Machine Learning for nuclear matrix element (NME)

庞龙刚

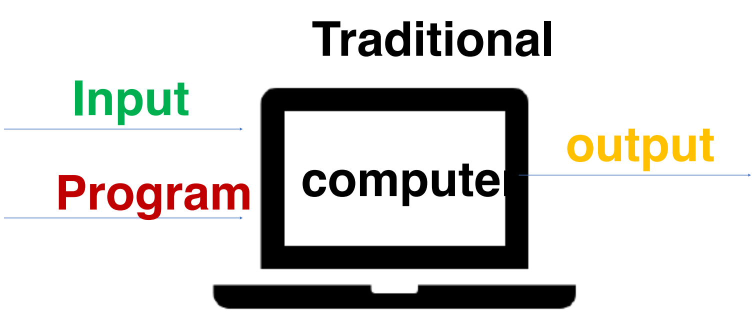
华中师范大学

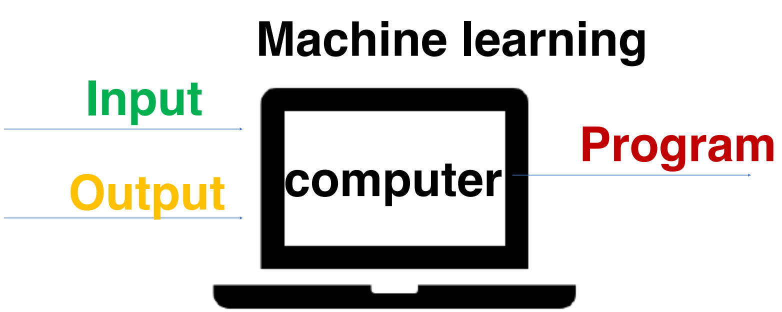
2021.5.23 - 中山大学 - 无中微子双beta衰变研讨会

CONTENTS

What is machine learning

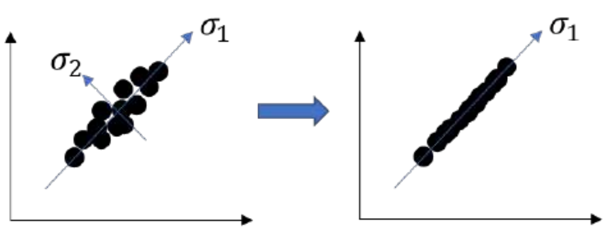
- A collection of algorithms (PCA, SVM, Random Forest, Boosting Trees, Neural Network …) that let computer learn patterns themselves.
- Keywords: Data driven; Functional; Optimize; Software 2.0;
- Minimize $loss\left[f(x,\ \theta),\ y\right] \rightarrow f$



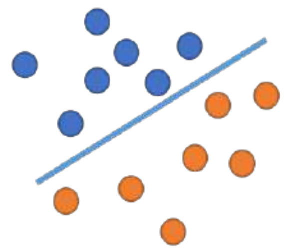


Applications of machine learning

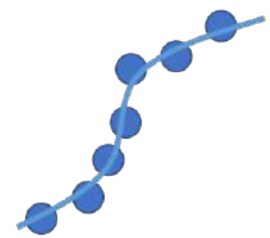
Dimensionality reduction(PCA, tSNE)



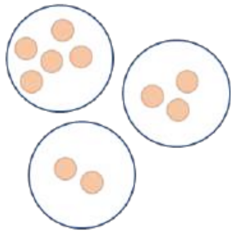
Classification



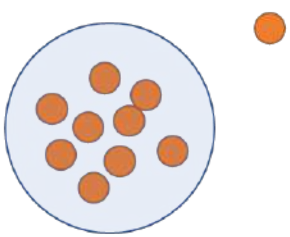
Regression



Clustering



Anomaly detection



Generation

Train with GAN, VAE or Flow model

Generate with given conditions

5

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

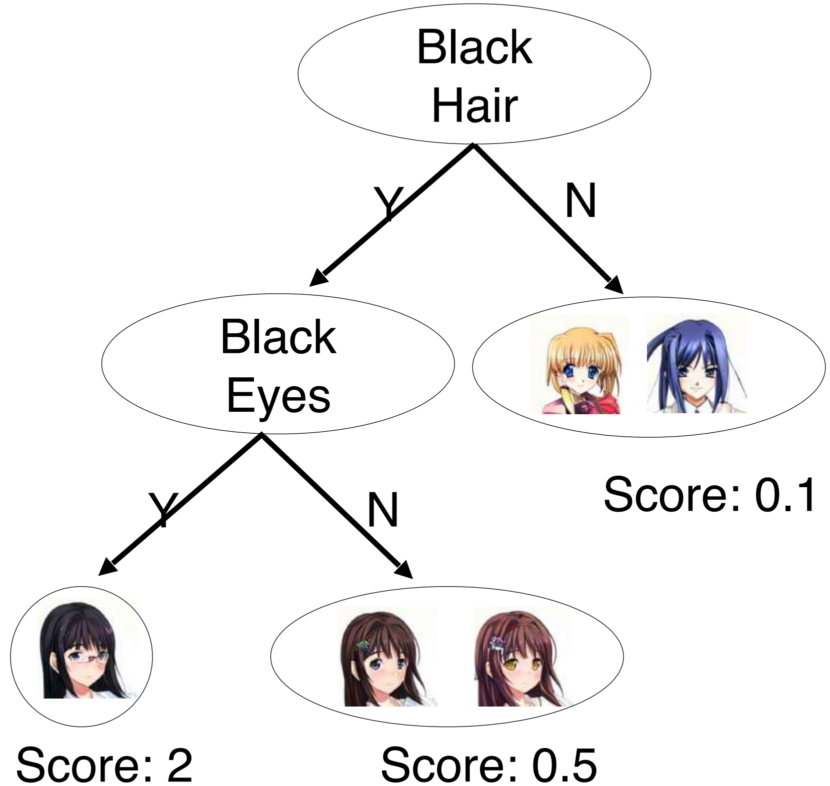
ML for HEP

- 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

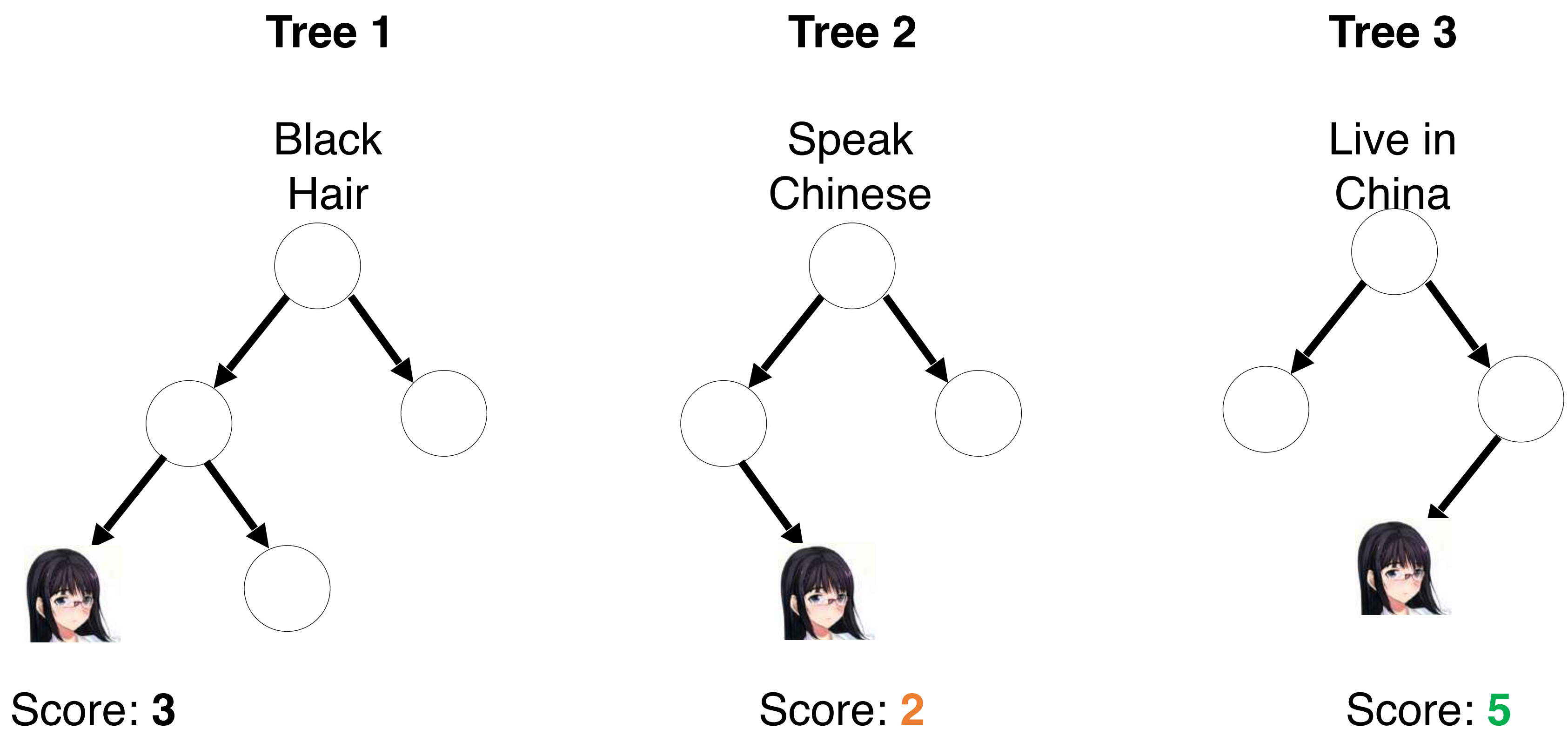
A single decision tree

Task: Asian girl?



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

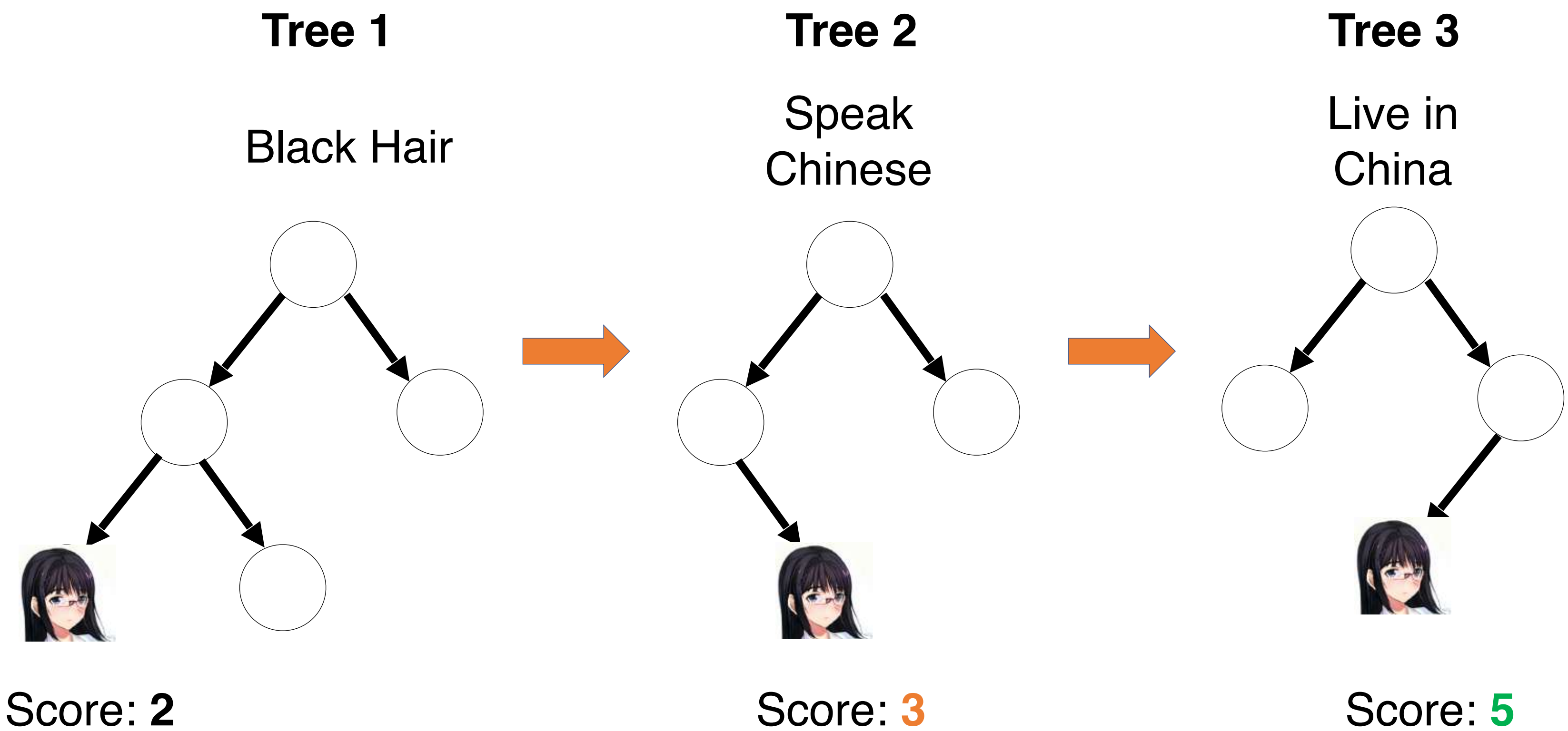
Ensemble of trees: random forest (in parallel)



IsAsian( **) = 3 + 2 + 5**

Low variance

Ensemble: boosted decision tree (in sequential)

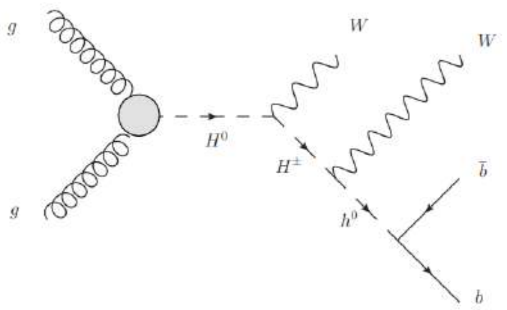


Low bias.

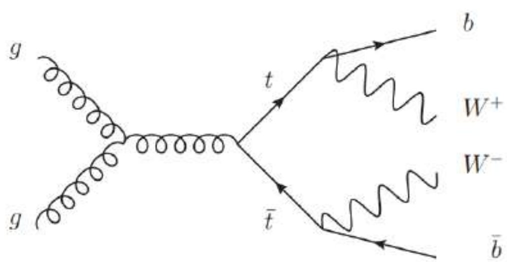
Train to correct the residual of the previous tree.

Higgs identification using deep learning (DL)

Signal



(a)



(b)

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*."

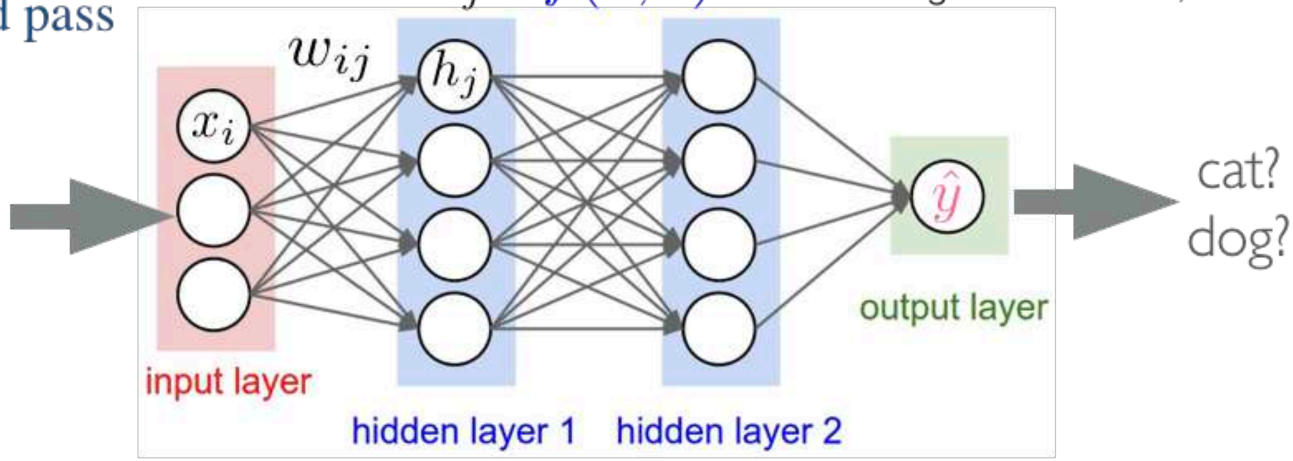
P.Baldi,P.Sadowski,& D.Whiteson, Nature Commun.5, 4308 (2014)

DL: neural net with multiple hidden layers

b_j $\quad f(x, \theta)$

Fig from CS231N, Stanford

Forward pass

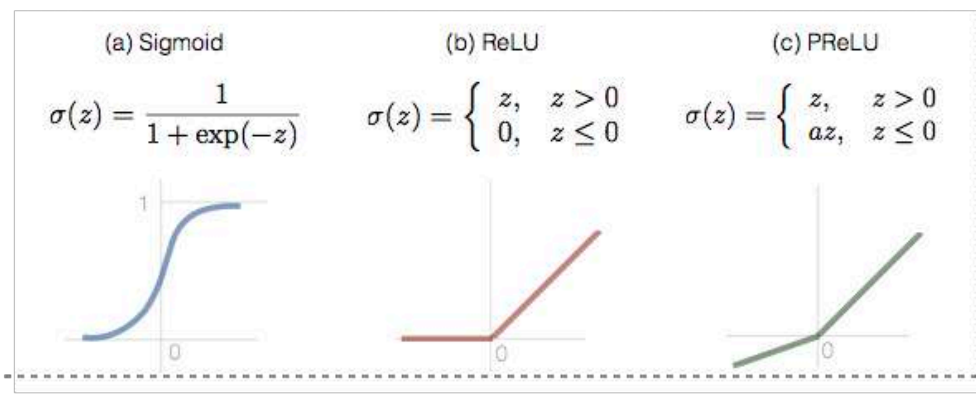


Linear operation

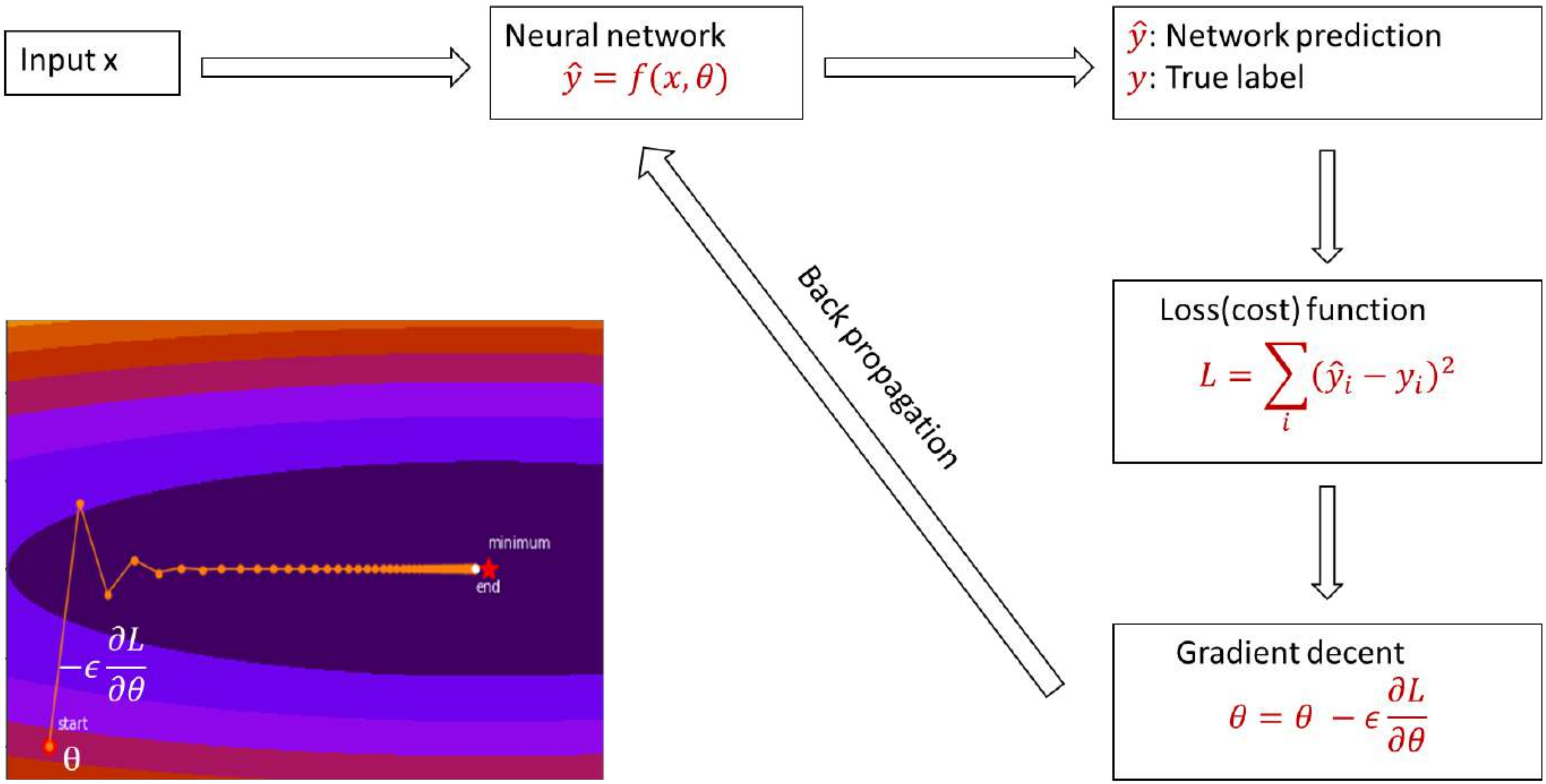
$$z_j = \sum_{i=1}^{N} x_i w_{ij} + b_j$$

scaling, rotating, boosting, changing dimensions

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

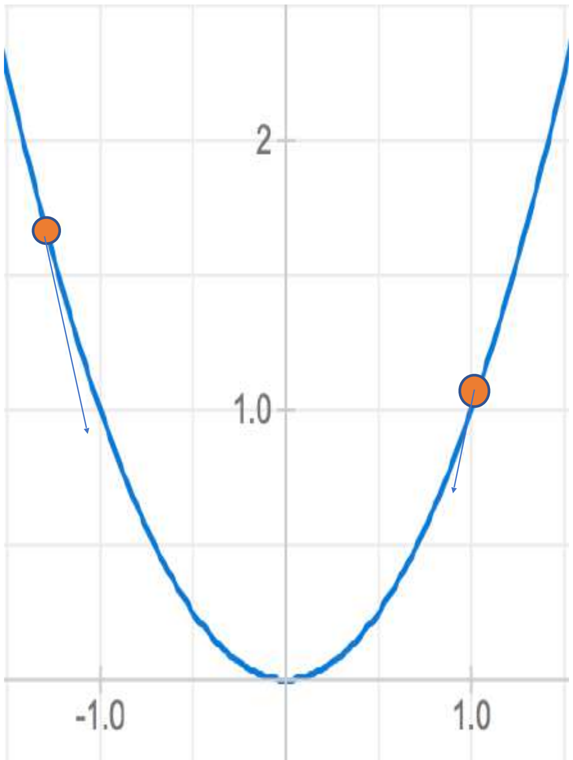


How does deep neural network learn: back propagation



SGD in 1D

$$\theta = \theta - lr \cdot \nabla_\theta L(\theta)$$



Simple example

Equation: $\mathrm{L}(\theta) = \theta^2$

Gradient: $\nabla_\theta \mathrm{L}(\theta) = 2\theta$

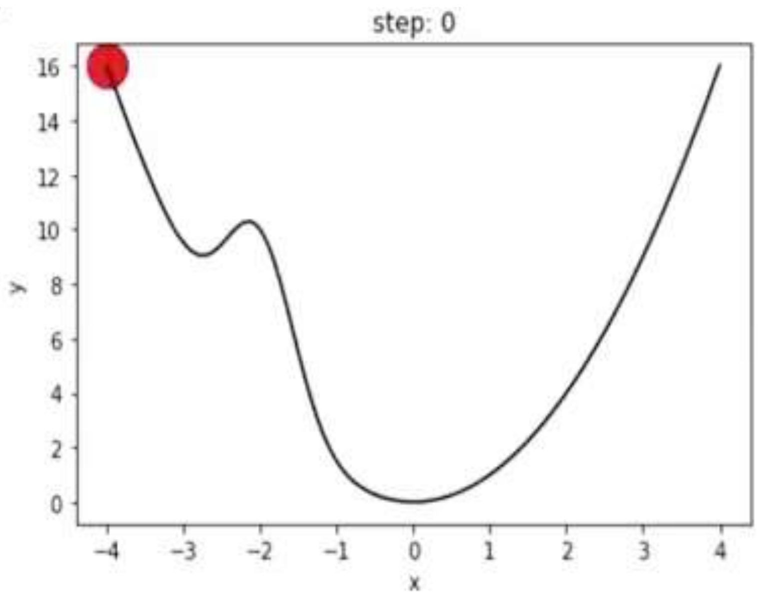
$$-lr \ \nabla_\theta L(\theta) = -2lr\ \theta \begin{cases} > 0, & \theta < 0 \\ < 0, & \theta > 0 \end{cases}$$

lr : learning rate, a small positive number

SGD + Momentum

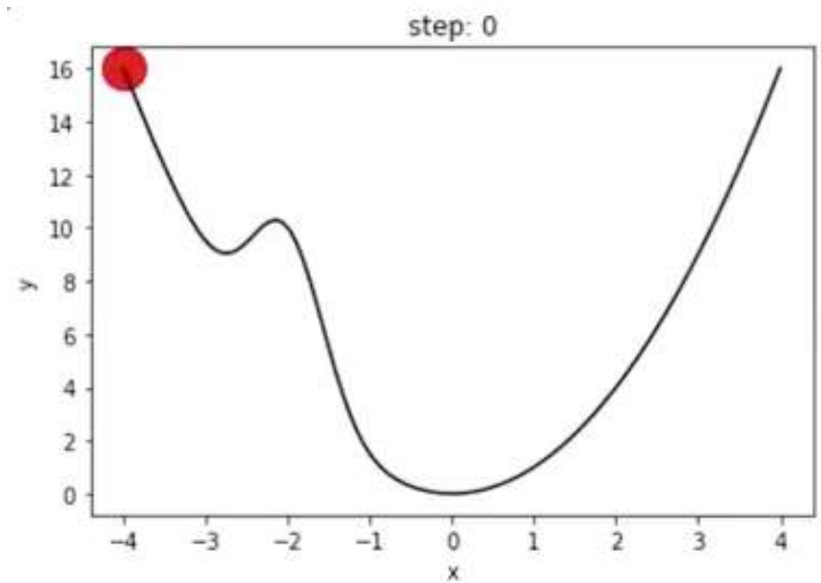
SGD

$$\theta = \theta - lr \cdot \nabla_\theta L(\theta)$$

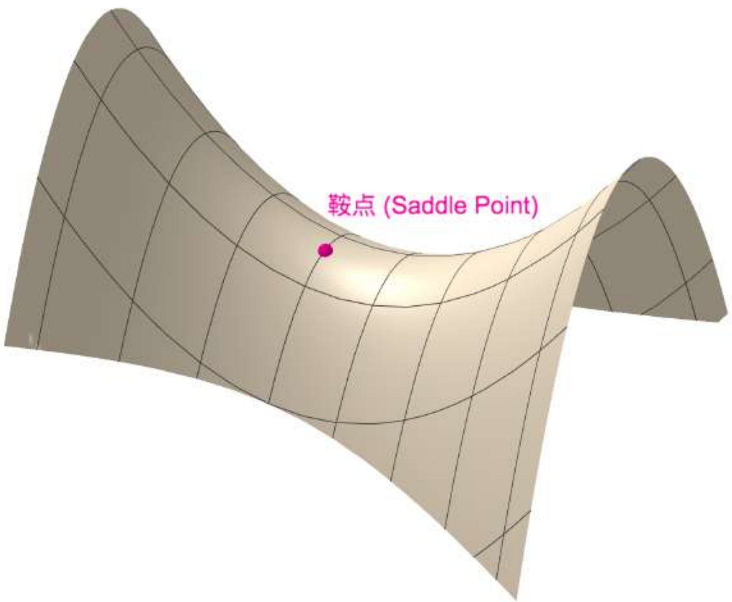


SGD + Momentum

$$v_t = \beta v_{t-1} + lr \nabla_\theta L(\theta)$$
$$\theta = \theta - v_t$$



Trap in local minimum? No



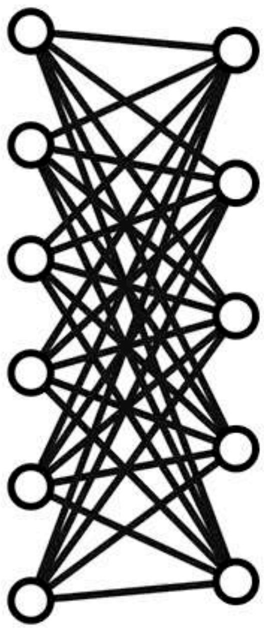
Quora Session: one theoretical puzzle is why the type of non-convex optimization that needs to be done when training deep neural nets seems to work reliably.

LeCun states:
Local minima do not arise in very high dimensional space

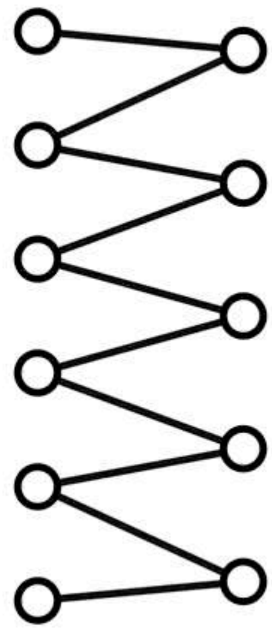
It's hard to build a box in 100 million dimensions.

P(Local Minimum) ~ 0.5^n
where n is the num of trainable parameters
Usually n > 1 million

Convolution Network (reduce parameters)

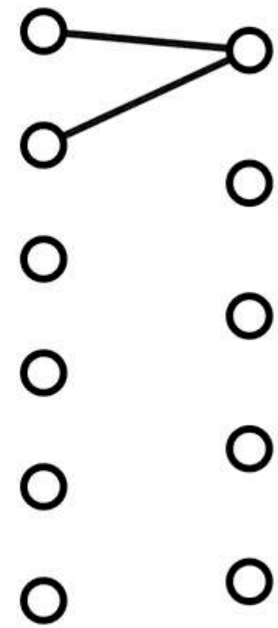


Densely connected



Locally connected

1D convolution

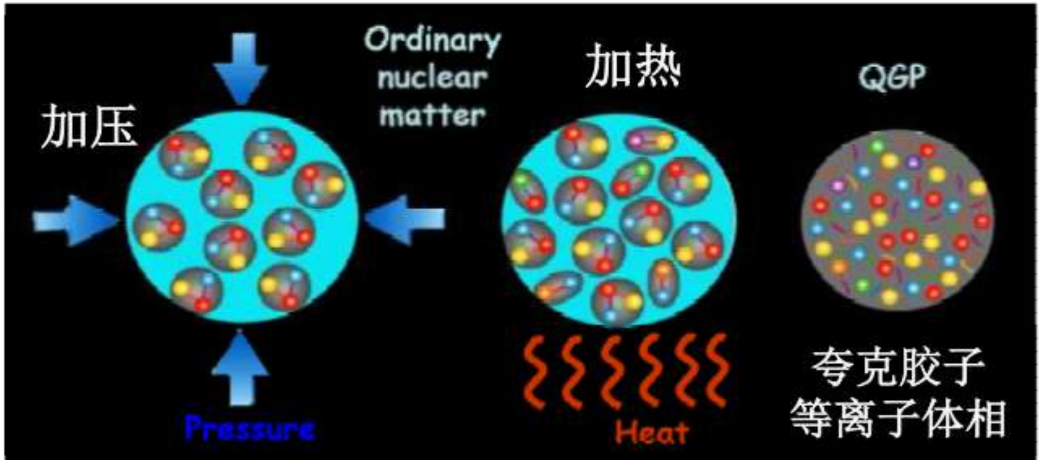


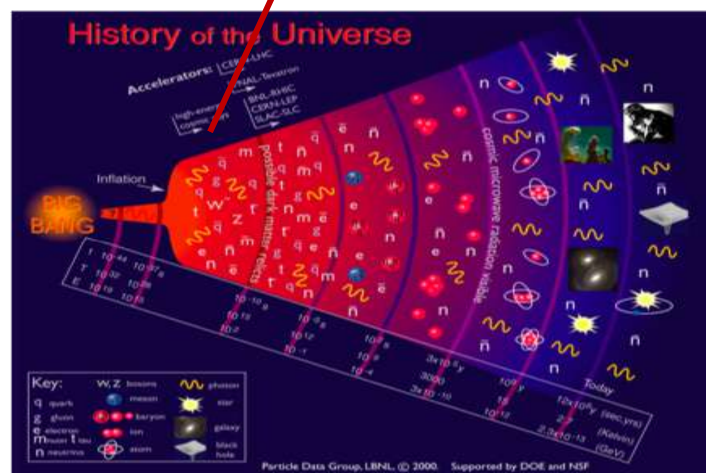
Locally connected
and sharing weights

CONTENTS

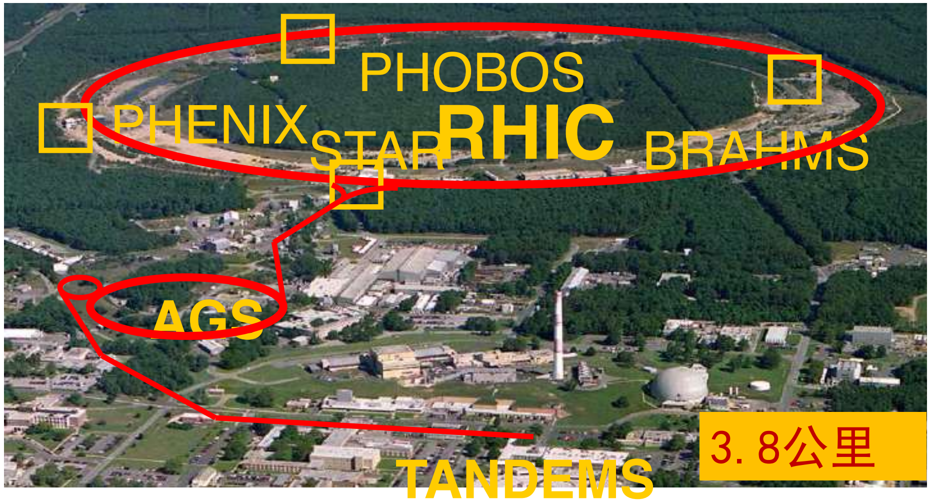
- What’s Machine Learning (ML)
- ML for high energy nuclear physics
- ML for atomic mass prediction
- Outlook： ML for nuclear matrix element
    - Variational wave function using deep learning
    - Quantum-classical hybrid computing

QUARK GLUON PLASMA





相对论重离子碰撞实验



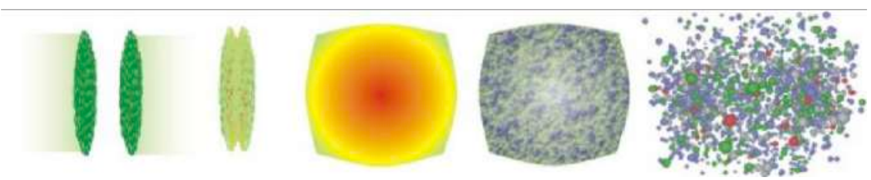
～99.99%光速

RHIC 美国**BNL**国家实验室：

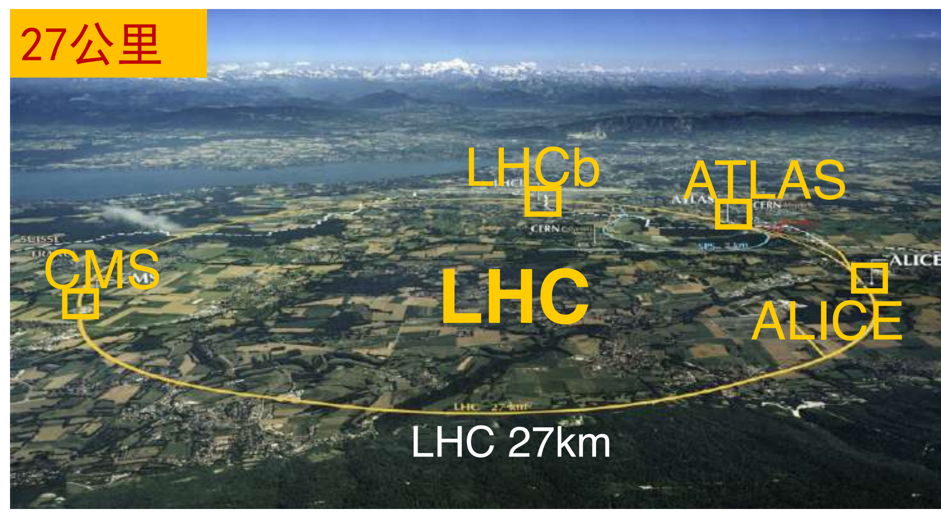
- 30多个国家/地区
- 一千多名科学家和工程师

LHC 欧洲核子中心

- **100**多个国家**/**地区
- 一万多名科学家和工程师

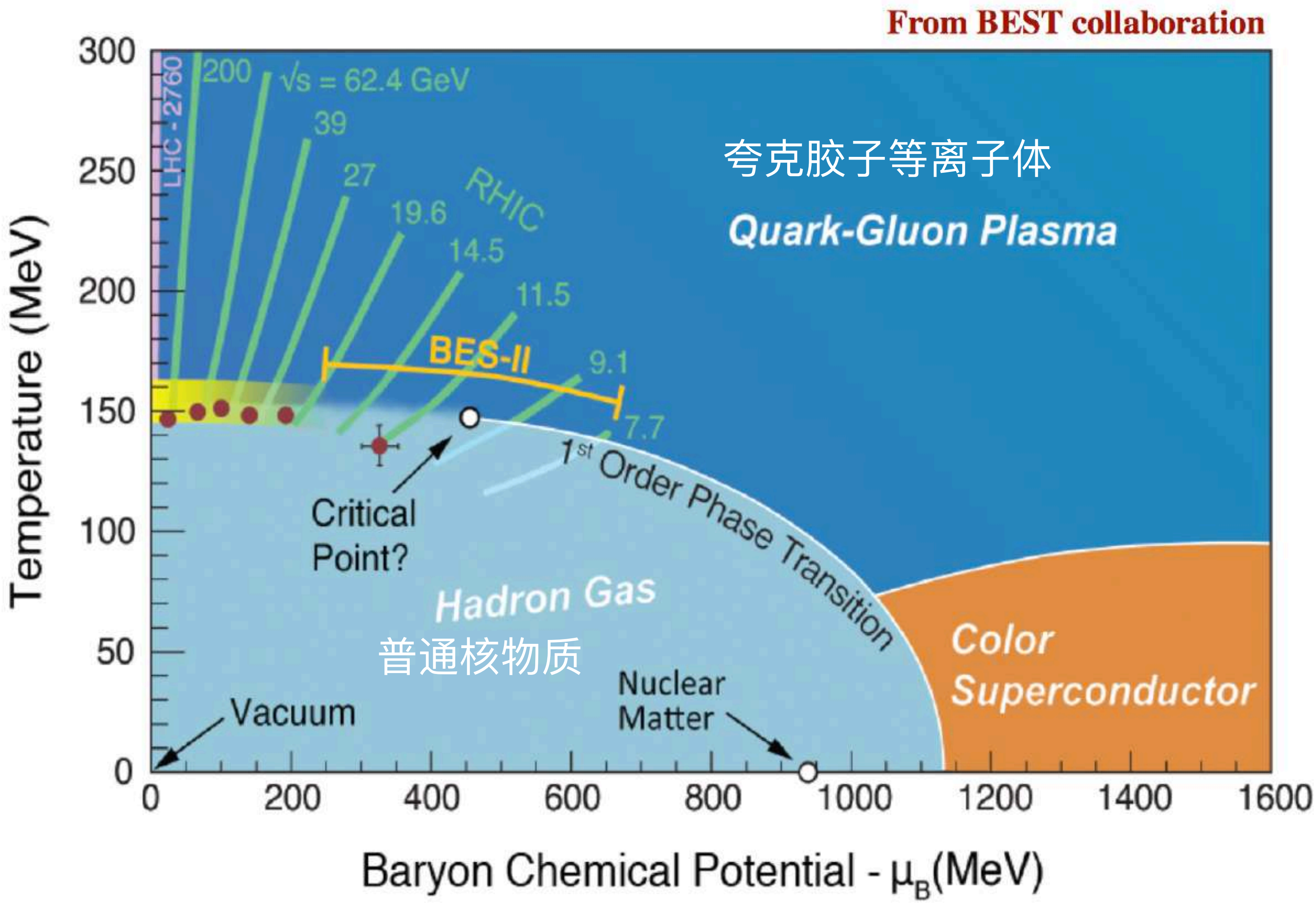


实验主要目标：核物质新形态，核物质相变临界点



～99.9999%光速

QCD Phase diagram

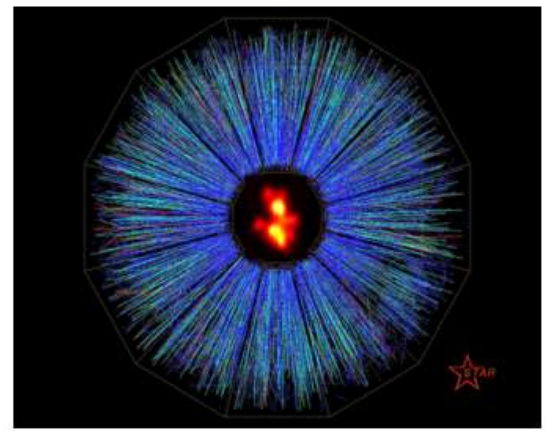


Forward process in high energy nuclear physics

Non-equilibrium dynamical evolution： For given initial condition and nuclear equation of state, solve relativistic hydrodynamics numerically and compare the final hadron spectra from model with experimental measurements.

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

$$T^{\mu\nu} = (\varepsilon + P)u^\mu u^\nu - P g^{\mu\nu} + \pi^{\mu\nu}$$



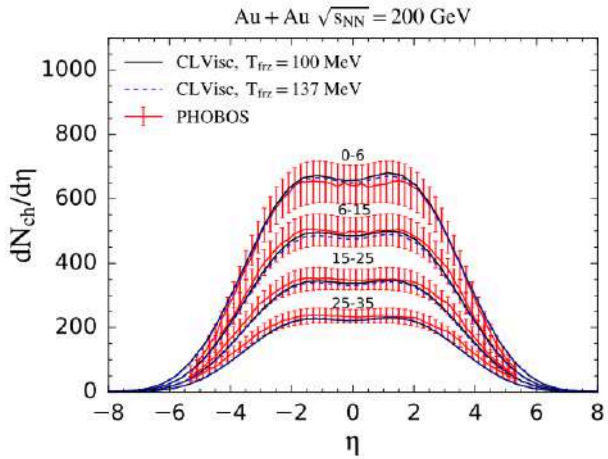
Where $T^{\mu\nu}$ is energy-momentum tensor， ε is energy density， $P$is pressure given by EoS， u^μ is fluid velocity， $g^{\mu\nu}$ is metric and $\pi^{\mu\nu}$ is shear stress tensor.

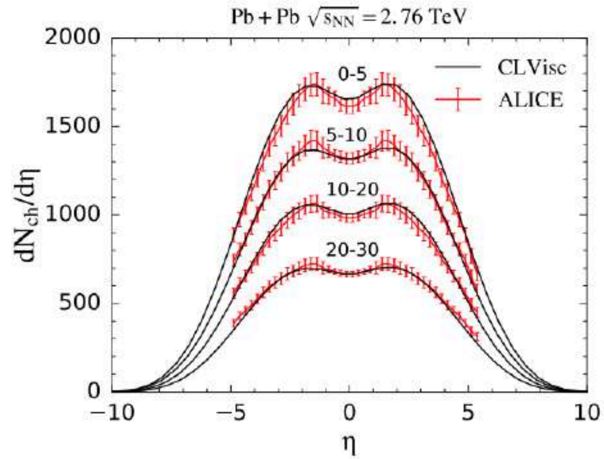
We developed CLVisc which is a (3+1)D viscous hydro parallelized on GPU using OpenCL (100 times speed up)

庞龙刚，H.Petersen,XN Wang, PRC97(2018)no.6,064918

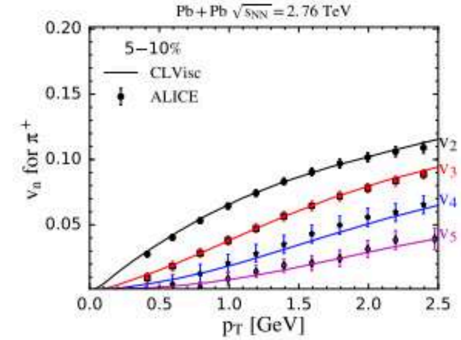
CLVisc vs heavy ion collision data

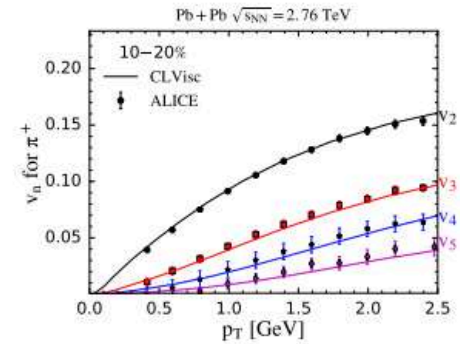
Longitudinal momentum distribution

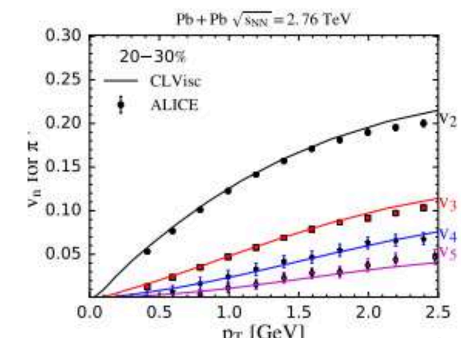




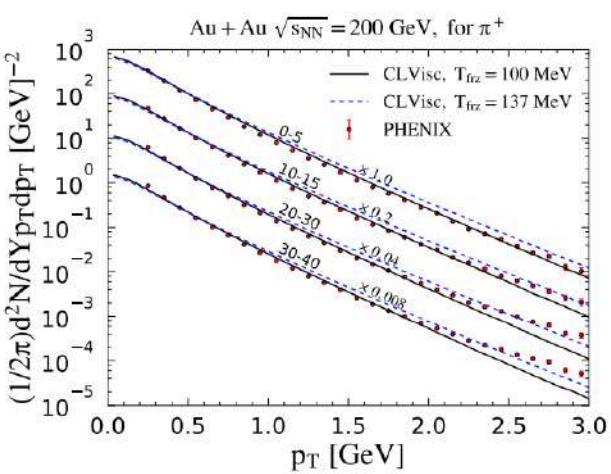
Fourier decomposition coef.
for azimuthal angle

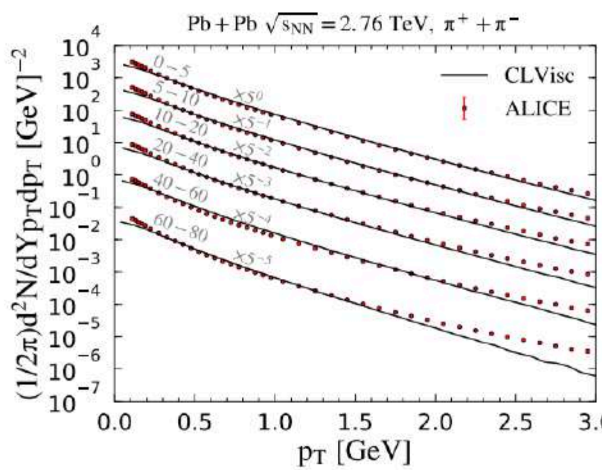






Transverse momentum distribution





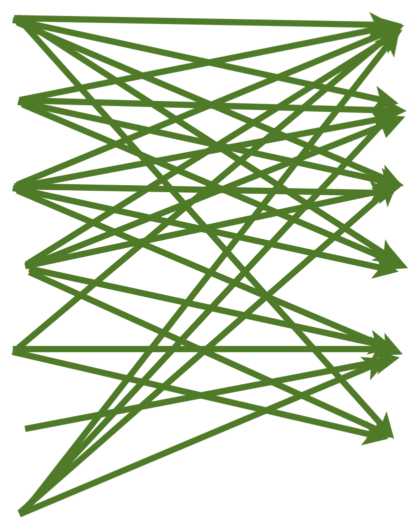
Inverse problem：decode QCD EoS and initial state from data

模型参数

人工特征

Model Parameter:

eqn. of state
shear viscosity
initial state
pre-equilibrium dynamics
thermalization time
quark/hadron chemistry
particlization/freeze-out

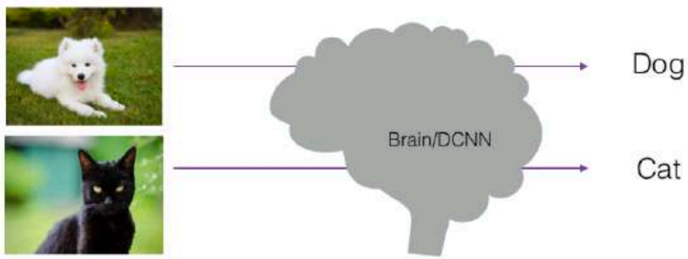


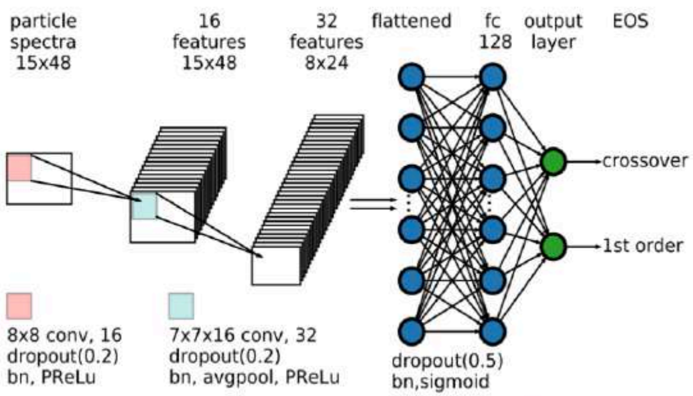
experimental data:

π/K/P spectra
yields vs. centrality & beam
elliptic flow
HBT
charge correlations & BFs
density correlations

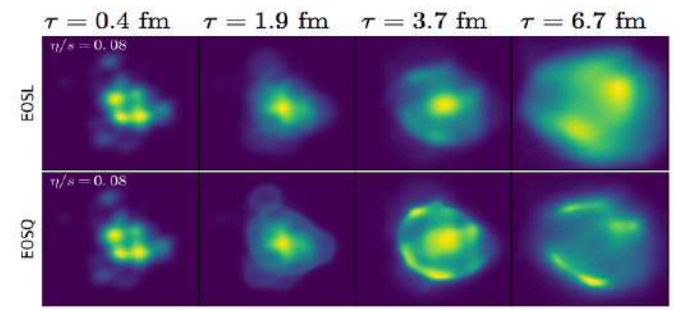
1. Entangled features and physical parameters
2. Degenerate output

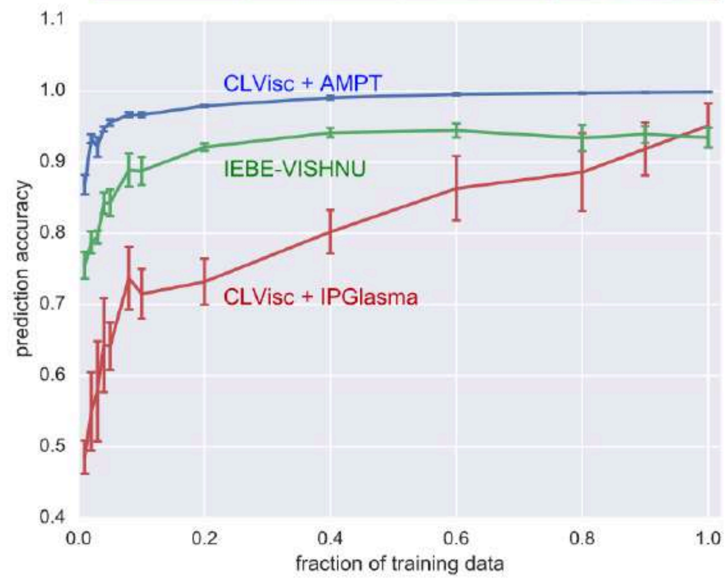
Deep Learning for nuclear EoS





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



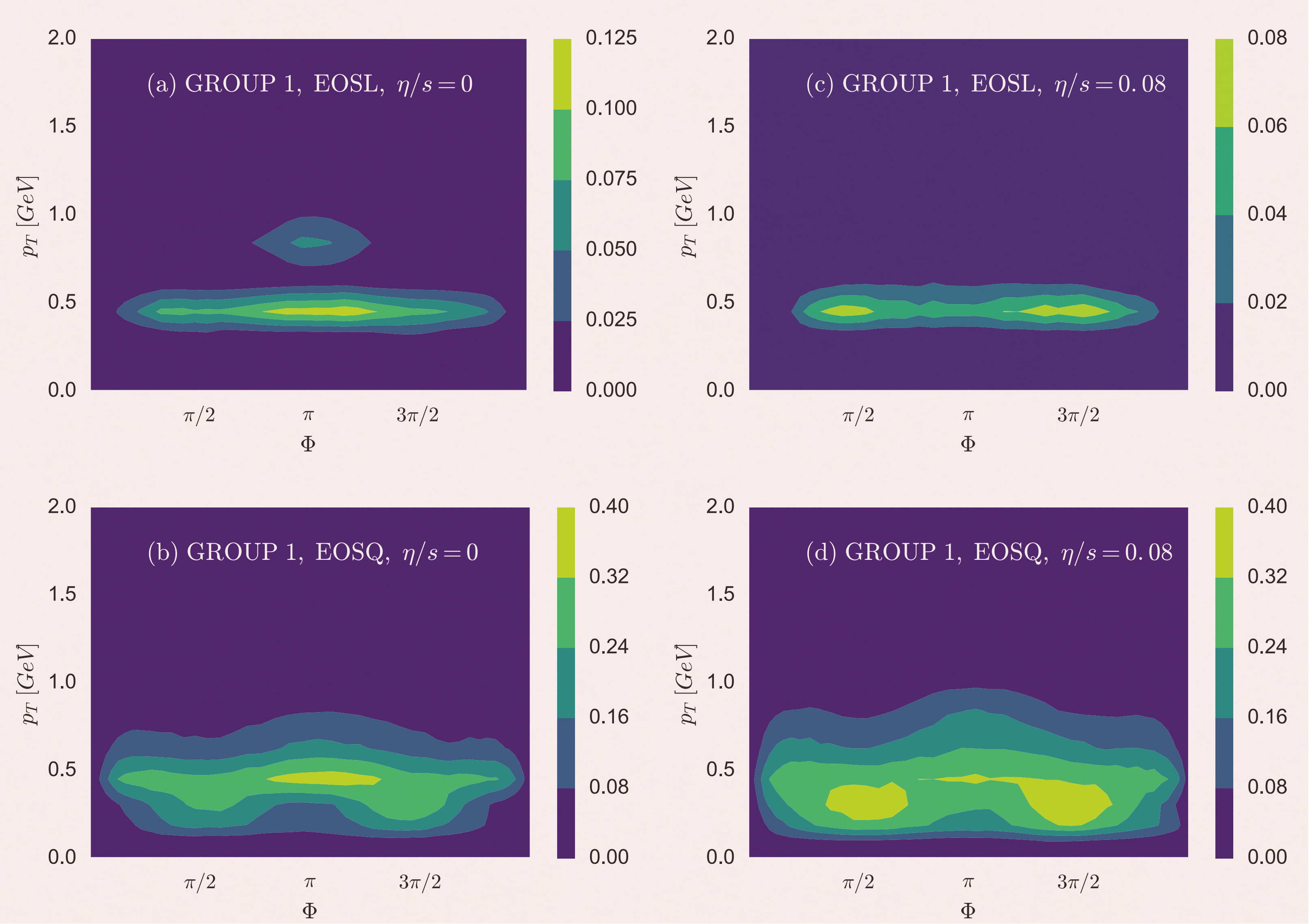


(1) Nuclear EoS is encoded in the final state output

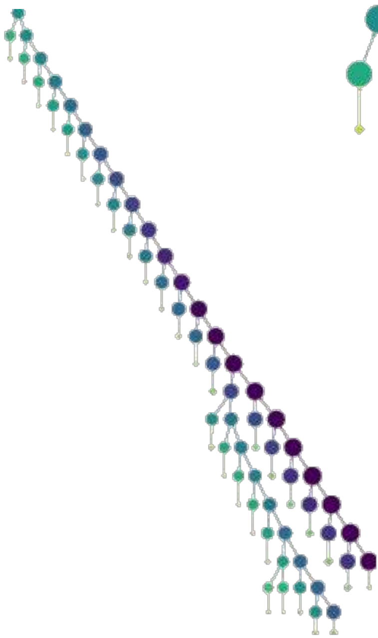
(2) Deep learning helps to decode this signal

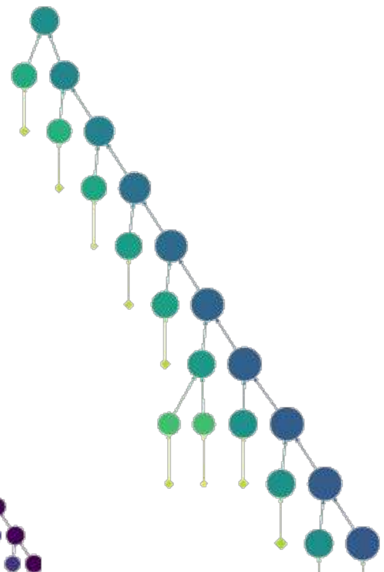
Long Gang Pang, K. Zhou, N. Su, H. Petersen, H. Stoecker, X-N Wang, Nature communications (2018)

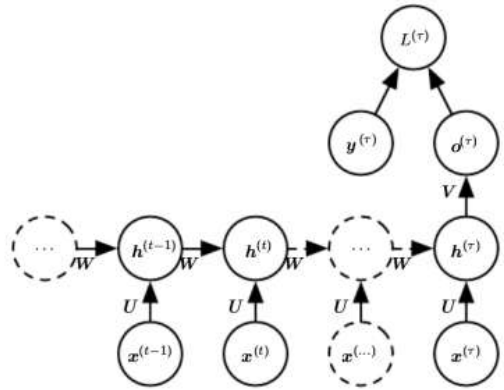
ML interpretability：the most important region for EoS classification

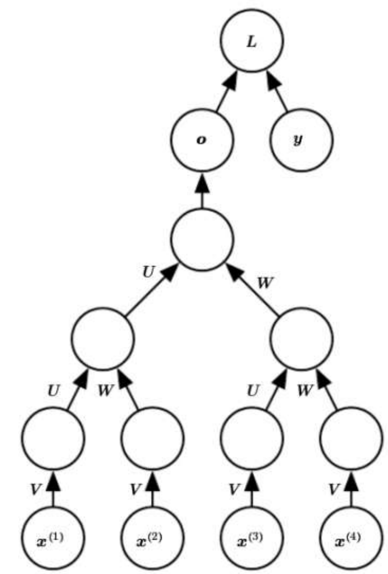


Optimal network for given data structure







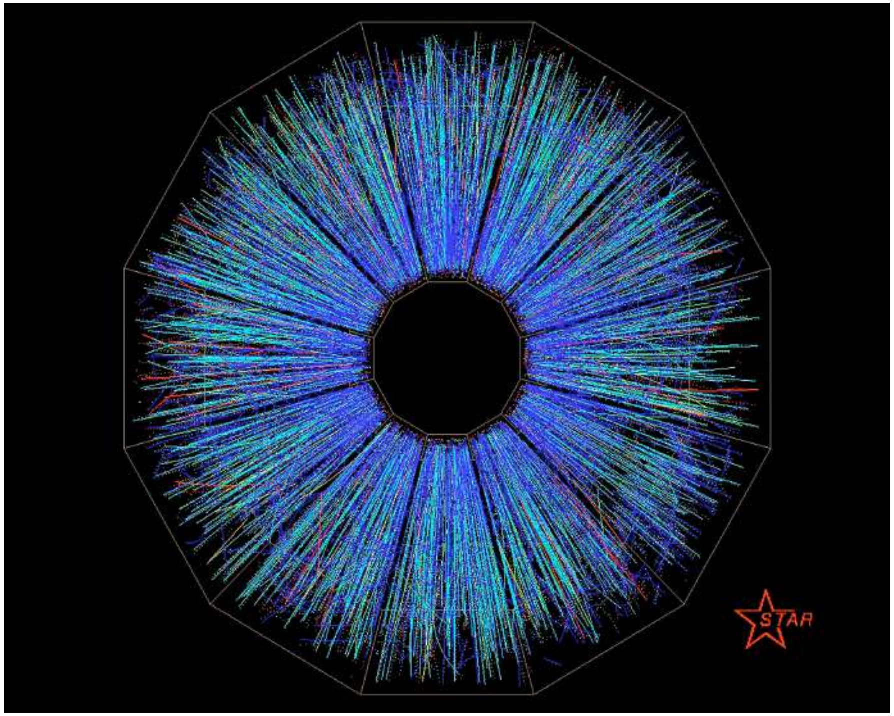


Gluon 喷注 **Quark 喷注** **Recurrent Net** **Recursive Net**

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

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

Better network architecture for particles



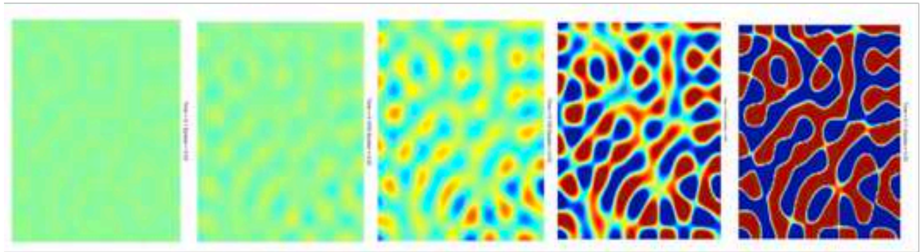
- Images: histograms
    - (px, py) or (pt, phi)
    - (px, py, pz)
    - (pt, phi, eta)
- Point cloud: particle list

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
…				

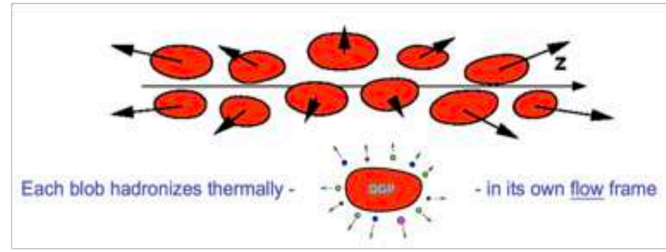
Convolution network is optimal for images
Point Cloud network is better for particle in momentum space

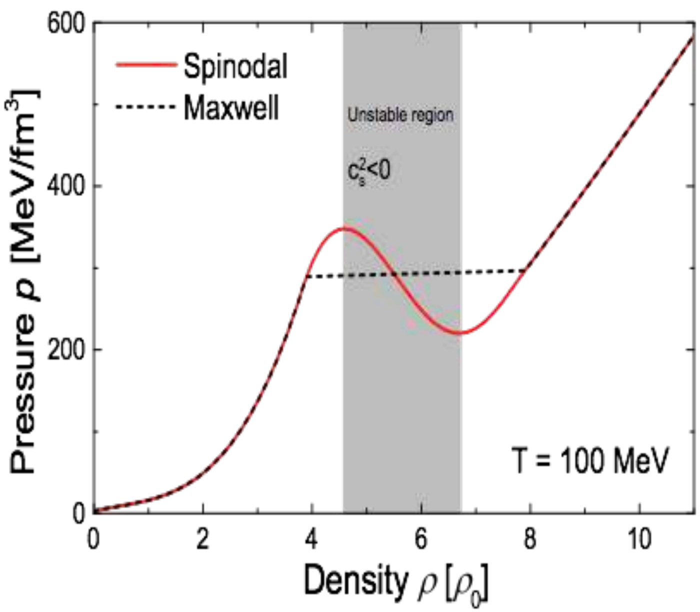
Point cloud network for EoS classification

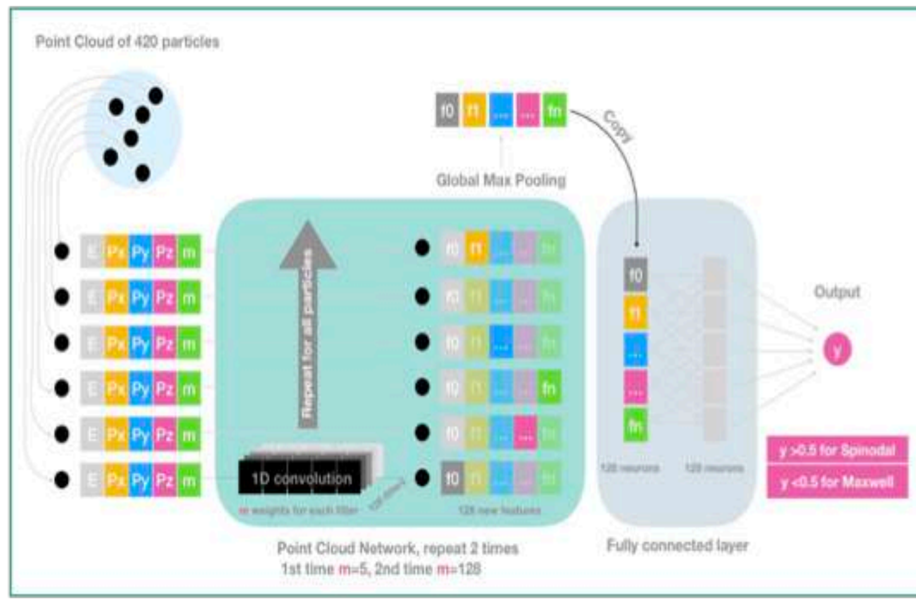
wiki



J. Randrup, INT 2009







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

Dynamical edge convolution to capture more local correlations

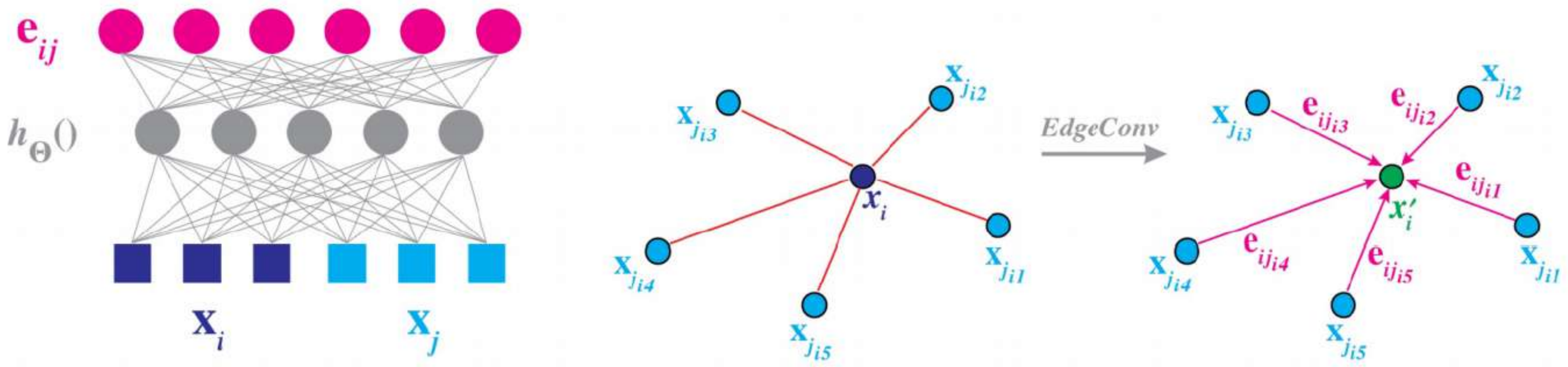


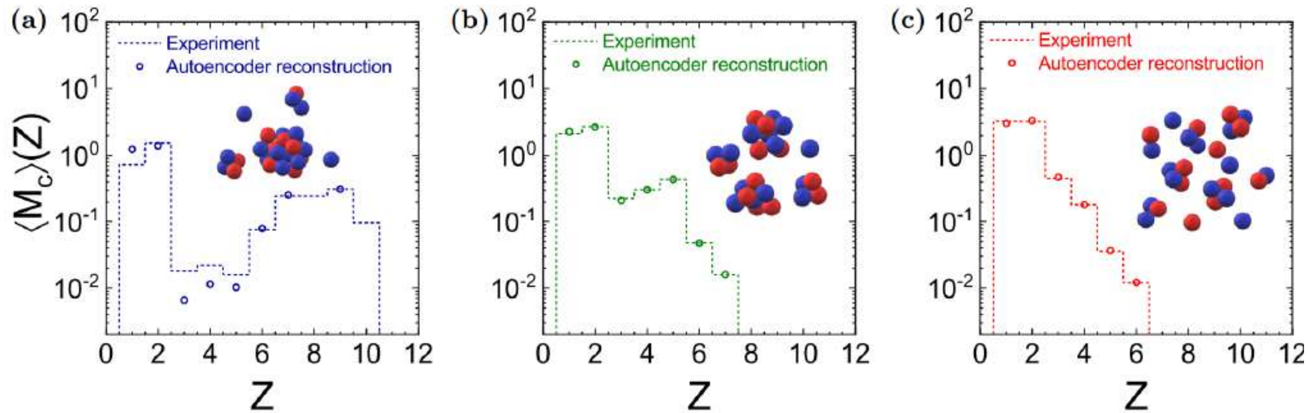
Fig. 2. **Left**: Computing an edge feature, $\mathbf{e}_{ij}$ (top), from a point pair, $\mathbf{x}_i$ and $\mathbf{x}_j$ (bottom). In this example, $h_\Theta()$ is instantiated using a fully connected layer, and the learnable parameters are its associated weights. **Right**: The EdgeConv operation.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

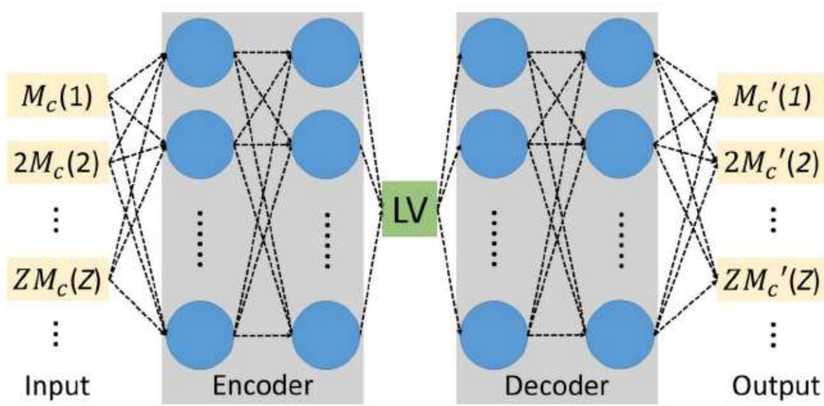
Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E. Sarma, Michael M. Bronstein, and Justin M. Solomon

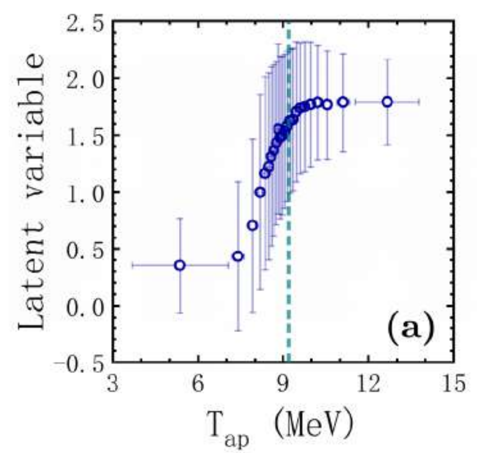
Latent Variable as liquid-gas phase transition order parameter

Nuclear liquid-gas phase transition with machine learning

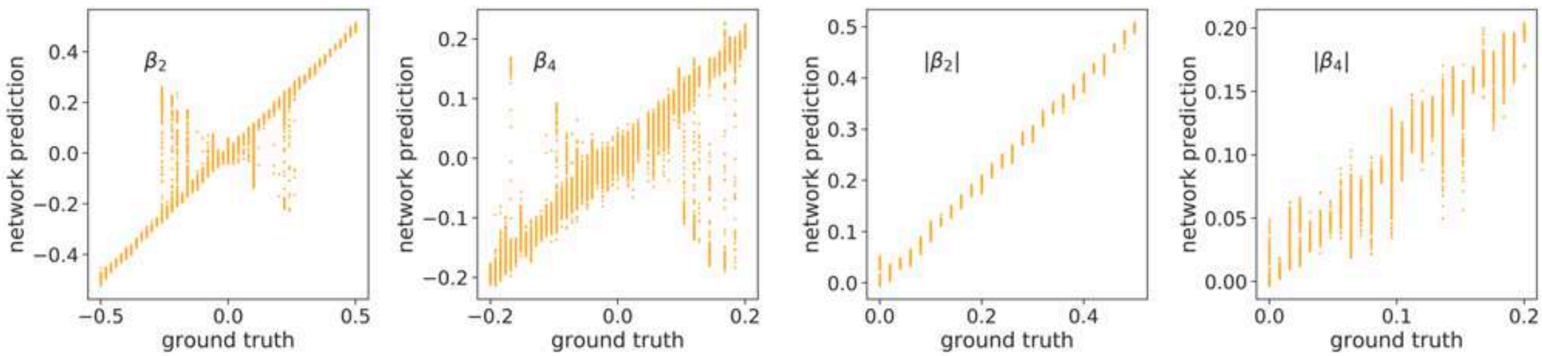
Rui Wang,[1,2,*] Yu-Gang Ma,[1,2,†] R. Wada,[3] Lie-Wen Chen,[4] Wan-Bing He,[1] Huan-Ling Liu,[2] and Kai-Jia Sun[3,5]
[1] *Key Laboratory of Nuclear Physics and Ion-beam Application (MOE), Institute of Modern Physics, Fudan University, Shanghai 200433, China*

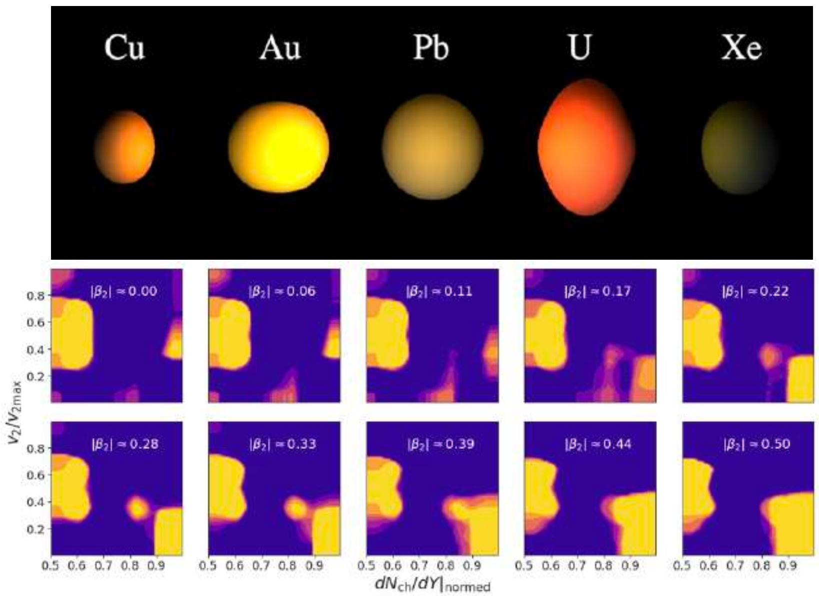


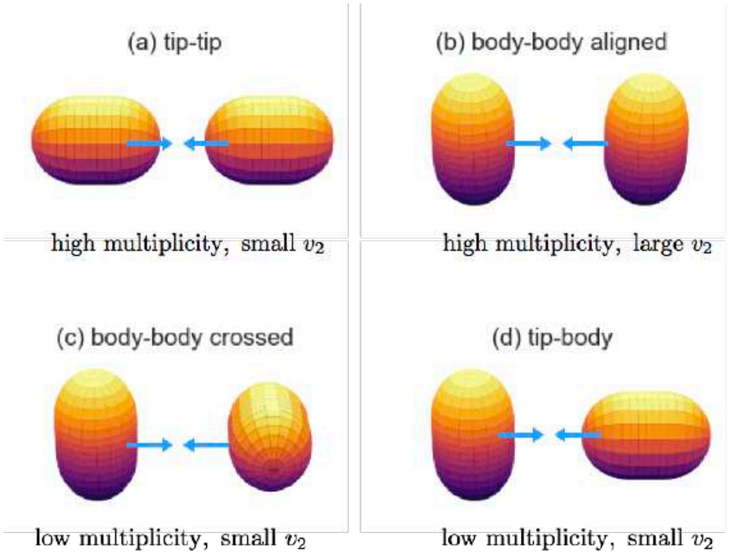




Nuclear deformation using HIC







34-layer residual network predicts the absolute value of nuclear deformation

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

Deep learning relativistic hydro

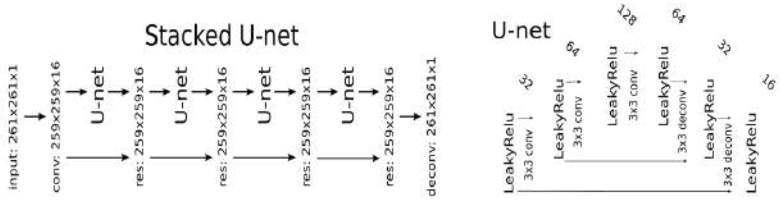
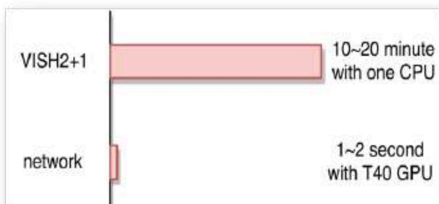
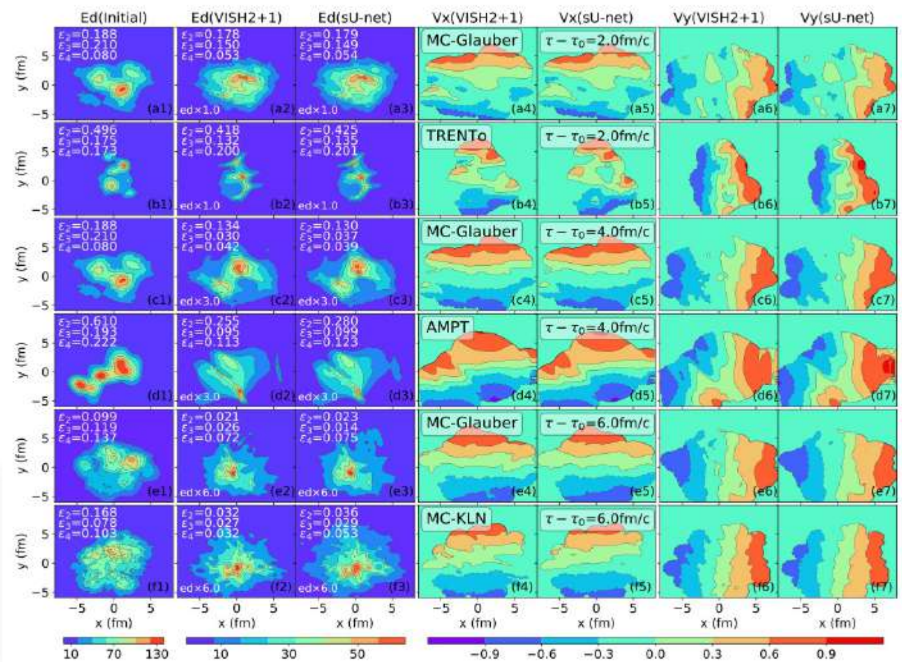


FIG. 1: An illustration of the encode-decode network, stacked U-net, which consists of the input and out layers and four residual U-net blocks. The right figure shows the U-net structure, and the depth of the hidden layer is written on the top of them.

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

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

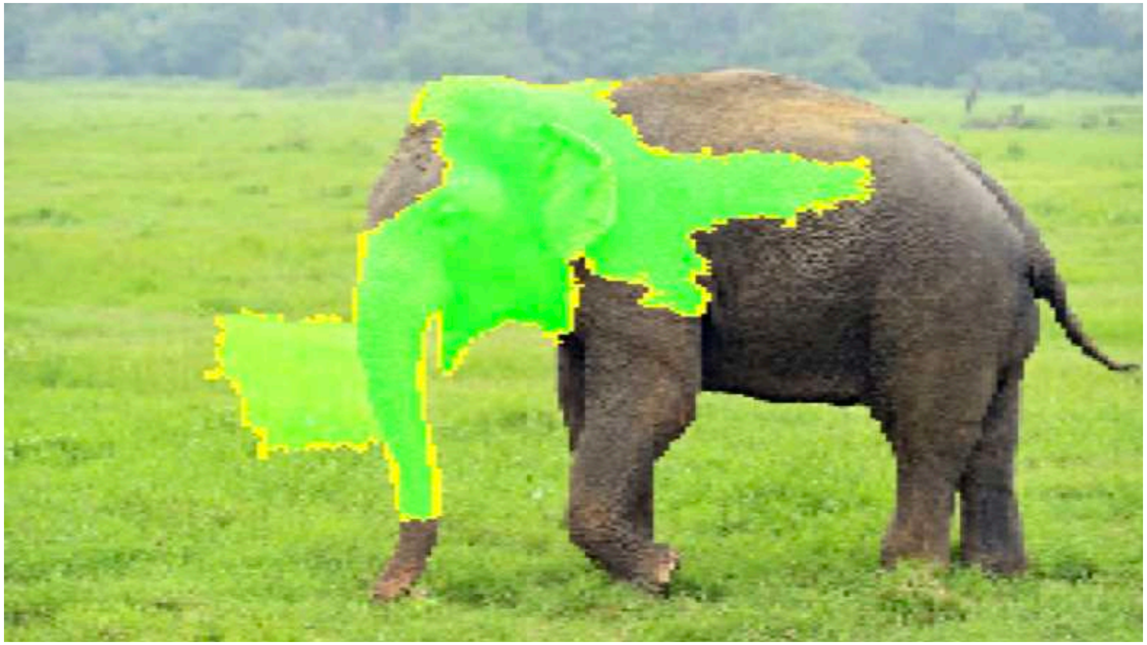




600 times speed up .vs. 60 times speed up on GPU

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

ML interpretability：1. ablation



将输入图像中的像素或超像素替换掉, 输入神经网络，观察网络预测结果的变化，按变化幅度制作重要性地图。

E.g., LIME, prediction difference analysis

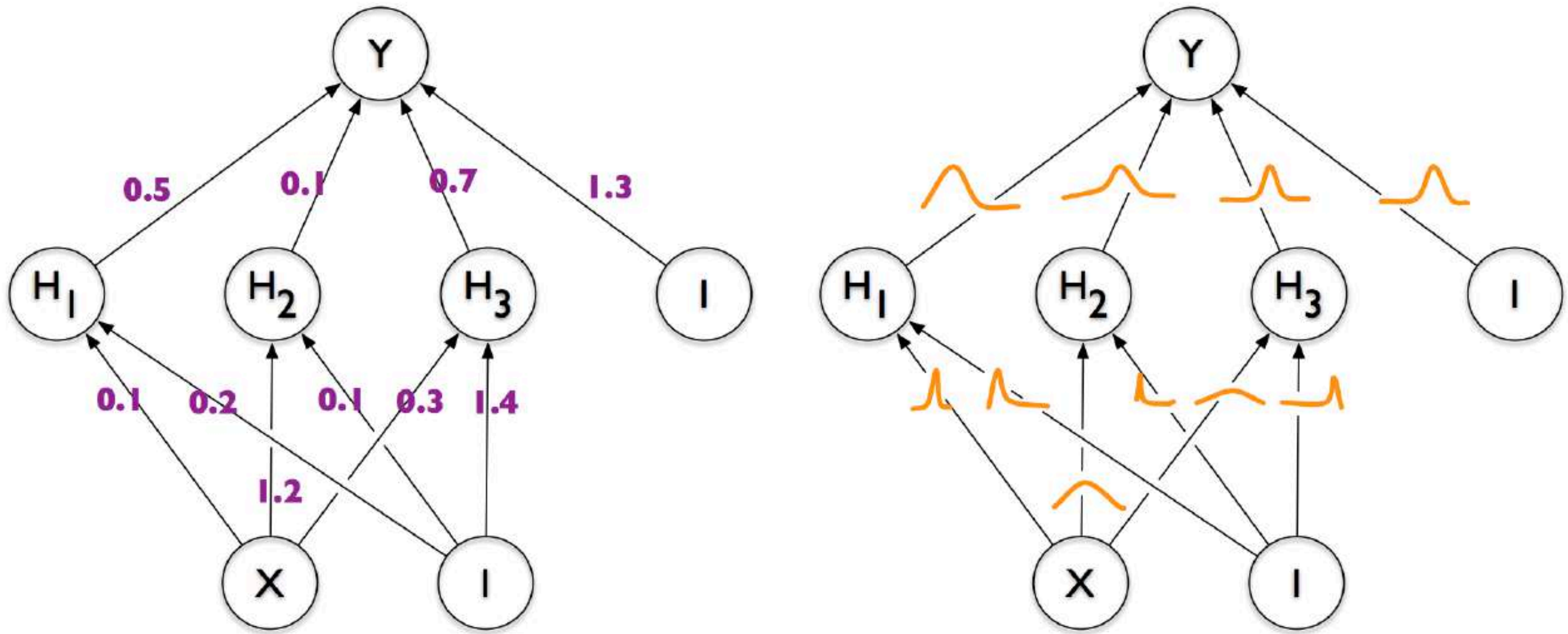
Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "why should I trust you?"

ML interpretability：2. class activation map

将最后一个卷积层（学到了抽象、整体的特征）的特征地图映射到输入图像中

B. Zhou, A. Khosla, Lapedriza. A., A. Oliva, and A. Torralba. Learning Deep Features for Discriminative Localization. *CVPR*, 2016.

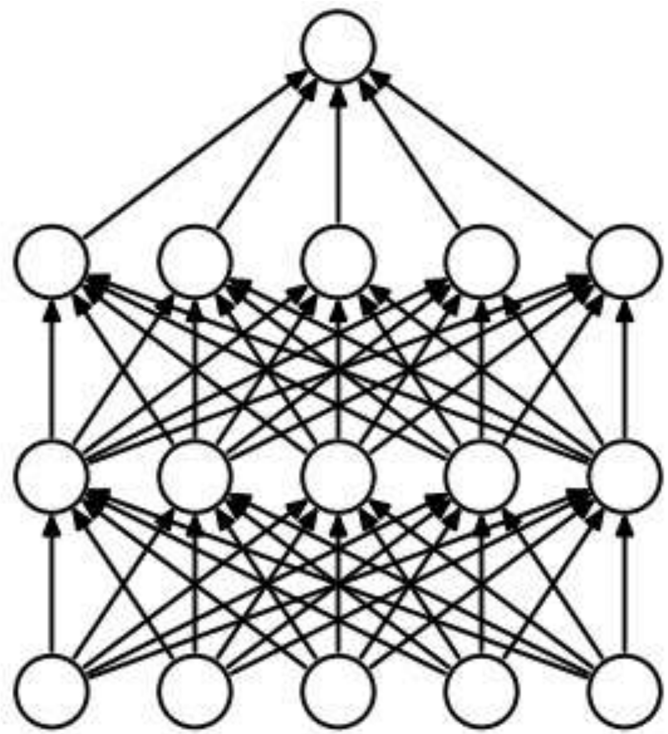
Uncertainty measure：1. Bayes Neural Network



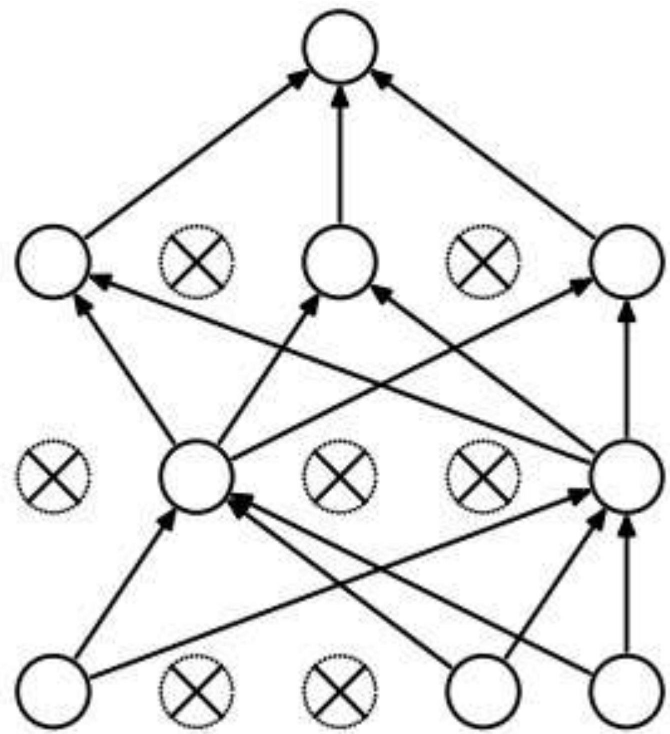
Weight Uncertainty in Neural Networks 2015

Replace weights with distributions，to get ensemble of infinity number of networks

Uncertainty measure：2. Monte Carlo Dropout



(a) Standard Neural Net



(b) After applying dropout.

Apply dropout during both training and testing.

Ensemble of networks through dropout

CONTENTS

- What’s Machine Learning (ML)
- ML for high energy nuclear physics
- ML for atomic mass prediction
- Outlook： ML for nuclear matrix element
  - Variational wave function using deep learning
  - Quantum-classical hybrid computing

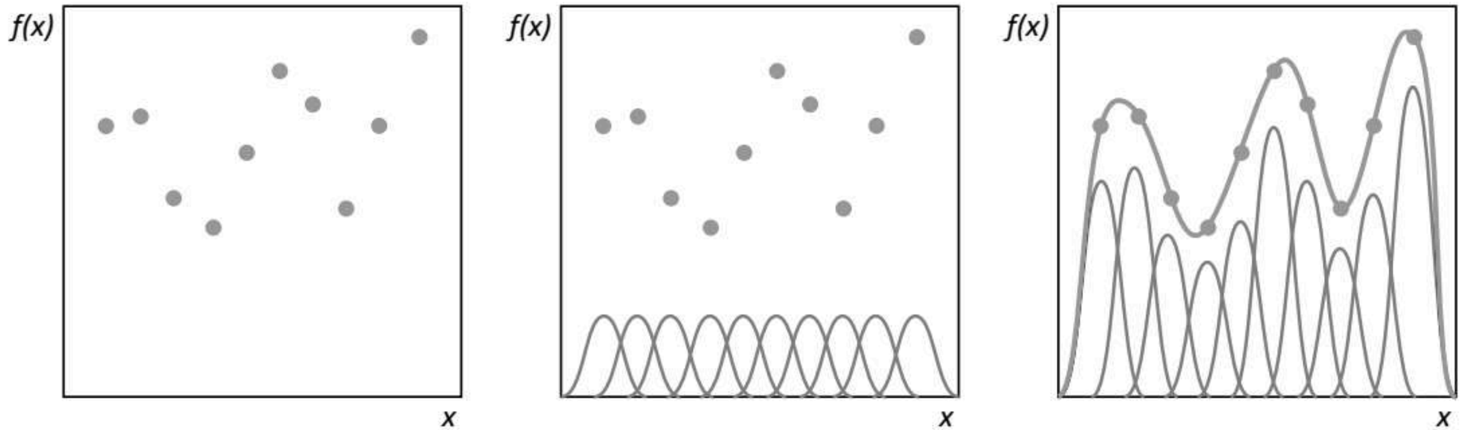
Nuclear mass prediction

- **WS4+RBF (RMS 170 keV):** **Ning Wang, Min Liu, Xizhen Wu and Jie Meng, Phys. Lett. B 734 (2014) 215**

TABLE II: Rms deviations between data and predictions from the WS4 formula (in keV). The line $\sigma(M)$ refers to all the 2353 measured masses in AME2012, the line $\sigma(M_{\rm new})$ to the measured masses of 219 "new" nuclei in AME2012, the line $\sigma(M_{0.1})$ to the masses of 286 nuclei with $|I-I_0|>0.1$, the line $\sigma(S_n)$ to all the 2199 measured neutron separation energies S_n, the line $\sigma(Q_\alpha)$ to the α-decay energies of 46 super-heavy nuclei ($Z \geq 106$) [14]. The corresponding results of WS3 model are also presented for comparison. WS4$^{\mathbf{RBF}}$ denotes that the radial basis function (RBF) corrections [46] are combined in the WS4 calculations.

	WS3	WS4	WS4$^{\mathbf{RBF}}$
$\sigma(M)$	335	298	170
$\sigma(M_{\rm new})$	424	346	155
$\sigma(M_{0.1})$	516	444	215
$\sigma(S_n)$	273	258	251
$\sigma(Q_\alpha)$	248	238	237

RBF/KRR for nuclear mass prediction



- Radial Basis Function is a powerful way for function interpolation
- Usually RBF are trained on mass residual

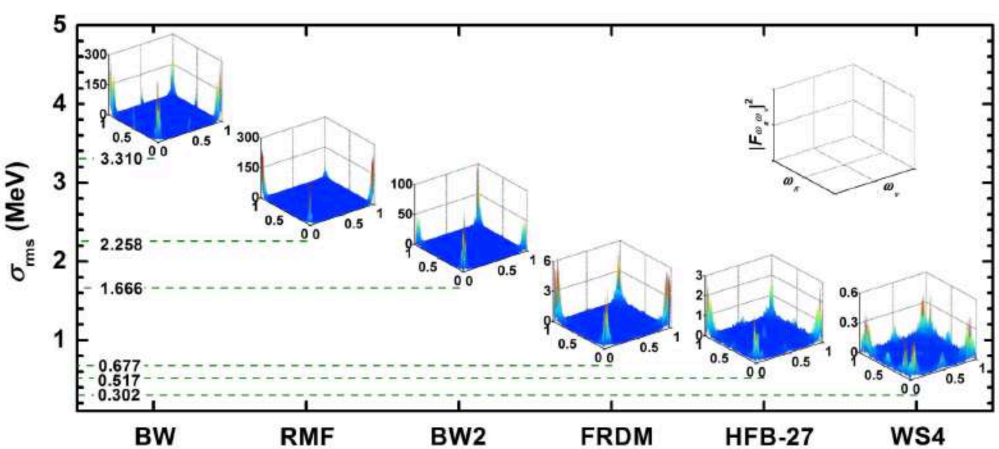
$$\delta(Z,N) = M_{\rm exp} - M_{\rm th}$$

WS4 + RBF + FOURIER TRANSFORM

High precision nuclear mass predictions towards a hundred kilo-electron-volt accuracy

Zhongming Niu[a], Haozhao Liang[b], Baohua Sun[c], Yifei Niu[d], Jianyou Guo[a], and Jie Meng[c,e,f]

$$F_{kl} = \frac{1}{Z_\mathrm{m} N_\mathrm{m}} \sum_{Z=8}^{Z_\mathrm{m}} \sum_{N=8}^{N_\mathrm{m}} (M_\mathrm{exp}^{Z,N} - M_\mathrm{th}^{Z,N}) e^{-i2\pi\left[\frac{(k-1)(Z-1)}{Z_\mathrm{m}} + \frac{(l-1)(N-1)}{N_\mathrm{m}}\right]}, \quad (1)$$



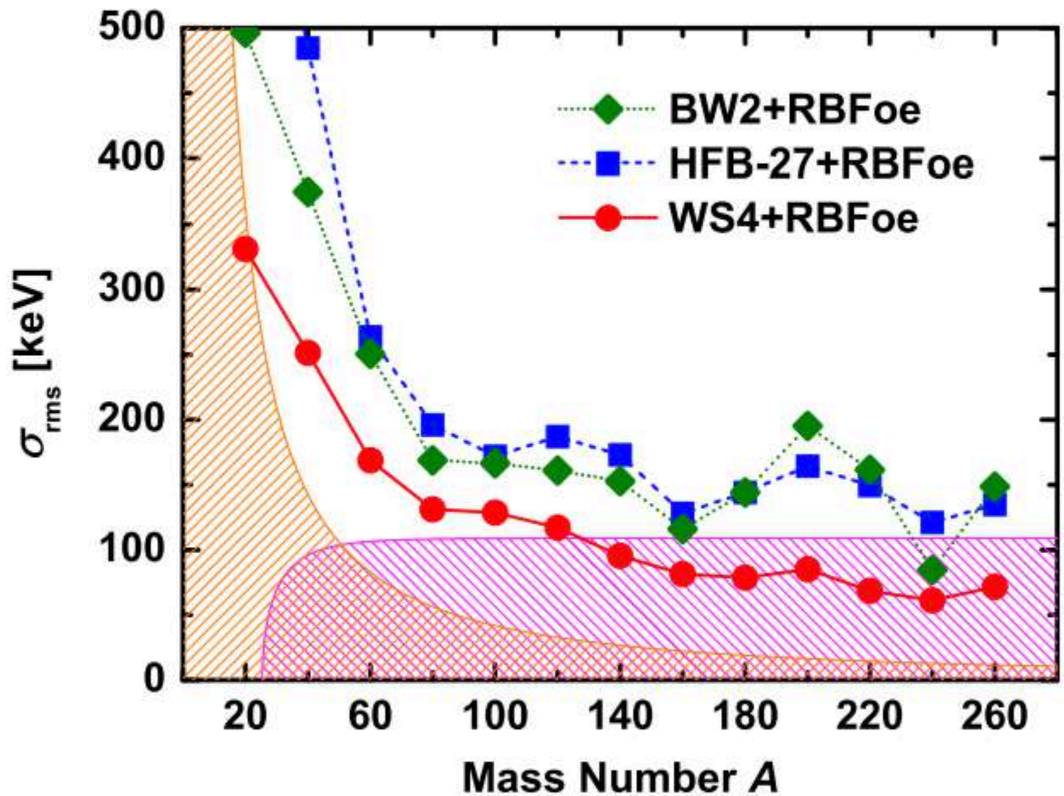
1. The deviations are large for low Z and N frequencies for all semi-empirical models, corrected by RBF

2. High frequency part for odd-even nuclei due to pairing effect, difficult for RBF

WS4 + RBF + Odd-even correction: RMS 138 keV

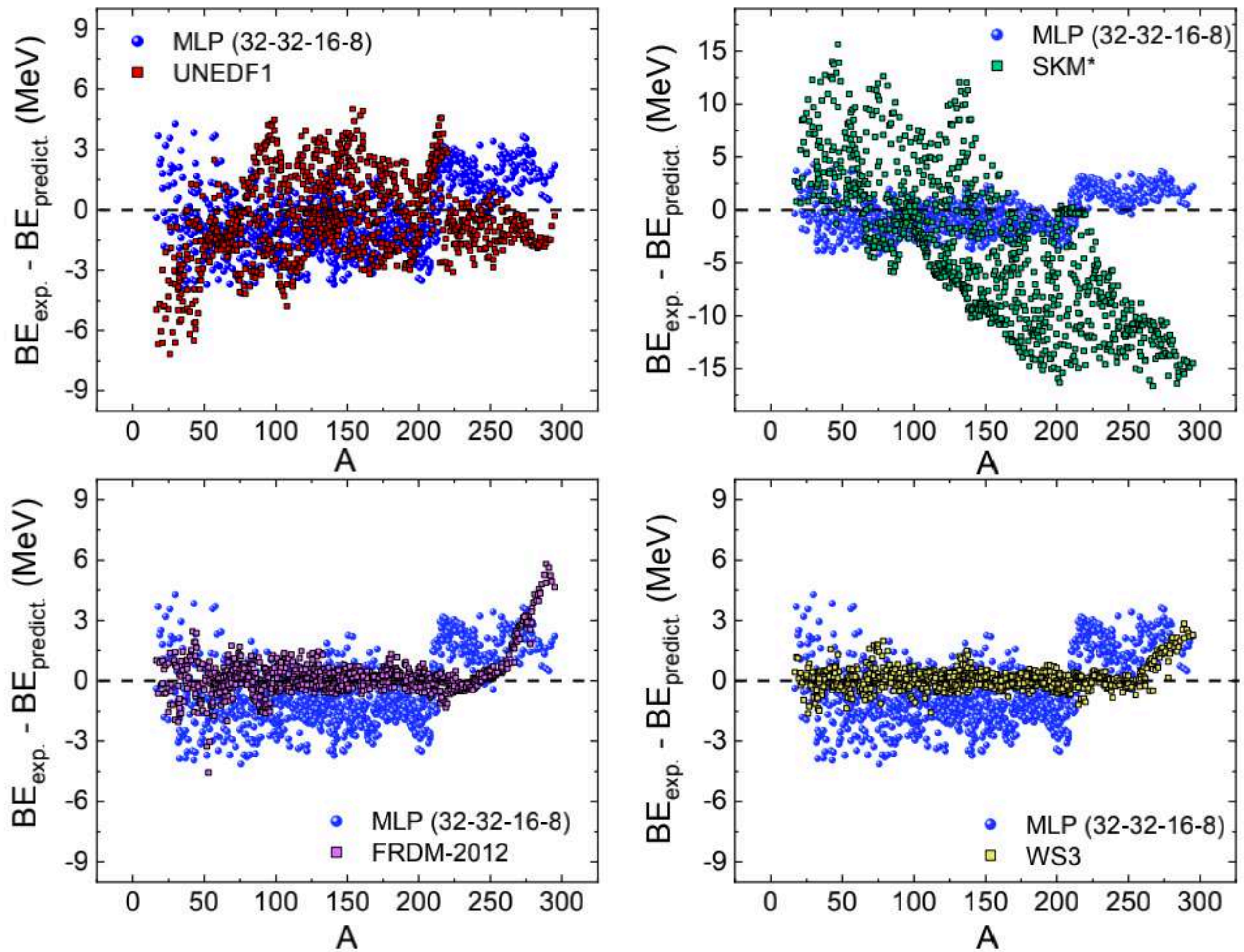
High precision nuclear mass predictions towards a hundred kilo-electron-volt accuracy

Zhongming Niu[a], Haozhao Liang[b], Baohua Sun[c], Yifei Niu[d], Jianyou Guo[a], and Jie Meng[c,e,f]



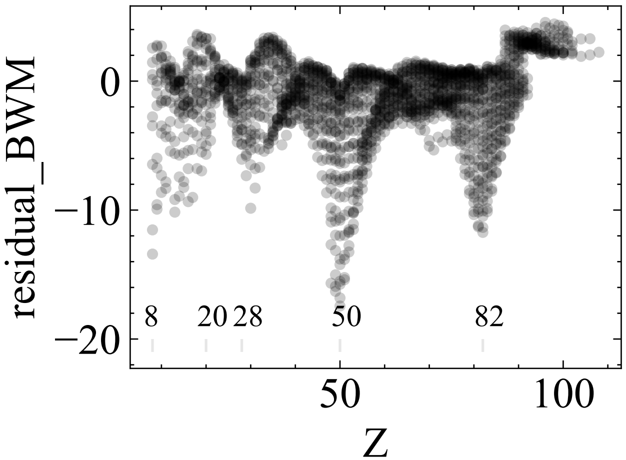
Odd-even staggering of nuclear mass are high frequency, considered separately.

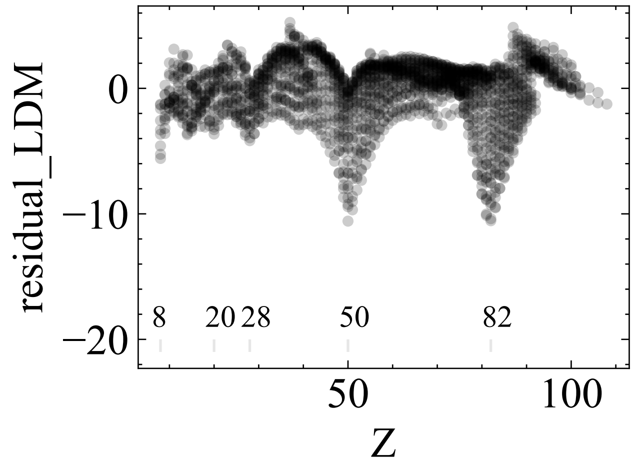
MLP (4-hidden layers)

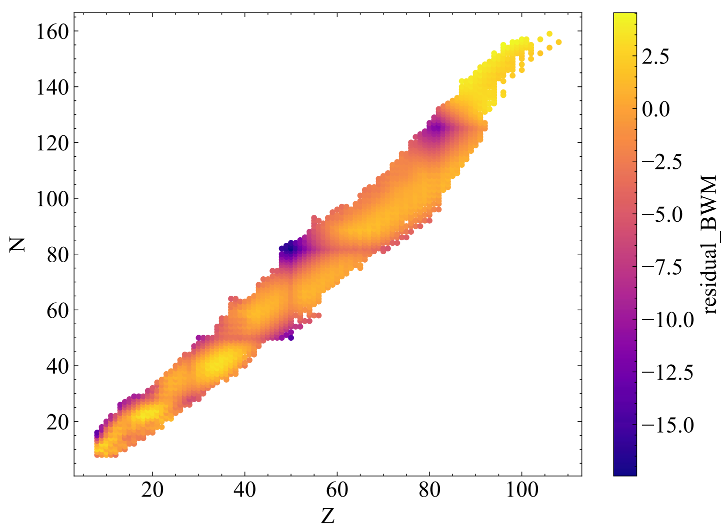


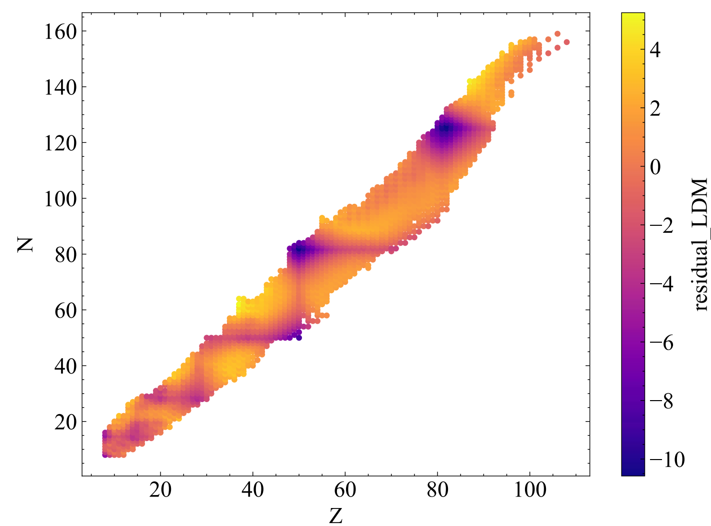
Nuclear mass predictions using neural networks: application of the multilayer perceptron (2021): RMS > 1 MeV

Our result (preliminary)： training data



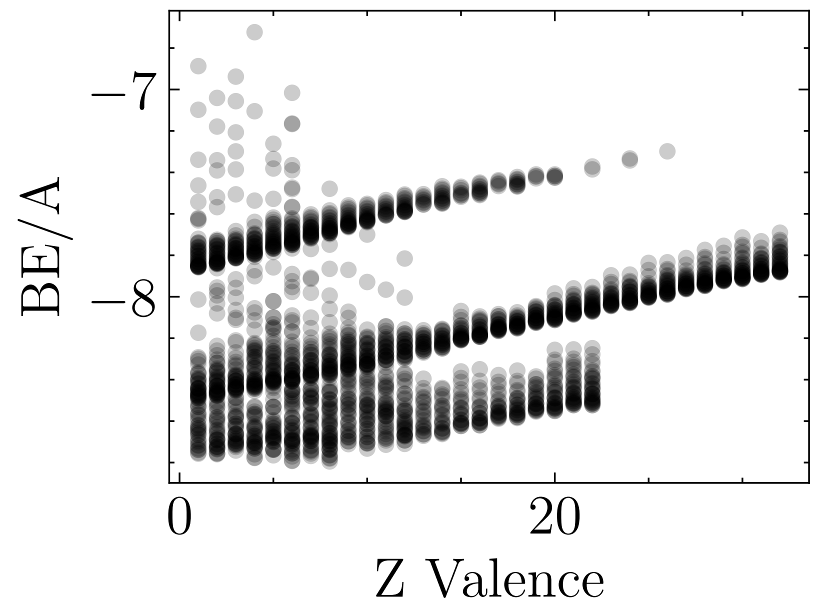






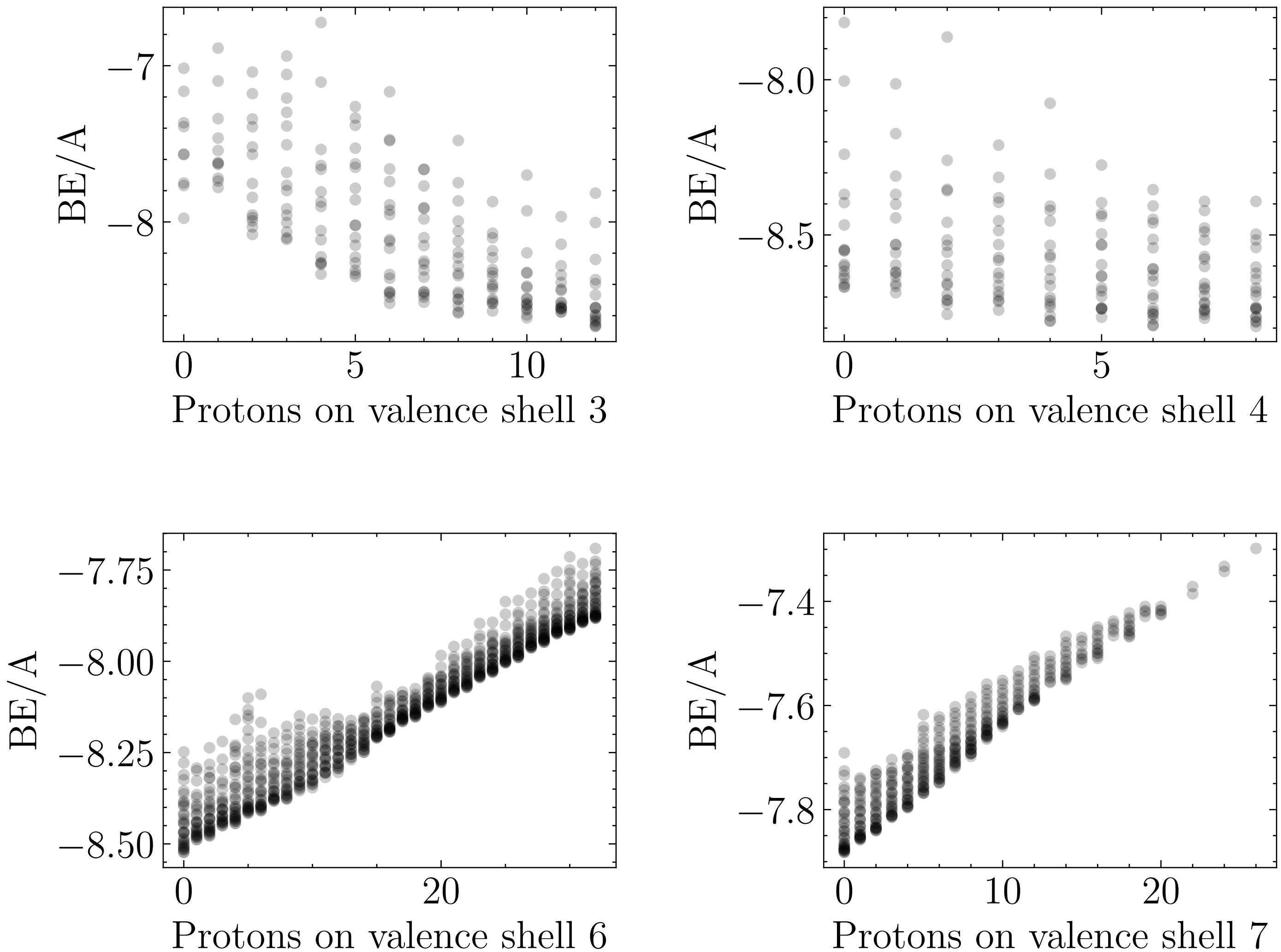
Correlation analysis

$$C_{12} = \frac{\langle \delta x_1 \delta x_2 \rangle}{\sqrt{\langle \delta x_1^2 \rangle \langle \delta x_2^2 \rangle}}$$

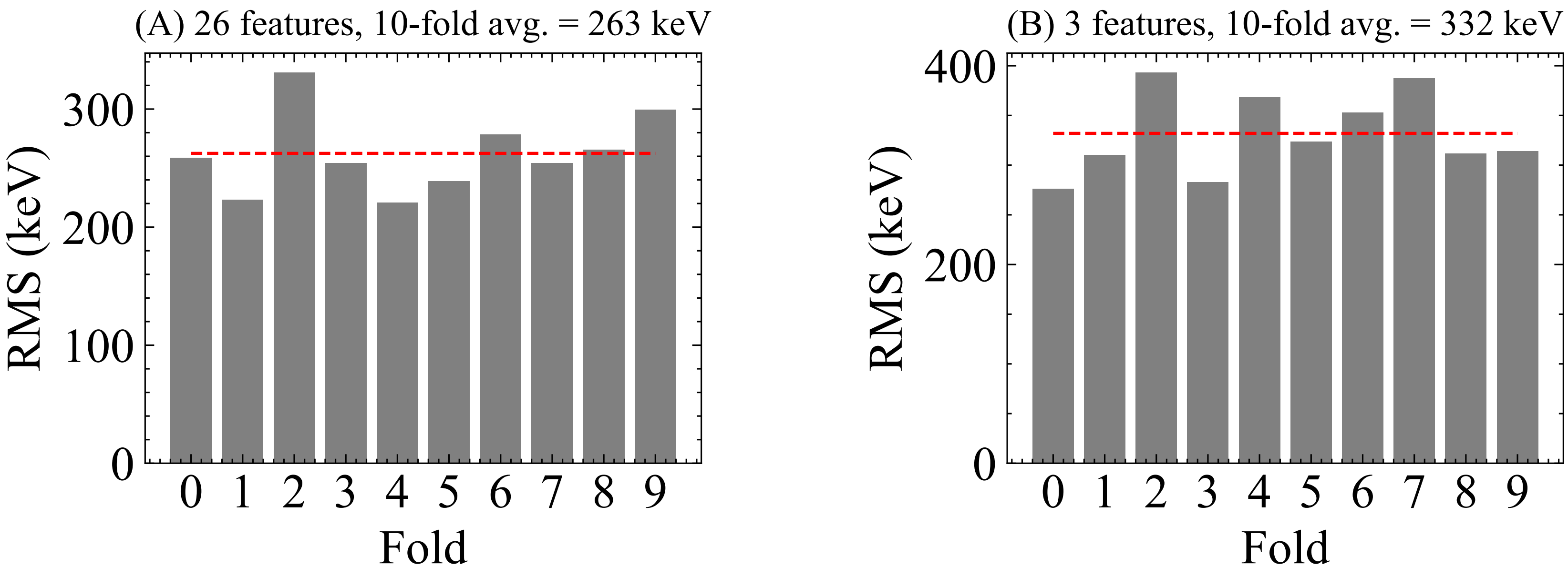


- Pearson correlation analysis found strong correlation between the mass residual, Magic Numbers and number of valence nucleons

Correlation analysis

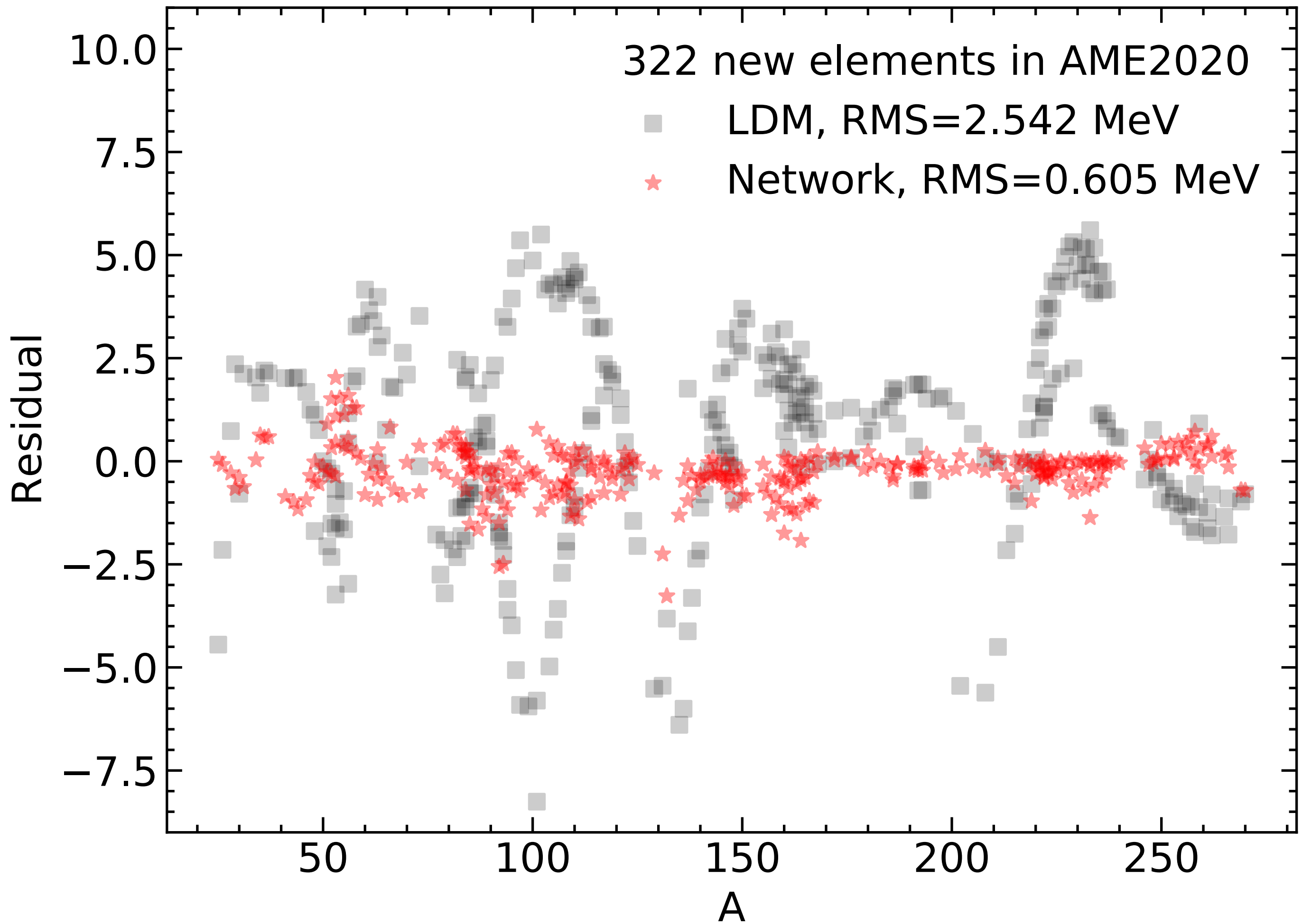


Mass residual prediction



- 26-feature input: Z, N, A, 7 Z shells, 8 N shells, pairing term, surface term, volume term, magic Z, magic N, valence Z, valence N
- For 3 features, Z, N, A

New data from AME2020 which are not in training dataset



Other studies

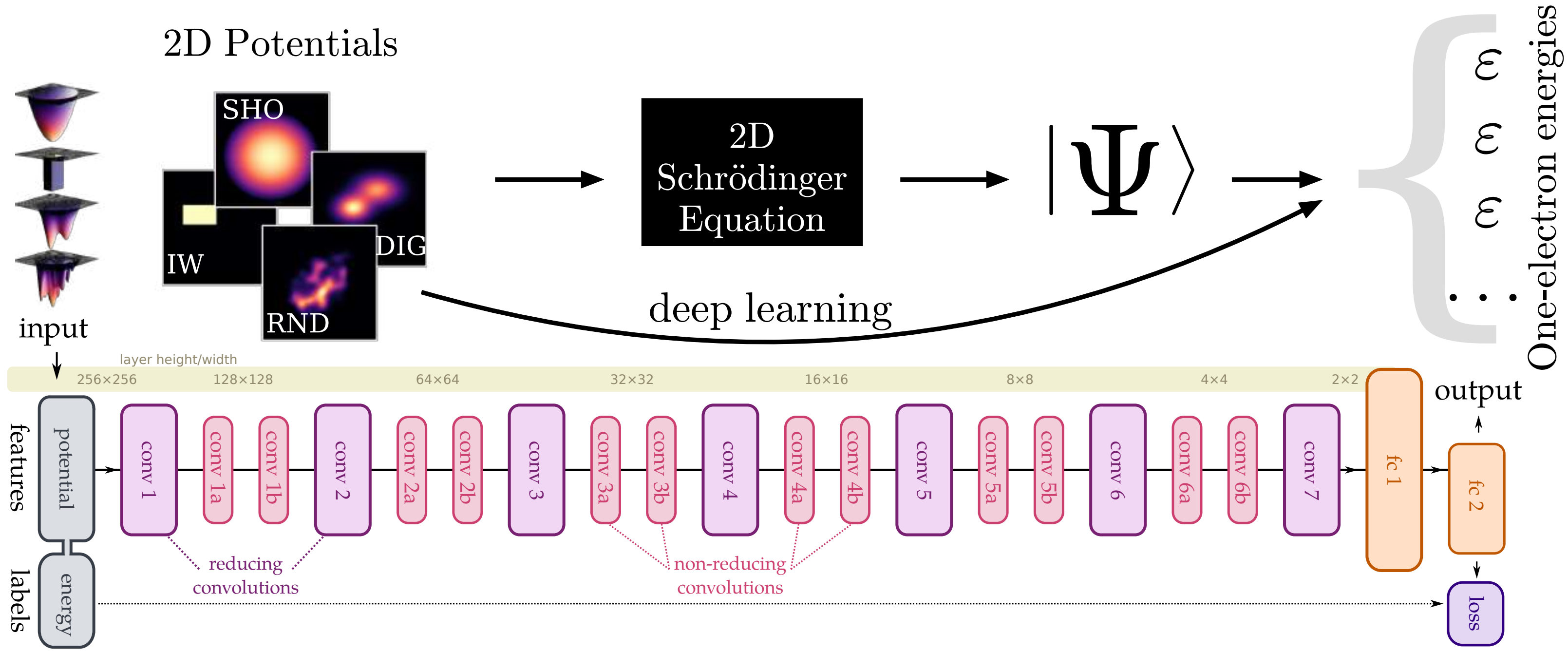
1. Nuclear mass predictions based on Bayesian neural network approach with pairing and shell effects (Z. M. Niu and H. Z. Liang)
2. NIU ZM, FANG J, NIU YF. Comparative study of radial basis function and Bayesian neural network approaches in nuclear mass predictions. 2019. Physical Review C. 100.
3. New extrapolation method for predicting nuclear masses (C.Ma, M.Bao, Z.M. Niu, Y.M.Zhao, A. Arima)
4. Machine learning the nuclear mass (Zepeng Gao, Y.J. Wang, H.L. Lv, Q.F. Li, C.W. Shen)
5. Bayesian extraction of incomplete fission yield (ZiAo Wang, J.C. Pei, Y. Liu, Q.Yu)
6. The description of giant dipole key parameters with multi-task neural network (JingHu Bai, Z.M. Niu, B.Y. Sun, Y.F. Niu)
7. …

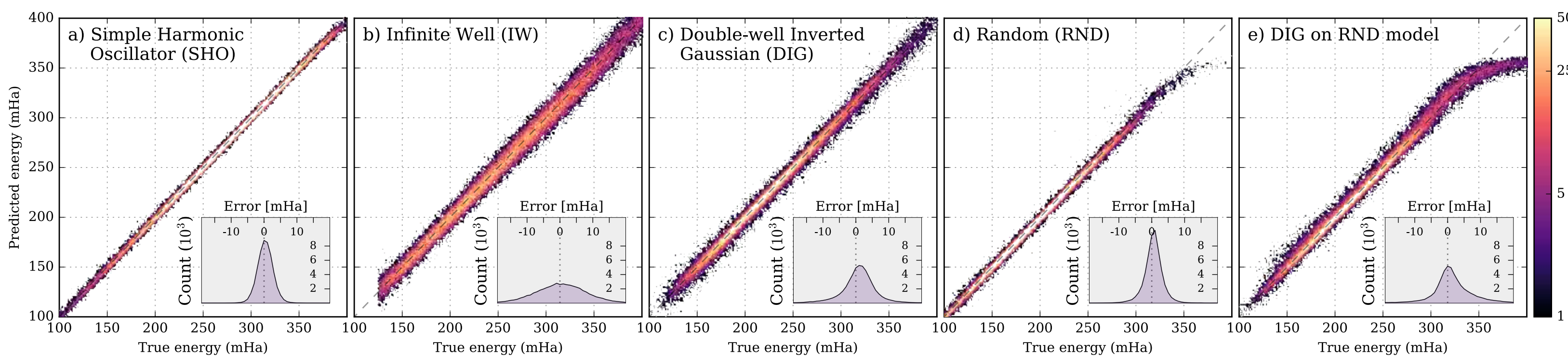
CONTENTS

- What's Machine Learning (ML)
- My field: high energy nuclear physics
- ML for atomic mass prediction
- Outlook： ML for nuclear matrix element
    - Variational wave function using deep learning
    - Speed up MCMC sampling for lattice cal.

DEEP LEARNING FOR NUCLEAR MATRIX ELEMENTS THROUGH
(1) WAVE FUNCTIONS
(2) SAMPLING

Solving Schrodinger Eqs: no wf



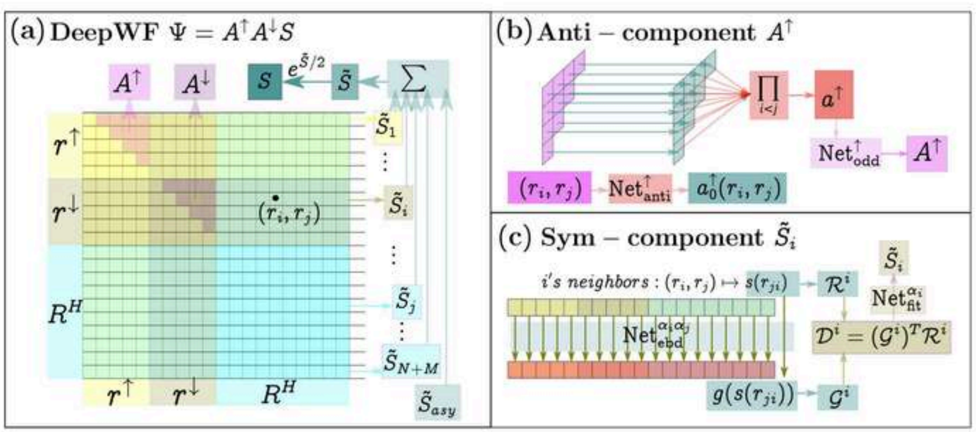


, by K. Mills, M. Spanner, Tamblyn (February 7, 2017)

DeepWF: anti-symmetric trial wave-function using neural network

Solving many-electron Schrödinger equation using deep neural network

JiequnHan LinfengZhang WeinanE



$$\Psi(\boldsymbol{r};\boldsymbol{R}) = S(\boldsymbol{r};\boldsymbol{R})A^{\uparrow}\left(\boldsymbol{r}^{\uparrow}\right)A^{\downarrow}\left(\boldsymbol{r}^{\downarrow}\right)$$

$$\boldsymbol{a}^{\uparrow}\left(\boldsymbol{r}^{\uparrow}\right) = \prod_{1\leq i<j\leq N_{\uparrow}} \boldsymbol{a}_0^{\uparrow}\left(\boldsymbol{r}_i,\boldsymbol{r}_j\right)$$

Build physical a prior into the neural network, e.g., anti-symmetric, vortical free, divergence free, translational invariant (equivalent), rotational symmetry

$$\boldsymbol{a}_0^{\uparrow}\left(\boldsymbol{r}_i,\boldsymbol{r}_j\right) = \mathrm{Net}^{\uparrow}_{\mathrm{anti}}\left(\boldsymbol{r}_i,\boldsymbol{r}_j,|\,r_{ji}\,|\right) - \mathrm{Net}^{\uparrow}_{\mathrm{anti}}\left(\boldsymbol{r}_j,\boldsymbol{r}_i,|\,r_{ji}\,|\right)$$

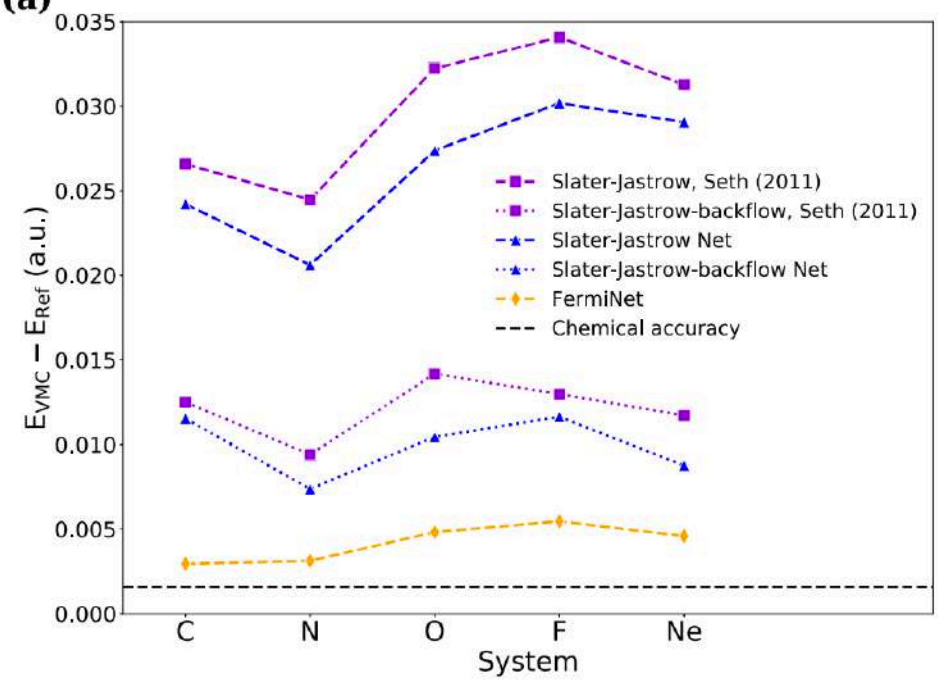
Fermi-Net:

Ab initio solution of the many-electron Schrödinger equation with deep neural networks

David Pfau,[*,†] James S. Spencer,[*] and Alexander G. D. G. Matthews
DeepMind, 6 Pancras Square, London N1C 4AG, United Kingdom

W. M. C. Foulkes
Department of Physics, Imperial College London, South Kensington Campus, London SW7 2AZ, United Kingdom

(a)

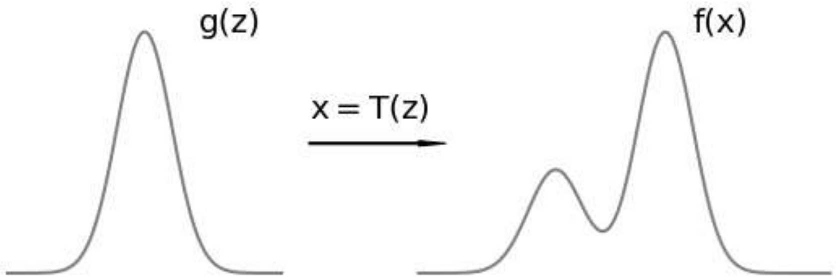


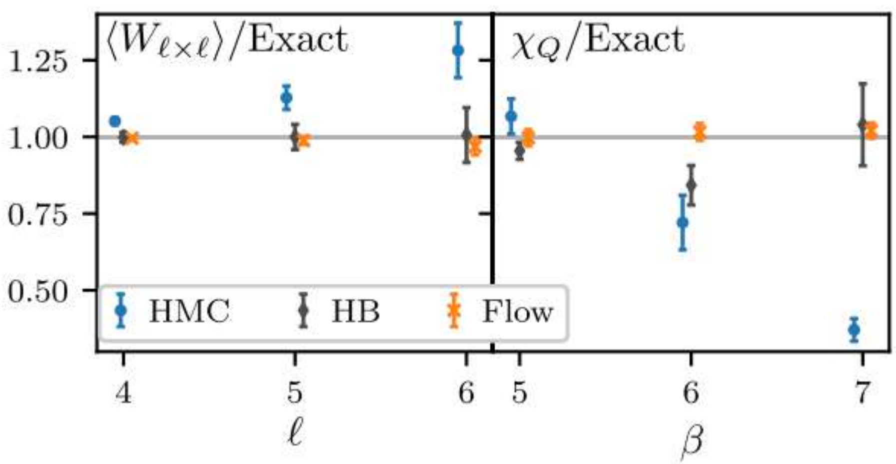
- Similar idea as DeepWF
- Avoids the use of a finite basis set
- Accuracy close to chemical accuracy (1 kcal/mol)

Flow model + MCMC for Lattice

Equivariant Flow-Based Sampling for Lattice Gauge Theory

Gurtej Kanwar,[1] Michael S. Albergo,[2] Denis Boyda,[1] Kyle Cranmer,[2] Daniel C. Hackett,[1] Sébastien Racanière,[3] Danilo Jimenez Rezende,[3] and Phiala E. Shanahan[1]





- Flow model: reversible, bijection , generation model

- Use flow model to speed up sampling that are gauge invariant by construction

Summary

1. Machine learning is widely used in high and low energy nuclear physics
2. It becomes popular in ab initio calculations in computational chemistry
3. Deep learning may help the NME calc in 2 ways
   - Trial wave function using network
   - Speed up sampling