Identifying QCD transition in a hybrid model with deep learning

Yi-Lun Du (FIAS/NJU/UiB)

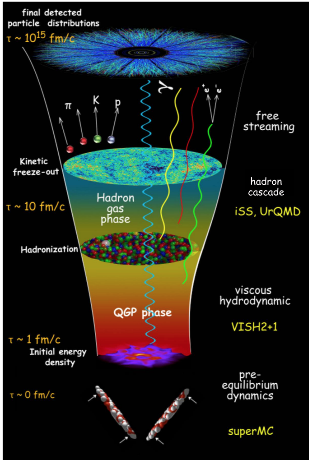
with K. Zhou, J. Steinheimer, A. Motornenko(FIAS), L-G. Pang, X-N. Wang(CCNU/LBL), H-S. Zong(NJU), and H. Stöcker(FIAS/GU/GSI)

Quark matter 2019 Nov 4-9, @ Wuhan

Outline

- Motivation: employ the deep learning technique
- Review: pure hydrodynamics study
- Hybrid model: hydro + cascade (arXiv:1910.11530)
- Summary & Outlook

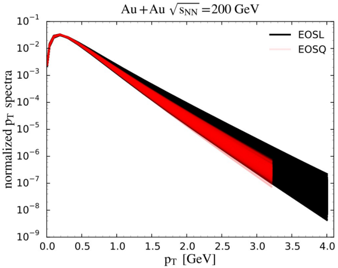
Evolution of the fireball

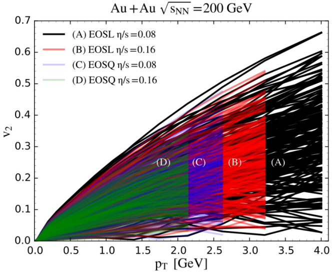


C. Shen et al., Comput. Phys. Commun. 199, 61 (2016)

- Initial condition: Glauber, CGC...
- Equilibration time τ_0
- Strongly coupled fluid described by viscous hydrodynamics - Input: EoS of QCD, viscousity η/s...
- T_freeze-out / switching
- Hadronic cascade (UrQMD, SMASH)

Conventional observables

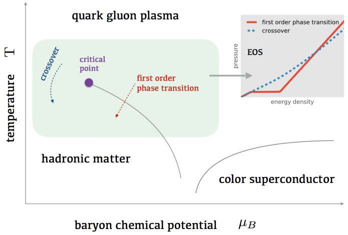


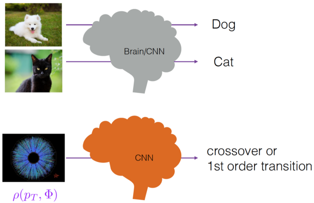


Strongly depends on initial stage fluctuations and other parameters

Analogy with pattern recognition

QCD transition and quark-gluon plasma





CAUTION: model (e.g. event generators) dependence in training

Hydrodynamics: Training dataset

Final Spectra for charged pions at mid-rapidity : $\rho(p_T, \Phi) \equiv \frac{dN_i}{dY p_T dp_T d\Phi} = g_i \int_\sigma p^\mu d\sigma_\mu f_i$

	TRAINING DATASET	$\eta/s = 0$		$\eta/s = 0.08$	
		EOSL	EOSQ	EOSL	EOSQ
RHIC	Au-Au $\sqrt{s_{NN}} = 200\,\text{GeV}$	7435	5328	500	500
LHC	Pb-Pb $\sqrt{s_{NN}} = 2.76\,\text{TeV}$	4967	2828	500	500

- CLVisc 3+1 D viscous hydrodynamics with AMPT initial conditions
- τ_0 is 0.4 fm for Au-Au STAR and 0.2 fm for Pb-Pb
- T_freeze-out is 137 MeV

~22000 events, doubled by left-right flipping along \phi,
10% for validation during the training

CLVisc (3+1) D viscous hydro package: L.-G. Pang, Q. Wang, and X.-N. Wang, Phys. Rev. C 86, 024911 (2012)

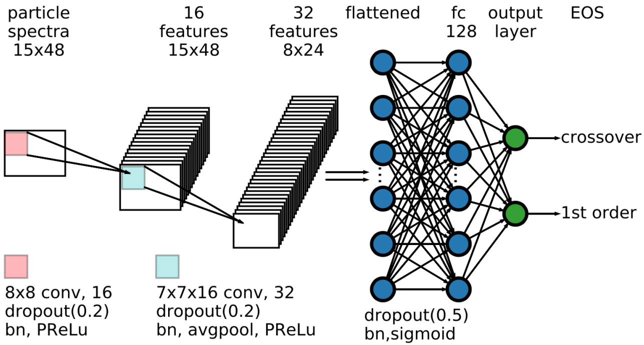
Testing dataset

TESTING DATASET GROUP 1 : iEBE-VISHNU + MC-Glauber						
Centrality:	$\eta/s \in [0, 0.05]$		$\eta/s \in (0.05, 0.10]$		$\eta/s = (0.10, 0.16]$	
10-60%	EOSL	EOSQ	EOSL	EOSQ	EOSL	EOSQ
Au-Au $\sqrt{s_{\rm NN}} = 200$ GeV	650	850	900	750	200	950
Pb-Pb $\sqrt{s_{\rm NN}} = 2.76$ TeV	500	650	600	644	499	150

TESTING DATASET GROUP 2 : CLVisc + IP-Glasma		
Au-Au $\sqrt{s_{\rm NN}} = 200$ GeV, $b \lesssim 8$ fm	EOSL	EOSQ
$\eta/s = 0$	4164	4752
$\eta/s = 0.08$	1173	864

- iEBE-VISHNU (C. Shen et al., Comput. Phys. Commun. 199, 61 (2016)): (2+1) D viscous hydro package with different initial condition (MC-Glauber)
- $\tau_0 = 0.6$ fm; $\eta/s \in [0, 0.16]$
- T_freeze-out: $\in$ [115, 142] MeV for iEBE-VISHNU; 137 MeV for CLVisc+IP-Glasma

DCNN architecture



Validation & Testing results

TESTING DATA	GROUP 0	GROUP 1	GROUP 2
Number of events	4000	7343	10953
Accuracy	$99.88 \pm 0.04\%$	$93.46 \pm 1.35\%$	$93.91 \pm 3.92\%$

- on average $\sim$ 95% prediction accuracy $\rightarrow$ the trained CNN model identifies the type of QCD transition solely(!) from the raw spectra
- The performance is ROBUST against: initial conditions, η/s, τ_0, T_freeze-out $\rightarrow$ model independent!

Nature Commun 9, 210 (2018)

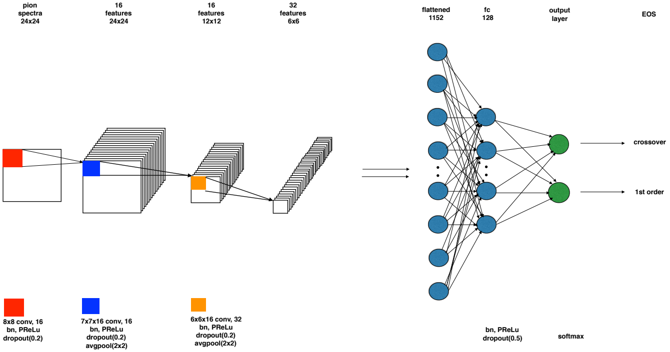
Hybrid model: more realistic circumstances

Couple (2+1) D viscous hydro model (VISHNew) with hadronic cascade model (UrQMD), where probabilistic scatterings and resonance decays are involved.

* Smooth sampled distributions from hydrodynamics $\rightarrow$
  - Event-by-event spectra, with $T_{sw} = 137$ MeV
  - Cascade-coarse-grained spectra, with $T_{sw} = 137$ MeV
  - Events-fine-averaged spectra, with $T_{sw} = 137$ MeV

  - Event-by-event spectra, with $T_{sw} > 150$ MeV
  - Cascade-coarse-grained spectra, with $T_{sw} > 150$ MeV
  - Events-fine-averaged spectra, with $T_{sw} > 150$ MeV

Deeper CNN architecture



There are overall 203194 trainable and 120 non-trainable parameters in our neural network.

Training data, with $T_{sw} = 137$ MeV

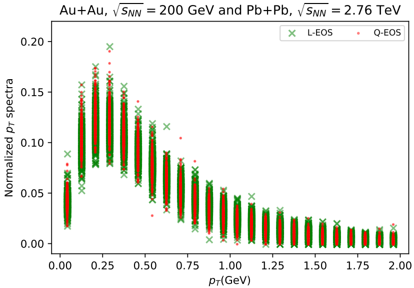
TRAINING DATASET1		
Centrality bin	EOSL	EOSQ
4%-5%	2539	2540
14%-15%	1022	1024
20%-21%	2814	2816
30%-31%	2560	2560
40%-41%	1024	1024
50%-51%	896	1024

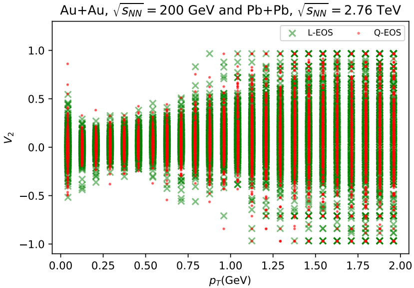
TABLE I: Training dataset 1: numbers of $\rho(p_T, \Phi)$ generated by the iEBE-VISHNU package with the Glauber initial conditions in the centrality range 0-60%. Ratio of shear viscosity to entropy density $\eta/s = 0.08$, equilibration time $\tau_0 = 0.5$ fm/c. The freeze-out temperature is set to be 137 MeV. Pb-Pb $\sqrt{S_{NN}} = 2.76$ TeV.

TRAINING DATASET2		
Centrality bin	EOSL	EOSQ
0%-1%	979	1024
10%-11%	2560	2560
20%-21%	1024	1024
30%-31%	1024	1024
40%-41%	2560	2560
50%-51%	2816	2816

TABLE II: Training dataset 2: numbers of $\rho(p_T, \Phi)$ generated by the iEBE-VISHNU package with the Glauber initial conditions in the centrality range 0-60%. Ratio of shear viscosity to entropy density $\eta/s = 0.00$, equilibration time $\tau_0 = 0.4$ fm/c. The freeze-out temperature is set to be 137 MeV. Au-Au $\sqrt{S_{NN}} = 200$ GeV.

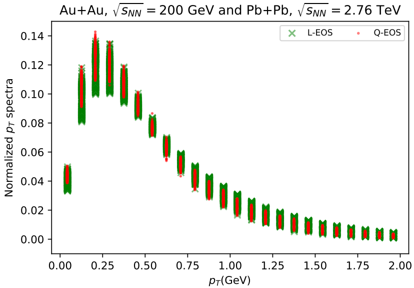
Event-by-event p_T spectra and elliptic flow v_2

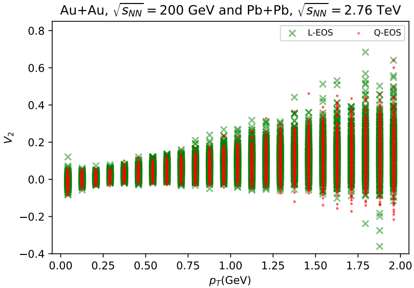




These events are generated in different centrality bins with $T_{sw} = 137$ MeV in two collision systems.

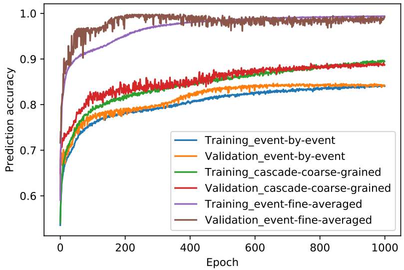
30-events-fine-averaged p_T spectra and elliptic flow v_2

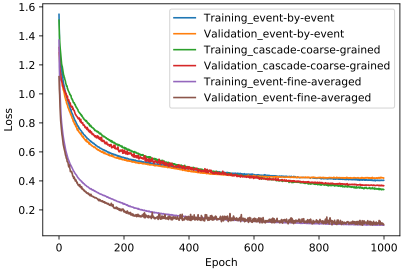




These events are generated in different centrality bins with $T_{sw} = 137$ MeV in two collision systems.

Prediction accuracy and loss





A clear hierarchy is observed in the prediction accuracy when using three different spectra as input for the network, which are around 80%, 90% and 99%, respectively.

Testing results, with event-fine-averaged spectra

PREDICTIVE ACCURACY FOR TESTING DATASETS 1								
Centrality bin	$\sqrt{s_{NN}}$ [TeV]	Ini. Cond.	τ_0 (fm/c)	η/s	T_{sw}	L-EOS	Q-EOS	Accuracy
15%-16%	Au+Au 0.2	MC-G	0.4	0.00	141 MeV	512	512	89.1%
15%-16%	Au+Au 0.2	MC-G	0.4	0.00	140 MeV	2560	2560	95.6%
45%-46%	Au+Au 0.2	MC-G	0.6	0.12	130 MeV	1024	1024	100%
7%-8%	Pb+Pb 2.76	MC-G	0.6	0.12	130 MeV	1280	1279	99.8%
17%-18%	Pb+Pb 2.76	MC-G	0.6	0.12	130 MeV	2560	2560	98.1%
25%-26%	Pb+Pb 2.76	MC-G	0.6	0.12	130 MeV	2560	2560	97.4%
25%-26%	Pb+Pb 2.76	MC-G	0.6	0.16	130 MeV	1024	1024	97.8%

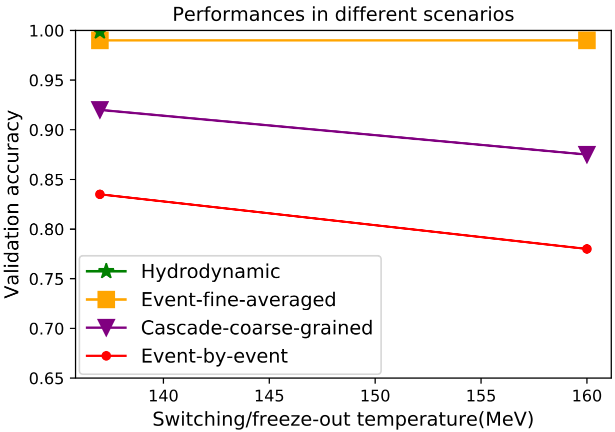
Table B.4: Predictive accuracy on the testing datasets 1: 30-events-fine-averaged spectra $\rho_a(p_T, \Phi)$ generated with MC-Glauber initial conditions and different $\sqrt{s_{NN}}$, η/s, τ_0, and T_{sw} in the centrality range 0-50%.

PREDICTIVE ACCURACY FOR TESTING DATASETS 2								
Centrality bin	$\sqrt{s_{NN}}$ [TeV]	Ini. Cond.	τ_0 (fm/c)	η/s	T_{sw}	L-EOS	Q-EOS	Accuracy
15%-16%	Au+Au 0.2	MCKLN	0.6	0.12	137 MeV	512	256	98.6%
35%-36%	Au+Au 0.2	MCKLN	0.6	0.12	142 MeV	896	896	99.4%
10%-11%	Pb+Pb 2.76	MCKLN	0.6	0.12	142 MeV	150	150	100%
25%-26%	Pb+Pb 2.76	MCKLN	0.6	0.12	137 MeV	256	256	84.4%

Table B.5: Predictive accuracy on the testing datasets 2: 30-events-fine-averaged spectra $\rho_a(p_T, \Phi)$ generated with MCKLN initial conditions and the different $\sqrt{s_{NN}}$, η/s, τ_0, and T_{sw} in the centrality range 0-40%.

The validation and testing accuracy can be achieved up to 99% and 96%, respectively, on average, which are robust against different initial condition, η/s, τ_0, T_{fo}, $\sqrt{S_{NN}}$.

Overall results



Summary

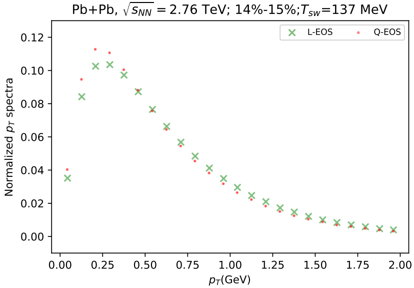
- The predictive power of CNN by training event-by-event spectra decreases down to 80% (compared with pure hydro case 99%), due to stochastic particlization, hadronic cascade and resonance decays.
- The predictive power of CNN by training cascade-coarse-grained or events-fine-averaged spectra improves profoundly with good generalizability.
- With freeze-out T increased, more hadronic cascade are involved, the predictive power of CNN by training event-by-event and cascade-coarse-grained decrease slightly.
- Existence of an "EoS-meter" on identifying the QCD phase transition from the final spectrum $\rho(p_T, \phi)$ even though probabilistic hadronic rescatterings and resonance decays are included.

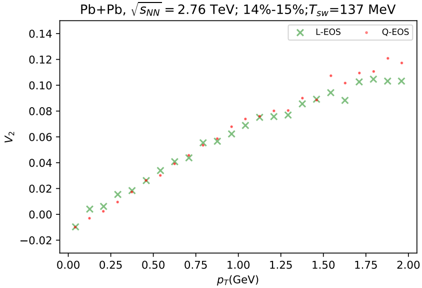
Outlook

- Verify model dependence!
- Extraction of transport coefficient η/s
- Classification of initial conditions

Thank you for your attention!

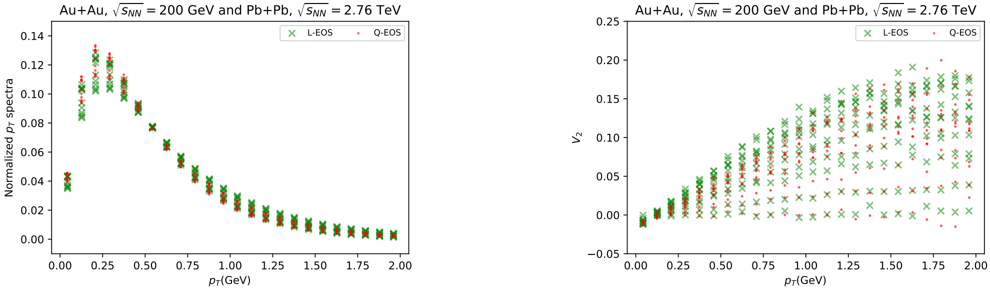
All-events-fine-averaged p_T spectra and elliptic flow v_2





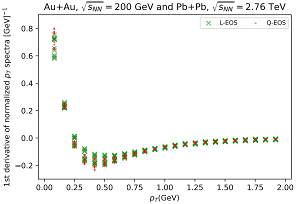
These events are generated in centrality bin 14-15% with $T_{sw} = 137$ MeV in one collision system.

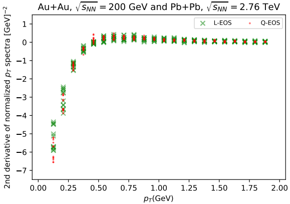
All-events-fine-averaged p_T spectra and elliptic flow v_2

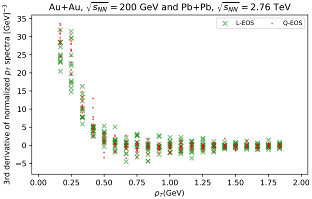


These events are generated in different centrality bins with $T_{sw} = 137$ MeV in two collision systems.

Derivatives of all-events-fine-averaged p_T spectra

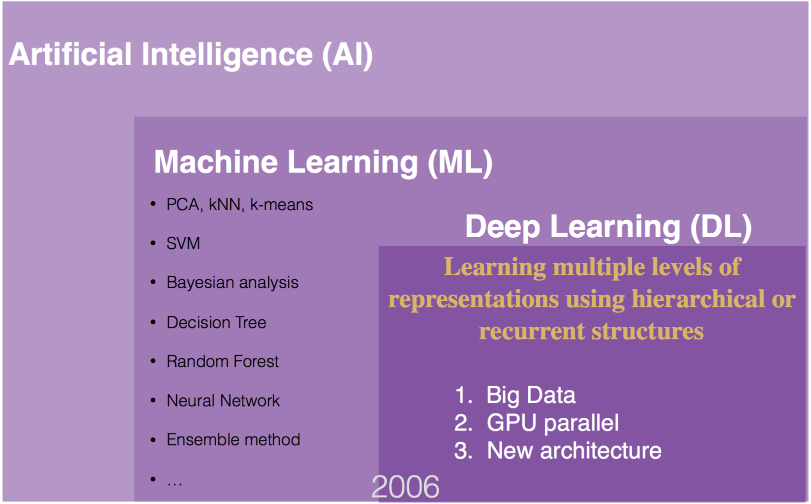






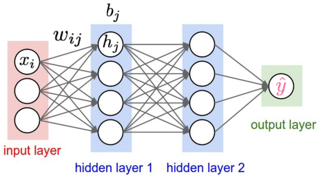
These events are generated in different centrality bins with $T_{sw} = 137$ MeV in two collision systems.

What's deep learning?

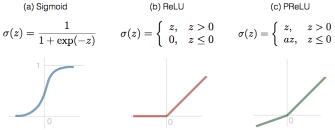


Geoffrey Hinton

“hello world” example of deep neural network



Fully Connected Network (FCN)



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

Linear operations:rotating, boosting,... increasing or decreasing dimensions

$$h_j = \sigma(z_j)$$

Nonlinear activation function: correlation/links

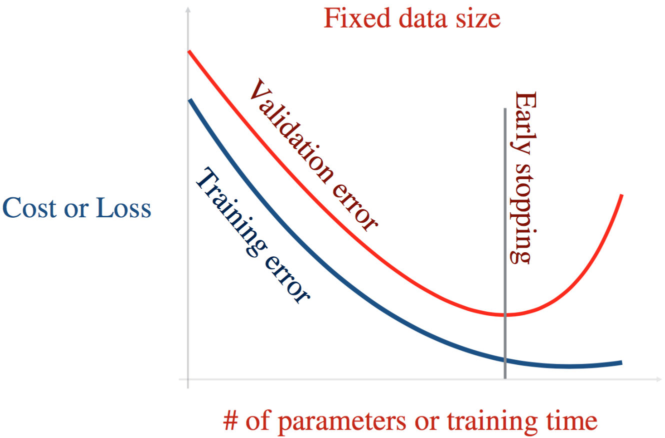
$$l(\theta) = \sum_i (\hat{y}_i - y_i)^2$$

Mean square error (simplest loss function) with $\hat{y}_i$ the predicted value and y_i the true value

$$\theta' = \theta - \epsilon \frac{\partial l(\theta)}{\partial \theta}$$

SGD for parameter update to minimize loss function

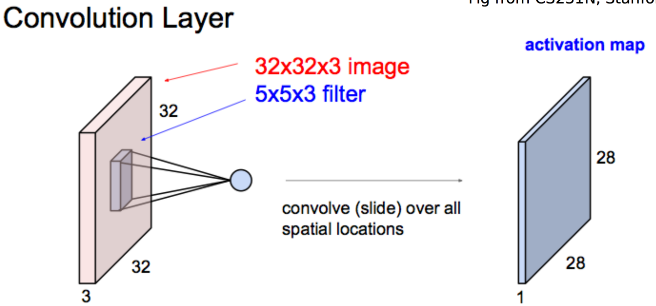
Overfitting problem



Too many parameters may easily overfit training dataset

Convolution neural network

Fig from CS231N, Stanford



Advantage: scaling, rotating, translation invariant features can be learned since only subregion is connected to the filter/kernel which scan the whole input to feel the 2D/3D structure and local statistics.

CNN vs FCN in image recognition

- Local features are broken in flattened representation
- 2D/3D filters with certain weights scan the whole picture to capture different features in neighborhood: edges, specific color, blurring noise...
- Convolutional and Pooling layer: reduce No. of dimension and parameters
- The deeper layer goes, the more abstract features can be learned
- FCN: converted to class of input data and improve the generalizablity - "firewall"