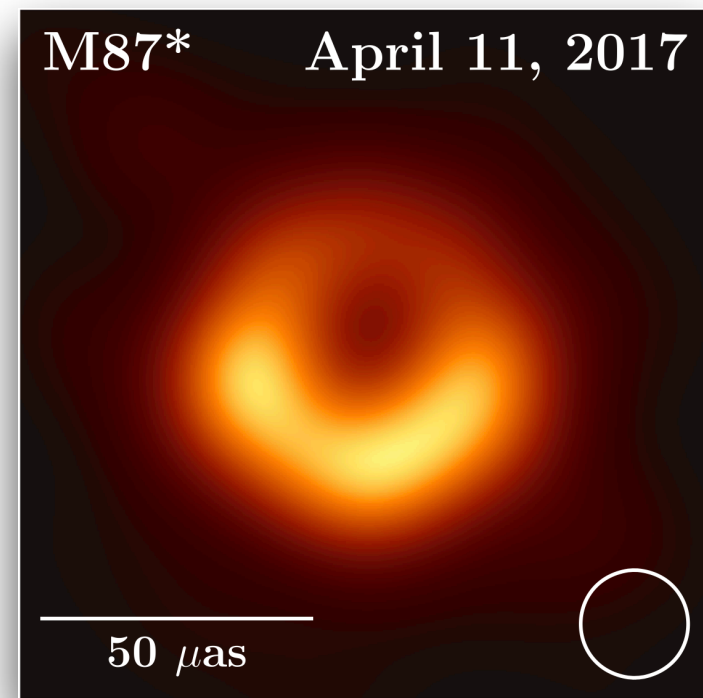
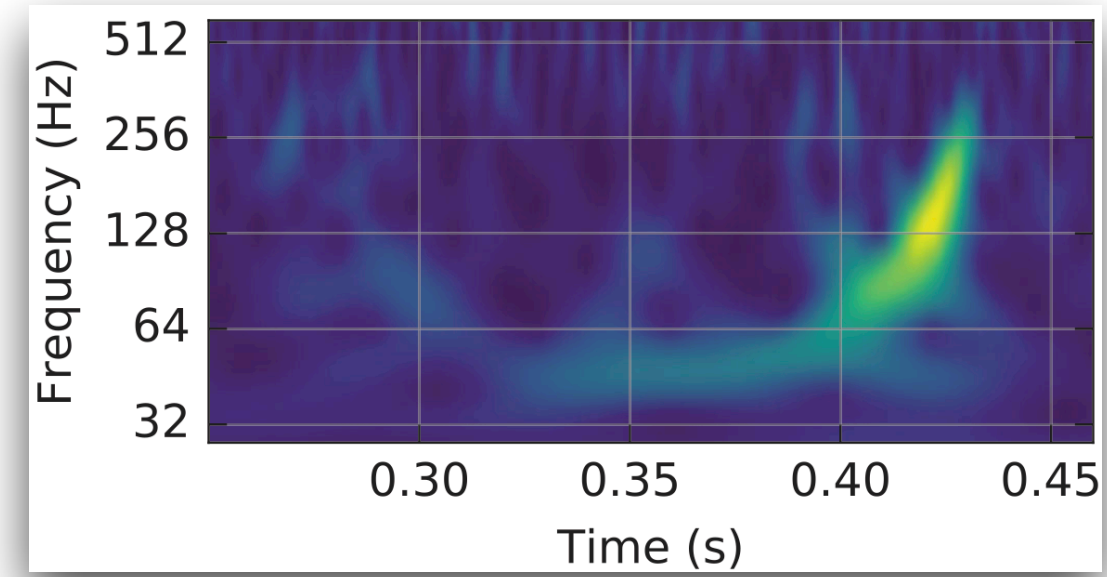


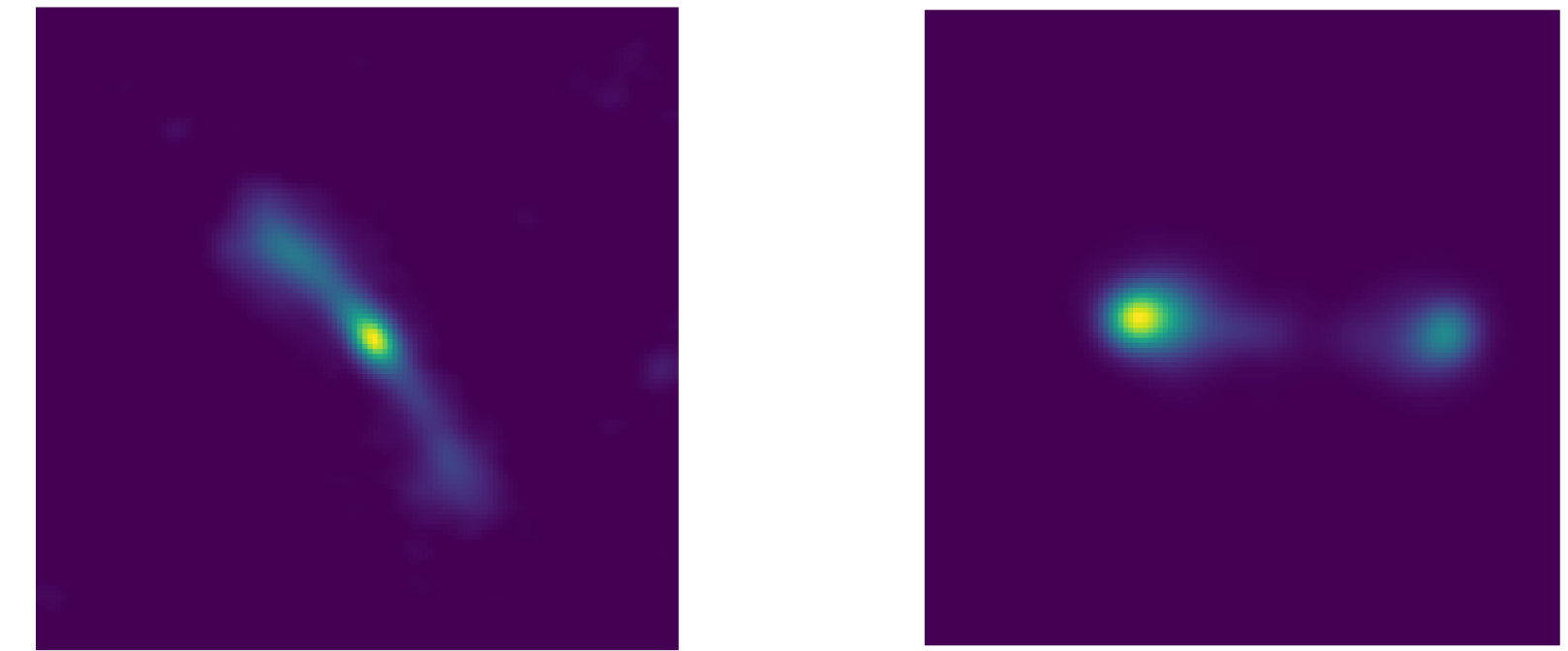
# Classification of Radio Sources Through Self-Supervised Representation Learning



arXiv:1906.11238



arXiv:1602.03837

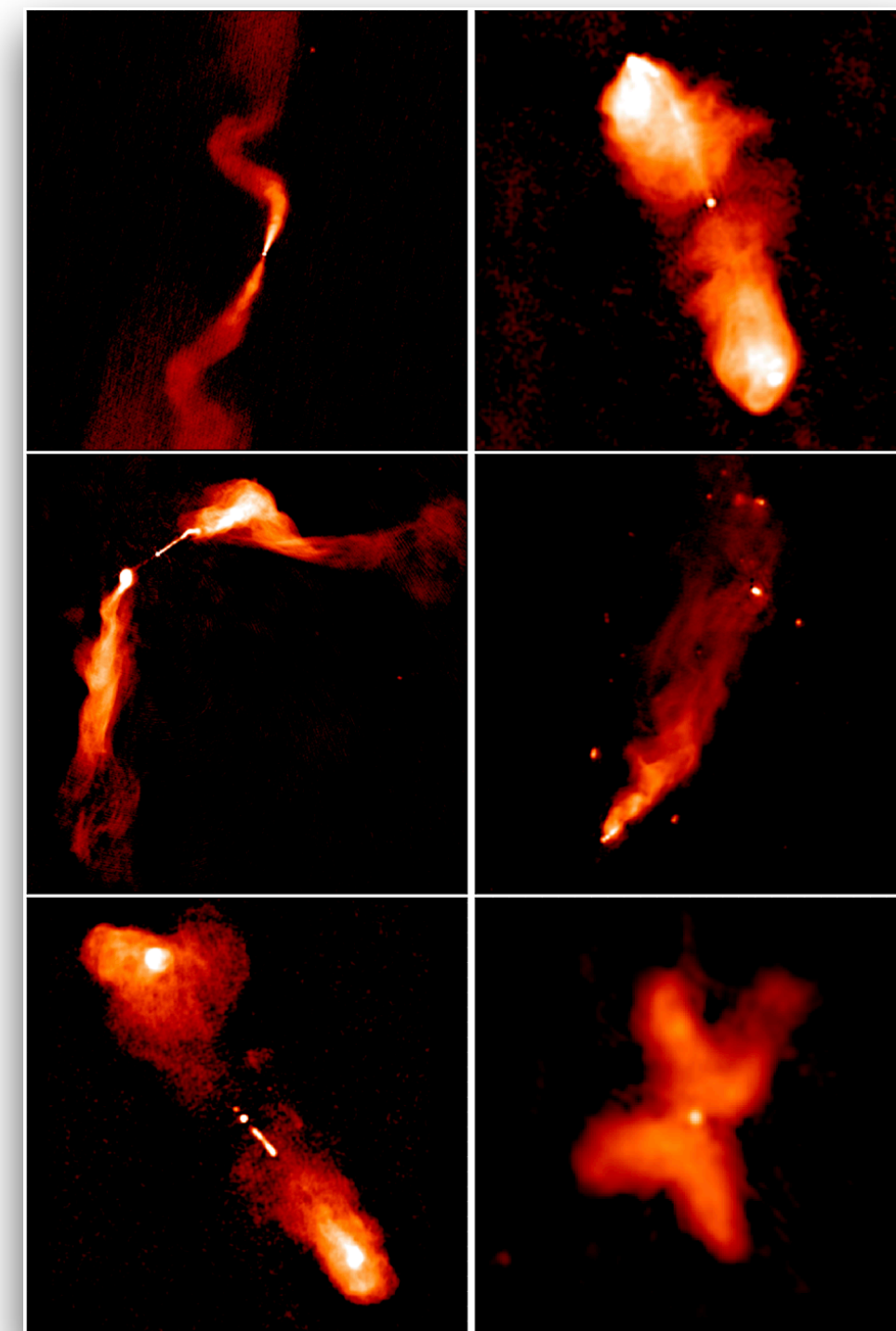


FRI

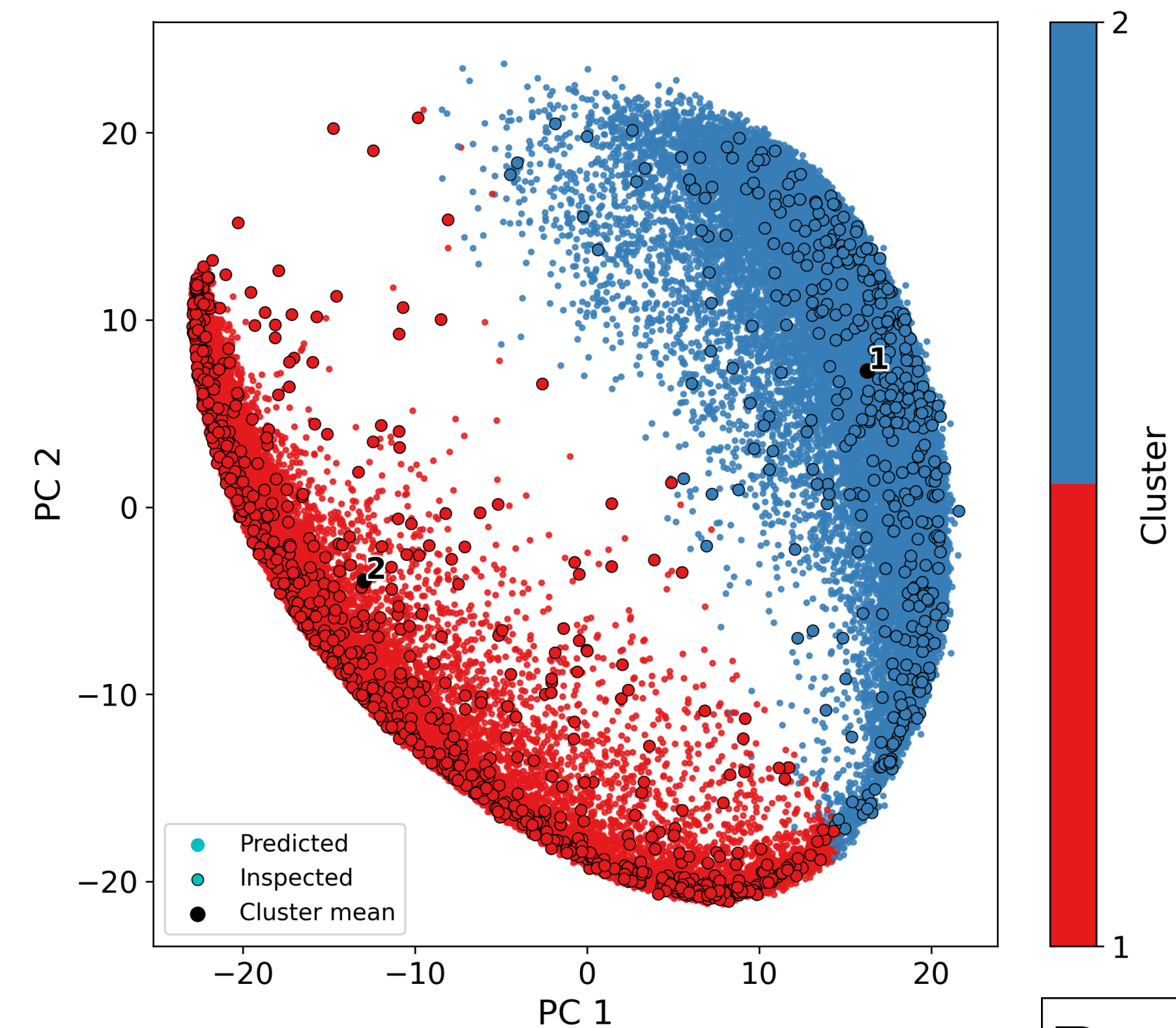
FR II



J. Davelaar 2018

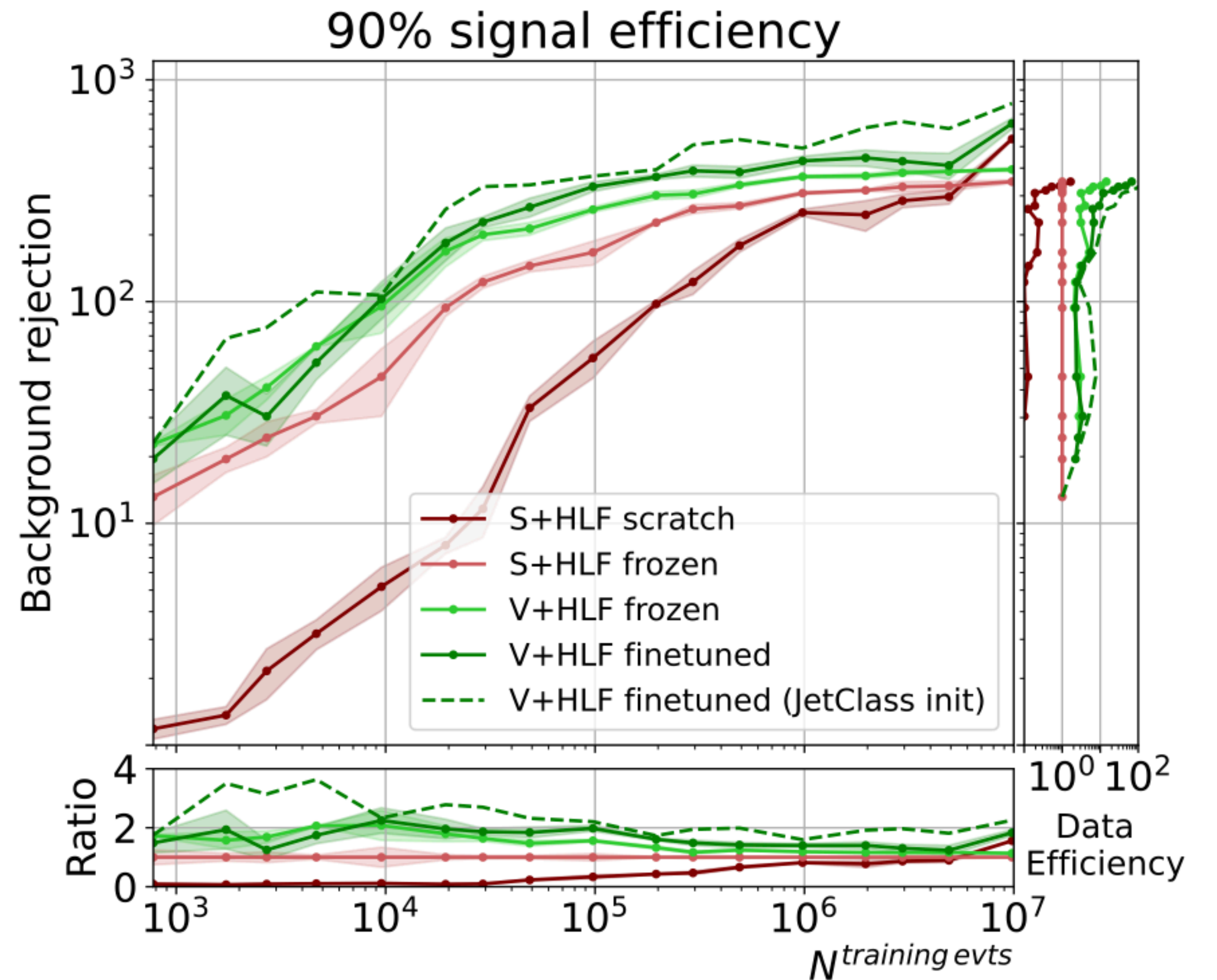


arXiv:2003.06137



# Finetuning Foundation Models

for Joint Analysis Optimization



Matthias Vigl, Nicole Hartman, Lukas Heinrich

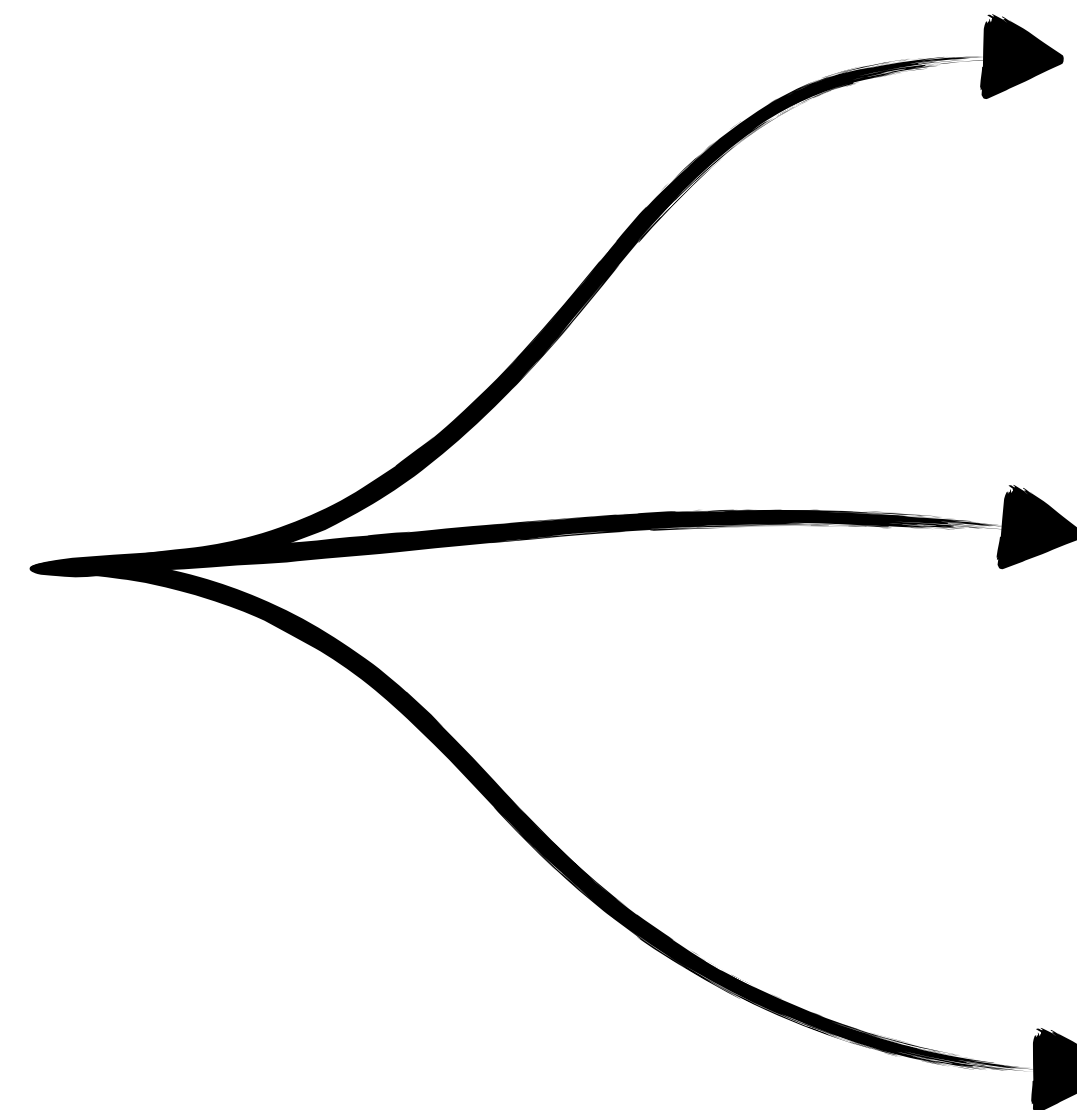
Modality 1

Modality 2

Modality 3



Joint  
Embedding /  
Representation

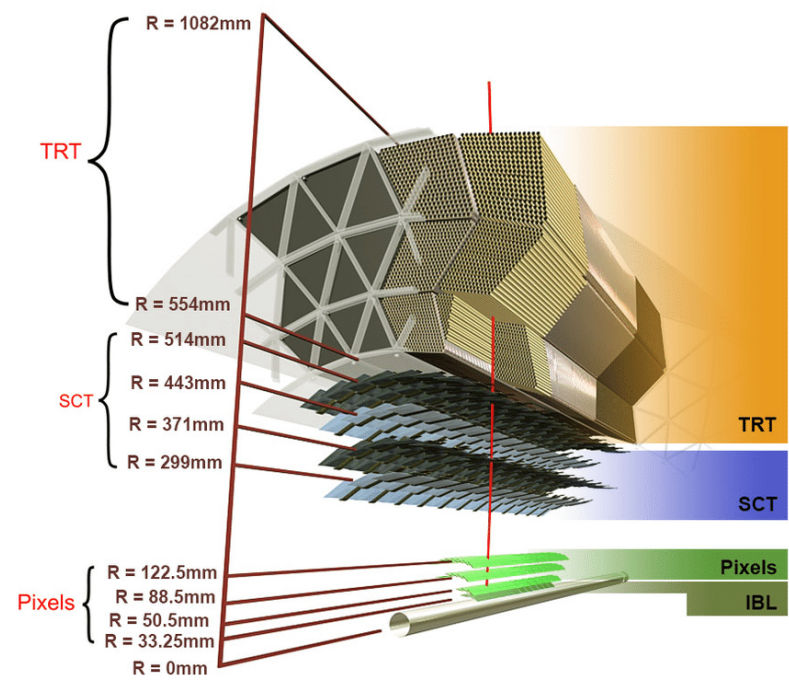


Use Case 1

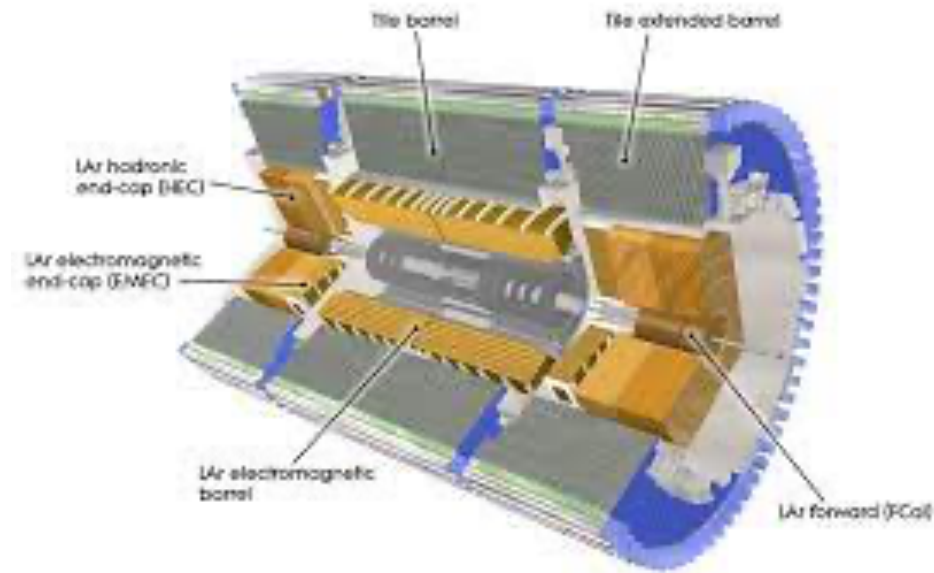
Use Case 2

Use Case 3

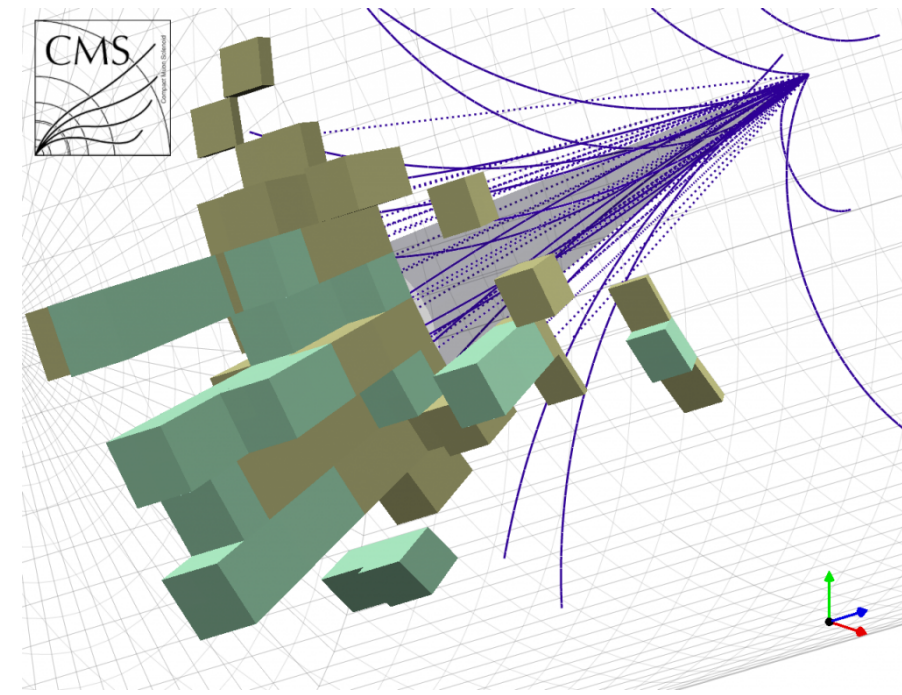
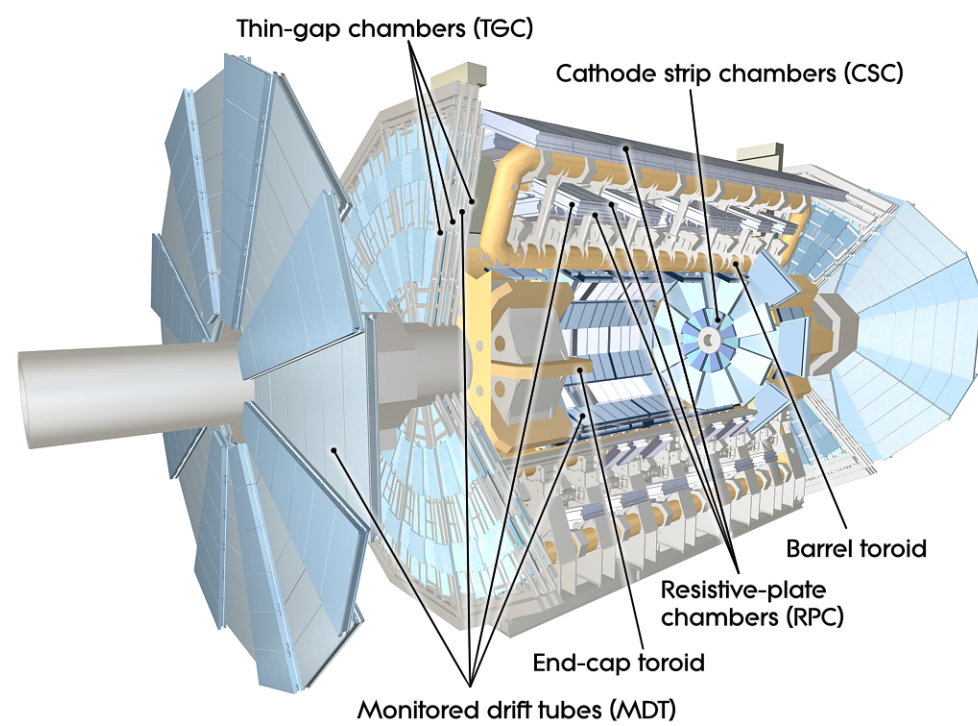
# Tracking Data



# Calorimeter

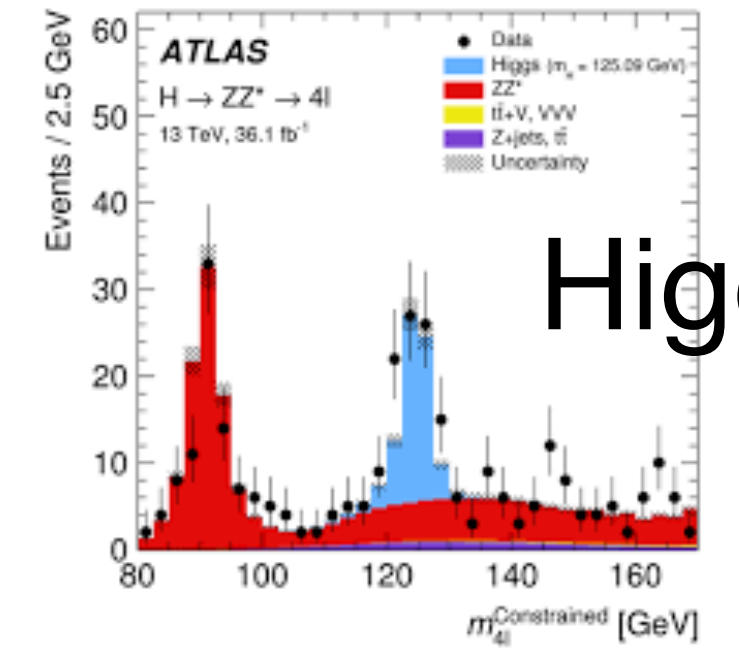


# Muon Data

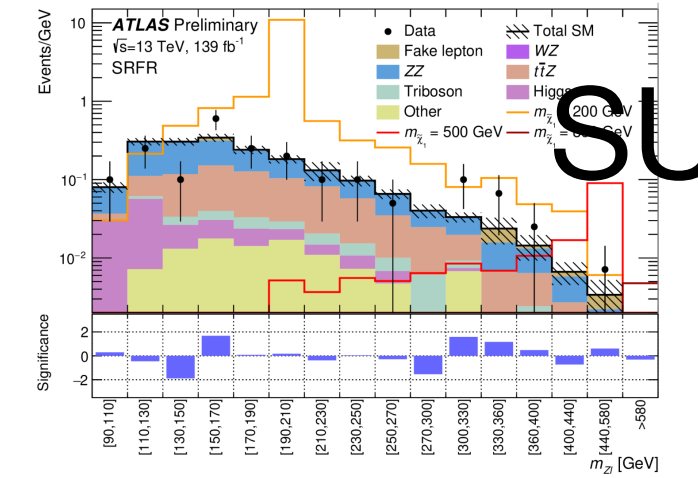


# Reco

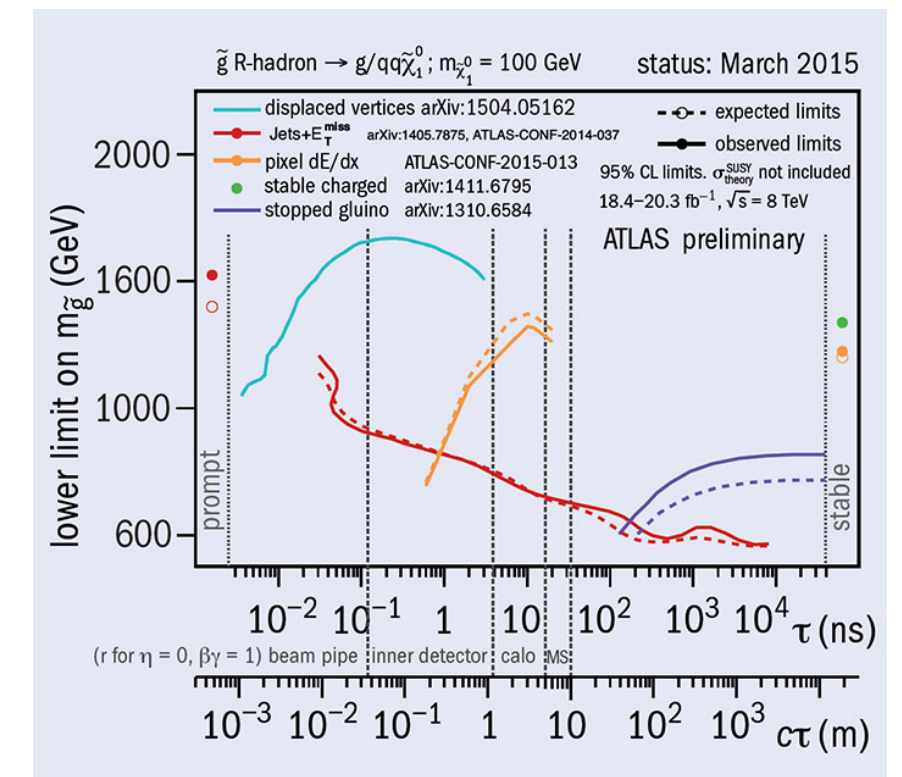
Reconstructed Event



Higgs



SUSY

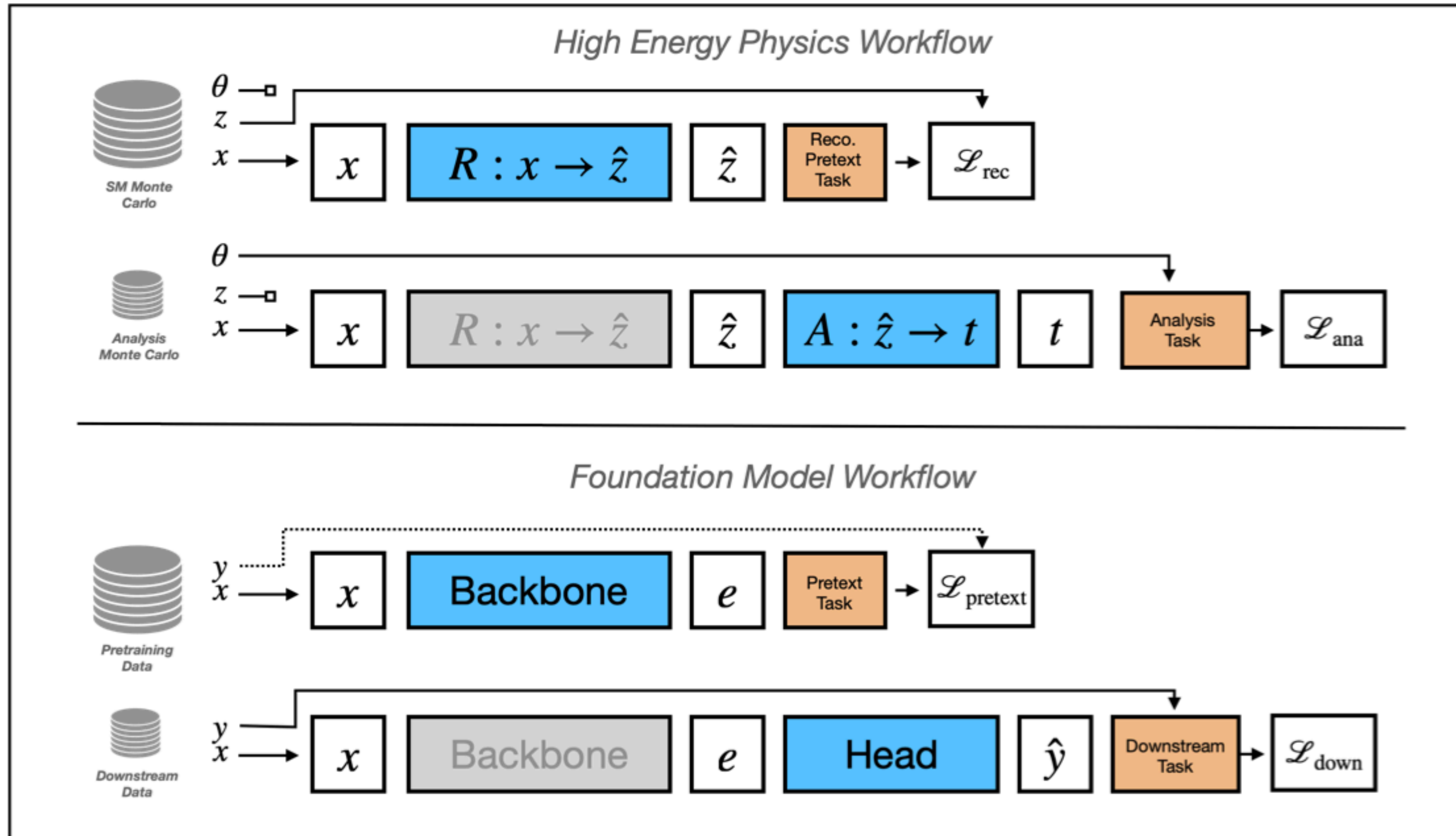


Exotic Particles

*In many ways we've always had a foundation model for general purpose experiments..*

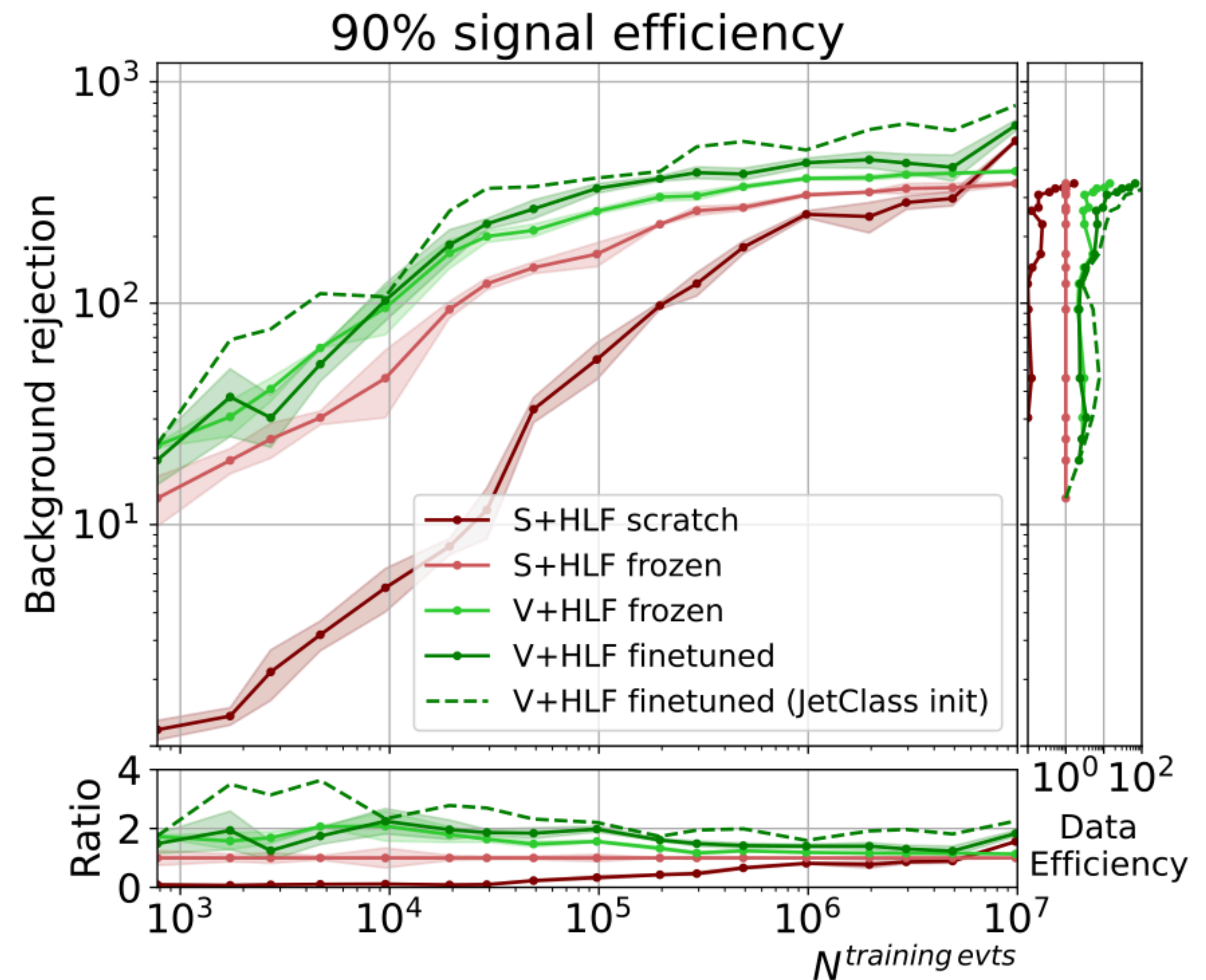
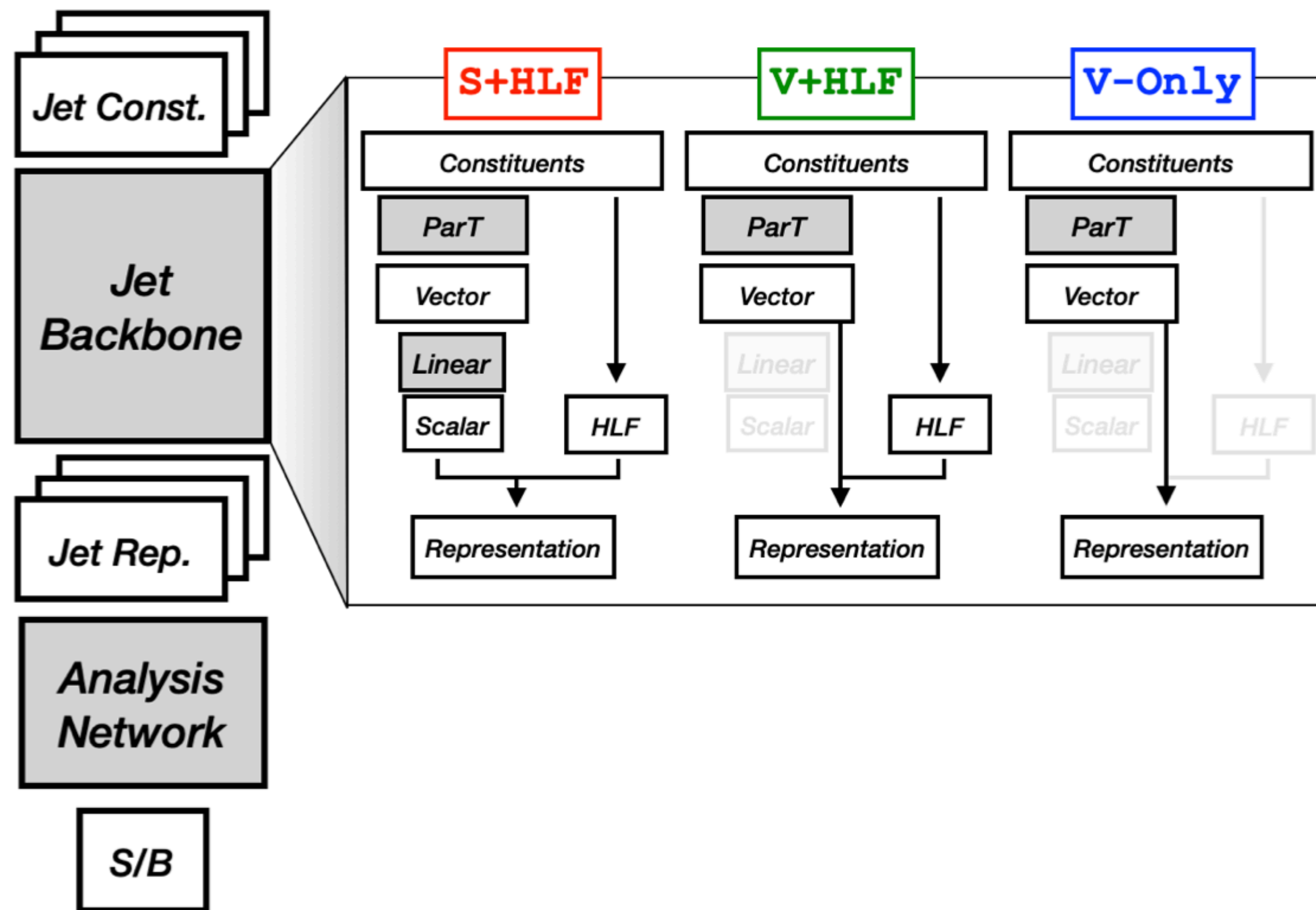
Some more detail here: [\[Slides\]](#)

# Key Property: Finetuning!



# It works!

*Finetuning & other workflows from Foundation Model research translate to particle physics and can lead to 100x more data-efficient models*



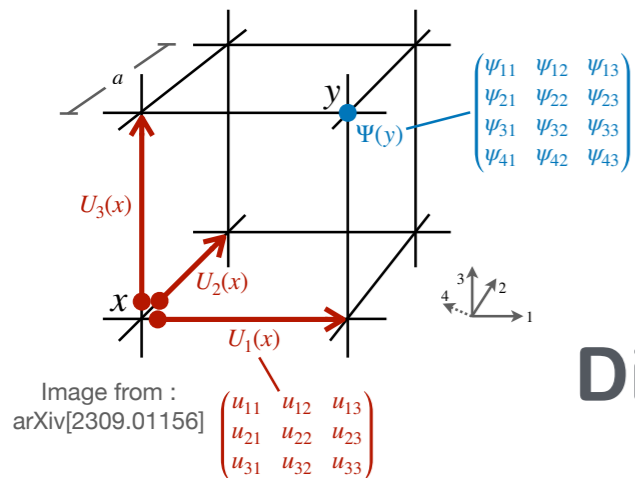
# ML Unfolding for error reduction in Lattice QCD observables

EuCAIFCon 2024

1<sup>st</sup> May 2024

Poster # 77  
Thu : 12-15

Simran Singh - Postdoc @ Bielefeld LGT group



**Lattice QCD** - currently our best probe for understanding **low energy QCD**

Discretise space-time and move to Euclidean space



Sample Gauge configurations from a probability distribution



$$Z = \int \mathcal{D}U \det M_f^n e^{-S_G}$$

Compute observables on the generated configurations  $\mathcal{O}(\sim (1 - 10)K)$

$$\langle \mathcal{O} \rangle \sim \frac{\partial \ln Z}{\partial K} \sim \text{Tr} \left( M_f^{-1} \frac{\partial M_f}{\partial K} \right)$$

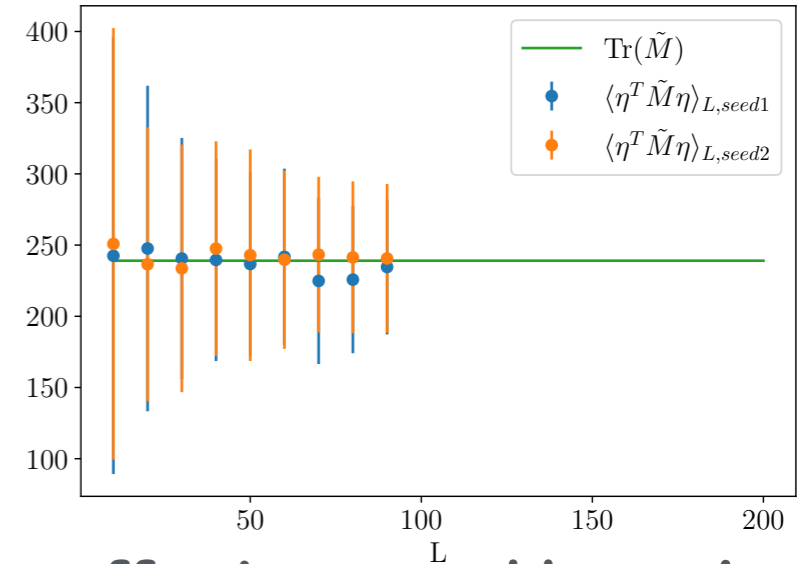
Typical size of Fermion matrices :  $N_\sigma^3 \times N_\tau \times N_c \times 4$ , can go up to  $\sim O(10^7 - 10^9)$

# Random Noise Method

No access to individual matrix elements - only matrix vector products !

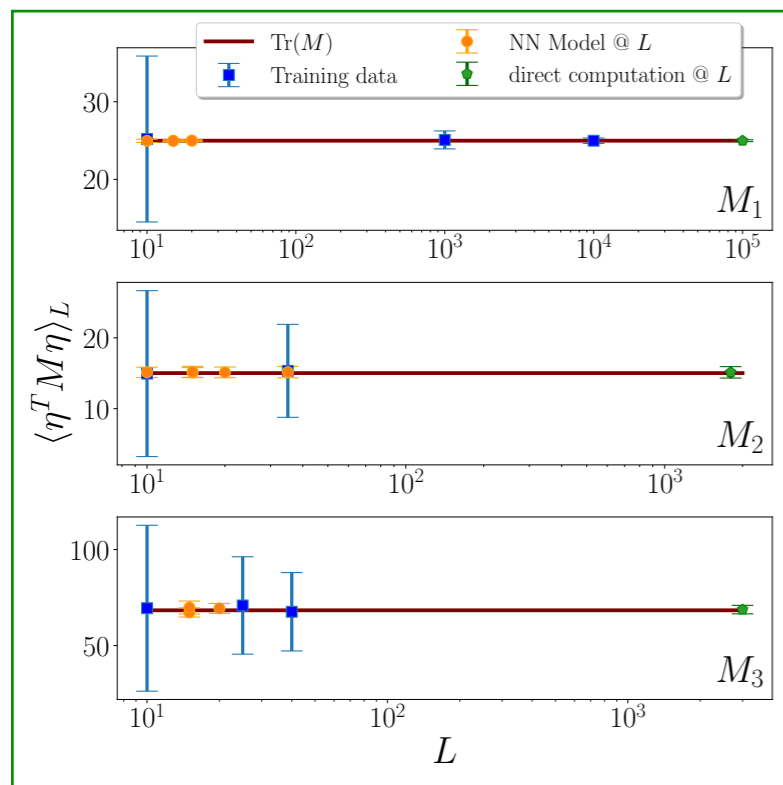
Choose  $L$  randomly drawn vectors  $\eta$  satisfying certain conditions to get

$$\langle \eta^T M \eta \rangle_L \simeq \text{Tr} M + \mathcal{O} \left( \frac{f(M)}{\sqrt{L}} \right)$$



## Data Unfolding via NN

**Question 1** : Can we train a **NN** to learn the systematic effect caused by using only finite such random vectors - given for some observable the **true distribution** with very large  $L$  and **measured distribution** with very small  $L$  ?



**Question 2** : How does this generalise to different Matrices?

Poster # 77  
on Thu !

**ML Unfolding for error reduction in Lattice QCD observables**  
 Simran Singh, Universität Bielefeld, Germany

UNIVERSITÄT BIELEFELD  
 Faculty of Physics

**Computing observables in lattice QCD**

- In lattice QCD, fermions integrated out to give  $Z = \int \mathcal{D}U \det M_f^n e^{-S_0}$
- Observables  $\leftarrow$  derivatives of  $\ln Z$ , e.g. quark number density:  $\frac{\partial \ln \det M_f}{\partial \mu_f} = \text{Tr} \left( M_f^{-1} \frac{\partial M_f}{\partial \mu_f} \right)$
- Typical size of Fermion matrices:  $N_f^2 \times N_s \times N_c \times 4$ , can go up to  $\sim O(10^7 - 10^9)$
- No direct access to matrix elements, only vector products.

**Random Noise method**

- Current state of art method to compute these traces based on drawing random vectors from a distribution satisfying  $\langle \eta_i \rangle = \lim_{L \rightarrow \infty} \frac{1}{L} \sum_{m=1}^L \eta_i^m = 0$  &  $\langle \eta_i \eta_j \rangle = \lim_{L \rightarrow \infty} \frac{1}{L} \sum_{m=1}^L \eta_i^m \eta_j^m = \delta_{ij}$
- Trace of  $M$  is then given by:  $\langle \eta^T M \eta \rangle_L \simeq \text{Tr} M + \mathcal{O} \left( \frac{f(M)}{\sqrt{L}} \right)$
- Only true in the limit of sampling infinite such randomly drawn vectors.

**ML based Unfolding: Try to un-learn the effect of finite # of random sources**

- Goal: To reconstruct the underlying "true" distribution from observations, which are smeared by "limited experimental" resolution.
- Equivalent to solving the inverse problem:  $F_{\text{observed}}(x) = \int K(x,y) * F_{\text{true}}(y) dy$ , with the goal to learn  $K^{-1}(x,y)$ , given  $F_{\text{true}}$  &  $F_{\text{observed}}$
- Can one adapt this to reduce error on the estimate for  $\text{Tr} M$  by asking whether we can train a sequential NN on two distributions: one with very small number of sources  $F_{\text{observed}}$  and the other with very large number of sources  $F_{\text{true}}$ ?
- Could we then perform measurements on different matrices with small number of sources and apply this transformation to estimate their true trace?

**Results of experiments**

- Tests performed on matrices with different structures - sparse and dense, different sizes - 100, 1000, 10000, Poisson and normally distributed elements.
- Re-training has to be done but needs fewer resources!

Re-training on  $M_1$  ( $L = 10$  &  $L = 10^3$ )  
 First training on  $M_1$  ( $L = 10$  &  $L = 10^3$ )  
 Re-training on  $M_2$  with  $L = 10$  &  $L = 35$   
 Re-training on  $M_3$  with  $L = 10$  &  $L = \{20, 25\}$

Next step: Apply to real QCD data ...