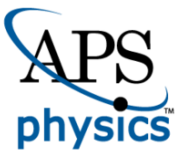

Bayesian calibration of a hybrid nuclear collision model using p-Pb and Pb-Pb data at energies available at the CERN Large Hadron Collider

J. Scott Moreland, Jonah E. Bernhard, and Steffen A. Bass

Phys. Rev. C **101**, 024911 — Published 21 February 2020

DOI: 10.1103/PhysRevC.101.024911

Bayesian calibration of a hybrid nuclear collision model using *p*-Pb and Pb-Pb data at energies available at the CERN Large Hadron Collider

J. Scott Moreland, Jonah E. Bernhard, and Steffen A. Bass
Department of Physics, Duke University, Durham, NC 27708-0305
(Dated: February 3, 2020)

We posit a unified hydrodynamic and microscopic description of the quark-gluon plasma (QGP) produced in ultrarelativistic p-Pb and Pb-Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV and evaluate our assertion using Bayesian inference. Specifically, we model the dynamics of both collision systems using initial conditions with parametric nucleon substructure, a pre-equilibrium free streaming stage, event-by-event viscous hydrodynamics, and a microscopic hadronic afterburner. Free parameters of the model which describe the initial state and QGP medium are then simultaneously calibrated to fit charged-particle yields, mean p_T, and flow cumulants. We argue that the global agreement of the calibrated model with the experimental data strongly supports the existence of hydrodynamic flow in small collision systems at ultrarelativistic energies, and that the flow produced develops at length scales smaller than a single proton. Posterior estimates for the model's input parameters are obtained, and new insights into the temperature dependence of the QGP transport coefficients and event-by-event structure of the proton are discussed.

I. INTRODUCTION

Ultrarelativistic nuclear collisions between one light-ion and one heavy-ion, e.g. ^{3}He-Au and p-Pb collisions, generate dense, compact sources of nuclear matter which produce long-range multiparticle correlations that are strikingly similar to the correlations observed in heavy-ion collisions where collectivity is commonly explained by the existence of hydrodynamic flow [1–4]. This observation suggests that hydrodynamic behavior could be manifest in small droplets of quark-gluon-plasma (QGP) [5, 6], and that flow even might develop at length scales smaller than a single proton [7].

Hydrodynamic models of ultrarelativistic nuclear collisions are complicated by a number of theoretical unknowns including the detailed geometry of the QGP initial conditions, the strength and duration of pre-equilibrium dynamics, the temperature dependence of QGP transport coefficients, and the boundaries of hydrodynamic applicability [8–11]. In general, these theoretical uncertainties tend to grow with decreasing system size, where emergent physics at sub-femtometer length scales becomes important to describe bulk properties of the produced system.

One method for reducing theoretical uncertainties is to test model calculations by varying the species of colliding nuclei at a single beam energy [12–19]. Since initial condition and hydrodynamic models generally factorize the structure of the colliding nuclei from the subsequent time dynamics of the collision, a proposed theory framework can be validated by testing its predictions for multiple collision systems using a single self-consistent set of model parameters, where only the nuclear structure in the model is permitted to vary.

Typically, the macroscopic structure of heavy nuclei, characterized e.g. by an atomic mass number and set of Woods-Saxon coefficients [20, 21], is regarded as a known input to hydrodynamic models which contributes negligible uncertainty to simulation predictions, outweighed by large uncertainties in modeling initial energy deposition and off-equilibrium dynamics [8, 22–25]. The geometry of light-ions, meanwhile, is naturally more sensitive to the detailed size and shape of individual protons and neutrons inside the nucleus, which may fluctuate event-by-event and differ significantly from the round blobs typically used to approximate nucleons in heavy-ion collisions [7, 26–29]. These nucleon substructure properties are difficult to measure and calculate from first principles and hence contribute significant uncertainty to model predictions of small systems.

Early substructure studies replaced round nucleons with composite nucleons, described by a few salient model parameters, in order to investigate the effect of each parameter on simulated observables [26, 30–33]. These sensitivity studies were able to identify cause and effect relationships between model inputs and outputs, but lacked the ability to constrain nucleon substructure parameters in any kind of global or systematic fashion. It quickly became apparent that numerous substructure implementations might be compatible with available data, and that additional work would be required to identify observables which are particularly sensitive to the average size, shape, and fluctuations of the nucleon.

Several such observables have been identified in proton-proton and proton-lepton scattering data. Measurements by the TOTEM collaboration at $\sqrt{s} = 7$ TeV, for instance, found an unexpected dip in the inelasticity density of p-p collisions at zero impact parameter [34]. It was later realized that this depression, or so-called hollowness effect, in the p-p inelastic collision profile [35] can be explained by the existence of correlated domains inside the proton, and that aspects of these domains, such as their size and correlation strength, may be constrained by comparing model predictions to inelastic p-p measurements [36, 37].

Independently, studies of coherent and incoherent J/ψ production based on a color dipole picture of vector meson production were used to simultaneously constrain both the average color charge density of the proton as well as its event-by-event fluctuations in a saturation based framework [38–41]. Initial condition studies using the IP-Glasma model of Color Glass Condensate effective field theory [42] simultaneously demonstrated that these color charge fluctuations leave a lasting imprint on the small-x gluon distribution of the proton and hence the initial geometry of QGP energy deposition [29]. In addition, it was recently shown that hydrodynamic simulations using IP-Glasma initial conditions with color charge fluctuations calibrated to fit coherent and incoherent J/ψ diffraction measured by the H1 and Zeus experiments at HERA [40, 41] provide a good description of collectivity in small and large collision systems [43].

Model parameters, such as those calibrated by the aforementioned studies, are of course always in some degree of tension. For instance, fitting one observable may require parameter values that degrade the quantitative description of some other observable. Similarly, parameters which provide an optimal description of small-system observables may lead to a sub-optimal description of heavy-ion observables or *vice versa*. It is thus import to look at the experimental data holistically, and to use model calibration methods which (1) explore all parameter combinations and (2) compare model predictions to all relevant experimental measurements in a statistically rigorous fashion.

With these considerations in mind, we present progress toward a fully global analysis of p-Pb and Pb-Pb bulk observables at $\sqrt{s_{\rm NN}} = 5.02$ TeV using a model calibration framework known as Bayesian parameter estimation. We begin in Sec. II by constructing a nuclear collision model for p-Pb and Pb-Pb collisions using initial conditions with parametric nucleon substructure and transport dynamics described by a pre-equilibrium free streaming stage,

viscous hydrodynamics, and microscopic Boltzmann transport. In Sec. III, we calibrate free parameters of the model to fit charged-particle yields, mean p_T, and anisotropic flow cumulants of *both* collision systems at $\sqrt{s_{\rm NN}} = 5.02$ TeV, and finally, in Secs. IV and V, we present posterior results for the model input parameters and comment on the implications for hydrodynamic descriptions of small collision systems.

II. NUCLEAR COLLISION MODEL

We employ a multi-stage hybrid transport model that uses relativistic viscous hydrodynamics to describe the QGP medium and microscopic Boltzmann transport to simulate the dynamics of the system after hadronization [44, 45]. The hydrodynamic initial conditions are provided by a modified version of the TRENTo model [46] with additional parameters to vary the number and size of hot spots inside the nucleon. Each initial condition profile is free streamed to the hydrodynamic starting time and matched onto the hydrodynamic energy-momentum tensor using the Landau matching procedure [47, 48]. Many of the components of the present model have been documented in previous studies [45, 46, 49]; we review each component here for completeness.

A. Initial state

We model the initial state of p-Pb and Pb-Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV using the boost-invariant TRENTo model [46]. Generally speaking, the initial three-dimensional energy density deposited by relativistic nuclear collisions is *not* boost-invariant. The longitudinal energy density fluctuates both locally point-to-point in the transverse plane as well as globally event-by-event due to asymmetries in the sampled density of participant matter [50, 51]. Nevertheless, boost-invariance has been shown to be a good approximation for both large and small collision systems when hydrodynamic observables are calculated from particles that are detected close to midrapidity [17].

The TRENTo model operates in the ultrarelativistic limit with a Lorentz factor $\gamma \gg 1$ such that each nucleus appears as a thin sheet of nuclear density in the laboratory frame. The sheets of colliding nuclear density penetrate and pass through each other in time $\tau_{\rm overlap} \sim D_{\rm nucl}/(\gamma\,\beta_z)$ in the laboratory frame, where $D_{\rm nucl}$ is the diameter of the nucleus in its rest frame, γ is the usual Lorentz factor of the accelerated ions, and β_z is their velocity along the beam axis. The resulting nuclear overlap time $\tau_{\rm overlap} \lesssim 0.1$ fm/c at top RHIC and LHC energies, and thus it is reasonable to neglect the initial transverse dynamics which occur while the nuclei pass through each other. We therefore assume that the collision produces all secondary particles at a uniform proper time $\tau_0 \ll 1$ fm/c, and that it deposits energy at midrapidity which is a function of the locally varying transverse density inside each nucleus.

Consider the collision of two nucleons, labeled A and B, with three-dimensional densities ρ_A and ρ_B in their local rest frames. The nucleon-nucleon overlap function

$$T_{nn}(\mathbf{b}) = \int d^2x_\perp \int dz_1\, \rho_A(\mathbf{x}_\perp, z_1) \int dz_2\, \rho_B(\mathbf{x}_\perp - \mathbf{b}, z_2) \tag{1}$$

describes the eikonal overlap of the two nucleons at impact parameter $\mathbf{b}$ in the transverse plane $\mathbf{x}_\perp$, orthogonal to the beam axis coordinates $z_{1,2}$. Here we assume that each nucleon is comprised of smaller constituents—e.g. valence quarks, sea quarks, and small-x gluons—which may collide to produce secondary particles and contribute to the observed inelastic nucleon-nucleon cross section.

Within a picture of independent pairwise collisions between the constituents, a Glauber model model may be used to calculate the probability $P_{\rm coll}$ that the two nucleons collide inelastically at impact parameter $\mathbf{b}$. In the limit when the number of constituents is large, it yields the particularly simple form

$$P_{\rm coll}(\mathbf{b}) = 1 - \exp[-\sigma_{\rm eff}\, T_{nn}(\mathbf{b})], \tag{2}$$

where $\sigma_{\rm eff}$ is the effective cross section for inelastic collisions between each pair of constituents, determined by fitting the inelastic proton-proton cross section

$$\sigma_{pp}^{\rm inel} = \int d^2b\, P_{\rm coll}(\mathbf{b}) \tag{3}$$

at the specified collision energy $\sqrt{s_{\rm NN}}$. We tune $\sigma_{\rm eff}$ in the present work to fit the experimental inelastic nucleon-nucleon cross section $\sigma_{\rm NN}^{\rm inel} = 7.0$ fm^2 at $\sqrt{s_{\rm NN}} = 5.02$ TeV for comparison with our chosen datasets [52]. The resulting TRENTo inelastic nucleon-nucleon cross section agrees with the experimental value to better than 2% accuracy, as verified by one of the model's standard unit tests.

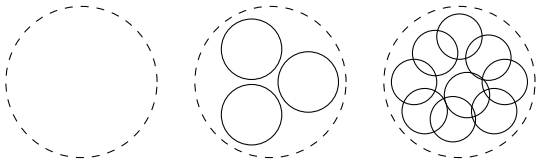

FIG. 1 Schematic of plausible nucleon shapes. The sketch on the left shows a spherically symmetric nucleon (dashed line), while the middle and right illustrations depict a fluctuating nucleon with three and nine constituents respectively (solid lines).

The nucleon densities $\rho_{A,B}$ in Eq. (1) are commonly modeled using a spherically symmetric distribution. For instance, the original implementation of the TRENTo model uses Gaussian nucleons, largely because it yields a simple analytic solution to Eq. (2). Needless to say, such approximations are admittedly crude and may have a significant effect on the dynamics of small collision systems where the nucleon size is comparable to the size of the produced QGP medium.

A number of previous studies have investigated the effects of deformed or "lumpy" nucleons. One common implementation is a superposition of three valence quarks, typically described by Gaussian or exponential form factors [7, 12, 26, 29, 32, 33]. The corresponding nucleon density is then assumed to be that of predominantly small-x gluons, seeded by the distribution of color charge in each of the three valence quarks.

In this work, we pursue a less restrictive and more parametric description of the nucleon where the number of substructure degrees of freedom is uncertain as depicted in Fig. 1. We model each nucleon's three-dimensional density $\rho_{A,B}$ as a sum of n_c independent constituents

$$\rho_{A,B}(\mathbf{x}) = \frac{1}{n_c}\sum_{i=1}^{n_c}\rho_c(\mathbf{x}-\mathbf{x}_i), \tag{4}$$

where each constituent density ρ_c is described by a Gaussian distribution

$$\rho_c(\mathbf{x}) = \frac{1}{(2\pi w_c^2)^{3/2}}\exp\left(-\frac{|\mathbf{x}|^2}{2w_c^2}\right) \tag{5}$$

of variable width w_c. The constituent positions $\mathbf{x}_i$ are sampled independently (ignoring correlations) from a Gaussian probability distribution

$$P(\mathbf{x}_i) = \frac{1}{(2\pi r_{cp}^2)^{3/2}}\exp\left(-\frac{|\mathbf{x}_i-\mathbf{x}_n|^2}{2\,r_{cp}^2}\right), \tag{6}$$

where r_{cp} is a free parameter that varies the dispersion of the constituent positions $\mathbf{x}_i$ about each nucleon position $\mathbf{x}_n$. As a matter of convenience, we sample the nucleon positions *before* determining the constituent positions. This leads to small discrepancies between the designated nucleon positions and the actual position of each nucleon's center-of-mass, owing to fluctuations in the constituent positions. The parameter r_{cp} should thus be interpreted with care. It is a computational sampling radius, *not* the Gaussian width of the sampled nucleons in their center-of-mass frame.

The two nucleons A, B are assigned a random impact parameter offset $\mathbf{b}$, and Eq. (2) is used to sample their inelastic collision probability $P_{\rm coll}(\mathbf{b})$. Note, this collision probability has no direct knowledge of the individual constituent positions; it is only *indirectly* sensitive to the constituent positions via their effect on the nucleon densities $\rho_{A,B}$. The constituents are merely used as a mechanism to deform and fluctuate each nucleon profile.

This is an important distinction between the present model and a similar nucleon substructure implementation known as the participant quark model, which allows for a subset of quarks (constituents) to participate inside a single nucleon [32, 53]. The nucleon, unlike the nucleus, cannot produce semi-stable spectator fragments in a high-energy collision. Any spectator quarks produced by a wounded quark model would be colored objects that necessarily contribute to secondary particle production. We correspondingly require that the nucleons in Eq. (4) participate as singular objects, such that all spectator matter discarded by the simulation is appropriately color-neutral and inert.

Hereafter, we switch to Milne coordinates $(\mathbf{x}_\perp, \eta_s, \tau)$, where $\mathbf{x}_\perp$ is a Cartesian coordinate (x, y) in the transverse plane, $\eta_s = \frac{1}{2}\log[(t+z)/(t-z)]$ is the system's spacetime rapidity, and $\tau = \sqrt{t^2-z^2}$ its proper time. Assuming the nucleons collide at the sampled impact parameter $\mathbf{b}$, we assign each nucleon a participant thickness

$$\tilde{T}_{A,B}(\mathbf{x}_\perp) \equiv \int dz\,\tilde{\rho}_{A,B}(\mathbf{x}_\perp \pm \mathbf{b}/2, z), \tag{7}$$

which projects its *fluctuated* nucleon density

$$\tilde{\rho}_{A,B}(\mathbf{x}) \equiv \frac{1}{n_c} \sum_{i=1}^{n_c} \gamma_i \, \rho_c(\mathbf{x} - \mathbf{x}_i) \tag{8}$$

onto the transverse plane $\mathbf{x}_\perp$ centered at midrapidity. This fluctuated nucleon density $\tilde{\rho}_{A,B}$ equals the nucleon density $\rho_{A,B}$ in Eq. (4) with additional weights γ_i sampled from a gamma distribution with unit mean and variance $1/k$. Such *ad hoc* random weights are necessary to describe the large fluctuations observed in minimum bias proton-proton collisions, although their exact physical origin is not well understood.

The resulting nucleon fluctuation variance naturally falls like $\sim 1/n_c$, where n_c is the number of constituents. This means that the natural range for the constituent fluctuations is larger when the number of constituents is larger. We therefore reparametrize the constituent fluctuations using

$$\sigma_{\text{fluct}} = 1/\sqrt{k \, n_c}, \tag{9}$$

where k is the shape parameter (inverse variance) of the gamma distribution weights and n_c is the number of nucleon constituents.

The energy density e deposited at midrapidity $\eta_s = 0$ and proper time $\tau_0 \ll 1$ fm/c is then some function

$$e(\mathbf{x}_\perp, \eta_s = 0, \tau_0) = f(\tilde{T}_A, \tilde{T}_B) \tag{10}$$

of the local participant thicknesses $\tilde{T}_A$ and $\tilde{T}_B$. A natural first guess for this mapping is an arithmetic mean

$$e(\mathbf{x}_\perp, \eta_s = 0, \tau_0) = \text{const} \times \frac{\tilde{T}_A + \tilde{T}_B}{2}, \tag{11}$$

equal to a participant or "wounded nucleon" model for initial energy deposition up to meaningless factor of two in the overall normalization constant. This wounded nucleon scaling was in fact one of the first such mappings used as a proxy for initial particle production in relativistic heavy-ion collisions [54]. It was subsequently realized, however, that the wounded nucleon model predicts the wrong scaling for charged-particle production as a function of collision centrality [55] and hence the wrong scaling for midrapidity energy deposition as a function of participant density.[1]

A simple remedy is to replace the arithmetic mean of the wounded nucleon model with a more flexible parametrization

$$e(\mathbf{x}_\perp, \eta_s = 0, \tau_0) = \text{const} \times M_p(\tilde{T}_A, \tilde{T}_B), \tag{12}$$

where M_p is a family of functions known as the generalized means

$$M_p(x, y) = \left(\frac{x^p + y^p}{2} \right)^{1/p}. \tag{13}$$

This parametrization introduces a dimensionless parameter p which varies the scaling behavior of initial energy deposition at midrapidity. For certain discrete p values, it reduces to well known functional forms such as the arithmetic, geometric, and harmonic means:

$$M_p(x, y) = \begin{cases} \max(x, y) & p \to +\infty, \\ (x + y)/2 & p = +1, \quad \text{(arithmetic)} \\ \sqrt{xy} & p = 0, \quad \text{(geometric)} \\ 2\,xy/(x + y) & p = -1, \quad \text{(harmonic)} \\ \min(x, y) & p \to -\infty. \end{cases} \tag{14}$$

Note, the form of Eq. (12) differs somewhat from our previous studies [45, 46], which parametrized the system's *entropy* density using the generalized mean ansatz:

$$s(\mathbf{x}_\perp, \eta_s = 0, \tau_0) \propto M_p(\tilde{T}_A, \tilde{T}_B). \tag{15}$$

---

[1] Typically, wounded nucleon scaling is used to parametrize the entropy density s, but the shortcomings of the model are nevertheless the same when parametrizing the system's energy density e. Both parametrizations underpredict the steep rise of particle production observed in central collisions.

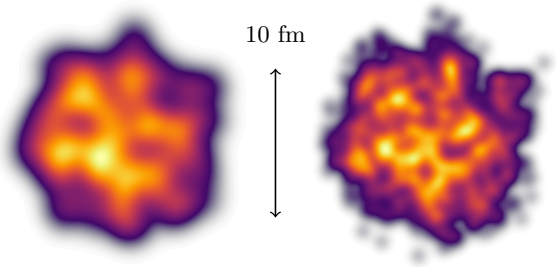

FIG. 2 Effect of nucleon substructure on the nuclear thickness function $T(x,y) \equiv \int dz\, \rho(x,y,z)$ of a ^{208}Pb nucleus. The nucleus on the left has Gaussian nucleons of width 0.8 fm, while the nucleus on the right has composite nucleons, each containing six constituents of width 0.4 fm.

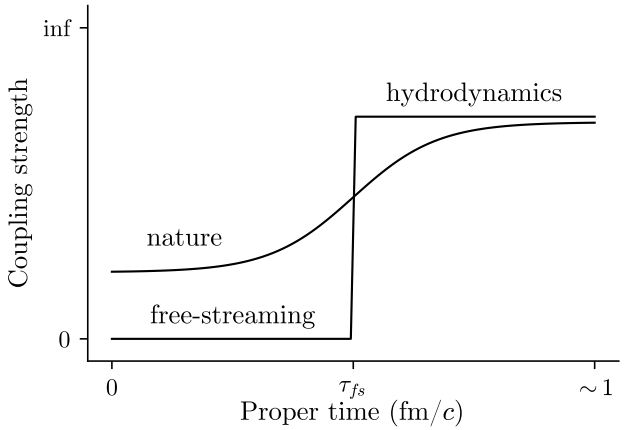

FIG. 3 Cartoon of the free streaming approximation for hydrodynamic initialization. The initial state is free streamed for proper time $\tau_{\rm fs}$ (zero coupling) before it is matched to hydrodynamics (strong coupling). This piecewise evolution approximates the more realistic scenario expected in nature where the medium's coupling strength smoothly changes as a function of time.

Our motivation for reinterpreting the left-side of this equation as an energy density in the present work is to enable the application of pre-equilibrium free streaming equations (described shortly) which require the system's initial energy density as input. This modification is permissible since the TRENTo model is formulated using particle yield observations and the approximate scaling relation $dN_{\rm ch}/d\eta \propto M_p(\tilde{T}_A, \tilde{T}_B)$, which remains valid for both static entropy density and free streamed energy density initialization.

Up to this point we have restricted our attention to a single nucleon-nucleon collision. Equation (12) is a purely local function of nuclear density in the transverse plane and hence it should, in principle, be equally valid for any pair of colliding nuclei at sufficiently high beam energy. The TRENTo model readily generalizes from individual nucleon-nucleon collisions to arbitrary nucleus-nucleus collisions by summing the participant thicknesses $\tilde{T}_{A,B}$ in Eq. (7) over all nucleons which participate in one or more inelastic collisions. The only modeling difference between p-p, p-Pb, and Pb-Pb collisions is the number and the position of the nucleons. Figure 2 shows the effect of adding nucleon substructure to a generic lead nucleus. Additional fluctuations emerge over the length scale of a nucleon, but the macroscopic geometry of the nucleus is largely unchanged.

When applying the model to heavy-ions, we sample nucleon positions from a Woods-Saxon density distribution subject to a minimum distance criteria $|\mathbf{x}_i - \mathbf{x}_j| > d_{\rm min}$ between all pairs of nucleons i, j. The minimum distance algorithm, first described in Ref. [49], uses a simple trick to resample the nucleon positions without modifying the target Woods-Saxon radial distribution. We first presample the radii of all nucleons in a given nucleus and sort them in ascending order. We then sample the solid angles of each nucleon one-by-one, starting with the nucleon closest to the center of the nucleus and working our way outwards. If a sampled nucleon position is too close to any of its previously placed neighbors, its solid angle is resampled—but not its radial coordinate—until the minimum distance criteria is satisfied. Similar methods could be used to model correlations between individual constituents inside each nucleon, although the implementation would be somewhat tedious.

B. Pre-equilibrium dynamics

There are of course two limiting cases for the strength of interactions inside the QGP medium immediately after the collision: infinitely weak coupling where the secondary partons free stream without interacting, and infinitely strong coupling where the fluid's inter-particle mean-free-path effectively vanishes. Realistically, one expects the initial parton interactions to lie somewhere between these two extremes. We therefore choose to model the QGP's initial off-equilibrium dynamics using a simple step-function approximation, depicted in Fig. 3, which free streams the initial state for proper time $\tau_{\rm fs}$ (zero coupling) before instantaneously switching to viscous hydrodynamics (strong coupling) [24, 47]. The free parameter $\tau_{\rm fs}$ allows us to parametrically vary the *time averaged* coupling strength during the initial stage of the collision.

The parametric energy deposition ansatz in Eq. (12) does not provide any information about the initial masses or momenta of particles produced in the collision. In general, these details will affect the dynamics predicted by the collisionless Boltzmann equation

$$p^\mu \partial_\mu f(x,p) = 0 \tag{16}$$

through its dependence on the underlying distribution function $f(x,p)$. Equation (16), however, simplifies for a boost-invariant gas of massless noninteracting partons with locally isotropic transverse momentum distributions. Subject to these assumptions, the energy-momentum tensor $T^{\mu\nu}(\mathbf{x}_\perp, \eta_s = 0, \tau)$ at transverse coordinate $\mathbf{x}_\perp$ and time $\tau > \tau_0$ equals [24, 47]

$$T^{\mu\nu}(\mathbf{x}_\perp, \eta_s = 0, \tau) = \frac{\tau_0}{\tau} \int_{-\pi}^{\pi} d\phi_p \, \hat{p}^\mu \hat{p}^\nu e(\mathbf{x}_\perp - (\tau - \tau_0)\hat{\mathbf{p}}_\perp, \eta_s = 0, \tau_0), \tag{17}$$

where $\hat{p}^\mu = (1, \cos\phi_q, \sin\phi_q, 0)$ and $\hat{\mathbf{p}}_\perp = (\cos\phi_q, \sin\phi_q)$ are momentum unit vectors. Here we assume that the time $\tau_0 \ll 1$ fm/c is small and define $\tau_{\rm fs} \equiv \tau - \tau_0 \sim \tau$. We also combine the unknown constant in Eq. (12) with the pre-factor τ_0 in Eq. (17) to produce a single normalization factor Norm = const $\times \tau_0$ with units of energy. The solution (17) is then decomposed in hydrodynamic form

$$T^{\mu\nu} = e u^\mu u^\nu - (P + \Pi)\Delta^{\mu\nu} + \pi^{\mu\nu}, \tag{18}$$

where e and P are the energy density and pressure in the local fluid rest frame, u^μ is the local fluid velocity, $\Delta^{\mu\nu} = g^{\mu\nu} - u^\mu u^\nu$ is the projector onto the space orthogonal to u^μ, and Π and $\pi^{\mu\nu}$ are the bulk pressure and shear stress tensor respectively. We then solve for the energy density e and fluid velocity u^μ using the Landau matching condition which defines the fluid rest frame velocity as the time-like eigenvector of $T^{\mu\nu}$ with energy density e as its eigenvalue

$$T^{\mu\nu} u_\nu = e u^\mu. \tag{19}$$

The initial bulk and shear corrections are finally solved for by subtracting the ideal pressure from the total pressure to find Π, then solving for $\pi^{\mu\nu}$ using Eq. (18):

$$\Pi = -\frac{1}{3}{\rm Tr}(\Delta_{\mu\nu} T^{\mu\nu}) - P, \tag{20}$$

$$\pi^{\mu\nu} = T^{\mu\nu} - e u^\mu u^\nu + (P + \Pi)\Delta^{\mu\nu}. \tag{21}$$

This procedure provides initial values for $T^{\mu\nu}$, u^μ, Π, and $\pi^{\mu\nu}$ which conserve energy and are consistent with the underlying hydrodynamic equation of state. We therefore expect it to provide a more realistic description of the initial stages of the collision as compared to a previous study using the T_RENTo initial condition model which set Π, $\pi^{\mu\nu}$, and u^μ initially to zero [45].

C. Hydrodynamics

After free streaming for proper time $\tau_{\rm fs}$, we transition to viscous hydrodynamics which solves the conservation equations

$$\partial_\mu T^{\mu\nu} = 0 \tag{22}$$

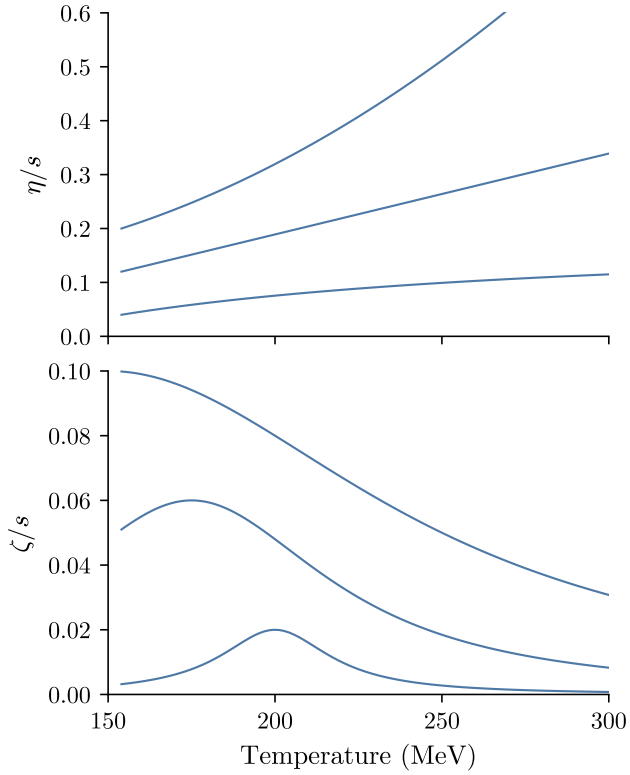

FIG. 4 Degrees of freedom in the temperature dependent shear and bulk viscosity parametrizations. Lines are chosen for illustrative purposes only and do not represent all possible variability. For instance, η/s could have a large slope and negative curvature, or ζ/s could have a large max and narrow width, neither of which are depicted above.

for the hydrodynamic energy-momentum tensor $T^{\mu\nu}$ expressed in Eq. (18) using a set of second-order Israel-Stewart equations formulated in the 14-moment approximation [56–59]. This produces a pair of relaxation-type equations

$$\tau_\Pi \Pi + \dot{\Pi} = -\zeta\theta - \delta_{\Pi\Pi}\Pi\theta + \lambda_{\Pi\pi}\pi^{\mu\nu}\sigma_{\mu\nu}, \tag{23a}$$

$$\begin{aligned}\tau_\pi \dot{\pi}^{\langle\mu\nu\rangle} + \pi^{\mu\nu} = 2\eta\sigma^{\mu\nu} - \delta_{\pi\pi}\pi^{\mu\nu}\theta + \phi_7\pi_\alpha^{\langle\mu}\pi^{\nu\rangle\alpha} \\ - \tau_{\pi\pi}\pi_\alpha^{\langle\mu}\sigma^{\nu\rangle\alpha} + \lambda_{\pi\Pi}\Pi\sigma^{\mu\nu},\end{aligned} \tag{23b}$$

for the bulk pressure Π and shear-stress $\pi^{\mu\nu}$. We model the shear viscosity η and bulk viscosity ζ as unknown temperature dependent quantities and fix the remaining transport coefficients $\{\tau_\Pi, \delta_{\Pi\Pi}, \lambda_{\Pi\pi}, \tau_\pi, \delta_{\pi\pi}, \phi_7, \tau_{\pi\pi}, \lambda_{\pi\Pi}\}$ using analytic results derived in the limit of small but finite masses [60].

The hydrodynamic equations of motion are necessarily closed using an equation of state (EoS) to relate the energy density e and pressure P of the produced medium. We use a parametrization for $P(e)$ that matches a hadron resonance gas EoS at low temperature to a lattice QCD EoS at high temperature by smoothly connecting their trace anomalies in the interval $165 \leq T \leq 200$ MeV [49]. For the lattice EoS, we use a calculation by the HotQCD collaboration for (2+1)-flavor QCD which was extrapolated to the continuum limit [61]. Recent developments in lattice QCD now enable calculations in (2+1+1)-flavors [62], i.e. with thermalized charm quarks, and the additional charm flavor has been shown to visibly affect predictions of p_T-differential flow observables [63]. Investigating this sensitivity would thus be a natural target for future improvements to the present work.

We parametrize the temperature dependence of the QGP viscosities in order to marginalize over their uncertainty when calibrating to data. For the specific shear viscosity η/s, we use a modified linear ansatz

$$(\eta/s)(T) = (\eta/s)_{\rm min} + (\eta/s)_{\rm slope} \cdot (T - T_c) \left(\frac{T}{T_c}\right)^{(\eta/s)_{\rm crv}}, \tag{24}$$

where $(\eta/s)_{\rm min}$, $(\eta/s)_{\rm slope}$, and $(\eta/s)_{\rm crv}$ are tunable parameters, and $T_c = 0.154$ GeV is the pseudocritical transition temperature of the HotQCD EoS. Meanwhile, for the specific bulk viscosity ζ/s, we use an unnormalized Cauchy

distribution

$$(\zeta/s)(T) = \frac{(\zeta/s)_{\rm max}}{1 + \left(\dfrac{T - (\zeta/s)_{T_{\rm peak}}}{(\zeta/s)_{\rm width}}\right)^2}, \tag{25}$$

described by a tunable maximum $(\zeta/s)_{\rm max}$, temperature width $(\zeta/s)_{\rm width}$, and temperature location parameter $(\zeta/s)_{T_{\rm peak}}$. Figure 4 shows several of the possible curves parametrized by Eqs. (24) and (25), although infinitely more are possible.

The aforementioned hydrodynamic equations are solved numerically using the boost-invariant VISH2+1 viscous hydrodynamics code [44, 64]. We vary the spatial grid's maximum width $x_{\rm max}$ event-by-event to accommodate systems of varying size and determine the optimal spatial step dx and time step $d\tau$ for each set of model parameters to balance trade-offs between numerical accuracy and computation time (see Appendix A). Although these details are somewhat mundane, they are critically important to the present study, since the computation time scales with the number of spacetime cells $n_{\rm cell} \sim n_x^2\, n_\tau$, and this quantity grows rapidly when n_x and n_τ are large, as is typically the case for simulations with nucleon substructure.

D. Particlization and Boltzmann transport

We evolve the system hydrodynamically down to a pre-specified switching isotherm $T_{\rm switch}$ at which point the medium is converted into particles using the Cooper-Frye formula [65]

$$E\frac{dN_i}{d^3p} = \frac{g_i}{(2\pi)^3}\int_\Sigma f_i(x,p)\, p^\mu\, d^3\sigma_\mu, \tag{26}$$

where i is an index over species, f_i is the distribution function of that species, and $d^3\sigma_\mu$ is a volume element of the isothermal hypersurface Σ defined by $T_{\rm switch}$. Thermal particles are then sampled in the rest frame of each fluid cell according to a Bose-Einstein or Fermi-Dirac distribution at zero baryon chemical potential

$$f(m,p) = \frac{1}{\exp(\sqrt{m^2+p^2}/T) \mp 1}, \tag{27}$$

where m is the mass of the sampled particle, p is its momentum, and T is the temperature of the fluid cell.

Traditionally, particlization models have sampled resonances using each particle's pole mass in Eq. (27). This approximation, however, is somewhat crude and has been known to underpredict pion production, particularly at low p_T [66–68]. We thus follow Ref. [49], and sample particles with a *distribution* of masses

$$f(p) = \int dm\, \mathcal{P}(m)\, f(m,p), \tag{28}$$

where $\mathcal{P}(m)$ is modeled by a Breit-Wigner distribution

$$\mathcal{P}(m) \propto \frac{\Gamma(m)}{(m-m_0)^2 + \Gamma(m)^2/4}. \tag{29}$$

Here m_0 is the resonance's Breit-Wigner mass, and $\Gamma(m)$ is its mass-dependent width, for which we use the simple form

$$\Gamma(m) = \Gamma_0\sqrt{\frac{m-m_{\rm min}}{m_0-m_{\rm min}}}, \tag{30}$$

where Γ_0 is the usual Breit-Wigner width, and $m_{\rm min}$ is a production threshold equal to the total mass of the lightest decay products. We tabulate the values of $\{\Gamma_0, m_0, m_{\rm min}\}$ for all particles and sample the masses of each particle during particlization [69]. The resonances are then passed to a hadronic transport model, described shortly, which simulates subsequent scatterings and decays.

When the viscous terms $\pi^{\mu\nu}$ and Π are nonzero in Eq. (18), the distribution function f must be modified to preserve the continuity of $T^{\mu\nu}$ as the system transitions from hydrodynamics to Boltzmann transport. We perform

the appropriate modification using a general method which transforms the momentum vector *inside* the distribution function [70]

$$p_i \to p_i' = p_i + \sum_j \lambda_{ij}\, p_j, \tag{31}$$

$$\lambda_{ij} = (\lambda_{\rm shear})_{ij} + \lambda_{\rm bulk}\, \delta_{ij}, \tag{32}$$

where λ_{ij} is a linear transformation matrix consisting of a traceless shear part and a bulk part which is proportional to the identity matrix.

We use for the shear viscous correction the form [70]

$$(\lambda_{\rm shear})_{ij} = \frac{\tau}{2\eta}\pi_{ij}, \tag{33}$$

with a value for η/τ obtained from the noninteracting hadron resonance gas model

$$\frac{\eta}{\tau} = \frac{1}{15T}\sum_{sp} g \int \frac{d^3p}{(2\pi)^3}\frac{p^4}{E^2} f_0(1 \pm f_0), \tag{34}$$

where the sum runs over all species in the hadron gas, and g and f_0 are the degeneracy factor and equilibrium distribution function of each species respectively.

For the bulk viscous correction, we use a novel procedure developed in Ref. [49]. The total kinetic pressure of the system is

$$P + \Pi = \sum_{\rm sp} g \int \frac{d^3p}{(2\pi)^3}\frac{p^2}{3E} f(p). \tag{35}$$

For a given bulk pressure, we rescale the momentum p inside the distribution function $f(p) \to f(p + \lambda_{\rm bulk}\, p)$ and adjust the parameter $\lambda_{\rm bulk}$ to match the total pressure on the left side of Eq. (35). This substitution of course also modifies the energy density

$$e = \sum_{\rm sp} g \int \frac{d^3p}{(2\pi)^3} E f(p), \tag{36}$$

and so a fugacity term $z_{\rm bulk}$ is introduced which modifies the yield of all particles by the same overall factor to compensate. The full transformation is then given by $f(p) \to z_{\rm bulk}\, f(p + \lambda_{\rm bulk}\, p)$, where the parameters $\lambda_{\rm bulk}$ and $z_{\rm bulk}$ are determined numerically for each value of the bulk pressure.

Once the fluid is converted into particles, we simulate its subsequent microscopic dynamics using the Ultra-relativistic Quantum Molecular Dynamics (UrQMD) transport model [71, 72]. It solves the microscopic Boltzmann equation

$$\frac{df_i(x,p)}{dt} = \mathcal{C}_i(x,p), \tag{37}$$

where f_i is the distribution function for species i, and $\mathcal{C}_i$ is its microscopic collision kernel. The model propagates all produced hadrons along classical trajectories and simulates their scatterings, resonance formations, and decays until the last interactions cease.

One primary advantage of using a microscopic transport model such as UrQMD as an afterburner, is that it realistically simulates the system break-up when the mean-free-path becomes large relative to the system size. This dilute limit is expected to play a significant role in small collision systems where the produced medium is smaller and shorter lived.

III. PARAMETER ESTIMATION

The nuclear collision model constructed in Sec. II includes a number of free parameters $\mathbf{x}$ which describe the initial state, pre-equilibrium dynamics, and hydrodynamic medium. Given values for the parameters $\mathbf{x}$, the model may be used to predict a vector of simulated observables $\mathbf{y}_{\rm model}$. For example, $\mathbf{y}_{\rm model}$ might be a vector consisting of charged-particle yields in different centrality bins. The physics model thus represents a vector-valued function $\mathbf{y}_{\rm model} = f(\mathbf{x})$ which maps the parameter values $\mathbf{x}$ to the model observables $\mathbf{y}_{\rm model}$.

The goal of this work is to estimate the true model parameters $\mathbf{x}_\star$ provided some evidence that our model predictions $\mathbf{y}_{\rm model}$ describe experimental measurements $\mathbf{y}_{\rm expt}$. The problem involves three distinct components:

TABLE I Input parameter ranges for the physics model.

Parameter	Description	Range
Norm	Normalization factor	9–28 GeV
p	Energy deposition parameter	-1 to $+1$
σ_{fluct}	Nucleon fluctuation std. dev.	0–2
r_{cp}	Constituent position radius	0–1.2 fm
n_c	Number of constituents	1–9
w_c	Width of constituents	0.2–1.2 fm
d_{min}	Minimum inter-nucleon dist.	0–1.7 fm
τ_{fs}	Free streaming time	0.1–1.5 fm/c
$(\eta/s)_{\text{min}}$	Minimum value of η/s (at T_c)	0–0.2
$(\eta/s)_{\text{slope}}$	Slope of η/s above T_c	0–8 GeV^{-1}
$(\eta/s)_{\text{crv}}$	Curvature of η/s above T_c	-1 to $+1$
$(\zeta/s)_{\text{max}}$	Maximum value of ζ/s	0–0.1
$(\zeta/s)_{\text{width}}$	Width of ζ/s peak	0–0.1 GeV
$(\zeta/s)_{T_{\text{peak}}}$	Temp. of ζ/s maximum	0.150–0.200 GeV
T_{switch}	Switching/particlization temp.	0.135–0.165 GeV

1. H_f: the hypothesis that the nuclear collision model f formulated in this work provides a realistic description of reality,
2. $H_{\mathbf{x}}$: the hypothesis that the model parameters $\mathbf{x}$ are the true model parameters $\mathbf{x}_\star$ of f, and
3. E: the evidence provided by the model $\mathbf{y}_{\text{model}}$, the experimental data $\mathbf{y}_{\text{expt}}$, and their corresponding uncertainties.

As a practical matter, we always assume that hypothesis H_f is correct, meaning there are no glaring flaws in our chosen theoretical framework. This is a significant assumption; the application of hydrodynamic simulations to small collision systems is speculative, and our conclusions are conditional on the framework making sense.

Subject to this assumption, we can apply Bayes' theorem to evaluate the hypothesis $H_{\mathbf{x}}$, *given* the evidence provided by E. Simplifying notation and writing $H_{\mathbf{x}}$ as just $\mathbf{x}$, Bayes' theorem yields

$$P(\mathbf{x}|E) \propto P(E|\mathbf{x})\, P(\mathbf{x}). \tag{38}$$

The left-side of this expression is the *posterior*: the probability of $\mathbf{x} = \mathbf{x}_\star$ given the experimental evidence E. On the right-side there are two separate terms. The first term $P(E|\mathbf{x})$ is the *likelihood* function: the probability of observing the evidence E provided that $\mathbf{x} = \mathbf{x}_\star$, and the second term $P(\mathbf{x})$ is the *prior*: an estimate of the probability of hypothesis $\mathbf{x} = \mathbf{x}_\star$ in the absence of evidence E.

We assume that the likelihood function in Eq. (38) is described by a multivariate normal distribution

$$P(E|\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^m \det \Sigma}} \exp\left(-\frac{1}{2}\Delta\mathbf{y}^\intercal \Sigma^{-1} \Delta\mathbf{y}\right), \tag{39}$$

where $\Delta\mathbf{y} = \mathbf{y}_{\text{model}}(\mathbf{x}) - \mathbf{y}_{\text{expt}}$ is a vector of size m, equal to the discrepancy of the model and experiment, and $\Sigma = \Sigma_{\text{model}}(\mathbf{x}) + \Sigma_{\text{expt}}$ is the *total* covariance matrix, equal to the sum of a modeling component $\Sigma_{\text{model}}(\mathbf{x})$ and an experimental component Σ_{expt} which account for all known sources of uncertainty in the simulated and measured observables.

A. Parameter design and observables

For the prior $P(\mathbf{x})$, we specify ranges, i.e. minimum and maximum values, for each parameter which are listed in Table I. We assume the prior distribution is constant and nonzero within each specified range and zero otherwise. The selected parameter ranges are intentionally wide to avoid clipping the calibrated posterior. For example, a previous analysis of the TRENTo model [45] found $p \approx 0$, but we use a prior range $p \in [-1, 1]$ to account for differences in the present model, e.g. nucleon substructure, which could modify its posterior. One exception is the constituent number

TABLE II Experimental data used to calibrate the model.

Pb-Pb $\sqrt{s_{\rm NN}} = 5.02$ TeV	p-Pb $\sqrt{s_{\rm NN}} = 5.02$ TeV
Charged-particle multiplicity $dN_{\rm ch}/d\eta$, $\lvert\eta\rvert < 0.5$ [73]	Charged-particle multiplicity $dN_{\rm ch}/d\eta$, $\lvert\eta\rvert < 1.4$ [74]
Two-particle flow cumulants $v_n\{2\}$ for $n = 2, 3, 4$, $\lvert\eta\rvert < 0.8$, charged-particles, $\lvert\Delta\eta\rvert > 1$, $0.2 < p_T < 5.0$ GeV [75]	Two-particle flow cumulants $v_n\{2\}$ for $n = 2, 3$, $\lvert\eta\rvert < 2.4$, charged-particles, $\lvert\Delta\eta\rvert > 2$, $0.3 < p_T < 3.0$ GeV [76]
	Charged-particle mean p_T, $0.15 < p_T < 10$ GeV, $\lvert\eta\rvert < 0.3$ [77]

n_c which we limit for practical considerations. Recall that each constituent fluctuates independently, weighted by a gamma random variable. Hence for constituent numbers $n_c \gg 1$, the fluctuations average out, and the resulting nucleon fluctuations vanish. To counteract this effect, the constituent fluctuation variance must increase as n_c increases. Eventually, these required fluctuations become unreasonably large. We find that for $n_c < 10$, the energy density fluctuations are reasonable, and hence we limit n_c to this prior range.

The likelihood function (39) provides evidence for (or against) the model parameters $\mathbf{x}$ by comparing the model predictions $\mathbf{y}_{\rm model}$ to experimental data $\mathbf{y}_{\rm expt}$. We focus on simple experimental observables in the present study which are sensitive to the bulk properties of the produced medium. We calculate for each set of model parameters the following observables at midrapidity:

- Charged-particle multiplicity $dN_{\rm ch}/d\eta$.
- Identified particle yields dN/dy of pions, kaons, and protons.
- Transverse energy production $dE_T/d\eta$.
- Charged-particle mean transverse momentum $\langle p_T \rangle$ ($0.15 < p_T < 10$ GeV).
- Identified particle mean transverse momentum $\langle p_T \rangle$ of pions, kaons, and protons.
- Mean transverse momentum fluctuations $\delta p_T/\langle p_T \rangle$ (charged-particles, $0.15 < p_T < 2.0$ GeV).
- Two-particle flow cumulants $v_n\{2\}$ for $n = 2, 3, 4$ (charged-particles, $0.2 < p_T < 5.0$ GeV for ALICE, and $0.3 < p_T < 3.0$ GeV for CMS).
- Four-particle flow cumulant $v_2\{4\}$ (charged-particles, $0.2 < p_T < 5.0$ GeV).
- Symmetric cumulants SC$(4, 2)$ and SC$(3, 2)$.

Each observable is calculated from the list of final state particles produced by UrQMD using the same methods applied by experiment. We generally match the kinematic cuts of all measurements with two exceptions: we use a larger rapidity interval $|\eta| < 0.8$ than experiment for some boost-invariant observables to improve our finite particle statistics, and we do not apply a rapidity gap, e.g. $|\Delta\eta| > 1$, between pairs of particles when calculating the two-particle cumulant $v_n\{2\}$ since we already oversample particles from each hydrodynamic event, and this oversampling suppresses non-flow correlations.

At the time of this writing, many of the aforementioned experimental observables are not yet published for p-Pb and Pb-Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV. We therefore restrict our calibration to the subset of measured and published observables listed in Table II. Notably absent from this list are the four-particle cumulants $v_n\{4\}$ at $\sqrt{s_{\rm NN}} = 5.02$ TeV despite being measured and published. Unfortunately, the four-particle cumulants require minimum-bias event statistics an order of magnitude larger than those used in this work. We therefore refrain from *calibrating* on the four-particle cumulants, although we do show calculations of the Pb-Pb four-particle cumulant $v_2\{4\}$ later in the text, using a single set of calibration parameters.

Most of the calibration observables listed in Table II are calculated as a function of collision centrality, where centrality is defined using some measure of the underlying event activity, e.g. the charged-particle yield in a given rapidity window. When calculating these observables, we generate $\mathcal{O}(10^4)$ minimum-bias events at each design point and divide the events into centrality bins using the charged-particle yield at midrapidity, similar to the procedure used by experiment.

However, for some observables such as p-Pb mean p_T [77] and flow cumulants $v_n\{k\}$ [76], the experiments use a special high-multiplicity trigger to select rare, ultra-central events according to the number of charged-particles produced $N_{\rm ch}$ or detector tracks offline $N_{\rm trk}^{\rm offline}$. These high-multiplicity bins are too selective for our modest minimum bias event sample, and so a different procedure is required. We exploit, for this purpose, the approximate monotonic

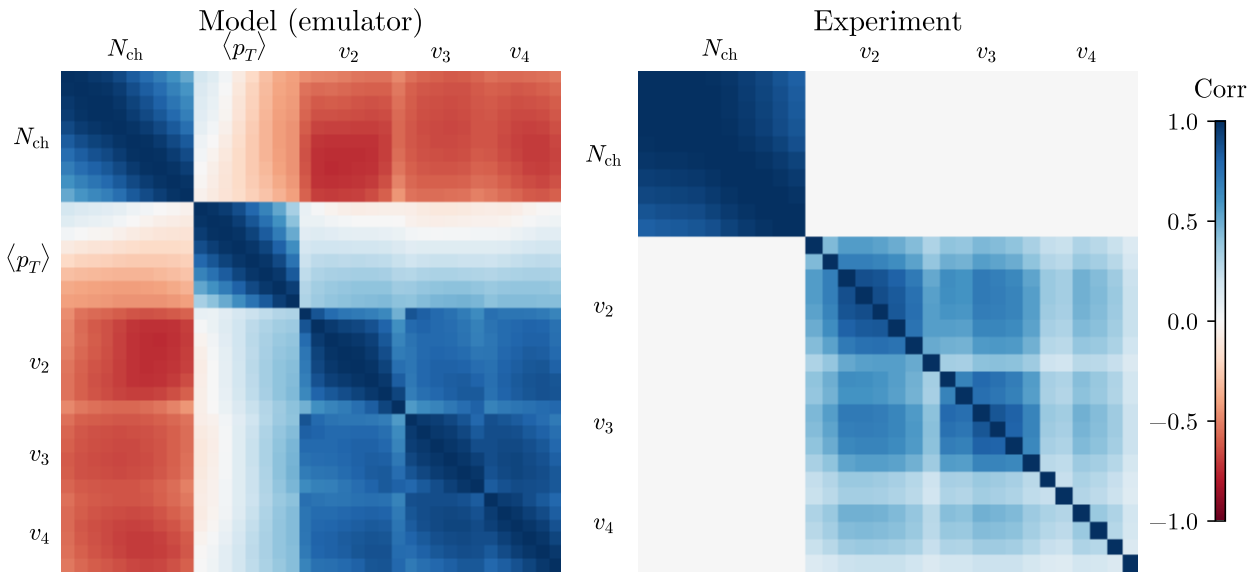

FIG. 5 Visualization of the Pb-Pb correlation matrix $\mathrm{corr}(y_i, y_j) = \mathrm{cov}(y_i, y_j)/(\sigma_i \sigma_j)$ for the model (emulator) at a random point in parameter space (left-side) and for the experimental data (right-side). Each cell represents an observable in a single centrality bin. Experimental statistical and systematic errors are from ALICE [73, 75]. The experimental correlation structure is modeled using Eq. (48).

relation between each event's initial energy density

$$\left.\frac{dE}{d\eta_s}\right|_{\eta_s=0} = \tau_0 \int d^2x_\perp e(\mathbf{x}_\perp, \eta_s = 0, \tau_0) \tag{40}$$

and its final charged-particle density $(dN_{\rm ch}/d\eta)|_{\eta=0}$ at midrapidity.

Consider, for example, a single multiplicity bin $[N_{\rm ch}^{\rm low}, N_{\rm ch}^{\rm high}]$ which selects events from a minimum bias event sample with $p_T^{\rm min} < p_T < p_T^{\rm max}$ and $|\eta| < \eta^{\rm max}$. Let $\langle N_{\rm ch} \rangle$ denote the average charged-particle multiplicity of these events. We first rescale the experimental multiplicity bin edges

$$[N_{\rm ch}^{\rm low}, N_{\rm ch}^{\rm high}] \to \left[\frac{N_{\rm ch}^{\rm low}}{\langle N_{\rm ch}\rangle}, \frac{N_{\rm ch}^{\rm high}}{\langle N_{\rm ch}\rangle}\right] \tag{41}$$

in order to reexpress each bin edge as a unitless variable. These bin edges are then associated with a pair of energy bin edges

$$\left[\frac{E^{\rm min}}{\langle E\rangle}, \frac{E^{\rm max}}{\langle E\rangle}\right] \leftrightarrow \left[\frac{N_{\rm ch}^{\rm low}}{\langle N_{\rm ch}\rangle}, \frac{N_{\rm ch}^{\rm high}}{\langle N_{\rm ch}\rangle}\right], \tag{42}$$

where $E \propto (dE/d\eta_s)|_{\eta_s=0}$ is the midrapidity energy of a single event in the desired kinematic range, and $\langle E \rangle$ is the corresponding average energy over the full minimum bias event sample.

Finally, we mimic the method used by experiment and apply Eq. (42) to select rare high-multiplicity events from a continuous stream of minimum-bias TRENTo events satisfying the correct bin edges. This of course means that, in addition to running a large sample of minimum-bias events for centrality binned observables, we must also generate (much like experiment) a separate sample of multiplicity triggered events. In practice, we use a few hundred to a few thousand events per multiplicity bin, depending on the type of observable.

B. Experimental uncertainties

We also take stock of the statistical and systematic errors reported by each experiment and incorporate their uncertainty into the likelihood covariance matrix

$$\Sigma = \Sigma_{\rm model} + \Sigma_{\rm expt} \tag{43}$$

appearing in Eq. (39), which includes uncertainty contributions from both the model $\Sigma_{\rm model}$ and experimental data $\Sigma_{\rm expt}$. The experimental contribution to the covariance $\Sigma_{\rm expt}$ is further broken down into statistical and systematic components

$$\Sigma_{\rm expt} = \Sigma^{\rm stat}_{\rm expt} + \Sigma^{\rm sys}_{\rm expt}. \tag{44}$$

The statistical errors in $\Sigma^{\rm stat}_{\rm expt}$ are uncorrelated, and thus its covariance matrix is diagonal

$$\Sigma^{\rm stat}_{\rm expt} = {\rm diag}[(\sigma^{\rm stat}_1)^2, (\sigma^{\rm stat}_2)^2, \ldots (\sigma^{\rm stat}_m)^2], \tag{45}$$

where $\sigma^{\rm stat}_i$ is the statistical uncertainty of observable y_i in the observable vector $\mathbf{y}_{\rm expt} = (y_1, \ldots, y_m)$. The systematic errors, meanwhile, are typically correlated, but the correlation structure is not reported by the experiments so we assert a reasonable form. We can expand the systematic covariance as

$$(\Sigma^{\rm sys}_{\rm expt})_{ij} = \rho_{ij}\sigma_i\sigma_j, \tag{46}$$

where σ_i and σ_j are the systematic errors of observables y_i and y_j respectively, and ρ_{ij} is the Pearson correlation coefficient

$$\rho_{ij} = \frac{{\rm cov}(y_i, y_j)}{\sigma_i\sigma_j} \tag{47}$$

between observable y_i and y_j, which satisfies $\rho_{ij} = 1$ for $i = j$ and $|\rho_{ij}| \leq 1$ for $i \neq j$. We assume that each observable is correlated across different centrality/multiplicity bins, and uncorrelated with observables of a different type, e.g. correlations between yields and flows. This is a crude simplifying assumption, but it is better than neglecting the correlation structure of the experimental data entirely.

For the correlation structure between different observable bins, we assert the simple Gaussian form

$$\rho^{\rm sys}_{ij} = \exp\left[-\frac{1}{2}\left(\frac{b_i - b_j}{\ell}\right)^2\right], \tag{48}$$

where b_i and b_j are the midpoints of two observable bins of a single type (centrality or relative multiplicity), and ℓ is a correlation length which describes how quickly the observable bins decorrelate as the distance between the bins increases. We use centrality correlation lengths $\ell = 100$ for all of the centrality binned Pb-Pb observables and $\ell = 30$ for the centrality binned p-Pb charged-particle yield $dN_{\rm ch}/d\eta$. The p-Pb mean p_T and flow observables, meanwhile, use relative multiplicity bins $N_{\rm ch}/\langle N_{\rm ch}\rangle$ and $N^{\rm offline}_{\rm trk}/\langle N^{\rm offline}_{\rm trk}\rangle$ which necessitate a smaller correlation length $\ell = 5$. We show an example correlation matrix

$${\rm corr}(y_i, y_j) = {\rm cov}(y_i, y_j)/(\sigma_i\sigma_j) \tag{49}$$

for the Pb-Pb experimental data constructed using Eq. (48) on the right-side of Fig. 5. Here y_i denotes an element of the experimental data $\mathbf{y}_{\rm expt}$ and σ_i its corresponding uncertainty.

C. Model emulator

In principle, one could calculate the likelihood function in Eq. (39) directly, e.g. by running the model a large number of times at a given parameter point $\mathbf{x}$ to calculate the model observables $\mathbf{y}_{\rm model}(\mathbf{x})$ from the ensemble of simulated events, but in practice such a procedure would be intractable. The model is computationally intensive to evaluate, and thousands of events are required to calculate the simplest observables at a single parameter point. Moreover, we need to evaluate Eq. (39) numerous times in order to sample the multidimensional posterior distribution so that the samples may be histogrammed and visualized.

We therefore follow an established framework for computationally intensive models and train an emulator to act as a fast surrogate for the full physics simulation [78–80]. The emulator enables essentially instantaneous predictions for $\mathbf{y}_{\rm model} = f(\mathbf{x})$ and allows us to sample the posterior distribution millions of times. In order to train the emulator, we first generate a scaffolding of the parameter space using a maximin Latin hypercube design [81] to distribute 500 points throughout our 15-dimensional parameter space according to the parameter ranges in Table I. We then run minimum-bias and multiplicity triggered p-Pb and Pb-Pb events at each design point and calculate the model observables from the ensemble of events.

Specifically, let X denote the $d \times n$ design matrix of $d = 500$ training points, where each training point is a vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ of the $n = 15$ model parameters in Table I. Similarly, let Y denote the corresponding $d \times m$ observables matrix, where each row of Y is a vector $\mathbf{y}_{\text{model}} = (y_1, y_2, \ldots, y_m)$ of m model observables. The emulator operates strictly on model outputs, so let's temporarily drop the subscript on $\mathbf{y}_{\text{model}}$ to declutter the notation. Our goal is to train an emulator for the physics model f using the discrete observations $f : X \mapsto Y$.

We use for this purpose a specific type of emulator known as a Gaussian process (GP) emulator [82]. The advantage of using GPs is that they provide an estimate of their own uncertainty which allows us to account for this uncertainty when constructing the covariance matrix Σ in Eq. (39). One quirk of GPs is that they are restricted to scalar-valued functions, i.e. functions of one output, whereas we require an emulator for vector-valued functions with multiple outputs. This restriction is commonly circumvented using principal component analysis (PCA): a general procedure which transforms a set of correlated variables $\mathbf{y} = (y_1, y_2, \ldots, y_m)$ into a new basis representation $\mathbf{z} = (z_1, z_2, \ldots, z_m)$ where the linear correlations between z_i, z_j vanish for all $i \neq j \in m$ [83]. Independent GPs can then be used to emulate each $z \in \mathbf{z}$ since the variables $(z_1, \ldots, z_m)$ are linearly uncorrelated. The emulated vector $\mathbf{z}$ is then easily reexpressed in the basis of $\mathbf{y}$ through its inverse transformation.

We first preprocess our model observables by centering and scaling each column of Y (single observable) to zero mean and unit variance to generate a standardized observable matrix $\hat{Y}$. PCA is then used to reexpress each row-vector $\hat{\mathbf{y}}_i$ of $\hat{Y}$ (all standardized observables at a single design point) in the new principal component basis

$$\hat{\mathbf{y}}_i = \sum_{j=1}^{m} z_{ij} \mathbf{v}_\mathbf{j}, \tag{50}$$

where $\hat{\mathbf{y}}_i$ are the standardized observables of the i-th row-vector (design point) of matrix $\hat{Y}$, and z_{ij} and $\mathbf{v}_j$ are the coefficients and vectors of its j-th principal component.

The principal components are reported in order of explained variance, with the first principal component vector $\mathbf{v}_1$ accounting for the most variance in $\hat{Y}$, and the last principal component vector $\mathbf{v}_m$ accounting for the least. We then train a set of independent GPs $\{z_i = \text{gp}_i(\mathbf{x})\}$ to predict the first k principal components $(z_1, \ldots, z_k)$ as a function of the model parameters $\mathbf{x}$ which vary across the design X. For the present study, we use $k = 7$ principal components when emulating the p-Pb system and $k = 8$ principal components when emulating Pb-Pb, chosen to describe 99.5% of the total observed variance of each system.

The GPs are essentially fancy interpolators applied to the model's training points and PCA transformed observables. Each GP reports a mean value $z(\mathbf{x})$ as well as an estimated error $\delta z(\mathbf{x})$ which accounts for statistical noise in the training data and interpolation error between the design points. Once the GPs are trained, we can predict the observables $\mathbf{y}_{\text{model}}$ at parameter point $\mathbf{x}$ by transforming the vector of principal components

$$\mathbf{z}(\mathbf{x}) = [z_1(\mathbf{x}), z_2(\mathbf{x}), \ldots, z_k(\mathbf{x})] \tag{51}$$

back to physical space.

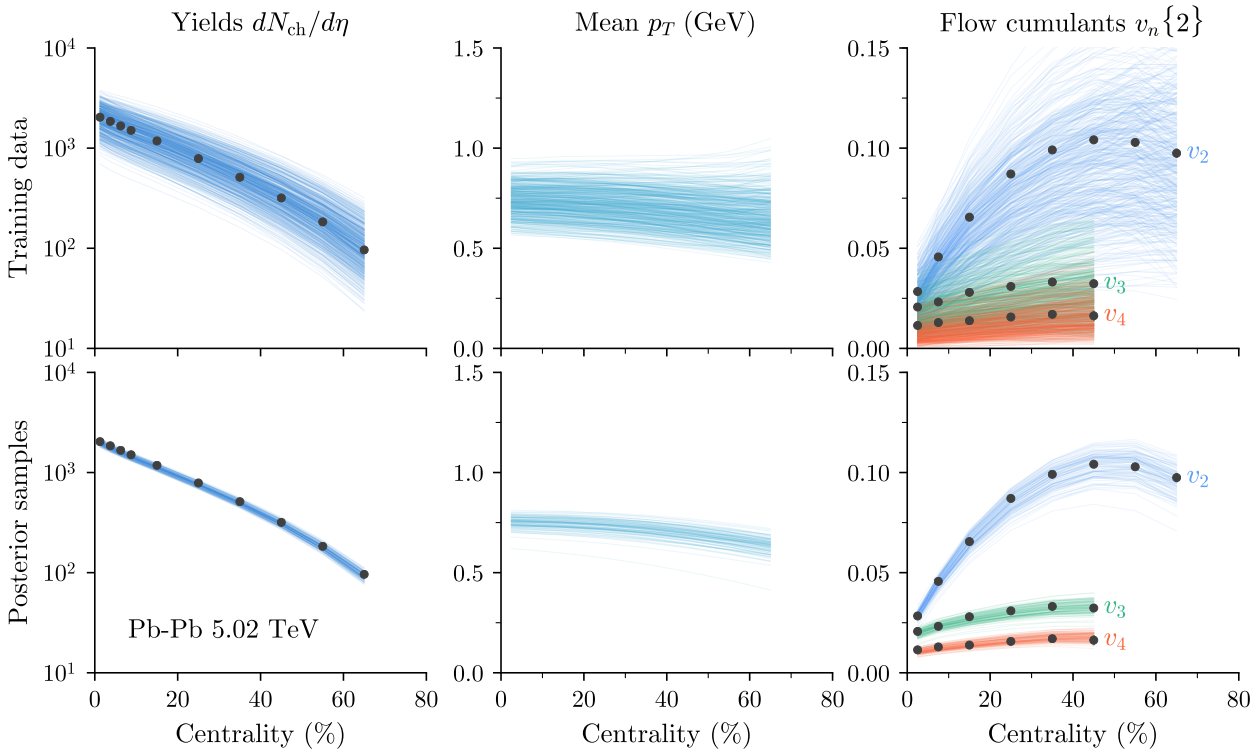

FIG. 6 Simulated observables compared to experimental data for Pb-Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV. Top row: explicit model calculations (no emulator) for each of the $d = 500$ design points; bottom row: emulator predictions for $n = 100$ random samples drawn from the posterior. Points with error bars are experimental data from ALICE with statistical and systematic errors added in quadrature [73, 75].

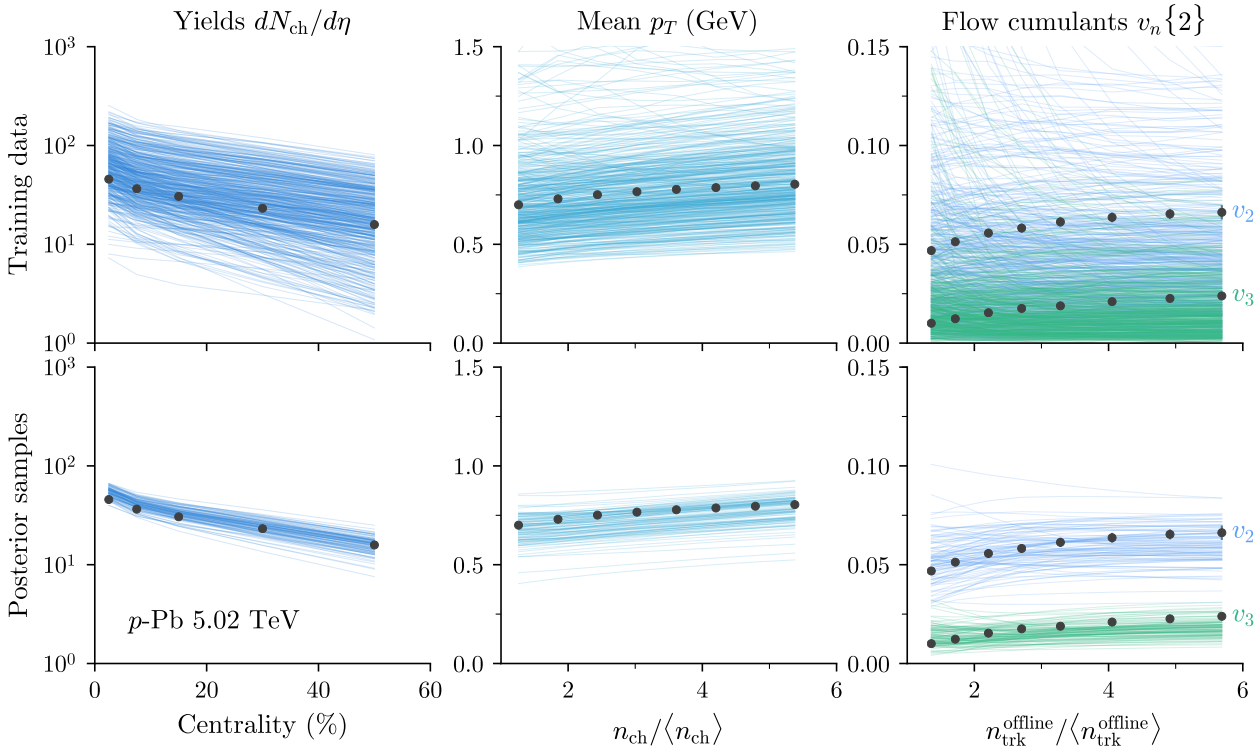

FIG. 7 Same as Fig. 6 but for p-Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV. Note that multiplicity bins are used for mean p_T and flow cumulant observables to match the bins used by experiment. Experimental data are from ALICE [74, 77] and CMS [76].

Similarly, we can construct the covariance matrix of the observables in PCA space

$$\mathrm{cov}(z_i, z_j) = \mathrm{diag}[(\delta z_1)^2, (\delta z_2)^2, \ldots, (\delta z_k)^2], \qquad (52)$$

and transform it back to physical space as well to obtain the covariance matrix $\Sigma_{\rm model}$ of the model observables at parameter point $\mathbf{x}$.

The resulting emulator therefore predicts both a mean prediction $\mathbf{y}_{\rm model}(\mathbf{x})$ and an uncertainty covariance matrix $\Sigma_{\rm model}(\mathbf{x})$ which accounts for multiple sources of model and emulator uncertainty, including the truncation error expected from using a finite number of principal components $k < m$. The model covariance matrix $\Sigma_{\rm model}$ includes so-called *known*-unknowns such as statistical error and emulator interpolation error, but not *unknown*-unknowns such as the overall validity of small-system hydrodynamics, i.e. things which lack a unified consensus or are difficult to quantify. We show in Fig. 5 the resulting Pb-Pb correlation matrix $\mathrm{corr}(y_i, y_j)$ for the model (emulator) at a random parameter point $\mathbf{x}$ in the design space (left-side), along side the same correlation matrix for the experimental data (right-side) discussed previously. For additional information on the model emulator, we direct the reader to Appendix B which includes several validation tests of the emulator prediction accuracy.

D. Bayesian calibration

In order to calibrate the model on two different collision systems, we expand the likelihood function (39) into a joint likelihood

$$P(E|\mathbf{x}) = P(E_{\text{Pb-Pb}}|\mathbf{x}) \cdot P(E_{p\text{-Pb}}|\mathbf{x}), \qquad (53)$$

where E subsumes all evidence from the p-Pb and Pb-Pb collision systems. We then perform Markov-chain Monte Carlo (MCMC) importance sampling on the posterior distribution $P(\mathbf{x}|E)$ in Eq. (38) to draw random samples of the "true" model parameters $\mathbf{x}_\star$, given the evidence provided by the model predictions and the experimental data [84, 85]. For this we use an affine-invariant sampler which uses a large ensemble of interdependent walkers [84, 85] and allow the MCMC chain to burn-in before generating $\mathcal{O}(10^7)$ posterior samples.

IV. RESULTS

We show the simulated and emulated model observables (thin colored lines) for Pb-Pb collisions in Fig. 6 and for p-Pb collisions in Fig. 7 at $\sqrt{s_{\rm NN}} = 5.02$ TeV compared to experimental data from the CMS [76] and ALICE collaborations [73–75, 77]. The top row of each figure shows explicit model calculations at each of the $d = 500$ design points (training data), while the bottom row shows emulator predictions for $n = 100$ random parameter samples drawn from the Bayesian posterior (sampled from the MCMC chain). Each column shows a different class of observable. The charged-particle yield $dN_{\rm ch}/d\eta$ is shown on the left, mean p_T is in the middle, and two-particle flow cumulants $v_n\{2\}$ for $n = 2, 3, 4$ are on the right. The Pb-Pb mean p_T and p-Pb $v_4\{2\}$ datasets are missing and hence are omitted from the present calibration.

Notice the large spread of the observables calculated at the training points (top row of each figure). The design is constructed to vary each parameter across a wide range of values, specified in Table I, and hence the corresponding model calculations are equally uncertain. We also point out that there is considerably more variance in the p-Pb training data than the Pb-Pb training data. The p-Pb yields, mean p_T, and flow cumulants all vary wildly within the chosen parameter ranges. For instance, we can turn the p-Pb flows completely *off* with suitably chosen parameters which is not possible in the Pb-Pb system. Evidently the p-Pb model predictions are far more sensitive to modeling uncertainties.

Conversely, the calibrated (posterior sampled) emulator predictions (bottom row of Figs. 6 and 7) are far better constrained and nicely track the experimental data points. We emphasize here that the posterior parameter values are obtained from a *simultaneous* calibration to p-Pb and Pb-Pb data, and thus they are self-consistent between the two systems. The spread in the posterior samples reflects different sources of model and experimental uncertainty as well as tension in the optimal fit parameters which describe each observable. We demonstrate later in the text that a single set of model parameters well describes all of the calibration data, and thus we believe that much of the spread in the posterior samples is uncertainty contributed by our emulator.

We also note that although the p-Pb posterior samples have a somewhat larger spread than the Pb-Pb samples, the percentage uncertainty of the p-Pb emulator is similar to that of the Pb-Pb emulator, and thus the difference is likely due to the larger variance of the p-Pb training data. The uncertainty in the posterior distribution could thus be improved by running the calibration with more design points or with a narrower range of parameter values to increase the density of the training points and reduce interpolation uncertainty.

We now direct our attention to Fig. 8 which shows the main result of this work, the posterior distribution of the model input parameters. Recall that the posterior $P(\mathbf{x}|E)$ is the probability that our hypothesis $\mathbf{x} = \mathbf{x}_\star$ is correct, given the evidence E provided by the predictions of the model and the experimental data. The present posterior has 15 dimensions, one for each parameter listed in Table I, and thus its joint distribution cannot be summarized by one figure alone. We therefore sample the distribution and histogram the samples to project the distribution onto one or two dimensions at a time.

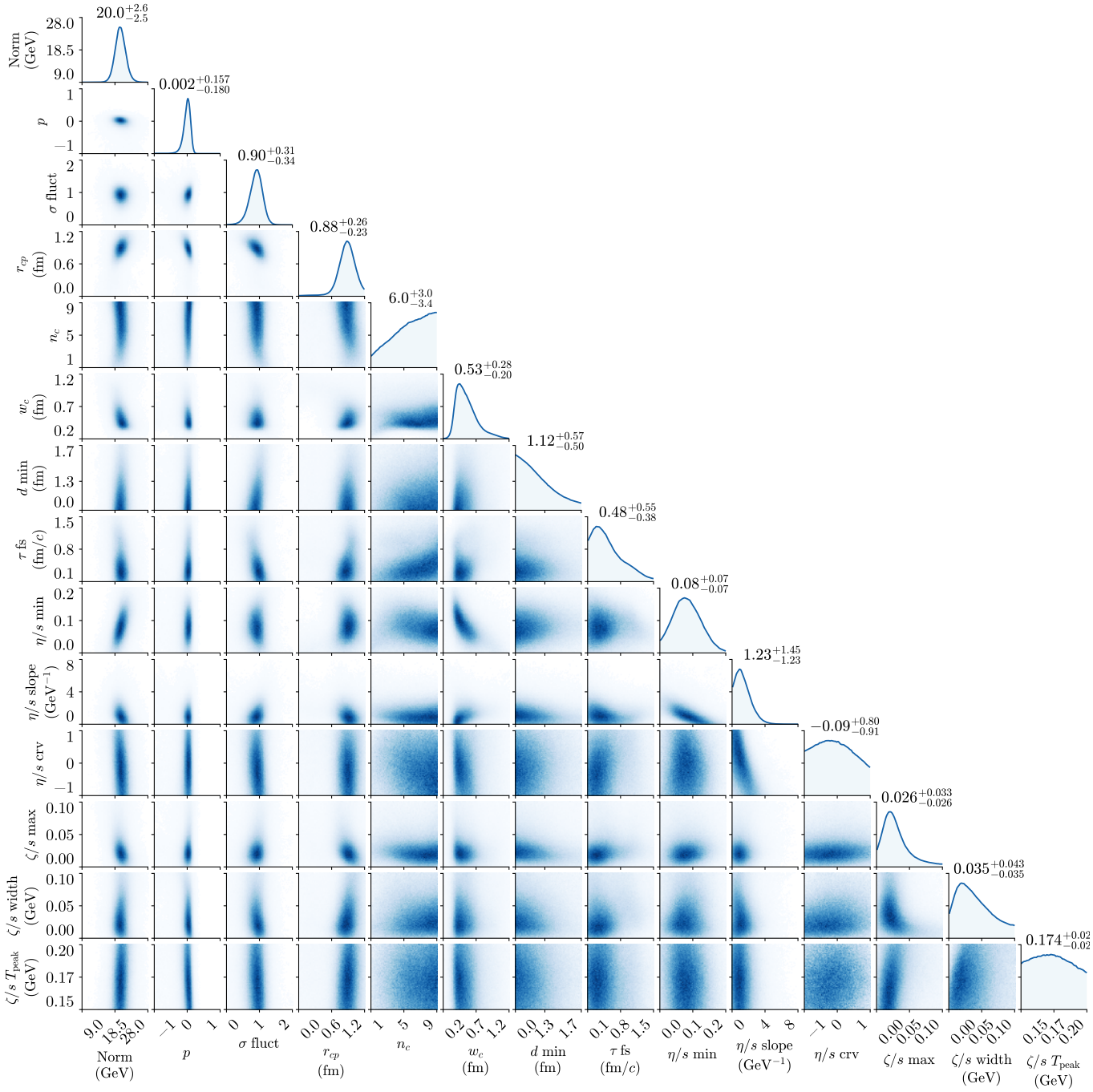

FIG. 8 Bayesian posterior distribution of the model input parameters. The diagonal panels show the marginalized distributions of individual model parameters, while off-diagonal panels show the joint distributions for pairs of model parameters, visualizing their correlations. The marginalized distribution medians and 90% credible intervals are annotated along the diagonal. Posterior distribution on the switching temperature $T_{\rm switch}$ is flat and not shown.

TABLE III Posterior parameter estimates corresponding to Fig. 8. The reported values are for the distribution median and 90% highest posterior density credible interval.

Initial condition / Pre-eq		QGP medium	
Norm	$20.0^{+2.6}_{-2.5}$ GeV	$(\eta/s)_{\rm min}$	$0.08^{+0.07}_{-0.07}$
p	$0.002^{+0.157}_{-0.180}$	$(\eta/s)_{\rm slope}$	$1.23^{+1.45}_{-1.23}$ GeV^{-1}
$\sigma_{\rm fluct}$	$0.90^{+0.31}_{-0.34}$	$(\eta/s)_{\rm crv}$	$-0.09^{+0.80}_{-0.91}$
r_{cp}	$0.88^{+0.26}_{-0.23}$ fm	$(\zeta/s)_{\rm max}$	$0.026^{+0.033}_{-0.026}$
n_c	$6.0^{+3.0}_{-3.4}$	$(\zeta/s)_{\rm width}$	$0.035^{+0.043}_{-0.035}$ GeV
w_c	$0.53^{+0.28}_{-0.20}$ fm	$(\zeta/s)_{T_{\rm peak}}$	$0.174^{+0.020}_{-0.024}$ GeV
$d_{\rm min}$	$1.12^{+0.57}_{-0.50}$ fm	$T_{\rm switch}$	$0.148^{+0.013}_{-0.013}$ GeV
$\tau_{\rm fs}$	$0.48^{+0.55}_{-0.38}$ fm/c		

Each diagonal panel is the distribution of a *single* model parameter (marginalized over all others), and each lower-diagonal panel is the joint distribution of a *pair* of model parameters (marginalized over all others). We also report numeric estimates for each parameter's median value and 90% credible interval and annotate their values along the distribution diagonal (see Table III). For example, the fictitious parameter estimate $q = 2.45^{+0.20}_{-0.15}$ reports a median value $\tilde{q} = 2.45$ and 90% credible interval $2.30 < q < 2.65$.

A. Initial condition properties

The T_RENTo normalization factor Norm $= 20.0^{+2.6}_{-2.5}$ and energy deposition parameter $p = 0.002^{+0.157}_{-0.180}$ are well constrained by the present analysis. Moreover, Figs. 6 and 7 show that the model predictions using these values nicely describe both the p-Pb and Pb-Pb calibration observables. While it would not be surprising to fit one or two of these observables using a narrow range of parameter values, the quality of the combined fit (more on this later) and the number of observables described is highly non-trivial. For example, consider the ratio of the p-Pb charged-particle yield to the Pb-Pb charged-particle yield. As the energy deposition parameter p trends toward positive (negative) infinity, particle production scales like the maximum (minimum) of the two nuclear thickness functions. This has a much stronger effect on the highly asymmetric p-Pb system than it does on the Pb-Pb system; hence the parameter p strongly affects the ratio of the two average yields.

It is therefore compelling that $p \approx 0$ correctly describes the charged-particle yield $dN_{\rm ch}/d\eta$ of both systems, while simultaneously describing the centrality dependence of $v_n\{k\}$, an observable which is also known to strongly depend on p [49]. Specifically, this value $p \approx 0$ corresponds to an energy deposition mapping proportional to the geometric mean of participant nuclear thickness

$$e(\mathbf{x}_\perp, \eta_s = 0, \tau_0) \propto \sqrt{\tilde{T}_A\, \tilde{T}_B}. \tag{54}$$

We caution, however, that this specific analytic form should not be interpreted too literally. For instance, a generalized mean described by $p = 0.05$ is well within our 90% credible interval, but it does not equal the geometric mean of Eq. (54). We also note that this scaling is somewhat different than the scaling obtained from previous analyses of the T_RENTo model, which parametrized the *entropy* density using a framework which assumed instant thermalization and zero pre-equilibrium flow. Evidently, both prescriptions prefer geometric mean scaling, but each prescription leads to a somewhat different interpretation of the initially produced quantity.

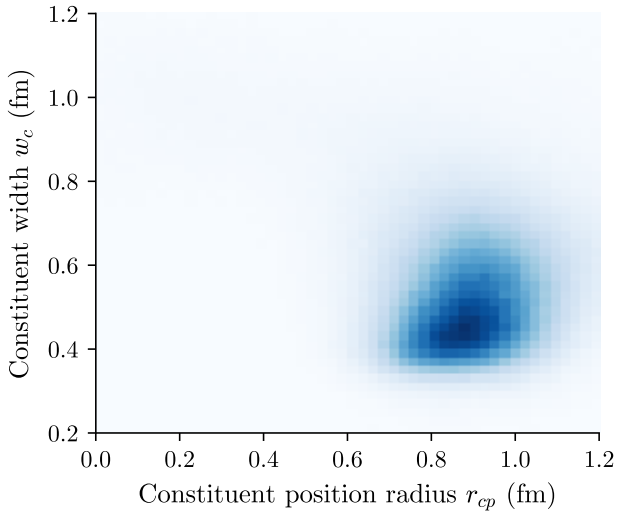

FIG. 9 Posterior distribution (blue histogram) for the constituent position sampling radius r_{cp} and constituent width w_c. The prior for r_{cp} and w_c spans the full plot range.

Continuing down the diagonal in Fig. 8, we see that the constituent position sampling radius $r_{cp} = 0.88^{+0.26}_{-0.23}$ fm, and the constituent width $w_c = 0.53^{+0.28}_{-0.20}$ fm. Figure 9 shows the joint posterior distribution of both parameters, illustrating the constraining power of the Bayesian analysis. While the sampling radius r_{cp} varies the size of the nucleons, we caution that its specific meaning should be interpreted with care; it specifies a computational *sampling* radius, not a physical nucleon width. Consider, for instance, a single nucleon with $n_c = 2$ constituents. If the two constituents positions land on the same side of the nucleon, the effective nucleon size will be smaller than the Gaussian sampling radius r_{cp}. Despite this idiosyncrasy, one can easily define a physical nucleon width in the nucleon center-of-mass frame *ex post facto*, given specific values for the sampling radius r_{cp}, constituent width w_c, and constituent number n_c.

For example, using the posterior distribution's median values, $r_{cp} = 0.88$ fm, $n_c = 6$, and $w_c = 0.53$ fm, we can generate a large ensemble of random nucleon con-

figurations and average their density in each nucleon's center-of-mass frame. The resulting ensemble-averaged nucleon density

$$\langle \rho_n(\mathbf{x}) \rangle = \frac{1}{(2\pi w^2)^{3/2}} \exp\left(-\frac{|\mathbf{x}|^2}{2w^2}\right) \tag{55}$$

is described by a single Gaussian of width $w = 0.96$ fm. This nucleon width is consistent with a previous estimate $w = 0.96^{+0.04}_{-0.05}$ fm obtained by a similar Bayesian analysis of Pb-Pb collisions at $\sqrt{s_{\rm NN}} = 2.76$ and 5.02 TeV using a physics model without nucleon substructure [49].

This is perhaps the single largest difference between our work and the conclusions of recent saturation-based calculations which constrained the event-by-event fluctuations of the proton using a color dipole picture of vector meson production [38, 39]. Those studies find that the measured coherent and incoherent J/Ψ spectra at HERA prefer a compact gluon distribution inside each nucleon, with a Gaussian width $w_g \approx 0.4$ fm which is roughly *half* the Gaussian width preferred by our analysis. Evidently, it may be necessary to place an informative prior on our nucleon substructure parameters in order to resolve the apparent tension between our parameter values and those needed to describe DIS measurements at HERA.

Additional constraints on the proton size and shape are provided by the proton-proton inelasticity density $G_{\rm inel}(b) = d^2\sigma_{\rm inel}(b)/d^2b$, which measures the proton-proton inelastic collision probability as a function of its impact parameter. We compare in Fig 10 the normalized proton-proton inelasticity density $G_{\rm inel}(b)/G_{\rm inel}(0)$ at $\sqrt{s_{\rm NN}} = 5.02$ TeV predicted by T_RENTo using the present study's posterior median parameters with a calculation that extracted the proton-proton inelasticity density using a parametrization fit to proton-proton differential scattering data at $\sqrt{s_{\rm NN}} = 7$ TeV [37]. Our Bayesian median estimate (orange line) prefers a larger width for the proton-proton normalized inelasticity density compared to that of reference [37], and this trend is *opposite* what one would expect based on the difference in beam energy alone. This suggests that our posterior estimate for the proton radius is somewhat oversized. Nevertheless, it is fascinating that the present Bayesian estimate is as close as it is, given that the model is fit to quantities which are not typically used to extract the proton-proton inelasticity density.

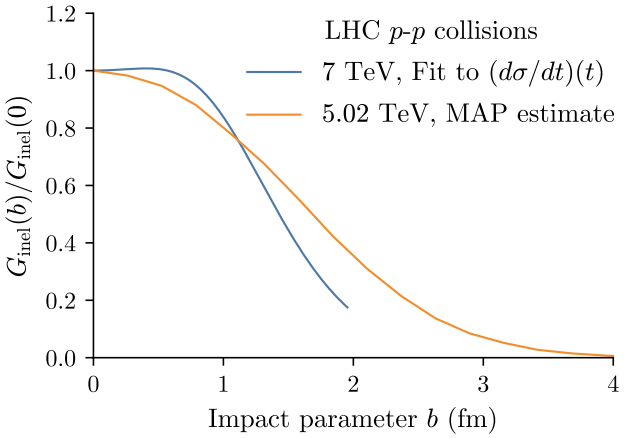

FIG. 10 Estimates of the normalized proton-proton inelasticity density $G_{\rm inel}(b) = d^2\sigma_{\rm inel}/d^2b$ at two different LHC beam energies. Blue line: Inelasticity density obtained by fitting a parametric form to LHC data at $\sqrt{s_{\rm NN}} = 7$ TeV [37]. Orange line: T_RENTo predictions at $\sqrt{s_{\rm NN}} = 5.02$ TeV using median parameters from the present study's Bayesian posterior.

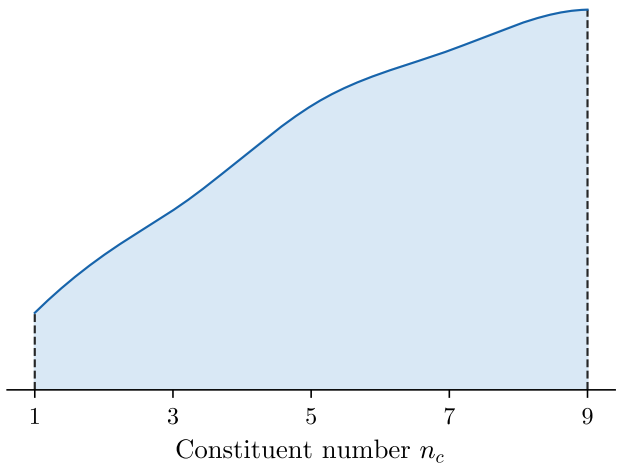

FIG. 11 Posterior distribution for the number of nucleon constituents n_c determined by the analysis. The parameter n_c is an integer (discrete) variable at every design point, but the emulator interpolation produces a posterior distribution which is continuous.

Moving on, we direct our attention to the constituent number n_c, shown enlarged in Fig. 11. The distribution is not sharply peaked, and hence we refrain from quoting a distribution median and 90% credible interval. Note, however, that the posterior clearly favors $n_c > 1$ constituents. This is not surprising. The T_RENTo model mimics saturation-based initial condition models [45], and saturation models tend to produce proton-sized fireballs in p-Pb collisions [86]. When the proton is spherically symmetric, the resulting QGP is also largely symmetric and thus produces very little anisotropic flow. Saturation-based models are therefore unable to describe the significant anisotropic flow measured in central p-Pb collisions without nucleon substructure, or alternatively, some other source of additional correlations [87].

B. Transport properties

In this section, we compare several of our posterior estimates to those obtained from a similar Bayesian analysis in Ref. [49] which used an (almost) identical version of the present physics model. The only modeling difference is the inclusion of nucleon substructure in the present

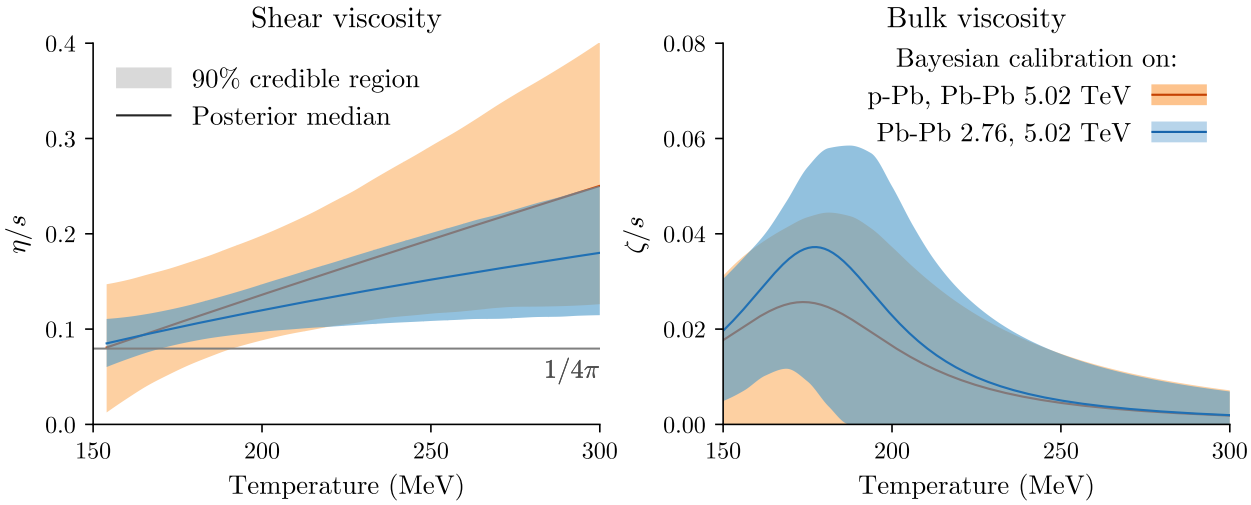

FIG. 12 Left figure: estimated temperature dependence of the QGP specific shear viscosity $(\eta/s)(T)$ determined by the present Bayesian analysis of p-Pb and Pb-Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV (orange line/band) compared to a previous Bayesian analysis of Pb-Pb collisions at $\sqrt{s_{\rm NN}} = 2.76$ and 5.02 TeV (blue line/band) [49]. The lines are the medians of each posterior distribution, and the bands are their 90% credible regions. Right figure: same as before, but for the temperature dependence of the QGP specific bulk viscosity $(\zeta/s)(T)$.

study which was absent in Ref. [49]. Several calibration details, however, are different between the two analyses. This work used a modest number of p-Pb and Pb-Pb observables at $\sqrt{s_{\rm NN}} = 5.02$ TeV (limited by availability), whereas Ref. [49] calibrated on a much larger number of Pb-Pb observables at $\sqrt{s_{\rm NN}} = 2.76$ and 5.02 TeV.

The posterior free streaming time $\tau_{\rm fs} = 0.48^{+0.55}_{-0.38}$ fm/c obtained in this work is significantly smaller than the previous estimate $\tau_{\rm fs} = 1.16^{+0.29}_{-0.25}$ fm/c quoted in Ref. [49]. We point out that the present study is missing several important observables which could affect the estimated free streaming time, e.g. the Pb-Pb mean p_T and mean p_T fluctuations at $\sqrt{s_{\rm NN}} = 5.02$ TeV. Nevertheless, it appears that the inclusion of nucleon substructure significantly reduces the maximum allowed free streaming time, although more work is needed to establish if this is indeed the case.

We also compare in Fig. 12 our estimates for the temperature dependence of the QGP specific shear viscosity $(\eta/s)(T)$ and bulk viscosity $(\zeta/s)(T)$ with those of Ref. [49]. The lines are the distribution medians, and the bands are their 90% credible regions. The results of this work are shown in orange, and the results of Ref. [49] are shown in blue. In general, our estimates are broader and less certain but otherwise self-consistent. Evidently, the combined analysis of Pb-Pb data at $\sqrt{s_{\rm NN}} = 2.76$ and 5.02 TeV in Ref. [49] provides a better constraint on the QGP viscosities which is not surprising given the additional observables and multiple beam energies studied. The p-Pb data, meanwhile, does not appear to provide strong viscous constraints.

C. Verification of high-probability parameters

We verified the emulator and tested the accuracy of our physics model framework using a single set of high-probability parameters selected from the Bayesian posterior. These parameters, listed in Table IV, are the approximate "best fit" values of the calibrated model, commonly referred to as the *maximum a posteriori* (MAP) estimate

$$\mathbf{x}_{\rm MAP} \equiv \arg\max_{\mathbf{x}} P(\mathbf{x}|E). \tag{56}$$

We then ran $\mathcal{O}(10^6)$ minimum-bias and multiplicity triggered events using the MAP estimate $\mathbf{x}_{\rm MAP}$ and computed all of the model observables listed in Sec. III A. The resulting model calculations are shown in Fig. 13 alongside experimental data from CMS [76] and ALICE [73–75, 77]. The left and right columns show the results for the p-Pb and Pb-Pb collision systems respectively, and each row shows a different group of related observables.

The global agreement of the MAP model calculations with the experimental data is generally quite good. The largest tension is observed in the two-particle cumulants $v_2\{2\}$ and $v_3\{2\}$ of the p-Pb system, although even that tension is only about 10–15%. Quite remarkably, the model perfectly describes the shape of the p-Pb and Pb-Pb two-particle correlations which is strong evidence that these correlations are hydrodynamic in origin.

Moreover, we obtain an excellent description of the p-Pb mean p_T, although this fit is somewhat less meaningful since we are unable to calibrate on the Pb-Pb mean p_T simultaneously (data is not yet available). Additionally, the model provides a simultaneous description of the

TABLE IV *Maximum a posteriori* (MAP) parameters determined from the posterior distribution and used to generate Fig. 13. The posterior distribution on the particlization temperature $T_{\rm switch}$ is flat (agnostic), so we fix it's value using Ref. [49].

Initial condition / Pre-eq		QGP medium	
Norm	20. GeV	$(\eta/s)_{\rm min}$	0.11
p	0.0	$(\eta/s)_{\rm slope}$	1.6 GeV^{-1}
k	0.19	$(\eta/s)_{\rm crv}$	-0.29
n_c	6	$(\zeta/s)_{\rm max}$	0.032
r_{cp}	0.81 fm	$(\zeta/s)_{\rm width}$	0.024 GeV
w_c	0.43 fm	$(\zeta/s)_{T_{\rm peak}}$	0.175 GeV
$d_{\rm min}$	0.81 fm	$T_{\rm switch}$	0.151 GeV
$\tau_{\rm fs}$	0.37 fm/c		

p-Pb and Pb-Pb charged-particle yields using a single energy deposition parameter $p = 0$. This is the *exact same* generalized mean p-value supported by multiple previous studies [45, 46, 49, 50]. Evidently, this scaling continues to hold for initial conditions with sizable nucleon substructure.

We also present calculations for several observables which were omitted from the calibration due to missing experimental data and the statistical limitations of our training data. Here our MAP event sample is several orders of magnitude larger so the statistics are no issue. The bottom-right panel of Fig. 13 shows our model calculation for the Pb-Pb four-particle elliptic flow cumulant $v_2\{4\}$ along with the measured data points from ALICE [75]. We see that the MAP estimate nicely describes the measured $v_2\{4\}$ data which is encouraging since this particular observable was never used to calibrate the model.

The relative mean p_T fluctuation $\delta p_T/\langle p_T\rangle$ is another important bulk observable to test the predictions of the calibrated model. It measures the *dynamical* component of event-by-event mean p_T fluctuations, quantified by the two-particle correlator

$$(\delta p_T)^2 = \langle\langle (p_{T,i} - \langle p_T\rangle)(p_{T,j} - \langle p_T\rangle)\rangle\rangle. \tag{57}$$

The inner-average in Eq. (57) runs over all pairs of particles i, j in the same event, the outer average runs over all events in a given bin (centrality or multiplicity), and the symbol $\langle p_T\rangle$ denotes the usual mean transverse momentum of particles in the bin. The observable is typically presented in terms of the dimensionless ratio $\delta p_T/\langle p_T\rangle$, which quantifies the strength of the dynamical fluctuations in units of the average transverse momentum $\langle p_T\rangle$.

We show the MAP estimate predictions for the p-Pb and Pb-Pb relative mean p_T fluctuations $\delta p_T/\langle p_T\rangle$ at $\sqrt{s_{\rm NN}} = 5.02$ TeV in the third row of Fig. 13. For the Pb-Pb system, we use centrality bins and for the p-Pb system we use the same relative multiplicity bins used for the p-Pb charged-particle mean p_T. The relative mean p_T fluctuations have been shown to be particularly sensitive to the existence of nucleon substructure [88], and thus it would be interesting to ultimately include this observable in the calibration when the data becomes available.

Lastly, we compute the symmetric cumulants ${\rm SC}(m,n)$ for the Pb-Pb collision system at $\sqrt{s_{\rm NN}} = 5.02$ TeV which quantify correlations between event-by-event fluctuations of the flow harmonics of different order [89, 90]

$$\begin{aligned}{\rm SC}(m,n) &= \langle\langle\cos[m(\phi_1-\phi_3)+n(\phi_2-\phi_4)]\rangle\rangle \\ &\quad - \langle\langle\cos[m(\phi_1-\phi_2)]\rangle\rangle\langle\langle\cos[n(\phi_1-\phi_2)]\rangle\rangle \\ &\approx \langle v_m^2 v_n^2\rangle - \langle v_m^2\rangle\langle v_n^2\rangle. \end{aligned}\tag{58}$$

We show these model predictions in Fig. 14 along with the *normalized* symmetric cumulants

$${\rm NSC}(m,n) = {\rm SC}(m,n)/\langle v_m^2\rangle\langle v_n^2\rangle, \tag{59}$$

which are expected to be less sensitive to the medium response and more sensitive to the properties of the initial state. The solid lines are the MAP estimate of the present study, and the dashed lines are the MAP estimate of Ref. [49] which did not include nucleon substructure and was calibrated on Pb-Pb collisions at $\sqrt{s_{\rm NN}} = 2.76$ and 5.02 TeV. We observe that the gap between ${\rm SC}(3,2)$ and ${\rm SC}(4,2)$ is generally wider in the present analysis than in Ref. [49], as is the gap between the normalized symmetric cumulants ${\rm NSC}(3,2)$ and ${\rm NSC}(4,2)$.

We emphasize that multiple aspects of the two analyses are different such as the collision systems and beam energies considered, the observables which were included in each calibration, and the existence of nucleon substructure in the model. Thus we can only speculate what might have caused the large difference in the MAP estimate for the symmetric flow cumulants. Two reasonable culprits would be the inclusion of nucleon substructure and the large difference in the preferred pre-equilibrium free streaming time determined by the two studies.

V. SUMMARY AND CONCLUSIONS

Relativistic heavy-ion collisions produce long-range multiparticle correlations which are commonly explained by the existence of hydrodynamic flow [9]. This narrative is evidenced by the global, self-consistent, and highly non-trivial quantitative agreement of hydrodynamic models with a large number of heavy-ion bulk observables [45, 91, 92]. Naturally, such descriptions rely on the validity of hydrodynamic approximations, and these approximations begin to break down in the so-called dilute limit where discrete particle degrees-of-freedom dominate and continuous field descriptions of the medium cease to make sense. Tell-tale signatures of hydrodynamic collectivity were thus generally expected to vanish in smaller nuclear collision systems, e.g. p-p and p-Pb collisions, where the number of produced particles is orders of magnitude smaller than a typical Pb-Pb collision.

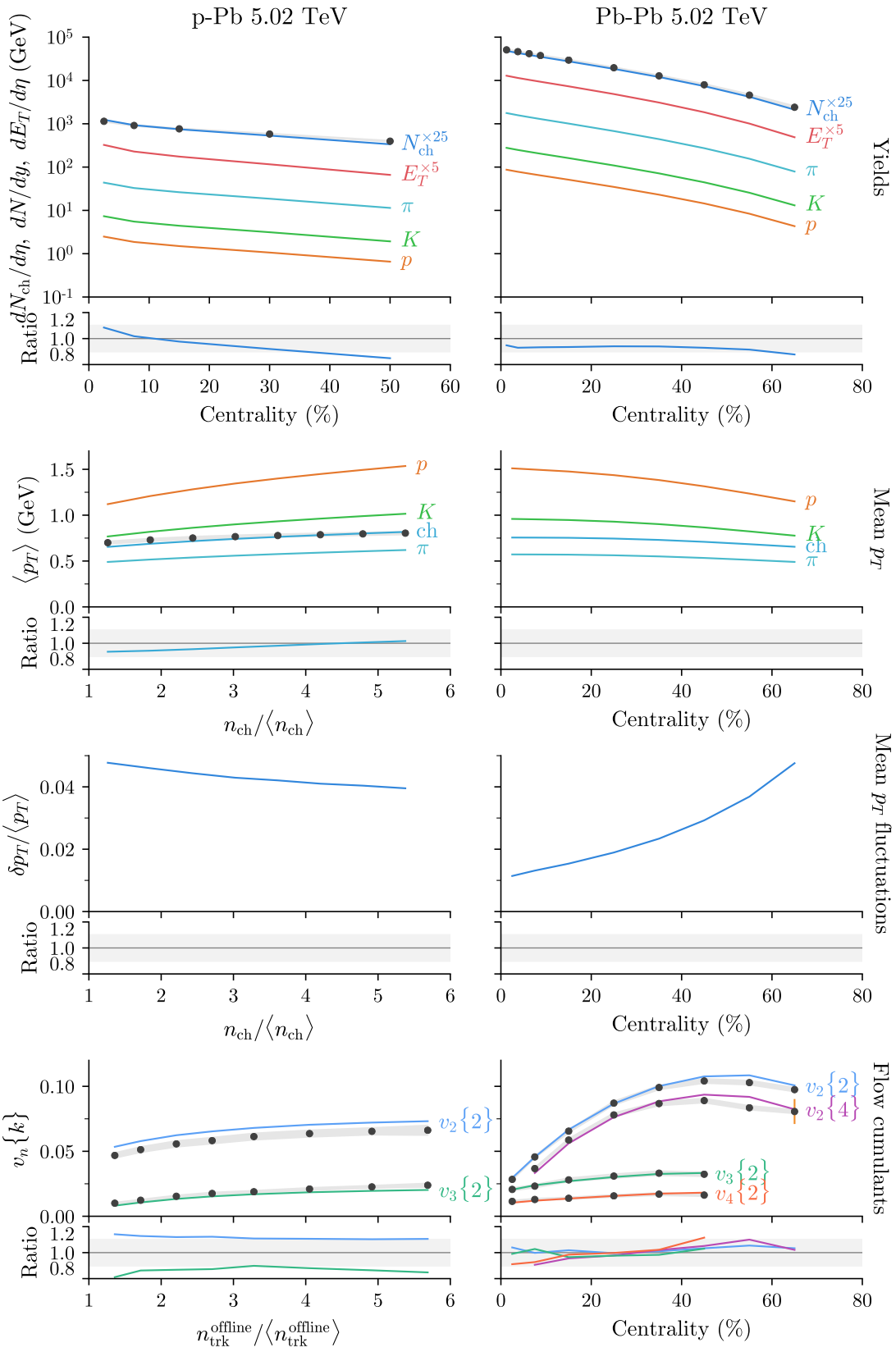

FIG. 13 Model calculations using the *maximum a posterior* (MAP) parameters compared to experiment. Colored lines are model calculations for p-Pb collisions (left) and Pb-Pb collisions (right) at $\sqrt{s_{\rm NN}} = 5.02$ TeV. Points with error bars are the experimental data with statistical uncertainties, and gray bands their corresponding systematic uncertainties, from CMS [76] and ALICE [73–75, 77]. The sub-axes show the ratio of model over data where available with gray bands indicating ±10%.

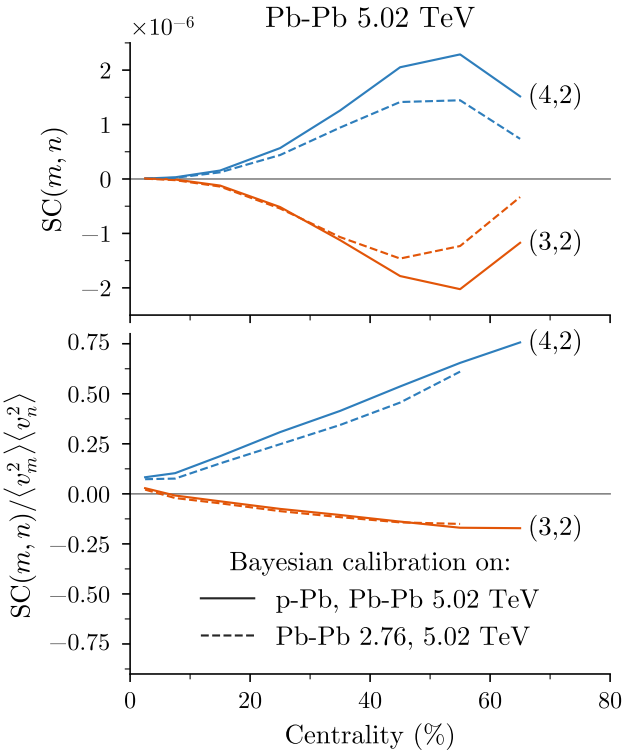

FIG. 14 Model calculations of the symmetric cumulants (top) and normalized symmetric cumulants (bottom) for Pb-Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV using the *maximum a posteriori* (MAP) parameters. The solid lines are the MAP estimate of the present analysis (with nucleon substructure), and the dashed lines are the MAP estimate of Ref. [49] (without nucleon substructure) which was calibrated on Pb-Pb observables at $\sqrt{s_{\rm NN}} = 2.76$ and 5.02 TeV. In general, most model parameters are somewhat different between the two studies.

These expectations were upended, however, when long-range multiparticle correlations were detected in high-multiplicity p-Pb collisions and found to be similar in magnitude to those observed in Pb-Pb collisions [1–3]. Nuclear collision systems which were previously thought to be too small for hydrodynamic flow, were subsequently found to generate the same collectivity used to justify hydrodynamic flow in heavy-ion collisions. It is thus natural to wonder if a single unified hydrodynamic framework might be able to describe p-Pb and Pb-Pb bulk observables simultaneously.

In this work, we performed a semi-exhaustive search for a unified description of p-Pb and Pb-Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV using Bayesian methods to rigorously calibrate and constrain free parameters of a flexible nuclear collision model based on viscous hydrodynamics. The goal of our study was two fold. First, we aimed to establish whether or not our hydrodynamic framework was able to describe both collision systems simultaneously. And second, in the event that the former was true, we wished to obtain estimates for the *true* parameters of our model given the assumptions of our framework and the evidence provided by the model predictions and the experimental data.

We built, for this purpose, a flexible multi-stage nuclear collision model characterized by a number of free parameters which vary theoretically uncertain aspects of the framework such as the QGP initial conditions and hydrodynamic transport properties. For the QGP initial conditions, we employed a modified version of the T_RENTo model [46] which adds new parameters to vary the fluctuating size and shape of each nucleon. Specifically, we modeled each nucleon as a cluster of n_c constituents (hot spots), where each constituent is described by a Gaussian density of width w_c. The constituent positions were each sampled randomly (without correlations) from a Gaussian radial distribution of width r_{cp} centered about each pre-defined nucleon position.

The transport dynamics of the collision were simulated using a pre-equilibrium free streaming stage followed by boost-invariant viscous hydrodynamics for hot and dense regions of the fireball and a microscopic hadronic afterburner for the relatively dilute corona. We parametrized various sources of uncertainty in each stage of the collision including the duration of the pre-equilibrium free streaming stage, the temperature dependence of the QGP shear and bulk viscosities, and the particlization temperature used to switch from a hydrodynamic description to microscopic Boltzmann transport.

With the full evolution model in hand, we applied Bayesian methods which were developed to estimate the parameters of computationally intensive models [78–80]. We first constructed a scaffolding of $n = 500$ parameters points distributed throughout our 15-dimensional parameter space and evaluated the nuclear collision model using $\mathcal{O}(10^4)$ events at each parameter point. The ensemble of events was then used to calculate a large number of experimental observables at each design point and train Gaussian process emulators to interpolate the model predictions as a function of the input parameters. Finally, we used Markov chain Monte Carlo (MCMC) importance sampling to explore the parameter space and draw samples from the Bayesian posterior distribution for the *true* values of the model parameters, given our emulated model predictions, the experimental data, and their associated uncertainties. The model calibration process is summarized by Figs. 6 and 7, and the resulting posterior distribution for the model input parameters is shown in Fig. 8. We also evaluated the model predictions using a single set of high-probability parameters in Fig. 13. With these results, we are able to address the two primary goals of the study.

First, we demonstrated in Fig. 13 the existence of a single set of model parameters which can simultaneously describe p-Pb and Pb-Pb charged-particle yields, mean p_T, and flow cumulants at $\sqrt{s_{\rm NN}} = 5.02$ TeV. The excellent quantitative agreement of the model calculation with the experimental data is strong evidence for a unified hydrodynamic description of p-Pb and Pb-Pb collisions at ultrarelativistic energies. Moreover, the modifications

to the physics model which were required to obtain this agreement are generally modest; one must simply replace Gaussian nucleons with composite nucleons of several or more constituents.

Second, we obtained a posterior distribution for the model input parameters in Fig. 8 and reported quantitative estimates for these parameters in Table III. Here we summarize our key findings about the model:

1. Using TRENTo initial conditions followed by a pre-equilibrium free streaming stage, we find that the initially produced energy density scales like the geometric mean of participant nuclear thickness; see Eq. (54). Previous Bayesian studies have reported similar scaling for the initially produced *entropy* density [45, 50], although these studies assumed static initialization, an approximation which neglects the initial values of u^μ, $\pi^{\mu\nu}$ and Π at the hydrodynamic starting time.
2. We find that nucleon substructure is necessary to simultaneously describe p-Pb and Pb-Pb bulk observables. However, we observe no strong preference for a specific number of constituents n_c (hot spots) inside the nucleon. In particular, we find no evidence to support the specific number $n_c = 3$ which is commonly used in the literature.
3. The present Bayesian analysis prefers larger nucleons, $w \approx 1$ fm, in agreement with a similar Bayesian analysis calibrated to Pb-Pb data at $\sqrt{s_{\rm NN}} = 2.76$ and 5.02 TeV using a hybrid model without nucleon substructure [49]. We note that our result is in significant tension with an estimate for the effective nucleon width based on the gluon distribution extracted from HERA data [93]. Our model calculations also predict a broader proton-proton inelasticity density at $\sqrt{s_{\rm NN}} = 5.02$ TeV than supported by the data [37]. This suggests that our reconstructed protons are somewhat oversized.
4. We obtain an estimate $w_c = 0.53^{+0.28}_{-0.20}$ fm for the Gaussian width of the constituent hot spots inside each nucleon. This width is considerably larger than the length scales typically associated with nucleon substructure. However, it is natural to wonder if this estimate is oversized, given the aforementioned comments regarding our extracted nucleon width.
5. We report an estimate $\tau_{\rm fs} = 0.48^{+0.55}_{-0.38}$ fm/c for the model's pre-equilibrium free streaming time. This is significantly shorter than the estimate obtained from a similar Bayesian study in Ref. [49], which found $\tau_{\rm fs} = 1.16^{+0.29}_{-0.25}$ fm/c. It is not clear whether the difference is a result of nucleon substructure or the different observables used to calibrate each analysis.
6. We compare in Fig. 12 our estimate for the temperature dependence of the QGP specific shear and bulk viscosities to those of Ref. [49], which performed a Bayesian calibration to Pb-Pb bulk observables at $\sqrt{s_{\rm NN}} = 2.76$ and 5.02 TeV using a physics model without nucleon substructure. The two studies are in good agreement, although Ref. [49] obtains a more precise estimate for $(\eta/s)(T)$, likely due to the additional beam energies and observables included and an enhanced sensitivity of larger collision systems to the QGP viscosity.
7. We make predictions in Figs. 13 and 14 for several quantities which were not included in the model calibration, including the identified yields, transverse energy, symmetric cumulants, and mean p_T fluctuations at $\sqrt{s_{\rm NN}} = 5.02$ TeV. Interestingly, our MAP estimate for the Pb-Pb symmetric cumulants at $\sqrt{s_{\rm NN}} = 5.02$ TeV are significantly different than those estimated in Ref. [49]. This could be a direct (or indirect) result of including nucleon substructure in the model calibration.

The present study would benefit from a number of improvements. Perhaps the most obvious target for improvement is the absence of several important experimental datasets. Specifically, we are missing the transverse energy, identified particle yields, and the mean p_T fluctuations of both collision systems, as well the charged-particle mean p_T for the Pb-Pb system. These observables would certainly influence the quality of the combined fit and correspondingly our estimates for the model parameters.

Similarly, the results would greatly benefit from additional beam energies and collision systems. For example, future studies should incorporate proton-proton collisions at top LHC energies, as well as the numerous collisions systems studied at top RHIC energy. Such data would undoubtedly improve constraints on the model parameters and would enable more stringent tests of the calibrated model predictions. The RHIC data may also help elucidate the beam-energy dependence of the model parameters which would be worth investigating. We leave these improvements for future studies.

All software used in this work is open source:

- TRENTo with nucleon substructure (C++) [94]
- Pre-equilibrium free streaming (Python) [95]
- VISH2+1 hydrodynamics (Fortran) [96]
- FRZOUT particle sampler (Python) [97]
- UrQMD microscopic transport model (Fortran) [98]
- DukeQCD event generator wrapper (Python) [99]
- Bayesian parameter estimation (Python) [100]

ACKNOWLEDGMENTS

JSM thanks Ulrich Heinz, Jamie Nagle, Berndt Müller, Weiyao Ke, Björn Schenke, and Heikki Mäntysaari for

helpful discussions and clarifying comments. This research was completed using 6 million CPU hours provided by the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231. JSM and SAB are supported by the U.S. Department of Energy Grant No. DE-FG02-05ER41367 and JEB by NSF Grant No. NSF-ACI-1550225.

[1] S. Chatrchyan *et al.* (CMS), Phys. Lett. **B718**, 795 (2013), arXiv:1210.5482 [nucl-ex].
[2] B. Abelev *et al.* (ALICE), Phys. Lett. **B719**, 29 (2013), arXiv:1212.2001 [nucl-ex].
[3] G. Aad *et al.* (ATLAS), Phys. Rev. Lett. **110**, 182302 (2013), arXiv:1212.5198 [hep-ex].
[4] A. Adare *et al.* (PHENIX), Phys. Rev. Lett. **115**, 142301 (2015), arXiv:1507.06273 [nucl-ex].
[5] P. Bozek, Phys. Rev. **C85**, 014911 (2012), arXiv:1112.0915 [hep-ph].
[6] P. Bozek and W. Broniowski, Phys. Rev. **C88**, 014903 (2013), arXiv:1304.3044 [nucl-th].
[7] B. Schenke and R. Venugopalan, Phys. Rev. Lett. **113**, 102301 (2014), arXiv:1405.3605 [nucl-th].
[8] H. Niemi, *Proceedings, 24th International Conference on Ultra-Relativistic Nucleus-Nucleus Collisions (Quark Matter 2014): Darmstadt, Germany, May 19-24, 2014*, Nucl. Phys. **A931**, 227 (2014).
[9] R. Derradi de Souza, T. Koide, and T. Kodama, Prog. Part. Nucl. Phys. **86**, 35 (2016), arXiv:1506.03863 [nucl-th].
[10] J.-Y. Ollitrault and F. G. Gardim, *Proceedings, 23rd International Conference on Ultrarelativistic Nucleus-Nucleus Collisions : Quark Matter 2012 (QM 2012): Washington, DC, USA, August 13-18, 2012*, Nucl. Phys. **A904-905**, 75c (2013), arXiv:1210.8345 [nucl-th].
[11] H. Song, *Proceedings, 23rd International Conference on Ultrarelativistic Nucleus-Nucleus Collisions : Quark Matter 2012 (QM 2012): Washington, DC, USA, August 13-18, 2012*, Nucl. Phys. **A904-905**, 114c (2013), arXiv:1210.5778 [nucl-th].
[12] A. Adare *et al.* (PHENIX), Phys. Rev. **C93**, 024901 (2016), arXiv:1509.06727 [nucl-ex].
[13] B. Schenke, P. Tribedy, and R. Venugopalan, Phys. Rev. **C89**, 064908 (2014), arXiv:1403.2232 [nucl-th].
[14] C. Aidala *et al.* (PHENIX), (2018), arXiv:1805.02973 [nucl-ex].
[15] A. Adare *et al.* (PHENIX), Phys. Rev. **C97**, 064904 (2018), arXiv:1710.09736 [nucl-ex].
[16] L. Adamczyk *et al.* (STAR), Phys. Rev. Lett. **115**, 222301 (2015), arXiv:1505.07812 [nucl-ex].
[17] C. Shen, J.-F. Paquet, G. S. Denicol, S. Jeon, and C. Gale, Phys. Rev. **C95**, 014906 (2017), arXiv:1609.02590 [nucl-th].
[18] C. Aidala *et al.* (PHENIX), Phys. Rev. Lett. **120**, 062302 (2018), arXiv:1707.06108 [nucl-ex].
[19] A. Adare *et al.* (PHENIX), Phys. Rev. Lett. **98**, 162301 (2007), arXiv:nucl-ex/0608033 [nucl-ex].
[20] P. Moller, J. Nix, W. Myers, and W. Swiatecki, Atomic Data and Nuclear Data Tables **59**, 185 (1995).
[21] H. D. Vries, C. D. Jager, and C. D. Vries, Atomic Data and Nuclear Data Tables **36**, 495 (1987).
[22] H. Song, S. A. Bass, U. Heinz, T. Hirano, and C. Shen, Phys. Rev. **C83**, 054910 (2011), [Erratum: Phys. Rev.C86,059903(2012)], arXiv:1101.4638 [nucl-th].
[23] E. Retinskaya, M. Luzum, and J.-Y. Ollitrault, Phys. Rev. **C89**, 014902 (2014), arXiv:1311.5339 [nucl-th].
[24] J. Liu, C. Shen, and U. Heinz, Phys. Rev. **C91**, 064906 (2015), [Erratum: Phys. Rev.C92,no.4,049904(2015)], arXiv:1504.02160 [nucl-th].
[25] A. Kurkela, *Proceedings, 25th International Conference on Ultra-Relativistic Nucleus-Nucleus Collisions (Quark Matter 2015): Kobe, Japan, September 27-October 3, 2015*, Nucl. Phys. **A956**, 136 (2016), arXiv:1601.03283 [hep-ph].
[26] K. Welsh, J. Singer, and U. W. Heinz, Phys. Rev. **C94**, 024919 (2016), arXiv:1605.09418 [nucl-th].
[27] J. S. Moreland, J. E. Bernhard, W. Ke, and S. A. Bass, *Proceedings, 26th International Conference on Ultra-relativistic Nucleus-Nucleus Collisions (Quark Matter 2017): Chicago, Illinois, USA, February 5-11, 2017*, Nucl. Phys. **A967**, 361 (2017), arXiv:1704.04486 [nucl-th].
[28] B. Schenke and R. Venugopalan, *Proceedings, 24th International Conference on Ultra-Relativistic Nucleus-Nucleus Collisions (Quark Matter 2014): Darmstadt, Germany, May 19-24, 2014*, Nucl. Phys. **A931**, 1039 (2014), arXiv:1407.7557 [nucl-th].
[29] S. Schlichting and B. Schenke, Phys. Lett. **B739**, 313 (2014), arXiv:1407.8458 [hep-ph].
[30] S. S. Adler *et al.* (PHENIX), Phys. Rev. **C89**, 044905 (2014), arXiv:1312.6676 [nucl-ex].
[31] J. T. Mitchell, D. V. Perepelitsa, M. J. Tannenbaum, and P. W. Stankus, Phys. Rev. **C93**, 054910 (2016), arXiv:1603.08836 [nucl-ex].
[32] W. Broniowski, P. Bozek, and M. Rybczynski, *Proceedings, 10th International Workshop on Critical Point and Onset of Deconfinement (CPOD 2016): Wrocław, Poland*, Acta Phys. Polon. Supp. **10**, 513 (2017), arXiv:1611.00250 [nucl-th].
[33] P. Bozek, W. Broniowski, and S. Chatterjee, *Proceedings, 9th Workshop "Excited QCD" 2017: Sintra, Portugal, May 7-13, 2017*, Acta Phys. Polon. Supp. **10**, 1091 (2017), arXiv:1707.04420 [nucl-th].
[34] G. Antchev *et al.* (TOTEM), EPL **95**, 41001 (2011), arXiv:1110.1385 [hep-ex].
[35] E. R. Arriola and W. Broniowski, Few-Body Systems **57**, 485 (2016).
[36] J. L. Albacete, H. Petersen, and A. Soto-Ontoso, Phys. Rev. **C95**, 064909 (2017), arXiv:1612.06274 [hep-ph].
[37] J. L. Albacete and A. Soto-Ontoso, Phys. Lett. **B770**, 149 (2017), arXiv:1605.09176 [hep-ph].
[38] H. Mantysaari and B. Schenke, Phys. Rev. Lett. **117**, 052301 (2016), arXiv:1603.04349 [hep-ph].
[39] H. Mantysaari and B. Schenke, Phys. Rev. **D94**, 034042 (2016), arXiv:1607.01711 [hep-ph].
[40] F. D. Aaron *et al.* (ZEUS, H1), JHEP **01**, 109 (2010), arXiv:0911.0884 [hep-ex].

[41] H. Abramowicz *et al.* (ZEUS, H1), Eur. Phys. J. **C75**, 580 (2015), arXiv:1506.06042 [hep-ex].
[42] B. Schenke, P. Tribedy, and R. Venugopalan, Phys. Rev. Lett. **108**, 252301 (2012), arXiv:1202.6646 [nucl-th].
[43] B. Schenke, C. Shen, and P. Tribedy, in *27th International Conference on Ultrarelativistic Nucleus-Nucleus Collisions (Quark Matter 2018) Venice, Italy, May 14-19, 2018* (2018) arXiv:1807.05205 [nucl-th].
[44] C. Shen, Z. Qiu, H. Song, J. Bernhard, S. Bass, and U. Heinz, Comput. Phys. Commun. **19**, 61 (2016), arXiv:1409.8164 [nucl-th].
[45] J. E. Bernhard, J. S. Moreland, S. A. Bass, J. Liu, and U. Heinz, Phys. Rev. **C94**, 024907 (2016), arXiv:1605.03954 [nucl-th].
[46] J. S. Moreland, J. E. Bernhard, and S. A. Bass, Phys. Rev. **C92**, 011901(R) (2015), arXiv:1412.4708 [nucl-th].
[47] W. Broniowski, W. Florkowski, M. Chojnacki, and A. Kisiel, Phys. Rev. **C80**, 034902 (2009), arXiv:0812.3393 [nucl-th].
[48] U. W. Heinz and J. Liu, *Proceedings, 25th International Conference on Ultra-Relativistic Nucleus-Nucleus Collisions (Quark Matter 2015): Kobe, Japan, September 27-October 3, 2015*, Nucl. Phys. **A956**, 549 (2016), arXiv:1512.08276 [nucl-th].
[49] J. E. Bernhard, *Bayesian parameter estimation for relativistic heavy-ion collisions*, Ph.D. thesis, Duke U. (2018-04-19), arXiv:1804.06469 [nucl-th].
[50] W. Ke, J. S. Moreland, J. E. Bernhard, and S. A. Bass, Phys. Rev. **C96**, 044912 (2017), arXiv:1610.08490 [nucl-th].
[51] P. Bozek, W. Broniowski, and J. Moreira, Phys. Rev. **C83**, 034911 (2011), arXiv:1011.3354 [nucl-th].
[52] B. Abelev *et al.* (ALICE), Phys. Rev. Lett. **110**, 032301 (2013), arXiv:1210.3615 [nucl-ex].
[53] V.V. Anisovich and Yu.M. Shabelsky and V.M. Shekhter, Nuclear Physics B **133**, 477 (1978).
[54] A. Bialas, M. Bleszynski, and W. Czyz, Nucl. Phys. **B111**, 461 (1976).
[55] D. Kharzeev and M. Nardi, Phys. Lett. **B507**, 121 (2001), arXiv:nucl-th/0012025 [nucl-th].
[56] W. Israel and J. M. Stewart, Annals Phys. **118**, 341 (1979).
[57] W. Israel and J. Stewart, Phys. Lett. **A58**, 213 (1976).
[58] G. S. Denicol, H. Niemi, E. Molnar, and D. H. Rischke, Phys. Rev. **D85**, 114047 (2012), [Erratum: Phys. Rev.D91,no.3,039902(2015)], arXiv:1202.4551 [nucl-th].
[59] G. S. Denicol, T. Koide, and D. H. Rischke, Phys. Rev. Lett. **105**, 162501 (2010), arXiv:1004.5013 [nucl-th].
[60] G. S. Denicol, S. Jeon, and C. Gale, Phys. Rev. **C90**, 024912 (2014), arXiv:1403.0962 [nucl-th].
[61] A. Bazavov, T. Bhattacharya, C. DeTar, H.-T. Ding, S. Gottlieb, R. Gupta, P. Hegde, U. M. Heller, F. Karsch, E. Laermann, L. Levkova, S. Mukherjee, P. Petreczky, C. Schmidt, C. Schroeder, R. A. Soltz, W. Soeldner, R. Sugar, M. Wagner, and P. Vranas (HotQCD Collaboration), Phys. Rev. D **90**, 094503 (2014).
[62] S. Borsanyi *et al.*, Nature **539**, 69 (2016), arXiv:1606.07494 [hep-lat].
[63] J. Noronha-Hostler and C. Ratti, (2018), arXiv:1804.10661 [nucl-th].
[64] H. Song and U. W. Heinz, Phys. Rev. **C77**, 064901 (2008), arXiv:0712.3715 [nucl-th].
[65] F. Cooper and G. Frye, Phys. Rev. D **10**, 186 (1974).
[66] J. Sollfrank, P. Koch, and U. W. Heinz, Z. Phys. **C52**, 593 (1991).
[67] P. Huovinen, P. M. Lo, M. Marczenko, K. Morita, K. Redlich, and C. Sasaki, Phys. Lett. **B769**, 509 (2017), arXiv:1608.06817 [hep-ph].
[68] V. Vovchenko, M. I. Gorenstein, and H. Stoecker, (2018), arXiv:1807.02079 [nucl-th].
[69] C. Patrignani and P. D. Group, Chinese Physics C **40**, 100001 (2016).
[70] S. Pratt and G. Torrieri, Phys. Rev. **C82**, 044901 (2010), arXiv:1003.0413 [nucl-th].
[71] S. A. Bass *et al.*, Prog. Part. Nucl. Phys. **41**, 255 (1998), [Prog. Part. Nucl. Phys.41,225(1998)], arXiv:nucl-th/9803035 [nucl-th].
[72] M. Bleicher *et al.*, J. Phys. **G25**, 1859 (1999), arXiv:hep-ph/9909407 [hep-ph].
[73] J. Adam *et al.* (ALICE), Phys. Rev. Lett. **116**, 222302 (2016), arXiv:1512.06104 [nucl-ex].
[74] J. Adam *et al.* (ALICE), Phys. Rev. **C91**, 064905 (2015), arXiv:1412.6828 [nucl-ex].
[75] J. Adam *et al.* (ALICE), Phys. Rev. Lett. **116**, 132302 (2016), arXiv:1602.01119 [nucl-ex].
[76] S. Chatrchyan *et al.* (CMS), Phys. Lett. **B724**, 213 (2013), arXiv:1305.0609 [nucl-ex].
[77] B. B. Abelev *et al.* (ALICE), Phys. Lett. **B727**, 371 (2013), arXiv:1307.1094 [nucl-ex].
[78] A. O'Hagan, Rel.Engin.Sys.Safety **91**, 1290 (2006).
[79] D. Higdon, J. Gattiker, B. Williams, and M. Rightley, J.Amer.Stat.Assoc. **103**, 570 (2008).
[80] D. Higdon, J. D. McDonnell, N. Schunck, J. Sarich, and S. M. Wild, J.Phys. **G42**, 034009 (2015), arXiv:1407.3017 [physics.data-an].
[81] M. D. Morris and T. J. Mitchell, J.Stat.Plan.Inf. **43**, 381 (1995).
[82] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* (MIT Press, Cambridge, MA, 2006).
[83] M. E. Tipping and C. M. Bishop, Neural Comput. **11**, 443 (1999).
[84] J. Goodman and J. Weare, Comm.App.Math.Comp.Sc. **5**, 65 (2010).
[85] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, PASP **125**, 306 (2013), arXiv:1202.3665 [astro-ph.IM].
[86] A. Bzdak, B. Schenke, P. Tribedy, and R. Venugopalan, Phys. Rev. **C87**, 064906 (2013), arXiv:1304.3403 [nucl-th].
[87] B. Schenke, *Proceedings, 26th International Conference on Ultra-relativistic Nucleus-Nucleus Collisions (Quark Matter 2017): Chicago, Illinois, USA, February 5-11, 2017*, Nucl. Phys. **A967**, 105 (2017), arXiv:1704.03914 [nucl-th].
[88] P. Bozek and W. Broniowski, Phys. Rev. **C96**, 014904 (2017), arXiv:1701.09105 [nucl-th].
[89] A. Bilandzic, C. H. Christensen, K. Gulbrandsen, A. Hansen, and Y. Zhou, Phys. Rev. **C89**, 064904 (2014), arXiv:1312.3572 [nucl-ex].
[90] J. Adam *et al.* (ALICE), Phys. Rev. Lett. **117**, 182301 (2016), arXiv:1604.07663 [nucl-ex].

[91] H. Niemi, K. J. Eskola, and R. Paatelainen, Phys. Rev. **C93**, 024907 (2016), arXiv:1505.02677 [hep-ph].
[92] C. Gale, S. Jeon, B. Schenke, P. Tribedy, and R. Venugopalan, Phys. Rev. Lett. **110**, 012302 (2013), arXiv:1209.6330 [nucl-th].
[93] A. H. Rezaeian, M. Siddikov, M. Van de Klundert, and R. Venugopalan, Phys. Rev. **D87**, 034002 (2013), arXiv:1212.2974 [hep-ph].
[94] J. S. Moreland, J. E. Bernhard, and S. A. Bass, "TRENTo initial condition model with nucleon substructure," https://github.com/morelandjs/trento-substructure (2016).
[95] J. E. Bernhard, "Freestreaming and Landau matching for boost-invariant hydrodynamic initial conditions," https://github.com/Duke-QCD/freestream (2015).
[96] H. Song, Z. Qiu, C. Shen, J. Liu, and U. Heinz, "The Ohio State University viscous hydrodynamics code for relativistic heavy-ion collisions," https://u.osu.edu/vishnu/ (2016), We use a slightly modified version of the hydro code by J. Bernhard, https://github.com/jbernhard/osu-hydro.
[97] J. E. Bernhard, "Particlization model (Cooper-Frye sampler) for relativistic heavy-ion collisions," https://github.com/Duke-QCD/frzout (2016).
[98] UrQMD collaboration, "Ultra-relativistic quantum molecular dynamics," http://urqmd.org (2015), UrQMD is a large, complex model written by many contributors. Please see the license file on the official website for more information. We use a slightly modified version of the software by J. Bernhard, https://github.com/jbernhard/urqmd-afterburner.
[99] J. S. Moreland and J. E. Bernhard, "Nuclear collision event generator," https://github.com/morelandjs/hic-eventgen (2017), This is a forked and modified version of original code written by J. Bernhard.
[100] J. S. Moreland and J. E. Bernhard, "Bayesian parameter estimation of nuclear collisions with nucleon substructure," https://github.com/morelandjs/hic-param-est-qm18 (2017), This is a forked and modified version of original code written by J. Bernhard.

Appendix A: Event-by-event grid resizing

The boost-invariant VISH2+1 hydrodynamics code used in this work [44, 64] runs on a Cartesian transverse grid specified by a maximum grid size $x_{\max}$ and grid step width dx which fix the transverse grid extent $-x_{\max} < x < x_{\max}$ and number of grid cells along each dimension $n_x = 2\,x_{\max}/dx$. In general, the maximum grid size $x_{\max}$ should be set large enough to contain the full spacetime evolution of the event. This means that the truncation of $T^{\mu\nu}$ at the boundaries of the grid should never modify the final-state observables. We enforce this requirement by finding an energy density cutoff $e_{\min}$ for which the matter $e < e_{\min}$ can be effectively discarded without significantly modifying the simulation observables. We then fix the maximum grid size $x_{\max}$ such that it fully encloses the isotherm $T = T(e_{\min})$ for the full lifetime of the fireball.

We find that we can quickly estimate the maximum transverse radius $R_{\max}$ of the spacetime hypersurface $T = T(e_{\min})$ by running the event on a coarse-grained spatial grid with one-third the spatial resolution we would otherwise require to resolve typical hydrodynamic observables such as mean p_T and flows. The simulation time of a

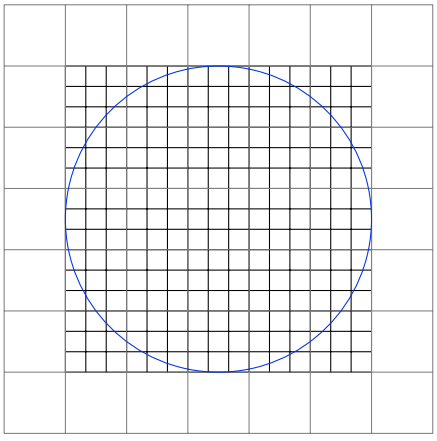

FIG. 15 Diagram of the adaptive grid resizing algorithm (not drawn to scale). Each initial condition event is first run on a very large coarse-grained mesh (large gray grid) of one-third the spatial resolution otherwise required to measure hydrodynamic observables. We then measure the maximum transverse radius $R_{\max}$ (blue circle) of the hypersurface defined by the temperature isotherm $T = T(e_{\min})$, where $e_{\min}$ is the largest energy density which can be truncated without modifying the hydrodynamic observables calculated from the event. Finally, the initial condition event is rerun on a smaller and finer mesh (smaller black grid) with three-times the cell density of the pre-run event and a smaller transverse extent $-R_{\max} < x < R_{\max}$.

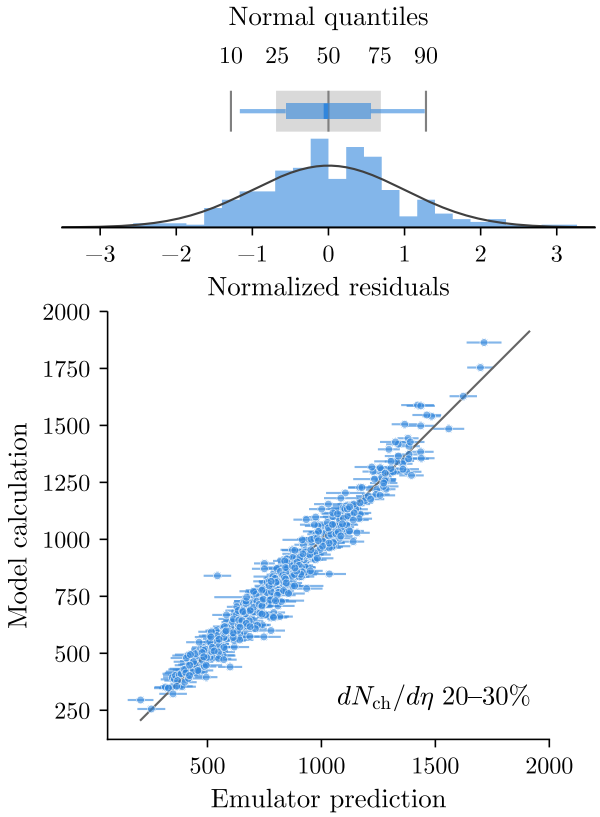

FIG. 16 Example emulator validation for one observable, the Pb-Pb charged-particle yield $dN_{\rm ch}/d\eta$ in the 20-30% centrality class. We use the k-fold cross validation method (explained in the text) to partition the model inputs X and outputs Y into training and validation data. The scatter plot on the left shows the emulator predictions and one sigma error bars (x-axis) against explicit model calculations (y-axis). Perfect emulator/model agreement is indicated by the black like $y_{\rm pred} = y_{\rm obs}$. The histogram on the right shows that the errors are properly accounted for, i.e. the normalized residuals follow a normal distribution with unit variance and zero mean.

single VISH2+1 event scales like $\sim n_x^3$ since $dx \propto d\tau$, and thus our "pre-run" event requires only $\sim$1/27th the time of a production event. We therefore start by running a coarse-grained event on an excessively large grid for *every* minimum-bias event to estimate $R_{\rm max}$, then rerun the same event on a thrice finer grid with a trimmed spatial extent $x_{\rm max} = R_{\rm max}$. See Fig. 15 for a simple diagram of the procedure.

In practice, we find that event-by-event grid resizing leads to a massive speed increase for minimum bias events compared to using a single fixed grid for the entire minimum bias sample. This is because the maximum transverse size of each event varies dramatically, from a few fm in peripheral Pb-Pb collisions to 50 fm or more in central Pb-Pb collisions. The procedure should generalize to other hydrodynamic codes.

Appendix B: Emulator validation

The emulator is a surrogate for the full physics simulation which generates probabilistic predictions for the model observables $\mathbf{y}_{\rm model}$ at a given point $\mathbf{x}$. Here we validate these probabilistic predictions using a method known as k-fold cross validation. We first randomly partition our $d = 500$ training points into $k = 20$ equal sized subsamples or "folds". One of the subsamples is used to validate the emulator and the remaining $k-1$ subsamples are used to train it. The process is then repeated for each of the subsamples so that we end up validating on all of the training data.

Figure 16 shows a scatter plot of the emulator predictions with one-sigma error bars (x-axis) against explicit model calculations (y-axis). Perfect emulator and model agreement is indicated by the black line $y_{\rm pred} = y_{\rm obs}$. If the emulator errors are properly accounted for, then the normalized residuals $z = (y_{\rm pred} - y_{\rm obs})/\sigma_{\rm pred}$ sample a unit normal distribution

$$P(z) \sim \mathcal{N}(\mu = 0, \sigma = 1). \tag{B1}$$

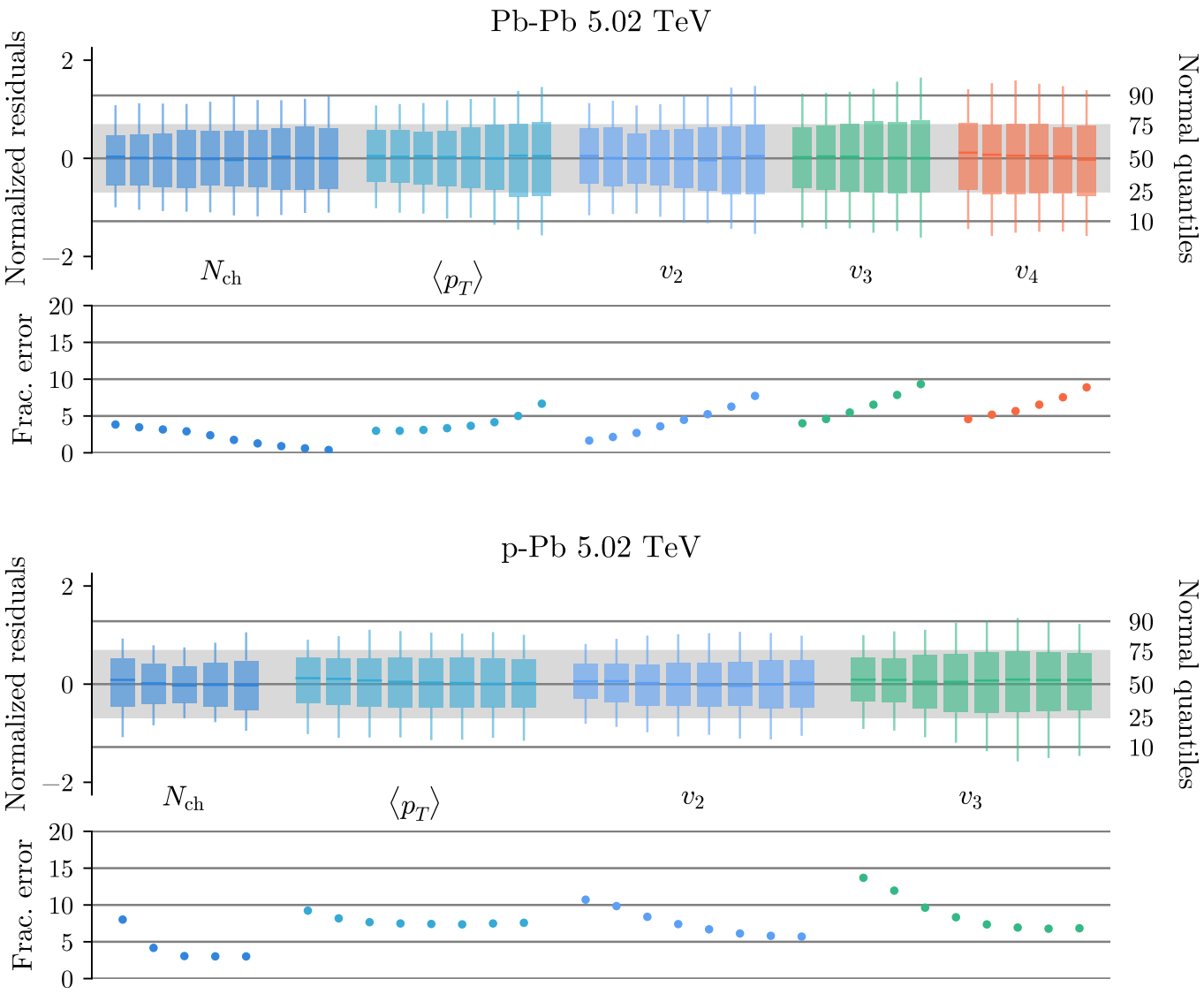

FIG. 17 Emulator validation for the Pb-Pb collision system (top) and p-Pb collision system (bottom) at $\sqrt{s_{\rm NN}} = 5.02$ TeV. The "piano keys" in the top row of each figure are horizontally stacked box plots for the normalized residuals of each model observable. The boxes are 50% interquartile ranges and whiskers are the 90% interquantiles. The bottom row of each figure is the RMS fractional error defined by Eq. (B1).

This comparison is shown by the histogram and box plot on the right side of Fig. 16. The emulator error is clearly significant, but it is also properly modeled, as indicated by the agreement between the normalized residuals and the unit normal distribution on the right (black curve). Moreover, since we include this uncertainty in the likelihood covariance matrix (39), we expect our results to be robust to the emulator limitations. This is an important point that bears repeating. The emulator uncertainty does not erode the veracity of the posterior distribution if it is correctly modeled and accounted for.

More generally, we can perform the validation test in Fig. 16 for *every* observable $y \in \mathbf{y}$ and check that each observable's normalized residuals $z = (y_{\rm pred} - y_{\rm obs})/\sigma_{\rm pred}$ follow a unit normal distribution. This test is applied to the p-Pb and Pb-Pb collision systems in Fig. 17. The top row of each figure shows a box-plot for the normalized residuals of each observable compared to the quantiles of a unit normal distribution. The thin horizontal black lines correspond to the 10th and 90th percentiles of a unit normal distribution, and the gray band its interquartile range. These visual references should be compared to the whiskers and interquartile range respectively of each box plot, analogous to the comparison test of Fig. 16. The emulators generally behave as expected, although the validation is better for the Pb-Pb system than the p-Pb system. For instance, the p-Pb charged-particle yield $dN_{\rm ch}/d\eta$ uncertainties are over predicted. It is not immediately clear why this would be the case, but the MAP observables in Fig. 13 are

in good agreement with their emulator predictions which suggests it should not be a grave concern.

We also show in Fig. 17 an estimate of the emulator error magnitude. This error is expressed in terms of the unitless variable

$$\hat{z} = \frac{y_{\rm pred} - y_{\rm obs}}{(\Delta y)_{99\%}}, \tag{B2}$$

where $(\Delta y)_{99\%}$ is 99% of the full variability of y across the design. Thus $\hat{z}$ can be thought of as a fractional emulator error relative to the full design variability. The bottom row of each figure shows the root-mean-square (RMS) value of $\hat{z}$. We see that RMS $\hat{z}$ ranges from a few percent for most observables to a maximum value of 15% for the p-Pb triangular flow $v_3\{2\}$ in the lowest multiplicity bin. This suggests that the present analysis would benefit the most from more p-Pb events, in particular, from more multiplicity triggered events which are used to calculate the flows.