

Uncovering the temperature dependence of the QGP jet transport parameter using the information field

The Berkeley Symposium on Hard Probes and Beyond, LBNL, Aug 18-19, 2022

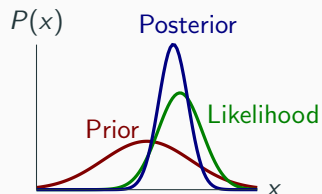
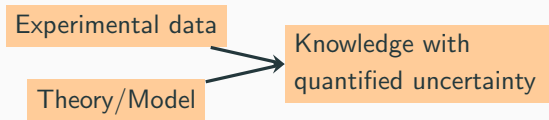
Man Xie (SCNU), Weiyao Ke (LANL), Hanzhong Zhang (CCNU), and Xin-Nian Wang (LBNL).

arXiv:2206.01340 and arXiv:XXXX.XXXX

Aug 18, 2022



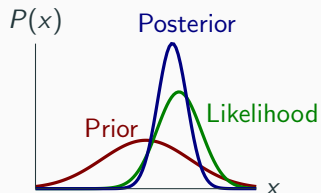
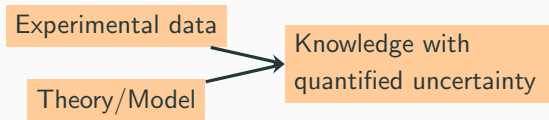
Statistical inference of physical parameters in high-energy nuclear physics



- The problem has been pretty well formulated with the Bayes' Theorem:

$$\underbrace{P_1(p|\text{exp})}_{\text{Posterior}} = \underbrace{E^{-1}(\text{exp})}_{\text{Evidence}} \underbrace{L(\text{exp}|p)}_{\text{Likelihood}} \underbrace{P_0(p)}_{\text{Prior}}$$

Statistical inference of physical parameters in high-energy nuclear physics



- The problem has been pretty well formulated with the Bayes' Theorem:

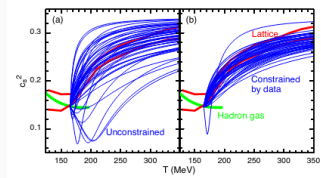
$$\underbrace{P_1(p|\text{exp})}_{\text{Posterior}} = \underbrace{E^{-1}(\text{exp})}_{\text{Evidence}} \underbrace{L(\text{exp}|p)}_{\text{Likelihood}} \underbrace{P_0(p)}_{\text{Prior}}$$

- Prior is often less “popular” than posterior, **but it can bias posterior when “Likelihood” is weak.**
 - For finite # of parameters, enlarge the prior to maintain the generality.
 - **But what is a general prior for functional quantities with infinite DoF.**

In fact, many interested quantities are functional objects

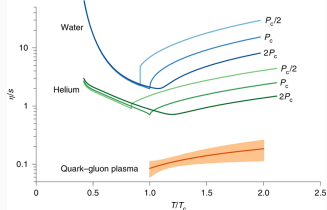
QCD Equation of State

PRL114(2015)202301



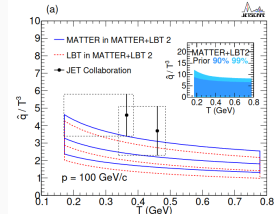
QGP shear and bulk viscosity

Nat.Phys.15(2019)111365



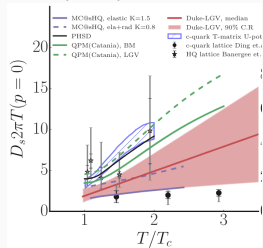
QGP jet transport parameter

PRC 104 (2021) 024905

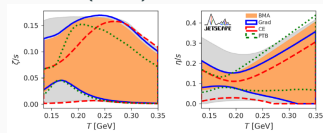


Heavy-quark diffusion coefficient

PR97(2018)014907



PRL126(2021)242301



Parton distribution functions

PRD96(2017)014015

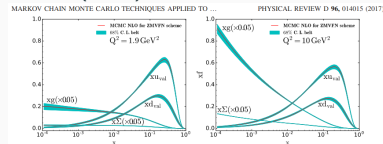


FIG. 11. PDFs of valence quarks ($u_{val}, d_{val}, s_{val}$), sea quarks ($\Sigma = \bar{s} + \bar{u}^2 + \bar{s} + \bar{c} + \bar{t}$, with $\bar{c} = 0$ for $Q^2 < m_c^2$), and gluon ($xy f(x)$) using MCMC at the scale $Q^2 = 1.9$ GeV² (left) and $Q^2 = 10$ GeV² (right) obtained with the ZMVFN scheme. The gluon and the sea distributions are scaled down by a factor of 20. The experimental uncertainties (68% confidence limit as defined in the text from the probability density of PDFs) are represented by the green-shaded region.

$$\text{Obs} = \text{Model}[F(x), \dots]$$

Our methods so far: use explicit parametrizations for $F(x)$

QGP shear & bulk viscosity

Duke: Nat. Phys. 15(2019)111365

$$\frac{\eta}{s}(T) = a + b(T - T_c) \left(\frac{T}{T_c} \right)^c$$
$$\frac{\zeta}{s}(T) = \frac{a}{1 + \left(\frac{T-b}{c} \right)^2}$$

JETSCAPE PRC103(2021)054904

$$\frac{\eta}{s}(T) = a + b(T - d)\Theta(d - T) + c(T - c)\Theta(T - c)$$
$$\frac{\zeta}{s}(T) = \frac{a}{1 + \left(\frac{T-b}{c[1+d\text{sign}(T-b)]} \right)^2}$$

Then, apply Bayes theorem to individual parameters.

Jet transport parameter

JETSCAPE PRC 104 (2021) 024905

$$\frac{\hat{q}}{T^3} = \# \left\{ a \frac{\ln(E/\Lambda) - \ln b}{\ln^2(E/\Lambda)} + c \frac{\ln(E/T) - \ln d}{\ln^2(ET/\Lambda^2)} \right\}$$
$$\frac{\hat{q}}{T^3} = \# \left\{ \Theta(Q - Q_0) a \frac{\ln(Q/\Lambda) - \ln(Q_0/\lambda)}{\ln^2(Q/\Lambda)} + c \frac{\ln(E/T) - \ln d}{\ln^2(ET/\Lambda^2)} \right\}$$

Parton distribution/fragmentation function

JAM22, 2205.00999

We generically parametrize the collinear functions $h_1(x)$, $F_{FT}(x, x)$, $H_1^{\perp(1)}(z)$, $\tilde{H}(z)$ as

$$F^q(x) = \frac{N_q x^{a_q} (1-x)^{b_q} (1 + \gamma_q x^{\alpha_q} (1-x)^{\beta_q})}{\text{B}[a_q+2, b_q+1] + \gamma_q \text{B}[a_q+\alpha_q+2, b_q+\beta_q+1]},$$

$\times k_T$ parametrization.

What can go wrong with explicit parametrizations?

Jet quenching as an example: parton energy loss is sensitive an integrated quantity

$$\int_{\tau_0}^{\infty} d\tau \alpha_s \hat{q} \tau \ln \frac{E}{\tau m_D^2} \rightarrow \int_{T_c}^{T_{\max}} \frac{1}{2} \frac{d\tau^2}{dT} \ln \frac{E}{\tau m_D^2} \hat{q}(T) dT$$

Explicit parametrization $\hat{q}(T; a, b, c, d \dots)$

What can go wrong with explicit parametrizations?

Jet quenching as an example: parton energy loss is sensitive an integrated quantity

$$\int_{\tau_0}^{\infty} d\tau \alpha_s \hat{q} \tau \ln \frac{E}{\tau m_D^2} \rightarrow \int_{T_c}^{T_{\max}} \frac{1}{2} \frac{d\tau^2}{dT} \ln \frac{E}{\tau m_D^2} \hat{q}(T) dT$$

Explicit parametrization $\hat{q}(T; a, b, c, d \dots)$

- Introduce long-range correlations: change certain parameter affect $\hat{q}(T)$ at all T .
- **Undermine the generality**, e.g., RHIC data constraining $\hat{q}(T \lesssim 0.35 \text{ GeV})$ also restricts the prior at high- T . May introduce tension combining with LHC data ($T \lesssim 0.5 \text{ GeV}$).

What can go wrong with explicit parametrizations?

Jet quenching as an example: parton energy loss is sensitive an integrated quantity

$$\int_{\tau_0}^{\infty} d\tau \alpha_s \hat{q} \tau \ln \frac{E}{\tau m_D^2} \rightarrow \int_{T_c}^{T_{\max}} \frac{1}{2} \frac{d\tau^2}{dT} \ln \frac{E}{\tau m_D^2} \hat{q}(T) dT$$

Explicit parametrization $\hat{q}(T; a, b, c, d \dots)$

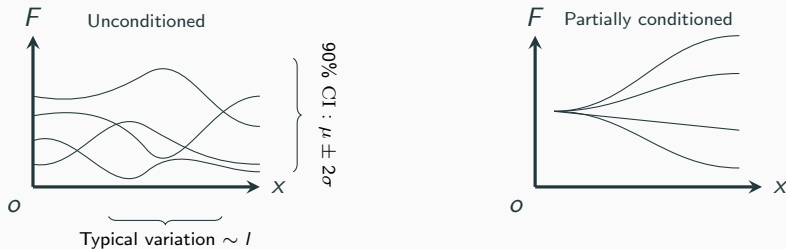
- Introduce long-range correlations: change certain parameter affect $\hat{q}(T)$ at all T .
 - **Undermine the generality**, e.g., RHIC data constraining $\hat{q}(T \lesssim 0.35 \text{ GeV})$ also restricts the prior at high- T . May introduce tension combining with LHC data ($T \lesssim 0.5 \text{ GeV}$).
- Often contain **high-degree of nonlinear correlations**
 - Bayesian analysis in HIC heavily relies on Machine-learning (ML) acceleration, i.e., a fast interpolation of $R_{AA}(a, b, c, d)$:
$$a, b, c, d \xrightarrow[\text{parametrization}]{\text{Complicated}} \hat{q} \xrightarrow[\text{models}]{\text{well-behaved}} R_{AA} \dots$$

This looks unnecessary and troublesome for ML.

Functional inference from an *Information Field (IF)* perspective

Treat $\hat{q}(T)$ as an “unconditioned random field”:

- Gaussian random field, $\langle F(x) \rangle = \mu(x)$, $\langle \delta F(x) \delta F(x') \rangle = C(x, x')$, linear correlations only.
- A common correlation function: $C(x, x') = \sigma^2 \exp \left\{ -\frac{|x-x'|^2}{2L^2} \right\}$, infinite-degree smoothness.



★ μ, L, σ controls what type of functions are favored: $P_0[F] = e^{-\frac{1}{2} \int dx dx' \delta F(x) C^{-1}(x, x') \delta F(x')}$

★ When $F(x)$ is constrained at x_1 , prior at x_2 is “unaffected” for $|x_1 - x_2| \gg l$.

The essence: 1) get rid of long-range correlation; 2) a smooth control of $\hat{q}(T)$

Statistical inference problem from field perspective, not really new...

Field-theory approach to infer PDF from N -point function PRL77(1996)4693, hep-ph/9808474v1

Field Theories for Learning Probability Distributions

William Bialek,¹ Curtis G. Callan,² and Steven P. Strong¹

¹NEC Research Institute, 4 Independence Way, Princeton, New Jersey 08540

²Department of Physics, Princeton University, Princeton, New Jersey 08544

(Received 25 July 1996)

Imagine being shown N samples of random variables drawn independently from the same distribution. What can you say about the distribution? In general, of course, the answer is nothing, unless you have some prior notions about what to expect. From a Bayesian point of view one needs an *a priori* distribution on the space of possible probability distributions, which defines a scalar field theory. In one dimension, free field theory with a normalization constraint provides a tractable formulation of the problem, and we discuss generalizations to higher dimensions. [S0031-9007(96)01804-2]

Functional statistical inference of parton distributions

Vipul Periwal

Department of Physics, Princeton University, Princeton, New Jersey 08544

Bialek, Callan and Strong have recently given a solution of the problem of determining a continuous probability distribution from a finite set of experimental measurements by formulating it as a one-dimensional quantum field theory. This report applies an extension of their formalism to the inference of functional parton distributions from scattering data.

The prior distribution, $P[Q(x)]$, should capture our prejudice that the distribution $Q(x)$ is smooth, so $P[Q(x)]$ must penalize large gradients, as in conventional field theories. To have a field variable $\phi(x)$ that takes on a full range of real values ($-\infty < \phi < \infty$), we write

$$Q(x) = \frac{1}{\ell_0} \exp[-\phi(x)], \quad (4)$$

where ℓ_0 is an arbitrary length scale. Then we take ϕ to be a free scalar field with a constraint to enforce normalization of $Q(x)$. Thus $\phi(x)$ is chosen from a probability distribution

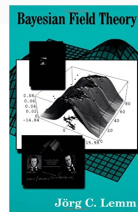
$$P_\ell[\phi(x)] = \frac{1}{Z} \exp\left[-\frac{\ell}{2} \int dx (\partial_x \phi)^2\right] \times \delta\left[1 - \frac{1}{\ell_0} \int dx e^{-\phi(x)}\right], \quad (5)$$

physics/9912005v3

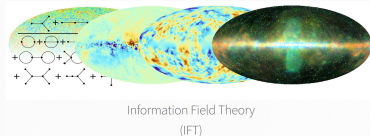
Bayesian Field Theory Nonparametric Approaches to Density Estimation, Regression, Classification, and Inverse Quantum Problems

Jörg C. Lemm^{*}

Institut für Theoretische Physik I
Universität Münster
Wilhelm-Klemm-Str.9
D-48149 Münster, Germany



Applications in astrophysics (Information Field Theory Group at the Max-Planck-Institute for Astrophysics)



Information field theory (IFT) is information theory, logic under uncertainty, applied to fields. A field can be any quantity defined over some space, such as the air temperature over Europe, the magnetic field strength in the Milky Way, or the matter density in the Universe. IFT describes how data and knowledge can be used to infer field properties. Mathematically it is a statistical field theory and exploits many of the tools developed for such. Practically, it is a framework for signal processing and image reconstruction. IFT is fully Bayesian. How else can infinitely many field degrees of freedom be constrained by finite data? It can be used without the knowledge of Feynman diagrams. There is a full toolbox of methods. It reproduces many known well working algorithms. This should be reassuring. And, there were certainly previous works in a similar spirit, like Bayesian Field Theory (BFT). See below for IFT & BFT publications and previous works. Anyhow, in many cases IFT provides novel rigorous ways to extract information from data.

▶ link

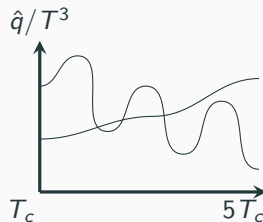
Apply to determine the temperature-dependent jet transport parameter

The random field:

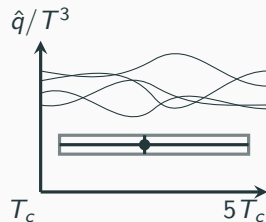
$$F(x) = \ln(\hat{q}/T^3).$$
$$x = \ln(T)$$

- $\ln(\hat{q})$ ensures positivity. Not a Gaussian random field anymore.
- $\langle F(x) \rangle = \text{const. } \mu$.
- $x = \ln T$: no special scales at high temperature.

★ In principle, μ, σ, L are arbitrary and to be marginalized by Bayes Theorem, but many choices fails our prior belief:



- ✗ L too short, priors favors "oscillating" functions.
- ✗ L cannot be too large, or $\hat{q}/T^3 \approx \text{const.}$

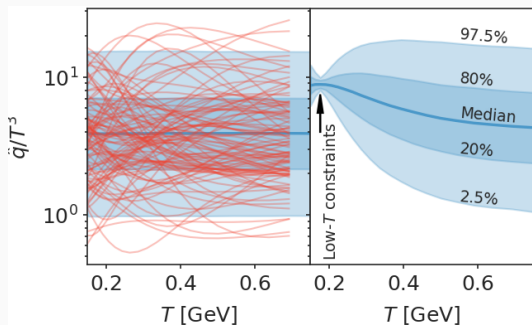


- ✗ 90% bands $\mu \pm 2\sigma$ does not cover the results of, e.g., $\text{const-}\hat{q}/T^3$ analysis.

We will fix $\mu = 1.36$, $\sigma = 0.70$ (translates into $\langle \hat{q}/T^3 \rangle = 5$, $\sigma\{\hat{q}/T^3\} = 4$), and $L = \ln 2$.

The choices are indeed subtle, we will come back to this with a toy model.

The information field prior of \hat{q}/T^3 .



Left: prior range of random field $0.9 \lesssim \hat{q}/T^3 \lesssim 15$, covers most of previous studies. **Red lines**: 100 random realizations.

Right: even if low- T \hat{q} is tightly constrained, high-temperature prior is “unaffected” (other than smoothness requirement).

The model: NLO + Higher-Twist (H-T) modified fragmentation

- Higher-twist gluon emission spectra

$$\frac{dN_{q,g \rightarrow g}}{dzdk_T^2} = \frac{\alpha_s C_A}{2\pi} P_{q,g \rightarrow g}(z) \int \frac{p \cdot u}{p^0} d\Delta\tau \int \frac{dk_T^2}{k_T^2 (k_T^2 + \mu_D^2)} \hat{q}_{F,A}(\tau, \vec{r} + \Delta\vec{n}) 4 \sin^2 \left[\frac{k_T^2 \Delta\tau}{4z(1-z)E} \right].$$

- Energy loss of the leading parton: $\Delta z = \int z \frac{dN_{q,g \rightarrow g}}{dzdk_T^2} dzdk_T^2$.
- Average number of induced soft gluon emission: $\langle N_g^d \rangle = \int \frac{dN_{q,g \rightarrow g}}{dzdk_T^2} dzdk_T^2$.
- Build the modified fragmentation function in the QGP:

$$D_{h/i}^{AA}(z_i, \mu^2) = (1 - e^{-\langle N_g \rangle}) \left[\frac{z'}{z} D_{h/i}^{\text{vac}}(z', \mu^2) + \langle N_g \rangle \frac{z_g'}{z} D_{h/g}^{\text{vac}}(z_g', \mu^2) \right] + e^{-\langle N_g \rangle} D_{h/i}^{\text{vac}}(z, \mu^2)$$

$p + p$: RMP59(1987)465, RMP67(1995)157, CT14 PDF PRD95(2017)034003 with KKP FF NPB582(2000)514

$A + A$: Nuclear shadowing from EPPS16 EPJC77(2017)163 + b -dependence JPG37(2010)105109, PRC61(2000)044904, PRC69(2004)034908, Modified FF from higher-twist PRL98(2007)212301, PRL103(2009)032302, PRC70(2004)031901

Single-hadron spectra & modifications

$$\frac{d\sigma^h}{dy_h d^2 p_T^h} = \sum_{abcd} \int dx_a dx_b f_a(x_a, \mu^2) f_b(x_b, \mu^2) \frac{1}{\pi z_c} \frac{d\sigma_{ab \rightarrow cd}}{d\hat{t}} D_{h/c}(z_c, \mu^2).$$

Di-hadron spectra & modifications

$$\frac{d\sigma^{hh}}{[dy d^2 p_T]_{h_1} [dy d^2 p_T]_{h_2}} = \sum_{abcd} \int dz_c dz_d f_a(x_a, \mu^2) f_b(x_b, \mu^2) \frac{x_a x_b}{\pi z_c^2 z_d^2} \frac{d\sigma_{ab \rightarrow cd}}{d\hat{t}} D_{h_1/c}(z_c, \mu^2) D_{h_2/d}(z_d, \mu^2) \delta^{(2)}(\vec{p}_T^c + \vec{p}_T^d).$$

γ -hadron spectra & modifications

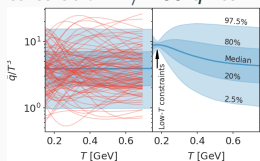
$$\frac{d\sigma^{\gamma h}}{[dy d^2 p_T]_{\gamma} [dy d^2 p_T]_h} = \sum_{abd} \int dz_d f_a(x_a, \mu^2) f_b(x_b, \mu^2) \frac{x_a x_b}{\pi z_d^2} \frac{d\sigma_{ab \rightarrow \gamma d}}{d\hat{t}} D_{h/d}(z_d, \mu^2) \delta^2(\vec{p}_T^{\gamma} + \frac{\vec{p}_T^h}{z_d}).$$

More recently: modification of fragmentation functions in γ -jet

Calibration to a global data set

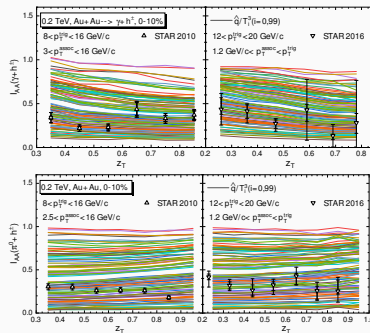
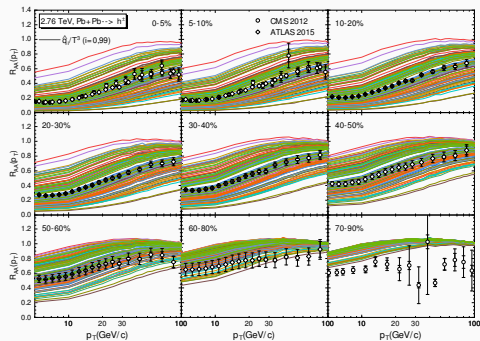
Training data: NLO+HT

calculation w/ 100 \hat{q} realizations.



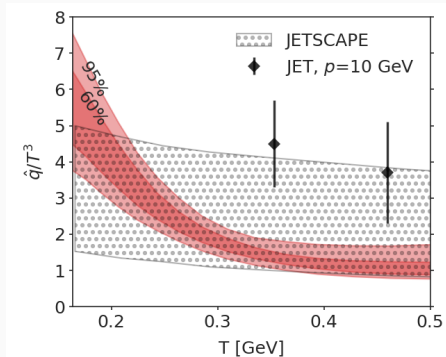
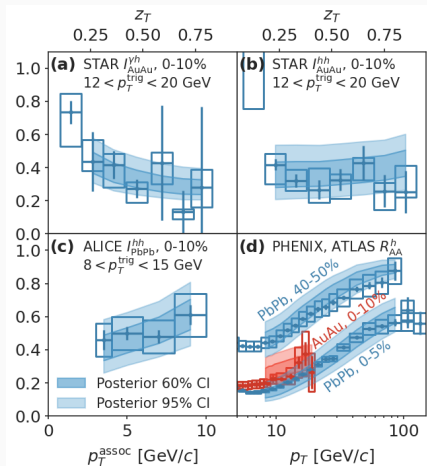
Experimental data:

- R_{AA}^0 in 0-5%, 5-10%, 10-20%, 20-30%, 30-40%, 40-50% in Au+Au @0.2 TeV PHENIX, PRL101(2008)232301.
- $R_{AA}^{h^\pm}$ in 0-5%, 5-10%, 10-20%, 20-30%, 30-40%, 40-50%, in Pb+Pb @2.76 TeV ATLAS, JHEP09(2015)050.
- $R_{AA}^{h^\pm}$ in 0-5%, 5-10%, 10-20%, 10-30%, 20-30%, 30-50% in Pb+Pb@5.02 TeV CMS, JHEP04(2017)039
- $I_{AA}^{\gamma h^\pm}$ and $I_{AA}^{\pi^0 h^\pm}$ in 0-10% Au+Au@0.2 TeV STAR, PLB760(2016)689-696.
- $I_{AA}^{h^\pm h^\pm}$ in 0-10% Pb+Pb@2.76 TeV ALICE, PRL108(2012)092301; CMS, NPA904-905(2013)451-454



and more ...

Calibration to the real-world data



Red: this work.

JET PRC90(2014)014909, JETSCAPE PRC104(2021)024905

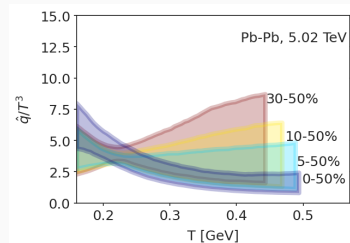
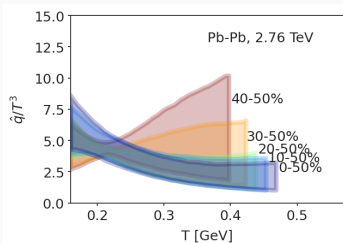
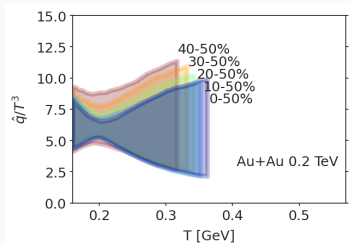
◁ Only a subset of all calibrated data is shown.

△ A strong T -dependence. Where does it come from?

Progressive constraining power from low T to high T (using R_{AA})

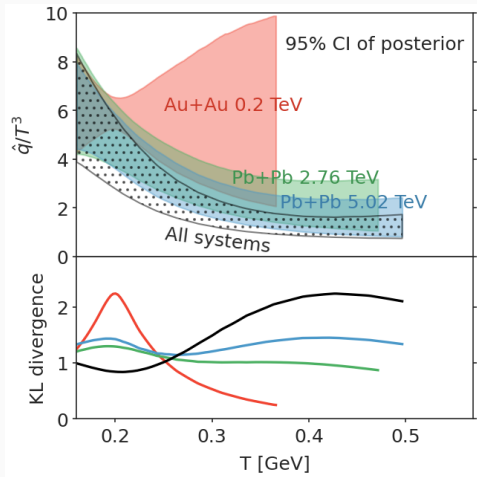
IF prior suppresses long-range correlation, what's left is the true experimental constraint!

Incremental constraining power from peripheral to central data



- Only data within $XX - YY\%$ centrality range are used for each calibration.
- Each band is drawn up to the highest temperature reached in hydro simulations.

Incremental constraining power from RHIC to LHC data.



- One can safely combine data from different centrality and \sqrt{s} .
- ★ With explicit parameterizations, one can still separate data constraints from the prior limitation by computing the K-L divergence:

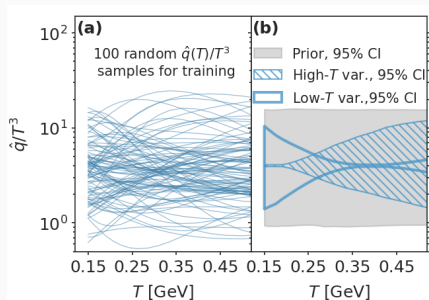
$$KL = \int dx P_1(x) \ln \frac{P_1(x)}{P_0(x)}$$

Quantify the information gain from prior (P_0) to posterior (P_1).

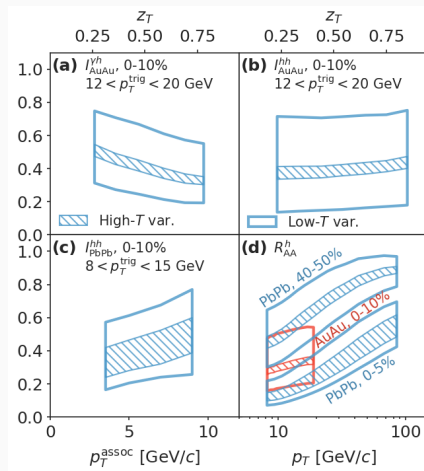
Information field approach allows an easy data sensitivity test

Generate two samples of $\hat{q}(T)$

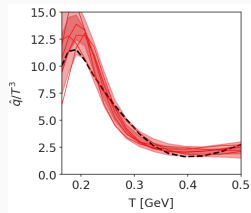
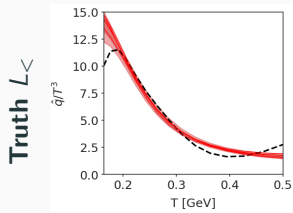
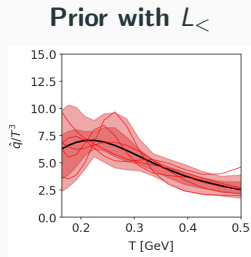
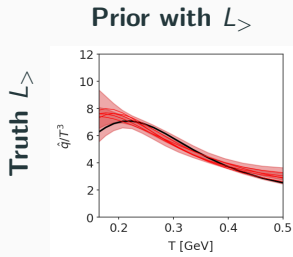
- High- T variation: tightly fix \hat{q} for $T \sim 0.165$ GeV.
- Low- T variation: tightly fix \hat{q} for $T \sim 0.375$ GeV.



Make ensemble predictions to observables to visualize which observable is more sensitive to high/low- T region.



What is the role of the correlation length L in random field prior?



Try two different $L_> = \ln 2$ & $L_< = \ln 1.3$ within a “toy model”.

- Extract truth with $L_>$ or $L_<$ using a prior random field with $L_>$ or $L_<$.
- Ideally, one should match $L = L_*$, where L_* is the (T -)resolution of the experimental data.

In PRL77(1996)4693: infer PDF $P(x)$ from N -point correlators, L_* can be obtained analytically

$$\text{Prior} \sim \exp \left[-\frac{L}{2} \int dx \phi (-i\partial_x)^2 \phi \right]$$

$$\ell_* \propto N^{1/3} \left[\frac{\int dx P^{1/2}}{\int dx (\partial_x \ln P)^2} \right]^{2/3}.$$

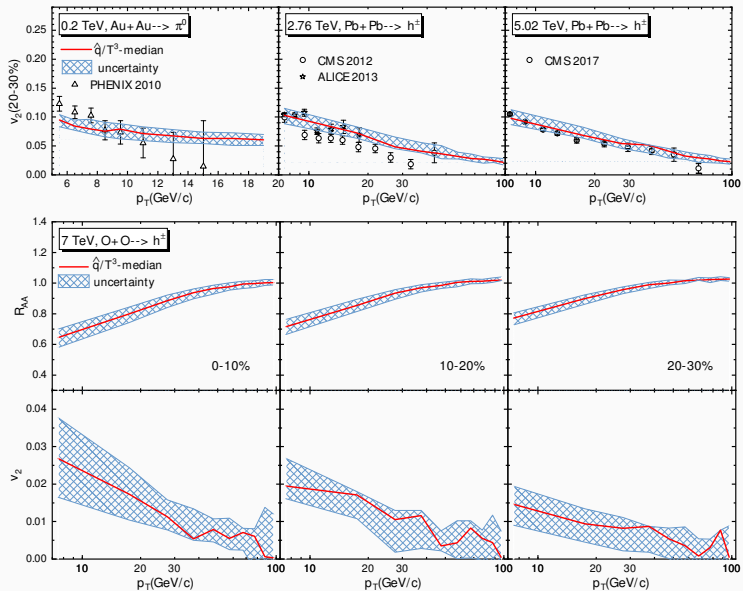
For us, we simply take the peripheral bin in Au-Au@200 GeV and guess $L_* \sim \ln \frac{T_{\max}}{T_{\min}} = \ln 2$.

Summary and Future

- The choice of prior is of fundamental importance to functional inference.
 - Propose information fields as a reasonable prior of physical functions.
 - Get rid of long-range correlations to fully expose the constraining power of experiments.
- Application to inferring the temperature dependence of \hat{q}/T^3 .
 - Demonstrate how the T -dependence of \hat{q}/T^3 are progressively determined from peripheral to central data & from RHIC to LHC data.
 - Information fields is also useful for data sensitivity analysis.
- Need improved understanding of the random-field hyper-parameters (especially L).
- Hope it is useful for broader topics in nuclear physics $\hat{q}, \eta/s, \zeta/s, f_{i/p}(x, k_T)$...

Questions?

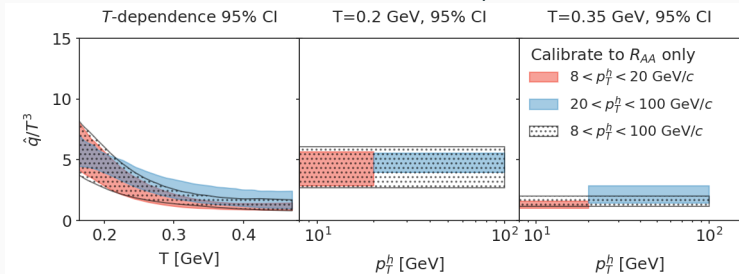
Prediction for v_2 and future O+O@LHC



Momentum dependence?

We neglected the momentum dependence of \hat{q} in this proof-of-principle study using IF method.

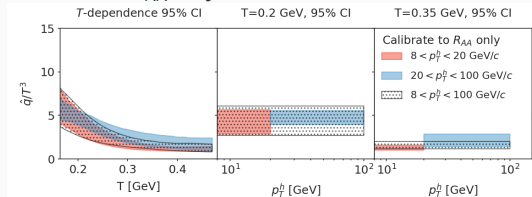
- Future: represent $\hat{q}(T, p)$ as a 2D random field.
- For now: calibrate to data with different p_T^h cuts to estimate the impact of p -dependence.



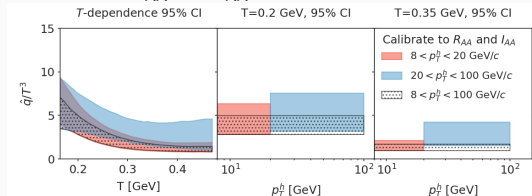
- Calibration using $8 < p_T^h < 20$ GeV and $20 < p_T^h < 100$ GeV R_{AA} data are consistent at low temperature, suggestive difference at high temperature.
- However, both are consistent with the uncertainty bands from the momentum-independent analysis.

Impact of h - h and γ - h correlations

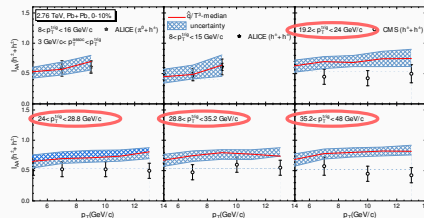
Calibrate to R_{AA} only



Calibrate to R_{AA} and I_{AA}



- No significant tightening of the uncertainty band.
- Uncertainty even gets larger when calibrate to CMS data with $p_T^{\text{trig}} > 19.2$ GeV, caused by a tension between thep. and exp.



- Also, exp uncertainty $I_{AA} > R_{AA}$.

Explore the role of hyperparameters with a toy model

Choices of $\mu(x)$ and σ are straightforward: $\mu \pm 2\sigma$ defines the high-probability prior region. The correlation length L deserves more attention. Three length scales in the problem:

- L of the prior and the training data.
- The typical variation length L^* of the truth \hat{q} .
- The temperature resolution of the data L_M . Ideally, $L, L^* \lesssim L_M$

$$\text{From } \Delta E_{\text{rad}}^{\text{approx}} \sim \frac{\alpha_s C_R}{\pi} \int d\tau \hat{q}(\tau) \tau \ln \frac{E}{m_D^2 \tau} \xrightarrow[\tau T^3 = \text{const}]{\text{Bjorken medium}} \int_{x_{\min} = \ln T_c}^{x_{\max} = \ln T_{\max}} e^{F(x) - 3x} dx, \quad F = \frac{\hat{q}(T)}{T^3}$$

We use integration limits for 40 – 50% data RHIC to estimate $L_M = \ln \frac{T_{\max}}{T_c} \approx \ln 2$.

Use the a toy model $R_{AA} = dN(p_T + \Delta E_{\text{rad}}^{\text{approx}}) / dN(p_T)$: Now we consider two scenarios

Closure tests:

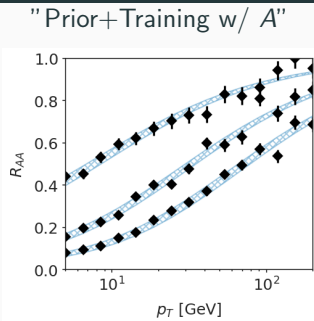
A : $L_A = \ln 2$, same as previous analysis.

B : $L_B = \ln 1.3 \approx 0.38 L_A$.

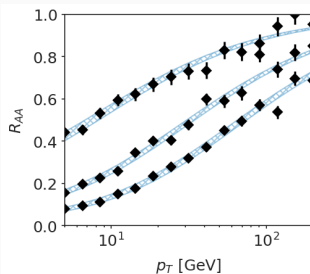
- Prior A calibrated to truth function sampled from B.
- Prior B calibrated to truth function sampled from A.

Posterior of Observables

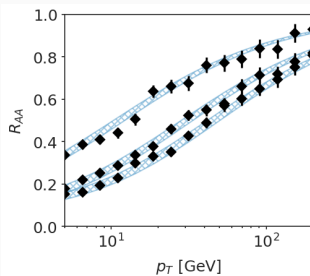
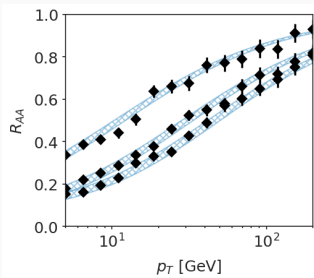
"Pseudodata from A"



"Prior+Training w/ B"



"Pseudodata from B"



Pseudodata: generated w/ the truth \hat{q} sampled from either prior A or B

- Three "centralities",
 $T_{\max} = 0.3, 0.4, 0.5$ GeV.
- 5% level uncorrelated uncertainty.

Both choice of correlation length can describe the pseudodata from either prior A or B .

Calculations in $p+p$ collisions

Renormalization scales $\mu \propto p_T$ at tuned at RHIC and LHC energies to describe the single-hadron, $h-h$ and $\gamma-h$ spectra in $p+p$.

