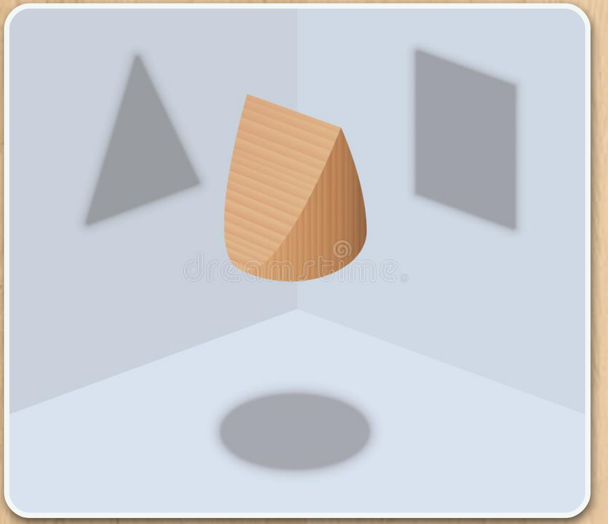
Bayesian constraints on initial condition in HI Collisions

Yi Chen (MIT)

INT Workshop 23-1a, Feb 7 2023



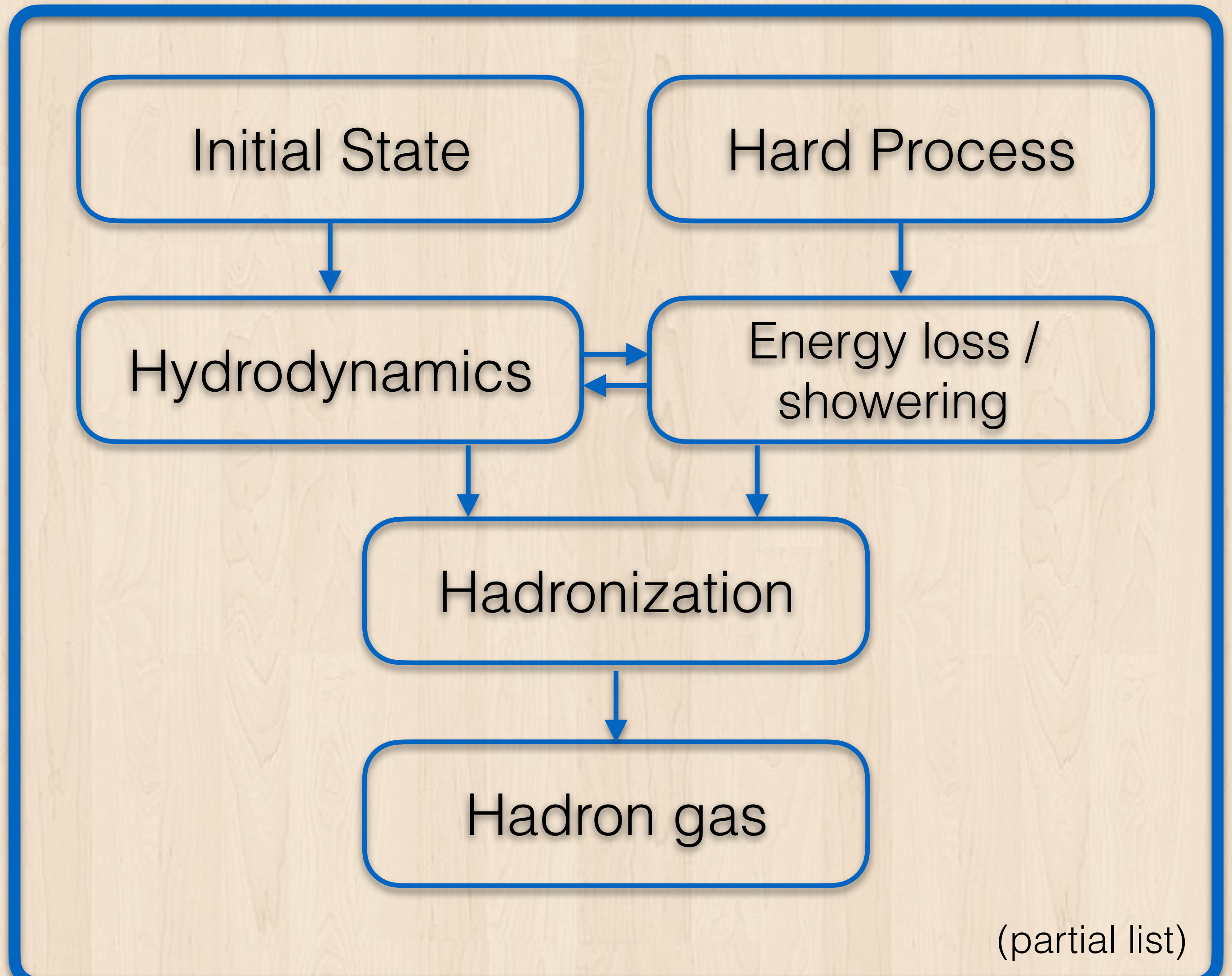
The MITHIG's work is supported by US DOE-NP

Prologue: Bayesian Analysis

Also see talk by Christian Drischler and Xilin Zhang on Monday & J-F week 1

The problem of Heavy Ions

MODEL



Problem of code-base alleviated by e.g. JETSCAPE framework

Large amount of moving pieces → parameters

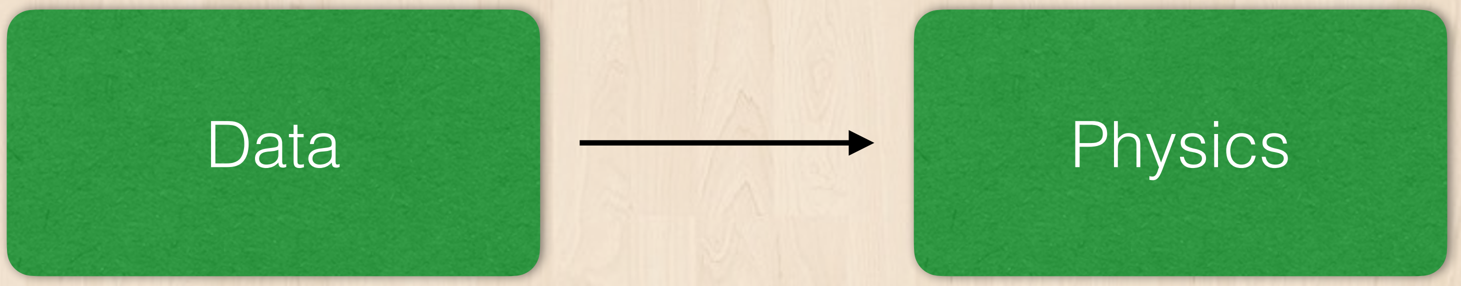
Often different codes

Computing intensive

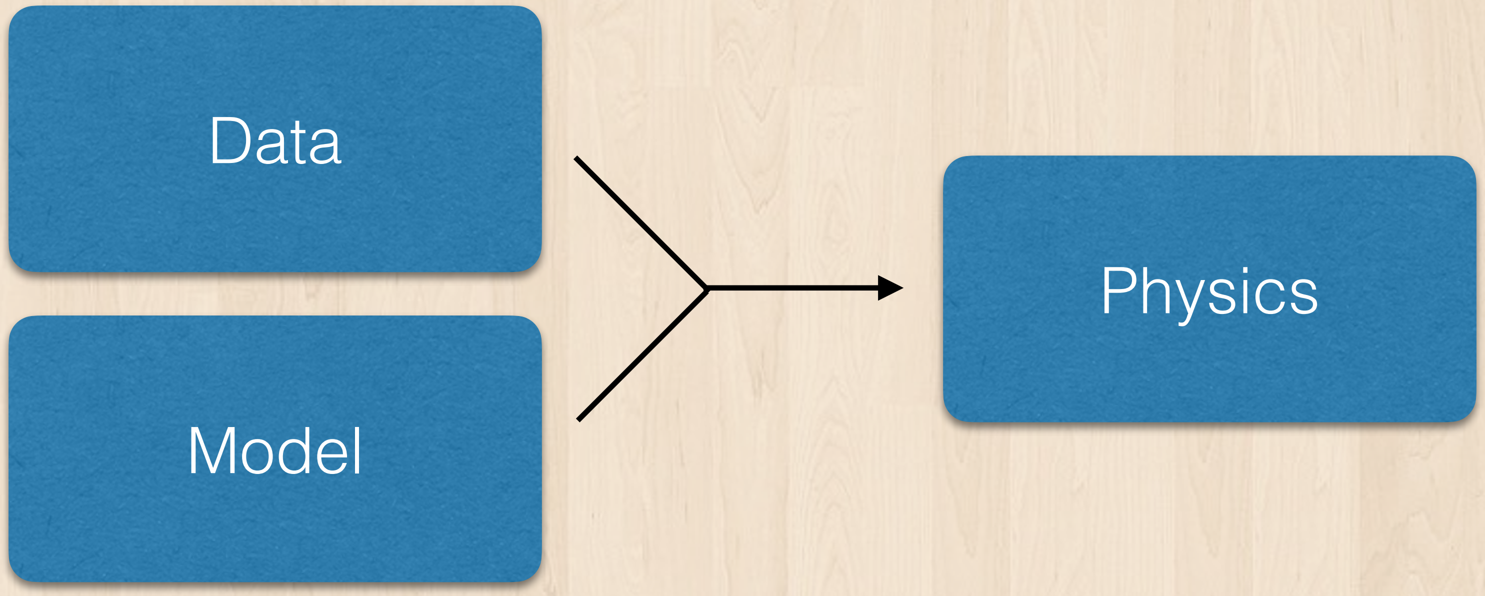
Quickly becomes non-practical

Rigorous model-data comparison

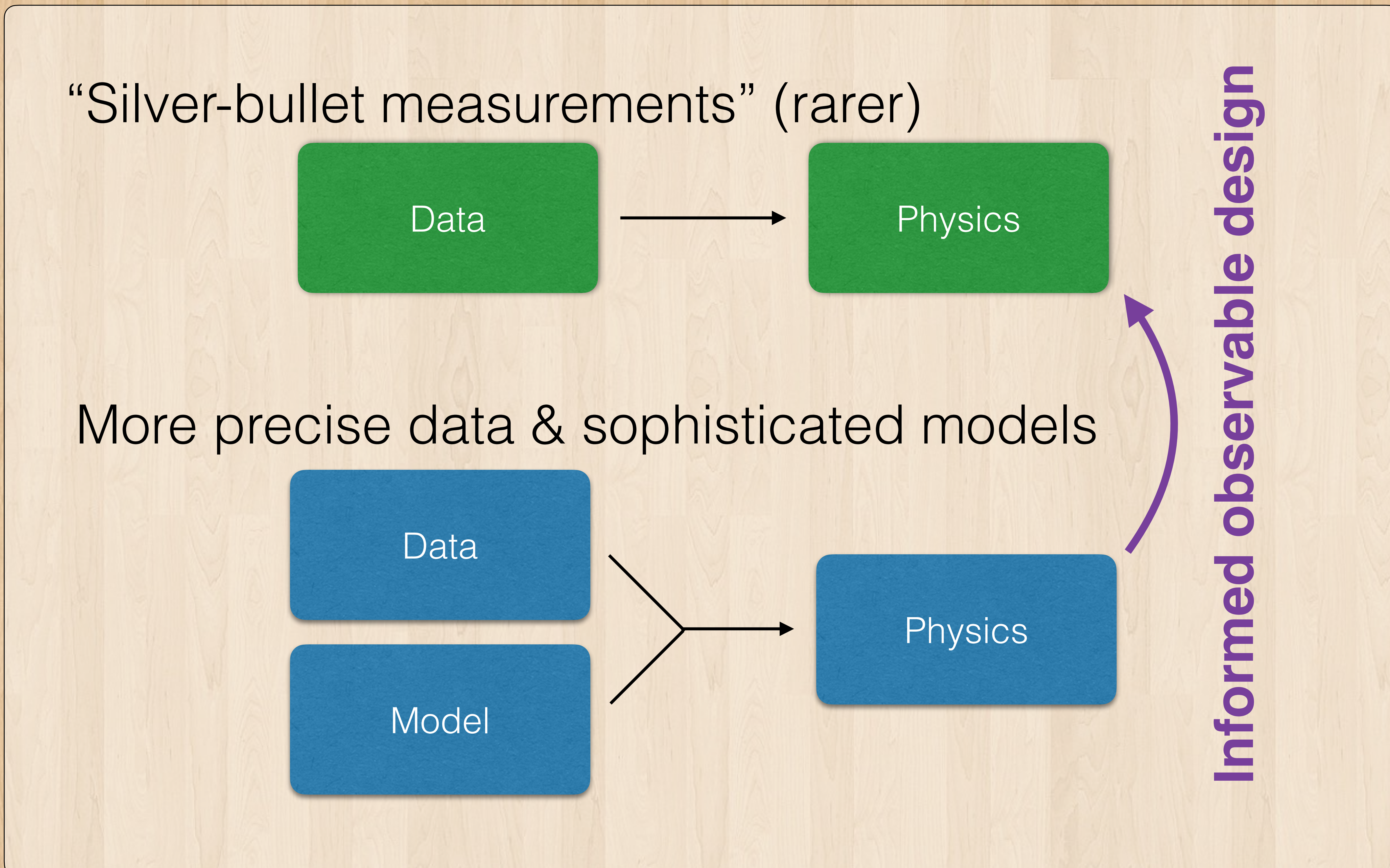
“Silver-bullet measurements” (rarer)



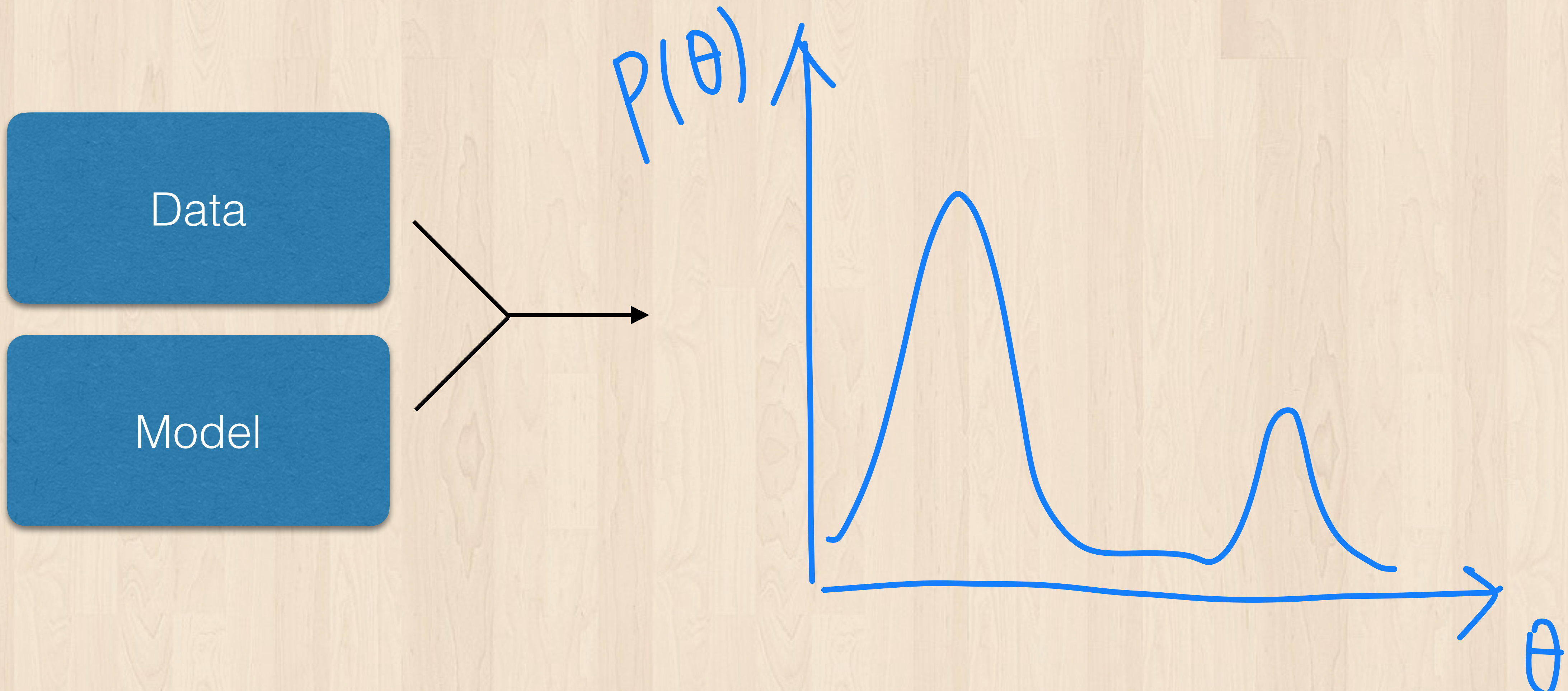
More precise data & sophisticated models



Rigorous model-data comparison



The basic idea



$P(\theta)$ = how "likely" θ is true

Encodes information from data and we can learn about the parameters

The Bayes' formalism

Bayesian likelihood

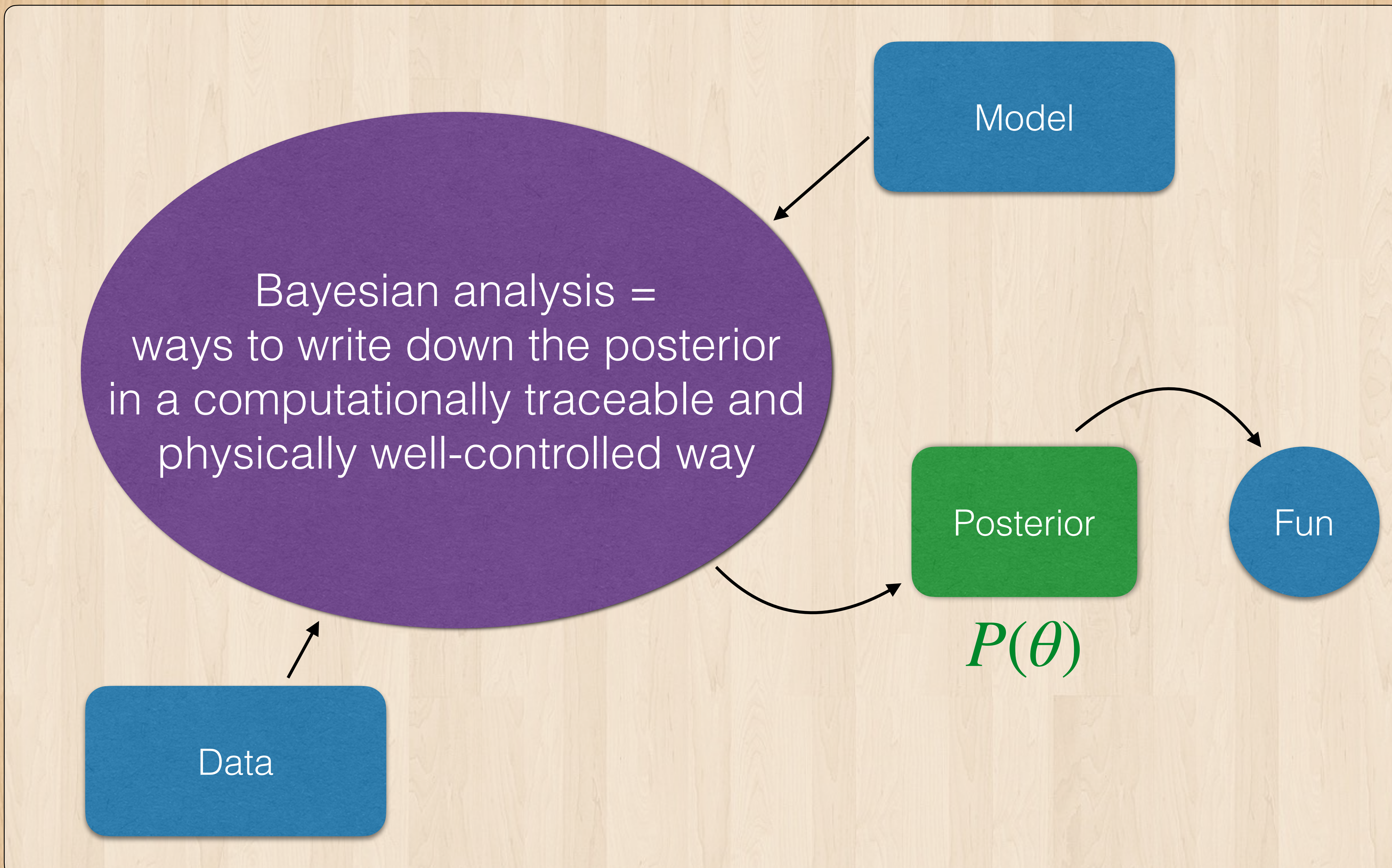
Prior knowledge

$$P(\vec{\theta}\,|\,\text{data}) = \frac{P(\text{data}\,|\,\vec{\theta})P(\vec{\theta})}{P(\text{data})}$$

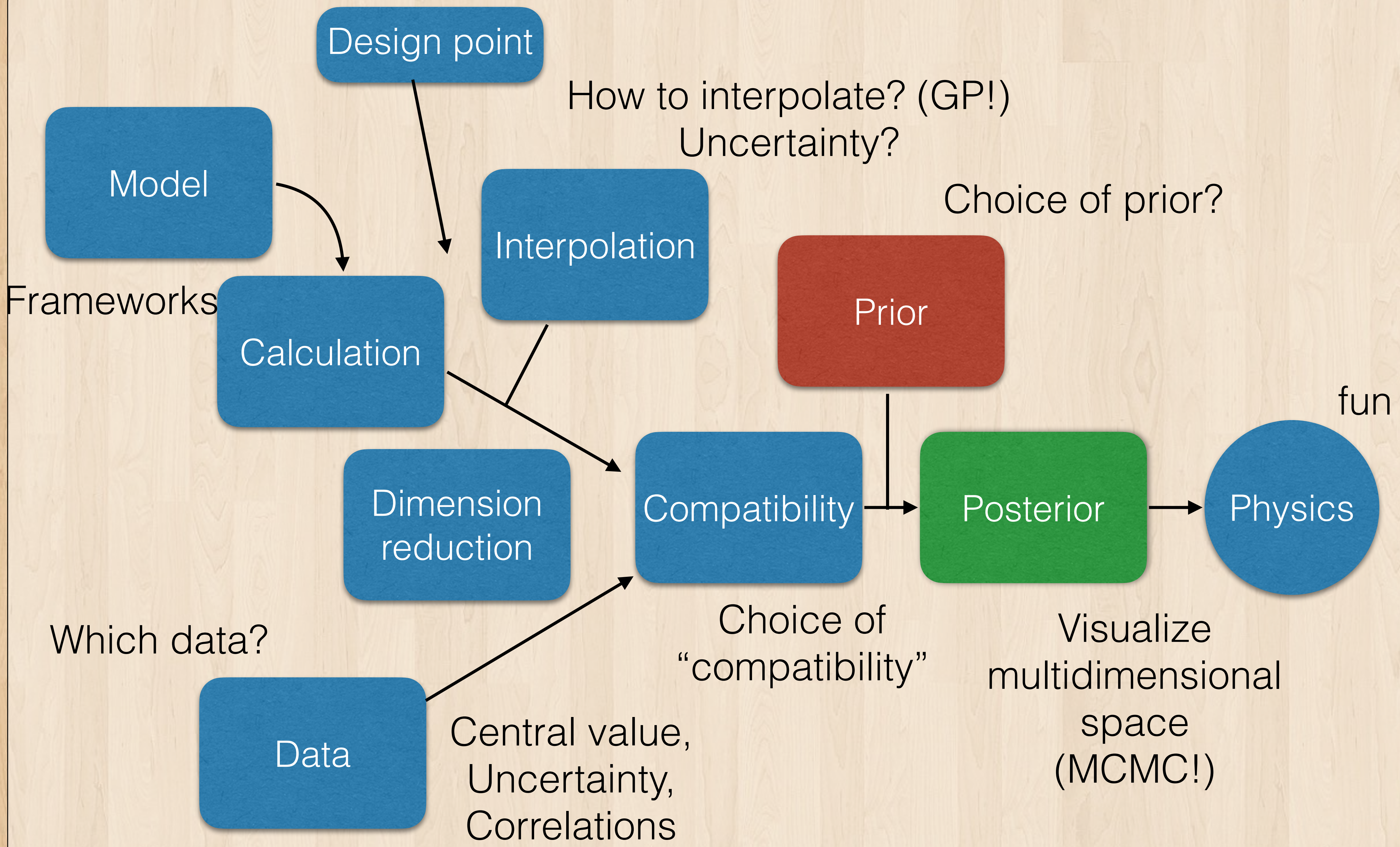
Bayesian evidence

Posterior: probability density of parameter $\vec{\theta}$ being “true” given the observed data

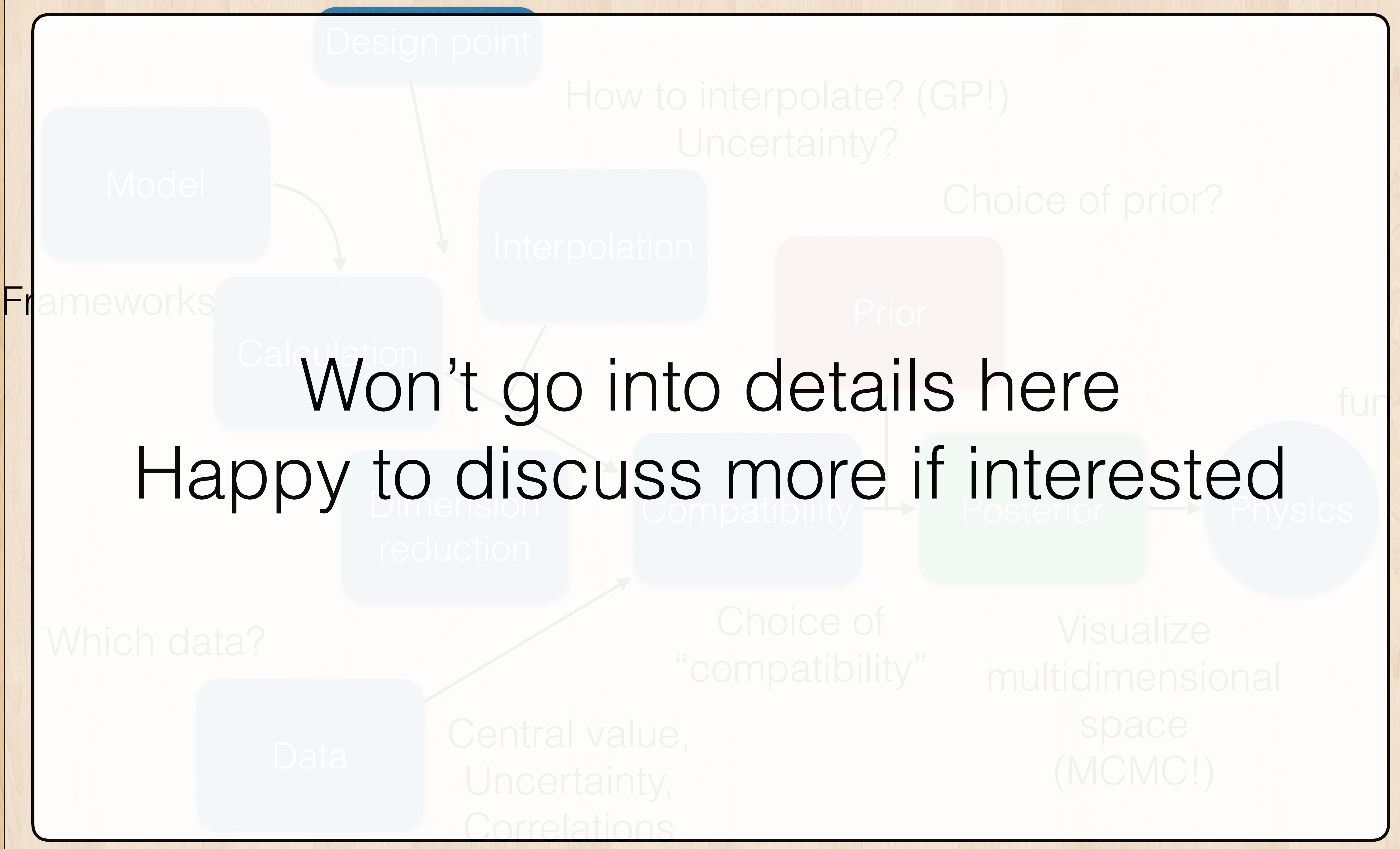
The Bayesian analysis



Under the hood…



Under the hood…



Won't go into details here
Happy to discuss more if interested

Advantages of the approach

- Computing requirements do not scale directly with volume of parameter space
  - Great for tackling complex problems that are hard to solve otherwise
- Rigorous control of analysis precision
  - Systematically improvable if higher precision on calculation is required

Limitations on the approach

- Requires a good enough model to begin with
- The analysis look for “best fit” within the parameter space associatedc with the model
- Computing intensive: we can do a lot more but some things are still a bit out of reach with current methods

Some recent efforts

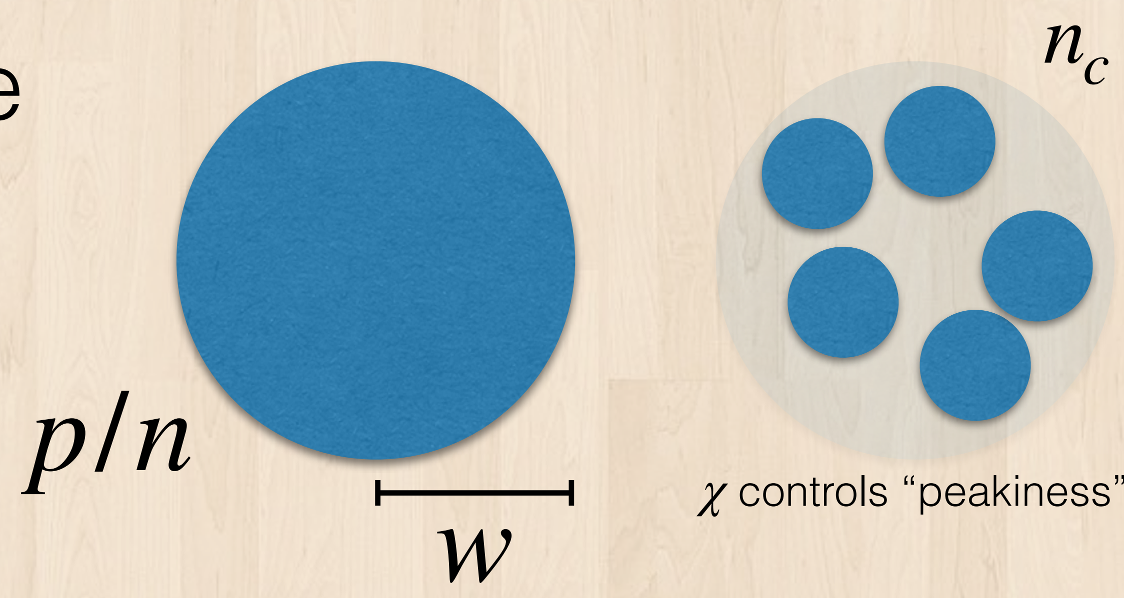
See also Shuzhe Shi talk Monday on Ru/Zr studies
See also Wilke van der Schee talk Monday on neutron skin

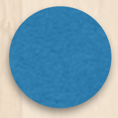
Initial state modeling variations

Nucleon location within nucleus generally sampled with Woods-Saxon with minimum distance $d_{\min}$

(Some use pregenerated profile, e.g. ^{16}O)

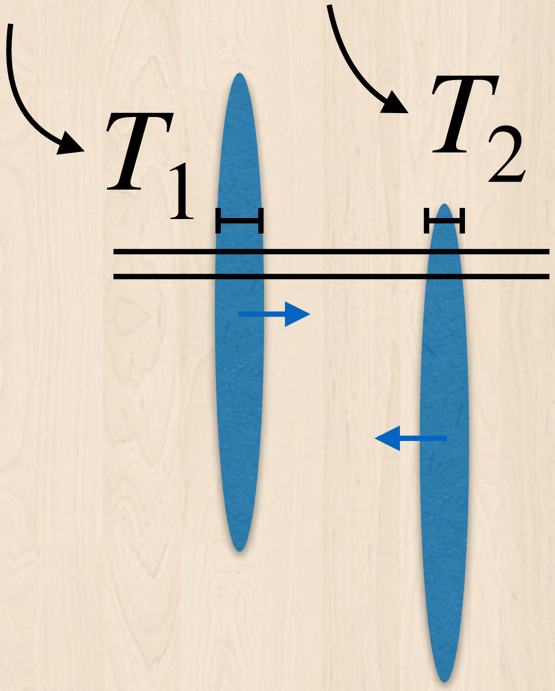
With & without substructure
Transverse profile



 = Gaussian

The Trento Ansatz

Thickness function: how much "interacting stuff" as a function of transverse location



Trento Ansatz

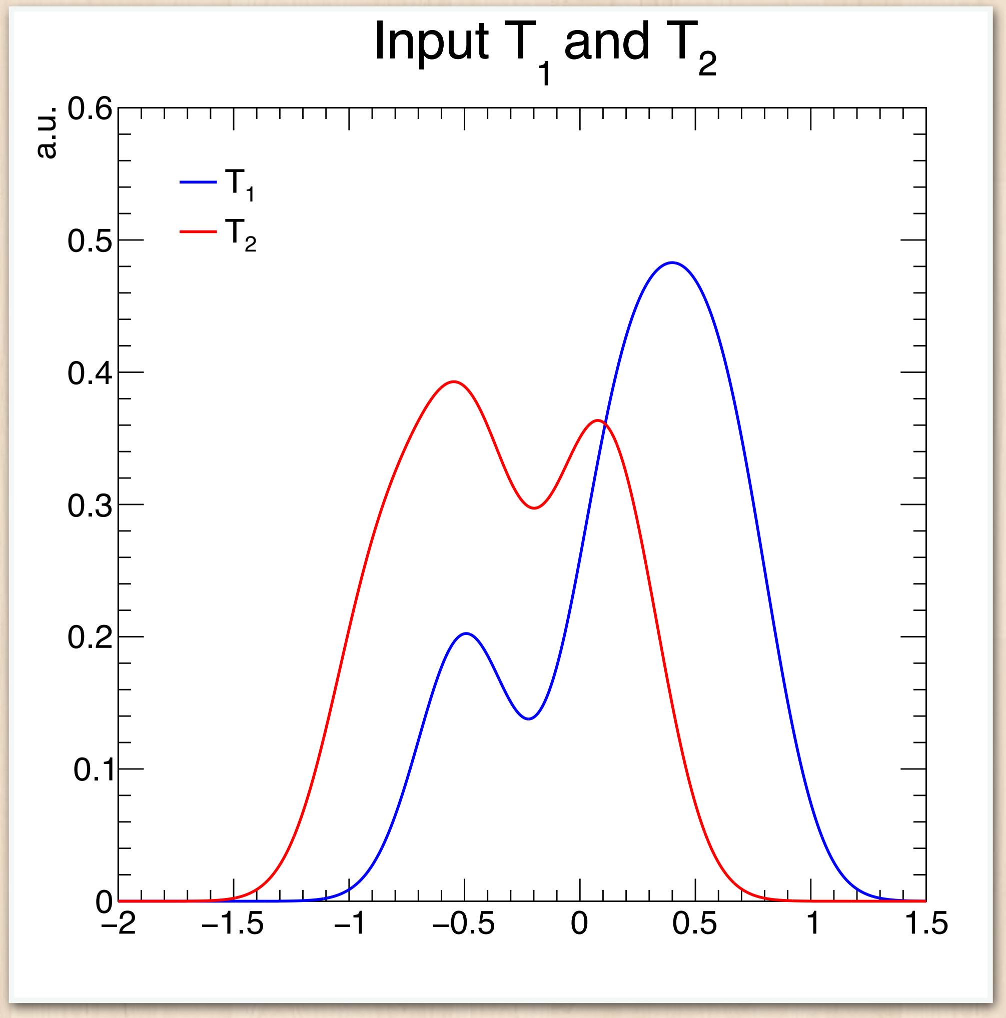
$$\text{entropy} \propto T_{12} = \left(\frac{T_1^p + T_2^p}{2} \right)^{1/p}$$

$\min(T_1, T_2)$	$\sqrt{T_1 T_2}$	$(T_1 + T_2)/2$	$\max(T_1, T_2)$	
$-\infty$	0	1	∞	p

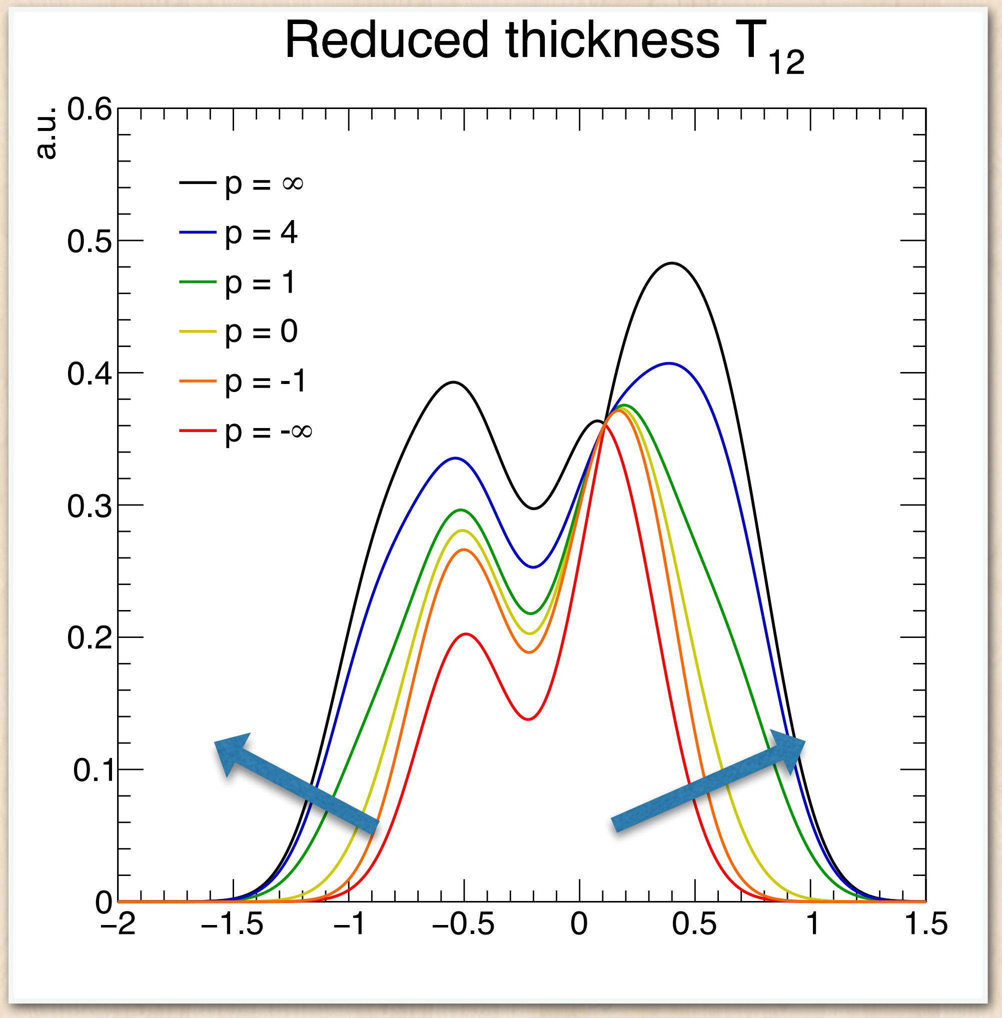
See also W. Ke talk last week

The Trento Ansatz: example

Toy example



Larger p ~more diffuse



$$\min(T_1, T_2) \quad \sqrt{T_1 T_2} \quad (T_1 + T_2)/2 \quad \max(T_1, T_2)$$

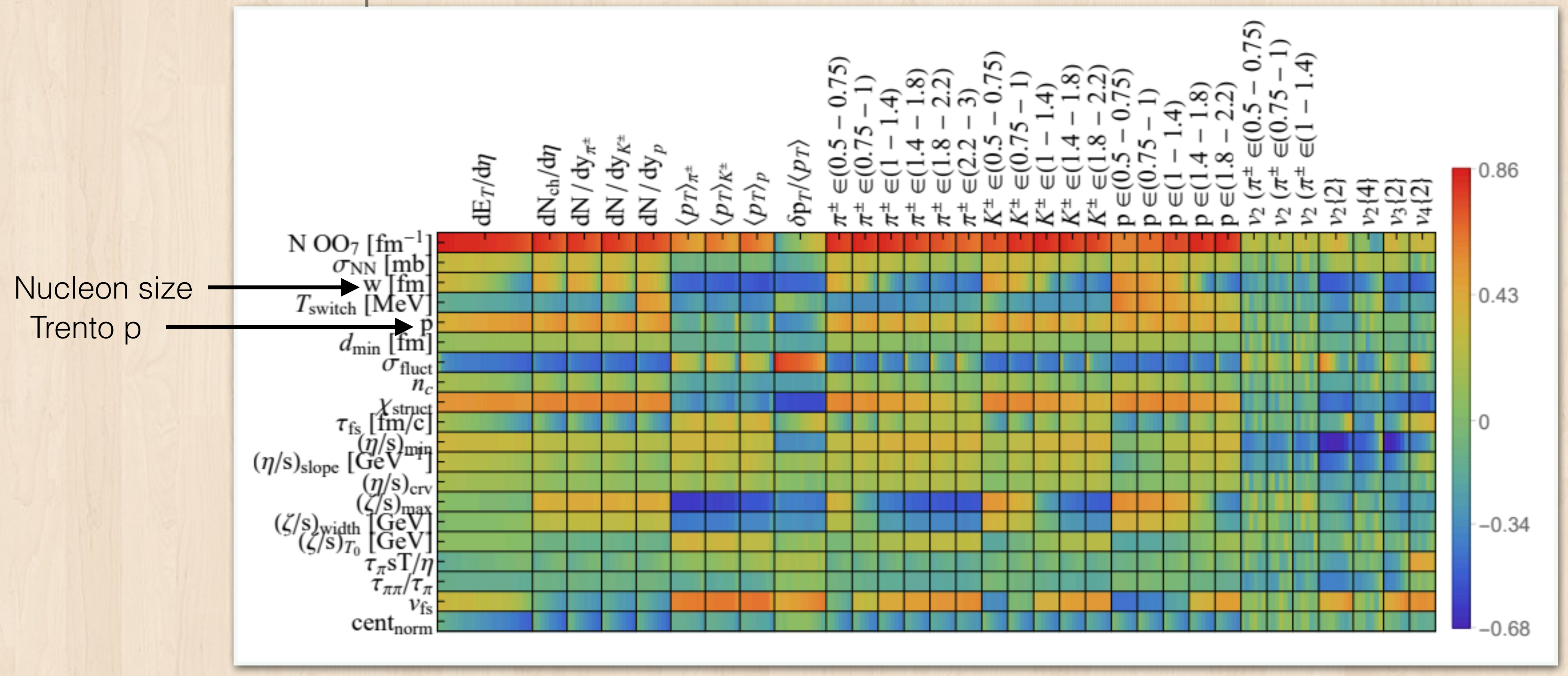
$$-\infty \longrightarrow 0 \longrightarrow 1 \longrightarrow \infty \quad p$$

1412.4708

Sensitivity analysis

Explore sensitivity of parameters to observables

Example



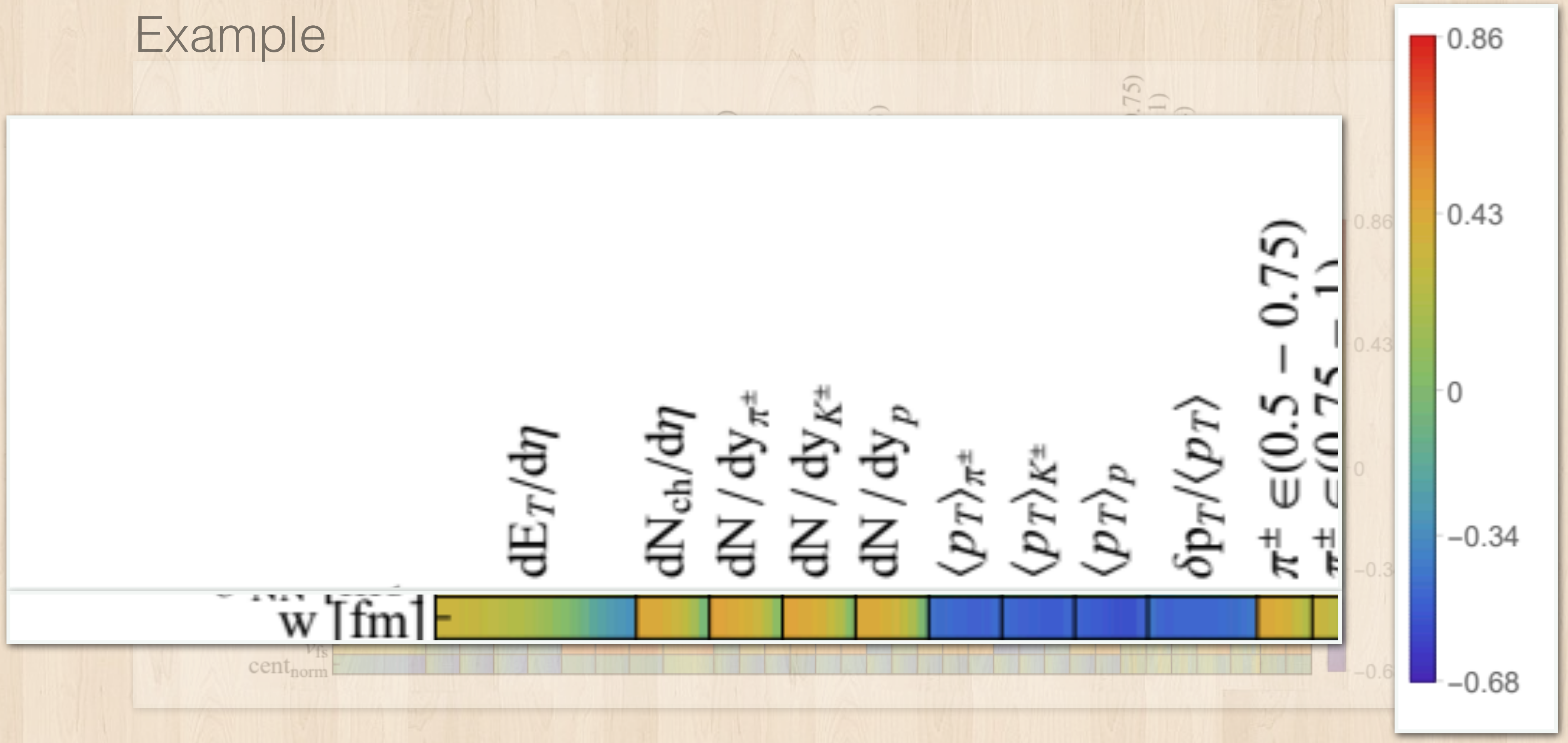
Build physics intuition and guide future efforts

2010.15134

Sensitivity analysis

Explore sensitivity of parameters to observables

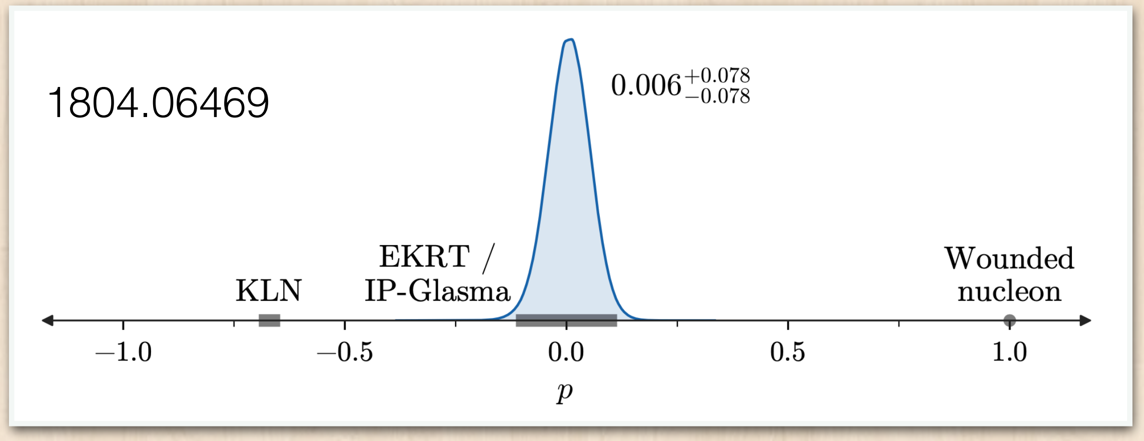
Example



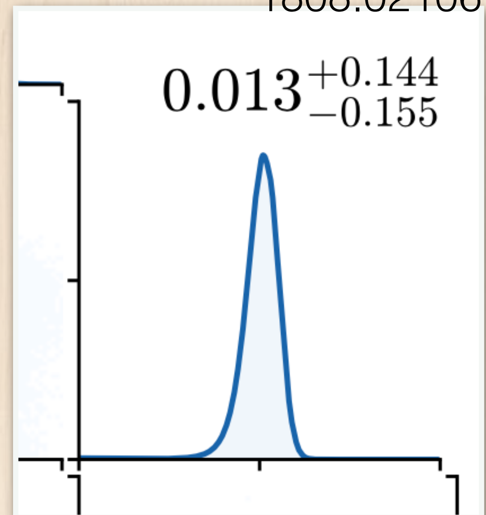
Build physics intuition and guide future efforts

2010.15134

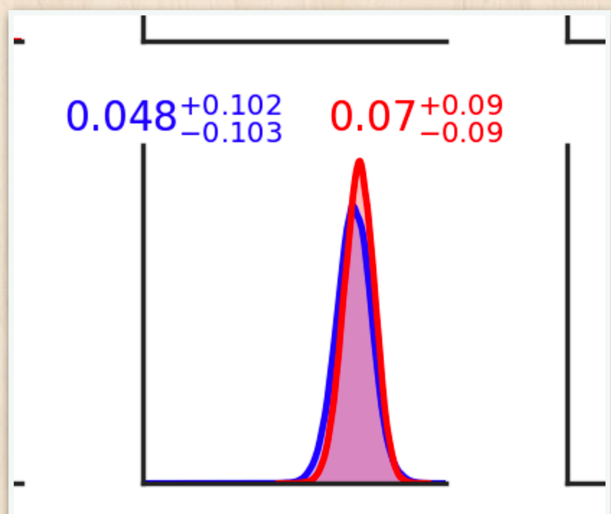
The parameter p: examples



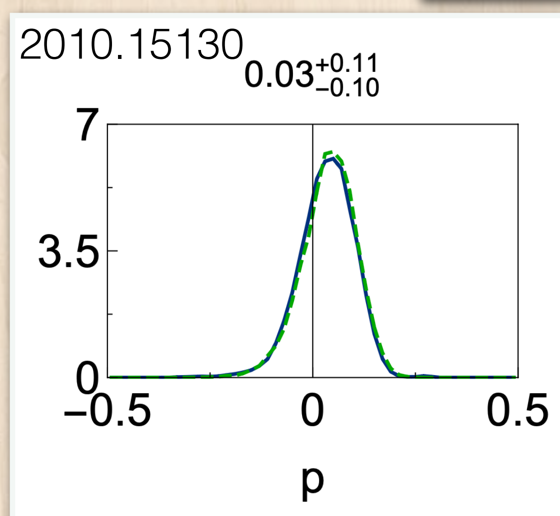
1808.02106

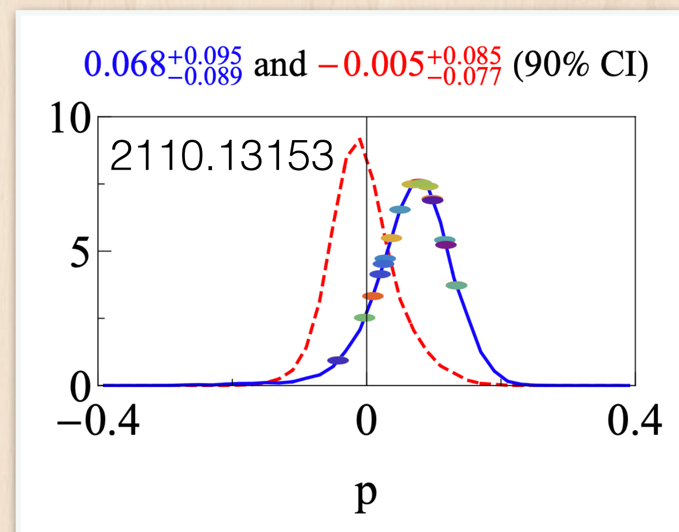


$$T_{12} = \left(\frac{T_1^p + T_2^p}{2}\right)^{1/p}$$

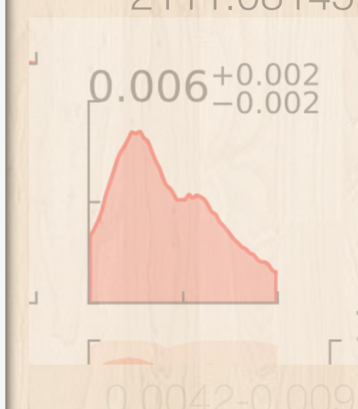


2011.01430





2111.08145



The Trento p parameter seems quite consistent across the board?

Two flavors of parameter p

- Original parametrization

$$\frac{dS}{dy} \propto \left(\frac{T_1^p + T_2^p}{2} \right)^{1/p} \xrightarrow{p=0} (T_1 T_2)^{1/2}$$

- Some work choose to use

$$\frac{dE}{d\eta} \propto \left(\frac{T_1^p + T_2^p}{2} \right)^{1/p} \xrightarrow{p=0} (T_1 T_2)^{1/2}$$

→ more diffuse in general for same $p = 0$

Generalization of the Ansatz

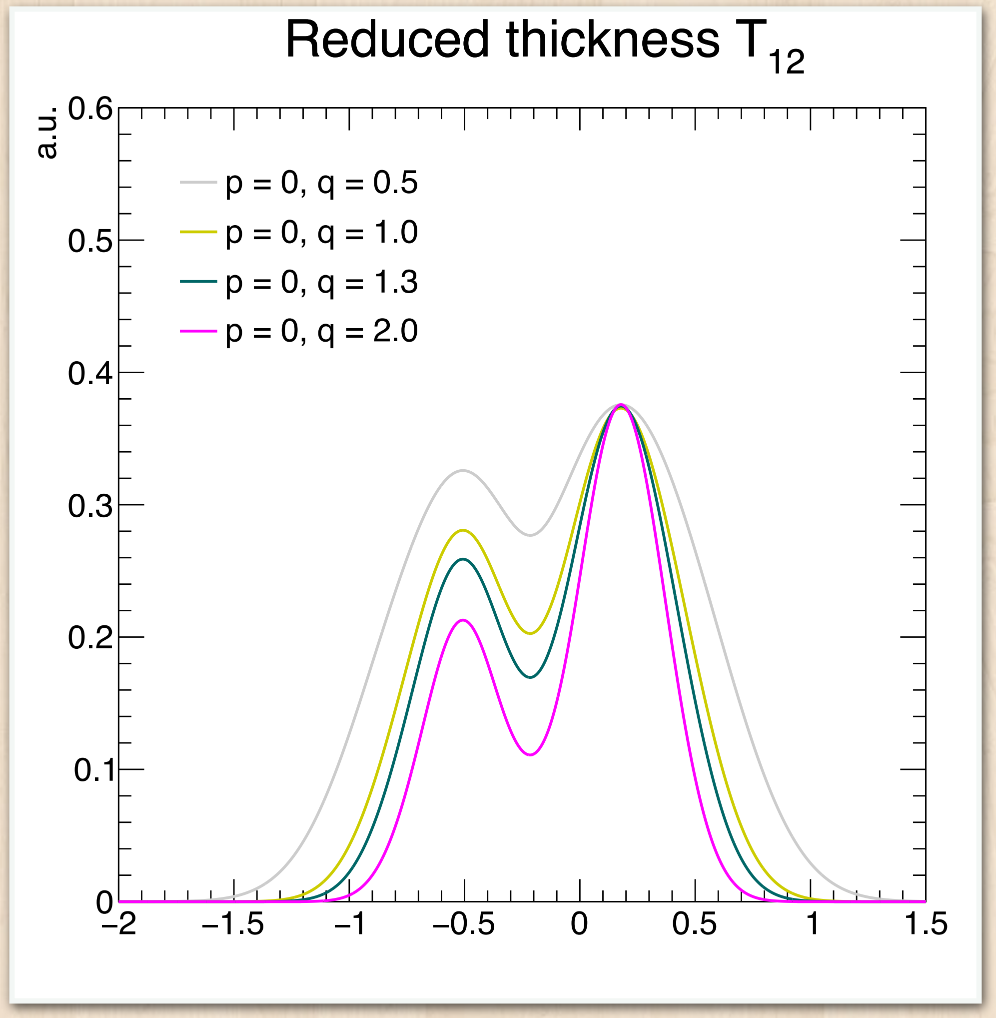
Participant scaling $T_{12} \sim T_1 T_2$ not included in original Ansatz

Generalized:

$$\frac{dS}{dy} \propto T_{12} = \left(\frac{T_1^p + T_2^p}{2} \right)^{q/p}$$

$q = 2$ (& $p = 0$) $\rightarrow T_{12} \sim T_1 T_2$ disfavored

The toy example from before



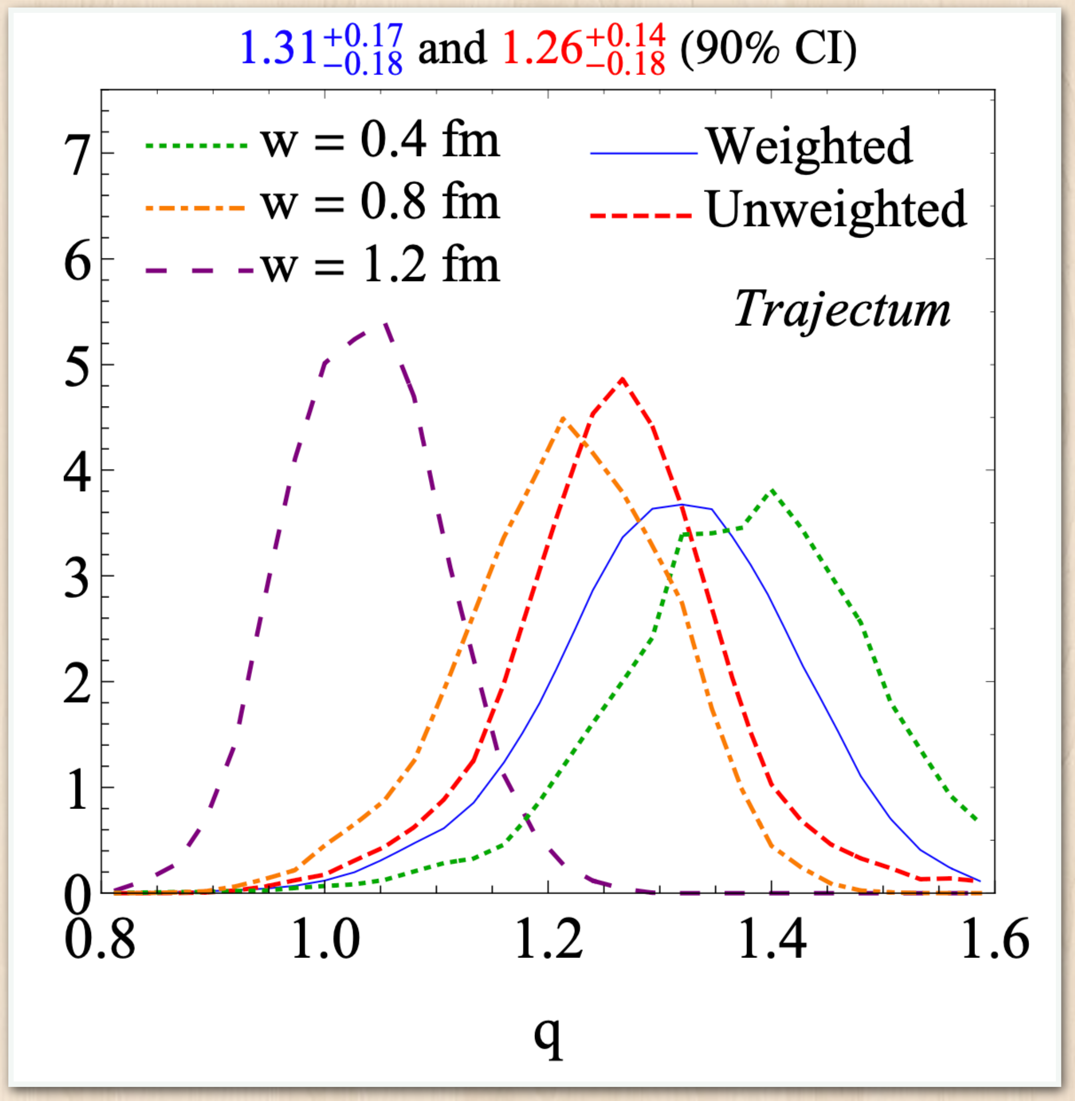
~sharper for larger q

Generalization of the Ansatz

Participant scaling $T_{12} \sim T_1 T_2$ not included in original Ansatz

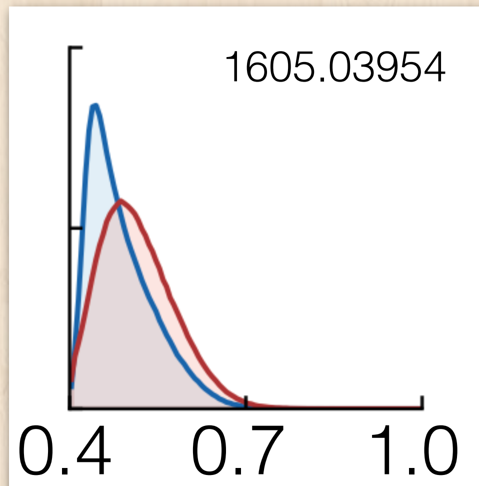
Generalized:

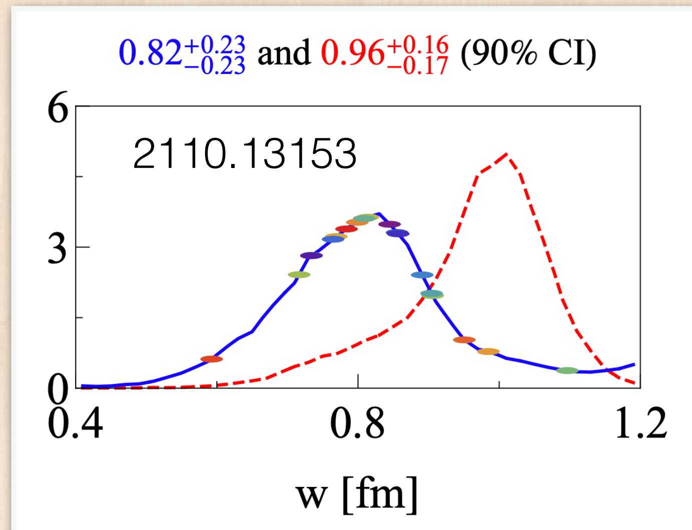
$$\frac{dS}{dy} \propto T_{12} = \left(\frac{T_1^p + T_2^p}{2} \right)^{q/p}$$

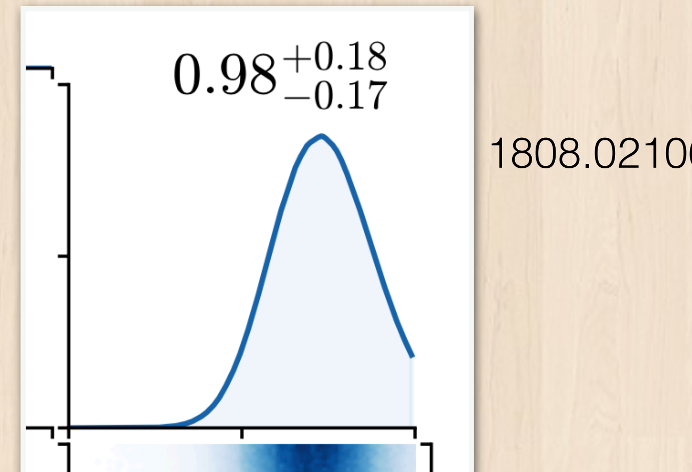


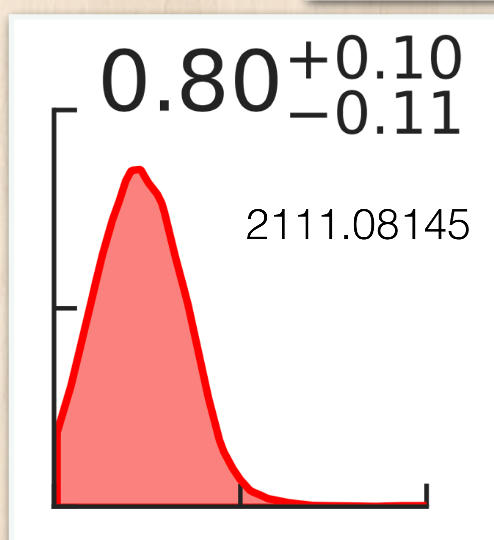
$q = 2$ (& $p = 0$)
$\rightarrow T_{12} \sim T_1 T_2$ disfavored

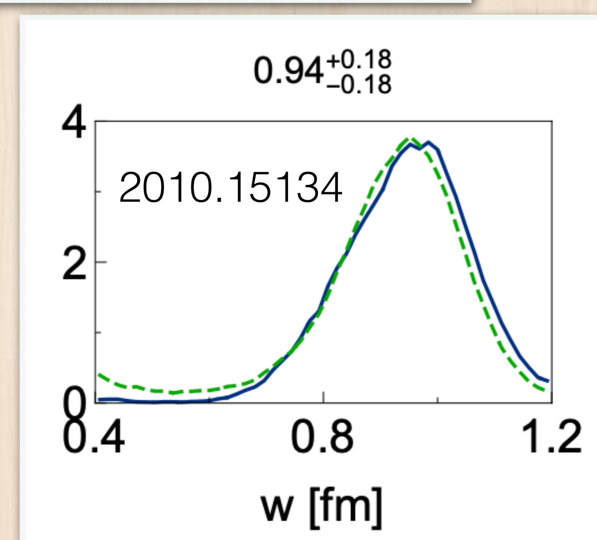
What about nucleon width?

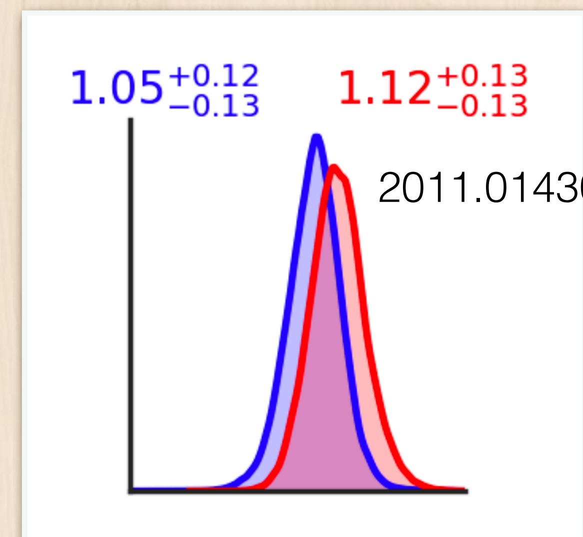












Discrepancy observed across analyses

Also they are huge!?

Proton charge radius ~0.84 fm
IP-Glasma uses a smaller value

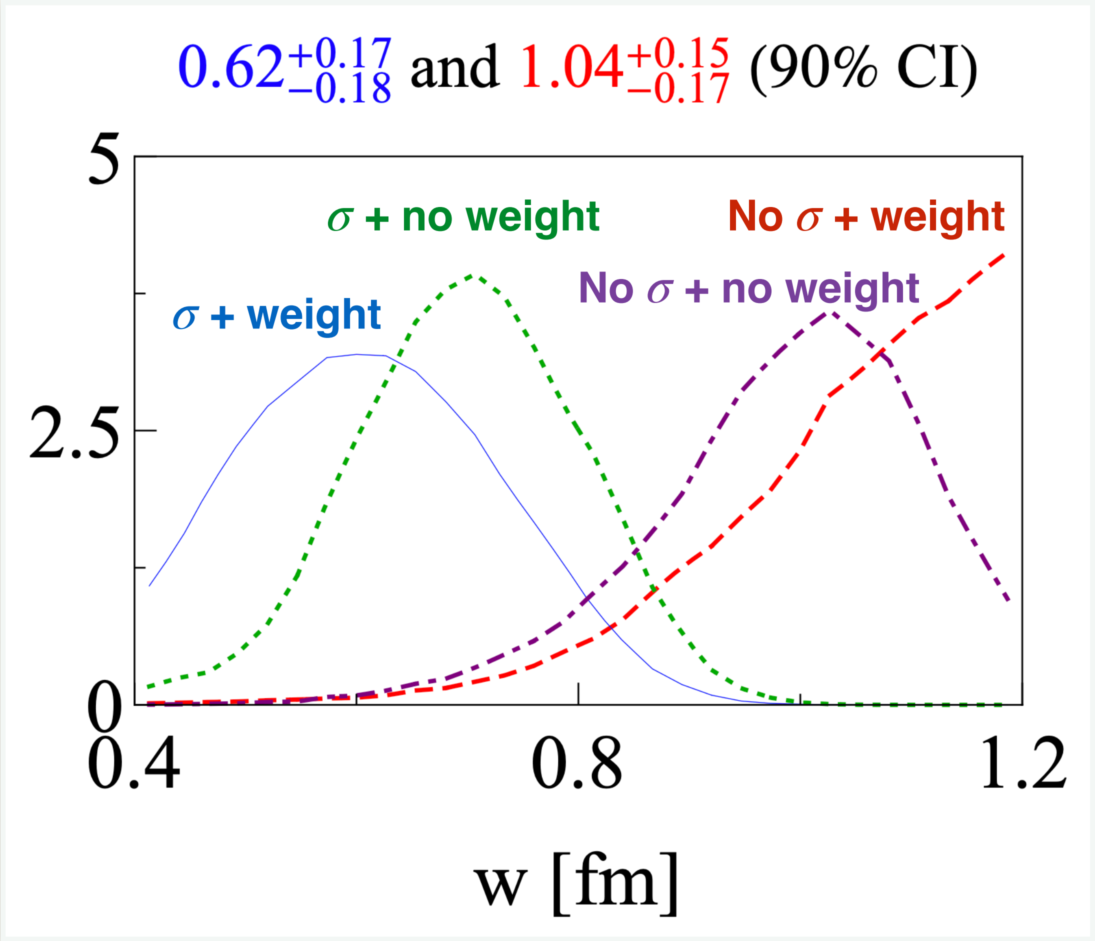
See also 2208.06839

A study with σ_{pA} and σ_{AA}

In addition to using the usual observables, add also total inelastic cross section σ_{pA} and σ_{AA}

larger w
→ diffuse nucleon
→ smaller cross section

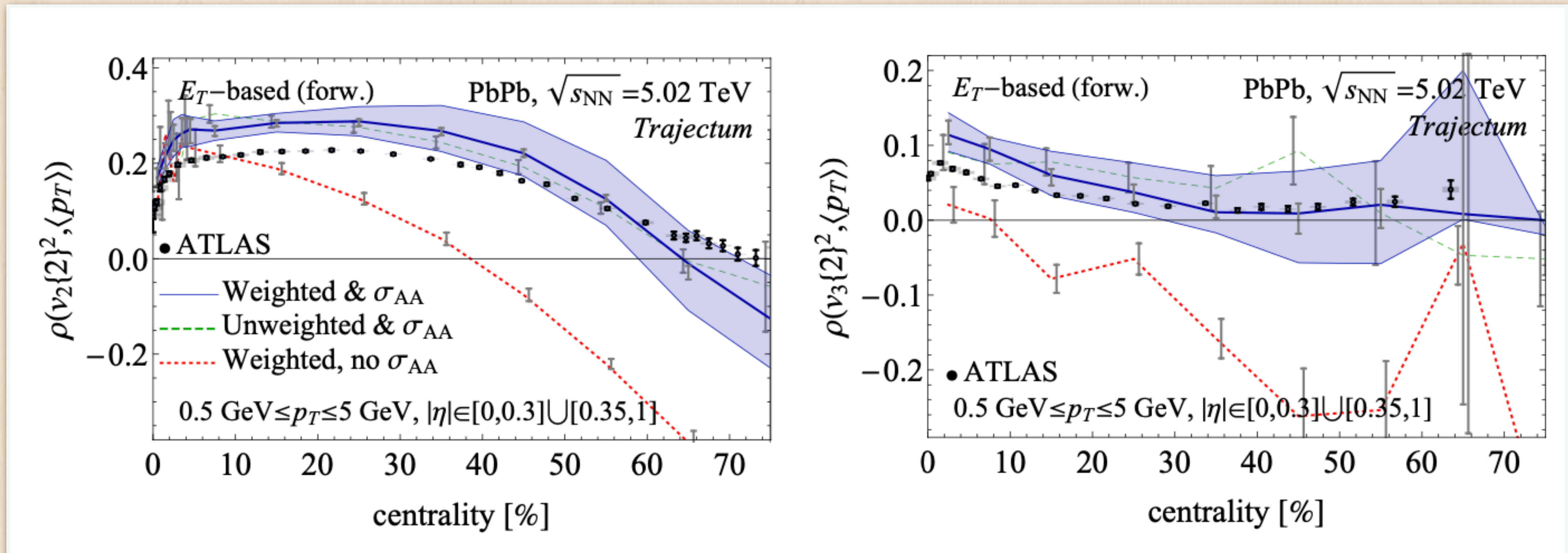
Additionally weight observable based on "trust": ones we believe should model better are weighted more heavily



2206.13522

A study with σ_{pA} and σ_{AA}

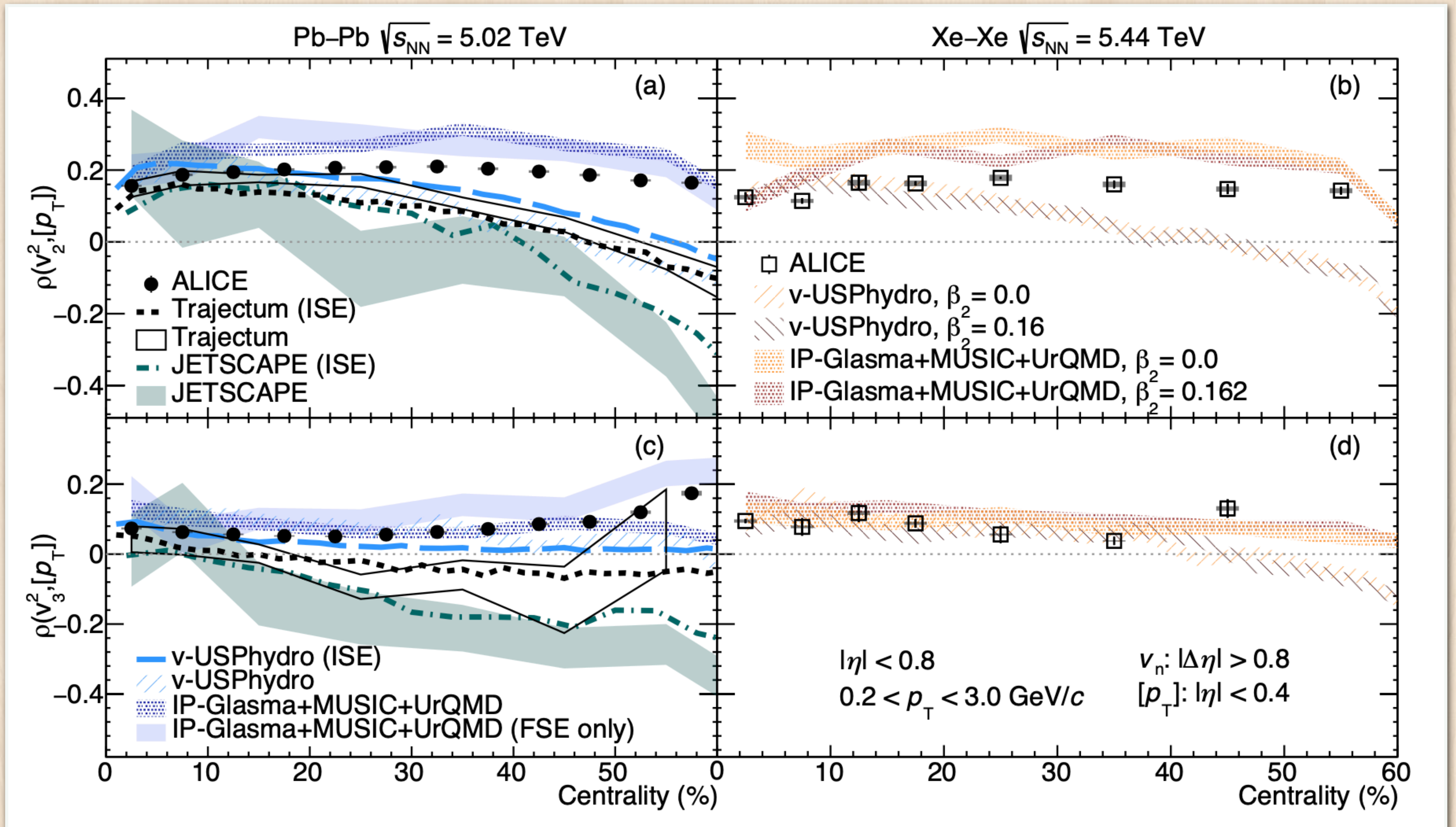
Testing the Bayesian analysis outcome
on correlation observables



Indeed including cross section improves the description

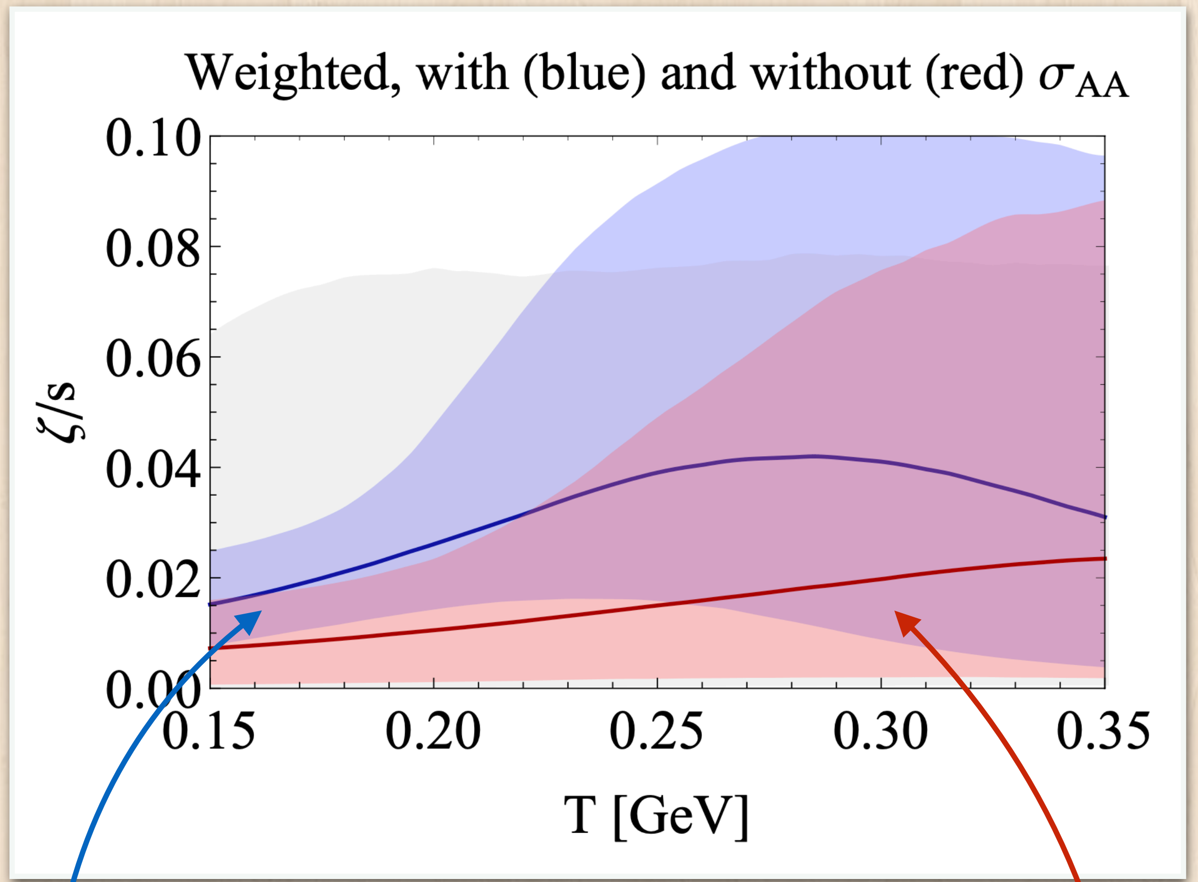
2206.13522

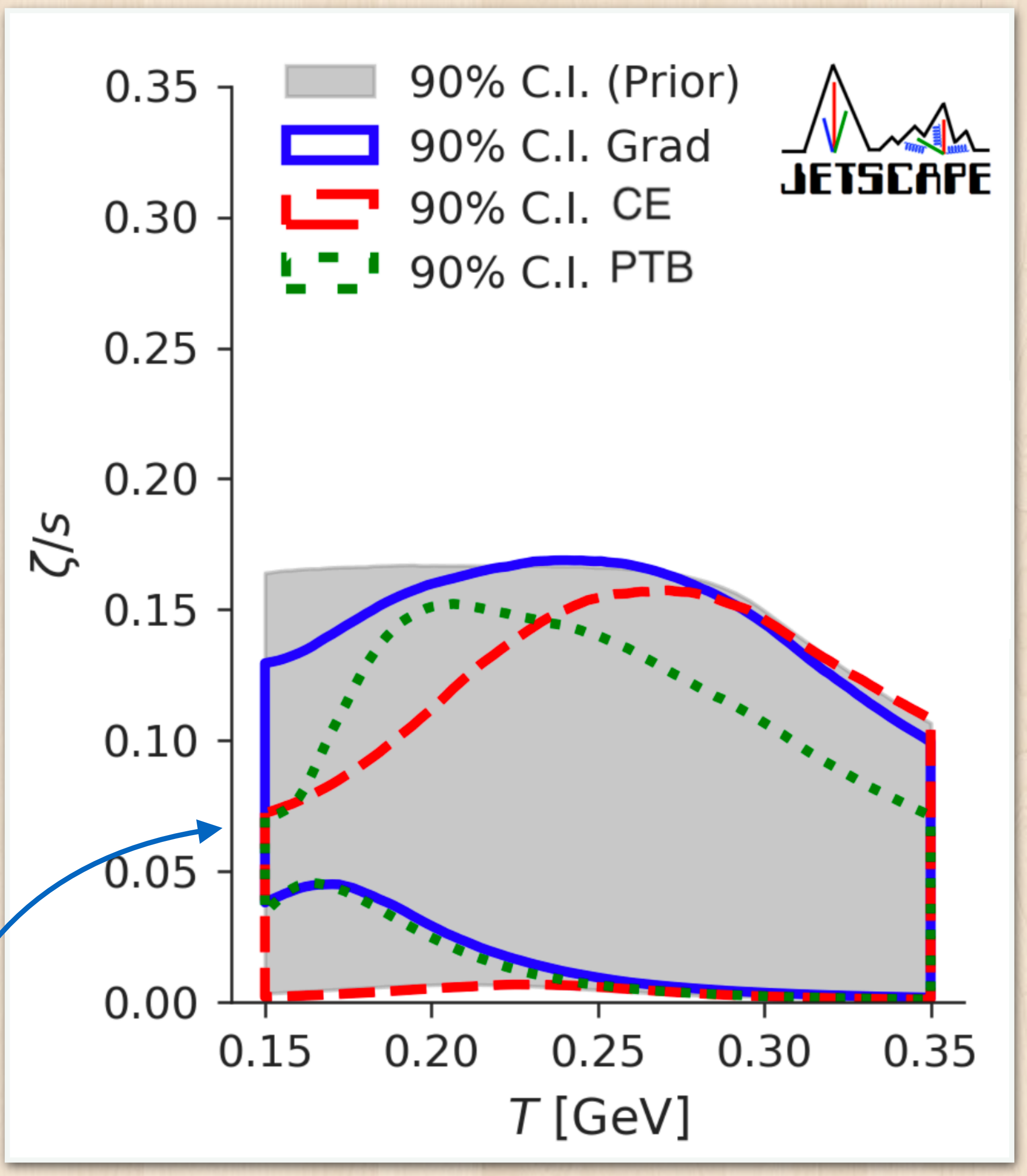
cf. $(p_T - v_n^2)$ correlation before



2111.06106

Effect on viscosity





With σ_{AA}/σ_{pA}, w ~ 0.5 fm

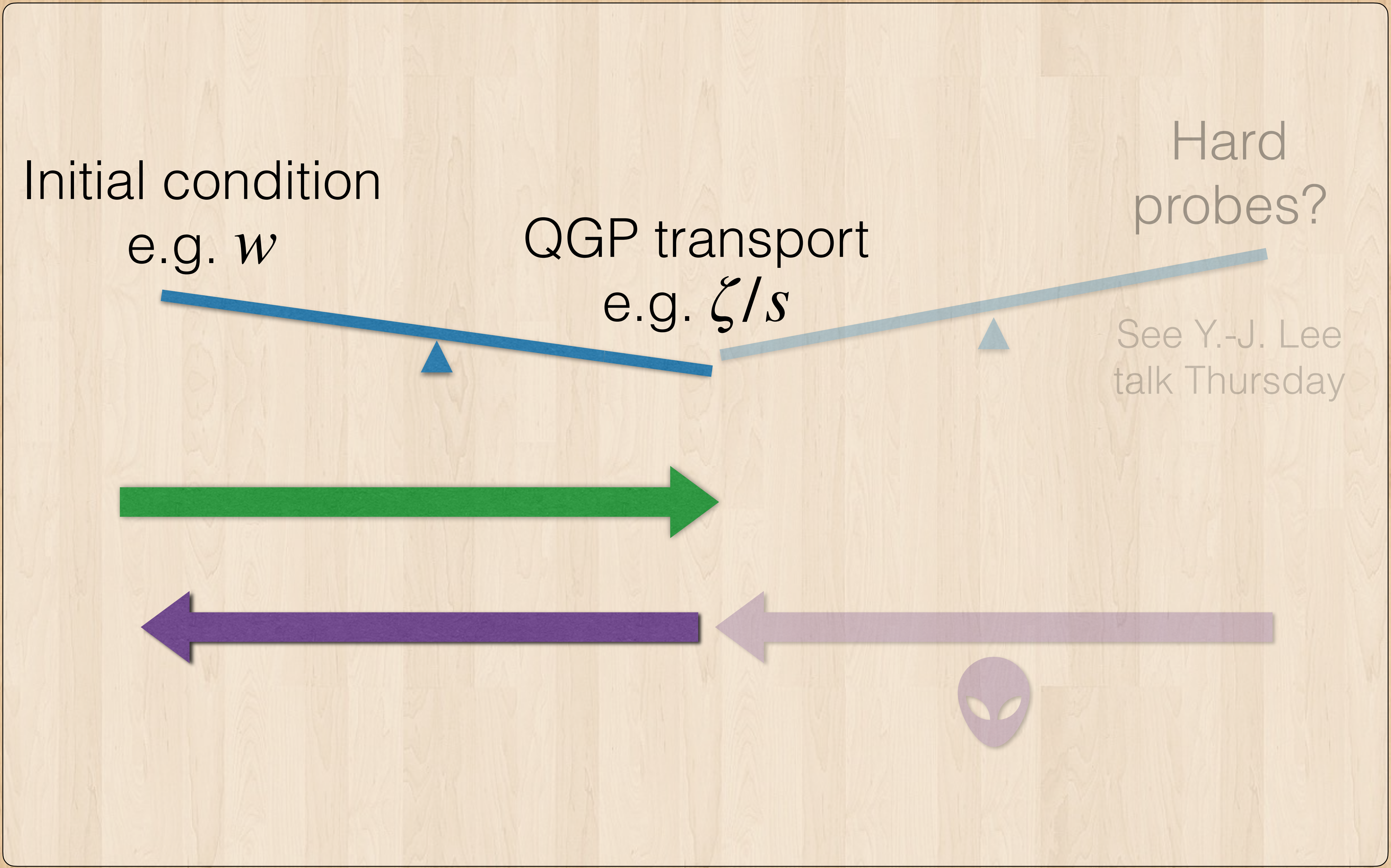
Without σ_{AA}/σ_{pA}, w ~ 1.0 fm

Compensating effect rippling through QGP parameters

2206.13522

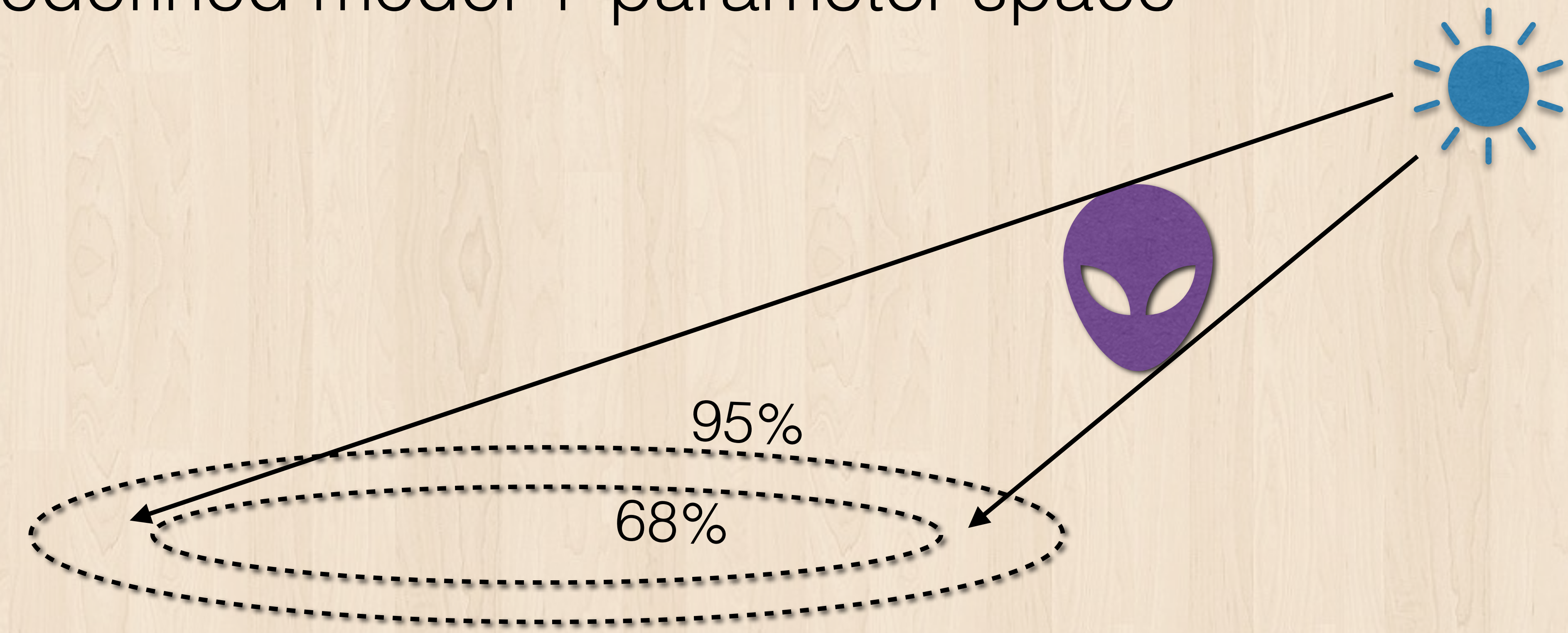
2011.01430

Opportunity



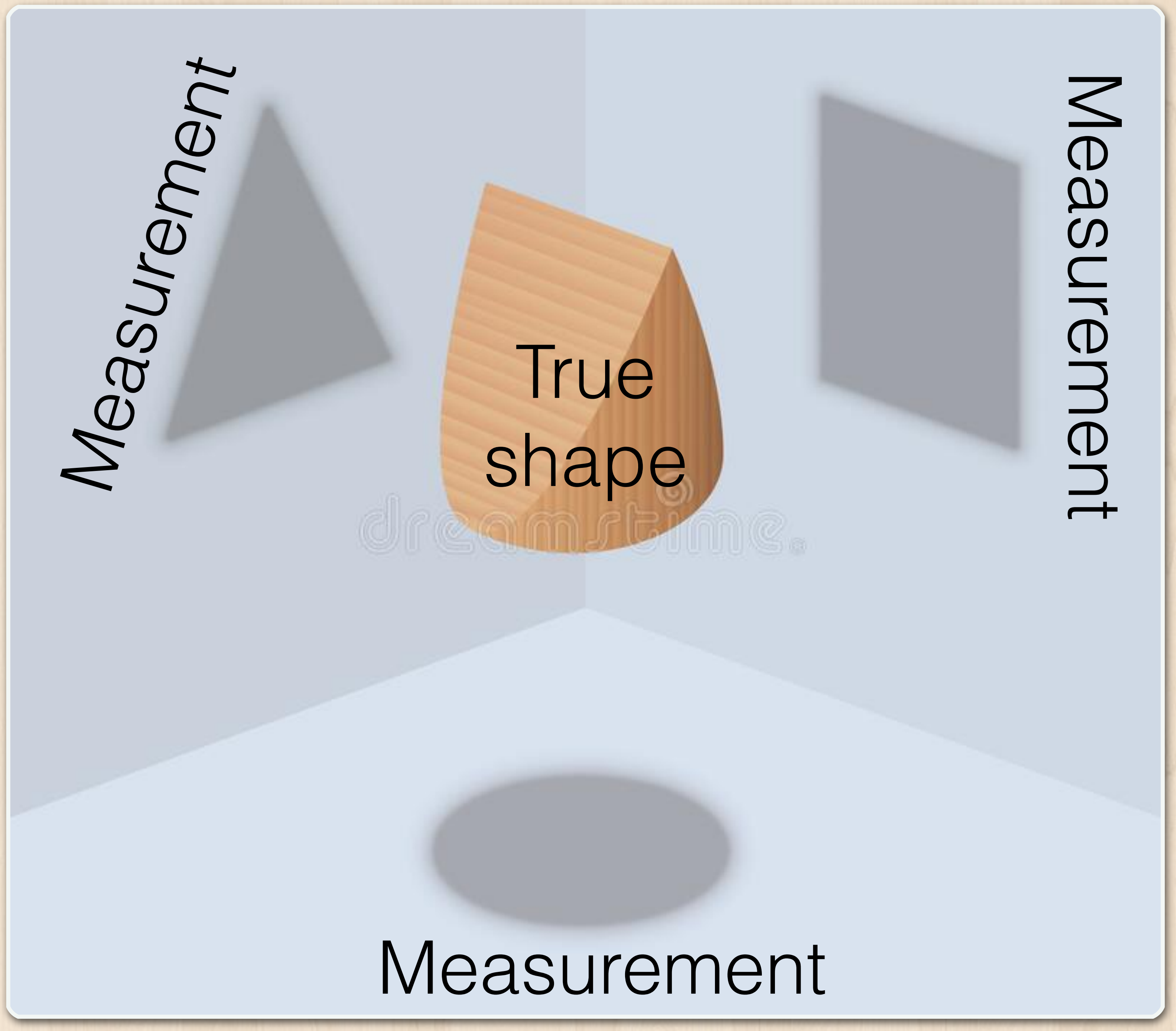
Notes

- Remember one of the main features of the Bayesian analysis: it searches for the best parameters within a predefined model + parameter space



- Only by systematically including/designing/checking more and more observables and physics into models can we hope to see the full picture

Measurements vs truth

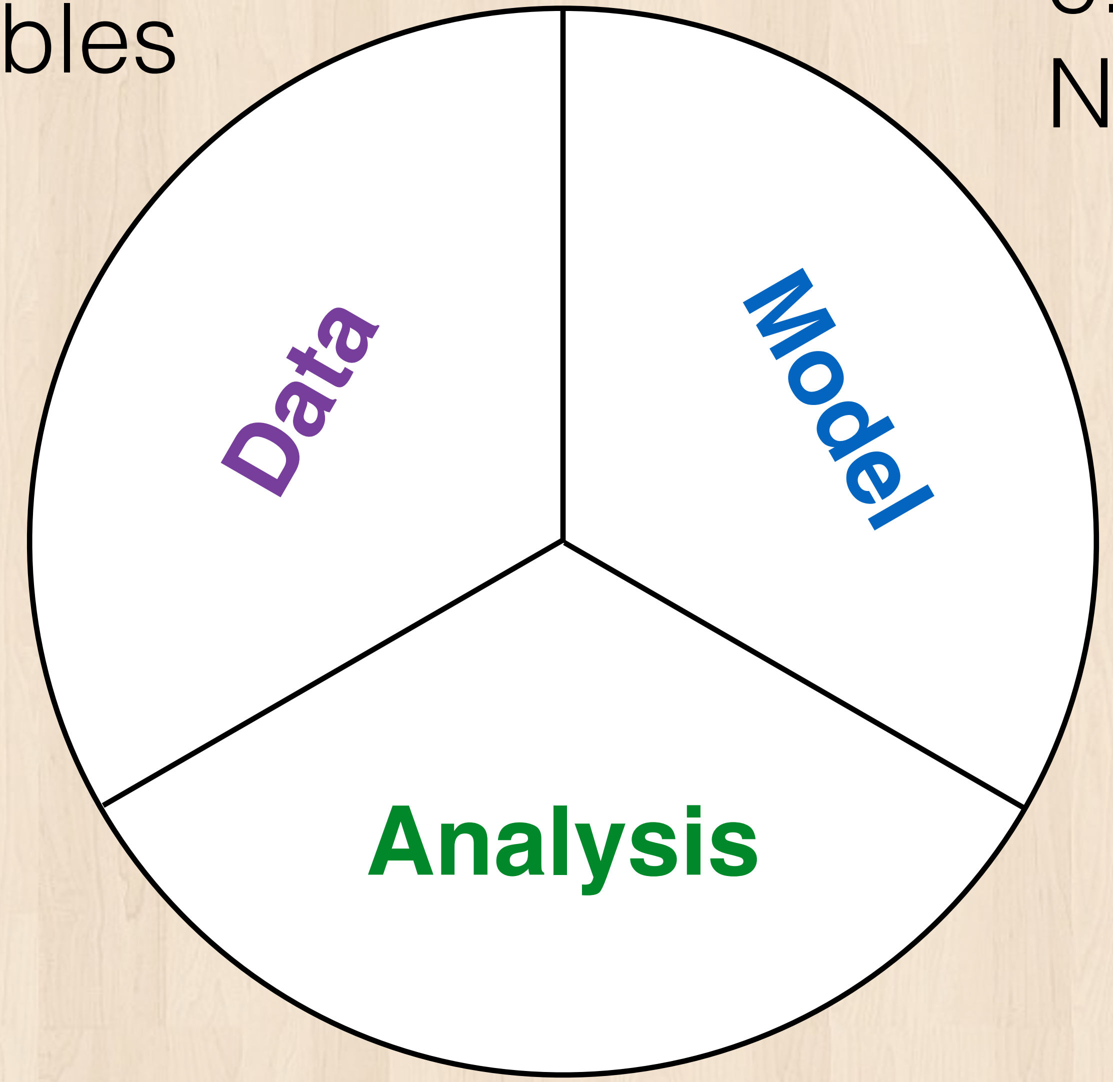


Looking foward

Many things can be improved

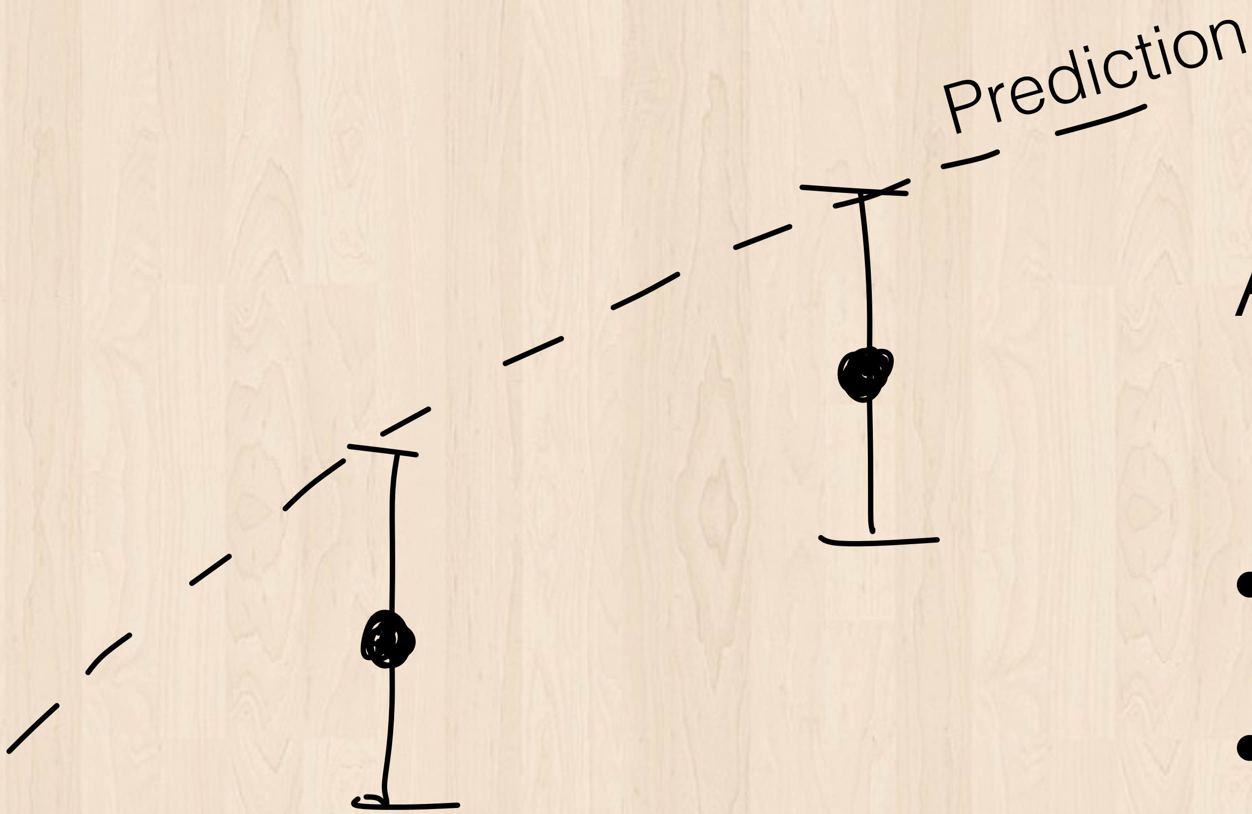
e.g.
New observables
Uncertainty
reporting

e.g.
Newer models



e.g.
better utilization of
computing resources

Data uncertainty correlation



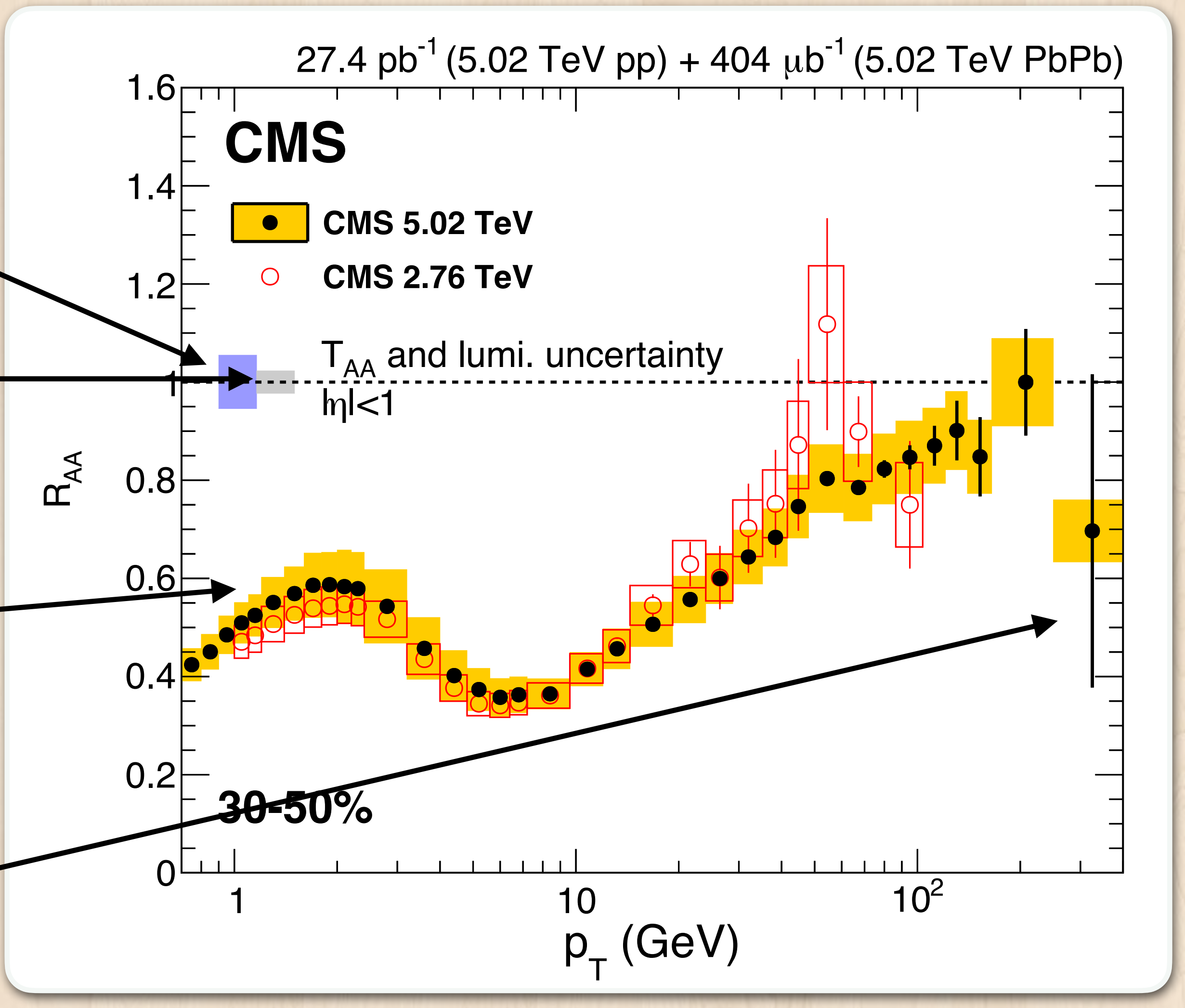
Agreement depends on uncertainty correlation

- Fully Correlated: 1σ
- Non-correlated: 2σ
- Anti-correlated: >2σ

Correlation is key!

Faithfully capturing the correlation is **crucial**

Capture Correlations



T_{AA}

Luminosity

Other Systematic Uncertainties

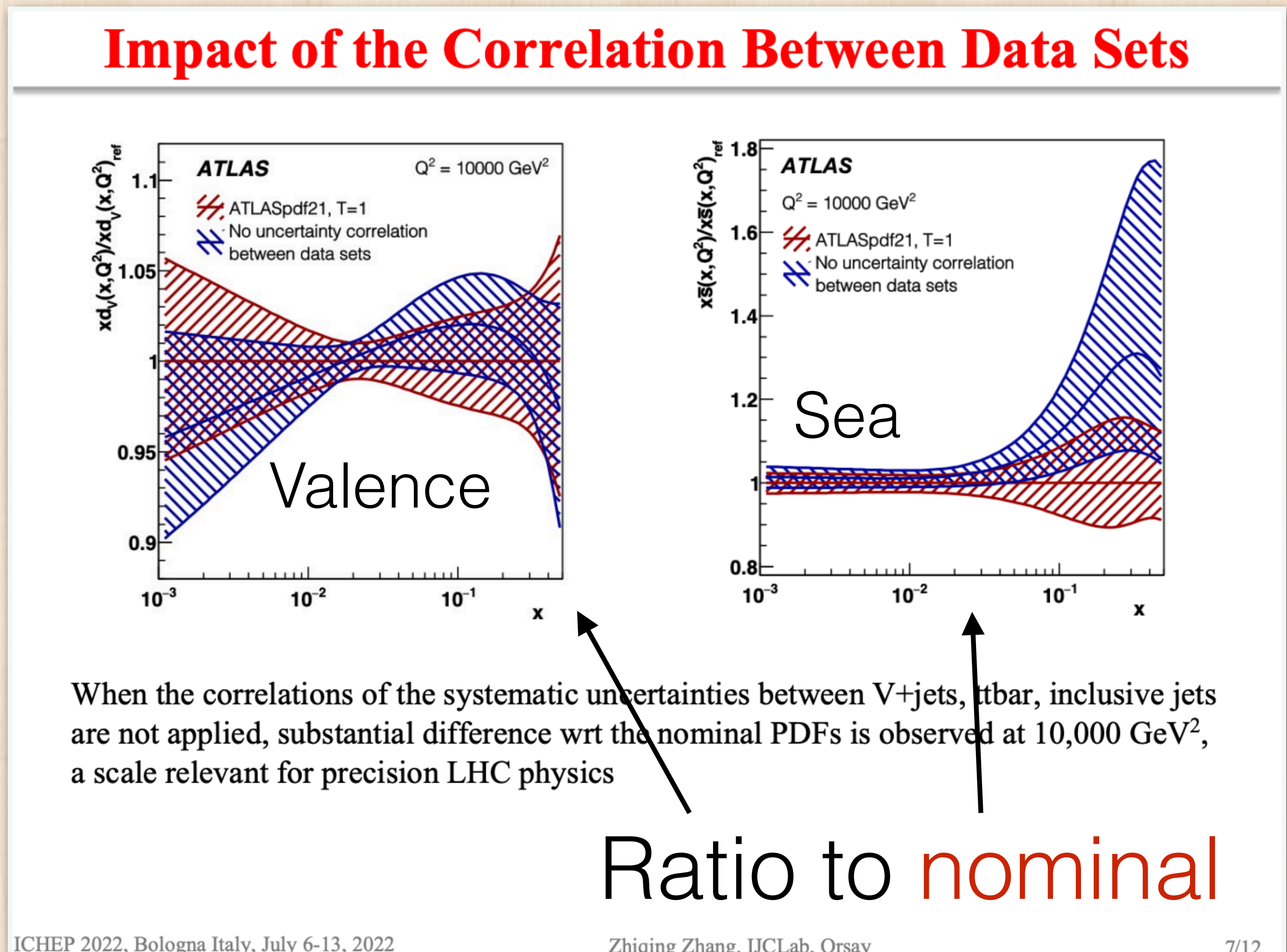
Statistical Uncertainty

Many uncertainties with different correlations

As much as reasonably possible!

<u>More information from experiments will be nice</u>

cf. pdf fits & statistics

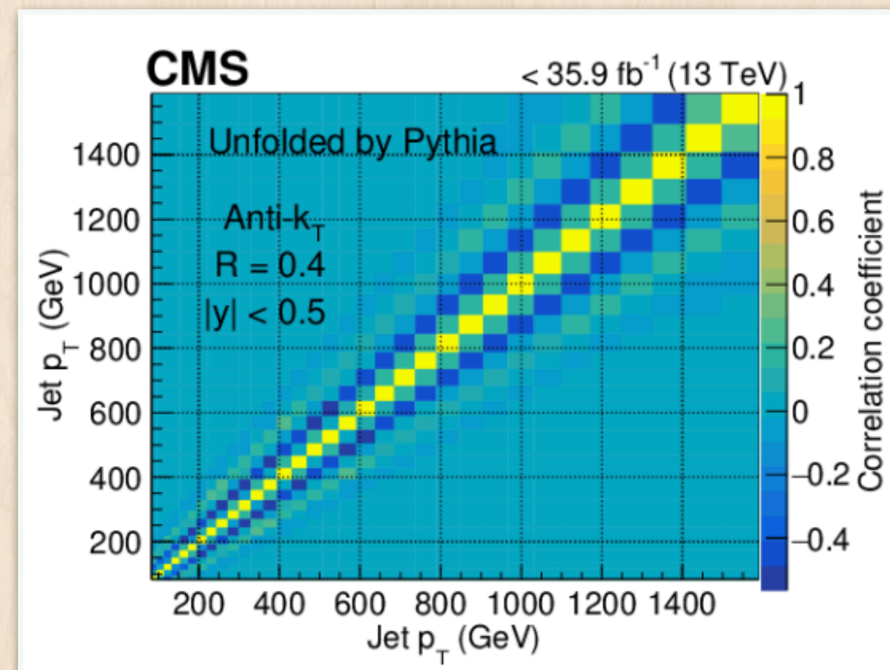


Effect of **inter-dataset** correlation

Dip at ~resolution

Wake at 2x resolution

etc.

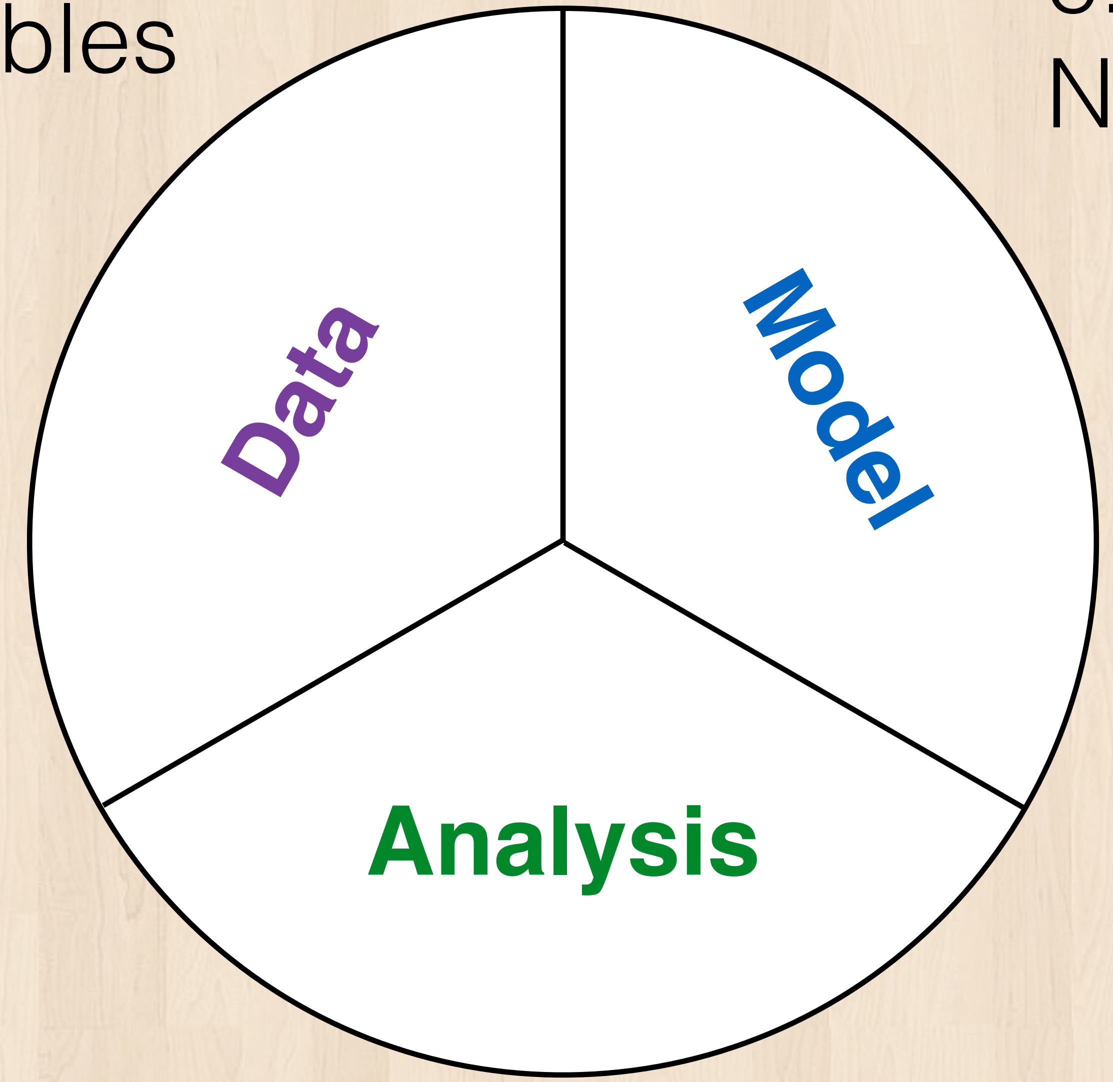


Common feature

After unfolding

Many things can be improved

e.g.
New observables
Uncertainty reporting



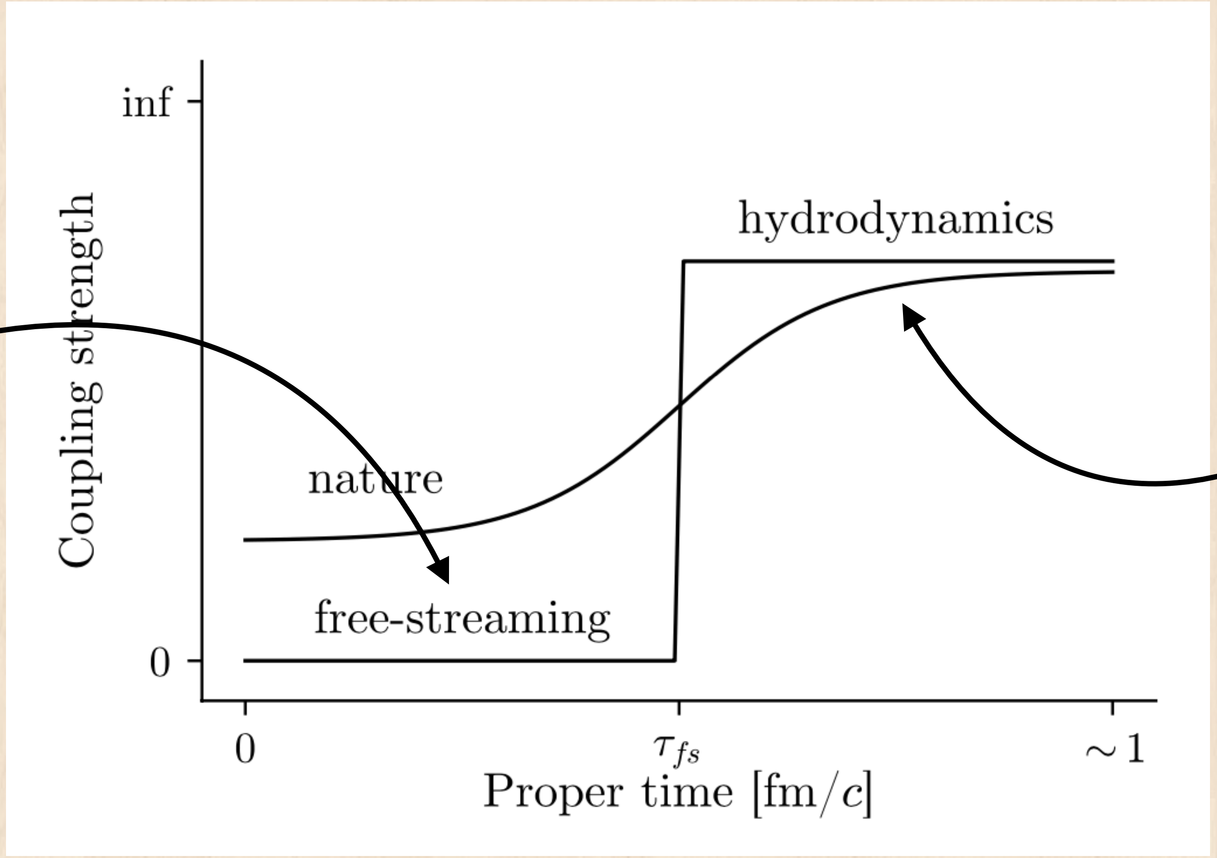
e.g.
Newer models

e.g.
better utilization of computing resources

Interface?

Current efforts split things up into different phases

“Pre-hydro”



“Hydro”

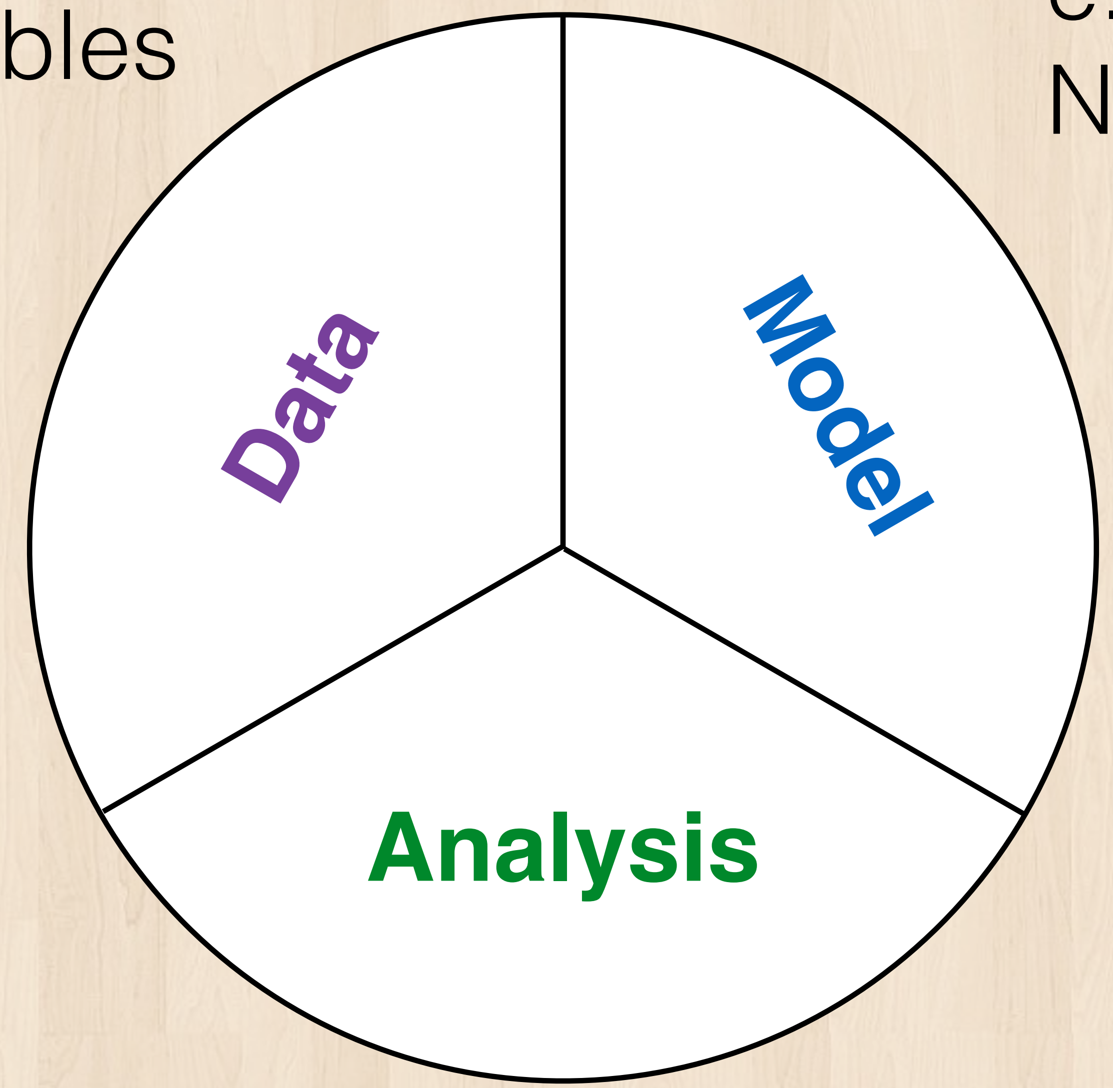
What are the implications?
Challenge for modeling

1808.02106

Many things can be improved

e.g.
New observables
Uncertainty reporting

e.g.
Newer models



e.g.
better utilization of computing resources

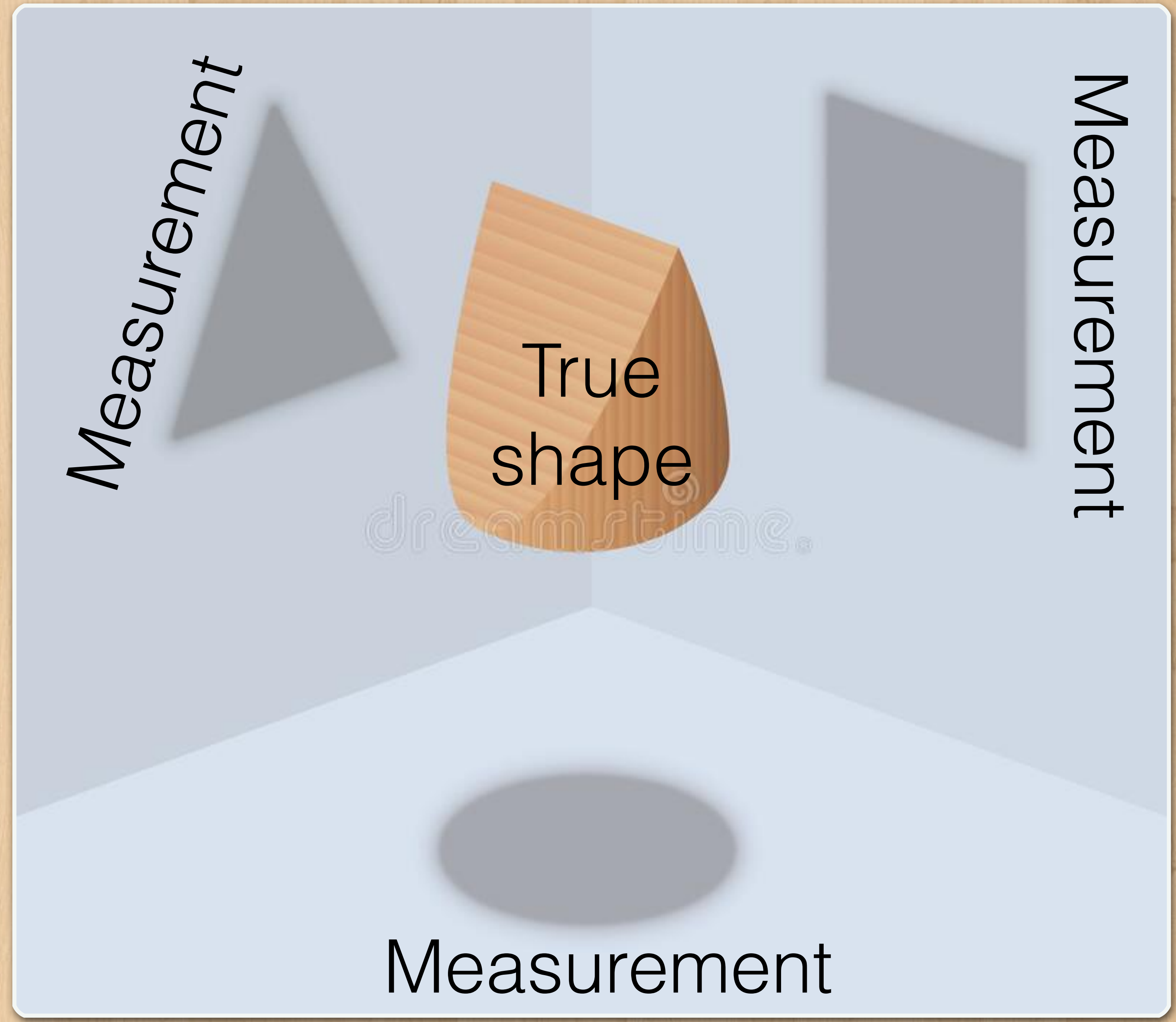
Analysis advancements

- How to perform the analysis with a similar precision but with a smaller amount of computing resources?
  - Many interesting developments!
- Great opportunity for cross talk among different physics subfields and statistics/CS communities

Concluding Remarks

Concluding remarks

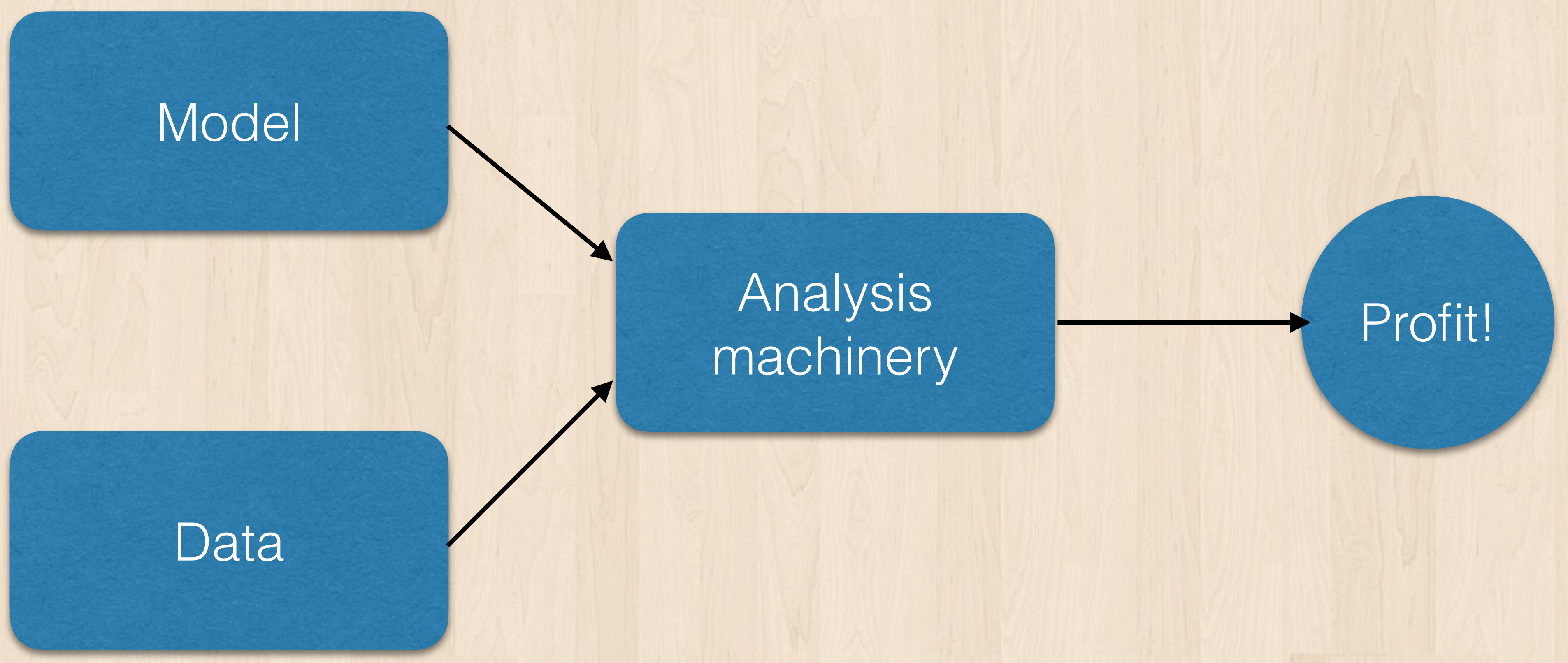
- Bayesian analysis is a powerful tool to help us distill more nuanced information from data
- A number of efforts in recent years extracting initial conditions and QGP transport parameters
  - Trento-based initial conditions
  - Interesting constraint observed
  - Ripple effect across parameters
  - Check obtained result on as many observables as possible
- Feedback for observable design is important


Measurement
True shape
Measurement
Measurement

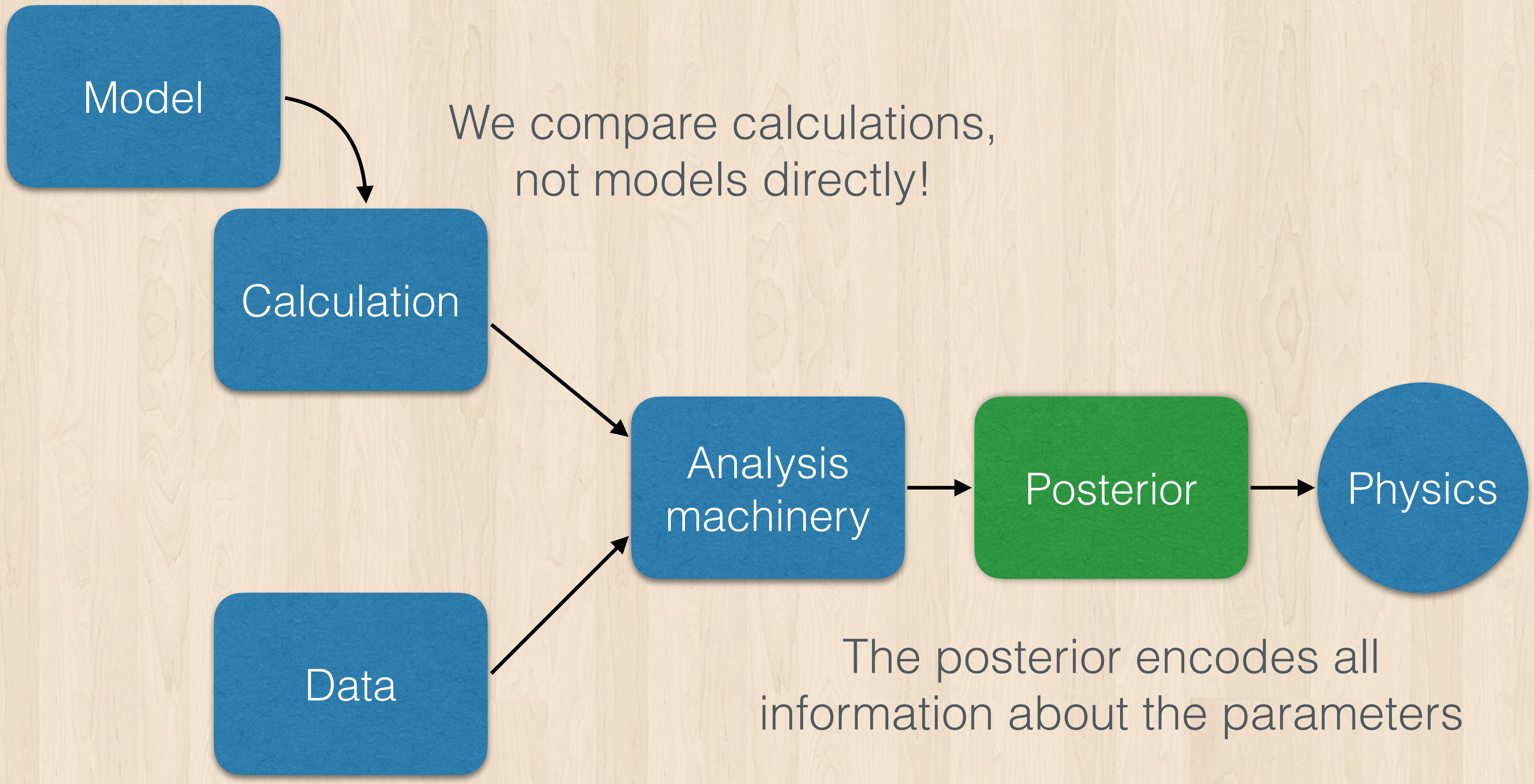
Backup Slides Ahead

Analysis in a nutshell

The 2000 ft. view



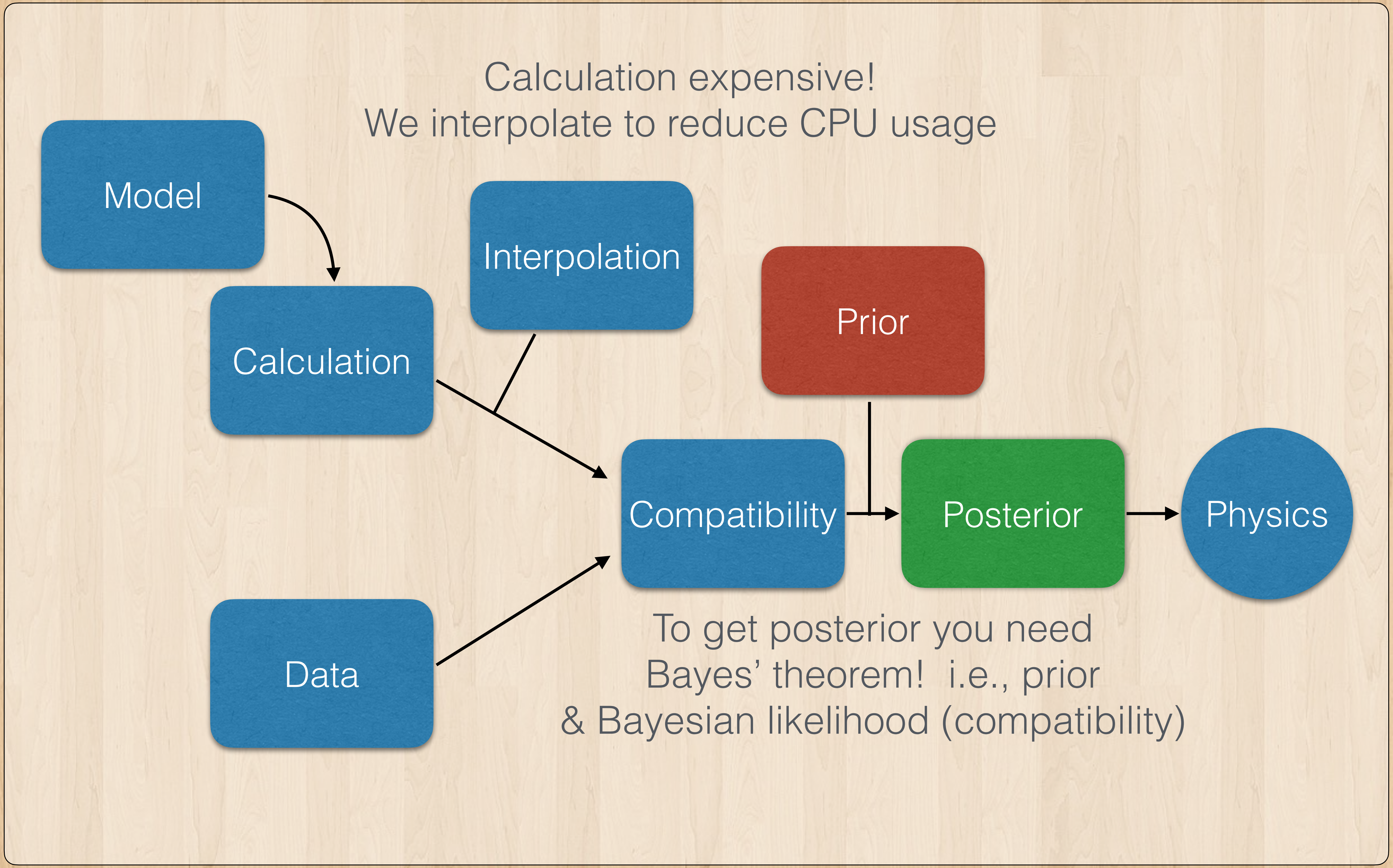
The 200 ft. view



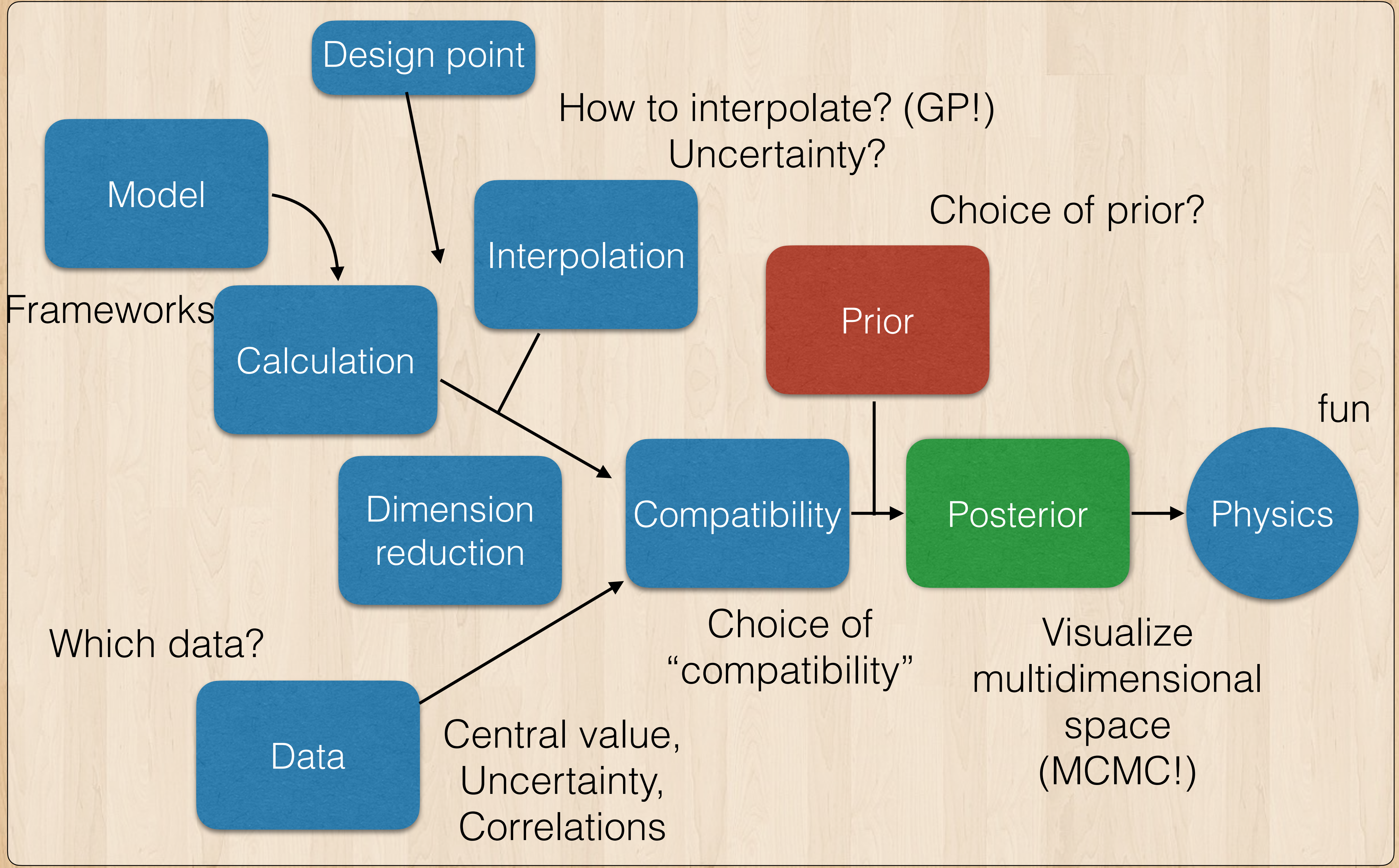
We compare calculations, not models directly!

The posterior encodes all information about the parameters

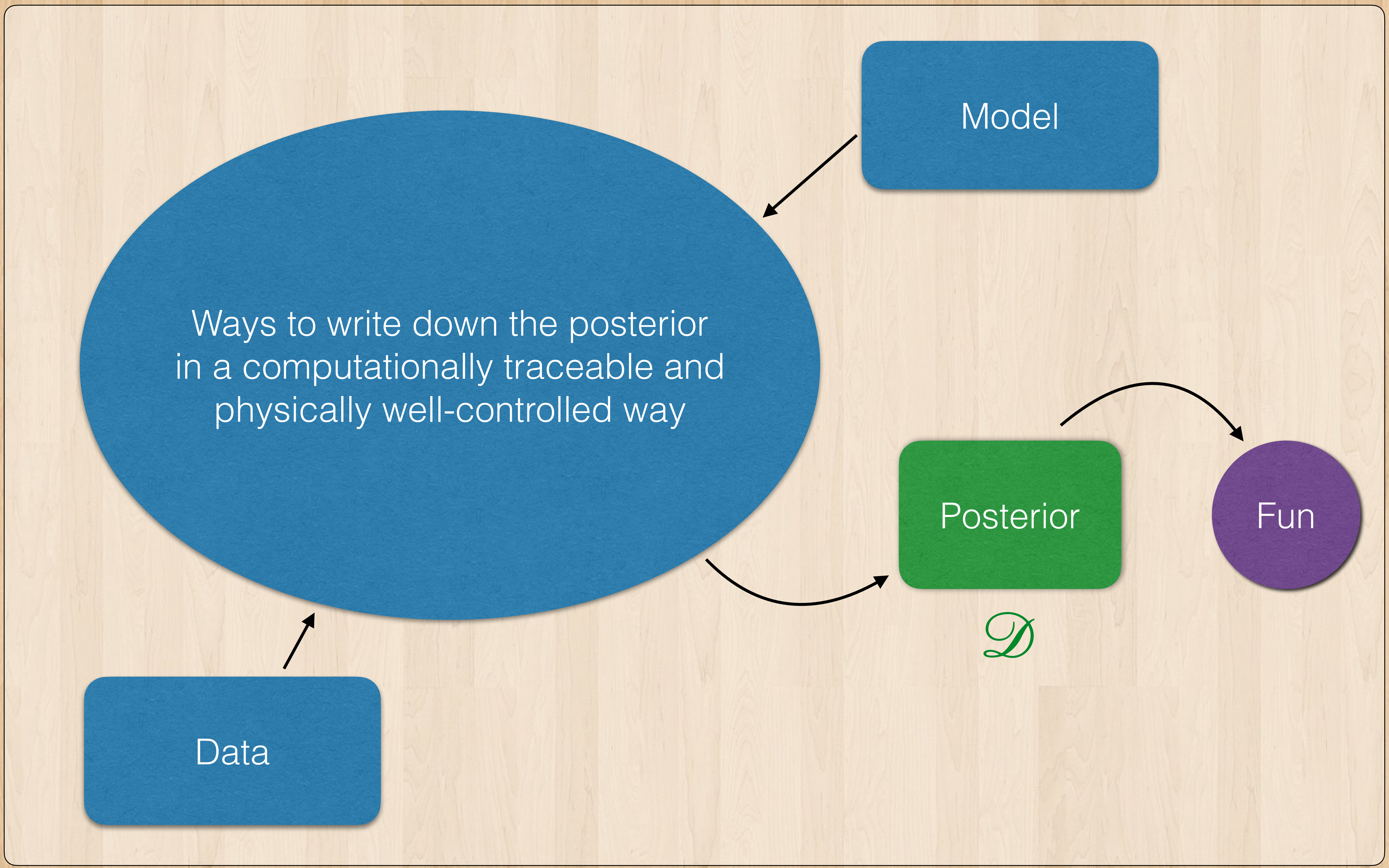
The 20 ft. view



The 14 ft. view



Zoom back out

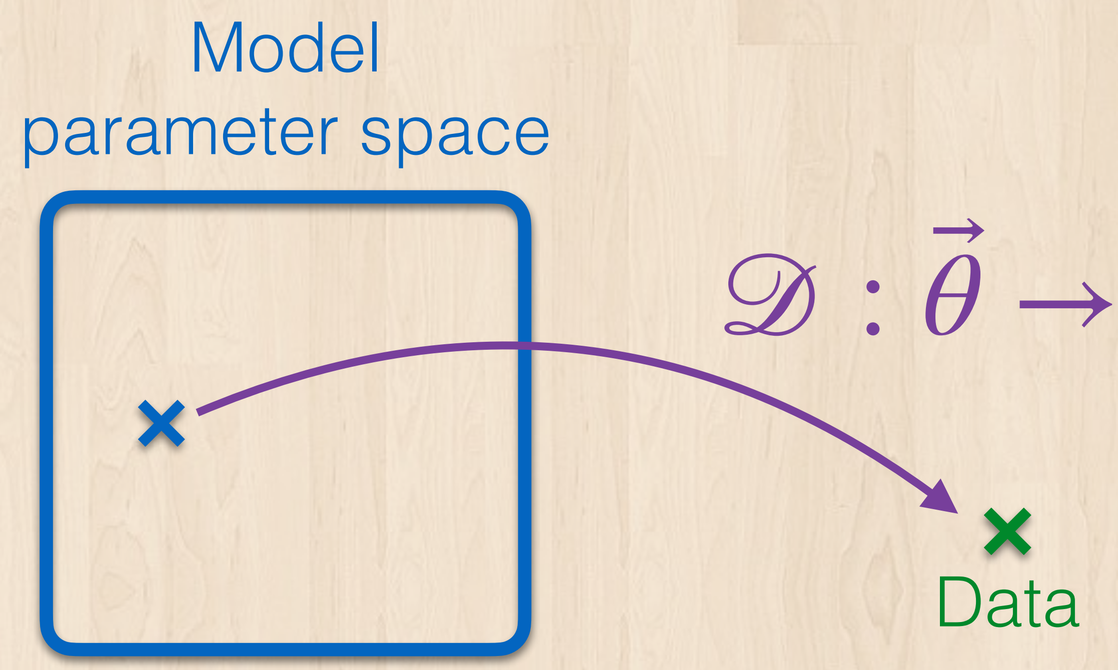


Bayesian analysis: operational view

Bayesian analysis: operational view

Function $\mathscr{D}$ maps parameter point to a "distance" to the data

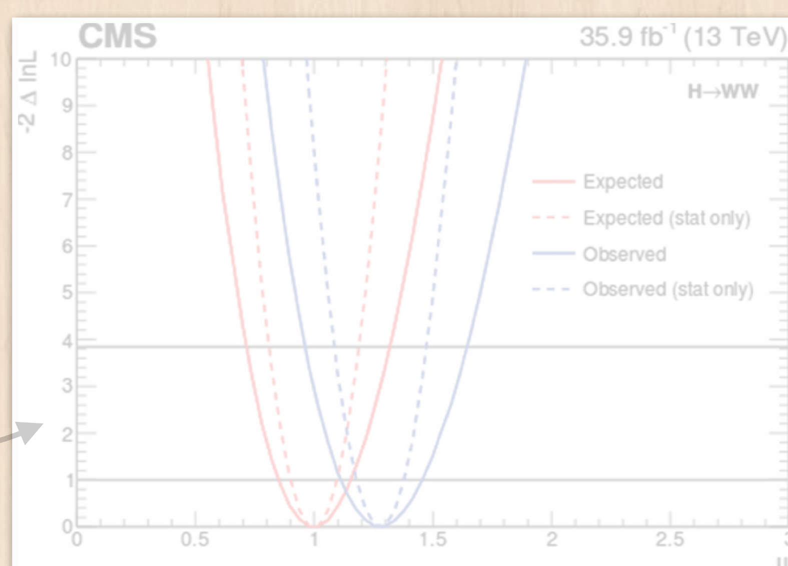
Contains all physics we want to extract



$$\mathscr{D} : \vec{\theta} \rightarrow \mathbb{R}$$

$\mathscr{D}$ = the posterior function in the Bayes' formalism

Similar in spirit

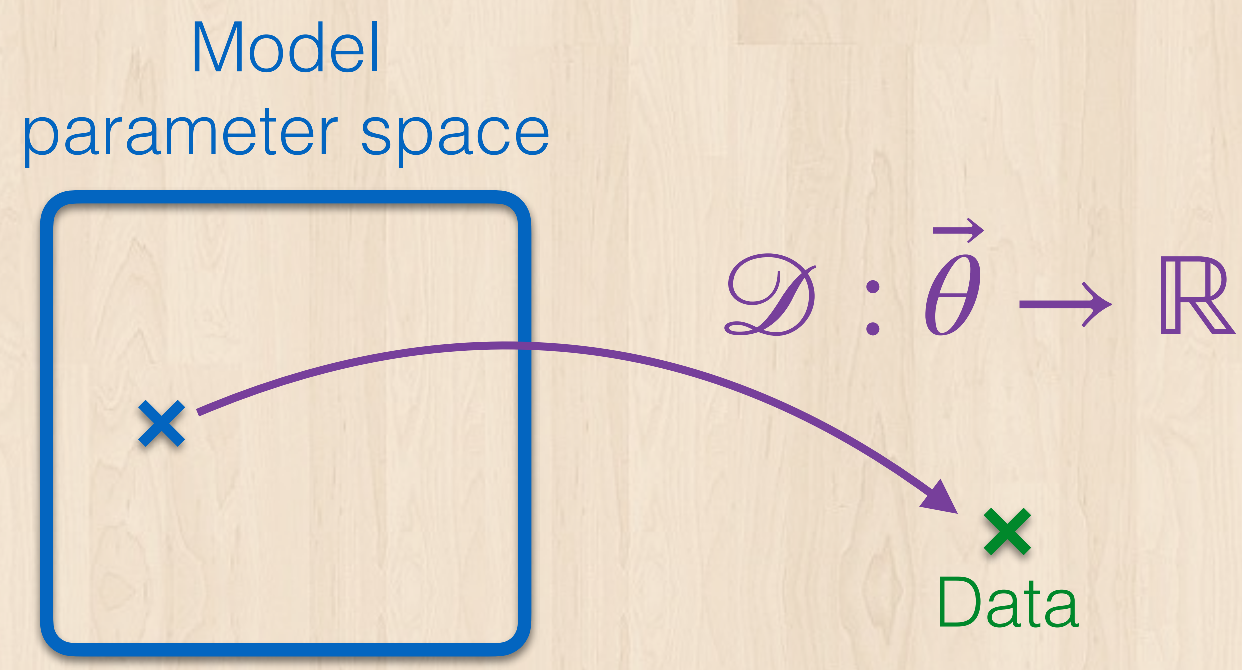


Bayesian analysis provides a way to get to $\mathscr{D}$ efficiently

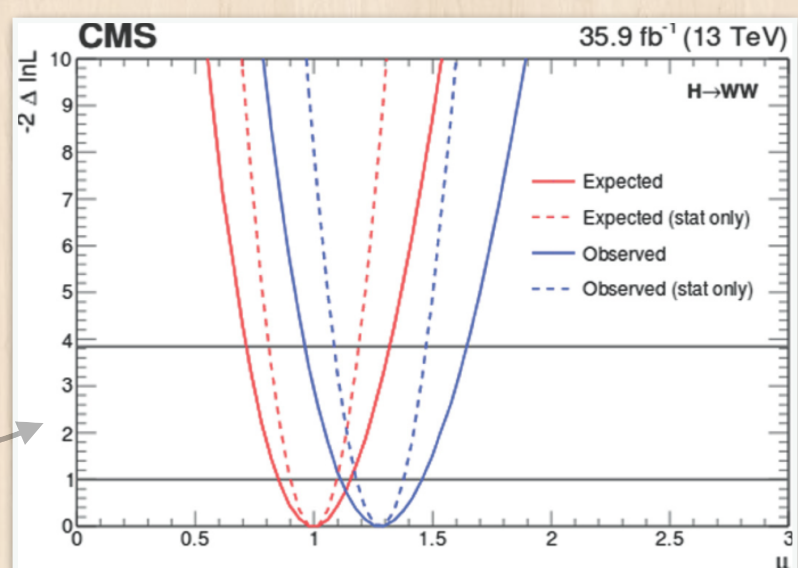
Bayesian analysis: operational view

Function $\mathscr{D}$ maps parameter point to a “distance” to the data

Contains all physics we want to extract



Similar in spirit



$\mathscr{D}$ = the posterior function in the Bayes’ formalism

Bayesian analysis provides a way to get to $\mathscr{D}$ efficiently

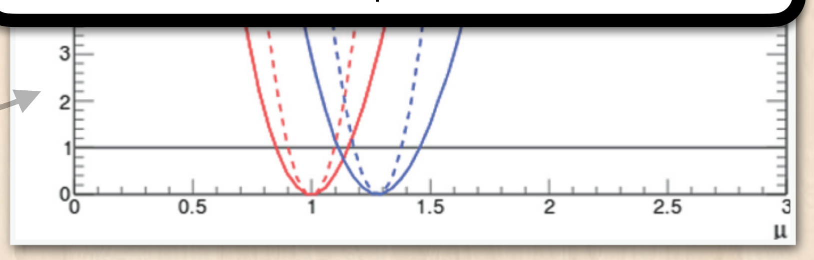
Bayesian analysis: operational view

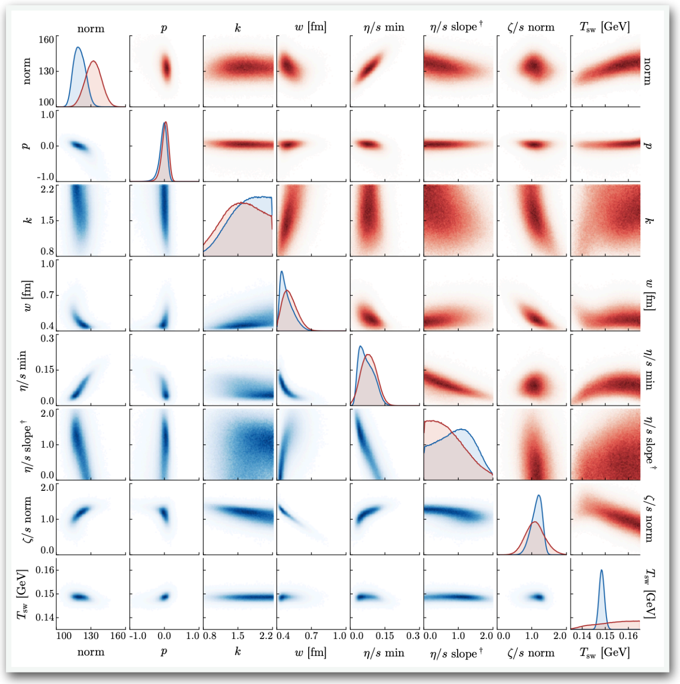
Function $\mathscr{D}$ maps parameter point to a "distance" to the data

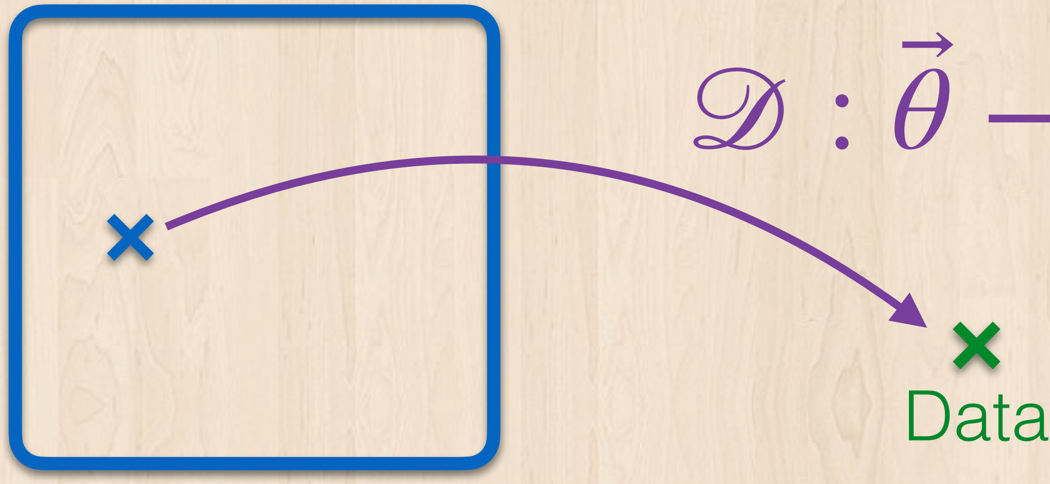
Con

we

Similar in spirit




Example for illustrations



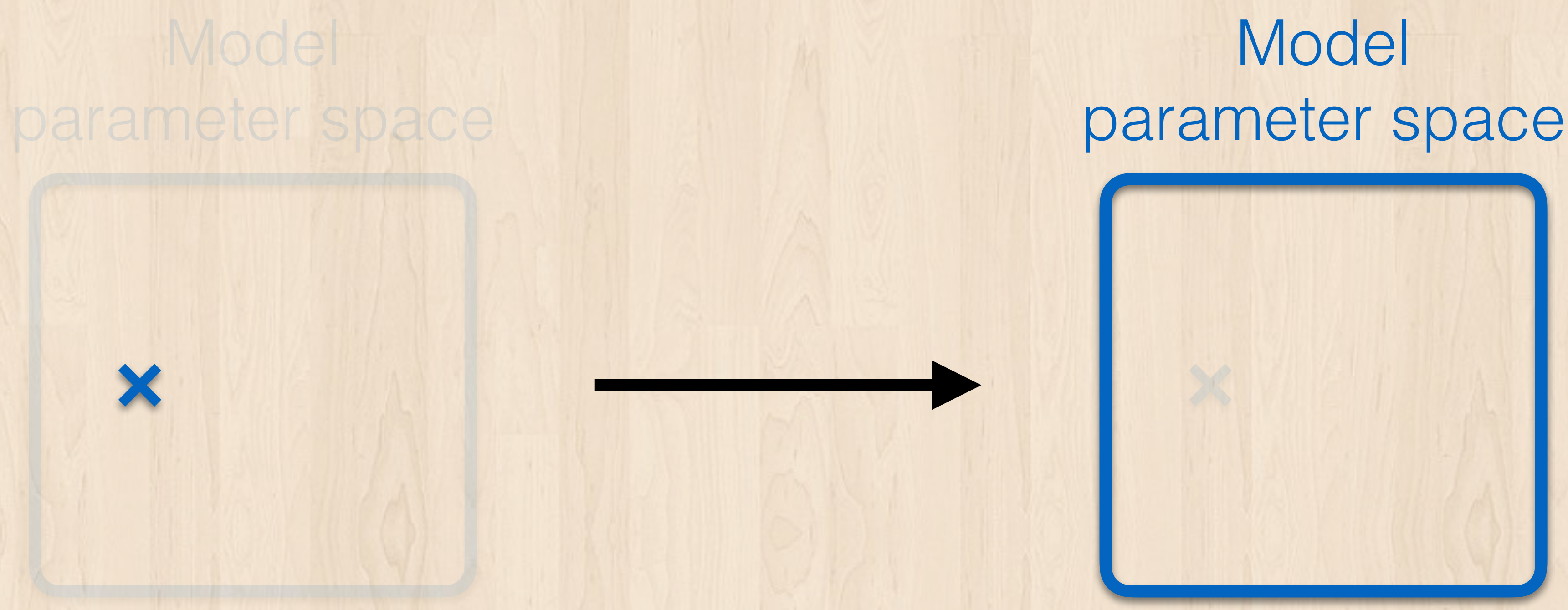
$$\mathscr{D} : \vec{\theta} \to \mathbb{R}$$

$\mathscr{D}$ = the posterior function in the Bayes' formalism

Bayesian analysis provides a way to get to $\mathscr{D}$ efficiently

PRC 94, 024907 (2016)

Conceptual shift



Instead of single parameter, we analyze the model parameter space as a whole

Chance to test models instead of parameters

ideas

Recent developments

Examples

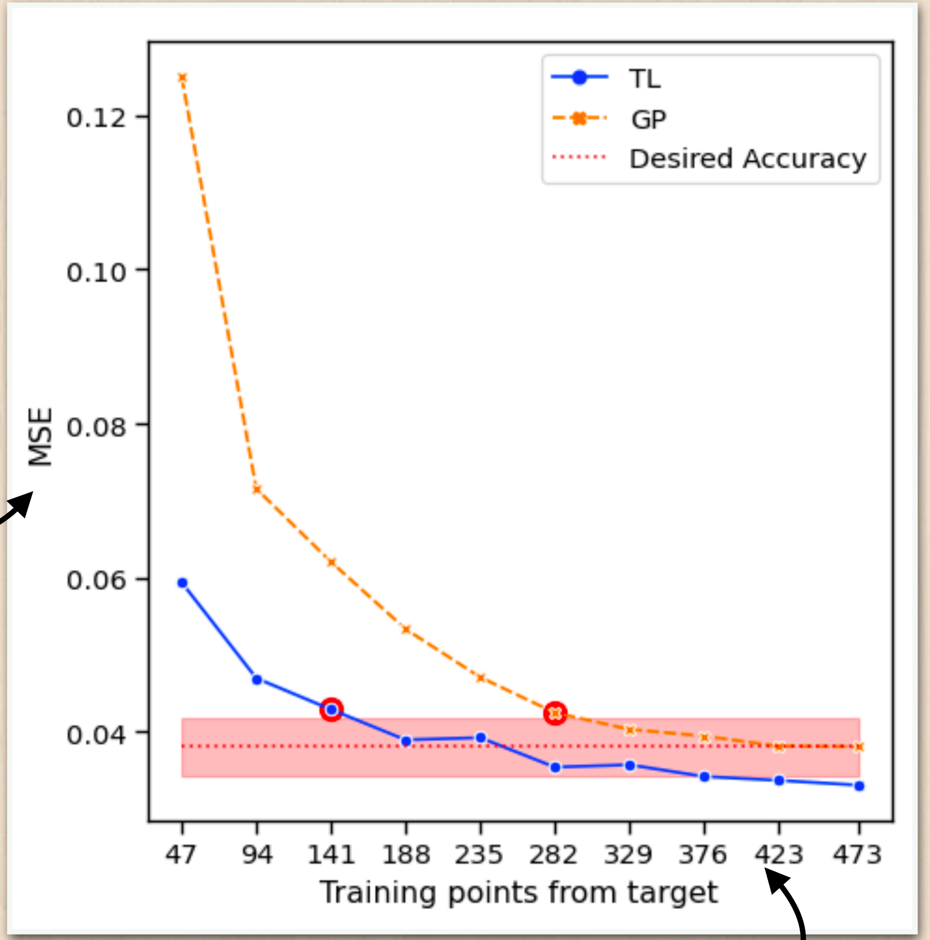
Transfer learning

In addition to the nominal analysis, many developments in the analysis side as well

Transfer analysis "knowledge" across similar tasks

Case study: transfer from 2.76 TeV Pb+Pb to 200 GeV Au+Au

Precision



Amount of computing needed

PRC 105, 034910 (2022)

Multi-fidelity approach

Model 1

Cheap to run
Does not capture full physics

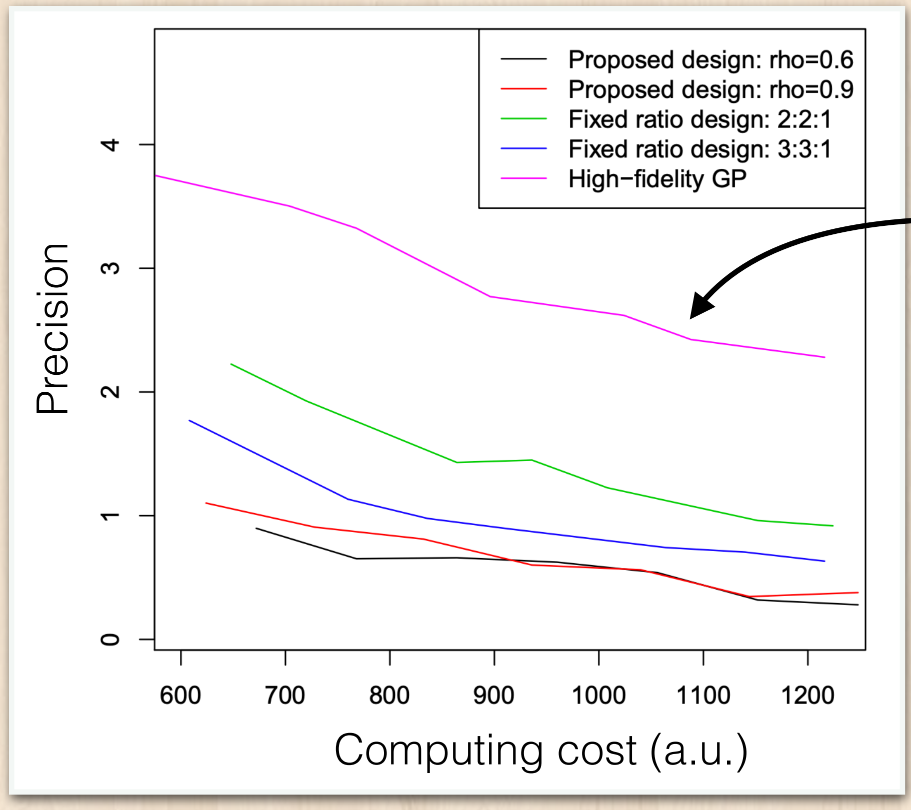
e.g. LO

Model 2

Expensive to run
Precise

e.g. NLO, or even NNLO

Strategy: use model 1 to learn the "big structure" and model 2 to refine



Use only model 2

Multi-fidelity approach

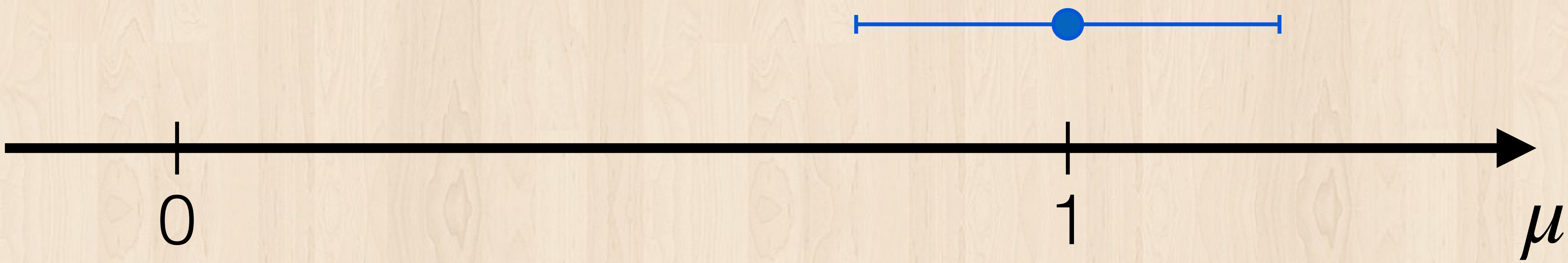
Reduces CPU cost needed to achieve same level of precision

arXiv: 2108.00306

Uncertainty: tails

What about the tails?

Compatibility of 1 ± 0.25 to 0?

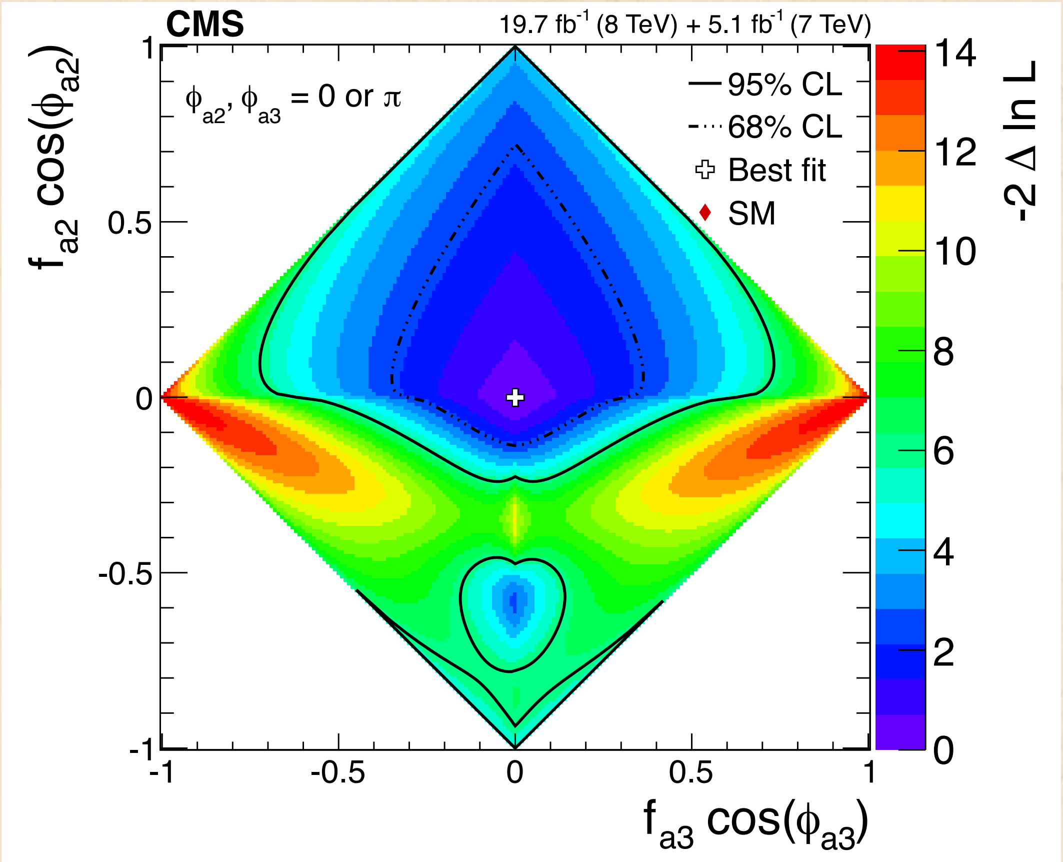


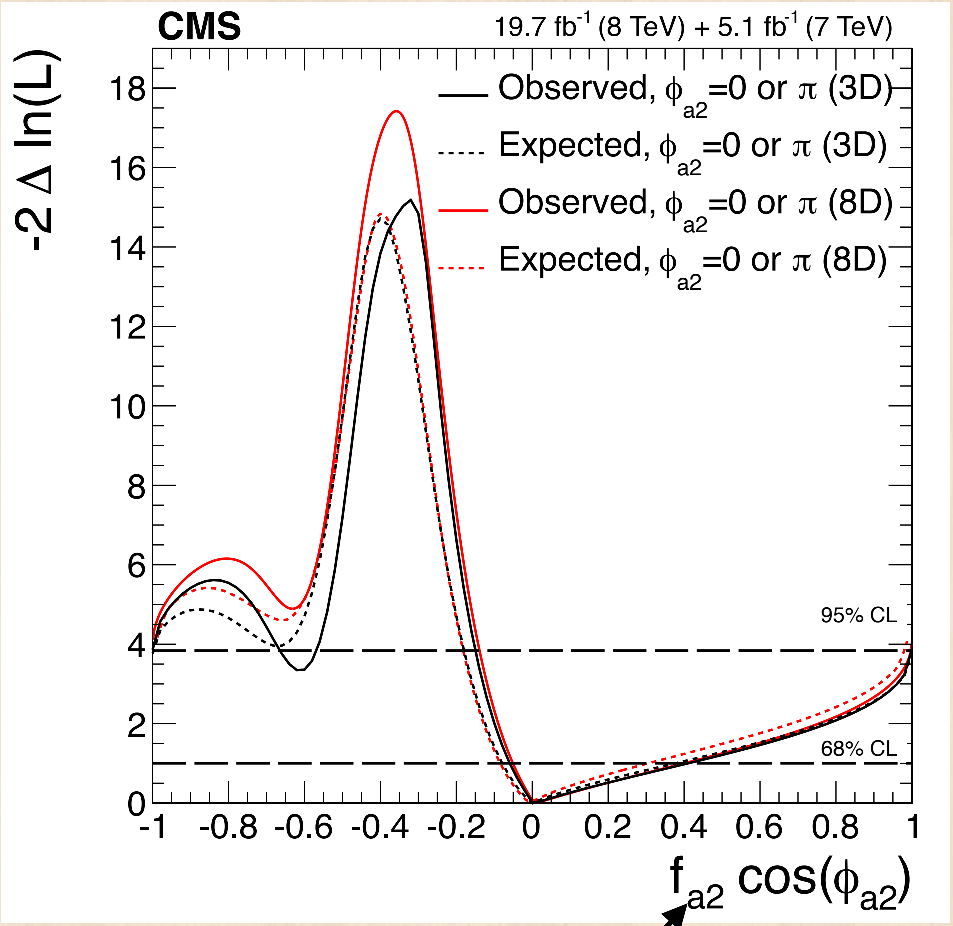
We don't know! There is not enough information

Especially important for small-error measurements
(For example flow, etc)

Example of nontrivial tail

$$\mathscr{L} \sim c\left(HZ_\mu Z^\mu + a_2 HZ_{\mu\nu}Z^{\mu\nu} + a_3 HZ_{\mu\nu}\tilde{Z}^{\mu\nu}\right) + \ldots$$





Size of CP-odd HZZ term

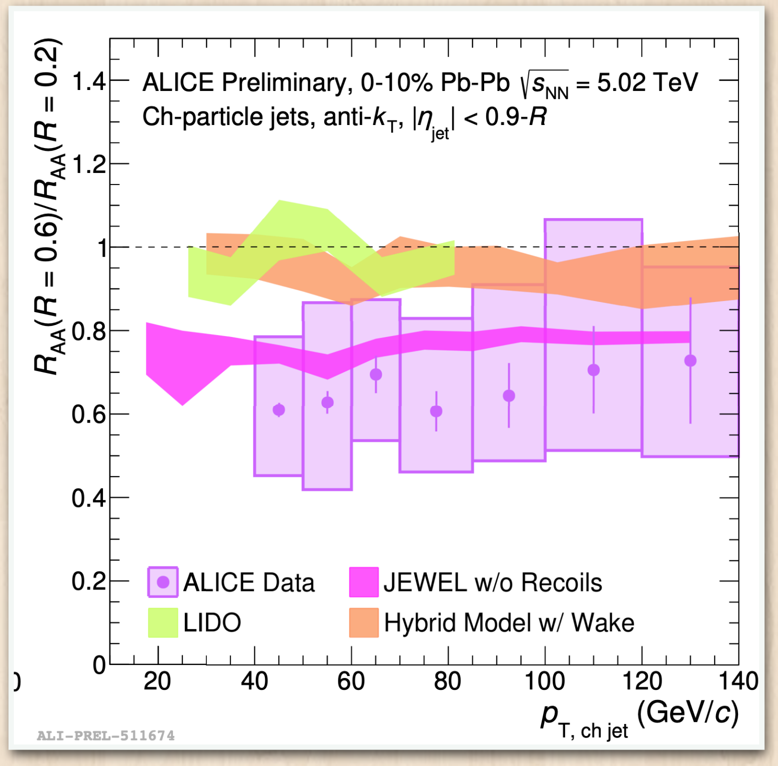
Size of higher order CP-even HZZ term

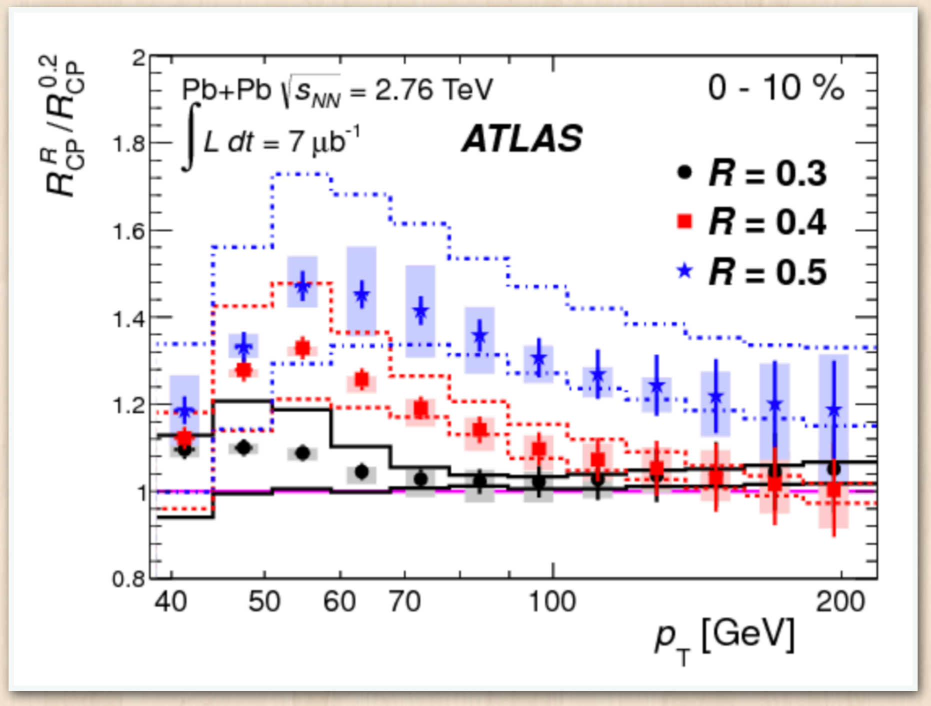
Guesses and missed opportunites

- Missing information needs to be specified as **guesses**
- Guesses need to be checked and varied!
  - **Extra uncertainties** in the extracted results
- Food for thought for experiments: how much information to provide? (or, how much time to invest in this?)
  - Otherwise a lot of **missed opportunities**

Other miscellaneous things

Data choice





Important to pick a scope and include ALL eligible data

*unless there are known issues (ps. tension doesn't count)

High chance of bias if only a subset is used

Generators

- What Bayesian analysis does is to find the region of phase space matching the best to the data/truth
- If generator does not have required physics it’s easy to misinterpret the result
  - Case for better vacuum shower modeling (for example)
- Ratios help but not everything is multiplicative

Example new observable

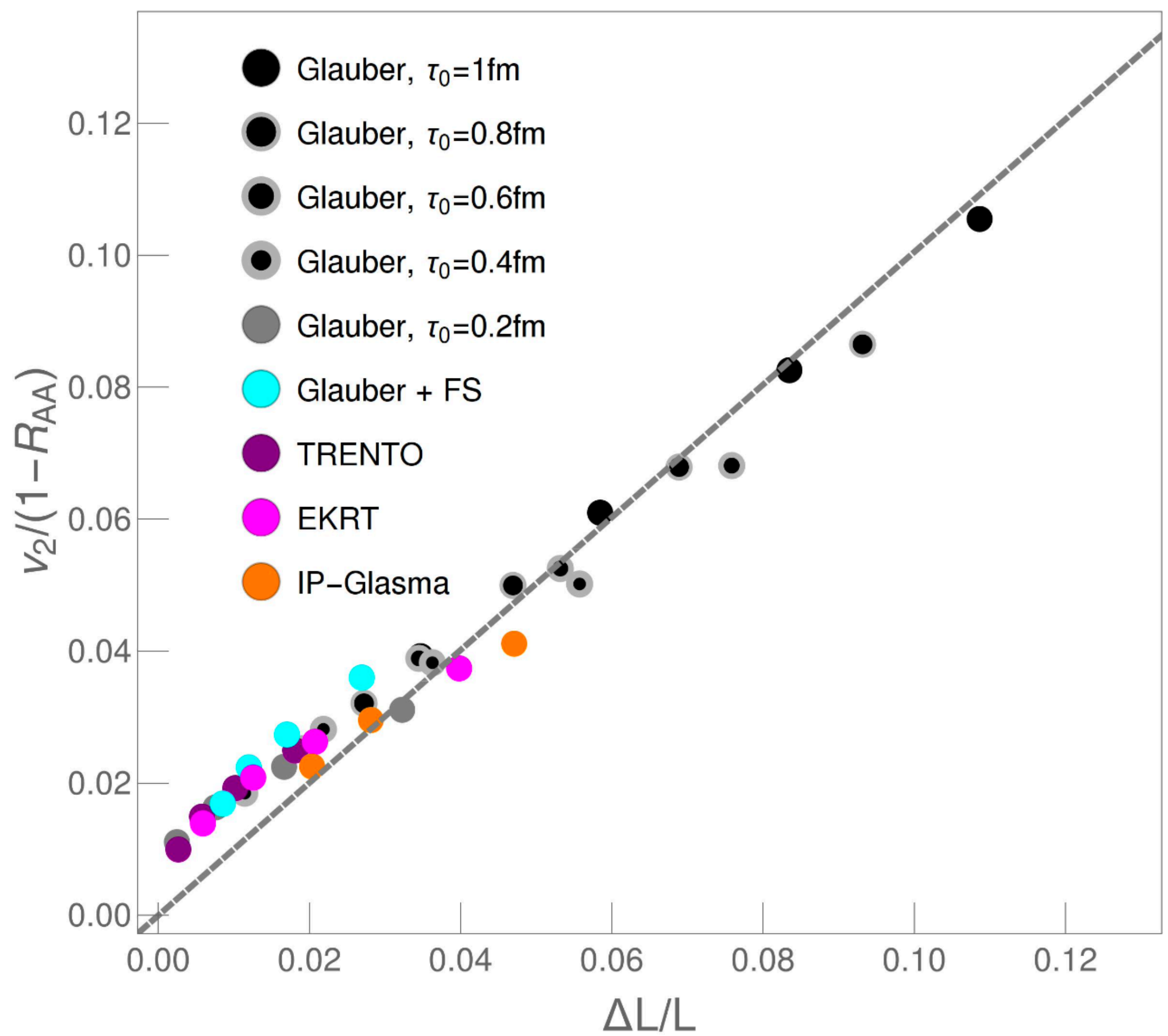


FIG. 4: Charged hadron $v_2/(1-R_{AA})$ as a function of path-length anisotropies $\Delta L/L$, for various centrality classes and temperature profiles. The value of transverse momentum is fixed at $p_\perp = 100$ GeV. The linear fit yields a slope of approximately 1.

Trento p

$$\tilde{T}_R = \begin{cases} \max(\tilde{T}_A, \tilde{T}_B) & p \to +\infty, \\ (\tilde{T}_A + \tilde{T}_B)/2 & p = +1, \quad \text{(arithmetic)} \\ \sqrt{\tilde{T}_A \tilde{T}_B} & p = 0, \quad \text{(geometric)} \\ 2\,\tilde{T}_A \tilde{T}_B / (\tilde{T}_A + \tilde{T}_B) & p = -1, \quad \text{(harmonic)} \\ \min(\tilde{T}_A, \tilde{T}_B) & p \to -\infty. \end{cases}$$

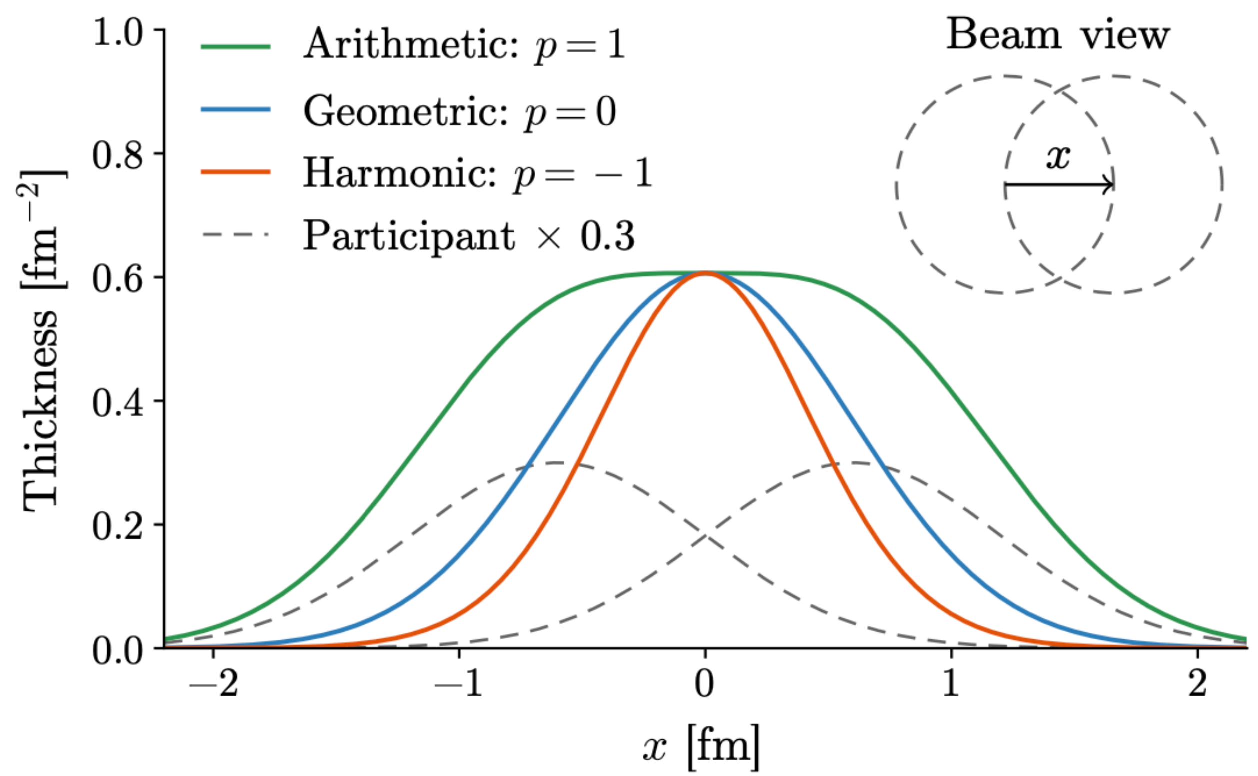


Figure 3.1 Reduced thickness of a pair of nucleon participants. The nucleons collide with a nonzero impact parameter along the x-direction as shown in the upper right. The gray dashed lines are one-dimensional cross sections of the participant nucleon thickness functions $\tilde{T}_A, \tilde{T}_B$, and the colored lines are the reduced thickness $\tilde{T}_R$ for $p = 1, 0, -1$ (green, blue, orange).

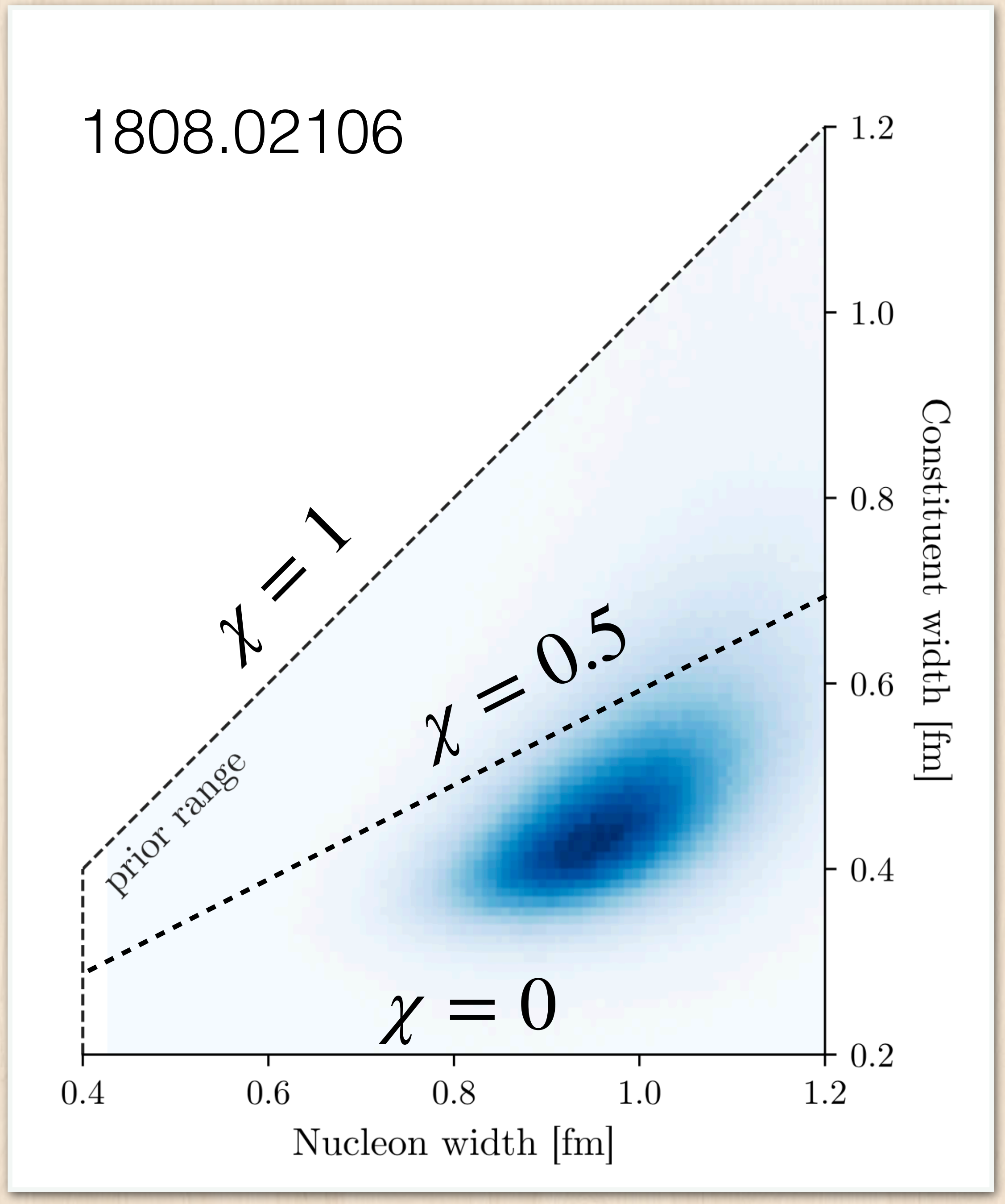
Modeling improvements

- 3D Trento initial condition
- The X-SCAPE project
- Improved modeling of nucleus/constituent radial profile (moving away from simple Gaussian)
- …

Inputs to Bayesian

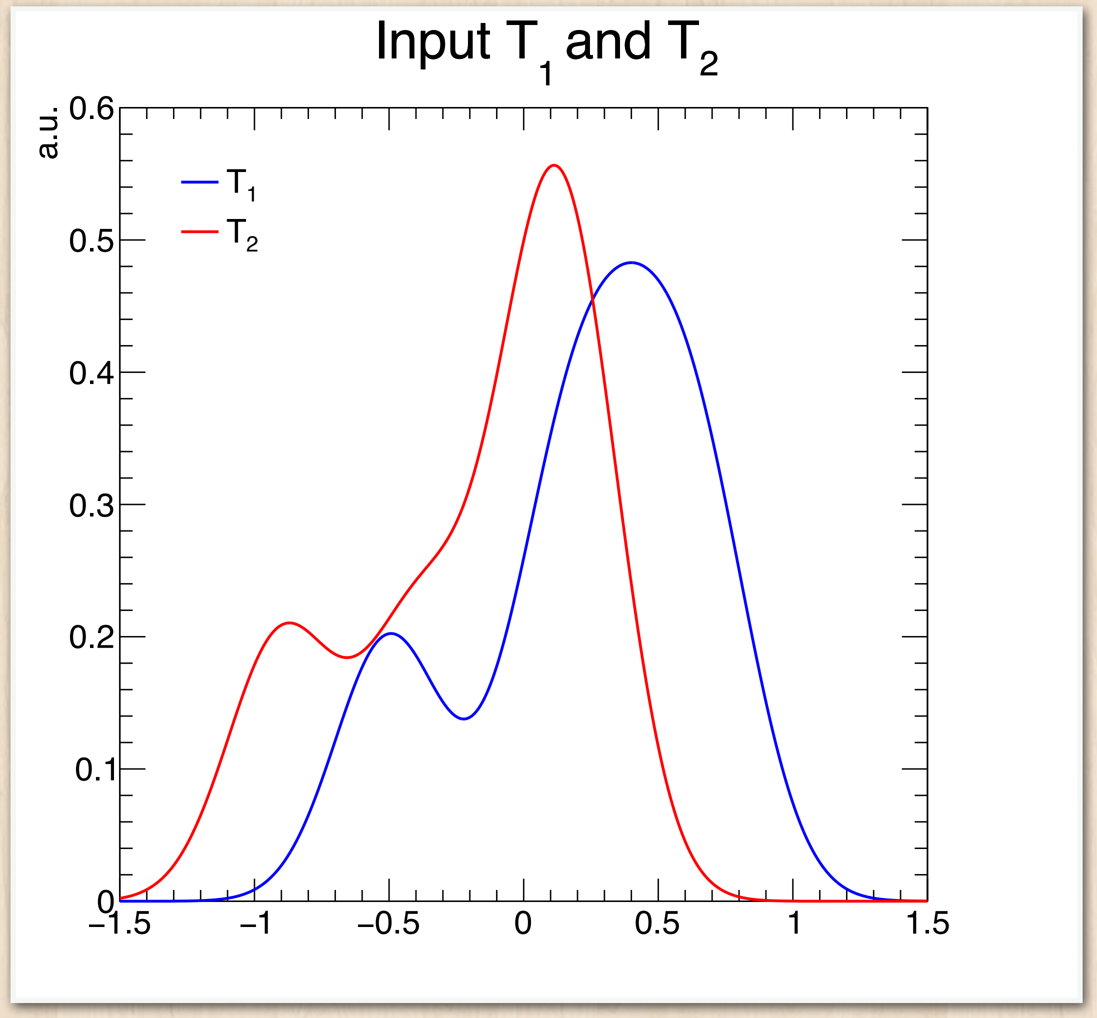
- Bayesian analysis is useful for uncovering complex correlations between different parameters and measurement — but —
  - Equally important, we should <u>also</u> include "pure" observables that are sensitive to only small amount of parameters
  - As well as less model-dependent observables
- Then we design more observables to feed back into the loop

Nuclear size vs substructure

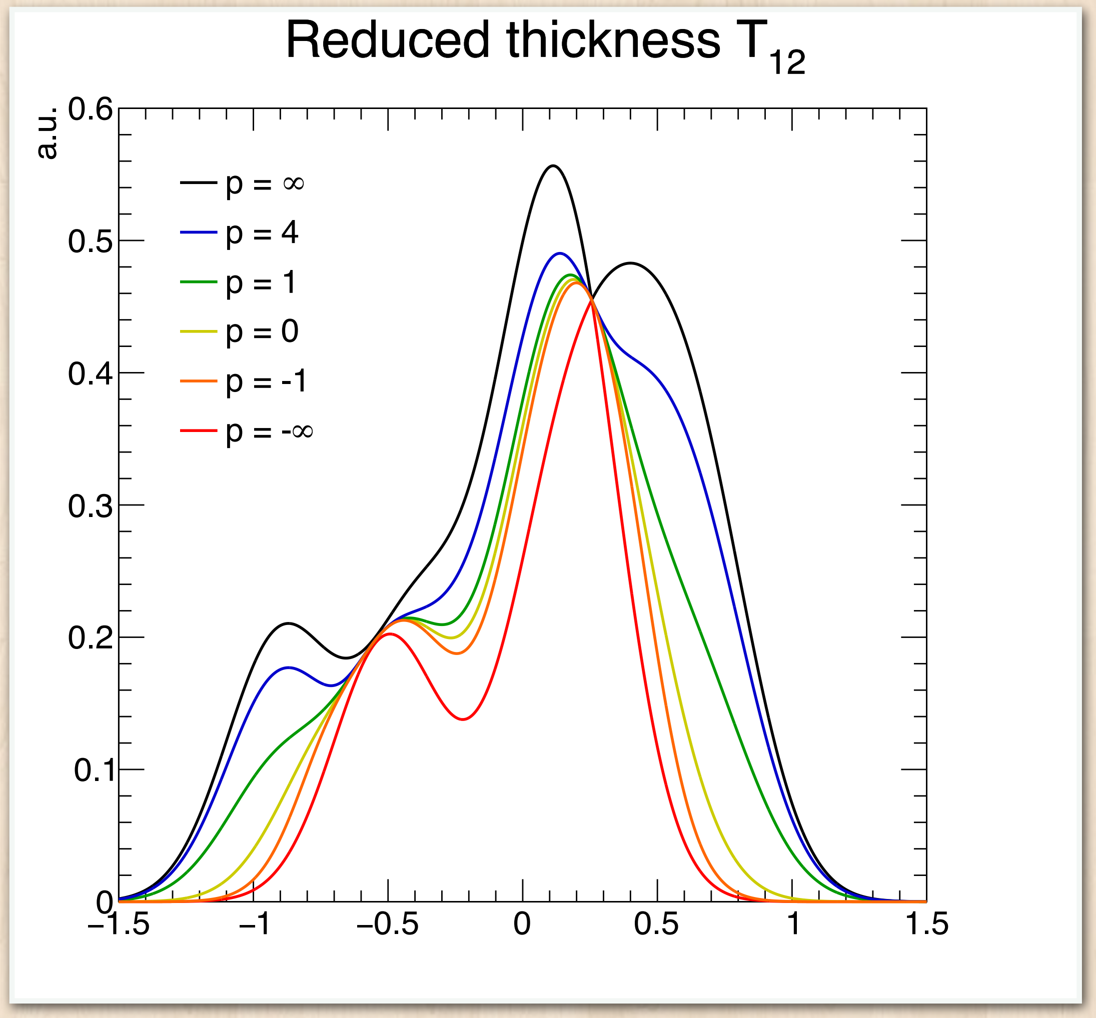


The Trento Ansatz: example

Toy example



Lower p ~sharper

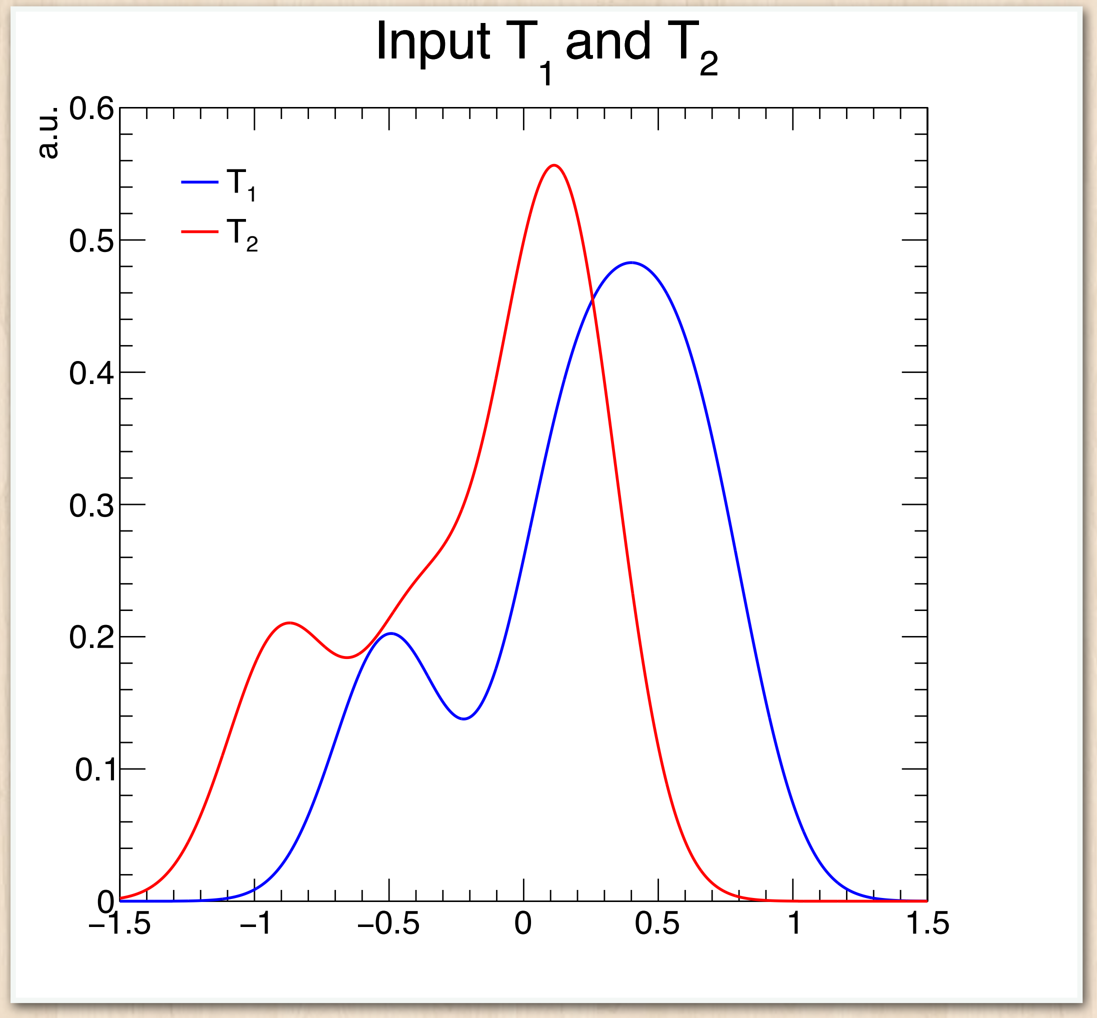


$\min(T_1, T_2)$ $\sqrt{T_1 T_2}$ $(T_1 + T_2)/2$ $\max(T_1, T_2)$

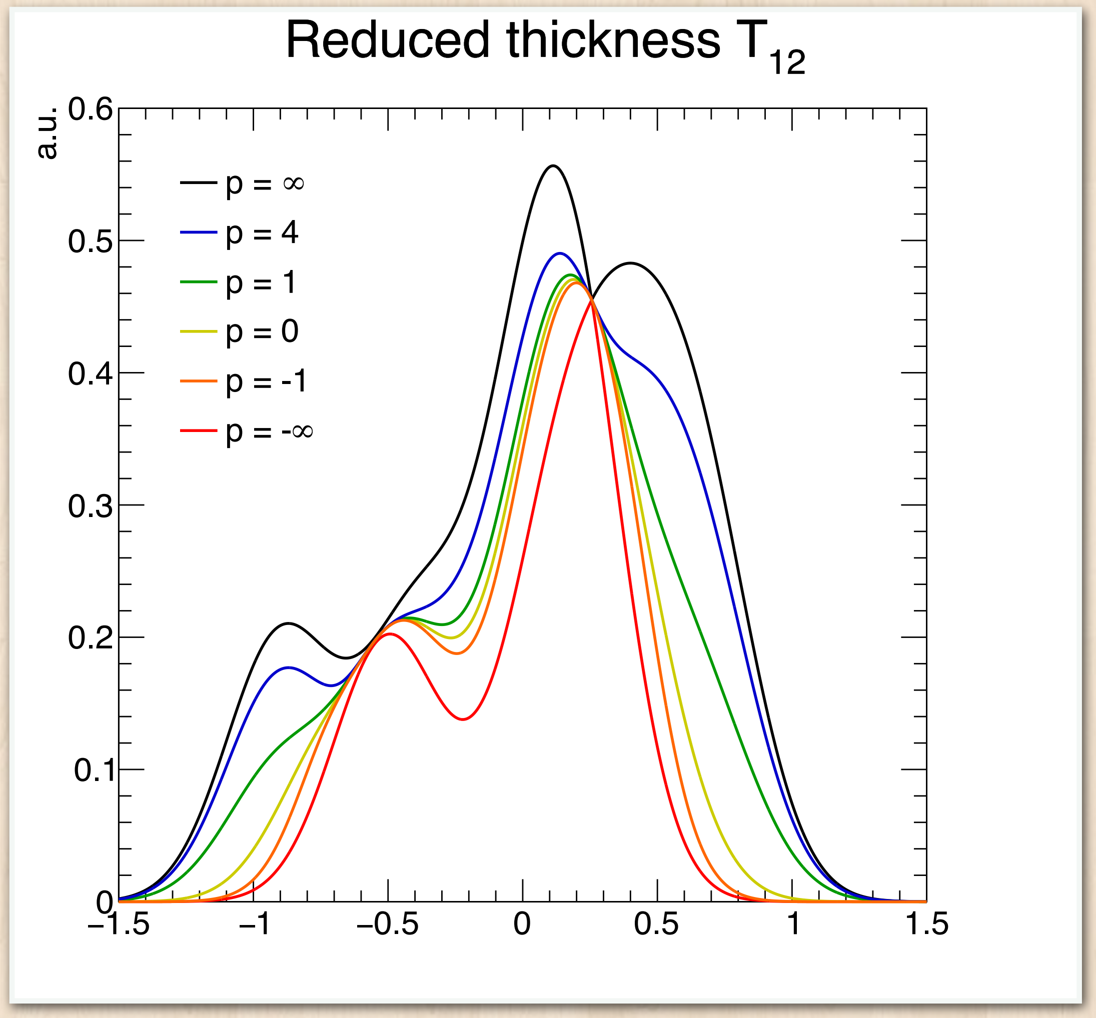
p: $-\infty$ → 0 → 1 → ∞

The Trento Ansatz: example

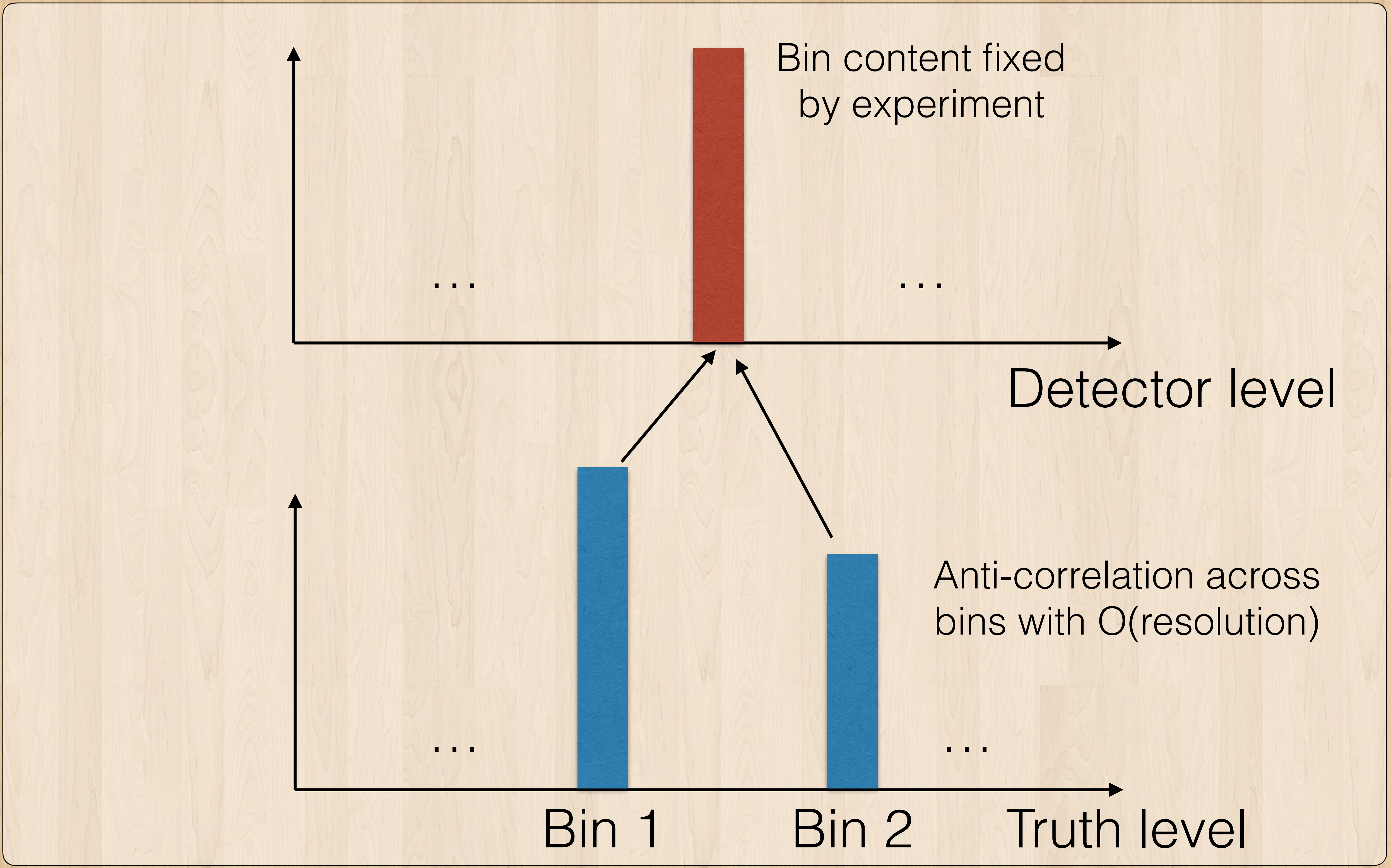
Toy example



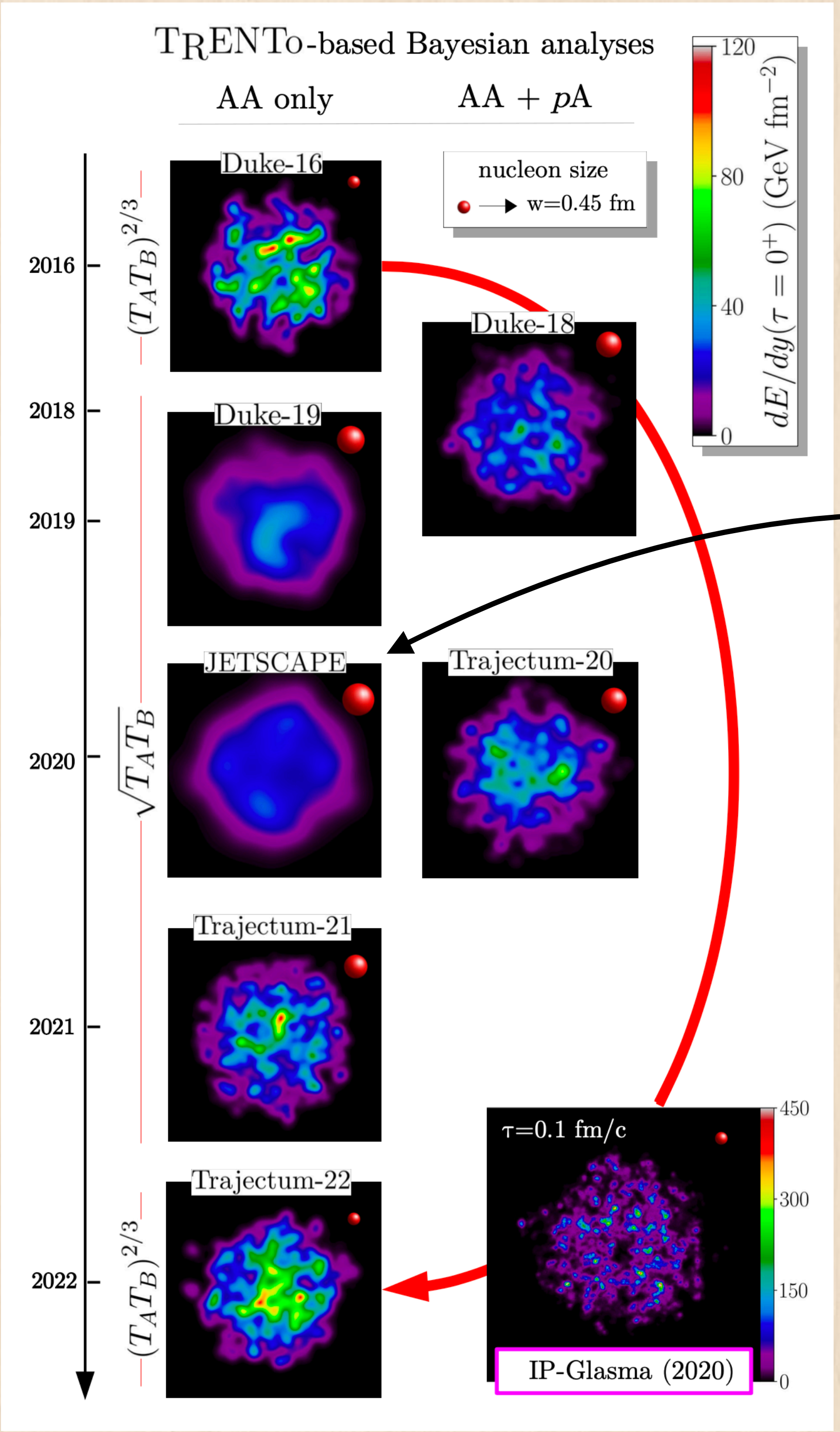
Larger q ~sharper



Feed-down vs feed-up



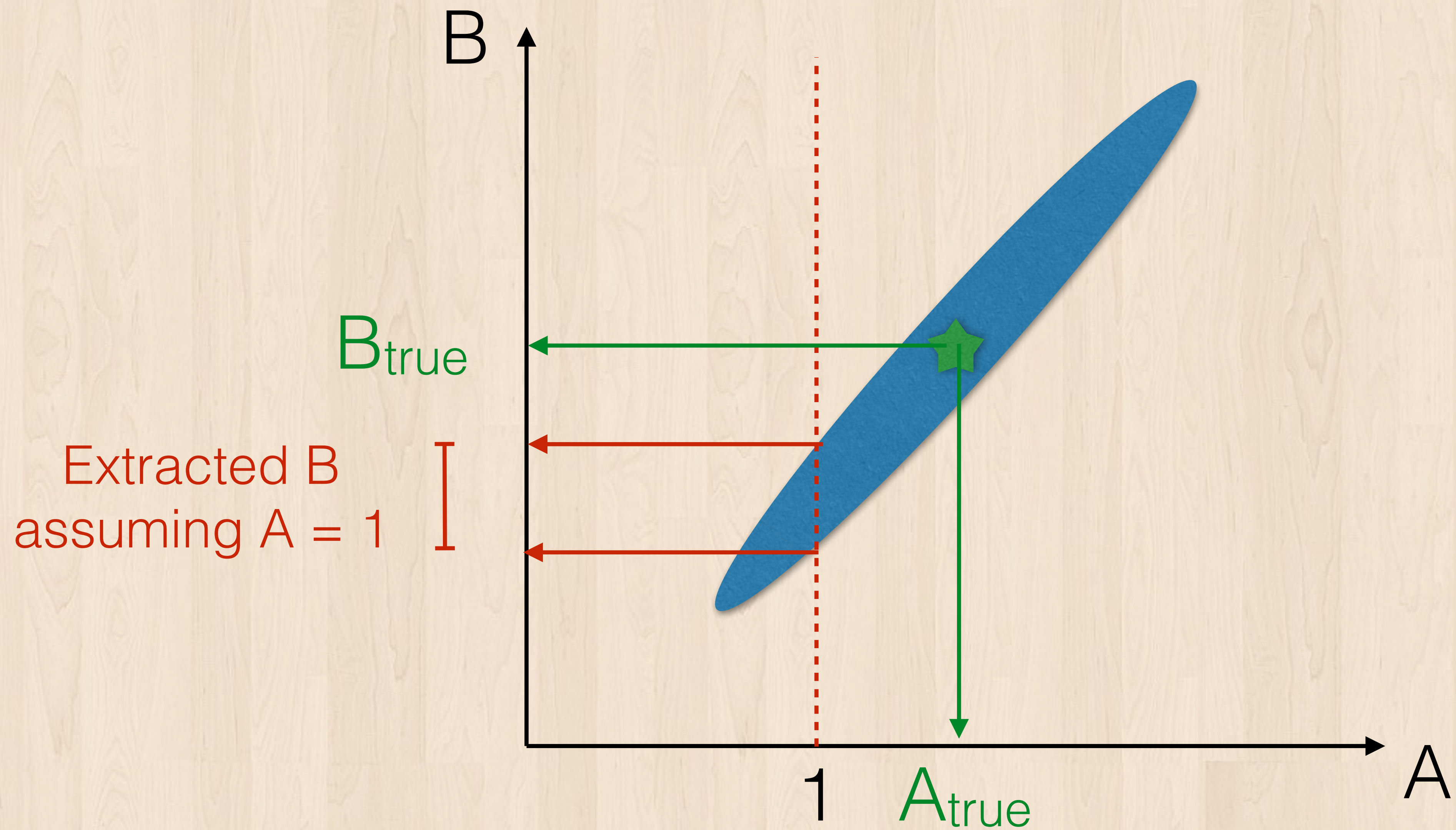
Nice illustration from G. Giacalone



$w \sim 1$ & no substructure

2208.06839

Correlation + model assumption



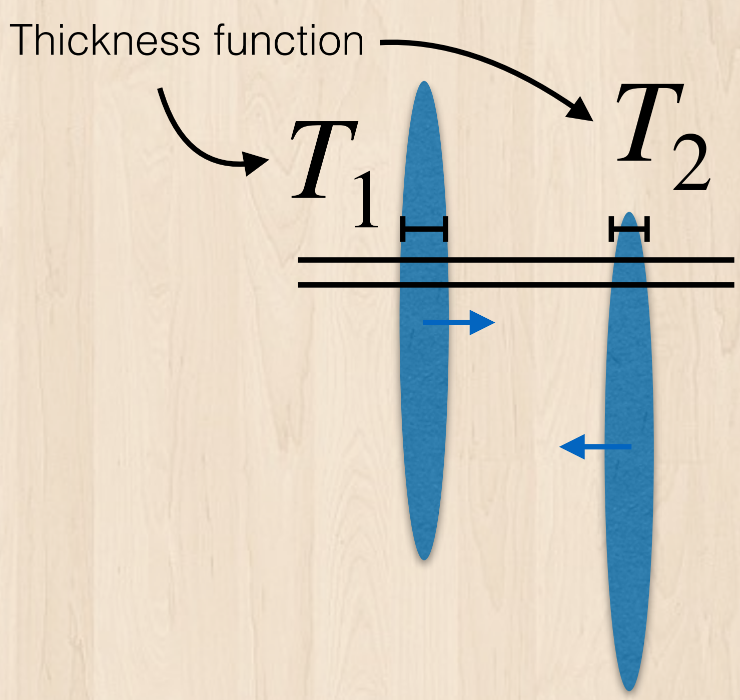
The Trento Ansatz

First sample to determine if nucleons collide

$$prob. = 1 - \exp\left(-\sigma_{gg}\int \rho_1(\vec{x})\rho_2(\vec{x})d\vec{x}\right)$$

If so, the nucleon adds to the nucleus' thickness function

(with a gamma-distributed random weight for extra fluctuation)



Trento Ansatz

$$\text{entropy} \propto T_{12} = \left(\frac{T_1^p + T_2^p}{2}\right)^{1/p}$$

$\min(T_1, T_2)$	$\sqrt{T_1 T_2}$	$(T_1 + T_2)/2$	$\max(T_1, T_2)$
$-\infty$	0	1	∞

p

1412.4708

See also W. Ke talk last week