




## Detecting the chiral magnetic effect via deep learning

Yuan-Sheng Zhao <sup>1,\*</sup> Lingxiao Wang <sup>2,†</sup> Kai Zhou <sup>2,‡</sup> and Xu-Guang Huang<sup>1,3,§</sup>

<sup>1</sup>Physics Department and Center for Particle Physics and Field Theory, Fudan University, Shanghai 200433, China

<sup>2</sup>Frankfurt Institute for Advanced Studies, Ruth Moufang Strasse 1, D-60438, Frankfurt am Main, Germany

<sup>3</sup>Key Laboratory of Nuclear Physics and Ion-beam Application (MOE), Fudan University, Shanghai 200433, China



(Received 10 June 2021; revised 17 August 2022; accepted 18 October 2022; published 3 November 2022)

The search for chiral magnetic effect (CME) in heavy-ion collisions has attracted long-term attention. Multiple observables are proposed, but all suffer large background contaminations. In this study, we construct an end-to-end *CME-meter* based on a deep convolutional neural network. After being trained over a dataset generated by a multiphase transport model, the *CME-meter* shows high accuracy in recognizing the CME-featured charge separation from the final-state pion spectra. It also exhibits remarkable robustness to diverse conditions including different collision energies, centralities, and elliptic flow backgrounds. In extrapolation tests, the *CME-meter* is validated in isobaric collisions and different model simulation, showing robust generalization performance. Moreover, based on variational approaches, we utilize the *DeepDream* method to derive the most responsive CME-spectrum that demonstrates the physical contents the machine learned.

DOI: [10.1103/PhysRevC.106.L051901](https://doi.org/10.1103/PhysRevC.106.L051901)

**Introduction.** Quantum chromodynamics (QCD) is the standard theory describing the physics of the strong interaction. Among the studies on QCD, the proposal of using chiral magnetic effect (CME) to reveal the vacuum structure of QCD is of great importance [1–3]. It predicts that in hot and dense quark-gluon plasma (QGP), the topological fluctuations of gluon fields can cause imbalance between the number of left-handed and right-handed quarks and this difference can induce an electric current under external magnetic field.

High-energy heavy ion collisions (HICs) provide an environment for CME to take place. However, QGP and the strong magnetic field required for giving rise to CME exist only in the early stages of the collisions. To retrieve the information of possible CME from the final-state hadrons, multiple observables were proposed [4–8], such as the  $\gamma$ -correlator (see definition in below). However, due to the large contributions of elliptic flow and other background noises [9–11], these observables cannot clearly recognize CME or its induced charge separation (CS) in QGP along the magnetic-field direction.

Although it is difficult to detect CME through specific observables, analyzing the final-state hadronic spectrum as a whole in the sense of big data may help reveal hidden

fingerprints of CME. Deep learning is a branch of machine learning with superiority in recognizing patterns with complex correlations [12,13]. With the hierarchical structure of neural networks, deep learning is particularly effective in tackling complex systems that cannot be easily handled by conventional techniques. Recently, significant progress has been made in applying deep learning to physics studies, including nuclear physics [14–22], particle physics [23–27], and condensed matter physics [28–32]. In this study, we explore the possibility of using deep learning to determine whether there are detectable final-state signals of CME that survive the collision dynamics and background interference.

**Methods.** In this section, we introduce a deep learning model containing convolutional neural networks (CNNs) to detect the CME signals in HIC systems. The architecture of the neural network is shown in Fig. 1. Details are in the Supplemental Material [57] with also a connection to the physics-informed neural network concept discussion [33,34]. For the purpose of supervised learning, we prepare training data from the string melting a multiphase transport (AMPT) model.<sup>1</sup> To implement the CME in the AMPT model, we adopt a global CS scheme first employed for Au + Au collisions in Ref. [44]: the  $y$ -components of momenta of a fraction of downward moving  $u$  quarks and upward moving  $\bar{u}$  quarks are switched, likewise for  $\bar{d}$  and  $d$  quarks.<sup>2</sup> This gives a

\*yczhao19@fudan.edu.cn

†lwang@fias.uni-frankfurt.de

‡Corresponding author: zhou@fias.uni-frankfurt.de

§Corresponding author: huangxuguang@fudan.edu.cn

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP<sup>3</sup>.

<sup>1</sup>The AMPT model [35] is a transport model which is widely used to simulate the evolution of both partonic and hadronic matter in HICs and has been proven to be successful in describing the experimental data of harmonic flows [36–39], global polarization [40–42], QCD phase transitions [43], and so on.

<sup>2</sup>Here, “upward” and “downward” refer to the  $y$ -axis which is perpendicular to the reaction plane.

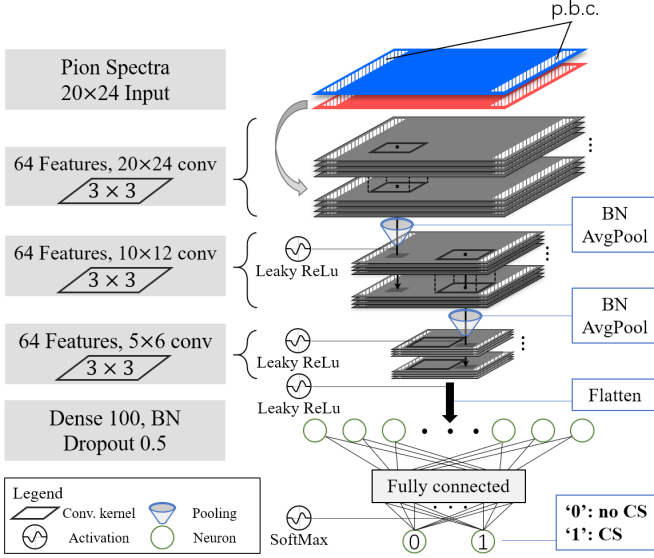


FIG. 1. The convolutional neural network architecture with  $\pi^+$  and  $\pi^-$  spectra  $\rho^\pm(p_T, \phi)$  as input.

right-dominant CME event, and for a left-dominant one we switch quarks with the momenta opposite to those in the right-dominant case, or just rotate the event along the  $z$ -axis (beam direction) by  $180^\circ$ . The CS fraction  $f$  is introduced as

$$f = \frac{N_{\uparrow(\downarrow)}^\pm - N_{\downarrow(\uparrow)}^\mp}{N_{\uparrow(\downarrow)}^\pm + N_{\downarrow(\uparrow)}^\mp}, \quad (1)$$

where the superscript labels the sign of charge and the subscript labels the direction of momentum along the  $y$ -axis. Events with  $f = 0\%$  belong to “no CS” class and are labeled as “0” while events with  $f > 0\%$  are in “CS” class and labeled as “1”.

With supervised learning, the deep CNNs are trained to distinguish the two classes from the labeled data. As for the input to the CNNs, we prepare from AMPT simulation series of two-dimensional (2-D) spectra  $\rho^\pm(p_T, \phi)$  of charged pions ( $\pi^+$  or  $\pi^-$ ) in the final state with 20 transverse momentum  $p_T$ -bins and 24 azimuthal angle  $\phi$ -bins.<sup>3</sup> The dataset consists of Au + Au collision events at  $\sqrt{s_{NN}} = 7.7, 11.5, 14.5, 19.6, 27, 39, 62.4,$  and 200 GeV, with CS fraction  $f = 0\%$  or  $f > 0\%$ , all divided into six centrality bins in the range 0 to 60%. Each species of collision conditions contains 50 000 events. To reduce fluctuations, 100 events with the same collision condition and dominant chirality are randomly picked out. Their pion spectra are averaged and normalized to form a single sample for the training. Meanwhile, such averaging also suppresses backgrounds and sets a prerequisite to the experimental feasibility of the model, which will be discussed in the following sections. To preserve the mirror symmetry, every sample is accompanied by its flipped copy along the

<sup>3</sup>We maintain the periodic boundary condition (*p.b.c.*) in  $\phi$  in all the convolution layers in the network, so as not to lose correlations near  $\phi = 0$  and  $2\pi$ , while  $p_T$  follows the original boundary condition of the convolution layer.

TABLE I. The validation accuracy of the well-trained model in confronting unseen events with the same set and different centralities.

Model	0–10%	10–20%	20–30%	30–40%	40–50%	50–60%
0% + 5%	79.6%	80.8%	79.2%	77.9%	77.8%	75.2%
0% + 10%	91.4%	94.1%	93.9%	94.4%	93.3%	92.8%

$y$ -axis. It can be viewed as exchanging the initial distribution of nucleons between the projectile and target nuclei, which naturally provides data augmentation and reduces redundancy in training the CNNs. To eliminate the ambiguity of introducing the CME under various conditions, we take two species of dataset with different CS fractions to train the deep CNNs. They are  $f = 5\%$  and  $10\%$  “CS” events mixed with the equivalent amount of “no CS” events. The corresponding well-trained CNN models are named as (0% + 5%) and (0% + 10%), respectively, in Table I, with their performance on recognizing the CS signal also shown. The validation accuracy of the (0% + 5%) model is less than the other one, which, however, could ameliorate overfitting [45] as demonstrated in the Supplemental Material [57]. It indicates that larger CS fraction is recognized better by the network, being manifested as more distinctiveness. In spite of the discrepancy between the two models, their performance is robust against various collision conditions, such as the  $\sqrt{s_{NN}}$  and centrality, see Table I for centrality dependency. The performance of the network also reflects that the CS signals are not totally diminished or contaminated after the collision dynamics and visible to our network-constructed CME-meter.

*Results and discussion- a. Extrapolations.* In this section, we investigate the generalization abilities of the trained NN and how it can serve as a *CME-meter*. The output of the network contains two components ( $P_0, P_1$ ) for each input spectrum with  $P_0 + P_1 = 1$ .  $P_1$  is identified as the probability that the network regards the input spectrum to be with CME. This will be an important indicator in the following. Besides the shown robustness against  $\sqrt{s_{NN}}$  and centrality, a positive correlation of the NN performance with the CS fraction  $f$  is also expected and demonstrated by Fig. 2. As  $f$  increases,

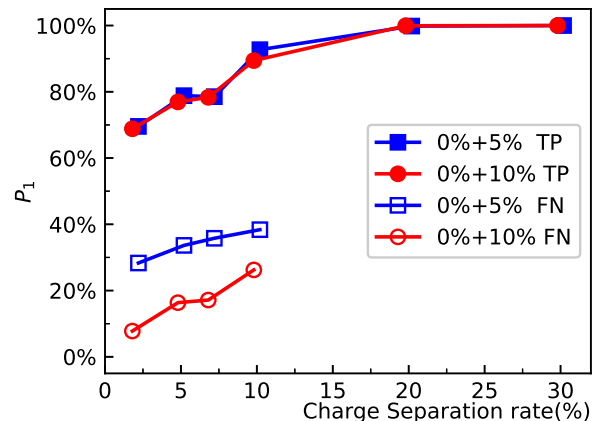


FIG. 2. The network prediction probability on datasets with mixed collision energies and centralities along with varying CS rate.

TABLE II. The results of the (0% + 10%) model on the isobaric collision systems (Ru + Ru and Zr + Zr at 200 GeV).

Centrality	0–10%	10–20%	20–30%	30–40%	40–50%	50–60%
$R_{\text{iso}}$	9.95%	12.99%	8.13%	13.84%	19.67%	10.47%
$\sigma_{R_{\text{iso}}}$	2.52%	2.05%	2.03%	1.87%	2.25%	2.64%

the  $P_1$  of the two NNs also increase, both for true-positive and false-negative cases. The reasonable extrapolation of  $P_1$  to different  $f$  indicates that the CME strength (or  $f$ ) is aligned to  $P_1$  by the NN, thus it is also related to the CS signal intensity.

An extrapolation to the other collision system is also realized. The networks are trained to recognize the CS signal in Au + Au collisions, whereas they are also qualified for  $^{96}_{40}\text{Zr} + ^{96}_{40}\text{Zr}$  and  $^{96}_{44}\text{Ru} + ^{96}_{44}\text{Ru}$ , the isobar collisions [46–50] recently proposed specifically for the CME search. Although being deformed for the nuclei which just has about half the mass of Au, the two systems well verify our trained NN, as shown in Table II. Physically, it is because four more protons in Ru may induce a larger magnetic field and thus cause a larger CS signal in Ru + Ru collisions. The quantity presented is

$$R_{\text{iso}} = 2 \times \frac{\langle \text{logit}(P_1^{\text{Ru}}) \rangle - \langle \text{logit}(P_1^{\text{Zr}}) \rangle}{\langle \text{logit}(P_1^{\text{Ru}}) \rangle + \langle \text{logit}(P_1^{\text{Zr}}) \rangle}, \quad (2)$$

and  $\sigma_{R_{\text{iso}}}$  is its error. The function  $\text{logit}(x) = \log[x/(1-x)]$  is used to restore the derivative at the saturation region of SOFTMAX, the last activation of our NN. The angle bracket means the average over samples.  $R_{\text{iso}}$  indicates a distinguishable difference between the two isobaric collision systems caused by  $P_1^{\text{Ru}} > P_1^{\text{Zr}}$ . The portability of NN to isobaric systems benefits from the normalization we carry out in training as to highlight the pion distribution instead of scale. This offers evidence for applicability of NN to a series of collision systems.

The generalization of the trained CNN is also validated on different model simulations. Events generated with

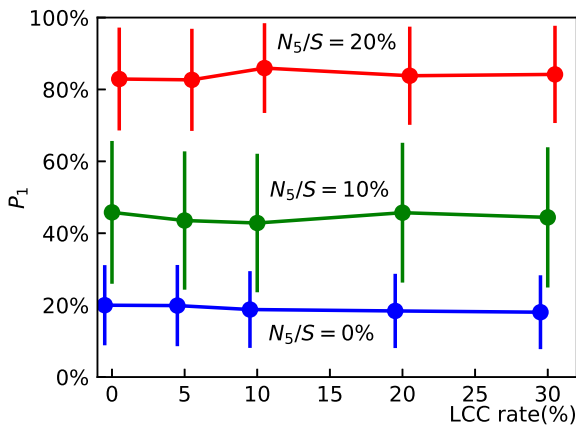


FIG. 3. Test on AVFD samples of (0% + 10%).  $N_5/S$  controls the CME strength, LCC rate is local charge conservation contamination.

anomalous-viscous fluid dynamics (AVFD)<sup>4</sup> are tested here, see Fig. 3.<sup>5</sup> As  $N_5/S$  increases,  $P_1$  rises just as concluded also from Table I, indicating the positive correlation between them. Moreover,  $P_1$  remains robust against the contaminations from LCC even up to 30%, where almost half of the final particles are from resonance decay. More details are in Supplemental Material [57].

*CME-meter.* In the above, we demonstrated the trained network can efficiently decode the CS information from  $\rho^\pm(p_T, \phi)$ . In this sense, the network provides a meter to measure the CME occurring probability in HICs. In the following, we investigate the comparison between  $P_1$  and the  $\gamma$ -correlator to reveal the coherent account of the *CME-meter*. The  $\gamma$ -correlator measures the event-by-event two-particle azimuthal correlation of charged hadrons, which is considered sensitive to CME [4]. It is defined as  $\gamma_{\text{same}} = \langle \cos(\phi_\alpha^{(\pm)} + \phi_\beta^{(\pm)} - 2\Phi_R) \rangle$  or  $\gamma_{\text{opp}} = \langle \cos(\phi_\alpha^{(\pm)} + \phi_\beta^{(\mp)} - 2\Phi_R) \rangle$  for correlation between the same or opposite charges, where  $\phi_\alpha^{(\pm)}$  is the azimuthal angle of particle  $\alpha$  with positive or negative charge,  $\Phi_R$  is the azimuthal angle of the reaction plane ( $\Phi_R = 0$  in this work), and  $\langle \cdot \rangle$  represents the average over particles in the event. To subtract charge-independent backgrounds one often uses  $\Delta\gamma = \gamma_{\text{same}} - \gamma_{\text{opp}}$ .

The histograms of  $P_1$  for “no CS” and “CS” cases are shown in Fig. 4(a). As a comparison, the histogram of  $\Delta\gamma$  on single events in Fig. 4(b) do not show a distinct difference between the two cases.  $\Delta\gamma$  on averaged events composed of 100 single events are also calculated (see the Supplemental Material [57]), showing around ten times smaller width than the single-event histogram for both cases and around five times smaller in the mean values. This means for averaged events, the signal from  $\Delta\gamma$  also suppresses, together with the backgrounds. In contrast,  $P_1$  can recognize the two cases, indicating that possible background contamination may mask the  $\gamma$ -correlator, but does not disturb the *CME-meter*. Such contamination is potentially proportional to the elliptic flow  $v_2$ . In fact, previous studies revealed that  $v_2$ -driven backgrounds can strongly interfere with the CME signal in the  $\gamma$ -correlator because both the magnetic field and  $v_2$  have similar centrality dependence [53,54]. Thus,  $v_2$ -induced  $\Delta\gamma$  can emerge in both “CS” and “no CS” events, making  $\Delta\gamma$  difficult to distinguish these two cases.

To further examine whether the *CME-meter* will be influenced by  $v_2$ , we depict  $P_1$  versus  $v_2/N$  ( $N$  is the multiplicity) in Fig. 5. It shows that CNN retains robust against the  $v_2$  background in identifying the CME signal. Specifically,  $P_1$  of “CS” events changes inconspicuously with  $v_2/N$  increasing, except for small  $v_2/N$ .

<sup>4</sup>Anomalous-viscous fluid dynamics (AVFD) implements the fluid dynamical evolution of chiral fermion currents on top of the neutral bulk background described by the VISH2 + 1 hydrodynamic simulations and plays as an active role in recent simulations of many anomalous effects in HICs [6,50–52].

<sup>5</sup>Samples are averaged among events from different hydrodynamic evolution, where  $\sqrt{s_{NN}} = 200$  GeV, centrality is mixed within 30–60%.

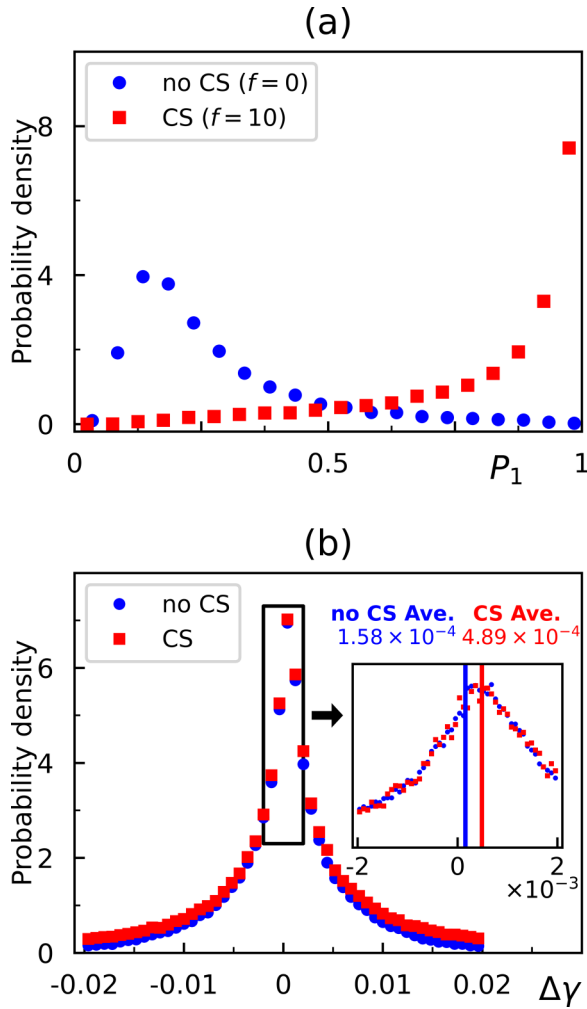


FIG. 4. (a)  $P_1$  histogram calculated by (0% + 10%) on averaged events for Au + Au  $\sqrt{s_{NN}} = 200$  GeV, centrality 40–50%. (b)  $\Delta\gamma$  histogram for single events. The average of  $\Delta\gamma$  are marked with lines in the zoom-in plot. For clarity some data points in (b) are moved.

*b. Hypothesis test.* Although events-averaging for the training set introduces background suppression manually, it provides a clear CS signal for constructing a sensitive *CME-meter*. In addition, the reaction plane of each event in averaging is aligned beforehand. To deploy the *CME-meter* into experimental measurements, one needs to reconstruct the reaction plane first, which can be reached through measuring final particle correlations and inevitably include finite resolution and backgrounds. A detailed discussion about how the event plane reconstruction influences  $P_1$  shows in the Supplemental Materials, where the trained meter can recognize CS signals under restricted reconstruction. With better reconstruction where events averaging is closer to our set-up,  $P_1$  is less affected by the background, and represents the CS signal more faithfully.

However, with a hypothesis test perspective, on single-event measurements the *CME-meter* also holds experimental feasibility. For a fixed finite number of events  $M$ , one can assume a large-enough “residual chirality,” which can be detected through our *CME-meter* if CS effects are believed

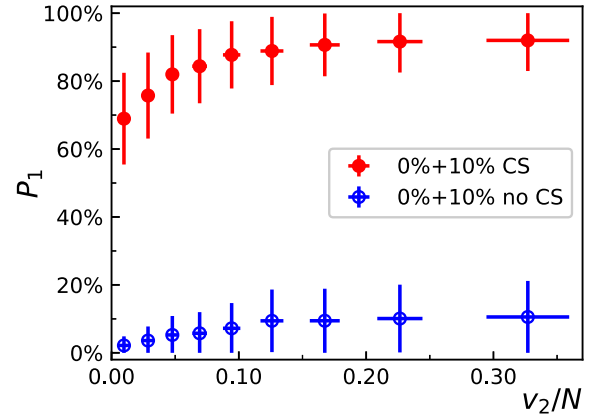


FIG. 5.  $P_1$  at different  $v_2/N$ , here  $N$  is the multiplicity at midrapidity (Au + Au at 200 GeV).

exiting in HICs. If *CME* exists in HICs, the residual chirality of an averaged event will behave like  $P(P_1)$  in the  $f \neq 0$  case, otherwise it will obey the distribution of the  $f = 0$  as shown in Fig. 4. To establish a reasonable estimation on  $P(P_1)$  for testing  $M$  events, we treat  $f$  as a latent variable representing *CME* in a single event, by the initial CS fraction we use in this work. Moreover, because of event-by-event fluctuation, we set  $f$  to be a random variable subject to a *Gaussian* distribution,  $f \sim N(\mu, \sigma^2)$ , where  $\mu$  is the mean of the latent variable  $f$  and shall be around 0. For the variance  $\sigma$ , we give an estimation based on Fig. 1 in the Supplementary Material(A) [57]. From  $\gamma$ -correlators, we can estimate the average of  $|f|$  around 8%,

$$8\% = \langle |f| \rangle = \int |f| N^>(|f|; \sigma^2) d|f|, \quad (3)$$

here  $N^>(|f|; \sigma^2)$  is the half normal distribution and  $|f|$  is a positive-definite variable because the  $\gamma$ -correlator is independent of the sign of  $f$ . Solving Eq. (3) gives  $\sigma \sim 0.2$ . Because we adopt averaged events in preparing the *CME-meter*, the way to compose a  $\rho(f_{\text{eff}})$  from single events  $\{\rho(f_i)\}$  becomes crucial, where  $f_{\text{eff}}$  is the residual chirality of averaged events. We can choose the arithmetic mean as

$$\frac{1}{M} \sum_i \rho(f_i) = \rho\left(\frac{1}{M} \sum_i f_i\right) = \rho(f_{\text{eff}}). \quad (4)$$

Therefore, the distribution of  $|f_{\text{eff}}|$  can be achieved as  $f_{\text{eff}} \sim N(\mu/M, M\sigma^2/M^2) = N(0, 0.04/M)$  with  $F_{\text{eff}} \equiv |f_{\text{eff}}| \sim N^>(0.04/M)$ . The conditional probability  $P(P_1 | F_{\text{eff}})$  can be approximated as a *Beta* distribution,  $\text{Be}(x; \alpha, \beta)$ .<sup>6</sup> After

<sup>6</sup>The *Beta* distribution is  $\text{Be}(x; \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}$ , with  $\alpha$  and  $\beta$  the parameters of the *Beta* distribution, and  $\Gamma$  is Gamma function. To describe  $P(P_1 | F_{\text{eff}})$  at any  $F_{\text{eff}}$ , we assume  $\alpha$  and  $\beta$  are functions of  $F_{\text{eff}}$ , and fit several sets of  $(\alpha, \beta)$  from the fitted beta distribution with polynomial (for  $\alpha$ ) and *Softplus* (for  $\beta$ , to reach proper asymptotic behavior around  $P_1 = 1$ ).

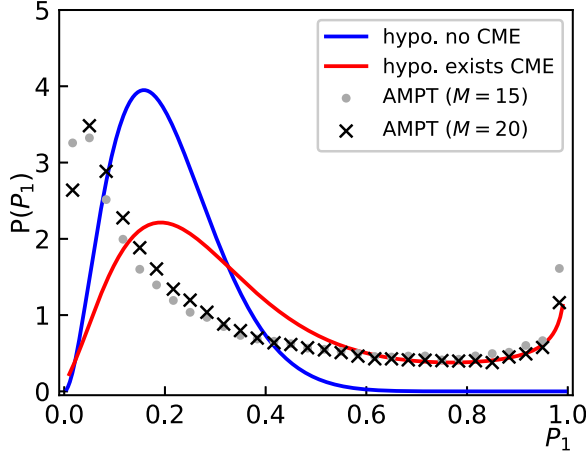


FIG. 6.  $P(P_1)$  of AMPT events from two different average numbers show as gray circles ( $M = 15$ ) and black crosses ( $M = 20$ ). The red and blue curves are from predictions of two different hypotheses.

parametrizing  $P(P_1 | F_{\text{eff}})$ ,  $P(P_1)$  is derived as

$$\begin{aligned} P(P_1) &= \int P(P_1 | F_{\text{eff}}) P(F_{\text{eff}}) dF_{\text{eff}} \\ &= \int_0^\infty \text{Be}(P_1; \alpha(F_{\text{eff}}), \beta(F_{\text{eff}})) N^> \left( \frac{0.04}{M} \right) dF_{\text{eff}}. \end{aligned} \quad (5)$$

It should be noticed that the integral variable is  $F_{\text{eff}}$  because  $NN$  responds the same to different signs of  $f_{\text{eff}}$ . The numerical results are shown in Fig. 6.  $P(P_1)$  of the “existing CME” has an obvious rise around  $P_1 = 1$  compared to the “no CME” case, which suggests a nonzero probability of composing a large residual chirality. With a smaller  $M$ , the width of  $f_{\text{eff}}$  becomes larger, which allows one to get a visible  $P_1$ . In Fig. 6 we also present results of randomly mixing left- and right-handed events generated by AMPT, where  $M = 20$  or  $15$ . From the large- $P_1$  area, it is consistent with our hypothesis test analysis, which indicates that the trained  $NN$  is capable of recognizing charge separation with less averaged events.

*c. Interpretable deep learning for CME.* The prediction  $P_1[\rho^\pm(p_T, \phi)]$  from the well-trained network could be understood as a CME-signal response to the spectrum  $\rho^\pm(p_T, \phi)$ , which can be utilized to find the most responsive features via the variational scrutinization

$$\frac{\delta P_1[\rho^\pm(p_T, \phi)]}{\delta \rho^\pm(p_T, \phi)} = 0. \quad (6)$$

Specifically, with the pion spectrum to be a variational *Ansatz*, we start from a flat spectrum  $\rho^\pm(p_T, \phi) = 1/X$  with  $X = 480$  the total number of pixels of the spectrum, which derives  $P_1 = 0$ , and gradually tune the functional form of  $\rho^\pm(p_T, \phi)$  with the variational target to maximize  $P_1[\rho^\pm(p_T, \phi)]$ , that is, to approach  $P_1 = 1$ . Note that the trained *CME-meter* network is fixed, through which gradient of its output with respect to its input,  $\delta P_1[\rho^\pm(p_T, \phi)] / \delta \rho^\pm(p_T, \phi)$ , can be evaluated via back propagation and is provided as the guidance for the above spectrum tuning. The resultant “ground state”

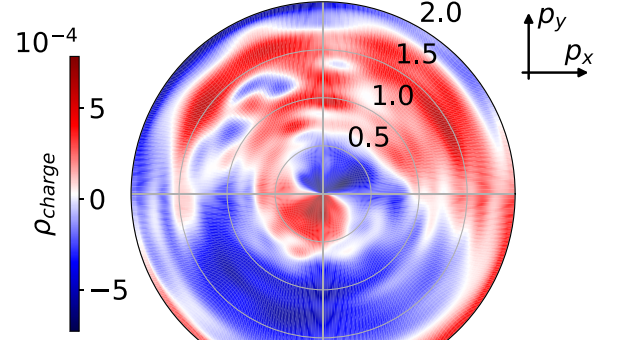


FIG. 7. *DeepDream* map for the (0% + 10%) model.

$\rho_0^\pm(p_T, \phi)$  could disclose the crucial patterns manifesting the CS signal in the perspective of the trained network. The above procedure is the so-called *DeepDream* method [55], in which the variational tuning is implemented as gradient ascent algorithms. In Fig. 7, the “ground state” pion spectrum tuned from the *DeepDream* method is visualized (also see Supplemental Material C [57]). Although this spectrum may neither be real nor physical, it shows the “CME pattern” that the network would response most dramatically. Similar visualization method was also demonstrated in image recognition tasks [56]. The  $\rho_0^\pm(p_T, \phi)$  explains the following basic features:

- (1) *Charge conservation.* During variation procedures, the charge conservation is reasonably reserved, from examination via tracking the charge density of the spectrum.
- (2) *Dipole structure.* The “ground state” spectrum from *DeepDream* variation intuitively displays the CS pattern. In the low  $p_T$  regime (center), the pion distribution induces an electric current, or a dipole downward, nevertheless it presents an opposite current which is larger in the high  $p_T$  regime ( $p_T \sim 3$  GeV).

It should be mentioned that  $\rho_0^\pm(p_T, \phi)$  derived from *DeepDream* is a virtual spectrum whose local properties depend on the AMPT simulation and the well-trained neural networks. However, it offers a reliable way to evaluate the effectiveness of  $P_1$  in detecting CME and helps us reveal the physical contents the machine learns.

*Summary.* In this study, we propose a deep convolutional neural network (CNN) model to detect the CME signal in the simulated dataset from a multiphase transport model. With two different charge separation fractions (5% and 10%), the machine is trained to recognize the CME signal under supervision. Besides accuracy and robustness against collision conditions, it is worth noting that this well-trained machine, encoded with the knowledge of CME in the final state of HICs, can extrapolate such a pattern to different charge separation fractions, collision systems, or even different model simulation (AVFD). The well-trained machine also provides a powerful meter to quantify the CME, remaining insensitive to the backgrounds dominated by the elliptic flow  $v_2$

and LCC in comparison with the conventional  $\gamma$ -correlator. Furthermore, a characteristic distribution of  $\max\{P_1(\phi)\}$  is given based on hypothesis test with reasonable reaction plane reconstructions, which can be a criterion of whether CME existing in HICs. In the end, *DeepDream*, a method used to visualize the patterns learned by CNNs, is applied as a validation test of adopting  $P_1$  to detect the CME. It helps us drill the physical knowledge hidden in the well-trained machine, including charge conservation and special charge distribution.

*Acknowledgments.* We acknowledge discussions with D. Kharzeev, R. Lacey, and X. N. Wang. The work is supported by NSFC through Grant No. 12075061 and Shanghai NSF through Grant No. 20ZR1404100 (Y.-S.Z. and X.-G.H.), by the AI grant at FIAS of SAMSON AG, Frankfurt (L.W. and K.Z.), by the BMBF under the ErUM-Data project (K.Z.), and by the NVIDIA Corporation with the generous donation of NVIDIA GPU cards for the research (K.Z.).

- 
- [1] D. Kharzeev, R. D. Pisarski, and M. H. G. Tytgat, *Phys. Rev. Lett.* **81**, 512 (1998).
- [2] D. E. Kharzeev, L. D. McLerran, and H. J. Warringa, *Nucl. Phys. A* **803**, 227 (2008).
- [3] K. Fukushima, D. E. Kharzeev, and H. J. Warringa, *Phys. Rev. D* **78**, 074033 (2008).
- [4] S. A. Voloshin, *Phys. Rev. C* **70**, 057901 (2004).
- [5] W.-T. Deng and X.-G. Huang, *Phys. Lett. B* **742**, 296 (2015).
- [6] N. Magdy, S. Shi, J. Liao, N. Ajitanand, and R. A. Lacey, *Phys. Rev. C* **97**, 061901(R) (2018).
- [7] J. Zhao, H. Li, and F. Wang, *Eur. Phys. J. C* **79**, 168 (2019).
- [8] A. H. Tang, *Chin. Phys. C* **44**, 054101 (2020).
- [9] A. Bzdak, V. Koch, and J. Liao, *Phys. Rev. C* **83**, 014905 (2011).
- [10] S. Schlichting and S. Pratt, *Phys. Rev. C* **83**, 014913 (2011).
- [11] F. Wang, *Phys. Rev. C* **81**, 064902 (2010).
- [12] J. Schmidhuber, *Neural Networks* **61**, 85 (2015).
- [13] Y. LeCun, Y. Bengio, and G. Hinton, *Nature (London)* **521**, 436 (2015).
- [14] L.-G. Pang, K. Zhou, N. Su, H. Petersen, H. Stöcker, and X.-N. Wang, *Nat. Commun.* **9**, 210 (2018).
- [15] K. Zhou, G. Endrődi, L.-G. Pang, and H. Stöcker, *Phys. Rev. D* **100**, 011501(R) (2019).
- [16] Z. Liu, W. Zhao, and H. Song, *Eur. Phys. J. C* **79**, 870 (2019).
- [17] H. Huang, B. Xiao, H. Xiong, Z. Wu, Y. Mu, and H. Song, *Nucl. Phys. A* **982**, 927 (2019).
- [18] Y.-L. Du, K. Zhou, J. Steinheimer, L.-G. Pang, A. Motornenko, H.-S. Zong, X.-N. Wang, and H. Stöcker, *Eur. Phys. J. C* **80**, 516 (2020).
- [19] M. Omana Kuttan, J. Steinheimer, K. Zhou, A. Redelbach, and H. Stoecker, *Phys. Lett. B* **811**, 135872 (2020).
- [20] L. Jiang, L. Wang, and K. Zhou, *Phys. Rev. D* **103**, 116023 (2021).
- [21] S. Shi, K. Zhou, J. Zhao, S. Mukherjee, and P. Zhuang, *Phys. Rev. D* **105**, 014017 (2022).
- [22] L. Wang, S. Shi, and K. Zhou, *Phys. Rev. D* **106**, L051502 (2022).
- [23] P. Baldi, P. Sadowski, and D. Whiteson, *Nat. Commun.* **5**, 4308 (2014).
- [24] P. Baldi, P. Sadowski, and D. Whiteson, *Phys. Rev. Lett.* **114**, 111801 (2015).
- [25] J. Barnard, E. N. Dawe, M. J. Dolan, and N. Rajcic, *Phys. Rev. D* **95**, 014018 (2017).
- [26] P. Broecker, J. Carrasquilla, R. G. Melko, and S. Trebst, *Sci. Rep.* **7**, 8823 (2017).
- [27] A. Radovic, M. Williams, D. Rousseau, M. Kagan, D. Bonacorsi, A. Himmel, A. Aurisano, K. Terao, and T. Wongjirad, *Nature (London)* **560**, 41 (2018).
- [28] J. Carrasquilla and R. G. Melko, *Nat. Phys.* **13**, 431 (2017).
- [29] E. P. L. van Nieuwenburg, Y.-H. Liu, and S. D. Huber, *Nat. Phys.* **13**, 435 (2017).
- [30] Z.-Y. Han, J. Wang, H. Fan, L. Wang, and P. Zhang, *Phys. Rev. X* **8**, 031012 (2018).
- [31] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, *Rev. Mod. Phys.* **91**, 045002 (2019).
- [32] L. Wang, Y. Jiang, L. He, and K. Zhou, [arXiv:2005.04857](https://arxiv.org/abs/2005.04857).
- [33] M. Raissi, P. Perdikaris, and G. E. Karniadakis, *J. Comput. Phys.* **378**, 686 (2019).
- [34] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, *Nat. Rev. Phys.* **3**, 422 (2021).
- [35] Z.-W. Lin, C. M. Ko, B.-A. Li, B. Zhang, and S. Pal, *Phys. Rev. C* **72**, 064901 (2005).
- [36] Z.-W. Lin, *Phys. Rev. C* **90**, 014904 (2014).
- [37] J. Xu and C. M. Ko, *Phys. Rev. C* **83**, 021903(R) (2011).
- [38] R. S. Bhalerao, J.-Y. Ollitrault, and S. Pal, *Phys. Rev. C* **88**, 024909 (2013).
- [39] D.-X. Wei, X.-G. Huang, and L. Yan, *Phys. Rev. C* **98**, 044908 (2018).
- [40] H. Li, L.-G. Pang, Q. Wang, and X.-L. Xia, *Phys. Rev. C* **96**, 054908 (2017).
- [41] Y. Sun and C. M. Ko, *Phys. Rev. C* **96**, 024906 (2017).
- [42] D.-X. Wei, W.-T. Deng, and X.-G. Huang, *Phys. Rev. C* **99**, 014905 (2019).
- [43] X. Jin, J. Chen, Z. Lin, G. Ma, Y. Ma, and S. Zhang, *Sci. China Phys. Mech. Astron.* **62**, 11012 (2019).
- [44] G.-L. Ma and B. Zhang, *Phys. Lett. B* **700**, 39 (2011).
- [45] P. Mehta, M. Bukov, C.-H. Wang, A. G. R. Day, C. Richardson, C. K. Fisher, and D. J. Schwab, *Phys. Rep.* **810**, 1 (2019).
- [46] S. A. Voloshin, *Phys. Rev. Lett.* **105**, 172301 (2010).
- [47] W.-T. Deng, X.-G. Huang, G.-L. Ma, and G. Wang, *Phys. Rev. C* **94**, 041901(R) (2016).
- [48] W.-T. Deng, X.-G. Huang, G.-L. Ma, and G. Wang, *Phys. Rev. C* **97**, 044901 (2018).
- [49] H.-J. Xu, X. Wang, H. Li, J. Zhao, Z.-W. Lin, C. Shen, and F. Wang, *Phys. Rev. Lett.* **121**, 022301 (2018).
- [50] S. Shi, H. Zhang, D. Hou, and J. Liao, *Phys. Rev. Lett.* **125**, 242301 (2020).
- [51] S. Shi, H. Zhang, D. Hou, and J. Liao, *Nucl. Phys. A* **982**, 539 (2019).
- [52] N. Magdy, S. Shi, J. Liao, P. Liu, and R. A. Lacey, *Phys. Rev. C* **98**, 061902(R) (2018).

- [53] X.-G. Huang, *Rep. Prog. Phys.* **79**, 076302 (2016).
- [54] D. E. Kharzeev, J. Liao, S. A. Voloshin, and G. Wang, *Prog. Part. Nucl. Phys.* **88**, 1 (2016).
- [55] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, New York, 2015), pp. 1–9.
- [56] L. Gatys, A. Ecker, and M. Bethge, *J. Vision* **16**, 326 (2016).
- [57] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevC.106.L051901> for more information on the neural network setup details; for more information on the validation and test accuracy demonstration; and for more information on extrapolation check.