Bayesian constraints on the initial stage using *Trajectum*

Govert Nijs

February 6, 2023

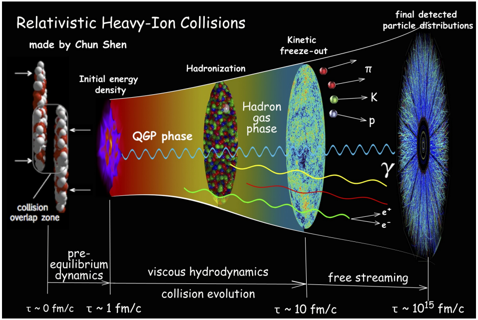
Based on:

- GN, van der Schee, 2206.13522
- GN, van der Schee, 2302.xxxxx
- Giuliano Giacalone, GN, van der Schee, 230x.xxxxx

Massachusetts Institute of Technology

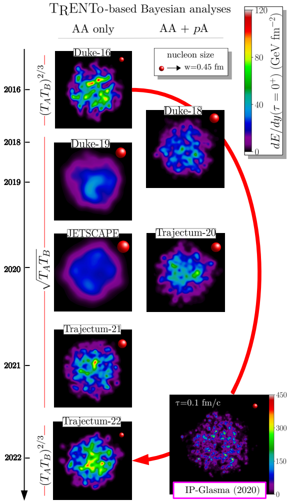
The status of the field

- The general picture of the stages of a heavy ion collision is known.
- We now want to understand each part in more detail.
- Making progress from first principles is challenging.
- We use a data-driven approach.
  - Fit a model with many parameters to a wide range of experimental data.
  - The parameters themselves are in many cases interesting quantities.
- Requirements:
  - Fast simulation code.
  - Bayesian methods.



[Sorensen, Shen, 2014]

Our evolving understanding of the initial state



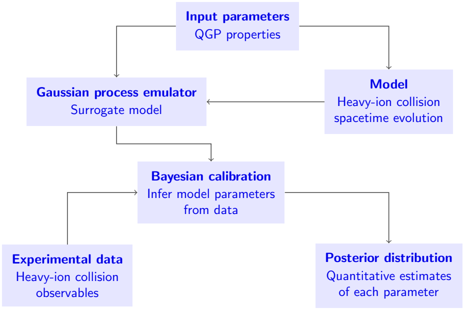
- Initial conditions took a circle journey since 2016:
  - Small nucleon size at first, then larger, now small again.
  - Energy deposition went from $T^{00} \propto (\mathcal{T}_A\mathcal{T}_B)^{2/3}$ to $T^{00} \propto \sqrt{\mathcal{T}_A\mathcal{T}_B}$, and now back to $T^{00} \propto (\mathcal{T}_A\mathcal{T}_B)^{2/3}$.
  - Pre-hydrodynamic stage increased in complexity from no dynamics, to free streaming, and now to a parameterized interpolation between weak and strong coupling.
- Progress was enabled by Bayesian analysis.
- We focus on the latest of these analyses: *Trajectum*-22.

[Giacalone, 2208.06839]

Massachusetts Institute of Technology

Bayesian analysis

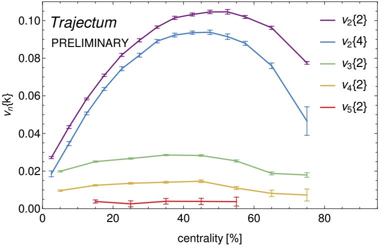
- We want to fit 23 parameters to 653 data points.
- Two problems:
  - Even the fastest models are too slow.
  - The parameter space is large.
- The first problem is solved by replacing the model with an emulator trained on model simulations.
- The second problem is solved by using Markov Chain Monte Carlo (MCMC), which samples the posterior using importance sampling.



[Bernhard, 1804.06469]

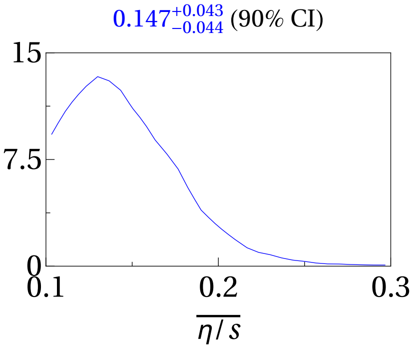
Trajectum

- New heavy ion code developed in Utrecht/MIT/CERN.
- Contains initial stage, hydrodynamics and freeze-out, as well as an analysis suite.
- Easy to use, example parameter files distributed alongside the source code.
- Fast, fully parallelized.
    - Figure (20k oversampled PbPb events at 2.76 TeV) computes on a laptop in 21h.
    - Bayesian analysis requires $\mathcal{O}(1000)$ similar calculations to this one.
- Publicly available at sites.google.com/view/govertnijs/trajectum/.

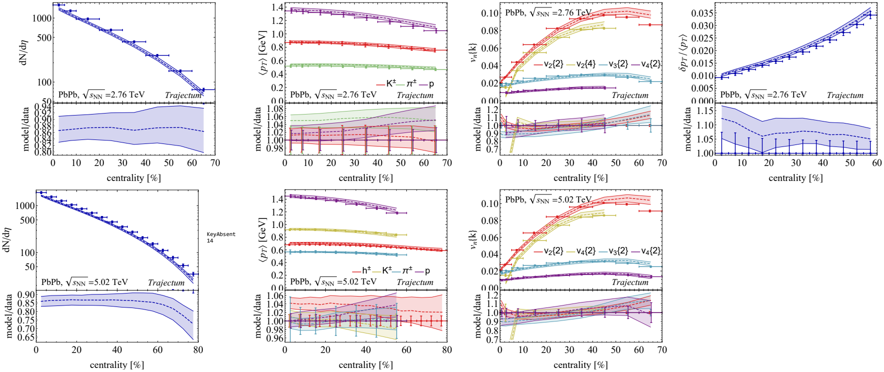


Posterior distribution of parameters

- MCMC yields large amount of parameter sets distributed according to the posterior.
- 1D projections show the preferred values for a single parameter.
- Computing observables with randomly drawn parameters from the posterior propagates the parameter uncertainties, including correlations.



A selection of data used



Massachusetts Institute of Technology

TRENTo initial conditions

- Nucleons A and B become *wounded* with probability

$$P_{\text{wounded}} = 1 - \exp\left(-\sigma_{gg} \int d\mathbf{x}\, \rho_A(\mathbf{x}) \rho_B(\mathbf{x})\right), \quad \rho_A \propto \exp\left(\frac{-|\mathbf{x} - \mathbf{x}_A|^2}{2w^2}\right).$$

- Each wounded nucleon desposits energy into its nucleus's *thickness function* $\mathcal{T}_{A/B}$:

$$\mathcal{T}_{A/B} = \sum_{i \in \text{wounded A/B}} \gamma \exp(-|\mathbf{x} - \mathbf{x}_i|^2 / 2w^2),$$

with γ drawn from a gamma distribution with mean 1 and standard deviation σ_{fluct}.

- Actual formulas slightly modified because each nucleon has n_c constituents.

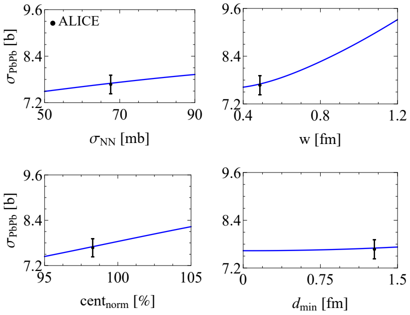
[Moreland, Bernhard, Bass, 1412.4708, 1808.02106]

Massachusetts Institute of Technology

The cross-section σ_{AA} for different nucleon widths

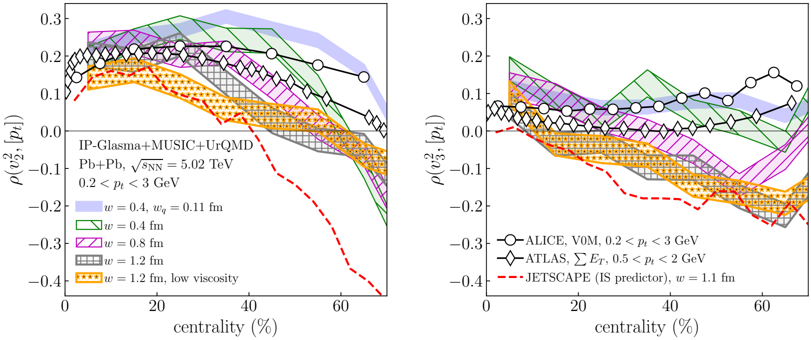
- The cross-section depends strongly on the nucleon width w and the centrality normalization $\text{cent}_{\text{norm}}$.
- ALICE finds: $7.67 \pm 0.24\,\text{b}$.
- Cross-section measurement seems to require smaller w than earlier analyses.
- Basic observable: models should get this right.

[ALICE, 2204.10148; ALICE-PUBLIC-2022-004]



Massachusetts Institute of Technology

$\rho(v_2^2, \langle p_T \rangle)$ for different nucleon widths

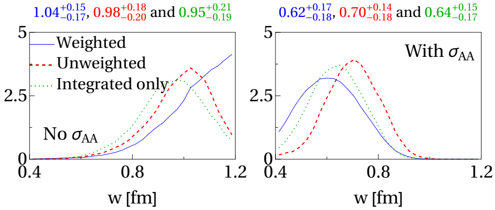


- The correlation between v_2^2 and $\langle p_T \rangle$ is sensitive to the nucleon width w.
- Smaller w is preferred.
- This is a statistically challenging observable.

[Giacalone, Schenke, Shen, 2111.02908]

Including the pPb and PbPb cross sections in the analysis

- Including the pPb and PbPb cross sections in the fit lowers w from 1 fm to 0.6 fm.
- Smaller width is now compatible with our knowledge of the proton.
- Result is robust under various fitting scenarios.



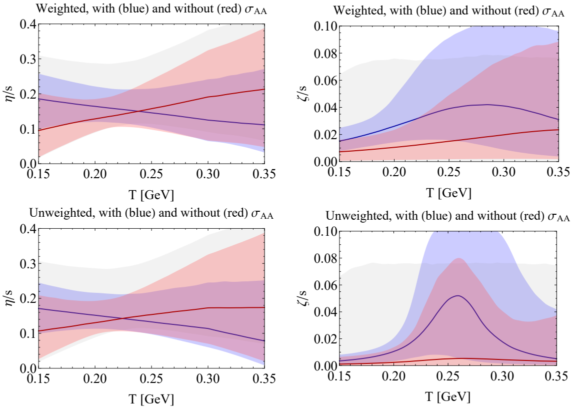
	σ_{PbPb}[b]	$\sigma_{p\text{Pb}}$[b]
with σ_{AA}	8.02 ± 0.19	2.20 ± 0.06
without σ_{AA}	8.95 ± 0.36	2.48 ± 0.10
ALICE/CMS	7.67 ± 0.24	2.06 ± 0.08

[ALICE, 2204.10148; ALICE-PUBLIC-2022-004; CMS, 1509.03893]

Implication for viscosities

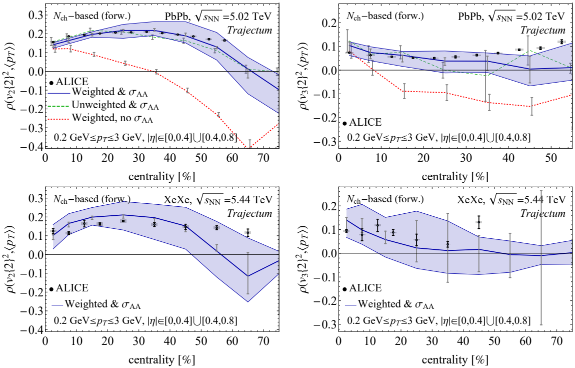
- Smaller nucleons imply larger radial flow.
- Specific bulk viscosity ζ/s increases to compensate.
- Including σ_{AA} reverses the preferred slope of specific shear viscosity η/s.
- Similar findings in IP-Glasma based Bayesian analysis presented at Quark Matter.

[Heffernan, Jeon, Gale, Paquet, to appear]



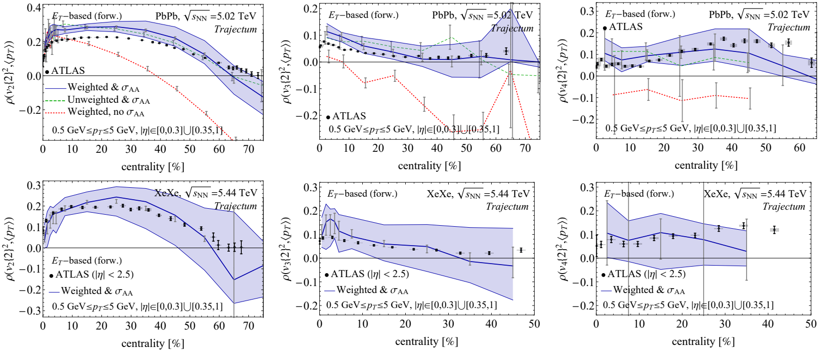
Implication for $\rho(v_2^2, \langle p_T \rangle)$ (ALICE)

- We can use the full posterior to propagate uncertainties from parameters to observables.
- Much improved agreement with ALICE for $\rho(v_2^2, \langle p_T \rangle)$.



Massachusetts Institute of Technology

Implication for $\rho(v_2^2, \langle p_T \rangle)$ (ATLAS)



- Still some tension with ATLAS:
  - Kinematic cuts are different, probably needs 3+1D simulations to resolve.
  - Important to match the precise experimental procedure.

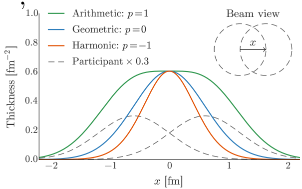
Massachusetts Institute of Technology

The TRENTo phenomenological ansatz

- The standard TRENTo formula combines thickness functions of the two nuclei $\mathcal{T}_A$ and $\mathcal{T}_B$ into a *reduced thickness* $\mathcal{T}$, interpreted as an energy density:

$$\mathcal{T} \propto \left(\frac{\mathcal{T}_A^p + \mathcal{T}_B^p}{2} \right)^{1/p},$$

  with p a parameter.



[Moreland, Bernhard, Bass, 1412.4708]

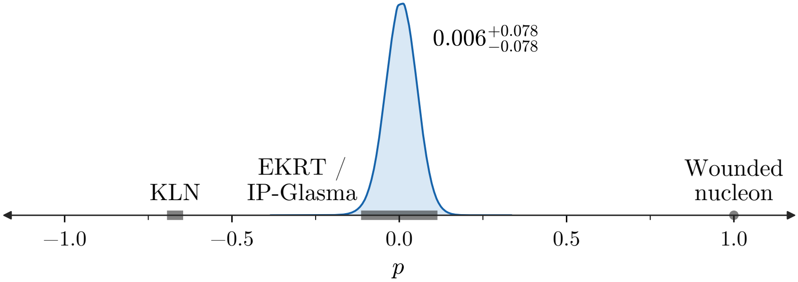
- Some useful limits:

p	-1	0	1
$\mathcal{T}$	$\frac{2}{\frac{1}{\mathcal{T}_A}+\frac{1}{\mathcal{T}_B}}$	$\sqrt{\mathcal{T}_A\mathcal{T}_B}$	$\frac{\mathcal{T}_A+\mathcal{T}_B}{2}$

- Binary scaling $\mathcal{T} = \mathcal{T}_A\mathcal{T}_B$ is not available.

Massachusetts Institute of Technology

The power of Bayesian analysis



- Can test theories for the initial state with TRENTo, in this case by comparing their scaling behavior.
- General workflow for testing theories/questions:
  - Introduce parameter(s) which parameterize the question.
  - Confront the generalized model with data using Bayesian analysis.
  - Read off the posterior distribution for the parameter(s).

[Bernhard, 1804.06469]

Massachusetts Institute of Technology

The q parameter

- We make the following modification to the TRENTo formula:

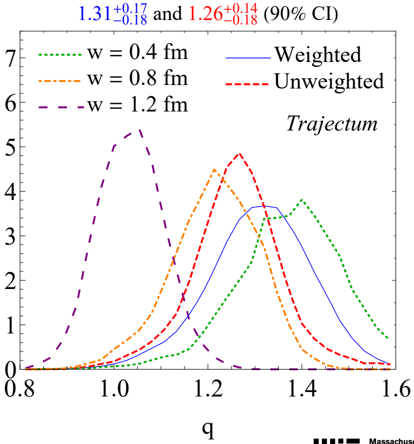
$$\mathcal{T} \propto \left(\frac{\mathcal{T}_A^p + \mathcal{T}_B^p}{2} \right)^{q/p},$$

introducing the parameter q.

- We now include *binary scaling* as a limit when $p = 0$, $q = 2$.
- Assuming approximate conformality of the equation of state, we can also interpret the right hand side as an *entropy density* by setting $q = 4/3$.

Posterior distribution for q

- Binary scaling ($q = 2$) is strongly disfavored.
- Fixing the nucleon width w at different values has a large effect on the fitted value for q.
- Fixing $w = 0.4\,\mathrm{fm}$ favors $q \approx 4/3$.
- Weighted distribution is close to $w = 0.4\,\mathrm{fm}$ distribution.



Comparing to IP-Glasma

- IP-Glasma scales as follows:

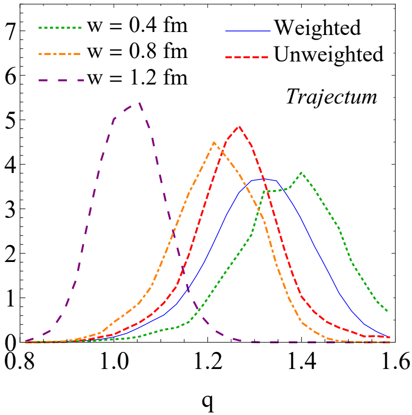
$$\mathcal{T} \propto \frac{\mathcal{T}_A \mathcal{T}_B (2\mathcal{T}_A^2 + 7\mathcal{T}_A\mathcal{T}_B + 2\mathcal{T}_B^2)}{(\mathcal{T}_A + \mathcal{T}_B)^{5/2}}.$$

- If $\mathcal{T}_A \approx \mathcal{T}_B$, this reduces to

$$\mathcal{T} \propto (\mathcal{T}_A\mathcal{T}_B)^{3/4}.$$

- This corresponds to $q = 1.5$.
- IP-Glasma is compatible with our posterior.

$1.31^{+0.17}_{-0.18}$ and $1.26^{+0.14}_{-0.18}$ (90% CI)



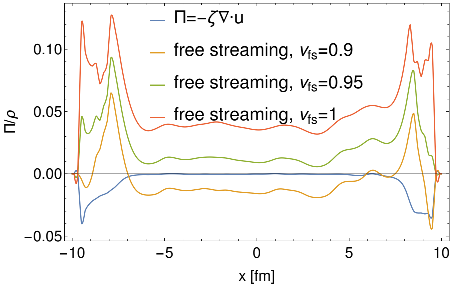
[Borghini et al., 2209.01176]

Strongly coupled pre-hydrodynamic stage: early effort

- In AdS/CFT simulations of the initial stage, the shear stress and bulk pressure quickly relax to their 'hydro' values:

$$\pi^{\mu\nu} = 2\eta\sigma^{\mu\nu}, \quad \Pi = -\zeta\nabla\cdot u.$$

- In free streaming however, the initialization of $\pi^{\mu\nu}$ and Π is qualitatively different.
- Use free streaming velocity as a proxy for this difference.



[van der Schee, Romatschke, Pratt, 1307.2539; GN, van der Schee, Gürsoy, Snellings, 2010.15134]

Massachusetts Institute of Technology

Free streaming pre-hydrodynamic stage

- TRENTo creates matter at proper time $\tau = 0^+$.
- Propagate the matter using free streaming:

$$T^{\mu\nu}(x, y, \tau_{\text{hyd}}) = \frac{1}{\tau_{\text{hyd}}} \int d\phi \, \hat{p}^\mu \hat{p}^\nu \mathcal{T}(x - \tau_{\text{hyd}} \cos\phi, y - \tau_{\text{hyd}} \sin\phi),$$

with

$$\hat{p}^\mu = \begin{pmatrix} 1 & \cos\phi & \sin\phi \end{pmatrix},$$

giving us the stress tensor $T^{\mu\nu}$ at proper time $\tau = \tau_{\text{hyd}}$.

- Here τ_{hyd} is the time at which hydrodynamics is started.
- The factor $1/\tau_{\text{hyd}}$ is due to longitudinal expansion.

The r_{hyd} parameter

- We compute the hydrodynamic values for $\pi^{\mu\nu}$ and Π explicitly from the velocity u^μ:

$$\pi^{\mu\nu}_{\text{hyd}} = 2\eta\sigma^{\mu\nu}, \qquad \Pi_{\text{hyd}} = -\zeta\nabla\cdot u.$$

- We then mix the hydrodynamic values with the free streaming values and initialize hydrodynamics with

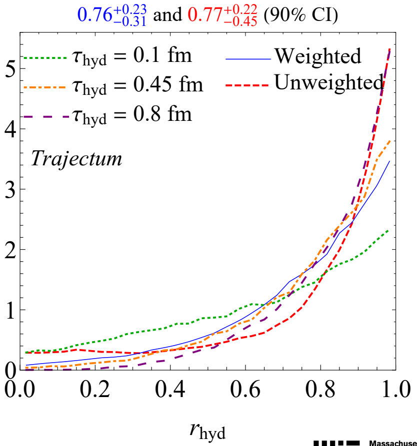
$$\pi^{\mu\nu} = r_{\text{hyd}}\pi^{\mu\nu}_{\text{hyd}} + (1 - r_{\text{hyd}})\pi^{\mu\nu}_{\text{fs}},$$
$$\Pi = r_{\text{hyd}}\Pi_{\text{hyd}} + (1 - r_{\text{hyd}})\Pi_{\text{fs}},$$

with $r_{\text{hyd}} \in [0, 1]$ interpolating between the two scenarios.

Massachusetts Institute of Technology

Posterior distribution for r_{hydro}

- $r_{\text{hyd}} = 1$ is strongly favored over $r_{\text{hyd}} = 0$, implying a preference for strongly coupled pre-hydrodynamic stage.
- Preference also becomes stronger for larger hydro initialization time τ_{hyd}.
- One can see this as model averaging, albeit cheaper since we can interpolate between models with a continuous parameter.



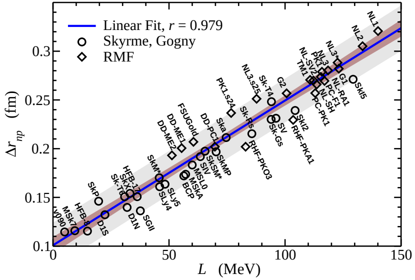
Massachusetts Institute of Technology

Neutron skin

- In a nucleus, neutrons sit further from the center than protons.
- This is quantified with the neutron skin thickness

$$\Delta r_{np} = \langle r^2 \rangle_n^{1/2} - \langle r^2 \rangle_p^{1/2}.$$

- Neutron skin can be used to obtain the symmetry energy L, which is relevant for neutron stars:



$$\Delta r_{np}\ [\text{fm}] = 0.101 + 0.00147 L\ [\text{MeV}].$$

[Viñas, Centelles, Roca-Maza, Warda, 1308.1008]

Massachusetts Institute of Technology

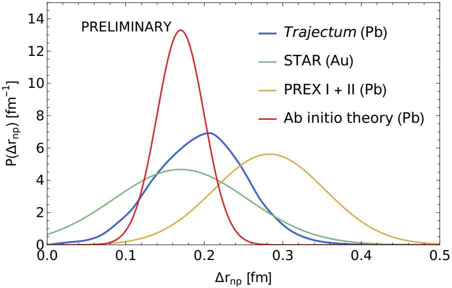
Bayesian analysis result using LHC data

- We draw nucleons in a nucleus from a Woods-Saxon distribution

$$\rho_{\mathrm{WS}}(r) \propto \frac{1}{1+\exp\left(\frac{r-R}{\sigma}\right)},$$

  where R and σ are parameters, which we take to be different for protons and neutrons.
- Proton distribution is well-constrained by electron scattering experiments, so we vary the neutron σ.

- Resulting posterior for Δr_{np} is compatible with STAR, PREX and ab initio nuclear theory.



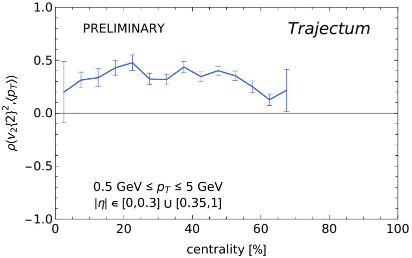
[STAR, 2204.01625; PREX, 2102.10767; Hu et al., *Nat. Phys.* **18**, 1196–1200 (2022)]

Conclusions

- Describing the experimental cross-section requires smaller nucleon width.
- Binary scaling in TRENTo is strongly disfavored.
- Reduced thickness $\mathcal{T}$ should be interpreted as an entropy density.
- Scaling behavior of TRENTo is compatible with IP-Glasma.
- Our fit favors a strongly coupled pre-hydrodynamic stage.
- The neutron skin thickness can be extracted using heavy ion data.

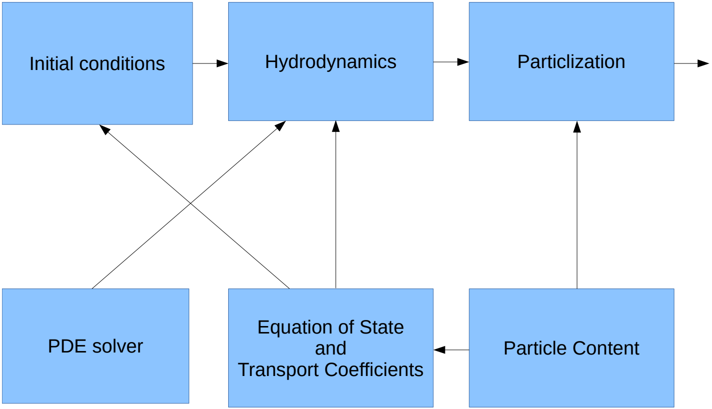
Outlook

- Much improved statistics: can now fit to $\rho(v_2\{2\}^2, \langle p_T \rangle)$ directly.
- Bayesian analysis with 3+1D simulations.
- Nuclear structure with ^{16}O and ^{20}Ne.
- Interpolating between TRENTo scaling and using the IP-Glasma scaling directly.



Massachusetts Institute of Technology

Components of *Trajectum*



Hydrodynamics

- Define ($g^{\mu\nu} = \mathrm{diag}(1, -1, -1, -1)$):

$$\Delta^{\mu\nu} = g^{\mu\nu} - u^\mu u^\nu, \quad \nabla^\mu = \Delta^{\mu\nu}\partial_\nu, \quad D = u^\mu \nabla_\mu, \quad \sigma^{\mu\nu} = \nabla^{\langle\mu} u^{\nu\rangle},$$

  with $\langle\rangle$ symmetrizing and removing the trace.
- We solve viscous hydrodynamics without currents, i.e.

$$\partial_\mu T^{\mu\nu} = 0, \quad T^{\mu\nu} = e u^\mu u^\nu - (P + \Pi)\Delta^{\mu\nu} + \pi^{\mu\nu},$$

- $\pi^{\mu\nu}$ and Π follow the 14-moment approximation:

$$\begin{aligned} -\tau_\pi \Delta^\mu_\alpha \Delta^\nu_\beta D\pi^{\alpha\beta} &= \pi^{\mu\nu} - 2\eta\sigma^{\mu\nu} + \delta_{\pi\pi}\pi^{\mu\nu}\nabla\cdot u \\ &\quad - \phi_7 \pi^{\langle\mu}_\alpha \pi^{\nu\rangle\alpha} + \tau_{\pi\pi}\pi^{\langle\mu}_\alpha \sigma^{\nu\rangle\alpha} - \lambda_{\pi\Pi}\Pi\sigma^{\mu\nu}, \\ -\tau_\Pi D\Pi &= \Pi + \zeta\nabla\cdot u + \delta_{\Pi\Pi}\nabla\cdot u\Pi - \lambda_{\Pi\pi}\pi^{\mu\nu}\sigma_{\mu\nu}. \end{aligned}$$

Massachusetts Institute of Technology

Particlization

- At the freeze-out temperature T_{sw}, we turn the fluid back into particles.
- Particles are sampled thermally, and boosted with the fluid velocity u^{μ}.
- We use the PTB prescription to match $\pi^{\mu\nu}$ and Π across the transition, so that $T^{\mu\nu}$ is smooth.
- After particlization, we use SMASH as a hadronic afterburner.

[Pratt, Torrieri, 1003.0413; Bernhard, 1804.06469]

Need to go to second order

- First order scheme fails: combination of large T_RENTo norm N, small hydro initialization time τ_{hyd} and large specific shear viscosity η/s causes extreme particle yields, ruining the emulator.
- Need to go to second order, which penalizes large initial values for $\pi^{\mu\nu}$ and Π.
- Use full 14-moment approximation:

$$-\tau_\pi \Delta^\mu_\alpha \Delta^\nu_\beta D\pi^{\alpha\beta} = \pi^{\mu\nu} - 2\eta\sigma^{\mu\nu} + \delta_{\pi\pi}\pi^{\mu\nu}\nabla\cdot u - \phi_7 \pi^{\langle\mu}_\alpha \pi^{\nu\rangle\alpha} + \tau_{\pi\pi}\pi^{\langle\mu}_\alpha \sigma^{\nu\rangle\alpha} - \lambda_{\pi\Pi}\Pi\sigma^{\mu\nu},$$

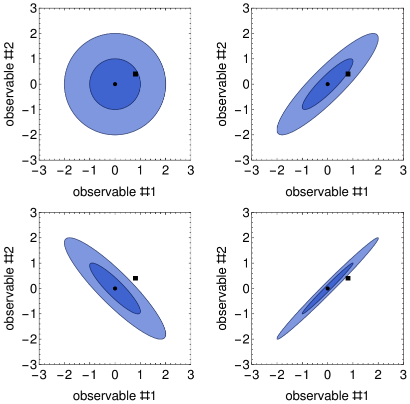
$$-\tau_\Pi D\Pi = \Pi + \zeta\nabla\cdot u + \delta_{\Pi\Pi}\nabla\cdot u\Pi - \lambda_{\Pi\pi}\pi^{\mu\nu}\sigma_{\mu\nu},$$

where we set the left hand side to zero, and solve for $\pi^{\mu\nu}$ and Π.

Massachusetts Institute of Technology

Why weights?

- Higher p_T, higher centralities are harder to model theoretically.
- Experimental correlation matrix is not available.
  - Figure shows 1σ and 2σ regions for $\rho \in \{0, 0.9, -0.9, 0.99\}$, with standard deviations the same.
  - Same difference between theory and experiment can be within 1σ or outside of 2σ depending on ρ.
  - Correlated observable classes can be over/underimportant for the Bayesian analysis.



Definition of weights

- In the bayesian analysis, the probability of the data given the parameter point x is given by:

$$P(D|x) = \frac{1}{\sqrt{(2\pi)^m \det \Sigma}} \exp\left(-\frac{1}{2}(y - y_{\text{exp}})^T \Sigma^{-1} (y - y_{\text{exp}})\right),$$

  with y the vector of observables computed from x, y_{exp} the vector of the corresponding experimental data, and Σ the combined theory/experiment covariance matrix.

- We define weights by replacing

$$P(D|x) = \frac{1}{\sqrt{(2\pi)^m \det \Sigma}} \exp\left(-\frac{1}{2}(y - y_{\text{exp}})^T \omega \Sigma^{-1} \omega (y - y_{\text{exp}})\right),$$

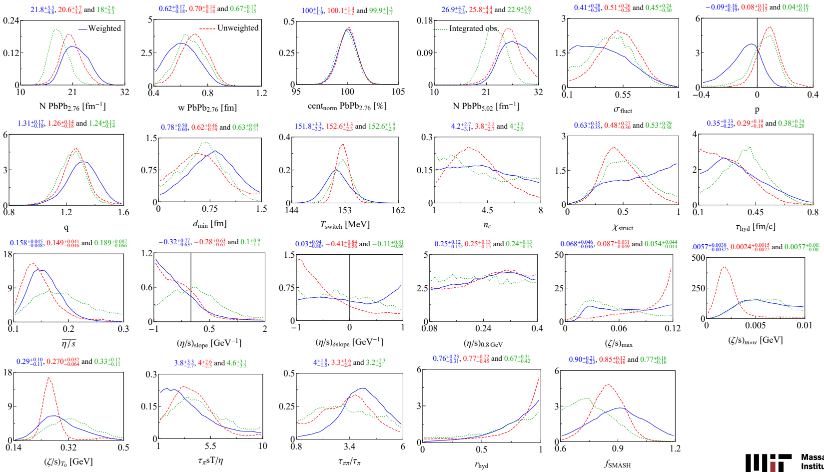
  where ω is the diagonal matrix containing the weight for each observable.

Choice of weights

- We choose for weights ω:
  - 1/2 for every particle identified observable.
  - 1/2 for p_T-differential observables, and an additional $(2.5 - p_T[\text{GeV}])/1.5$ if $p_T > 1\,\text{GeV}$.
  - $(100 - c[\%])/50$ if the centrality class c is beyond 50%.
- Weighting only worsens the average discrepancy slightly.
- Distribution of discrepancies makes more sense.

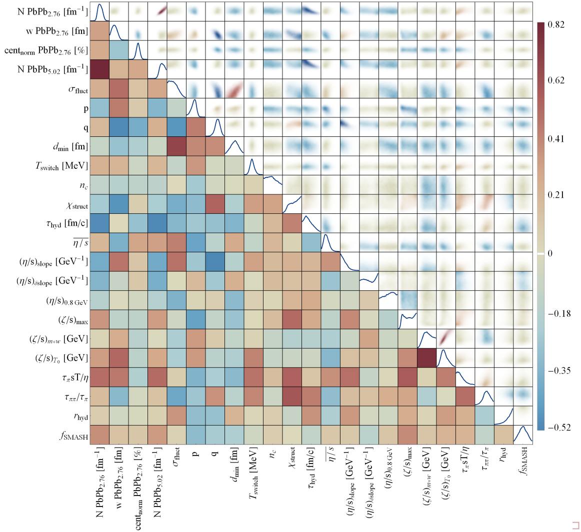
	$\langle (y_{\text{theory}} - y_{\text{experiment}})/\sigma \rangle$				$\bar{\omega}$
	σ_{AA} & ω	ω	σ_{AA}	neither	
$dN_{\text{ch}}/d\eta$	0.55	0.60	1.23	1.22	1.00
$dN_{\pi^\pm,k^\pm,p^\pm}/dy$	0.76	0.70	0.60	0.57	0.48
$dE_T/d\eta$	1.59	1.51	0.82	0.77	0.48
$\langle p_T \rangle_{\text{ch},\pi^\pm,K^\pm,p^\pm}$	0.66	0.60	0.88	0.72	0.46
$\delta p_T/\langle p_T \rangle$	0.56	0.62	0.51	0.58	0.49
$v_n\{k\}$	0.58	0.51	0.54	0.49	1.00
$d^2N_{\pi^\pm}/dy\,dp_T$	1.19	1.07	0.86	0.92	0.20
$d^2N_{K^\pm}/dy\,dp_T$	1.41	1.27	0.79	0.73	0.20
$d^2N_{p^\pm}/dy\,dp_T$	1.35	1.21	0.73	0.67	0.25
$v_2^{\pi^\pm}(p_T)$	0.81	0.74	0.46	0.44	0.19
$v_2^{K^\pm}(p_T)$	0.92	0.89	0.55	0.55	0.19
$v_2^{p^\pm}(p_T)$	0.49	0.47	0.34	0.35	0.25
$v_3^{\pi^\pm}(p_T)$	0.65	0.57	0.69	0.57	0.24
average	0.89	0.83	0.69	0.66	
σ_{AA}	1.13	3.80	1.53	3.40	1.00

How much do weights change the posteriors?

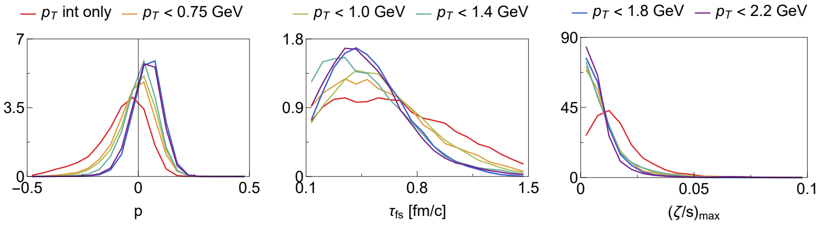


Massachusetts Institute of Technology

Correlations between parameters



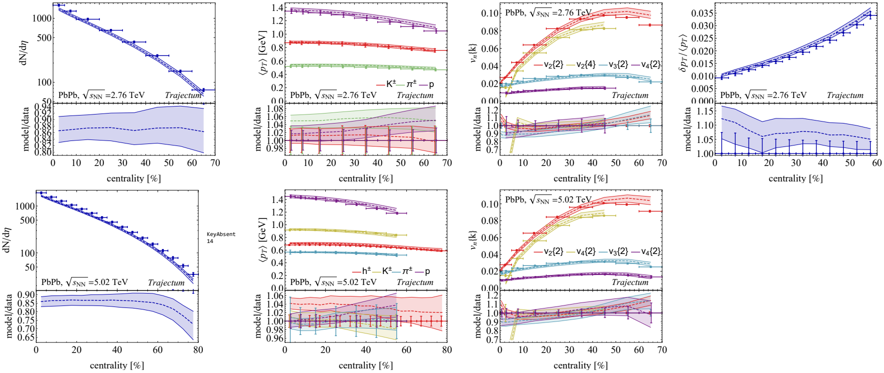
Bulk viscosity over entropy density ζ/s



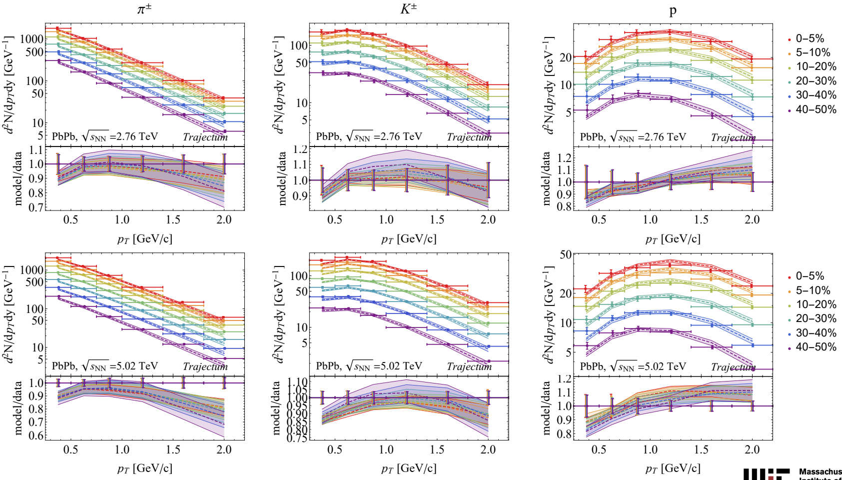
- Lower bulk viscosity than other groups.
- This is mostly due to the inclusion of p_T-differential observables.
- It is important to fit to as wide a range of data as possible (within reason).
- We varied the highest p_T bin included to check that our result was robust.

[GN, van der Schee, Gürsoy, Snellings, 2010.15130]

Data used in our most recent fit: integrated observables



Data used in our most recent fit: spectra



Massachusetts Institute of Technology

Data used in our most recent fit: p_T-differential v_2