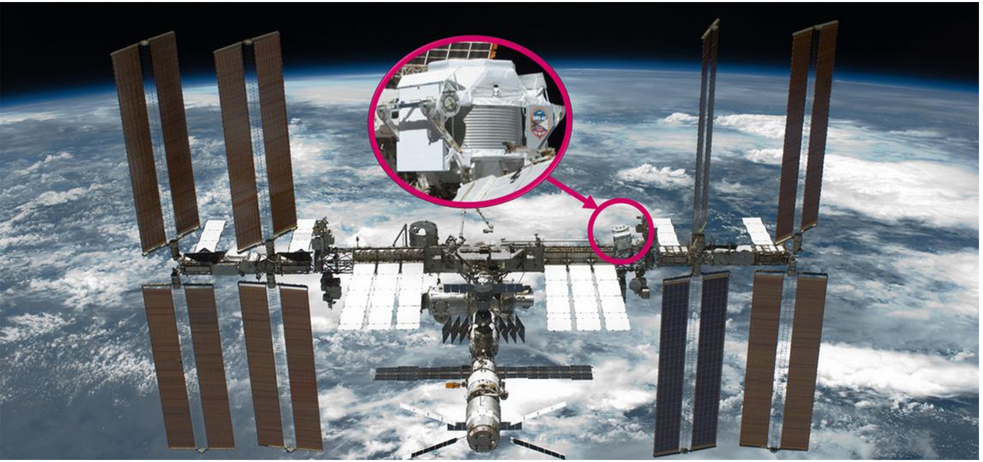
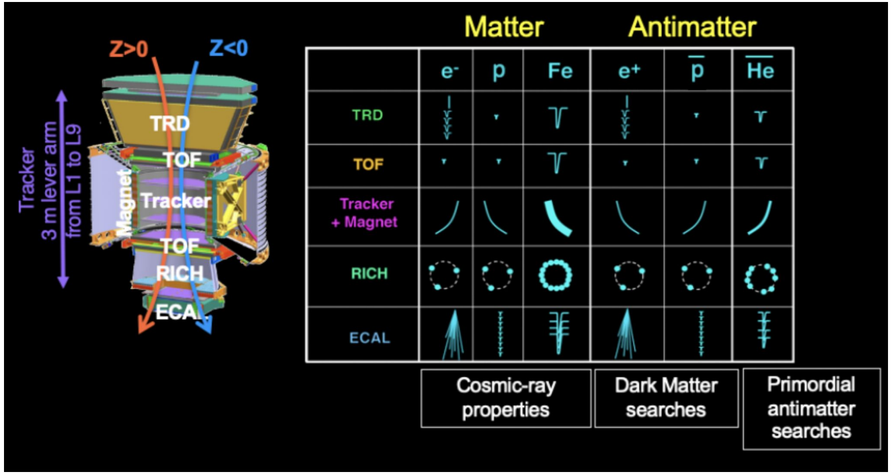
Machine-learning analysis of cosmic-ray nuclei data from the AMS-02 experiment on the International Space Station (ISS)

Shahid Khan



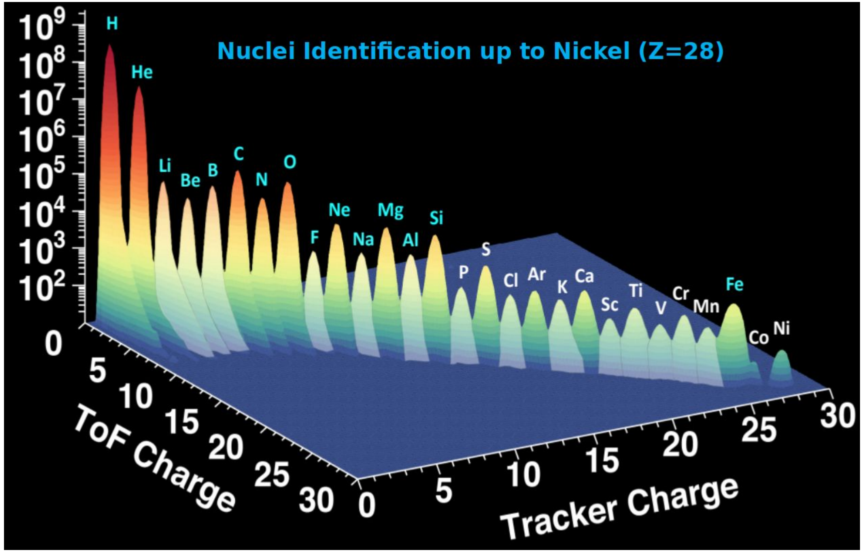
~7000 kg AMS-02 on the ISS



AMS-02 detector components and their use cases for the identification of various charges

Data and MC Mixture

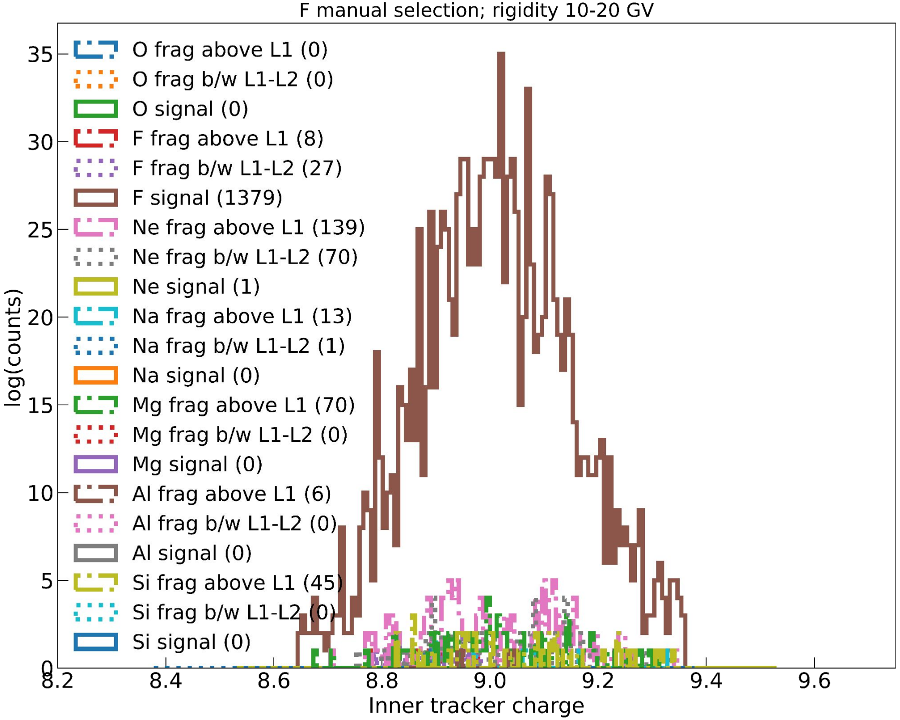
- Around 100 variables from tracker, TOF and TRD
- MC simulations and AMS-02 ISS experimental data with tight manual selections applied have been mixed to create train (equal abundance of all nuclei) and test (natural abundance of nuclei) data sets
- For each nuclei species 3 different classes are created
  - Fragmented above L1
  - Fragmented between L1 and L2
  - Non-fragmented and those which fragment between L2 and L8
- Benchmarking with ML algorithms: Multi-layered perception, Convolutional Neural Networks, Transformers, and XGBoost
- XGBoost showed better performance because of the tabular nature of the data. source:



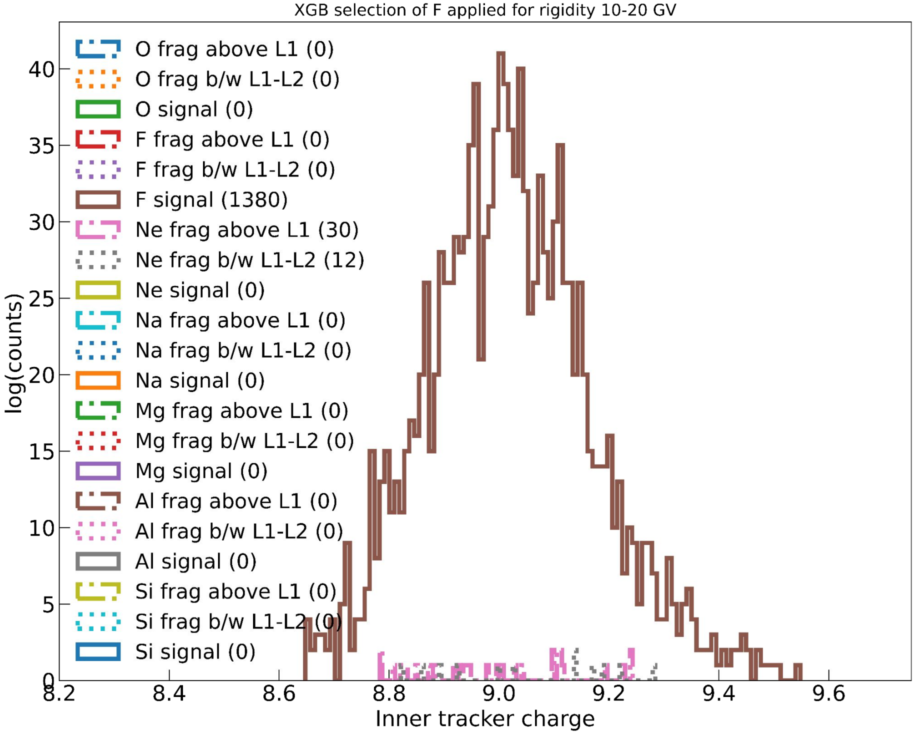
Nuclei identification with tracker, TOF

Comparison of fragments in the selections

- The manually selected Fluorine candidates have contamination from heavier nuclei
- For ML selected Fluorine the contamination mostly comes from the neighboring Ne



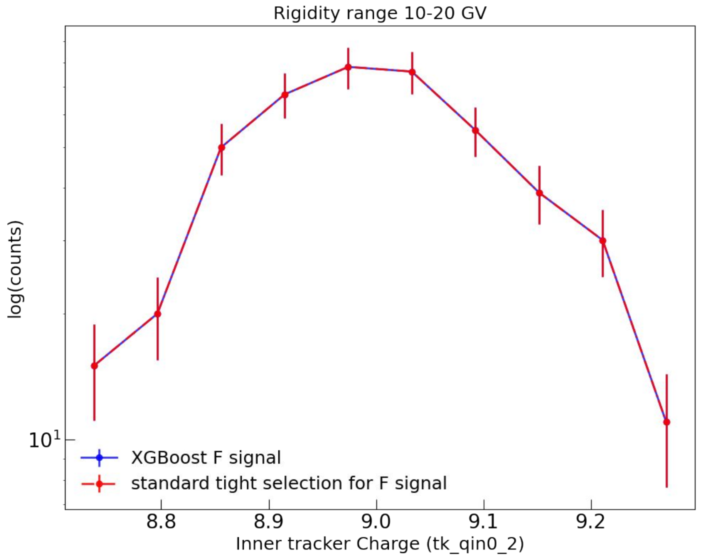
Manually optimized selection



XGBoost optimized selection

Model deployment on AMS-02 ISS data

- Test data, applied tight standard selection on AMS-02 experimental data from ISS, is taken here
- ML selections are applied and the performance of hypercube selection and ML selection is compared here



Summary

- ML performance was tested for nuclei selection of the AMS-02 simulated data and ISS data with tight selections applied
- Comparison between standard AMS-02 standard selection and ML model for Fluorine
  - At the same efficiency of standard selection, higher purity was achieved with ML model
  - ML based selection can be varied easily and is non-linear

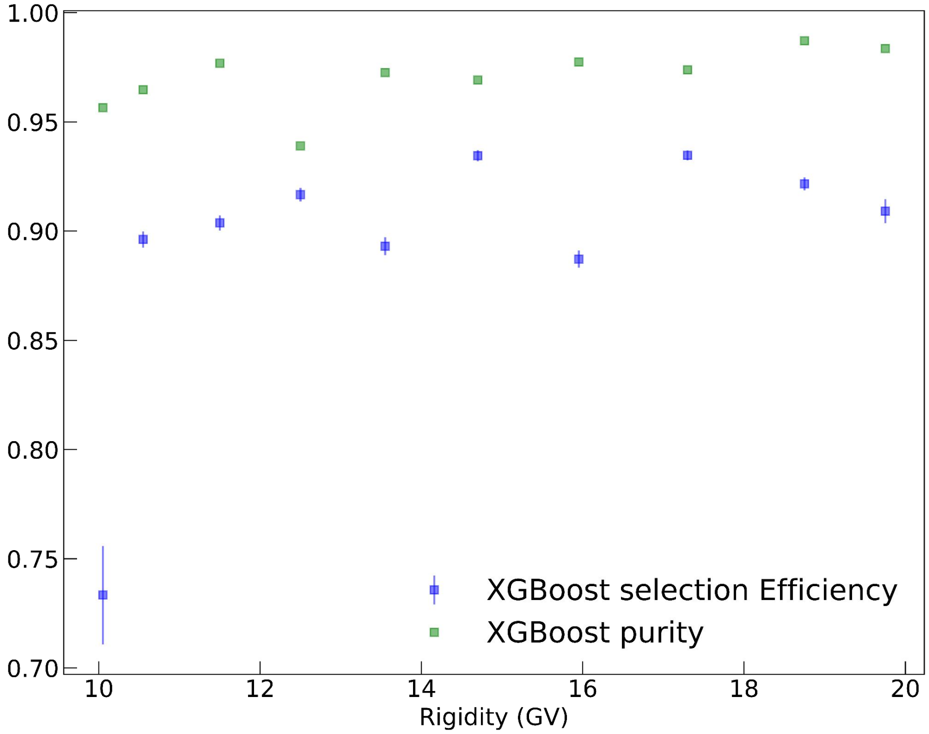
Outlook

- Efficiency correction for selection
- Systematic evaluation

Poster on Thursday

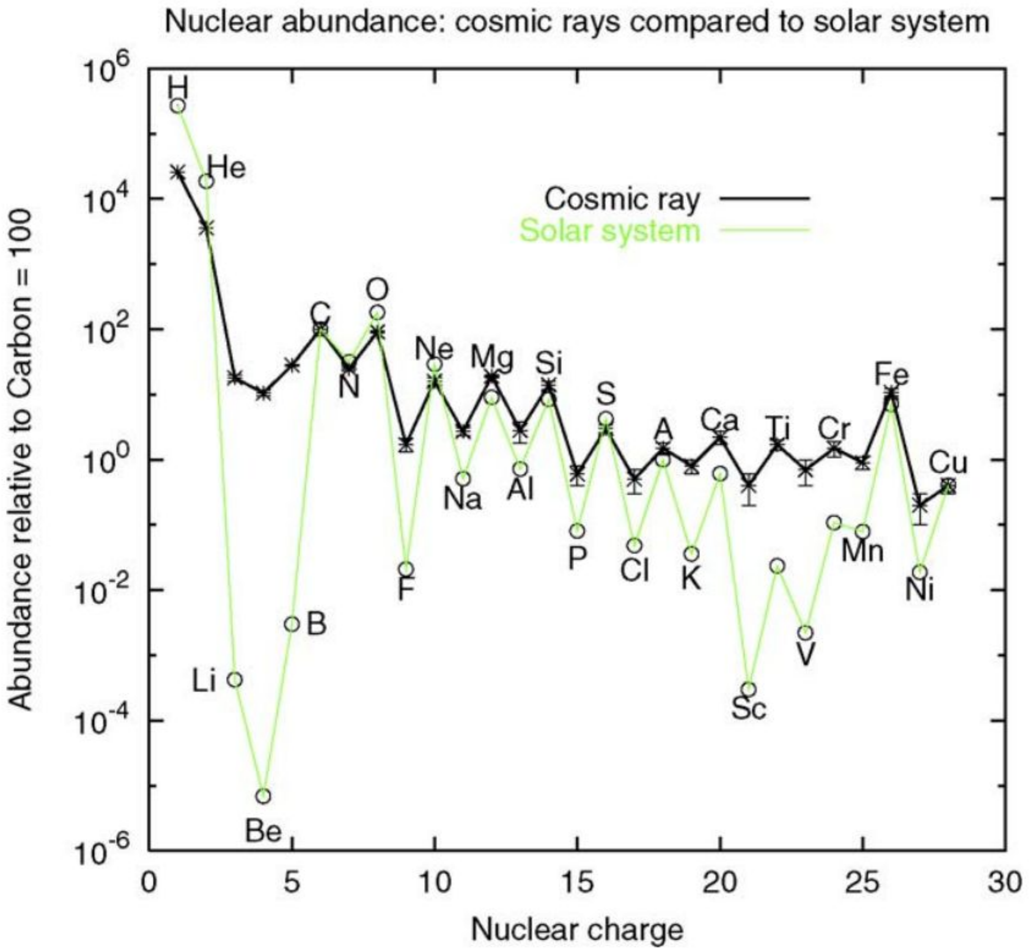
Back up

Bin wise visualization



$$Efficiency = \varepsilon = \frac{signal\ after\ selection}{signal\ before\ selection}$$

$$purity = \frac{signal}{signal + background}$$



https://doi.org/https://doi.org/10.1016/j.nuclphysa.2005.01.024

UPPSALA UNIVERSITET

Generating Lagrangians for particle theories

Eliel Camargo-Molina
Yong Sheng Koay
Rikard Enberg
Stefano Moretti

What:

Can a transformer model generate a Lagrangian given a set of quantum fields?

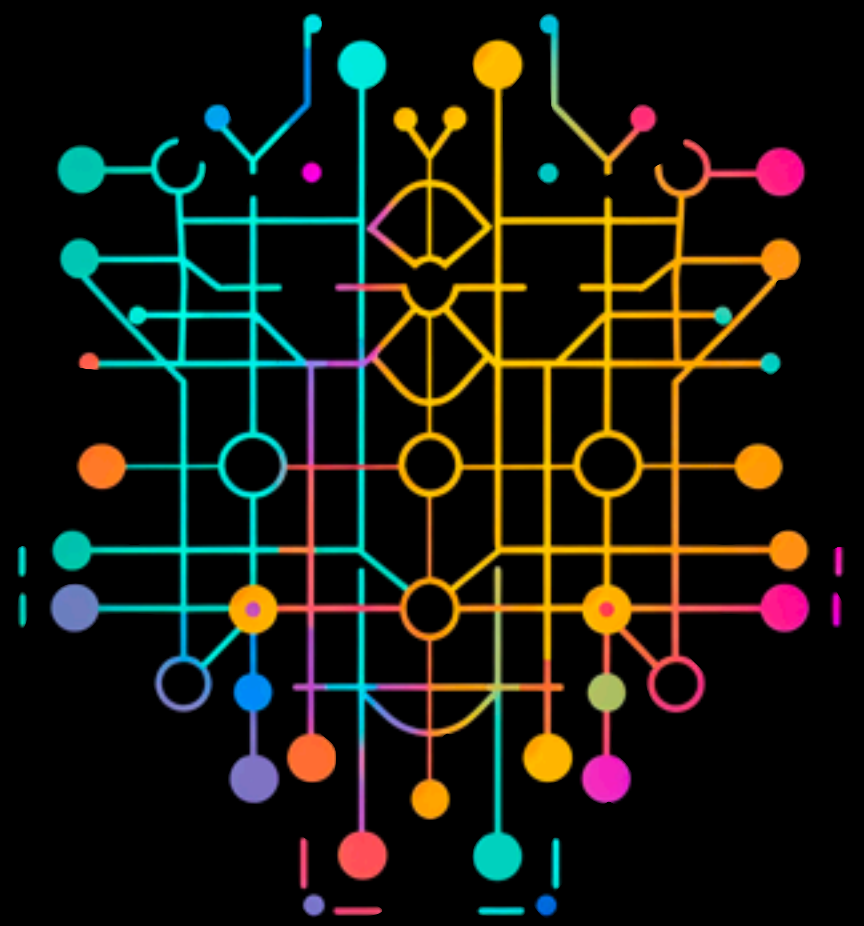
$\mathrm{SU}(3)\ \mathrm{SU}(2)\ \mathrm{U}(1)$

$\phi_{(3,2,0)}$

$\varphi_{(1,2,\frac{1}{2})}$

$\psi_{L\,(1,2,\frac{1}{3})}$

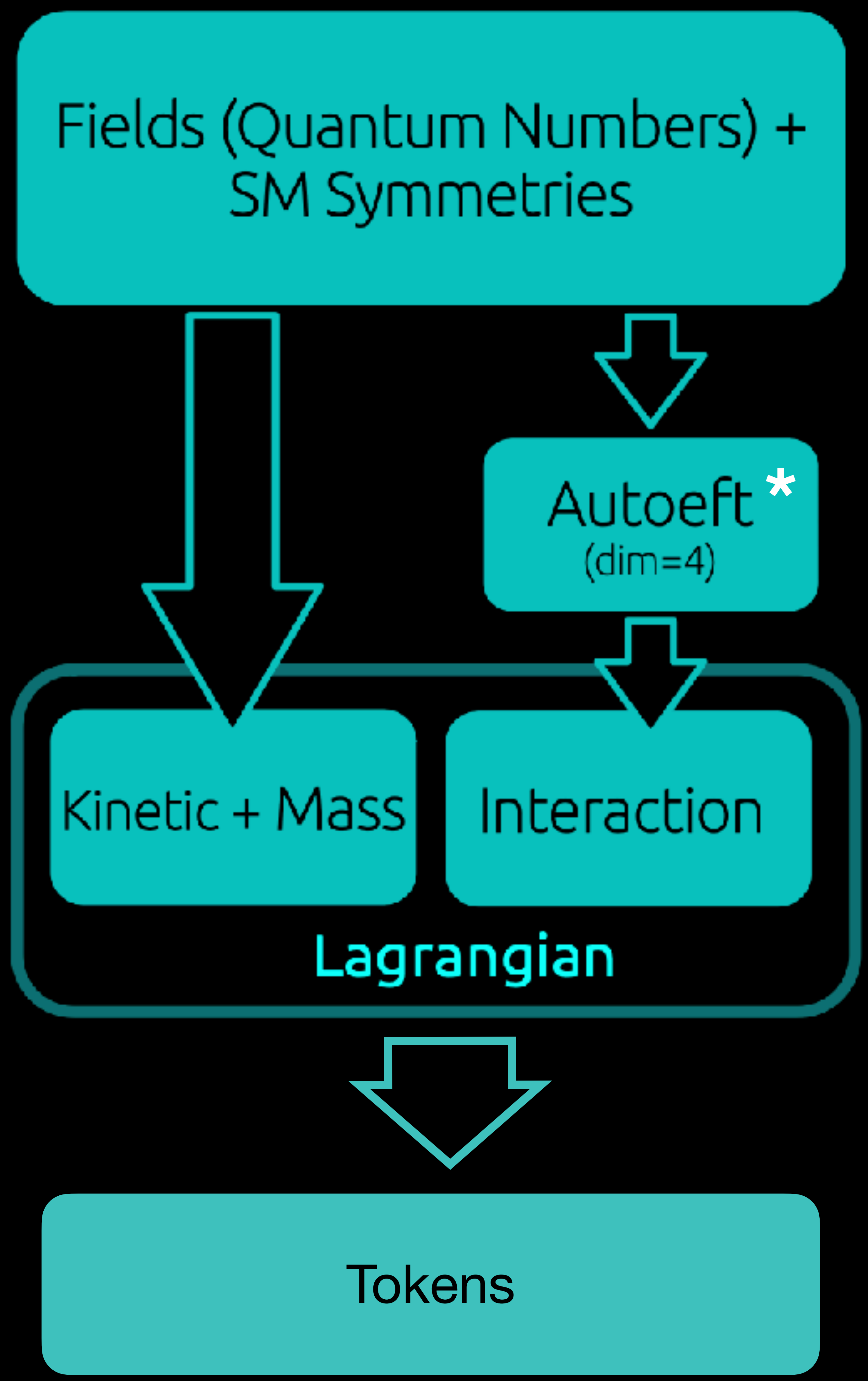
$\psi_{R\,(1,1,\frac{-1}{6})}$

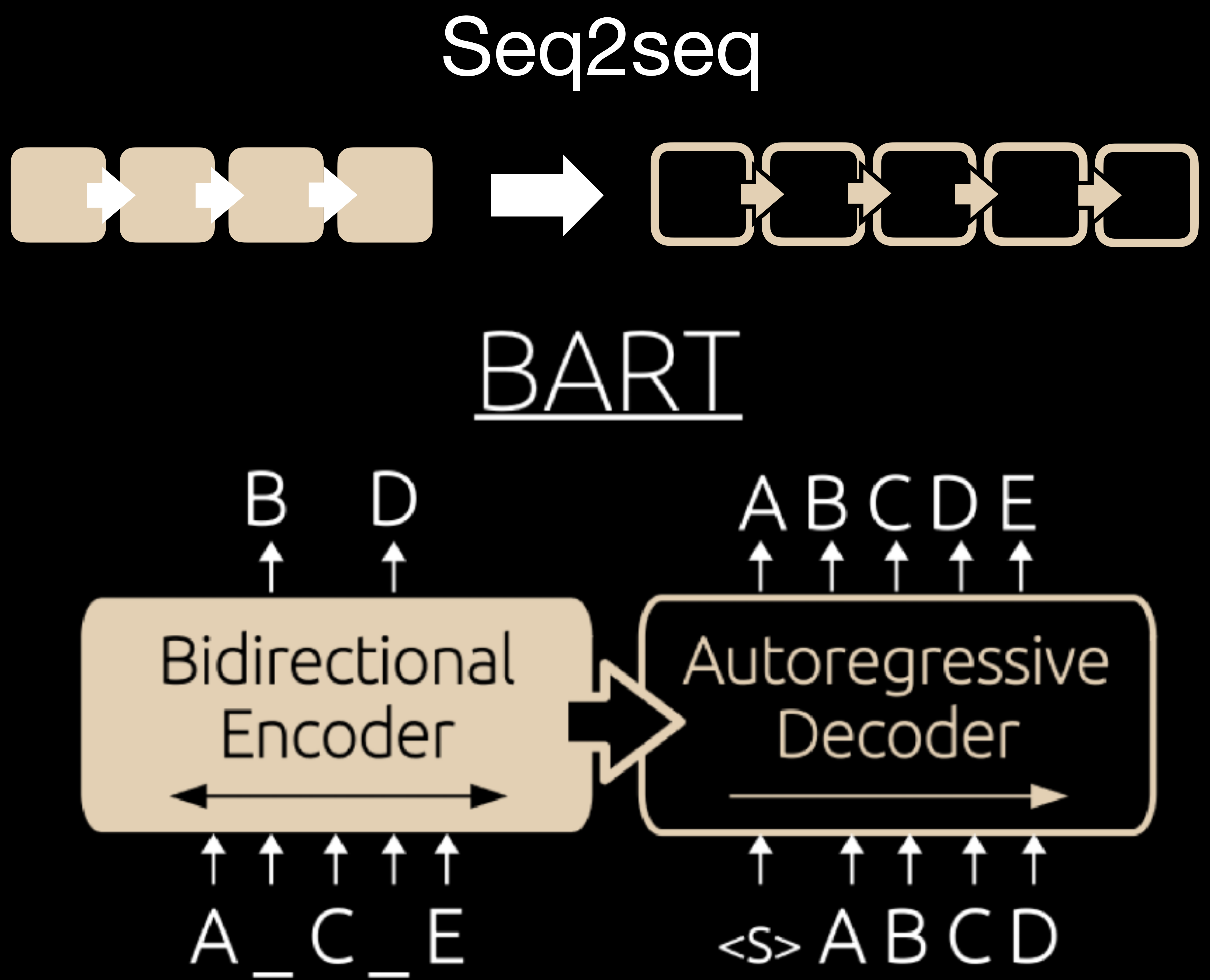


$$
\begin{aligned}
\mathcal{L} =& (D_\mu \phi)^\dagger (D^\mu \phi) + (D_\mu \varphi)^\dagger (D^\mu \varphi) \\
&+ i\bar{\psi}_L \gamma^\mu D_\mu \psi_L + i\bar{\psi}_R \gamma^\mu D_\mu \psi_R \\
&- \lambda_\phi (\phi^\dagger \phi)^2 - \lambda_\varphi (\varphi^\dagger \varphi)^2 \\
&- \lambda_1 \phi^\dagger \phi\, \varphi^\dagger \varphi - \lambda_2 \phi^\dagger \phi\, \varphi^\dagger \varphi \\
&- y \varphi \bar{\psi}_L \psi_R + h.c. + \ldots
\end{aligned}
$$

SU(2) SU(2)

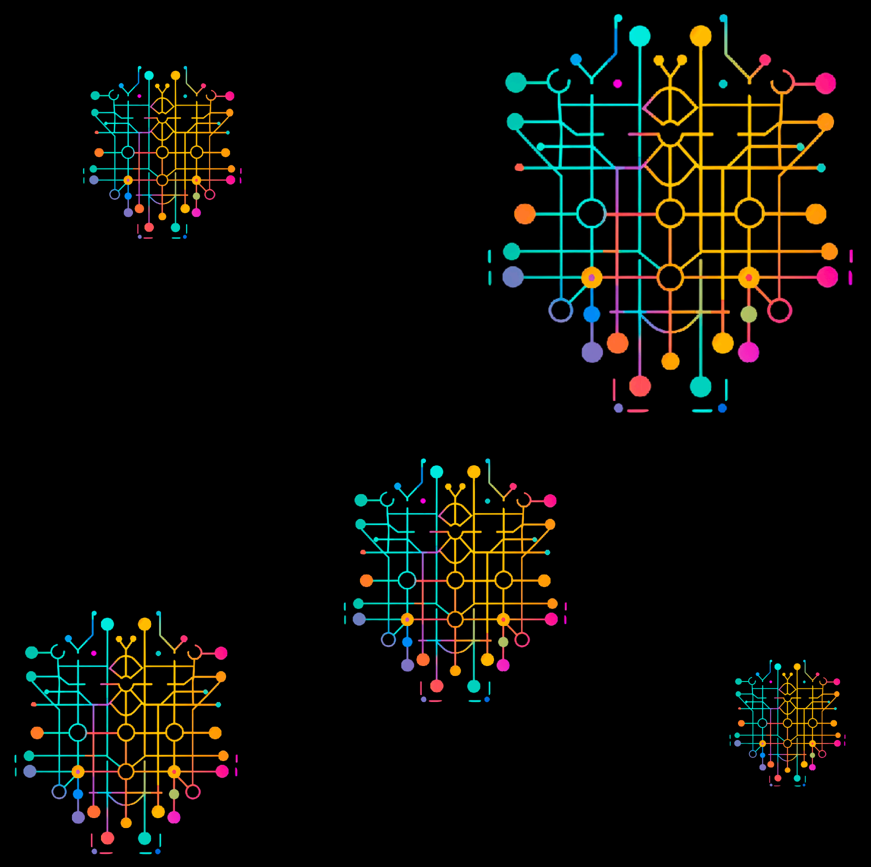
How: Train a BART model from scratch for a seq2seq task with carefully constructed dataset and a good tokenization





* 2309.15783 [Harlander, Schaaf]

How: Train a BART model from scratch for a seq-to-seq task with carefully constructed dataset and a good tokenization



Distribution of quantum numbers matters!

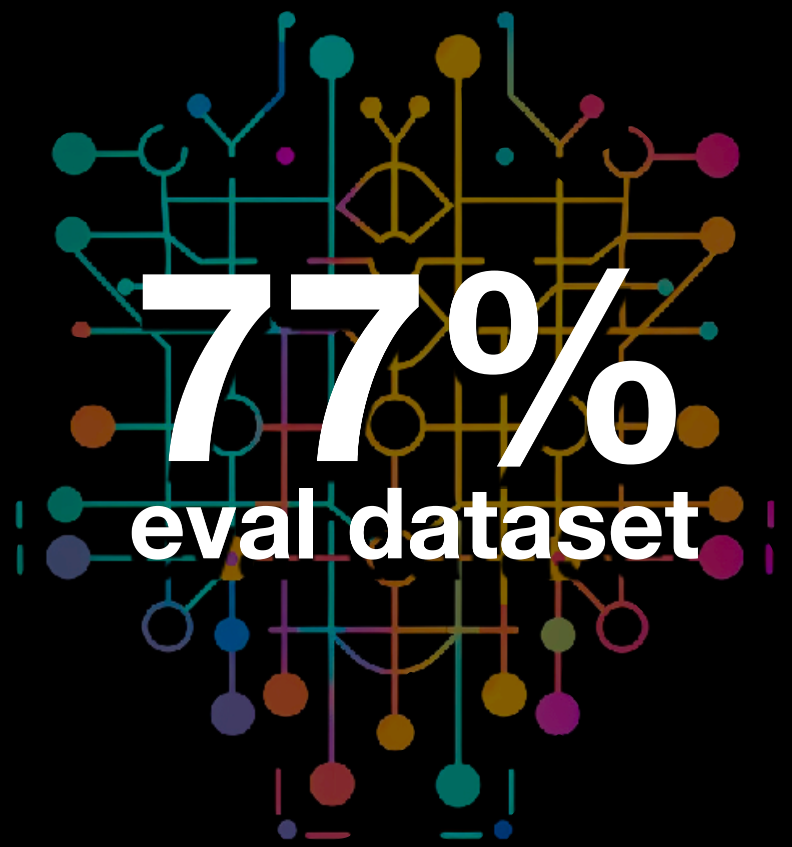
Some tokenizations are bad (e.g. too long, too loosy …)

17K Lagrangians

~110M pars

37K Lagrangians

~350M pars

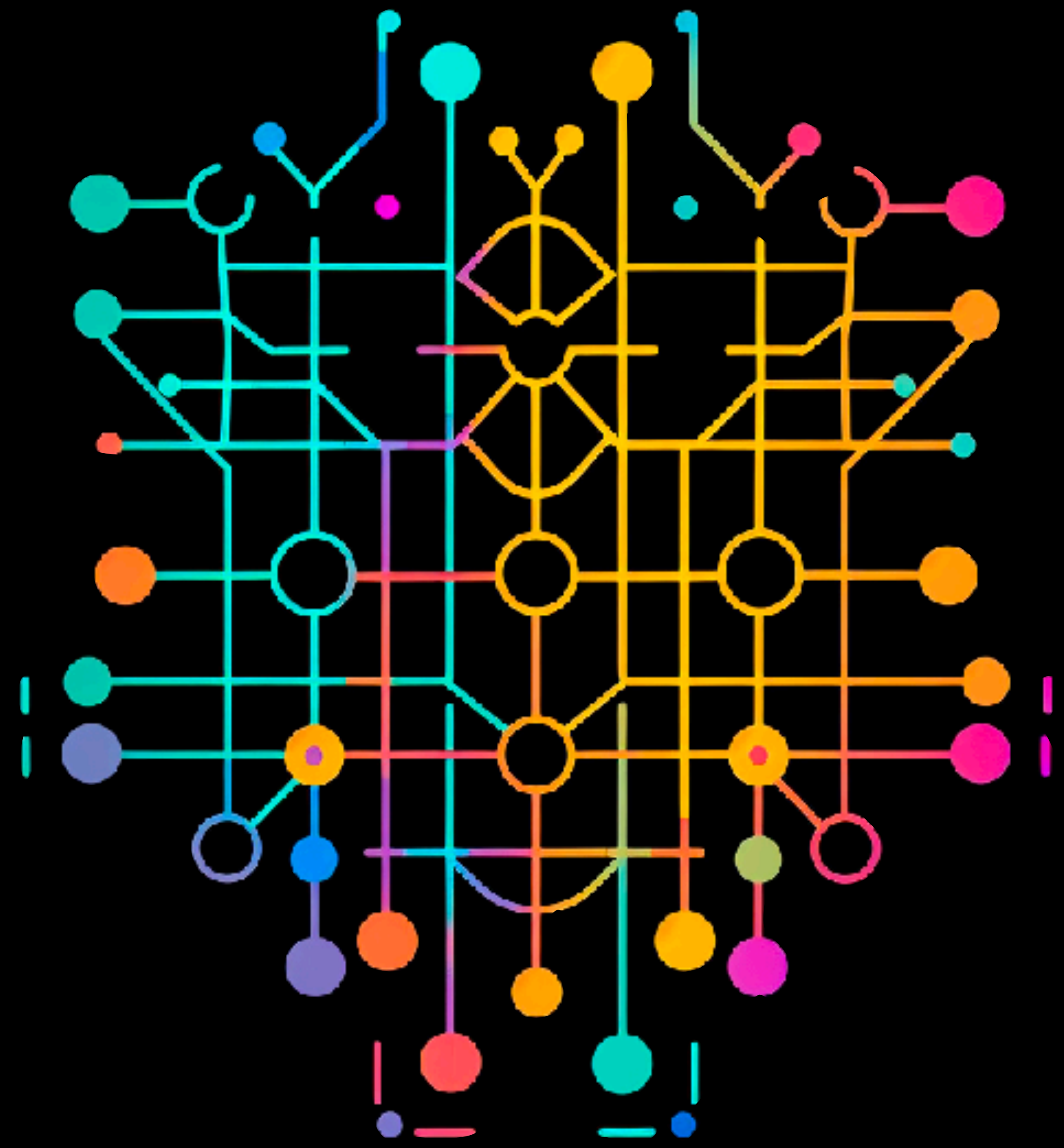


Mostly scalars and some fermions

Errors because of tokenization and length

NEXT STEP

~1B pars

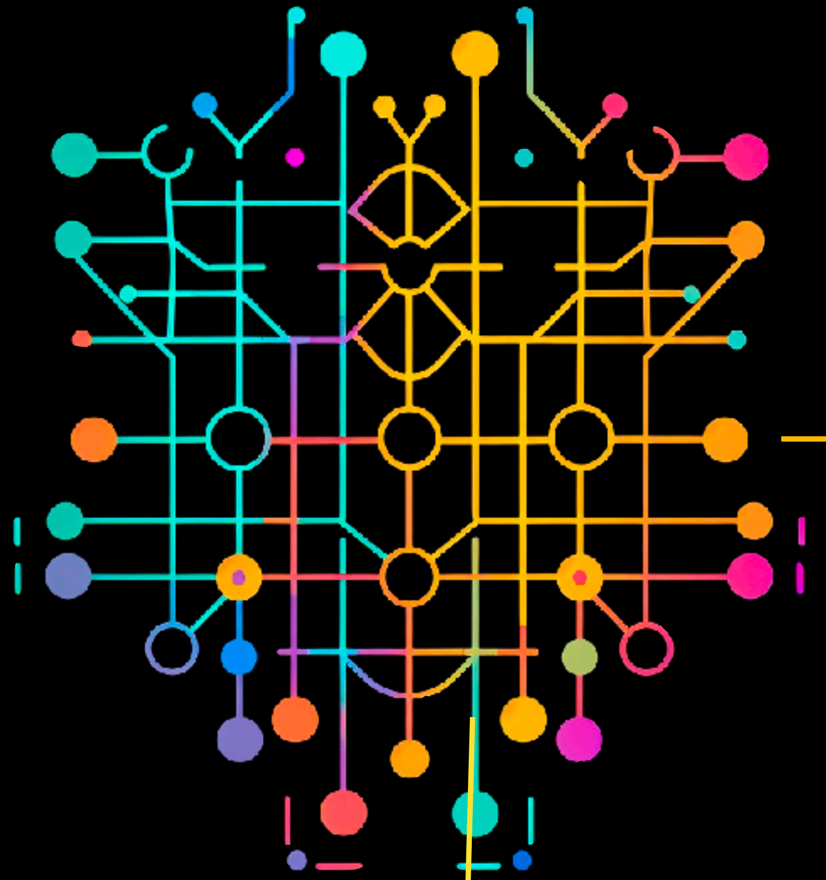


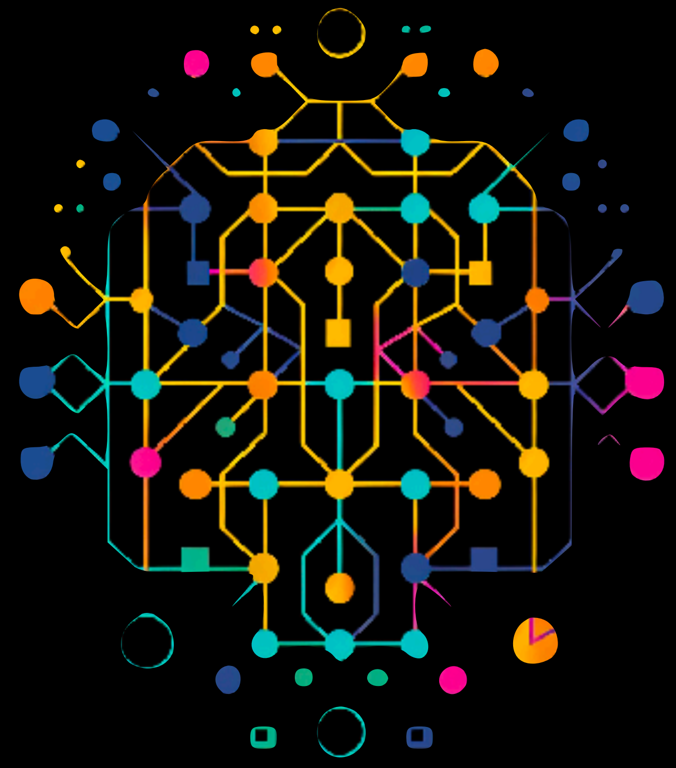
Larger datasets

Three strategies for distribution over quantum numbers

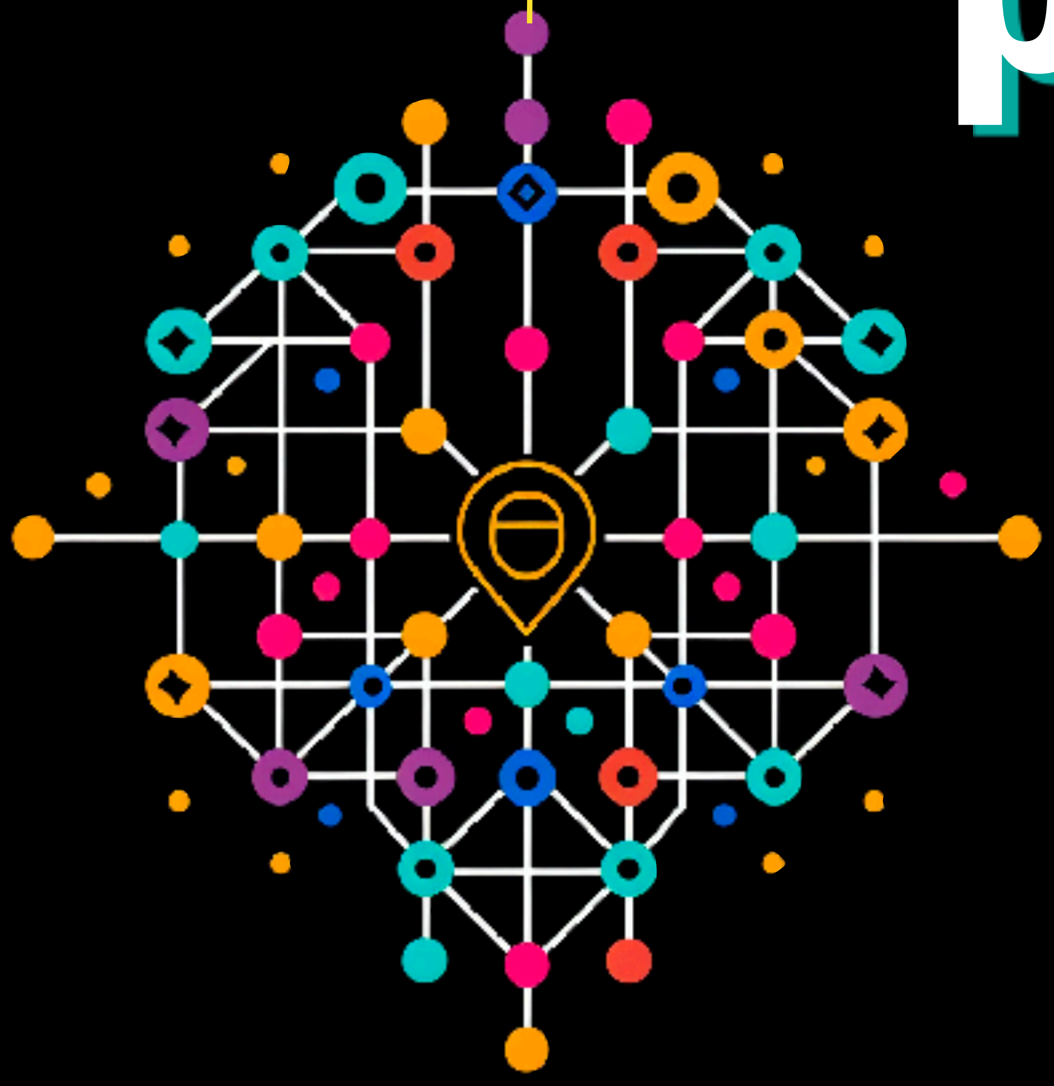
More fields

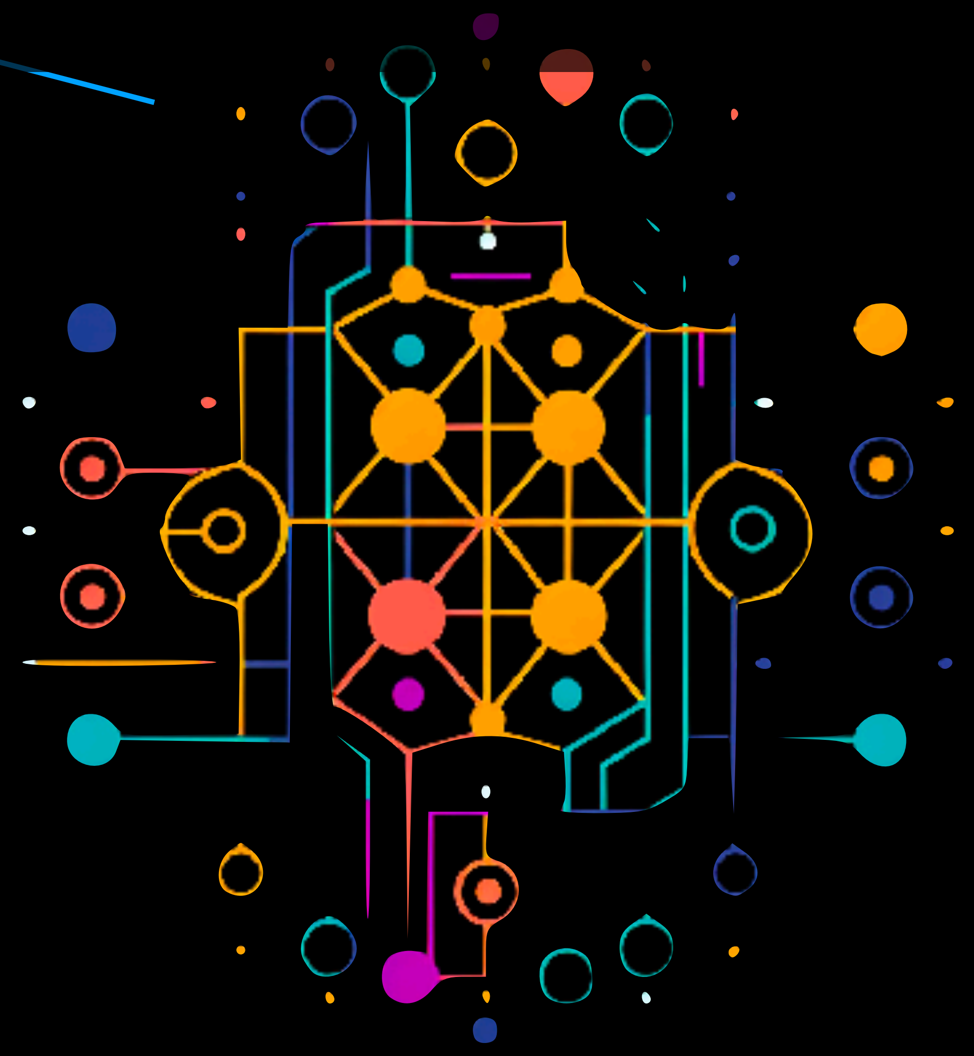
Why:

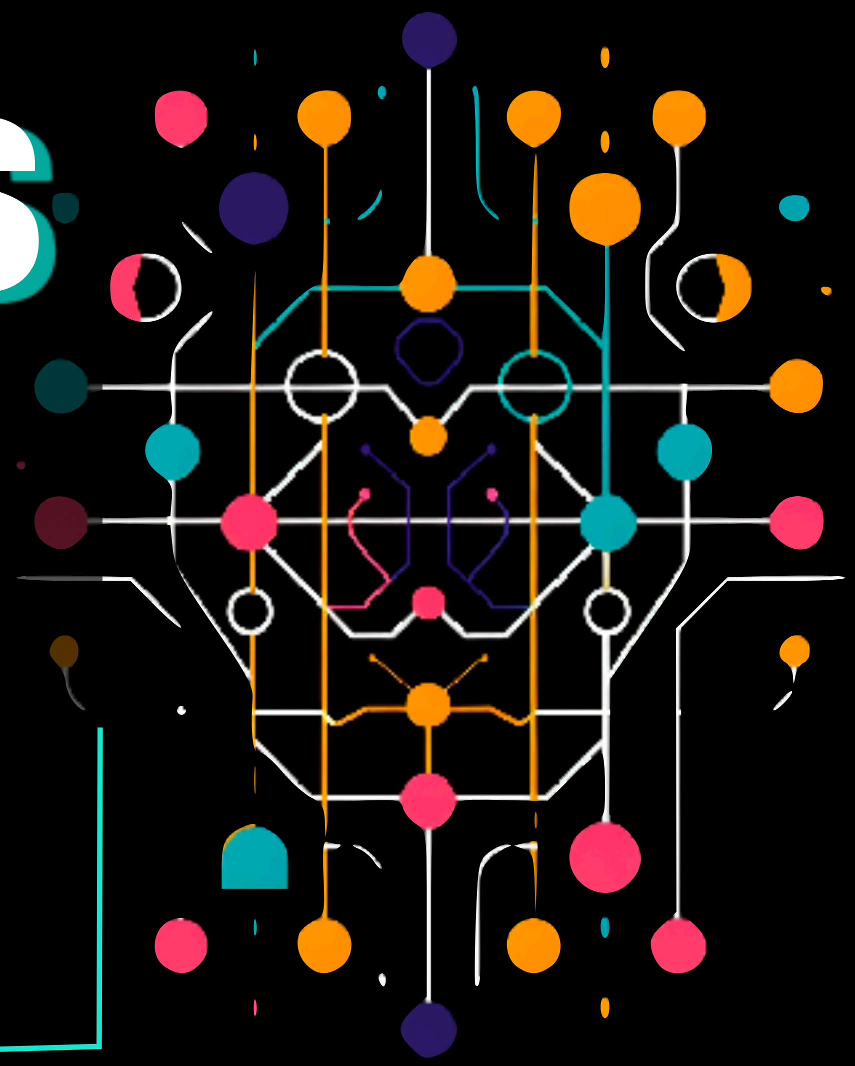




Foundational Model for cosmo+particle physics phenomenology







LORENTZ-EQUIVARIANT GEOMETRIC ALGEBRA TRANSFORMERS FOR HIGH-ENERGY PHYSICS

Víctor Bresó Pla

In collaboration with Jonas Spinner, Johann Brehmer, Pim de Haan, Tilman Plehn & Jesse Thaler

UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

EuCAIFCon 2024

MACHINE LEARNING APPROACHES FOR HIGH ENERGY PHYSICS

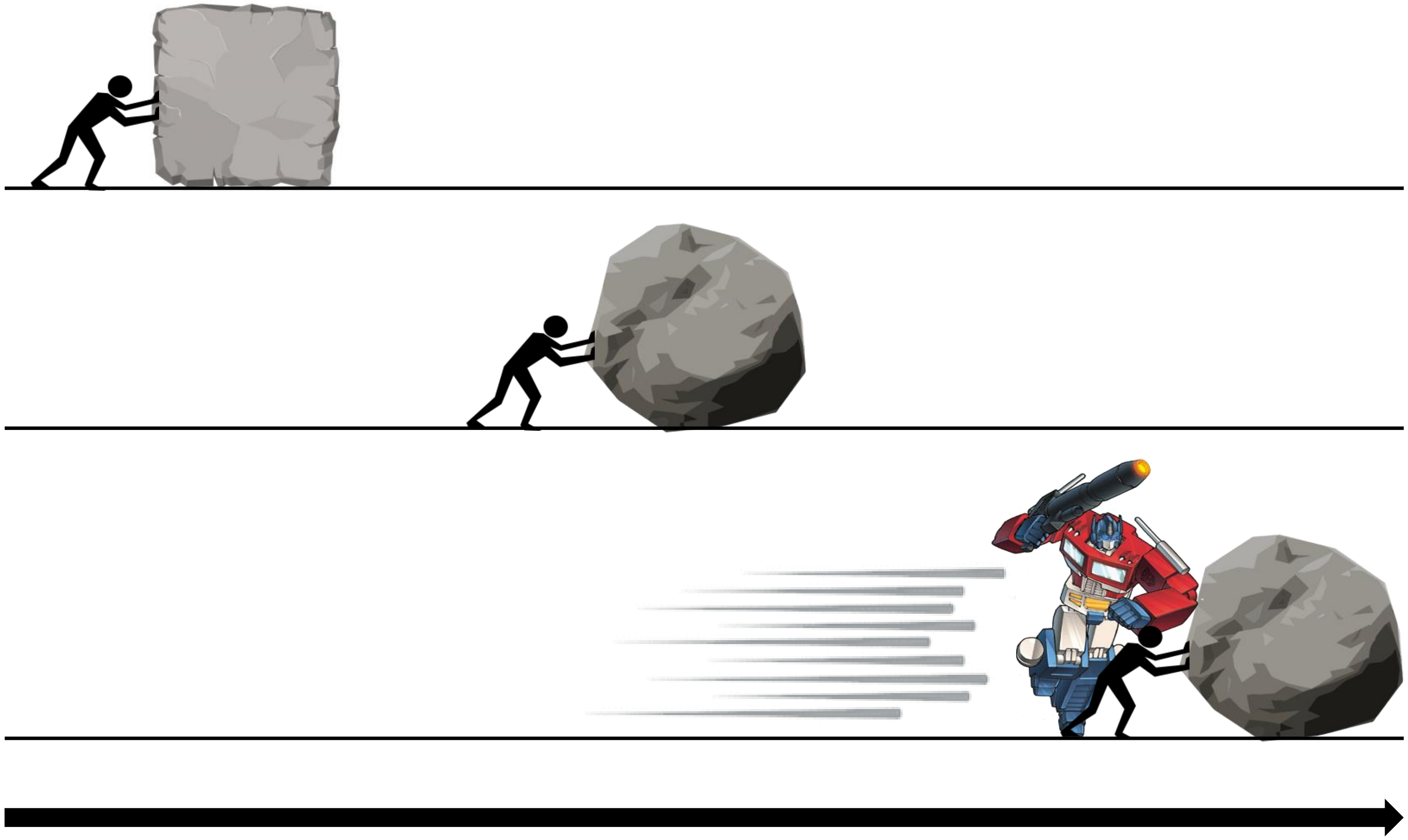
Out of the box
model

+

Symmetry
awareness

+

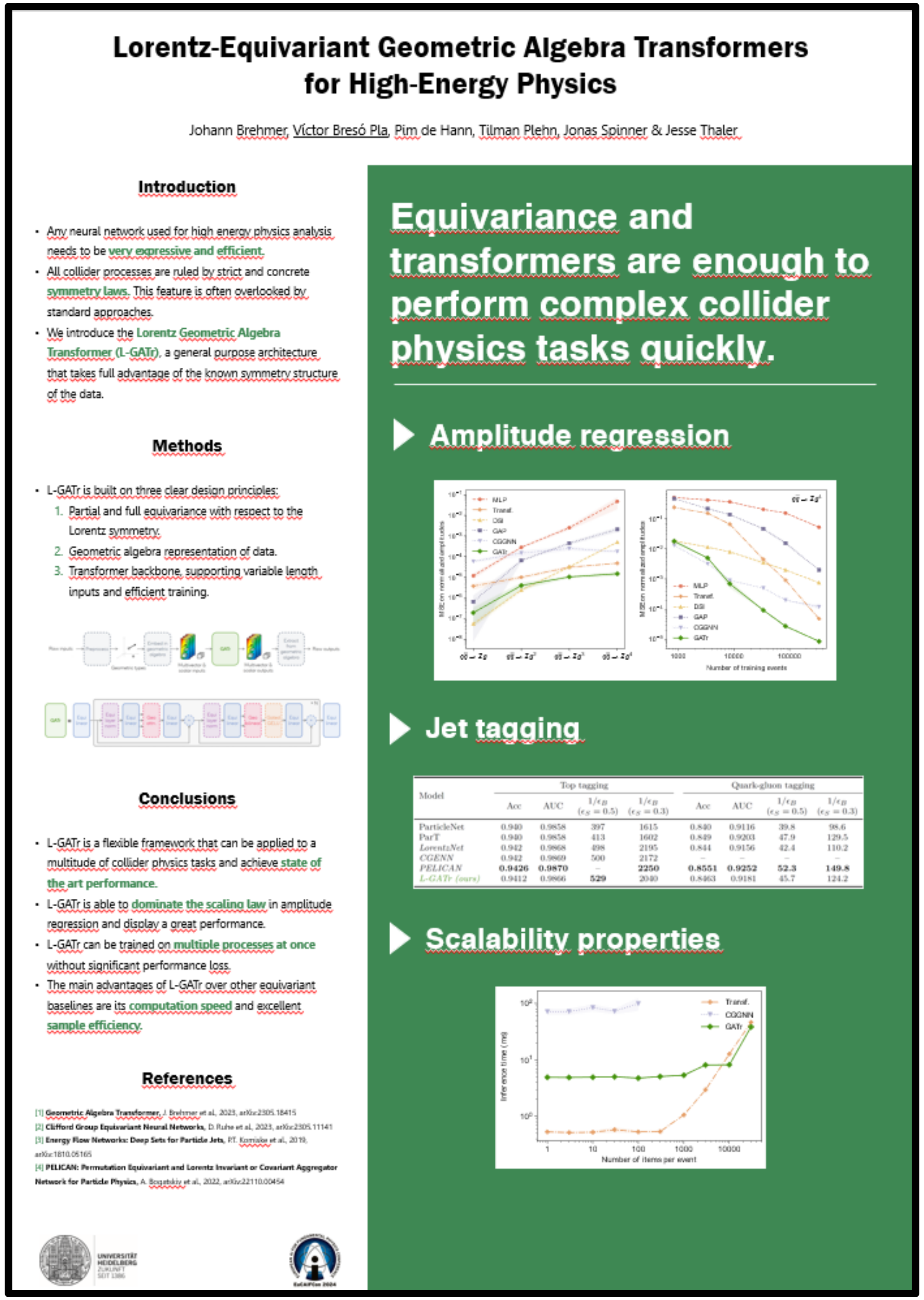
Transformer
backbone



Network output quality and efficiency

- Introducing the **Lorentz Geometric Algebra Transformer (L-GATr)**

- Main results:
  1. We achieve **state of the art performance** for multiple collider physics tasks
  2. L-GATr can learn the features of **multiple processes simulatenously**
  3. L-GATr is **faster** and more **memory efficient** than other equivariant baselines



HOW CAN WE TURN CLASSIFIERS INTO ANOMALY DETECTORS?

Sascha Caron (Nikhef, Radboud U), José Enrique García Navarro (IFIC, CSIC-UV), María Moreno Llácer (IFIC, CSIC-UV), Polina Moskvitinaa (Nikhef, Radboud U), Mats Rovers (Radboud U., Nikhef), **Adrián Rubio Jiménez (IFIC, CSIC-UV)**, Roberto Ruiz de Austri (IFIC, CSIC-UV), Zhongyi Zhang (Bonn U.).

MOTIVATION AND STRATEGY

Motivation

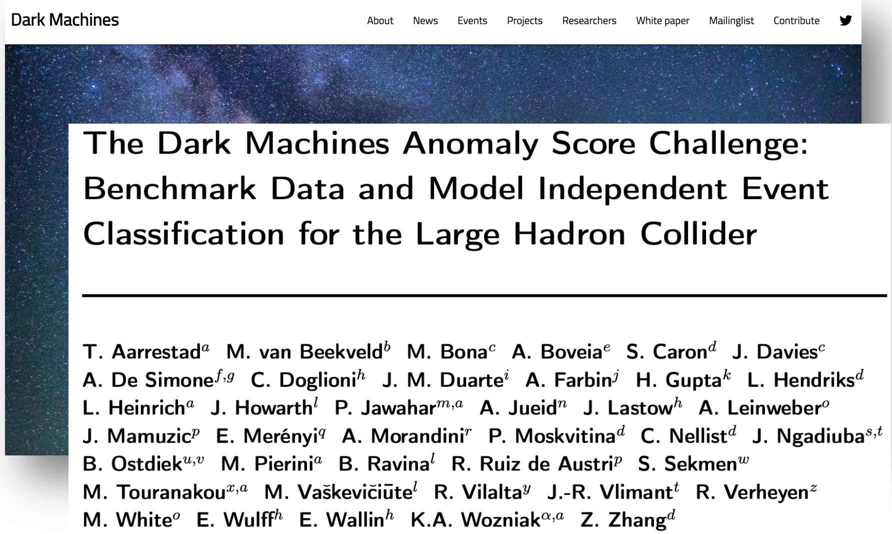
- The most powerful architectures for supervised classification learn the physical information more efficiently.
- But… **how can we turn them into anomaly detectors and how good are they?**

Strategy

- Adaptation of 2-3 different classifier architectures with 3 methods to detect anomalies (8 models).
- No network optimisation (or minimal) was performed to avoid biases.

DarkMachines dataset

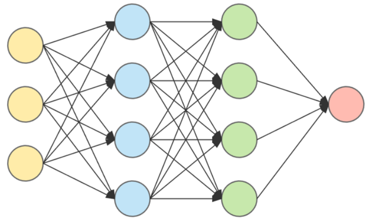
- Open data: Zenodo link to dataset from anomaly score challenge.
- Event generation: *proton-proton* collisions at 13 TeV .
- Detector simulation: simplified card for ATLAS detector at CERN.
- Reconstructed particles (objects): jets, b-tagged jets, charged leptons, photons.
- Low level variables: object type, the four-momentum of the objects and the missing transverse momentum of the event.



ARCHITECTURES AND TECHNIQUES

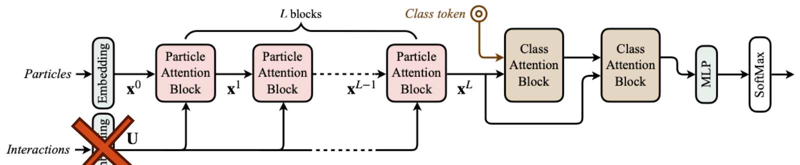
Architectures

Multi-Layer Perceptron (MLP)



Particle Transformer (ParT)

https://arxiv.org/abs/2211.05143



(a) Particle Transformer

No pairwise interactions

ParT+ SM couplings

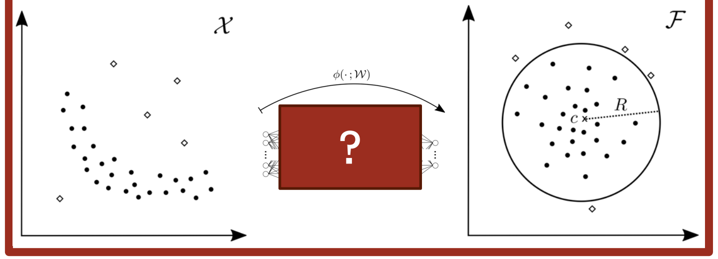
- Pairwise interactions
  - $\ln(m^2_{ij})$
  - $\ln(\Delta R_{ij})$
  - Physical information from Standard Model: couplings.

Developed by this group

Techniques

Deep Support Vector Data Description (dSVDD)

- Add an output layer with certain dimensions.
- Training: minimise distance to a centre in the hypersphere (anomaly score).
- Outliers are considered anomalies.
- Make ensemble for different dimensions.

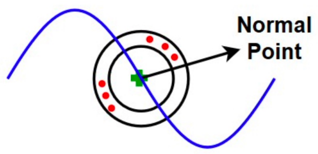


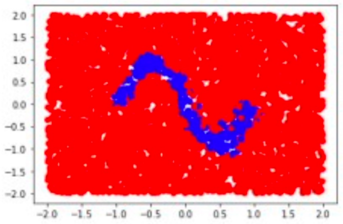
Deep Robust One-Class Classification (DROCC)

- Background is assumed to lie in a low-dimensional manifold.
- Anomalous background events are generated and their location in the manifold is searched with an adversarial training.

$$\sum_{i=1}^{n}\left[\ell(f_\theta(x_i),1)+\mu \max_{\substack{\tilde{x}_i\in\\ N_i(r)}} \ell(f_\theta(\tilde{x}_i),-1)\right]$$

- Weakly supervised implementation

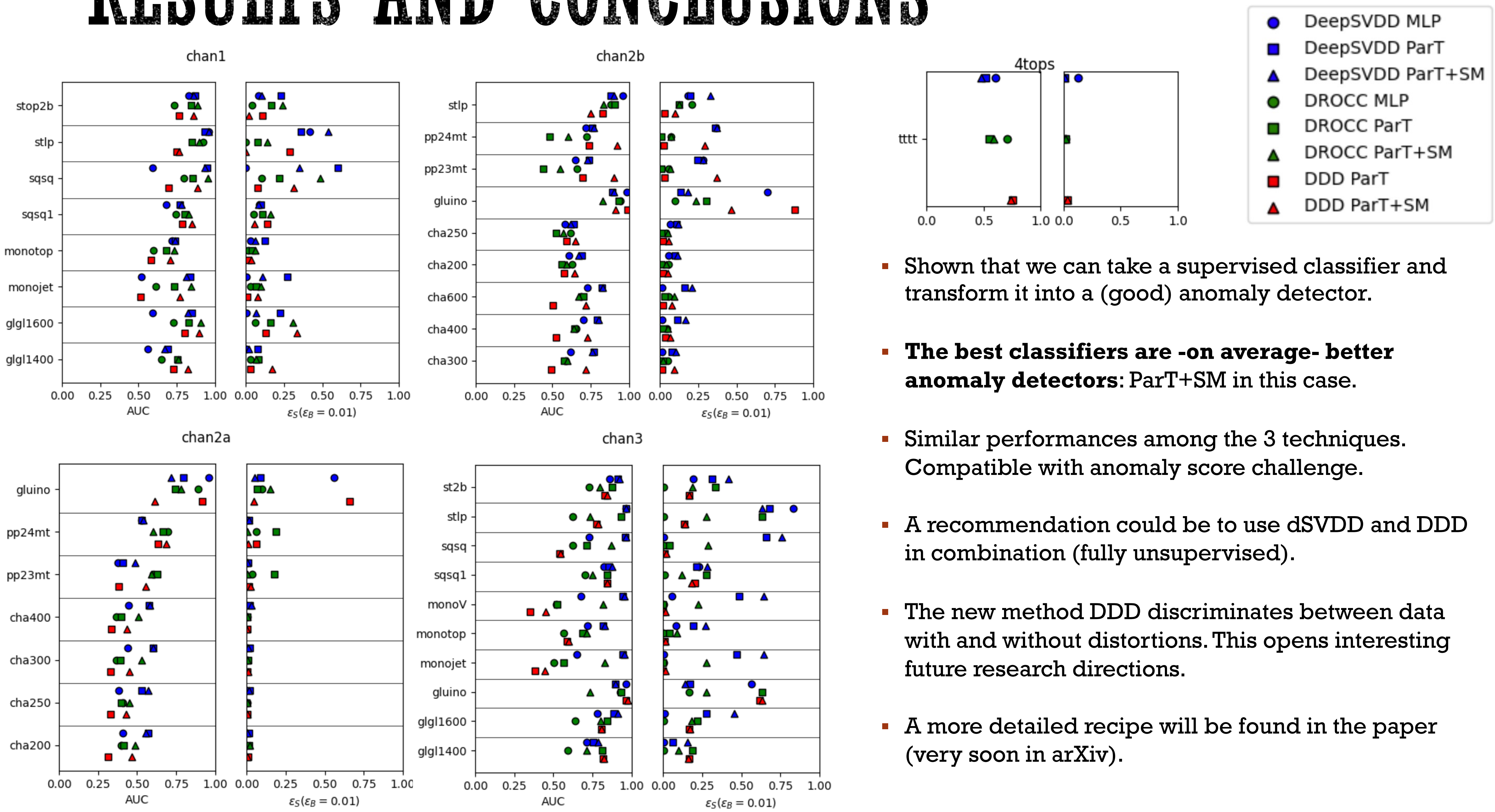




Discriminant distorsion detection (DDD)

- New technique developed for this study.
- Anomalies look like distorted background.
- Distorted training dataset is created:
  - Smearing kinematic variables with a gaussian.
  - Adding or removing objects.
- Train: discriminate *distorted bkg* vs *bkg.*
- Models with AUCs ~ 0.7-0.8 are picked up for testing on signals. Ensemble was made.

RESULTS AND CONCLUSIONS



- Shown that we can take a supervised classifier and transform it into a (good) anomaly detector.
- **The best classifiers are -on average- better anomaly detectors**: ParT+SM in this case.
- Similar performances among the 3 techniques. Compatible with anomaly score challenge.
- A recommendation could be to use dSVDD and DDD in combination (fully unsupervised).
- The new method DDD discriminates between data with and without distortions. This opens interesting future research directions.
- A more detailed recipe will be found in the paper (very soon in arXiv).

5

BACK-UP

CHANNELS AND SIGNALS

- Channel 1 (214k SM and 38k BSM):
  - $H_T \geq 600$ GeV .
  - $E_{Tmiss} \geq 200$ GeV.
  - $E_{Tmiss}/HT \geq 0.2$.
  - At least 4 (b)-jets with $p_T > 50$ GeV.
  - 1 (b)-jet with $p_T > 200$ GeV.
- Channel 2a (20k SM and 11k BSM):
  - $E_{Tmiss} > 50$ GeV.
  - $N_{lep} >= 3$ (where $p_{Tlep} > 15$ GeV).
- Channel 2b (340k SM and 90k BSM):
  - $E_{Tmiss} > 50$ GeV.
  - $N_{lep} >= 2$ (where $p_{Tlep} > 15$ GeV).
  - HT > 50 GeV.
- Channel 3 (8.5M SM and 1M BSM):
  - $E_{Tmiss} > 100$ GeV.
  - $H_T > 600$ GeV.

BSM process	Channel 1	Channel 2a	Channel 2b	Channel 3
Z' + monojet	$\times$	$\times$		$\times$
$Z' + W/Z$				$\times$
Z' + single top	$\times$			$\times$
Z' in lepton-violating $U(1)_{L_\mu - L_\tau}$		$\times$	$\times$	
$\not{R}$-SUSY stop-stop	$\times$		$\times$	$\times$
$\not{R}$-SUSY squark-squark	$\times$			$\times$
SUSY gluino-gluino	$\times$	$\times$	$\times$	$\times$
SUSY stop-stop	$\times$			$\times$
SUSY squark-squark	$\times$			$\times$
SUSY chargino-neutralino		$\times$	$\times$	
SUSY chargino-chargino			$\times$	