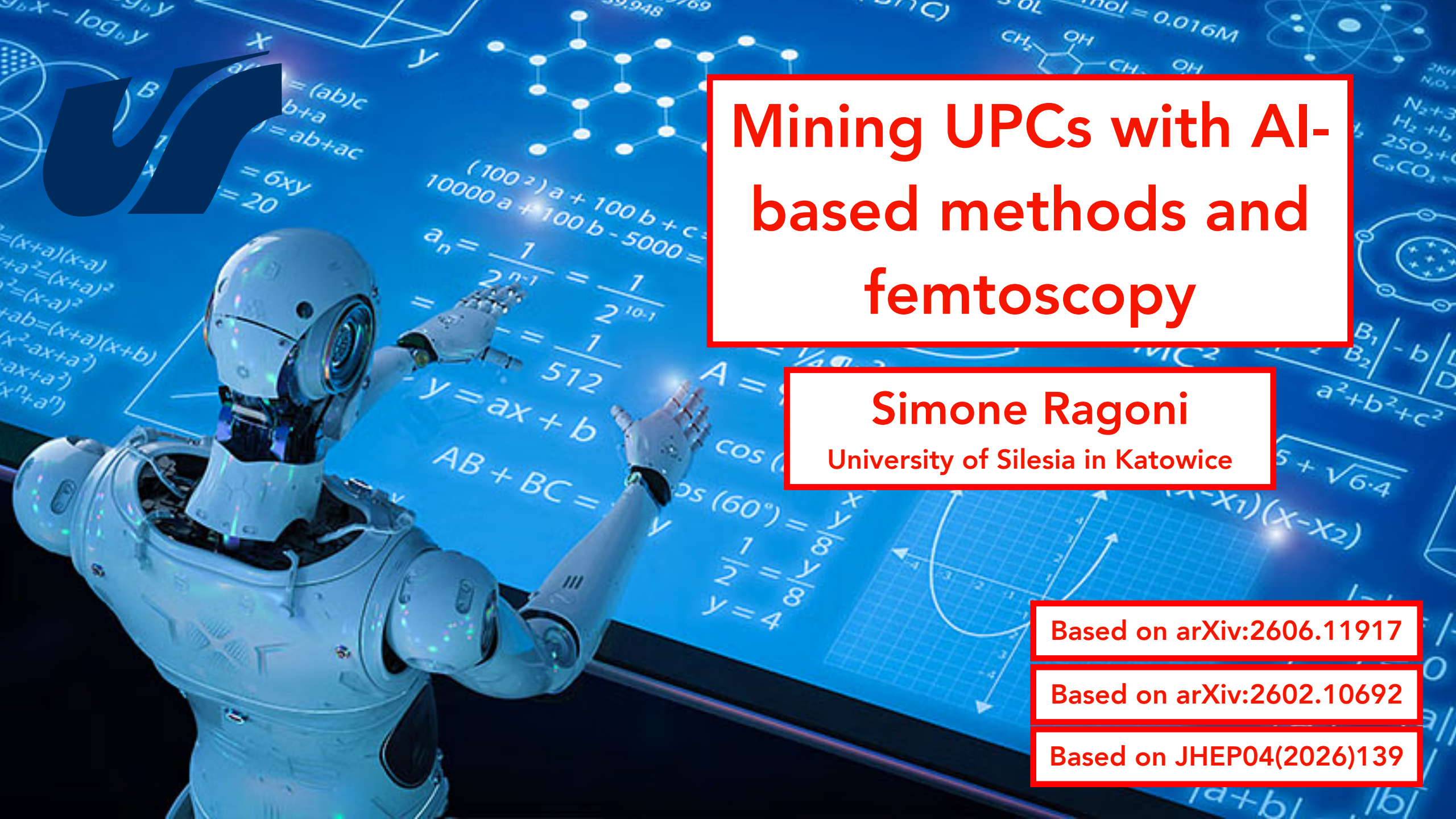




Mining UPCs with AI-based methods and femtoscopy

Simone Ragoni

University of Silesia in Katowice

Based on arXiv:2606.11917

Based on arXiv:2602.10692

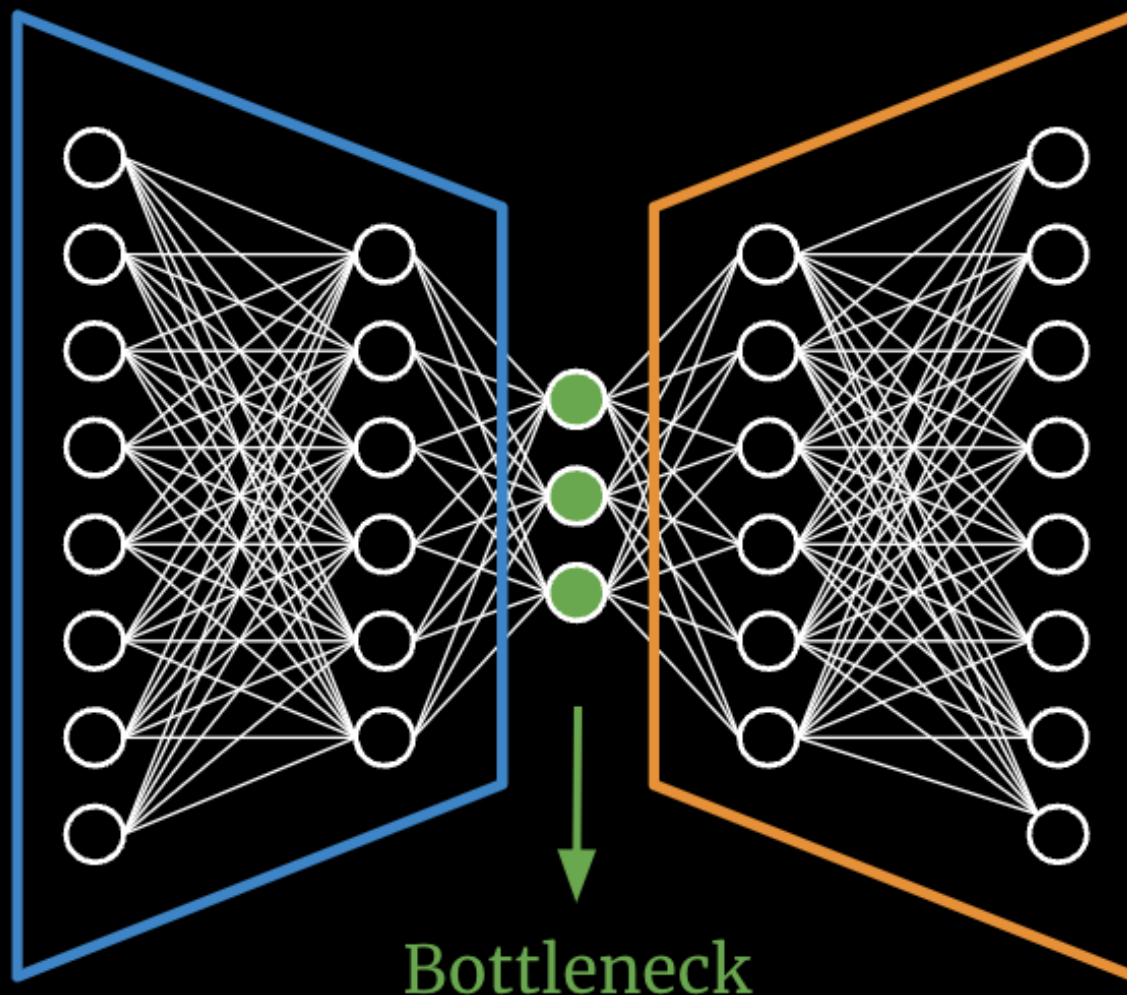
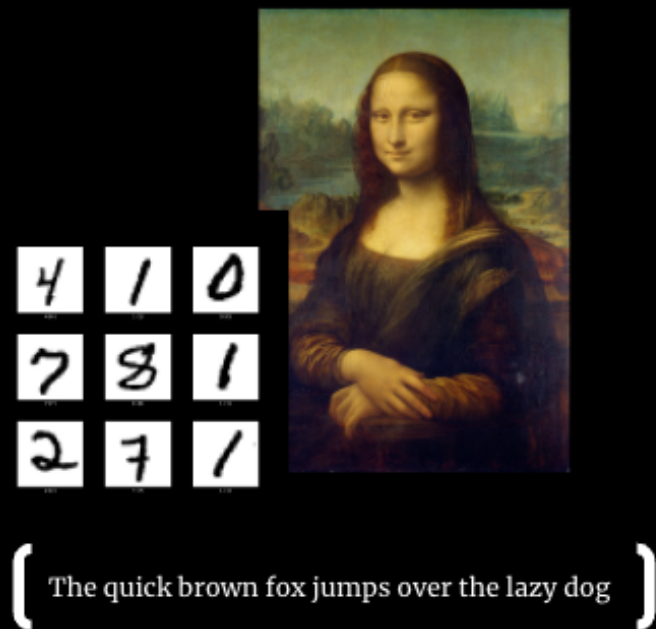
Based on JHEP04(2026)139



Accelerating discovery

Encoder

Decoder

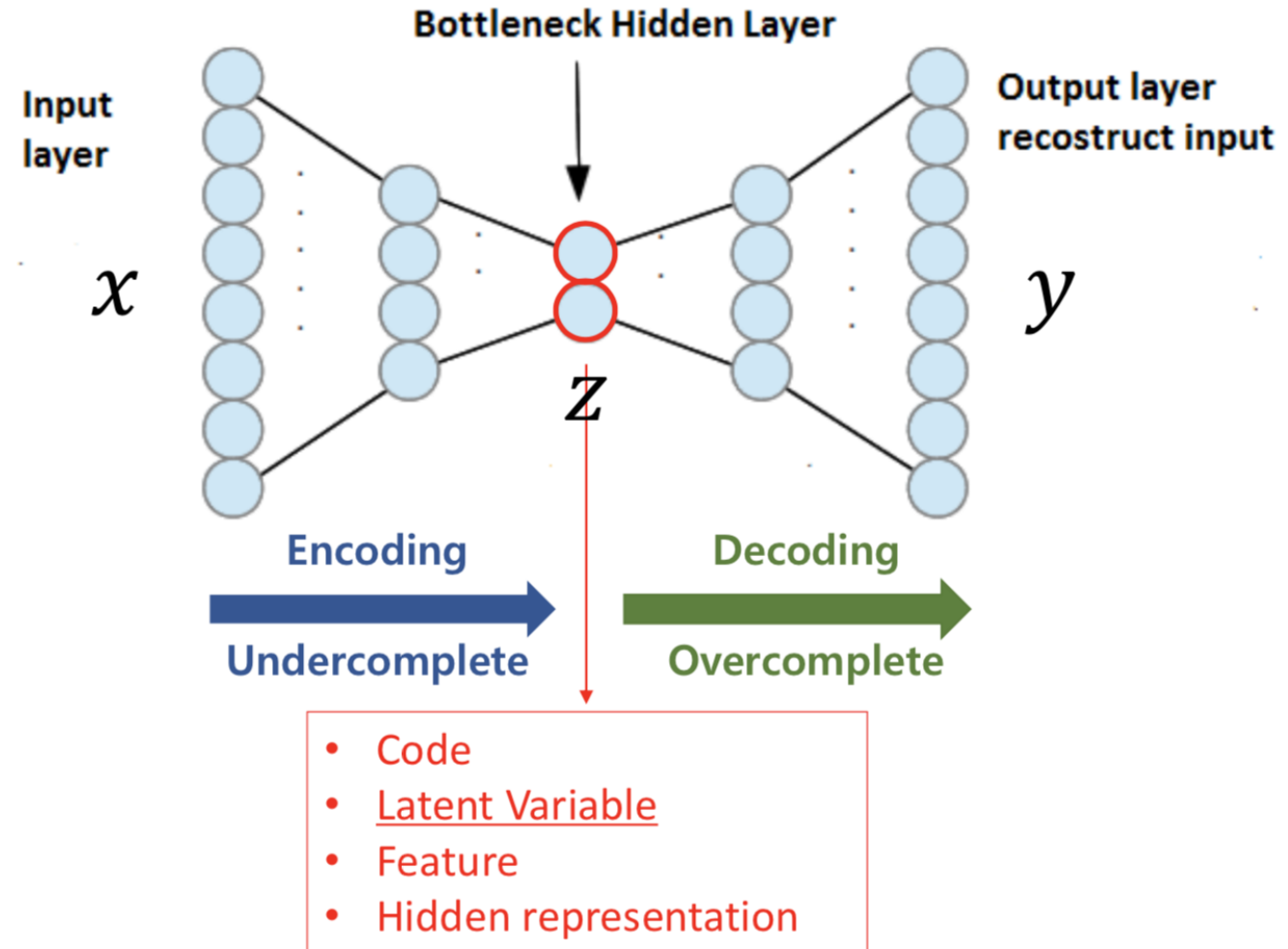


Disclaimer: no ALICE data have been used at all for the purpose of this presentation, but data simulated by us based on realistic conditions

Intro to Autoencoders



- Autoencoders are trained to reduce the dimensionality of the input in an efficient way
- They allow for data compression (achieving up to 60% without overtraining)
- They capture the important features of the data
- By reconstructing the encoded data, it is possible to look for errors in the decoding
- Suitable for anomaly detection

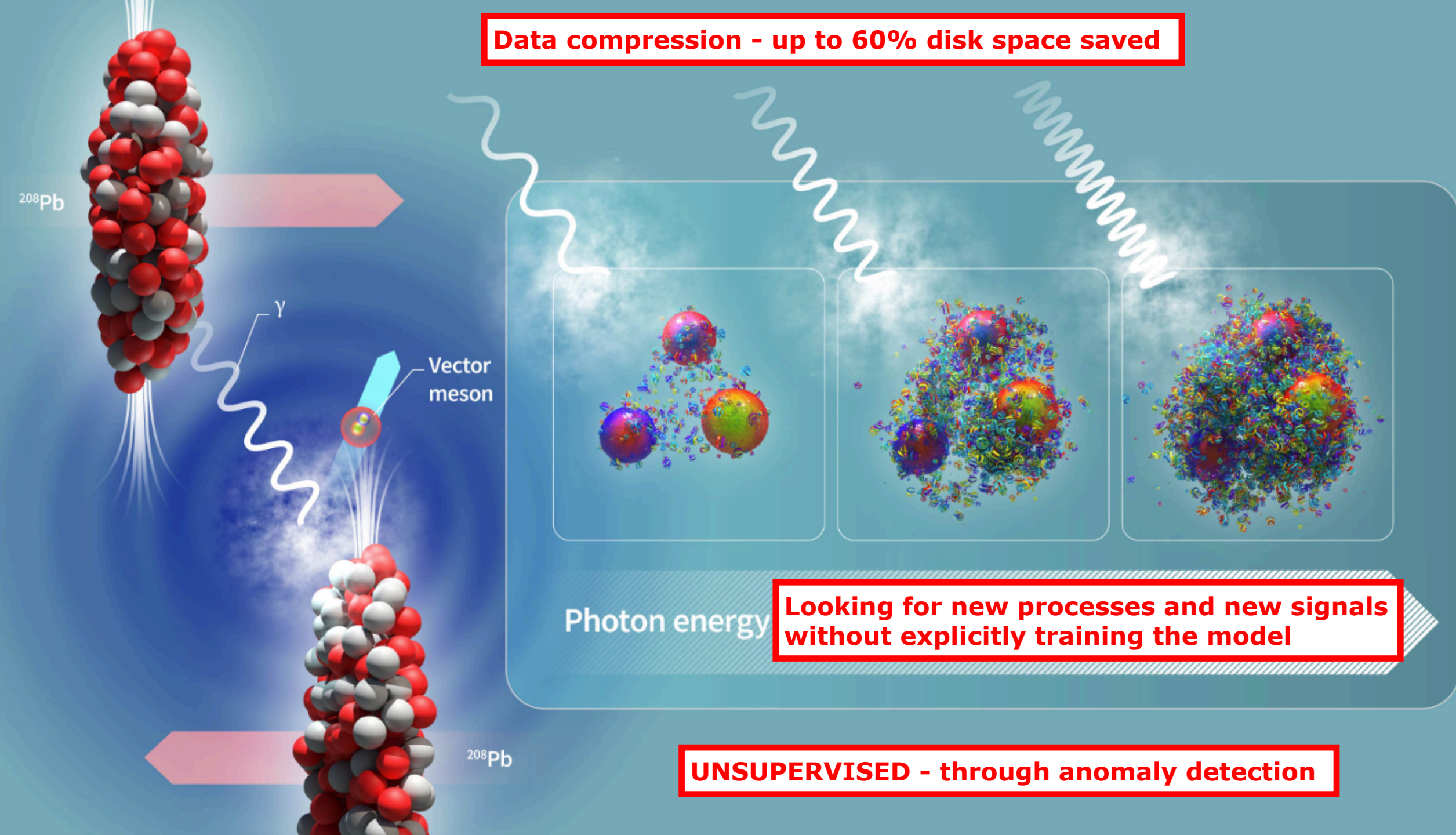


UPC AT LHC AND EIC

ANOMALY DETECTION



Data compression - up to 60% disk space saved



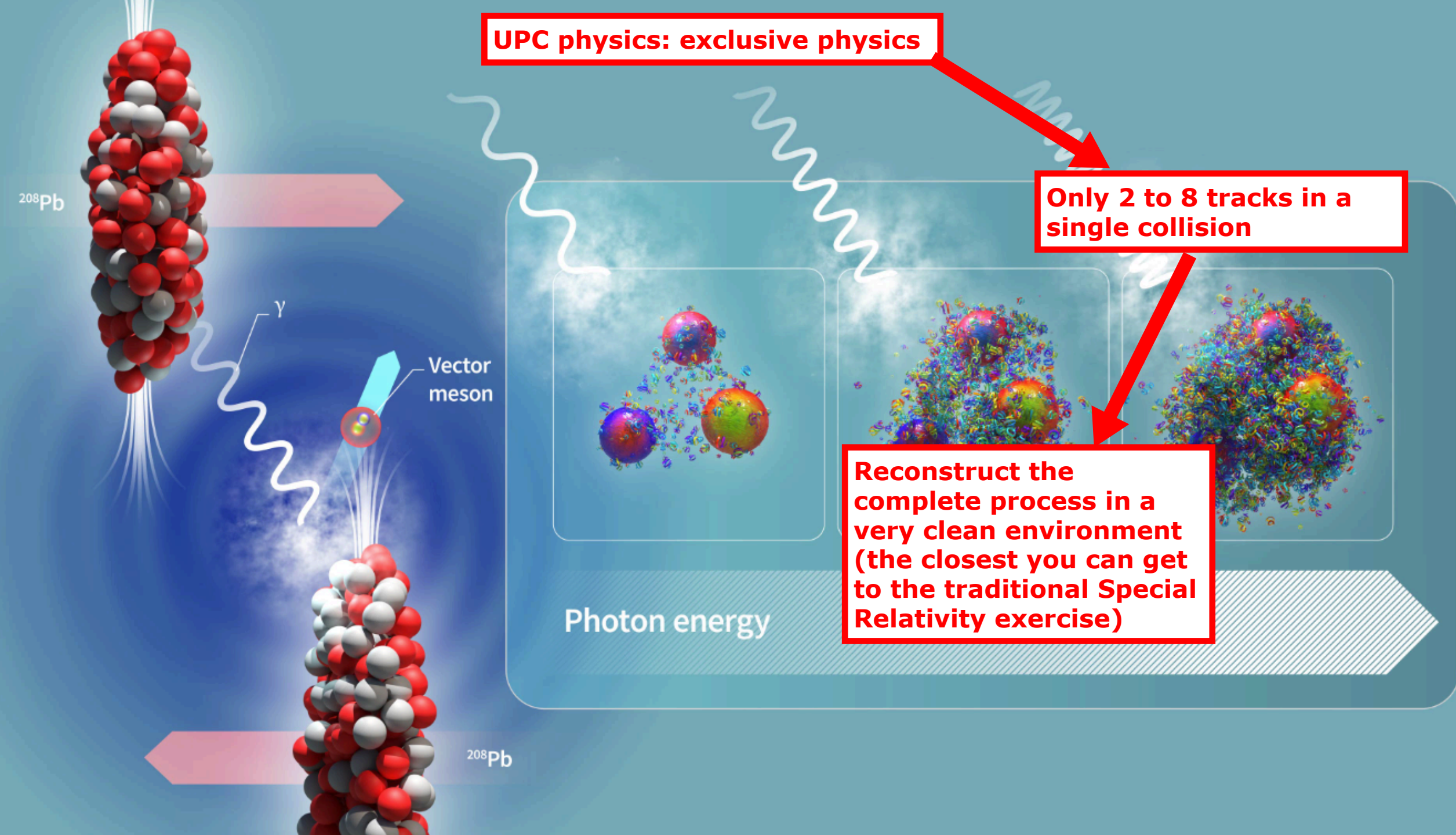
**Looking for new processes and new signals
without explicitly training the model**

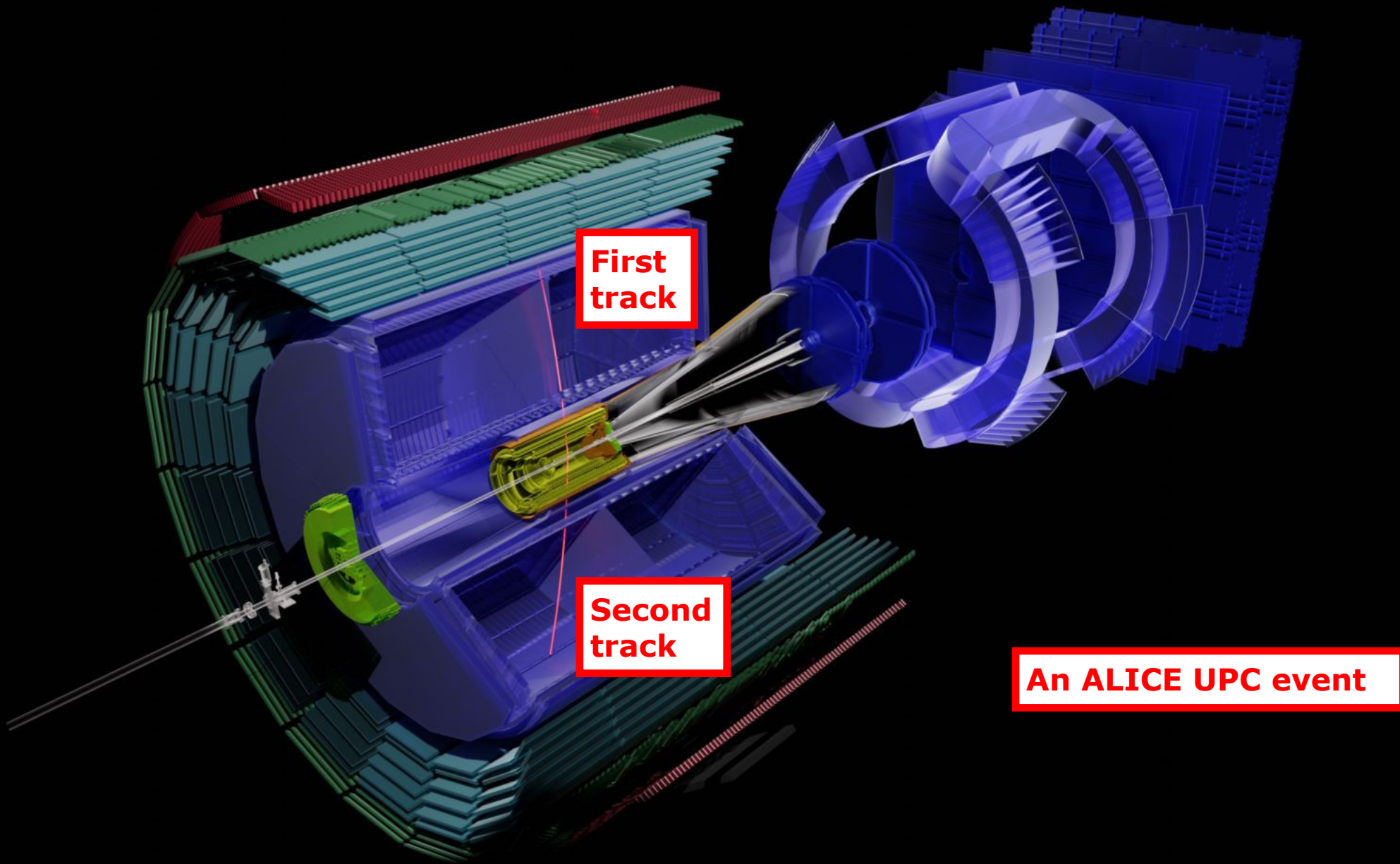
UNSUPERVISED - through anomaly detection

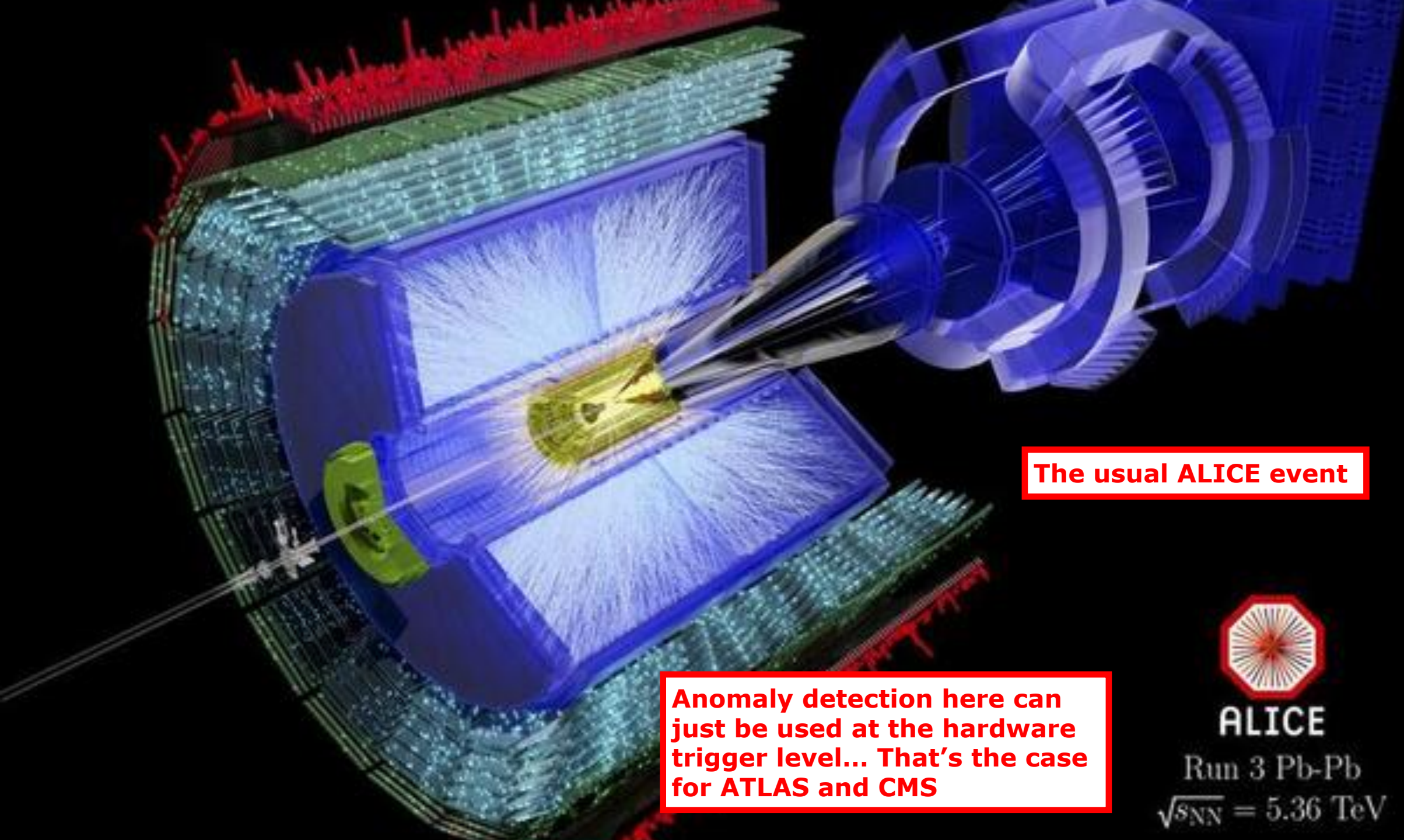
UPC physics: exclusive physics

Only 2 to 8 tracks in a single collision

Reconstruct the complete process in a very clean environment (the closest you can get to the traditional Special Relativity exercise)







The usual ALICE event

Anomaly detection here can just be used at the hardware trigger level... That's the case for ATLAS and CMS



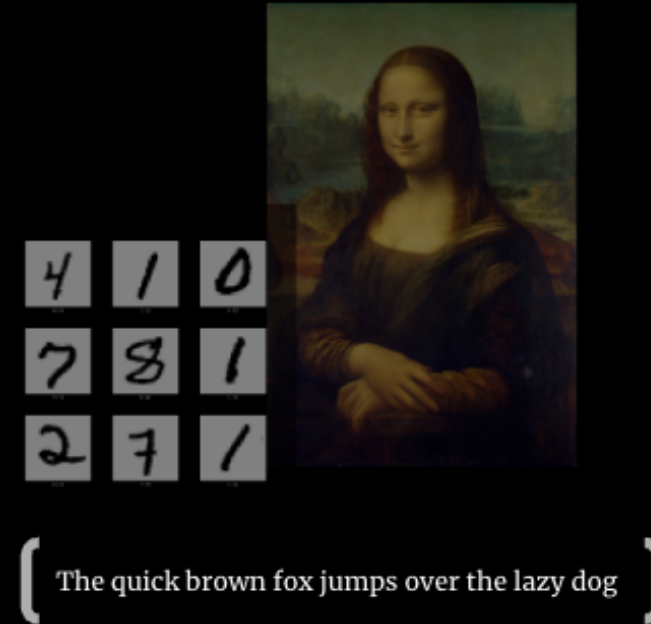
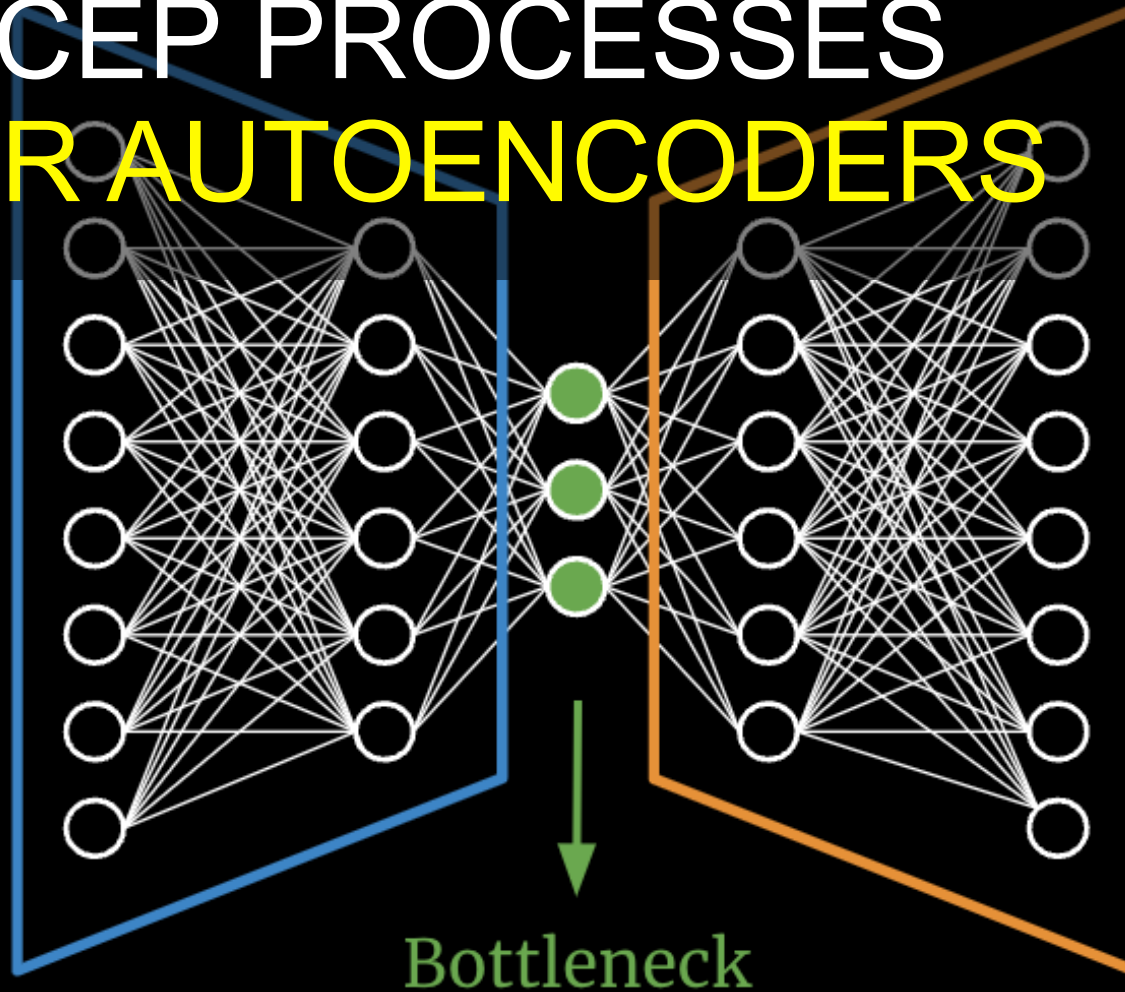
ALICE

Run 3 Pb-Pb
 $\sqrt{s_{NN}} = 5.36 \text{ TeV}$

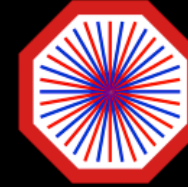
Encoder

Decoder

UPC AND CEP PROCESSES NICE FOR AUTOENCODERS



**In a different way, at
the analysis level!
With all the
reconstructed event**



ALICE

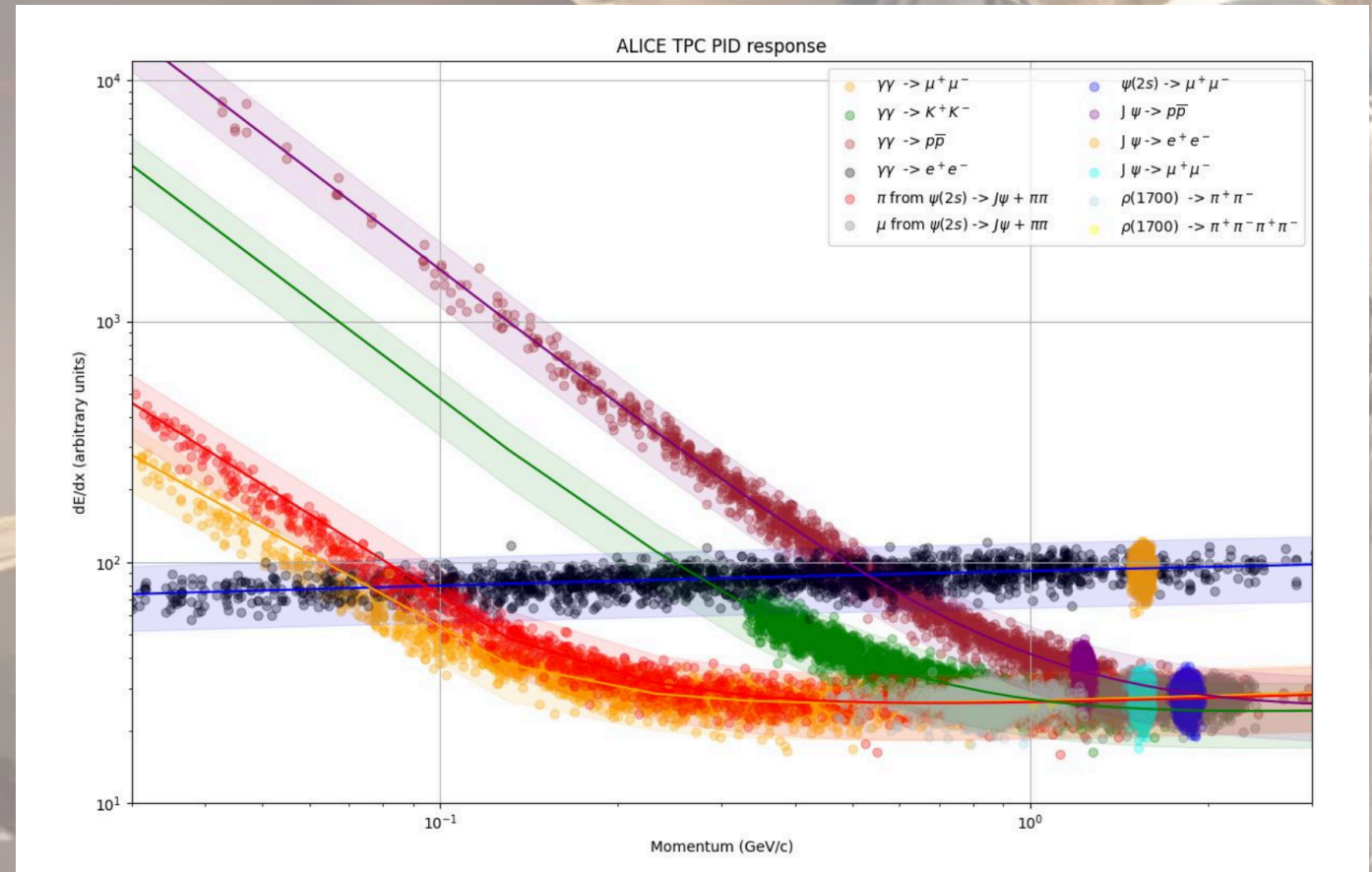
Pb-Pb Run 2
 $\sqrt{s_{NN}} = 5.02 \text{ TeV}$

... what happens in a real life situation instead?

Remember: this is a strategy to do all possible searches for New Physics in a single go!

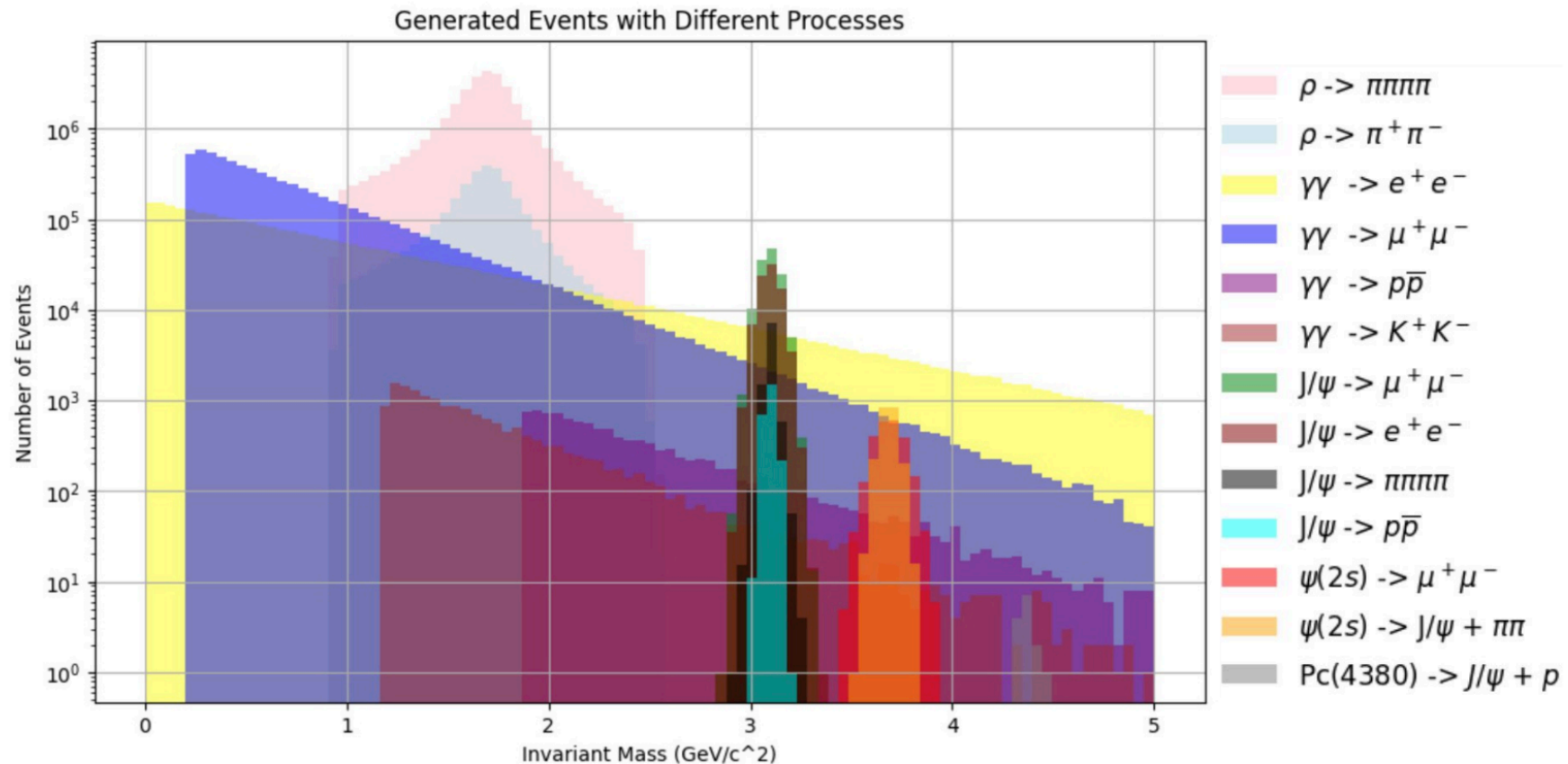
Simulating real detector conditions

- We start generating all the *exclusive UPC* processes measured until now by ALICE
- We apply kinematic cuts
- And PID

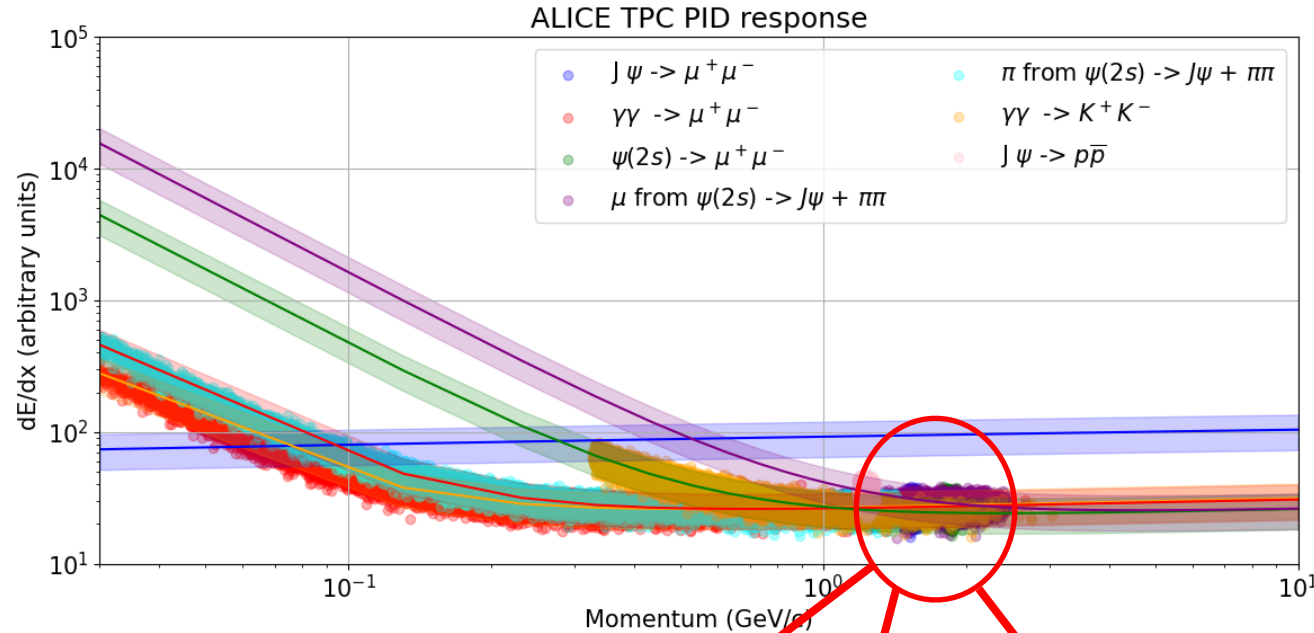


Simulating real detector conditions

- We start generating all the *exclusive UPC* processes measured until now by ALICE
- We ap
- And PI



Two strategies for autoencoders



FIRST DESIGN: a particle is assigned an identity based on a $N\sigma$ cut

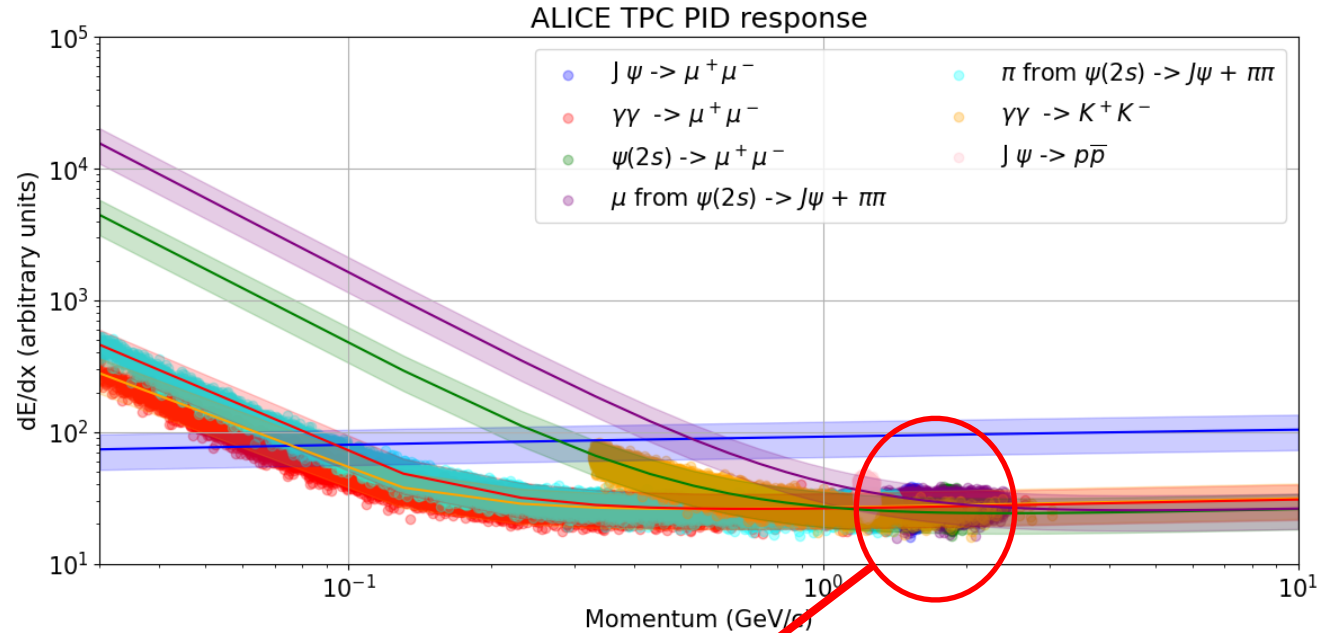
This means that the same particle may be assigned multiple identities

This is because processes may overlap each other

Event ID	N Tracks	N Electron candidates	N Muon candidates	N Pion candidates	N Kaon candidates	N Proton candidates
0	2	0	2	2	2	1
1	2	0	2	2	2	1
2	2	0	2	1	1	1

The total number of tracks is needed to constrain the misidentification

Two strategies for autoencoders



SECOND DESIGN: no specific identity is assigned

We just give each possible PID response value

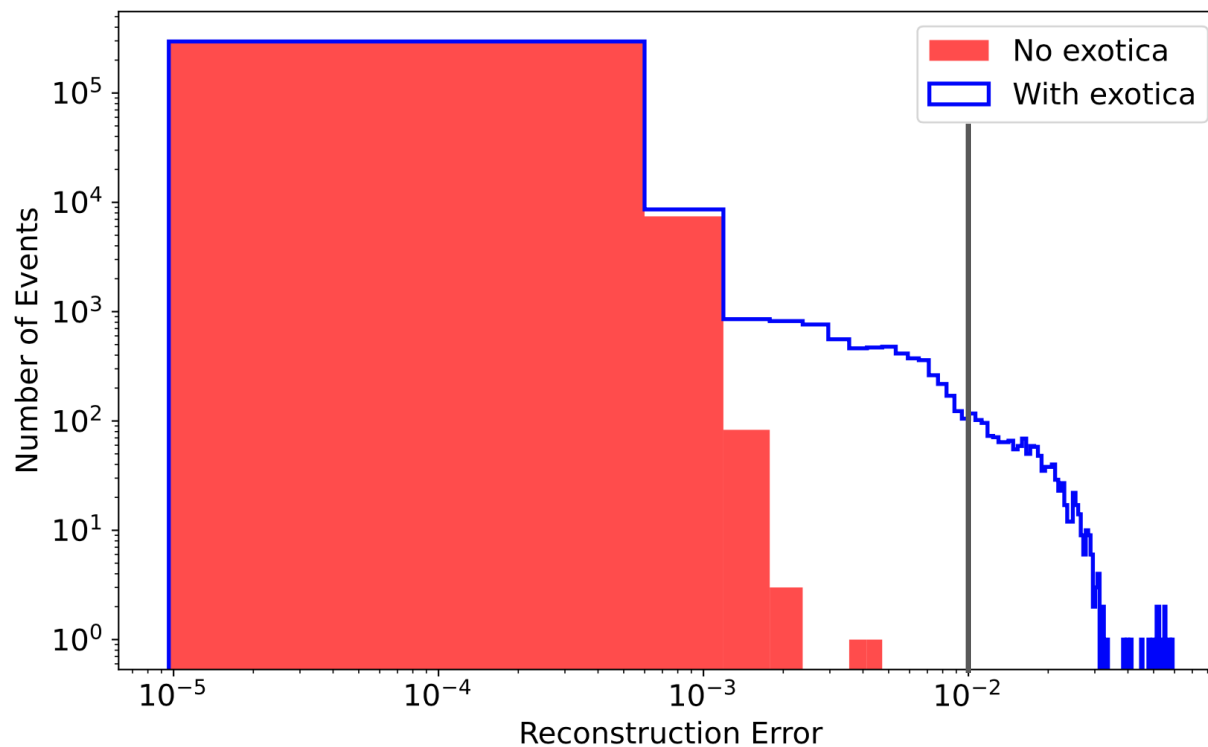
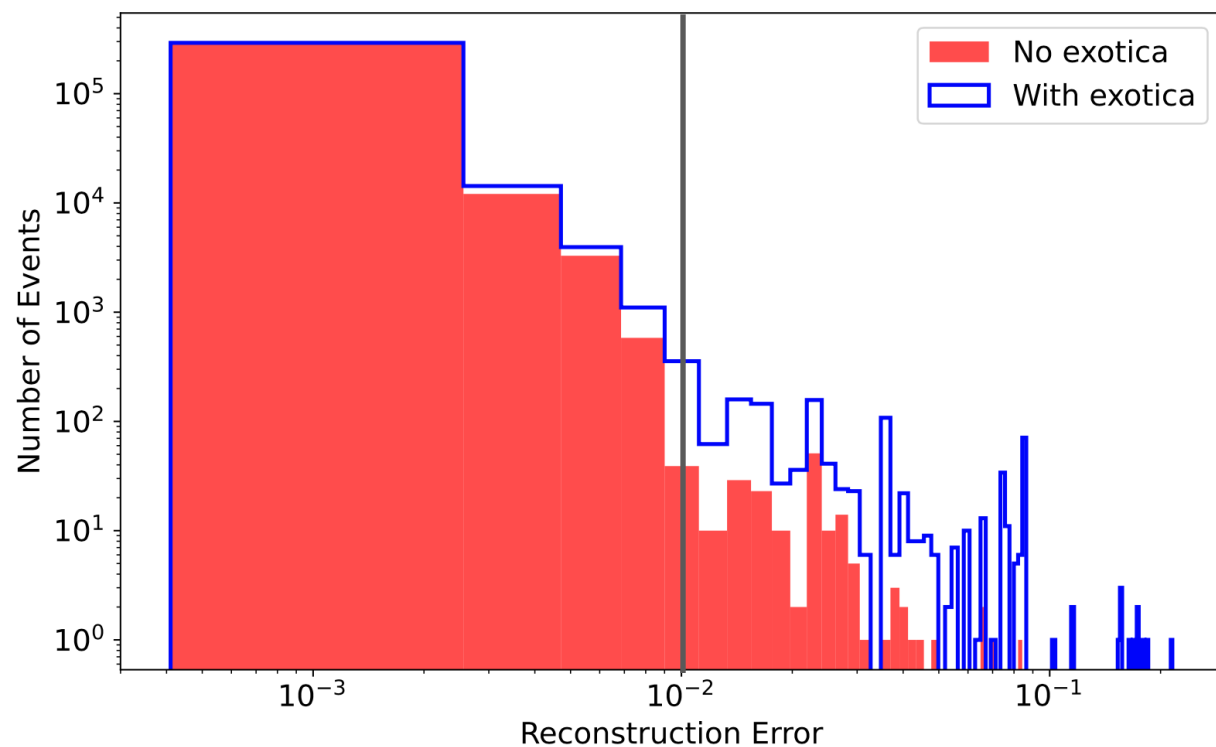
Event ID	N Tracks	$N\sigma_e^1$	$N\sigma_\mu^1$	$N\sigma_\pi^1$	$N\sigma_K^1$	$N\sigma_p^1$	$N\sigma_e^2$	$N\sigma_\mu^2$	...	$N\sigma_p^4$	$N\sigma_K^4$
0	2	2.370	0.020	0.041	0.349	0.389	2.389	0.039	...	10	10
1	2	2.293	0.248	0.315	0.643	0.160	2.376	0.119	...	10	10
2	2	2.337	0.097	0.160	0.469	0.310	2.330	0.019	...	10	10

The total number of tracks is needed to constrain the numbers of inputs

Testing on a Realistic Dataset

$$\text{MSE} = \frac{1}{\sigma} \sum_{i=1}^{\sigma} (x_i - \hat{x}_i)^2$$

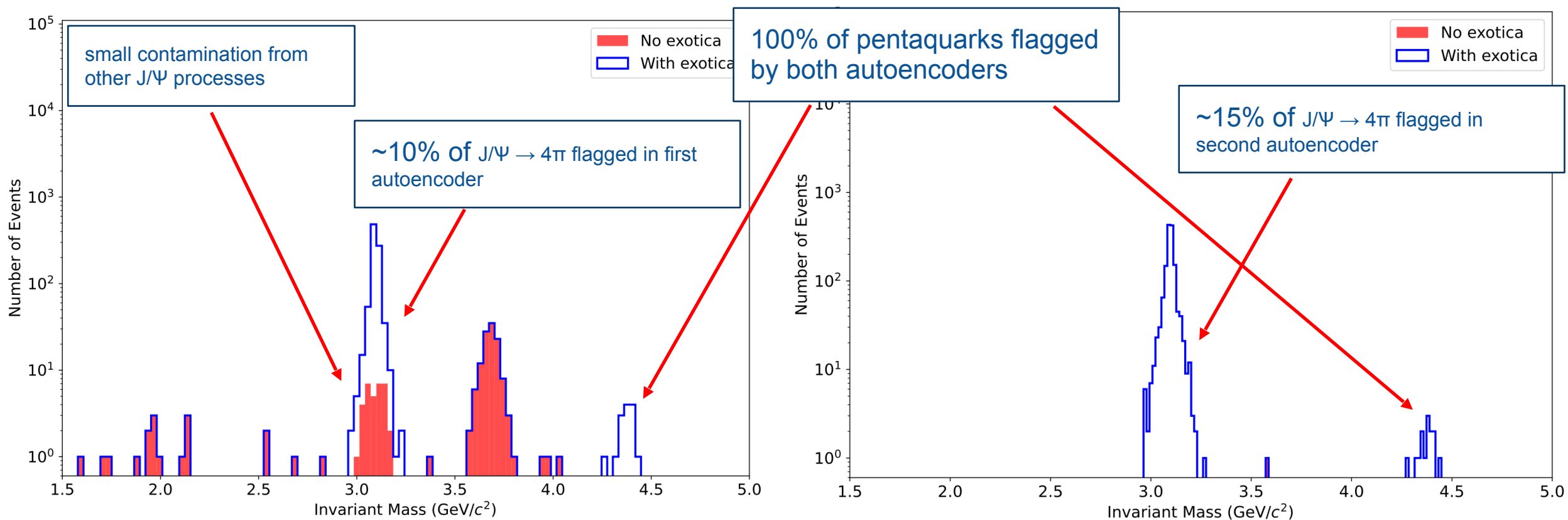
- Anomalous events extend to higher MSE values



Distribution of MSE values for the first and second autoencoders respectively

Testing on a Realistic Dataset

- Invariant masses of events with MSE values > 0.01



Invariant mass distributions for events flagged as anomalous for the first and second autoencoders respectively



ALICE

Pb-Pb Run 2
 $\sqrt{s_{NN}} = 5.02 \text{ TeV}$

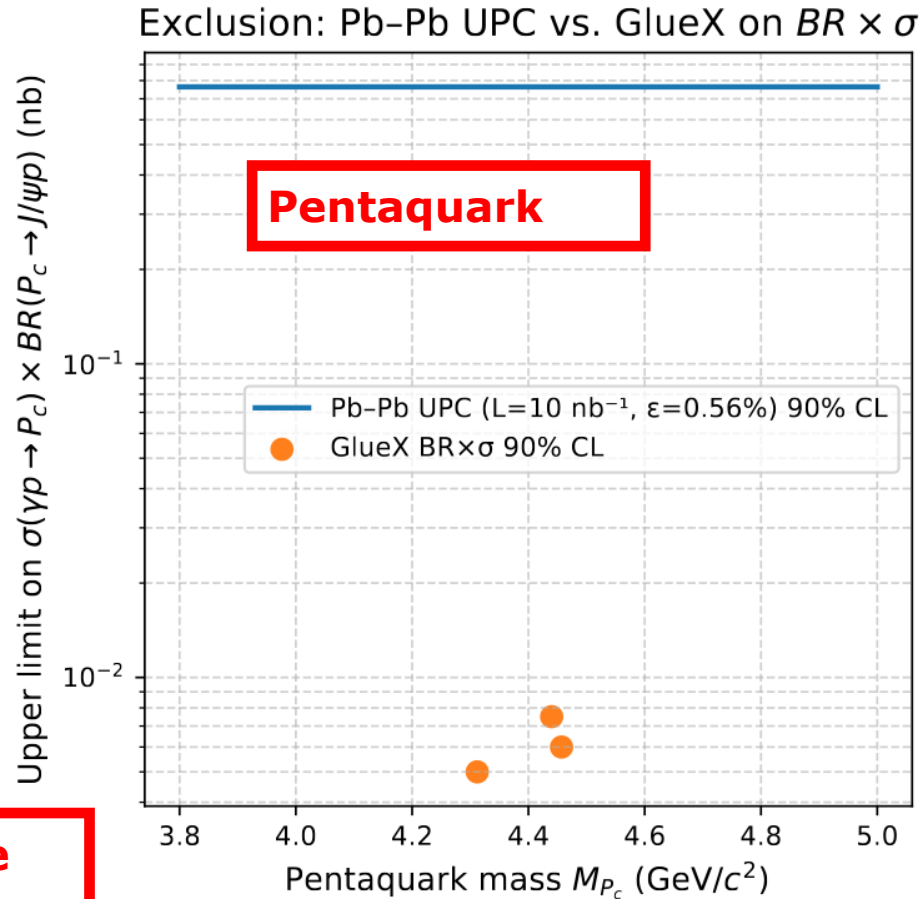
What if you do not find
anything after applying the
autoencoder technique?

**No worries! It becomes
a good exclusion limit**



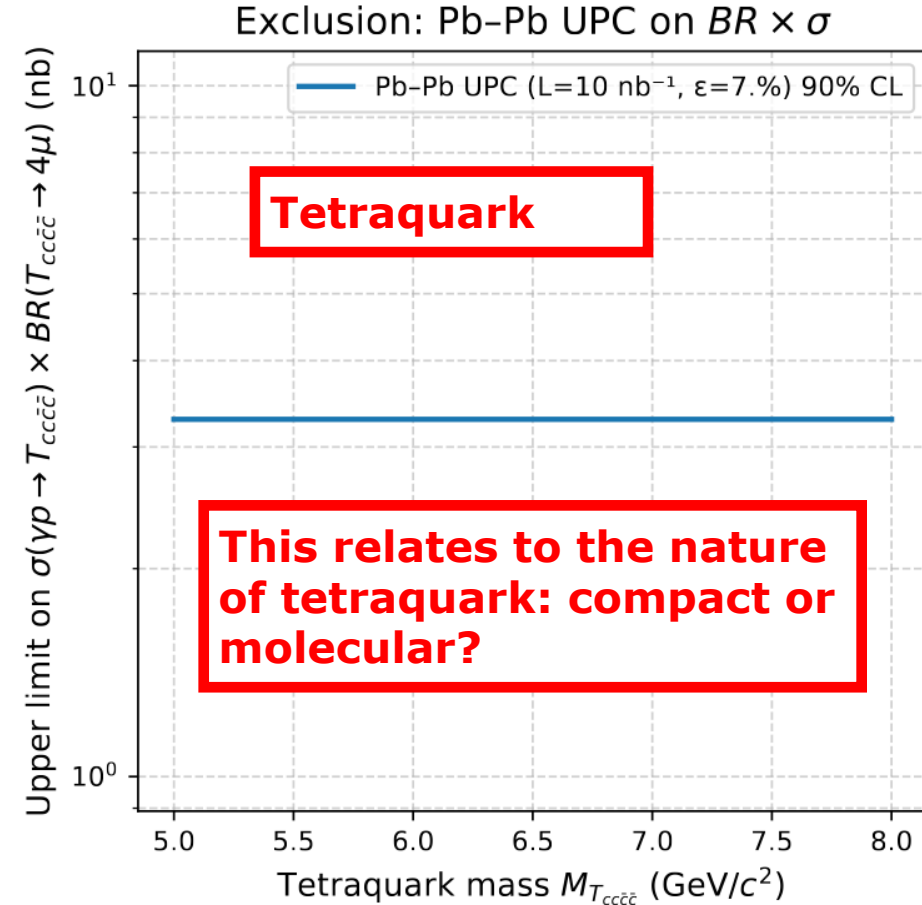
Exclusion Limits

No one has measured tetraquark production in UPCs yet



Limits in the case of no background (as it turns out from the autoencoder)

GlueX has more stringent limits but we can extend them with this technique



This relates to the nature of tetraquark: compact or molecular?



ALICE

Pb-Pb Run 2
 $\sqrt{s_{NN}} = 5.02 \text{ TeV}$

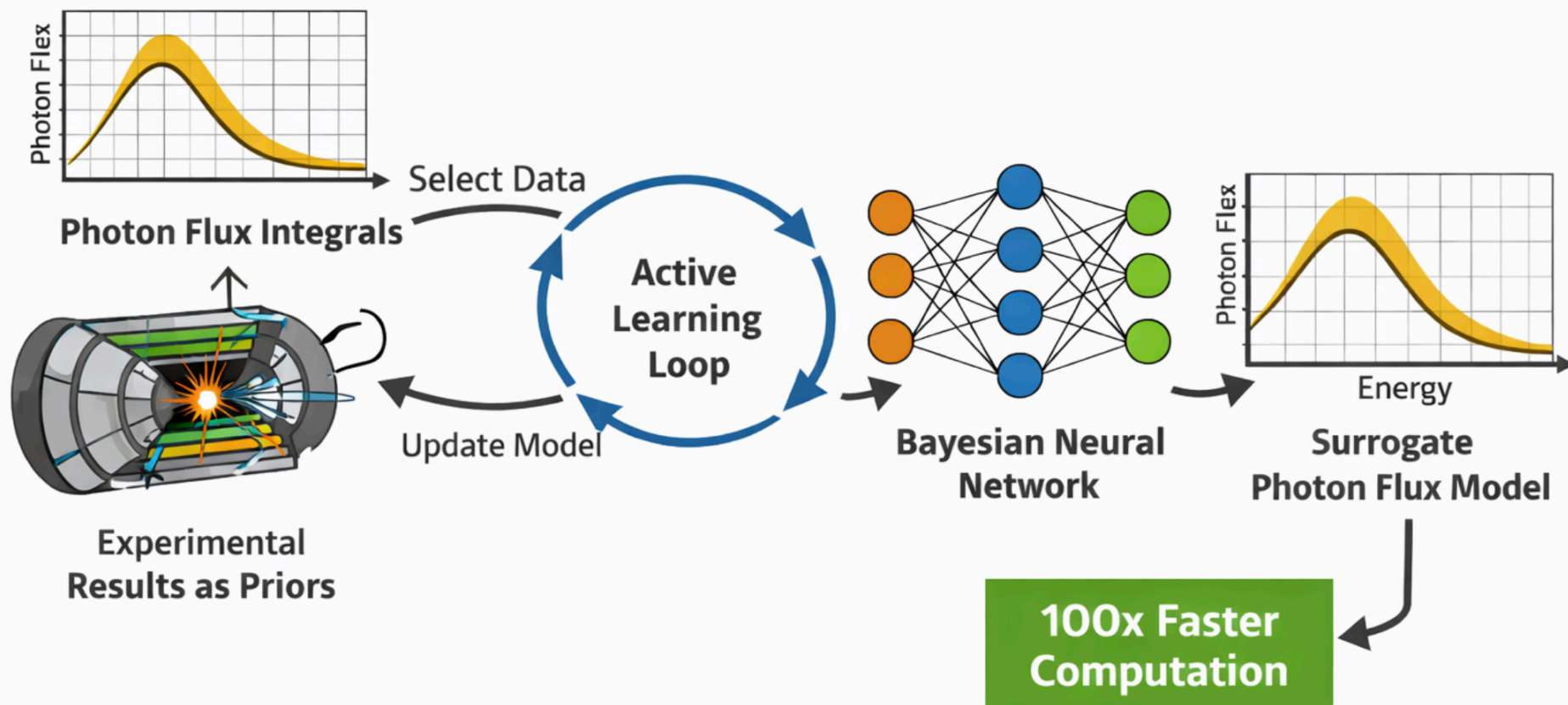
**Brianna Kinkaid's MSc thesis:
successfully defended on
June 12th - committee very
impressed too**

- Using autoencoders for the simultaneous analysis of the entire UPC dataset - in all possible decay channels! (Normally each channel is examined individually)
- Useful also for exclusion limits for New Physics





Accelerating modelling



Based on arXiv:2602.10692



Target for training (photon flux)

$$N(\omega, b) = \int_0^{2\pi} d\phi \int_0^{R_{\text{nuc}}} R dR \rho_{2D}(R) n_{\text{pt}}(\omega, |\mathbf{C}|)$$

$$\mathbf{x} = [\ln \omega, b, \Delta R, a] \in \mathbb{R}^4$$

Feature vector (simplified picture)

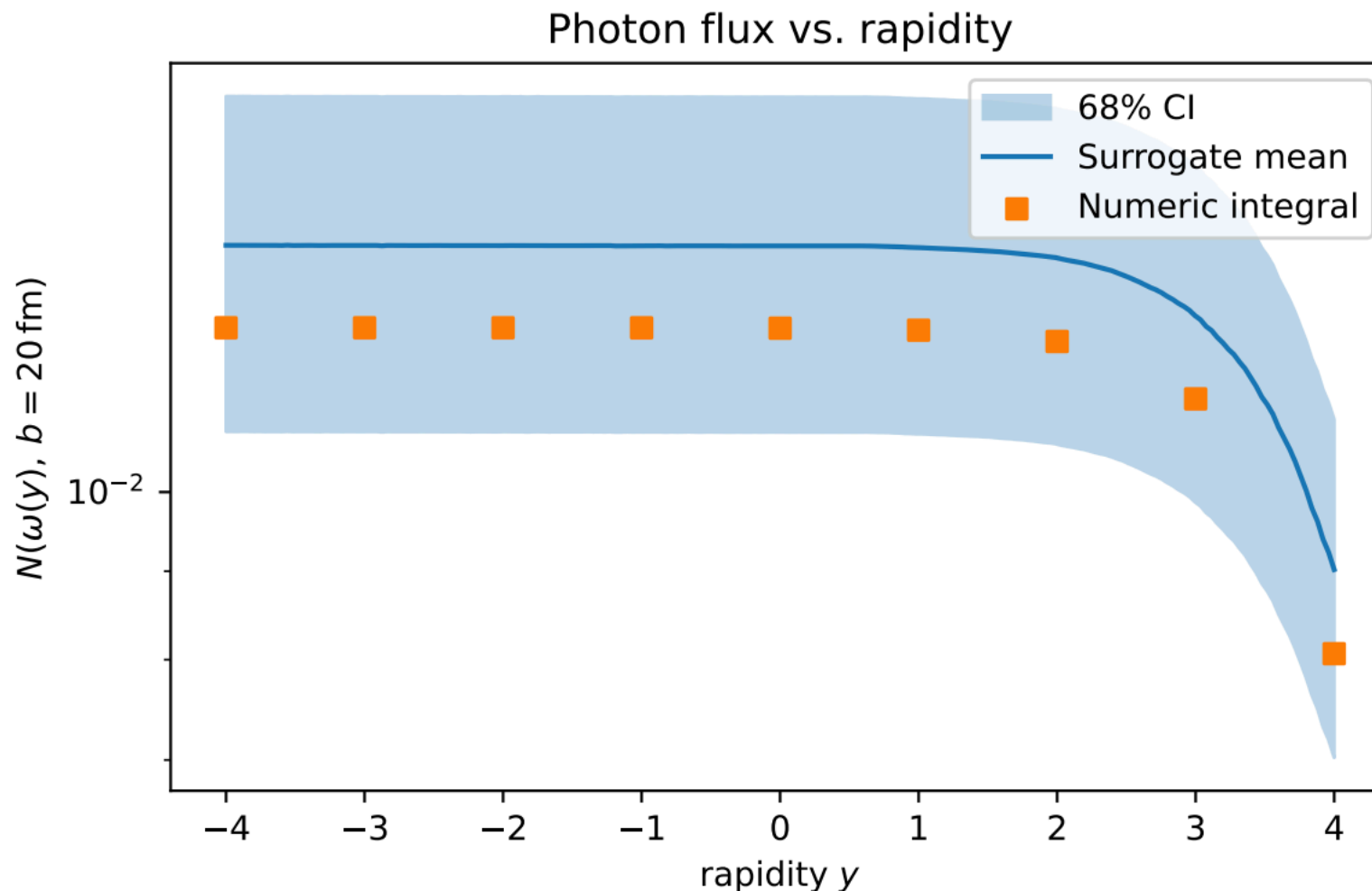


$$\Delta R \sim \mathcal{N}(0.283, 0.071^2) \text{ fm}$$

$$\Delta a \sim \mathcal{N}(0, 0.015^2) \text{ fm}$$

$$\mathbf{x} = [\ln \omega, b, \Delta R, a] \in \mathbb{R}^4$$

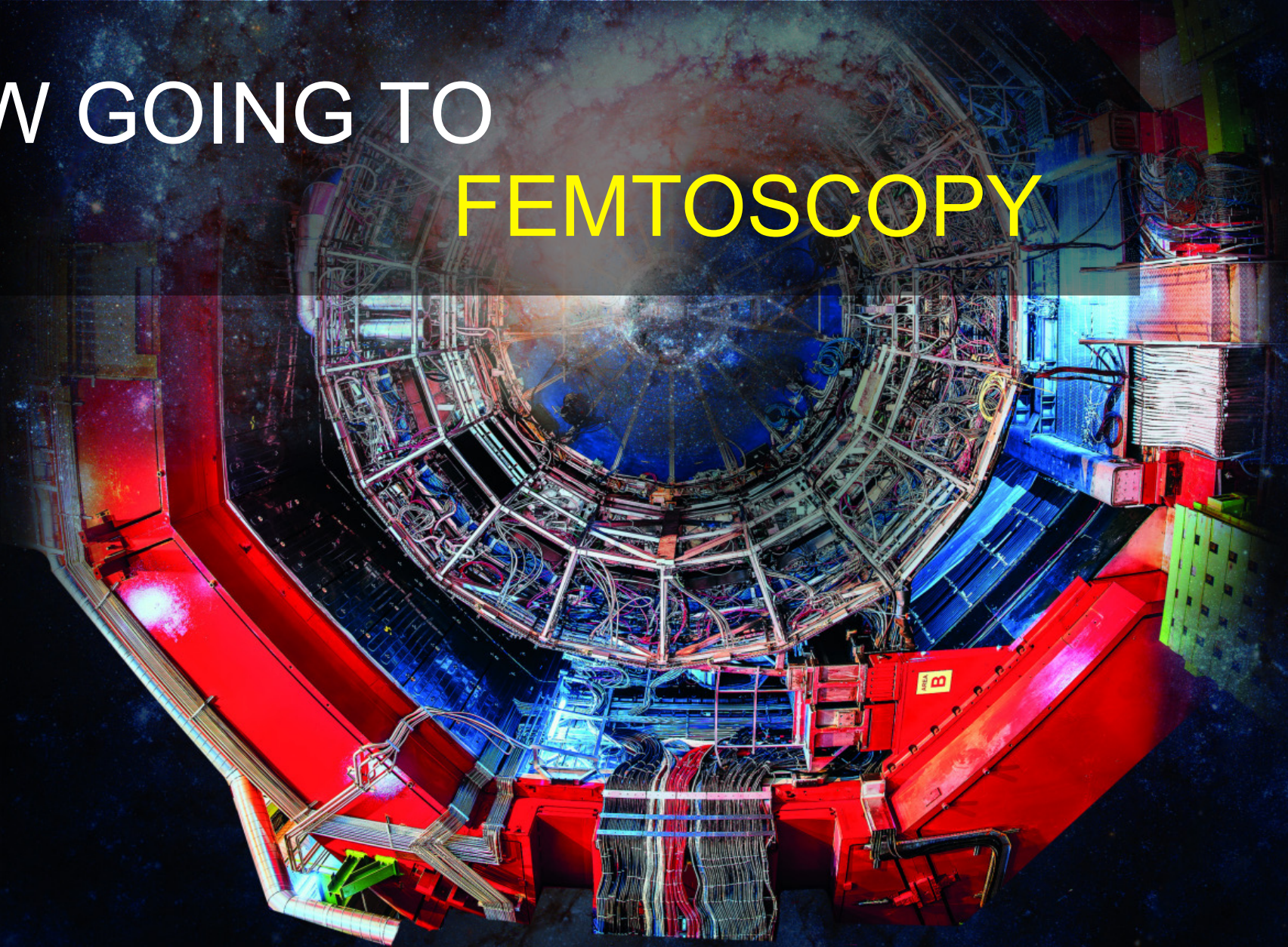
Feature vector (simplified picture)



**Gives
uncertainty of
the photon flux
in a Bayesian
approach &
provides
 $O(100)$
acceleration**

**Iterative procedure to produce the
training sample (reducing the
number of needed integrals)**

NOW GOING TO
FEMTOSCOPY

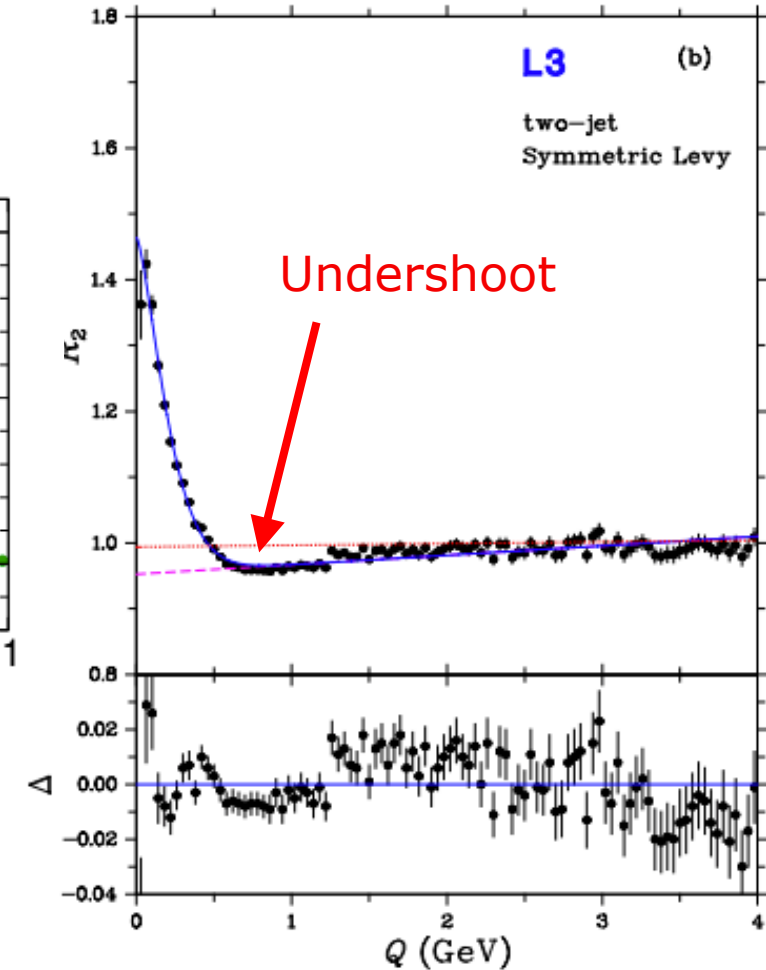
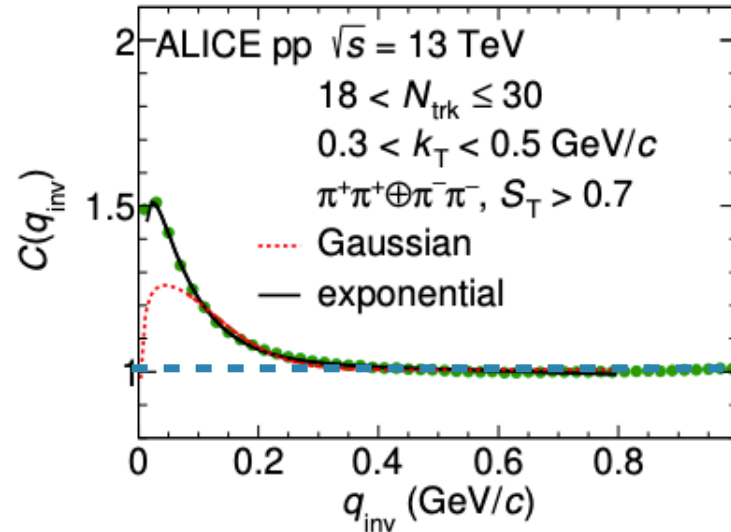


You can build correlation functions out of everything



Correlation functions:

- Have been measured in ee, pp, pPb, PbPb, etc
- Link between ee and the other collision systems is still unclear
- Undershoot typical of ee (what is its nature?)

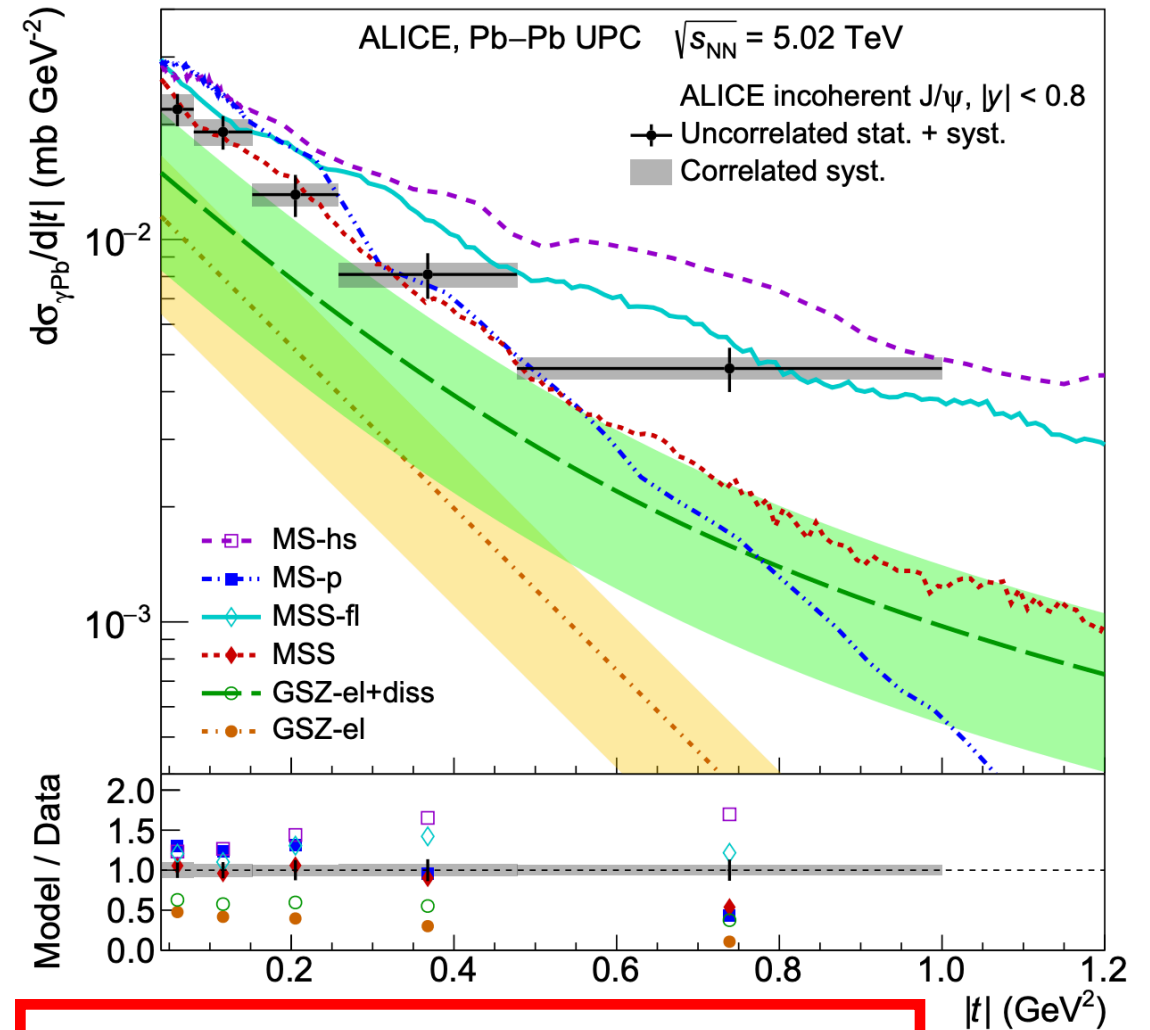


Subnucleonic structures



Our result:

- Demonstrated the presence of subnucleon fluctuations
- First attempt to search for saturation effects in heavy nuclei using fluctuations
- Can we do more?



Phys.Rev.Lett. 132 (2024) 16, 162302

First considerations on the UPC source



Still a lot of unknowns:

- Nature of the emission
- Is it coherent or fully incoherent? No “traditional” medium
- Is it really a “proper” source?
- What about the link to e.g. a Gaussian source?

$$\tilde{S}(r^*, x) \propto \underbrace{\left[\int_{b > 2R_A} d^2b \int d\omega N_\gamma(\omega, b) \right]}_{\text{EPA photon flux}} \times \underbrace{S_{\text{shad}}(x, b)}_{\text{shadowing}} \underbrace{\int_0^1 dz |\Psi_{\gamma \rightarrow q\bar{q}}(r^*, z)|^2}_{\text{dipole}} \underbrace{\left[1 - e^{-r^{*2} Q_s^2(x)/4} \right]}_{\text{gluon saturation}}$$



Bjorken-x

Using a proxy to the DIS formalism:

- Using $m_T = \sqrt{m^2 + p_T^2}$
- $x \approx \frac{2m_T}{\sqrt{s_{NN}}} \cdot e^{\pm y}$, with factor 2 for the meson pair
- Possible to measure pion pairs at midrapidity and with the MFT
- $x \sim 10^{-4}, 10^{-6}$



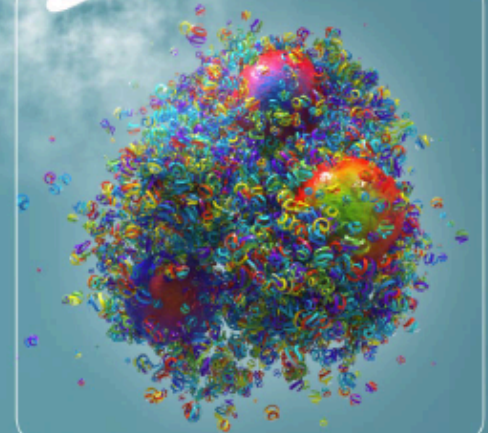
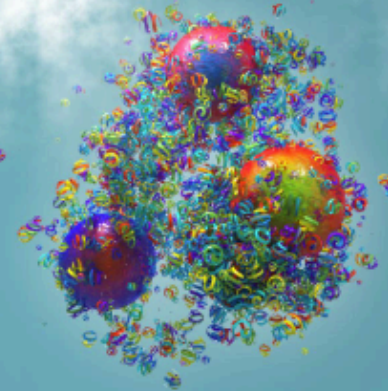
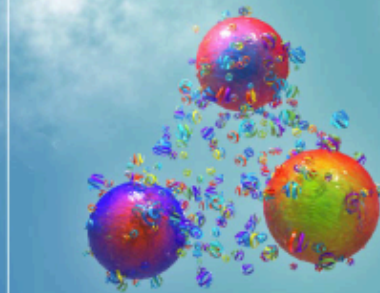
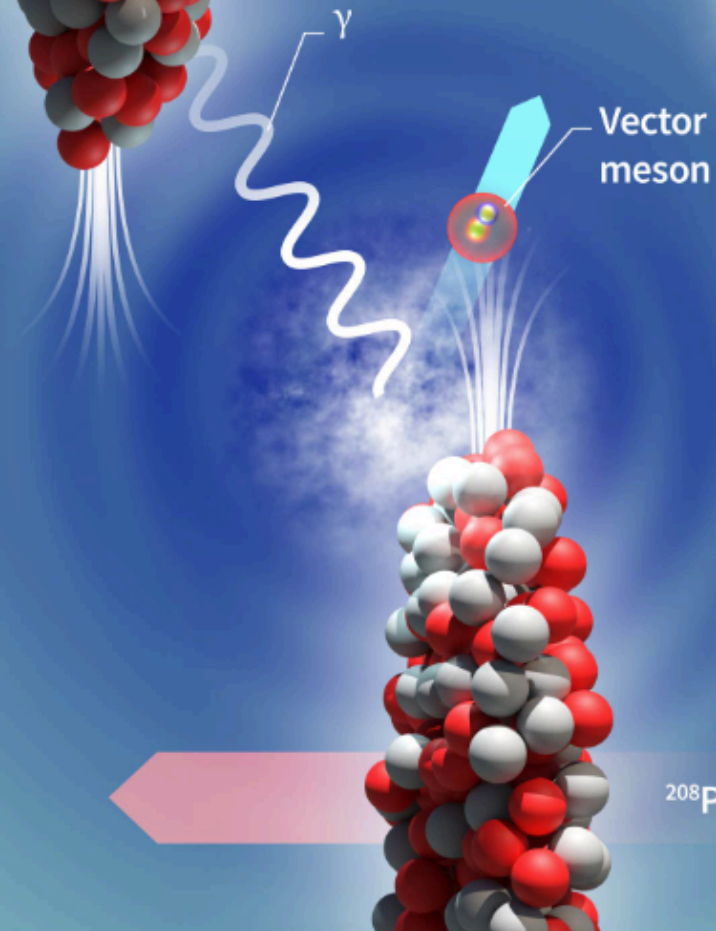
Femto radius vs Bjorken-x

In the weak-saturation limit:

- We can expand the saturation term
- $R_{weak} = \sqrt{5/6}R_0$
- Roughly independent from Bjorken-x

FIRST RESULTS FROM STARLIGHT+DPMJET

^{208}Pb



Photon energy

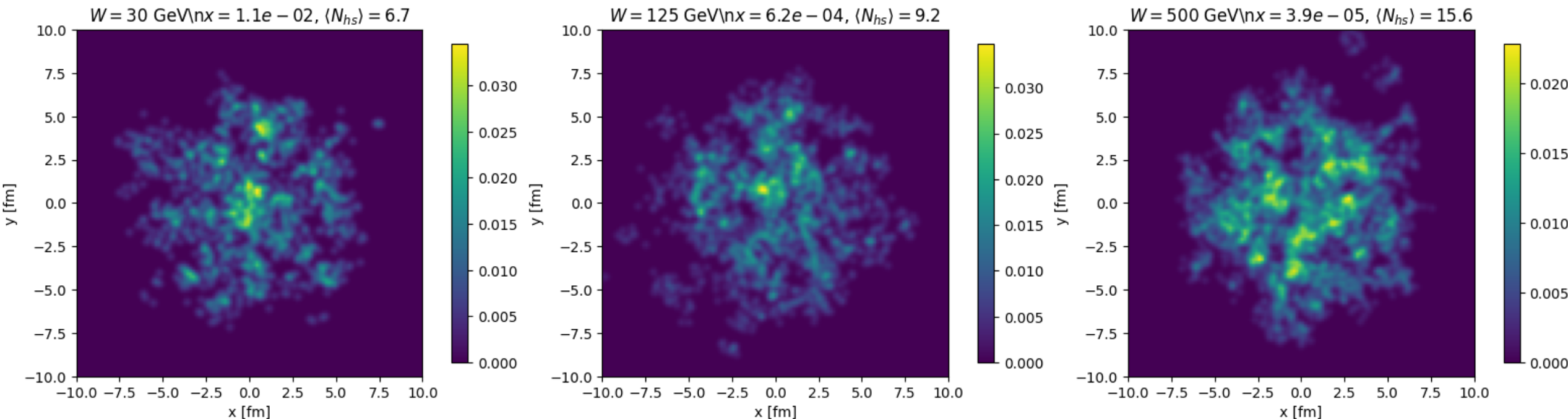
^{208}Pb

Hot spots



Hot spots implemented following the approach in *Phys.Lett.B* 852 (2024) 138613:

- Impact of localised emitting regions?





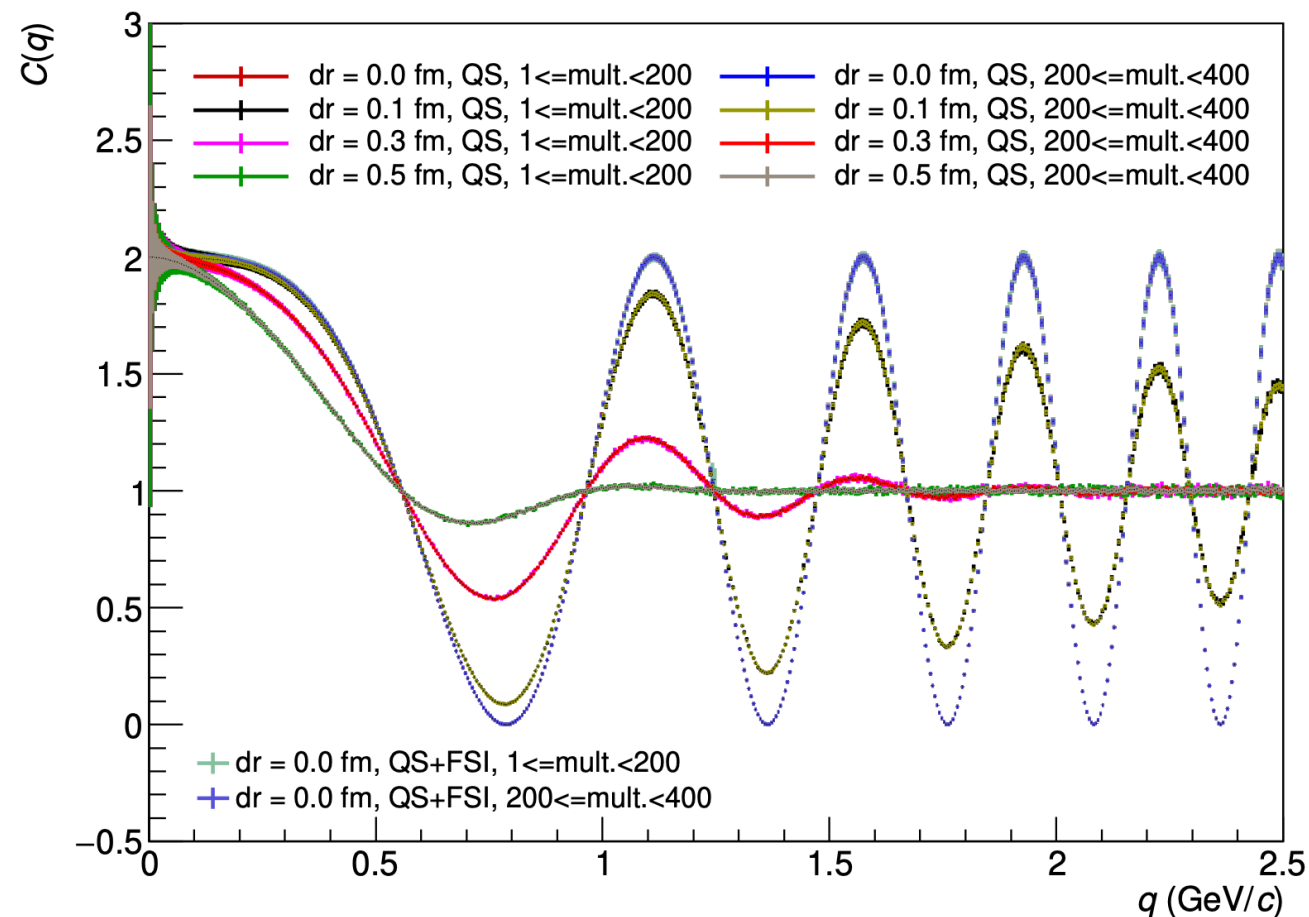
STARlight+DPMJET+afterburner

Characteristics:

- From 0 displacement up to 0.5 fm
- Multiplicities at generated level

Afterburner:

- STARlight+DPMJET shows full oscillatory trend
- Damping introduced by a finite-size source
- Direct link to size of localised emitting regions?
- Link to CGC?



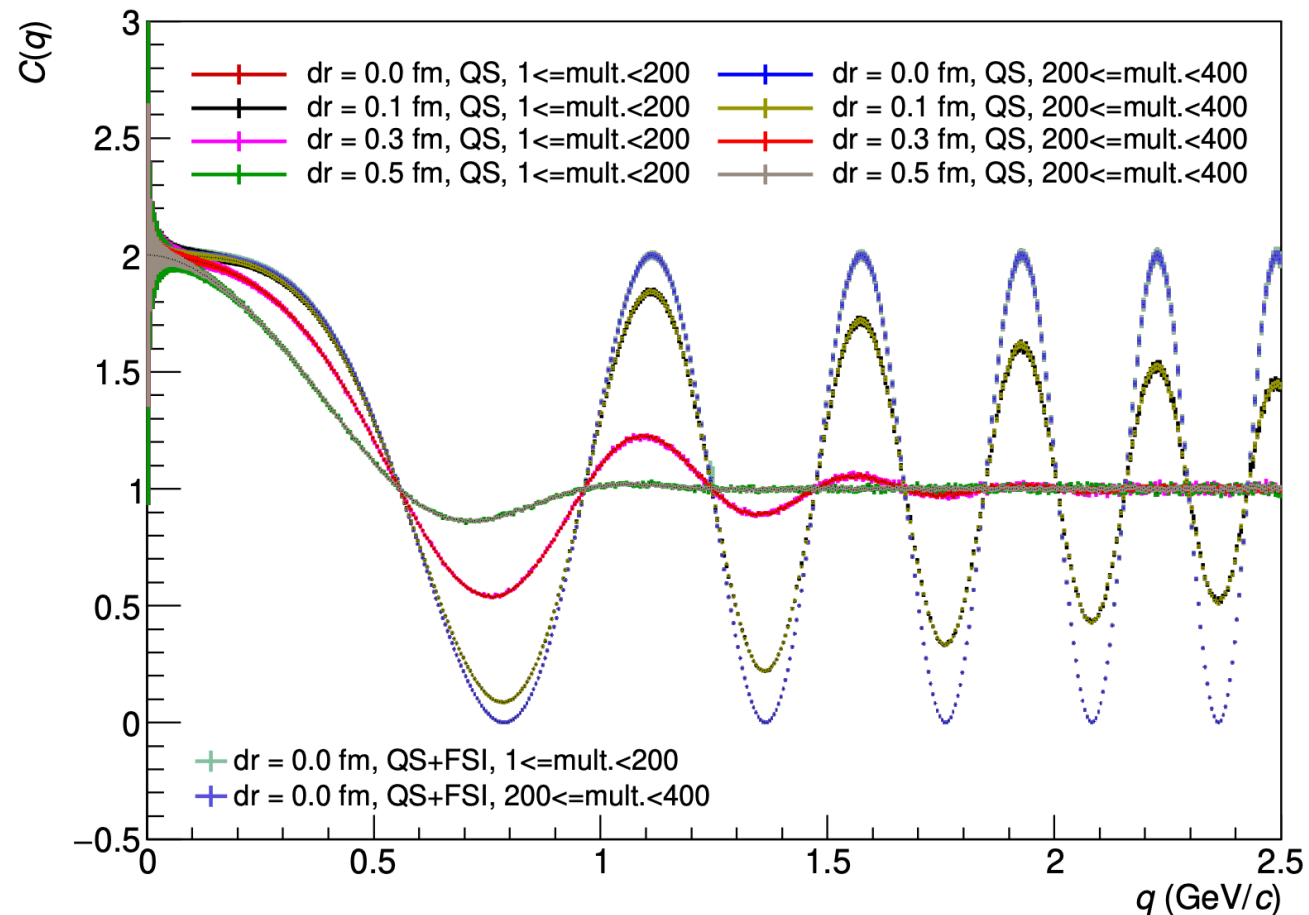
arXiv:2606.11917

STARlight+DPMJET+afterburner

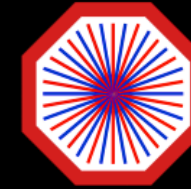


Characteristics:

- Dip at 0.7 fm corresponds to a spatial scale $d \sim \frac{\pi \hbar c}{q} \approx 0.9 \text{ fm}$
- DPMJET string fragmentation scale?
- BUT introducing additional displacements brings down the damping
- Femto sensitive to sub-fermi structures $\rightarrow$ hopes for hot spots



arXiv:2606.11917



ALICE



- First attempt to bring HBT to UPCs
- Still lot of work to do but:
- Link to size of subnucleonic structures