~~Axions~~, Strong lensing and Deep learning

Towards convincing dark matter discoveries in astrophysical data

Christoph Weniger, University of Amsterdam

Alex Cole, Adam Coogan, Kosio Karchev, Ben Miller, Noemi Anau Montel

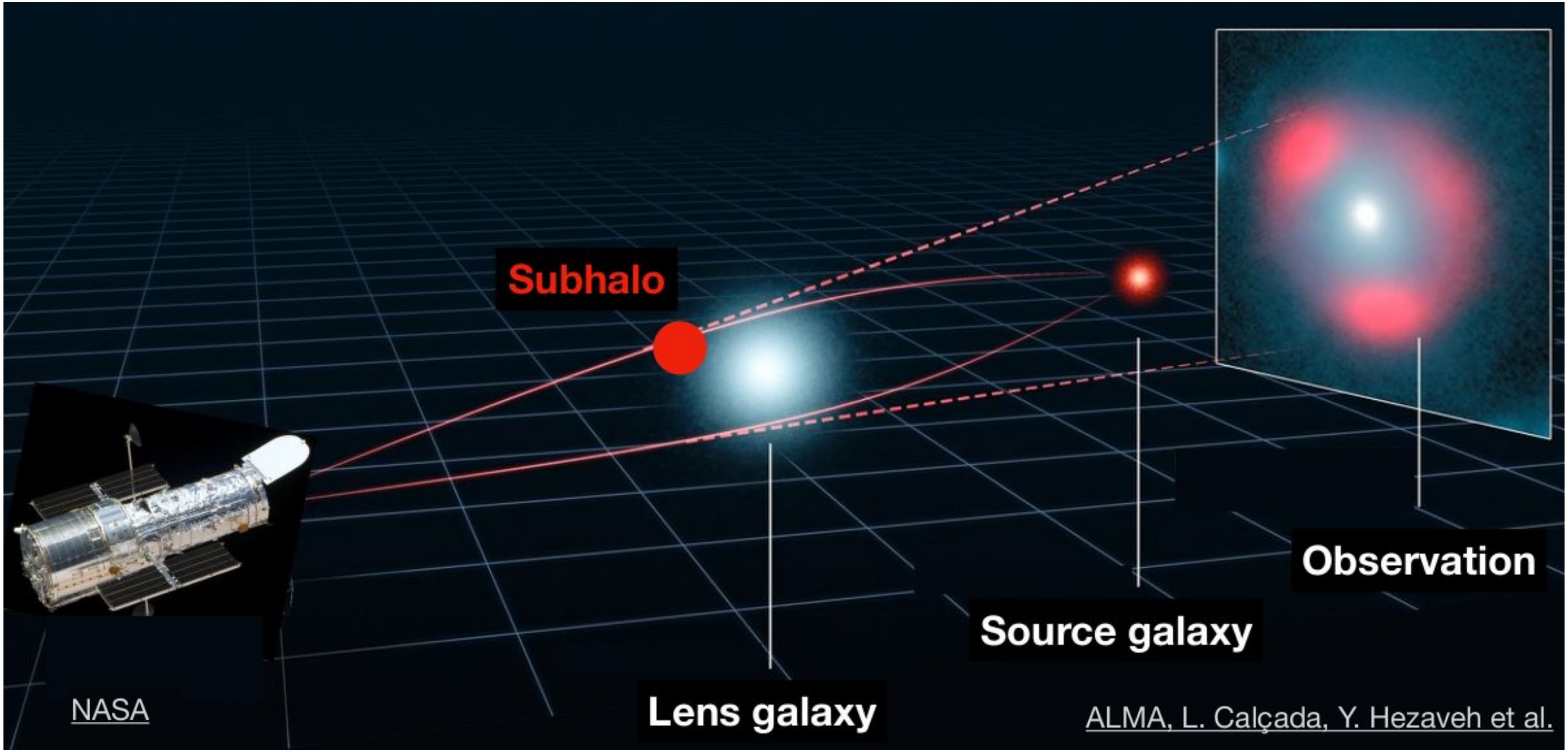
Axion talk (youtube, Sam Witte, TAUP2021)

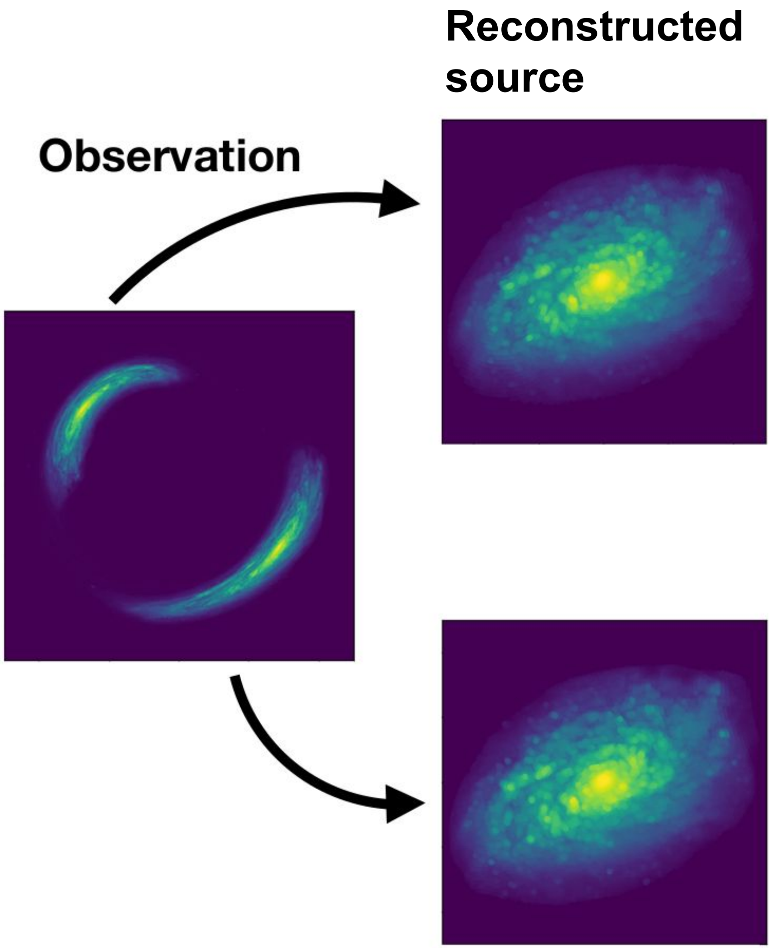
GRAPPA
GRavitation AstroParticle Physics Amsterdam

European Research Council

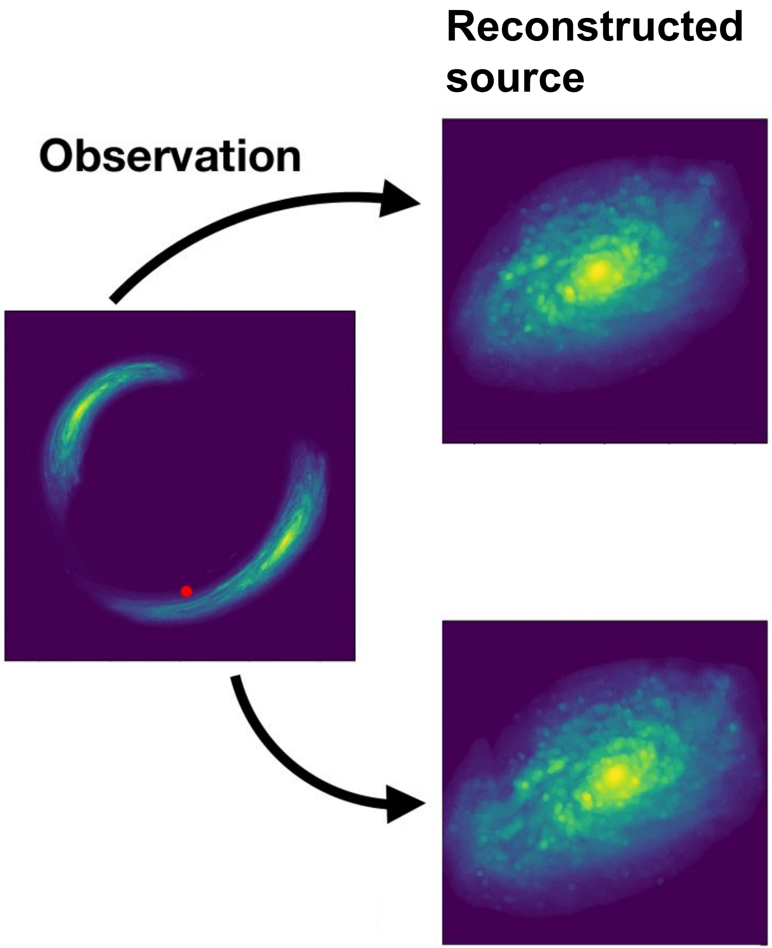
PANIC2021 Conference
8 September 2021

Strong galaxy-galaxy lensing



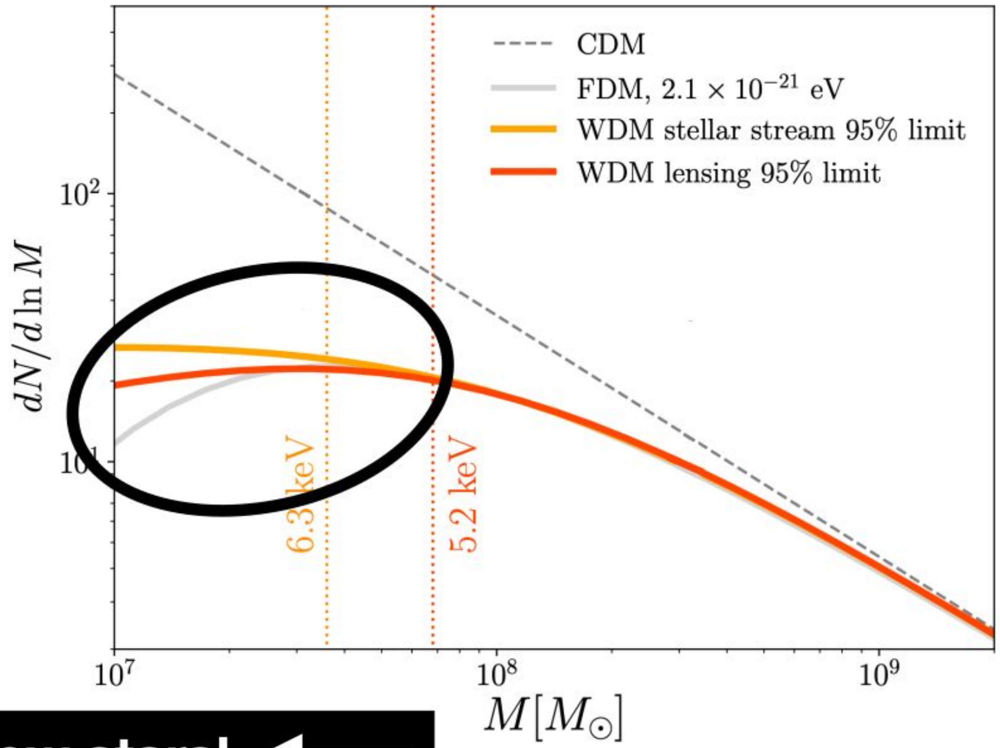


Slide credit: Adam Coogan



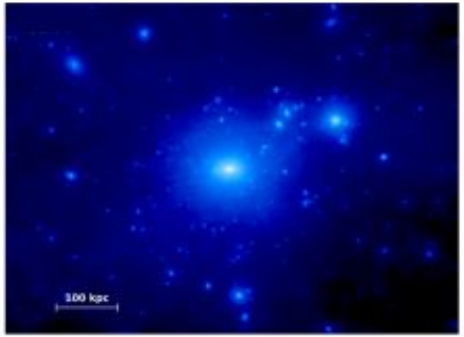
Slide credit: Adam Coogan

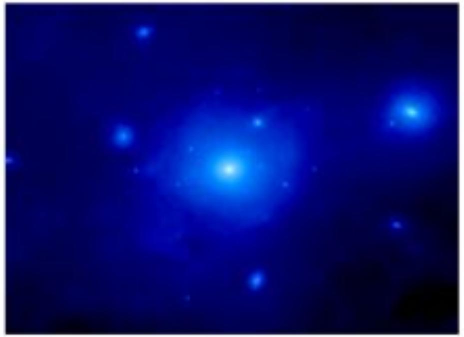
Subhalo mass function

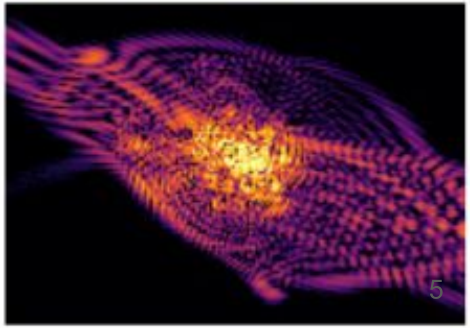


Few stars! ←

Schutz 2020, PRD 101, 123026
Fitts et al 2017, MNRAS 471 3







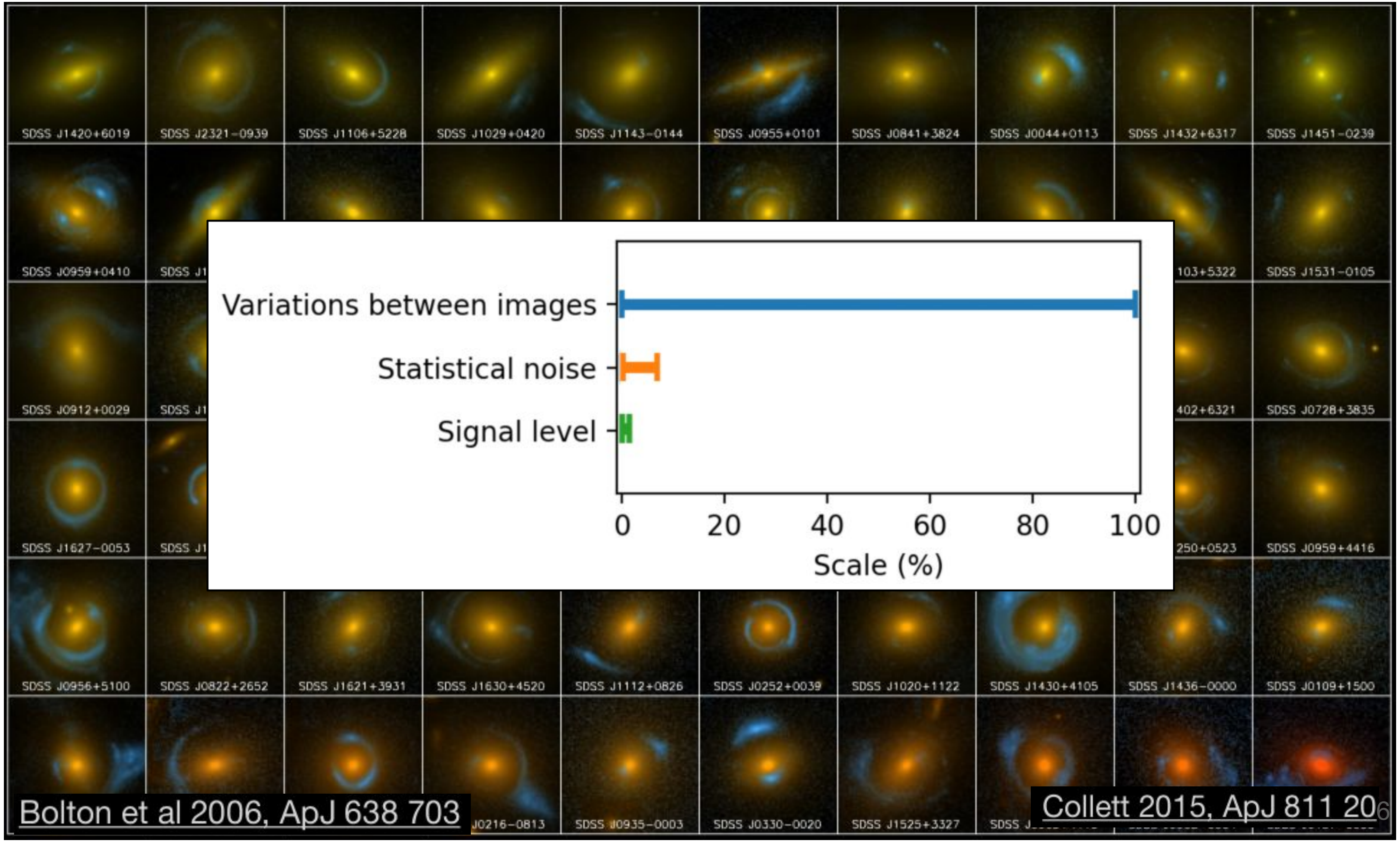
Slide credit: Adam Coogan

Strong lensing images

Present:
~60 lenses
(mostly HST)

Near future:
>150.000 lenses

(JWST, Euclid,
Rubin Obs., ELT)



Marginal posteriors

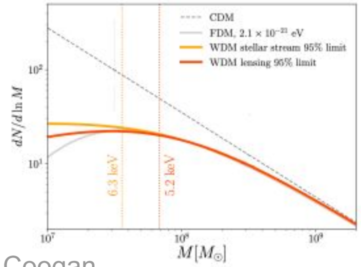
Physics modeling: $p(x \mid \theta)$ → **Inverse problem:** $p(\theta \mid x_0)$

But we want *marginal posteriors:*

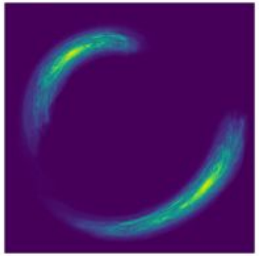
$$p(\theta \mid x_0) = \int d^{n}\eta \, p(\theta, \eta \mid x_0)$$

10^3 - 10^5

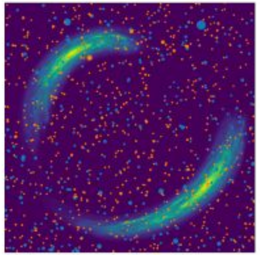
E.g.: mass function cutoff



Observation

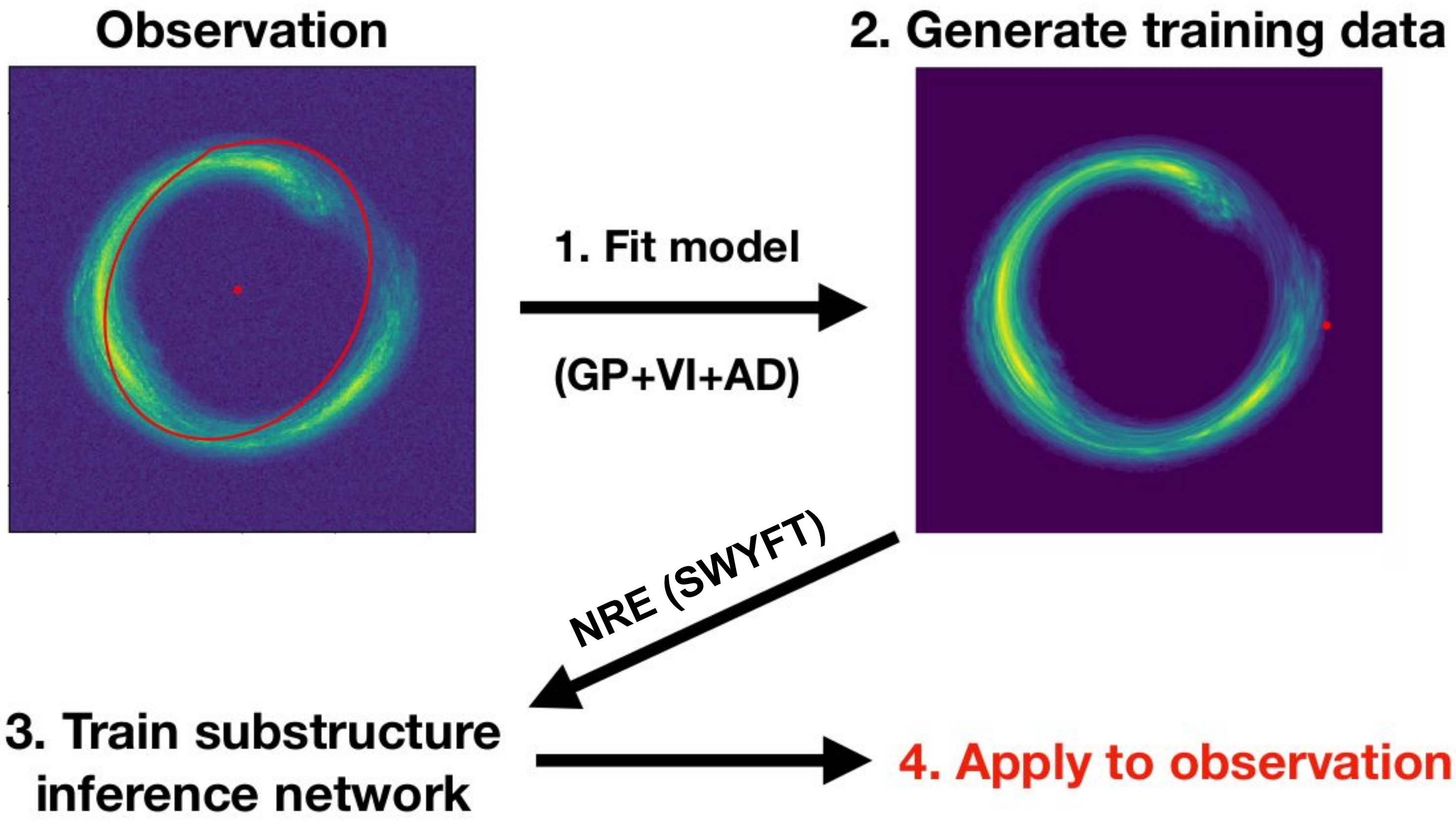


E.g.: subhalo parameters



Slide credit: Adam Coogan

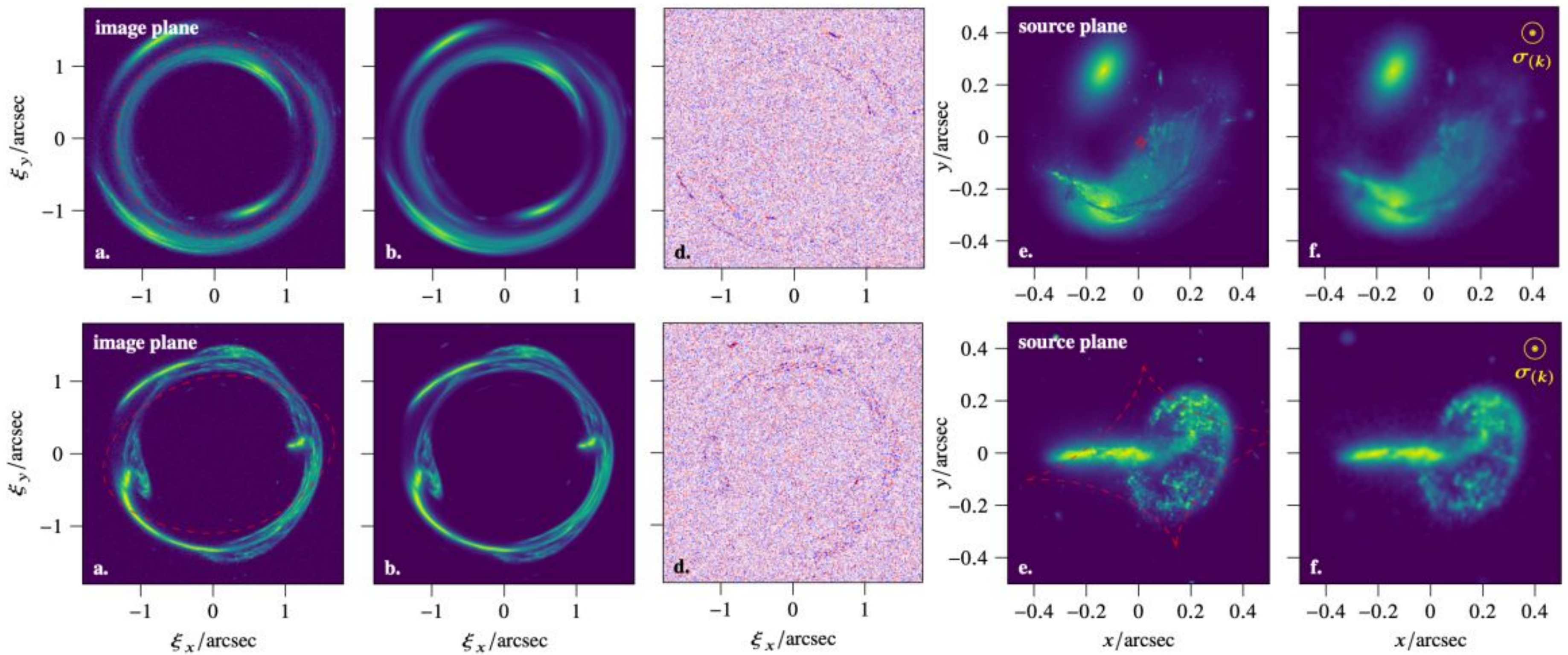
Our strategy: Targeted neural inference



Slide credit: Adam Coogan

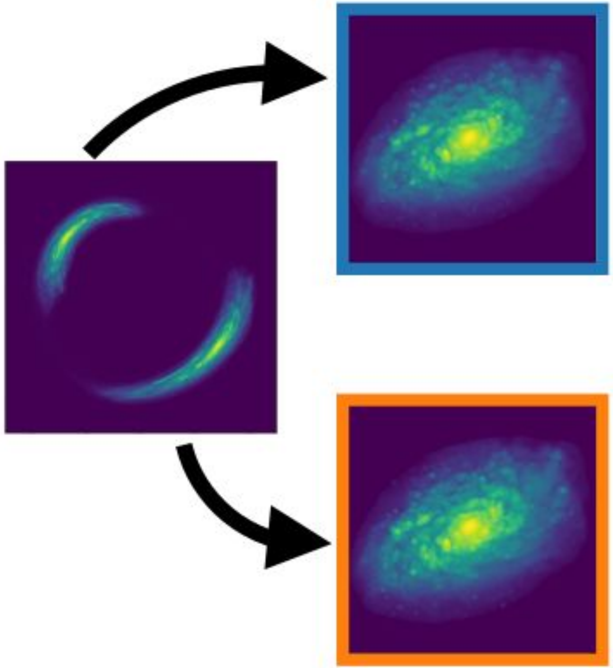
Coogan+ 2010.07032

Step 1: Lens and source fit using variational inference



Karchev+ 2105.09465

Step 1: **GP**+VI+AD



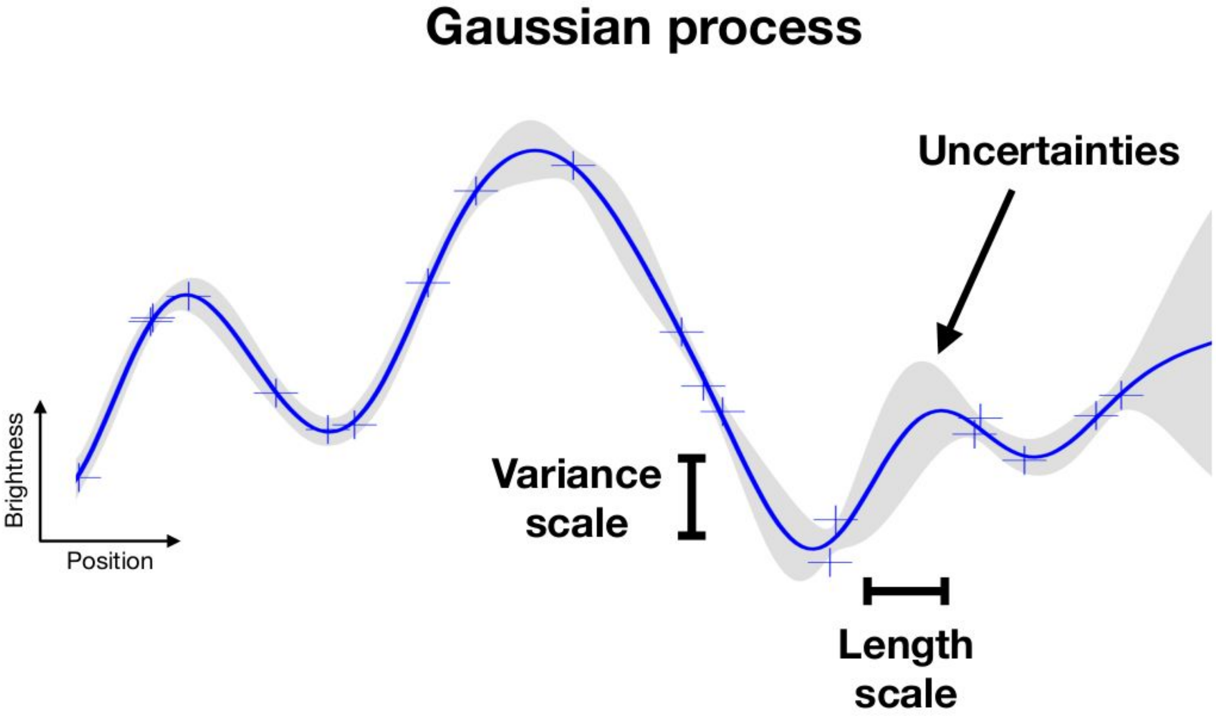
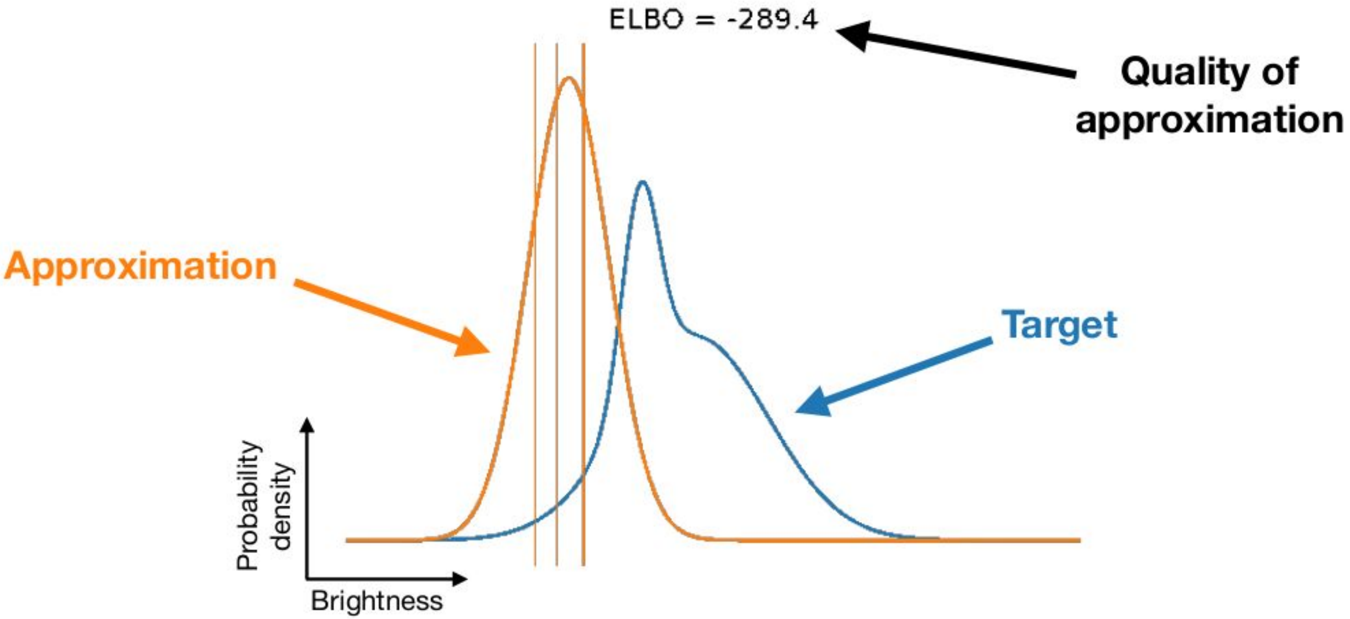


Image credit: Adam Coogan

Step 1: GP+**VI**+AD

Approx. marginal likelihood with variational inference



Slide credit: Adam Coogan

Step 1: GP+VI+**AD**

Automatic differentiation

```
def f(x):
  return sin(x)
```

→

```
grad(f)(1.1)
>>> 0.4536
```

Could've done that by hand. But this is harder:

```
def g(x):
  for _ in range(100):
    x = sin(x)
  return x
```

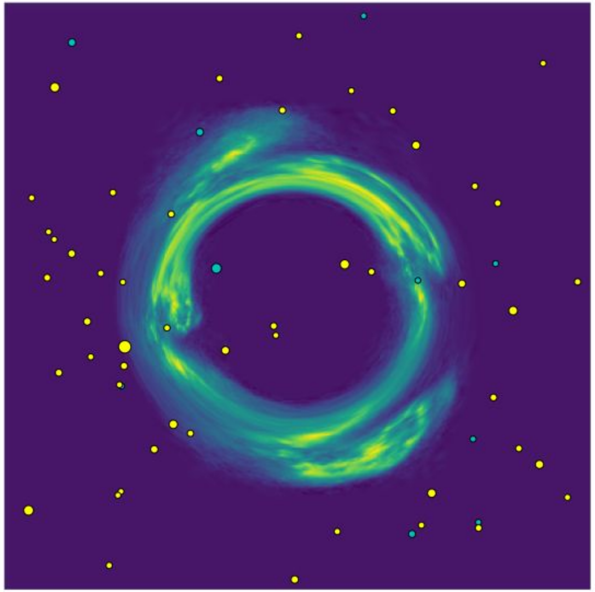
→

```
grad(g)(1.1)
>>> 0.0026
```

Core ML tech with many implementations

Slide credit: Adam Coogan

Step 2: Targeted training of inference network for the inference of subhalo population properties

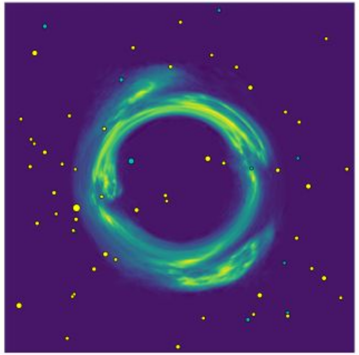


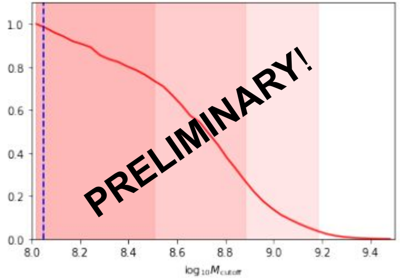
Nuisance parameters are marginalised via random sampling

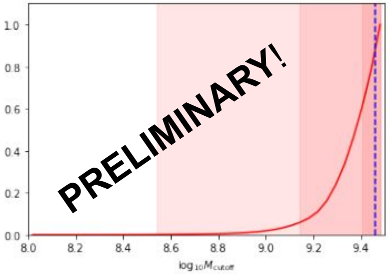
$$\vec{x}_{\text{image}}, M_{\text{cutoff}} \sim \underbrace{p(\vec{x}_{\text{image}} \,|\, \vec{z}_l, \vec{z}_s, \vec{z}_{\text{sub}})}_{\text{Simulator}} \; \underbrace{p(\vec{z}_l) p(\vec{z}_s)}_{\text{Priors from fit}} \; \underbrace{p(\vec{z}_{\text{sub}} \,|\, M_{\text{cutoff}})}_{\text{Sub. population}} \; p(M_{\text{cutoff}})$$

Slide credit: Noemi Anau Montel

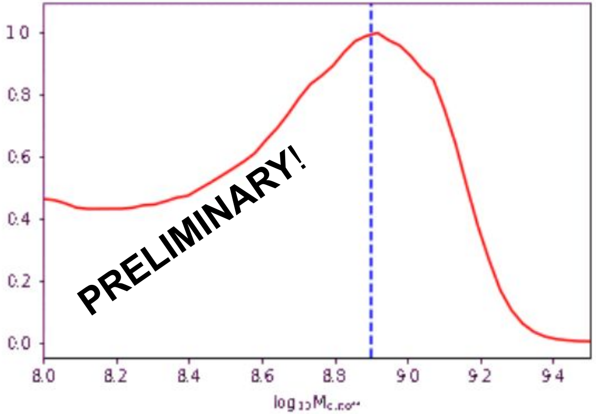
Results: Constraints on subhalo population (in mock data)







Combining observations (30)



Details:

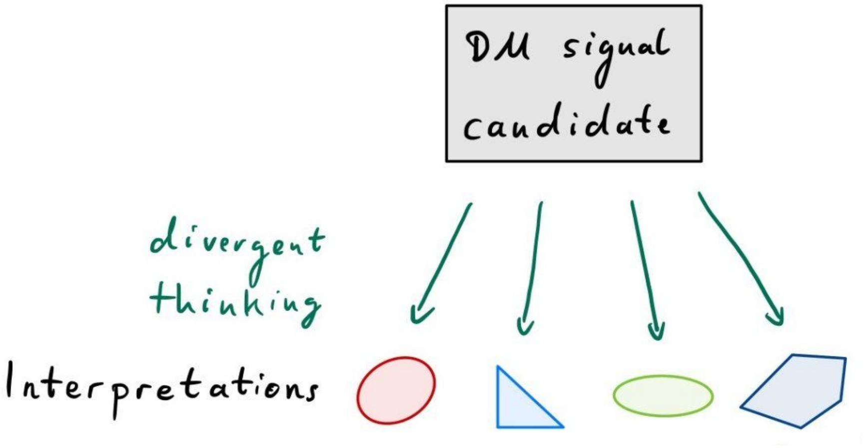
- Fit mock observation (several hours)
  - Noise = 1
  - SNR = 30
- Generate targeted training data (~40 minutes for 20000 samples)
  - 10 subhalos and 50 l.o.s. halos in mass range $[10^8, 10^{9.5}]\ M_\odot$
  - DM cutoff scale in the same range
- Train inference network (~30 minutes)

Montel+ in prep.

What lessons for indirect dark matter searches?

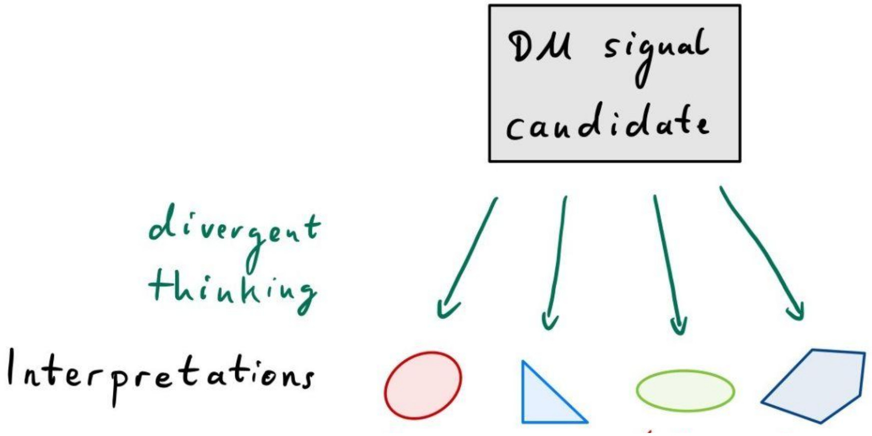
Indirect dark matter searches are broken

What should happen

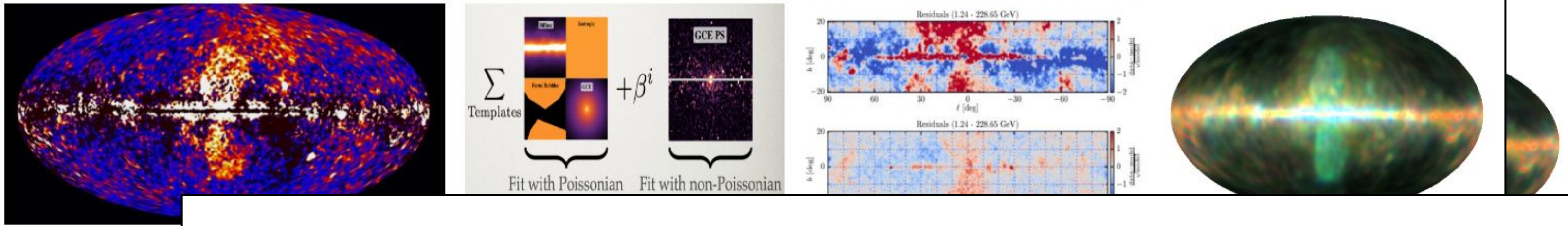


Indirect dark matter searches are broken

What *usually* happens

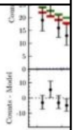


All analyses are defined by their compromises

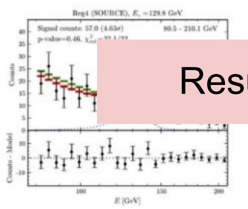


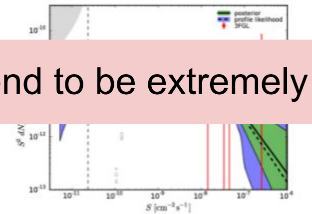
Templ

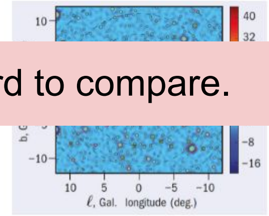
- $p(\mathbf{x}_{\mathrm{pi0}}, \boldsymbol{\eta}_{\mathrm{dif}}$
- $p(\mathbf{A}) \rightarrow \prod$
- Poisson →

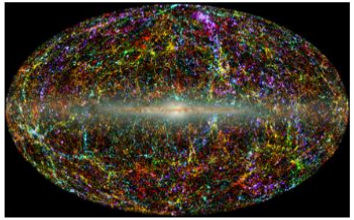


L









Results tend to be extremely hard to compare.

Line searches

- No spatial information, only summary energy spectra
- ROI is small energy range
- Approximate backgrounds with power-law
- "Trade systematics for statistical uncertainties"

One-point statistics

- Model likelihood of count statistics histogram rather than likelihood of skymaps
- Mask difficult sky regions
- Focus on single energy band

Wavelet filtering

- Put data with complex diffuse backgrounds through a matched (wavelet) filter
- Otherwise similar to One-point statistics analysis

Cross-correlation studies

- Project data onto tracers of DM mass
- Neglect effects of local Universe
- Sometimes neglect correlation between overlapping tracers

A Bayesian perspective

Independently of the analysis technique, everything relevant for analyzing a piece of data can be in principle summarized in a huge probabilistic model.

Probabilistic model

$$p(\mathbf{x}_{\text{data}}, \boldsymbol{\theta}_{\text{phys.}}, \boldsymbol{\theta}_{\text{instr.}}, \boldsymbol{\theta}_{\text{misc.}}) = p(\mathbf{x}_{\text{data}} | \boldsymbol{\theta}_{\text{phys.}}, \boldsymbol{\theta}_{\text{instr.}}, \boldsymbol{\theta}_{\text{misc.}}) p(\boldsymbol{\theta}_{\text{phys.}}, \boldsymbol{\theta}_{\text{instr.}}, \boldsymbol{\theta}_{\text{misc.}})$$

Physical simulators

Instrument simulation

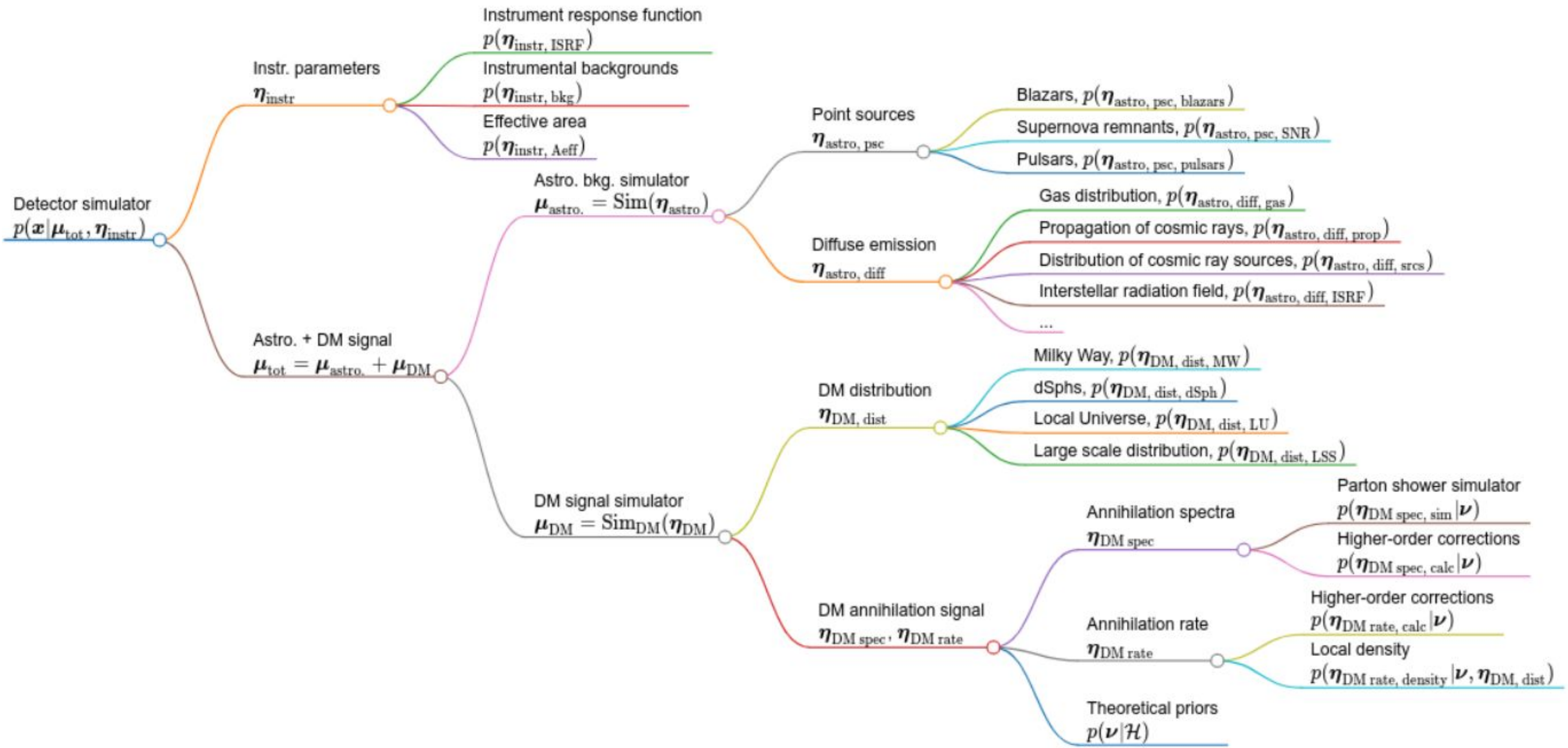
Complete accounting for all known unknowns

Problem: Higher model realism

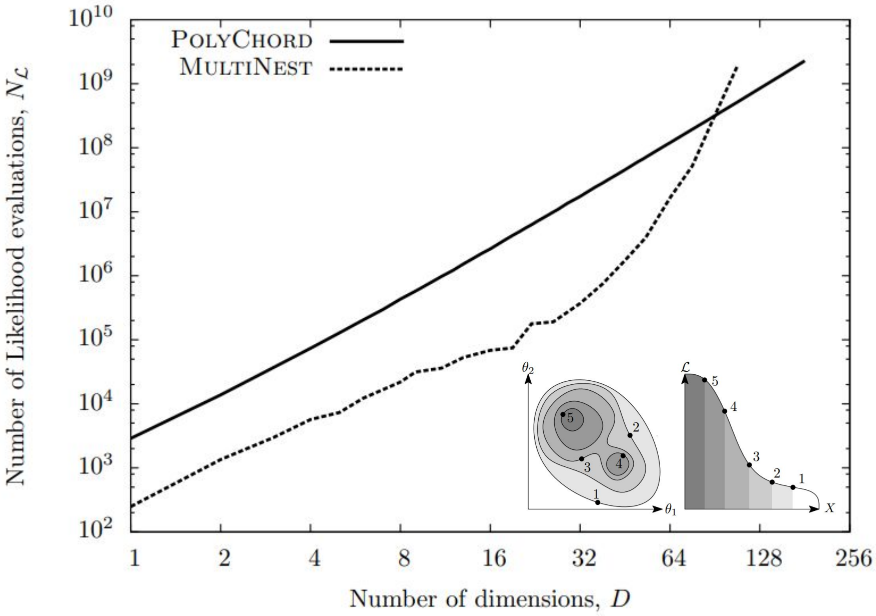
More parameters

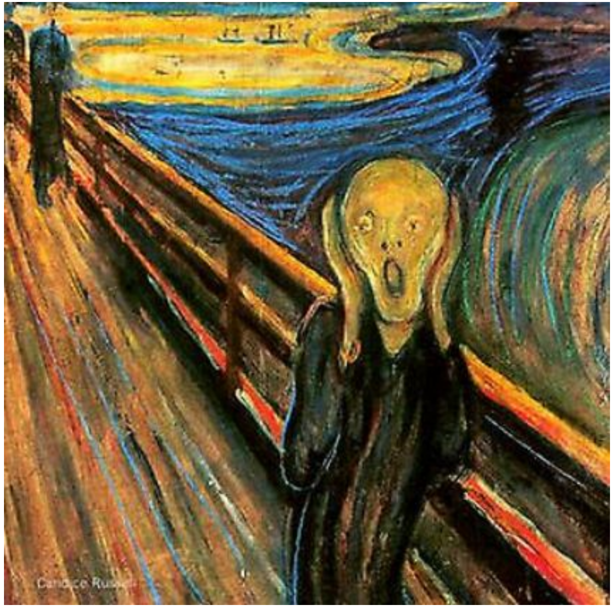
Higher per-simulation costs

A graphical model for Fermi LAT data (illustration)



A high fidelity analysis of complex data is hard with commonly used (likelihood-based) techniques

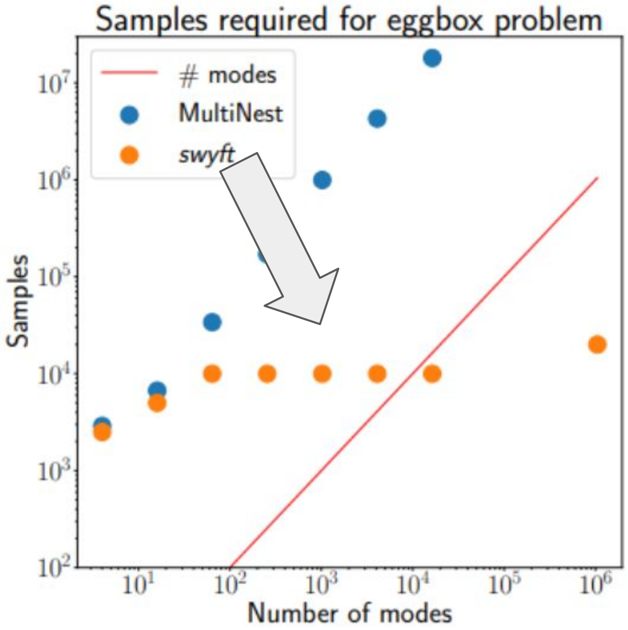


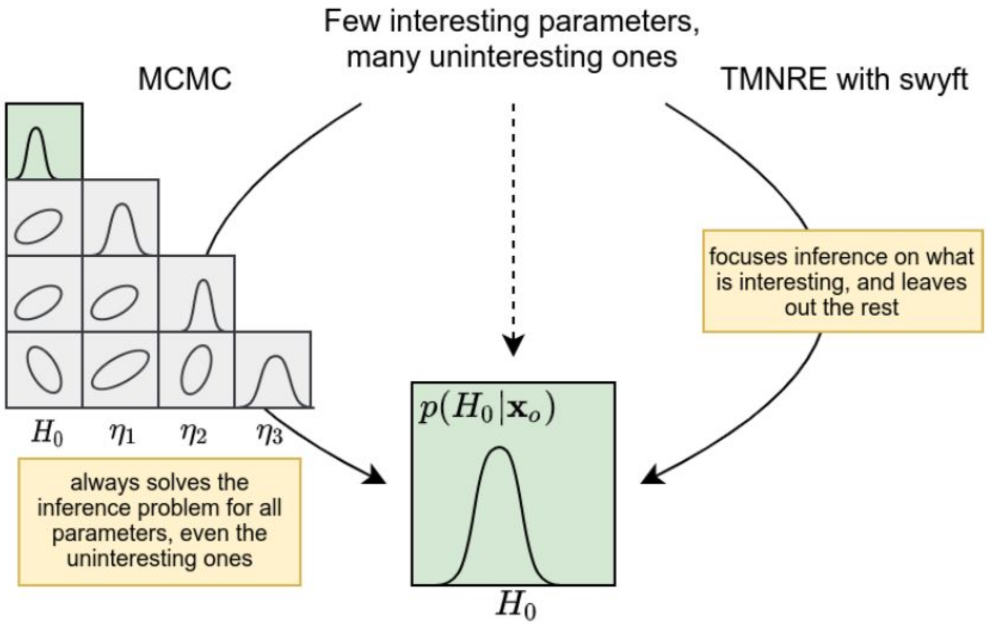


Handley+ 1506.00171

Don't PANIC

Cutting to the chase with neural simulation-based inference





Miller+ 2011.13951, Miller+ 2107.01214, Cole+ in prep.

Simulation-based inference

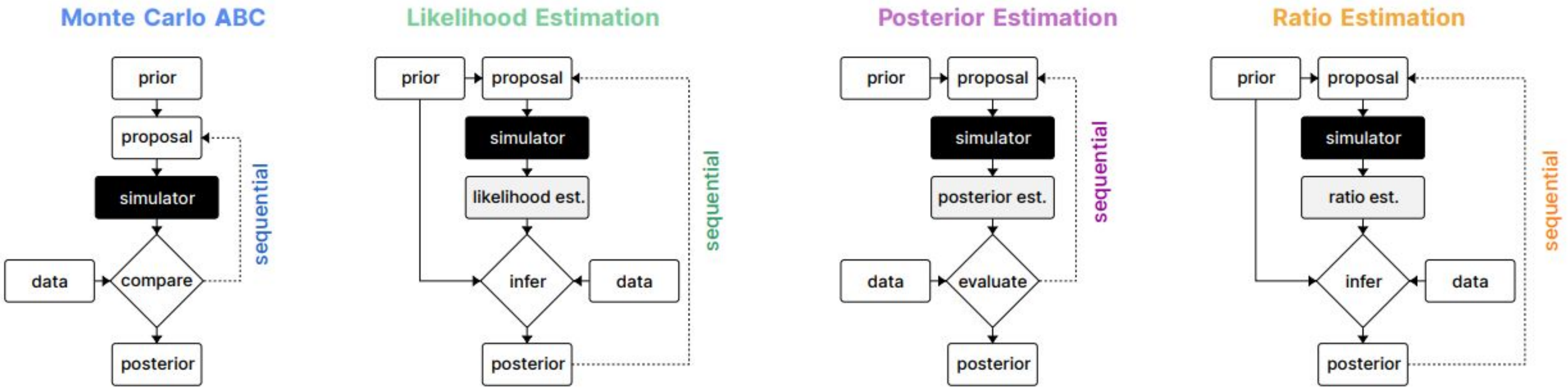
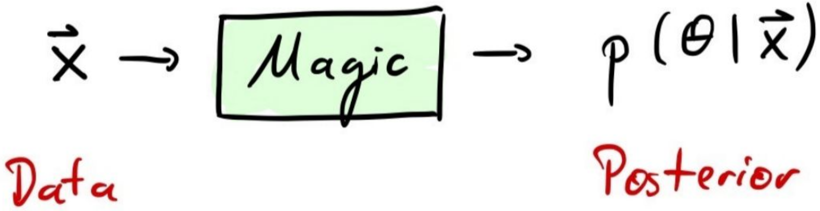
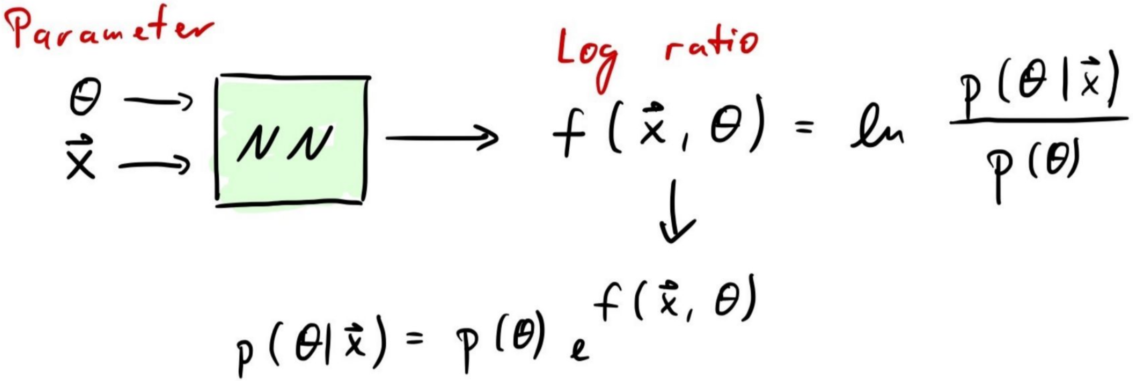


Image: Lueckermann+ 2101.04653

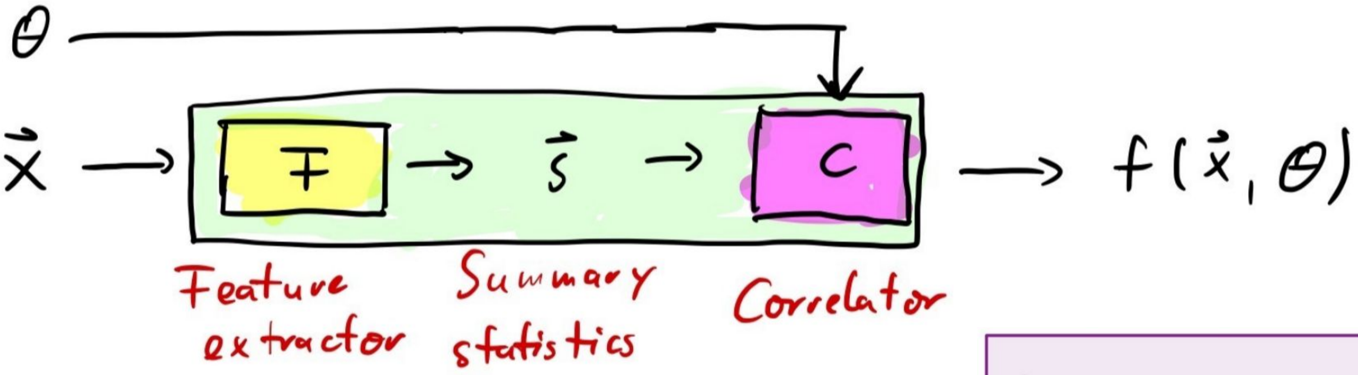
Some relevant papers:
Cranmer+ 1911.01429
Durkan+ 2002.03712
Papamakarios+ 1605.06376
Tran+ 1702.08896
Alsing+ 1903.00007
Hermans+ 1903.04057
Miller+ 2011.13951
Miller+ 2107.01214
...

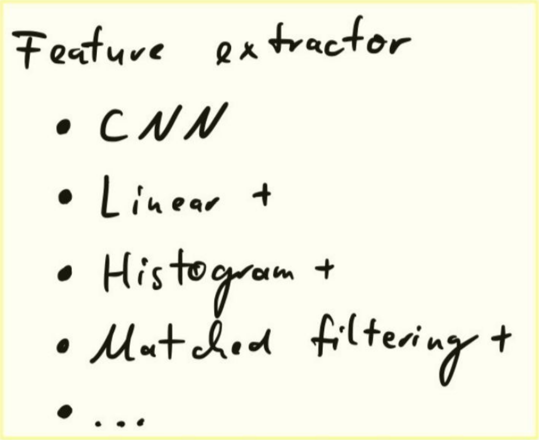
Truncated marginal neural ratio estimation with SWYFT





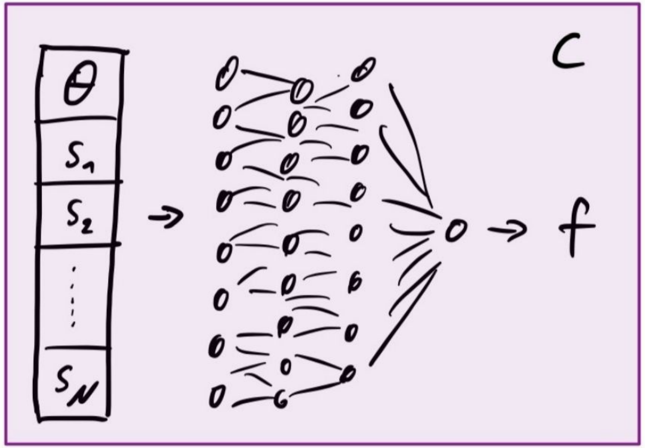
Truncated marginal neural ratio estimation with SWYFT



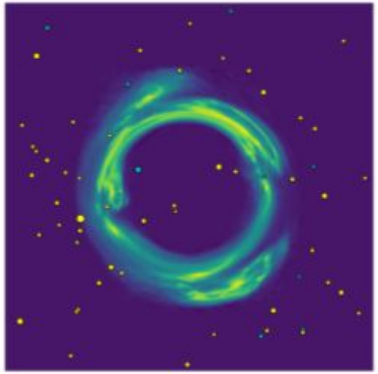


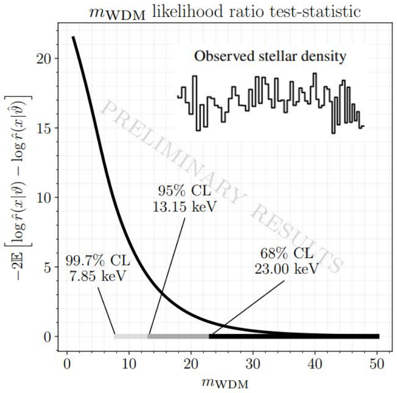
https://github.com/undark-lab/swyft

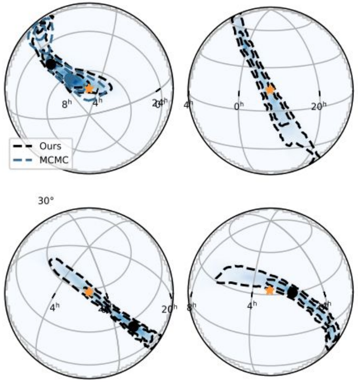
```
pip install swyft
```

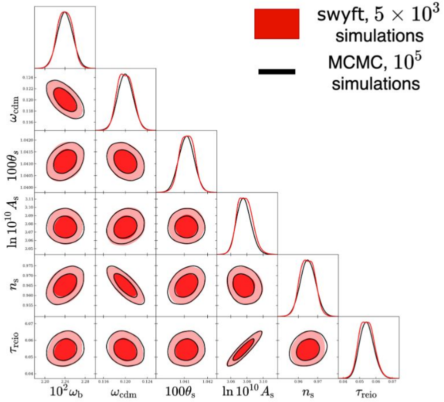


What can this do for us?









Hermans+ 2011.14923

Delaunoy+ 2010.12931

Cole+ in prep

Strong lensing

→ Constrain on DM mass

> 100.000 nuisance parameters

Analysis of GD-1 stream

→ Lower limit on DM mass

> 100 nuisance parameters

Gravitational waves

→ Instant localization

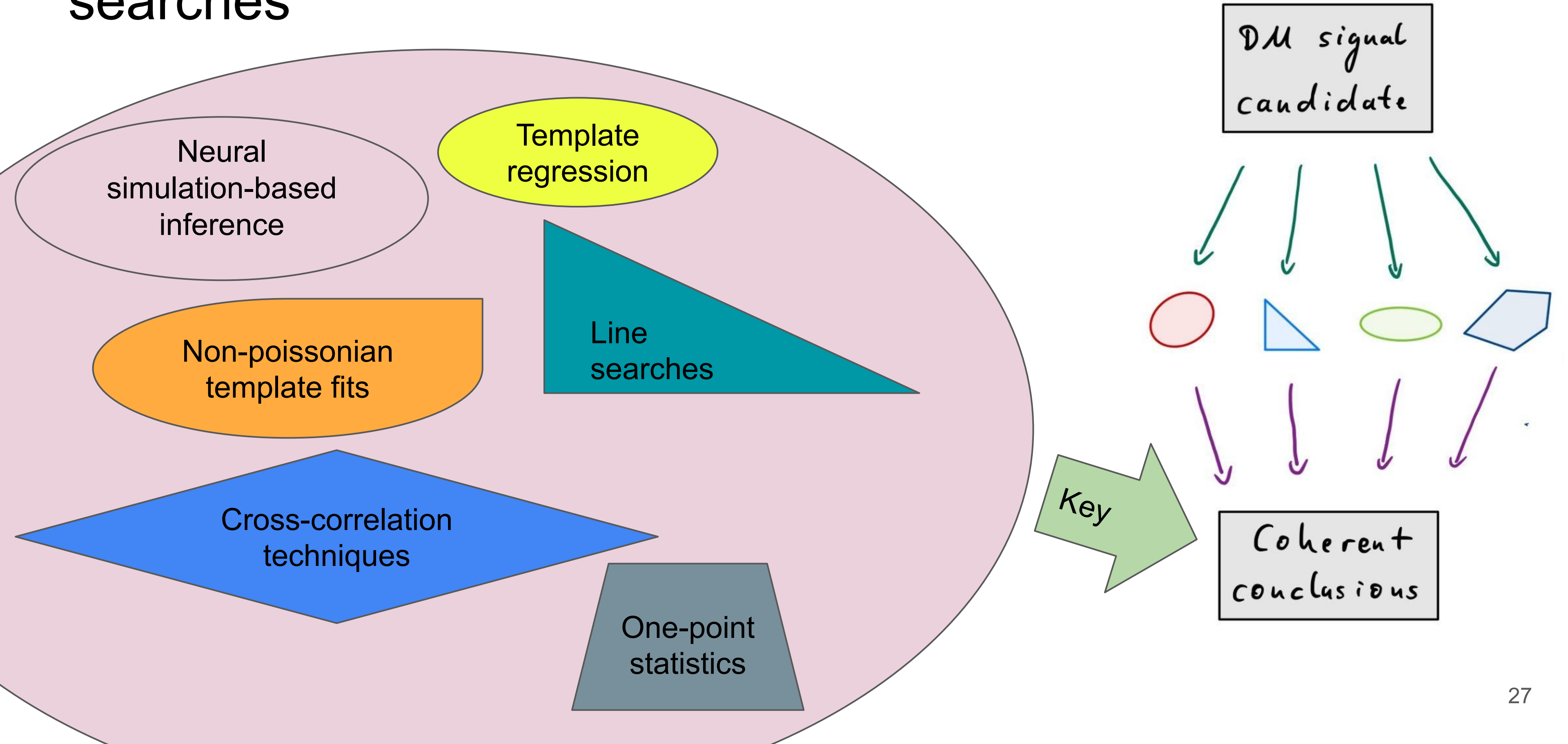
10-20 nuisance parameters

Cosmology

→ Cosmological parameters

10-20 nuisance parameters

Simulation-based inference as key for conclusive DM searches



Conclusions

Conclusions

- We developed a new **analysis pipeline for strong lensing searches for DM substructure** images
  - Step 1: Fitting images with end-to-end differentiable models and variational inference
  - Step 2: Targeted training of inference networks to extract population information about small scale structure
  - The method works (and is transferable to other data analysis problems). Right now sensitivity down to O(1e8 Msol) on mock ELT images. Much more to come.
- **Deep learning can be key for convincing dark matter discoveries** in astrophysical data
  - Classical analysis methods enforce compromises that make results hard to compare.
  - Neural simulation-based inference enables us to consider much more complete and realistic models.
  - Huge potential for the community to combine forces and converge on the interpretation of dark matter signal candidates.
- With **SWYFT** we provide a “batteries included” open source tool **for neural simulation-based inference** (steep learning curve, but high gain).

Thank you!

Backup slides

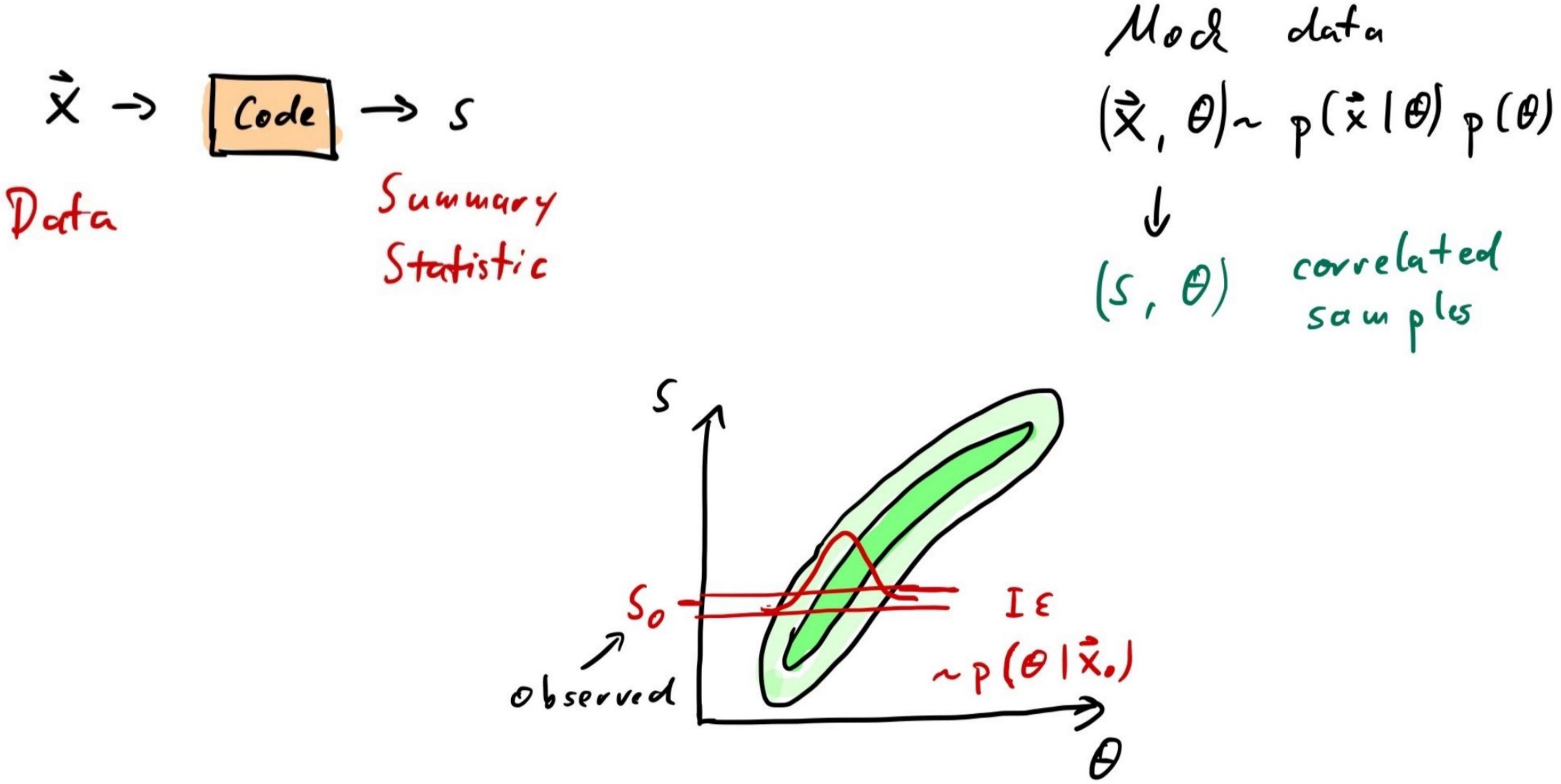
Cursed by the dimensionality of your nuisance space?

Wasted by Markov chains that reject your simulations?

Exhausted from messing with simplistic models, because your inference algorithm cannot handle the truth?

Try swyft for some pain relief.

Compare with: Approximate Bayesian Computation



Detection of individual subhalos

- We train a network to estimate marginal posteriors
- We handle models with hundreds of thousand of parameters