THE EUROPEAN
PHYSICAL JOURNAL C

Regular Article - Theoretical Physics

Identifying the nature of the QCD transition in relativistic collision of heavy nuclei with deep learning

Yi-Lun Du[1,2,3,4,a], **Kai Zhou**[1,b], **Jan Steinheimer**[1], **Long-Gang Pang**[5,6,7], **Anton Motornenko**[1,2], **Hong-Shi Zong**[3,8,9], **Xin-Nian Wang**[5,6,7], **Horst Stöcker**[1,2,10]

[1] Frankfurt Institute for Advanced Studies, Giersch Science Center, 60438 Frankfurt am Main, Germany
[2] Institut für Theoretische Physik, Goethe Universität Frankfurt, 60438 Frankfurt am Main, Germany
[3] Department of Physics, Nanjing University, Nanjing 210093, China
[4] Department of Physics and Technology, University of Bergen, 5007 Bergen, Norway
[5] Nuclear Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[6] Physics Department, University of California, Berkeley, CA 94720, USA
[7] Key Laboratory of Quark and Lepton Physics (MOE) and Institute of Particle Physics, Central China Normal University, Wuhan 430079, China
[8] Nanjing Proton Source Research and Design Center, Nanjing 210093, China
[9] Department of physics, Anhui Normal University, Wuhu 241000, Anhui, China
[10] GSI Helmholtzzentrum für Schwerionenforschung, 64291 Darmstadt, Germany

Abstract Using deep convolutional neural network (CNN), the nature of the QCD transition can be identified from the final-state pion spectra from hybrid model simulations of heavy-ion collisions that combines a viscous hydrodynamic model with a hadronic cascade "after-burner". Two different types of equations of state (EoS) of the medium are used in the hydrodynamic evolution. The resulting spectra in transverse momentum and azimuthal angle are used as the input data to train the neural network to distinguish different EoS. Different scenarios for the input data are studied and compared in a systematic way. A clear hierarchy is observed in the prediction accuracy when using the event-by-event, cascade-coarse-grained and event-fine-averaged spectra as input for the network, which are about 80%, 90% and 99%, respectively. A comparison with the prediction performance by deep neural network (DNN) with only the normalized pion transverse momentum spectra is also made. High-level features of pion spectra captured by a carefully-trained neural network were found to be able to distinguish the nature of the QCD transition even in a simulation scenario which is close to the experiments.

1 Introduction

The dynamics of the strong interactions between quarks and gluons, governing the properties of hot and dense nuclear matter, can be described by the theory of QCD. It predicts that, if the temperature of strongly-interacting matter becomes large enough, a new state of matter is formed in which quarks and gluons can roam freely and are not confined in the hadrons anymore. This state of matter is called the quark-gluon plasma (QGP). Lattice QCD has established that the transition from a hadron gas to the QGP is a smooth crossover at a high temperature $T \sim 140 - 180$ MeV and low net baryon density [1–3]. A variety of theoretical models, such as the Dyson–Schwinger equations model [4–7], the (Polyakov loop-) Nambu–Jona–Lasinio model [8–12] and the quark-meson coupling model [13–15] also predict the existence of a first-order phase transition that occurs at low temperature and moderate to large net baryon densities.

Relativistic heavy ion experiments have been carried out at the SIS18 [16], at the AGS [17] and at the SPS [18] in the fixed target mode and at the Relativistic Heavy Ion Collider (RHIC) [19] as well as at the Large Hadron Collider (LHC) [20] in the collider mode. The forthcoming Facility for Anti-proton and Ion Research (FAIR) [21,22] and the Nuclotron-based Ion Collider fAcility (NICA) [23] will provide unprecedented intensities and luminosities for future studies. The main goal of these large experiments is to search for signals for the QCD phase transition and study the properties of QGP in nucleus-nucleus collisions. Due to the transience of the heavy ion collision dynamics, the QCD medium bulk properties can't be directly observed in experiment. A strategy to identify the signals of QGP is to compare sophisticated model simulations with varying parameter sets and

different equations of state (with and without a phase transition) with experimental data such as particle spectra and correlation functions. Currently some observables, for example, anisotropic flow [24–27], directed flow [28,29] and fluctuations of particle multiplicities [30–33], are conjectured as most sensitive to the appearance of a phase transition. However, no disentangled mapping between these observables and this specific bulk property of the QCD medium from others, has been obtained so far. Then it's necessary to call for modern data analysis methods like Bayesian analysis or the deep neural network approach.

The Bayesian analysis [34–36] applies a global fitting to a set of different observables for parameter estimation. In Ref. [36], the crossover type EoS was employed in the hybrid hydrodynamic framework and the event-averaged experimental data (e.g. particle yields, momentum distribution and flow) were used to infer the temperature-dependent shear and bulk viscosity of nuclear matter and other parameters at the same time. These estimated temperature-dependent viscosities are their marginal distributions by integrating out other parameters, respectively. One way to constrain these bulk properties of nuclear matter better is to fit more data or make use of more information from data. On one hand, one can employ higher-dimensional raw data instead of the integrated one to fit. On the other hand, the event-by-event fluctuation may contain more information as well.

In this work, we will explore the feasibility of identifying QCD EoS from event-by-event high-dimensional raw hadron spectra in high energy nucleus-nucleus collisions using the tools and techniques of Deep Learning (DL). DL was developed to capture highly-correlated features from big data [37,38]. It has achieved tremendous success in a wide variety of applications, like image processing, natural language processing, computer vision, medical imaging, medical information processing, and other interesting fields. These have inspired physicists to adopt the technique to tackle physical problems of great complexity. A lot of progresses have been made in nuclear physics [39–45], lattice field theory [46–50], particle physics [51–55], astrophysics [56–58] and condensed matter physics [59–65].

For our exploration with DL method here, the purpose is to find out a disentangled mapping between observed final raw spectra and the EoS type for the medium. We vary different parameters, including shear viscosity, equilibration time, freeze-out temperature, etc., to enforce the neural network to explore if it can find a direct mapping from event-by-event high-dimensional raw spectra to the EoS type which can be immune to other parameters' 'interference' in certain ranges. As long as we can find such a mapping, its straightforward to infer information about the EoS type from the measured data in experiment as the detector simulation or calibration is also considered for further study.

The great advantage of the DL method over conventional ones is its ability to extract hidden features from highly dynamical, rapidly evolving and complex non-linear systems, like in relativistic heavy ion collisions. Conventional observables rely on human's design and are usually low-dimensional projections of the high-dimensional raw data. When one uses only part of these projected information to constrain the properties of nuclear matter, the estimated value are prone to be dependent on the specific model setup (e.g. other untuned parameters in the fitting) and the chosen observables. Instead DL methods can be used to explore distinct mappings and to construct observables from the full high-dimensional raw data for the classification task at hand. Recently, a deep CNN classifier was developed as an effective "EoS-meter", an excellent tool for revealing the nature of the QCD transition with a high predictive accuracy $\sim 95\%$ in hadron spectra from a pure hydrodynamic study [39].

The present work studies the performance of a CNN to identify the EoS trained and tested with hadron spectra from a more realistic simulation of heavy ion collisions. The generalizability of the method is explored by considering well established dynamics in the state-of-the-art simulation models. First of all, the hadronic rescattering, after the hydrodynamics evolution, is taken into account in the simulation via a hadronic cascade. Consequently, the event-by-event final-state pion spectra are discrete instead of smooth as in hydrodynamic simulations. Secondly, the resonance decays are included, which also contribute to the pion spectra. Due to the finite number of particles, the discrete event-by-event pion spectra will have significant fluctuations that might overwhelm correlations one is looking for. We will develop modified DL-tools with CNN to identify the EoS in this more complex and more realistic dynamic scenario.

This paper is organized as follows: Sect. 2 introduces the hybrid simulation model. Section 3 discusses the neural network and the methods of the data pre-processing. Section 4 presents the performance of the trained CNN in different scenarios and comparisons with that of a fully-connected deep neural network (DNN). Finally, Sect. 5 summarizes the results and gives the conclusions. A gives the details of the neural network structure. B shows the simulated data and predictive performance on testing datasets by the trained neural network. C visualizes the training datasets in B with traditional observables.

2 Micro–macro hybrid model of relativistic heavy-ion collisions

The modeling of relativistic heavy-ion collision is mostly done by following a "standard prescription" for the spatio-temporal evolution of the collision dynamics. The initial state of the matter right after the violent collision is described

by the "color glass condensate", which consists of frozen primordial gluons and is assumed to isotropize within 1 fm/c [66–70]. These gluons may evolve rapidly in accordance with the classical Yang–Mills equation. A few fm/c later, they can achieve approximate local thermal equilibrium [71,72] and may exist briefly as a Yang–Mills gluon plasma, which may quickly expand nearly isentropically due to the high initial temperature. The total entropy and energy are not yet distributed over quark–anti-quark degrees of freedom. Subsequently, quarks are produced by gluon–gluon collisions [67–70], forming a strongly coupled quark-gluon plasma (sQGP). The dynamical evolution of that QGP can be described approximately by macroscopic dissipative hydrodynamics [73–78]. Viscous corrections are included to describe some of the remaining deviation from local isotropy and thermal equilibrium. The EoS of the hot QGP medium, the constitutive element used to close the hydrodynamic equations, is one crucial input. As the medium expands and cools quasi-isentropically, the quark-gluon fluid will go through a smooth crossover, or hypothetically in this work as a control experiment, a first order phase transition. The nature of the QCD transition strongly affects the hydrodynamic evolution [79]. Different forms of transitions are associated with different pressure gradients which consequently lead to different expansion rates. As the matter becomes more dilute, it will form an expanding non-equilibrium hadronic matter with important final state effects. For instance, final absorption of the products of the resonance decays in the hadronic matter can substantially change the yields of the hadrons observed by the experimental detectors. This evolution of the hadronic matter can be successfully described by microscopic hadron cascade models [80–82].

To generate the data for the training of the CNN, we use the iEBE-VISHNU hybrid model [83], which can perform event-by-event simulations of relativistic heavy-ion collisions at different energies. Major components of this hybrid model include an initial condition generator (SuperMC), a (2+1)D second-order event-by-event viscous hydrodynamic simulator (VISHNew), a particle sampler (iSS) and a hadron cascade "afterburner" simulator (UrQMD).

This hybrid model uses either the Monte-Carlo Glauber (MC-G) [84–86] or the Monte–Carlo Kharzeev–Levin–Nardi (MCKLN) [87,88] model to generate the fluctuating initial conditions in the SuperMC module. The collision centrality can be set up as needed, based on the assumption that, on average, the final charged hadron multiplicity, dN_{ch}/dy, is directly proportional to the initially produced total entropy in the transverse plane $dS/dy|_{y=0}$. The effect of viscous heating will cause a spread in the final dN_{ch}/dy, which is considered small (2-3%) for a given $dS/dy|_{y=0}$.

The simulation with the hydrodynamic package VISH-New uses two different EoSs: (1) the crossover type EoS, based on a lattice-QCD parametrization [89], denoted as L-
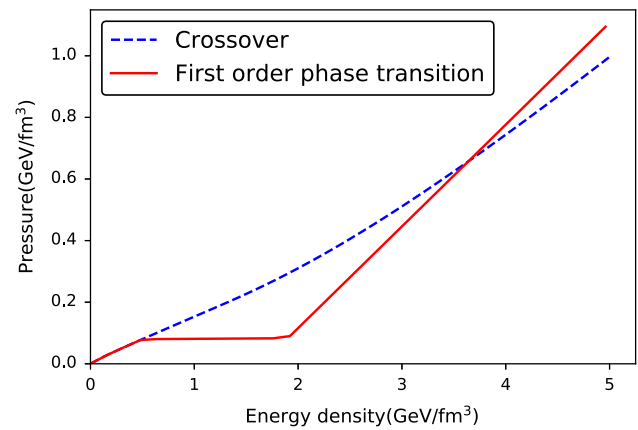

Fig. 1 Two different EoSs are implemented in the hydrodynamic simulation, as functions of the energy density. A crossover, based on a lattice QCD parametrization is compared with a first order phase transition with a transition temperature $T_c = 165$ MeV, obtained by a Maxwell construction. It is assumed that the baryon-chemical potential is exactly $\mu_B = 0$ throughout the whole simulation

EOS; (2) the first order type EoS with a Maxwell construction [90] between a hadron resonance gas and an ideal gas of quarks and gluons, as Q-EOS. The transition temperature is $T_c = 165$ MeV. These two EoSs are depicted in Fig. 1.

After the hydrodynamic evolution, the fluid fields are projected via the Cooper-Frye formula into particles, which will then be further propagated in a hadronic cascade, the Ultrarelativistic Quantum Molecular Dynamics (UrQMD) model. In UrQMD, a non-equilibrium transport model, resonance decays and hadronic rescatterings are included in the simulation. In contrast to the hydrodynamic evolution, which is governed by the conservation of energy and momentum with the EoS, shear viscosity η, bulk viscosity ξ, particles are assumed to be in asymptotic states and the trajectories are given by straight-lines between the collisions in the hadronic cascade. The hadronic cascade evolution is not deterministic since the processes involve certain randomness, e.g., scattering angle, scattering probability and decay probabilities. Furthermore, the effects of finite number of particles, i.e., thermal fluctuations, are included since the cascade propagates the discrete particles instead of the average densities.

This hybrid model with some adjustable parameters can fit experimental data on final hadron spectra. These parameters include: the equilibration time τ_0, which defines the point when the local thermal equilibration is reached and the hydrodynamics evolution starts, the ratio of the shear viscosity to the entropy density η/s, and the freeze-out temperature T_{sw}, which defines the switch from the hydrodynamic evolution to the hadronic cascade.

We vary the model parameters in the generation of the training datasets to allow the neural network to capture the intrinsic features encoded in the EoS, instead of those biased by the specific setup of other physical uncertainties. This

would require many events simulations for hundreds of different parameter combinations and centrality selections, to make sure that the neural network gains a sufficient generalizabilty. However, in practice this is impossible. Hence we focus on systematic changes of these parameters and study the performance of the network whence it reaches the boundary of these parameter values.

3 Neural network and data pre-processing

In Ref. [39], the DL-tool engine with a CNN has been shown to classify successfully the EoS in pure hydrodynamical simulations, on an event-by-event basis with a $\sim 95\%$ accuracy. To apply this strategy to real experimental data, it's crucial to perform realistic simulations with hadronic "afterburner" and resonance decays. In the present paper, the DL-tool engine is constructed for more realistic simulations of heavy ion collisions. The CNN architecture used here is similar to that discussed in Ref. [39]. We refer to that paper for technical details. An introduction to this new CNN network is presented in detail in Fig. 4 in A.

The input $\rho(p_T, \Phi) \equiv dN_\pi/dy\,dp_T\,d\Phi$ to this neural network is a histogram of the number of pions with 24 p_T-bins and 24 Φ-bins. p_T denotes the transverse momenta of observed pions in the final state, while Φ denotes the azimuthal angles. Only pions with $p_T \leq 2$GeV, rapidity $|y| \leq 1$ and $\Phi \in [0, 2\pi]$ are accepted and accounted in the histogram.

In general, training or learning algorithms benefit a lot from pre-processing of the datasets. The input to the neural network used here, pion spectra $\rho(p_T, \Phi)$, is a 24×24 matrix. One refers to each matrix element as one "feature" and each matrix as one "sample". The pre-processing of the input data can be applied in a feature-wise (per feature) or sample-wise (per input sample) manner.

In the feature-wise standardization, the input $\rho(p_T, \Phi)$ of all the training samples are pre-processed in a sample-interdependent manner. Each feature is subtracted with the mean over all training samples and is divided by their standard deviation. In this way, all features are centered around zero and have variances of the same order. Thus it is prevented that one feature with larger variance dominates the objective function over other features. The transformation is saved and then will be applied in the testing samples. With this standardization, the testing data should be simulated in one of the same collision systems as the training data, since the multiplicity in different collision systems differ a lot.

In the sample-wise standardization, or min–max normalization, the input $\rho(p_T, \Phi)$ are pre-processed in a sample-independent manner. Each 24×24 matrix can be rescaled to have a zero mean and a unit variance, or to a specific range, such as $[-\frac{1}{2}, \frac{1}{2}]$, respectively. The latter choice is used in Ref. [39] with success.

Our training results show that feature-wise standardization does always perform better than the other two sample-wise methods. Hence we will show in the following only the results of the feature-wise standardization.

4 Training and testing results

A systematic analysis of the performance of the above described CNN is presented for hybrid modeling for relativistic heavy-ion collisions. Here an important aspect is the generalizability of the trained CNN model in the testing stage. The overfitting of the network to the training data will be checked on the validation data which are generated with the same physical parameter set in modeling the training data. The testing is performed on the testing datasets which are generated with different physical parameter sets in modeling the training data. The generalizability of the CNN model with respect to different physical parameter sets is studied systematically. In the previous study with pure hydrodynamics [39], the training data are generated with a viscous (3+1)D hydrodynamics model, CLVisc [77], with AMPT initial conditions [91], while the testing data are generated with a viscous (2+1)D hydrodynamics model, VISHNew, with Monte-Carlo Glauber initial conditions, which are used in a hybrid model in this work for the training data generation instead. However, here we find that, even in the pure hydrodynamic study, reversing the simulation models for training and testing data generation will obtain a testing accuracy only about 70%, from which we suspect some superiority of $(3 + 1)$D hydrodynamics model with AMPT initial conditions over other ones. Thus in this work, we would not be able to discuss the generalizability of the CNN model with respect to different hybrid simulation models.

4.1 Hybrid model with late transition to cascade

The CNN in the previous study [39] was directly trained using primordial pion spectra, obtained from a numerical integration of the Cooper–Frye formula over the freeze-out hypersurface in the hydrodynamics. In such a scenario, one neglects the fluctuations due to the finite number of hadrons. In addition, a significant portion of pions originating from resonance decays also need to be taken into account. In this section, we study the influence of the aforementioned effects on the predictive power of the CNN. To see the influence of the finite number of particles and resonance decays, we first assume a late transition from hydrodynamics to the UrQMD cascade by taking the switching temperature the same value as the hydrodynamics freeze-out temperature used in Ref. [39], $T_{sw} = 137$ MeV. In this scenario, the duration and influence of the hadronic cascade are significantly diminished and we are left with the effects of the

finite number of particles and resonance decays as compared to the pure hydrodynamics modeling.

4.1.1 Event-by-event input, switch at $T_{sw} = 137$ MeV

In this sub-scenario, the event-by-event pion spectra $\rho(p_T, \phi)$ are taken as the input to the CNN. 12 training datasets are generated by the iEBE-VISHNU hybrid model with the fluctuating MC-Glauber initial condition and 6 different fine centrality bins with 1% width in the centrality range 0-60% in two collision systems, respectively. We set the ratio of the shear viscosity to entropy density as $\eta/s = 0.08$ and 0.00, the equilibration time as $\tau_0 = 0.5$ and 0.4 fm/c in the collision systems Pb+Pb $\sqrt{s_{NN}} = 2.76$ TeV and Au+Au $\sqrt{s_{NN}} = 200$ GeV, respectively. The details of the datasets are shown in Tables 1 and 2 in B. About 44000 events with two different EoSs are generated in total. Figure 6 in C shows the event-by-event normalized p_T spectra and the elliptic flow v_2 as a function of p_T of these training datasets with two EoSs. These two one-dimensional traditional observables are non-distinguishable by the human eye with respect to the EoSs. Thus it's not trivial to identify the EoS from just final-state pion p_T spectra. The negative elliptic flow v_2 in Fig. 6 shows that there are great fluctuations in the event-by-event spectra.

The validation accuracy is found to be about 83.5% after 1000 epochs training. This validation accuracy indicates that high-level correlations are extracted from the two-dimensional pion spectra $\rho(p_T, \phi)$ to identify the EoS. However, it is significantly lower than that in pure hydrodynamics modeling [39], where a validation accuracy up to 99% was obtained. This implies that the fluctuations due to the finite number of particles and resonance decays overwhelm some correlation information from the early dynamics to the final-state particle spectra and thus result in the "overlap" between these two types of event-by-event spectra with different EoSs, which hinders the discrimination between them.

4.1.2 Cascade-coarse-grained input, switch at $T_{sw} = 137$ MeV

To mitigate the effect of fluctuation due to the finite number of particles and resonance decays, we average the pion spectra over a certain number of events. In the model simulations one can repeat the hadronic cascade for any number of times for the same hydrodynamic evolution. Then the pion spectra averaged over these simulations are taken as the input for training, which will be called "cascade-coarse-grained input". We would like to find out whether such an event averaging will improve the network performance due to the statistics enhancement or worsen it due to the information loss.

In this sub-scenario, 2 training datasets are generated by the iEBE-VISHNU hybrid model with the fluctuating MC-Glauber initial condition in the centrality range 0–50%. The details are shown in Table 3 in B. In total, 15747 events are generated with two different EoSs. The hadronic cascade is repeated 30 times after each hydrodynamics evolution. The spectra averaged over these 30 events are taken as the input to the network. The validation accuracy with these cascade-coarse-grained spectra $\rho_c(p_T, \phi)$ can achieve about 92%. One can see that such averaging over cascade-stage is beneficial in identifying the EoS information in early dynamics from the final-state particle spectra. This means that the statistics matters a lot for using particle spectra to decode the EoS information.

4.1.3 Event-fine-averaged input, switch at $T_{sw} = 137$ MeV

One drawback of the above average procedure is that the separation of collision dynamics into hydrodynamic and hadronic cascade stage is purely theoretical. Thus from a realistic point of view, an averaging procedure based on experimentally controllable event filtering is preferable. In this sub-scenario, spectra are averaged within the same fine centrality bin (with 1% width) instead, which will be called "event-fine-averaged input" in the following. To be specific, we average the spectra of 30 random events within the same fine centrality bin in Tables 1 and 2 as the input to the network to accumulate the statistics. Figure 7 in C shows the 30-events-fine-averaged normalized p_T spectra and the elliptic flow v_2 as a function of p_T of these training datasets with two EoSs. These two one-dimensional traditional observables are still not distinguishable by eye. By comparing with the corresponding event-by-event observables as shown in Fig. 6, one can see that the fluctuations are significantly reduced in the 30-events-fine-averaged spectra. This manner of averaging reduces the fluctuations from the initial conditions besides that from hadronic cascade and resonance decays. Consequently, a surprisingly obvious improvement for the CNN performance in classifying the two types of EoS is made. The validation accuracy reaches about 99% with the 30-events-fine-averaged spectra $\rho_a(p_T, \phi)$ after 1000 epochs training, a value similar to that in the pure hydrodynamic case [39]. In principal, one can include more datasets generated in different fine centrality bins for training. However, we confirm that it's enough to use the datasets simulated only in 6 representative fine centrality bins as in Tables 1 and 2, respectively, for training, since the predictive performances on the datasets simulated in other unselected fine centrality bins are as high as the training accuracy. This demonstrates that non-trivial high-level correlations which are independent of the centrality bins are learned by the neural network.

After the training with validation, the trained network is confronted with the testing data, which are generated with
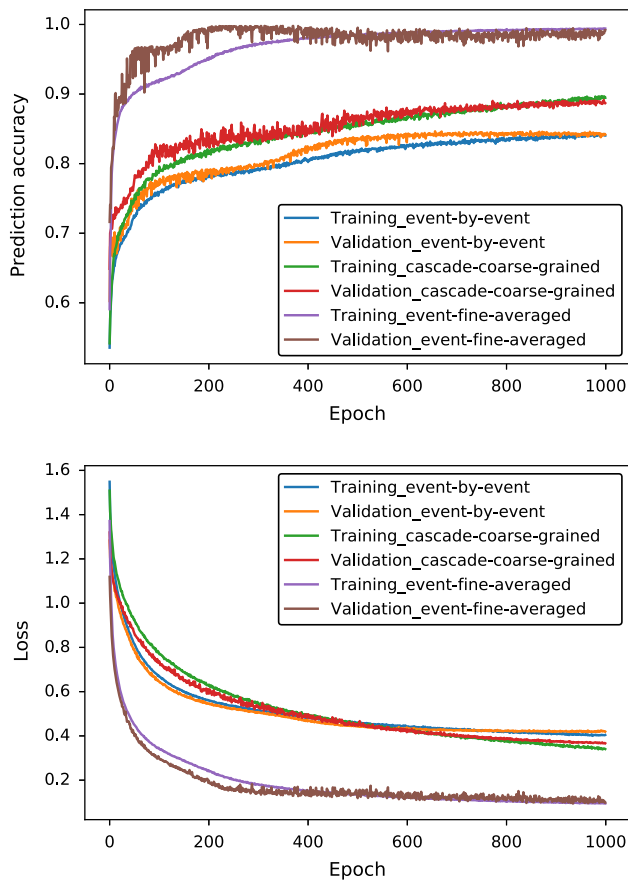
Fig. 2 Training and validation accuracy (upper panel) and loss (lower panel) in three different sub-scenarios with switching temperature $T_{sw} = 137$ MeV. These three sub-scenarios refer to the 30-events-fine-averaged spectra (purple and brown), the cascade-coarse-grained spectra (red and green) as well as the event-by-event spectra (blue and orange)

different physical parameter sets in simulations to explore the network's generalizability. In Tables 4 and 5 in B we show the predictive performance of the neural network trained with the 30-events-fine-averaged spectra. A testing accuracy 95% on average is obtained on the testing data simulated in the centrality range 0–50% with MC-Glauber or MCKLN initial conditions. This evidently demonstrates that the trained neural network is robust against different model setups such as initial conditions, τ_0, η/s and T_{sw} in a range between [130, 142] MeV. We observe a slight centrality dependence of the predictive accuracy in the collision system Pb+Pb $\sqrt{s_{NN}} = 2.76$ TeV, which decreases for more peripheral events.

4.1.4 A hierarchy of the accuracy in the above sub-scenarios

Figure 2 shows the training and validation accuracy (upper panel) and loss (lower panel), respectively, by the CNN with

the same setup for the first 1000 epochs in three aforementioned sub-scenarios. In each sub-scenario, training and validation accuracy (loss) are still close after 1000 epochs training, which implies that over-fitting is avoided. Besides, the network has not been sufficiently trained in the cascade-coarse-grained sub-scenario after 1000 epochs as the accuracy (loss) is still increasing (decreasing).

A clear hierarchy of the prediction accuracy is observed when the averaging is performed over more and more stages of the simulated dynamics. The CNN with event-by-event spectra gives the lowest accuracy, while the one with the 30-events-fine-averaged spectra gives the highest one, which is as high as in the pure hydrodynamic study [39].

4.2 Hybrid model with early transition to cascade

The scenario with early transition from hydrodynamics to hadronic cascade in hybrid modeling is in accordance with a widely used choice of the switching temperature $T_{sw} > 150$ MeV. This scenario is different from the one discussed in the previous subsection in two aspects. Firstly, the higher switching temperature decreases the contribution from the primordial pions which are directly emitted from the hydrodynamic evolution, and increases the contribution from resonance decays. Secondly, the elongated duration of the hadronic cascade stage may further blur out the imprint of the phase transition encoded in the final-state particle spectra. In the following, we will study how a higher switching temperature affects the performance of the CNN in three aforementioned sub-scenarios, respectively.

4.2.1 Event-by-event input, switch at $T_{sw} > 150$ MeV

In this sub-scenario, 9 training datasets are generated by the iEBE-VISHNU hybrid model with the fluctuating MC-Glauber initial condition in the centrality range 0–50%. The switching temperature is $T_{sw} = 160$ MeV. Two different values for the equilibration time τ_0 and the ratio of shear viscosity to entropy η/s are used in the simulations. The details are shown in Tables 6 and 7 in B. In total, about 60000 events are generated with two different EoS types.

The validation accuracy is found to be about 78% for the CNN trained with these event-by-event spectra as input. This validation accuracy is lower than that in the sub-scenario with late transition (switching temperature $T_{sw} = 137$ MeV). This decrease in the validation accuracy can be understood as a result of the increased contribution from resonance decays and the elongated duration of the hadronic rescattering.

4.2.2 Cascade-coarse-grained input, switch at $T_{sw} > 150$ MeV

In this sub-scenario, the cascade-coarse-grained pion spectra $\rho_c(p_T, \phi)$ are taken as the input to the CNN. 2 training datasets are generated in analogy to the previous late transition case, by the iEBE-VISHNU hybrid model with the fluctuating MC-Glauber initial condition in the centrality range 0–50% with the hadronic cascade simulated 30 times individually after each hydrodynamic evolution. The switching temperature T_{sw} is set to be 155 or 160 MeV. The details are shown in Table 8 in B. About 24000 events with two different EoSs are generated in total. The validation accuracy is found to be 87.5% at most, which is also lower than that in previous sub-scenario with late transition to cascade.

4 testing datasets are generated in this sub-scenario as shown in Table 9 in B in the centrality range 0–50%. Both MC-Glauber and MCKLN initial conditions are used, and simulation parameters are varied from the training datasets to check the generalizability of the CNN. After training and validating the neural network, the testing accuracy on these datasets is 83% on average, which is slightly lower than the validation accuracy.

4.2.3 Event-fine-averaged input, switch at $T_{sw} > 150$ MeV

In this sub-scenario, the 30-events-fine-averaged spectra for training is explored with the switching temperature $T_{sw} = 160$ MeV. This input is generated by the average over the spectra of 30 independent events within the same fine centrality bins (with 1% width) shown in Tables 6 and 7. The validation accuracy can also reach up to 99% in this sub-scenario as in the previous late transition one. The testing accuracy is up to 95% on average on the testing datasets as shown in Table 10 in the B. We also observe a slight centrality dependence of the predictive accuracy in the collision system Au+Au $\sqrt{s_{NN}} = 200$ GeV, which decreases for more peripheral events.

Its also interesting to further check the performance of the neural network on the testing datasets which employ temperature-dependent shear viscosities. Here taking this sub-scenario for example, we evaluated the network's prediction accuracy on the testing datasets in Table 11 where four temperature-dependent shear viscosities are employed in hybrid simulations as shown in Fig. 5 (labelled as 1–4, respectively). The first two are taken from Ref. [92], while the last two are taken from the Bayesian analysis estimations [35,36], respectively). The results show that the performance is robust against the setup of these temperature-dependent shear viscosities as compared with Table 10.

4.3 Comparison with fully-connected deep neural network

As already discussed in Sect. 4.1, the event-by-event and 30-events-fine-averaged normalized p_T spectra and elliptic flow v_2 with two different EOS from all centrality bins in Tables 1 and 2, as shown in Figs. 6 and 7, respectively, are non-distinguishable within the range of event-by-event fluctuations. However, one can observe that the peaks of the normalized p_T spectra with Q-EOS are higher than that with L-EOS on the whole. In Figs. 8, 9 and 10 in C, we show the event-by-event, 30-events-fine-averaged and all-events-fine-averaged normalized p_T spectra (left panel) and elliptic flow v_2 (right panel) solely from centrality bin 14–15% in Pb+Pb collision $\sqrt{s_{NN}} = 2.76$ TeV in Table 1, respectively. Within the same centrality bin one can see that the all-events-fine-averaged normalized p_T spectra are distinguishable with respect to different EOSs, 30-events-fine-averaged normalized p_T spectra are almost distinguishable from certain p_T bins, while the event-by-event normalized p_T spectra are still not. In Fig. 11 in C, we show the all-events-fine-averaged normalized p_T spectra (upper left panel) and elliptic flow v_2 (upper right panel) as well as the first (lower left panel), second (lower middle panel) and third (lower right panel) derivatives of the normalized p_T spectra from all centrality bins in Tables 1 and 2. One can see that these all-events-fine-averaged normalized p_T spectra are not distinguishable again by the human eye. Their derivatives are also helpful to distinguish the EoS in certain p_T bins, which might lead us to construct novel observables from normalized p_T spectra in the future. Inspired with this observation, we use the normalized p_T spectra as the input to a fully-connected DNN to distinguish the EOSs as a first try. In this case, the normalized p_T spectra are regarded as a whole instead of isolated points at each p_T bin as regarded by the human eye, and high-level correlations including but not limited to high-order derivatives can be extracted supervisely.

We train a fully-connected DNN[1] with the event-by-event normalized p_T spectra from all centrality bins in Tables 6 and 7 as the input. The validation accuracy is about 74%, which is below that by CNN with two-dimensional spectra, about 78%. Here the correlations are not very strong in both cases due to the fluctuations from the particlization and "afterburner". When the 30-events-fine-averaged normalized p_T spectra are taken as the input instead, the validation accuracy is about 97%, which is also a little below that by CNN with two-dimensional spectra, about 99%. Here the correlations are very strong in both cases. As for the

---

[1] This fully-connected DNN consists of two hidden dense layers of size 128 and 256, respectively, and each is followed by a dropout [93] (with a rate of 0.5) and PReLu activation layer [94]. These two dense layers are initialized with "He normal" initializer [94] and constrained with L2 regularization [95].
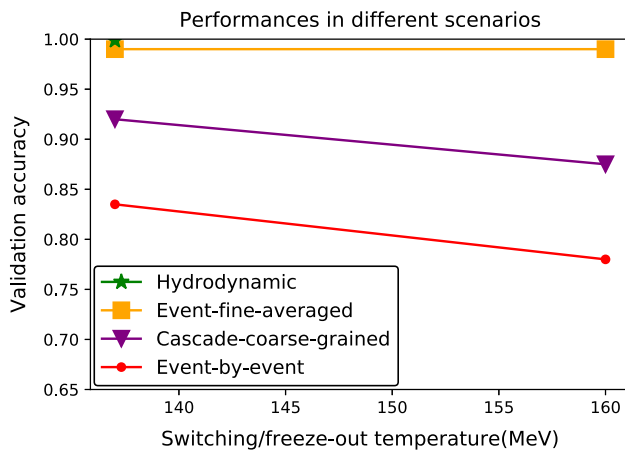
Fig. 3 Comparison between the validation accuracy in all the different sub-scenarios studied. The green star depicts the pure hydrodynamic result [39]. The orange square, the purple triangle and the red filled circle symbols depict the results for the 30-events-fine-averaged, cascade-coarse-grained and event-by-event spectra, respectively, in different switching temperatures

testing accuracy, CNN with two-dimensional spectra outperforms fully-connected DNN with one-dimensional spectra by about 8% with 30-events-fine-averaged spectra. Apparently, in the above cases, fully-connected DNN with one-dimensional normalized p_T spectra can capture the main correlations, while CNN with two-dimensional spectra performs better and improves the generalizability.

When the event-by-event normalized p_T spectra from all centrality bins in Tables 1 and 2 with $T_{sw} = 137$ MeV and in Tables 6 and 7 with $T_{sw} = 160$ MeV are taken as the input to the fully-connected DNN, the validation accuracy is about 62%, which is much lower than that by CNN with two-dimensional spectra, about 69%. This shows that when physical parameters in the simulation model vary a lot in the generation of the training data, the normalized p_T spectra are more difficult to distinguish and CNN with two-dimensional spectra will outperform fully-connected DNN with one-dimensional normalized p_T spectra.

5 Summary and conclusion

We extended a previous exploratory study on identifying EoS in the modeling of heavy ion collisions from hadron spectra using DL technique [39]. In this extended study, we consider more realistic hybrid modeling for heavy-ion collisions, where hadronic cascade "afterburner" with finite number of particles and resonance decays are properly taken into account. In the hybrid modeling the final-state particle spectra are histograms containing large fluctuations and thus are different from those in the previous study [39], which are smooth hadron spectra from Cooper-Frye prescription with

perfect statistics. Fig. 3 summarizes the predictive performances on the validation datasets in the above exploratory studies of different sub-scenarios.

We have demonstrated that, after the hydrodynamic evolution, stochastic particlization, hadronic cascade and resonance decays, the information about EoS in early dynamics is preserved in the final-state pion spectra, from the perspectives of deep CNN, as shown in Fig. 3. The event-by-event input for the network can reveal the EoS-type information with about 80% classification accuracy in binary classification setup.

The downward trend for the performance of network in validation with respect to the switching temperature in Fig. 3, implies that more stochasticity from the resonance decays and the elongated hadronic cascade will diminish the correlation between the EoS information in the early dynamics and the final-state particle spectra. This is in accordance with the common physical interpretation.

Finally, the hierarchy of the validation accuracy in different sub-scenarios in Fig. 3 shows that proper enhancement of statistics and reduction of fluctuations from either the final hadronic dynamics or together with the initial conditions in the input data are found to facilitate the revealing of the EoS information by the network from final-state particle spectra.

In conclusion, deep CNN can decode the imprint of the EoS in hydrodynamic evolution (encoded within the phase transition dynamics) on the final-state pion spectra from heavy-ion collisions. The good performance of the network does demonstrate that this "EoS-encoder" works. The fingerprint of the early dynamics of the bulk matter is not washed out by the evolution even when stochasticity is increased due to the hadronization and sequential hadron dynamics. Deep CNN provides an effective decoding method to extract high-level correlations from two-dimensional final-state pion spectra, which are immune to different physical factors, such as centrality bins. In relatively simple cases, fully-connected deep neural network can also identify the EoS from normalized pion p_T spectra with close validation accuracy as CNN does, which can lead us to discover new observables sensitive to EoS from normalized pion p_T spectra. The generalizability of the learned features with respect to other simulation models also depends on the simulation model for the training data generation. In the present study, the training data is generated with well tested iEBE-VISHNU (VISHNew + UrQMD) hybrid model. In the future we will explore how to capture the features which can be generalized to the testing data from other models as well as experimental data. Possible applications of the framework developed here can be extended to classifying fluctuating initial conditions, extracting transport coefficients of QCD matter, analysis of real experimental data filtering and pre-processing, and detector calibration.

Data Availability Statement This manuscript has no associated data or the data will not be deposited. [Authors' comment: The data that support the training and testing of the neural networks and the plots within this paper as well as other findings of this study are available from the corresponding author upon request].

Appendix A: Neural network structure

Figure 4 shows the neural network architecture. We use three convolutional layers and one subsequent fully-connected layer. All the convolutional layers and the fully-connected one are followed by a batch normalization [96], PReLu activation [94], dropout [93] (with a rate of 0.2 and 0.5, respectively) and average pooling (of pool size 2×2, following last two convolutional layers only) layer, one by one. There are 16, 16, 32 filters of size 8×8, 7×7 and 6×6, respectively, in these three convolutional layers, scanning through the input $\rho(p_T, \Phi)$, or the previous layers, and creating 16, 16, 32 features of size 24×24, 24×24, 12×12, respectively. The weight and bias matrix of these convolutional layers are initialized with "He normal" initializer [94], i.e. truncated normal distribution with zero mean and standard deviation $\sqrt{2/N_{in}}$ where N_{in} is the number of input units in the weight tensor. They are constrained with L2 regularization [95]. Each neuron in a convolutional layer does connect only locally to a small chunk of neurons in the previous layer by a convolution operation. This is a key reason for the suc-

| pion spectra 24x24 | 16 features 24x24 | 16 features 12x12 | 32 features 6x6 | | flattened 1152 | fc 128 | output layer | EOS |

crossover

1st order

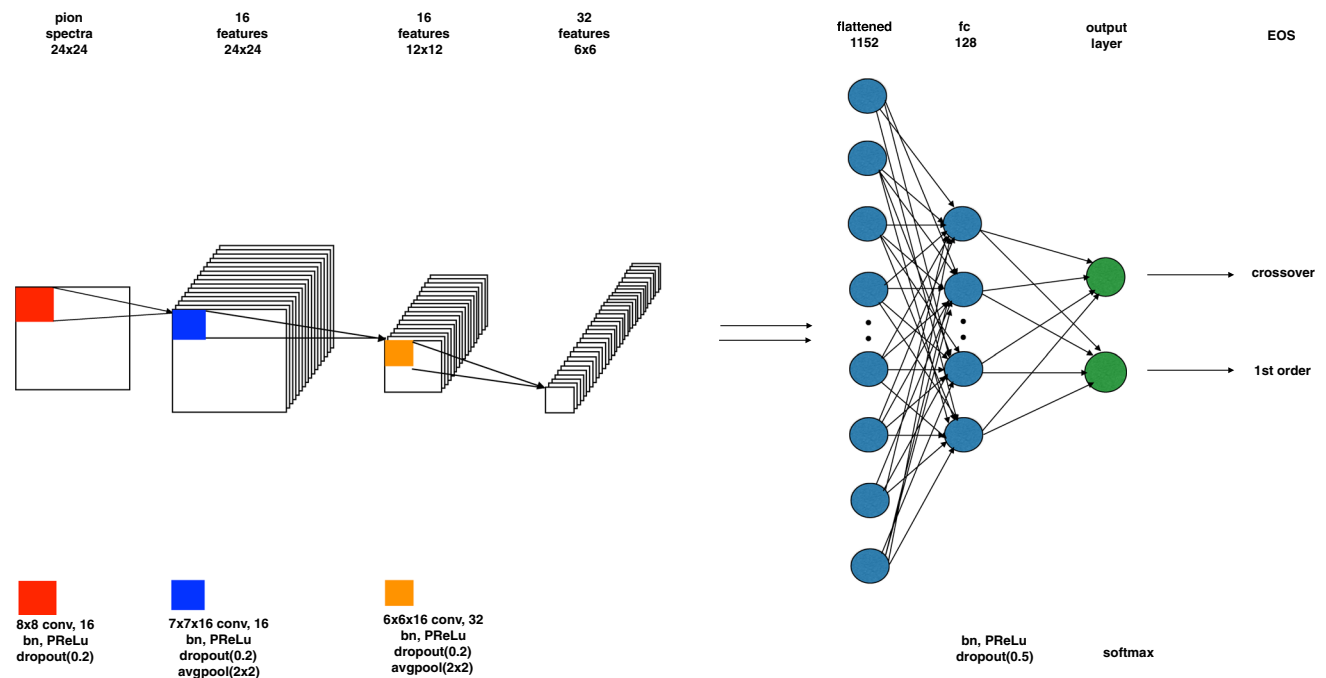| 8x8 conv, 16 bn, PReLu dropout(0.2) | 7x7x16 conv, 16 bn, PReLu dropout(0.2) avgpool(2x2) | 6x6x16 conv, 32 bn, PReLu dropout(0.2) avgpool(2x2) | | bn, PReLu dropout(0.5) | softmax |

Fig. 4 The architecture of our convolution neural network (CNN) for identifying the QCD transition type by using pion spectra with 24 transverse momentum p_T bins and 24 azimuthal angle Φ bins

cess of the CNN architecture. Dropout, batch normalization, PReLU and L2 regularization, all work together to prevent overfitting, which will generate model-parameter-dependent features from the training dataset and thus hinder the generalizability of the method. The resulting 32 features of size 6×6 from the last average pooling layer are flattened and connected to a 128-neuron fully-connected layer. The output layer is another fully-connected layer with softmax activation and 2 special neurons which indicate the type of the EoS. There are overall 203194 trainable and 120 non-trainable parameters in the present neural network.

The supervised learning is performed in tackling this binary classification task with the L-EOS case, labeled by (1, 0), and the Q-EOS case, labeled by (0, 1). The difference between the true label and the predicted label from the two output neurons is quantified by the cross entropy [97], which plays the role of the loss function $l(\theta)$, where θ are the trainable parameters of the neural network. The training minimizes the loss function by updating $\theta \rightarrow \theta - \delta\theta$. Here $\delta\theta = \alpha \partial l(\theta)/\partial\theta$, where α is the learning rate, with initial value 0.0001, which is adaptively changed by the AdaMax method [98].

The architecture is built by Keras with a Theano backend. The training datasets are fed into the network in batches with an empirically selected size of 128. One traversal of all the batches in the training datasets is called one epoch. The training datasets are reshuffled before each epoch to speed-up the convergence. The neural network is trained with 1000 epochs. The model parameters are saved to a new checkpoint whenever a smaller validation loss is encountered.

Appendix B: Collection of the training data and predictions on the testing data

See Tables 1, 2 and Fig. 5.

Table 1 Training datasets 1: numbers of event-by-event spectra $\rho(p_T, \Phi)$ computed by the iEBE-VISHNU hybrid model with the MC-Glauber initial conditions in the centrality range 0–60%. The ratio of shear viscosity to entropy density $\eta/s = 0.08$. The equilibration time $\tau_0 = 0.5$ fm/c. The switching temperature $T_{sw} = 137$ MeV. The collision system is Pb + Pb at $\sqrt{s_{NN}} = 2.76$ TeV

Training datasets 1		
Centrality bin	L-EOS	Q-EOS
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 2 Training datasets 2: numbers of event-by-event spectra $\rho(p_T, \Phi)$ computed by the iEBE-VISHNU hybrid model with the MC-Glauber initial conditions in the centrality range 0–60%. The ratio of shear viscosity to entropy density $\eta/s = 0.00$. The equilibration time $\tau_0 = 0.4$ fm/c. The switching temperature $T_{sw} = 137$ MeV. The collision system is Au + Au at $\sqrt{s_{NN}} = 200$ GeV

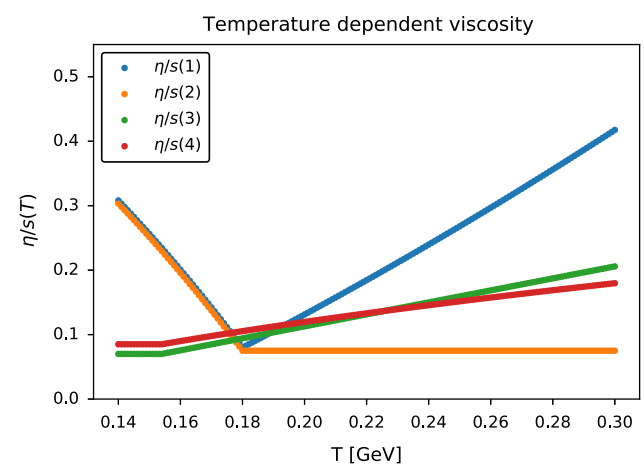| Training datasets 2		
Centrality bin	L-EOS	Q-EOS
0%–1%	979	1024
10%–11%	2560	2560
20%–21%	1024	1024
30%–31%	1024	1024
40%–41%	2560	2560
50%–51%	2816	2816

Fig. 5 Four temperature-dependent viscosities from Refs. [35,36,92]

Appendix C: Traditional observables from the training data

See Tables 3, 4, 5, 6, 7, 8, 9, 10 and 11 and Figs. 6, 7, 8, 9, 10 and 11.

Table 3 Training datasets 3: numbers of cascade-coarse-grained spectra $\rho_c(p_T, \Phi)$ computed by the iEBE-VISHNU hybrid model with the MC-Glauber initial conditions in the centrality range 0–50%

Training datasets 3							
Centrality bin	$\sqrt{s_{NN}}$ [TeV]	Ini. Cond.	τ_0 (fm/c)	η/s	T_{sw} [MeV]	L-EOS	Q-EOS
0%–50%	Au + Au 0.2	MC-G	0.4	0.16	137	3990	4096
0%–50%	Pb + Pb 2.76	MC-G	0.6	0.08	137	3830	3835

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

Table 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%

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

Table 6 Training datasets 4: numbers of event-by-event spectra $\rho(p_T, \Phi)$ computed by the iEBE-VISHNU hybrid model with the MC-Glauber initial conditions in the centrality range 0–50%. The ratio of shear viscosity to entropy density $\eta/s = 0.08$. The equilibration time $\tau_0 = 0.6$ fm/c. The switching temperature $T_{sw} = 160$ MeV. The collision system is Pb + Pb at $\sqrt{s_{NN}} = 2.76$ TeV

Training datasets 4		
Centrality bin	L-EOS	Q-EOS
15%–16%	2560	2560
20%–21%	2560	2560
34%–35%	3840	3840
44%–45%	3840	3840

Table 7 Training datasets 5: numbers of event-by-event spectra $\rho(p_T, \Phi)$ computed by the iEBE-VISHNU hybrid model with the MC-Glauber initial conditions in the centrality range 0–50%. The ratio of shear viscosity to entropy density $\eta/s = 0.16$. The equilibration time $\tau_0 = 0.4$ fm/c. The switching temperature $T_{sw} = 160$ MeV. The collision system is Au + Au at $\sqrt{s_{NN}} = 200$ GeV

Training datasets 5		
Centrality bin	L-EOS	Q-EOS
10%–11%	2560	2560
15%–16%	2560	2560
25%–26%	2560	2560
34%–35%	3840	3840
44%–45%	3840	3840

Table 8 Training datasets 6: numbers of cascade-coarse-grained spectra $\rho_c(p_T, \Phi)$ computed by the iEBE-VISHNU hybrid model with the MC-Glauber initial conditions in the centrality range 0–50%

Training datasets 6							
Centrality bin	$\sqrt{s_{NN}}$ [TeV]	Ini. Cond.	τ_0 (fm/c)	η/s	T_{sw} [MeV]	L-EOS	Q-EOS
0%–50%	Au + Au 0.2	MC-G	0.4	0.16	155	4608	4608
0%–50%	Au + Au 0.2	MC-G	0.4	0.00	155	3072	3072
0%–50%	Au + Au 0.2	MC-G	0.4	0.16	160	9724	9724
0%–50%	Pb + Pb 2.76	MC-G	0.6	0.08	155	5770	5521

Table 9 Predictive accuracy on testing datasets 3: cascade-coarse-grained spectra $\rho_c(p_T, \Phi)$ generated with the different $\sqrt{s_{NN}}$, initial conditions, η/s, τ_0, and T_{sw} in the centrality range 0–50%

Predictive accuracy for testing datasets 3								
Centrality bin	$\sqrt{s_{NN}}$ [TeV]	Ini. Cond.	τ_0 (fm/c)	η/s	T_{sw} [MeV]	L-EOS	Q-EOS	Accuracy
0%–50%	Au + Au 0.2	MCKLN	0.5	0.08	160	128	128	82%
0%–50%	Au + Au 0.2	MC-G	0.5	0.08	160	128	128	82.28%
0%–50%	Pb + Pb 2.76	MC-G	0.6	0.08	160	256	256	85%
0%–50%	Pb + Pb 2.76	MC-G	0.4	0.16	155	118	118	84.32%

Table 10 Predictive accuracy on testing datasets 4: 30-events-fine-averaged spectra $\rho_a(p_T, \Phi)$ generated with the different $\sqrt{s_{NN}}$, initial conditions, η/s, τ_0, and T_{sw} in the centrality range 0–30%

Predictive accuracy for testing datasets 4								
Centrality bin	$\sqrt{s_{NN}}$ [TeV]	Ini. Cond.	τ_0 (fm/c)	η/s	T_{sw} [MeV]	L-EOS	Q-EOS	Accuracy
15%–16%	Au + Au 0.2	MCKLN	0.6	0.16	160	640	640	95.59%
10%–11%	Au + Au 0.2	MC-G	0.4	0.12	160	2560	2560	100%
15%–16%	Au + Au 0.2	MC-G	0.4	0.12	160	2560	2560	99.8%
20%–21%	Au + Au 0.2	MC-G	0.4	0.12	160	2560	2560	94.9%
15%–16%	Au + Au 0.2	MC-G	0.4	0.00	155	2560	2560	74.86%
20%–21%	Au + Au 0.2	MC-G	0.4	0.12	155	1792	1792	88.8%
15%–16%	Pb + Pb 2.76	MC-G	0.6	0.08	155	2560	2560	99.99%
20%–21%	Pb + Pb 2.76	MC-G	0.6	0.08	155	2560	2560	99.78%

Table 11 Predictive accuracy on testing datasets 5: 30-events-fine-averaged spectra $\rho_a(p_T, \Phi)$ generated with the different $\sqrt{s_{NN}}$, initial conditions, temperature-dependent $\eta/s(T)$, τ_0, and T_{sw} in the centrality range 0–30%

| Predictive accuracy for testing datasets								
Centrality bin	$\sqrt{s_{NN}}$ [TeV]	Ini. Cond.	τ_0 (fm/c)	$\eta/s(T)$	T_{sw} [MeV]	L-EOS	Q-EOS	Accuracy
10%–11%	Au + Au 0.2	MC-G	0.4	1	160	512	512	100%
10%–11%	Au + Au 0.2	MC-G	0.4	2	160	512	512	100%
10%–11%	Au + Au 0.2	MC-G	0.4	3	160	512	512	100%
10%–11%	Au + Au 0.2	MC-G	0.4	4	160	512	512	100%
15%–16%	Au + Au 0.2	MC-G	0.4	4	160	512	512	99.51%
20%–21%	Au + Au 0.2	MC-G	0.4	3	160	512	512	98.34%
10%–11%	Au + Au 0.2	MCKLN	0.6	3	160	512	512	98.04%
10%–11%	Pb + Pb 2.76	MC-G	0.6	3	155	512	512	99.80%
25%–26%	Pb + Pb 2.76	MC-G	0.6	4	160	512	512	99.90%
35%–36%	Pb + Pb 2.76	MC-G	0.6	3	155	512	512	86.72%
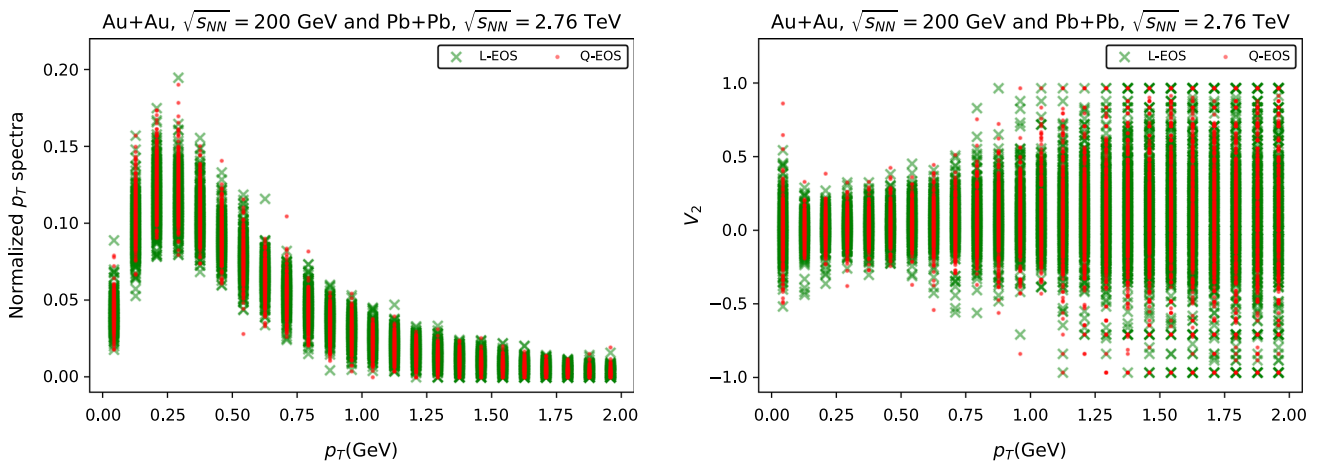
Fig. 6 Event-by-event normalized p_T spectra $dN/Ndydp_T$ (left panel) and elliptic flow v_2 as a function of p_T (right panel) of the training datasets in Tables 1 and 2 with two EoSs. Vertical discrepancy is event-by-event fluctuations. The green cross and the red point symbol depict the observables with L-EOS and Q-EOS, respectively. These events are generated in different centrality bins with $T_{sw} = 137$ MeV in two collision systems
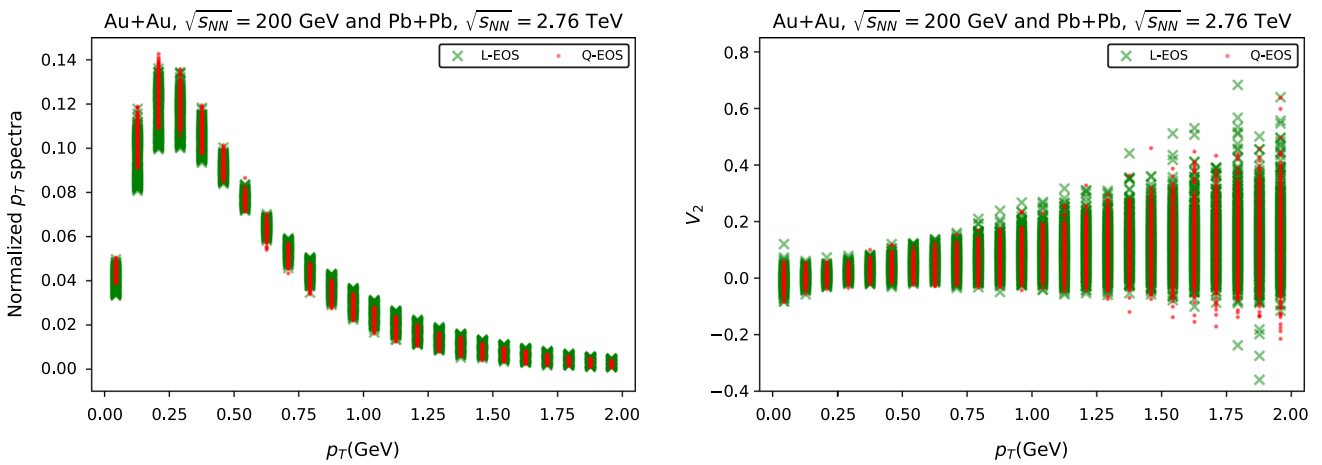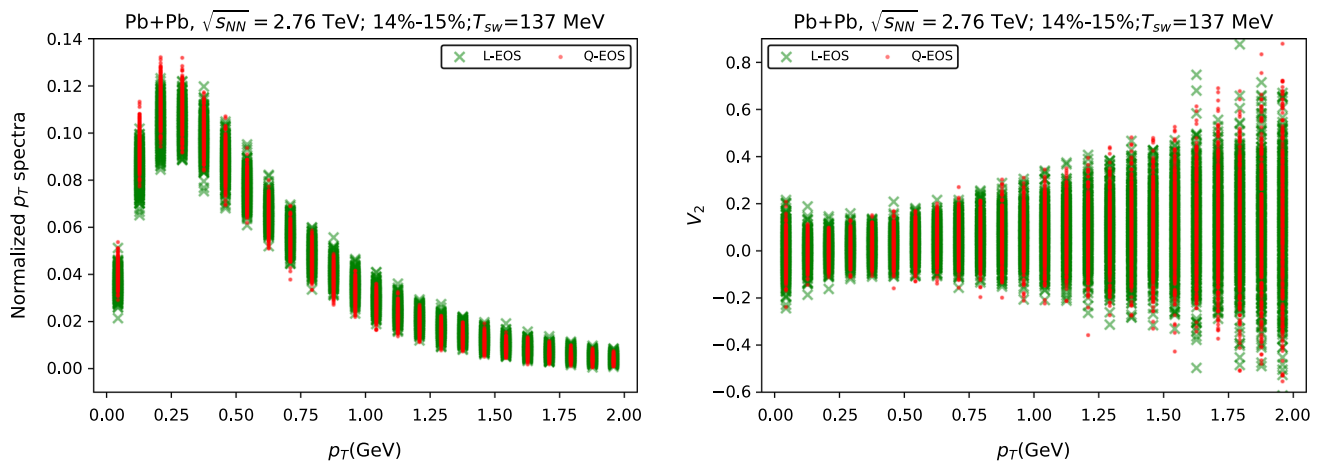

Fig. 7 Same as Fig. 6 but for 30-events-fine-averaged normalized p_T spectra $dN/Ndydp_T$ (left panel) and elliptic flow v_2 as a function of p_T (right panel)

Fig. 8 Event-by-event normalized p_T spectra $dN/Ndydp_T$ (left panel) and elliptic flow v_2 as a function of p_T (right panel) of the training datasets in Table 1 with two EoSs. The green cross and the red point symbol depict the observables with L-EOS and Q-EOS, respectively. These events are generated in centrality bin 14%–15% with $T_{sw} = 137$ MeV in two collision systems
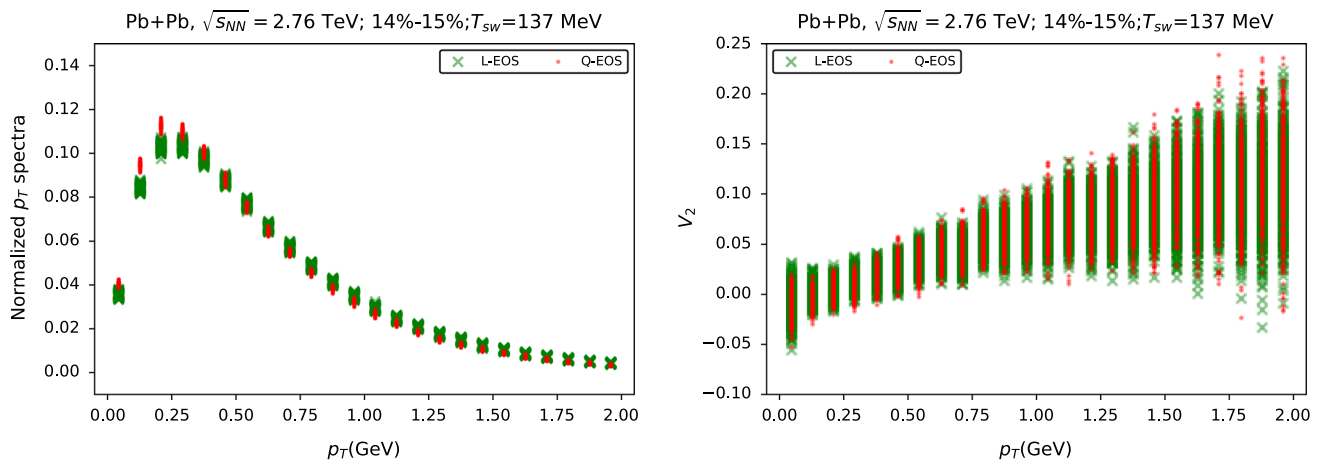

Fig. 9 Same as Fig. 8 but for 30-events-fine-averaged normalized p_T spectra $dN/Ndydp_T$ (left panel) and elliptic flow v_2 as a function of p_T (right panel)
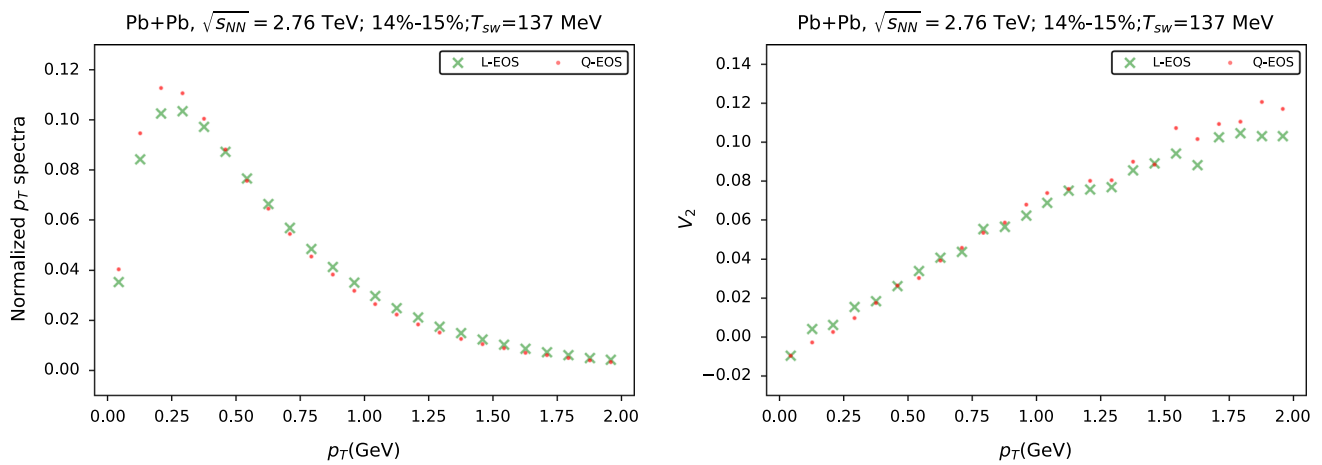
Fig. 10 All-events-fine-averaged normalized p_T spectra $dN/Ndydp_T$ (left panel) and elliptic flow v_2 as a function of p_T (right panel) of the training datasets in Table 1 with two EoSs. The green cross and the red point symbol depict the observables with L-EOS and Q-EOS, respectively. These events are generated in centrality bin 14%–15% with $T_{sw} = 137$ MeV in two collision systems
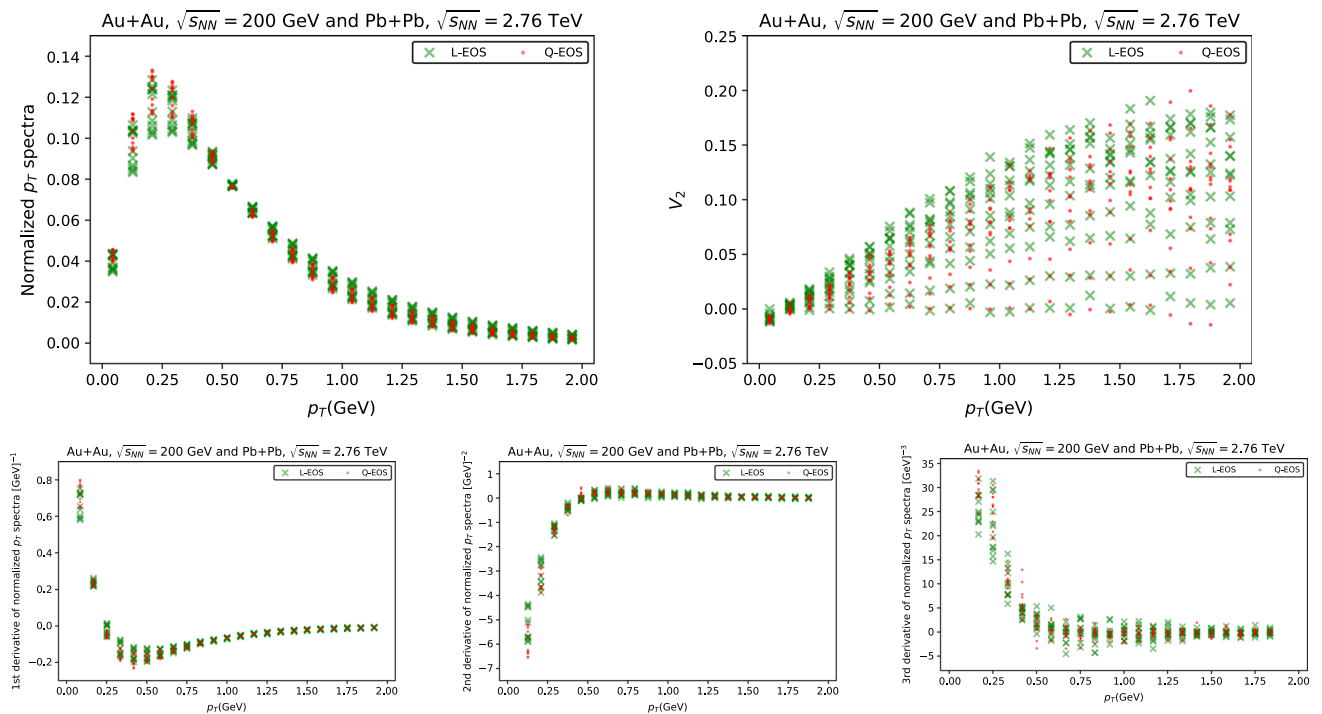

Fig. 11 All-events-fine-averaged normalized p_T spectra $dN/Ndydp_T$ (upper left panel) and elliptic flow v_2 as a function of p_T (upper right panel) and the first, second and third derivative of these normalized p_T spectra (lower panel) of the training datasets in Tables 1 and 2 with two EoSs. The green cross and the red point symbol depict the observables with L-EOS and Q-EOS, respectively. These events are generated in different centrality bins with $T_{sw} = 137$ MeV in two collision systems

References

1. Y. Aoki, G. Endrodi, Z. Fodor, S.D. Katz, K.K. Szabo, Nature **443**, 675 (2006). https://doi.org/10.1038/nature05120. arXiv:hep-lat/0611014

2. Y. Aoki, Z. Fodor, S.D. Katz, K.K. Szabo, Phys. Lett. B **643**, 46 (2006). https://doi.org/10.1016/j.physletb.2006.10.021. arXiv:hep-lat/0609068

3. F. Karsch, *Lectures on quark matter. Proceedings, 40. International Universitaetswochen for theoretical physics, 40th Winter School, IUKT 40: Schladming, Austria, March 3–10, 2001*, Lect. Notes Phys. **583**, 209 (2002). https://doi.org/10.1007/3-540-45792-5_6.0 arXiv:hep-lat/0106019

4. S.-X. Qin, L. Chang, H. Chen, Y.-X. Liu, C.D. Roberts, Phys. Rev. Lett. **106**, 172301 (2011)

5. C.S. Fischer, J. Luecker, Phys. Lett. B **718**, 1036 (2013)

6. C. Shi, Y.-L. Wang, Y. Jiang, Z.-F. Cui, H.-S. Zong, JHEP **07**, 014 (2014). https://doi.org/10.1007/JHEP07(2014)014

7. C. Shi, Y.-L. Du, S.-S. Xu, X.-J. Liu, H.-S. Zong, Phys. Rev. D **93**, 036006 (2016)

8. P. Costa, M. Ruivo, C. De Sousa, Phys. Rev. D **77**, 096001 (2008)

9. P. Costa, C. De Sousa, M. Ruivo, H. Hansen, Europhys. Lett. **86**, 31001 (2009)

10. K. Fukushima, Phys. Rev. D **77**, 114028 (2008). https://doi.org/10.1103/PhysRevD.77.114028

11. W.-J. Fu, Z. Zhang, Y.-X. Liu, Phys. Rev. D **77**, 014006 (2008). https://doi.org/10.1103/PhysRevD.77.014006

12. Y.-L. Du, Z.-F. Cui, Y.-H. Xia, H.-S. Zong, Phys. Rev. D **88**, 114019 (2013). https://doi.org/10.1103/PhysRevD.88.114019

13. B.-J. Schaefer, J.M. Pawlowski, J. Wambach, Phys. Rev. D **76**, 074023 (2007)

14. D. Nickel, Phys. Rev. D **80**, 074025 (2009)

15. V. Skokov, B. Friman, K. Redlich, Phys. Rev. C **83**, 054904 (2011)

16. N. Tahir, A. Adonin, C. Deutsch, V. Fortov, N. Grandjouan, B. Geil, V. Grayaznov, D. Hoffmann, M. Kulish, I. Lomonosov, V. Mintsev, D. Ni, D. Nikolaev, A. Piriz, N. Shilkin, P. Spiller, A. Shutov, M. Temporal, V. Ternovoi, S. Udrea, D. Varentsov, Nucl. Instrum. Methods Phys. Res. **544**, 16 (2005). Proceedings of the 15th International Symposium on Heavy Ion Inertial Fusion https://doi.org/10.1016/j.nima.2005.01.178

17. G. Rai, N. Ajitanand, J. Alexander, M. Anderson, D. Best, F. Brady, T. Case, W. Caskey, D. Cebra, J. Chance et al., Nucl. Phys. A **661**, 162 (1999)

18. J.-P. Lansberg, S. Brodsky, F. Fleuret, C. Hadjidakis, Few-Body Syst. **53**, 11 (2012)

19. J. Adams, M. Aggarwal, Z. Ahammed, J. Amonett, B. Anderson, D. Arkhipin, G. Averichev, S. Badyal, Y. Bai, J. Balewski et al., Nucl. Phys. A **757**, 102 (2005)

20. B. Muller, J. Schukraft, B. Wyslouch, Ann. Rev. Nucl. Part. Sci. **62**, 361 (2012)

21. B. Friman, C. Höhne, J. Knoll, S. Leupold, J. Randrup, R. Rapp, P. Senger, *The CBM Physics Book: Compressed Baryonic Matter in Laboratory Experiments*, vol. 814 (Springer, New York, 2011)

22. T. Ablyazimov, A. Abuhoza, R. Adak, M. Adamczyk, K. Agarwal, M. Aggarwal, Z. Ahammed, F. Ahmad, N. Ahmad, S. Ahmad et al., Eur. Phys. J. A **53**, 60 (2017)

23. A. Sissakian, A. Sorin, N. collaboration et al., J. Phys. G: Nucl. Part. Phys. **36**, 064069 (2009)

24. M. Luzum, P. Romatschke, Phys. Rev. C **78**, 034915 (2008)

25. J. Hofmann, H. Stoecker, U.W. Heinz, W. Scheid, W. Greiner, Phys. Rev. Lett. **36**, 88 (1976). https://doi.org/10.1103/PhysRevLett.36.88

26. H. Stoecker, W. Greiner, Phys. Rep. **137**, 277 (1986)

27. C.M. Hung, E.V. Shuryak, Phys. Rev. Lett. **75**, 4003 (1995). https://doi.org/10.1103/PhysRevLett.75.4003. arXiv:hep-ph/9412360 [hep-ph]

28. D.H. Rischke, Y. Pursun, J.A. Maruhn, H. Stoecker, W. Greiner, Acta Phys. Hung. **A1**, 309 (1995). arXiv:nucl-th/9505014 [nucl-th]

29. J. Brachmann, S. Soff, A. Dumitru, H. Stoecker, J.A. Maruhn, W. Greiner, L.V. Bravina, D.H. Rischke, Phys. Rev. C **61**, 024909 (2000). https://doi.org/10.1103/PhysRevC.61.024909. arXiv:nucl-th/9908010 [nucl-th]

30. M.A. Stephanov, K. Rajagopal, E.V. Shuryak, Phys. Rev. Lett. **81**, 4816 (1998). https://doi.org/10.1103/PhysRevLett.81.4816. arXiv:hep-ph/9806219 [hep-ph]

31. M.A. Stephanov, Phys. Rev. Lett. **102**, 032301 (2009). https://doi.org/10.1103/PhysRevLett.102.032301. arXiv:0809.3450 [hep-ph]

32. V. Koch, in *Relativistic Heavy Ion Physics* (Springer, New York, 2010) pp. 626–652

33. J. Steinheimer, J. Randrup, Phys. Rev. Lett. **109**, 212301 (2012). https://doi.org/10.1103/PhysRevLett.109.212301. arXiv:1209.2462 [nucl-th]

34. S. Pratt, E. Sangaline, P. Sorensen, H. Wang, Phys. Rev. Lett. **114**, 202301 (2015)

35. J.E. Bernhard, J.S. Moreland, S.A. Bass, J. Liu, U. Heinz, Phys. Rev. C **94**, 024907 (2016)

36. J.E. Bernhard, J.S. Moreland, S.A. Bass, Nat. Phys. **15**, 1113 (2019)

37. J. Schmidhuber, Neural Netw **61**, 85 (2015)

38. Y. LeCun, Y. Bengio, G. Hinton, Nature **521**, 436 (2015)

39. L.-G. Pang, K. Zhou, N. Su, H. Petersen, H. Stcker, X.-N. Wang, Nat. Commun. **9**, 210 (2018). https://doi.org/10.1038/s41467-017-02726-3. arXiv:1612.04262 [hep-ph]

40. H. Huang, B. Xiao, H. Xiong, Z. Wu, Y. Mu, H. Song, (2018). arXiv:1801.03334 [nucl-th]

41. Y.-T. Chien, R. K. Elayavalli, arXiv:1803.03589 (2018)

42. J.E. Bernhard, J.S. Moreland, S.A. Bass, J. Liu, U. Heinz, Phys. Rev. C **94**, 024907 (2016). https://doi.org/10.1103/PhysRevC.94.024907

43. R. Utama, W.-C. Chen, J. Piekarewicz, J. Phys. G **43**, 114002 (2016)

44. R. Haake, arXiv:1709.08497 (2017)

45. S. Pratt, E. Sangaline, P. Sorensen, H. Wang, Phys. Rev. Lett. **114**, 202301 (2015). https://doi.org/10.1103/PhysRevLett.114.202301

46. K. Zhou, G. Endrdi, L.-G. Pang, H. Stcker, Phys. Rev. D **100**, 011501 (2019). https://doi.org/10.1103/PhysRevD.100.011501. arXiv:1810.12879 [hep-lat]

47. J.M. Urban, J.M. Pawlowski, (2018). arXiv:1811.03533 [hep-lat]

48. Y. Mori, K. Kashiwa, A. Ohnishi, PTEP **2018**, 023B04 (2018). https://doi.org/10.1093/ptep/ptx191. arXiv:1709.03208 [hep-lat]

49. P.E. Shanahan, D. Trewartha, W. Detmold, Phys. Rev. D **97**, 094506 (2018). https://doi.org/10.1103/PhysRevD.97.094506. arXiv:1801.05784 [hep-lat]

50. A. Tanaka, A. Tomiya, (2017), arXiv:1712.03893 [hep-lat]

51. P. Baldi, P. Sadowski, D. Whiteson, Nat. Commun **5** (2014)

52. P. Baldi, P. Sadowski, D. Whiteson, Phys. Rev. Lett. **114**, 111801 (2015)

53. J. Searcy, L. Huang, M.-A. Pleier, J. Zhu, Phys. Rev. D **93**, 094033 (2016)

54. J. Barnard, E.N. Dawe, M.J. Dolan, N. Rajcic, Phys. Rev. D **95**, 014018 (2017)

55. I. Moult, L. Necib, J. Thaler, J. High Energy Phys. **2016**, 153 (2016)

56. K. Schawinski, C. Zhang, H. Zhang, L. Fowler, G.K. Santhanam, Mon. Not. R. Astron. Soc. **467**, L110 (2017)

57. M. Erdmann, J. Glombitza, D. Walz, Astropart. Phys. **97**, 46 (2018)

58. Y. Fujimoto, K. Fukushima, K. Murase, Phys. Rev. D **98**, 023019 (2018). https://doi.org/10.1103/PhysRevD.98.023019

59. P. Mehta, D. J. Schwab, arXiv:1410.3831 [stat.ML] (2014)

60. J. Carrasquilla, R. G. Melko, Nat. Phys. (2017)

61. G. Carleo, M. Troyer, Science **355**, 602 (2017)

62. G. Torlai, R.G. Melko, Phys. Rev. B **94**, 165134 (2016)
63. P. Broecker, J. Carrasquilla, R. G. Melko, S. Trebst, arXiv:1608.07848 [cond-mat.str-el] (2016)
64. K. Chng, J. Carrasquilla, R.G. Melko, E. Khatami, Phys. Rev. X **7**, 031038 (2017)
65. Y.-Z. You, Z. Yang, X.-L. Qi, Phys. Rev. B **97**, 045153 (2018)
66. W. Florkowski, R. Ryblewski, M. Strickland, Nucl. Phys. A **916**, 249 (2013)
67. Y.V. Kovchegov, E. Levin, *Quantum Chromodynamics at High Energy* (Cambridge University Press, Cambridge, 2012)
68. H. Stoecker et al., J. Phys. G **43**, 015105 (2016). https://doi.org/10.1088/0954-3899/43/1/015105. arXiv:1509.00160 [hep-ph]
69. H. Stoecker et al., Proceedings, 4th International Symposium on Strong Electromagnetic Fields and Neutron Stars (STARS2015): Havana, Cuba, May 10–16, 2015. Astron. Nachr. **336**, 744 (2015). https://doi.org/10.1002/asna.201512252. arXiv:1509.07682 [hep-ph]
70. V. Vovchenko, I.A. Karpenko, M.I. Gorenstein, L.M. Satarov, I.N. Mishustin, B. Kaempfer, H. Stoecker, Phys. Rev. C **94**, 024906 (2016). https://doi.org/10.1103/PhysRevC.94.024906. arXiv:1604.06346 [nucl-th]
71. Z. Xu, K. Zhou, P. Zhuang, C. Greiner, Phys. Rev. Lett. **114**, 182301 (2015). https://doi.org/10.1103/PhysRevLett.114.182301. arXiv:1410.5616 [hep-ph]
72. K. Zhou, Z. Xu, P. Zhuang, C. Greiner, Phys. Rev. D **96**, 014020 (2017). https://doi.org/10.1103/PhysRevD.96.014020. arXiv:1703.02495 [hep-ph]
73. U. Heinz, in *Relativistic Heavy Ion Physics* (Springer, New York, 2010) pp. 240–292
74. P. Romatschke, Int. J Mod. Phys. E **19**, 1 (2010)
75. H. Song, U. Heinz, Phys. Rev. C **77**, 064901 (2008)
76. B. Schenke, S. Jeon, C. Gale, Phys. Rev. Lett. **106**, 042301 (2011)
77. L. Pang, Q. Wang, X.-N. Wang, Phys. Rev. C **86**, 024911 (2012)
78. L.-G. Pang, Y. Hatta, X.-N. Wang, B.-W. Xiao, Phys. Rev. D **91**, 074027 (2015)
79. H. Stoecker, Nucl. Phys. A **750**, 121 (2005)
80. S.A. Bass, M. Belkacem, M. Bleicher, M. Brandstetter, L. Bravina, C. Ernst, L. Gerland, M. Hofmann, S. Hofmann, J. Konopka et al., Prog. Part. Nucl. Phys. **41**, 255 (1998)
81. M. Bleicher, E. Zabrodin, C. Spieles, S.A. Bass, C. Ernst, S. Soff, L. Bravina, M. Belkacem, H. Weber, H. Stoecker et al., J. Phys. G **25**, 1859 (1999)
82. J. Steinheimer, J. Aichelin, M. Bleicher, H. Stoecker, Phys. Rev. C **95**, 064902 (2017). https://doi.org/10.1103/PhysRevC.95.064902. arXiv:1703.06638 [nucl-th]
83. C. Shen, Z. Qiu, H. Song, J. Bernhard, S. Bass, U. Heinz, Comput. Phys. Commun. **199**, 61 (2016)
84. W. Broniowski, M. Rybczyński, P. Bożek, Comput. Phys. Commun. **180**, 69 (2009)
85. B. Alver, M. Baker, C. Loizides, P. Steinberg, arXiv:0805.4411 (2008)
86. C. Loizides, J. Nagle, P. Steinberg, SoftwareX **1**, 13 (2015)
87. D. Kharzeev, E. Levin, M. Nardi, Phys. Rev. C **71**, 054903 (2005)
88. D. Kharzeev, E. Levin, M. Nardi, Nucl. Phys. A **747**, 609 (2005)
89. P. Huovinen, P. Petreczky, Nucl. Phys. A **837**, 26 (2010)
90. J. Sollfrank, P. Huovinen, M. Kataja, P. Ruuskanen, M. Prakash, R. Venugopalan, Phys. Rev. C **55**, 392 (1997)
91. Z.-W. Lin, C.M. Ko, B.-A. Li, B. Zhang, S. Pal, Phys. Rev. C **72**, 064901 (2005). https://doi.org/10.1103/PhysRevC.72.064901. arXiv:nucl-th/0411110 [nucl-th]
92. H. Niemi, G.S. Denicol, P. Huovinen, E. Molnar, D.H. Rischke, Phys. Rev. Lett. **106**, 212302 (2011)
93. N. Srivastava, G.E. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, J. Mach. Learn. Res. **15**, 1929 (2014)
94. K. He, X. Zhang, S. Ren, J. Sun, in *Proceedings of the IEEE international conference on computer vision***2015**) pp. 1026–1034
95. A. Y. Ng, in *Proceedings of the twenty-first international conference on Machine learning* (organization ACM, 2004) p. 78
96. S. Ioffe, C. Szegedy, International Conference on Machine Learning 448–456 (2015)
97. S. Kullback, R.A. Leibler, Ann. Math. Stat. **22**, 79 (1951)
98. D. P. Kingma, J. Ba, arXiv:1412.6980 (2014)