EPJ Web of Conferences **364**, 02005 (2026) https://doi.org/10.1051/epjconf/202636402005
Quark Matter 2025

HEIDi: A deep generative framework for ultra-fast, event-by-event heavy-ion simulations

Manjunath Omana Kuttan[1,2,*], *Jan* Steinheimer[3,1], *Kai* Zhou[4,1], *Tom* Reichert[5], *Marcus* Bleicher[6,7], and *Horst* Stoecker[1,3,6]

[1]Frankfurt Institute for Advanced Studies, Ruth-Moufang-Str. 1, D-60438 Frankfurt am Main, Germany
[2]Xidian-FIAS International Joint Research Center, Giersch Science Center, D-60438 Frankfurt am Main, Germany
[3]GSI Helmholtzzentrum für Schwerionenforschung GmbH, Planckstr. 1, D-64291 Darmstadt, Germany
[4]School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen, P.R. China
[5]Department of Physics, Duke University, Durham, NC 27708-0305, USA
[6]Institut für Theoretische Physik, Goethe Universität Frankfurt, Max-von-Laue-Str. 1, D-60438 Frankfurt am Main, Germany
[7]Helmholtz Research Academy Hesse for FAIR (HFHF), GSI Helmholtz Center for Heavy Ion Physics, Campus Frankfurt, Max-von-Laue-Str. 12, 60438 Frankfurt, Germany

Abstract. *HEIDi*, a deep learning-based conditional diffusion model for ultra-fast generation of event-by-event heavy-ion collision output is introduced. Trained on UrQMD outputs, *HEIDi* is shown to generate point clouds of collision output particles, that accurately reproduce distributions of multiplicity and momentum across 26 different hadron species in UrQMD. Compared to UrQMD cascade simulations, *HEIDi* achieves a speedup factor of 100. These results demonstrate the potential of *HEIDi* as a versatile AI tool for both theoretical studies and experimental analyses.

1 Introduction

Relativistic heavy-ion collisions at the experimental facilities worldwide provide a unique opportunity to explore the properties of hot and/or dense nuclear matter in the laboratory. The experimental programmes of STAR-FXT and STAR-BES at RHIC, CBM at SIS-100, HADES at SIS-18, CEE at HIAF and MPD at NICA are of particular interest as they probe the least explored high density, moderate temperature region of the conjectured Quantum Chromodynamics (QCD) phase diagram. This is the region of the phase diagram where a possible first-order phase transition from hadronic to partonic matter as well as an associated critical end point, are most likely to occur.

At finite densities, first-principle lattice QCD calculations become intractable because of the fermionic sign problem. Therefore, effective model descriptions that closely simulate the dynamics and evolution of the system created in experiments are essential to study high density QCD matter created in heavy-ion collisions. An unknown density and momentum-dependent Equation of State (EoS) governs the evolution of the system created in collisions.

*e-mail: manjunath@fias.uni-frankfurt.de

The EoS is an essential input for both microscopic transport models with mean-field potentials [1, 2] and state of the art hybrid models with an intermediate hydrodynamic stage [3]. Non trivial-feaures of the phase diagram such as a phase transition can be implemented in these models by adjusting the parameters of the EoS. Conversely, identifying a phase transition or the critical point involves systematic comparisons of model predictions with experimental data to constrain the EoS.

Attempts to constrain the dense nuclear matter EoS with only a limited set of observables has shown that a more comprehensive analysis is necessaary for unambiguous results [4, 5]. Such a comprehensive Bayesian analysis would include a wide range of observables, such as multi-differential flow and multi-particle correlation spectra. However, the high computational expense of event-by-event models used to simulate moderate energy heavy-ion collisions severely limits the feasibility of a comprehensive Bayesian analysis.

To address this challenge, we present a generative AI solution that accelerates event-by-event heavy-ion collision simulations. Generative models are machine learning methods capable of learning the underlying distribution of a dataset and subsequently generating new samples from it. Deep learning methods [6, 7] including generative models [8–10] have successfully outperformed conventional approaches in a variety of tasks within high-energy nuclear physics. However, a complete event-by-event generative model capable of producing the full list of hadrons emitted in a heavy-ion collision event, including their four-momenta and hadron types, was introduced only recently in [11, 12]. Here, we provide a brief overview of this framework, named *HEIDi* (**H**eavy-ion **E**vents through **I**ntelligent **Di**ffusion), highlighting its novel features and potential applications.

2 HEIDi: Heavy-ion Events through Intelligent Diffusion

As proof of concept, *HEIDi* is trained to generate the collision output of the microscopic model UrQMD for central Au-Au collisions at 10 *A*GeV. From the choice of data structure to the design of model architecture, the framework has been developed to be flexible enough to be readily adapted to emulate other physics models or detector-response simulations, and to be extended to incorporate additional physics scenarios.

As a natural representation of both detector outputs and event-by-event collision output data, *HEIDi* uses a point cloud structure to represent collision events. Each event is a list of particle vectors where each vector corresponds to a final state particle. A particle vector contains the three components of momentum (p_x, p_y and p_z) and its particle species ID in one-hot encoded format. Thus, an event can be represented as a 2-D array where each row is a particle and each column corresponds to an attribute of the particle. The number of rows is fixed at 1084, which is larger than the maximum event multiplicity in the training data. Events containing fewer particles are zero-padded to maintain this fixed size. The array has 32 columns. the first three store the momentum components, while the remaining 29 encode particle species information of 26 distinct hadron species, spectator protons and neutrons, as well as dummy particles introduced to preserve the fixed dimensionality of the point cloud.

HEIDi employs a conditional probabilistic diffusion model based on [13], to generate point clouds of collision event outputs. Diffusion models learn by progressively adding controlled random noise to the data until the original event becomes indistinguishable from samples drawn from a simple distribution, such as a Gaussian. The model then learns to predict the noise added at each step. Once trained, the setup allows us to generate new samples by drawing from the simplified distribution and iteratively removing the predicted noise to recover a point cloud that represents the collision event output. Additionally, the diffusion model is conditioned on a latent vector generated by a normalizing flow model. This flow model learns to construct latent representations that encode various correlations within an event , which are necessary for generating realistic events.

The model was trained on 18000 collision events and its performance was evaluated on 50000 events. A detailed description of the network structure, the mathematical background, and the evaluation tests can be found in [11, 12].

3 Results

After training, the collision outputs generated by *HEIDi* were compared with UrQMD events to assess the quality of generation. The model successfully produced realistic, "UrQMD-like" collision outputs. The mean event multiplicity for various hadron types as well as their multiplicity distributions showed very good agreement with UrQMD results. Moreover, the momentum distributions of particles in *HEIDi* generated events accurately reproduce those from UrQMD. The transverse momentum distributions of selected hadrons from *HEIDi* and UrQMD events are compared in Figure 1. Remarkably, using a relatively small training set of just 18000 events, the model is able to learn the probabilities over five orders in magnitude. With the exception of deviations at very small p_T, *HEIDi* demonstrates excellent agreement with UrQMD. A similar deviation is observed in the rapidity distributions of these hadrons at rapidities close to zero. This effect can be traced to an excess of very low-momentum particles in the *HEIDi* output, which in turn causes the discrepancies observed in both the p_T and rapidity spectra.

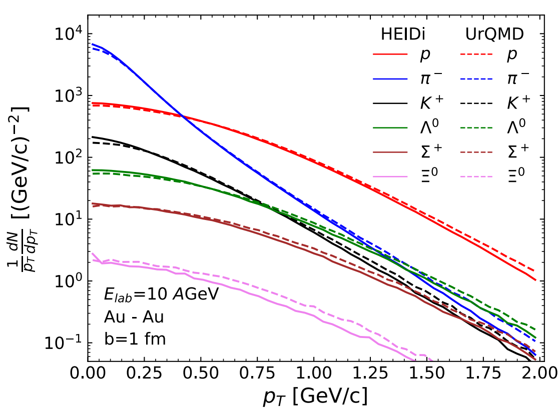

Figure 1. Transverse momentum distributions of selected hadrons for Au-Au collisions at 10 *A*GeV with impact parameter b=1 fm. *HEIDi* results are shown as solid lines while UrQMD results are shown as dotted lines. Apart from the deviation at very low p_T values, *HEIDi* successfully reproduces the UrQMD p_T distributions for different hadrons species.

While it is important to identify the source of the deviation at small p_T values, it only leads to a small difference to the total yield. Moreover, for experimental studies, a low momentum cutoff is usually applied to simulation data as detectors often have limited acceptance and efficiency for low-momentum particles.

HEIDi accelerates event-by-event heavy-ion simulations by exploiting the massive parallelization capabilities of Graphics Processing Units (GPU). On a single CPU core, the runtimes of *HEIDi* and UrQMD cascade model are comparable, at about 3 seconds per event.

However, on an NVIDIA A100 GPU with 40 GB of memory, *HEIDi* achieves a runtime of roughly 30 milliseconds per event, corresponding to a speedup by a factor of 100, and is far more energy efficient than running the same workload on 100 CPU cores in parallel. For more sophisticated models such as the hybrid version of UrQMD, which includes an intermediate hydrodynamic stage, one can expect speedups of at least five orders of magnitude on GPUs and about three orders of magnitude even on a CPU.

4 Outlook

As a deep generative framework capable of ultrafast event-by-event collision output generation, *HEIDi* represents an important step toward developing the first foundation model for heavy-ion collisions. As *HEIDi* generates output in point cloud format, it can be readily adapted to accelerate any other event-by-event model or detector simulations. This flexibility and generation speed make *HEIDi* an excellent tool capable of calculating any observable to perform a comprehensive Bayesian inference of the underlying physics such as the EoS. As a differentiable surrogate of a physics model, *HEIDi* also supports novel gradient based approaches for various parameter estimation tasks. Furthermore, the latent condition vector can be extended to include different collision parameters such as beam energy, impact parameter or the EoS itself as inputs. With such extensions and with careful treatment of the observed deviations from the UrQMD outputs, *HEIDi* has the potential to become an inevitable tool for heavy-ion physics with broad applications in both theoretical studies and experimental analyses.

This work is supported by the the BMBF under the KISS project (M.O.K, K.Z), the CUHK-Shenzhen university development fund under grant No. UDF01003041 and UDF03003041 (K.Z), and Shenzhen Peacock fund under No. 2023TC0179 (K.Z) and the Walter Greiner Gesellschaft zur Förderung der physikalischen Grundlagenforschung e.V. through the Judah M. Eisenberg Laureatus Chair at the Goethe Universität Frankfurt am Main (H.S). T.R. gratefully acknowledges support from The Branco Weiss Fellowship - Society in Science, administered by the ETH Zürich.

References

[1] M. Omana Kuttan, A. Motornenko, J. Steinheimer, H. Stoecker, Y. Nara, M. Bleicher, A chiral mean-field equation-of-state in UrQMD: effects on the heavy ion compression stage, Eur. Phys. J. C **82**, 427 (2022), 2201.01622. 10.1140/epjc/s10052-022-10400-2

[2] J. Steinheimer, T. Reichert, Y. Nara, M. Bleicher, Momentum dependent potentials from a parity doubling CMF model in UrQMD: results on flow and particle production, J. Phys. G **52**, 035103 (2025), 2410.01742. 10.1088/1361-6471/adab0b

[3] H. Petersen, J. Steinheimer, G. Burau, M. Bleicher, H. Stöcker, A Fully Integrated Transport Approach to Heavy Ion Reactions with an Intermediate Hydrodynamic Stage, Phys. Rev. C **78**, 044901 (2008), 0806.1695. 10.1103/PhysRevC.78.044901

[4] M. Omana Kuttan, J. Steinheimer, K. Zhou, H. Stoecker, QCD Equation of State of Dense Nuclear Matter from a Bayesian Analysis of Heavy-Ion Collision Data, Phys. Rev. Lett. **131**, 202303 (2023), 2211.11670. 10.1103/PhysRevLett.131.202303

[5] D. Oliinychenko, A. Sorensen, V. Koch, L. McLerran, Sensitivity of Au+Au collisions to the symmetric nuclear matter equation of state at 2–5 nuclear saturation densities, Phys. Rev. C **108**, 034908 (2023), 2208.11996. 10.1103/PhysRevC.108.034908

[6] M. Omana Kuttan, J. Steinheimer, K. Zhou, A. Redelbach, H. Stoecker, A fast centrality-meter for heavy-ion collisions at the CBM experiment, Phys. Lett. B **811**, 135872 (2020), 2009.01584. 10.1016/j.physletb.2020.135872

[7] M. Omana Kuttan, K. Zhou, J. Steinheimer, A. Redelbach, H. Stoecker, An equation-of-state-meter for CBM using PointNet, JHEP **21**, 184 (2020), `2107.05590`. 10.1007/JHEP10(2021)184

[8] J.A. Sun, L. Yan, C. Gale, S. Jeon, An end-to-end generative diffusion model for heavy-ion collisions, arXiv (2024), `2410.13069`.

[9] D. Torbunov, Y. Huang, M. Lin, Y. Ren, Y. Go, T. Rinn, H. Yu, B. Viren, J. Huang, Effectiveness of denoising diffusion probabilistic models for fast and high-fidelity whole-event simulation in high-energy heavy-ion experiments, Phys. Rev. C **110**, 034912 (2024), `2406.01602`. 10.1103/PhysRevC.110.034912

[10] H. Huang, B. Xiao, Z. Liu, Z. Wu, Y. Mu, H. Song, Applications of deep learning to relativistic hydrodynamics, Phys. Rev. Res. **3**, 023256 (2021), `1801.03334`. 10.1103/PhysRevResearch.3.023256

[11] M. Omana Kuttan, K. Zhou, J. Steinheimer, H. Stöcker, Ultra fast, event-by-event heavy-ion simulations for next generation experiments, arXiv (2025), `2502.16330`.

[12] M. Omana Kuttan, K. Zhou, J. Steinheimer, H. Stöcker, Towards a foundation model for heavy-ion collision experiments through point cloud diffusion, arXiv (2024), `2412.10352`.

[13] S. Luo, W. Hu, Diffusion Probabilistic Models for 3D Point Cloud Generation, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2021), pp. 2837–2845