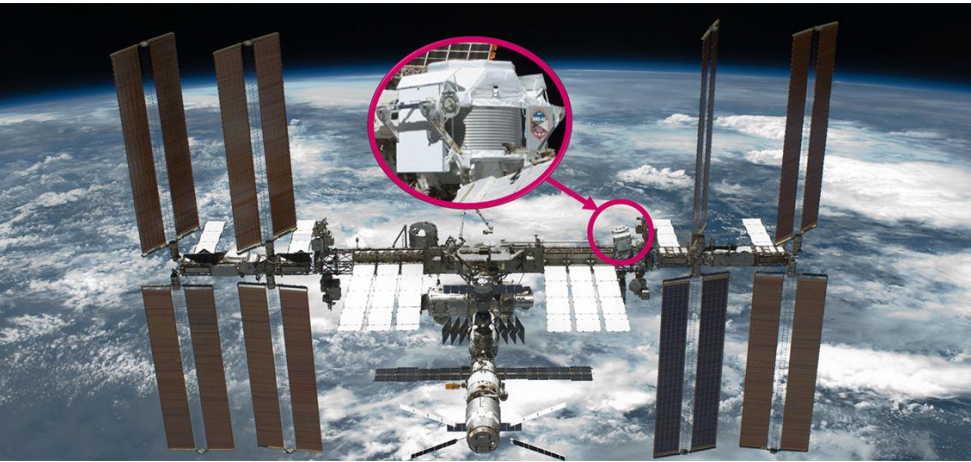
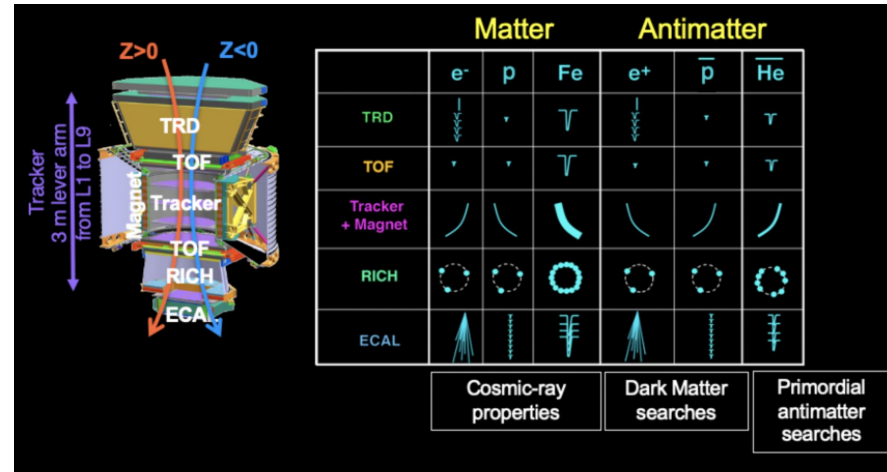


Machine-learning analysis of cosmic-ray nuclei data from the AMS-02 experiment on the International Space Station (ISS)

Shahid Khan



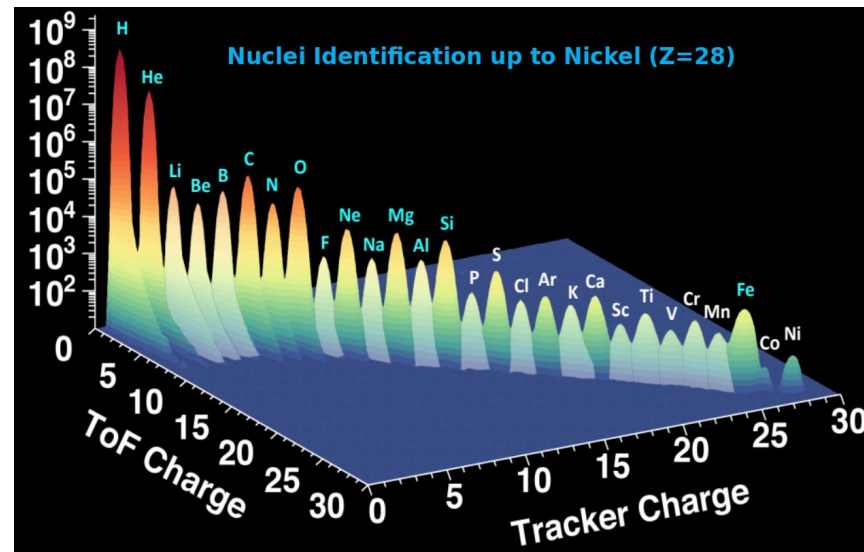
~7000 kg AMS-02 on the ISS



AMS-02 detector components and their use cases₁ for the identification of various charges

Data and MC Mixture

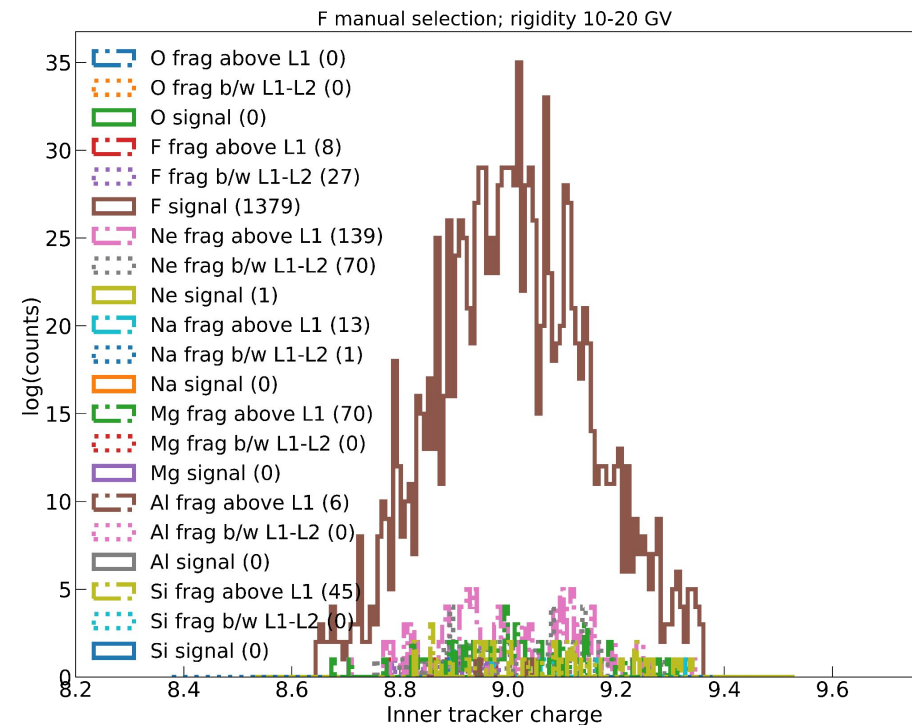
- Around 100 variables from tracker, TOF and TRD
- MC simulations and AMS-02 ISS experimental data with tight manual selections applied have been mixed to create train (equal abundance of all nuclei) and test (natural abundance of nuclei) data sets
- For each nuclei species 3 different classes are created
 - 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:



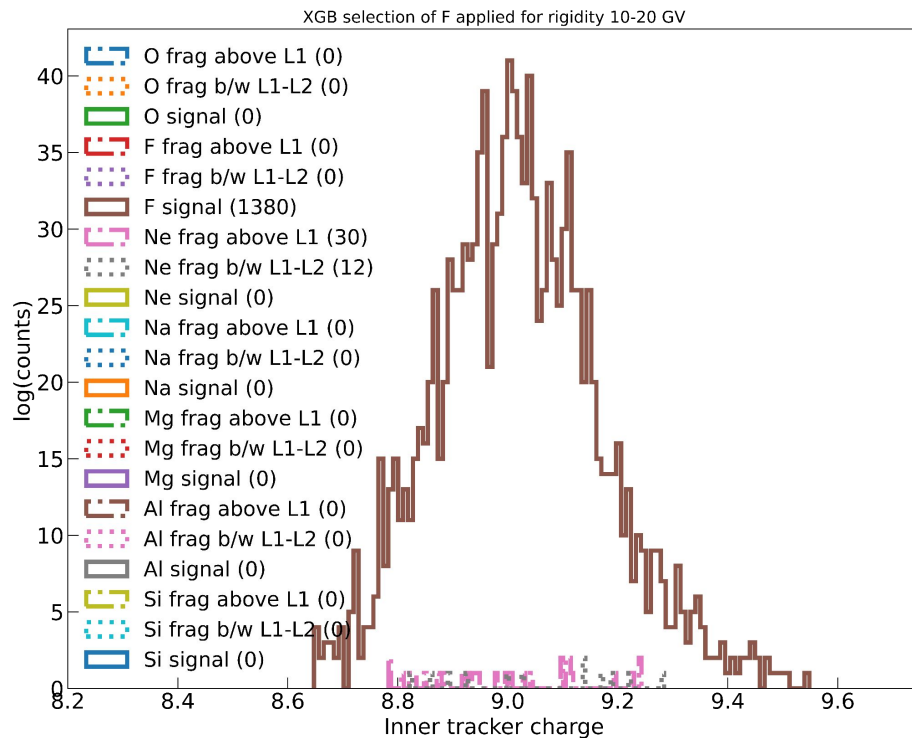
Nuclei identification with tracker, TOF

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



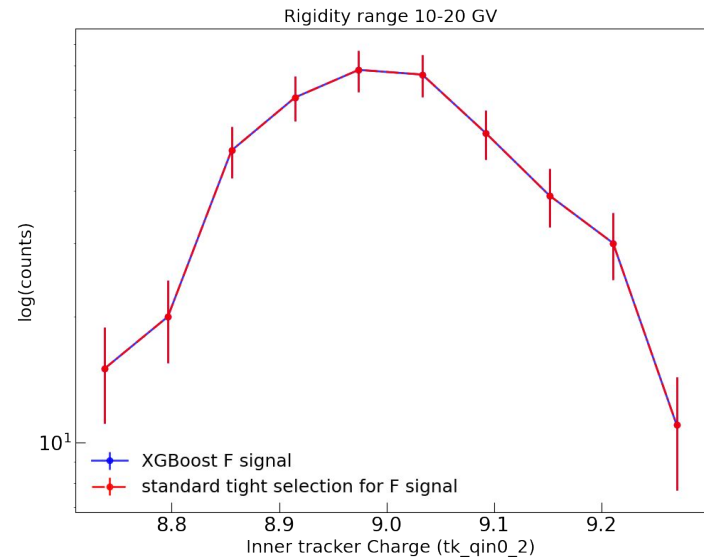
Manually optimized selection



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

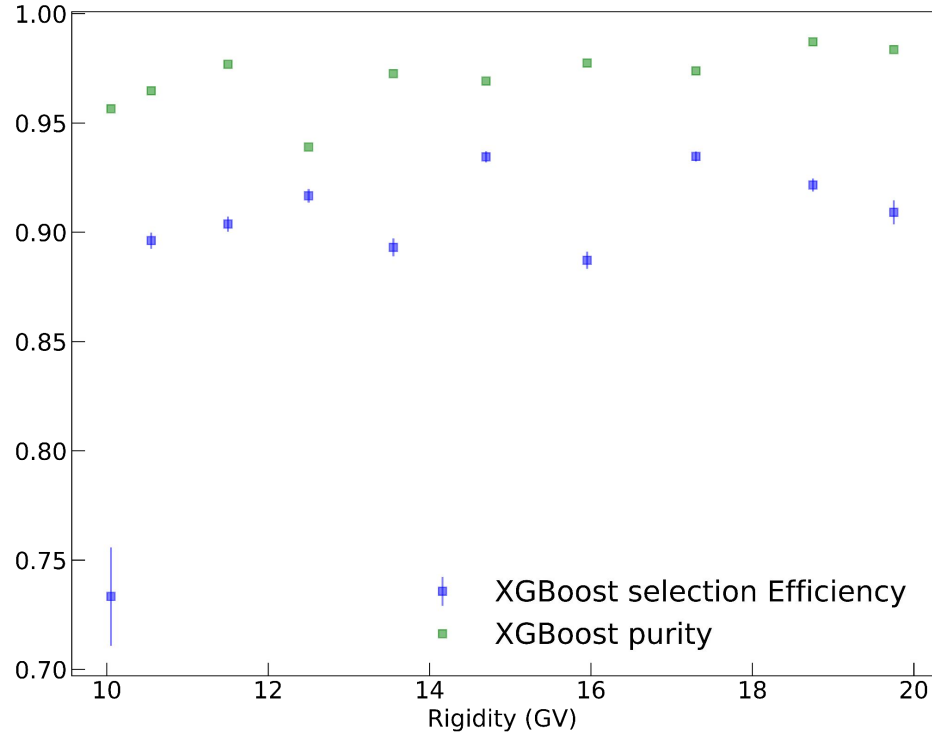
Outlook

- Efficiency correction for selection
- Systematic evaluation

Poster on
Thursday

Back up

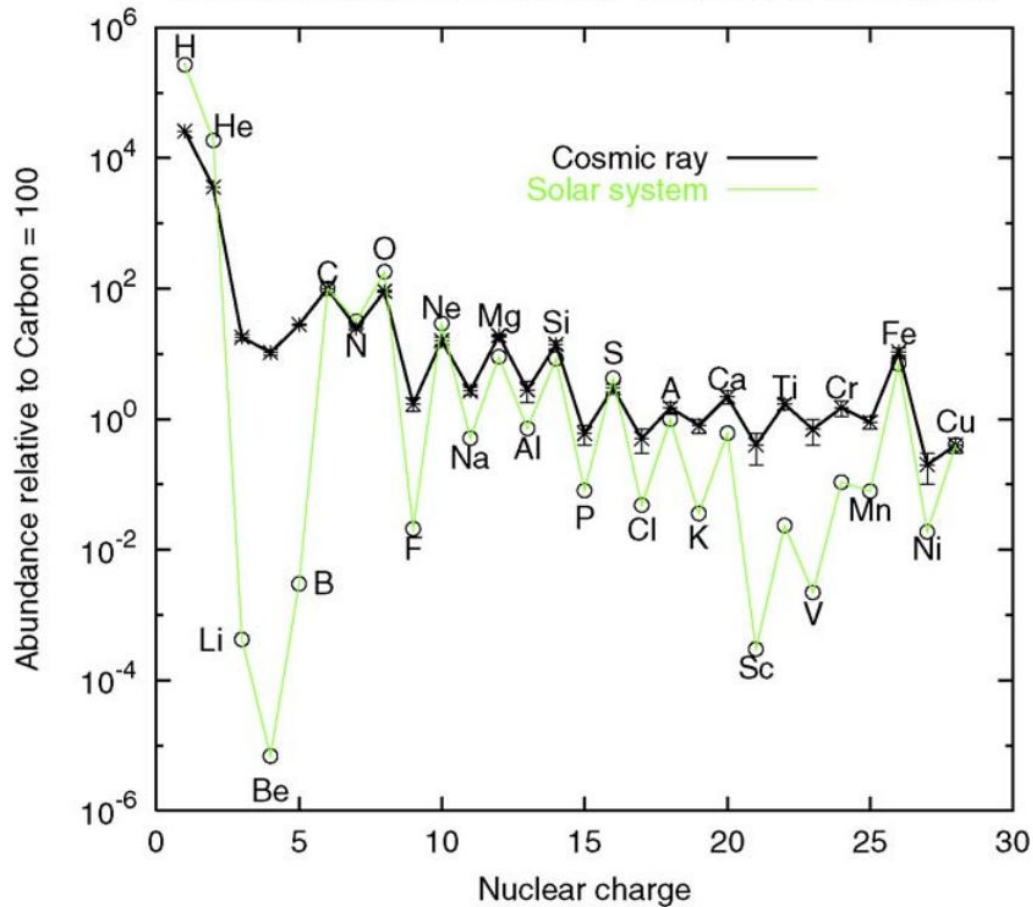
Bin wise visualization



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

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

Nuclear abundance: cosmic rays compared to solar system



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

Generating Lagrangians for particle theories

Elieel Camargo-Molina
Yong Sheng Koay
Rikard Enberg
Stefano Moretti

What:

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



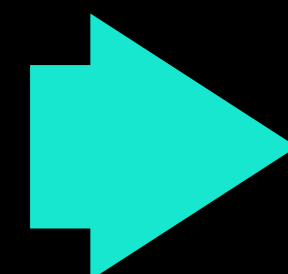
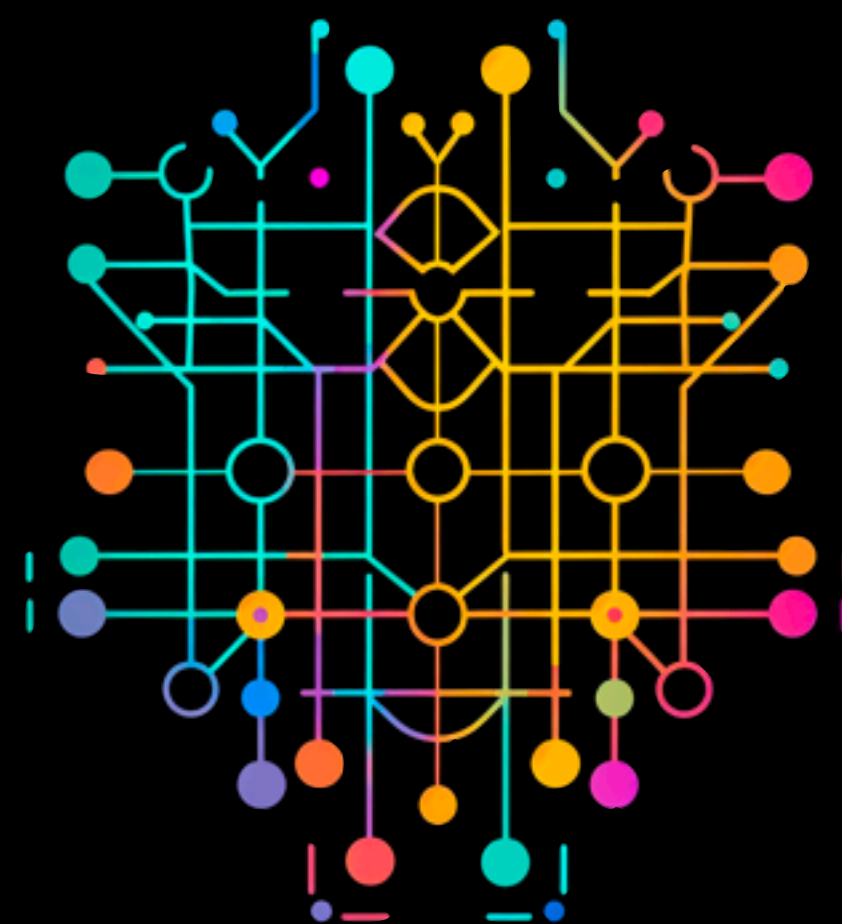
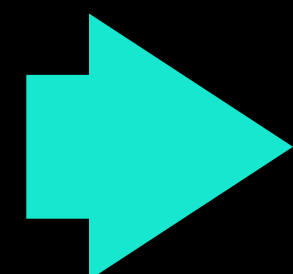
$SU(3) SU(2) U(1)$

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

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

$$\psi_L_{(1,2,\frac{1}{3})}$$

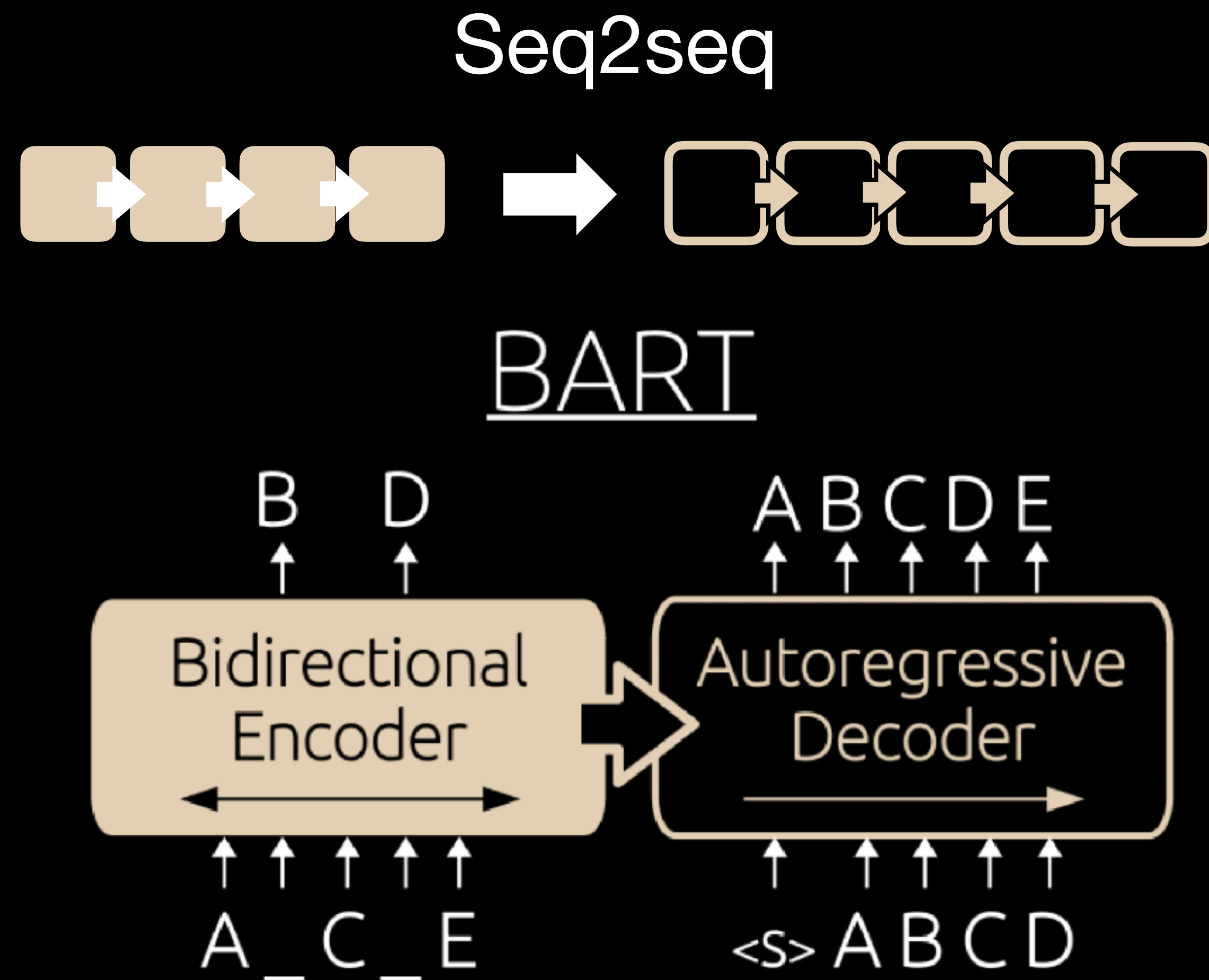
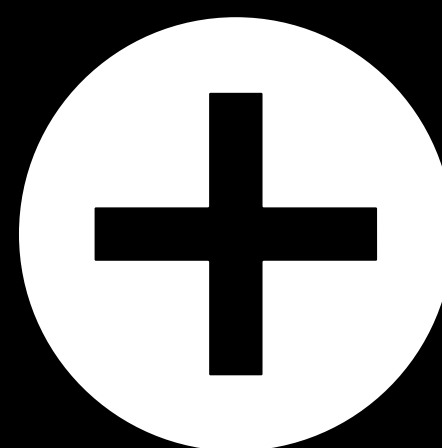
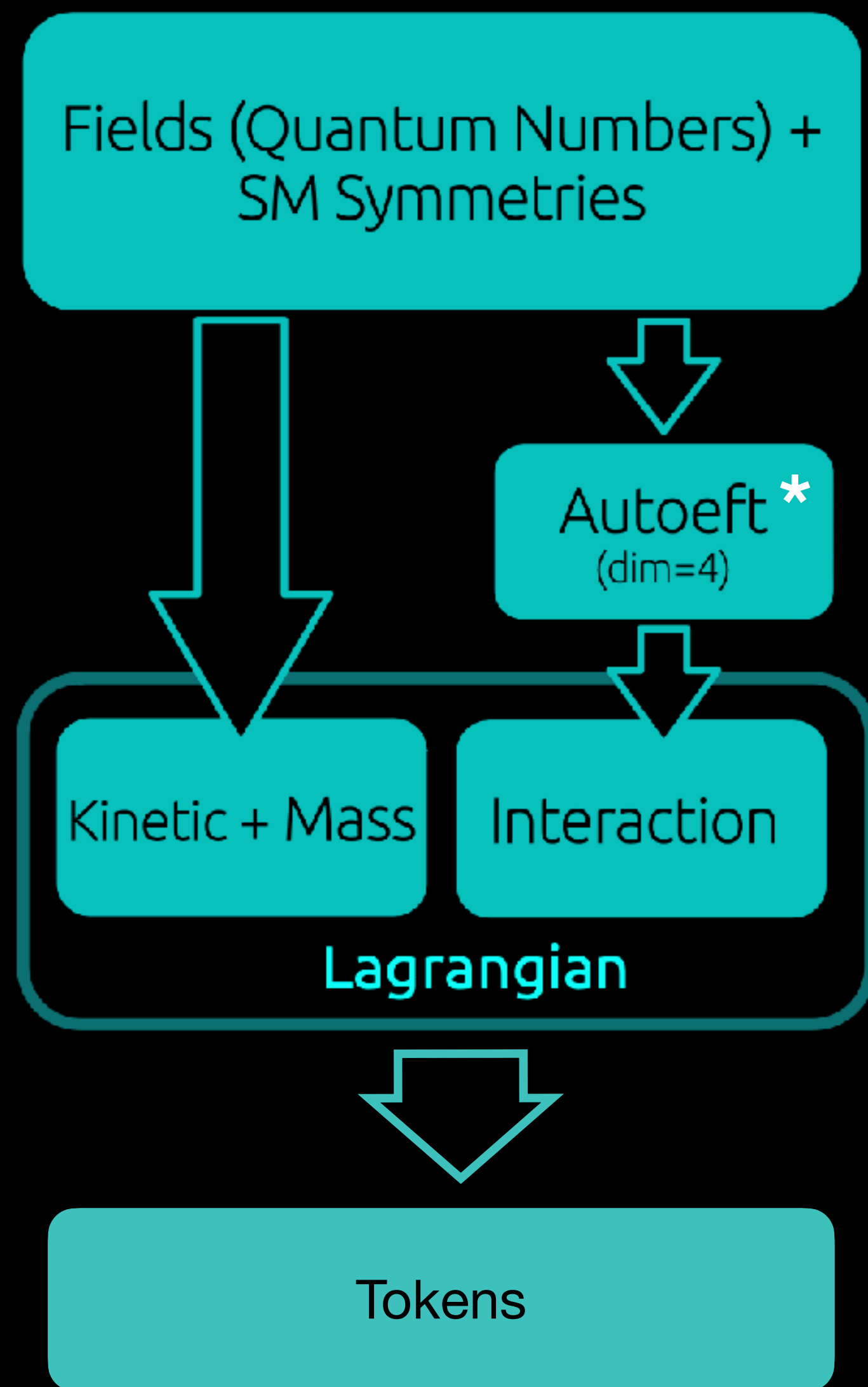
$$\psi_R_{(1,1,-\frac{1}{6})}$$



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

How:

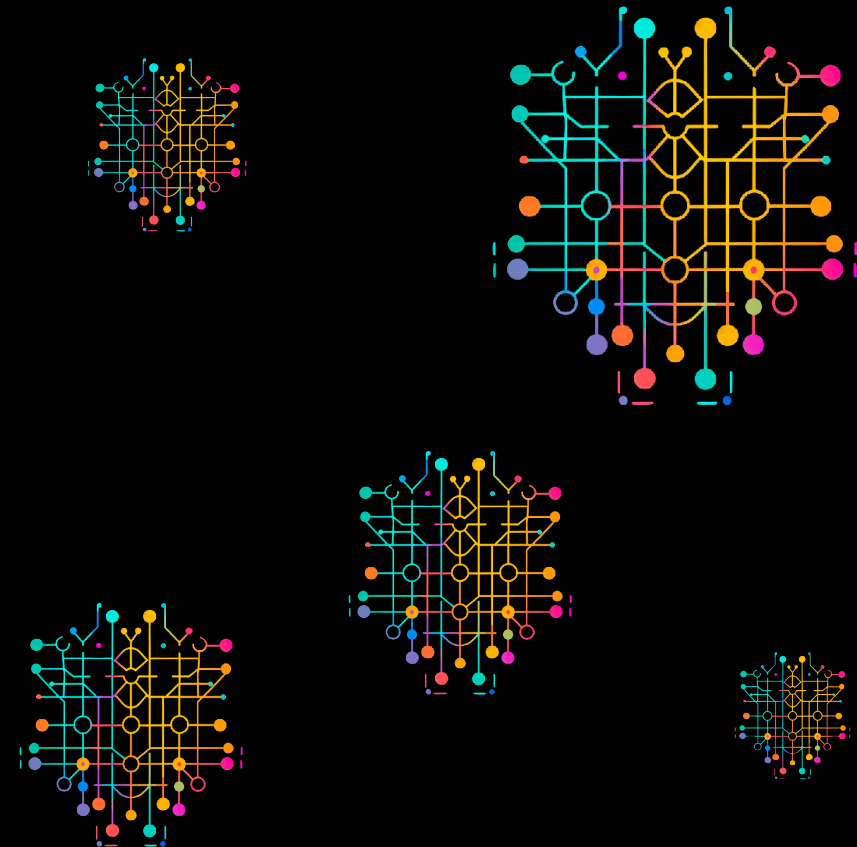
Train a BART model from scratch for a seq2seq task with carefully constructed dataset and a good tokenization



* 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 loopy ...)

17K Lagrangians
~110M pars



Mostly scalars and some fermions

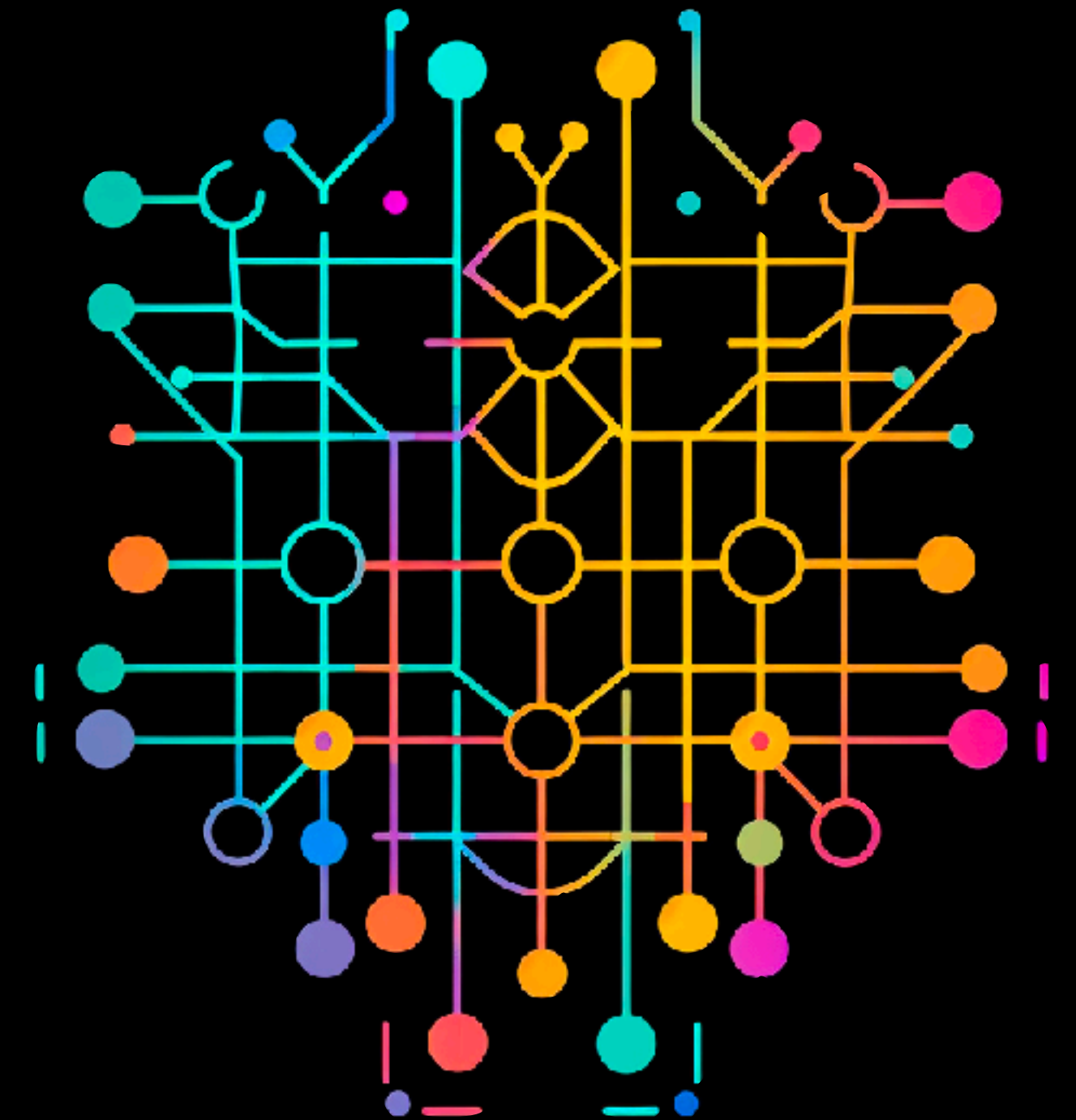
Errors because of tokenization and length

37K Lagrangians
~350M pars



Three strategies for distribution over quantum numbers

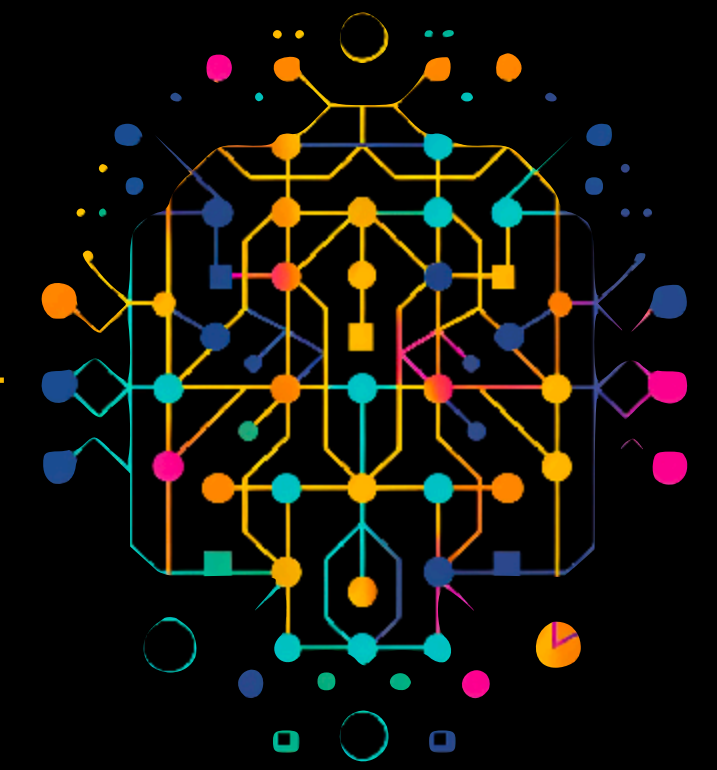
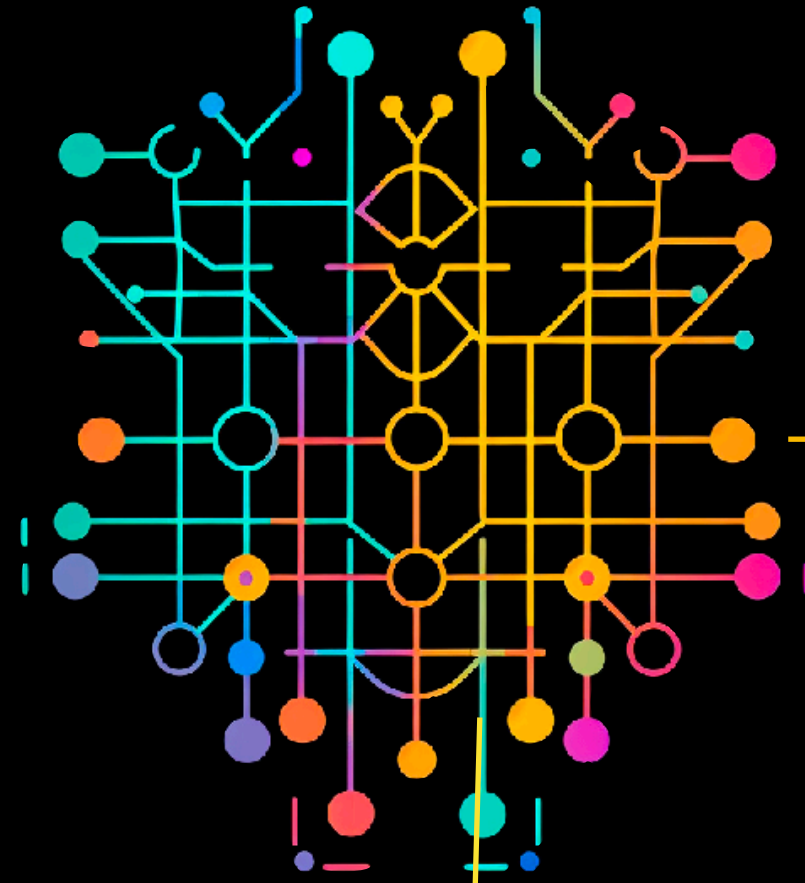
NEXT STEP
~1B pars



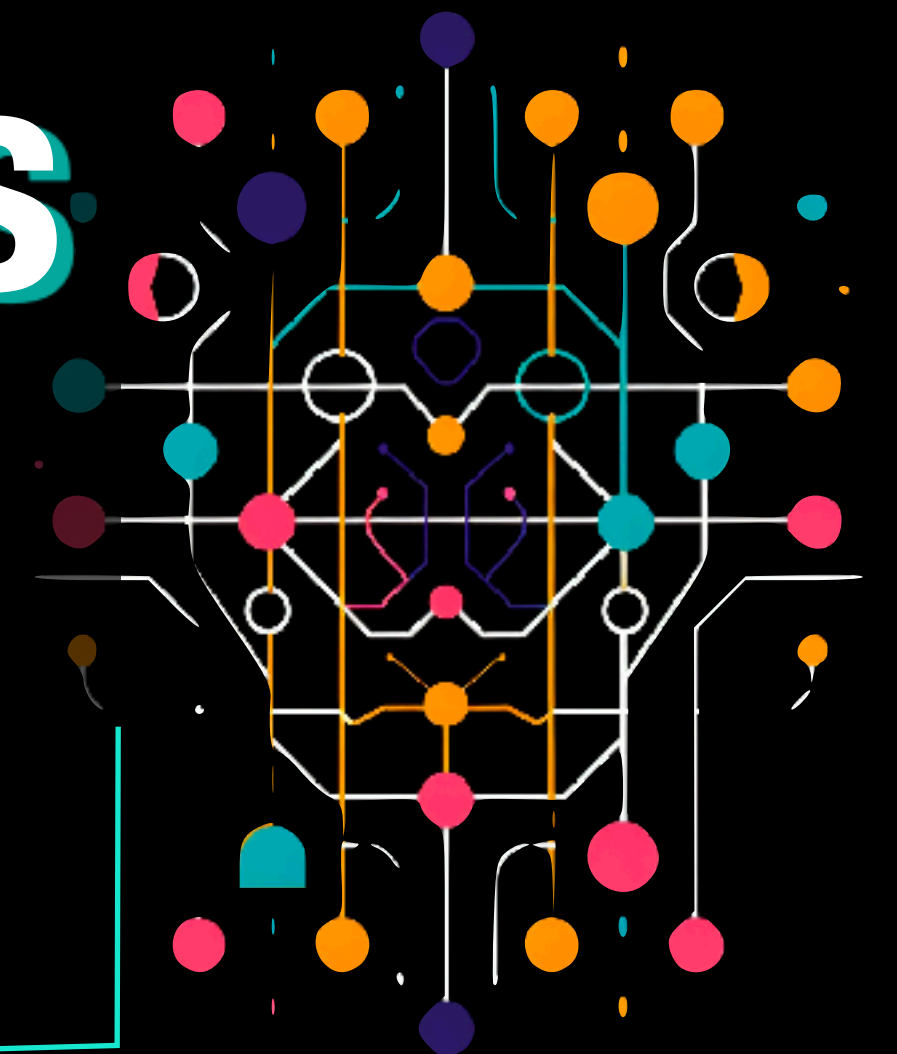
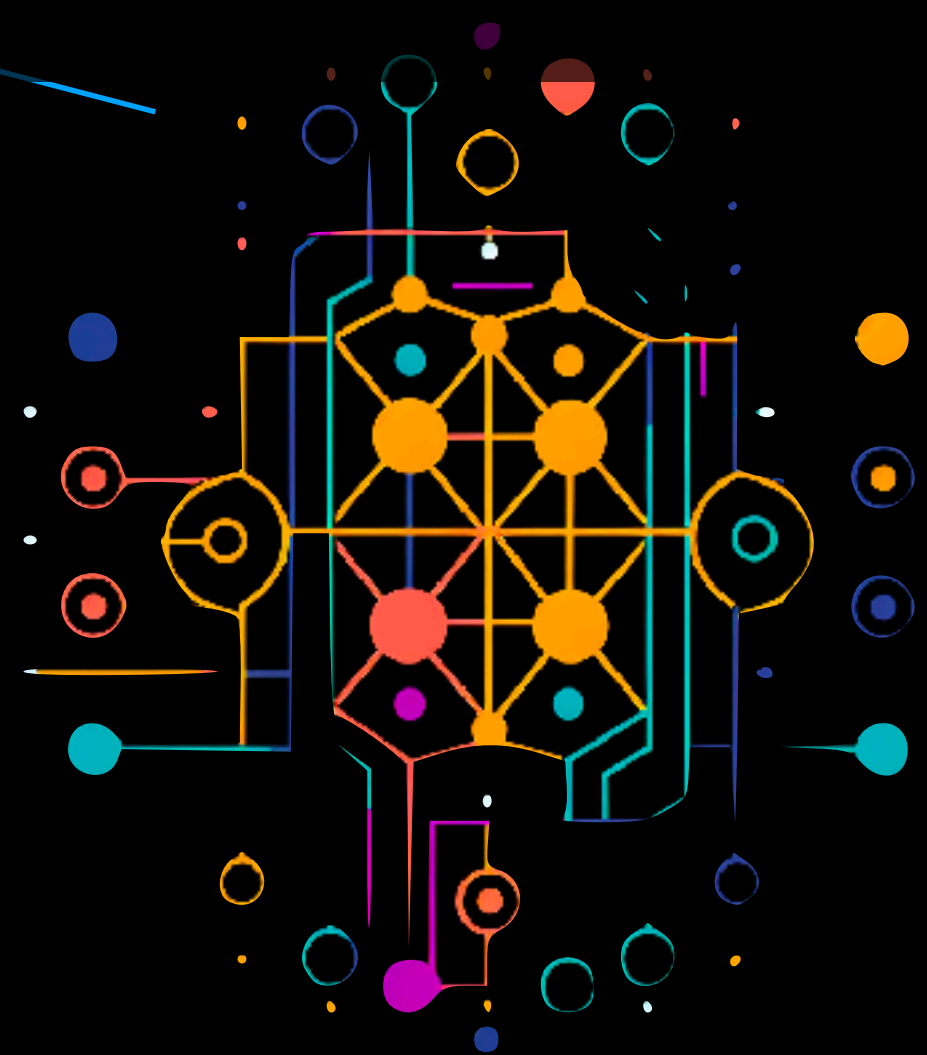
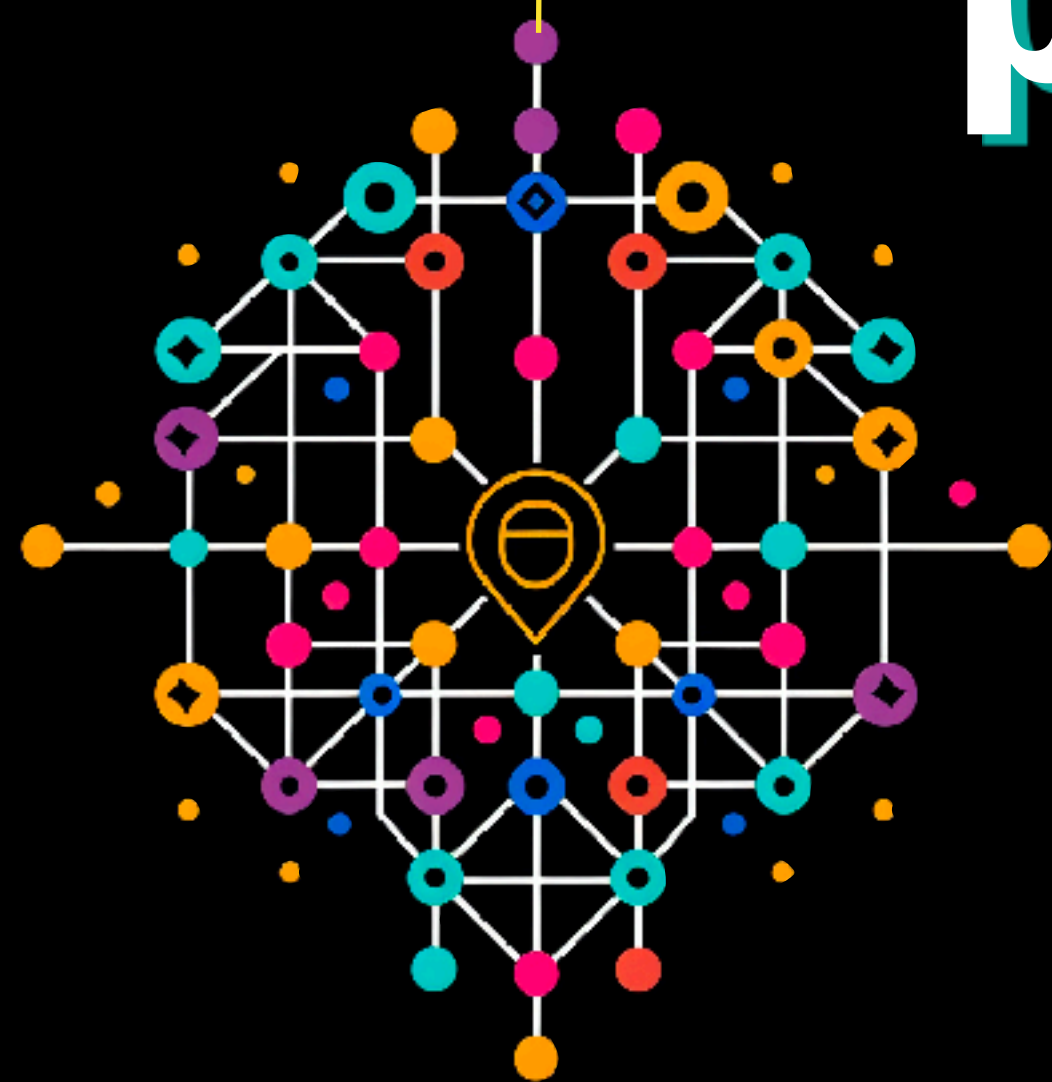
Larger datasets

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



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

Lorentz-Equivariant Geometric Algebra Transformers for High-Energy Physics

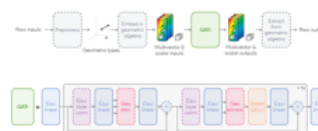
Johann Brehmer, Víctor Bresó Pla, Pim de Hann, Tilman Plehn, Jonas Spinner & Jesse Thaler

Introduction

- Any neural network used for high energy physics analysis needs to be very expressive and efficient.
- All collider processes are ruled by strict and concrete symmetry laws. This feature is often overlooked by standard approaches.
- We introduce the Lorentz Geometric Algebra Transformer (L-GATr), a general purpose architecture that takes full advantage of the known symmetry structure of the data.

Methods

- L-GATr is built on three clear design principles:
 1. Partial and full equivariance with respect to the Lorentz symmetry.
 2. Geometric algebra representation of data.
 3. Transformer backbone, supporting variable length inputs and efficient training.



Conclusions

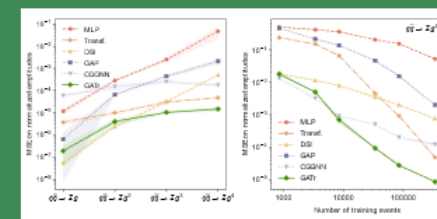
- L-GATr is a flexible framework that can be applied to a multitude of collider physics tasks and achieve state of the art performance.
- L-GATr is able to dominate the scaling law in amplitude regression and display a great performance.
- L-GATr can be trained on multiple processes at once without significant performance loss.
- The main advantages of L-GATr over other equivariant baselines are its computation speed and excellent sample efficiency.

References

- [1] Geometric Algebra Transformer, J. Brehmer et al., 2023, arXiv:2305.18415
- [2] Clifford Group Equivariant Neural Networks, D. Buco et al., 2023, arXiv:2305.11141
- [3] Energy Flow Networks: Deep Sets for Particle Jets, P. de Hann et al., 2019, arXiv:1810.05165
- [4] PELICAN: Permutation Equivariant and Lorentz Invariant or Covariant Aggregator Network for Particle Physics, A. Bognányi et al., 2022, arXiv:2211.04654

Equivariance and transformers are enough to perform complex collider physics tasks quickly.

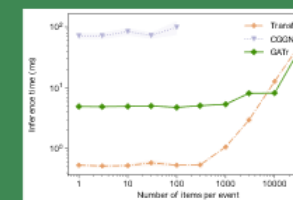
▶ Amplitude regression



▶ Jet tagging

Model	Top tagging				Quark-gluon tagging			
	Acc	AUC	$1/\epsilon_{\text{sig}}$ ($\epsilon_{\text{bg}} = 0.5$)	$1/\epsilon_{\text{sig}}$ ($\epsilon_{\text{bg}} = 0.3$)	Acc	AUC	$1/\epsilon_{\text{sig}}$ ($\epsilon_{\text{bg}} = 0.5$)	$1/\epsilon_{\text{sig}}$ ($\epsilon_{\text{bg}} = 0.3$)
ParticleNet	0.940	0.9858	397	1615	0.840	0.9116	39.8	96.6
ParT	0.940	0.9858	413	1602	0.849	0.9203	47.9	129.5
LorentzNet	0.942	0.9868	498	2195	0.841	0.9156	42.4	110.2
CGNN	0.942	0.9869	500	2172				
PELICAN	0.9426	0.9870	2250	2250	0.8551	0.9252	52.3	149.8
L-GATr (ours)	0.9412	0.9866	529	2030	0.8463	0.9181	45.7	124.2

▶ Scalability properties



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 Moskvitina (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.

Dark Machines

[About](#) [News](#) [Events](#) [Projects](#) [Researchers](#) [White paper](#) [Mailinglist](#) [Contribute](#) 

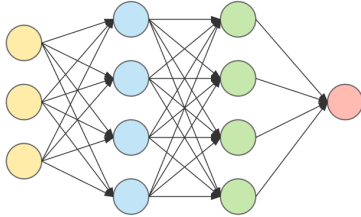
The Dark Machines Anomaly Score Challenge:
Benchmark Data and Model Independent Event
Classification for the Large Hadron Collider

T. Aarrestad^a M. van Beekveld^b M. Bona^c A. Boveia^e S. Caron^d J. Davies^c
A. De Simone^{f,g} C. Doglioni^h J. M. Duarteⁱ A. Farbin^j H. Gupta^k L. Hendriks^d
L. Heinrich^a J. Howarth^l P. Jawahar^{m,a} A. Jueidⁿ J. Lastow^h A. Leinweber^o
J. Mamuzic^p E. Merényi^q A. Morandini^r P. Moskvitina^d C. Nellist^d J. Ngadiuba^{s,t}
B. Ostdiek^{u,v} M. Pierini^a B. Ravina^l R. Ruiz de Austri^p S. Sekmen^w
M. Touranakou^{x,a} M. Vaškeviciute^l R. Vilalta^y J.-R. Vlimant^t R. Verheyen^z
M. White^o E. Wulff^h E. Wallin^h K.A. Wozniak^{α,a} Z. Zhang^d

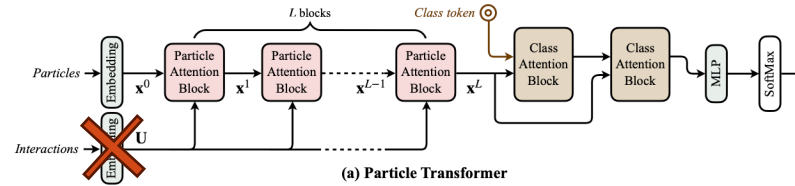
ARCHITECTURES AND TECHNIQUES

Architectures

Multi-Layer Perceptron (MLP)



Particle Transformer (ParT)



No pairwise interactions

<https://arxiv.org/abs/2211.05143>

ParT+ SM couplings

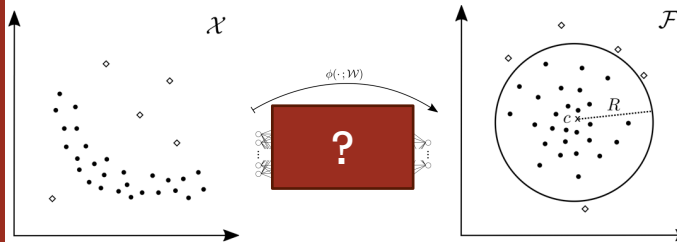
- Pairwise interactions
 - $\ln(m_{ij}^2)$
 - $\ln(\Delta R_{ij})$
- Physical information from Standard Model: couplings.

Developed by this group

Techniques

Deep Support Vector Data Description (dSVDD)

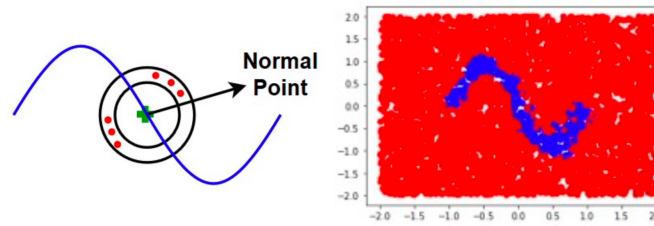
- Add an output layer with certain dimensions.
- Training: minimise distance to a centre in the hypersphere (anomaly score).
- Outliers are considered anomalies.
- Make **ensemble** for different dimensions.



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.

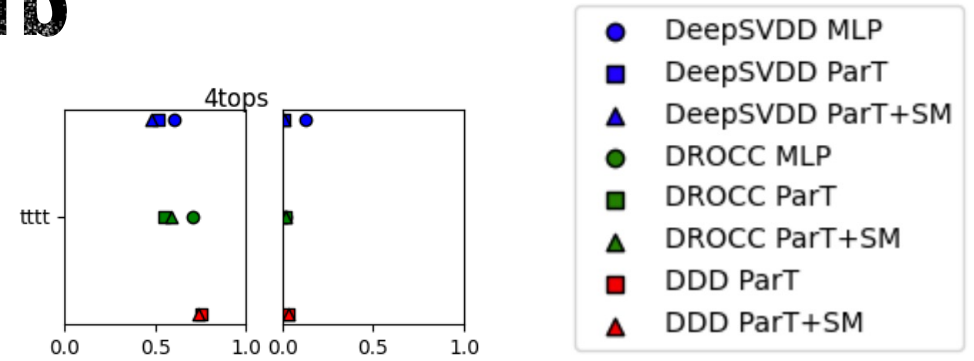
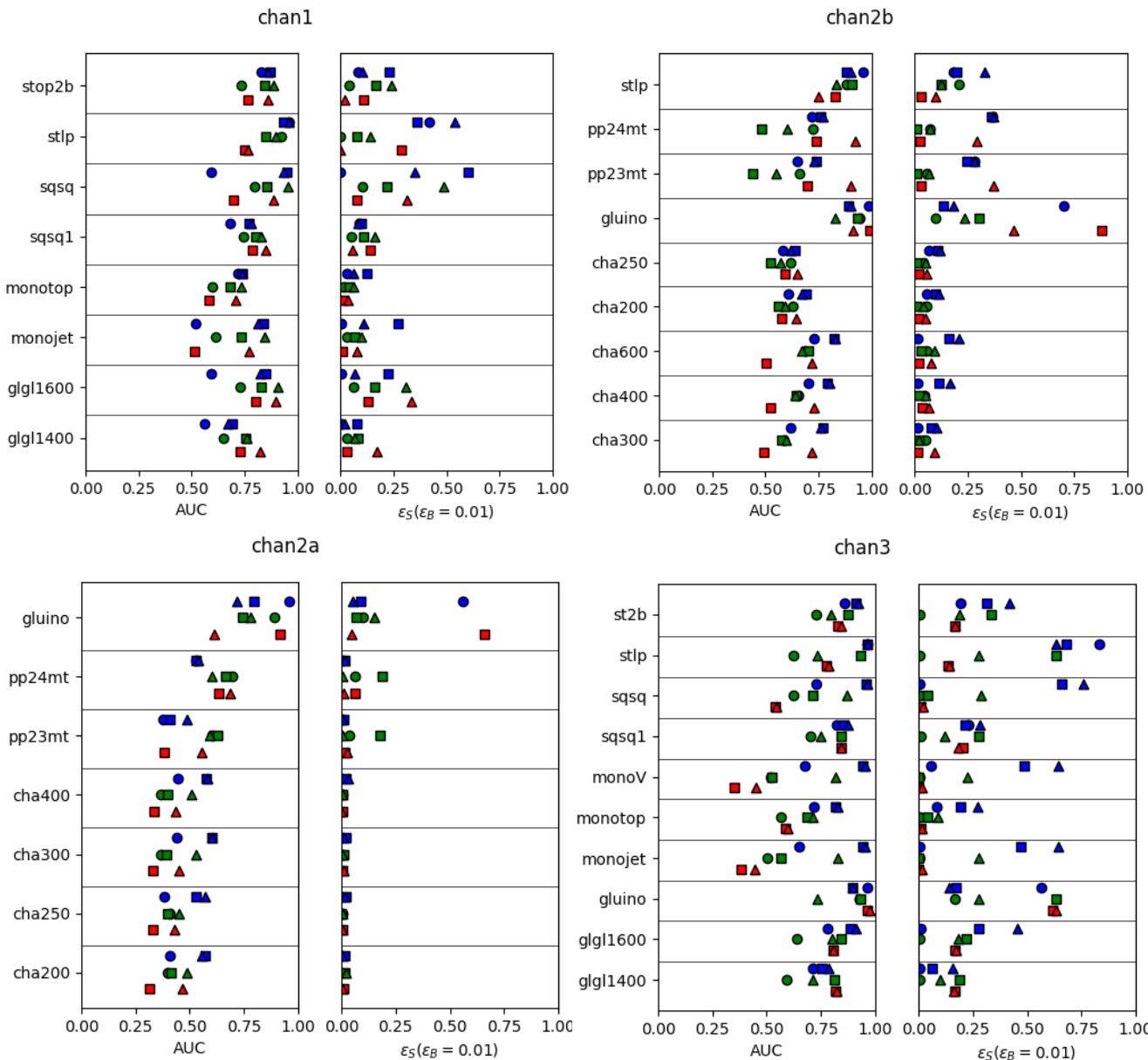
$$\sum_{i=1}^n [\ell(f_{\theta}(x_i), 1) + \mu \max_{\tilde{x}_i \in N_i(r)} \ell(f_{\theta}(\tilde{x}_i), -1)]$$
- Weakly supervised implementation



Discriminant distortion 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 $\sim 0.7-0.8$ 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).



BACK-UP



CHANNELS AND SIGNALS

- Channel 1 (214k SM and 38k BSM):

- $H_T \geq 600$ GeV .
- $E_{T\text{miss}} \geq 200$ GeV.
- $E_{T\text{miss}}/HT \geq 0.2$.
- 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):

- $E_{T\text{miss}} > 50$ GeV.
- $N_{\text{lep}} \geq 3$ (where $p_{T\text{lep}} > 15$ GeV).

- Channel 2b (340k SM and 90k BSM):

- $E_{T\text{miss}} > 50$ GeV.
- $N_{\text{lep}} \geq 2$ (where $p_{T\text{lep}} > 15$ GeV).
- $HT > 50$ GeV.

- Channel 3 (8.5M SM and 1M BSM):

- $E_{T\text{miss}} > 100$ GeV.
- $H_T > 600$ GeV.

BSM process	Channel 1	Channel 2a	Channel 2b	Channel 3
$Z' + \text{monojet}$	×	×		×
$Z' + W/Z$				×
$Z' + \text{single top}$	×			×
Z' in lepton-violating $U(1)_{L_\mu - L_\tau}$		×	×	
\cancel{R} -SUSY stop-stop	×		×	×
\cancel{R} -SUSY squark-squark	×			×
SUSY gluino-gluino	×	×	×	×
SUSY stop-stop	×			×
SUSY squark-squark	×			×
SUSY chargino-neutralino		×	×	
SUSY chargino-chargino			×	