

# An introduction to simulation-based inference

Advanced Machine Learning  
March 10, 2022

Gilles Louppe  
[g.louppe@uliege.be](mailto:g.louppe@uliege.be)



Thanks!



Kyle Cranmer



Juan Pavez



Johann Brehmer



Joeri Hermans



Antoine Wehenkel



Arnaud Delaunoy



Norman Marlier



Francois Rozet



Malavika Vasist



Christophe Weniger



Siddarth Mishra-Sharma



Lukas Heinrich



Atilim Güneş Baydin

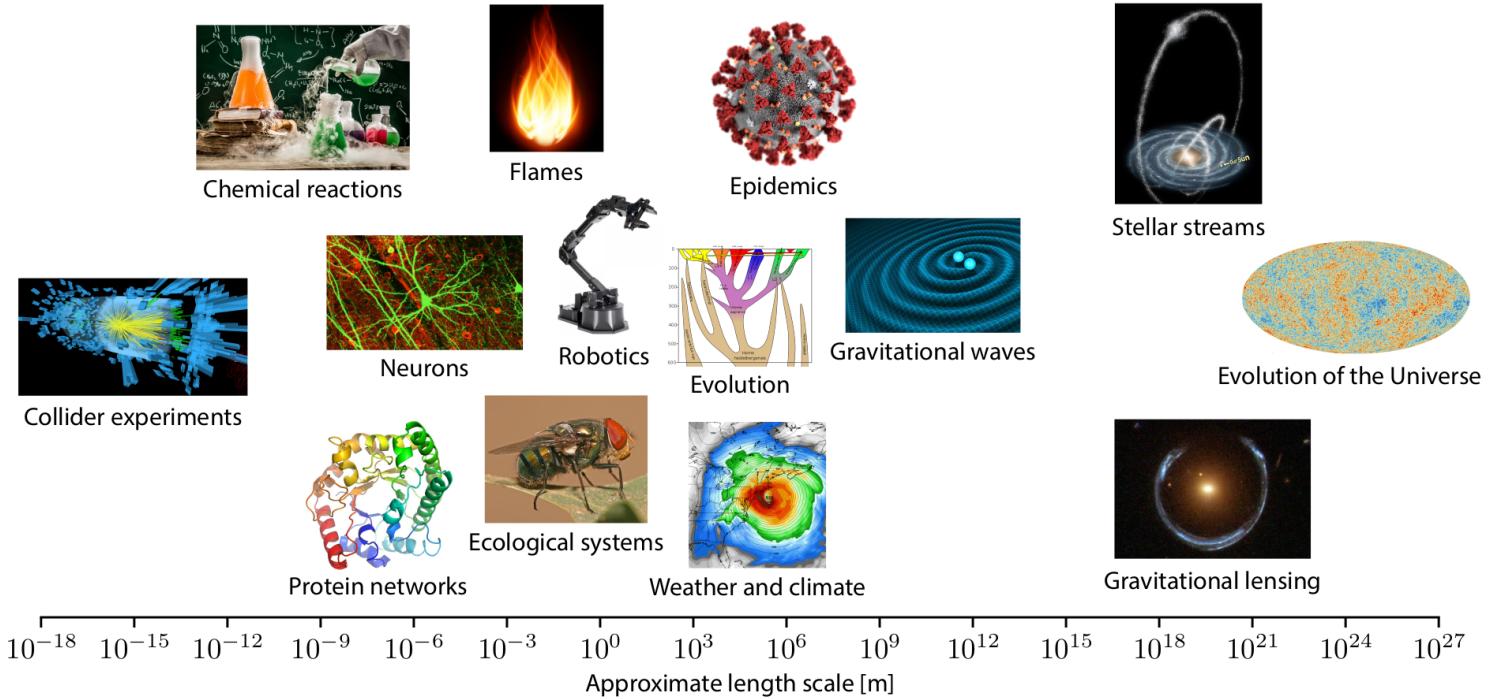


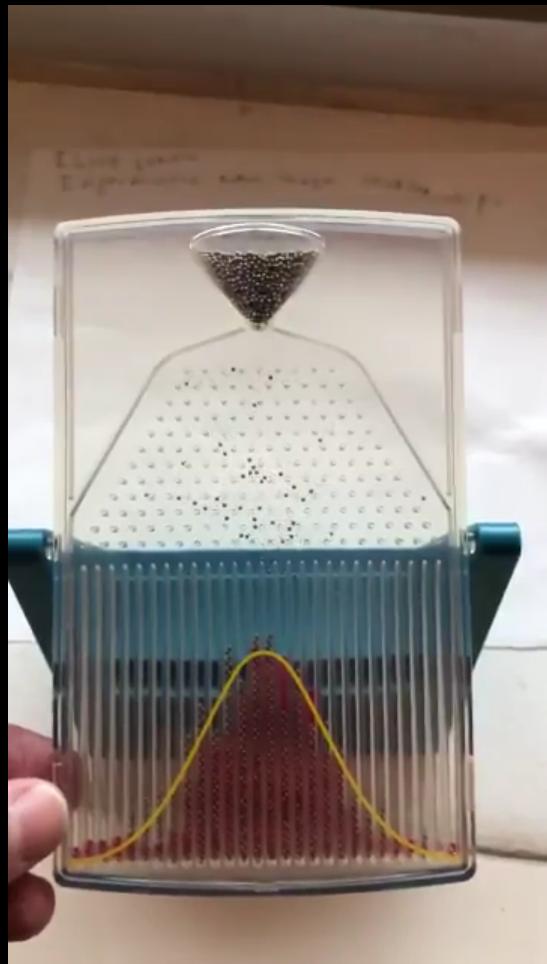
Gianfranco Bertone



Nil Banik

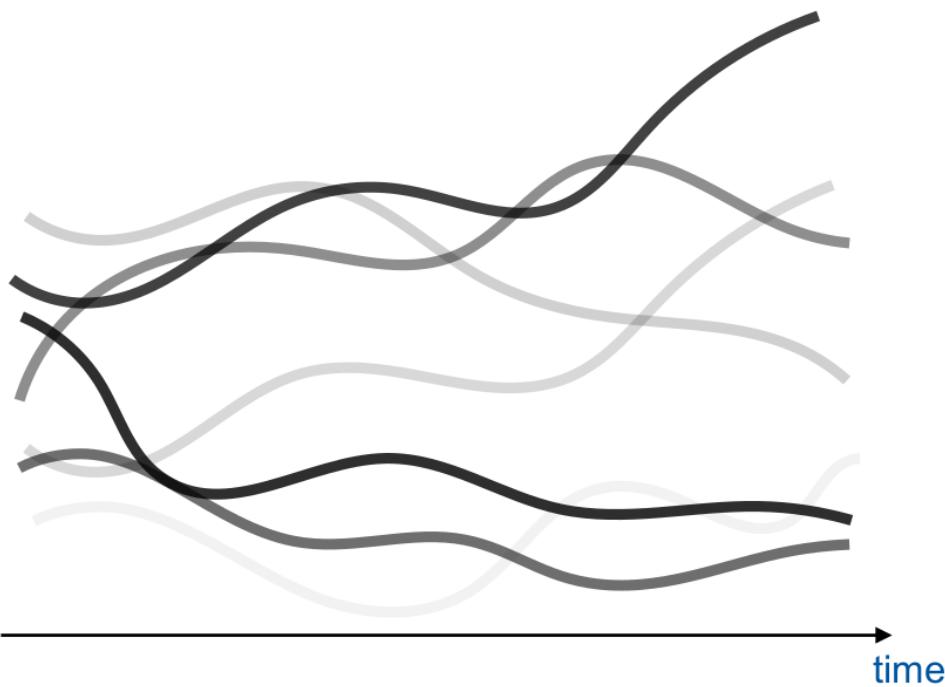
# Scientific simulators





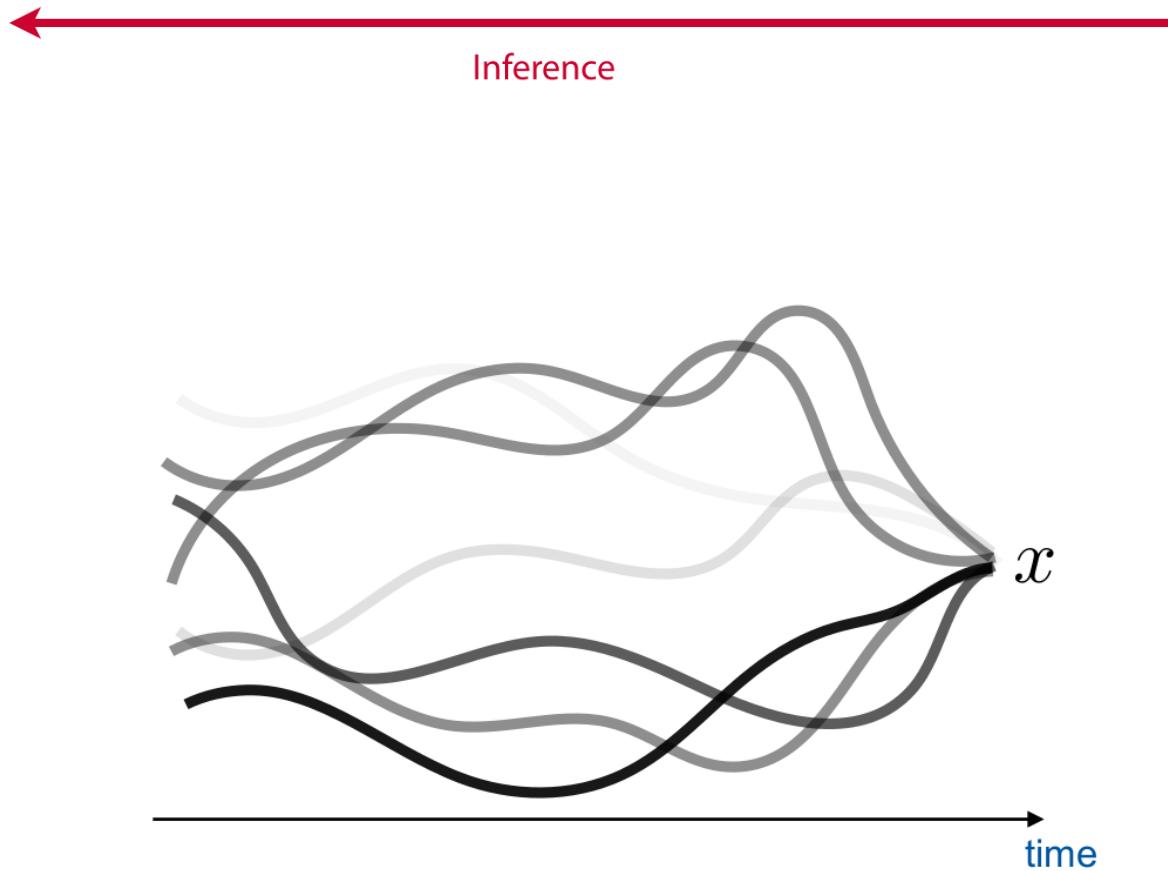


Prediction



time

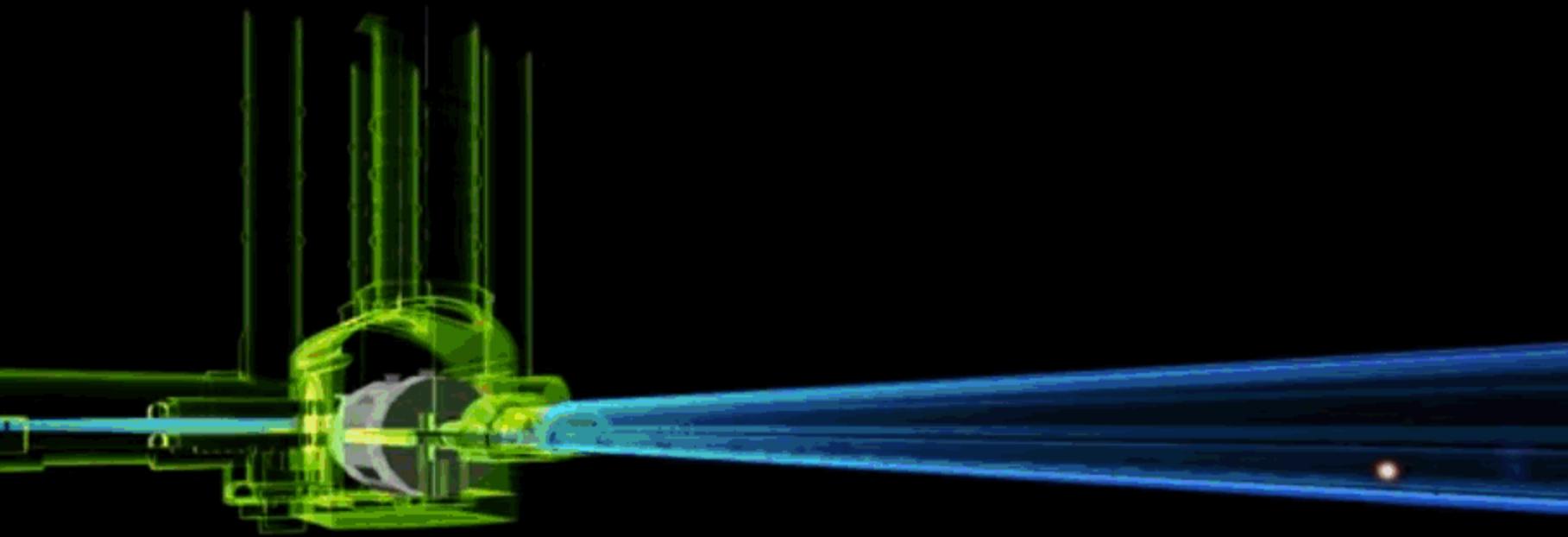
$$\theta, z, x \sim p(\theta, z, x)$$

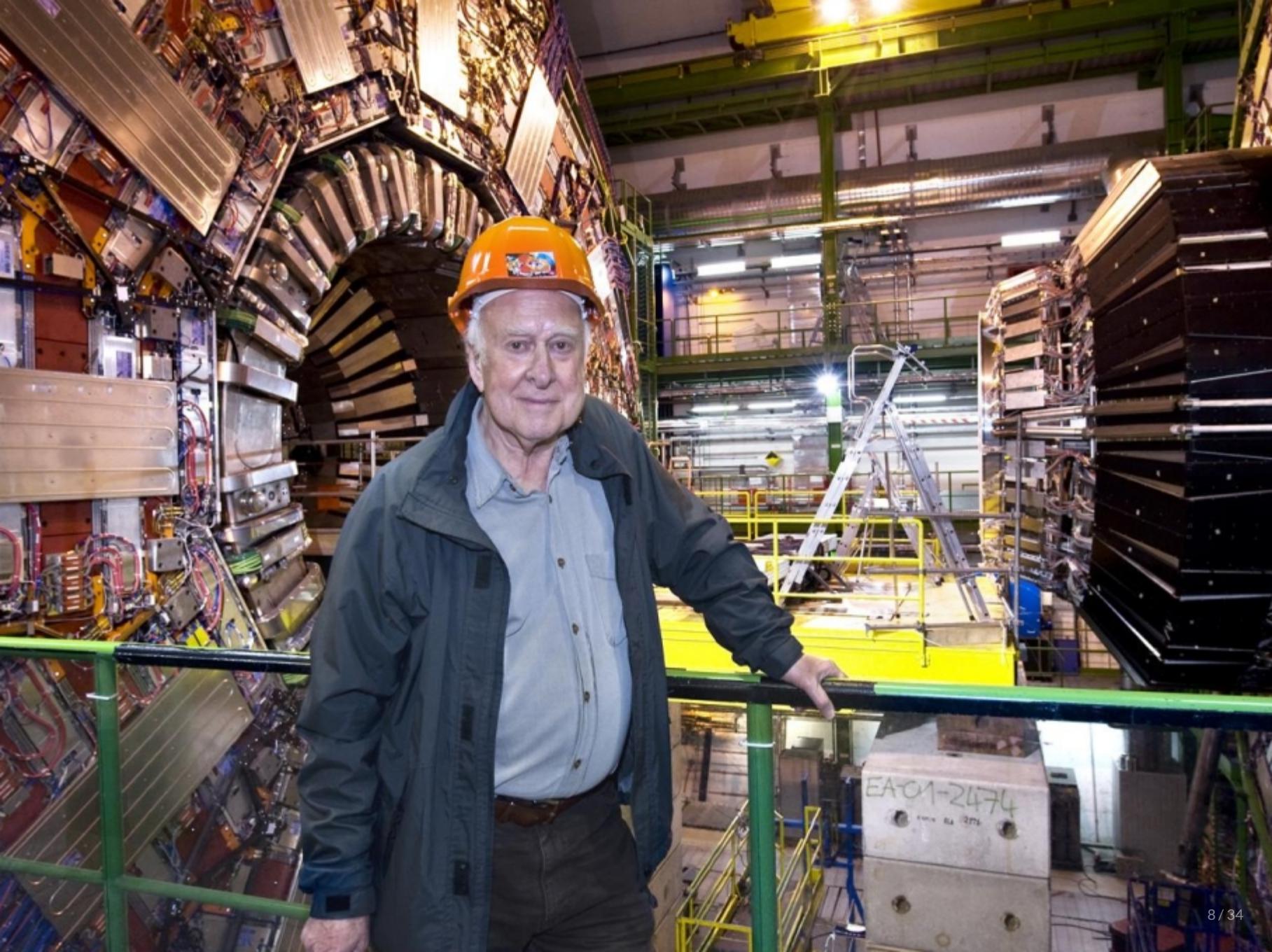


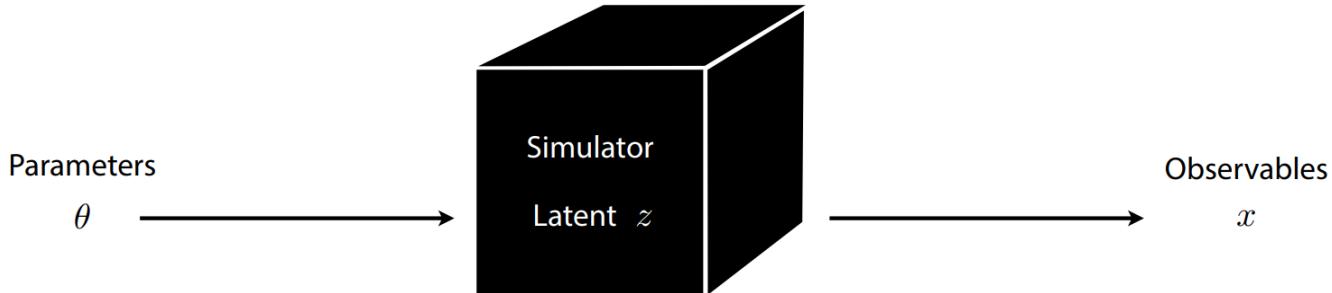
This results in the likelihood  $p(x|\theta) = \int p(x, z|\theta) dz$  to be intractable.

# The case of particle physics

$$\begin{aligned}
\mathcal{L}_{SM} = & -\frac{1}{2}\partial_\mu g_\mu^a \partial_\nu g_\mu^a - g_s f^{abc} \partial_\mu g_\nu^a g_\mu^b g_\nu^c - \frac{1}{4}g_\mu^2 f^{abc} f^{acd} g_\mu^b g_\nu^c g_\mu^d g_\nu^e - \partial_\nu W_\mu^+ \partial_\nu W_\mu^- - \\
& M^2 W_\mu^+ W_\mu^- - \frac{1}{2}\partial_\nu Z_\mu^0 \partial_\nu Z_\mu^0 - \frac{1}{2c_w^2} M^2 Z_\mu^0 Z_\mu^0 - \frac{1}{2}\partial_\mu A_\nu \partial_\mu A_\nu - ig s_w (\partial_\nu Z_\mu^0 (W_\mu^+ W_\nu^- - \\
& W_\nu^+ W_\mu^-) - Z_\mu^0 (W_\mu^+ \partial_\nu W_\mu^- - W_\nu^+ \partial_\nu W_\mu^+) + Z_\mu^0 (W_\mu^+ \partial_\nu W_\mu^- - W_\nu^- \partial_\nu W_\mu^+)) - \\
& ig s_w (\partial_\nu A_\mu (W_\mu^+ W_\nu^- - W_\nu^+ W_\mu^-) - A_\nu (W_\mu^+ \partial_\nu W_\mu^- - W_\nu^- \partial_\nu W_\mu^+) + A_\mu (W_\nu^+ \partial_\nu W_\mu^- - \\
& W_\nu^- \partial_\nu W_\mu^+)) - \frac{1}{2}g^2 W_\mu^+ W_\mu^- W_\nu^+ W_\nu^- + \frac{1}{2}g^2 W_\mu^+ W_\nu^+ W_\mu^- W_\nu^- + g^2 c_w^2 (Z_\mu^0 W_\mu^+ Z_\nu^0 W_\nu^- - \\
& Z_\mu^0 Z_\nu^0 W_\mu^+ W_\nu^-) + g^2 s_w^2 (A_\mu W_\mu^+ A_\nu W_\nu^- - A_\mu A_\nu W_\mu^+ W_\nu^-) + g^2 s_w c_w (A_\mu Z_\nu^0 (W_\mu^+ W_\nu^- - \\
& W_\nu^+ W_\mu^-) - 2 A_\mu Z_\mu^0 W_\nu^+ W_\nu^-) - \frac{1}{2} \partial_\mu H \partial_\mu H - 2 M^2 \alpha_h H^2 - \partial_\mu \phi^+ \partial_\mu \phi^- - \frac{1}{2} \partial_\mu \phi^0 \partial_\mu \phi^0 - \\
& \beta_h \left( \frac{2M^2}{g^2} + \frac{2M}{g} H + \frac{1}{2}(H^2 + \phi^0 \phi^0 + 2\phi^+ \phi^-) \right) + \frac{2M^4}{g^2} \alpha_h - \\
& g \alpha_h M (H^3 + H \phi^0 \phi^0 + 2H \phi^+ \phi^-) - \\
& \frac{1}{8} g^2 \alpha_h (H^4 + (\phi^0)^4 + 4(\phi^+ \phi^-)^2 + 4(\phi^0)^2 \phi^+ \phi^- + 4H^2 \phi^+ \phi^- + 2(\phi^0)^2 H^2) - \\
& g M W_\mu^+ W_\mu^- H - \frac{1}{2} g \frac{M}{c_w^2} Z_\mu^0 Z_\mu^0 H - \\
& \frac{1}{2} i g (W_\mu^+ (\phi^0 \partial_\mu \phi^- - \phi^- \partial_\mu \phi^0) - W_\mu^- (\phi^0 \partial_\mu \phi^+ - \phi^+ \partial_\mu \phi^0)) + \\
& \frac{1}{2} g (W_\mu^+ (H \partial_\mu \phi^- - \phi^- \partial_\mu H) + W_\mu^- (H \partial_\mu \phi^+ - \phi^+ \partial_\mu H)) + \frac{1}{2} g \frac{1}{c_w} (Z_\mu^0 (H \partial_\mu \phi^0 - \phi^0 \partial_\mu H) + \\
& M (\frac{1}{c_w} Z_\mu^0 \partial_\mu \phi^0 + W_\mu^+ \partial_\mu \phi^- + W_\mu^- \partial_\mu \phi^+) - ig \frac{s_w^2}{c_w} M Z_\mu^0 (W_\mu^+ \phi^- - W_\mu^- \phi^+) + ig s_w M A_\mu (W_\mu^+ \phi^- - \\
& W_\mu^- \phi^+) - ig \frac{1-2c_w^2}{2c_w} Z_\mu^0 (\phi^+ \partial_\mu \phi^- - \phi^- \partial_\mu \phi^+) + ig s_w A_\mu (\phi^+ \partial_\mu \phi^- - \phi^- \partial_\mu \phi^+) - \\
& \frac{1}{4} g^2 W_\mu^+ W_\mu^- (H^2 + (\phi^0)^2 + 2\phi^+ \phi^-) - \frac{1}{8} g^2 \frac{1}{c_w^2} Z_\mu^0 Z_\mu^0 (H^2 + (\phi^0)^2 + 2(2s_w^2 - 1)^2 \phi^+ \phi^-) - \\
& \frac{1}{2} g^2 \frac{s_w^2}{c_w} Z_\mu^0 \phi^0 (W_\mu^+ \phi^- - W_\mu^- \phi^+) - \frac{1}{2} i g^2 \frac{s_w^2}{c_w} Z_\mu^0 H (W_\mu^+ \phi^- - W_\mu^- \phi^+) + \frac{1}{2} g^2 s_w A_\mu \phi^0 (W_\mu^+ \phi^- + \\
& W_\mu^- \phi^+) + \frac{1}{2} i g^2 s_w A_\mu H (W_\mu^+ \phi^- - W_\mu^- \phi^+) - g^2 \frac{s_w}{c_w} (2c_w^2 - 1) Z_\mu^0 A_\mu \phi^+ \phi^- - \\
& g^2 s_w^2 A_\mu A_\mu \phi^+ \phi^- + \frac{1}{2} i g_s \lambda_{ij} (\bar{q}_i^\sigma \gamma^\mu q_j^\sigma) g_a^a - \bar{e}^\lambda (\gamma \partial + m_e^\lambda) e^\lambda - \bar{\nu}^\lambda (\gamma \partial + m_\nu^\lambda) \nu^\lambda - \bar{u}^\lambda (\gamma \partial + \\
& m_u^\lambda) u^\lambda - \bar{d}_j^\lambda (\gamma \partial + m_d^\lambda) d_j^\lambda + ig s_w A_\mu ((-\bar{e}^\lambda \gamma^\mu e^\lambda) + \frac{2}{3} (\bar{u}_j^\lambda \gamma^\mu u_j^\lambda) - \frac{1}{3} (\bar{d}_j^\lambda \gamma^\mu d_j^\lambda)) + \\
& \frac{ig}{4c_w} Z_\mu^0 \{(\bar{\nu}^\lambda \gamma^\mu (1 + \gamma^5) \nu^\lambda) + (\bar{e}^\lambda \gamma^\mu (4s_w^2 - 1 - \gamma^5) e^\lambda) + (\bar{d}_j^\lambda \gamma^\mu (\frac{4}{3}s_w^2 - 1 - \gamma^5) d_j^\lambda) + \\
& (\bar{u}_j^\lambda \gamma^\mu (1 - \frac{8}{3}s_w^2 + \gamma^5) u_j^\lambda)\} + \frac{ig}{2\sqrt{2}} W_\mu^+ ((\bar{\nu}^\lambda \gamma^\mu (1 + \gamma^5) U^{lep}{}_{\lambda\kappa} e^\kappa) + (\bar{u}_j^\lambda \gamma^\mu (1 + \gamma^5) C_{\lambda\kappa} d_j^\kappa)) + \\
& \frac{ig}{2\sqrt{2}} W_\mu^- ((\bar{e}^\kappa U^{lep\dagger}{}_{\kappa\lambda} \gamma^\mu (1 + \gamma^5) \nu^\lambda) + (\bar{d}_j^\kappa C_{\kappa\lambda}^\dagger \gamma^\mu (1 + \gamma^5) u_j^\lambda)) + \\
& \frac{ig}{2M\sqrt{2}} \phi^+ (-m_e^\kappa (\bar{\nu}^\lambda U^{lep}{}_{\lambda\kappa} (1 - \gamma^5) e^\kappa) + m_\nu^\kappa (\bar{\nu}^\lambda U^{lep}{}_{\lambda\kappa} (1 + \gamma^5) e^\kappa) + \\
& \frac{ig}{2M\sqrt{2}} \phi^- (m_e^\lambda (\bar{e}^\lambda U^{lep\dagger}{}_{\lambda\kappa} (1 + \gamma^5) \nu^\kappa) - m_\nu^\kappa (\bar{e}^\lambda U^{lep\dagger}{}_{\lambda\kappa} (1 - \gamma^5) \nu^\kappa) - \frac{g}{2} \frac{m_e^\lambda}{M} H (\bar{\nu}^\lambda \nu^\lambda) - \\
& \frac{g}{2} \frac{m_\nu^\lambda}{M} H (\bar{e}^\lambda e^\lambda) + \frac{ig}{2} \frac{m_e^\lambda}{M} \phi^0 (\bar{\nu}^\lambda \gamma^5 \nu^\lambda) - \frac{ig}{2} \frac{m_\nu^\lambda}{M} \phi^0 (\bar{e}^\lambda \gamma^5 e^\lambda) - \frac{1}{4} \bar{\nu}_\lambda M_{\lambda\kappa}^R (1 - \gamma_5) \bar{\nu}_\kappa - \\
& \frac{1}{4} \bar{\nu}_\lambda M_{\lambda\kappa}^R (1 - \gamma_5) \bar{\nu}_\kappa + \frac{ig}{2M\sqrt{2}} \phi^+ (-m_d^\kappa (\bar{u}_j^\lambda C_{\lambda\kappa} (1 - \gamma^5) d_j^\kappa) + m_u^\lambda (\bar{u}_j^\lambda C_{\lambda\kappa} (1 + \gamma^5) d_j^\kappa) + \\
& \frac{ig}{2M\sqrt{2}} \phi^- (m_d^\lambda (\bar{d}_j^\lambda C_{\lambda\kappa}^\dagger (1 + \gamma^5) u_j^\kappa) - m_u^\kappa (\bar{d}_j^\lambda C_{\lambda\kappa}^\dagger (1 - \gamma^5) u_j^\kappa) - \frac{g}{2} \frac{m_e^\lambda}{M} H (\bar{u}_j^\lambda u_j^\lambda) - \\
& \frac{g}{2} \frac{m_\nu^\lambda}{M} H (\bar{d}_j^\lambda d_j^\lambda) + \frac{ig}{2} \frac{m_e^\lambda}{M} \phi^0 (\bar{u}_j^\lambda \gamma^5 u_j^\lambda) - \frac{ig}{2} \frac{m_\nu^\lambda}{M} \phi^0 (\bar{d}_j^\lambda \gamma^5 d_j^\lambda) + \bar{G}^a \partial^2 G^a + g_s f^{abc} \partial_\mu \bar{G}^a G^b g_c^+ + \\
& \bar{X}^+ (\partial^2 - M^2) X^+ + \bar{X}^- (\partial^2 - M^2) X^- + \bar{X}^0 (\partial^2 - \frac{M^2}{c_w^2}) X^0 + \bar{Y} \partial^2 Y + ig c_w W_\mu^+ (\partial_\mu \bar{X}^0 X^- - \\
& \partial_\mu \bar{X}^+ X^0) + ig s_w W_\mu^+ (\partial_\mu \bar{Y} X^- - \partial_\mu \bar{X}^+ Y) + ig c_w W_\mu^- (\partial_\mu \bar{X}^- X^0 - \\
& \partial_\mu \bar{X}^0 X^+) + ig s_w W_\mu^- (\partial_\mu \bar{X}^- Y - \partial_\mu \bar{Y} X^+) + ig c_w Z_\mu^0 (\partial_\mu \bar{X}^+ X^+ - \\
& \partial_\mu \bar{X}^- X^-) + ig s_w A_\mu (\partial_\mu \bar{X}^+ X^+ - \\
& \partial_\mu \bar{X}^- X^-) - \frac{1}{2} g M \left( \bar{X}^+ X^+ H + \bar{X}^- X^- H + \frac{1}{c_w^2} \bar{X}^0 X^0 H \right) + \frac{1-2c_w^2}{2c_w} ig M (\bar{X}^+ X^0 \phi^+ - \bar{X}^- X^0 \phi^-) + \\
& \frac{1}{2c_w} ig M (\bar{X}^0 X^- \phi^+ - \bar{X}^0 X^+ \phi^-) + ig M s_w (\bar{X}^0 X^- \phi^+ - \bar{X}^0 X^+ \phi^-) + \\
& \frac{1}{2} ig M (\bar{X}^+ X^+ \phi^0 - \bar{X}^- X^- \phi^0) .
\end{aligned}$$



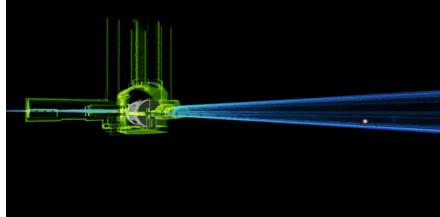




## SM with parameters $\theta$

## Simulated observables $x$

Real observations  $x_{\text{obs}}$



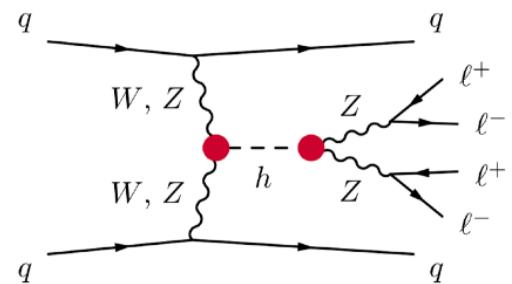
Latent variables

Parameters  
of interest

Parton-level  
momenta

Theory  
parameters

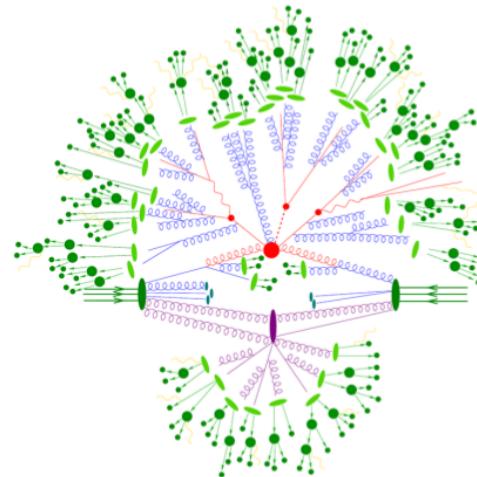
$$z_p \leftarrow \theta$$



Latent variables      Parameters  
of interest

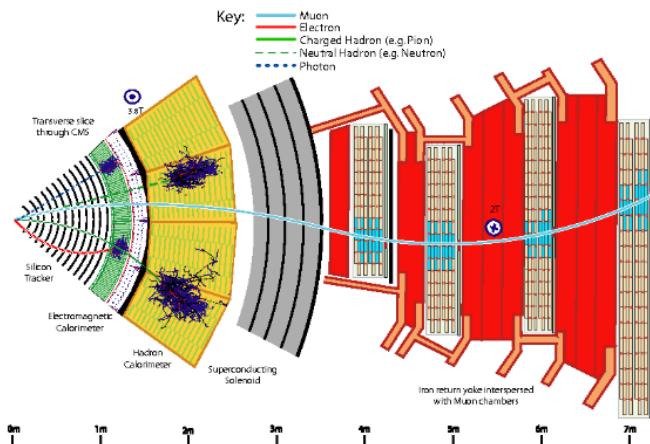
Shower      Parton-level      Theory  
splittings      momenta      parameters

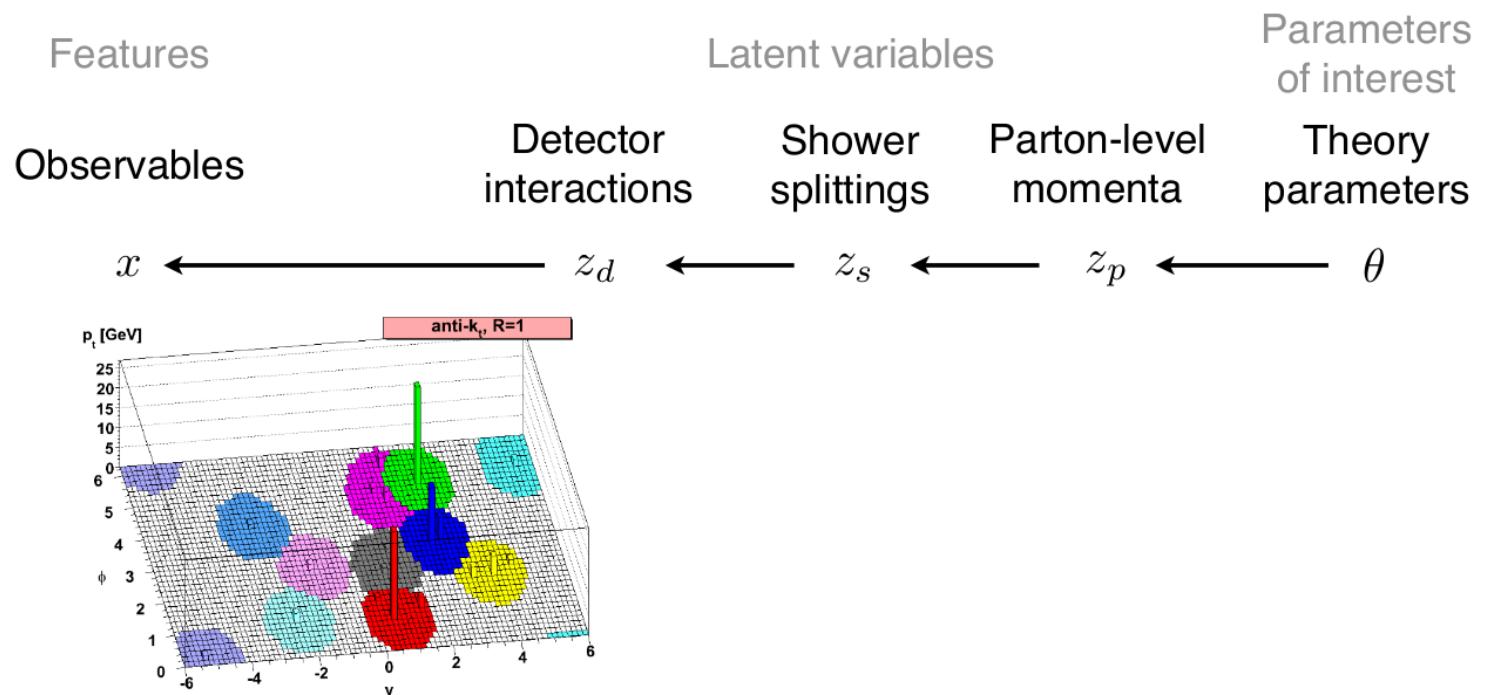
$$z_s \leftarrow z_p \leftarrow \theta$$



Latent variables		Parameters of interest	
Detector interactions	Shower splittings	Parton-level momenta	Theory parameters

$$z_d \leftarrow z_s \leftarrow z_p \leftarrow \theta$$





[Image source: M. Cacciari,  
G. Salam, G. Soyez 0802.1189]

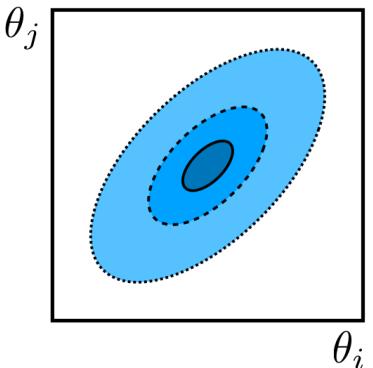
$$p(x|\theta) = \underbrace{\iiint}_{\text{intractable}} p(z_p|\theta)p(z_s|z_p)p(z_d|z_s)p(x|z_d)dz_p dz_s dz_d$$

# Inference, the frequentist (physicist's) way

The Neyman-Pearson lemma states that the likelihood ratio

$$r(x|\theta_0, \theta_1) = \frac{p(x|\theta_0)}{p(x|\theta_1)}$$

is the **most powerful test statistic** to discriminate between a null hypothesis  $\theta_0$  and an alternative  $\theta_1$ .



## IX. On the Problem of the most Efficient Tests of Statistical Hypotheses.

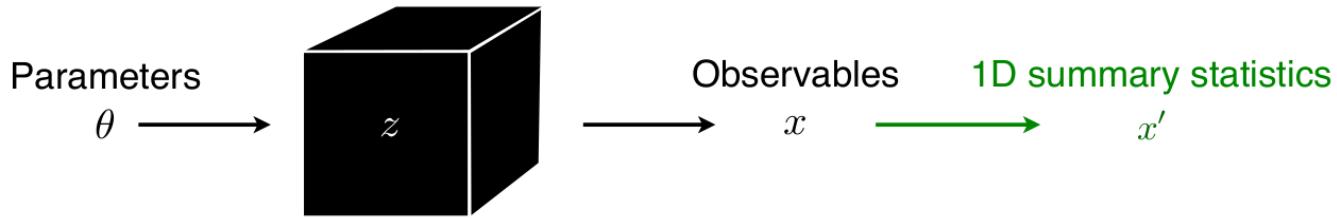
By J. NEYMAN, Nencki Institute, Soc. Sci. Lit. Varsoviensis, and Lecturer at the Central College of Agriculture, Warsaw, and E. S. PEARSON, Department of Applied Statistics, University College, London.

(Communicated by K. PEARSON, F.R.S.)

(Received August 31, 1932.—Read November 10, 1932.)

### CONTENTS.

	PAGE.
I. Introductory . . . . .	289
II. Outline of General Theory . . . . .	293
III. Simple Hypotheses . . . . .	



Define a projection function  $s : \mathcal{X} \rightarrow \mathbb{R}$  mapping observables  $x$  to a summary statistic  $x' = s(x)$ .

Then, **approximate** the likelihood  $p(x|\theta)$  with the surrogate  $\hat{p}(x|\theta) = p(x'|\theta)$ .

From this it comes

$$\frac{p(x|\theta_0)}{p(x|\theta_1)} \approx \frac{\hat{p}(x|\theta_0)}{\hat{p}(x|\theta_1)} = \hat{r}(x|\theta_0, \theta_1).$$

## Wilks theorem

Consider the test statistic

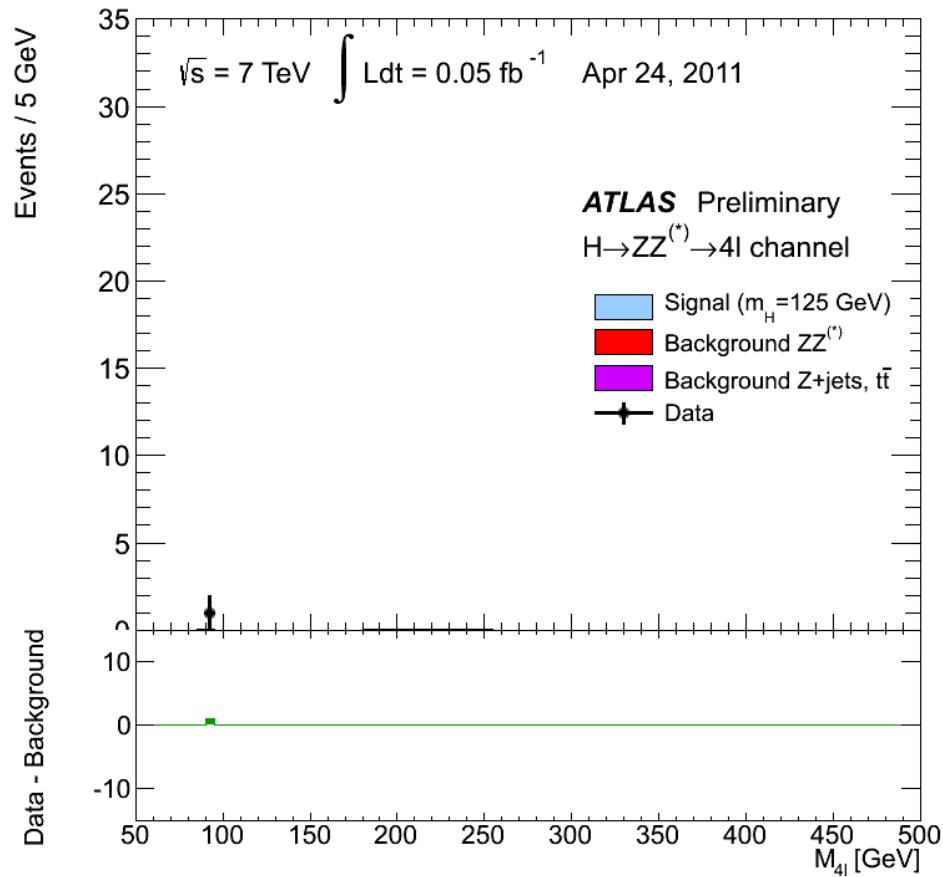
$$q(\theta) = -2 \sum_x \log \frac{p(x|\theta)}{p(x|\hat{\theta})} = -2 \sum_x \log r(x|\theta, \hat{\theta})$$

for a fixed number  $N$  of observations  $\{x\}$  and where  $\hat{\theta}$  is the maximum likelihood estimator.

When  $N \rightarrow \infty, q(\theta) \sim \chi_2$ .

Therefore (and provided the assumptions apply!), an observed value  $q_{\text{obs}}(\theta)$  translates directly to a p-value that measures the confidence with which  $\theta$  can be excluded:

$$p_\theta \equiv \int_{q_{\text{obs}}(\theta)}^{\infty} p(q|\theta) dq = 1 - F_{\chi_2}(q_{\text{obs}}(\theta)).$$



Discovery of the Higgs boson

# Inference algorithms

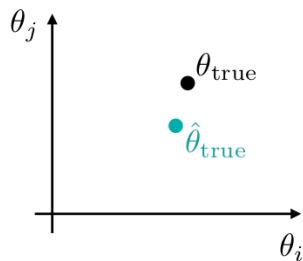
# Problem statement

Start with

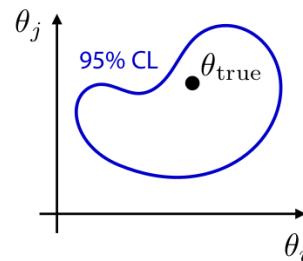
- a simulator that lets you generate  $N$  samples  $x_i \sim p(x_i | \theta_i)$ ,
- observed data  $x_{\text{obs}} \sim p(x_{\text{obs}} | \theta_{\text{true}})$ ,
- a prior  $p(\theta)$ .

Then,

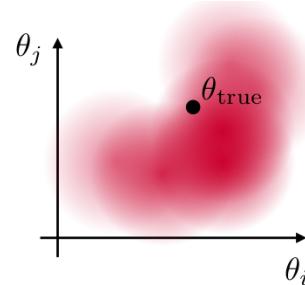
a) estimate  $\theta_{\text{true}}$   
(e.g., MLE)



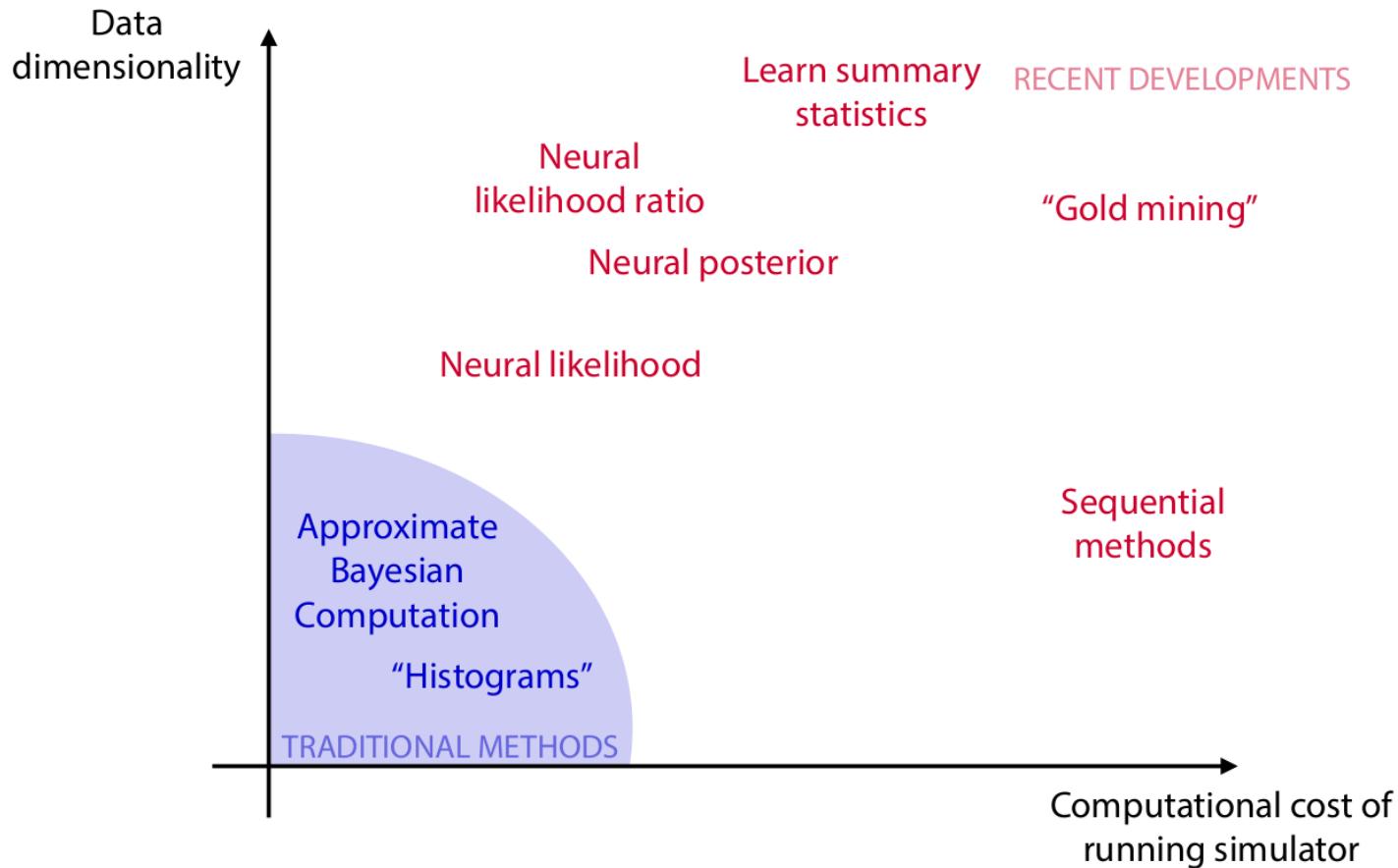
b) construct  
confidence sets



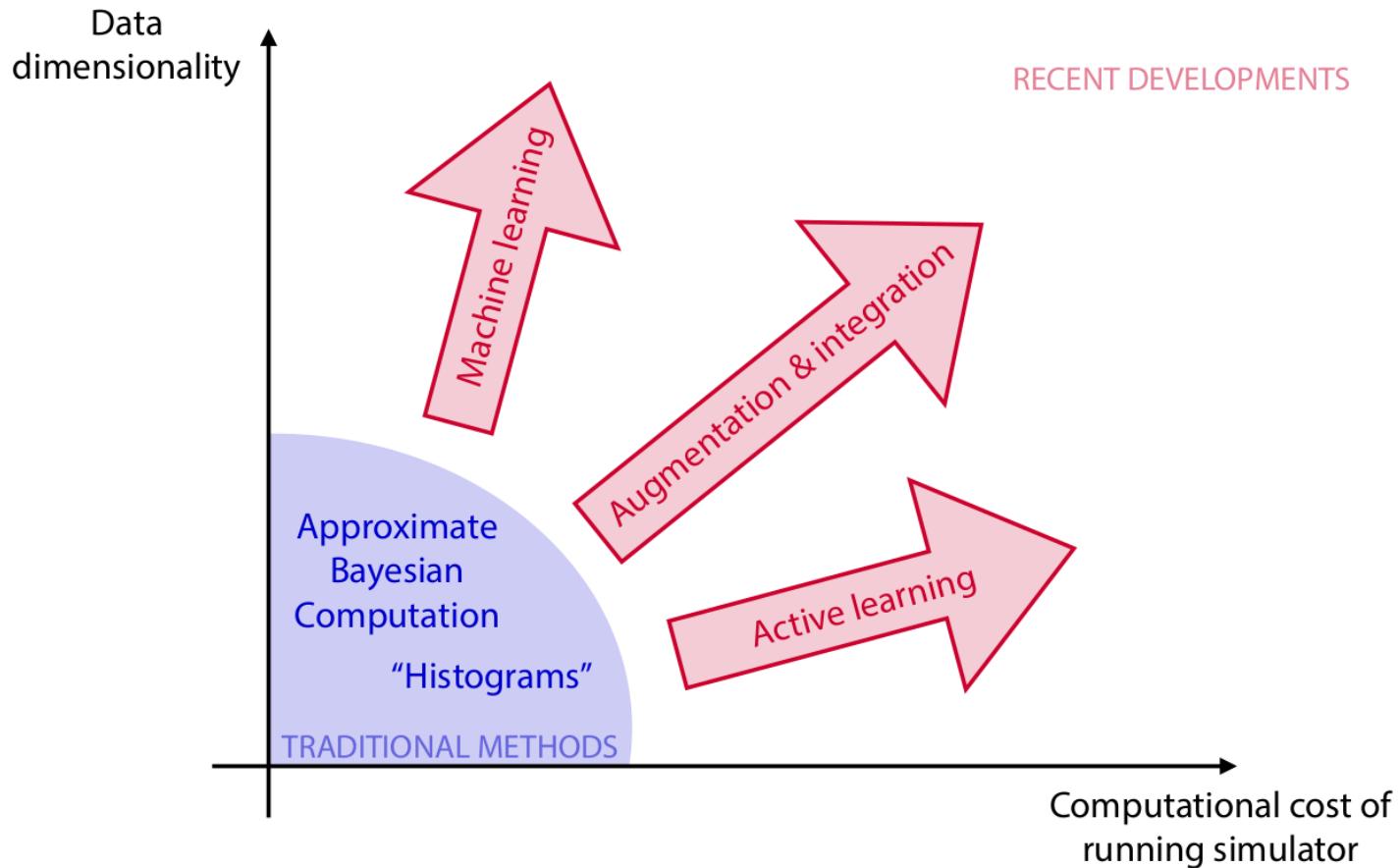
c) estimate the posterior  
 $p(\theta | x_{\text{obs}})$



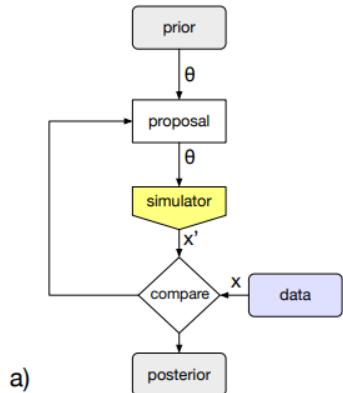
# Inference algorithms



# Inference algorithms

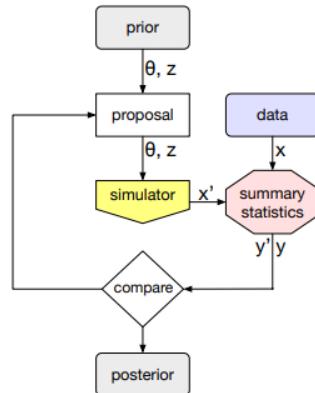


Approximate Bayesian Computation  
with Monte Carlo sampling



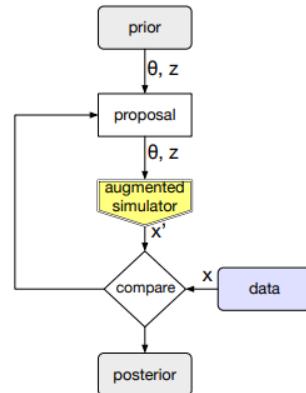
a)

Approximate Bayesian Computation  
with learned summary statistics



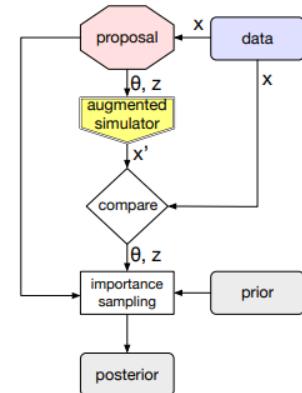
b)

Probabilistic Programming  
with Monte Carlo sampling



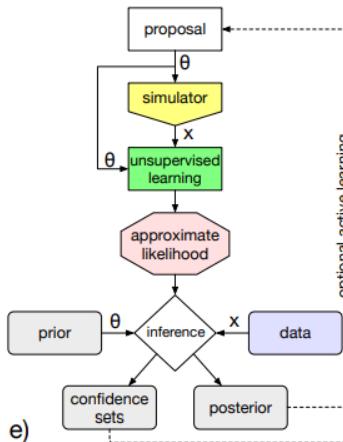
c)

Probabilistic Programming  
with Inference Compilation



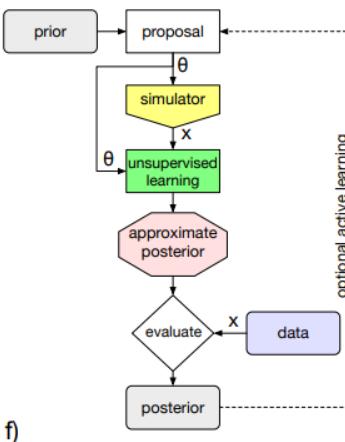
d)

Amortized likelihood



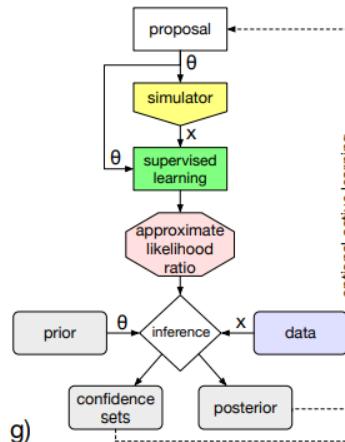
e)

Amortized posterior



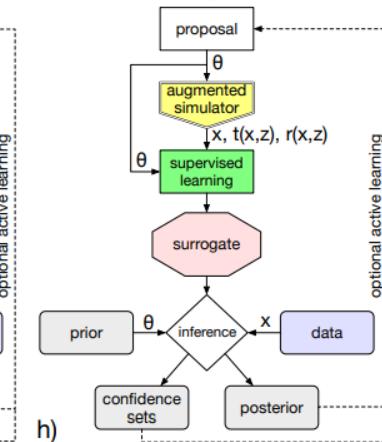
f)

Amortized likelihood ratio



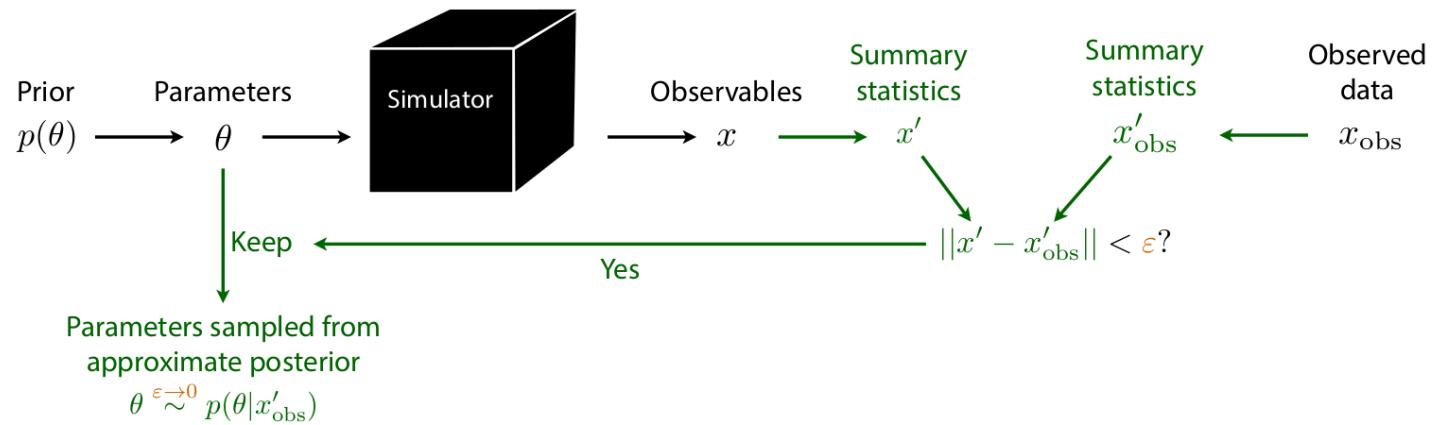
g)

Amortized surrogates  
trained with augmented data



h)

# Approximate Bayesian Computation (ABC)



## Issues

- How to choose  $x'?$   $\epsilon?$   $|| \cdot ||?$
- No tractable posterior.
- Need to run new simulations for new data or new prior.

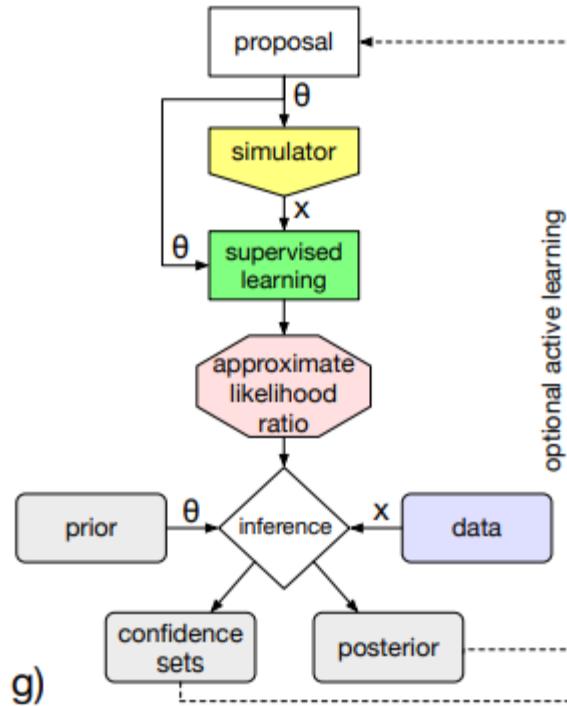
# Neural Ratio Estimation (NRE)

The Bayes rule can be rewritten as

$$\begin{aligned} p(\theta|x) &= \frac{p(x|\theta)p(\theta)}{p(x)} \\ &= r(x|\theta)p(\theta) \\ &\approx \hat{r}(x|\theta)p(\theta) \end{aligned}$$

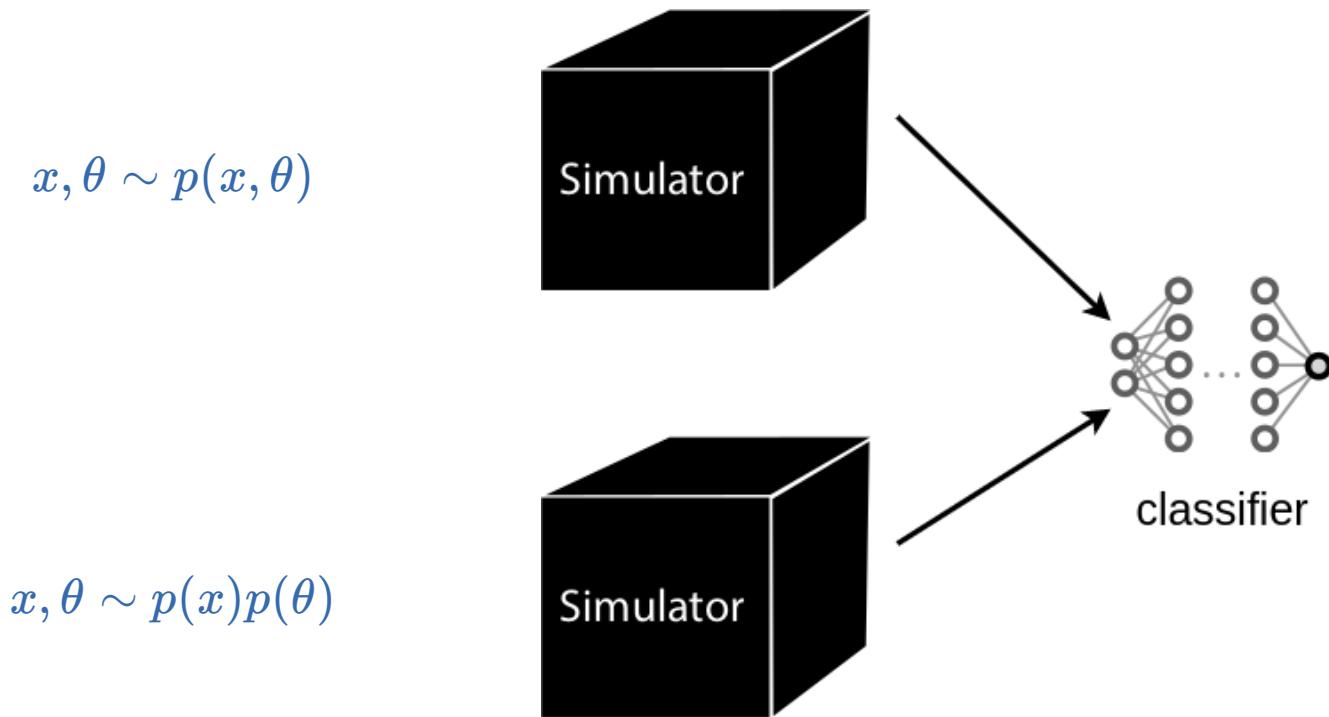
where  $r(x|\theta) = \frac{p(x|\theta)}{p(x)}$  is the likelihood-to-evidence ratio.

**Amortized likelihood ratio**



## The likelihood ratio trick

The ratio can be learned with machine learning, even if neither the likelihood nor the evidence can be evaluated!



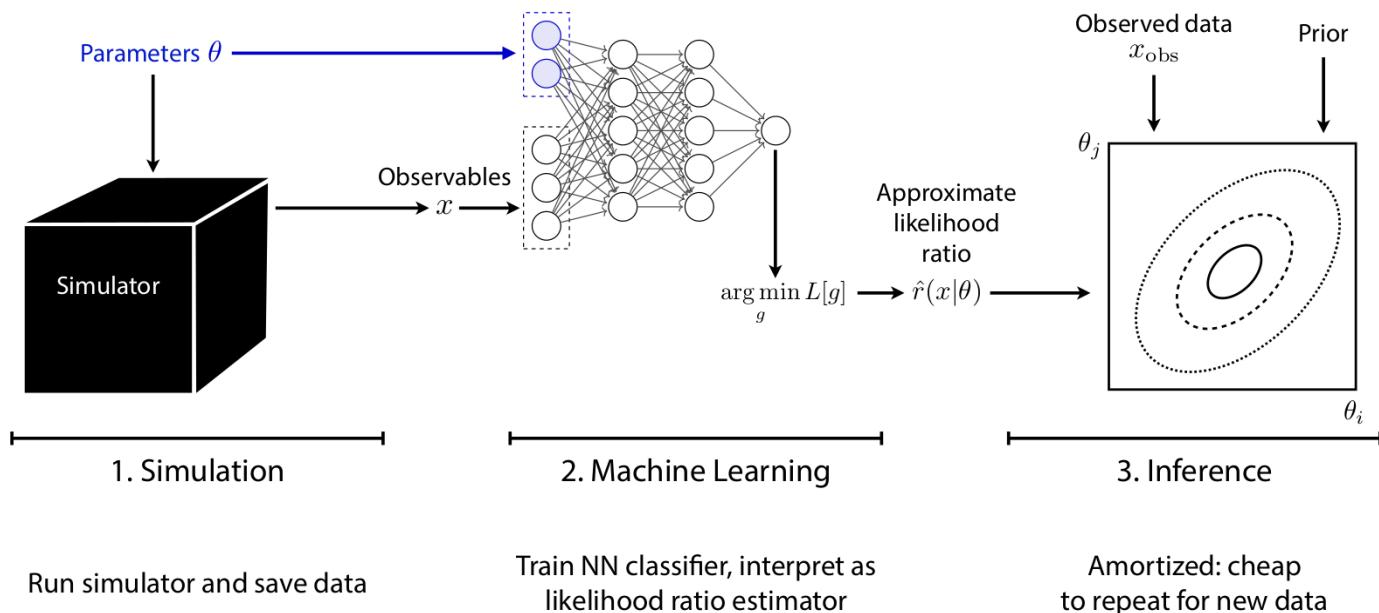
The solution  $\hat{d}$  found after training approximates the optimal classifier

$$d(x, \theta) \approx d^*(x, \theta) = \frac{p(x, \theta)}{p(x, \theta) + p(x)p(\theta)}.$$

Therefore,

$$r(x|\theta) = \frac{p(x|\theta)}{p(x)} = \frac{p(x, \theta)}{p(x)p(\theta)} \approx \frac{d(x, \theta)}{1 - d(x, \theta)} = \hat{r}(x|\theta).$$

## Inference

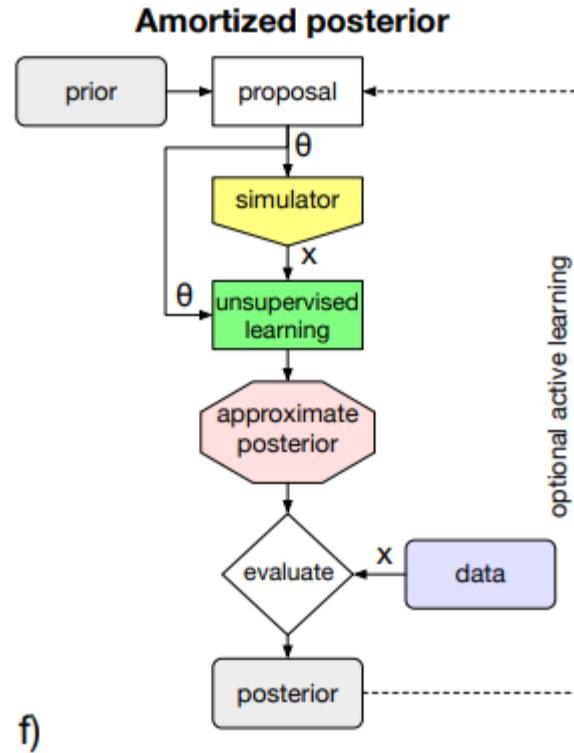


# Neural Posterior Estimation (NPE)

Use variational inference to directly estimate the posterior:

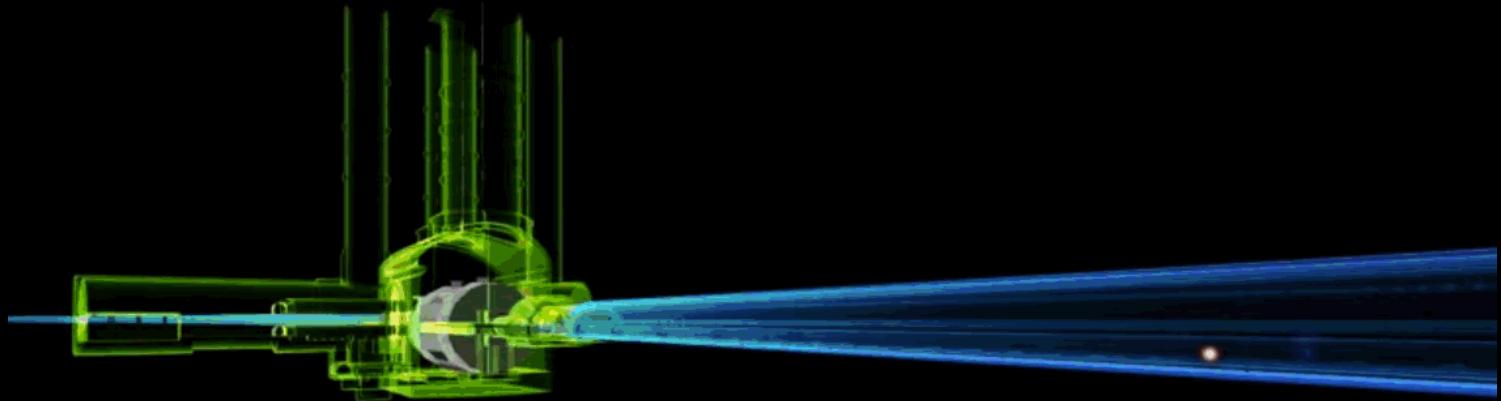
$$\min_{q_\phi} \mathbb{E}_{p(x)} [\text{KL}(p(\theta|x) || q_\phi(\theta|x))]$$

where  $q_\phi$  is a neural density estimator, such as a normalizing flow.

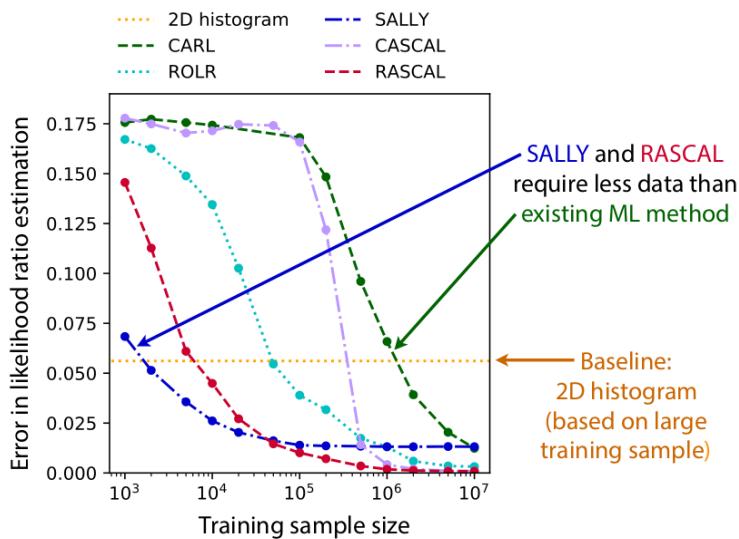


f)

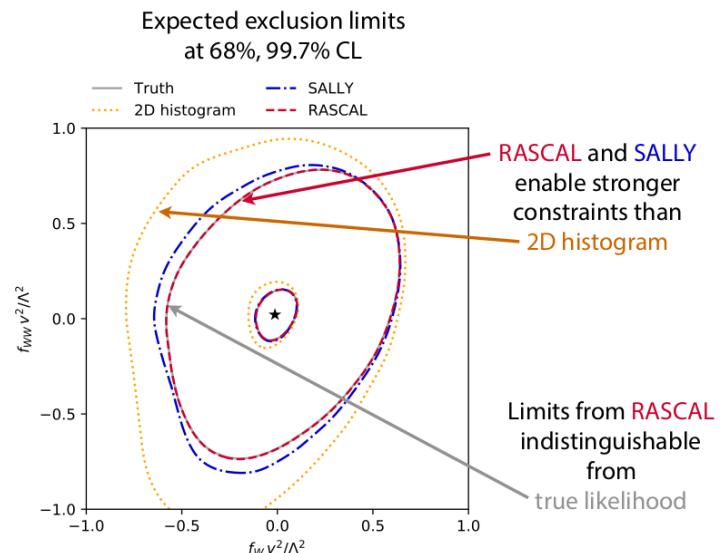
# Showtime!



## Case 1: Hunting new physics at particle colliders



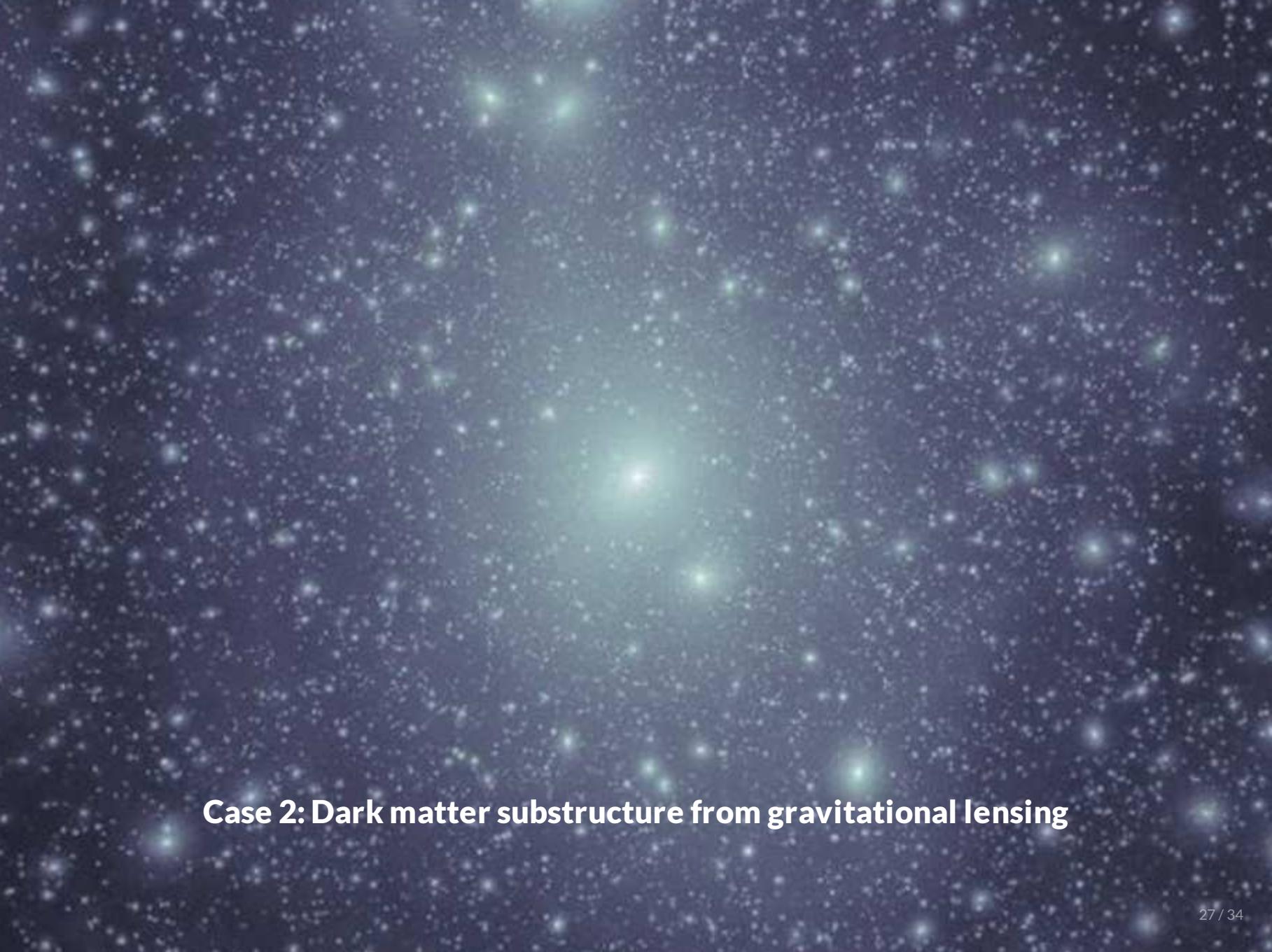
SALLY and RASCAL require less data than existing ML method  
Baseline:  
2D histogram (based on large training sample)



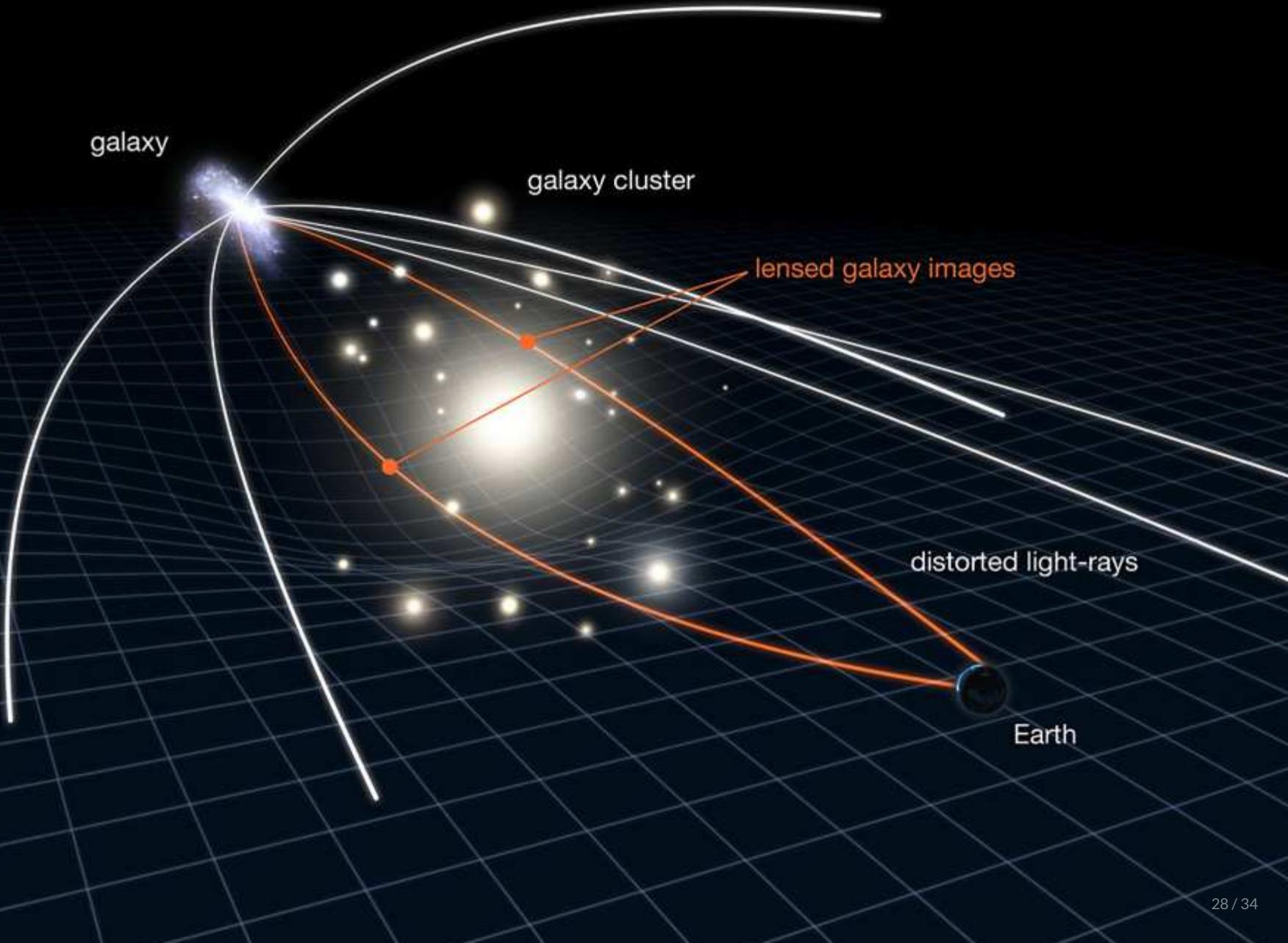
RASCAL and SALLY enable stronger constraints than 2D histogram  
Limits from RASCAL indistinguishable from true likelihood

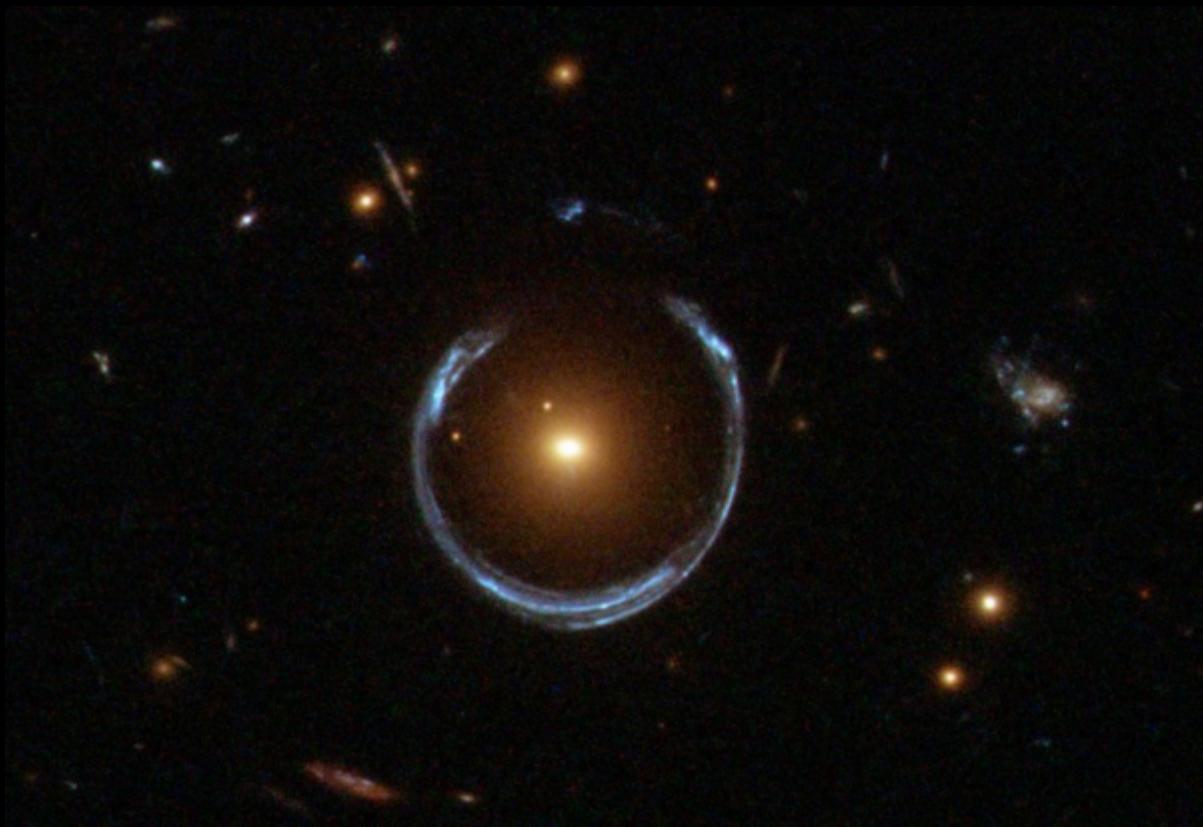
With enough training data, NRE gets the likelihood-ratio statistic right.

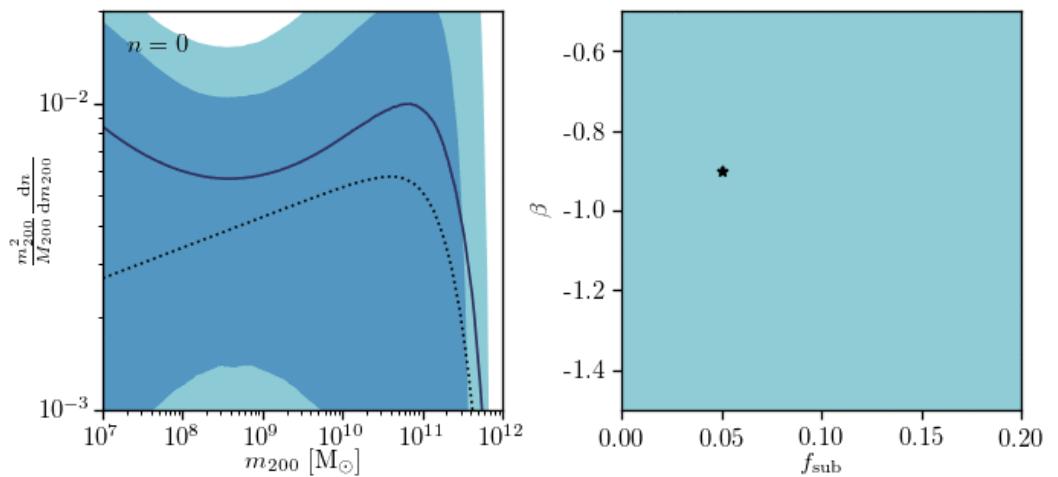
Using more information from the simulator improves sample efficiency substantially.



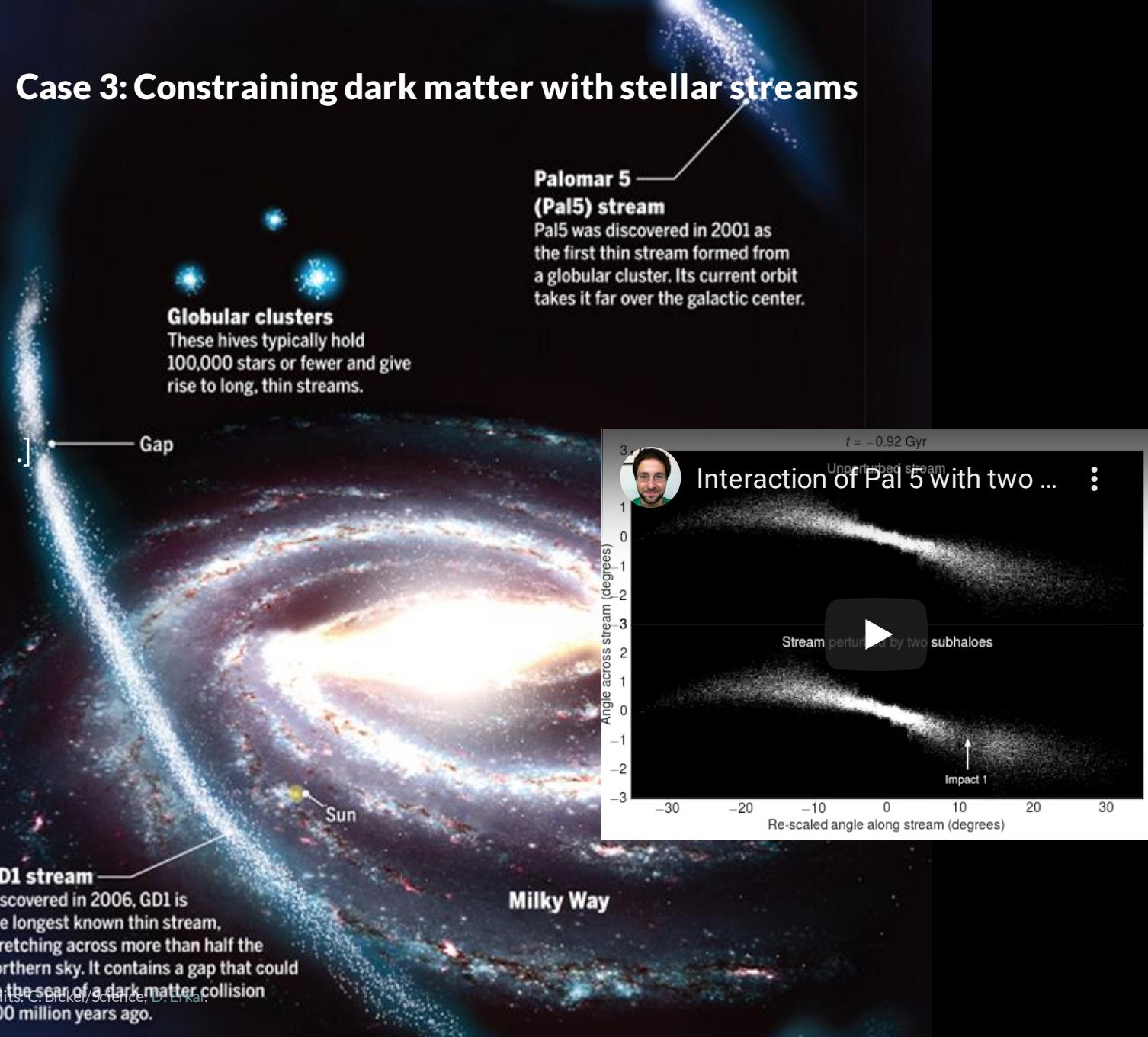
**Case 2: Dark matter substructure from gravitational lensing**

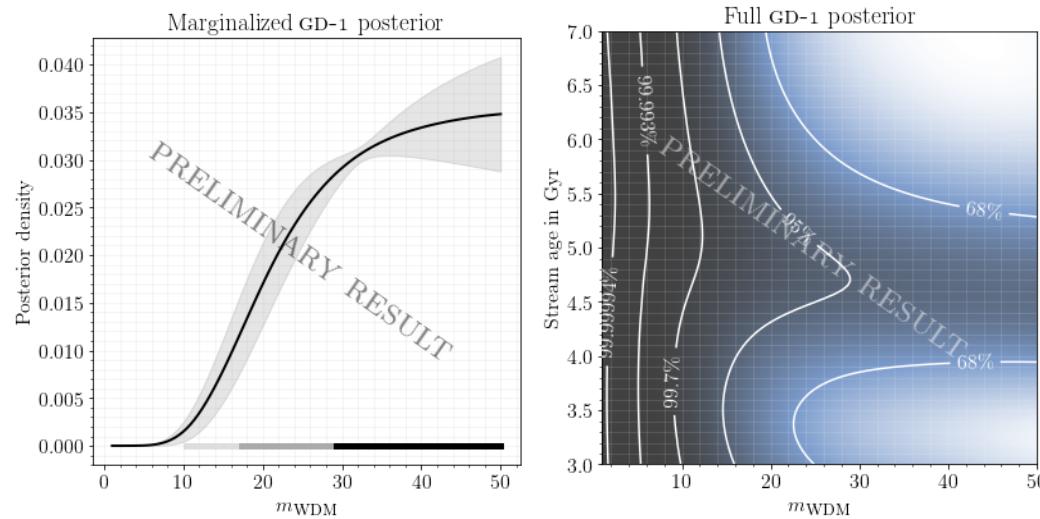
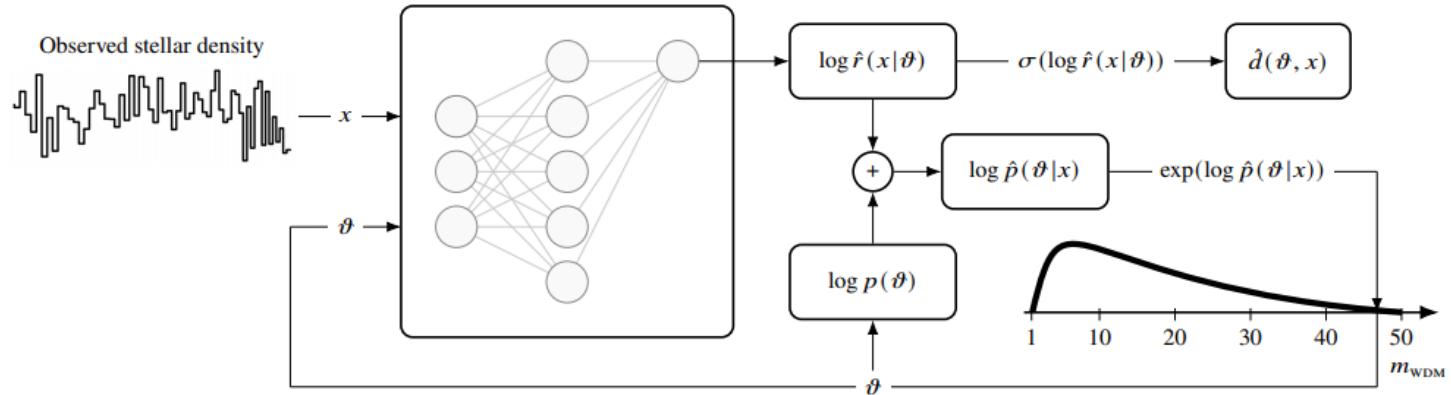






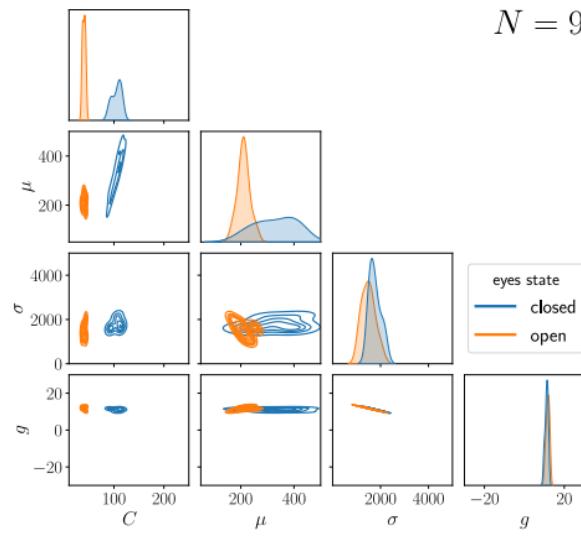
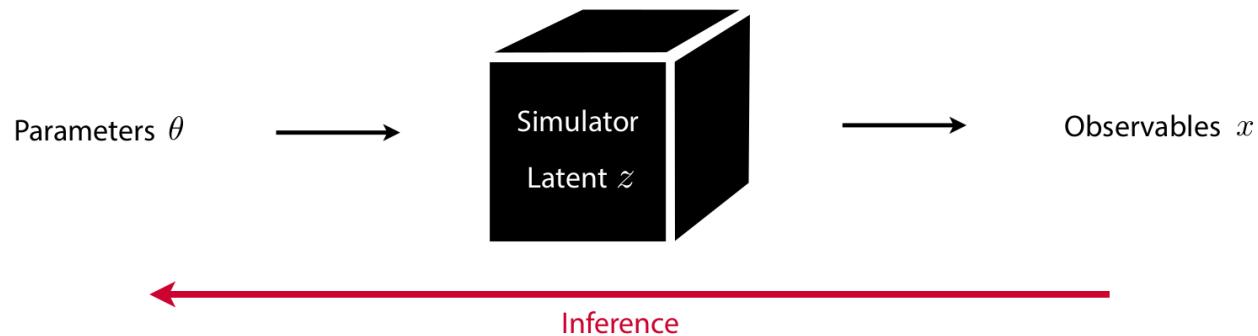
## Case 3: Constraining dark matter with stellar streams



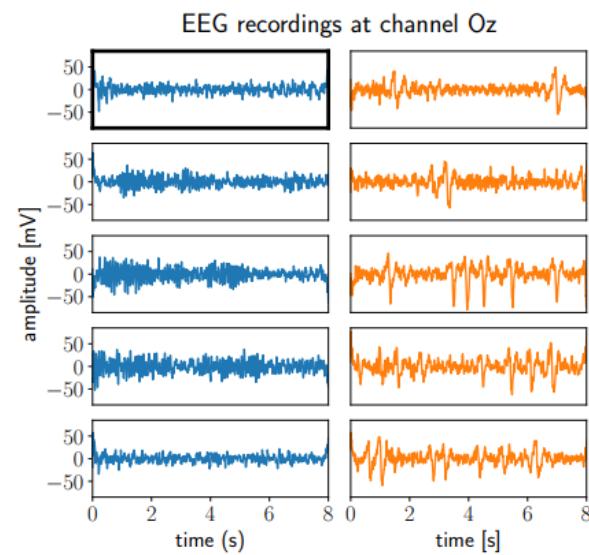


Preliminary results for GD-1 suggest a **preference for CDM over WDM**.

## Case 4: Inverting a neural mass model (Jansen and Rit)



$N = 9$



# Summary

Simulation-based inference is a major evolution in the statistical capabilities for science, enabled by advances in machine learning.

Need to efficiently generate simulated data and use it to train ML components.

The end.