Initial Conditions of Bulk Matter in Ultrarelativistic Nuclear Collisions

by

J. Scott Moreland

Department of Physics
Duke University

Date: ________________________

Approved:

__________________________________________
Steffen A. Bass, Supervisor

__________________________________________
Berndt Müller

__________________________________________
Christopher Walter

__________________________________________
Alfred T. Goshaw

__________________________________________
John E. Thomas

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Physics in The Graduate School of Duke University

2019
ABSTRACT

Initial Conditions of Bulk Matter in Ultrarelativistic Nuclear Collisions

by

J. Scott Moreland

Department of Physics
Duke University

Date: __________________________

Approved:

_____________________
Steffen A. Bass, Supervisor

_____________________
Berndt Müller

_____________________
Christopher Walter

_____________________
Alfred T. Goshaw

_____________________
John E. Thomas

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Physics in The Graduate School of Duke University

2019
Abstract

Dynamical models based on relativistic fluid dynamics provide a powerful tool to extract the properties of the strongly-coupled quark-gluon plasma (QGP) produced in the first $\sim 10^{-23}$ seconds of an ultrarelativistic nuclear collision. The largest source of uncertainty in these model-to-data extractions is the choice of theoretical initial conditions used to model the distribution of energy or entropy at the hydrodynamic starting time.

Descriptions of the QGP initial conditions are generally improved through iterative cycles of testing and refinement. Individual models are compared to experimental data; the worst models are discarded and best models retained. Consequently, successful traits (assumptions) are passed on to subsequent generations of the theoretical landscape. This so-called bottom-up approach correspondingly describes a form of theoretical trial and error, where each trial proposes an *ab initio* solution to the problem at hand.

A natural complement to this strategy, is to employ a top-down or data-driven approach which is able to reverse engineer properties of the initial conditions from the constraints imposed by the experimental data. In this dissertation, I motivate and develop a parametric model for initial energy and entropy deposition in ultrarelativistic nuclear collisions which is based on a family of functions known as the generalized means. The ansatz closely mimics the variability of *ab initio* calculations and serves as a reasonable parametric form for exploring QGP energy and entropy deposition assuming imperfect knowledge of the complex physical processes which lead to its creation.

With the parametric model in hand, I explore broad implications of the proposed ansatz
using recently adapted Bayesian methods to simultaneously constrain properties of the ini-
tial conditions and QGP medium using experimental data from the Large Hadron Collider. 
These analyses show that the QGP initial conditions are highly constrained by available 
measurements and provide evidence of a unified hydrodynamic description of small and 
large nuclear collision systems.
I dedicate this dissertation to my mother and father and to Erin, for their unyielding love, encouragement, and unwavering support.
Contents

Abstract iv

List of Tables ix

List of Figures x

Acknowledgments xiv

1 Introduction 1

2 Ultrarelativistic nuclear collisions 6
   2.1 Characterizing hot and dense nuclear matter .............................. 6
   2.2 Hadron collider experiments ............................................... 17
   2.3 Hydrodynamic computer simulations ...................................... 29
   2.4 The initial condition problem .............................................. 36

3 Initial conditions of bulk matter 43
   3.1 Approximations in the high-energy limit .................................. 45
   3.2 Nuclear structure ............................................................. 49
   3.3 Inelastic nuclear cross sections .......................................... 55
   3.4 Energy and entropy deposition ............................................. 69
   3.5 The TRENTo model ........................................................... 80
List of Tables

3.1 Estimates of the parameters $R$ and $a$ for the two-parameter Fermi distrib... 52
3.2 Approximate best-fit fluctuation parameters $k$ and normalizations for each... 88

5.1 Input parameter ranges for the nuclear collision model. 147
5.2 Experimental data to be compared with model calculations. 147
5.3 Estimated parameter values (medians) and uncertainties (90% credible in... 151
5.4 High-probability parameters chosen based on the posterior distributions... 157
5.5 Maximum a posteriori (MAP) model parameters from Ref. [221]. 170
5.6 Generating function cumulant parametrizations used in the present analy... 183
5.7 Input parameter ranges for the nuclear collision model. 187
5.8 Experimental data used to calibrate the model parameters. 188
5.9 High-probability maximum a posterior (MAP) parameters selected from... 195
5.10 Input parameter ranges for the nuclear collision model. 212
5.11 Experimental data used to calibrate the model parameters. 213
5.12 Posterior parameter estimates corresponding to Fig. 5.35. The reported... 221
5.13 High-probability parameters selected from the posterior distribution and... 230
List of Figures

2.1 Schematic of the QCD phase diagram in the temperature and baryon che... 7
2.2 Lattice equation of state for (2+1)-flavor QCD (u, d, and s quarks) at zero... 10
2.3 The shear viscosity $\eta$ applies a force that opposes shearing flows (left),... 13
2.4 Computer rendering of the ALICE detector experiment [59]. 18
2.5 Diagram illustrating the relationship between the pseudorapidity $\eta$ and... 20
2.6 Two nuclei move in and out of the plane of the page. Arrow is the collision... 21
2.7 Centrality classes defined by the ALICE experiment using the sum of am... 22
2.8 Left: Particle yields for various hadrons predicted by the statistical hadro... 23
2.9 Left: Spatially deformed initial energy density profile in the transverse... 24
2.10 Left two figures: Differential elliptic flow $v_2$ for various hadron species... 26
2.11 Di-jet event (teal cones) superimposed on the initial transverse energy de... 27
2.12 Top: Two-particle azimuthal distributions for minimum-bias and central... 28
2.13 Qualitative space-time diagram of a relativistic heavy-ion collision. 30
2.14 Hybrid model computer simulation of a typical heavy-ion collision event.... 33
2.15 Elliptically deformed spatial profile (left) generates elliptically deformed... 33
2.16 Eccentricity-scaled elliptic flow $v_2/\varepsilon_2$ plotted versus the charged-particle... 35
2.17 Left: Root-mean-square anisotropic flow coefficients $\sqrt{\langle v_n \rangle^2}$ as a function... 37
2.18 Probability distributions for the scaled anisotropic flows $v_n/\langle v_n \rangle$ and scal... 38
2.19 Hydrodynamic calculations using EKRT initial conditions [114] for the... 41
3.1 Nuclear overlap times $\tau_{\text{overlap}}$ measured in the lab frame for Pb-Pb collisi... 46
3.2 Cartoon of the local nuclear density overlap in an ultrarelativistic nuclear... 47
3.3 In the ultrarelativistic limit, causality reduces the collision dynamics to a... 48
3.4 Examples of lead, xenon, and copper isotopes generated by the Woods-... 51
3.5 Geometric interpretation of the cross section for a collision of two billiard... 56
3.6 Left: Participant nucleon density (heat map) from the analytic Glauber... 64
3.7 Inelastic nucleon-nucleon collision probability $P_{\text{coll}}$ as a function of the... 66
3.8 Monte Carlo Glauber model predictions for the number of nucleon partici... 67
3.9 Left: Local participant thickness functions $\tilde{T}_A$ and $\tilde{T}_B$. Right: Average... 70
3.10 Charged-particle density $dN_{\text{ch}}/d\eta$ as a function of collision centrality for... 72
3.11 Charged-particle density $dN_{\text{ch}}/d\eta$ as a function of collision centrality for... 78
3.12 Realistic transverse entropy density profiles predicted by the generalized... 79
3.13 Stages of the T\textsc{R}\textsc{E}\textsc{NT}o initial condition model: 1) sample nucleon positio... 81
3.14 Profiles of the initial thermal distribution predicted by the KLN (left),... 84
3.15 Eccentricity harmonics $\varepsilon_2$ and $\varepsilon_3$ as a function of impact parameter $b$ for... 85
3.16 Multiplicity distributions for proton-proton, proton-lead, and lead-lead co... 88
3.17 Average charged particle density per participant nucleon pair... 89
3.18 Left and middle plots: Eccentricity harmonics $\varepsilon_2$ and $\varepsilon_3$ as a function of... 90
3.19 Comparison of tip-tip and side-side uranium-uranium collisions. Schemat... 91
3.20 Ellipticity $\varepsilon_2$ as a function of normalized charged-particle multiplicity... 92
3.21 Cross section of the reduced thickness function for a pair of nucleon parti... 93
3.22 Reduced thickness for a pair of nucleon participants using a harmonic me... 94

4.1 Left: $5 \times 5$ factorial design. Right: $5 \times 5$ Latin square. The Latin square... 101
4.2 Latin hypercube design with $d = 100$ design points and $n = 3$ three para... 102
4.3 Simulated observables compared to the experimental data. Blue lines are... 103
4.4 GPs trained on several different datasets (black symbols) using the covari... 108
4.5 Left: Scatter plot of 1000 vectors $Y = (y_1, y_2, \ldots, y_{1000})$ randomly sampl... 113
4.6 The reference figure on the left shows a GP (blue line and error band)... 120
4.7 Left: Emulator predictions for observable $y_5$ scatter plotted against the... 123
4.8 Correlation matrix, $\text{corr}(y_i, y_j) = \text{cov}(y_i, y_j)/(\sigma_i\sigma_j)$, visualized for the mo... 127
4.9 Simulated observables (blue lines) compared to the experimental data (bl... 132
4.10 Bayesian posterior distribution for the three model input parameters $x_1,... 133

5.1 Some example curves demonstrating the variability of the temperature-... 143
5.2 Scatter plot of the predicted (emulated) model output versus the observed... 148
5.3 Model and emulator predictions (colored lines) compared to experimental... 150
5.4 Posterior distribution for the model parameters, calibrated including iden... 152
5.5 Posterior distribution on the TRENTo generalized mean parameter $p$ com... 154
5.6 Posterior estimate for the temperature dependence of the QGP specific... 156
5.7 Explicit model calculations using the high-probability parameters listed in... 158
5.8 Cartoon of the coupling strength for the first $\sim 1$ fm/c of the collision. The... 161
5.9 Degrees of freedom in the temperature dependent shear and bulk viscosity... 163
5.10 Diagram of the adaptive grid resizing algorithm (not to scale). Each initial... 165
5.11 Posterior distribution for the model input parameters of Ref. [221]. Diag... 171
5.12 Model calculations (colored lines) for Pb-Pb collisions at two beam energi... 172
5.13 Posterior on the TRENTo generalized mean parameter $p$ compared to the... 174
5.14 Posterior estimates for the temperature dependence of the QGP specific... 175
5.15 Experimentally measured values for the inelastic nucleon-nucleon cross se... 176
5.16 Pb-Pb charged-particle density $dN_{ch}/d\eta$ for 0–10% central events at... 177
5.17 Model calculations (colored lines) using the MAP parameters from Ref... 178
5.18 Rapidity-dependence of initial entropy deposition shown for different valu... 184
5.19 Initial entropy density for an example Pb-Pb event (top) and p-Pb event... 185
5.20 Posterior distribution for the model input parameters of Ref. [157]. Diag... 192
5.21 Left and middle: Pseudorapidity dependence of the charged-particle dens... 193
5.22 Anisotropic flow cumulants $v_2\{2\}$ and $v_3\{2\}$ at midrapidity $|\eta| < 0.8$ plot... 196
5.23 Rapidity-dependence of the flow cumulants $v_2\{2\}$, $v_3\{2\}$, and $v_2\{4\}$ plott... 197
5.24 Twisting of the second-order event plane angle $\Psi_2$ as a function of pseudo... 198
5.25 Two-particle factorization ratio $r_n(\eta^a, \eta^b)$ for Pb-Pb collisions at $\sqrt{s_{NN}} = ... 200
5.26 Top panels: Symmetric cumulants SC(4, 2) (blue) and SC(3, 2) (green)... 202
5.27 Root-mean-square elliptic flow coefficient $\langle v_2^2 \rangle^{1/2}$ plotted as a function of... 206
5.28 Thickness function $T(x_\perp) = \int dz \rho_n(x_\perp, z)$ for three example nucleon den...

5.29 Effect of nucleon substructure on the nuclear thickness function $T(x_\perp) = ...$

5.30 Visualization of the Pb-Pb correlation matrix...

5.31 Example emulator validation for one observable, the Pb-Pb charged...

5.32 Emulator validation for the Pb-Pb collision system (top) and $p$-Pb collision...

5.33 Simulated observables compared to experimental data for Pb-Pb collisions...

5.34 Same as Fig. 5.33 but for $p$-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. Multiplicity...

5.35 Bayesian posterior distribution of the model input parameters. The diago...

5.36 Left: Joint posterior distribution of the constituent sampling radius $r$ and...

5.37 Left figure: Estimated temperature dependence of the QGP specific shear...

5.38 Model calculations using the *maximum a posterior* (MAP) parameters co...

5.39 Relative mean transverse momentum fluctuations $\delta p_T/\langle p_T \rangle$ plotted for...

5.40 Model calculations of the symmetric cumulants (top) and normalized sym...

xiii
Acknowledgments

This work was supported by the Department of Energy National Nuclear Security Administration Stewardship Science Graduate Fellowship (DOE NNSA SSGF) under grant number DE-FC52-08NA28752, and by the U.S. Department of Energy (DOE) under grant number DE-FG02-05ER41367.

Computing resources were provided by the Open Science Grid (OSG), funded by the DOE and National Science Foundation (NSF), and by the National Energy Research Scientific Computing Center (NERSC), funded by the DOE.
Introduction

A central scientific endeavor is to investigate the reducible nature of matter—to classify its elementary quanta and understand its fundamental interactions. This search is aided by particle accelerators, fantastic machines that collide together nature's smallest particles in search of hidden substructure and unifying symmetries. The crowning jewel of this effort is the so-called Standard Model of particle physics which describes the strong, weak, and electromagnetic forces observed in nature.

Quantum chromodynamics (QCD), the theory of the strong nuclear force, explains the zoo of strongly interacting particles produced by high-energy nuclear collisions as combinations of two or more fundamental particles known as quarks. Each quark carries color charge—analogous to the more familiar electric charge of classical electromagnetism—and interacts by exchanging particles known as gluons which mediate the strong force.

A property of QCD known as color confinement stipulates that free quarks can never be observed in nature; quarks may only combine to form color-neutral bound states known as hadrons, of which the proton and neutron are just two examples. Although the existence of quark and gluon degrees of freedom cannot be observed directly, their presence has been inferred by examining the properties of final-state hadrons produced by energetic nuclear collisions.

One of the primary goals of the high-energy nuclear physics community is to understand the emergent behavior which arises from fundamental quark and gluon interactions over
different time and distance scales. This encompasses both the complex dynamics which occur inside a relativistic nuclear collision, as well as other more exotic nuclear phenomena such as the primordial interactions of quarks and gluons shortly after the big bang. This particular endeavor is distinguished from the more general effort to specify the fundamental forces and elementary particles of the Standard Model, in that it seeks to understand bulk properties of quark-gluon matter, i.e. attributes of the aggregate substance and not just its individual components.

Relativistic nuclear collisions are a powerful experimental tool to study quark and gluon interactions experimentally, because unlike the one-of-a-kind event which produced the big bang, high-energy particle physics experiments are repeatable and configurable. They therefore provide an experimental sand box to develop and test theoretical ideas. This general concept of using high-energy collisions to study the bulk properties of nuclear matter dates back to the early 1950’s, when Landau proposed a hydrodynamic description of hadronic collisions [1]. His general argument followed a simple line of reasoning. When two nucleons collide at relativistic energies, they release a large amount of energy into a very small volume which may be viewed from the center of mass frame of the colliding nucleon pair. If the collision energy is sufficiently high, the resulting density of secondary particles will be large, and their mean free path will be short relative to the system size. The resulting interparticle interactions will thus be governed by statistical laws, and the produced fireball will expand hydrodynamically until the mean free path of the particles becomes comparable to that of the system size. The system will then ultimately break up and disintegrate into a shower of separate particles [2].

Landau’s original hydrodynamic model never described the quanta of a nuclear collision in terms of quarks and gluons; in fact, the existence of these particles was not even postulated until nearly a decade later [3, 4]. His hydrodynamic model would, however, ultimately lay the groundwork for a new way of thinking about fundamental interactions between quarks and gluons in the context of a thermalized fluid. This modern hydrodynamic picture of relativistic nuclear collisions began to emerge when Gross, Wilczek, and
Politzer discovered asymptotic freedom in 1973: a phenomenon that predicts a weakening of the strong interaction between quarks as the quarks get closer together [5, 6]. Their finding had broad phenomenological implications, and it lead to the realization that quarks and gluons would become liberated in high energy nuclear collisions to produce a new state of deconfined matter subsequently referred to as *quark-gluon plasma* or QGP for short [7, 8].

About a decade later, Bjorken famously synthesized these ideas and developed a revolutionary model of relativistic nuclear collisions which remains largely accurate to this day [9]. The ideas were based on Landau’s model of ideal hydrodynamics. Bjorken’s insight was to apply additional symmetries to the problem in order to derive simple solutions for the hydrodynamic equations of motion. These equations allowed Bjorken to elucidate the space-time evolution of the collision and provide estimates for its initial energy density and temperature. From these estimates he reasoned that it was likely the produced system would be in the deconfined QGP phase.

In the years that followed, hydrodynamic modeling of nuclear collisions grew from a nascent qualitative science into a quantitative one. Viscosity was added to the simulations [10–16]. Crude estimates for the QGP energy density and pressure were replaced with realistic calculations derived from first principles [17, 18]. Models were updated to include event-by-event fluctuations in the density of initial nuclear matter [19], and descriptions of dilute regions of the collision were also greatly improved [20–24]. The refined simulations began to accurately reproduce and even predict a large number of seemingly unrelated experimental observables, substantiating the veracity of the hydrodynamic framework.

Modern hydrodynamic computer models allow researchers to simulate the full time history of the QGP produced in relativistic nuclear collisions in all its gory detail. The models recreate events exactly as they are believed to occur inside the detector and output simulated observables that can be directly compared to experimental data. Free parameters of the framework such as its dissipative transport coefficients are then calibrated to optimally reproduce experimental measurements in order to infer intrinsic properties of the produced
matter.

In this manner, data-driven methods are used to extract fundamental properties of hot and dense nuclear matter which are not directly accessible to first principle calculations due to the complexity of the system’s microscopic dynamics. The accuracy of these model-based QGP parameter extractions is of course limited by the fidelity of the simulations. If any aspect of the simulation is incorrectly modeled, it will generally affect the inferred values of the model parameters. Estimating these QGP parameters with quantitative uncertainty thus involves a careful accounting of all sources of potential error in the assumed framework.

The hydrodynamic initial conditions—which describe the energy density and flow velocity of the QGP medium at the hydrodynamic starting time $\sim 1 \text{ fm/c}$ after the nuclei first collide—are the single largest source of uncertainty impeding the extraction of QGP medium properties by comparing simulation predictions to data. They are simulated using a variety of different computer models, and there is no unified consensus regarding their correct theoretical treatment. Different initial condition models generally predict different descriptions of the QGP space-time evolution and hence prefer different values for the QGP medium parameters. Their understanding is thus a limiting factor when using models to reverse engineer properties of the produced matter.

The QGP initial conditions are therefore important for two separate reasons. First, they are interesting in their own right. They evolve out of a highly chaotic dynamical process which tests our current understanding of nuclear matter under extreme conditions. Second, they provide a necessary ingredient for dynamical simulations of the collision. If the initial conditions are incorrectly modeled, the simulation predictions will be misleading, and all derivative conclusions will be tenuous at best. In this latter sense, the initial conditions act as a nuisance parameter.

Ultimately, one seeks a correct first principles description of the QGP initial conditions as it would appropriately address both of these objectives. Deriving the QGP initial conditions from first principles, however, is exceptionally challenging. QCD is so difficult to solve in practice, that \textit{ab initio} initial condition calculations only exist for approximations of QCD
and related quantum field theories. These calculations generally involve different starting assumptions and hence result in descriptions of the QGP initial conditions which are always in some degree of mutual tension.

Such *ab initio* calculations are commonly refined through iterative cycles of trial and error. Individual theoretical assumptions are tested by comparing model predictions to experimental data. Successful assumptions are then passed on to subsequent iterations of the theoretical landscape and problematic assumptions discarded. Each step of the validation process is slow and typically involves significant computational effort. Model-to-data comparison has thus emerged as a rich field of research in and of itself.

Generally speaking, these efforts describe a so-called bottom-up approach that searches for a solution to the problem derived from deeper fundamental laws. In this dissertation, I apply an alternative, albeit complementary, approach to study the QGP initial conditions which addresses the problem from the opposite direction. I start with the observations of the experimental data and work backwards to infer the requisite starting point of hydrodynamic simulations. This data-driven or top-down approach is commonly known as solving the inverse problem.

Data-driven methods naturally require a new way of thinking about the QGP initial conditions, one that embraces theoretical uncertainty instead of fighting it. For this purpose, I develop an extremely simple parametric model of the QGP initial conditions which is flexible enough to span a wide range of reasonable theoretical descriptions. In this sense, I create a meta-model for the landscape of mutually incompatible theory calculations. With the parametric model in hand, I then proceed to rigorously constrain its free parameters with experimental data, using Bayesian methods recently developed for heavy-ion collisions. I find that the functional form of the QGP initial conditions is highly constrained by existing measurements, regardless of the theoretical uncertainty surrounding the details of its derivation. This eliminates, to a large degree, the confounding uncertainty introduced by different microscopic models of the initial conditions, enabling quantitative QGP parameter estimates with meaningful uncertainty and unprecedented precision.
Ultrarelativistic nuclear collisions

Shortly after the quark-gluon nature of nuclear matter was discovered in the 1960s and 1970s, physicists began to seriously consider the idea of using high-energy nuclear collisions to study the properties of nuclear matter at extreme temperatures and densities [25]. It was believed that heavy-ion collisions, e.g. two gold nuclei, would maximize the produced matter’s lifetime and system size, thereby enhancing the QGP’s effect on final state observables. Simple estimates based on the energy released per unit rapidity in nucleon-nucleon collisions indicated that relativistic heavy-ion collisions could reasonably attain energy densities in excess of ~3 GeV/fm$^3$, conditions which were generally expected to be sufficient to produce thermalized matter in the deconfined QGP phase [9].

2.1 Characterizing hot and dense nuclear matter

Motivated in part by these general ideas, the US and international nuclear theory communities invested significant resources over the next few decades developing ultrarelativistic heavy-ion programs at the Relativistic Heavy-ion Collider (RHIC) located in Brookhaven, New York and the Large Hadron Collider (LHC) situated on the border of France and Switzerland. I’ll discuss these experiments in more detail shortly. First, I want to explain some of the big picture questions which these programs sought to address.

Broadly, the goal of these investments is to quantify the bulk properties of hot and dense nuclear matter. These properties can be subdivided into two general categories: equilibrium
properties which characterize the matter’s steady-state behavior and dynamical properties which describe its response to deviations from equilibrium.

### 2.1.1 Equilibrium properties

The equilibrium properties of a substance depend on the conditions of its static environment. For example, water is a liquid at room temperature and atmospheric pressure, while it exists as a solid and gas at other temperature and pressure combinations. This information is typically plotted as a phase diagram which illustrates the pressure and temperature combinations needed to reproduce each phase of matter.

The same general picture is also used to classify different phases of nuclear matter. Figure 2.1 shows the current picture of the QCD phase diagram as a function of temperature $T$ and baryon chemical potential $\mu_B$, a quantity related to the net baryon density (net imbalance of matter and antimatter). At low temperatures and baryon densities, nuclear matter exists as hadrons, color-neutral combinations of two or three bound-quark states,
while at higher temperatures and/or baryon densities, these hadrons “melt” to form the deconfined QGP phase. Lattice QCD calculations (see below) have established that the transition from the hadronic phase to the QGP phase is a smooth crossover at $\mu_B = 0$. Meanwhile, at larger $\mu_B$, the transition is expected to become first-order [26], although it is not yet clear from first principles where this should occur. I’ll briefly summarize now the basis of our current theoretical understanding at small baryon chemical potential. RHIC and LHC collisions produce almost equal parts matter and antimatter at $\sqrt{s_{NN}} > 100$ GeV, so zero net baryon density is a very good approximation for the collisions studied in this dissertation.

2.1.1.1 Equation of state

At each point in the QCD phase diagram, the equilibrium properties of nuclear matter are quantified by an equation of state (EoS), specifying the energy density, entropy density, and pressure (among other quantities) at fixed temperature and baryochemical potential. At zero baryochemical potential (left edge of figure 2.1), the QCD EoS is rigorously calculable using non-perturbative methods based on the Feynman path integral approach. The key to this method is the realization that the density operator $\hat{\rho} = e^{-\beta \hat{H}}$ resembles a time-evolution operator $e^{i\hat{H}t}$ if one replaces $\beta$ (inverse temperature) with imaginary time $\tau = -it$. Therefore, by substituting $t \rightarrow i\tau$, the path integral formulation of the field theory can be made to resemble a partition function $Z = \text{tr} \hat{\rho}$, thereby specifying the system’s statistical properties in thermodynamic equilibrium.

The partition function can be evaluated using lattice QCD, an algorithm to discretize the path integral onto a hypercubic lattice of $N_\sigma^3 N_\tau$ space-time points, where $N_\sigma$ and $N_\tau$ are the number of steps used to discretize the spatial and temporal dimensions respectively. These lattice sites are separated by lattice spacing $a$ which relates the number of grid steps
\( N_\sigma, N_\tau \) to the simulation’s effective equilibrium temperature and volume

\[
T = 1/(aN_\tau),
\]

\[
V = 1/(aN_\sigma)^3.
\]

The calculation is repeated for different grid dimensions to vary the system’s equilibrium temperature and grid resolution. Then, the results of successively finer grids are extrapolated to the continuum limit to remove finite lattice effects.

Lattice calculations are typically presented in terms of the trace of the stress energy tensor \( \Theta^{\mu\nu} \), equal to the difference of the energy density and three times the pressure. This quantity is commonly referred to as the trace anomaly or interaction measure because it measures the deviation of the fluid from the conformal EoS. Defined on the lattice, the trace anomaly is related to the total derivative of \( \log Z \) with respect to the lattice spacing \( a \):

\[
\Theta^{\mu\nu} = -\frac{T}{V} \frac{d \log Z}{d \log a}.
\]

Scaled by powers of the temperature \( T \), the trace anomaly \( \Theta^{\mu\nu} \) forms a dimensionless interaction measure

\[
I = \frac{\Theta^{\mu\nu}(T)}{T^4} = e - 3P.
\]

The thermodynamic pressure is then calculated from the interaction measure using the relation

\[
\frac{P(T)}{T^4} = \frac{P_0}{T_0^4} + \int_{T_0}^{T} \frac{dT'}{T'} I(T'),
\]

where \( P_0 \) and \( T_0 \) are a reference pressure and temperature, typically calculated from the hadron resonance gas model. The energy density \( e \) and entropy density \( s \) are then easily
Figure 2.2  Lattice equation of state for (2+1)-flavor QCD (u, d, and s quarks) at zero net baryon density calculated by the HotQCD collaboration (colored) and Wuppertal-Budapest collaboration (gray) [17, 18].

obtained from the thermodynamic relations

\[
e(T) T^{-4} = I(T) + 3 \frac{P(T)}{T^4}, \tag{2.6}
\]

\[
s(T) T^{-3} = \frac{e(T) + P(T)}{T^4}. \tag{2.7}
\]

Figure 2.2 shows the trace anomaly, pressure, and entropy density divided by powers of the temperature for (2+1)-flavor QCD (u, d, and s quarks) at zero net baryon density obtained from lattice calculations performed by two independent collaborations. The gray bands are calculations by the Wuppertal-Budapest collaboration using the stout fermion action [17], and the colored bands are calculations by the HotQCD collaboration using the HISQ/tree action [18]. Both collaborations observe a smooth crossover phase transition to the QGP phase located at the pseudocritical temperature \( T_c \sim 150-155 \text{ MeV} \) at \( \mu_B = 0 \). Considering the complexity of each calculation, the agreement between the two groups is a remarkable accomplishment.

Recent developments in lattice QCD include more precise estimates for the QGP pseudocritical temperature \( T_c = 156.5 \pm 1.5 \text{ MeV} \) [27] and new calculations in (2+1+1)-flavors, i.e. with thermalized charm quarks [28]. The addition of charm quarks modifies the trace
anomaly at very high temperatures, but the corrections are modest for $T \lesssim 400$ MeV. Therefore, the QCD EoS is generally considered to be well constrained by first-principles theory at vanishing net baryon density.

These calculations, however, describe just one edge of the QCD phase diagram at zero baryon chemical potential $\mu_B$. As I mentioned previously, an outstanding question facing the nuclear physics community is whether QCD switches from a smooth crossover at $\mu_B = 0$ to a first-order transition at some $\mu_B > 0$, as predicted by multiple theories [29]. This feature in the phase diagram is known as the critical point (see figure 2.1).

Strictly speaking, it is not yet possible to calculate the QCD EoS at significant baryon density on the lattice due to the existence of the fermion sign problem [30]. Nevertheless, several lattice-based methods exist to calculate the QCD EoS in the presence of a small quark potential. For example, one can take derivatives of quark and gluonic observables with respect to $\mu_B$ to calculate the leading-order Taylor expansion of the theory at the edge of the phase diagram $\mu_B = 0$ [31]. The truncated Taylor expansion can then be used to extrapolate to small baryon densities $\mu_B > 0$. Alternatively, the QCD EoS can be solved for imaginary quark potentials, thereby circumventing the sign problem, and analytically continued to real $\mu_B$ [32–34]. The QCD EoS at nonzero baryon density is therefore an evolving picture, and an ongoing area of theoretical research.

### 2.1.2 Dynamical properties

To this point, I've only discussed the steady-state properties of bulk nuclear matter at fixed temperature and chemical potential. Such ideal systems, however, seldom exist in nature. The physical processes that produce QGP matter are typically violent and far from equilibrium. Ultrarelativistic heavy-ion collisions, for example, produce small ($10^{-14}$ m), short-lived ($10^{-23}$ s) QGP fireballs that rapidly expand and cool, tracing complex trajectories through the QCD phase diagram.

The collision’s dynamical evolution contains additional information—specific to the form of matter—which is not specified by the QCD EoS. Therefore, it is important to supplement
thermodynamic measures with additional numbers to characterize these properties.

2.1.2.1 Dissipative hydrodynamics

Hydrodynamics is a mathematical framework that describes the response of a system to small perturbations from local thermal equilibrium, constructed by applying basic conservation laws to gradient expansions of the stress-energy tensor $T^{\mu \nu}$. It relates the system’s extended non-equilibrium dynamics to the properties of its locally equilibrated matter. Expanded to first-order in gradients of the fluid flow velocity, the stress-energy tensor can be written as

$$T^{\mu \nu} = (e + P)u^{\mu}u^{\nu} - P g^{\mu \nu} + \pi^{\mu \nu} - \Delta^{\mu \nu} \Pi,$$  \hspace{1cm} (2.8)

where $e$ and $P$ are the energy density and pressure in the local fluid rest frame, $u^{\mu}$ is the local fluid velocity, and $\Delta^{\mu \nu} = g^{\mu \nu} - u^{\mu}u^{\nu}$ is the projector onto the space orthogonal to $u^{\mu}$. The terms $\pi^{\mu \nu}$ and $\Pi$, meanwhile, are the first-order shear and bulk viscous corrections to the zeroth-order theory which I’ll describe shortly.

The hydrodynamic equations of motion are obtained from equation (2.8) by applying energy-momentum and charge conservation,

$$\partial_{\mu}T^{\mu \nu} = 0 \quad \text{and} \quad \partial_{\mu}j^{\mu} = 0,$$  \hspace{1cm} (2.9)

to the energy-momentum tensor $T^{\mu \nu}$ and charge-current $j^{\mu} = nu^{\mu}$ in combination with an EoS $P = P(e)$ and initial conditions for $e$, $u^{\mu}$, $\Pi$, and $\pi^{\mu \nu}$. Typically for heavy-ion collisions, the charge-current is associated with the system’s baryon density $n$. Throughout this dissertation, I study ultrarelativistic nuclear collisions with vanishing net baryon density, so I’ll neglect discussing this latter conserved current.

The shear viscous pressure tensor $\pi^{\mu \nu}$ and bulk pressure $\Pi$ apply dissipative corrections to the stress-energy tensor $T^{\mu \nu}$. In relativistic Navier-Stokes theory, these viscous terms
can be further decomposed in the form \cite{35}

\[ \pi^{\mu\nu} = 2\eta \Delta^{\mu\nu\alpha\beta} \partial_\alpha u_\beta \quad \text{and} \quad \Pi = -\zeta \partial_\mu u^\mu, \]  

(2.10)

where

\[ \Delta^{\mu\nu\alpha\beta} = \frac{1}{2} (\Delta^{\mu\alpha} \Delta^{\nu\beta} + \Delta^{\nu\alpha} \Delta^{\mu\beta}) - \frac{1}{3} \Delta^{\mu\nu} \Delta^{\alpha\beta} \]  

(2.11)

is a symmetric direct product of projection operators orthogonal to \( u^\mu \) \cite{36}. The quantities \( \eta \) and \( \zeta \) multiplying each term are hydrodynamic transport coefficients. They are free parameters of the theory describing fundamental dynamical properties of the fluid.

**Remark.** When discretized on a grid, the first-order Navier-Stokes equations generate superluminal hydrodynamic modes which render the numerical scheme unstable. Therefore, in practice, the gradient expansion is implemented at second-order to maintain stability. I’ll introduce the second-order equations of motion later in subsection 5.1.1.2. The second-order equations introduce additional transport coefficients, but it is reasonable to expect their effect on the system dynamics to be much smaller than the first-order coefficients. Indeed, it has been shown, for example, that the system is relatively agnostic to the value of the second-order relaxation time transport coefficient \( \tau_\pi \) \cite{37, 38}.

### 2.1.2.2 First-order transport coefficients

The shear viscosity \( \eta \) and bulk viscosity \( \zeta \) (natural units fm\(^{-3}\)) describe the fluid’s dissipative corrections at leading order. Determining these transport coefficients for QCD matter is therefore a primary goal of fundamental importance. Both coefficients are generally
expected to depend on the temperature $T$ and baryon chemical potential $\mu_B$.

In the hydrodynamic equations, the viscosities appear as dimensionless ratios, $\eta/s$ and $\zeta/s$, where $s$ is the fluid entropy density. These so-called specific viscosities are generally more interesting and meaningful than the unscaled $\eta$ and $\zeta$ values, because they describe the magnitude of stresses inside the medium relative to its natural scale.

**Shear viscosity** The shear viscosity $\eta$ applies a force that opposes shearing flows in the fluid (see figure 2.3), converting the damped motion to heat. Microscopically, it describes how well the fluid transmits momentum across adjacent layers of fluid flow. In weakly-coupled kinetic theory, the specific shear viscosity $\eta/s$ relates to the inter-particle mean free path $\lambda_{\text{mfp}}$ [39–41]:

$$\eta/s \propto T\lambda_{\text{mfp}}. \quad (2.12)$$

Therefore a larger mean free path (weaker coupling) corresponds to a larger value of $\eta/s$. Conversely, in the strongly-coupled limit, $\eta/s$ vanishes and the matter behaves like a “perfect fluid” with minimal resistance to shearing flow. It’s important to note that quasi-particle descriptions of the fluid only make sense up to some maximum coupling strength. Beyond this point, the particles’ mean free path $\lambda_{\text{mfp}}$ becomes smaller than their de Broglie wavelength $1/T$, at which point the notion of quasi-particles is ill defined [36]. Such strongly-coupled systems are thus fundamentally field-like.

**Bulk viscosity** The bulk viscosity $\zeta$ introduces an effective pressure $\Pi$ that modifies the ideal pressure $P$. When the local fluid velocity divergence $\partial_\mu u^\mu$ is positive, this effective pressure is negative and *vice versa*. The bulk viscosity therefore opposes radial expansion and compression (see figure 2.3). Microscopically, the mechanisms that explain bulk viscosity are complicated. However, they generally relate to a certain reconfiguration energy needed for the fluid to expand or contract. The bulk viscosity of a diatomic gas, for example, is nonzero due to the exchange of molecular energy between translational and rotational
degrees of freedom [42]. For scale invariant\(^1\) theories, the bulk viscosity of the system must vanish. However, QCD is known to break scale invariance, particularly near the QGP phase transition, so the bulk viscosity of QCD matter could be large near \(T_c\).

### 2.1.2.3 Jet & hard-probe interactions

Hydrodynamic transport coefficients describe the medium's bulk interactions among its constituents. It’s also interesting to study interactions between the hydrodynamic medium and highly energetic probes that are initially far from equilibrium. For example, suppose I shoot an energetic quark through an infinite brick of equilibrated QGP matter. There are many interesting questions that I might ask, for example:

- How does the quark scatter inside the medium and lose energy?
- How does the quark deflect perpendicular to its direction of motion and diffuse inside the medium?
- What is the path-length dependence of its energy loss?
- How does the medium absorb the energy that is lost by the quark?

These types of questions broadly pertain to a subfield of the QGP research effort dedicated to studying jets and hard-probes. The term jet refers to a highly energetic cone of hadrons and other material ejected by an initial hard-scattering process, while the term hard-probe usually refers to a single energetic particle (possibly inside a jet), e.g. a high-momentum charm quark which traverses the medium. This subject matter is beyond the scope of the present work, so I will not delve into it here. For an overview, see [43–45]. Nevertheless, for the sake of completeness, I’ll describe a few of the primary quantities that jet and hard-probe studies seek to measure. These coefficients are similar in importance to the specific shear viscosity \(\eta/s\) and bulk viscosity \(\zeta/s\) used to quantify the properties of bulk matter interactions.

\(^1\)A scale invariant system is one that appears self-similar at all scales. For example, an equilateral triangle is scale invariant.
The majority of interactions between the probe and the medium are soft small-angle scatterings which each transfer a small amount of momentum from the fluid to the probe such that the fractional change of the probe’s momentum is small. The probe’s response to these soft kicks is summarized by the Fokker-Planck equation [46, 47]. Much like hydrodynamics, the Fokker-Planck equation introduces several transport coefficients which specify important properties of the probe-medium interaction.

**Drag coefficient**  One fundamental measure of the probe-medium interactions is the longitudinal drag coefficient [48]

\[ \eta_D = -\frac{1}{p_L} \frac{dp_L}{dt}, \]  
\[ (2.13) \]

where \( p_L \) is the longitudinal component of the probe momentum \( p \). This quantity measures the percentage longitudinal momentum loss per unit time. It is sensitive to the stopping power of the medium, and hence the coupling strength between the probe and the locally equilibrated QGP.

**Longitudinal momentum broadening**  The second-moment of the longitudinal momentum transfer distribution is quantified by the longitudinal broadening coefficient [48]

\[ \dot{q}_L = \frac{d}{dt} \langle (\Delta p_L)^2 \rangle, \]  
\[ (2.14) \]

defined as the typical longitudinal momentum kick squared per unit time incurred by the probe as it traverses the medium.

**Transverse momentum broadening**  Perhaps the most studied transport parameter is the transverse momentum broadening coefficient \( \dot{q} \), defined as the typical transverse momentum kick squared per unit time incurred by a jet or hard-probe as it traverses the QGP medium [45, 48, 49]:

\[ \dot{q} = \frac{d}{dt} \langle (\Delta p_{\perp})^2 \rangle. \]  
\[ (2.15) \]
It is expected to measure important properties of hot and dense QCD matter, such as its coupling strength (strong vs weak) and its constituent nature (quasi-particles vs non-localized fields) [50]. As such, it is considered a fundamental QCD quantity of primary interest.

2.2 Hadron collider experiments

To this point, I’ve described the QGP largely theoretically, as something believed to exist based on our current knowledge of QCD. How do we know that it actually exists? The primary experimental evidence for the QGP’s existence is provided by ultrarelativistic nuclear collisions conducted at RHIC and the LHC which I mentioned briefly. These facilities are massive, each involving thousands of scientists and numerous nuclear collision experiments.

**Relativistic Heavy-ion Collider (RHIC)** This circular accelerator collides primarily heavy-ions, but also protons and light-ions, at center-of-mass energies per nucleon pair\(^2\) ranging from \(\sqrt{s_{NN}} = 7.7\) to 200 GeV [51]. It is a lower energy collider than the LHC, but it has several unique advantages which make it an excellent probe of the QGP. For instance, it supports longer heavy-ion operation times, and its beam is highly configurable, enabling researchers to study numerous collision partners and beam energies.

**Large Hadron Collider (LHC)** Like RHIC, the LHC is a large circular hadron collider. A distinctive feature of the LHC is its unprecedented beam energy. To date, it has run proton-proton collisions up to \(\sqrt{s_{NN}} = 13\) TeV [52–55] and heavy-ion collisions up to \(\sqrt{s_{NN}} = 5.02\) TeV [56–58]. This is over an order of magnitude larger than the highest energies achieved at RHIC. Heavy-ion collisions, however, are a smaller fraction of the overall physics program at the LHC compared to RHIC, so fewer collision systems and beam energies have been studied.

The experimentalists running these colliders are able to directly control two quantities:

---

\(^2\)Beam energies are commonly measured using \(\sqrt{s_{NN}}\), equal to the total energy of each colliding nucleon pair in its center-of-mass frame.
the species of the colliding nuclei and the energy of the collision. They fix these quantities, accelerate two counter-rotating circular beams of nuclei, and perform measurements on the random collisions that occur between the accelerated ions. Isolated events are then selected from the stream of detector activity, using an experimental trigger to identify the existence of individual inelastic nuclear-nuclear collisions. These raw unfiltered events form a minimum-bias sample, i.e. an unbiased subsample drawn from the population of all equal probability inelastic collision events.

The nuclear collision events are measured by one of several detectors situated on each beam. Each detector is a large apparatus wrapped around the symmetry axis of the beam pipe which captures the flux of particles generated by each collision event. Both facilities have multiple detectors, each managed by an independent experimental collaboration sharing the name of the detector. RHIC has the BRAHMS, PHENIX, PHOBOS, and STAR detectors, while the LHC has ALICE, ATLAS, CMS, and LHCb.

These detectors vary in their design. Each is specifically optimized to perform a certain task. For example, ALICE (A Large Ion Collider Experiment) is optimized to detect the tens of thousands of particles produced by a lead-lead collision, specifically those particles emitted with low momentum. Figure 2.4 shows a computer rendering of the ALICE experiment. Notice the two people in the lower left corner to appreciate the sense of scale.
2.2.1 Event properties

The properties of the particles produced by each collision are determined using an ensemble of particle trackers and energy calorimeters layered around the nominal interaction vertex. These detector components allow the experiments to measure properties of each particle (when possible) such as its momentum, charge, mass, and particle type. These raw particle properties are then post-processed into experimental observables which describe features of the event sample.

2.2.1.1 Kinematic variables

Consider a collision in which the nuclei move through a beam aligned with the $\hat{z}$ direction. In high-energy particle physics, it is common to specify each particle’s four-momentum in a transformed coordinate system:

$$(E, p_x, p_y, p_z) = (m_T \cosh y, p_T \cos \phi, p_T \sin \phi, m_T \sinh y), \quad (2.16)$$

$$m_T = \sqrt{E^2 - p_z^2}, \quad (2.17)$$

$$y = \frac{1}{2} \log \frac{E + p_z}{E - p_z}, \quad (2.18)$$

$$p_T = \sqrt{p_x^2 + p_y^2}, \quad (2.19)$$

$$\phi = \text{atan2}(p_y, p_x), \quad (2.20)$$

where $m_T$ is the particle’s transverse mass, $y$ is its rapidity, $p_T$ is its average transverse momentum, and $\phi$ is its azimuthal angle in the plane orthogonal to the beam axis.

Note, the transverse mass $m_T$ and the rapidity $y$ both require knowledge of the particle’s total energy $E$ which depends on its mass $m$. This information is often inaccessible for technical reasons, so typically the experiments replace the rapidity $y$ with a similar quantity known as the pseudorapidity $\eta$. It is defined as

$$\eta = -\log[\tan(\theta/2)] = \frac{1}{2} \log \frac{|\mathbf{p}| + p_z}{|\mathbf{p}| - p_z}, \quad (2.21)$$
where $\theta$ is the momentum vector’s polar angle with respect to the beam axis, i.e. $\cos \theta = \frac{p_z}{|p|}$. For massless particles, the rapidity $y$ and the pseudorapidity $\eta$ are equivalent. They are also equivalent at midrapidity, i.e. for $y = \eta = 0$. This quantity is convenient because the particle’s polar angle $\theta$ is easily measured inside the detector.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure2.5}
\caption{Diagram illustrating the relationship between the pseudorapidity $\eta$ and the polar angle $\theta$ [60].}
\end{figure}

Figure 2.5 visualizes the relationship between the particle’s pseudorapidity $\eta$ and its polar angle $\theta$. For $y = \eta = 0$, the particle emerges orthogonal to the beam axis. This two-dimensional $(x, y)$ plane is thus commonly referred to as the transverse plane. Meanwhile, for $\eta = \infty$, the particle remains inside the beam pipe. Thus, due to detector limitations, it is only possible for the experiments to measure particles out to some maximum rapidity.

### 2.2.1.2 Collision centrality

Once the beam is running, there is no way to control the orientation of the collisions. Each pair of nuclei collides randomly, separated by an impact parameter $b$ in the transverse plane, defined as the distance between the two nuclei’s centers of mass at the moment of closest approach; see figure 2.6.

In principle, it would be useful to measure the collision’s properties as a function of the impact parameter $b$. However, this quantity cannot be directly measured, so it is typically replaced with a related observable known as centrality.

The collision centrality is defined by sorting all events in a minimum-bias event sample according to some measure of the underlying event activity (see below). Once the events
are sorted, they are partitioned into equal sized bins, where each bin is associated with some percentage of the overall event sample. For example, if the events are partitioned into $n = 10$ equal sized bins by their event activity, then the bin with the highest event activity is the 0–10% centrality class.

The definition of the underlying event activity used to sort the events varies from experiment to experiment. A common choice is to measure some proxy for the event’s charged-particle yield in a given rapidity window. For example, the ALICE experiment commonly defines the collision centrality according to the sum of amplitudes in the detector’s VZERO scintillators, covering $2.8 < \eta < 5.1$ (VZERO-A) and $-3.7 < \eta < -1.7$ (VZERO-C), signals which are monotonically related to the charged-particle yield [61]. Figure 2.7 shows an example of this centrality binning procedure applied to Pb-Pb collision data measured by the ALICE experiment [62].

### 2.2.2 Signatures of the quark-gluon plasma

Now that I’ve broadly motivated and described heavy-ion collision experiments at RHIC and the LHC, I want to summarize some of their key results, particularly those results which evidence the production of the QGP. This subsection is not meant to be an exhaustive list; doing so would require far more than a few pages. Rather, these are several experimental observations that are commonly cited when discussing QGP formation. Ultimately, I will explain at the end of the chapter that these features are collectively explained by a standard hydrodynamic model of relativistic heavy-ion collisions. Once I’ve motivated and explained this model, I’ll be able to frame the central problem addressed by this dissertation.
2.2.2.1 Thermal particle yields

One intriguing indicator that ultrarelativistic heavy-ion collisions produce QGP is provided by statistical hadronization models. These models calculate hadron yields in nuclear collisions by sampling particles from a common chemical freeze-out surface at fixed temperature $T$ and baryon chemical potential $\mu_B$, i.e. by sampling from an emitter in thermal equilibrium. The observed particle yields are consequently assumed to arise from the decay of fully equilibrated hadronic matter comprising all known hadron states. The model is then calibrated to optimally fit the data by adjusting the temperature $T$ and chemical potential $\mu_B$ of the emitter, together with its freeze-out volume. A detailed description of this approach is presented in [63].

The left side of figure 2.8 shows the particle yields predicted by such a model [64], calibrated on and compared to Pb-Pb collision data at $\sqrt{s_{NN}} = 2.76$ TeV measured by ALICE [65–70]. The fit obtains a hadronization temperature $T = 156$ MeV and baryon chemical potential $\mu_B = 0$ MeV, which is in perfect agreement with the location of the pseudocritical transition temperature $T_c = 156.5 \pm 1.5$ MeV at $\mu_B = 0$ predicted by lattice
QCD [27]. Meanwhile, a study of the energy-dependence of the fit parameter \( T \) presented on the right-side of figure 2.8 shows that the chemical freeze-out temperature increases as a function of beam energy before flat-lining at \( \sqrt{s_{NN}} \sim 10 \) GeV [71]. This suggests that the hadron resonance gas cannot be heated above some maximum temperature, presumably the temperature of the QGP phase transition.

2.2.2.2 Collective flow

Perhaps the most famous observation associated with QGP formation is the existence of collective flow. Prior to the first ultrarelativistic heavy-ion collisions at RHIC, many believed that QGP would behave like a weakly-coupled gas characterized by a large mean free path. Assuming particle production occurs independently at different points inside the heavy-ion collision, this conjecture would imply final hadron yields that are weakly correlated with respect to the azimuthal angle \( \phi \). The only significant azimuthal correlations would arise from jets and other hard scatterings which produce back-to-back showers of particles close to midrapidity.

However, the first measurements at RHIC revealed a very different picture of the col-

![Figure 2.8](image_url)

**Figure 2.8** Left: Particle yields for various hadrons predicted by the statistical hadronization model calibrated to fit Pb-Pb collisions at \( \sqrt{s_{NN}} = 2.76 \) TeV [64] using experimental data from ALICE [65–70]. Right: Best fit statistical hadronization parameter \( T \) as a function of beam energy \( \sqrt{s_{NN}} \) [71].
The particles produced by each collision were found to be strongly correlated with respect to the azimuthal angle $\phi$, and these correlations persisted far from midrapidity [72] at odds with weakly-coupled predictions [73, 74]. The signal was consistent with a strongly-coupled picture of the collision, in which the QGP flows like a nearly inviscid liquid.

To understand how hydrodynamic flow gives rise to these correlations, consider a generic collision between two highly relativistic nuclei as shown in figure 2.9. When these nuclei collide, they generate an initial transverse energy density profile (left) which is spatially deformed due to the “almond” shape of the overlap region at nonzero impact parameter (see figure 2.6) and the fluctuations of nucleon positions inside each nucleus. These spatial inhomogeneities create pressure gradients along the radial direction which vary as a function of the azimuthal angle $\phi$, producing stronger radial expansion along some directions and less along others. This generates an azimuthally anisotropic flow field which preferentially emits particles in the direction of strongest fluid flow (middle), imparting this signal on the final azimuthal hadron distribution $dN/d\phi$ (right). In essence, hydrodynamics converts spatial anisotropy into momentum anisotropy, which also shows up in the detector as a particle yield anisotropy.

Experimentally, this yield anisotropy is quantified by expanding the azimuthal particle distribution $dN/d\phi$ as a function of the azimuthal angle $\phi$: $dN/d\phi \propto 1 + A_2 \cos(2\phi) + A_4 \cos(4\phi) + \cdots$ where $A_n$ are the azimuthal harmonics. The measured $A_n$ values are consistent with the strongly-coupled hydrodynamical picture.
distribution as a Fourier series \[75–77\]

\[
\frac{dN}{d\phi} \propto 1 + 2 \sum_{n=1}^{\infty} v_n \cos[n(\phi - \Psi_n)],
\]  

(2.22)

where \(\Psi_n\) is the phase or “event plane” angle, equal to the direction of maximum final-state particle density. Here the number \(n\) indexes the order of the harmonic. The first harmonic \(v_1\) is called directed flow, the second harmonic \(v_2\) elliptic flow, the third harmonic \(v_3\) triangular flow, and so on.

These coefficients are calculated using the relation

\[
v_n = \langle \langle \cos[n(\phi - \Psi_n)] \rangle \rangle,
\]

(2.23)

where the double angular brackets mean averaging over all particles in a given event, then averaging over all events in a given event class selected to satisfy certain centrality, rapidity, and transverse momentum requirements. When the flow is calculated as a function of transverse momentum \(p_T\) using narrow \(p_T\)-bins it is called differential flow, and when it is calculated using all particles irrespective of their \(p_T\) over a wide kinematic range, it is referred to as integrated flow.

Remarkably, early RHIC experiments showed that the produced collectivity is best understood if it is assumed to develop from a flowing liquid of deconfined quarks, rather than a super hot gas of hadrons \[78–80\]. This preference for quark degrees of freedom is illustrated in figure 2.10.

First, look at the far left plot which shows the \(p_T\)-differential elliptic flow \(v_2(p_T)\) for various hadron species in minimum-bias Au-Au collisions at RHIC. The mass splitting visible among the different species is a characteristic signature of hydrodynamic flow. If the mass-ordering of \(v_2\) is driven by hydrodynamic pressure gradients, then the differential \(v_2\) of each particle should scale with the transverse kinetic energy \(KE_T = m_T - m\), where \(m_T\) is the particle’s transverse mass.

The second figure from the left shows the differential elliptic flow \(v_2\) plotted against the
transverse kinetic energy $KE_T$. Notice how the elliptic flow curves split into two branches. The upper branch contains all the baryons (three-quark states) while the lower branch contains all the mesons (two-quark states). Presumably, the baryons carry more elliptic flow because they carry one extra quark than the mesons.

Finally, look at the figure on the far right which shows both these quantities divided by the number of valence quarks in each particle. Suddenly, all of the differential flow measurements collapse to a single curve. This signifies that the elliptic flow is carried by individual quarks, and that the elliptic flow is transmitted from the quarks to the hadrons by the hadronization process after the flow has already developed. This strongly evidences the creation of a fluid comprised of free flowing quarks.

2.2.2.3 Jet quenching

When two nuclei collide at ultrarelativistic energies, the quarks and gluons inside the nuclei occasionally scatter at large angles, producing two or more energetic partons carrying very large transverse momenta, anywhere from one to several orders of magnitude larger than the typical transverse particle momentum inside the event. These energetic partons penetrate the produced QGP medium and fragment into softer particles, emerging from the interaction region as columnated sprays of nuclear matter known as jets.
As each penetrating jet moves through the QGP medium, it images the properties of the produced matter analogous to an x-ray radiograph. If the QGP is strongly-coupled, each jet is expected to lose significant energy to the medium via induced gluon radiation such that the final jet is strongly modified or “quenched”. The existence of jet quenching is therefore a key prediction of a strongly-coupled QGP. Presumably, this effect should depend on fundamental properties of interest such the color-charge density of the QGP and its short-distance structure [81].

Naturally, if a hard-scattering process produces back-to-back jets near the periphery of the fireball, with one jet moving into the medium and the other moving out of it, then the jet moving into the medium (away-side jet) should be more strongly modified than the jet moving out of it (near-side jet). This setup is depicted in figure 2.11. One way to test this hypothesis, is to measure two-particle azimuthal correlations, using a high-$p_T$ trigger particle to orient the correlation function relative to the dominant jet.

Figure 2.12 shows such a test applied to $p$-$p$, $d$-$Au$, and $Au$-$Au$ collisions at $\sqrt{s_{NN}} = 200$ GeV by the STAR collaboration [82]. The quantity plotted is the two-particle azimuthal distribution

$$D(\Delta \phi) = \frac{1}{N_{\text{trigger}}} \frac{dN}{d(\Delta \phi)},$$

constructed by correlating a high-$p_T$ trigger particle with angle $\phi_a$ and transverse momen-
Figure 2.12  Top: Two-particle azimuthal distributions for minimum-bias and central \(d\)-Au collisions, and for \(p-p\) collision. Bottom: Comparison to the distribution for central Au-Au collisions [82].

The constant \(N_{\text{trigger}}\) is the number of selected trigger particles, and the quantity \(\Delta \phi = \phi_a - \phi_b\) is the azimuthal angle between each particle pair.

First, look at the top panel of the figure which shows this two-particle azimuthal distribution for minimum-bias and central \(d\)-Au collisions, and for minimum-bias \(p-p\) collisions. All three distributions show a sharp peak at \(\Delta \phi = 0\), corresponding to particles that are emitted at small angles with respect to the high-\(p_T\) trigger particle. There is also a second peak centered on \(\Delta \phi = \pi\), which is somewhat smaller in stature and smeared out. Now look at the bottom figure, which shows the \(p-p\) and \(d\)-Au two-particle azimuthal distributions compared to the same distribution for central Au-Au collisions. In the Au-Au system, the \(\Delta \phi = 0\) peak is clearly visible, but the \(\Delta \phi = \pi\) peak is absent.

This result is naturally explained by the existence of a strongly-coupled QGP. The peak at \(\Delta \phi = 0\) is produced by particles emitted from the near-side jet, while the peak at \(\Delta \phi = \pi\) is produced by the away-side jet. In a di-jet event, the initial partons are produced back-to-back so the jets are separated by 180°. The near-side jet is produced closer to the surface...
of the QGP fireball, so it escapes with little modification, while the away-side jet plows into
the medium where it is strongly quenched. Given that the jet loses several GeV of energy
as it traverses a relatively short distance, this observation corroborates that the matter
is strongly-coupled. Meanwhile, independent studies show that electromagnetic probes,
e.g. direct photons and $Z$ bosons, show no evidence of jet quenching [83, 84]. Hence, the
opaqueness of the matter appears specific to particles which interact via the strong force,
consistent with the picture of QGP formation.

2.3 Hydrodynamic computer simulations

Hydrodynamic computer simulations are a powerful tool to refine our current understanding
of hot and dense nuclear matter. These simulations recreate entire nuclear collision events,
exactly as they are believed to occur inside the detector, and output virtual particles that
can be post-processed and analyzed using the same methods applied to the experimen-
tal data. Important QGP medium parameters, e.g. the QGP specific shear viscosity $\eta/s$
and specific bulk viscosity $\zeta/s$, are then extracted by tuning their values to maximize the
agreement of the simulation with experiment.

Hydrodynamic computer simulations vary in their exact implementation, but they gen-
erally follow a canonical framework which is constantly being updated and refined. This
section briefly summarizes the current picture of the hydrodynamic framework and explains
how it can be used to extract QGP transport coefficients. Finally, I conclude by discussing
the largest obstacle limiting the precision of these simulation-based extractions, the so-called
QGP “initial condition problem”, which is the subject of this dissertation research.

2.3.1 Space-time picture of a single event

Consider two nuclei barreling toward each other at nearly the speed of light inside the beam
pipe, as visualized by the space-time diagram figure 2.13.
2.3.1.1 Preparing the nuclei

Hydrodynamic computer simulations begin by preparing the two ions for a simulated collision. The ions are constructed by sampling their three-dimensional nucleon densities, mimicking the spatial fluctuations seeded by the ultimate collapse of each nuclear wave function. The nuclei are given a random rotation and impact parameter offset and boosted to their respective beam velocities, causing each ion to appear as a Lorentz contracted disk in the stationary lab frame. The Lorentz factor is about half the value of the center-of-mass energy per nucleon pair when expressed in units of GeV. Thus for a collision at \( \sqrt{s_{NN}} = 200 \) GeV, each nucleus is contracted by \( \gamma \sim 100 \) along its direction of motion.

2.3.1.2 Initial state

If the sampled impact parameter offset is sufficiently small, the two nuclei interpenetrate and briefly overlap. This convolves the three-dimensional density of each nucleus, depositing tremendous energy in the process. The produced secondary matter fills the space between the receding ion fragments and forms an extended tube of deconfined quarks and gluons characterized by very small baryon density. This initial overlap process is so brief, \( \Delta \tau_{\text{overlap}} \lesssim 0.1 \text{ fm}/c \), that computer models commonly assume it to happen instantaneously. Simulations therefore start by calculating the matter’s energy or entropy density at some early time \( \tau_0 \lesssim 1 \text{ fm}/c \) shortly after the nuclei interpenetrate. Alternatively, more advanced simulations calculate all components of the initial stress-energy tensor \( T^{\mu\nu} \) [85, 86].
2.3.1.3 Pre-equilibrium evolution

The stress-energy tensor of the initially produced matter is locally anisotropic and far from equilibrium. These conditions preclude the direct application of hydrodynamics at very early times $\tau_{\text{hydro}} \lesssim 1 \text{ fm/c}$. Pre-equilibrium transport models based on strongly and weakly-coupled effective field theories are therefore used to evolve the system forward in time until the local stress-energy tensor more closely resembles the form predicted by second-order hydrodynamics [85, 87–89]. Computer simulations that properly model the pre-equilibrium stage of the collision are a relatively recent development, so often hydrodynamic models evolve the system to the hydrodynamic starting time using simple free-streaming approximations [90, 91] or they opt to skip the pre-equilibrium stage entirely.

2.3.1.4 Hydrodynamic evolution

The pre-equilibrium phase is then matched to viscous hydrodynamics to simulate the space-time evolution of the QGP liquid. The hydrodynamic simulation is provided initial conditions for the energy density $e$, fluid velocity $u^\mu$, and shear and bulk viscous corrections $\pi^{\mu\nu}$ and $\Pi$, an EoS from lattice QCD, and values for the temperature-dependent QGP transport coefficients $\eta/s$ and $\zeta/s$.\(^3\) The hydrodynamic equations of motion are then solved numerically on a discretized grid.

Hydrodynamic simulations vary in their approximations and numerical schemes. One common variant of the framework applies a simplifying symmetry known as boost-invariance which asserts Lorentz invariance to boosts along the beam direction. In his seminal paper on relativistic heavy-ion collisions, Bjorken argued that boost-invariance should hold for ultrarelativistic heavy-ion collisions, since the nuclei are already so highly boosted ($\gamma \gtrsim 100$), that the collision will appear essentially identical to any observer in a moderately boosted reference frame [9].

This assumed symmetry reduces (3+1) space-time dimensions to (2+1) dimensions and

\(^3\)If the initial fluid velocity $u^\mu$ and viscous corrections $\pi^{\mu\nu}$ and $\Pi$ are not provided by the initial condition model, they are typically set to zero. This approximation is known as static initialization.
dramatically simplifies the hydrodynamic equations of motion. Boost-invariant hydrodynamic codes therefore run an order of magnitude faster than their three-dimensional counterparts. In this dissertation, I perform calculations using both boost-invariant and three-dimensional hydro codes. Boost-invariance generally works well near midrapidity $|\eta| \lesssim 1$ [92, 93], but it is a poor approximation if used to analyze particles detected at moderate to large rapidities.

2.3.1.5 Particlization and hadronic evolution

After $O(10)$ fm/c of hydrodynamic evolution, the medium cools past the QGP transition temperature and freezes into individual hadrons. These emitted particles continue to scatter and decay, then eventually decouple and free stream into the detector. Hydrodynamic mean-field approximations begin to break down as the system disintegrates, so the hydrodynamic evolution is commonly spliced onto microscopic kinetic theory which is better suited to handle the system’s non-equilibrium break-up.

This hybrid model prescription [20, 22, 23] converts the fluid to hadrons assuming thermal particle emission from a pre-specified switching isotherm $T_{\text{switch}}$, typically required to lie near the pseudocritical temperature $T_c$ in order to fit the observed particle yields. Once the fluid is “particlized”, its subsequent interactions are modeled by the Monte Carlo implementation of the Boltzmann equation which follows each hadron microscopically until the last interactions cease and the system freezes out, yielding a list of final particle data for each event. For a visualization, see figure 2.14 which shows several snapshots of a typical event simulated using the hybrid model framework.

2.3.2 Extracting QGP transport coefficients

The QGP transport coefficients can be inferred from hydrodynamic simulations, by analyzing their effect on bulk particle properties. Typically, this is accomplished by identifying key observables which are particularly sensitive to a given parameter of interest. The parameter’s true value is then inferred by adjusting its assumed value until the simulation
optimally agrees with experiment.

One notable example of this procedure, is the use of the flow harmonics $v_n$ to constrain the QGP specific shear viscosity $\eta/s$. Recall that these harmonics $v_n$ measure the final particle distribution’s anisotropy with respect to the azimuthal angle $\phi$. Elliptic flow $v_2$ measures its ellipticity, triangular flow $v_3$ measures its triangularity, and so on.

These final-state momentum anisotropies originate as initial-state spatial anisotropies. Crudely speaking, hydrodynamics converts spatial anisotropy into flow anisotropy. For example, if the initial state is elliptically deformed, its hydrodynamic evolution will generate elliptic flow (see figure 2.15). Similarly, triangular profiles generate triangular flow, quadrangular profiles generate quadrangular flow, etc.

![Figure 2.14](image1)

**Figure 2.14** Hybrid model computer simulation of a typical heavy-ion collision event. Figure is adapted from a visualization originally constructed by H. Petersen for the MADAII collaboration.

![Figure 2.15](image2)

**Figure 2.15** Elliptically deformed spatial profile (left) generates elliptically deformed momentum profile (right).

Much like the final momentum anisotropy, the initial spatial anisotropy can be quantified
by its azimuthal harmonics

\[ \varepsilon_n \varepsilon^{i \Phi} = \frac{\int dx \, dy \, r^n e^{i \Phi} \rho(x, y)}{\int dx \, dy \, r^n \rho(x, y)}, \tag{2.25} \]

where \( \varepsilon_n \) is the eccentricity harmonic of order \( n \), \( \Phi \) is its phase angle, and \( \rho \) is the density profile of interest, typically assumed to be the event’s transverse energy or entropy density. Generally speaking, initial profiles with large \( \varepsilon_n \) generate large \( v_n \). Linear scaling \( v_n \propto \varepsilon_n \) is observed for \( n = 2, 3 \) in heavy-ion collisions to good approximation [94–96], but the scaling breaks down for \( n > 3 \) due to non-linear mode mixing [94].

The QGP specific shear viscosity \( \eta/s \) governs the efficiency with which the hydrodynamic evolution converts spatial anisotropy into momentum anisotropy. Hence, it is directly related to the ratio \( v_n/\varepsilon_n \) which quantifies the flow that’s produced per unit eccentricity. Small values of \( v_n/\varepsilon_n \) correspond to large shear viscosities and large values of \( v_n/\varepsilon_n \) correspond to small shear viscosities. Note, the flow anisotropies \( v_n \) are directly measurable, while the initial state eccentricities \( \varepsilon_n \) are not; they can only be estimated theoretically. Therefore, \( \eta/s \) extractions are directly limited by one’s ability to calculate the QGP initial conditions precisely.

Figure 2.16 shows the eccentricity-scaled elliptic flow \( v_2/\varepsilon_2 \) (second harmonic) plotted versus the charged-particle density per unit overlap area \((1/S) dN_{ch}/dy\) for Au-Au collisions at \( \sqrt{s_{NN}} = 200 \text{ GeV} \) using two different models for the QGP initial conditions (left and right plots). The symbols are calculations using the experimentally measured elliptic flow \( v_2 \) and charged-particle density \( dN_{ch}/dy \), while the colored lines are constructed using simulated values for these quantities, calculated for several different values of the QGP specific shear viscosity \( \eta/s \). The eccentricity \( \varepsilon_2 \) and root-mean-square overlap area \( S = \sqrt{\langle x^2 \rangle \langle y^2 \rangle} \), meanwhile, are provided by the respective initial condition model. The panel on the left shows an extraction using MC-KLN initial conditions [101, 102], and the panel on the right shows an extraction using MC-Glauber initial conditions [103]. It’s not important that I describe these models in detail at the moment—suffice to say, each initial condition model predicts different eccentricities \( \varepsilon_n \).
Figure 2.16 Eccentricity-scaled elliptic flow $v_2/\varepsilon_2$ plotted versus the charged-particle density per unit overlap area $(1/S) dN_{ch}/dy$ [97]. Symbols are constructed using the experimentally measured elliptic flow mean $\langle v_2 \rangle$ [98] and two-particle cumulant $v_2\{2\}$ [99], along with the charged-particle density $dN_{ch}/dy$ [100], while colored lines are constructed using simulated values for these quantities, calculated for several values of the specific shear viscosity $\eta/s = 0.0$, 0.08, 0.16, and 0.24. The eccentricity $\varepsilon_2$ and overlap area $S = \sqrt{\langle x^2 \rangle \langle y^2 \rangle}$ are obtained from the initial condition model. Results are shown for the MC-KLN initial conditions [101, 102] (left) and MC-Glauber initial conditions [103] (right).

The ratio $v_2/\varepsilon_2$ can be thought of as a “ruler” which measures the fluid’s specific shear viscosity $\eta/s$. The experimentally extracted viscosity is read from the plot by matching the symbols with the colored lines, each corresponding to a specific value of $\eta/s$. Hence, the extraction based on MC-KLN initial conditions obtains $\eta/s \sim 0.2$, while the extraction based on MC-Glauber initial conditions obtains $\eta/s \sim 0.08$ (each with large errors). The authors of the study were therefore able to conclude that the QGP specific shear viscosity for $T_c < T \lesssim 2T_c$ lies within the range $0.08 < (\eta/s)_\text{QGP} < 0.20$, with the remaining uncertainty arising from insufficient theoretical control over the initial source eccentricity $\varepsilon_2$. Consequently, the primary means to improve this estimate is to reduce the model’s systematic initial condition uncertainty. This is one example of what I refer to as the initial condition problem.

The initial condition problem is, of course, far more general than the relationship between the elliptic flow, shear viscosity, and eccentricity. The initial conditions strongly affect essentially every model output, so their uncertainty is strongly correlated with the uncertainty of the inferred medium properties. For example, if a given initial condition model...
predicts QGP energy densities which are too compact, the simulation will expand more explosively than it should and require an artificially large bulk viscosity to compensate.

2.4 The initial condition problem

To date, there exist numerous theoretical models for the QGP initial conditions, of which the MC-Glauber and MC-KLN models are but two examples. Different initial condition models generally predict different energy density and flow velocity profiles, so their hydrodynamic evolutions consequently prefer different values of the QGP transport coefficients. Studies of the initial condition and QGP medium properties are thus inextricably linked.

The most straightforward procedure to reduce the list of mutually incompatible theory calculations is to validate candidate models using sensitive experimental observables. Each initial condition model typically includes several free parameters which can be tuned to selectively fit one or two observables at a time, so it is important to test models self-consistently using a large cross section of the available experimental data. Presumably, the correct model will reproduce all observables within the realm of its applicability, assuming the subsequent hydrodynamic evolution is well understood.

2.4.1 Ab initio theory calculations

Over the last decade, tremendous progress has been made in understanding the initial stages of ultrarelativistic nuclear collisions. The discovery process has been accelerated by several important theoretical developments, resulting in a handful of credible bottom-up initial condition approaches based on approximations of QCD and related field theories. This subsection summarizes two such models which have demonstrated broad agreement with the experimental data, far surpassing the MC-Glauber and MC-KLN models mentioned previously. I should emphasize that this is not meant to be an exhaustive list of all credible initial condition models, and I apologize to the authors whose work is not discussed.
2.4.1.1 IP-Glasma model

One *ab initio* model which successfully describes a large number of experimentally measured bulk observables is IP-Glasma [85, 104]. This model obtains the QGP initial conditions from Color Glass Condensate effective field theory, by combining the impact-parameter dependent saturation model (IP-Sat) [105, 106] with the classical Yang-Mills description of initial gluon fields. Color Glass Condensate (CGC) effective field theory is a general theoretical framework which describes the small-$x$ behavior of the hadronic wavefunction in QCD [107]. In this approach, the system’s large-$x$ color-charge degrees of freedom act as static sources for small-$x$ gauge fields $A^\mu$. At high energies, the density of produced partons at small-$x$ becomes large, leading to a saturation of the parton distribution function which occurs at the characteristic saturation momentum $Q_s$.

The IP-Glasma model starts by sampling the positions of nucleons within each nucleus from a Fermi distribution (more on this later). Once the nucleon positions are known, the IP-Sat model provides the saturation scale $Q_s^2(x, b_\perp)$ as a function of Bjorken $x$ and the transverse impact parameter $b_\perp$ relative to each nucleon’s center. The color-charge density

\[ 4 \text{Bjorken } x = Q^2/(2p \cdot q) \] is a common variable in deep-inelastic scattering related to the fraction of the proton momentum carried by a certain parton. Here $p$ is the incoming proton momentum, $q$ is its momentum transfer with the probe, and $Q^2 = -q^2$. 

---

**Figure 2.17** Left: Root-mean-square anisotropic flow coefficients $\sqrt{\langle v_n^2 \rangle}$ as a function of centrality calculated from IP-Glasma initialized hydrodynamic simulations (lines) [86] compared to the experimentally measured two-particle flow cumulants $v_n\{2\}$ measured by ALICE (symbols) [108]. Right: Root-mean-square anisotropic flow coefficients $\sqrt{\langle v_n^2 \rangle}$ as a function of transverse momentum $p_T$ (lines) compared to experimental data from ATLAS (symbols) using the event-plane method [109]. All calculations are for $\eta/s = 0.2$ and $\zeta/s = 0$. 

---

37
squared per unit transverse area \( g^2\mu^2 \) is then assumed to be proportional to the saturation scale \( Q_s^2 \).

For a nucleus with \( A \) nucleons, the quantity \( g^2\mu_A^2(x, x_\perp) \) is obtained for each nucleus by adding the color-charge contributed by each nucleon. Provided this mean square color-charge density, random color charges \( \rho^a \) are sampled from the Gaussian distribution

\[
\langle \rho^a_{A(B)}(x_\perp) \rho^b_{A(B)}(y_\perp) \rangle = \delta^{ab}\delta^{(2)}(x_\perp - y_\perp)g^2\mu_{A,B}^2(x, x_\perp), \quad (2.26)
\]

for nucleus \( A \) and \( B \).

After this sampling, the random color-charge distribution of each nucleus is used to calculate the electric and magnetic color fields by solving the classical Yang-Mills equations

\[
[D_\mu, F^{\mu\nu}] = J^\nu, \quad (2.27)
\]
where $F^{\mu\nu}$ is the field strength tensor and $J^{\nu}$ is the color current density, calculated from each Lorentz contracted sheet of boosted color-charge density. Finally, the QGP’s initial energy density $e$ and flow velocity $u^\mu$ are calculated from the produced gluon fields evolved to a pre-specified hydrodynamic starting time shortly after the collision.

IP-Glasma is perhaps best known as the first initial condition model to correctly reproduce the first few harmonics of the azimuthal flow anisotropy $v_n$ generated by heavy-ion collisions [86]. Figure 2.17 shows the root-mean-square anisotropic flow coefficient $\sqrt{\langle v_n^2 \rangle}$ for $n = 2, 3, 4, 5$ plotted as a function of collision centrality (left) and as a function of transverse momentum $p_T$ (right) for Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV compared to experimental data from ALICE [108] and ATLAS [109]. The model provides a superb description of these observables, suggesting that a proper modeling of the eccentricity harmonics $\varepsilon_n$ is achieved.

At the time, this level of agreement with the data was truly unprecedented.

Even more impressive, the IP-Glasma model correctly describes the full probability distribution $P(v_n)$ of each flow harmonic as a function of collision centrality [86]. In other words, the model doesn’t just describe one moment of the flow distribution, it correctly describes its non-trivial shape as well. Figure 2.18 shows IP-Glasma initialized hydrodynamic calculations for the mean-scaled eccentricity distribution $P(\varepsilon_n / \langle \varepsilon_n \rangle)$ and mean-scaled flow distribution $P(v_n / \langle v_n \rangle)$ [86] compared to the corresponding flow distributions measured by ATLAS [110]. The model calculations nicely track the experimental data, validating the assumptions of the framework.

The IP-Glasma initial condition model is generally well tested, and has been compared to numerous other experimental observables at RHIC and LHC energies as well, including e.g. the centrality-dependence of the charged-particle yield and mean transverse momenta [111, 112]. Extractions of $(\eta/s)(T)$ and $(\zeta/s)(T)$ obtained using IP-Glasma initial conditions vary somewhat in the literature due to the specifics of each analysis; however, recent estimates [111, 112] find good agreement with the data using an effective specific shear viscosity $\eta/s \sim 0.095–0.12$ and a temperature-dependent specific bulk viscosity $(\zeta/s)(T)$ which peaks near $T \sim 165–180$ MeV and obtains a maximum value $(\zeta/s)_{\text{max}} \sim 0.24–0.3$. 
2.4.1.2 EKRT model

The recently updated NLO EKRT model, which combines next-to-leading-order (NLO) collinearly factorized pQCD minijet production with a conjecture for low-\(p_T\) gluon saturation, is another highly successful initial condition model named for its original authors Eskola, Kajantie, Ruuskanen, and Tuominen [113, 114]. In this approach, the collision deposits energy in the form of low-\(p_T\) partons (predominantly gluons) and high-\(p_T\) minijets which are separated by a transverse momentum scale \(p_0 \gg \lambda_{\text{QCD}}\).

Consider two nuclei, labeled \(A\) and \(B\), with three-dimensional nuclear densities \(\rho_A\) and \(\rho_B\) respectively. Assume the nuclei collide with impact parameter vector \(b\) in the transverse plane \(x_\perp = (x, y)\). Let \(T_A = \int dz \rho_A(x_\perp, z)\) define the transverse density of nucleus \(A\), and assume \(T_B\) follows accordingly. For a given beam energy \(\sqrt{s_{\text{NN}}}\), the initial transverse-area density of minijet transverse energy, \(dE_T/\text{d}^2x_\perp\), produced perturbatively into a rapidity window \(\Delta y\) above the transverse momentum cut-off \(p_0\) is given by

\[
\frac{dE_T}{\text{d}^2x_\perp} = T_A(x_\perp) T_B(x_\perp - b) \sigma\langle E_T\rangle_{p_0, \Delta y, \beta}, \tag{2.28}
\]

where \(\sigma\langle E_T\rangle_{p_0, \Delta y, \beta}\) is the \(E_T\)-weighted minijet cross section computed from NLO pQCD. This quantity depends on the transverse momentum cut-off \(p_0\), the width of the rapidity interval \(\Delta y\), and a phenomenological parameter \(\beta\) which controls the minimum transverse energy \(E_T\) allowed in \(\Delta y\). For a detailed formulation of \(\sigma\langle E_T\rangle\), see [114, 115].

Here it is assumed that only minijets with transverse momenta \(p_T > p_0\) contribute significantly to \(dE_T/\text{d}^2x_\perp\). Below the transverse momentum cut-off \(p_0\), contributions from \((3 \to 2)\) and higher-order partonic processes begin to dominate conventional \((2 \to 2)\) processes causing the parton density to saturate. This condition leads to the saturation criteria [116]

\[
\frac{dE_T}{\text{d}^2x_\perp} = \frac{K_{\text{sat}}}{\pi} p_0^3 \Delta y, \tag{2.29}
\]

where \(K_{\text{sat}}\) is an unknown normalization constant determined by the fitting the experimentally measured charged-particle density \(dN_{\text{ch}}/d\eta\) using a single narrow centrality interval.
Equations (2.28) and (2.29) are finally equated and solved numerically to determine the transverse momentum cut-off \( p_0 \) where the soft-gluon production saturates. Provided this saturation momentum \( p_{\text{sat}} \equiv p_0 \), the local energy density \( e \) at the local formation time \( \tau_s(x_\perp) = 1/p_{\text{sat}}(x_\perp) \) at midrapidity follows from equation (2.29):

\[
e(x_\perp, \tau_s) = \frac{dE_T(p_{\text{sat}})}{d^2x_\perp} \frac{1}{\tau_s \Delta y} = \frac{K_{\text{sat}}}{\pi} \left[ p_{\text{sat}}(x_\perp) \right]^4.
\]

This energy density is then evolved to a universal proper time \( \tau_0 = 0.2 \text{ fm/c} \) using one-dimensional Bjorken hydrodynamics. The EKRT model does not provide the initial flow velocity \( u^\mu \) or shear corrections \( \pi^{\mu\nu} \) and \( \Pi \), so these additional components are typically set to zero.

Figure 2.19 shows EKRT initialized hydrodynamic calculations for the centrality-dependence of the midrapidity charged-particle density \( dN_{\text{ch}}/d\eta \) (left), identified-particle mean \( p_T \) (middle), and two-particle flow cumulants \( v_n\{2\} \) for \( n = 2, 3, \) and \( 4 \) (right) using several different specific shear viscosity parametrizations (lines) compared to experimental data from ALICE (symbols) [62, 65, 108]. The model provides an excellent description of these observables, and also explains several other observables not pictured including the experimentally measured anisotropic flow probability distributions [117] and event-plane correlations [118]. See reference [114] for a comprehensive overview.
To extract the QGP specific shear viscosity, the authors of reference [114] ran EKRT initialized hydrodynamic simulations with several different piecewise-linear \((\eta/s)(T)\) parametrizations (assuming zero bulk viscosity) and calculated numerous RHIC and LHC flow observables. Of the parametrizations that they tested, the two that provided the best overall description of the data were a constant (flat) parametrization \(\eta/s = 0.2\), and a sloped parametrization with a small hadronic viscosity and a minimum specific shear viscosity \((\eta/s)_{\text{min}} = 0.12\) located at \(T = 150\) MeV.

Presumably, these preferred specific shear viscosities would also change in the presence of non-zero bulk viscosity, which has been shown to affect extracted shear viscosity estimates [119, 120]. Therefore, it is difficult at the present time to directly compare the viscosities extracted by IP-Glasma and EKRT initial conditions. The estimates are obviously different, but it is not yet clear how much should be attributed to the initial conditions versus other components of the hydrodynamic simulation framework.

### 2.4.2 Case for a new approach

The IP-Glasma and EKRT models have significantly improved our current theoretical understanding of the initial stages of the collision. However, neither model describes the experimental data perfectly within the realm of its applicability, so it stands to reason that neither model is complete. This residual modeling error is a form of systematic uncertainty which biases current estimates of the QGP transport coefficients.

In the next chapter, I motivate and develop a complementary top-down approach for studying the QGP initial conditions using the constraints provided by the experimental data. This will allow me to investigate the correlated effect of initial condition uncertainties on QGP parameter estimates, and it will allow me to independently validate the effective scaling predicted by the IP-Glasma and EKRT initial condition frameworks. I start by deconstructing the initial condition problem into its simplest form.
Initial conditions of bulk matter

Every simulation needs a starting point. For hydrodynamic simulations of relativistic nuclear collisions, the starting point is the energy density $\epsilon$, fluid velocity $u^\mu$, and initial values of the bulk correction $\Pi$ and shear correction $\pi^{\mu\nu}$ at the hydrodynamic starting time. Generally speaking, models of the QGP initial conditions strive to be parameter free, predictive, and established on a firm theoretical footing. The holy grail would be an initial condition model that is elegantly derived from first principles, void of free parameters, and in perfect agreement with experimental measurements, barring the existence of confounding model errors. This idealized description would effectively eliminate the uncertainty in the QGP initial conditions and enable simulation-based extractions of fundamental QGP properties with unprecedented precision.

Over the past decade, theoretical progress has brought the field closer to this ultimate goal. In section 2.3, I discussed two of the more successful \textit{ab initio} theoretical calculations, the so-called EKRT [113] and IP-Glasma [85] initial condition models which are based on general concepts of gluon saturation physics. There are of course many other theoretical models which have been proposed in the literature, but these two models in particular have arguably reproduced the largest swath of experimental data using a rather small (albeit non-zero) number of free parameters.

These models, of course, do not provide all of the answers. It remains unclear, for exam-
ple, to what extent the IP-Glasma and EKRT frameworks are mutually compatible. While both theoretical models are based on similar ideas, their theoretical and computational implementations diverge in subtle ways which are difficult to quantify. Moreover, the experimental data can only validate the result of each model calculation. It thus becomes difficult to assess the veracity of competing initialization frameworks when the candidate models provide comparable descriptions of global experimental measurements. Additionally, it is not fully understood why these models reproduce certain experimental measurements which other models fail to describe. In order to address this question, it is important to identify the essential and non-essential features of each initial condition model which are needed to describe the data. Hydrodynamic simulations, however, often blur cause-and-effect relationships which makes it difficult to enumerate evidence for (or against) individual theoretical assumptions.

While the IP-Glasma and EKRT models provide global descriptions of soft-sector observables in relativistic nuclear collisions which are—all things considered—quite good, their descriptions of the data are of course imperfect. Often imperfections reflect missing features, e.g. nuclear structure modifications, which are easily added to the models without modifying their essential substance. It is of course also likely that at least some of the observed tension is attributable to errors in the adopted frameworks themselves. This is only natural; theoretical models are rarely perfect, and modeling errors are unavoidable.

Parameters of the EKRT and IP-Glasma models are generally fixed by their respective theoretical frameworks. In this sense, they are rigid models. When such models fail to describe the experimental data, there is little one can do to resolve the observed tension short of reworking each calculation. Initial condition errors are often reabsorbed by hydrodynamic model parameters when calibrating simulations to describe experimental data. For instance, if an initial condition model generates too little radial flow, the simulation may prefer a smaller QGP bulk viscosity than it should to compensate as I mentioned before. In this manner, initial condition errors propagate through the entire simulation framework. Hydrodynamic parameter estimates are thus often (and rightly) criticized for being highly
dependent on the choice of initial conditions.

In this chapter, I propose an alternative approach to *ab initio* theory calculations, which seeks to reverse engineer the properties of the QGP initial conditions using systematic model-to-data comparison. I develop for this purpose a new parametric model of the QGP initial conditions which is designed to be flexible. This flexibility allows the model to mimic specific theory calculations as well as interpolate between them. It describes, in this sense, a sort of meta-model which spans a semi-exhaustive space of reasonable theoretical descriptions. I then constrain free parameters of the model using top-down data-driven methods that rigorously account for different sources of uncertainty in the hydrodynamic framework. The method hence claims to know very little about the QGP initial conditions *a priori* in order to see what can be learned from the data and hydrodynamic framework alone. Such conclusions are thus less model dependent, and more robust to theoretical uncertainties.

This chapter is intended for the pragmatist. My goal is to explain the QGP initial conditions simply, using notation that is readily expressed as computer code. I also make a concerted effort to describe all relevant components of the initial conditions, including those components which are often neglected in the literature because they are deemed theoretically uninteresting, or because they are relatively generic.

### 3.1 Approximations in the high-energy limit

Throughout this dissertation, I apply approximations which are only valid in the so-called ultrarelativistic limit, i.e. collisions where the nuclei are Lorentz contracted by $\gamma \gtrsim 100$ along their direction of motion in the lab frame. This definition is somewhat arbitrary, but I will explain why it is necessary in a moment, and it will become clear why this choice is a reasonable cutoff.

Consider, for example, two identical spherical nuclei, each with radius $R$, that move
Figure 3.1  Nuclear overlap times $\tau_{\text{overlap}}$ measured in the lab frame for Pb-Pb collisions at several RHIC and LHC beam energies. Based on figure from [121].

with velocities $\pm \beta_z$ along the $\hat{z}$ direction. Each nucleus is Lorentz contracted by a factor

$$\gamma = \frac{1}{\sqrt{1 - \beta_z^2}}$$  \hspace{1cm} (3.1)

along its direction of motion and thus has a diameter $D = 2R/\gamma$ along the $\hat{z}$ direction when viewed from the lab frame. Assuming the two nuclei collide head on, they will pass through each other after an overlap time

$$\tau_{\text{overlap}} = \frac{2R}{\gamma \beta_z} = \frac{2R}{\sinh(y_{\text{beam}})},$$  \hspace{1cm} (3.2)

in the lab frame, where $y_{\text{beam}} = \text{acosh}(\sqrt{s_{\text{NN}}}/(2m_p))$ is the rapidity of the beam [121]. Here $\sqrt{s_{\text{NN}}}$ is the center of mass energy per nucleon pair of the accelerated ions, and $m_p = 0.938$ GeV is the proton mass. Several example overlap times are shown in figure 3.1 for Pb-Pb collisions at $\sqrt{s_{\text{NN}}} = 200, 2760,$ and $5020$ GeV, beam energies which are used at RHIC and the LHC. For all three of these beam energies, $\gamma \geq 100$ and $\tau_{\text{overlap}} \lesssim 0.1 \text{ fm/c}$.

I now argue that these overlap times are sufficiently short to neglect transverse dynamics which occur while the nuclei pass through each other. Let’s consider a single vertex for an interaction between two partons, each located on the leading edge of the colliding nuclei. When these primary partons scatter, they produce secondary partons which emerge from
their interaction vertex with some velocity \( \beta \leq 1 \). If \( \tau_{\text{overlap}} \lesssim 0.1 \text{ fm/c} \), all partons involved in the interaction—secondary or otherwise—may propagate for an equivalent amount of time as the nuclei continue to interpenetrate.

The absolute farthest each parton can move from its original interaction vertex in this time is \( \Delta x_{\text{max}} = 0.1 \text{ fm} \), and hence the same is true for its displacement in the transverse plane. Given that hydrodynamics is an effective theory which (in its kinetic formulation) averages thermal quantities over length scales of the interparticle mean free path, density fluctuations over distances of \( \ell \lesssim 0.1 \text{ fm} \) should not significantly affect the bulk dynamics of the system at much larger scales [122]. I therefore assume that the interacting matter moves along straight-line trajectories parallel to the beam axis as it pierces each nucleus. This is a central assumption of my work, and it underlies nearly all of the approximations that follow.

Consider now, a single straight-line trajectory defined by \( (x, y) = (x', y') \), which is parallel to the beam axis and pierces the interaction region of the collision, as depicted by the dashed line in figure 3.2. Moreover, let \( e_0 \) define the three-dimensional energy density deposited by the collision immediately after the nuclei pass through each other. The energy density \( e_0 \) deposited along this line at very early times is causally disconnected from all positions \( (x, y, z) \) in each colliding nucleus where

\[
(x - x')^2 + (y - y')^2 > \tau_{\text{overlap}}^2.
\]  

(3.3)
In the ultrarelativistic limit, causality reduces the collision dynamics to a one-dimensional problem at each transverse grid location. The resulting picture is analogous to that of two colliding trains, where each train is a stack of Lorentz contracted nucleons. The secondary matter shown on the right is produced locally and is some function of the projectile and target densities $\rho_A$ and $\rho_B$ in each nucleus.

In other words, the collision dynamics can only traverse distances less than or equal to the collision’s age. For ultrarelativistic collisions with $\gamma \geq 100$ and $\tau_{\text{overlap}} \lesssim 0.1 \text{ fm/c}$, the energy density $e_0(x, y, z)$ is essentially insensitive to all transverse coordinates $(x', y') \neq (x, y)$.

The aforementioned ultrarelativistic limit consequently factorizes the dynamics which occur at different points in the transverse plane immediately after first impact. Each point in the transverse plane thus describes an independent one-dimensional initialization problem. The situation is conceptually analogous at each transverse grid location to a head-on collision between two oncoming trains of variable length, where each train represents a stack of Lorentz contracted nucleons.

This analogue is depicted in figure 3.3 which shows a stack of three nucleons barreling down the beam axis to collide head-on with a stack of four nucleons moving in the opposite direction. Let’s assume that these stacks fully interpenetrate after an overlap time $\tau_{\text{overlap}} < 0.1 \text{ fm/c}$ as is generally the case for nuclear collisions at top RHIC and LHC energies. Our goal is to determine the energy (or entropy) which is deposited by the collision shortly after this overlap time into a small volume element centered on the collision epicenter.

This highly simplified one-dimensional picture of a relativistic nuclear collision may be further subdivided into three distinct modular components which each describe different aspects of the collision problem:

i) Prior to the collision, the state of the system is described by the density of nuclear matter $\rho_A$ and $\rho_B$ in each nucleus which passes through the transverse coordinate of
interest. Following the freight train analogy, these densities represent the number of boxcars in each train and hence the total energy and momentum carried toward the collision epicenter. The first component of the problem therefore describes nuclear structure.

ii) Nucleons are quantum objects which interact probabilistically. Even for the seemingly head-on collision in figure 3.3, there is a small chance the nucleons interpenetrate without interacting. The existence of an inelastic nucleus-nucleus collision hence presupposes the existence of one or more inelastic nucleon-nucleon collisions. The second component of the problem thus describes the collision’s inelastic nuclear cross sections.

iii) Given the density of participant matter in each nucleus which passes through a certain transverse grid location, one must ultimately determine the energy or entropy deposited by the collision into a small volume element centered at that point. The last component of the problem hence describes local energy and entropy deposition.

I now proceed to describe each of these modular components in detail. Collectively, they form a framework for modeling the energy and entropy deposited by the collision immediately after the two nuclei interpenetrate.

### 3.2 Nuclear structure

The starting point of every initial condition model is the three-dimensional density $\rho(x)$ of nuclear matter in the rest frame of each colliding nucleus. Strictly speaking, this density is governed by quantum mechanics, and thus it is characterized by a multi-body nuclear wave function $\Psi_N$ subject to the normalization condition

$$\int d^3x_1 \ldots d^3x_N |\Psi_N(x_1, \ldots, x_N)|^2 = N, \quad (3.4)$$

where $N$ is the total number of protons and neutrons inside the nucleus. The act of the collision collapses the wave function and samples inside each nucleus a set of discrete nucleon
where each nucleon is itself an extended object described by its own multi-body wave function consisting of smaller quark and gluon degrees of freedom. The density of nuclear matter probed by the collision is thus given by

$$\rho(x) = \sum_{i=1}^{N} \rho_n(x - x_i),$$

(3.6)

where $x_i$ is the position of each nucleon, $\rho_n$ is its sampled nucleon density, and $N$ is the number of nucleons in the nucleus. For the moment, I choose to ignore nucleon density fluctuations and blithely model each nucleon density $\rho_n$ as a generic blob described by a three-dimensional Gaussian distribution

$$\rho_n(x) = \frac{1}{(2\pi w^2)^{3/2}} \exp \left( -\frac{|x|^2}{2w^2} \right),$$

(3.7)

with a free parameter $w$ which varies the nucleon’s effective size. Here I’ve chosen a Gaussian for convenience; it is a crude simplifying assumption commonly used in the literature [114, 123], but it is by no means realistic. I will revisit this approximation later in subsection 5.3.1.

### 3.2.1 Sampling heavy-ions

The distribution used to sample the nucleon positions is typically inferred from electron scattering experiments which measure the average charge-density of the nucleus [124]. This charge-density is carried by the distribution of protons inside the nucleus, which is generally somewhat different than the distribution of neutrons [125, 126]. However, for reasons which I will explain shortly, the difference between the two distributions is rather small, and thus the measured charge-density distribution is commonly used to sample the positions of all nucleons inside the nucleus.

Electron scattering experiments use the Born approximation to relate the charge-density
distribution of the nucleus \( \rho(r) \) to the differential cross section \( d\sigma/d\Omega \) of electrons scattered off it [127]. The radial charge-density of a spherical heavy nucleus is commonly modeled by a so-called two-parameter Fermi (2PF) distribution

\[
\rho(r) = \frac{\rho_0}{1 + \exp \left( \frac{r-R}{a} \right)},
\]

(3.8)

where \( \rho_0 \) is the local charge-density inside the nucleus, \( R \) is the half-height radius of the nucleus, and \( a \) is a diffuseness parameter which gives the nucleus a soft edge.

Deviations from spherical symmetry are then added using spherical harmonics \( Y_{\ell}^{m}(\theta, \phi) \) to deform the half-height radius of the distribution such that \( R \rightarrow R'(\theta) \). For example, \( Y_{2}^{0} \) and \( Y_{4}^{0} \) harmonics are commonly used to deform \( R \) along the polar angle \( \theta \) according to

\[
R'(\theta) = R \left[ 1 + \beta_2 Y_{2}^{0}(\theta) + \beta_4 Y_{4}^{0}(\theta) \right],
\]

(3.9)

where \( \beta_2 \) and \( \beta_4 \) are dimensionless coefficients that control the degree of deformation. For instance, \( {^{238}}U \) is roughly shaped like a rugby ball and is described by deformation parameters \( \beta_2 = 0.28 \) and \( \beta_4 = 0.093 \) [128, 129]. I list the 2PF and deformation parameters for several common heavy nuclei at RHIC and the LHC in table 3.1.

Each heavy-ion is then modeled—to first approximation—by sampling independent nucleon positions from the 2PF distribution equation (3.8), using the experimentally measured
Table 3.1  Estimates of the parameters $R$ and $a$ for the two-parameter Fermi distribution (3.8) used to model the charge-density of heavy nuclei. Values are shown for several common isotopes used at RHIC and the LHC along with their corresponding deformation parameters $\beta_2$ and $\beta_4$. The $^{238}$U parameters are somewhat contentious so several sets are listed.

<table>
<thead>
<tr>
<th>Name</th>
<th>$A$</th>
<th>$R$ [fm]</th>
<th>$a$ [fm]</th>
<th>$\beta_2$</th>
<th>$\beta_4$</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copper</td>
<td>63</td>
<td>4.2</td>
<td>0.596</td>
<td>0.162</td>
<td>-0.006</td>
<td>[124, 130]</td>
</tr>
<tr>
<td>Xenon</td>
<td>129</td>
<td>5.36</td>
<td>0.59</td>
<td>0.162</td>
<td>-0.003</td>
<td>[131–133]</td>
</tr>
<tr>
<td>Gold</td>
<td>197</td>
<td>6.38</td>
<td>0.535</td>
<td>-0.131</td>
<td>-0.031</td>
<td>[124, 130]</td>
</tr>
<tr>
<td>Lead</td>
<td>208</td>
<td>6.62</td>
<td>0.546</td>
<td>—</td>
<td>—</td>
<td>[124, 131]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.81</td>
<td>0.550</td>
<td>0.280</td>
<td>0.093</td>
<td>[128, 129]</td>
</tr>
<tr>
<td>Uranium</td>
<td>238</td>
<td>6.86</td>
<td>0.420</td>
<td>0.265</td>
<td>0.000</td>
<td>[128, 134]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.67</td>
<td>0.440</td>
<td>0.280</td>
<td>0.093</td>
<td>[131, 135]</td>
</tr>
</tbody>
</table>

half-height radius $R$, diffuseness parameter $a$, and (when necessary) deformation parameters $\beta_2$ and $\beta_4$ using supplementary equation (3.9). Figure 3.4 shows, for illustration purposes, nucleon configurations for lead, xenon, and copper isotopes generated using this method.

There are several subtleties which are worth mentioning that pertain to sampling the nucleon positions. First, as I mentioned before, the charge-density distribution is not the same as the nucleon density. Studies have shown that in spherical, neutron rich nuclei, the radial distribution of neutrons is generally somewhat larger than the radial distribution of protons [126, 136]. This difference between the neutron and proton distributions is often expressed as a neutron skin thickness $\Delta r_{np}$, defined as the difference between the root-mean-square radii of the neutron and proton distributions. Recent measurements of $^{208}$Pb nuclei found $\Delta r_{np} = 0.15 \pm 0.03\,(\text{stat.})^{+0.01}_{-0.03}\,(\text{sys.})$ fm [126], while previous estimates report an even larger effect $\Delta r_{np} = 0.33^{+0.16}_{-0.18}$ fm [136].

Second, due to the finite size of nucleons inside the nucleus, the measured radial density is somewhat larger than the radial density of nucleon centers [137]. As a crude analogy, imagine a pepperoni pizza. If you distribute the pepperoni centers all the way out to the edge of the pizza, the pepperoni will extend past the edge of the crust. These nuclear densities are related by the convolution equation

$$
\rho(x) = \int d^3x_0 \hat{\rho}(x_0) \rho_n(x - x_0), \tag{3.10}
$$
where \( \rho(x) \) is the observed radial density, \( \hat{\rho}(x_0) \) is the density of nucleon positions, and \( \rho_n(x) \) is the assumed density profile of the nucleon. Consequently, if one naively samples the positions of nucleons according to a given target density distribution, the resulting ensemble-averaged nucleon density will be larger than desired. It turns out, however, that correcting this artificial swelling tends to negate the effect of accounting for the neutron skin. I therefore choose to ignore both effects when modeling various nuclei.

3.2.2 Nucleon correlations in the nucleus

When nucleon positions are sampled independently, it’s possible that two or more nucleons land on top of one another. This is of course unrealistic. Repulsive forces between the nucleons introduce short-range correlations \([138]\) which discourage mutual overlap. Nucleon correlations are commonly added to equation (3.8) by imposing a minimum distance criteria \([61, 131, 139]\)

\[ |x_i - x_j| > d_{\text{min}}, \tag{3.11} \]

between all pairs of nucleons \( i, j \) in the sampled nucleus.

This minimum distance constraint is regularly implemented using basic rejection sampling; nucleon positions are sampled one-by-one, and candidate positions are rejected if they place a nucleon too close to any of its previously sampled neighbors. Rejecting samples in this manner, however, modifies the target radial distribution and leads to an artificial swelling of the nucleus, similar to the effect caused by equation (3.10).

This swelling is commonly remedied by readjusting the parameters of the 2PF distribution for every value of \( d_{\text{min}} \) \([139]\). Unfortunately, such adjustments are cumbersome, and they often fail to recover the target radial distribution when \( d_{\text{min}} \) is large. Here, I describe a simple algorithm, developed by fellow graduate student Jonah Bernhard \([140]\), which implements the minimum distance criteria in equation (3.11) without modifying the target radial distribution.

First, consider a spherically symmetric heavy-ion with \( A \) nucleons described by the radial density \( \rho(r) \). The algorithm starts by sampling the radii \((r_1, r_2, \ldots, r_A)\) for all \( A \) nucleon
positions. These radii are then sorted in ascending order, and a pair of spherical angles \((\theta_i, \phi_i)\) is sampled from a distribution of uniform solid angle \(d\Omega = \sin \theta \, d\theta \, d\phi\) one-by-one for each nucleon position radius \(r_i\). If a sampled pair of spherical angles places a nucleon too close to any of its previously placed neighbors, \((\theta_i, \phi_i)\) is resampled (but not \(r_i\)) until the minimum distance criteria (3.11) is satisfied. The algorithm will attempt to relocate each nucleon 1000 times by resampling \((\theta_i, \phi_i)\) until it gives up and leaves it in its last sampled position. Note, the radii are sampled once and are never resampled; only the angles of each nucleon are resampled. The target radial distribution is thus perfectly preserved.

The algorithm also works for deformed nuclei such as \(^{238}\text{U}\), although, for deformed nuclei, only the azimuthal angle \(\phi_i\) may be resampled since \(\rho\) is no longer constant as a function of polar angle \(\theta\). This restricts the available phase space to readjust each nucleon position, and the algorithm breaks down for smaller values of \(d_{\text{min}}\). Despite this limitation, the algorithm works well (encounters limited failures) up to \(d_{\text{min}} = 1.7\) fm for spherically symmetric nuclei and \(d_{\text{min}} = 1.5\) fm for deformed nuclei.

### 3.2.3 Sampling light-ions

Experiments also commonly study collisions of light-ions including protons, deuterons, and helium-3 nuclei. These ions are too small to be modeled by a two-parameter Fermi distribution, so their nucleon position distributions are modeled on a case-by-case basis.

The deuteron is a loosely bound system consisting of one proton and one neutron. Its structure is commonly modeled using the Hulthén wave function

\[
\phi_d(r_{pn}) = \left(\frac{\alpha \beta (\alpha + \beta)}{2 \pi (\alpha - \beta)^2}\right)^{1/2} \left(\frac{e^{-\alpha r_{pn}} - e^{-\beta r_{pn}}}{r_{pn}}\right),
\]

where \(r_{pn}\) is the distance between the proton and neutron, \(\alpha = 0.228\) fm\(^{-1}\), and \(\beta = 1.18\) fm\(^{-1}\) [141, 142]. The corresponding density \(\rho(r_{pn})\) is simply proportional to the square of this wave function, and thus

\[
\rho(r_{pn}) \propto \left(\frac{e^{-\alpha r_{pn}} - e^{-\beta r_{pn}}}{r_{pn}}\right)^2.
\]
In order to sample each nucleon position, we make the substitution \( r_{pn} \rightarrow 2r \), where \( r \) is half the distance between the proton and neutron. Then we sample the position of the first nucleon in the center of mass frame according to the radial distribution

\[
\rho(r) \propto \left( \frac{e^{-2\alpha r} - e^{-2\beta r}}{r} \right)^2.
\]  

Once the position of the first nucleon is known, the second nucleon is placed directly across from it at a distance \( r_{pn} = 2r \) [131].

Unfortunately, the three-body system of the helium-3 nucleus is far too complicated to model using a simple analytic form. When simulating collisions of \(^3\)He nuclei, I sample and randomly rotate pre-tabulated nucleon positions determined from computer simulations based on Green’s-function Monte Carlo [143]. I do not show any results for collisions of deuteron or \(^3\)He nuclei in this dissertation, but the nuclei are nevertheless implemented in the computer code developed later in this chapter to facilitate future comparisons to RHIC data.

### 3.3 Inelastic nuclear cross sections

Consider now two particles which are simultaneously shot into opposite ends of a cylindrical pipe. Assume that each particle moves through the pipe along a straight-line trajectory, parallel to its axis of symmetry. Moreover, assume that each particle enters the pipe at a random entry point. At some moment, the particles either collide inside the pipe or pass by each other unscathed.

Naturally, this collision probability depends on the properties of the particles and the properties of the pipe. It’s easy to reason the dependence on the pipe geometry, namely \( P_{\text{coll}} \propto \mathcal{A}^{-1} \), where \( \mathcal{A} \) is the cross sectional area of the cylinder. The proportionality constant \( \sigma = P_{\text{coll}} \mathcal{A} \) thus depends strictly on the properties of each particle. This proportionality constant has units of area, is called a cross section, and is often used as a proxy for particle collision probabilities.
Figure 3.5  Geometric interpretation of the cross section for a collision of two billiard balls of radius $R$. One billiard ball moves into the page, and the other moves out of the page (solid circles). The line connecting their centers is the impact parameter $b_{\text{max}} = 2R$ corresponding to their maximal point of contact. Their geometric cross section is the area enclosed by the dashed circle $\sigma = \pi (2R)^2$.

Experimentally, the cross section $\sigma$ is determined by measuring the average collision rate between beams of particles. It is defined by the relation

$$\frac{dN}{dt} = \sigma \times \mathcal{L},$$

where $dN/dt$ is the number of collisions per unit time, and $\mathcal{L}$ is the beam luminosity, defined as the number of particles passing through the beam’s cross sectional area $A$ per unit time [144].

Theoretically, the cross section is defined a number of different ways. In this dissertation, I focus exclusively on the geometric cross section, defined as the area

$$\sigma = \int d^2 b \, P_{\text{coll}}(b).$$

Here $b$ is the impact parameter of the two-body system, defined as the vector between each particle’s center of mass at the moment of closest approach, and $P_{\text{coll}}(b)$ is the probability of a collision at a given impact parameter. One can easily verify that this geometric definition of the cross section also agrees with the experimental definition using Monte Carlo methods to sample random collisions between the particles. Equation (3.16) thus connects the experimentally measured cross section to the impact parameter dependent collision probability.

This geometric definition of the cross section is best explained by a simple example.
Consider for this purpose a classical collision between two billiard balls, each of radius $R$. Their hard-sphere collision probability is given by

$$P_{\text{coll}}(b) = \begin{cases} 1 & \text{if } b < 2R, \\ 0 & \text{otherwise.} \end{cases}$$  \hspace{1cm} (3.17)

The resulting geometric cross section is thus

$$\sigma = \int_0^\infty 2\pi b \, db \, P_{\text{coll}}(b) = \int_0^{2R} 2\pi b \, db = 4\pi R^2. \hspace{1cm} (3.18)$$

Equivalently, this area may be written in terms of the maximum impact parameter $b_{\text{max}} = 2R$ between the two spheres which produces a collision,

$$\sigma = \pi b_{\text{max}}^2. \hspace{1cm} (3.19)$$

One sees that the classical billiard ball cross section is simply the area defined by the maximum point of contact between the two spheres as depicted in figure 3.5.

Conversely, given the experimentally measured value for the cross section $\sigma$, one can easily calculate the radius of the hard-sphere and its impact parameter dependent collision probability:

$$P_{\text{coll}}(b) = \begin{cases} 1 & \text{if } b < \sqrt{\sigma/\pi}, \\ 0 & \text{otherwise.} \end{cases}$$  \hspace{1cm} (3.20)

This hard-sphere model of the cross section, while admittedly crude, is thus commonly used in computer models of high-energy particle collisions. It is also referred to as the black-disk approximation, since it is functionally equivalent to a classical collision of two solid disks.

Up to this point, I’ve been rather vague about the definition of a collision event and similarly nondescript about the types of particles involved. I now direct my attention collisions of ultrarelativistic nucleons, specifically inelastic nucleon-nucleon collisions, i.e. collisions where at least one of the nucleons becomes excited or breaks up. The aforementioned black-disk model is commonly used to sample inelastic nucleon-nucleon interactions, but
what about the inelastic multi-body interaction of two heavy-nuclei?

### 3.3.1 Glauber model of nuclear cross sections

Inelastic nuclear cross sections are commonly described using a model of nucleus-nucleus collisions developed by Glauber [145–147]. His model provides a theoretical foundation which relates pairwise nucleon-nucleon cross sections to the overall cross section of a larger nucleus. I cover in this section practical aspects of the Glauber model which are relevant to nuclear collision simulations. For a more detailed overview of the Glauber model including its historical origins, see reference [103].

Consider a collision of two heavy-ions, one with \( A \) nucleons and the other with \( B \) nucleons (labeled \( A \) and \( B \) respectively), which collide with impact parameter \( b \) in the transverse plane. Moreover, assume that each nucleus contains an average density of nucleons \( \rho_A \) and \( \rho_B \) which are normalized so that \( \int d^3x \rho_N(x) = N \), where \( N \) is the number of nucleons in that nucleus. The nuclear thickness function

\[
T_A(x_\perp) = \int dz \rho_A(x_\perp, z)
\]

(3.21)

describes the nucleon density in nucleus \( A \) which penetrates each transverse coordinate \( x_\perp \) during the collision. Thickness function \( T_B(x_\perp) \) follows in a similar fashion.

Now, let \( i \) index a single nucleon in nucleus \( A \), and let \( j \) index a single nucleon in nucleus \( B \). I assert that each pair of nucleons \( i,j \) collide inelastically with probability

\[
P_{i,j}^{\text{coll}}(b) = \frac{\sigma_{\text{inel}}^{NN}}{AB} \int d^2x T_A(x_\perp) T_B(x_\perp - b).
\]

(3.22)

To see why this is the case, let’s revisit the simple example discussed at the beginning of the section where each density \( \rho_{A,B} \) describes a single proton distributed uniformly inside a beam pipe of cross sectional area \( A \). The corresponding nuclear thickness functions are
determined by equation (3.21):

\[
T_{A,B} = \begin{cases} 
1/A & \text{if } r < \sqrt{A/\pi}, \\
0 & \text{otherwise,}
\end{cases}
\]  

(3.23)

where \( r \) is the transverse distance from the symmetry axis of the pipe. Assume that both thickness functions are centered in the middle of the beam pipe, i.e. \( b = 0 \) fm, so their impact parameter offset may be ignored. The collision probability of the two-nucleon system, defined by equation (3.22), equals

\[
P_{\text{coll}}^{i,j} = \frac{\sigma_{\text{NN}}^{\text{inel}}}{A} \int_0^{\sqrt{A/\pi}} 2\pi r dr = \frac{\sigma_{\text{NN}}^{\text{inel}}}{A}.
\]  

(3.24)

Equation (3.22) thus recovers the original definition of the cross section \( \sigma = A P_{\text{coll}} \) as desired. This nucleon-nucleon collision probability may be further simplified by defining a new quantity

\[
T_{AB}(b) \equiv \int d^2 x T_A(x_\perp) T_B(x_\perp - b),
\]  

(3.25)

known as the nuclear overlap function. The inelastic nucleon-nucleon collision probability may then be expressed as

\[
P_{\text{coll}}^{i,j}(b) = \frac{\sigma_{\text{NN}}^{\text{inel}} T_{AB}(b)}{AB}.
\]  

(3.26)

We can now proceed to calculate the total inelastic nucleus-nucleus cross section \( \sigma_{\text{AB}}^{\text{inel}} \). It is related to the probability of observing at least one pairwise nucleon-nucleon collision between nucleus \( A \) and \( B \), equal to

\[
P_{\text{coll}}^{\text{inel}}(b) = 1 - [1 - P_{i,j}^{\text{coll}}(b)]^{AB}
\]  

(3.27)

\[
= 1 - \left[ 1 - \frac{\sigma_{\text{NN}}^{\text{inel}} T_{AB}(b)}{AB} \right]^{AB}.
\]  

(3.28)

The inelastic nucleus-nucleus cross section subsequently follows from the definition of the
geometric cross section (3.16). This yields

\[
\sigma_{\text{inel}}^{AB}(b) = \int d^2b \left\{ 1 - \left[ 1 - \frac{\sigma_{\text{inel}}^{\text{NN}} T_{AB}(b)}{AB} \right]^{AB} \right\}.
\] (3.29)

Noting that \( \sigma_{\text{inel}}^{\text{NN}} T_{AB} \ll AB \) in a typical heavy-ion collision, we can use the fact that 

\[
\lim_{n \to \infty} (1 + x/n)^n = \exp(x)
\]

for small \( x \) to further simplify equation (3.29):

\[
\sigma_{\text{inel}}^{AB}(b) \approx \int d^2b \left\{ 1 - \exp\left[-\sigma_{\text{inel}}^{\text{NN}} T_{AB}(b)\right] \right\}.
\] (3.30)

Equations (3.29) and (3.30) thus reexpress, as desired, the inelastic nucleus-nucleus cross section in terms of the inelastic nucleon-nucleon cross section.

### 3.3.2 Counting participants and binary collisions

The Glauber model describes nucleus-nucleus collisions as a sum of all possible nucleon-nucleon collisions. Previously, it was shown that a single nucleon pair \( i, j \) collides with probability

\[
P_{i,j}^{\text{coll}}(b) = \frac{\sigma_{\text{inel}}^{\text{NN}} T_{AB}(b)}{AB}.
\] (3.31)

The probability of observing \( n \) such pairwise collisions is described by a binomial distribution

\[
P(n, b) = \binom{AB}{n} \left( \frac{\sigma_{\text{inel}}^{\text{NN}} T_{AB}(b)}{AB} \right)^n \left[ 1 - \frac{\sigma_{\text{inel}}^{\text{NN}} T_{AB}(b)}{AB} \right]^{AB-n},
\] (3.32)

where the prefactor is the number of combinations forming \( n \) pairwise collisions, the first term is the probability of observing \( n \) pairwise collisions, and the second term is the probability of observing \( AB - n \) pairwise misses.

The average number of collisions is then given by

\[
N_{\text{coll}}(b) = \sum_{n=1}^{AB} n P(n, b) = \sigma_{\text{inel}}^{\text{NN}} T_{AB}(b),
\] (3.33)

which follows from the mean of the binomial distribution. These pairwise inelastic nucleon
collisions are commonly called binary collisions.

In a similar fashion, one can estimate the average number of nucleons which participate in one of more pairwise collisions. The fraction of nuclear thickness $T_A(x_\perp)$ which pierces the target thickness $T_B(x_\perp - b)$ without interacting is given by

$$F^\text{miss}_A(x_\perp, b) = \left[ 1 - \frac{\sigma^{\text{inel}}_{NN} T_B(x_\perp - b)}{B} \right]^B.$$  \hfill (3.34)

Correspondingly, the complementary fraction that is struck is

$$F^\text{part}_A(x_\perp, b) = 1 - \left[ 1 - \frac{\sigma^{\text{inel}}_{NN} T_B(x_\perp - b)}{B} \right]^B.$$  \hfill (3.35)

The average number of participant nucleons in nucleus $A$ and $B$ is therefore

$$N^\text{part}_A(b) = \int d^2 x T_A(x_\perp) F^\text{part}_A(x_\perp, b) + \int d^2 x T_B(x_\perp - b) F^\text{part}_B(x_\perp, b),$$  \hfill (3.36)

which when expanded yields

$$N^\text{part}_A(b) = \int T_A(x_\perp) \left\{ 1 - \left[ 1 - \frac{\sigma^{\text{inel}}_{NN} T_B(x_\perp - b)}{B} \right]^B \right\} d^2 x +$$

$$\int T_B(x_\perp - b) \left\{ 1 - \left[ 1 - \frac{\sigma^{\text{inel}}_{NN} T_A(x_\perp)}{A} \right]^A \right\} d^2 x.$$  \hfill (3.37)

These participant nucleons are commonly called wounded nucleons. All remaining nucleons are called spectators since they rapidly leave the interaction region of the collision without meaningfully contributing to the dynamics of the produced fireball.

### 3.3.3 Monte Carlo Glauber model

The Glauber model just described employed several approximations to simplify the multi-body nature of the problem. The nuclei were modeled as smooth, cloud-like densities of nuclear matter $\rho_A$ and $\rho_B$. Nucleon positions were assumed to be uncorrelated inside the nucleus and unspecified. Each thickness function $T_A$ and $T_B$ was hence averaged over
all possible nucleon positions sampled by the collision. The implementation consequently neglected all effects which arise from event-by-event fluctuations in the positions of nucleons within each nucleus.

While it is difficult to account for these effects analytically, it is easy to incorporate them in numerical computer simulations. This section describes an extension of the original Glauber model, known as the Monte Carlo Glauber model or MC-Glauber model for short. It is more realistic than the analytic Glauber model, and hence it is commonly used when calculating quantities like $\sigma_{AB}^{inel}, N_{part}$, and $N_{coll}$. It is also easier to explain and more intuitive.

Consider as before a collision between two ions, labeled $A$ and $B$, which are shot down a beam pipe aligned with the $\hat{z}$ direction. Let $\{x_i^\perp\}$ denote the transverse positions of nucleons in nucleus $A$ and $\{x_j^\perp\}$ the transverse positions of nucleons in nucleus $B$ determined by the collapse of each wave function. Assume for the moment that each pair of nucleons $i, j$ collides inelastically according to the black-disk collision profile, equation (3.20), separated by impact parameter $b_{ij} = |x_i^\perp - x_j^\perp|$ in the transverse plane.

The Monte Carlo Glauber model loops over all pairs of nucleons and samples their inelastic collision probability, determined by the measured inelastic nucleon-nucleon cross section $\sigma_{NN}^{inel}$. This asserts that every collision is independent of the previous collisions. If a pair of nucleons collide, they are both labeled participants and the number of binary collisions is incremented. Pseudo-code for the algorithm is shown below.

```python
binary_collisions = 0
nucleon_participants = 0

for nucleon_i in nucleus_A:
    for nucleon_j in nucleus_B:
        xi = nucleon_i.position
        xj = nucleon_j.position
        bij = distance(xi, xj)
```

62
if bij < sqrt(sigma_nn / pi):
    nucleon_i.isParticipant = true
    nucleon_j.isParticipant = true

    binary_collisions += 1

for nucleon_i in nucleus_A:
    if nucleon_i.isParticipant:
        nucleon_participants += 1

for nucleon_j in nucleus_B:
    if nucleon_j.isParticipant:
        nucleon_participants += 1

Expressed as code, the Monte Carlo Glauber model is exceptionally simple. It merely defines an impact-parameter dependent collision probability for each nucleon pair and then applies it to all pairs of nucleons independently. Figure 3.6 shows the average density and positions of participant nucleons predicted by the analytic and Monte Carlo Glauber models respectively. Both panels show a single Pb-Pb collision with impact parameter $b = 7$ fm using an inelastic nucleon-nucleon cross section $\sigma_{NN}^{\text{inel}} = 6.4$ fm$^2$. The difference between the two models is obvious and striking. Evidently, fluctuations in the positions of the nucleons within each nucleus are large, and thus one expects the local density of participant matter to vary significantly from event to event.

### 3.3.4 Modeling the nucleon-nucleon interaction profile

Unlike the analytic Glauber model, the Monte Carlo Glauber model is sensitive to the impact-parameter dependence of the inelastic nucleon-nucleon collision probability. I previously assumed for this purpose a black-disk interaction profile (3.20) which is commonly used in the literature. This is, however, a crude assumption which only satisfies the most
basic requirement of the interaction, namely $\int d^2 b P_{\text{coll}}(b) = \sigma_{\text{NN}}^{\text{inel}}$. Nearly everything that is known about protons and neutrons indicates that they do not collide like hard spheres.

A far more realistic interaction profile may be obtained by modeling each nucleon as a composite object consisting of smaller constituents. The analytic Glauber model may then be used to calculate the multi-body collision probability as a function of the nucleon-nucleon impact parameter [148, 149]. Recall that for a heavy-ion collision, this probability (3.28) was

$$P_{AB}^{\text{coll}}(b) = 1 - \left[ 1 - \frac{\sigma_{\text{NN}}^{\text{inel}} T_{AB}(b)}{A B} \right]^{AB} ,$$

(3.38)

where $A$ and $B$ were the number of nucleons in each nucleus, $T_{AB}$ was the nuclear overlap function (3.25), and $\sigma_{\text{NN}}^{\text{inel}}$ was the inelastic nucleon-nucleon cross section. The analogous collision probability for two nucleons is thus

$$P_{nn}^{\text{coll}}(b) = 1 - \left[ 1 - \frac{\sigma_{\text{eff}}^{\text{inel}} T_{nn}(b)}{N^2} \right]^{N^2} ,$$

(3.39)

where $N$ is the average number of constituents inside each nucleon, $T_{nn}$ is the nucleon-nucleon overlap function, and $\sigma_{\text{eff}}^{\text{inel}}$ is the effective cross section between the constituents.
The content of each nucleon includes valence quarks, sea quarks and gluons. At high energy, the gluon density grows and the total number of visible constituents $N \gg 1$ is large. In this many-body limit, $\sigma_{\text{eff}}^{\text{inel}} T_{nn} \ll N^2$, and so

$$P_{nn}^{\text{coll}}(b) = 1 - \exp[-\sigma_{\text{eff}}^{\text{inel}} T_{nn}(b)]. \quad (3.40)$$

Each nucleon is, generally speaking, some blob of fluctuating constituent density. To first approximation, this density can be modeled as a three-dimensional Gaussian (3.7) of width $w$. The nucleon thickness is then

$$T_n(x_\perp) = \frac{1}{2\pi w^2} \exp \left( - \frac{|x_\perp|^2}{2w^2} \right), \quad (3.41)$$

and the nucleon-nucleon overlap function is correspondingly

$$T_{nn}(b) = \int d^2x T_n(x_\perp) T_n(x_\perp - b) \quad (3.42)$$

$$= \frac{1}{4\pi w^2} \exp \left( - \frac{b^2}{4w^2} \right). \quad (3.43)$$

Plugging this expression into equation (3.40) and using the definition of the geometric cross section $\int d^2b P_{\text{coll}}(b) = \sigma_{\text{NN}}^{\text{inel}}$, one finds that

$$\sigma_{\text{NN}}^{\text{inel}} = \int_0^{\infty} 2\pi b db \left\{ 1 - \exp \left[ - \frac{\sigma_{\text{eff}}^{\text{inel}}}{4\pi w^2} \exp \left( - \frac{b^2}{4w^2} \right) \right] \right\}. \quad (3.44)$$

Given a certain nucleon width $w$ and inelastic cross section $\sigma_{\text{NN}}^{\text{inel}}$, this equation may be solved numerically to calculate the effective cross section parameter $\sigma_{\text{eff}}^{\text{inel}}$ which, together with equation (3.40), specifies the inelastic nucleon-nucleon collision probability $P_{nn}^{\text{coll}}(b)$ at each impact parameter.

In practice, the relation may be solved by truncating the integral at a maximum impact parameter $b_{\text{max}} = Cw$, expressed as some number of nucleon widths $C$. After appropriate
change of variables, it may be written
\[
\frac{\sigma_{\text{NN}}^{\text{inel}}}{4\pi w^2} = \frac{C^2}{4} + \operatorname{Ei}\left(-e^{-A^2/4} \frac{\sigma_{\text{eff}}^{\text{inel}}}{4\pi w^2}\right) - \operatorname{Ei}\left(-\frac{\sigma_{\text{eff}}^{\text{inel}}}{4\pi w^2}\right),
\]
where \(\operatorname{Ei}\) is the exponential integral. This transcendental equation is then easily solved using a standard root finding algorithm. Throughout this work \(C = 6\) is used, i.e. the maximum nucleon-nucleon impact parameter is sampled out to six nucleon widths.

Figure 3.7 compares the black-disk nucleon interaction profile to the analytic Glauber nucleon profile for several values of the nucleon width \(w\). When the nucleon width \(w\) is small, its density is compact, and the effective cross section parameter \(\sigma_{\text{eff}}^{\text{inel}}\) determined by equation (3.45) is large. The nucleon thus becomes opaque and asymptotically approaches the black-disk limit. Equation (3.40) is thus a more general case of the black-disk limit that naturally accommodates nucleons of different widths.

Figure 3.8 shows the average number of nucleon participants and binary collisions predicted by the model as a function of collision centrality, calculated for different models of the nucleon-nucleon collision profile. The black-disk interaction profile (black dashed line) is compared to a Glauber interaction profile (colored lines) using Gaussian nucleons of various width \(w\). While the different collision profiles predict roughly the same number of participants, they predict dramatically different numbers of binary collisions as illustrated.
Figure 3.8  Monte Carlo Glauber model predictions for the number of nucleon participants (left) and binary collisions (right) as a function of collision centrality for Pb-Pb collisions assuming $\sigma_{NN}^{\text{inel}} = 6.4$ fm$^2$ and $d_{\text{min}} = 1$ fm. Each line shows the prediction of a different nucleon-nucleon collision profile; black dashed-lines are the black-disk nucleon collision profile (3.20), and colored lines are the Glauber-like collision profile (3.40) using Gaussian nucleons of variable width $w$. Sub-panels show the ratio of each calculation to the black-disk model.

by the ratio plot in the bottom-right panel. The difference is $\sim$10% for $w = 0.5$ fm, and it grows as large as $\sim$30% for $w = 0.9$ fm.

The nucleon-nucleon collision profile is thus an important source of uncertainty which should be accounted for when estimating the number of nucleon participants and binary collisions. Throughout this work, I use the analytic Glauber interaction profile (3.40) for sampling nucleon-nucleon collisions because it is unequivocally more realistic than the black-disk approximation, and because it contains the black-disk approximation as a specific limiting case. Moreover, it allows me to vary the nucleon width $w$ over a large range of values while maintaining a sensible nucleon collision profile.

### 3.3.5 Constructing the participant thickness functions

The Monte Carlo Glauber model is a useful tool for calculating the QGP initial conditions because it samples inelastic collisions between each pair of nucleons. Each simulated nuclear collision event can then be categorized as elastic or inelastic based on the existence (or non-existence) of at least one inelastic nucleon-nucleon collision, roughly mimicking the inelastic event selection used by experiment.
Even when the inelastic event trigger is satisfied, usually only a small fraction of nucleons in each nucleus collide inelastically. The remaining nucleons disassociate from their parent nuclei and are ejected at large rapidities. These spectators have negligible effect on the energy and entropy deposited at midrapidity so their contribution can be safely ignored. Therefore, it is only necessary to model the participant density of each nucleus.

This participant density is easily calculated in the local rest frame of nucleus $A, B$ by summing the density of each participant nucleon

$$\rho_{A,B}^{\text{part}}(x) = \sum_{i=1}^{N_{A,B}} \rho_n(x - x_i \pm b/2), \quad (3.46)$$

over $N_{A,B}$, the total number of participants in nucleus $A, B$ respectively. The function $\rho_n$ under the summation is the nucleon density specified by equation (3.7), $x_i$ is the location of each nucleon in the nuclear center of mass frame, and $b$ is the impact parameter vector separating each nuclear center of mass. Hence, for a simple collision involving two protons, each participant density consists of a single Gaussian shifted by the appropriate impact parameter offset. In a larger nucleus-nucleus collision, the number and distribution of participants fluctuates event-by-event, leading to fluctuations in the collision’s overall energy and entropy deposition.

This fluctuating nuclear density, however, does not account for all the observed variance in the resulting event activity. Proton-proton collisions, for example, evidence large multiplicity fluctuations [150–152] which cannot be explained by differences in the geometric overlap density alone. It is therefore common to multiply each nucleon density using a randomly sampled weight factor [153–155]. The resulting fluctuated participant density may then be written as

$$\tilde{\rho}_{A,B}^{\text{part}}(x) = \sum_{i=1}^{N_{A,B}} \gamma_i \rho_n(x - x_i \pm b/2), \quad (3.47)$$

where $\gamma_i$ is the random weight used to fluctuate each nucleon (note, I’ve added a tilde to $\tilde{\rho}_{A,B}^{\text{part}}$ to indicate the presence of the additional fluctuations). Throughout this work, I sample the weights $\gamma_i$ from a gamma distribution with unit mean and variance $1/k$. 

68
where \( k \) is the so-called gamma shape parameter. I've chosen this distribution because particlization fluctuations are Poissonian, and the folding of the gamma distribution with a Poisson distribution yields a negative binomial distribution which is commonly used to fit the multiplicity distribution of high-energy proton-proton and nucleus-nucleus collisions [150, 156].

Equation (3.47) describes the three-dimensional density of participant matter in the rest frame of each nucleus. Following the convention used in Glauber model, I project this density onto the transverse plane \( x_\perp \) orthogonal to the beam axis to calculate the density of participant matter visible to a probe piercing the transverse coordinate \( x_\perp \):

\[
\tilde{T}_{A,B}(x_\perp) = \int dz \tilde{\rho}_{A,B}(x_\perp, z).
\] (3.48)

I call this quantity the participant thickness function. It equals the canonical thickness function defined by equation (3.21) with a few modifications: it excludes spectator matter, it re-weights each nucleon density by a gamma random variable, and it has the impact parameter offset \( \pm b/2 \) already baked in. The participant thickness functions \( \tilde{T}_A, \tilde{T}_B \) provide almost all the information that is needed to model the initial conditions at midrapidity. All that remains is to specify a scalar mapping which translates this information to an energy or entropy density profile.

### 3.4 Energy and entropy deposition

Consider again the situation depicted in figure 3.9 which was discussed at the beginning of the chapter. Three nucleons barrel down a beam pipe to collide head-on with four nucleons moving in the opposite direction. Suppose these nucleons collide inelastically and convert some of their energy into secondary matter, e.g., liberated quarks and gluons, which subsequently interact and undergo hydrodynamic expansion.

In the previous section, I characterized the initial state of the collision just before impact by two fields \( \tilde{T}_A, \tilde{T}_B \), which describe the density of participant matter in each nucleus
projected onto the plane orthogonal to the beam axis. The collision depicted in figure 3.9 is similarly characterized by two participant densities $\tilde{T}_A, \tilde{T}_B$, each describing the transverse density of three and four stacked nucleons respectively. Since I am interested in modeling the collision hydrodynamically, suppose that these participant densities rapidly interact to produce a fluid which is in local thermal equilibrium shortly after the nucleons interpenetrate. Moreover, assume that this fluid is approximately boost-invariant near midrapidity as discussed in subsection 2.3.1.

Generally speaking, the three-dimensional fluid produced by the collision is not boost-invariant. The longitudinal density of produced matter 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 [157, 158]. Nevertheless, boost-invariance has been shown to be a good approximation when analyzing particles detected close to midrapidity [92]. I therefore assume approximate boost-invariance and work in Milne coordinates $(x_\perp, \eta_s, \tau)$, where $x_\perp$ specifies a Cartesian coordinate $(x, y)$ in the transverse plane, $\eta_s = \frac{1}{2} \ln[(t + z)/(t - z)]$ is the system’s space-time rapidity, and $\tau = \sqrt{t^2 - z^2}$ is its proper time.

Consider now the infinitesimal volume element $dV = d^2x_\perp \tau_0 d\eta_s$ centered on the space-time coordinate $(x_\perp, \eta_s = 0, \tau_0)$, where $\tau_0$ is some time shortly after the nucleons interpenetrate (right-side of figure 3.9). Additionally, let $e_0 = e(x_\perp, \eta_s = 0, \tau_0)$ and $s_0 = s(x_\perp, \eta_s = 0, \tau_0)$ denote the energy and entropy density inside the cell at this time, averaged over an ensemble of collisions with the same values of $\tilde{T}_A$ and $\tilde{T}_B$. The collision correspondingly maps the initial state of the system just before impact, characterized by $\tilde{T}_A$.
and $\tilde{T}_B$, to some new state immediately after impact, characterized by its energy density $e_0$ or entropy density $s_0$. These mappings describe scalar functions

$$f_e : \tilde{T}_A, \tilde{T}_B \mapsto e_0 \quad \text{and} \quad f_s : \tilde{T}_A, \tilde{T}_B \mapsto s_0,$$

which provide an effective description of early-time dynamics. The goal of this section is to parametrize the functions $f_e$ and $f_s$. Each parametrization need not arise from a first-principles calculation, but it must obey basic physical constraints. Specifically, it should be:

- Symmetric under interchange of $\tilde{T}_A$ and $\tilde{T}_B$. The average system at midrapidity is invariant to reflections through the transverse plane.

- Conserving of total energy. Let $E_0$ be the total energy deposited into the volume $dV = d^2x_\perp d\eta s$ at time $\tau_0$. This energy is bounded above by $E_{\text{tot}} = (\sqrt{s_{\text{NN}}}/2)(\tilde{T}_A + \tilde{T}_B) d^2x_\perp$, the total energy carried by all nucleons interacting with that volume element.

- Monotonically increasing as a function of $\tilde{T}_A$ and $\tilde{T}_B$. Denser, more energetic collisions should deposit more energy and entropy on average.

Subject to these stipulations, we can conjecture a reasonable ansatz for the initial energy density $e_0$ or entropy density $s_0$ as a function of the participant thickness functions $\tilde{T}_A$ and $\tilde{T}_B$. Perhaps the simplest such function is a sum:

$$\begin{cases} e_0 \\ s_0 \end{cases} = \text{Norm} \times (\tilde{T}_A + \tilde{T}_B),$$

where Norm is an energy-dependent normalization constant carrying the appropriate units. The term in parentheses is simply the local participant or “wounded” nucleon density

$$n_{\text{part}} = \tilde{T}_A + \tilde{T}_B,$$
Figure 3.10  Charged-particle density \( dN_{\text{ch}}/d\eta \) as a function of collision centrality for Pb-Pb collisions at \( \sqrt{s_{\text{NN}}} = 2.76 \text{ TeV} \). Colored lines are model predictions based on wounded nucleon scaling \( dN_{\text{ch}}/d\eta \propto N_{\text{part}} \), binary collision scaling \( dN_{\text{ch}}/d\eta \propto N_{\text{coll}} \), and the two-component ansatz (3.54) with \( \alpha = 0.08 \). Black symbols show experimental data from the ALICE experiment [62]. The bottom plot shows the ratio of the model over the data (gray band \( \pm 10\% \)), and circles below the x-axis show the approximate nuclear overlap at two centralities.

so this ansatz is commonly referred to as the wounded nucleon model. It effectively assigns each nucleon a finite amount of energy or entropy which is fully deposited by its first nucleon-nucleon collision.

Indeed, this simple ansatz was actually one of the first models used in the literature to describe the initial state of nucleus-nucleus collisions. The idea dates back to a seminal paper by Bialas et al. [159] which conjectured a linear relation between the number of produced particles and the number of inelastic nucleon participants. This conjecture was supported by measurements of Pb-Pb collisions at the Super Proton Synchrotron (SPS) which showed that the transverse energy density \( dE_T/d\eta \) and charged-particle multiplicity \( dN_{\text{ch}}/d\eta \) both scale linearly with the number of participant nucleons [160]. Ideal boost-invariant hydrodynamics conserves the system’s energy and entropy per unit rapidity, so naturally this seemed to suggest that \( e_0 \) and \( s_0 \) scaled with the wounded nucleon density as well.

It was later realized, however, that the wounded nucleon model fails to reproduce the
centrality dependence of particle production observed by experiment. For example, consider the so-called entropy variant of the wounded nucleon model, $s_0 \propto n_{\text{part}}$. In ideal boost-invariant hydrodynamics, the midrapidity charged-particle yield $dN_{\text{ch}}/d\eta$ scales linearly with the initial entropy density $dS/d\eta_s$ at fixed freeze-out temperature and chemical potential [161]:

$$dN_{\text{ch}}/d\eta \propto dS/d\eta_s. \quad (3.52)$$

Hence the conjectured scaling $s_0 \propto n_{\text{part}}$ is easily verified by comparing the predicted centrality dependence of $N_{\text{part}}$ against the measured centrality dependence of $dN_{\text{ch}}/d\eta$. Figure 3.10 shows this comparison for Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. The blue line is the aforementioned scaling $dN_{\text{ch}}/d\eta \propto N_{\text{part}}$ scaled up to fit the data, and the black symbols with errors are data from the ALICE collaboration [62]. The wounded nucleon model significantly underpredicts the steep rise in particle production observed at small centralities. A similar line of reasoning can also be used to invalidate the alternative wounded nucleon variant $e_0 \propto n_{\text{part}}$; see reference [161]. The wounded nucleon model is therefore excluded by the data.

### 3.4.1 Two-component ansatz

Motivated by this finding, it was suggested that particle production may receive an additional contribution from hard processes which scale like the number of binary collisions $N_{\text{coll}}$ [162, 163]. Unlike the participant number $N_{\text{part}}$ which grows approximately linearly with increasing nuclear thickness, the binary collision number grows quadratically,

$$N_{\text{coll}} \propto \int d^2x_\perp T_A(x_\perp)T_B(x_\perp). \quad (3.53)$$

The orange line in figure 3.10 shows the curve $dN_{\text{ch}}/d\eta \propto N_{\text{coll}}$ predicted by binary collision scaling. It suffers from the opposite problem of participant scaling. The number of binary collisions rises much faster than the density of produced particles. Such observations led to
the creation of the so-called mixed or two-component ansatz \[164, 165\]

\[
\begin{aligned}
\begin{cases}
\epsilon_0 \\
n_0
\end{cases} &= \text{Norm} \times \left( \frac{1 - \alpha}{2} n_{\text{part}} + \alpha n_{\text{coll}} \right),
\end{aligned}
\]

which linearly interpolates between the local wounded nucleon density \(n_{\text{part}}\) and the local binary collision density \(n_{\text{coll}}\) using a dimensionless parameter \(\alpha\). The parameter \(\alpha\) is tuned once at each collision energy by fitting the centrality dependence of the charged-particle yield \[166\] after which it remains fixed. The model was particularly appealing because it could exactly fit the centrality dependence of the charged-particle density \(dN_{\text{ch}}/d\eta\) at midrapidity as illustrated by the green line in figure 3.10.

I argue now that this fit is a red herring, and that the two-component ansatz is wrong. There are in fact many problems with the ansatz; I’ll explain the simplest one. Imagine a stack of \(N\) projectile nucleons colliding head-on with a stack of \(N\) target nucleons. Moreover, assume for simplicity’s sake that all of the nucleons collide with each other. Let \(S\) denote the total entropy deposited into the rapidity interval \(|\eta| < 1\) immediately after the collision.

According to the two-component ansatz (3.54)

\[
S/N \propto (1 - \alpha) + \alpha N.
\]

This implies that I can make the entropy deposited per nucleon \(S/N\) arbitrarily large by increasing \(N\), the number of columnated nucleons in each stack, at fixed beam energy \(\sqrt{s_{\text{NN}}}\). Similarly, since the energy density grows faster than the entropy density \(e \sim s^{4/3}\), I can make \(E/N\), the energy deposited per nucleon, arbitrarily large as well. This is clearly absurd as it violates energy conservation in the dense-dense limit at fixed beam energy. Note, this violation will occur for any value of \(\alpha\), so long as \(N\) is sufficiently large.

The two-component model fails to conserve energy because it treats collision \(N\) the same as collision \(N + 1\). Presumably, the incremental energy and entropy deposited by each collision is somewhat smaller each time on average \[167\]. The average energy (or
entropy) deposited by the collision should therefore be boundable by participant scaling. Functionally, this means there exists some constant $C$ such that

$$f(\tilde{T}_A, \tilde{T}_B) \leq C(\tilde{T}_A + \tilde{T}_B), \quad \forall \tilde{T}_A, \tilde{T}_B.$$  \hfill (3.56)

This stipulation, however, is somewhat puzzling. Figure 3.10 seems to require a term which rises faster than $N_{\text{part}}$ with increasing density. How can this also be true?

### 3.4.2 Generalized mean ansatz

There is, in fact, a superior ansatz for QGP energy and entropy deposition which alleviates this problem. Let’s back up for a moment and return to the wounded nucleon ansatz in equation (3.50). This function is equivalent to the arithmetic mean of participant nucleon density

$$\begin{align*}
\langle \epsilon_0 \rangle & = \text{Norm} \times \frac{\tilde{T}_A + \tilde{T}_B}{2}, \\
\langle s_0 \rangle & = \text{Norm} \times s_0
\end{align*}$$  \hfill (3.57)

up to a meaningless factor of two in the denominator, which may be extracted from the normalization coefficient.

Evidently, this scaling fails to describe the centrality dependence of charged-particle production, but what about other types of means? With this in mind, I replace the arithmetic mean of the wounded nucleon model with a more flexible parametrization [168]

$$\begin{align*}
\langle \epsilon_0 \rangle & = \text{Norm} \times M_p(\tilde{T}_A, \tilde{T}_B), \\
\langle s_0 \rangle & = \text{Norm} \times s_0
\end{align*}$$  \hfill (3.58)

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

$$M_p(x, y) = \left( \frac{x^p + y^p}{2} \right)^{1/p}.$$  \hfill (3.59)

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

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

(3.60)

although it is not limited to these values. The parameter $p$ is a continuous variable which
can take any value $-\infty < p < \infty$.

The generalized mean has a number of interesting properties which make it suitable for
the present context:

- It is symmetric: $M_p(x, y) = M_p(y, x)$. This is required; the average energy or entropy
  produced at midrapidity should naturally be invariant under reflections $\tilde{T}_A \leftrightarrow \tilde{T}_B$.

- It is bounded: $M_p(x, y) \leq x + y$ for all positive $x$ and $y$. Therefore I can always choose
  a normalization that satisfies equation (3.56), ensuring that the available energy is not
  violated in the dense-dense limit (the same is not true for the two-component ansatz).

- It is scale-invariant: $M_p(c\tilde{T}_A, c\tilde{T}_B) = cM_p(\tilde{T}_A, \tilde{T}_B)$. I’ll explain the significance of this
  property later in the chapter.

The transverse field $M_p(\tilde{T}_A, \tilde{T}_B)$ carries units of fm$^{-2}$, the same as $\tilde{T}_A, \tilde{T}_B$, so I refer to
this quantity as the collision’s reduced thickness function

$$T_R \equiv M_p(\tilde{T}_A, \tilde{T}_B),$$

(3.61)

so named because it takes two thickness functions $\tilde{T}_A$ and $\tilde{T}_B$ and reduces them to a third
thickness function, similar to a reduced mass. Indeed, for generalized mean parameter
$p = -1$, the reduced thickness function and the reduced mass are algebraically equivalent.
Let me return now to the puzzle which I posed at the end of the last section. Figure 3.10 appears to require a term which grows faster than \( N_{\text{part}} \) with increasing density. Is this really true? No, it’s not true. At least it’s not necessarily true. It turns out that that generalized mean ansatz (3.58) also describes the centrality dependence of charged-particle production, and it does so without invoking the problematic binary collision term. Consider for instance, the entropy variant of the generalized mean ansatz

\[
  s_0 \propto M_p(\tilde{T}_A, \tilde{T}_B),
\]

(3.62)

defined by equation (3.58). In ideal boost-invariant hydrodynamics, the final charged-particle density scales linearly with the initial entropy density, i.e. \( dN_{\text{ch}}/d\eta \propto dS/d\eta_s \), so equation (3.62) implies that

\[
  dN_{\text{ch}}/d\eta \propto \int d^2x_\perp T_R(x_\perp),
\]

(3.63)

where \( T_R \) is the reduced thickness function defined by equation (3.61). The participant thickness functions \( \tilde{T}_A, \tilde{T}_B \) are easily simulated using equation (3.48), so I can test this prediction against experimental measurements.

Figure 3.11 shows the centrality dependence of charged-particle production \( dN_{\text{ch}}/d\eta \) predicted by equation (3.62) for Pb-Pb collisions at \( \sqrt{s_{\text{NN}}} = 2.76 \) TeV compared to experimental data from ALICE [62]. The blue line shows an arithmetic mean \( p = 1 \), the orange line shows a geometric mean \( p = 0 \), and the green line shows a harmonic mean \( p = -1 \). The participant thickness functions \( \tilde{T}_A, \tilde{T}_B \) were calculated using a nucleon width \( w = 0.6 \) fm and inelastic nucleon-nucleon cross section \( \sigma_{\text{NN}}^{\text{inel}} = 6.4 \) fm\(^2\), although these details do not strongly affect the results as shown.

Notice how the generalized mean parameter \( p \) varies the centrality dependence of charged-particle production, similar to the effect of \( \alpha \) in the two-component model. Smaller, more negative values of \( p \) pull the generalized mean towards the minimum of the two thickness functions (3.60). This attenuates entropy deposition in asymmetric regions of the
These asymmetric regions tend to be prevalent in peripheral collisions which occur at large impact parameters. Hence decreasing $p$ (making it more negative) suppresses particle production in peripheral events at large centralities. Conversely, the effect of the binary collision fraction in the two-component ansatz is to enhance particle production in central events. The two ansatzes thus predict markedly different scaling behavior.

These differences are imparted on the initial profiles of energy and entropy predicted by each model. Figure 3.12 compares the transverse profiles predicted by the two-component and generalized mean ansatzes—here assumed to be entropy densities—for a realistic Pb-Pb event at $\sqrt{s_{NN}} = 2.76$ TeV. The nucleon positions, cross sections and all other aspects of each calculation are identical, except for the chosen deposition ansatz.

First, look at the top row of the figure which shows the predictions of the generalized mean ansatz for $p = +1, 0, \text{ and } -1$. Smaller, more negative $p$ values compress the transverse profile along the impact parameter direction (here aligned with the $x$-axis), leading
Figure 3.12  Realistic transverse entropy density profiles predicted by the generalized mean ansatz \((3.58)\) with \(p = +1, 0, \) and \(-1\) (top row) compared to those predicted by the two-component ansatz \((3.54)\) with \(\alpha = 0, 0.1,\) and \(0.2\) (bottom row). Nucleon positions describe a peripheral Pb-Pb event at \(\sqrt{s_{NN}} = 2.76\) TeV. Each collision is identical except for the varied deposition ansatz.

to smaller, more elliptic fireballs. Meanwhile, the bottom row of the figure shows the predictions of the two-component ansatz using binary collision fractions \(\alpha = 0, 0.1,\) and \(0.2\). As the binary collision fraction \(\alpha\) increases, the profile becomes more concentrated in the center of the collision, but the overall ellipticity of the fireball changes mildly. Also notice, that for \(p = +1\) and \(\alpha = 0\), both ansatzes predict wounded nucleon scaling. Hence the far left profiles are identical.

The qualitative geometric scaling predicted by the generalized mean ansatz is of course nothing new. Models based on Color Glass Condensate (CGC) effective field theory have predicted similar scaling behavior for quite some time; for example, see reference [169]. It explains why saturation-based models generally predict profiles of energy and entropy with larger ellipticities [85, 169]. However, to the best of my knowledge, the generalized mean ansatz is the first model to effectively parametrize this degree of freedom on an event-by-event basis.
3.5 The TRENTo model

I want to demonstrate now that the generalized mean ansatz is able to reproduce important features of the experimental data. Before I proceed though, let me summarize each step described thus far which I’ve illustrated in figure 3.13. These steps complete the specification of the so-called TRENTo model [168] which stands for Reduced Thickness Event-by-event Nuclear Topology\(^1\). The name pays homage to Trento, Italy where fellow graduate student Jonah Bernhard and I first conceived the formalism. The model is summarized as follows:

1. First, I sample nucleon positions in the rest frame of each nucleus. Heavy-ion nucleon positions are sampled using a deformed two-parameter Fermi distribution (3.8), while light-ion nucleon positions are modeled on a case-by-case basis. Optionally, when modeling heavy-ions, nucleon positions are sampled subject to a minimum distance criteria \(|x_i - x_j| > d_{\text{min}}\) which mimics repulsive interactions between individual pairs of nucleons. I then randomly sample an impact parameter offset \(b\) between the centroids of the two colliding nuclei and shift each set of nucleon positions by \(\Delta x_{\perp} = \pm b/2\).

2. Once the nucleon positions are determined, inelastic nucleon-nucleon collisions are sampled using an impact parameter dependent nucleon-nucleon collision profile (3.40) adapted from the analytic Glauber model, where each nucleon density \(\rho_n(x)\) is assumed to be a three-dimensional Gaussian distribution (3.7) of one-sigma width \(w\). This determines the subset of nucleons in each nucleus which collide inelastically.

3. The density of each nucleon \(\rho_n\) is then summed over \(N_{A,B}\), the total number of participants in each nucleus, to produce a three-dimensional participant density field

\[
\rho_{A,B}^{\text{part}}(x) = \sum_{i=1}^{N_{A,B}} \gamma_i \rho_n(x - x_i \pm b/2),
\]

\(^1\)Topology is a technical term that describes the study of geometric properties and spatial relations unaffected by continuous deformations. Geometry would be a more appropriate word for our acronym, but we liked the name TRENTo so we used the scale-invariant property of the generalized mean to justify our abuse of terminology.
1. Nucleon positions
2. Nucleon participants
3. Participant thicknesses
4. Reduced thickness

**Figure 3.13** Stages of the TRENTo initial condition model: 1) sample nucleon positions, 2) determine which nucleons participate inelastically, 3) construct the participant thickness functions $\tilde{T}_A$ and $\tilde{T}_B$, and 4) apply the generalized mean ansatz to calculate the reduced thickness function $T_R = M_p(\tilde{T}_A, \tilde{T}_B)$ which is renormalized to furnish the initial energy or entropy density.

which includes additional nucleon weights $\gamma_i$ sampled from a gamma distribution with unit mean and variance $1/k$. This participant density is finally projected onto the transverse plane $x_\perp$ by integrating over $z$ to construct two participant thickness functions

$$\tilde{T}_{A,B}(x_\perp) = \int dz \tilde{\rho}_{A,B}^\text{part}(x_\perp, z). \quad (3.65)$$

These participant thickness functions describe the fluctuated density of participant matter in each nucleus “seen” by a probe that moves parallel to the beam axis and pierces the transverse coordinate $x_\perp$.

4. Finally, I set the initial energy density $e_0$ or entropy density $s_0$ proportional to the reduced thickness function

$$\begin{pmatrix} e_0 \\ s_0 \end{pmatrix} = \text{Norm} \times T_R, \quad (3.66)$$

defined as the generalized mean $T_R = M_p(\tilde{T}_A, \tilde{T}_B)$ of the participant thickness functions $\tilde{T}_A, \tilde{T}_B$. The normalization prefactor is tuned once at a given beam energy $\sqrt{s_{NN}}$, after which it remains constant for all collision systems at the same energy.

In principle, the only way to rigorously test the model is to evolve it through a realistic transport model such as viscous hydrodynamics. The output of the simulation can then be used to assess the predictions of the model compared to the experimental data. Indeed, I will do exactly this later in chapter 5. However, such calculations are laborious, and they
often involve considerable computing resources. Hence, before I perform such an analysis, let me first discuss properties of the initial state which do not require a full-blown transport simulation to study.

3.5.1 Mimicking existing models

One of the primary strengths of the TRENTo model is its flexibility. It can mimic a large number of initial condition models proposed in the literature, including (but not limited to) specific calculations in Color Glass Condensate effective field theory. This section demonstrates this flexibility for several well-known initialization models by comparing each model’s prediction for the initial entropy density $s_0$ as a function of $\tilde{T}_A, \tilde{T}_B$. The following text and figures are adapted from my own contributions to


which I’ve lightly edited to conform to the notation used throughout this dissertation.

As discussed previously, one of the simplest and oldest models of heavy-ion initial conditions is the wounded nucleon model which deposits energy or entropy for each nucleon which engages in one or more inelastic collisions [159]. Consider, for example, the entropy variant of the wounded nucleon model $s_0 \propto n_{\text{part}}$, which may be expressed in terms of the participant thickness functions $\tilde{T}_A, \tilde{T}_B$ in its Monte Carlo formulation [170–173] as

$$s_0 \propto \tilde{T}_A + \tilde{T}_B. \quad (3.67)$$

This form is exactly equivalent the generalized mean ansatz $s_0 \propto M_p(\tilde{T}_A, \tilde{T}_B)$ using $p = 1$, so this model is a specific subcase of the TRENTo model.

More sophisticated calculations of the entropy density $s_0$ can be derived from Color Glass Condensate (CGC) effective field theory. A common implementation of a CGC based
saturation picture is the KLN model [174–176], in which entropy deposition at the QGP thermalization time can be approximated from the produced gluon density, \( s_0 \propto N_g \), where

\[
\frac{dN_g}{dy d^2 p_{\perp}} \sim Q_{s,\min}^2 \left[ 2 + \log \left( \frac{Q_{s,\max}^2}{Q_{s,\min}^2} \right) \right],
\]

and \( Q_{s,\max} \) and \( Q_{s,\min} \) denote the larger and smaller values of the two saturation momenta in opposite nuclei at any fixed position in the transverse plane [101]. In the original formulation of the KLN model, the two saturation scales are proportional to the local participant nucleon density in each nucleus, \( Q_{s,A}^2 \propto \tilde{T}_A \), and the entropy density can be recast as

\[
s_0 \sim \tilde{T}_{\min} \left[ 2 + \log(\tilde{T}_{\max}/\tilde{T}_{\min}) \right],
\]

where \( \tilde{T}_{\min}, \tilde{T}_{\max} \) are the minimum and maximum of the two participant thickness functions respectively.

Another saturation model which attracted interest after it successfully described an extensive list of experimental particle multiplicity and flow observables [114, 116] is the previously discussed EKRT model, which combines collinearly factorized pQCD minijet production with a simple conjecture for gluon saturation [113, 177]. The energy density predicted by the model after a pre-thermal Bjorken free streaming stage is given by

\[
e_0 \sim \frac{K_{\text{sat}}}{\pi} p_{\text{sat}}^3 (K_{\text{sat}}, \beta; T_A, T_B),
\]

where the saturation momentum \( p_{\text{sat}} \) depends on the nuclear thickness functions \( T_A \) and \( T_B \), as well as two phenomenological model parameters \( K_{\text{sat}} \) and \( \beta \). Calculating the saturation momentum in the EKRT formalism is computationally intensive, and hence—in its Monte Carlo implementation—the model parametrizes the saturation momentum \( p_{\text{sat}} \) to facilitate efficient event sampling [114]. The energy density in equation (3.70) can then be recast as an entropy density using the thermodynamic relation \( s \sim e^{3/4} \) to compare it with the previous models.
Figure 3.14  Profiles of the initial thermal distribution predicted by the KLN (left), EKRT (middle), and wounded nucleon (right) models (dashed black lines) compared to a generalized mean with different values of the parameter $p$ (solid blue lines). Staggered lines show different slices of the initial entropy density $s_0$ as a function of the participant nucleon density $\tilde{T}_A$ for several values of $\tilde{T}_B = 1, 2, 3 \text{ fm}^{-2}$. The EKRT mapping is shown with model parameters $K = 0.64$ and $\beta = 0.8$ [114]. Entropy normalization is arbitrary. Figure and caption are from [178].

Note that equation (3.70) is expressed as a function of nuclear thickness $T$ which includes contributions from all nucleons in the nucleus, as opposed to the participant thickness $\tilde{T}$. In order to express initial condition mappings as functions of a common variable one could, e.g. relate $\tilde{T}$ and $T$ using an analytic wounded nucleon model. The effect of this substitution on the EKRT model is small, as the mapping deposits zero entropy if nucleons are non-overlapping, effectively removing them from the participant thickness function. We thus replace $T$ with $\tilde{T}$ in the EKRT model and note that similar results are obtained by recasting the wounded nucleon, KLN, and TRENTo models as functions of $T$ using standard Glauber relations.

Figure 3.14 shows the midrapidity entropy density $s_0$ predicted by each of the aforementioned models as a function of $\tilde{T}_A, \tilde{T}_B$. Each line is a one-dimensional cut through the two-dimensional surface $s_0(\tilde{T}_A, \tilde{T}_B)$. We fix $\tilde{T}_B = 1, 2, 3 \text{ fm}^{-2}$ and vary $\tilde{T}_A$ from 0 to 4 fm$^{-2}$ to simulate typical values probed by a heavy-ion collision. The dashed lines are the entropy densities $s_0$ of the KLN, EKRT and wounded nucleon (WN) models, and the solid lines are those predicted by the generalized mean ansatz, using $p$ values tuned to fit each model. The figure illustrates the ability of the generalized mean ansatz to reproduce different initial condition calculations and quantifies differences among them in terms of the generalized mean parameter $p$. The KLN model, for example, is well-described by $p \sim -0.67$, the EKRT
Figure 3.15  Eccentricity harmonics $\varepsilon_2$ and $\varepsilon_3$ as a function of impact parameter $b$ for Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV calculated from IP-Glasma and TRENTo initial conditions. IP-Glasma events are evaluated after $\tau = 0.4$ fm/$c$ classical Yang-Mills evolution [85]; TRENTo events after $\tau = 0.4$ fm/$c$ free streaming [90, 91] and using parameters $p = 0 \pm 0.1$, $k = 1.6$, and nucleon width $w = 0.4$ fm to match IP-Glasma [184]. Figure and caption are from [178].

model corresponds to $p \sim 0$, and the wounded nucleon model is precisely reproduced by $p = 1$. Smaller, more negative values of $p$ pull the generalized mean toward a minimum function and hence correspond to models with more extreme gluon saturation effects.

The three models considered in figure 3.14 are by no means an exhaustive list of proposed initial condition models, see e.g. references [173, 179–183]. Notably absent, for instance, is the highly successful IP-Glasma model which combines IP-Sat CGC initial conditions with classical Yang-Mills dynamics to describe the full pre-equilibrium evolution of produced glasma fields [85, 86, 104]. The IP-Glasma model lacks a simple analytic form for initial energy (or entropy) deposition at the QGP thermalization time and so it cannot be directly compared to the generalized mean ansatz. In lieu of such a comparison, we examined the geometric properties of the IP-Glasma and TRENTo models through their eccentricity harmonics $\varepsilon_n$.

We generated a large number of TRENTo events using entropy deposition parameter $p = 0$, Gaussian nucleon width $w = 0.4$ fm, and fluctuation parameter $k = 1.6$, which were previously shown to reproduce the ratio of ellipticity and triangularity in IP-Glasma [168]. We then free streamed [90, 91] the events for $\tau = 0.4$ fm/$c$ to mimic the weakly coupled pre-equilibrium dynamics of IP-Glasma and match the evolution time of both models. Fi-
nally, we calculated the eccentricity harmonics $\varepsilon_2$ and $\varepsilon_3$ weighted by energy density $e(x,y)$ according to the definition

$$\varepsilon_n e^{inc} = -\frac{\int dx \, dy \, r^n e^{inc} e(x,y)}{\int dx \, dy \, e(x,y)},$$  \hspace{1cm} (3.71)$$

where the energy density is the time-time component of the stress-energy tensor after the free streaming phase, $T^{00}$. The resulting eccentricities, pictured in figure 3.15, are in good agreement for all but the most peripheral collisions, where sub-nucleonic structure becomes important. This similarity suggests that TRENTo with $p \sim 0$ can effectively reproduce the scaling behavior of IP-Glasma, although a more detailed comparison would be necessary to establish the strength of correspondence illustrated in figure 3.14.

Needless to say, there are also models in the literature which cannot be approximated by the generalized mean ansatz. The generalized mean ansatz is scale-invariant, i.e.

$$M_p(cT_A, cT_B) = cM_p(T_A, T_B)$$ \hspace{1cm} (3.72)$$

for any nonzero constant $c$, so the parametrization can only mimic initial condition models which scale accordingly. Note, this property is clearly broken by the binary collision term $n_{coll}$ in the two-component ansatz (3.54). This property may seem overly restrictive. However, I’ll show somewhat later in the next section that even small deviations from scale invariance produce tension with the data. The scale-invariant postulate thus appears to be a necessary requirement for physically reasonable models.²

### 3.5.2 Application to experimental data

This section demonstrates TRENTo’s ability to simultaneously describe a wide range of collision systems, using approximate scaling laws to compare the model predictions with experimental data. The text and figures in this subsection section are adapted from one of

²Recently, it has come to light that the IP-Glasma model predicts local energy density deposition which scales approximately like the product of nuclear thickness functions $e_0 \propto T_A T_B$ [41, 185]. This functional form clearly breaks scale invariance. Understanding this feature of the model and its compatibility with the data is currently the subject of active investigation.
my publications,


which I’ve lightly edited for clarity and formatting. Note, the TRENTo reduced thickness parameter \( p \), gamma fluctuation parameter \( k \), and nucleon profile \( \rho_n \) used here are not rigorously constrained—doing so requires advanced Bayesian statistical machinery which I’ll introduce in the next chapter. Therefore, the results of this section do not necessarily represent the best possible fit of the model to data.

In this publication, we compared the TRENTo model with experimentally measured multiplicity distributions using a three-stage model for particle production, similar to that of reference [155], in which the final multiplicity arises from a convolution of the initial entropy deposited by the collision, viscous entropy production during hydrodynamic evolution, and statistical hadronization at freeze-out. The average charged-particle multiplicity \( \langle N_{ch} \rangle \) after hydrodynamic evolution is roughly proportional to the total initial entropy [37] and hence to the integrated reduced thickness via equation (3.66):

\[
\langle N_{ch} \rangle \propto \int dx \, dy \, T_R. \tag{3.73}
\]

Then, assuming independent particle emission at freeze-out, the final number of charged particles is Poisson distributed [186, 187], i.e. \( P(N_{ch}) = \text{Poisson}(\langle N_{ch} \rangle) \). The folding of the Poisson fluctuations with the gamma weights for each participant yields a negative binomial distribution [155], which has historically been used to fit proton-proton multiplicity fluctuations (see subsection 3.3.5).

Following this prescription, we generated a large ensemble of minimum-bias events, integrated their \( T_R \) profiles, rescaled them by an overall normalization constant, and sampled a Poisson number for the multiplicity of each event. The left panel of figure 3.16 shows the \( N_{ch} \) distributions for proton-proton simulations with reduced thickness parameter \( p = 1, 0, \ldots \).
Figure 3.16  Multiplicity distributions for proton-proton, proton-lead, and lead-lead collisions. The histograms are TRENTo results for reduced thickness parameter $p = −1$ (top, orange), $p = 0$ (middle, blue), and $p = 1$ (bottom, green), with approximate best-fit fluctuation parameters $k$ and normalizations given in table 3.2. The shaded bands show the sensitivity from varying $k$ by $±30\%$. Data points (triangles, squares, circles) are experimental distributions from ALICE [152, 188] offset by powers of ten for comparison with the model.

$\int dz \rho_n = \frac{1}{2\pi B} \exp \left( -\frac{x^2 + y^2}{2B} \right)$

(3.74)

with effective area $B = (0.6 \text{ fm})^2$. We tuned the fluctuation parameter $k$ for each value of $p$ to qualitatively fit the experimental proton-proton distribution [152], and additionally varied $k$ by $±30\%$ to explore the sensitivity of the model to the gamma participant weights. For proton-lead and lead-lead collisions [188] (middle and right panels), we used identical model parameters except for the overall normalization factor, which was allowed to vary independently across collision systems to account for differences in beam energy and kinematic cuts (annotated in the figure). The $k$ values and normalizations are provided in table 3.2.

Table 3.2  Approximate best-fit fluctuation parameters $k$ and normalizations for each $p$ value and collision system in figure 3.16.

<table>
<thead>
<tr>
<th>$p$</th>
<th>$k$</th>
<th>p-p norm</th>
<th>p-Pb norm</th>
<th>Pb-Pb norm</th>
</tr>
</thead>
<tbody>
<tr>
<td>+1</td>
<td>0.8</td>
<td>9.7</td>
<td>7.0</td>
<td>13.</td>
</tr>
<tr>
<td>0</td>
<td>1.4</td>
<td>19.</td>
<td>17.</td>
<td>16.</td>
</tr>
<tr>
<td>−1</td>
<td>2.2</td>
<td>24.</td>
<td>26.</td>
<td>18.</td>
</tr>
</tbody>
</table>
Figure 3.17  Average charged particle density per participant nucleon pair \((dN_{ch}/d\eta)/(N_{part}/2)\) at midrapidity as a function of the participant number \(N_{part}\) for Pb-Pb, p-Pb, and Au-Au systems at various collision energies. Lines are TRENTo calculations with generalized mean parameter \(p = 0\), and symbols are data from PHENIX [128] and ALICE [56, 62]. The average minimum bias participant number for p-Pb is shifted for clarity. Figure and caption are from [178].

The model is able to reproduce the experimental proton-proton distribution for each value of \(p\), provided \(k\) is appropriately tuned. Varying the best-fit \(k\) value (by \(\pm 30\%\)) has a noticeable effect on proton-proton and proton-lead systems, especially in the high-multiplicity tails, but is less important in lead-lead collisions, where the gamma weights are averaged over many participant nucleons. Each \(p\) value also yields a reasonable fit to the shapes of the proton-lead and lead-lead distributions, although lead-lead appears to favor \(p \approx 0\). Note that the normalizations for \(p = 1\) (wounded nucleon model) in proton-lead and lead-lead collisions (table 3.2) are not self-consistent, since proton-lead requires roughly half the normalization as lead-lead, even though the experimental data were measured at a higher beam energy.

In a somewhat later publication [178], we extended the yield comparison for \(p = 0\) to include additional beam energies. Figure 3.17 shows the charged-particle density per participant pair \((dN_{ch}/d\eta)/(N_{part}/2)\) at midrapidity as a function of participant number for proton-lead, gold-gold, and lead-lead collisions with beam energies spanning several orders in magnitude [178]. The colored lines are TRENTo model calculations using \(p = 0\), and the black symbols are experimental data from PHENIX [128] and ALICE [62]. The TRENTo
Figure 3.18  Left and middle plots: Eccentricity harmonics $\varepsilon_2$ and $\varepsilon_3$ as a function of centrality for reduced thickness parameters $p = 1, 0, -1$ (green, blue, orange). The shaded bands show the sensitivity from varying $k$ by $\pm 30\%$ from the values in table 3.2. Right plot: Ratio of the rms eccentricities $\sqrt{\langle \varepsilon_2^2 \rangle}/\sqrt{\langle \varepsilon_3^2 \rangle}^{0.6}$ against the allowed region (grey band) and the ratio computed by IP-Glasma (circles) [189]. Note that the axes have different ranges in the ratio plot.

calculations provides a superb fit to the experimental data, even fitting the average yield of the $p$-Pb data point using the same overall normalization as the Pb-Pb system at the same beam energy, consistent with previous observations.

In reference [168], we also examined the sensitivity of the eccentricity harmonics $\varepsilon_n$ to different values of the generalized mean parameter $p$ and fluctuation factor $k$. The reduced thickness function eccentricity equals

$$\varepsilon_n e^{i\phi} = -\frac{\int dx \, dy \, r^n e^{i\phi} T_R}{\int dx \, dy \, r^n T_R}.$$  \hspace{1cm} (3.75)

Figure 3.18 shows ellipticity $\varepsilon_2$ and triangularity $\varepsilon_3$ as a function of centrality using the same lead-lead data as in figure 3.16. There is a clear trend of increasing eccentricity (particularly $\varepsilon_2$) with decreasing $p$. As $p$ decreases, the generalized mean (3.58) attenuates entropy production in asymmetric regions of the collision, accentuating the elliptical overlap shape in non-central collisions and enhancing their eccentricity. Meanwhile, varying the fluctuation parameter $k$ has limited effect.

In addition, we performed the test proposed by [189], which uses flow data and hydrodynamic calculations to determine an experimentally allowed band for the ratio of root-mean-square eccentricities $\sqrt{\langle \varepsilon_2^2 \rangle}/\sqrt{\langle \varepsilon_3^2 \rangle}^{0.6}$ as a function of centrality. Among the initial condition models available at that time, only IP-Glasma consistently falls within the allowed region. As shown in the right panel of figure 3.18, the TRENTo model with $p = 0$ (geometric mean)
yields excellent agreement with the allowed band and is similar to IP-Glasma.

As a final novel application, we explained how the generalized mean ansatz resolves an apparent puzzle in uranium-uranium collisions at RHIC. Unlike e.g. gold and lead, uranium nuclei have a highly deformed prolate spheroidal shape, so uranium-uranium collisions may achieve maximal overlap via two distinct orientations: “tip-tip”, in which the long axes of the spheroids are aligned with the beam axis and the overlap area is circular; or “side-side”, where the long axes are perpendicular to the beam axis and the overlap area is elliptical, as shown in figure 3.19. Hence side-side collisions will in general have larger initial-state ellipticity $\varepsilon_2$ and final-state elliptic flow $v_2$ than tip-tip.

In the two-component Glauber model, tip-tip collisions produce more binary nucleon-nucleon collisions than side-side, so tip-tip collisions have larger charged-particle multiplicity $N_{\text{ch}}$. Therefore, the most central uranium-uranium events are dominated by tip-tip collisions with maximal $N_{\text{ch}}$ and small $v_2$, while side-side collisions have a smaller $N_{\text{ch}}$ and somewhat larger $v_2$. This predicted drop in elliptic flow as a function of $N_{\text{ch}}$ is known as the “knee” [190]. Data from STAR on uranium-uranium collisions, however, exhibits no evidence of a knee [191, 192], at odds with the predictions of the two-component model. It has been proposed that fluctuations could wash out the knee [193], but a detailed flow analysis showed that it would still be visible [194].

The data therefore imply that multiplicity is independent of the number of binary collisions, justifying the scale-invariant condition (3.72), postulated during the construction of

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{figure3.19}
\caption{Comparison of tip-tip and side-side uranium-uranium collisions. Schematics are shown from a side view and looking down the beam axis, and the following quantities are compared: ellipticity $\varepsilon_2$, number of participating nucleons $N_{\text{part}}$, and number of binary nucleon-nucleon collisions $N_{\text{coll}}$.}
\end{figure}
the reduced thickness ansatz (3.61). Due to this assumed scale invariance, TRENTo predicts roughly the same number of charged particles in tip-tip and side-side uranium-uranium collisions. As shown in figure 3.20, the slope of $\varepsilon_2$ as a function of $N_{\text{ch}}$ is approximately equal for uranium-uranium and gold-gold, in contrast to the two-component Glauber model which predicts a much steeper slope for uranium. Short of conducting a full hydrodynamic analysis, TRENTo appears to be more consistent with STAR data than the two-component model, and behaves similarly to IP-Glasma [195].

3.5.3 Proton-proton collision geometry

The TRENTo model makes no distinction between small and large collision systems, so it can also be used to predict the profiles of energy (or entropy) produced by a single proton-proton collision. The framework is admittedly strained—I crudely modeled each nucleon as a Gaussian blob—nevertheless, it should provide some qualitative insight into the macroscopic geometry of the produced fireball.

Figure 3.21 shows a cross section of the reduced thickness function $T_R$ for the collision of two Gaussian protons at non-zero impact parameter with the gamma fluctuations (3.47)
Figure 3.21  Cross section of the reduced thickness function for a pair of nucleon participants. The nucleons collide with a nonzero impact parameter along the $x$-direction as shown in the upper right. The black dashed lines are one-dimensional cross sections of the participant nucleon thickness functions $\tilde{T}_A, \tilde{T}_B$, and the colored lines are the reduced thickness $T_R$ for $p = 1, 0, -1$ (green, blue, orange). Figure and caption are from [168].

turned off. Each colored line is the generalized mean for a different value of $p$. The green line is an arithmetic mean $p = 1$, the blue line is a geometric mean $p = 0$, and the orange line is a harmonic mean $p = -1$. The black dashed lines outline each nucleon participant, scaled down by a factor of three for visibility. Meanwhile, figure 3.22 shows the same reduced thickness functions plotted as heatmaps in the transverse plane. As before, the black dashed lines mark the outline of each nucleon participant.

Notice how the reduced thickness becomes more sharply peaked in the overlap region as $p \to -\infty$, producing a highly elliptic fireball. Conversely, for $p \to +\infty$, the fireball grows in size and becomes oblong, reproducing the underlying density of each Gaussian participant. For the intermediate value $p = 0$, the energy (or entropy) produced by the collision is a Gaussian located midway between the two interacting protons.

Intriguingly, Bzdak et al. pondered these exact three proton-proton collision scenarios two years before the generalized mean ansatz was even conceived; see figure 1 in reference [196]. The TRENTo model explains these scenarios as certain subcases of the overall energy and entropy deposition mapping. Hence, if we can rigorously constrain the parameter $p$, it might help to lift the veil on the distribution of energy and entropy deposited by individual
Figure 3.22 Reduced thickness for a pair of nucleon participants using a harmonic mean $p = -1$ (left), geometric mean $p = 0$ (middle), and arithmetic mean $p = 1$ (right). The black dashed circles are contours showing the location of each Gaussian nucleon participant. Geometry parallels figure 3.21.

proton-proton collisions. The next chapter describes some advanced statistical machinery which will help rigorously pin down the parameters of the TRENTo initial condition model, enabling state-of-the-art extractions of the QGP transport coefficients.
Quark-gluon plasma produced in the laboratory is not directly observable. Relativistic nuclear collisions produce small droplets of the substance, and these droplets expand and cool rapidly, disintegrating into showers of separate particles on the order of $\sim 10^{-23}$ seconds. All direct physical traces of the QGP are thus long gone by the time the collision is measured inside the detector. Studying the QGP is hence a forensic endeavor. Experimentalists measure final state properties of the collision, and theorists work to reconstruct the causal factors which explain their findings. Such problems are called inverse problems because the results of the process are known but its root causes are not.

Inverse problems are particularly challenging to solve when information is lost or scrambled during the process of interest. For example, if someone kicks over a sandcastle, it’s difficult to reconstruct the sandcastle from a picture of its demolished state. The same is also true for a relativistic nuclear collision event. These events generate considerable entropy, and hence—according to the second law of thermodynamics—their processes are irreversible. This prevents one from simply running the dynamics of a nuclear collision simulation in reverse.

That’s not to say, however, that all information is lost during an irreversible process. If someone kicked over a sandcastle, I could still estimate its total mass and get a rough idea of its initial shape. If I wanted to be more precise, I could build lots of different sandcastles
and kick each one over until I found one that resembled the demolished sandcastle I sought to reconstruct, i.e. I could attempt to solve the sandcastle inverse problem using brute force trial-and-error.

Trial-and-error is particularly useful for A/B testing, i.e. testing whether causal factors are better explained by process A or process B. For example, it was originally unknown whether the QGP would behave like a strongly coupled liquid or a weakly coupled gas. Researchers simulated the space-time dynamics of heavy-ion collisions using both descriptions and found that strong coupling provides a far superior description of the data [73, 74].

This is, of course, an over simplification of a more nuanced problem. The QGP is neither infinitely strongly coupled nor infinitely weakly coupled but something inbetween. The relevant problem-space is hence continuous, not discrete. We can therefore frame the problem more rigorously if we imagine a meta-model of a single parameter $x$ which varies the simulation’s coupling strength. For example, suppose that for $x \sim 0$ the meta-model is weakly coupled, and for $x \gg 1$ it is strongly coupled. Additionally, suppose there is some simulated observable $y$ which is sensitive to the coupling strength $x$. Solving the inverse problem amounts to estimating the value of the coupling strength $x$, given some estimate (experimental measurement) for the simulated observable $y = f(x)$.

This one-parameter model is merely a pedagogical example. Most nuclear collision models include multiple parameters $x = (x_1, x_2, \ldots, x_n)$. For example, one parameter might vary the QGP’s average shear viscosity while another varies its average bulk viscosity. They also generally predict more than one observable. For instance, a model might simulate a large vector of observables $y = (y_1, y_2, \ldots, y_m)$ describing, for example, the system’s charged-particle yield in different centrality bins. The model calibration problem hence seeks to estimate the model parameters $x$ using the simulation predictions $y = f(x)$ and their global fit to experimental measurements, described by some third vector $y_e$.

The model parameters $x$ typically correlate among each other and affect multiple observables. Hence, if one wishes to obtain a global fit to the data, the only option is to fit all parameters simultaneously. If there are only one or two model parameters, it is often
possible to tune their values using a brute force approach or manually by hand, but when
the number of parameters is large, this becomes infeasible. A graduate student could easily
spend an entire lifetime guessing and checking parameters if they live in a high-dimensional
space.

The experimental data \( y_e \) is also inherently uncertain. Measurement errors include, for
example, statistical and systematic uncertainties such as finite event statistics and imper-
fect detector response. The existence of these uncertainties naturally limits the ability to
estimate \( x \) precisely. As a simple example, consider a “physics model” consisting of one
parameter \( x \) and one observable \( y \) trivially related by

\[
y = cx,
\]  

(4.1)

where \( c = 0.1 \) is a fixed coefficient that specifies the model. Moreover, assume that \(-1 < x < 1\) is bounded for physical reasons. Suppose I measure the observable \( y \) and find \( y_e = 0.0^{+0.5}_{-0.5} \).

Can I conclude anything interesting about the parameter \( x \)? Sadly, no I cannot. The model
maps the parameter range \(-1 < x < 1\) to the observable range \(-0.1 < y < 0.1\) which
is well within the experimental uncertainty band. Essentially all allowed values of \( x \) are
equally supported by the data! However, if I blindly invert the function \( y = f(x) \) ignoring
this measurement uncertainty, I find a single preferred value \( x = 0 \). This conclusion is of
course nonsense; there is nothing statistically special about this one value. Measurement
uncertainties are hence a crucial component of the inverse problem.

In this chapter, I describe a general statistical framework developed to address the
aforementioned challenges of quantitative model-to-data comparison. Commonly known as
Bayesian parameter estimation, the procedure describes a powerful methodology to con-
strain the parameters of computationally intensive computer models. I should emphasize
that I did not develop the framework myself. The groundwork was developed over time by
numerous contributors [197–199]. I also did not adapt the framework to heavy-ion collisions.
That effort was the subject of previous dissertations [140, 200] and numerous published pa-
pers [38, 201–203]. However, I use the framework extensively in this dissertation to study
the properties of the QGP initial conditions [157, 178, 204], and a firm understanding of Bayesian parameter estimation is essential to communicate these results.

The goal of this chapter is to introduce Bayesian parameter estimation using a simple example to capture the essential features of the methodology. Disclaimer: in this chapter and throughout this dissertation, I use a suite of open source software tools developed by fellow Duke graduate student Jonah Bernhard. His dissertation, *Applying Bayesian parameter to heavy-ion collisions* [140], is the primary resource for this chapter, and many of my graphics are generated using his own original software. For a more thorough discussion of Bayesian parameter estimation and its application to heavy-ion collisions, I recommend reading his dissertation.

4.1 Problem statement

The case study for this chapter is a simple vector-valued function $y = f(x)$ with three input parameters $x = (x_1, x_2, x_3)$ and ten output observables $y = (y_1, y_2, \ldots, y_{10})$. This mapping $f : x \mapsto y$ is assumed to model some physical process in nature. The vector $x$ describes unknown attributes of the process, and the vector $y$ describes their measurable consequences.

Suppose that the function $f$ is described by the simple linear form

$$y = x_1 u + (x_2 + x_3) v,$$  \hspace{1cm} (4.2)

where $u = (1, 1, \ldots, 1)$ is a vector of ones, and $v = (0.05, 0.15, \ldots, 0.95)$ is a vector that uniformly partitions the interval $[0, 1]$, both with ten elements. Moreover, suppose that we are unable to evaluate this function exactly, but we can approximate it with some computer model $f_{\text{model}} : x \mapsto y + \delta y$, where $\delta y$ is a vector of uncorrelated random noise sampled from the normal distribution

$$\delta y_i = \mathcal{N}(\mu, \sigma) \quad \text{for } i = 1, 2, \ldots, 10,$$  \hspace{1cm} (4.3)

with mean $\mu = 0$ and standard deviation $\sigma = 0.02$. Finally, assume that this function $f_{\text{model}}$
is slow to evaluate, e.g. suppose it takes one processor hour of computation time to evaluate
the function once.

Now imagine the following scenario. Someone secretly writes down

\[ x_1 = 0.3, \quad x_2 = 0.5, \quad \text{and} \quad x_3 = 0.7 \]  

(4.4)
on a piece of paper. This someone is the omniscient creator, and they feed these parameter
values through the physical process modeled by the function \( y = f(x) \). An experimentalist
then measures the output of this process and reports their measurement \( y_e \) along with its
uncertainties, quantified by the covariance matrix

\[
(S_e)_{ij} = \text{cov}(y_i, y_j) \\
= E[(y_i - \langle y_i \rangle)(y_j - \langle y_j \rangle)],
\]  

(4.5)

where \( y_i \) and \( y_j \) are two observables from the vector \( y_e \), and \( E[y] \) is the expected value of
\( y \). You are finally given the data \( y_e \) and \( \Sigma_e \) and tasked with estimating the true physical
parameters \( x_{true} = (0.3, 0.5, 0.7) \) using the computer model \( y = f_{\text{model}}(x) \). This chapter
describes how Bayesian parameter estimation can be used to solve the problem.

4.2 Parameter space

The aforementioned computer model maps a three-dimensional vector \( x \) to a ten-
dimensional vector \( y \). Here we’ve assume that \( f_{\text{model}} \) is a linear equation, but imagine
instead that it is something far more complicated, e.g. a large scale computer simulation
that maps QGP initial condition parameters to simulated hadronic observables. The func-
tion \( f_{\text{model}} \) is thus assumed to be messy, non-algebraic, and slow to evaluate. If we want to
figure out what it does, we need to do so empirically.

With this in mind, imagine that you set out to explore the mapping \( f_{\text{model}} : x \mapsto y + \delta y \).
You do not know the true parameters \( x_{true} \), but you are told by the omniscient creator
that each parameter $x_1, x_2, x_3 \in [0, 1]$. You therefore decide to evaluate the function $f_{model}$ at lots of different parameter points $X = (x_1, x_2, \ldots, x_d)$ uniformly distributed inside the unit-cube $[0, 1]^3$. To distribute these points, you assign each parameter $x_1$, $x_2$, and $x_3$ to one of $k$ possible values uniformly distributed between 0 and 1. You then take the Cartesian product of all possible parameter combinations. This produces a matrix

$$X = \left\{ \begin{array}{c}
(x_1)_1 \\
(x_1)_2 \\
\vdots \\
(x_1)_k
\end{array} \right\} \times \left\{ \begin{array}{c}
(x_2)_1 \\
(x_2)_2 \\
\vdots \\
(x_2)_k
\end{array} \right\} \times \left\{ \begin{array}{c}
(x_3)_1 \\
(x_3)_2 \\
\vdots \\
(x_3)_k
\end{array} \right\},$$

(4.6)

with $d = k^3$ rows and $n = 3$ columns. The row vectors $(x_1, x_2, \ldots, x_d)$ are called design points, and the matrix $X$ is called a design. This particular prescription for distributing the design points describes what’s known as a factorial design.

For the sake of our example, let’s say that $k = 100$, a seemingly reasonable number for the problem at hand. The resulting factorial design will have $d = 100 \times 100 \times 100 = 10^6$ design points. If our function takes one processor hour to evaluate, the entire design will require $\sim 100$ processor years! Clearly, we need to cap the number of design points at something reasonable and find a way to distribute these points more sensibly throughout the design space.

An algorithm known as Latin hypercube sampling is commonly used for this purpose [205, 206]. The Latin hypercube is a generalization of a Latin square to three or more dimensions. A Latin square is a $d \times d$ grid filled with $d$ symbols (points) where each symbol appears once (and only once) in each row and column (it should be familiar to anyone who has played a game of Sudoku). The algorithm hence distributes the symbols (points) uniformly across each row and column space. Figure 4.1 shows an example of a Latin square compared to a factorial design in two dimensions.

The Latin hypercube extends this idea for a square to an $n$-dimensional unit hypercube $[0, 1]^n$. Unlike the factorial design where $d$ points-per-dimension requires $d^n$ total points,
Figure 4.1  Left: $5 \times 5$ factorial design. Right: $5 \times 5$ Latin square. The Latin square contains one (and only one) point in each row and column. Both designs distribute their points uniformly across each dimension.

the Latin hypercube only requires $d$ total points. It therefore scales linearly with the desired points-per-dimension, not exponentially. At this point you might be wondering, “Why not take the $n$-th root of the total number of design points and distribute them according to a factorial design? What’s the advantage of the Latin hypercube algorithm?”

Often there exist model parameters that are inactive. This means that a parameter $x$ has no affect on the observables $y$. For example, let’s imagine in our example that $x_1$ and $x_2$ are inactive and that $x_3$ is active, i.e. $x_3$ is the only parameter that affects the observables $y$. Suppose that we create two separate designs: a $10 \times 10 \times 10$ factorial design with $d = 10^3$ parameter points, and a Latin hypercube design with $d = 10^3$ parameter points.

The parameters $x_1$ and $x_2$ are assumed to have no effect on the model so we can ignore their values. This leaves 10 distinct values for $x_3$ in the factorial design and $10^3$ distinct values for $x_3$ in the Latin hypercube design. This means that we wasted 99% of our design points repeating the same 10 parameter values in the factorial design! The Latin hypercube never repeats a parameter value so this redundancy is never an issue.

Returning to the problem at hand, let’s generate a Latin hypercube design for the three model parameters $x_1, x_2,$ and $x_3$ listed in equation (4.2). I generate, for this purpose, $d = 100$ parameter points using a publicly available Latin hypercube implementation written in R [207]. In addition to satisfying the Latin hypercube requirements, it also attempts to optimize the sample by maximizing the minimum distance between every pair of design points (maximin criteria). Note, the present parameters $x_1, x_2, x_3 \in [0, 1]$ lie inside a unit
Latin hypercube design with $d = 100$ design points and $n = 3$ three parameters $x_1$, $x_2$, and $x_3$ distributed in the interval $x_i \in [0, 1]$. The figure shows the design projected onto two dimensions, $x_1$ and $x_2$, with the marginal distributions (histograms) for each dimension shown above and to the right of the design.

hypercube $[0, 1]^3$, but this is generally not the case. It is therefore typically necessary to scale and shift the Latin hypercube design along each dimension to fit the desired parameter ranges.

Figure 4.2 shows the resulting three-dimensional design projected onto the two-dimensional subspace spanned by $x_1$ and $x_2$. The blue histogram at the top of the figure shows the marginal distribution of parameter $x_1$, and the blue histogram on the right shows the marginal distribution of parameter $x_2$. These distributions are uniform, illustrating a fundamental property of the Latin hypercube sampling procedure. Note that $d = 100$ points is overkill for this example; most physical problems only require $O(10)$ points per parameter dimension [208].

With the design in hand, I can then evaluate the computer model at each design point

$$f_{\text{model}} : (x_1, x_2, \ldots, x_d) \mapsto (y_1, y_2, \ldots, y_d), \quad (4.7)$$

to produce a $d \times m$ matrix $Y = (y_1, y_2, \ldots, y_d)$ of simulated model observables. Per specification, the computer model requires one processor hour to evaluate one point. The entire Latin hypercube design therefore requires 100 processor hours. Note, this number is
Figure 4.3 Simulated observables compared to the experimental data. Blue lines are explicit model calculations \( y = (y_1, y_2, \ldots, y_{10}) \) for each of the 100 design points, and black symbols with error bars are the mean \( y_e \) and standard deviation \( \sigma_e = \text{diag}(\Sigma_e)^{1/2} \) of the experimental measurement. The experimental covariance matrix \( \Sigma_e \) is described later in the chapter, Eq. (4.54).

actually reasonable compared to the 100 processor years required by the factorial design.

Figure 4.3 shows the resulting computer model output at each design point (blue lines) compared to the experimental data (black symbols). Recall that the experimental data is the output of the model using a certain set of true parameters, \( x_{\text{true}} = (0.3, 0.5, 0.7) \). The black bars on each symbol (barely visible) are the experimental uncertainties, equal to \( \sigma_e = \text{diag}(\Sigma_e)^{1/2} \), where \( \Sigma_e \) is the experimental covariance matrix (I'll specify this matrix later in the chapter). Notice the slight wiggle that is visible in each model calculation caused by the statistical noise term in equation (4.3). The model outputs also have a large visual spread resulting from the varied parameter combinations sampled by the design matrix \( X \). Evidently, our design scaffolding nicely covers the data, and there appear to be some parameter values which describe the experimental data, although it is not yet clear what these values are.

4.3 Computer model emulator

Seeing as we’ve already spent 100 processor hours evaluating the computer model at each design point, we can’t afford to evaluate it any more times. If we want to calculate \( y' = f_{\text{model}}(x') \) at some new parameter point \( x' \), we’ll need to leverage our previous observations \( f_{\text{model}} : X \mapsto Y \). Emulators are essentially fancy interpolators that serve this purpose.
Broadly speaking, an emulator is a black box that accepts the same inputs and predicts the same outputs as the model that it emulates. It is first trained or “calibrated” to reproduce a finite set of input-output observations. Once trained, the emulator acts as a fast surrogate for the full model calculation, enabling essentially instantaneous predictions at arbitrary points in parameter space.

4.3.1 Gaussian processes

There are many different types of emulators in the literature, each having its own advantages and disadvantages. When performing Bayesian parameter estimation, the convention is to use a specific type of emulator known as a Gaussian process (GP) emulator [209].

GP emulators have several strengths which make them popular for Bayesian parameter estimation:

- They are non-parametric: GPs do not assume a specific algebraic form for the underlying function of interest.
- They readily generalize to functions with multiple input variables, i.e. they have multivariate support.
- They predict a distribution of values. GPs return a mean prediction as well as an uncertainty estimate.

This last point is key. Bayesian parameter estimation seeks to estimate the model parameters $\mathbf{x}$ with quantitative uncertainty. Emulators are naturally imperfect, and their prediction error biases parameter estimates. In order to preserve the integrity of our parameter estimates, it’s imperative that we account for our emulator error. GPs also have one significant limitation:

- They only support single-valued functions, $y = f(\mathbf{x})$.

There is, however, a simple procedure that can be used to circumvent this issue, enabling the application of GPs to vector-valued functions. I’ll postpone describing the details of
this procedure till later in the chapter; for now, it suffices to say that such a work-around exists. I’ll therefore assume that we are only dealing with single-valued output for now, and I’ll return to the issue of vector-valued output later.

A GP generalizes a multivariate normal distribution from the space of vectors to the space of functions. We say that a \( k \)-dimensional random vector \( \mathbf{y} = (y_1, y_2, \ldots, y_k) \) is normally distributed if

\[
P(\mathbf{y}) = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \exp \left( -\frac{1}{2} (\mathbf{y} - \mu)^T \Sigma^{-1} (\mathbf{y} - \mu) \right).
\]

This normal distribution is parametrized by a mean vector \( \mu \) and a positive semi-definite covariance matrix \( \Sigma_{ij} = \text{cov}(y_i, y_j) \) where \( y_i \) and \( y_j \) are two elements of the vector \( \mathbf{y} \). Hereafter, I’ll write this equation more compactly using the shorthand notation

\[
\mathbf{y} \sim \mathcal{N}(\mu, \Sigma),
\]

to signify a normal random vector \( \mathbf{y} \) with mean \( \mu \) and covariance \( \Sigma \).

Crudely speaking, we can also think of an arbitrary function as a vector. For example, given some single-valued multivariate function \( y = f(x) \), we can evaluate the function at a discrete number of points

\[
f : (x_1, x_2, \ldots, x_k) \mapsto (y_1, y_2, \ldots, y_k),
\]

to approximate its behavior over some finite region of its domain. Taking the number (and density) of evaluation points to infinity, this discrete realization converges to the function \( y = f(x) \), assuming it is continuous. A function is thus something like a vector of infinite length. Given enough evaluation points, it contains essentially all of the same information.

Abstractly, a GP is a multivariate normal distribution over functions, i.e. it samples a continuous curve (or hypersurface) instead of a discrete vector. This random curve is defined by a collection of random variables, and these variables are constructed such that every finite
number of them shares a multivariate normal distribution. Numerical implementations naturally cannot generate continuous functions, so in practice these curves are discretized.

Let \( X = (x_1, x_2, \ldots, x_k) \) be some points that discretize each curve. Given these points, a GP predicts a random vector of corresponding function values \( y_p = (y_1, y_2, \ldots, y_k) \) that is distributed according to a multivariate normal distribution

\[
y_p \sim \mathcal{N}(\mu, \Sigma).
\] (4.11)

This property also implies that the GP output \( y_p \) at a single point \( x \) is a normal random variable with some mean value \( \mu \) and variance \( \sigma^2 = \text{cov}(y_p, y_p) \).

Before a GP can be sampled to make predictions, it must be conditioned on some training data to determine the values of \( \mu \) and \( \Sigma \) in equation (4.11). This conditioning process requires three additional ingredients:

1. a list of training inputs, \( X_t = (x_1, x_2, \ldots, x_d) \), where each input is an \( n \)-dimensional vector,

2. a list of training outputs, \( y_t = (y_1, y_2, \ldots, y_d) \), where each output is a single number, and

3. an assumed covariance function \( k(x, x') = \text{cov}(y_p(x), y_p(x')) \) which describes the similarity of the predicted emulator outputs \( y_p(x) \) and \( y_p(x') \) as a function of the emulator inputs \( x \) and \( x' \).

Note, I’ve used a subscript \( t \) to label training data and a subscript \( p \) to label prediction data following the notation in reference [140], e.g. \( y_t \) is a training point output, and \( y_p \) is an emulator prediction. This will help me distinguish between both types of variables when they appear side-by-side. I’ll stick to this convention throughout the remainder of the chapter.

Expressed as a single function call, we can now write the emulator prediction \( y_p \) at the
parameter point $\mathbf{x}$ compactly using the following notation

$$y_p = \mathcal{GP}(\mathbf{x}; X_t, \mathbf{y}_t, k(\mathbf{x}, \mathbf{x}')).$$  \hfill (4.12)

The arrays $X_t$ and $\mathbf{y}_t$ on the right-side are the training data used to condition the emulator, while $k(\mathbf{x}, \mathbf{x}')$ is the GP covariance function. The conditioning procedure that calculates $\mu$ and $\Sigma$ for use in equation (4.11) given these variables is somewhat technical, and I won’t describe it here, but a detailed discussion can be found in [140, 209]. Rather, let me now pivot to some example visualizations which should make things more clear.

For the sake of simplicity, let me temporarily restrict our attention to single-variate GPs which are easier to visualize than multivariate GPs. Suppose that you are given the training outputs $\mathbf{y}_t = (y_1, y_2, \ldots, y_d)$ at several training inputs $\mathbf{x}_t = (x_1, x_2, \ldots, x_d)$. Moreover, suppose that you are also given the simple covariance function

$$k(x, x') = \sigma_f^2 \exp \left( -\frac{|x - x'|^2}{2\ell^2} \right) + \sigma_n^2 \delta(x - x'),$$  \hfill (4.13)

for the similarity of $y_p(x)$ and $y_p(x')$ as a function of $x$ and $x'$. The right-side of this equation consists of two separate terms. The term on the left is a squared-exponential kernel with autocovariance $\sigma_f^2$ and correlation length $\ell$, and the term on the right is an uncorrelated white noise kernel with variance $\sigma_n^2$. The squared exponential term asserts that the function is smoothly varying, and the white noise term allows for some additional emulator wiggle room at each training point. This general two-component form is a common choice for real world applications, and I’ll use a similar variant of it later in the text.

Figure 4.4 shows several GP realizations conditioned to various datasets (black symbols) using the covariance function (4.13). The emulators are trained using a GP module bundled with scikit-learn, a Python machine learning library [210]. The emulator mean predictions are shown as solid blue lines and their $\pm 2\sigma$ uncertainties as shaded blue bands. The software automatically estimates maximum likelihood values for the covariance hyperparameters $\sigma_f$, $\ell$, and $\sigma_n$ which are annotated along the top of each figure. For more
Figure 4.4  GPs trained on several different datasets (black symbols) using the covariance function (4.13). The dark blue line is the emulator mean prediction, and the shaded band is ±2σ uncertainty. The maximum likelihood covariance hyperparameters $\sigma_f$, $\ell$, and $\sigma_n$ are listed above each figure.

information on the maximum likelihood hyperparameter calibration, see chapter 4 section 3.3 of reference [140].

Notice how the covariance hyperparameters $\sigma_f$, $\ell$, and $\sigma_n$ affect the behavior of each GP. For example, the left panel shows training data that’s noisy, but slowly varying over the full range of the plot. The preferred correlation length $\ell$ and noise term $\sigma_n$ are hence large. Conversely, in the middle figure, the data points bend and twist rapidly over the plot range with very little apparent randomness. The correlation length, $\ell$ and noise term $\sigma_n$ are therefore small. The figure on the right, meanwhile, shows an intermediate example combining features of the first two examples.

I want to pause for a moment to explain how impressive this is. I did not tell the GP what kind of function it was emulating. Nor did I specify its covariance hyperparameters, e.g. the correlation length $\ell$ or statistical noise level $\sigma_n$. I merely provided the form of the GP covariance function and conservative bounds for each hyperparameter. The GP module was able to infer the optimal hyperparameters using its own internal maximum likelihood optimization routine. Remarkably, it determined a sensible mean prediction and uncertainty band for three dramatically different datasets. This is what people mean when they refer to GPs as flexible non-parametric functions. Such flexibility is naturally vital if the framework is to generalize.
4.3.2 Procedure for multiple correlated outputs

GP emulators are scalar-valued functions. Our toy-model in equation (4.2), however, is vector-valued. It has three input parameters \( x = (x_1, x_2, x_3) \) and ten outputs \( y = (y_1, y_2, \ldots, y_{10}) \). This section describes a standard procedure which is used to apply GP emulators to vector-valued functions.

Consider, for the moment, the following vector-valued function

\[
\begin{pmatrix}
y_1 \\
y_2
\end{pmatrix} = \begin{pmatrix}
3x_1 \\
2x_2
\end{pmatrix}.
\tag{4.14}
\]

This function consists of two independent equations, \( y_1 = 3x_1 \) and \( y_2 = 2x_2 \). These equations are decoupled, so we are free to train two independent GP emulators, one for each scalar-valued equation. Now consider the slightly more complicated function

\[
\begin{pmatrix}
y_1 \\
y_2
\end{pmatrix} = \begin{pmatrix}
x_1 + x_2 \\
x_1 - x_2
\end{pmatrix}.
\tag{4.15}
\]

This example consists of two coupled equations \( y_1 = x_1 + x_2 \) and \( y_2 = x_1 - x_2 \). Both equations depend on the parameters \( x_1 \) and \( x_2 \) so they cannot be emulated independently.

If we want to model this function with two independent emulators, we need to first transform the function into a new basis representation where the resulting equations are decoupled. Let’s define two orthogonal variables, \( z_1 = (y_1 + y_2)/2 \) and \( z_2 = (y_1 - y_2)/2 \) expressed as functions of \( y_1 \) and \( y_2 \). This transformation describes a linear operation

\[
\begin{pmatrix}
z_1 \\
z_2
\end{pmatrix} = \begin{pmatrix}
1/2 & 1/2 \\
1/2 & -1/2
\end{pmatrix} \begin{pmatrix}
y_1 \\
y_2
\end{pmatrix},
\tag{4.16}
\]

on the vector \( y = (y_1, y_2) \). In the transformed basis representation, equation (4.15) becomes

\[
\begin{pmatrix}
z_1 \\
z_2
\end{pmatrix} = \begin{pmatrix}
x_1 \\
x_2
\end{pmatrix}.
\tag{4.17}
\]
The variables $z_1$ and $z_2$ are now decoupled, i.e. they vary independently as functions of $x_1$ and $x_2$. This allows us to train two separate GPs

\[
\begin{pmatrix}
z_{p1} \\
z_{p2}
\end{pmatrix} = \begin{pmatrix}
GP_1(x_1, x_2) \\
GP_2(x_1, x_2)
\end{pmatrix},
\tag{4.18}
\]

to emulate the function inputs and outputs. Recall that the subscript $p$ here means that the variable is an emulator prediction.

Suppose now that we want emulate the vector $y = (y_1, y_2)$. We simply emulate $z_1$ and $z_2$ using equation (4.18) then transform the emulated variables $z_{p1}$ and $z_{p2}$ back to the original basis

\[
\begin{pmatrix}
y_{p1} \\
y_{p2}
\end{pmatrix} = \begin{pmatrix}
1 & 1 \\
1 & -1
\end{pmatrix} \begin{pmatrix}
z_{p1} \\
z_{p2}
\end{pmatrix},
\tag{4.19}
\]

When expanded, this yields

\[
y_{p1} = GP_1(x_1, x_2) + GP_2(x_1, x_2),
\tag{4.20}
\]
\[
y_{p2} = GP_1(x_1, x_2) - GP_2(x_1, x_2).
\tag{4.21}
\]

Needless to say, realistic vector-valued functions are far more complicated than this simple example. When a function has many correlated outputs, we need to use a more sophisticated procedure to decouple the output variables.

### 4.3.2.1 Principal component analysis

Before proceeding, let me briefly define some additional notation which will be important for the discussion that follows. Throughout this dissertation I’ve used regular type lowercase letters to denote scalars, bolded lowercase letters to denote vectors, and regular type capital letters to denote matrices. This means that $y$ is a scalar, $\mathbf{y}$ is a vector, and $Y$ is a matrix. When two objects share the same letter, e.g. $\mathbf{y}$ and $Y$, it’s been assumed that they are related, i.e. $\mathbf{y}$ is some vector from $Y$. 
In this section, I’ll also need to distinguish between row vectors and column vectors. If not otherwise specified, it should be assumed that \( i \) is a row index and \( j \) is a column index. Moreover, indices that are not specified are assumed to be arbitrary. For example, given some matrix \( Y \):

- \( y_i \) is a row vector,
- \( y_j \) is a column vector,
- \( y_{ij} \) is the element in the \( i \)th row of some column \( j \),
- \( y_{ij} \) is the element in the \( j \)th column of some row \( i \), and
- \( y_{ij} \) is the element in row \( i \) and column \( j \).

With this comment about notation out of the way, let’s return to the topic at hand, functions with many correlated output variables.

Bayesian parameter estimation commonly uses a statistical procedure known as principal component analysis (PCA) to decouple correlated model outputs. Suppose that you are given a \( d \times m \) matrix \( Y \). Let the variables \( y_j \) and \( y_{j'} \) denote the matrix elements in columns \( j \) and \( j' \) of two random row vectors \( y_i \) and \( y_{i'} \). Now assume that these variables are non-trivially correlated:

\[
\text{corr}(y_j, y_{j'}) \neq 0, \quad \text{for } j \neq j'.
\]  

(4.22)

The PCA transformation changes the basis of \( Y \) to produce a new \( d \times m \) matrix \( Z \). It sends each row vector \( y_i = (y_{i1}, y_{i2}, \ldots, y_{im}) \) to a new row vector \( z_i = (z_{i1}, z_{i2}, \ldots, z_{id}) \), where each element \( z_{ij} \) is a linear combination of the elements of \( y_i \). The transformation is constructed such that it removes all linear correlations between the different columns of \( Z \). For example, if \( z_j \) and \( z_{j'} \) are the values in columns \( j \) and \( j' \) of two random row vectors, \( z_i \) and \( z_{i'} \), then

\[
\text{corr}(z_j, z_{j'}) = 0, \quad \text{for } j \neq j'.
\]  

(4.23)

This linear transformation is described by an \( m \times m \) matrix \( V \) which multiplies \( Y \) from
the right,

$$Z = YV.$$  \hspace{1cm} (4.24)

The coefficients \((z_1, z_2, \ldots, z_d)\) of each row vector \(z\) are called the principal components (PCs) of the transformation, and they are ranked in order of explained variance—that is to say that each PC describes the maximal variance possible subject to the constraint that its basis vector remains orthogonal to the basis vectors of the preceding components. The first variable \(z_1\) is called the first principal component, the second variable \(z_2\) the second principal component, and so on.

The transformation matrix \(V\) is calculated using the singular value decomposition (SVD) of the matrix \(Y\). Since we are only concerned with real-valued output, let’s assume that \(Y\) is real. Taking the SVD of the real matrix \(Y\) then yields

$$Y = U\Sigma V^T,$$  \hspace{1cm} (4.25)

where \(U\) is a \(d \times d\) orthogonal matrix, \(\Sigma\) is a \(d \times m\) rectangular diagonal matrix, and \(V\) is an \(m \times m\) orthogonal matrix. The diagonal entries of \(\Sigma\) are called the singular values of \(Y\), while the columns of \(U\) and the columns of \(V\) are called the left and right singular vectors.

Multiplying both sides of this equation by \(V\), we see that

$$YV = U\Sigma,$$  \hspace{1cm} (4.26)

which performs the desired PCA decomposition. The columns of the matrix \(Z = U\Sigma\) are uncorrelated and sorted in order of decreasing variance. The matrix \(V\) of right singular vectors is hence the desired PCA transformation matrix.

This is all very nice, but it’s not particularly illuminating. In order to see why PCA is useful, let’s apply the transformation to a simple dataset in order to visualize its effect.
Consider, for this purpose, the bivariate normal distribution

\[ P(y) = \frac{1}{2\pi \sqrt{|\Sigma_y|}} \exp \left( -\frac{1}{2} (y - \mu)^T \Sigma_y^{-1} (y - \mu) \right), \tag{4.27} \]

of a random two-component vector \( y = (y_1, y_2) \). This distribution is parametrized by a mean vector \( \mu \) and a covariance matrix

\[ \Sigma_y = \begin{pmatrix} \text{cov}(y_1, y_1) & \text{cov}(y_1, y_2) \\ \text{cov}(y_2, y_1) & \text{cov}(y_2, y_2) \end{pmatrix}. \tag{4.28} \]

If you are confused by the subscript \( y \) on the covariance matrix, its purpose will become clear in a moment. Consider now some arbitrary parameters for this distribution, namely

\[ \mu = (0, 0) \quad \text{and} \quad \Sigma_y = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}. \tag{4.29} \]

Let’s sample the distribution 1000 times and concatenate the samples into a \( 1000 \times 2 \) matrix \( Y = (y_1, y_2, \ldots, y_{1000}) \).

The left-side of figure 4.5 shows the scatter plot of these samples. Each blue symbol is a
random vector $\mathbf{y} = (y_1, y_2)$ sampled from the distribution. The covariance matrix (4.29) has non-zero off-diagonal elements, so $y_1$ and $y_2$ are linearly correlated. Let’s apply the PCA transformation now to the matrix $Y$ and plot the result. The resulting PC basis vectors, $\mathbf{z}_1$ and $\mathbf{z}_2$, are plotted as black arrows on top of the original vectors $(y_1, y_2, \ldots, y_{1000})$. The PC basis is orthogonal by construction, so $\mathbf{z}_1 \perp \mathbf{z}_2$. The first PC vector points in the direction of largest variance, while the second PC vector points in the direction of second largest variance.

Now look at the middle-panel of figure 4.5. This panel shows the result of the PCA transformation applied to the matrix $Y$. Each vector $\mathbf{y} = (y_1, y_2)$ is sent to a new vector $\mathbf{z} = (z_1, z_2)$, where linear correlations between $z_1$ and $z_2$ vanish. The PCA transformation is linear, so we can easily reexpress the covariance matrix $\Sigma_y$ in the basis of the transformed variables. It equals

$$\Sigma_z = V^T \Sigma_y V, \quad (4.30)$$

where $V$ is the PCA transformation matrix defined by equation (4.25). The PCs of different order are uncorrelated, so this matrix is diagonal. Specifically,

$$\Sigma_z = \text{diag}(\sigma_{z_1}^2, \sigma_{z_2}^2), \quad (4.31)$$

where $\sigma_{z_1}^2 = 1.8$ and $\sigma_{z_2}^2 = 0.2$ are the variances along the directions of PCs $z_1$ and $z_2$ respectively. Notice that $\sigma_{z_1}^2 \geq \sigma_{z_2}^2$ as required: the PCs of the transformation are always sorted in order of decreasing variance.

Often, when applying PCA to real world problems, it is common to whiten the PC vectors. Whitening divides each PC by its standard deviation to produce a covariance matrix that is unit diagonal. Specifically, it sends the $d \times m$ matrix of PCs

$$Z \rightarrow ZW, \quad (4.32)$$
where $W$ is an $m \times m$ whitening matrix

$$W = \text{diag}(\sigma_{z_1}^{-1}, \sigma_{z_2}^{-1}, \ldots, \sigma_{z_m}^{-1}),$$

(4.33)

and $\sigma_{z_j}$ is the standard deviation of the $j$th column of $Z$. Combining this with equation (4.24), the full PCA transformation with whitening is

$$Z = YVW.$$  

(4.34)

The right-side of figure 4.5 shows the result of the whitened PCA transformation applied to the matrix $Y$. Notice how the whitened PCs now have unit variance. Whitening is thus commonly used to standardize the input of a downstream machine learning estimator.

**Remark.** Whitening is applied internally by the *sklearn* PCA transformer. I always work with the whitened PCs in this dissertation, and it is cumbersome to keep track of the whitened and non-whitened PCs separately. Therefore, I define $Z$ to be the matrix of whitened PCs and $V$ to be the matrix which performs the PCA transformation with whitening.

### 4.3.3 Emulator calibration

Now that I’ve described how PCA can be used to decouple our model outputs, let me return to the original task at hand. I’ve digressed from this problem, so let me briefly jog your memory. We’ve been tasked with estimating the parameters of a vector-valued function $y = f_{\text{model}}(x)$ described by equation (4.2). The function is a computer model with three inputs $x = (x_1, x_2, x_3)$ and ten outputs $y = (y_1, y_2, \ldots, y_{10})$. Several experimentalists have measured the physical process $f : x \mapsto y$ modeled by the function $f_{\text{model}}$, and they’ve reported their measurement $y_e$ along with a covariance matrix $\Sigma_e$. The physical process has some true parameters $x_{\text{true}} = (0.3, 0.5, 0.7)$ which have been kept secret, and we’ve been tasked with estimating these unknown parameters using the experimental data.

In section 4.2 we evaluated the computer model at $d = 100$ parameter points $X =$
(x_1, x_2, \ldots, x_d) selected to fill the space of reasonable parameter combinations. This required significant computing resources (the model is slow), so we’ve decided to emulate the discrete mapping f_{model} : X \mapsto Y. Once this emulator is trained, we’ll be able to rapidly predict the model output y = f_{model}(x) at any point x in the parameter space. This will enable a thorough exploration of the three-dimensional parameter space using Bayesian methods. I’ll now describe how this emulator is assembled and calibrated using the existing training data.

First, let’s take our d \times m matrix Y of simulated model outputs, whose d = 100 rows correspond to design points and d = 10 columns to model observables, and center the data by subtracting off the mean of each column

\[ y_j \rightarrow y_j - \langle y_j \rangle. \tag{4.35} \]

This is a necessary preprocessing step for PCA, and it is applied automatically by the scikit-learn PCA transformer [210]. Next, let’s divide out the standard deviation \( \sigma_{y_j} \) of each observable column \( y_j \)

\[ y_j - \langle y_j \rangle \rightarrow \frac{y_j - \langle y_j \rangle}{\sigma_{y_j}}, \tag{4.36} \]

so that every column has unit variance. This is another common machine learning procedure that places every observable (column) on the same general scale. Without it, I would be able to change the form of the PCA decomposition simply by changing the units of each observable which is clearly undesirable. Now, let’s define \( \hat{Y} \) to be the matrix of scaled observables

\[ \hat{Y} = (Y - \bar{Y})S^{-1}, \tag{4.37} \]

constructed by applying the aforementioned transformation to each of the columns of \( Y \). Here I use the d \times m matrix \( \hat{Y}_{ij} = \langle y_j \rangle \) to recenter the observables, and the m \times m diagonal matrix

\[ S = \text{diag}(\sigma_{y_1}, \sigma_{y_2}, \ldots, \sigma_{y_m}), \tag{4.38} \]
to divide out their standard deviations $\sigma_{y_j}$ and rescale them.

The next step is to use PCA to transform the $d \times m$ matrix $\hat{Y}$ into a new $d \times m$ matrix $Z$ of whitened PCs

$$Z = \hat{Y}V. \quad (4.39)$$

The first column $z_1$ of $Z$ is the first PC at every design point, the second column $z_2$ is the second PC at every design point, and so on. Each PC column vector $z_j$ is linearly uncorrelated with the other PC column vectors $z_{j'} \neq z_j$, so we can train a single GP to emulate each column separately. Collectively, this forms a probabilistic vector-valued emulator:

$$z_p \sim \left( \begin{array}{c} z_{p1} \\
                z_{p2} \\
                \vdots \\
                z_{p10} \end{array} \right) = \left( \begin{array}{c} GP_1(x; X_t, z_{t1}, k(x, x')) \\
                GP_2(x; X_t, z_{t2}, k(x, x')) \\
                \vdots \\
                GP_{10}(x; X_t, z_{t10}, k(x, x')) \end{array} \right), \quad (4.40)$$

for the PC vector $z_p = (z_{p1}, z_{p2}, \ldots, z_{p10})$ as a function of the model input parameters $x = (x_1, x_2, x_3)$. Here $X_t$ is the $d \times m$ matrix of training inputs, $z_{tj}$ is the $d \times 1$ vector of training outputs of the $j$th PC, and $k(x, x')$ is the GP covariance function. Note, technically $z_p$ is an $1 \times 10$ row vector here. I’ve simply written $z_p$ as a column vector for notational convenience.

In principle, we could train ten independent GP emulators—one for each PC—but ten PCs would be overkill for the problem at hand. The first two PCs describe 91.728% and 8.008% of the model’s total output variance respectively, while PCs 3–10 describe the remaining 0.264%. Emulating the first two PCs therefore reproduces essentially all of the model variance. While the first two PCs include meaningful variance, e.g. variance that arises from sensitivity to the model input parameters, PCs 3–10 include meaningless variance, i.e. they are essentially random statistical noise. This noise results from the statistical noise of the model, introduced by equation (4.3). PCA is an information preserving transformation so this statistical noise has to go somewhere. PCs 3–10 are essentially a projection of this noise onto random orthogonal vectors which complete the basis of the
linear transformation.

GP implementations will occasionally fail to determine the correct hyperparameters when fitting noisy data, so it is generally safer to model the especially noisy PCs using the sample mean and sample variance of each PC over the entire design. Hence, if a PC is pure noise, it samples a normal distribution $\mathcal{N}(\mu, \sigma^2)$ with mean $\mu = 0$ and variance $\sigma^2 = 1$, the mean and variance of each PC after whitening. This is a conservative approximation that over predicts the emulator and model uncertainties. Equation (4.40) may then be written as

$$z_p \sim \begin{pmatrix} z_{p1} \\ z_{p2} \\ z_{p3} \\ \vdots \\ z_{p10} \end{pmatrix} = \begin{pmatrix} \mathcal{GP}_1(x; X_t, z_{t1}, k(x, x')) \\ \mathcal{GP}_2(x; X_t, z_{t2}, k(x, x')) \\ \mathcal{N}(0, 1) \\ \vdots \\ \mathcal{N}(0, 1) \end{pmatrix}, \quad (4.41)$$

using $\mathcal{N}(0, 1)$ to replace each noise-dominated GP emulator prediction.

Suppose now that I want to sample the emulator prediction $y_p$ at some parameter point $x$. The first step is to sample the emulated PC vector $z_p(x)$ defined by equation (4.41). I can then revert the PC transformation (4.39) and the scaling and centering transformations (4.37) in sequence yielding,

$$y_p = z_p V^T S + \tilde{y}, \quad (4.42)$$

where $\tilde{y}$ is now a vector since we are recentering a single model output.

Similarly, if I want to sample the predicted emulator output at $k$ parameter points, I can sample $z_p$ at those points and concatenate the samples into a $k \times 10$ matrix $Z_p$, where each row of $Z_p$ is the PC vector at a single point. I can then transform all of the PC vectors in one pass,

$$Y_p = Z_p V^T S + \tilde{Y}. \quad (4.43)$$

Moreover, I can also use this procedure to directly calculate the mean emulator output
\( \langle y_p \rangle \) at a certain parameter point \( x \). The mean vector in PC space is

\[
\langle z_p(x) \rangle = (\mu_1(x), \mu_2(x), 0, \ldots, 0),
\]

where \( \mu_1(x) \) and \( \mu_2(x) \) are the means predicted by GP1 and GP2 respectively. The other components are zero since they sample a normal distribution with zero mean and unit variance. Equation (4.42) can then be used to transform the mean vector \( \langle z_p(x) \rangle \) into the mean vector \( \langle y_p(x) \rangle \), equal to the average emulator prediction at the point \( x \).

### 4.3.4 Emulator uncertainty

Additionally, we can also transform the emulator uncertainty on \( z_p(x) \) to calculate the emulator uncertainty on \( y_p(x) \). The variance of the PC vector \( z_p(x) \) is given by

\[
\text{var}(z_p(x)) = (\sigma_{z_1}^2(x), \sigma_{z_2}^2(x), 1, \ldots, 1),
\]

where \( \sigma_{z_1}^2(x) \) and \( \sigma_{z_2}^2(x) \) are the predictive variances of PCs \( z_{p1} \) and \( z_{p2} \) respectively at the point \( x \). Here I’ve dropped the subscript \( p \) from the right-side of equation (4.45) because I’m running out of room. The other PCs have zero mean and unit variance across the entire design, so their variances are set to one. The individual components of \( z_p \) are all linearly uncorrelated, so the resulting covariance matrix is diagonal:

\[
\Sigma_z(x) = \text{diag}[(\sigma_{z_1}^2(x), \sigma_{z_2}^2(x), 1, \ldots, 1)].
\]

Again, \( \sigma_{z_1}^2(x) \) is the predictive variance that’s returned by the GP emulator at a single parameter point, not to be confused with \( \sigma_{z_i}^2 \), the variance of that PC over all design points. The former is generally smaller than unity, while the latter is unity by construction; recall that each PC is whitened to unit variance.

The PCA transformation is linear, so the covariance matrix (4.46) is easily reexpressed in the original basis of the model observables \( y \). Performing this change of coordinates...
Figure 4.6  The reference figure on the left shows a GP (blue line and error band) conditioned on some training points (black symbols) sampled from a function (red line) with added statistical noise. The setup in the middle figure is identical to that of the reference figure, but with larger statistical noise added to the training data. The figure on the right, meanwhile, is also identical to the reference figure, but this time with a large chunk of training points removed from the middle.

yields

\[ \Sigma_y(x) = Q \Sigma_z(x) Q^T, \]  

(4.47)

where \( Q = S^{-1}V \) is the pair of transformations performed by the scaling matrix \( S \) (4.37) and the whitened PCA transformation \( V \) (4.39). The centering matrix \( \bar{Y} \) does not contribute here since the overall mean has no effect on the covariance matrix. Note, while \( \Sigma_z \) is diagonal, the matrix \( \Sigma_y \) is generally non-diagonal. Applying the inverse PCA transformation reintroduces all the observable correlations that were originally removed by the forward PCA transformation.

So where does the emulator uncertainty come from? Generally speaking, the emulator is only as good as its training data. If the training data is noisy, or there are too few parameter points, then the emulator uncertainty will be large. To get a better feel for why this happens, let’s look at a specific example using a GP emulator in one dimension. The left panel of figure 4.6 shows some training data (black points) sampled from a function (red dashed line) with added statistical noise. A GP is then trained to emulate the black points using the covariance function from equation (4.13). The blue line is the GP’s mean prediction, and the blue band is the region containing \( \pm 2\sigma \) uncertainty.

Using this training data and GP as a reference point, let’s increase the statistical noise of the training data. The middle panel of figure 4.6 shows the qualitative effect of this
change. The emulator uncertainty clearly grows to accommodate the larger variance of the training data. In a similar fashion, we can also investigate what happens when the emulator is forced to interpolate over a larger distance. The right panel of figure 4.6 shows the effect of removing a large chunk of training points from the middle of the dataset to create a gap. This causes the GP uncertainty to swell in the middle of the gap where its predictions are furthest from the neighboring training points.

The overall emulator uncertainty is therefore reduced by either running the computer model more times to reduce statistical noise in the model output at each design point, or by using more design points to produce a denser scaffolding of the parameter space. Both require more computing resources, so there is generally a trade off to be made in optimizing each variable separately. I do not analyze these trade offs in the present dissertation, but it would nevertheless an interesting topic for future study.

4.3.5 Emulator validation

The emulator will serve as a stand-in for the physics model so it is imperative to check that it is working correctly. The emulator is validated by checking its predictions against new inputs and outputs that were excluded from the calibration process. In the parlance of machine learning, we say that this validation data is out of sample. Generally speaking, the emulator is working properly if it makes statistically accurate statements. This is somewhat loaded terminology, so let me elaborate with a concrete example.

Imagine that you have a model which generates probabilistic predictions for the value of some variable $y$. Suppose, for example, that this variable is the point-total of two teams in a basketball game. Moreover, assume that each prediction is a normal random variable $\mathcal{N}(\mu, \sigma)$, parametrized by some mean value $\mu$ and standard deviation $\sigma$.

Now imagine that you run this basketball model to predict the outcome of $k$ different games which are each part of some future tournament. These predictions form a list of $k$
normal distributions

\[
predictions = [\mathcal{N}(\mu_1, \sigma_1), \mathcal{N}(\mu_2, \sigma_2), \ldots, \mathcal{N}(\mu_k, \sigma_k)], \tag{4.48}
\]

where \(\mathcal{N}(\mu_i, \sigma_i)\) is the probabilistic prediction of the \(i\)th game. After the tournament ends, you record the point-total of each game and construct a list of corresponding observations

\[
observations = [y_1, y_2, \ldots, y_k], \tag{4.49}
\]

where each observation is the outcome of a single prediction.

If the basketball point-total prediction model is truthful, then each outcome \(y_i\) is a “sample” of its predicted distribution \(\mathcal{N}(\mu_i, \sigma_i)\). Moreover, if \(y_i\) is a sample of \(\mathcal{N}(\mu_i, \sigma_i)\), then the transformed variable

\[
z_i = \frac{y_i - \mu_i}{\sigma_i}, \tag{4.50}
\]

is normally distributed with zero mean and unit variance. In the literature, this variable is commonly referred to as a \(z\)-score.

Hence, if I calculate the \(z\)-score of every model prediction, the \(z\)-scores should populate a normal distribution with zero mean and unit variance. I can easily check this property by calculating the \(z\)-scores and histogramming their distribution. If the \(z\)-scores deviate strongly from a normal distribution, then the predictions are statistically inaccurate. Note, statistical accuracy has a very specific meaning. I can make a prediction that claims to know nothing at all, but with very large error bars, and have it be statistically accurate. In other words, it’s ok if the emulator is imperfect so long as it returns a reasonable estimate of its own uncertainty.

Returning to the example that’s the subject of this chapter, let’s apply the aforementioned \(z\)-score test to the emulator constructed in the previous section. Unfortunately, there’s no simple analogue for a \(z\)-score test in multiple dimensions, so I’ll have to validate each model output \((y_1, y_2, \ldots, y_{10})\) separately. This is somewhat tedious and repetitive, so let me demonstrate the procedure for just one observable. For no particular reason, let’s
look at the fifth emulated model output $y_5$.

Figure 4.7 shows two complementary tests of the emulator prediction accuracy. The panel on the left shows a scatter plot of the model and emulator predictions at 100 random validation points $x_1, x_2, x_3 \in [0, 1]$. Each blue symbol is the model’s observed value plotted against the emulator’s predicted value calculated using the same input parameters. The emulator predictions are close to the true model values, so each point falls close to the thin gray line which indicates perfect agreement. The horizontal error bars on the predicted values of $y_5$ are $\pm 2\sigma$ emulator uncertainties, so we’d expect $\sim 95\%$ of these error bars to touch the gray line. This is of course difficult to eye ball, so let’s plot the distribution of $z$-scores which serves as a more sensitive test.

The right panel of figure 4.7 shows the distribution of $z$-scores (blue histogram) for observable $y_5$ calculated from the emulator and model predictions. Here, I’ve used significantly more validations points, $x \in [0, 1]^3$, to produce a smooth distribution. This histogram should be compared to the orange curve which is a normal distribution with zero mean and unit variance. The agreement here is excellent. Evidently, the emulator uncertainties are correct estimated. This means that we will be able to properly account for this uncertainty
when using Bayesian parameter estimation to calculate the unknown model parameters.

4.4 Applying Bayes’ theorem

This section describes how Bayesian statistics is used to solve the inverse problem, enabling statistically accurate statements to be made about the three unknown model parameters \((x_1,x_2,x_3)\) which are the subject of this chapter. The following explanation of Bayesian parameter estimation is adapted from one of my publications J. S. Moreland, J. E. Bernhard, and S. A. Bass, “Bayesian calibration of a hybrid nuclear collision model using p-Pb and Pb-Pb data from the LHC,” Submitted for publication, arXiv:1808.02106 [nucl-th],

which I’ve edited for both content and clarity.

4.4.1 Bayesian posterior

Ultimately, the goal of this chapter is to estimate the true parameters of our model, \(x_{\text{true}}\), provided some evidence that the model predictions describe a set of experimental measurements \(y_e\). The problem involves three distinct components:

1. \(H_f\): the hypothesis that the assumed model \(y = f(x)\) provides a realistic description of the physical process that it simulates,

2. \(H_x\): the hypothesis that \(x\) are the true parameters \(x_{\text{true}}\) of the physical process modeled by the function \(f\), and

3. \(E\): the evidence provided by the model, the experimental data, and their associated uncertainties.

As a practical matter, one always asserts the truth of hypothesis \(H_f\). This means that there are no glaring flaws in the chosen theoretical model framework, i.e. there exist some model parameters \(x\) where the model provides a sensible description of reality. This is a significant
assumption, and all of our results are predicated on it. If the model is completely bogus, the constraints on the parameters are meaningless as well.

Subject to this assumption, we can apply Bayes’ theorem to evaluate hypothesis \( H_x \) given the evidence provided by \( E \). Simplifying my notation and writing \( H_x \) as just \( x \), Bayes’ theorem yields

\[
P(x|E) \propto P(E|x) P(x). \tag{4.51}
\]

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

### 4.4.1.1 Likelihood function

Let’s assume that the likelihood function \( P(E|x) \) in equation (4.51) is described by a multivariate normal distribution:

\[
P(E|x) = \frac{1}{\sqrt{(2\pi)^m \det \Sigma}} \exp \left[ -\frac{1}{2} \Delta y(x)^T \Sigma^{-1}(x) \Delta y(x) \right], \tag{4.52}
\]

where \( \Delta y = y_m(x) - y_e \) is a vector of length \( m \) equal to the discrepancy of the model and experiment, and

\[
\Sigma = \Sigma_m(x) + \Sigma_e, \tag{4.53}
\]

is a total covariance matrix, equal to the sum of a modeling component \( \Sigma_m(x) \) and an experimental component \( \Sigma_e \) which account for all known sources of uncertainty in the simulated and measured observables.

**Remark.** When two uncertainties are uncorrelated their variances add. The same general rule also applies when combining two covariance matrices.

Our model is agonizingly slow to run directly, so we’ve trained an emulator to replace the model calculation \( y_m(x) \) with an emulator prediction \( y_m^{\text{emu}}(x) \). The covariance matrix
\( \Sigma_m(x) \) is therefore the covariance matrix of our emulator, \( \Sigma_{\text{emu}}(x) \), which subsumes all quantifiable sources of model and emulator uncertainty. Namely, it accounts for the interpolation uncertainty introduced by using a finite number of emulator training points as well as the statistical uncertainty introduced by the noise in the simulation outputs.

To complete the specification of equation (4.53), I’ll also need to specify an experimental covariance matrix \( \Sigma_e \) for use in our example problem. Generally speaking, the experimental covariance consists of separate statistical and systematic contributions,

\[
\Sigma_e = \Sigma_{\text{stat}} + \Sigma_{\text{sys}}.
\] (4.54)

The statistical errors in \( \Sigma_{\text{stat}} \) are uncorrelated, so its covariance matrix is diagonal:

\[
\Sigma_{\text{stat}} = \text{diag}[(\sigma_{y_1}^{\text{stat}})^2, (\sigma_{y_2}^{\text{stat}})^2, \ldots, (\sigma_{y_m}^{\text{stat}})^2].
\] (4.55)

Here \( \sigma_{y_j}^{\text{stat}} \) is the statistical uncertainty of observable \( y_j \) in the experimental observable vector \( y_e = (y_1, y_2, \ldots, y_m) \). For the purpose of our example, let’s assume that \( \sigma_{y_j}^{\text{stat}} = 0.03 \) for every observable component.

The systematic errors in \( \Sigma_{\text{sys}} \), meanwhile, are correlated so the matrix is generally non-diagonal. Given an arbitrary covariance, we can decompose it into the form

\[
\text{cov}(y_i, y_j) = \rho_{ij} \sigma_i \sigma_j,
\] (4.56)

where \( \sigma_i \) and \( \sigma_j \) are the standard deviations of \( y_i \) and \( y_j \) respectively, and \( \rho_{ij} \) is their Pearson correlation coefficient:

\[
\rho_{ij} = \frac{\text{cov}(y_i, y_j)}{\sigma_i \sigma_j},
\] (4.57)

satisfying \( \rho_{ij} = 1 \) for \( i = j \) and \( |\rho_{ij}| \leq 1 \) for \( i \neq j \). I’ll use this decomposition now to define a systematic covariance matrix for our example problem.

Let’s assume that the systematic standard deviation \( \sigma_{y_i}^{\text{sys}} = 0.03 \) for every component \( y_i \). Let’s also assume that each pair of observables, \( y_i \) and \( y_j \), has a systematic error correlation...
Figure 4.8 Correlation matrix, $\text{corr}(y_i, y_j) = \text{cov}(y_i, y_j)/(\sigma_i\sigma_j)$, visualized for the model (left-side) and the experimental data (right-side). Each colored cell is the correlation coefficient of a pair of observables $y_i$ and $y_j$. Cooler colors indicate positive correlations and warmer colors indicate negative correlations. Observables are trivially correlated with themselves, so the diagonal is unity.

Coefficient

$$\rho_{ij}^{\text{sys}} = \exp \left[ -\frac{1}{2} \left( \frac{v_i - v_j}{l} \right)^2 \right], \quad (4.58)$$

where $v = (0.05, 0.15, \ldots, 0.95)$ is a ten component vector that evenly partitions the interval $[0, 1]$, and $l = 0.5$ is a fixed correlation length. This functional form strongly correlates two observables $y_i$ and $y_j$ if they occupy proximate elements of the vector $y_e$. For example, the observables $y_1$ and $y_2$ are strongly correlated, while $y_1$ and $y_{10}$ are not. The correlation length $l$ controls the extent of the correlation. As $l \to \infty$, the systematic errors become perfectly correlated, and as $l \to 0$ the systematic errors become perfectly uncorrelated.

Figure 4.8 compares the correlation structure of the model emulator (left-side) to that of the experimental data (right-side). Each colored cell is the correlation coefficient of a pair of observables:

$$\text{corr}(y_i, y_j) = \frac{\text{cov}(y_i, y_j)}{\sigma_i\sigma_j}, \quad (4.59)$$

defined as the covariance of $y_i$ and $y_j$ divided by $\sigma_i$ and $\sigma_j$, the standard deviation of each observable. Note, the correlation matrix is shown here instead of the covariance matrix because the correlation coefficient is bounded between -1 and 1. This makes it significantly easier to visualize the matrix elements using a single heatmap.
Looking at the left-side of figure 4.8, we see that the emulator’s correlation matrix is essentially diagonal. This makes sense. When I constructed the model, I added some additional statistical noise to the model outputs. This statistical noise is uncorrelated, so it only appears in the diagonal elements of the correlation matrix. Every model output is perfectly correlated with itself, so the diagonal elements are unity by construction. Presumably the emulator also includes some interpolation uncertainty which is likely correlated, but evidently this interpolation uncertainty is small compared the model’s overall statistical uncertainty. Generally speaking, this will not always be the case.

Compare this emulator correlation matrix to the experimental correlation matrix depicted on the right-side of figure 4.8. The experimental correlation matrix includes contributions from both a statistical component and a systematic component. The statistical errors are naturally uncorrelated so they only contribute to the matrix diagonal. The systematic errors, meanwhile, are correlated, so they introduce nonzero off-diagonal elements in the correlation matrix, producing the soft blue gradient that is visible in the figure. This systematic correlation uncertainty was modeled using a finite correlation length, so the correlation fades as the element-wise distance between the components increases, i.e. \( y_1 \) and \( y_2 \) are more strongly correlated than \( y_1 \) and \( y_{10} \).

### 4.4.1.2 Specifying a prior

Now that I’ve specified the likelihood function \( P(E|x) \), I need to specify the prior \( P(x) \) to complete the right-side of equation (4.51). The Bayesian prior describes the initial uncertainty on the model parameters \( x \), absent the evidence provided by the likelihood function. Specifically, it is a probability distribution of the model inputs parameters. Regions of parameter space where the prior probability density is large indicate likely regions for the true parameters \( x_{\text{true}} \). Conversely, if the prior probability density is small, then those regions of parameter space are unlikely to contain the true model parameters \( x_{\text{true}} \).

Prior specification is a hotly contested topic in Bayesian statistics. If a prior is too restrictive, i.e. specific, it can bias the posterior in harmful ways. For example, suppose I
choose a prior that’s a Dirac delta function

\[ P(x) = \delta(x - x'), \]  

where \( x' \) is some specific set of parameter values. The Bayesian posterior is proportional to the product of the likelihood function and the prior, so the resulting posterior \( P(x|E) \) will be zero for all values of \( x \neq x' \). This essentially forces the posterior to equal the prior. In other words, I learn nothing from the likelihood function at all because I’ve purported to know everything there is to know with my choice of prior.

Alternatively, there are times when it makes sense to use a restrictive prior. For example, if it’s known that a parameter cannot be negative, then the prior distribution should be zero for all negative parameter values. This asserts that there is 0% chance the true parameter value \( x_{\text{true}} \) is negative.

In the present example, we’ve been told that each parameter \( (x_1, x_2, x_3) \) lies inside a unit cube, \([0, 1]^3\) so the relevant prior distribution is

\[ P(x_1, x_2, x_3) = \begin{cases} 1 & \text{if } 0 \leq x_i \leq 1 \text{ for all } x_i, \\ 0 & \text{otherwise}. \end{cases} \]  

This is, by no accident, the same range of values that I used to construct the design matrix \( X \). There is no need to train the emulator outside this region, since the prior is zero there anyhow. If the prior is zero, the posterior also will be zero regardless of the value of the likelihood function.

### 4.4.2 Importance Sampling

Now that I’ve specified the likelihood \( P(E|x) \) and the prior \( P(x) \), I can proceed to calculate the Bayesian posterior

\[ P(x|E) \propto P(E|x)P(x). \]  

129
Written in this way, I can only calculate the Bayesian posterior up to an overall normalization factor. However, only relative probabilities will matter in the steps that follow, so I can disregard the value of this factor.

In the present example, the Bayesian posterior distribution $P(x|E)$ is three-dimensional, i.e. it has one dimension for each parameter $x_1$, $x_2$, and $x_3$. The plane of this page is two-dimensional, so I’ll ultimately have to project this distribution down onto a lower number of dimensions. One common procedure to do this is to sample the distribution a large number of times. Once the distribution is sampled, I can histogram the samples using a single parameter value $x_i$ to project the joint posterior distribution onto that dimension. In a similar fashion, I can also histogram a pair of parameters $x_i$ and $x_j$ to visualize their correlations.

Sampling a multidimensional probability distribution is a difficult task in and of itself. Bayesian posterior distributions may include as many as ten parameters or more, so accept-reject sampling is out of the question; this basic sampling algorithm fails spectacularly when the number of dimensions is large. The canonical procedure for sampling a multi-dimensional probability distribution is to use Markov chain Monte Carlo importance sampling, commonly referred to as MCMC for short.

Given a target probability distribution, the MCMC sampling procedure constructs a sequence of random values known as a chain. When this chain is sufficiently long, one can obtain a sample of the target distribution by observing the random values at the end of the chain. The simplest and most commonly used MCMC algorithm is the Metropolis-Hastings (M-H) algorithm. M-H samples the probability distribution using a series of proposed updates which are either accepted or rejected to produce the next sample in the chain. The algorithm has a number of walkers which wander the parameter space. At any given moment, each walker occupies a single parameter point $x$. The algorithm then proposes a new parameter point $x'$ which is obtained by perturbing the parameter $x$ in a random
direction. The walker then moves to the proposed parameter point \( x' \) with probability

\[
\alpha = \min \left( 1, \frac{P(x')}{P(x)} \right),
\]

referred to as the proposal acceptance fraction. If the proposal is rejected, the walker stays in its current location. If it is accepted, the walker moves to the new location. Since this update step only depends on the relative probability \( P(x')/P(x) \), it is not necessary that the target probability distribution is normalized. Every time the walker position is updated, the new position is appended to the end of the MCMC chain. The update step is then repeated many times to build up the length of the chain. Eventually, the samples at the end of the chain form a random sample of the target probability distribution.

The Bayesian parameter estimation framework [140] used in this dissertation uses emcee [211], an affine-invariant MCMC ensemble sampler implemented in Python that uses a large number of interdependent walkers [212, 213]. The algorithm is qualitatively similar to the original M-H algorithm, but it generally converges much faster to a stationary set of samples from the target distribution. This reduces the amount of “burn-in” steps which must be discarded from the beginning of the MCMC chain to ensure the same level of statistical accuracy.

### 4.5 Visualizing the posterior distribution

The final step of Bayesian parameter estimation is to visualize the multidimensional posterior distribution defined by equation (4.51). Let’s use the emcee ensemble sampler to draw \( \mathcal{O}(10^7) \) parameter samples \( \{x_i\} \) from the posterior distribution of our example problem. Using these samples, we’ll be able to visualize the constraints provided by the model, the experimental data, and all associated uncertainties.

Figure 4.9 shows various model calculations (blue lines) compared to the experimental data (black symbols with error bars). The plot on the left shows the model output \( y = f_{\text{model}}(x) \) at each of the \( d = 100 \) design points (same as figure 4.3), while the plot on...
the right shows the mean emulator output $\mathbf{y} = f_{\text{emu}}(\mathbf{x})$ at 100 different parameter points randomly selected from the Bayesian posterior.

We initially assumed that each parameter $\mathbf{x} \in [0,1]^3$, so the model outputs on the left span a wide range of values. This visual spread is therefore the uncertainty of the assumed prior (4.61). The plot on the right, meanwhile, shows the constraining power of the evidence, i.e. the refinement to the prior provided the model, the experimental data, and all uncertainties. These calculations are now tightly clustered around the experimental data, giving us confidence that the model is in fact a good representation of the ground truth. Of course, the experimental data $\mathbf{y}_e$ was constructed by running the model using a certain set of true parameters, $\mathbf{y}_e = f_{\text{model}}(\mathbf{x}_{\text{true}})$, so it should be no surprise that this is the case.

Finally, let’s address the original problem that we were tasked with solving at the beginning of this chapter. I’ll now present the posterior estimate for the model parameters $\mathbf{x}$. Figure 4.10 visualizes the three-dimensional Bayesian posterior distribution $P(\mathbf{x}|E)$. Each diagonal panel is the marginal distribution of a single model parameter constructed by integrating the posterior distribution over all other parameters. For example, the marginal distribution for parameter $x_1$ is

$$P(x_1) = \int dx_2 dx_3 P(x_1, x_2, x_3).$$  \hspace{1cm} (4.64)
Figure 4.10 Bayesian posterior distribution for the three model input parameters $x_1$, $x_2$, and $x_3$. 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. Red symbols and lines indicate the true value of each model parameter.

The off-diagonal panels, meanwhile, are the joint posterior distribution for a pair of model parameters, visualizing their correlations. They are similarly constructed by marginalizing (integrating) over all but two model parameters. Note, the probability distributions in figure 4.10 are histograms, so there’s no need to actually integrate, we simply bin the variables of interest.

The black text above each diagonal panel is the marginalized posterior distribution’s median value and 90% highest posterior density (HPD) credible interval. This latter quantity is defined as the narrowest parameter interval containing 90% of the posterior density. Roughly speaking, this means that we expect $x_{\text{true}}$ to land inside the interval 90% of the time. The estimates are presented as a median value plus-minus the distance to each edge of the credible interval. For example, the estimate $x_1 = 0.306^{+0.061}_{-0.061}$ corresponds to a median
value $\bar{x}_1 = 0.306$ and 90% credible interval $0.245 < x_1 < 0.367$. Also shown, for reference, are the true model parameters $x_1 = 0.3$, $x_2 = 0.5$, and $x_3 = 0.7$. These values are plotted as a red dot on each joint posterior distribution and as a red line on each marginal distribution.

There are several important features that should be noticed when looking at this figure. First, the posterior distribution of $x_1$ is tightly constrained, while the estimates for $x_2$ and $x_3$ span a large range of values. This result is easily understood if we look at the original function that was modeled, equation (4.2), which I’ve written below:

$$y = x_1 u + (x_2 + x_3)v.$$  \hfill (4.65)

Recall that $u = (1, 1, \ldots, 1)$ is a vector of ones, and $v = (0.05, 0.15, \ldots, 0.95)$ is a vector that uniformly partitions the interval $[0, 1]$. Inspecting this function, we see that parameter $x_1$ shifts the vector $y$ vertically up and down, while the parameters $x_2$ and $x_3$ vary its slope. The effect of parameter $x_1$ is unique, while $x_2$ and $x_3$ are degenerate; all combinations of $x_2$ and $x_3$ have the same effect if their sum is held constant.

This illustrates a very important property of the inverse problem, namely information loss. If two parameters produce the same effect, then their values cannot be constrained independently. Look now at the joint posterior distribution of parameters $x_2$ and $x_3$. This joint distribution is highly correlated: $x_2$ can be small if $x_3$ is large and vice versa, but both parameters cannot be large (or small) at the same time. This is, of course, exactly what we’d expect from looking at equation (4.65). Clearly, it is their sum, $x_2 + x_3$, which is constrained by the data.

Before I conclude this chapter, I want to make one more statement about the Bayesian posterior distribution. Evidently, we did a pretty good job determining $x_1$, but our estimates were considerably less certain about the individual values of $x_2$ and $x_3$. This is not a failure! This is merely a fact of life. Our goal here was obviously to say as much as humanly possible about each parameter, but we also wanted to do so truthfully. Sometimes the correct answer is “I don’t know” or “I can’t tell”, and that’s perfectly acceptable. In fact, in this case, it was the only correct answer.
In the next chapter, I will use the Bayesian parameter estimation framework developed in [140] to study the TRENTo initial condition model which is the focus of this dissertation. When I constructed the model, I tried to assume as little as possible. The model is intentionally flexible, constructed to interpolate a subspace of all initialization models including (but not limited to) specific calculations in Color Glass Condensate effective field theory. This flexibility is parametrized by a handful of model parameters whose values are uncertain. Bayesian parameter estimation is useful, in large part, because it allows one to marginalize over this uncertainty, i.e. the constraint on any one parameter accounts for the uncertainty of all other parameters. This enables robust statements about the initial conditions and QGP medium parameters which would not otherwise be possible.
5

Model calibration and comparison to data

Ultimately, the goal of this dissertation is to study the QGP initial conditions through the lens of the inverse problem. While most calculations seek to derive the QGP initial conditions from first-principles or approximations thereof, I want to approach the problem from the opposite direction. Namely, I want to use the experimental data and the predictions of relativistic fluid dynamics to place robust constraints on the QGP initial conditions without assuming a narrow theoretical formalism for their derivation. This constitutes what is broadly referred to as a top-down or data-driven approach.

Section 3.5 described a parametric model of the QGP initial conditions called TRENTo developed for this purpose. The model was designed to be maximally flexible, enabling future analyses to determine the correct form of the initial conditions from their resulting description of the data and not the other way around. It therefore describes a sort of meta-model that encompasses a large subspace of reasonable theoretical descriptions.

The method that I will use to study this initial condition model is hypothesis testing. First, I fix the parameters of the TRENTo model and propose a certain parametric description of the initial state. Then, I evolve the system forward in time using relativistic viscous hydrodynamics and microscopic Boltzmann transport to simulate the final state of the system as it is observed inside the detector. Finally, I evaluate my initial hypothesis by
comparing the model’s simulated output to experimental data.

The problem is challenging for several reasons. First, the initial conditions are but one part of a multi-stage simulation. In addition to the handful of parameters needed to describe the initial state, I’ll need several more to account for uncertainties in the QGP medium properties. These parameters typically correlate among each other and affect multiple observables so I cannot tune them individually by hand. And second, the model is computationally intensive. Evaluating just a single set of model parameters requires hundreds of CPU hours. Therefore, I cannot afford to brute force the problem by running the model countless times.

In chapter 4, I described a general statistical framework known as Bayesian parameter estimation designed to handle these problems. In this chapter, I present several Bayesian studies of the TRENTo initial condition model using a software package for Bayesian parameter estimation developed by fellow graduate student Jonah Bernhard [140]. I should emphasize again that I did not write the Bayesian framework. Each project is also a collaborative effort involving multiple contributors. My own specific contribution level correspondingly varies from project to project.

Generally speaking, each study applies Bayesian parameter estimation using the same sequence of steps:

1. First, we construct a dynamical model of a relativistic nuclear collision event. The model components change slightly from study to study, but the overall framework is more or less the same. We use relativistic hydrodynamics to model the hot and dense early stage of the collision and microscopic Boltzmann transport to simulate individual particle interactions once the QGP freezes out into hadrons.

2. Next, we identify a number of free parameters in the model framework which are uncertain. These parameters are the objects of interest in each study. We place a conservative prior distribution on each model parameter and run the computer model at lots of different parameter points uniformly selected from the prior distribution.
3. For each analysis, we identify a set of physical observables which we use to build evidence for (or against) the aforementioned parameter values. We calculate these observables using the model predictions at every evaluation point and use these predictions to train an emulator. Once the emulator is trained to reproduce the input-output relationship of the model at each evaluation point, it can be used to make essentially instantaneous predictions at new points in parameter space.

4. With the trained emulator in hand, we calculate an overall likelihood function for each set of model parameters that quantifies the global compatibility of the experimental data with the model predictions at that point. This likelihood function is then multiplied by the prior to furnish the Bayesian posterior probability for the model parameters up to an overall normalization constant.

5. Finally, we use MCMC to sample and visualize the Bayesian posterior distribution. This posterior is the central result of each analysis, and it includes a wealth of information about the unknown model parameters. We therefore conclude each study with a detailed discussion of the posterior distribution.

The following chapter results are divided into three sections which address different aspects of the QGP initial condition problem. In the first section, I examine the QGP initial conditions at midrapidity using boost-invariant approximations. Specifically, I describe and discuss the first “proof of concept” Bayesian analysis of the TRENTo initial condition model, calibrated to Pb-Pb collision data at $\sqrt{s_{NN}} = 2.76$ TeV. I also briefly summarize a greatly improved analysis of Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ and 5.02 TeV, and use the results of the improved analysis to generate predictions for Xe-Xe collisions at $\sqrt{s_{NN}} = 5.44$ TeV.

In the second section, I relax the boost-invariant approximation to explore the three-dimensional structure of the collision. My coauthors and I construct a flexible parametrization for the rapidity-dependence of the QGP initial conditions and use this parametrization to extend the TRENTo model beyond midrapidity. The model is then matched to a dynamical simulation consisting of three-dimensional viscous hydrodynamics and microscopic
Boltzmann transport. Using the model, we perform a Bayesian analysis of $p$-Pb and Pb-Pb collisions at the LHC and present posterior results for the three-dimensional structure of the produced plasma.

Finally in the third section, I investigate the effect of nucleon substructure on hydrodynamic bulk observables. I model this substructure using three parameters to vary the number, size, and distribution of constituent sources inside each nucleon. This parametric nucleon substructure is then added to a dynamical model consisting of TRENTo initial conditions, pre-equilibrium free streaming, boost-invariant viscous hydrodynamics, and microscopic Boltzmann transport. Free parameters of the model which describe the initial state and QGP medium are then simultaneously calibrated to fit bulk observables in $p$-Pb and Pb-Pb collisions at $\sqrt{s_{\text{NN}}} = 5.02$ TeV. Posterior estimates for the nucleon substructure parameters are obtained, and implications for hydrodynamic flow in small collision systems discussed. I’ll now describe each of these projects in detail, following the general sequence of steps outlined on the previous page.

5.1 Bulk matter at midrapidity

The following describes the first Bayesian analysis using TRENTo initial conditions. It is based off the publication


which combines my own contributions in initialization theory [168] with the Bayesian parameter estimation framework developed by coauthor Jonah Bernhard [140]. He wrote the Bayesian parameter estimation software used for this project, ran the hydrodynamic events, and performed the primary data analysis. I contributed to the design of the analysis (parameters, observables, model components, etc.), co-developed the TRENTo model used
in the analysis, and was involved in multiple aspects of the data analysis and manuscript preparation.

Prior to this publication, we had already found compelling evidence that TRENTo would be able to simultaneously and self-consistently describe hydrodynamic bulk observables. For instance, in figure 3.16, we showed that model was able to describe \( p-p \), \( p-Pb \), and \( Pb-Pb \) charged particle multiplicity distributions using a single set of model parameters. Additionally, in figure 3.18 we showed that these same parameters were also those that were needed to describe the relative magnitude of the second and third eccentricity harmonics, a quantity that was found to be strongly constrained by ALICE flow data [189]. Moreover, unlike the two-component ansatz (3.54) which predicts a knee-shaped structure in ultra-central U-U elliptic flow data 3.20, the generalized mean ansatz (3.59) predicts that \( v_2 \) should flat-line (e.g. see figure 3.20) in qualitative agreement with measurements by the STAR collaboration [191]. These bread crumbs all suggested that our parametric approach was sensible.

Nevertheless, these indicators were really just that, indicators. The only reliable way to validate the initial condition model is to run it through a full hydrodynamic simulation. The TRENTo model is parametric, so naturally we didn’t want to just check a single set of model parameters, we wanted to check every set of model parameters. Thus, we used the Bayesian parameter estimation framework [38] recently developed by Jonah Bernhard et al. to estimate the TRENTo initial condition and QGP medium parameters using the predictions of a realistic hybrid transport model [214].

5.1.1 Boost-invariant nuclear collision model

This study was conducted using on an older version of our nuclear collision model that’s missing many of the bells and whistles included in our later studies. The study’s parameter estimates are therefore outdated, superseded by the parameter estimates of our more recent work. I show them here purely for historical context. The nuclear collision model consisted of the following components.
5.1.1.1 Initial conditions

We used the TRENTo model to initialize the hydrodynamic simulations. The version of the model used at the time was largely identical to the current version described in chapter 3 with one small difference. It did not yet include the inter-nucleon minimum distance algorithm described in subsection 3.2.2 to sample correlated nucleon positions. The nucleon positions were therefore sampled from a standard uncorrelated Woods-Saxon distribution [131]. Hence the nucleon minimum distance parameter $d_{\text{min}}$ was absent from the analysis.

In this particular study, we initialized the hydrodynamic medium using the entropy variant of the generalized mean ansatz, $s_0 \propto T_R$, where $T_R$ is the reduced thickness function (3.61). The initial flow velocity $u^\mu$ was set to zero at the hydrodynamic starting time as well as all viscous correction terms, $\pi^{\mu\nu}$ and $\Pi$. We therefore assumed instant thermalization at the hydrodynamic starting time $\tau_0$ and parametrized the initial entropy density $s = dS/(d^2x_\perp \tau_0 d\eta)$ according to

$$
 s(x_\perp; \eta_s = 0, \tau_0) = \frac{\text{Norm}}{\tau_0} \times \left[ \frac{\hat{T}_A(x_\perp)^p + \hat{T}_B(x_\perp)^p}{2} \right]^{1/p},
$$

where Norm is a dimensionless normalization constant, $\tau_0 = 0.6$ fm/$c$ is the hydrodynamic starting time, and $\hat{T}_A$ and $\hat{T}_B$ are the participant thickness functions defined by equation (3.48).

5.1.1.2 Boost-invariant viscous hydrodynamics

This study also debuted an upgraded version of VISH2+1 [215], a boost-invariant viscous hydrodynamics code developed by our collaborators at Ohio State. These upgrades included support for fluctuating event-by-event initial conditions [214] and bulk viscous corrections with shear-bulk coupling [216]. The code has been extensively validated, and it reproduces semi-analytic solutions for ideal hydrodynamics to excellent precision [214].

It’s numerical implementation solves the second-order Israel-Stewart equations [217, 218] in the so-called 14-moment approximation. This formalism produces a set of relaxation-type
for the shear viscosity $\eta$ and bulk viscosity $\zeta$ which are parametrized below. All other transport coefficients were fixed using analytic results derived from the Boltzmann equation near the conformal limit [219].

The hydrodynamic equations of motion were closed using a modern EoS based on lattice calculations at zero baryon density published by the HotQCD collaboration [18]. These calculations were then blended with a hadron resonance gas EoS in the temperature interval $110 \leq T \leq 130$ TeV using a smoothstep interpolation function [220]. This matching procedure was done differently in subsequent studies [204, 221], although the effect of the difference is likely small. See figure 3 in [220] and figure 3.9 in [140] for a direct comparison of the two different methods.

The study sought to estimate both the properties of the initial conditions as well as the properties of the produced QGP medium. So we parametrized the temperature dependence of the QGP shear and bulk viscosities expressed as dimensionless ratios $\eta/s$ and $\zeta/s$ where $s$ is the entropy density. We used a piecewise linear parametrization for the shear viscosity

$$(\eta/s)(T) = \begin{cases} 
(\eta/s)_{\text{min}} + (\eta/s)_{\text{slope}}(T - T_c) & T > T_c, \\
(\eta/s)_{\text{hrg}} & T \leq T_c,
\end{cases}$$

which is constant below the temperature $T_c$ and linearly rising above it. The constant $T_c = 0.154$ GeV was then fixed using the pseudocritical transition temperature of the HotQCD EoS, motivated by studies which demonstrate that $\eta/s$ has a minimum near the QCD transition temperature [222–224]. The constants $\eta/s$ hrg, min, and slope, meanwhile, were treated as variable model input parameters to be determined by the Bayesian analysis.
Figure 5.1 Some example curves demonstrating the variability of the temperature-dependent shear viscosity parametrization (5.3) and the bulk viscosity parametrization (5.4). Lines are chosen for illustrative purposes only and do not represent all possible variability.

For the specific bulk viscosity $\zeta/s$, we used the parametrization from references [119, 225]

$$\begin{align*}
(\zeta/s)(T) &= \begin{cases} 
C_1 + \lambda_1 \exp[(x-1)/\sigma_1] & T < T_a \\
+ \lambda_2 \exp[(x-1)/\sigma_2] & T_a \leq T \leq T_b \\
A_0 + A_1 x + A_2 x^2 & T > T_b \\
C_2 + \lambda_3 \exp[-(x-1)/\sigma_3] + \lambda_4 \exp[-(x-1)/\sigma_4] & T > T_b
\end{cases}
\end{align*}$$

with $x = T/T_0$ and coefficients:

$$
C_1 = 0.03, \quad C_2 = 0.001, \\
A_0 = -13.45, \quad A_1 = 27.55, \quad A_2 = -13.77, \\
\sigma_1 = 0.0025, \quad \sigma_2 = 0.022, \quad \sigma_3 = 0.025, \quad \sigma_4 = 0.13, \\
\lambda_1 = 0.9, \quad \lambda_2 = 0.22, \quad \lambda_3 = 0.9, \quad \lambda_4 = 0.25, \\
T_0 = 0.18 \text{ GeV}, \quad T_a = 0.995T_0, \quad T_b = 1.05T_0.
$$

We fixed the peak temperature $T_0 = 180$ MeV chosen to lie somewhat above the EoS critical temperature and scaled the overall magnitude of the curve using a tunable normalization.
factor \((\zeta/s)_{\text{norm}}\). Hence for \((\zeta/s)_{\text{norm}} = 1\), our bulk viscosity parametrization reproduced
the curve used in reference [119, 225]. Figure 5.1 shows some of the temperature-dependent
specific viscosities \((\eta/s)(T)\) and \((\zeta/s)(T)\) describable using each parametrization.

5.1.1.3 Particlization and Boltzmann transport

The study employed a hybrid transport model that simulated the hot and dense phase of
the collision using viscous hydrodynamics and the cooler, more dilute regions of the fireball
using microscopic Boltzmann transport [20, 22, 23]. The two phases were separated by a
pre-specified switching temperature \(T_{\text{switch}}\). Matter above this temperature was simulated
hydrodynamically, and matter below this temperature was simulated using the microscopic
transport model. To preserve the continuity of energy and momentum at the interface
between the two regimes, the hydrodynamic medium was converted into particles as it
cooled below the \(T_{\text{switch}}\) isotherm.

Particles were sampled from this isotherm using the standard Cooper-Frye algorithm
[226]

\[
E \frac{dN_i}{d^3p} = \frac{g_i}{(2\pi)^3} \int_{\sigma} f_i(p)p^\mu d^3\sigma_\mu, \quad (5.5)
\]

where \(g_i\) and \(f_i\) are the degeneracy and distribution function of particle species \(i\), and \(d^3\sigma_\mu\)
is a volume element of the isothermal hypersurface \(\sigma\) defined by \(T_{\text{switch}}\).

Following convention, we decomposed the distribution function into an ideal part \(f_0\) and
a viscous correction \(\delta f\). We modeled the shear contribution to \(\delta f\) using the common form
[227]

\[
\delta f_{\text{shear}} = f_0(1 \pm f_0) \frac{1}{2T^2(e + P)} p^\mu p^\nu \pi_{\mu\nu}, \quad (5.6)
\]

where \(f_0\) is the ideal Bose-Einstein or Fermi-Dirac distribution, \(T\) is the fluid cell tempera-
ture, \(e\) is its energy density, \(P\) is its ideal pressure, and \(\pi_{\mu\nu}\) is its shear stress tensor in the
fluid rest frame.

The bulk viscous correction to the distribution function, meanwhile, was completely
neglected in this study. Our justification for this choice at the time was two fold. First,
our chosen bulk viscosity parametrization (5.4) predicted a small value for $\zeta/s$ at most reasonable particlization temperatures $0.135 \leq T_{\text{switch}} \leq 0.165$ GeV, so we expected this correction to be small. And second, the various algorithms used to implement bulk viscous corrections at the time predicted different behavior when either the bulk pressure $\Pi$ or the momentum $p$ are large [228, 229]. Therefore, we decided to neglect bulk viscous corrections until a better algorithm could be implemented. This choice precluded a quantitative study of the QGP bulk viscosity which we emphasized in the publication at the time [178]. Indeed, I will show somewhat later in this chapter that estimates for the temperature dependence of $\zeta/s$ are quite different if the bulk viscous corrections at freeze-out are appropriately accounted for.

These particles were then fed into the Ultrarelativistic Quantum Molecular Dynamics (UrQMD) model [230, 231] which simulates the particles microscopically until they stop interacting. UrQMD models individual particle collisions using Monte Carlo techniques to solve the Boltzmann equation

$$\frac{df_i(x,p)}{dt} = C_i(x,p),$$

(5.7)

where $f_i(x,p)$ is the distribution function of particle species $i$, and $C_i(x,p)$ is its collision kernel. The model propagates particles along classical trajectories and simulates their scatterings, resonance formations, and decays. This produces a list of final particle data, where each particle has an identification number, four-momentum, and four-position at the moment of last interaction. The particle data for each event was finally processed to calculate the model observables used in the analysis.

5.1.2 Parameter design and observables

This study estimated the joint posterior distribution for nine model parameters used in the construction of the aforementioned nuclear collision model. Four parameters varied the QGP initial conditions modeled by TRENTo:

1. the overall normalization factor for initial entropy deposition,
2. the generalized mean parameter $p$ for the scaling of initial entropy deposition as a function of nuclear thickness,

3. the gamma shape parameter $k$ controlling nucleon multiplicity fluctuations, and

4. the Gaussian nucleon width $w$ determining initial state granularity.

Meanwhile, another five parameters varied the properties of the hybrid model simulation:

5–7. three parameters ($\eta/s \text{ hrg, min, and slope}$) for the temperature dependence of the QGP shear viscosity,

8. one parameter $\zeta/s$ norm for the overall normalization and magnitude of the QGP bulk viscosity, and

9. a particlization temperature $T_{\text{switch}}$ that defined the isotherm for Cooper-Frye particle emission.

We assigned each parameter the conservative range of prior values listed in table 5.1 and sampled 300 parameter points inside the resulting nine-dimensional parameter space using a maximin Latin hypercube design. We then ran $O(10^4)$ minimum-bias Pb-Pb collisions at $\sqrt{s_{\text{NN}}} = 2.76$ TeV at each parameter point. Each nuclear collision event consisted of a single fluctuated initial condition profile and hydrodynamic simulation followed by numerous hadronic afterburner oversamples, i.e. repeated iterations of the Cooper-Frye emission and UrQMD evolution. The completed events were then partitioned into centrality classes according to their final charged-particle multiplicity at midrapidity.

Using this event data, we calculated the charged-particle yield $dN_{\text{ch}}/d\eta$, identified-particle yields $dN/dy$, identified-particle mean $p_T$, and two-particle flow cumulants $v_n\{2\}$ for $n = 2, 3,$ and 4. Each observable was calculated for a number of different centrality bins selected to match the experiment. Table 5.2 summarizes these Pb-Pb $\sqrt{s_{\text{NN}}} = 2.76$ TeV observables, including the kinematic cuts and experimental data, which were taken from the ALICE experiment [65, 108].
Table 5.1  Input parameter ranges for the nuclear collision model.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Prior range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Norm</td>
<td>Normalization factor</td>
<td>100–250</td>
</tr>
<tr>
<td>$p$</td>
<td>Entropy deposition parameter</td>
<td>−1 to +1</td>
</tr>
<tr>
<td>$k$</td>
<td>Multiplicity fluct. shape</td>
<td>0.8–2.2</td>
</tr>
<tr>
<td>$w$ [fm]</td>
<td>Gaussian nucleon width</td>
<td>0.4–1.0</td>
</tr>
<tr>
<td>$\eta/s$ hrg</td>
<td>Const. shear viscosity below $T_c$</td>
<td>0.3–1.0</td>
</tr>
<tr>
<td>$\eta/s$ min</td>
<td>Minimum value of $\eta/s$ (at $T_c$)</td>
<td>0–0.3</td>
</tr>
<tr>
<td>$\eta/s$ slope [GeV$^{-1}$]</td>
<td>Slope of $\eta/s$ above $T_c$</td>
<td>0–2</td>
</tr>
<tr>
<td>$\zeta/s$ norm</td>
<td>Prefactor for $(\zeta/s)(T)$</td>
<td>0–2</td>
</tr>
<tr>
<td>$T_{\text{switch}}$ [GeV]</td>
<td>Switching / particlization temperature</td>
<td>0.135–0.165</td>
</tr>
</tbody>
</table>

These measurements are of course a small cross section of the data that ALICE and the other experiments have to offer. We chose these specific observables as a starting point because they are statistically cheap to compute and because nuclear collision models often struggle to describe them simultaneously, e.g. see reference [38]. I’ll describe a far more grandiose study later in this chapter which used a much larger cross section of the available experimental data.

Table 5.2  Experimental data to be compared with model calculations.

<table>
<thead>
<tr>
<th>Observable</th>
<th>Particle species</th>
<th>Kinematic cuts</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yields $dN/dy$</td>
<td>$\pi^\pm$, $K^\pm$, $p\bar{p}$</td>
<td>$</td>
<td>y</td>
</tr>
<tr>
<td>Mean transverse momentum $\langle p_T \rangle$</td>
<td>$\pi^\pm$, $K^\pm$, $p\bar{p}$</td>
<td>$</td>
<td>y</td>
</tr>
<tr>
<td>Two-particle flow cumulants $v_n{2}$</td>
<td>all charged</td>
<td>$0.2 &lt; p_T &lt; 5.0$ GeV</td>
<td>[108]</td>
</tr>
</tbody>
</table>

5.1.3 Bayesian parameter estimation

Bayesian parameter estimation was then used to estimate the values of the nine model parameters listed in table 5.1 using the experimental data in table 5.2 and the predictions of our nuclear collision model. We trained, for this purpose, an emulator using the simulated observables predicted by the nuclear collision model at each design point (see section 4.3 for details). Once the emulator was trained, we validated the emulator results by comparing the emulator predictions to explicit model calculations using an independent set of parameter points which were excluded from the calibration. Figure 5.2 shows this validation test.
applied to several of the model observables used in the analysis. The results demonstrate the accuracy of the emulator, as evidenced by the proximity of each point to the diagonal gray line indicating perfect emulator and model agreement.

Before I proceed to present the results of the analysis, a few comments are in order about its treatment of uncertainties. Generally speaking, there are two sources of quantifiable uncertainty which should be accounted for in the analysis. First, there is the experimental uncertainty quantified by the covariance matrix $\Sigma_e$, consisting of individual statistical and systematic contributions. And second, there is the model (emulator) uncertainty quantified by the covariance matrix $\Sigma_m$. It includes statistical uncertainty in the model outputs and interpolation uncertainty arising from limited training data. The overall likelihood function is then calculated using the total covariance matrix (4.53), equal to their sum, $\Sigma = \Sigma_m + \Sigma_e$.

Previously, in subsection 4.4.1.1, I described how to calculate the likelihood function equation (4.52) in the physical observable basis, i.e. in the basis of $y$, following the derivation outlined in [140]. In the study I am discussing now, the likelihood function was calculated in the basis of the PCA transformed observables, i.e. in the basis of $z$. Expressed in this form, the likelihood function becomes

$$P(E|x) \propto \exp \left\{-\frac{1}{2} [z_m(x) - z_e]^T \Sigma_{z}^{-1} [z_m(x) - z_e] \right\}, \quad (5.8)$$

Figure 5.2 Scatter plot of the predicted (emulated) model output versus the observed (true) model output for the pion yield (left), mean $p_T$ (middle) and elliptic flow cumulant $v_2\{2\}$ (right). Diagonal line indicates perfect agreement.
where $z_m(x)$ and $z_e$ are the PCA transformed model (emulator) observables and experimental observables respectively, and $\Sigma_z$ is their total covariance matrix. Here I’ve discarded the overall normalization prefactor when writing equation (5.8) because it is not necessary for MCMC sampling. A 10% fractional uncertainty was then placed on the PCs following

$$\Sigma_z = \text{diag}(\sigma_z^2 z_e^2), \quad (5.9)$$

using $\sigma_z = 0.10$. This was a rough approximation intended to conservatively account for various sources of uncertainty in the experimental data, model calculations, and emulator predictions. Recently, more advanced methods have been developed to rigorously estimate the likelihood covariance matrix [221]. At the time, approximation (5.9) was merely a stop-gap.

We placed a uniform prior on each model parameter which was constant within the design ranges listed in table 5.1 and zero outside them. We then applied Bayes’ theorem (4.51) to calculate the posterior probability density at each parameter point using the likelihood function and the prior. Finally, we used the affine-invariant MCMC ensemble sampler \texttt{emcee} [211] to draw samples from the posterior distribution, first generating $O(10^6)$ burn-in samples to allow the chain to equilibrate followed by $O(10^7)$ production samples to visualize the posterior.

### 5.1.4 Posterior parameter estimates

Figure 5.3 shows simulated model observables (colored lines) compared to the experimental data (black symbols). The top row of the figure shows explicit model calculations at each of the 300 design points. Their large visual spread is a result of each parameter’s initial design range (see table 5.1) reflecting the prior uncertainty in the true parameter values. Now direct your attention to the bottom row of the figure. These lines are emulator predictions using 100 parameter combinations randomly sampled from the posterior. The lines are now tightly clustered around the experimental data, reflecting the uncertainty of the posterior distribution and the trade-offs that are made when fitting all of the observables.
Figure 5.3  Model and emulator predictions (colored lines) compared to experimental data (black symbols and error bars). Top row: explicit model calculations at each of the 300 design points used to train the emulator. Bottom row: emulator predictions using 100 parameter combinations randomly drawn from the posterior. Left column shows identified particle yields $dN/dy$ for pions, kaons, and protons, middle column shows mean $p_T$ for pions, kaons, and protons, and right column shows the two-particle flow cumulants $v_n\{2\}$ for charged particles. Experimental data is from the ALICE collaboration [65, 108].

The agreement of the posterior sampled emulator predictions is really quite impressive, matching the accuracy of the most advanced dynamical models at the time [85, 86, 114, 119]. The result is testament to the remarkable success of the hydrodynamic standard model which is the foundation of our model-to-data comparison. Moreover, it validates the general parametric initial condition approach used throughout this dissertation. The agreement of the calibrated model with the data is evidence that the TRENTo model is able to well reproduce the data using appropriately chosen parameters. Of course, this does not necessarily mean the model is correct or anything of that nature—our inference could still be contaminated by errors lurking elsewhere in the framework. Rather, it demonstrates that the model survives a significant test of its prediction accuracy.

The larger scope of this work is of course not to merely validate the model, but to learn about the underlying physics of the collision using the constraints on each parameter imposed by the experimental data and the assumptions of our framework. Figure 5.4 presents
Table 5.3  Estimated parameter values (medians) and uncertainties (90% credible intervals) from
the posterior distributions calibrated to identified and charged particle yields (middle and right
columns, respectively). The distribution for $T_{\text{switch}}$ based on charged particles is essentially flat, so
we do not report a quantitative estimate.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Identified</th>
<th>Charged</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normalization</td>
<td>120.$^{+8.}_{-8.}$</td>
<td>132.$^{+11.}_{-11.}$</td>
</tr>
<tr>
<td>$p$</td>
<td>$-0.02^{+0.16}_{-0.18}$</td>
<td>$0.03^{+0.16}_{-0.17}$</td>
</tr>
<tr>
<td>$k$</td>
<td>1.7$^{+0.5}_{-0.5}$</td>
<td>1.6$^{+0.6}_{-0.6}$</td>
</tr>
<tr>
<td>$w$ [fm]</td>
<td>0.48$^{+0.10}_{-0.07}$</td>
<td>0.51$^{+0.10}_{-0.09}$</td>
</tr>
<tr>
<td>$\eta/s$ min</td>
<td>0.07$^{+0.05}_{-0.04}$</td>
<td>0.08$^{+0.05}_{-0.05}$</td>
</tr>
<tr>
<td>$\eta/s$ slope [GeV$^{-1}$]</td>
<td>0.93$^{+0.65}_{-0.92}$</td>
<td>0.65$^{+0.77}_{-0.65}$</td>
</tr>
<tr>
<td>$\zeta/s$ norm</td>
<td>1.2$^{+0.2}_{-0.3}$</td>
<td>1.1$^{+0.5}_{-0.5}$</td>
</tr>
<tr>
<td>$T_{\text{switch}}$ [GeV]</td>
<td>0.148$^{+0.002}_{-0.002}$</td>
<td>—</td>
</tr>
</tbody>
</table>

the primary result of this study, a visualization of the Bayesian posterior distribution for
the model input parameters. Each diagonal panel is the marginal distribution of a single
model parameter, and each off-diagonal panel is the joint distribution of a pair of model
parameters visualizing their correlations. There are, in fact, two posterior distributions in
this figure. The blue lower-triangle is the posterior that’s obtained when pion, kaon, and
proton yields $dN/dy$ are included in the calibration, and the red upper-triangle is the pos-
terior that’s obtained when these identified yields are replaced by the charged particle yield
$dN_{\text{ch}}/d\eta$ instead. The reason for showing both posteriors will be explained in a moment.

Table 5.3 also lists the median and 90% highest posterior density (HPD) credible interval
for each model parameter. This interval is defined in section 4.5 as the smallest interval
containing 90% of the parameter’s posterior density. Roughly speaking, we expect the
true parameter values to fall within these ranges 90% of the time assuming the model is
exact and that all uncertainties are accounted for. There is a lot of information to unpack
here, so I’ll start with the marginal distributions in figure 5.4. These are the red and blue
histograms located on the diagonal. I’ll begin with the top-left corner and work my way to
the bottom-right.
Figure 5.4  Posterior distribution for the model parameters, calibrated including identified particle yields (blue, lower triangle) and charged particle yields (red, upper triangle). Each diagonal panels is the marginal distribution of a single parameter, and each off-diagonal panel is the joint distribution of a pair of parameters visualizing their correlations. The parameter \((\eta/s)_{hrg}\) has been dropped from the posterior distribution for reasons which are explained in the text.
5.1.4.1 Initial condition parameters

The initial entropy normalization is unsurprisingly well constrained. Here we see that the normalization is different if we calibrate to fit the identified particle yields (blue histograms) or the charged particle yields (red histograms). This is because the model cannot fit the pion, kaon, and proton yields simultaneously. So it generally tends to underpredict the pion yield in order to better fit the kaons and protons. This was our original motivation for calibrating with and without the identified particle yields included.

The next diagonal panel shows the marginal distribution for the generalized mean parameter $p$ which controls the scaling of initial entropy deposition as a function of nuclear thickness, equation (5.1). Here we see a peak centered at $p = 0$. Recall that this value corresponds to a geometric mean

$$s(x_\perp; \eta_s = 0, \tau_0) = \text{Norm} \frac{\tau_0}{\tau_0} \times \sqrt{\bar{T}_A \bar{T}_B},$$

(5.10)

where $\bar{T}_A$ and $\bar{T}_B$ are the participant thickness functions. Note, this is the exact same $p$ value supported by our preliminary examination of charged particle multiplicity distributions, figure 3.16, and eccentricity harmonics, figure 3.18. This consensus is strong indication of universal scaling.

Previously in subsection 3.5.1, I showed that the $p$ parameter smoothly interpolates different classes of initial condition models; the wounded nucleon model is equivalent to arithmetic mean scaling $p = 1$, the EKRT and IP-Glasma models behave similarly to geometric mean scaling $p \sim 0$, and the KLN model is closely fit by $p = -0.67$. Figure 5.5 shows an expanded view of the Bayesian posterior distribution on $p$ with black symbols and bands marking the effective $p$ values needed to describe each of the aforementioned models. The EKRT and IP-Glasma models lie squarely in the peak of the posterior distribution, while the KLN and wounded nucleon models are considerably outside, corroborating the findings of several previous model validation studies [38, 114, 232].

The posterior distribution on the parameter $k$, meanwhile, is inconclusive. Recall that
Figure 5.5 Posterior distribution on the TRENTo generalized mean parameter $p$ compared to the effective values of $p$ needed to mimic the KLN, EKRT, IP-Glasma, and Wounded nucleon models. These values are explained in subsection 3.5.1.

this parameter is the inverse variance of the Gamma distribution used to sample nucleon-nucleon fluctuations; see equation (3.48). In hindsight, we’ve realized there is a problem with this parametrization. The only way to turn the fluctuations off is to send $k$ to infinity. Our prior tests a finite range of $k$ values, so it never tests the model predictions with the fluctuations turned completely off. We’ve fixed this issue in later studies by replacing the parameter $k$ with the standard deviation of the fluctuations $\sigma_{\text{fluct}} = 1/\sqrt{k}$. We’ll use this modified parametrization in the next study.

Continuing down the diagonal, the Gaussian nucleon width $w$ is narrowly peaked mostly within the range 0.4–0.6 fm. This estimate is quantitatively consistent with the value for the gluonic nucleon width extracted from deep inelastic scattering data at HERA [233–235], and it is also similar to the values commonly used in studies of the IP-Glasma and EKRT models [85, 114]. However, this estimate is also somewhat puzzling in hindsight because we obtained much larger estimates for $w$ in our subsequent Bayesian analyses. Evidently the preferred nucleon width is sensitive to the details of the dynamical model. I’ll return to this puzzle later in the chapter.

5.1.4.2 QGP medium parameters

The next three parameters $(\eta/s)_{\min}$, $(\eta/s)_{\text{slope}}$, and $(\zeta/s)_{\text{norm}}$ all control the temperature dependence of the QGP viscosity. Here, we’ve dropped the parameter $(\eta/s)_{\text{hrg}}$ from the posterior distribution because it has no effect on the model, i.e. its posterior distribution is completely flat. The $(\eta/s)_{\text{hrg}}$ parameter controls the specific shear viscosity of the hy-
drodynamic simulation below the temperature $T_c = 0.154$ GeV, so it only affects the model output if we particlize the fluid well below that temperature. The model generally prefers a particlization (switching) temperature near $T_c$ (see below), so the $(\eta/s)_{\text{hrg}}$ parameter ultimately has little effect.

The posterior on $(\eta/s)_{\text{min}}$ is mostly peaked between 0 and 0.15. Its temperature dependent slope $(\eta/s)_{\text{slope}}$, meanwhile, is far broader and extends to either edge of its prior; only large slopes $(\eta/s)_{\text{slope}} > 2$ GeV appear to be excluded. These two marginal distributions, however, paint an incomplete picture of the posterior on $(\eta/s)(T)$. The joint posterior distribution of $(\eta/s)_{\text{min}}$ and $(\eta/s)_{\text{slope}}$ reveals that both parameters are inversely correlated. The posterior allows for a small value of $(\eta/s)_{\text{min}}$ and a large value of $(\eta/s)_{\text{slope}}$ or vice versa, but both cannot be large (or small) at the same time. Hence, most of the uncertainty in $(\eta/s)_{\text{slope}}$ arises from this degeneracy. This is exactly the type of relationship that is difficult to discover without using Bayesian parameter estimation.

There is, alternatively, a better way to visualize the posterior estimate for the temperature dependence of the specific shear viscosity $(\eta/s)(T)$. We modeled this function above the critical temperature $T_c$ using the linear ansatz

$$
(\eta/s)(T) = (\eta/s)_{\text{min}} + (\eta/s)_{\text{slope}}(T - T_c),
$$

(5.11)

with two parameters $(\eta/s)_{\text{min}}$ and $(\eta/s)_{\text{slope}}$. Now that we have the joint posterior distribution for these parameters, we can sample the function defined by equation (5.11). Concatenating these samples into a single list, we can then calculate a credible interval at each value of the temperature to form a credible region (CR). Figure 5.6 applies this method to visualize our estimate for the temperature dependence of the specific shear viscosity for $T \geq T_c$. The gray band is the region spanned by the prior distribution on the shear viscosity parameters, the blue band is the posterior’s 90% CR, and the blue line is its median. We’ve also superimposed the well-known KSS bound [236–238] from AdS-CFT holography for reference purposes.

Perhaps the most interesting feature of this figure is its characteristic bow-tie shape.
Figure 5.6  Posterior estimate for the temperature dependence of the QGP specific shear viscosity $\eta/s$. The trapezoidal gray region is the prior range explored by the parametrization (5.11). The blue band is the parametrization’s posterior 90% HPD credible region, and the blue line its median. The horizontal gray line is the KSS bound $\eta/s \geq 1/4\pi$ [236–238].

The posterior distribution narrows slightly above $\sim$200 MeV and fans out again on either side. Presumably this is the temperature that the medium evolution is most sensitive to. It suggests that we may be able to get a better handle on the temperature dependence of the QGP viscosity by including more beam energies and hence more handles at different temperature points on this graph. Indeed, I will summarize the results of a study which does exactly that in the next section.

The posterior on the bulk viscosity $(\zeta/s)_{\text{norm}}$ parameter (scaling prefactor for equation (5.4)), meanwhile, is clearly peaked just above one. Therefore, the results agree with leaving the bulk viscosity parametrization unscaled, as in reference [119]. Given the previously mentioned short-comings in our treatment of bulk viscosity, namely neglecting bulk corrections at particlization and the lack of a dynamical pre-equilibrium phase, we refrained from making any quantitative statements about the temperature dependence of bulk viscosity. Suffice to say, the analysis clearly prefers non-zero bulk viscosity.

The last model parameter is the Cooper-Frye particlization temperature $T_{\text{switch}}$. This parameter sets the temperature of thermal particle emission, and it strongly affects the relative particle abundances. Therefore it is primarily constrained by the pion, kaon, and proton yields $dN/dy$. This is easily verified by removing these yields from the calibration.
(red posterior distribution). Once the yields are removed, the constraint on the particlization temperature vanishes. Evidently, there was no single particlization temperature which could fit the pion, kaon, and proton yields simultaneously, but $T_{\text{switch}} \sim 148$ MeV provided the best overall compromise.

### 5.1.5 Verification of high-probability parameters

As a final test of the emulator prediction accuracy and our model’s veracity, we ran $\mathcal{O}(10^6)$ events using a single combination of high-probability parameters selected from the mode of the posterior probability distribution. Roughly speaking, these values demonstrate the single best-fit attainable by the model. Table 5.4 lists these parameter values and figure 5.7 shows the resulting model predictions compared to the experimental data. The identified and charged particle yields $dN/dy$ and $dN_{\text{ch}}/d\eta$ are shown on the left, the identified mean $p_T$ in the middle, and the two-particle flow cumulants $v_n\{2\}$ on the right. The inset below each figure shows the ratio of the model calculations to the data, and the gray band indicates $\pm 10\%$ error.

**Table 5.4** High-probability parameters chosen based on the posterior distributions and used to generate Fig. 5.7. Pairs of values separated by slashes are based on identified / charged particle yields, respectively. Single values are the same for both cases.

<table>
<thead>
<tr>
<th>Initial condition / Pre-eq</th>
<th>QGP medium</th>
</tr>
</thead>
<tbody>
<tr>
<td>norm 120. / 129.</td>
<td>$\eta/s$ min 0.08</td>
</tr>
<tr>
<td>$p$ 0.0</td>
<td>$\eta/s$ slope 0.85 / 0.75 GeV$^{-1}$</td>
</tr>
<tr>
<td>$k$ 1.5 / 1.6</td>
<td>$\zeta/s$ norm 1.25 / 1.10</td>
</tr>
<tr>
<td>$w$ 0.43 / 0.49 fm</td>
<td>$T_{\text{switch}}$ 0.148 GeV</td>
</tr>
</tbody>
</table>

The agreement of all observables is $\sim 10\%$ or better with two notable exceptions. The pion, kaon, and proton yields $dN/dy$ are off by $\sim 10$–$20\%$, particularly in the most peripheral centrality bins, and the two-particle flow cumulants $v_n\{2\}$ deviate from the data in the 0–5% centrality range. Both issues are known problems for this class of model. All nitpicking aside, the model performance is really quite remarkable. For instance, the total charged particle production is essentially perfect all the way out to 80% centrality, suggesting that the tension in the identified particle yields is a problem with the medium chemistry, not
Figure 5.7  Explicit model calculations using the high-probability parameters listed in table 5.4. Solid lines are calculations using parameters based on the identified particle posterior, dashed lines are calculations using parameters based on the charged particle posterior, and black symbols with errors are experimental data from ALICE [65, 108]. Top row: Charged or identified particle yields $dN/dy$ or $dN_{ch}/d\eta$ (left), identified particle mean $p_T$ (middle), and two-particle flow cumulants $v_n(2)$ (right). Bottom row: Ratio of the model to the experimental data.

with initial entropy deposition. It’s also impressive that the model correctly describes the nuanced shape of each mean $p_T$ and flow cumulant curve. For example, the mean $p_T$ of pions, kaons, and protons is more separated in $0–10\%$ centrality range than it is in the $60–70\%$ centrality range for both the model calculation and experiment. The model also nails the centrality dependence of the anisotropic flow harmonics, observables that are notoriously sensitive to the geometry of the initial conditions and the temperature dependence of the hydrodynamic transport coefficients.

I should reemphasize, however, that the aforementioned results include several caveats. We did not account for any uncertainty in the pre-equilibrium dynamics that describe the first moments of the collision prior to the onset of hydrodynamic expansion, asserting instead instant thermalization at the hydrodynamic starting time $\tau_0$. We also chose a rather restrictive parametrization for the temperature dependence of the specific bulk viscosity, and we completely neglected the bulk viscous corrections to the particle distribution at freeze-out. Moreover, we took several short-cuts when estimating the likelihood covariance matrix which degrade the strict interpretation of our uncertainty estimates. These defects are the reason why I initially referred to the study as a proof of concept.
5.1.6 Improved boost-invariant nuclear collision model

This subsection describes several improvements to the boost-invariant nuclear collision model used in the previous study. The most significant changes to the model are the addition of a pre-equilibrium free streaming stage [239] and a newly developed particle sampler [240] which were added by Jonah Bernhard. The following describes these upgrades among others which distinguish the dynamical model from its previous version. I’ll also describe some computational tricks which greatly reduced the time needed to run the model. These modifications are relevant to results which I will present later in the chapter.

5.1.6.1 Initial conditions

The updated model includes two important changes to TRENTo:

1. The nucleon minimum distance algorithm described in subsection 3.2.2 was added to study the effect of nucleon correlations on our sampled lead nuclei. The algorithm imposes a minimum distance criterion

   \[ |x_i - x_j| > d_{\text{min}} \]  \hspace{1cm} (5.12)

   between all pairs of nucleons \( i, j \), where \( d_{\text{min}} \) is a variable model parameter. This algorithm is constructed to preserve the desired Woods-Saxon radial distribution so there is no artificial swelling of the nucleus with increasing \( d_{\text{min}} \).

2. Second, we reinterpreted the output of the TRENTo model as an initial energy density profile \( e = dE/(d^2x_\perp d\tau_0 d\eta_s) \), written as

   \[ e(x_\perp; \eta_s = 0, \tau_0) = \text{const} \times \left[ \frac{\tilde{T}_A(x_\perp)^p + \tilde{T}_B(x_\perp)^p}{2} \right]^{1/p}, \]  \hspace{1cm} (5.13)

   where \( \text{const} \) is an overall normalization factor with units of GeV/fm.

   Here we opted to parametrize the initial energy density (rather than the initial entropy density) since its dynamics are exactly calculable in the weakly-coupled non-interacting
limit, subject to certain simplifying assumptions [90, 91]. This so-called free streaming approximation provides a more realistic starting point for the hydrodynamic equations of motion, as I’ll explain below.

### 5.1.6.2 Pre-equilibrium evolution

The nuclear collision model was updated to free stream the initial energy density modeled by equation (5.13) to the hydrodynamic starting time using a procedure developed in [90, 91]. The free streaming evolution was implemented using FREESTREAM, a publicly available Python code written by J. Bernhard [239]. The following excerpt summarizes the free streaming procedure in my own words. This text also appears in reference [221].

The initially deposited matter modeled by equation (5.13) is expected to rapidly approach the conditions of hydrodynamic applicability over a timescale of $\tau \sim 1 \text{ fm}/c$, although the details of this evolution are the subject of ongoing investigation [87, 241]. There are, of course, two natural limiting cases for the initial strength of the medium interactions: infinitely weak coupling, where the matter free streams without interacting; and infinitely strong coupling, where the inter-particle mean free path effectively vanishes. Realistically, one expects the initially produced medium to lie somewhere between these two extremes. The updated model therefore free streams the energy density $e(x_\perp, \eta_s = 0, \tau_0)$ until a variable hydrodynamic starting time $\tau_{fs} > \tau_0$ to modulate the time-averaged strength of medium interactions prior to the onset of hydrodynamic expansion. Figure 5.8 shows a cartoon of this evolution.

The free streaming evolution is easily calculable for a boost-invariant gas of massless partons with locally isotropic transverse momenta [90, 91]. Under these conditions, the energy-momentum tensor at transverse position $x_\perp$ and time $\tau > \tau_0$ is given by

$$T^{\mu\nu}(x_\perp, \eta_s = 0, \tau) = \frac{\tau_0}{\tau} \int_{-\pi}^{\pi} d\phi_p \hat{p}^\mu \hat{p}^\nu e(x_\perp - (\tau - \tau_0)\hat{p}_\perp, \eta_s = 0, \tau_0),$$  \hspace{1cm} (5.14)

where $\hat{p}^\mu = (1, \cos \phi_p, \sin \phi_p, 0)$ and $\hat{p}_\perp = (\cos \phi_p, \sin \phi_p)$ are momentum unit vectors. At top RHIC and LHC energies, the nuclear interpenetration time $\tau_0 \ll 1 \text{ fm}/c$ so the free
streaming time $\tau_{fs} \equiv \tau - \tau_0 \approx \tau$. One can also combine the normalization constant in equation (5.13) with the prefactor $\tau_0$ in equation (5.14) to yield a single normalization factor, $\text{Norm} = \text{const} \tau_0$, with units of energy, which varies the overall magnitude of energy deposition for all collisions at a given beam energy.

At time $\tau_{fs}$, the energy-momentum tensor can be decomposed into hydrodynamic form

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

(5.15)

in order to match the free streamed energy-momentum tensor with viscous relativistic hydrodynamics. Here $e$ and $P$ are the energy density and pressure in the local fluid rest frame, $u^\mu$ is the local fluid velocity, $\Delta^{\mu \nu} = g^{\mu \nu} - u^\mu u^\nu$ is the projector onto the space orthogonal to $u^\mu$, and $\pi^{\mu \nu}$ and $\Pi$ are the shear and bulk viscous pressures, respectively. The Landau matching condition furnishes the energy density $e$ and fluid velocity $u^\mu$ as the timelike eigenvalue and eigenvector of the energy-momentum tensor:

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

(5.16)

The equilibrium pressure $P = P(e)$ can then obtained from the equation of state (see

![Figure 5.8](image.png)

**Figure 5.8** Cartoon of the coupling strength for the first $\sim 1 \text{ fm/c}$ of the collision. The solid line shows a reasonable time-evolution for the coupling strength in nature, and the dashed line shows the free streaming approximation used in this work.
below), and the bulk pressure from the difference with the total pressure

\[ P + \Pi = -\frac{1}{3} \Delta_{\mu\nu} T^{\mu\nu}. \] (5.17)

Finally, the shear pressure tensor can be obtained by rearranging equation (5.15)

\[ \pi^{\mu\nu} = T^{\mu\nu} - eu^{\mu} u^{\nu} + (P + \Pi) \Delta^{\mu\nu}, \] (5.18)

since everything on the right-hand side is now known. This provides all the information necessary to initialize the hydrodynamic equations of motion.

### 5.1.6.3 Boost-invariant viscous hydrodynamics

The upgraded transport model uses the exact same boost-invariant VISH2+1 hydrodynamics code with shear-bulk coupling [214–216] described in subsection 5.1.1 with a slight modification to the equation of state (EoS). As before, the EoS was constructed by blending lattice QCD calculations at high-temperature [18] with a hadron resonance gas calculation at low temperatures. However, in the updated model, the blending procedure is performed somewhat differently. It matches the trace anomalies \( I(T) = (e - 3p)/T^4 \) of each calculation in the interval between 165 and 200 MeV using a Krogh polynomial, which ensures continuity of the functions and their first several derivatives; see reference [140].

The updated model also uses new parametrizations for the temperature dependent shear and bulk viscosities [140]. For the shear viscosity, it uses

\[
(\eta/s)(T) = \begin{cases} 
(\eta/s)_{\text{min}} + (\eta/s)_{\text{slope}} (T - T_c) \cdot (T/T_c)^{(\eta/s)_{\text{crv}}} & T > T_c, \\
(\eta/s)_{\text{hrg}} & T \leq T_c,
\end{cases}
\] (5.19)

which upgrades equation (5.3) by adding a dimensionless curvature parameter \((\eta/s)_{\text{crv}}\) to vary the second derivative of the \((\eta/s)(T)\) above the critical temperature \(T_c = 0.154\) GeV. Values \((\eta/s)_{\text{crv}} > 1\) introduce positive curvature, while values \((\eta/s)_{\text{crv}} < 1\) introduce neg-
Figure 5.9 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, $\eta/s$ could have a large slope and negative curvature, or $\zeta/s$ could have a large max and narrow width, neither of which are depicted above. Note, $y$-axis limits are different for the shear and bulk viscosity plots.

...ative curvature. Meanwhile, for $(\eta/s)_{crv} = 1$, the curvature is turned off, reproducing equation (5.3).

The bulk viscosity parametrization was completely overhauled using an unscaled Cauchy distribution

$$\zeta(T) = \frac{(\zeta/s)_{max}}{1 + \left(\frac{T - (\zeta/s)T_0}{(\zeta/s)_{width}}\right)^2},$$

(5.20)

where $(\zeta/s)_{max}$ is the maximum value of $\zeta/s$, $(\zeta/s)T_0$ is the location of the peak (units of temperature), and $(\zeta/s)_{width}$ is the width of the peak (also units of temperature). Figure 5.9 shows several of the possible $(\eta/s)(T)$ and $(\zeta/s)(T)$ curves parametrized by equations (5.19) and (5.20).

Additionally, we modified the simulation’s hydrodynamic grid. In the previous version of the transport model, we used a square transverse grid $|x| < x_{max}$ of fixed maximum extent $x_{max} = 13$ fm and step width $dx = 0.1$ fm along each dimension. These are the VISH2+1 defaults. In general, the grid should be made large enough to contain the entire event but no larger. If the grid is too large, precious computation time is wasted, and if the grid is too small, the spacetime evolution of the event will be clipped by the edge of the grid.
The updated model uses a newly developed procedure, visualized in figure 5.10, to vary the maximum grid size $x_{\text{max}}$ event-by-event, allowing large events to run with large grids and small events with small grids. This procedure is as follows:

1. Generate a minimum-bias TrENTo event using a square transverse grid with $x_{\text{max}} = 15$ fm and $dx = 0.15 \times w$, where $w$ is the nucleon width (I’ll explain this choice shortly). Call this the fine grid.

2. Free stream the event to time $\tau_{\text{fs}}$ to calculate the energy density $e$, flow velocity $u^\mu$, and viscous corrections $\pi^{\mu\nu}$ and $\Pi$ on the fine grid.

3. Enlarge these fine grids to $x_{\text{max}} = 27$ fm by padding each with zeros. Then take every third grid cell along each dimension to resolve the same event on a thrice coarser grid. Call this the coarse grid.

4. Run the large coarse grid through the hydrodynamic simulation, setting the viscosities to zero, and calculate the smallest transverse radius $R_{\text{max}}$ that fully encloses the prespecified isotherm $T = T_{\min}$ for the full lifetime of the event.

5. Trim the original fine grid using $x_{\text{max}} \to R_{\text{max}}$. Then rerun the event (with viscosity on) using the trimmed fine grid.

This procedure works because the coarse ideal hydro event runs dramatically faster than the thrice finer viscous hydro event. The numerical computation time for VISH2+1 scales like $\sim n_x^3$, where $n_x$ is the number of grid cells along either spatial dimension, so the thrice coarser grid runs $\sim 27$ times faster. Obviously, if the spatial step width $dx$ is too large, numerical viscosity will spoil the simulation accuracy. However, the maximum truncation radius $R_{\text{max}}$ can be estimated using grids which are far coarser than those needed to calculate typical observables. For the truncation isotherm, we found that $T_{\min} = 110$ MeV leads to a good compromise between numerical accuracy and speed, resulting in numerical errors of a few percent or less for most observables at LHC energies. Of course, we could lower $T_{\min}$ further still, but it would likely hurt our simulation accuracy since it would limit our simulation statistics.
Figure 5.10  Diagram of the adaptive grid resizing algorithm (not 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 $R_{\text{max}}$ (dashed circle) of the hypersurface defined by the temperature isotherm $T = T_{\text{min}}$. The temperature $T_{\text{min}}$ is a prespecified isotherm, below which the event can be truncated without modifying the simulated observables. 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 $x_{\text{max}} = R_{\text{max}}$. Figure and caption are adapted from [204].

In a similar fashion, the hydrodynamic grid should be fine enough to calculate simulated observables accurately but no finer. Generally speaking, smaller nucleon widths $w$ require finer grids, since they produce smaller transverse structures in the initial energy density profile. Thus, we decided to replace the fixed grid step width $w = 0.1$ fm in the previous version of the model with a variable step width $dx = 0.15 \times w$, equal to 15% of the chosen nucleon width. This leads to slight improvements in the simulation accuracy and faster event times when the nucleon width $w$ is large.

While somewhat technical and mundane, these grid modifications cut the CPU time needed to run the model in half. The projects discussed in this dissertation required $O(10)$ million CPU hours to execute, so these savings are significant. This grid resizing method is also completely general and should work for a number of different applications. Indeed, I’ll use this exact same method in my final study where I study nuclear collisions with nucleon substructure.
5.1.6.4 Particlization and Boltzmann transport

One issue with the previous model described in 5.1.1 was that it completely neglected bulk viscous corrections when converting the hydrodynamic fluid into particles using the Cooper-Frye algorithm. This precluded quantitative statements about the QGP bulk viscosity, and it also put a caveat on other parameter estimates that are strongly correlated with bulk viscosity. For example, figure 5.4 shows that the Gaussian nucleon width $w$ and bulk viscosity normalization $(\zeta/s)_{\text{norm}}$ are inversely correlated. Hence, one should interpret our previous estimate for the nucleon width $w$ with care.

Collaborator Jonah Bernhard addressed this issue by developing FRZOUT, a new computer model for Cooper-Frye particle emission [240]. His model includes, among other features, a novel method to implement bulk viscous corrections to the particle distribution functions. It addresses many of the issues which led us to neglect bulk corrections in the first place, such as their stability against large bulk pressures, enabling a more realistic treatment of the interface between the hydrodynamic and microscopic transport models. The following is a short summary of his model [140].

The FRZOUT model starts by sampling particles from the isothermal spacetime hypersurface $\sigma$ defined by the hydro-to-micro switching temperature $T_{\text{switch}}$. Particles are emitted from the hypersurface $\sigma$ with momentum distributions described by the Cooper-Frye formula [226]

$$E^i dN_i = g_i \frac{1}{(2\pi)^3} \int_{\sigma} f_i(p) p^\mu d^3 \sigma_\mu, \quad (5.21)$$

where $g_i$ and $f_i$ are the degeneracy and distribution function of particle species $i$, and $d^3 \sigma_\mu$ is a volume element on the hypersurface $\sigma$.

The particles are sampled in the local rest frame of each fluid cell, using either a Bose-Einstein or Fermi-Dirac thermal distribution

$$f(m, p) = \frac{1}{\exp(\sqrt{m^2 + p^2}/T) + 1}, \quad (5.22)$$

where $m$ is the mass of the particle, $p$ is its momentum in the rest frame of the fluid cell,
and $T$ is the fluid cell temperature. If the particle is a resonance, it samples a distribution of masses $P(m)$. The distribution $f(p)$ is then calculated by integrating out this mass dependence, yielding

$$f(p) = \int dm P(m) f(m, p). \quad (5.23)$$

Typically, computational implementations neglect the mass-width of resonances, using instead $P(m) = \delta(m - m_0)$ for every particle where $m_0$ is the mass where the resonance peaks. Indeed, our previous study did not include the effect of each resonance width. However, such approximations are crude and are known to affect the particle yields, particularly at low $p_T$ [242, 243].

The FRZOUT model therefore samples resonances with a distribution of masses, modeled by a Breit-Wigner distribution

$$P(m) \propto \frac{\Gamma(m)}{(m - m_0)^2 + \Gamma(m)^2 / 4}, \quad (5.24)$$

where $m_0$ is the usual Breit-Wigner mass, and $\Gamma$ is its mass dependent width. This width is then parametrized using the simple form

$$\Gamma(m) = \Gamma_0 \sqrt{\frac{m - m_{\text{min}}}{m_0 - m_{\text{min}}}}, \quad (5.25)$$

where $\Gamma_0$ is the usual Breit-Wigner width and $m_{\text{min}}$ is a minimum mass threshold equal to the sum of the masses of the lightest decay products. The particle properties $\Gamma_0, m_0, m_{\text{min}}$ are provided by the Particle Data Group [244].

In hydrodynamic theory, the energy-momentum tensor is described by equation (5.15), while in kinetic theory it equals

$$T^{\mu \nu} = \sum_i g_i \int \frac{d^3 p}{(2\pi)^3} \frac{p^\mu p^\nu}{E} f_i(p), \quad (5.26)$$

where the summation runs over all particle species. Comparing these two forms, it is clear that the distribution function $f_i$ must be modified as a function of the shear pressure $\pi^{\mu \nu}$.
and bulk pressure $\Pi$, to preserve the continuity of $T^{\mu\nu}$ as the system transitions from hydrodynamics to Boltzmann transport.

The FRZOUT model applies these viscous corrections using a general procedure developed by Pratt and Torrieri to transform the sampled particle momenta inside the distribution function [245]

$$p_i \rightarrow p_i' = p_i + \sum_j \lambda_{ij} p_j,$$  \hspace{1cm} (5.27)

using a linear transformation matrix

$$\lambda_{ij} = (\lambda_{\text{shear}})_{ij} + \lambda_{\text{bulk}} \delta_{ij},$$  \hspace{1cm} (5.28)

composed of a traceless shear contribution $\lambda_{\text{shear}}$ and a bulk contribution $\lambda_{\text{bulk}}$ that’s proportional to the identity matrix $\delta_{ij}$.

When the shear pressure is small relative to the ideal pressure, the shear transformation equals

$$(\lambda_{\text{shear}})_{ij} = \frac{\tau}{2\eta} \pi_{ij},$$  \hspace{1cm} (5.29)

where $\pi_{ij}$ is the shear stress in the local fluid rest frame, and $\tau/\eta$ is the ratio of the relaxation time to the shear viscosity. In the non-interacting hadron resonance gas model, the inverse of this later quantity equals

$$\frac{\eta}{\tau} = \frac{1}{15T} \sum_i g_i \int \frac{d^3p}{(2\pi)^3} \frac{p^4}{E^2} f_i(1 \pm f_i),$$  \hspace{1cm} (5.30)

where the summation runs over all species in the gas, and $f_i$ is the equilibrium Fermi-Dirac or Bose-Einstein distribution of that species.

For the bulk viscous corrections, the FRZOUT model applies a novel numerical algorithm developed in [140]. The total kinetic pressure of the gas in kinetic theory is

$$P + \Pi = \sum_i g_i \int \frac{d^3p}{(2\pi)^3} \frac{p^2}{3E} f_i(p),$$  \hspace{1cm} (5.31)
where \( P \) is the fluid’s ideal pressure, and \( \Pi \) is its bulk pressure. For a given fluid bulk pressure \( \Pi \), the algorithm rescales the momentum inside the distribution function \( f(p) \rightarrow f(p + \lambda_{\text{bulk}}p) \) and adjusts the parameter \( \lambda_{\text{bulk}} \) to match the total pressure on the left-side of the equation. This modification also alters the fluid’s energy density

\[
e = \sum_i g_i \int \frac{d^3p}{(2\pi)^3} E f_i(p),
\]

so the algorithm introduces a second parameter \( z_{\text{bulk}} \)—assumed to be the same for all particle species—that rescales the particle yield \( f(p) \rightarrow z_{\text{bulk}}f(p) \) to compensate. The parameters \( z_{\text{bulk}} \) and \( \lambda_{\text{bulk}} \) are determined numerically for each value of the bulk pressure to match the fluid cell’s pressure and energy density.

As before, the sampled particles are passed to the UrQMD transport model which simulates their individual microscopic interactions until the last interactions cease and the medium “freezes out”, producing a list of particle IDs and four-momenta for each simulated event. There were no significant changes to our implementation of UrQMD compared to the previous iteration of the model.

5.1.7 Predictions for Xe-Xe collisions at 5.44 TeV

To this point, I have only demonstrated that the TRENTo initial condition model is postdictive, i.e. given some calibration data, the model can be tuned to fit that data. Naturally, our hope is that once the model is calibrated, it will also be able to predict observables that are out of sample, i.e. measurements not included in the calibration data. This subsection presents some previously unpublished TRENTo model predictions for Xe-Xe collisions at \( \sqrt{s_{\text{NN}}} = 5.44 \) TeV using the updated version of our nuclear collision model which I just described.

5.1.7.1 Constraints from Pb-Pb collisions at 2.76 and 5.02 TeV

I fix the TRENTo and QGP medium parameters for the Xe-Xe predictions using the results of a recently published Bayesian analysis of Pb-Pb bulk observables at \( \sqrt{s_{\text{NN}}} = 2.76 \) and
5.02 TeV [178]. This reference builds upon the previously described analysis [178] and is the capstone of J. Bernhard’s dissertation work on Bayesian parameter estimation [140]. My own contributions to this updated global analysis are meaningful, but they are not significant enough to warrant a full description of the project in its entirety here. Rather, let me highlight some key differences compared to the previous Bayesian analysis which I just described:

- The study combines Pb-Pb data from two beam energies at the LHC, $\sqrt{s_{NN}} = 2.76$ and 5.02 TeV.
- It adds transverse energy and mean $p_T$ fluctuations to the list of calibration observables.
- It uses the significantly upgraded nuclear collision model discussed in subsection 5.1.6. Notable additions include a pre-equilibrium free streaming evolution [239] and a new particle sampler [240].
- It significantly improves the treatment of uncertainties in [178].

Figure 5.11 shows the fifteen dimensional posterior distribution obtained from the analysis, and figure 5.12 shows the calibrated model predictions compared to experimental data, using parameters selected from the mode of the posterior distribution which are listed in table 5.5. The mode parameter point maximizes the posterior probability density and characterizes the best possible agreement of the model with the data. It is also commonly referred to as the maximum a posteriori (MAP) estimate.

<table>
<thead>
<tr>
<th>Table 5.5</th>
<th>Maximum a posteriori (MAP) model parameters from Ref. [221].</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial condition / Pre-eq</td>
<td>QGP medium</td>
</tr>
<tr>
<td>Norm 2.76 TeV</td>
<td>13.94 GeV</td>
</tr>
<tr>
<td>Norm 5.02 TeV</td>
<td>18.38 GeV</td>
</tr>
<tr>
<td>$p$</td>
<td>0.007</td>
</tr>
<tr>
<td>$\sigma_{\text{fluct}}$</td>
<td>0.918</td>
</tr>
<tr>
<td>$w$</td>
<td>0.956 fm</td>
</tr>
<tr>
<td>$d_{\text{min}}$</td>
<td>1.27 fm</td>
</tr>
<tr>
<td>$\tau_{\text{fs}}$</td>
<td>1.16 fin/c</td>
</tr>
</tbody>
</table>

170
Figure 5.11 Posterior distribution for the model input parameters of Ref. [221]. Diagonal panels are the marginal distributions of individual parameters and off-diagonal panels are the joint distributions for pairs of parameters.
Figure 5.12  Model calculations (colored lines) for Pb-Pb collisions at two beam energies, $\sqrt{s_{NN}} = 2.76$ TeV (solid lines) and $\sqrt{s_{NN}} = 5.02$ TeV (dashed lines), using the MAP parameters from Ref. [221], listed in table 5.5. All experimental data (symbols) are from ALICE [56, 62, 65, 108, 246–248]. The subpanel below each plot shows the ratio of the model over the experimental data. Figure is from Ref. [221].
In principle, these MAP parameters should also describe Xe-Xe collisions at LHC energies as well, with one small caveat. The TRENTo model has several parameters that scale strongly with the collision energy, so we’ll need to extrapolate these parameters from the Pb-Pb energy $\sqrt{s_{NN}} = 5.02$ TeV to the Xe-Xe energy $\sqrt{s_{NN}} = 5.44$ TeV. I’ll describe this extrapolation in a moment. First, I want to comment briefly on figures 5.11 and 5.12.

There are two results in figure 5.11 which are particularly striking: the marginal posterior distribution on the generalized mean parameter $p$, and the joint posterior distribution of the QGP viscosity parameters.

The analysis obtains a sharply peaked posterior on the TRENTo generalized mean parameter $p$ centered essentially at zero. Figure 5.13 shows this posterior distribution compared to the effective $p$ values needed to fit the KLN, EKRT, IP-Glasma, and wounded nucleon models determined in subsection 3.5.1. Note, our interpretation of this parameter is somewhat different than before. Originally in reference [178], we parametrized the initial entropy density using the generalized mean ansatz (3.58)

$$s_0 \propto M_p(\tilde{T}_A, \tilde{T}_B),$$

(5.33)

assuming static profiles at the hydrodynamic starting time with $u^\mu = 0$. However in this study, we parametrized the initial energy density

$$e_0 \propto M_p(\tilde{T}_A, \tilde{T}_B).$$

(5.34)

This energy density was free streamed to the hydrodynamic starting time and matched to viscous hydrodynamics using the Landau matching procedure. Evidently, both prescriptions prefer $p \sim 0$, leading to similar geometric mean scaling

$$\left\{ \begin{array}{l} e_0 \\ s_0 \end{array} \right\} \propto \sqrt{\tilde{T}_A \tilde{T}_B},$$

(5.35)

as a function of nuclear thickness. Suffice to say, our strict interpretation of the parameter...
p depends on the assumed matching procedure.

The study also presents the first Bayesian estimate for the temperature dependence of the QGP specific shear and bulk viscosities extracted from hydrodynamic simulations with rigorously calculated uncertainties. These uncertainties account for the model’s finite statistical error, the emulator’s systematic interpolation error, and the experiment’s statistical and systematic errors. Moreover, by marginalizing over the TRENTo model parameters, the analysis also accounts for the correlated effect of the QGP initial condition uncertainties.

Figure 5.14 visualizes the posterior estimates for the specific shear viscosity \((\eta/s)(T)\) and specific bulk viscosity \((\zeta/s)(T)\). The QGP shear viscosity prefers a minimum value \((\eta/s)_{\text{min}} = 0.085^{+0.026}_{-0.025}\) and a gently rising or flat slope \((\eta/s)_{\text{slope}} = 0.83^{+0.83}_{-0.83}\) GeV\(^{-1}\) in agreement with reference \([178]\). Note, we restricted the slope parameter to positive values so we excluded negative slopes \textit{a priori}. The bulk viscosity, meanwhile, can be either tall and narrow, or short and broad. Evidently, the important quantity is something like the integral of the bulk viscosity curve which scales like the product of the \((\zeta/s)_{\text{width}}\) and \((\zeta/s)_{\text{max}}\) parameters.

Finally, I want to reflect on the global agreement of the model’s MAP estimate with the experimental data from ALICE \([56, 62, 65, 108, 246–248]\) which is plotted figure 5.12. The hybrid nuclear collision model consisting of TRENTo initial conditions, free stream-
Figure 5.14  Posterior estimates for the temperature dependence of the QGP specific shear viscosity \( \eta/s \) (left) and bulk viscosity \( \zeta/s \) (right). The blue band is the parametrization’s posterior 90% HPD credible region, and the blue line is its median. Here the prior covers essentially the entire region of the figure so it is not explicitly shown. The horizontal gray line (left) is the KSS bound \( \eta/s \geq 1/4\pi \) [236–238]. Figure is from Ref. [140, 221].

VISH2+1 boost-invariant viscous hydrodynamics, FRZOUT, and UrQMD microscopic hadronic transport describes nearly all calibration observables at the 10% level or better. Indeed many of the observables agree at the level of a few percent. This is a truly remarkable agreement of theory with experiment.

5.1.7.2 Extrapolating predictions to 5.44 TeV

Let me return now to the topic of this subsection, generating hydrodynamic model predictions for Xe-Xe collisions at \( \sqrt{s_{\text{NN}}} = 5.44 \) TeV. I’ll use the Bayesian MAP parameters (table 5.5) determined by reference [221] to fix the parameters of our Xe-Xe predictions. The TRENTo model’s inelastic nucleon-nucleon cross section \( \sigma_{\text{NN}}^{\text{inel}} \) and overall normalization factor vary as a function of the beam energy, so we’ll need to scale these values to \( \sqrt{s_{\text{NN}}} = 5.44 \) TeV to account for the slight difference in beam energy compared to our Pb-Pb calibration data.

We fix the inelastic nucleon-nucleon cross section \( \sigma_{\text{NN}}^{\text{inel}} \) at 5.44 TeV by interpolating previous cross section measurements at \( \sqrt{s_{\text{NN}}} = 0.2, 2.76, 5.02 \) and 7 TeV [128, 144, 249, 250]. Figure 5.15 shows the value of the cross section at each energy fit to a second-order
Figure 5.15  Experimentally measured values for the inelastic nucleon-nucleon cross section $\sigma_{NN}^{\text{inel}}$ at $\sqrt{s_{NN}} = 0.2, 2.76, 5.02, \text{and } 7 \text{ TeV}$ [128, 144, 249, 250] plotted as a function of $\log(\sqrt{s_{NN}})$ (black symbols) and fit with a second-order polynomial (black line). The extrapolated value for the inelastic nucleon-nucleon cross section at $\sqrt{s_{NN}} = 5.44 \text{ TeV}$ (red circle) equals $\sigma_{NN}^{\text{inel}} = 7.07 \text{ fm}^2$.

polynomial as a function of $\log(\sqrt{s_{NN}})$. Using this fit, we estimate that the inelastic nucleon-nucleon cross section at $5.44 \text{ TeV}$ is $\sigma_{NN}^{\text{inel}} = 7.07 \text{ fm}^2$. Then, we estimate the relative increase of the TRENTo normalization factor from 5.02 to 5.44 TeV as follows:

1. Fit the experimental Pb-Pb charged-particle density $dN_{ch}/d\eta$, 0–10% centrality, at $\sqrt{s_{NN}} = 2.76$ and 5.02 TeV [56, 62] with a power law.

2. Predict the Pb-Pb charged-particle density $dN_{ch}/d\eta$, 0–10% centrality, at $\sqrt{s_{NN}} = 5.44 \text{ TeV}$ using the power law fit.

3. Run TRENTo Pb-Pb events at 5.02 and 5.44 TeV and calculate the normalization factor which fits the target charged-particle density $dN_{ch}/d\eta$, 0–10% centrality, at each beam energy, using the approximate scaling relation

$$dN_{ch}/d\eta = \text{norm} \times \int d^2 x_T R(x_T),$$

as depicted in figure 5.16. Finally, divide the normalization factor at 5.44 TeV by the normalization factor at 5.02 TeV to calculate their ratio. This predicts a normalization that’s $\sim 2\%$ larger for the Xe-Xe events.
Figure 5.16  Pb-Pb charged-particle density $dN_{ch}/d\eta$ for 0–10% central events at $\sqrt{s_{NN}} = 2.76$ and 5.02 TeV (black square and diamond) measured by ALICE [56, 62], fit with a power law (black line) and used to predict the same charged-particle density at $\sqrt{s_{NN}} = 5.44$ TeV (red circle). Legend lists the normalizations, equation (5.36), needed to fit the charged-particle density at each collision energy.

After adjusting the 5.02 TeV normalization in table 5.5 for the slight increase in beam energy, we run $O(10^6)$ minimum-bias Xe-Xe events using the MAP parameters from reference [221] and calculate the same observables as before. Figure 5.17 shows our MAP estimate (prediction) for Xe-Xe collisions at 5.44 TeV (dashed lines) plotted on top of our MAP estimate (postdiction) for Pb-Pb collisions at 2.76 TeV (dotted lines) and 5.02 TeV (solid lines). We also show, for reference purposes, the Pb-Pb experimental data (filled and open symbols) used to calibrate the MAP parameters [56, 62, 65, 108, 246–248]. The subpanel below each figure shows the ratio of each model calculation to the Pb-Pb model calculation at 5.02 TeV. Caution, this is a ratio of two model calculations, not the ratio of the model to data as previously plotted in figure 5.12. These ratio plots predict several interesting features for the Xe-Xe collision system.

We see that changing the collision configuration from Xe-Xe to Pb-Pb essentially rescales the centrality dependence of charged-particle production by a constant overall factor. Comparing the Xe-Xe system at 5.44 TeV to the Pb-Pb system at 5.02 TeV, we also predict a large enhancement for the triangular flow cumulant $v_3\{2\}$, particularly at small centralities, and a similarly large suppression for $v_2\{4\}$. The identified mean $p_T$’s, meanwhile, change very little as a function of beam energy and system size. Finally, we predict a striking
Figure 5.17  Model calculations (colored lines) using the MAP parameters from Ref. [221] compared to experimental data (black and white symbols) from ALICE [56, 62, 65, 108, 246–248]. The Pb-Pb model calculations at 2.76 TeV (dotted) and 5.02 TeV (solid) are postdictions, calibrated to fit the experimental data, while the Xe-Xe model calculations at 5.44 TeV (dashed) are predictions based on the calibrated parameters.
enhancement of the mean $p_T$ fluctuations $\delta p_T/\langle p_T \rangle$ for the Xe-Xe system at 5.44 TeV compared to the Pb-Pb systems at 2.76 and 5.02 TeV. Presumably, this observable is sensitive to event-by-event fluctuations which depend on the number of nucleons inside each nucleus.

5.2 Bulk matter far from midrapidity

In the previous section, I described several Bayesian studies of the initial condition and QGP medium properties at midrapidity using boost-invariant approximations. This section describes an exploratory study which relaxes these approximations to analyze the full three-dimensional structure of the produced fireball. The following content is based on the publication


which presents a study that I co-developed with primary author Weiyao Ke. I helped conceive, design, and execute the study, but I did not write the software or perform the primary data analysis.

While ab initio theoretical calculations have made tremendous progress determining the QGP initial conditions at midrapidity [85, 114], the search for a comprehensive theory which describes the full three-dimensional structure of the produced plasma remains an outstanding challenge [251]. Consider, for example, the average charged-particle yield $dN_{ch}/d\eta$ which measures the collision’s single-particle distribution. Multiple theories nicely describe the centrality dependence of this observable at midrapidity [56], but there exists (to date) no compelling explanation for its rapidity dependence. Theoretical calculations of the QGP initial conditions at nonzero rapidity therefore tend to be somewhat speculative.

Data-driven Bayesian methods provide an attractive, complementary approach to bottom-up theory calculations [140]. In this work, we parametrized the three-dimensional structure of the QGP initial conditions and constrained the model’s parameters using multi-
plicity observables measured in \( p-\text{Pb} \) and \( \text{Pb-\text{Pb}} \) collisions at the LHC [252–255]. The optimized initial condition model was then compared to a number of experimental observables outside our calibration dataset—pseudorapidity-dependent anisotropic flows, event-plane decorrelations, and flow correlations—to assess its veracity. Our results shed light on the three-dimensional structure of the QGP initial conditions and provide useful guidance for \textit{ab initio} theory calculations. They also determine realistic three-dimensional profiles of QGP entropy density which can be used to calculate hydrodynamic backgrounds for hard-probe and jet energy-loss calculations.

5.2.1 Nuclear collision model in 3+1 dimensions

This study employed a hybrid transport model qualitatively similar to the model described in subsection 5.1.1. We used relativistic hydrodynamics to simulate the hot and dense QGP phase of the collision and microscopic Boltzmann transport to simulate the cooler hadron resonance gas. However, unlike the boost-invariant model described in subsection 5.1.1, we modeled the collision dynamics in all three spatial dimensions. This required two significant changes to the nuclear collision model: we parametrized the rapidity dependence of local entropy deposition to extend the TRENTo initial condition model from two to three spatial dimensions, and we replaced the boost-invariant VISH2+1 hydrodynamic code [215] with the 3+1 dimensional hydrodynamic code vHLLE [256]. The specifics of this three-dimensional hybrid nuclear collision model are summarized below.

5.2.1.1 Initial conditions

We modeled the three-dimensional entropy density \( s(x_\perp, \eta_s, \tau_0) \) at the hydrodynamic starting time \( \tau_0 \) as the product of two functions

\[
s(x_\perp, \eta_s, \tau_0) = f(x_\perp) \times g(x_\perp, \eta_s),
\]

where the second function satisfies \( g(x_\perp, 0) = 1 \). The function \( f(x_\perp) \) models the initial transverse entropy density at midrapidity \( \eta_s = 0 \), and the function \( g(x_\perp, \eta_s) \) models the
shape of the initial entropy density profile as a function of spacetime rapidity $\eta_s$.

Following previous work [178], we parametrized the midrapidity transverse entropy density $f(x_\perp)$ using the entropy variant of the generalized mean ansatz

$$f(x_\perp) \propto \left[ \frac{T_A(x_\perp)^p + T_B(x_\perp)^p}{2} \right]^{1/p},$$

(5.38)

where $T_A, T_B$ are the participant thickness functions of nucleus $A, B$ respectively. Meanwhile, for the rapidity-dependent mapping $g(x_\perp, \eta_s)$, we constructed a flexible parametrization using a cumulant generating function characterized by a variable mean, standard deviation, and skewness. I’ll describe this rapidity parametrization in a moment. First, let me clarify our use of spacetime rapidity $\eta_s$, pseudorapidity $\eta$, and rapidity $y$.

Assuming, to first approximation, that the initially produced particles are massless and free streaming along the $\hat{z}$ direction, then

$$\frac{z}{t} = \frac{p}{|p|}. \quad (5.39)$$

This approximation allows us to identify the spacetime rapidity $\eta_s$ with the pseudorapidity $\eta$,

$$\eta_s = \frac{1}{2} \log \left( \frac{t + z}{t - z} \right) \sim \eta = \frac{1}{2} \log \left( \frac{|p| + p_z}{|p| - p_z} \right), \quad (5.40)$$

from which it follows that $g(x_\perp, \eta_s) \approx g(x_\perp, \eta)$. Furthermore, we can perform a change of variables from pseudorapidity $\eta$ to rapidity $y$ using the relations

$$g(x_\perp, \eta) \, d\eta = g(x_\perp, y) \, dy, \quad (5.41a)$$

$$\frac{dy}{d\eta} = \frac{J \cosh \eta}{\sqrt{1 + J^2 \sinh^2 \eta}}, \quad (5.41b)$$

where $J \approx \langle p_T \rangle / \langle m_T \rangle$ is a free parameter that characterizes the entropy density’s effective particle composition. It follows that the spacetime rapidity profile $g(x_\perp, \eta_s)$ can be written
as

\[ g(x_\perp, \eta_s) \approx g(x_\perp, y) \frac{dy}{d\eta}, \quad (5.42) \]

with \( dy/d\eta \) provided by equation (5.41).

We modeled the rapidity dependence of the function \( g(x_\perp, y) \) on the right-side of equation (5.42) at each transverse coordinate \( x_\perp \) by parametrizing its cumulants as functions of \( \tilde{T}_A(x_\perp) \) and \( \tilde{T}_B(x_\perp) \). Once these cumulants are known, we can reconstruct the function \( g(x_\perp, y) \) by taking the inverse Fourier transform of the cumulant generating function

\[ g(x_\perp, y) = \mathcal{F}^{-1}\{\tilde{g}(x_\perp, k)\}, \quad (5.43a) \]

\[ \log \tilde{g} = i\mu k - \frac{1}{2} \sigma^2 k^2 - \frac{1}{6} i\gamma \sigma^3 k^3 + \ldots. \quad (5.43b) \]

The function is then normalized, \( g(x_\perp, 0) = 1 \), so the midrapidity entropy density is unmodified.

Different rapidity dependent initial condition models are described by different parametrizations of the generating function cumulants. For example, the authors in reference [257] investigated two models which “shifted” and “tilted” the rapidity profile \( g \). The shifted model varied the mean of the rapidity profile, and the tilted model strongly varied its skewness. Generally speaking, however, all of the cumulants of the rapidity profile could be nonzero, and all of them could vary as functions of \( \tilde{T}_A, \tilde{T}_B \).

We therefore parametrized the first three cumulants of the rapidity profile as functions of \( \tilde{T}_A, \tilde{T}_B \). These cumulants describe the distribution’s mean \( \mu \), standard deviation \( \sigma \), and skewness \( \gamma \). We parametrized the mean \( \mu \) as

\[ \mu = \mu_0 y_{cm} \quad (5.44) \]

where \( \mu_0 \) is a dimensionless parameter, and \( y_{cm} = \frac{1}{2} \log(\tilde{T}_A/\tilde{T}_B) \) is the center-of-mass rapidity of the colliding matter. For the distribution’s standard deviation \( \sigma \), we asserted a constant \( \sigma = \sigma_0 \), and for its skewness \( \gamma \), we explored two parametrizations: a relative
skewness model

\[ \gamma = \gamma_0 \left( \frac{\tilde{T}_A - \tilde{T}_B}{T_A - T_B} \right), \quad (5.45) \]

and an absolute skewness model

\[ \gamma = \gamma_0 \left( \frac{\tilde{T}_A - \tilde{T}_B}{\tilde{T}_0} \right), \quad (5.46) \]

with \( \tilde{T}_0 = 1 \text{ fm}^{-2} \). Additional cumulants can be added to the generating function, but they increase the model complexity. Table 5.6 summarizes the parametrizations used for each generating function cumulant.

**Table 5.6** Generating function cumulant parametrizations used in the present analysis. The variable \( y_{cm} = \frac{1}{2} \log(\tilde{T}_A/\tilde{T}_B) \), and the constant \( \tilde{T}_0 = 1 \text{ fm}^{-2} \).

<table>
<thead>
<tr>
<th>Distribution cumulant:</th>
<th>Model variant</th>
<th>mean ( \mu )</th>
<th>std. ( \sigma )</th>
<th>skewness ( \gamma )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative skewness</td>
<td>( \mu_0 y_{cm} )</td>
<td>( \sigma_0 )</td>
<td>( \gamma_0 \left( \frac{\tilde{T}_A - \tilde{T}_B}{T_A + T_B} \right) )</td>
<td></td>
</tr>
<tr>
<td>Absolute skewness</td>
<td>( \mu_0 y_{cm} )</td>
<td>( \sigma_0 )</td>
<td>( \gamma_0 \left( \frac{\tilde{T}_A - \tilde{T}_B}{\tilde{T}_0} \right) )</td>
<td></td>
</tr>
</tbody>
</table>

There is, unfortunately, one significant problem with the aforementioned generating function approach. The left panel of figure 5.18 shows several example functions \( g \) generated by equation (5.43) using different values of the skewness \( \gamma \). When \( \gamma \) is large, the distribution is ill behaved and turns negative at large rapidities. The conditions to ensure a positive-definite Fourier transform are involved, so instead we introduced a regulation term

\[ \gamma \to \gamma \exp \left( -\frac{1}{2} \sigma^2 k^2 \right), \quad (5.47) \]

to suppress spurious behavior at large rapidities. The right panel of figure 5.18 shows the function with the regulator applied. It suppresses negative regions while maintaining monotonic scaling of the skewness. In realistic calculations, we set all negative entropy densities to zero.

Figure 5.19 shows the initial entropy density profile \( s(x_{\perp}, \eta_s, \tau_0) \) at the hydrodynamic
starting time $\tau_0$ generated by the model for an example Pb-Pb event (top) and $p$-Pb event (bottom) sliced along $\eta_s = 0$ (left) and $x = 0$ (right), using typical parameter values for $\mu_0, \sigma_0,$ and $\gamma_0$ annotated in the figure caption. There are large entropy density fluctuations in the $(x, y)$ plane resulting from local nucleon density fluctuations. These fluctuations generate momentum density anisotropies and hence significant forward-backward rapidity fluctuations which are visible in the $(\eta_s, y)$ plane. Note, these parameters are not yet optimized, but I’ll show somewhat later that they are in fact close to the model’s best fit parameters.

5.2.1.2 Hybrid transport model

We replaced the boost-invariant VISH2+1 hydrodynamic code [215] with vHLLE [256], a 3+1D viscous hydrodynamics code with shear and bulk corrections. The vHLLE code was initialized using the entropy density $s(x_\perp, \eta_s, \tau_0)$ in Milne coordinates at the hydrodynamic starting time $\tau_0$. Following reference [178], we assumed instant thermalization at the hydrodynamic starting time $\tau_0$, setting the initial flow velocity $u^\mu$, shear correction $\pi^{\mu \nu}$, and bulk correction $\Pi$ identically to zero. We also implemented the modern lattice based equation of state described in subsection 5.1.1.
Figure 5.19  Initial entropy density for an example Pb-Pb event (top) and p-Pb event (bottom) sliced along $\eta_s = 0$ (left) and $x = 0$ (right). Events are simulated using the relative skewness model in table 5.6 with $\mu_0 = 1$, $\sigma_0 = 3$, and $\gamma_0 = 6$ along with midrapidity parameters from [178].

Unlike the boost-invariant studies in section 5.1, we opted to fix the model’s hydrodynamic medium parameters. Ideally, we would parametrize and constrain these parameters concurrently with the model’s initial condition parameters. However, three-dimensional viscous hydrodynamic simulations require an order of magnitude more computing resources than boost-invariant hydrodynamic simulations. This makes it difficult to calibrate on statistically intensive observables such as the two-particle flow cumulants $v_n\{2\}$ which are generally needed to constrain the QGP medium parameters. In light of this issue, we calibrated the model using azimuthally integrated quantities, e.g. charged-particle densities $dN_{ch}/d\eta$, which are statistically cheap to compute.

At the time of this study, the vHLLE code did not implement terms in the hydrodynamic equations of motion which couple the shear correction $\pi^{\mu\nu}$ to the bulk correction $\Pi$. There were also a number of challenges which limited our ability to apply bulk viscous corrections to the fluid’s distribution function at particlization; e.g. see section 5.1.1. These issues precluded a realistic treatment of bulk viscosity, so we opted to turn the bulk viscosity completely off, i.e. we set $\zeta/s = 0$ for all calculations. For the QGP shear viscosity $\eta/s$, we assumed a constant (temperature-independent) value, hand tuned in the range 0.17–0.28.
for each calculation to compensate for our simplistic treatment of viscous corrections.

All other aspects of the model follow the specification of subsection 5.1.1: we converted the fluid into particles along a fixed hydro-to-micro switching isotherm $T_{\text{switch}} = 0.154$ GeV, we applied shear viscous corrections to the particle distribution function using equation (5.6), and we modeled the hadronic evolution below the switching temperature using the UrQMD microscopic transport model.

5.2.2 Parameter design and observables

The study estimated the joint posterior distribution for nine initial condition parameters. Five parameters modeled the entropy density at midrapidity:

1–2. two overall normalization factors for initial entropy deposition; one for Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV and one for p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV,

3. the generalized mean parameter $p$ for the scaling of initial entropy deposition at midrapidity,

4. the gamma shape parameter $k$ equal to the inverse variance of the random weights used to fluctuate each nucleon density, and

5. the Gaussian nucleon width $w$ to specify initial state granularity.

Meanwhile, another four parameters modeled its rapidity dependence:

6–8. three coefficients $\mu_0$, $\sigma_0$, and $\gamma_0$ to parametrize the local rapidity distribution’s mean, standard deviation, and skewness, and

9. one parameter $J$ to specify the pseudorapidity Jacobian.

We assigned each parameter the conservative range of prior values listed in table 5.7 and sampled $d = 100$ space filling parameter points within these ranges using a maximin Latin hypercube design. We then executed $4 \times 10^3$ minimum bias Pb-Pb events at $\sqrt{s_{NN}} = 2.76$ TeV and $10^4$ minimum bias p-Pb events at $\sqrt{s_{NN}} = 5.02$ TeV at each parameter point.
Each nuclear collision event was run on a coarse spacetime grid in ideal (non-viscous) mode to greatly reduce computational overhead. Finally, we divided our minimum-bias events into centrality bins according to the charged-particle multiplicity $N_{ch}$ in a given acceptance $\eta_{\text{min}} < \eta < \eta_{\text{max}}$. We used $|\eta| < 0.8$ for Pb-Pb collisions and $-4.9 < \eta < -3.1$ for $p$-Pb collisions, selected to match our calibration data.

Table 5.7  Input parameter ranges for the nuclear collision model.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Prior range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Norm Pb-Pb</td>
<td>Normalization factor $\sqrt{s_{NN}} = 2.76$ TeV</td>
<td>140–190</td>
</tr>
<tr>
<td>Norm $p$-Pb</td>
<td>Normalization factor $\sqrt{s_{NN}} = 5.02$ TeV</td>
<td>150–200</td>
</tr>
<tr>
<td>$p$</td>
<td>Generalized mean parameter</td>
<td>$-0.3$–$0.3$</td>
</tr>
<tr>
<td>$k$</td>
<td>Multiplicity fluct. shape</td>
<td>1.0–5.0</td>
</tr>
<tr>
<td>$w$ [fm]</td>
<td>Gaussian nucleon width</td>
<td>0.4–0.6</td>
</tr>
<tr>
<td>$\mu_0$</td>
<td>Rapidity shift mean coeff.</td>
<td>0.0–1.0</td>
</tr>
<tr>
<td>$\sigma_0$</td>
<td>Rapidity width std. coeff.</td>
<td>2.0–4.0</td>
</tr>
<tr>
<td>$\gamma_0$</td>
<td>Rapidity skewness coeff.</td>
<td>0.0–10.0 (rel)</td>
</tr>
<tr>
<td>$J$</td>
<td>Pseudorapidity Jacobian param.</td>
<td>0.6–0.9</td>
</tr>
</tbody>
</table>

Running the hydrodynamic simulations on a coarse spacetime grid in ideal mode precluded a meaningful comparison to anisotropic flow observables, so we calibrated the model using azimuthally integrated multiplicity observables which are less sensitive to the hydrodynamic viscosity and grid resolution. We calculated the rapidity dependence of:

- the charged-particle density $dN_{ch}/d\eta$ measured in various centrality bins for Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV and $p$-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV, and

- the root-mean-square of the two-particle pseudorapidity correlation coefficient $a_1$ for Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. I’ll define this observable momentarily.

The first observable is sensitive to the rapidity dependence of the ensemble-averaged entropy density profile, while the second is sensitive to its event-by-event rapidity fluctuations. Table 5.8 lists the experimental data used to calibrate each of these observables, taken from the ALICE [252, 253] and ATLAS collaborations [254, 255]. Let me explain now what I mean by the two-particle pseudorapidity correlation coefficient $a_1$.

Previously in subsection 2.2.2, I described how to project the event-by-event azimuthal
particle distribution $dN/d\phi$ onto a basis of orthogonal harmonics. In a similar fashion, we can project the event-by-event pseudorapidity distribution $dN/d\eta$ within the acceptance $[-Y,Y]$ onto a basis of orthonormal polynomials [258]

$$\frac{dN}{d\eta} = \langle \frac{dN}{d\eta} \rangle \left[ 1 + \sum_{n=0}^{\infty} a_n T_n \left( \frac{\eta}{Y} \right) \right], \quad (5.48)$$

using, for example, the normalized Legendre polynomials [255, 259]

$$T_n(x) = \sqrt{n + 1/2} P_n(x). \quad (5.49)$$

These normalized Legendre polynomials can then be used to expand the two-particle correlation function

$$C(\eta_1, \eta_2) = \frac{\langle N(\eta_1) N(\eta_2) \rangle}{\langle N(\eta_1) \rangle \langle N(\eta_2) \rangle}, \quad (5.50)$$

where $N(\eta) \equiv dN/d\eta$ is the multiplicity density at $\eta$, and $\langle N(\eta) \rangle$ is the average multiplicity at $\eta$ for a given event class.

Expressed as a sum of symmetrized Legendre polynomials, the two-particle correlation function $C(\eta_1, \eta_2)$ equals

$$C(\eta_1, \eta_2) = 1 + \sum_{m,n=0}^{\infty} \langle a_m a_n \rangle T_{mn}(\eta_1, \eta_2), \quad (5.51a)$$

$$T_{mn}(\eta_1, \eta_2) = \frac{T_m(\eta_1) T_n(\eta_2) + T_m(\eta_2) T_n(\eta_1)}{2}. \quad (5.51b)$$

**Table 5.8** Experimental data used to calibrate the model parameters.

<table>
<thead>
<tr>
<th>Collision system</th>
<th>Observable</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pb-Pb, 2.76 TeV</td>
<td>Charged-particle pseudorapidity dist. $dN_{ch}/d\eta$</td>
<td>[252, 253]</td>
</tr>
<tr>
<td></td>
<td>Two-particle pseudorapidity corr. $\sqrt{\langle \sigma_t^2 \rangle}$</td>
<td>[255]</td>
</tr>
<tr>
<td>$p$-Pb, 5.02 TeV</td>
<td>Charged-particle pseudorapidity dist. $dN_{ch}/d\eta$</td>
<td>[254]</td>
</tr>
</tbody>
</table>
Rearranging terms, this expansion can be written as

\[ C(\eta_1, \eta_2) = 1 + \frac{1}{2} \langle a_0 a_0 \rangle + \frac{1}{\sqrt{2}} \sum_{n=1}^{\infty} \langle a_0 a_n \rangle (T_n(\eta_1) + T_n(\eta_2)) \]

\[ + \sum_{m,n=1}^{\infty} \langle a_m a_n \rangle \frac{T_m(\eta_1)T_n(\eta_2) + T_m(\eta_2)T_n(\eta_1)}{2}. \quad (5.52) \]

The last term containing coefficients \( \langle a_m a_n \rangle \) with \( m, n > 0 \) describes the system’s dynamical rapidity fluctuations at fixed centrality. It can be isolated from the preceding terms by dividing the correlation function \( C(\eta_1, \eta_2) \) by its projection onto the \( \eta_1 \) and \( \eta_2 \) dimensions:

\[ C_N(\eta_1, \eta_2) = \frac{C(\eta_1, \eta_2)}{C_1(\eta_1)C_2(\eta_2)}, \quad (5.53a) \]

\[ C_{1,2}(\eta_1, \eta_2) = \frac{1}{2Y} \int_{-Y}^{Y} C(\eta_1, \eta_2) \, d\eta_{1,2}. \quad (5.53b) \]

This rescaling sends the \( \langle a_0 a_0 \rangle \) and \( \langle a_0 a_n \rangle \) coefficients to zero, yielding

\[ C_N(\eta_1, \eta_2) = 1 + \sum_{m,n=1}^{\infty} \langle a_m a_n \rangle \frac{T_m(\eta_1)T_n(\eta_2) + T_m(\eta_2)T_n(\eta_1)}{2}. \quad (5.54) \]

Finally, the terms \( \langle a_m a_n \rangle \) with \( m, n > 0 \) are projected out of the renormalized correlation function using the expansion

\[ C_N(\eta_1, \eta_2) = 1 + \frac{3}{2} \langle a_1^2 \rangle \frac{\eta_1 \eta_2}{Y^2} + \ldots. \quad (5.55) \]

Each coefficient \( \langle a_m a_n \rangle \) receives contributions from long-range correlations, such as those introduced by initial state rapidity fluctuations, as well as short-range correlations, such as those caused by resonance decays. Recent work has shown that short-range correlations from resonance decays are a significant contribution to the \( \langle a_m a_n \rangle \) signal, while the QGP viscosity has a much smaller effect [260]. Generally speaking, one can either model the short-range correlations or subtract their effect from both the model and the experimental data [261].
The hadronic transport model UrQMD properly accounts for short-range correlations, so we opted to compare to experimental data with the short-range correlations included [255].

5.2.3 Bayesian parameter estimation

Bayesian parameter estimation was then applied to estimate the values of the nine initial condition parameters listed in table 5.7 using the predictions of our nuclear collision model and the experimental data listed in table 5.8. We followed the procedure described in section 4.3 and trained an emulator to reproduce the model’s input-output mapping using the model calculations at each design point. This emulator was trained to reproduce the first \( q = 6 \) principal components (PCs) of the \( dN_{ch}/d\eta \) data, and the first \( q = 4 \) PCs of the rms \( a_1 \) data. Collectively, these PCs explain 99.5% of the observed variance across the computer experiment design.

We constructed the Bayesian likelihood from the product of three terms

\[
P(E|x) = P(E^{p-Pb}_{dN_{ch}/d\eta}|x) \cdot P(E^{Pb-Pb}_{dN_{ch}/d\eta}|x) \cdot P(E^{Pb-Pb}_{rms a_1}|x),
\]

(5.56)

where each term quantifies the experimental evidence provided by a certain class of observable. The first two terms on the right quantify the evidence provided by the \( p-Pb \) and \( Pb-Pb \) \( dN_{ch}/d\eta \) data, while the last term quantifies the evidence provided by the \( Pb-Pb \) rms \( a_1 \) observable.

Following reference [178], we calculated each likelihood function \( P(E|x) \) in the PCA transformed observable basis

\[
P(E|x) \propto \exp \left\{ -\frac{1}{2} [z_m(x) - z_e]^T \Sigma_z^{-1} [z_m(x) - z_e] \right\},
\]

(5.57)

where \( z_m(x) \) is the vector of PCA transformed model (emulator) observables, \( z_e \) is the vector of PCA transformed experimental observables, and \( \Sigma_z \) is their total covariance matrix. We used a covariance matrix in the PC space proportional to the identity matrix \( \Sigma_z = \sigma I \), corresponding to 5%, 10%, and 20% relative error on the total variance of the \( p-Pb \) \( dN_{ch}/d\eta \),
Pb-Pb $dN_{\text{ch}}/d\eta$, and Pb-Pb rms $a_1$ observables. This was a rough approximation, similar to reference [178], designed to conservatively account for various sources of uncertainty in the experimental data, model calculations, and emulator predictions. We elected to give more weight to charged-particle yields $dN_{\text{ch}}/d\eta$ relative to the rms $a_1$ coefficient because they measure the single-particle distribution which is more fundamental than the two-particle distribution. We also gave more weight to the $p$-Pb yields relative to the Pb-Pb yields, since they are more sensitive to the asymmetry parameters of the model.

In this study, we placed an informative prior $P(x)$ on the generalized mean parameter $p$, equal to the posterior distribution obtained from reference [178]. For all other model parameters, we used a uniform prior which was constant within the design ranges listed in table 5.7 and zero outside them. The Bayesian posterior distribution was then calculated from the likelihood and the prior using equation (4.51). Finally, we used the affine-invariant MCMC sampler emcee to draw samples from the Bayesian posterior distribution. We ran $O(10^5)$ burn-in steps to allow the chain to equilibrate followed by $O(10^6)$ production steps.

5.2.4 Posterior parameter distribution

Figure 5.20 shows the Bayesian posterior distribution for the relative skewness model (blue lower triangle) and absolute skewness model (red upper triangle) parametrized by table 5.6. I’ll start, as before, by describing the marginal distributions shown on the figure diagonal, starting in the upper-left corner and working my way to the bottom-right. As I traverse these marginal distributions, I’ll also comment on some of the interesting correlations contained in the joint posterior distributions visualized by the off-diagonal elements.

The posterior on the Pb-Pb normalization $N_{\text{Pb-Pb}}$ is broad for both the relative and absolute skewness models. Evidently, we should have increased the upper bound on the $N_{\text{Pb-Pb}}$ prior since its posterior is clipped. Unsurprisingly, the Pb-Pb normalization is strongly correlated with the $p$-Pb normalization. The $p$-Pb normalization $N_{p-Pb}$, meanwhile, peaks in the middle of the design range, with the absolute skewness model preferring somewhat larger values than the relative skewness model.
Figure 5.20 Posterior distribution for the model input parameters of Ref. [157]. Diagonal panels are the marginal distributions of individual parameters and off-diagonal panels are the joint distributions for pairs of parameters. The blue lower triangle is the posterior for the relative skewness model, and the red upper triangle is the posterior for the absolute skewness model.
Figure 5.21  Left and middle: Pseudorapidity dependence of the charged-particle density $dN_{ch}/d\eta$ for Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV (left) and $p$-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV (middle). Colored bands cover the model emulator’s mean posterior prediction $\pm$2 standard deviations, and black symbols are experimental data from ALICE [252, 253] and ATLAS [254]. Right: Two-particle pseudorapidity correlations quantified by $\sqrt{\langle a_1^2 \rangle}$ plotted versus collision centrality. Colored bands are same as before; black symbols are data from ATLAS [255]. The colored lines, meanwhile, are results from full event-by-event viscous hybrid model simulations using selected parameters from the Bayesian posterior.

There’s not much to say about the generalized mean parameter $p$ that has not already been said. Recall that this parameter controlled the scaling of initial entropy deposition as a function of nuclear thickness at midrapidity (3.58). We assigned this parameter an informative prior peaked near zero using the results of reference [178], so the peak on our posterior is not particularly informative; it merely reflects our chosen prior.

Interestingly, the nucleon fluctuation parameter $k$ is tightly constrained. This parameter specifies the inverse variance of the gamma random variables used to fluctuate each nucleon’s contribution to the participant thickness function (3.47). Hence, small $k$ values enhance local nuclear density fluctuations, while large $k$ values suppress them. Both the relative and absolute skewness models prefer $k \sim 2$, although it is unclear what drives this constraint.

The posterior on the Gaussian nucleon width $w$, meanwhile, is less clear. The relative skewness model prefers significantly larger nucleons than the absolute skewness model. The Gaussian nucleon width strongly affects the average asymmetry of nuclear density fluctuations, so it is not surprising that this parameter correlates with the skewness of the rapidity distribution. This suggests that one should be careful when interpreting the posterior on the Gaussian nucleon width as it is highly sensitive to our modeling assumptions.
The next three parameters $\mu_0$, $\sigma_0$, and $\gamma_0$ parametrize the cumulants of the rapidity distribution as functions of $\tilde{T}_A, \tilde{T}_B$. Their precise values are not particularly meaningful, but we do observe some interesting features. For example, the relative skewness model prefers a small rapidity shift, while the absolute skewness model prefers a large rapidity shift close to the nucleon center-of-mass rapidity $y_{cm}$. We also observe a strong correlation between the width of the rapidity distribution $\sigma_0$ and the value of the Jacobian parameter $J$. Finally, we observe that the relative and absolute skewness models prefer sharply peaked distributions for $\gamma_0$. It’s not interesting that these distributions peak in different locations; each model’s parametrization of the skewness is different. However, it is interesting that the skewness is tightly constrained for both parametrizations. Evidently, the data appears to prefer rapidity profiles with nonzero skewness.

Figure 5.21, meanwhile, compares both models (colored bands) to the experimental calibration data (black symbols) using parameters sampled from the Bayesian posterior distribution. Each colored band covers the model emulator’s mean posterior prediction $\pm 2$ standard deviations. First, direct your attention to the left two panels which show the Pb-Pb and $p$-Pb charged-particle densities $dN_{ch}/d\eta$. Both models nicely describe the centrality and rapidity dependence of $dN_{ch}/d\eta$, illustrating the flexibility of the cumulant generating function. Now, look at the far right panel which shows the centrality dependence of two-particle pseudorapidity correlations, quantified by the $\sqrt{\langle a_1^2 \rangle}$ observable. Both models describe this observable to 20% accuracy in the 0–50% centrality range, but they underestimate its value at larger centralities. Recently, it was shown that the microscopic model HIJING [262] reproduces rms $a_1$ for $N_{part} < 80$, while it overestimates its value for more central collisions [255]. This suggests that microscopic and hadronic models are complementary in understanding the longitudinal rapidity fluctuations.

5.2.5 Verification of high-probability parameters

This subsection evaluates our three-dimensional hybrid model’s “best fit” predictions for new observables not included in the calibration process, using maximum a posteriori (MAP)
parameters (table 5.9) selected from the mode of the posterior distribution. In the parlance of machine learning, the following predictions are said to be out of sample. These verification tests will help us assess the hidden assumptions of our model framework.

5.2.5.1 Anisotropic flow cumulants

First, we checked the centrality and rapidity dependence of the model’s two-particle flow cumulants $v_n\{k\}$, calculated using the Q-cumulant method described in [263], which are highly sensitive to the initial distribution of matter in the $(x, y)$ plane. Fitting the anisotropic flow cumulants is therefore a sensitive test of the QGP initial condition geometry.

Previously, in section 5.1, I showed that the boost-invariant T\(\text{R}^{\text{ENT}}\)o model nicely describes the two-particle flow cumulants $v_n\{2\}$ for $n = 2, 3,$ and 4 at midrapidity $|\eta| < 0.8$. Our rapidity-dependent initial condition model includes the boost-invariant T\(\text{R}^{\text{ENT}}\)o model as a specific subcase, so the present analysis should also describe the experimentally measured two-particle flow cumulants at midrapidity provided suitably chosen parameters.

Figure 5.22 shows the present study’s three-dimensional hybrid model predictions (colored lines) for the two-particle flow cumulants $v_2\{2\}$ and $v_3\{2\}$ at midrapidity $|\eta| < 0.8$ plotted as a function of centrality for Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV compared to experimental data (black symbols) from ALICE [108]. As expected, our three-dimensional model well describes the experimental data at midrapidity in agreement with similar model results based on boost-invariant approximations [178].
Figure 5.22 Anisotropic flow cumulants $v_2\{2\}$ and $v_3\{2\}$ at midrapidity $|\eta| < 0.8$ plotted as a function of collision centrality for Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. Colored lines are three-dimensional viscous hybrid model calculations using the relative (solid) and absolute (dashed) skewness models for initial entropy deposition, and symbols are experimental data from ALICE [108].

Now that we are confident that our three-dimensional model reproduces the results of our boost-invariant model at midrapidity, we can leverage the posterior constraints of the present analysis to predict each observable’s rapidity dependence. Figure 5.23 shows the rapidity dependence of the flow cumulants $v_2\{2\}$, $v_3\{2\}$, and $v_2\{4\}$ calculated in various centrality bins using the same particle selection criteria applied by the experiment. The colored lines are the calculations of our three-dimensional hybrid model, and the black symbols are experimental data from ALICE [264].

The flow cumulants peak at midrapidity and decrease at forward/backward rapidity, producing a triangular shape similar to the ALICE data. Generally speaking, the model describes the data quite well, although the description is somewhat worse at large rapidities where the model tends to over predict the magnitude of each flow harmonic. Recently, it was shown that the pseudorapidity dependence of $v_n(\eta)$ is highly sensitive to the hadronic shear viscosity [260], so figure 5.23 corroborates the effective shear viscosity determined by UrQMD transport dynamics.

It’s also worth noting that we have neglected the effects of nonzero baryon density. For example, the initial conditions impart a rapidity-dependent baryon current which has been
shown to affect final state observables [121, 265]. The QCD equation of state also varies as a function of baryochemical potential which we have completely omitted [31–33, 265–268]. Presumably, these effects would improve our model’s description of the data at large rapidities where baryon density effects are most important.

5.2.5.2 Event-plane decorrelation

Recall that the anisotropic flow harmonics \( v_n \) describe the modulation of the azimuthal single-particle distribution [75–77]

\[
\frac{dN}{d\phi} \propto 1 + 2 \sum_{n=1}^{\infty} v_n \cos[n(\phi - \Psi_n)],
\]

(5.58)
relative to the phase

$$\Psi_n = \frac{1}{n} \arctan2(\langle \sin n\phi \rangle, \langle \cos n\phi \rangle),$$

commonly known as the event-plane angle. In general, the event-plane angle $\Psi_n$ may change as a function of rapidity $\eta$ in a single event due to longitudinal rapidity fluctuations and finite particle effects. Consequently, the event-plane angles $\Psi_n(\eta_1 \pm \delta\eta)$ and $\Psi_n(\eta_2 \pm \delta\eta)$ are expected to decorrelate as the gap between $\eta_1$ and $\eta_2$ increases; e.g. see figure 5.24. This effect produces energy and entropy density profiles that are twisted or “torqued” along the pseudorapidity direction [158].

One method to study the single-particle anisotropic flow harmonics $v_n$ experimentally is to measure two-particle azimuthal correlations [75–77]

$$\frac{dN_{\text{pair}}}{d\Delta\phi} \propto 1 + 2 \sum_{n=1}^{\infty} V_{n\Delta} \cos(n\Delta\phi),$$

where $\Delta\phi = \phi_a - \phi_b$ is the relative angle between each particle pair ($a$ and $b$) in the event. If the particles are independently correlated with a common source, e.g. a hydrodynamic flow field, then the two-particle distribution factorizes into a product of single-particle distributions, so that

$$V_{n\Delta} = v_n^a v_n^b.$$ 

If, however, the particles $a$ and $b$ sample different underlying particle distributions, e.g. different flow fields, then the factorization breaks. For example, if the fireball is torqued,
then particles emitted with different rapidities $\eta^a \neq \eta^b$ will sample different flow fields so that $V_n(\Delta) \neq v_n^a v_n^b$.

The CMS collaboration measured the pseudorapidity dependence of this event-plane decorrelation effect in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV using the $\eta$-dependent factorization ratio $r_n$ [270], defined as

$$r_n(\eta^a, \eta^b) \equiv \frac{V_n(\Delta)(-\eta^a, \eta^b)}{V_n(\Delta)(\eta^a, \eta^b)},$$

$$V_n(\eta^a, \eta^b) = \langle \langle \cos(n \Delta \phi) \rangle \rangle,$$

where the inner average means averaging over all particle pairs in a given event, and the outer average means averaging over all events in a given centrality class. Here, three rapidity bins, $\pm \eta^a$ and $\eta^b$, are used to remove the contamination of short-range jet-like two-particle correlations. The resulting $\eta$-dependent factorization ratio $r_n(\eta^a, \eta^b)$ equals unity if the factorization holds, and is expected to be smaller than unity in the presence of rapidity fluctuations.

Figure 5.25 shows our calibrated model calculations for the factorization ratio $r_n(\eta^a, \eta^b)$ plotted as a function of $\eta^a$ using $0.3 < p_T^a < 3.0$ GeV and $p_T^b > 0$ GeV, with two different rapidity cuts for $\eta^b$ selected to match the experiment: $3.0 < \eta^b < 4.0$ (left figure) and $4.4 < \eta^b < 5.0$ (right figure). The factorization ratios are shown for the second harmonic $n = 2$ (blue lines), the third harmonic $n = 3$ (orange lines) and for six centrality classes (figure panels). We show calculations for both the relative skewness model (solid lines) and absolute skewness model (dashed lines).

The model reproduces the general shape and approximate magnitude of the factorization breaking $r_n(\eta^a, \eta^b)$ for $n = 2, 3$ plotted as a function of the pseudorapidity $\eta^a$ for multiple centrality bins. It correctly describes, for instance, the hierarchy of $r_2$ and $r_3$, and it reproduces the magnitude of the factorization breaking for non-central events provided $3.0 < \eta^b < 4.0$. However, the agreement with the data also leaves room for improvement. The model fails to describe the factorization breaking for 0–5% centrality events, and it
over predicts the strength of the factorization breaking when one of the particles is selected from $4.4 < |\eta^b| < 5.0$, i.e. far from midrapidity.

Needless to say, some discrepancy at large rapidities is to be expected. Our model is constructed to extrapolate well-developed initial conditions at midrapidity to nonzero rapidity using the constraints of azimuthally integrated multiplicity observables. Naturally, this extrapolation gradually loses its predictive power for fine-structure flow observables as one moves far from midrapidity. The observed tension indicates that our extrapolation is breaking down. Clearly, future improvements to the model are needed at large rapidities.

**5.2.5.3 Symmetric cumulants**

As a final test of the calibrated model, we investigated the correlations between anisotropic flow harmonics of different order which have been shown to provide additional constraints
on the QGP initial conditions [95]. Experimentally, these correlations can be quantified by the symmetric cumulants $SC(m, n)$ [271], defined as

$$SC(m, n) = \langle \langle \cos[m(\phi_1 - \phi_3) + n(\phi_2 - \phi_4)] \rangle \rangle - \langle \langle \cos[m(\phi_1 - \phi_2)] \langle \cos[n(\phi_1 - \phi_2)] \rangle \rangle \rangle$$

$$= \langle v^2_m v^2_n \rangle - \langle v^2_m \rangle \langle v^2_n \rangle,$$

(5.63)

where the double average means averaging over particles in each event and then averaging over all events in a given centrality class. This quantity is positive if $v_m$ and $v_n$ are correlated, zero if they are uncorrelated, and negative if they are anti-correlated. We also calculated the normalized symmetric cumulants

$$NSC(m, n) = \frac{SC(m, n)}{\langle v^2_m \rangle \langle v^2_n \rangle},$$

(5.64)

which divide out the magnitudes of $\langle v^2_m \rangle$ and $\langle v^2_n \rangle$. Previous studies show that the $NSC(3, 2)$ observable is sensitive mainly to the initial conditions, while $NSC(4, 2)$ is sensitive to both the initial conditions and QGP medium properties [95, 272].

Figure 5.26 shows the symmetric cumulants $SC(m, n)$ (top row) and normalized symmetric cumulants $NSC(m, n)$ (bottom row) calculated for the Pb-Pb system at $\sqrt{s_{NN}} = 2.76$ TeV using the relative skewness model (left column) and absolute skewness model (right column). The blue lines/bands are $(m, n) = (4, 2)$ and the green lines/bands are $(m, n) = (3, 2)$. In addition, we also calculated each observable using two different sets of kinematic cuts (solid and dashed lines) described below.

First, we calculated each observable at midrapidity using the same kinematic cut $|\eta| < 0.8$ used by the ALICE experiment [272]. These model calculations are the solid lines in figure 5.26. They nicely reproduce the centrality dependence of the symmetric and normalized symmetric cumulants measured by ALICE (black symbols), although we underpredict the magnitudes of the measured symmetric cumulants at large centralities. Curiously, we obtained a worse description of this observable in reference [221], although
it is unclear why. Determining the culprit which explains this discrepancy would be an interesting project for future work.

Next, we calculated each observable using two reference particles sampled from $|\eta| < 0.8$ and two particles of interest sampled from $2.5 < |\eta| < 3.5$. This new observable quantifies correlations between different flow harmonics separated by a large rapidity gap. Therefore we expect it to be sensitive to the rapidity dependence of the initial conditions. The dashed lines in figure 5.26 show our predictions for this novel observable which has not yet been measured. Selecting the particles of interest from $2.5 < |\eta| < 3.5$ suppresses the magnitude of the symmetric cumulants at large centralities, but it has no meaningful effect on the normalized symmetric cumulants. The invariance of the normalized symmetric cumulant to

![Figure 5.26](image_url)
the rapidity of the particles of interest is therefore a testable prediction which can be used to validate the assumptions of our framework.

5.3 Bulk matter in small collision systems

Clearly, hydrodynamics well describes heavy-ion collisions at ultrarelativistic energies, but what about smaller collision systems, e.g. $p$-$p$ and $p$-Pb collisions, which generate nucleon sized QGP droplets and fewer particles per unit rapidity? Do these small fireballs expand hydrodynamically as well? This section presents the results of an exploratory study of hydrodynamic behavior in small collision systems using a modified version of the TRENTo initial condition model described in section 3.5. The following text and figures appear in one of my publications,

J. S. Moreland, J. E. Bernhard, and S. A. Bass, “Bayesian calibration of a hybrid nuclear collision model using $p$-Pb and Pb-Pb data from the LHC”, Submitted for publication, 1808.02106 [nucl-th],

which I’ve lightly edited to fit the format and layout of this dissertation.

Relativistic heavy-ion collisions produce long-range multiparticle correlations which are commonly explained by the existence of hydrodynamic flow [36]. 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 [86, 114, 178]. 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 always 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.

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 [273–275]. 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 $\sqrt{s_{\text{NN}}} = 5.02$ 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 experimental data.

### 5.3.1 Boost-invariant model with nucleon substructure

This study used an updated version of our nuclear collision model largely identical to the model which I described in subsection 5.1.6. It consisted of:

- TRENTo initial conditions,
- pre-equilibrium free streaming and Landau matching,
- boost-invariant viscous hydrodynamics with shear and bulk coupling,
- a modern lattice-based QCD equation of state,
- Cooper-Frye particlization with shear and bulk viscous corrections using the FRZOUT sampler, and
- UrQMD to simulate microscopic hadronic interactions below the pre-specified hydro-to-micro switching temperature.
The study also made one important addition to the nuclear collision model. We replaced Gaussian nucleons in the TRENTo model with “lumpy” nucleons characterized by several free parameters to vary the fluctuating size and shape of constituent degrees of freedom inside each nucleon. I’ll explain the specifics of this nucleon substructure extension shortly. First, let me motivate why this additional physics is relevant in the first place.

5.3.1.1 Case for nucleon substructure

IP-Glasma [85] is one of the most theoretically sound and phenomenologically successful models for the initial conditions of ultrarelativistic nuclear collisions. The model describes, for instance, numerous bulk observables in heavy-ion collisions at RHIC and LHC energies [276]. Thus, when it was first observed that high-multiplicity $p$-Pb collisions generate flow-like signatures—not definitive proof of flow, but tantalizing indicators—it was natural to check the predictions of hydrodynamic simulations using IP-Glasma initial conditions against said measurements.

Schenke and Venugopolan performed this test in reference [277] using the MUSIC viscous hydrodynamics code [119, 278, 279]. While IP-Glasma coupled to MUSIC well reproduces numerous heavy-ion bulk observables, they found that the model significantly underpredicts the elliptic flow $v_2$ observed in $p$-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV as shown in figure 5.27. This observed tension could mean one of several things:

1. Hydrodynamics is the correct framework to study $p$-Pb collisions, but IP-Glasma is flawed.

2. IP-Glasma is essentially correct, and the application of hydrodynamics to $p$-Pb collisions is flawed. Perhaps some additional non-hydrodynamic correlations are required to describe the data.

3. IP-Glasma coupled to hydrodynamics is a sensible framework to study $p$-Pb collisions, but the IP-Glasma model is incomplete.

Confronted with this list of possibilities, the authors suggested that eccentric nucle-
ons might offer an attractive solution to the problem. In its original formulation, the IP-Glasma model uses round, spherically symmetric nucleons. This symmetry produces roundish blobs of initial energy density in $p$-$p$ and $p$-$Pb$ collisions [196] and hence small initial state eccentricities. Conversely, deformed nucleons generate significantly larger $p$-$p$ and $p$-$Pb$ eccentricities, enhancing the anisotropic flow predicted for small systems. Since Pb-$Pb$ collisions are only weakly sensitive to nucleon substructure [281], this solution could potentially resolve the $p$-$Pb$ discrepancy without spoiling the model’s simultaneous description of heavy-ion data. Indeed, recent studies of the IP-Glasma initial condition model using nucleon substructure show promising descriptions of $p$-$Pb$ bulk observables [282] and work is on-going to extend these comparisons to heavy-ion observables which were previously fit by IP-Glasma model calculations without nucleon substructure [283].

Much like IP-Glasma, the calibrated T\RENTTo model predicts roundish blobs of energy (or entropy) in $p$-$p$ and $p$-$Pb$ collisions. Indeed, when the generalized mean parameter $p = 0$, the T\RENTTo model predicts energy (or entropy) density profiles for $p$-$p$ collisions that are exactly Gaussian; see figure 3.22. Certainly, such similarities are to be expected if the
bottom-up approach of IP-Glasma and the top-down approach of TREnto are describing the same underlying physics. Therefore, it stands to reason that TREnto will also need nucleon substructure if our hydrodynamic model is to have any chance of describing $p$-$p$ and Pb-Pb bulk observables simultaneously.

5.3.1.2 Nucleon substructure in the literature

Unlike heavy-ions for which there exists an established procedure to randomly sample nuclear density configurations (see section 3.2), the nucleon’s fluctuating structure is poorly understood from first principles and difficult to measure experimentally. Nucleon substructure implementations in the literature have thus often employed simplistic models, replacing round protons with composite protons described by a few salient model parameters, in order to investigate the effect of each model parameter on simulated observables [284–288]. Sensitivity studies such as these have been able to identify cause and effect relationships between model inputs and outputs, but they lack the ability to constrain nucleon substructure parameters in any kind of global or systematic fashion. Multiple nucleon substructure descriptions appear to be qualitatively consistent with the usual list of hydrodynamic bulk observables (yields, mean $p_T$’s and flows), so new observables are needed to discriminate mutually incompatible model assumptions.

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 [289]. It was later realized that this depression, or so-called hollowness effect in the $p$-$p$ inelastic collision profile [290], 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 [291, 292].

Independently, studies of coherent and incoherent $J/\psi$ 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 [293–296]. Initial condition studies using the IP-Glasma model of Color Glass Condensate effective field theory [85] 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 [297]. In addition, it was recently shown that hydrodynamic simulations using IP-Glasma initial conditions with color charge fluctuations calibrated to fit coherent and incoherent \( J/\psi \) diffraction measured by the H1 and Zeus experiments at HERA [295, 296] provide a good description of collectivity in small and large collision systems [283].

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 important 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 experimental measurements in a statistically rigorous fashion.

5.3.1.3 Parametric nucleon substructure

From a Bayesian perspective, there is nothing unique about nucleon substructure uncertainty compared to the other sources of initial condition uncertainty already parametrized by the TRENTo model. Indeed, the study which I will discuss now is essentially identical to that of reference [221] with one significant change to the nuclear collision model. We replaced one-parameter round nucleons with three-parameter lumpy nucleons.

Specifically, we modeled each lumpy nucleon density as a sum of \( n_c \) independent constituent densities:

\[
\rho_n(x) = \frac{1}{n_c} \sum_{i=1}^{n_c} \gamma_i \rho_c(x - x_i),
\]

(5.65)

where \( \gamma_i \) is a random weight sampled from a gamma distribution with unit mean and
Figure 5.28  Thickness function $T(x_{⊥}) = \int dz \rho_n(x_{⊥}, z)$ for three example nucleon densities $\rho_n$ generated by the parametric substructure model. These example nucleons show a small subset of all possible variability.

variance $1/k$ as before, and $\rho_c$ is a Gaussian constituent density

$$\rho_c(x) = \frac{1}{(2\pi v^2)^{3/2}} \exp \left( -\frac{|x|^2}{2v^2} \right)$$  \hspace{1cm} (5.66)

of variable width $v$. The constituent positions $x_i$ were sampled independently (ignoring correlations) from a Gaussian radial distribution

$$\rho_{\text{pos}}(x_i; x_n) = \frac{1}{(2\pi r^2)^{3/2}} \exp \left( -\frac{|x_i - x_n|^2}{2r^2} \right),$$ \hspace{1cm} (5.67)

where $x_n$ is the position of each nucleon, and $r$ is a parameter which varies the sampling radius of the constituent positions about each nucleon position. This allows the model to generate a diverse range of initial nucleon shapes as shown in figure 5.28.

As a matter of convenience, we sampled the nucleon positions before determining the constituent positions. This creates small discrepancies between the designated nucleon positions $x_n$ and the actual position of each nucleon’s center-of-mass, owing to fluctuations in the constituent positions $x_i$. The sampling radius $r$ should thus be interpreted with care. It is not the Gaussian width of the sampled nucleons in their center-of-mass frame.

Once each lumpy nucleon is sampled, the nucleon density is treated as a singular field $\rho_n$. The constituents are merely used as a mechanism to add deformity and fluctuations to this field. This distinguishes the T\textsc{rento} model from the wounded quark model which calculates individual quark participants, depositing for each wounded quark a fixed amount of energy or entropy [287, 298, 299]. Protons and neutrons are bound by the strong force,
Figure 5.29  Effect of nucleon substructure on the nuclear thickness function $T(x_\perp) = \int dz \rho(x_\perp, z)$ 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.

so I’d argue that there is no sensible concept of spectator quarks like there is for spectator nucleons. If one constituent collides inelastically, the entire nucleon is ripped apart and contributes inelastically.

The TRENTo model essentially treats nucleus-nucleus collisions as a superposition of individual proton-proton collisions so nucleon substructure modifies the overall thickness function of a macroscopic nucleus as well. Figure 5.29 shows the effect of nucleon substructure on the thickness function of a generic lead nucleus. Additional fluctuations emerge over the length scale of a nucleon, but the macroscopic geometry of the nucleus is largely unchanged. We therefore expect nucleon substructure to have a small effect on heavy-ion collisions, but it is impossible to know for certain without running the calculations directly.

5.3.2 Parameter design and observables

This study estimated the joint posterior distribution for fifteen model parameters. Seven parameters varied the initial conditions modeled by TRENTo:

1. the overall normalization factor for initial energy deposition (same for p-Pb and Pb-Pb),

2. the generalized mean parameter $p$ for the scaling of initial energy deposition as a function of nuclear thickness,
3. the number of nucleon constituents $n_c$, 
4. the sampling radius for the nucleon constituent positions $r$, 
5. the nucleon constituents’ width $v$, 
6. the nucleons’ effective fluctuation standard deviation $\sigma_{\text{fluct}} = (n_c k)^{-1/2}$, where $k$ is the inverse variance of the gamma random variable used to fluctuate each constituent density, and 
7. the cube of the minimum inter-nucleon distance $d_{\text{min}}^3$. 

Meanwhile, another eight parameters varied the properties of the hybrid model simulation: 

8. the pre-equilibrium free streaming time $\tau_{fs}$, 
9–11. three parameters ($\eta/s$ min, slope, and curvature) for the temperature dependence of the QGP shear viscosity, 
12–14. another three parameters ($\zeta/s$ max, width, and location) for the temperature dependence of the QGP bulk viscosity, and 
15. a particlization temperature $T_{\text{switch}}$ that defined the isotherm for Cooper-Frye particle emission. 

We assigned each parameter the conservative range of prior values listed in table 5.10 and used Latin hypercube sampling to distribute 500 design points uniformly throughout the fifteen dimensional parameter space. The selected parameter ranges were chosen to be intentionally wide to avoid clipping the calibrated posterior. For example, even though references [178, 221] found $p \sim 0$, we decided to 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 $n_c$ which we limited 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
Table 5.10 Input parameter ranges for the nuclear collision model.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Norm</td>
<td>Normalization factor</td>
<td>9–28 GeV</td>
</tr>
<tr>
<td>p</td>
<td>Energy deposition parameter</td>
<td>−1 to +1</td>
</tr>
<tr>
<td>σ_{fluct}</td>
<td>Nucleon fluctuation std. dev.</td>
<td>0–2</td>
</tr>
<tr>
<td>r</td>
<td>Constituent sampling radius</td>
<td>0–1.2 fm</td>
</tr>
<tr>
<td>n_{c}</td>
<td>Number of nucleon constituents</td>
<td>1–9</td>
</tr>
<tr>
<td>v</td>
<td>Constituent width</td>
<td>0.2–1.2 fm</td>
</tr>
<tr>
<td>d_{3}^{\text{min}}</td>
<td>Minimum inter-nucleon distance</td>
<td>0–4.9 fm³</td>
</tr>
<tr>
<td>τ_{fs}</td>
<td>Free streaming time</td>
<td>0.1–1.5 fm/c</td>
</tr>
<tr>
<td>(η/s)_{\text{min}}</td>
<td>Minimum value of η/s (at T_{c})</td>
<td>0–0.2</td>
</tr>
<tr>
<td>(η/s)_{\text{slope}}</td>
<td>Slope of η/s above T_{c}</td>
<td>0–8 GeV⁻¹</td>
</tr>
<tr>
<td>(η/s)_{\text{crv}}</td>
<td>Curvature of η/s above T_{c}</td>
<td>−1 to +1</td>
</tr>
<tr>
<td>(ζ/s)_{\text{max}}</td>
<td>Maximum value of ζ/s</td>
<td>0–0.1</td>
</tr>
<tr>
<td>(ζ/s)_{\text{width}}</td>
<td>Width of ζ/s peak</td>
<td>0–0.1 GeV</td>
</tr>
<tr>
<td>(ζ/s)_{T_0}</td>
<td>Temperature of ζ/s maximum</td>
<td>150–200 MeV</td>
</tr>
<tr>
<td>T_{\text{switch}}</td>
<td>Switching/particlization temp.</td>
<td>135–165 MeV</td>
</tr>
</tbody>
</table>

Variance must increase as \( n_{c} \) increases. Eventually, these fluctuations become unreasonably large. We found that for \( n_{c} < 10 \), the energy density fluctuations are reasonable, and hence we limited \( n_{c} \) to this prior range.

We executed \( \mathcal{O}(10^4) \) minimum-bias and multiplicity-triggered \( p\)-Pb and Pb-Pb events at \( \sqrt{s_{NN}} = 5.02 \) TeV at each design point, and calculated the model predictions for a number of experimental observables at midrapidity:

- the charged-particle density \( dN_{\text{ch}}/d\eta \) \([56, 300]\) and identified-particle densities \( dN/dy \) of pions, kaons, and protons,
- the transverse energy production \( dE_T/d\eta \),
- the mean transverse momentum \( \langle p_T \rangle \) for charged-particles, pions, kaons, and protons \([301]\),
- the mean transverse momentum fluctuations \( \delta p_T/\langle p_T \rangle \) (defined shortly),
- the two-particle flow cumulants \( v_n\{2\} \) for \( n = 2, 3, 4 \) and the four-particle flow cumulant \( v_2\{4\} \) for charged-particles \([248, 280]\), and
- the symmetric cumulants \( \text{SC}(4, 2) \) and \( \text{SC}(3, 2) \).

We matched the kinematic cuts of each observable to experiment with two exceptions: we
used a larger rapidity interval $|\eta| < 0.8$ for some boost-invariant observables to improve our finite particle statistics, and we did 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 dissertation writing, many of the aforementioned experimental observables are not yet available for $p$-Pb and Pb-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. Therefore, we restricted our calibration to the subset of measured and published observables listed in table 5.11. Notably absent from this list are the four-particle cumulants $v_n(4)$ at $\sqrt{s_{NN}} = 5.02$ 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. Therefore we refrained from calibrating on the four-particle cumulants, although I’ll show calculations of the four-particle cumulant $v_2(4)$ later in this text, using a single set of calibration parameters.

Most of the calibration observables listed in table 5.11 are presented 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. To calculate these observables, we generated $\mathcal{O}(10^4)$ minimum-bias events at each design point and divided 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 $p_T$ [301] and flow cumulants $v_n(k)$

<table>
<thead>
<tr>
<th>Collision system</th>
<th>Observable</th>
<th>Rapidity cut</th>
<th>Momentum cut</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pb-Pb, 5.02 TeV</td>
<td>Yield $dN_{ch}/d\eta$</td>
<td>$</td>
<td>\eta</td>
<td>&lt; 0.5$</td>
</tr>
<tr>
<td></td>
<td>Flow cumulants $v_n(2)$, $n = 2, 3, 4$</td>
<td>$</td>
<td>\eta</td>
<td>&lt; 0.8$</td>
</tr>
<tr>
<td>p-Pb, 5.02 TeV</td>
<td>Yield $dN_{ch}/d\eta$</td>
<td>$</td>
<td>\eta</td>
<td>&lt; 1.4$</td>
</tr>
<tr>
<td></td>
<td>Mean transverse momentum $p_T$</td>
<td>$</td>
<td>\eta</td>
<td>&lt; 0.3$</td>
</tr>
<tr>
<td></td>
<td>Flow cumulants $v_n(2)$, $n = 2, 3$</td>
<td>$</td>
<td>\eta</td>
<td>&lt; 2.4$</td>
</tr>
</tbody>
</table>
[280], the experiments used a special high-multiplicity trigger to select rare, ultra-central events according to the number of charged-particles produced $N_{\text{ch}}$ or detector tracks offline $N_{\text{trk}}$. These high-multiplicity bins are too selective for our modest minimum-bias event sample, and so a different procedure is required. We exploited, for this purpose, the approximate monotonic relation between each event’s initial transverse energy density

$$\left. \frac{dE_T}{d\eta} \right|_{\eta_s=0} = \tau_0 \int d^2x_\perp e(x_\perp, \eta_s = 0, \tau_0),$$

(5.68)

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

Consider, for example, a single multiplicity bin $[N_{\text{ch}}^{\text{low}}, N_{\text{ch}}^{\text{high}}]$ which selects events from a minimum-bias event sample with $p_T^{\text{min}} < p_T < p_T^{\text{max}}$ and $|\eta| < \eta^{\text{max}}$. Let $\langle N_{\text{ch}} \rangle$ denote the average charged-particle multiplicity of these events. We first rescaled the experimental multiplicity bin edges $[N_{\text{ch}}^{\text{low}}, N_{\text{ch}}^{\text{high}}] \rightarrow \left[ \frac{N_{\text{ch}}^{\text{low}}}{\langle N_{\text{ch}} \rangle}, \frac{N_{\text{ch}}^{\text{high}}}{\langle N_{\text{ch}} \rangle} \right]$, in order to reexpress each bin edge as a unitless variable. These bin edges were then associated with a pair of transverse energy bin edges

$$\left[ \frac{E_T^{\text{min}}}{\langle E_T \rangle}, \frac{E_T^{\text{max}}}{\langle E_T \rangle} \right] \leftrightarrow \left[ \frac{N_{\text{ch}}^{\text{low}}}{\langle N_{\text{ch}} \rangle}, \frac{N_{\text{ch}}^{\text{high}}}{\langle N_{\text{ch}} \rangle} \right],$$

(5.70)

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

Finally, we mimicked the method used by experiment and applied equation (5.70) 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 also had to generate (much like experiment) a separate sample of multiplicity-triggered events. In practice, we used a few hundred to a few thousand events per multiplicity bin, depending on the type of observable.
5.3.2.1 Treatment of uncertainties

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

$$\Sigma = \Sigma_m + \Sigma_e$$

(5.71)

described in subsection 4.4.1.1, which includes uncertainty contributions from both the model emulator $\Sigma_m$ and the experimental data $\Sigma_e$. The experimental contribution to the covariance $\Sigma_e$ can be further broken down into its statistical and systematic components

$$\Sigma_e = \Sigma_{e,\text{stat}} + \Sigma_{e,\text{sys}}.$$ (5.72)

The statistical errors in $\Sigma_{e,\text{stat}}$ are uncorrelated, so its covariance matrix is diagonal:

$$\Sigma_{e,\text{stat}} = \text{diag}[(\sigma_{1,\text{stat}})^2, (\sigma_{2,\text{stat}})^2, \ldots, (\sigma_{m,\text{stat}})^2],$$ (5.73)

where $\sigma_{i,\text{stat}}$ is the statistical uncertainty of observable $y_i$ in the experimental observable vector $y_e = (y_1, \ldots, y_m)$. The systematic errors, meanwhile, are typically correlated, but the correlation structure is not reported by the experiments so we asserted a reasonable form. We expanded the systematic covariance matrix as

$$(\Sigma_{e,\text{sys}})_{ij} = \rho_{ij} \sigma_i \sigma_j,$$ (5.74)

where $\sigma_i$ and $\sigma_j$ are the systematic errors of observables $y_i$ and $y_j$ respectively, and $\rho_{ij}$ is the Pearson correlation coefficient between observable $y_i$ and $y_j$:

$$\rho_{ij} = \frac{\text{cov}(y_i, y_j)}{\sigma_i \sigma_j},$$ (5.75)

which satisfies $\rho_{ij} = 1$ for $i = j$ and $|\rho_{ij}| \leq 1$ for $i \neq j$. We assumed that each observable is correlated across different centrality/multiplicity bins, and uncorrelated with observables.
Figure 5.30 Visualization of the Pb-Pb correlation matrix \( \text{corr}(y_i, y_j) = \text{cov}(y_i, y_j) / (\sigma_i \sigma_j) \) 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 [56, 248]. The experimental correlation structure is modeled using equation (5.76).

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 asserted a simple Gaussian form

\[
\rho_{ij}^{\text{sys}} = \exp \left[ -\frac{1}{2} \left( \frac{b_i - b_j}{l} \right)^2 \right], \tag{5.76}
\]

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 used centrality correlation lengths \( l = 100 \) for all of the centrality binned Pb-Pb observables and \( l = 30 \) for the centrality binned \( p\)-Pb charged-particle yield \( dN_{\text{ch}}/d\eta \). The \( p\)-Pb mean \( p_T \) and flow observables, meanwhile, use relative multiplicity bins \( N_{\text{ch}} / \langle N_{\text{ch}} \rangle \) and \( N_{\text{trk}} / \langle N_{\text{trk}} \rangle \) which necessitate a smaller correlation length \( l = 5 \). The right-side of figure 5.30 shows an example correlation matrix

\[
\text{corr}(y_i, y_j) = \text{cov}(y_i, y_j) / (\sigma_i \sigma_j), \tag{5.77}
\]
for the Pb-Pb experimental data constructed using equation (5.76). Here $y_i$ denotes an element of the experimental data $y_e$ and $\sigma_i$ its corresponding uncertainty. The correlation matrix is block diagonal, with each block representing the correlations within a single class of observable.

For the model covariance matrix $\Sigma_m$, we used the procedure described in subsection 4.3.4 to calculate the emulator covariance in principal component space and transform it to the physical observable space. This covariance matrix accounts for the statistical uncertainty caused by our limited event statistics and the interpolation uncertainty caused by our limited number of design points. Figure 5.30 shows the resulting Pb-Pb correlation matrix $\text{corr}(y_i, y_j)$ for the model (emulator) at a random parameter point $x$ in the design space (left-side) along side the same correlation matrix for the experimental data (right-side) discussed previously.

### 5.3.3 Bayesian parameter estimation

We calibrated the model on two different collision systems by expanding the likelihood, equation (4.52), into a joint likelihood

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

where $E$ subsumes all evidence from the $p$-Pb and Pb-Pb collision systems and $x$ is shorthand for the hypothesis that $x = x_{\text{true}}$. We then used emcee [211] to perform Markov-chain Monte Carlo (MCMC) importance sampling on the posterior distribution $P(x|E)$ defined by equation (4.51), assuming a flat prior on each parameter that was constant within the design range 5.10 and zero outside it. The MCMC chain was allowed to “burn-in” before generating $O(7)$ posterior samples.

We validated the emulator predictions using a method known as k-fold cross validation. The $d = 500$ training points were randomly partitioned into $k = 20$ equal sized subsamples or “folds”. For each partition, we used one subsample to validate the emulator and the remaining $k - 1$ subsamples to train it. The process was then repeated for each of the
Figure 5.31 Example emulator validation for one observable, the Pb-Pb charged-particle yield $dN_{\text{ch}}/d\eta$ 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 line $y_{\text{pred}} = y_{\text{obs}}$. 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.

This comparison is shown by the histogram and box plot on the right side of figure 5.31. 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 included this uncertainty in the likelihood covariance matrix (4.52), 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.
Figure 5.32  Emulator validation for the Pb-Pb collision system (top) and p-Pb collision system (bottom) at $\sqrt{s_{NN}} = 5.02$ 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 equation (5.79).
Figure 5.32 applies this z-score validation test to every observable $y \in \mathbf{y}$ for the $p$-Pb and Pb-Pb collision systems. 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 figure 5.31. The emulators generally behave as expected, although the $p$-Pb uncertainties are noticeably over estimated. This is not necessarily a problem, but it is suboptimal. Evidently our $p$-Pb emulator is somewhat more accurate than it purports to be. Therefore our posterior uncertainty estimates are conservative.

We also show in figure 5.32 an estimate of the emulator error magnitude. This error is expressed as a fraction

$$f_{err} = \frac{y_{pred} - y_{obs}}{(\Delta y)_{99\%}},$$  (5.80)

where $(\Delta y)_{99\%}$ is 99% of the full variability of $y$ across the design. Thus $f_{err}$ 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 $f_{err}$. It 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. There are at least two ways which we could reduce this error: we could run more $p$-Pb events to reduce our finite statistical error, or we could add more design points to suppress our systematic emulator interpolation error.

### 5.3.4 Posterior parameter estimates

Figures 5.33 and 5.34 show model calculations for the Pb-Pb and $p$-Pb systems respectively at $\sqrt{s_{NN}} = 5.02$ TeV (thin colored lines) compared to experimental data from the CMS [280] and ALICE collaborations [56, 248, 300, 301]. 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
Table 5.12  Posterior parameter estimates corresponding to Fig. 5.35. The reported values are for the distribution median and 90% HPD credible interval.

<table>
<thead>
<tr>
<th>Initial condition / Pre-eq</th>
<th>QGP medium</th>
</tr>
</thead>
<tbody>
<tr>
<td>Norm</td>
<td>(η/s)$<em>{\text{min}}$ 0.08$^{+0.07}</em>{-0.07}$</td>
</tr>
<tr>
<td>$p$</td>
<td>(η/s)$<em>{\text{slope}}$ 1.24$^{+1.46}</em>{-1.24}$ GeV$^{-1}$</td>
</tr>
<tr>
<td>$\sigma_{\text{fluct}}$</td>
<td>(η/s)$<em>{\text{crv}}$ $-0.09^{+0.80}</em>{-0.91}$</td>
</tr>
<tr>
<td>$r$</td>
<td>(ζ/s)$<em>{\text{max}}$ 0.026$^{+0.032}</em>{-0.026}$</td>
</tr>
<tr>
<td>$n_c$</td>
<td>(ζ/s)$<em>{\text{width}}$ $0.035^{+0.043}</em>{-0.035}$ GeV</td>
</tr>
<tr>
<td>$v$</td>
<td>(ζ/s)$<em>{T_0}$ $0.174^{+0.020}</em>{-0.024}$ GeV</td>
</tr>
<tr>
<td>$d_{\text{min}}$</td>
<td>$T_{\text{switch}}$ $0.149^{+0.013}_{-0.014}$ GeV</td>
</tr>
<tr>
<td>$\tau_{\text{fs}}$</td>
<td>$0.47^{+0.55}_{-0.37}$ fm/c</td>
</tr>
</tbody>
</table>

of observable. The charged-particle yield $dN_{\text{ch}}/d\eta$ 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 5.10, 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 each figure) 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. I’ll demonstrate later in the text that a single set of model parameters well describes all of the
Figure 5.33  Simulated observables compared to experimental data for Pb-Pb collisions at \( \sqrt{s_{NN}} = 5.02 \) 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. Black symbols are experimental data from ALICE with statistical and systematic errors added in quadrature [56, 248].

Figure 5.34  Same as Fig. 5.33 but for \( p\)-Pb collisions at \( \sqrt{s_{NN}} = 5.02 \) TeV. Multiplicity bins are used for mean \( p_T \) and flow cumulant observables to match the bins used by experiment. Experimental data are from ALICE [300, 301] and CMS [280].
calibration data. Therefore it appears 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.

Figure 5.35 shows the main result of this work, the fifteen dimensional posterior distribution for the model input parameters. Table 5.12 also lists each parameter’s median and 90% HPD credible interval. I’ll proceed as before by describing the constraints on each model parameter one-by-one, starting with the initial condition parameters located in the top-left corner of the posterior diagonal and ending with the QGP medium parameters located in the bottom-right.

5.3.4.1 Initial condition properties

The TRENTo normalization factor $\text{Norm} = 20.0^{+2.6}_{-2.5}$ and generalized mean energy deposition parameter $p = 0.002^{+0.157}_{-0.180}$ are well constrained by the present analysis. Moreover, figures 5.33 and 5.34 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 generalized mean 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.
Figure 5.35  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.
Figure 5.36  Left: Joint posterior distribution for the constituent sampling radius \( r \) and constituent width \( v \). The prior range for \( r \) and \( v \) spans the full plot range. The posterior distribution, shown in blue, indicates the preferred values for \( r \) and \( v \) determined by the analysis. Right: Marginal posterior distribution on the constituent number \( n_c \) which was varied in the initial range \( n_c \in [1, 9] \).

It is therefore compelling that \( p \sim 0 \) correctly describes the charged-particle yield \( dN_{ch}/d\eta \) of both systems while simultaneously describing the centrality dependence of \( v_n \{k\} \), an observable which is also known to strongly depend on \( p \) [140]. Specifically, this value corresponds to an energy deposition mapping proportional to the geometric mean of participant nuclear thickness

\[
e(\mathbf{x}_\perp, \eta_s = 0, \tau_0) \propto \sqrt{\hat{T}_A \hat{T}_B}. \tag{5.81}
\]

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 in equation (5.81). We also note that this scaling is somewhat different than the scaling obtained by reference [178], 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.

Continuing down the diagonal in figure 5.35, we see that the constituent sampling radius
r and the constituent width v are both tightly constrained. Figure 5.36 shows the joint posterior distribution of both parameters (left side), plotted for the region scanned by the Bayesian prior. This figure suggests that we can infer the nucleon’s fluctuating size and shape from the collective properties of bulk particle production, a feat largely unimaginable a decade ago. While the sampling radius r 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 constituent positions land on the same side of the nucleon, the effective nucleon size will be smaller than the Gaussian sampling radius r. Despite this idiosyncrasy, one can easily define a physical nucleon width in the nucleon center-of-mass frame \textit{ex post facto}, given specific values for the sampling radius r, constituent width v, and constituent number \( n_c \).

For example, using the posterior distribution’s median values, \( r = 0.88 \) fm, \( n_c = 6 \), and \( v = 0.52 \) fm, we can generate a large ensemble of random nucleon configurations and average their density in each nucleon’s center-of-mass frame. The resulting ensemble-averaged nucleon density

\[
\langle \rho_n(x) \rangle = \frac{1}{(2\pi w^2)^{3/2}} \exp \left( -\frac{|x|^2}{2w^2} \right),
\]

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 \( \sqrt{s_{NN}} = 2.76 \) and 5.02 TeV using a physics model without nucleon substructure [140].

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 [293, 294]. Those studies find that the measured coherent and incoherent \( J/\Psi \) 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. Alternatively, it is also possible that the fluctuations probed by coherent and incoherent $J/\Psi$ production are different than those probed by minimum-bias particle production.

Moving on, we redirect our attention to the posterior on the constituent number $n_c$ shown enlarged in figure 5.36 (right side). 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. Sans nucleon substructure, saturation based models produce proton-sized fireballs in $p$-Pb collisions characterized by small eccentricities [196]. Saturation-based models are therefore unable to describe the significant flow measured in high-multiplicity $p$-Pb collisions without nucleon substructure, or alternatively, some other source of additional correlations [302]. It’s also worth noting that we see no special preference for $n_c = 3$ constituents as commonly used in the literature. In fact, larger constituent numbers $n_c$ generally improve the description of the data.

5.3.4.2 QGP medium properties

It’s interesting to compare the posterior estimates for the shear and bulk viscosities obtained by this study to those of reference [221] which used an (almost) identical version of the present physics model. The only modeling difference is the inclusion of nucleon substructure in the present study which was absent from reference [221]. Several calibration details, however, are different between the two analyses. This work used a modest number of $p$-Pb and Pb-Pb observables at $\sqrt{s_{NN}} = 5.02$ TeV (limited by availability), whereas reference [221] calibrated on a much larger number of Pb-Pb observables at $\sqrt{s_{NN}} = 2.76$ and 5.02 TeV.

The posterior free streaming time $\tau_{fs} = 0.47^{+0.55}_{-0.37}$ fm/$c$ obtained in this work is significantly smaller than the estimate $\tau_{fs} = 1.16^{+0.29}_{-0.25}$ fm/$c$ obtained by reference [221]. 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 $\sqrt{s_{NN}} = 5.02$ 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 figure 5.37 our estimates for the temperature dependence of the QGP specific shear viscosity $(\eta/s)(T)$ and bulk viscosity $(\zeta/s)(T)$ with those of reference [221]. 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 reference [221] are shown in blue. In general, our estimates are broader and less certain than reference [221] but otherwise self-consistent. Evidently, the combined analysis of Pb-Pb data at $\sqrt{s_{NN}} = 2.76$ and 5.02 TeV in reference [221] 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 used in this study, meanwhile, does not appear to provide any unique viscous constraints.
Figure 5.38 Model calculations using the maximum a posteriori (MAP) parameters compared to experiment. Colored lines are model calculations for p-Pb collisions (left) and Pb-Pb collisions (right) at √sNN = 5.02 TeV. Points with error bars are the experimental data with statistical uncertainties, and gray bands their corresponding systematic uncertainties, from CMS [280] and ALICE [56, 248, 300, 301]. The sub-axes show the ratio of model over data where available with gray bands indicating ±10%.
5.3.4.3 Verification of high-probability parameters

Finally, we verified the emulator and tested the accuracy of our physics model framework using parameters selected from the mode of the Bayesian posterior distribution, listed in table 5.13. Recall that these parameters characterize the approximate “best fit” values of the calibrated model, and are commonly referred to as the maximum a posteriori (MAP) estimate. We then ran $O(10^6)$ minimum-bias and multiplicity-triggered events using the MAP estimate and computed all of the model observables listed in section 5.3.2. The resulting model calculations are shown in figure 5.38 alongside experimental data from CMS [280] and ALICE [56, 248, 300, 301]. 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.

Table 5.13 High-probability parameters selected from the posterior distribution and used to generate Fig. 5.38. The posterior distribution on the particlization temperature $T_{\text{switch}}$ is flat (agnostic), so we fix it’s value using Ref. [221].

<table>
<thead>
<tr>
<th></th>
<th>Initial condition / Pre-eq</th>
<th>QGP medium</th>
</tr>
</thead>
<tbody>
<tr>
<td>Norm</td>
<td>20. GeV</td>
<td>($\eta/s)_{\text{min}}$</td>
</tr>
<tr>
<td>$p$</td>
<td>0.0</td>
<td>($\eta/s)_{\text{slope}}$</td>
</tr>
<tr>
<td>$k$</td>
<td>0.19</td>
<td>($\eta/s)_{\text{crv}}$</td>
</tr>
<tr>
<td>$n_c$</td>
<td>6</td>
<td>($\zeta/s)_{\text{max}}$</td>
</tr>
<tr>
<td>$r$</td>
<td>0.81 fm</td>
<td>($\zeta/s)_{\text{width}}$</td>
</tr>
<tr>
<td>$v$</td>
<td>0.43 fm</td>
<td>($\zeta/s)_{T_0}$</td>
</tr>
<tr>
<td>$d_{\text{min}}$</td>
<td>0.81 fm</td>
<td>$T_{\text{switch}}$</td>
</tr>
<tr>
<td>$\tau_{fs}$</td>
<td>0.37 fm/c</td>
<td></td>
</tr>
</tbody>
</table>

The global agreement of the MAP model calculations with the experimental data is very 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 $p$-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 \([140, 157, 168, 178]\). 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 figure 5.38 shows our model calculation for the four-particle elliptic flow cumulant \( v_2 \{4 \} \) along with the measured data points from ALICE \([248]\). 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

\[
(\delta p_T)^2 = \langle \langle (p_T, i - \langle p_T \rangle)(p_T, j - \langle p_T \rangle) \rangle \rangle. \tag{5.83}
\]

The inner-average in equation (5.83) runs over all pairs of particles \( i, j \) in the same event, the
Figure 5.40  Model calculations of the symmetric cumulants (top) and normalized symmetric cumulants (bottom) for Pb-Pb collisions at $\sqrt{s_{NN}} = 5.02$ 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. [140] (without nucleon substructure) which was calibrated on Pb-Pb observables at $\sqrt{s_{NN}} = 2.76$ and 5.02 TeV. In general, most model parameters are somewhat different between the two studies.

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 dynamical fluctuations in units of the average transverse momentum $\langle p_T \rangle$.

Figure 5.39 shows the MAP estimate predictions for the $p$-Pb and Pb-Pb relative mean $p_T$ fluctuations $\delta p_T / \langle p_T \rangle$ at $\sqrt{s_{NN}} = 5.02$ TeV. For the Pb-Pb system, we used centrality bins, and for the $p$-Pb system we used 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 [303], and thus it would be interesting to ultimately include this observable in the calibration when the data becomes available.

Finally, we present in figure 5.40 our calculations for the symmetric cumulants $SC(m, n)$ (left) and normalized symmetric cumulants $NSC(m, n)$ (right) for $(m, n) = (4, 2)$ (blue lines) and $(3, 2)$ (orange lines), which were previously defined in subsection 5.1.5. Recall that these observables quantify correlations between event-by-event fluctuations of the flow harmonics of different order [271, 272]. The solid lines are calculations using the current model and
MAP estimate, while the dashed lines are calculations using the model and MAP estimate of reference [221], calibrated to Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ and 5.02 TeV without nucleon substructure. We observe that the gap between SC(3, 2) and SC(4, 2) is generally wider in the present analysis than in reference [140], as is the gap between the normalized symmetric cumulants NSC(3, 2) and NSC(4, 2).

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. Therefore, we can only speculate what might have caused the large difference in the symmetric flow cumulants between the two analyses. 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. Future studies of the symmetric cumulants therefore promise to provide additional constraints on the initial condition and QGP medium parameters.

5.4 Topics for future study

This dissertation shed new light on the QGP initial condition problem, but the problem itself is hardly solved. As I conclude this results chapter, I would like to identify some areas of concern and outline new ideas for future improvements to the present work. The following topics are those which I consider to be the highest priority.

5.4.1 TRENTo model

Nuclear structure  The starting point of every initial condition model is a Monte Carlo generator for the nuclear density inside the nucleus. If this nuclear density is miscalculated, then it can severely hamstring the model predictions. Initial condition models typically use low-energy nuclear structure measurements to model the nucleus at high-energy. This is a somewhat worrisome approximation that should be investigated further. Presumably, if the inelastic cross section grows strongly with collision energy, the size of the nucleus could as well.
Multiplicity fluctuations Minimum-bias proton-proton collisions produce large multiplicity fluctuations which cannot be described by geometric effects alone. To account for these fluctuations, I multiplied each nucleon (or constituent) by an \textit{ad hoc} random weight following previous work [153–155]. Personally, I find this prescription somewhat unsettling because it associates the fluctuations with the participant matter instead of with the dynamical process itself. What is the physical origin of these multiplicity fluctuations?

Generalized mean ansatz The generalized mean ansatz is flexible, but it is still a one parameter function. It is possible that we’ve missed a degree of freedom that is important to describe the data. The merits of the generalized mean ansatz rest heavily on its assumption of scale invariance, equation (3.72), so it is important to test the validity of this assertion. If future experiments show that scale invariance breaks, then the generalized mean ansatz should be modified or discarded. Alternatively, one could try expanding the energy density \( e(\tilde{T}_A, \tilde{T}_B) \) as a sum of symmetric orthogonal functions.

5.4.2 Calibration methodology

New collision systems I have dedicated the vast majority of this dissertation to studying \( p\)-Pb, Xe-Xe, and Pb-Pb collisions. There are, of course, many other collision systems which have been studied by experiments at RHIC and the LHC, including \( p-p, d-Au, ^3He-Au, Cu-Cu, Cu-Au, Au-Au, \) and \( U-U \) to name a few. I’ve presented limited results for a few of these systems, but I have not yet studied them at great length. Once the TRENTo model is calibrated at a specific beam energy, it should be able to describe all collision systems without additional fine tuning. Therefore, this additional data should be used to validate (or invalidate) the model.

Parameter energy dependence In principle, the TRENTo parameters could all vary as functions of the beam energy \( \sqrt{s_{NN}} \). It would be interesting, for example, to see if we obtain different posterior estimates for the Gaussian nucleon width at different beam energies, or if we observe a beam energy dependence for the TRENTo generalized mean parameter \( p \).
This could signal the emergence of new physics at high-energies.

**Sizing the nucleon** We obtained relatively consistent posterior estimates for the TRENTo model parameters with one notable exception. The Gaussian nucleon width $\varinjudgeta$ varied considerably from study to study. To complicate matters further, our most recent posterior estimate [221] is incompatible with independent proton size measurements based on coherent and incoherent $J/\Psi$ production [293–296]. It would be helpful to add new observables to the model calibration which are directly sensitive to the fluctuating size and shape of the proton.

**Expanded nucleon substructure study** I presented an exploratory study of $p$-Pb and Pb-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV using an updated version of the TRENTo model with nucleon substructure [204]. Future work should expand this substructure analysis to include more collision systems and beam energies. Nucleon substructure could resolve several existing puzzles, such as the centrality dependence of mean transverse momentum fluctuations $\delta p_T/\langle p_T \rangle$ and symmetric flow cumulants $SC(m, n)$. 

235
Conclusion

Hydrodynamics is an indispensable tool to study the bulk properties of hot and dense nuclear matter produced by ultrarelativistic nuclear collisions. Notably, the framework has been used to extract dynamical properties of the strongly-coupled QGP liquid produced in the first $\sim 10^{-23}$ seconds of the collision, such as its specific shear viscosity $\eta/s$ and bulk viscosity $\zeta/s$. For many years, decades even, theoretical uncertainty in the simulation’s initial conditions has complicated efforts to rigorously constrain the QGP transport coefficients using heavy-ion collision data. This initial condition problem has correspondingly been identified as one of the primary factors impeding our understanding of QGP matter.

Over the last decade, theoretical models of the QGP initial conditions have improved greatly. *Ab initio* calculations based on general concepts of gluon saturation are now able to reproduce a wide variety of experimental bulk observables—including the collision’s final particle yields, mean $p_T$, and multiparticle correlations—to impressive precision. There remains, however, differences between these calculations, and there is no uniform consensus within the theoretical community regarding the validity of each approach. Moreover, it is generally fair to say that no single initial condition model describes the experimental data perfectly. Therefore, it is natural to wonder how this residual error affects extracted values of the QGP transport coefficients.

In this dissertation, I motivated and developed—in collaboration with Jonah Bernhard
and Steffen Bass—a simple parametric model for the QGP initial conditions called TRENTo [168], based on a family of functions known as the generalized means. This functional form parametrizes a semi-exhaustive subspace of all physically reasonable initial condition models, allowing the model to not only mimic certain calculations in the literature but also interpolate between them. Hence, it effectively represents a meta-model for the landscape of mutually incompatible theory calculations.

This parametric initial condition model was then embedded in a realistic hydrodynamic simulation and constrained using a Bayesian parameter estimation framework developed by Jonah Bernhard [140]. Leveraging the statistical machinery provided by his model-to-data comparison framework, we were able to show that the initial condition and QGP medium properties are simultaneously constrained by the experimental data [221]. This analysis resulted in the first quantitative estimates for \((\eta/s)(T)\) and \((\zeta/s)(T)\) obtained from hydrodynamic simulations with rigorously defined modeling uncertainties (see figure 5.14). Determining these quantities has been a primary goal of the RHIC and LHC heavy-ion programs and has been the subject of numerous published papers. Notably, we found that the QGP specific shear viscosity obtains a minimum value \((\eta/s)_{\text{min}} = 0.085^{+0.026}_{-0.025}\) at \(T_c = 154\) MeV which is conspicuously close to the conjectured KSS bound \(\eta/s \geq 1/4\pi\) [236–238].

Studies of the TRENTo initial condition model at midrapidity also suggest that the collision’s initial entropy density [178] or energy density [204, 221] scales approximately as

\[
\begin{pmatrix}
\epsilon_0 \\
S_0
\end{pmatrix} \propto \sqrt{\tilde{T}_A \tilde{T}_B},
\]

where \(\tilde{T}_A, \tilde{T}_B\) are the participant thickness functions of each nucleus defined by equation (3.48).

The specific meaning of this expression, i.e. whether the quantity on the left-side is an energy or entropy density, depends on the assumed hydrodynamic matching procedure. Although one should be careful interpreting this analytic form too literally—the analyses
constrain an approximate region of function space, not a specific analytic form—the result is so simple that it merits further investigation. Furthermore, we find that this scaling persists for different beam energies and nuclear collision systems.

This dissertation research further expands on the previous studies by performing additional exploratory analyses of the QGP initial conditions far from midrapidity and in small collision systems. In both cases, the present hydrodynamic model framework provides a compelling simultaneous description of bulk particle properties, corroborating the assumptions of the TRENTo initial condition framework and the broad success of hydrodynamic descriptions.

Perhaps most intriguingly, the small-system study revealed that a simultaneous quantitative description of \( p\text{-Pb} \) and \( \text{Pb-Pb} \) bulk observables at \( \sqrt{s_{\text{NN}}} = 5.02 \) TeV is obtainable if one simply replaces spherically symmetric round nucleons with deformed “lumpy” nucleons, parametrized by two additional degrees of freedom.

This observation evidences the hydrodynamic nature of small-system multi-particle correlations and suggests that a unified hydrodynamic description of small and large collision systems is emerging. While additional work is needed to establish if this is indeed the case, it is compelling that the TRENTo model with nucleon structure simultaneously describes the \( p\text{-Pb} \) and \( \text{Pb-Pb} \) collision systems using the exact same model parameters and model-to-data comparison methods.

I want to conclude by pausing to appreciate the remarkable success of the hydrodynamic framework pioneered by Landau, Bjorken, and the many other scientists which have contributed to its formulation over the years. Extracting the QGP specific shear viscosity to \( \pm 30\% \) accuracy from hydrodynamic simulations would have been unthinkable fifty years ago. Moreover, no one could have imagined that hydrodynamic simulations might evidence the fluctuating shape of constituent sources inside the proton. These developments indicate that hydrodynamic simulations are entering a new precision age.
References


241


245


[138] L. Frankfurt, M. Sargsian, and M. Strikman, “Recent observation of short range nucleon correlations in nuclei and their implications for the structure of nuclei and


[151] V. Khachatryan et al., “Charged particle multiplicities in pp interactions at $\sqrt{s} = 0.9$, 2.36, and 7 TeV”, JHEP 01, 079 (2011), arXiv:1011.5531 [hep-ex].


[177] K. J. Eskola, P. V. Ruuskanen, S. S. Rasanen, and K. Tuominen, “Multiplicities and transverse energies in central A A collisions at RHIC and LHC from pQCD,


253


[216] Bulk viscous corrections and shear-bulk coupling were implemented in VISH2+1 by J. Liu and U. Heinz. A first preliminary study involving a much more restricted set of fit parameters including bulk viscosity was presented at the Quark Matter 2015 conference [241].


259


261


Biography

John Scott Moreland earned his B.S. in Physics and Mathematics, with honors, from Ohio State University in 2011, and his Ph.D. in Physics from Duke University in 2019.