Reconstructing ALP properties with simulation-based inference

Alessandro Morandini
with Felix Kahlhoefer and Torben Ferber
based on [2308.01353]

Karlsruher Institut für Technologie

ML4Jets, Hamburg, 06.11.2023

Inverse problem in high-energy physics

Simulation: ✓

BSM parameters $\theta \longrightarrow$ detector measurements x

Inference: ✗

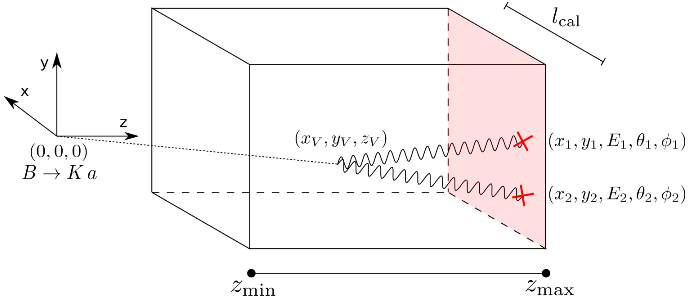
(real) detector measurements $x \longrightarrow$ BSM parameters θ

Why is it complicated? **No direct access to likelihood**

For instance at LHC:

$$p(x|\theta) = \int \mathrm{d}z_d \int \mathrm{d}z_s \int \mathrm{d}z_p \, p(x|z_d) p(z_d|z_s) p(z_s|z_p) p(z_p|\theta)$$

Beam dump sketch and input variables



Geometry of the detector: $z_{\text{min}} = 10\text{m}, z_{\text{max}} = 35\text{m}, l_{\text{cal}} = 2.5\text{m}$

Detector readout: hits x_i, y_i, energies E_i, photon direction θ_i, ϕ_i

Problem: inaccurate measurements

We see some signal events and group their features in $\mathbf{x}$,
what can we say about the ALP?

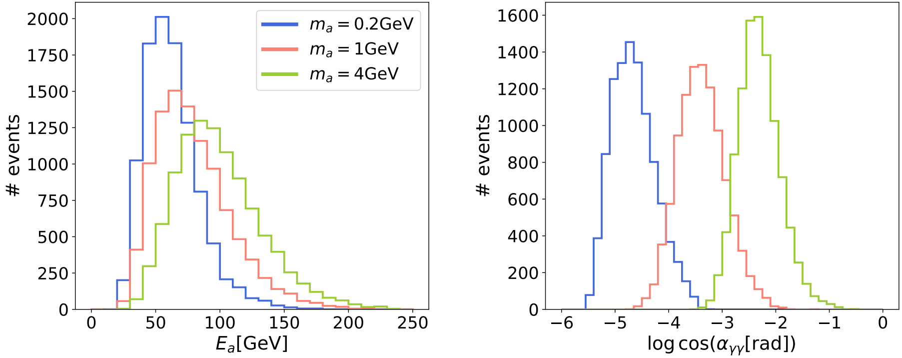
Perfect detector:

- $\mathbf{x} \longrightarrow m_{\gamma\gamma} \longrightarrow m_a$
- $\mathbf{x} \longrightarrow |\mathbf{V}|, \quad |\mathbf{V}| \sim \exp\left(-|\mathbf{V}|\frac{m_a}{|\mathbf{p}_a| c\tau_a}\right)$

Imperfect detector:

- if error is small, should work with $m_{\gamma\gamma}$
- for larger error, signal purity requirements and ad-hoc high-level observables can help
- other?

Variable distributions



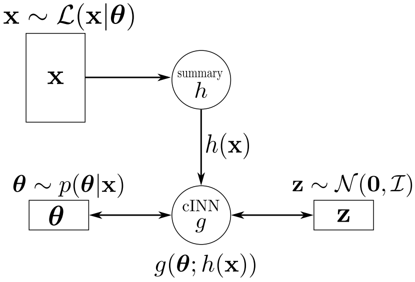
- Mass information also in energy distribution
- Angles useful to infer the mass (but hardest feature to measure)

Simulation-based inference

No direct access to likelihood?
Then use the simulations → simulation-based inference

One possible approach†: use ML to approximate likelihood/posterior

In our case: posterior learnt with conditional invertible neural network



†:

- Posterior/Likelihood
- Classifier/NF
- Summary or not

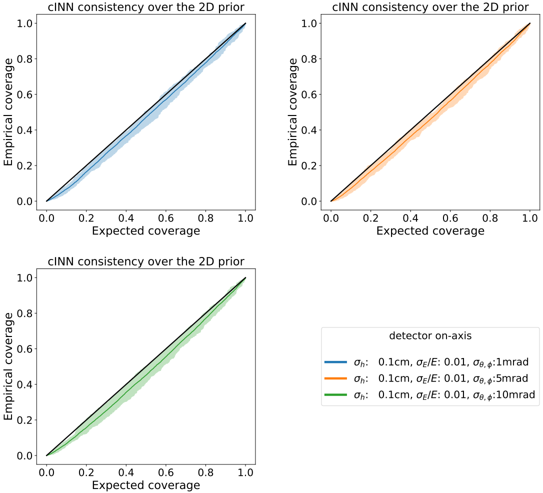
Technical details

Posterior → choose a prior:
$m_a \in [0.1\text{GeV}, 4.5\text{GeV}]$,
$c\tau_a \in [0.1\text{m}, 100\text{m}]$

Input parameters:
$m_a, c\tau_a / m_a$

Number of seen events: 3

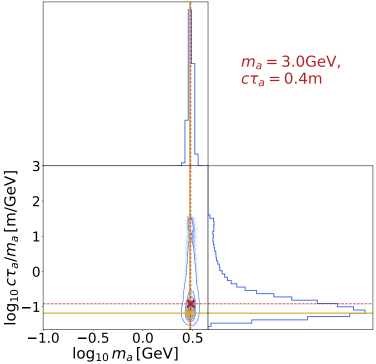
Do we trust the posterior?
Check pp-plot

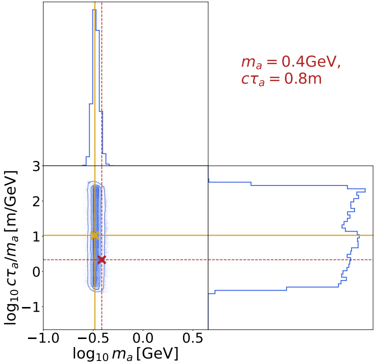


Posterior: role of mass and lifetime

Can we constrain both the mass and the lifetime?

Depends on detector geometry and lifetime value

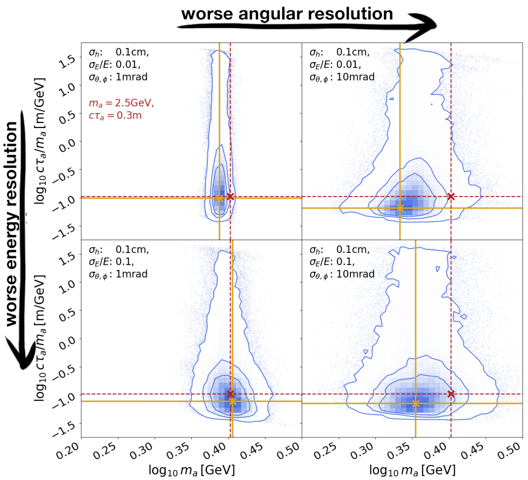




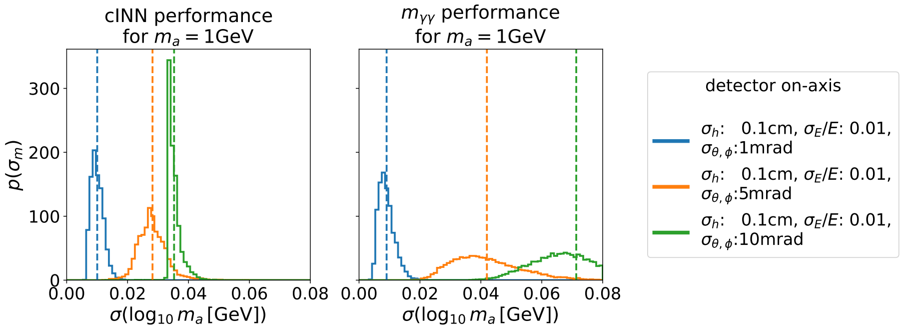
Posterior: role of detector resolution

What if we change the detector?
Then the uncertainties change

Feature resolution	Values scanned
$\sigma(E)/E$	$[0.01, 0.05, 0.1]$
$\sigma(h)$	$[0.1\text{cm}]$
$\sigma(\theta), \sigma(\phi)$	$[1\,\text{mrad}, 5\,\text{mrad}, 10\,\text{mrad}]$

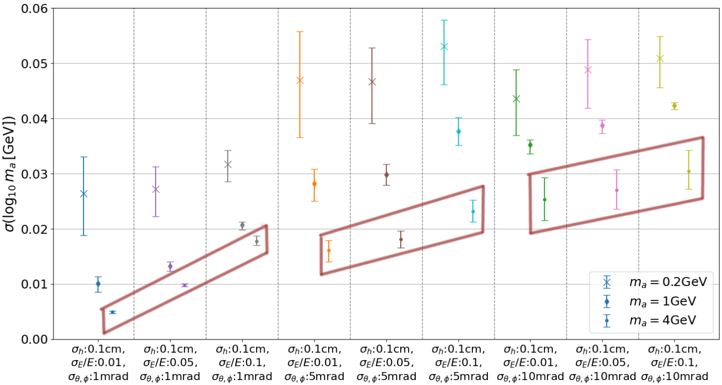


Performance for 1 GeV (role of angle resolution)



- Agreement with $m_{\gamma\gamma}$ for great resolution
- Considerable outperformance for worse angular resolutions

Performance of different detector setups (example)



- Major role played by angular resolution
- Interplay between energy and angular resolution

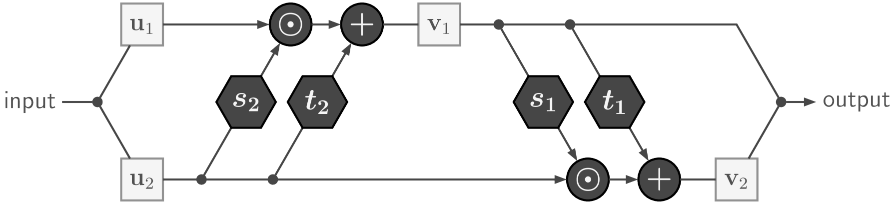
Conclusions

- We can derive fast and reliable parameter estimates **including their uncertainty**
- The considered algorithm **easily adapts** to different setups
- Performance comparisons allow **experimental design**

Thank you!

BACKUP

INN structure



$$v_1 = u_1 \odot \exp(s_1(u_2)) + t_1(u_2) \tag{20}$$
$$v_2 = u_2 \odot \exp(s_2(v_1)) + t_2(v_1) \tag{21}$$

$$u_2 = (v_2 - t_2(v_1)) \odot \exp(-s_2(v_1)) \tag{22}$$
$$u_1 = (v_1 - t_1(u_2)) \odot \exp(-s_1(u_2)) \tag{23}$$

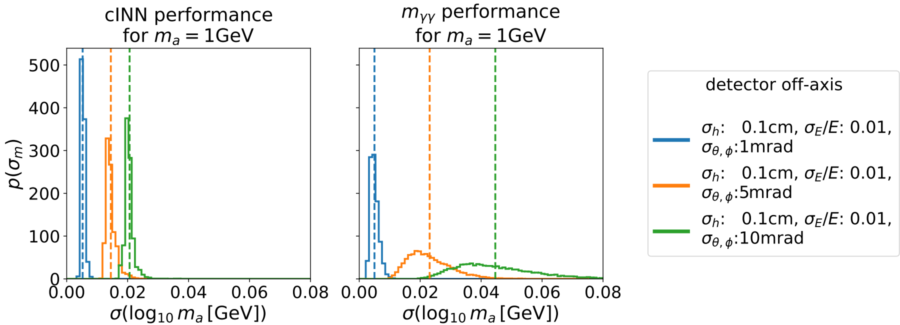
(formulas from BayesFlow and plot from Heidelberg)

INN structure

Coupling layers	
Number coupling layers	4
Hidden layers	[128, 128, 128, 128]
Hidden layers activation	ReLU
Output layer activation	linear
Summary network	
Output layer dimension	2
Hidden layers	[64, 64, 64, 64]
Hidden layers activation	LeakyReLU ($\alpha = 0.01$)
Output layer activation	linear
Training hyperparameters	
Max number epochs	500
Batch size	512
Initial learning rate	$5 \cdot 10^{-3}$
Decay rate	0.9 every 10 epochs
Early stopping	$\delta < 10^{-3}$ for 50 epochs

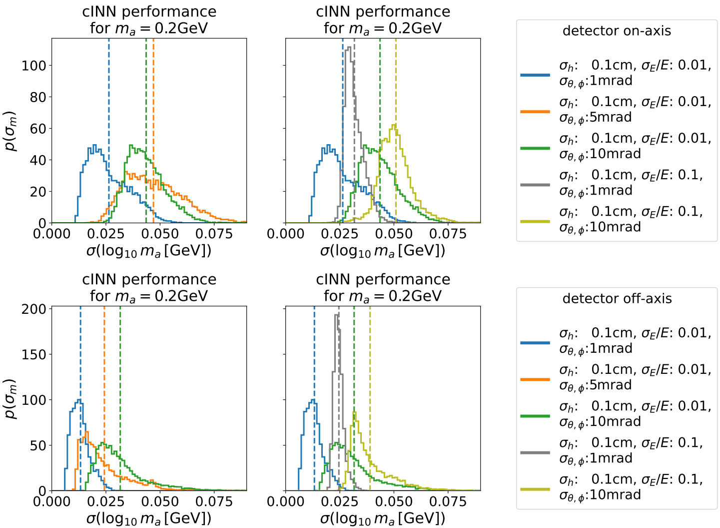
Table: Architecture of the summary network and of the cINN. The output of the summary network is fed into the coupling layers transformations. Since the summary network and the cINN are trained together, the training hyperparameters apply to both of them.

Performance for 1 GeV (displaced case)

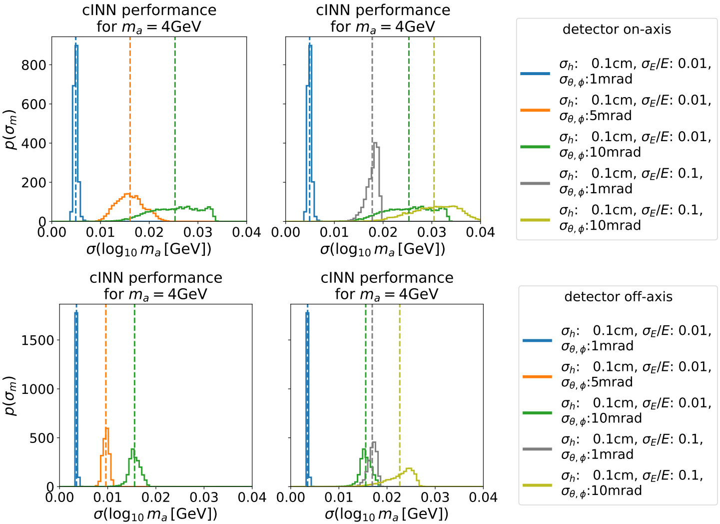


- Same conclusions as in non-displaced case
- Smaller parameter uncertainty (for same number of seen events!)

Performance for 0.2 GeV



Performance for 4 GeV



Performance of different detector setups