Productionizing Machine Learning Inference for Modern Cosmological Surveys

Matthew Ho (matthew.annam.ho@gmail.com)

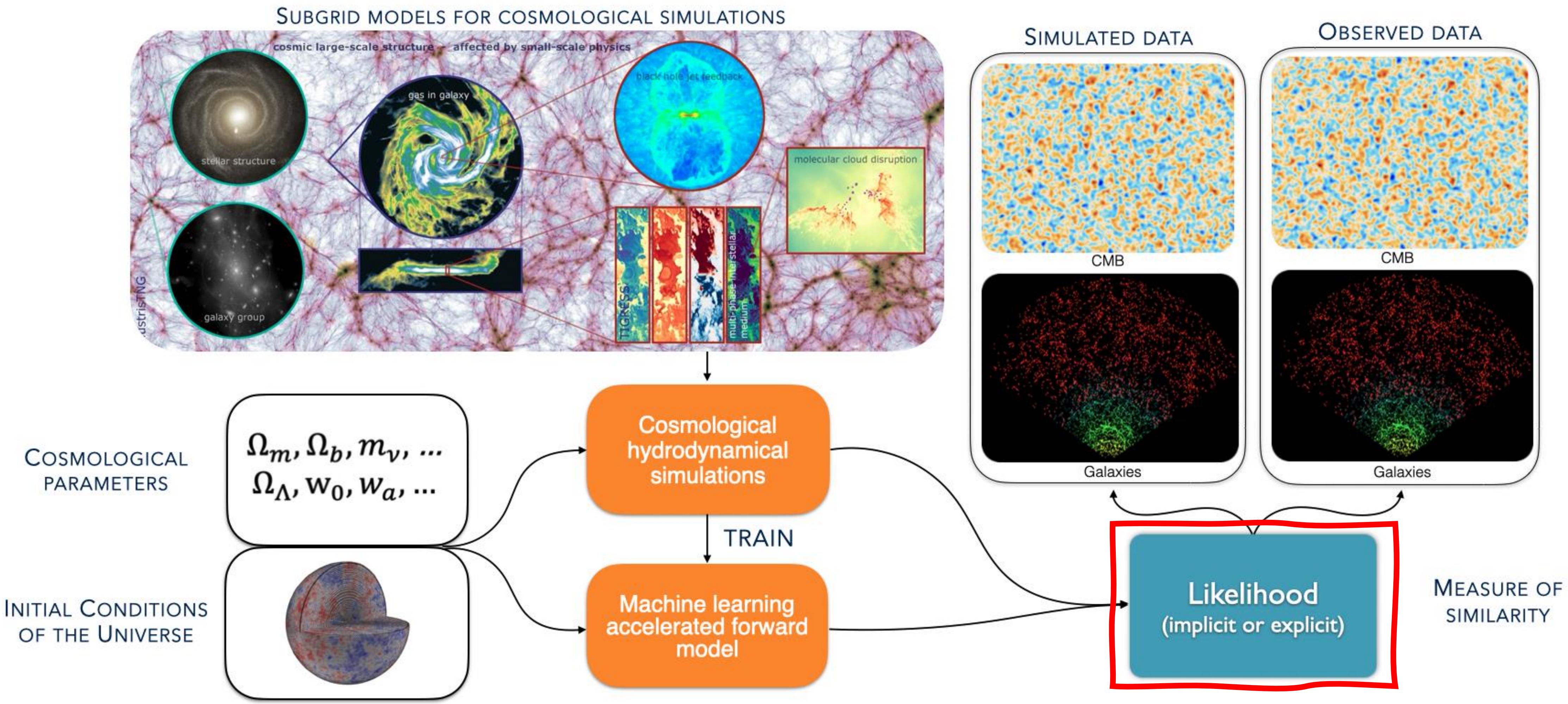
IAP

Columbia University

Cosmic Horizons – July 2026

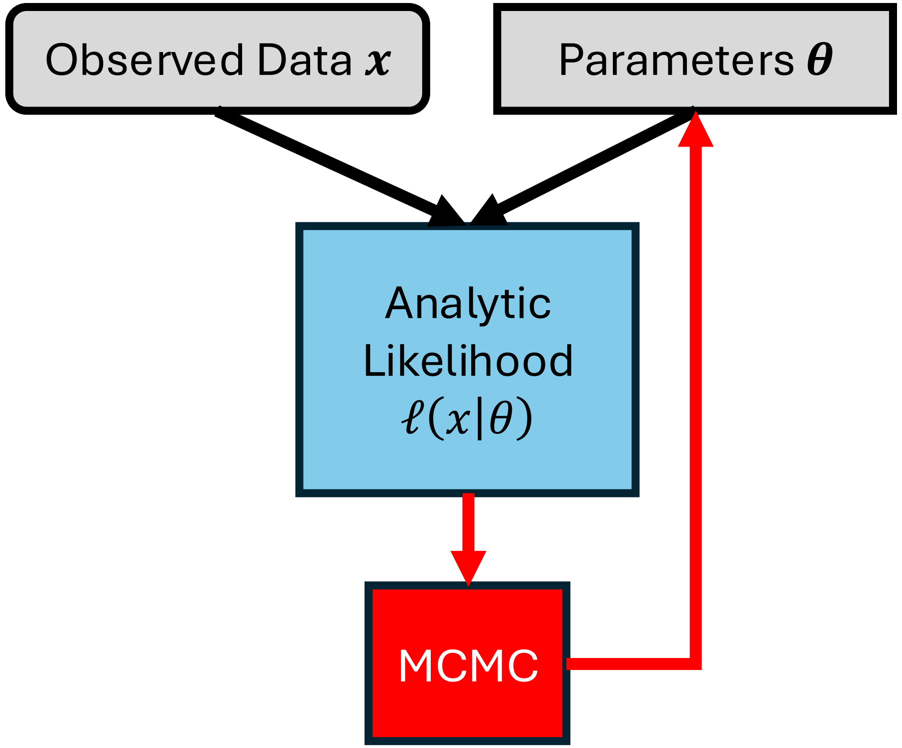
Learning The Universe

learning-the-universe.org

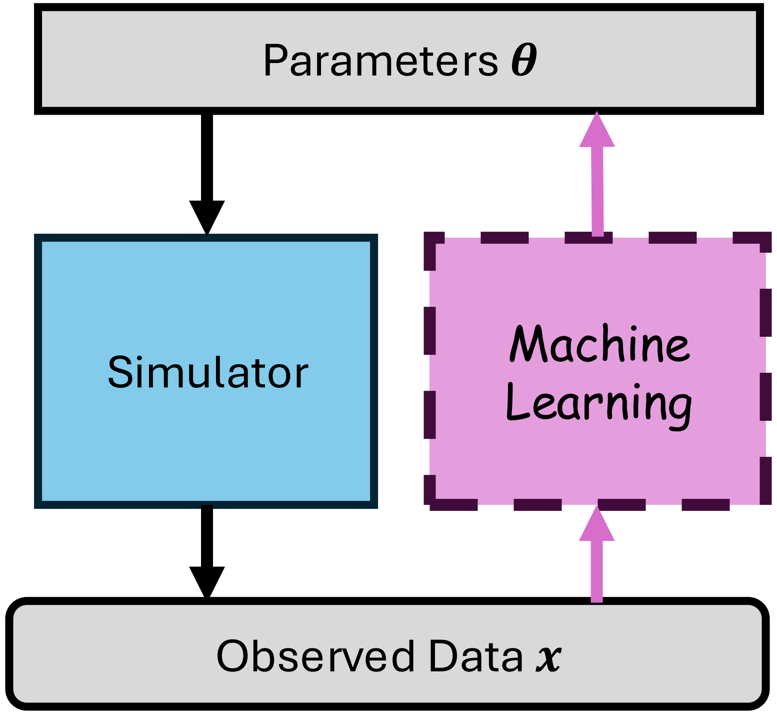


Simon Ding
ChangHoon Hahn
Shirley Ho
Christian Jespersen
Richard Stiskalek
Xiaosheng Zhao
Lucas Makinen
Axel Lapel
Adrian Bayer
Guilhem Lavaux
Benjamin Wandelt
Ce Sui
Matthew Ho
Leander Thiele
Rosa Malandrino
Greg Bryan
Pablo Lemos
Nicolas Chartier
Lucia Perez
Chirag Modi
Deaglan Bartlett
Shivam Pandey
Sammy Sharief
Ana Maria Delgado
Justin Alsing
Anirban Bairagi
Christopher Lovell
Carolina Cuesta-Lazaro
Francisco Villaescusa-Navarro
Laurence Perreault Levasseur

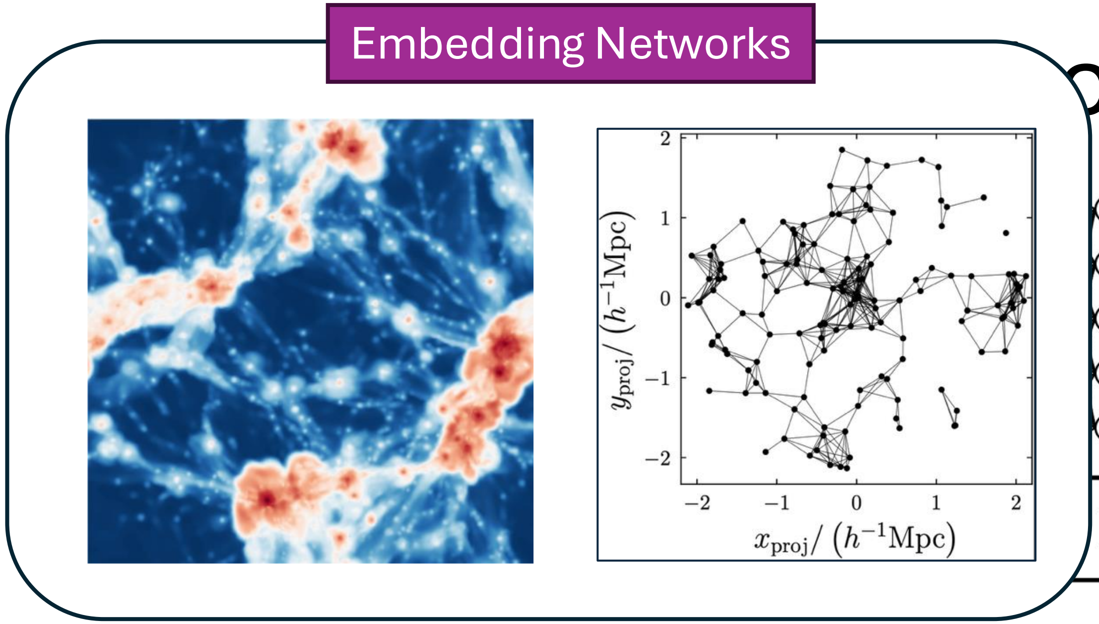
Explicit Likelihood Inference

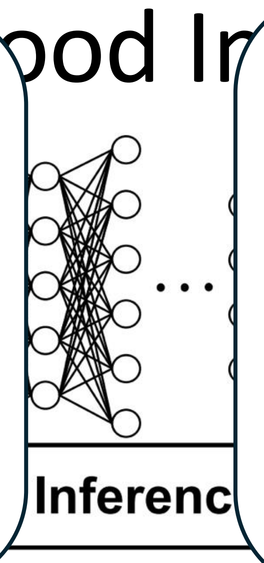


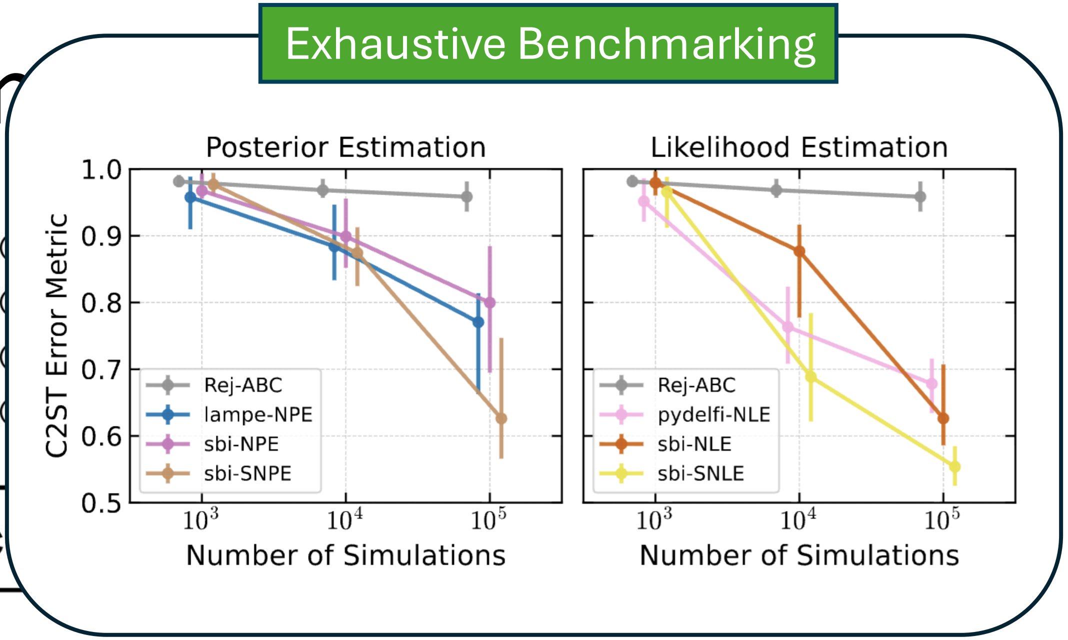
Implicit Likelihood Inference*



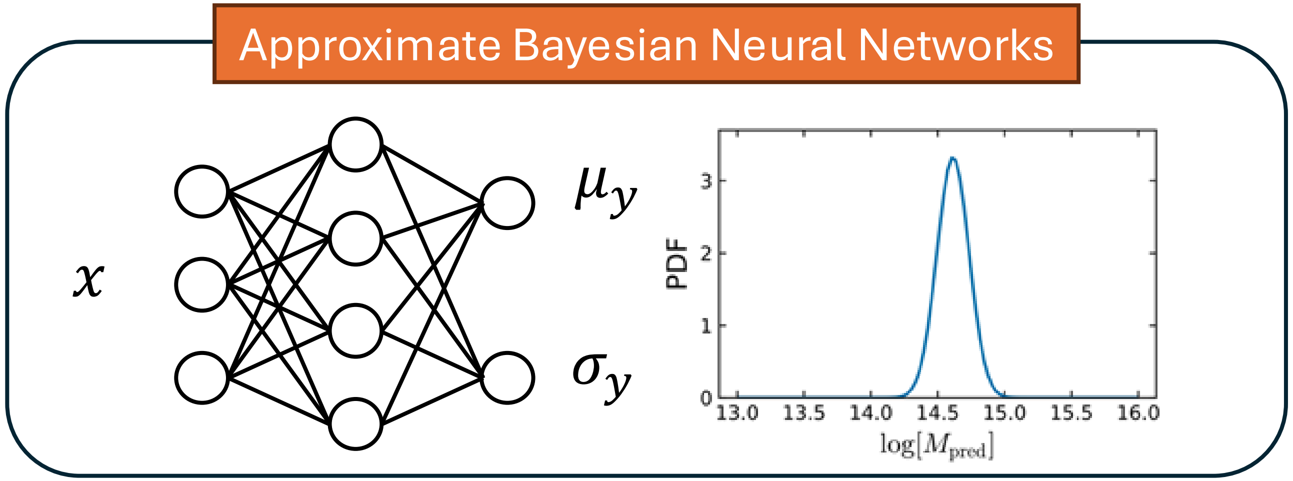
*also known as Simulation-Based Inference (SBI) or Likelihood-Free Inference (LFI)

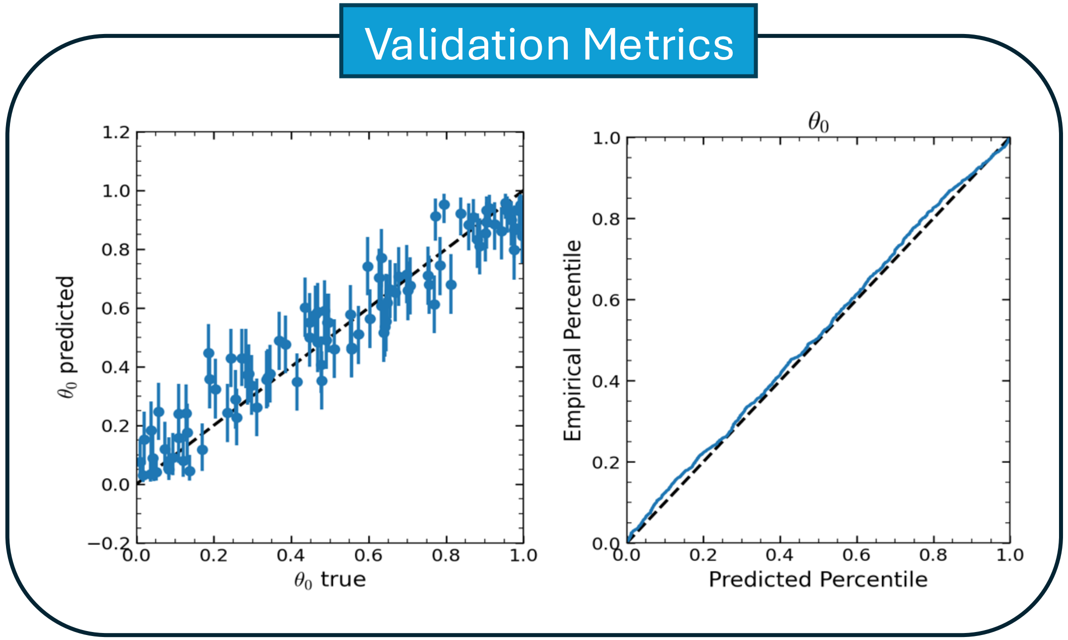






An **all-in-one** framework for implicit inference in astronomy.



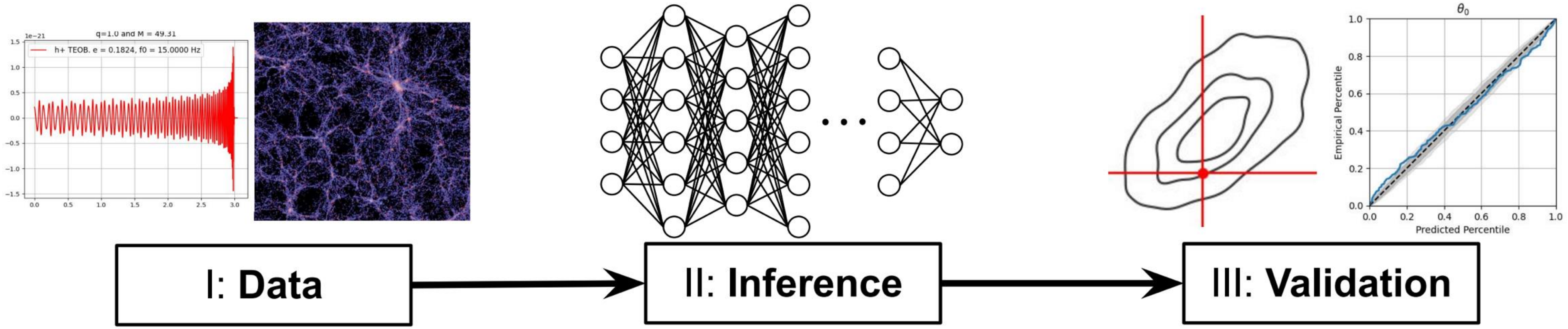


MH + 2024, arxiv:2402.05137

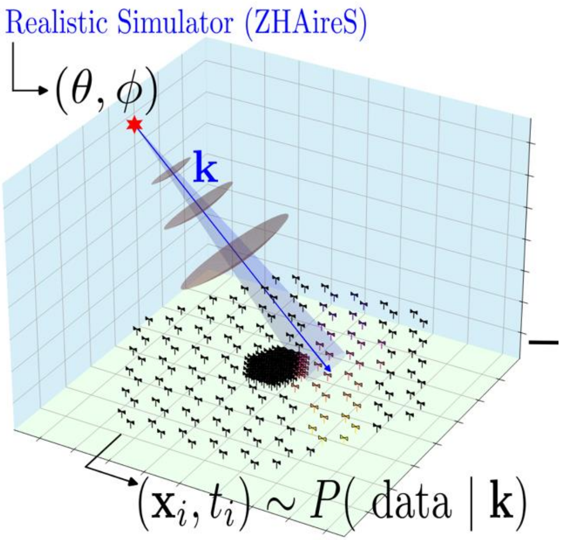
https://github.com/maho3/ltu-ili

*includes undergrads
or junior grad students

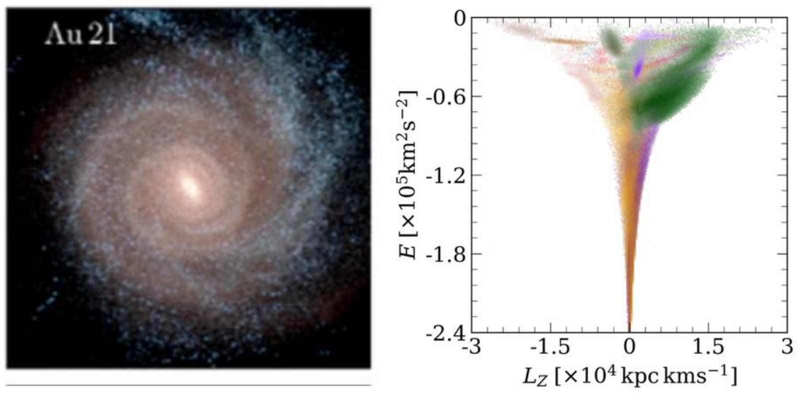
LtU Implicit Likelihood Inference [`ltu-ili`]



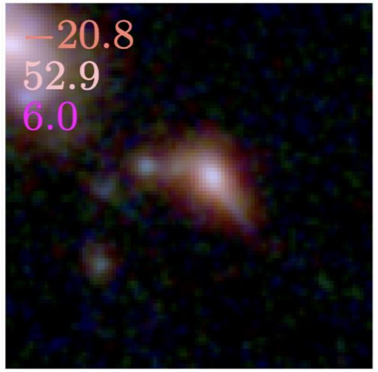
*GRAND Cosmic Rays
arxiv:2508.15991



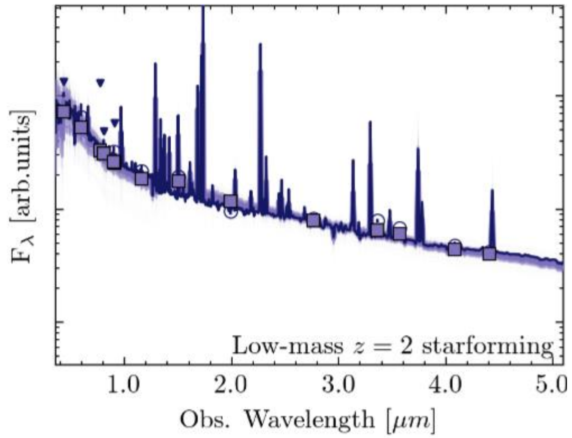
*Labeling Progenitors for
Galactic Archaeology
arxiv:2502.14972

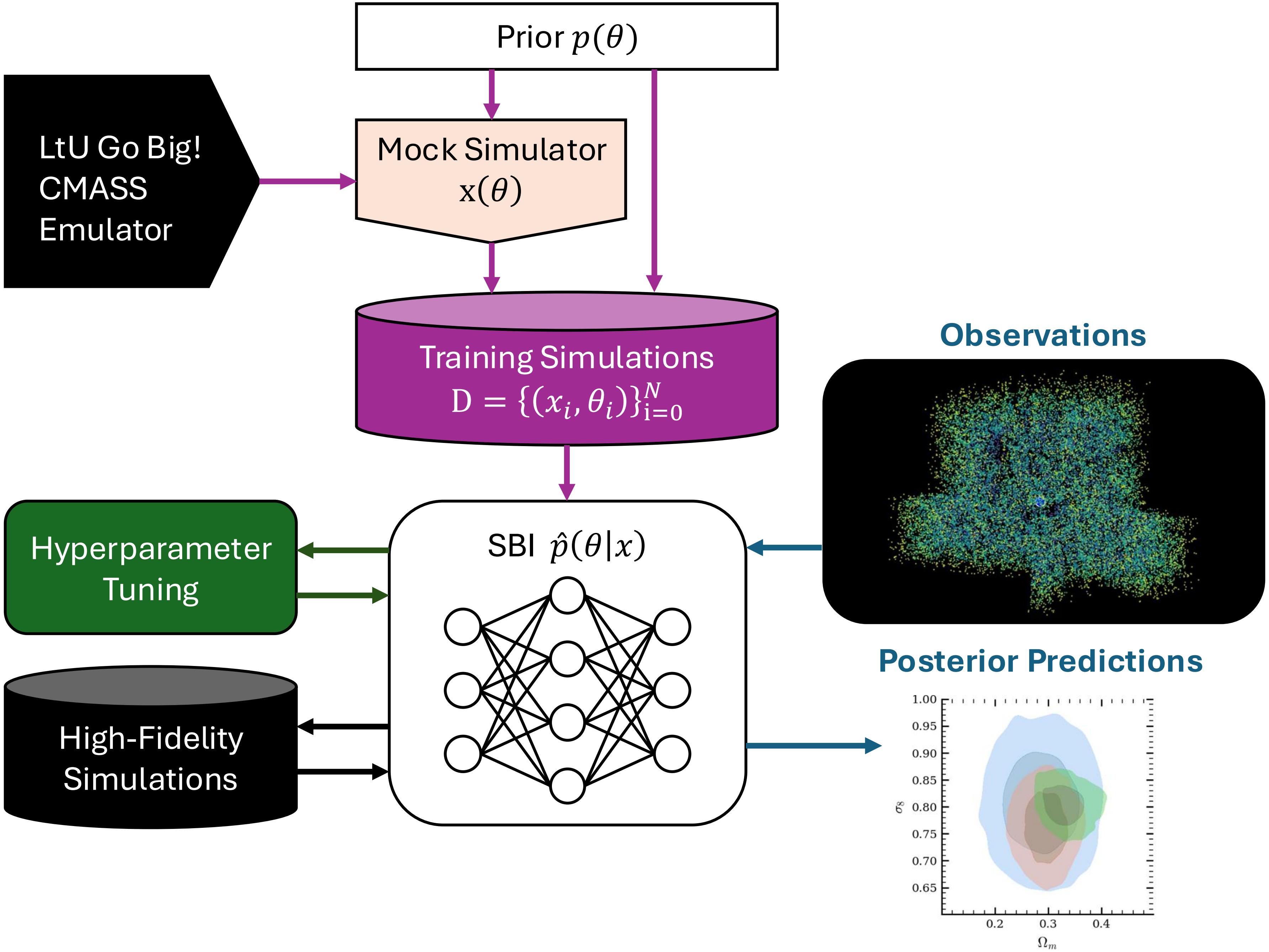


Measuring Reionization
with JWST
arxiv:2405.09720



*Fitting Stellar
Populations of Galaxies
arxiv:2511.10640





Assumptions

My models extract all of the information.

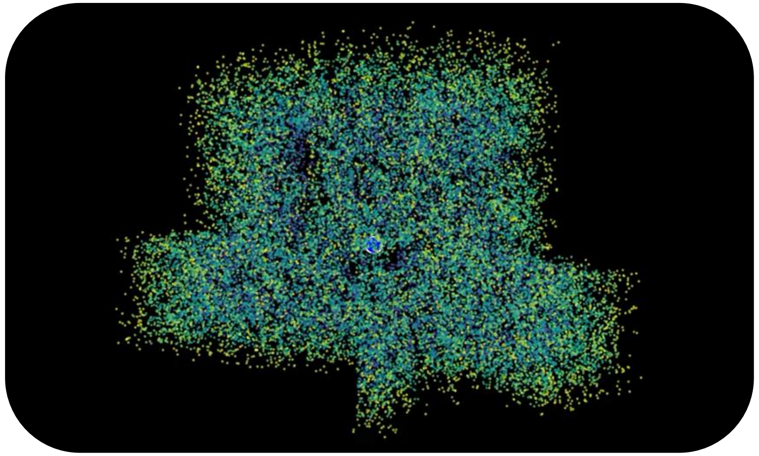
Wide, flat priors are uninformative and well-distributed.

Posterior coverage sufficiently captures uncertainty.

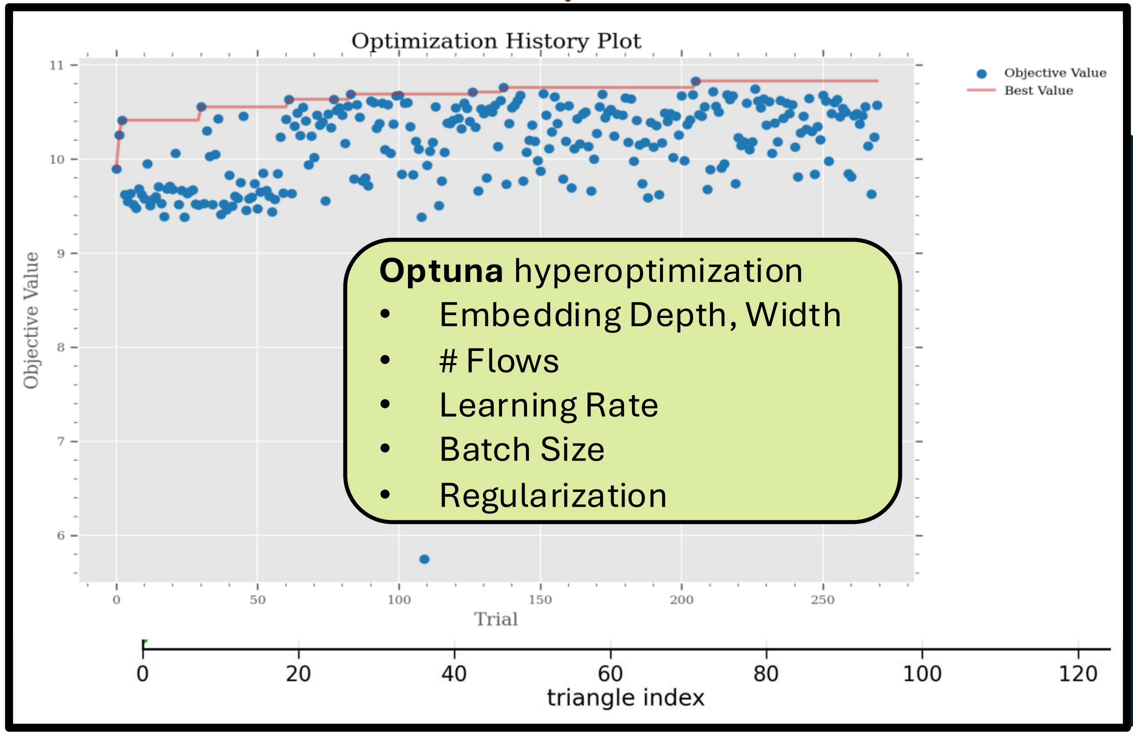
OOD robustness ensures that my model learns the correct physics.

“My models extract all of the information”

Observations



2x Training Data



Optuna hyperoptimization

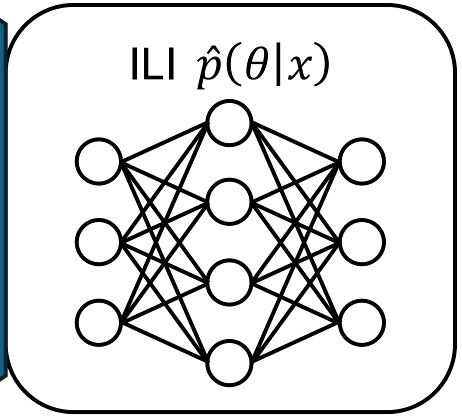
- Embedding Depth, Width
- # Flows
- Learning Rate
- Batch Size
- Regularization

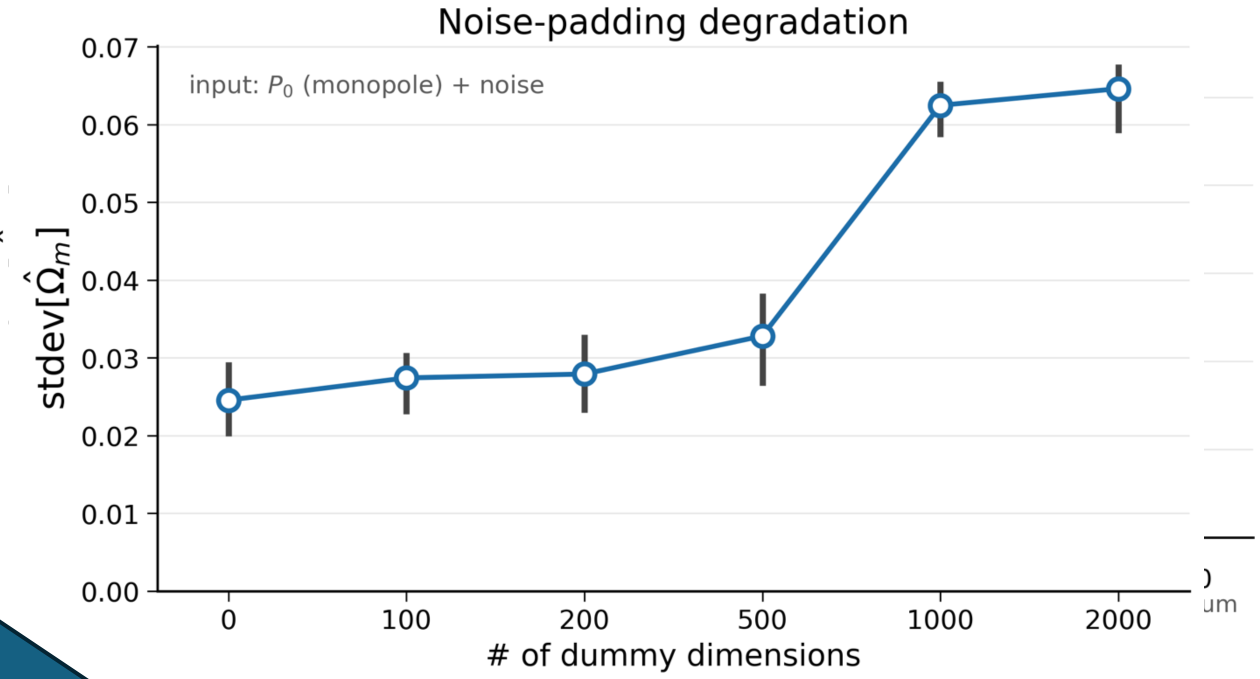
PCA,

FCN,

CNN,

Attn.





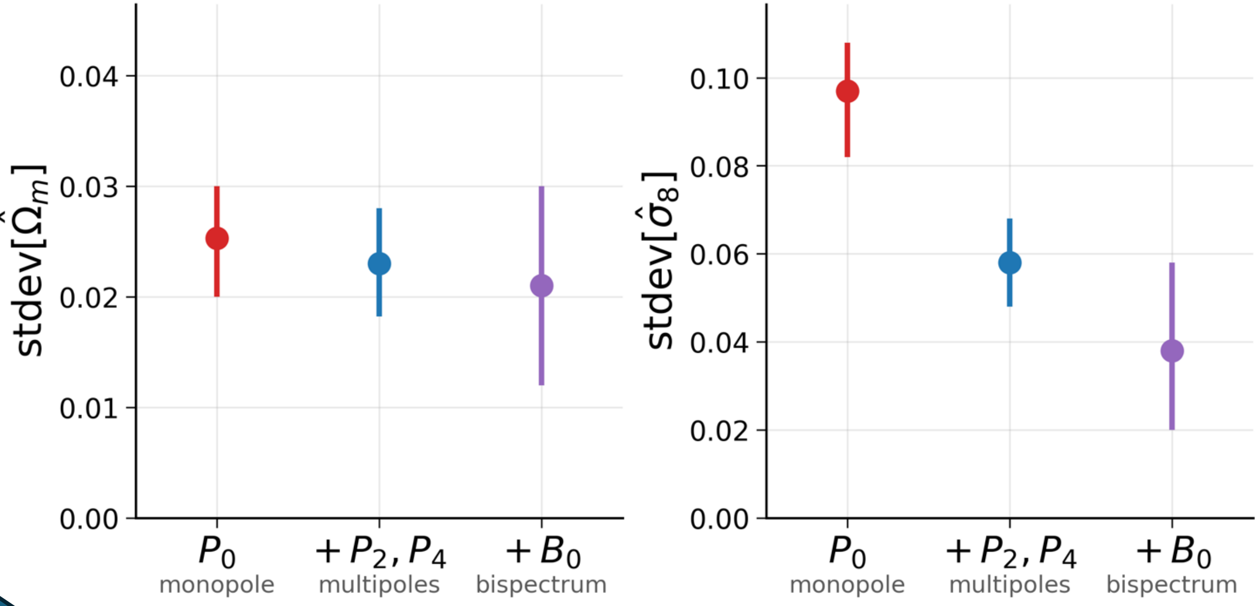
$$(\Omega_m, \sigma_8)$$

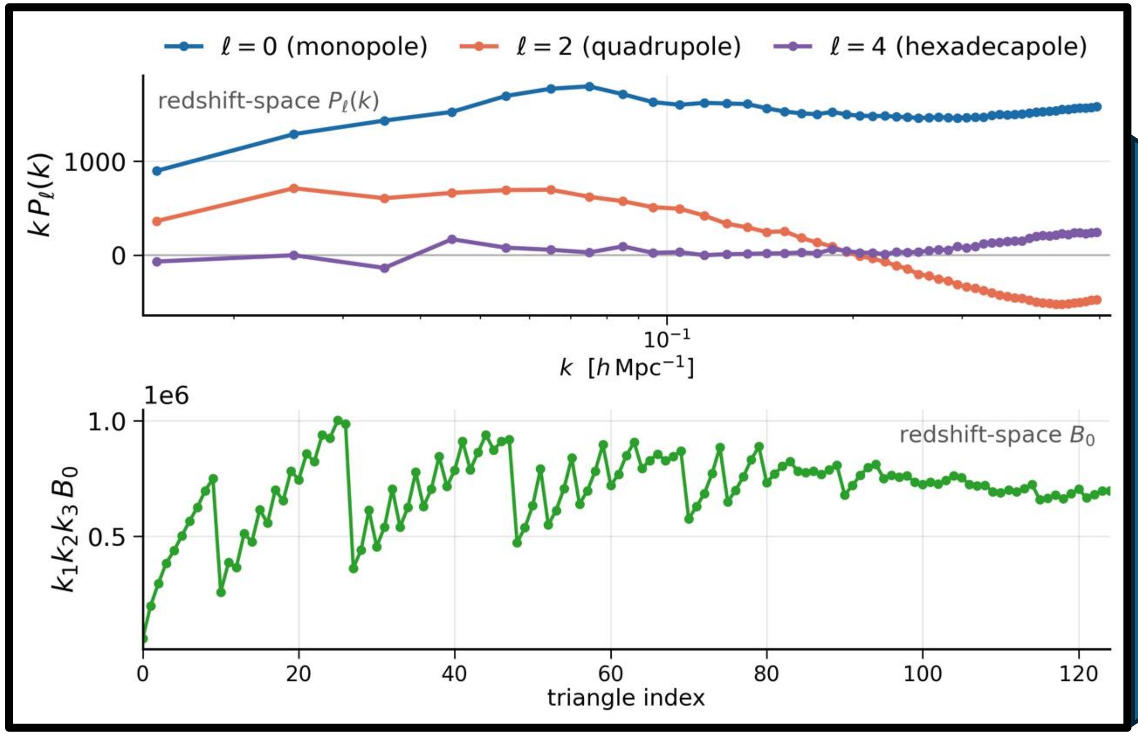
MH, Bartlett, LtU + in prep

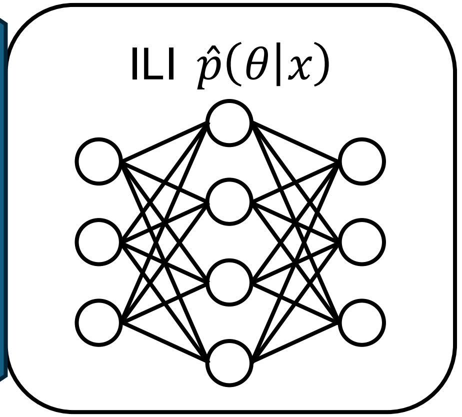
Changes from default **ltu-ili / sbi**

- \- Turn off EarlyStopping & weight reloading
- \+ Optuna hyperparameter tuning
- \+ Noisy data augmentation
- \+ CosineAnnealing learning scheduler
- \+ ‘Funnel’ embedding networks

“My models extract all of the information”





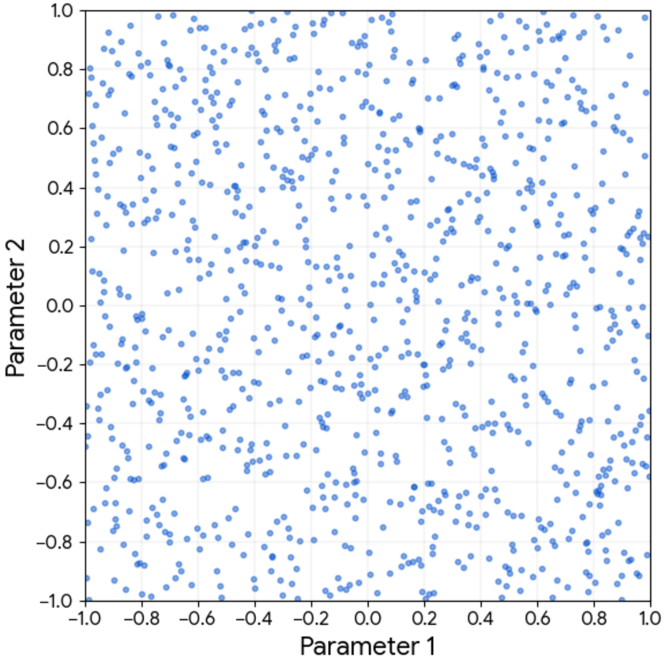


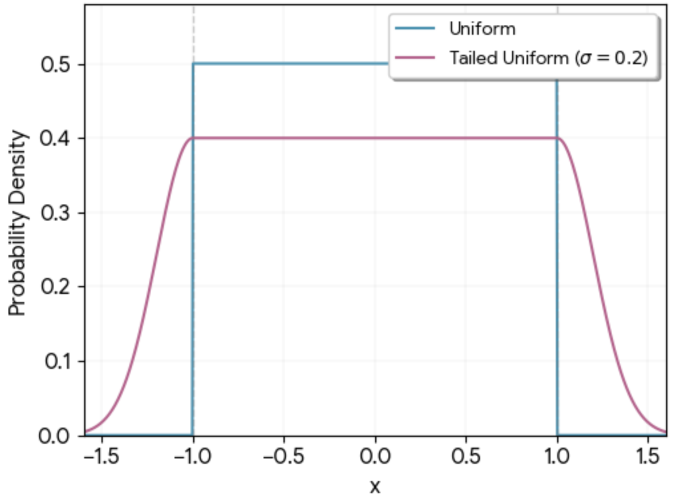
$$(\Omega_m, \sigma_8)$$

MH, Bartlett, LtU + in prep

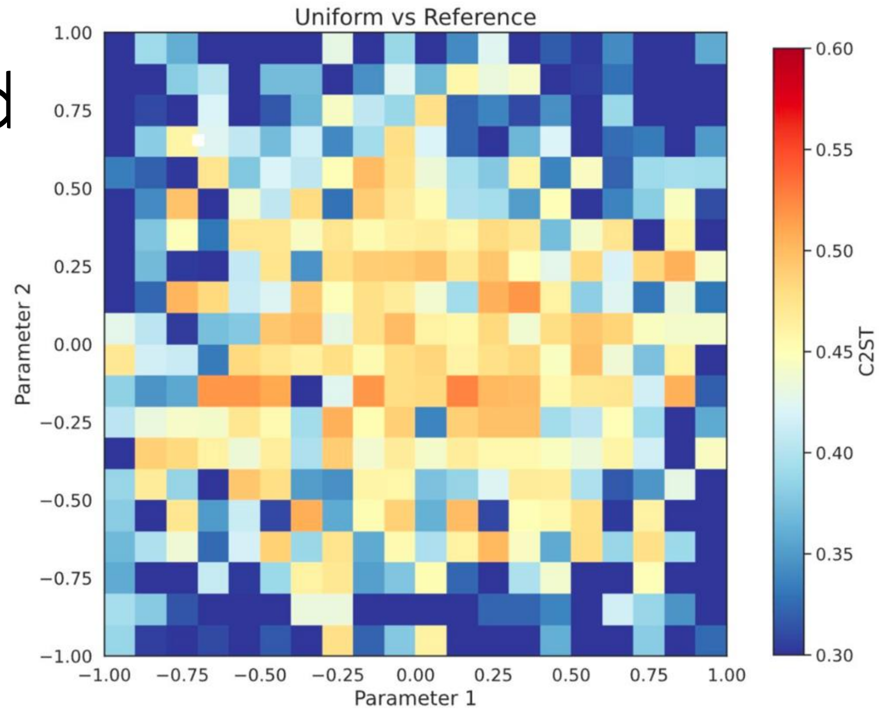
“Wide, flat priors are uninformative and well-distributed."

Uniform Latin Hypercube





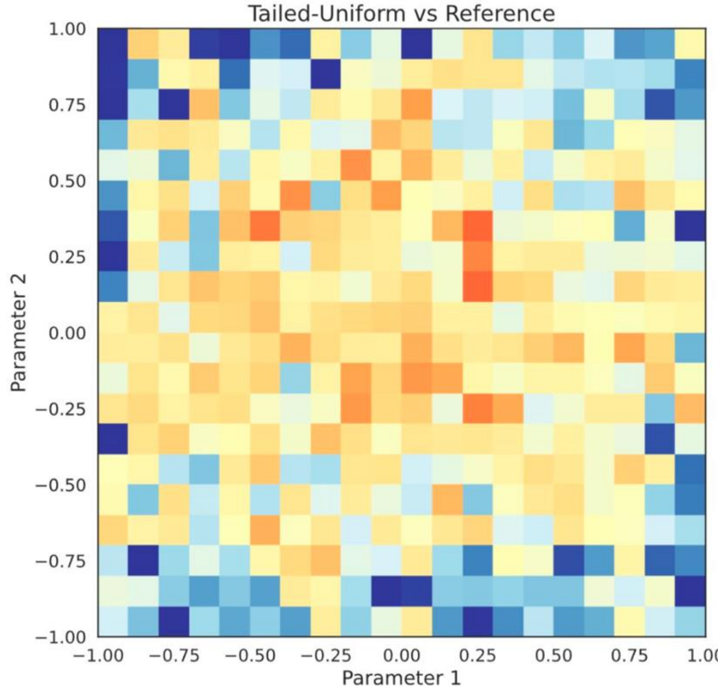
Tirapongprasert, **MH**+2026
arxiv:2601.17120, + longer paper in prep



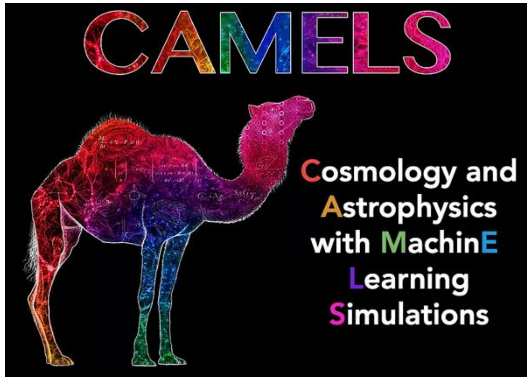
GOOD

C2ST Error Metric

BAD

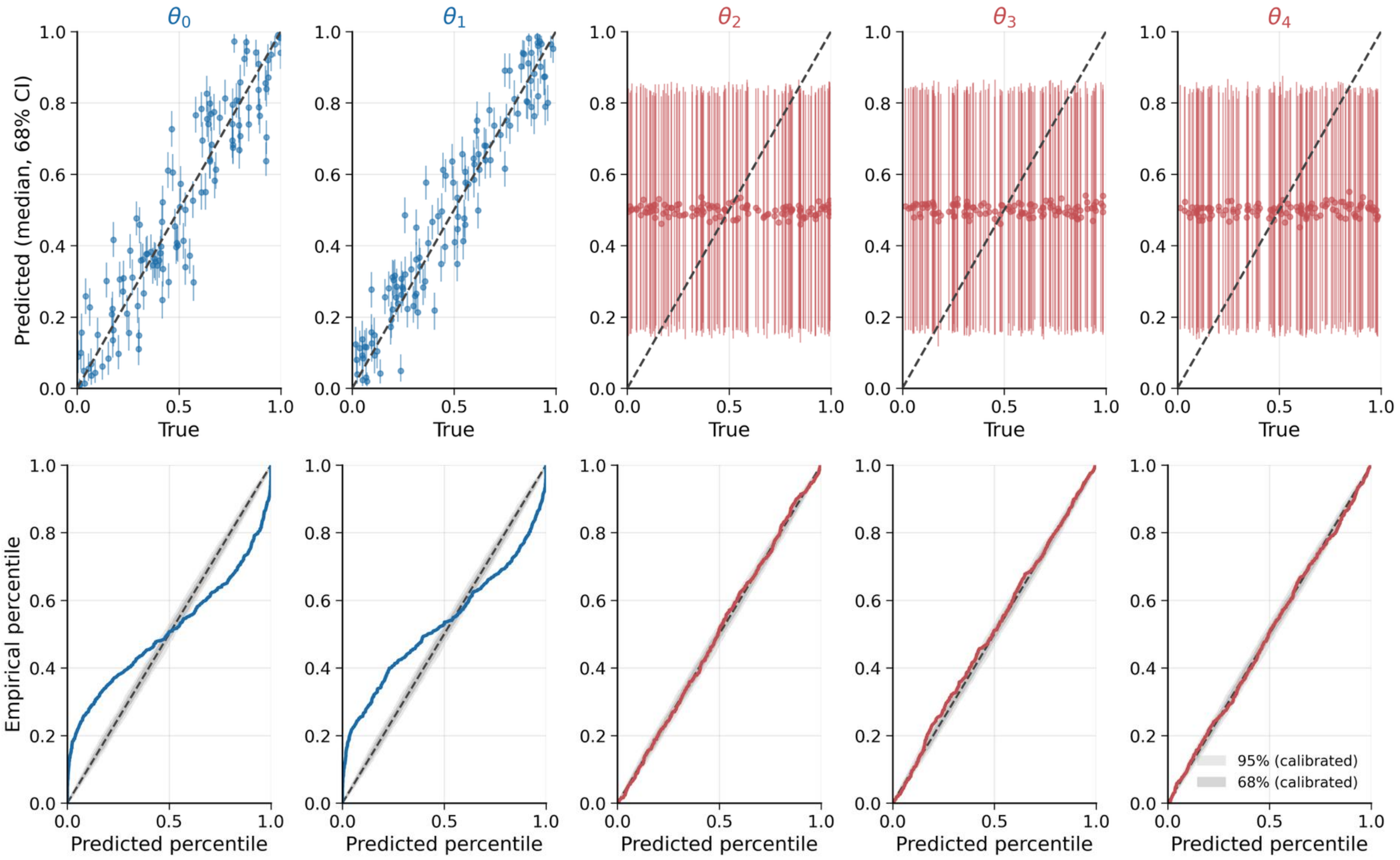


Quijote



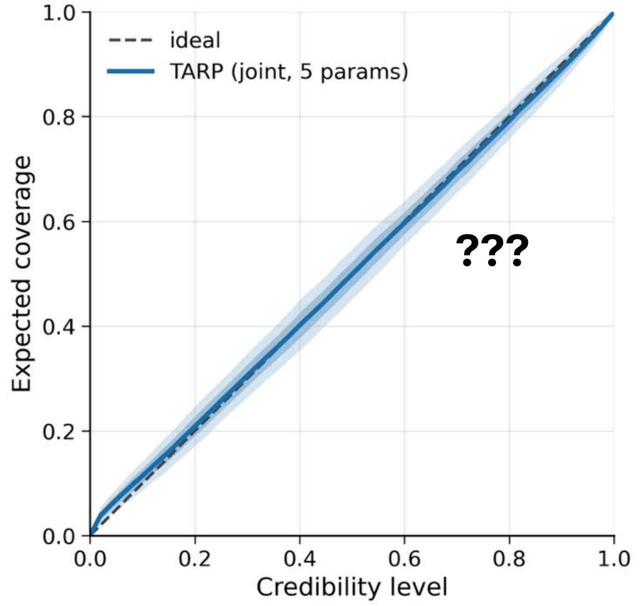
DREAMS

“Posterior coverage sufficiently captures uncertainty.”



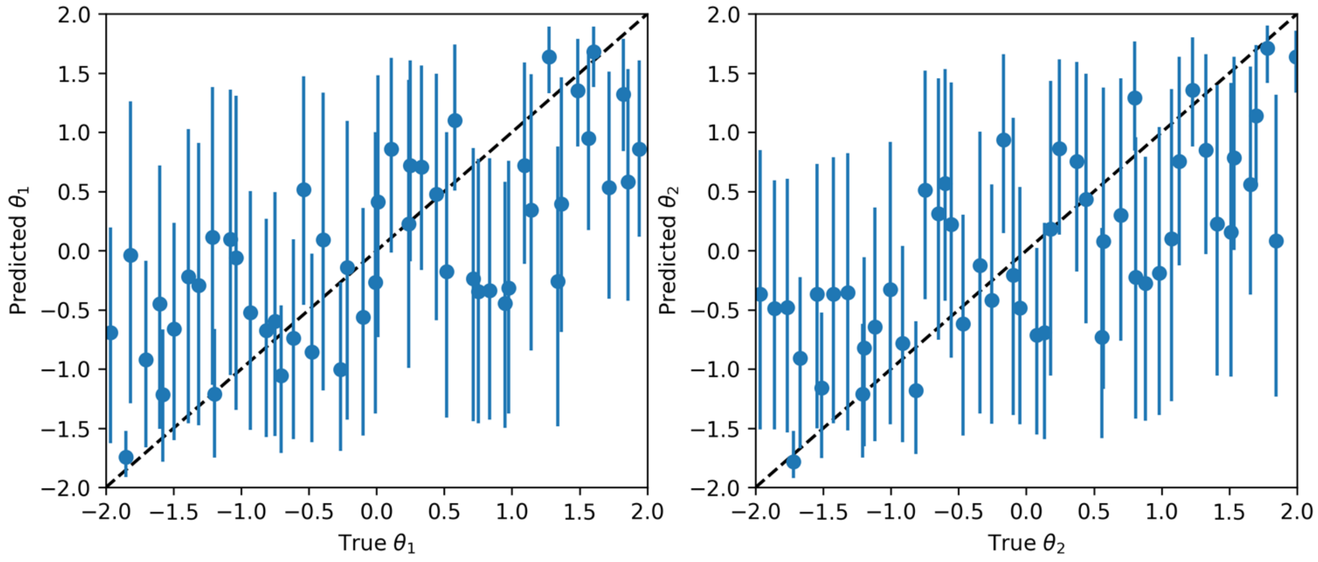
Underconfident!

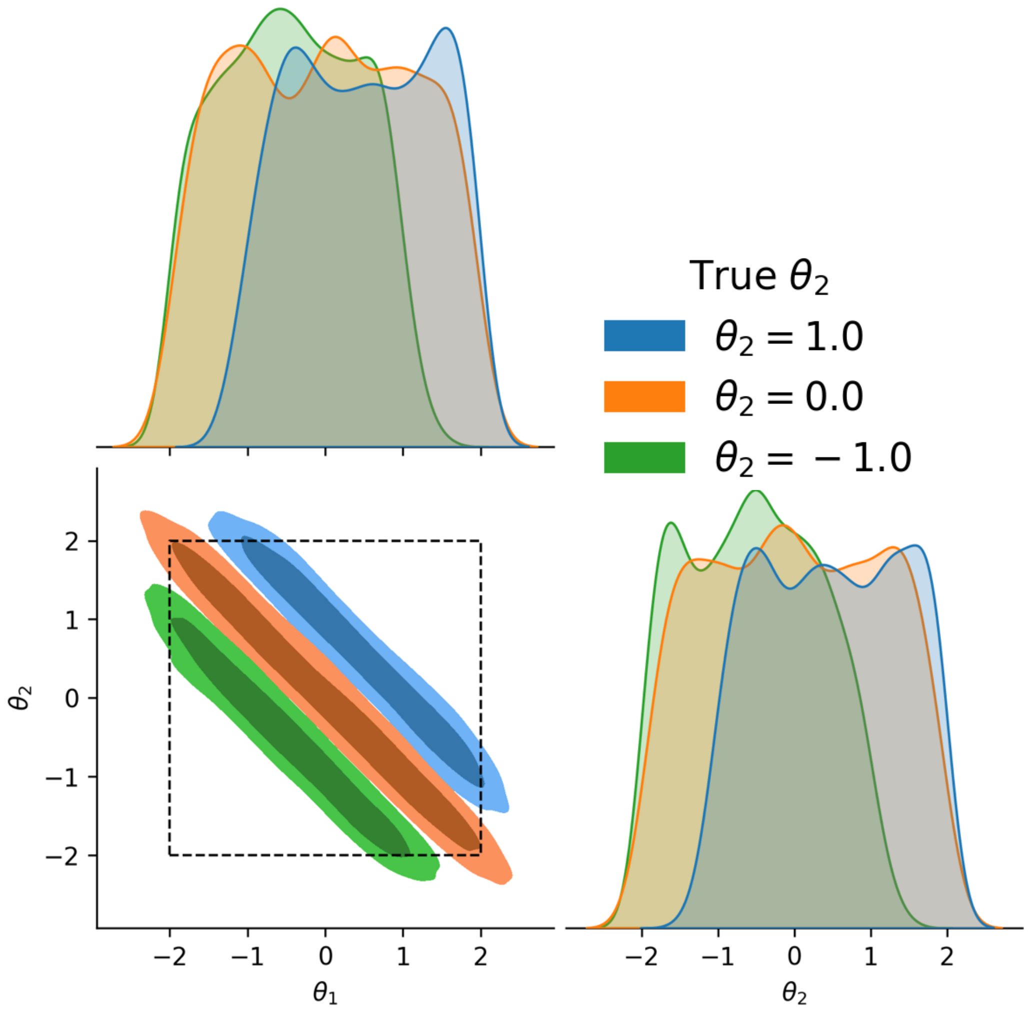
Multivariate Coverage



TARP Lemos+2023
arxiv:2302.03026

"Posterior coverage sufficiently captures uncertainty."





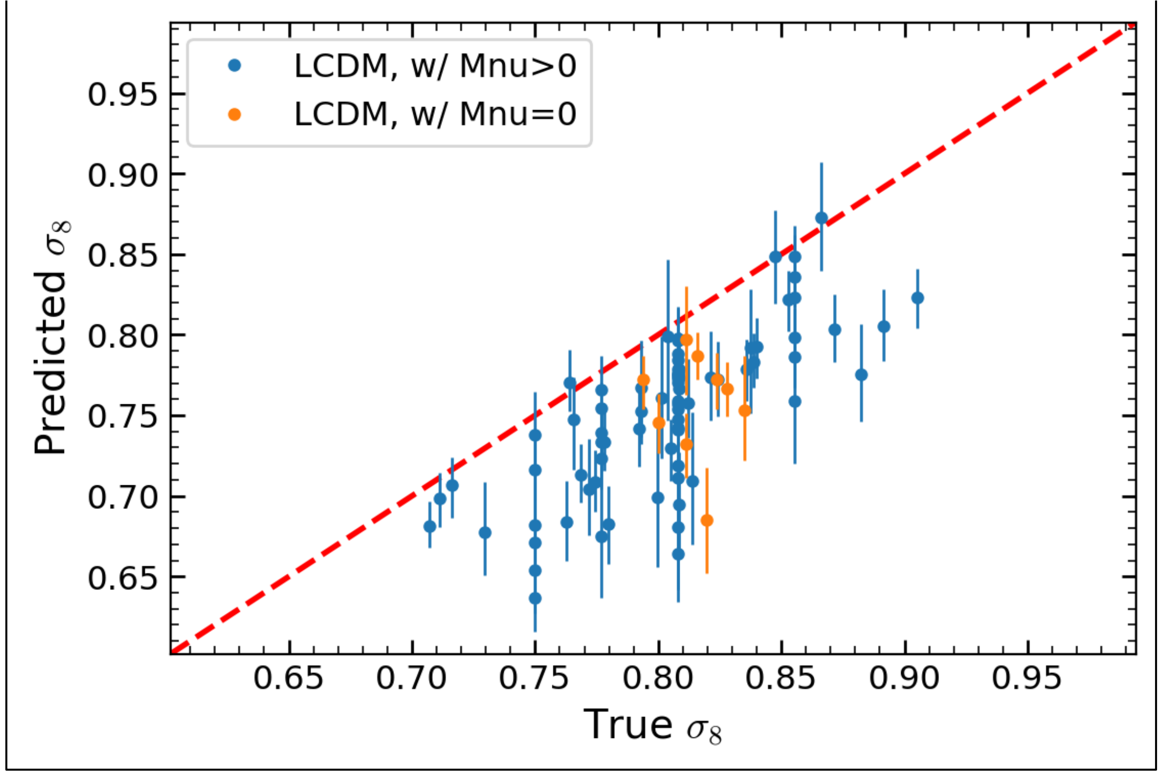
Marginals can hide higher-order degeneracies!

- Degeneracy Distillery (Makinen+2026– arxiv: 2606.23838)
- Degeneracy Detector (Tirapongprasert, **MH**+in prep)

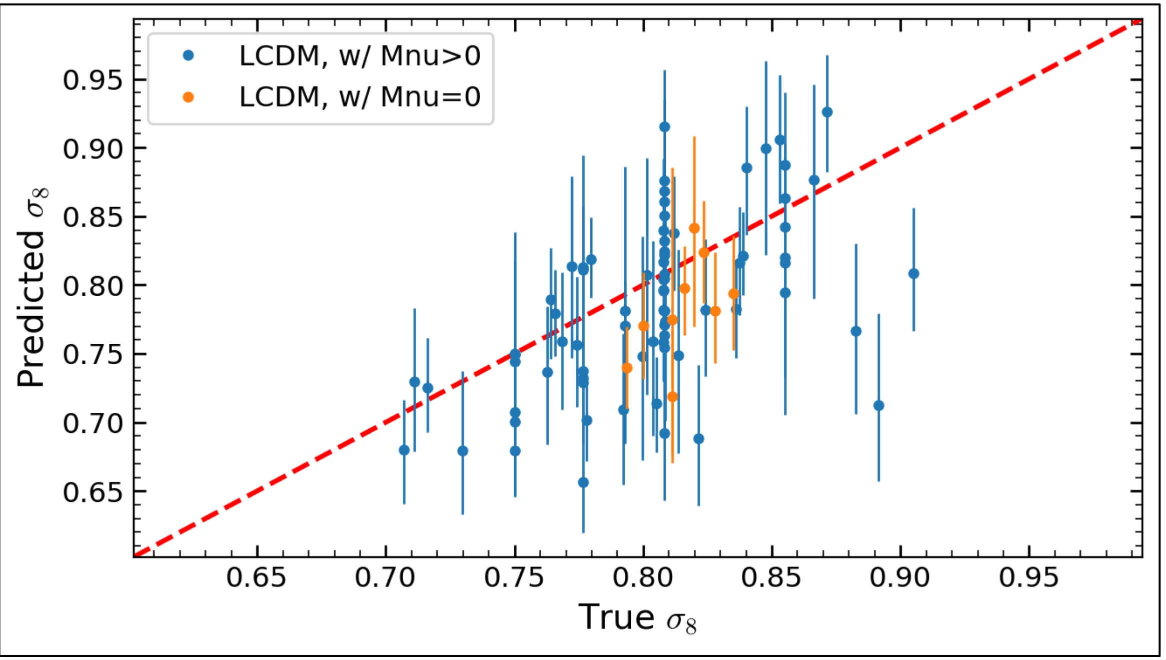
"OOD robustness ensures that my model learns the correct physics."

Train on: Quijote + HOD galaxies
Test on: Abacus + HOD galaxies
Using: $P_0(k)$ w/ RSDs

Even gravity-only simulations disagree with each other!

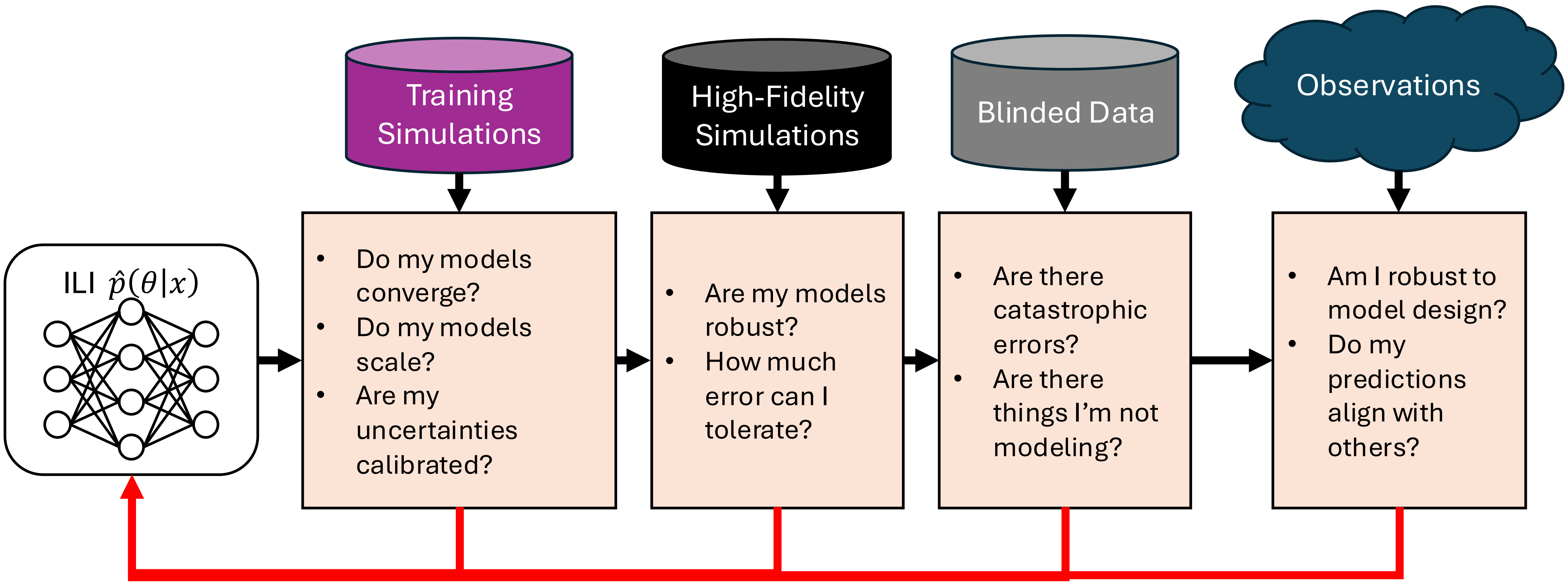


Incl. assembly + velocity bias



MH, Bartlett, LtU + in prep

Building a Pipeline for Observational ILI



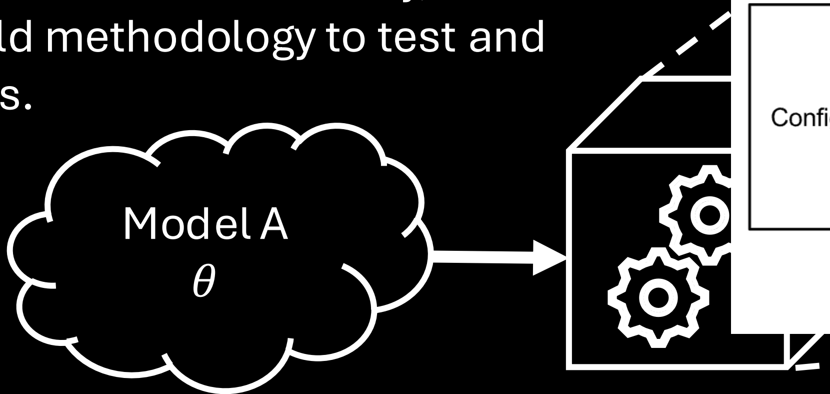
Conclusion

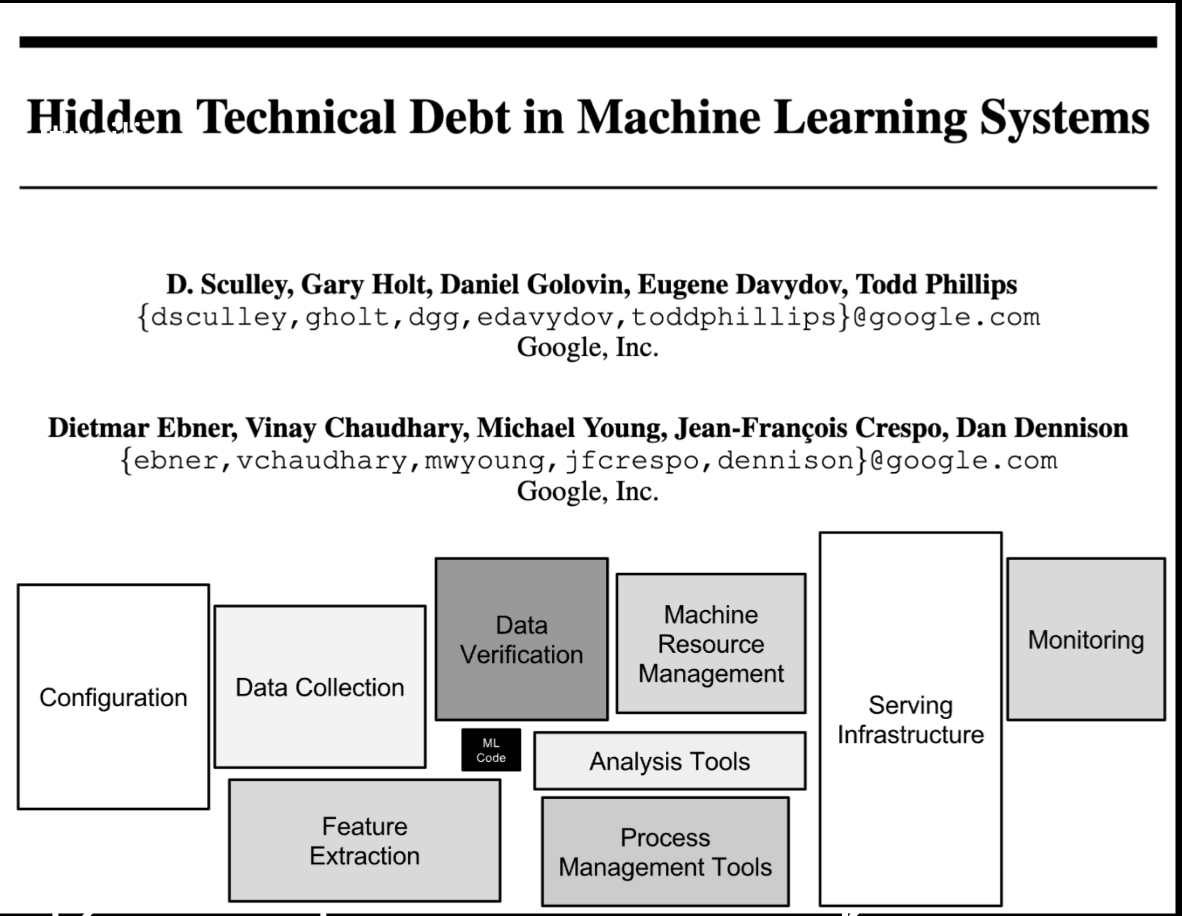
ILI is extremely powerful, but requires careful validation.

Even with best practices, we can't assume that:
- Our models extract all the information.
- Our training suites are optimal.
- Posterior coverage is sufficient.
- Robustness to high-fidelity simulations guarantees observational robustness.

Similar to the explicit likelihood community, we should continue to build methodology to test and guarantee assumptions.

- Model A is consistent with observations
- $P(\theta|x)$ are plausible parameters





Hidden Technical Debt in Machine Learning Systems

D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips
{dsculley,gholt,dgg,edavydov,toddphillips}@google.com
Google, Inc.

Dietmar Ebner, Vinay Chaudhary, Michael Young, Jean-François Crespo, Dan Dennison
{ebner,vchaudhary,mwyoung,jfcrespo,dennison}@google.com
Google, Inc.