Studying hadronization with Machine Learning techniques

Margaret Island Symposium 2022 on Vacuum Structure, Particles, and Plasmas

15-18 05 2022

GÁBOR BÍRÓ

biro.gabor@wigner.hu

Gergely Gábor Barnaföldi
Bence Tankó-Bartalis

WIGNER UNIVERSITY OF OXFORD

arXiv:2111.15655

Outline

- Machine Learning: the motivation
- Research goals
- Results
- Future directions
- Summary

Data, data, and more data

Large Hadron Collider data:

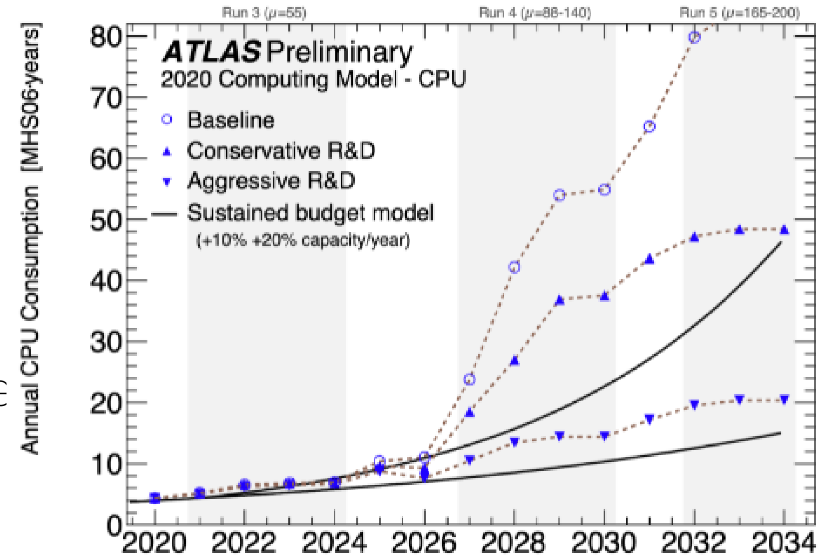
2021: 336 PB

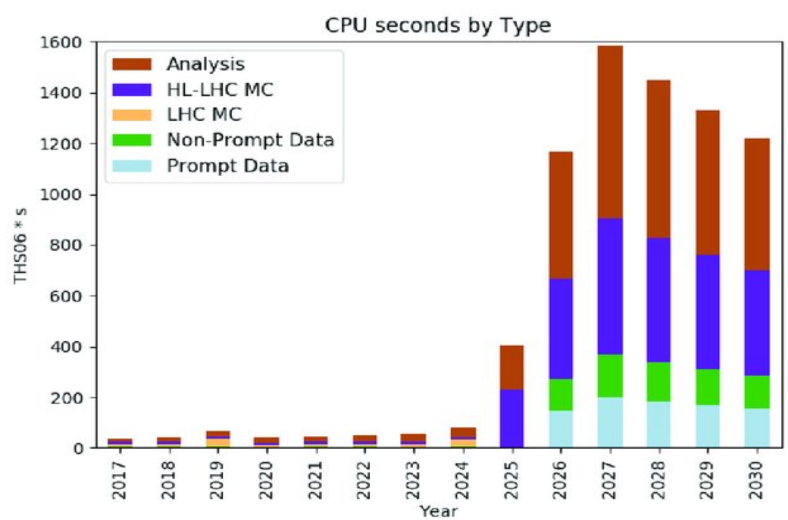
From 2022: 200+ PB/year

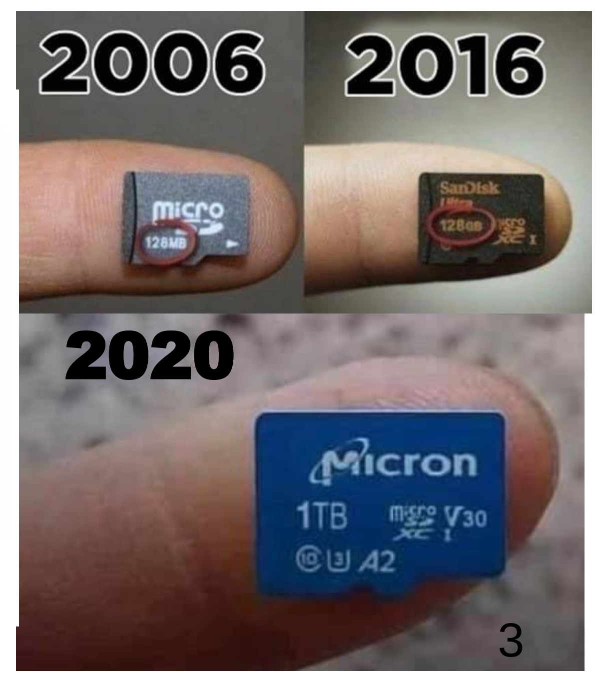
Simulations:

Computationally very expensive

1s LHC data ~ days of CPU time

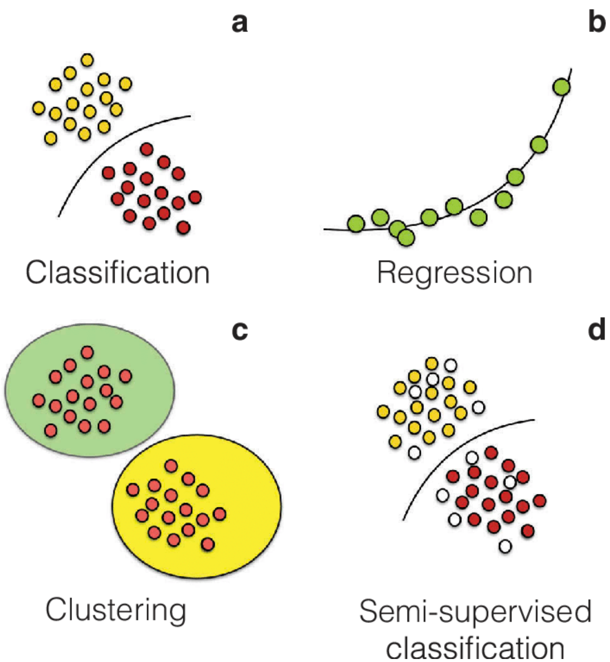


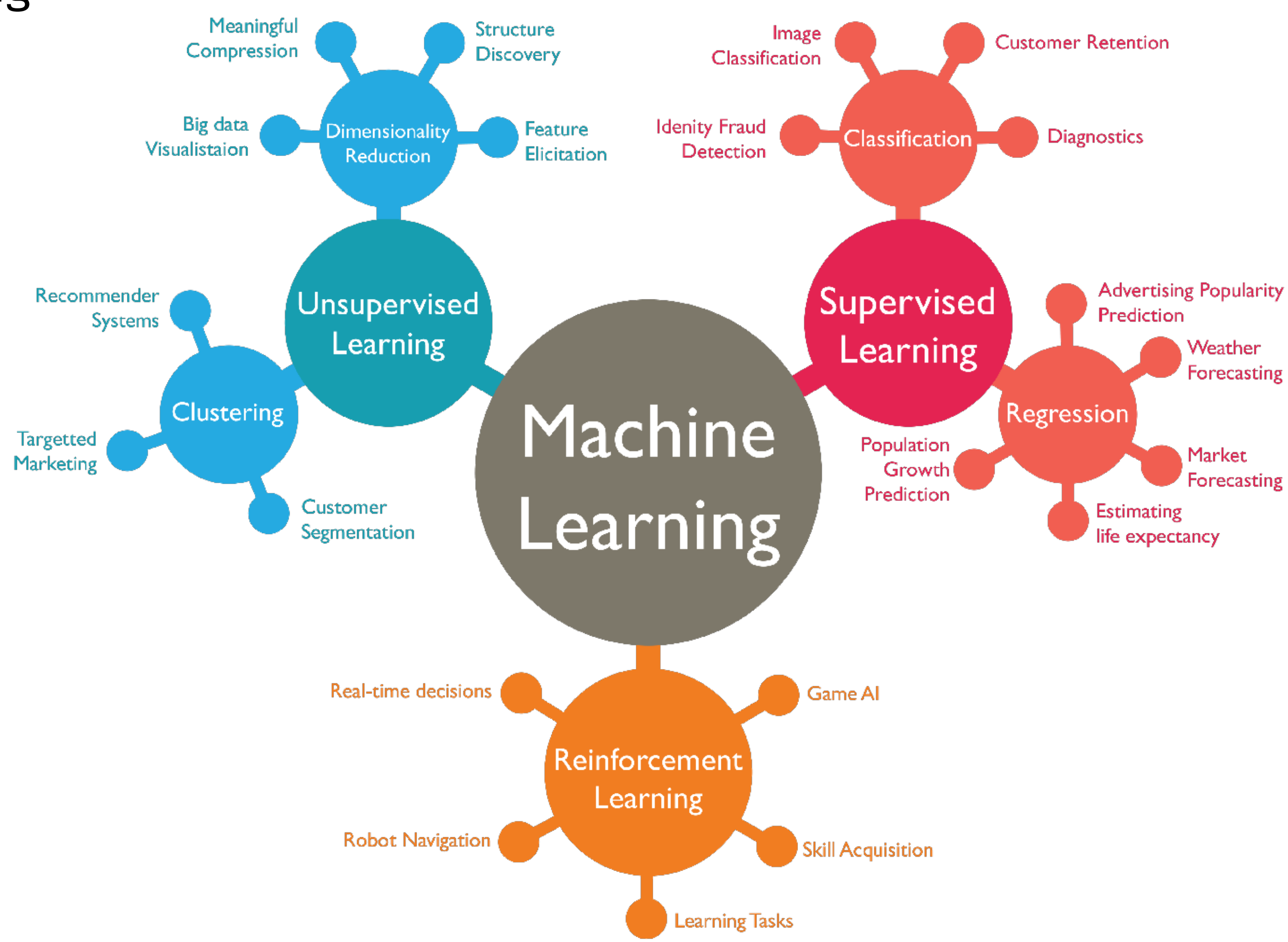




Machine learning

- Data driven decisions
- Automated analysis
- Perform tasks without being explicitly programmed to do so

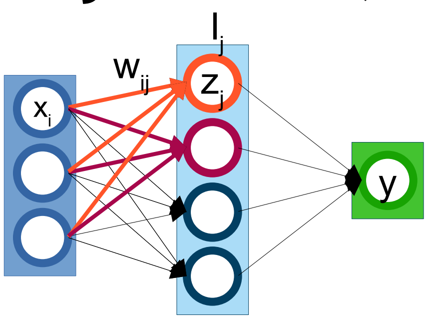




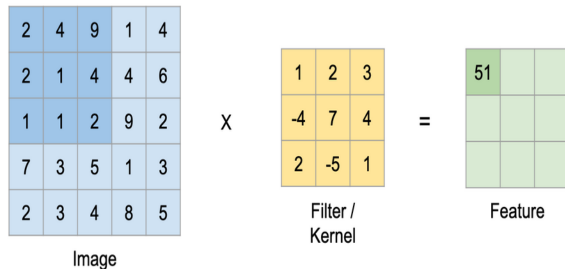
Basic building blocks of a neural network

Linear algebra: $$z_j = \sum_{i=1}^{N} x_i w_{ij} + b_j$$

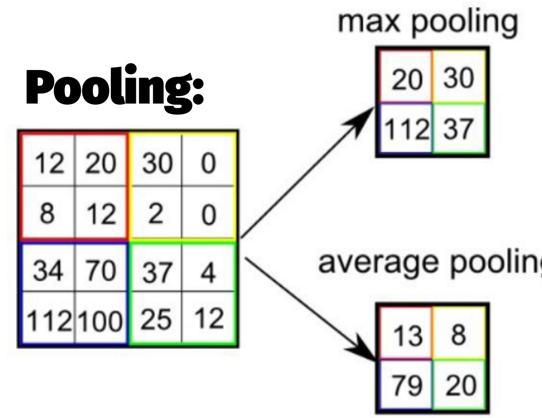
Fully connected (dense):



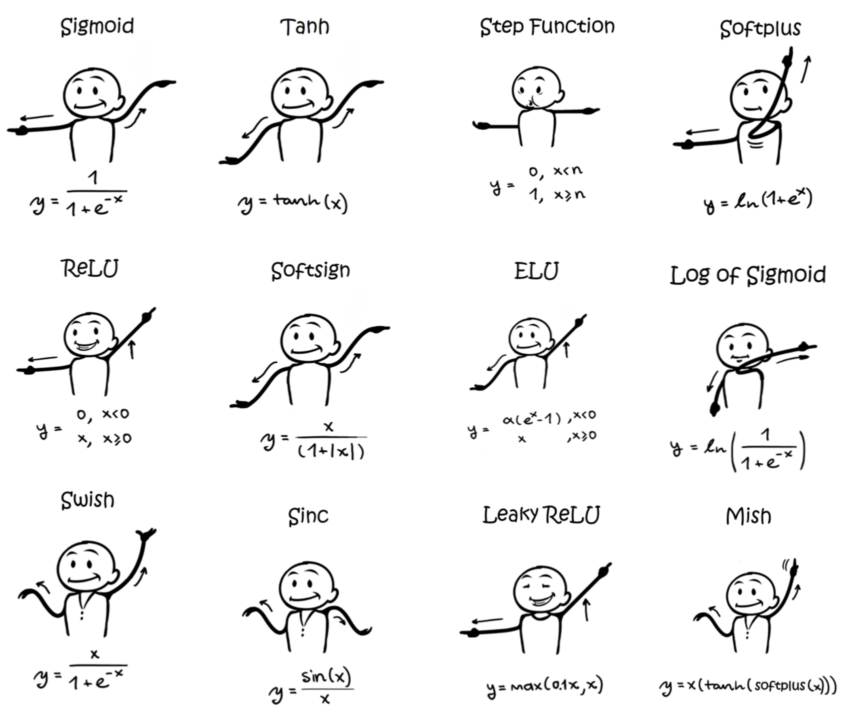
Convolutional:



Pooling:



Activation functions:



https://sefiks.com/2020/02/02/dance-moves-of-deep-learning-activation-functions/

Loss functions, optimizers...

Regression losses

- MeanSquaredError class
- MeanAbsoluteError class
- MeanAbsolutePercentageError class
- MeanSquaredLogarithmicError class
- CosineSimilarity class
- mean_squared_error function
- mean_absolute_error function
- mean_absolute_percentage_error function
- mean_squared_logarithmic_error function
- cosine_similarity function
- Huber class
- huber function
- LogCosh class
- log_cosh function

Probabilistic losses

- BinaryCrossentropy class
- CategoricalCrossentropy class
- SparseCategoricalCrossentropy class
- Poisson class

Hinge losses for "maximum-margin" classification

- Hinge class
- SquaredHinge class
- CategoricalHinge class

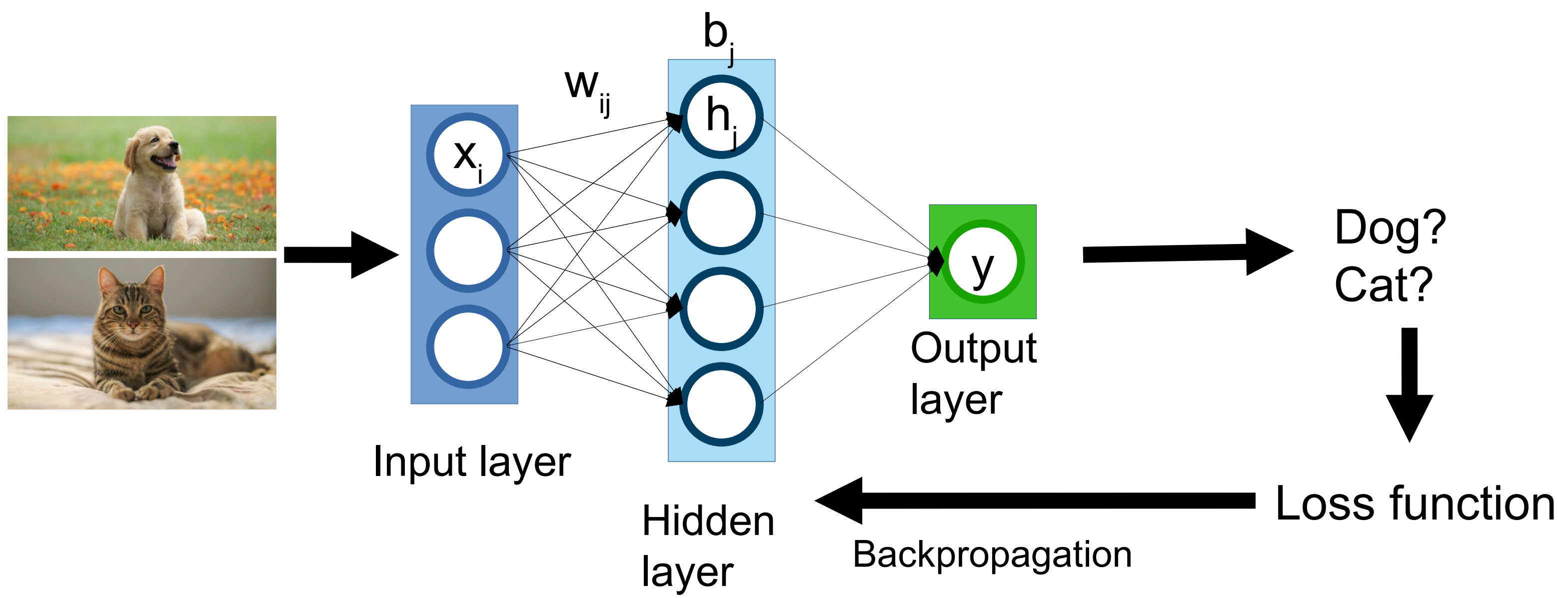
Available optimizers

- SGD
- RMSprop
- Adam
- Adadelta
- Adagrad
- Adamax
- Nadam
- Ftrl

https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

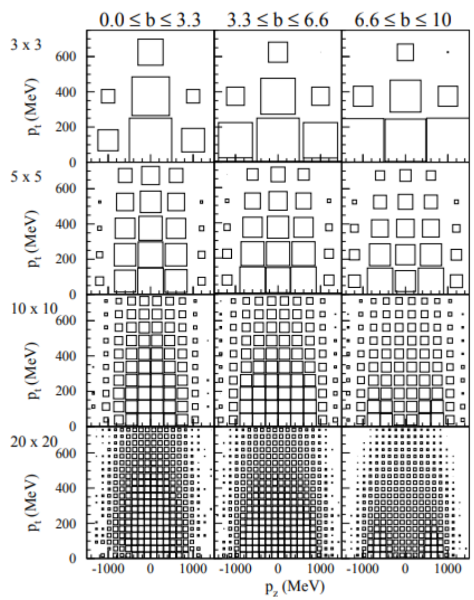
Example: FCNN

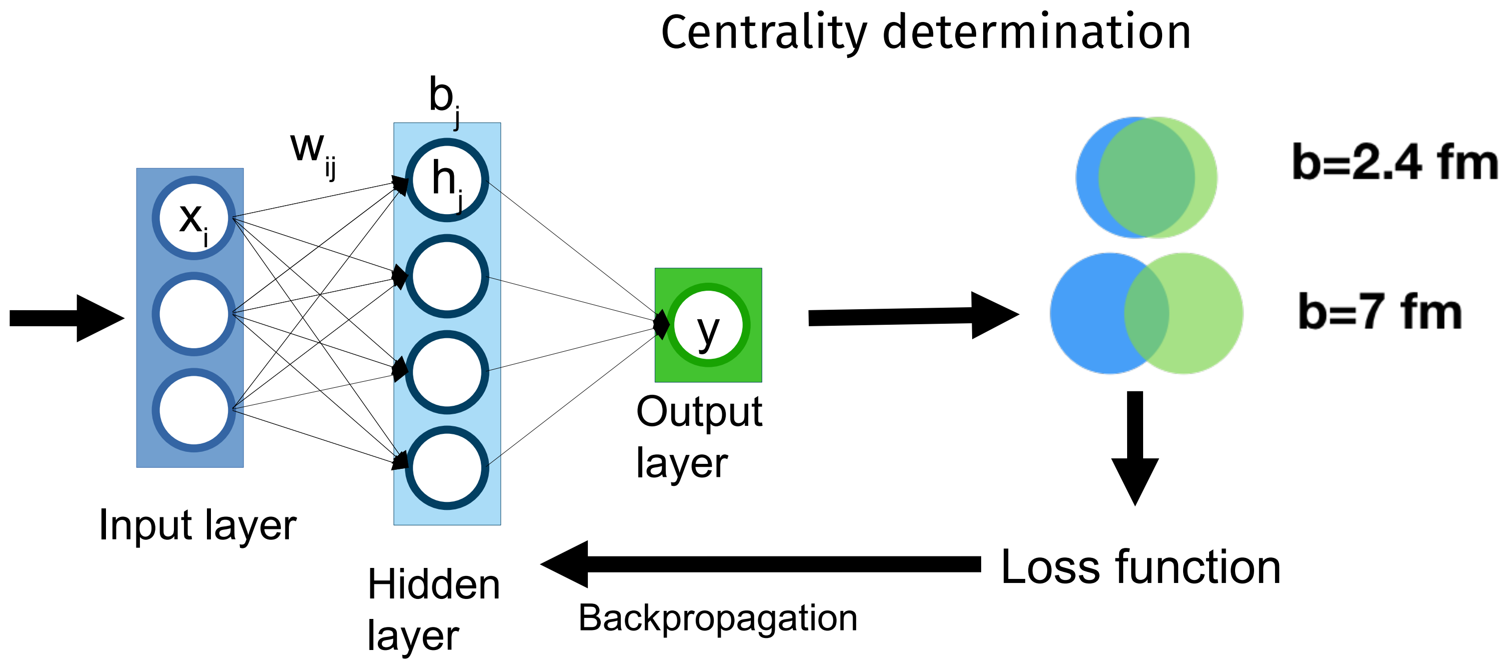
Simple classification



$$L_\delta(y, f(x)) = \begin{cases} \frac{1}{2}(y - f(x))^2, & |y - f(x)| \le \delta \\ \delta(|y - f(x)| - \frac{1}{2}\delta) \end{cases}$$

Example: FCNN





$$L_\delta(y, f(x)) = \begin{cases} \frac{1}{2}(y - f(x))^2, & |y - f(x)| \le \delta \\ \delta(|y - f(x)| - \frac{1}{2}\delta) \end{cases}$$

PRC.53.2358 (1996), Bass, S. A.; Bischoff, A.; Maruhn, J. A.; Stöcker, H.; Greiner, W.

Popular architectures

Classifiers

- AlexNet (Comm. ACM. 60 (6): 84–90, 2012)
- VGG16 (138M parameters, 23 layers, arXiv:1409.1556)
- ResNet (25M+ parameters, arXiv:1512.03385)
- DenseNet (8M parameters, 121 layers, arXiv:1608.06993)

Object detection

- (Fast(er)) R-CNN (arXiv:1311.2524, arXiv:1504.08083, arXiv:1506.01497)
- YOLO (arXiv:1506.02640)
- Detectron (github.com/facebookresearch/detectron2)

Regression

Autonomous vehicles

Decision trees

Transformers

Generative adversarial networks (https://bit.ly/2YMCFdy)

(Variational) autoencoders

...

Machine Learning in HEP

A Living Review of Machine Learning for Particle Physics

https://iml-wg.github.io/HEPML-LivingReview/

Matthew Feickert, Benjamin Nachman, arXiv:2102.02770

2021 May: **417** references

2021 November: **568** references

Today: **629** references

- Track reconstruction
- Quark/gluon jet separation
- Jet reconstruction
- Tuning Monte Carlo event generators
- GAN of detectors
- …

  - Accelerated Charged Particle Tracking with Graph Neural Networks on FPGAs
  - Particle Track Reconstruction using Geometric Deep Learning
  - Jet tagging in the Lund plane with graph networks [DOI]
  - Vertex and Energy Reconstruction in JUNO with Machine Learning Methods
  - MLPF: Efficient machine-learned particle-flow reconstruction using graph neural networks
  - 25th International Conference on Computing in High-Energy and Nuclear Physics
  - 25th International Conference on Computing in High-Energy and Nuclear Physics
  - Graph Neural Network for Object Reconstruction in Liquid Argon Time Projection Chambers
  - Instance Segmentation GNNs for One-Shot Conformal Tracking at the LHC
  - Charged particle tracking via edge-classifying interaction networks
  - Jet characterization in Heavy Ion Collisions by QCD-Aware Graph Neural Networks
  - Graph Generative Models for Fast Detector Simulations in High Energy Physics
  - Segmentation of EM showers for neutrino experiments with deep graph neural networks
- Sets (point clouds)
  - Energy Flow Networks: Deep Sets for Particle Jets [DOI]
  - ParticleNet: Jet Tagging via Particle Clouds [DOI]
  - ABCNet: An attention-based method for particle tagging [DOI]
  - Secondary Vertex Finding in Jets with Neural Networks
  - Equivariant Energy Flow Networks for Jet Tagging
  - Permutationless Many-Jet Event Reconstruction with Symmetry Preserving Attention Networks
  - Zero-Permutation Jet-Parton Assignment using a Self-Attention Network
  - Learning to Isolate Muons
  - Point Cloud Transformers applied to Collider Physics
- Physics-inspired basis
  - Automating the Construction of Jet Observables with Machine Learning [DOI]
  - How Much Information is in a Jet? [DOI]
  - Novel Jet Observables from Machine Learning [DOI]
  - Energy flow polynomials: A complete linear basis for jet substructure [DOI]
  - Deep-learned Top Tagging with a Lorentz Layer [DOI]
  - Resurrecting $b\bar{b}h$ with kinematic shapes
- W/Z tagging
  - Jet-images — deep learning edition [DOI]
  - Parton Shower Uncertainties in Jet Substructure Analyses with Deep Neural Networks [DOI]
  - QCD-Aware Recursive Neural Networks for Jet Physics [DOI]
  - Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques [DOI]
  - Boosted W and Z tagging with jet charge and deep learning [DOI]
  - Supervised Jet Clustering with Graph Neural Networks for Lorentz Boosted Bosons [DOI]
  - Jet tagging in the Lund plane with graph networks [DOI]
  - A $W^\pm$ polarization analyzer from Deep Neural Networks
- $H\rightarrow b\bar{b$}
  - Automating the Construction of Jet Observables with Machine Learning [DOI]
  - Boosting $H\to b\bar b$ with Machine Learning [DOI]
  - Interaction networks for the identification of boosted $H \rightarrow b\overline{b}$ decays [DOI]
  - Interpretable deep learning for two-prong jet classification with jet spectra [DOI]
  - Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques [DOI]
  - Disentangling Boosted Higgs Boson Production Modes with Machine Learning
  - Benchmarking Machine Learning Techniques with Di-Higgs Production at the LHC
  - The Boosted Higgs Jet Reconstruction via Graph Neural Network
  - Extracting Signals of Higgs Boson From Background Noise Using Deep Neural Networks
  - Learning to increase matching efficiency in identifying additional b-jets in the $\text{t}\bar{\text{t}}\text{b}\bar{\text{b}}$ process
- quarks and gluons
  - Quark versus Gluon Jet Tagging Using Jet Images with the ATLAS Detector
  - Deep learning in color: towards automated quark/gluon [DOI]
  - Recursive Neural Networks in Quark/Gluon Tagging [DOI]
  - DeepJet: Generic physics object based jet multiclass classification for LHC experiments
  - Probing heavy ion collisions using quark and gluon jet substructure
  - JEDI-net: a jet identification algorithm based on interaction networks [DOI]
  - Quark-Gluon Tagging: Machine Learning vs Detector [DOI]
  - Towards Machine Learning Analytics for Jet Substructure [DOI]
  - Quark Gluon Jet Discrimination with Weakly Supervised Learning [DOI]

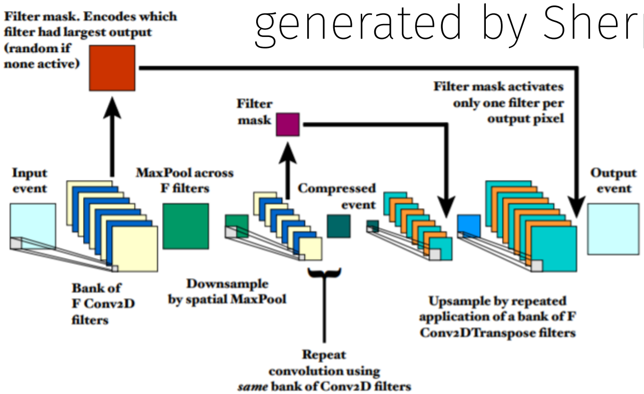
- Classification
  - Parameterized classifiers
    - Parameterized neural networks for high-energy physics [DOI]
    - Approximating Likelihood Ratios with Calibrated Discriminative Classifiers
    - E Pluribus Unum Ex Machina: Learning from Many Collider Events at Once
    - Jet images
      - How to tell quark jets from gluon jets
      - Jet-Images: Computer Vision Inspired Techniques for Jet Tagging [DOI]
      - Playing Tag with ANN: Boosted Top Identification with Pattern Recognition [DOI]
      - Jet-images — deep learning edition [DOI]
      - Quark versus Gluon Jet Tagging Using Jet Images with the ATLAS Detector
      - Boosting $H\to b\bar b$ with Machine Learning [DOI]
      - Learning to classify from impure samples with high-dimensional data [DOI]
      - Parton Shower Uncertainties in Jet Substructure Analyses with Deep Neural Networks [DOI]
      - Deep learning in color: towards automated quark/gluon [DOI]
      - Deep-learning Top Taggers or The End of QCD? [DOI]
      - Pulling Out All the Tops with Computer Vision and Deep Learning [DOI]
      - Reconstructing boosted Higgs jets from event image segmentation
      - An Attention Based Neural Network for Jet Tagging
      - Quark-Gluon Jet Discrimination Using Convolutional Neural Networks [DOI]
      - Learning to Isolate Muons
      - Deep learning jet modifications in heavy-ion collisions
    - Event images
      - Topology classification with deep learning to improve real-time event selection at the LHC [DOI]
      - Convolutional Neural Networks with Event Images for Pileup Mitigation with the ATLAS Detector
      - Boosting $H\to b\bar b$ with Machine Learning [DOI]
      - End-to-End Physics Event Classification with the CMS Open Data: Applying Image-based Deep Learning on Detector Data to Directly Classify Collision Events at the LHC [DOI]
      - Disentangling Boosted Higgs Boson Production Modes with Machine Learning
      - Identifying the nature of the QCD transition in relativistic collision of heavy nuclei with deep learning [DOI]
    - Sequences
      - Jet Flavor Classification in High-Energy Physics with Deep Neural Networks [DOI]
      - Topology classification with deep learning to improve real-time event selection at the LHC [DOI]
      - Jet Flavour Classification Using DeepJet [DOI]
      - Development of a Vertex Finding Algorithm using Recurrent Neural Network
      - Sequence-based Machine Learning Models in Jet Physics
    - Trees
      - QCD-Aware Recursive Neural Networks for Jet Physics [DOI]
      - Recursive Neural Networks in Quark/Gluon Tagging [DOI]
    - Graphs
      - Neural Message Passing for Jet Physics
      - Graph Neural Networks for Particle Reconstruction in High Energy Physics detectors
      - Probing stop pair production at the LHC with graph neural networks [DOI]
      - Pileup mitigation at the Large Hadron Collider with graph neural networks [DOI]
      - Unveiling CP property of top-Higgs coupling with graph neural networks at the LHC [DOI]
      - JEDI-net: a jet identification algorithm based on interaction networks [DOI]
      - Learning representations of irregular particle-detector geometry with distance-weighted graph networks [DOI]
      - Interpretable deep learning for two-prong jet classification with jet spectra [DOI]
      - Neural Network-based Top Tagger with Two-Point Energy Correlations and Geometry of Soft Emissions [DOI]
      - Probing triple Higgs coupling with machine learning at the LHC
      - Casting a graph net to catch dark showers [DOI]
      - Graph neural networks in particle physics [DOI]
      - Distance-Weighted Graph Neural Networks on FPGAs for Real-Time Particle Reconstruction in High Energy Physics [DOI]
      - Supervised Jet Clustering with Graph Neural Networks for Lorentz Boosted Bosons [DOI]
      - Track Seeding and Labelling with Embedded-space Graph Neural Networks
      - Graph neural network for 3D classification of ambiguities and optical crosstalk in scintillator-based neutrino detectors [DOI]

Parton shower and hadronization

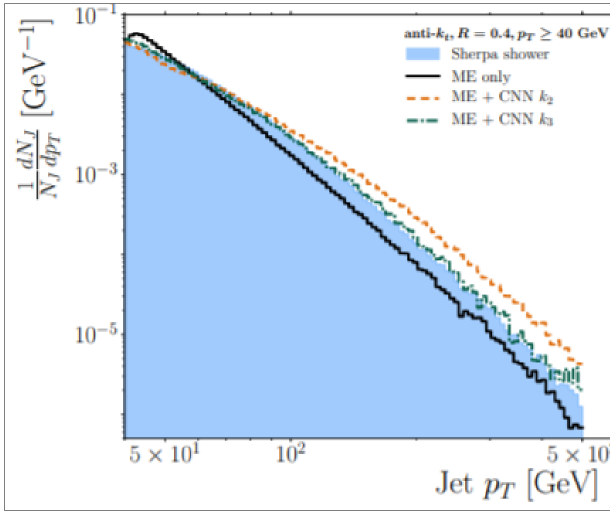
The goal of this study

J.W. Monk: Deep Learning as a Parton Shower (arXiv:1807.03685)

Dataset: 500 000 QCD pp event @ 7 TeV, generated by Sherpa



parameter	model k_2	model k_3
Kernel size, k	2	3
Input image size, N	64	81
Size of filter bank, F	9	7
Levels of decomposition	5	3
Regularisation, λ	500	300
Learning rate	5×10^{-5}	1×10^{-5}
Loss weight w_1	5	4
Loss weight w_2	2	2
Loss weight w_3	1	1
Total number of trained weights	72	126

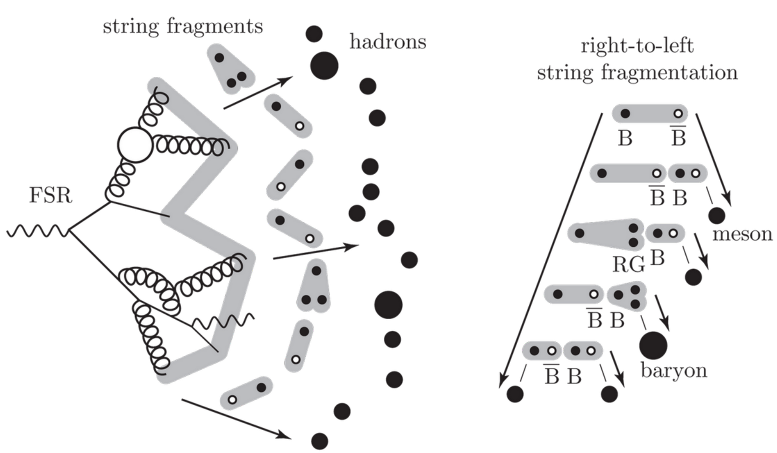


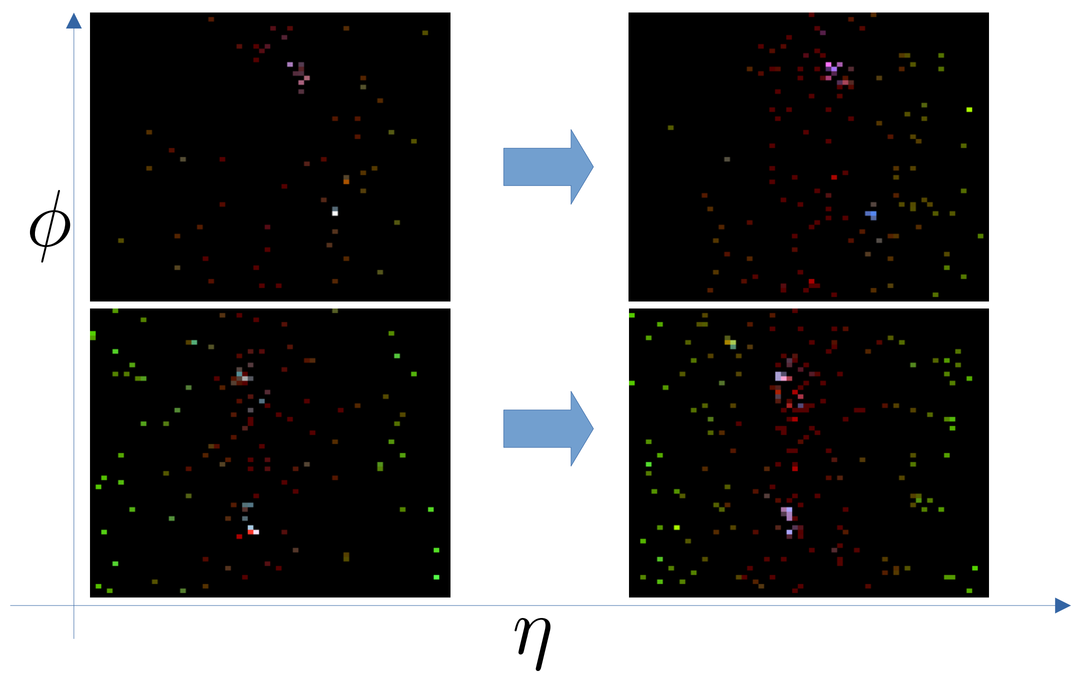
Hadronization

Partons → hadrons

Non-perturbative process

Lund-fragmentation (Comput.Phys.Commun. 27 (1982) 243)





Train and validation sets

Monte Carlo data: Pythia 8.303

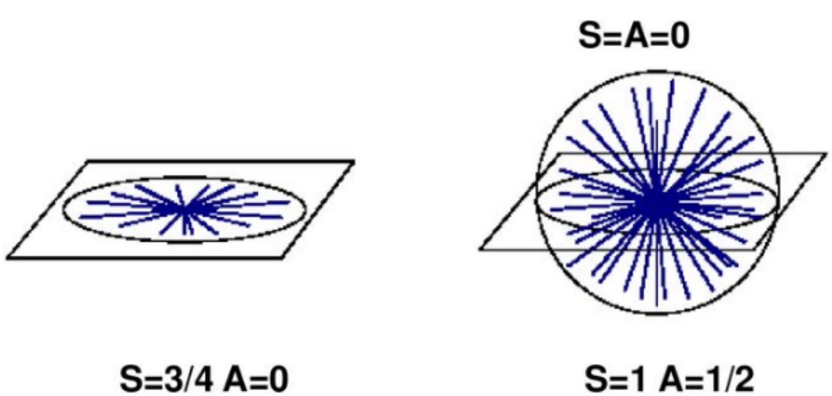
Monash tune

Selection:

- All final particles with $|y| < \pi$
- At least 2 jets
  - Anti-k_T
  - R=0.6
  - p_T>40 GeV

Event number:

- Train: 150 000
- Validation: 150 000
- ~20 GB raw data



Input:

Parton level

Discretized in the (y, ϕ) plane: p_x, p_y, p_z, E, m, multiplicity

$y \in [\pi, \pi]$, 62 bins

$\phi \in [0, 2\pi]$ 31 bins

Hadron level output:

(Charged) event multiplicity, (tr-)sphericity, jet p_T, -mass, -width, -multiplicity

$$M_{xyz} = \sum_i \begin{pmatrix} p_{xi}^2 & p_{xi}p_{yi} & p_{xi}p_{zi} \\ p_{yi}p_{xi} & p_{yi}^2 & p_{yi}p_{zi} \\ p_{zi}p_{xi} & p_{zi}p_{yi} & p_{zi}^2 \end{pmatrix}$$

Eigenvalues: $\lambda_1 > \lambda_2 > \lambda_3$ $\sum_i \lambda_i = 1$

Sphericity: $S = \frac{3}{2}(\lambda_2 + \lambda_3)$

Transverse sphericity: $S_\perp = \frac{2\lambda_2}{\lambda_1 + \lambda_2}$

Models

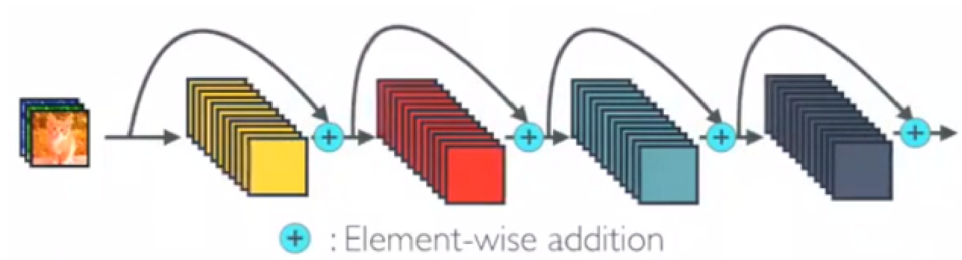
Stacking more layers: solve complex problems more efficiently, get highly accurate results

BUT:

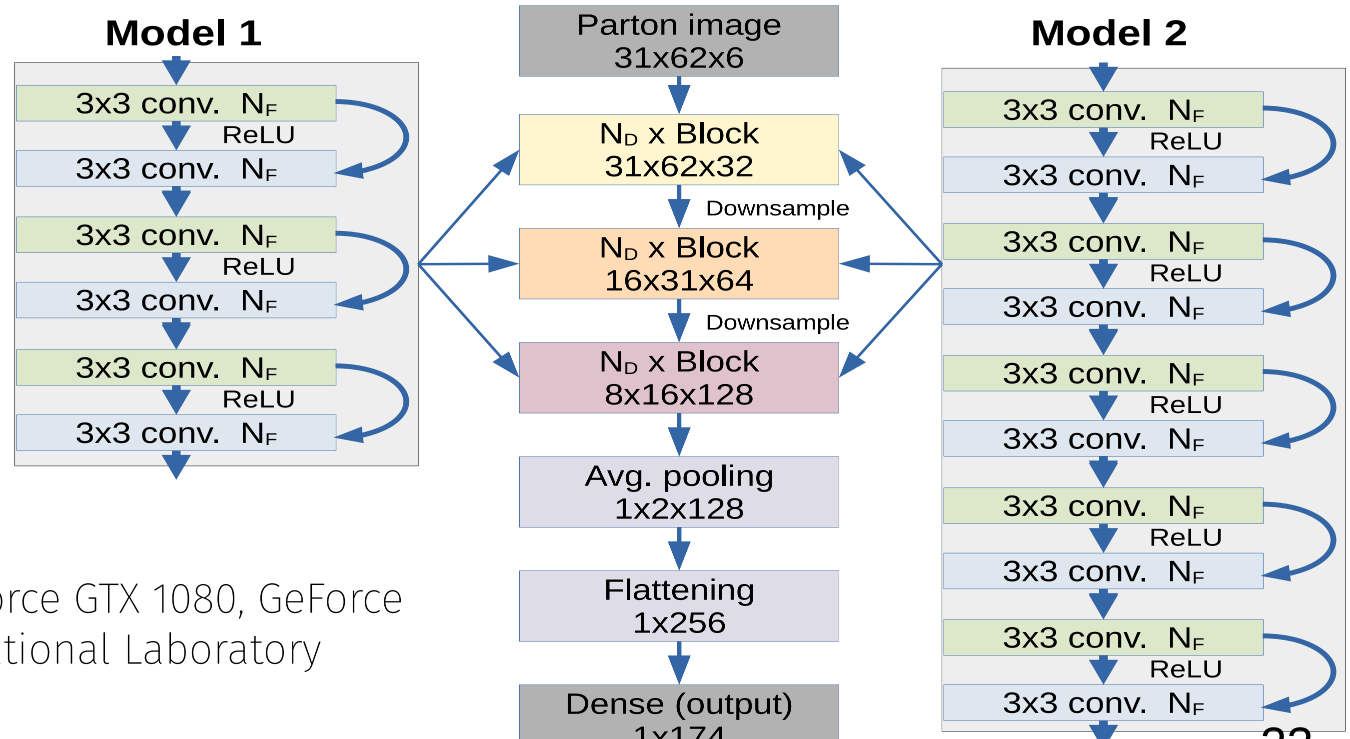
Vanishing/exploding gradients (not to confuse with overfitting)

ResNet

Residual blocks with “skip connections”



	Model 1	Model 2
Trainable parameters	1.13 M	1.90 M



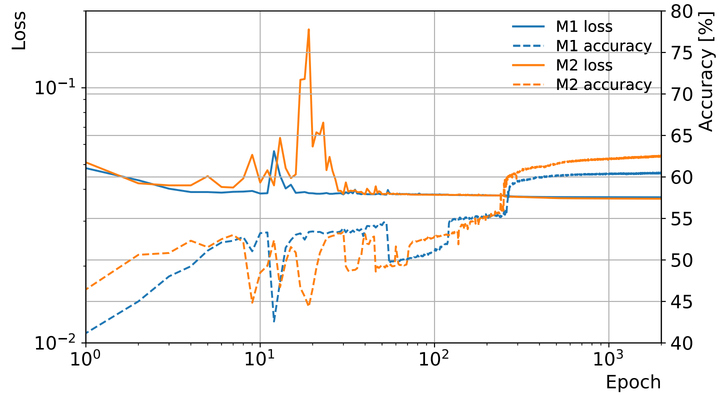
Used hardwares: Nvidia Tesla T4, GeForce GTX 1080, GeForce GTX 980 @ Wigner Scientific Computational Laboratory

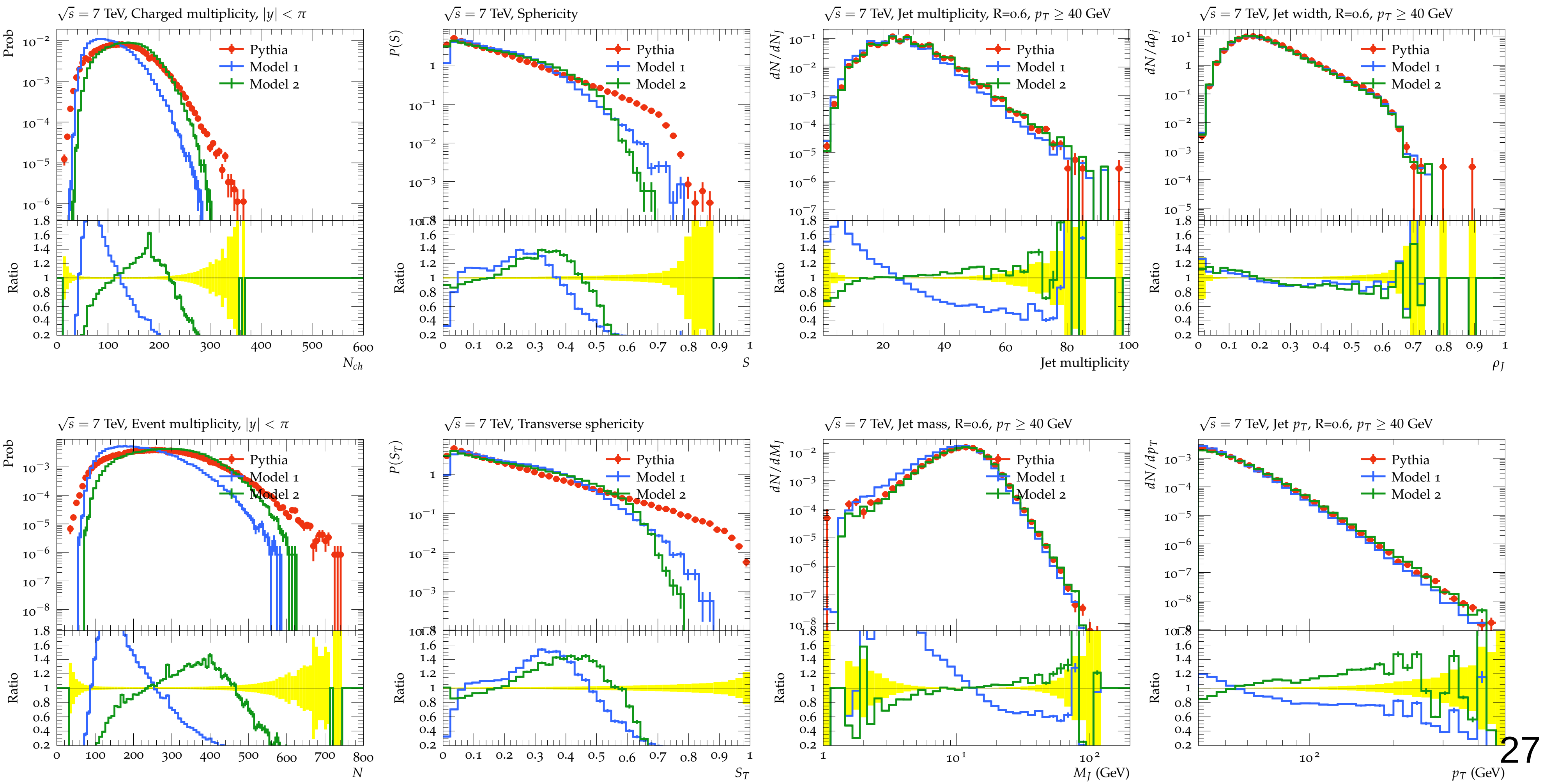
Framework: Tensorflow 2.4.1, Keras 2.4.0

Results

Proton-proton @ 7 TeV, Training + Validation

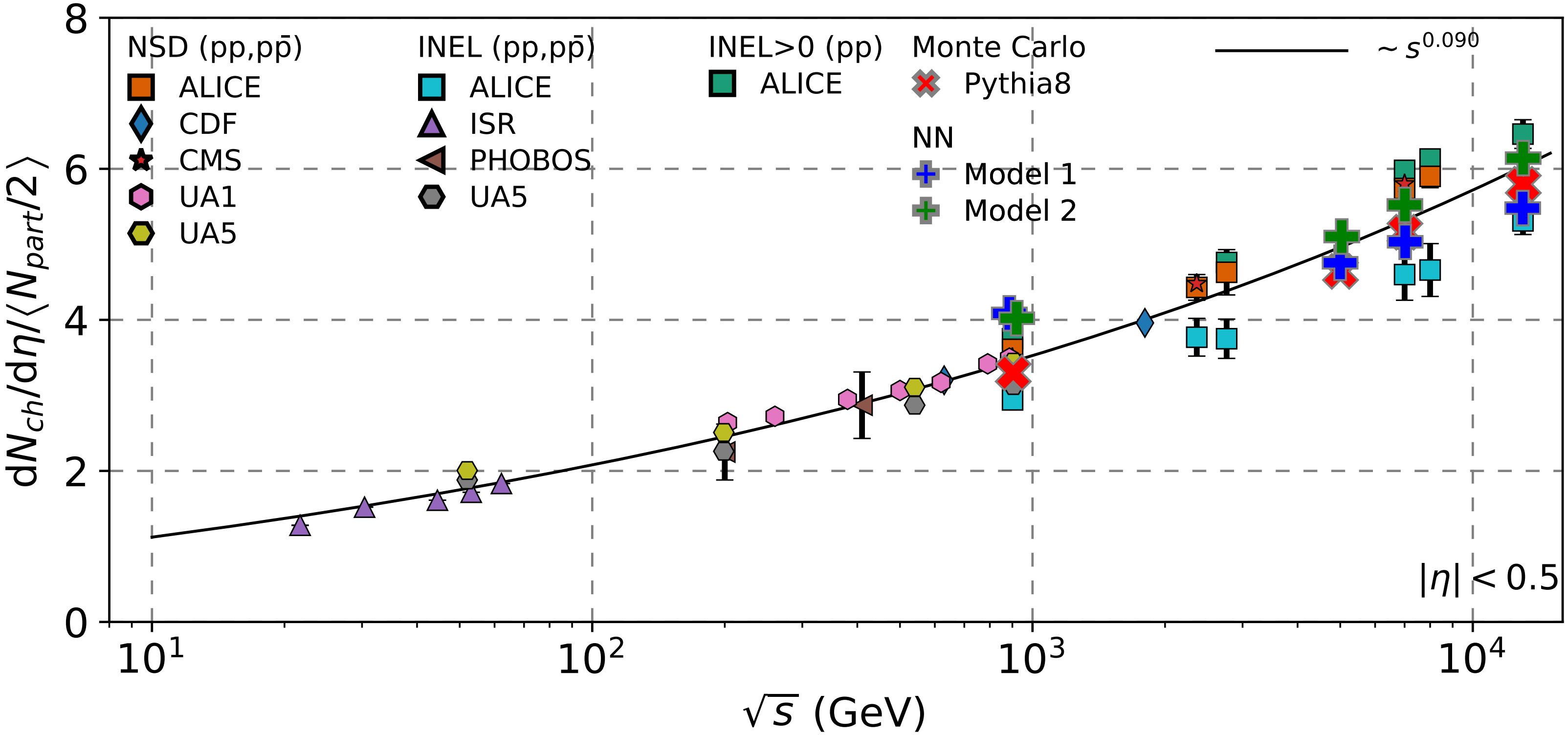
	Model 1	Model 2
Trainable parameters	1.13 M	1.90 M





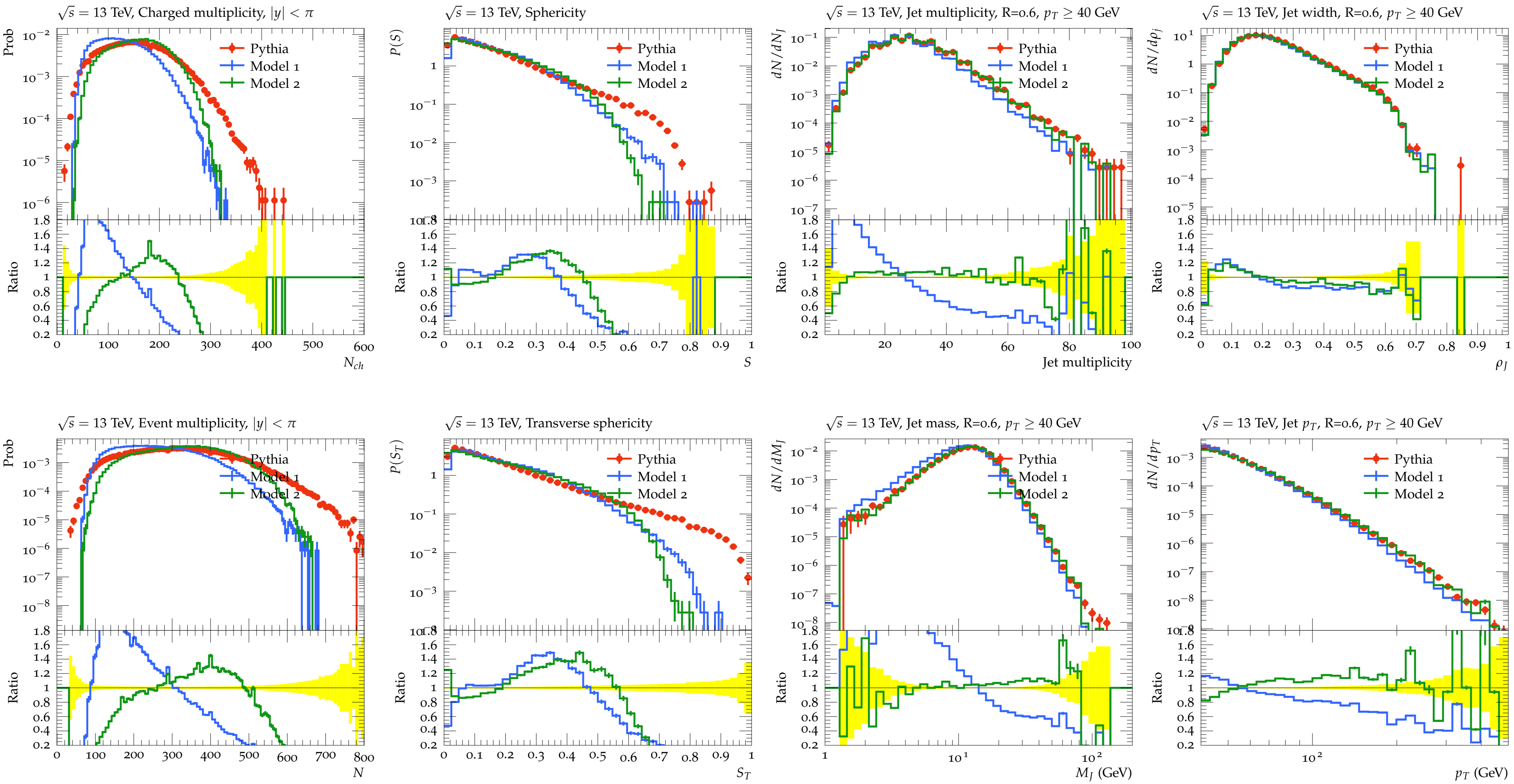
Center-of-mass energy scaling

	Model 1	Model 2
Trainable parameters	1.13 M	1.90 M



Prediction at other CM energies

$$\sqrt{s} = 13\ \mathrm{TeV}$$



How to do it more effectively?
(WIP)

Dimensionality

Input:

Parton level

Discretized in the (y, ϕ) plane: p_x, p_y, p_z, E, m, multiplicity

$$\left.\begin{matrix} y \in [\pi, \pi], & \text{62 bins} \\ \phi \in [0, 2\pi] & \text{31 bins} \end{matrix}\right\} := M$$

$\mathcal{O}(10^3 - 10^4)$ Total pixels

vs

$\mathcal{O}(10^2)$ Pixels with information

Reduction with Singular Value Decomposition:

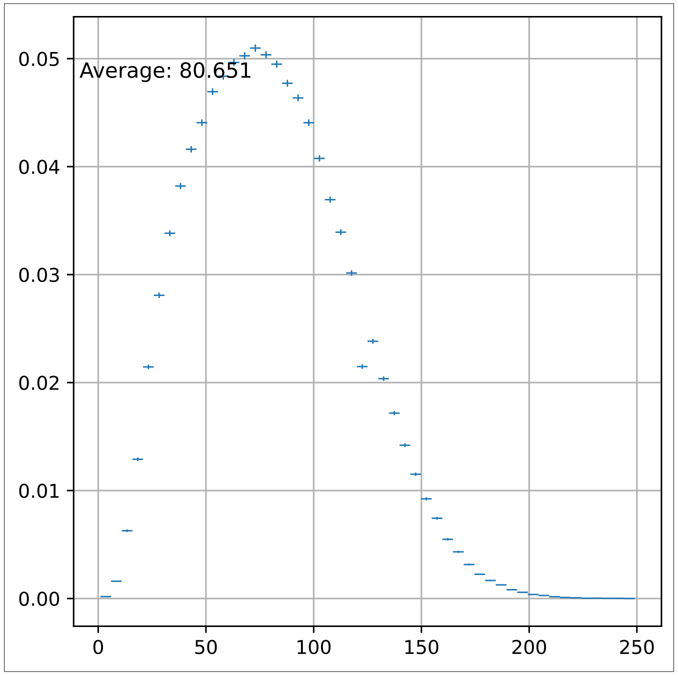
$$M_{n\times m} = U_{n\times n}\Sigma_{n\times m}V^T_{m\times m}$$

- Unitarity
- Ordered by importance
- Guaranteed to exist, unique

$$M \approx \sum_{i=1}^{r} \sigma_i u_i v_i^T + \mathcal{O}(\epsilon), \quad r \leq \min\{n, m\}$$

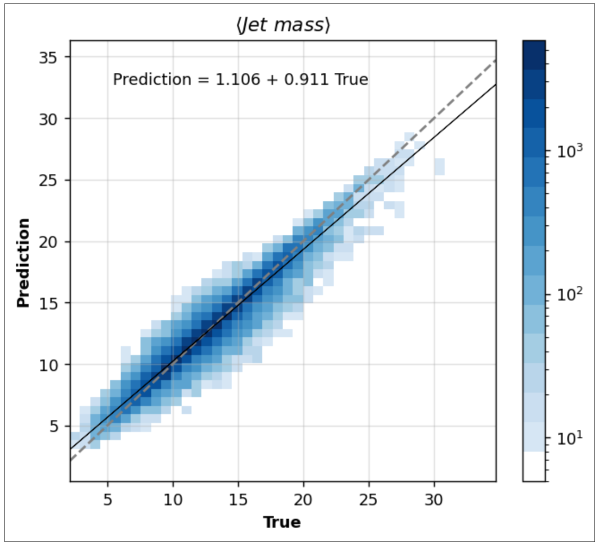
Reduce the input to $\mathcal{O}(10^2)$

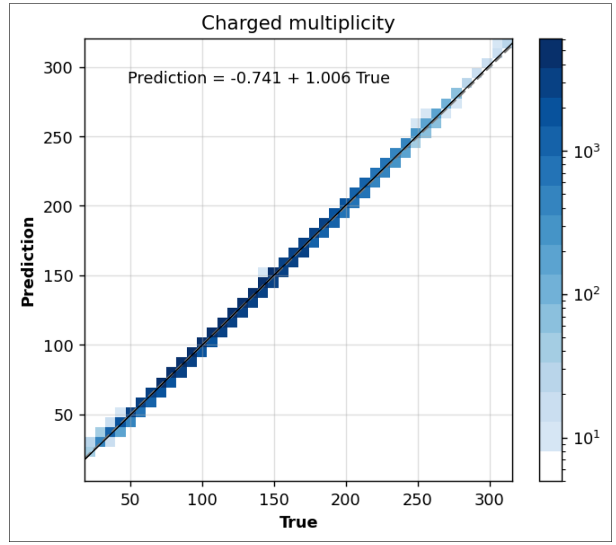
doi:10.1007/BF02288367

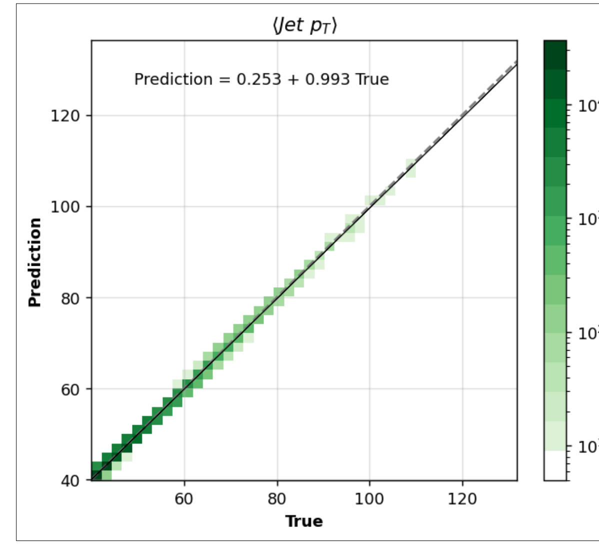


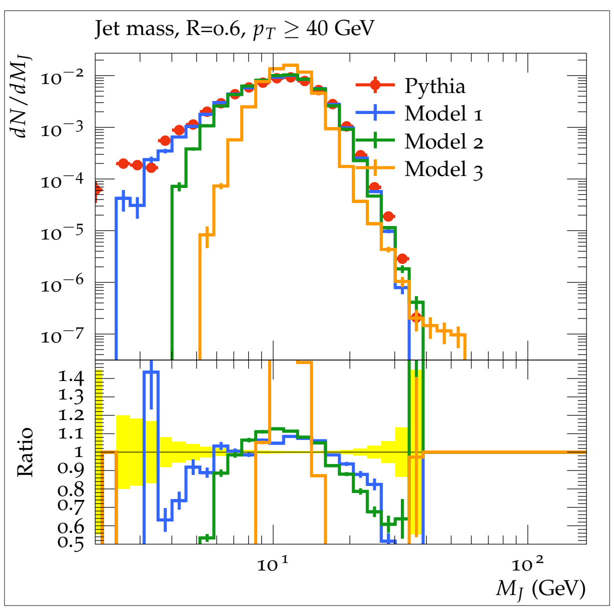
Data-Driven Science and Engineering (S. L. Brunton, J. N. Kutz)

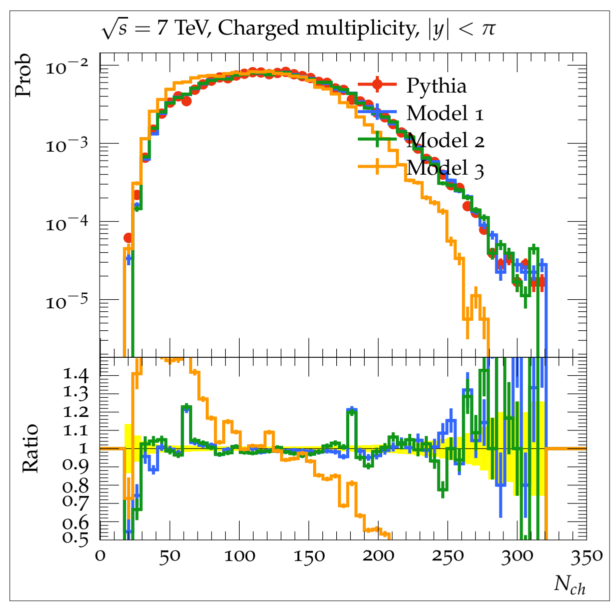
Dimensionality (work in progress)

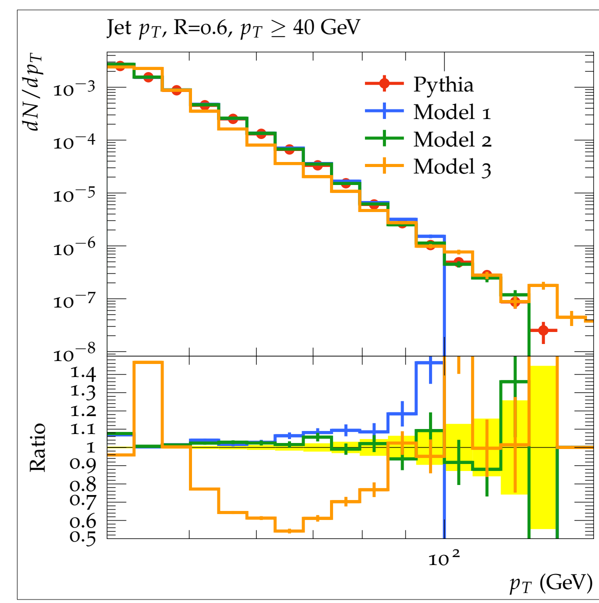












Summary

Traditional computer vision algorithms capture the main features of high-energy event variables successfully

Generalization to other CM energies: multiplicity scaling

Prospects

Dimensional reduction

Various architectures (hyperparameter fine-tuning)

Other observables (p_T, rapidity, particle species)

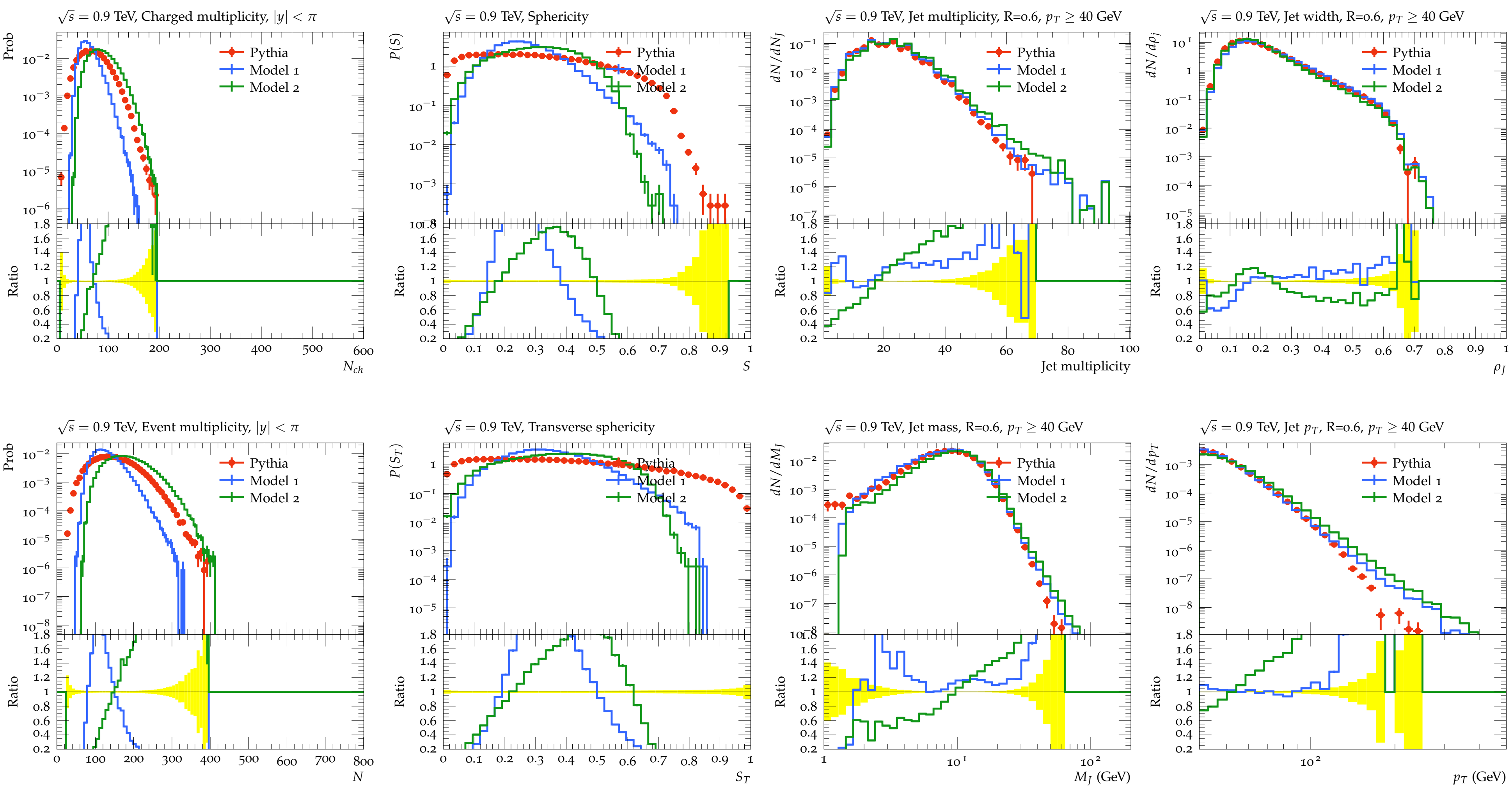
Heavy ion (centralities, collective effects)

Thank you for your attention!

The research was supported by OTKA grants K135515, K123815, NKFIH 2019-2.1.6-NEMZKI-2019-00011, 2020-2.1.1-ED-2021-00179, the **Wigner Scientific Computational Laboratory** (former Wigner GPU Laboratory) and NKFIH within the framework of the MILAB Artificial Intelligence National Laboratory Program.

Prediction at other CM energies

$$\sqrt{s} = 900\ \text{GeV}$$



Prediction at other CM energies

$$\sqrt{s} = 5.02\ \mathrm{TeV}$$