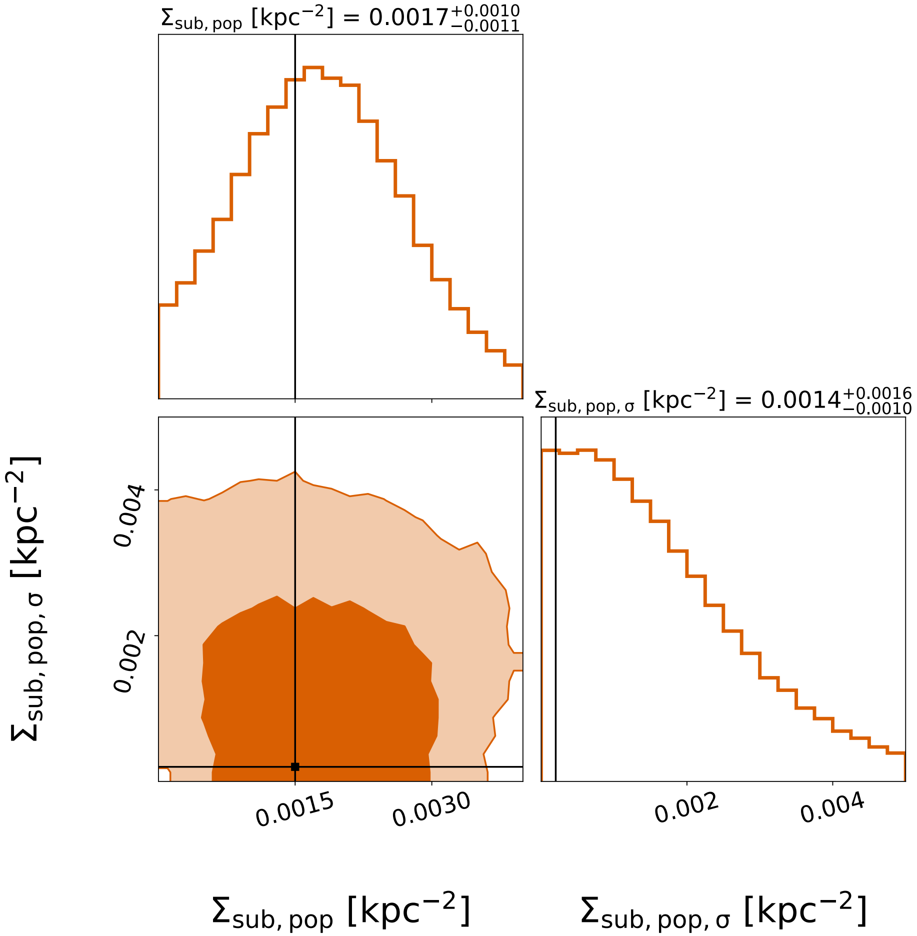
SNPE Comparison

- *Are we still limited by the same power-law scaling, or do we accelerate learning?*
  - On our 'difficult' parameter-of-interest, sequential achieves performance gains equivalent to over **three orders-of-magnitude more images**
- *Are we more data-efficient at the population level?*

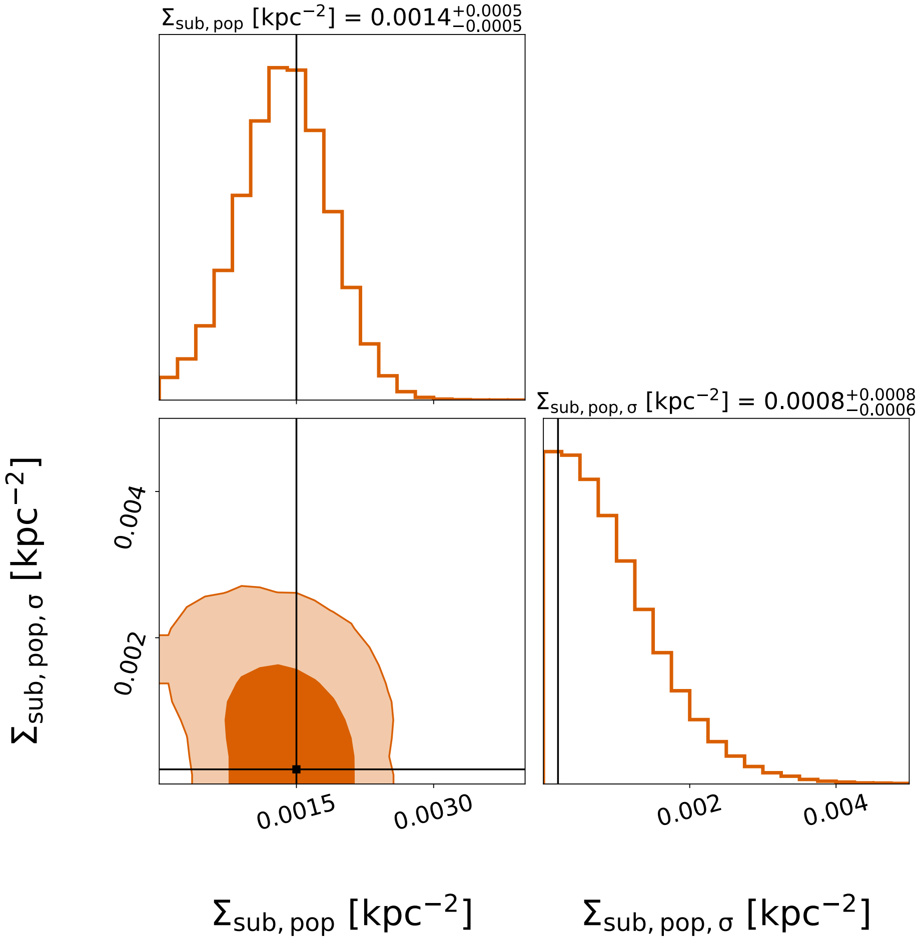
[Wagner-Carena et al., 2024]()

Hierarchical Inference - NPE

NPE: 10 Lenses



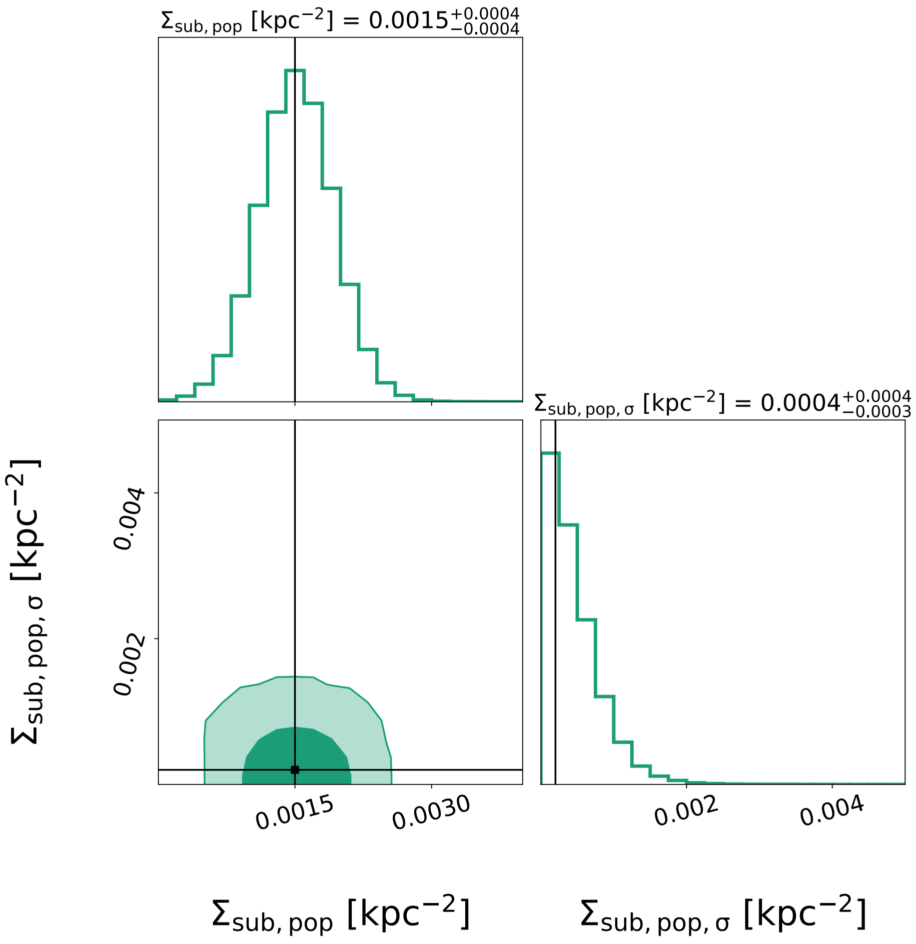
NPE: 30 Lenses



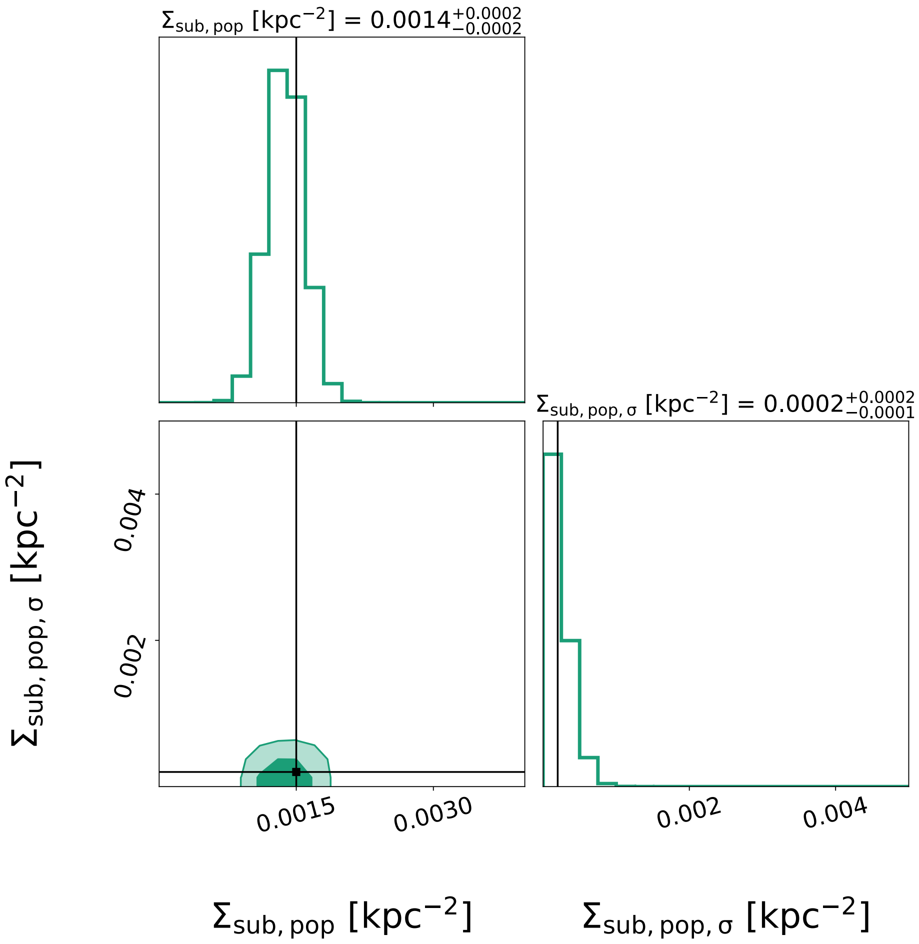
Wagner-Carena et al., 2024

Hierarchical Inference - SNPE

SNPE: 10 Lenses

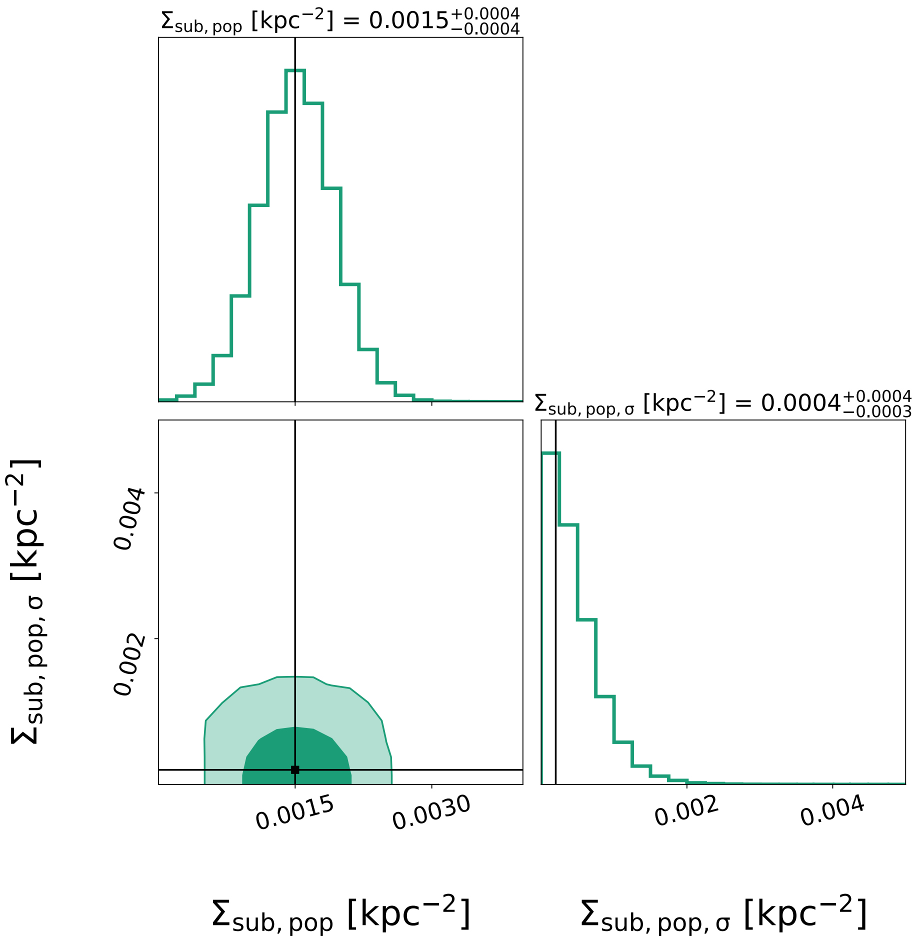


SNPE: 30 Lenses

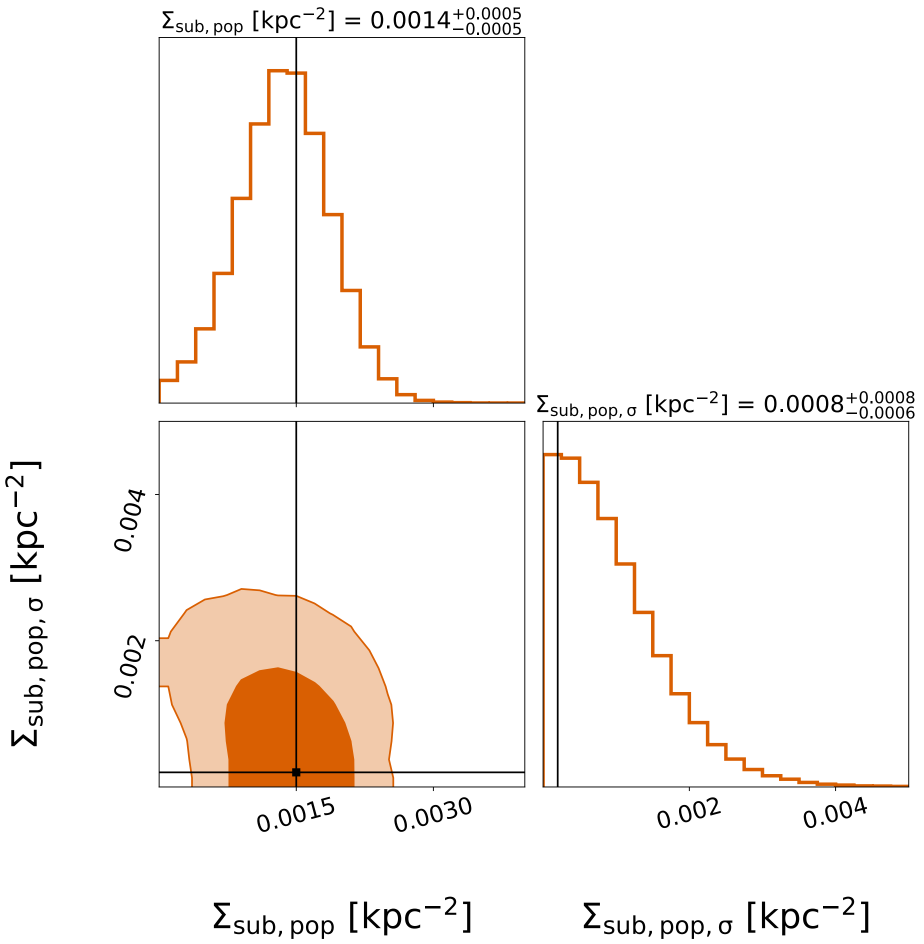


Hierarchical Inference - Comparison

SNPE: 10 Lenses

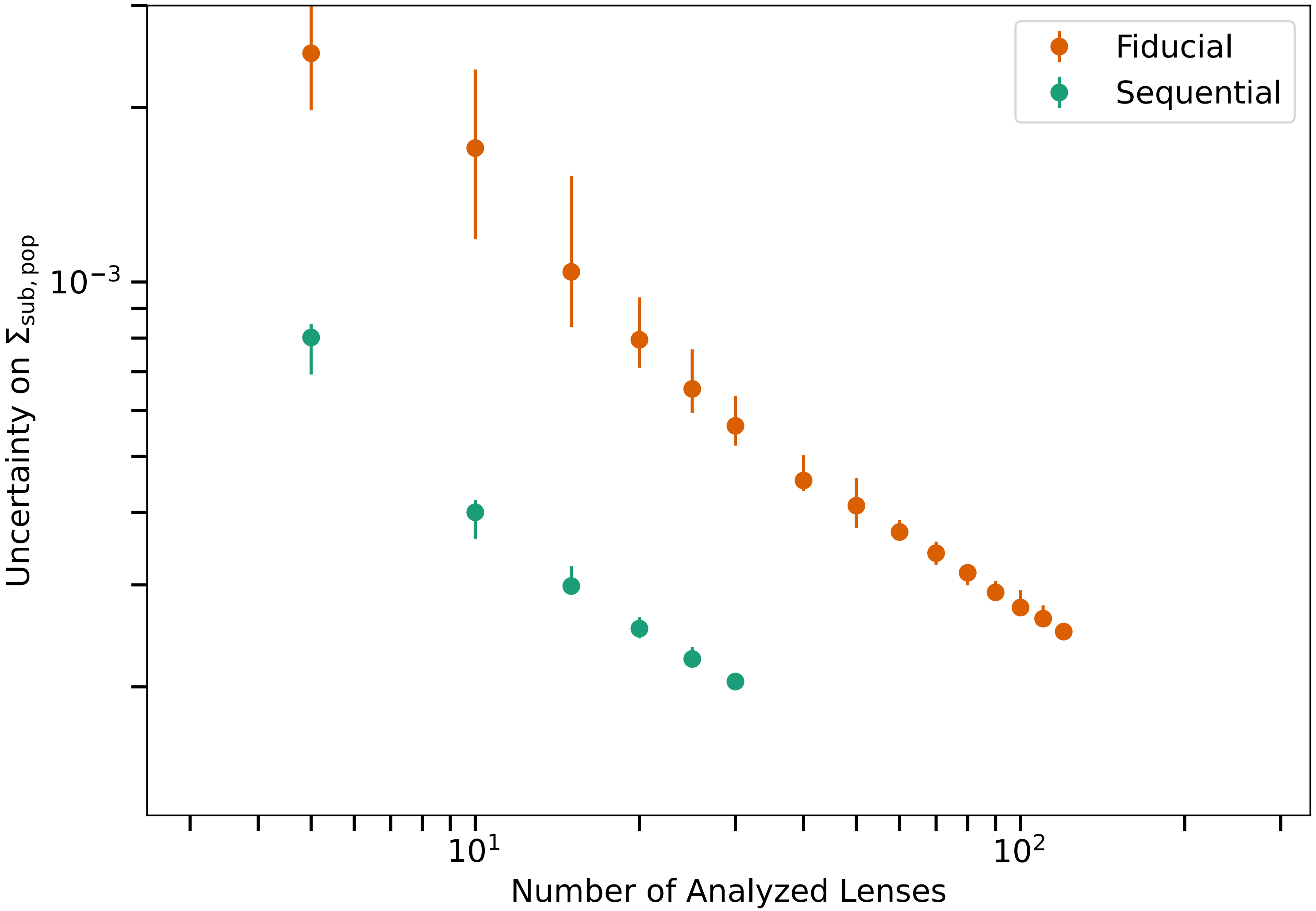


NPE: 30 Lenses



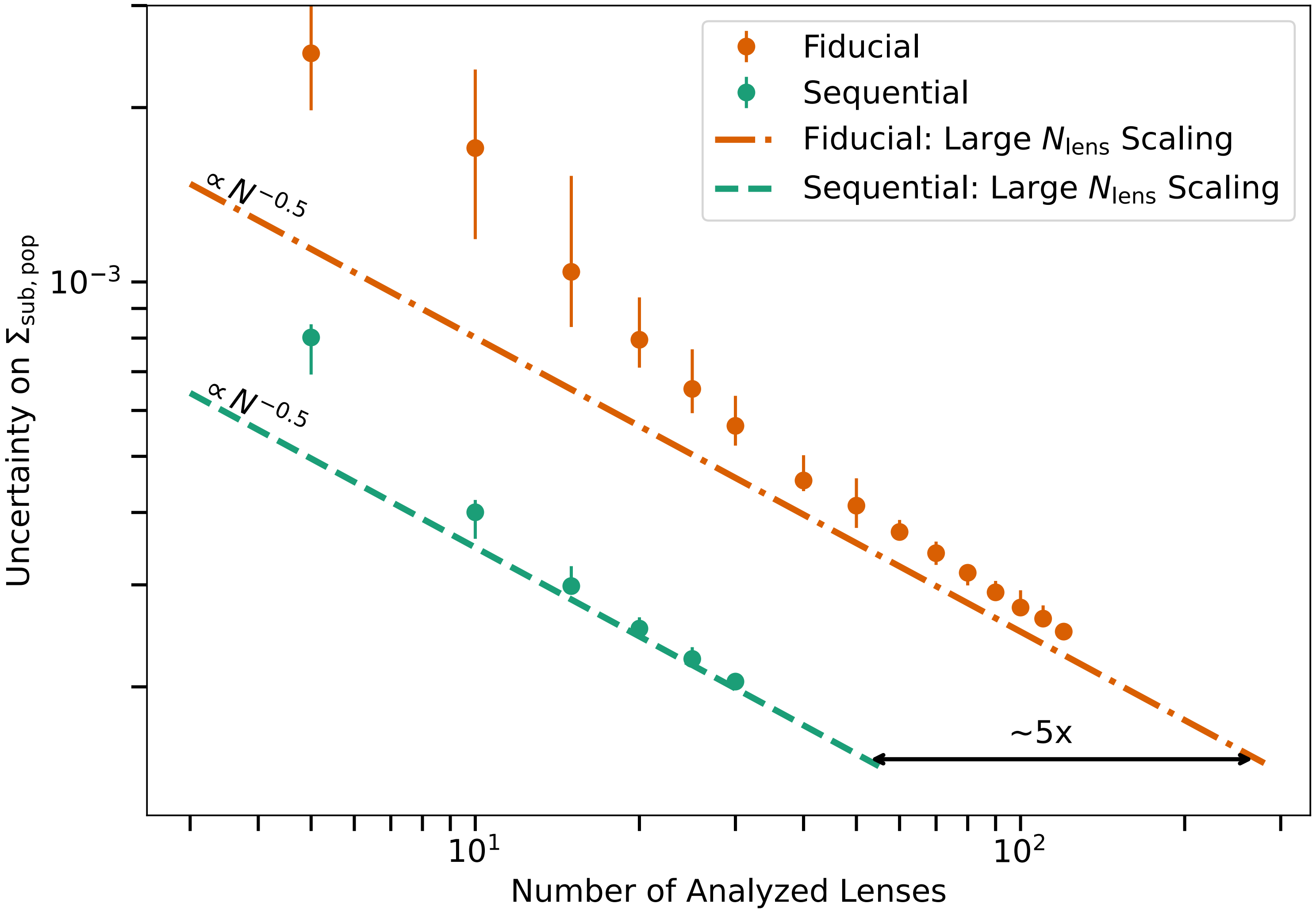
Wagner-Carena et al., 2024

Hierarchical Inference - Comparison



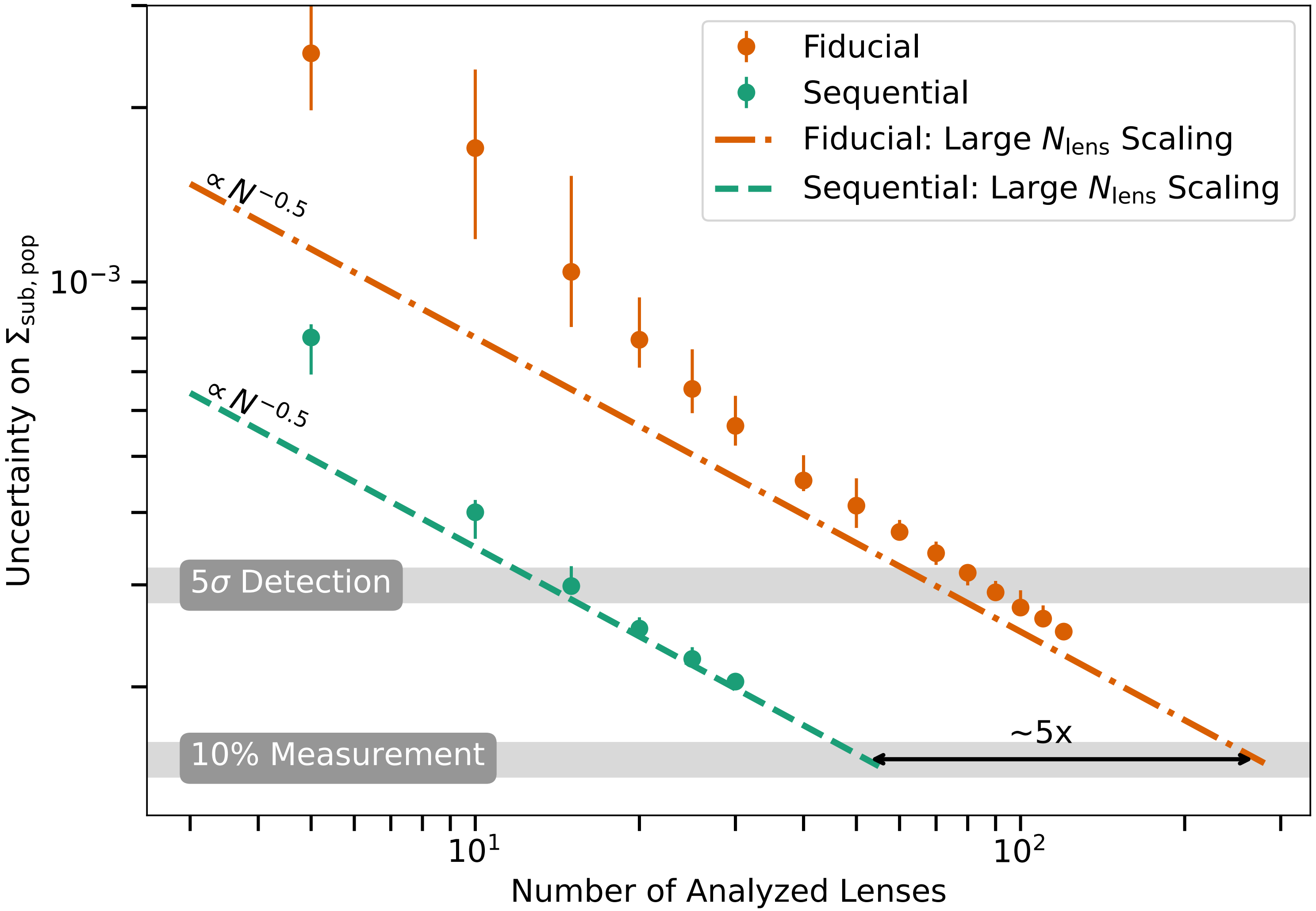
[Wagner-Carena et al., 2024]

Hierarchical Inference - Comparison



Wagner-Carena et al., 2024

Hierarchical Inference - Comparison



Wagner-Carena et al., 2024

SNPE Comparison

- *Are we still limited by the same power-law scaling, or do we accelerate learning?*
  - On our 'difficult' parameter-of-interest, sequential achieves performance gains equivalent to over **three orders-of-magnitude more images**
- *Are we more data-efficient at the population level?*
  - Unbiased population constraints that are **~5x more efficient per lens**
  - ~50 lenses to produce a 10% measurement compared to ~300 lenses

Wagner-Carena et al., 2024

Conclusions

Strong Lensing

- We are not **data limited**, we are limited by our training sets
- Sequential methods drastically improve constraining power – similar improvements from the naïve approach would be **computationally untenable**

Wagner-Carena et al., 2024

Conclusions

Strong Lensing

- We are not **data limited**, we are limited by our training sets
- Sequential methods drastically improve constraining power – similar improvements from the naïve approach would be **computationally untenable**

Broader Implications

- The **quality** of our training sets is not just determined by **size**
- As we employ SBI, we need to treat training set generation and model optimization as **interconnected stages** of our analysis