Neural network enhanced Bayesian analysis of heavy-ion observables

Henry Hirvonen

Vanderbilt University

Precision Frontier of QCD Matter: Inference and Uncertainty Quantification, Wuhan

9th of September, 2025

H. Hirvonen, K. J. Eskola and H. Niemi, Phys. Rev. C 108, 034905 (2023)
H. Hirvonen, K. J. Eskola and H. Niemi, EPJ Web Conf. 296, 02002
J. Auvinen, K. J. Eskola, H. Hirvonen, and H. Niemi, work in progress

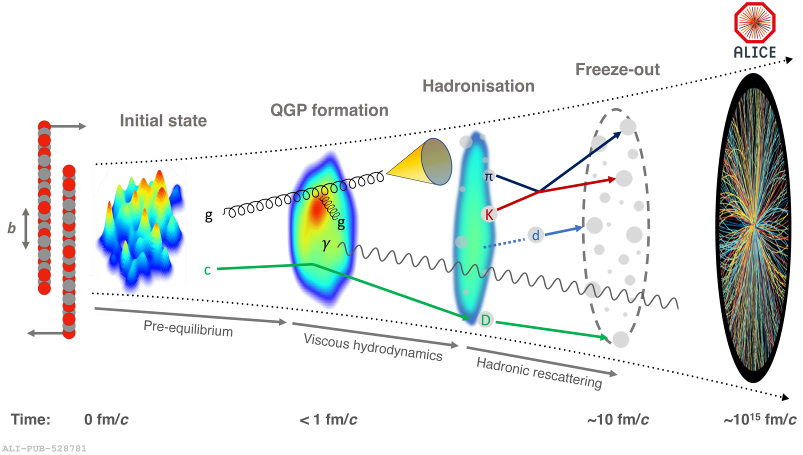
Motivation

- The goal of the heavy ion collisions is to study the matter properties of the quark-gluon plasma (QGP)
- Shear and bulk viscosities of QGP can be constrained from the measured data by means of Bayesian analysis
- A lot of Bayesian analyses have been done in recent years

  J. E. Bernhard, et al. Nature Phys. 15 1113–1117 (2019)
  J. S. Moreland, et al. Phys. Rev. C 101 024911 (2020)
  G. Nijs, et al. Phys. Rev. C 103 054909 (2021)
  M. R. Heffernan, et al. Phys.Rev.C 109 6, 065207 (2024)
  J. E. Parkkila et al. Phys.Rev.C 111 4, 044903 (2025)
  and many more
- Gaussian process emulators are used to reduce this computation cost significantly
  - With the emulators typical cost $\sim$1-50 million CPU hours
  - Can this be reduced further?
- Related work:

  J. Sun, et al. arXiv:2410.13069 [nucl-th] (2024)
  M. O. Kuttan, et al. arXiv:2502.16330 [hep-ph] (2025)

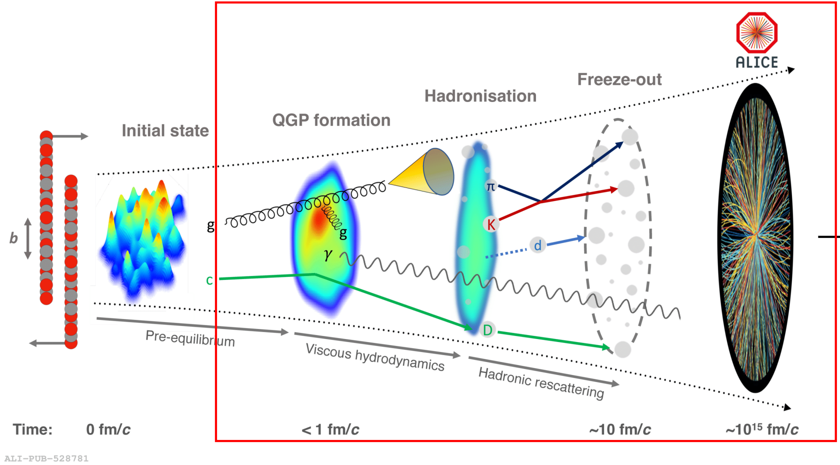
Modeling relativistic heavy-ion collisions



⟶ Observable,
e.g. v_n

ALICE collaboration, Eur. Phys. J. C84, 813

Modeling relativistic heavy-ion collisions



⟶ Observable, e.g. v_n

Computationally expensive part $\sim$ 30 min/event

$\Longrightarrow$ Replace with neural network

ALICE collaboration, Eur. Phys. J. C84, 813

The Model

- The initial energy density from pQCD+saturation based EbyE-EKRT

  H. Niemi et al. Phys.Rev.C 93 (2016) 2, 024907

- Boost-invariant Israel-Stewart viscous fluid dynamics with shear and bulk viscosities
- Equation of state: The s95p parametrization of the QCD equation of state with a chemical freeze-out at temperature $T_{\rm chem}$
- The fluid is converted into the particle spectrum by calculating the Cooper-Frye integral at the decoupling surface determined by dynamical freeze-out conditions
  - Decays of unstable particles are taken into account

Dynamical freeze-out

- Fluid dynamics applicable when expansion rate $(\theta) \lesssim$ scattering rate $(1/\tau_\pi)$ and mean free path $(\tau_\pi) \lesssim$ size of the system (R)
  $\Longrightarrow$ Dynamical decoupling conditions:

Knudsen number

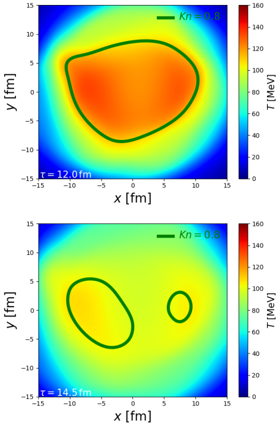
$$\mathrm{Kn} \equiv \frac{\text{exp. rate}}{\text{scat. rate}} = \tau_\pi \theta = C_{\mathrm{Kn}}$$

Global size of the system

$$\frac{\gamma \tau_\pi}{R} = C_R, \quad R = \sqrt{A/\pi}$$

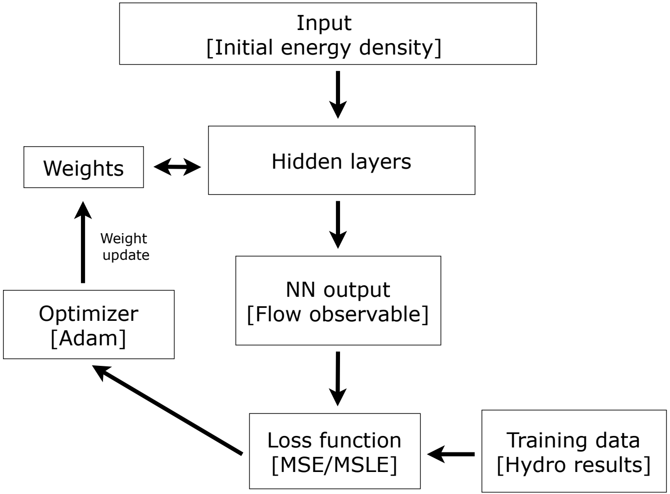
- C_{Kn} and C_R are free parameters, fitted from data
- A is the area in which $\mathrm{Kn} < C_{Kn}$ and T $<$ 150 MeV
- Allow multiple separate areas with different R

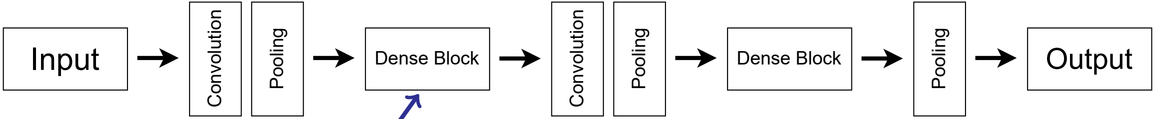
H. Hirvonen et al. Phys.Rev.C 106, 044913 (2022)

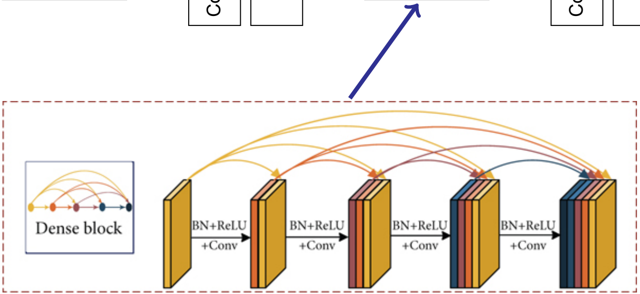


Neural Network for energy density

- Use NN to predict midrapidity flow observables from initial $e(x, y)$ Event-by-Event
- The weights in the hidden layers are trained to minimize the loss function
- Training data is for one fixed viscosity parametrization

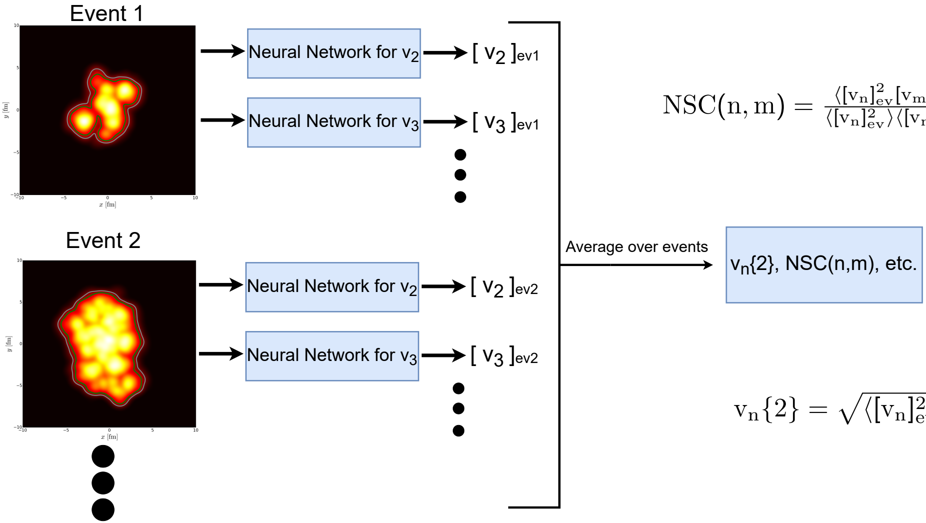






- Layers structure is implemented as a modified version of DenseNet
  G. Huang et al. arXiv:1608.06993
  - Very deep network structure with 128 convolutional layers
  - Each layer gets additional input from all previous layers to prevent feature loss and vanishing gradients
  - In total of $\approx$ 5.4M trainable parameters

Workflow



$$\mathrm{NSC(n,m)} = \frac{\langle [v_n]^2_{ev} [v_m]^2_{ev} \rangle}{\langle [v_n]^2_{ev} \rangle \langle [v_m]^2_{ev} \rangle} - 1$$

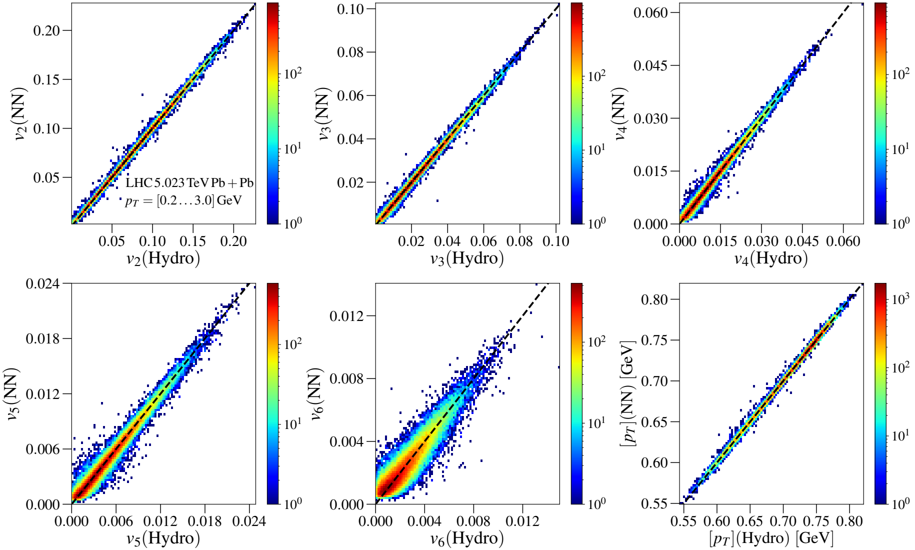
$$v_n\{2\} = \sqrt{\langle [v_n]^2_{ev} \rangle}$$

Training - Fixed model parameters

- Separate network trained for each p_T-integrated observable: $v_2, v_3, v_4, v_5, v_6, [p_T], dN_{ch}/d\eta$
- Each network trained with multiple different p_T ranges for an observable
- In total of 20000 training events: 5000 from each collision system
    - 200 GeV Au+Au
    - 2.76 TeV Pb+Pb
    - 5.023 TeV Pb+Pb
    - 5.44 TeV Xe+Xe (deformed nuclei)
- Training data augmented using random flips, rotations and translations
- Training one network takes ≈ 1 GPU hour with NVIDIA V100 32GB GPU
- After training, NN can generate $\sim 10^6$ events/hour with GPU
    - Factor of 10^5 faster than doing full simulations using CPU!

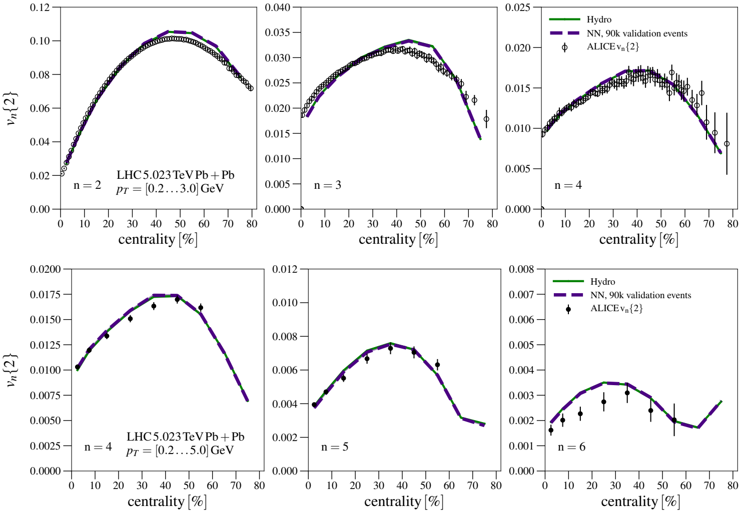
Validation tests: Errors with 90k validation events

H. Hirvonen et al. Phys. Rev. C 108, 034905 (2023)



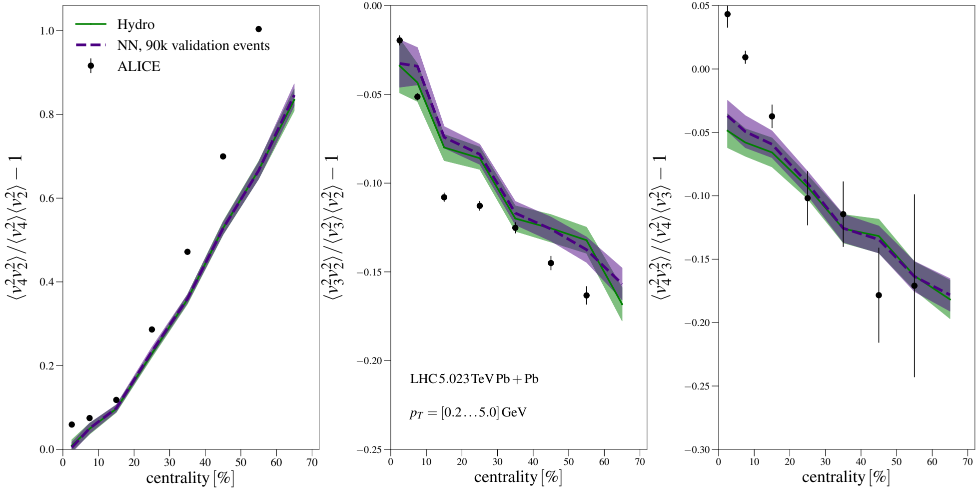
Validation tests: Flow coefficients

H. Hirvonen et al. Phys. Rev. C 108, 034905 (2023)



Validation tests: Four-particle flow correlations

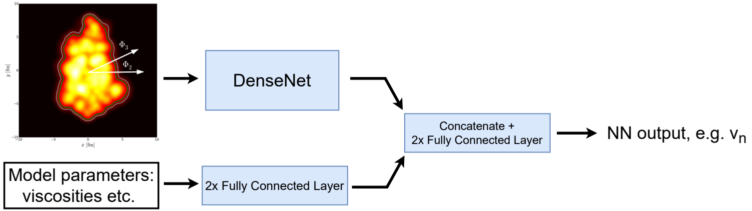
H. Hirvonen et al. Phys. Rev. C 108, 034905 (2023)



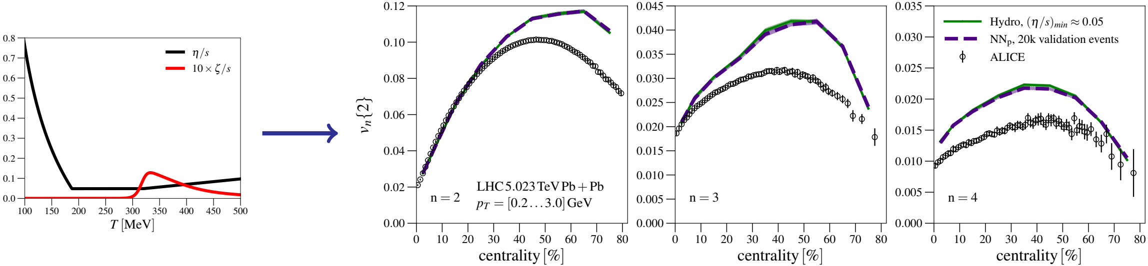
Model parameters as an input ($\mathrm{NN_p}$)

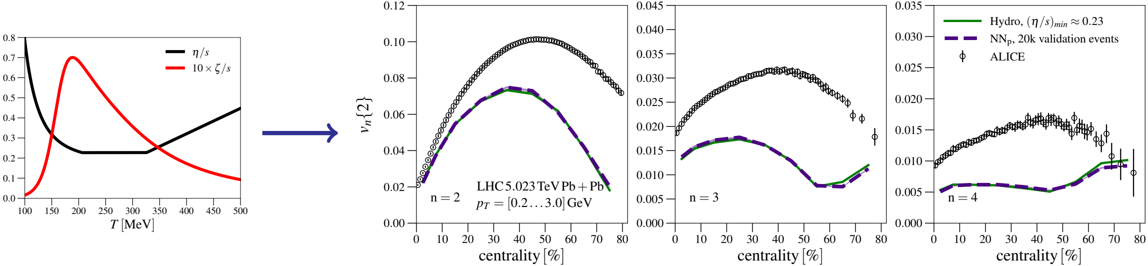
- Neural network can be extended to take model parameters as additional input:
  - 6 parameters describing $\eta/s(T)$
  - 4 parameters describing $\zeta/s(T)$
  - 3 parameters describing chemical and kinetic decoupling
- Training data consists of 160000 events
  - 1000 different parameter points sampled from Latin hypercube
  - 4 collision systems

$\Longrightarrow$ Very efficient: only 40 training events/parameter for each collision system

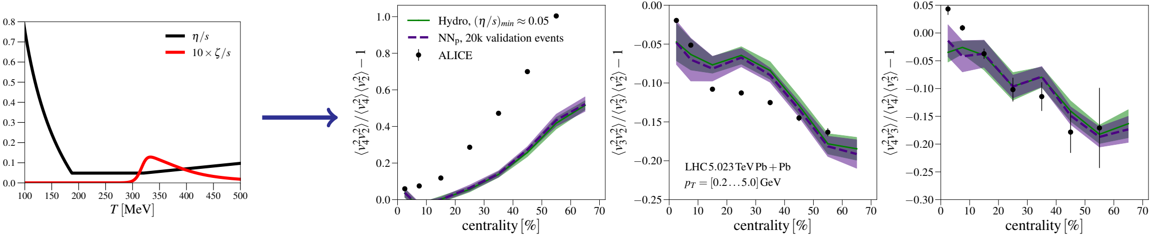


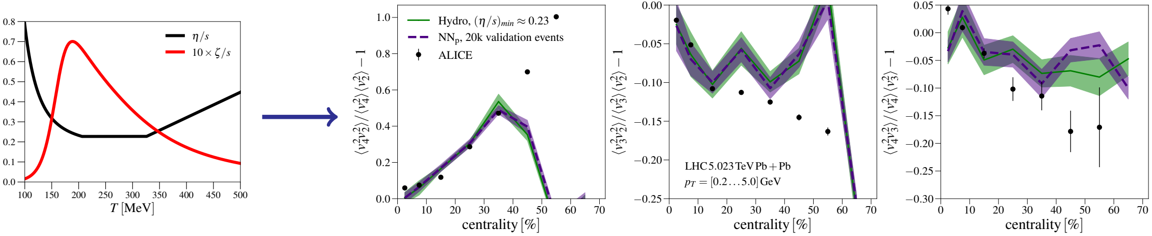
Validation with additional inputs: Flow coefficients





Validation with additional inputs: Normalized symmetric cumulants





Bayesian analysis

- Goal: Determine the model parameters $\vec{x}$ by comparing model output $\vec{y}$ against measurements $\vec{y}^{\mathrm{exp}}$ according Bayes' theorem:

$$P(\vec{x}|\vec{y}^{\mathrm{exp}}) \propto \mathcal{L}(\vec{y}^{\mathrm{exp}}|\vec{x})p(\vec{x}),$$

where prior $p(\vec{x})$ is chosen uniform and likelihood function is given by:

$$\mathcal{L}(\vec{y}^{\mathrm{exp}}|\vec{x}) = \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left(-\frac{1}{2}\left(\vec{y}\left(\vec{x}\right) - \vec{y}^{\mathrm{exp}}\right)^T \Sigma^{-1} \left(\vec{y}\left(\vec{x}\right) - \vec{y}^{\mathrm{exp}}\right)\right)$$

and Σ is the covariance matrix containing emulation, theoretical, and experimental uncertainties

- Posterior probability distribution sampled using Markov Chain Monte Carlo (MCMC)

The Input

Initial state: 2 parameters, K_{sat}, and σ
Shear Viscosity: 6 parameters, $(\eta/s)_{min}$, T_H, S_H, P_H, S_Q

$$\eta/s(T) = \begin{cases} (\eta/s)_{min} + S_H T\left(\frac{T}{T_H}\right)^{-P_H}, & T < T_H \\ (\eta/s)_{min}, & T_H \leq T \leq T_H + W_{\mathrm{min}} \\ (\eta/s)_{min} + S_Q(T - T_Q), & T > T_H + W_{\mathrm{min}}, \end{cases}$$

Bulk Viscosity: 4 parameters, $(\zeta/s)_{max}$, $T_{\zeta\mathrm{max}}$, $a_{\zeta/s}$, $(\zeta/s)_{width}$

$$\zeta/s(T) = \frac{(\zeta/s)_{max}}{1 + \left(\frac{T - T_{\zeta\mathrm{max}}}{w(T)}\right)^2}, w(T) = \frac{2(\zeta/s)_{width}}{1 + \exp\left(\frac{a_{\zeta/s}(T - T^{\zeta/s}_{max})}{(\zeta/s)_{width}}\right)},$$

Decoupling: 3 parameters T_{chem}, C_{kn}, and C_{R}

The Output

Four Collision systems:

Au+Au at 200 GeV, Pb+Pb at 2.76 TeV, Pb+Pb at 5.02 TeV, and Xe+Xe at 5.44 TeV

For each collision system, we include the following observables:

- Charged particle multiplicity $dN_{\mathrm{ch}}/d\eta$
- p_T integrated 2-particle flow coefficients $v_n\{2\}, n = 2, 3, 4$

For Au+Au and Pb+Pb systems, we also include:

- Identified particle multiplicities $dN/d\mathrm{y}$ of pions, kaons, and protons
- Average transverse momenta $\langle p_T \rangle$ of pions, kaons, and protons
- Normalized symmetric cumulant $\mathrm{NSC}(2,4) = \frac{\langle v_2^2 v_4^2 \rangle - \langle v_2^2 \rangle \langle v_4^2 \rangle}{\langle v_2^2 \rangle \langle v_4^2 \rangle}$

For the Xe+Xe system, we also include:

- Charged particle transverse momentum $\langle p_T \rangle_{\mathrm{ch}}$

Emulation

- Unfortunately, neural networks alone are still too slow to be used directly in MCMC
- Instead, use neural networks to generate training data for Gaussian process emulators, which can then be used in MCMC

Neural networks

Reduces needed simulations in one parameter point

GP emulator

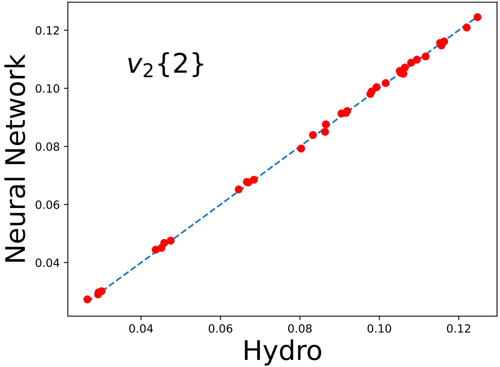
Reduces the number of parameter points that need to be simulated

- 40k events sampled with a neural network in 1000 parameter points for 4 collision systems
  - Computing time without neural networks $\sim 10^8$ CPU hours
  - Computing time with neural networks: $\approx 8 \times 10^4$ CPU hours (training data) + 320 GPU hours
- Neural networks do not provide uncertainty estimates, so only Gaussian process uncertainties are taken into account in the analysis

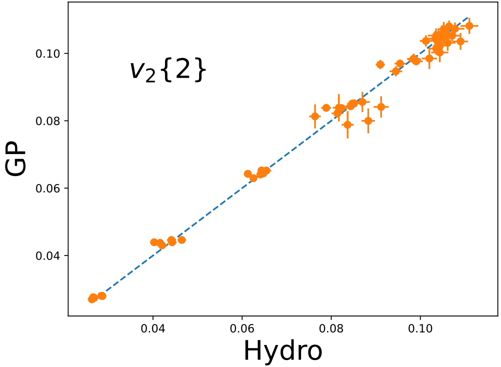
Emulation uncertainties

Compare emulators against simulations with different parameter values

Neural Network:



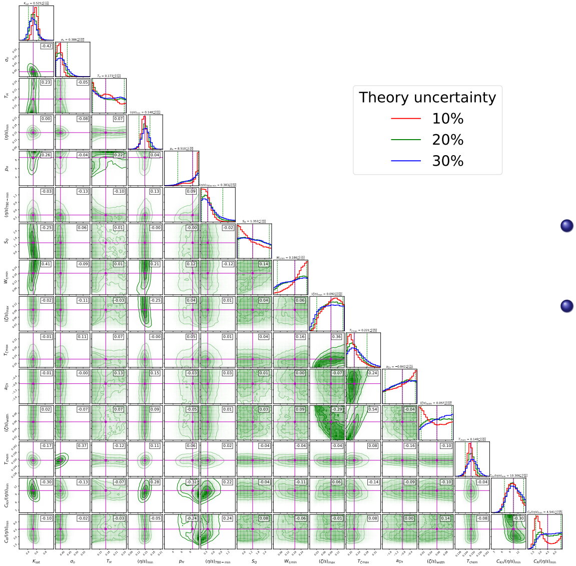
Gaussian Process:



Choose parameters from the training samples close to the MAP value

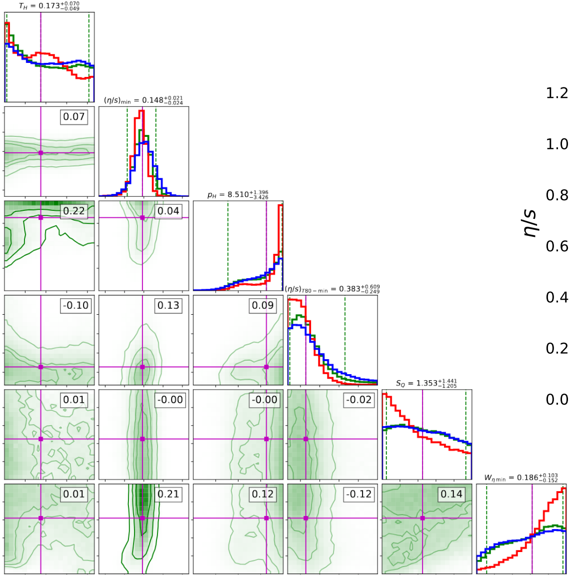
Sample parameters from the posterior distribution

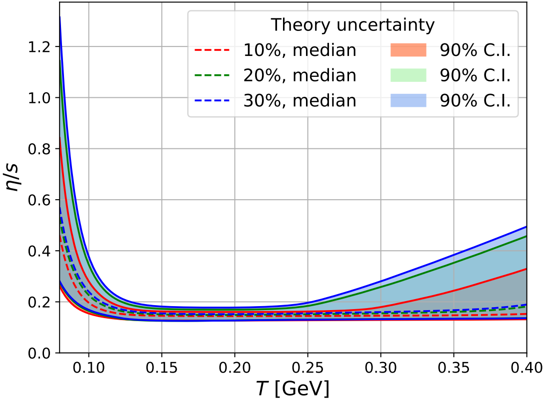
Posterior distribution



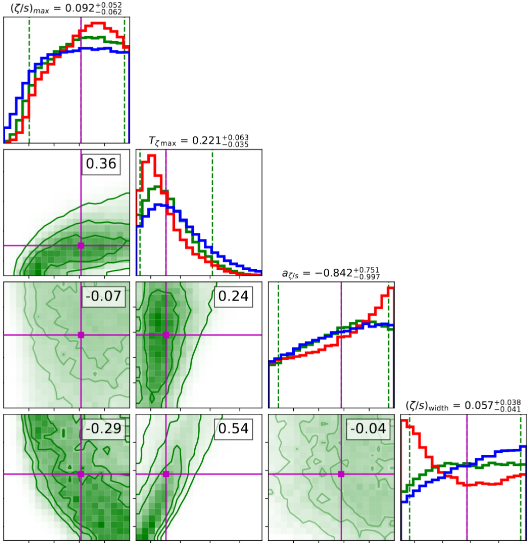
- Theoretical uncertainty is varied between 10-30%
- Results are quite robust for this variation

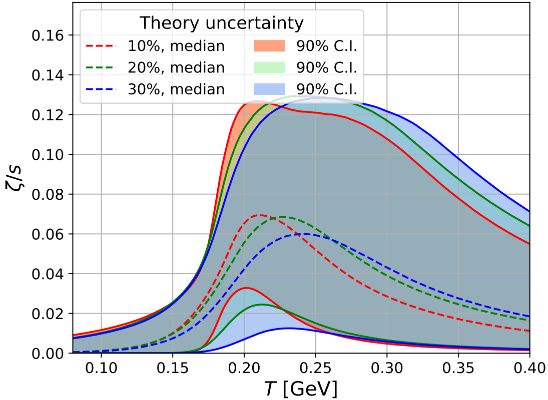
Shear viscositiy



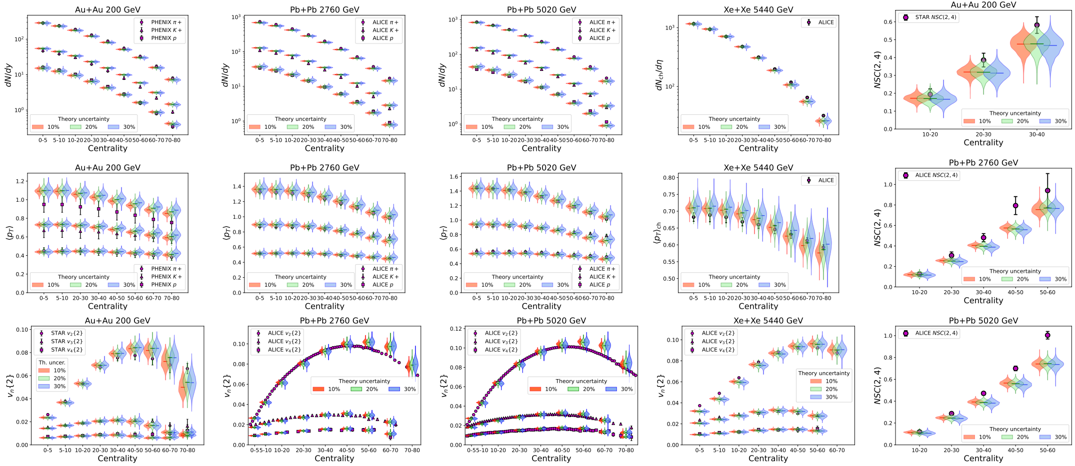


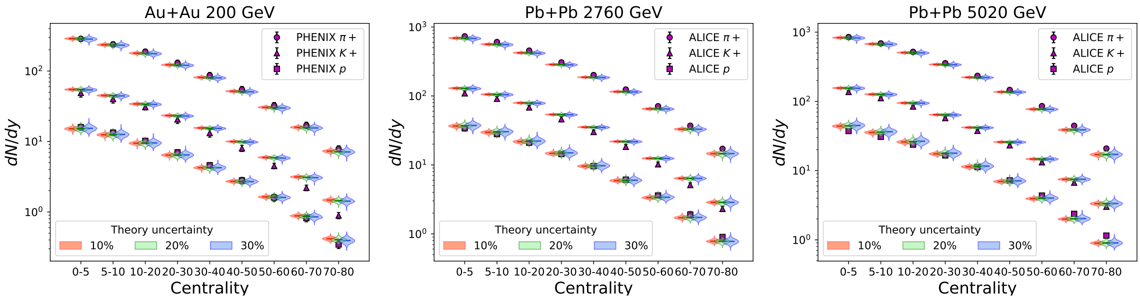
Bulk viscositiy





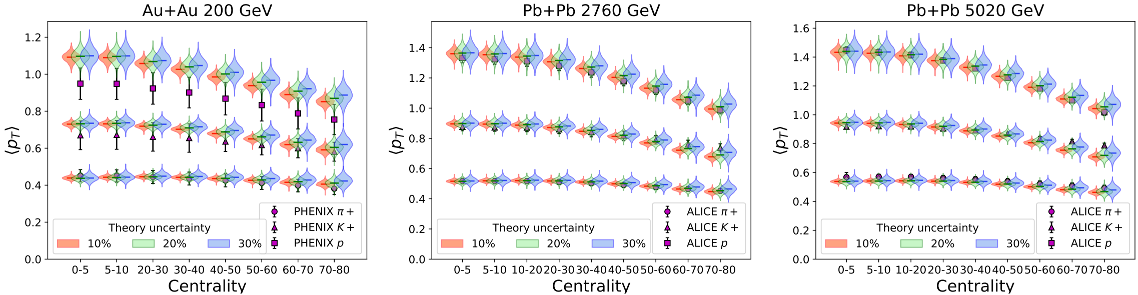
Observables





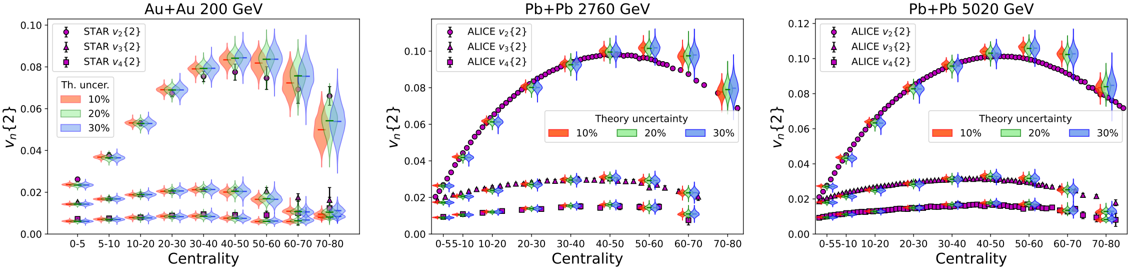
- Overall good agreement, but slight systematic overestimation of kaon multiplicities

Mean transverse momenta



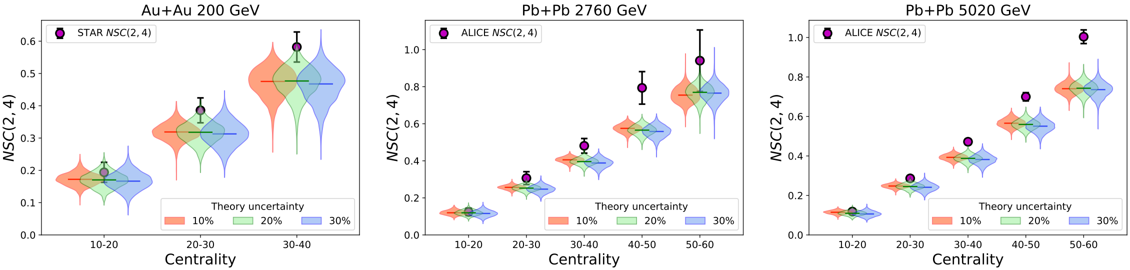
- Increasing theoretical uncertainty makes distributions wider, but the mean values are nearly unchanged

Flow coefficients



- Some deviations from the measurements in the peripheral collisions, but otherwise quite good agreement with the data

Normalized symmetric cumulants



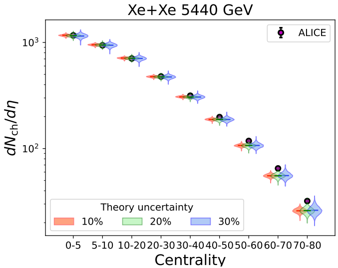
- Clear underestimation of correlations for non-central collisions
- Distributions are quite wide due to the statistical fluctuations
  - More simulated events needed for more precise analysis

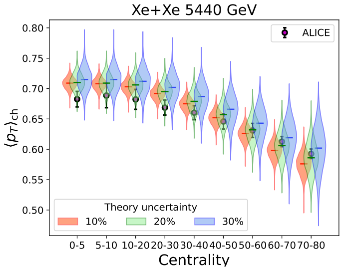
Outlook

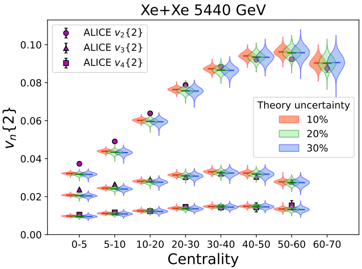
- Utilization of neural networks reduces computation time by many orders of magnitude
  - Maintains good accuracy for a variety of flow observables
  - Can be extended to take model parameters as additional input
  - A lot of possibilities in the future: 3+1D, other frameworks, etc.
- Combining neural networks with the GP emulators can reduce the computation time needed in Bayesian analyses to a fraction compared to current state-of-the-art methods
- Future avenues:
  - More statistics
  - Adaptive sampling routine
  - Automation
  - Multiple GP emulators with different kernels for more robust uncertainty estimates

Backup:

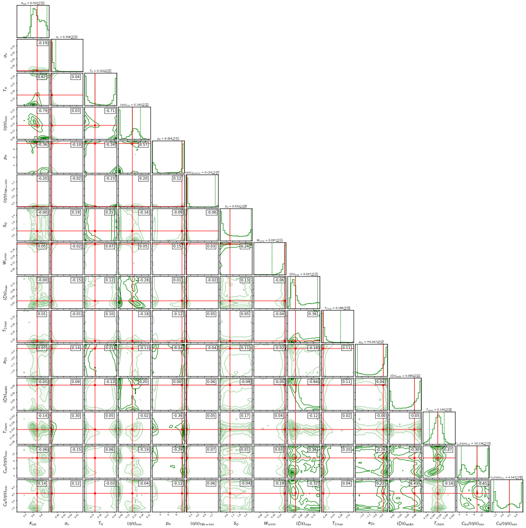
Xe+Xe observables

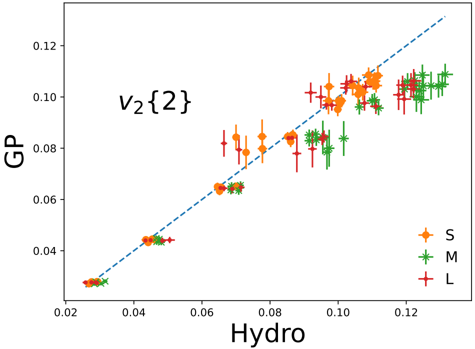






Posterior no theory error

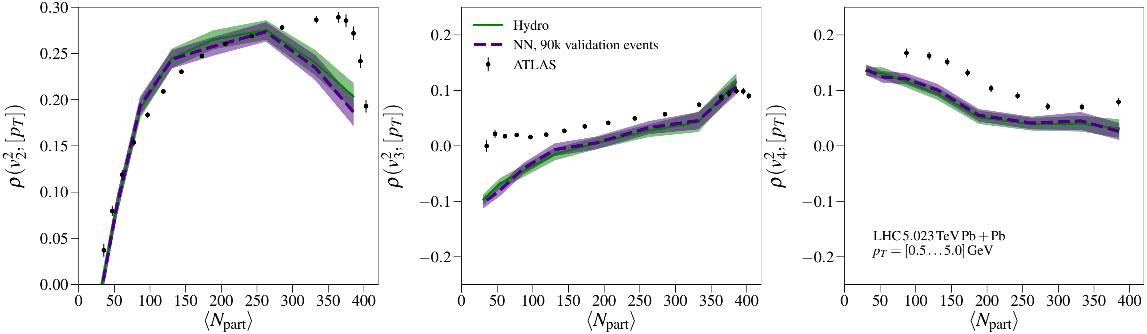




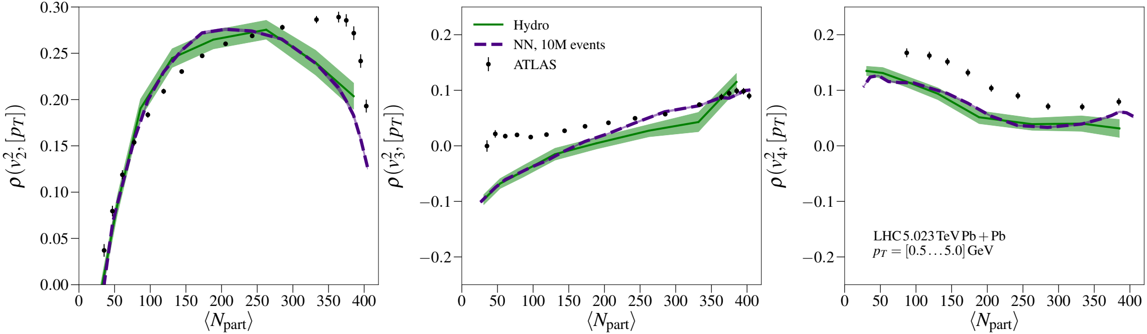
- Systematic errors in scenarios when $(\eta/s)_{\mathrm{min}}$ (M) or C_{Kn} (L) have values near prior boundaries

Validation: Flow-transverse-momentum correlations

H. Hirvonen et al. Phys. Rev. C 108, 034905 (2023)

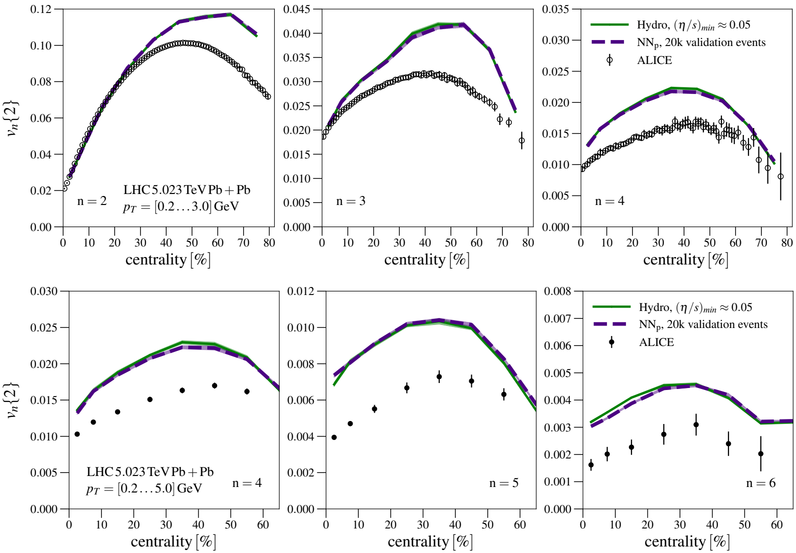


H. Hirvonen et al. Phys. Rev. C 108, 034905 (2023)

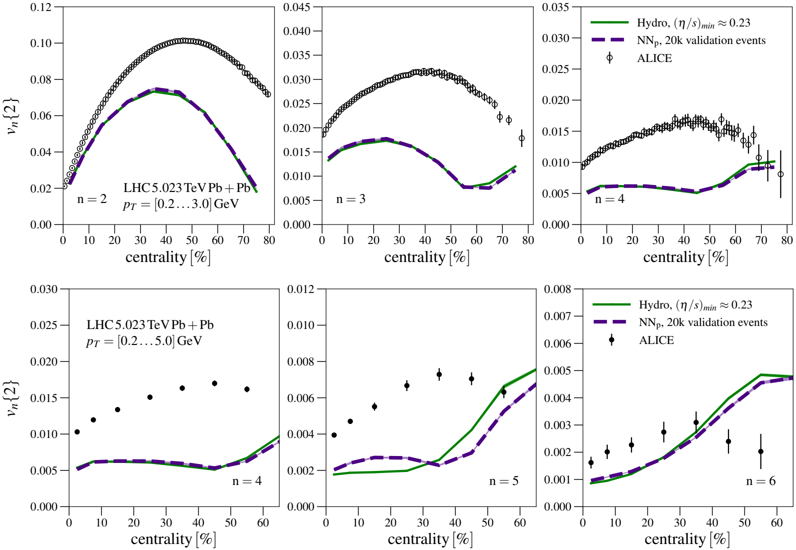


Validation: flow coefficients, low viscosity

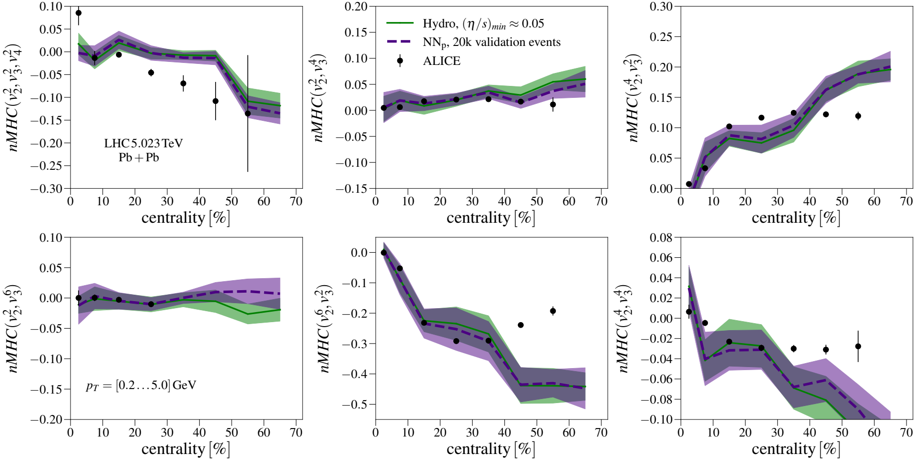
Work in progress!



Validation: flow coefficients, high viscosity



Validation: nMHC, low viscosity



Validation: nMHC, high viscosity