Bayesian inference in high-energy nuclear physics

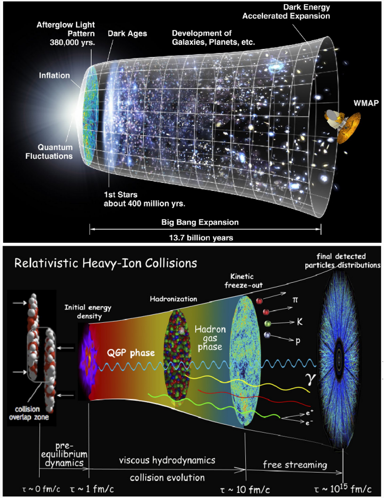
The 9th HuaDa QCD School 2021

Weiyao Ke
UCB & LBNL

Table of contents

Introduction

High-energy nuclear collisions “the little bang”



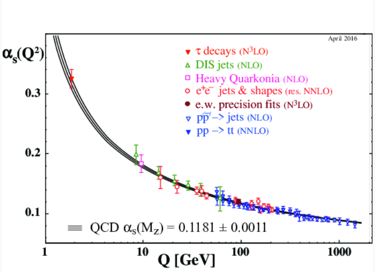
[NASA ↑, P. Sorensen & C. Shen ↓]

Why do we collide nuclei at extreme energies

- Evolution of universe from the Big bang: initial state $\rightarrow$ fast expansion $T\downarrow$ $\rightarrow$ Decoupling / freeze-out $\rightarrow$.
- Colliding heavy nuclei:
  - Initial temperature $T \approx 500$ MeV $\sim 10^{13}$ K. Approximately the temperature at $t = 10^-6$s of the universe.
  - Strongly interacting & expanding matter $\rightarrow$ freeze out.

From hadrons to quark-gluon plasma

Asymptotic freedom of quantum chromodynamics (QCD)

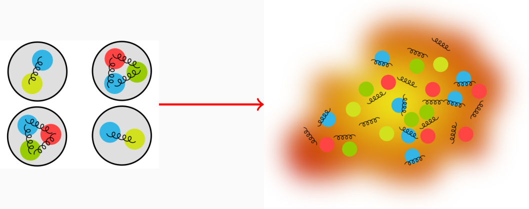


Coupling $\alpha_s = g^2/(4\pi)$ decreases in the perturbative regime. [Figure from PDG]

At sufficient high temperature / energy density $\alpha_s(3k_B T)$ becomes small

Hadrons

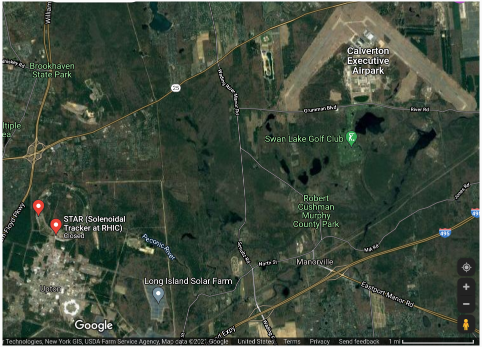
Quark-gluon plasma (QGP)



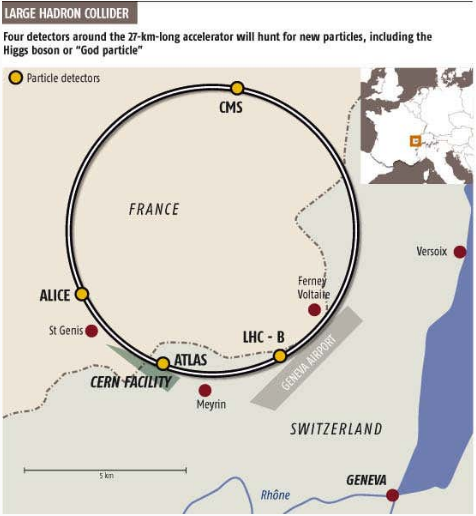
Quarks & gluons / color fields liberate from bound states [Figures from JS Moreland]

Create the little bang

The Relativistic Heavy-Ion Collider (RHIC) at the Brookhaven National Lab (BNL).
The Large Hadron Collider (LHC) at the European Organization for Nuclear Research (CERN).



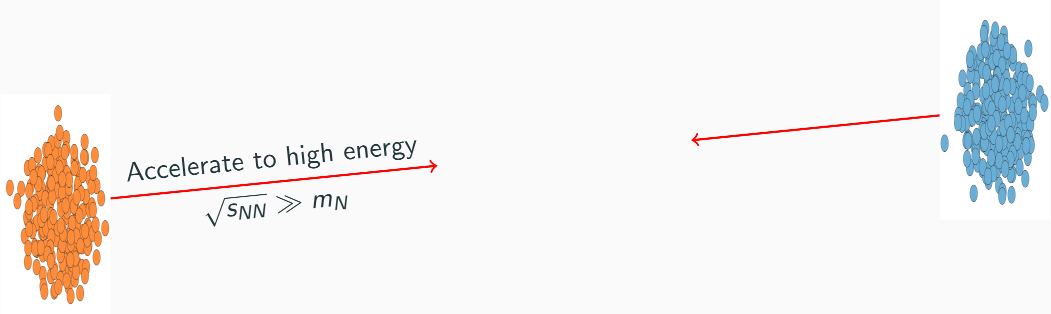
[From Google Map]



[From newscientist.com]

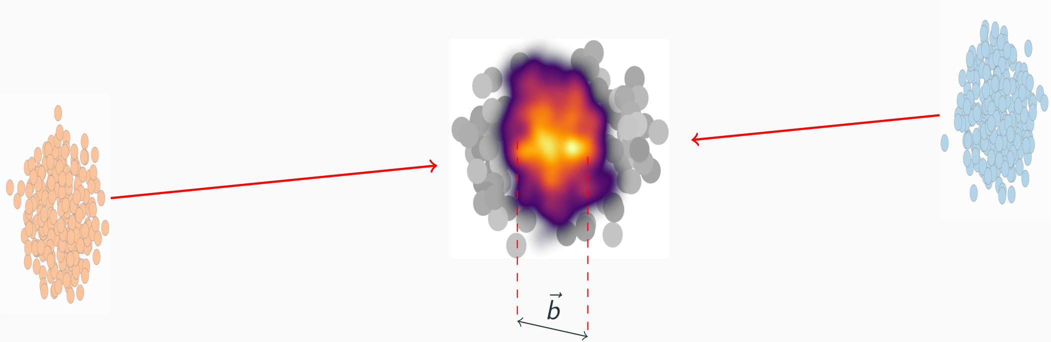
Basic pictures

- At high energy $\gamma = \sqrt{s_{NN}}/(2m_N) \gg 1$, strong Lorentz contraction in the beam direction.



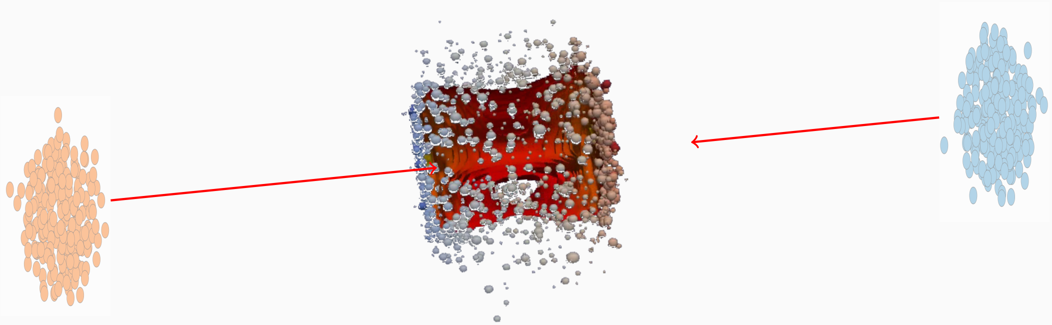
Basic pictures

- At high energy $\gamma = \sqrt{s_{NN}}/(2m_N) \gg 1$, strong Lorentz contraction in the beam direction.
- Almost instantaneous energy production in the interaction / overlapped region.
  $\Delta t_I \sim 2R/\gamma \ll r_N, \Lambda_{\mathrm{QCD}} \to$ nuclei collide with their internal d.o.f. freezes within δt_I.
  Nuclear configuration fluctuates event by event.



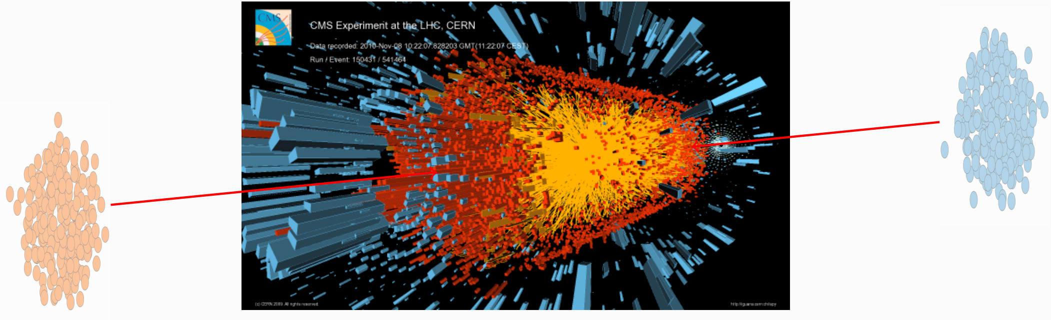
Basic pictures

- At high energy $\gamma = \sqrt{s_{NN}}/(2m_N) \gg 1$, strong Lorentz contraction in the beam direction.
- Almost instantaneous energy production in the interaction / overlapped region.
  $\Delta t_I \sim 2R/\gamma \ll r_N, \Lambda_{\mathrm{QCD}} \to$ nuclei collide with their internal d.o.f. freezes within δt_I.
  Nuclear configuration fluctuates event by event.
- Complicated many-body dynamics



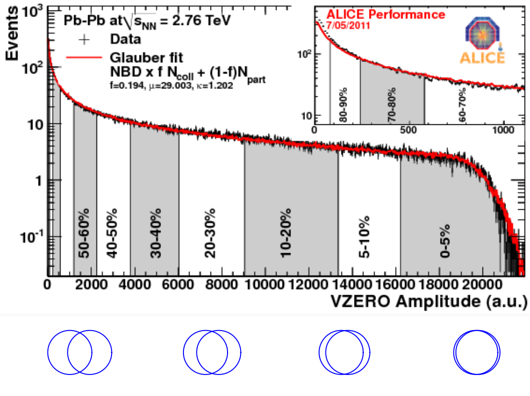
Final states:

- Up to 10^4 particles in central A-A collisions at top LHC energy.
- Most spatial information is lost. We only observe the momentum space & correlation.
- To learn that happens during the collision is a very hard inverse problem.



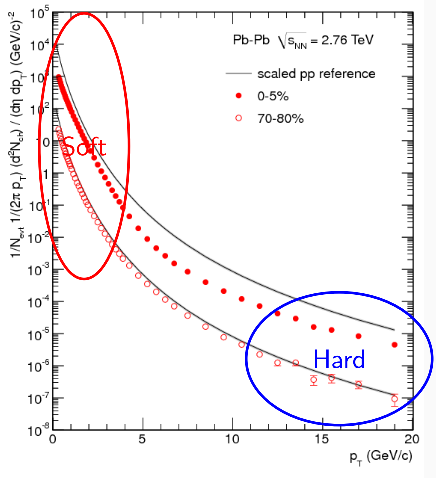
[Figure credit to the CMS collaboration]

Centraltiy classification



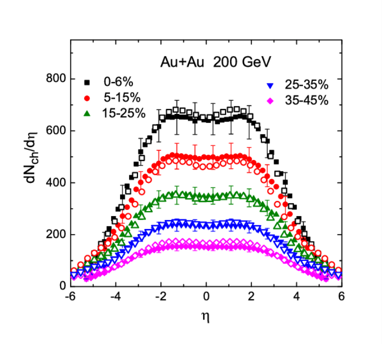
- In *A*-*A* collisions, particle production is strongly correlated with the geometry overlap.
- Sort events according to particle production. The percentile range / centrality (0-5%, 40-50%) is a good indicator of the average geometry in these collisions.
- This relation is model dependent, we cannot directly measure the impact parameter *b*.
- May not work in small collision systems.

Momentum distribution of particle



ALICE Collaboration PLB 696 (2011) 30-39

- Most particles are produced with small transverse momenta ($p_T \lesssim 3$ GeV).
- Hard particles are rare ($p_T \gtrsim 10$ GeV).
- Usually use rapidity/pseudorapidity instead of p_z: $\eta = \frac{1}{2}\ln\frac{|p|+p_z}{|p|-p_z}$. Soft particles $dN/d\eta$ displays a central plateau.

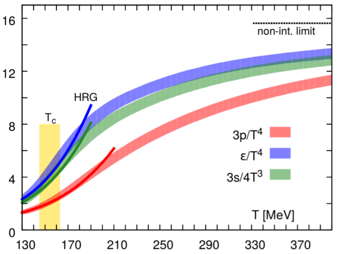


[PHOBOS]

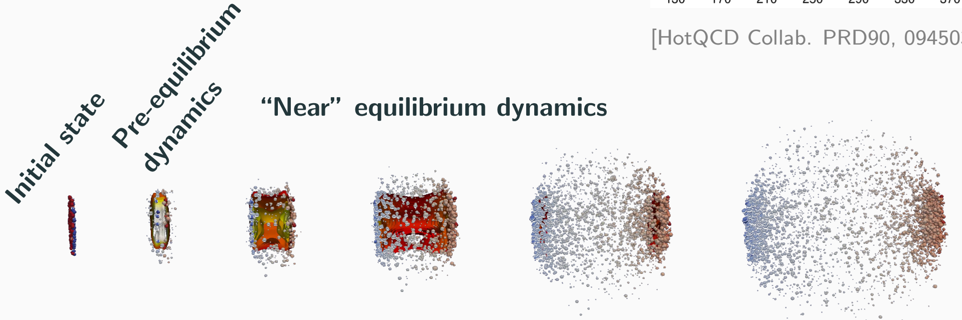
Our current understanding of the bulk of the particle (soft)

They dynamics is governed by several competing factors:

- Longitudinal expansion $1/\tau$.
- Collisions & many-body effect $1/\tau_{\rm coll} \propto d_{\rm eff}(T)T$.
- Pressure driven transverse expansion $1/R_T$.
- Freeze-out.



[HotQCD Collab. PRD90, 094503 (2014)]



Expansion drives the system out of equilibrium and cools down / hadronize

A multi-scale problem with a mult-stage modeling

A mutli-stage approach: with so many competing effects, one build specific model for each stage, staring from the dominant effect.

- Initial condition: neglect dynamics within $0^+ < \tau < \delta t_I = 2R/\gamma$.
- At early times, $\delta t_I < \tau < \tau_0$ longitudinal expansion dominates. Free-streaming + corrections from few interactions.
- Intermediate stage, collisions become frequent. Near equilibrium model with viscous hydrodynamics (Equation of state, viscosity, $\cdots$)
- Late stage, density goes down ($d(T)T^3$), system hadornize. Use Boltzmann equation for hadrons.

Multi-stage model

Advantage: simplified treatment in each stage and can be systematically improved.

Challenges:

- Contains many parameters and moving parts!
- Uncertainties from matching.
  - From early stage evolution to a classical hydrodynamics.
  - From hydrodynamic fields to hadron ensembles.

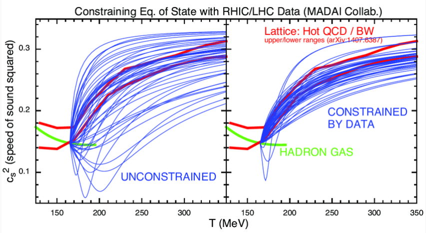
Nowadays, > 10 parameter $+$ unkown functions.

Interested parameter and nuiance parameters

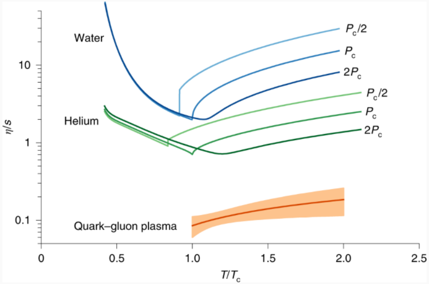
Parameters/functions of physical importance, also well-defined from first principle.

- Equation of state: equilibrium property of hot QCD.
- Transport coefficients, such as specific shear and bulk viscosity: dynamical properties.

Both are direct input to hydrodynamics, and has been constrained / extracted from data using Bayesian techniques:

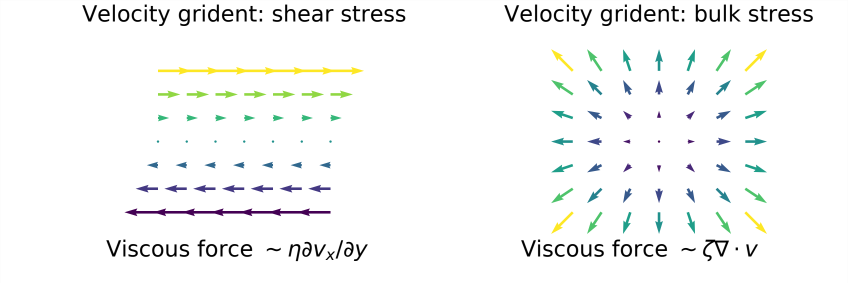


Bayesian constrained EoS of QGP from data v.s. lattice QCD. [S. Pratt et al, PRL 114, 202301 (2015)]



Bayesian extraction of η/s of QGP compared to other substances [Nature Physics 15, 1113–1117 (2019)]

Very interested parameters: specific shear and bulk viscosities



- Shear viscosity: the resistance that a fluid exert to shear strain. Direct probe of interaction strength at thermal scale $\eta/s \sim 1/[g^4 \ln(g \cdots)]$ (LO result)
- Bulk viscosity: the resistance to volume change $\leftrightarrow$ scale invariance ($L \to \lambda L$) breaking.
- How do they change with temperature?

Example of nuisance parameters:

Parameters that are not of immediate interest, lack of physical importance, model specific chocies, etc.

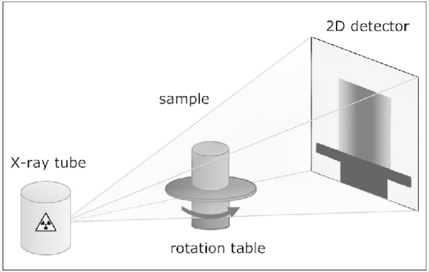
- Matching timescale between pre-equilibrium dynamics and hydrodynamics.
- Some initial condition related parameters.
- Cut-off, regulators, etc.

But they do contribute to the estimation of model uncertainty!

Similar situation for the study using hard particles / hard probes.

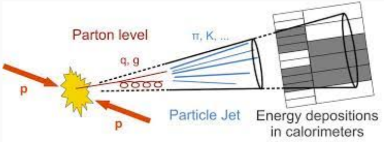
Tomography of medium using hard particles / jets

X-ray tomography. External probe of the internal structure of an object.



High-p_T hadron / jet (collimated spray of particles from parton dynamics) tomography of the nuclear medium.

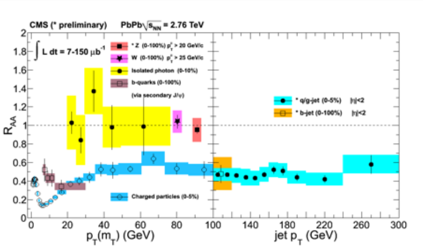
- Self generated probes.
- Both probe and medium undergo complex dynamics.



Tomography of medium using hard particles / jets

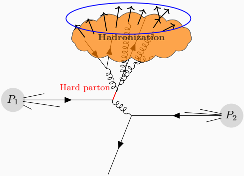
Medium properties imprint in the modification of the probes.
Most directly: less jet production at a fixed momentum p_T

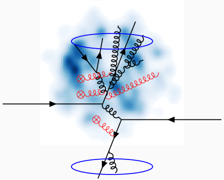
$$R_{AA} = \frac{d\sigma_{AA\to J/h}}{\langle N_{\rm coll}\rangle d\sigma_{pp\to J/h}} < 1$$



High-p_T hadron / jet (collimated spray of particles from parton dynamics) tomography of the nuclear medium.

- Self generated probes.
- Both probe and medium undergo complex dynamics.





Jet-medium interaction is often quantified by the jet transport parameter $\hat{q} = d\langle k_T^2\rangle/dL \to$ strength of gluon field in the medium.

The inverse problem for jet tomography is as challenging

The model part is even more complicated than the current bulk simulations.

Parameters

- $\hat{q}(x_B, Q^2; A)$, $\hat{q}(p, Q^2; T)$ or jet-medium coupling g_s.
- All the medium parameters!
- Cut-offs, regulators, etc.

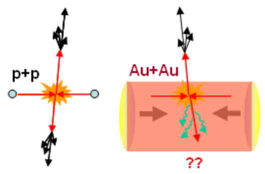
Observables

- Yield $R_{AA} = Y_{AA}/Y_{pp}/\langle N_{\mathrm{coll}}\rangle$.
- Correlations di-jet, hadron-jet, h-h.
- Internal & sub-structure of jets.

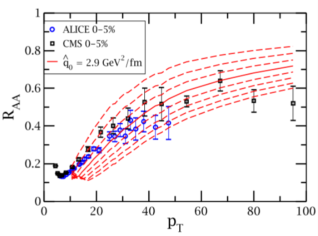
Many models chocies $\cdots \otimes \cdots$

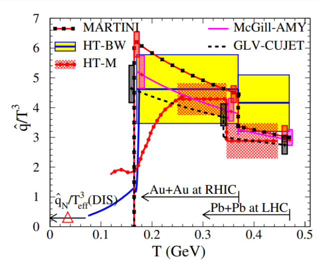
- Perturbative orders of initial production.
- Different assumptions on jet-medium interaction: few hard v.s. multiple soft; weakly v.s. strongly coupled.
- Different approximations to the medium-modified splitting functions.
- Different jet evolution equations.
- Different models for medium evolution.

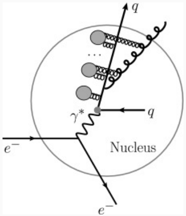
Powerful tool to understand QCD medium in both nuclear collisions and deep inelastic scatterings (DIS) on nucleus. Moving from χ^2 fit to Bayesian analysis (JETSCAPE).

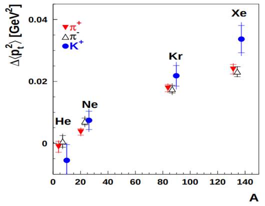


[JET Collaboration, χ^2-fit back in 2013 →]









[HERMES Collaboration]

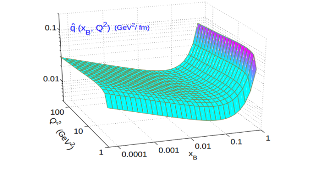


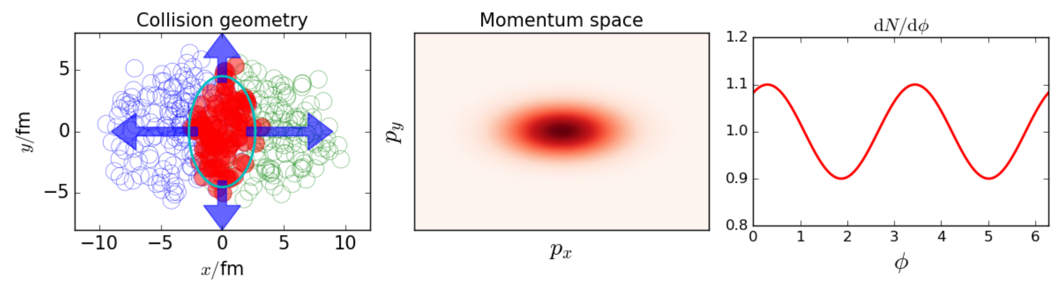
FIG. 4. The extracted $\hat{q}$ as functions of Bjorken x_B and scale Q^2.

[Peng Ru et al, PRD 103, 031901 (2021). ξ^2 fit.]

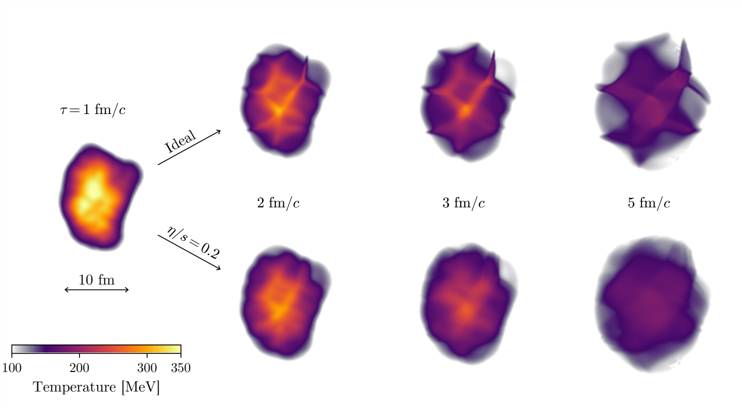
Bayesian inference applied to understand hot QCD matter

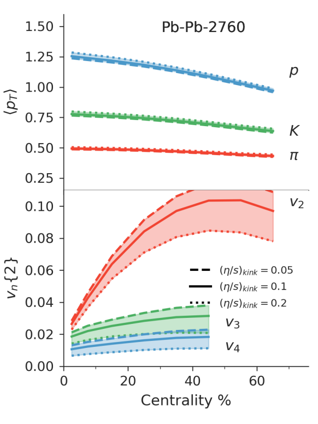
Some key observables: how do we detect the viscous effect?

- Non-central collision creates elliptic shaped blob of quark-gluon plasma.
- Hydrodynamic pressure gradient drives particles to accelerate in the radial direction.
- Initial eccentricity translates to momentum anisotropy $dN/d\phi \sim 1 + 2v_2 \cos(2(\phi - \Psi_2))$
- Approximately linear response $v_2 \approx k_{22}\epsilon_2$, k_{22} depends on viscosities of QGP.



Sensitity of v_n to shear viscosity

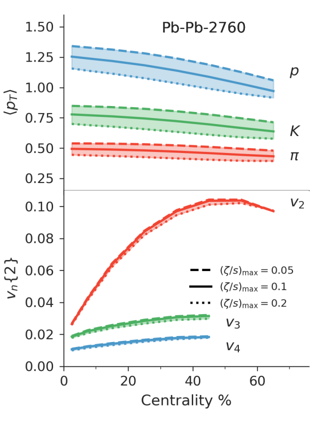




Sensitity of $\langle p_T \rangle$ to bulk viscosity

Main effect of bulk viscosity:

- Slow down the system that radially expands.
- Reduces the average velocity of particles $\rightarrow$ reduced mean transverse momentum $\langle p_T \rangle$.
- The effect is mass dependent.



It looks like we have two observables that can help to pin down η/s, η/s. However, the problem is much more complicated ...

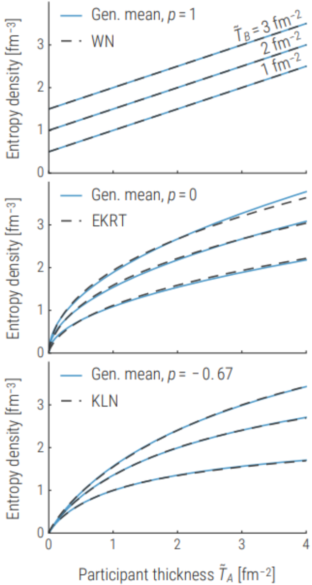
Major uncertainty from initial condition model

Currently, we don't know from first principle what the initial condition is. A alternative way is to parameterize a class of possible energy deposition relation:

$$e(x, y) \propto \left[\frac{T_A(x, y)^p + T_B^p(x, y)}{2} \right]^{1/p}$$

This is just a parametric function, but is shown to reproduce several widely used IC models.

Major uncertainty from initial condition model



- Wounded nucleon model

$$\frac{dS}{dy\, d^2r_\perp} \propto \tilde{T}_A + \tilde{T}_B$$

- EKRT model PRC 93, 024907 (2016)

  after brief free streaming phase

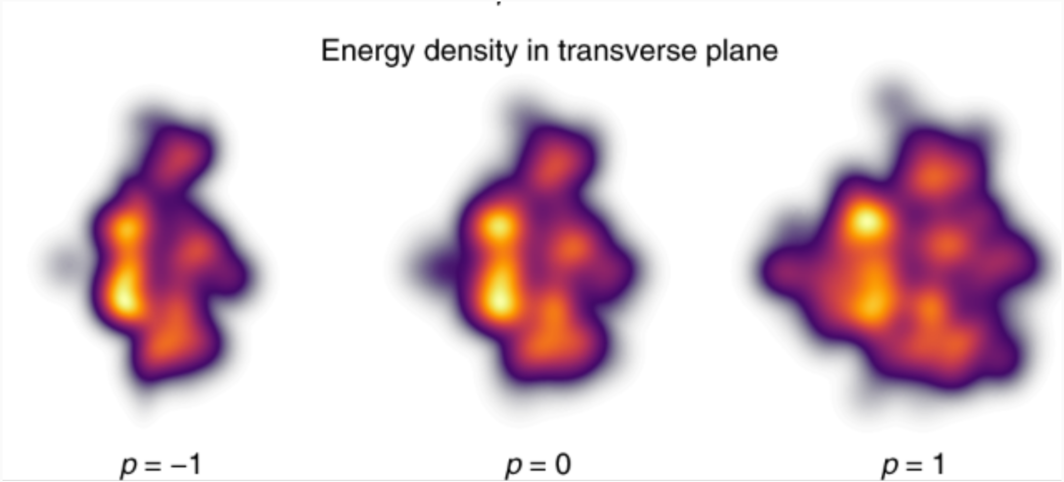
$$\frac{dE_T}{dy\, d^2r_\perp} \sim \frac{K_{\text{sat}}}{\pi} p_{\text{sat}}^3(K_{\text{sat}}, \beta; T_A, T_B)$$

- KLN model PRC 75, 034905 (2007)

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

[Slide credit to J. Scott Moreland]

Major uncertainty from initial condition model



The eccentricity varies a lot in different models. We need a simultaneous calibration of many features of the model to many observables:

- Maybe some observable help to constrain the initial condition.
- If not, propagate IC uncertainty to the interested quantity $\eta/s, \zeta/s$.

Marginalization and uncertainty propagation

$$v_2 = k_{22}(\eta/s)\epsilon_2$$

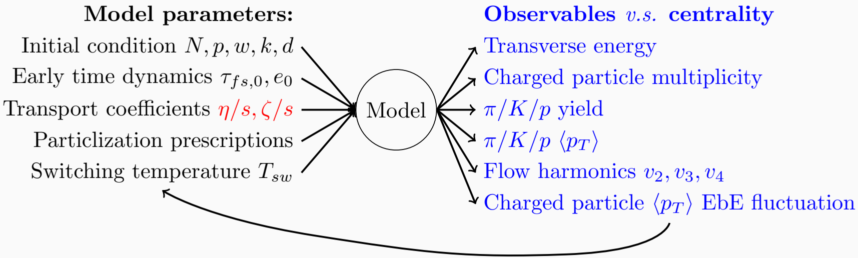
If we can somehow define an inverse problem:

$$\eta/s \quad \text{“}=\text{”} \quad \mathcal{F}(v_2, \epsilon_2)$$

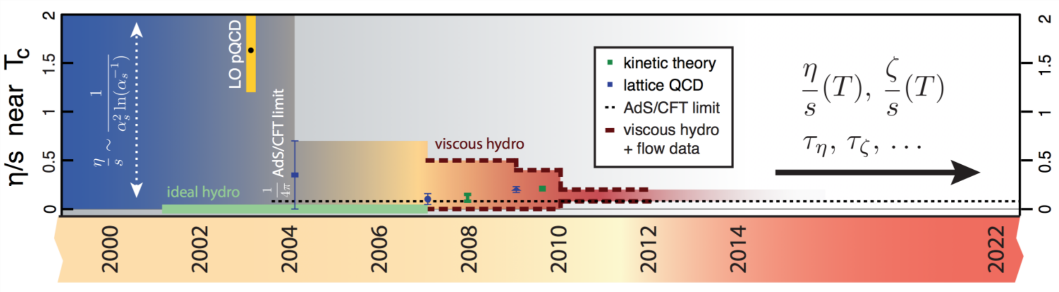
Marginalization: integrate out all possible variation of ϵ_2. The uncertainty of η/s comes from not only experimental uncertainty but also other under-constrained part of the model.

$$P(\eta/s) \sim \int P(v_2 + \delta v_2, \epsilon_2 + \delta_\epsilon) d\delta v_2 d\delta\epsilon$$

The real situation: a lot more parameters to be marginalized



Rigorous statistical procedure is essentail for progress



- 2000s: order of magnitude.
- 2004: strongly coupled theory $\eta/s = 1/(4\pi) + \cdots$.
- 2006-2013: eyeball fit with viscous hydro $(\eta/s)_{\rm eff} = 1 - 2$
- 2013–: Bayesian analysis. Simultaneous calibration of IC, $\eta/s(T)$, etc.
- 2016–: Temperature dependent shear and bulk viscosity. Refined model. Model uncertainty. Model averaging$\cdots$.

Statistical inference problem

1. A model $\mathcal{M}$: predict observables y at given input parameters x.
2. A prior belief of the distribution of true values of x: $P_0(\text{x}_{\text{true}})$
3. Make the measurement y_{exp}, and update the knowledge: $P_0 \rightarrow P(\text{x}_{\text{true}})$.
4. Marginalize over nuisance parameters $P(x^*) = \int P(x^*, \bar{x}) d\bar{x}$
    - x^* interested parameters
    - $\bar{x}$ nuisance parameters: not interest at this time, but is an essential part of model and contribute to uncertainty.

Bayesian Theorem (a recap)

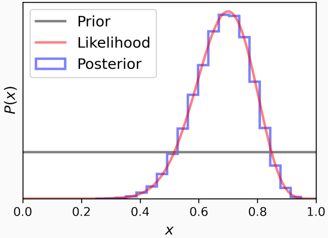
Bayes' theorem (from conditional prob: $P(A \bigcap B) = P(A|B)P(B) = P(B|A)P(A)$)

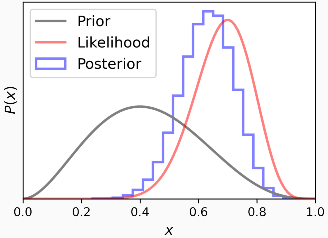
$$\underbrace{P(x_{\mathrm{true}}|\mathcal{M}, y_{\mathrm{exp}})}_{\text{Posterior}} = \frac{\overbrace{L(y_{\mathrm{exp}}|\mathcal{M}, x_{\mathrm{true}})}^{\text{Likelihood}}\overbrace{P_0(x_{\mathrm{true}})}^{\text{Prior}}}{\underbrace{\int L(x)P_0(x)dx}_{\text{Normalization (evidence)}}}$$

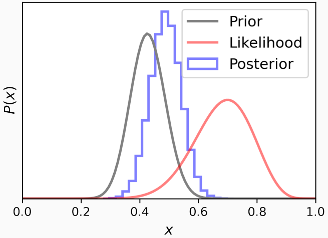
L is often unknown. Commonly assumed to take the form of a multivariate Gaussian:

$$\ln L = \frac{N}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma| - \frac{1}{2}\Delta y \Sigma^{-1} \Delta y^T, \quad \Delta y = y_{\mathrm{exp}} - y(x; \mathcal{M})$$

Prior brief and posterior probability distribution







What are the sources of uncertainty

Uncertainty covariance matrix Σ in the likelihood function:

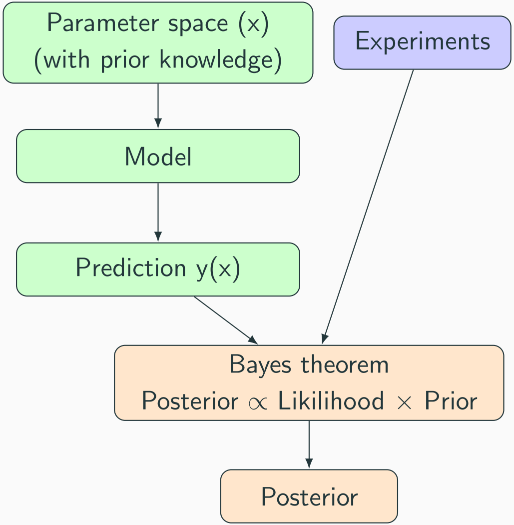
$$\ln L = \frac{N}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma| - \frac{1}{2}\Delta \mathrm{y}\Sigma^{-1}\Delta \mathrm{y}^T, \quad \Delta \mathrm{y} = \mathrm{y}_{\mathrm{exp}} - \mathrm{y}(\mathrm{x}; \mathcal{M})$$

$$\begin{aligned}\Sigma_{ij} &= \underbrace{\delta_{ij}[(\delta y_{\mathrm{stat}})_i^2 + (\delta y_{\mathrm{sys},0})_i^2] + \delta(y_{\mathrm{sys},\infty})_i\delta(y_{\mathrm{sys},\infty})_j + \delta(y_{\mathrm{sys,l}})_i\delta(y_{\mathrm{sys,l}})_j c(i,j;l)}_{\text{Experimental}} \\ &+ \sigma_{ij}^{\text{emulator}} \longleftarrow \text{Interpolation uncertainty, explained later} \\ &+ \sigma_{ij}^{\text{theory}} \longleftarrow \text{Model/theory imperfection, very hard}\end{aligned}$$

Experimental uncertainty

- Statistical & uncorrelated systematic uncertainty: $\delta y_{\mathrm{stat}}, \delta y_{\mathrm{sys},0}$ (zero correlation length).
- Fully correlated systematic uncertainty: $\delta y_{\mathrm{sys},\infty}$ (infinite correlation length).
- Partially correlated systematic uncertainty: $\delta y_{\mathrm{sys,l}}$ (finite correlation length).

For simple models that $y(x)$ is easy to compute:



Complex model: a high-dimenon problem

Take the medium evolution model in HIC as an example:

- Nowadays, > 10 parameters $+$ unknown functions $\eta/s(T), \zeta/s(T)$.
- These parameters are simultaneously constrained by hundreds of measurements.

Complex model: a time-consuming problem

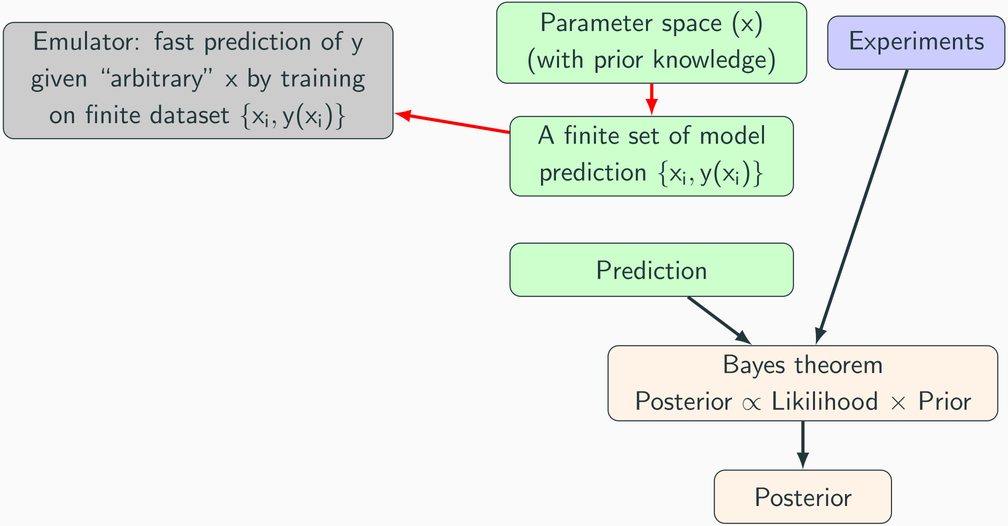
To compute observable at one parameter point:

- 10^4 events with randomized initial condition (multi-particle correlations require even more).
- 2+1D simulation: 0.5h/event. 3+1D simulation: 1day/event.
- If we evaluate the model on a 10^d grid in the parameter space $\rightarrow 10^d$ CPU year.
- To explore the posterior distribution, we should be able to evaluate the model at arbitrary many input points

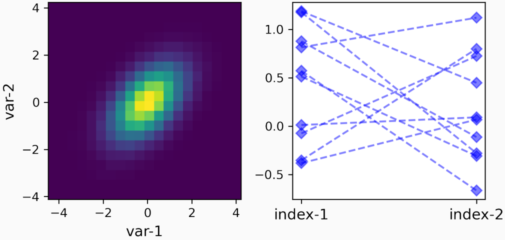
Complex model: non-linearity

- Usually, observable changes monotonically and smoothly with input parameters.
- However, complicated parametrization can result in a large degree of non-linearity.

For computationally intensive model



A class of non-parametic estimator: Gaussian emulators

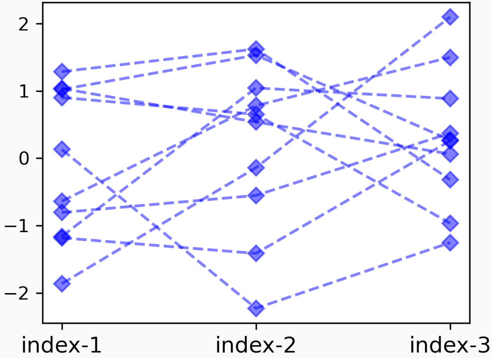


A 2D Gaussian with zero mean and

$$\sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$

off-diagonal controls how correlated (how close) the two output are.

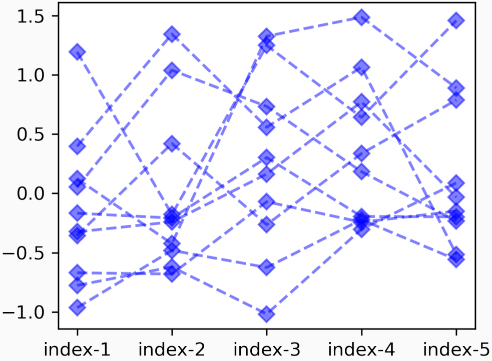
A class of non-parametic estimator: Gaussian emulators



A 3D Gaussian with zero mean and

$$\sigma = \begin{bmatrix} 1 & 0.5 & 0 \\ 0.5 & 1 & 0.5 \\ 0 & 0.5 & 1 \end{bmatrix}$$

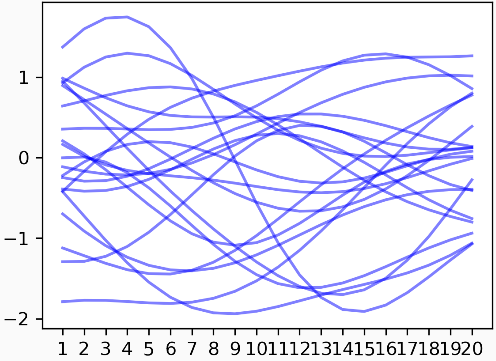
A class of non-parametic estimator: Gaussian emulators



A 5D Gaussian with zero mean and

$$\sigma = \begin{bmatrix} 1 & 0.7 & 0.5 & 0.3 & 0 \\ 0.7 & 1 & 0.7 & 0.5 & 0.3 \\ 0.5 & 0.7 & 1 & 0.7 & 0.5 \\ 0.3 & 0.5 & 0.7 & 1 & 0.7 \\ 0 & 0.3 & 0.5 & 0.7 & 1 \end{bmatrix}$$

A class of non-parametic estimator: Gaussian emulators

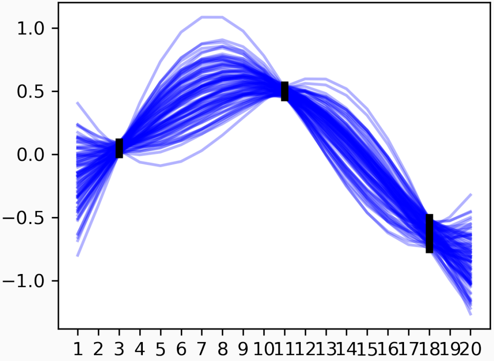


A 20D Gaussian with zero mean and

$$\sigma(x_i, x_j) = \sigma_0^2 \exp\left\{-\frac{(x_i - x_j)^2}{2L^2}\right\}$$

$N \to \inf$: Random functions with given variance and correlation length. (Or, 1D field with given 1-point and 2-point function)

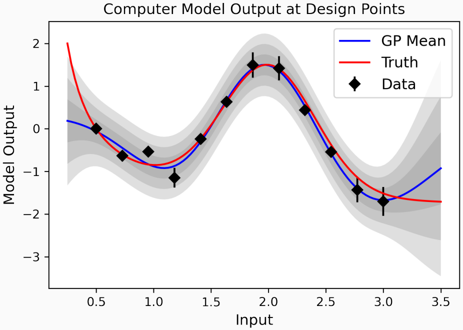
A class of non-parametic estimator: Gaussian emulators



Suppose we want to "interpolate" three points with some tolerance (black bars). Then, just select the subset of random functions that come close to these points.

→ an ensemble of random function forms a probabilistic inference of the underlying relation $y(x)$.

A class of non-parametic estimator: Gaussian emulators



Test on 1D scalar function $y(x)$. Easy generalization to scalar function with N-dim input $y(\vec{x})$.

All these can be formulated with tools of multivariate normal:

$$\begin{bmatrix} y(x') \\ y(x) \end{bmatrix} = \mathcal{N}\left(\mu = 0, \begin{bmatrix} K(x',x') & K(x',x) \\ K(x,x') & K(x,x) \end{bmatrix}\right)$$

Condition $y(x)$ on the training data $y(x_i) = y_i$

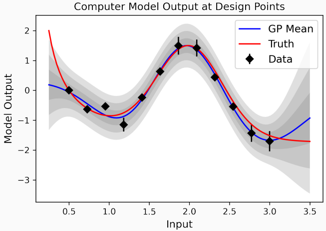
$$P(y(x')|y(x_i) = y_i) = \mathcal{N}(\mu, \sigma)$$

$$\mu = K(x', x_i)K^{-1}(x_i, x_j)y(x_j),$$

$$\sigma = K(x', x') - K(x', x_i)K^{-1}(x_i, x_j)K(x_j, x'))$$

Interpolate points from unknown functions with uncertainty quantification $y(x') = \mu(x') \pm \sigma(x')$

Interpolation uncertainty

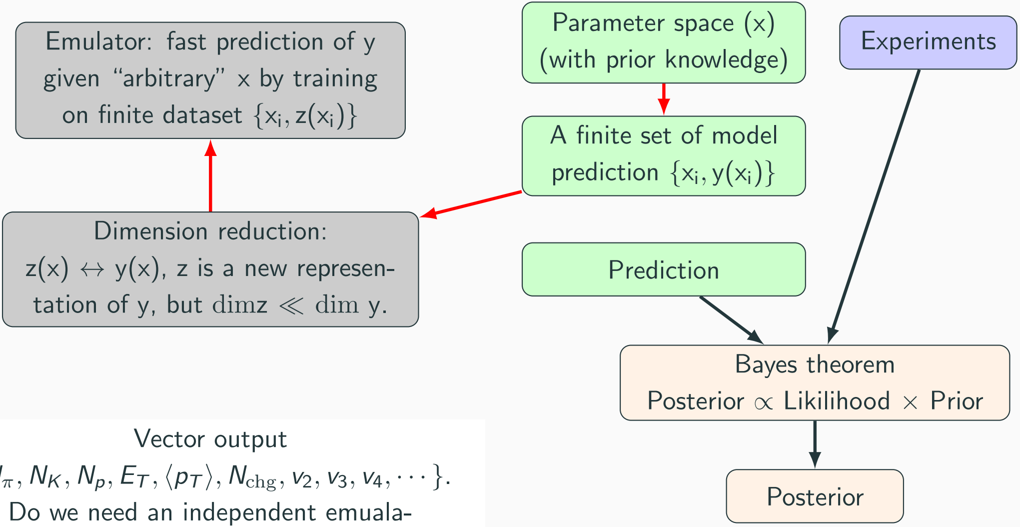


$$y(x) = \mu(x) \pm \sigma(x)$$

$$\Sigma_{ij} = \underbrace{\delta_{ij}[(\delta y_{\mathrm{stat}})_i^2 + (\delta y_{\mathrm{sys},0})_i^2] + \delta(y_{\mathrm{sys},\infty})_i \delta(y_{\mathrm{sys},\infty})_j + \delta(y_{\mathrm{sys},1})_i \delta(y_{\mathrm{sys},1})_j c(i,j;l)}_{\mathrm{Experimental}}$$
$$+ \ \sigma_{ij}^{\mathrm{emulator}} \longleftarrow \text{Interpolation uncertainty}$$
$$+ \ \sigma_{ij}^{\mathrm{theory}} \longleftarrow \text{Model/theory imperfection, still very hard}$$

In high-dimensional model, the interpolation uncertainty can actually be the dominant one!

For computational intensive models + high-dimensional output



Vector output

$\{N_\pi, N_K, N_p, E_T, \langle p_T \rangle, N_{\rm chg}, v_2, v_3, v_4, \cdots\}$.

Do we need an independent emualator to learn each "obs(params)"?

Dimensional reduction via Principal Component Analysis (PCA)

There are useful empirical correlations in the data. For example:

- Tune parameter to increase the initial-state energy density, then $N_{\pi,K,p}\uparrow$, $E_T\uparrow$, $N_{\mathrm{ch}}\uparrow$.
- Increase viscosity: $v_2\downarrow, v_3\downarrow, v_4\downarrow$.
- Given the same amount of initial energy: N_{ch} should anti-correlate with $\langle p_T\rangle$.

Clearly, we don't need less effective d.o.f. to describe these observable's dependence on input parameters.

PCA: use a few components from an emperical basis to represent data

If the set of functions that you care about can be approximated by keeping only a few terms, then this is a useful basis for expansion.

PCA: now we have a few hundreds' computation of $\vec{Y} = \{N_\pi, N_K, N_p, v_2, v_3, v_4, \cdots\}(\vec{p}_i), i = 1, 2, 3, \cdots$. Define the basis where new components are linearly independent of each other when averaged over all possible parameters:

$$O^T \mathrm{cov}(\vec{Y}, \vec{Y}) O \to \mathrm{diag}\{\lambda_1, \lambda_2, \cdots\}, \lambda_1 > \lambda_2 > \lambda_3 > \cdots .$$

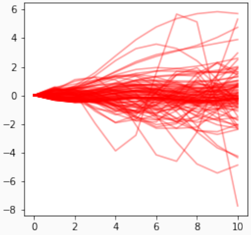
- The new basis $Z = O^T Y$ are linearly combinations of the original observables that are linearly uncorrelated from each other $\langle Z_i Z_j \rangle_p = \lambda_i \delta_{ij}$.
- Important: this does not remove non-linear correlation. One needs to check ensure the best performance.

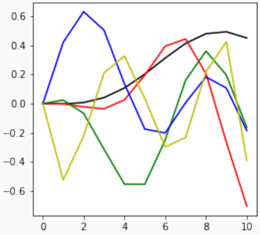
Dimensional reduction: simple demo

Consider a "model" with four parameters $0 < a, b, c, d < 1$, which generates 11 highly correlated outout labeled by $x = 0, 1, 2 \cdots 10$,

$$M(i) = ax^b \sin(cx + d)$$

Left: sample 100 sets of parameter (a, b, c, d) and plot the model outputs.
Right: the basis function corresponding to the first five principal components.



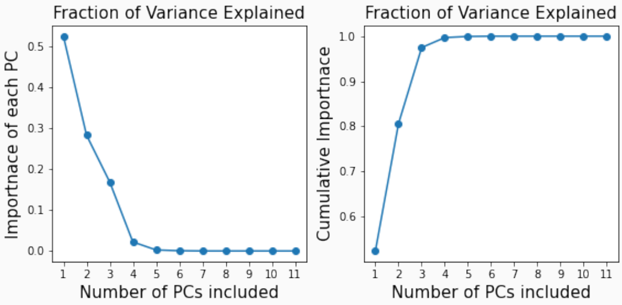


Dimensional reduction: simple demo

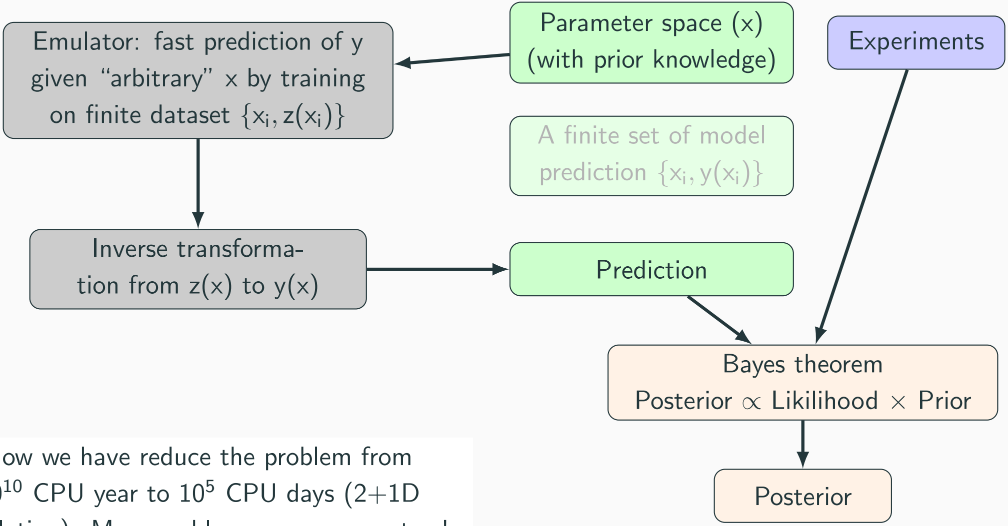
Consider a "model" with four parameters $0 < a, b, c, d < 1$, which generates 11 highly correlated outout labeled by $x = 0, 1, 2 \cdots 10$,

$$M(i) = ax^b \sin(cx + d)$$

Importance & cumulative importance of the PCs. Linear combination of the first five basis can describe the $M(i) = \sum_{i=1}^{5} p_i F_i(x) + \delta$ within the design range with good precision



The workflow of the emulator-assisted Bayesian analysis

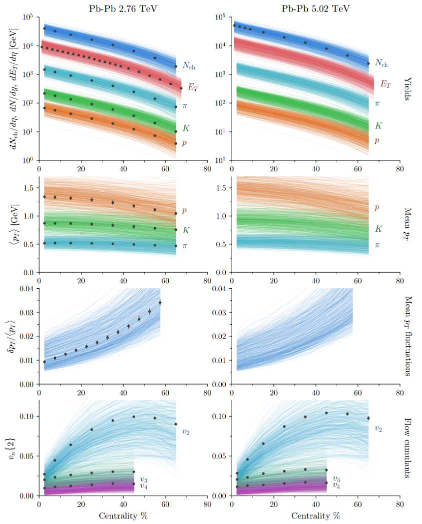


Now we have reduce the problem from 10^{10} CPU year to 10^5 CPU days (2+1D simulation). Manageable on supercomputers!

Parameters and observabels

Parameter	Description	Range
Norm	Normalization factor	8–20 (2.76 TeV) 10–25 (5.02 TeV)
p	Entropy deposition parameter	$-1/2$ to $+1/2$
σ_{fluct}	Multiplicity fluct. std. dev.	0–2
w	Gaussian nucleon width	0.4–1.0 fm
d^3_{min}	Minimum nucleon volume	0–1.7 fm^3
τ_{fs}	Free streaming time	0–1.5 fm/c
η/s hrg	Const. shear viscosity, $T < T_c$	0.1–0.5
η/s min	Shear viscosity at T_c	0–0.2
η/s slope	Slope above T_c	0–8 GeV^{-1}
η/s crv	Curvature above T_c	-1 to $+1$
ζ/s max	Maximum bulk viscosity	0–0.1
ζ/s width	Peak width	0–0.1 GeV
ζ/s T_0	Peak location	150–200 MeV
T_{switch}	Particlization temperature	135–165 MeV

[Jonah E. Bernhard Ph.D. dissertation]



Parameters and observabels

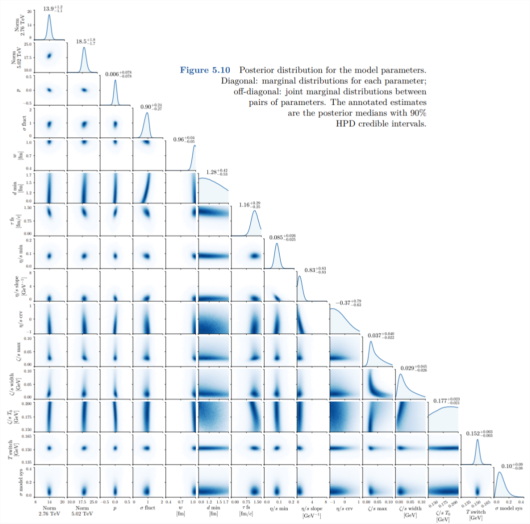
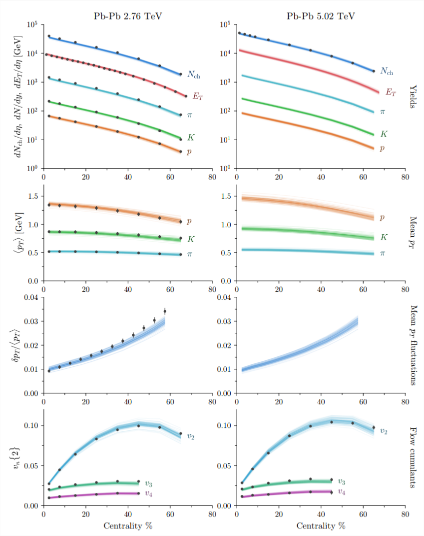
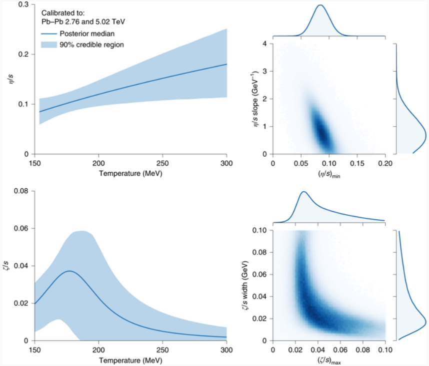


Figure 5.10 Posterior distribution for the model parameters. Diagonal: marginal distributions for each parameter; off-diagonal: joint marginal distributions between pairs of parameters. The annotated estimates are the posterior medians with 90% HPD credible intervals.



[Jonah E. Bernhard Ph.D. dissertation]

Shear and bulk viscosity



[Jonah E. Bernhard, J. Scott Moreland & Steffen A. Bass, Nature Physics volume 15, pages 1113–1117 (2019)]

$$\eta/s = (\eta/s)_{\text{min}} + (\eta/s)_{\text{slope}}(T - T_c)\left(\frac{T}{T_c}\right)^{(\eta/s)_{\text{curv}}}$$

$$\zeta/s = \frac{(\zeta/s)_{\text{max}}}{1 + (T - (\zeta/s)_{T_0})^2/(\zeta/s)^2_{\text{width}}}$$

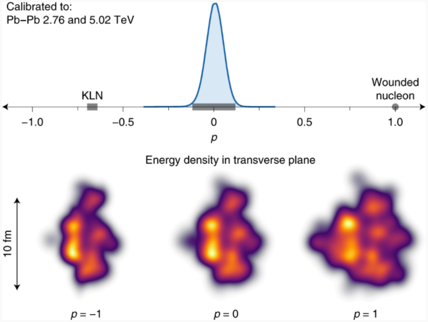
We are not really interested what the individual parameters in the parametrization are.

→ Marginalize over all of them and look at the 90% credible interval of $\eta/s(T)$ and $\eta/s(T)$.

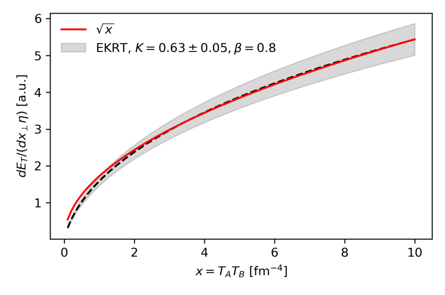
With high degree of confidence:

- QGP is strongly coupled $\eta/s = (1\cdots2)/(4\pi)$.
- QGP has a nonzero bulk viscosity!

Constrained initial condition



- The posterior suggests the data highly favors a specific type of energy deposition, $e(x,y) \propto \sqrt{T_A(x,y)T_B(x,y)} \propto$ local center-of-mass energy.
- This is numerically similar to certain models based on saturation physics, such as the EKRT model [PRC 93, 024907 (2016)].
- Wounded nucleon model, and KLN model (also saturation based) is disfavored.



Is this the end of story?

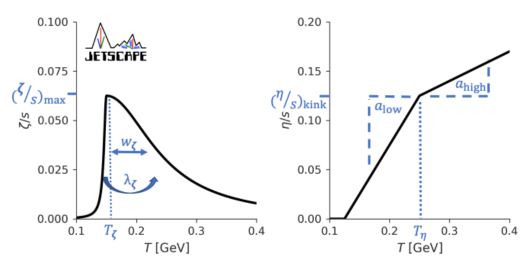
What is still missing?

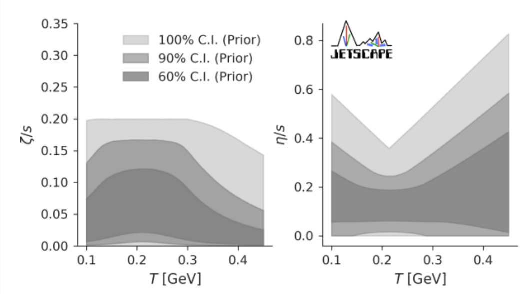
- Is the prior large enough?
- Is the paramerization general enough?
- How sensitive are the results to other model choices.
- How much confidence do we really have (need more validation and testing).

A careful revisit (JETSCAPE Collaboration): prior, model uncertainty, and closure test

Enlarge the prior distribution + more flexible paramerization

- Four-parameter each for $\eta/s(T)$ and $\zeta/s(T)$.
- The bulk viscosity does not have to be symmetric with respect to T_ζ.
- Shear does not necessarily approach minimum at T_c, and is allowed to decrease above T_c.





Model selection

One big problem of high-energy nuclear physics models are theoretical uncertainty.

Discrete model choices: (different basic assumptions, different approximations & truncation, different limit of the same theory, etc).

Bulk physics

- Use of different initial condition model.
- Hydro. v.s. full transport approach.
- Different schemes to particlize hydrodynamic fields into hadrons.

Jet physics

- Formula of in-medium jet evolution.
- Use of different bulk medium models.

Recent progress using Bayes factors: a concrete example

Partialize hydrodynamic fields into hadrons:

- Hydrodynamics fields $e(t, \vec{x}), u^\mu, \pi^{\mu\nu}, \Pi \to 10$ independent quantities ($\mu_b = 0$).
- Hadron momentum distribution $f_{\rm eq}(t, \vec{x}, \vec{p}) + \delta f_{\rm viscous}$ for each specie of hadrons.

The equilibrium part is known $f_{\rm eq} = 1/(e^{p\cdot u/T} \pm 1)$
To go from 10 numbers to $f_{\rm eq} + \delta f(p)$ largely depends on additional assumptions.

Different matching schemes from hydrodynamics to particle ensembles

- Grad 14-moment expansion:

$$\delta f(p) \propto A_\pi \pi^{\mu\nu} p_{\langle\mu} p_{\nu\rangle} + \Pi(A_T m_i^2 + A_E (p\cdot u)^2)$$

- 1^{st}-order Chapman-Enskog solution to the relaxation time approximation (RTA) Boltzmann equation.

$$\delta f \propto \frac{\pi_{\mu\nu} p^{\langle\mu} p^{\nu\rangle}}{2\beta_\pi (p\cdot u) T} + \frac{\Pi}{\beta_\Pi}\left(\frac{\mathcal{F}(p\cdot u)}{T^2} - \frac{\Delta_\mu \nu p^\mu p^\nu}{3(p\cdot u)T}\right)$$

- Modified f_{eq} approach Pratt-Torrieri-Bernhard/McNelis: rotate, stretch/squeeze, and scale the equilibrium distribution to match the viscous correction,

$$f_{\mathrm{eq}} + \delta f = \mathcal{Z} f_{eq}\left(p^i \to [(1+\frac{\Pi}{3\beta_\Pi})\delta_{ij} + \pi^{ij}]p_j,\, T \to T + \beta_\Pi^{-1}\Pi\mathcal{F}\right)$$

Details are complicated, but very different momentum and mass dependence of δf.

Use Bayes factor for model selection and averaging

- None of the above models is a first-principle QCD result. We can ask which one is prefered by data.

Bayes factor for comparing model "a" and "b": ratio of evidence

$$
\begin{aligned}
B_{M_a/M_b} &= \frac{P(y_{\rm exp}|M_a)}{P(y_{\rm exp}|M_b)} \\
P(y_{\rm exp}|M) &= \int \mathrm{Likilihood}(y_{\rm exp}|M, x_M)\mathrm{Prior}(x_M)dx_M, \ \text{for } M = M_a, M_b.
\end{aligned}
$$

Human interpretation:

$\log_{10}(B_{10})$	B_{10}	Evidence against H_0
0 to 1/2	1 to 3.2	Not worth more than a bare mention
1/2 to 1	3.2 to 10	Substantial
1 to 2	10 to 100	Strong
>2	>100	Decisive

[Robert E. Kass & Adrian E. Raftery (1995)]

Bayes factor for different particlization schemes

Result from the state-of-the-art model comparisons [JETSCAPE Collaboration, Phys.Rev.C 103 (2021) 5, 054904]

Model A	Model B	$\ln B_{A/B}$
Grad	CE	8.2 ± 2.3
Grad	PTB	1.4 ± 2.5
PTB	CE	6.8 ± 2.4

TABLE IV. A table of the logarithm of the Bayes factor $\ln B_{A/B}$ for each pair of viscous correction models and its integration uncertainty for the Grad, Chapman-Enskog (CE) and Pratt-Torrieri-Bernhard (PTB) viscous correction models.

Remember this table

$\log_{10}(B_{10})$	B_{10}	Evidence against H_0
0 to 1/2	1 to 3.2	Not worth more than a bare mention
1/2 to 1	3.2 to 10	Substantial
1 to 2	10 to 100	Strong
>2	>100	Decisive

- The Grad method (momentum expansion) is substantially favored over the PTB (modified equilibrium distribution).
- Both are decisive favored over the Chapman-Enskog solution of RTA Boltzmann equation[1].

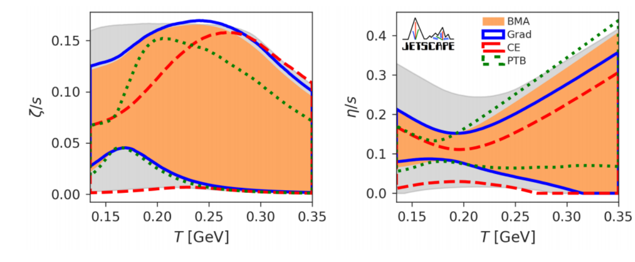
[1] This is mainly caused by their different ability to describe identified particle yield between π, K, p.

What are the impact on the QGP viscosities

Now that we have uncertainty in modeling choice / assumptions, we should update the uncertainty band of $\eta/s(T)$ and $\zeta/s(T)$ via marginalization.

Bayesian model averaging (MBA)

$$P_{\mathrm{BMA}}(x|y_{\mathrm{exp}}, \{M_i\}) = \sum_i \underbrace{P(x|y_{\mathrm{exp}}, M_i)}_{\text{Posterior for model ``}i\text{''}} \times \underbrace{P(y_{\mathrm{exp}}|M_i)}_{\text{Evidnece of model `}i\text{''}}$$



After model averaging (orange bands), the BMA posterior is dominated by the one with the highest evidence (Grad expansion).

[JETSCAPE Collaboration, Phys.Rev.Lett. 126 (2021) 24, 242301]

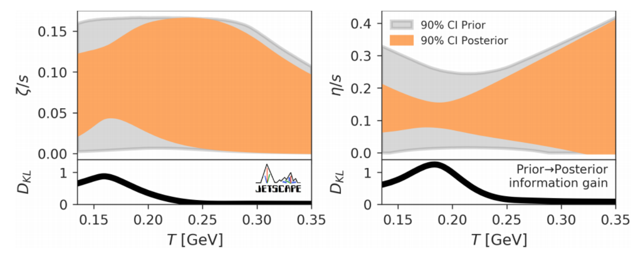
Quantify the information gain

- Note that we don't really learn anything new (compare to prior) at high temperature.
- To quantify this information gain, we use the "Kullback–Leibler divergence" (KL divergence, D_{KL}) to measure the functional distance between two distributions P_1 and P_2

$$D_{\mathrm{KL}}(P_1 \| P_2) \equiv \int dx P_1(x) \ln \frac{P_1(x)}{P_2(x)}, \text{ we take } P_1 = \text{Posterior}, \; P_2 = \text{Prior}.$$

- If $D_{\mathrm{KL}} = 0$, then the posterior is the same as our prior belief, nothing new...
- $D_{\mathrm{KL}} > 0$ signatures information gain from experimental data.
- What observable may grant increased sensitivity at high temperature?

Quantify the information gain



- $D_{KL}(T > 0.25\mathrm{GeV}) \approx 0$, little sensitivity to QGP transport properties at high T.
- Most information gain in $0.145 < T < 0.225$ GeV.

Reasons?

- Medium expand very fast and spend little time in the high-temperature region.
- With fast expansion, final observable is only sensitive to an "averaged η/s":

$$(\eta/s)_{\mathrm{eff}} = \frac{\int_{T_{\mathrm{sw}}}^{T_{\mathrm{max}}} \eta/s(T)/T^{\alpha} dT}{\int_{T_{\mathrm{sw}}}^{T_{\mathrm{max}}} 1/T^{\alpha} dT}$$

[Jean-Francois Paquet, Steffen A. Bass, Phys. Rev. C 102, 014903 (2020)]

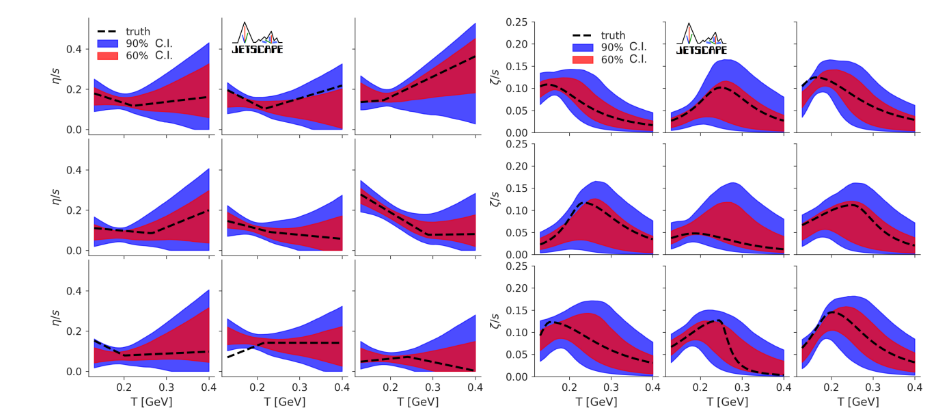
Closure test: how much can we trust the analysis in the best scenario?

A closure test

- Use the framework to calibrate on pseudodata, which is model calculation with known parameters.
- Compare the posterior to the true values.

This is a very conclusive test if the model is perfect. In the presence of model uncertainty, this is a weaker test.

Closure test on the extaction of bulk viscosities



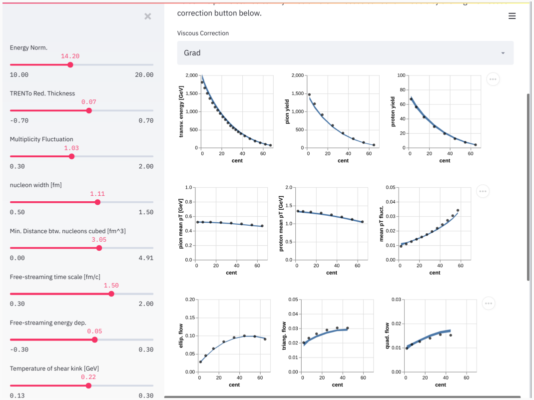
We generate nine different set of pseudodata:

- Dashed line: the true answer to $\eta/s(T)$(left) and $\zeta/s(T)$ (right).
- Blue/red bands: 90% & 60% credible region.
- The statistical analysis (if the model is perfect) works as expected.

How to publish the full results: publish the emulator online

Scientific papers are 2D objects. May not always be the best option to publish something that lives in higher dimension.

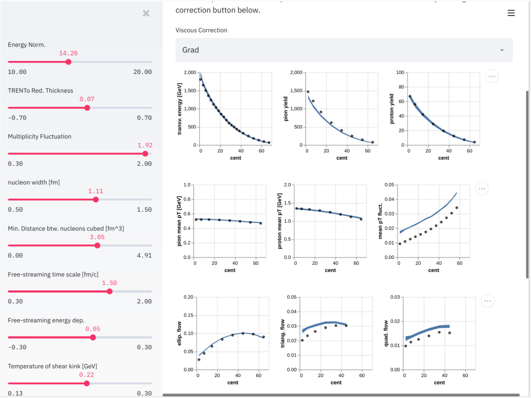
Checkout this interactive page `https://jetscape.org/sims/`.
Use the slides to see how each observable response to the change of each parameter



How to publish the full results: publish the emulator online

Scientific papers are 2D objects. May not always be the best option to publish something that lives in higher dimension.

Checkout this interactive page `https://jetscape.org/sims/`.
Use the slides to see how each observable response to the change of each parameter



How to publish the full results: publish the emulator online

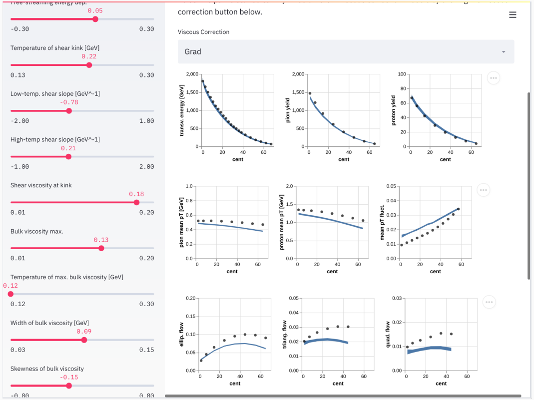
Scientific papers are 2D objects. May not always be the best option to publish something that lives in higher dimension.

Checkout this interactive page `https://jetscape.org/sims/`.
Use the slides to see how each observable response to the change of each parameter

How to publish the full results: publish the emulator online

Scientific papers are 2D objects. May not always be the best option to publish something that lives in higher dimension.

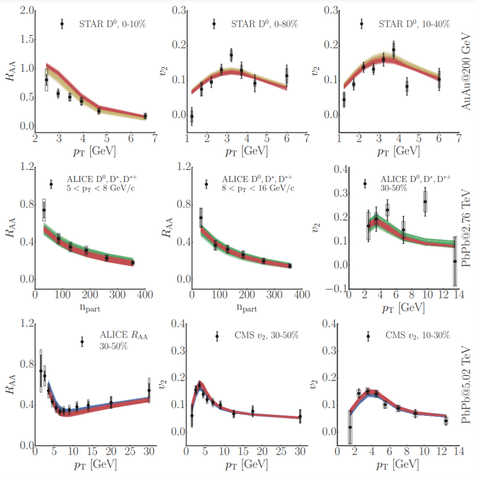
Checkout this interactive page `https://jetscape.org/sims/`.
Use the slides to see how each observable response to the change of each parameter



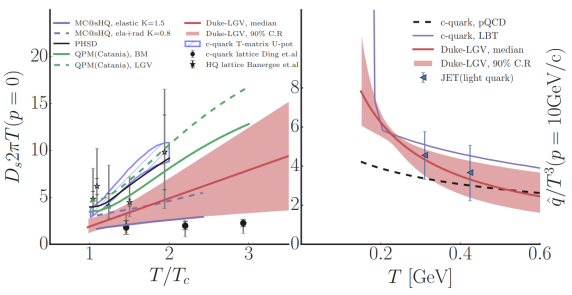
Application to jet quenching and jet transport coefficients

Inference of jet transport parmater from a single model

The first Bayesian analysis applied to the heavy-quark sector [Yingru Xu et al PRC97, 014907 (2018)]. Heavy quark dynamics modeled by a Langevin (drag and diffusion) process with recoil from radiated gluon.

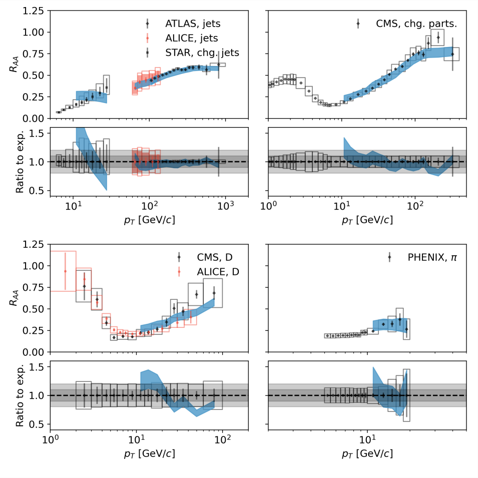


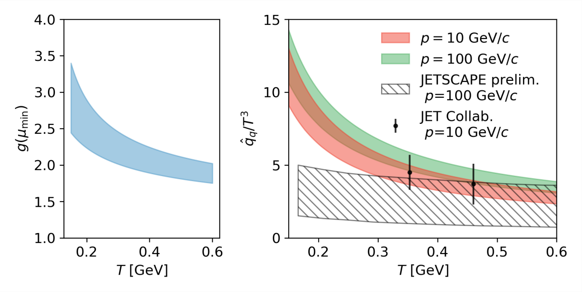
Global analysis on D-meson R_{AA} and momentum anisotropy.



Inference of jet transport parmater from a single model

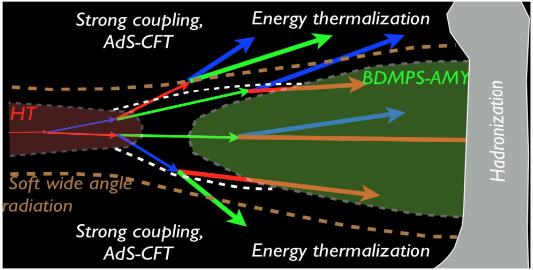
Global analysis on heavy / light hadron and full jet quenching at both RHIC and LHC [W Ke & X-N Wang JHEP 05 (2021) 041]. Consistency among jet and hadron observable.



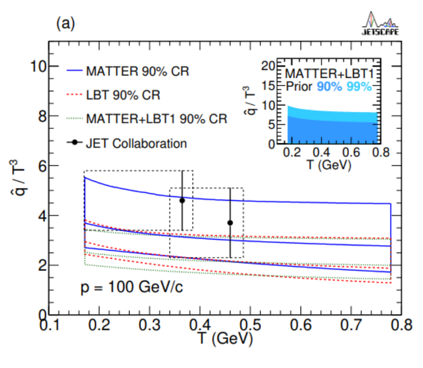


Recent Bayesian analysis using multi-stage model

- Different evolution equations / interaction mechanisms in different regions. High/low-virtuality region, High/low-energy region.
- Enable testing multiple model choices/combinations in the same environment.



[Figure credit to Abhijit Majumder]



Posterior of $\hat{q}$. Model-A(MATTER) or model-B (LBT) applied to the entire phase-space v.s. Matching model A+B in the phase-space. [JETSCAPE Collaboration PRC 104, 024905 (2021)]

Final remarks on functional prior & Summary

What is a reasonable prior?

Up to this point, we have always assume the parameter distribution has a uniform prior $P(x) = \frac{1(x_{\rm min}<x<x_{\rm max})}{x_{\rm max}-x_{\rm min}}$. Is the prior trivial? Not really...

Consider we use a, b, c to parameterize a function, such as $\zeta/s(T) = \frac{a}{(T-b)^2+c^2}$. An observable with a rather simple response: $\mathrm{obs} \propto \int_{T_1}^{T_2} \zeta/s(T) dT$

- a, b, c has independent, uniform distribution as Prior.
- The quantity of physical importance $\zeta/s(T)$ varies highly non-linearly within the design space! So is the observable!
- No matter what the posterior is, such a parametrization always suggests $\zeta(T \gg b)/s \to 0$.

When we try to extract unknown functions, such as $\eta/s(T), \hat{q}(x, Q^2), \mathrm{PDF}(x, Q_0)$

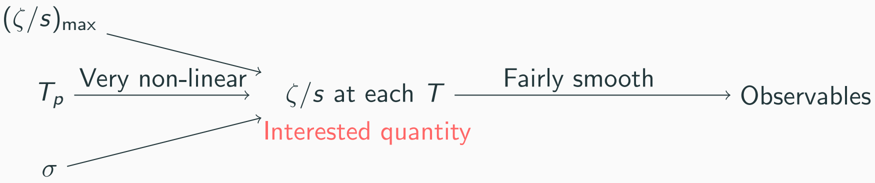
the parametrization itself is a very strong and informative prior!

Form of parametrization is a strong assumption on prior

In the Bayes extraction of continuous functions $\hat{q}(T)$, $\hat{q}(T, E)$, $\hat{q}(T, E, Q)$, and $\eta/s(T)$, $\cdots$.

- A given parametrization imposes strong correlation among the value of the function at different input.
- Parameters with clear physical meaning may not be “easy” for machine learning (emulator). For example:

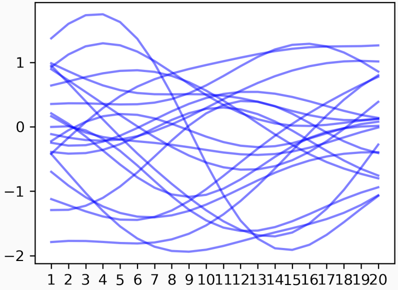
$$\zeta/s(T) = \frac{(\zeta/s)_{\max}}{1+(T/T_p-1)^2/\sigma^2}, \quad \Delta\hat{q} = \frac{AT^3}{(1+(E/aT)^p)(1+(T/bT_c)^q)}$$ [2]



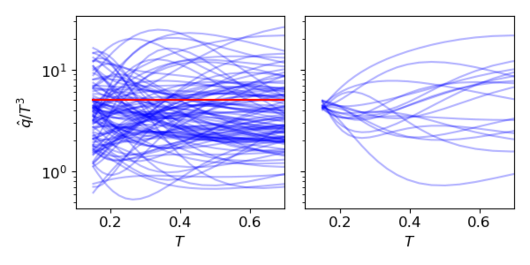
[2] Something more complicated that I tried in my dissertation

What can be a reasonable prior of unknown function

- We don't want to exclude any possible case.
- Assume there is no abrupt change v.s. input (with proper redefinition of input/output).
  - $f(x) = 1/x^{\lambda} = e^{-\lambda u}, u = \ln(x)$.
  - $g(x) = ax^3(1 + b\ln(x)) \to \tilde{g}(x) = g(x)/x^3$.
- Remember the Gaussian process that generates random functions?



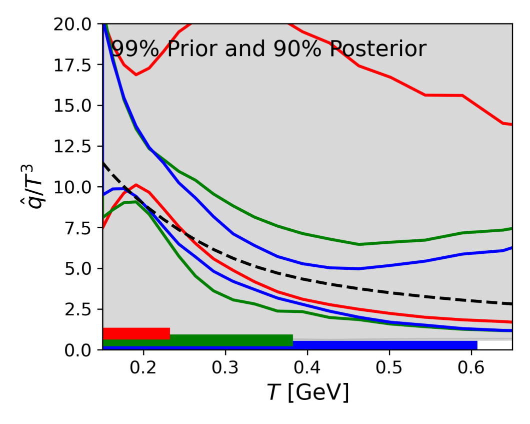
Use random function as prior



$$\langle \delta y(T) \delta y(T') \rangle = \sigma_0^2 \exp\{-\frac{(\ln T - \ln T')^2}{2L^2}\}$$

- One can control the range of variation with σ_0 and control the flexibility with L
- We don't really exclude any reasonable function. Any function is possible, though come with different probability.
- Data that determine the function in one region $\sim T$ does not affect the prior in other regions $\sim T', (|\ln(T/T')| \gg 1)$.

Test with a toy model $\Delta E/E \propto \int_{T_{\min}}^{T_{\max}} \hat{q}(T)/T^3 \frac{dT}{T}$



- The constraining power gradually increases with pseudodata covering higher temperature regions.
- This prevents tension from different collision energy due to a specific form of parametrization.

Not yet tested for more than 1D function, such as $\hat{q}(E, T)$, but should be straightforward.

Random function as functional prior

What is the prior of random function?

$$e^{-\frac{1}{2}\int dxdx' f(x)K^{-1}(x,x';L,\sigma)f(x')}$$

And the posterior:

$$e^{-\frac{1}{2}\int dxdx' f(x)K^{-1}(x,x';L,\sigma)f(x') - \ln(\mathrm{Likelihood}[f(x),X_i,y_{\mathrm{exp}};\mathcal{M}])}$$

Marginalization or prediction,

$$P(X_i) = \int D[f(x)]O[f(x),X_i;\mathcal{M}]e^{-\frac{1}{2}\int dxdx' f(x)K^{-1}(x,x';L,\sigma)f(x') - \ln(\mathrm{Likelihood}[f(x),X_i,y_{\mathrm{exp}};\mathcal{M}])}$$

For certain problems, one may also borrow ideas from field theory to analyze the posterior (information field theory IFT `https://wwwmpa.mpa-garching.mpg.de/~ensslin/research/research_IFT.html`)

Physics summary

With advanced statistical tools and physical modeling, we learned a lot in the past decade,

- Phenomenological constrained QCD EoS at high T corroborate lattice calculation.
- Temperature dependent shear and bulk viscosity $\rightarrow$ strongly-coupled nature of QGP and scale violation.
- Jet transport parameter in hot/cold QCD medium $\rightarrow$ drastic difference in color confined/deconfined matter.
- Constrained initial geometry of nuclear collisions.
- $\cdots$

Statistics summary

- Learning hot and cold nuclear matter from experimental data poses challenges
  - Complex multi-stage model.
  - Large parameter space & large and diverse dataset.
- The use of model emulator and dimension reduction techniques are essential to perform statistical analysis on these complex models.
- Bayesian inference provide a systematic way to incorporate both experimental and model uncertainty. Necessary for reliable extraction of interest QCD properties.
- Be careful with prior. Functional parametrization itself is a highly informative prior!
- Pay attention to model uncertainties. Use Bayes factors & model averaging to compare & combine various models.

Questions?