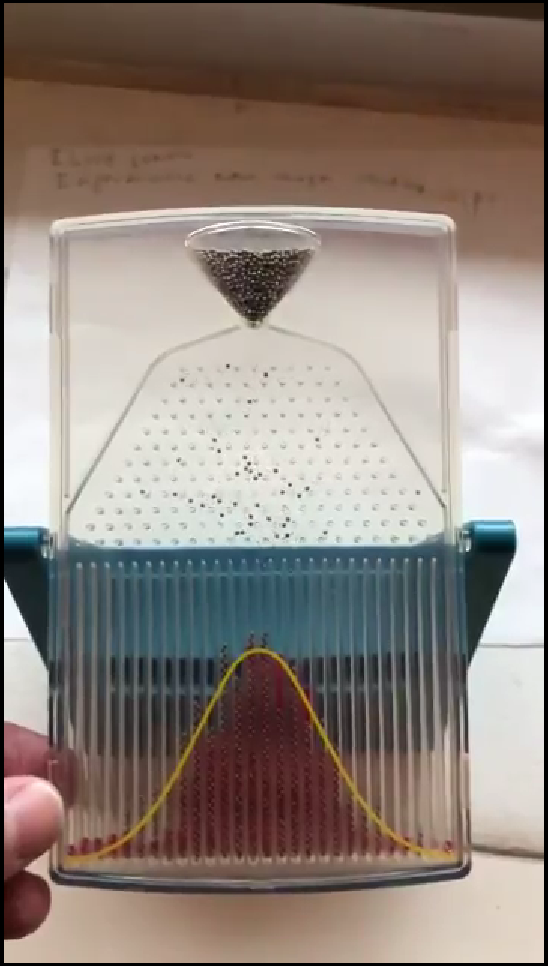
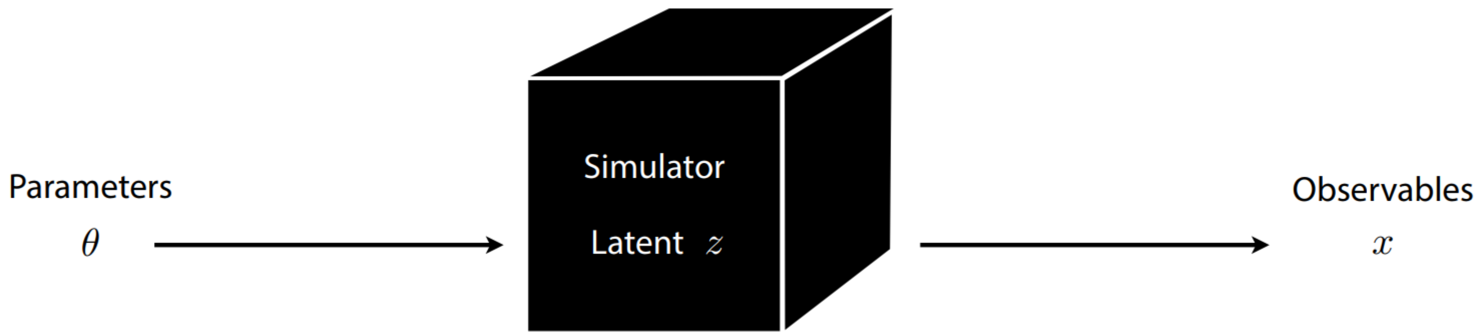
A short introduction to Neural Likelihood-free Inference for Physics

AMLD 2020
January 28, Lausanne, Switzerland

Gilles Louppe
g.louppe@uliege.be

A typical science experiment



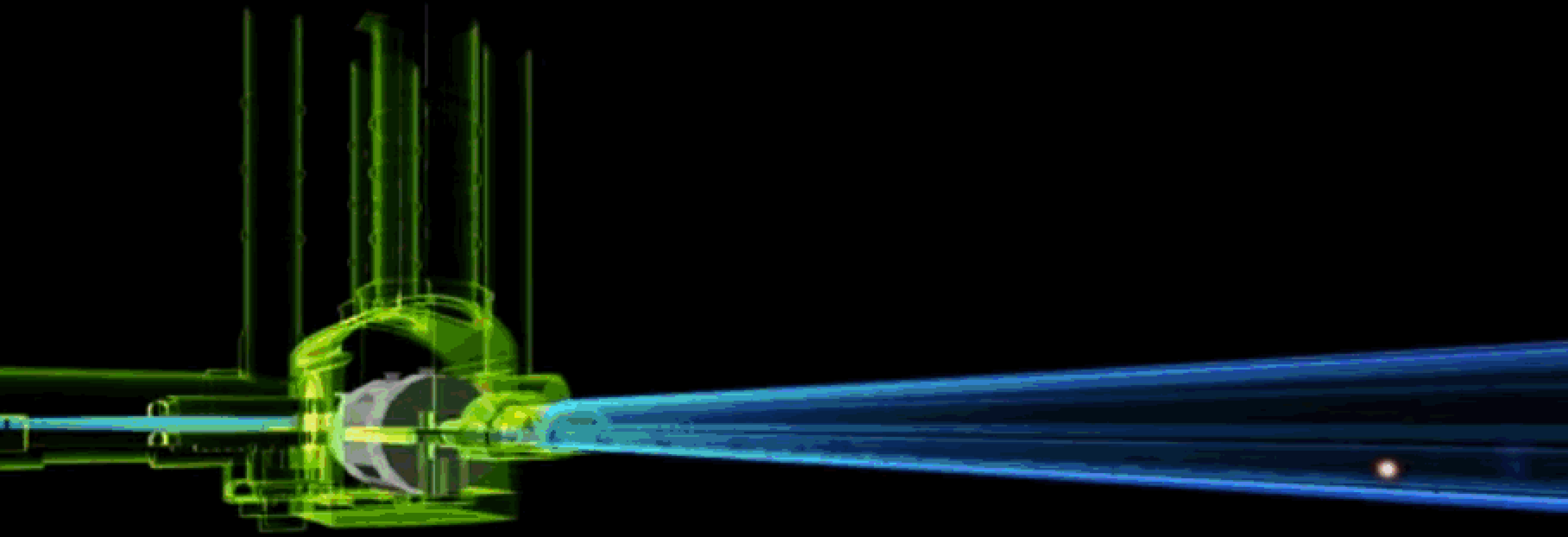


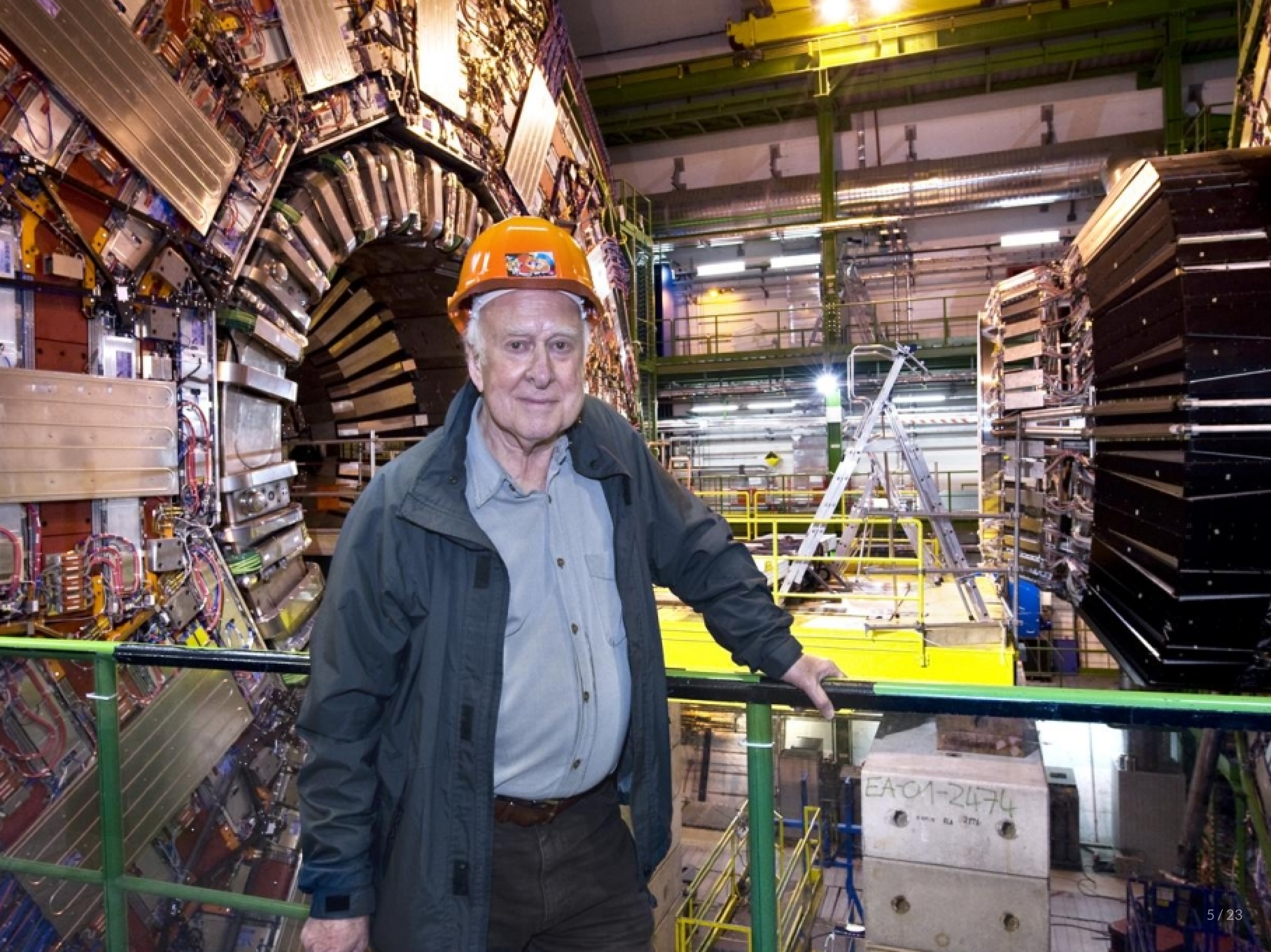
Prediction:
- Well-understood mechanistic model
- Simulator can generate samples

Inference:
- Likelihood function $p(x|\theta)$ is intractable
- Inference based on estimator $\hat{p}(x|\theta)$

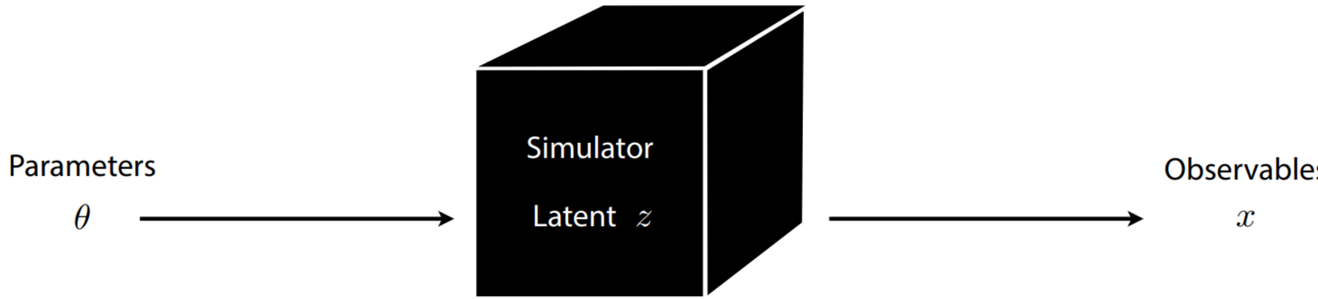
Credits: Johann Brehmer.

$$
\begin{gathered}
\mathcal{L}_{SM} = -\tfrac{1}{2}\partial_\nu g^a_\mu \partial_\nu g^a_\mu - g_s f^{abc}\partial_\mu g^a_\nu g^b_\mu g^c_\nu - \tfrac{1}{4}g_s^2 f^{abc}f^{ade}g^b_\mu g^c_\nu g^d_\mu g^e_\nu - \partial_\nu W^+_\mu \partial_\nu W^-_\mu - \\
M^2 W^+_\mu W^-_\mu - \tfrac{1}{2}\partial_\nu Z^0_\mu \partial_\nu Z^0_\mu - \tfrac{1}{2c_w^2}M^2 Z^0_\mu Z^0_\mu - \tfrac{1}{2}\partial_\mu A_\nu \partial_\mu A_\nu - igc_w(\partial_\nu Z^0_\mu(W^+_\mu W^-_\nu - \\
W^+_\nu W^-_\mu) - Z^0_\nu(W^+_\mu\partial_\nu W^-_\mu - W^-_\mu\partial_\nu W^+_\mu) + Z^0_\mu(W^+_\nu\partial_\nu W^-_\mu - W^-_\nu\partial_\nu W^+_\mu)) - \\
igs_w(\partial_\nu A_\mu(W^+_\mu W^-_\nu - W^+_\nu W^-_\mu) - A_\nu(W^+_\mu\partial_\nu W^-_\mu - W^-_\mu\partial_\nu W^+_\mu) + A_\mu(W^+_\nu\partial_\nu W^-_\mu - \\
W^-_\nu\partial_\nu W^+_\mu)) - \tfrac{1}{2}g^2 W^+_\mu W^-_\mu W^+_\nu W^-_\nu + \tfrac{1}{2}g^2 W^+_\mu W^-_\nu W^+_\mu W^-_\nu + g^2c_w^2(Z^0_\mu W^+_\mu Z^0_\nu W^-_\nu - \\
Z^0_\mu Z^0_\mu W^+_\nu W^-_\nu) + g^2 s_w^2(A_\mu W^+_\mu A_\nu W^-_\nu - A_\mu A_\mu W^+_\nu W^-_\nu) + g^2 s_w c_w(A_\mu Z^0_\nu(W^+_\mu W^-_\nu - \\
W^+_\nu W^-_\mu) - 2A_\mu Z^0_\mu W^+_\nu W^-_\nu) - \tfrac{1}{2}\partial_\mu H\partial_\mu H - 2M^2\alpha_h H^2 - \partial_\mu\phi^+\partial_\mu\phi^- - \tfrac{1}{2}\partial_\mu\phi^0\partial_\mu\phi^0 - \\
\beta_h\left(\tfrac{2M^2}{g^2} + \tfrac{2M}{g}H + \tfrac{1}{2}(H^2+\phi^0\phi^0+2\phi^+\phi^-)\right) + \tfrac{2M^4}{g^2}\alpha_h - \\
g\alpha_h M\left(H^3 + H\phi^0\phi^0 + 2H\phi^+\phi^-\right) - \\
\tfrac{1}{8}g^2\alpha_h\left(H^4 + (\phi^0)^4 + 4(\phi^+\phi^-)^2 + 4(\phi^0)^2\phi^+\phi^- + 4H^2\phi^+\phi^- + 2(\phi^0)^2H^2\right) - \\
gMW^+_\mu W^-_\mu H - \tfrac{1}{2}g\tfrac{M}{c_w^2}Z^0_\mu Z^0_\mu H - \\
\tfrac{1}{2}ig\left(W^+_\mu(\phi^0\partial_\mu\phi^- - \phi^-\partial_\mu\phi^0) - W^-_\mu(\phi^0\partial_\mu\phi^+ - \phi^+\partial_\mu\phi^0)\right) + \\
\tfrac{1}{2}g\left(W^+_\mu(H\partial_\mu\phi^- - \phi^-\partial_\mu H) + W^-_\mu(H\partial_\mu\phi^+ - \phi^+\partial_\mu H)\right) + \tfrac{1}{2}g\tfrac{1}{c_w}(Z^0_\mu(H\partial_\mu\phi^0 - \phi^0\partial_\mu H) + \\
M\left(\tfrac{1}{c_w}Z^0_\mu\partial_\mu\phi^0 + W^+_\mu\partial_\mu\phi^- + W^-_\mu\partial_\mu\phi^+\right) - ig\tfrac{s_w^2}{c_w}MZ^0_\mu(W^+_\mu\phi^- - W^-_\mu\phi^+) + igs_wMA_\mu(W^+_\mu\phi^- - \\
W^-_\mu\phi^+) - ig\tfrac{1-2c_w^2}{2c_w}Z^0_\mu(\phi^+\partial_\mu\phi^- - \phi^-\partial_\mu\phi^+) + igs_wA_\mu(\phi^+\partial_\mu\phi^- - \phi^-\partial_\mu\phi^+) - \\
\tfrac{1}{4}g^2W^+_\mu W^-_\mu\left(H^2+(\phi^0)^2+2\phi^+\phi^-\right) - \tfrac{1}{8}g^2\tfrac{1}{c_w^2}Z^0_\mu Z^0_\mu\left(H^2+(\phi^0)^2+2(2s_w^2-1)^2\phi^+\phi^-\right) - \\
\tfrac{1}{2}g^2\tfrac{s_w^2}{c_w}Z^0_\mu\phi^0(W^+_\mu\phi^- + W^-_\mu\phi^+) - \tfrac{1}{2}ig^2\tfrac{s_w^2}{c_w}Z^0_\mu H(W^+_\mu\phi^- - W^-_\mu\phi^+) + \tfrac{1}{2}g^2s_wA_\mu\phi^0(W^+_\mu\phi^- + \\
W^-_\mu\phi^+) + \tfrac{1}{2}ig^2s_wA_\mu H(W^+_\mu\phi^- - W^-_\mu\phi^+) - g^2\tfrac{s_w}{c_w}(2c_w^2-1)Z^0_\mu A_\mu\phi^+\phi^- - \\
g^2s_w^2A_\mu A_\mu\phi^+\phi^- + \tfrac{1}{2}ig_s\lambda^a_{ij}(\bar q^\sigma_i\gamma^\mu q^\sigma_j)g^a_\mu - \bar e^\lambda(\gamma\partial + m^\lambda_e)e^\lambda - \bar\nu^\lambda(\gamma\partial + m^\lambda_\nu)\nu^\lambda - \bar u^\lambda_j(\gamma\partial + \\
m^\lambda_u)u^\lambda_j - \bar d^\lambda_j(\gamma\partial + m^\lambda_d)d^\lambda_j + igs_wA_\mu\left(-(\bar e^\lambda\gamma^\mu e^\lambda) + \tfrac{2}{3}(\bar u^\lambda_j\gamma^\mu u^\lambda_j) - \tfrac{1}{3}(\bar d^\lambda_j\gamma^\mu d^\lambda_j)\right) + \\
\tfrac{ig}{4c_w}Z^0_\mu\{(\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^\lambda_j\gamma^\mu(\tfrac{4}{3}s_w^2-1-\gamma^5)d^\lambda_j) + \\
(\bar u^\lambda_j\gamma^\mu(1-\tfrac{8}{3}s_w^2+\gamma^5)u^\lambda_j)\} + \tfrac{ig}{2\sqrt2}W^+_\mu\left((\bar\nu^\lambda\gamma^\mu(1+\gamma^5)U^{lep}{}_{\lambda\kappa}e^\kappa) + (\bar u^\lambda_j\gamma^\mu(1+\gamma^5)C_{\lambda\kappa}d^\kappa_j)\right) + \\
\tfrac{ig}{2\sqrt2}W^-_\mu\left((\bar e^\kappa U^{lep\dagger}{}_{\kappa\lambda}\gamma^\mu(1+\gamma^5)\nu^\lambda) + (\bar d^\kappa_j C^\dagger_{\kappa\lambda}\gamma^\mu(1+\gamma^5)u^\lambda_j)\right) + \\
\tfrac{ig}{2M\sqrt2}\phi^+\left(-m^\kappa_e(\bar\nu^\lambda U^{lep}{}_{\lambda\kappa}(1-\gamma^5)e^\kappa) + m^\lambda_\nu(\bar\nu^\lambda U^{lep}{}_{\lambda\kappa}(1+\gamma^5)e^\kappa\right) + \\
\tfrac{ig}{2M\sqrt2}\phi^-\left(m^\lambda_e(\bar e^\lambda U^{lep\dagger}{}_{\lambda\kappa}(1+\gamma^5)\nu^\kappa) - m^\kappa_\nu(\bar e^\lambda U^{lep\dagger}{}_{\lambda\kappa}(1-\gamma^5)\nu^\kappa)\right) - \tfrac{g}{2}\tfrac{m^\lambda_\nu}{M}H(\bar\nu^\lambda\nu^\lambda) - \\
\tfrac{g}{2}\tfrac{m^\lambda_e}{M}H(\bar e^\lambda e^\lambda) + \tfrac{ig}{2}\tfrac{m^\lambda_\nu}{M}\phi^0(\bar\nu^\lambda\gamma^5\nu^\lambda) - \tfrac{ig}{2}\tfrac{m^\lambda_e}{M}\phi^0(\bar e^\lambda\gamma^5 e^\lambda) - \tfrac{1}{4}\bar\nu_\lambda M^R_{\lambda\kappa}(1-\gamma_5)\hat\nu_\kappa - \\
\overline{\tfrac{1}{4}\bar\nu_\lambda M^R_{\lambda\kappa}(1-\gamma_5)\hat\nu_\kappa} + \tfrac{ig}{2M\sqrt2}\phi^+\left(-m^\kappa_d(\bar u^\lambda_j C_{\lambda\kappa}(1-\gamma^5)d^\kappa_j) + m^\lambda_u(\bar u^\lambda_j C_{\lambda\kappa}(1+\gamma^5)d^\kappa_j\right) + \\
\tfrac{ig}{2M\sqrt2}\phi^-\left(m^\lambda_d(\bar d^\lambda_j C^\dagger_{\lambda\kappa}(1+\gamma^5)u^\kappa_j) - m^\kappa_u(\bar d^\lambda_j C^\dagger_{\lambda\kappa}(1-\gamma^5)u^\kappa_j)\right) - \tfrac{g}{2}\tfrac{m^\lambda_u}{M}H(\bar u^\lambda_j u^\lambda_j) - \\
\tfrac{g}{2}\tfrac{m^\lambda_d}{M}H(\bar d^\lambda_j d^\lambda_j) + \tfrac{ig}{2}\tfrac{m^\lambda_u}{M}\phi^0(\bar u^\lambda_j\gamma^5u^\lambda_j) - \tfrac{ig}{2}\tfrac{m^\lambda_d}{M}\phi^0(\bar d^\lambda_j\gamma^5 d^\lambda_j) + \bar G^a\partial^2G^a + g_sf^{abc}\partial_\mu\bar G^aG^bg^c_\mu + \\
\bar X^+(\partial^2-M^2)X^+ + \bar X^-(\partial^2-M^2)X^- + \bar X^0(\partial^2-\tfrac{M^2}{c_w^2})X^0 + \bar Y\partial^2Y + igc_wW^+_\mu(\partial_\mu\bar X^0X^- - \\
\partial_\mu\bar X^+X^0) + igs_wW^+_\mu(\partial_\mu\bar YX^- - \partial_\mu\bar X^+Y) + igc_wW^-_\mu(\partial_\mu\bar X^-X^0 - \\
\partial_\mu\bar X^0X^+) + igs_wW^-_\mu(\partial_\mu\bar X^-Y - \partial_\mu\bar YX^+) + igc_wZ^0_\mu(\partial_\mu\bar X^+X^+ - \\
\partial_\mu\bar X^-X^-) + igs_wA_\mu(\partial_\mu\bar X^+X^+ - \\
\partial_\mu\bar X^-X^-) - \tfrac12 gM\left(\bar X^+X^+H + \bar X^-X^-H + \tfrac{1}{c_w^2}\bar X^0X^0H\right) + \tfrac{1-2c_w^2}{2c_w}igM\left(\bar X^+X^0\phi^+ - \bar X^-X^0\phi^-\right) + \\
\tfrac{1}{2c_w}igM\left(\bar X^0X^-\phi^+ - \bar X^0X^+\phi^-\right) + igMs_w\left(\bar X^0X^-\phi^+ - \bar X^0X^+\phi^-\right) + \\
\tfrac12 igM\left(\bar X^+X^+\phi^0 - \bar X^-X^-\phi^0\right).
\end{gathered}
$$



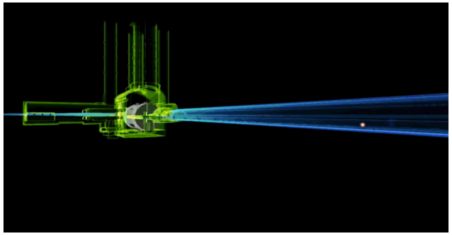

EA-01-2474

Particle physics

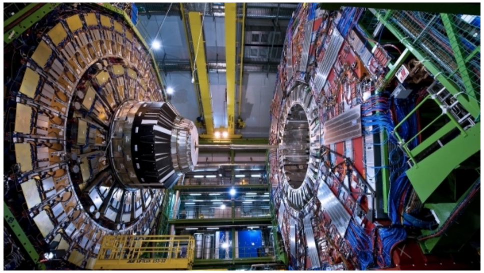


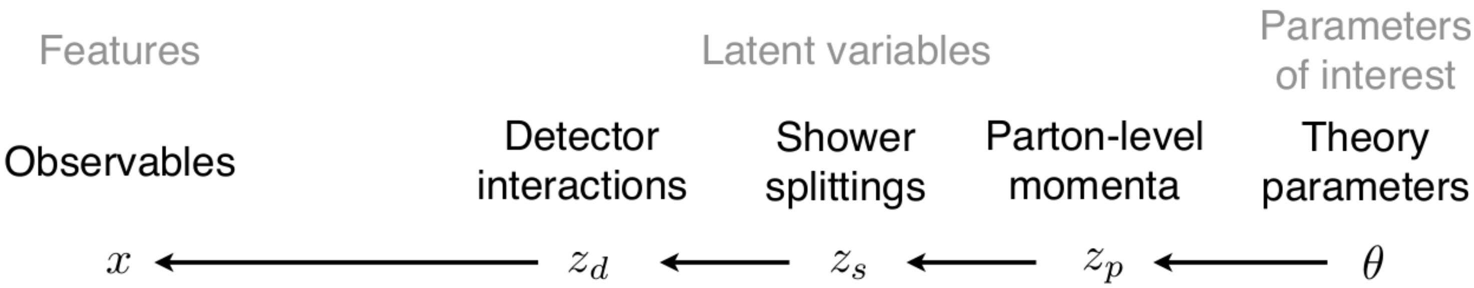
SM with parameters θ

Simulated observables x



Real observations x_{obs}





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

Ingredients

Statistical inference requires the computation of key ingredients, such as

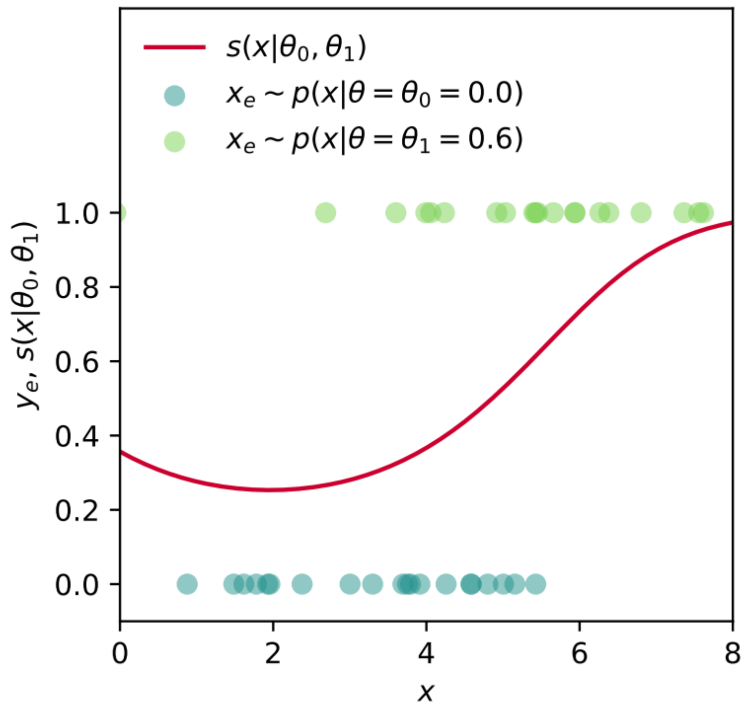
- the likelihood $p(x|\theta)$,
- the likelihood ratio $r(x|\theta_0, \theta_1) = \frac{p(x|\theta_0)}{p(x|\theta_1)}$,
- or the posterior $p(\theta|x)$.

In the simulator-based scenario, each of these ingredients can be approximated with modern machine learning techniques, even if none are tractable during training!

CARL

Supervised learning provides a way to automatically learn $p(x|\theta_0)/p(x|\theta_1)$:

- Let us consider a neural network classifier $\hat{s}$ tasked to distinguish $x_i \sim p(x|\theta_0)$ labelled $y_i = 0$ from $x_i \sim p(x|\theta_1)$ labelled $y_i = 1$.
- Train $\hat{s}$ by minimizing the cross-entropy loss.



Cranmer, Pavez and Louppe, 2015 [arXiv:1506.02169].

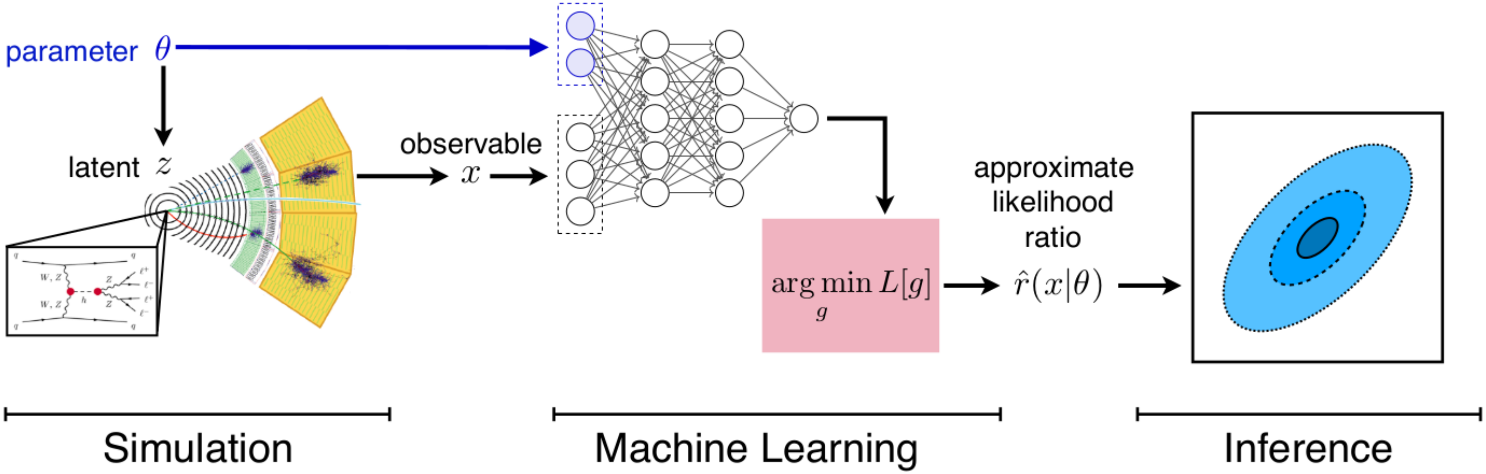
The solution $\hat{s}$ found after training approximates the optimal classifier

$$\hat{s}(x) \approx s^*(x) = \frac{p(x|\theta_1)}{p(x|\theta_0) + p(x|\theta_1)}.$$

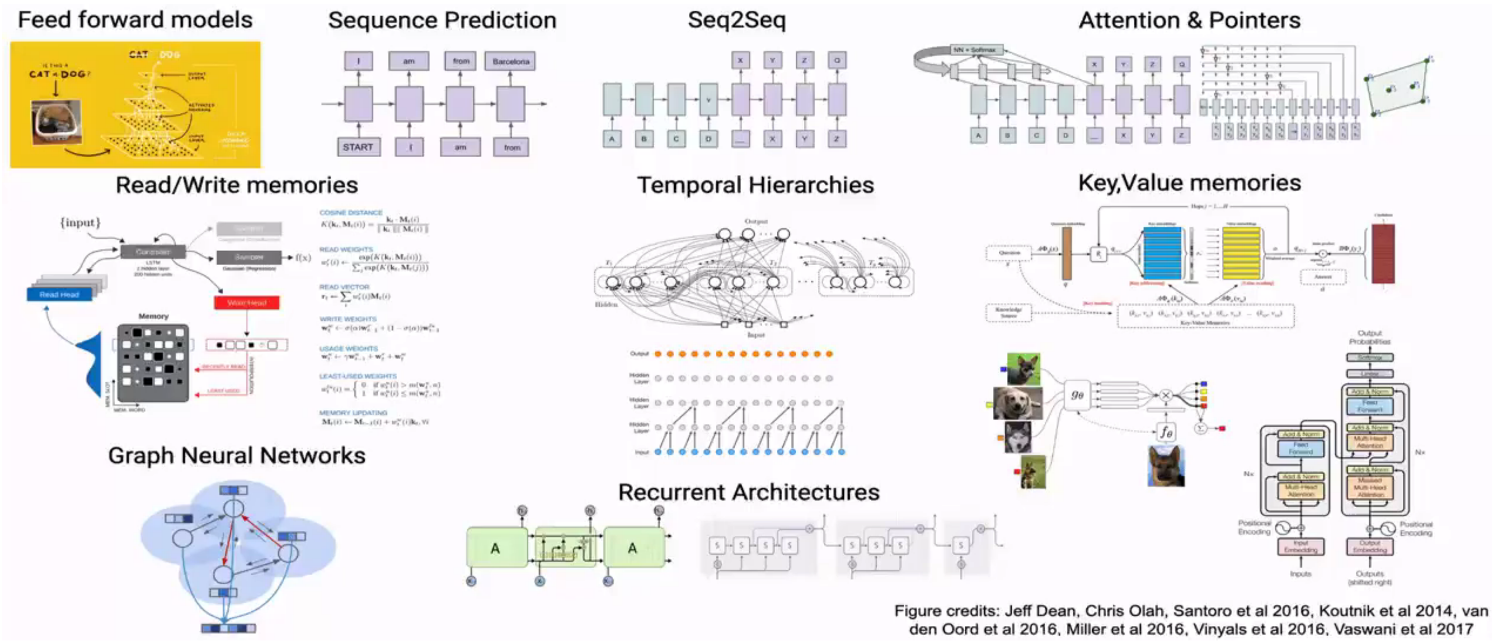
Therefore,

$$r(x|\theta_0, \theta_1) \approx \hat{r}(x|\theta_0, \theta_1) = \frac{1 - \hat{s}(x)}{\hat{s}(x)}.$$

Cranmer, Pavez and Louppe, 2015 [arXiv:1506.02169].



Cranmer, Pavez and Louppe, 2015 [arXiv:1506.02169].



Supervised classification is equivalent to likelihood ratio estimation, therefore the whole Deep Learning toolbox can be used for inference!

Credits: Oriol Vinyals, 2020.

There is more...

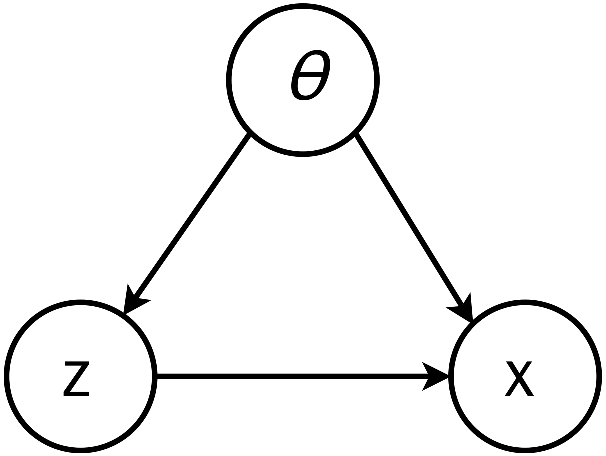
Method	Simulate	Extract $r(x,z)$	Extract $t(x,z)$	NN estimates	Asympt. exact	Generative
ROLR	$\theta_0 \sim \pi(\theta), \theta_1$	✓		$\hat{r}(x\|\theta_0, \theta_1)$	✓	
CASCAL	$\theta_0 \sim \pi(\theta), \theta_1$		✓	$\hat{r}(x\|\theta_0, \theta_1)$	✓	
ALICE	$\theta_0 \sim \pi(\theta), \theta_1$		✓	$\hat{r}(x\|\theta_0, \theta_1)$	✓	
RASCAL	$\theta_0 \sim \pi(\theta), \theta_1$	✓	✓	$\hat{r}(x\|\theta_0, \theta_1)$	✓	
ALICES	$\theta_0 \sim \pi(\theta), \theta_1$	✓	✓	$\hat{r}(x\|\theta_0, \theta_1)$	✓	
SCANDAL	$\theta \sim \pi(\theta)$		✓	$\hat{p}(x\|\theta)$	✓	✓
SALLY	θ_{ref}		✓	$\hat{t}(x\|\theta_{\text{ref}})$	in local approx.	
SALLINO	θ_{ref}		✓	$\hat{t}(x\|\theta_{\text{ref}})$	in local approx.	

Brehmer, Louppe, Pavez and Cranmer, 2018 [arXiv:1805.12244].

Bayesian inference

Bayesian inference = computing the posterior

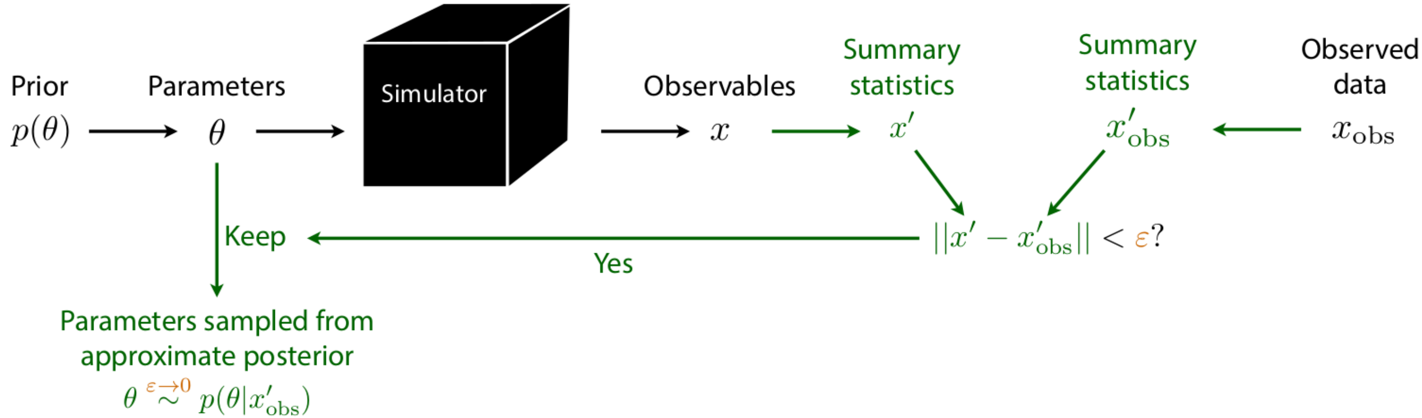
$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}.$$



Doubly intractable in the likelihood-free scenario:

- Cannot evaluate the likelihood $p(x|\theta) = \int p(x, z|\theta)dz$.
- Cannot evaluate the evidence $p(x) = \int p(x|\theta)p(\theta)d\theta$.

Approximate Bayesian Computation (ABC)



Issues

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

Credits: Johann Brehmer.

Amortizing Bayes

The Bayes rule can be rewritten as

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

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

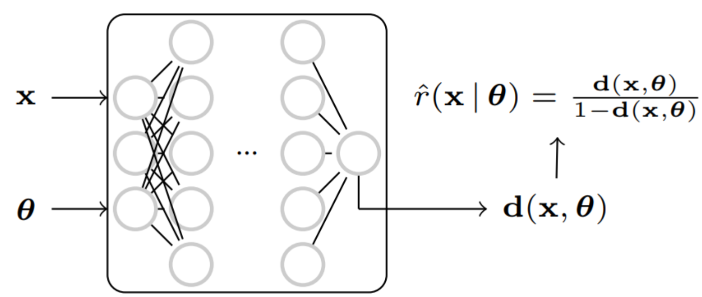
The likelihood-to-evidence ratio can be learned with a neural network tasked to distinguish $x \sim p(x|\theta)$ from $x \sim p(x)$.

This enables direct and amortized posterior evaluation.

Algorithm 1 Optimization of $\mathbf{d}(\mathbf{x}, \boldsymbol{\theta})$.

Inputs: Criterion ℓ (e.g., BCE)
Implicit generative model $p(\mathbf{x} \mid \boldsymbol{\theta})$
Prior $p(\boldsymbol{\theta})$
Outputs: Parameterized classifier $\mathbf{d}_\phi(\mathbf{x}, \boldsymbol{\theta})$
Hyperparameters: Batch-size M

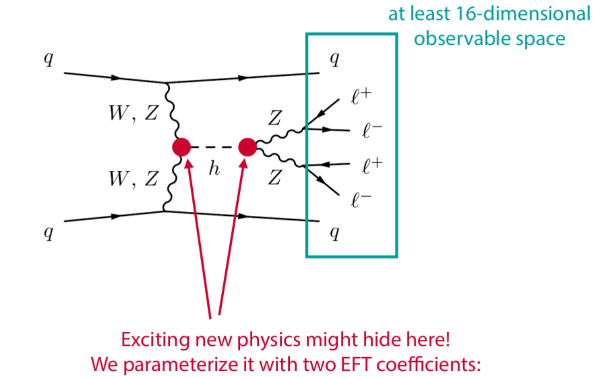
while not converged do
  Sample $\boldsymbol{\theta} \leftarrow \{\boldsymbol{\theta}_m \sim p(\boldsymbol{\theta})\}_{m=1}^M$
  Sample $\boldsymbol{\theta}' \leftarrow \{\boldsymbol{\theta}'_m \sim p(\boldsymbol{\theta})\}_{m=1}^M$
  Simulate $\mathbf{x} \leftarrow \{\mathbf{x}_m \sim p(\mathbf{x} \mid \boldsymbol{\theta}_m)\}_{m=1}^M$
  $\mathcal{L} \leftarrow \ell(\mathbf{d}_\phi(\mathbf{x}, \boldsymbol{\theta}),\ 1) + \ell(\mathbf{d}_\phi(\mathbf{x}, \boldsymbol{\theta}'),\ 0)$
  $\phi \leftarrow \text{OPTIMIZER}(\phi,\ \nabla_\phi \mathcal{L})$
end while
return $\mathbf{d}_\phi$



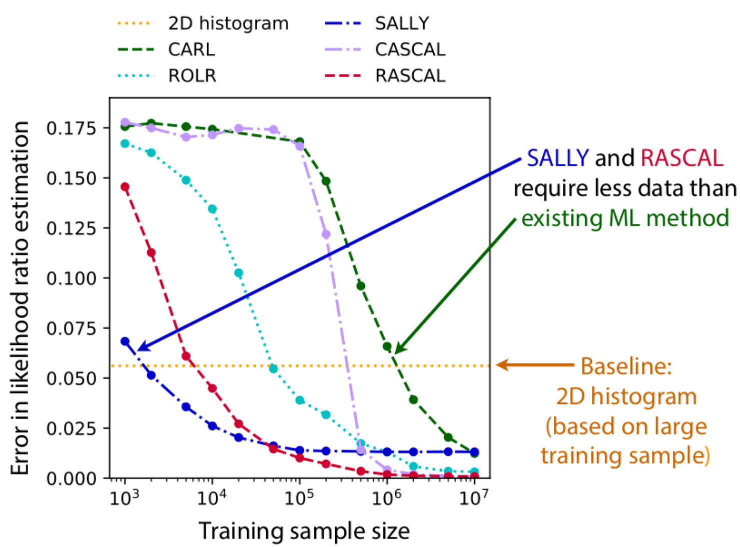
Showtime

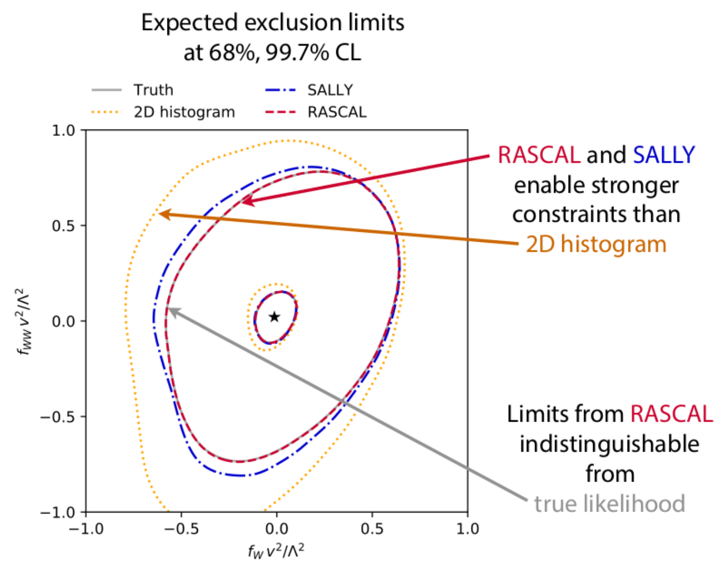
① Hunting new physics at particle colliders

The goal is to constrain two EFT parameters and compare against traditional histogram analysis.



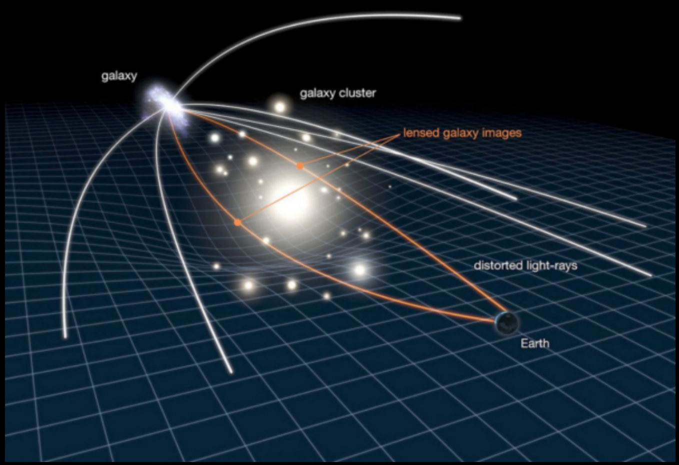
$$\mathcal{L} = \mathcal{L}_{\mathrm{SM}} + \underbrace{\frac{f_W}{\Lambda^2}}\,\underbrace{\frac{ig}{2}\,(D^\mu\phi)^\dagger\,\sigma^a\,D^\nu\phi\,W^a_{\mu\nu}}_{\mathcal{O}_W} - \frac{f_{WW}}{\Lambda^2}\,\underbrace{\frac{g^2}{4}\,(\phi^\dagger\phi)\,W^a_{\mu\nu}\,W^{\mu\nu\,a}}_{\mathcal{O}_{WW}}$$



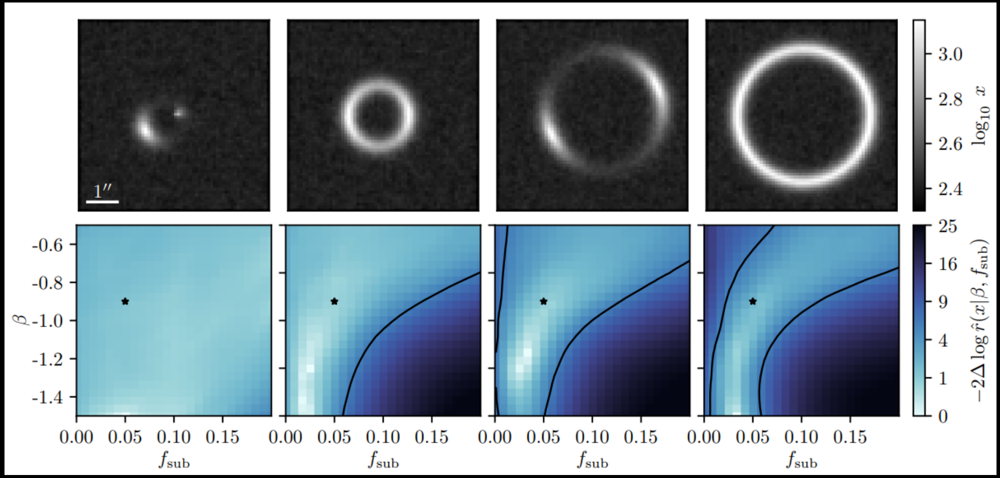


Brehmer, Cranmer, Louppe, and Pavez, 2018a [arXiv:1805.00020], 2018b [arXiv:1805.00013]; Brehmer, Louppe, Pavez and Cranmer, 2018 [arXiv:1805.12244].

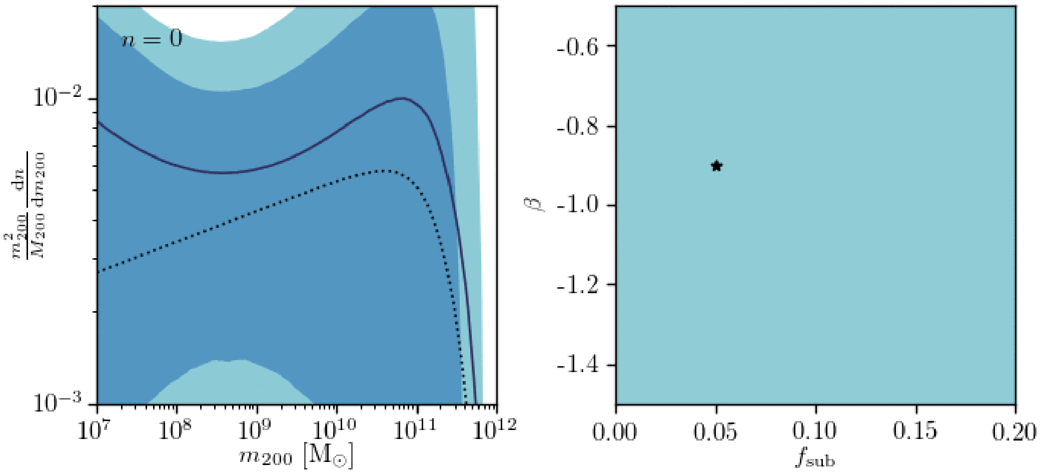
② Dark matter substructure from gravitational lensing



The number of dark matter subhalos and their mass and location lead to complex latent space of each image. The goal is the **inference of population parameters** β **and** f_{sub}.

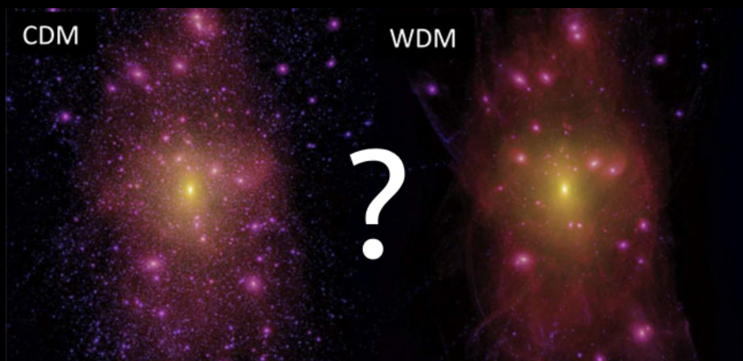


Brehmer, Mishra-Sharma, Hermans, Louppe, and Cranmer, 2019 [arXiv:1909.02005].



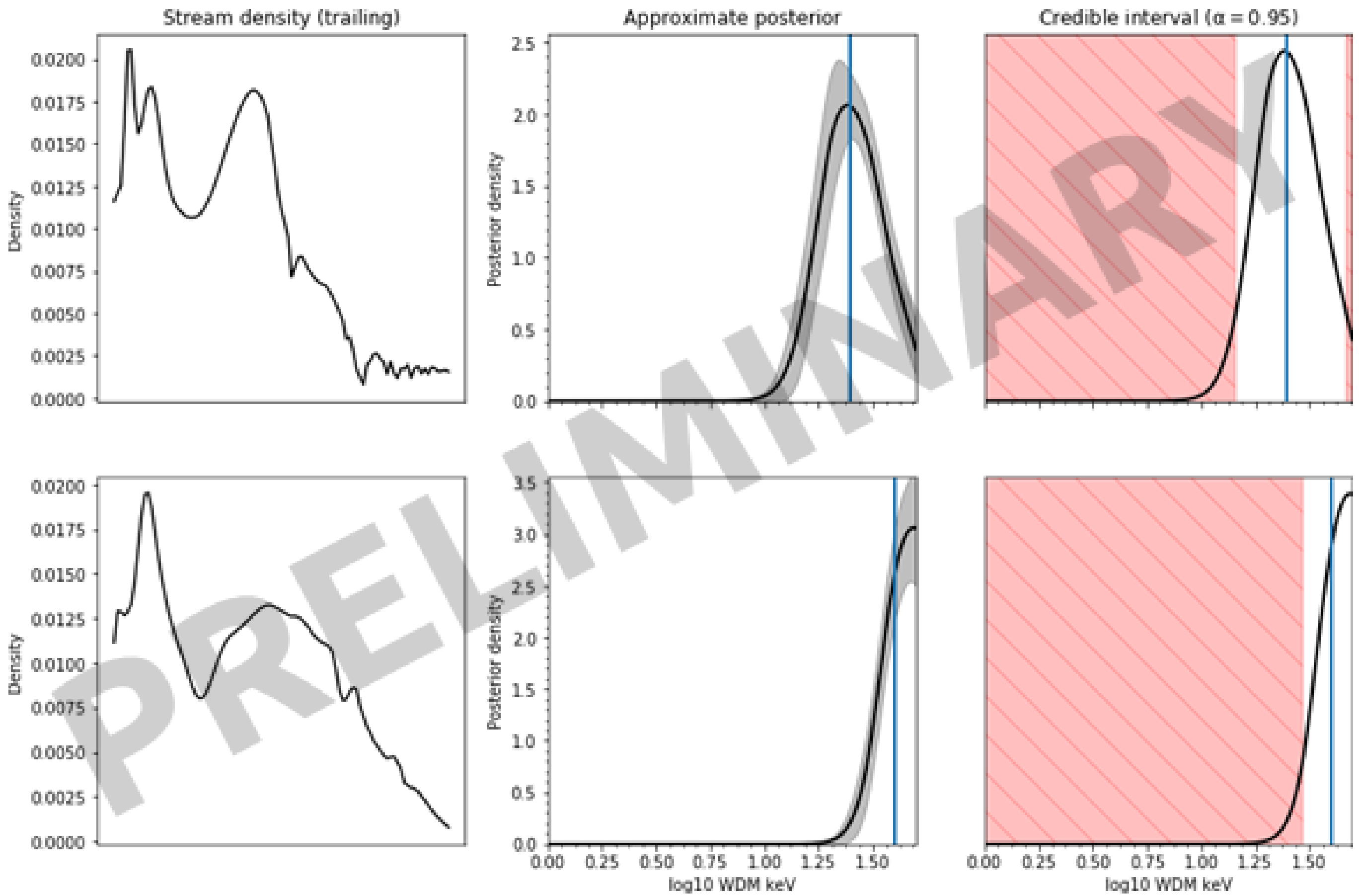
Brehmer, Mishra-Sharma, Hermans, Louppe, and Cranmer, 2019 [arXiv:1909.02005].

③ Constraining the WDM particle mass



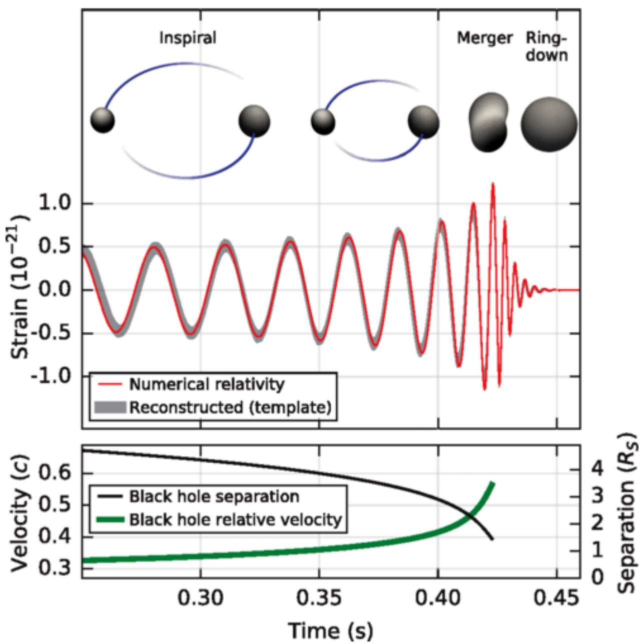
Dark matter subhalos cause disturbances in the density of stellar streams.

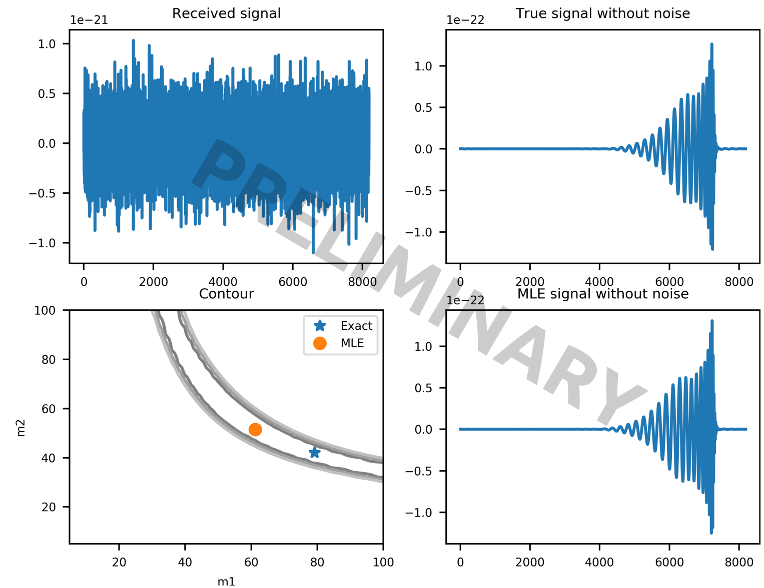
Therefore, observations of stellar streams may be used to **constrain the mass of the dark matter particle**.



Stream density (trailing)
Approximate posterior
Credible interval (α = 0.95)
Density
Posterior density
log10 WDM keV
PRELIMINARY

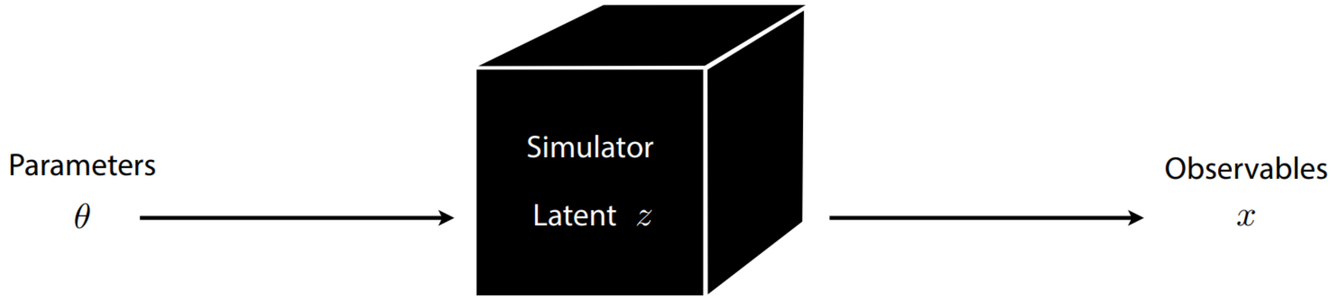
④ Fast parameter estimation for gravitational waves





Summary

- Much of modern science is based on "likelihood-free" simulations.
- The likelihood-ratio is central to many statistical inference procedures, regardless of your religion.
- Supervised learning enables likelihood-ratio estimation.
- Better likelihood-ratio estimates can be achieved by mining simulators.



Collaborators

Kyle Cranmer

Juan Pavez

Johann Brehmer

Joeri Hermans

Antoine Wehenkel

Arnaud Delaunoy

Siddarth Mishra-Sharma

References

- Cranmer, K., Brehmer, J., & Louppe, G. (2019). The frontier of simulation-based inference. arXiv preprint arXiv:1911.01429.
- Brehmer, J., Mishra-Sharma, S., Hermans, J., Louppe, G., Cranmer, K. (2019). Mining for Dark Matter Substructure: Inferring subhalo population properties from strong lenses with machine learning. arXiv preprint arXiv 1909.02005.
- Hermans, J., Begy, V., & Louppe, G. (2019). Likelihood-free MCMC with Approximate Likelihood Ratios. arXiv preprint arXiv:1903.04057.
- Stoye, M., Brehmer, J., Louppe, G., Pavez, J., & Cranmer, K. (2018). Likelihood-free inference with an improved cross-entropy estimator. arXiv preprint arXiv:1808.00973.
- Brehmer, J., Louppe, G., Pavez, J., & Cranmer, K. (2018). Mining gold from implicit models to improve likelihood-free inference. arXiv preprint arXiv:1805.12244.
- Brehmer, J., Cranmer, K., Louppe, G., & Pavez, J. (2018). Constraining Effective Field Theories with Machine Learning. arXiv preprint arXiv:1805.00013.
- Brehmer, J., Cranmer, K., Louppe, G., & Pavez, J. (2018). A Guide to Constraining Effective Field Theories with Machine Learning. arXiv preprint arXiv:1805.00020.
- Cranmer, K., Pavez, J., & Louppe, G. (2015). Approximating likelihood ratios with calibrated discriminative classifiers. arXiv preprint arXiv:1506.02169.

The end.