Machine Learning for HEP

Long-Gang Pang (庞龙刚)

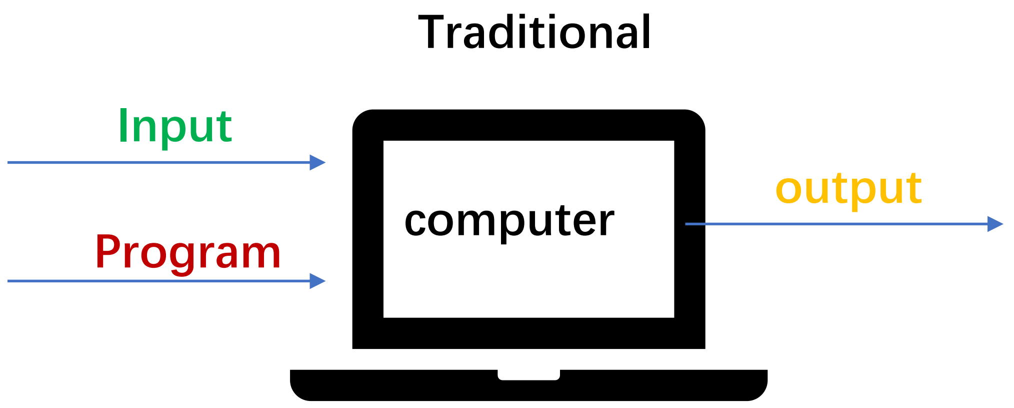
Central China Normal University

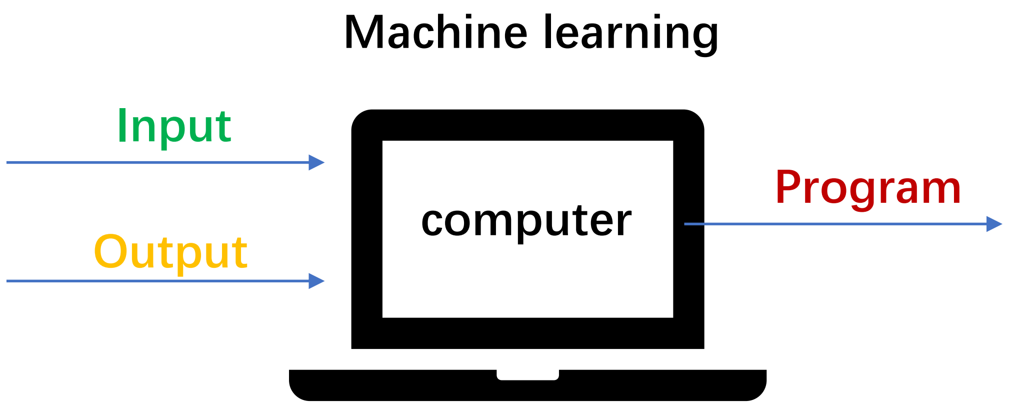
lgpang@mail.ccnu.edu.cn

第六届中国LHC物理会议

What is machine learning

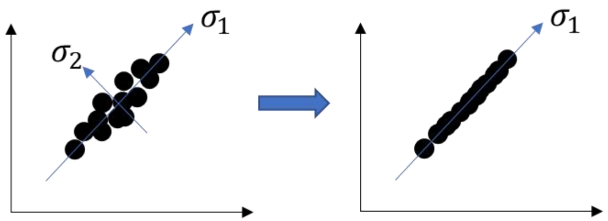
- **A collection of algorithms (PCA, SVM, Random Forest, Boosting Trees, Neural Network ⋯) that let computer learn patterns by themselves.**
- **Keywords: Data driven; Functional; Optimize; Software 2.0;**
- Minimize $loss[f(x,\theta),y] \rightarrow f$



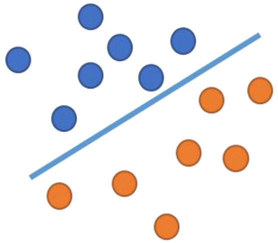


Applications of machine learning

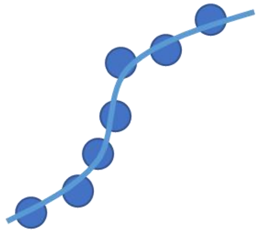
Dimensionality reduction(PCA, tSNE)



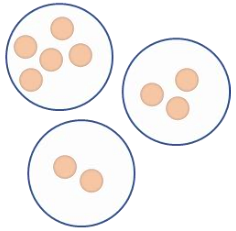
Classification



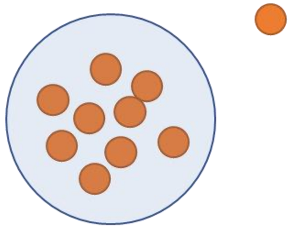
Regression



Clustering



Anomaly detection



Generation

Train with GAN, VAE or Flow model

Generate with given conditions

5

Examples from: https://make.girls.moe/

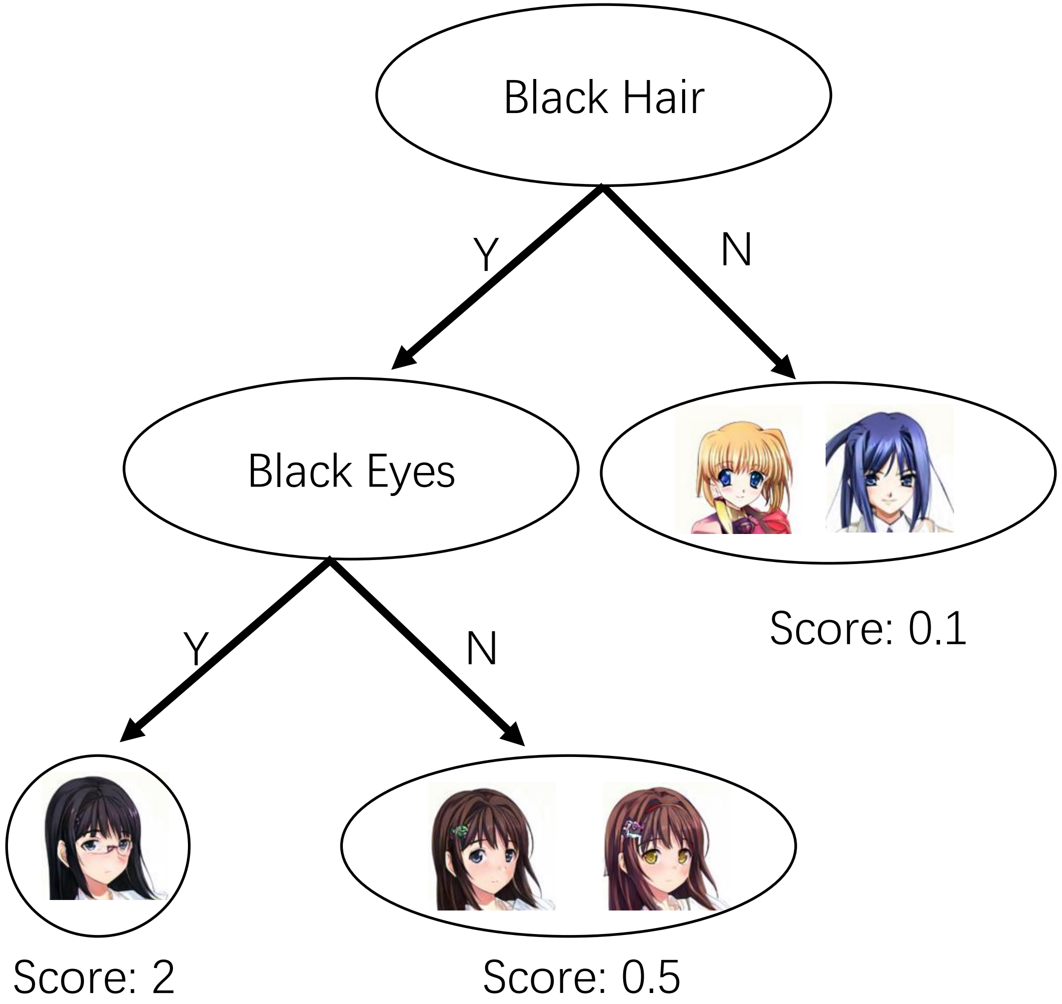
ML for HEP

- **In May 2014, ATLAS held Kaggle competition: Higgs Boson Machine Learning Challenge**
- **Goal: distinguish Higgs signal from exotic background**
- **The winner uses ensemble of neural networks**
- **In this competition, TianQi Chen and Tong He developed XGBoost, which became the most popular ML tool on Kaggle!**
- **Boosted trees and deep neural network are the most frequently used ML tools in HEP.**

Higgs Boson Discovery with Boosted Trees. TQ Chen and T He, HEPML 2014

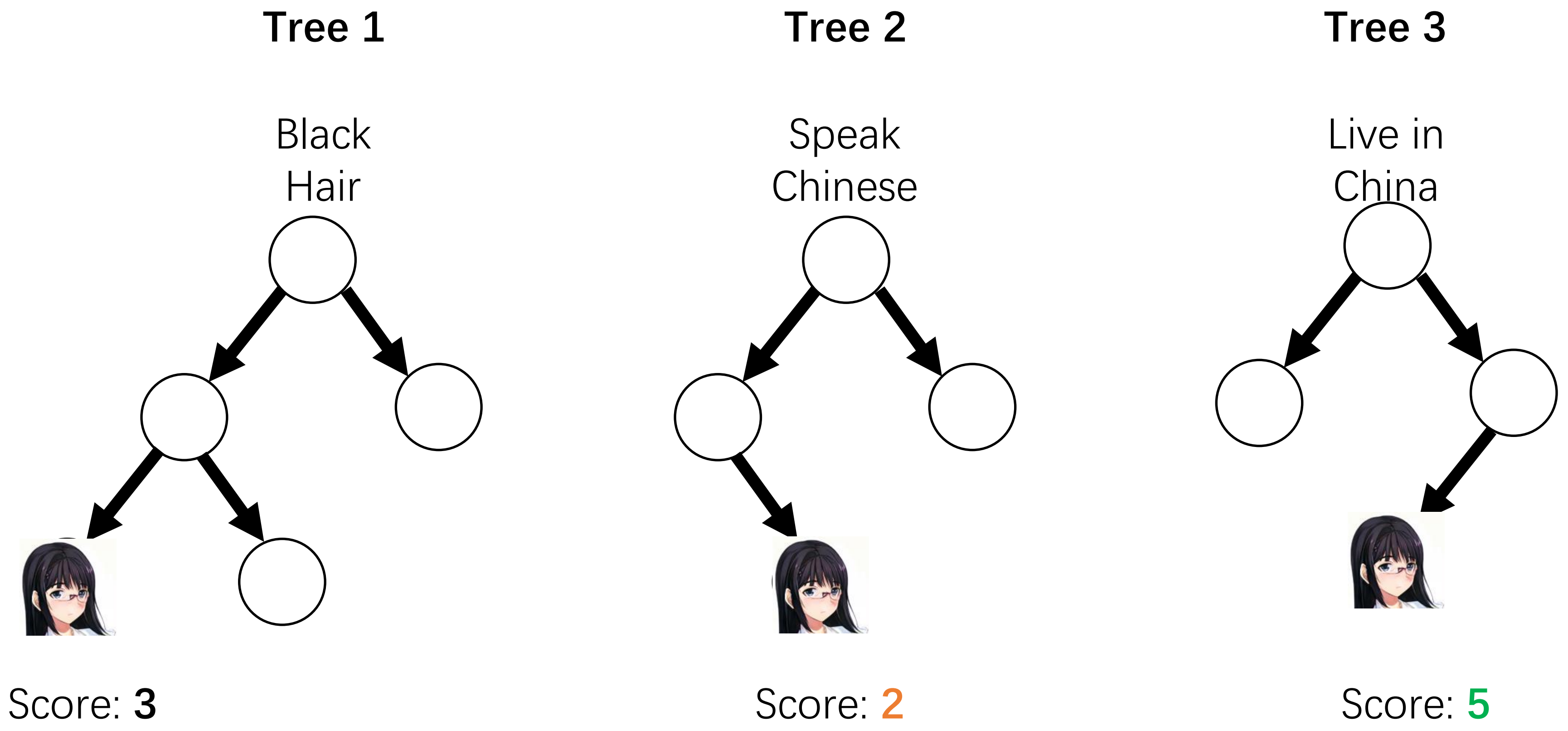
A single decision tree

Task: Asian girl?



Splitting nodes are chosen to minimize the MSE, entropy or Gini factor.

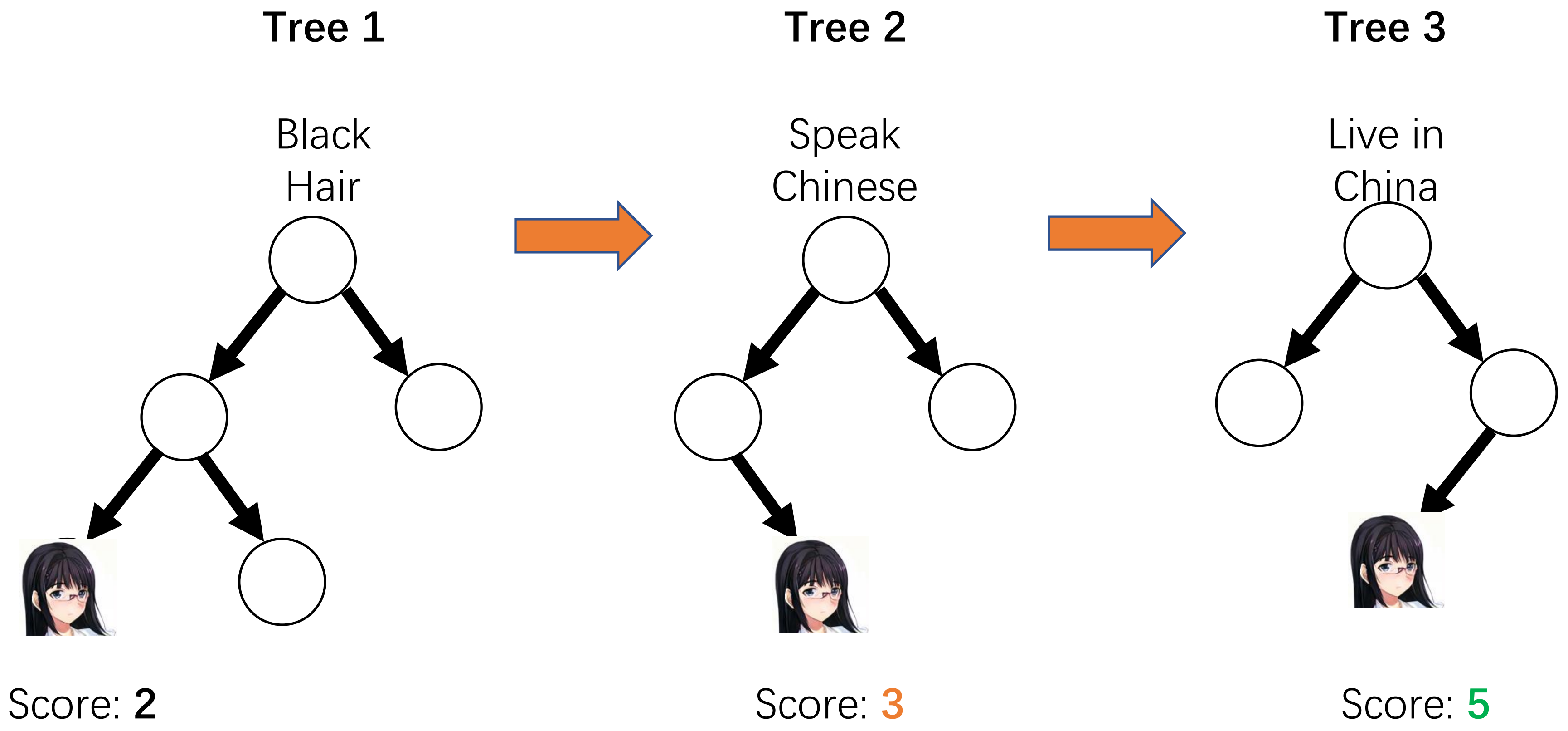
Ensemble of trees: random forest (in parallel)



IsAsian() = 3 + 2 + 5

Low variance

Ensemble: boosted decision tree (in sequential)

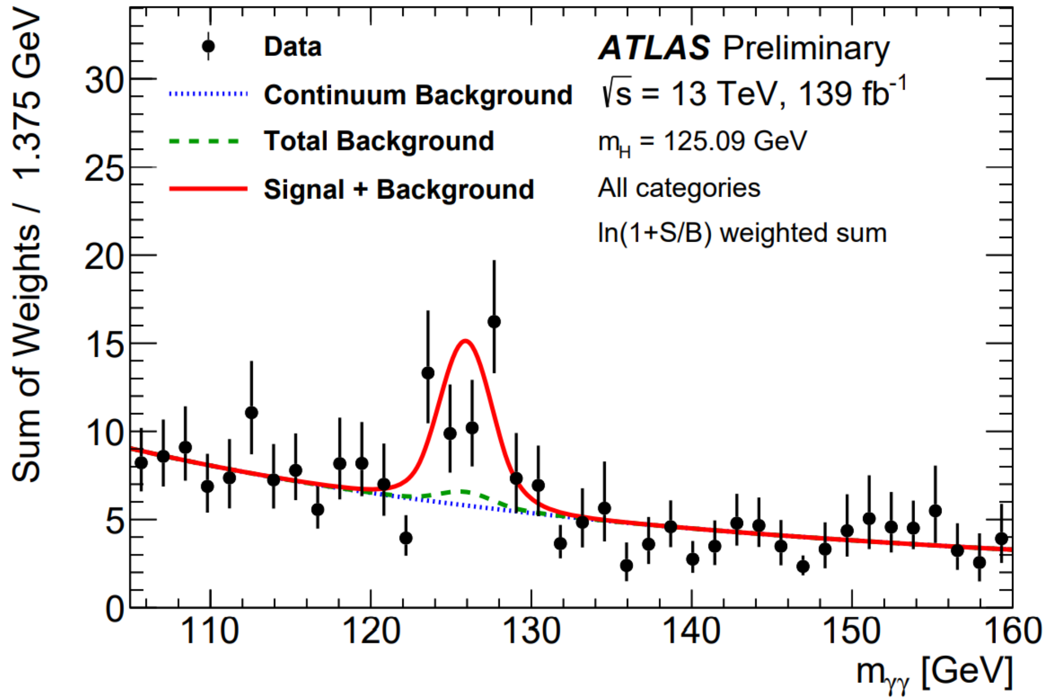


Low bias.

Improve the tree by training residual of the previous tree.

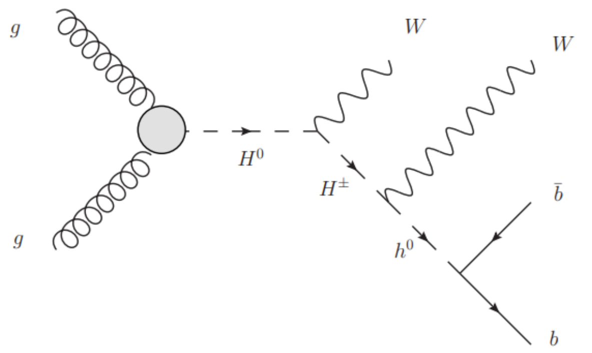
$t\bar{t}H$ identification using boosted decision tree (BDT)

- **Motivation: the coupling between the heaviest SM particle top quark and Higgs is important for tree level top Yukawa coupling.**
- **ATLAS use XGBoost to look for top associated Higgs.**
- **Signal:** $t\bar{t}H, H \to \gamma\gamma$
- **Noise:** $t\bar{t}\gamma\gamma$, **continuous gamma and non** $t\bar{t}H$ **Higgs events.**

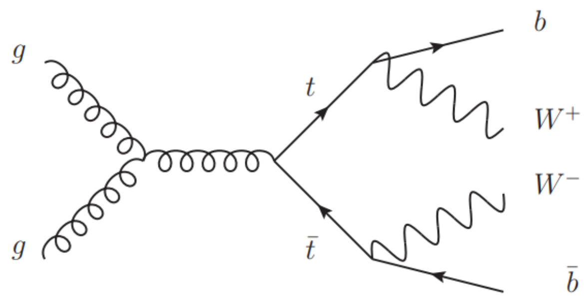


[ATLAS-CONF-2019-004]

Higgs identification using deep learning



Signal



Background

	AUC		
Technique	Low-level	High-level	Complete
BDT	0.73 (0.01)	0.78 (0.01)	0.81 (0.01)
NN	0.733 (0.007)	0.777 (0.001)	0.816 (0.004)
DN	0.880 (0.001)	0.800 ($<$ 0.001)	0.885 (0.002)

"*Our analysis shows that **recent advances in deep learning techniques may lift these limitations by automatically discovering powerful non-linear feature combinations** and providing better discrimination power than current classifiers – even when aided by manually-constructed features.*"

P.Baldi,P.Sadowski,& D.Whiteson, Nature Commun.5, 4308 (2014)

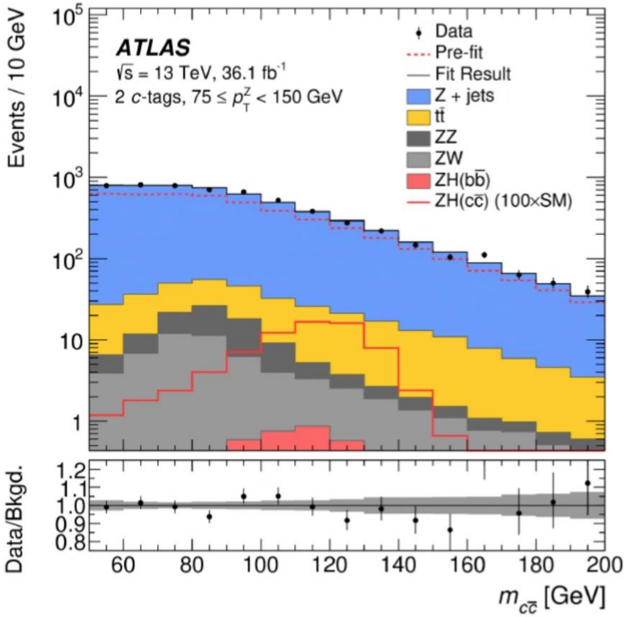
BDT vs DeepCSV

H→cc

PhysRevLett.
JHEP03(2020)131

- Very challenge: small BR, high background level, broad energy resolution (c-jet), difficult charm tagging …
- A general strategy is to use VH with lepton triggers and heavily apply ML techniques

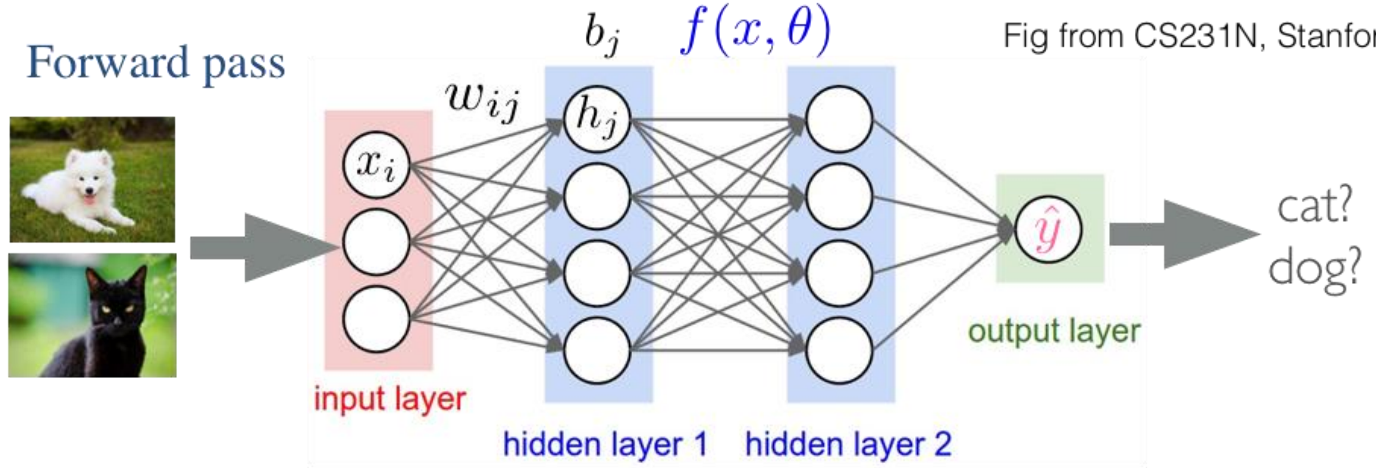
ATLAS: ZH→llcc	CMS: VH→llcc, ννcc, lνcc
Two same-flavor leptons Two R=0.4 resolved c-jets Categories: 1,2 c-tag and pTZ split at 150 GeV Simultaneous fit to m(cc) in 4 categories Two BDT 41% c-tagging eff 5% l-jet mistag rate 25% b-jet mistag rate	ee, $\mu\mu$, $e\nu$, $\mu\nu$, $\nu\nu$ Two R=0.4 resolved c-jets Multiple categories for signals and backgrounds separated by c-tagging scores Simultaneous fit to BDT scores in all categories DeepCSV (DNN) 28% c-tagging eff 4% l-jet mistag rate 15% b-jet mistag rate



C-tagging performance is crucial, especially in the b-jet rejection

What is deep neural network

Fig from CS231N, Stanford

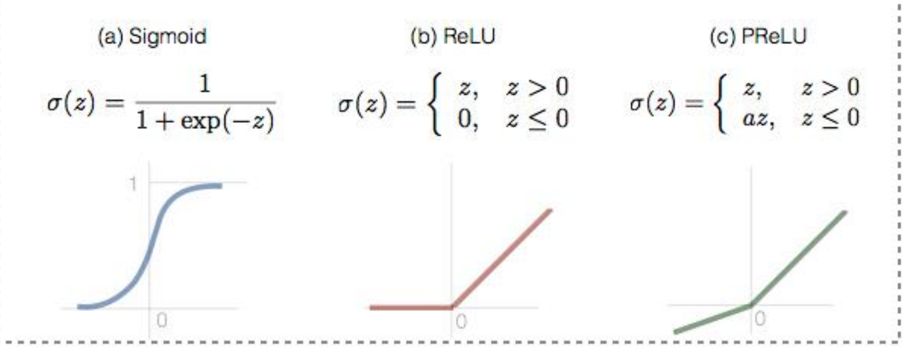


Linear operation

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

scaling, rotating, boosting, changing dimensions

Non-linear activation function $h_j = \sigma(z_j)$



DNN: artificial neural network with multiple hidden layers

How does deep neural network learn: back propagation

Input x → Neural network $\hat{y} = f(x, \theta)$ → $\hat{y}$: Network prediction; y: True label

↓

Loss(cost) function

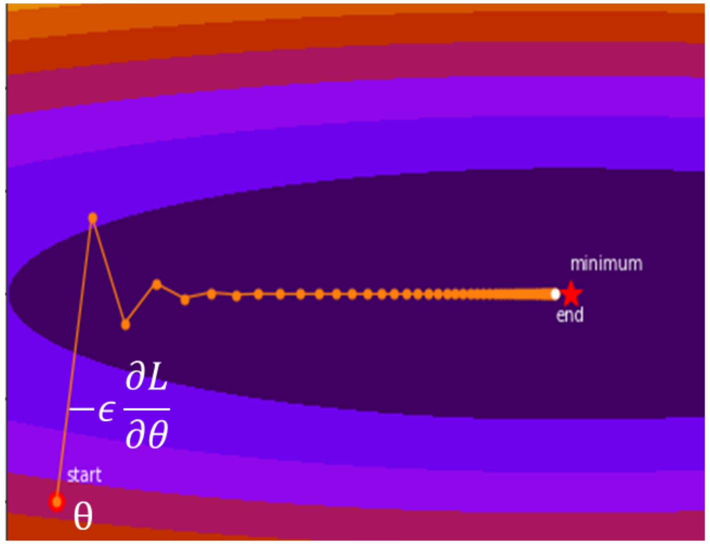
$$L = \sum_i (\hat{y}_i - y_i)^2$$

↓

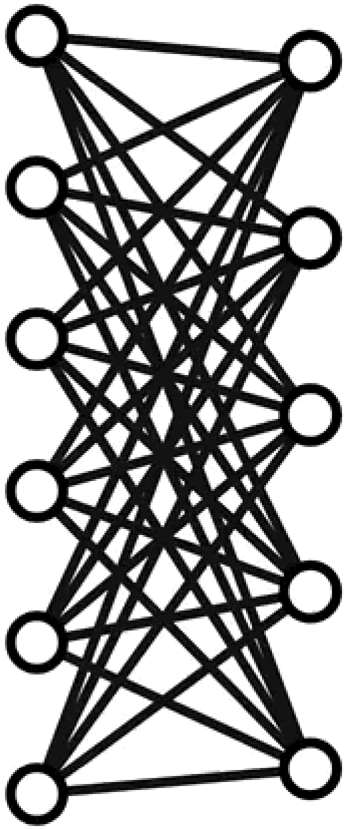
Gradient decent

$$\theta = \theta - \epsilon \frac{\partial L}{\partial \theta}$$

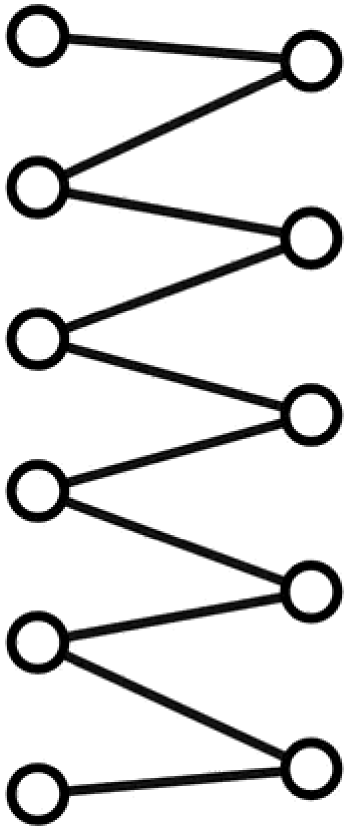
Back propagation



Convolution Network

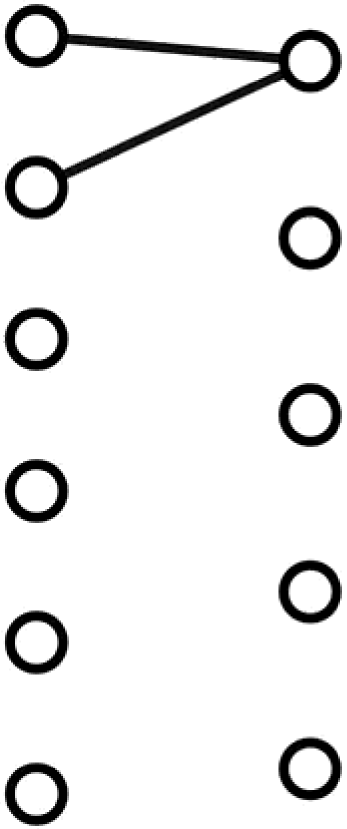


Densely connected



Locally connected

1D convolution



Locally connected
and sharing weights

CNN for jet tagging

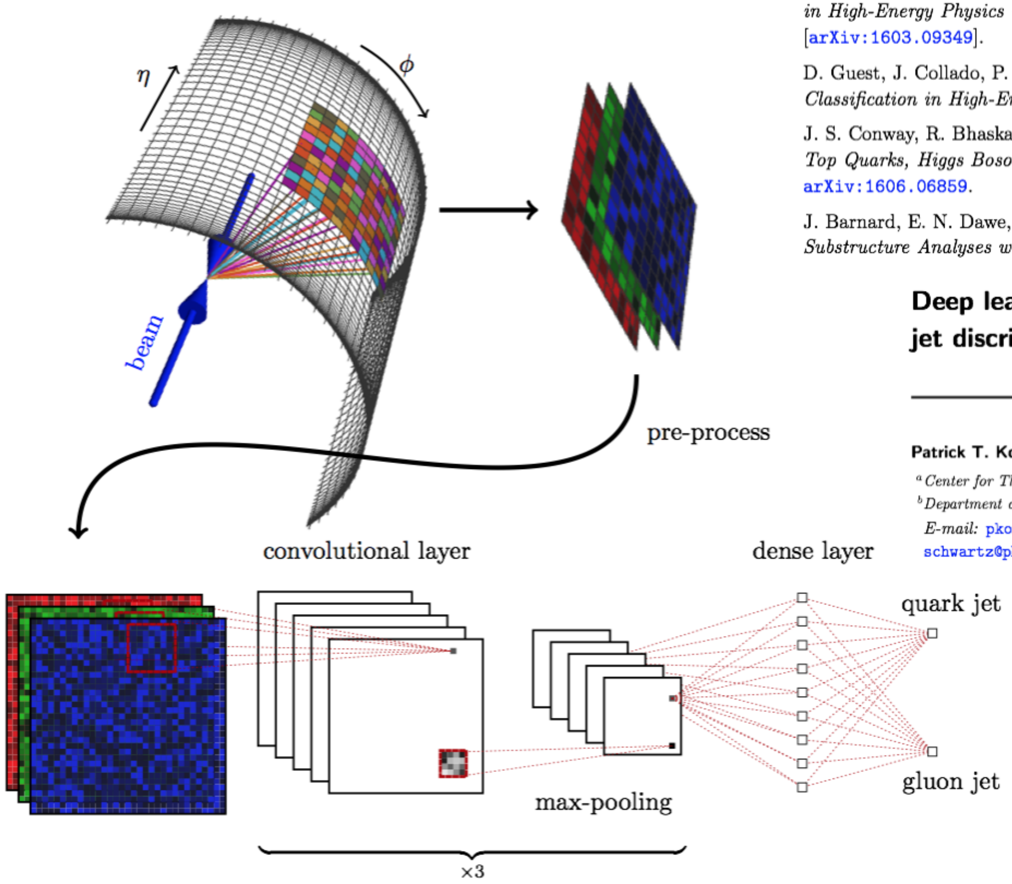


Figure 2: An illustration of the deep convolutional neural network architecture. The first layer is the input jet image, followed by three convolutional layers, a dense layer and an output layer.

P. Baldi, K. Bauer, C. Eng, P. Sadowski, and D. Whiteson, *Jet Substructure Classification in High-Energy Physics with Deep Neural Networks*, *Phys. Rev.* **D93** (2016), no. 9 094034, [arXiv:1603.09349].

D. Guest, J. Collado, P. Baldi, S.-C. Hsu, G. Urban, and D. Whiteson, *Jet Flavor Classification in High-Energy Physics with Deep Neural Networks*, arXiv:1607.08633.

J. S. Conway, R. Bhaskar, R. D. Erbacher, and J. Pilot, *Identification of High-Momentum Top Quarks, Higgs Bosons, and W and Z Bosons Using Boosted Event Shapes*, arXiv:1606.06859.

J. Barnard, E. N. Dawe, M. J. Dolan, and N. Rajcic, *Parton Shower Uncertainties in Jet Substructure Analyses with Deep Neural Networks*, arXiv:1609.00607.

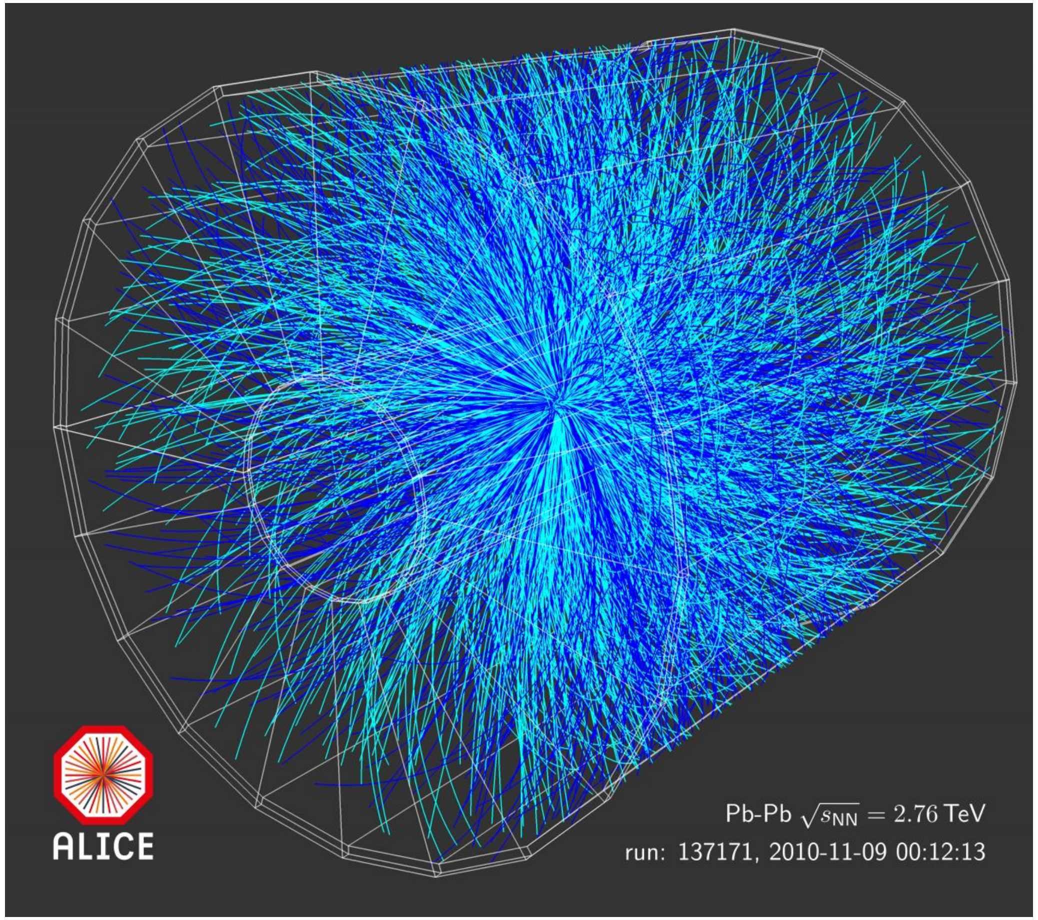
Deep learning in color: towards automated quark/gluon jet discrimination

Patrick T. Komiske,[a] Eric M. Metodiev,[a] and Matthew D. Schwartz[b]

[a] *Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[b] *Department of Physics, Harvard University, Cambridge, MA 02138, USA*
E-mail: pkomiske@mit.edu, metodiev@mit.edu, schwartz@physics.harvard.edu

Deep CNN match or outperform traditional jet observables.

Point Cloud in momentum space



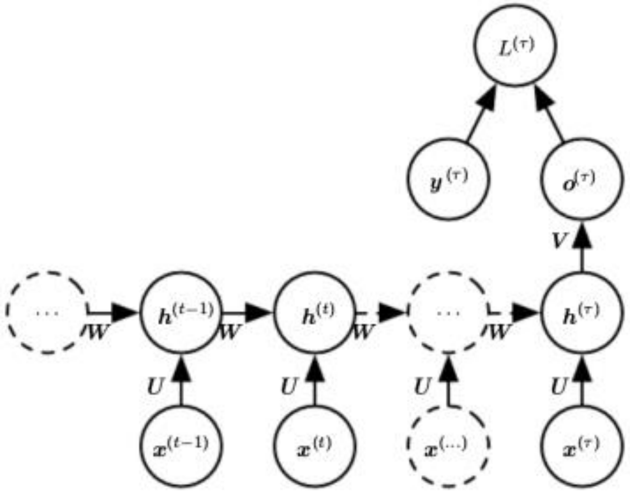
- Images: histograms
  - (px, py) or (pt, phi)
  - (px, py, pz)
  - (pt, phi, eta)
- Point cloud: particle list

E	Px	Py	Pz	pid
6.84	1.07	4.5	6.83	211
68.92	0.75	0.64	68.91	2212
40.4	0.06	0.54	40	321
…				

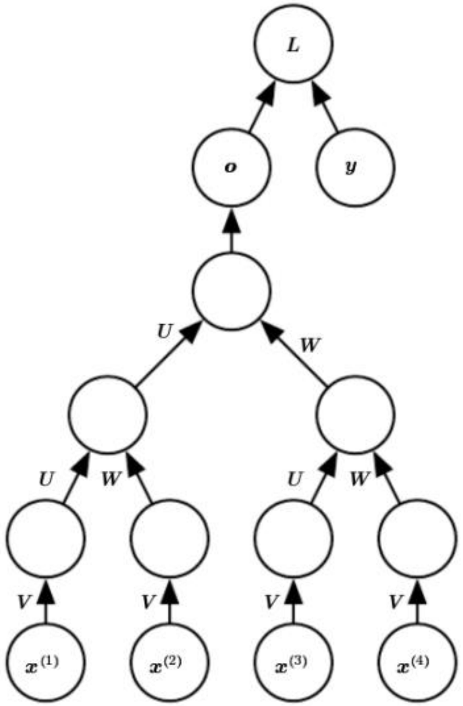
Recurrent and Recursive network for q/g jet tagging

Gluon Jet

Quark Jet



Recurrent Net



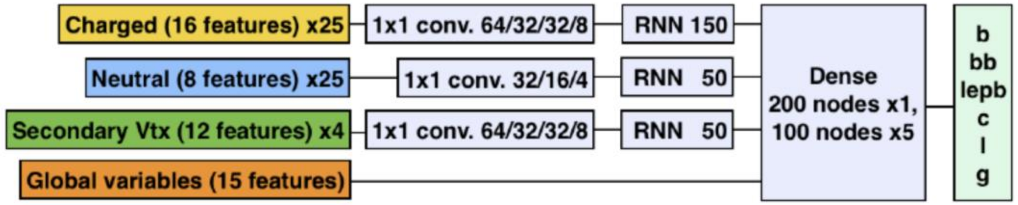
Recursive Net

[G. Louppe, K. Cho, C. Becot, K. Cranmer, arXiv: 1702.00748;
Taoli Cheng, 中国科学院大学, Comput Softw Big Sci (2018) 2: 3

$$\mathbf{h}_k = \sigma\left(W_h \begin{bmatrix} \mathbf{h}^{\text{jet}}_{k_L} \\ \mathbf{h}^{\text{jet}}_{k_R} \\ \mathbf{u}_k \end{bmatrix} + b_h\right)$$

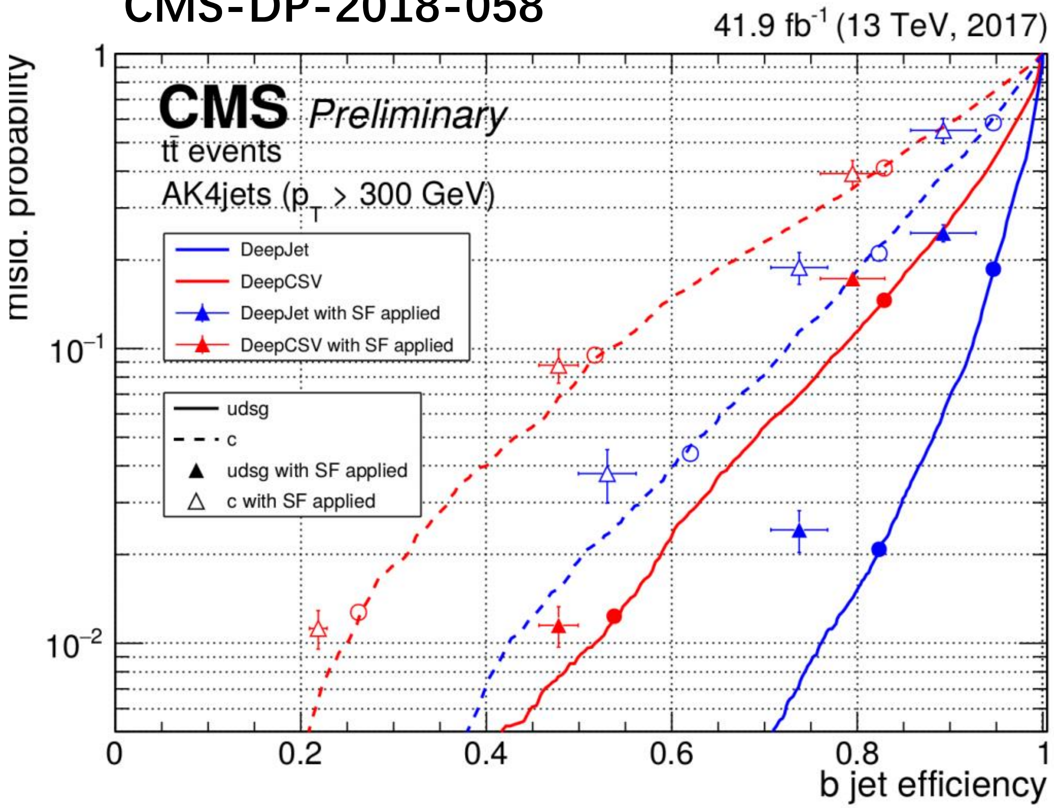
DeepJet to identify jets originating from b quarks

高能所： DeepJet



- DeepJet performs better than the previous widely used DeepCSV
- DeepCSV: 4 hidden layer fully connected network, 100 neurons per layer

CMS-DP-2018-058



Permutation symmetry: Particle/Energy flow network

FCN, RNN, RcNN all break the permutation symmetry

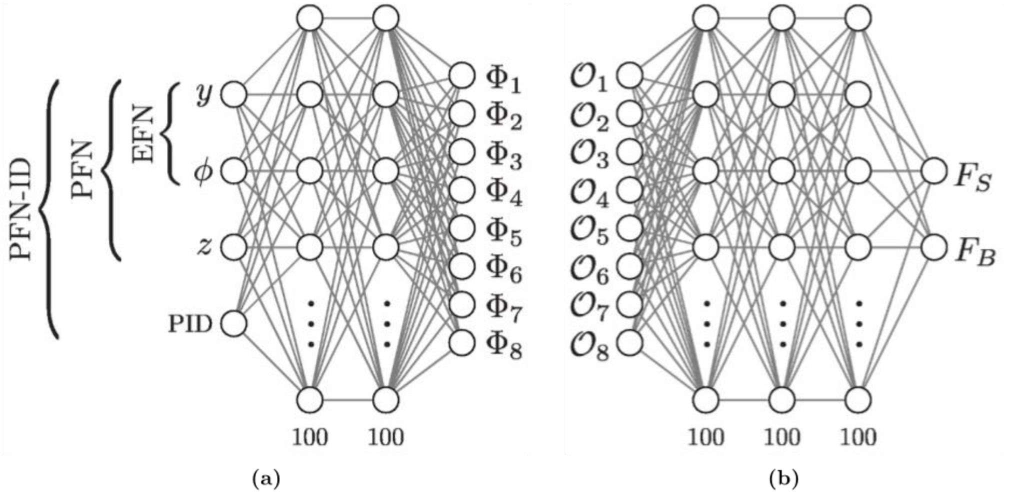
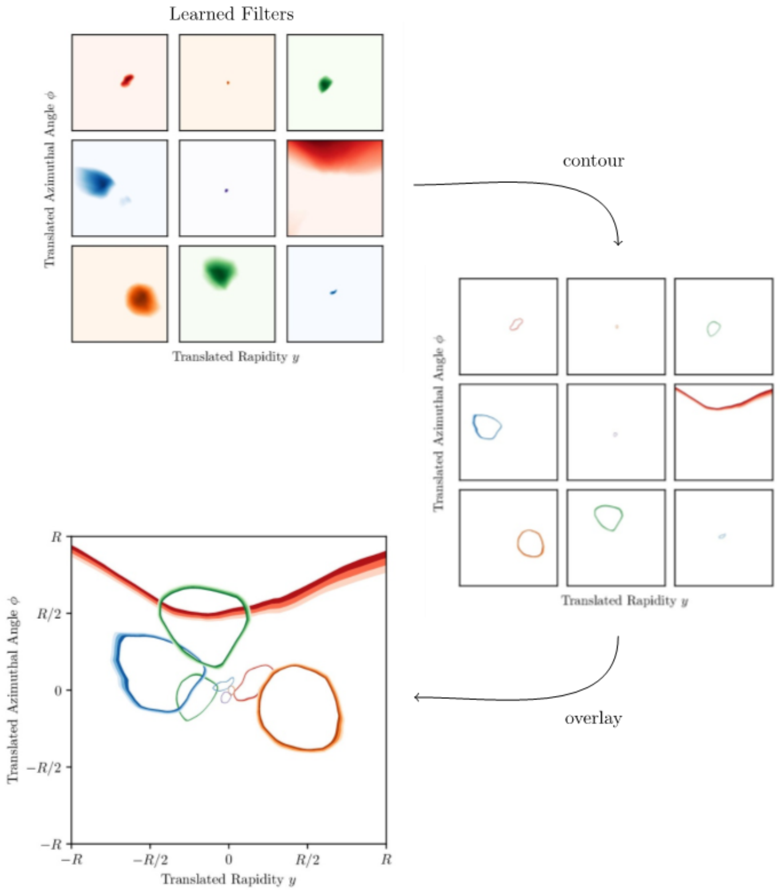


Figure 4. The particular dense networks used here to parametrize (a) the per-particle mapping Φ and (b) the function F, shown for the case of a latent space of dimension $\ell = 8$. For the EFN, the latent observable is $\mathcal{O}_a = \sum_i z_i \, \Phi_a(y_i, \phi_i)$. For the PFN family, the latent observable is $\mathcal{O}_a = \sum_i \Phi_a(y_i, \phi_i, z_i, \text{PID}_i)$, with different levels of particle-ID (PID) information. The output of F is a softmaxed signal (S) versus background (B) discriminant.



Extending point cloud network for jet classification and jet substructure studies.

JHEP 2018, 13, P.T. Komiske, E.M. Metodiev, and J. Thaler. 1810.00835 by Y.S. Lai

Graph CNN for point cloud – more local structure

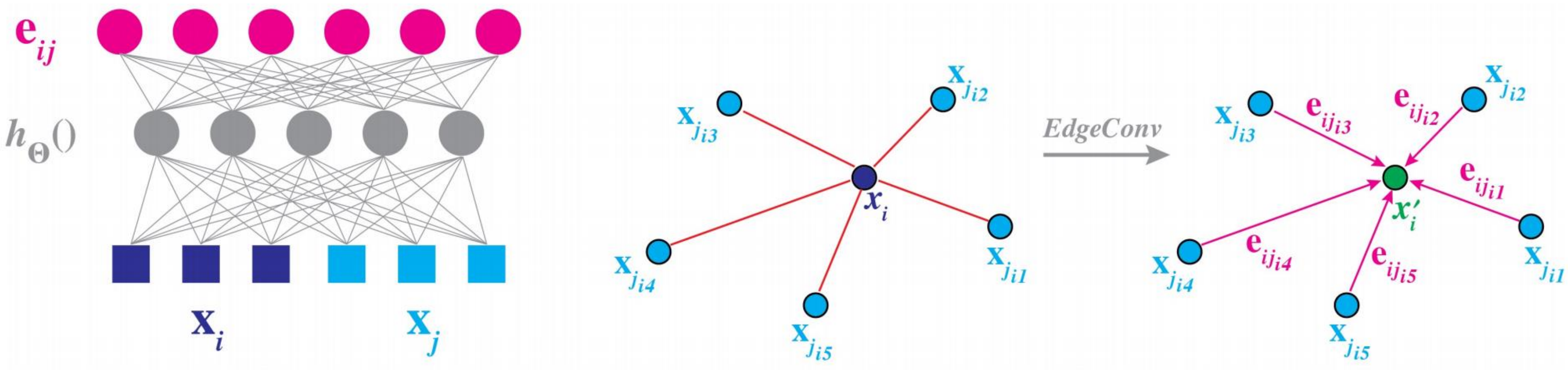
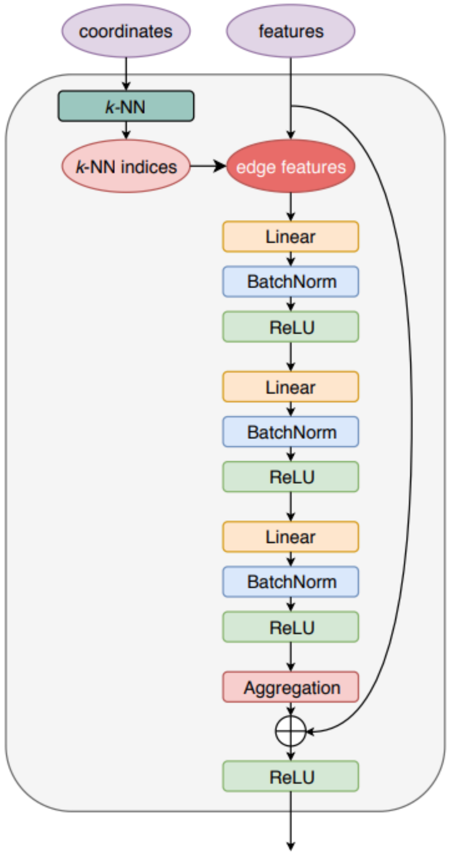


Fig. 2. **Left**: Computing an edge feature, $\mathbf{e}_{ij}$ (top), from a point pair, $\mathbf{x}_i$ and $\mathbf{x}_j$ (bottom). In this example, $h_\Theta()$ is instantiated using a fully connected layer, and the learnable parameters are its associated weights. **Right**: The EdgeConv operation. The output of EdgeConv is calculated by aggregating the edge features associated with all the edges emanating from each connected vertex.

Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E. Sarma, Michael M. Bronstein, and Justin M. Solomon Arxiv:1801.07829.

Jet tagging via ParticleNet

Edge Conv



$$D(x_i, x_j) = \sqrt{(p_{Ti} - p_{Tj})^2 + (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2} \,.$$

- Key Idea: **kNN + Edge Conv + PointCloud Net**
- **Better with PID**

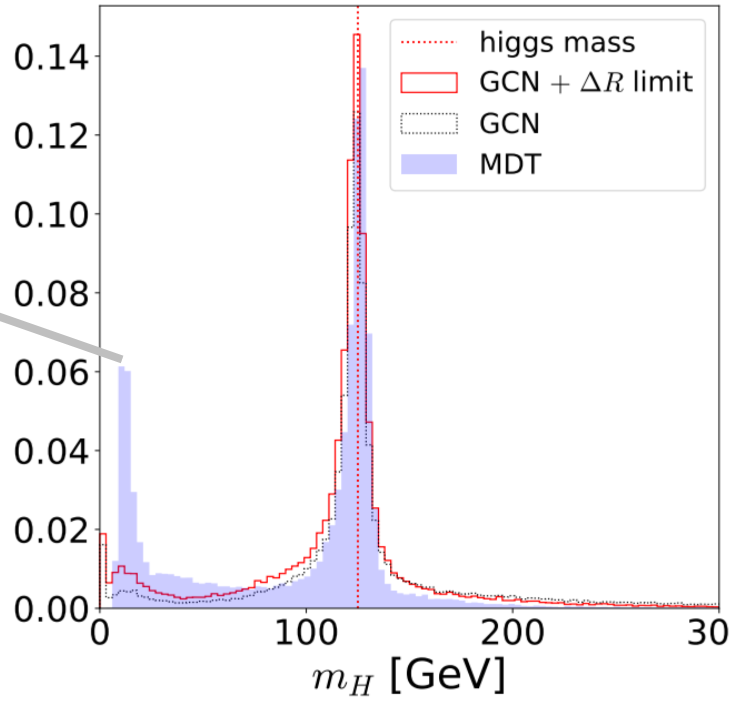
	Accuracy	AUC
ResNeXt-50	0.821	0.8960
P-CNN	0.818	0.8915
PFN	-	0.8911
ParticleNet-Lite	0.826	0.8993
ParticleNet	0.828	0.9014
P-CNN (w/ PID)	0.827	0.9002
PFN-Ex (w/ PID)	-	0.9005
ParticleNet-Lite (w/ PID)	0.835	0.9079
ParticleNet (w/ PID)	**0.840**	**0.9116**

Jet Tagging via Particle Clouds, HuiLin Qu and Loukas Gouskos, 2020

Graph CNN for boosted Higgs reconstruction

Reconstructed m_H from GCN is closer to 125 GeV

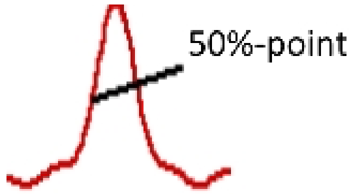
Fake signal suppressed

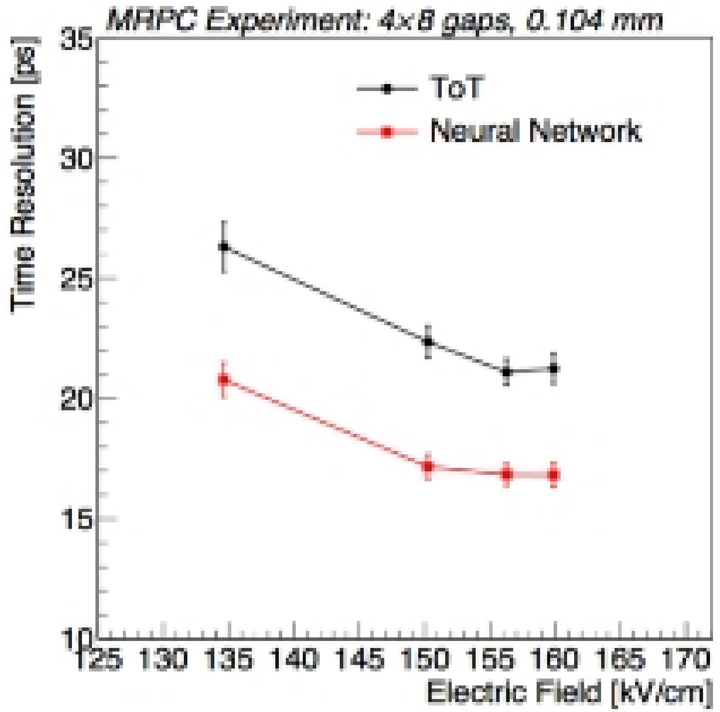


$p_T(H) > 300\ GeV$

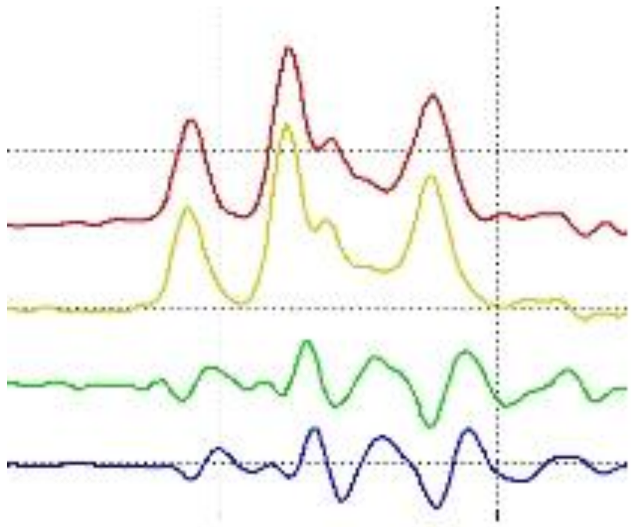
2003.11603, X. Ju et al.; 2010.05464, Jun Guo, Jinmian Li, and Tianjun Li;

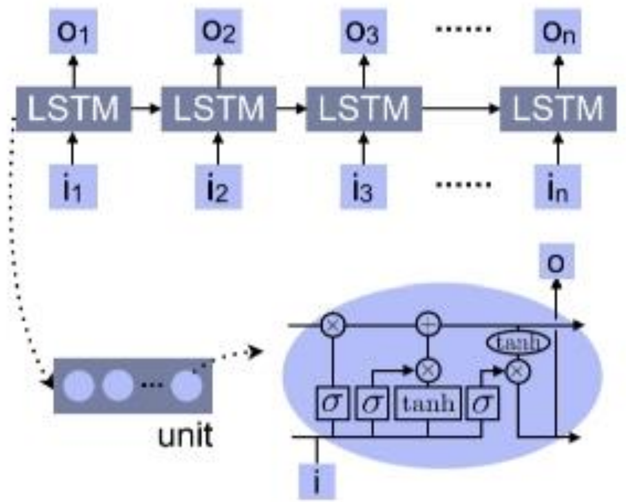
Recurrent net for MRPC and RPC time resolution regression





FY. Wang, RPC 2020





(b) Long Short Term Memory network

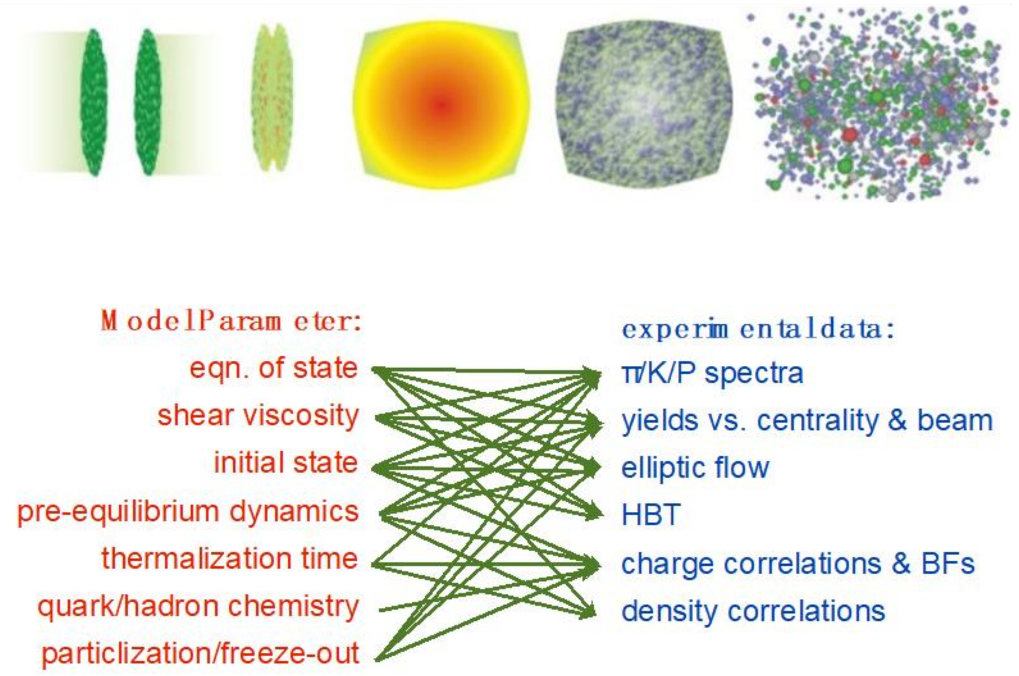
Input: RPC waveforms (18 ns * 4 channels * 2)

Network structure:
3-layer LSTM (hidden size ~700)
3-layer MLP in series
1-dim output: ToF_predict

XiangYu Xie, USTC
Talk: 2020-11-07

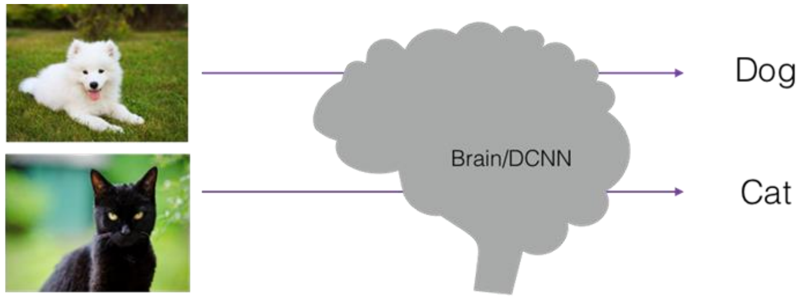
ML for Heavy Ion Collisions

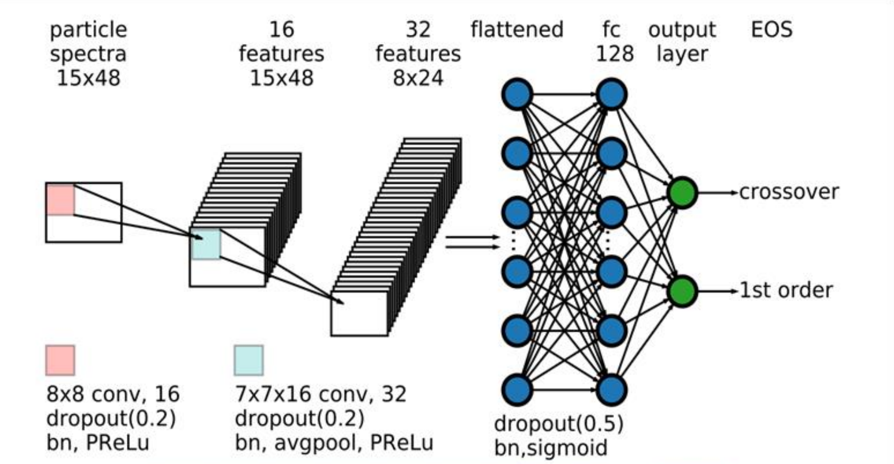
Fig originally from Steffen A. Bass



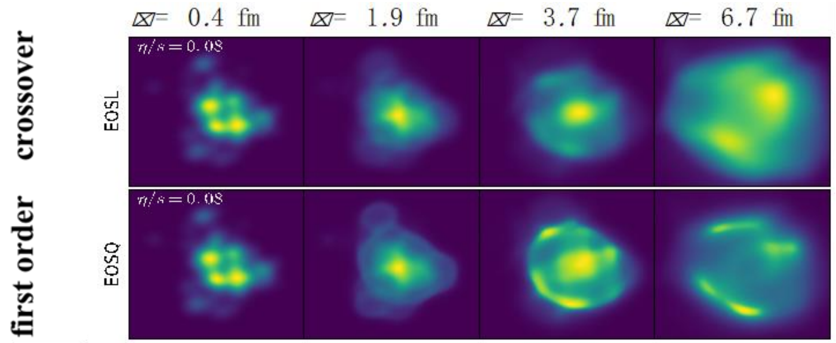
- **Many parameters contribute to the same observable**
- **How to constrain one physical parameter?**
- **Information survived?**
- **Encoded in final state?**
- **How to decode?**

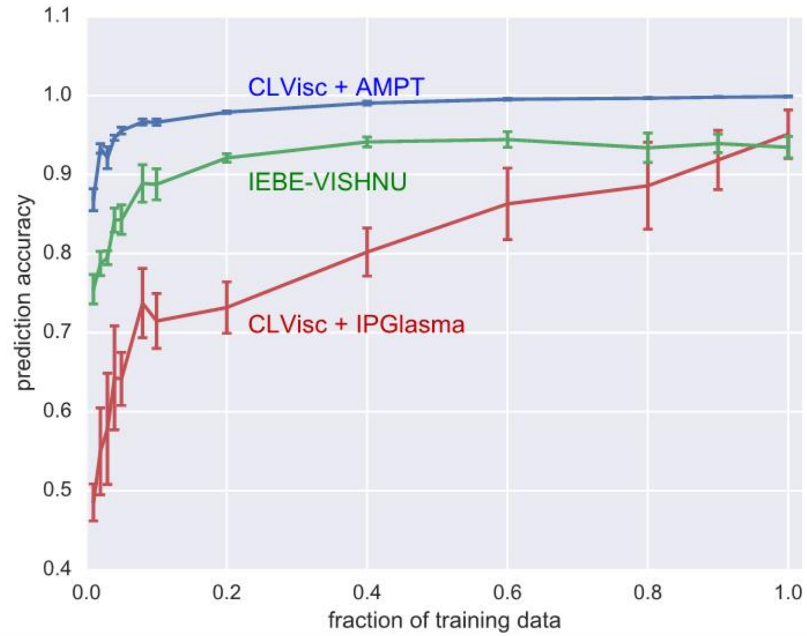
CNN for QCD Phase transition





$$l(\theta) = \underbrace{-\frac{1}{N}\sum_{i=1}^{N}[y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i)]}_{\text{cross entropy loss}} + \underbrace{\lambda||\theta||_2^2}_{\text{L2 regularization}}$$





Nature Communications 2018, LG. Pang, K.Zhou, N.Su, H.Petersen, H. Stoecker, XN. Wang.

Stacked U-net for relativistic hydrodynamics

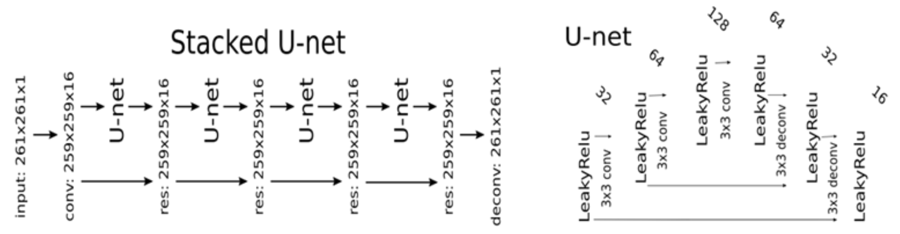
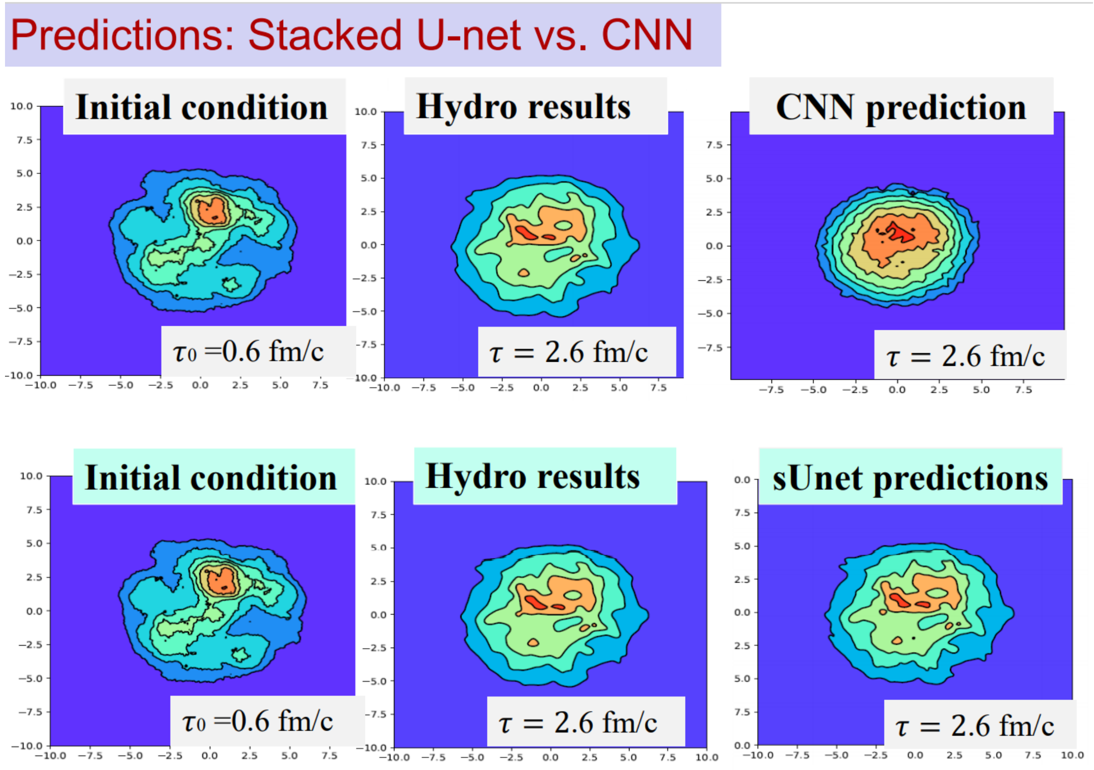


FIG. 1: An illustration of the encode-decode network, `stacked U-net`, which consists of the input and out layers and four residual U-net blocks. The right figure shows the U-net structure, and the depth of the hidden layer is written on the top of them.

The expansion of quark gluon plasma is learned in the image translation task using stacked UNET.

$$\nabla_\mu T^{\mu\nu} = 0$$

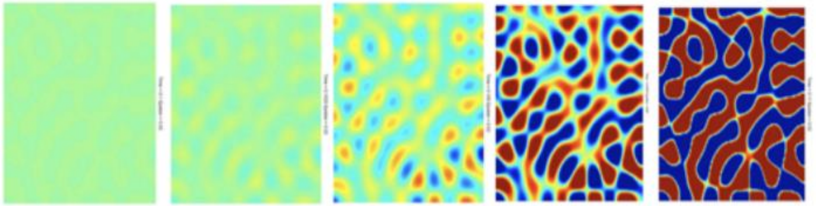


-sUnet is the proper directions that works

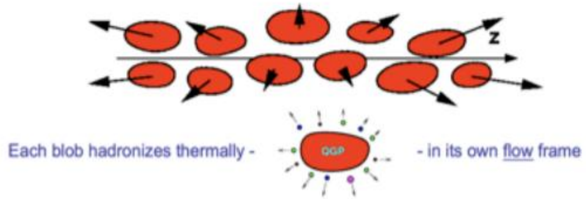
arXiv: 1801.03334; NPA2018, H.Huang, B.Xiao, H.Xiong, Z.Wu, Y. Mu and H.Song

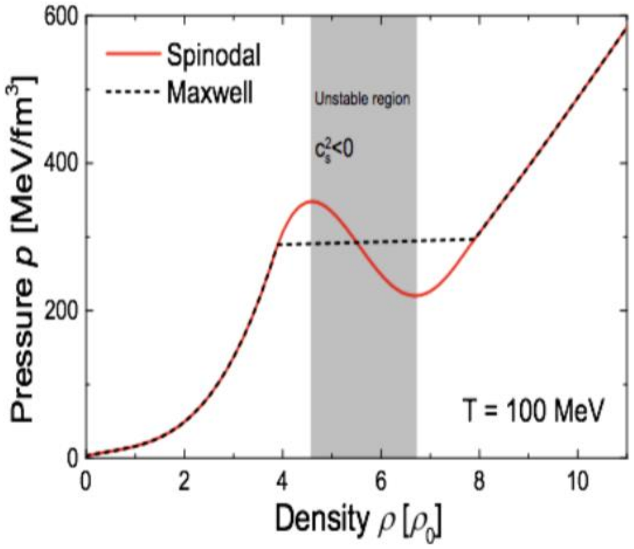
Point cloud network for QCD EoS

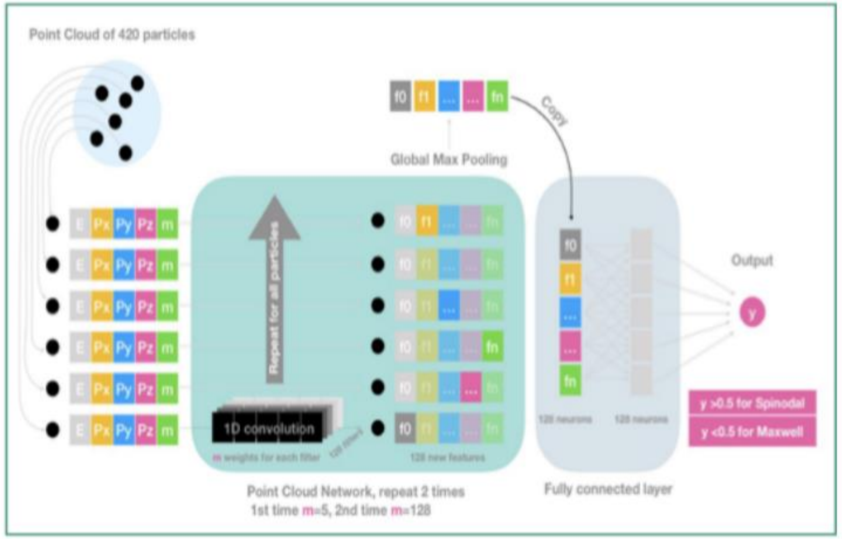
wiki



J. Randrup, INT 2009





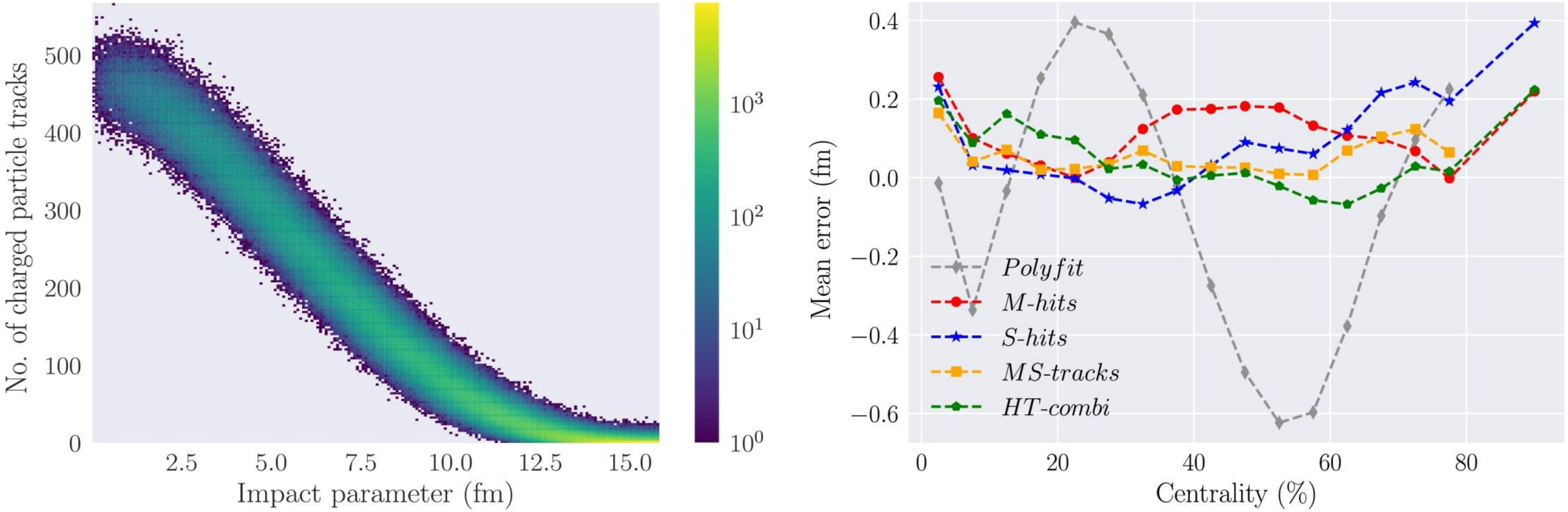


J. Steinheimer, L.G. Pang, K. Zhou, V. Koch, H. Stoecker, J. Randrup, 2019, JHEP

Point Cloud Net for impact parameter determination

A fast centrality-meter for heavy-ion collisions at the CBM experiment

Manjunath Omana Kuttan,[1,2,*] Jan Steinheimer,[1] Kai Zhou,[1,†] Andreas Redelbach,[1,3] and Horst Stoecker[1,2,4]
[1]*Frankfurt Institute for Advanced Studies, D-60438 Frankfurt am Main, Germany*



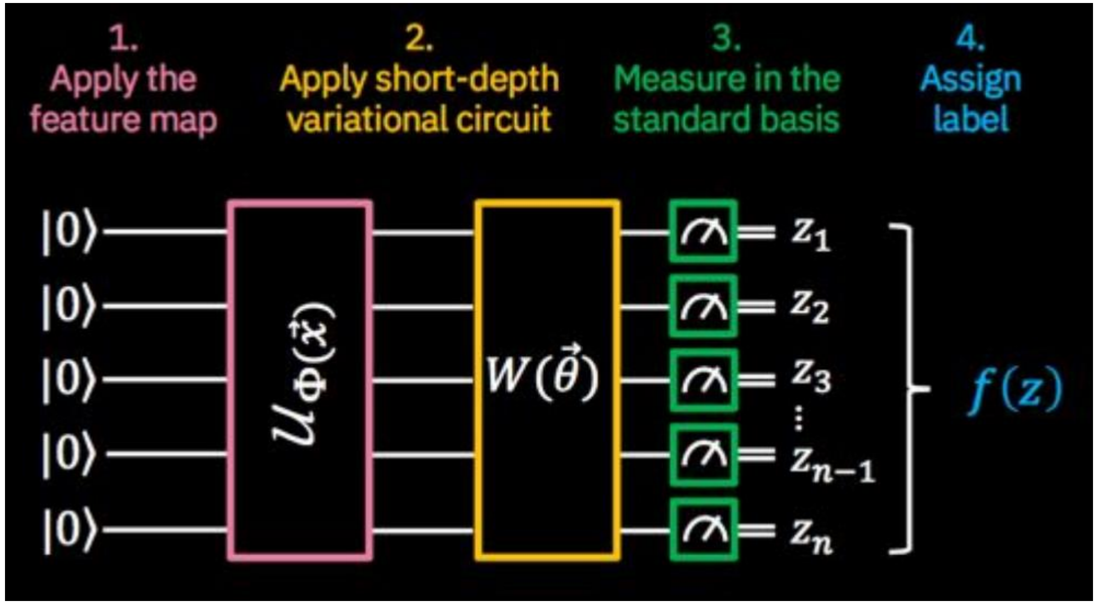
See also: 2008.11540 by Fupeng Li, Yongjia Wang, Hongliang Lv, Pengcheng Li, Qingfeng Li, and Fanxin Liu

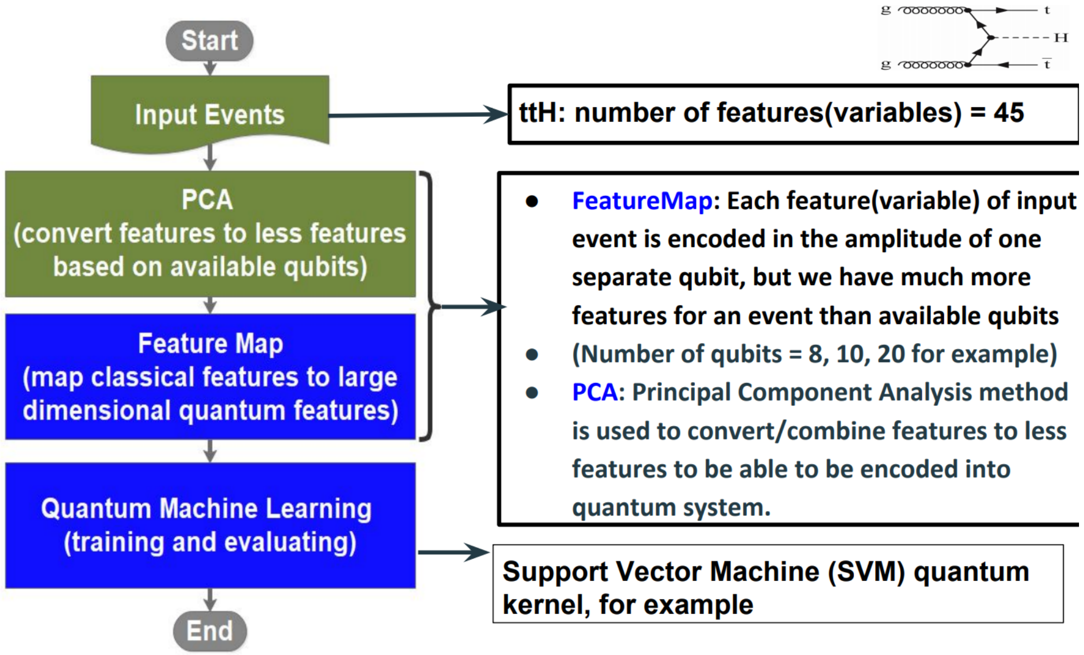
Quantum Machine Learning

- **Huge Hilbert space**
- **Before measuring, the states are in quantum parallel**
- **Period of quantum interference can be measured**
- **No classical correspondence for entanglements**
- **Feynman: simulate one quantum system with another**
- **Quantum Machine learning: the best of two worlds**

Quantum Machine Learning for HEP: QSVM

Many quantum machine learning algorithms are inspired by variational quantum eigen solver.





S. L. Wu and C. Zhou (U. Wisconsin) 40th International Conference on High Energy Physics July 28, 2020

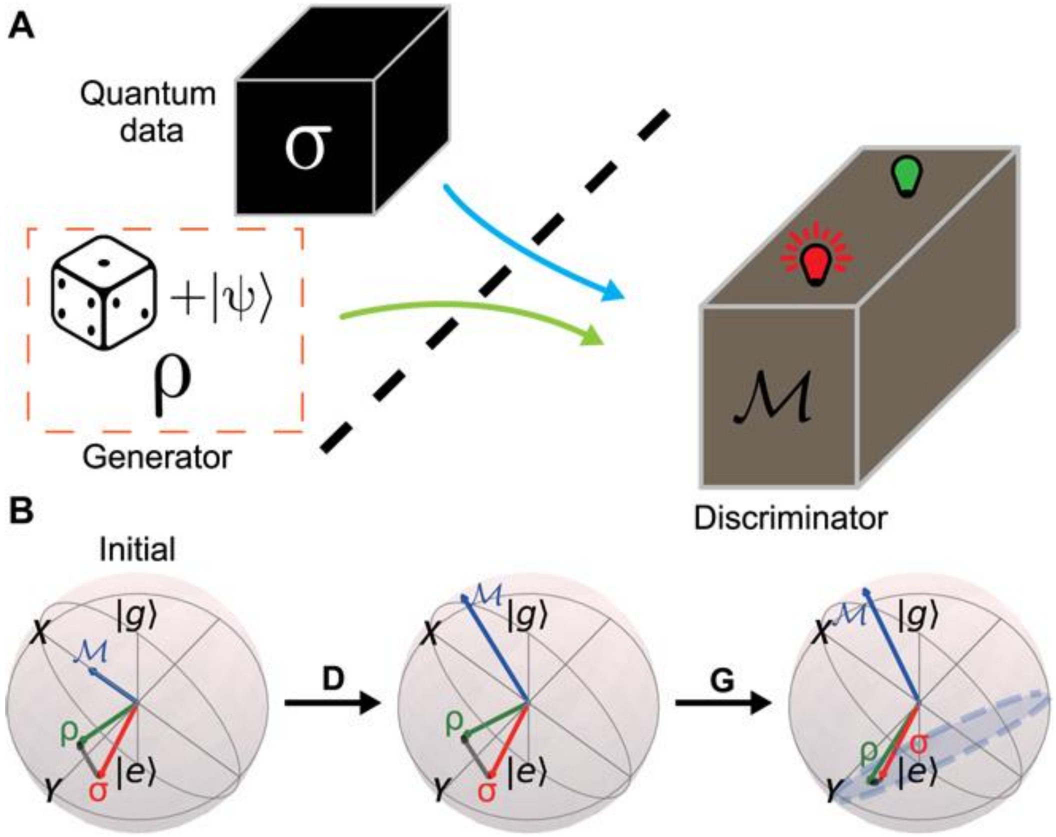
Quantum GAN

Quantum generative adversarial learning in a superconducting quantum circuit

Ling Hu[1,*], Shu-Hao Wu[2,*], Weizhou Cai[1], Yuwei Ma[1], Xianghao Mu[1], Yuan Xu[1], Haiyan Wang[1], Yipu Song[1], D...
+ See all authors and affiliations

Science Advances 25 Jan 2019:
Vol. 5, no. 1, eaav2761
DOI: 10.1126/sciadv.aav2761

- **Generative Adversarial Network (GAN) is quite successful in image and video generation.**
- **GAN is used as fast emulator of MC event generators (GEANT4).**
- **Real Quantum Generator!**



Summary

- ML plays important role in HEP and HIC
- ParticleNet/GCNN might be the way for HEP
- What has not been mentioned
    - Bayesian analysis
    - Attention network
    - Capsule network
    - Uncertainties
    - Interpretation
    - BDT, GAN and Flow models for high dim numerical integration
    - Various applications in detector design, pileup mitigation and event(track) reconstruction

The topics are selected and biased by the limitation of personal knowledge

Backups

Bayesian analysis

$$P(y \mid x)\, P(x) = P(x \mid y)\, P(y)$$

$$\Downarrow$$

$$P(y|x) = \frac{P(x|y)P(y)}{P(x)}$$

$$\Downarrow$$

$$P(y \mid x) \propto P(x \mid y)\, P(y)$$

x: data (observables designed by experts)
y: model parameters

P(y | x): Posterior distribution
P(x | y): Likelihood

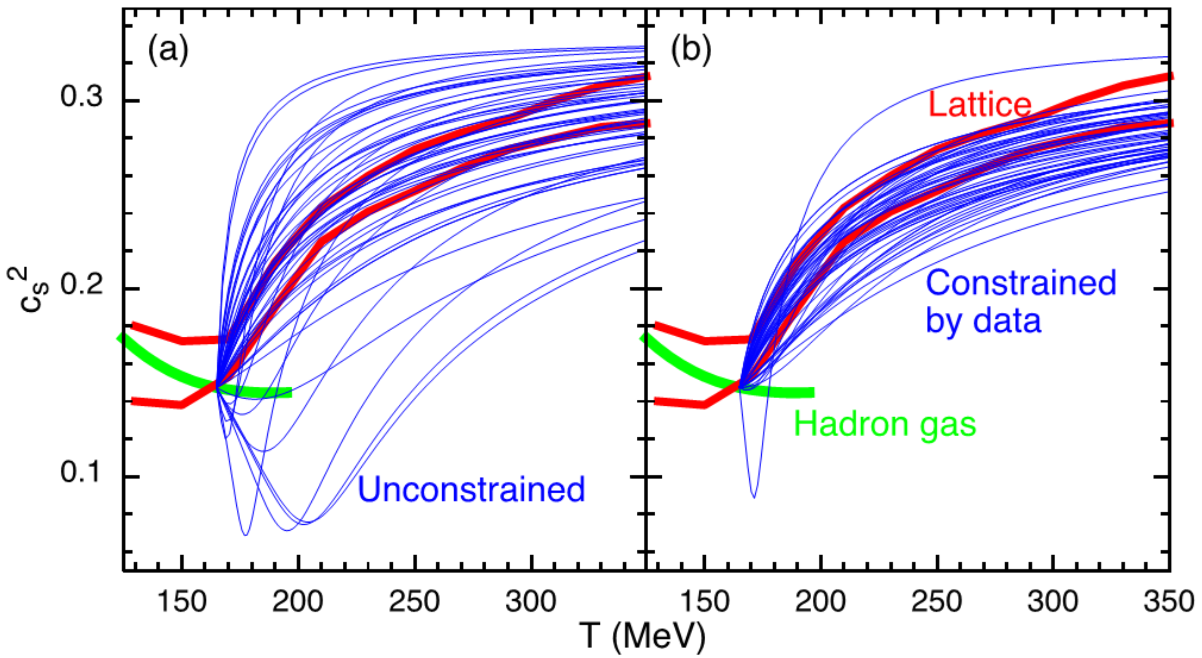
P(y): a prior

P$(x) = \int P(x|y)P(y)dy$: evidence

Metropolis importance sampling can sample unnormalized distributions.

Walk in parameter space of physical model.

Constrain QCD EoS



Data driven QCD EoS agrees with lattice QCD calculations.

PRL**114**, 202301 Scott Pratt, Evan Sangaline, Paul Sorensen, and Hui Wang

Constrain shear and bulk viscosity

Trento + iEBE-VISHNU + UrQMD

TABLE I. Input parameter ranges for the initial condition and hydrodynamic models.

Parameter	Description	Range
Norm	Overall normalization	100–250
p	Entropy deposition parameter	-1 to $+1$
k	Multiplicity fluct. shape	0.8–2.2
w	Gaussian nucleon width	0.4–1.0 fm
η/s hrg	Const. shear viscosity, $T < T_c$	0.3–1.0
η/s min	Shear viscosity at T_c	0–0.3
η/s slope	Slope above T_c	0–2 GeV^{-1}
ζ/s norm	Prefactor for $(\zeta/s)(T)$	0–2
T_{switch}	Particlization temperature	135–165 MeV

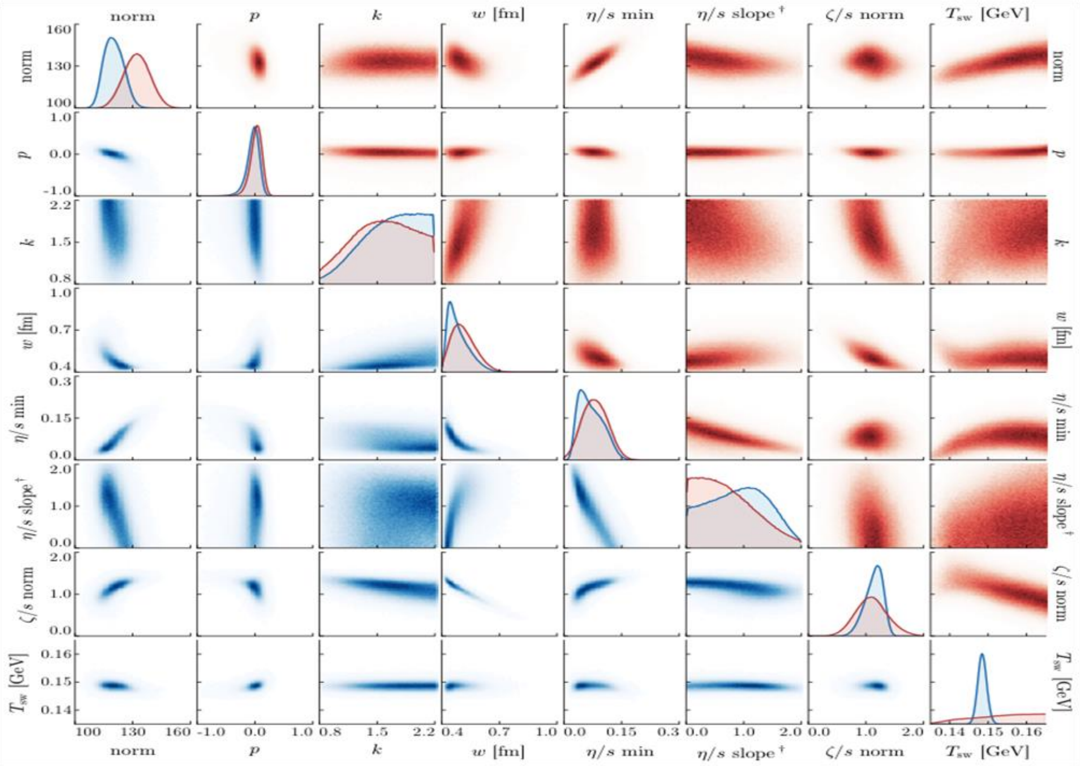
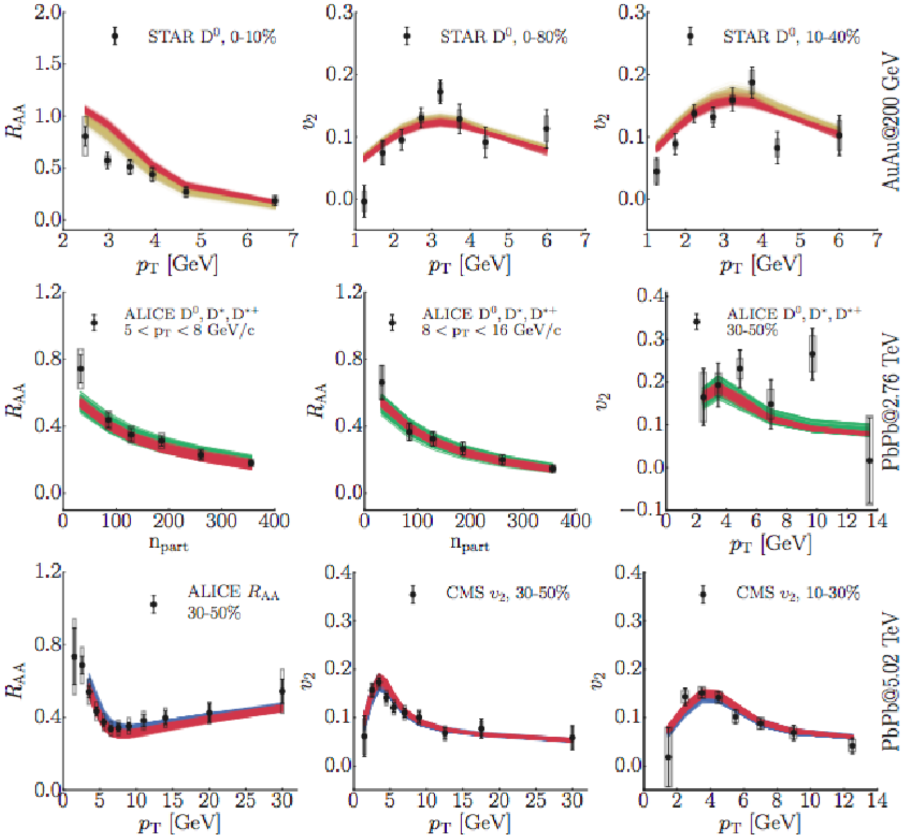


FIG. 7. Posterior distributions for the model parameters from calibrating to identified particles yields (blue, lower triangle) and charged particles yields (red, upper triangle). The diagonal has marginal distributions for each parameter, while the off-diagonal contains joint distributions showing correlations among pairs of parameters. †The units for η/s slope are [GeV^{-1}].

Clear non-zero $\frac{\eta}{s}$ and $\frac{\zeta}{s}$; Uncertain about $\frac{\eta}{s}$ slope; Fitting charged(red) .vs. identified(blue) particles

PRC 94.024907, J. E. Bernhard, J. Scott Moreland, S. A. Bass, J. Liu, U. Heinz

Constrain heavy quark diffusion coefficient



$$\frac{d\vec{p}}{dt} = -\eta_D(p)\vec{p} + \vec{\xi} + \vec{f}_g.$$

drag | fluctuation | gluon radiation

Spatial Diffusion Coefficient D_s

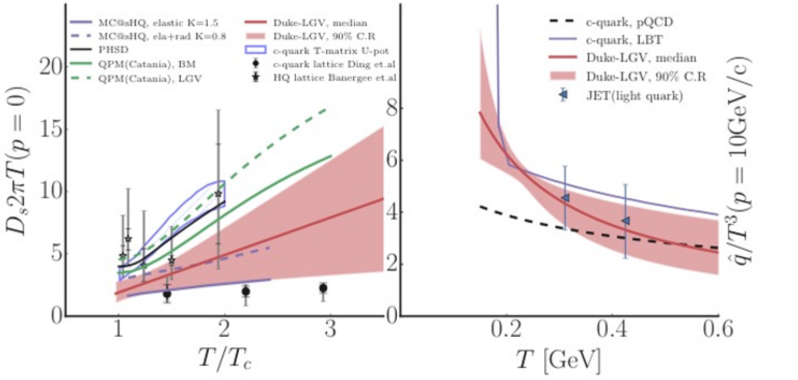


FIG. 12. (Color online) Comparison of the heavy quark diffusion coefficients across multiple approaches available in the literature. (**left**) spatial diffusion coefficient at zero momentum $D_s 2\pi T(p=0)$. (**right**) momentum diffusion coefficient $\hat{q}/T^3$ at $p = 10$ GeV.

PRC. **97 (2018)**, 014907, Yingru Xu, J.E. Bernhard, S.A. Bass and M. Nahrgang and S.S. Cao

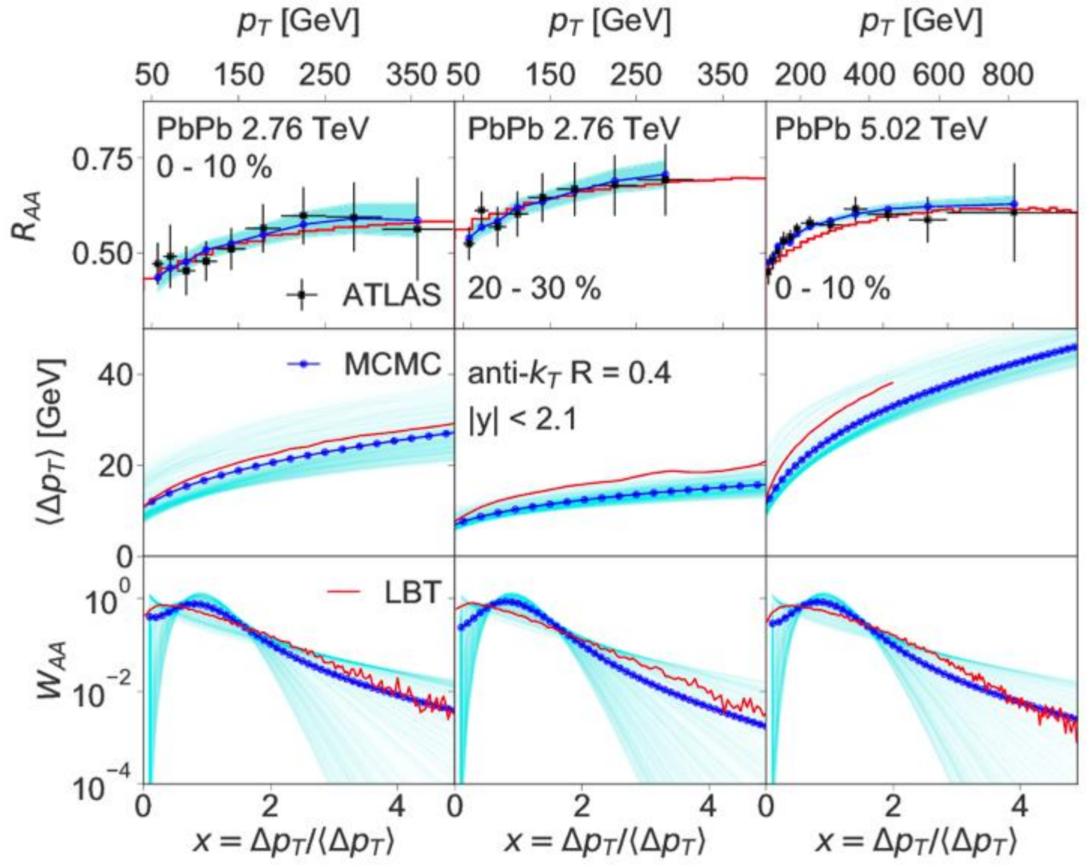
Constrain Jet energy loss distribution

LBT: Linear Boltzmann Transport model

$$p_a \cdot \partial f_a = \int \sum_{bcd} \prod_{i=b,c,d} \frac{d^3 p_i}{2E_i (2\pi)^3} (f_c f_d - f_a f_b) |\mathcal{M}_{ab\to cd}|^2 \times \frac{\gamma_b}{2} S_2(\hat{s}, \hat{t}, \hat{u}) (2\pi)^4 \delta^4(p_a + p_b - p_c - p_d) + \text{inelastic}, \quad (1)$$

Modification factor for Bayesian analysis

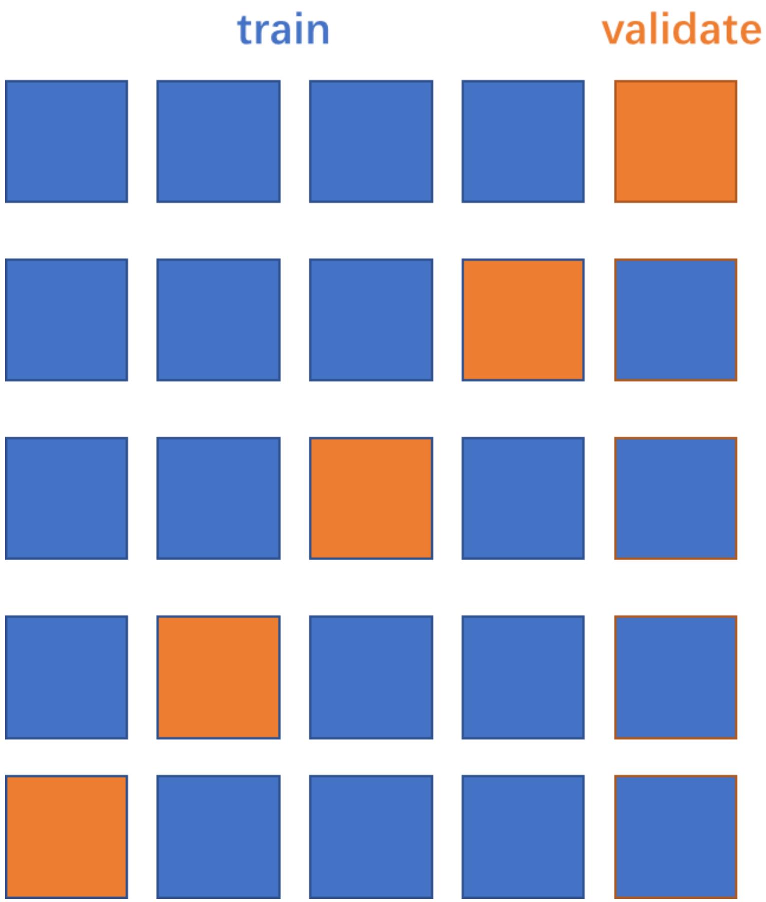
$$R_{AA}(p_T) \approx \frac{\int d\Delta p_T d\sigma_{pp}^{\rm jet}(p_T + \Delta p_T) W_{AA}(p_T + \Delta p_T \to p_T, R)}{d\sigma_{pp}^{\rm jet}(p_T)}.$$



PRL2019, Yayun He, L.G. Pang and X.N. Wang

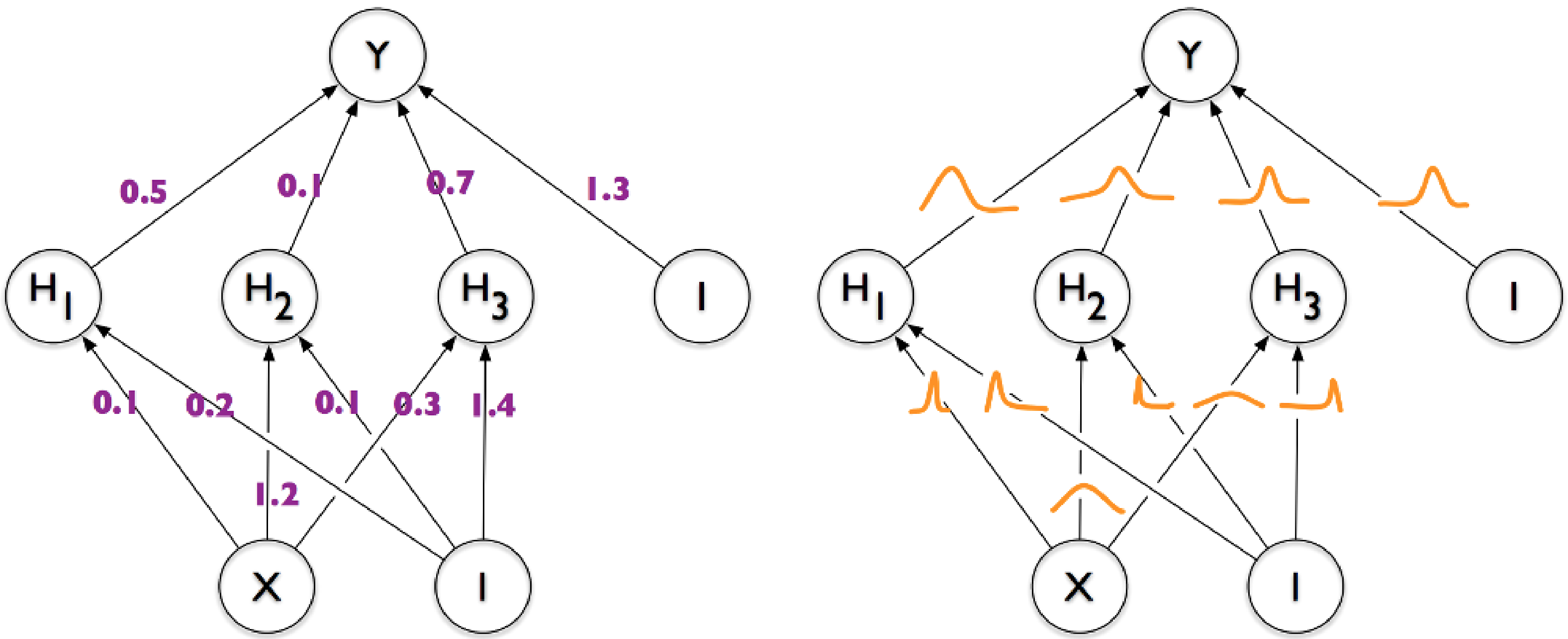
Uncertainty

K-fold Cross validation



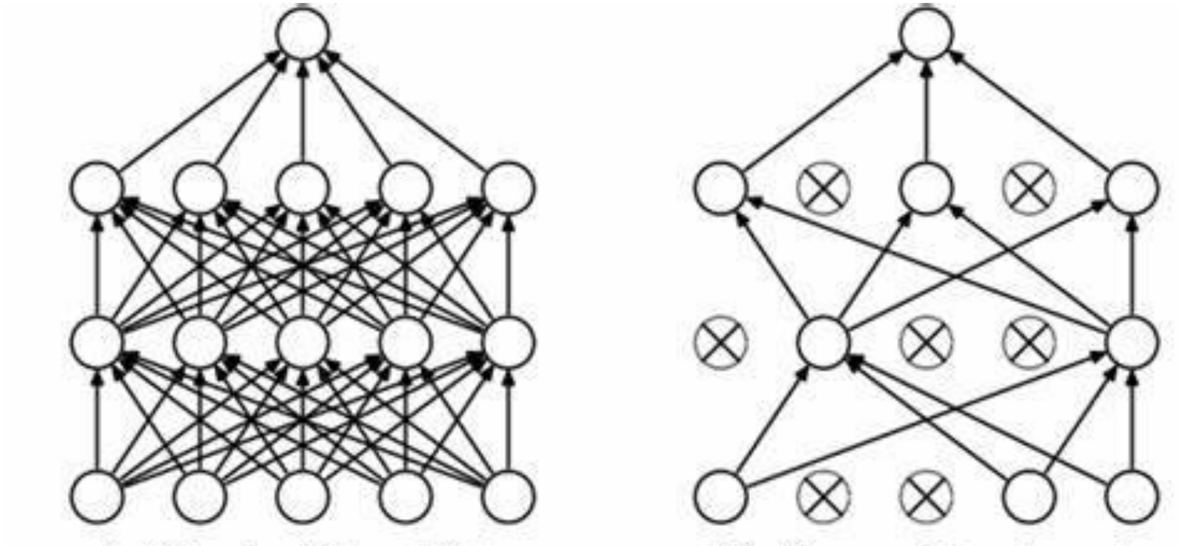
- Split data into K groups
- Train with K-1 groups
- Get K validation accuracy
- The uncertainty is the standard deviation of the K models
- Cons: training deep neural networks is computing intensive

Bayesian Neural Network



Weight Uncertainty in Neural Networks 2015

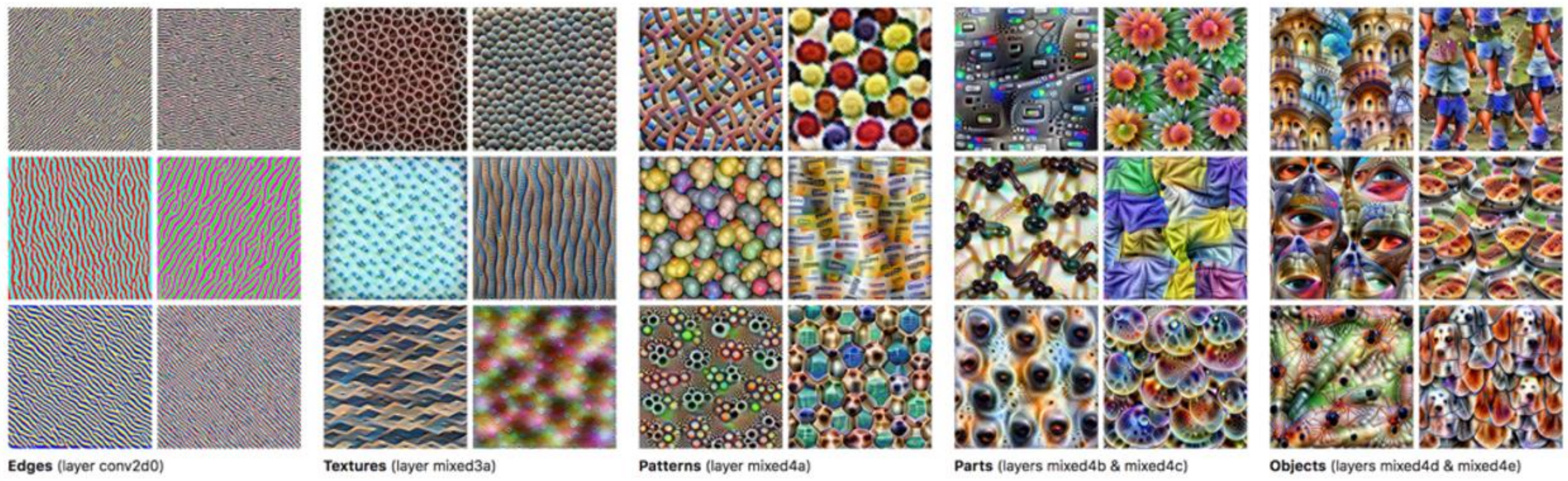
Monte Carlo dropout


(a) Standard Neural Net (b) After applying dropout.

Switch on dropout at testing stage to get an ensemble of networks.

Interpretable ML: global explanation

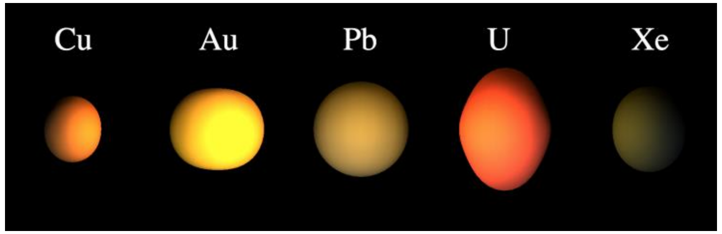
Olah, et al., "Feature Visualization", Distill, 2017.

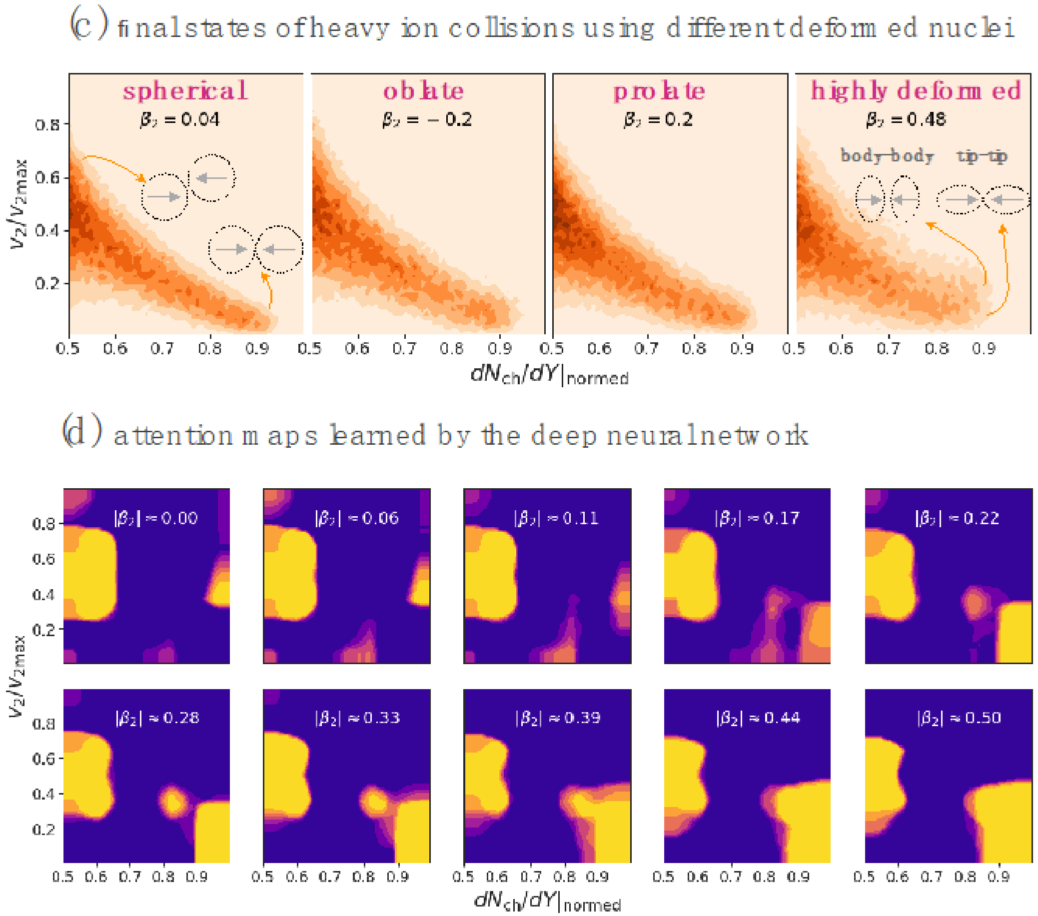


Global interpretation: staring with empty picture, visualize what has been learned by each neuron using gradient ascent.

Interpretable ML: local explanation

- **LIME; Surrogate; Activation map; Layer-wise relevance propagation, attention**





arXiv:1906.06429, L.-G. Pang, K. Zhou and X.-N. Wang