Common ground in classification (and generation?)

Gregor Kasieczka
(gregor.kasieczka@uni-hamburg.de)

IDT ErUM Meeting
2020-04-02

CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Emmy Noether-Programm
Deutsche Forschungsgemeinschaft DFG

PIER
Partnership of Universität Hamburg and DESY

Bundesministerium für Bildung und Forschung

Introduction

- Research into ML applications to specific physics problems as part of task area C
- Chance for wider ranging developments:
  - methods generalise beyond specific problem
  - recommendations and strategies

- Two common problems:

Classification

- Distinguish different observations, particles, galaxies,…
- Supervised - need labelled training data from simulation
  (or controlled observation)
- Straightforward to train and benchmark

Generative models

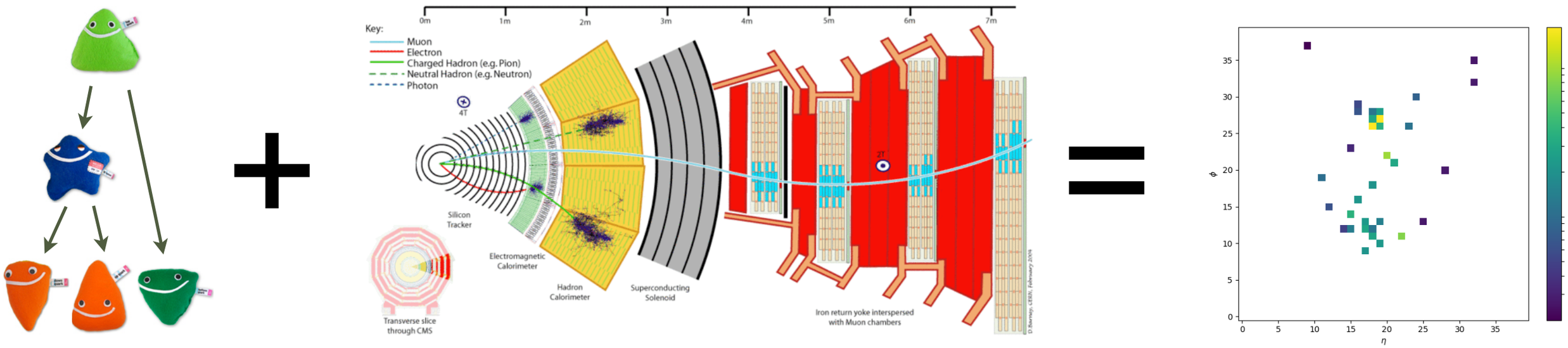
- Accelerate simulation of physics processes and detector responses
- Unsupervised - can train directly from data
- More difficult train, unclear how to benchmark
- *Of course this does not mean that it is not an interesting problem, only potentially more difficult to generalise for us*

Example:
Top Tagging Challenge

Heavy Resonance Tagging

- **Goal**: Distinguish decay products of heavy resonance (top quark, W/Z boson, Higgs boson) from other particles (light quark/gluon jets)
- Achieve by looking at substructure of jets in the detector

Top Quark *(Simulated) Detector* *Jet Image*

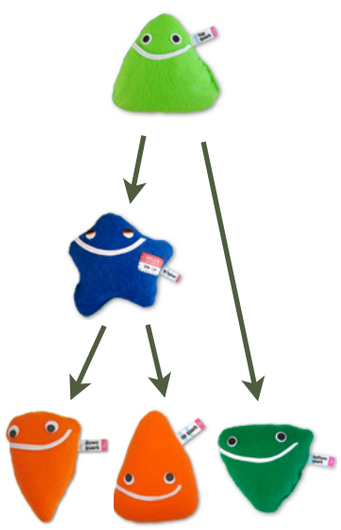


(ignoring parton shower, hadronisation,…)

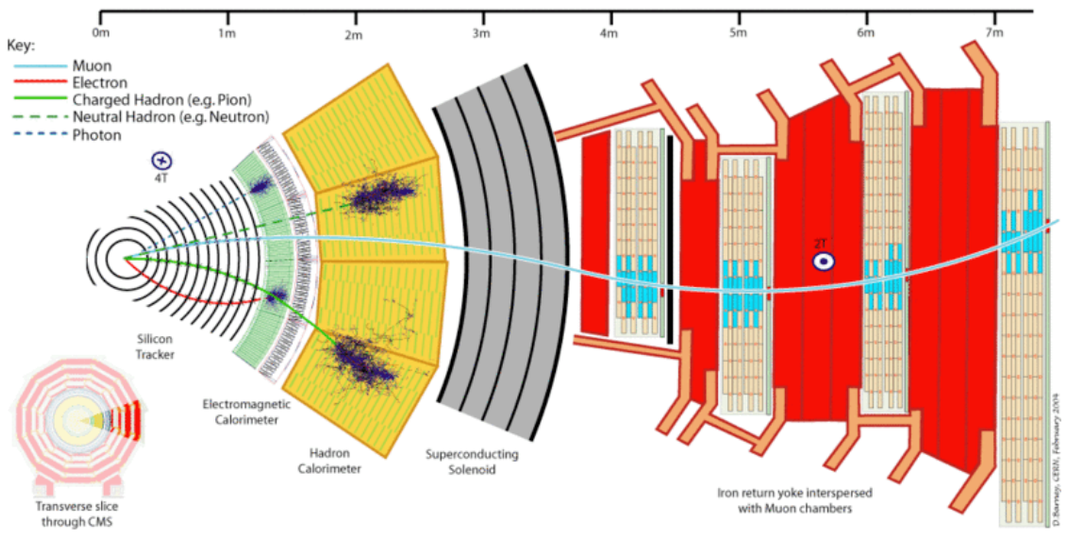
Heavy Resonance Tagging

- **Goal**: Distinguish decay products of heavy resonance (top quark, W/Z boson, Higgs boson) from other particles (light quark/gluon jets)
- Achieve by looking at substructure of jets in the detector

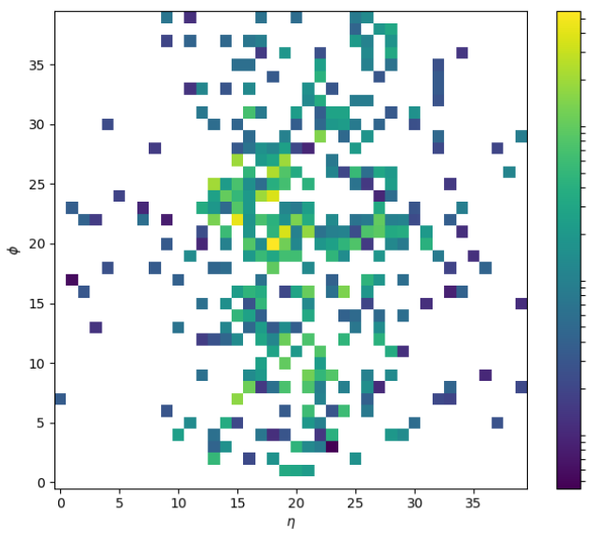
Top Quark *(Simulated) Detector* *10 Images*



+



=

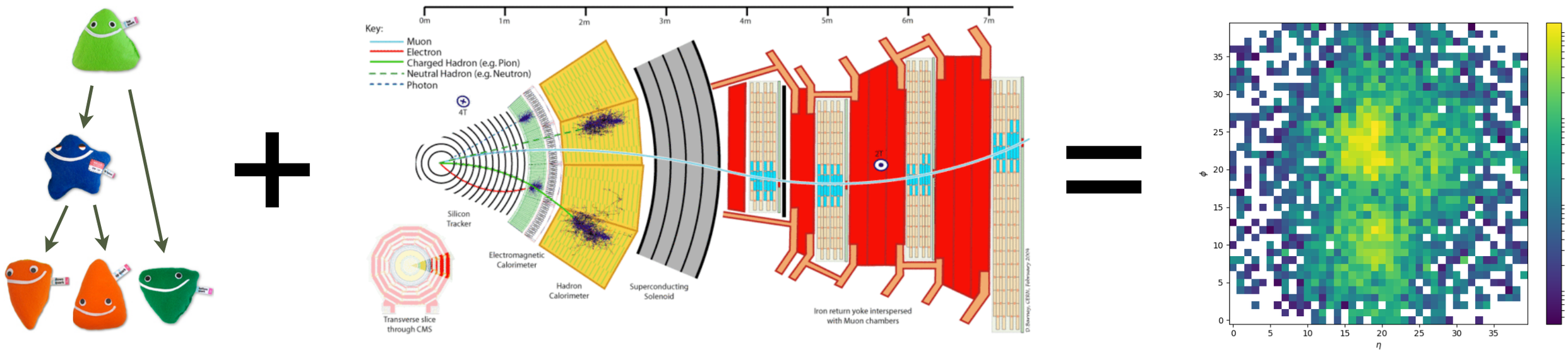


(ignoring parton shower, hadronisation,…)

Heavy Resonance Tagging

- **Goal**: Distinguish decay products of heavy resonance (top quark, W/Z boson, Higgs boson) from other particles (light quark/gluon jets)
- Achieve by looking at substructure of jets in the detector

Top Quark *(Simulated) Detector* *100 Images*

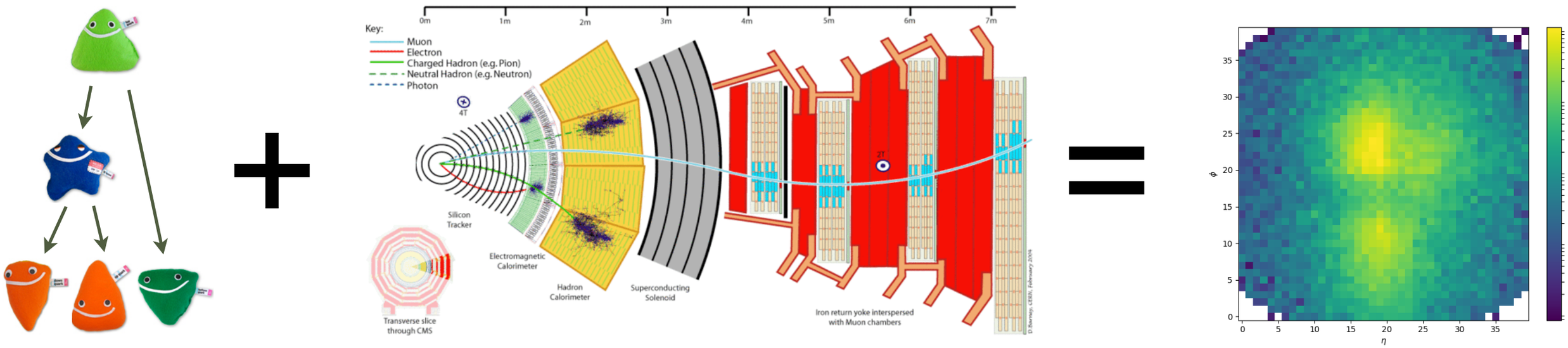


(ignoring parton shower, hadronisation,…)

Heavy Resonance Tagging

- **Goal**: Distinguish decay products of heavy resonance (top quark, W/Z boson, Higgs boson) from other particles (light quark/gluon jets)
- Achieve by looking at substructure of jets in the detector

Top Quark *(Simulated) Detector* *1000 Images*

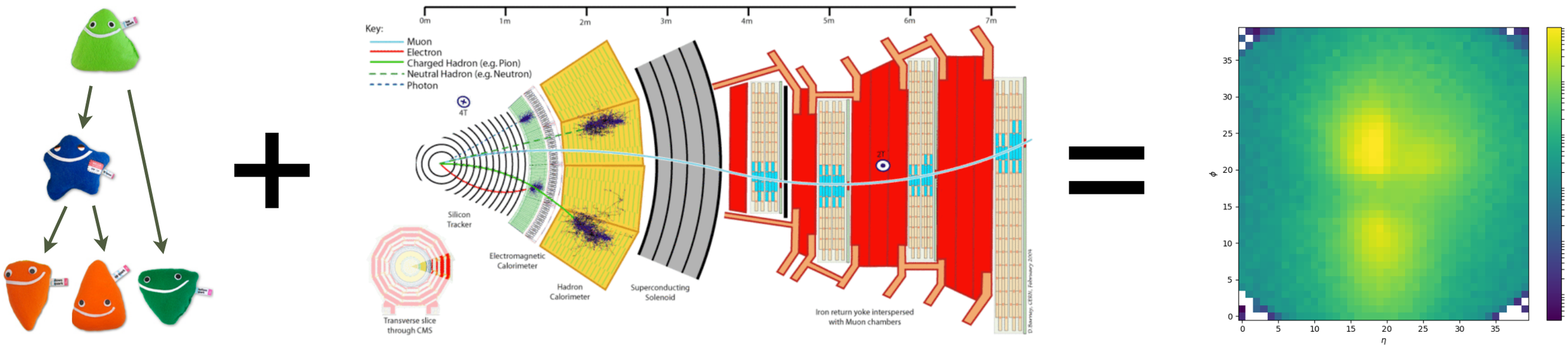


(ignoring parton shower, hadronisation,…)

Heavy Resonance Tagging

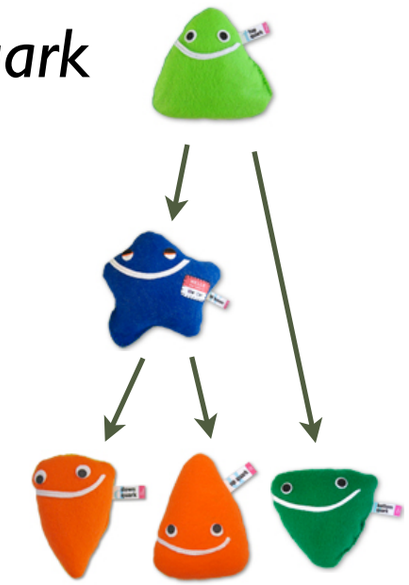
- **Goal**: Distinguish decay products of heavy resonance (top quark, W/Z boson, Higgs boson) from other particles (light quark/gluon jets)
- Achieve by looking at substructure of jets in the detector

Top Quark *(Simulated) Detector* *10000 Images*

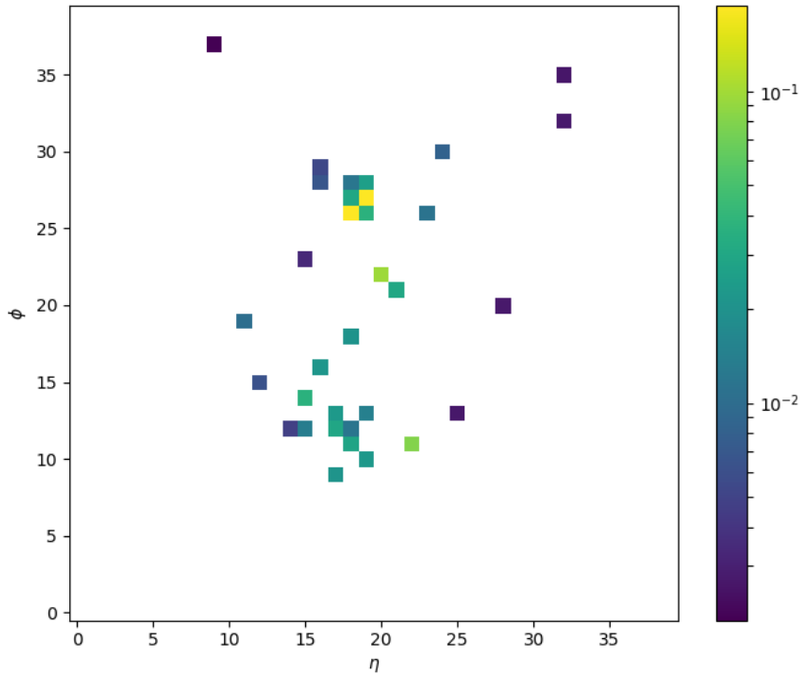


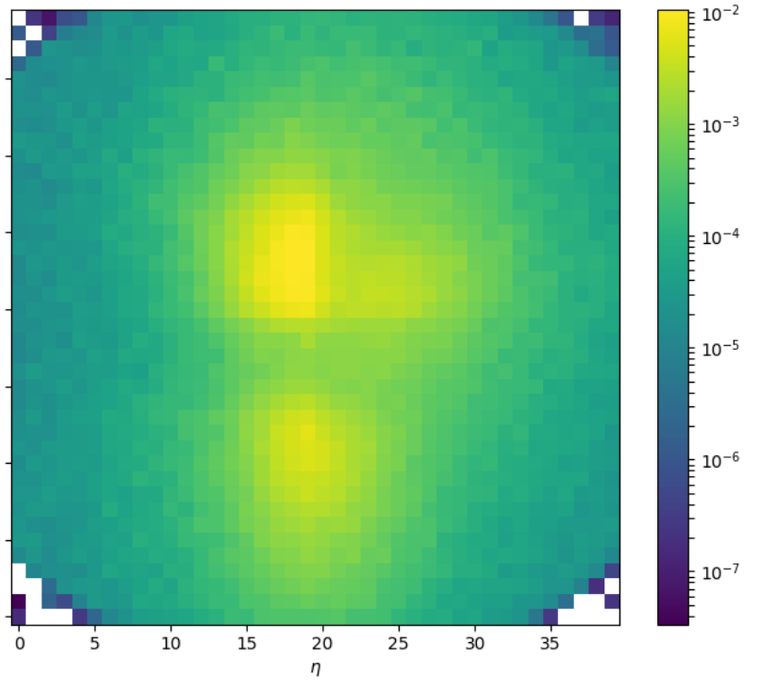
(ignoring parton shower, hadronisation,…)

Top Quark Jet



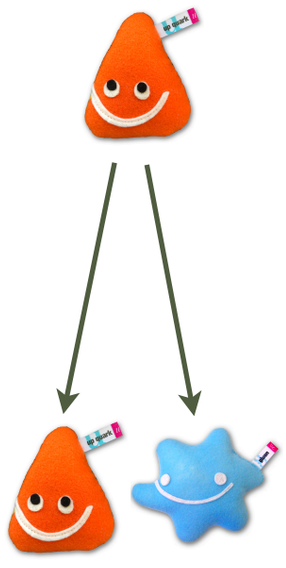
=



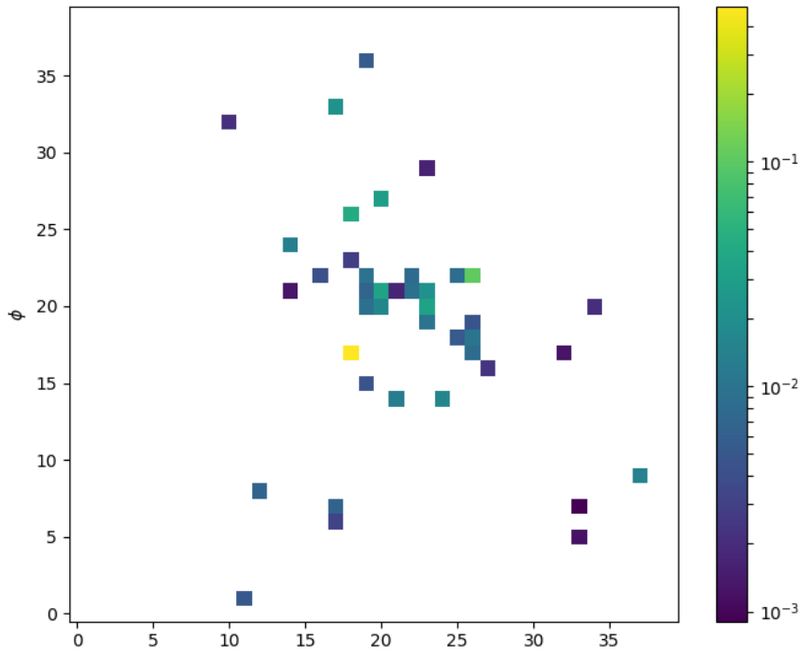


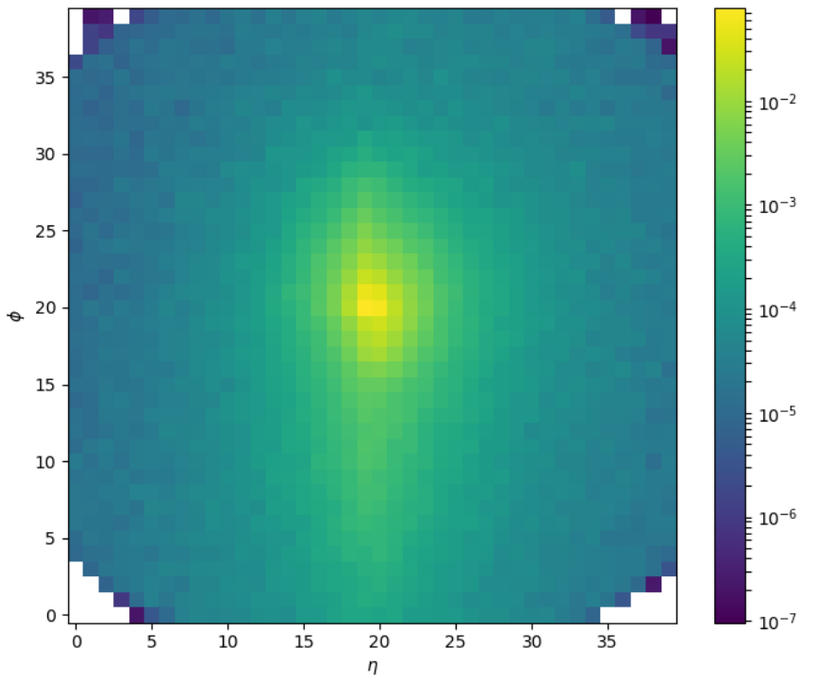
- *Binary classification task*
- *Fully supervised learning (using simulation)*
- *40x40 Pixels, E_T*
- *Perfectly suited for deep learning algorithms*

QCD Jet



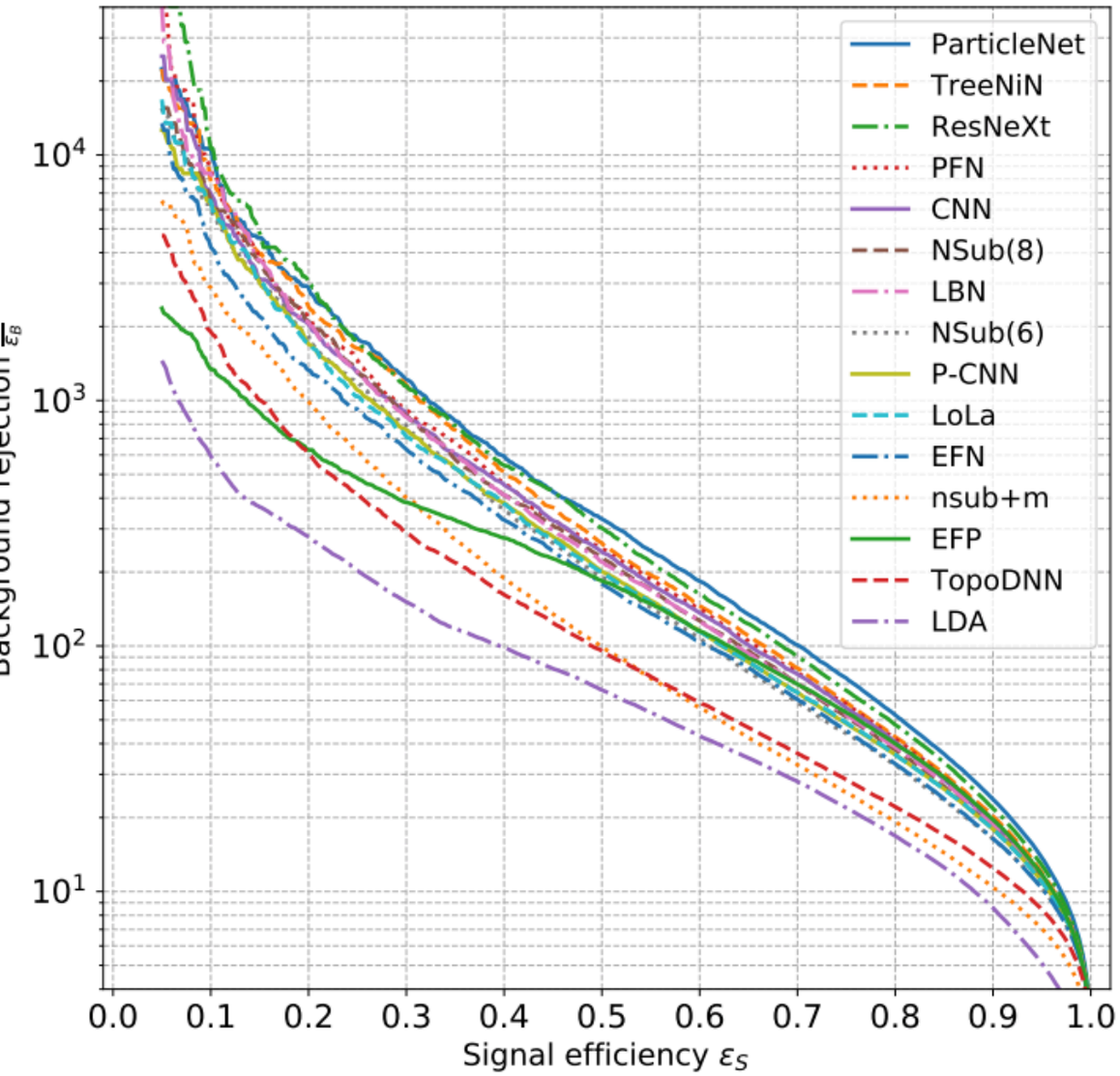
=





Architecture Comparison

Community performance comparison (toy dataset public): 1902.09914



- 1.2M simulated top quark and background events
- Great test-bed to compare different data representations
  - (and, of course, useful for new physics searches, top/Higgs measurements)
- Still surprising gains in performance
  - Although it needs to be seen how well these translate to data
- (Also developments in flavour tagging, not covered here)

Technical details

- Samples hosted on DESYCloud as .hdf5 files
- Up to 200 4-vectors per top-jet
  - Can also represent as images
- Total of ~1.6 GB
- Groups had access to all datasets - including final test
  trust that people do not abuse this
- Shared performance results and classification output on test-sample

	AUC	Acc	$1/\epsilon_B$ ($\epsilon_S = 0.3$)			#Param
			single	mean	median	
CNN [16]	0.981	0.930	914±14	995±15	975±18	610k
ResNeXt [31]	0.984	0.936	1122±47	1270±28	1286±31	1.46M
TopoDNN [18]	0.972	0.916	295±5	382± 5	378 ± 8	59k
Multi-body N-subjettiness 6 [24]	0.979	0.922	792±18	798±12	808±13	57k
Multi-body N-subjettiness 8 [24]	0.981	0.929	867±15	918±20	926±18	58k
TreeNiN [43]	0.982	0.933	1025±11	1202±23	1188±24	34k
P-CNN	0.980	0.930	732±24	845±13	834±14	348k
ParticleNet [47]	0.985	0.938	1298±46	1412±45	1393±41	498k
LBN [19]	0.981	0.931	836±17	859±67	966±20	705k
LoLa [22]	0.980	0.929	722±17	768±11	765±11	127k
LDA [54]	0.955	0.892	151±0.4	151.5±0.5	151.7±0.4	184k
Energy Flow Polynomials [21]	0.980	0.932	384			1k
Energy Flow Network [23]	0.979	0.927	633±31	729±13	726±11	82k
Particle Flow Network [23]	0.982	0.932	891±18	1063±21	1052±29	82k
GoaT	0.985	0.939	1368±140		1549±208	35k

Can use as template for a first IDT-ErUM overview

Challenges

- ErUM domains work with a diverse set of data representations
  - Different detector geometries, different types of experiments, theory calculations…
- But: several approaches should be flexible enough:
  - Fully connected
  - 1D convolutions
  - Graphs
  - (Images)
  - …?
- Developing a first set of general recommendations for non-ML-expert ErUM practitioners already would be valuable
  - ‘Physics’ performance
  - ‘Computational’ performance
  - Examples shared, eg on Vispa

What do we have?

- Datasets:
  - Top tagging reference sample
    1902.09914
    (Can either use directly or provide a slimmed version)
  - FIAS (Jan Steinheimer, Kai Zhou): Spinodal vs. Maxwell classification
    arXiv: 1906.06562
    160k 20x20 Grayscale images
  - FIAS (Jan Steinheimer, Kai Zhou): QCD transition: EOSL/EOSQ
    arXiv: 1910:11530
    24x24 2D histograms of pion spectra
  - Aaachen (Jonas Glombitza):
    Air shower classification: Proton vs Iron
    https://doi.org/10.1016/j.astropartphys.2017.10.006
    100k examples, ~7k channels
  - Upcoming: LMU (James Kahn)
    Belle II
- Algorithms:
  - Fully connected
  - Images
  - 1D Convolution
  - Graph Convolution
  - Spectral Graph Networks

If you have additional classification datasets
OR are can test (develop?) a specific algorithm:
Let me know

Next steps

Classification:

- Meet again in the next 1-2 weeks to go over data and coordinate next steps
  - Poll here: https://terminplaner4.dfn.de/LfXuLR11erQUqPMO
  - Already have critical mass of datasets and algorithms to go ahead!

Generative Models:

- Also interesting progress at multiple locations (Hamburg/DESY update from Erik later today)
- Exciting (but more challenging) potential follow-up

Thank you!