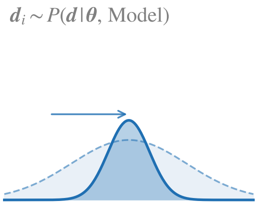
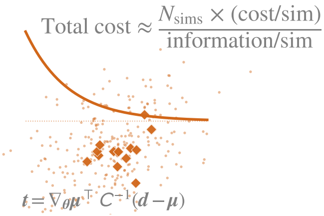
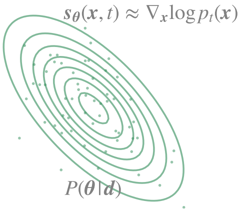






Overcoming the Simulation Bottleneck

Getting enough simulations for next-generation SBI analyses

Maximilian von Wietersheim-Kramsta

mwiet.github.io

SBI in Galaxy Evolution 2026, Kavli Institute for Cosmology, University of Cambridge

24th of June 2026

Institute for Computational Cosmology, Durham University

Institute for Computational Cosmology

Durham Centre for Extragalactic Astronomy

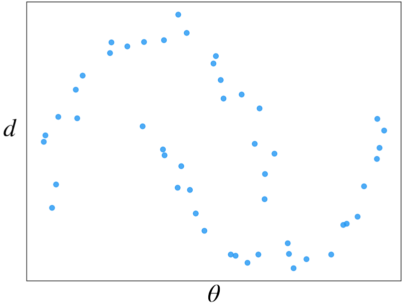
Durham University

The Challenges of Forward Modelling

The bottleneck

- Next-generation surveys (**Euclid, Rubin/LSST**, etc.) turn *statistics-limited* into *systematics- and model-limited* science. In this world, SBI is a natural framework.
- SBI needs **forward simulations**, often **multi-scale** and **high-dimensional**.

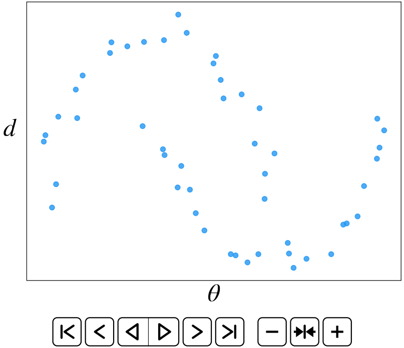
Simulator: $\mathbf{d}_i \sim P(\mathbf{d} \mid \boldsymbol{\theta}, \mathrm{Model})$



The bottleneck

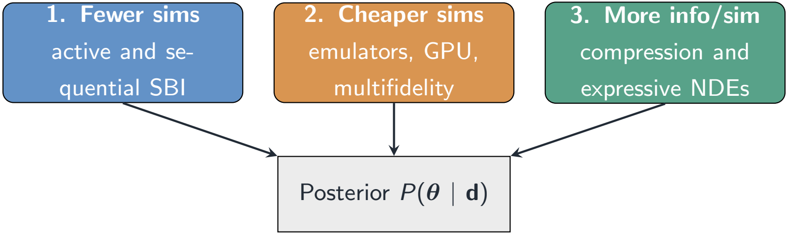
- Next-generation surveys (**Euclid, Rubin/LSST**, etc.) turn *statistics-limited* into *systematics- and model-limited* science. In this world, SBI is a natural framework.
- SBI needs **forward simulations**, often **multi-scale** and **high-dimensional**.
- A standard NPE needs $\sim 10^4$–10^5 sims (Lueckmann et al., 2021), yet one high-fidelity realisation costs 10^3–10^6 **CPU-h**.
  Their product is the bottleneck.

Simulator: $\mathbf{d}_i \sim P(\mathbf{d} \mid \boldsymbol{\theta}, \mathrm{Model})$



What drives the bottleneck?

$$\text{Total cost} \approx \frac{N_{\text{sims}} \times (\text{cost}/\text{sim})}{\text{information}/\text{sim}}$$



Fewer Simulations

Don’t simulate where you don’t need to

- Amortised SBI samples $\boldsymbol{\theta}$ from the prior: most sims land in low-posterior regions and are “wasted”.

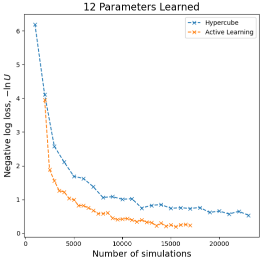
Don't simulate where you don't need to

- Amortised SBI samples $\boldsymbol{\theta}$ from the prior: most sims land in low-posterior regions and are "wasted".
- **Sequential/active SBI**: *simulate* $\rightarrow$ *train* $\rightarrow$ *propose* $\boldsymbol{\theta}$ *near the posterior* $\rightarrow$ *simulate*. Concentrate the budget where the data is informative.

Don't simulate where you don't need to

- Amortised SBI samples $\boldsymbol{\theta}$ from the prior: most sims land in low-posterior regions and are "wasted".
- **Sequential/active SBI**: *simulate* $\rightarrow$ *train* $\rightarrow$ *propose* $\boldsymbol{\theta}$ *near the posterior* $\rightarrow$ *simulate*. Concentrate the budget where the data is informative.

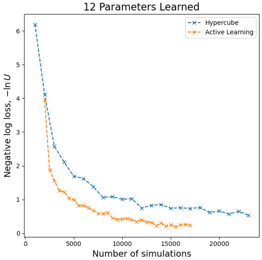
- **Order(s)-of-magnitude fewer sims** for the same posterior.



Lin et al. (2023) with DELFI

Don’t simulate where you don’t need to

- Amortised SBI samples θ from the prior: most sims land in low-posterior regions and are “wasted”.
- **Sequential/active SBI**: *simulate* $\rightarrow$ *train* $\rightarrow$ *propose* θ *near the posterior* $\rightarrow$ *simulate*. Concentrate the budget where the data is informative.

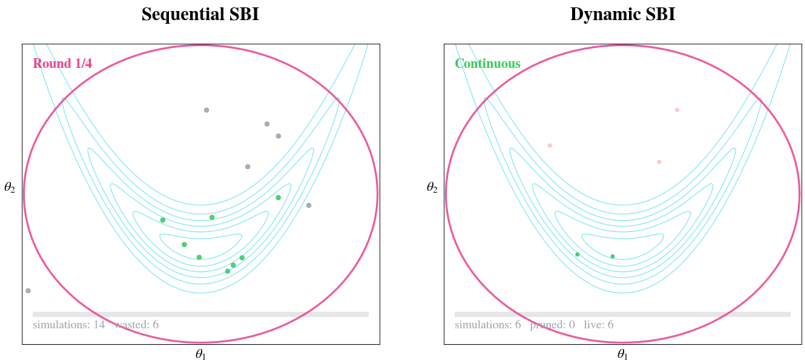


Lin et al. (2023) with DELFI

- **Order(s)-of-magnitude fewer sims** for the same posterior.
- Sequential estimators can be empirically *over-confident* (Hermans et al., 2022).

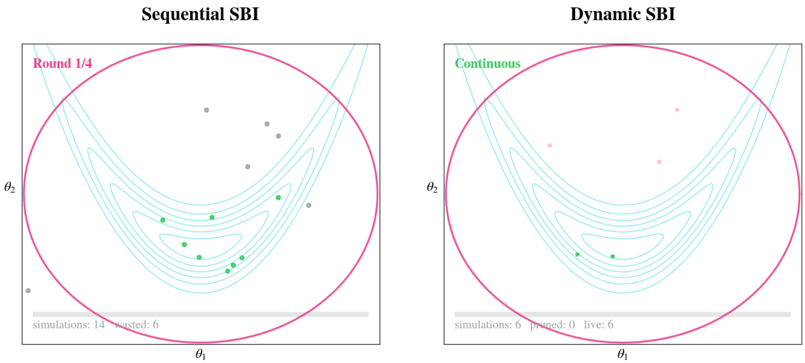
Fewer simulations: more recent implementations

Round-free, asynchronous: *Dynamic SBI* (Lyu et al., 2025) gives sequential gains without discrete rounds and is fully parallelisable.



Fewer simulations: more recent implementations

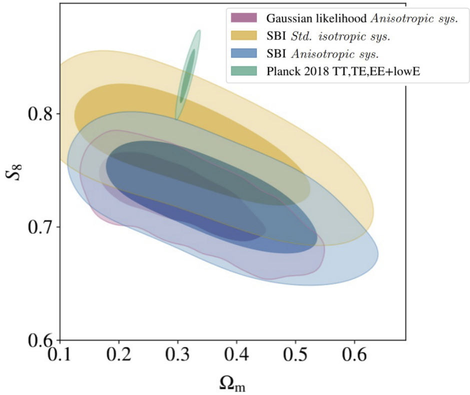
Round-free, asynchronous: *Dynamic SBI* (Lyu et al., 2025) gives sequential gains without discrete rounds and is fully parallelisable.



Round-based vs round-free SBI: wasted vs pruned simulations (Lyu et al., 2025).

Example: Kilo-Degree Survey's cosmic-shear cosmology SBI

$P(\theta|d)$: **Posterior**



Model: ΛCDM

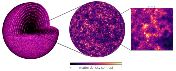
von Wietersheim-Kramsta et al. (2024)

- KiDS-1000 cosmic-shear SBI: an active-learning pipeline recovers the full **12-D posterior with** $\sim 10^4$ **simulations** (Lin et al., 2023).
- **KiDS-SBI**: end-to-end analysis, $S_8 = 0.731 \pm 0.033$ from $\sim$18 000 GLASS forward realisations (von Wietersheim-Kramsta et al., 2024).
- Sequential methods made wide-field weak-lensing SBI tractable. This is now being carried into KiDS-Legacy (Lin et al. in prep.).

Reuse

Pair many cheap sims with few expensive ones

- Most simulators come in a **fidelity ladder**, in the LSS example:
  analytic (virtually free)
  $\rightarrow$ statistical sims (lognormal)
  $\rightarrow$ semi-analytic/DMO
  $\rightarrow$ full N-body/hydro (expensive).



Low-fidelity tier: GLASS log-normal shells (Tessore et al., 2023).

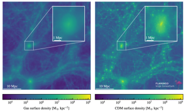
Analytic/log-normal
cheap, many

↓

Semi-analytic/DMO

↓

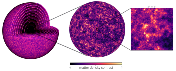
Full N-body/hydro
expensive, few



High-fidelity tier: FLAMINGO (Schaye et al., 2023).

Pair many cheap sims with few expensive ones

- Most simulators come in a **fidelity ladder**, in the LSS example:
  analytic (virtually free)
  $\rightarrow$ statistical sims (lognormal)
  $\rightarrow$ semi-analytic/DMO
  $\rightarrow$ full N-body/hydro (expensive).
- **Multifidelity SBI:** learn the bulk of the structure from the *cheap* tier, then *correct* with a handful of *high-fidelity* runs.



Low-fidelity tier: GLASS log-normal shells (Tessore et al., 2023).

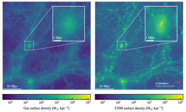
Analytic/log-normal
cheap, many

↓

Semi-analytic/DMO

↓

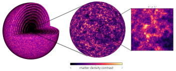
Full N-body/hydro
expensive, few



High-fidelity tier: FLAMINGO (Schaye et al., 2023).

Pair many cheap sims with few expensive ones

- Most simulators come in a **fidelity ladder**, in the LSS example:
  analytic (virtually free)
  $\rightarrow$ statistical sims (lognormal)
  $\rightarrow$ semi-analytic/DMO
  $\rightarrow$ full N-body/hydro (expensive).
- **Multifidelity SBI:** learn the bulk of the structure from the *cheap* tier, then *correct* with a handful of *high-fidelity* runs.
- Implemented as **transfer learning**: pre-train the compressor on low-fidelity sims, **fine-tune** on a small high-fidelity set (Krouglova et al., 2025).



Low-fidelity tier: GLASS log-normal shells (Tessore et al., 2023).

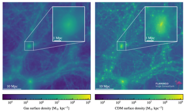
Analytic/log-normal
cheap, many

$\downarrow$

Semi-analytic/DMO

$\downarrow$

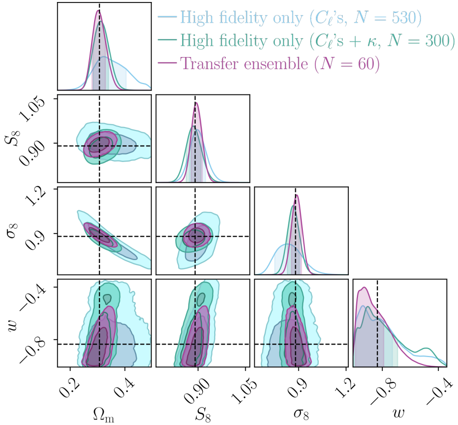
Full N-body/hydro
expensive, few



High-fidelity tier: FLAMINGO (Schaye et al., 2023).

Example: Field-level weak lensing with < 100 simulations

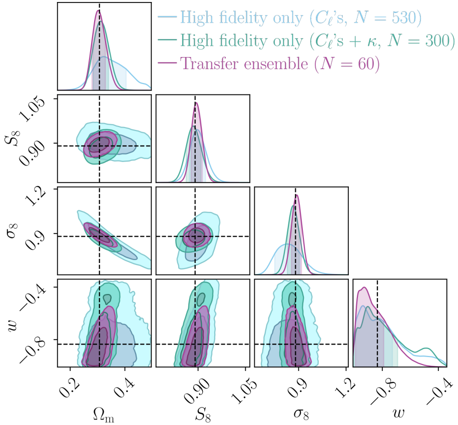
- Field-level inference extracts more than the power spectrum, but demands very high-fidelity sims.



Saoulis et al. (2025)

Example: Field-level weak lensing with < 100 simulations

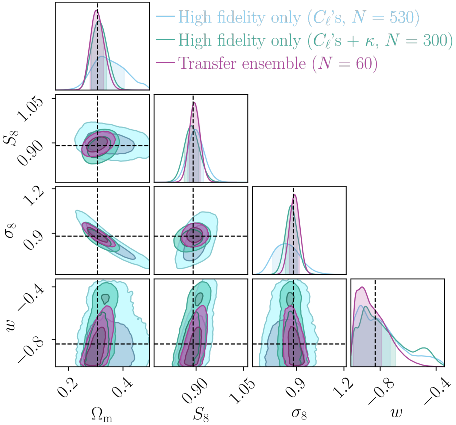
- Field-level inference extracts more than the power spectrum, but demands very high-fidelity sims.
- Pre-train neural compression on fast log-normal **GLASS** shear mocks (Tessore et al., 2023), then fine-tune on a small set of high-fidelity **N-body** sims.



Saoulis et al. (2025)

Example: Field-level weak lensing with < 100 simulations

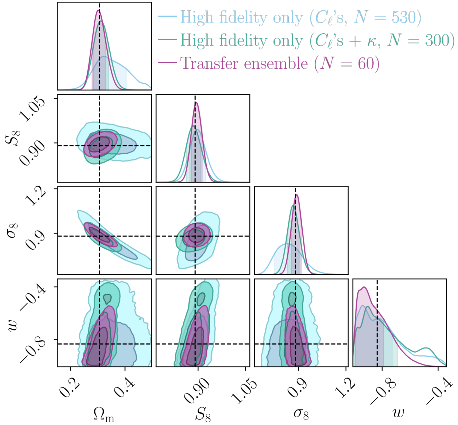
- Field-level inference extracts more than the power spectrum, but demands very high-fidelity sims.
- Pre-train neural compression on fast log-normal **GLASS** shear mocks (Tessore et al., 2023), then fine-tune on a small set of high-fidelity **N-body** sims.
- 60–100 high-fidelity sims give informative, well-calibrated posteriors.
  10 times fewer sims needed (Saoulis et al., 2026)



Saoulis et al. (2025)

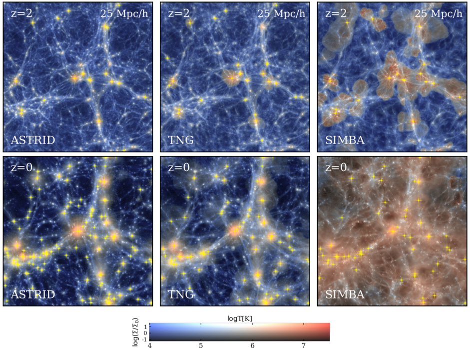
Example: Field-level weak lensing with < 100 simulations

- Field-level inference extracts more than the power spectrum, but demands very high-fidelity sims.
- Pre-train neural compression on fast log-normal **GLASS** shear mocks (Tessore et al., 2023), then fine-tune on a small set of high-fidelity **N-body** sims.
- 60–100 high-fidelity sims give informative, well-calibrated posteriors.
  10 times fewer sims needed (Saoulis et al., 2026)
- **More on this from Alex Saoulis on Thursday!**



Saoulis et al. (2025)

Generalising: robustness across simulation suites



Villaescusa-Navarro et al. (2021); Ni et al. (2023)

- Different hydro suites encode *different* sub-grid physics: discrete, incompatible simulators. One response is a continuous representation across them.
- Continuous baryonic-feedback representations **interpolate between** implementations and marginalise theoretical uncertainty (CAMELS-based; *Ming-Shau Liu, later today*).
- *Multifidelity in model space rather than resolution space.*

Cheaper

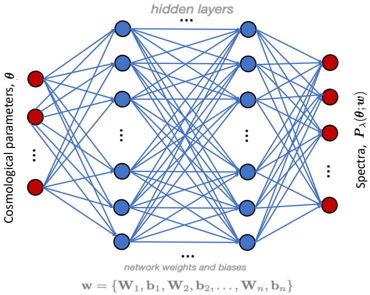
Emulators: bypass the expensive step

- Train a network to *replace* the expensive computation once, then call it free.

- **Cosmology:** e.g. **CosmoPower** emulates Boltzmann power spectra, $\mathcal{O}(10^4)\times$ speed-up (Spurio Mancini et al., 2022).

- **Galaxy evolution:** e.g. **Speculator** emulates SPS spectra and photometry, $\sim 10^3\times$ faster at percent accuracy (Alsing et al., 2020); emulating the galaxy-halo connection and whole hydro fields (*Carolina Cuesta-Lazaro, on Friday*).

- **Caveats:** the emulator's error is now *part of your model*, so propagate it; training costs an up-front sim budget; easy to accidentally be out-of-distribution.



Differentiable and GPU-native forward models

- **GPU + autodiff** (JAX, PyTorch) gets you more throughput, *gradients*, and faster sims.
- Cosmology: **jax-cosmo**, end-to-end differentiable, GPU, exact Fisher forecasts (Campagne et al., 2023).
- Galaxy evolution: e.g. **CERIDWEN** (*Amanda Stoffers, yesterday*), **L-Galaxies AD** (*Andrew Green, on Friday*), **DiffstarPop/Diffsky** (*Alex Alarcon, on Friday*).
- **Caveats:** migrating legacy CPU codes is costly and some physics is *not differentiable*.
- **Advantage:** can speed up statistical compression methods and/or enable other SBI methods such as score-based SBI.

More Information per Simulation

Optimal compression buys you simulations

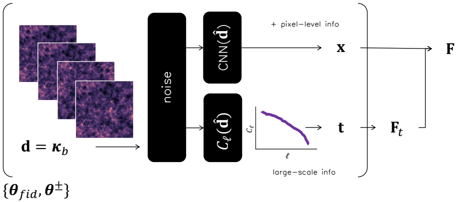
- Fewer, more-informative summaries
  $\Rightarrow$ lower-dimensional density estimation
  $\Rightarrow$ **fewer sims** for the same constraint.

Optimal compression buys you simulations

- Fewer, more-informative summaries
  $\Rightarrow$ lower-dimensional density estimation
  $\Rightarrow$ **fewer sims** for the same constraint.

- **Linear/score:** e.g. MOPED (Heavens et al., 2000), score compression to n_{param} numbers (Alsing and Wandelt, 2018), nuisance-hardened compression (Alsing and Wandelt, 2019), etc.

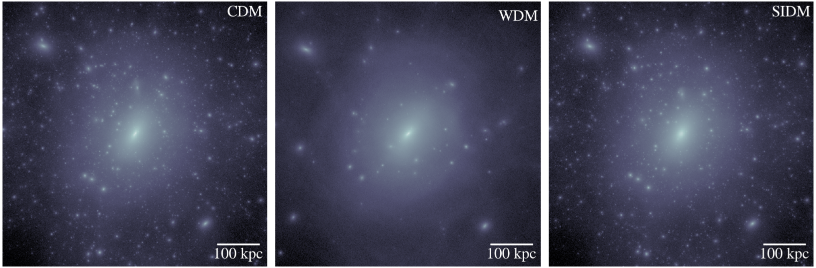
- **IMNNs** (Charnock et al., 2018); **hybrid summaries** fusing analytic and neural statistics via a joint Fisher loss (Makinen et al., 2025), reaching beyond the power spectrum.



Hybrid summaries (Makinen et al., 2025)

Example: Galaxy-Scale Strong Lensing

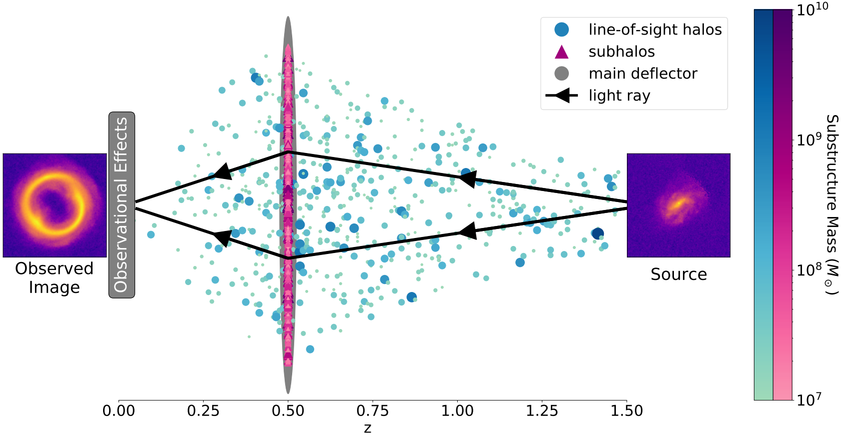
Example: galaxy-scale strong lensing for dark-matter substructure



Bullock and Boylan-Kolchin (2017)

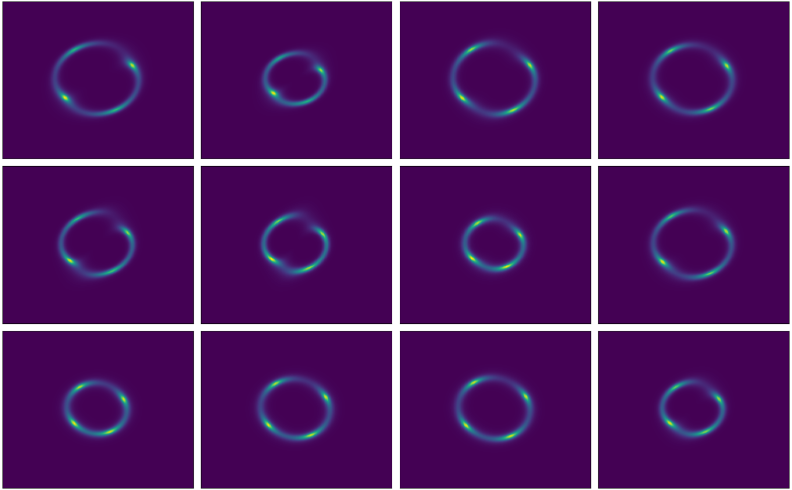
1. Infer cosmology-relevant physics and compare dark-matter models (CDM vs. WDM vs. SIDM) from galaxy-scale lenses.
2. The forward model carries a range of systematics:
   - Physical: macro lens (EPL, multipoles, shear), baryons, line-of-sight haloes, source light.
   - Observational: PSF, correlated noise, selection, resolution (HST, JWST...).
3. Many lenses, so amortised NPE is a natural fit.

The forward model



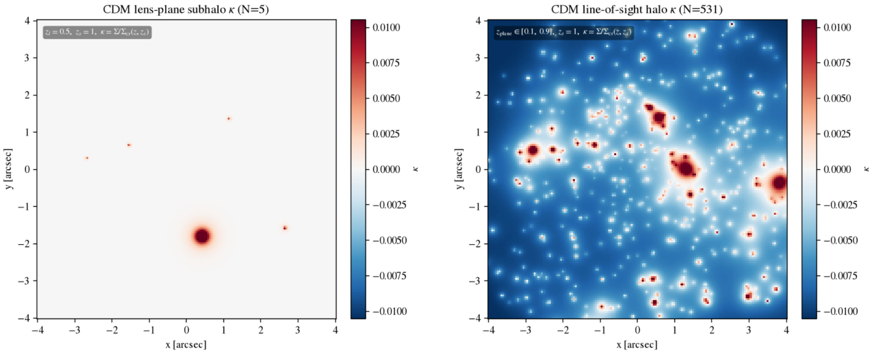
Wagner-Carena et al. (2024)

Drawing lenses from the simulator



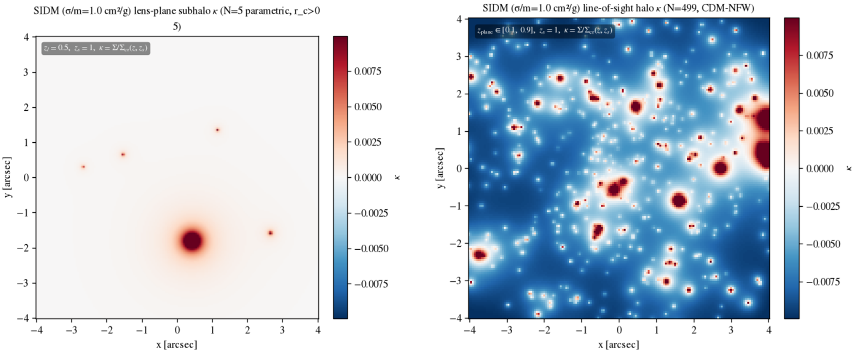
Example simulated lenses spanning the prior. von Wietersheim-Kramsta et al. (in prep.)

Substructure models



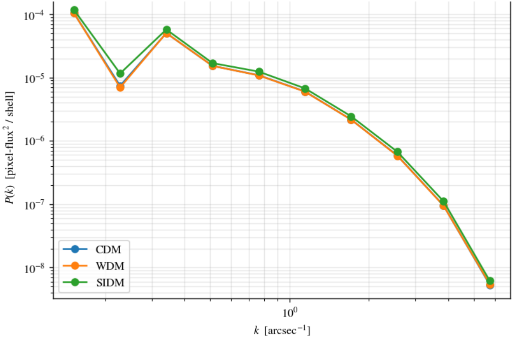
Cold dark matter. von Wietersheim-Kramsta et al. (in prep.)

Substructure models

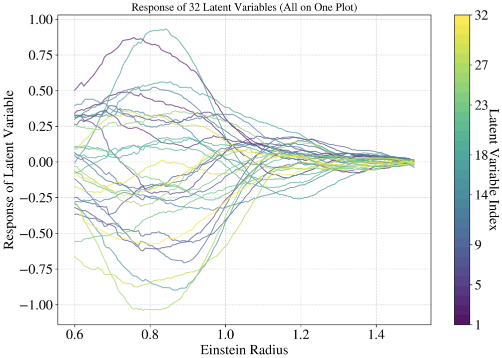


Self-interacting dark matter. von Wietersheim-Kramsta et al. (in prep.)

Compression: Where does the signal hide?



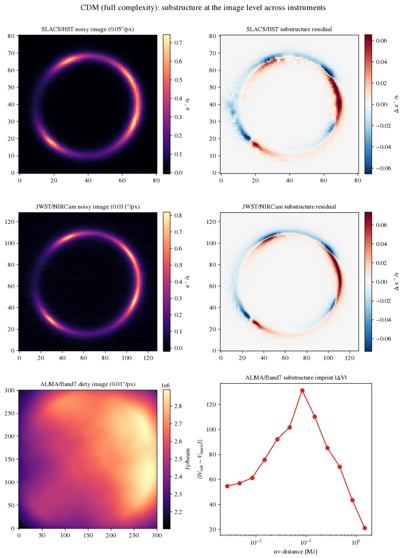
CDM/WDM/SIDM angular $P(k)$ of the lensed image nearly coincide.



Neural compression which incorporates the symmetries of the lens equation

Lessons applied here

- *Cheaper:* whole simulator is JIT-compiled.
- *More info/sim:* from $P(k)$ to CNN/Lensformer/hybrid compression of the arc.
- *Fewer:* sequential SBI.



One lens through HST, JWST and ALMA, with subhalo residual maps.

Conclusions

Which combination, when?

If your bottleneck is...	Use...	Watch out for...
Each simulation very expensive	Emulators; Multifidelity + transfer learning	emulator-error budget; low-fidelity bias
Huge, high-dim parameter space	Active/sequential SBI	over-confidence, so run coverage tests
Repeated, structured forward model	Emulators; GPU + differentiable models	migration cost; non-diff. components
Tiny, high-dim signal	IMNN/hybrid compression	misspecification, OOD

Outlooks

1. **Non-differentiable physics** in differentiable pipelines (nebular models, radiative transfer): surrogate, or learn around it?

Outlooks

1. **Non-differentiable physics** in differentiable pipelines (nebular models, radiative transfer): surrogate, or learn around it?

2. **Propagating emulator and low-fidelity error**: when does multifidelity quietly bias us?

Outlooks

1. **Non-differentiable physics** in differentiable pipelines (nebular models, radiative transfer): surrogate, or learn around it?

2. **Propagating emulator and low-fidelity error**: when does multifidelity quietly bias us?

3. **Shared infrastructure:** a community data challenge for SBI in galaxy evolution (*Luca Tortorelli, later today*); common tooling (`sbi`, GLASS, pop-cosmos, GalSBI, Synference).

Outlooks

1. **Non-differentiable physics** in differentiable pipelines (nebular models, radiative transfer): surrogate, or learn around it?

2. **Propagating emulator and low-fidelity error**: when does multifidelity quietly bias us?

3. **Shared infrastructure:** a community data challenge for SBI in galaxy evolution (*Luca Tortorelli, later today*); common tooling (`sbi`, GLASS, pop-cosmos, GalSBI, Synference).

4. **Amortise across problems**, not just data.

Outlooks

1. **Non-differentiable physics** in differentiable pipelines (nebular models, radiative transfer): surrogate, or learn around it?
2. **Propagating emulator and low-fidelity error**: when does multifidelity quietly bias us?
3. **Shared infrastructure:** a community data challenge for SBI in galaxy evolution (*Luca Tortorelli, later today*); common tooling (`sbi`, GLASS, pop-cosmos, GalSBI, Synference).
4. **Amortise across problems**, not just data.

The bottleneck is real but increasingly addressable, by combining and validating fewer, cheaper, more-informative simulations.

Thank you!

References

J. Alsing and B. Wandelt. Nuisance hardened data compression for fast likelihood-free inference. *MNRAS*, 2019.

J. Alsing, H. Peiris, J. Leja, et al. Speculator: Emulating stellar population synthesis for fast and accurate galaxy spectra and photometry. *ApJS*, 2020.

Justin Alsing and Benjamin Wandelt. Generalized massive optimal data compression. *MNRAS*, 476(1):L60–L64, May 2018. doi: 10.1093/mnrasl/sly029.

James S. Bullock and Michael Boylan-Kolchin. Small-Scale Challenges to the ΛCDM Paradigm. *ARA&A*, 55(1):343–387, August 2017. doi: 10.1146/annurev-astro-091916-055313.

J.-E. Campagne et al. Jax-cosmo: An end-to-end differentiable and gpu accelerated cosmology library. *The Open Journal of Astrophysics*, 2023.

Tom Charnock, Guilhem Lavaux, and Benjamin D. Wandelt. Automatic physical inference with information maximizing neural networks. *Phys. Rev. D*, 97(8):083004, April 2018. doi: 10.1103/PhysRevD.97.083004.

Alan F. Heavens, Raul Jimenez, and Ofer Lahav. Massive lossless data compression and multiple parameter estimation from galaxy spectra. *MNRAS*, 317(4):965–972, October 2000. doi: 10.1046/j.1365-8711.2000.03692.x.

J. Hermans, A. Delaunoy, F. Rozet, A. Wehenkel, V. Begy, and G. Louppe. A trust crisis in simulation-based inference? your posterior approximations can be unfaithful. *arXiv e-prints*, 2022.

Anastasia N. Krouglova, Hayden R. Johnson, Basile Confavreux, Michael Deistler, and Pedro J. Gonçalves. Multifidelity Simulation-based Inference for Computationally Expensive Simulators. *arXiv e-prints*, art. arXiv:2502.08416, February 2025. doi: 10.48550/arXiv.2502.08416.

Kiyam Lin, Maximilian von Wietersheim-Kramsta, Benjamin Joachimi, and Stephen Feeney. A simulation-based inference pipeline for cosmic shear with the Kilo-Degree Survey. *MNRAS*, 524(4):6167–6180, October 2023. doi: 10.1093/mnras/stad2262.

J.-M. Lueckmann, J. Boelts, D. S. Greenberg, P. J. Goncalves, and J. H. Macke. Benchmarking simulation-based inference. *AISTATS*, 2021.

X. Lyu et al. Dynamic simulation-based inference: Round-free sequential sbi with adaptive datasets. *arXiv e-prints*, 2025.

Lucas T. Makinen, Alan Heavens, Natalia Porqueres, Tom Charnock, Axel Lapel, and Benjamin D. Wandelt. Hybrid summary statistics: neural weak lensing inference beyond the power spectrum. *JCAP*, 2025(1):095, January 2025. doi: 10.1088/1475-7516/2025/01/095.

Yueying Ni, Shy Genel, Daniel Anglés-Alcázar, Francisco Villaescusa-Navarro, Yongseok Jo, Simeon Bird, Tiziana Di Matteo, Rupert Croft, Nianyi Chen, Natalí S. M. de Santi, Matthew Gebhardt, Helen Shao, Shivam Pandey, Lars Hernquist, and Romeel Dave. The CAMELS Project: Expanding the Galaxy Formation Model Space with New ASTRID and 28-parameter TNG and SIMBA Suites. *ApJ*, 959(2):136, December 2023. doi: 10.3847/1538-4357/ad022a.

A. Saoulis, others, and M. von Wietersheim-Kramsta. Field-level weak lensing cosmology with ¡100 simulations using multifidelity simulation-based inference. *arXiv e-prints*, 2026.

A. Saoulis et al. Transfer learning for multifidelity simulation-based inference in cosmology. *arXiv e-prints*, 2025.

Joop Schaye, Roi Kugel, Matthieu Schaller, John C. Helly, Joey Braspenning, Willem Elbers, Ian G. McCarthy, Marcel P. van Daalen, Bert Vandenbroucke, Carlos S. Frenk, Juliana Kwan, Jaime Salcido, Yannick M. Bahé, Josh Borrow, Evgenii Chaikin, Oliver Hahn, Filip Huško, Adrian Jenkins, Cedric G. Lacey, and Folkert S. J. Nobels. The FLAMINGO project: cosmological hydrodynamical simulations for large-scale structure and galaxy cluster surveys. *MNRAS*, 526(4):4978–5020, December 2023. doi: 10.1093/mnras/stad2419.

A. Spurio Mancini, D. Piras, J. Alsing, B. Joachimi, and M. P. Hobson. Cosmopower: emulating cosmological power spectra for accelerated bayesian inference. *MNRAS*, 2022.

Nicolas Tessore, Arthur Loureiro, Benjamin Joachimi, Maximilian von Wietersheim-Kramsta, and Niall Jeffrey. GLASS: Generator for Large Scale Structure. *The Open Journal of Astrophysics*, 6:11, March 2023. doi: 10.21105/astro.2302.01942.

F. Villaescusa-Navarro et al. The camels project: Cosmology and astrophysics with machine-learning simulations. *ApJ*, 2021.

Maximilian von Wietersheim-Kramsta, Kiyam Lin, Nicolas Tessore, Benjamin Joachimi, Arthur Loureiro, Robert Reischke, and Angus H. Wright. KiDS-SBI: Simulation-based inference analysis of KiDS-1000 cosmic shear. *A&A*, 694:A223, February 2024. doi: 10.1051/0004-6361/202450487.

S. Wagner-Carena et al. A strong gravitational lens is worth a thousand dark matter halos: Inference of galaxy-scale substructure. *ApJ*, 2024.