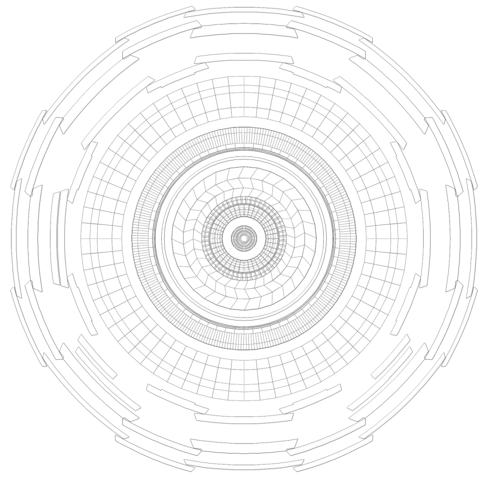
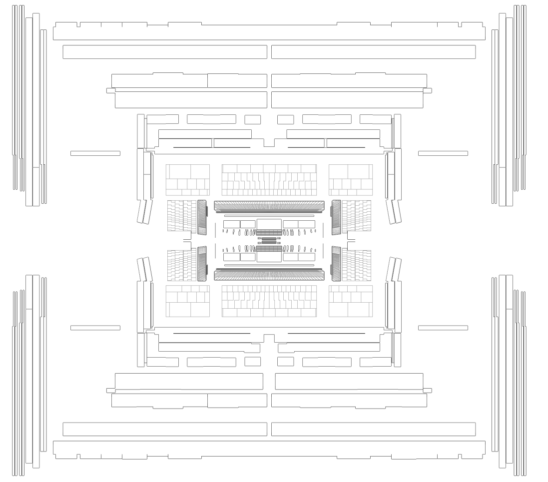
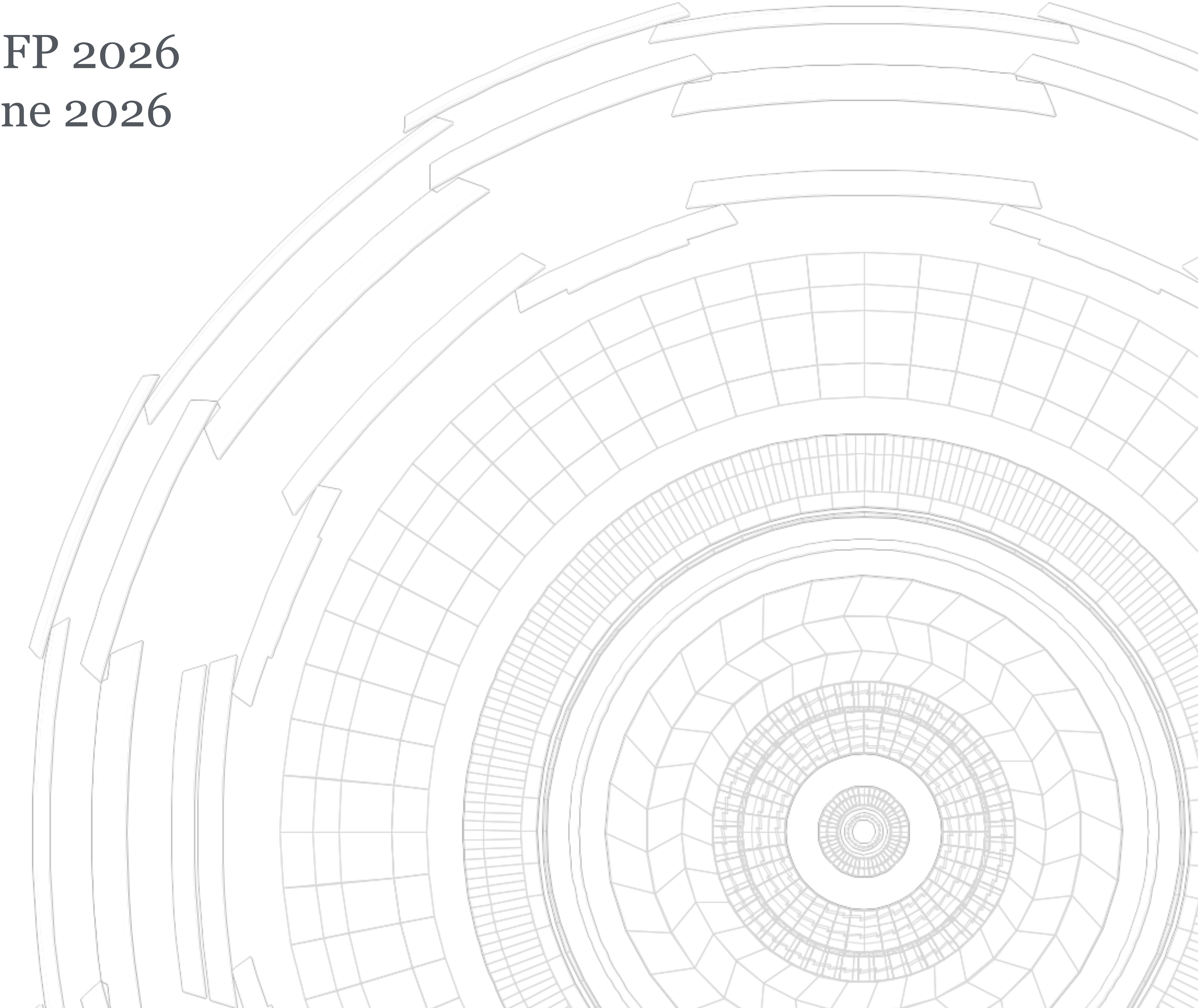


Collider Overview



Aishik Ghosh

ML4FP 2026
2 June 2026



Georgia Institute of Technology

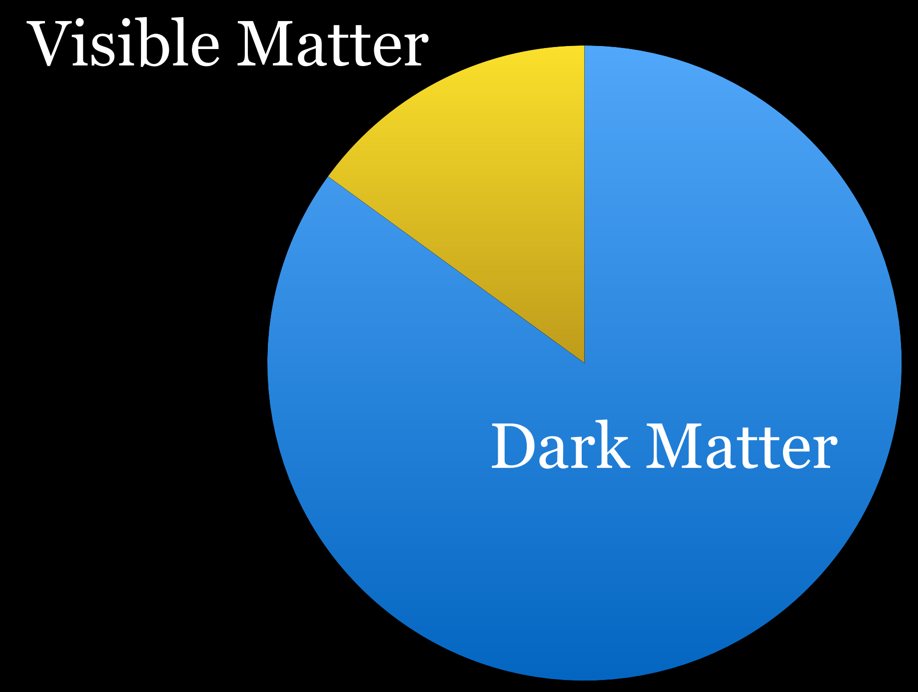
Prof-Aishik-Ghosh
aishikghosh.bsky.social
@Aishik_Ghosh_

Questions about the Universe ...

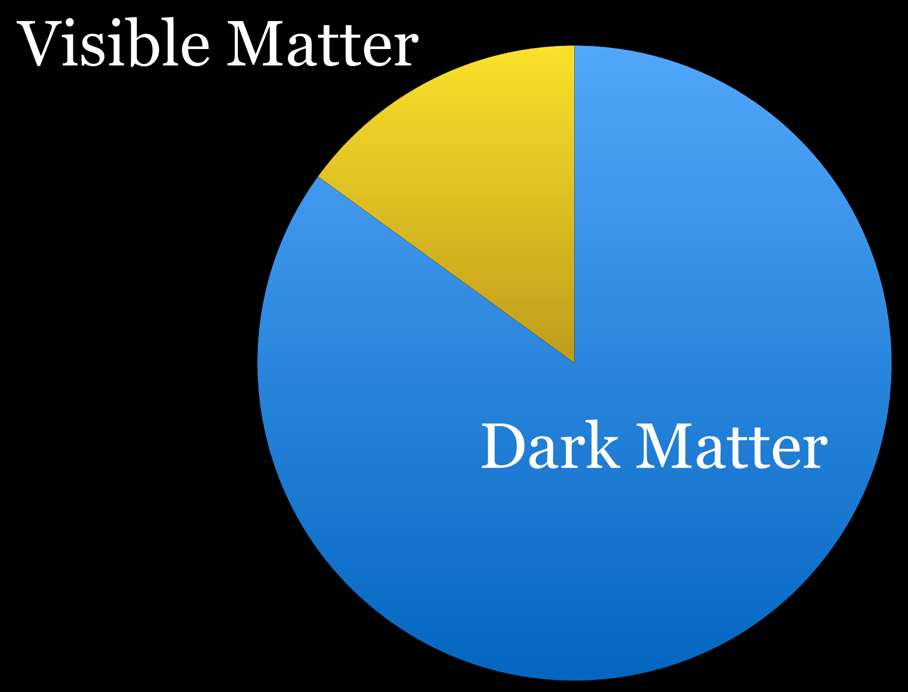
Questions about the Universe ...

How can we explain dark matter?

Why more matter than anti-matter in the Universe?



Questions about the Universe ...



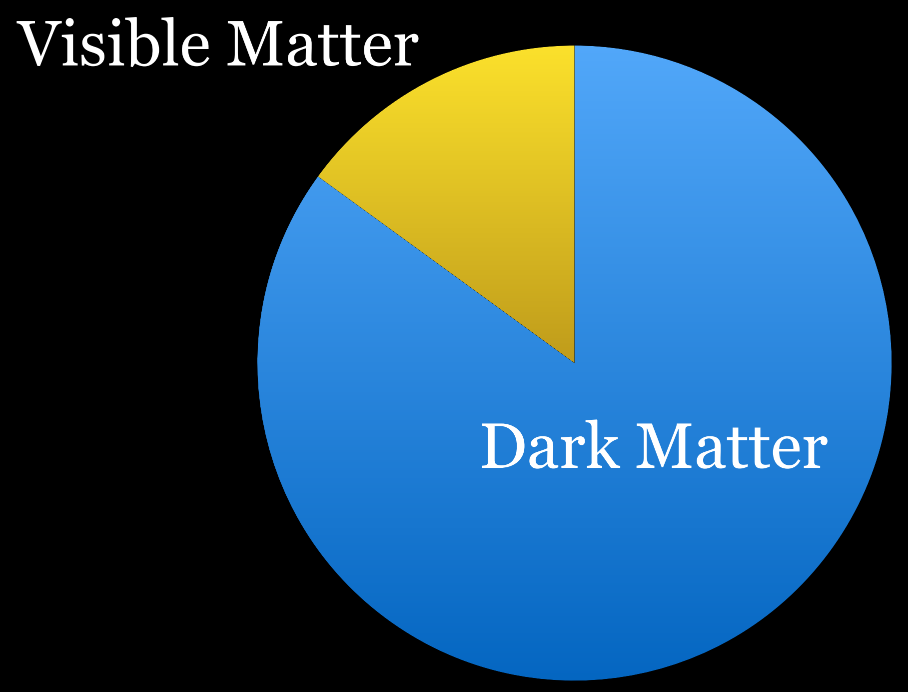
How can we explain dark matter?

Why more matter than anti-matter in the Universe?

Dark energy?

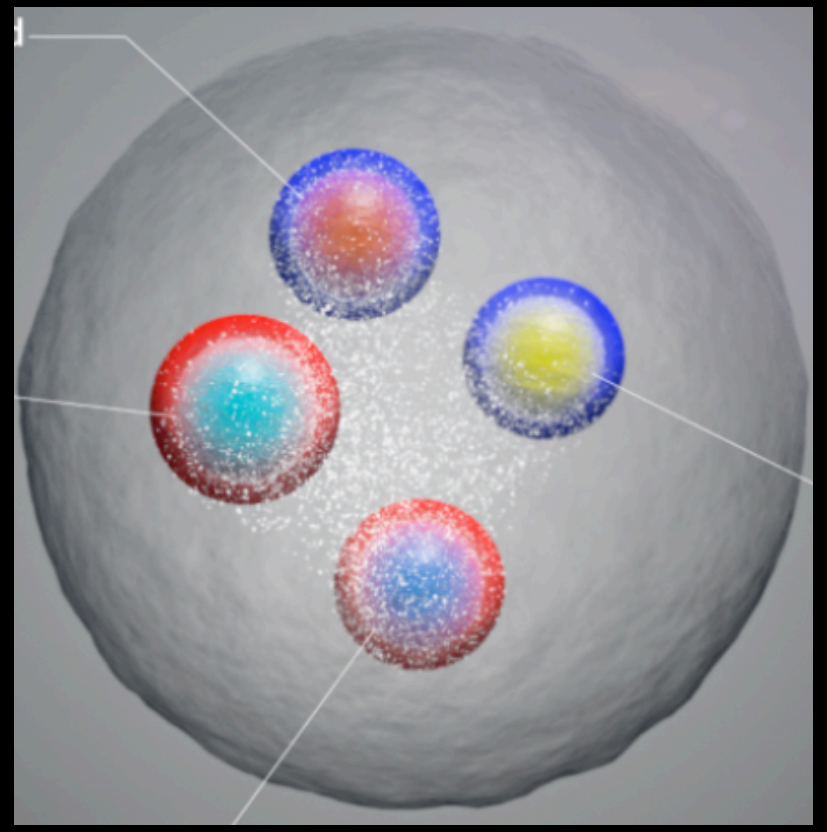
Questions about the Universe ...

How can we explain dark matter?



Why more matter than anti-matter in the Universe?

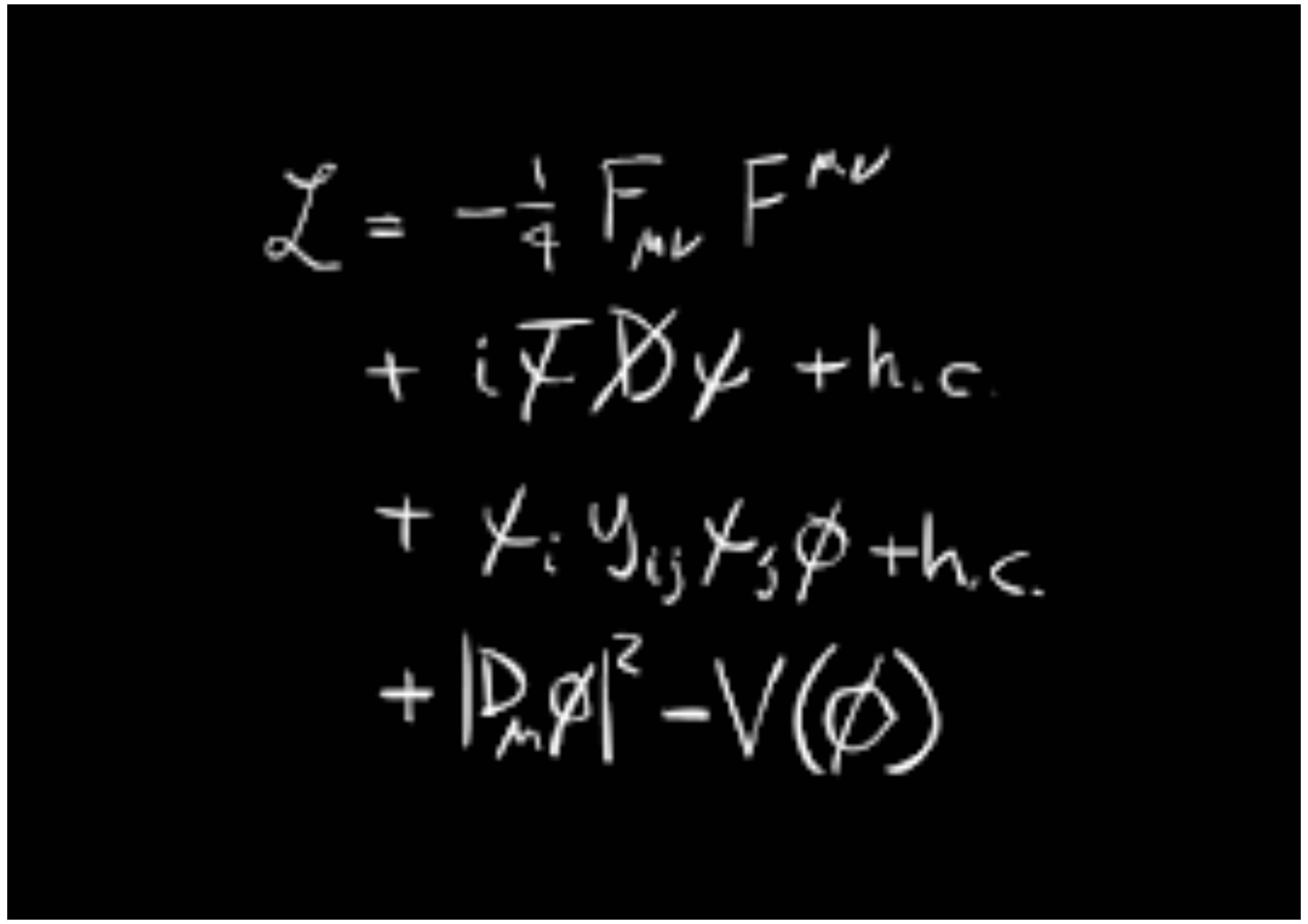
Dark energy?



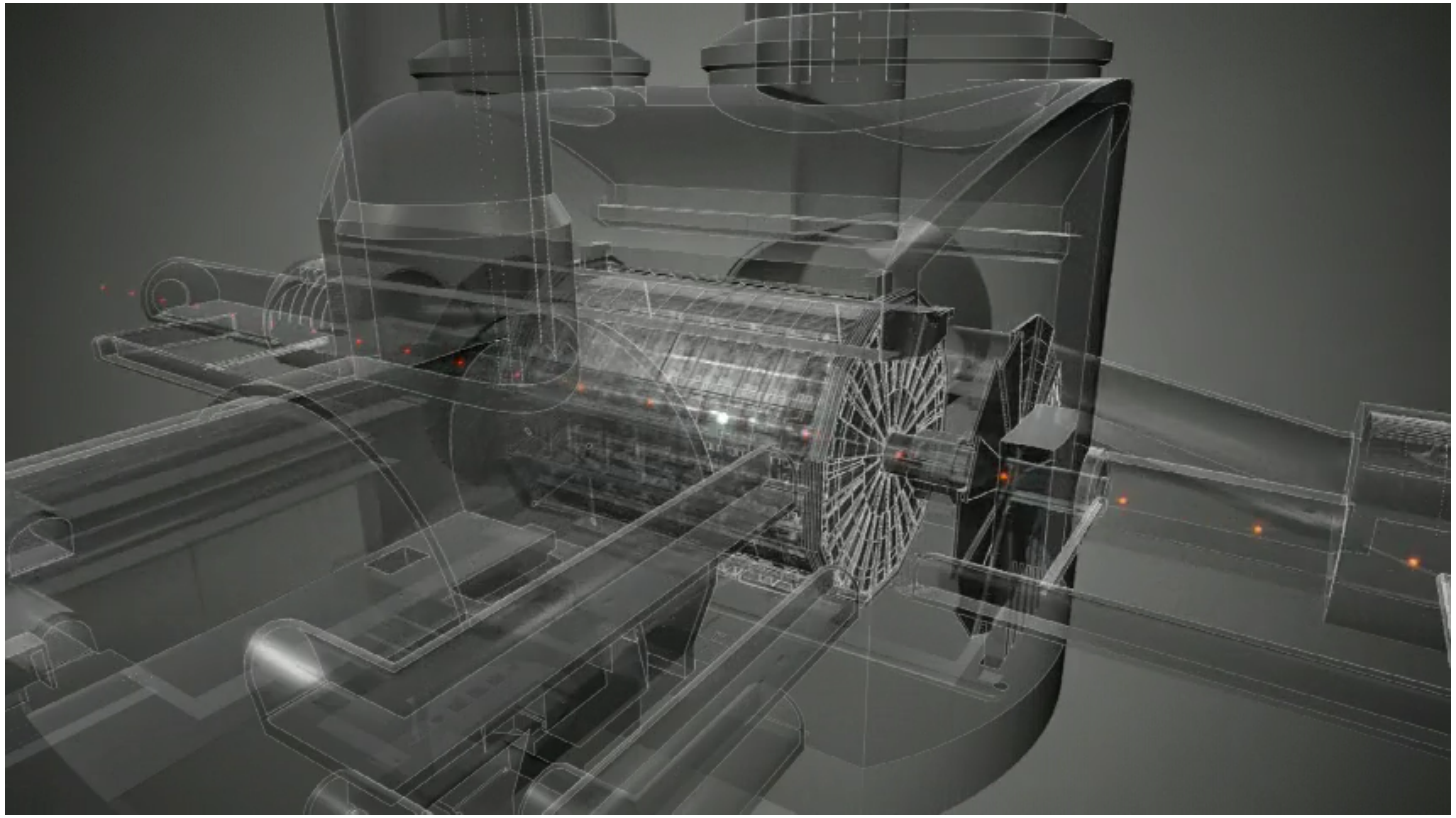
Are there new forces / particles to discover ?

Design theories,
make predictions

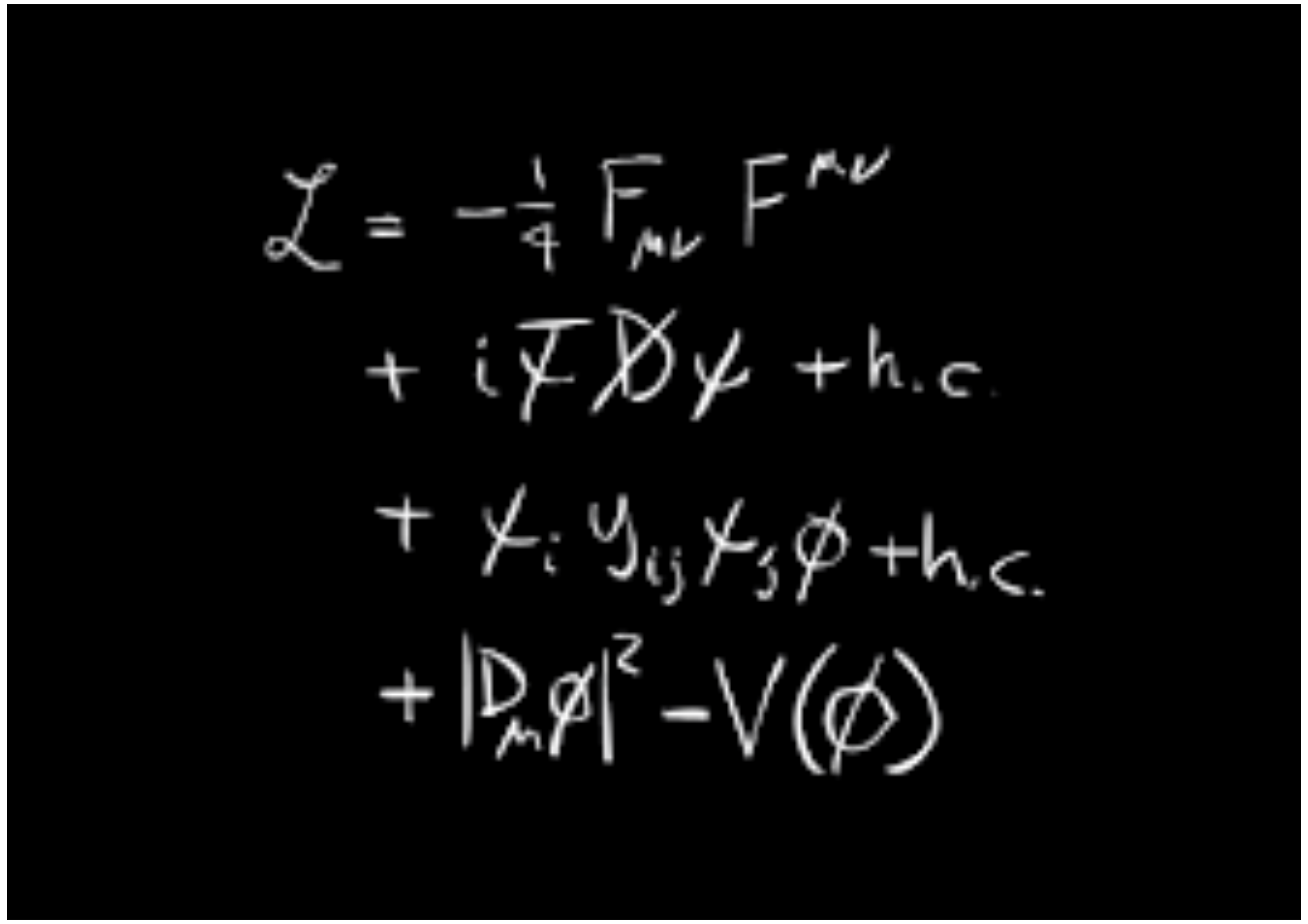
Design experiments,
test theories against data



$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + i\bar{\psi} \not{D} \psi + h.c. + \psi_i y_{ij} \psi_j \phi + h.c. + |D_\mu \phi|^2 - V(\phi)$$

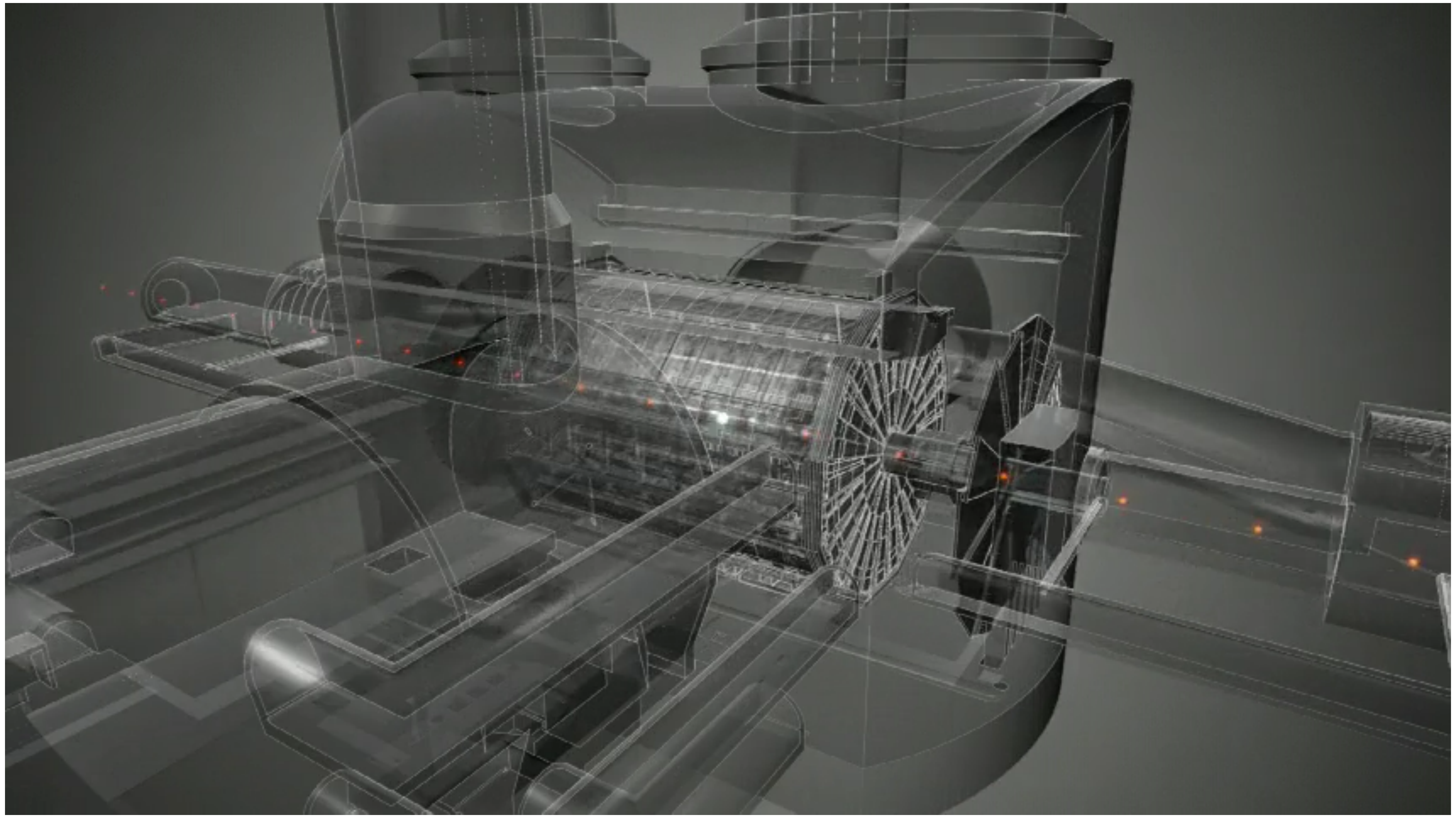


Design theories,
make predictions

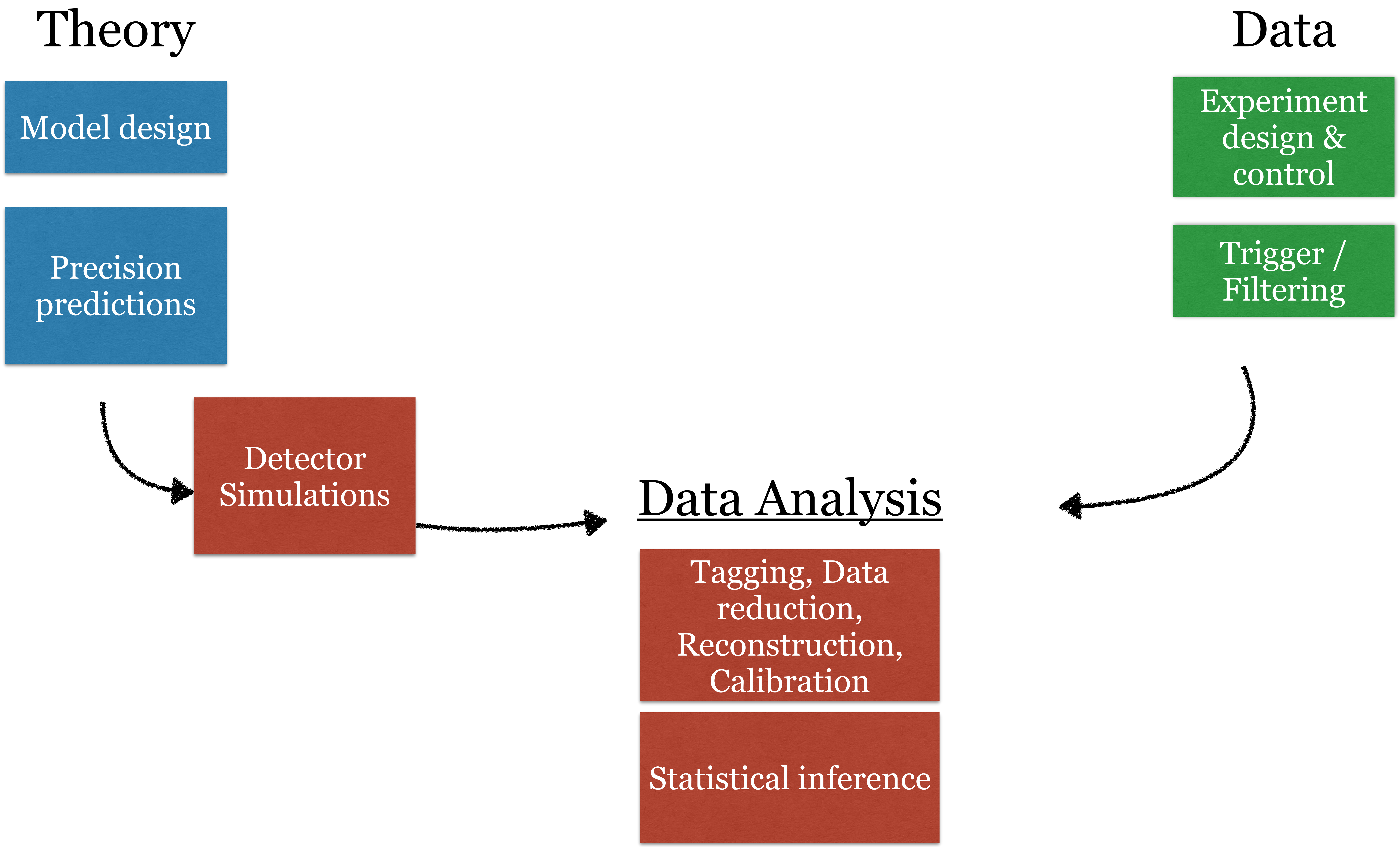


$$\begin{aligned}\mathcal{L} &= -\tfrac{1}{4} F_{\mu\nu} F^{\mu\nu} \\ &+ i\bar{\psi} \not{D} \psi + h.c. \\ &+ \psi_i y_{ij} \psi_j \phi + h.c. \\ &+ |D_\mu \phi|^2 - V(\phi)\end{aligned}$$

Design experiments,
test theories against data



ML across the scientific pipeline



Theory

Model design

Precision predictions

Detector Simulations

Data

Experiment design & control

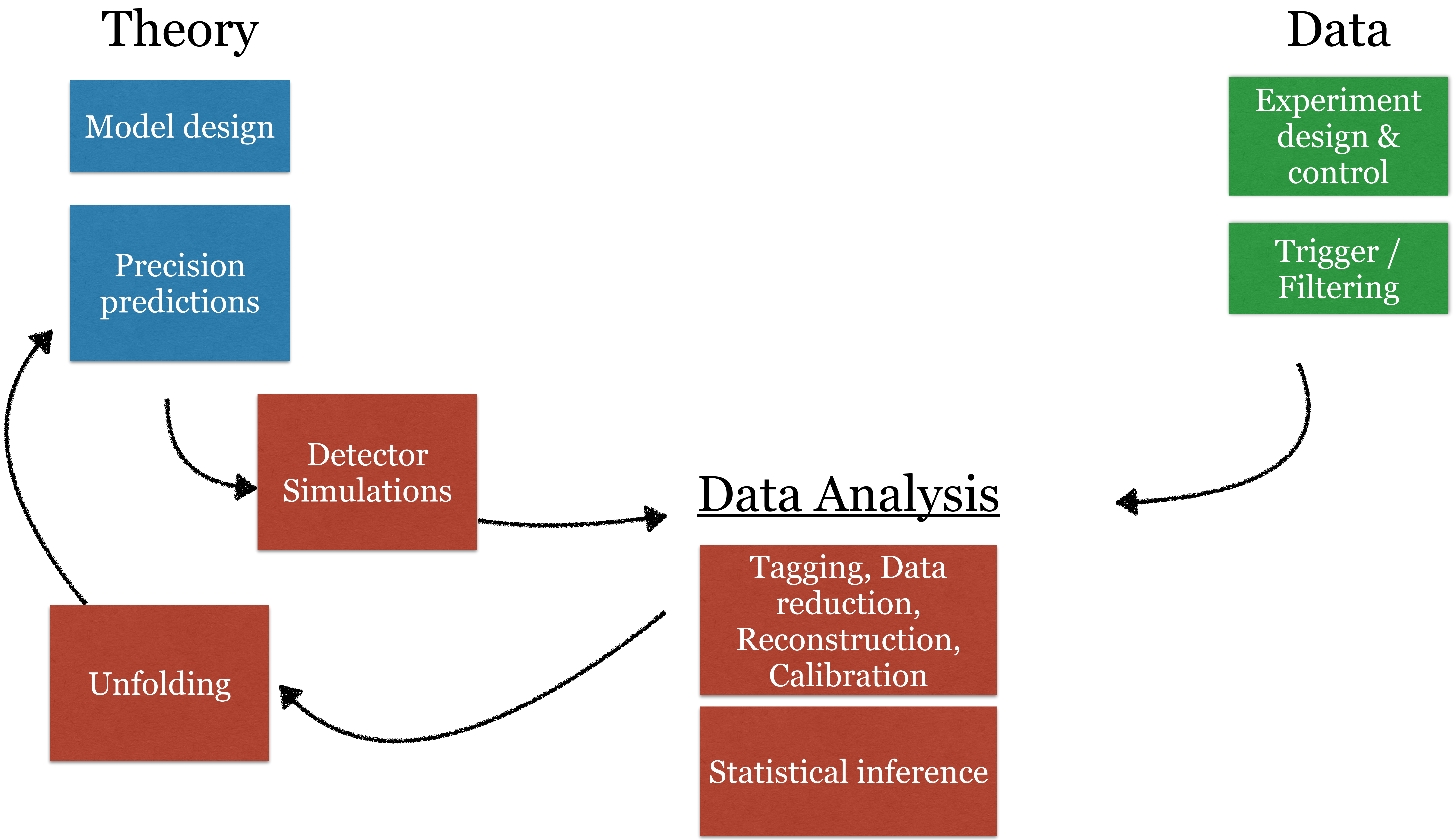
Trigger / Filtering

Data Analysis

Tagging, Data reduction, Reconstruction, Calibration

Statistical inference

ML across the scientific pipeline



Theory

Model design

Precision predictions

Data

Experiment design & control

Trigger / Filtering

Detector Simulations

Data Analysis

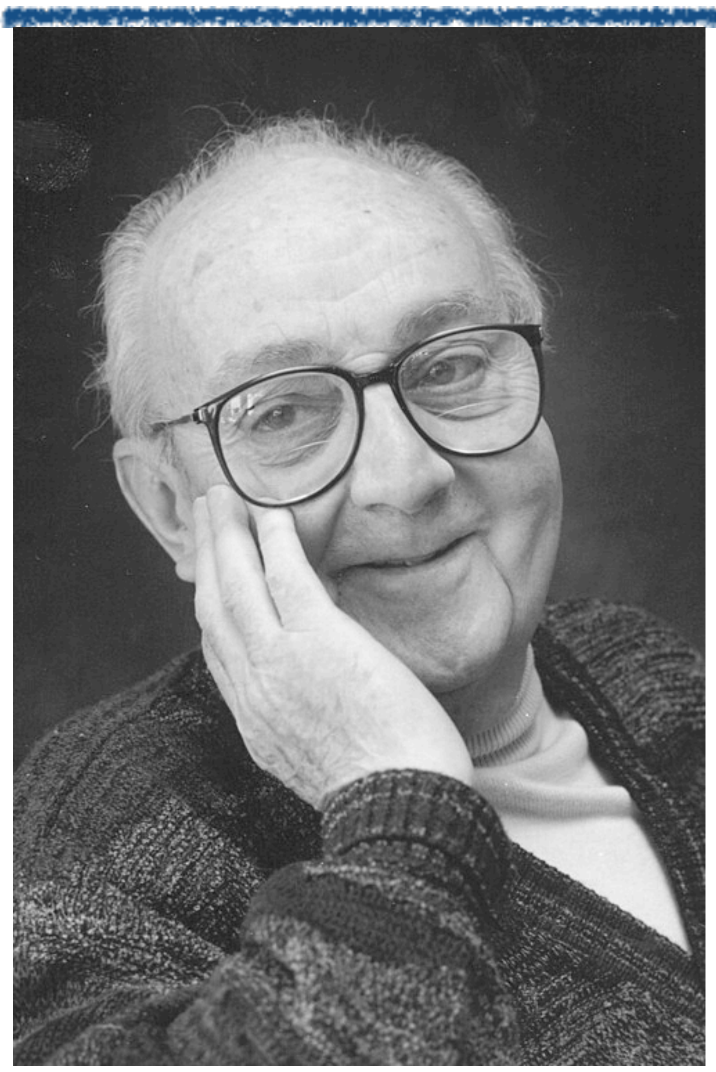
Tagging, Data reduction, Reconstruction, Calibration

Statistical inference

Unfolding

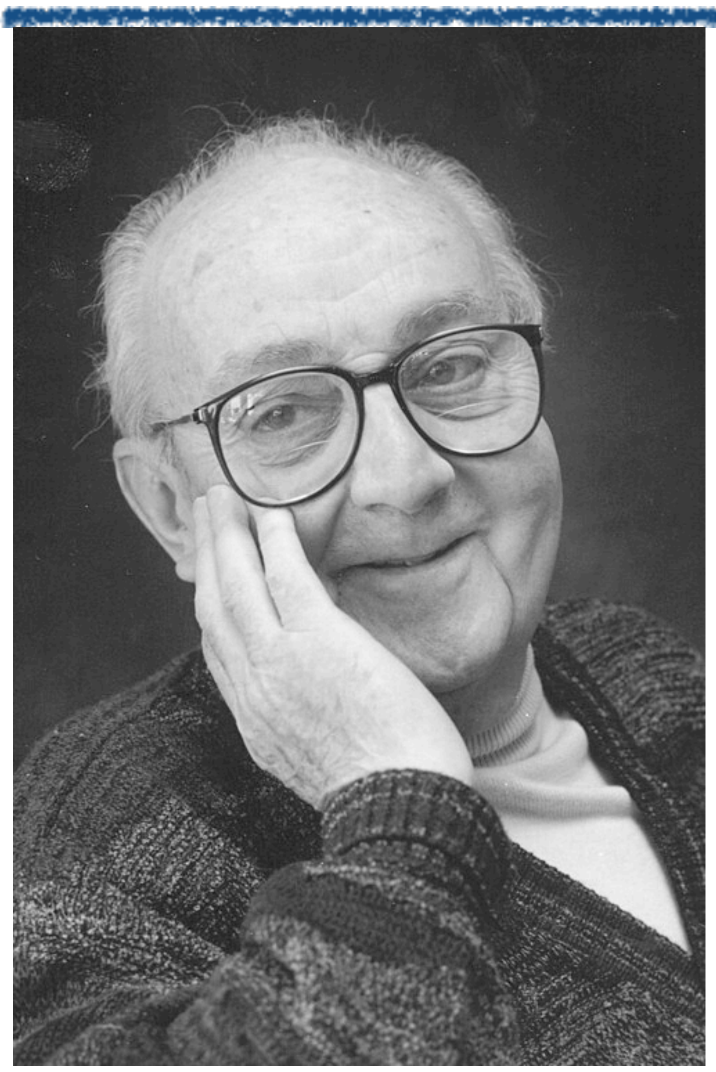
Physicists build mathematical models of the Universe

Physicists build mathematical models of the Universe



“All models are wrong, but some are useful”

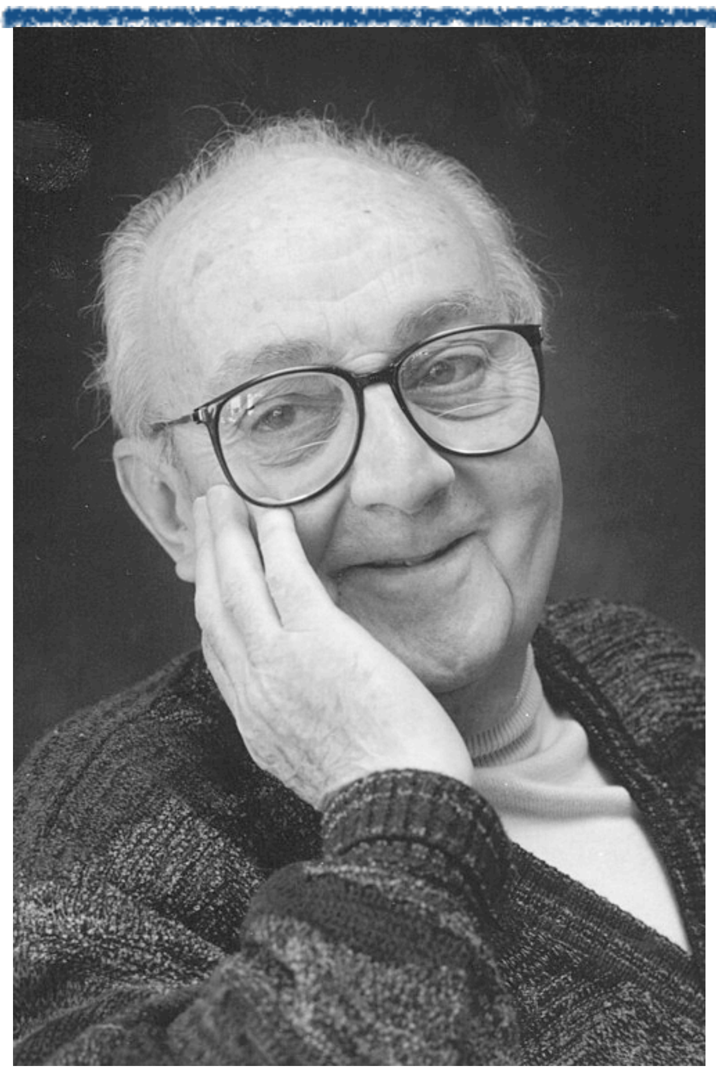
Physicists build mathematical models of the Universe



“All models are wrong, but some are useful”

What makes a useful model?: They are predictive

Physicists build mathematical models of the Universe

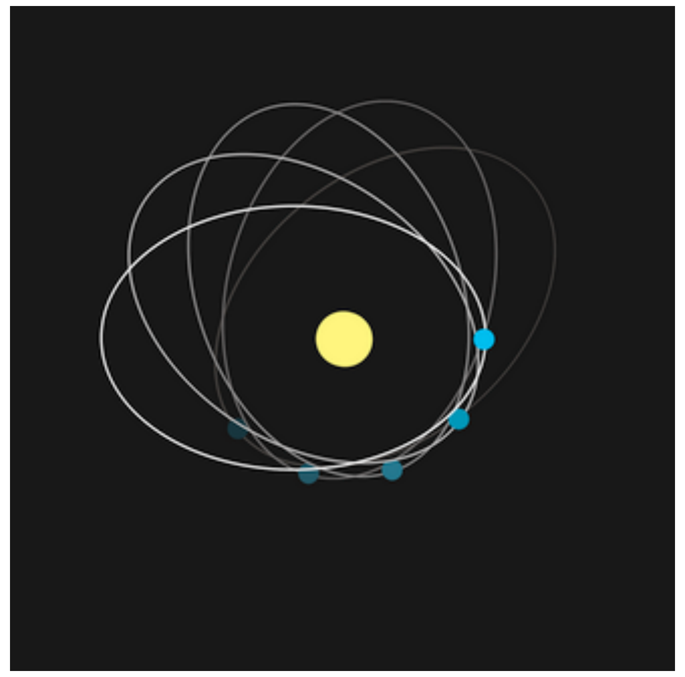


“All models are wrong, but some are useful”

What makes a useful model?: They are predictive

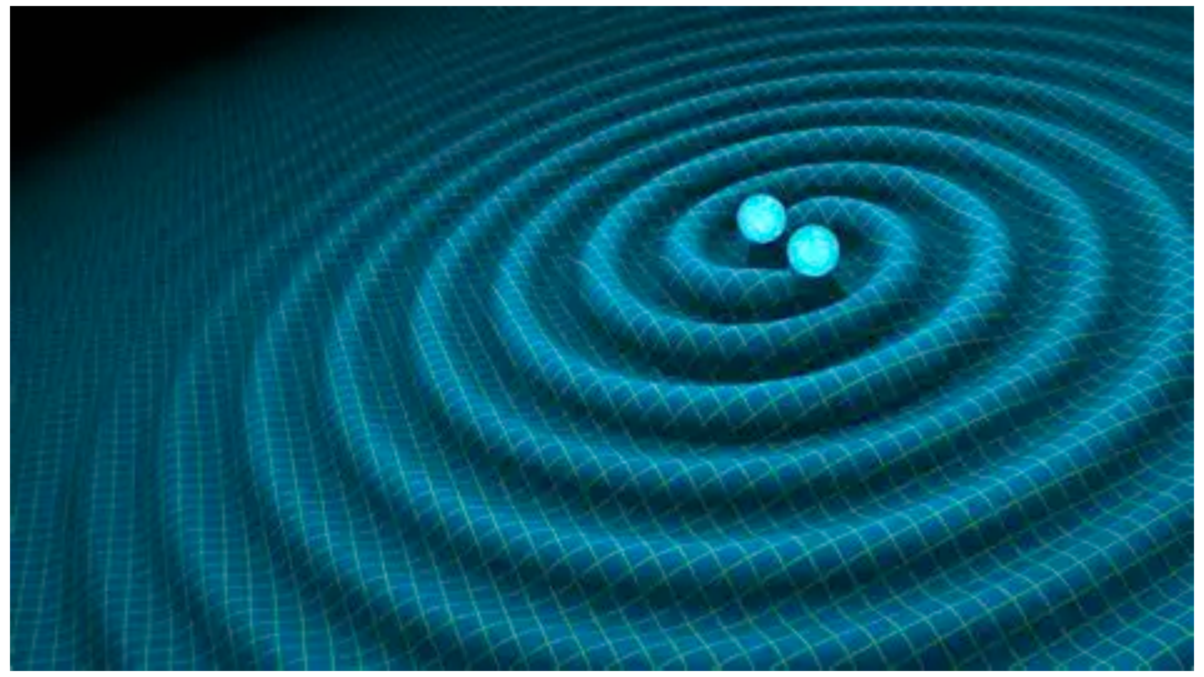
General Relativity reproduces Newtonian gravity in the right limits

But also predicts:



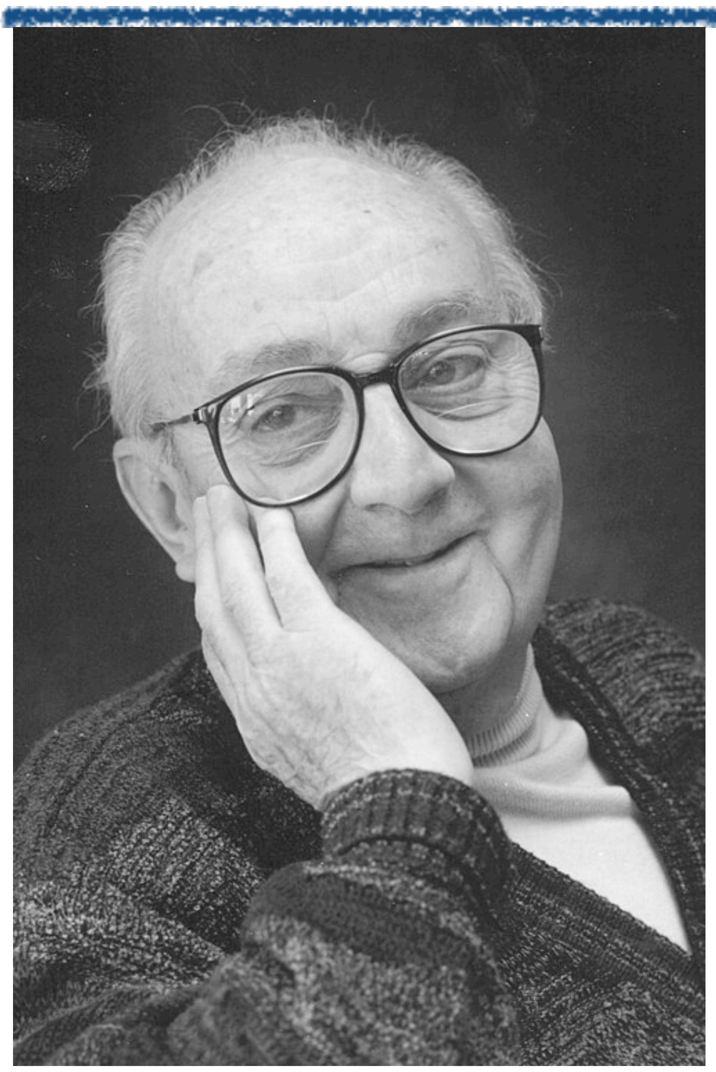
Perihelion of mercury

Black Holes



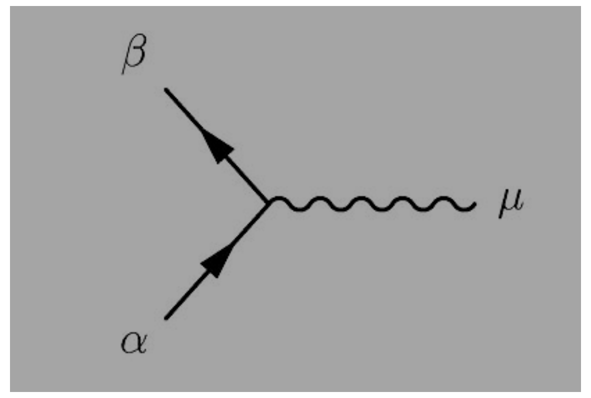
Gravitational Waves

Physicists build mathematical models of the Universe



“All models are wrong, but some are useful”

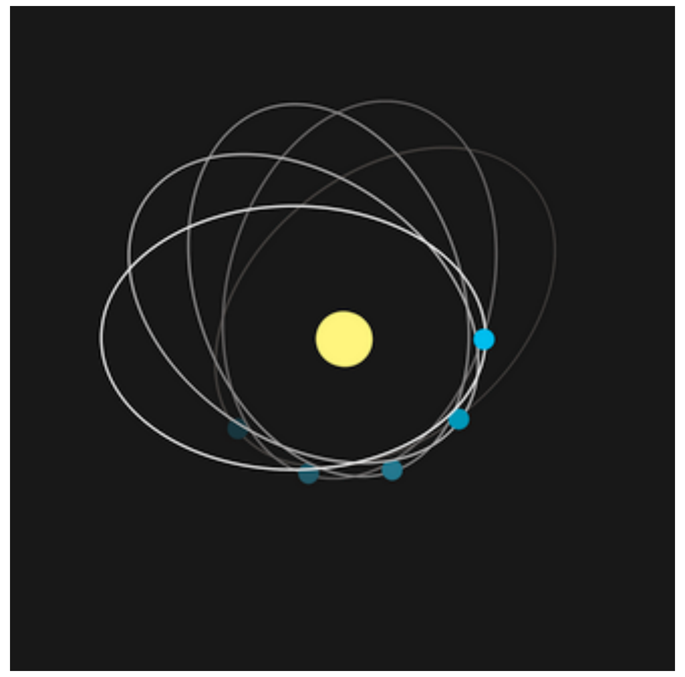
What makes a useful model?: They are predictive



Particle physics is more messy!

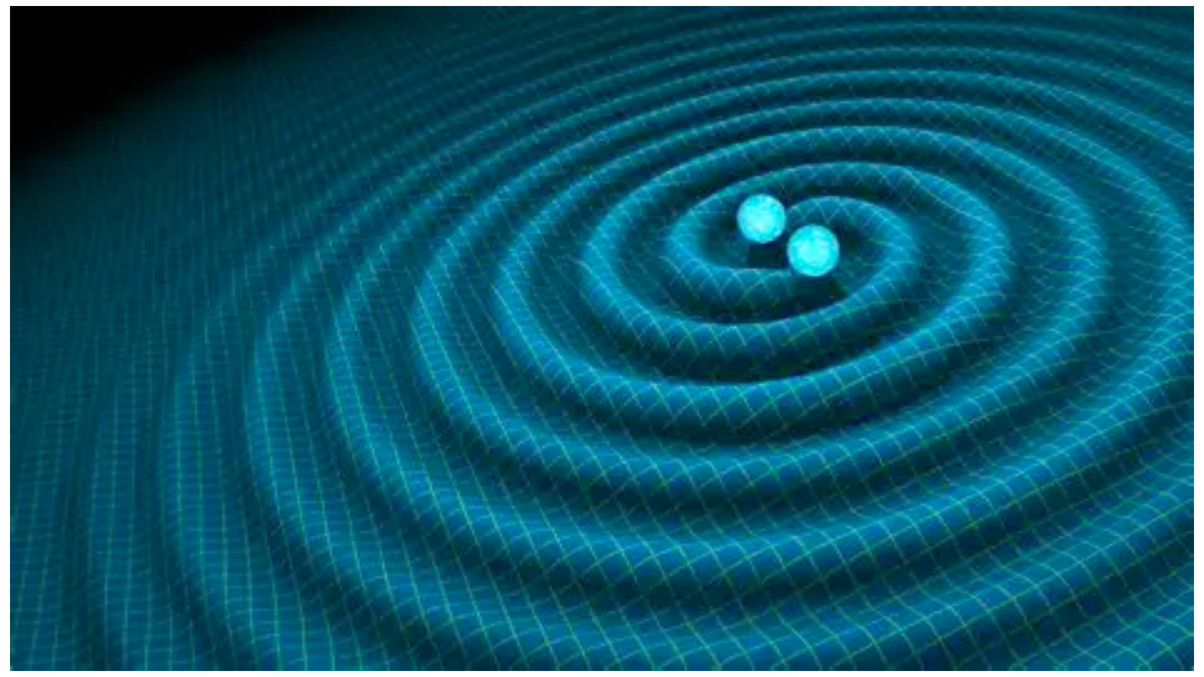
General Relativity reproduces Newtonian gravity in the right limits

But also predicts:



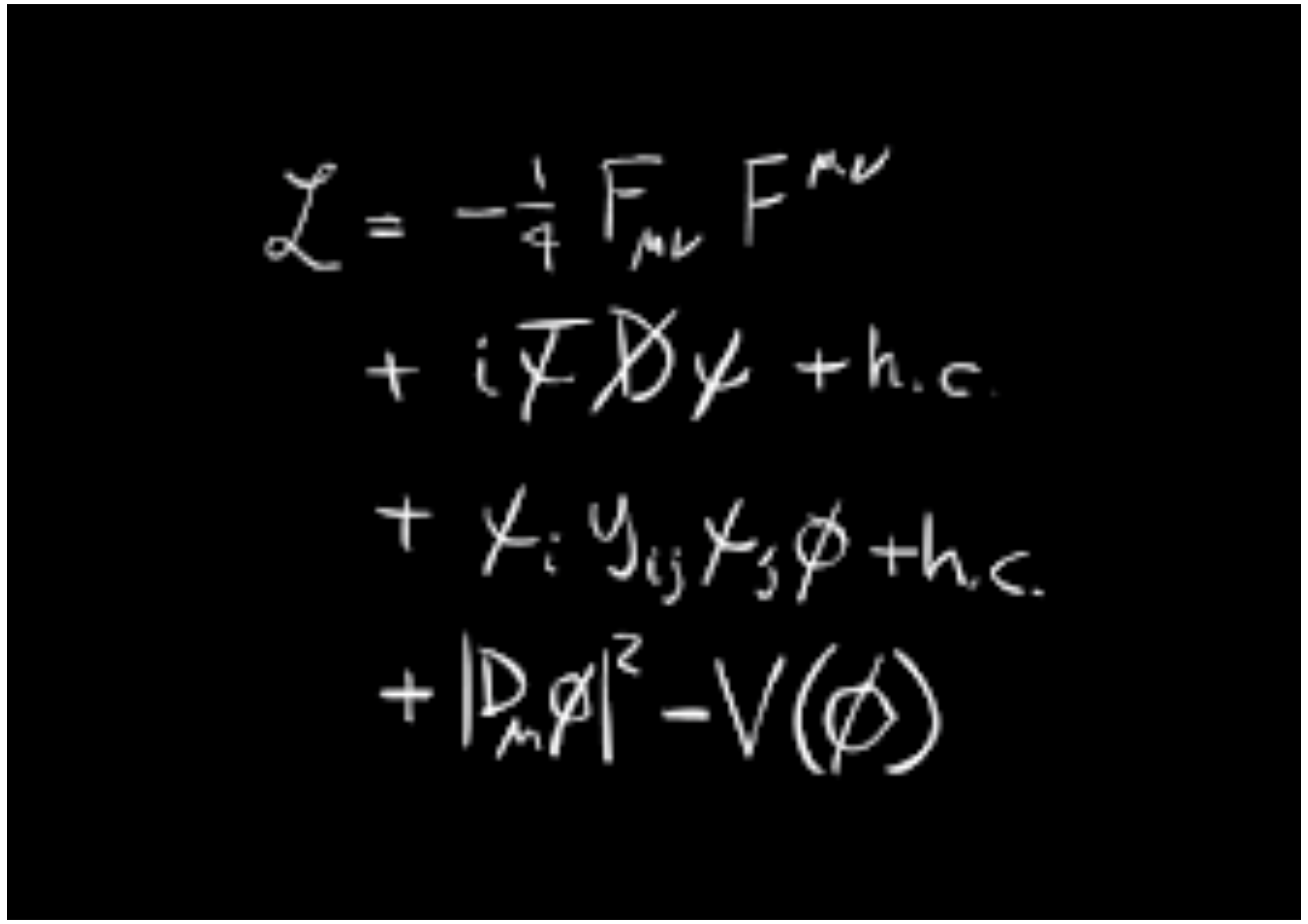
Perihelion of mercury

Black Holes



Gravitational Waves

Visible universe explained by this Lagrangian

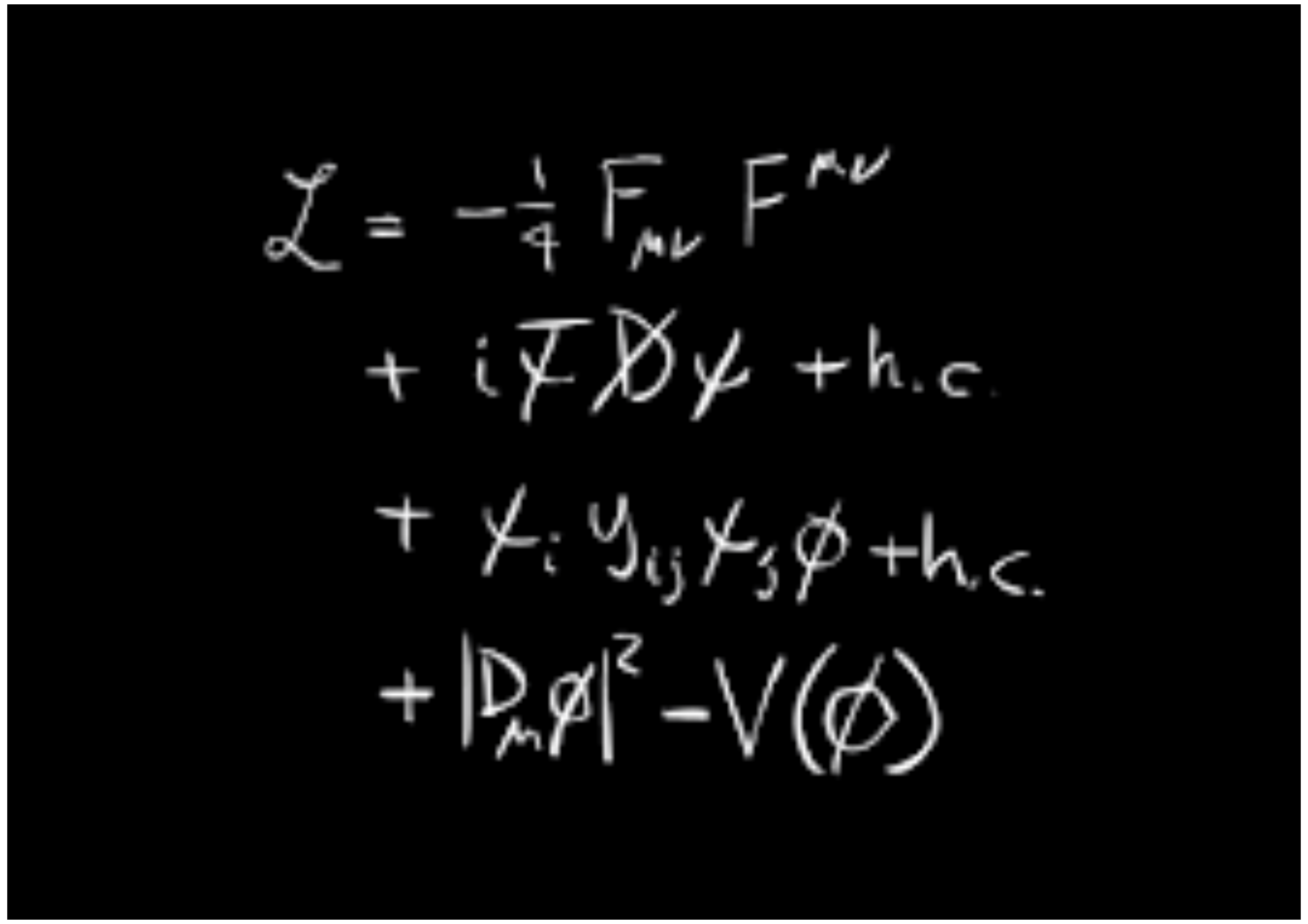


$$
\begin{aligned}
\mathcal{L} = & -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} \\
& + i\bar{\psi} \not{D} \psi + h.c. \\
& + \psi_i y_{ij} \psi_j \phi + h.c. \\
& + |D_\mu \phi|^2 - V(\phi)
\end{aligned}
$$

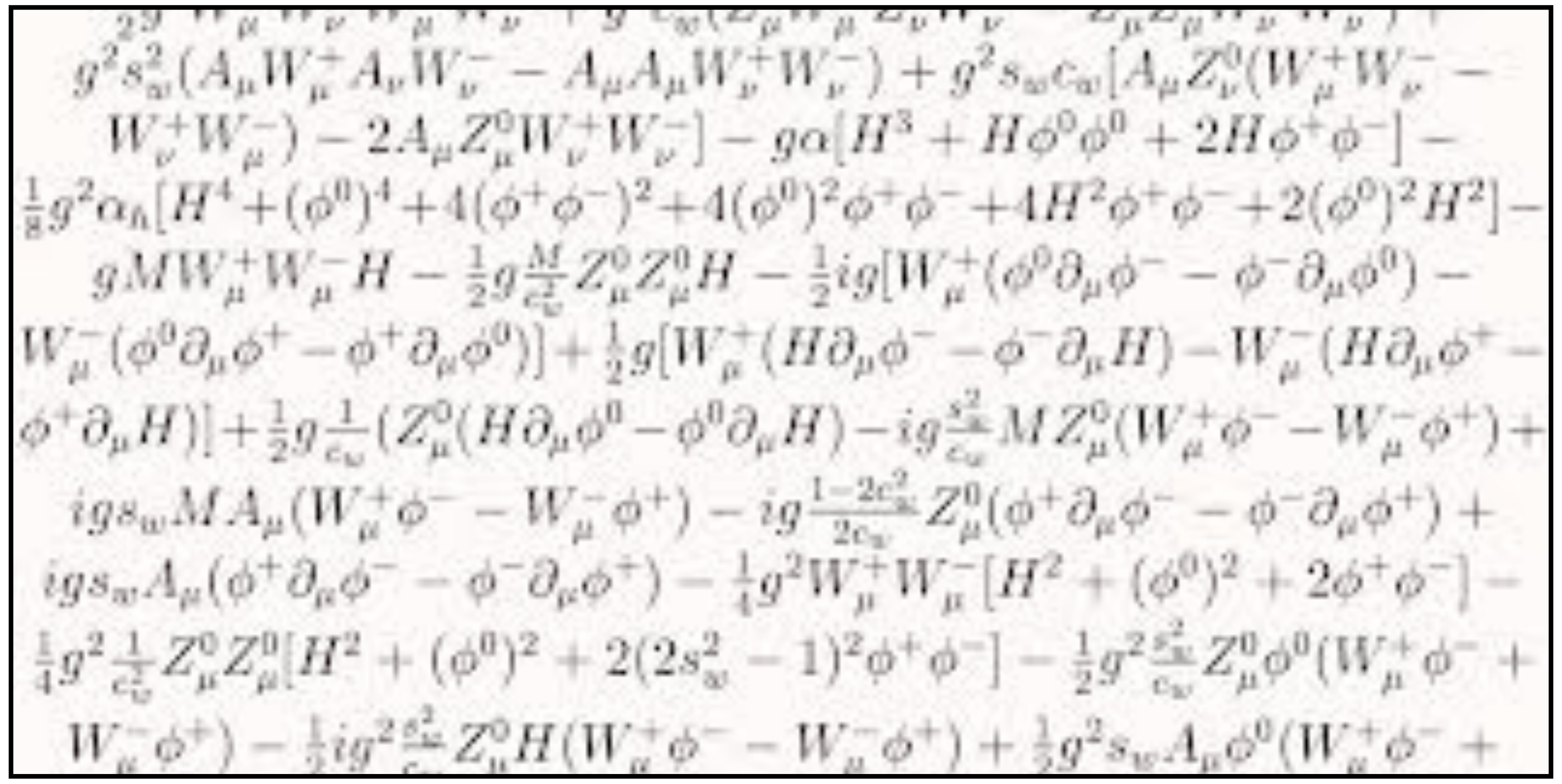
The Standard Model of Particle Physics

Visible universe explained by this Lagrangian

The Standard Model of Particle Physics



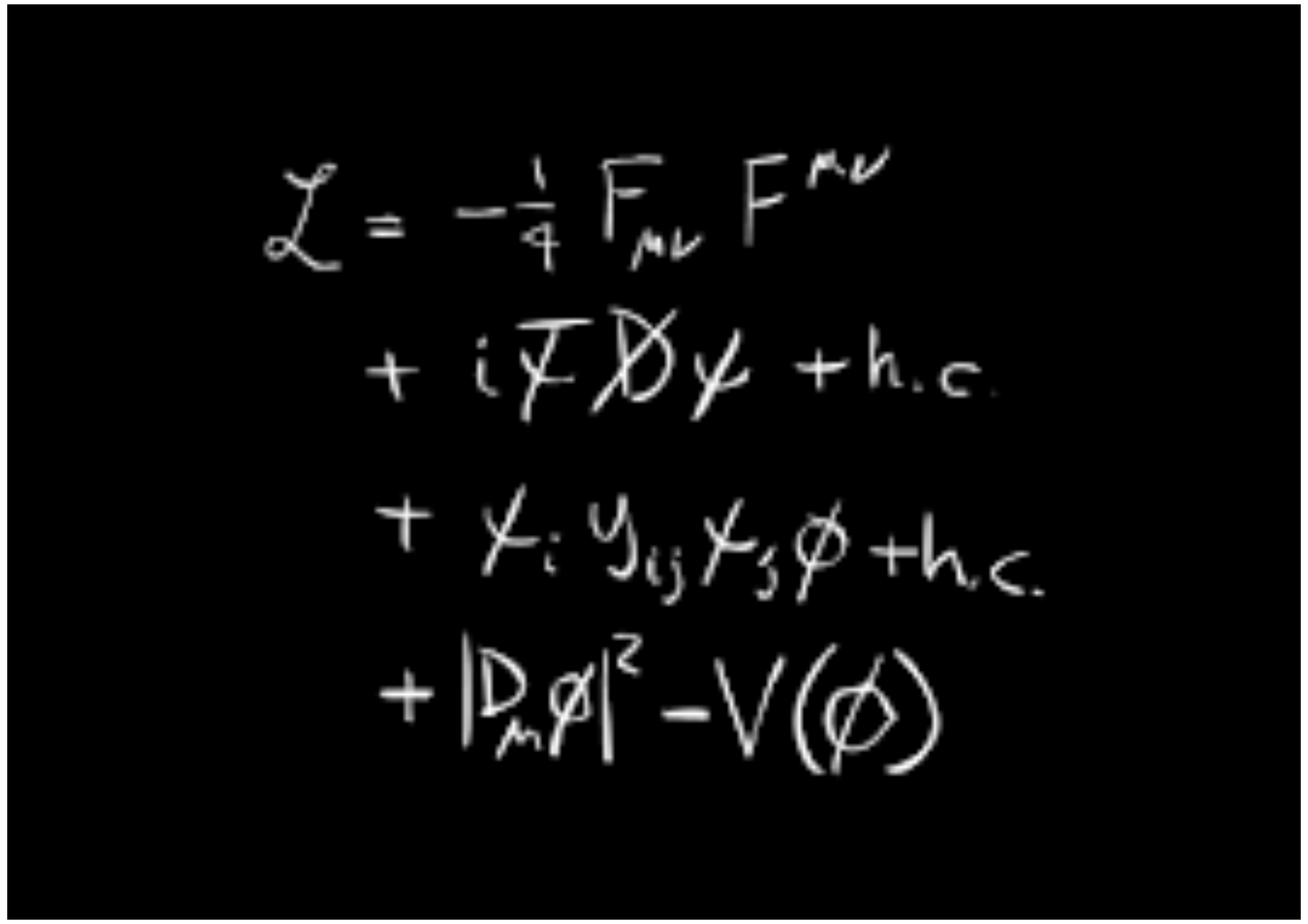
$$\begin{aligned}\mathcal{L} = &-\tfrac{1}{4} F_{\mu\nu} F^{\mu\nu} \\ &+ i\bar{\psi} \not{D} \psi + h.c. \\ &+ \psi_i y_{ij} \psi_j \phi + h.c. \\ &+ |D_\mu \phi|^2 - V(\phi)\end{aligned}$$



$$\begin{gathered}
g^2 s_w^2 (A_\mu W_\mu^+ A_\nu W_\nu^- - A_\mu A_\mu W_\nu^+ W_\nu^-) + g^2 s_w c_w [A_\mu Z_\nu^0 (W_\mu^+ W_\nu^- - \\
W_\nu^+ W_\mu^-) - 2 A_\mu Z_\mu^0 W_\nu^+ W_\nu^-] - g\alpha [H^3 + H\phi^0\phi^0 + 2H\phi^+\phi^-] - \\
\tfrac{1}{8} g^2 \alpha_h [H^4 + (\phi^0)^4 + 4(\phi^+\phi^-)^2 + 4(\phi^0)^2\phi^+\phi^- + 4H^2\phi^+\phi^- + 2(\phi^0)^2 H^2] - \\
g M W_\mu^+ W_\mu^- H - \tfrac{1}{2} g \tfrac{M}{c_w^2} Z_\mu^0 Z_\mu^0 H - \tfrac{1}{2} i g [W_\mu^+ (\phi^0 \partial_\mu \phi^- - \phi^- \partial_\mu \phi^0) - \\
W_\mu^- (\phi^0 \partial_\mu \phi^+ - \phi^+ \partial_\mu \phi^0)] + \tfrac{1}{2} g [W_\mu^+ (H \partial_\mu \phi^- - \phi^- \partial_\mu H) - W_\mu^- (H \partial_\mu \phi^+ - \\
\phi^+ \partial_\mu H)] + \tfrac{1}{2} g \tfrac{1}{c_w} (Z_\mu^0 (H \partial_\mu \phi^0 - \phi^0 \partial_\mu H) - i g \tfrac{s_w^2}{c_w} M Z_\mu^0 (W_\mu^+ \phi^- - W_\mu^- \phi^+) + \\
i g s_w M A_\mu (W_\mu^+ \phi^- - W_\mu^- \phi^+) - i g \tfrac{1-2c_w^2}{2c_w} Z_\mu^0 (\phi^+ \partial_\mu \phi^- - \phi^- \partial_\mu \phi^+) + \\
i g s_w A_\mu (\phi^+ \partial_\mu \phi^- - \phi^- \partial_\mu \phi^+) - \tfrac{1}{4} g^2 W_\mu^+ W_\mu^- [H^2 + (\phi^0)^2 + 2\phi^+\phi^-] - \\
\tfrac{1}{4} g^2 \tfrac{1}{c_w^2} Z_\mu^0 Z_\mu^0 [H^2 + (\phi^0)^2 + 2(2s_w^2 - 1)^2 \phi^+\phi^-] - \tfrac{1}{2} g^2 \tfrac{s_w^2}{c_w} Z_\mu^0 \phi^0 (W_\mu^+ \phi^- + \\
W_\mu^- \phi^+) - \tfrac{1}{2} i g^2 \tfrac{s_w^2}{c_w} Z_\mu^0 H (W_\mu^+ \phi^- - W_\mu^- \phi^+) + \tfrac{1}{2} g^2 s_w A_\mu \phi^0 (W_\mu^+ \phi^- +
\end{gathered}$$

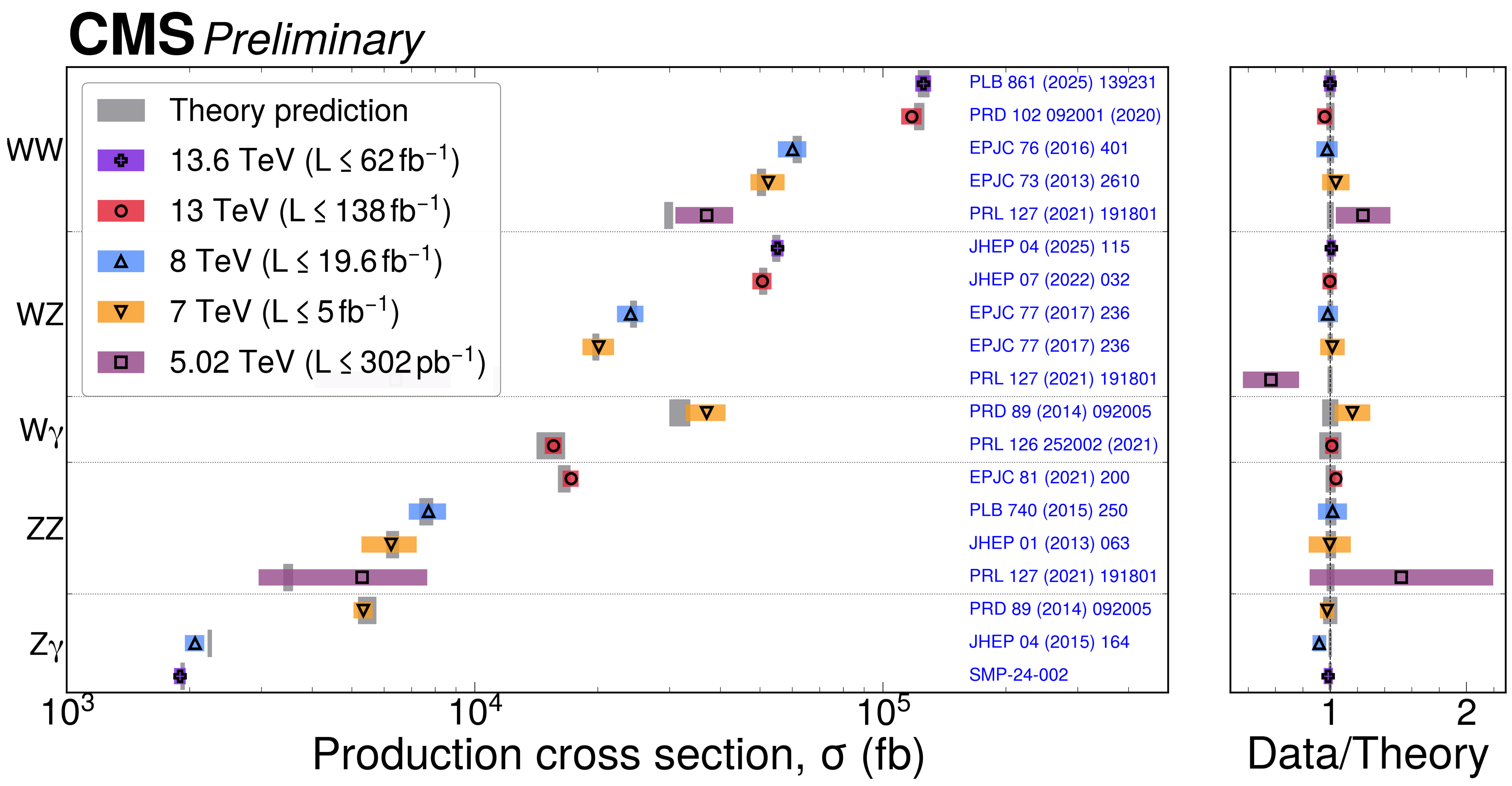
Same thing expanded…

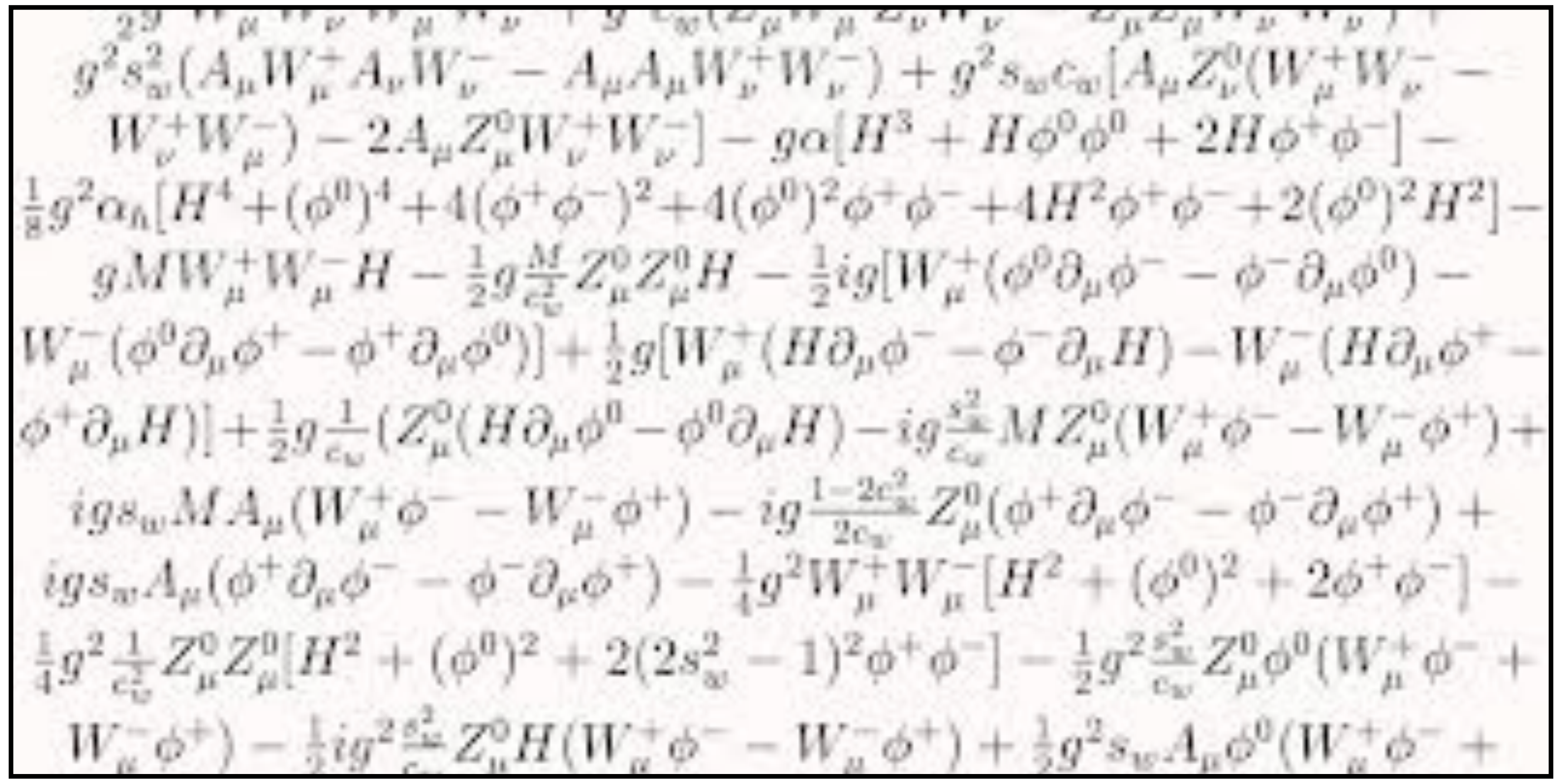
Visible universe explained by this Lagrangian



$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu}$$
$$+ i \bar{\psi} \not{D} \psi + h.c.$$
$$+ \psi_i y_{ij} \psi_j \phi + h.c.$$
$$+ |D_\mu \phi|^2 - V(\phi)$$

The Standard Model of Particle Physics

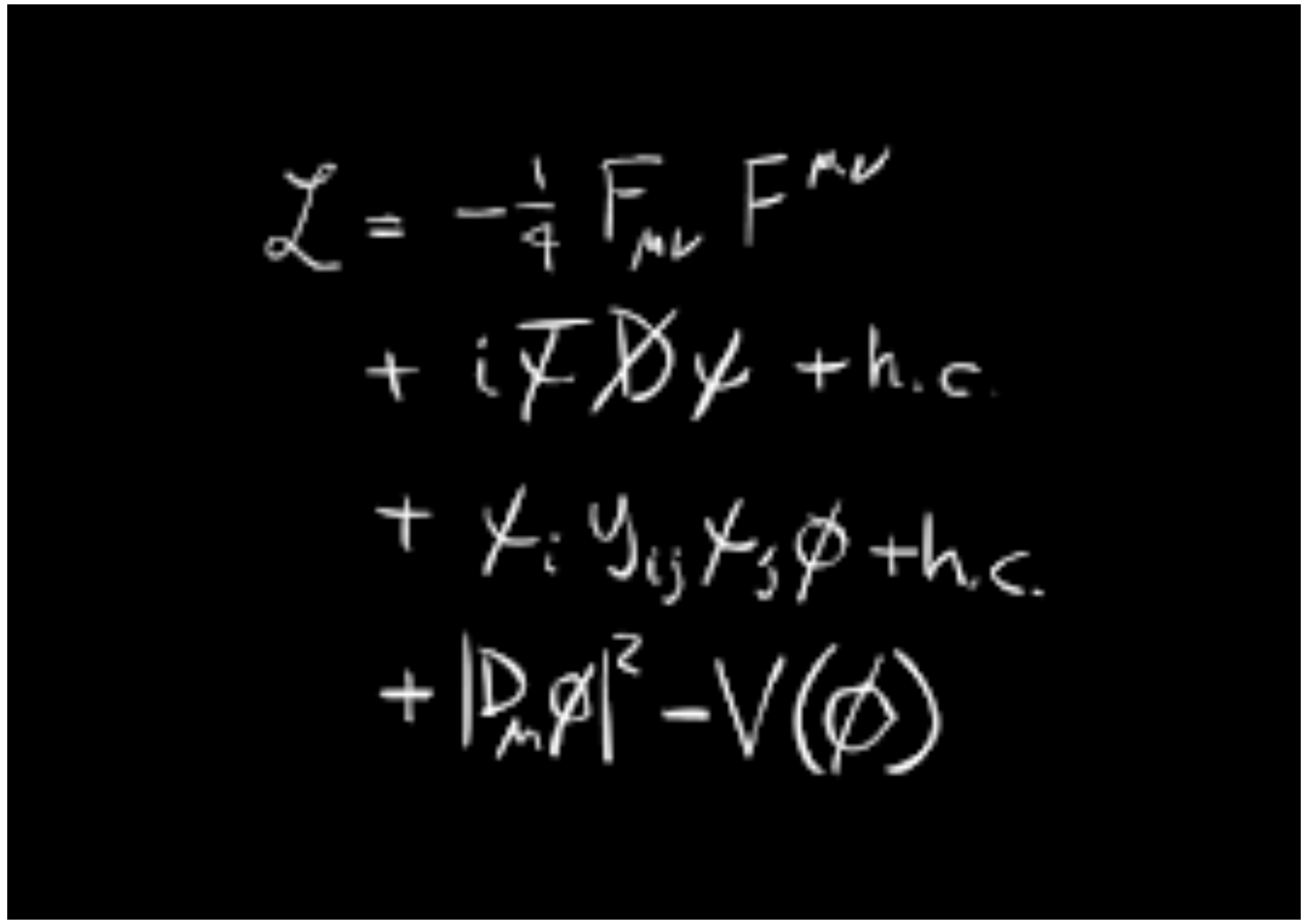




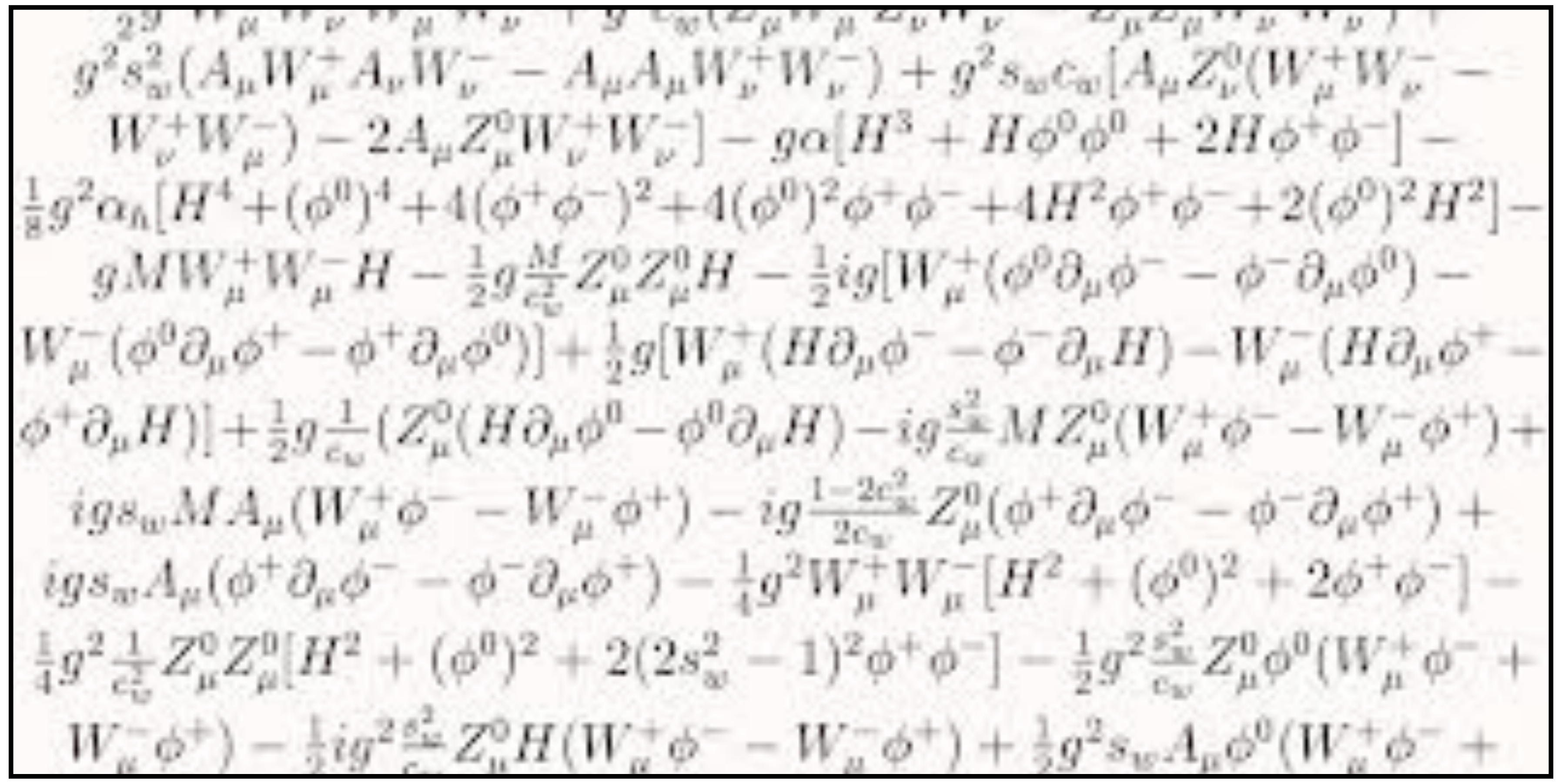
$$g^2 s_w^2 (A_\mu W_\mu^+ A_\nu W_\nu^- - A_\mu A_\mu W_\nu^+ W_\nu^-) + g^2 s_w c_w [A_\mu Z_\nu^0 (W_\mu^+ W_\nu^- -$$
$$W_\nu^+ W_\mu^-) - 2 A_\mu Z_\mu^0 W_\nu^+ W_\nu^-] - g\alpha [H^3 + H\phi^0\phi^0 + 2H\phi^+\phi^-] -$$
$$\frac{1}{8} g^2 \alpha_h [H^4 + (\phi^0)^4 + 4(\phi^+\phi^-)^2 + 4(\phi^0)^2\phi^+\phi^- + 4H^2\phi^+\phi^- + 2(\phi^0)^2 H^2] -$$
$$g M W_\mu^+ W_\mu^- H - \frac{1}{2} g \frac{M}{c_w^2} Z_\mu^0 Z_\mu^0 H - \frac{1}{2} i g [W_\mu^+ (\phi^0 \partial_\mu \phi^- - \phi^- \partial_\mu \phi^0) -$$
$$W_\mu^- (\phi^0 \partial_\mu \phi^+ - \phi^+ \partial_\mu \phi^0)] + \frac{1}{2} g [W_\mu^+ (H \partial_\mu \phi^- - \phi^- \partial_\mu H) - W_\mu^- (H \partial_\mu \phi^+ -$$
$$\phi^+ \partial_\mu H)] + \frac{1}{2} g \frac{1}{c_w} (Z_\mu^0 (H \partial_\mu \phi^0 - \phi^0 \partial_\mu H) - i g \frac{s_w^2}{c_w} M Z_\mu^0 (W_\mu^+ \phi^- - W_\mu^- \phi^+) +$$
$$i g s_w M A_\mu (W_\mu^+ \phi^- - W_\mu^- \phi^+) - i g \frac{1 - 2c_w^2}{2c_w} Z_\mu^0 (\phi^+ \partial_\mu \phi^- - \phi^- \partial_\mu \phi^+) +$$
$$i g s_w A_\mu (\phi^+ \partial_\mu \phi^- - \phi^- \partial_\mu \phi^+) - \frac{1}{4} g^2 W_\mu^+ W_\mu^- [H^2 + (\phi^0)^2 + 2\phi^+\phi^-] -$$
$$\frac{1}{4} g^2 \frac{1}{c_w^2} Z_\mu^0 Z_\mu^0 [H^2 + (\phi^0)^2 + 2(2s_w^2 - 1)^2 \phi^+\phi^-] - \frac{1}{2} g^2 \frac{s_w^2}{c_w} Z_\mu^0 \phi^0 (W_\mu^+ \phi^- +$$
$$W_\mu^- \phi^+) - \frac{1}{2} i g^2 \frac{s_w^2}{c_w} Z_\mu^0 H (W_\mu^+ \phi^- - W_\mu^- \phi^+) + \frac{1}{2} g^2 s_w A_\mu \phi^0 (W_\mu^+ \phi^- +$$

Same thing expanded...

All explained by this Lagrangian



$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu}$$
$$+ i\bar{\psi} \not{D} \psi + h.c.$$
$$+ \psi_i y_{ij} \psi_j \phi + h.c.$$
$$+ |D_\mu \phi|^2 - V(\phi)$$



$$\begin{gathered}
g^2 s_w^2 (A_\mu W_\mu^+ A_\nu W_\nu^- - A_\mu A_\mu W_\nu^+ W_\nu^-) + g^2 s_w c_w [A_\mu Z_\nu^0 (W_\mu^+ W_\nu^- - \\
W_\nu^+ W_\mu^-) - 2 A_\mu Z_\mu^0 W_\nu^+ W_\nu^-] - g\alpha [H^3 + H\phi^0\phi^0 + 2H\phi^+\phi^-] - \\
\frac{1}{8} g^2 \alpha_h [H^4 + (\phi^0)^4 + 4(\phi^+\phi^-)^2 + 4(\phi^0)^2\phi^+\phi^- + 4H^2\phi^+\phi^- + 2(\phi^0)^2 H^2] - \\
g M W_\mu^+ W_\mu^- H - \frac{1}{2} g \frac{M}{c_w^2} Z_\mu^0 Z_\mu^0 H - \frac{1}{2} ig [W_\mu^+ (\phi^0 \partial_\mu \phi^- - \phi^- \partial_\mu \phi^0) - \\
W_\mu^- (\phi^0 \partial_\mu \phi^+ - \phi^+ \partial_\mu \phi^0)] + \frac{1}{2} g [W_\mu^+ (H \partial_\mu \phi^- - \phi^- \partial_\mu H) - W_\mu^- (H \partial_\mu \phi^+ - \\
\phi^+ \partial_\mu H)] + \frac{1}{2} g \frac{1}{c_w} (Z_\mu^0 (H \partial_\mu \phi^0 - \phi^0 \partial_\mu H) - ig \frac{s_w^2}{c_w} M Z_\mu^0 (W_\mu^+ \phi^- - W_\mu^- \phi^+) + \\
ig s_w M A_\mu (W_\mu^+ \phi^- - W_\mu^- \phi^+) - ig \frac{1 - 2c_w^2}{2c_w} Z_\mu^0 (\phi^+ \partial_\mu \phi^- - \phi^- \partial_\mu \phi^+) + \\
ig s_w A_\mu (\phi^+ \partial_\mu \phi^- - \phi^- \partial_\mu \phi^+) - \frac{1}{4} g^2 W_\mu^+ W_\mu^- [H^2 + (\phi^0)^2 + 2\phi^+\phi^-] - \\
\frac{1}{4} g^2 \frac{1}{c_w^2} Z_\mu^0 Z_\mu^0 [H^2 + (\phi^0)^2 + 2(2s_w^2 - 1)^2 \phi^+\phi^-] - \frac{1}{2} g^2 \frac{s_w^2}{c_w} Z_\mu^0 \phi^0 (W_\mu^+ \phi^- + \\
W_\mu^- \phi^+) - \frac{1}{2} ig^2 \frac{s_w^2}{c_w} Z_\mu^0 H (W_\mu^+ \phi^- - W_\mu^- \phi^+) + \frac{1}{2} g^2 s_w A_\mu \phi^0 (W_\mu^+ \phi^- +
\end{gathered}$$

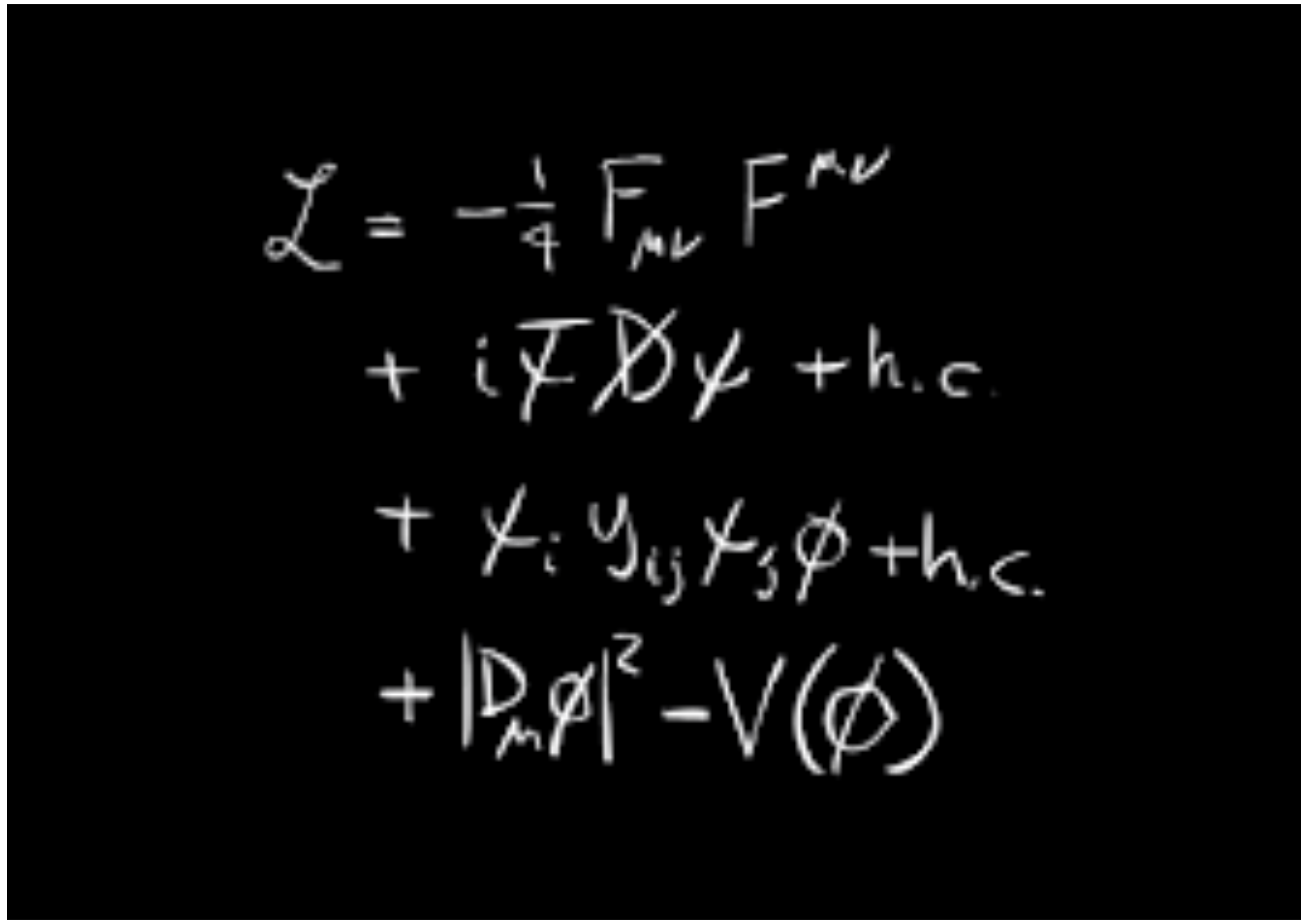
The Standard Model of Particle Physics

It’s deeply flawed:

- Can’t explain 85% of all matter (dark matter)
- Why more matter than anti-matter?
- Model parameters need incredible fine-tuning
- …
- **Why do neutrinos have mass and why do they oscillate?**

ν_τ | ν_e | ν_μ

All explained by this Lagrangian

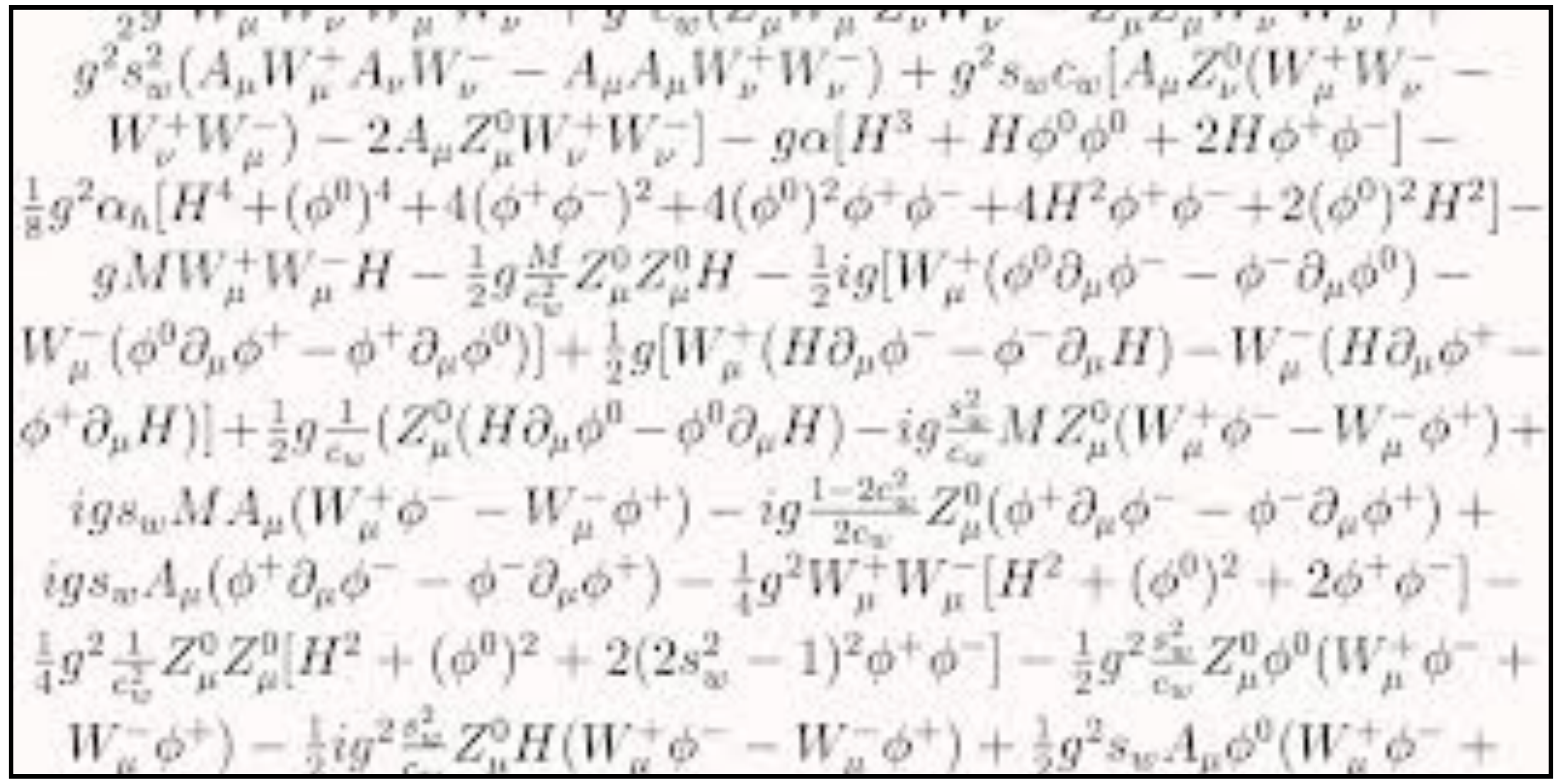


$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + i \bar{\psi} \not{D} \psi + h.c. + \psi_i y_{ij} \psi_j \phi + h.c. + |D_\mu \phi|^2 - V(\phi)$$

The Standard Model of Particle Physics

It’s deeply flawed:

- Can’t explain 85% of all matter (dark matter)
- Why more matter than anti-matter?
- Model parameters need incredible fine-tuning
- …
- **Why do neutrinos have mass and why do they oscillate?**



$$g^2 s_w^2 (A_\mu W_\mu^+ A_\nu W_\nu^- - A_\mu A_\mu W_\nu^+ W_\nu^-) + g^2 s_w c_w [A_\mu Z_\nu^0 (W_\mu^+ W_\nu^- - W_\nu^+ W_\mu^-) - 2 A_\mu Z_\mu^0 W_\nu^+ W_\nu^-] - g\alpha [H^3 + H\phi^0\phi^0 + 2H\phi^+\phi^-] -$$
$$\frac{1}{8} g^2 \alpha_h [H^4 + (\phi^0)^4 + 4(\phi^+\phi^-)^2 + 4(\phi^0)^2\phi^+\phi^- + 4H^2\phi^+\phi^- + 2(\phi^0)^2 H^2] -$$
$$g M W_\mu^+ W_\mu^- H - \frac{1}{2} g \frac{M}{c_w^2} Z_\mu^0 Z_\mu^0 H - \frac{1}{2} i g [W_\mu^+ (\phi^0 \partial_\mu \phi^- - \phi^- \partial_\mu \phi^0) -$$
$$W_\mu^- (\phi^0 \partial_\mu \phi^+ - \phi^+ \partial_\mu \phi^0)] + \frac{1}{2} g [W_\mu^+ (H \partial_\mu \phi^- - \phi^- \partial_\mu H) - W_\mu^- (H \partial_\mu \phi^+ -$$
$$\phi^+ \partial_\mu H)] + \frac{1}{2} g \frac{1}{c_w} (Z_\mu^0 (H \partial_\mu \phi^0 - \phi^0 \partial_\mu H) - i g \frac{s_w^2}{c_w} M Z_\mu^0 (W_\mu^+ \phi^- - W_\mu^- \phi^+) +$$
$$i g s_w M A_\mu (W_\mu^+ \phi^- - W_\mu^- \phi^+) - i g \frac{1 - 2c_w^2}{2c_w} Z_\mu^0 (\phi^+ \partial_\mu \phi^- - \phi^- \partial_\mu \phi^+) +$$
$$i g s_w A_\mu (\phi^+ \partial_\mu \phi^- - \phi^- \partial_\mu \phi^+) - \frac{1}{4} g^2 W_\mu^+ W_\mu^- [H^2 + (\phi^0)^2 + 2\phi^+\phi^-] -$$
$$\frac{1}{4} g^2 \frac{1}{c_w^2} Z_\mu^0 Z_\mu^0 [H^2 + (\phi^0)^2 + 2(2s_w^2 - 1)^2 \phi^+\phi^-] - \frac{1}{2} g^2 \frac{s_w^2}{c_w} Z_\mu^0 \phi^0 (W_\mu^+ \phi^- +$$
$$W_\mu^- \phi^+) - \frac{1}{2} i g^2 \frac{s_w^2}{c_w} Z_\mu^0 H (W_\mu^+ \phi^- - W_\mu^- \phi^+) + \frac{1}{2} g^2 s_w A_\mu \phi^0 (W_\mu^+ \phi^- +$$

Could we automatically design better theories and test them against data?

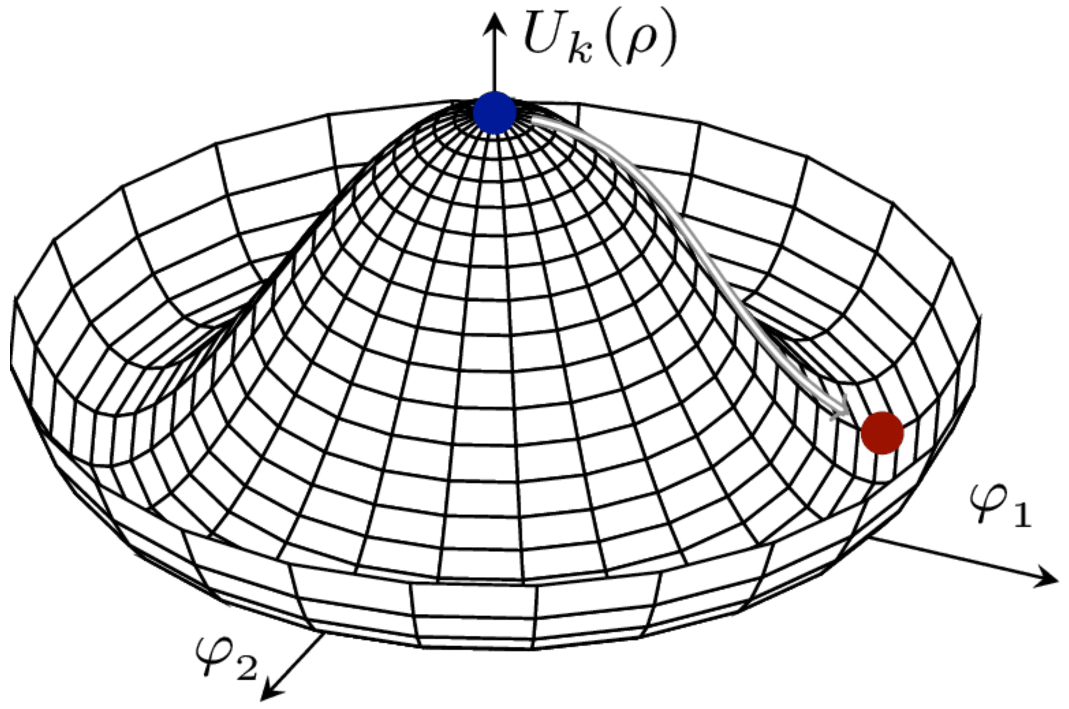
ν_τ | ν_e | ν_μ

The space of mathematical tricks we could use is enormous

In the old electroweak theory, **we didn't know how to give particles like Z and W bosons their mass** without violating important symmetries

The space of mathematical tricks we could use is enormous

In the old electroweak theory, **we didn't know how to give particles like Z and W bosons their mass** without violating important symmetries

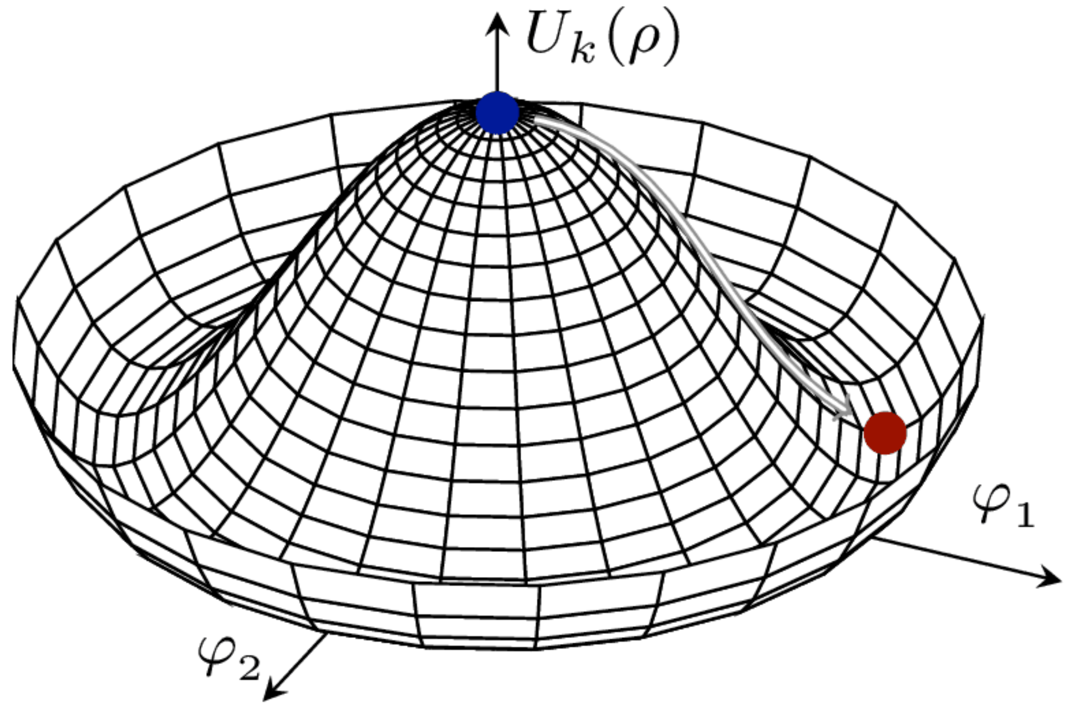


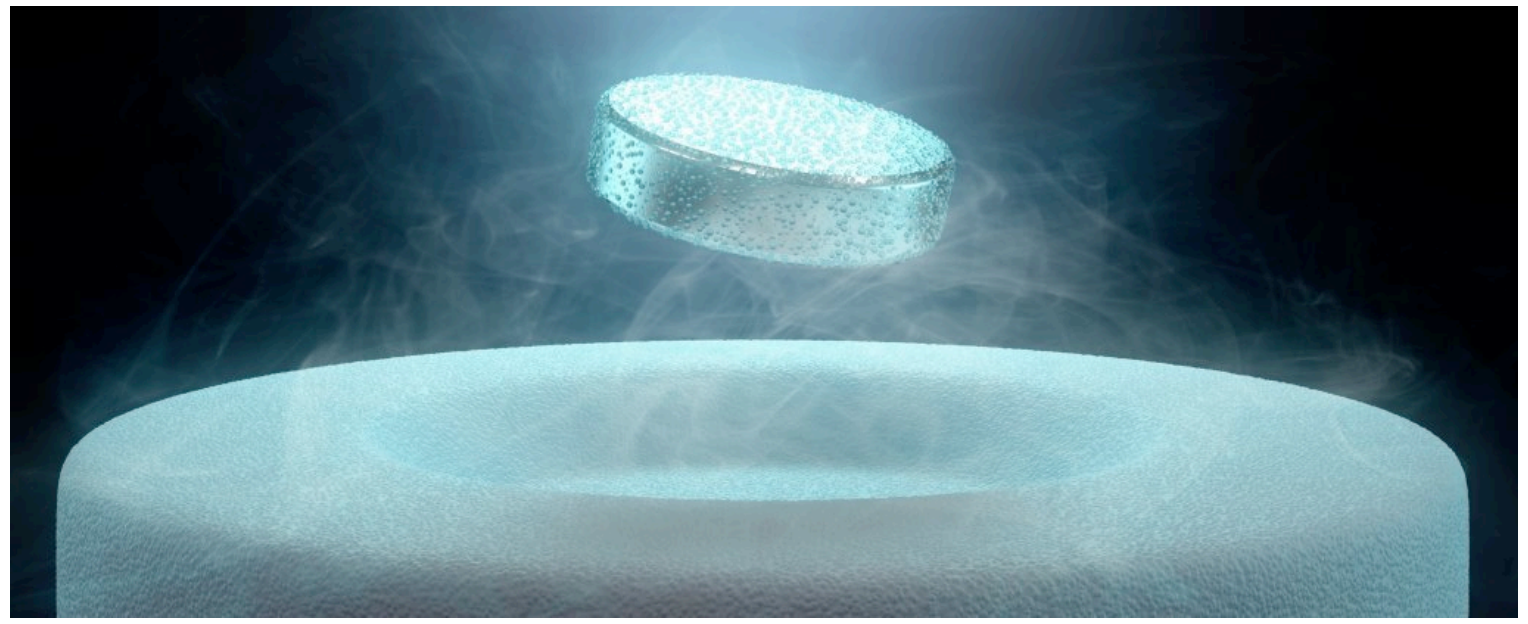
Higgs mechanism to the rescue:

- The core **mathematical trick is unintuitive**, **inspired by revelations from superconductivity research** in the 1950s!
- How long would particle physicists have take to figure it out on their own?

The space of mathematical tricks we could use is enormous

In the old electroweak theory, **we didn't know how to give particles like Z and W bosons their mass** without violating important symmetries



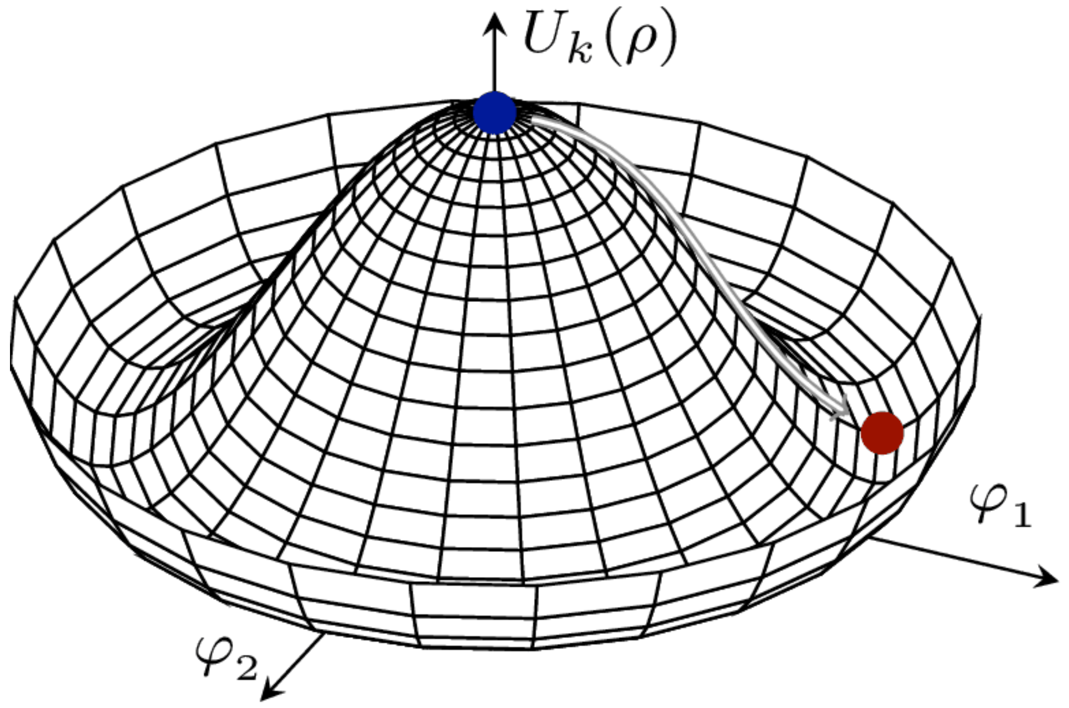


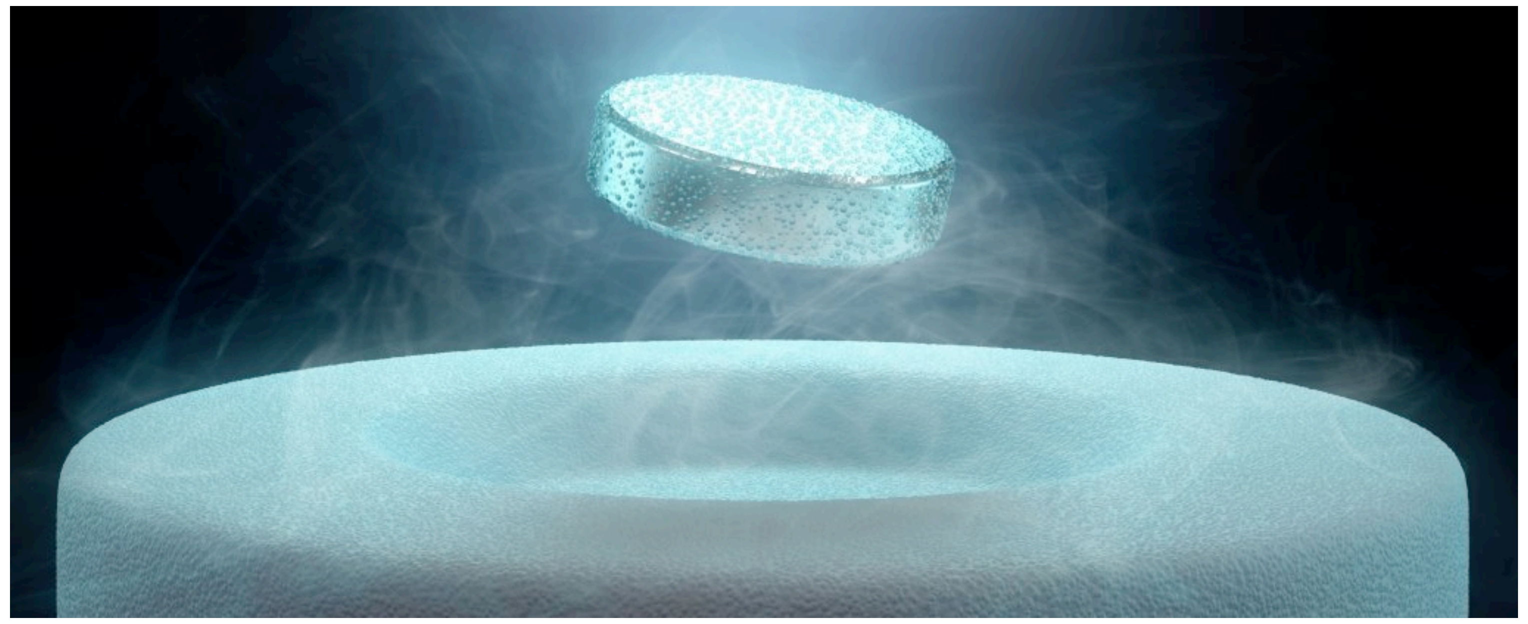
Higgs mechanism to the rescue:

- The core **mathematical trick is unintuitive**, **inspired by revelations from superconductivity research** in the 1950s!
- How long would particle physicists have take to figure it out on their own?

The space of mathematical tricks we could use is enormous

In the old electroweak theory, **we didn't know how to give particles like Z and W bosons their mass** without violating important symmetries





Higgs mechanism to the rescue:

- The core **mathematical trick is unintuitive**, **inspired by revelations from superconductivity research** in the 1950s!
- How long would particle physicists have take to figure it out on their own?

What if solutions to current open problems are in other unintuitive mathematical tricks?

Neutrinos remain massless in Standard Model, can’t explain their oscillations

Neutrinos remain massless in Standard Model, can’t explain their oscillations

A testing ground for automated theory design

ν_τ | ν_e | ν_μ

How to design your own particle physics theory ?

$$
\begin{aligned}
\mathcal{L} = & -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} \\
& + i \bar{\psi} \not{D} \psi + h.c. \\
& + \psi_i y_{ij} \psi_j \phi + h.c. \\
& + |D_\mu \phi|^2 - V(\phi)
\end{aligned}
$$

$\mathfrak{su}(2)$

$$\begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$$

Determined by choice of **fields** and **symmetries** they must obey

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

Fields sit inside symmetry group representations

$$\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$$

Adding fields: Akin to adding particles to your universe, gives you more free parameters, more flexibility. Trivial to explain data with 100s of free parameters

Adding symmetries: Adds structures and restrictions, reduces flexibility but makes the theory richer

A rich space of **discrete** mathematical choices

$\mathfrak{su}(2)$

$$\begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

$$\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$$

A rich space of **discrete** mathematical choices

$$\mathfrak{su}(3)$$

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad \lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}$$

$$\lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad \lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} \quad \lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}.$$

$$\mathfrak{su}(2)$$

$$\begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

$$\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$$

A rich space of **discrete** mathematical choices

$\mathfrak{su}(3)$

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad \lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}$$

$$\lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad \lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} \quad \lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}.$$

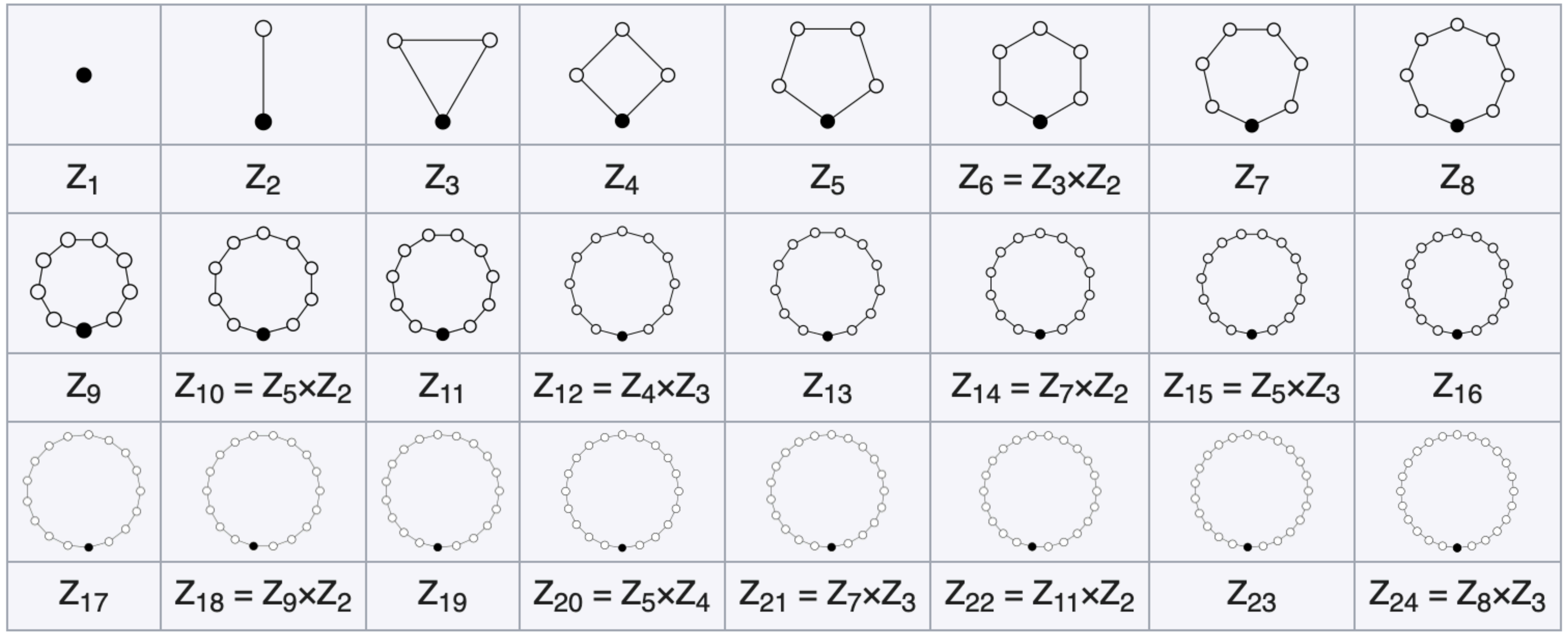
$\mathfrak{su}(2)$

$$\begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

$$\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$$

Cycle graphs up to order 24



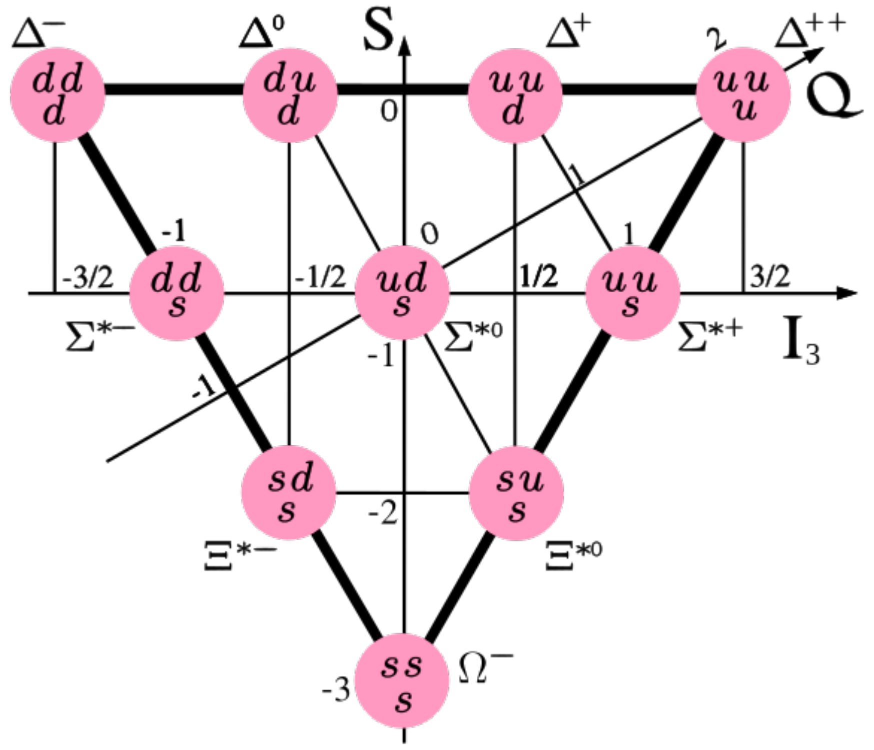
A rich space of **discrete** mathematical choices

$\mathfrak{su}(3)$

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad \lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}$$

$$\lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad \lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} \quad \lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}.$$



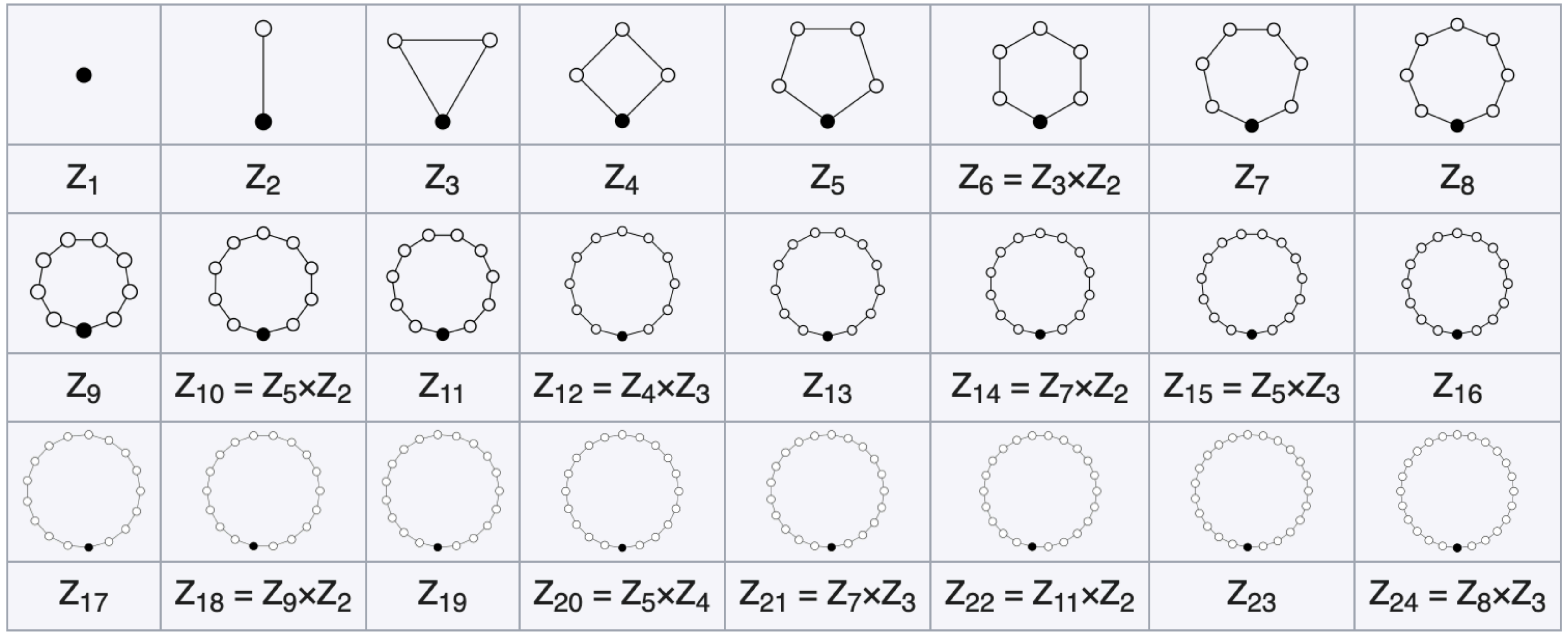
$\mathfrak{su}(2)$

$$\begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

$$\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$$

Cycle graphs up to order 24



Z_1	Z_2	Z_3	Z_4	Z_5	$Z_6 = Z_3 \times Z_2$	Z_7	Z_8
Z_9	$Z_{10} = Z_5 \times Z_2$	Z_{11}	$Z_{12} = Z_4 \times Z_3$	Z_{13}	$Z_{14} = Z_7 \times Z_2$	$Z_{15} = Z_5 \times Z_3$	Z_{16}
Z_{17}	$Z_{18} = Z_9 \times Z_2$	Z_{19}	$Z_{20} = Z_5 \times Z_4$	$Z_{21} = Z_7 \times Z_3$	$Z_{22} = Z_{11} \times Z_2$	Z_{23}	$Z_{24} = Z_8 \times Z_3$

A rich space of **discrete** mathematical choices

$\mathfrak{su}(3)$

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad \lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}$$

$$\lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad \lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} \quad \lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}.$$

Takes months to build intuition for how to use these symmetry groups and their representations

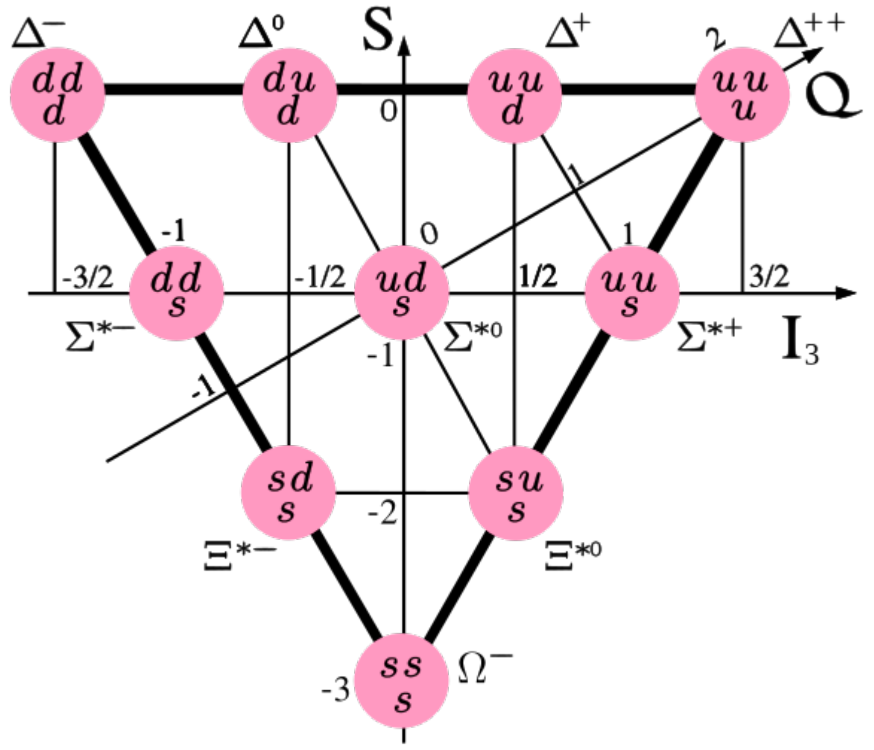
Risky: After months of study, you might hit a dead-end

$\mathfrak{su}(2)$

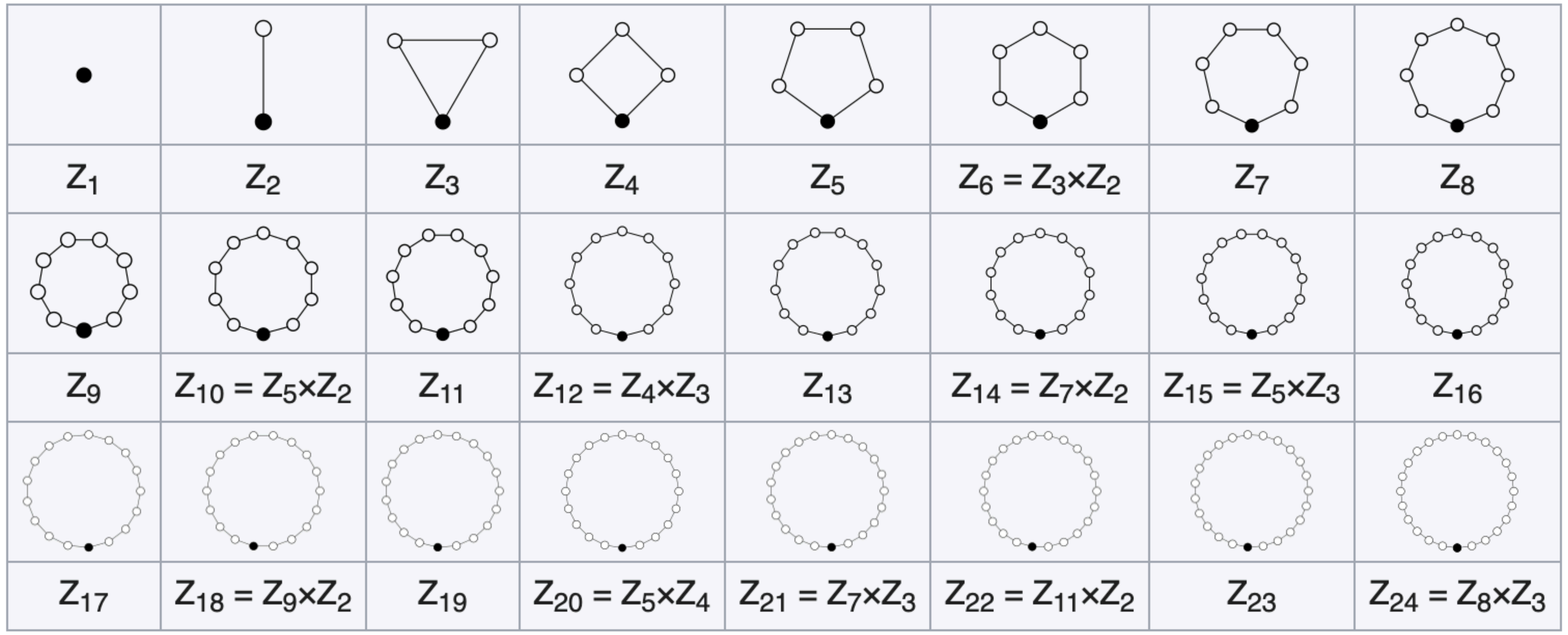
$$\begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

$$\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$$



Cycle graphs up to order 24

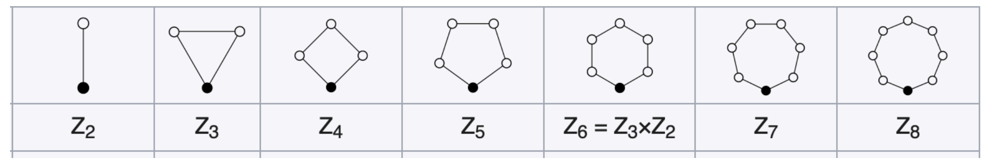


Baretz et al (inc. Ghosh): arXiv:2506.08080

Design end-to-end automated pipeline: Physics software to form building blocks

Fields & symmetries

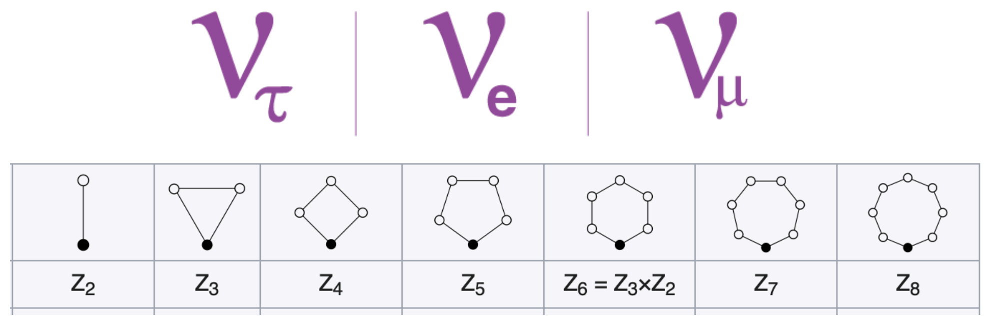
ν_τ | ν_e | ν_μ



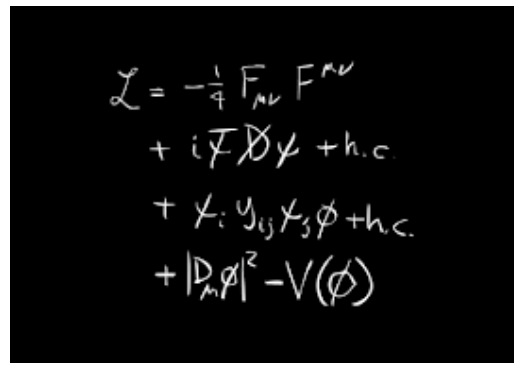
Baretz et al (inc. Ghosh): arXiv:2506.08080

Design end-to-end automated pipeline: Physics software to form building blocks

Fields & symmetries



Compute Lagrangian

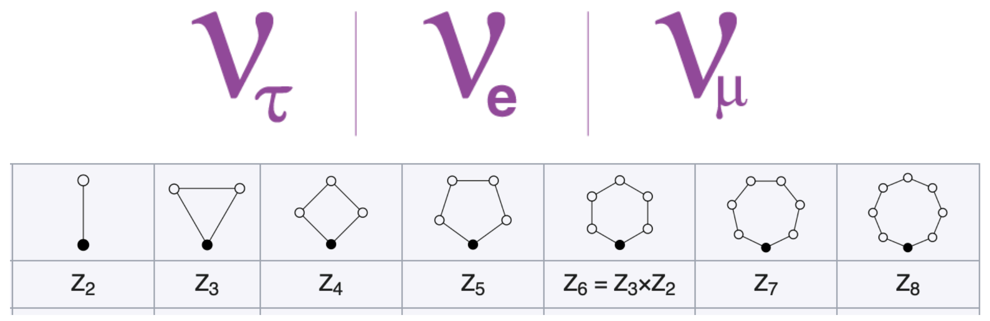


$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$$
$$+ i\bar{\psi}\not{D}\psi + h.c.$$
$$+ \psi_i y_{ij}\psi_j\phi + h.c.$$
$$+ |D_\mu\phi|^2 - V(\phi)$$

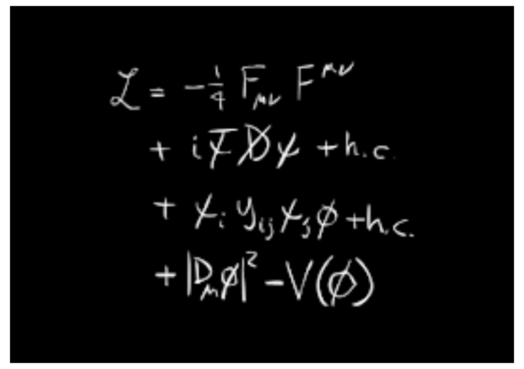
Baretz et al (inc. Ghosh): arXiv:2506.08080

Design end-to-end automated pipeline: Physics software to form building blocks

Fields & symmetries



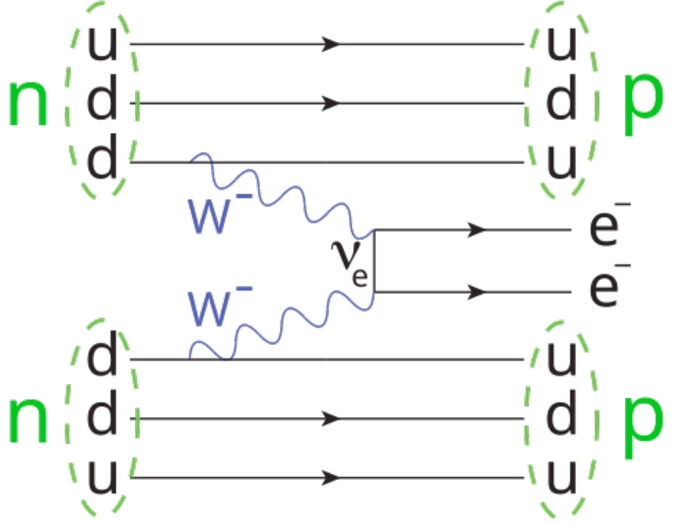
Compute Lagrangian



$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\bar{\psi}\not{D}\psi + h.c. + \psi_i y_{ij}\psi_j\phi + h.c. + |D_\mu\phi|^2 - V(\phi)$$

Compute testable predictions of theory

Check elegance of theory

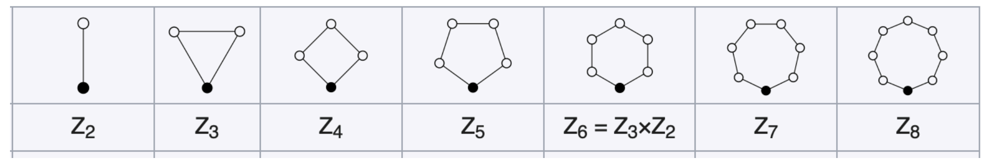


Baretz et al (inc. Ghosh): arXiv:2506.08080

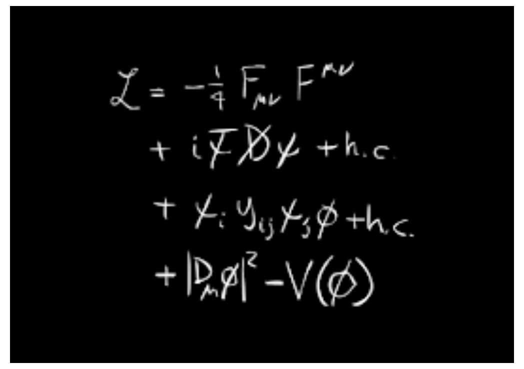
Design end-to-end automated pipeline: Physics software to form building blocks

Fields & symmetries

ν_τ | ν_e | ν_μ

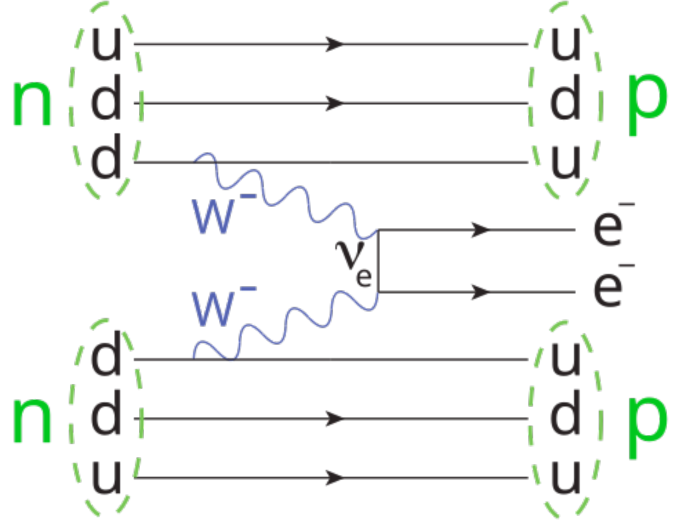


Compute Lagrangian

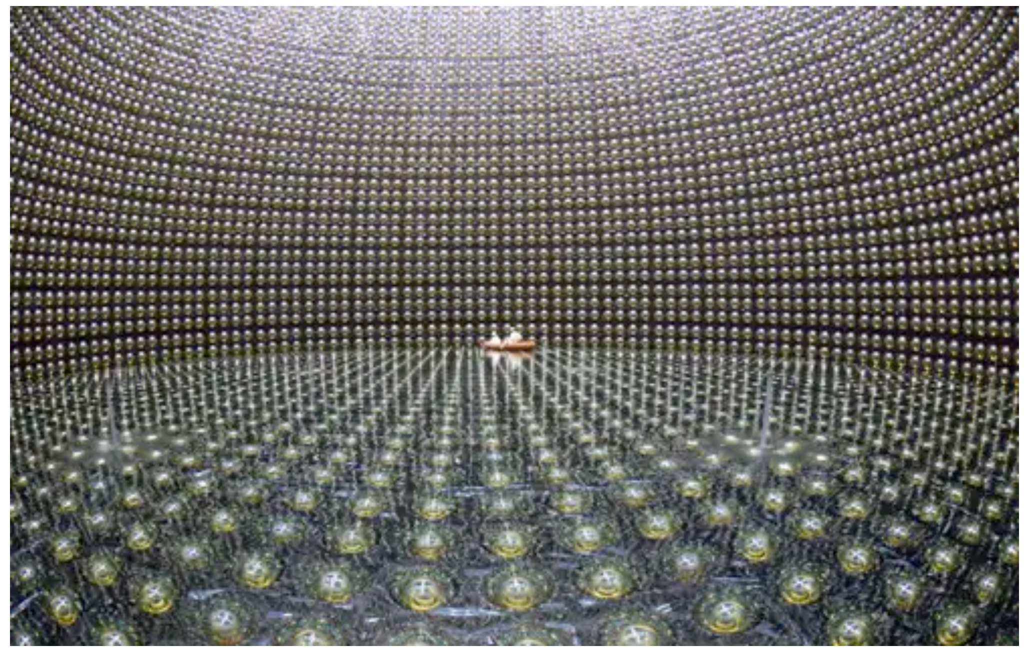


$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + i\bar{\psi} \not{D} \psi + h.c. + \psi_i y_{ij} \psi_j \phi + h.c. + |D_\mu \phi|^2 - V(\phi)$$

Compute testable predictions of theory
Check elegance of theory



Compare to observed data

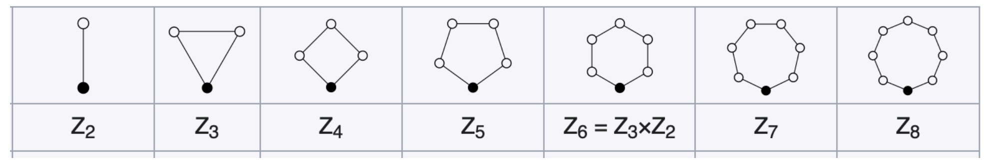


Baretz et al (inc. Ghosh): arXiv:2506.08080

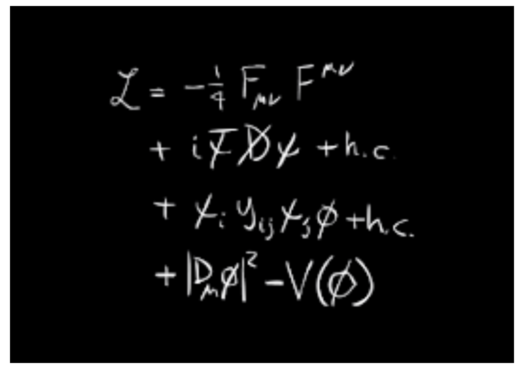
Design end-to-end automated pipeline: Physics software to form building blocks

Fields & symmetries

ν_τ | ν_e | ν_μ

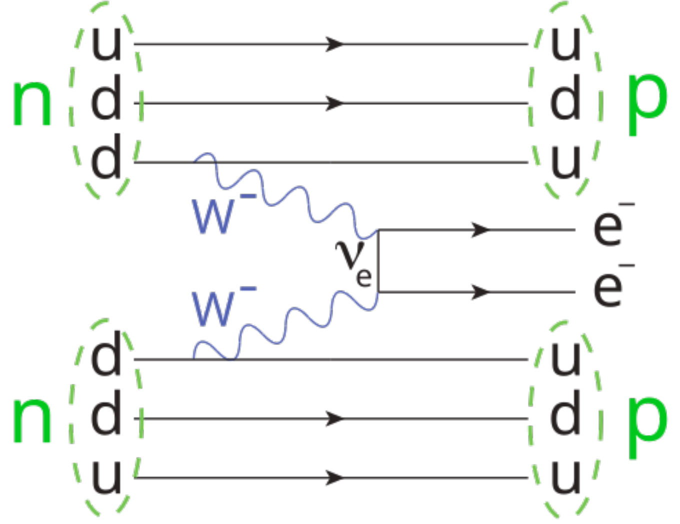


Compute Lagrangian

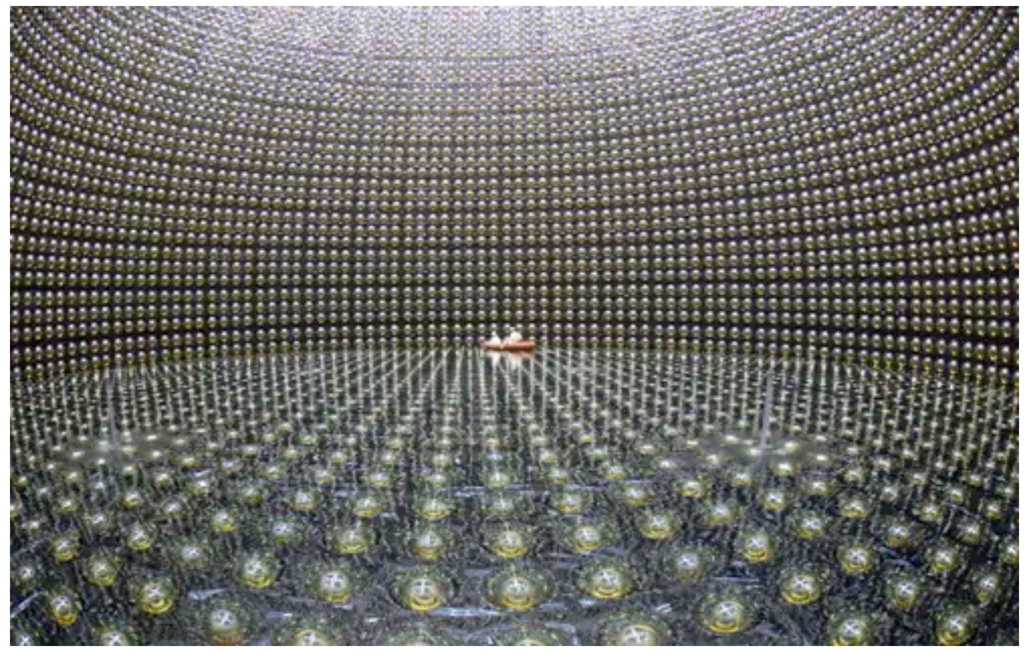


$$\mathcal{L} = -\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\bar{\psi}\not{D}\psi + h.c. + \psi_i y_{ij}\psi_j\phi + h.c. + |D_\mu\phi|^2 - V(\phi)$$

Compute testable predictions of theory
Check elegance of theory



Compare to observed data



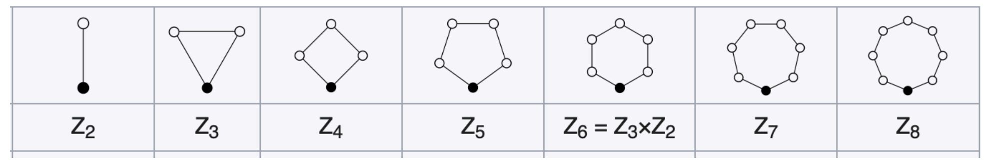
Evaluate fit, test fine-tuning of free parameters

Baretz et al (inc. Ghosh): arXiv:2506.08080

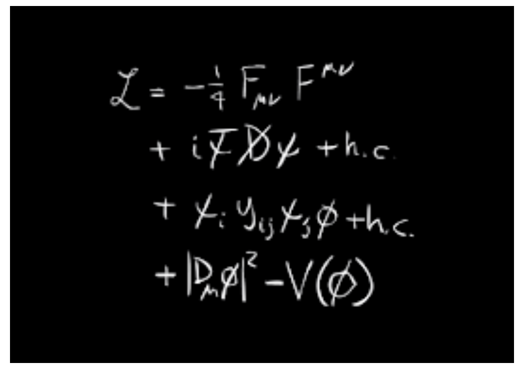
Design end-to-end automated pipeline: Physics software to form building blocks

Fields & symmetries

ν_τ | ν_e | ν_μ



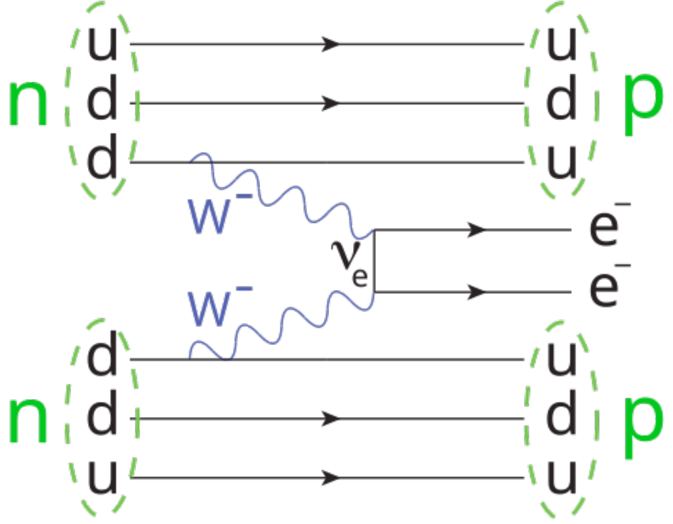
Compute Lagrangian



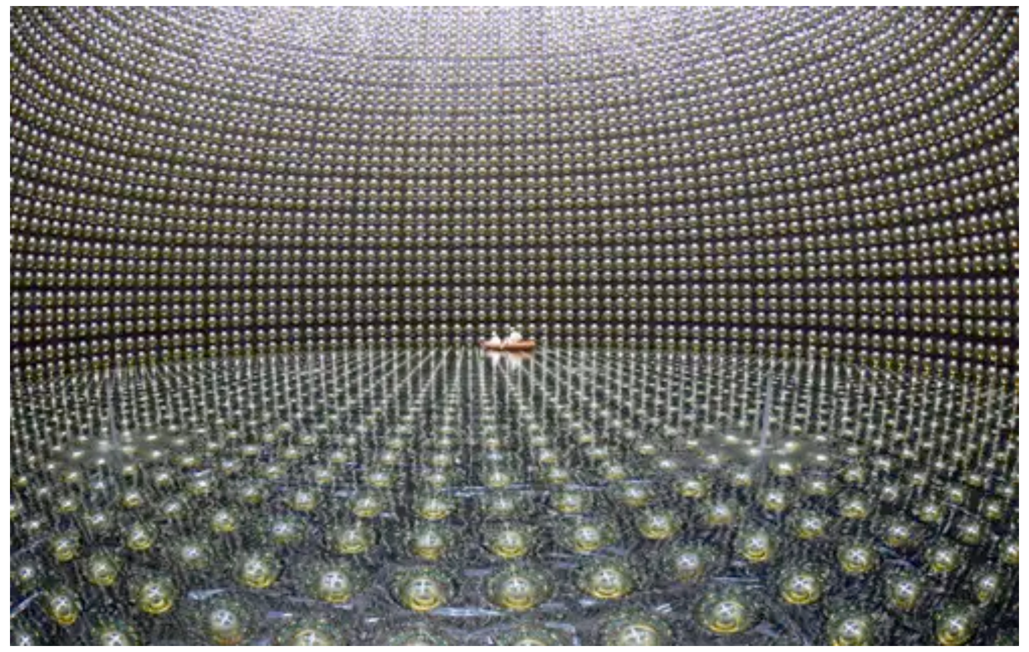
$$\begin{aligned}\mathcal{L} = &-\tfrac{1}{4} F_{\mu\nu} F^{\mu\nu} \\ &+ i\bar{\psi}\not{D}\psi + h.c. \\ &+ \psi_i y_{ij} \psi_j \phi + h.c. \\ &+ |D_\mu \phi|^2 - V(\phi)\end{aligned}$$

Compute testable predictions of theory

Check elegance of theory



Compare to observed data

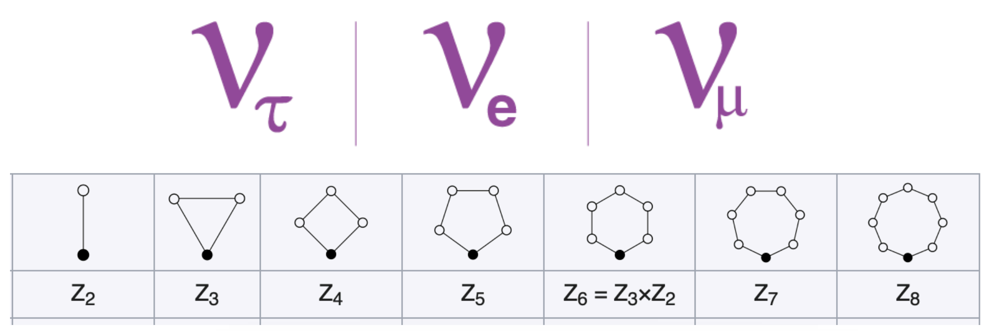


Evaluate fit, test fine-tuning of free parameters

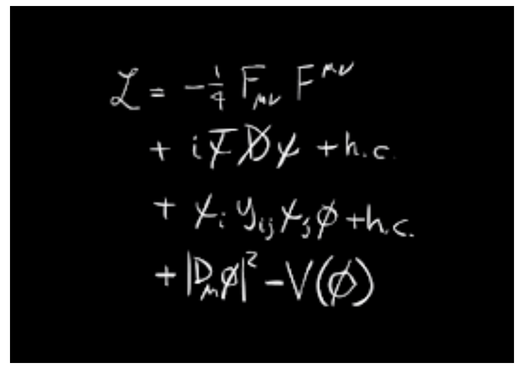
Baretz et al (inc. Ghosh): arXiv:2506.08080

Design end-to-end automated pipeline: Physics software to form building blocks

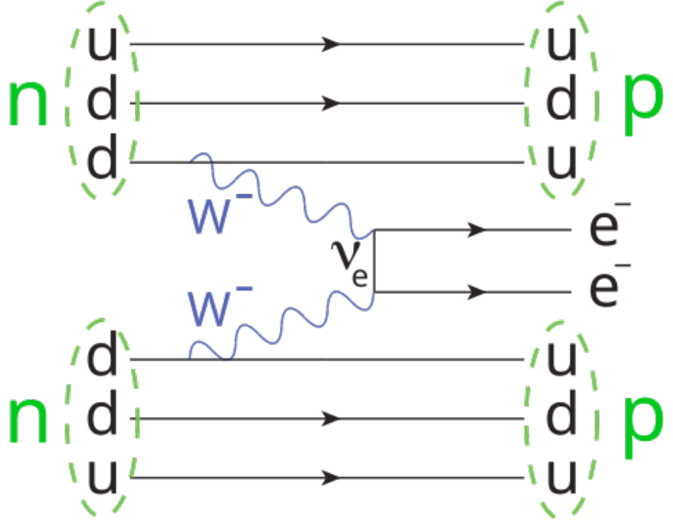
Fields & symmetries



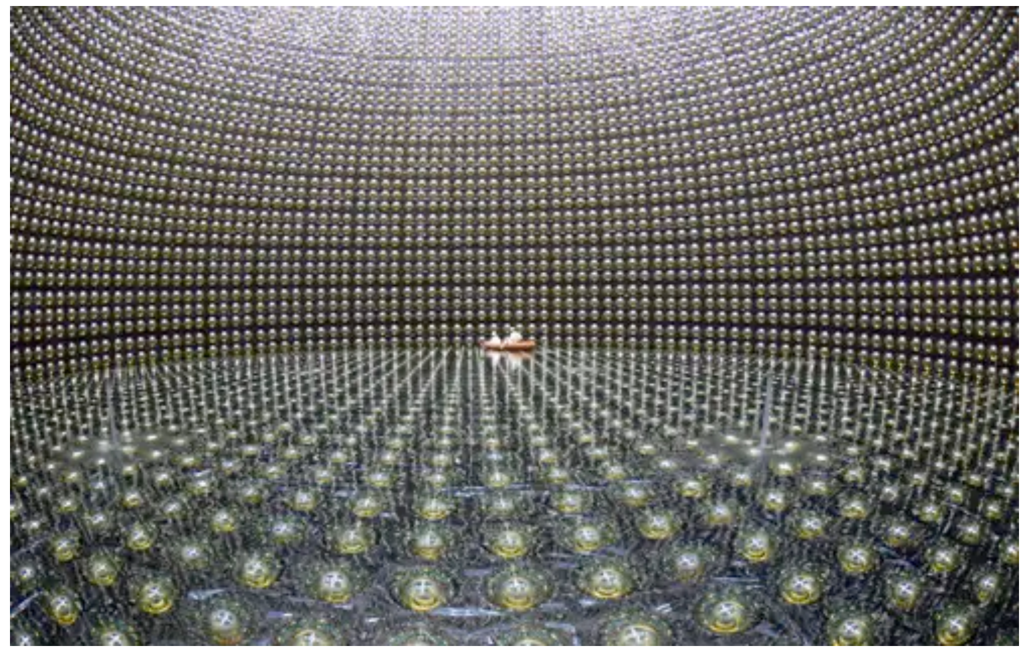
Compute Lagrangian



Compute testable predictions of theory
Check elegance of theory



Compare to observed data



Evaluate fit, test fine-tuning of free parameters

FlavorBuilder 0.3.1

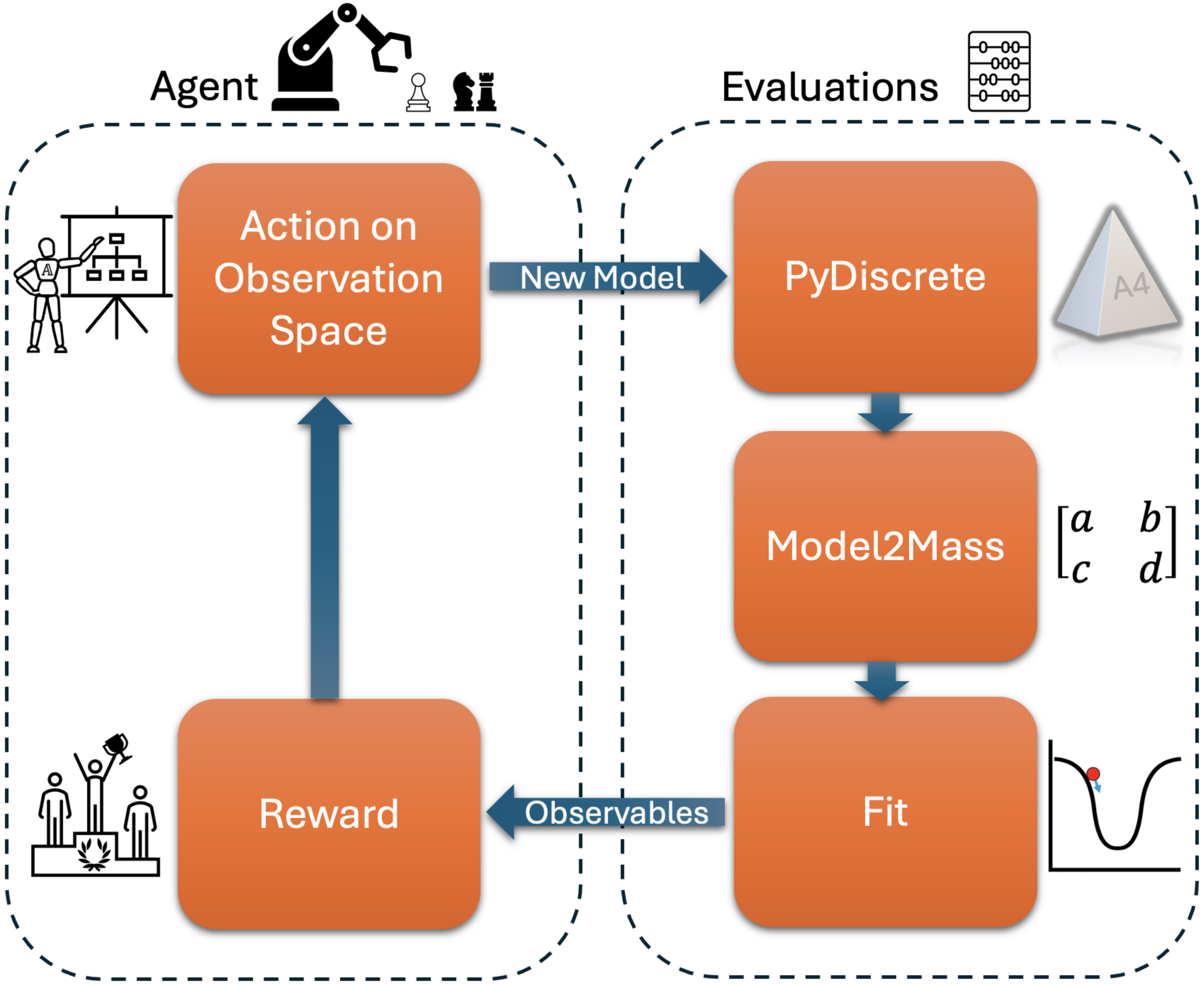
`pip install FlavorBuilder`

Knowledge from advisor's notes, slow Mathematica code
formalised into AI-ready workflow by Victoria Knapp-Perez

Victoria Kapp Perez

Baretz et al (inc. Ghosh): arXiv:2506.08080

Reinforcement-learning-assisted theory design

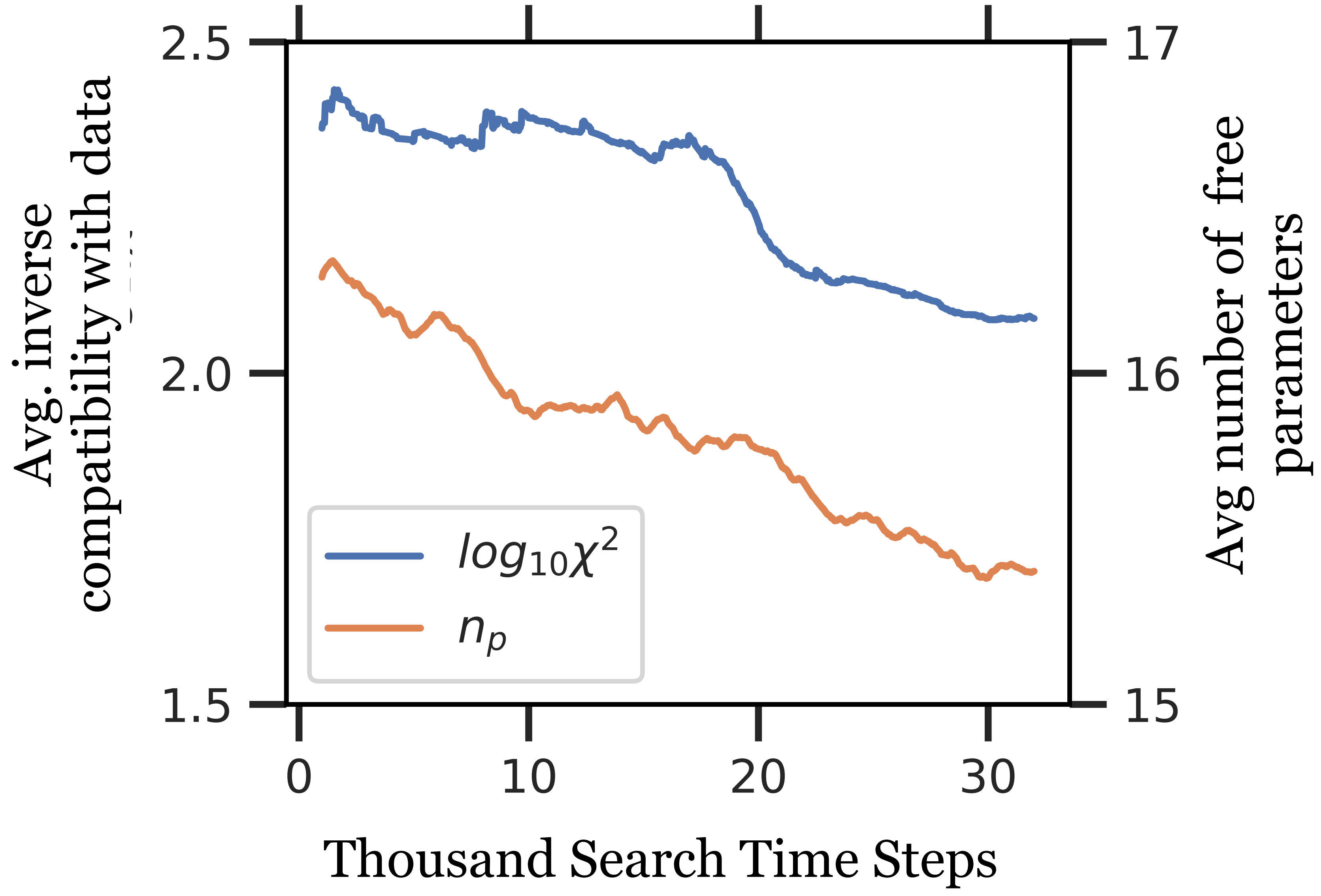


- RL agent proposes fields & symmetries
- Runs math software to calculate Lagrangian
- Run physics software to find predictions
- Compare to particle experiment data
  - Fit free parameters to data
  - Evaluate compatibility
  - Evaluate elegance of theory

Baretz et al (inc. Ghosh): arXiv:2506.08080

Agent learns to design good theories… too many!

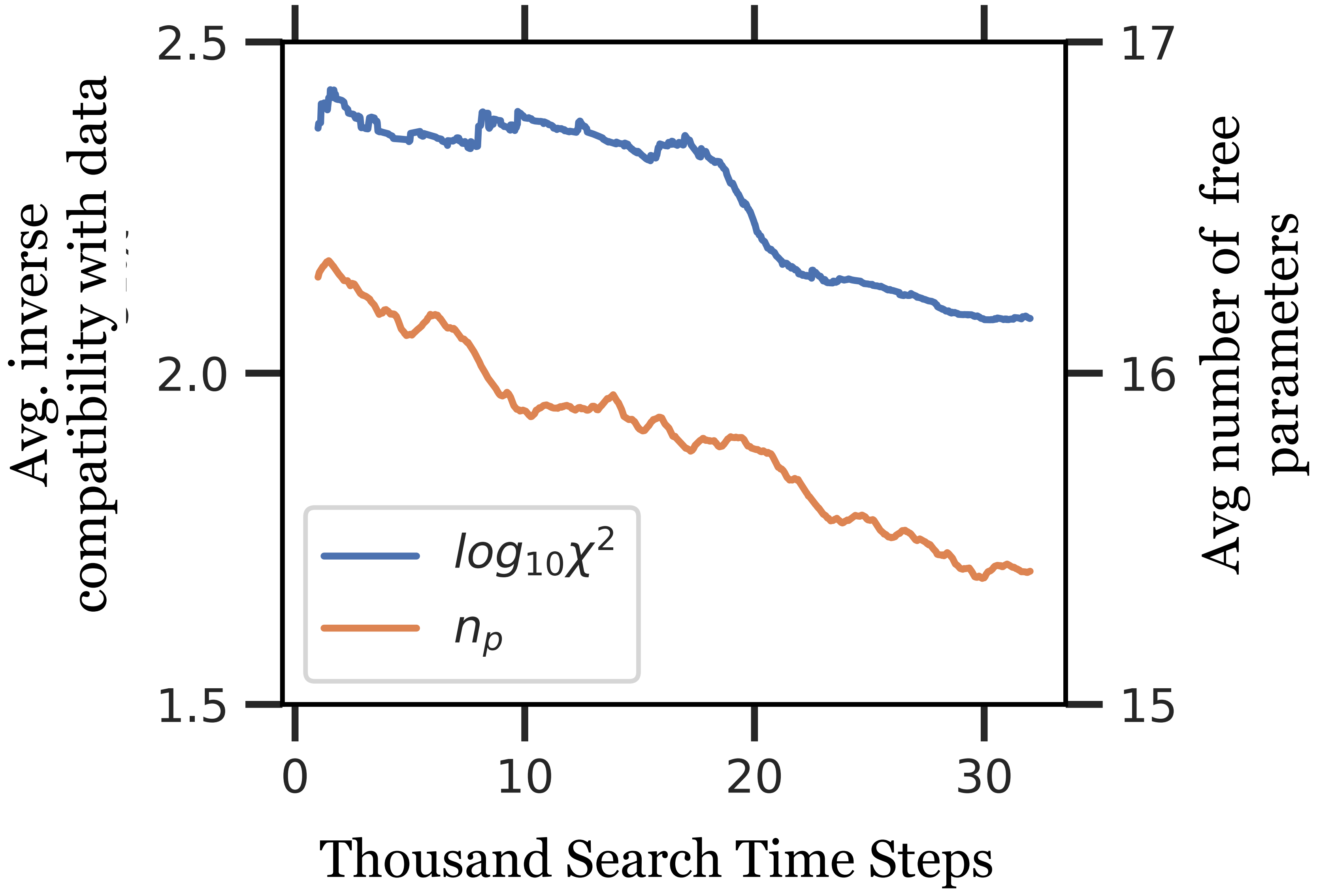
Agent learns to design good theories… too many!

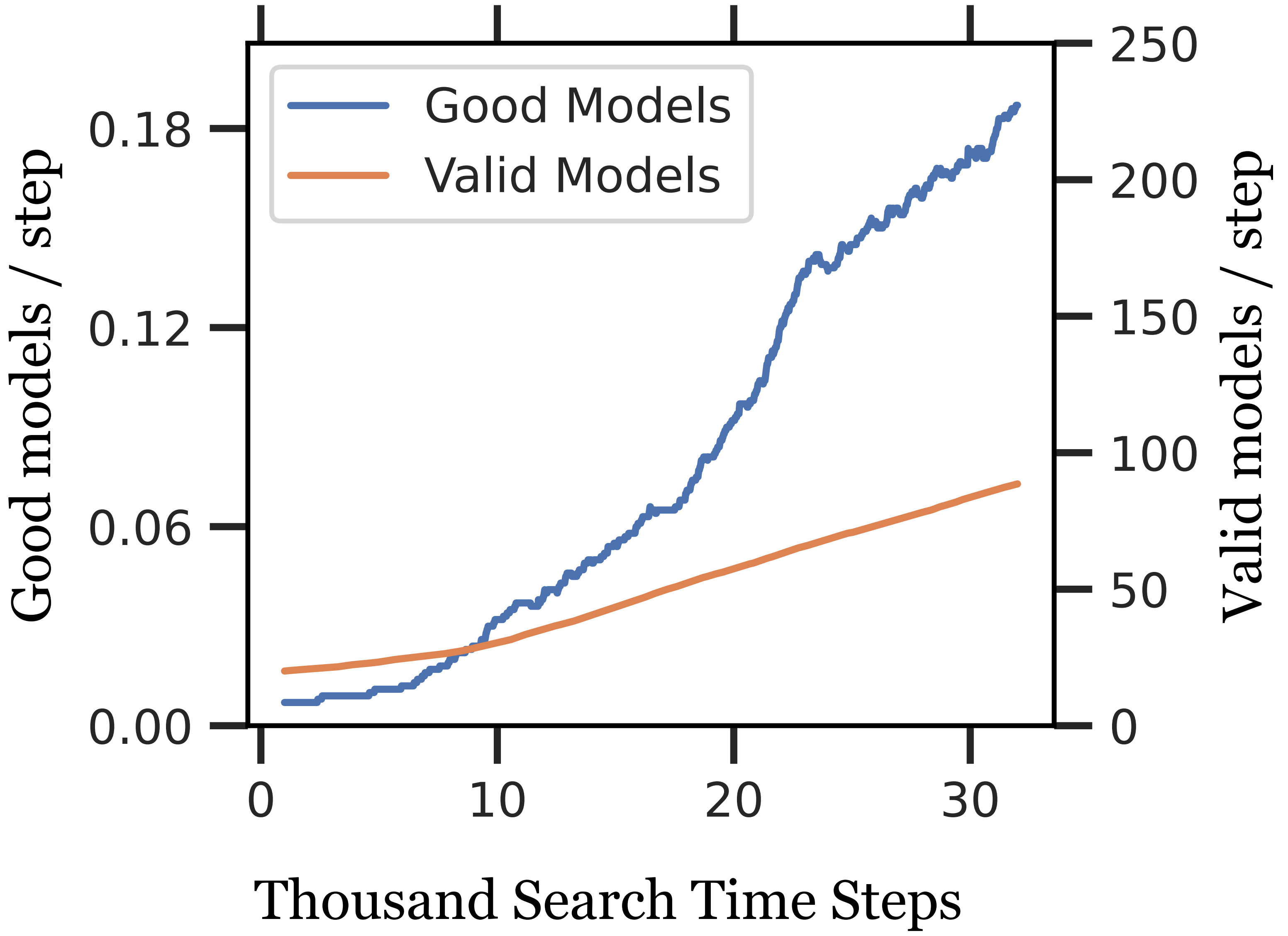


Searching in the most well-studied space of theories: $A_4 \times Z_4$

Baretz et al (inc. Ghosh): arXiv:2506.08080

Agent learns to design good theories… too many!

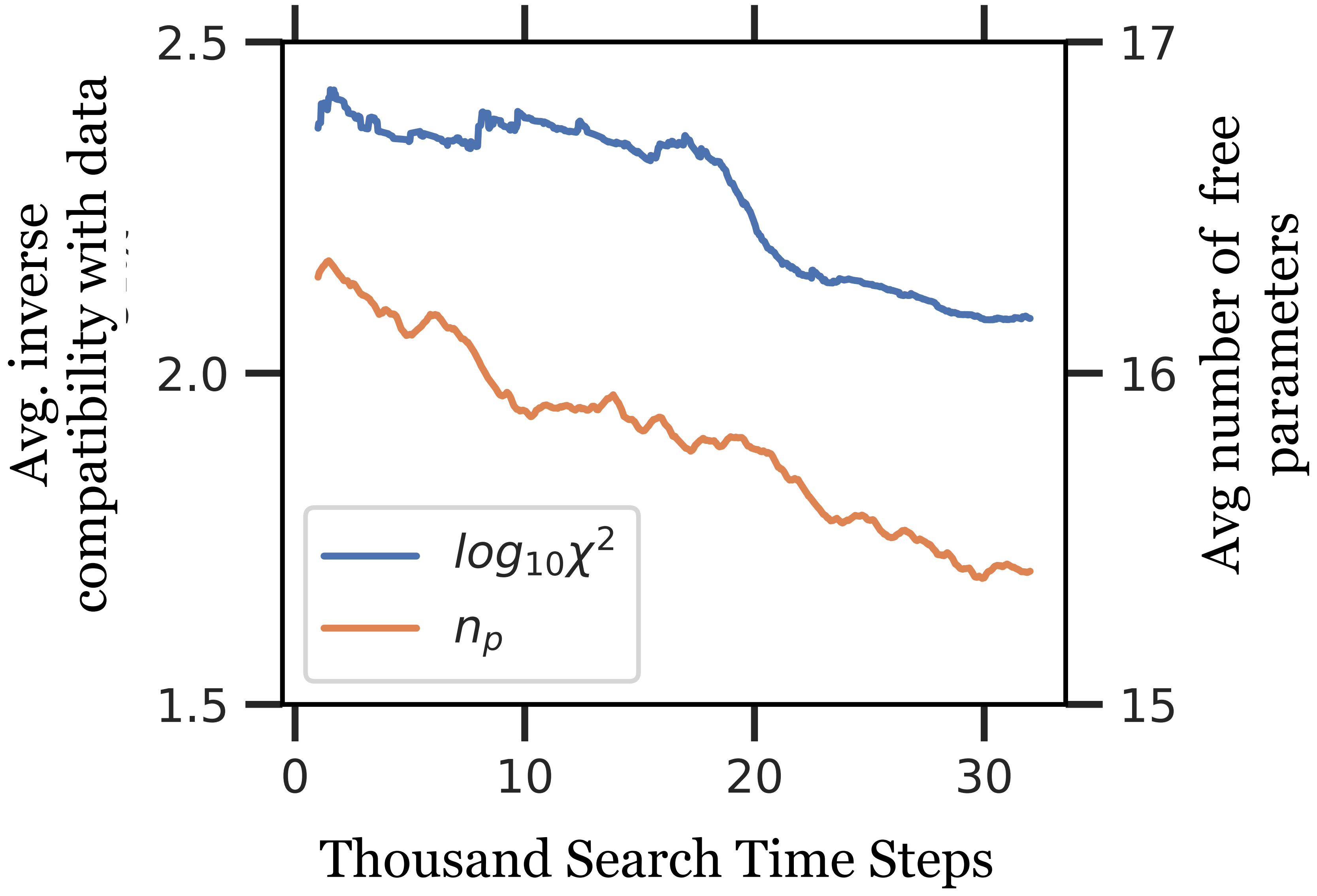


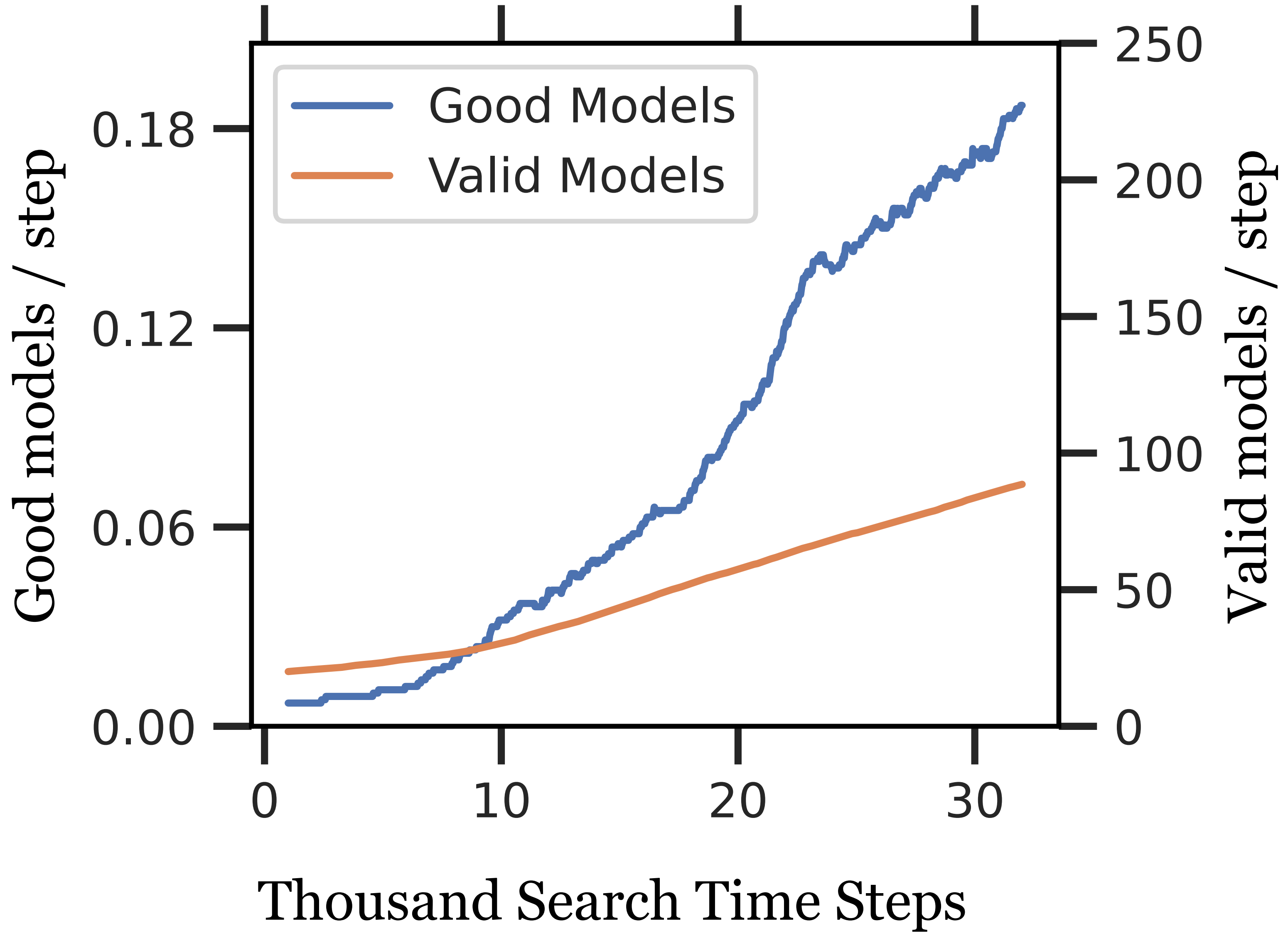


Searching in the most well-studied space of theories: $A_4 \times Z_4$

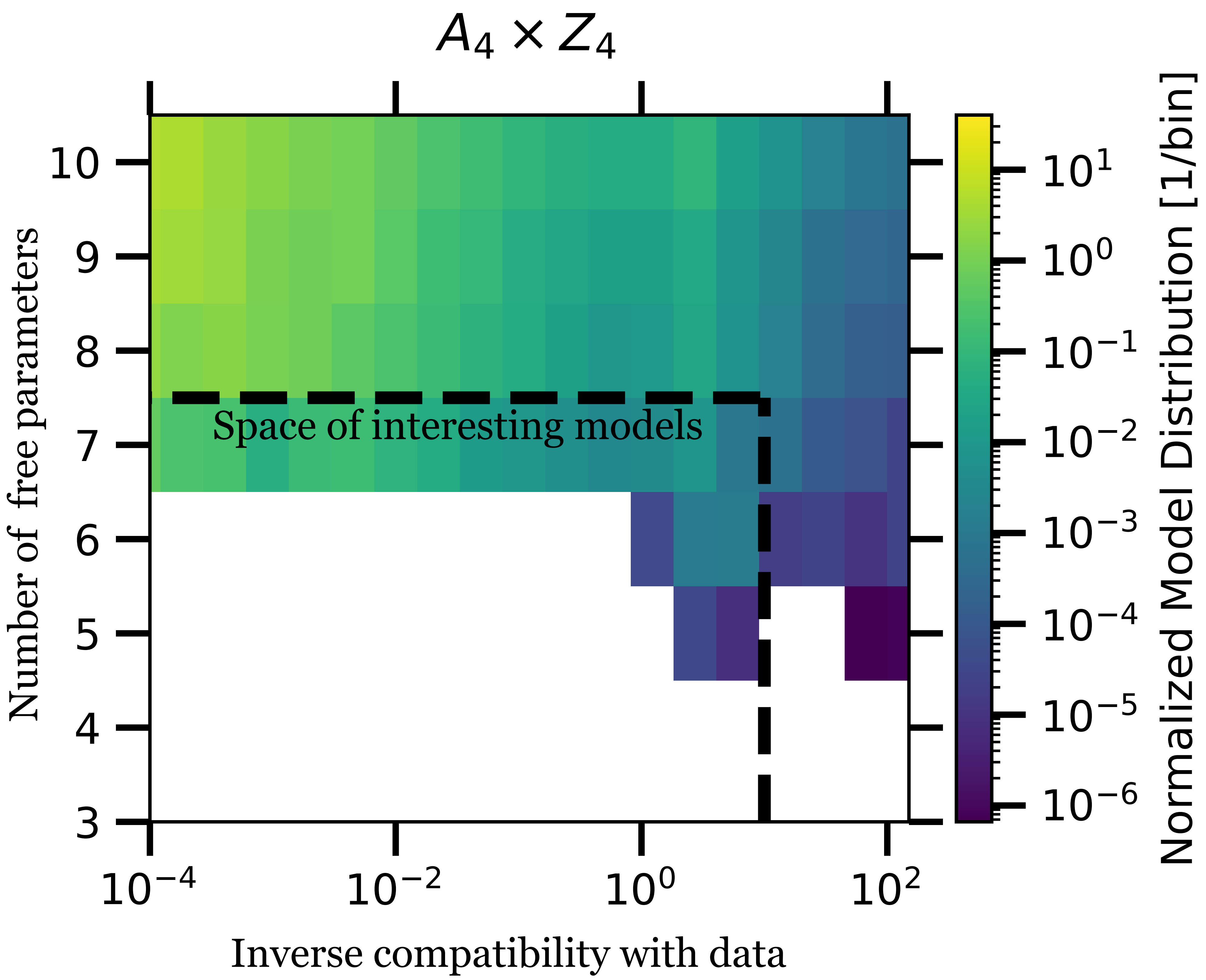
Baretz et al (inc. Ghosh): arXiv:2506.08080

Agent learns to design good theories... too many!





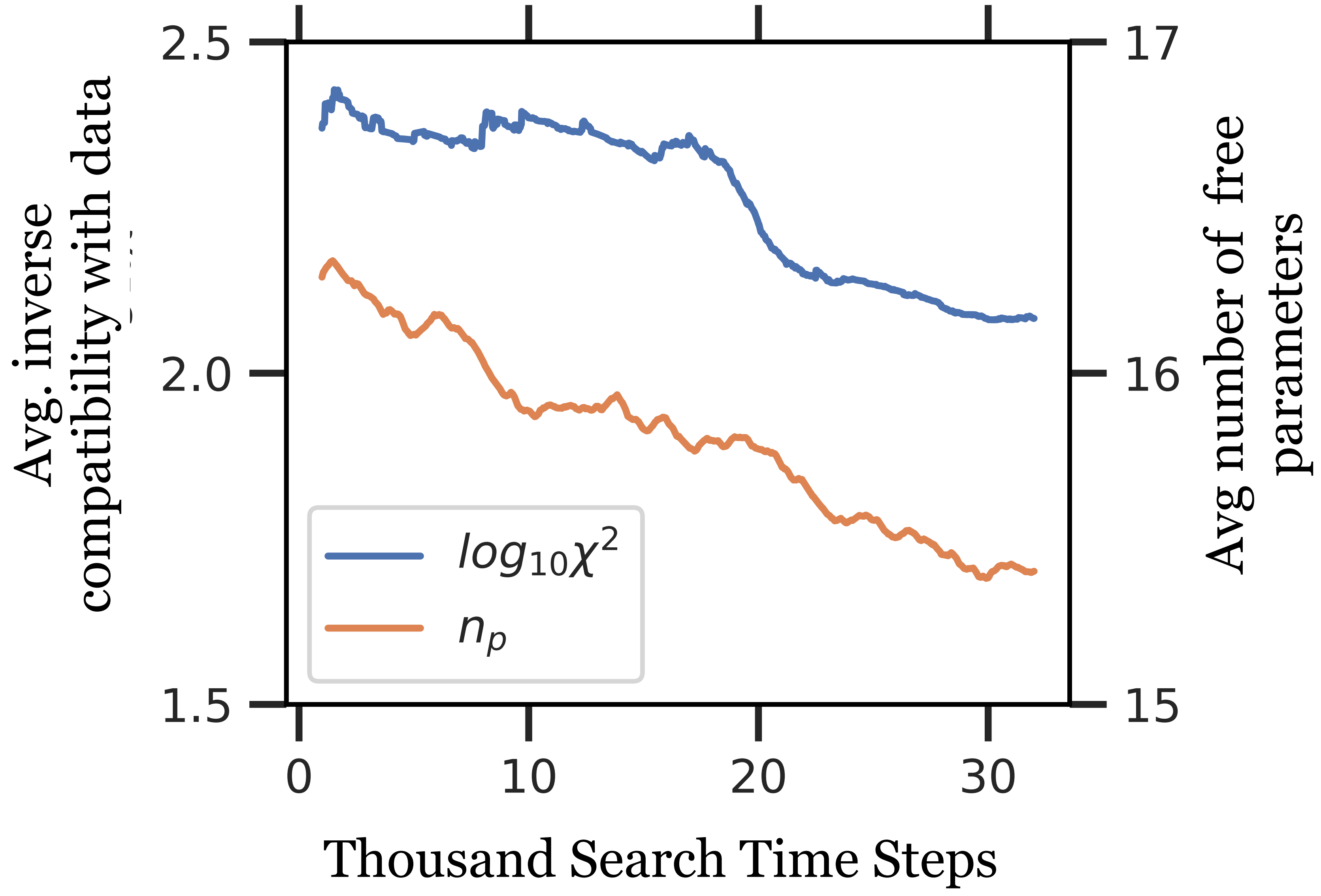
Definition of 'good models'

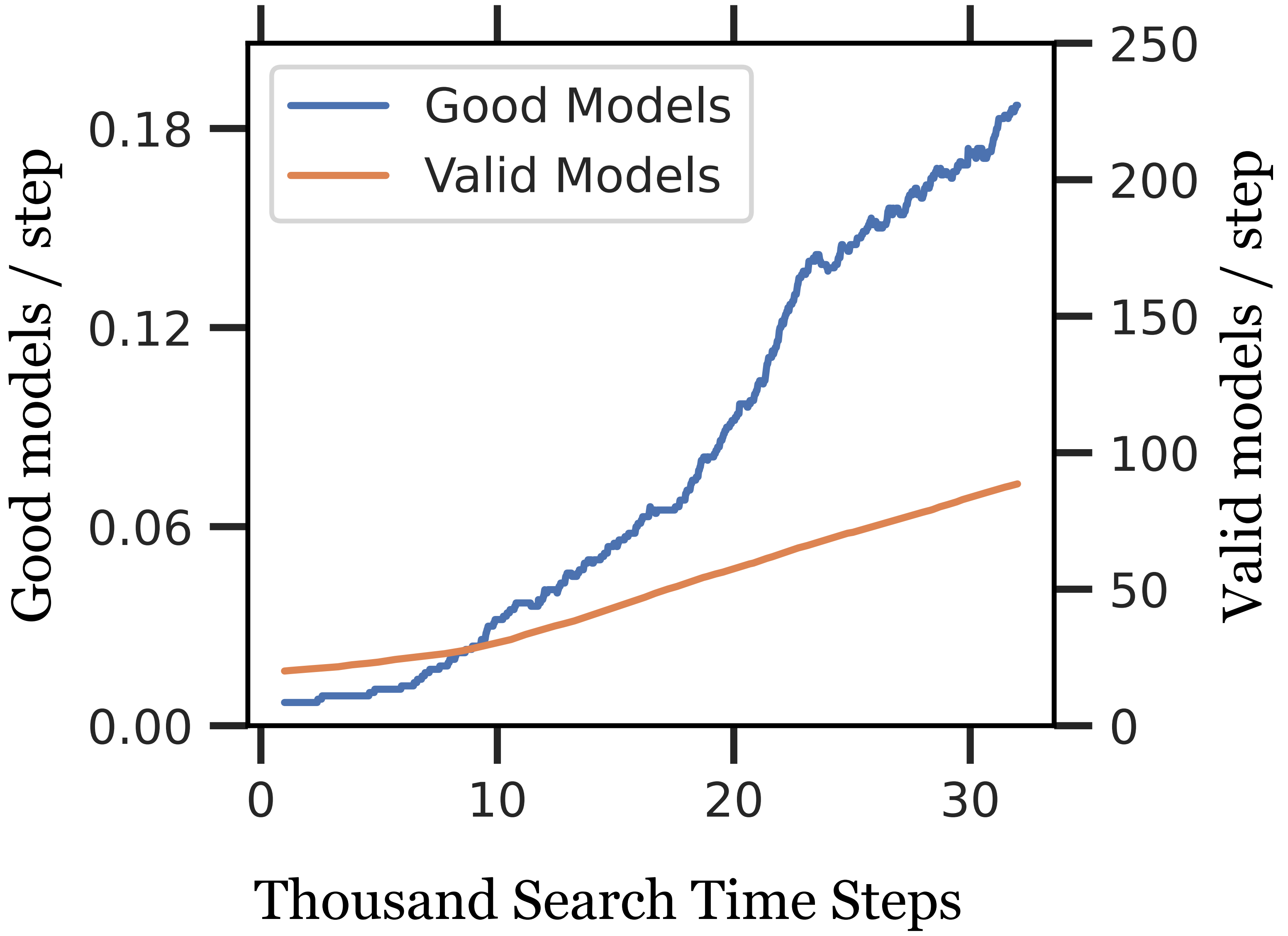


Searching in the most well-studied space of theories: $A_4 \times Z_4$

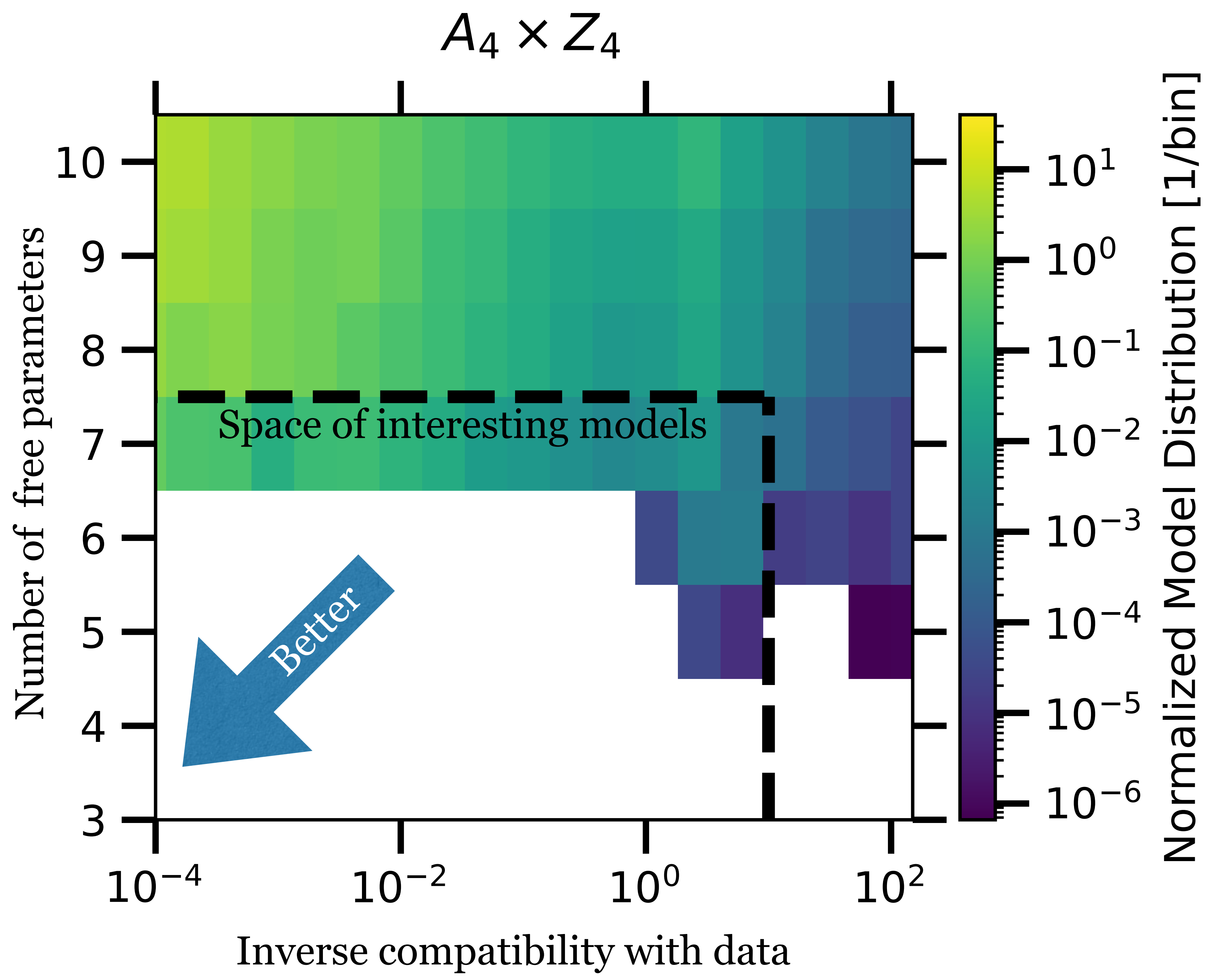
Re-discovers models of same form as human-designed models

Agent learns to design good theories… too many!





Definition of ‘good models’



Searching in the most well-studied space of theories: $A_4 \times Z_4$

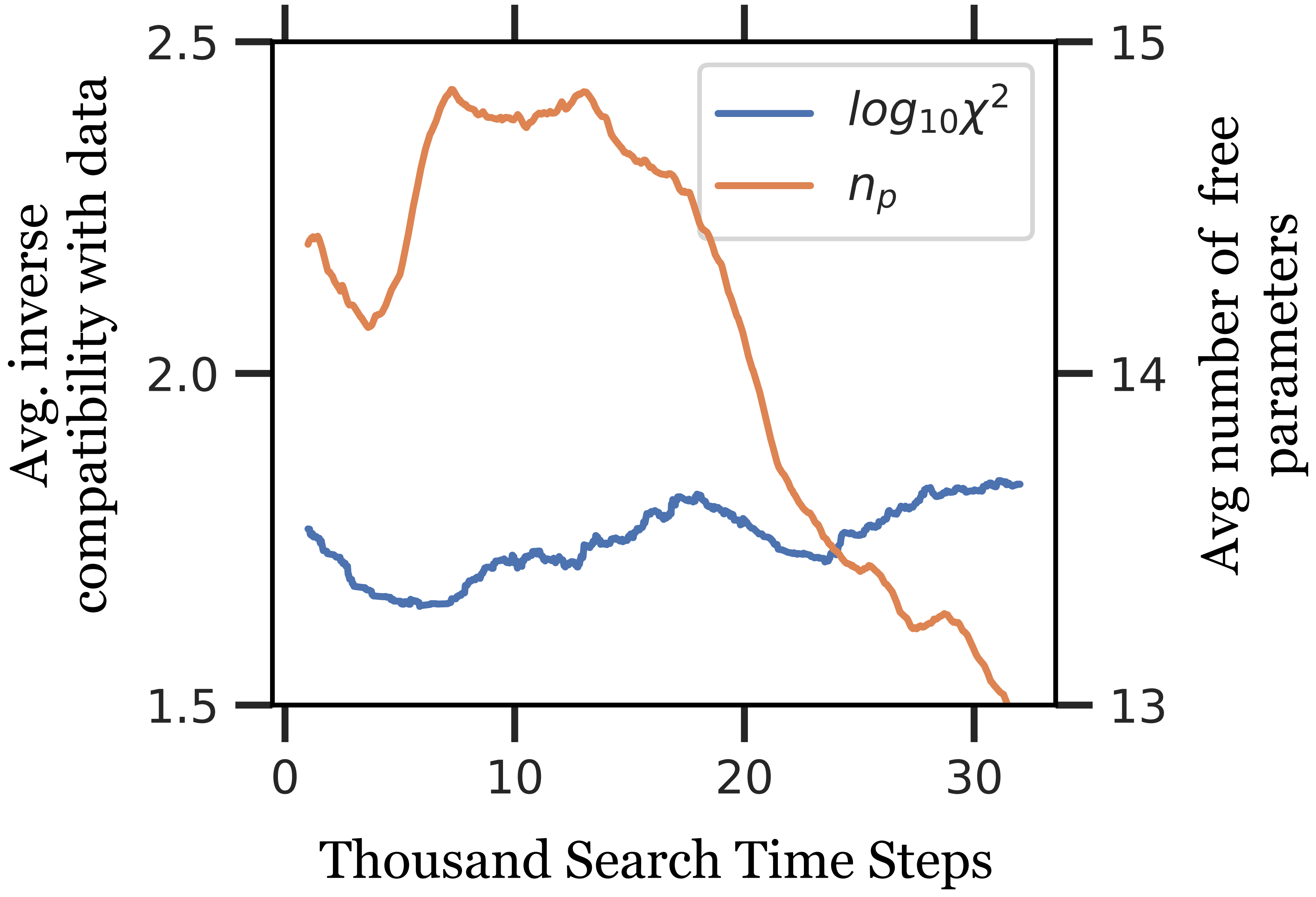
Re-discovers models of same form as human-designed models

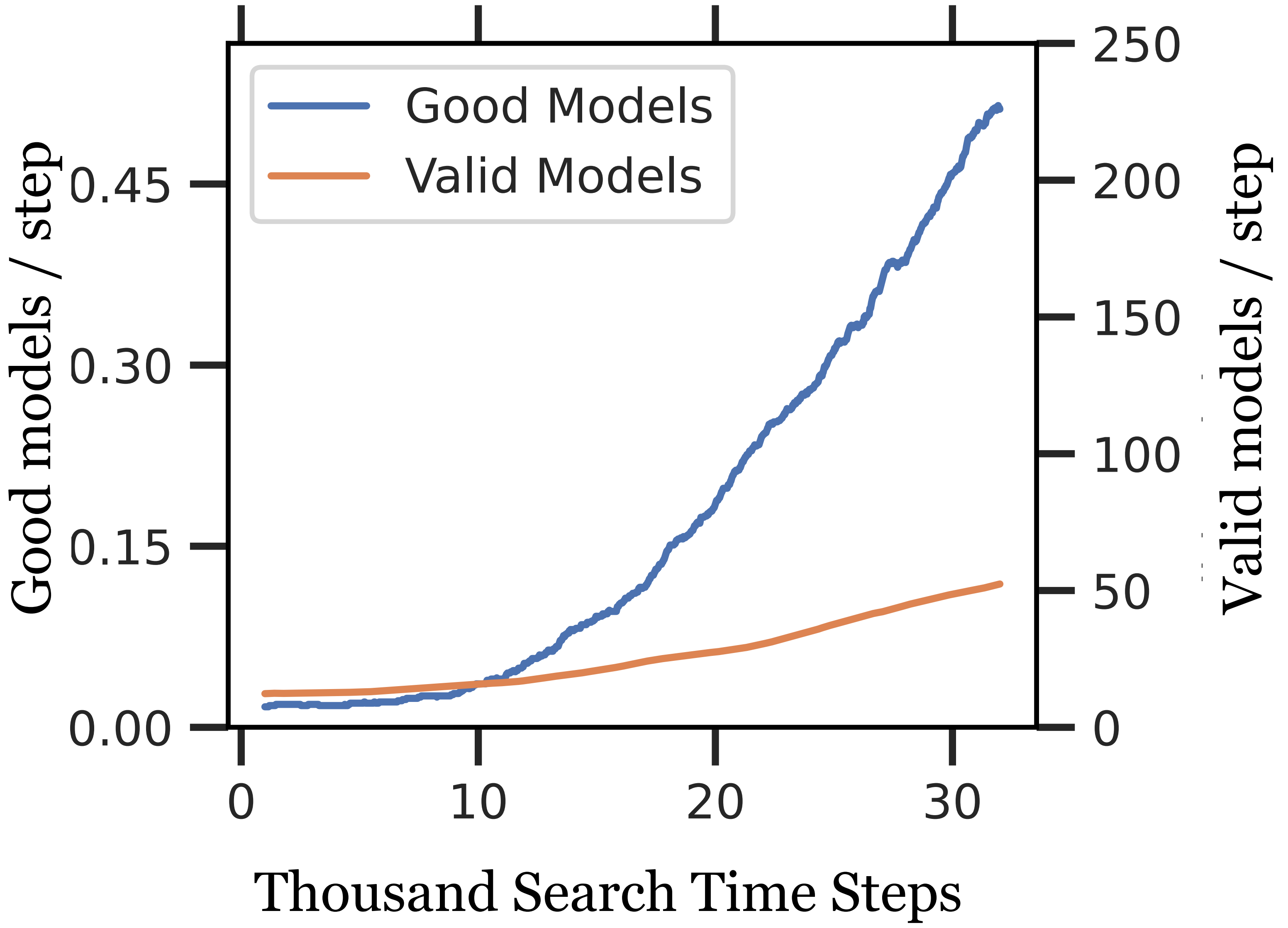
How about exploring new symmetry never previously studied by physicists?

$$T_{19} = \mathbb{Z}_{19} \rtimes \mathbb{Z}_3$$

Baretz et al (inc. Ghosh): arXiv:2506.08080

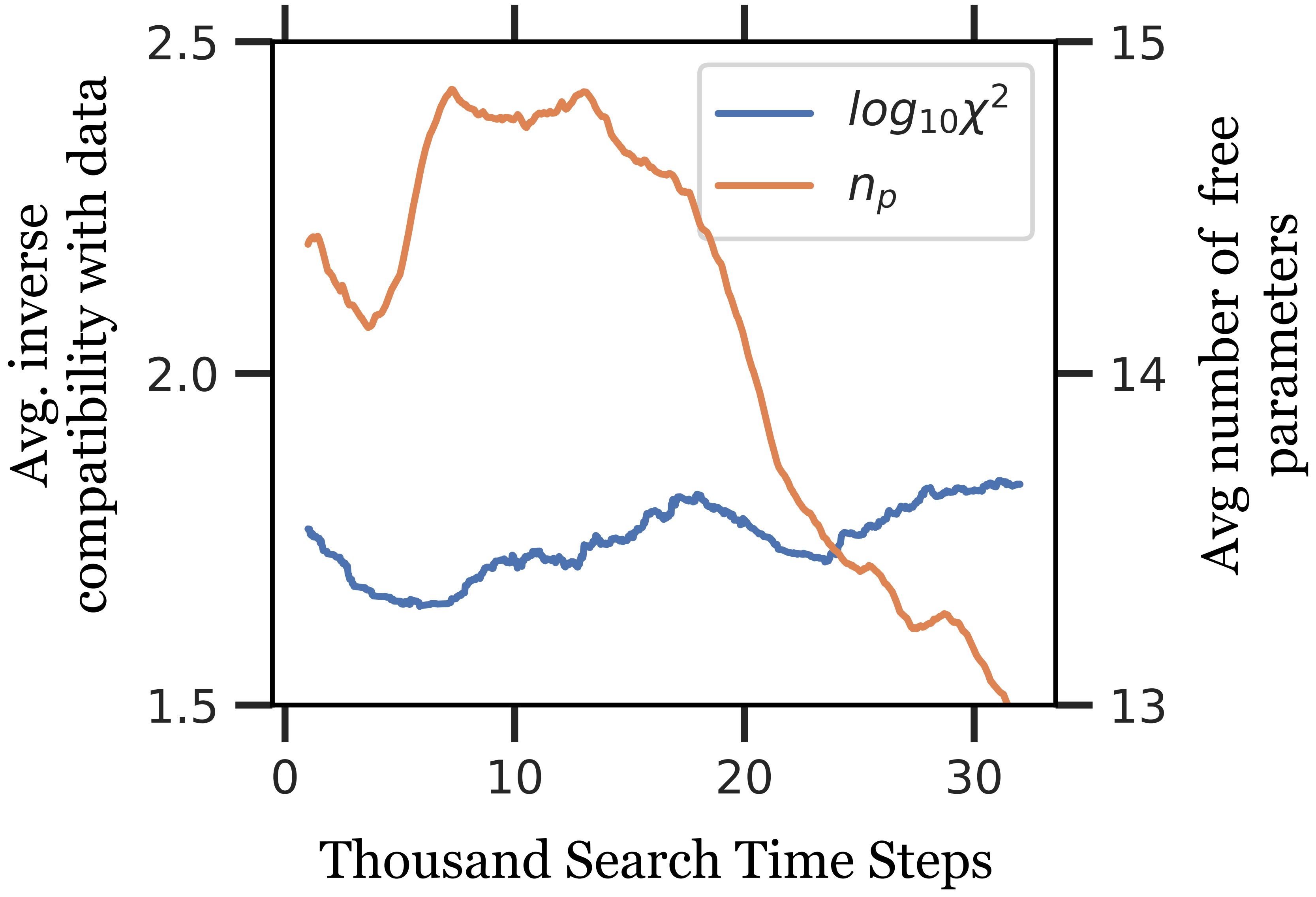
A new symmetry never before studied for neutrino model building

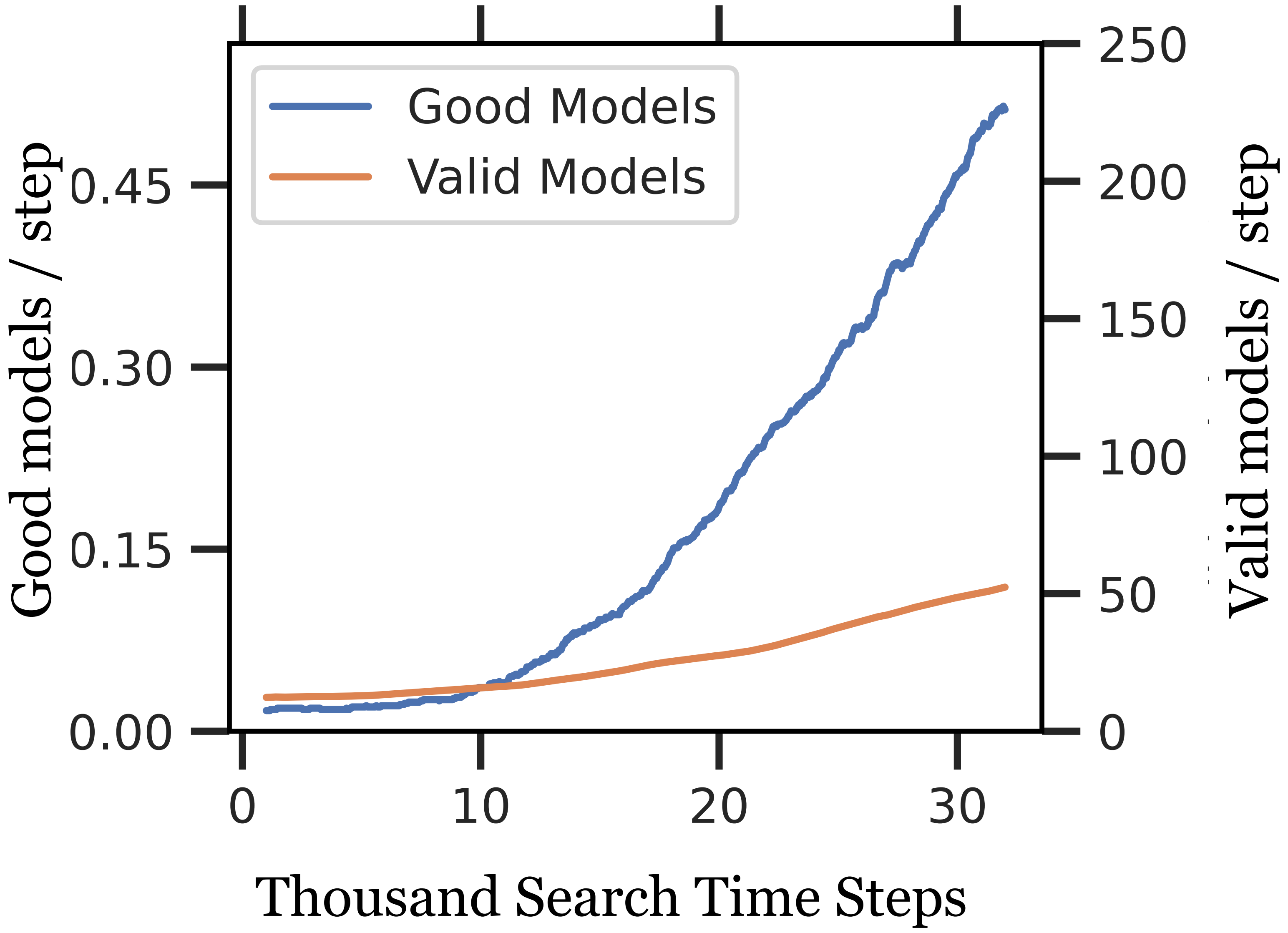




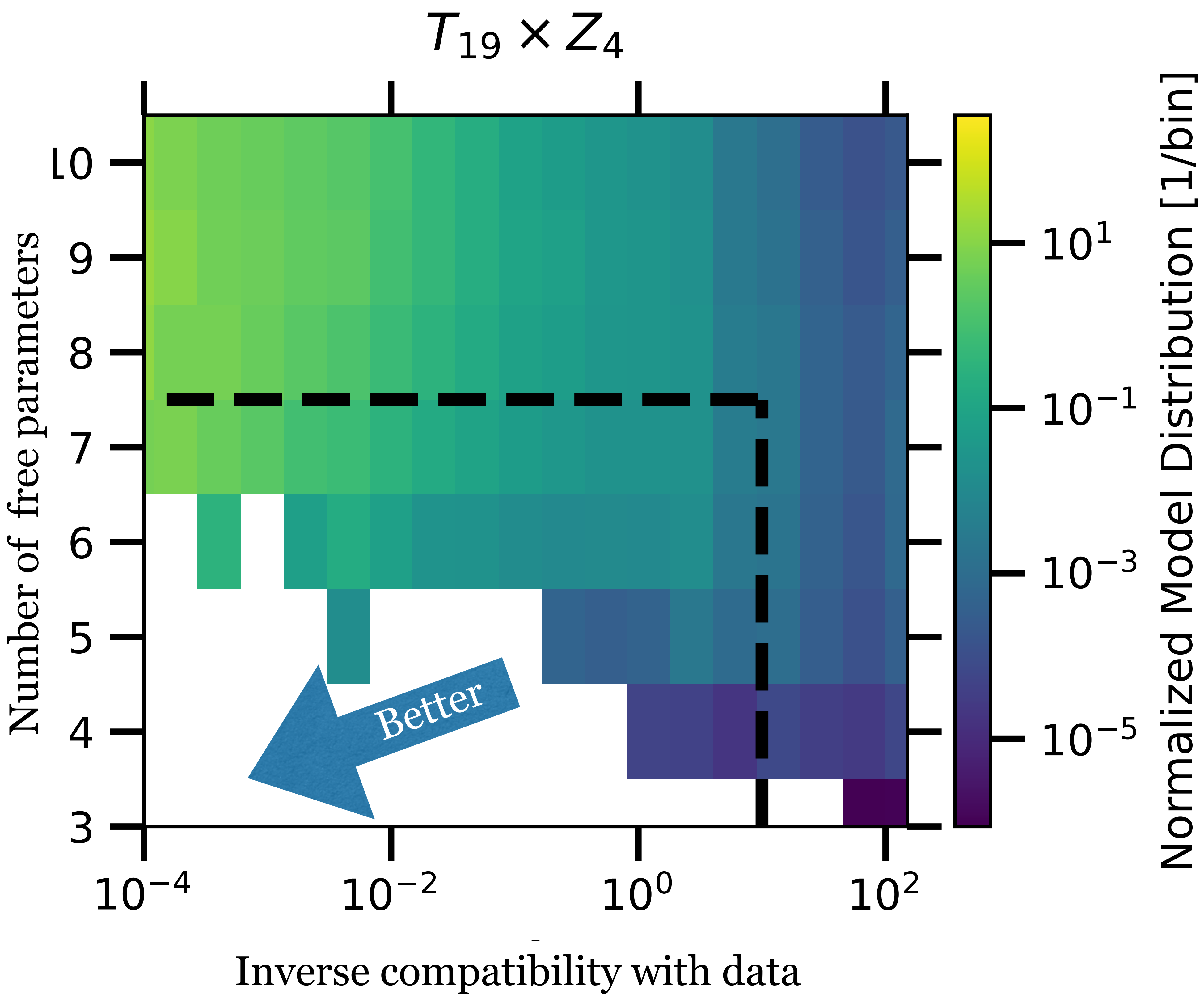
Baretz et al (inc. Ghosh): arXiv:2506.08080

A new symmetry never before studied for neutrino model building





Rich landscape of viable neutrino theories, never previously studied!



More details studies for these new theory models also performed

See paper for gory details: arXiv:2506.08080

Curious what an example new theory look like?

	Fields:	L	E	N	H_u	H_d	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5
Groups	T_{19}	$\mathbf{3}_1$	$\mathbf{3}_2$	$\bar{\mathbf{3}}_1$	$\mathbf{1}$	$\mathbf{1}''$	$\bar{\mathbf{3}}_2$	$\bar{\mathbf{3}}_3$	$\bar{\mathbf{3}}_2$	$\bar{\mathbf{3}}_3$	$\bar{\mathbf{3}}_3$
	$\mathbb{Z}_4$	3	2	0	0	0	3	2	3	1	3

Irreducible representation assigned to fields

Charges

See paper for gory details: arXiv:2506.08080

Curious what an example new theory look like?

5 Flavons ϕ_i

	Fields:	L	E	N	H_u	H_d	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5
Groups	T_{19}	$\mathbf{3}_1$	$\mathbf{3}_2$	$\bar{\mathbf{3}}_1$	$\mathbf{1}$	$\mathbf{1}''$	$\bar{\mathbf{3}}_2$	$\bar{\mathbf{3}}_3$	$\bar{\mathbf{3}}_2$	$\bar{\mathbf{3}}_3$	$\bar{\mathbf{3}}_3$
	$\mathbb{Z}_4$	3	2	0	0	0	3	2	3	1	3

Irreducible representation assigned to fields

Charges

See paper for gory details: arXiv:2506.08080

Curious what an example new theory look like?

Higgs doublet

5 Flavons ϕ_i

Fields:	L	E	N	H_u	H_d	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5
T_{19}	$\mathbf{3}_1$	$\mathbf{3}_2$	$\bar{\mathbf{3}}_1$	$\mathbf{1}$	$\mathbf{1}''$	$\bar{\mathbf{3}}_2$	$\bar{\mathbf{3}}_3$	$\bar{\mathbf{3}}_2$	$\bar{\mathbf{3}}_3$	$\bar{\mathbf{3}}_3$
$\mathbb{Z}_4$	3	2	0	0	0	3	2	3	1	3

Groups

Irreducible representation assigned to fields

Charges

See paper for gory details: arXiv:2506.08080

Curious what an example new theory look like?

Right handed neutrinos

Higgs doublet

5 Flavons ϕ_i

Groups

Fields:	L	E	N	H_u	H_d	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5
T_{19}	$\mathbf{3}_1$	$\mathbf{3}_2$	$\bar{\mathbf{3}}_1$	$\mathbf{1}$	$\mathbf{1}''$	$\bar{\mathbf{3}}_2$	$\bar{\mathbf{3}}_3$	$\bar{\mathbf{3}}_2$	$\bar{\mathbf{3}}_3$	$\bar{\mathbf{3}}_3$
$\mathbb{Z}_4$	3	2	0	0	0	3	2	3	1	3

Irreducible representation assigned to fields

Charges

See paper for gory details: arXiv:2506.08080

Curious what an example new theory look like?

3 Lepton singlets

Right handed neutrinos

Higgs doublet

5 Flavons ϕ_i

Groups

Fields:	L	E	N	H_u	H_d	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5
T_{19}	$\mathbf{3}_1$	$\mathbf{3}_2$	$\bar{\mathbf{3}}_1$	$\mathbf{1}$	$\mathbf{1}''$	$\bar{\mathbf{3}}_2$	$\bar{\mathbf{3}}_3$	$\bar{\mathbf{3}}_2$	$\bar{\mathbf{3}}_3$	$\bar{\mathbf{3}}_3$
$\mathbb{Z}_4$	3	2	0	0	0	3	2	3	1	3

Irreducible representation assigned to fields

Charges

See paper for gory details: arXiv:2506.08080

Curious what an example new theory look like?

3 Lepton singlets

Lepton doublets (L_1, L_2, L_3)

Right handed neutrinos

Higgs doublet

5 Flavons ϕ_i

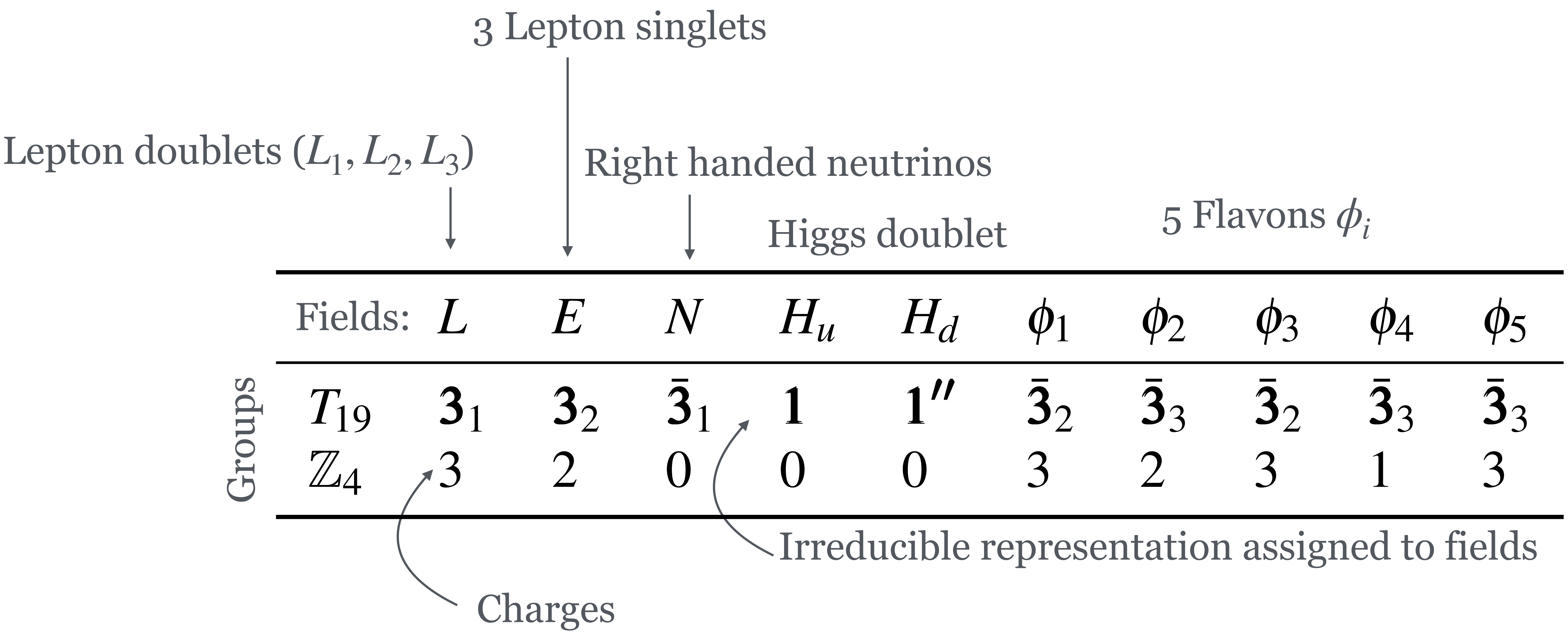
	Fields:	L	E	N	H_u	H_d	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5
Groups	T_{19}	$\mathbf{3}_1$	$\mathbf{3}_2$	$\bar{\mathbf{3}}_1$	$\mathbf{1}$	$\mathbf{1}''$	$\bar{\mathbf{3}}_2$	$\bar{\mathbf{3}}_3$	$\bar{\mathbf{3}}_2$	$\bar{\mathbf{3}}_3$	$\bar{\mathbf{3}}_3$
	$\mathbb{Z}_4$	3	2	0	0	0	3	2	3	1	3

Irreducible representation assigned to fields

Charges

See paper for gory details: arXiv:2506.08080

Curious what an example new theory look like?



Lepton doublets (L_1, L_2, L_3)

3 Lepton singlets

Right handed neutrinos

Higgs doublet

5 Flavons ϕ_i

	Fields:	L	E	N	H_u	H_d	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5
Groups	T_{19}	$\mathbf{3}_1$	$\mathbf{3}_2$	$\bar{\mathbf{3}}_1$	$\mathbf{1}$	$\mathbf{1}''$	$\bar{\mathbf{3}}_2$	$\bar{\mathbf{3}}_3$	$\bar{\mathbf{3}}_2$	$\bar{\mathbf{3}}_3$	$\bar{\mathbf{3}}_3$
	$\mathbb{Z}_4$	3	2	0	0	0	3	2	3	1	3

Irreducible representation assigned to fields

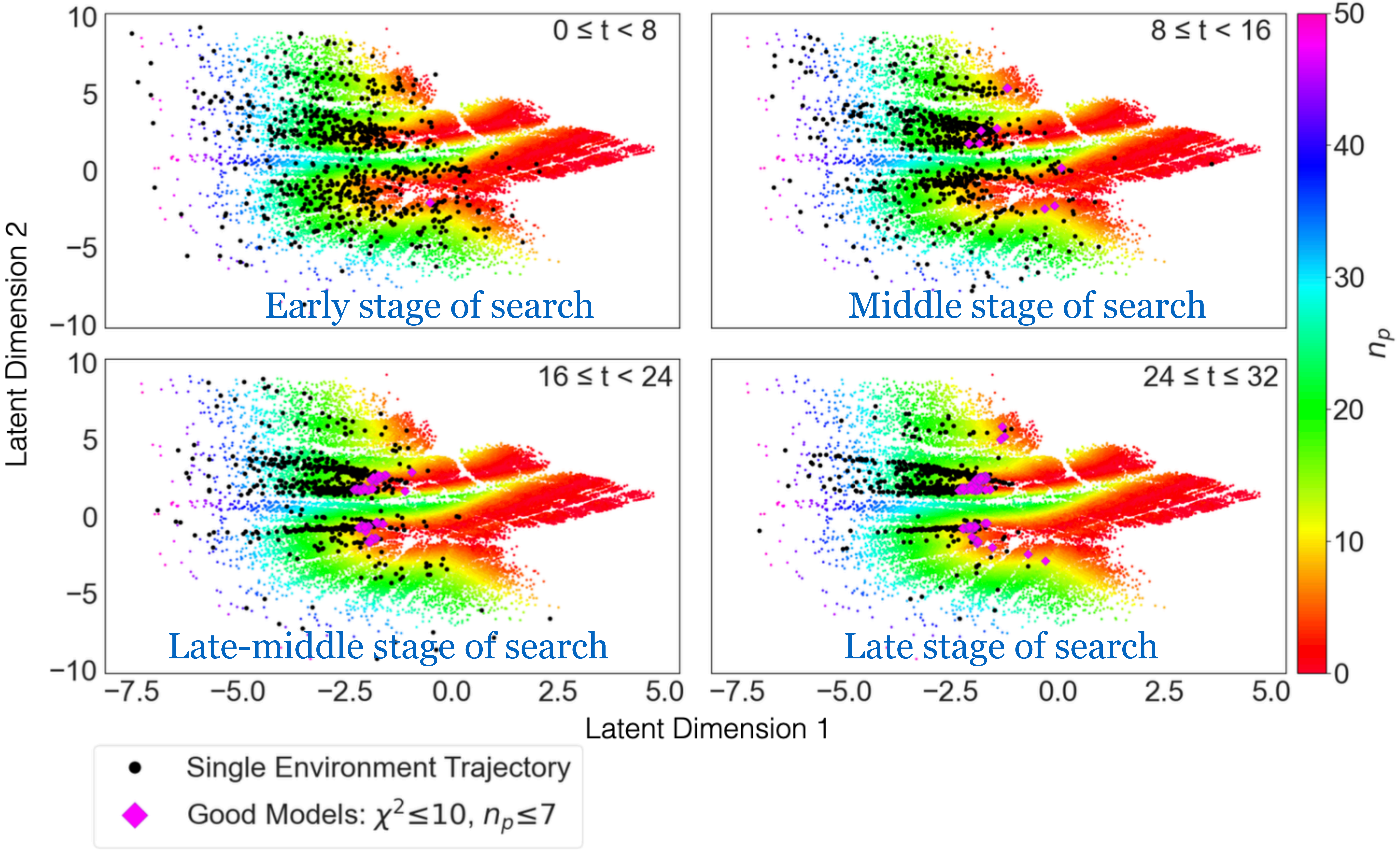
Charges

More recent work taking forward AI+workflows with commercial LLM agents and custom RL agents for such workflows by colleagues, eg. Alexander et al, Agarwal et al

Prior RL for BSM by Wojcik, et al

Deep dive into what the agent is doing over time

Visualised in 2D space created by an auto-encoder



Black dots are search trajectory for one agent in one environment

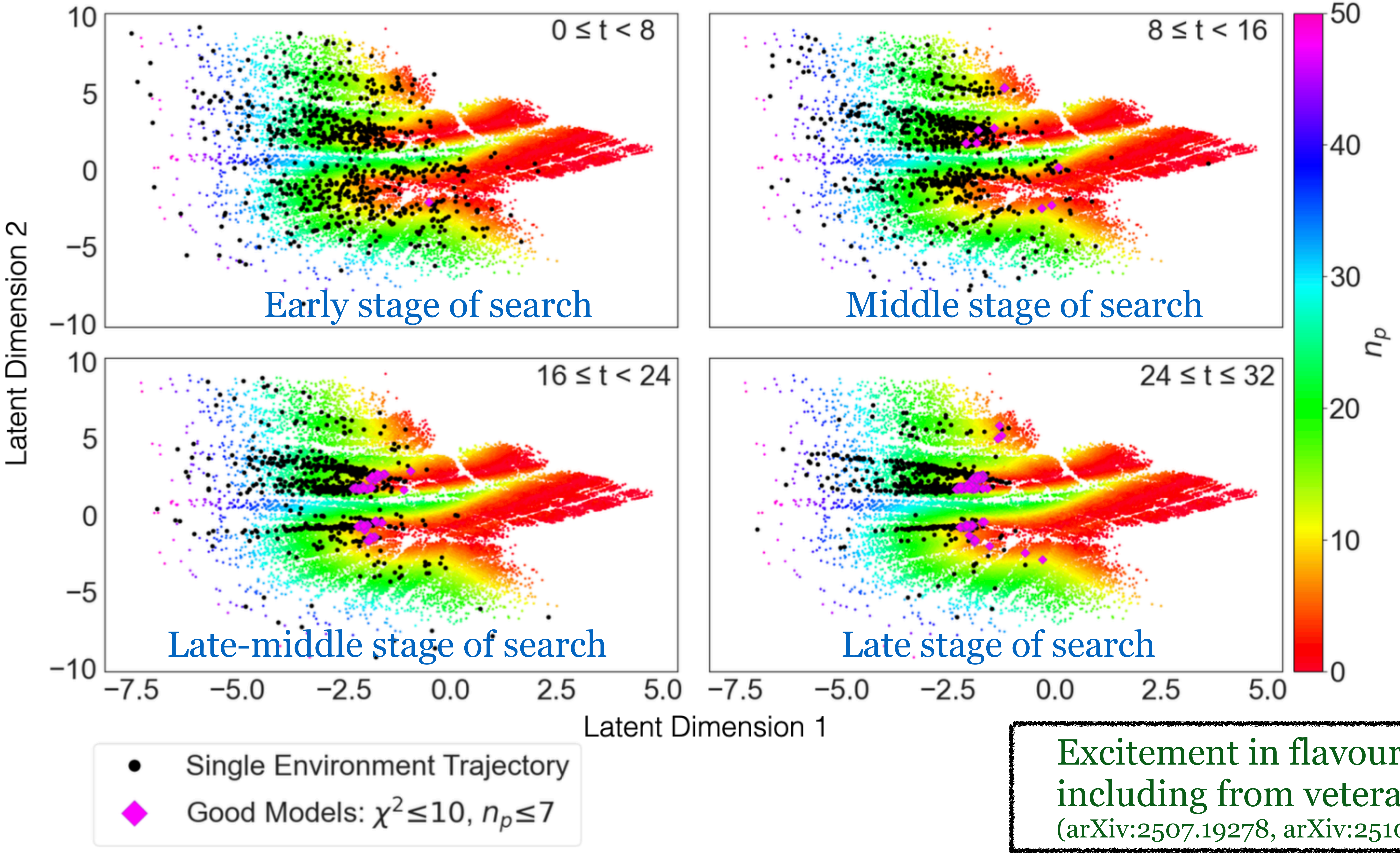
Pink are ‘good models’

Explores widely in the beginning, but realises that zero-parameter models are useless.

Focuses on few parameter models towards the end

Deep dive into what the agent is doing over time

Visualised in 2D space created by an auto-encoder



Black dots are search trajectory for one agent in one environment

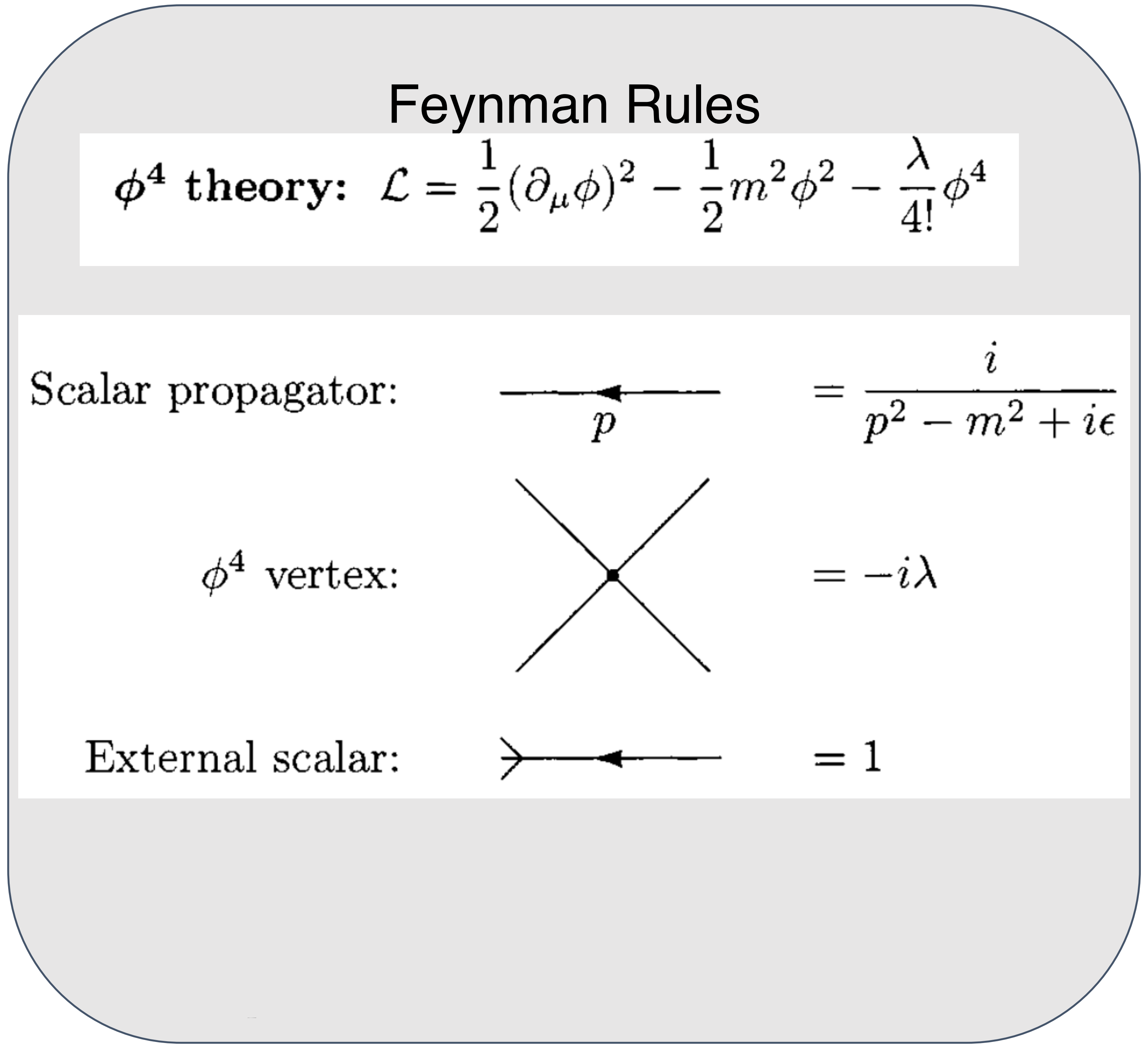
Pink are 'good models'

Explores widely in the beginning, but realises that zero-parameter models are useless.

Focuses on few parameter models towards the end

Excitement in flavour physics community, including from veterans of the field
(arXiv:2507.19278, arXiv:2510.25495)

AI to enhance precision QFT calculations



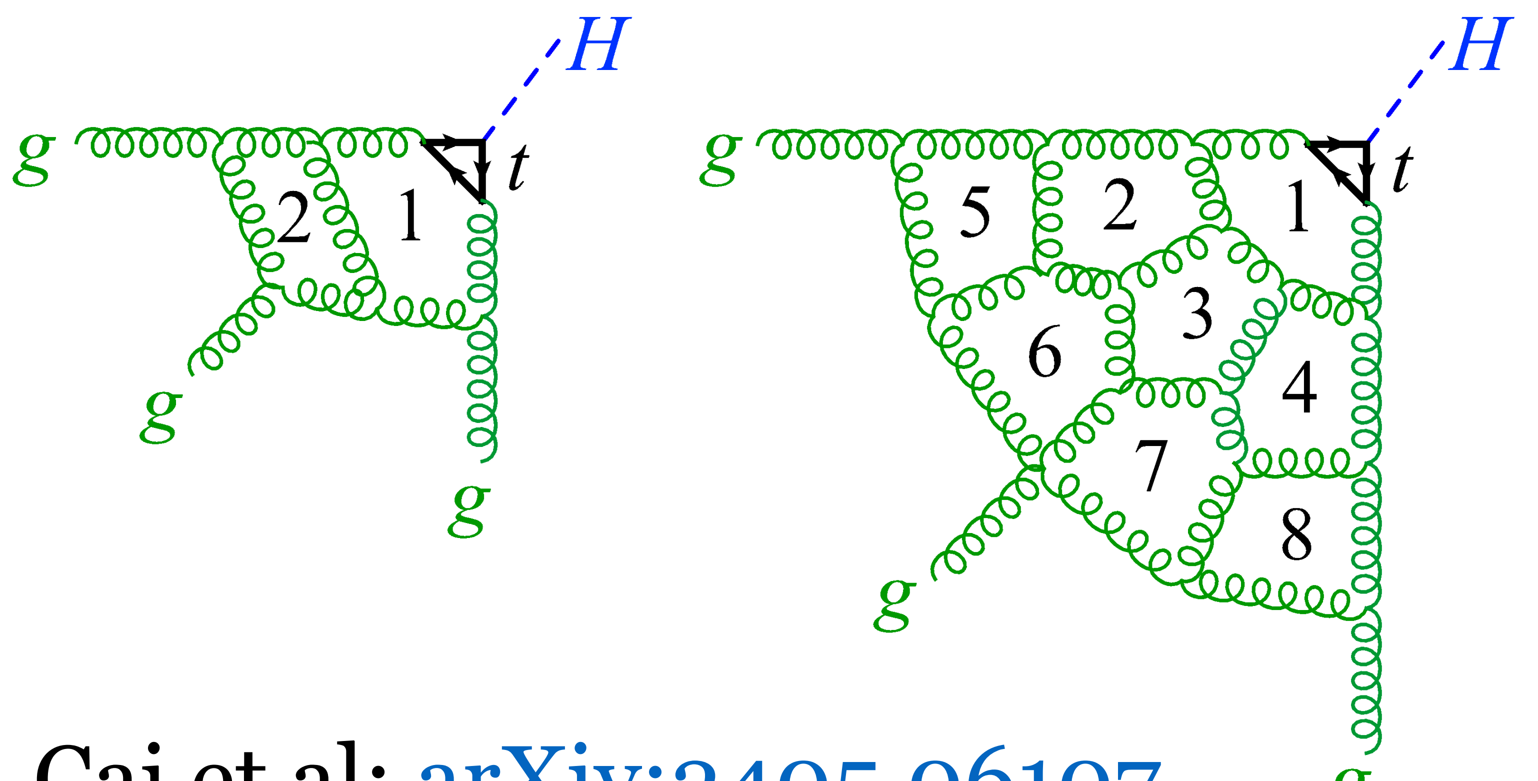
Feynman Rules

ϕ^4 theory: $\mathcal{L} = \frac{1}{2}(\partial_\mu\phi)^2 - \frac{1}{2}m^2\phi^2 - \frac{\lambda}{4!}\phi^4$

Scalar propagator: $= \frac{i}{p^2 - m^2 + i\epsilon}$

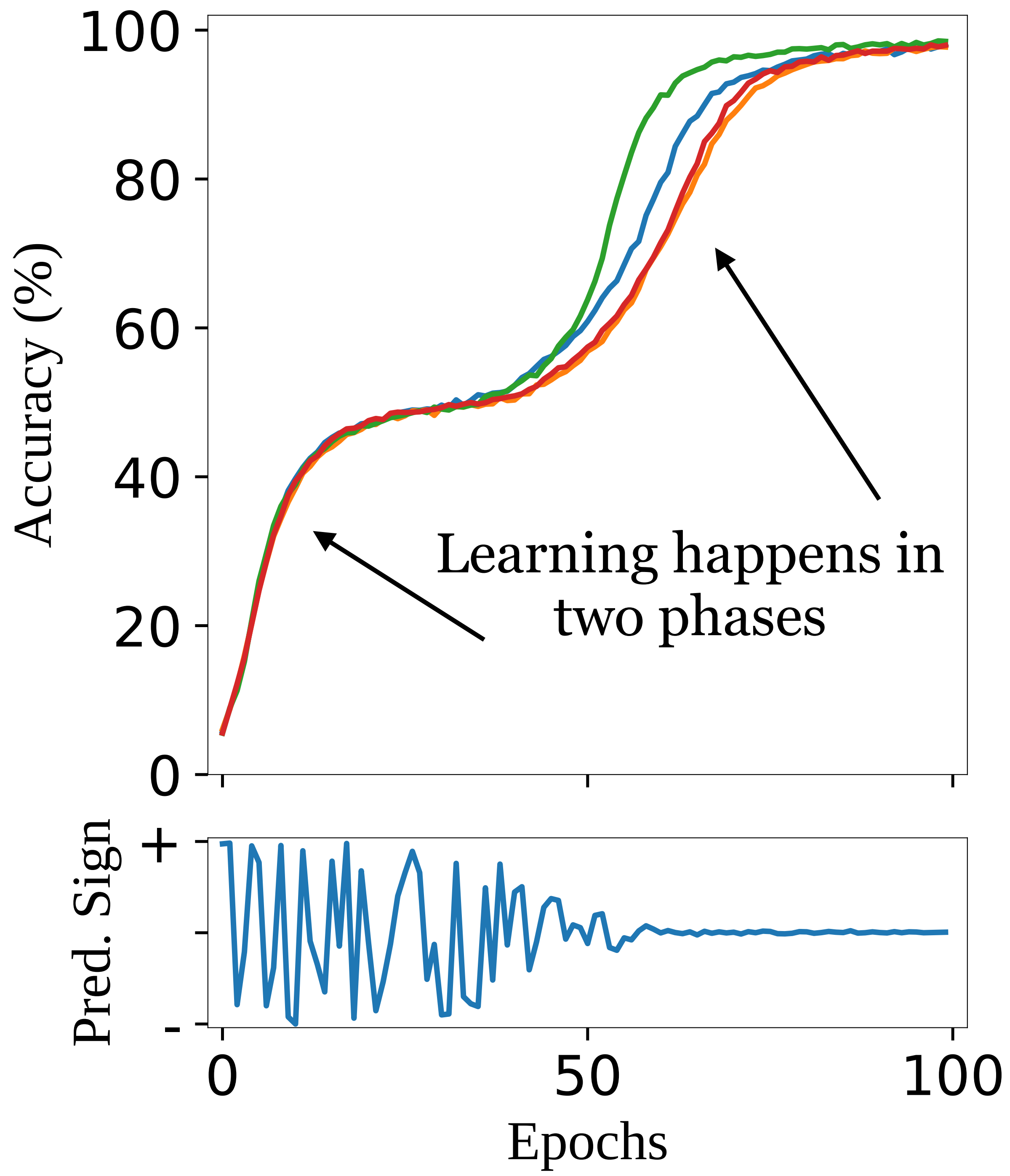
ϕ^4 vertex: $= -i\lambda$

External scalar: $= 1$



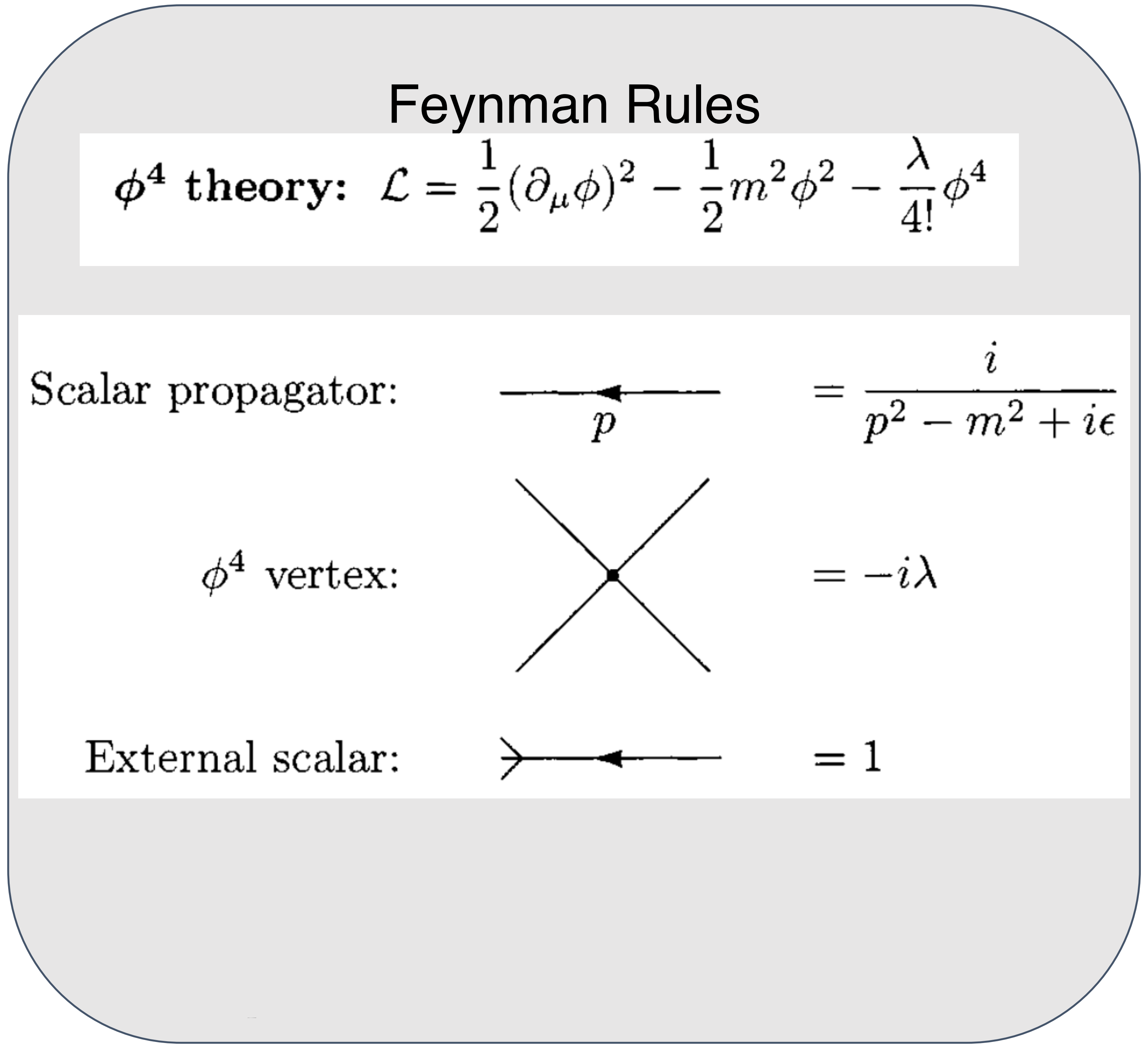
Cai et al: [arXiv:2405.06107](https://arxiv.org/abs/2405.06107)

Transformer to predict terms in amplitude solution



Learning happens in two phases

AI to enhance precision QFT calculations



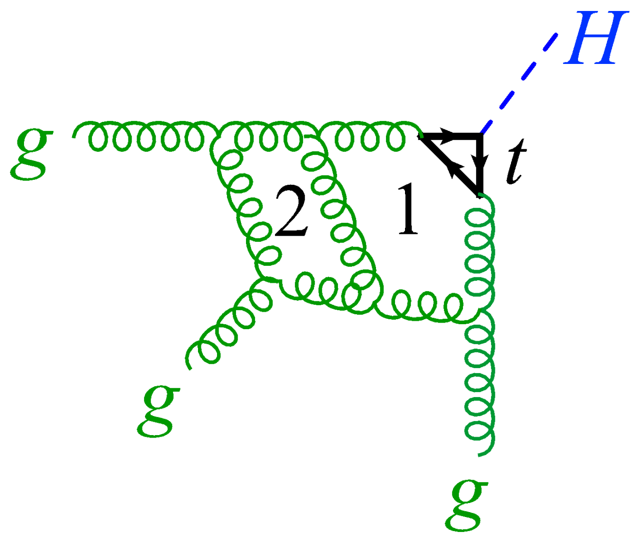
Feynman Rules

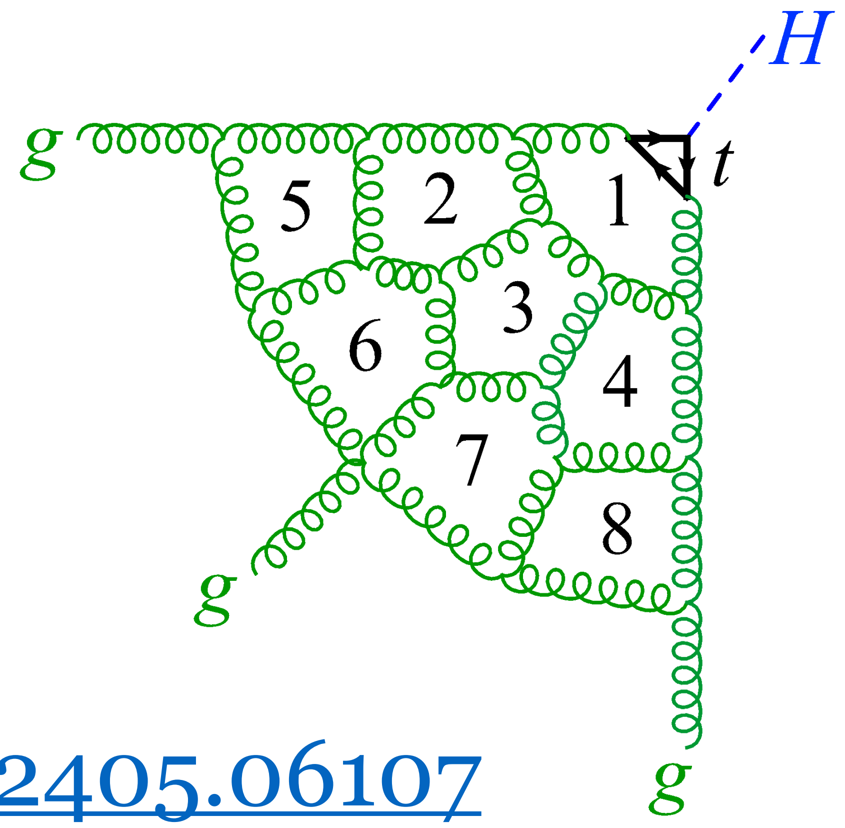
ϕ^4 **theory:** $\mathcal{L} = \frac{1}{2}(\partial_\mu\phi)^2 - \frac{1}{2}m^2\phi^2 - \frac{\lambda}{4!}\phi^4$

Scalar propagator: $= \frac{i}{p^2 - m^2 + i\epsilon}$

ϕ^4 vertex: $= -i\lambda$

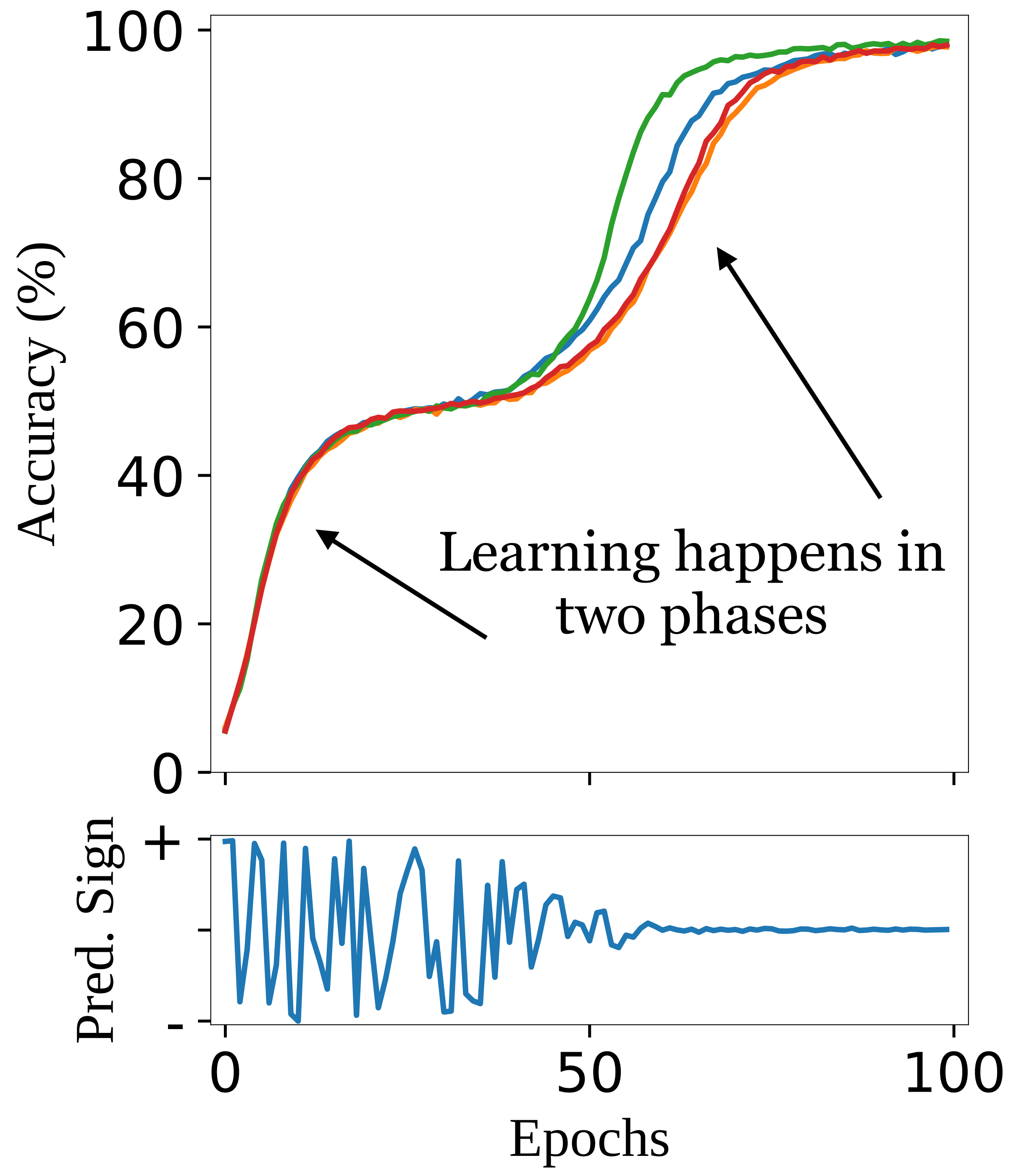
External scalar: $= 1$





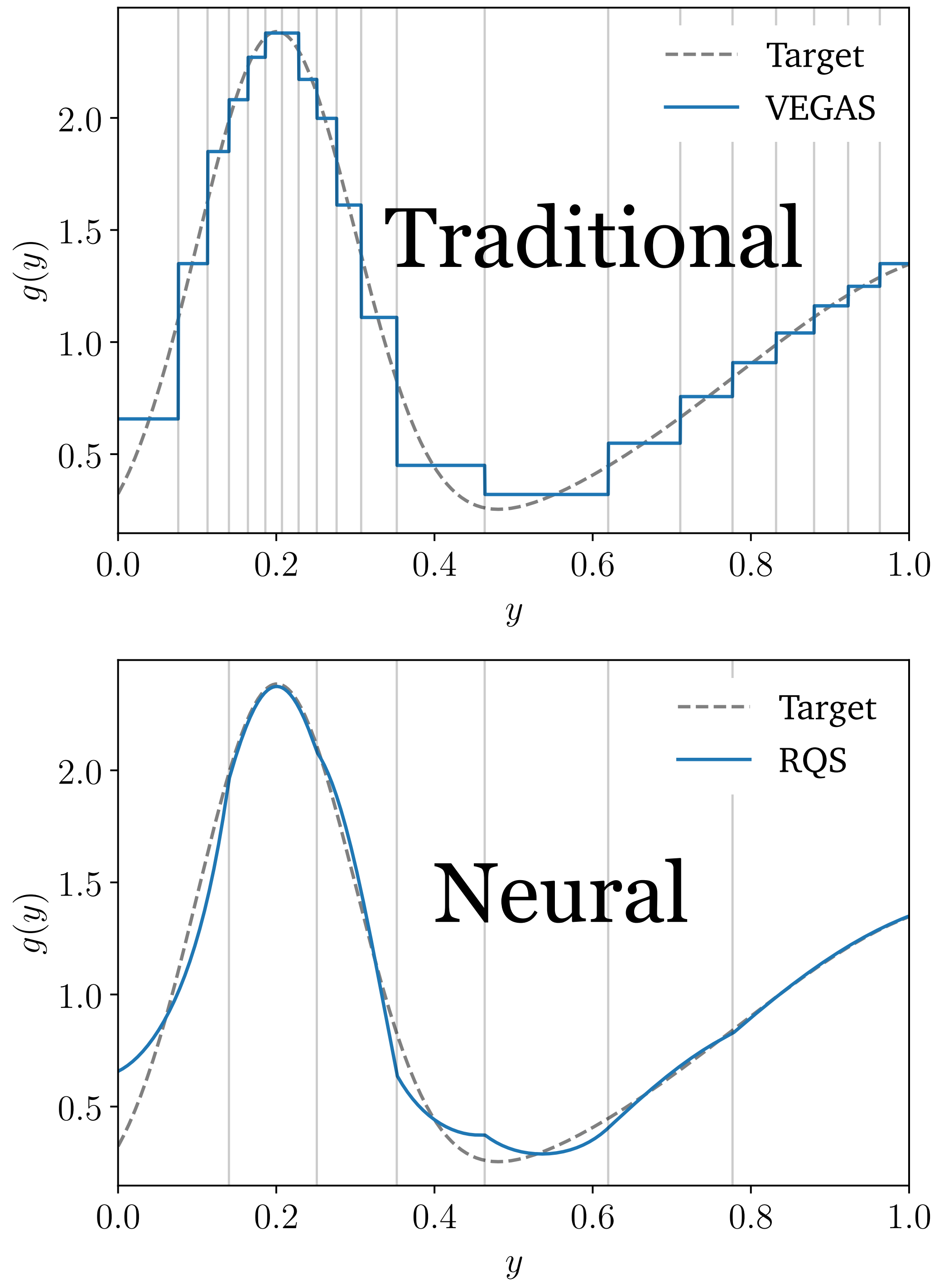
Cai et al: [arXiv:2405.06107]

Transformer to predict terms in amplitude solution



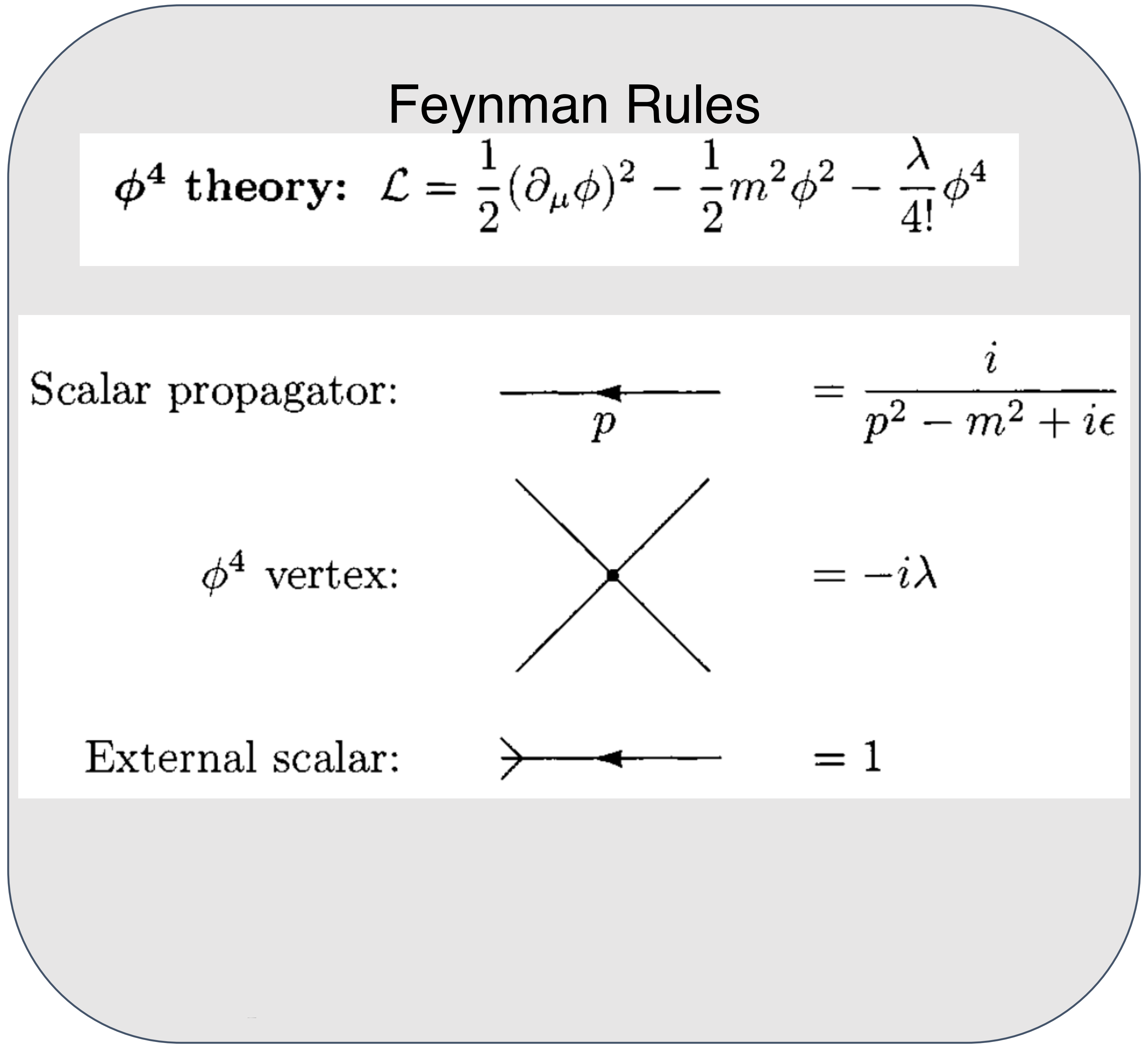
Learning happens in two phases

Improving the efficiency of event generators like MadGraph



Himel et al: [arXiv:2311.01548]

AI to enhance precision QFT calculations



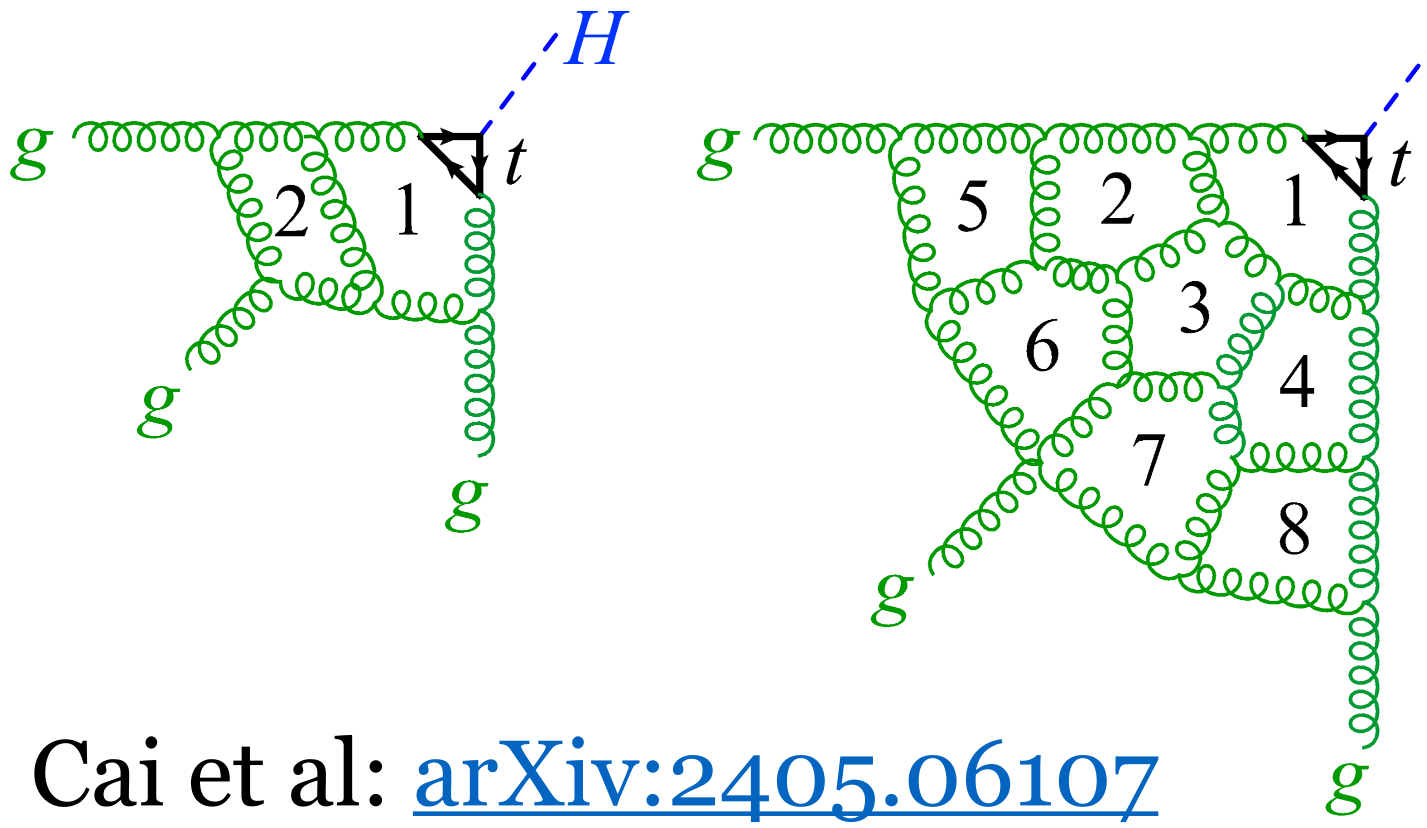
Feynman Rules

ϕ^4 theory: $\mathcal{L} = \frac{1}{2}(\partial_\mu \phi)^2 - \frac{1}{2}m^2\phi^2 - \frac{\lambda}{4!}\phi^4$

Scalar propagator: $= \frac{i}{p^2 - m^2 + i\epsilon}$

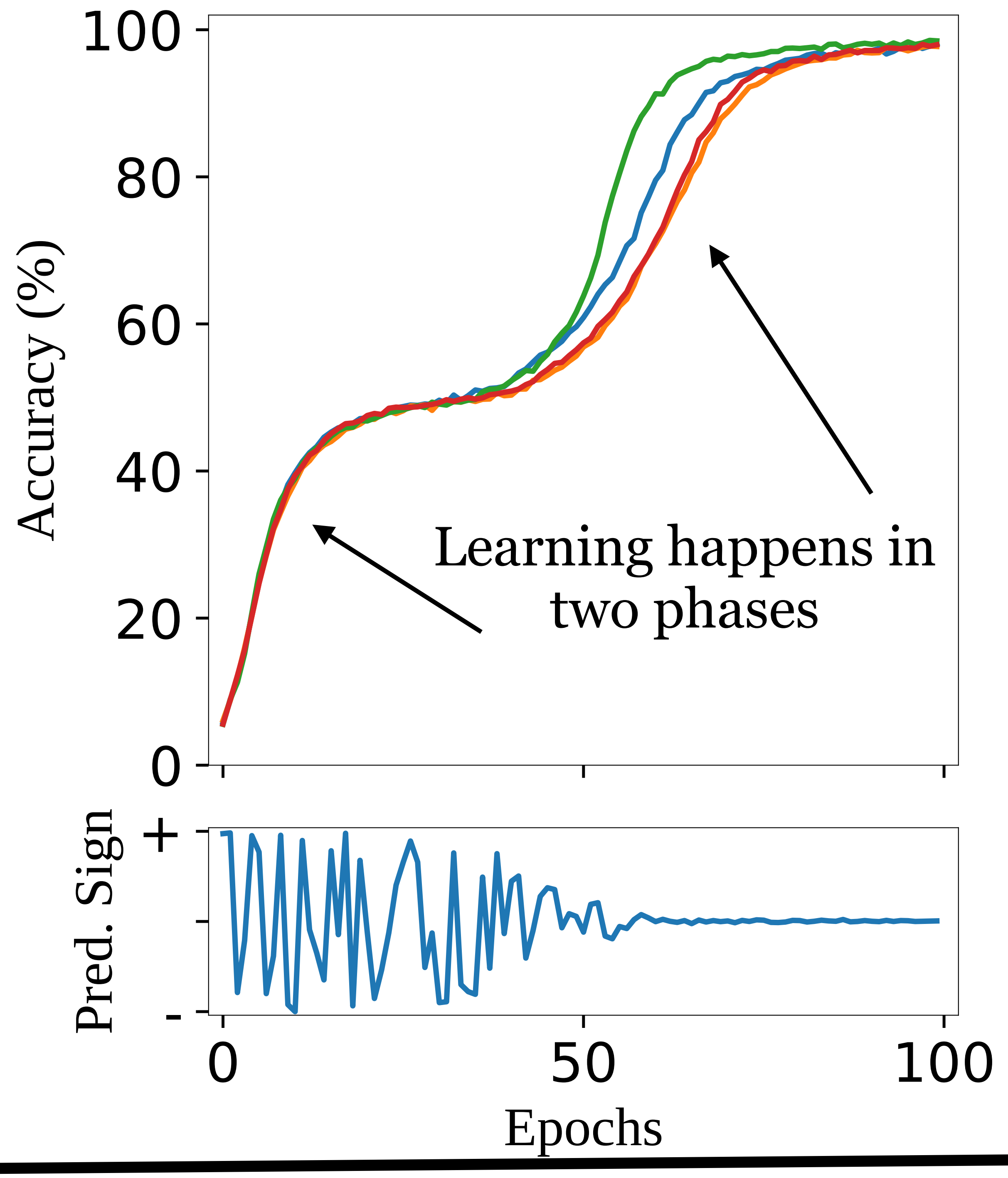
ϕ^4 vertex: $= -i\lambda$

External scalar: $= 1$



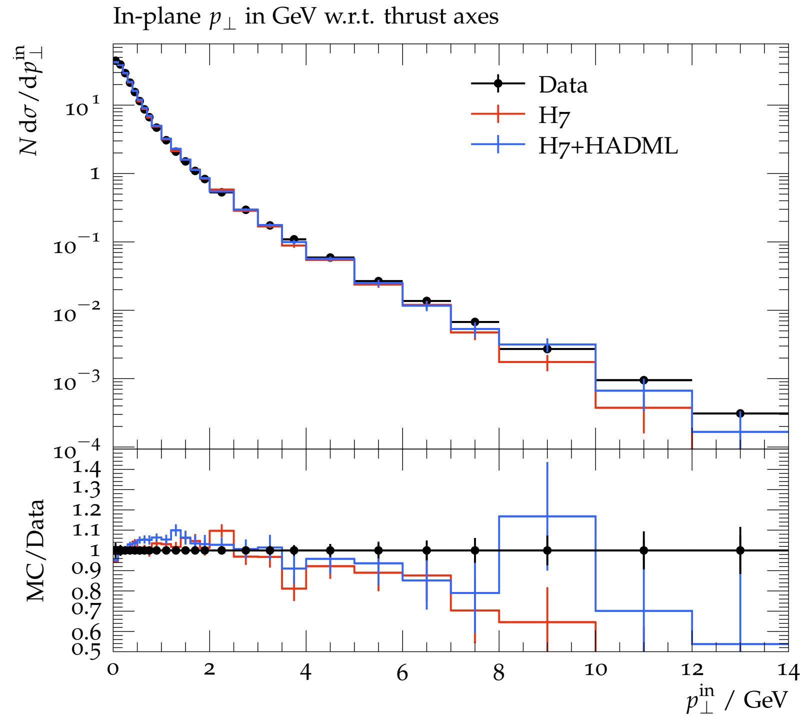
Cai et al: arXiv:2405.06107

Transformer to predict terms in amplitude solution



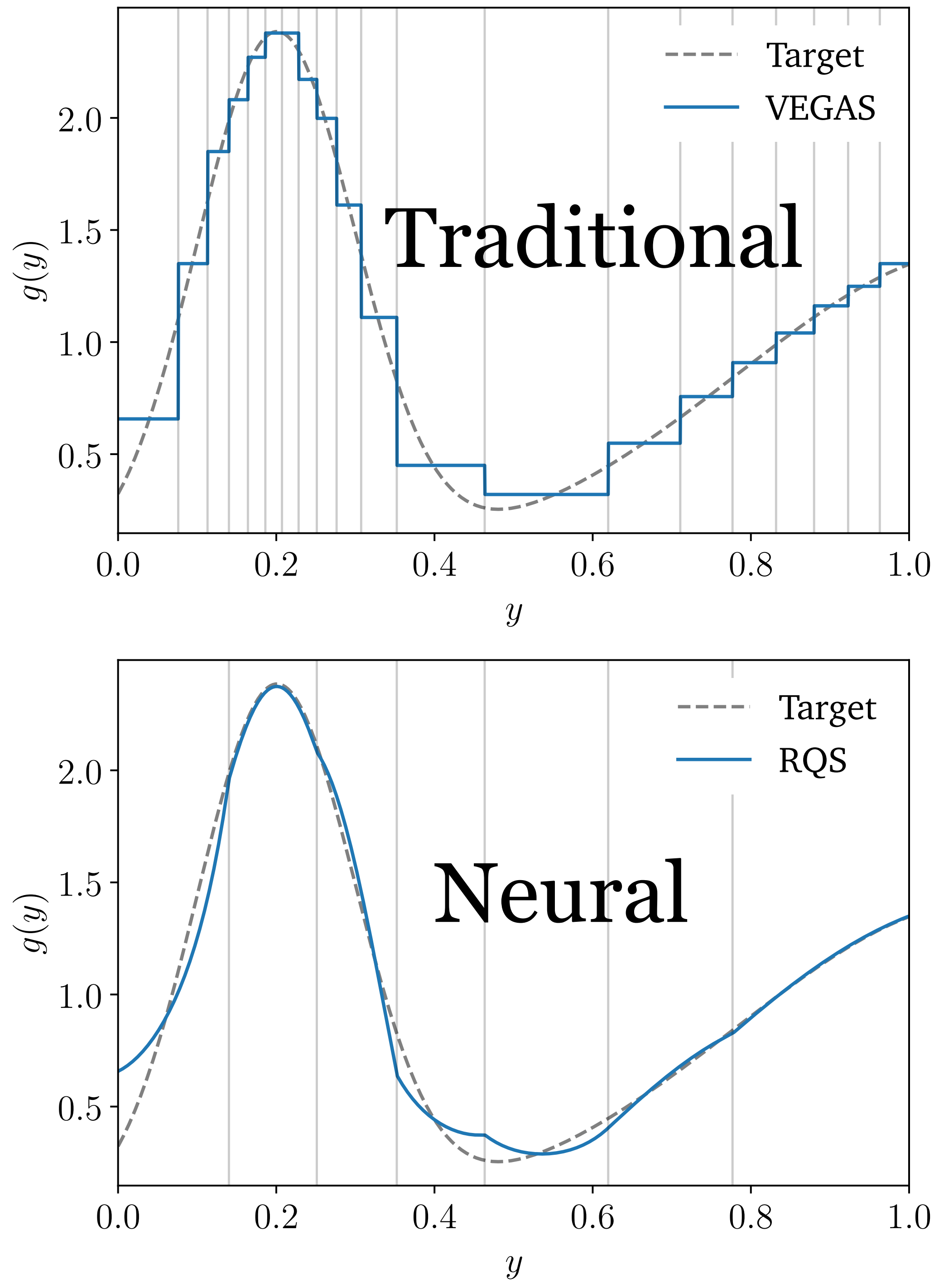
Learning happens in two phases

Learn QCD directly from data:



Ghosh, Ju et al: arXiv:2405.06107

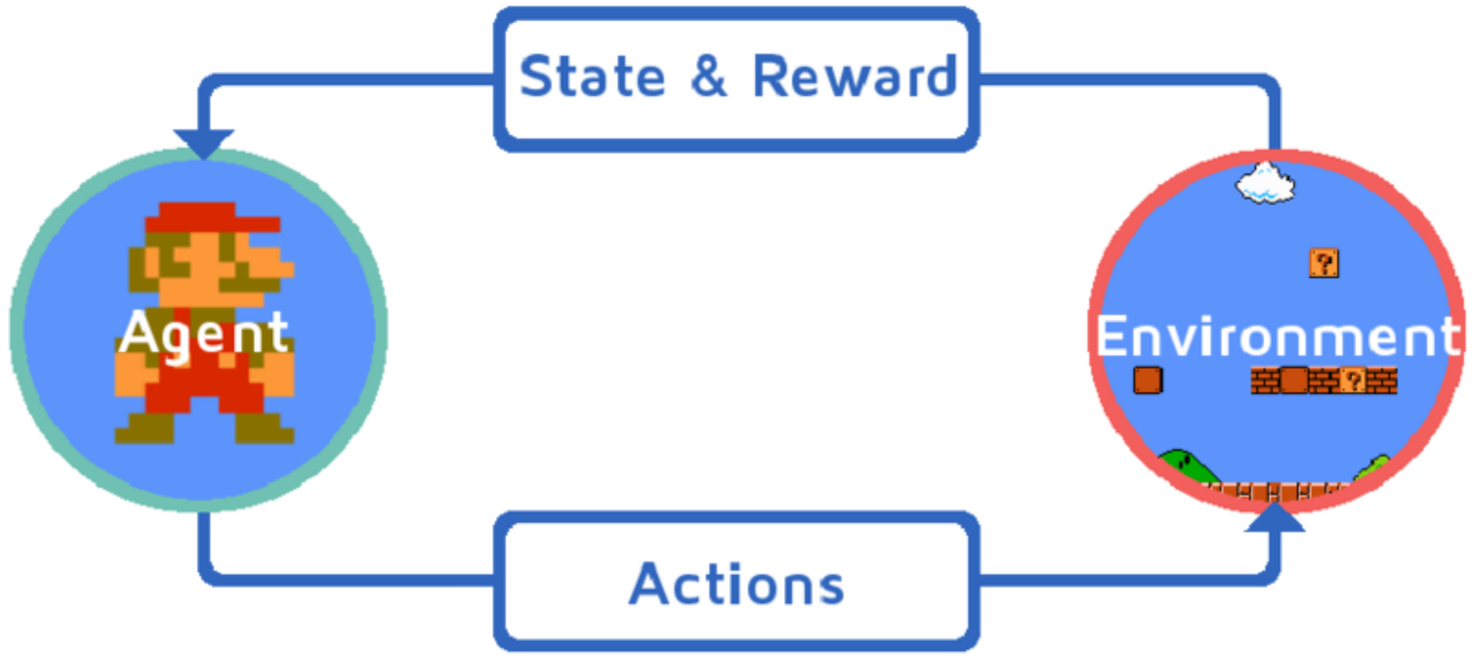
Improving the efficiency of event generators like MadGraph

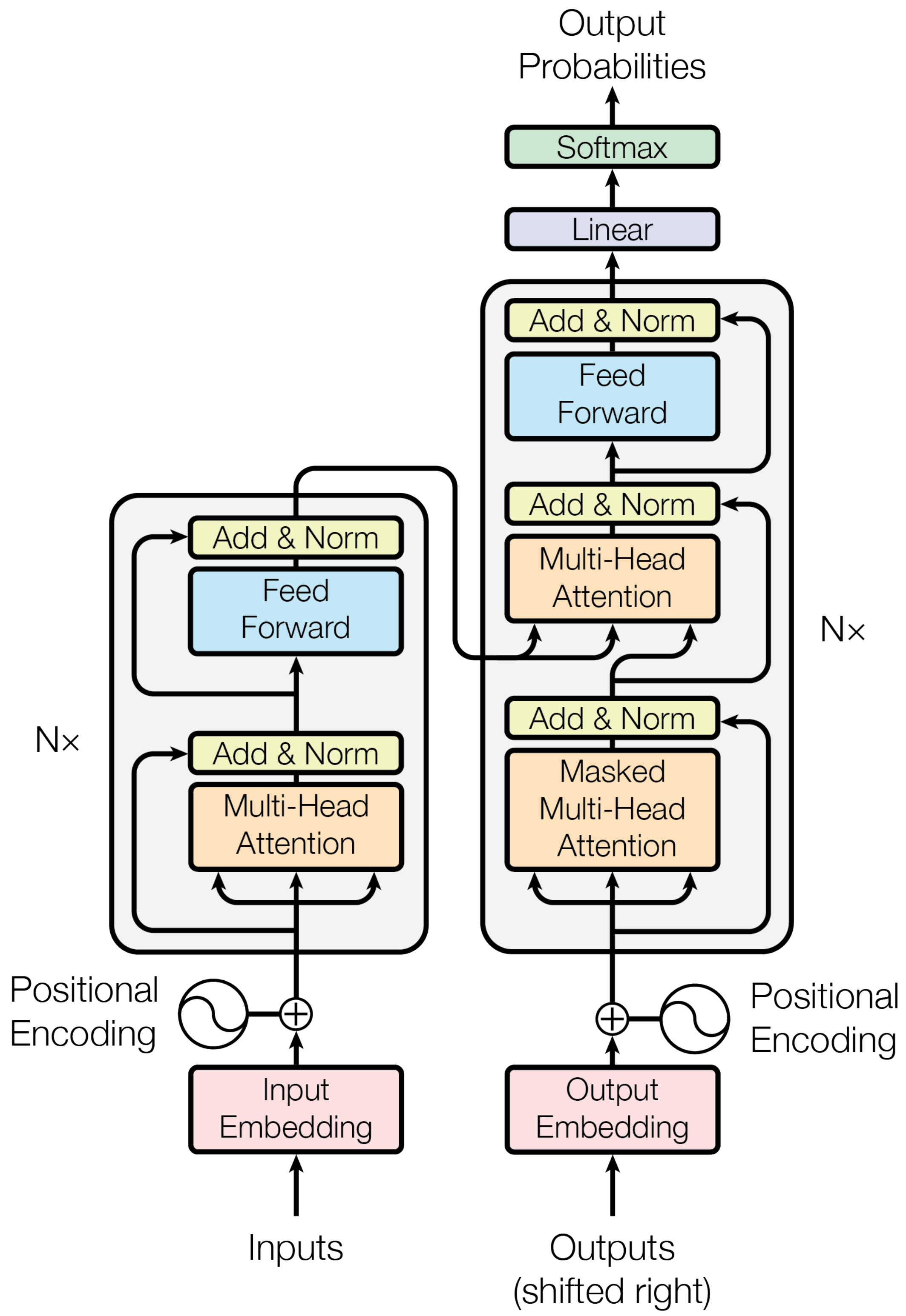


Himel et al: arXiv:2311.01548

What were the ML techniques?

Reinforcement Learning and LLM agents: No tutorial this year





Generative Models: Sofia's tutorial tomorrow!

Transformers: Aaron's tutorial on Thursday !

Hypothesis generation → hypothesis testing in experiments

First need to collect data

Trigger

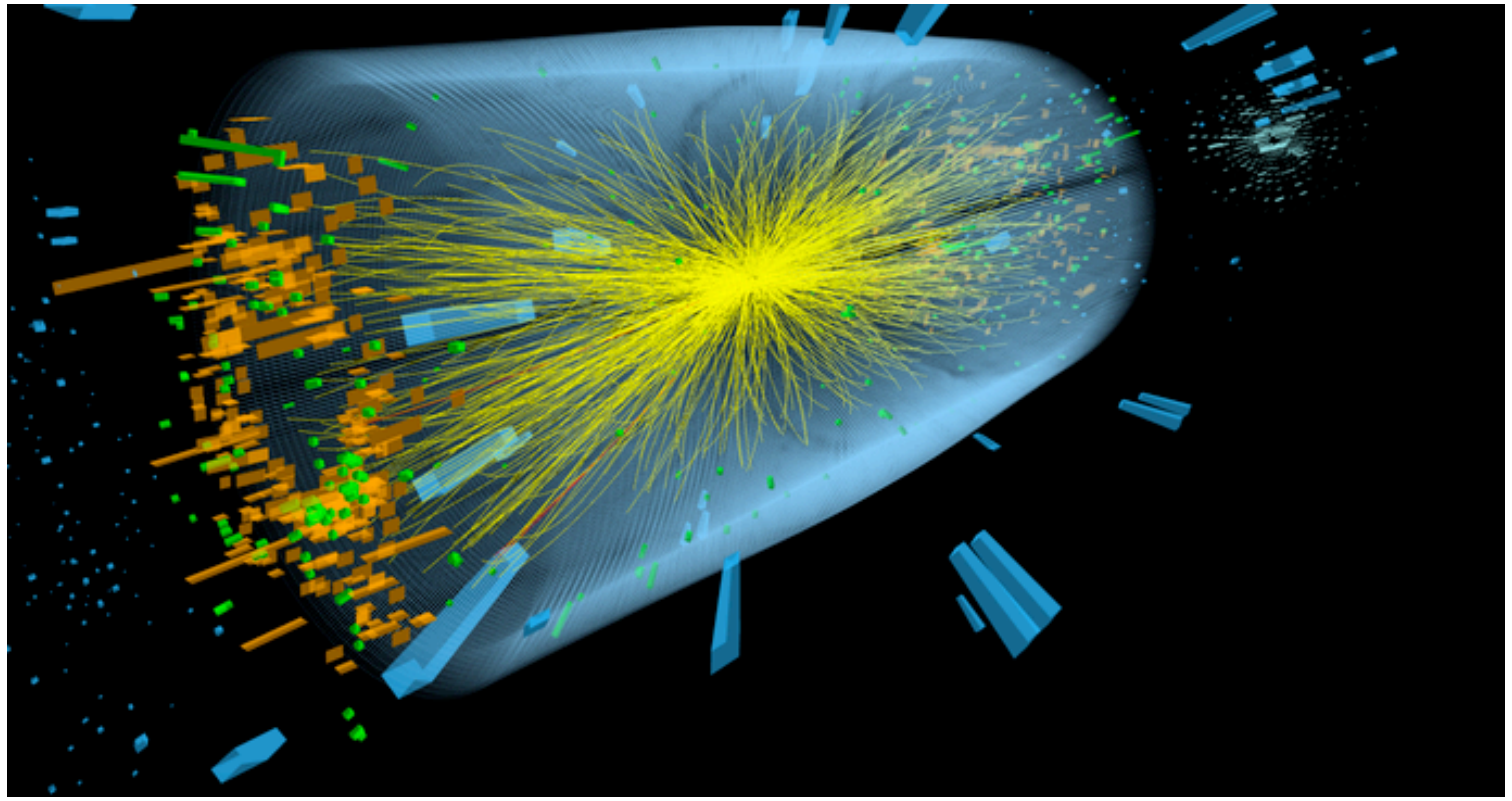


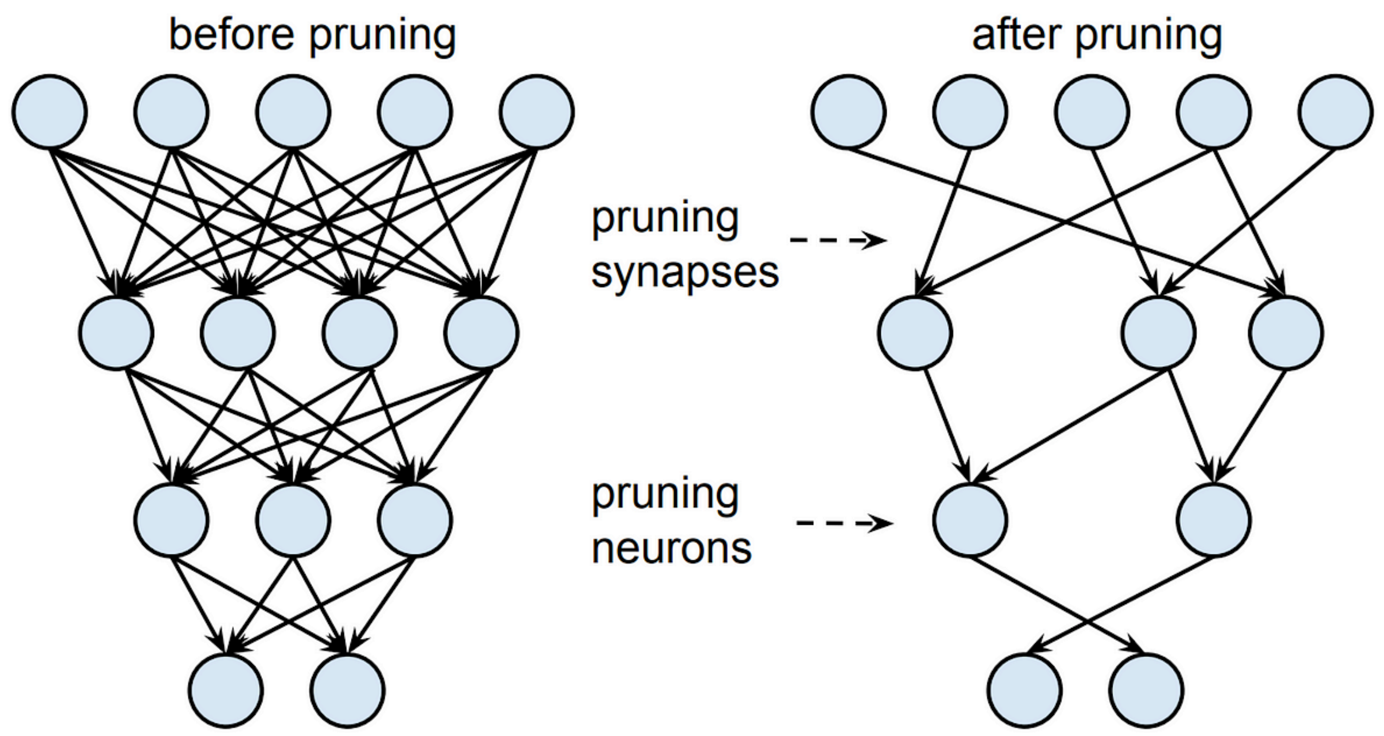
Image: Source

- 40 million collisions/second
- 1MB of data per collision (40 TB/s)
- We throw away 99.998% of the data using simple selections

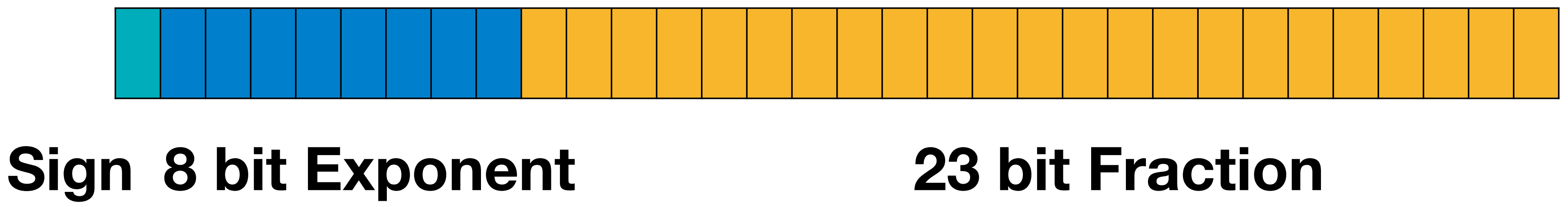
Neural Networks on GPUs too slow!

NNs at triggers

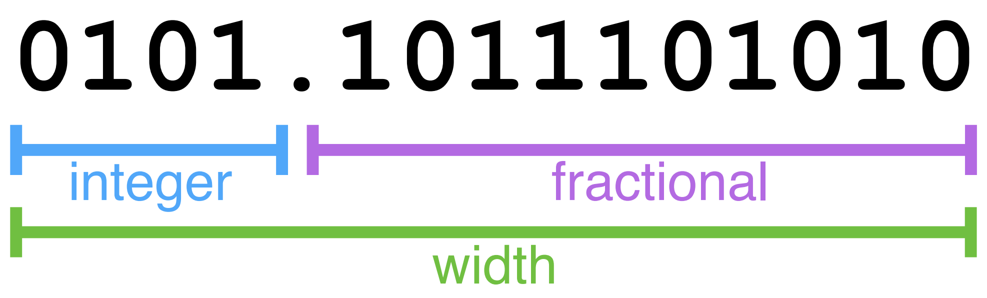
arxiv:2103.05579



$$L_{total}(w) = L_{BCE}(w) + \lambda \Sigma_i |w_i|$$



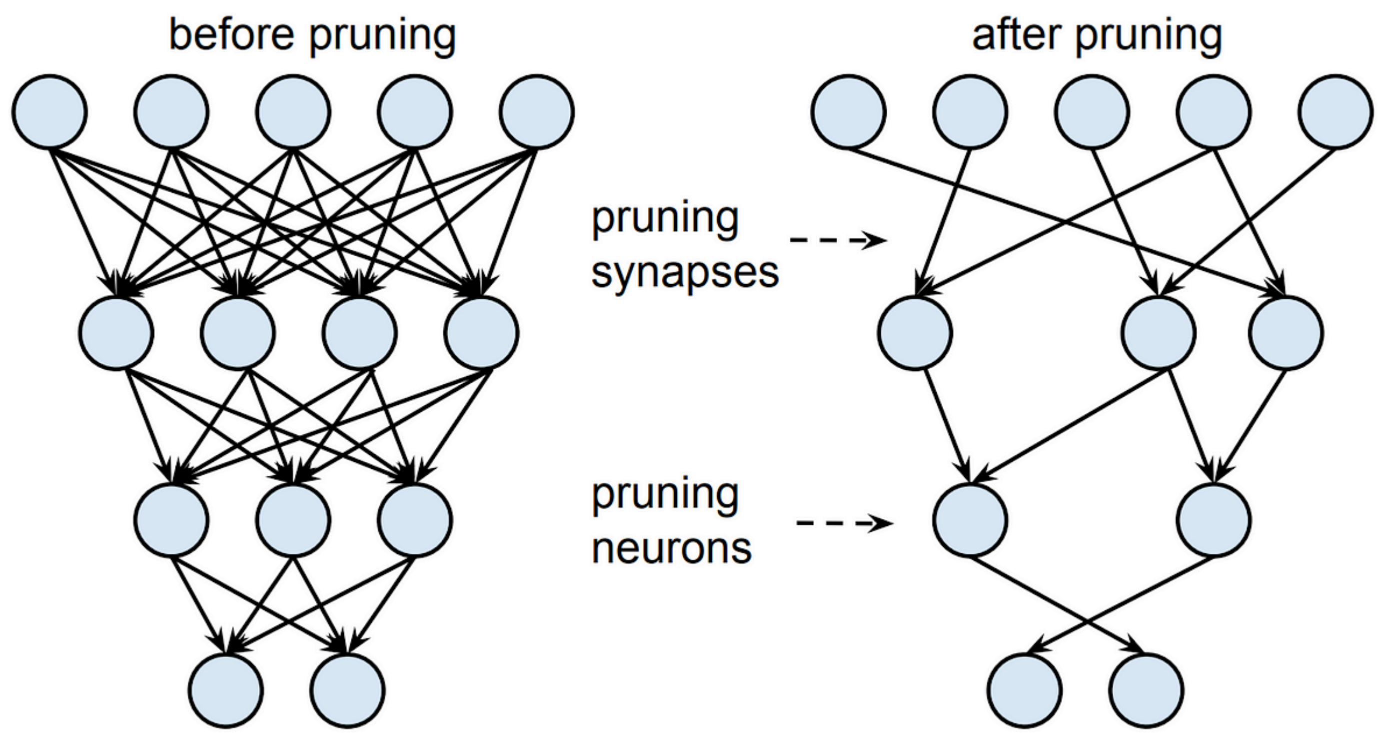
ap_fixed<width,integer>



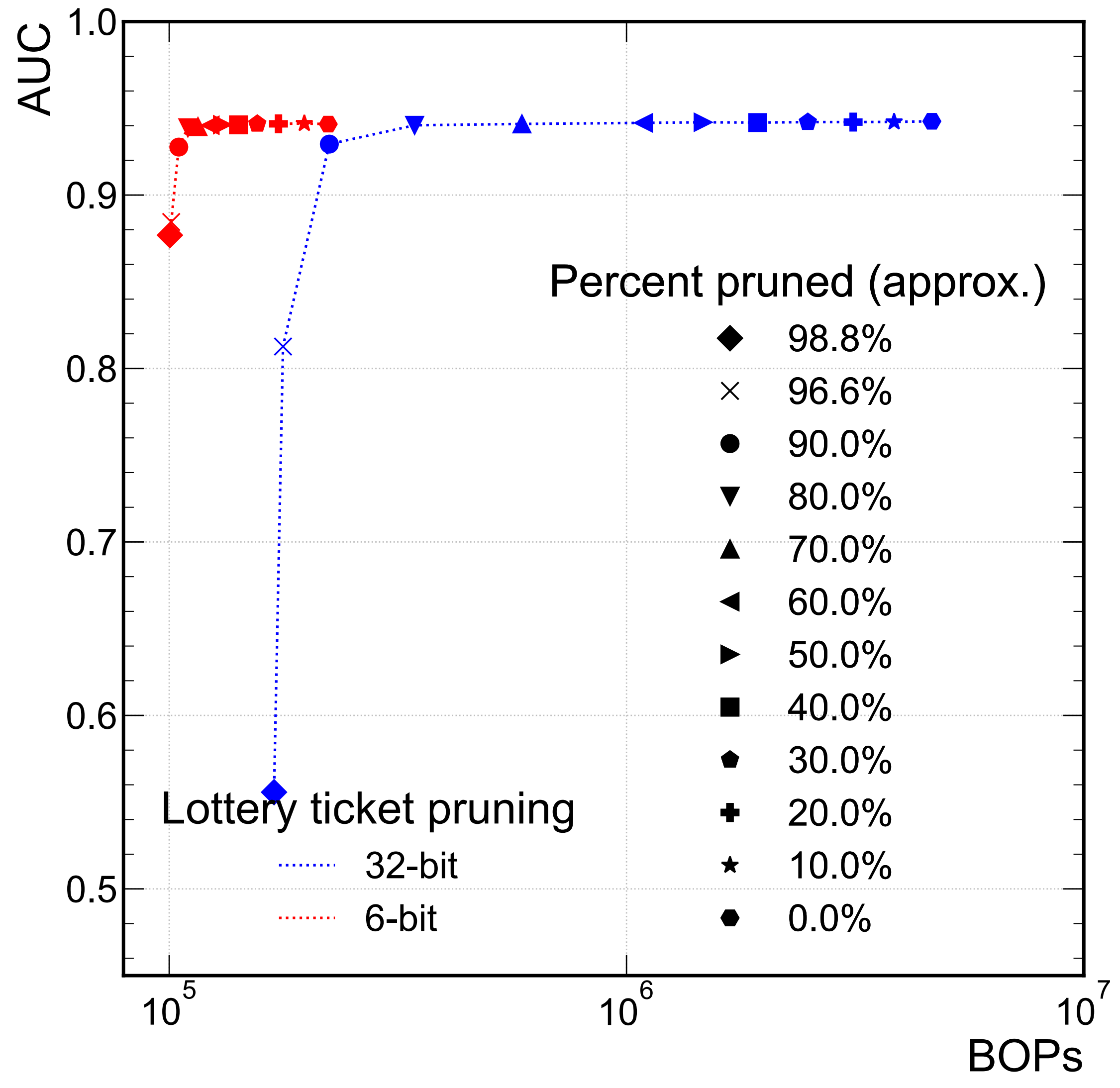
See in Elham’s lecture on Friday!

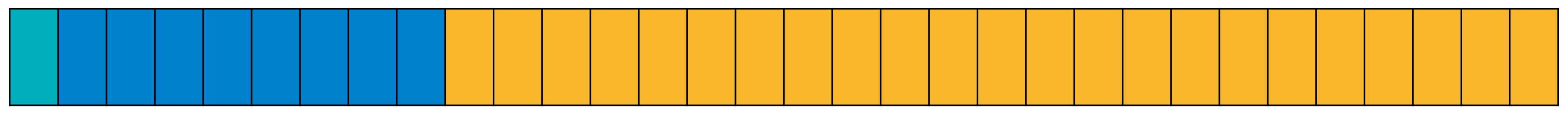
NNs at triggers

arxiv:2103.05579



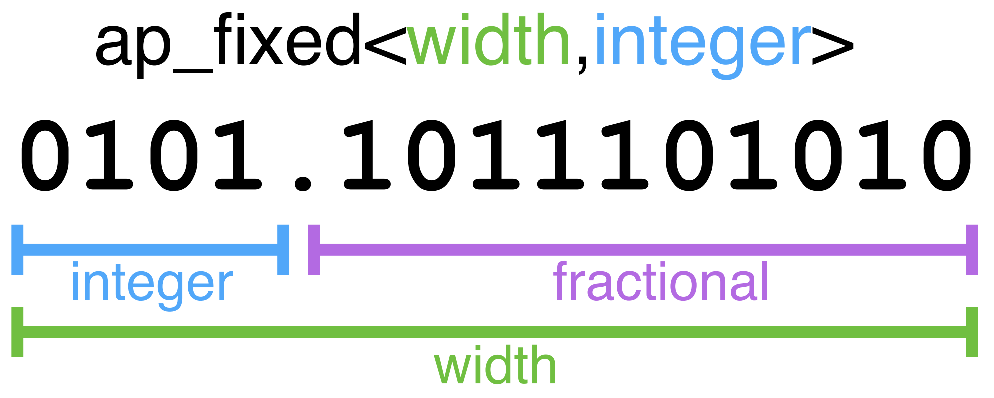
$$L_{total}(w) = L_{BCE}(w) + \lambda \Sigma_i |w_i|$$





You can get rid of most of the network, without losing much performance

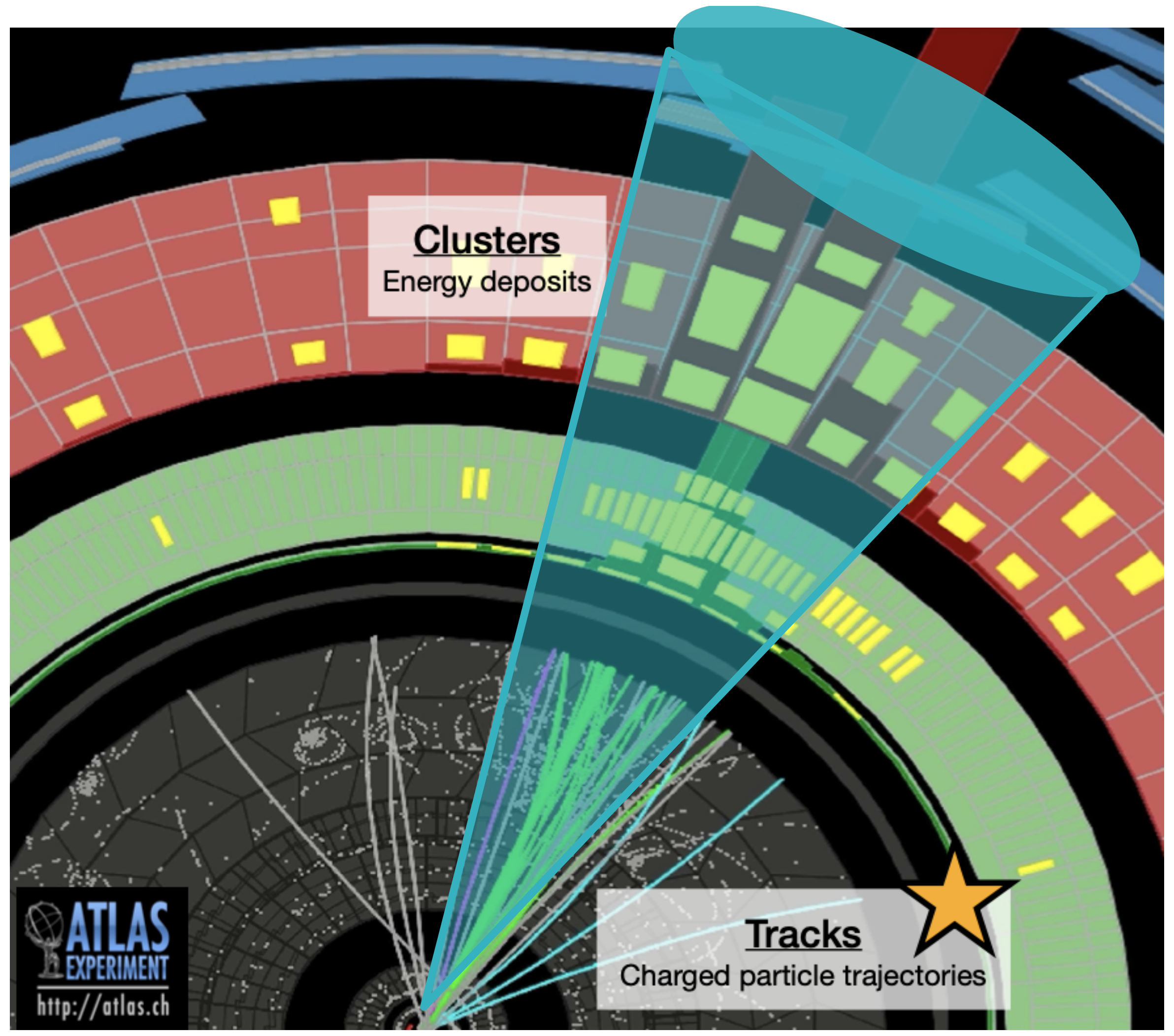
Kinda like most of raw collision data ;)



hls4ml

See in Elham's lecture on Friday!

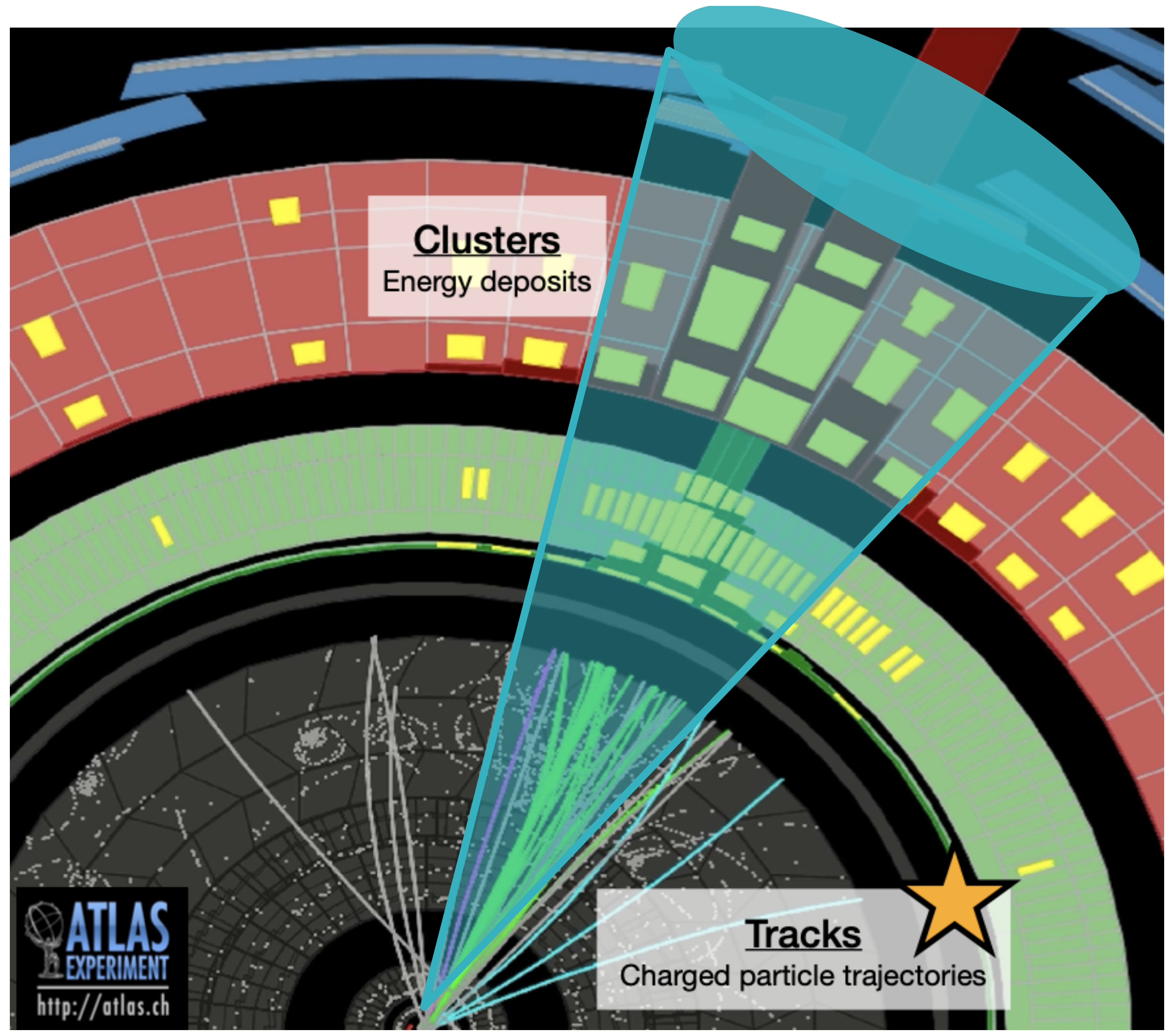
Making sense of the remaining mess by tagging jets



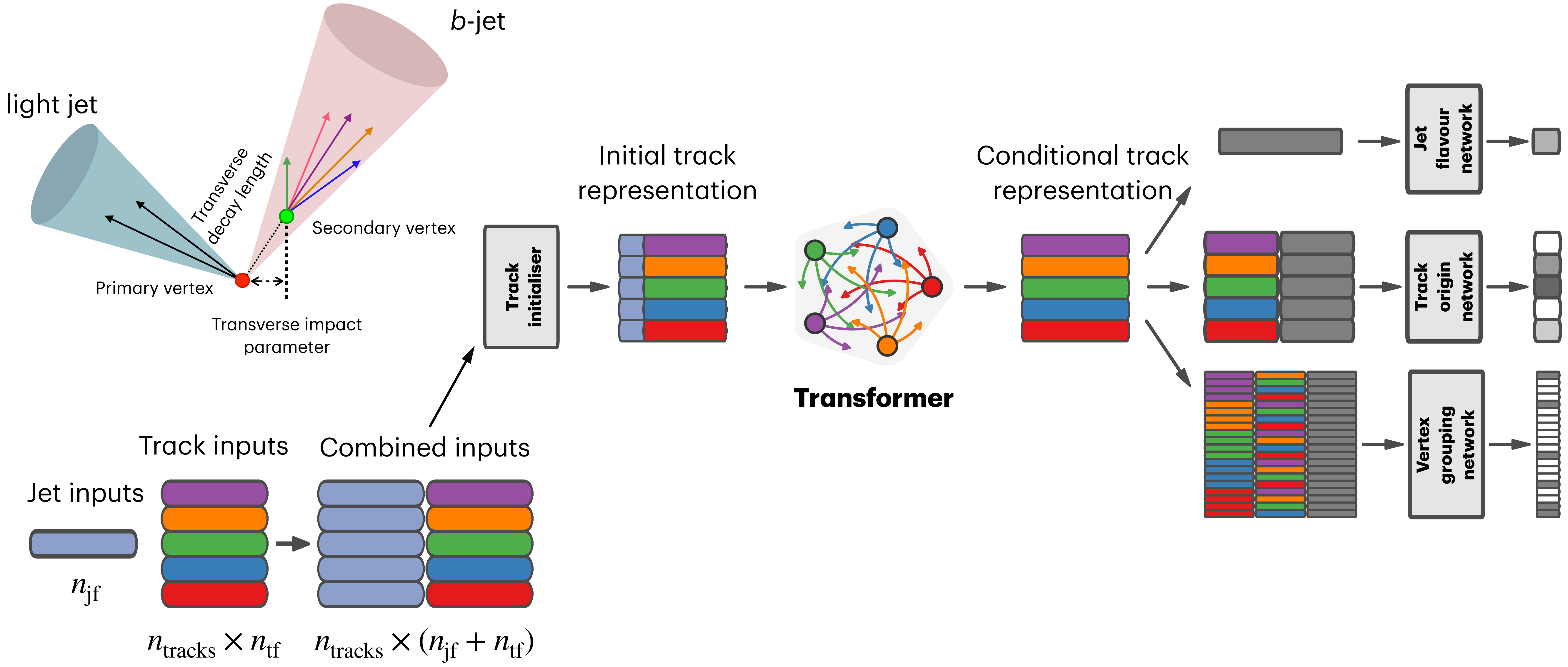
Courtesy Nicole Hartman

arxiv:2505.19689

Making sense of the remaining mess by tagging jets



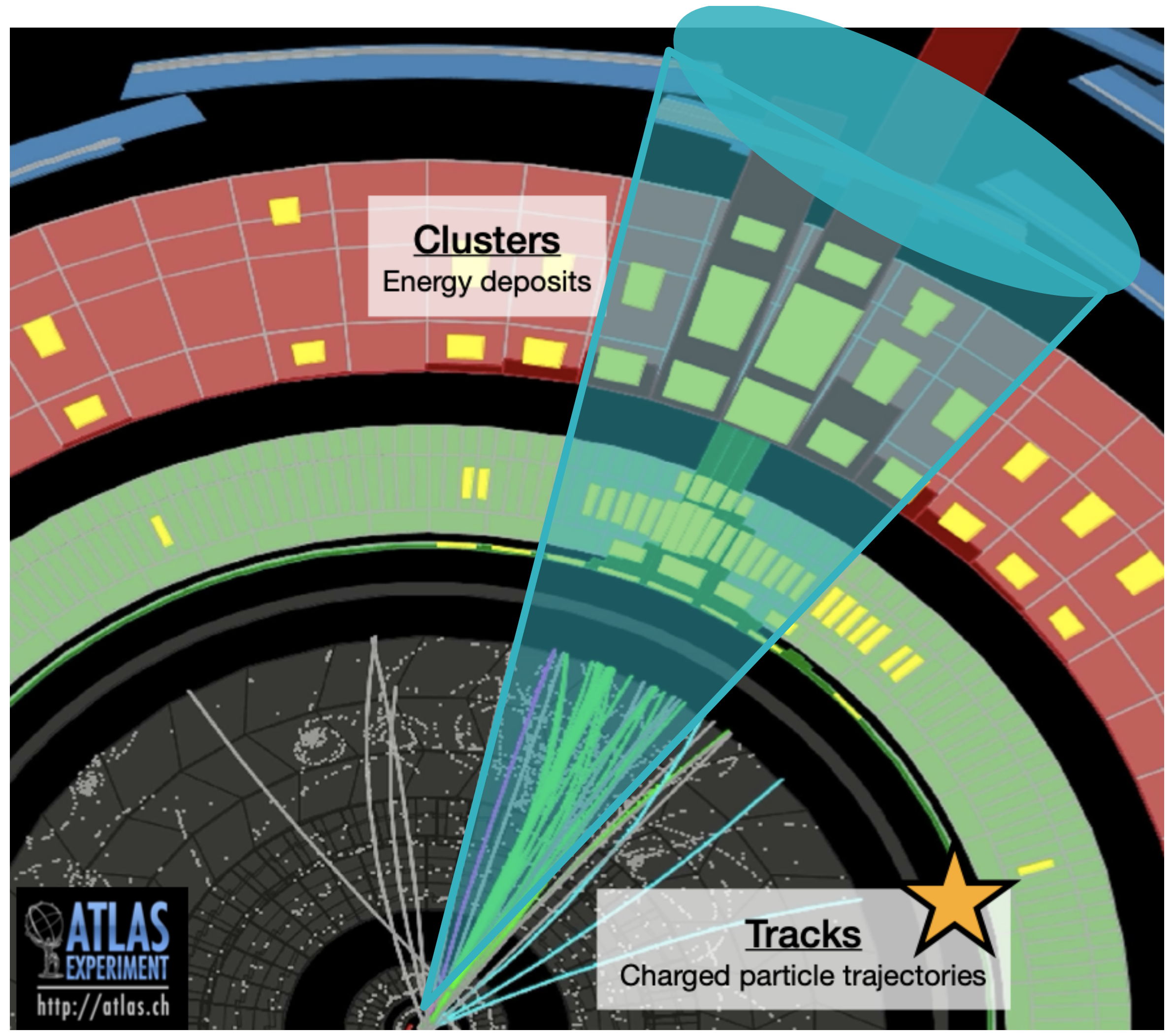
- Handle variable number of jets
- Calorimeter images tracker hits
- Multiple tagging objectives (b-tagging, c-tagging, τ-tagging, light jets),
- Auxiliary tasks, eg. track origin, vertex finding
- End-to-end training



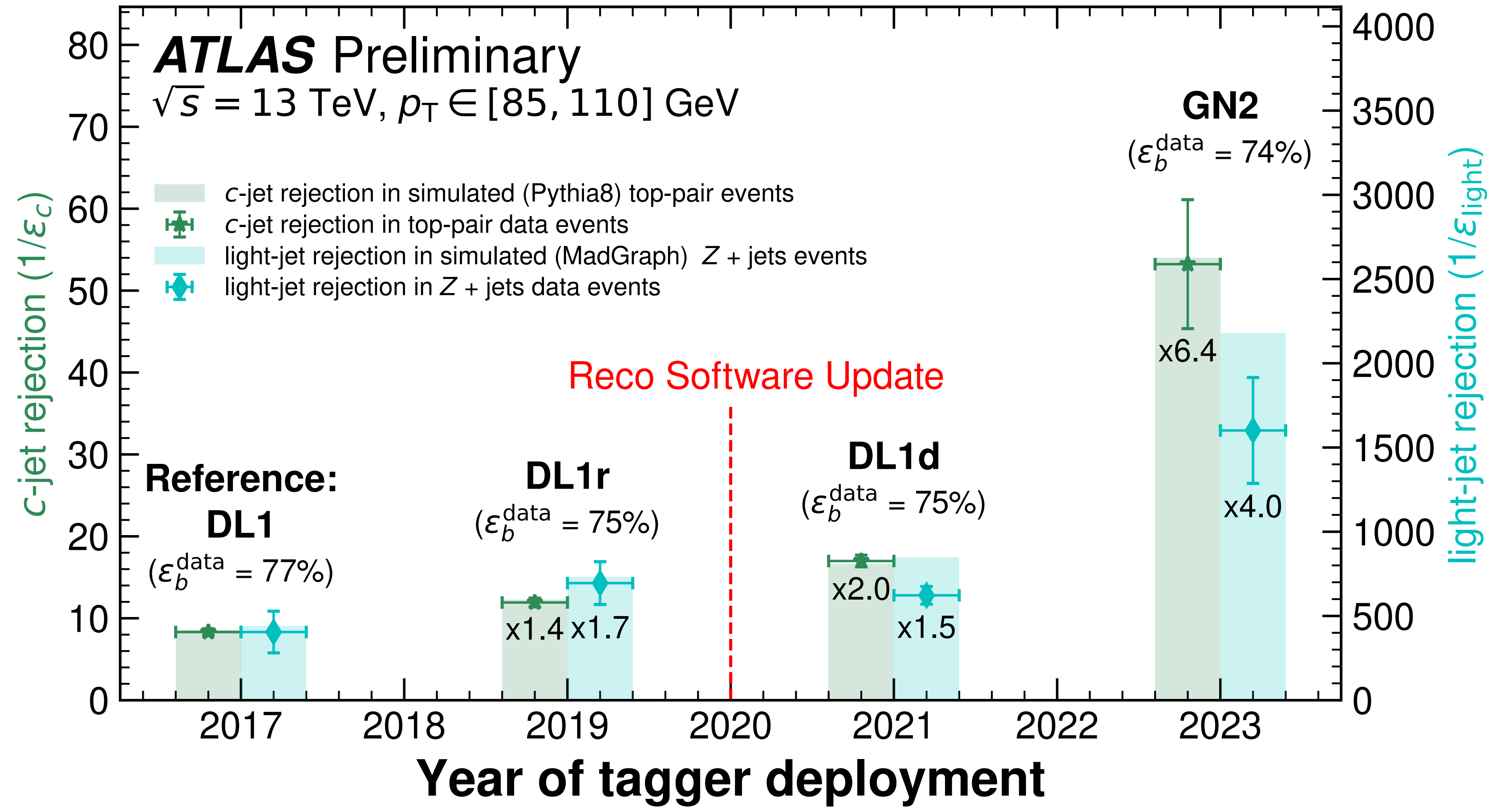
Courtesy Nicole Hartman

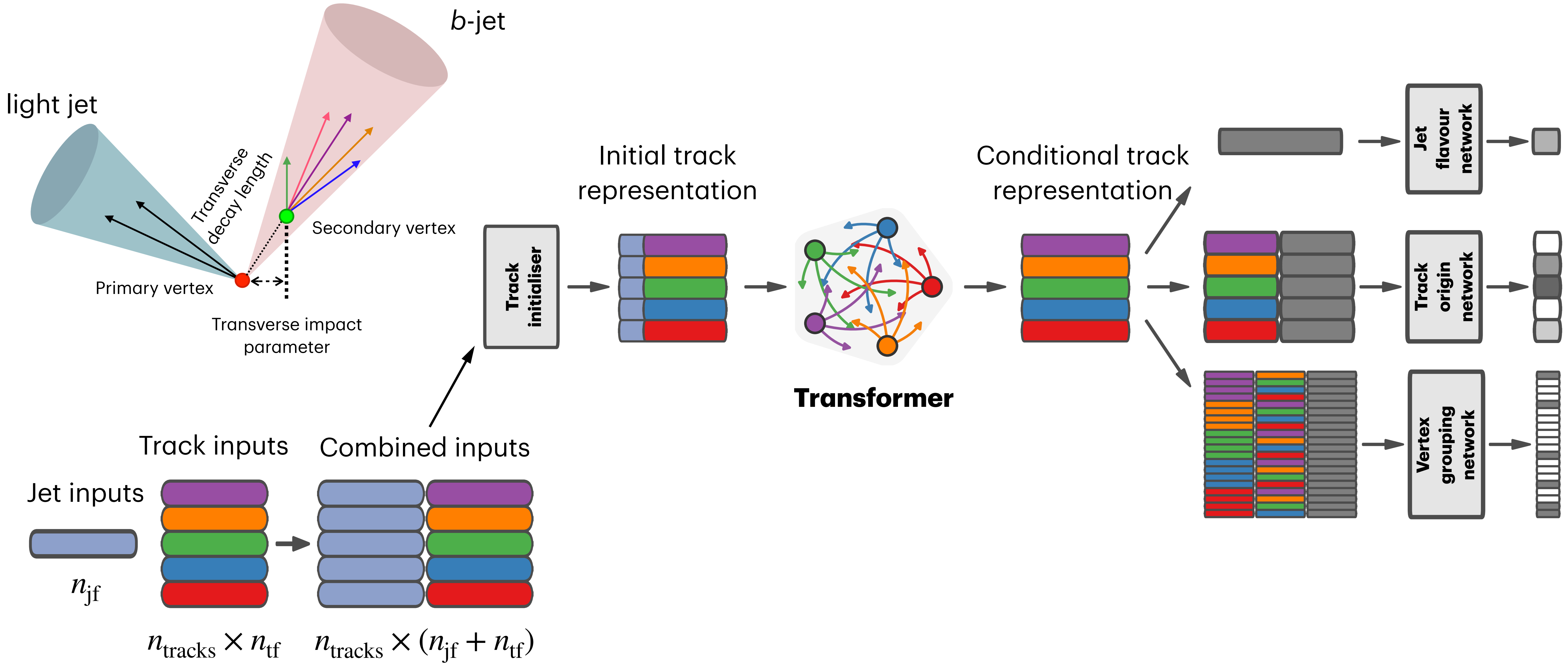
arxiv:2505.19689

Making sense of the remaining mess by tagging jets



- Handle variable number of jets
- Calorimeter images tracker hits
- Multiple tagging objectives (b-tagging, c-tagging, τ-tagging, light jets),
- Auxiliary tasks, eg. track origin, vertex finding
- End-to-end training

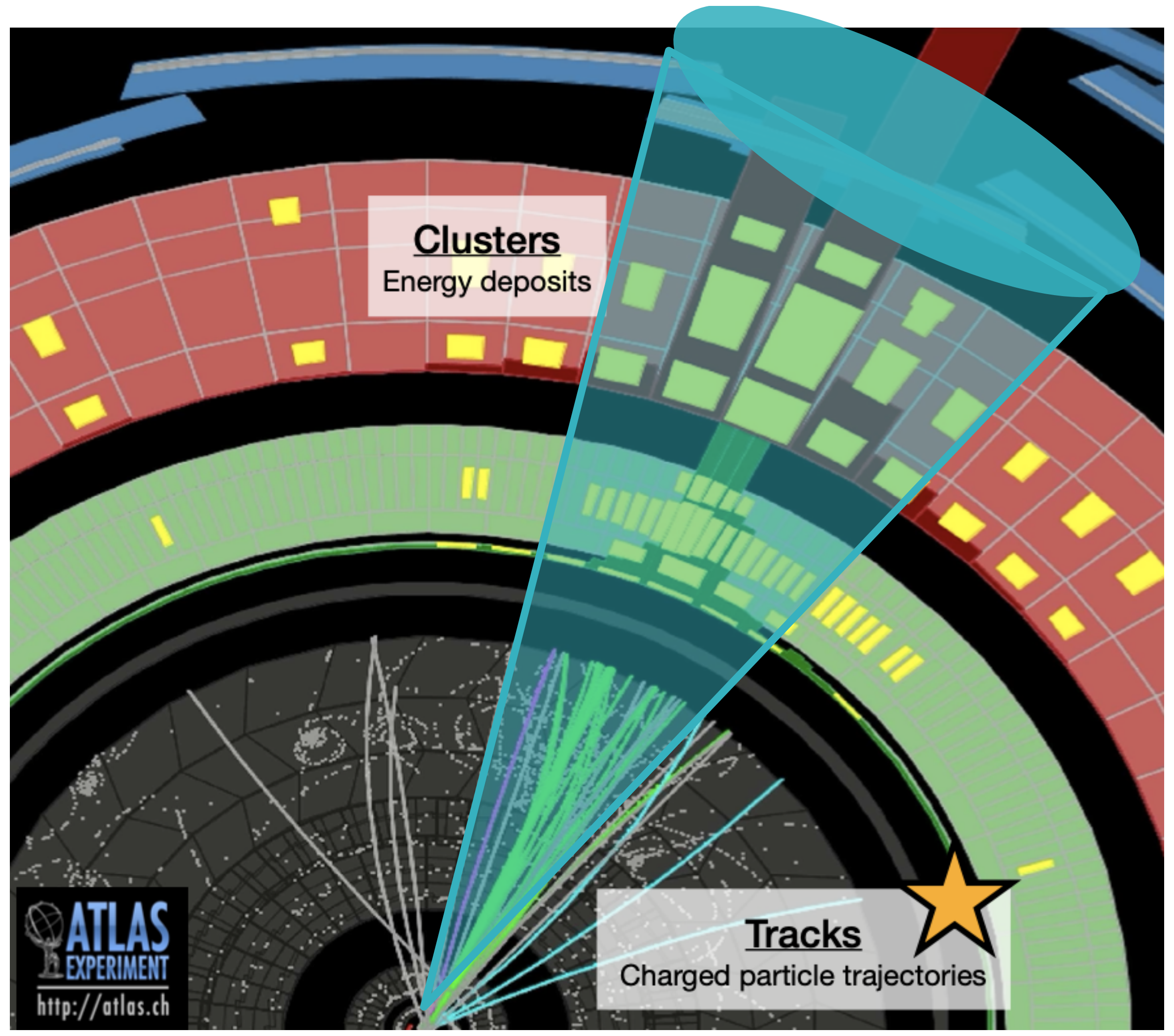




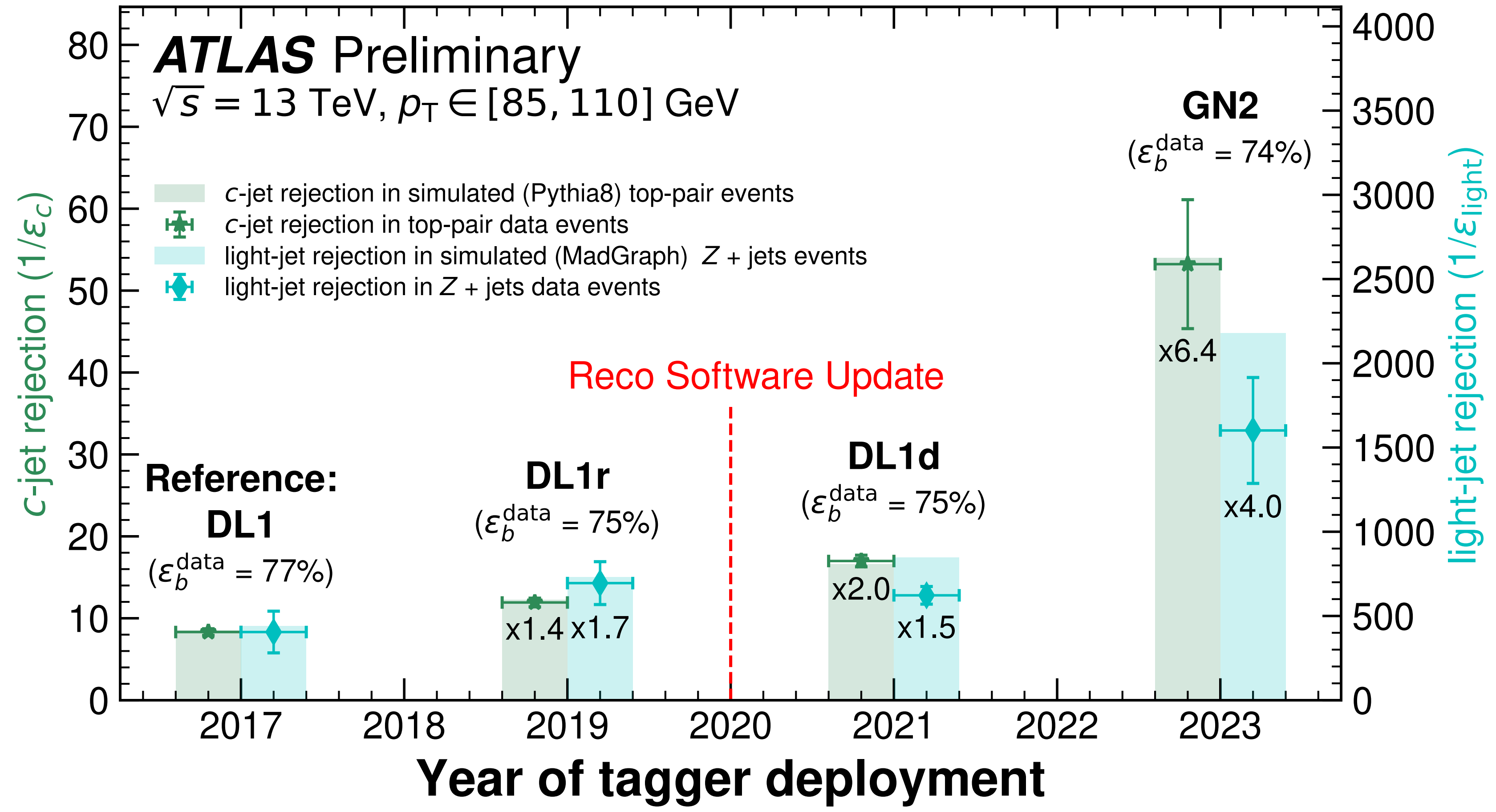
Courtesy Nicole Hartman

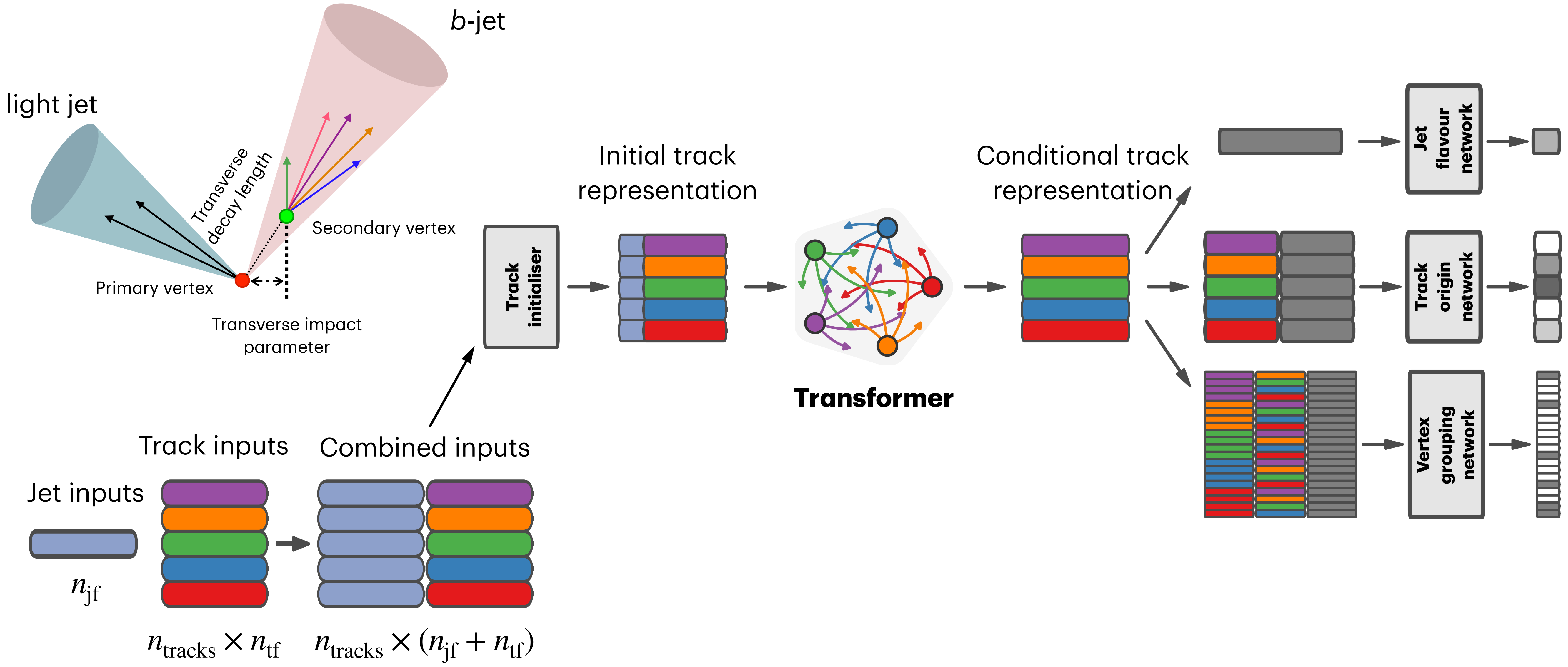
arxiv:2505.19689

Making sense of the remaining mess by tagging jets



- Handle variable number of jets
- Calorimeter images tracker hits
- Multiple tagging objectives (b-tagging, c-tagging, τ-tagging, light jets),
- Auxiliary tasks, eg. track origin, vertex finding
- End-to-end training



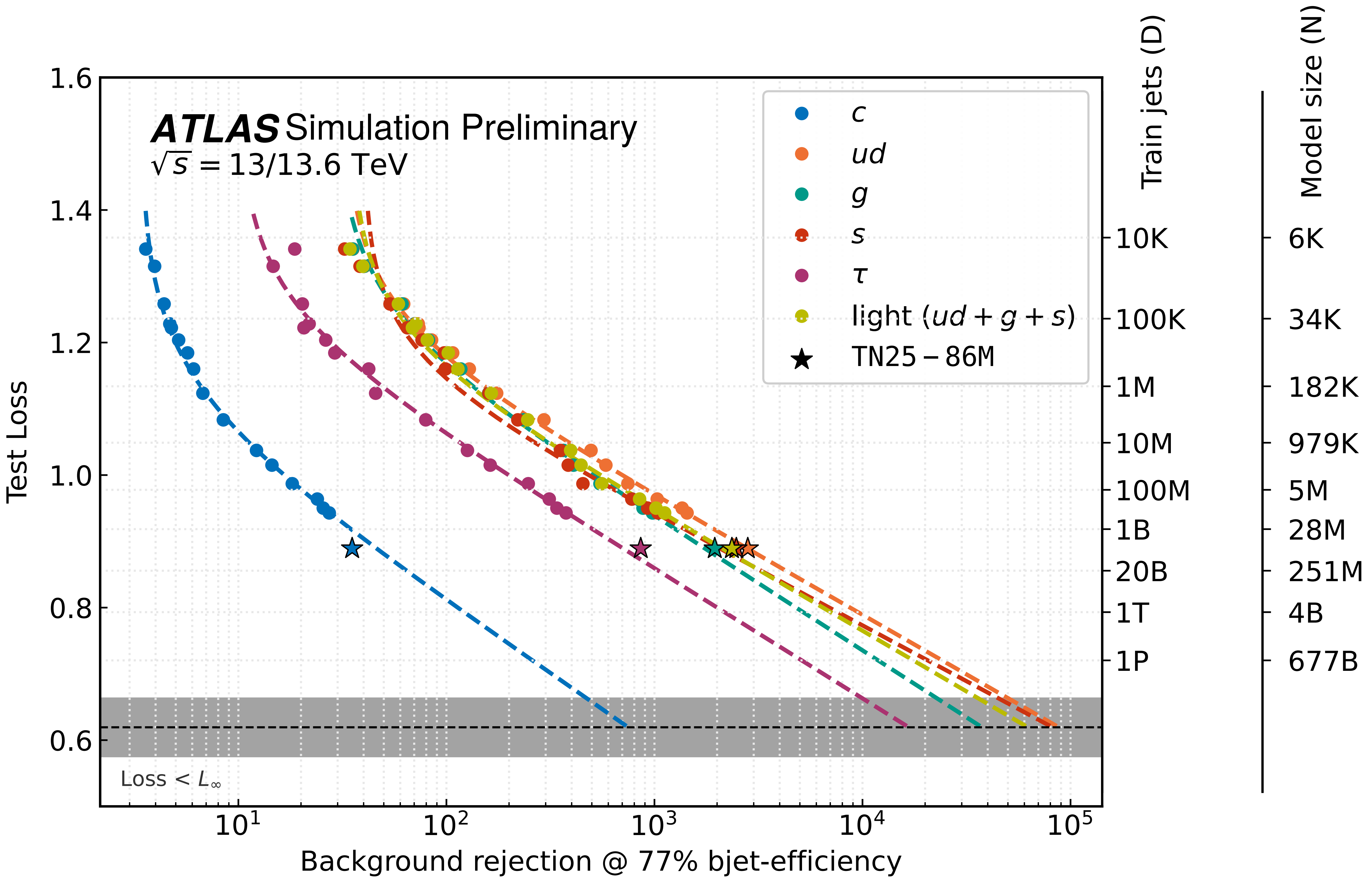


Do you think we squeezed all the juice out with our tagging algorithms now ?

Courtesy Nicole Hartman

arxiv:2505.19689

Can still get better with more data

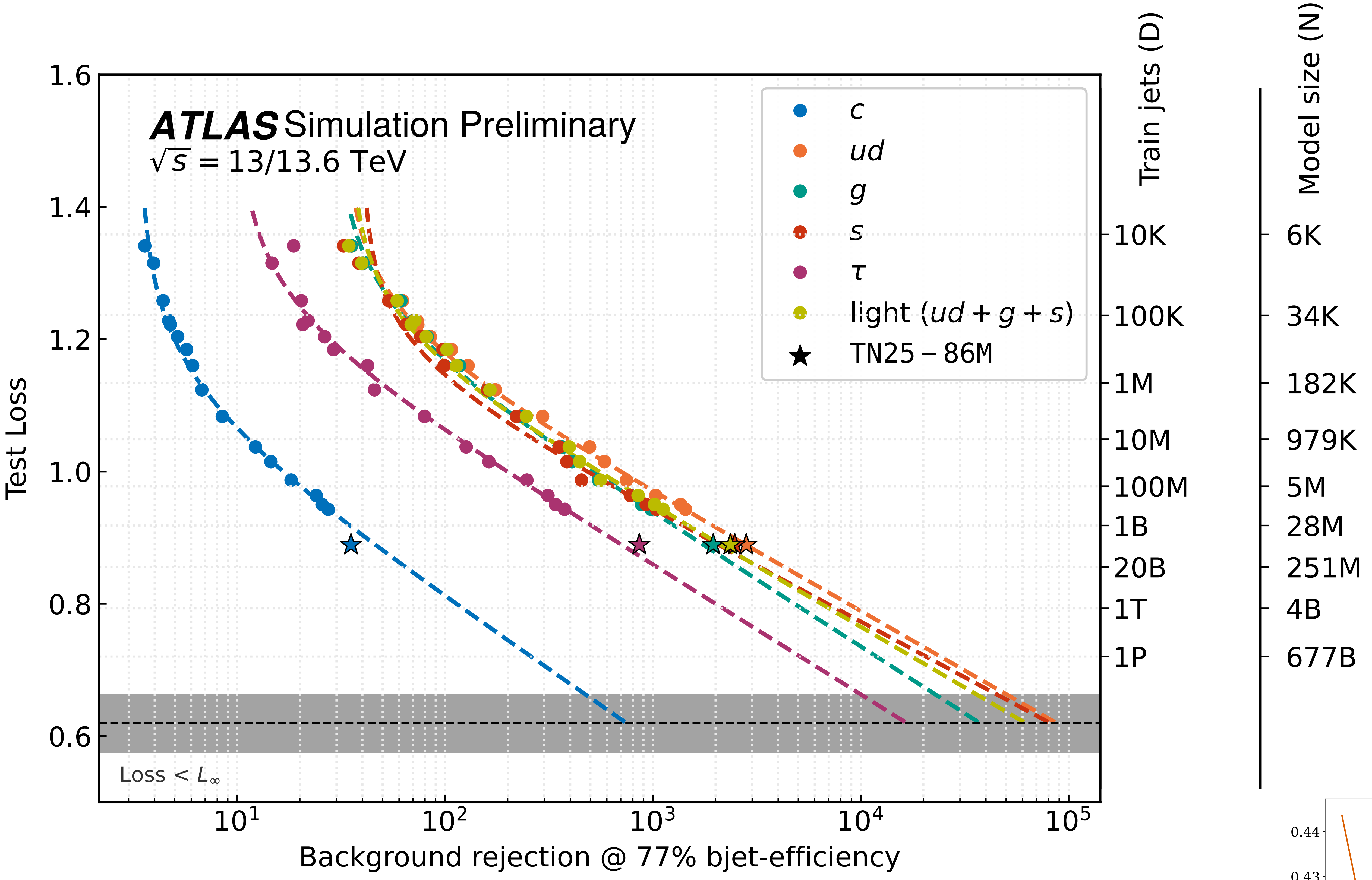


Scaling laws demonstrated with HEP data

Motivates cross-experiment and cross-task foundation models!

ATL-SOFT-PUB-2026-002

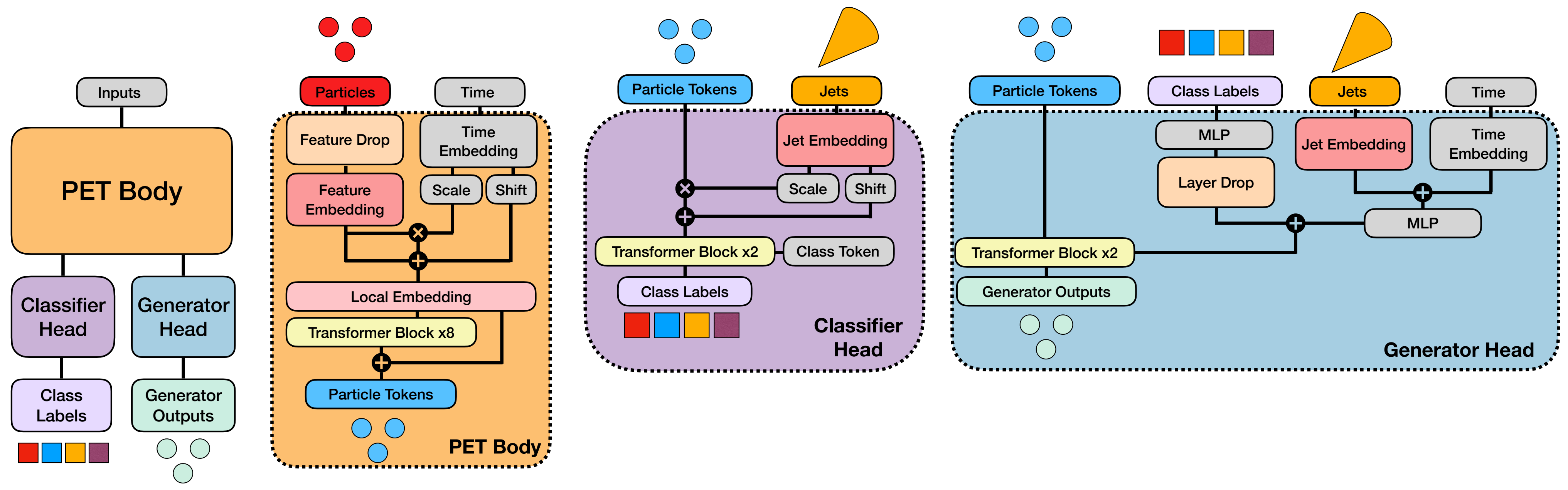
Can still get better with more data



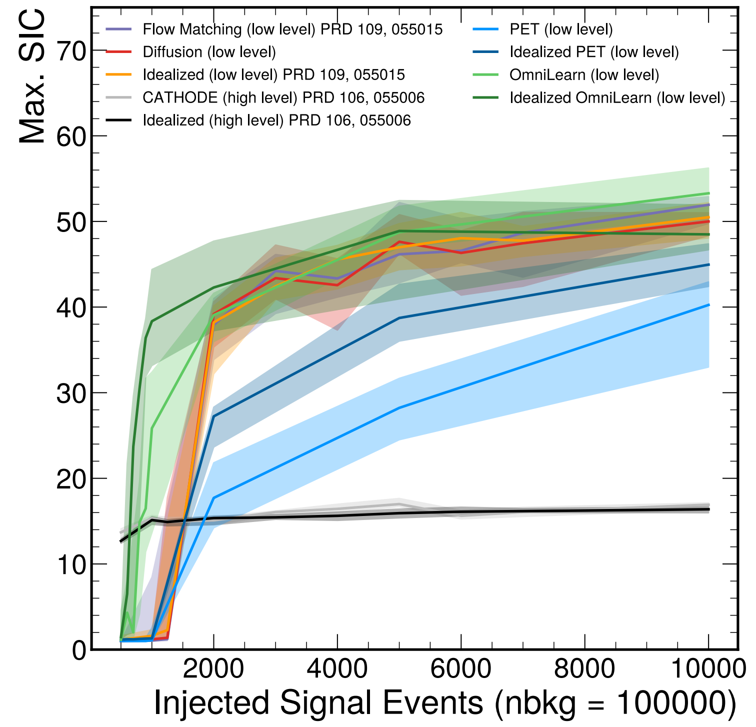
Scaling laws demonstrated with HEP data

Motivates cross-experiment and cross-task foundation models!

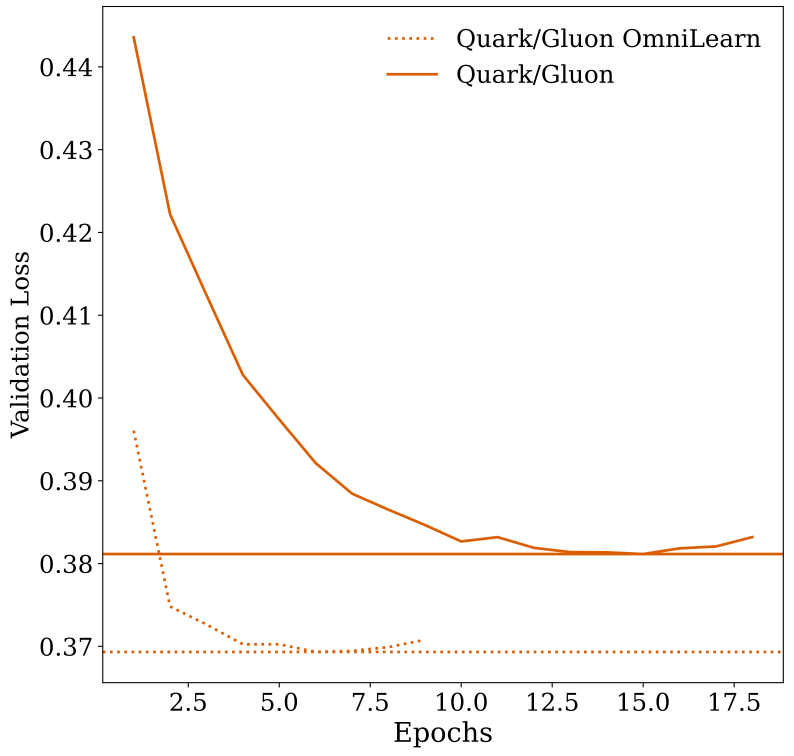
Eg. OmniLearn architecture



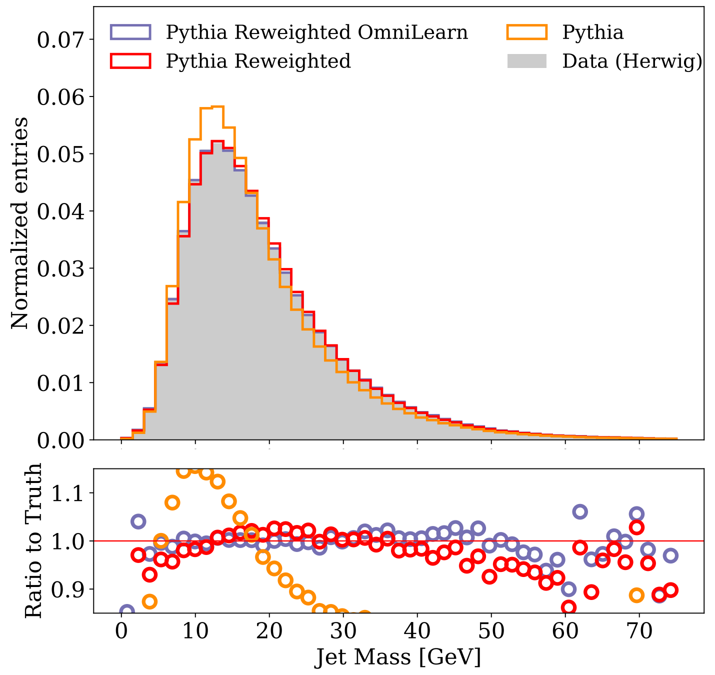
Anomaly detection



Jet tagging



Unfolding

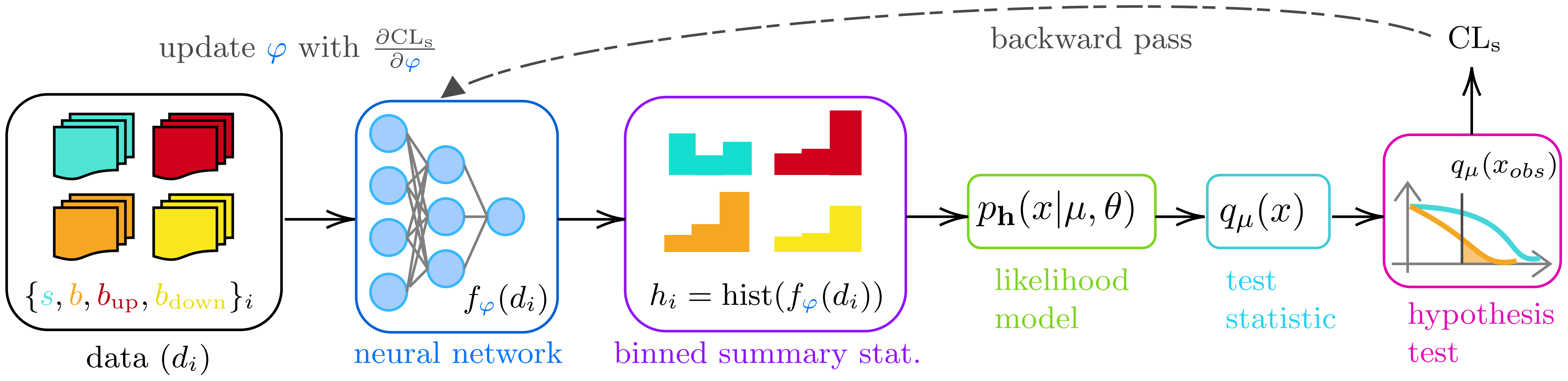


ATL-SOFT-PUB-2026-002

Mukuni et al

Or is the future more physics-informed ML?

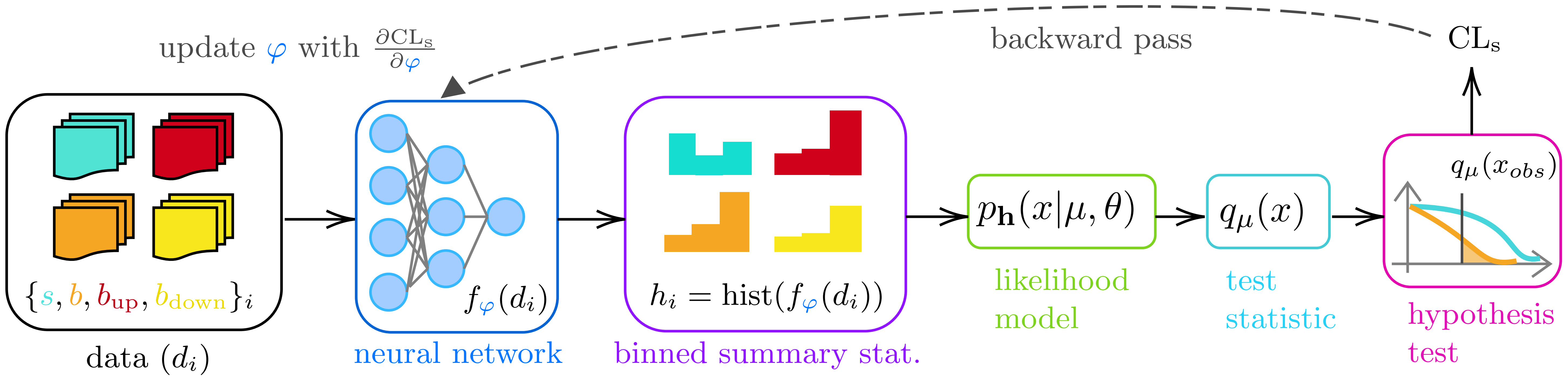
- End-to-end differentiable analysis takes the opposite route
- Takes traditional physics analysis and makes it a giant network



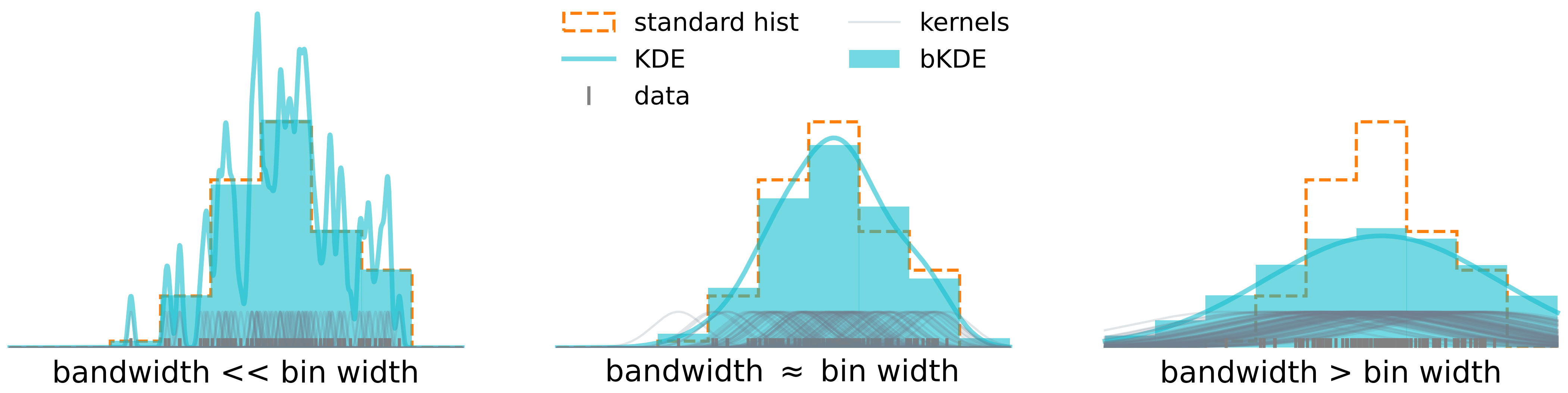
Simpson et al

Or is the future more physics-informed ML?

- End-to-end differentiable analysis takes the opposite route
- Takes traditional physics analysis and makes it a giant network



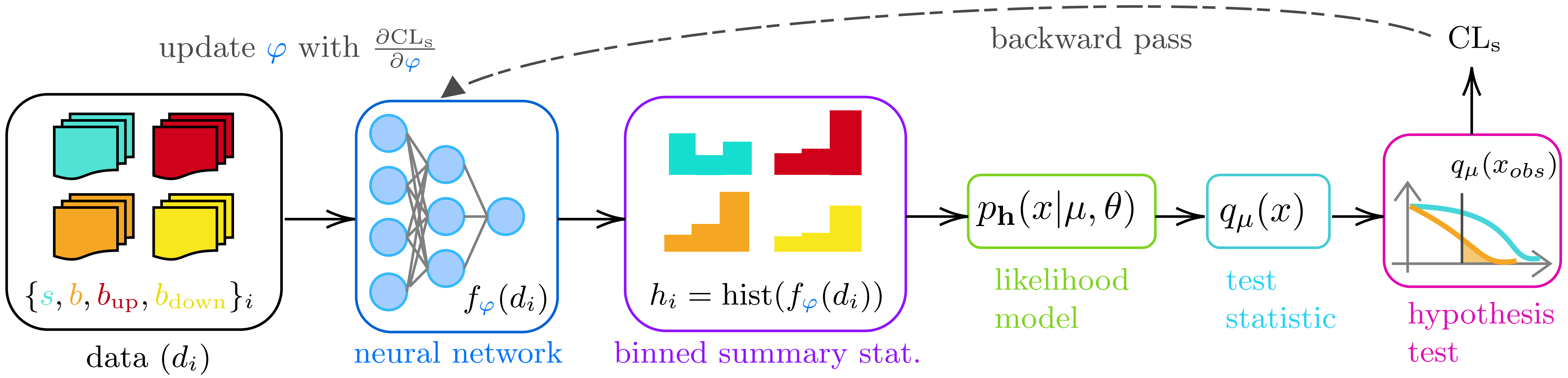
- Requires careful gradient relaxation techniques to handle non-differentaible operations

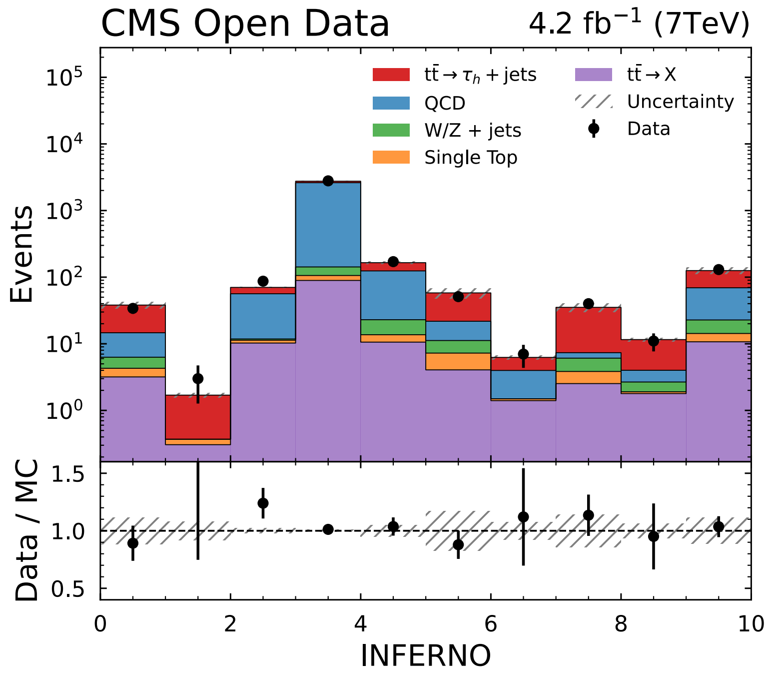


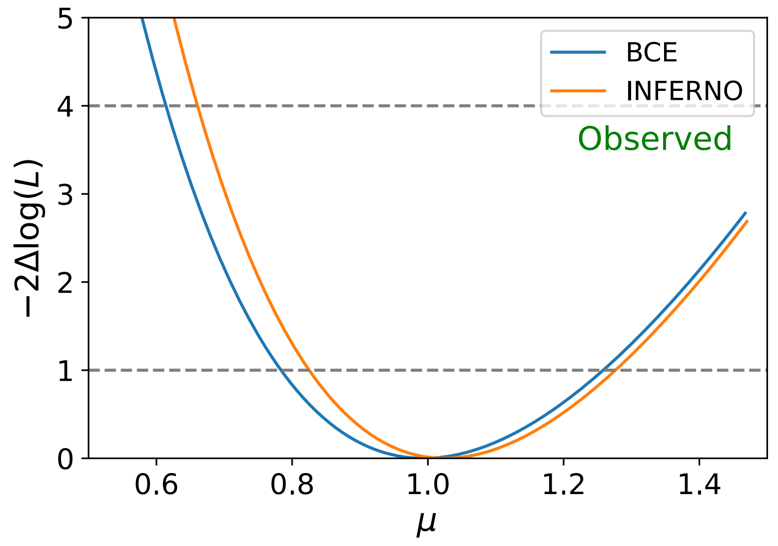
Simpson et al

Or is the future more physics-informed ML?

- End-to-end differentiable analysis takes the opposite route
- Takes traditional physics analysis and makes it a giant network



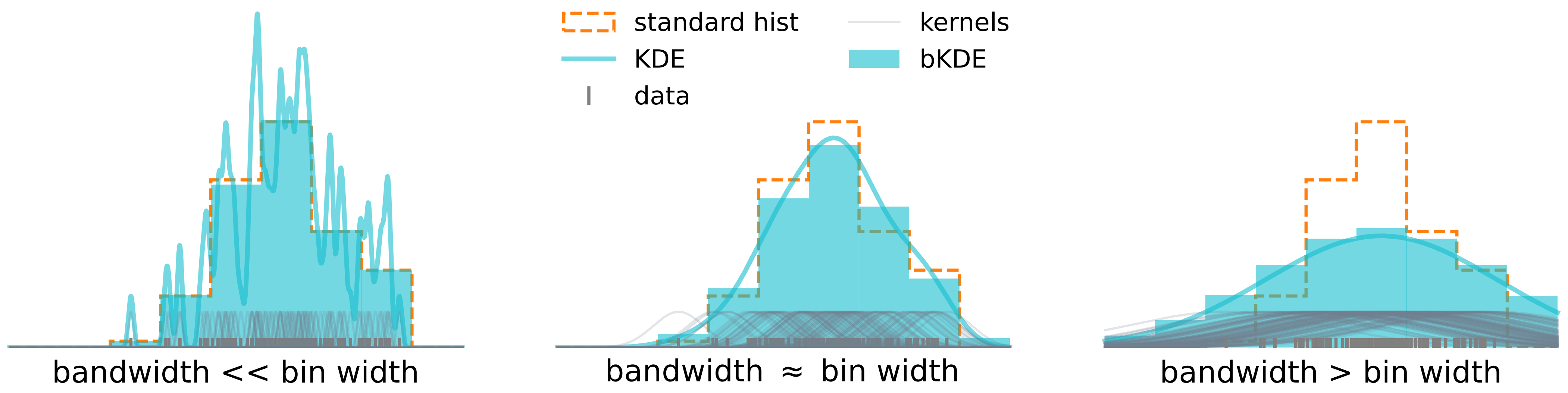




Layer et al

Let's you design more sensitive analyses

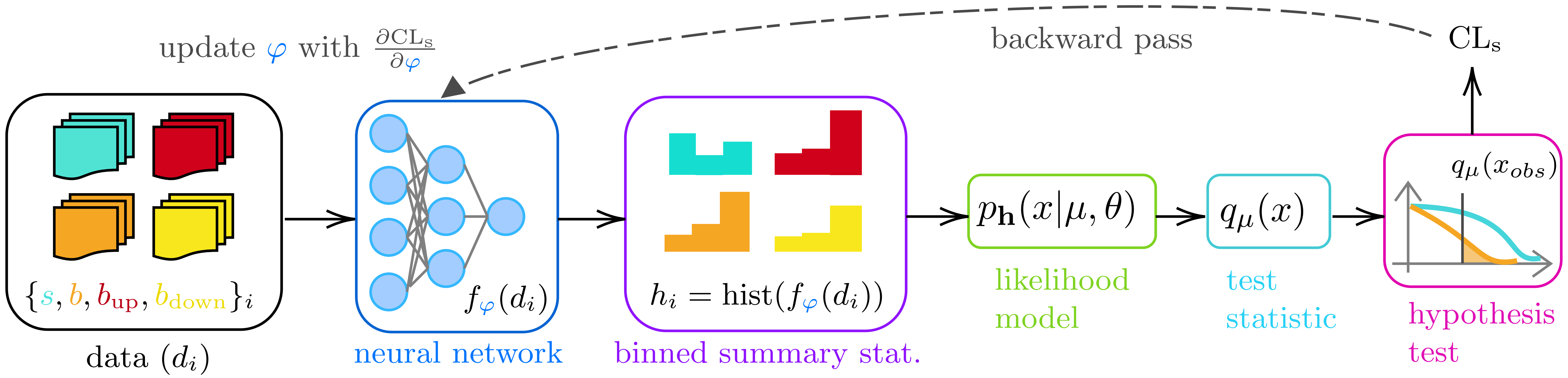
- Requires careful gradient relaxation techniques to handle non-differentaible operations



Simpson et al

Or is the future more physics-informed ML?

- End-to-end differentiable analysis takes the opposite route
- Takes traditional physics analysis and makes it a giant network



update φ with $\frac{\partial \mathrm{CL_s}}{\partial \varphi}$

backward pass

$\mathrm{CL_s}$

$\{s, b, b_{\mathrm{up}}, b_{\mathrm{down}}\}_i$

data (d_i)

$f_\varphi(d_i)$

neural network

$h_i = \mathrm{hist}(f_\varphi(d_i))$

binned summary stat.

$p_{\mathbf{h}}(x|\mu,\theta)$

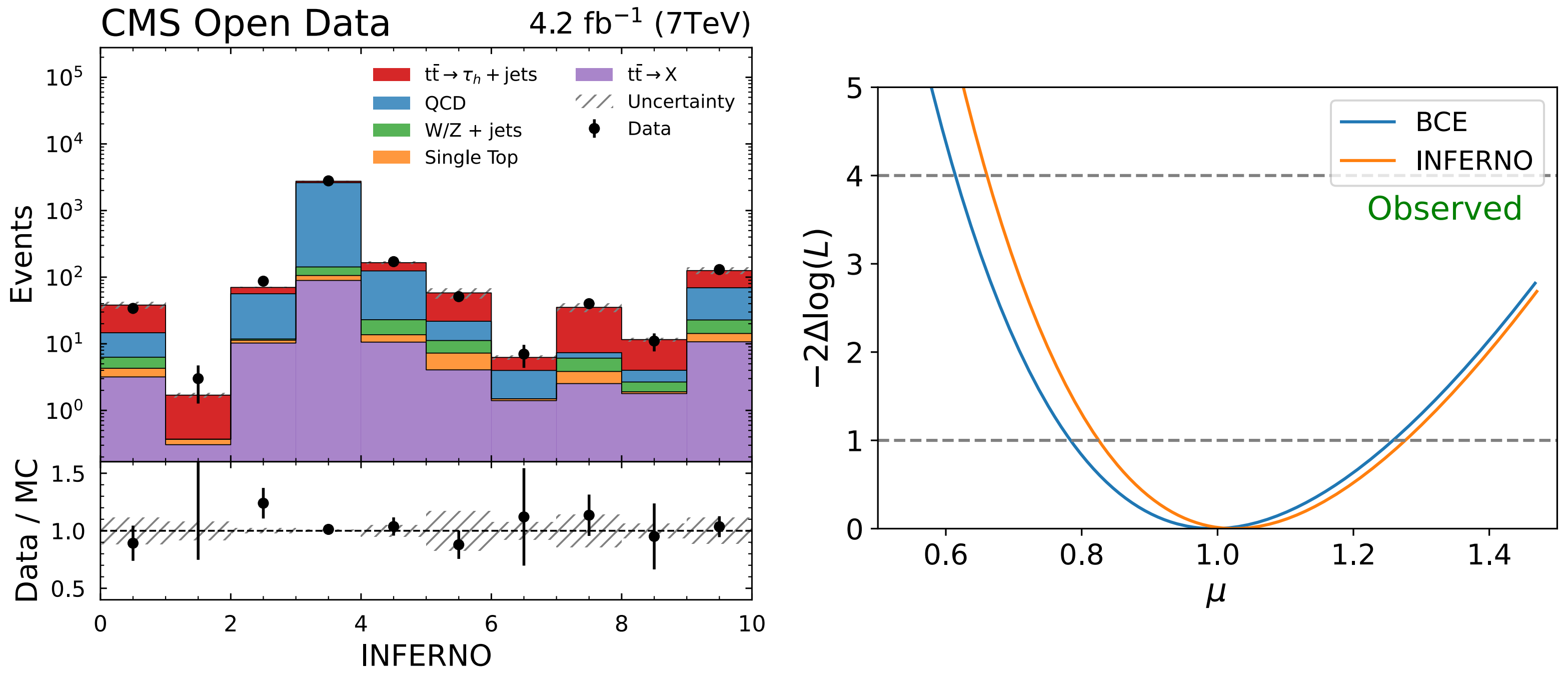
likelihood model

$q_\mu(x)$

test statistic

$q_\mu(x_{obs})$

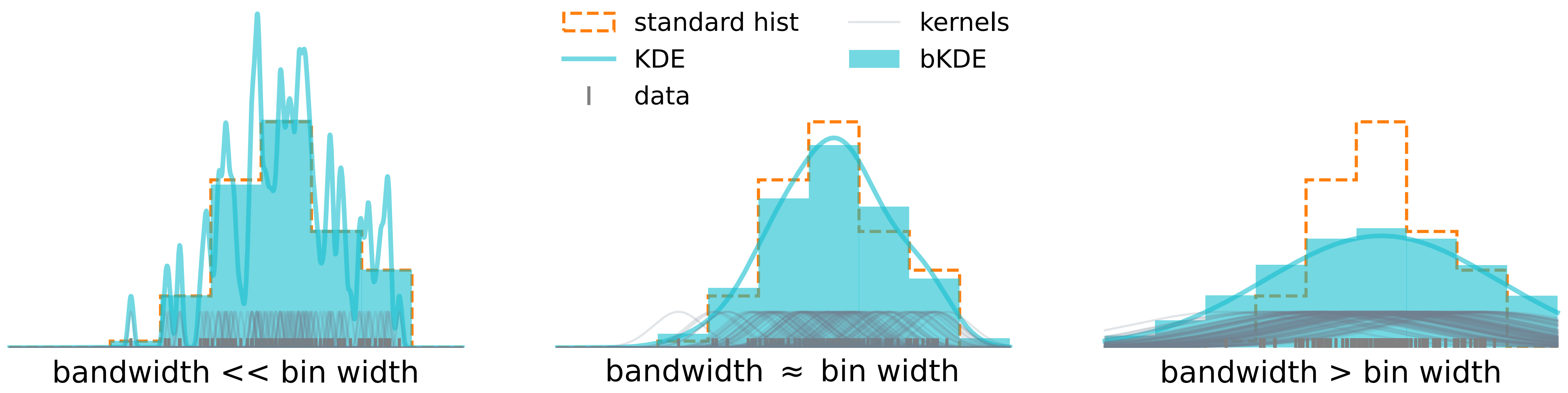
hypothesis test



Layer et al

Let’s you design more sensitive analyses

- Requires careful gradient relaxation techniques to handle non-differentaible operations

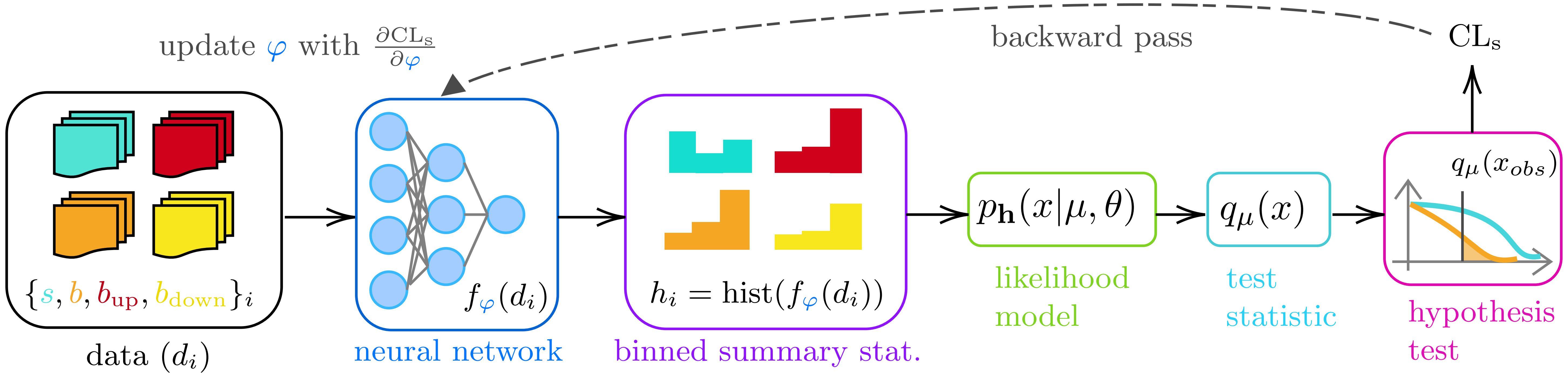


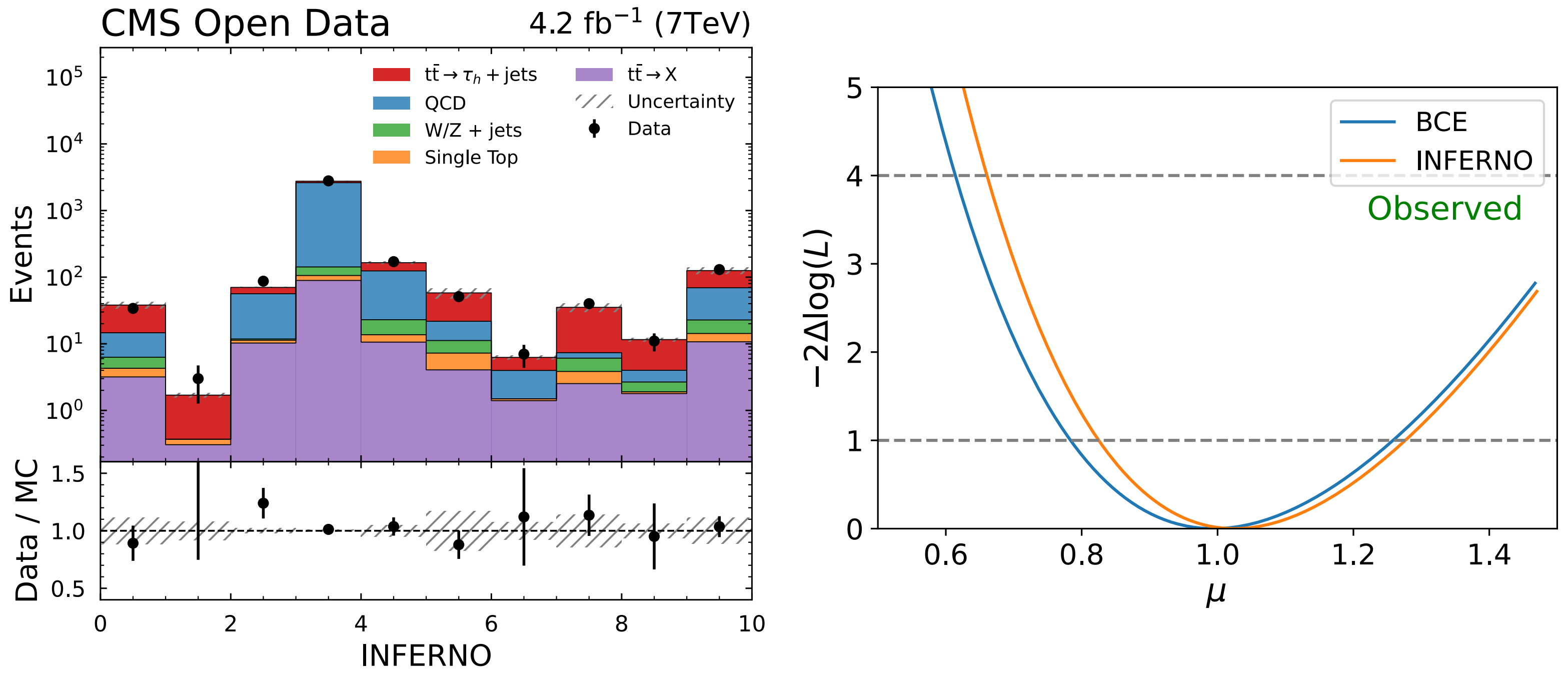
What do you think is the future?

Simpson et al

Or is the future more physics-informed ML?

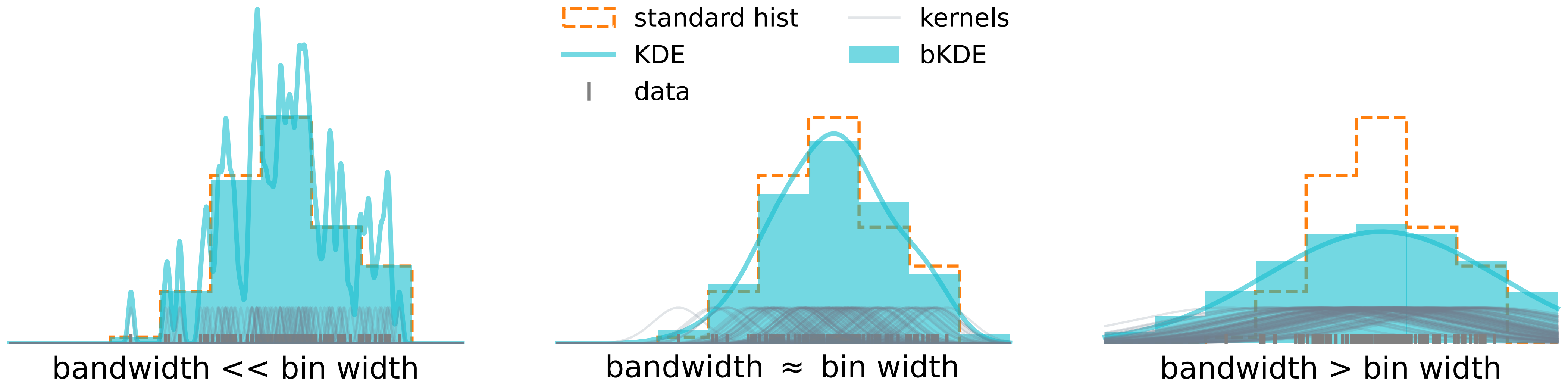
- End-to-end differentiable analysis takes the opposite route
- Takes traditional physics analysis and makes it a giant network





Layer et al

Let's you design more sensitive analyses



- Requires careful gradient relaxation techniques to handle non-differentaible operations

What do you think is the future?

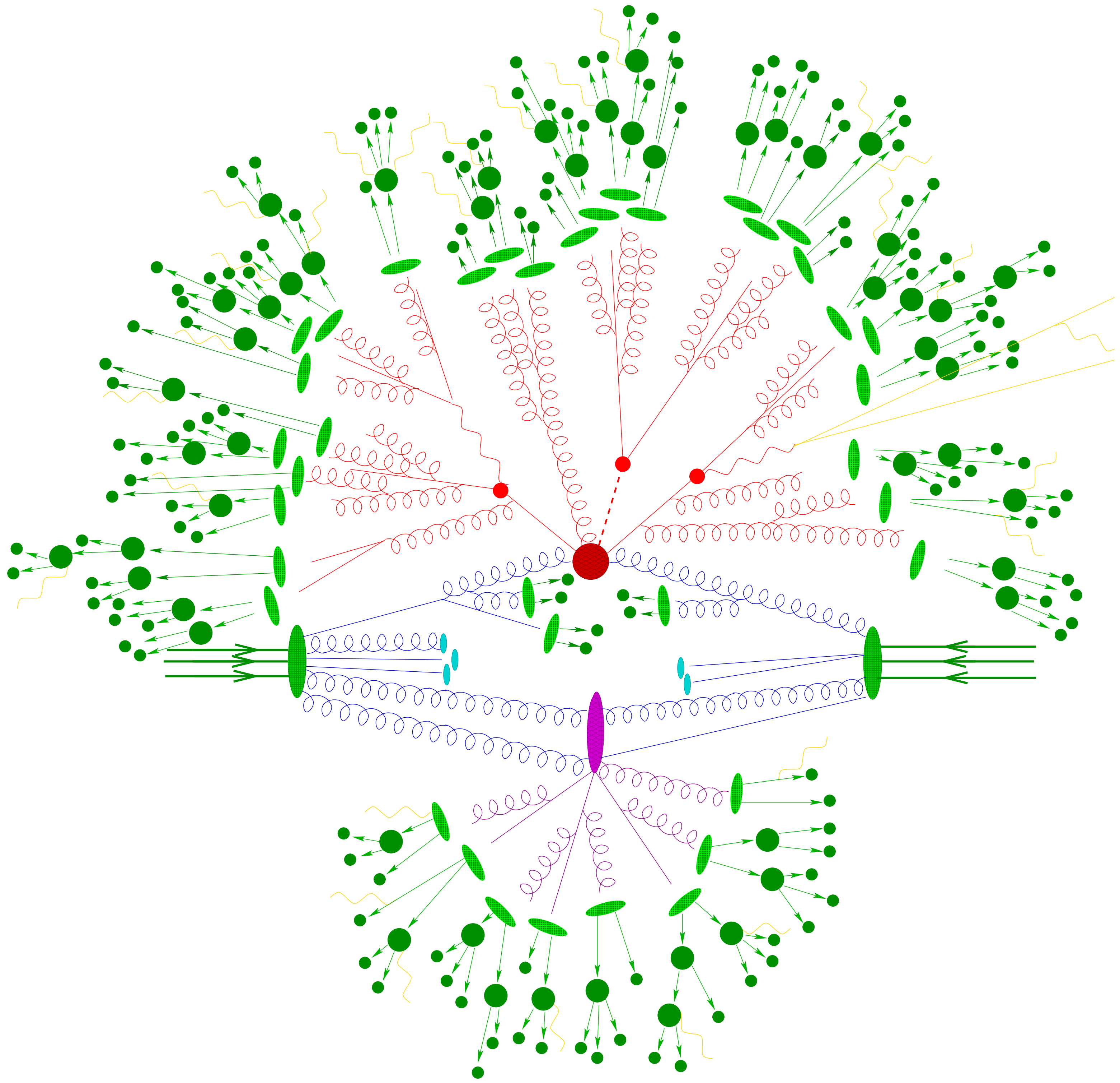
Learn in Yifan's diff prog talk!

Simpson et al

Hypothesis generation → hypothesis testing in experiments

First need to collect data and simulations

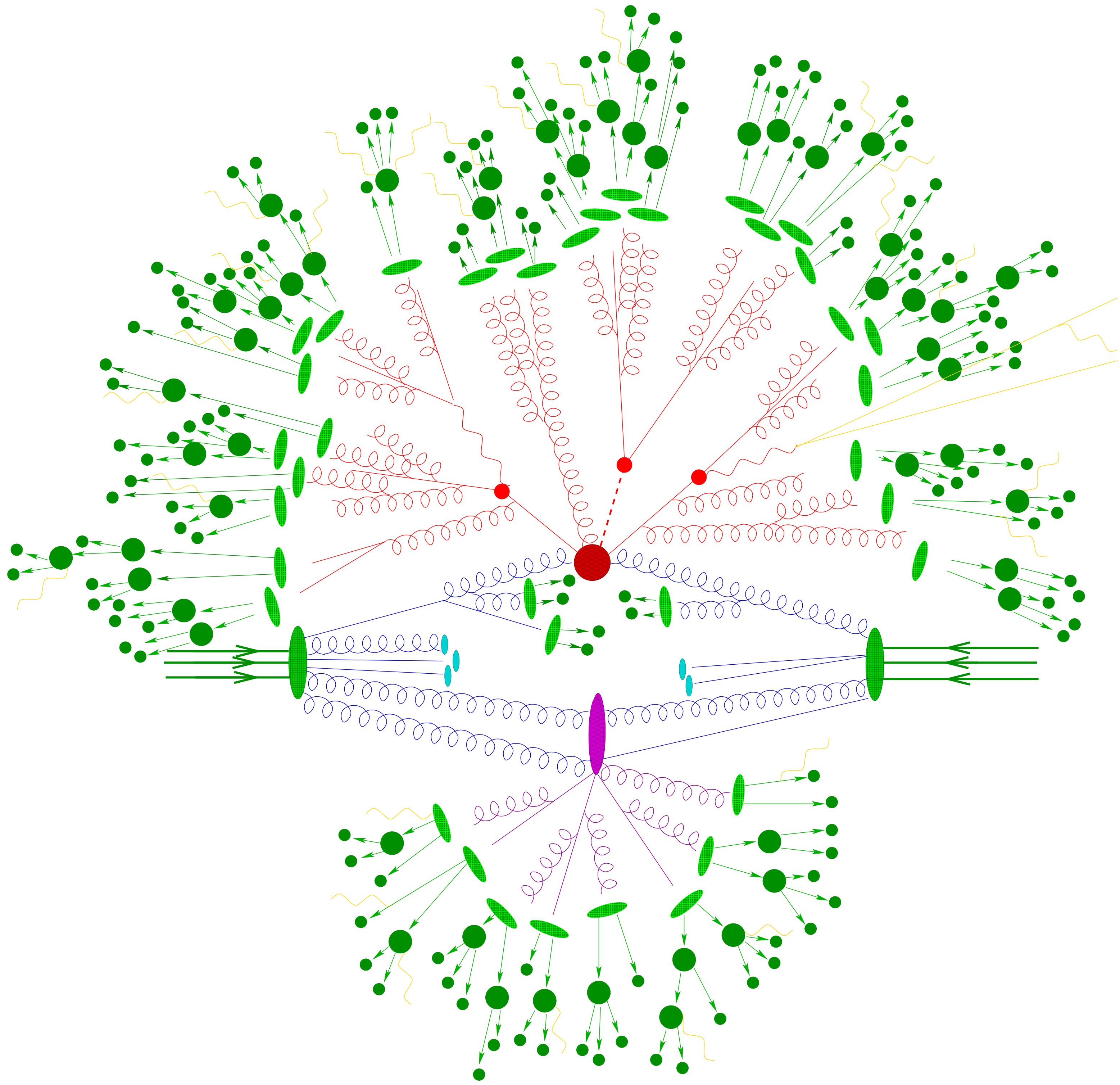
Detector Simulations



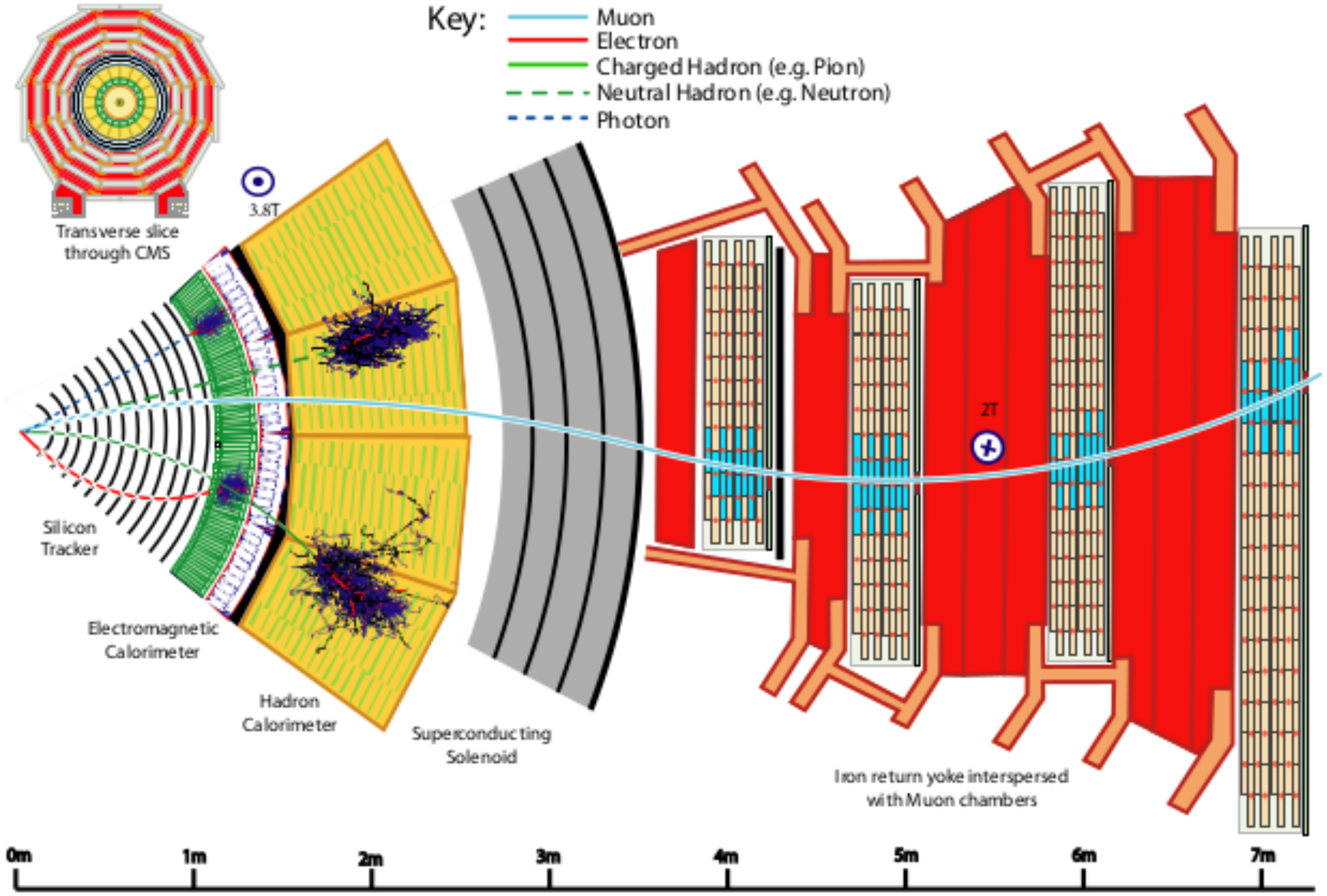
The physics

Detector interaction
With Geant4

Detector Simulations

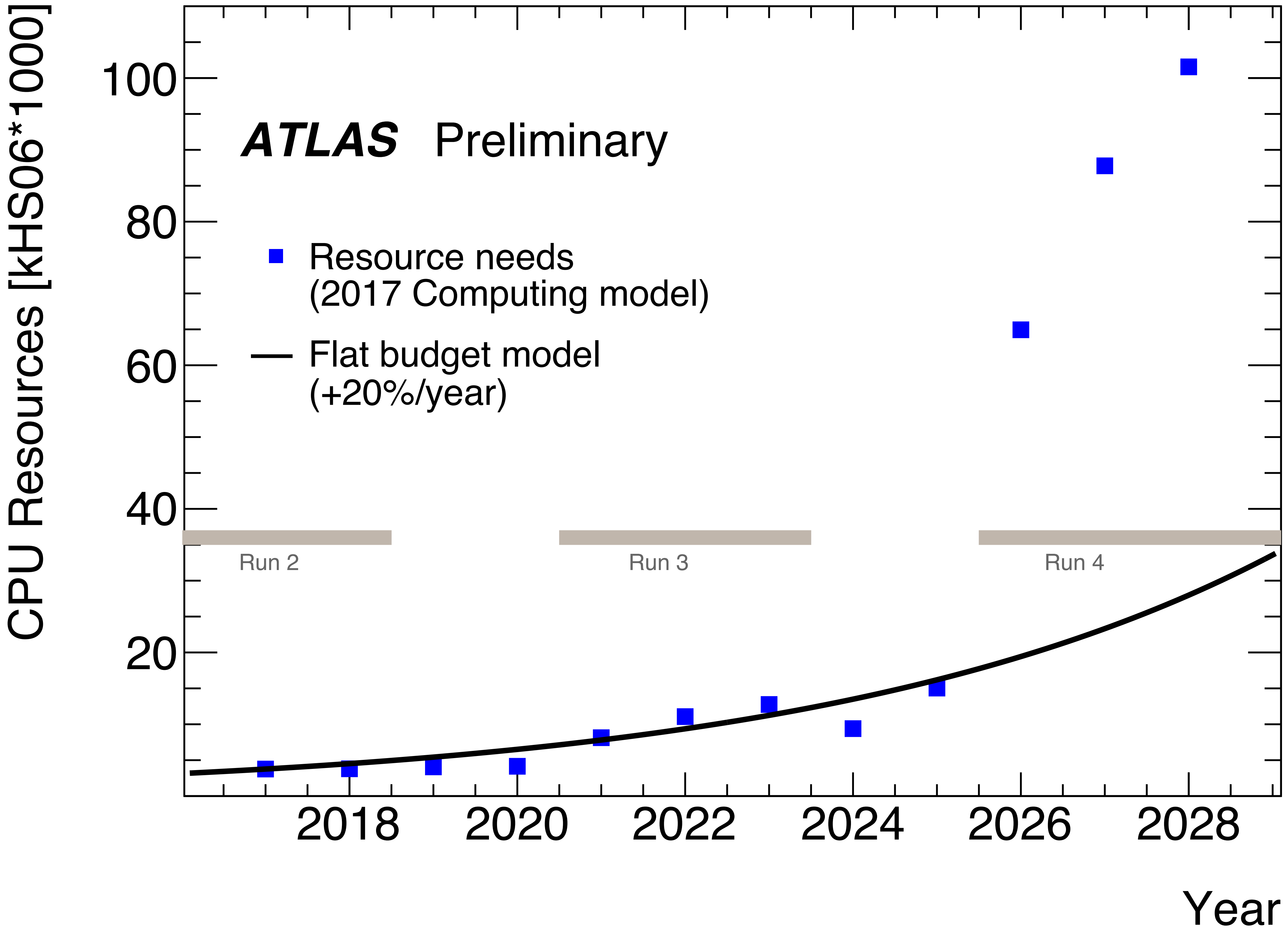


The physics



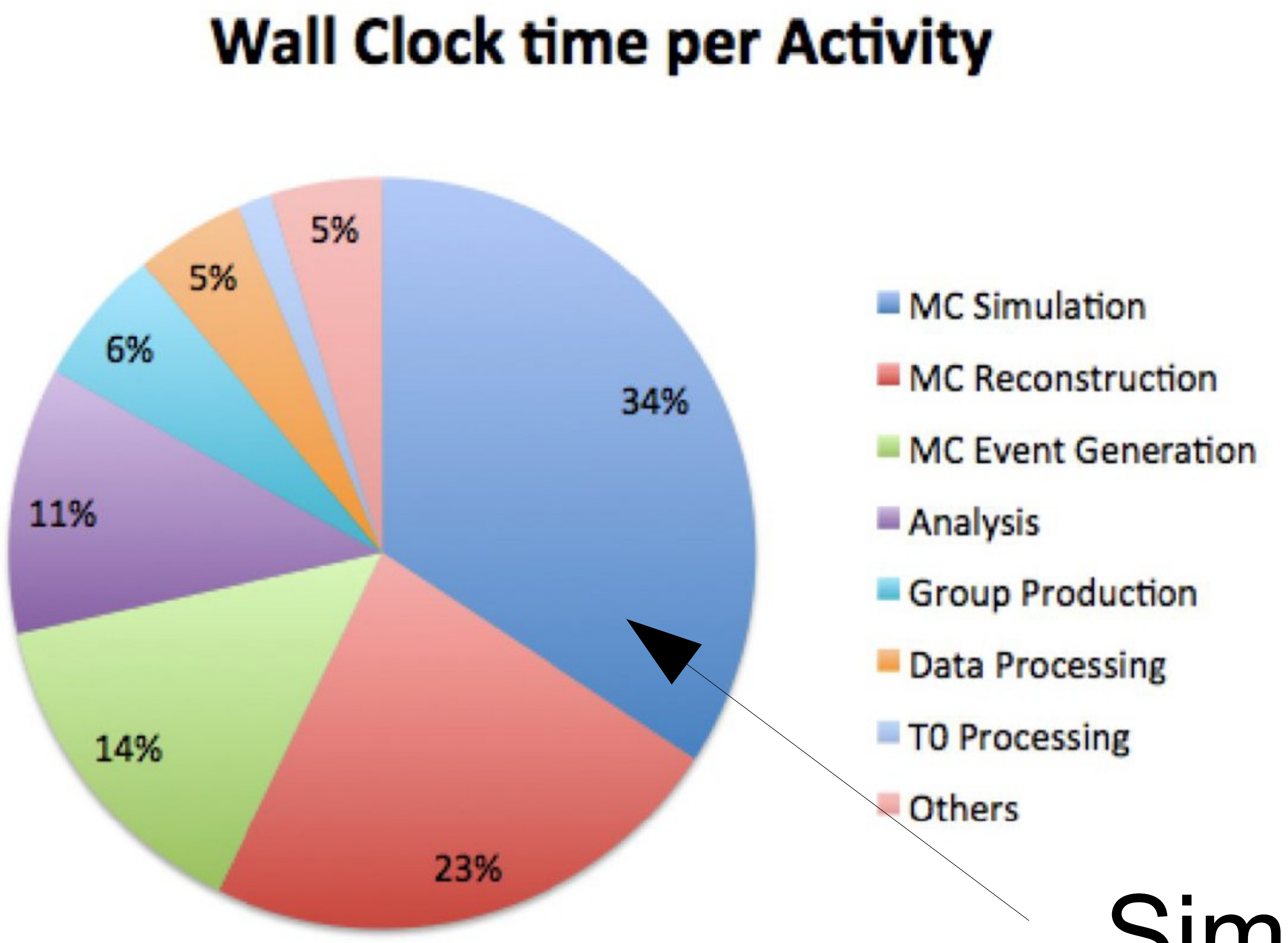
Detector interaction
With Geant4

Don’t have enough compute for calorimeter simulations



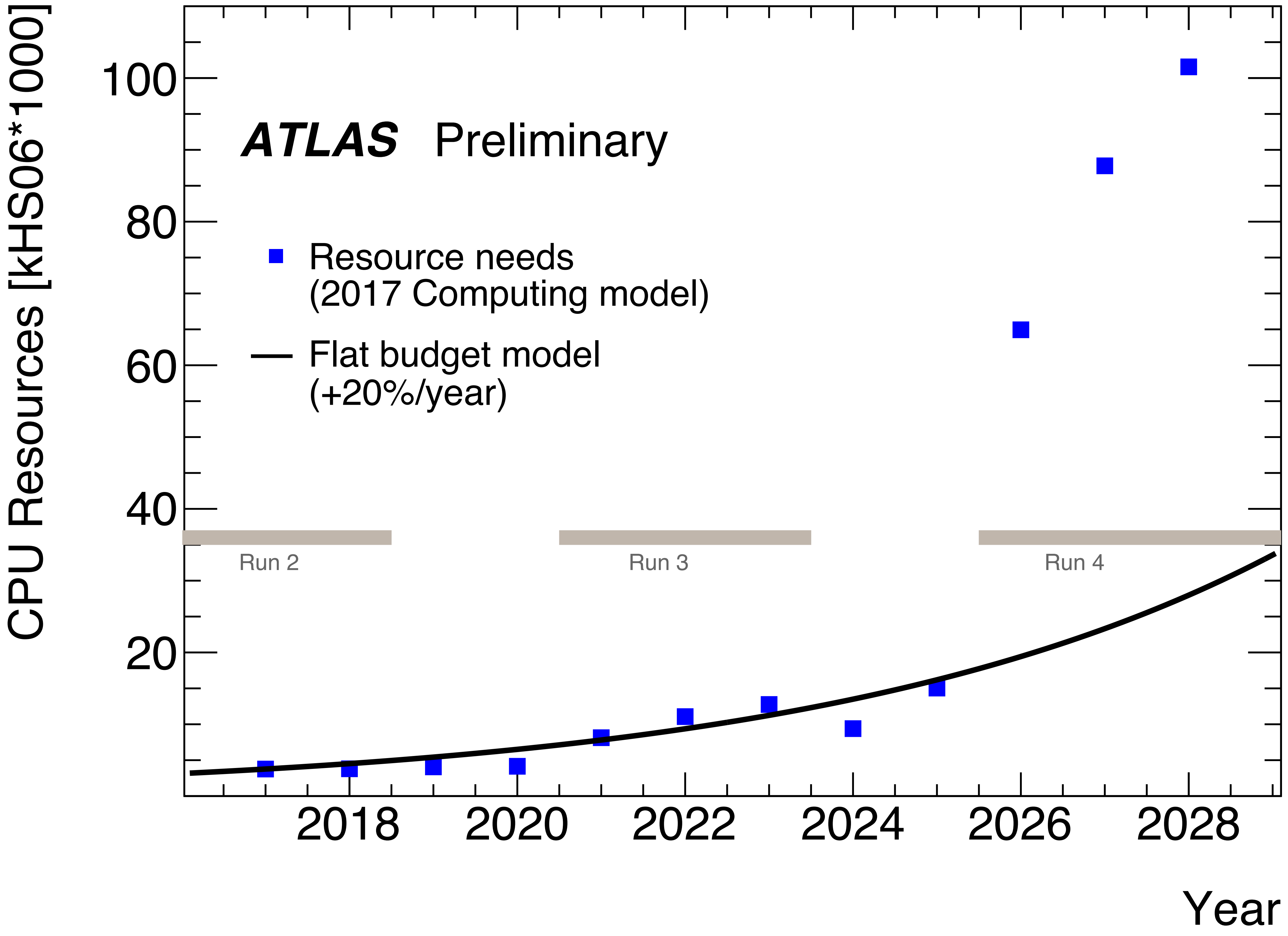
[Source](#) (might be old by now)

Don’t have enough compute for calorimeter simulations



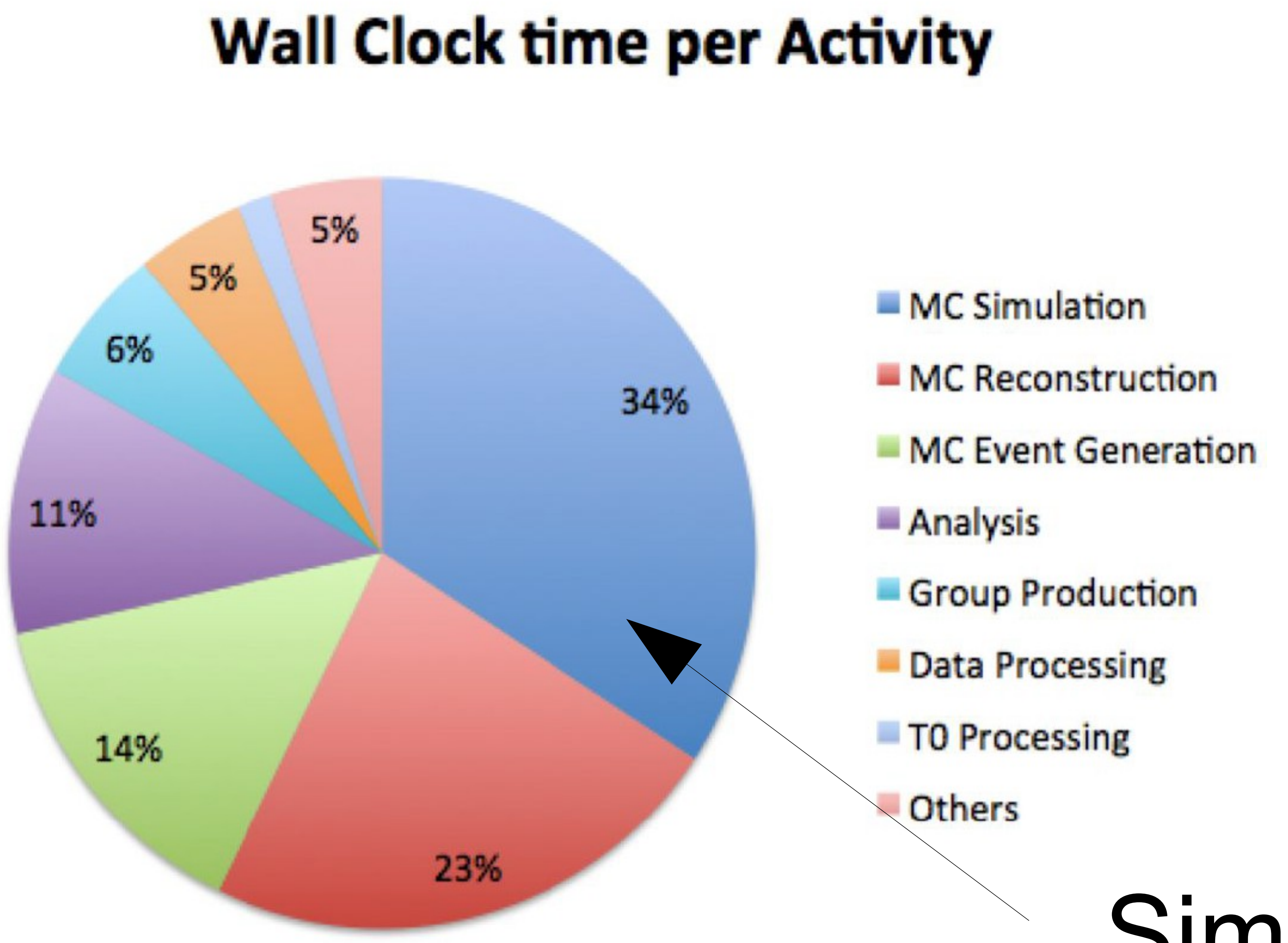
ATLAS 2016 numbers

Simulation!



Source (might be old by now)

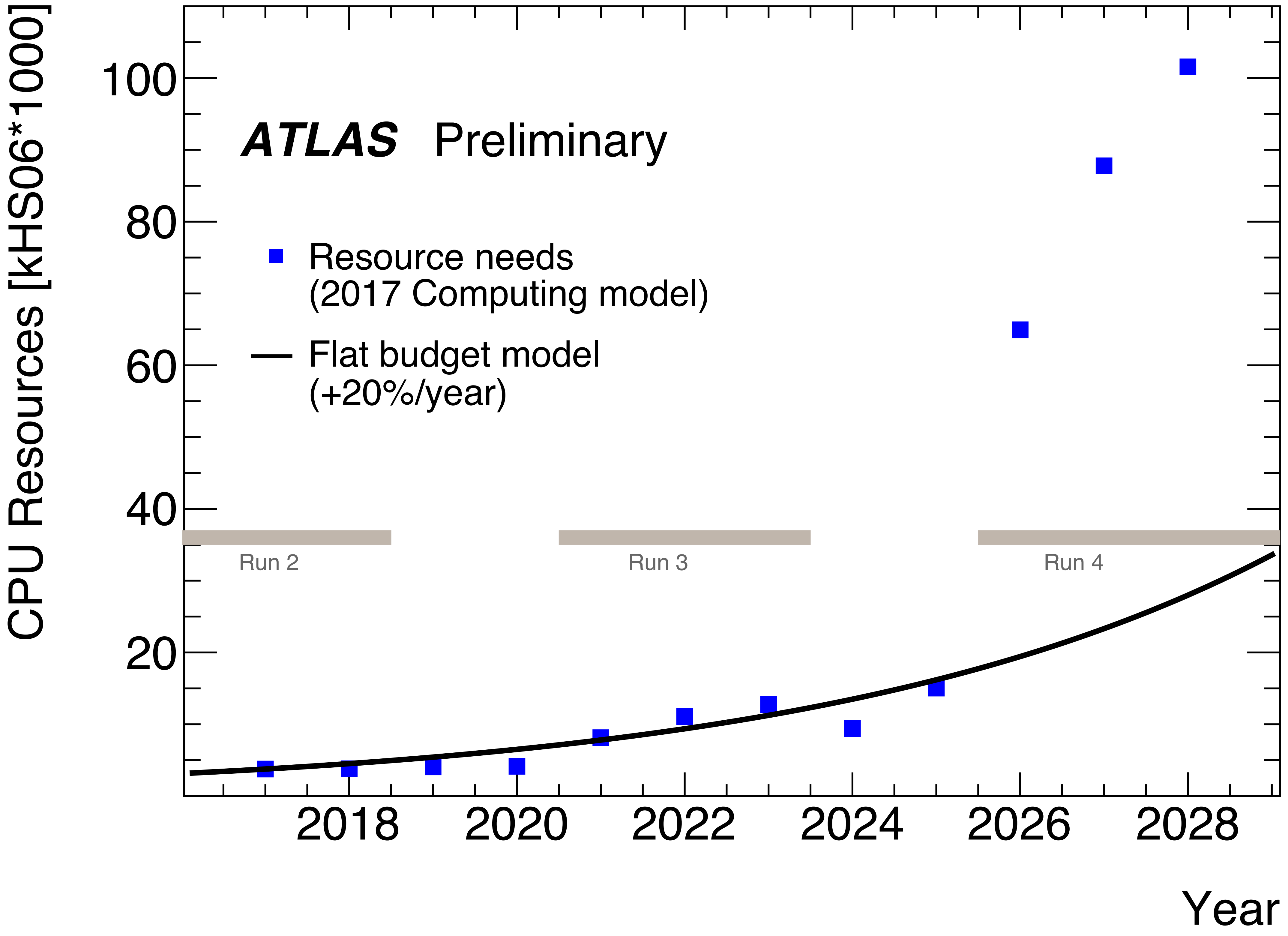
Don’t have enough compute for calorimeter simulations



ATLAS 2016 numbers

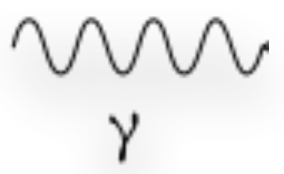
Simulation!

(~75% of that for calorimeter sim alone)

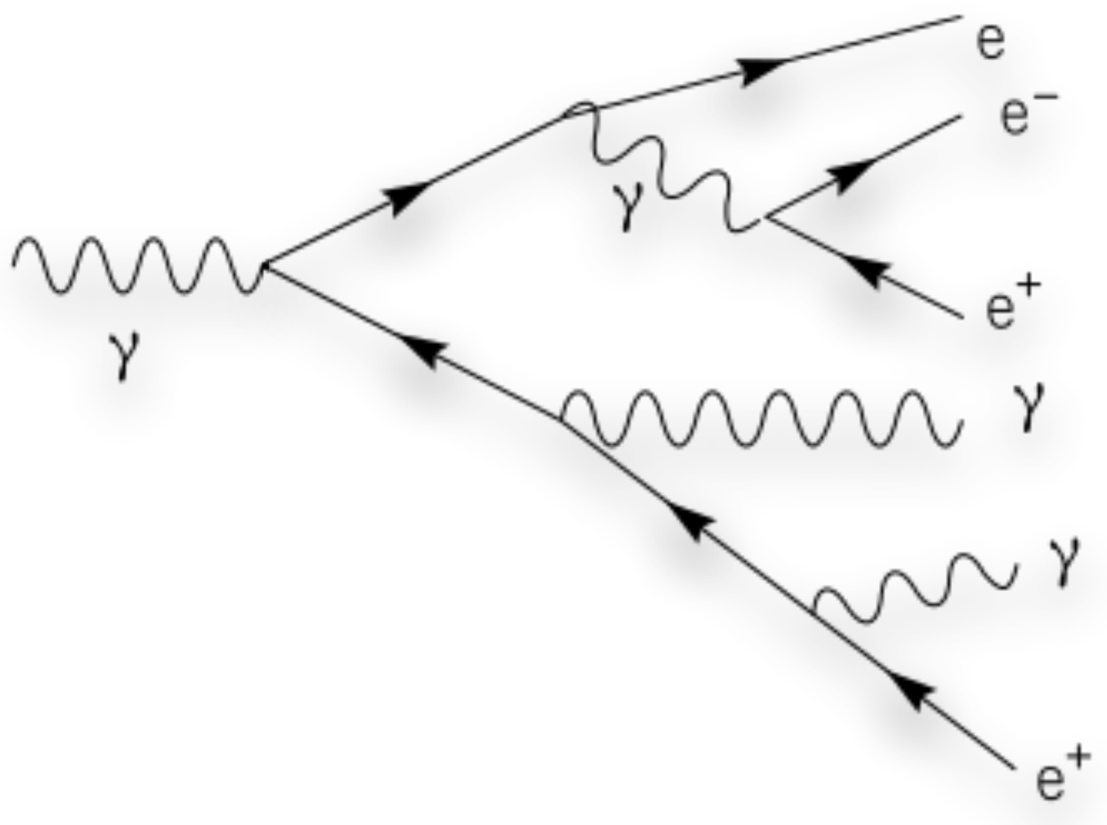


Source (might be old by now)

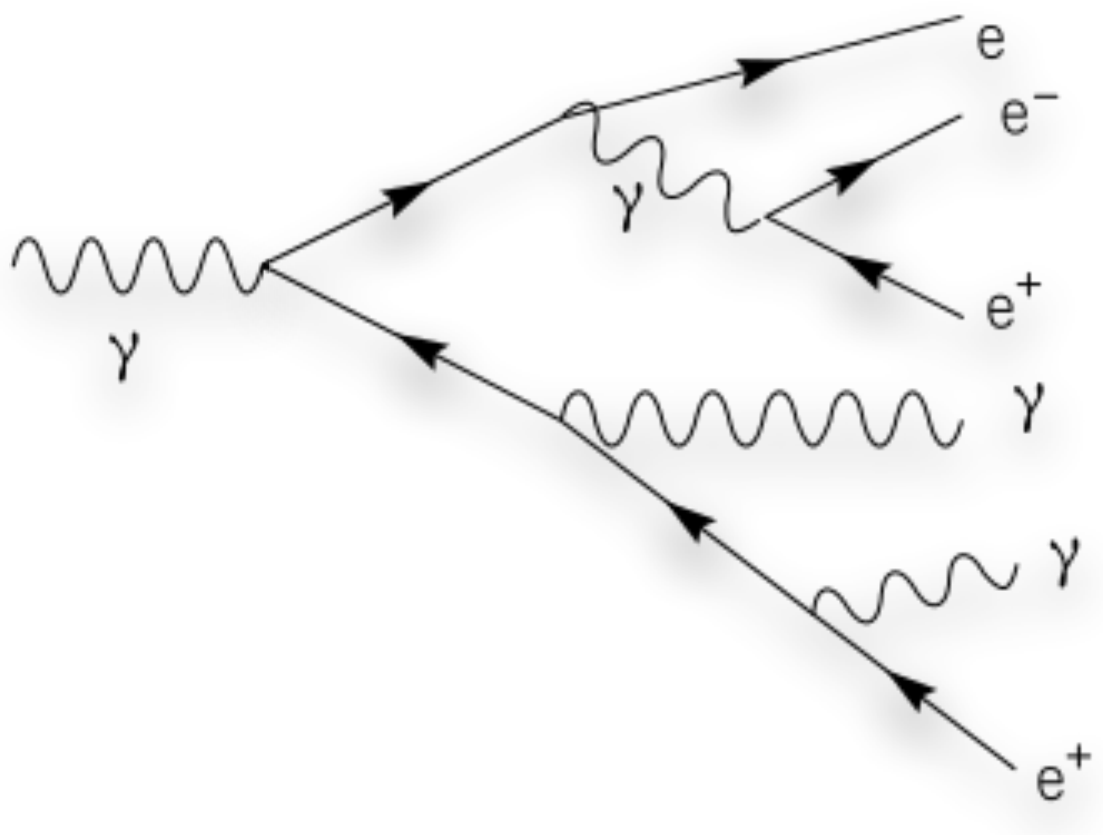
Calorimeter Response

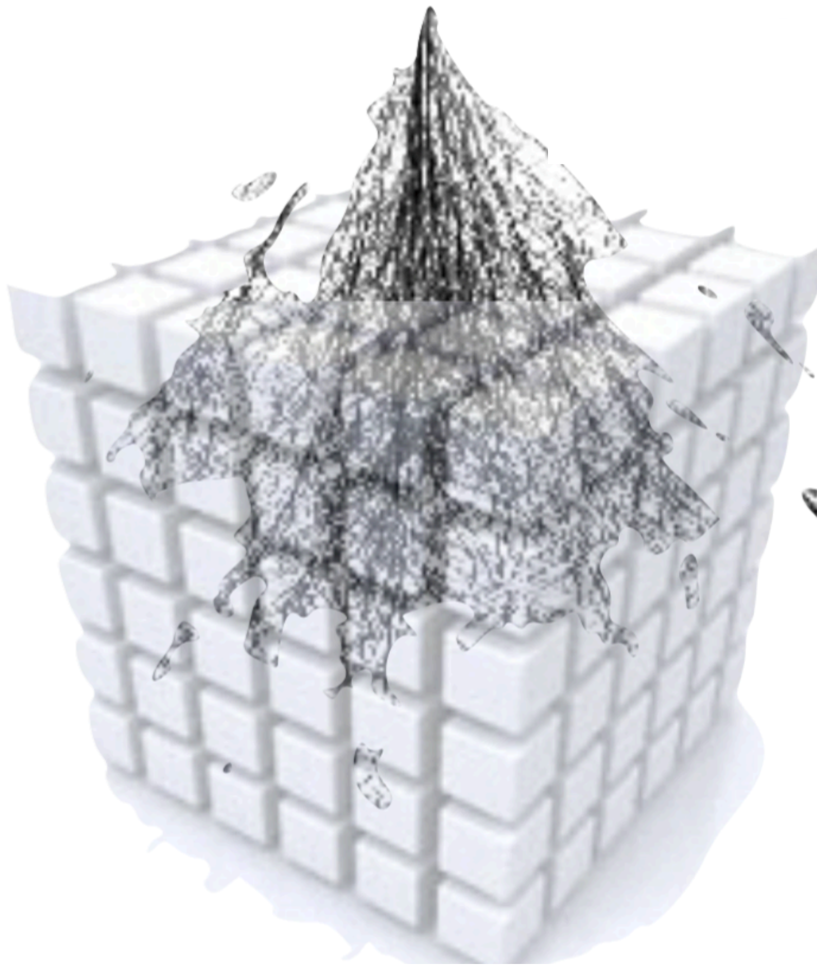


Calorimeter Response

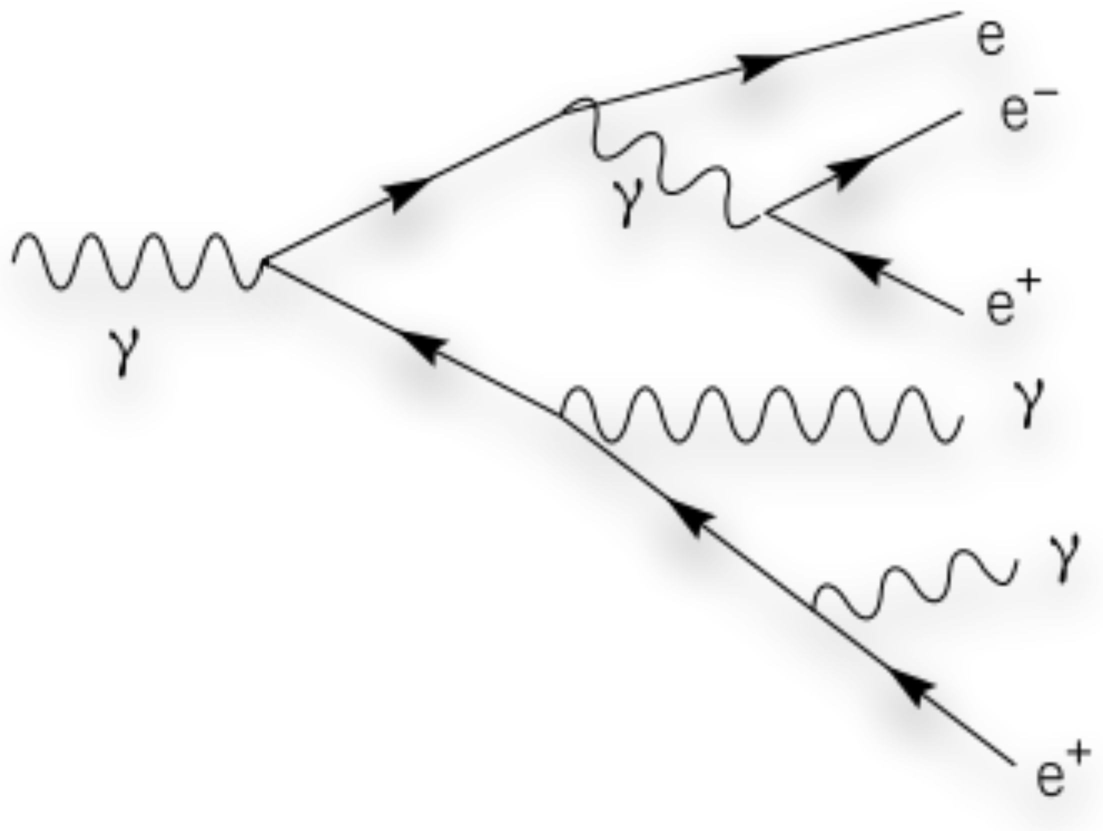


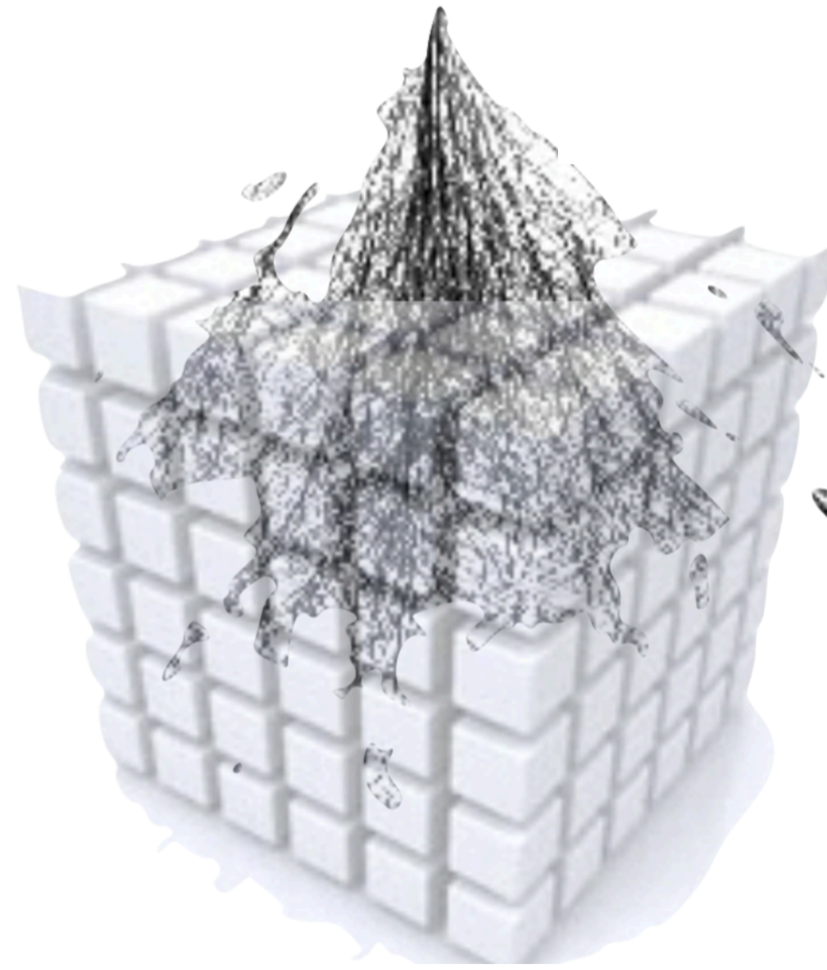
Calorimeter Response

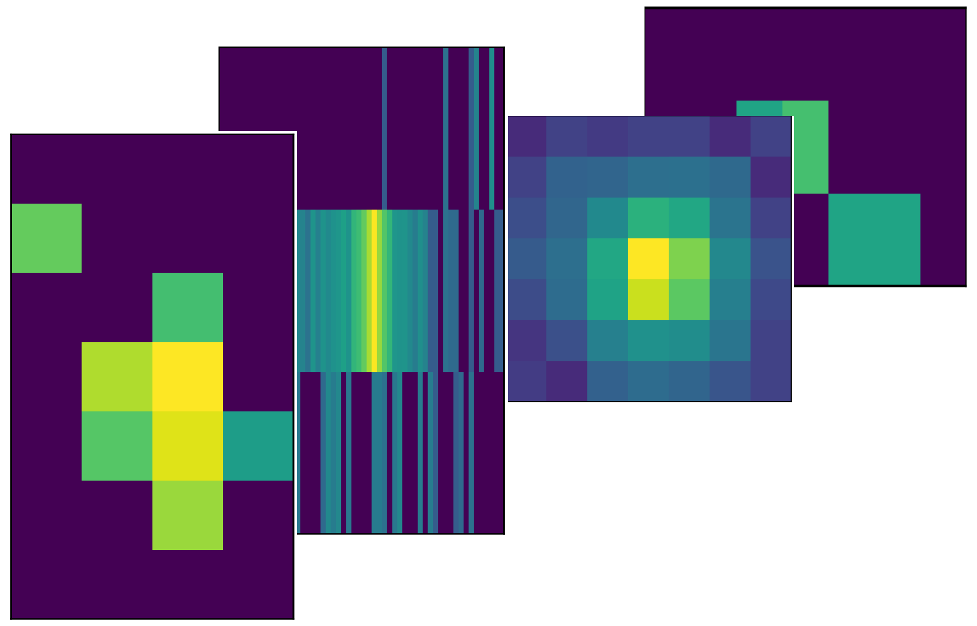




Calorimeter Response



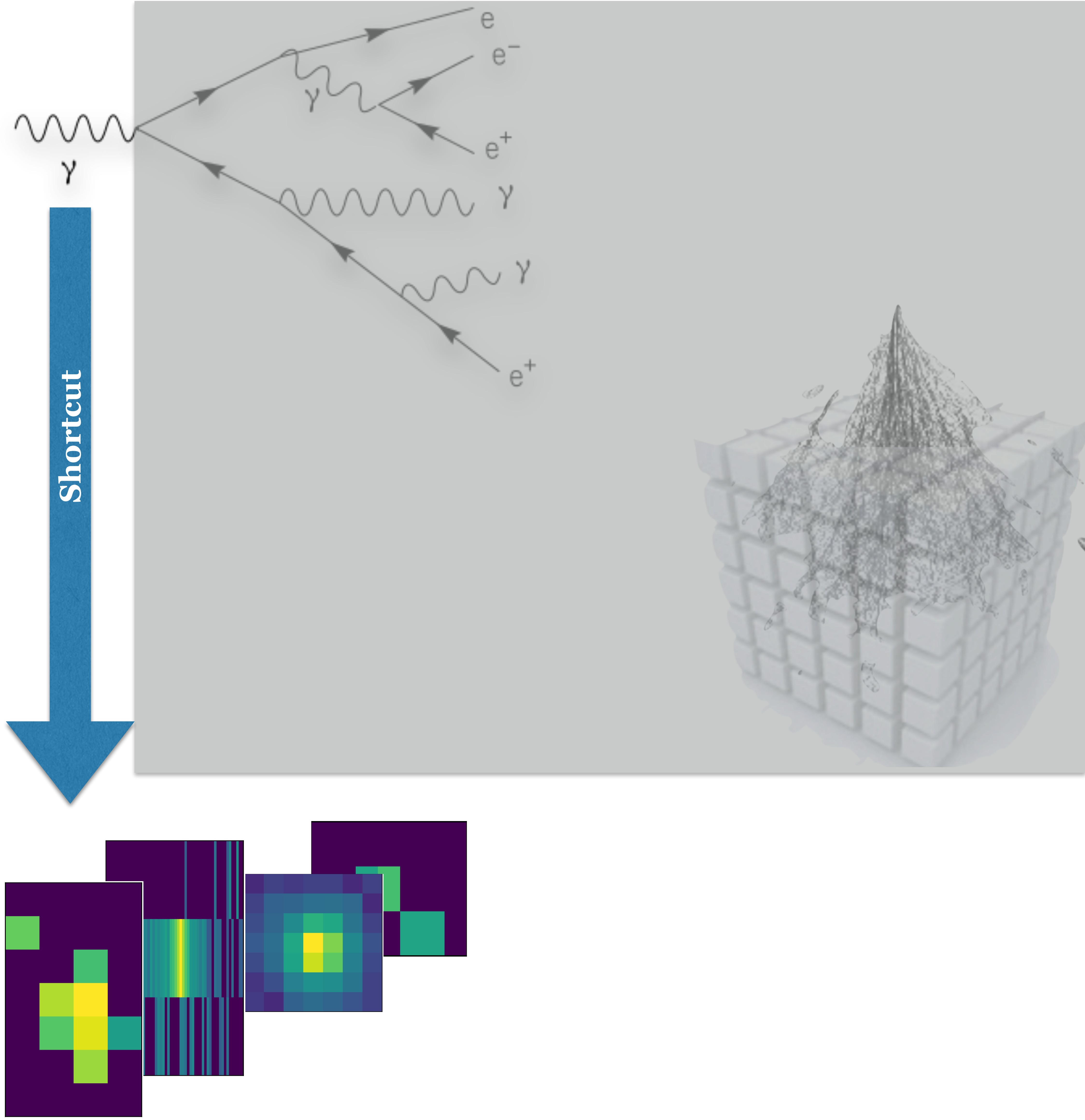




Calorimeter Response

Only final image recorded

Why track full time evolution of all particles ?



ATLAS Experiment Calorimeter Simulation (2019)

First to be integrated into full C++ simulation infrastructure

Details here here

ATLAS Experiment Calorimeter Simulation (2019)

First to be integrated into full C++ simulation infrastructure

- Trained GAN & VAE on **cropped 3D images in central region**
- Validated in ATLAS software using high level physics variables
- **Interpolation test:** Train on 9 log-spaced energy points, test on 10th
- **Speed:** Orders of magnitude faster than Geant4, no longer bottleneck in sim time
- **Memory:** Tiny memory footprint O(MBs) compared to O(GBs) for baseline

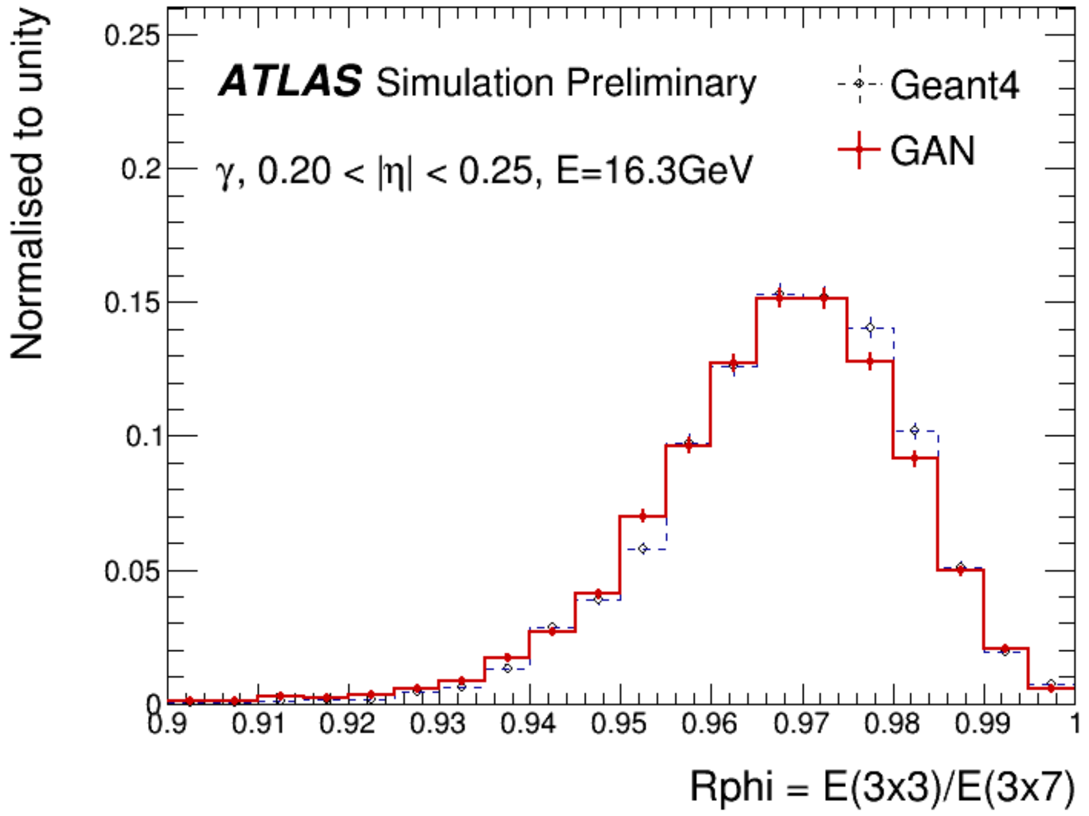
Details here here

ATLAS Experiment Calorimeter Simulation (2019)

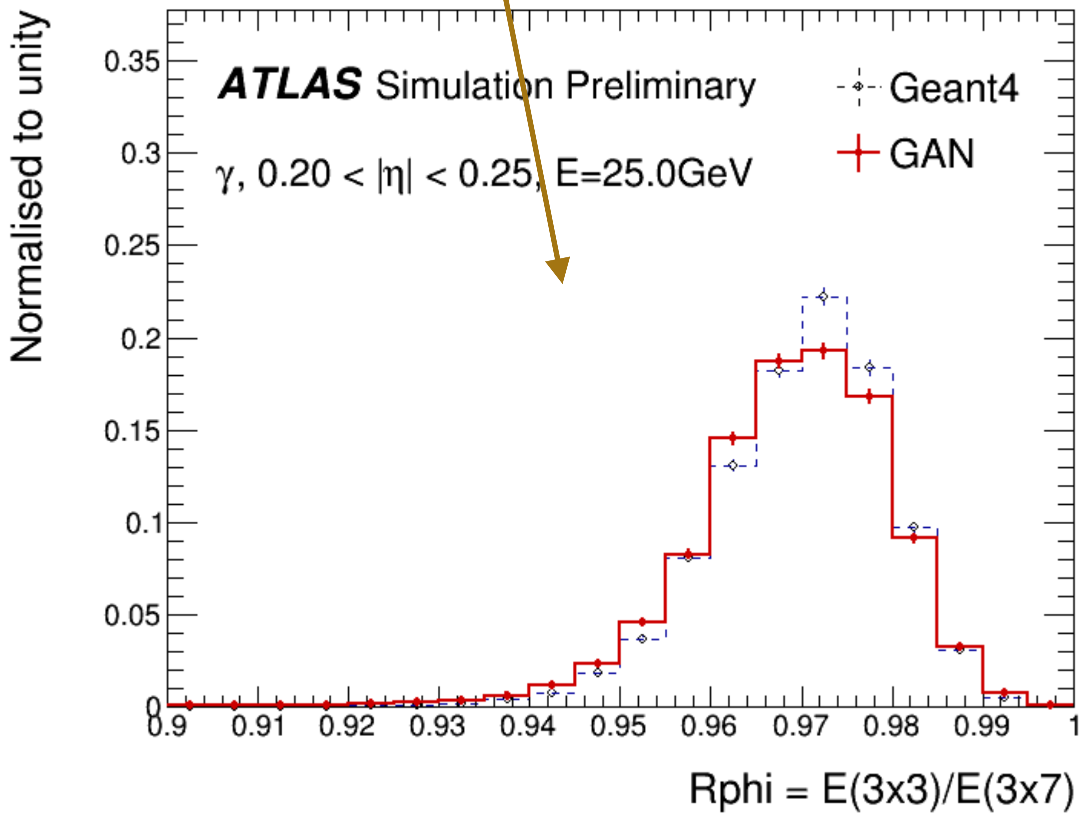
First to be integrated into full C++ simulation infrastructure

- Trained GAN & VAE on **cropped 3D images in central region**
- Validated in ATLAS software using high level physics variables
- **Interpolation test:** Train on 9 log-spaced energy points, test on 10th
- **Speed:** Orders of magnitude faster than Geant4, no longer bottleneck in sim time
- **Memory:** Tiny memory footprint O(MBs) compared to O(GBs) for baseline

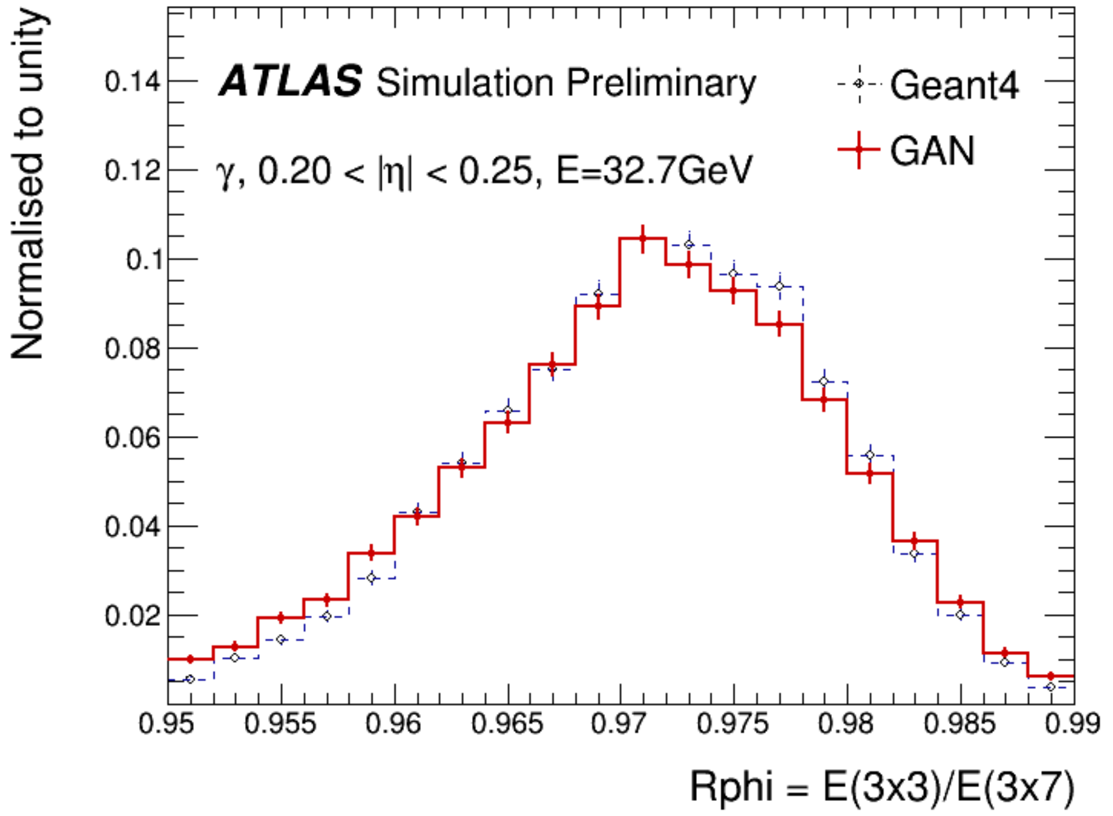
Interpolation test



16 GeV



25 GeV



32 GeV

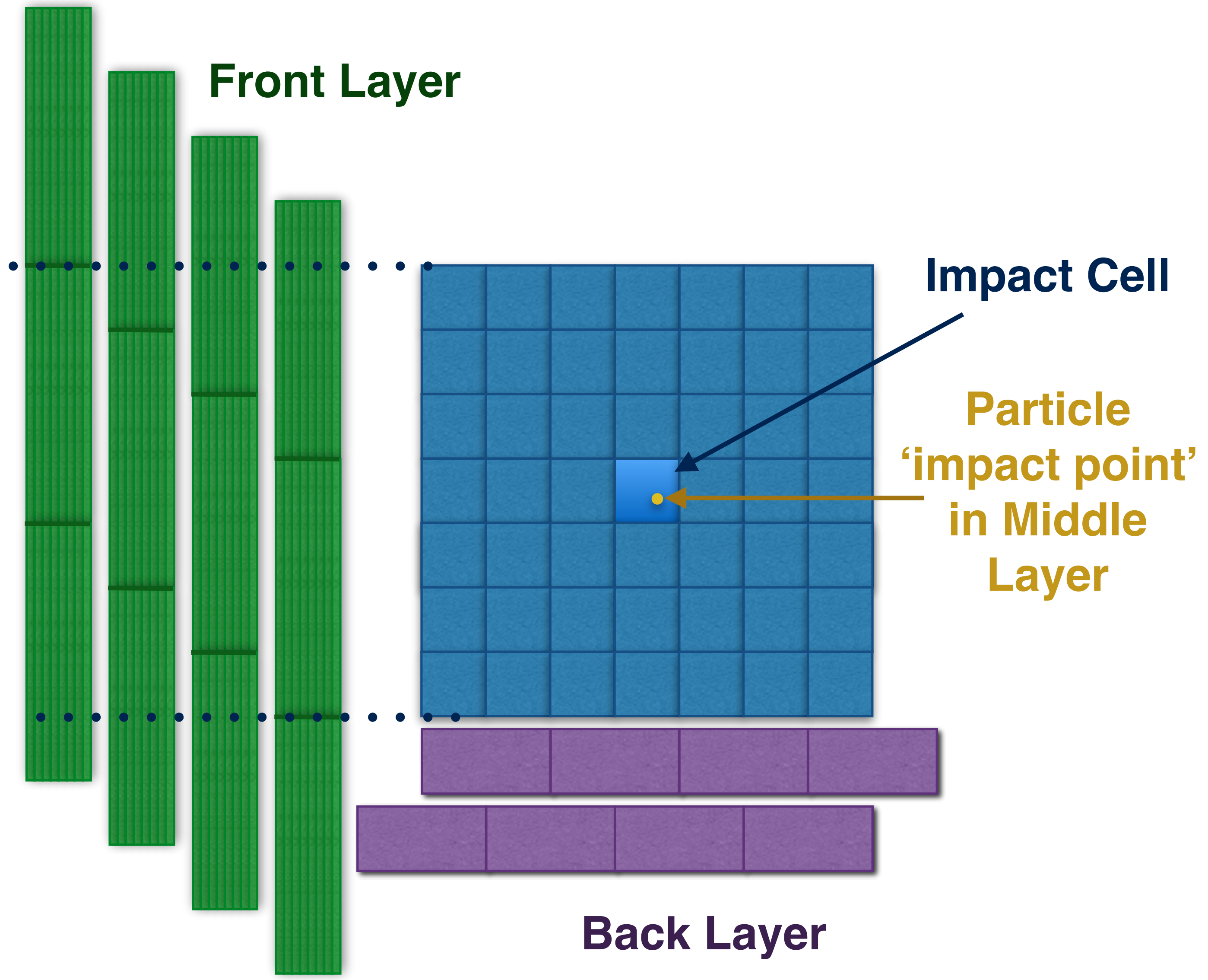
Details here here

Challenge: Condition on detector geometry

Alignment of cell edges in 3D changes from image to image

Cell sizes change from region to region in the detector

Conditioning on $4 \times 2 = 8$ configurations already challenging, cannot scale to full detector

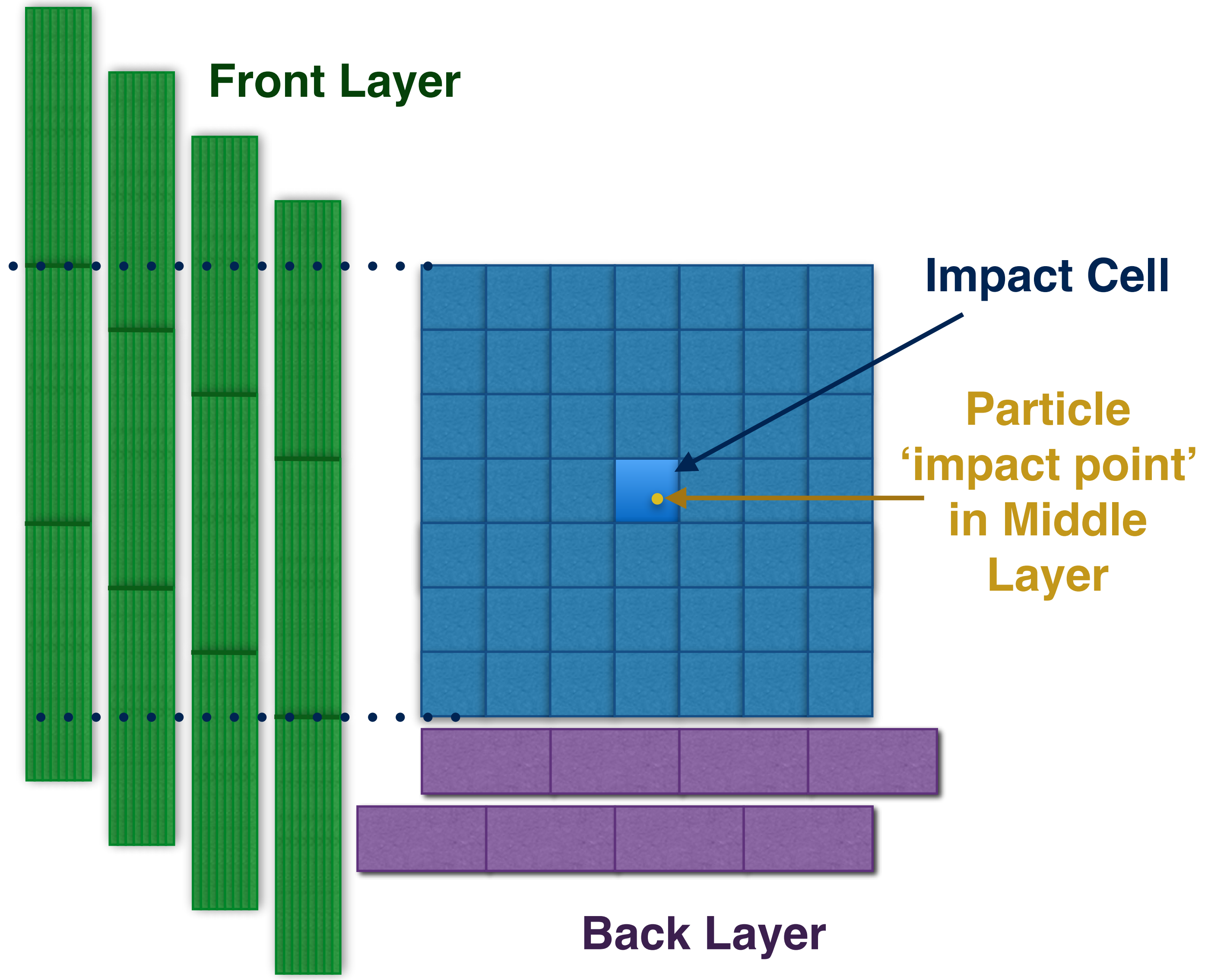


Challenge: Condition on detector geometry

Alignment of cell edges in 3D changes from image to image

Cell sizes change from region to region in the detector

Conditioning on $4 \times 2 = 8$ configurations already challenging, cannot scale to full detector



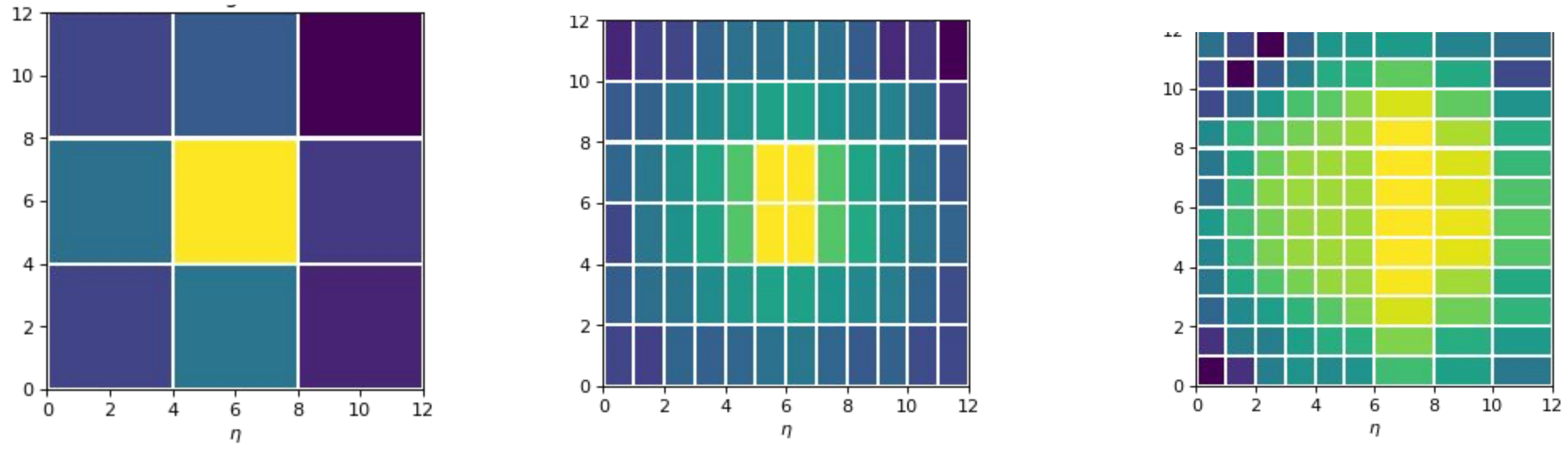
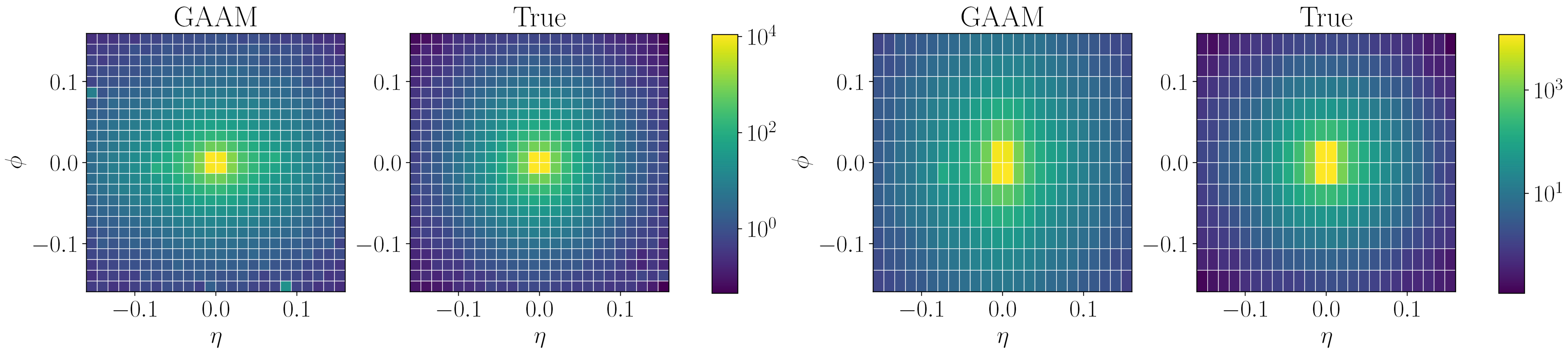


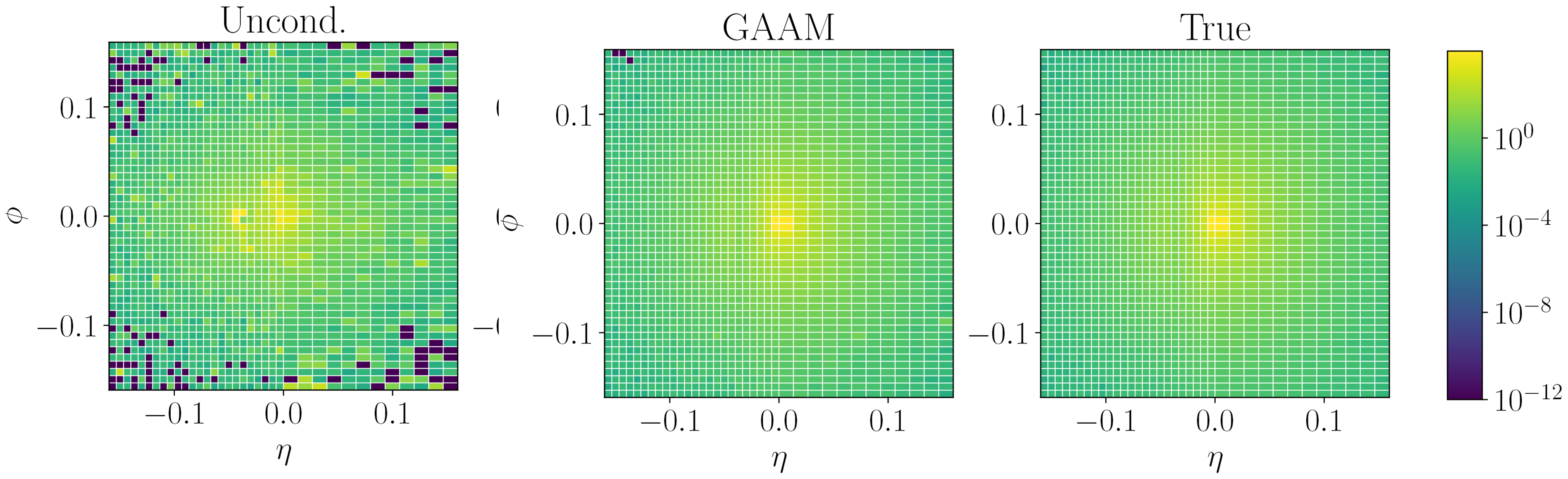
Image: Source

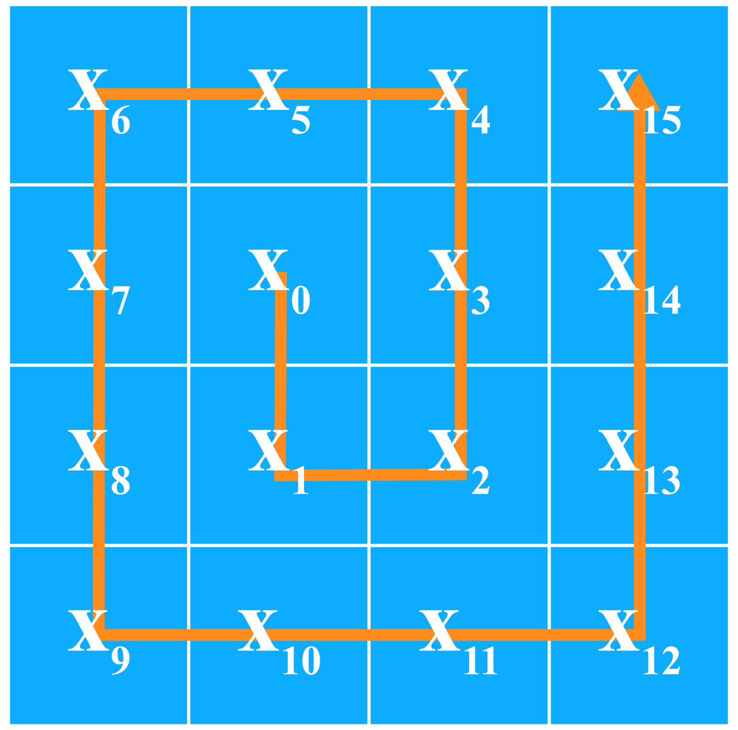
Geometry-aware autoregressive networks



(a) Middle layer - (24, 24)

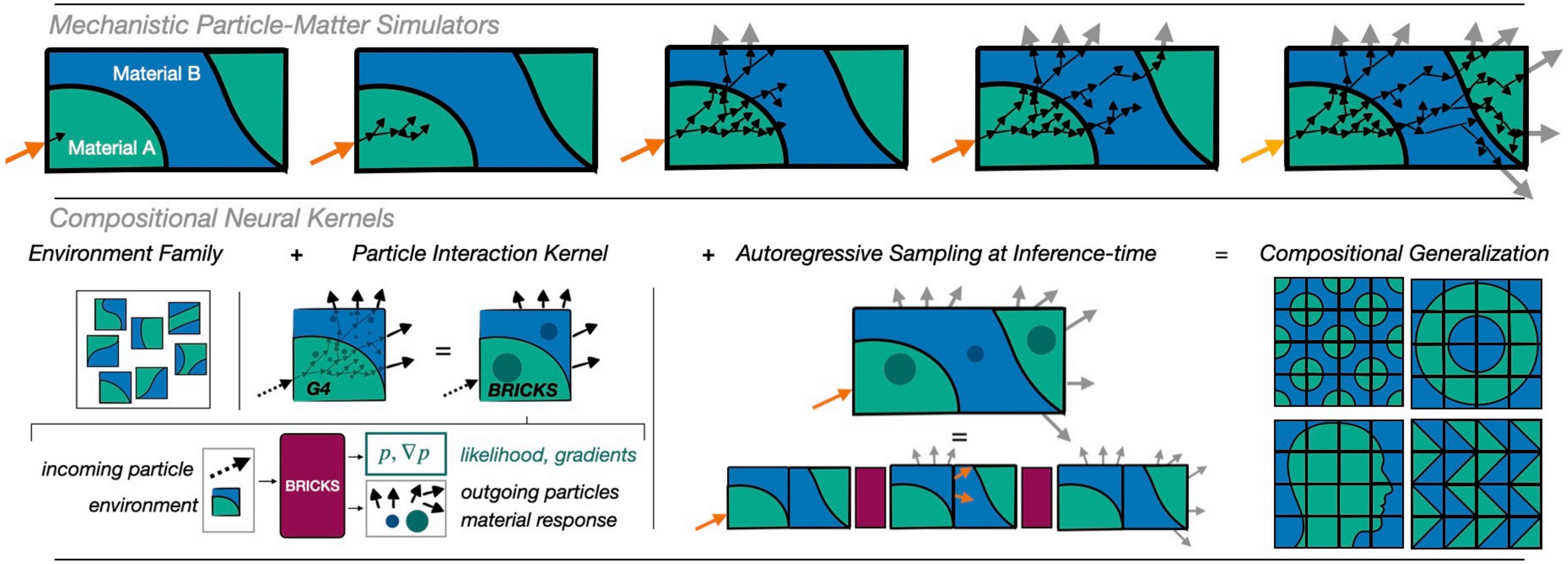
(b) Middle layer - (24, 12)





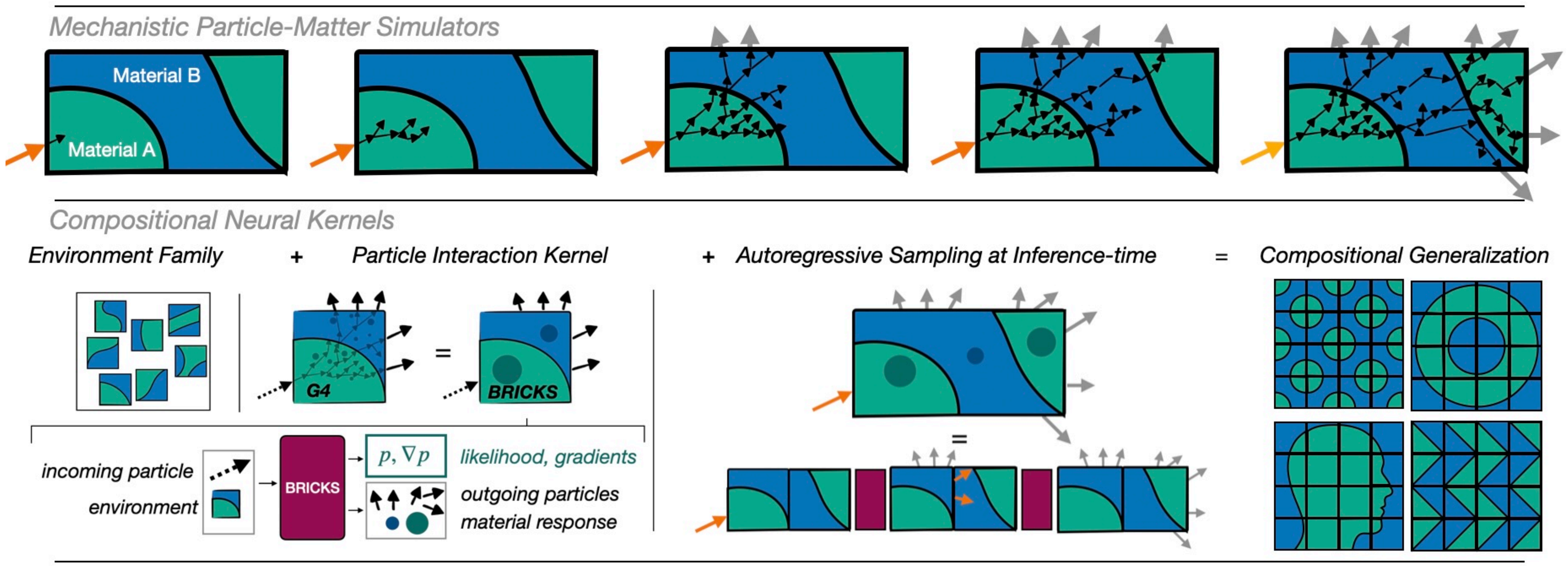
Liu et al (incl Ghosh)

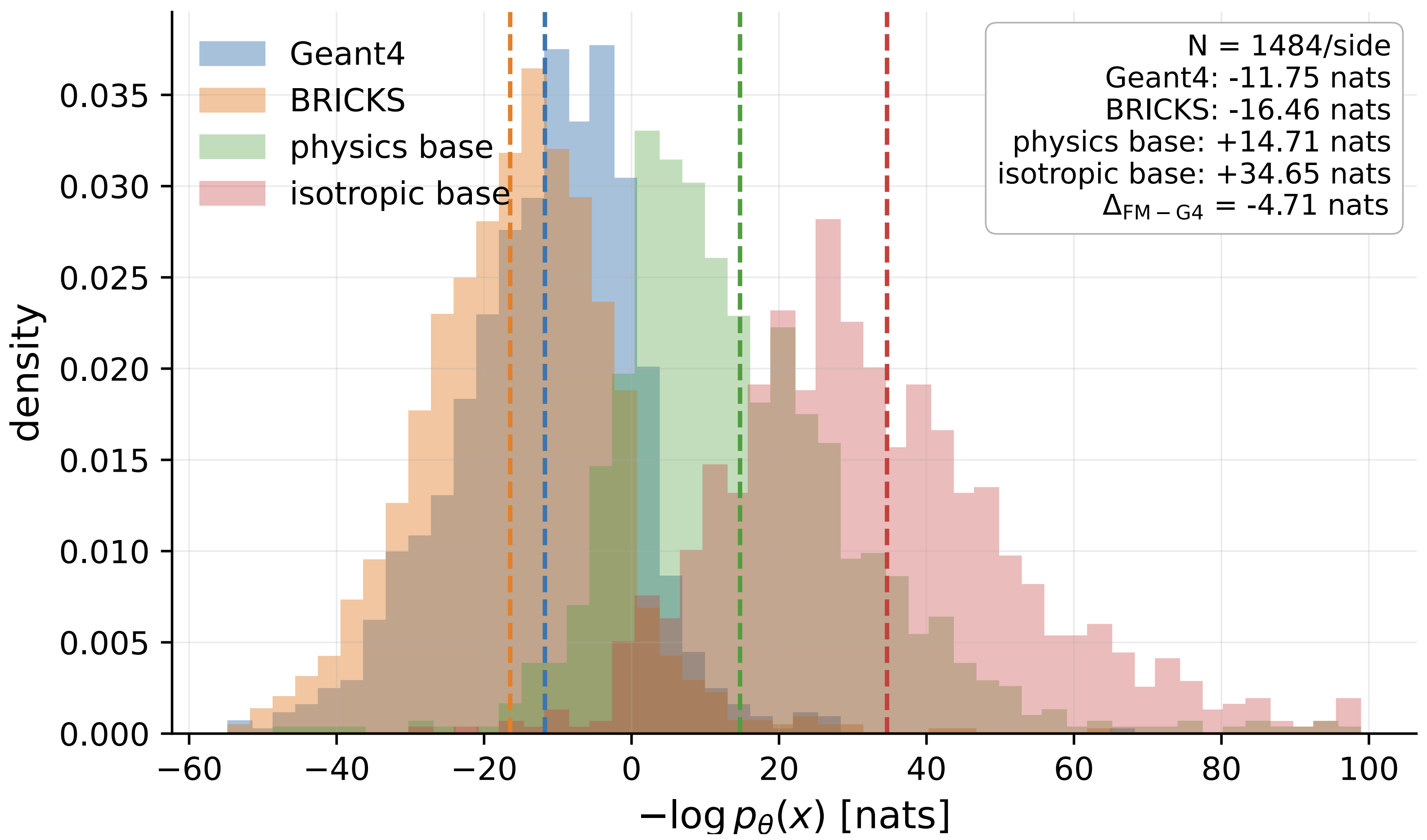
Bring back time-evolution through autoregressive networks



Hildebrandt et al

Bring back time-evolution through autoregressive networks

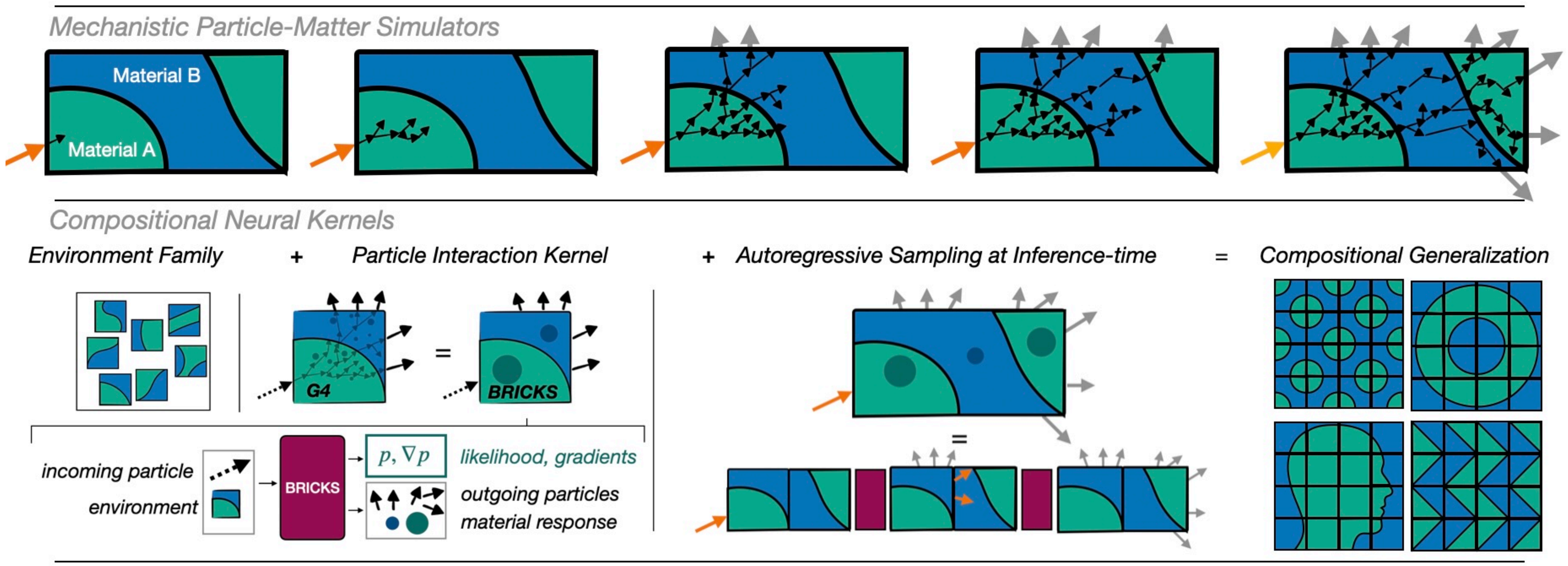


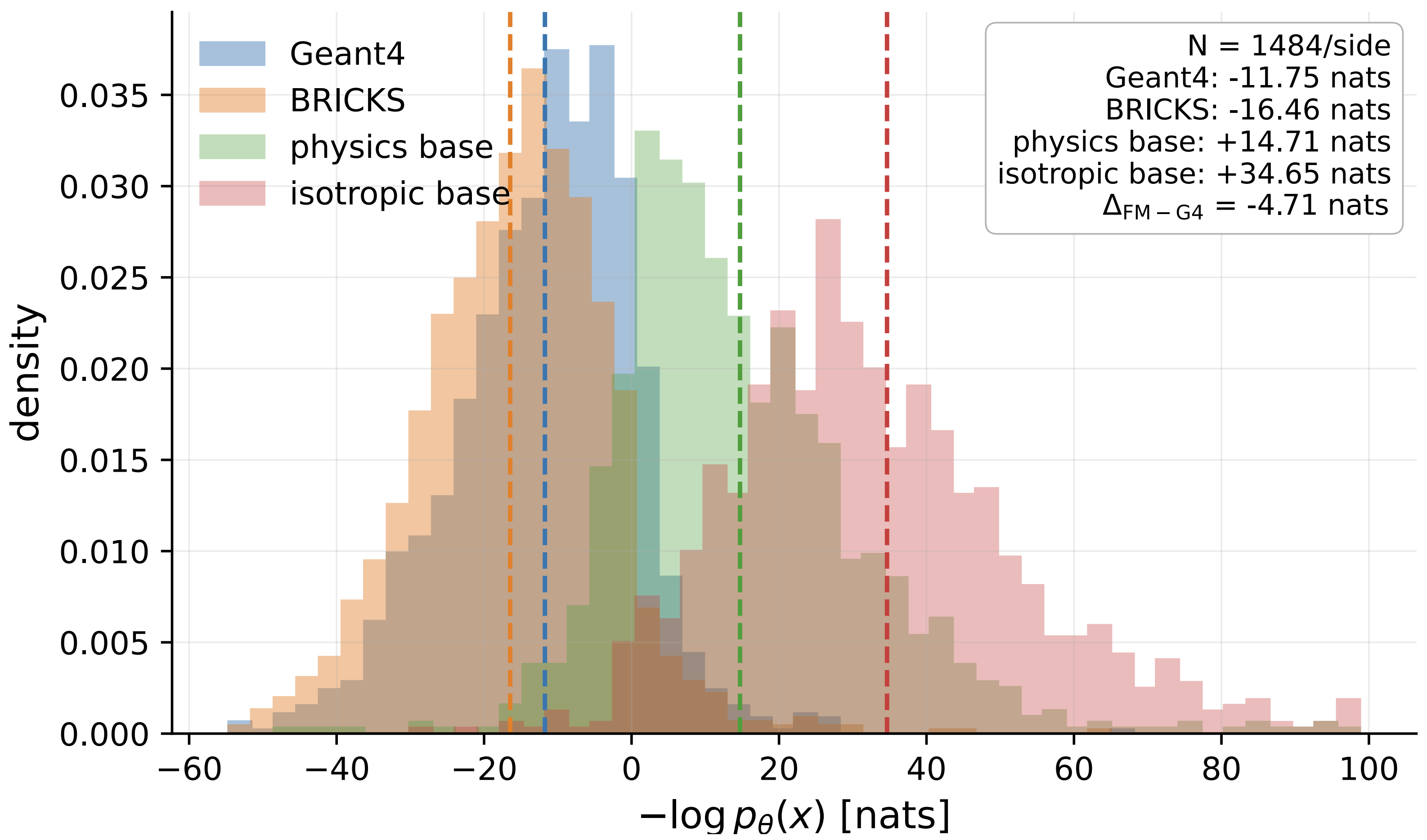


Hildebrandt et al

Could become a Geant4-like foundation model

Bring back time-evolution through autoregressive networks

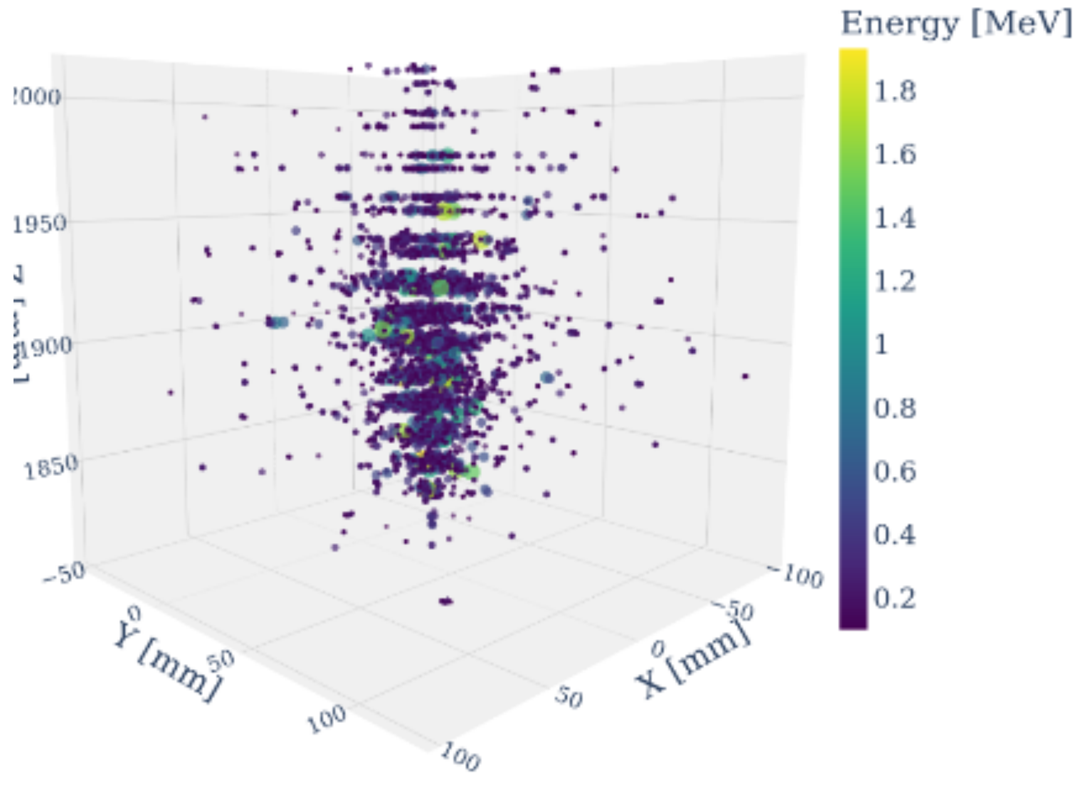


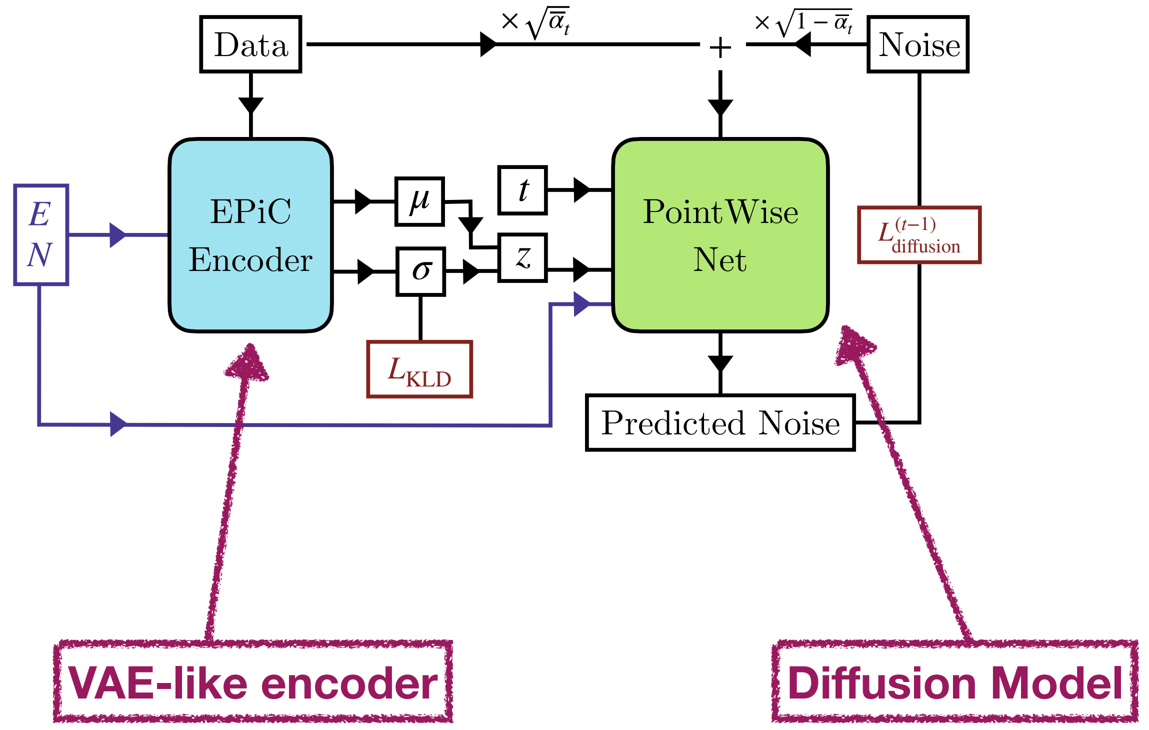


Hildebrandt et al

Could become a Geant4-like foundation model

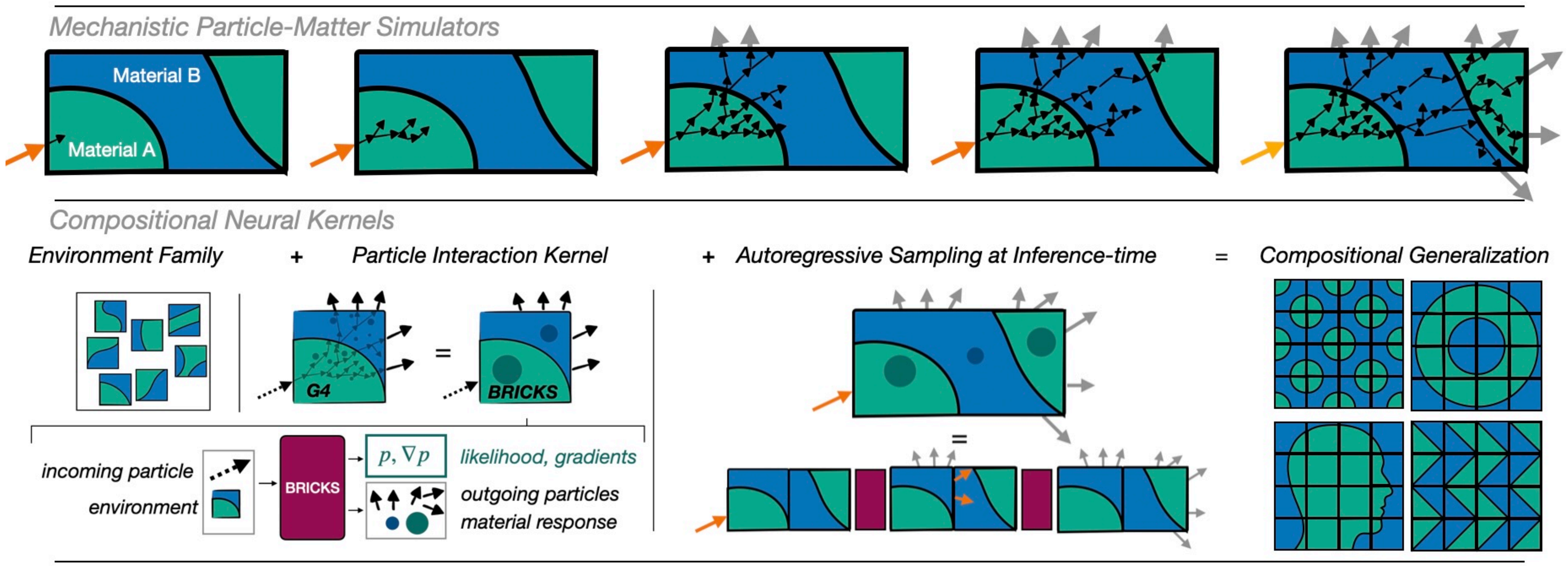
Or just generate point cloud level?

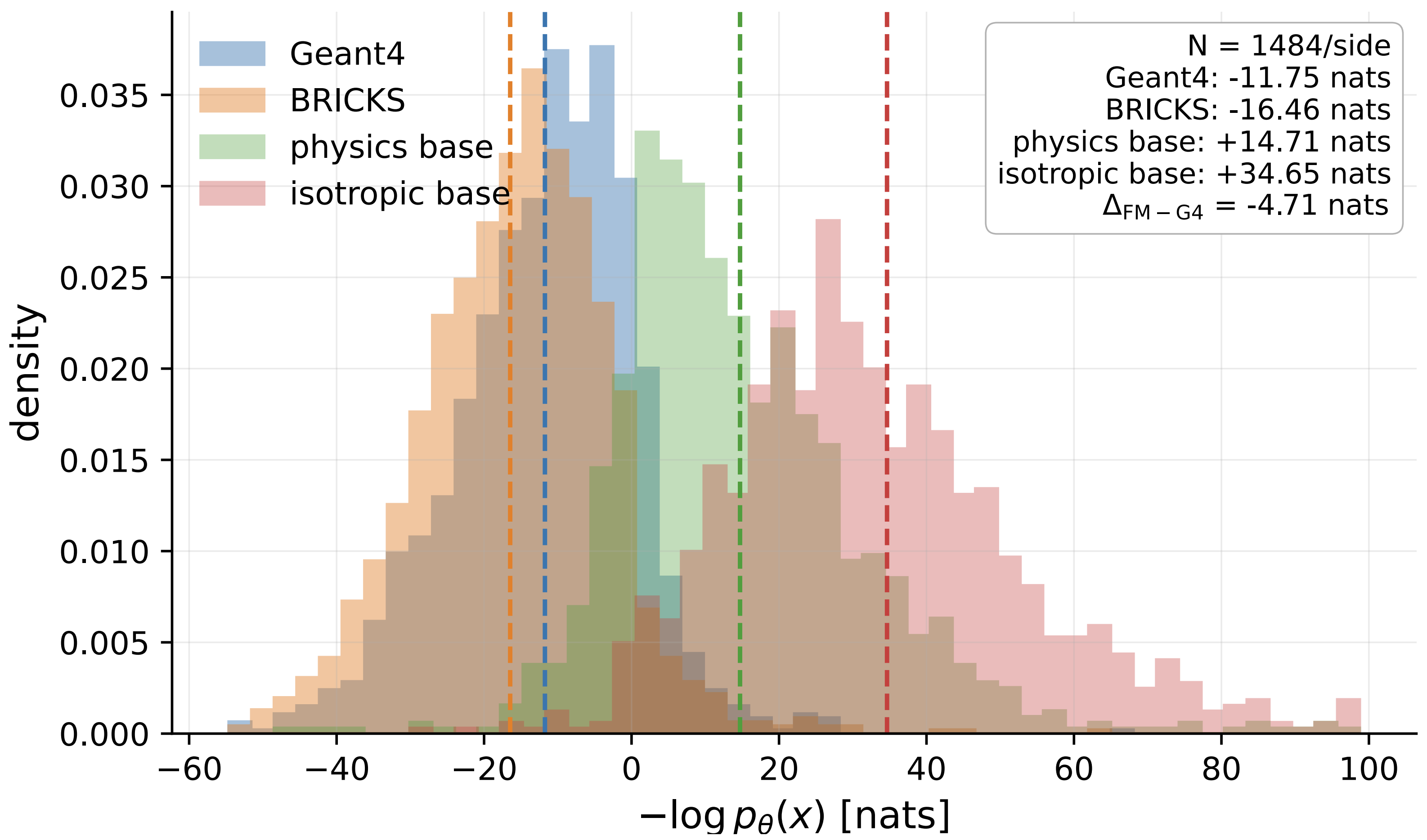




Buhmann et al

Bring back time-evolution through autoregressive networks



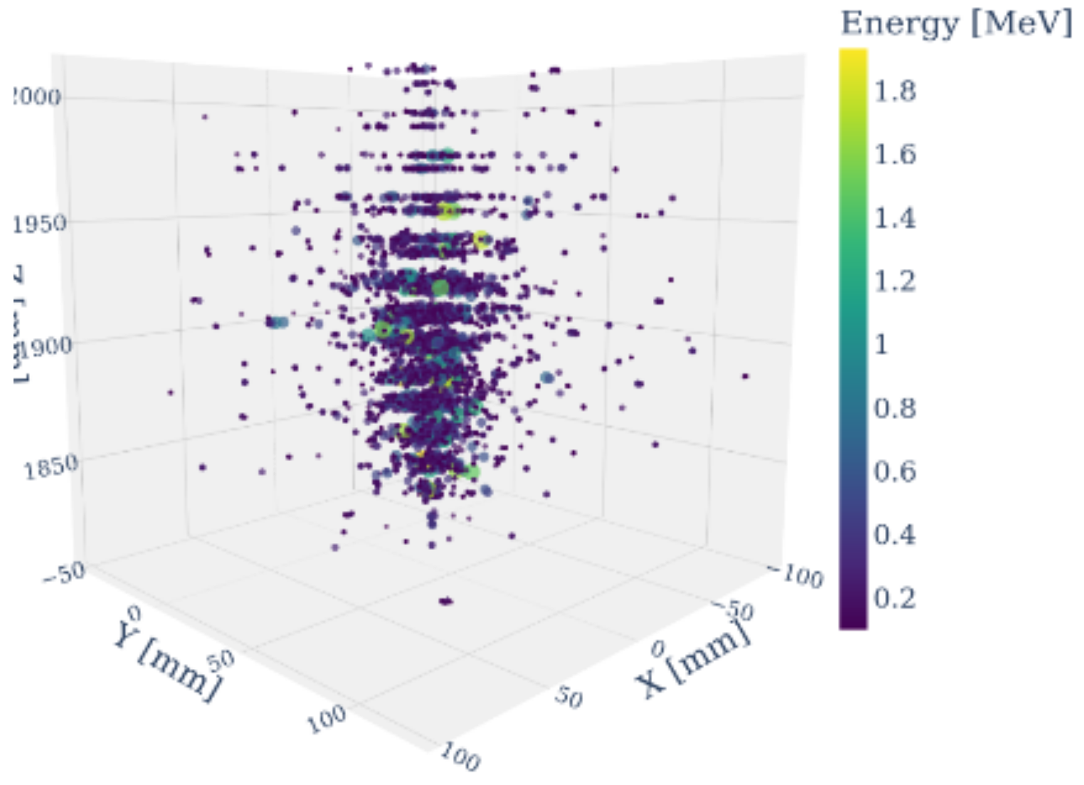


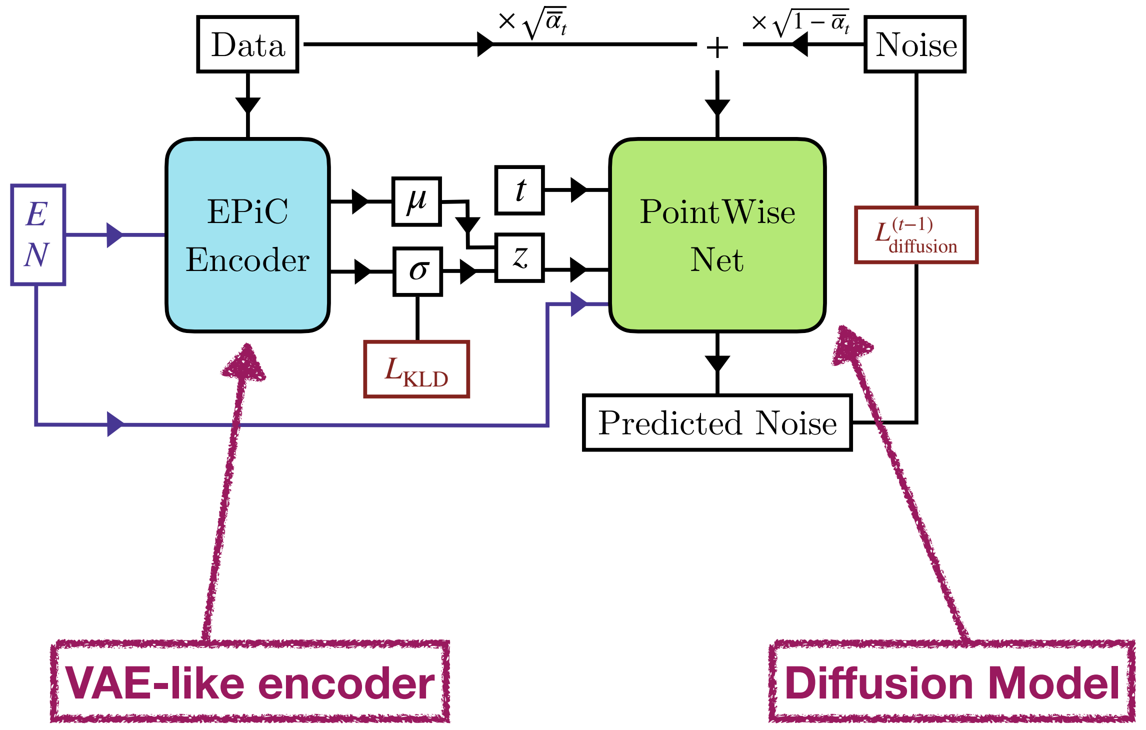
Hildebrandt et al

Could become a Geant4-like foundation model

Open question which approach will succeed longterm

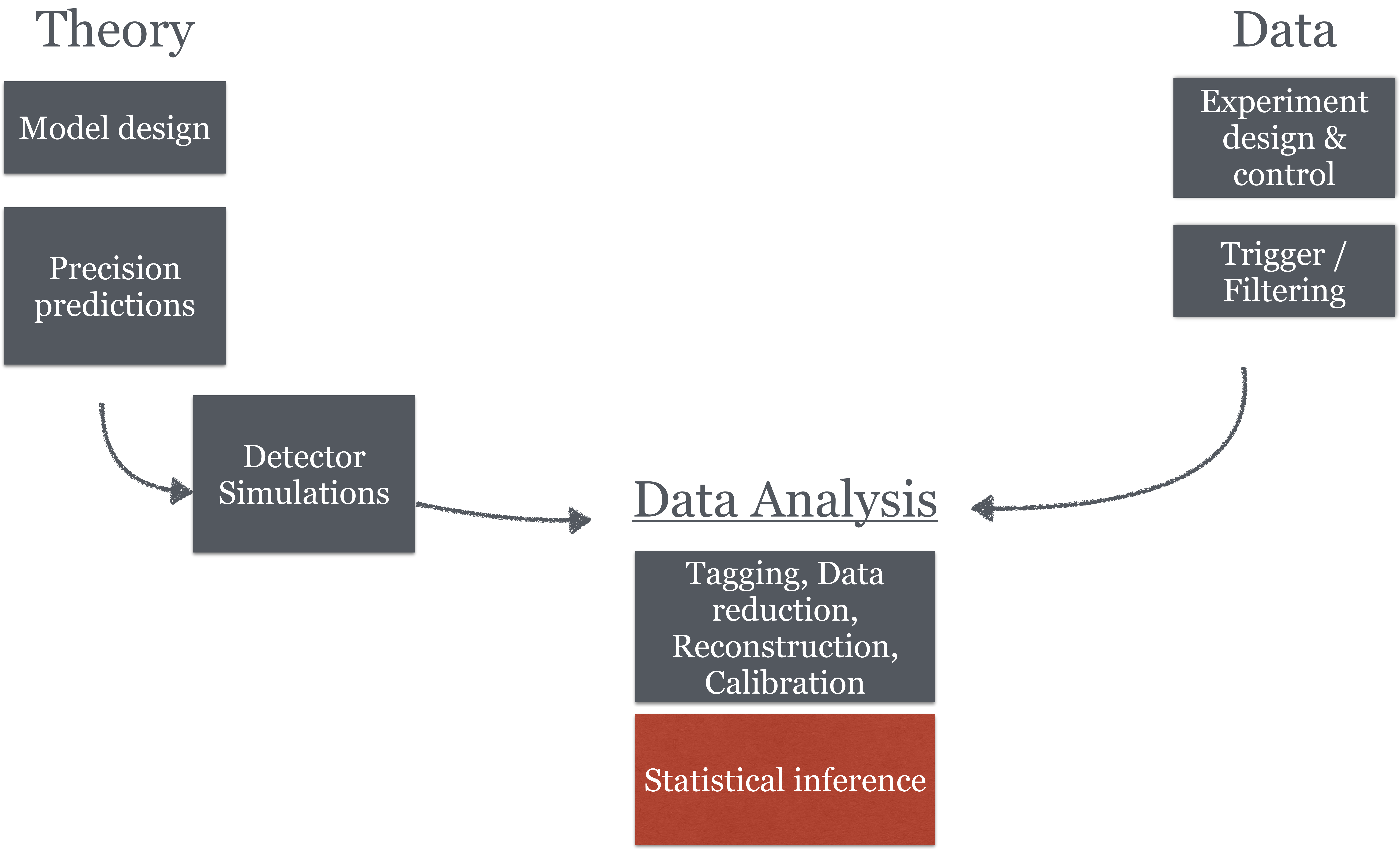
Or just generate point cloud level?





Buhmann et al

ML across the scientific pipeline



Theory

Model design

Precision predictions

Detector Simulations

Data

Experiment design & control

Trigger / Filtering

Data Analysis

Tagging, Data reduction, Reconstruction, Calibration

Statistical inference

Fully **implicit** assumptions
about the signal

NN architecture decides what
new physics looks like

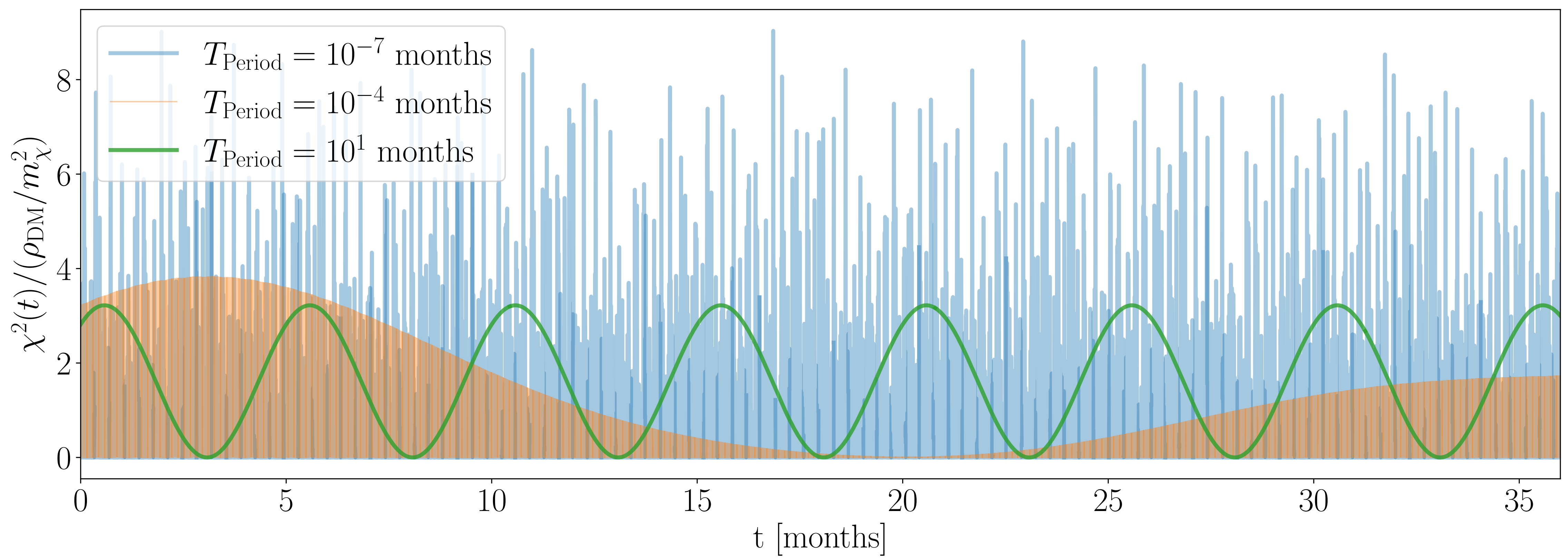
Fully **explicit** assumptions
about the signal

No ability to find
unexpected new physics

Eg. Time-dependent signals of new physics at the LHC:

You know time is important but you don’t know what the signal shape looks like.

Which approach do you choose?



Fully **implicit** assumptions about the signal

Fully **explicit** assumptions about the signal

NN architecture decides what new physics looks like

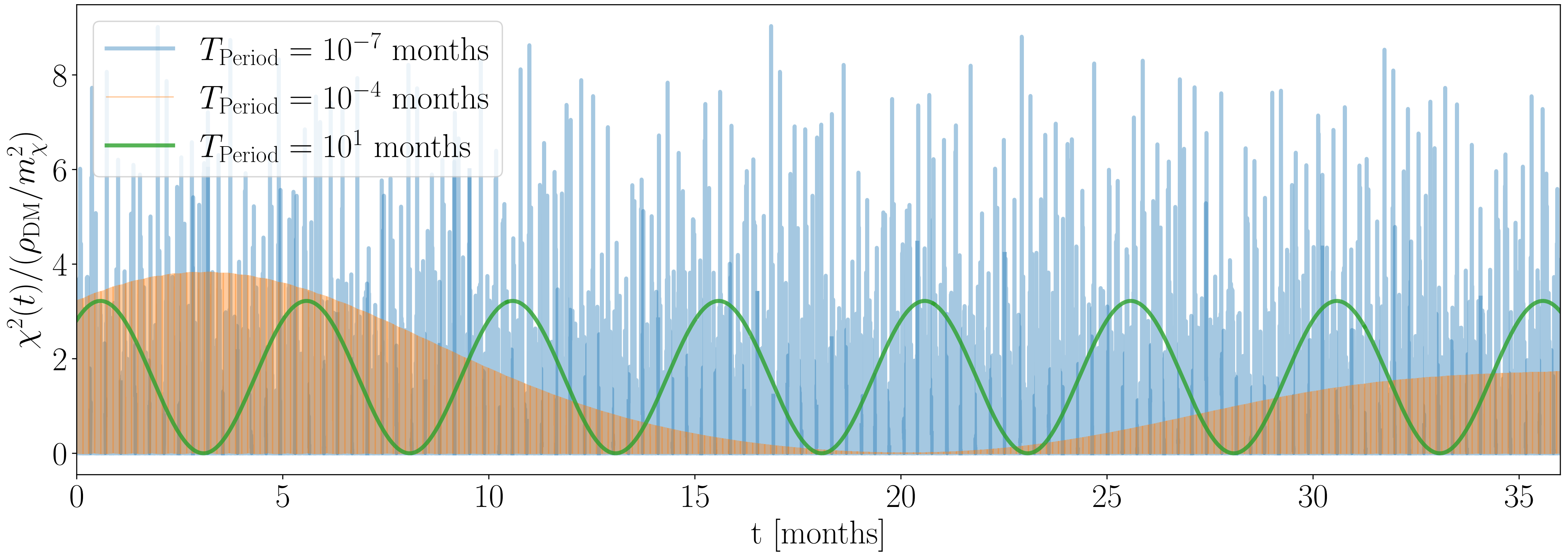
No ability to find unexpected new physics

Fieg et al (incl Ghosh)

Eg. Time-dependent signals of new physics at the LHC:

You know time is important but you don’t know what the signal shape looks like.

Which approach do you choose?



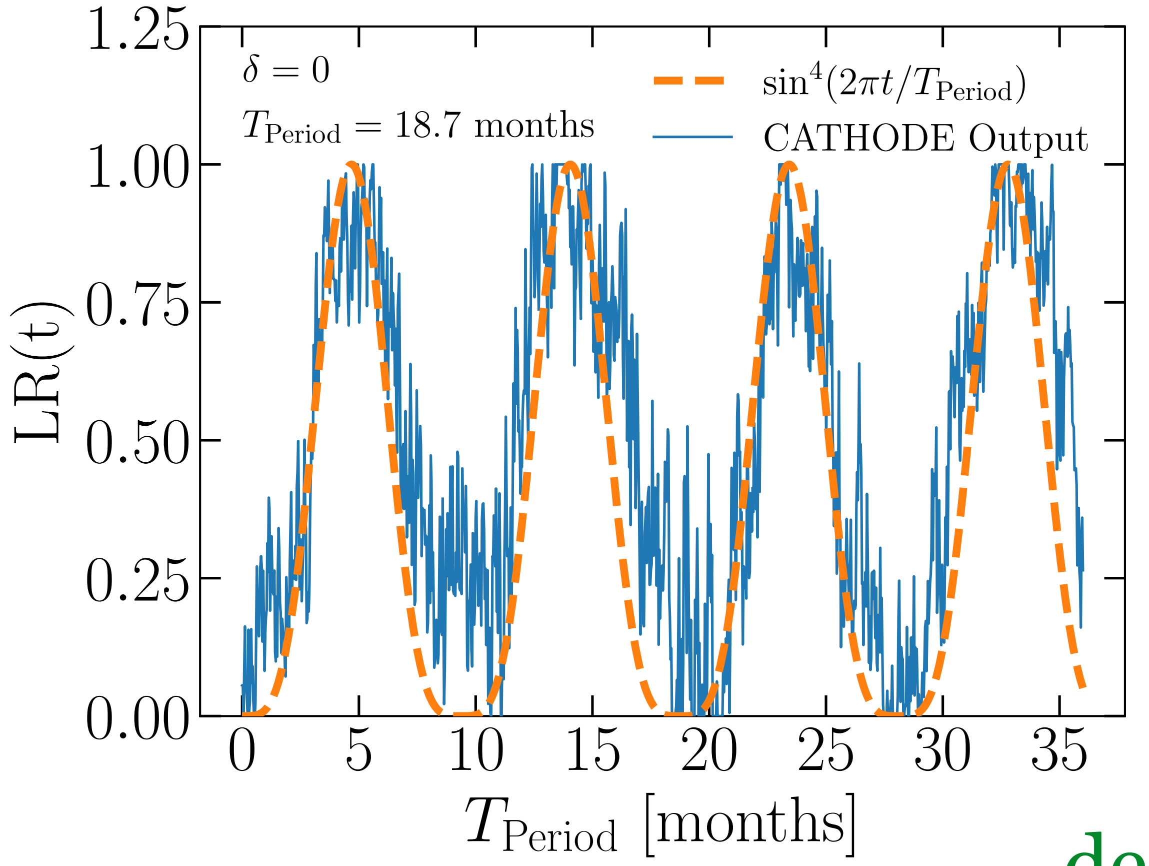
Fully **implicit** assumptions about the signal

Sweet spot?

Fully **explicit** assumptions about the signal

NN architecture decides what new physics looks like

Anomaly detection with physics priors



Knows about time dependence but finds the dependence from data

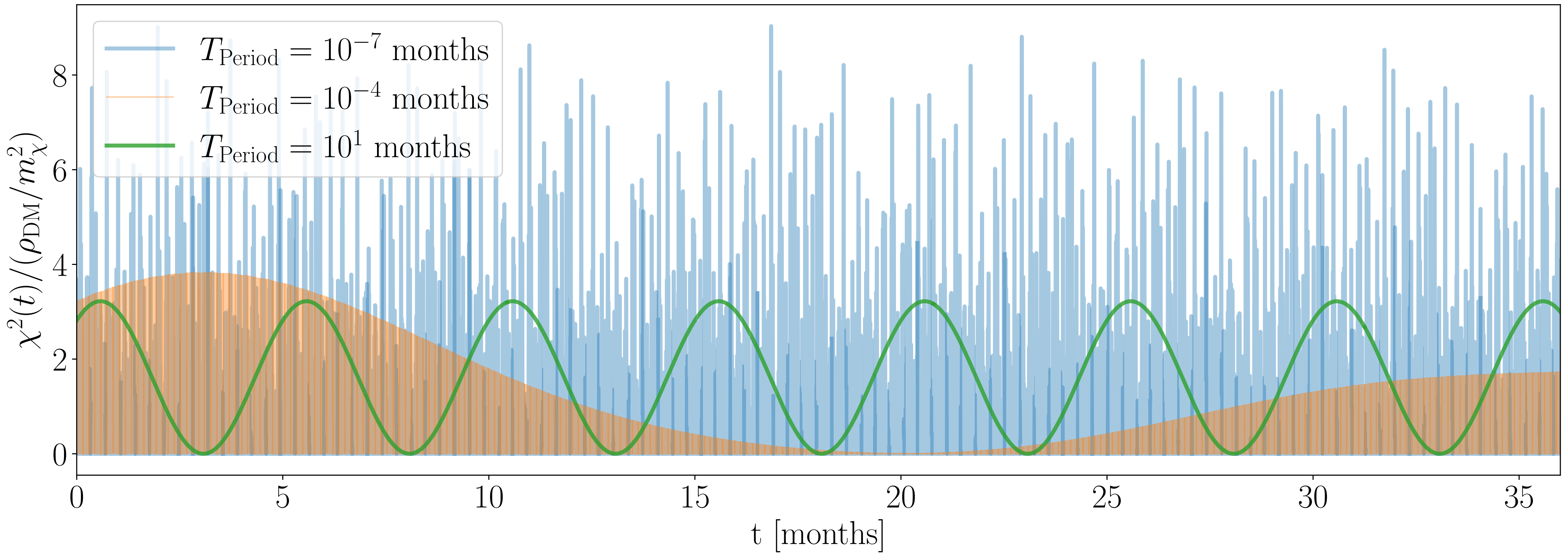
No ability to find unexpected new physics

Fieg et al (incl Ghosh)

Eg. Time-dependent signals of new physics at the LHC:

You know time is important but you don't know what the signal shape looks like.

Which approach do you choose?



Fully **implicit** assumptions about the signal

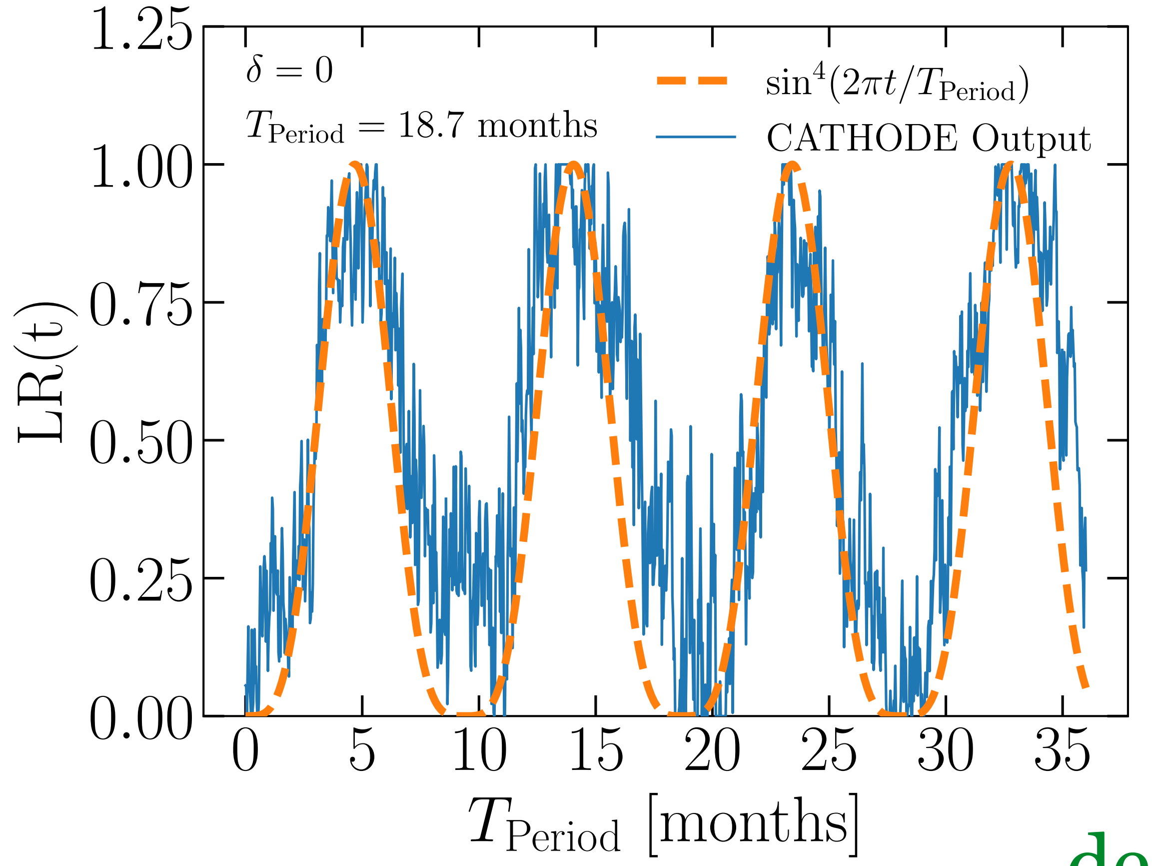
Sweet spot?

Fully **explicit** assumptions about the signal

NN architecture decides what new physics looks like

Anomaly detection with physics priors

No ability to find unexpected new physics



Learn more about AD in Melissa's talk Friday and talk to Dennis!

Knows about time dependence but finds the dependence from data

Fieg et al (incl Ghosh)

When testing well described hypotheses …

$$p(\text{theory}\,|\,\text{data}) = \frac{p(\text{data}\,|\,\text{theory})p(\text{theory})}{p(\text{data})}$$

When testing well described hypotheses …

What we all
want
(Posterior)

$$p(\text{theory} \mid \text{data}) = \frac{p(\text{data} \mid \text{theory})p(\text{theory})}{p(\text{data})}$$

When testing well described hypotheses …

What we all
want
(Posterior)

Likelihood

$$p(\text{theory}\,|\,\text{data}) = \frac{p(\text{data}\,|\,\text{theory})p(\text{theory})}{p(\text{data})}$$

When testing well described hypotheses …

What we all
want
(Posterior)

Likelihood

Prior

$$p(\text{theory} \mid \text{data}) = \frac{p(\text{data} \mid \text{theory})\,p(\text{theory})}{p(\text{data})}$$

Evidence

When testing well described hypotheses …

What we all
want
(Posterior)

Likelihood

Prior

$$p(\text{theory} \,|\, \text{data}) = \frac{p(\text{data} \,|\, \text{theory})\, p(\text{theory})}{p(\text{data})}$$

Evidence

Bayesians: Let’s calculate the posterior! → Credible intervals

Frequentists: We don’t know the prior, so our statements are about how confident we are in our procedure → Confidence intervals

ML for Detector Simulation

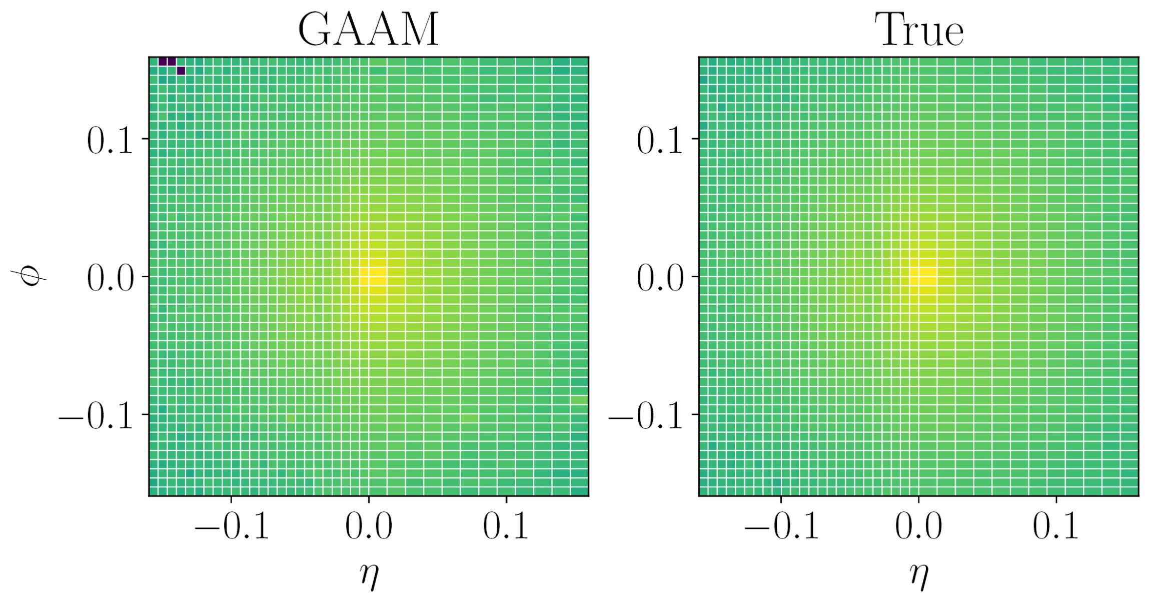


Image: arXiv:2305.11531

ML for Reconstruction

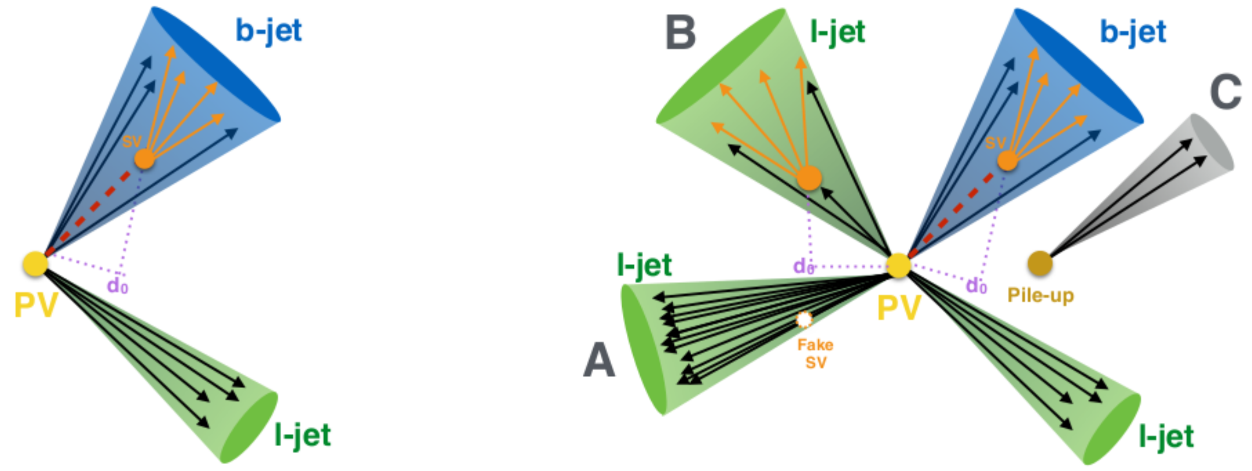


Image: USiegen group

ML for Signal vs Background

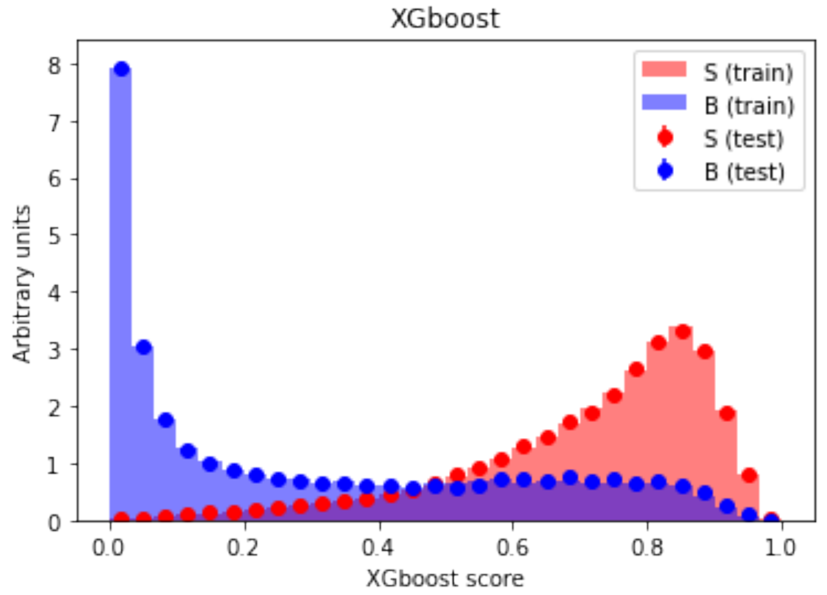


Image: ML4FP School

ML for Detector Simulation

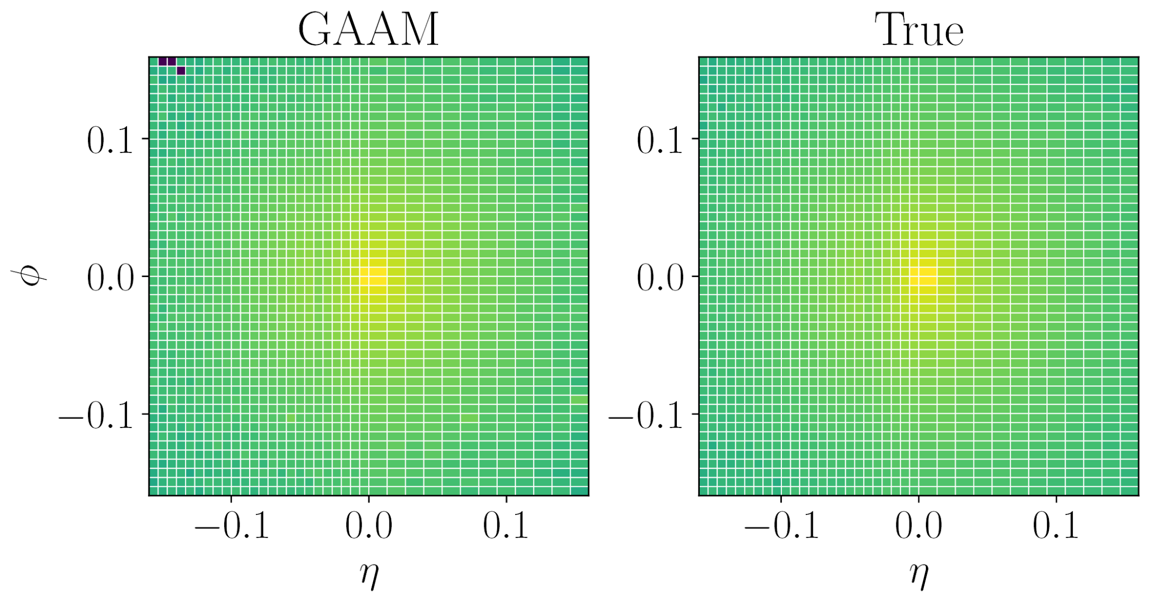


Image: arXiv:2305.11531

ML for Reconstruction

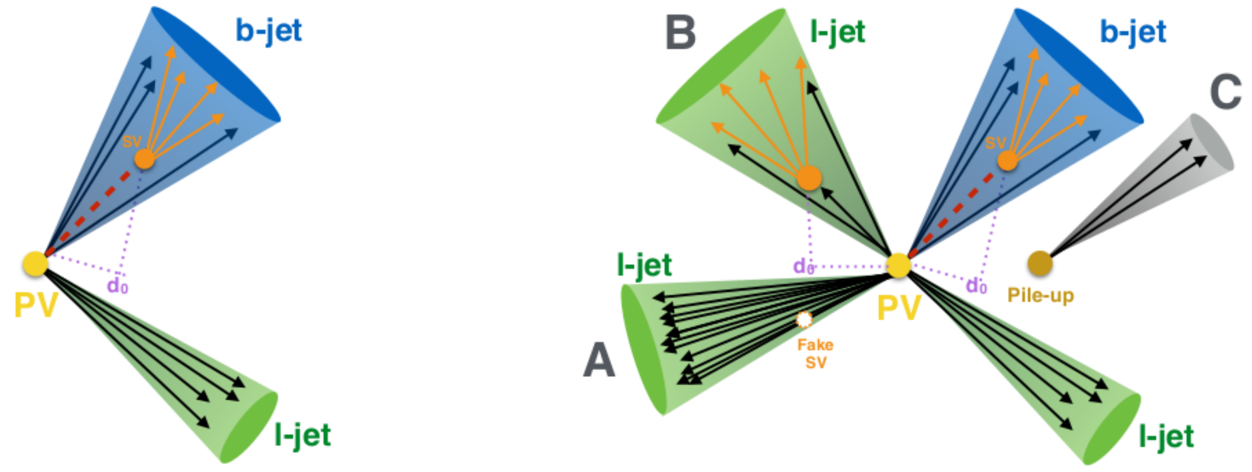


Image: USiegen group

ML for Signal vs Background

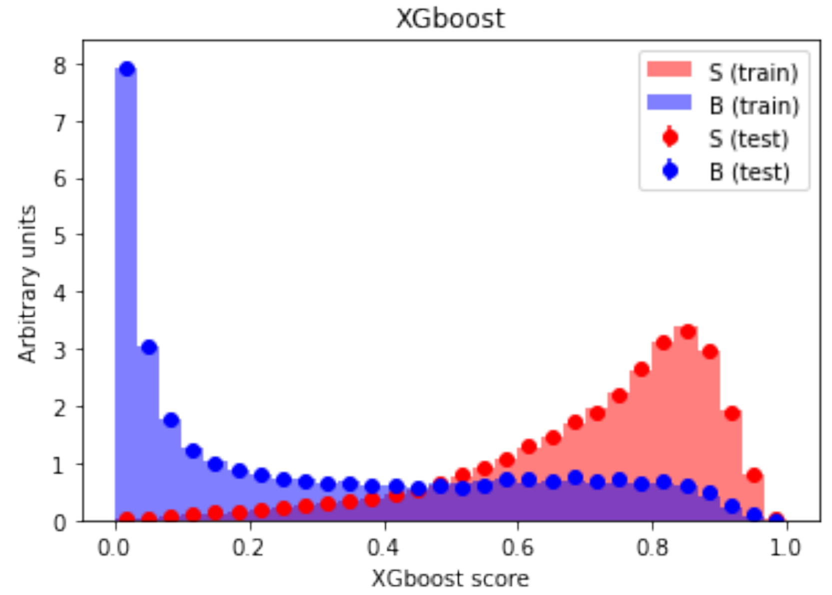


Image: ML4FP School

The final statistical inference step has remained the same

ML for Detector Simulation

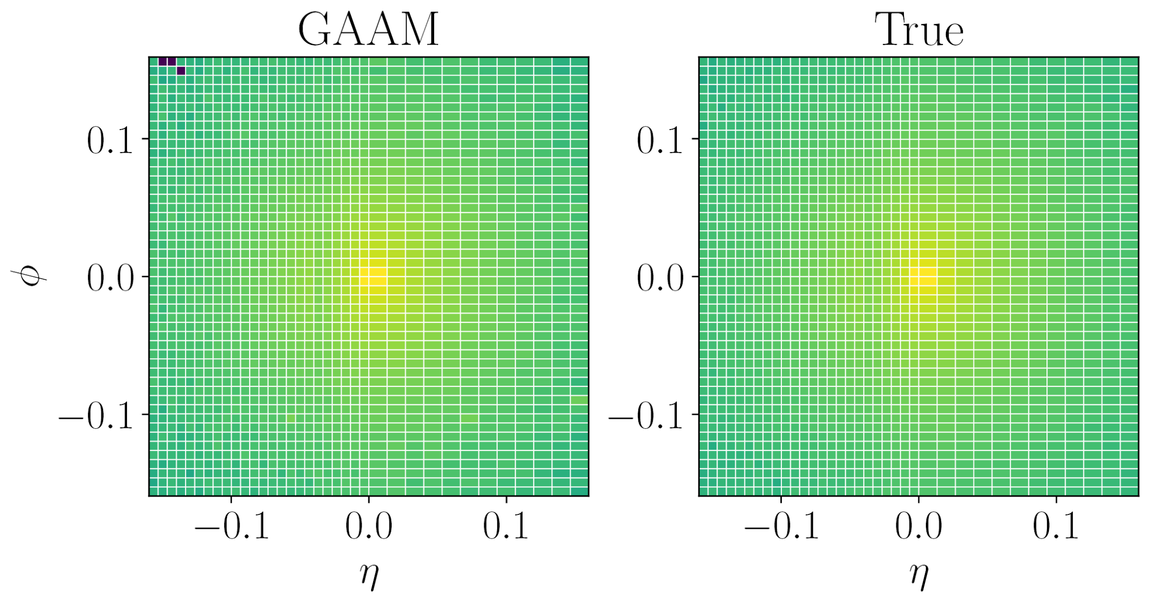


Image: arXiv:2305.11531

ML for Reconstruction

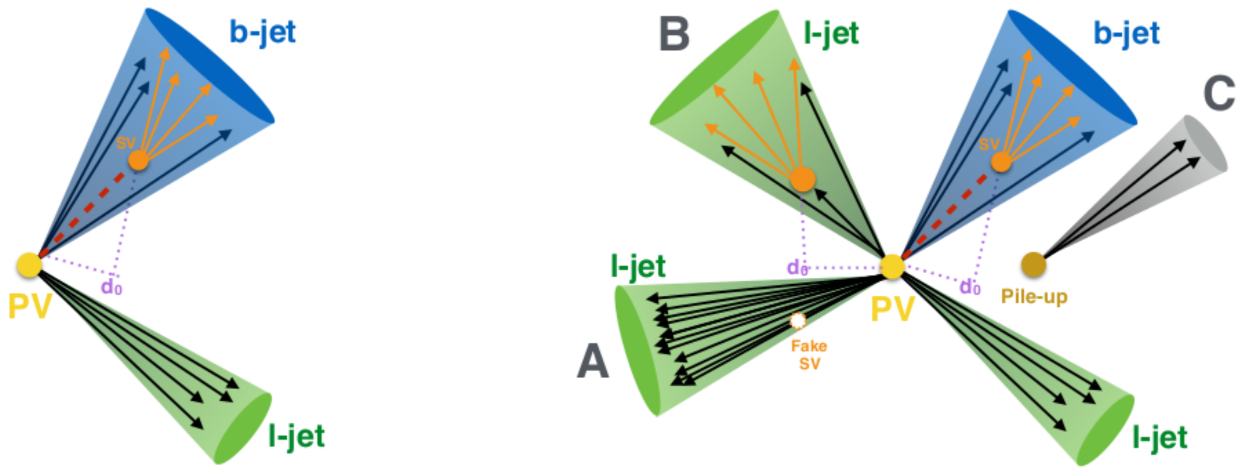


Image: USiegen group

ML for Signal vs Background

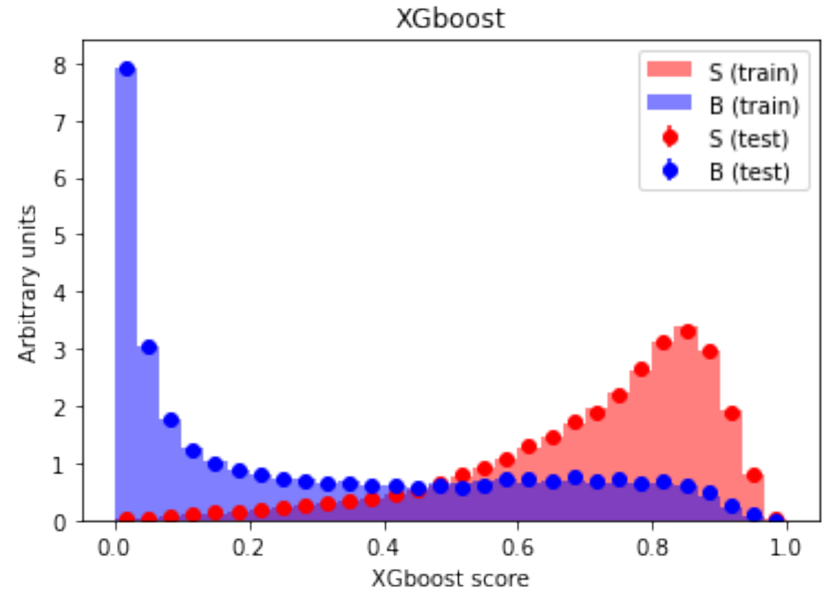
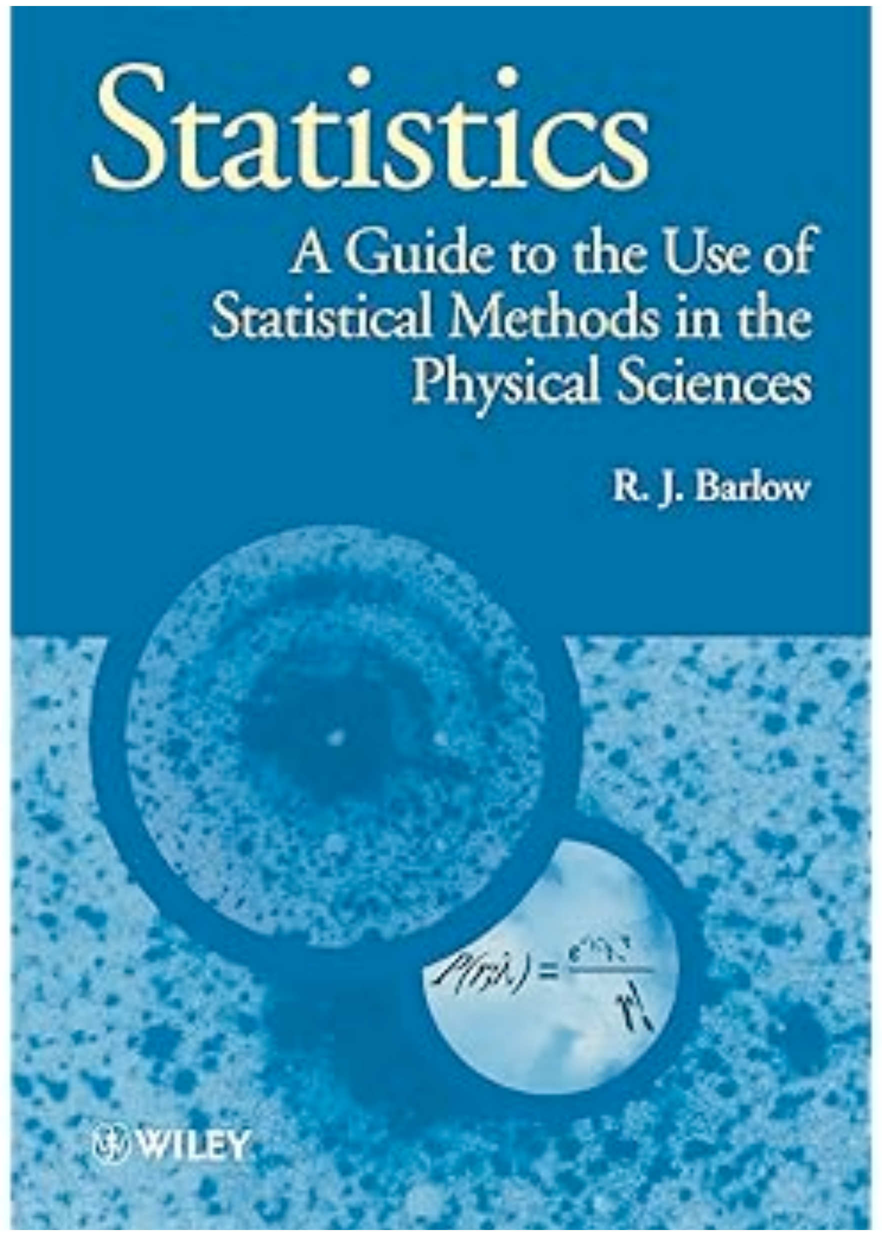
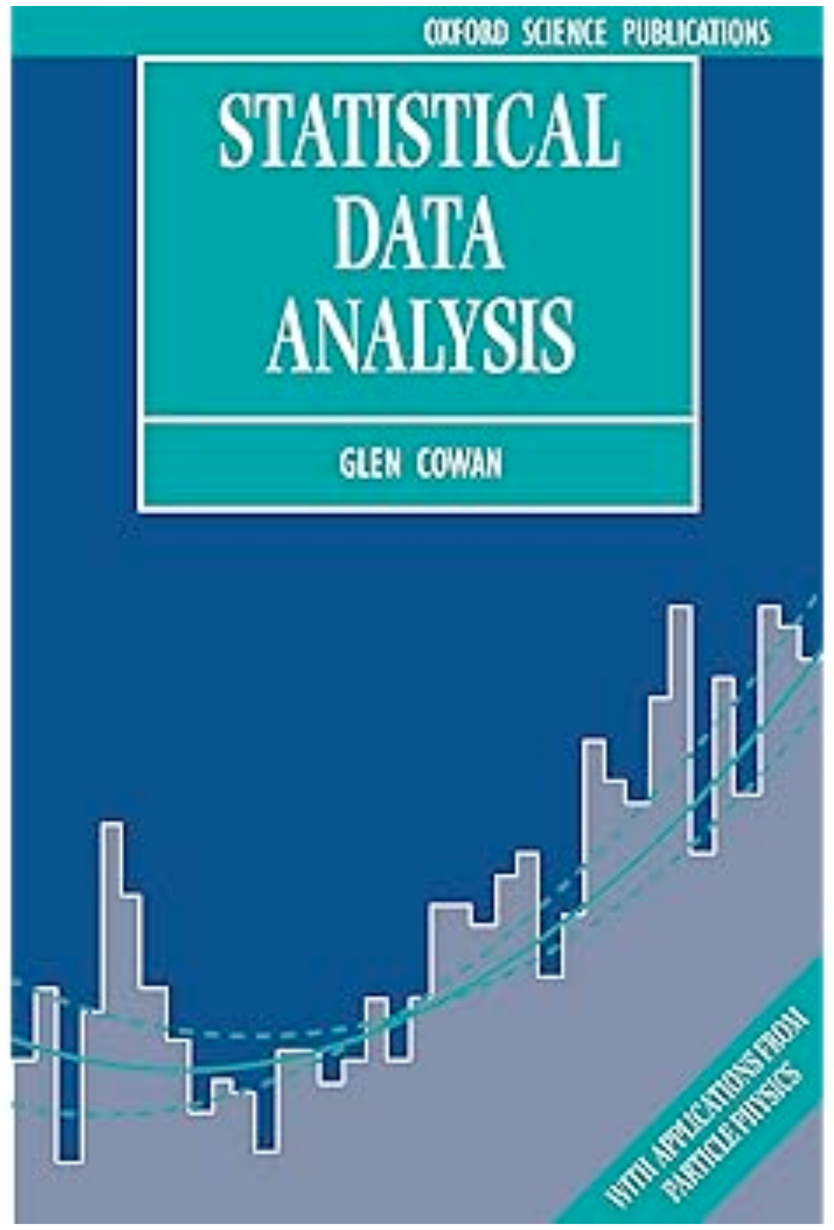


Image: ML4FP School

The final statistical inference step has remained the same



Publication date : November 1, 1993



Publication date : December 4, 1997

ML for Detector Simulation

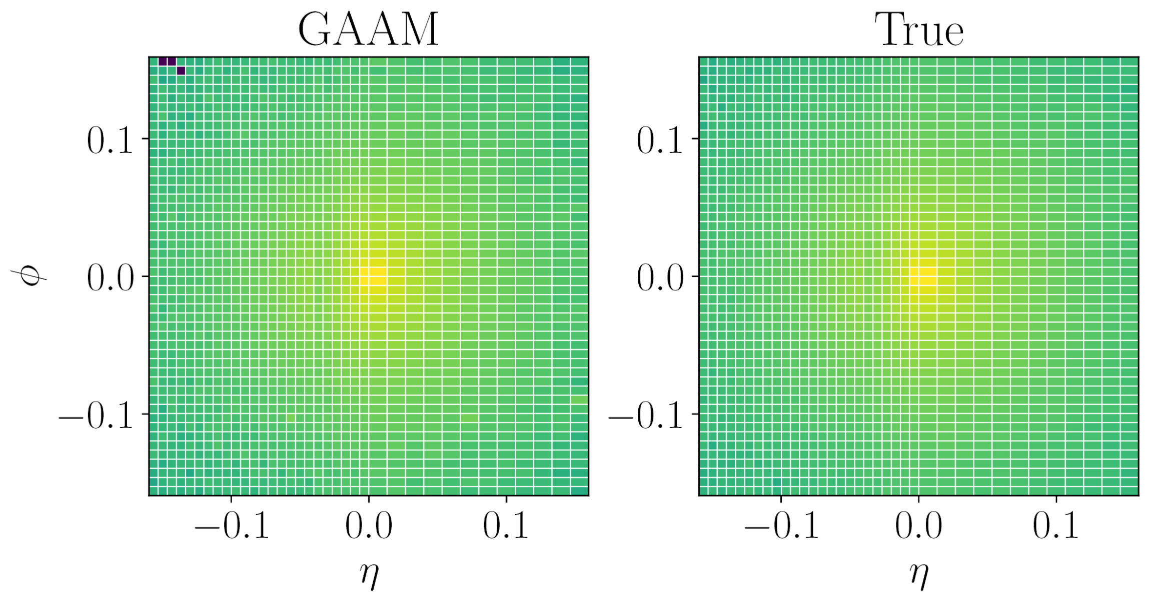


Image: arXiv:2305.11531

ML for Reconstruction

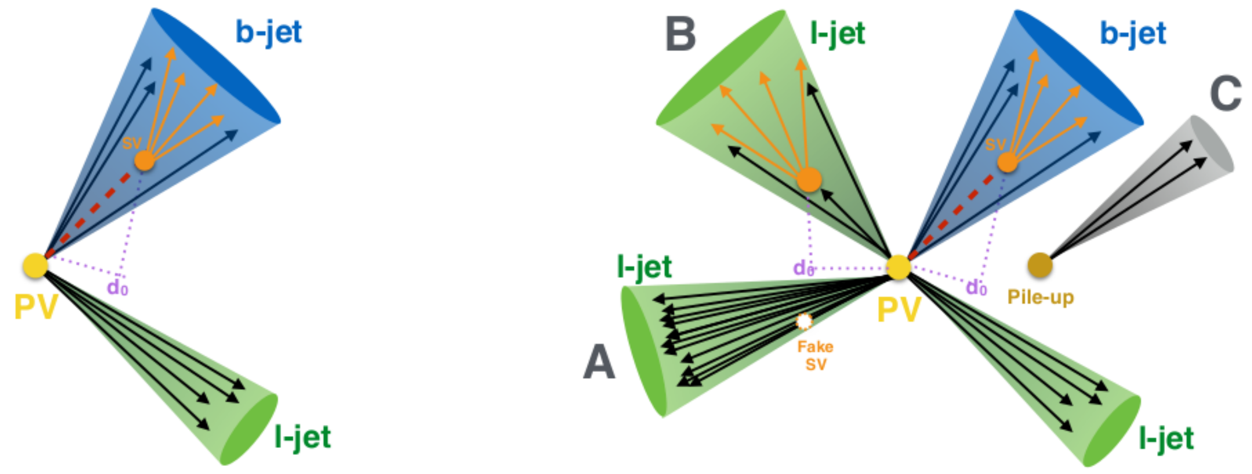


Image: USiegen group

ML for Signal vs Background

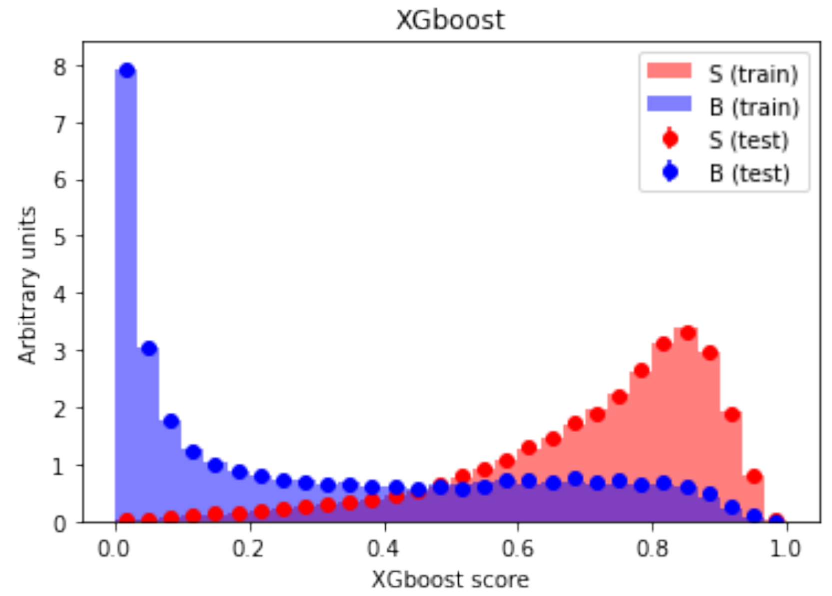
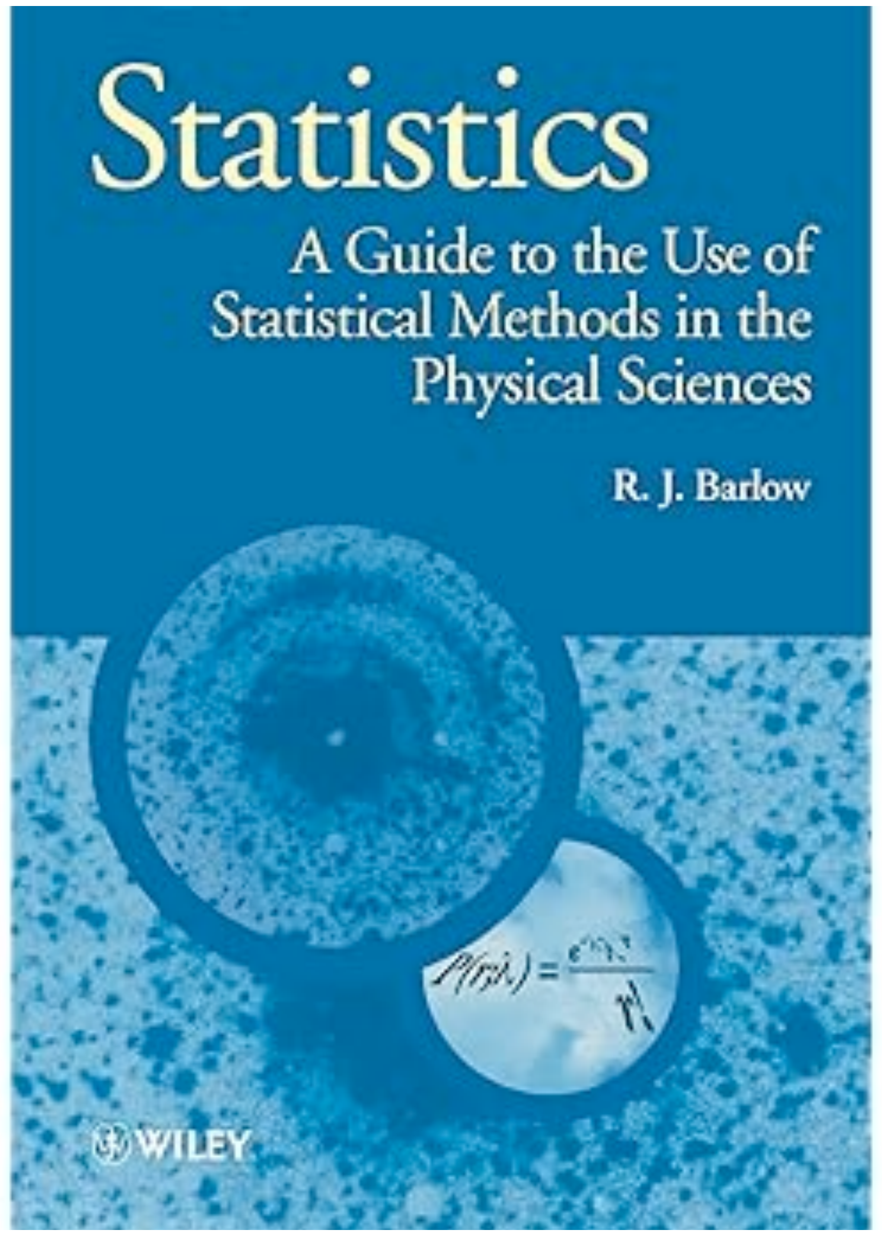


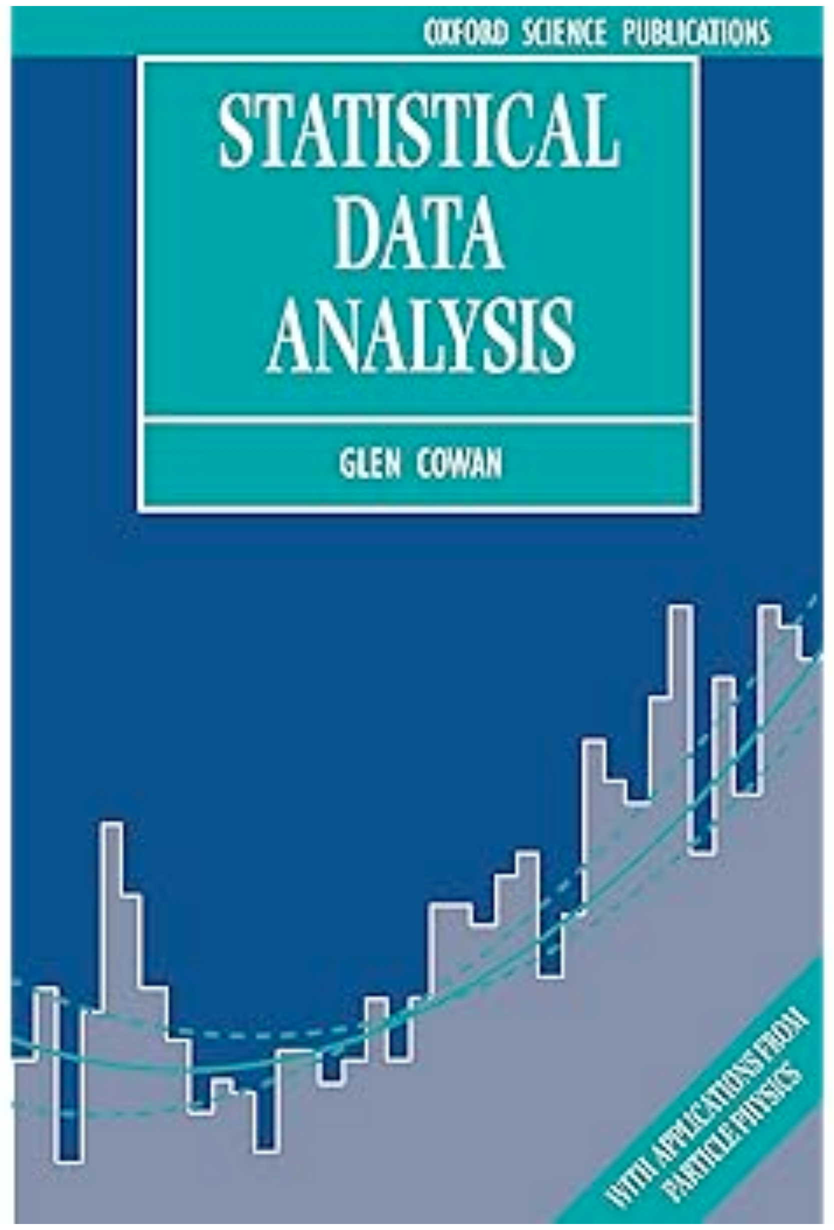
Image: ML4FP School

The final statistical inference step has remained the same



Publication date : November 1, 1993

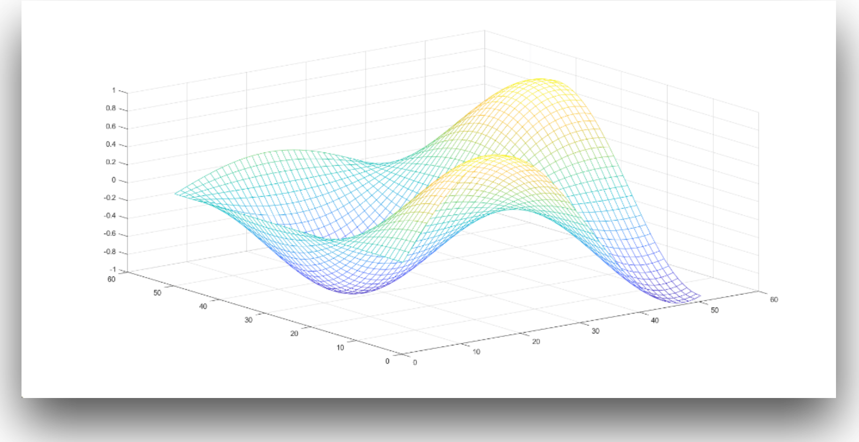
Time to modernise?



Publication date : December 4, 1997

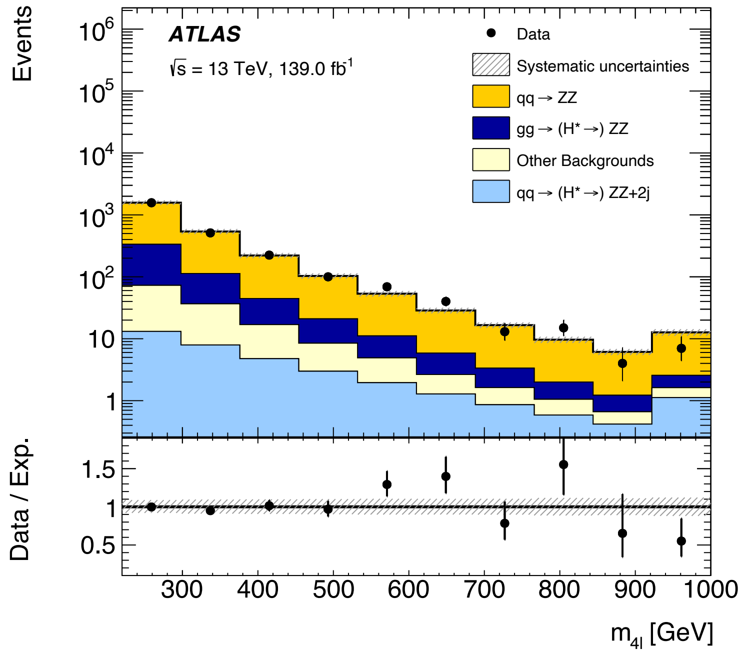
Perform likelihood ratio test using high-dimensional data

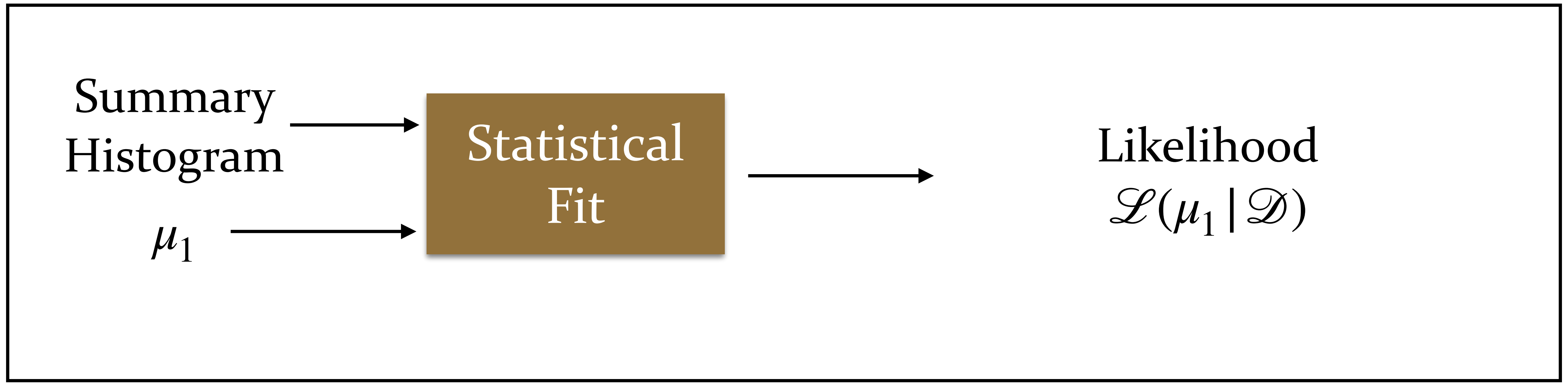
Traditional framework:



High-dim data

Summarisation to histogram





Summary Histogram

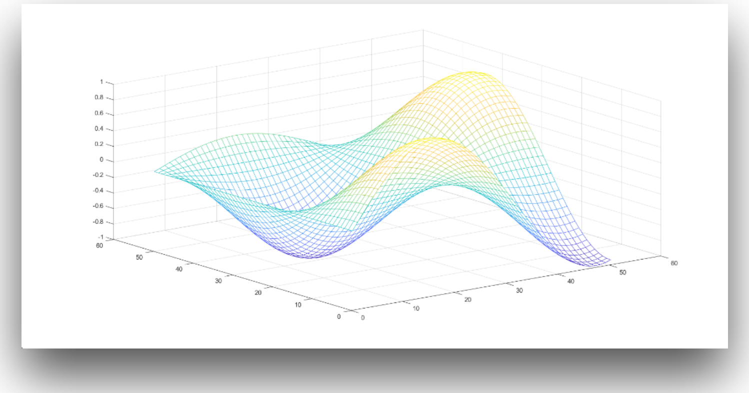
μ_1

Statistical Fit

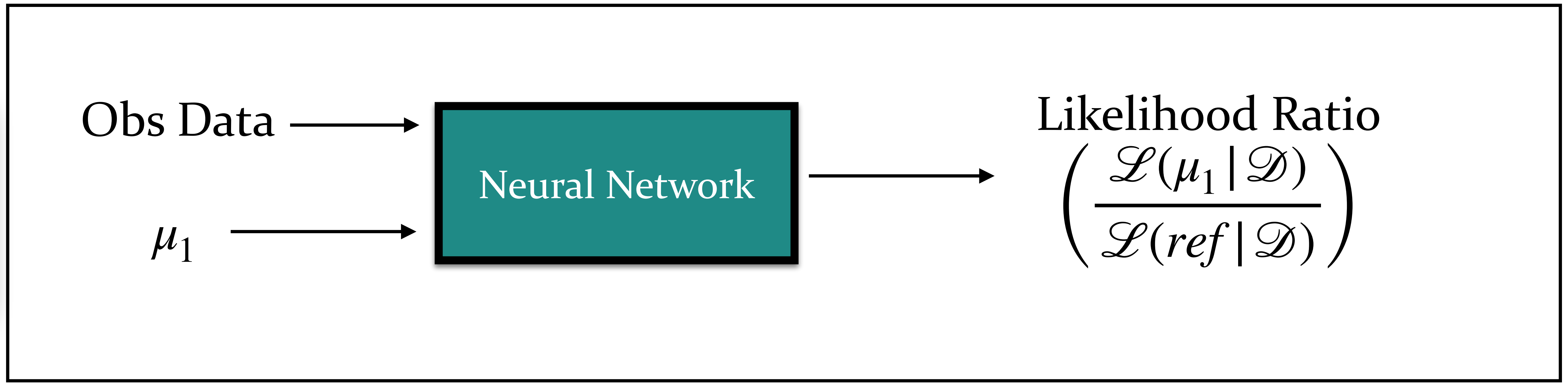
Likelihood $\mathscr{L}(\mu_1 | \mathscr{D})$

μ is now arbitrary parameter of interest(s)

Neural simulation-based inference framework:



High-dim data



Obs Data

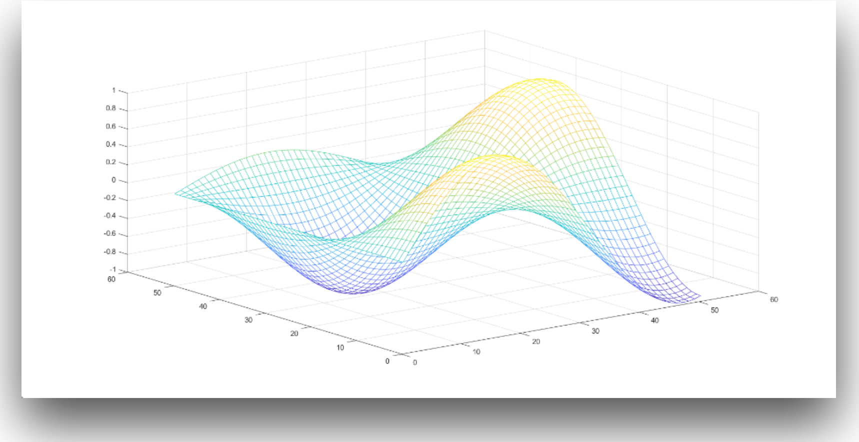
μ_1

Neural Network

Likelihood Ratio $\left(\dfrac{\mathscr{L}(\mu_1 | \mathscr{D})}{\mathscr{L}(ref | \mathscr{D})}\right)$

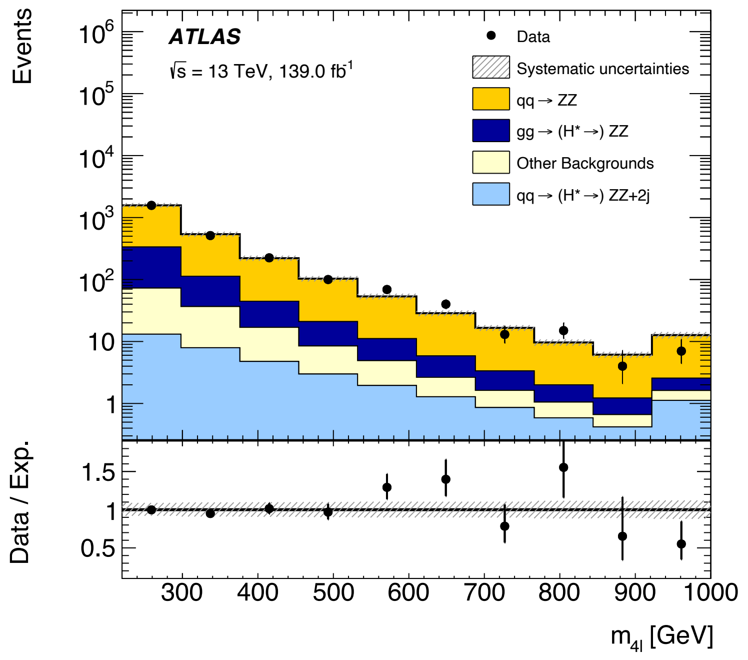
Perform likelihood ratio test using high-dimensional data

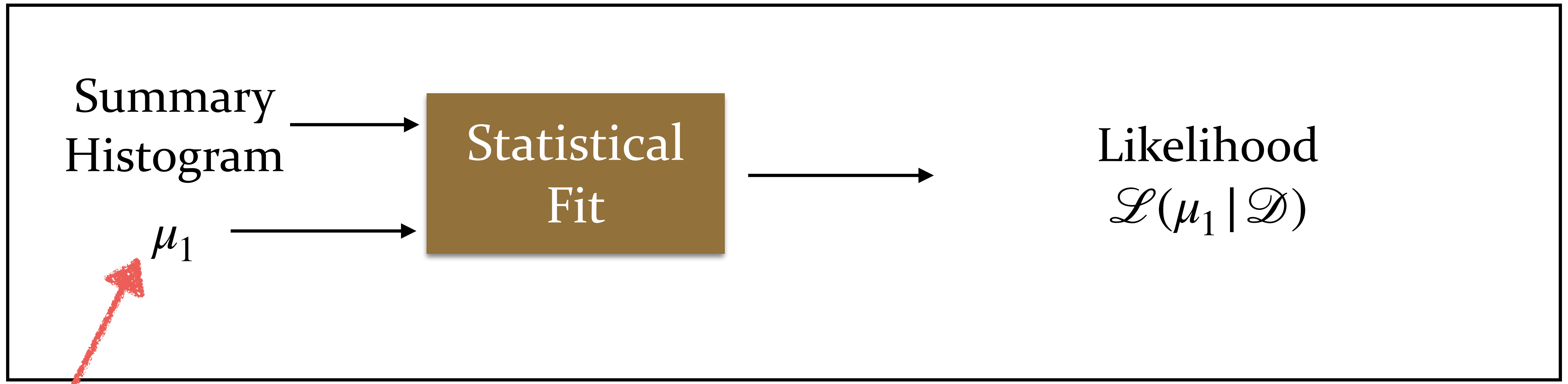
Traditional framework:



High-dim data

Summarisation to histogram





Summary Histogram

μ_1

Statistical Fit

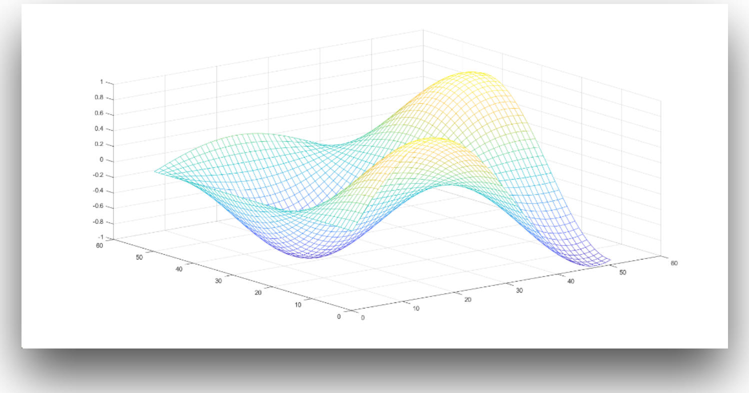
Likelihood

$\mathscr{L}(\mu_1 | \mathscr{D})$

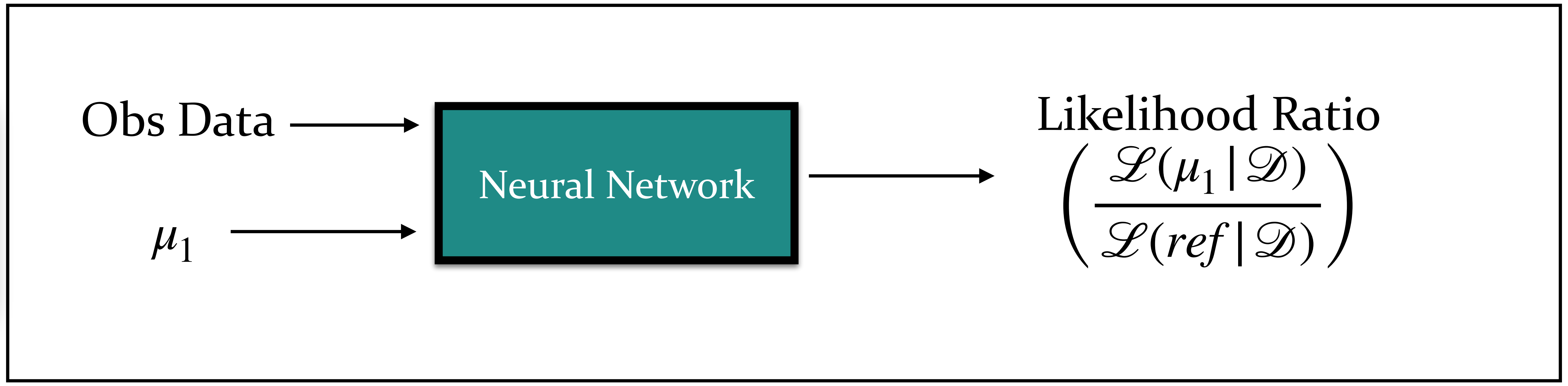
Hypothesis μ_1

μ is now arbitrary parameter of interest(s)

Neural simulation-based inference framework:



High-dim data



Obs Data

μ_1

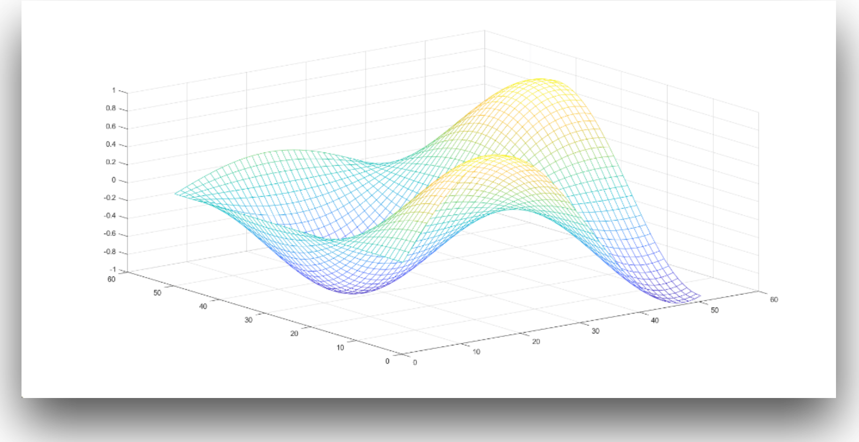
Neural Network

Likelihood Ratio

$$\left(\frac{\mathscr{L}(\mu_1 | \mathscr{D})}{\mathscr{L}(ref | \mathscr{D})} \right)$$

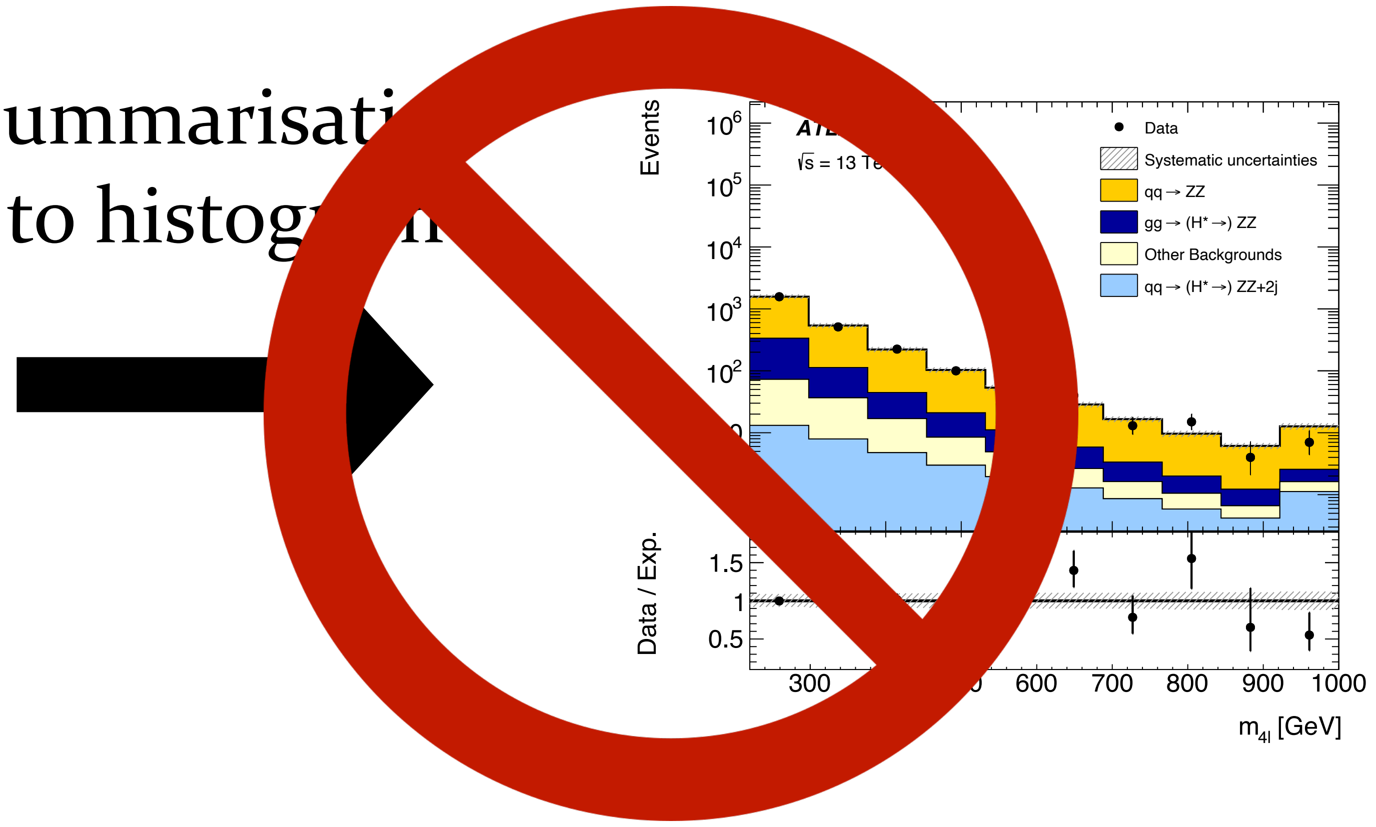
Perform likelihood ratio test using high-dimensional data

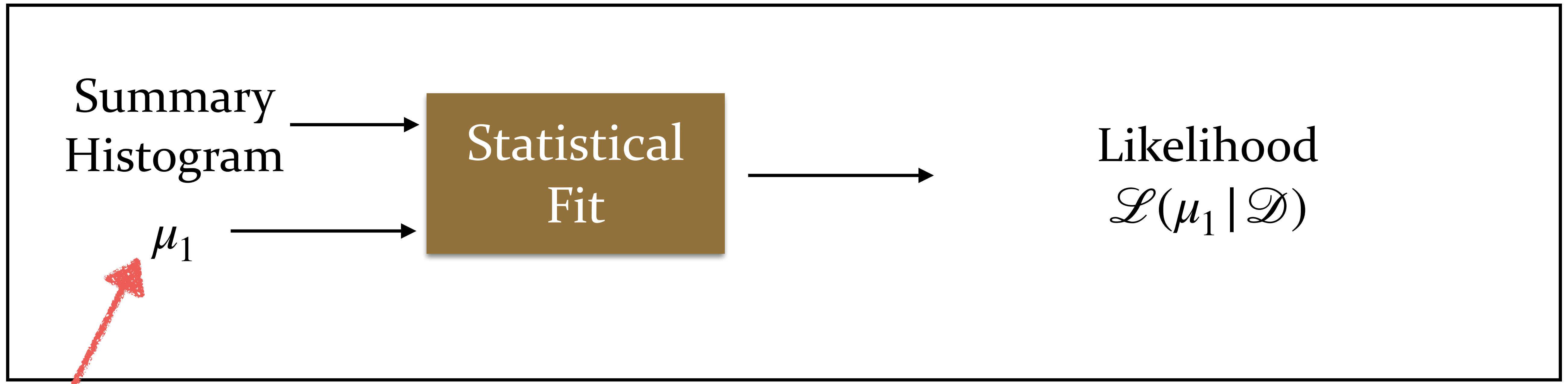
Traditional framework:



High-dim data

Summarisati
to histog





Summary Histogram

μ_1

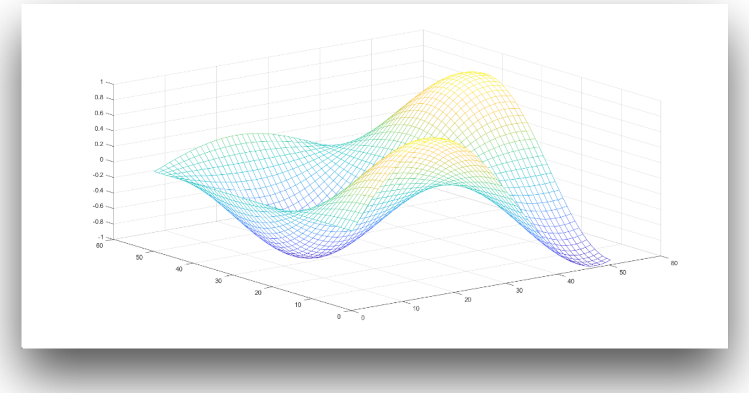
Statistical Fit

Likelihood $\mathscr{L}(\mu_1 | \mathscr{D})$

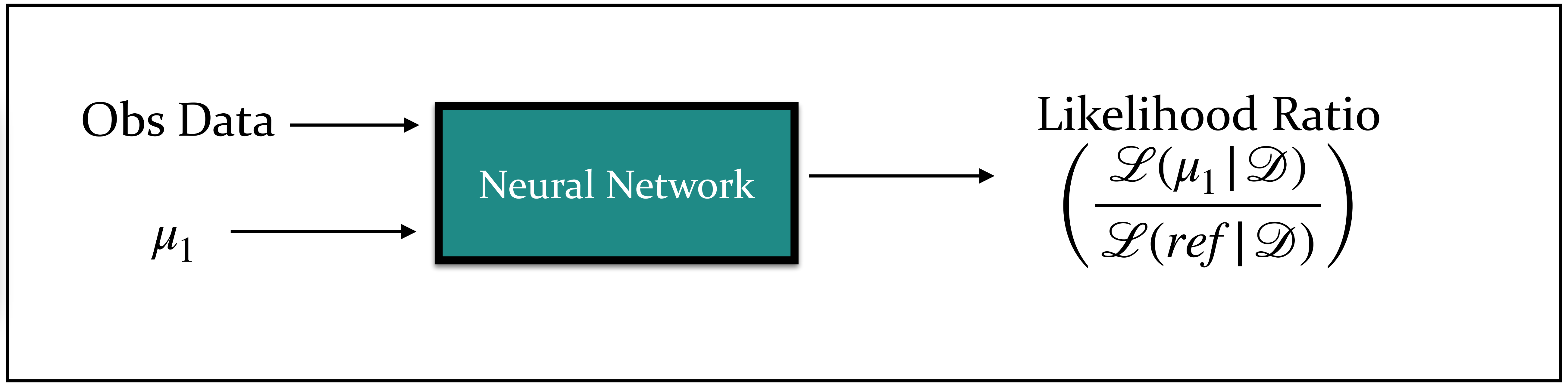
Hypothesis μ_1

μ is now arbitrary parameter of interest(s)

Neural simulation-based inference framework:



High-dim data



Obs Data

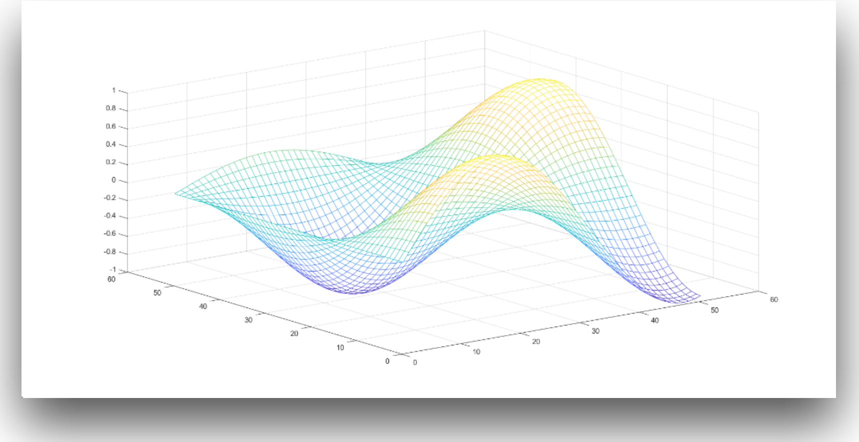
μ_1

Neural Network

Likelihood Ratio $\left(\frac{\mathscr{L}(\mu_1 | \mathscr{D})}{\mathscr{L}(ref | \mathscr{D})}\right)$

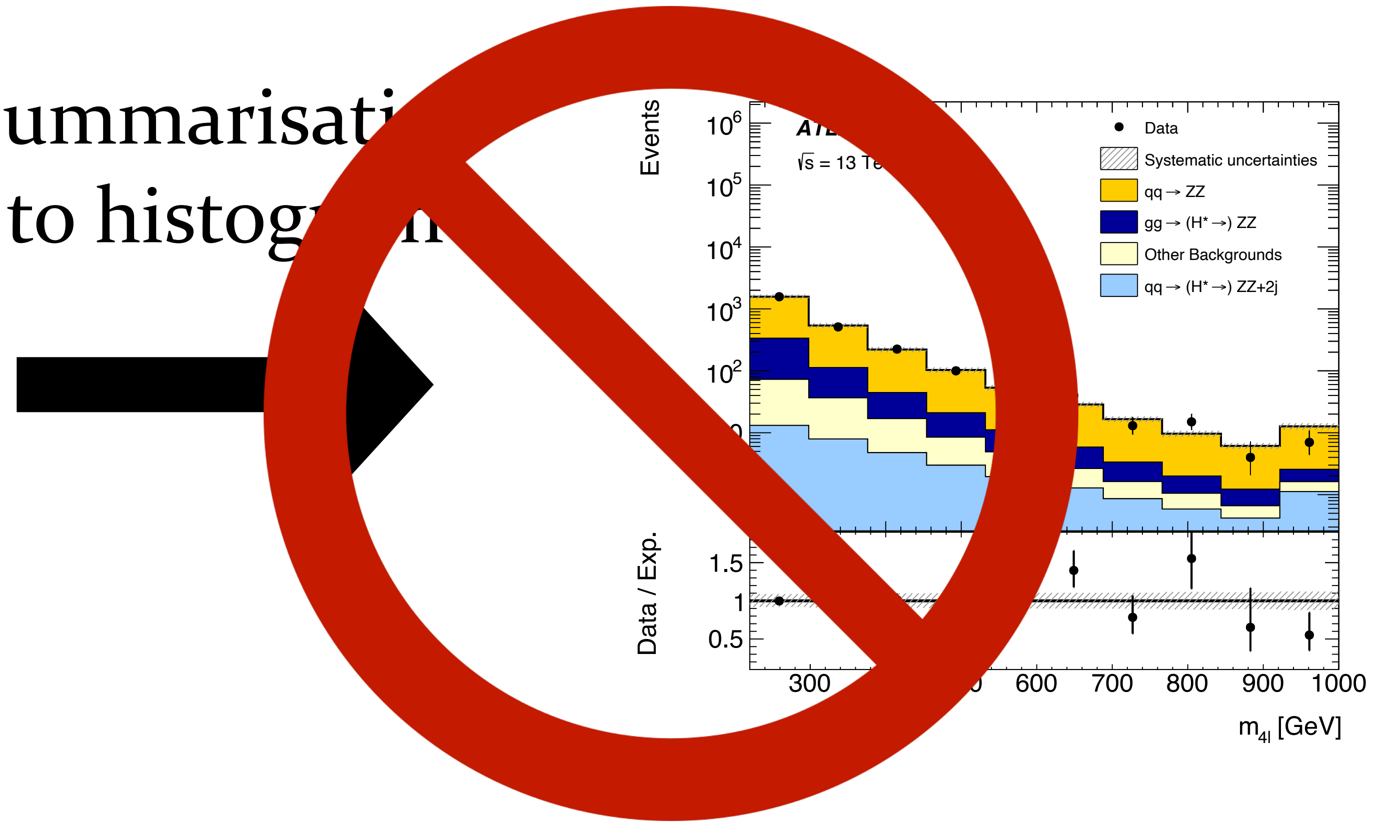
Perform likelihood ratio test using high-dimensional data

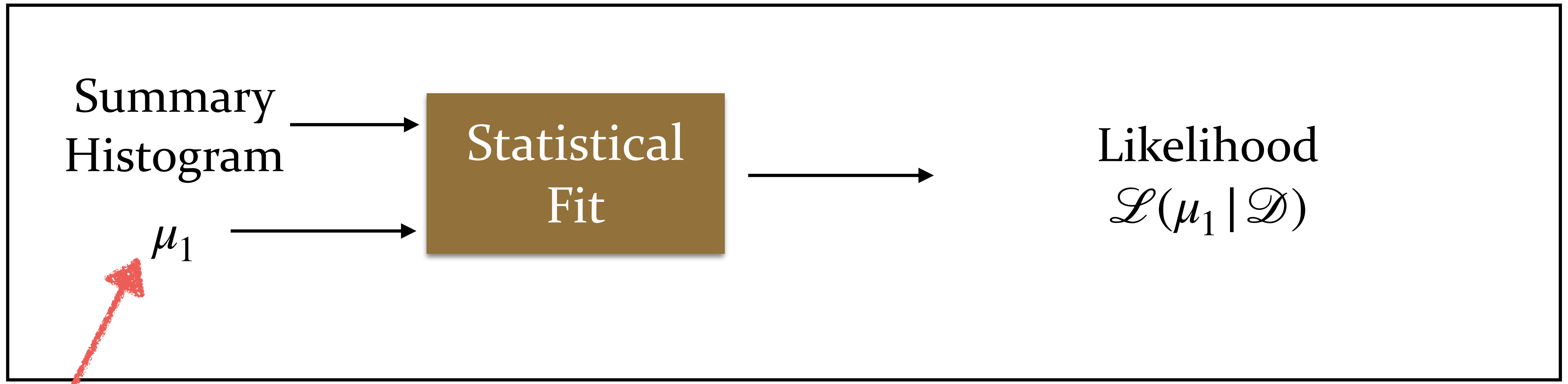
Traditional framework:



High-dim data

Summarisati
to histog





Summary Histogram → **Statistical Fit** → Likelihood

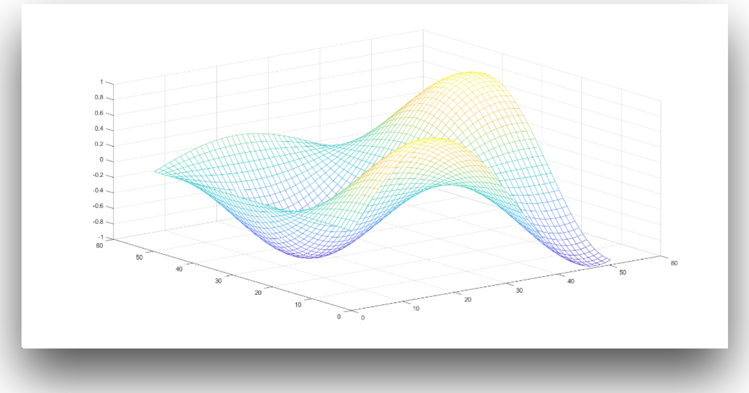
μ_1 →

$\mathscr{L}(\mu_1 | \mathscr{D})$

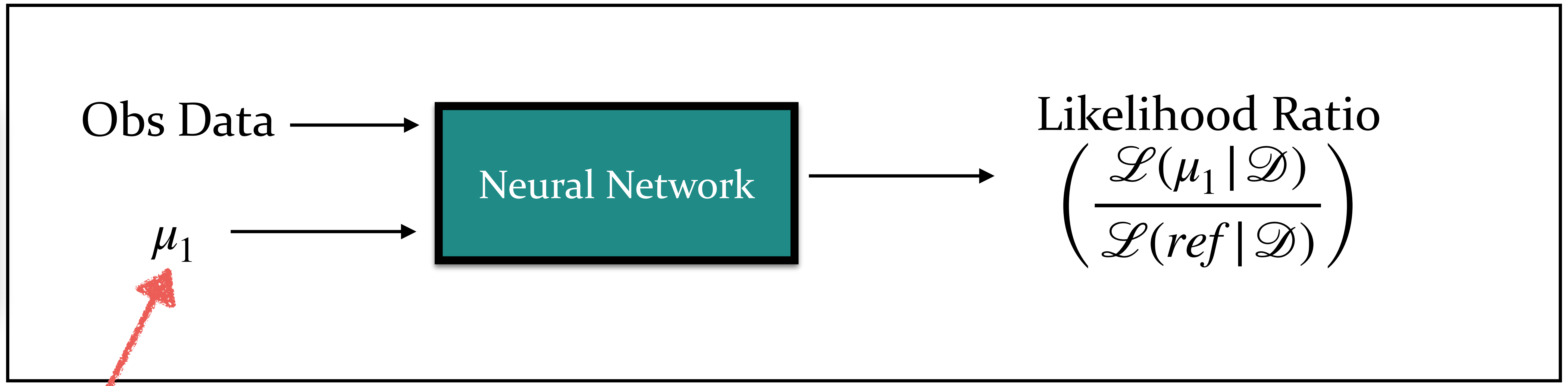
Hypothesis μ_1

μ is now arbitrary parameter of interest(s)

Neural simulation-based inference framework:



High-dim data



Obs Data → **Neural Network** → Likelihood Ratio

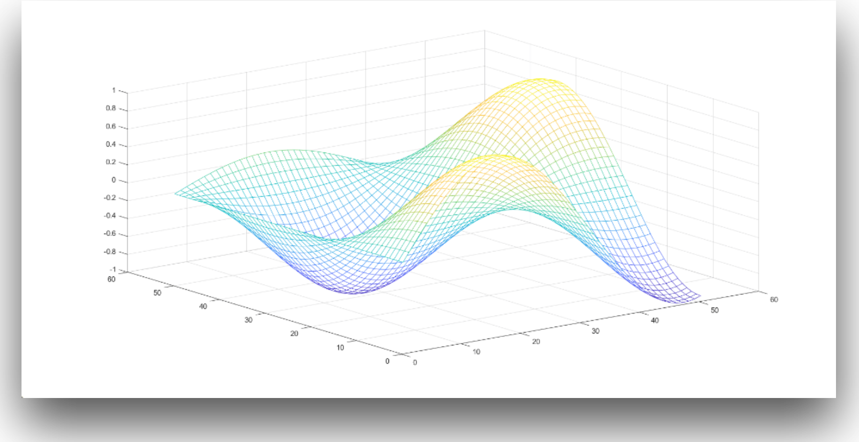
μ_1 →

$$\left(\frac{\mathscr{L}(\mu_1 | \mathscr{D})}{\mathscr{L}(ref | \mathscr{D})} \right)$$

Hypothesis μ_1

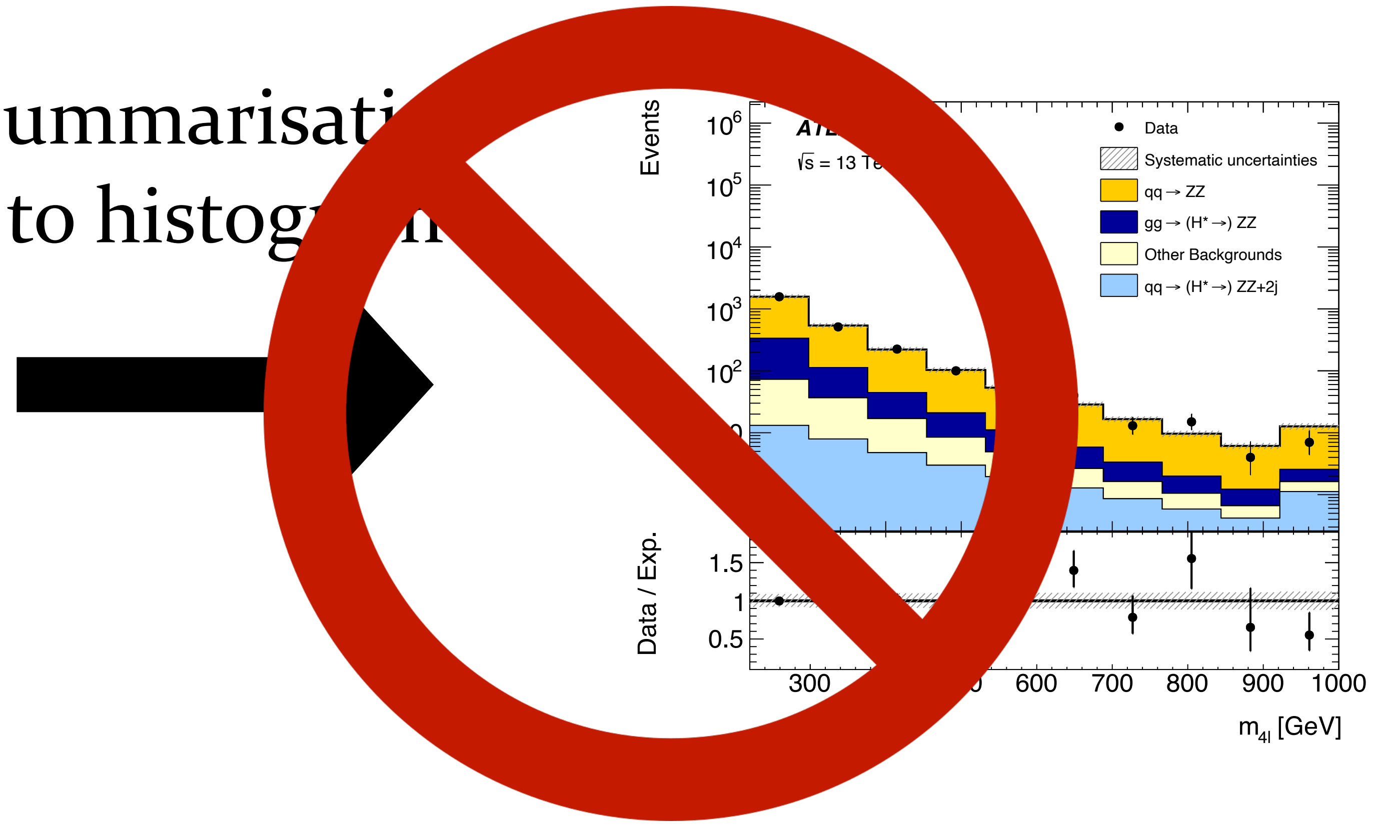
Perform likelihood ratio test using high-dimensional data

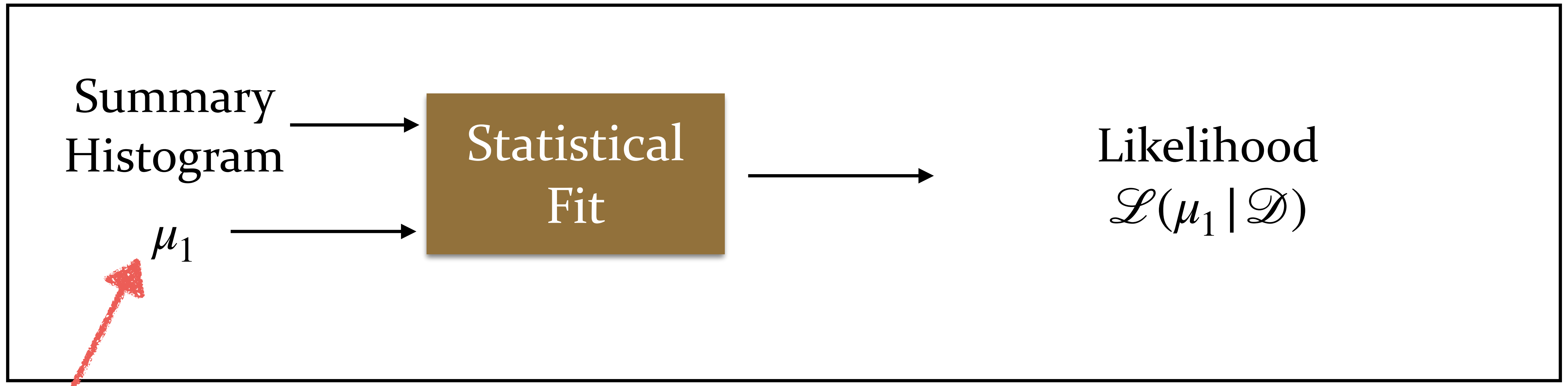
Traditional framework:



High-dim data

Summarisati
to histog





Summary Histogram →

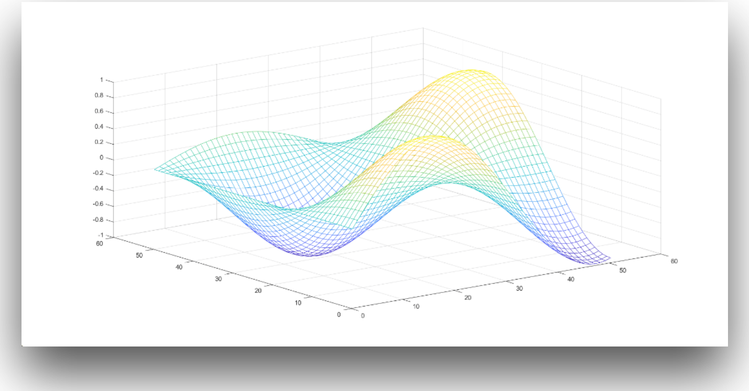
μ_1 →

Statistical Fit → Likelihood $\mathscr{L}(\mu_1 | \mathscr{D})$

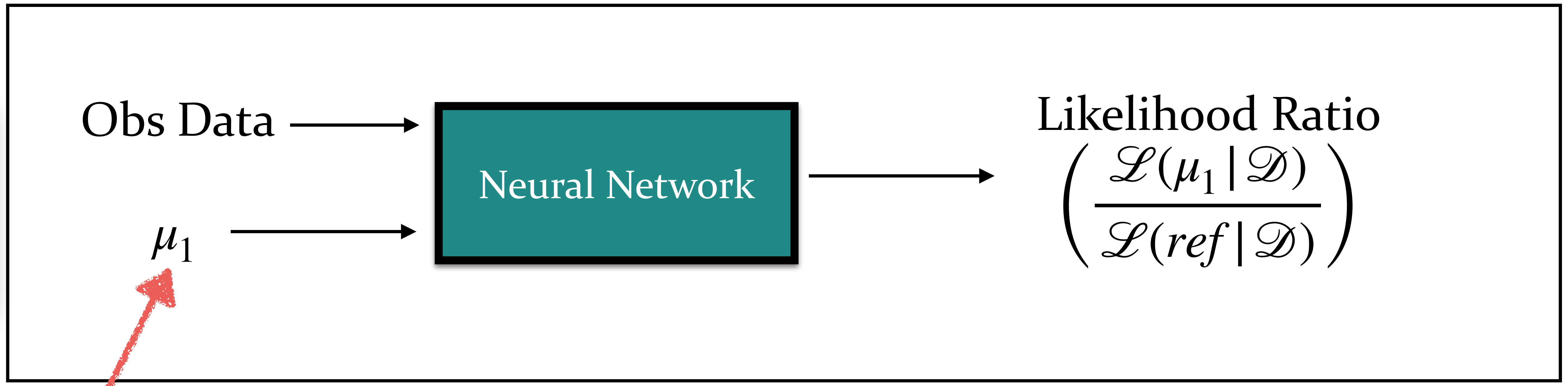
Hypothesis μ_1

μ is now arbitrary parameter of interest(s)

Neural simulation-based inference framework:



High-dim data



Obs Data →

μ_1 →

Neural Network →

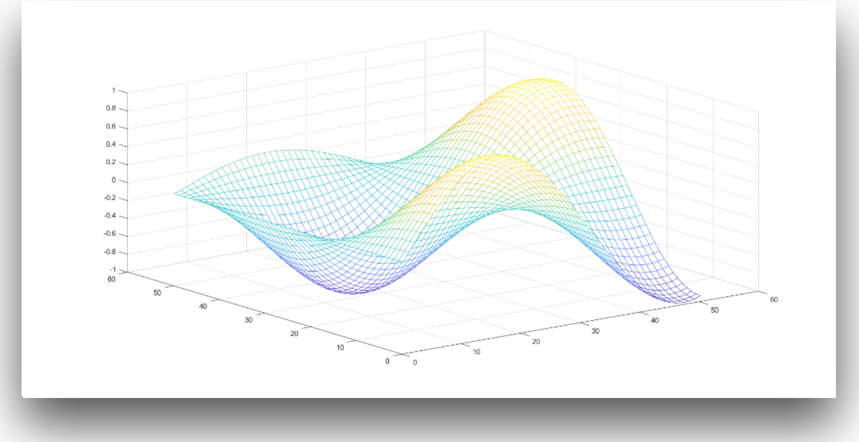
Likelihood Ratio
$$\left(\frac{\mathscr{L}(\mu_1 | \mathscr{D})}{\mathscr{L}(ref | \mathscr{D})}\right)$$

Hypothesis μ_1

Train on simulations, apply on data

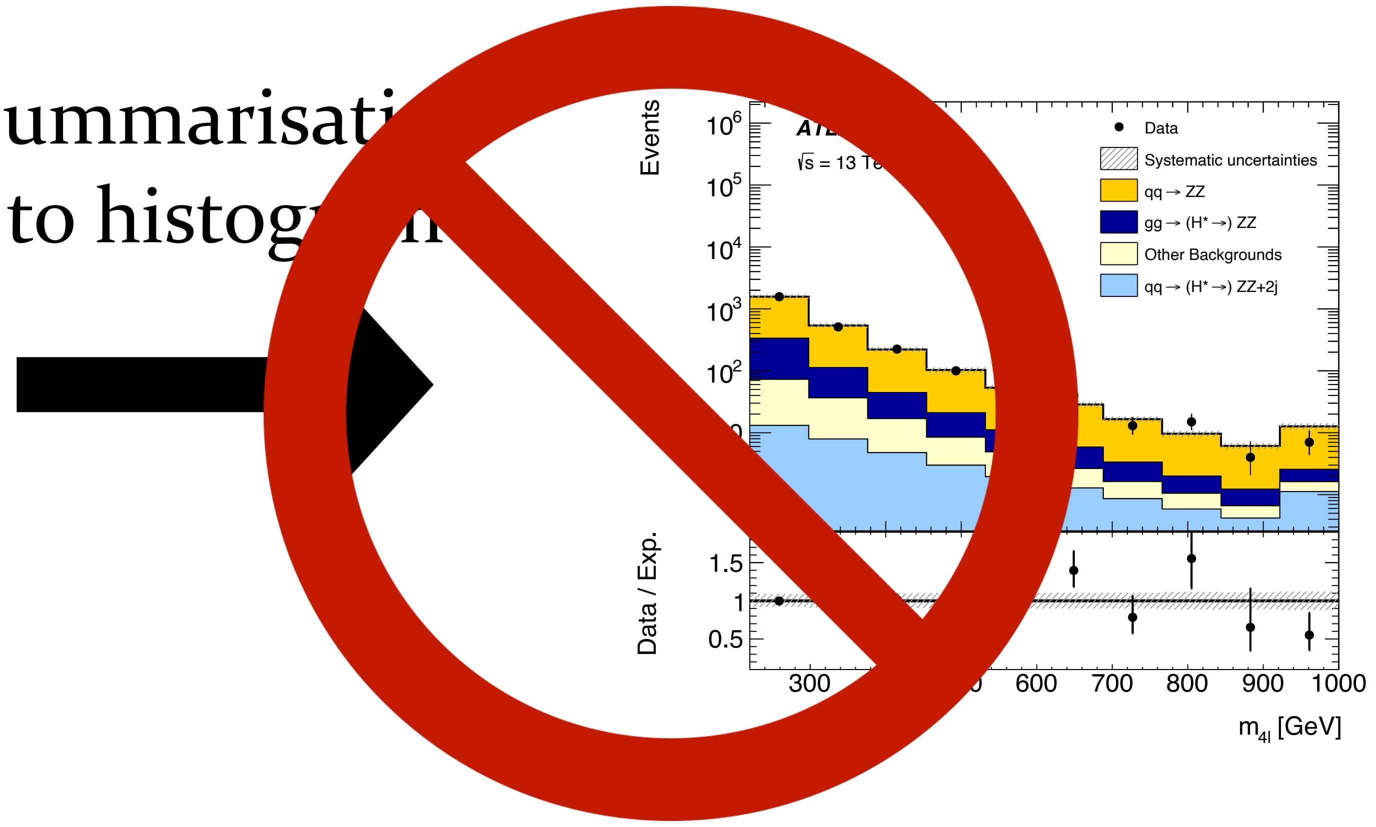
Perform likelihood ratio test using high-dimensional data

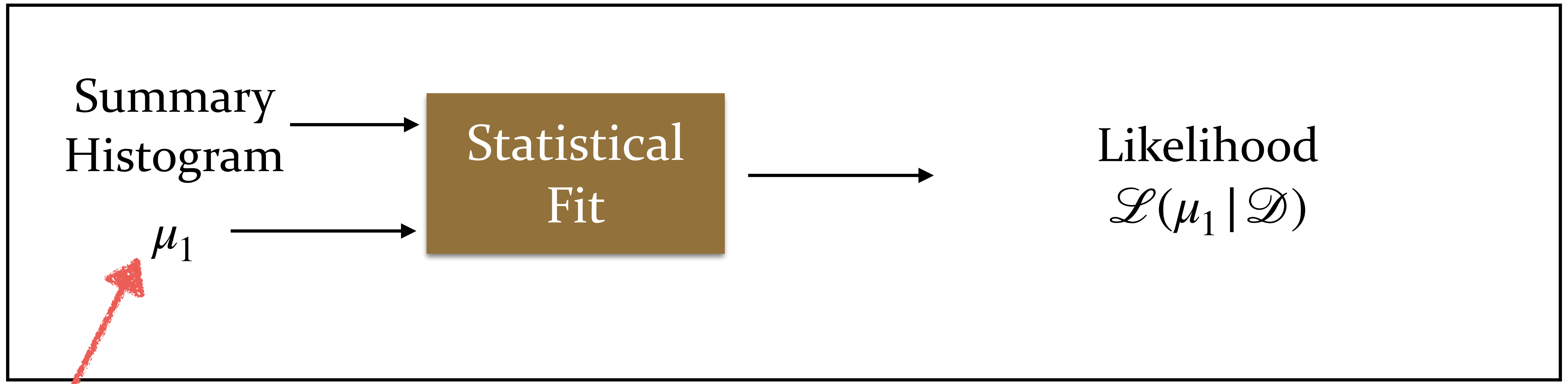
Traditional framework:



High-dim data

Summarisati
to histogr





Summary Histogram → Statistical Fit

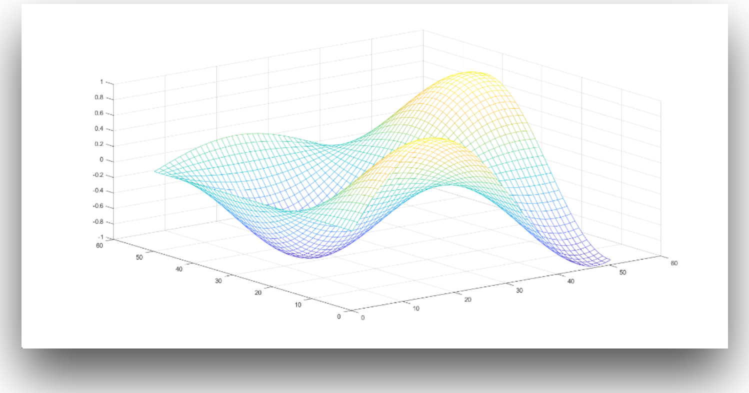
μ_1 → Statistical Fit

Statistical Fit → Likelihood $\mathscr{L}(\mu_1 | \mathscr{D})$

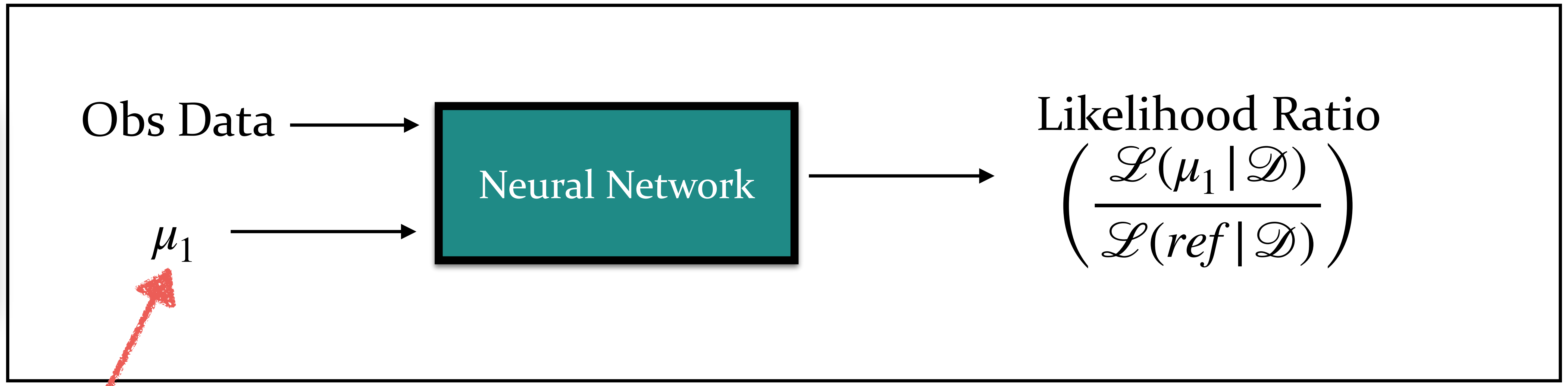
Hypothesis μ_1

μ is now arbitrary parameter of interest(s)

Neural simulation-based inference framework:



High-dim data



Obs Data → Neural Network

μ_1 → Neural Network

Neural Network → Likelihood Ratio $\left(\dfrac{\mathscr{L}(\mu_1 | \mathscr{D})}{\mathscr{L}(ref | \mathscr{D})} \right)$

See more in Simulation-based inference lecture by Eddie today!

Hypothesis μ_1

Train on simulations, apply on data

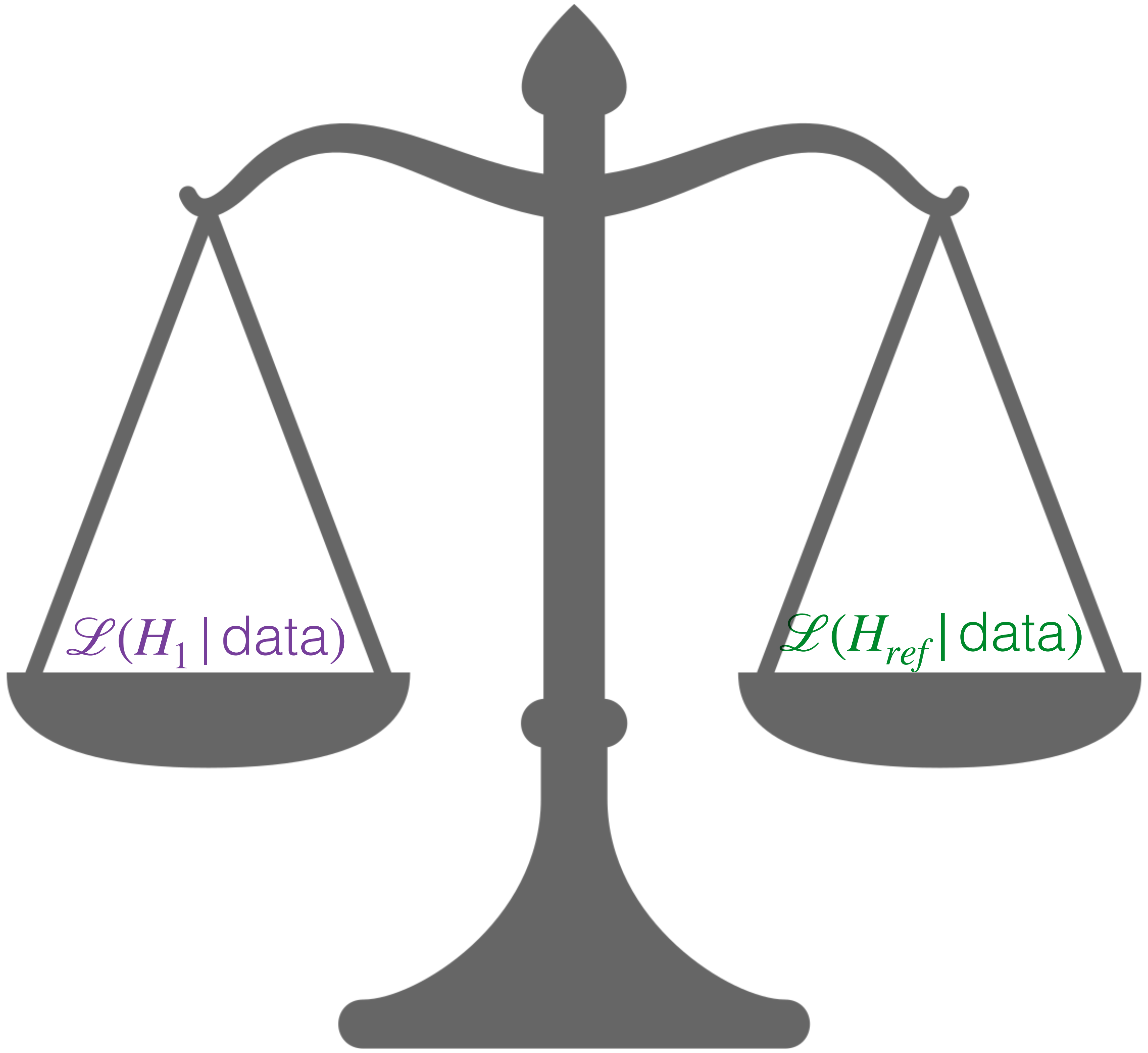
Until now, we have replaced individual pieces with ML in age-old likelihood ratio test

Do we dare question the test itself?

Until now, we have replaced individual pieces with ML in age-old likelihood ratio test

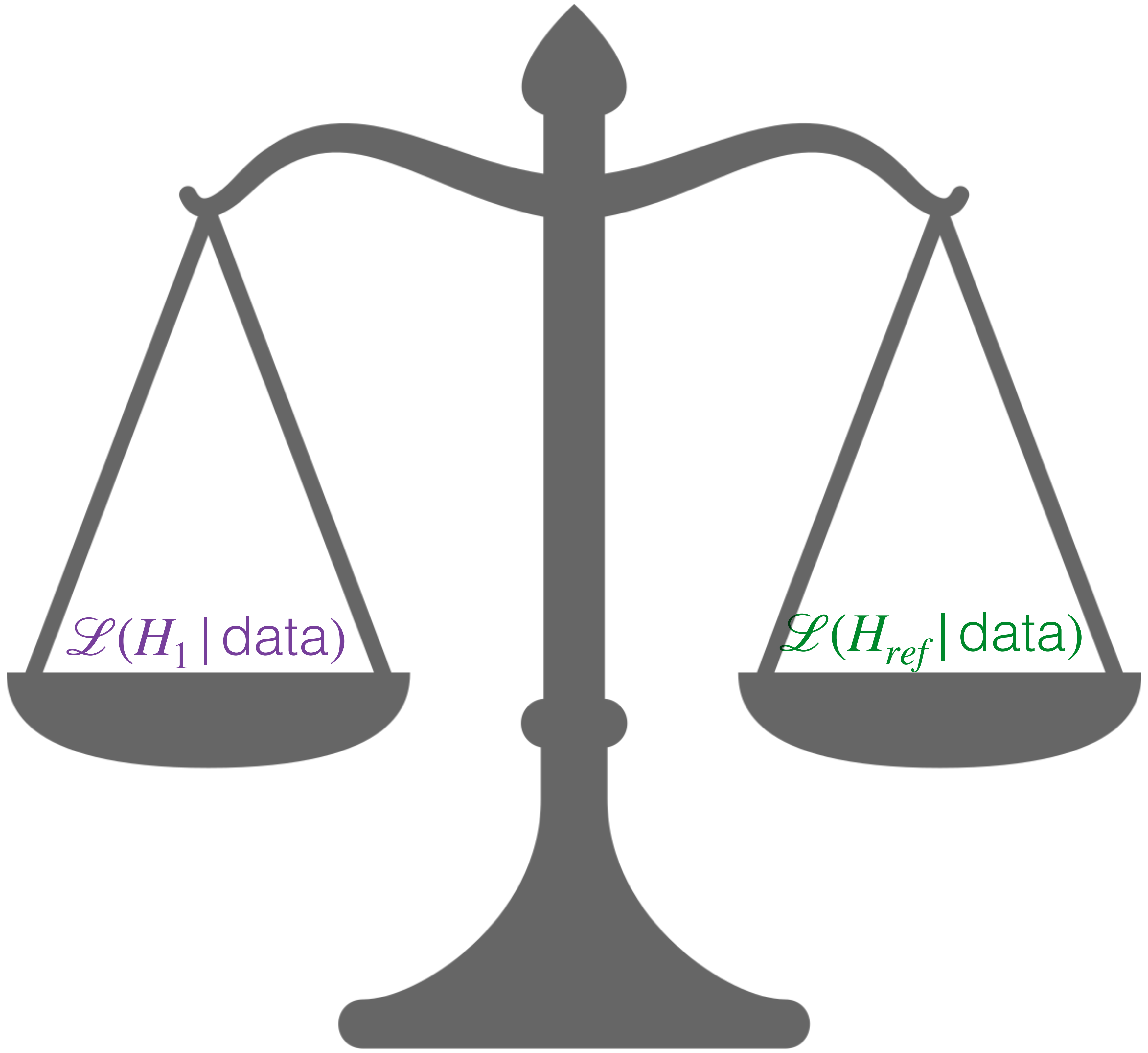
Do we dare question the test itself?

Hypothesis tests

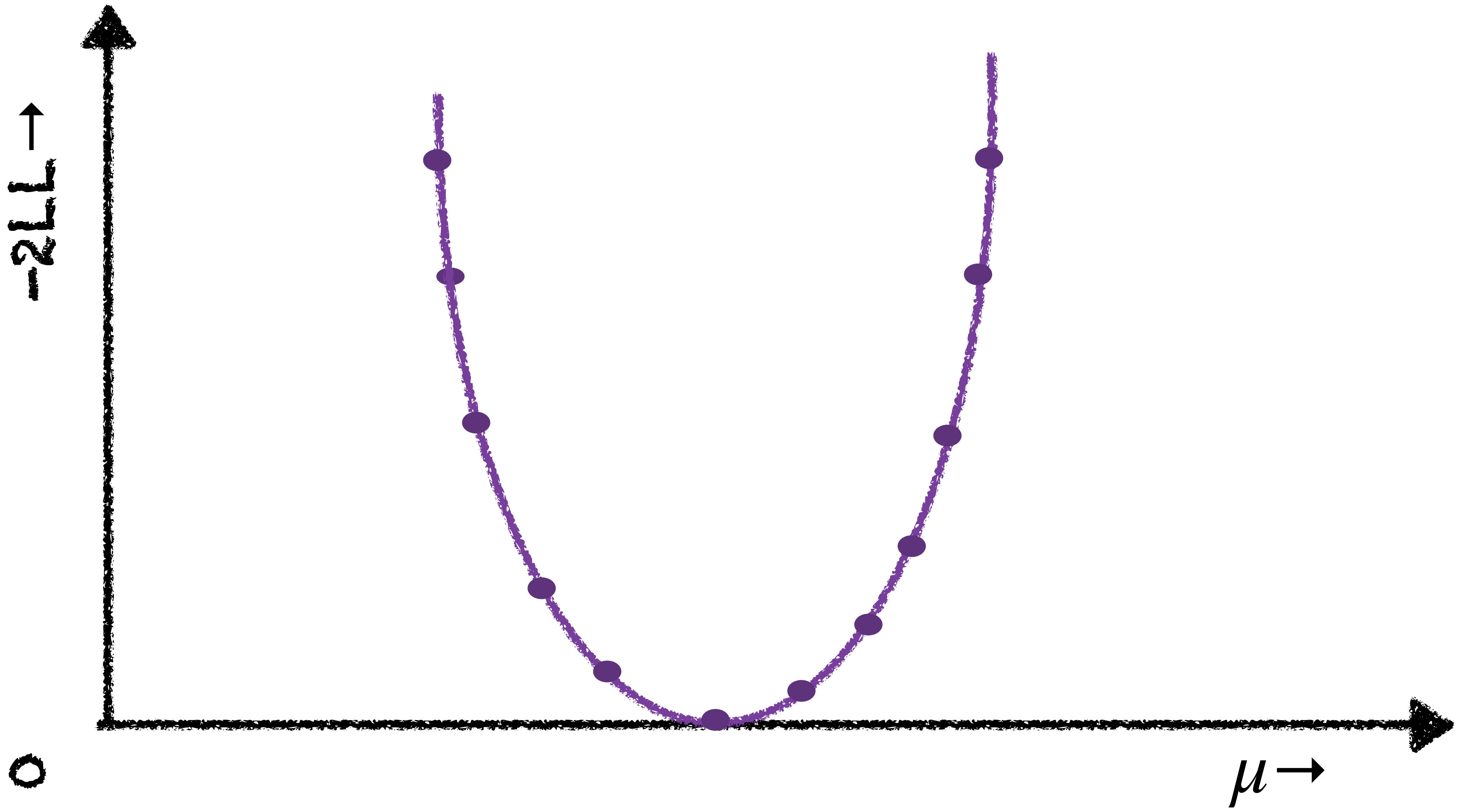


When comparing 2 hypotheses, guaranteed to be optimal test by Neyman-Person lemma

Hypothesis tests

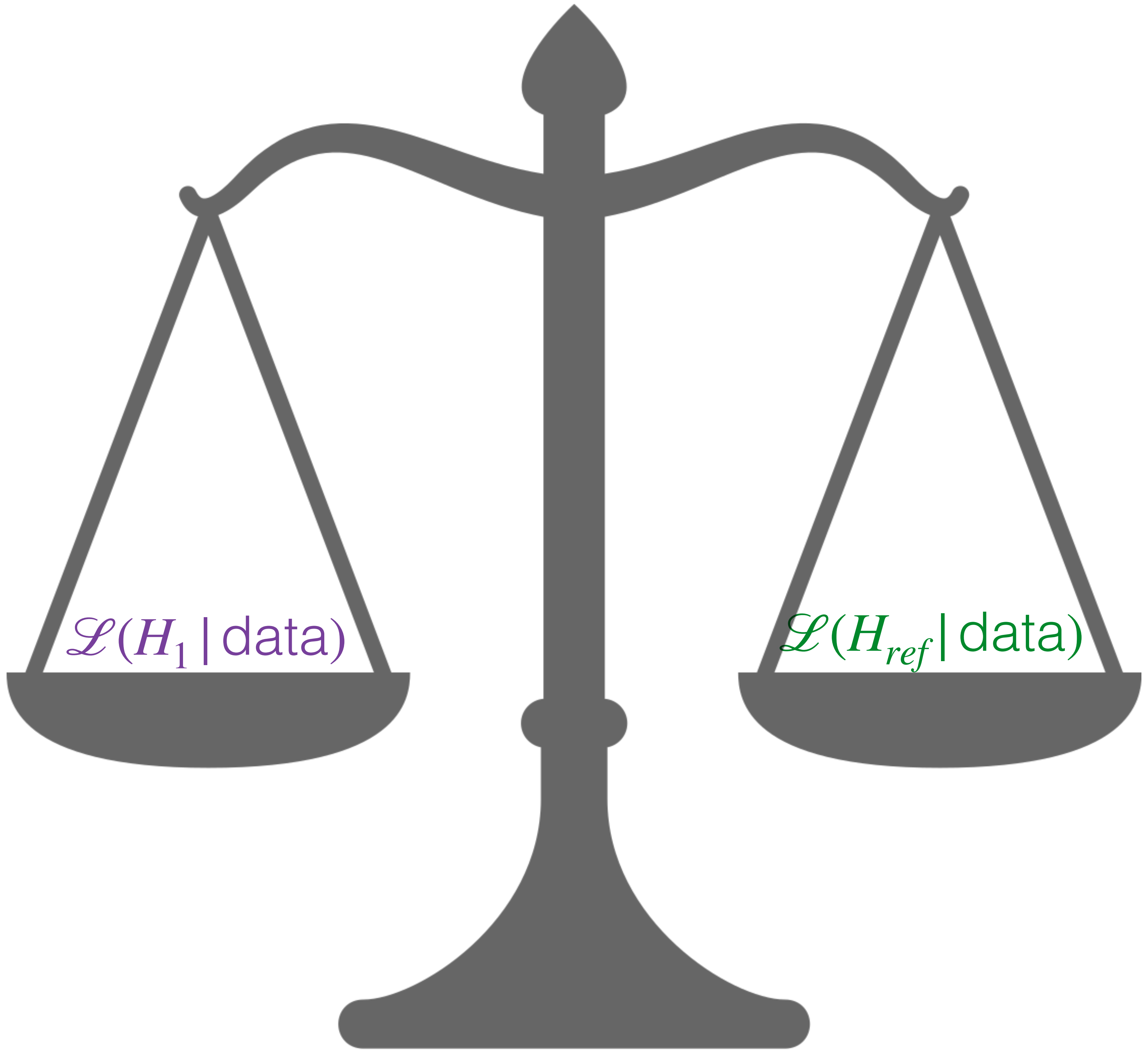


When comparing 2 hypotheses, guaranteed to be optimal test by Neyman-Person lemma

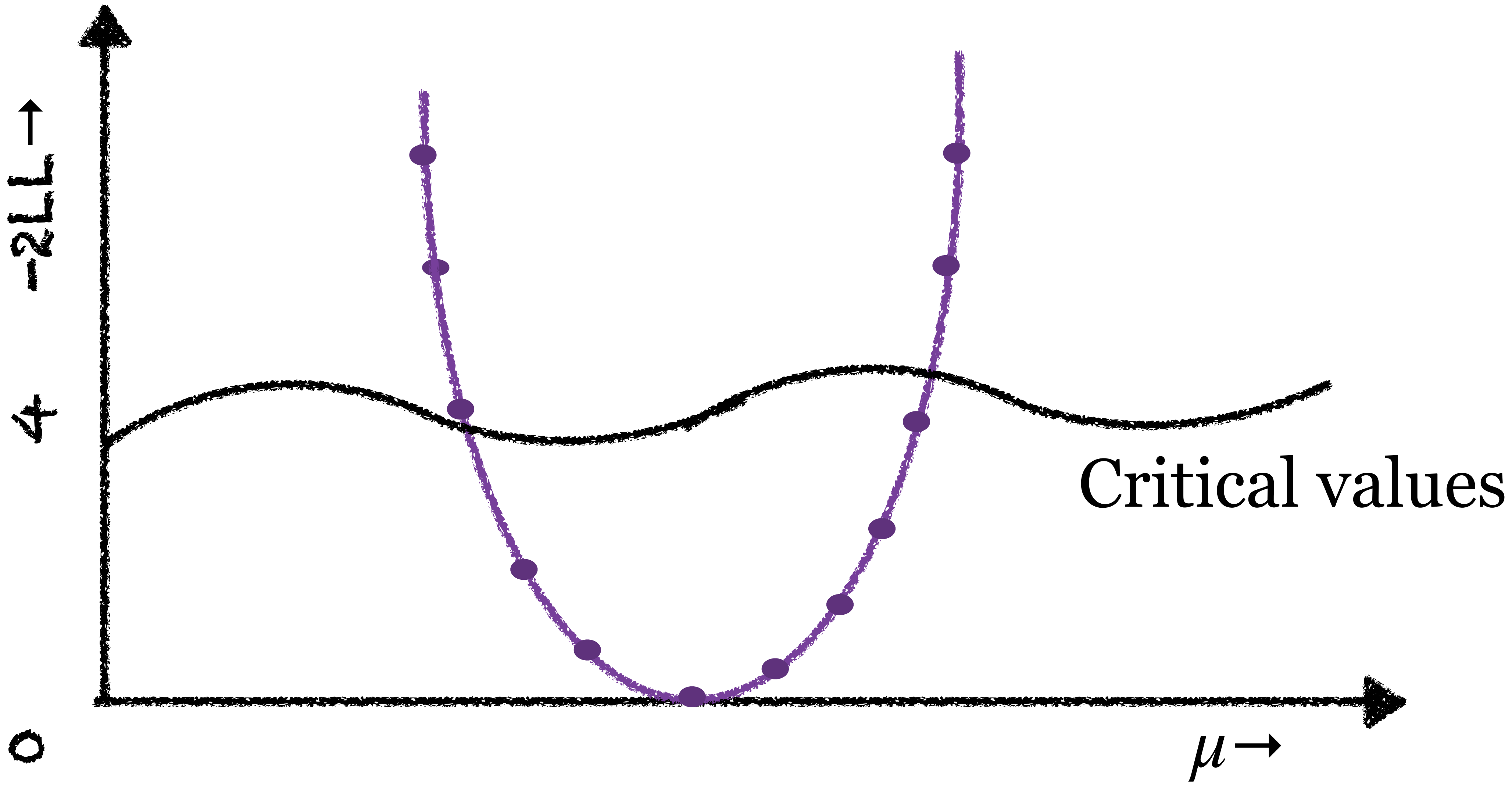


Parameter estimation (infinite hypotheses)

Hypothesis tests

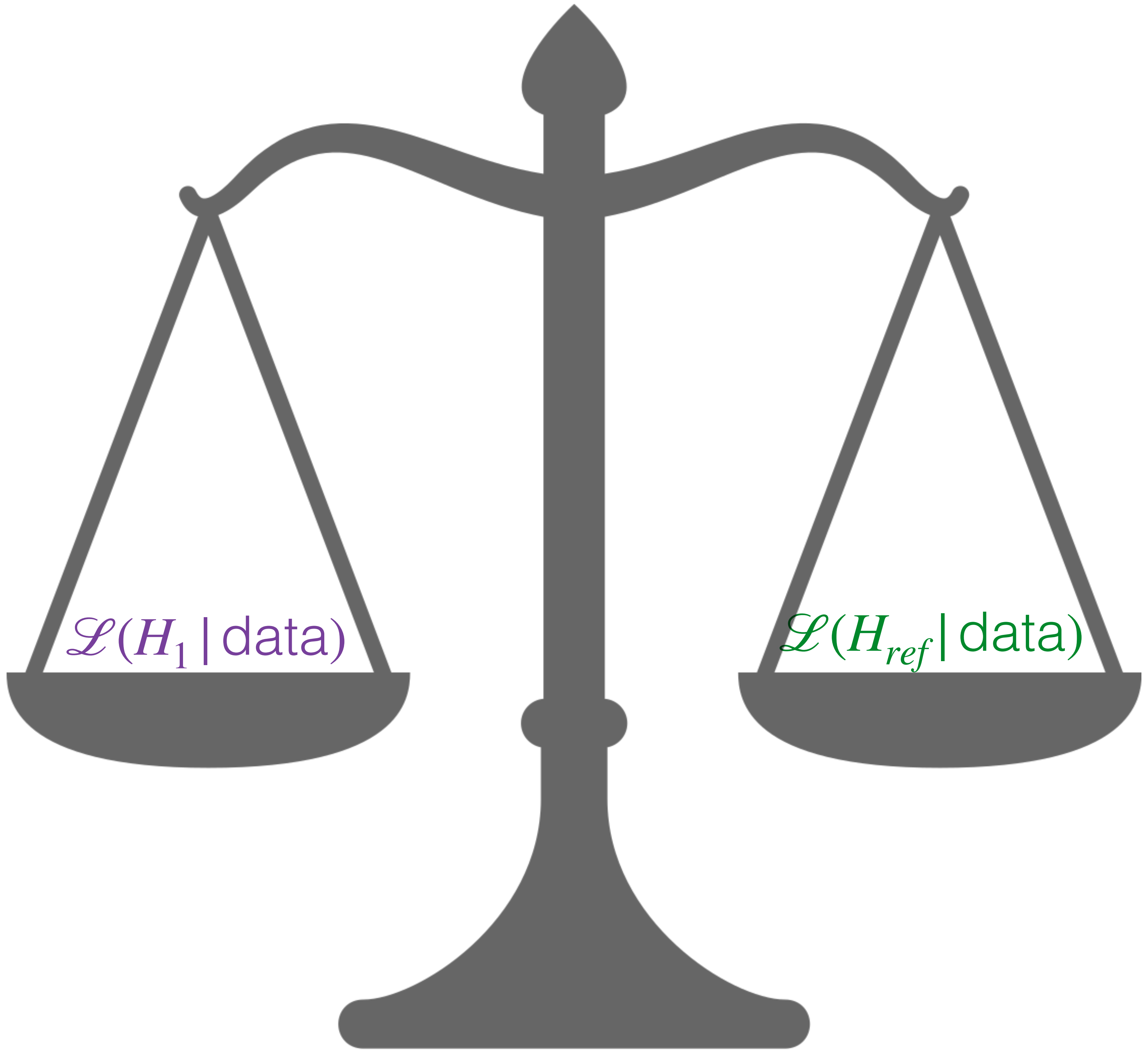


When comparing 2 hypotheses, guaranteed to be optimal test by Neyman-Person lemma

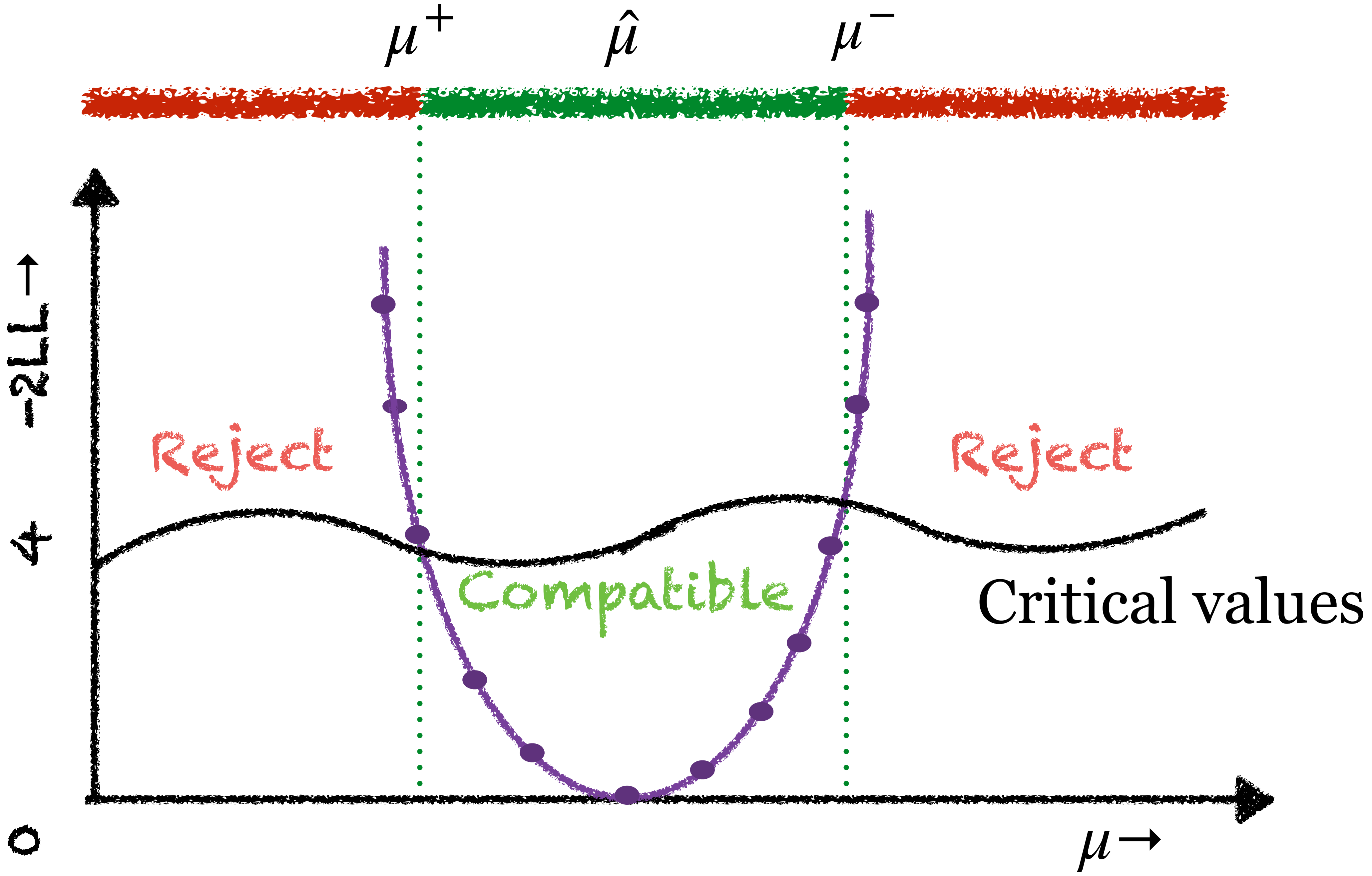


Parameter estimation (infinite hypotheses)

Hypothesis tests

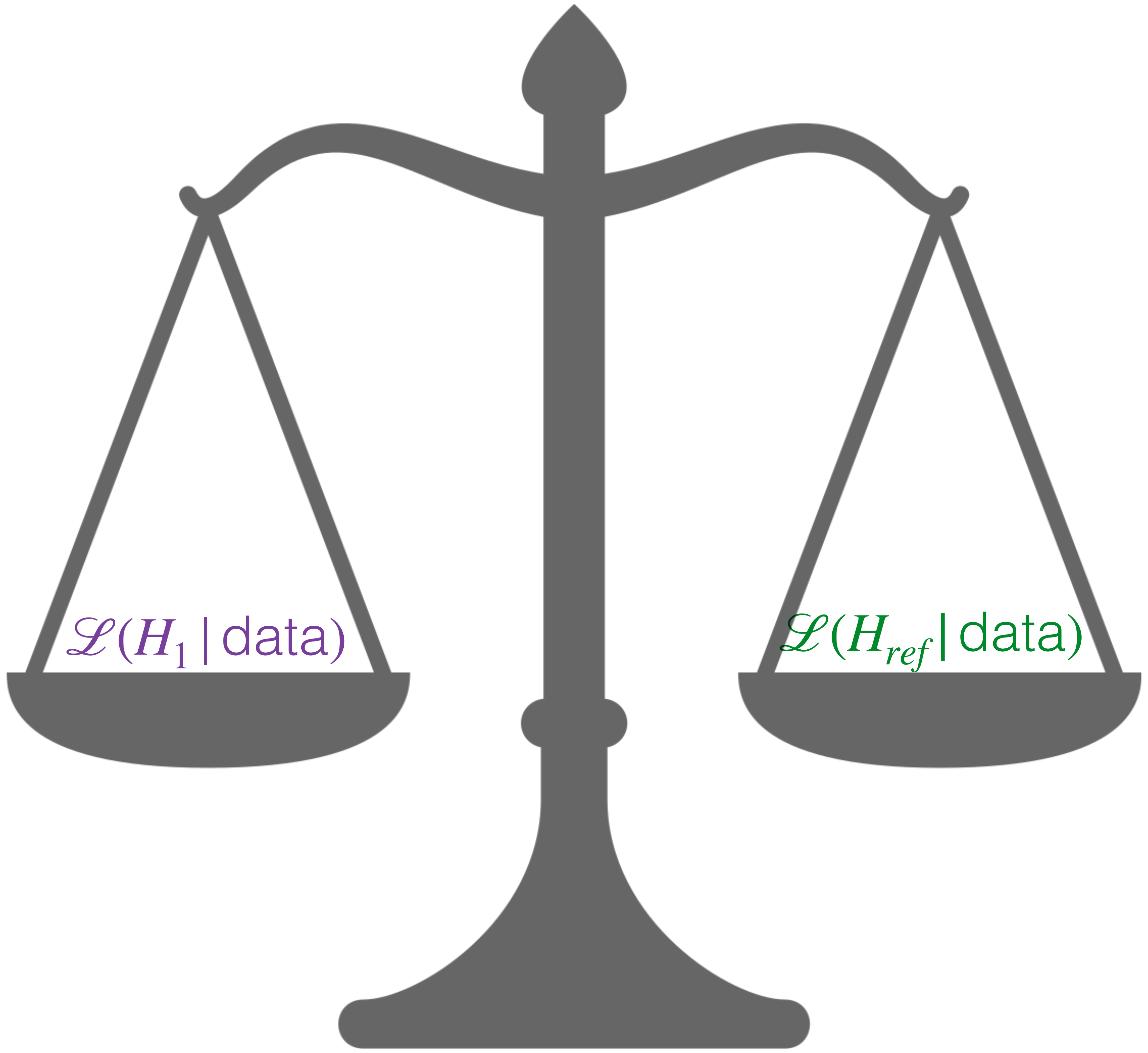


When comparing 2 hypotheses, guaranteed to be optimal test by Neyman-Person lemma



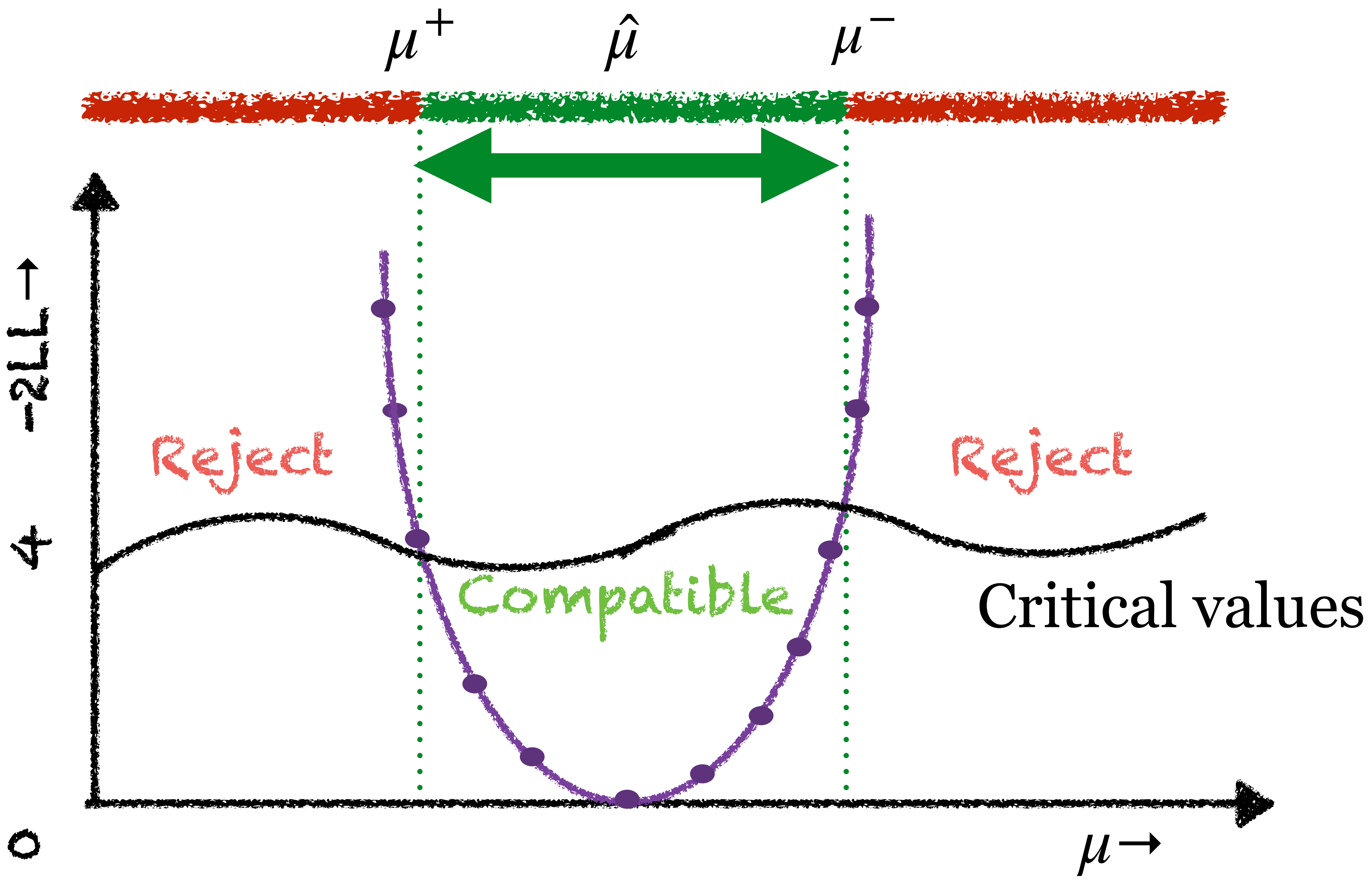
Parameter estimation (infinite hypotheses)

Hypothesis tests



When comparing 2 hypotheses, guaranteed to be optimal test by Neyman-Person lemma

We care about *length* of confidence intervals



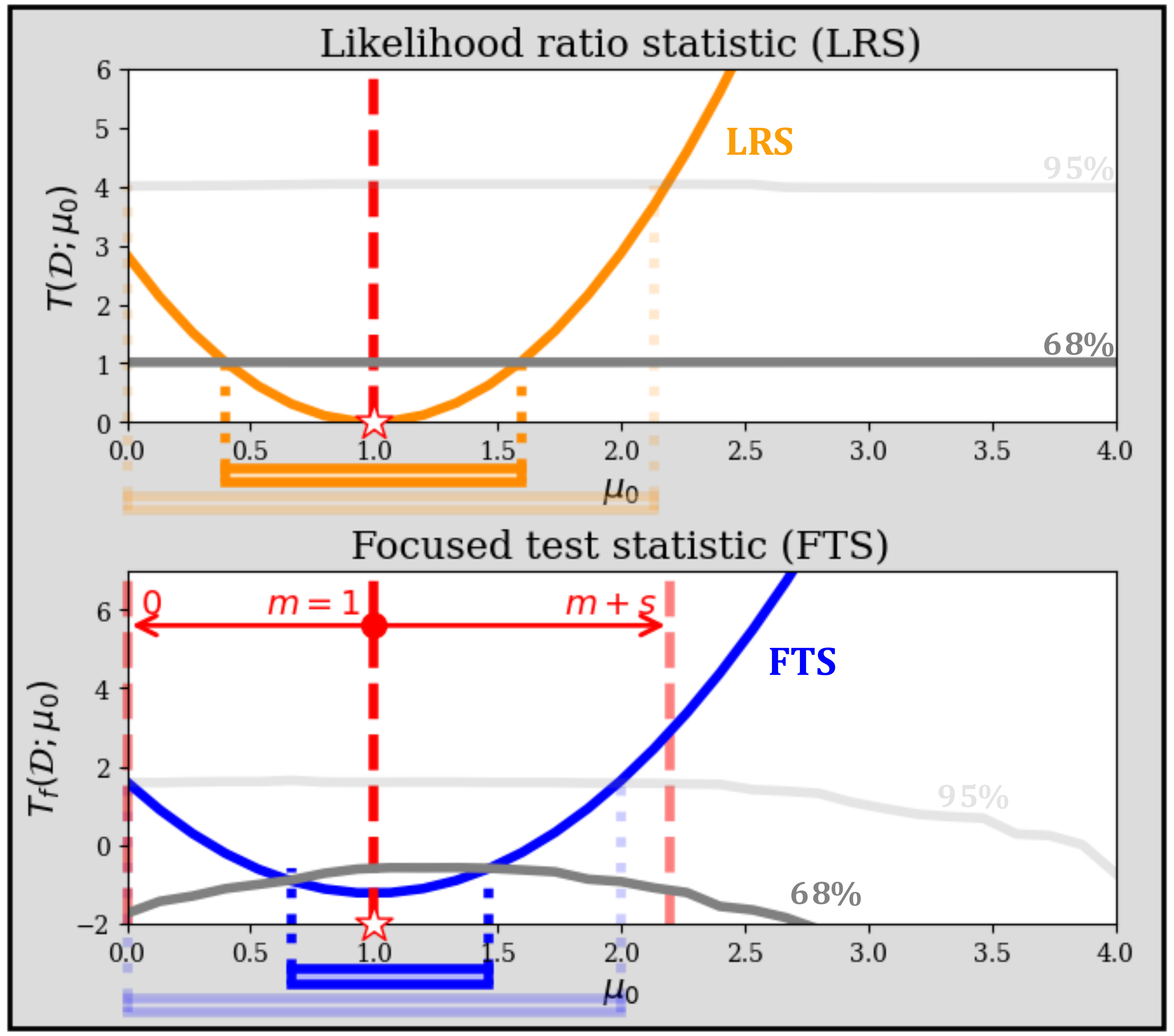
Parameter estimation (infinite hypotheses)

arXiv:2507.17831: Carzon, Ghosh, Izbicki, Lee, et al

You *can* beat the likelihood ratio test itself, with the help of ML

$$LRS(\mathscr{D};\mu_0) = -2\log\left(\frac{p(\mathscr{D}\,|\,\mu_0)}{\sup_{\mu\in\Theta} p(\mathscr{D}\,|\,\mu)}\right) \quad\Longrightarrow\quad FTS(\mathscr{D};\mu_0) = -2\log\left(\frac{p(\mathscr{D}\,|\,\mu_0)}{\int_{\Theta} p(\mathscr{D}\,|\,\mu)f(\mu)d\mu}\right)$$

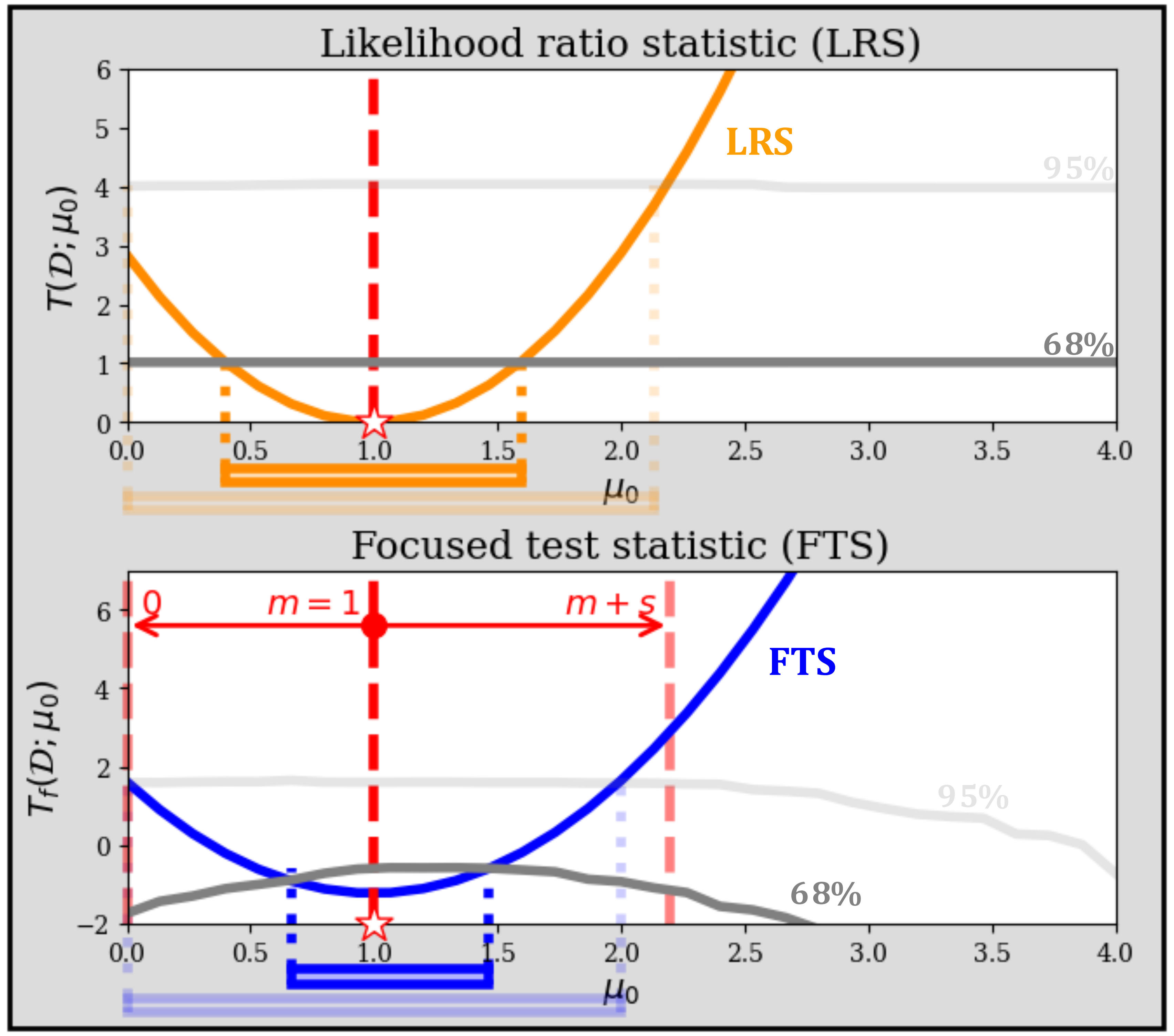
- Denominator in ‘Focused Test Statistic’ (FTS) knows about all alternate hypotheses



You *can* beat the likelihood ratio test itself, with the help of ML

$$LRS(\mathscr{D};\mu_0) = -2\log\left(\frac{p(\mathscr{D}\,|\,\mu_0)}{\sup_{\mu\in\Theta} p(\mathscr{D}\,|\,\mu)}\right) \longrightarrow FTS(\mathscr{D};\mu_0) = -2\log\left(\frac{p(\mathscr{D}\,|\,\mu_0)}{\int_\Theta p(\mathscr{D}\,|\,\mu)f(\mu)d\mu}\right)$$

- Denominator in ‘Focused Test Statistic’ (FTS) knows about all alternate hypotheses



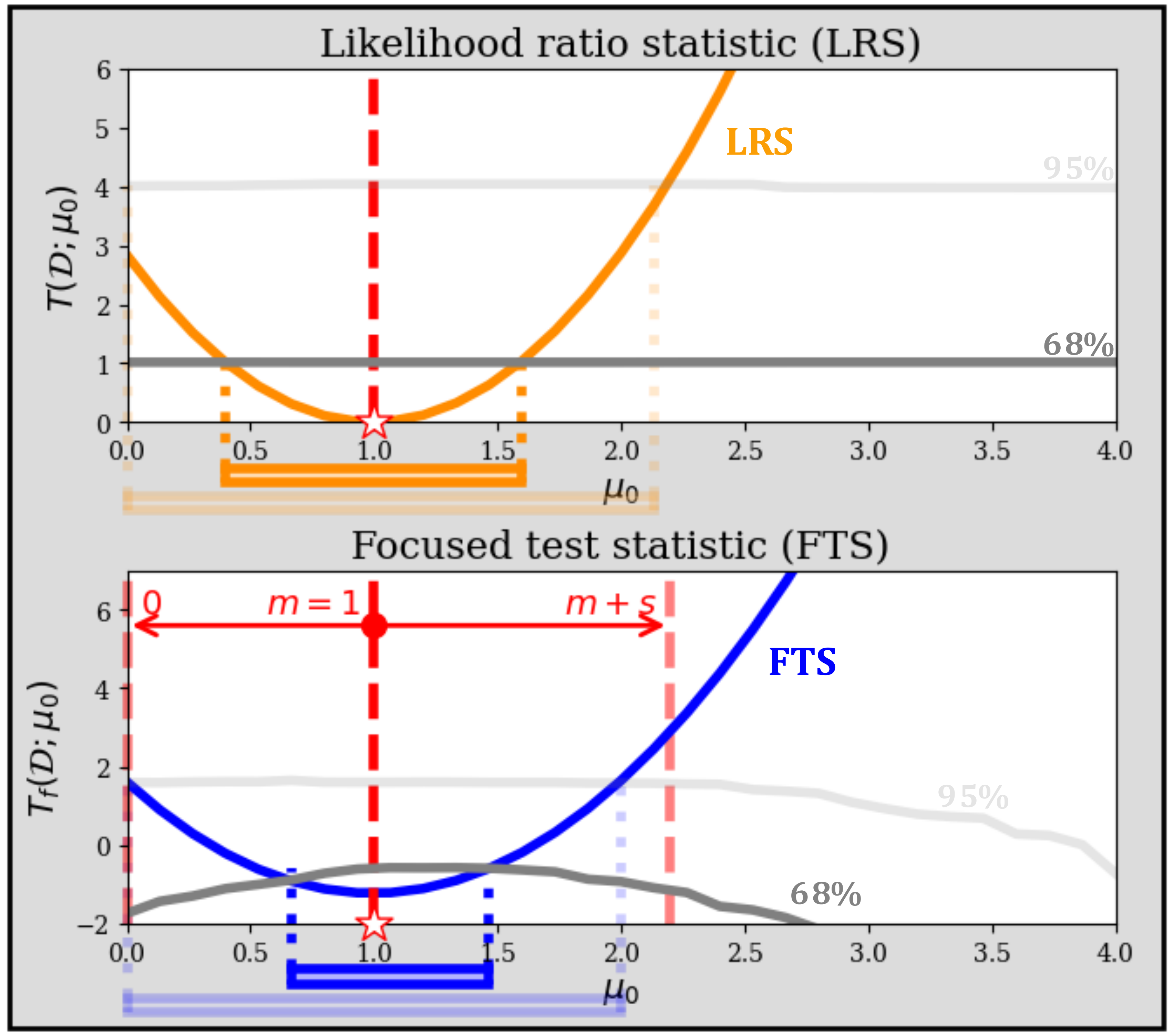
Shorter median length for confidence intervals with FTS

arXiv:2507.17831: Carzon, Ghosh, Izbicki, Lee, et al

You *can* beat the likelihood ratio test itself, with the help of ML

$$LRS(\mathscr{D};\mu_0) = -2\log\left(\frac{p(\mathscr{D}\,|\,\mu_0)}{\sup_{\mu\in\Theta} p(\mathscr{D}\,|\,\mu)}\right) \quad\Longrightarrow\quad FTS(\mathscr{D};\mu_0) = -2\log\left(\frac{p(\mathscr{D}\,|\,\mu_0)}{\int_\Theta p(\mathscr{D}\,|\,\mu)f(\mu)d\mu}\right)$$

- Denominator in ‘Focused Test Statistic’ (FTS) knows about all alternate hypotheses
- $f(\mu)$ focuses statistical power in meaningful regions of parameter space
    - Particularly useful in small sample / small signal regime
- Fast critical value estimation with ML



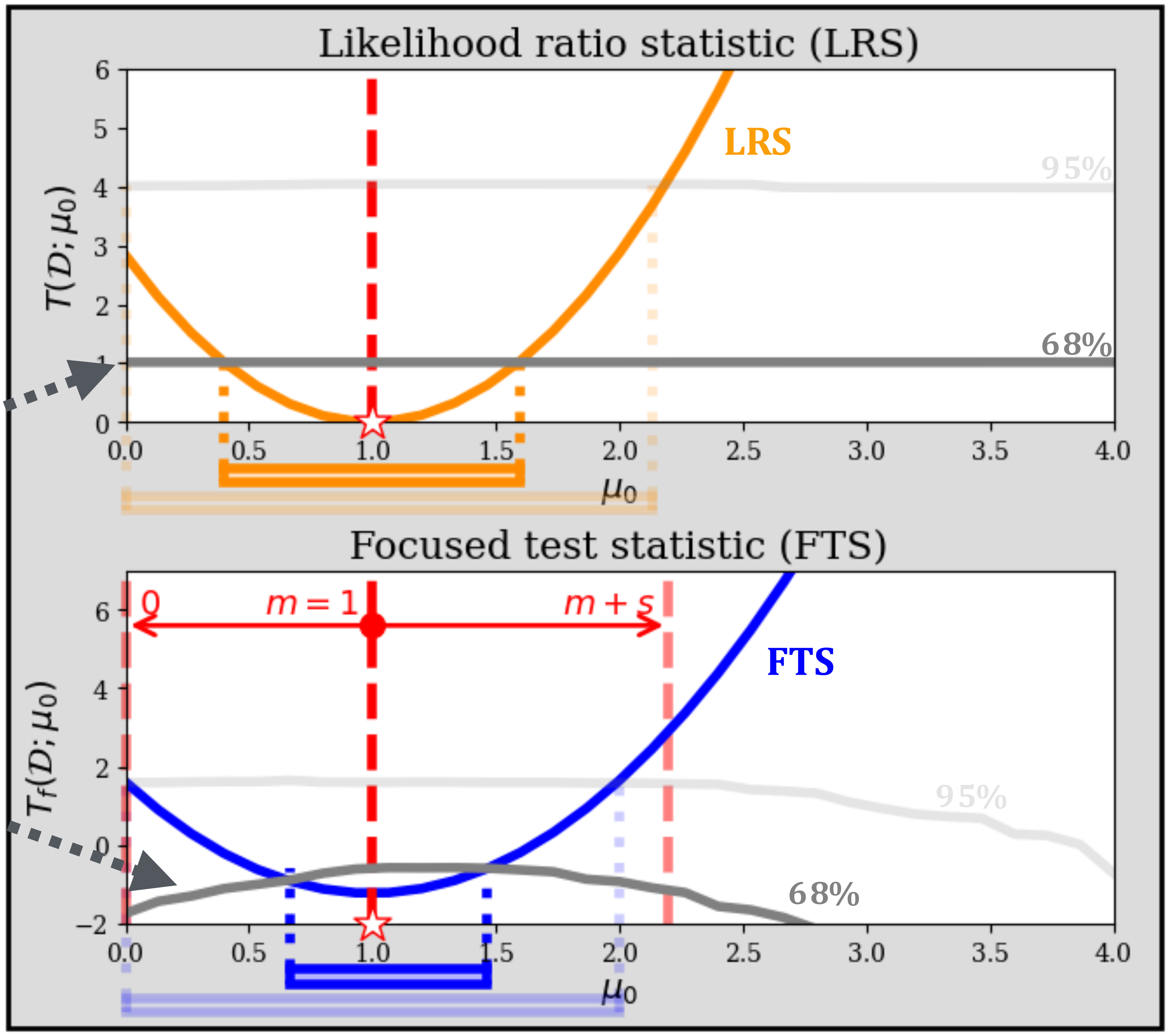
Shorter median length for confidence intervals with FTS

arXiv:2507.17831: Carzon, Ghosh, Izbicki, Lee, et al

You *can* beat the likelihood ratio test itself, with the help of ML

$$LRS(\mathscr{D};\mu_0) = -2\log\left(\frac{p(\mathscr{D}\,|\,\mu_0)}{\sup_{\mu\in\Theta} p(\mathscr{D}\,|\,\mu)}\right) \longrightarrow FTS(\mathscr{D};\mu_0) = -2\log\left(\frac{p(\mathscr{D}\,|\,\mu_0)}{\int_\Theta p(\mathscr{D}\,|\,\mu)f(\mu)d\mu}\right)$$

- Denominator in ‘Focused Test Statistic’ (FTS) knows about all alternate hypotheses
- $f(\mu)$ focuses statistical power in meaningful regions of parameter space
    - Particularly useful in small sample / small signal regime
- Fast critical value estimation with ML



Shorter median length for confidence intervals with FTS

funding gets harder to secure, principal investigators are in their office writing grants while the trainees get to do the cool stuff.

Bryan W. Jones is a retinal neuroscientist at the University of Pittsburgh in Pennsylvania.

AISHIK GHOSH

STUDENTS OVERTURN LONG-HELD ASSUMPTION

I have worked on experimental particle physics since 2015, searching for Higgs bosons at CERN, Europe's particle-physics lab near Geneva, Switzerland, and now also working on the Deep Underground Neutrino Experiment (DUNE) in the United States. For this research, there's one statistical test we've used for decades to confirm the existence of a new particle – the generalized likelihood ratio test (GLRT). This compares two models – a simple null hypothesis, which includes no new particle or matter being discovered, and a more complicated alternative model, which includes a new particle with many possible values of strength.

In December 2024, a couple of PhD students working with my collaborator, Ann Lee, a data scientist at Carnegie Mellon University in Pittsburgh, Pennsylvania, were confident they could disprove the assumption that the GLRT was optimal. In the corner of my mind, I hoped they would prove us wrong. I gave them one of the most famous Higgs boson data sets to play around with. By early 2025, they showed that, although our previous physics results weren't wrong, our use of the GLRT wasn't ideal because it assumed large sample sizes are always generated, which is often not the case. Instead, the test left valuable information on the table. That day was special. I was still sceptical and I went through a battery of checks because I had to go back to my community and defend the PhD students' work, but it was all correct. The paper is currently in review, receiving a great deal of scrutiny.

Together, we produced a statistical test that will drastically improve our ability to make discoveries in particle physics, for example in searches for a new particle such as dark matter, where we expect to see only a few signal events at best. As a scientist, I want deeply held beliefs to be questioned. It was a real shock to the particle-physics community. Young people find it exciting. Senior members are still highly sceptical, as they should be, but they are coming around. As the DUNE experiment comes online, with this new statistical model in place, we hope to make precise measurements about neutrinos much sooner than anticipated.

Aishik Ghosh is a fundamental physicist at the Georgia Institute of Technology in Atlanta.

RAFIK TAREK NEME GARRIDO

SHOCKING CORAL FIND

A couple of years ago, after a day of pouring rain, the water on the Caribbean coast of Colombia was crystal clear and my master's student, Jorge Mareno, managed to take pictures of corals that no one knew existed here. We could find no scientific reports of corals in the area. Typically, the water is pretty turbid because the Magdalena River, which flows from the south of the country to the Caribbean Sea, brings chemicals and pollutants. It's an ongoing ecological and social challenge, but these corals must be adapting to these conditions. We did a sampling campaign across three days with a boat, using environmental DNA to find areas where corals, sponges and fish successfully survive the conditions. Most of the records are completely new for the region. It's super gratifying.

Rafik Tarek Neme Garrido is an evolutionary biologist at the University of the North in Barranquilla, Colombia.

TIM CURRAN

BURN PREDICTIONS

In my group, we test the flammability of plant species using a barbecue. The results can help with fire-mitigation policies and with understanding the evolution of flammability. As part of an outreach activity, we host schoolchildren at the university who haven't had much exposure to academia before. We ask the kids to predict how a particular plant species will behave – for example, what characteristics will make it burn less or more – and then we see who is right. The kids get really into it. They ask amazing questions, the same kind that peer reviewers have asked us, including questioning our methodological assumptions, such as "why do you only blowtorch them for ten seconds?"

Most of the really good days doing science have been associated with young students having a light-bulb moment. In the rather

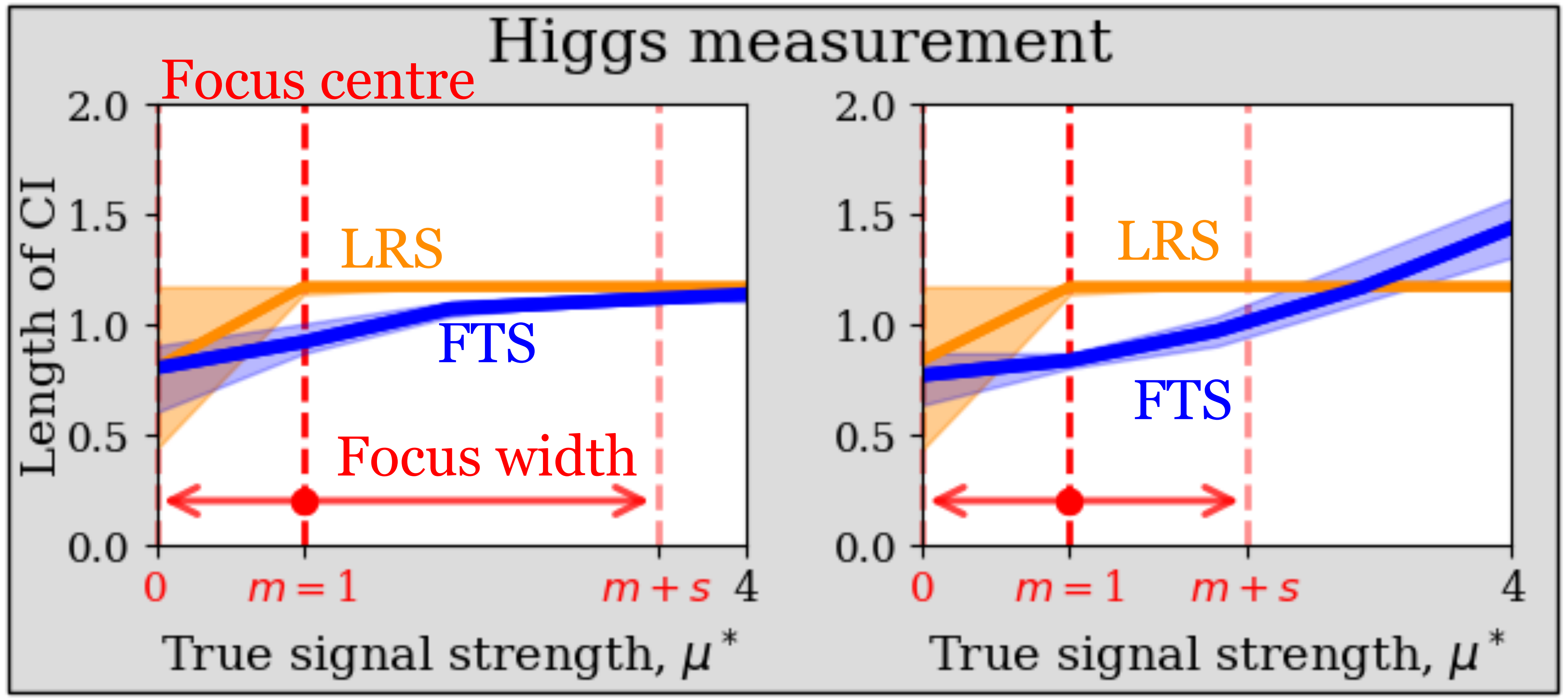
https://www.nature.com/articles/d41586-026-01479-8?utm_source=x&utm_medium=social&utm_campaign=nature&linkId=61827825

Robust to misspecified focus function

$H \rightarrow \tau\tau$ ATLAS simulated public benchmark dataset
Focus is around SM point

Robust to misspecified focus function

← Smaller is better



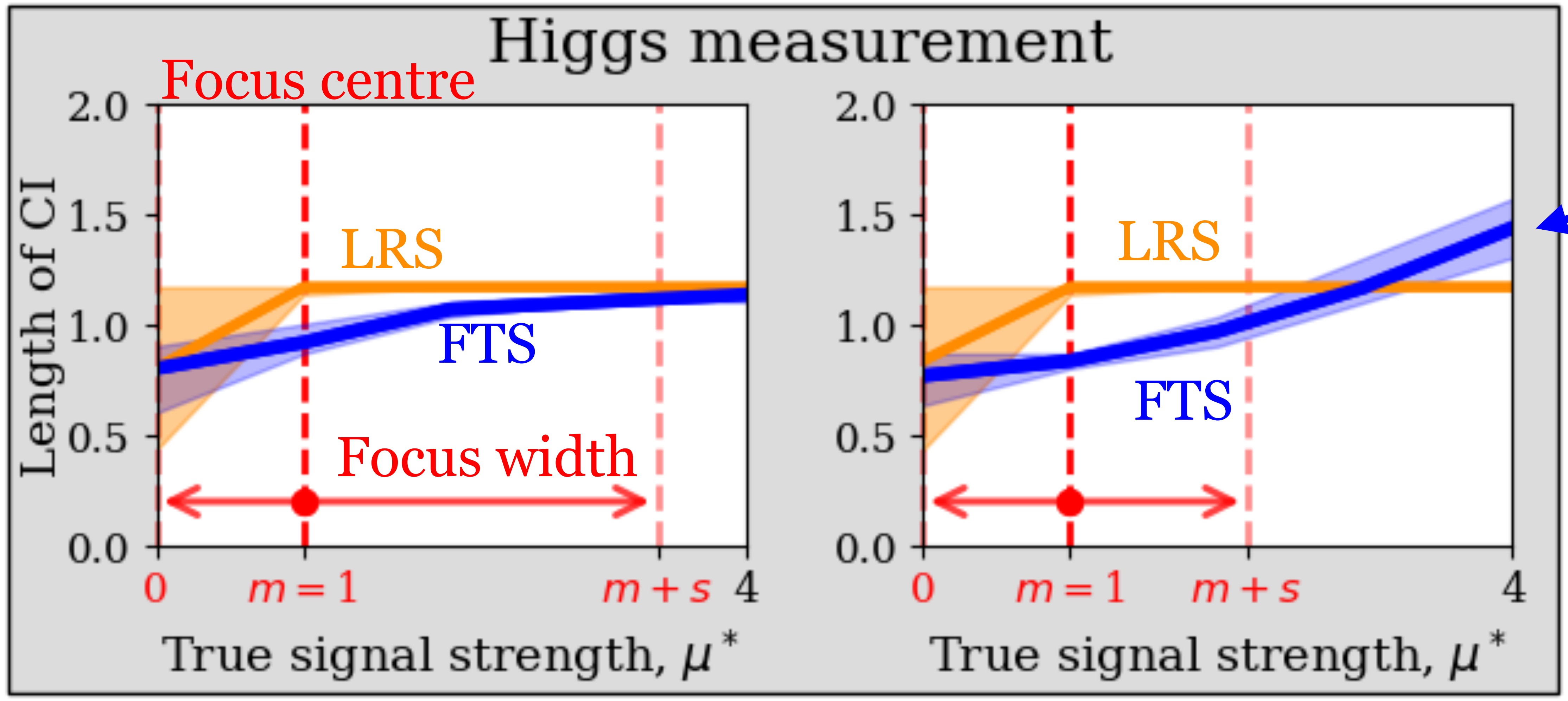
Wide focus $f(\mu)$

Narrow focus $f(\mu)$

$H \rightarrow \tau\tau$ ATLAS simulated public benchmark dataset
Focus is around SM point

Robust to misspecified focus function

← Smaller is better



Wide focus $f(\mu)$

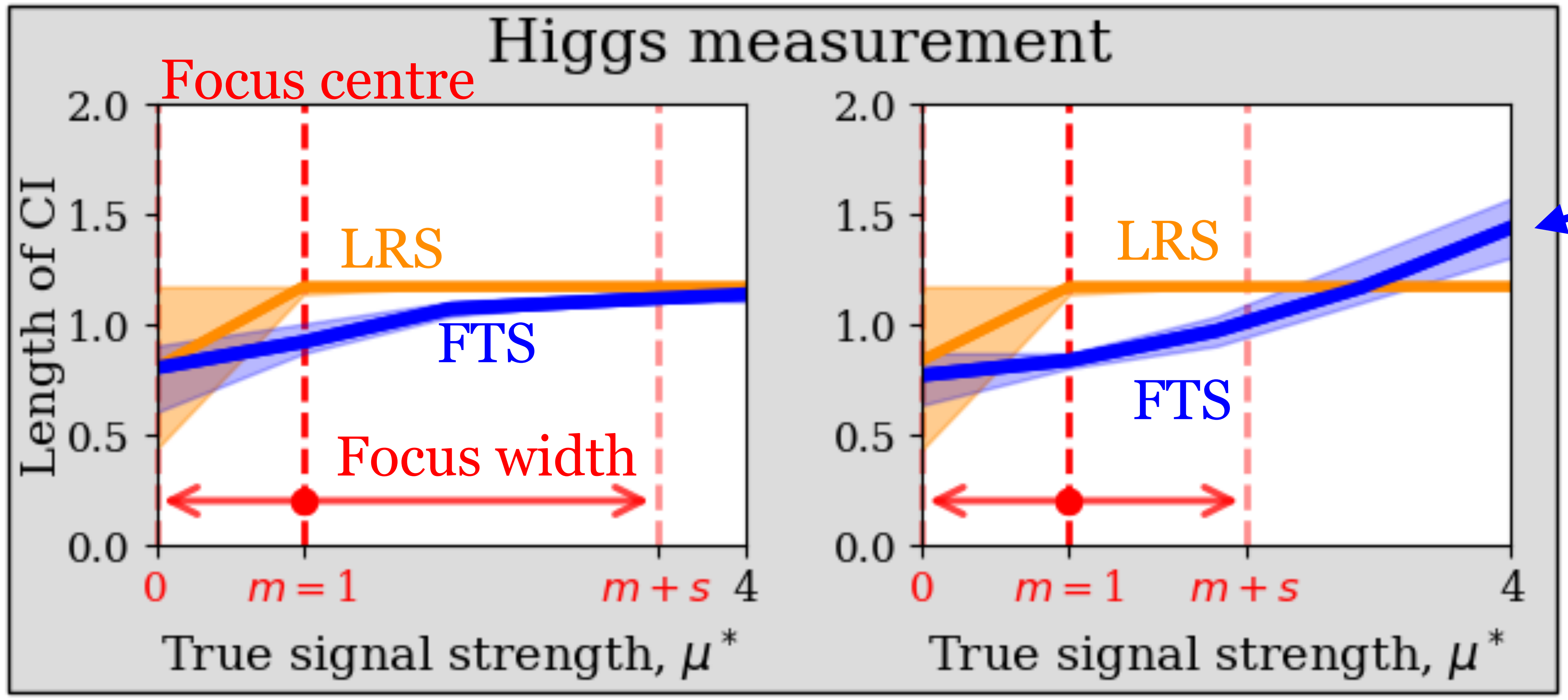
Narrow focus $f(\mu)$

FTS worse than LRS when true μ is very far away from focus region, but coverage of confidence intervals still correct

$H \to \tau\tau$ ATLAS simulated public benchmark dataset
Focus is around SM point

Robust to misspecified focus function

← Smaller is better



Wide focus $f(\mu)$

Narrow focus $f(\mu)$

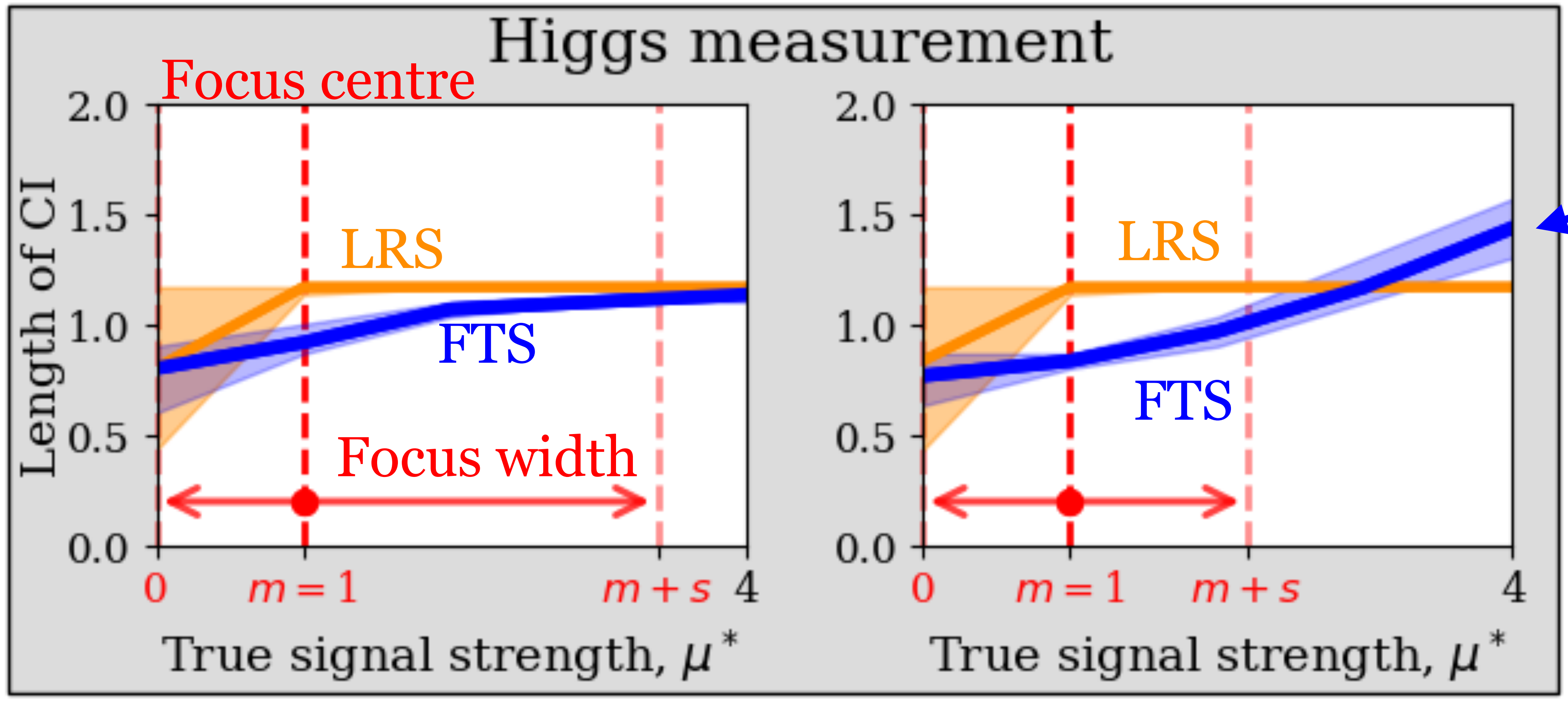
FTS worse than LRS when true μ is very far away from focus region, but coverage of confidence intervals still correct

$H \rightarrow \tau\tau$ ATLAS simulated public benchmark dataset
Focus is around SM point

Even when it is worse, the coverage of the confidence interval is still guaranteed by Neyman construction!

Robust to misspecified focus function

← Smaller is better



Wide focus $f(\mu)$

Narrow focus $f(\mu)$

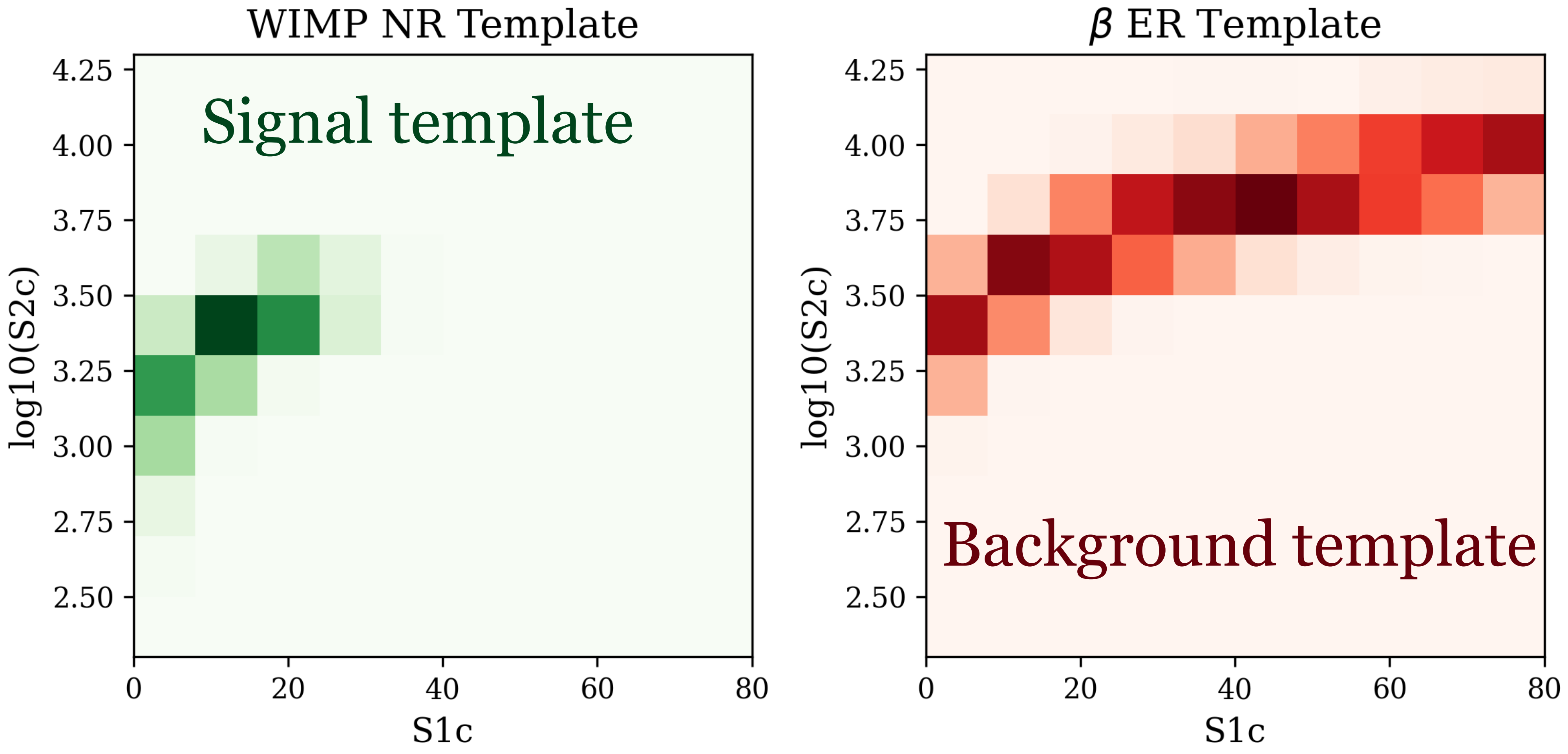
Which focus would you choose?

FTS worse than LRS when true μ is very far away from focus region, but coverage of confidence intervals still correct

$H \to \tau\tau$ ATLAS simulated public benchmark dataset
Focus is around SM point

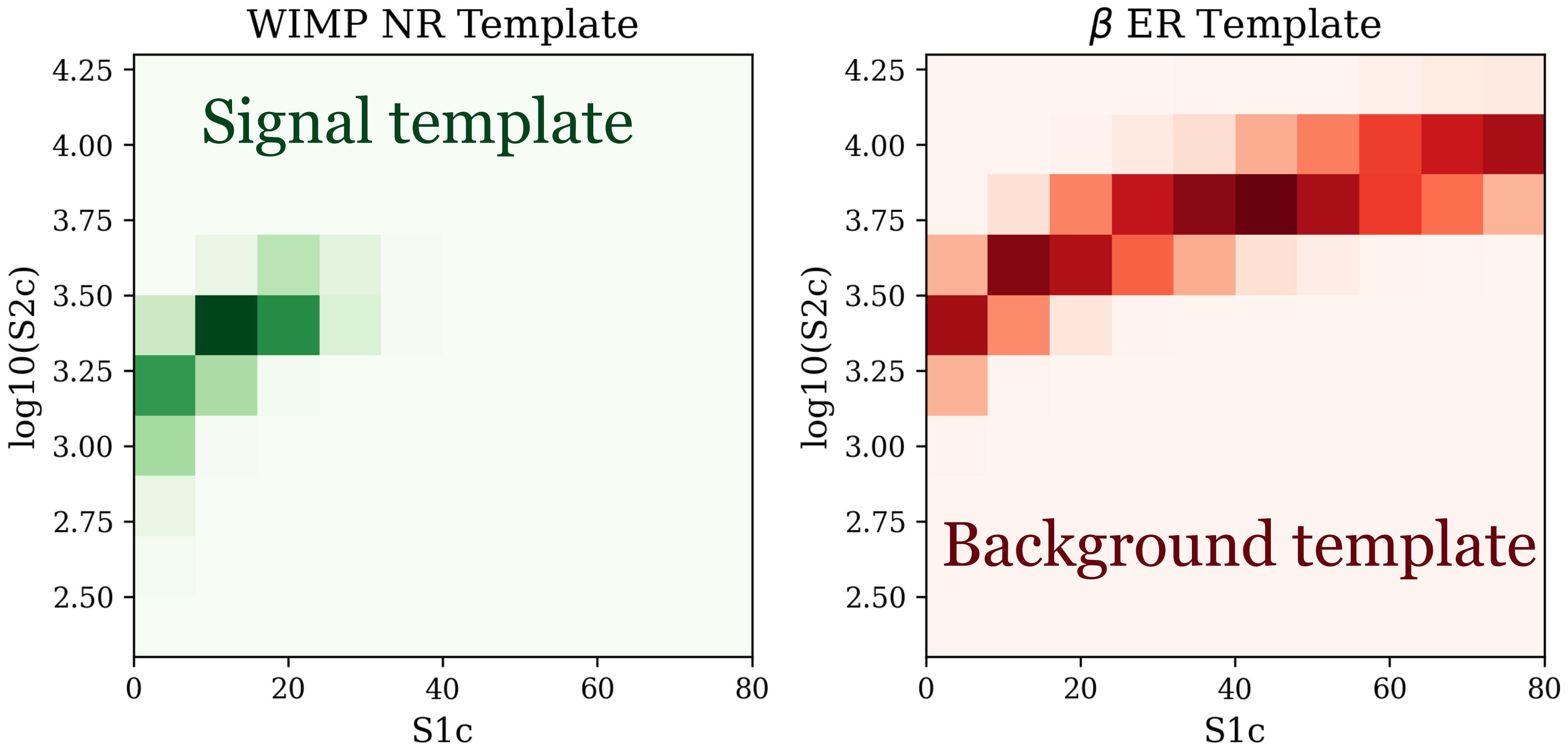
Even when it is worse, the coverage of the confidence interval is still guaranteed by Neyman construction!

Second case study: Dark matter search in LZ experiment



Simulated to mimic LZ data

Second case study: Dark matter search in LZ experiment

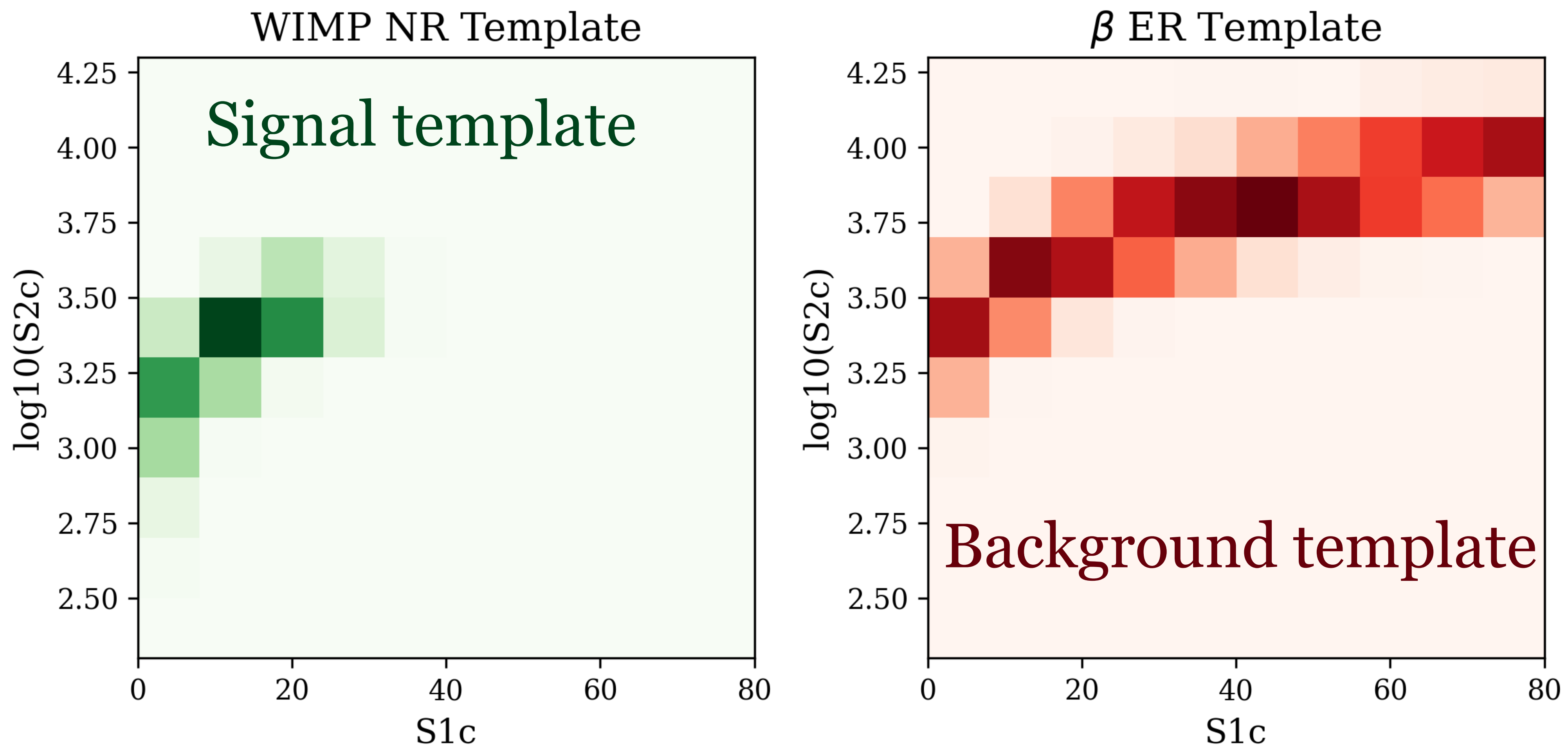


Simulated to mimic LZ data

Focus now placed at $\mu = 0$, we want to set upper limits on DM signal

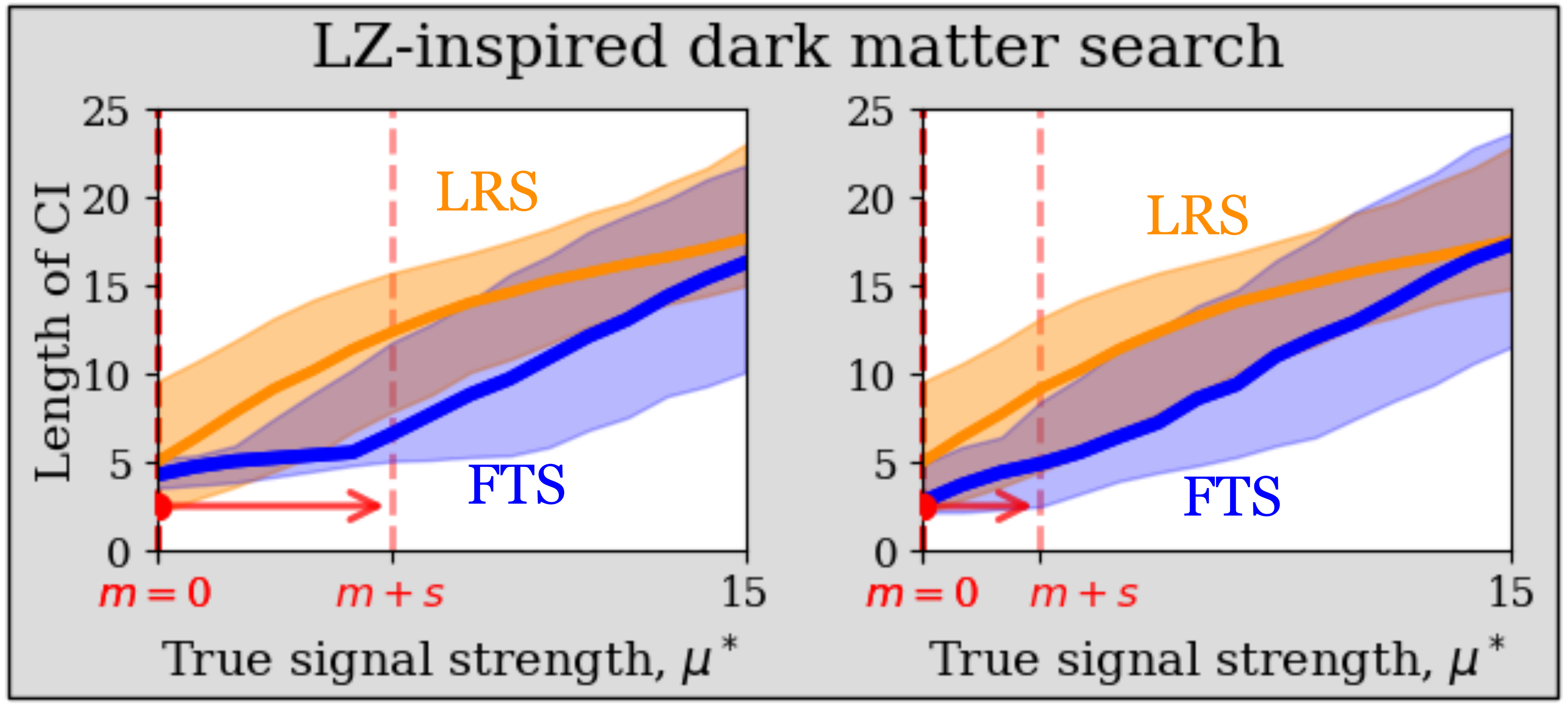
arXiv:2507.17831: Carzon, Ghosh, Izbicki, Lee, et al

Second case study: Dark matter search in LZ experiment



Simulated to mimic LZ data

Focus now placed at $\mu = 0$, we want to set upper limits on DM signal

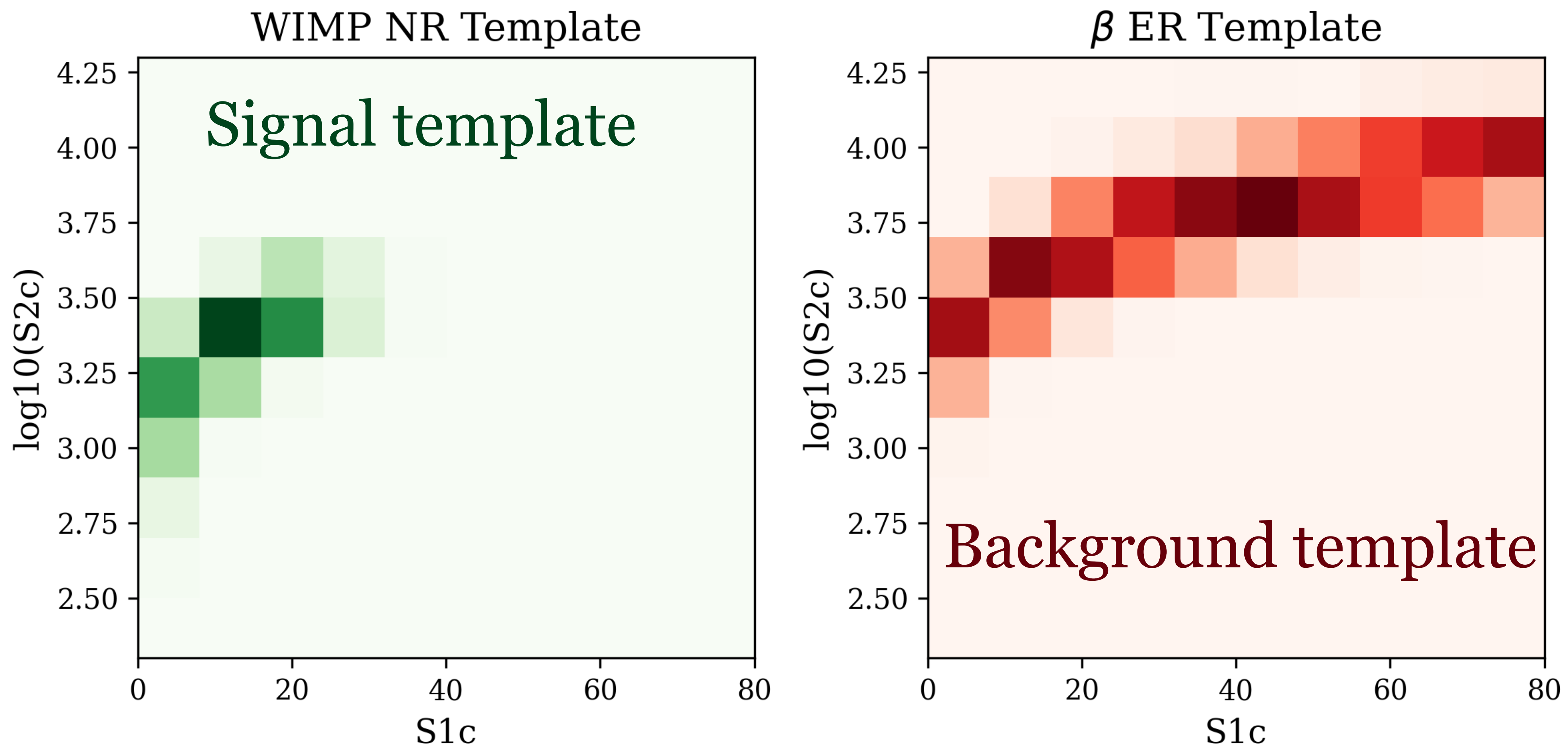


Wide focus

Narrow focus

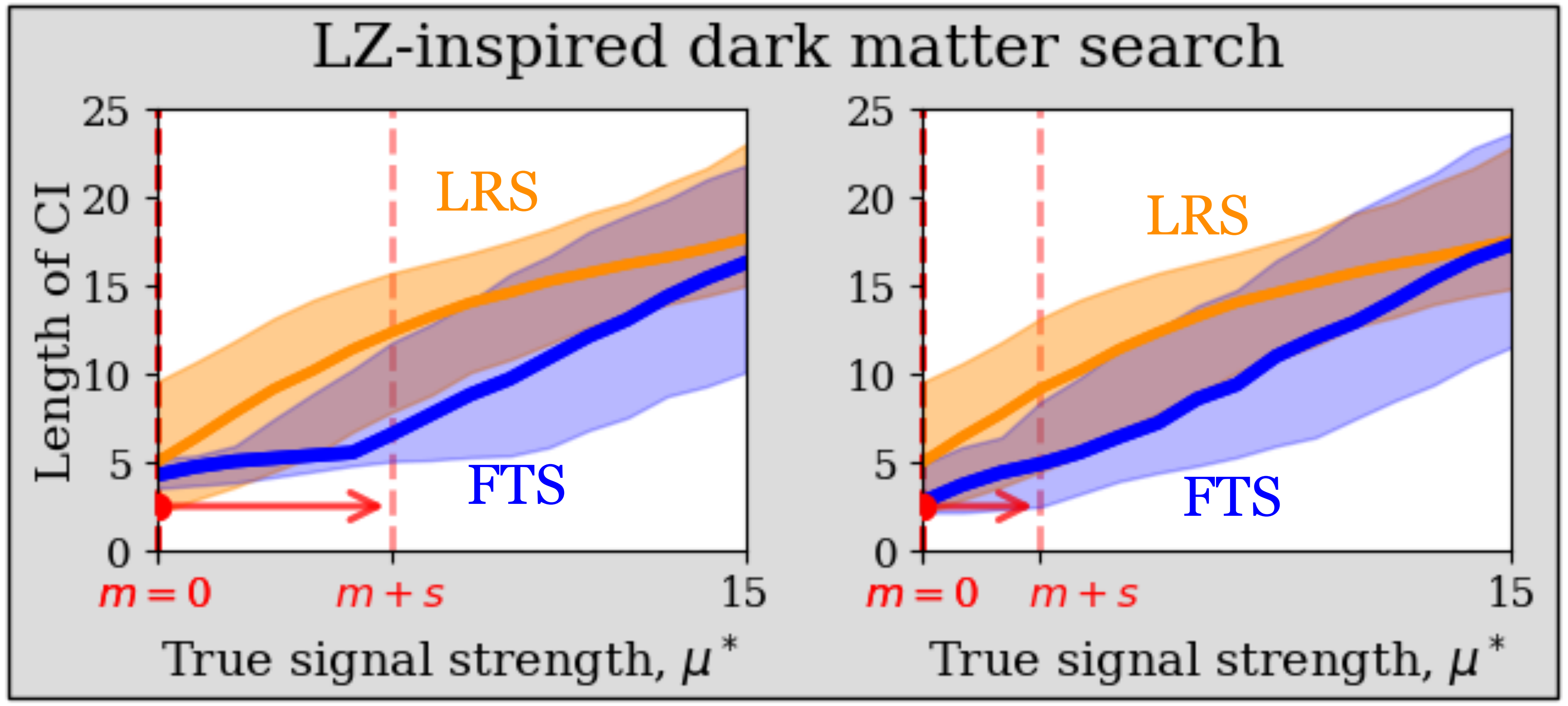
arXiv:2507.17831: Carzon, Ghosh, Izbicki, Lee, et al

Second case study: Dark matter search in LZ experiment



Simulated to mimic LZ data

Focus now placed at $\mu = 0$, we want to set upper limits on DM signal



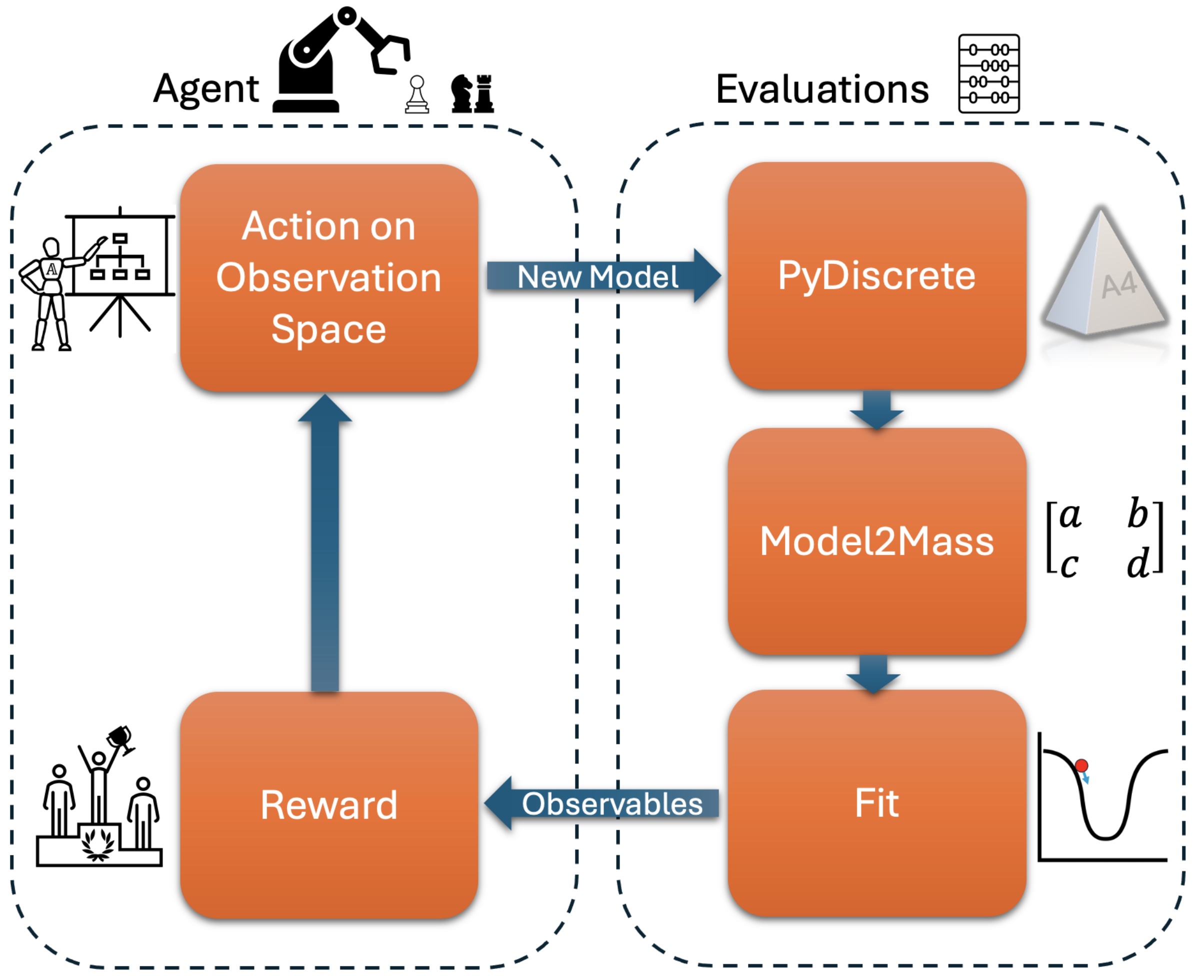
Wide focus

Narrow focus

FTS gains stable over large parameter space for both focus functions

More precision even when there is actually a signal in the observed data ($\mu > 0$)

Conclusion



Machine learning helps us do what was considered impossible, from theory design to extracting physics from tails of our distributions

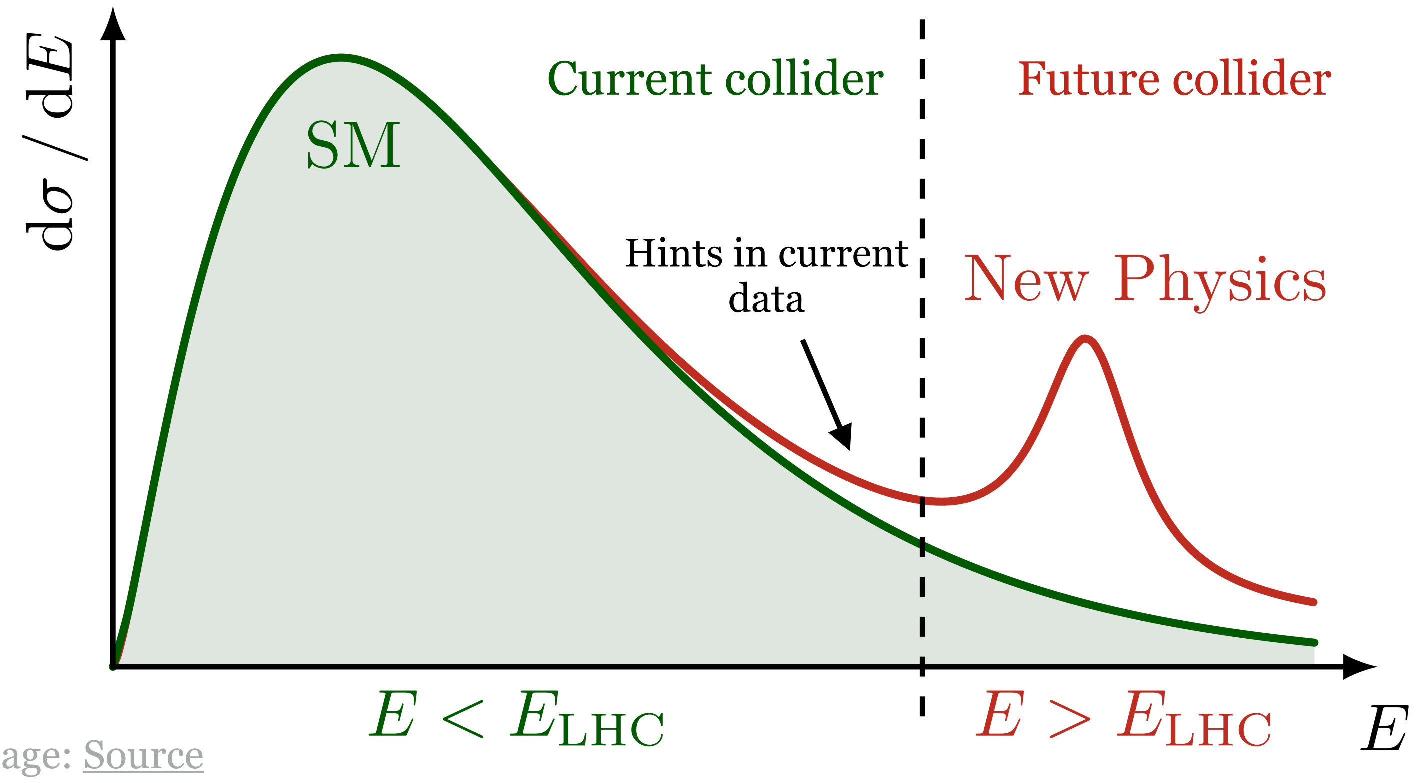


Image: Source

Thank you !

arXiv:2507.17831: Carzon, Ghosh, Izbicki, Lee, et al

FTS sets better lower bounds even with a focus centred at $\mu = 0$

arXiv:2507.17831: Carzon, Ghosh, Izbicki, Lee, et al

FTS sets better lower bounds even with a focus centred at $\mu = 0$

Consistently larger lower bounds if signal exits, smaller upper bounds when it doesn't

Lower bounds when there is a signal:

μ^*	LRS	FTS, wide	FTS, narrow
10.0	0.0 (0.0)	**2.08** (0.0)	1.60 (0.0)

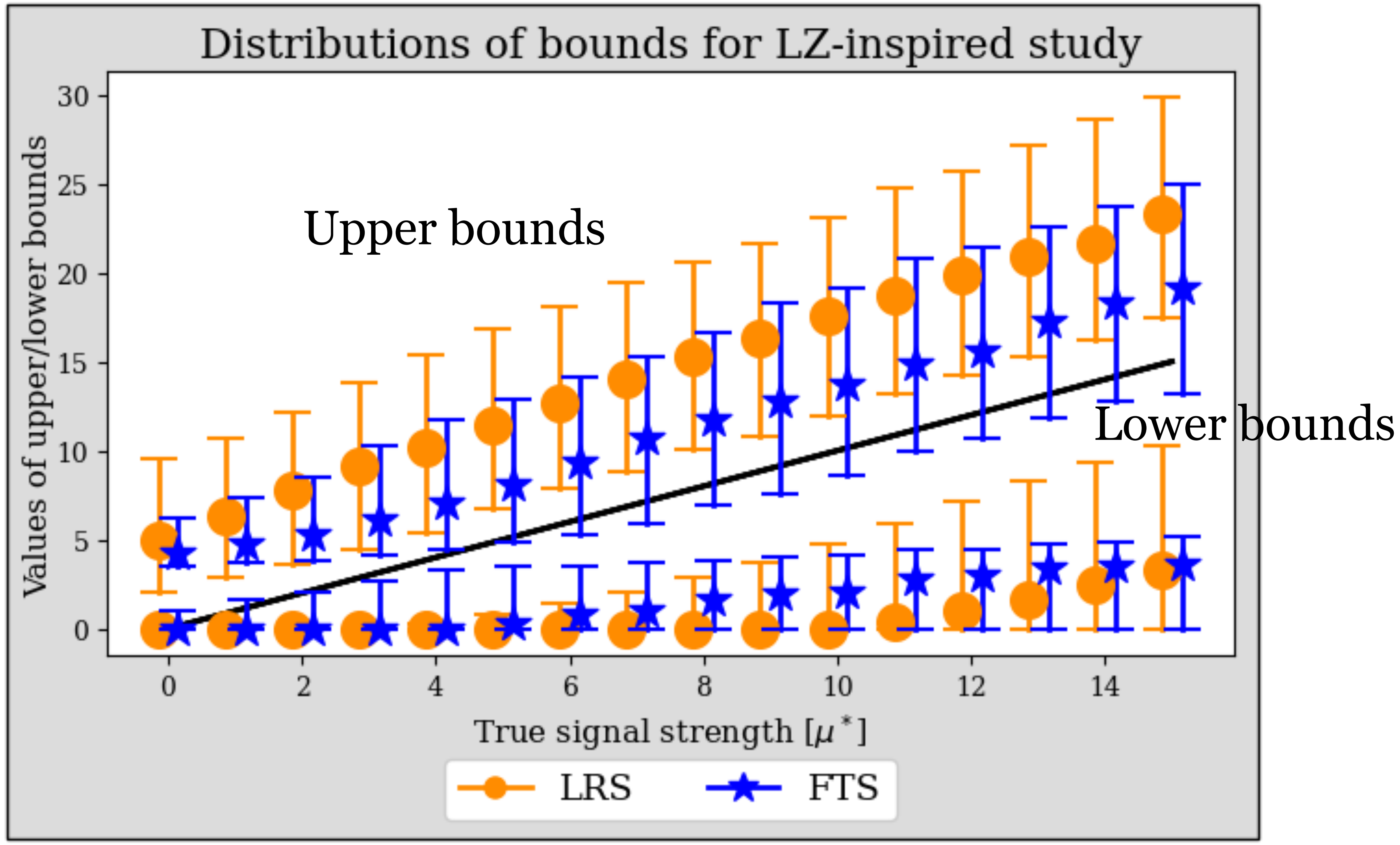
Upper bounds when there is no signal:

μ^*	LRS	FTS, wide	FTS, narrow
0.0	4.97 (12.59)	4.25 (**10.98**)	**2.73** (11.14)

Numbers represent length of confidence intervals for 1σ (2σ)

arXiv:2507.17831: Carzon, Ghosh, Izbicki, Lee, et al

FTS sets better lower bounds even with a focus centred at $\mu = 0$



Consistently larger lower bounds if signal exits, smaller upper bounds when it doesn't

Lower bounds when there is a signal:

μ^*	LRS	FTS, wide	FTS, narrow
10.0	0.0 (0.0)	**2.08** (0.0)	1.60 (0.0)

Upper bounds when there is no signal:

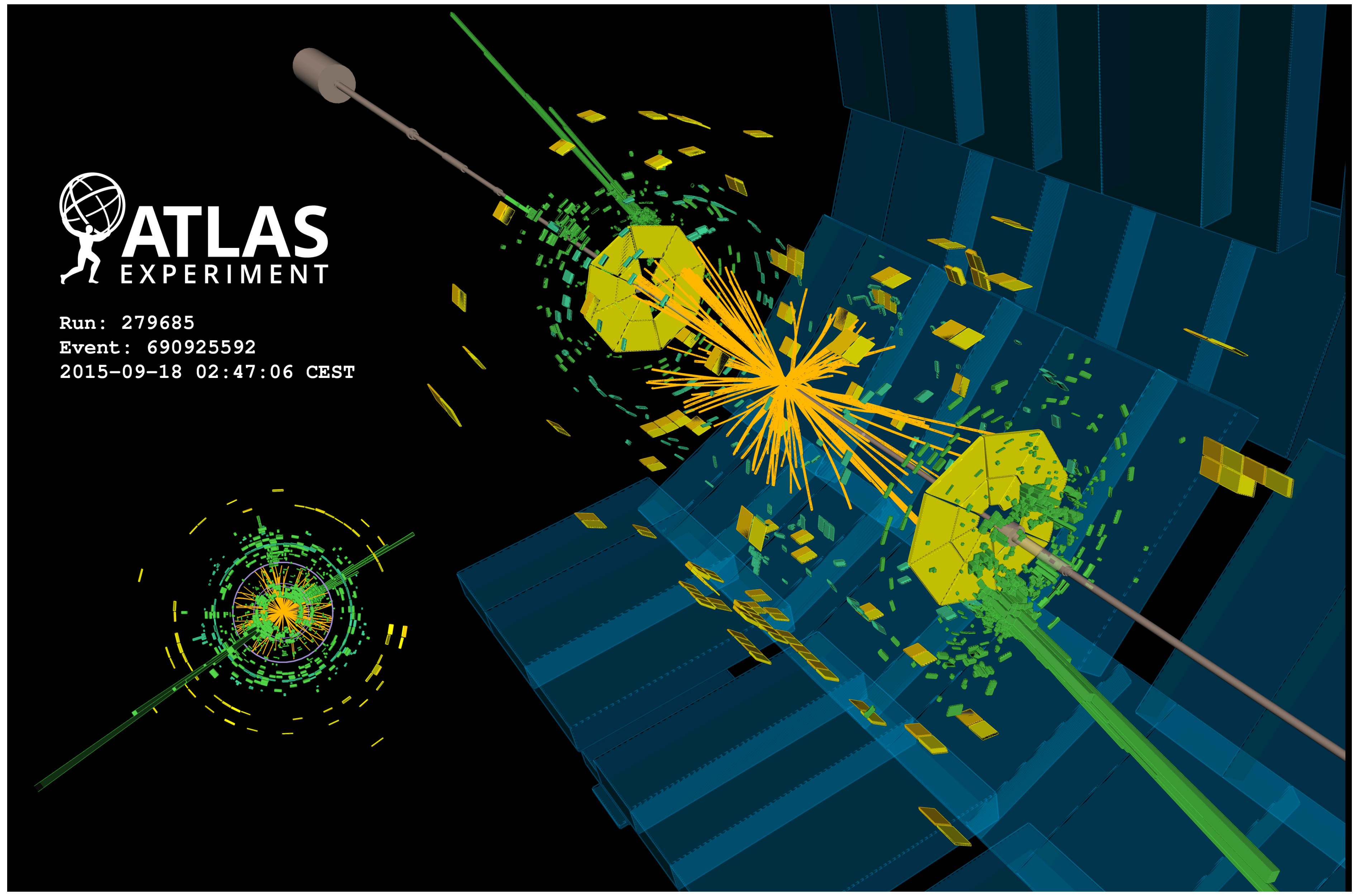
μ^*	LRS	FTS, wide	FTS, narrow
0.0	4.97 (12.59)	4.25 (**10.98**)	**2.73** (11.14)

Numbers represent length of confidence intervals for 1σ (2σ)

Traditional Approach: Design one sensitive observable

- Detector has O(100 million) sensors

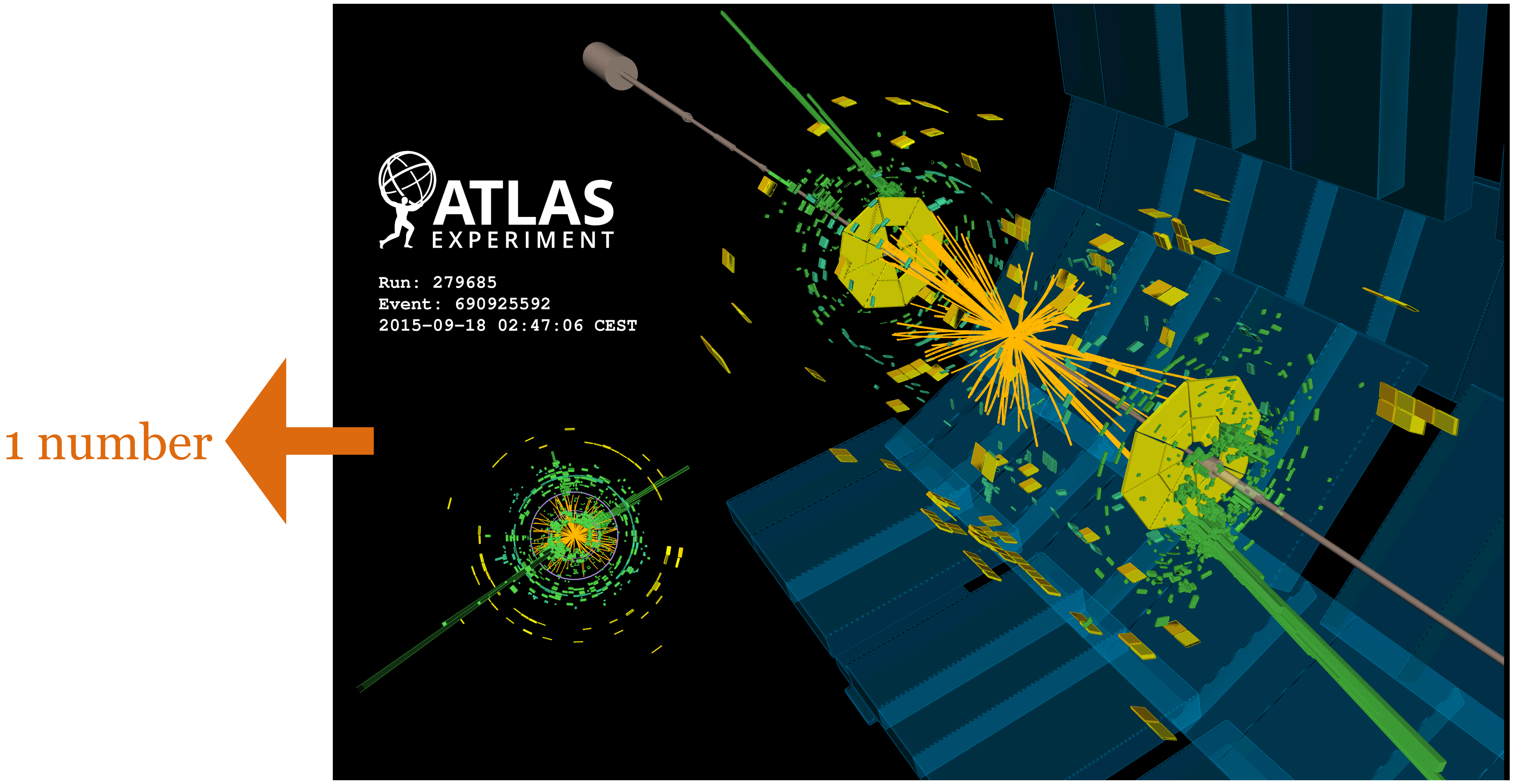
- Reconstruction pipeline, event selection
- Design one summary variable
    - Compression: O(100 million) → 1
- Build a 1-D histogram
- Calculate likelihoods with histograms



Traditional Approach: Design one sensitive observable

- Detector has O(100 million) sensors

- Reconstruction pipeline, event selection
- Design one summary variable
    - Compression: O(100 million) → 1
- Build a 1-D histogram
- Calculate likelihoods with histograms



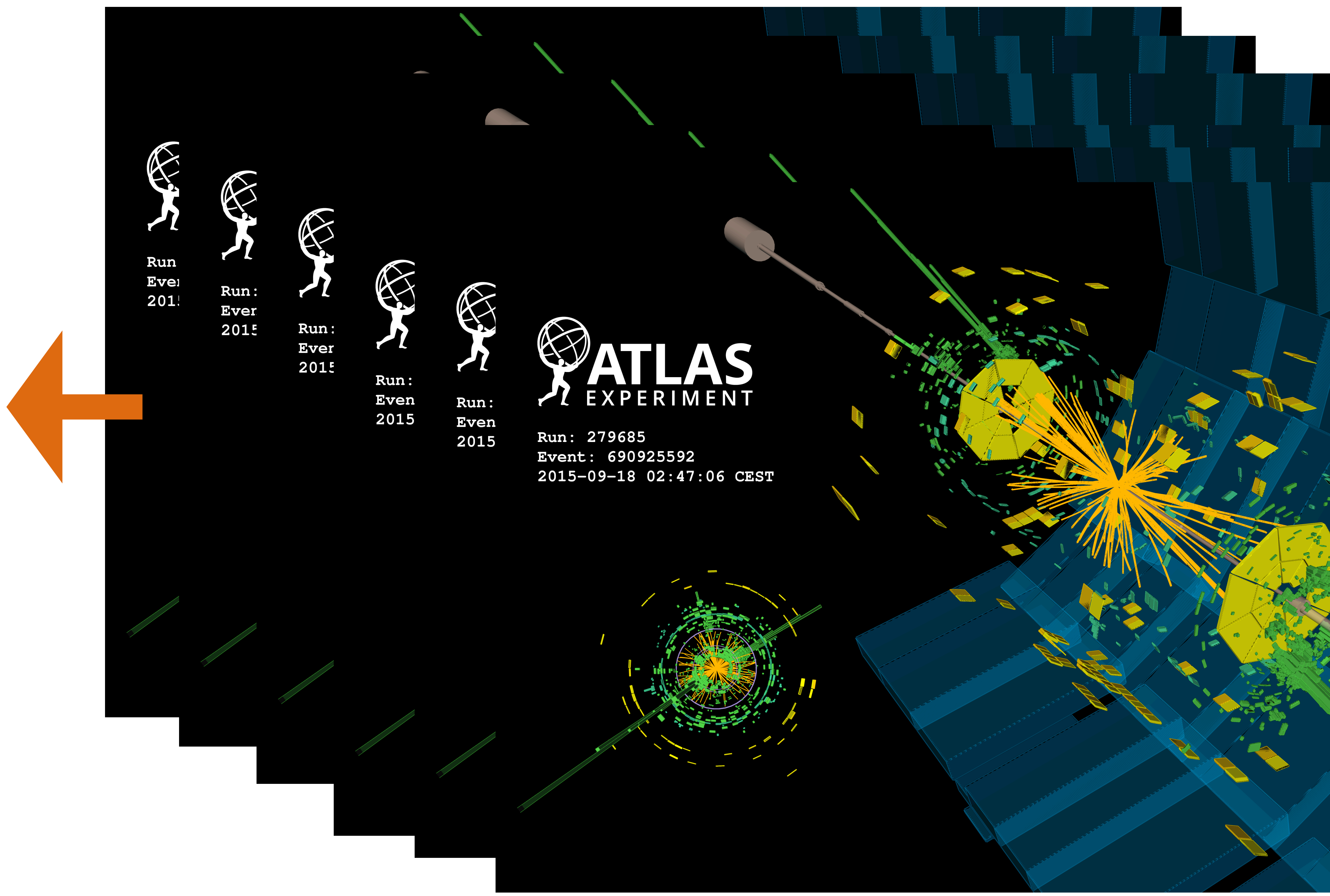
Traditional Approach: Design one sensitive observable

- Detector has O(100 million) sensors

- Reconstruction pipeline, event selection
- Design one summary variable
  - Compression: O(100 million) → 1
- Build a 1-D histogram
- Calculate likelihoods with histograms

Thousands of collision events

1 number

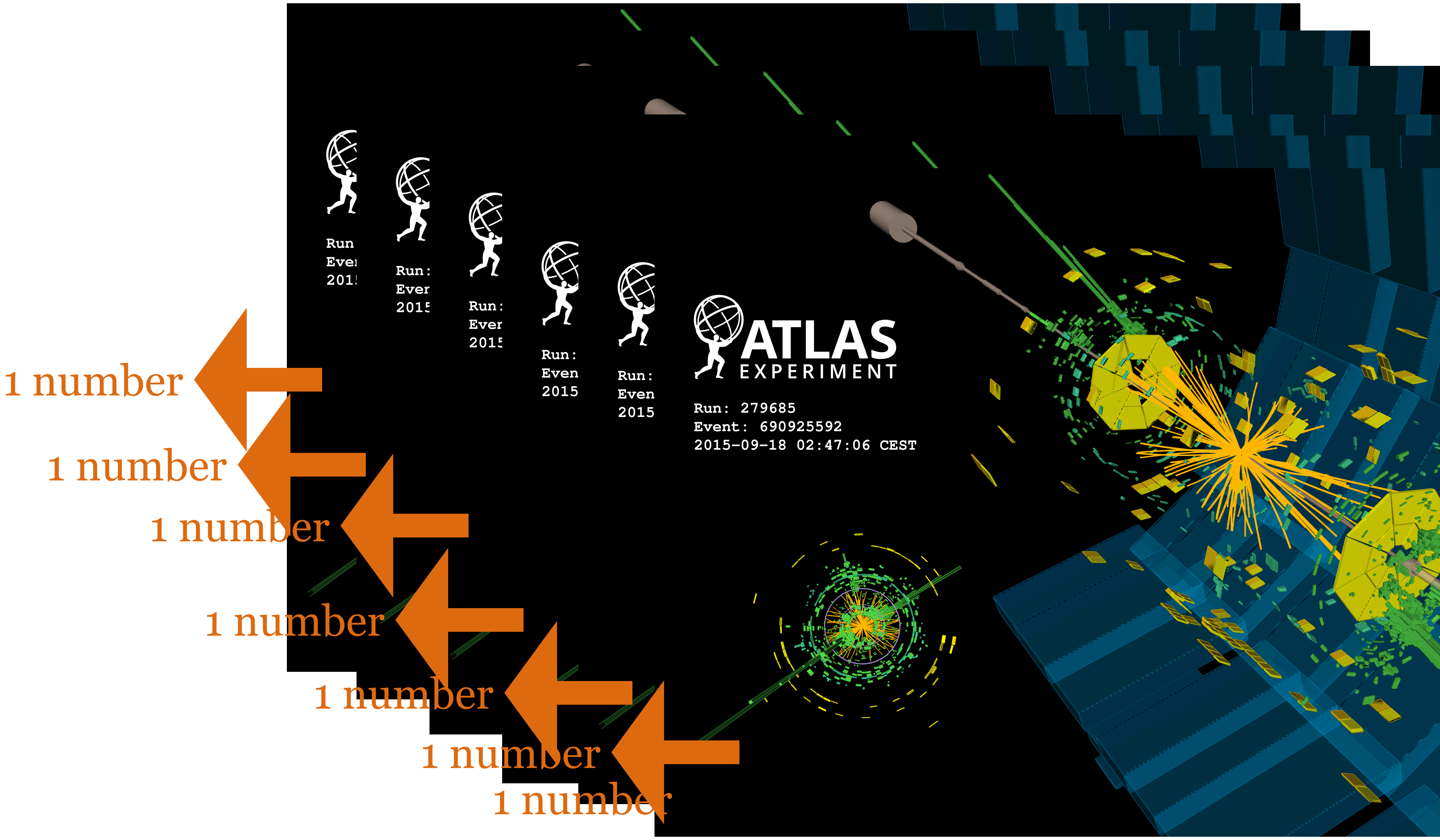


Traditional Approach: Design one sensitive observable

- Detector has O(100 million) sensors

- Reconstruction pipeline, event selection
- Design one summary variable
    - Compression: O(100 million) → 1
- Build a 1-D histogram
- Calculate likelihoods with histograms

Thousands of collision events

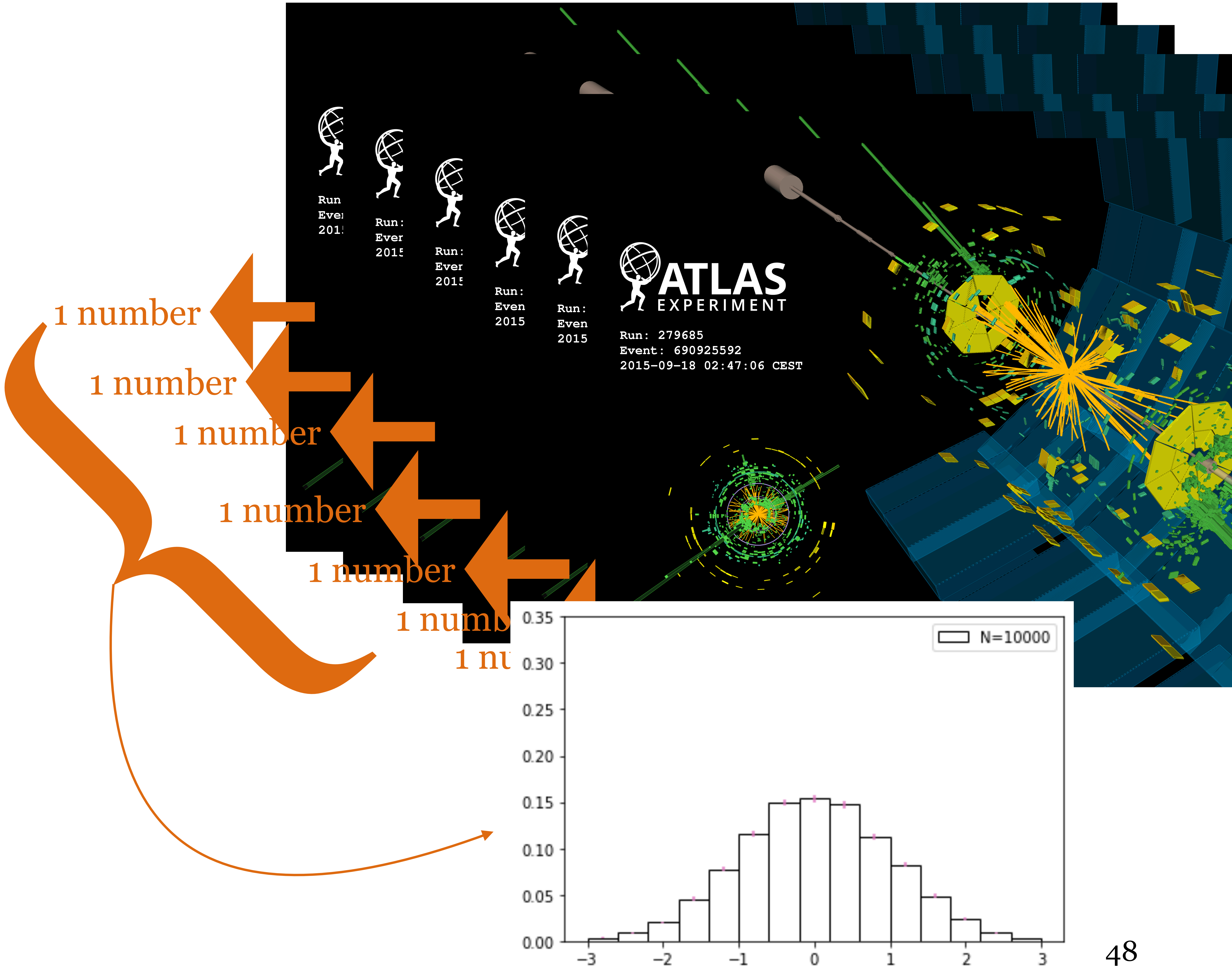


Traditional Approach: Design one sensitive observable

- Detector has O(100 million) sensors

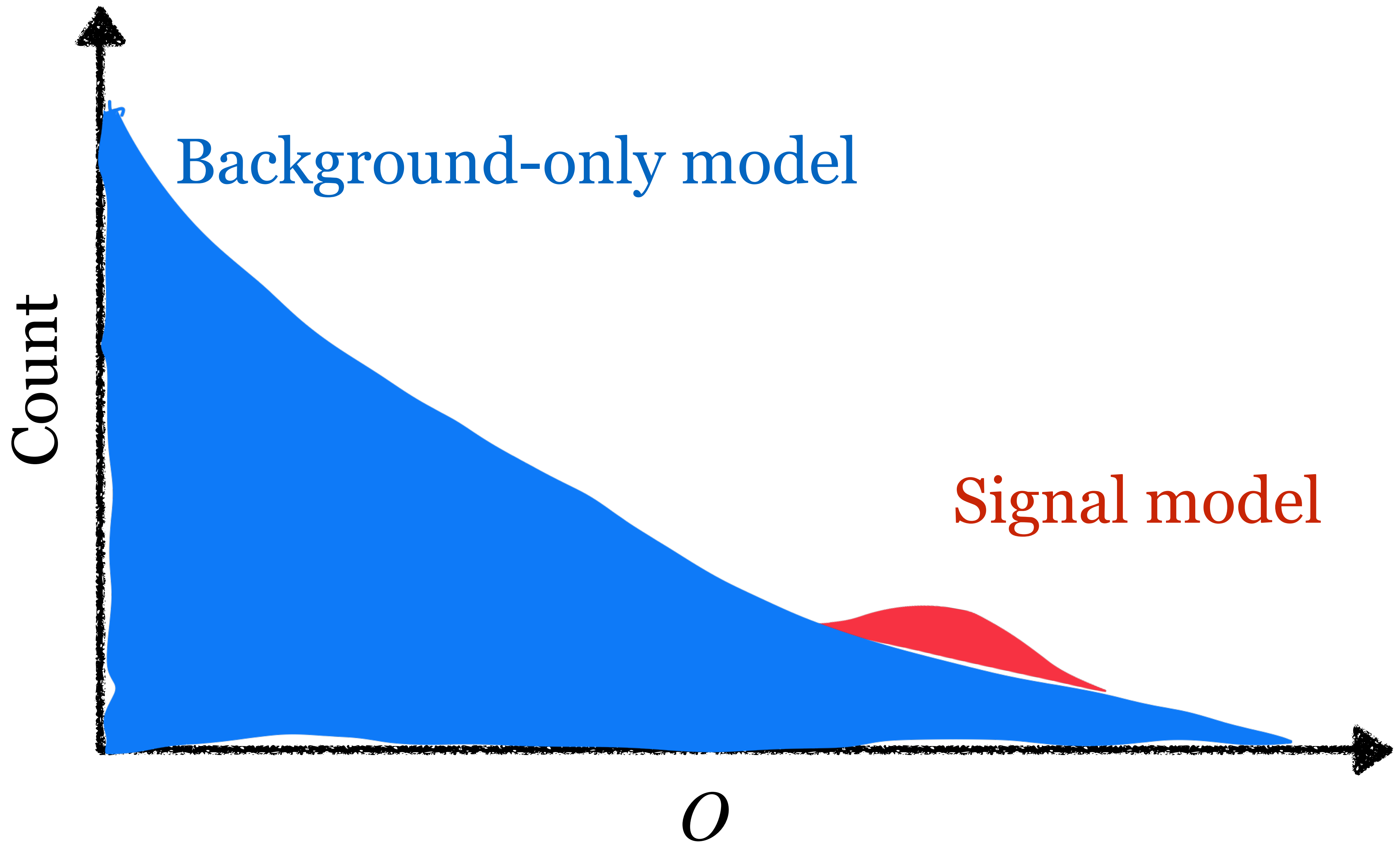
- Reconstruction pipeline, event selection
- Design one summary variable
  - Compression: O(100 million) → 1
- Build a 1-D histogram
- Calculate likelihoods with histograms

Thousands of collision events

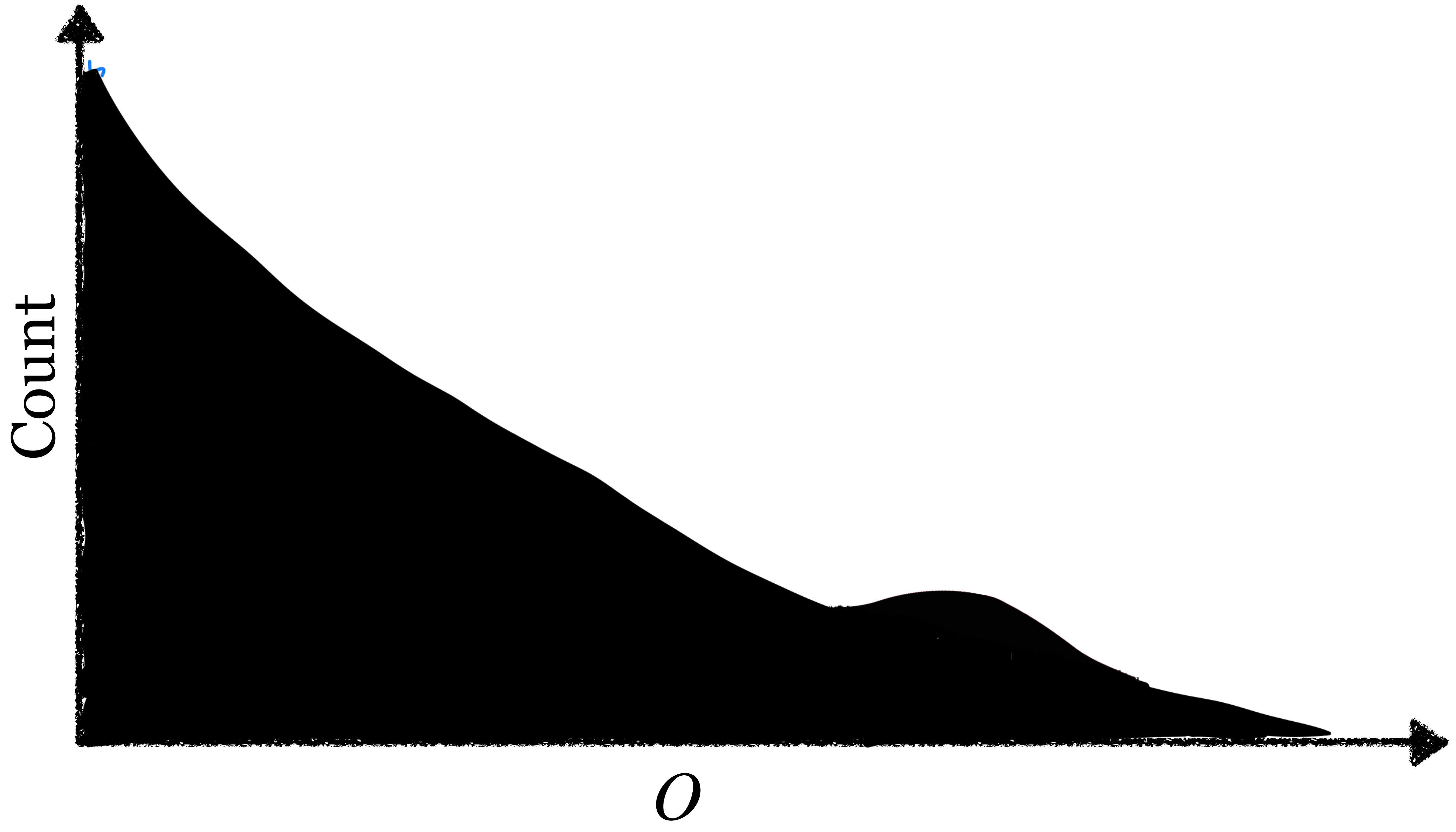


Probability density estimation with 1-D histogram

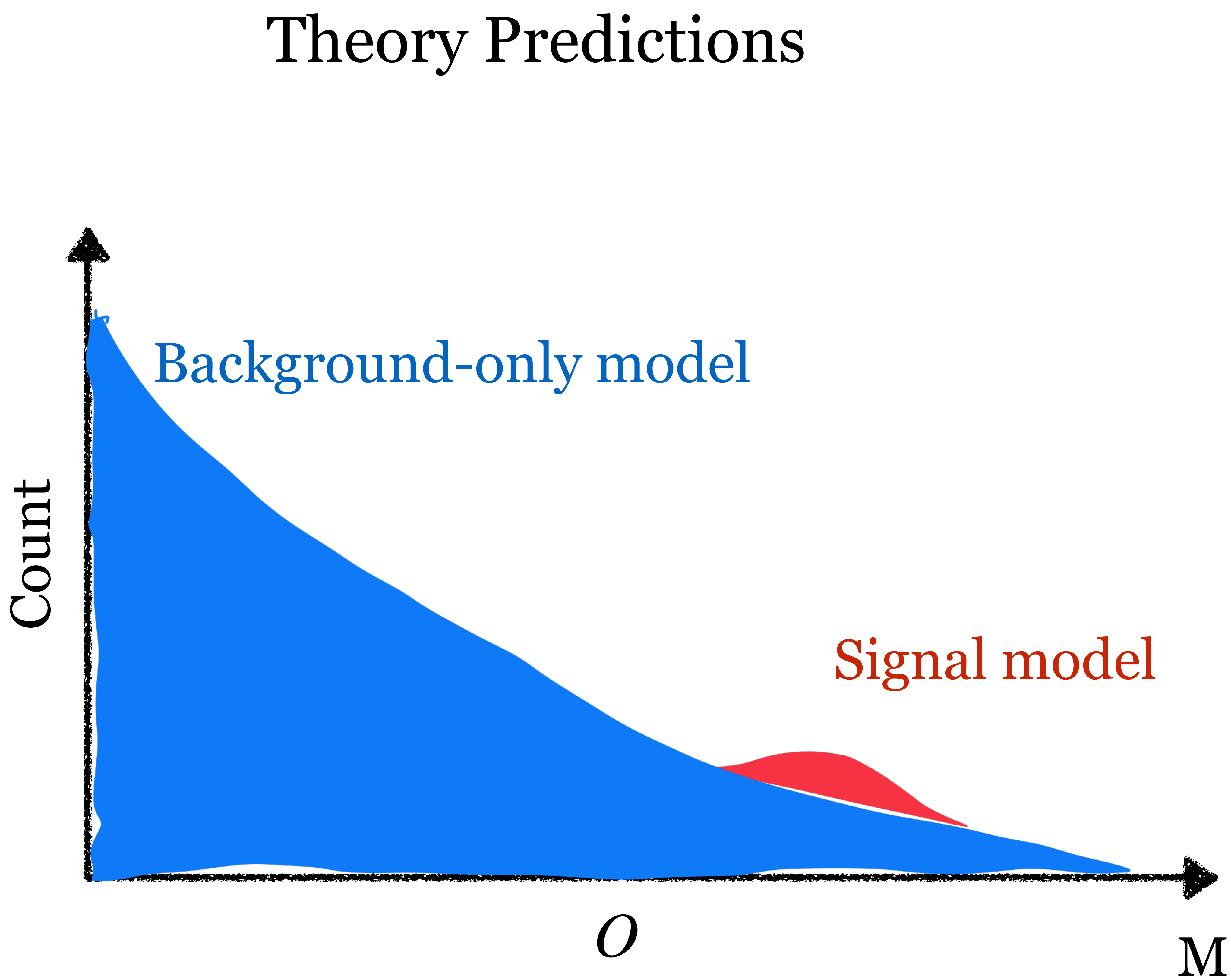
Theory Predictions

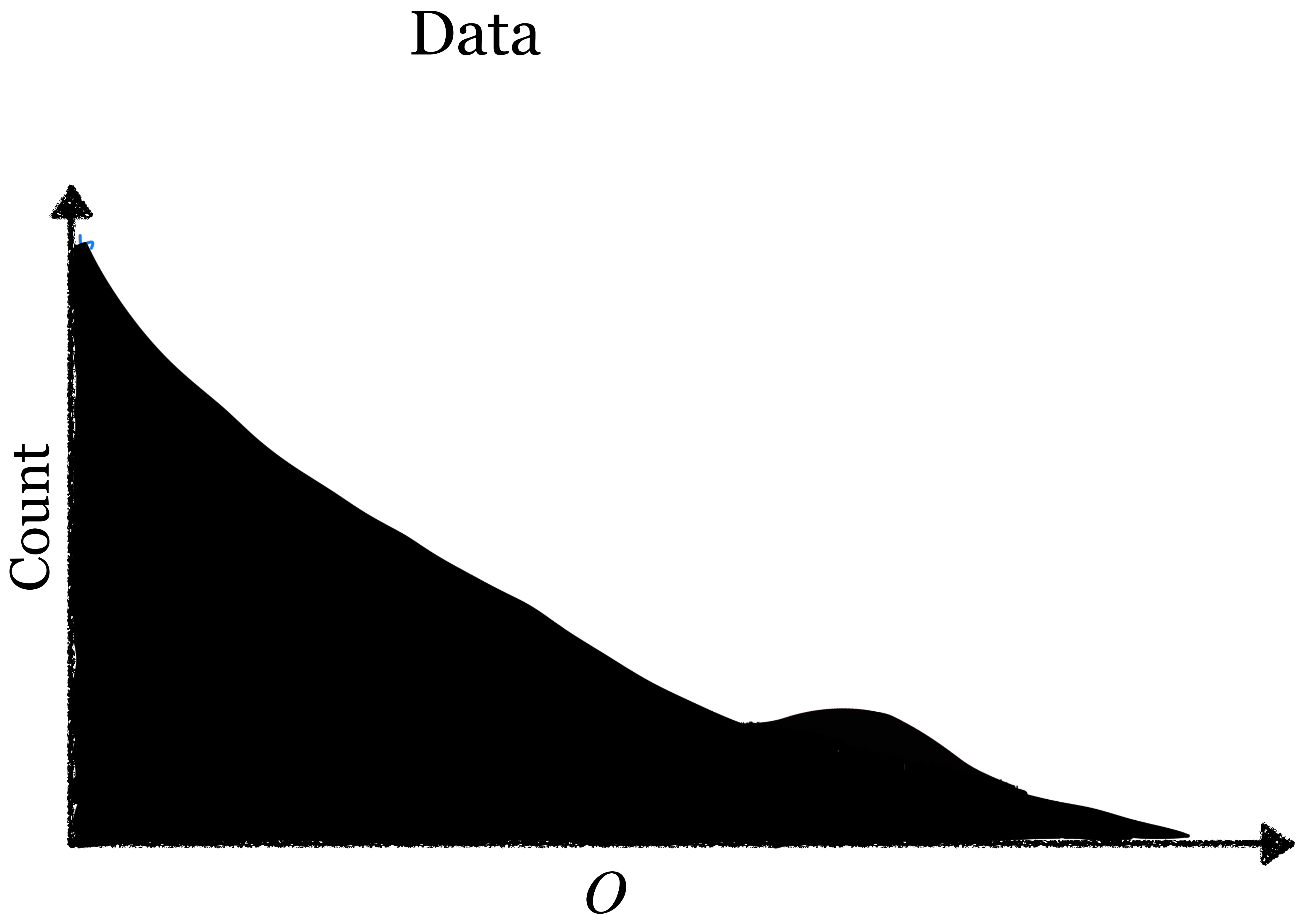


Data



Probability density estimation with 1-D histogram

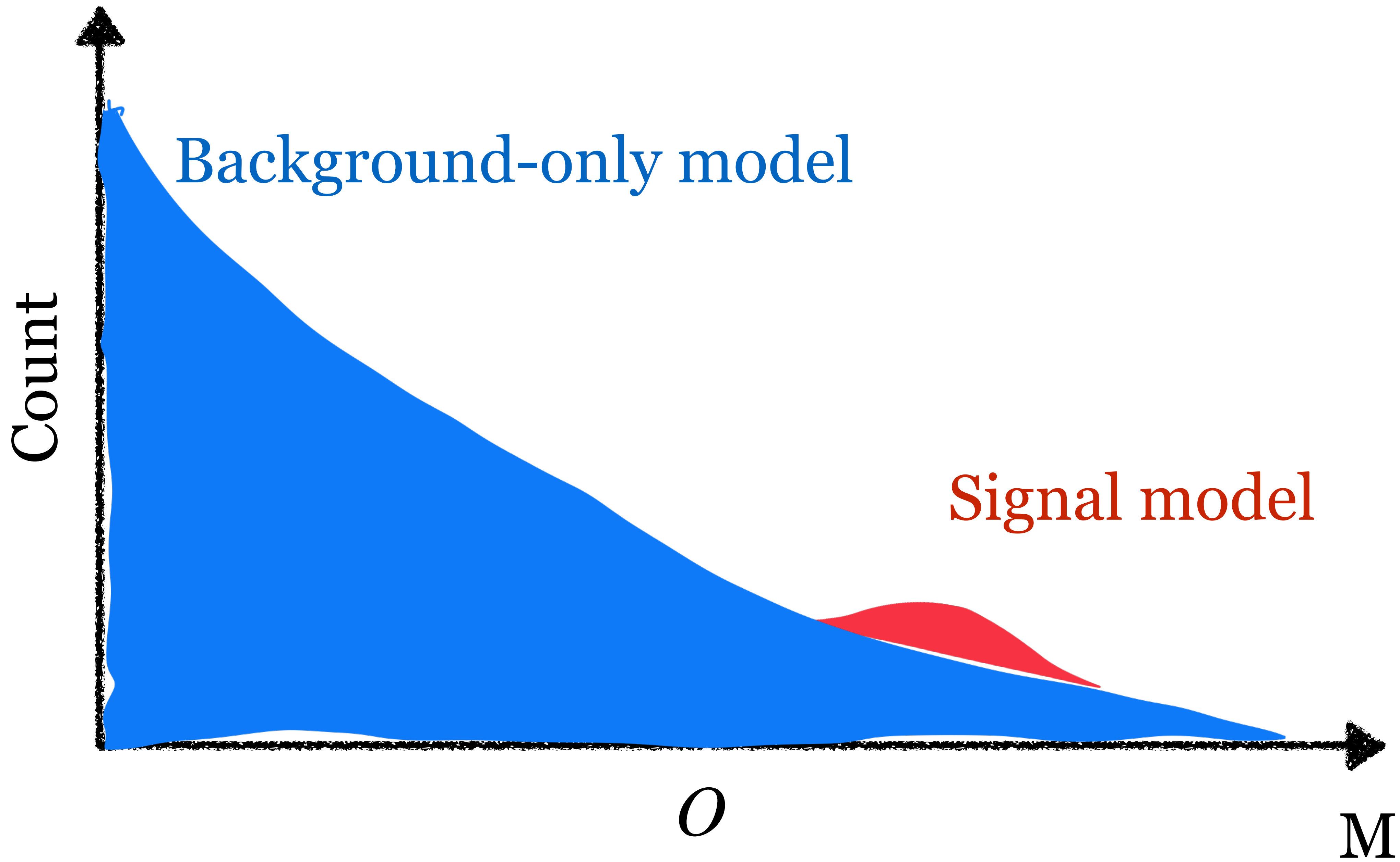




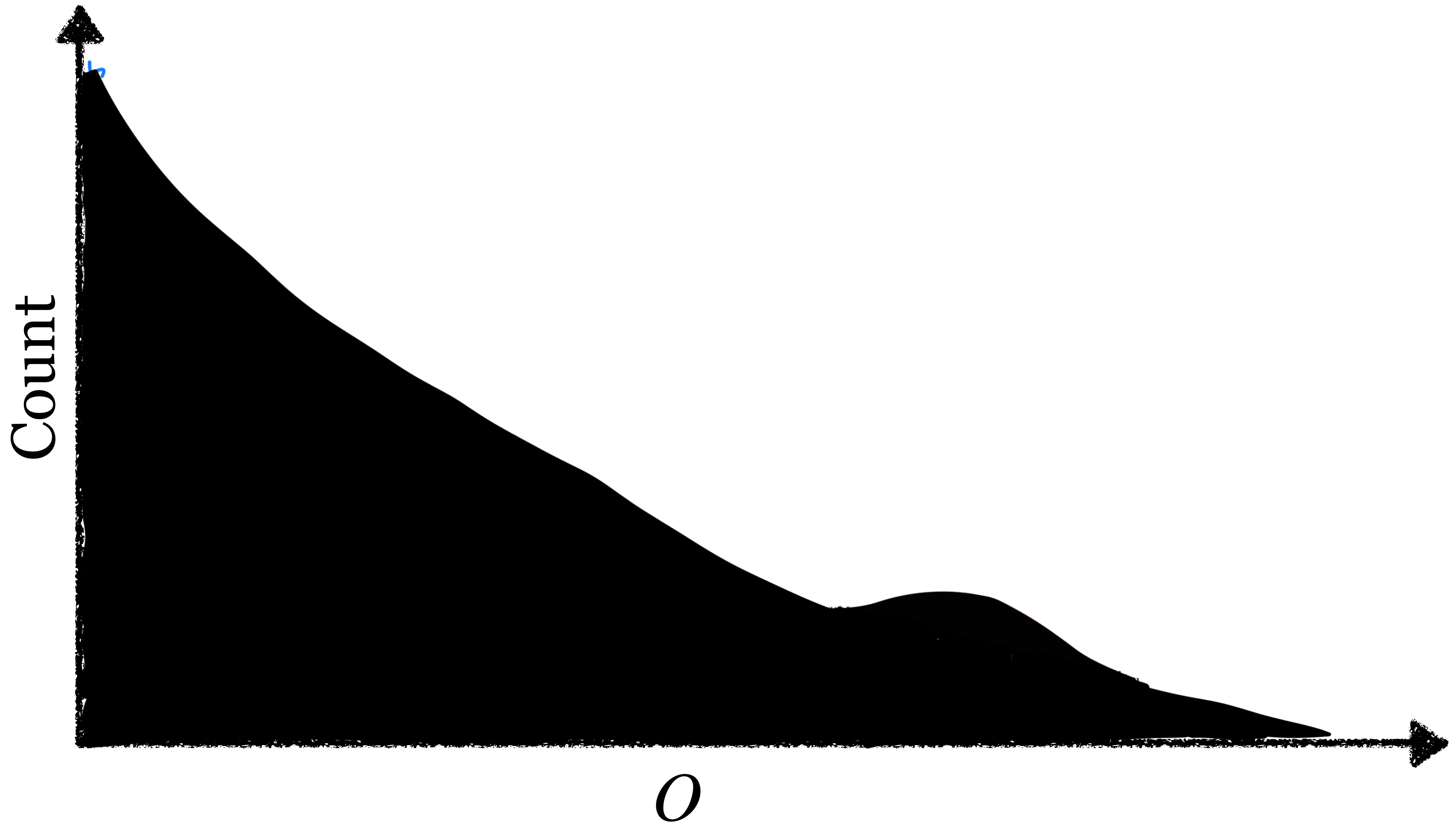
Measure ‘signal strength’ μ

Probability density estimation with 1-D histogram

Theory Predictions

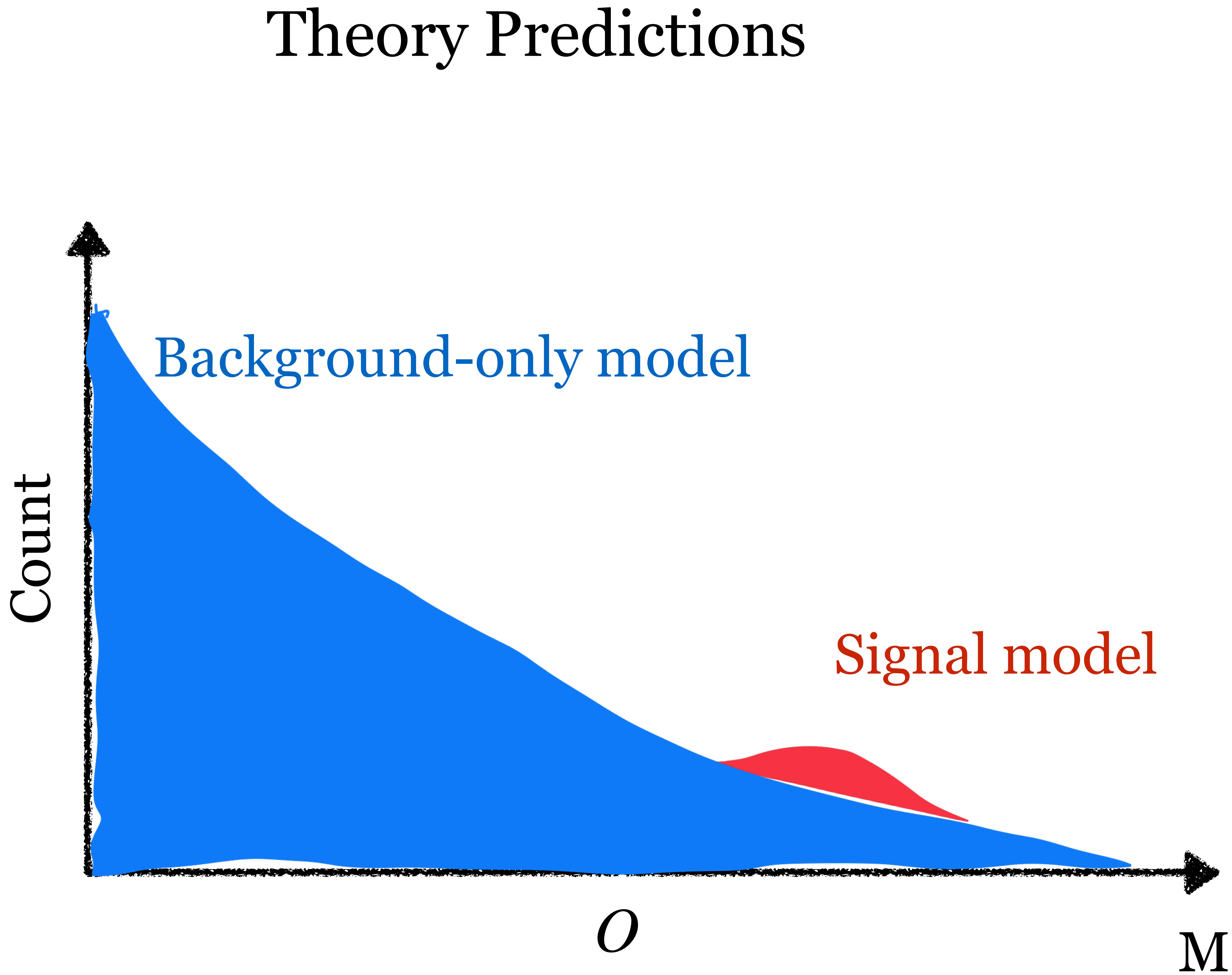


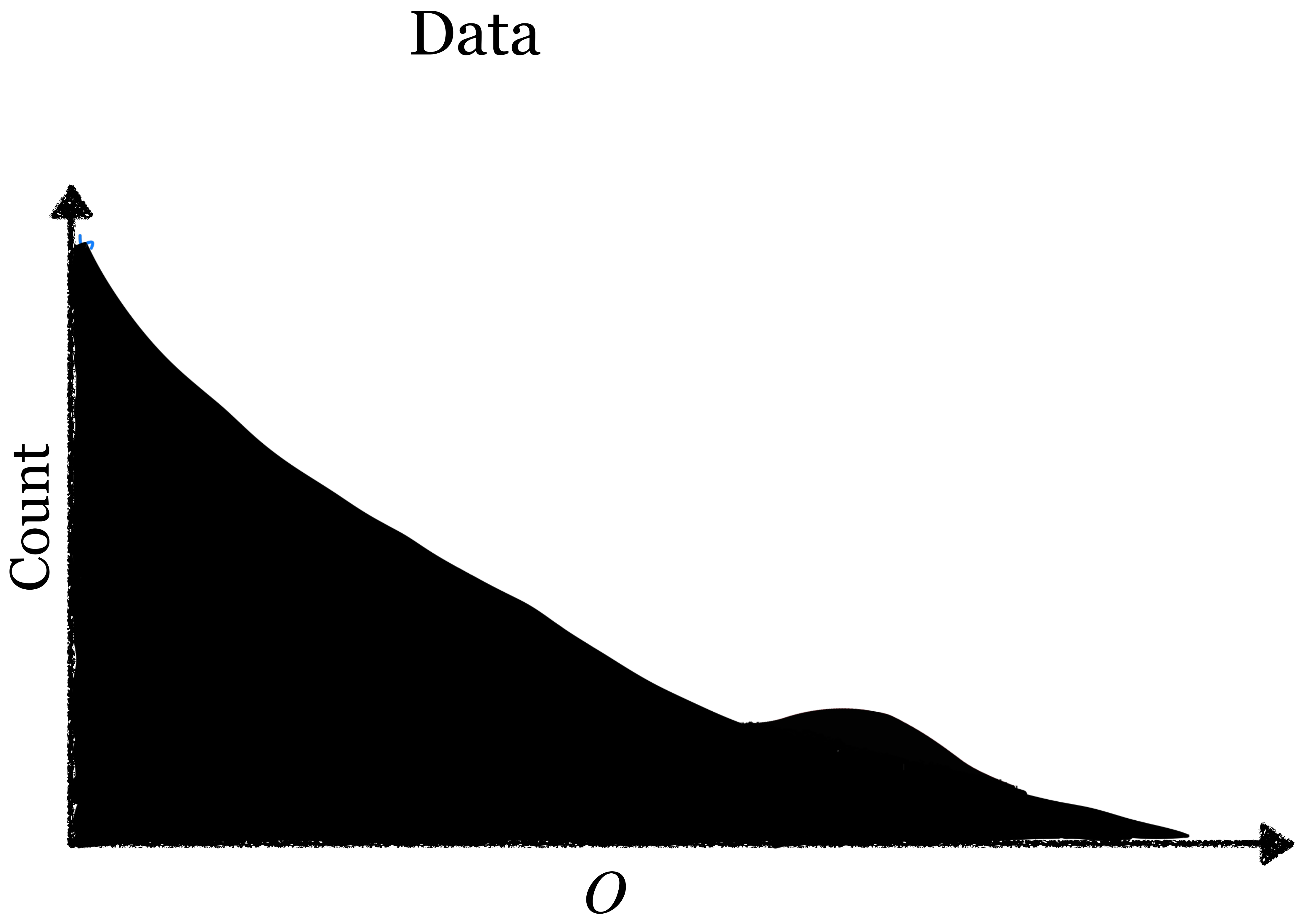
Data



Measure ‘signal strength’ μ

Probability density estimation with 1-D histogram

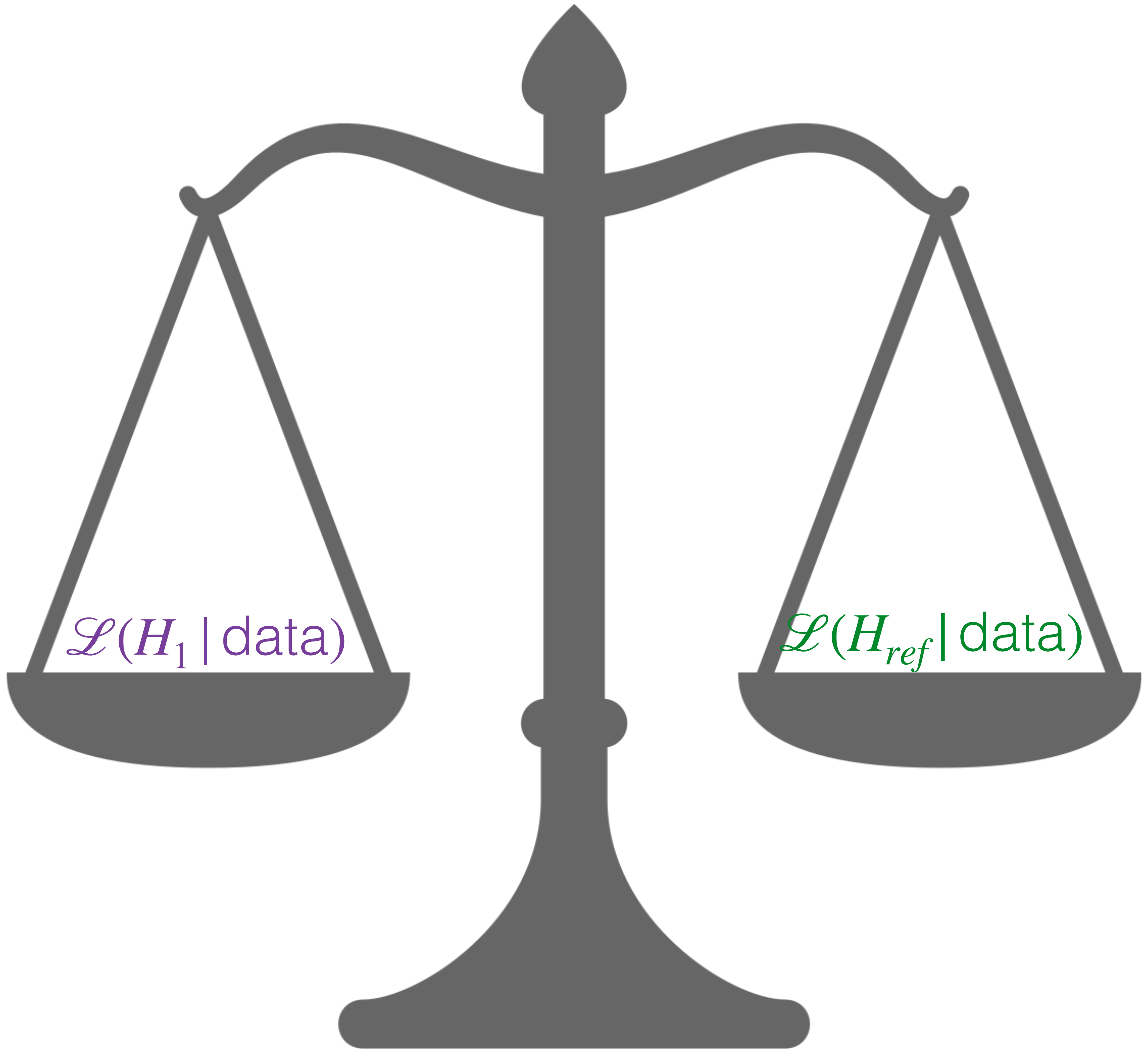




Measure ‘signal strength’ μ

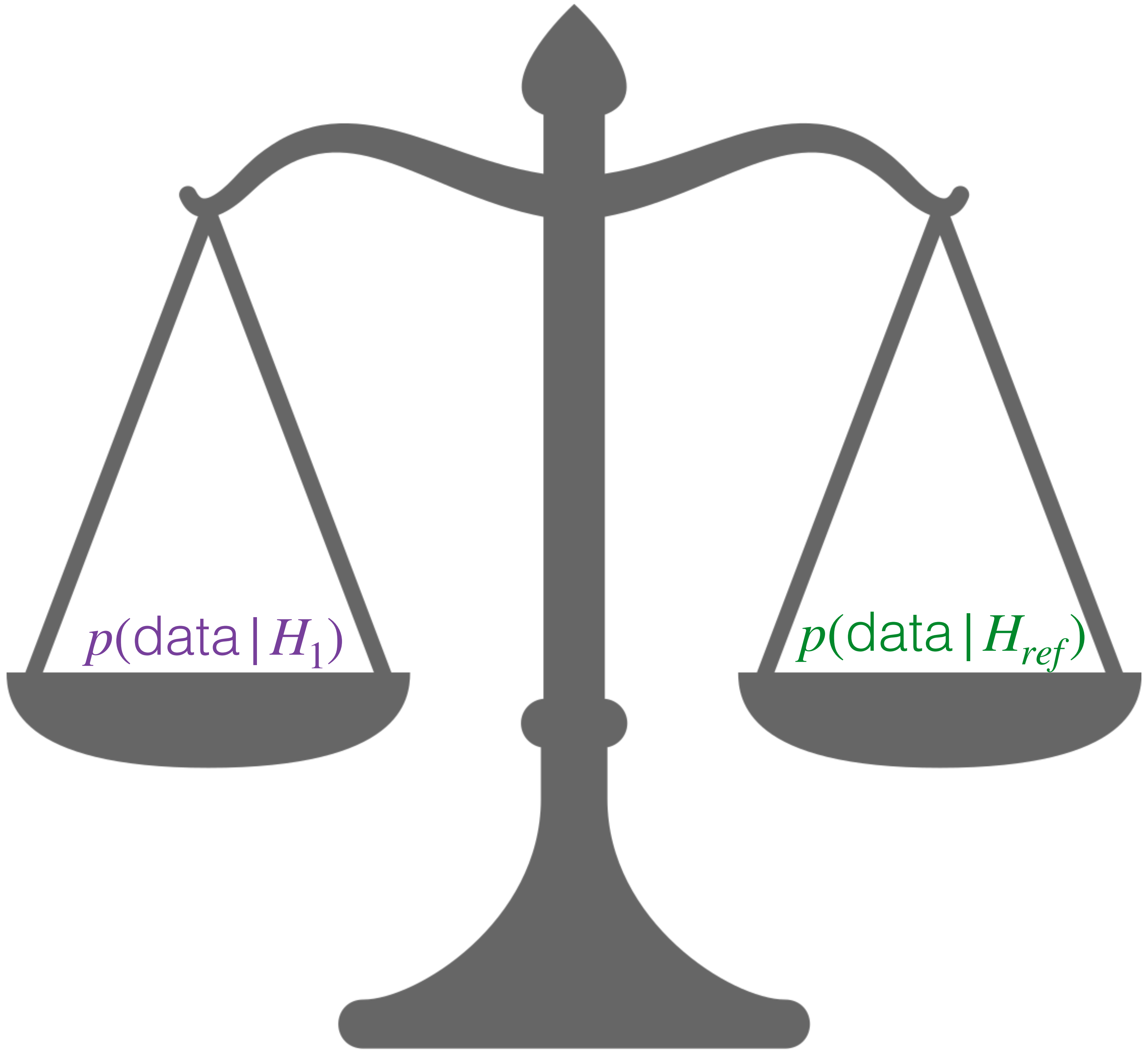
With histograms we can ask “Given the data, what is the likelihood of $\mu = 1$ hypothesis vs $\mu = 2$ hypothesis?”

Hypothesis tests



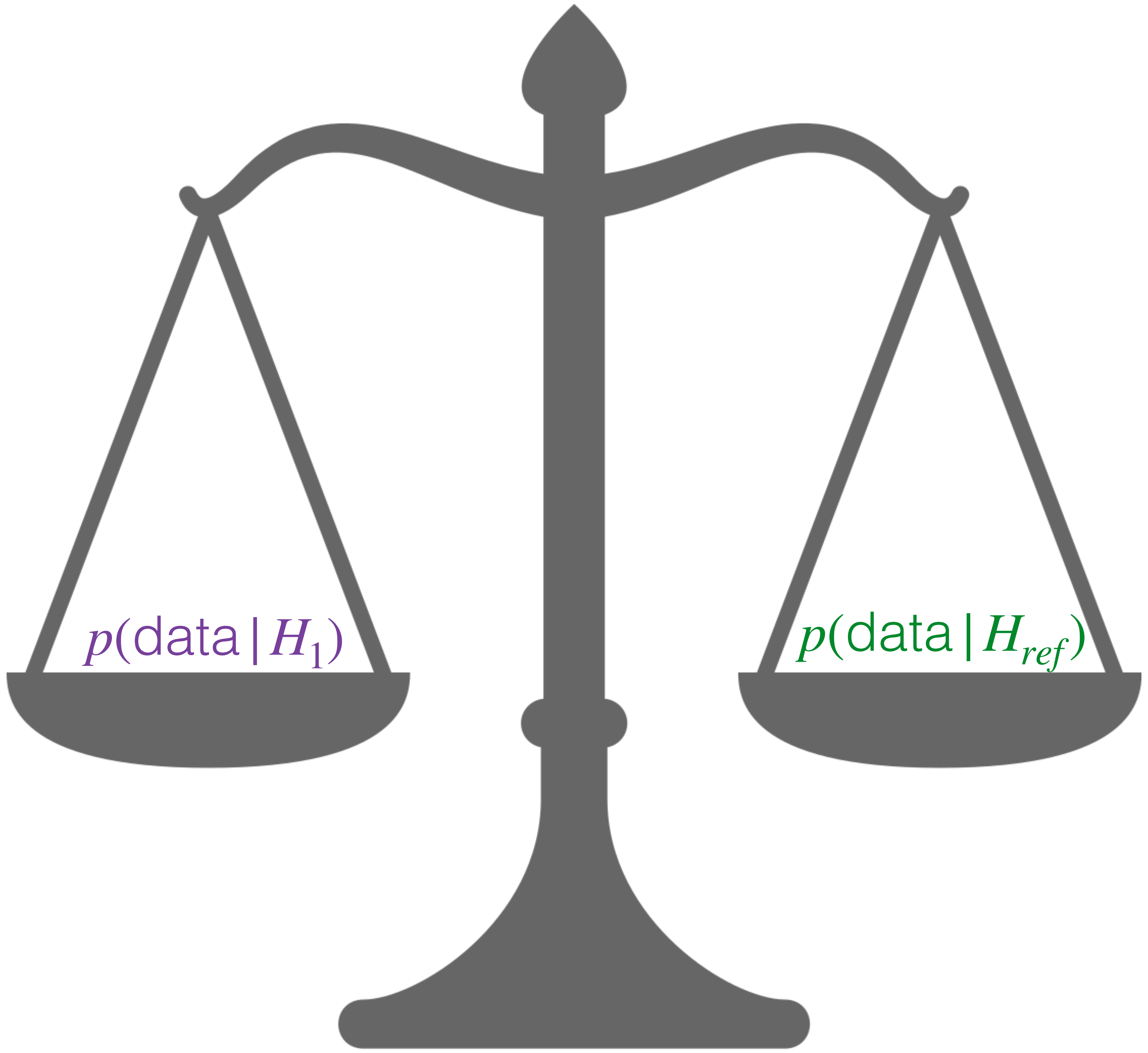
Hypothesis tests

$$\mathscr{L}(H_1 \,|\, \text{data}) = p(\text{data} \,|\, H_1)$$



Hypothesis tests

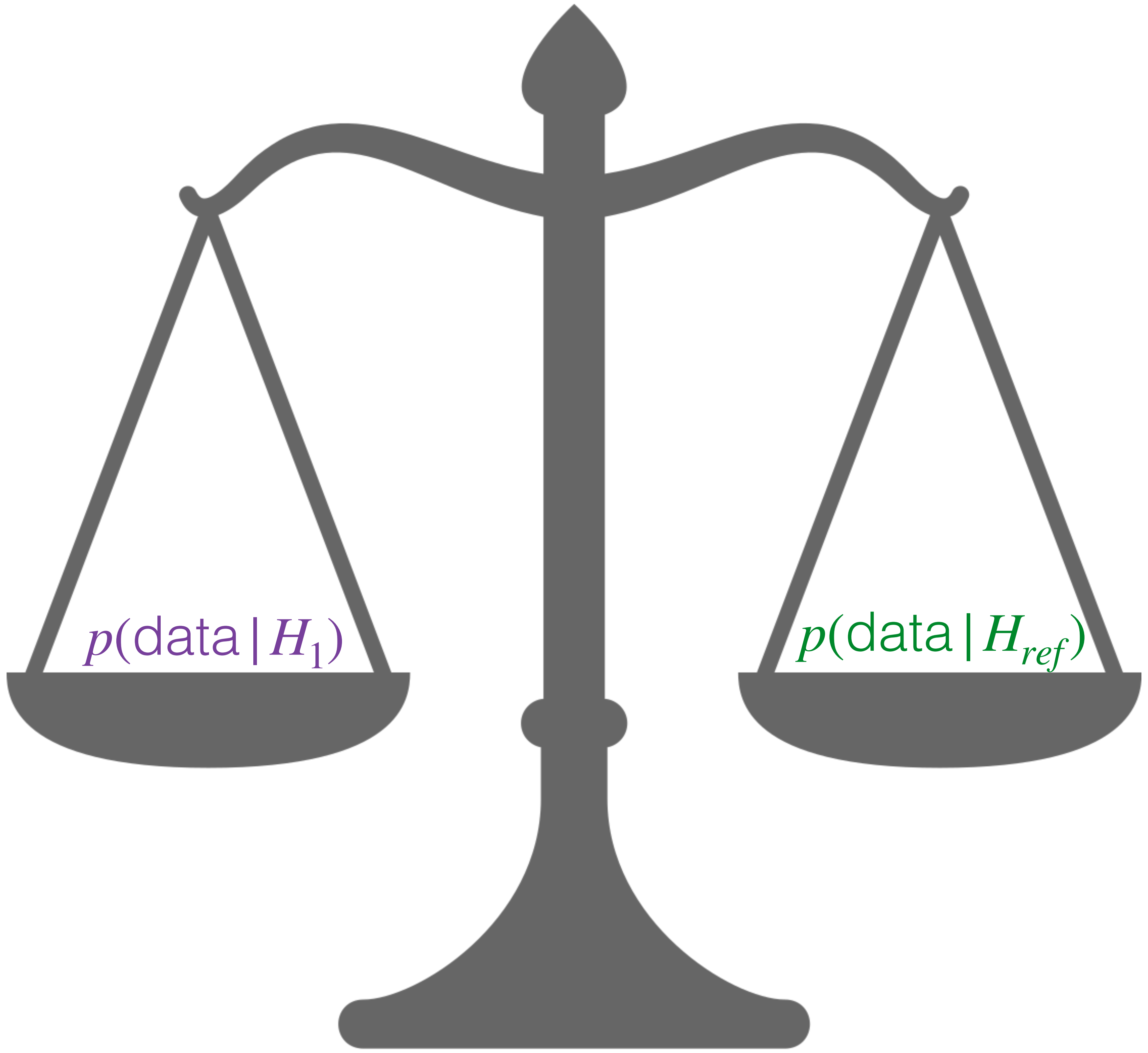
$$\mathscr{L}(H_1 | \text{data}) = p(\text{data} | H_1)$$



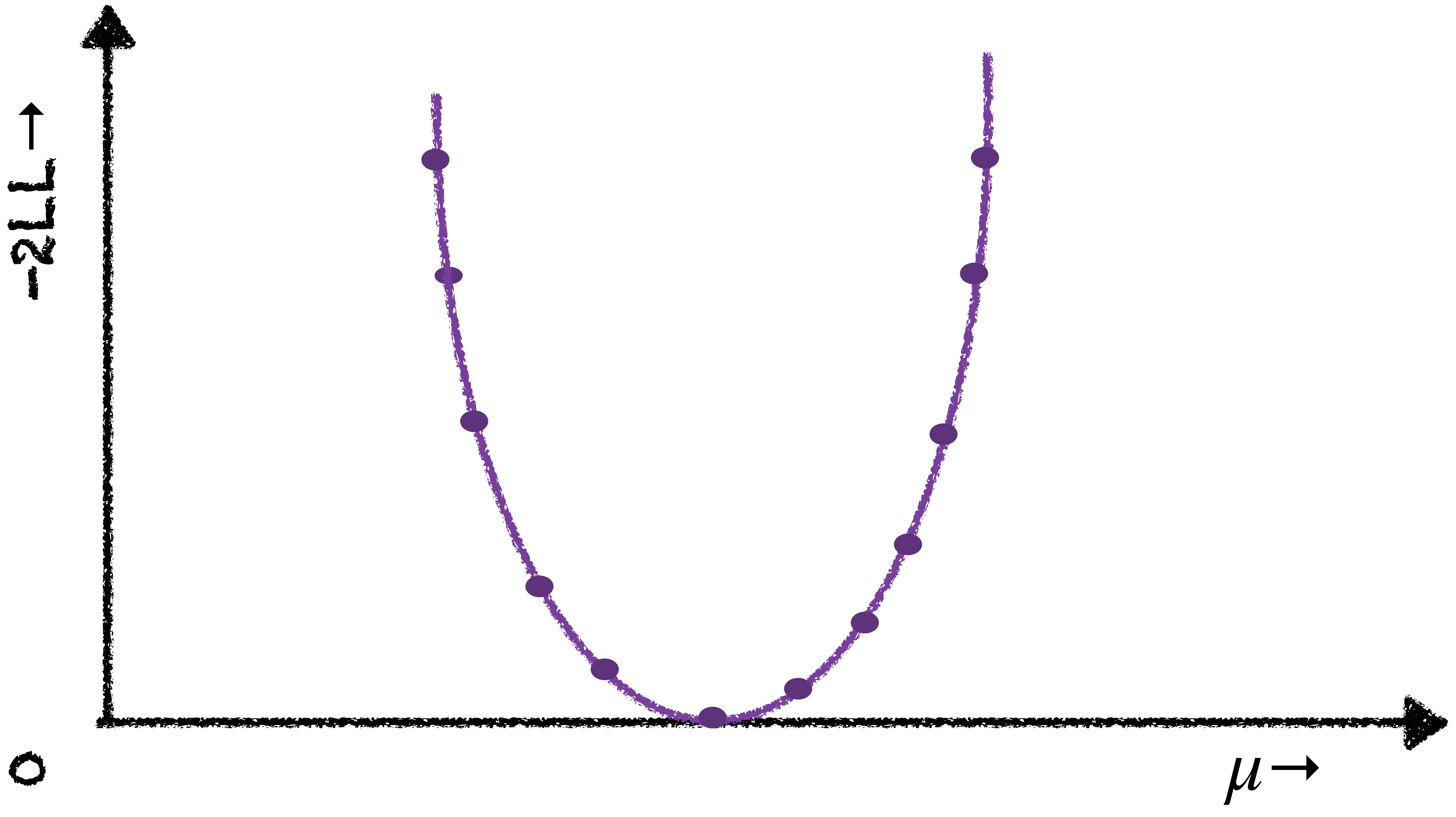
When comparing 2 hypotheses, guaranteed to be optimal test by Neyman-Person lemma

Hypothesis tests

$$\mathscr{L}(H_1 \,|\, \text{data}) = p(\text{data} \,|\, H_1)$$



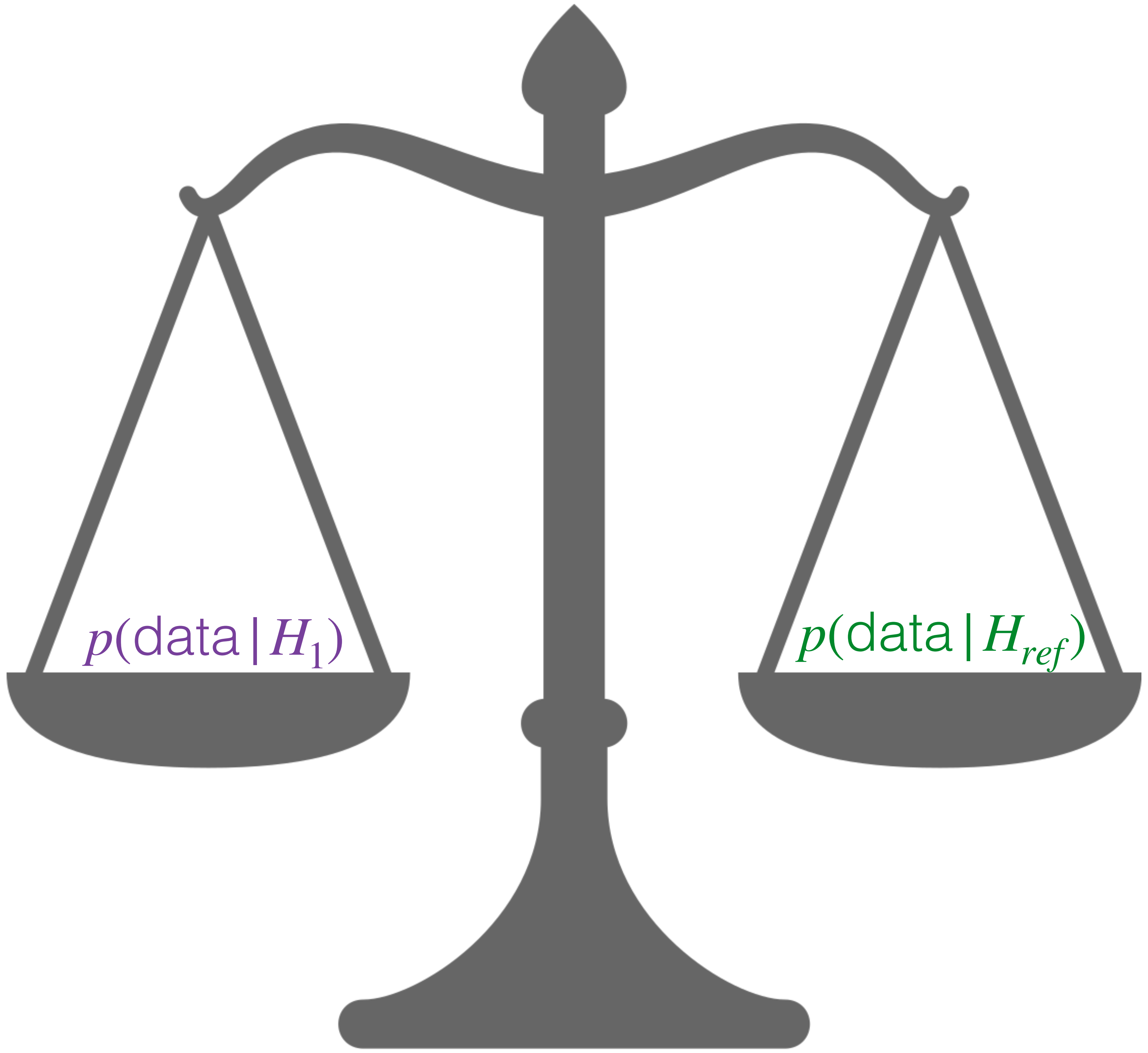
When comparing 2 hypotheses, guaranteed to be optimal test by Neyman-Person lemma



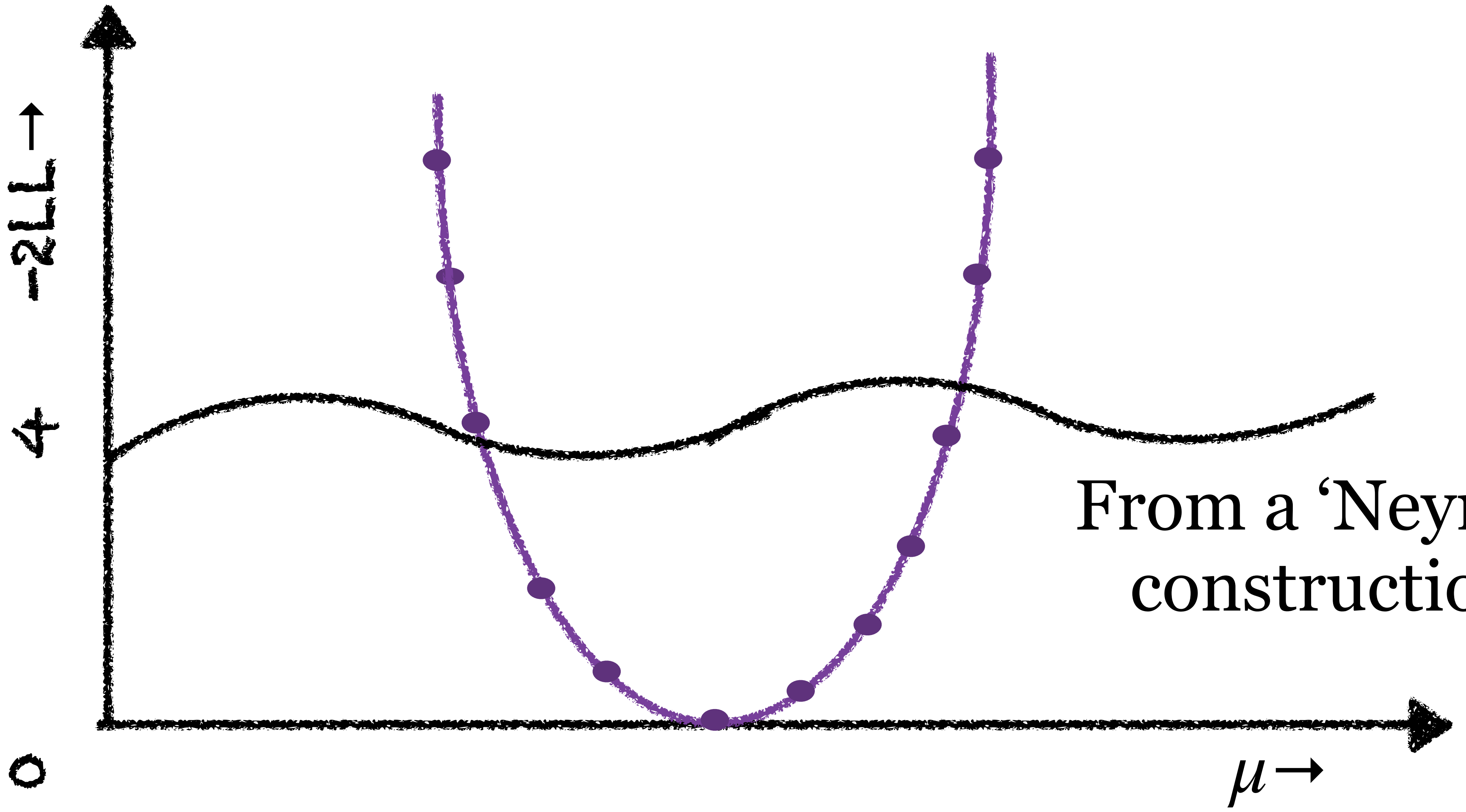
Parameter estimation (infinite hypotheses)

Hypothesis tests

$$\mathscr{L}(H_1 \,|\, \text{data}) = p(\text{data} \,|\, H_1)$$



When comparing 2 hypotheses, guaranteed to be optimal test by Neyman-Person lemma

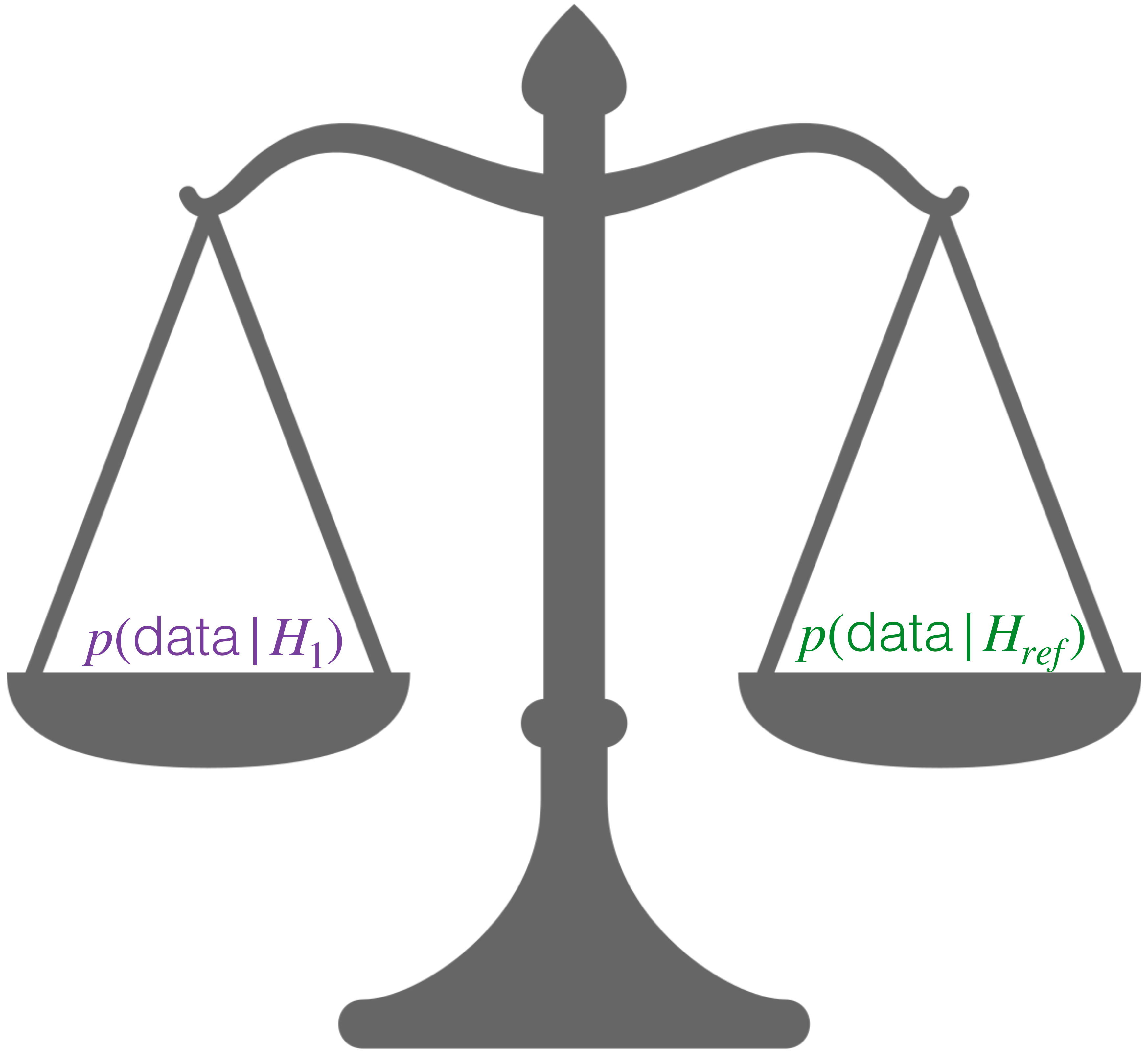


From a 'Neyman construction'

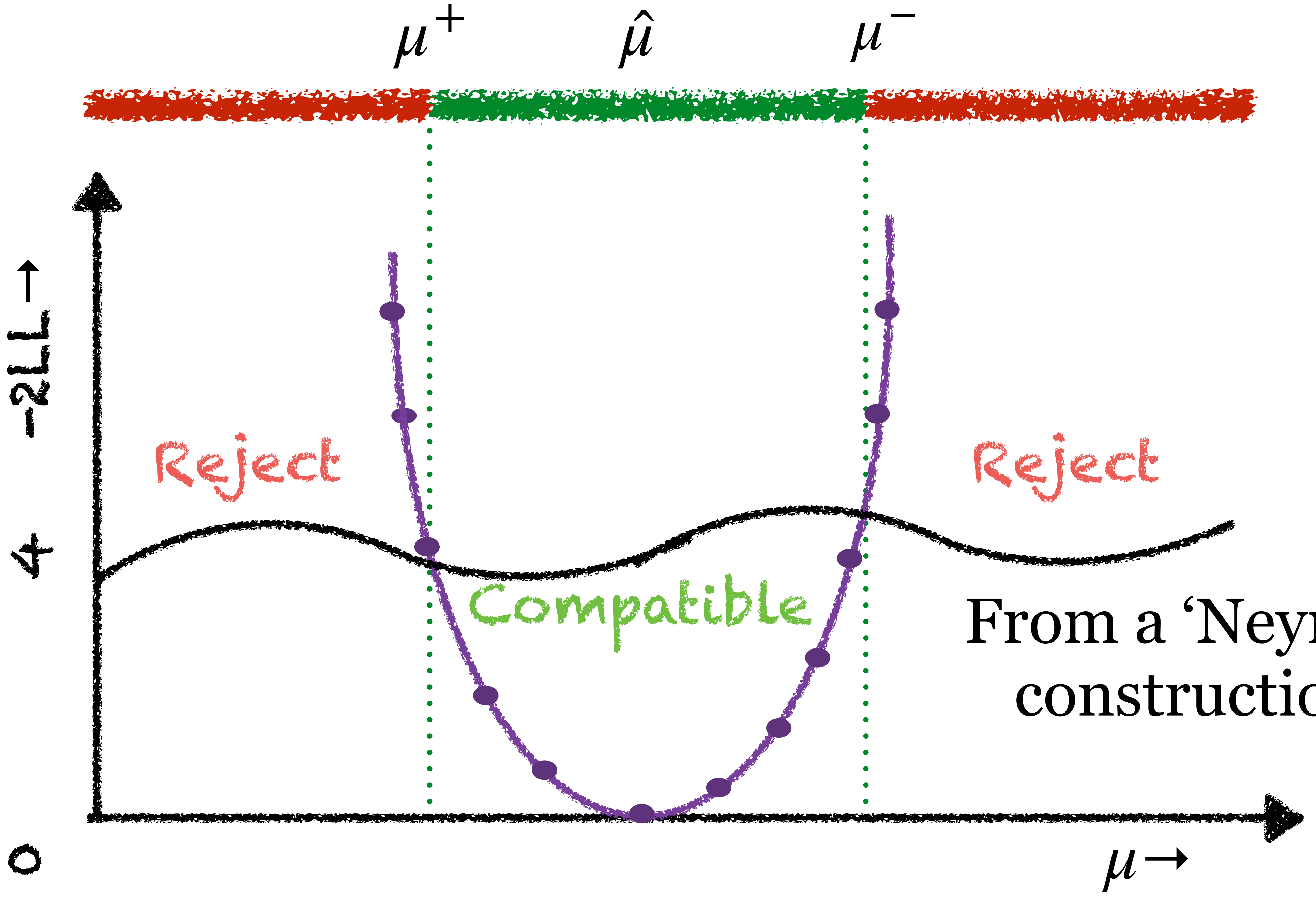
Parameter estimation (infinite hypotheses)

Hypothesis tests

$$\mathscr{L}(H_1 \,|\, \text{data}) = p(\text{data} \,|\, H_1)$$



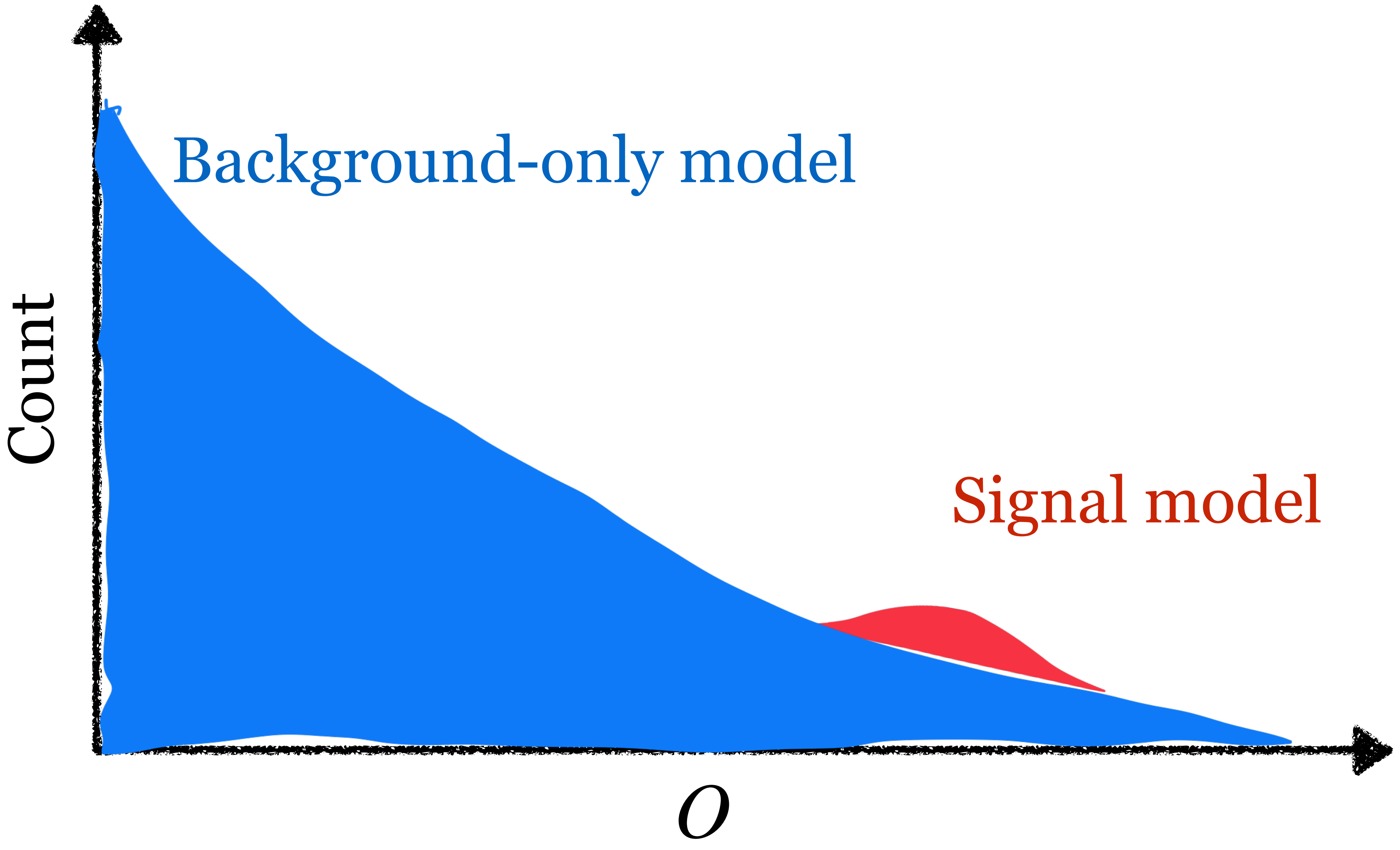
When comparing 2 hypotheses, guaranteed to be optimal test by Neyman-Person lemma



From a ‘Neyman construction’

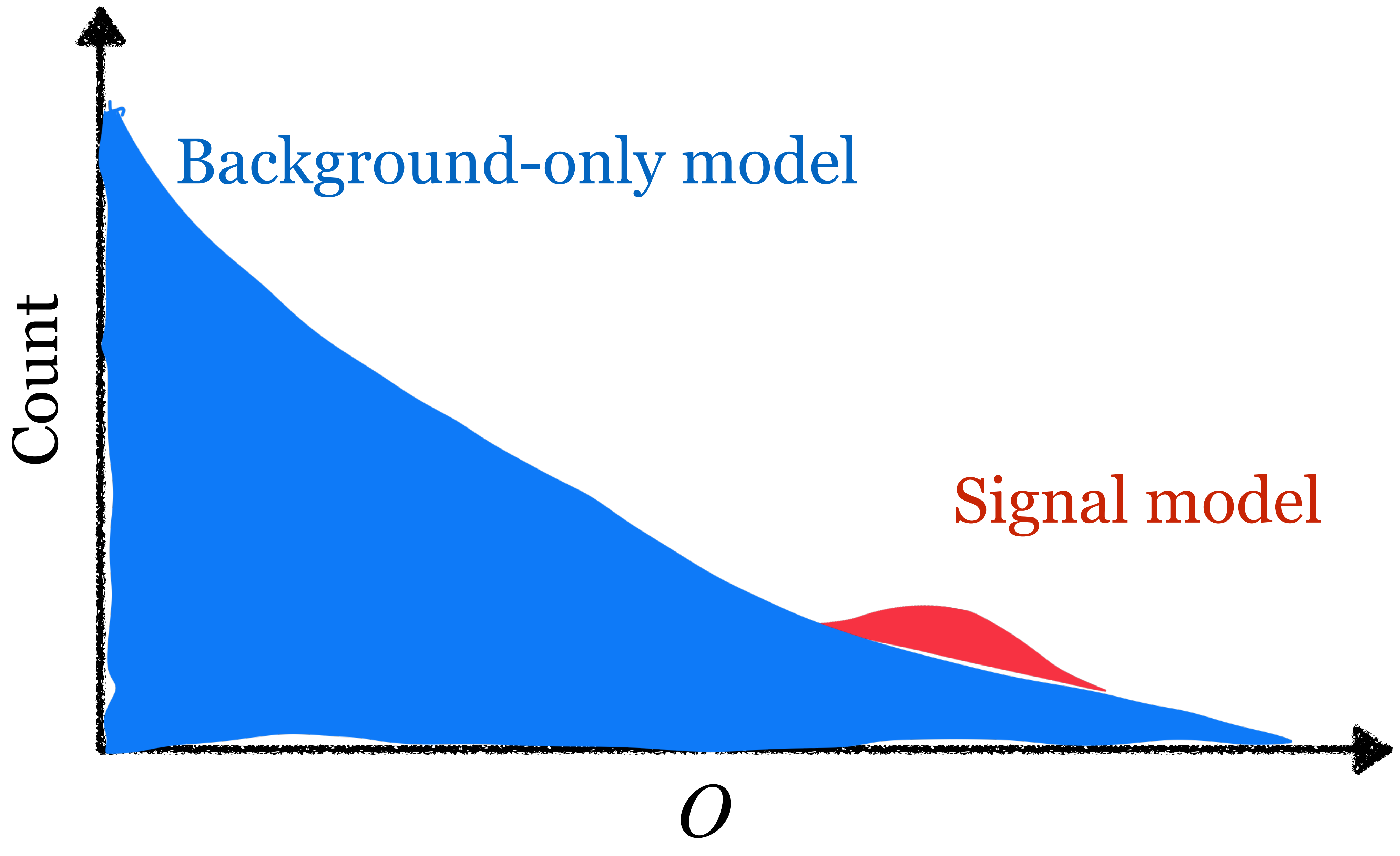
Parameter estimation (infinite hypotheses)

A new challenge: Quantum interference
Non-linear changes in kinematics

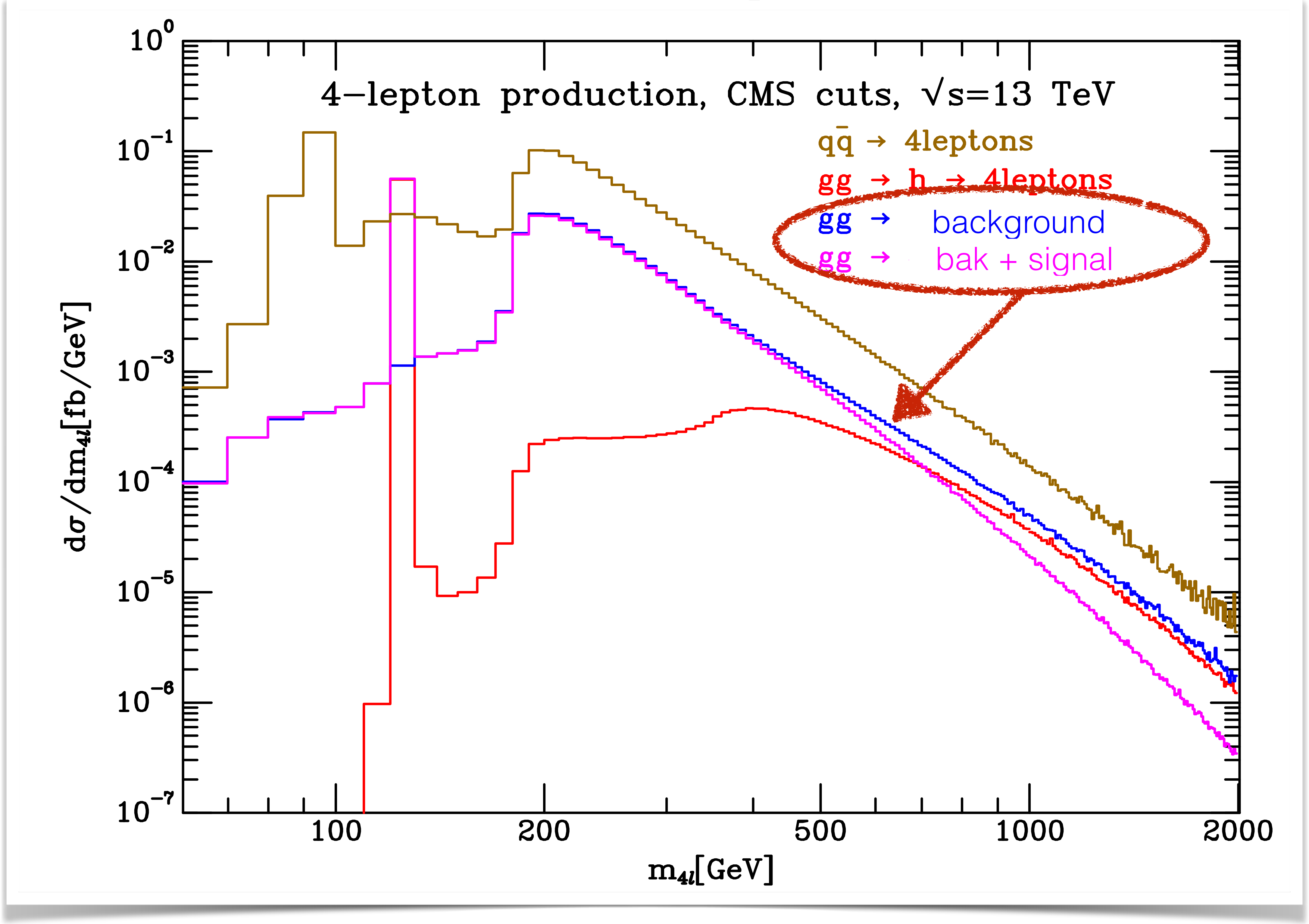


A new challenge: Quantum interference
Non-linear changes in kinematics

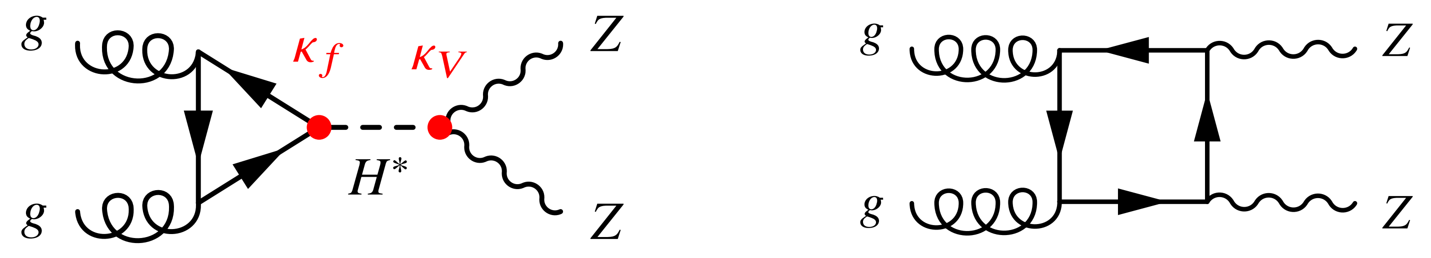
$$N_{exp} = \mu \cdot S + B + \sqrt{\mu} \cdot I$$



Campbell et al: arXiv:1311.3589



Quantum interference:

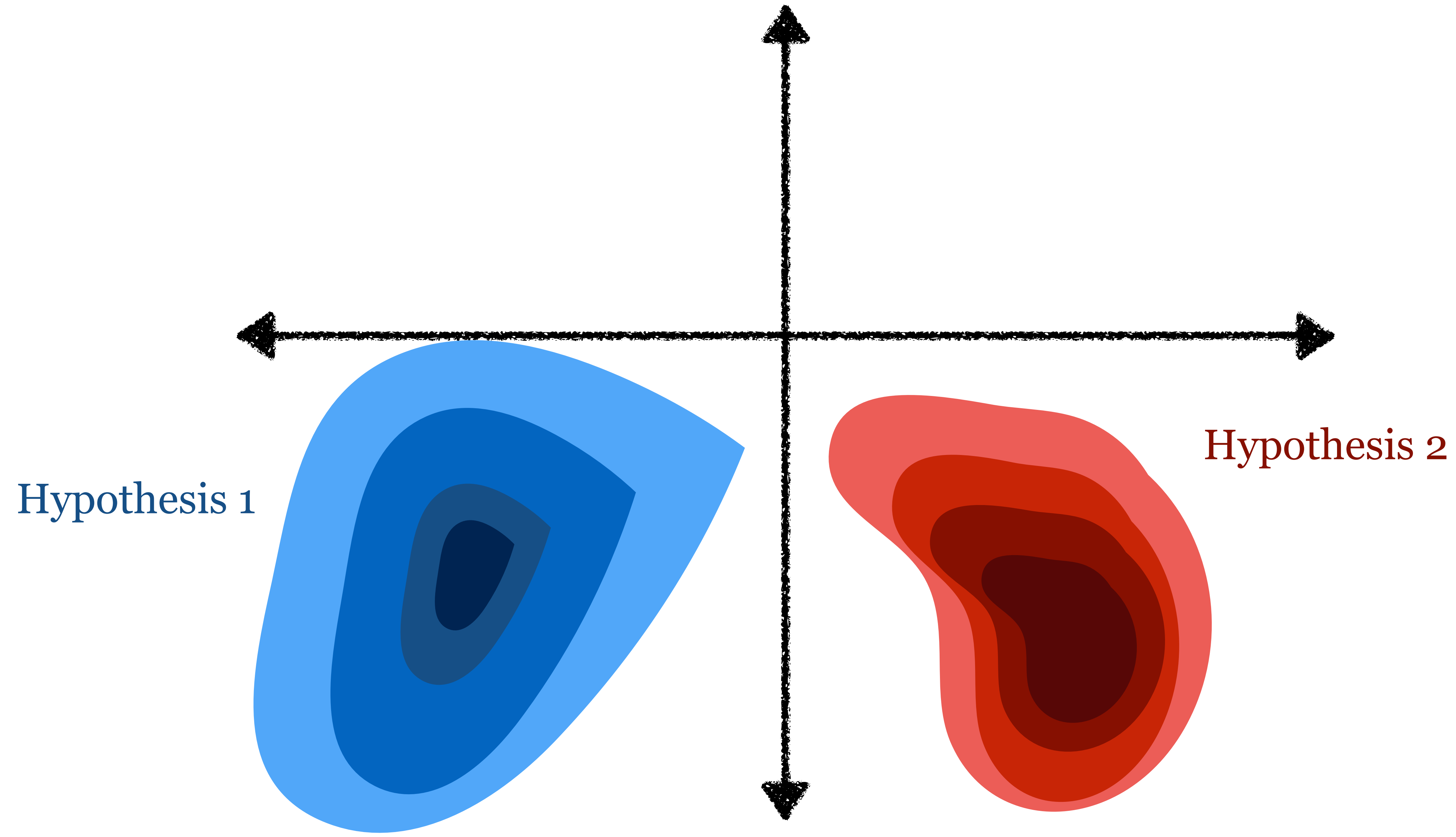


Any 1D histogram will be a lossy summary!

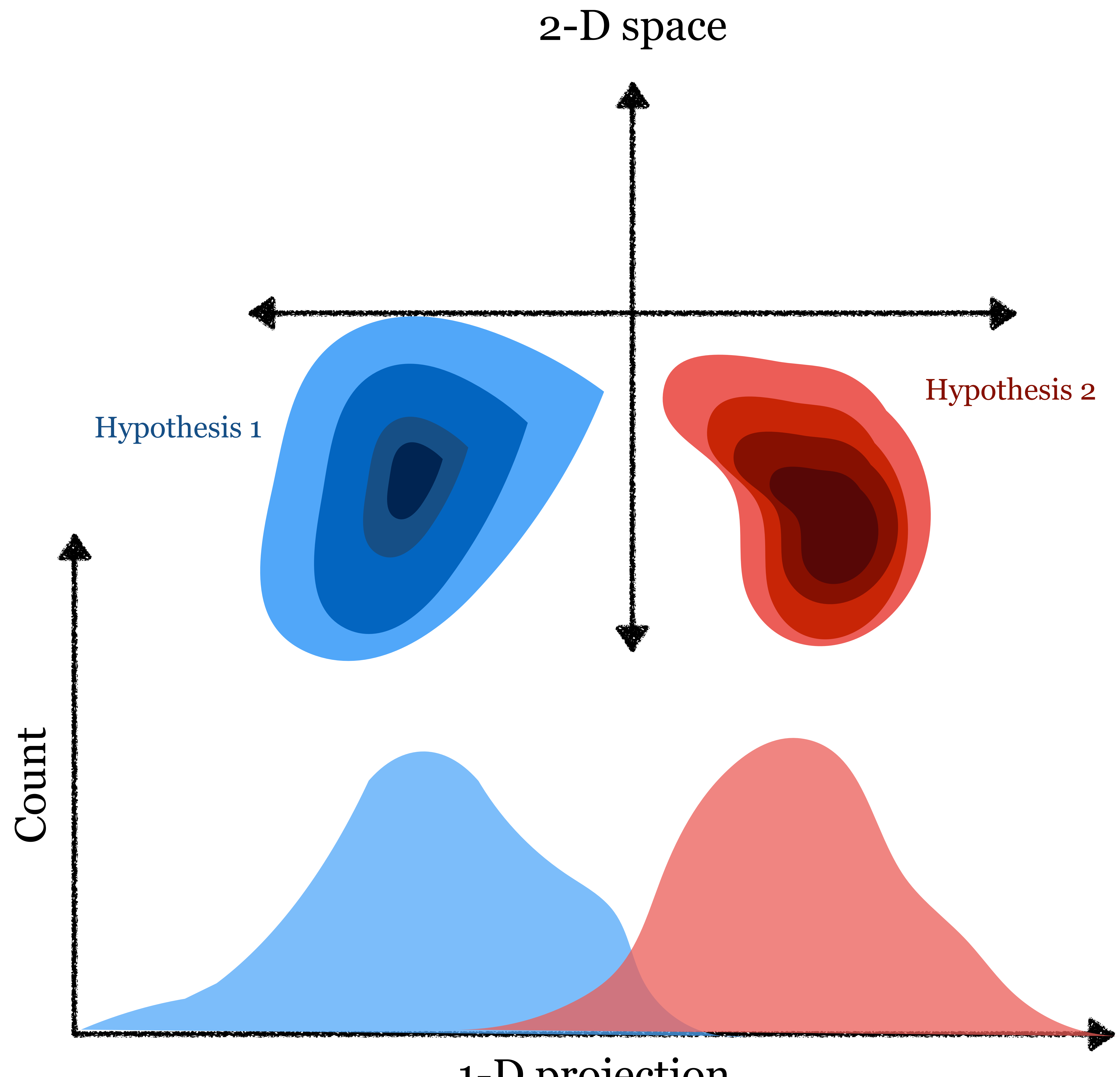
Can we always find a sufficient 1-D summary?

Can we always find a sufficient 1-D summary?

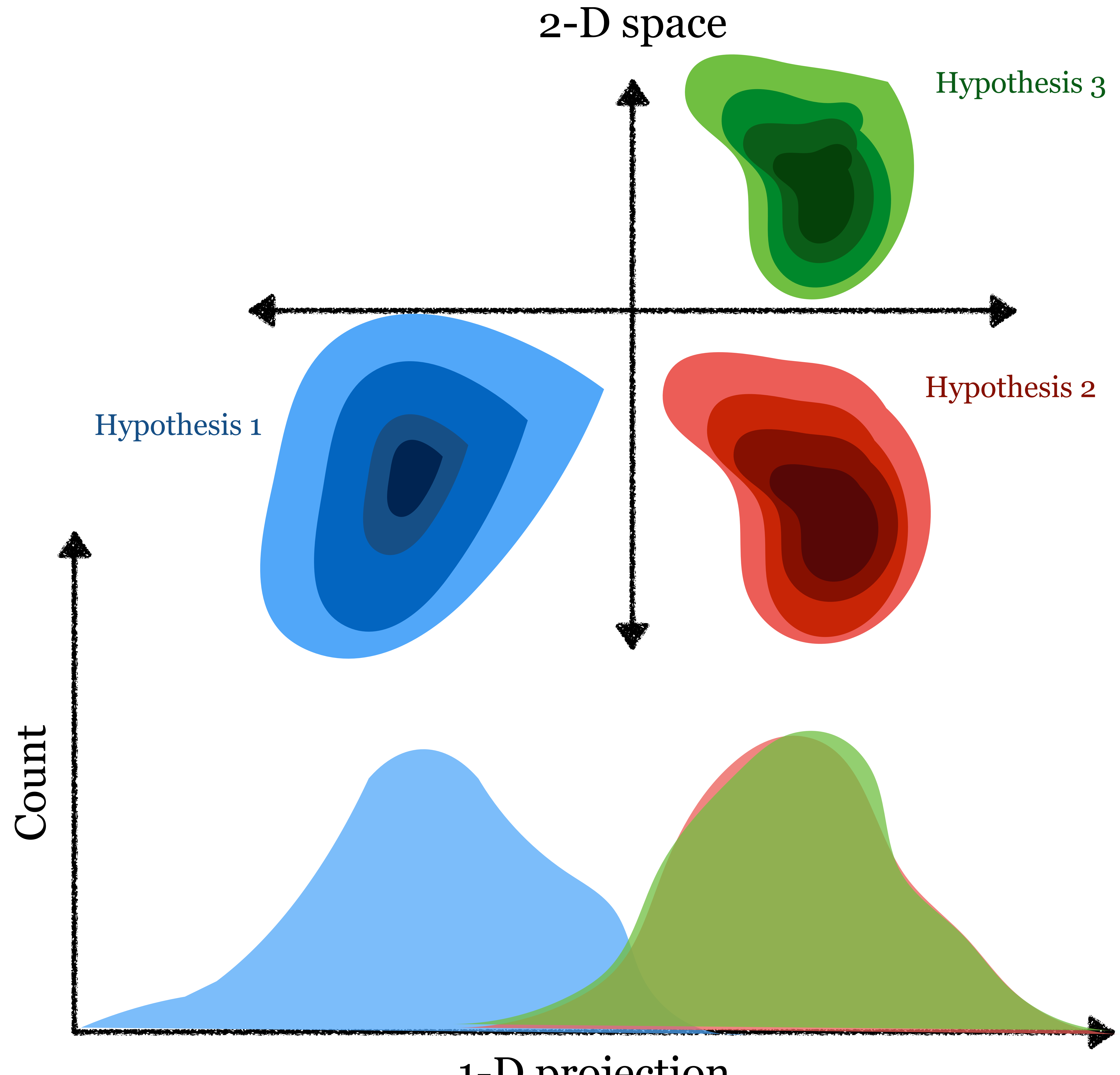
2-D space



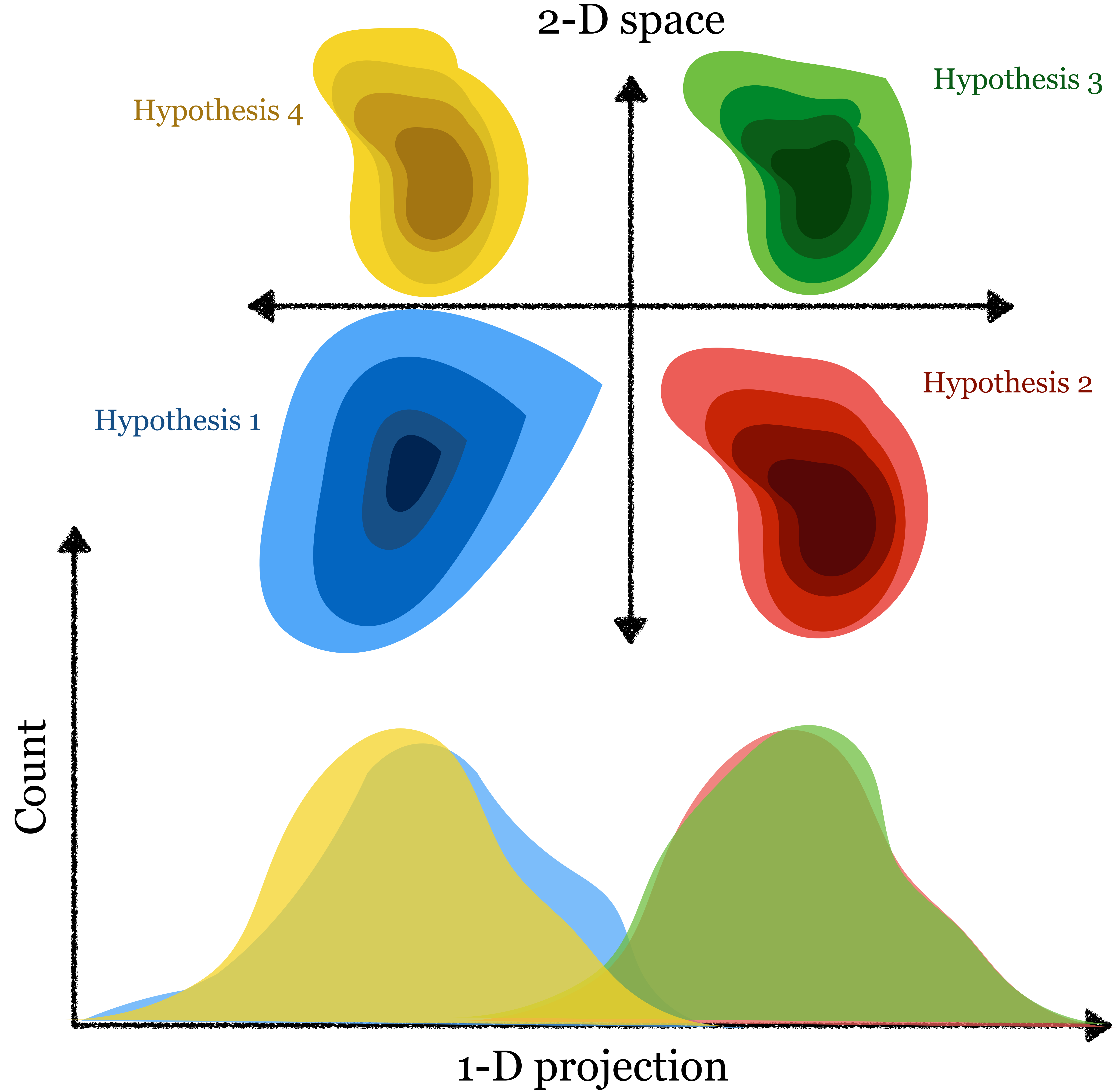
Can we always find a sufficient 1-D summary?



Can we always find a sufficient 1-D summary?



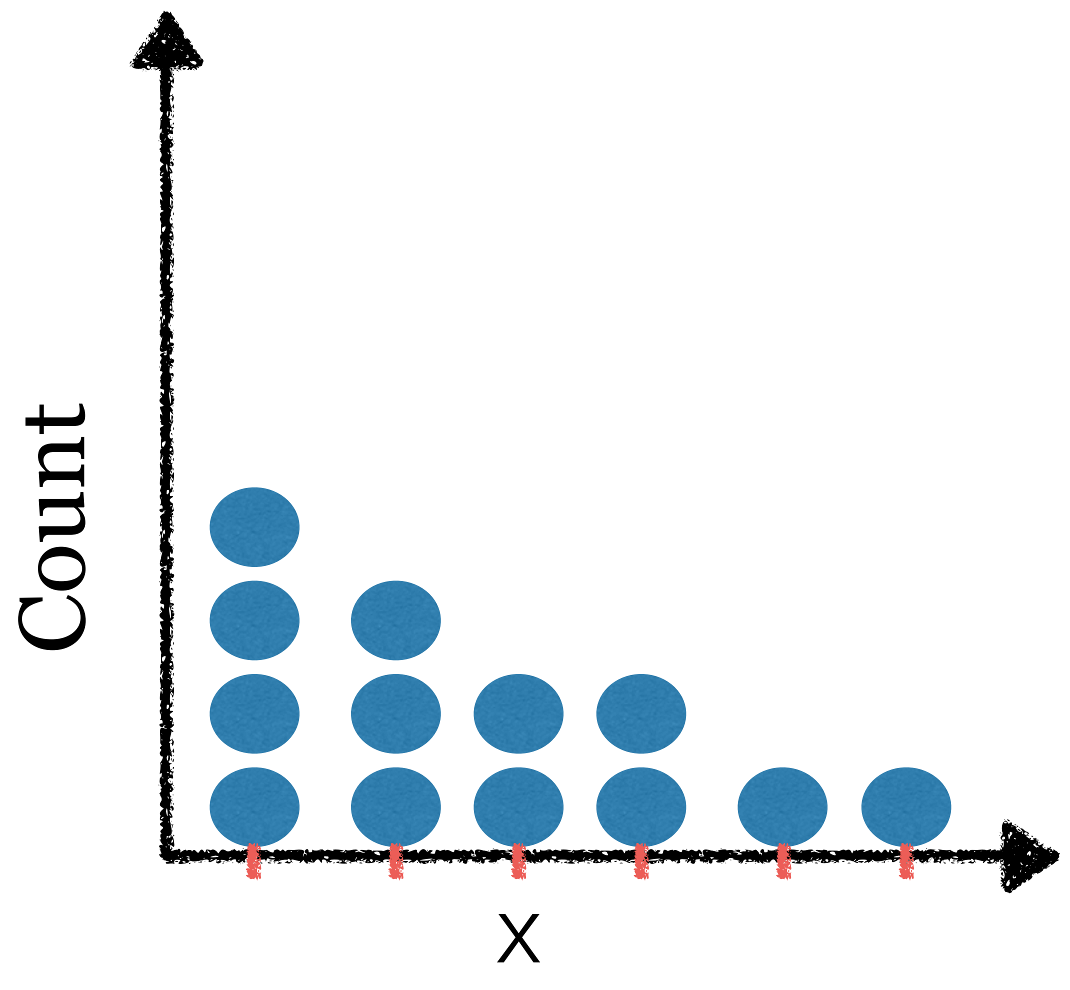
Can we always find a sufficient 1-D summary?



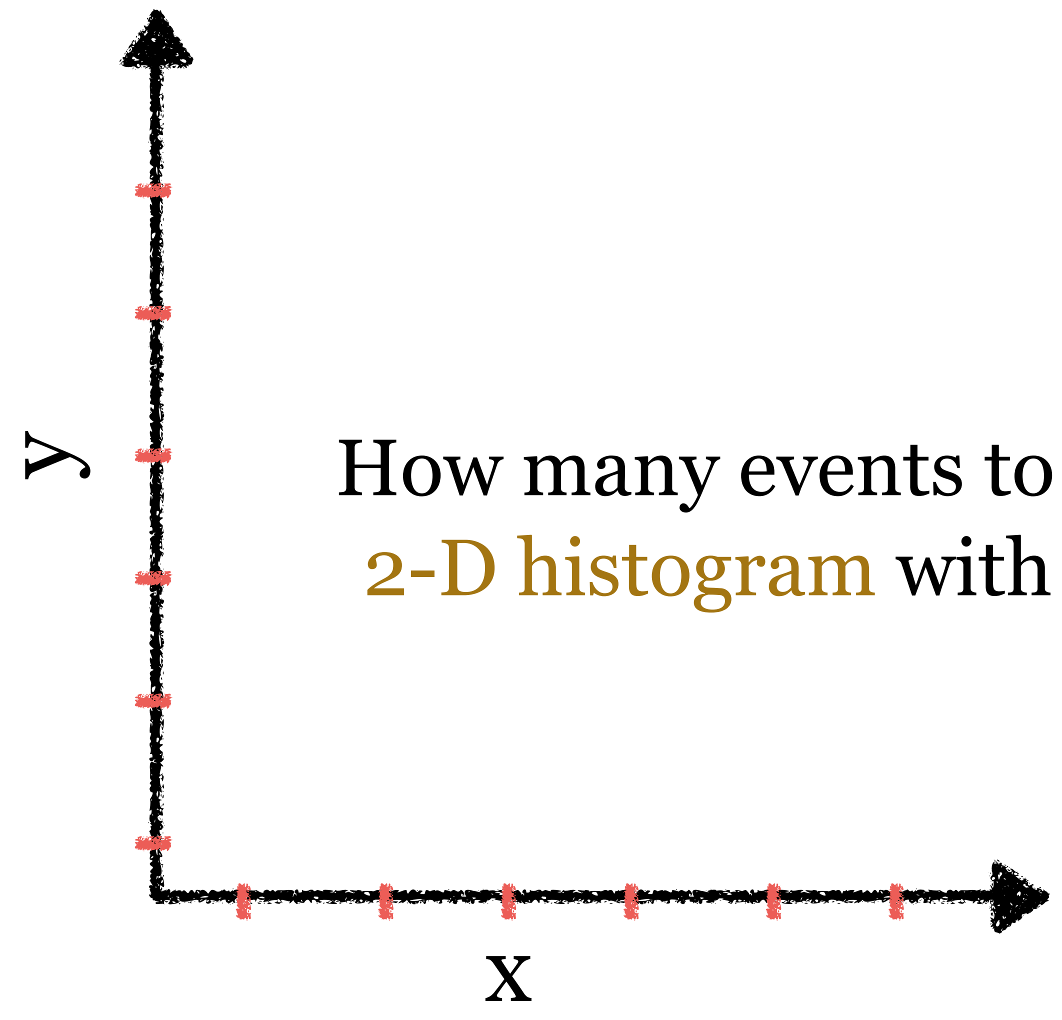
- Clearly separable in 2-D

- A 1-D sufficient summary statistic does not exist for several important particle physics measurements
  - Higgs Width: hal-02971995(p172): Ghosh et al
  - Higgs Self-coupling: arXiv:2507.02032: Ghosh et al
  - Systematic uncertainties: PRD 104.056026: Ghosh et al
  - Effective Field Theories: PRD 98.052004: Brehmer et al

But probability density estimation in higher dimensions is hard…



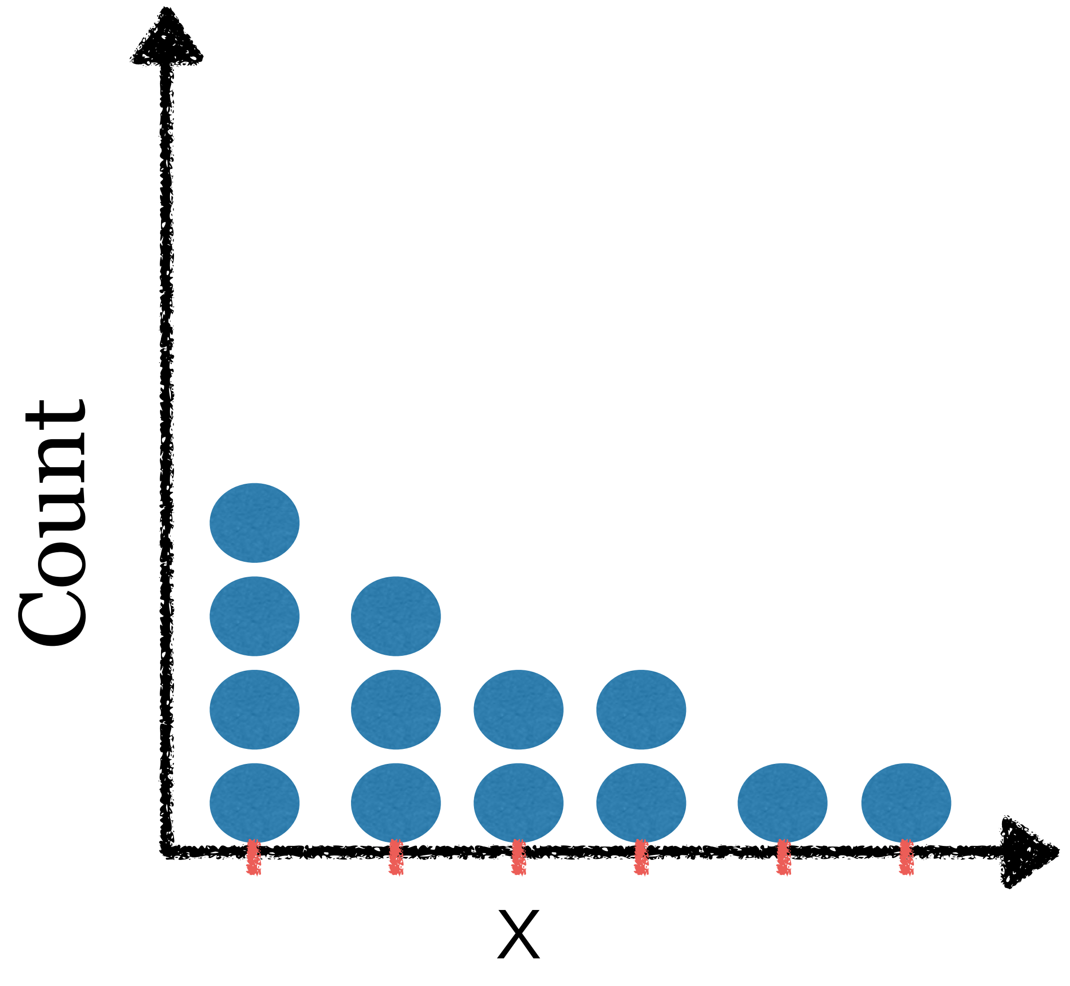
1-D histogram with 6 bins: few events enough to populate it



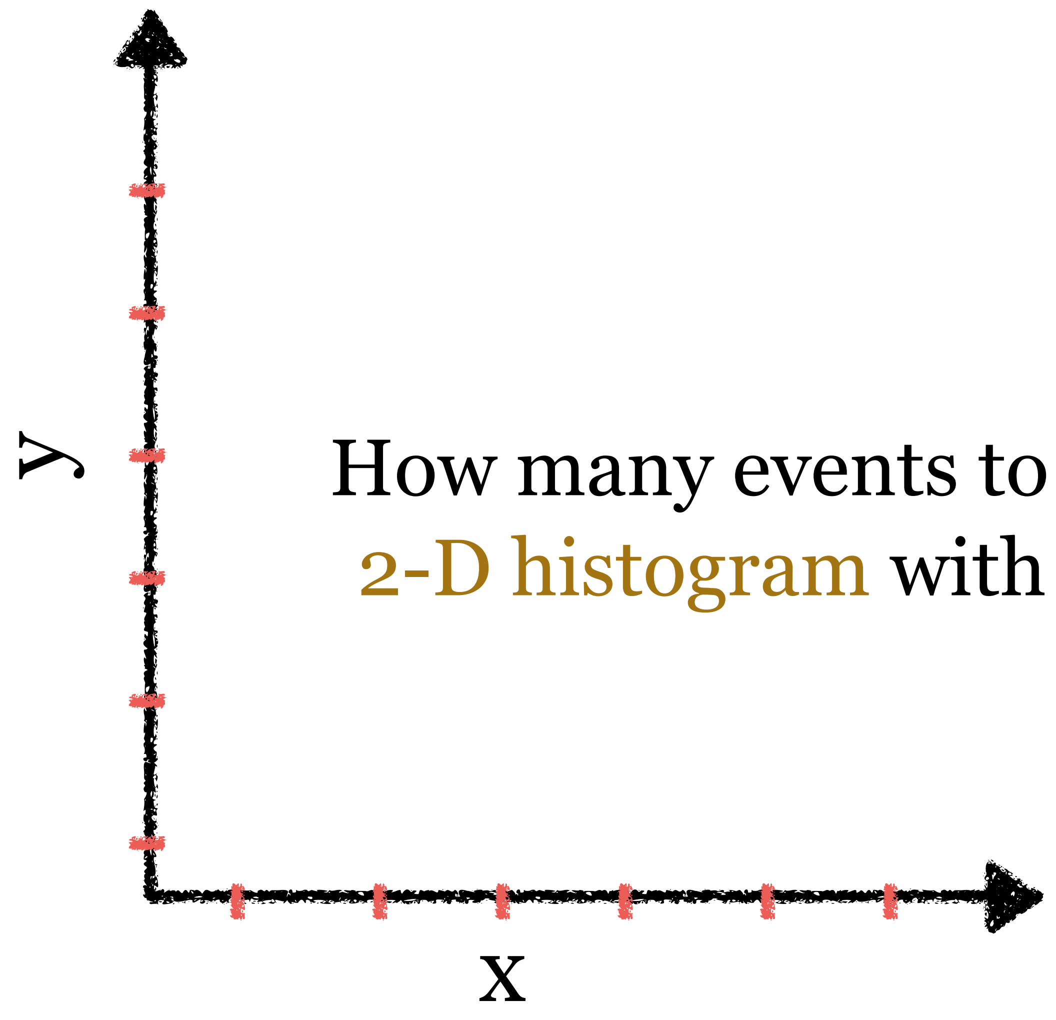
How many events to populate 2-D histogram with 6^2 bins ?

How many events for 50-D histogram with 6^{50} bins ?

But probability density estimation in higher dimensions is hard…



1-D histogram with 6 bins: few events enough to populate it



How many events to populate 2-D histogram with 6^2 bins ?

Curse of dimensionality

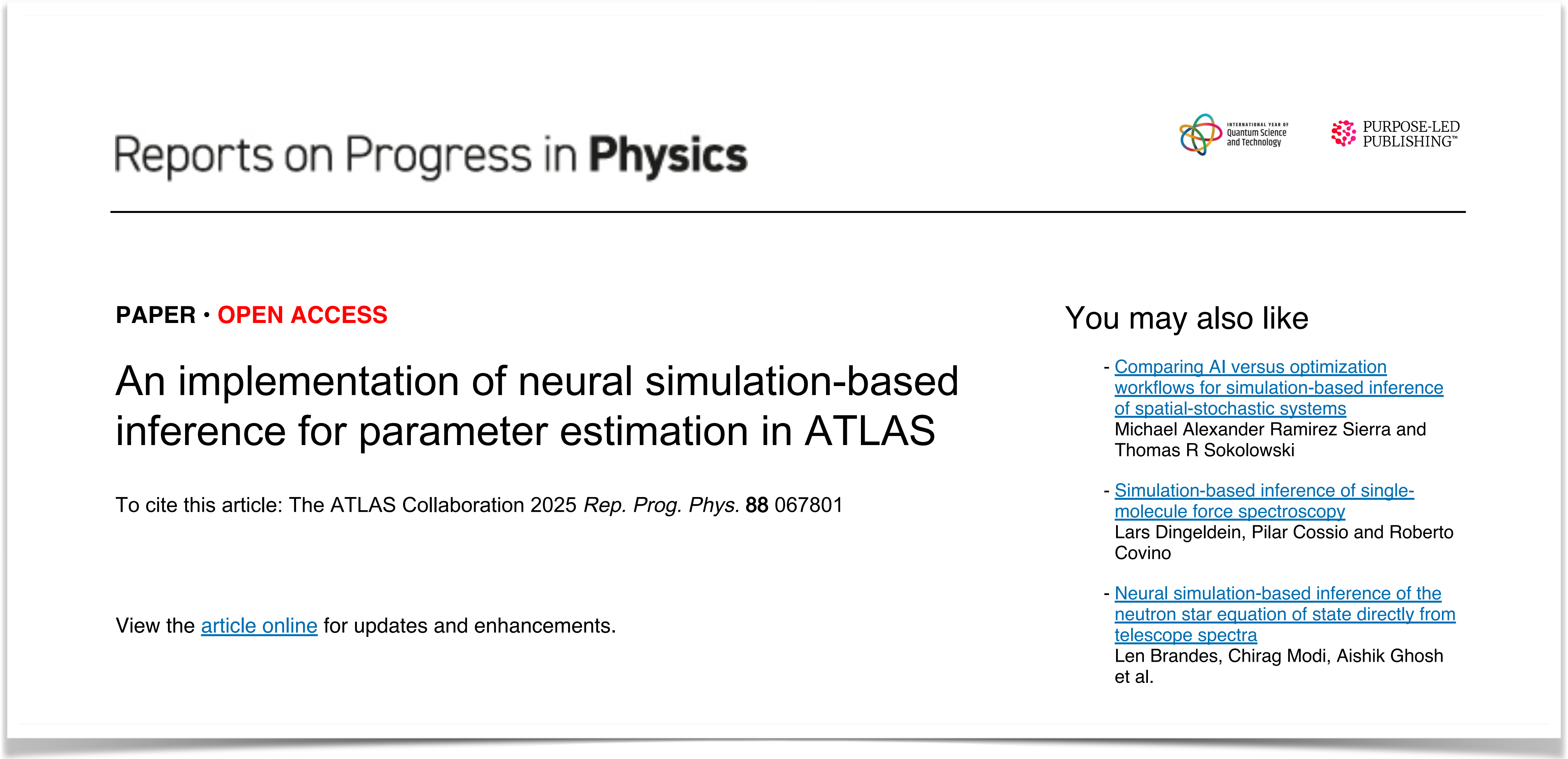
How many events for 50-D histogram with 6^{50} bins ?

Open problems to apply to Large Hadron Collider data:

- Robustness: Design and validation
- Systematic Uncertainties: Propagate them through network
- Interpretation of test statistic in high dimensions

Open problems to apply to Large Hadron Collider data:

Solved!



Reports on Progress in **Physics**

PURPOSE-LED PUBLISHING™

PAPER · **OPEN ACCESS**

An implementation of neural simulation-based inference for parameter estimation in ATLAS

To cite this article: The ATLAS Collaboration 2025 *Rep. Prog. Phys.* **88** 067801

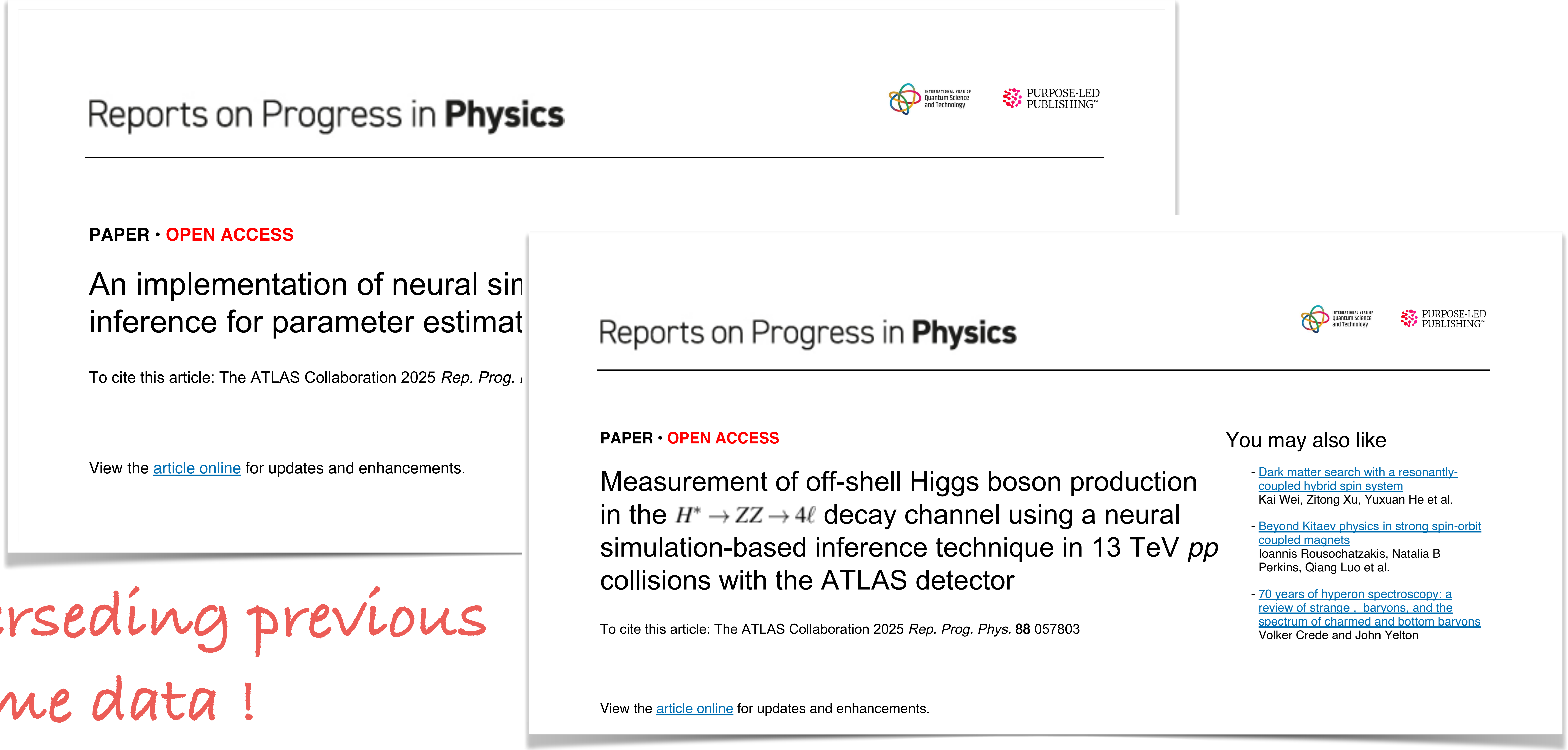
View the article online for updates and enhancements.

You may also like

- Comparing AI versus optimization workflows for simulation-based inference of spatial-stochastic systems
  Michael Alexander Ramirez Sierra and Thomas R Sokolowski
- Simulation-based inference of single-molecule force spectroscopy
  Lars Dingeldein, Pilar Cossio and Roberto Covino
- Neural simulation-based inference of the neutron star equation of state directly from telescope spectra
  Len Brandes, Chirag Modi, Aishik Ghosh et al.

Open problems to apply to Large Hadron Collider data:

Solved!



Reports on Progress in **Physics**

PURPOSE-LED PUBLISHING™

PAPER · **OPEN ACCESS**

An implementation of neural sin
inference for parameter estimat

To cite this article: The ATLAS Collaboration 2025 *Rep. Prog.*

View the article online for updates and enhancements.

Reports on Progress in **Physics**

PURPOSE-LED PUBLISHING™

PAPER · **OPEN ACCESS**

Measurement of off-shell Higgs boson production in the $H^* \to ZZ \to 4\ell$ decay channel using a neural simulation-based inference technique in 13 TeV *pp* collisions with the ATLAS detector

To cite this article: The ATLAS Collaboration 2025 *Rep. Prog. Phys.* **88** 057803

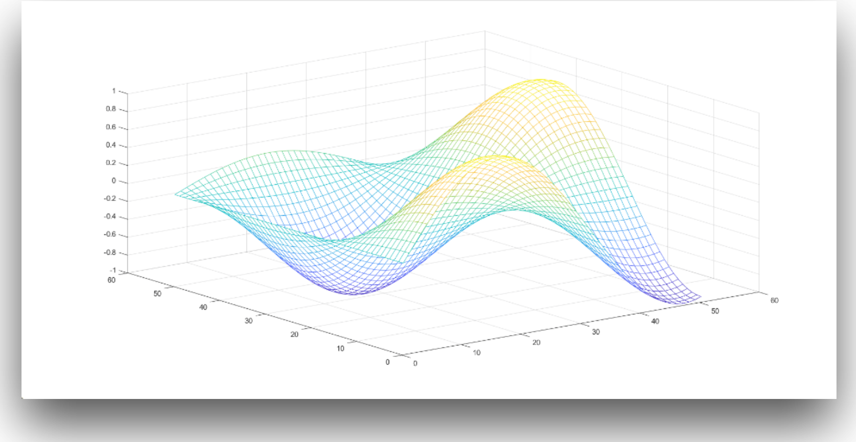
View the article online for updates and enhancements.

You may also like

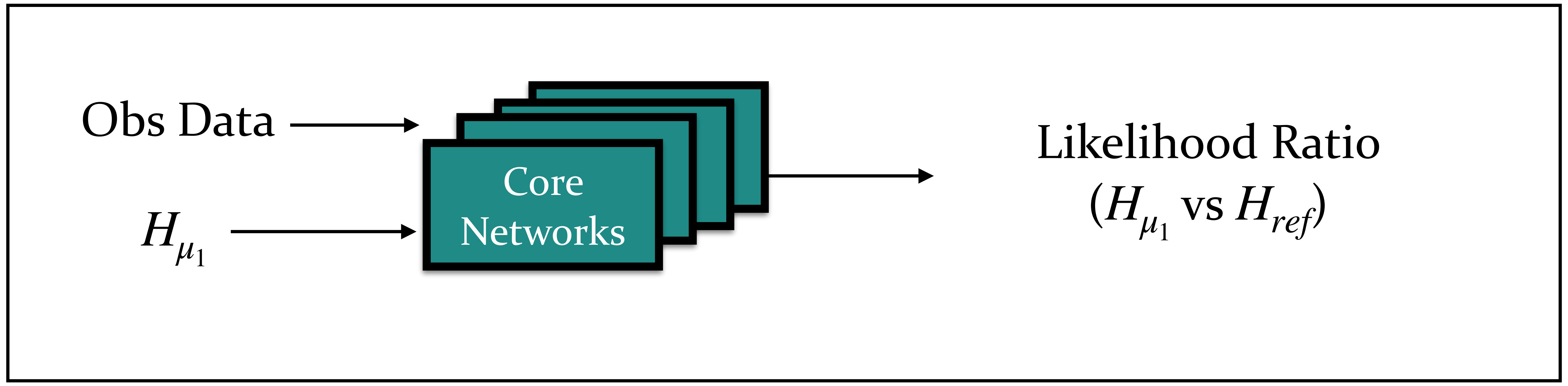
- Dark matter search with a resonantly-coupled hybrid spin system
  Kai Wei, Zitong Xu, Yuxuan He et al.
- Beyond Kitaev physics in strong spin-orbit coupled magnets
  Ioannis Rousochatzakis, Natalia B Perkins, Qiang Luo et al.
- 70 years of hyperon spectroscopy: a review of strange , baryons, and the spectrum of charmed and bottom baryons
  Volker Crede and John Yelton

Applied on Run2 data, superseding previous ATLAS paper on same data !

Combine physics knowledge with AI using 'semi-parametric' approach



$\mathcal{O}(16)$ observables



Obs Data

H_{μ_1}

Core Networks

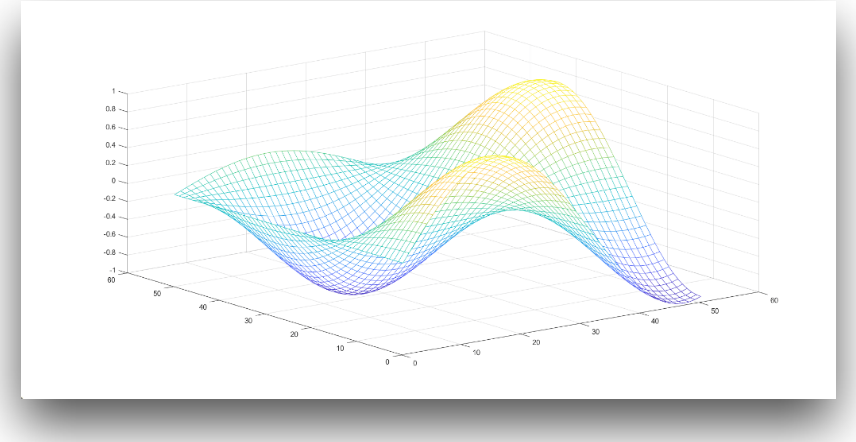
Likelihood Ratio

(H_{μ_1} vs H_{ref})

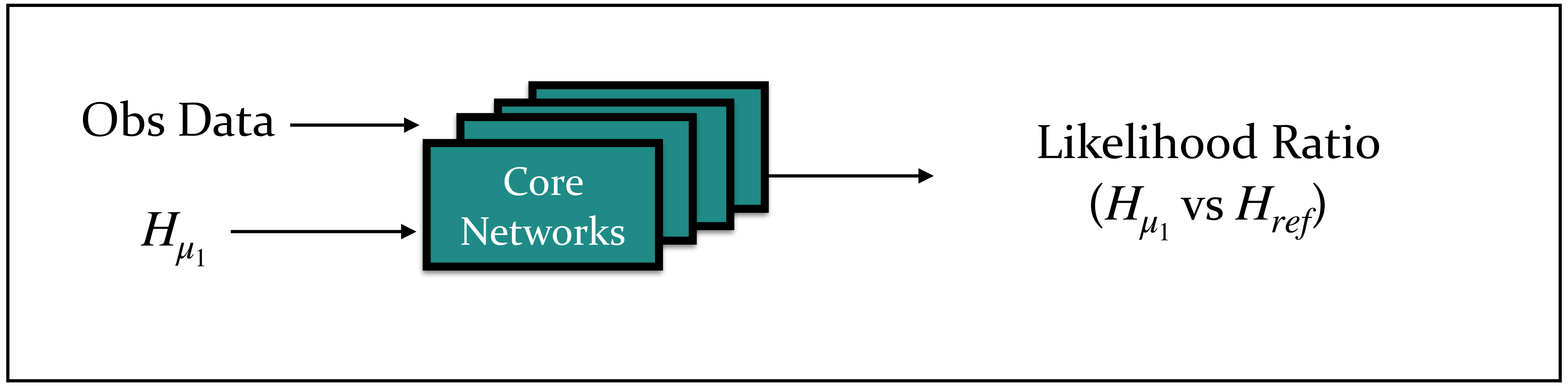
$$\frac{p(x_i \mid \mu, \alpha)}{p_{ref}(x_i)} = \frac{1}{\nu(\mu, \alpha)} \sum_j^C f_j(\mu) \cdot \nu_j \cdot \frac{p_j(x_i)}{p_{ref}(x_i)} \cdot \prod_k^{N_{syst}} G_j(\alpha_k) \cdot \frac{p_j(x_i, \alpha_k)}{p_j(x_i)}$$

ATLAS methods paper (inc. Ghosh): arXiv:2412.01600

Combine physics knowledge with AI using ‘semi-parametric’ approach



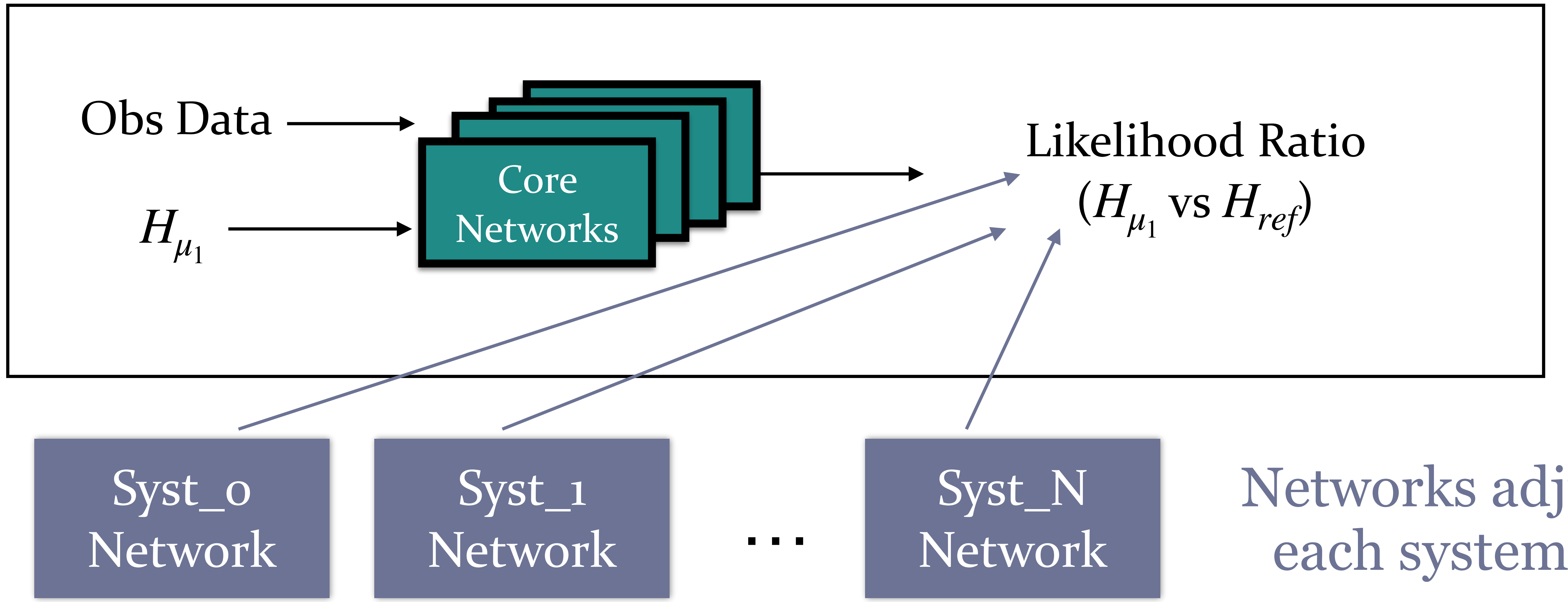
$O(16)$ observables



$$\frac{p(x_i\,|\,\mu,\alpha)}{p_{ref}(x_i)} = \frac{1}{\nu(\mu,\alpha)}\sum_{j}^{C} f_j(\mu)\cdot\nu_j\cdot\boxed{\frac{p_j(x_i)}{p_{ref}(x_i)}}\cdot\prod_{k}^{N_{syst}} G_j(\alpha_k)\cdot\frac{p_j(x_i,\alpha_k)}{p_j(x_i)}$$

ATLAS methods paper (inc. Ghosh): arXiv:2412.01600

Combine physics knowledge with AI using ‘semi-parametric’ approach



Obs Data

H_{μ_1}

Core Networks

Likelihood Ratio (H_{μ_1} vs H_{ref})

Syst_0 Network

Syst_1 Network

…

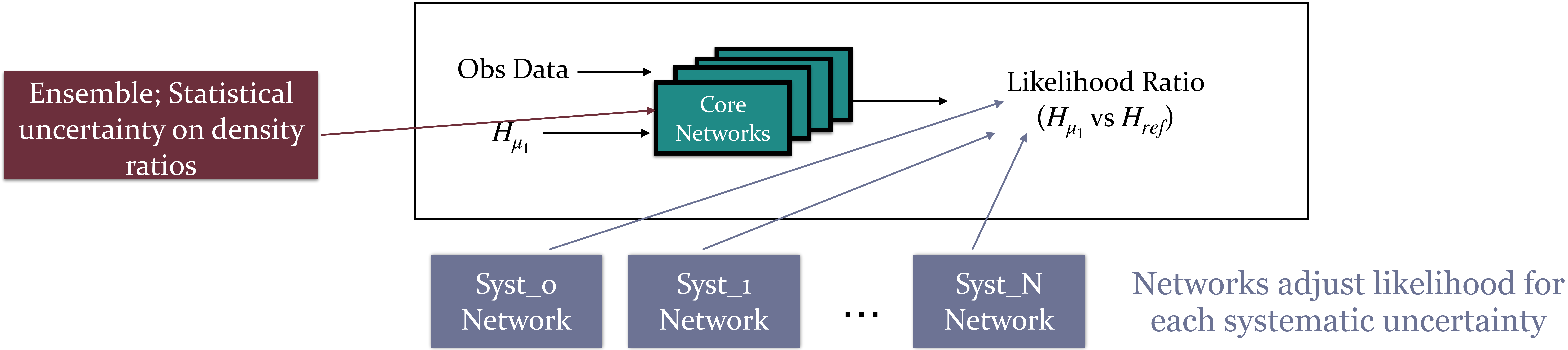
Syst_N Network

Networks adjust likelihood for each systematic uncertainty

$$\frac{p(x_i\,|\,\mu,\alpha)}{p_{ref}(x_i)} = \frac{1}{\nu(\mu,\alpha)}\sum_{j}^{C} f_j(\mu)\cdot\nu_j\cdot\frac{p_j(x_i)}{p_{ref}(x_i)}\cdot\prod_{k}^{N_{syst}} G_j(\alpha_k)\cdot\frac{p_j(x_i,\alpha_k)}{p_j(x_i)}$$

ATLAS methods paper (inc. Ghosh): arXiv:2412.01600

Combine physics knowledge with AI using ‘semi-parametric’ approach



Ensemble; Statistical uncertainty on density ratios

Obs Data

H_{μ_1}

Core Networks

Likelihood Ratio (H_{μ_1} vs H_{ref})

Syst_0 Network

Syst_1 Network

…

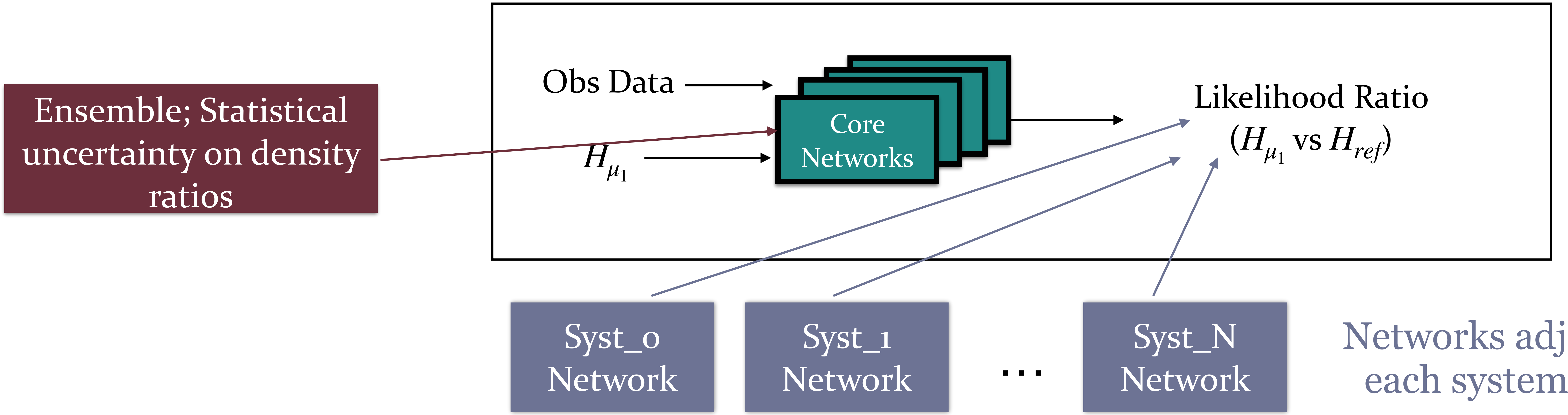
Syst_N Network

Networks adjust likelihood for each systematic uncertainty

$$\frac{p(x_i\,|\,\mu,\alpha)}{p_{ref}(x_i)} = \frac{1}{\nu(\mu,\alpha)} \sum_j^C f_j(\mu)\cdot\nu_j\cdot\frac{p_j(x_i)}{p_{ref}(x_i)}\cdot\prod_k^{N_{syst}} G_j(\alpha_k)\cdot\frac{p_j(x_i,\alpha_k)}{p_j(x_i)}$$

ATLAS methods paper (inc. Ghosh): arXiv:2412.01600

Combine physics knowledge with AI using ‘semi-parametric’ approach



Ensemble; Statistical uncertainty on density ratios

Obs Data

H_{μ_1}

Core Networks

Likelihood Ratio (H_{μ_1} vs H_{ref})

Syst_0 Network

Syst_1 Network

…

Syst_N Network

Networks adjust likelihood for each systematic uncertainty

$$\frac{p(x_i \,|\, \mu, \alpha)}{p_{ref}(x_i)} = \frac{1}{\nu(\mu, \alpha)} \sum_{j}^{C} f_j(\mu) \cdot \nu_j \cdot \boxed{\frac{p_j(x_i)}{p_{ref}(x_i)}} \cdot \prod_{k}^{N_{syst}} G_j(\alpha_k) \cdot \boxed{\frac{p_j(x_i, \alpha_k)}{p_j(x_i)}}$$

Training details

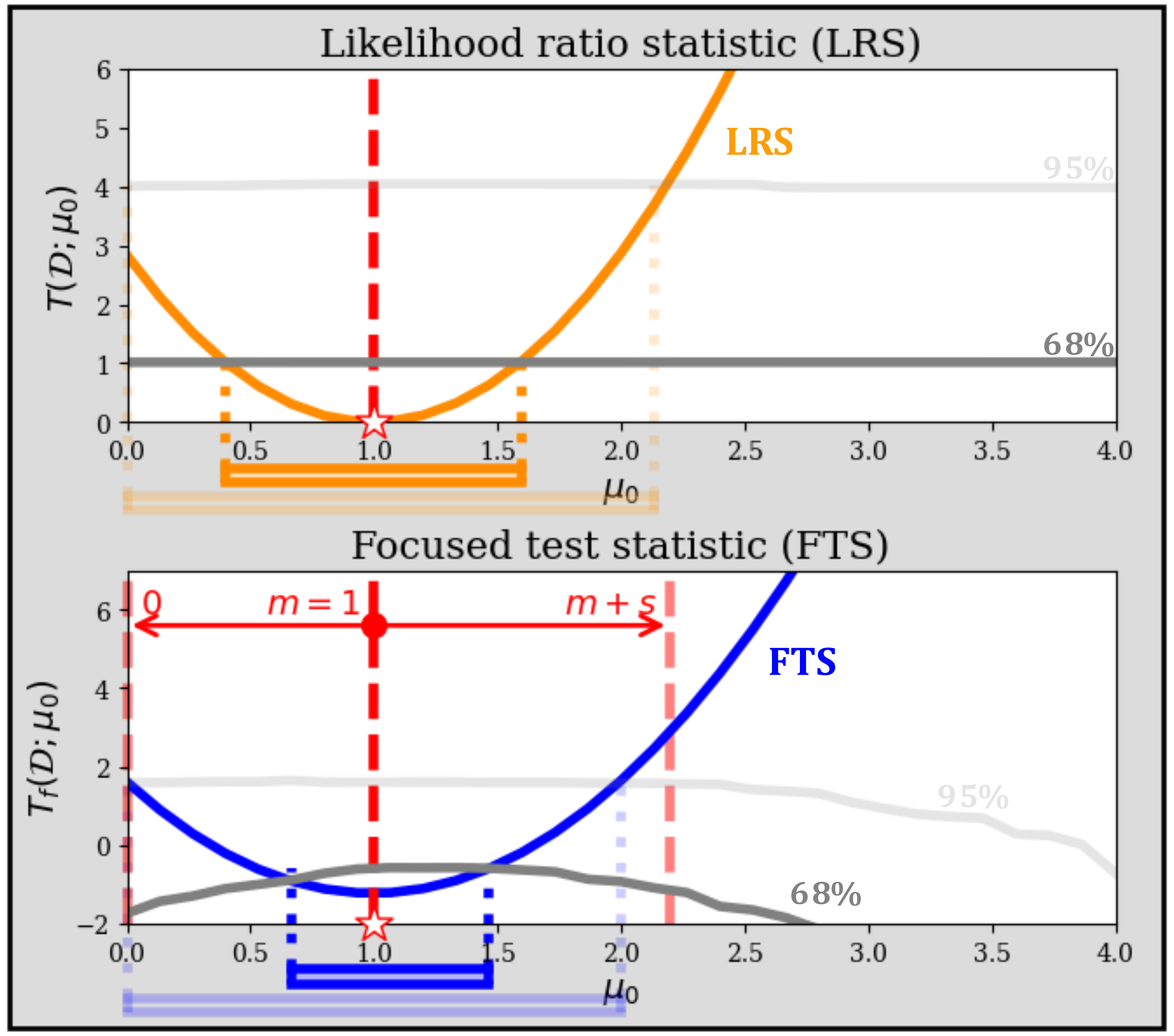
- ✦ Train $O(10^4)$ networks on TensorFlow
- ✦ Computing resources provided by Google, SMU, other HPC clusters
- ✦ Fits with JAX

arXiv:2507.17831: Carzon, Ghosh, Izbicki, Lee, et al

Challenging a deeply-held belief in particle physics

$$LRS(\mathscr{D};\mu_0) = -2\log\left(\frac{p(\mathscr{D}\,|\,\mu_0)}{\sup_{\mu\in\Theta} p(\mathscr{D}\,|\,\mu)}\right) \quad \longrightarrow \quad FTS(\mathscr{D};\mu_0) = -2\log\left(\frac{p(\mathscr{D}\,|\,\mu_0)}{\int_\Theta p(\mathscr{D}\,|\,\mu)f(\mu)d\mu}\right)$$

- Denominator in ‘Focused Test Statistic’ (FTS) knows about all alternate hypotheses

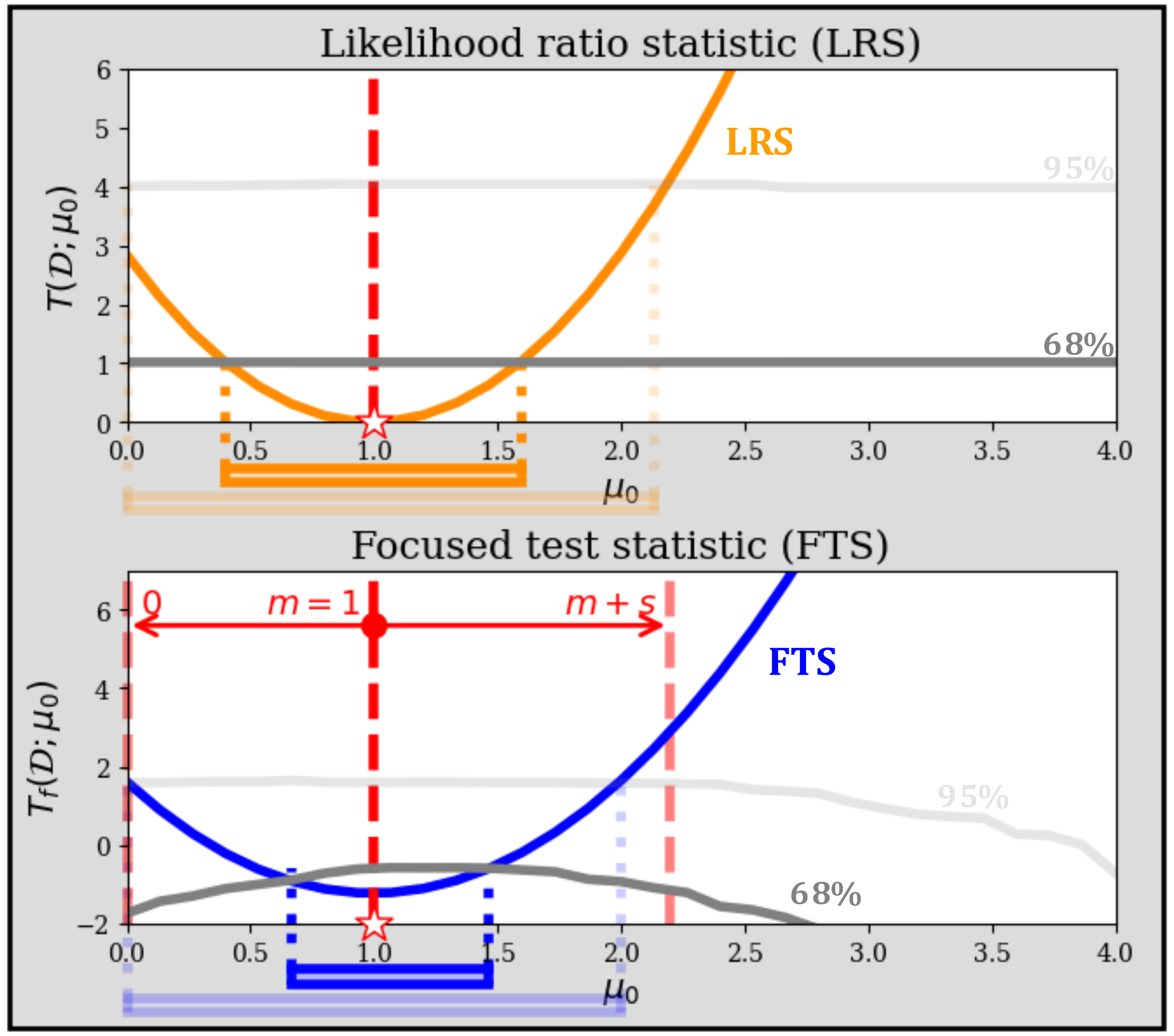


arXiv:2507.17831: Carzon, Ghosh, Izbicki, Lee, et al

Challenging a deeply-held belief in particle physics

$$LRS(\mathscr{D};\mu_0) = -2\log\left(\frac{p(\mathscr{D}\,|\,\mu_0)}{\sup_{\mu\in\Theta} p(\mathscr{D}\,|\,\mu)}\right) \quad \Longrightarrow \quad FTS(\mathscr{D};\mu_0) = -2\log\left(\frac{p(\mathscr{D}\,|\,\mu_0)}{\int_\Theta p(\mathscr{D}\,|\,\mu)f(\mu)d\mu}\right)$$

- Denominator in ‘Focused Test Statistic’ (FTS) knows about all alternate hypotheses

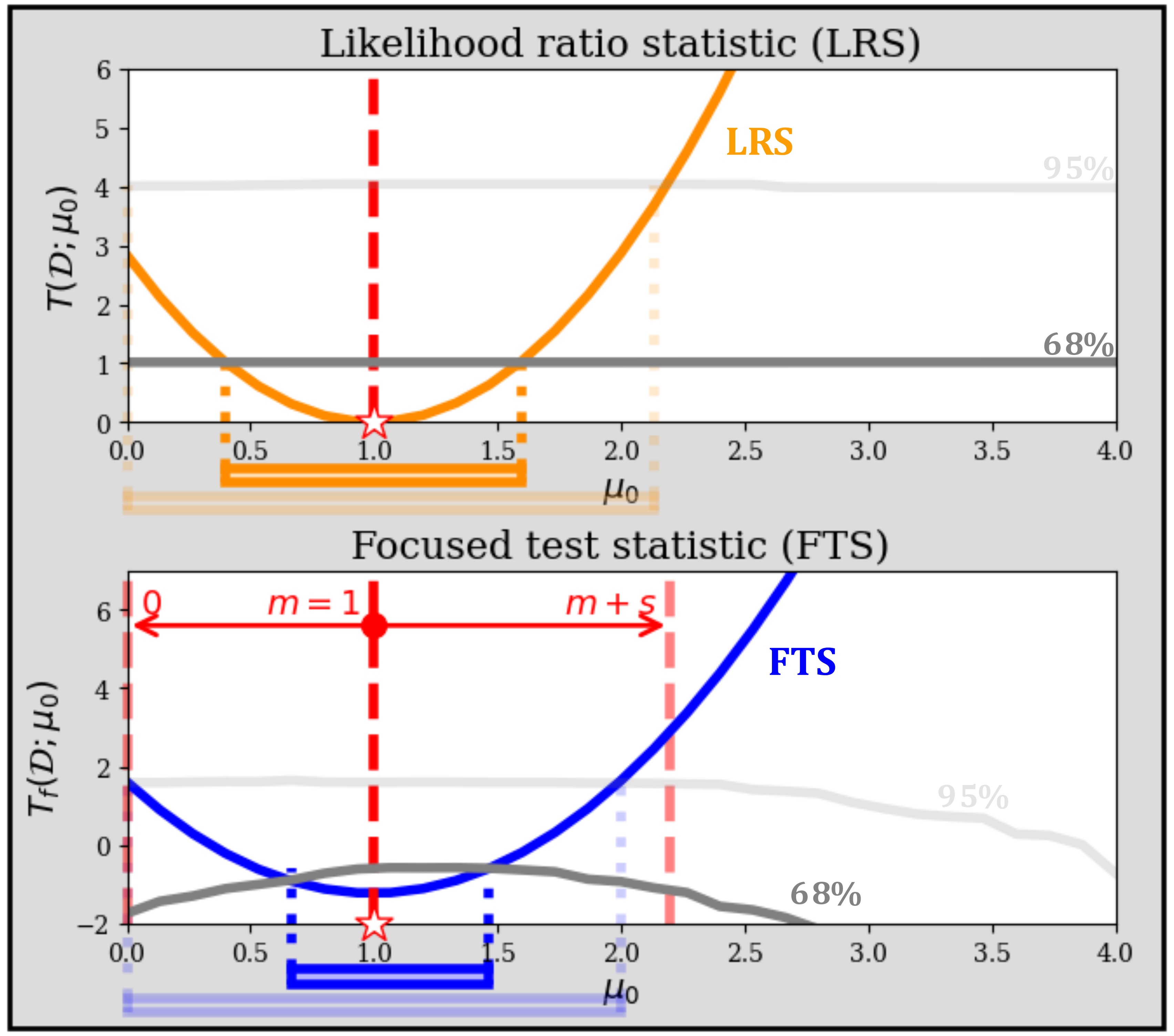


Shorter median length for confidence intervals with FTS even in ‘asymptotic regime’ where Wilks’ applies

Challenging a deeply-held belief in particle physics

$$LRS(\mathscr{D};\mu_0) = -2\log\left(\frac{p(\mathscr{D}\,|\,\mu_0)}{\sup_{\mu\in\Theta} p(\mathscr{D}\,|\,\mu)}\right) \longrightarrow FTS(\mathscr{D};\mu_0) = -2\log\left(\frac{p(\mathscr{D}\,|\,\mu_0)}{\int_\Theta p(\mathscr{D}\,|\,\mu)f(\mu)d\mu}\right)$$

- Denominator in ‘Focused Test Statistic’ (FTS) knows about all alternate hypotheses
- $f(\mu)$ focuses statistical power in meaningful regions of parameter space
    - Particularly useful in small sample / small signal regime
- Fast critical value estimation with ML



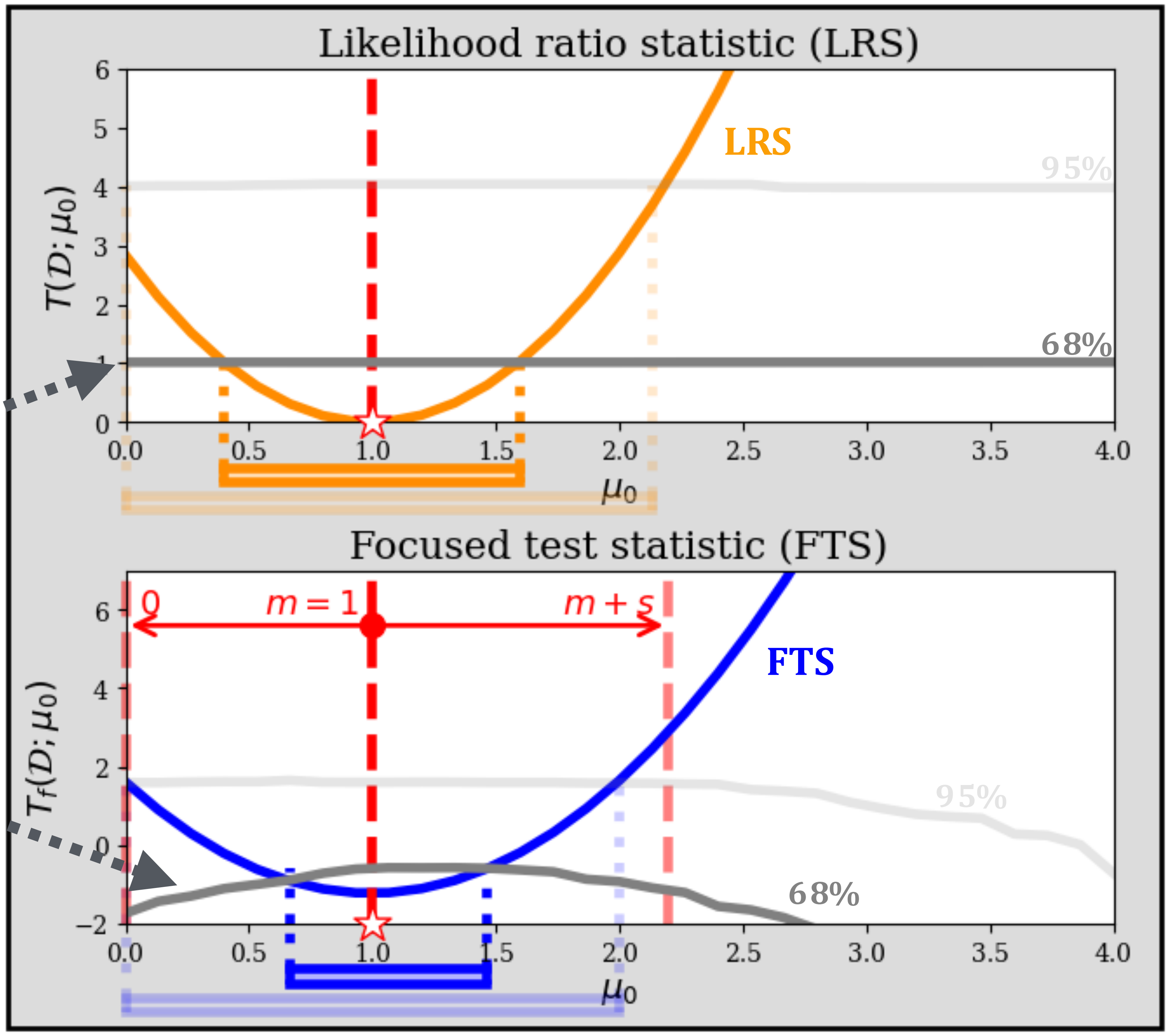
Shorter median length for confidence intervals with FTS even in ‘asymptotic regime’ where Wilks’ applies

[arXiv:2507.17831](https://arxiv.org/abs/2507.17831): Carzon, Ghosh, Izbicki, Lee, et al

Challenging a deeply-held belief in particle physics

$$LRS(\mathscr{D};\mu_0) = -2\log\left(\frac{p(\mathscr{D}\,|\,\mu_0)}{\sup_{\mu\in\Theta} p(\mathscr{D}\,|\,\mu)}\right) \longrightarrow FTS(\mathscr{D};\mu_0) = -2\log\left(\frac{p(\mathscr{D}\,|\,\mu_0)}{\int_\Theta p(\mathscr{D}\,|\,\mu)f(\mu)d\mu}\right)$$

- Denominator in ‘Focused Test Statistic’ (FTS) knows about all alternate hypotheses
- $f(\mu)$ focuses statistical power in meaningful regions of parameter space
    - Particularly useful in small sample / small signal regime
- Fast critical value estimation with ML



Shorter median length for confidence intervals with FTS even in ‘asymptotic regime’ where Wilks’ applies

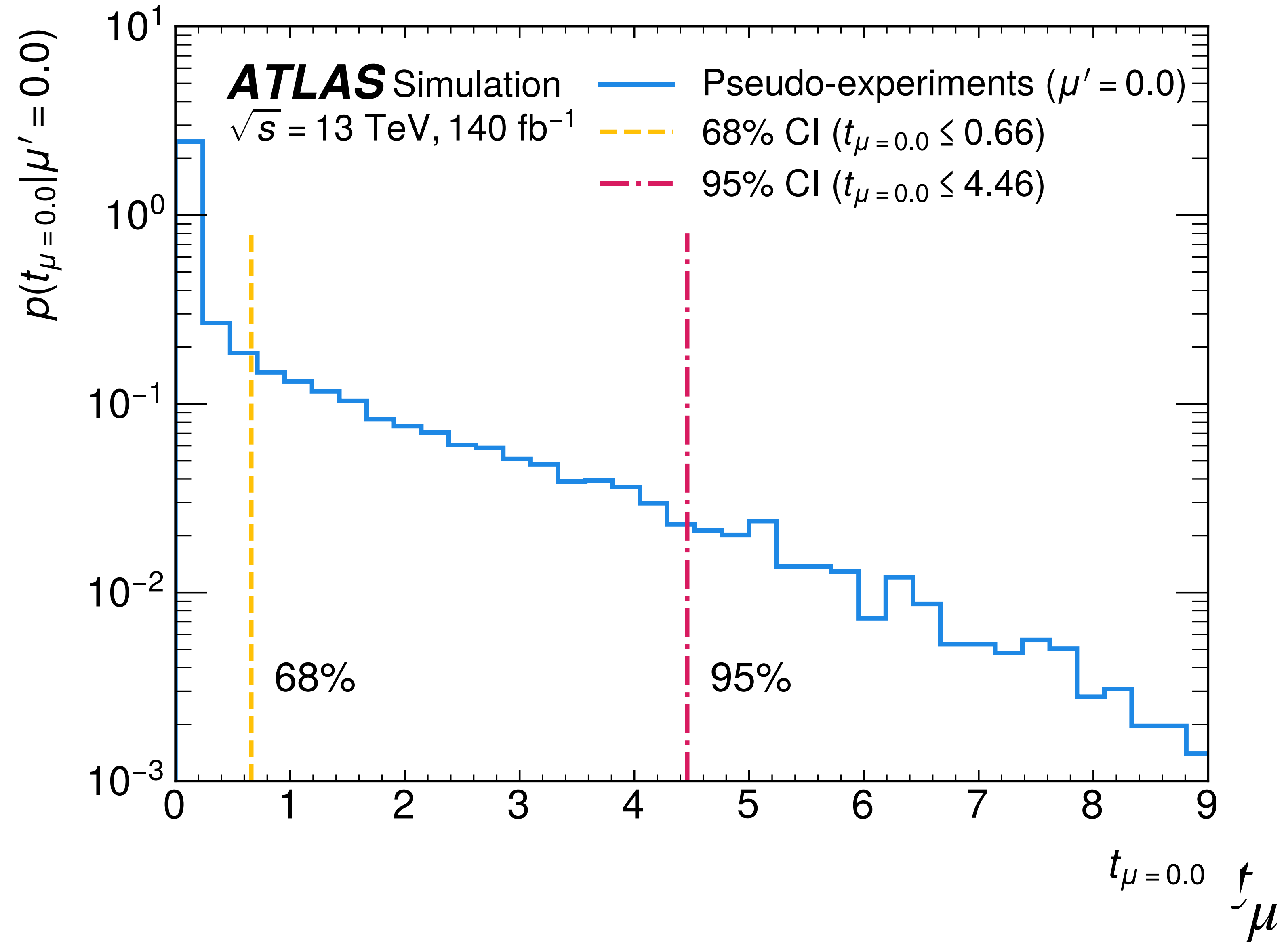
ATLAS NSBI paper: [arXiv:2412.01600](https://arxiv.org/abs/2412.01600)

Computational Challenge: Inverting the test

- Determine 68 % & 95 % CI empirically from this distribution
- Do it for each value of μ

Distribution of test statistic t_μ over thousands of simulated pseudo-experiments

True $\mu = 0$



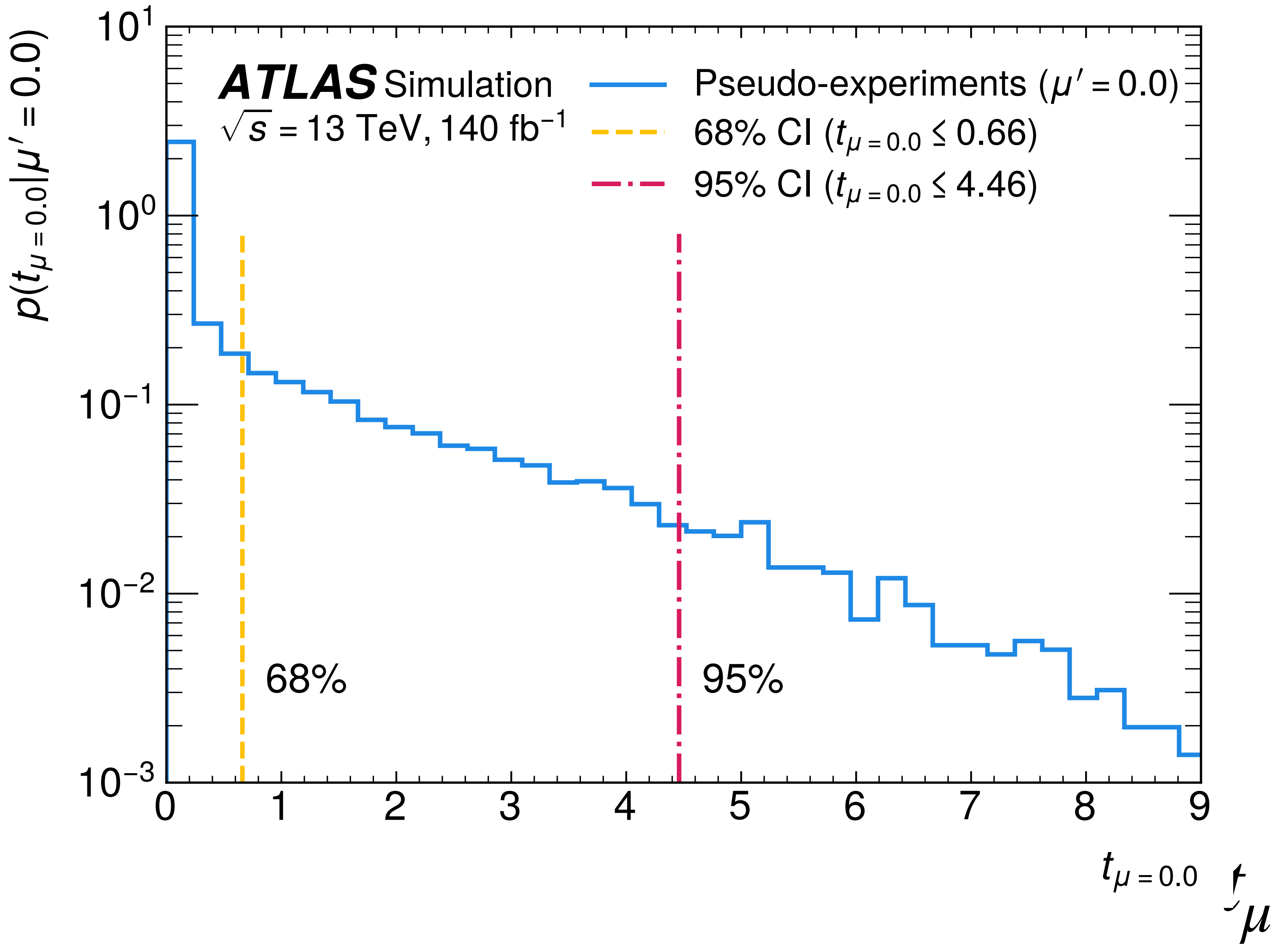
t_μ

Computational Challenge: Inverting the test

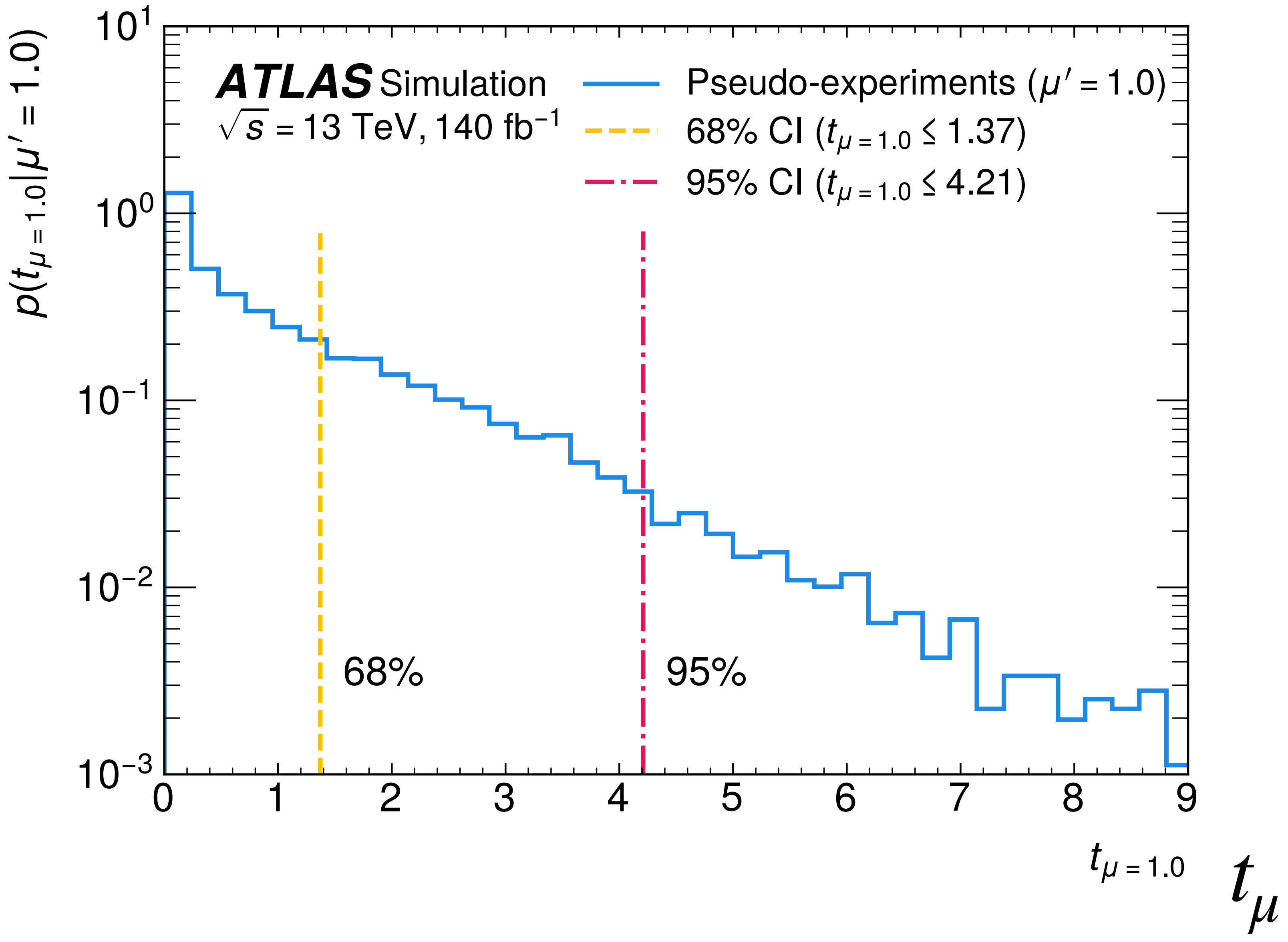
- Determine 68 % & 95 % CI empirically from this distribution
- Do it for each value of μ

Distribution of test statistic t_μ over thousands of simulated pseudo-experiments

True $\mu = 0$



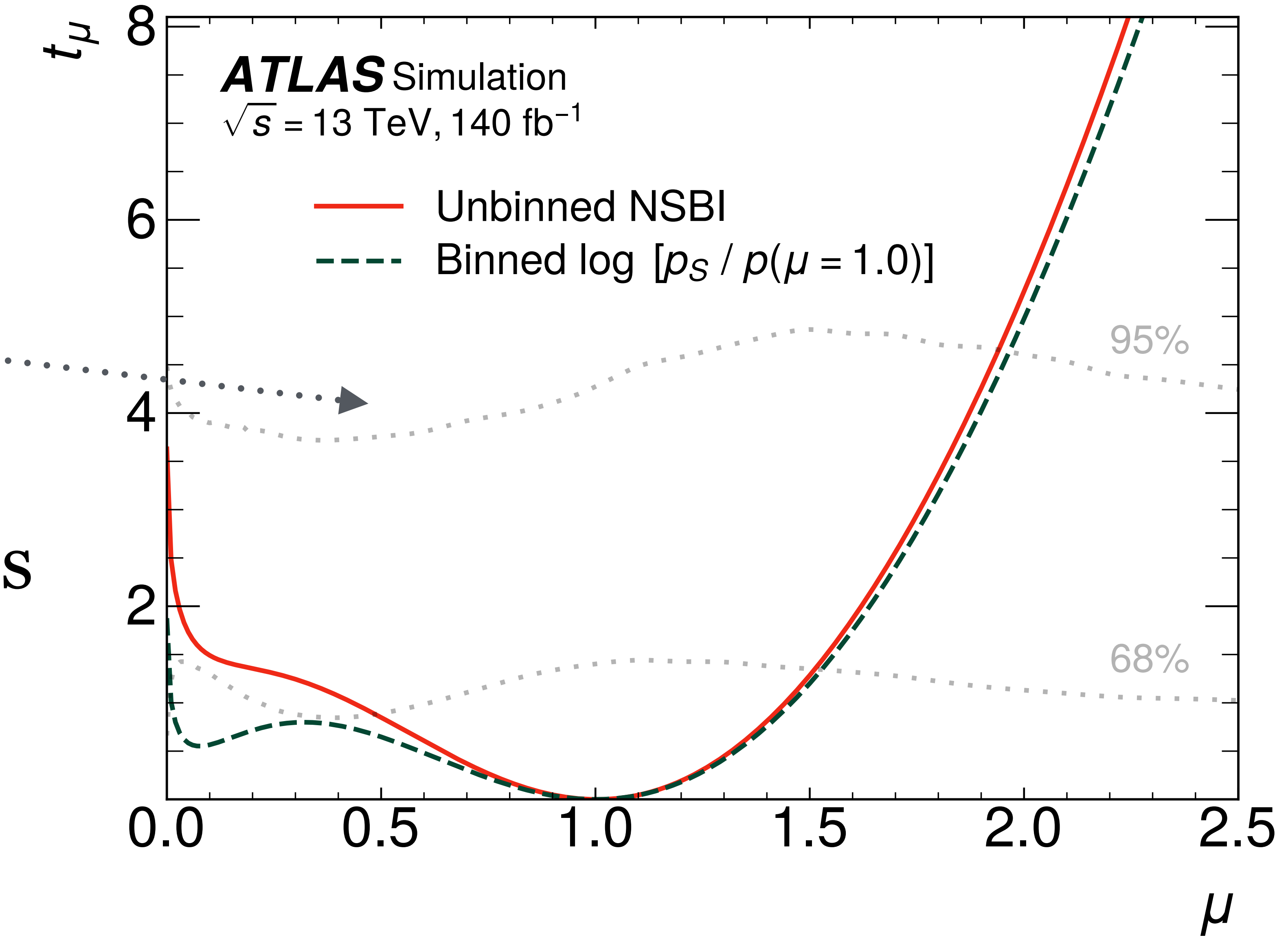
True $\mu = 1$



ATLAS NSBI paper: arXiv:2412.01600

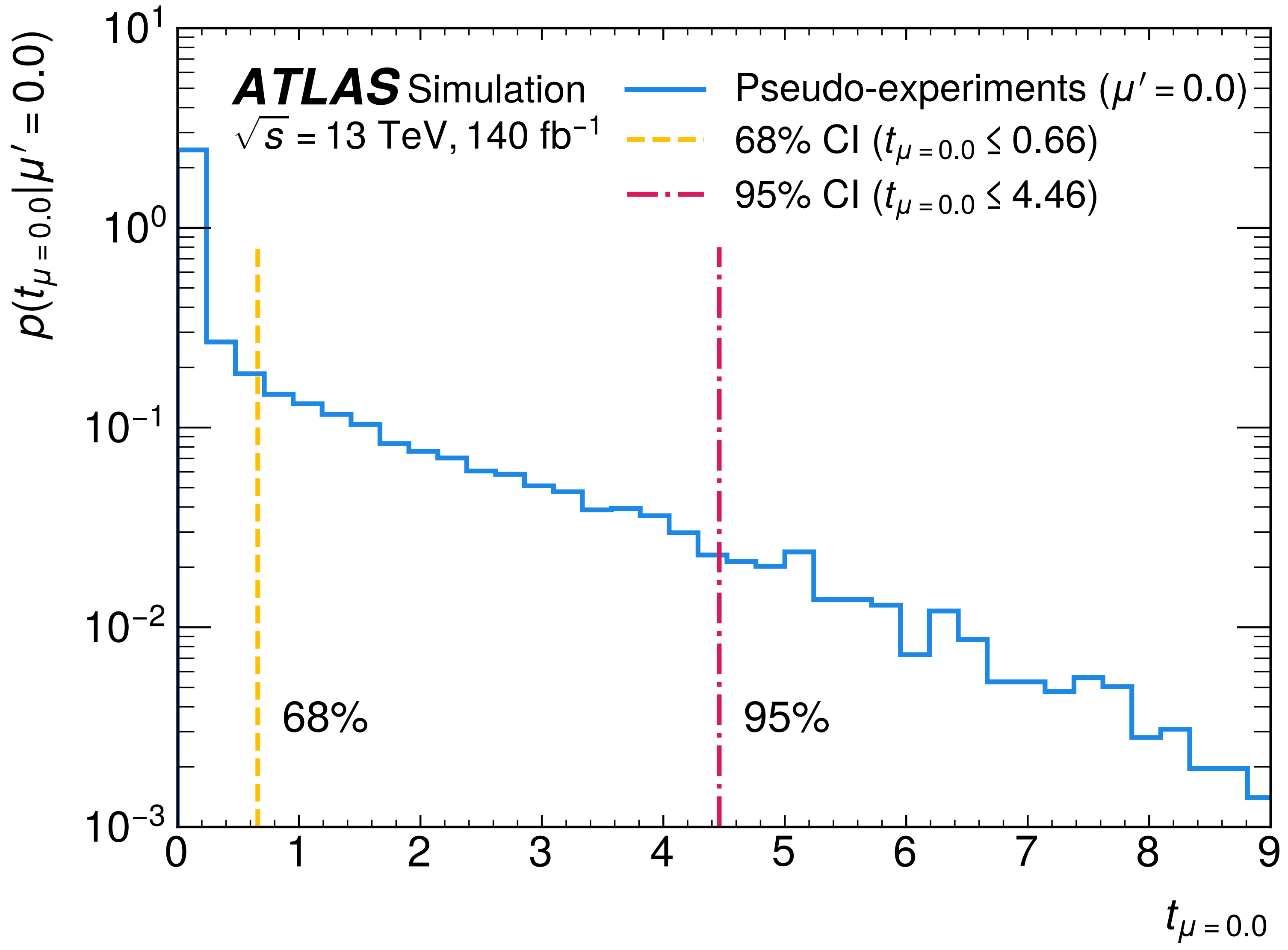
Computational Challenge: Inverting the test

- Determine 68 % & 95 % CI empirically from this distribution
- Do it for each value of μ



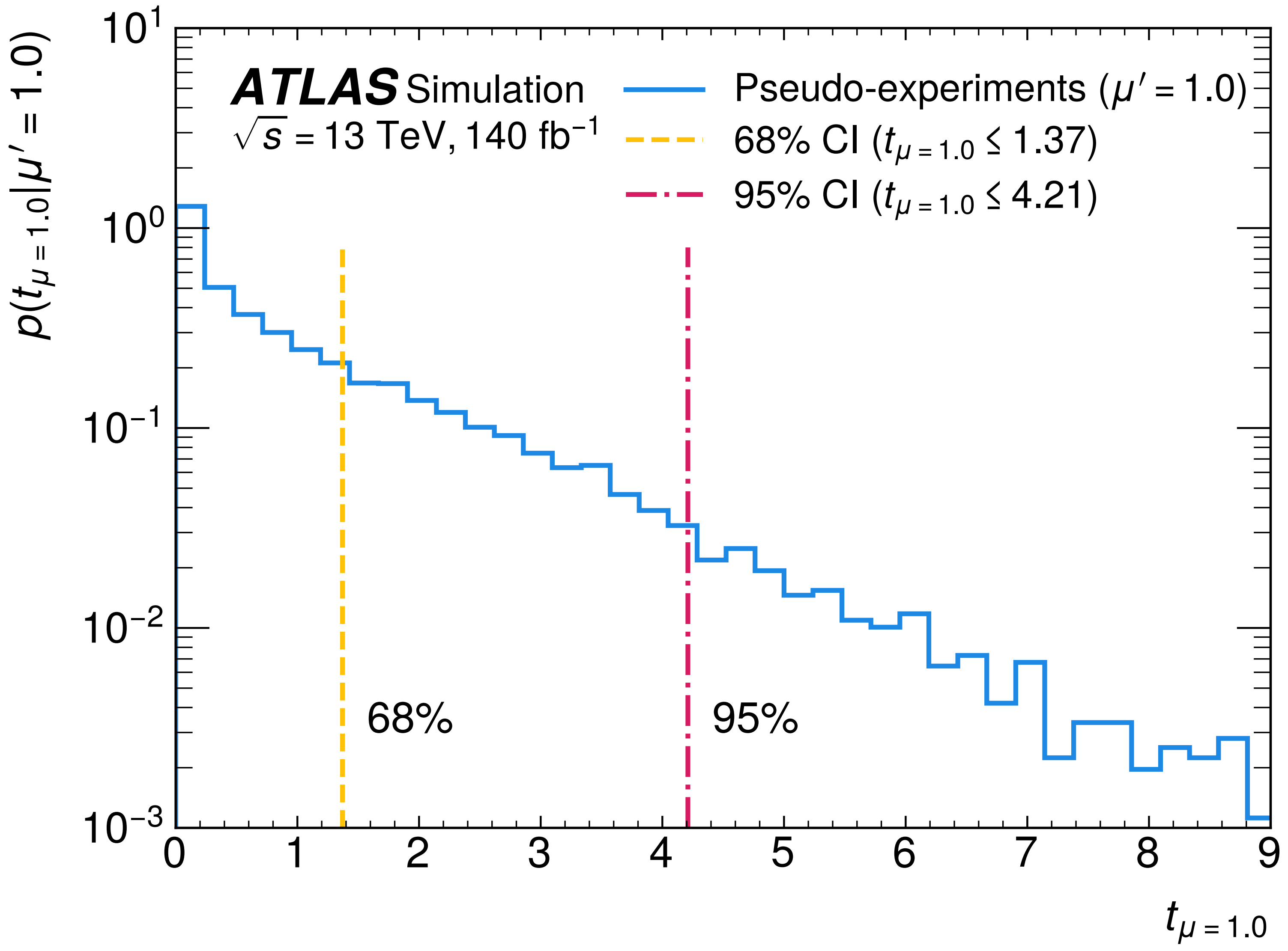
Distribution of test statistic t_μ over thousands of simulated pseudo-experiments

True $\mu = 0$



t_μ

True $\mu = 1$

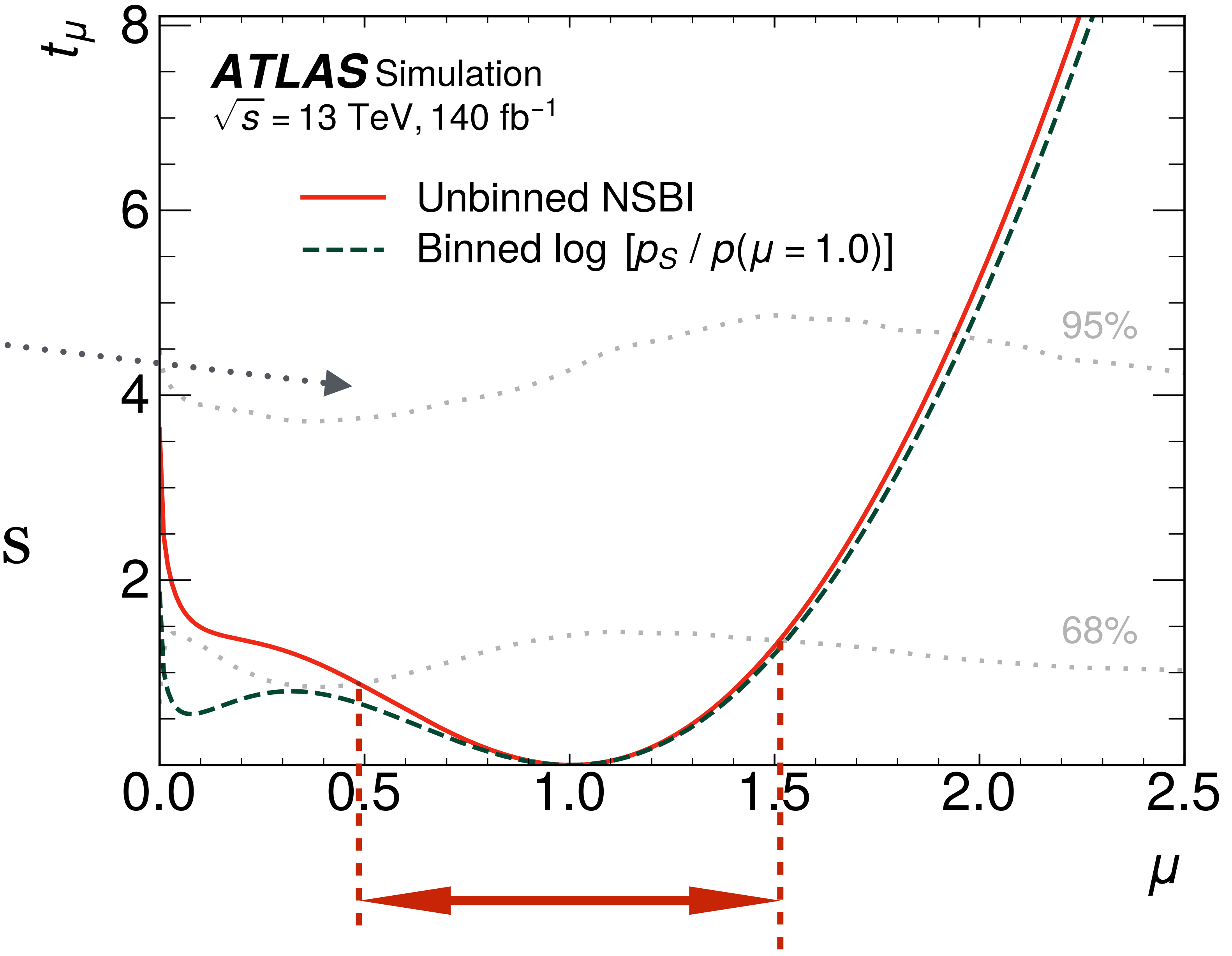


t_μ

ATLAS NSBI paper: [arXiv:2412.01600](https://arxiv.org/abs/2412.01600)

Computational Challenge: Inverting the test

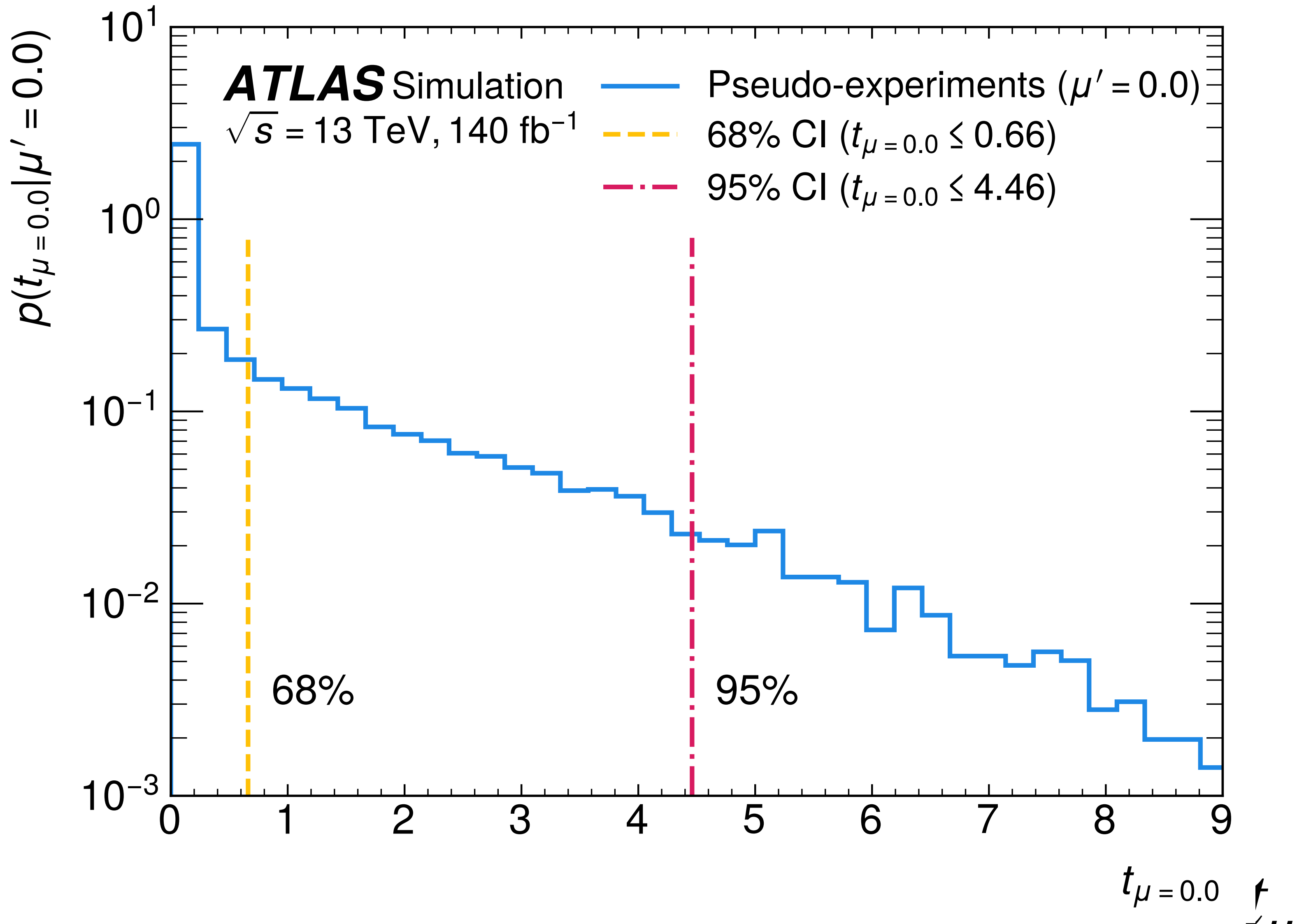
- Determine 68 % & 95 % CI empirically from this distribution
- Do it for each value of μ



Calibrated confidence interval

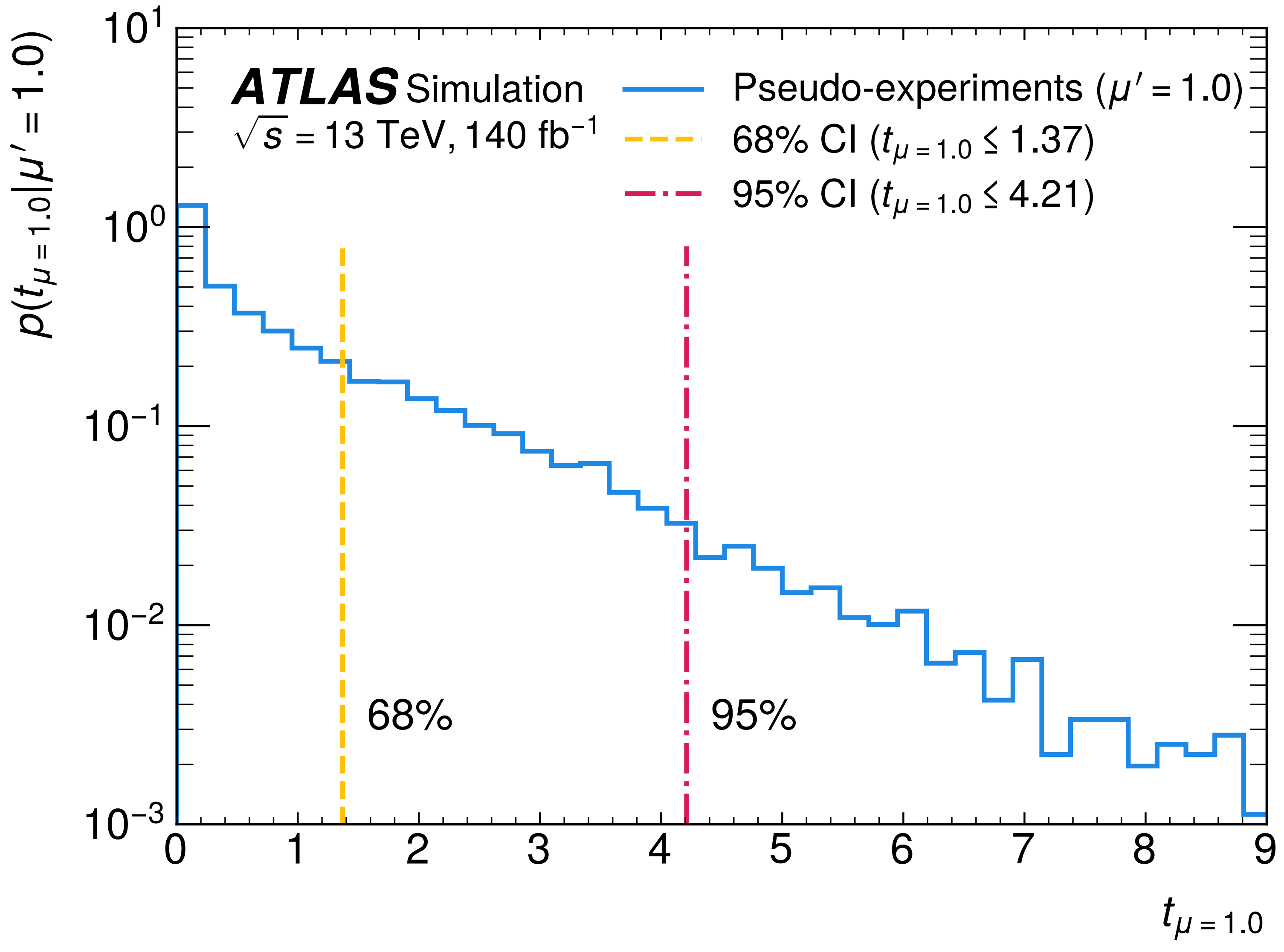
Distribution of test statistic t_μ over thousands of simulated pseudo-experiments

True $\mu = 0$



t_μ

True $\mu = 1$

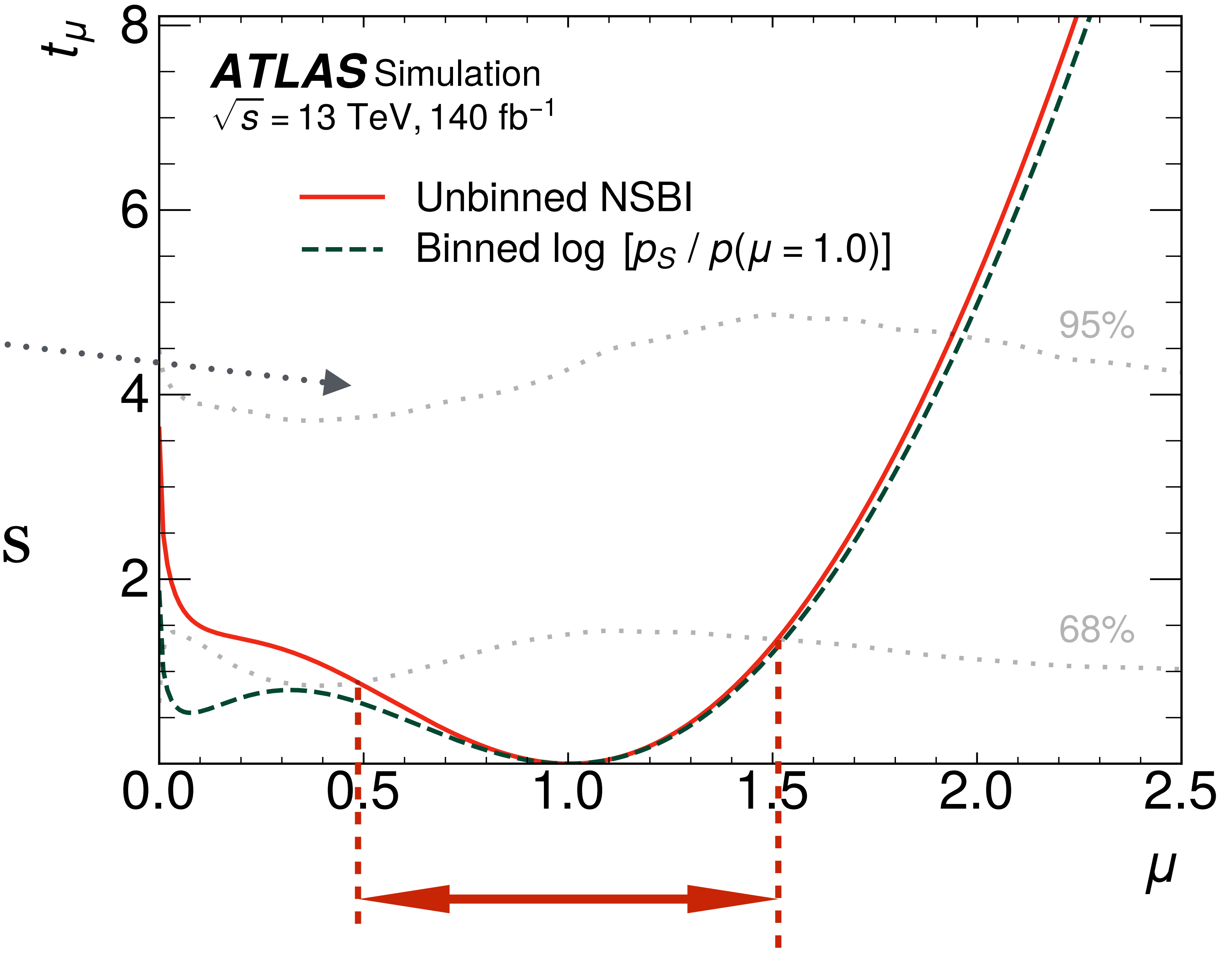


t_μ

ATLAS NSBI paper: [arXiv:2412.01600](https://arxiv.org/abs/2412.01600)

Computational Challenge: Inverting the test

- Determine 68 % & 95 % CI empirically from this distribution
- Do it for each value of μ

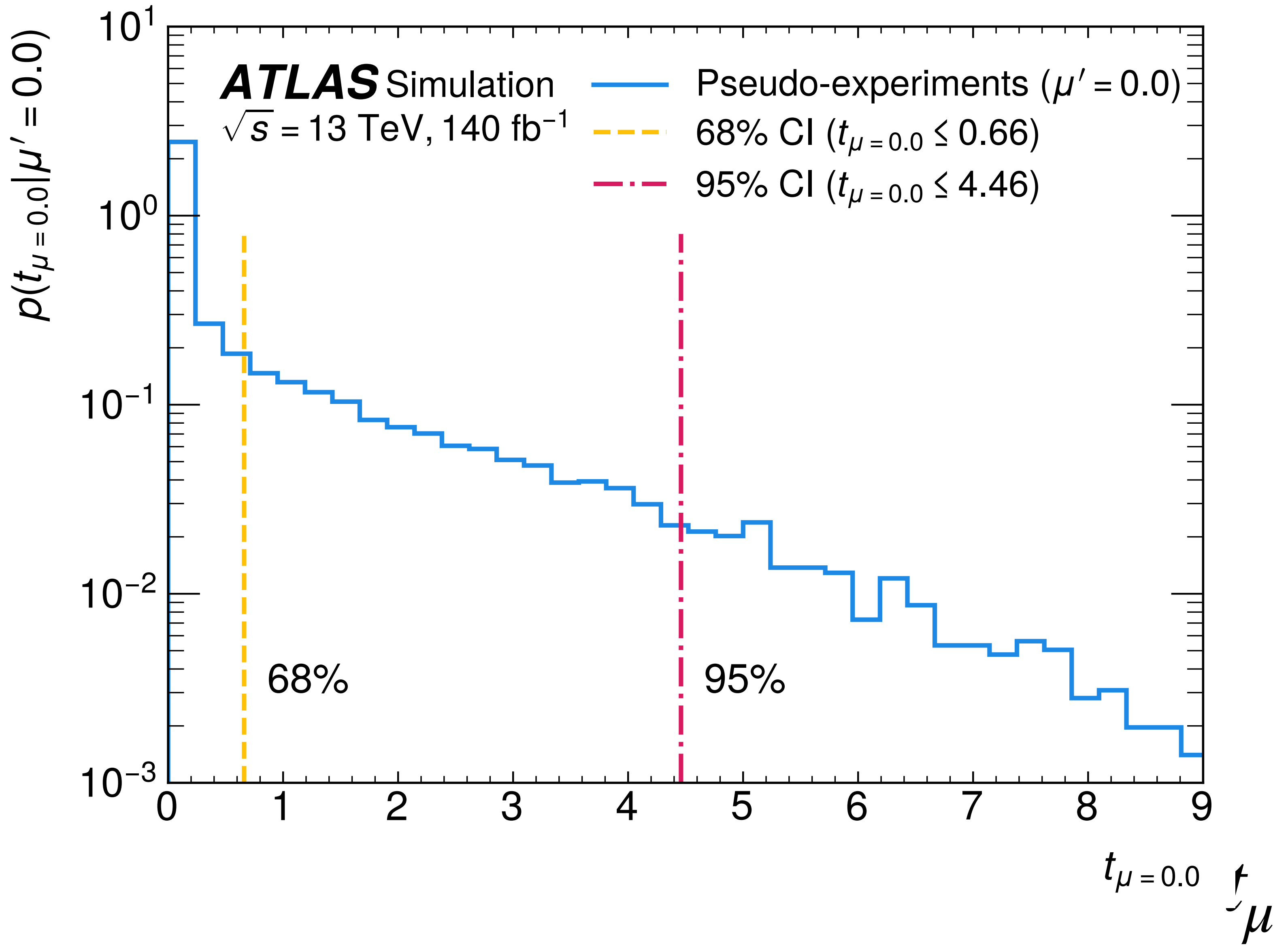


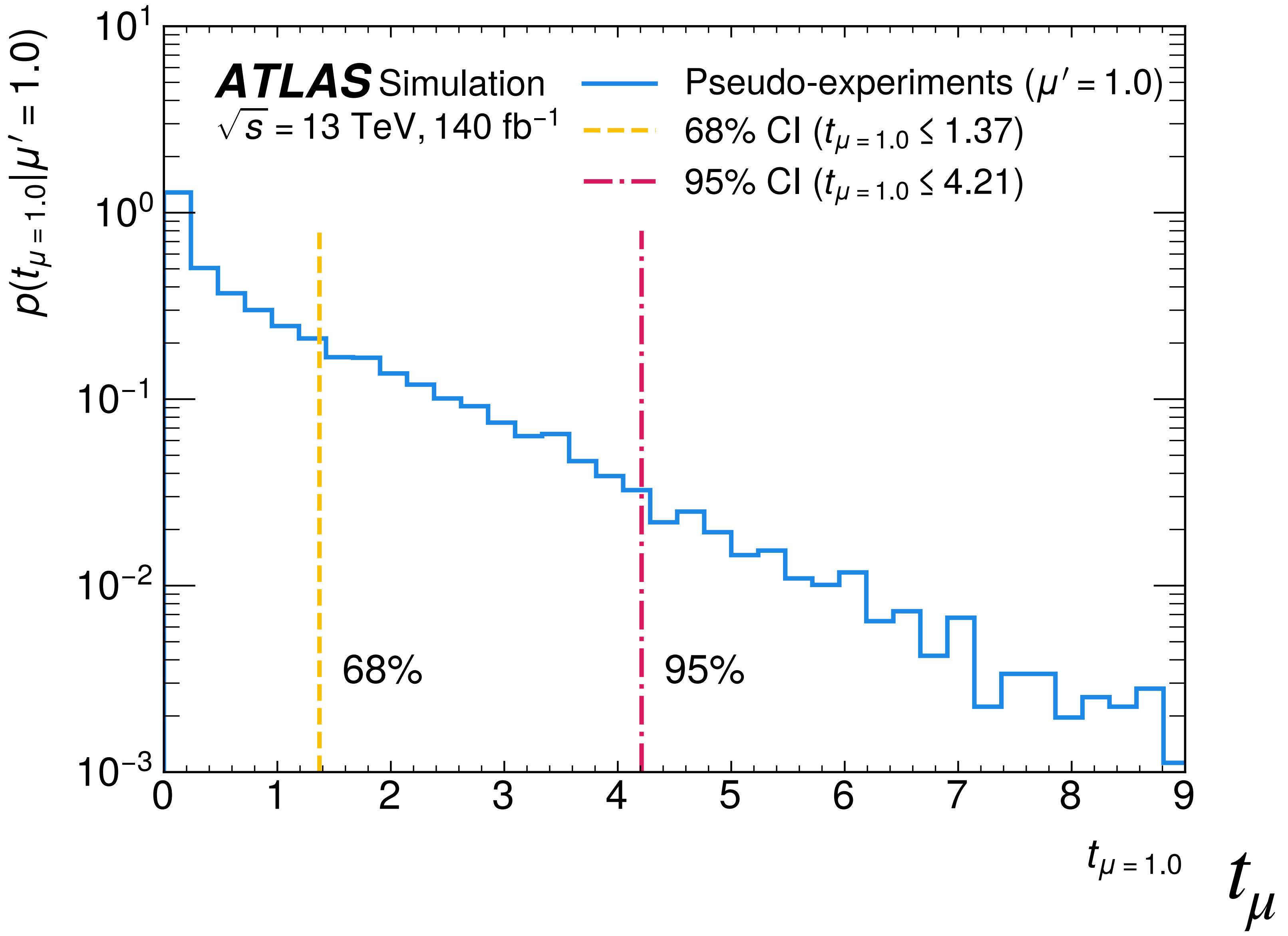
Calibrated confidence interval

Distribution of test statistic t_μ over thousands of simulated pseudo-experiments

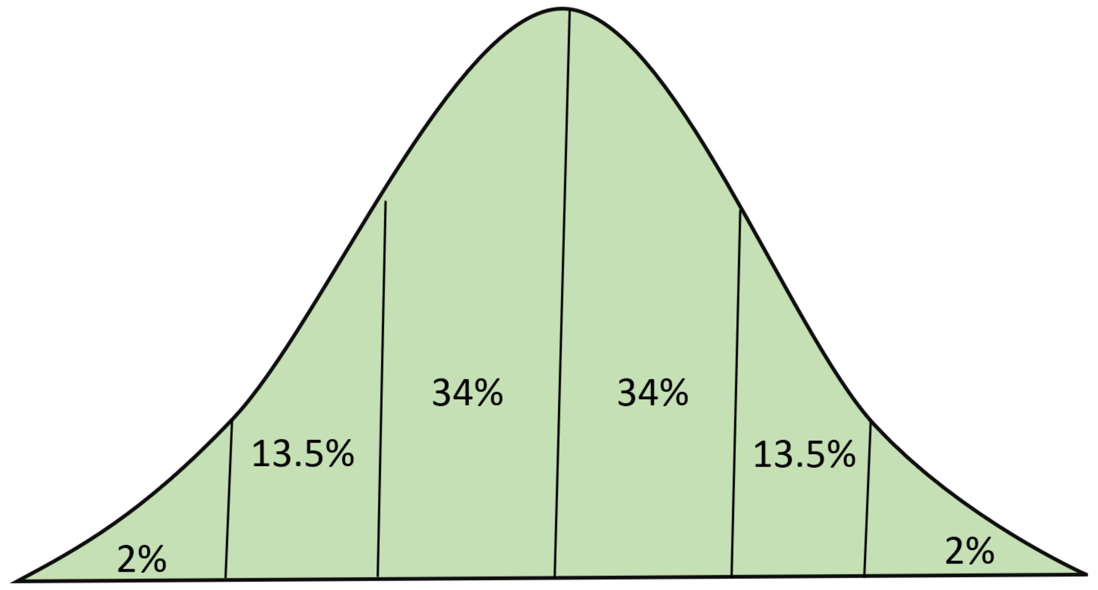
True $\mu = 0$

True $\mu = 1$

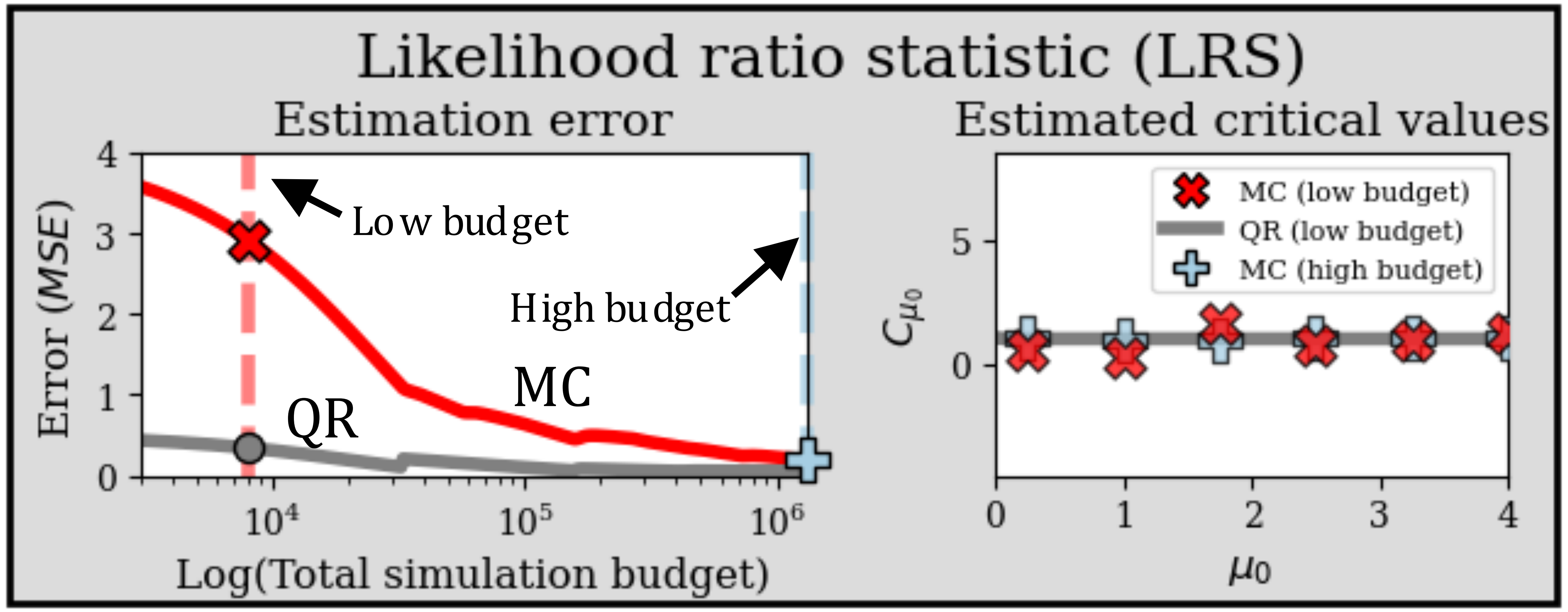




Task is to determine quantiles



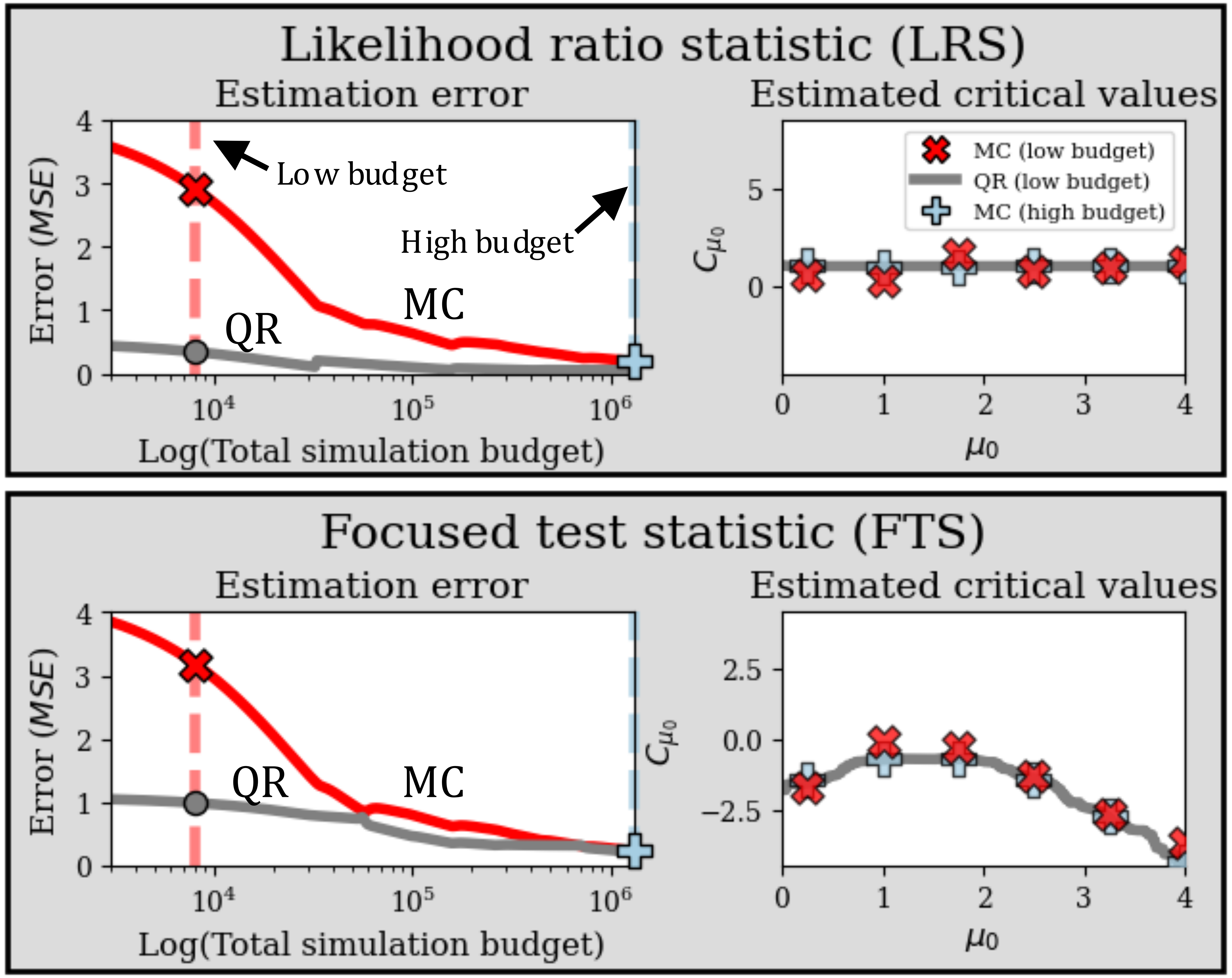
Perform efficiently with quantile regression!



Can do it faster and more accurately with ML technique known as 'quantile regression' (QR)

Perform efficiently with quantile regression!

Can do it faster and more accurately with ML technique known as 'quantile regression' (QR)



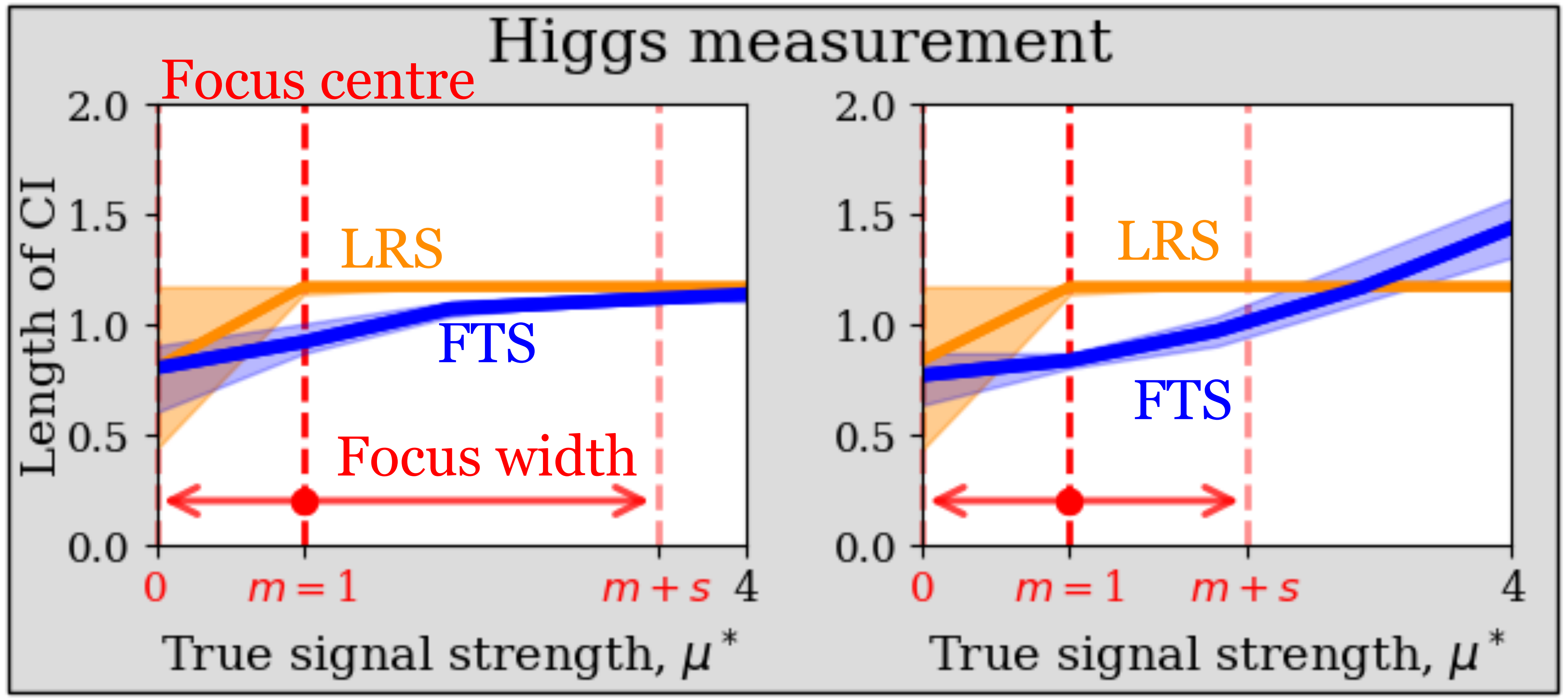
arXiv:2507.17831: Carzon, Ghosh, Izbicki, Lee, et al

Now ready to compare LRS to FTS on HiggsML benchmark dataset

$H \rightarrow \tau\tau$ ATLAS simulated public benchmark dataset

Now ready to compare LRS to FTS on HiggsML benchmark dataset

← Smaller is better



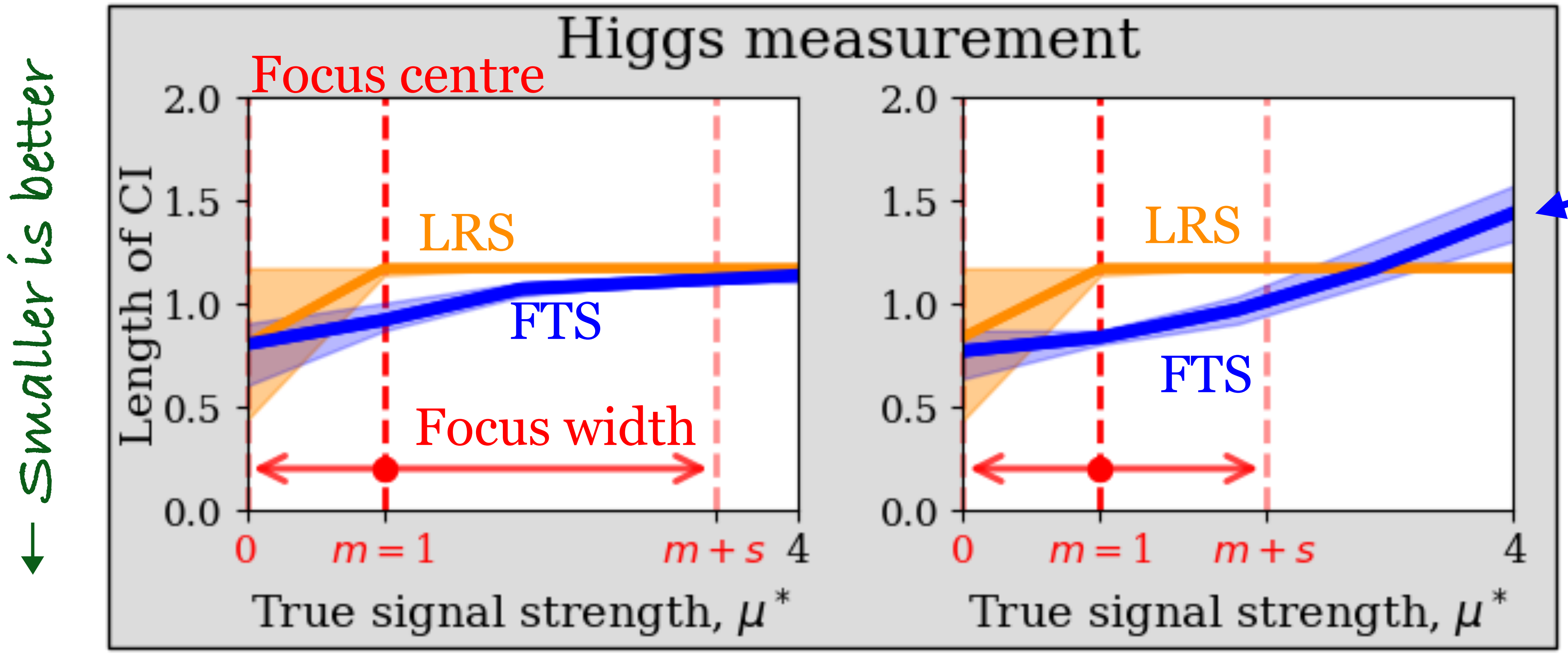
Wide focus $f(\mu)$

Narrow focus $f(\mu)$

$H \to \tau\tau$ ATLAS simulated public benchmark dataset

Now ready to compare LRS to FTS on HiggsML benchmark dataset

← Smaller is better



Wide focus $f(\mu)$

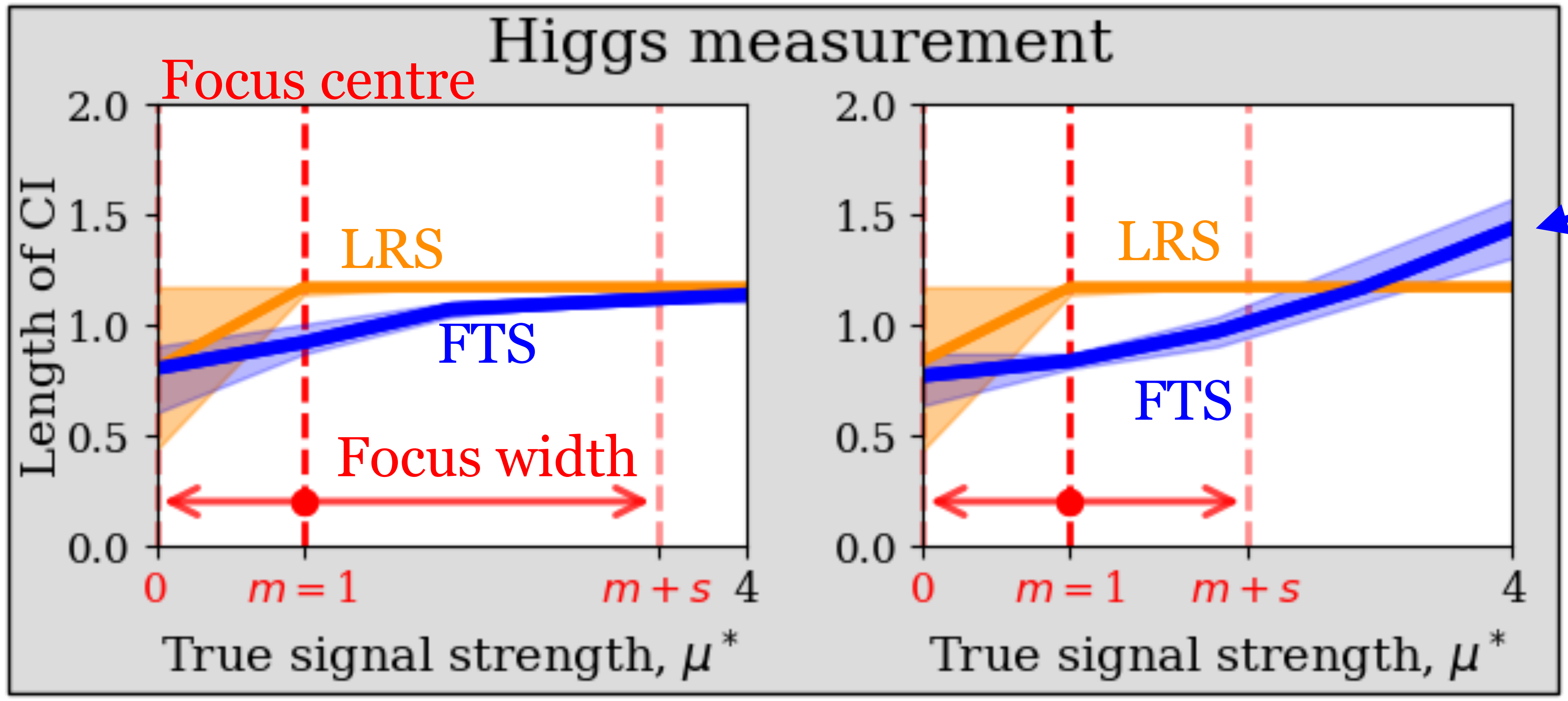
Narrow focus $f(\mu)$

FTS worse than LRS when true μ is very far away from focus region, but coverage of confidence intervals still correct

$H \to \tau\tau$ ATLAS simulated public benchmark dataset

Now ready to compare LRS to FTS on HiggsML benchmark dataset

← Smaller is better



Wide focus $f(\mu)$

Narrow focus $f(\mu)$

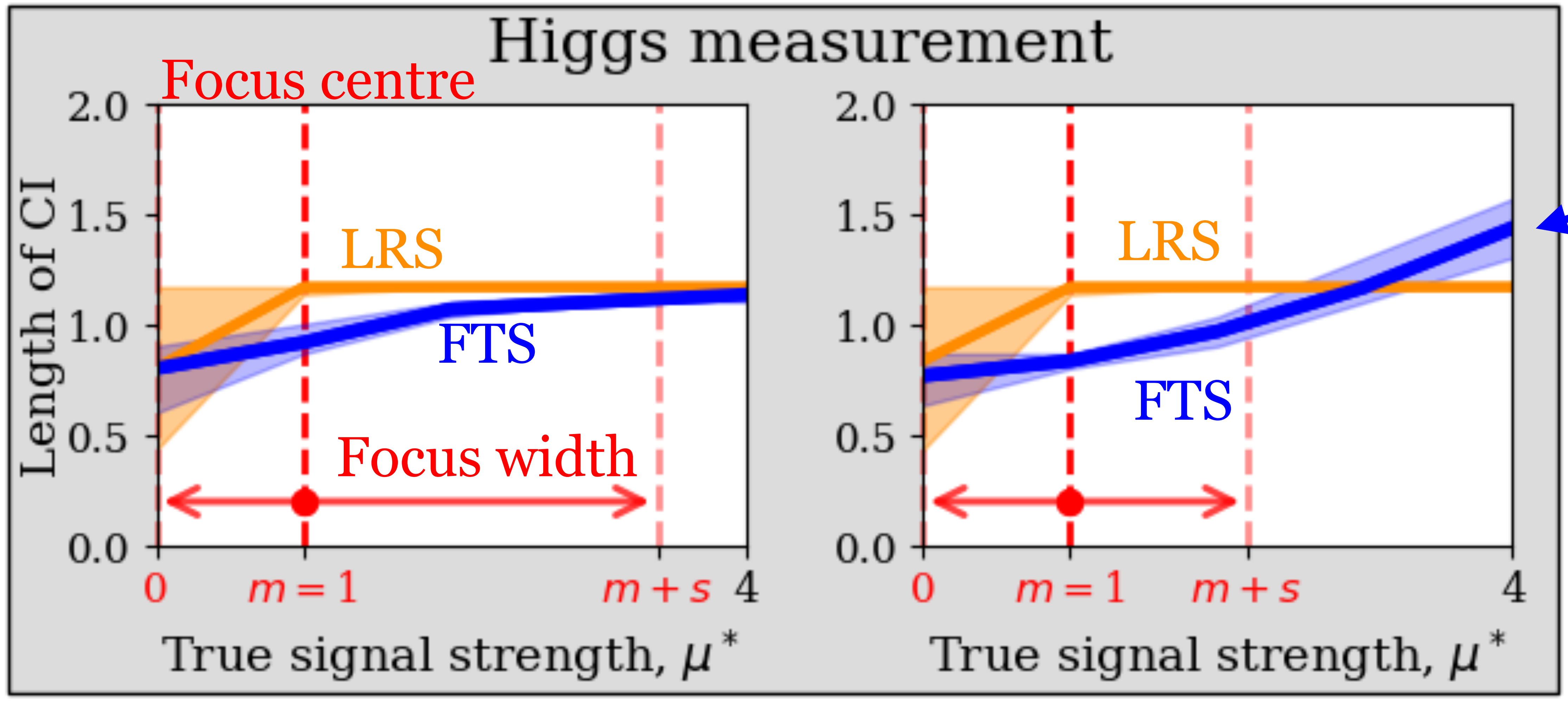
FTS worse than LRS when true μ is very far away from focus region, but coverage of confidence intervals still correct

$H \rightarrow \tau\tau$ ATLAS simulated public benchmark dataset

Even when it is worse, the coverage of the confidence interval is still guaranteed by Neyman construction!

Now ready to compare LRS to FTS on HiggsML benchmark dataset

← Smaller is better



Wide focus $f(\mu)$

Narrow focus $f(\mu)$

FTS worse than LRS when true μ is very far away from focus region, but coverage of confidence intervals still correct

$H \rightarrow \tau\tau$ ATLAS simulated public benchmark dataset

Which focus would you choose?

Even when it is worse, the coverage of the confidence interval is still guaranteed by Neyman construction!

funding gets harder to secure, principal investigators are in their office writing grants while the trainees get to do the cool stuff.

Bryan W. Jones is a retinal neuroscientist at the University of Pittsburgh in Pennsylvania.

AISHIK GHOSH

STUDENTS OVERTURN LONG-HELD ASSUMPTION

I have worked on experimental particle physics since 2015, searching for Higgs bosons at CERN, Europe's particle-physics lab near Geneva, Switzerland, and now also working on the Deep Underground Neutrino Experiment (DUNE) in the United States. For this research, there's one statistical test we've used for decades to confirm the existence of a new particle – the generalized likelihood ratio test (GLRT). This compares two models – a simple null hypothesis, which includes no new particle or matter being discovered, and a more complicated alternative model, which includes a new particle with many possible values of strength.

In December 2024, a couple of PhD students working with my collaborator, Ann Lee, a data scientist at Carnegie Mellon University in Pittsburgh, Pennsylvania, were confident they could disprove the assumption that the GLRT was optimal. In the corner of my mind, I hoped they would prove us wrong. I gave them one of the most famous Higgs boson data sets to play around with. By early 2025, they showed that, although our previous physics results weren't wrong, our use of the GLRT wasn't ideal because it assumed large sample sizes are always generated, which is often not the case. Instead, the test left valuable information on the table. That day was special. I was still sceptical and I went through a battery of checks because I had to go back to my community and defend the PhD students' work, but it was all correct. The paper is currently in review, receiving a great deal of scrutiny.

Together, we produced a statistical test that will drastically improve our ability to make discoveries in particle physics, for example in searches for a new particle such as dark matter, where we expect to see only a few signal events at best. As a scientist, I want deeply held beliefs to be questioned. It was a real shock to the particle-physics community. Young people find it exciting. Senior members are still highly sceptical, as they should be, but they are coming around. As the DUNE experiment comes online, with this new statistical model in place, we hope to make precise measurements about neutrinos much sooner than anticipated.

Aishik Ghosh is a fundamental physicist at the Georgia Institute of Technology in Atlanta.

RAFIK TAREK NEME GARRIDO

SHOCKING CORAL FIND

A couple of years ago, after a day of pouring rain, the water on the Caribbean coast of Colombia was crystal clear and my master's student, Jorge Mareno, managed to take pictures of corals that no one knew existed here. We could find no scientific reports of corals in the area. Typically, the water is pretty turbid because the Magdalena River, which flows from the south of the country to the Caribbean Sea, brings chemicals and pollutants. It's an ongoing ecological and social challenge, but these corals must be adapting to these conditions. We did a sampling campaign across three days with a boat, using environmental DNA to find areas where corals, sponges and fish successfully survive the conditions. Most of the records are completely new for the region. It's super gratifying.

Rafik Tarek Neme Garrido is an evolutionary biologist at the University of the North in Barranquilla, Colombia.

TIM CURRAN

BURN PREDICTIONS

In my group, we test the flammability of plant species using a barbecue. The results can help with fire-mitigation policies and with understanding the evolution of flammability. As part of an outreach activity, we host schoolchildren at the university who haven't had much exposure to academia before. We ask the kids to predict how a particular plant species will behave – for example, what characteristics will make it burn less or more – and then we see who is right. The kids get really into it. They ask amazing questions, the same kind that peer reviewers have asked us, including questioning our methodological assumptions, such as "why do you only blowtorch them for ten seconds?"

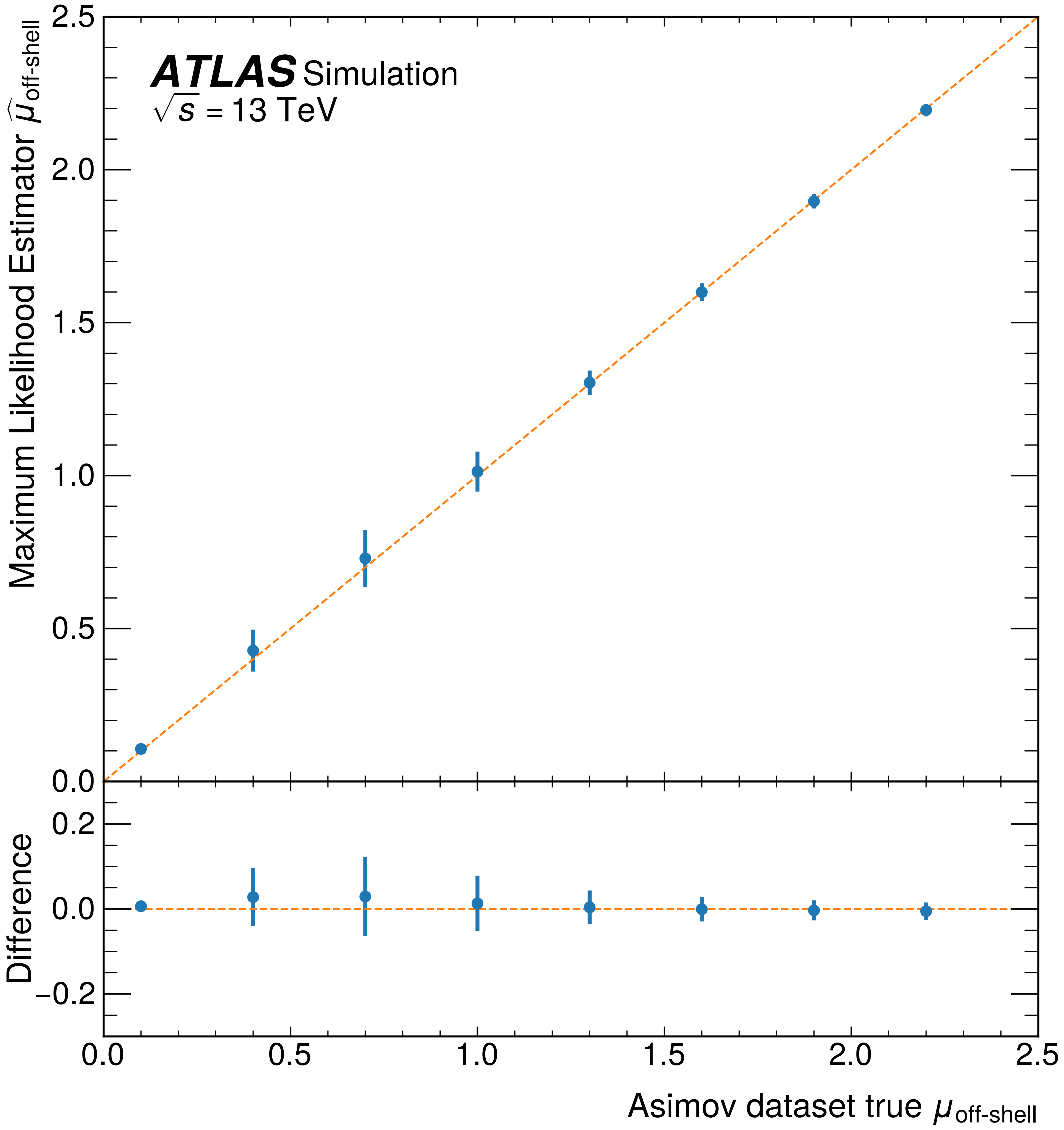
Most of the really good days doing science have been associated with young students having a light-bulb moment. In the rather

https://www.nature.com/articles/d41586-026-01479-8?utm_source=x&utm_medium=social&utm_campaign=nature&linkId=61827825

ATLAS Higgs width analysis paper: arXiv:2412.01548

Testing full analysis on simulated samples from different values of μ

Maximum likelihood estimate ($\hat{\mu}$)

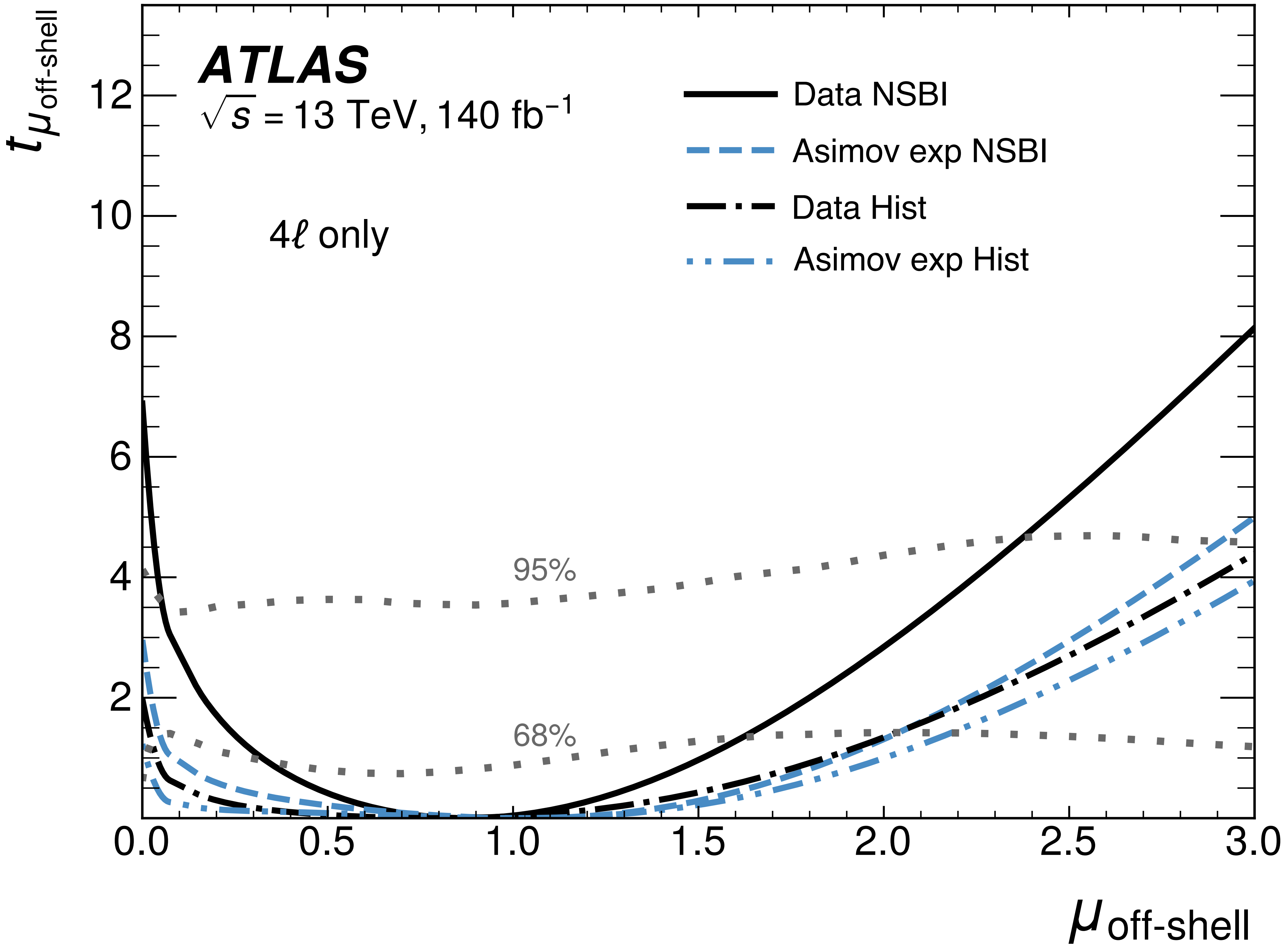


True μ

Ensure that true value is recovered by NSBI analysis on simulations

Application to a flagship Higgs measurement at LHC

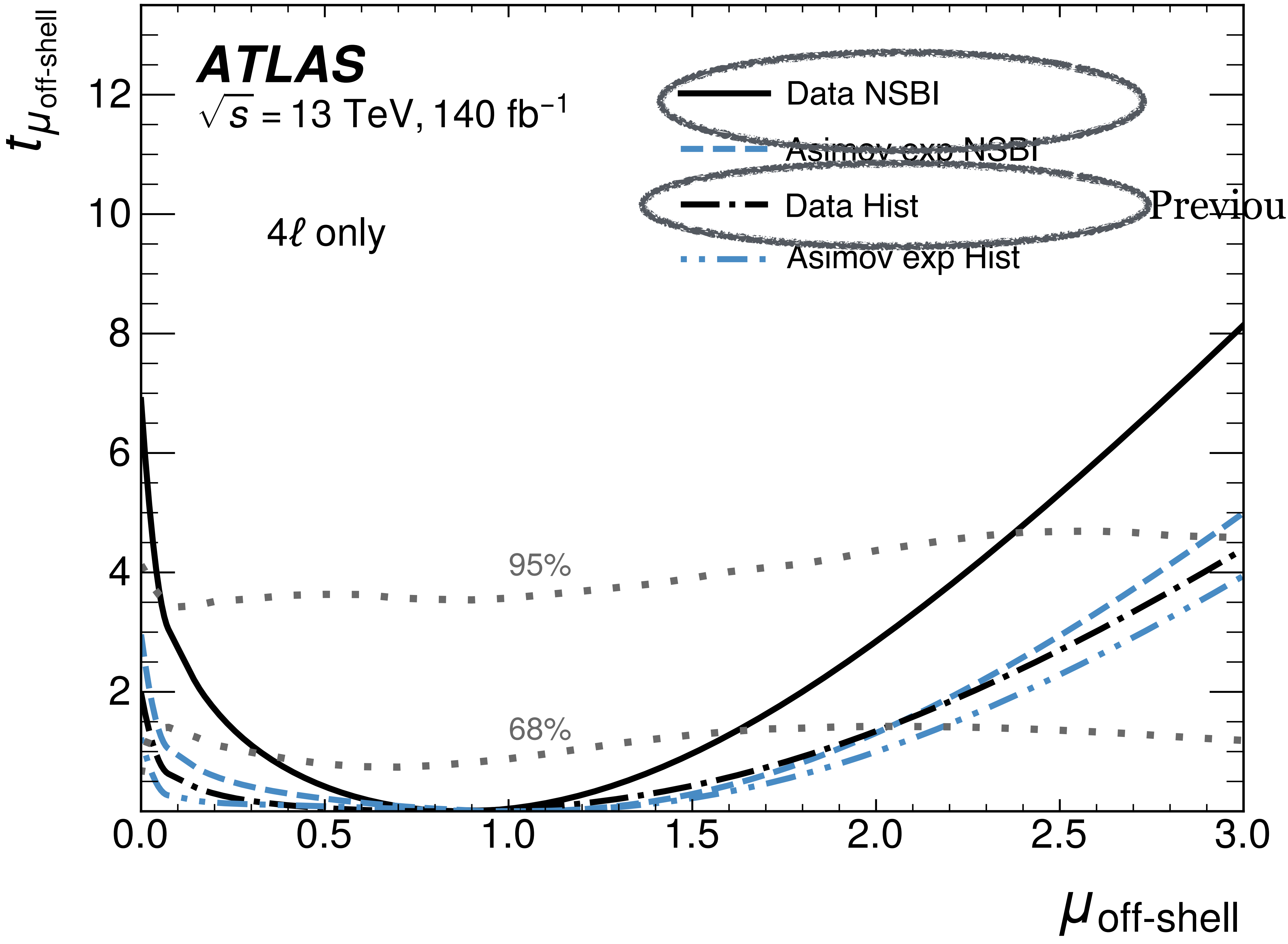
NSBI vs (published) histogram analysis



ATLAS Higgs width analysis paper: [arXiv:2412.01548](https://arxiv.org/abs/2412.01548)

Application to a flagship Higgs measurement at LHC

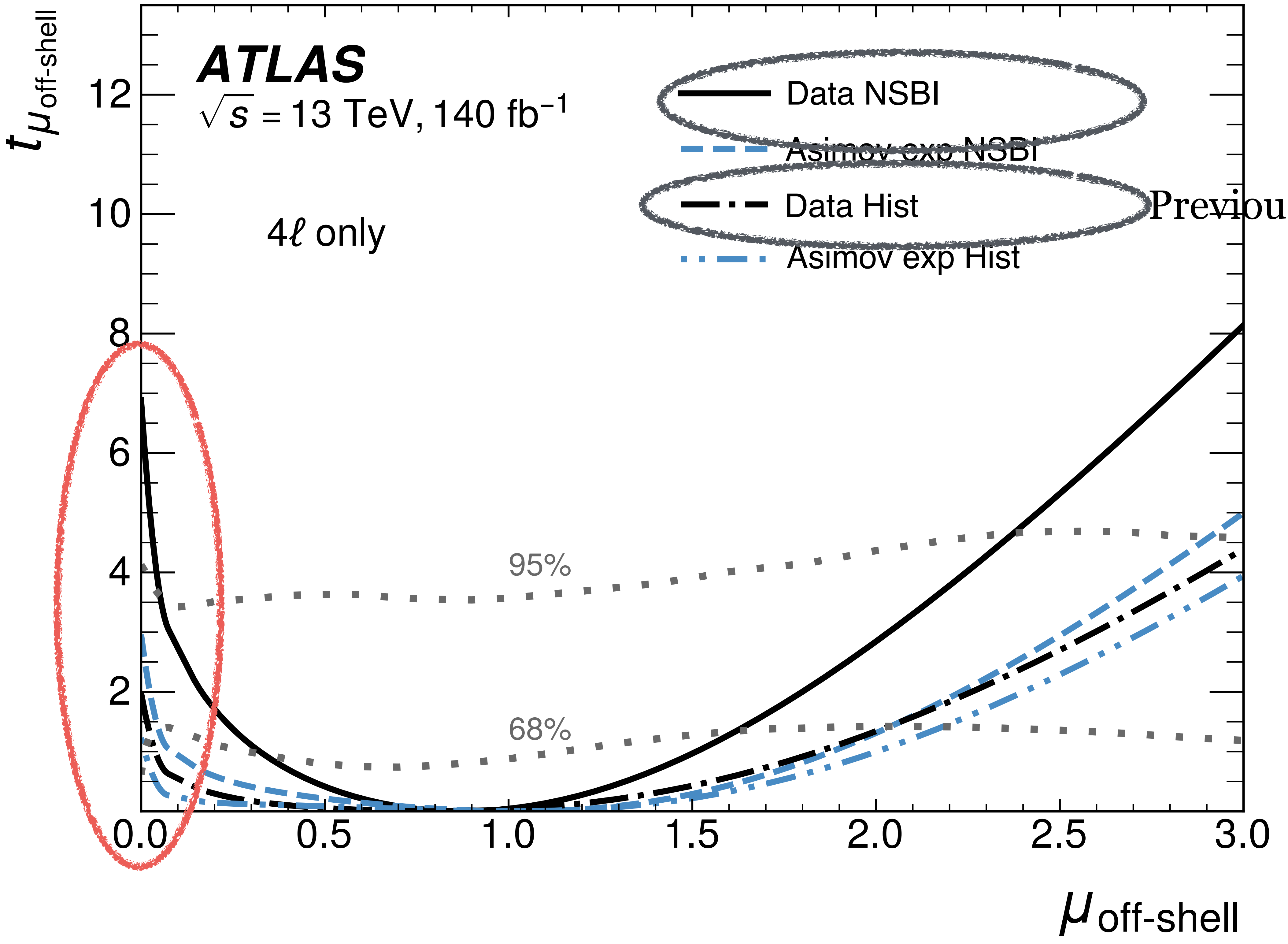
NSBI vs (published) histogram analysis



Previously published histogram paper

Application to a flagship Higgs measurement at LHC

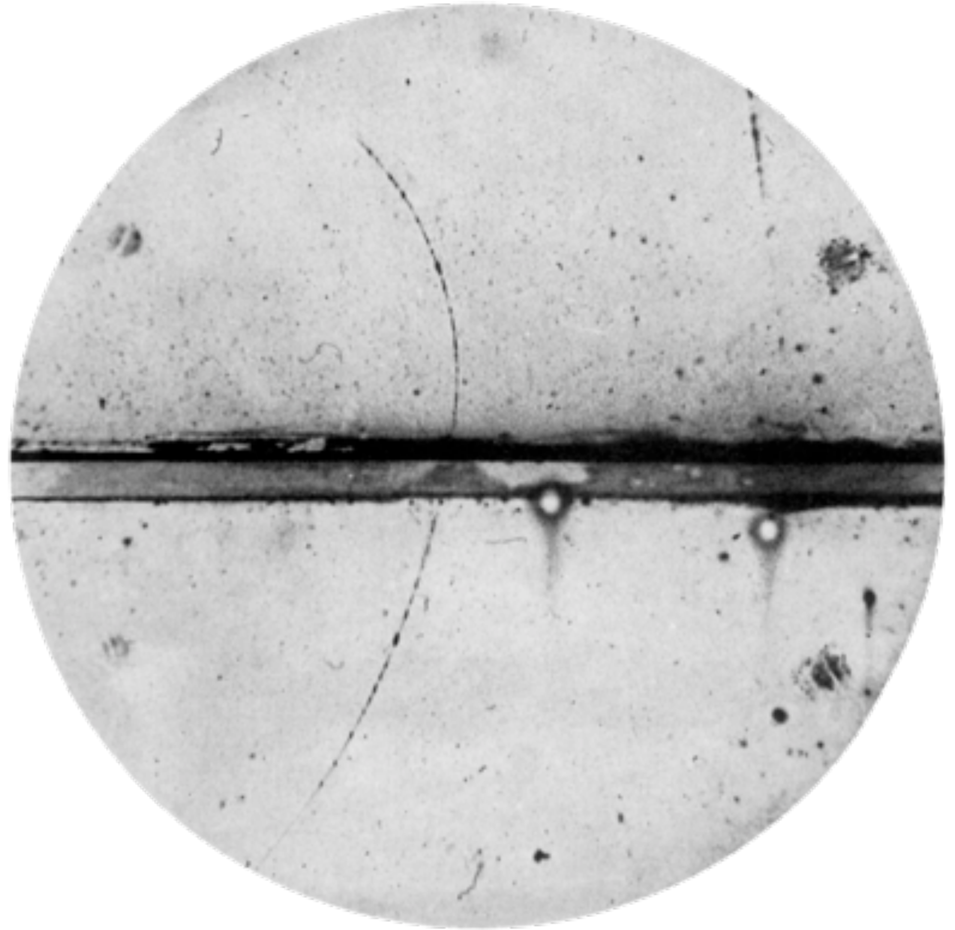
NSBI vs (published) histogram analysis



Previously published histogram paper

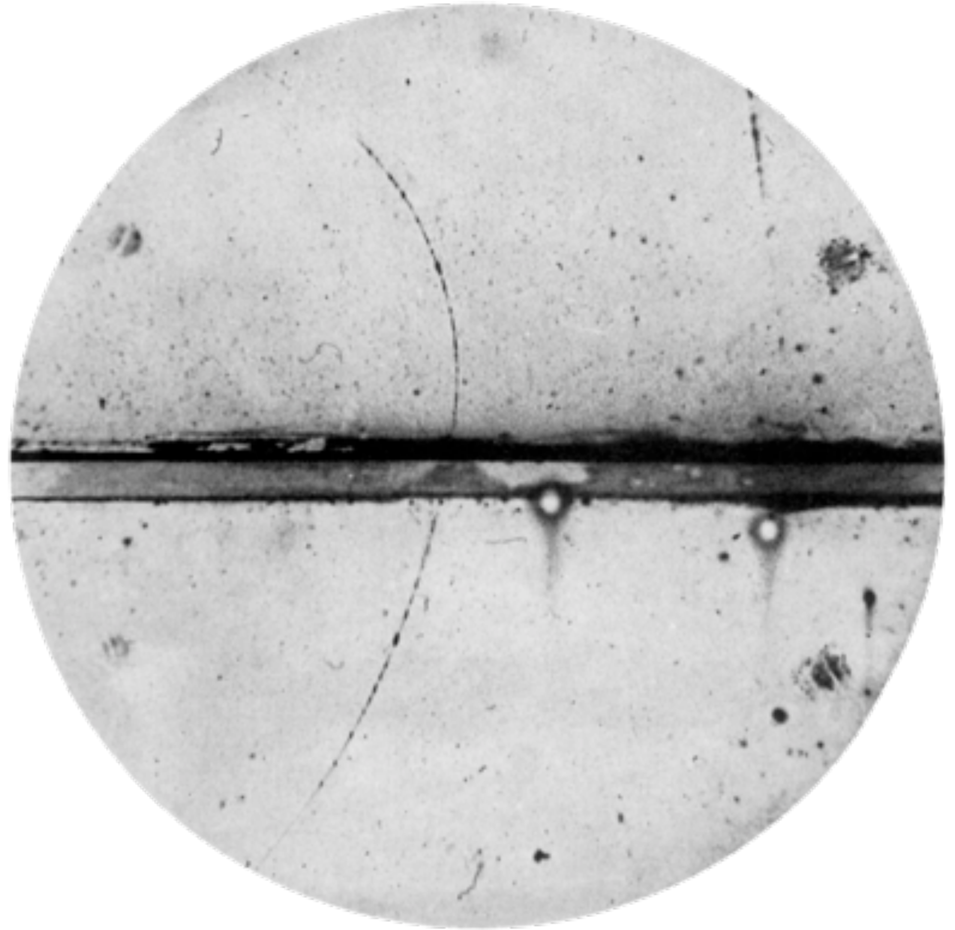
Unprecedented improvement in ability to reject null hypothesis!

Positron discovery (1930s)



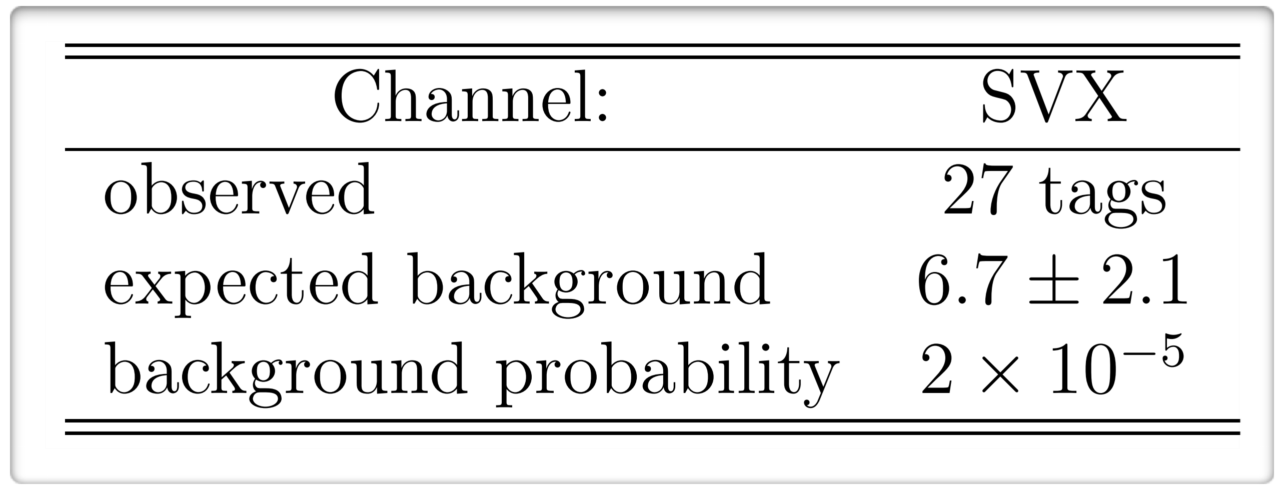
Single event

Positron discovery (1930s)



Single event

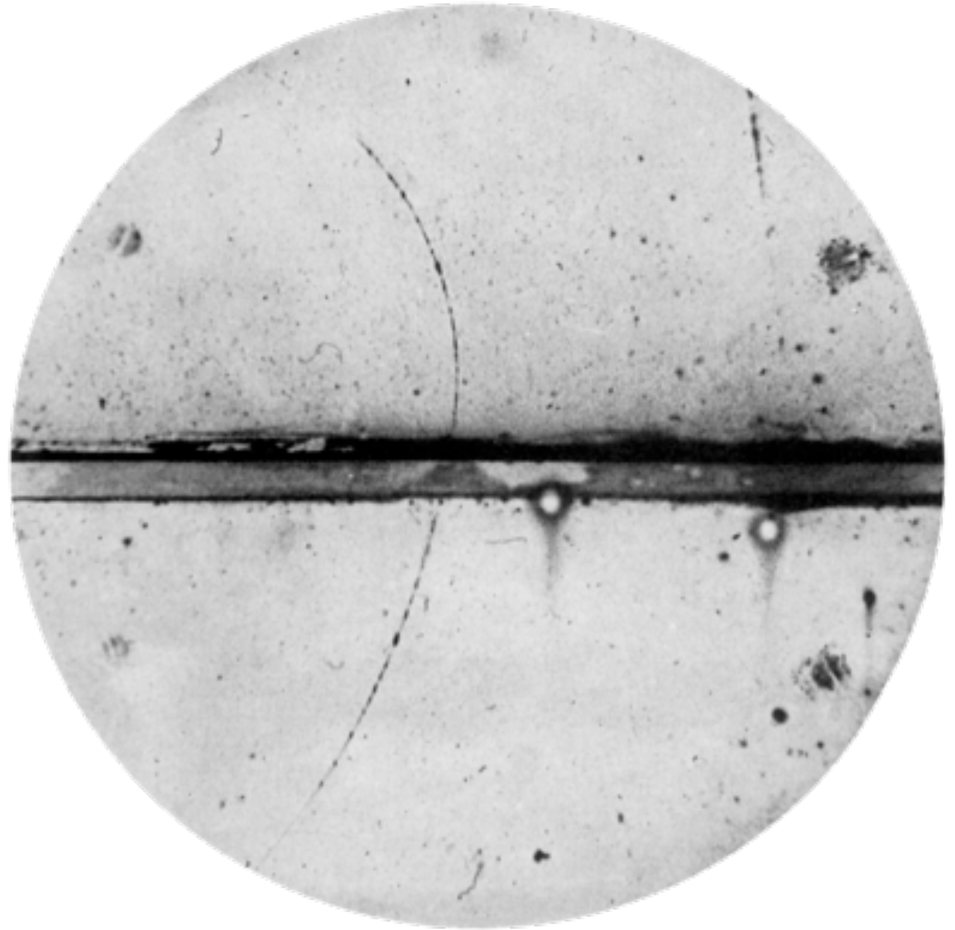
Top quark discovery (1990s)



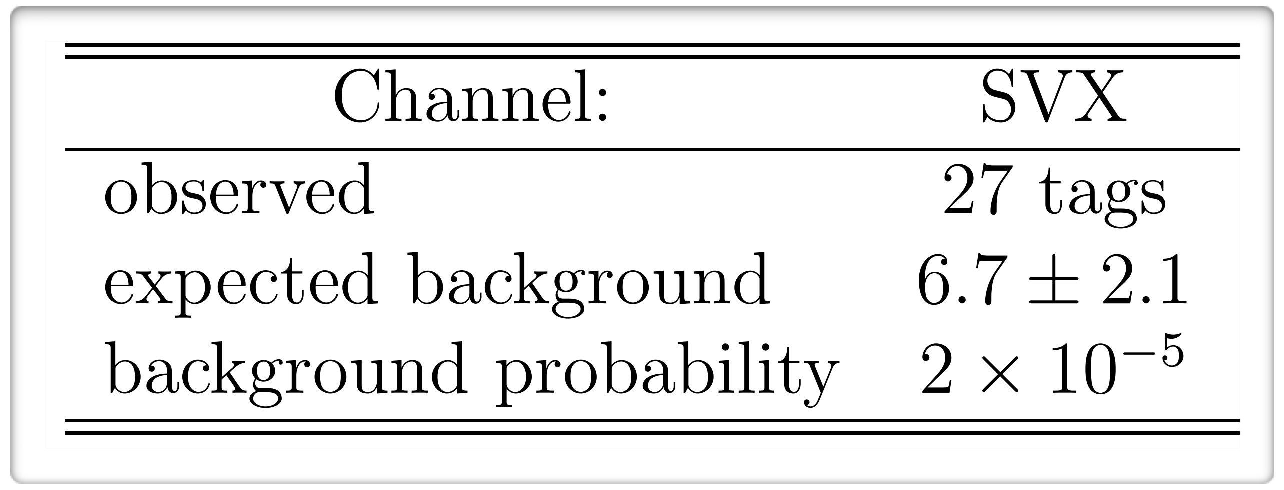
Channel:	SVX
observed	27 tags
expected background	6.7 ± 2.1
background probability	2×10^{-5}

Multiple events:
Cut-and-count

Positron discovery (1930s)

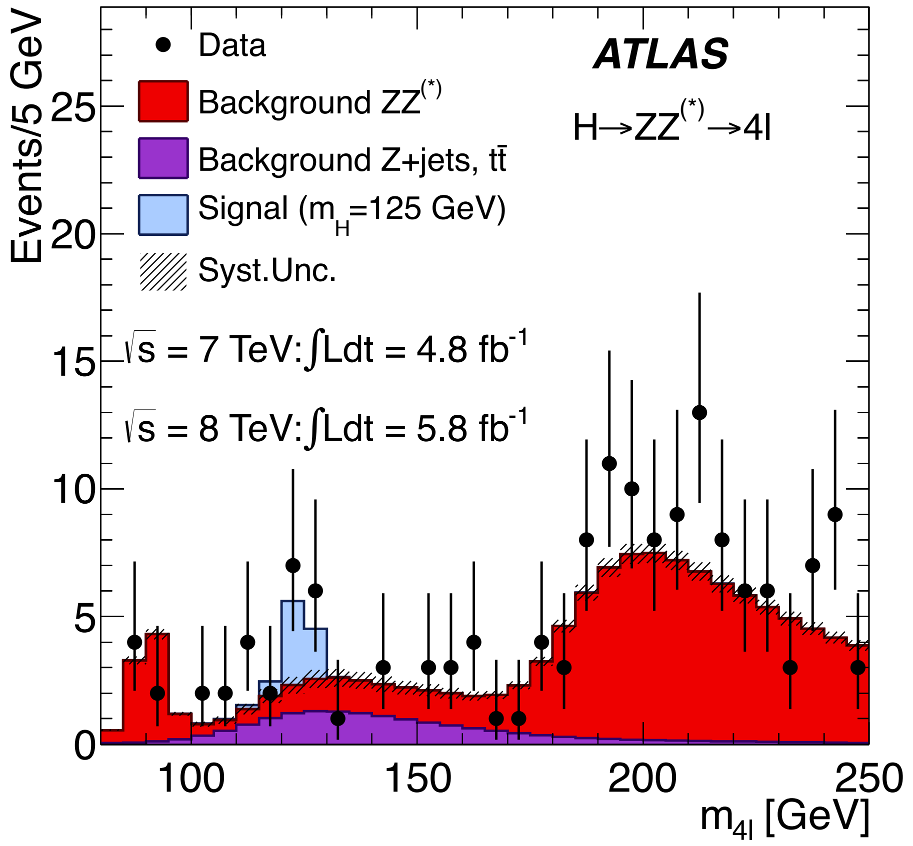


Top quark discovery (1990s)



Channel:	SVX
observed	27 tags
expected background	6.7 ± 2.1
background probability	2×10^{-5}

Higgs boson discovery (2010s)

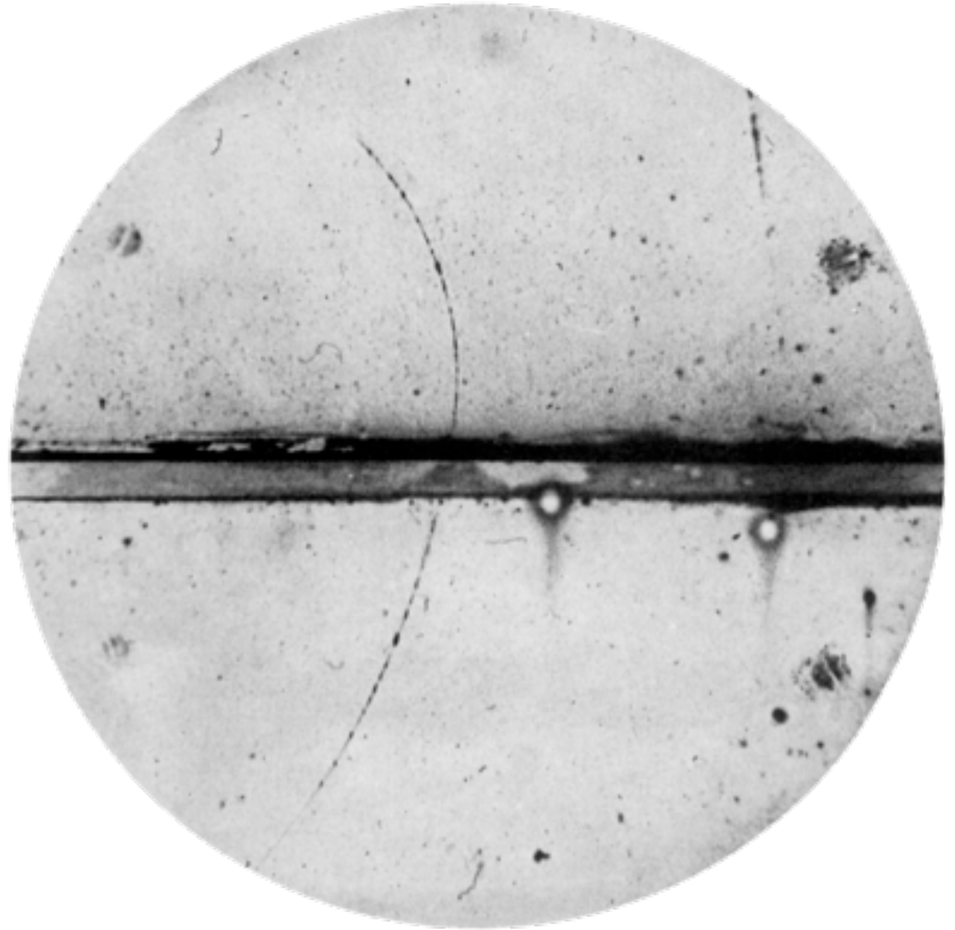


Single event

Multiple events:
Cut-and-count

Shape information:
Histogram

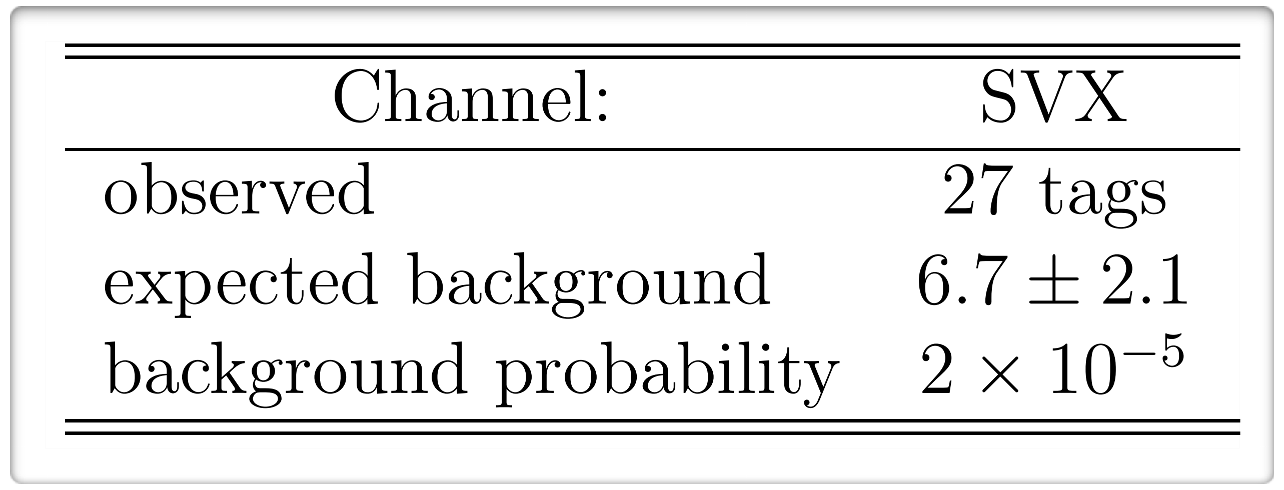
Positron discovery (1930s)



Single event

Image: Wikipedia / PhysRev.43.491

Top quark discovery (1990s)

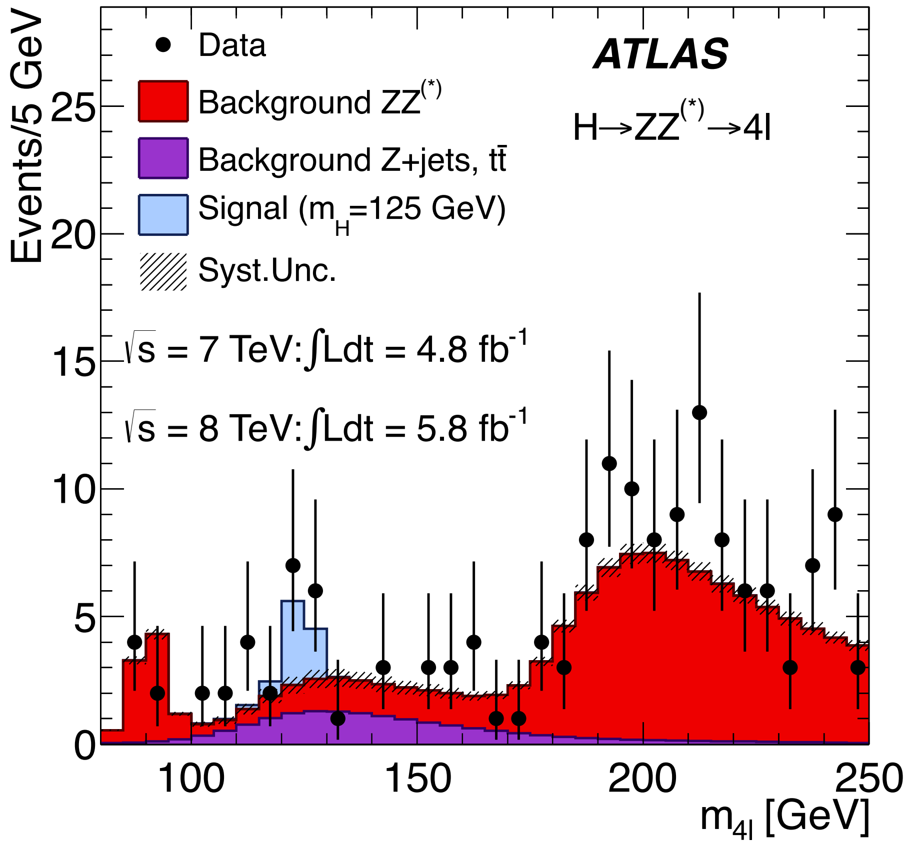


Channel:	SVX
observed	27 tags
expected background	6.7 ± 2.1
background probability	2×10^{-5}

Multiple events:
Cut-and-count

CDF Collaboration: arXiv:9503002

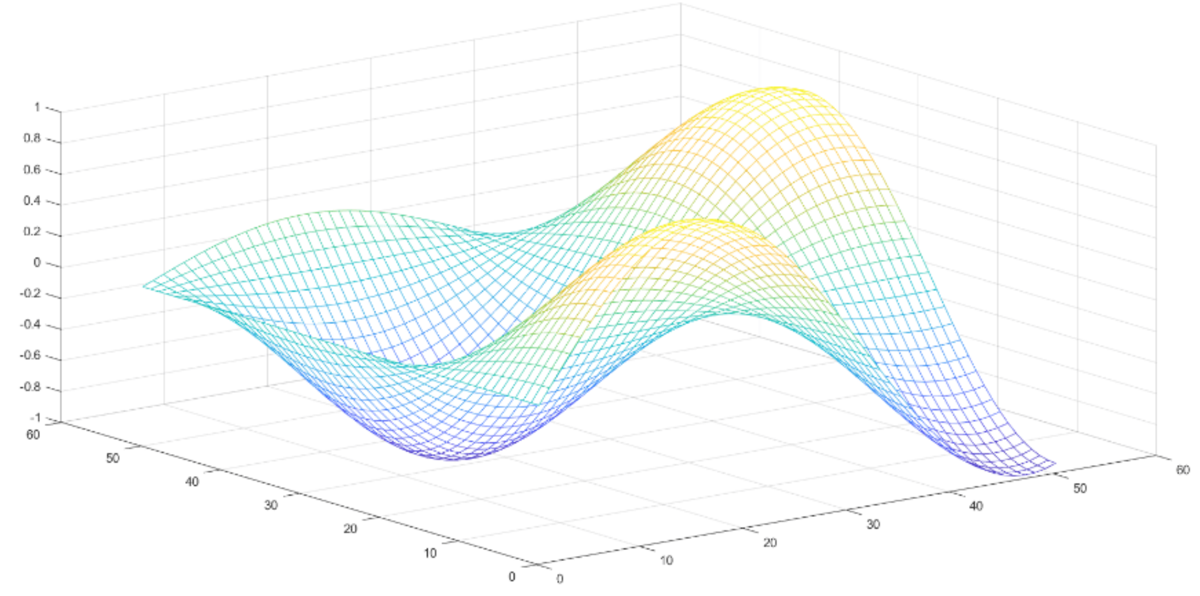
Higgs boson discovery (2010s)



Shape information:
Histogram

ATLAS Collaboration: arXiv:1207.7214

Future discovery (2020s ?)



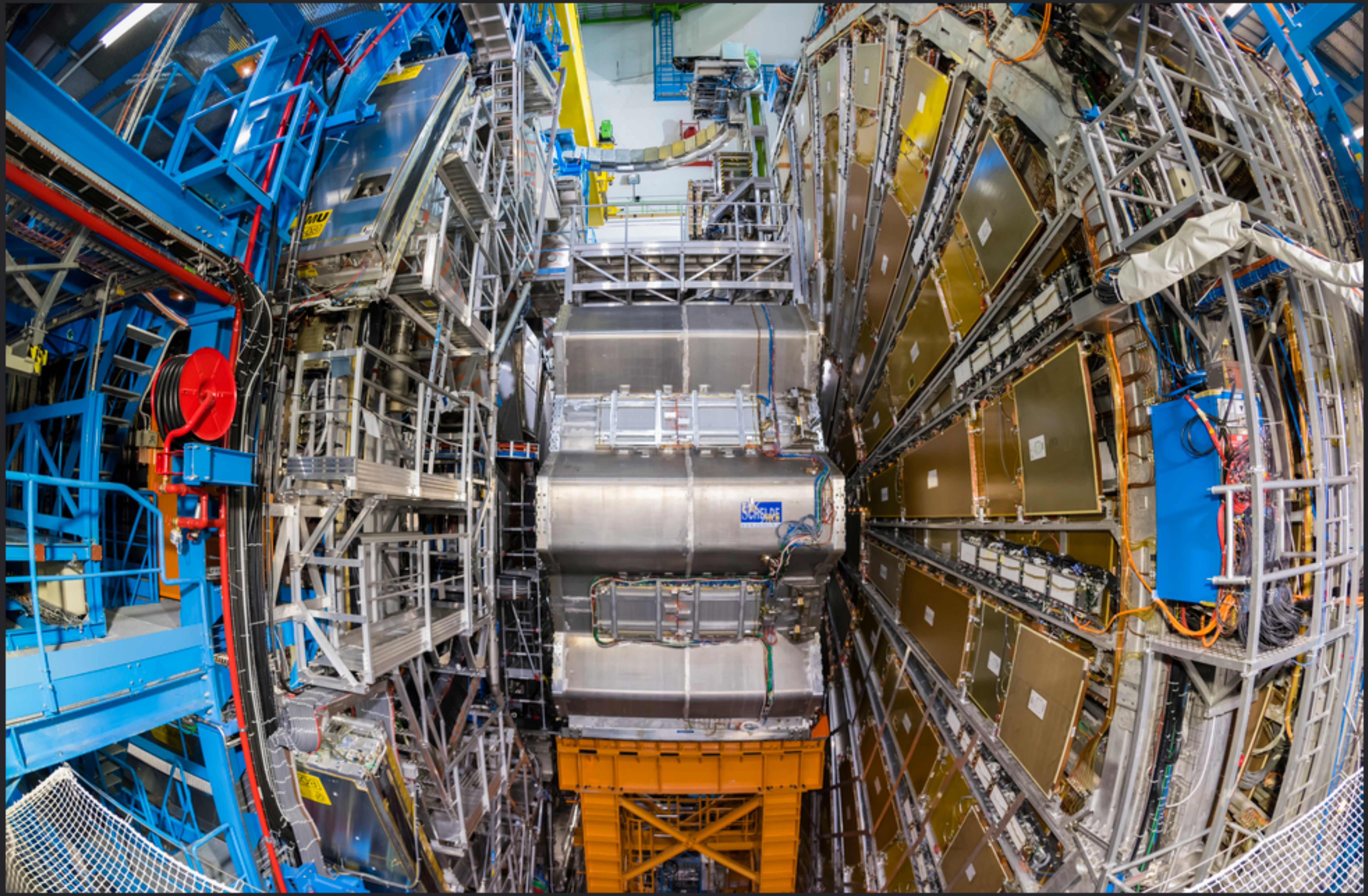
High-dim shape information,
continuous (i.e. unbinned):
Neural inference

A VERY FUN PROBLEM

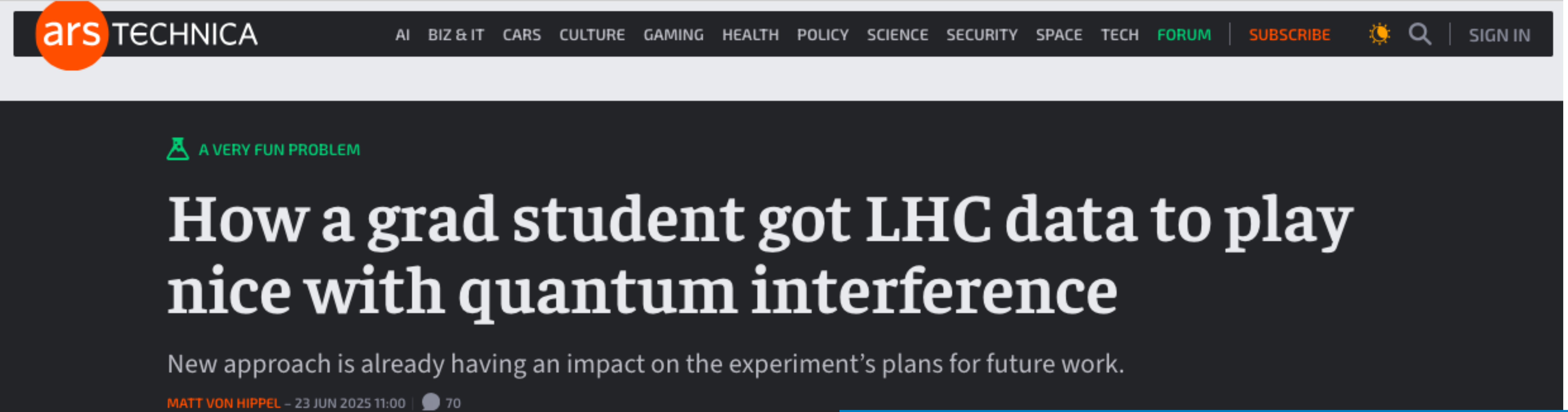
How a grad student got LHC data to play nice with quantum interference

New approach is already having an impact on the experiment’s plans for future work.

MATT VON HIPPEL – 23 JUN 2025 11:00 | 70



The ATLAS particle detector of the Large Hadron Collider (LHC) at the European Nuclear Research Center (CERN) in Geneva, Switzerland. Credit: EThamPhoto/Getty Images

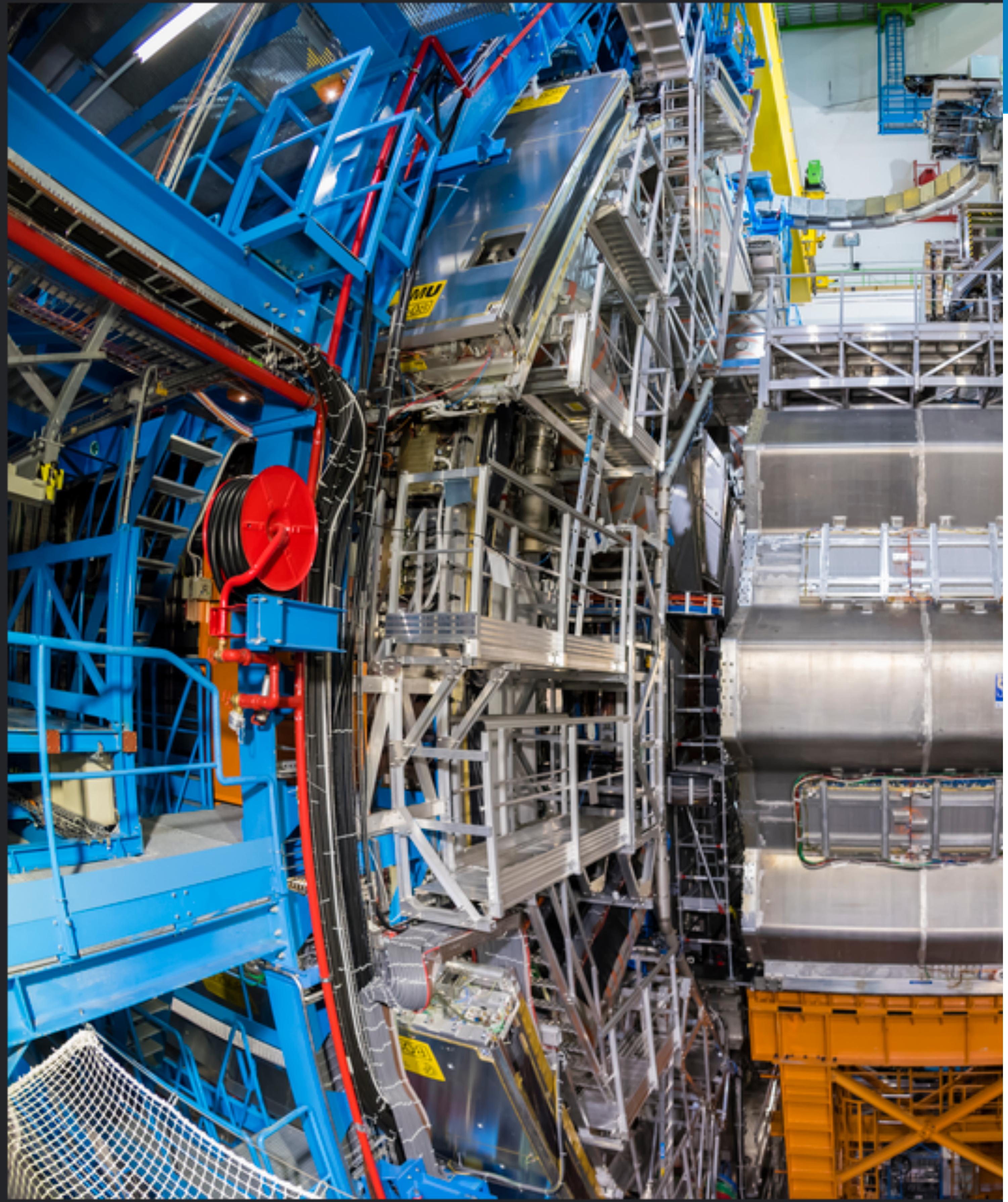


A VERY FUN PROBLEM

How a grad student got LHC data to play nice with quantum interference

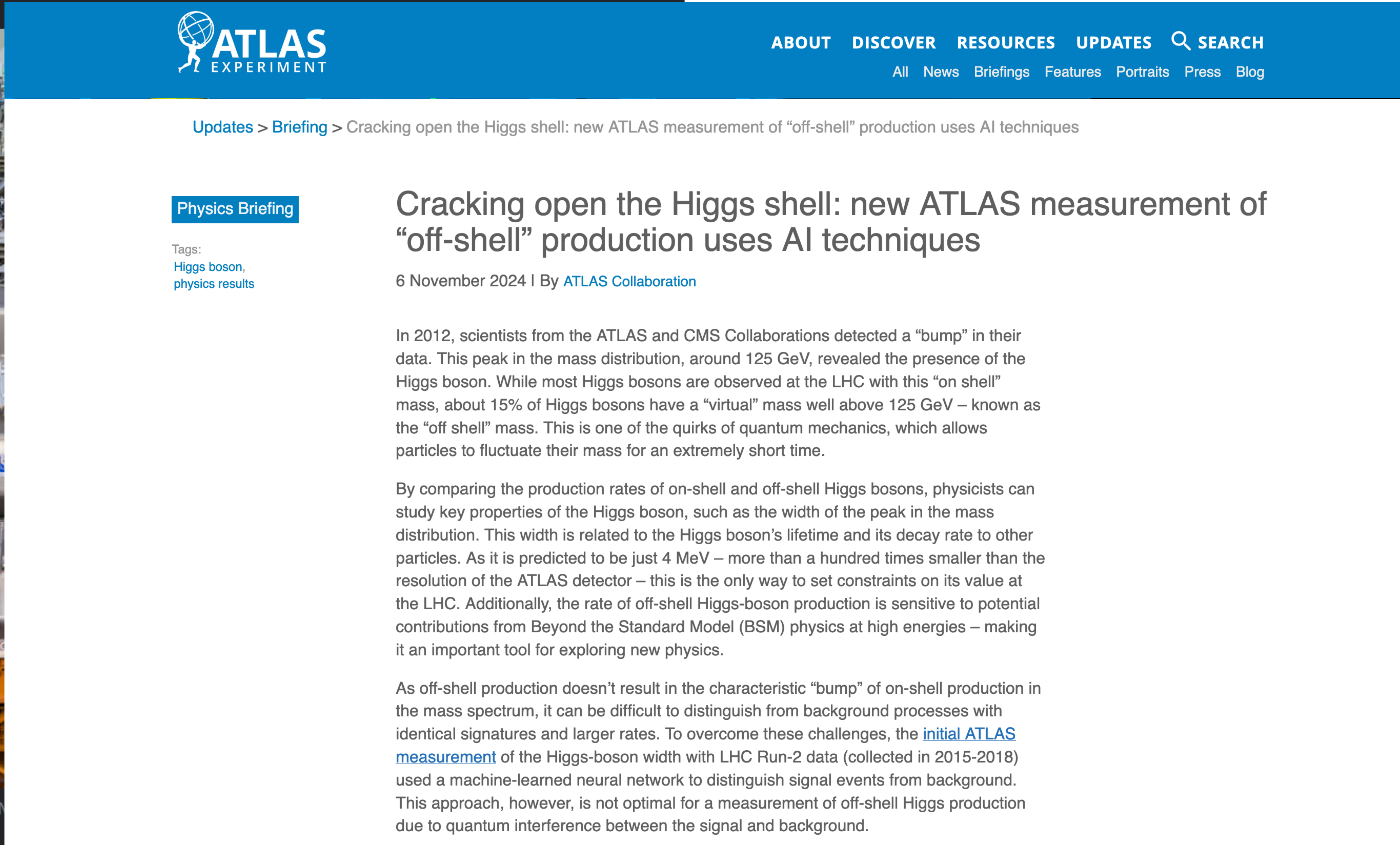
New approach is already having an impact on the experiment's plans for future work.

MATT VON HIPPEL – 23 JUN 2025 11:00 | 70



The ATLAS particle detector of the Large Hadron Collider (LHC) at the European N

EThamPhoto/Getty Images



Updates > Briefing > Cracking open the Higgs shell: new ATLAS measurement of "off-shell" production uses AI techniques

Physics Briefing

Tags:
Higgs boson,
physics results

Cracking open the Higgs shell: new ATLAS measurement of "off-shell" production uses AI techniques

6 November 2024 | By ATLAS Collaboration

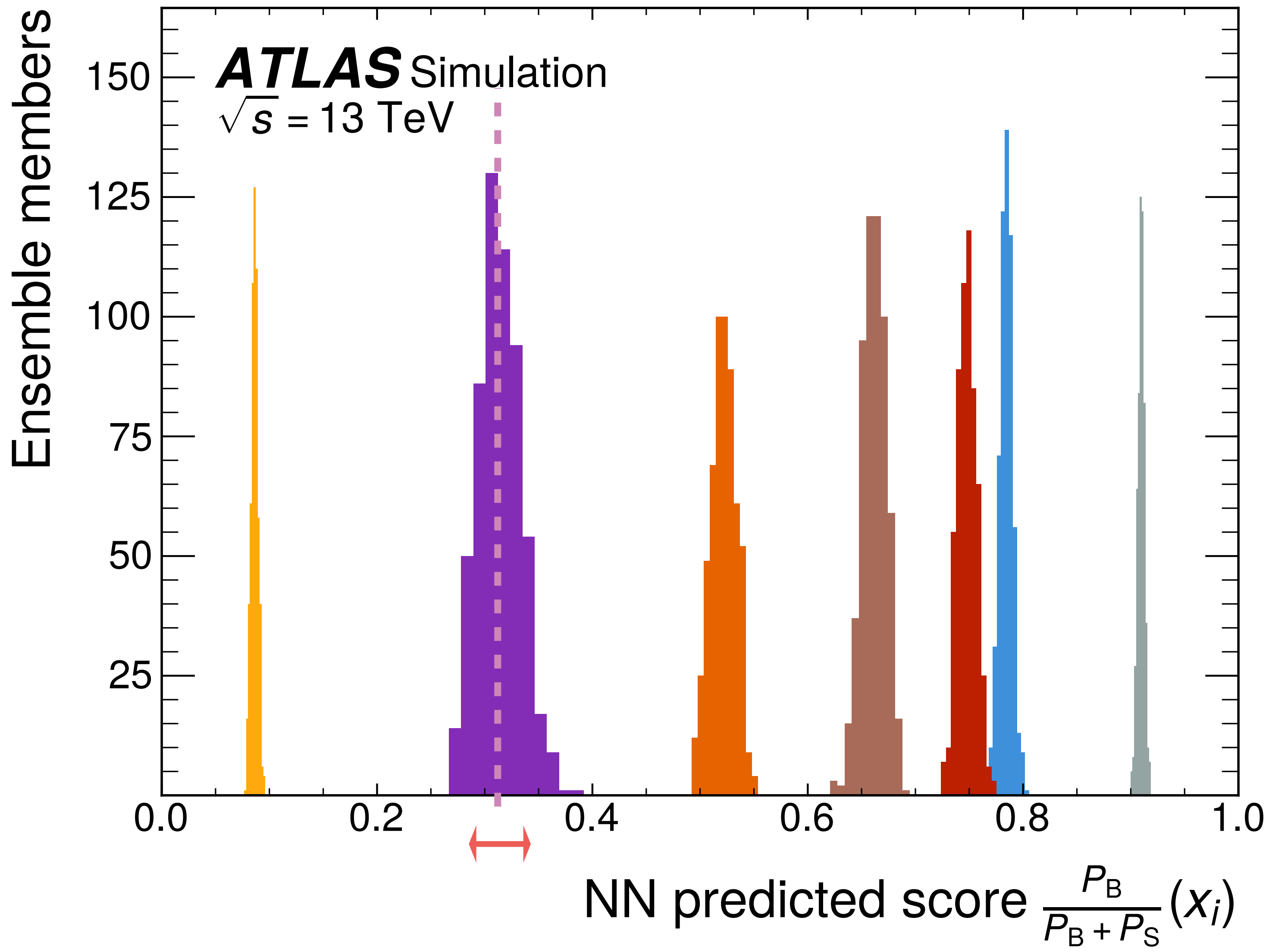
In 2012, scientists from the ATLAS and CMS Collaborations detected a "bump" in their data. This peak in the mass distribution, around 125 GeV, revealed the presence of the Higgs boson. While most Higgs bosons are observed at the LHC with this "on shell" mass, about 15% of Higgs bosons have a "virtual" mass well above 125 GeV – known as the "off shell" mass. This is one of the quirks of quantum mechanics, which allows particles to fluctuate their mass for an extremely short time.

By comparing the production rates of on-shell and off-shell Higgs bosons, physicists can study key properties of the Higgs boson, such as the width of the peak in the mass distribution. This width is related to the Higgs boson's lifetime and its decay rate to other particles. As it is predicted to be just 4 MeV – more than a hundred times smaller than the resolution of the ATLAS detector – this is the only way to set constraints on its value at the LHC. Additionally, the rate of off-shell Higgs-boson production is sensitive to potential contributions from Beyond the Standard Model (BSM) physics at high energies – making it an important tool for exploring new physics.

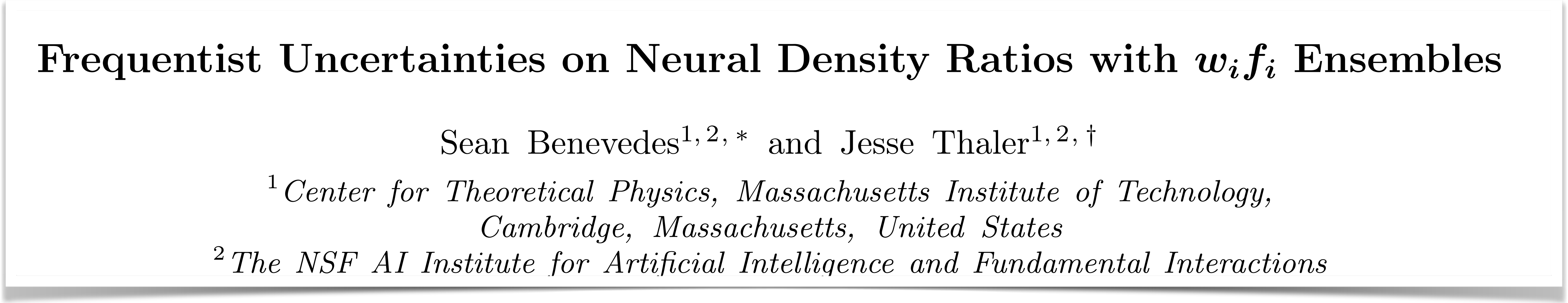
As off-shell production doesn't result in the characteristic "bump" of on-shell production in the mass spectrum, it can be difficult to distinguish from background processes with identical signatures and larger rates. To overcome these challenges, the initial ATLAS measurement of the Higgs-boson width with LHC Run-2 data (collected in 2015-2018) used a machine-learned neural network to distinguish signal events from background. This approach, however, is not optimal for a measurement of off-shell Higgs production due to quantum interference between the signal and background.

Uncertainties on the NN models themselves

Distribution of NN predictions for example events



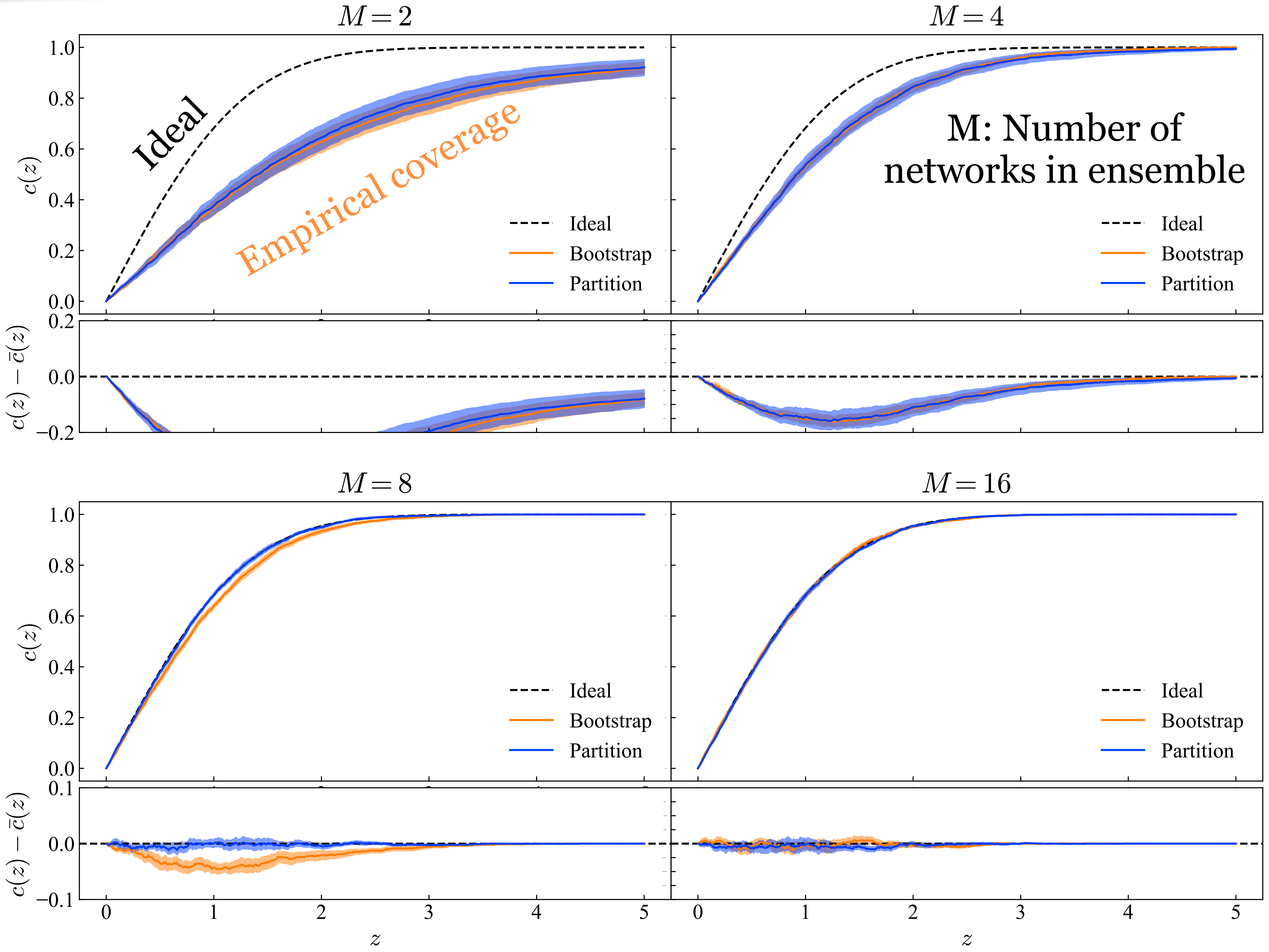
Uncertainties on the NN models themselves



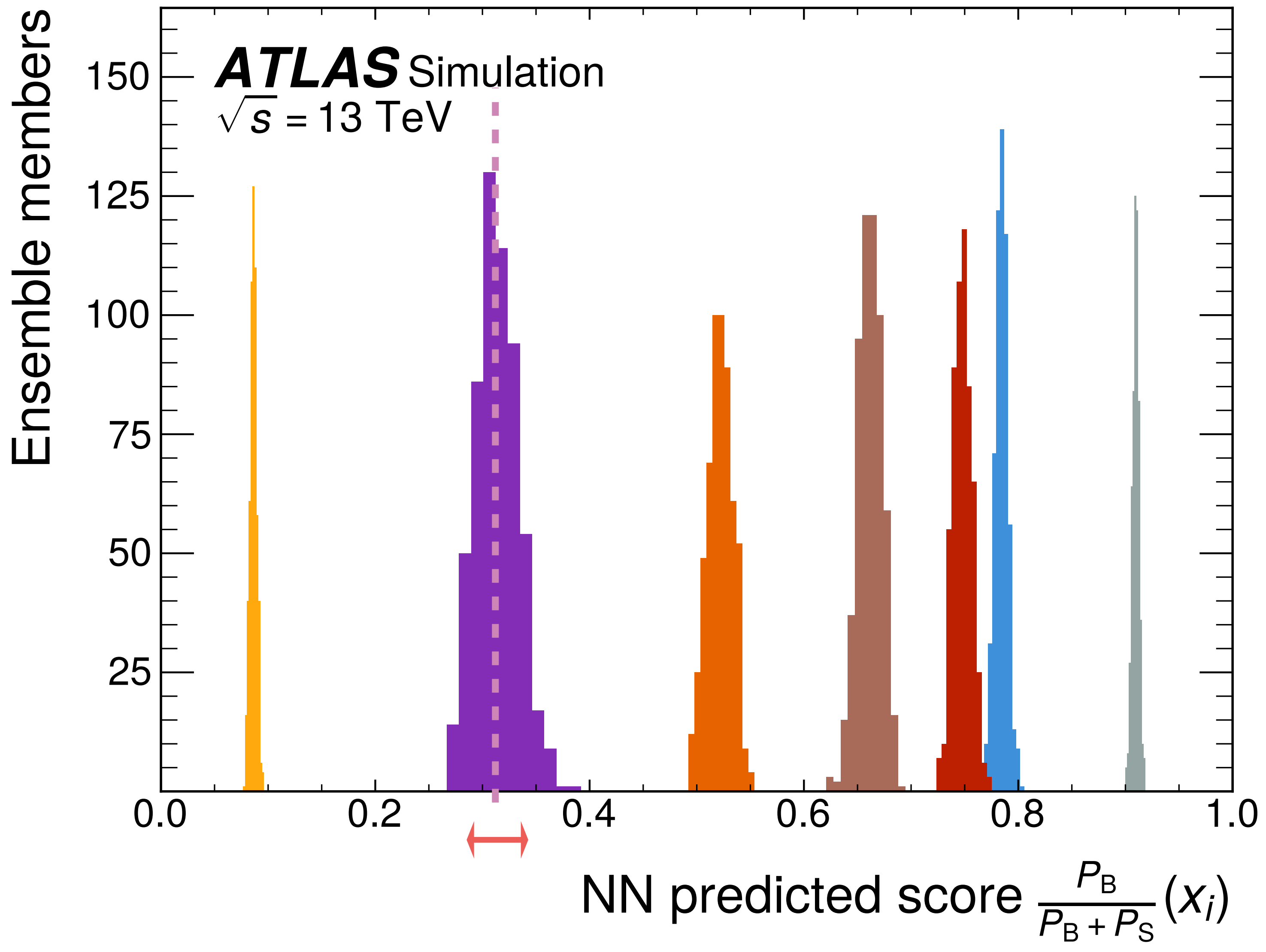
Frequentist Uncertainties on Neural Density Ratios with $w_i f_i$ Ensembles

Sean Benevedes[1, 2, *] and Jesse Thaler[1, 2, †]

[1] *Center for Theoretical Physics, Massachusetts Institute of Technology,*
Cambridge, Massachusetts, United States
[2] *The NSF AI Institute for Artificial Intelligence and Fundamental Interactions*



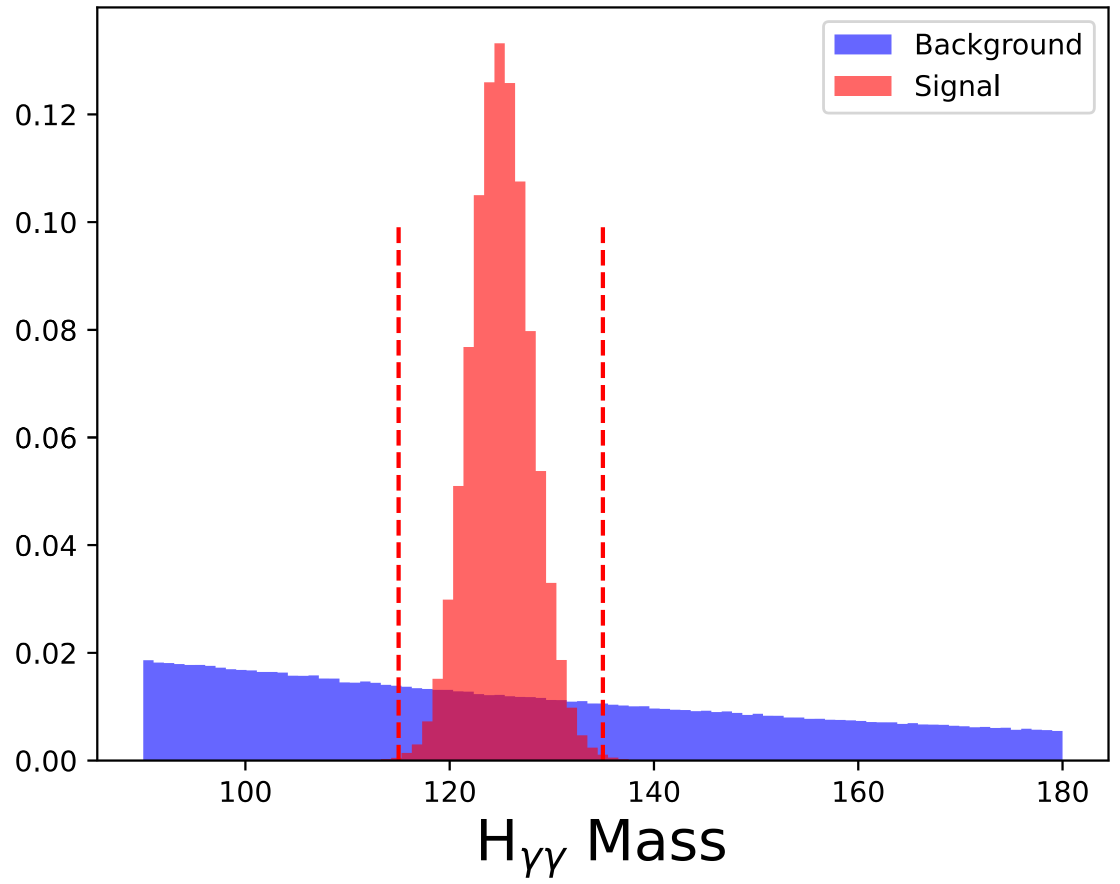
Distribution of NN predictions for example events

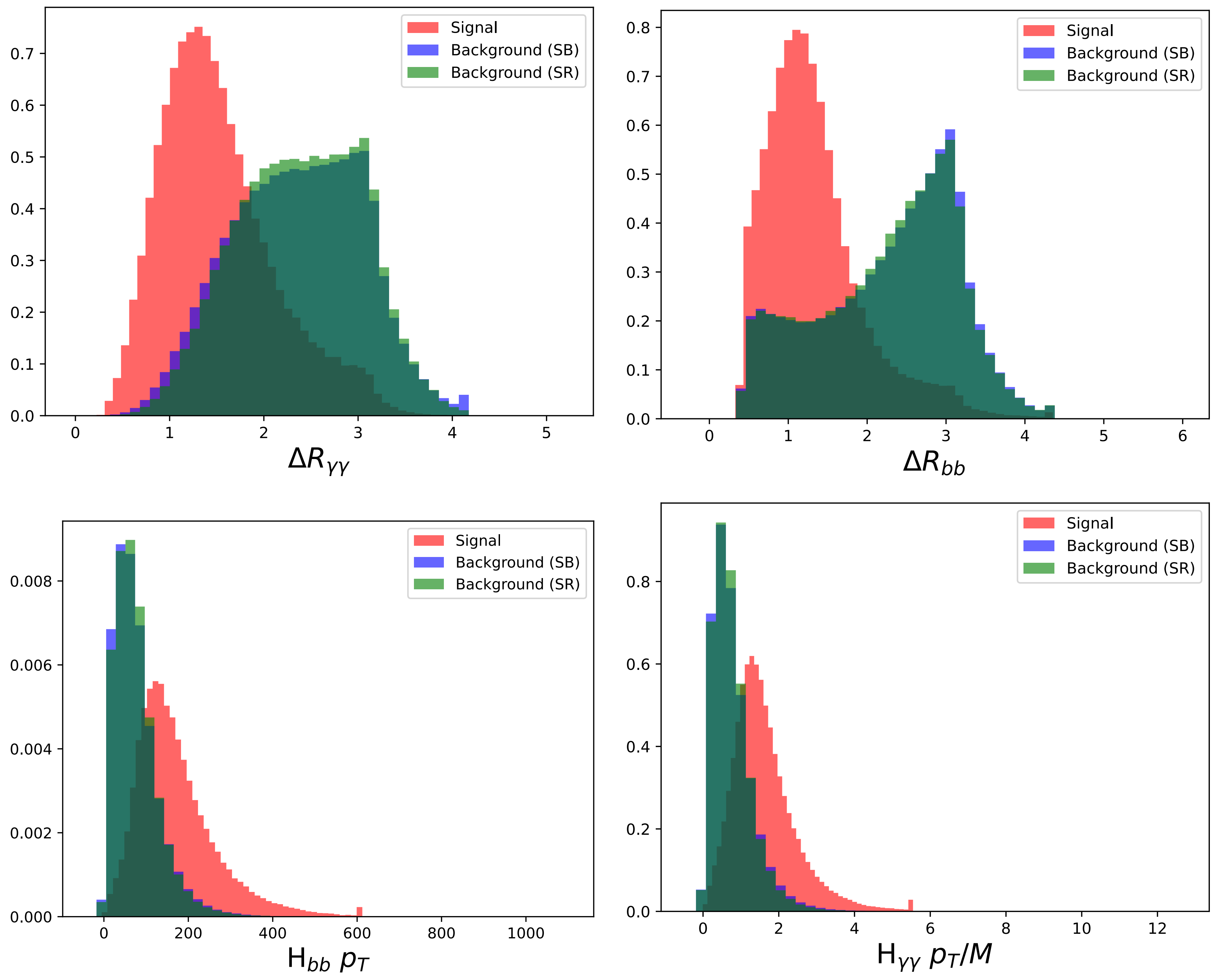


- Genuine 95% coverage, with much smaller ensembles
- Elegant, mathematically motivated method to estimate uncertainties

arXiv:2506.00113: Benevedes & Thaler

Unreliable simulator ?





- Training data can come from **control regions of real data**
- Data-driven background estimation techniques work with NSBI
- Amram & Szewc uses generative models instead of classifiers for NSBI

arXiv:2506.06438: Amram & Szewc

Until now, we have replaced individual pieces with ML in age-old likelihood ratio test

Do we dare question the test itself?

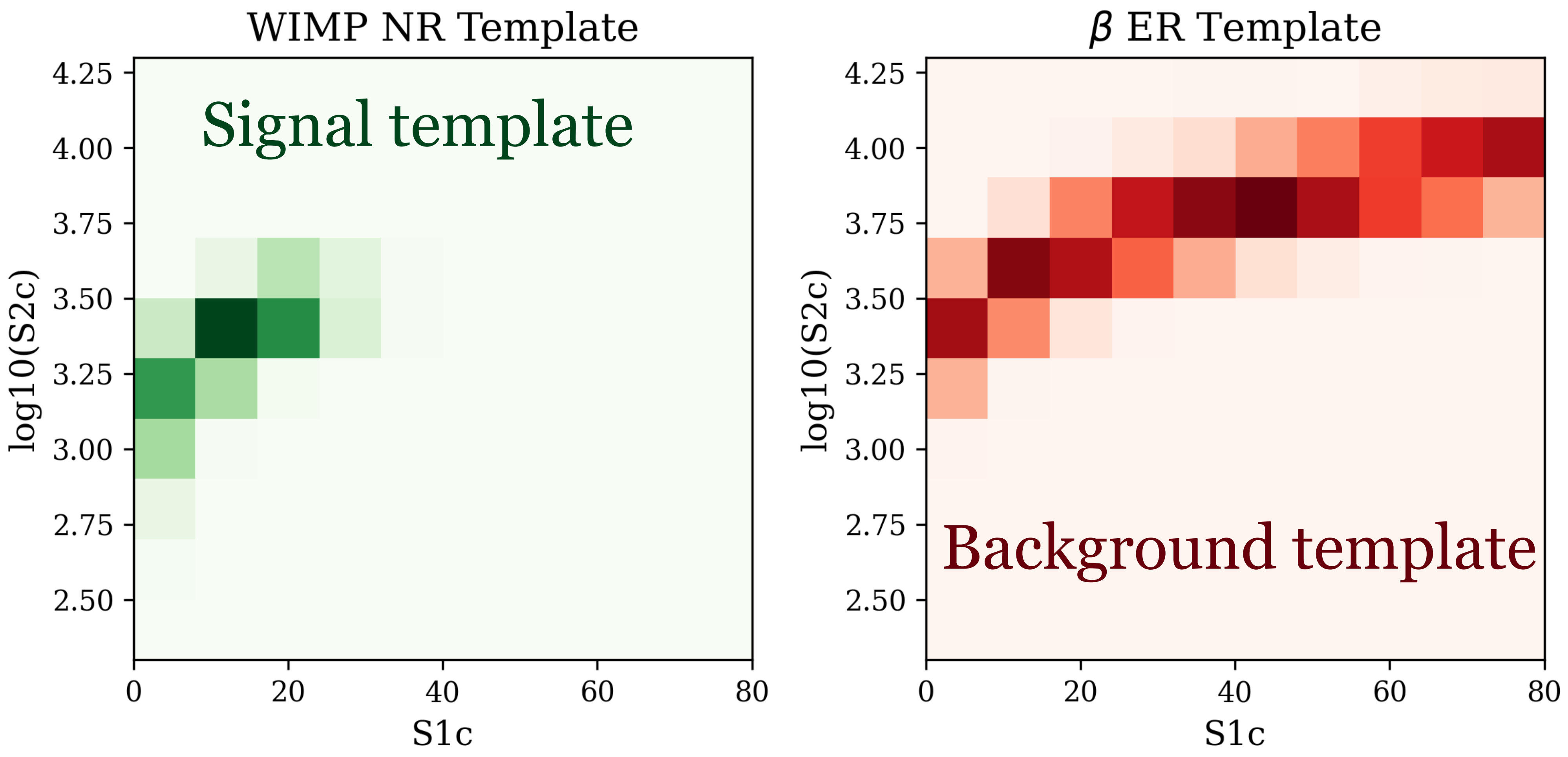
Until now, we have replaced individual pieces with ML in age-old likelihood ratio test

Do we dare question the test itself?

FTS backup slides

Now a completely different experiment, LZ for dark matter searches!

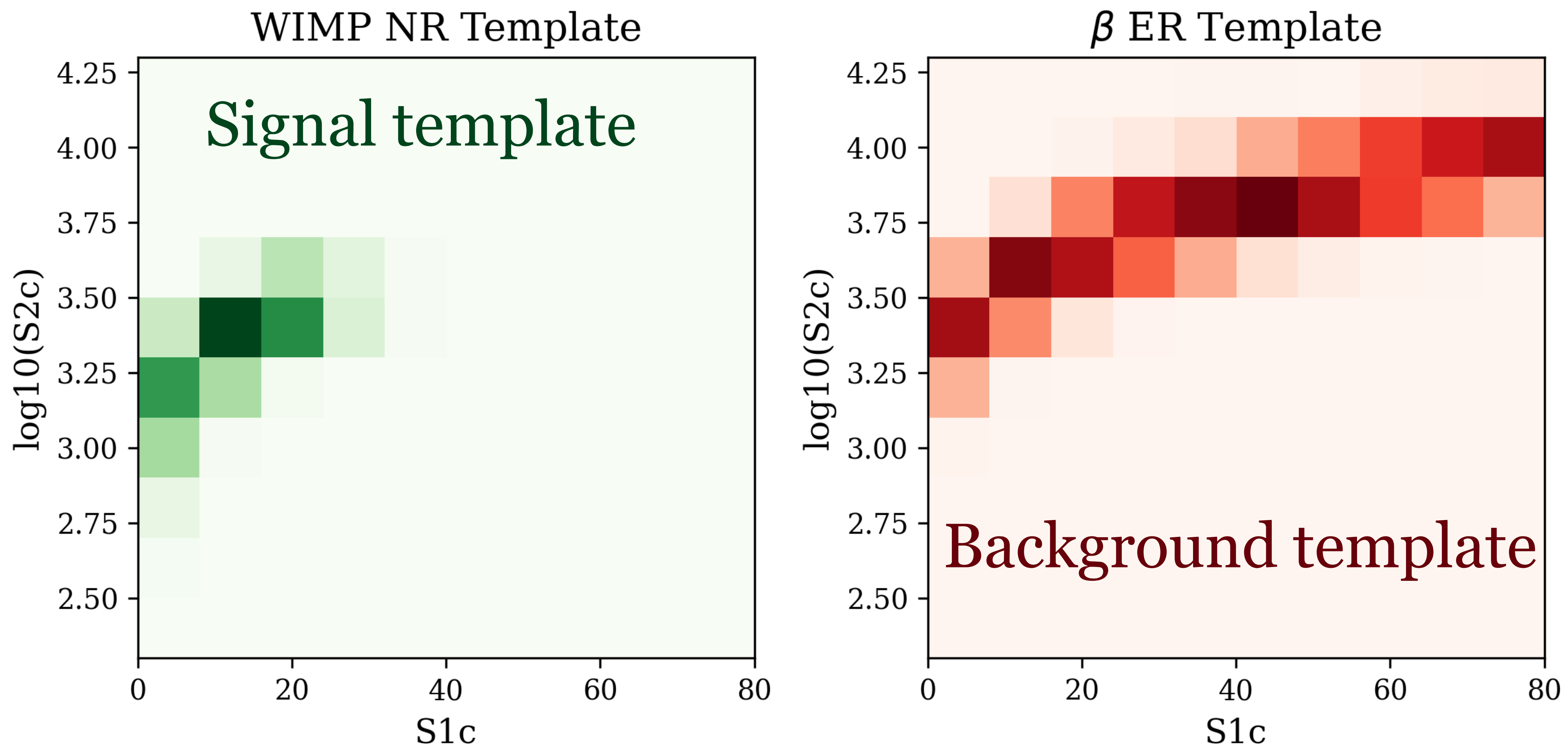
Second case study: WIMPs search in LZ experiment



Simulated to mimic latest LZ data

arXiv:2507.17831: Carzon, Ghosh et al

Second case study: WIMPs search in LZ experiment

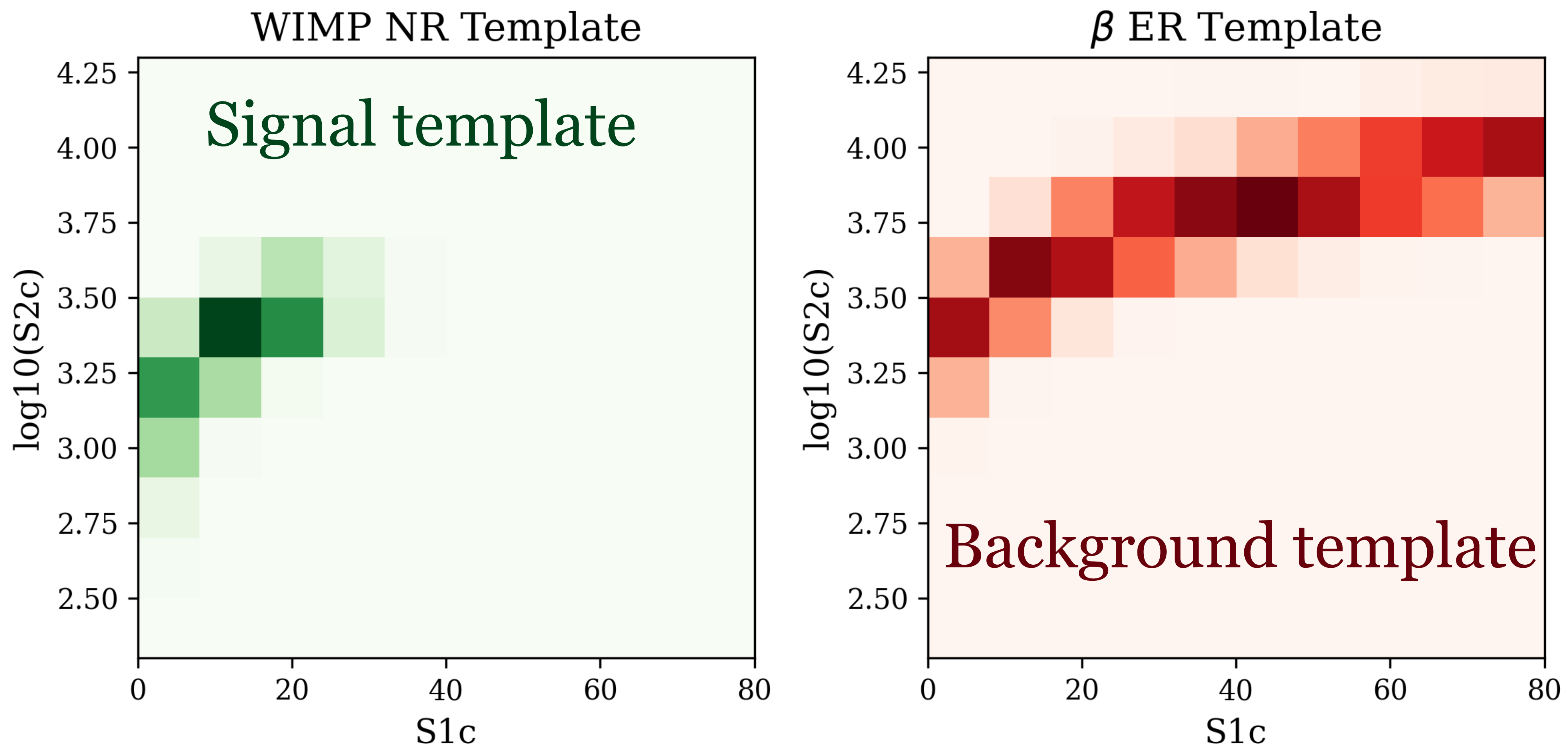


Simulated to mimic latest LZ data

Focus now placed at $\mu = 0$, we want to set upper limits on WIMP signal

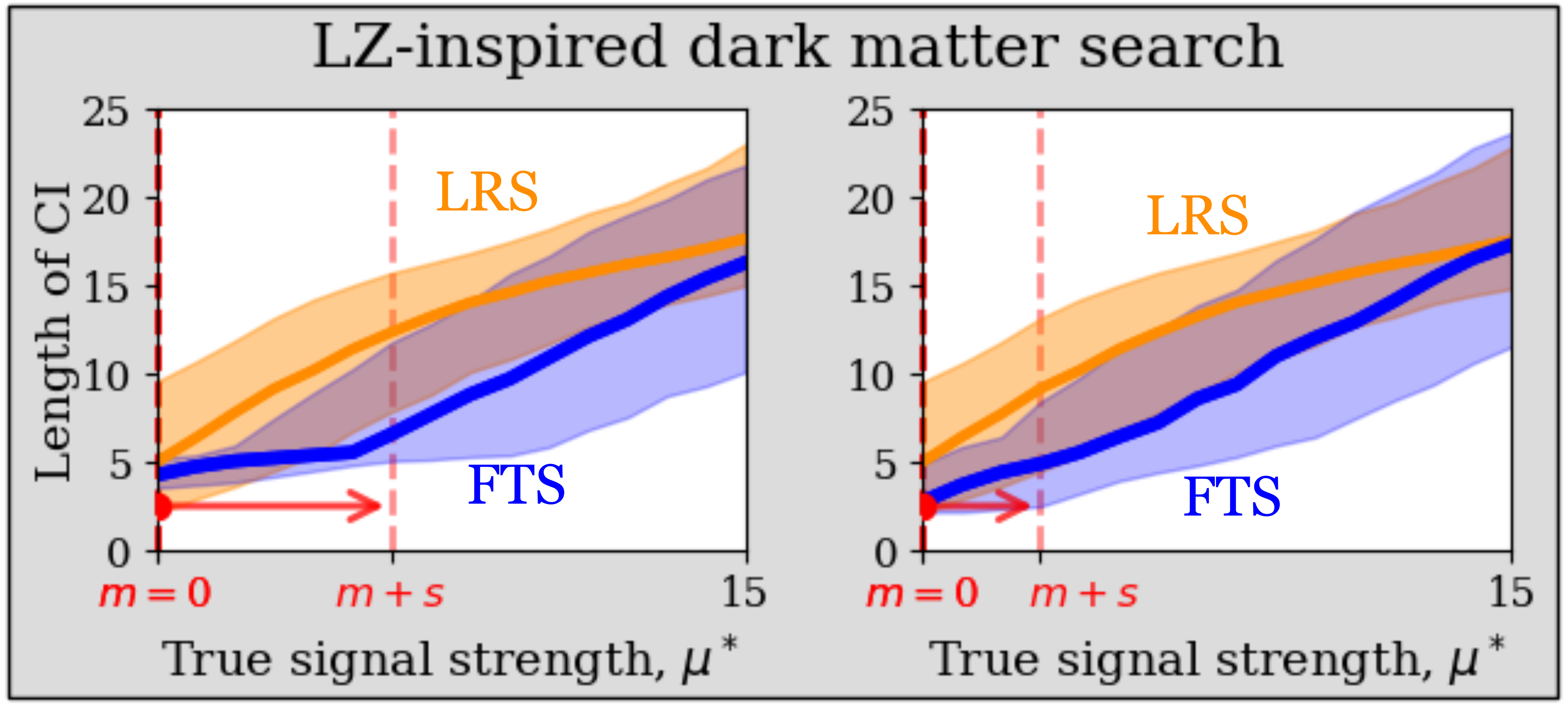
arXiv:2507.17831: Carzon, Ghosh et al

Second case study: WIMPs search in LZ experiment



Simulated to mimic latest LZ data

Focus now placed at $\mu = 0$, we want to set upper limits on WIMP signal

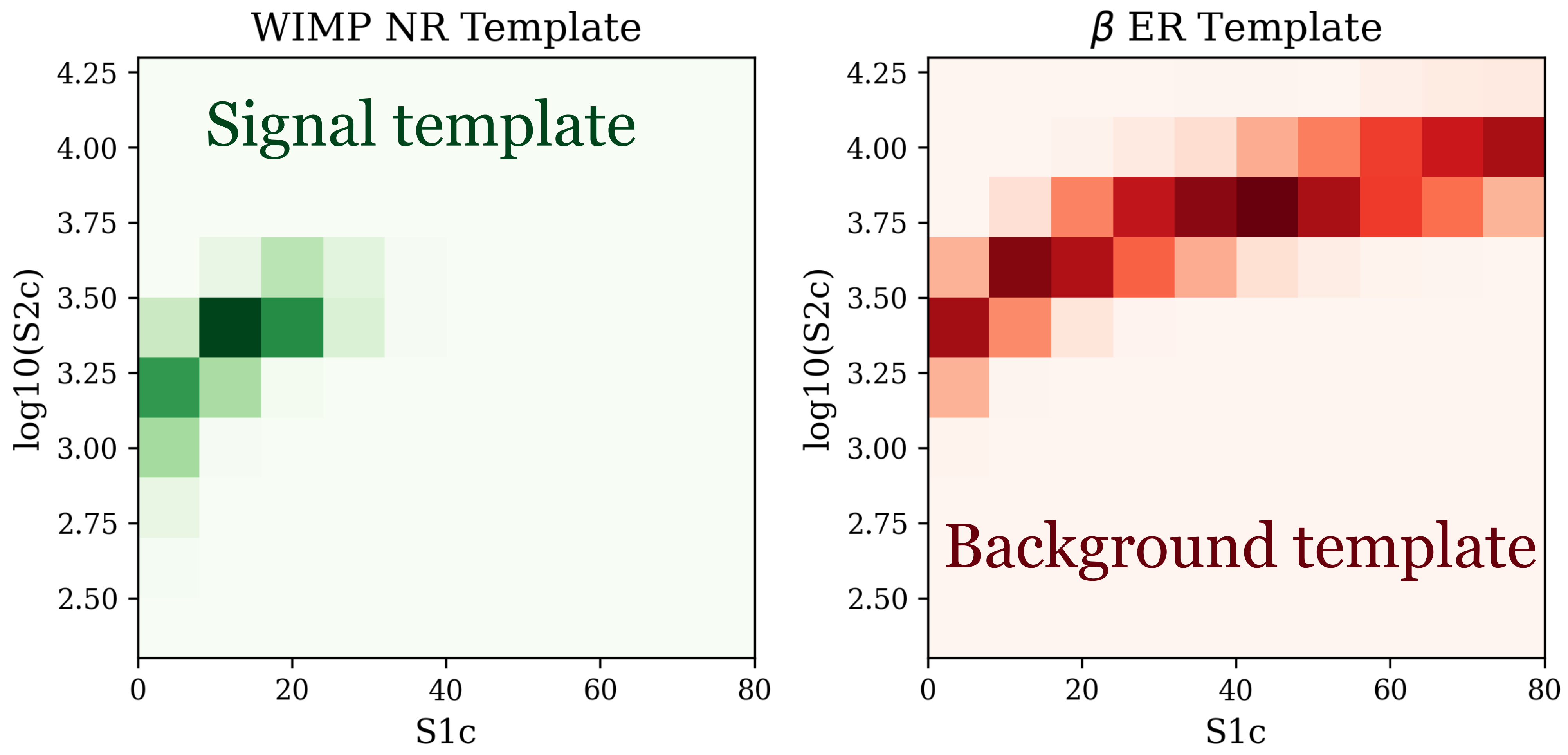


Wide focus

Narrow focus

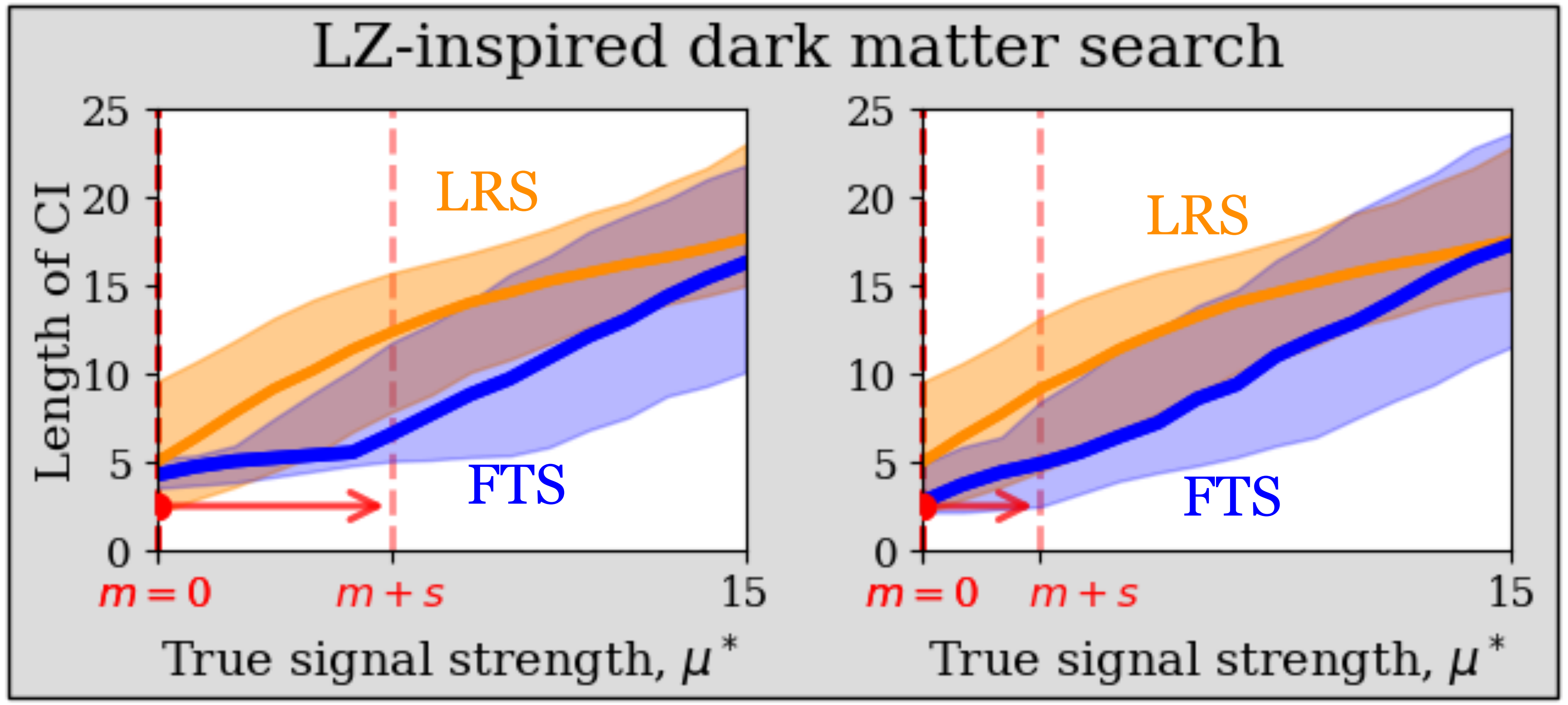
arXiv:2507.17831: Carzon, Ghosh et al

Second case study: WIMPs search in LZ experiment



Simulated to mimic latest LZ data

Focus now placed at $\mu = 0$, we want to set upper limits on WIMP signal



Wide focus

Narrow focus

FTS gains stable over large parameter space for both focus functions

More precision even when there is actually a signal in the observed data ($\mu > 0$)

arXiv:2507.17831: Carzon, Ghosh et al

Quantitative improvements with FTS

Using reconstructed Higgs mass as observable to construct histogram

Using visible energy as observable

Setting		Test statistic		
Experiment	μ^*	LRS	FTS-wide	FTS-narrow
Higgs (mass)	1.0	1.08 (2.01)	0.94 (1.89)	**0.85 (1.79)**
Higgs (vis. mass)	1.0	1.26 (2.31)	1.10 (2.17)	**0.98 (2.08)**
LZ-inspired	0.0	5.99 (13.87)	4.68 (11.47)	**3.89 (11.29)**
LZ-inspired	1.0	7.08 (15.18)	5.29 (12.82)	**4.68 (12.75)**

Numbers represent length of confidence intervals for 1σ (2σ)

arXiv:2507.17831: Carzon, Ghosh et al

Quantitative improvements with FTS

Using reconstructed Higgs mass as observable to construct histogram

Using visible energy as observable

Setting		Test statistic		
Experiment	μ^*	LRS	FTS-wide	FTS-narrow
Higgs (mass)	1.0	1.08 (2.01)	0.94 (1.89)	**0.85 (1.79)**
Higgs (vis. mass)	1.0	1.26 (2.31)	1.10 (2.17)	**0.98 (2.08)**
LZ-inspired	0.0	5.99 (13.87)	4.68 (11.47)	**3.89 (11.29)**
LZ-inspired	1.0	7.08 (15.18)	5.29 (12.82)	**4.68 (12.75)**

Numbers represent length of confidence intervals for 1σ (2σ)

Would be interesting to study gains for specific analyses!

arXiv:2507.17831: Carzon, Ghosh et al

How to choose a focus function?

Conservative strategy:

- Could choose based on previous measurement (ATLAS example used this strategy for ‘wide focus’)
- Could come from theoretical / conceptual argument, eg. An angle must be within $[0,2\pi]$

Optimised strategy:

- Study sensitivity on simulated samples for a range of focus
- Choose based on your physics priorities, eg. measurements and searches will have different strategies
- Must freeze focus function before unblinding data

Do you have more wacky ideas for a focus?

arXiv:2507.17831: Carzon, Ghosh et al

Lower and upper bounds for LZ example

arXiv:2507.17831: Carzon, Ghosh et al

Lower and upper bounds for LZ example

Consistently larger lower bounds if signal exits, smaller upper bounds when it doesn’t

Upper bounds when there is no signal:

μ^*	LRS	FTS, wide	FTS, narrow
0.0	4.97 (12.59)	4.25 (**10.98**)	**2.73** (11.14)

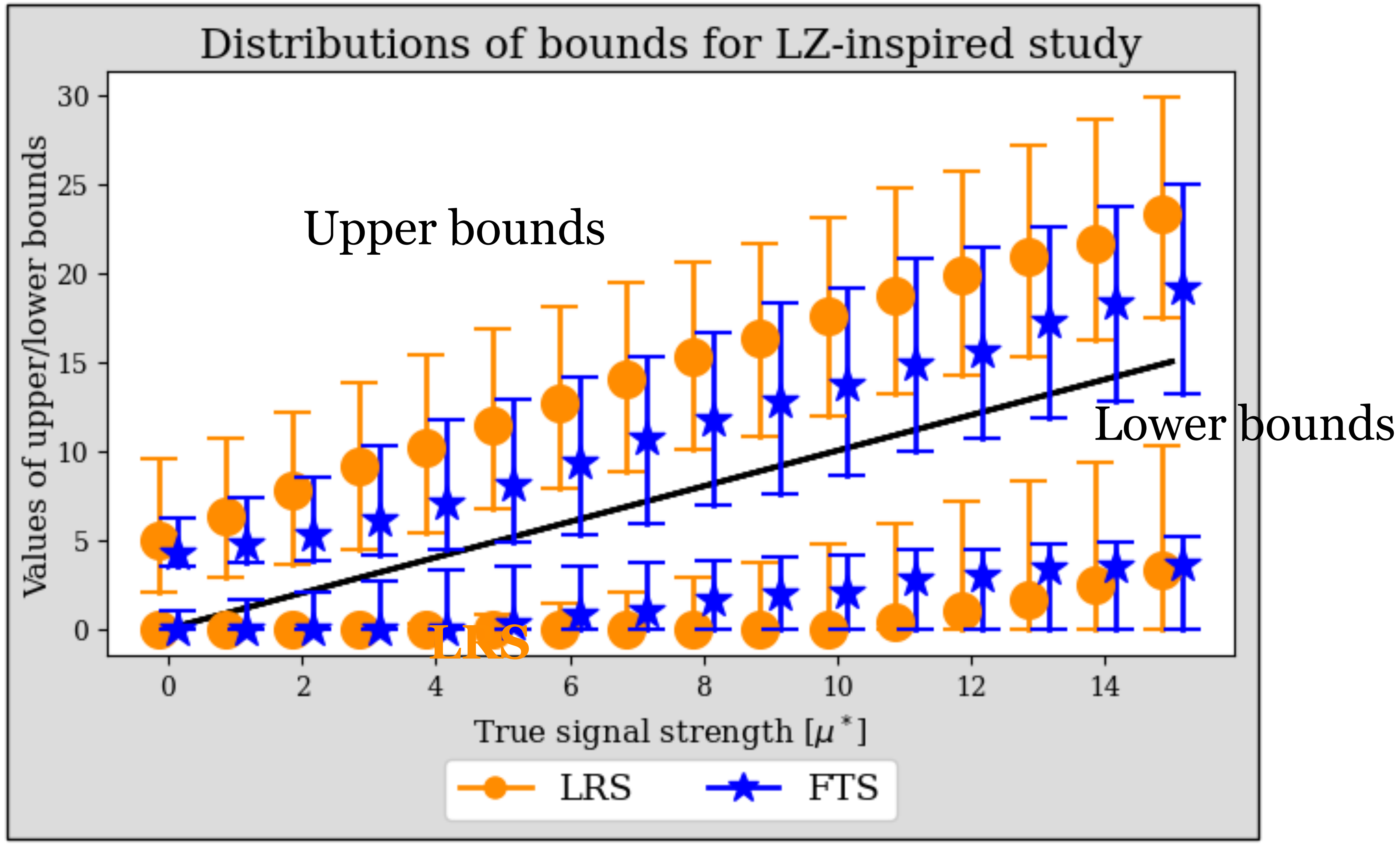
Lower bounds when there is a signal:

μ^*	LRS	FTS, wide	FTS, narrow
10.0	0.0 (0.0)	**2.08** (0.0)	1.60 (0.0)

Numbers represent length of confidence intervals for 1σ (2σ)

arXiv:2507.17831: Carzon, Ghosh et al

Lower and upper bounds for LZ example



Consistently larger lower bounds if signal exits, smaller upper bounds when it doesn’t

Upper bounds when there is no signal:

μ^*	LRS	FTS, wide	FTS, narrow
0.0	4.97 (12.59)	4.25 (**10.98**)	**2.73** (11.14)

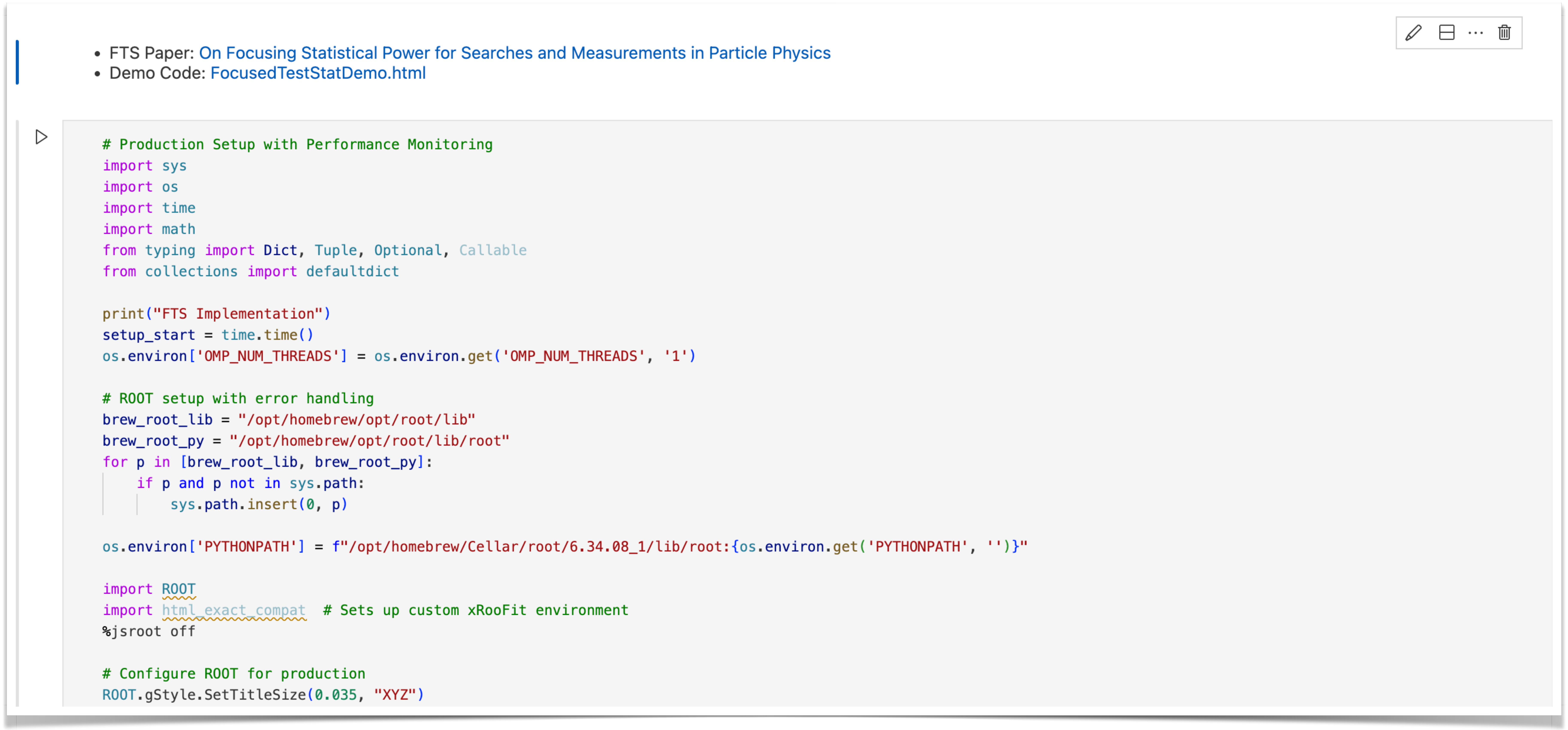
Lower bounds when there is a signal:

μ^*	LRS	FTS, wide	FTS, narrow
10.0	0.0 (0.0)	**2.08** (0.0)	1.60 (0.0)

Numbers represent length of confidence intervals for 1σ (2σ)

RooFit integration of FTS in development

FTS soon to be available in convenient software packages like RooFit



- FTS Paper: On Focusing Statistical Power for Searches and Measurements in Particle Physics
- Demo Code: FocusedTestStatDemo.html

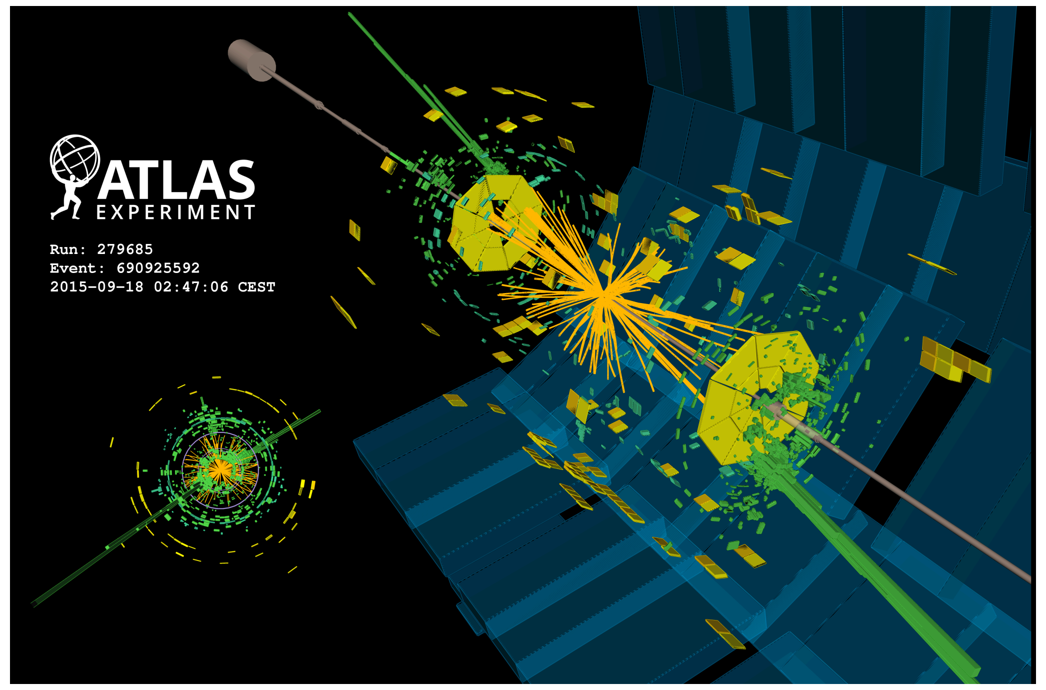
```
# Production Setup with Performance Monitoring
import sys
import os
import time
import math
from typing import Dict, Tuple, Optional, Callable
from collections import defaultdict

print("FTS Implementation")
setup_start = time.time()
os.environ['OMP_NUM_THREADS'] = os.environ.get('OMP_NUM_THREADS', '1')

# ROOT setup with error handling
brew_root_lib = "/opt/homebrew/opt/root/lib"
brew_root_py = "/opt/homebrew/opt/root/lib/root"
for p in [brew_root_lib, brew_root_py]:
    if p and p not in sys.path:
        sys.path.insert(0, p)

os.environ['PYTHONPATH'] = f"/opt/homebrew/Cellar/root/6.34.08_1/lib/root:{os.environ.get('PYTHONPATH', '')}"

import ROOT
import html_exact_compat  # Sets up custom xRooFit environment
%jsroot off

# Configure ROOT for production
ROOT.gStyle.SetTitleSize(0.035, "XYZ")
```

Currently led by students Jinbo Zhang & Daniel Zhang, based on xRooFit setup by Will Buttinger

What is a ‘Neyman construction’?

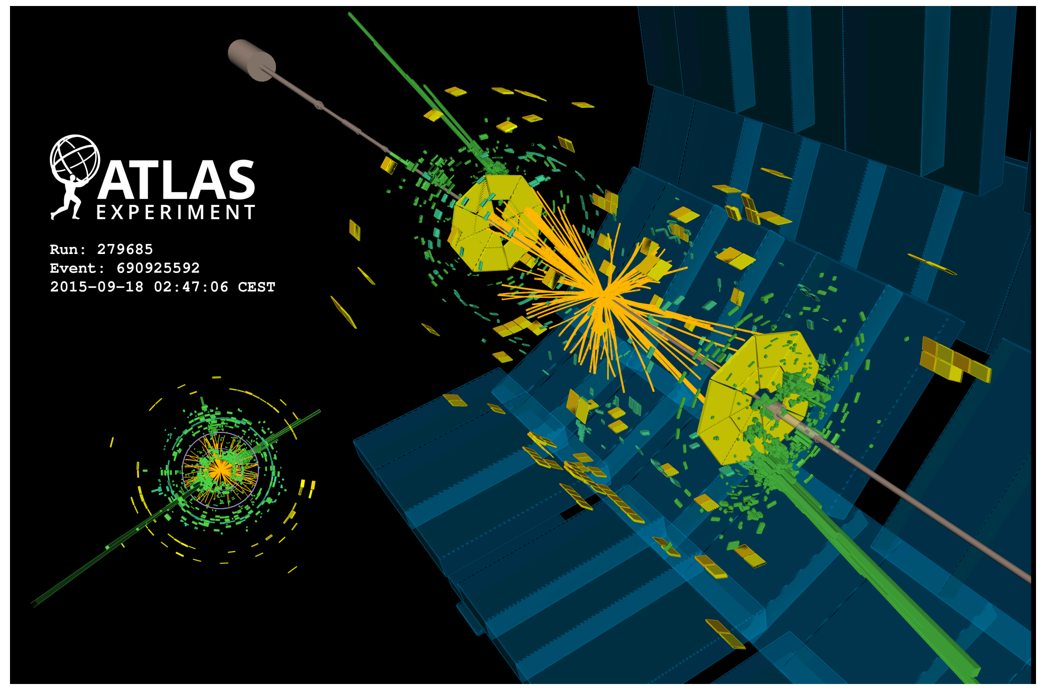
Obtaining the distribution of a test statistic

Raw data



Obtaining the distribution of a test statistic

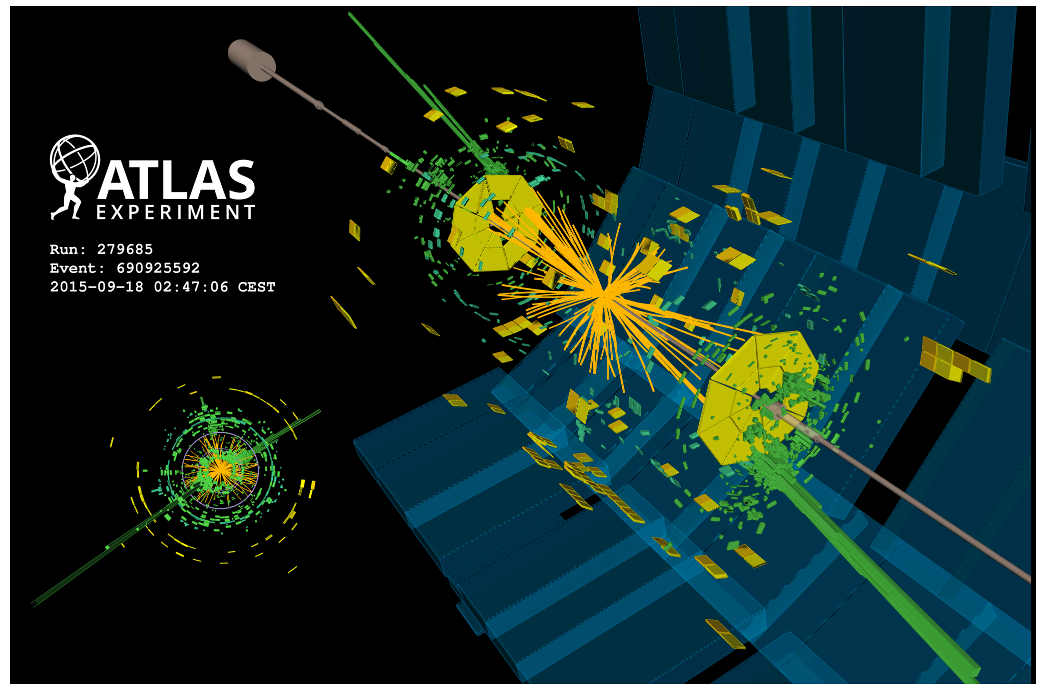
Raw data



Data processing

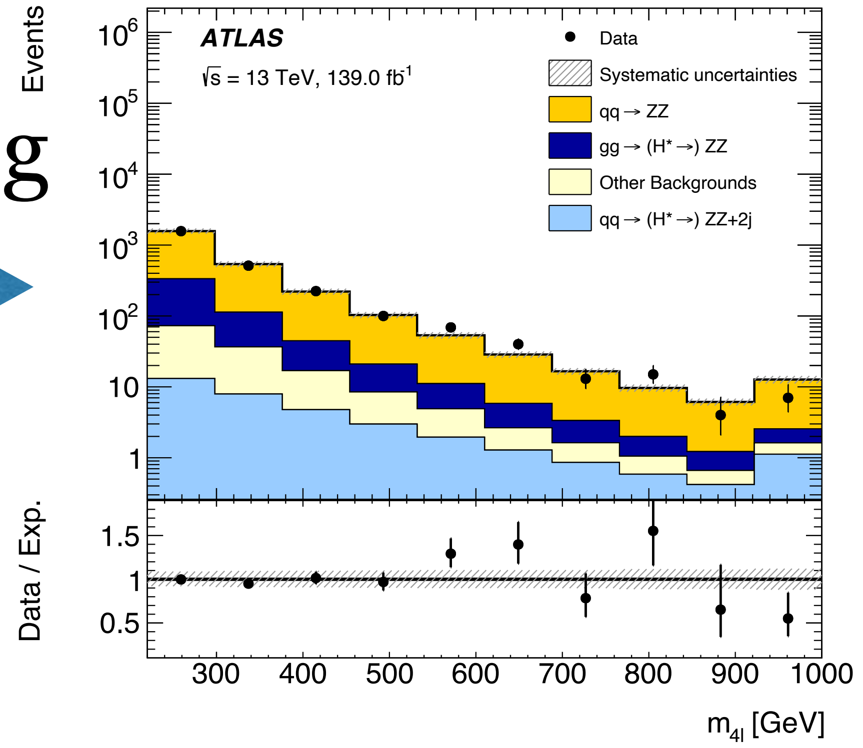
Obtaining the distribution of a test statistic

Raw data



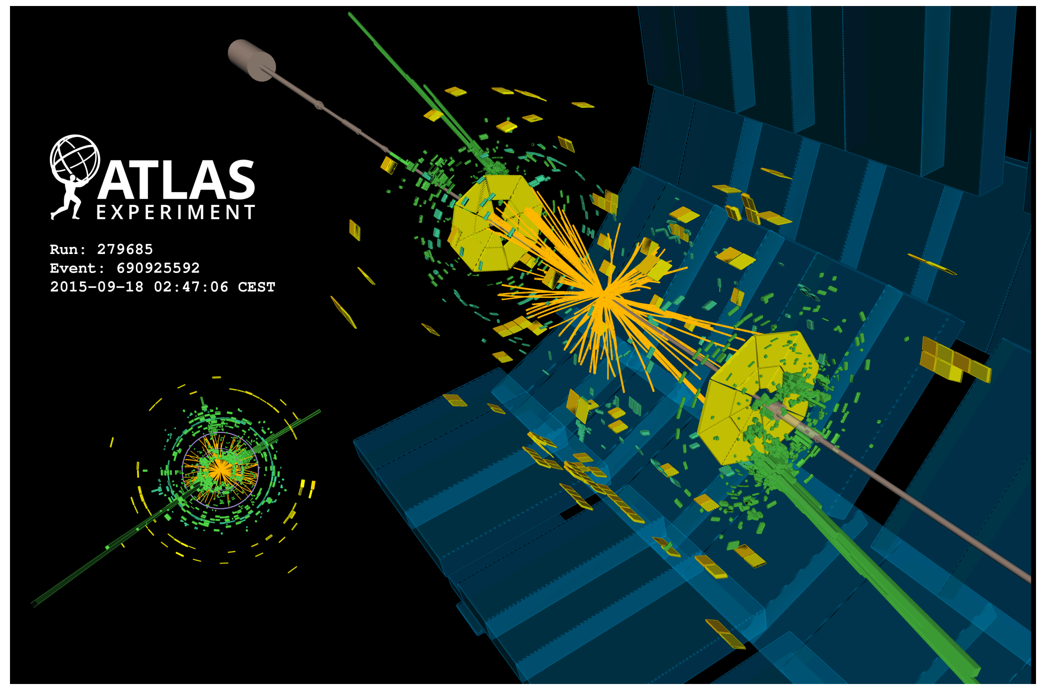
Data processing

Low-dimensional summary



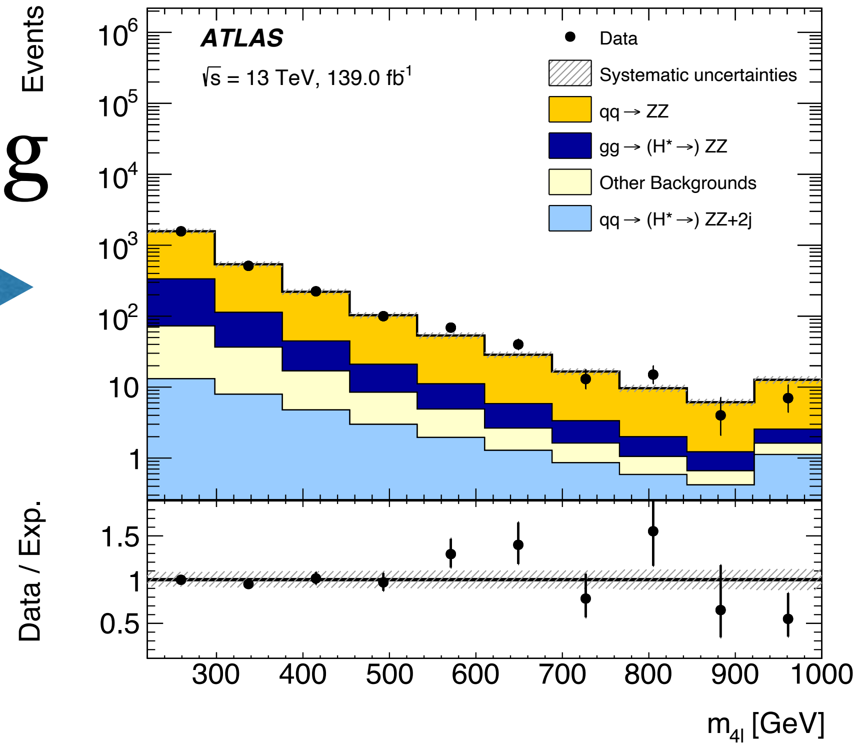
Obtaining the distribution of a test statistic

Raw data



Data processing

Low-dimensional summary

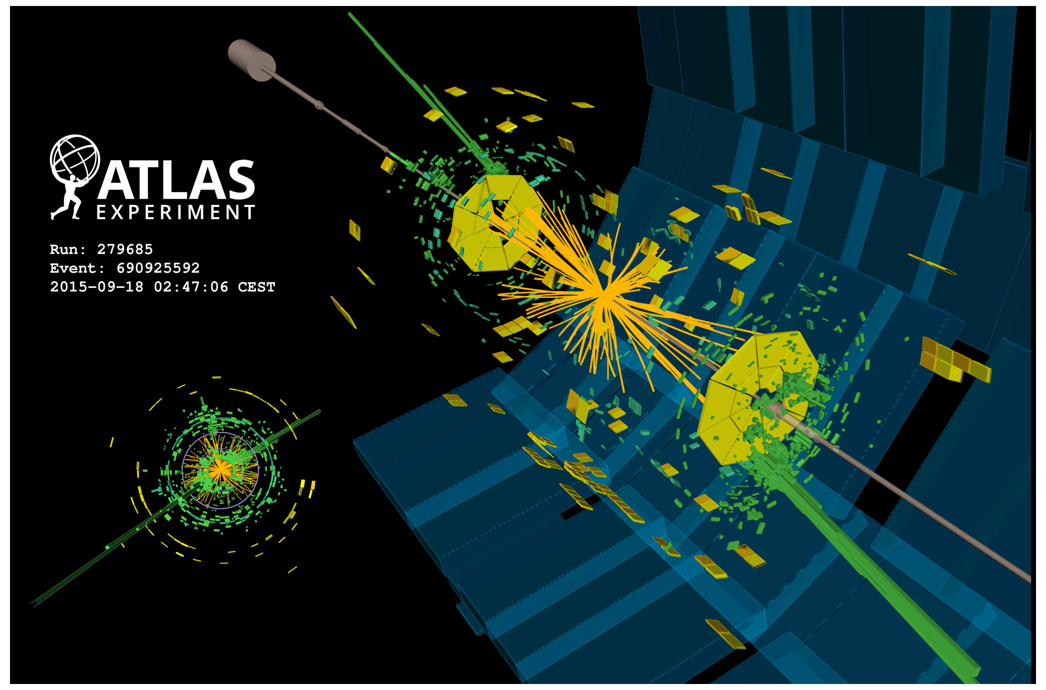


$$p_{bin}(n_{obs}|n_{exp}) = \frac{n_{exp}^{n_{obs}} \cdot e^{-n_{exp}}}{n_{obs}!}$$

Compare data to expectation, compute likelihood

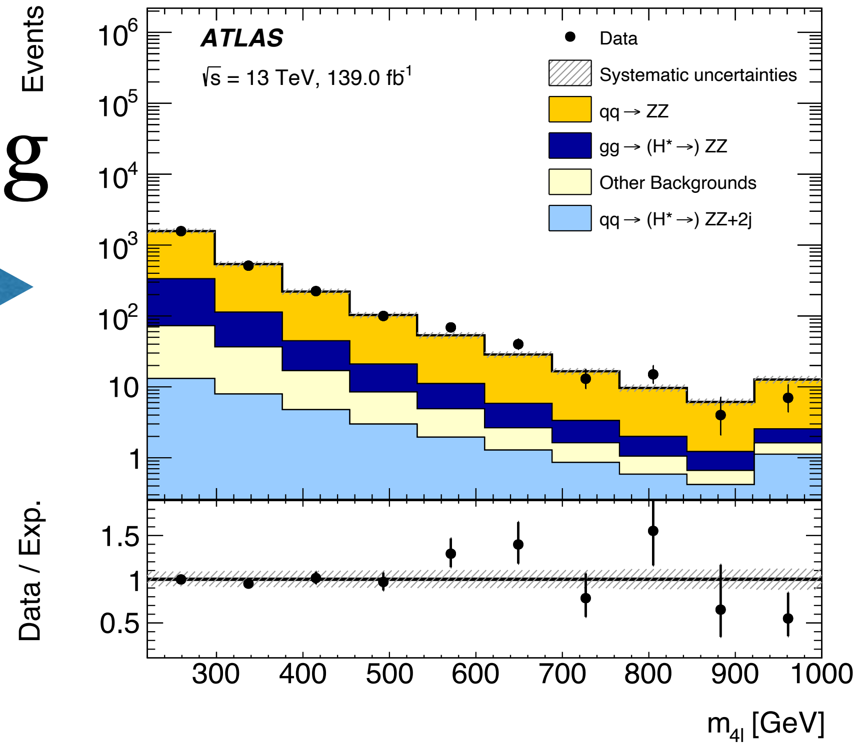
Obtaining the distribution of a test statistic

Raw data



Data processing

Low-dimensional summary



$$p_{bin}(n_{obs}|n_{exp}) = \frac{n_{exp}^{n_{obs}} \cdot e^{-n_{exp}}}{n_{obs}!}$$

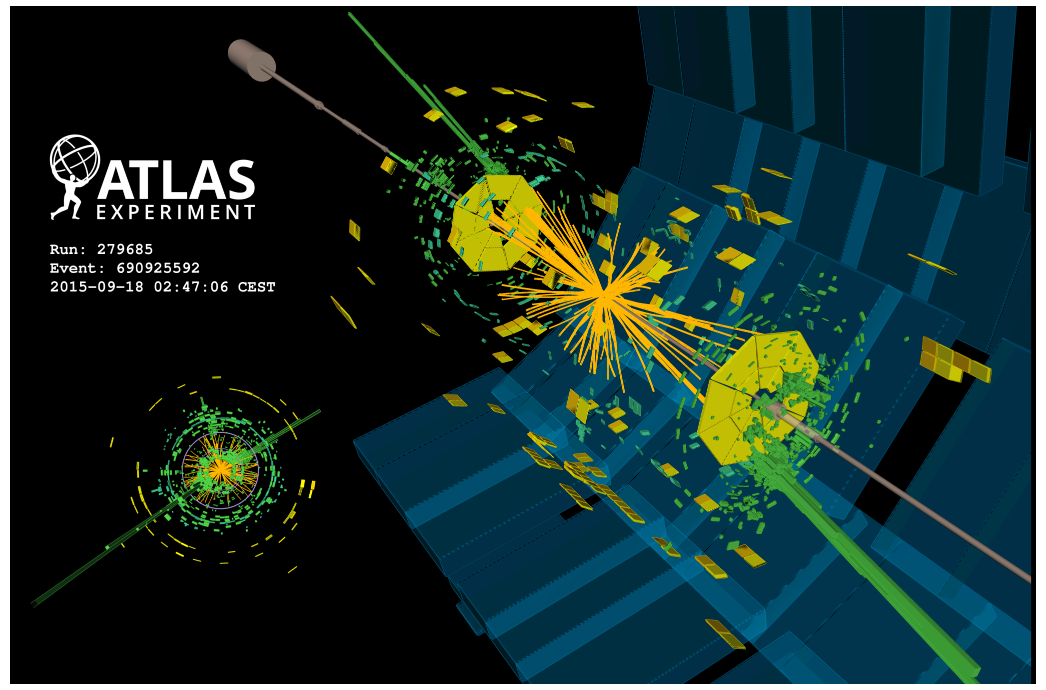
Compare data to expectation, compute likelihood

Obtain test statistic

$$t_\mu$$

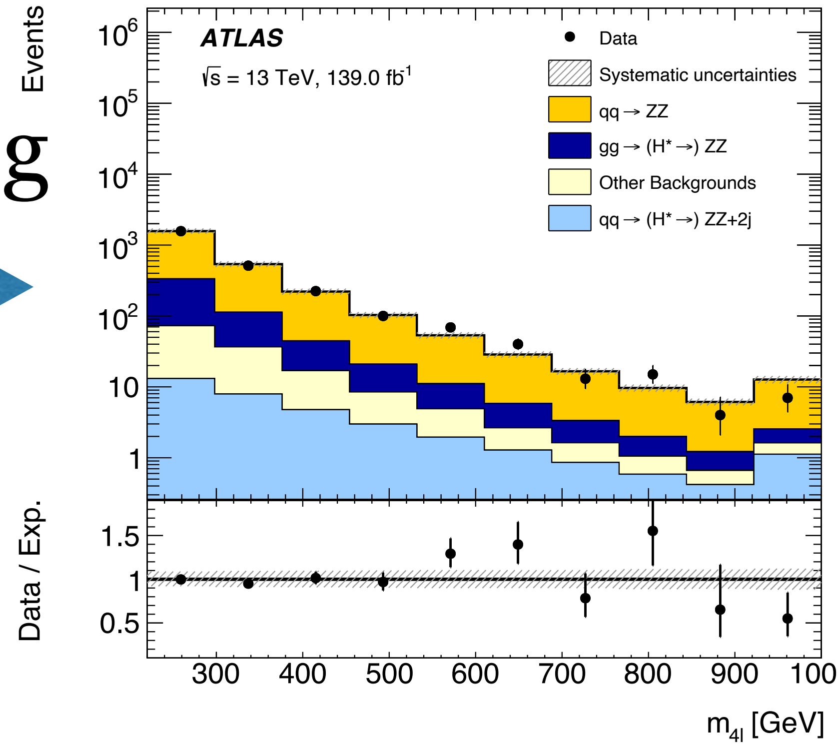
Obtaining the distribution of a test statistic

Raw data



Data processing

Low-dimensional summary



Obtain test statistic

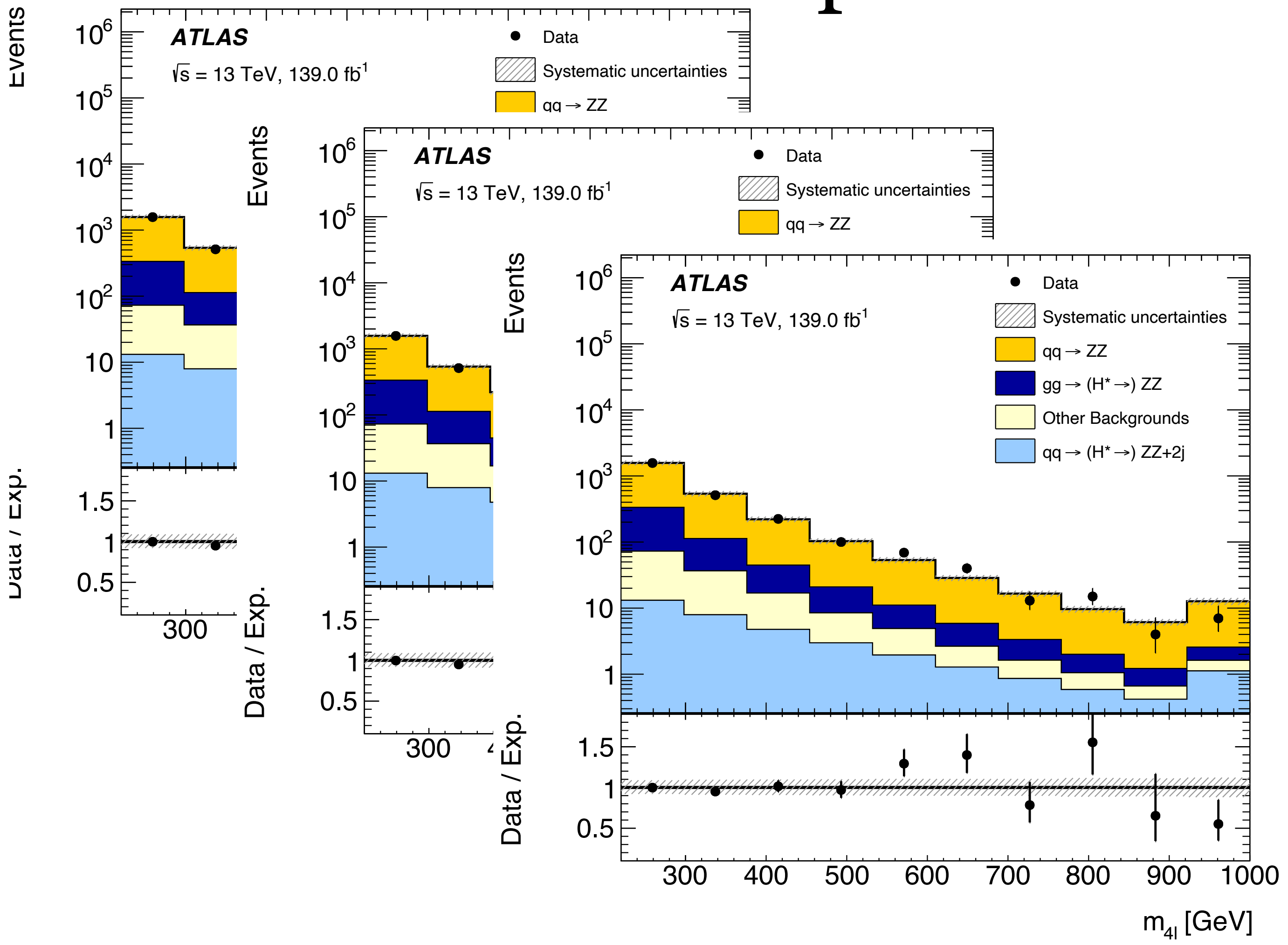
$$p_{bin}(n_{obs}|n_{exp}) = \frac{n_{exp}^{n_{obs}} \cdot e^{-n_{exp}}}{n_{obs}!}$$

Compare data to expectation, compute likelihood

$$t_\mu$$

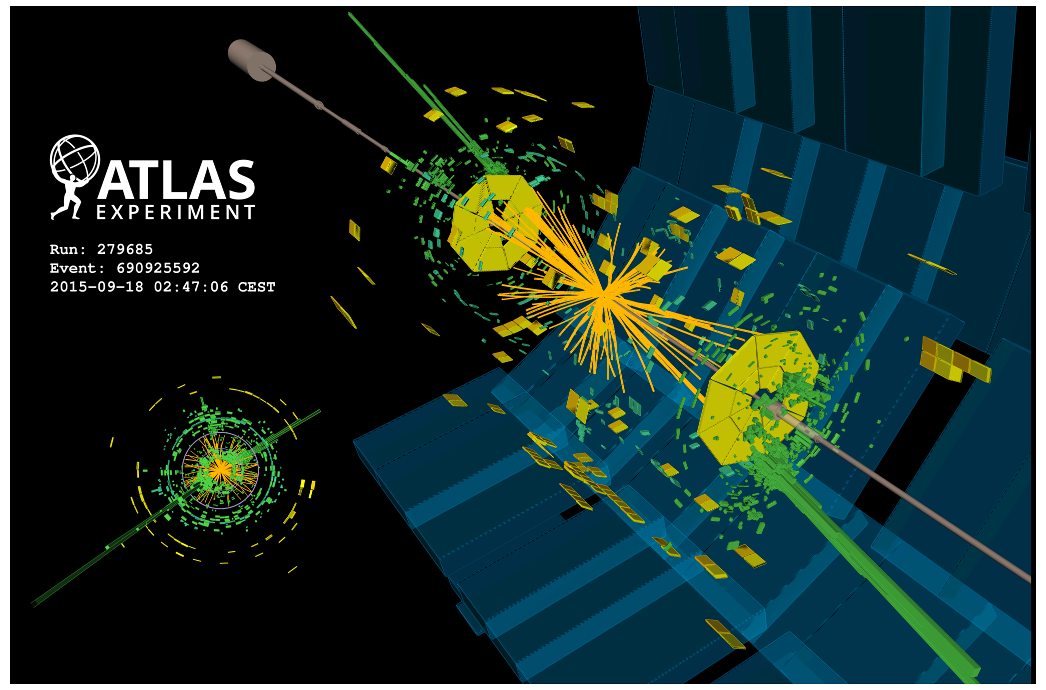
Create pseudo-experiments

From simulations for a fixed μ, eg. $\mu = 0$



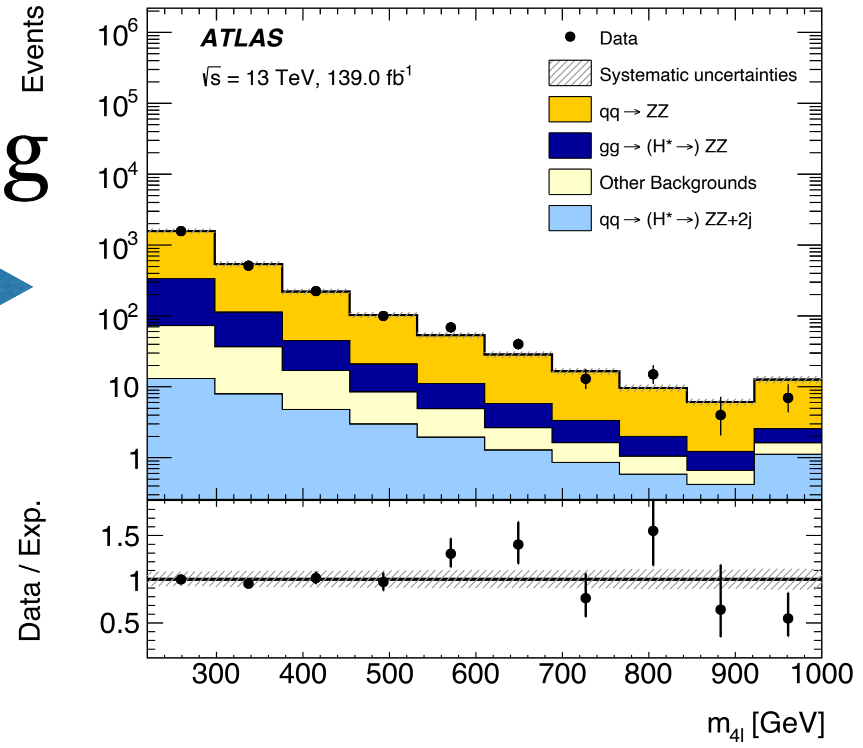
Obtaining the distribution of a test statistic

Raw data



Data processing

Low-dimensional summary



$$p_{bin}(n_{obs}|n_{exp}) = \frac{n_{exp}^{n_{obs}} \cdot e^{-n_{exp}}}{n_{obs}!}$$

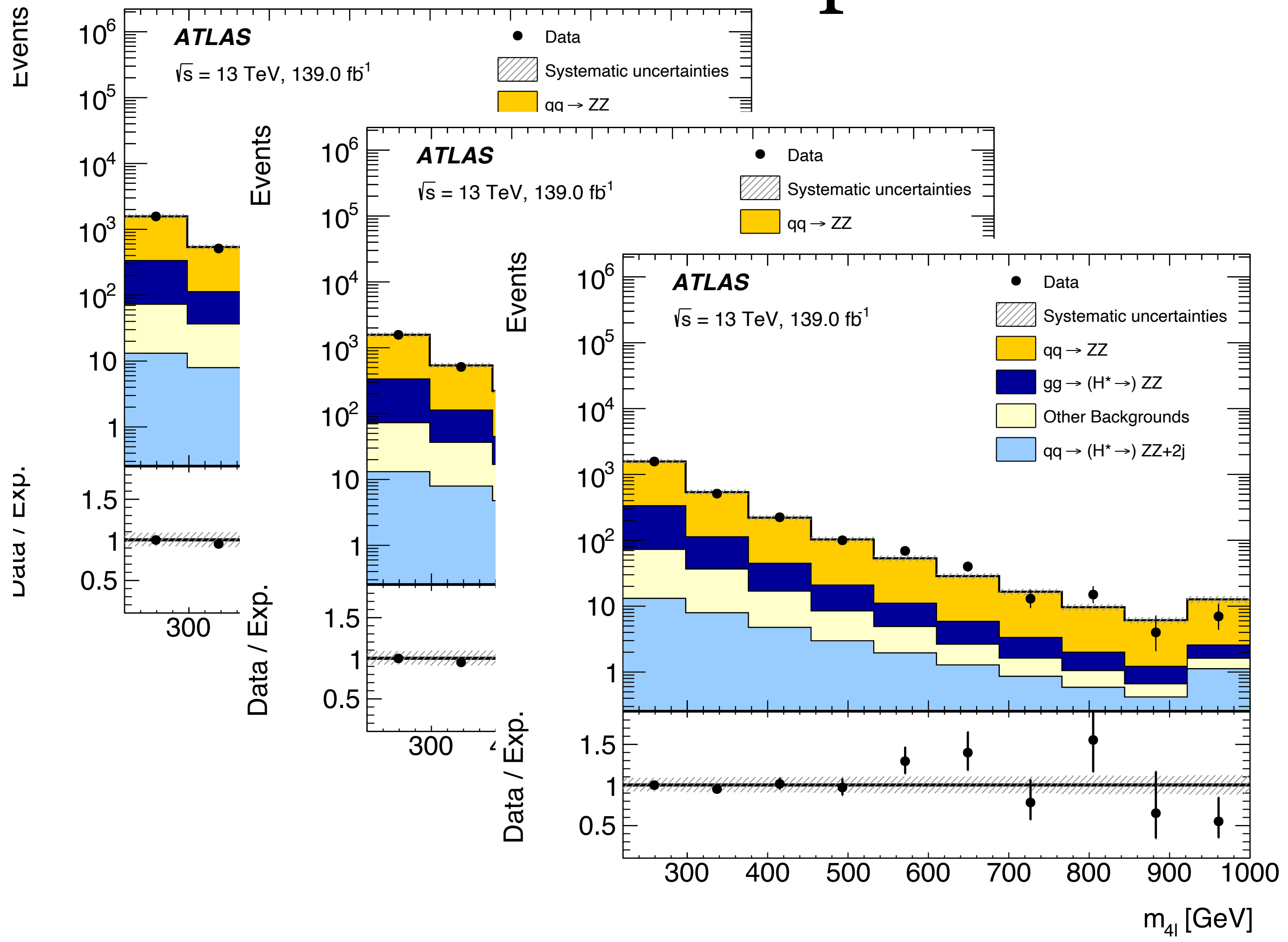
Compare data to expectation, compute likelihood

Obtain test statistic

$$t_\mu$$

From simulations for a fixed μ, eg. $\mu = 0$

Create pseudo-experiments

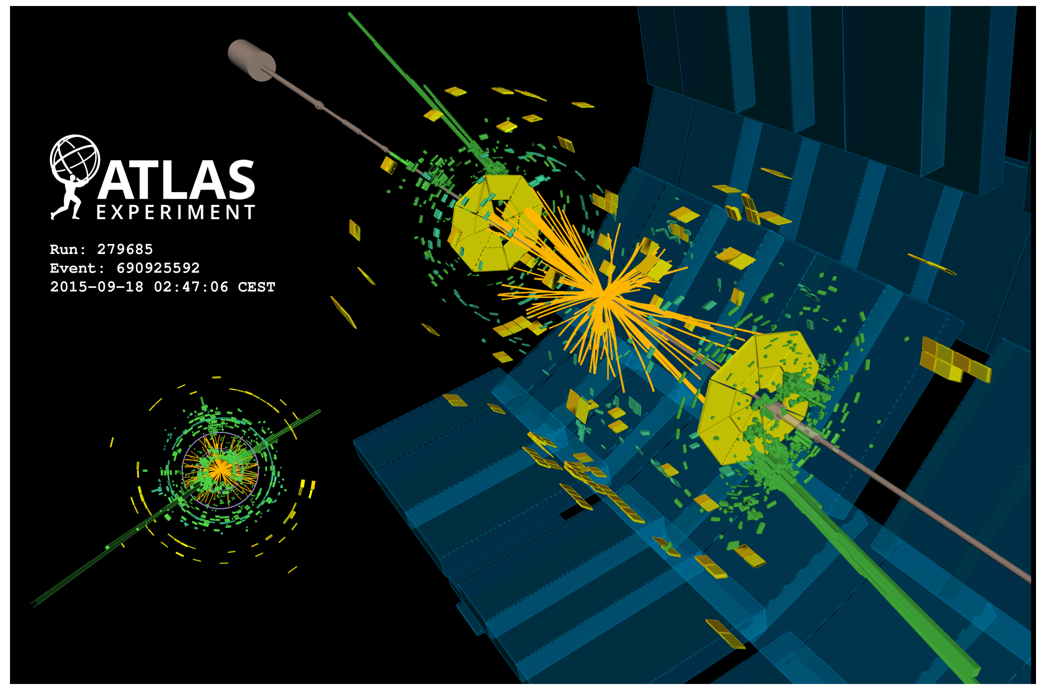


$$p_{bin}(n_{obs}|n_{exp}) = \frac{n_{exp}^{n_{obs}} \cdot e^{-n_{exp}}}{n_{obs}!}$$

Compare **pseudo-data** to expectation, compute likelihood

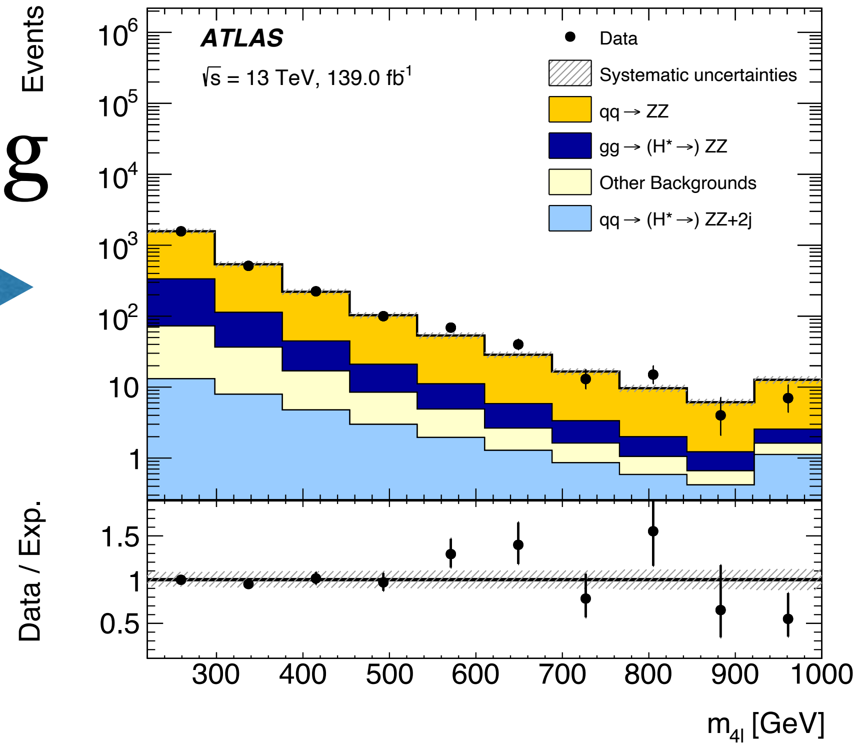
Obtaining the distribution of a test statistic

Raw data



Data processing

Low-dimensional summary



$$p_{bin}(n_{obs}|n_{exp}) = \frac{n_{exp}^{n_{obs}} \cdot e^{-n_{exp}}}{n_{obs}!}$$

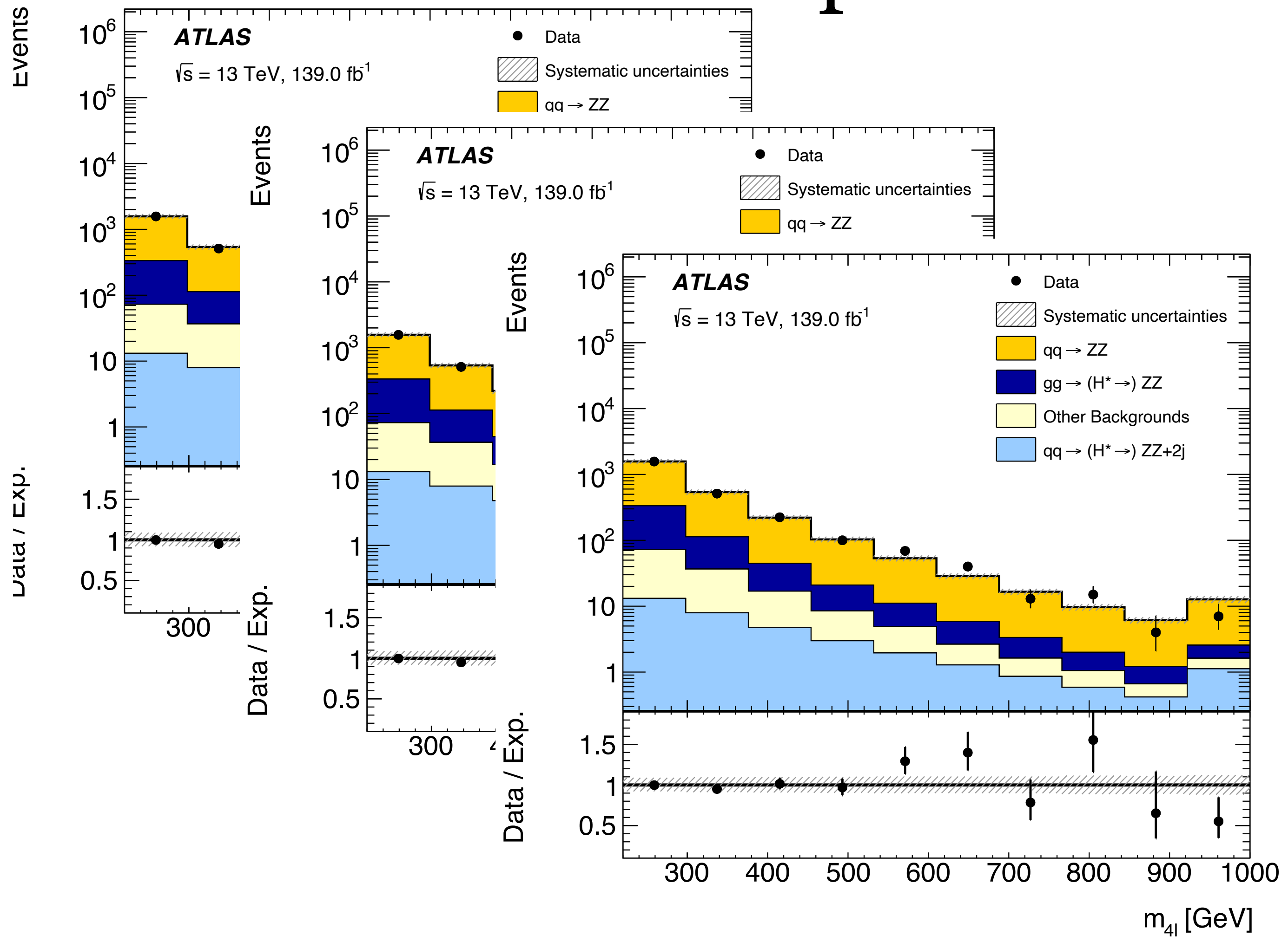
Compare data to expectation, compute likelihood

Obtain test statistic

$$t_\mu$$

Create pseudo-experiments

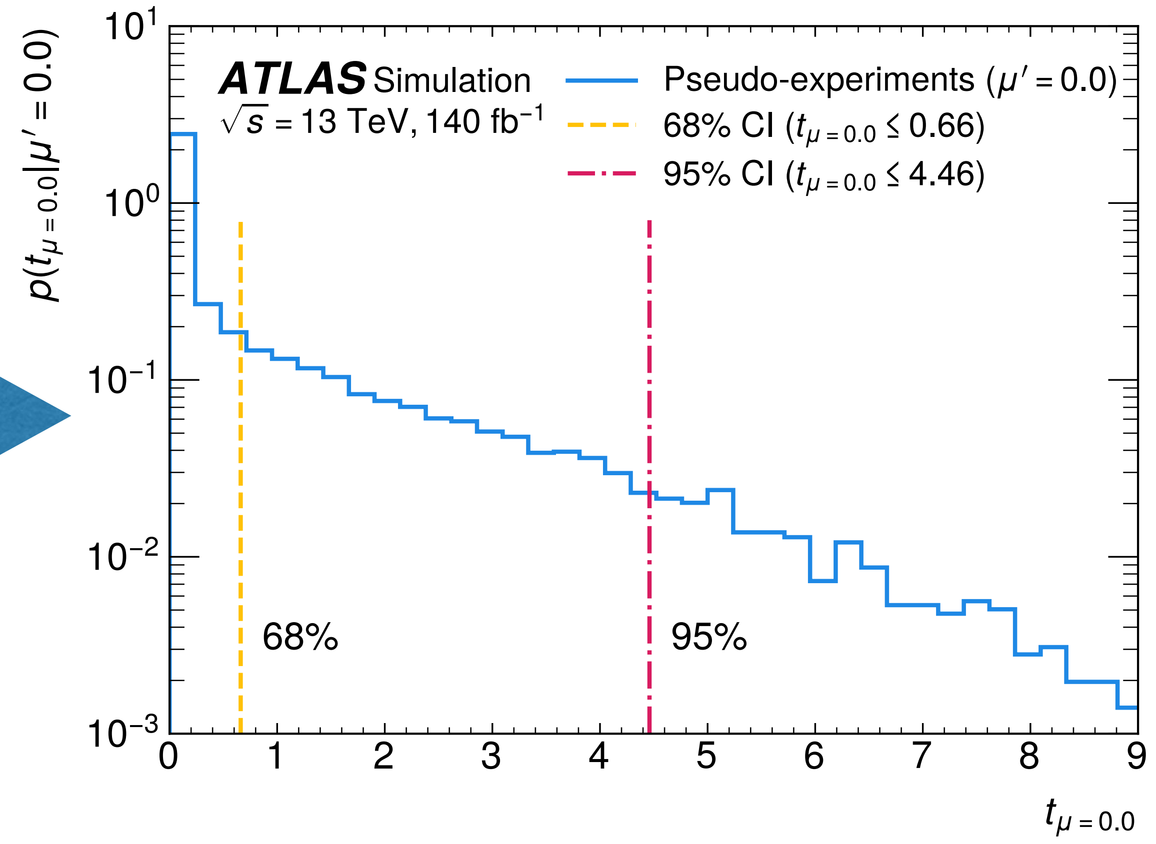
From simulations for a fixed μ, eg. $\mu = 0$



$$p_{bin}(n_{obs}|n_{exp}) = \frac{n_{exp}^{n_{obs}} \cdot e^{-n_{exp}}}{n_{obs}!}$$

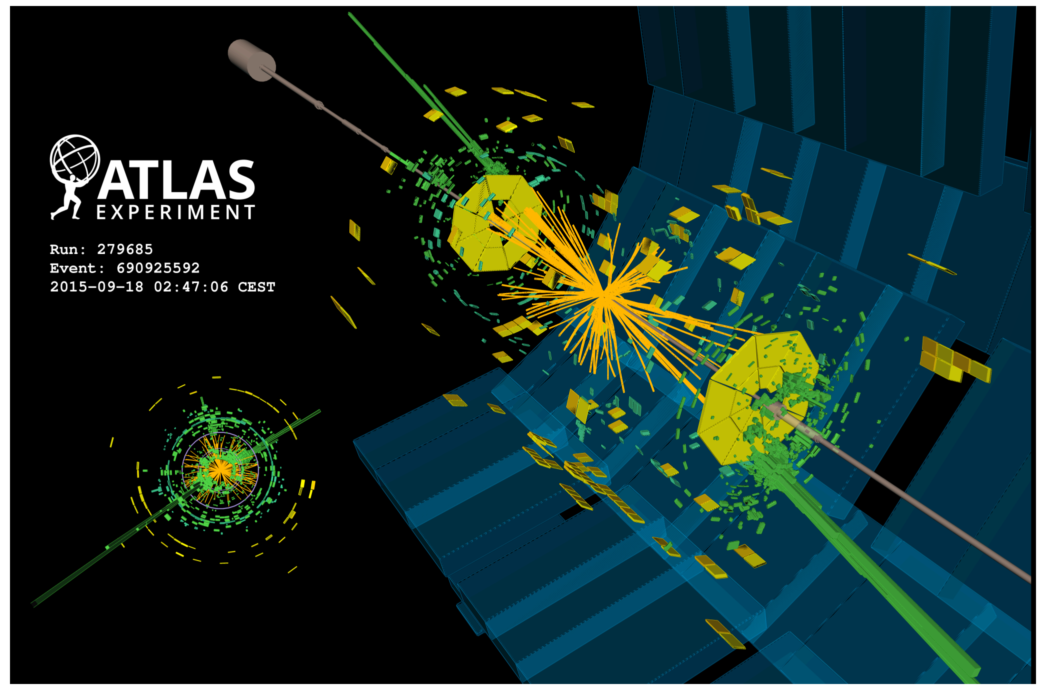
Compare **pseudo-data** to expectation, compute likelihood

Obtain **distribution of test statistic**



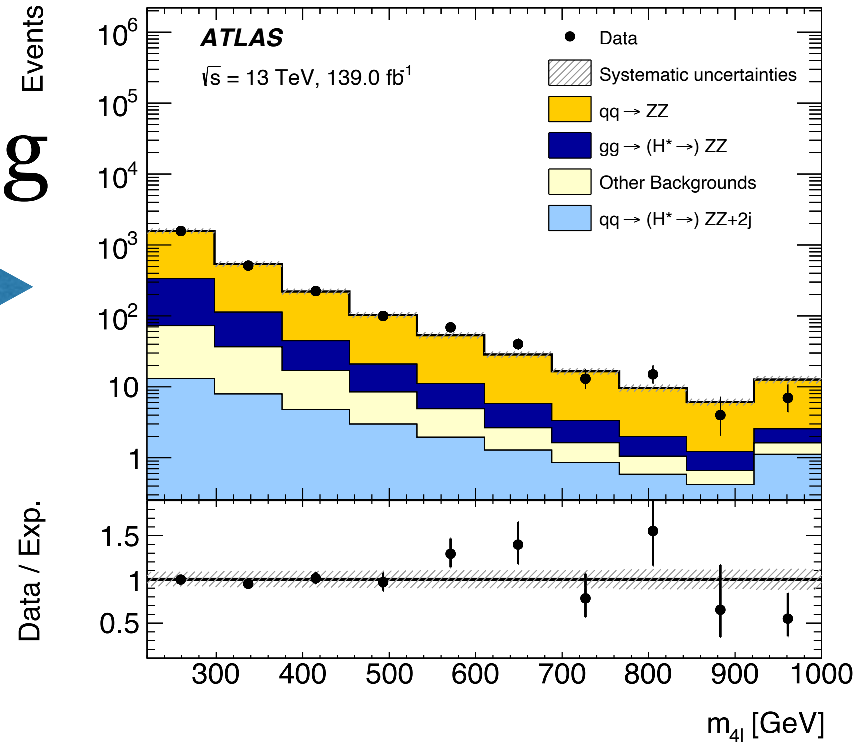
Obtaining the distribution of a test statistic

Raw data



Data processing

Low-dimensional summary



$$p_{bin}(n_{obs}|n_{exp}) = \frac{n_{exp}^{n_{obs}} \cdot e^{-n_{exp}}}{n_{obs}!}$$

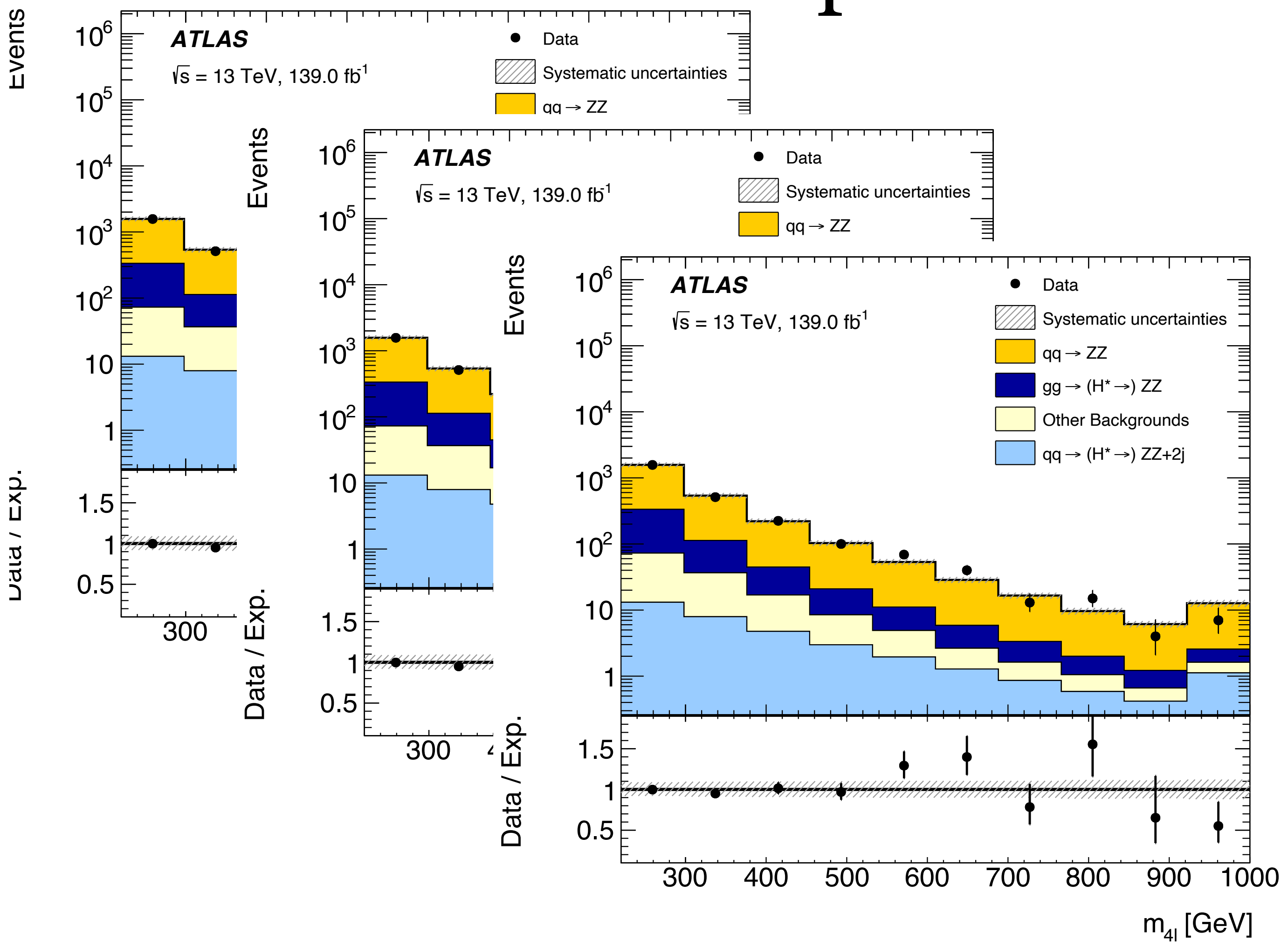
Compare data to expectation, compute likelihood

Obtain test statistic

$$t_\mu$$

From simulations for a fixed μ, eg. $\mu = 0$

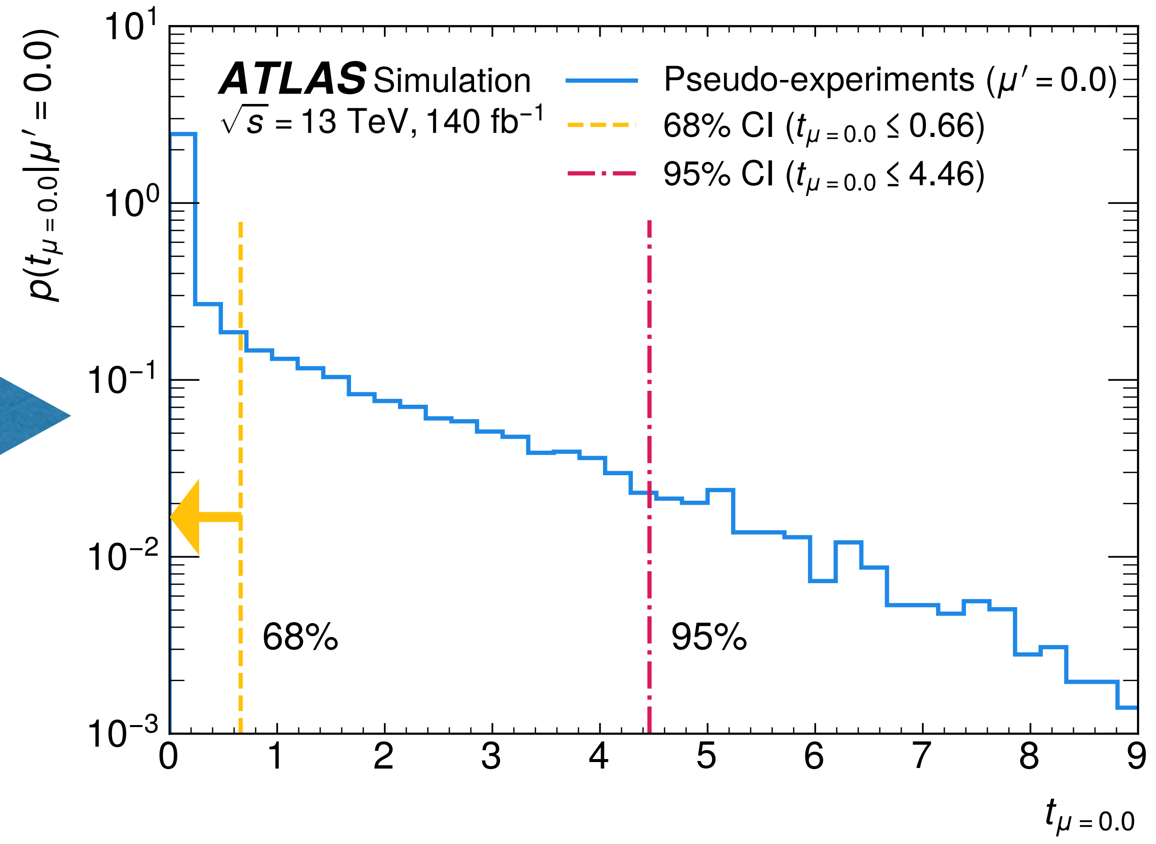
Create pseudo-experiments



$$p_{bin}(n_{obs}|n_{exp}) = \frac{n_{exp}^{n_{obs}} \cdot e^{-n_{exp}}}{n_{obs}!}$$

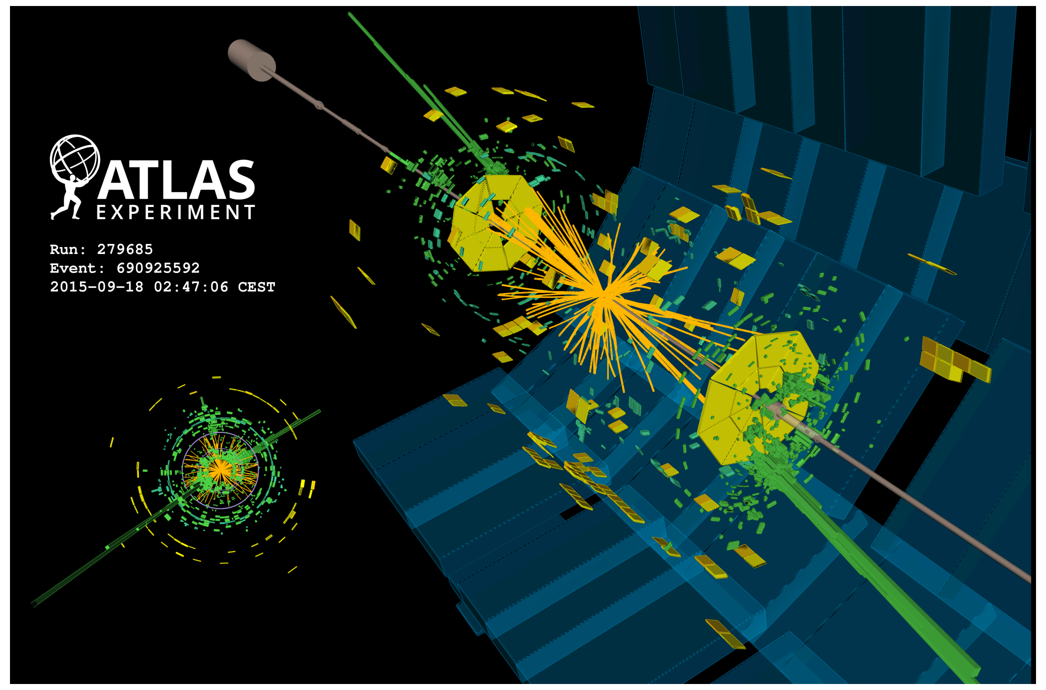
Compare **pseudo-data** to expectation, compute likelihood

Obtain **distribution of test statistic**



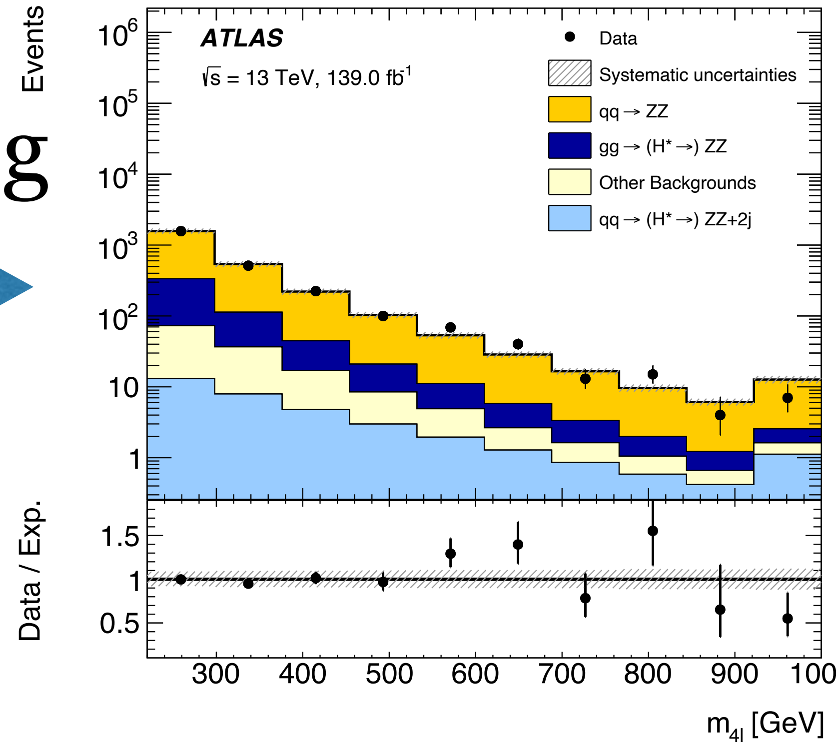
Obtaining the distribution of a test statistic

Raw data



Data processing

Low-dimensional summary



$$p_{bin}(n_{obs} | n_{exp}) = \frac{n_{exp}^{n_{obs}} \cdot e^{-n_{exp}}}{n_{obs}!}$$

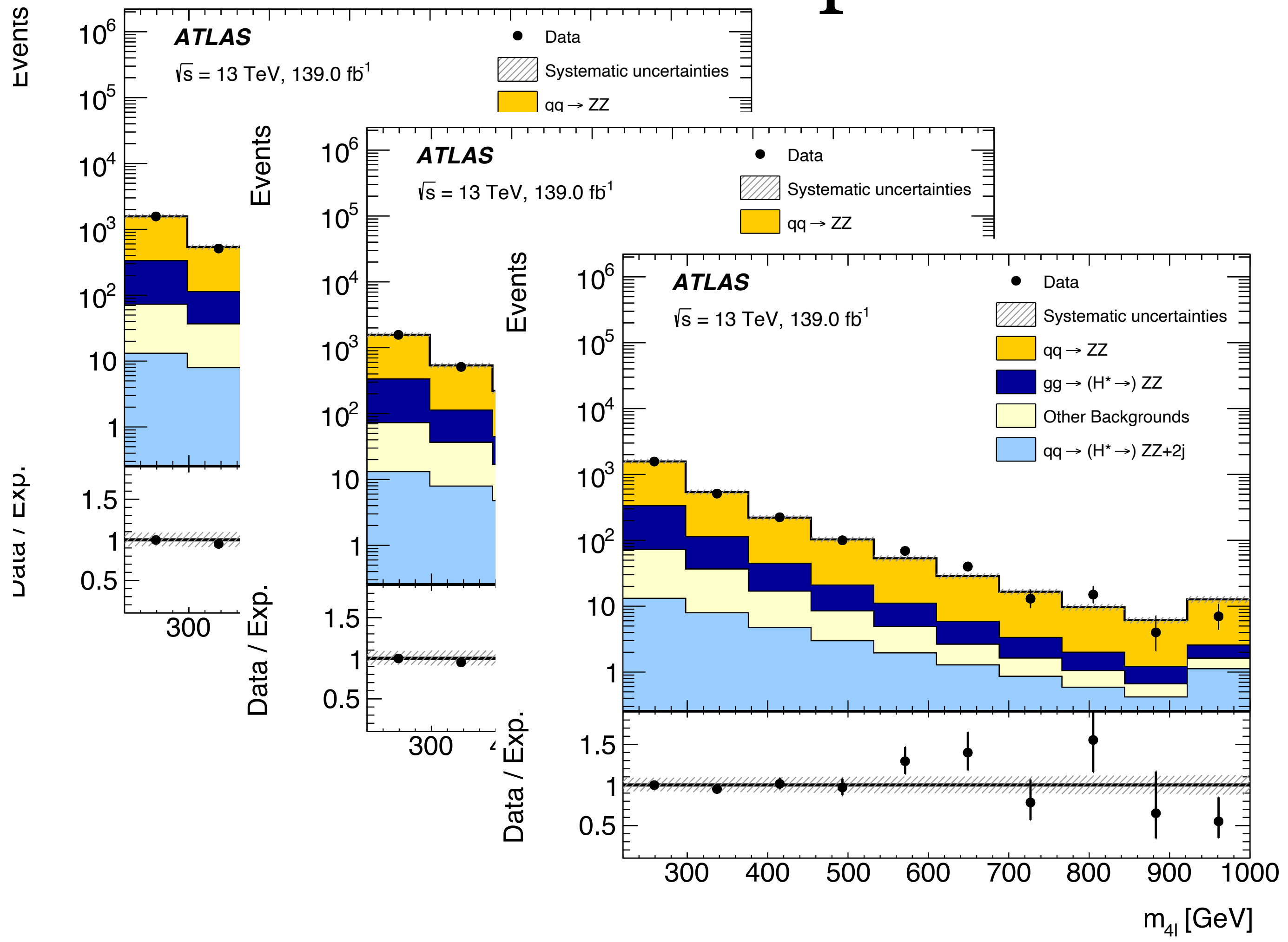
Compare data to expectation, compute likelihood

Obtain test statistic

$$t_\mu$$

From simulations for a fixed μ, eg. $\mu = 0$

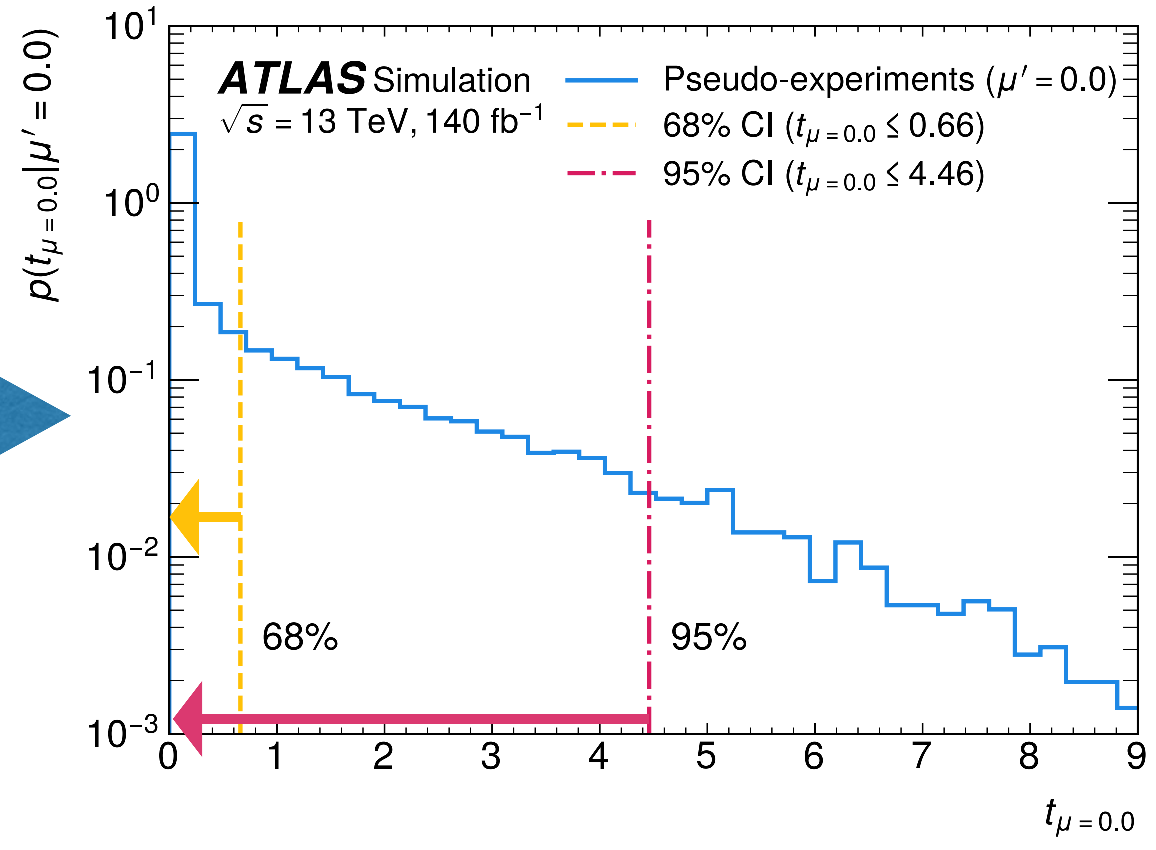
Create pseudo-experiments



$$p_{bin}(n_{obs} | n_{exp}) = \frac{n_{exp}^{n_{obs}} \cdot e^{-n_{exp}}}{n_{obs}!}$$

Compare **pseudo-data** to expectation, compute likelihood

Obtain **distribution of test statistic**

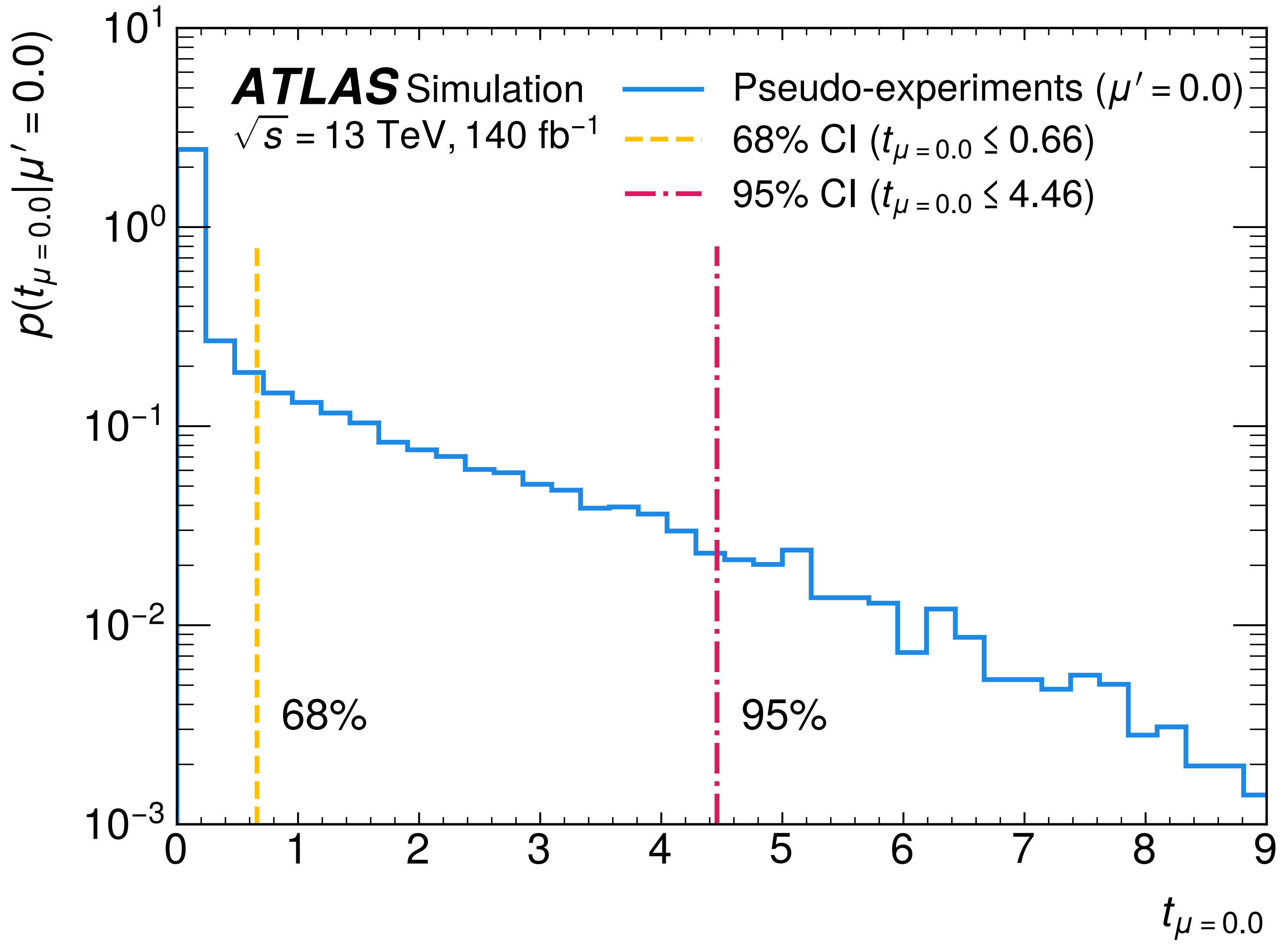


Calibrating confidence intervals

- Determine 68 % & 95 % CI empirically from this distribution
- Do it for each value of μ

Distribution of test statistic t_μ over thousands of simulated pseudo-experiments

True $\mu = 0$



t_μ

t_μ

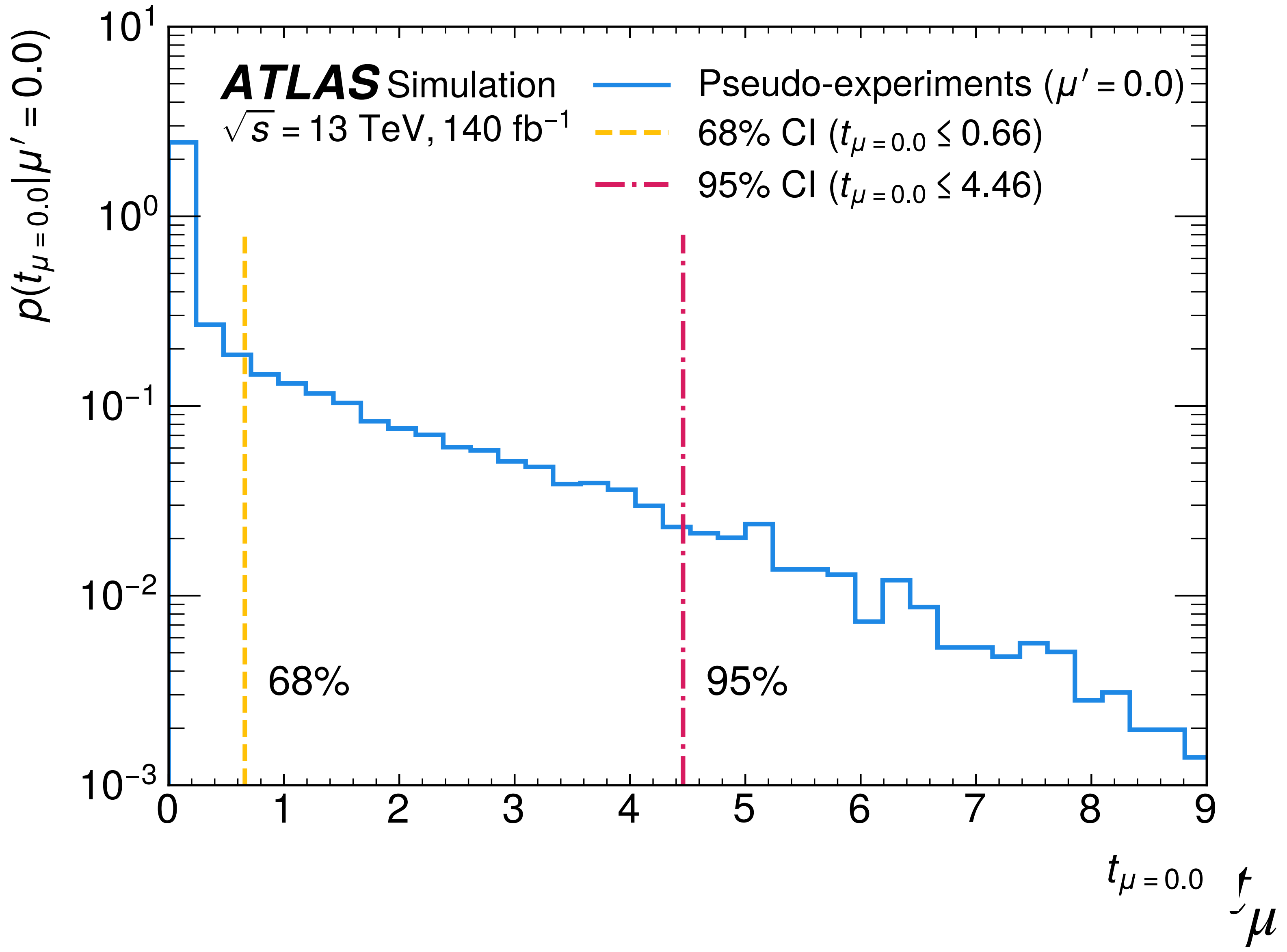
ATLAS NSBI paper: arXiv:2412.01600

Calibrating confidence intervals

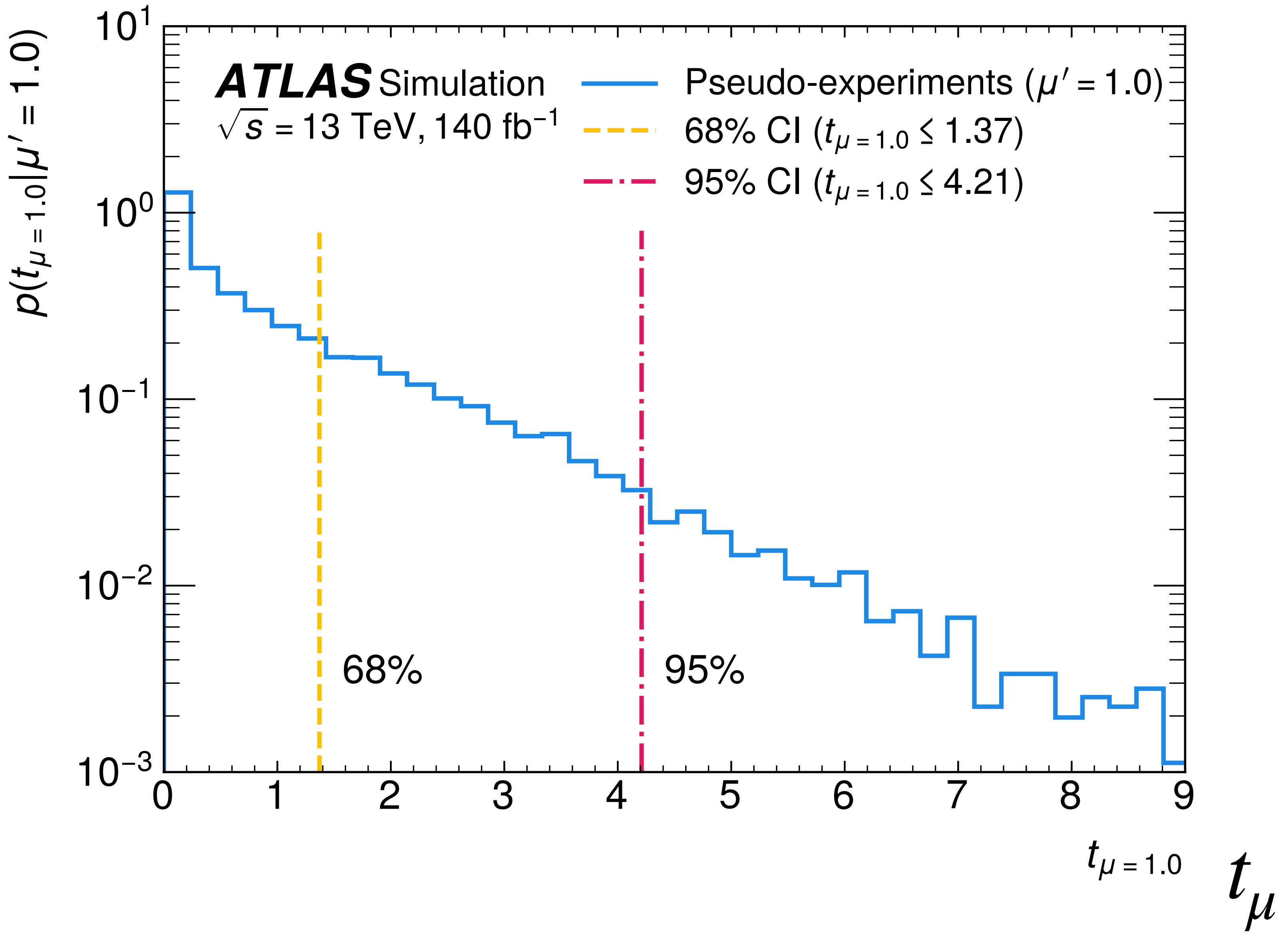
- Determine 68 % & 95 % CI empirically from this distribution
- Do it for each value of μ

Distribution of test statistic t_μ over thousands of simulated pseudo-experiments

True $\mu = 0$



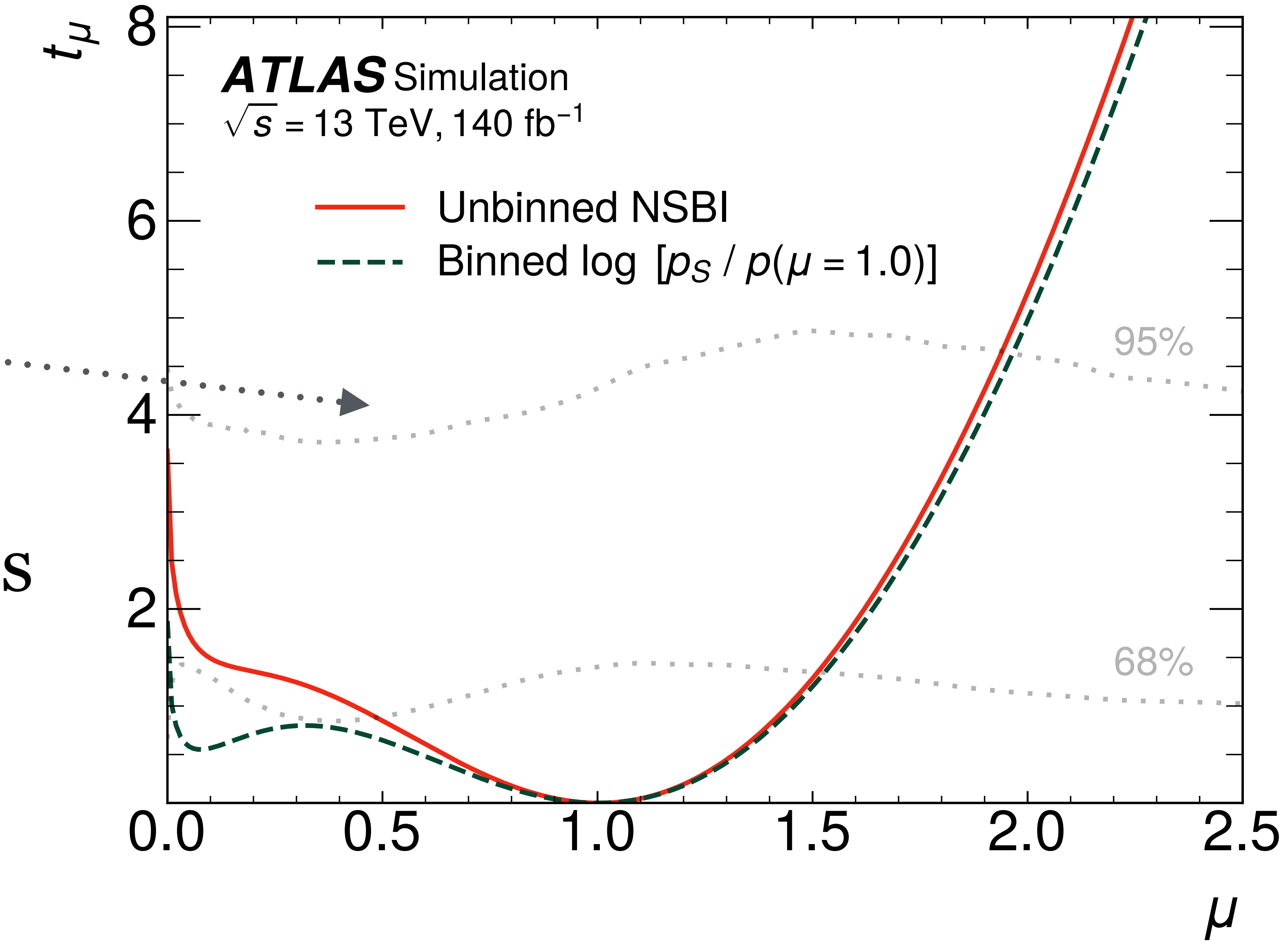
True $\mu = 1$



ATLAS NSBI paper: arXiv:2412.01600

Calibrating confidence intervals

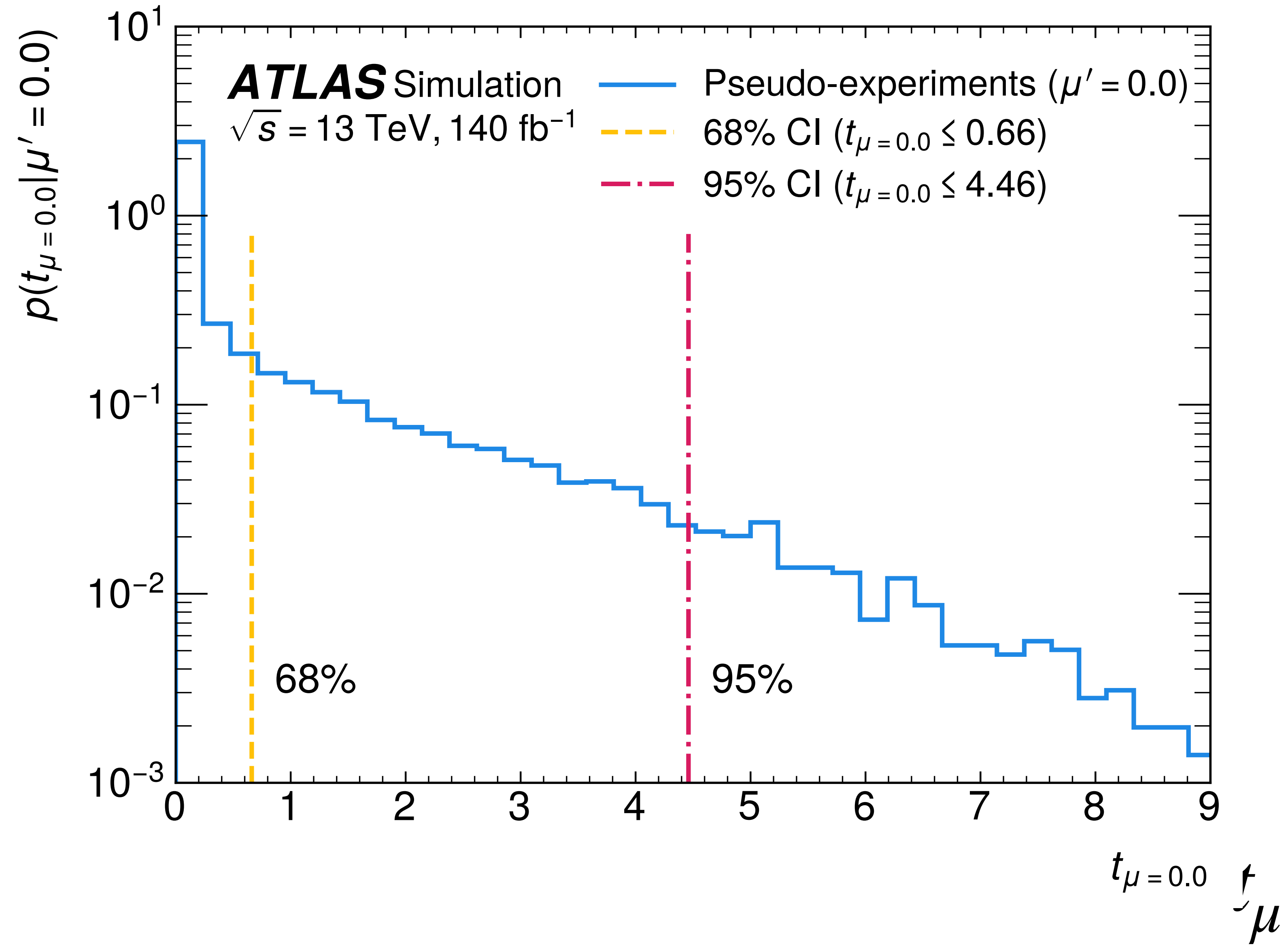
- Determine 68 % & 95 % CI empirically from this distribution
- Do it for each value of μ

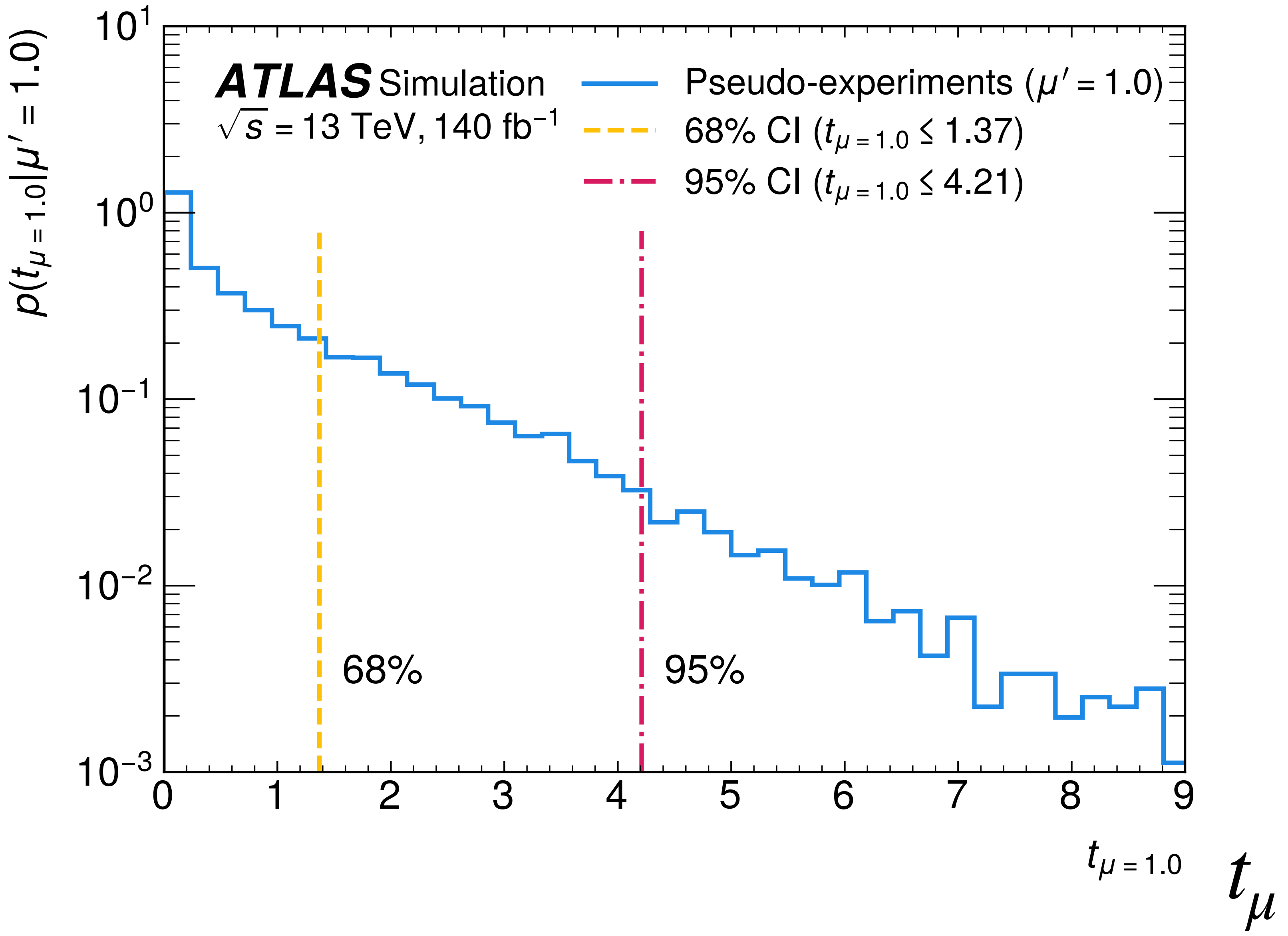


Distribution of test statistic t_μ over thousands of simulated pseudo-experiments

True $\mu = 0$

True $\mu = 1$

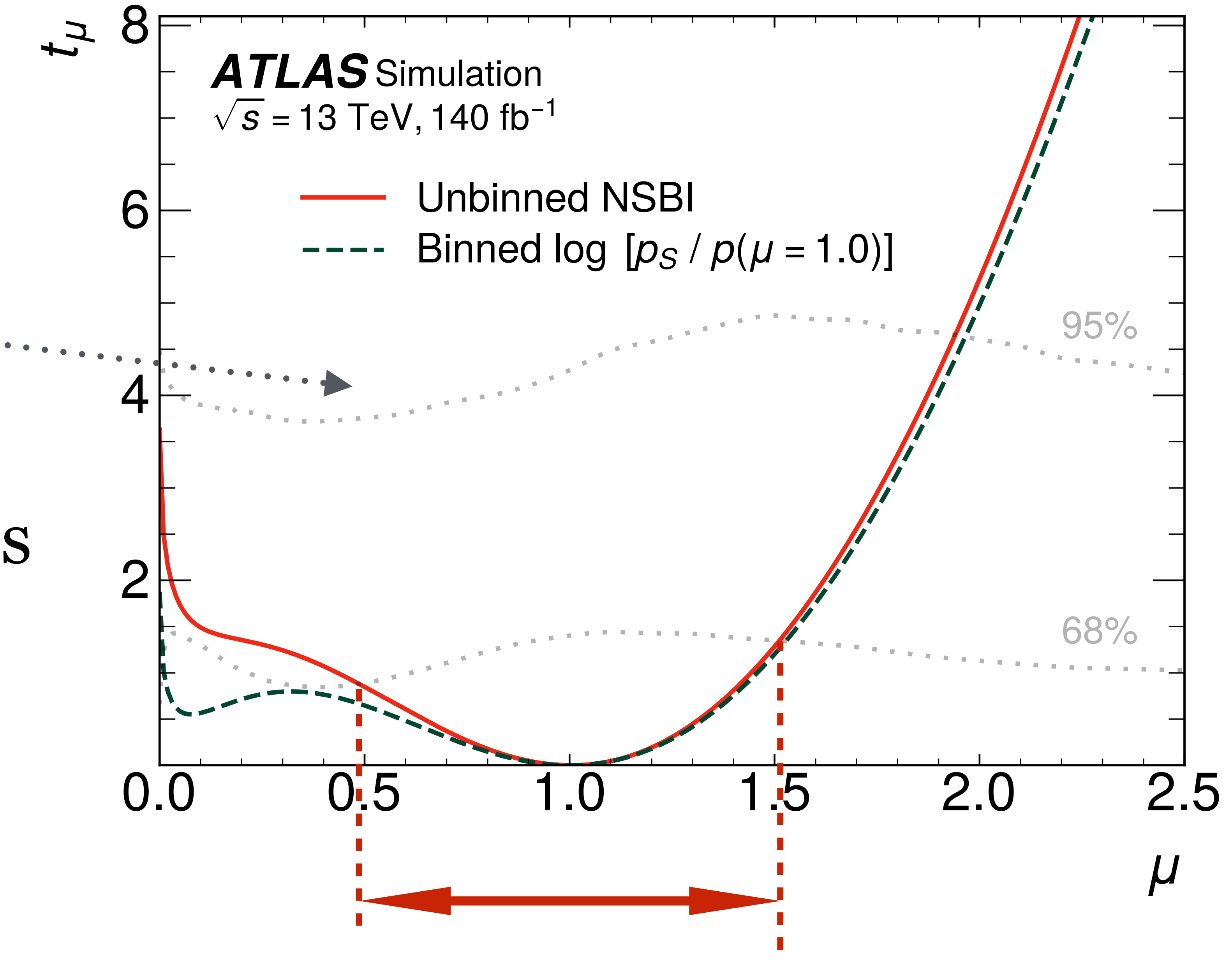




ATLAS NSBI paper: arXiv:2412.01600

Calibrating confidence intervals

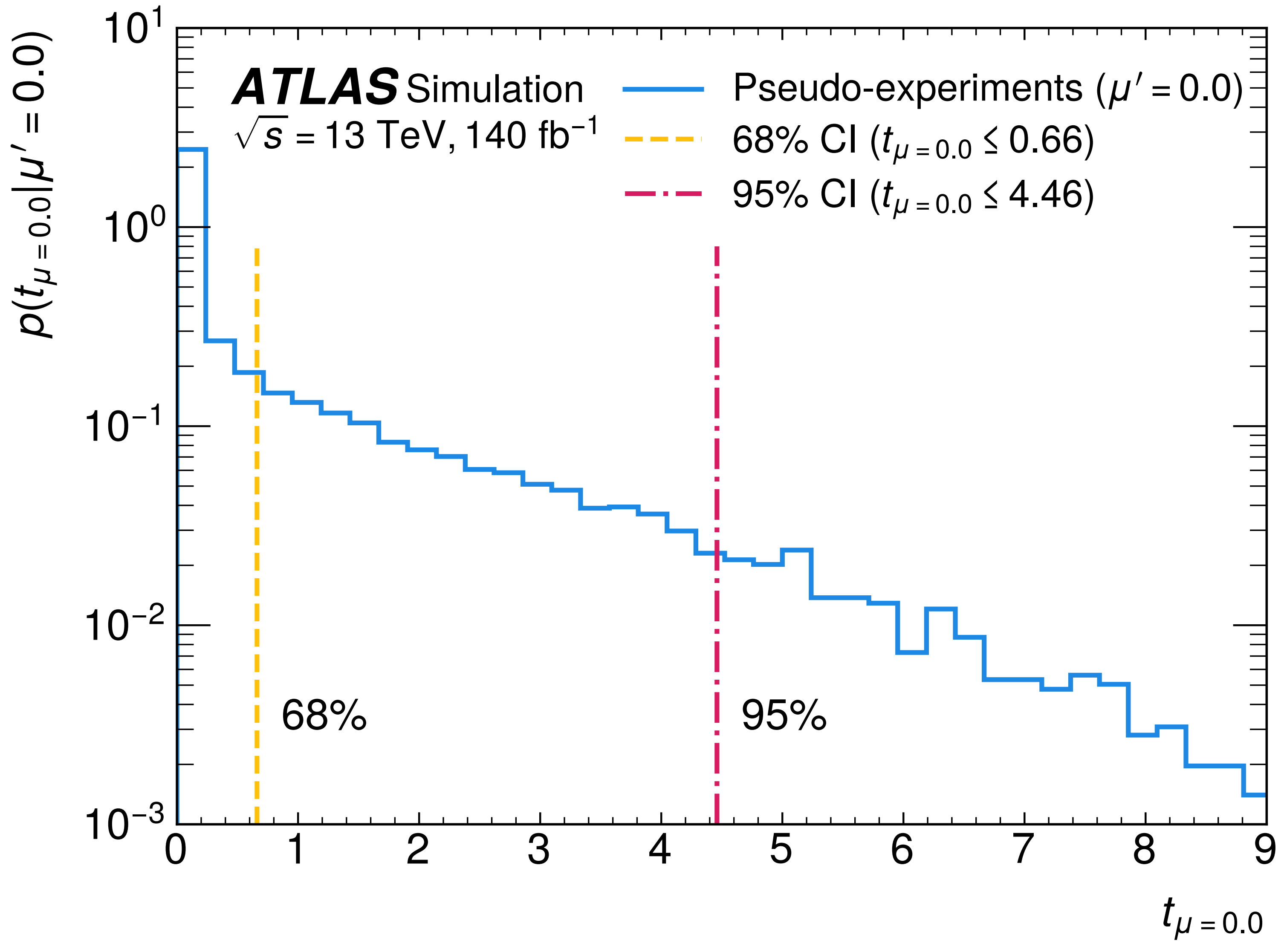
- Determine 68 % & 95 % CI empirically from this distribution
- Do it for each value of μ



Calibrated confidence interval

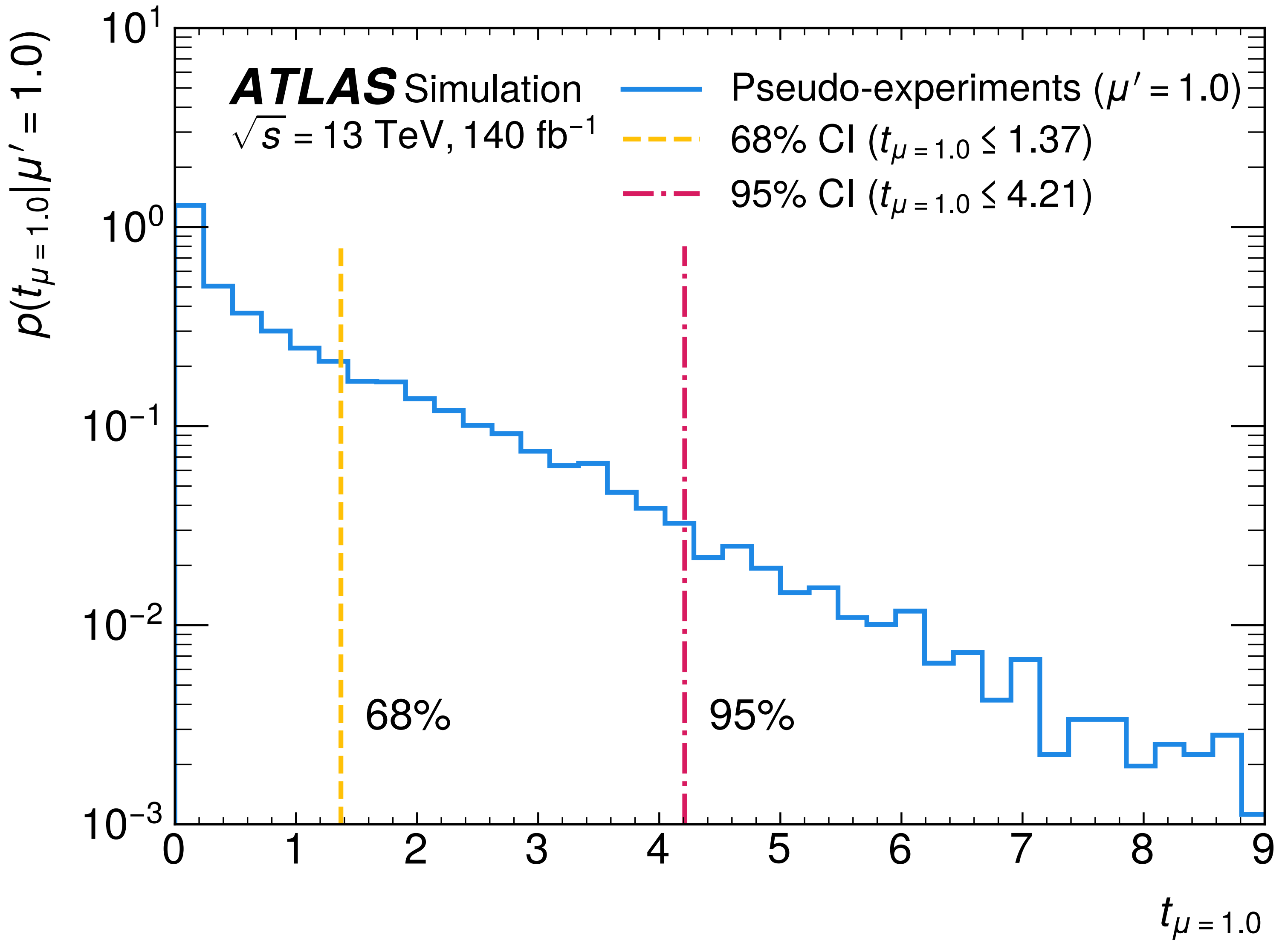
Distribution of test statistic t_μ over thousands of simulated pseudo-experiments

True $\mu = 0$



t_μ

True $\mu = 1$

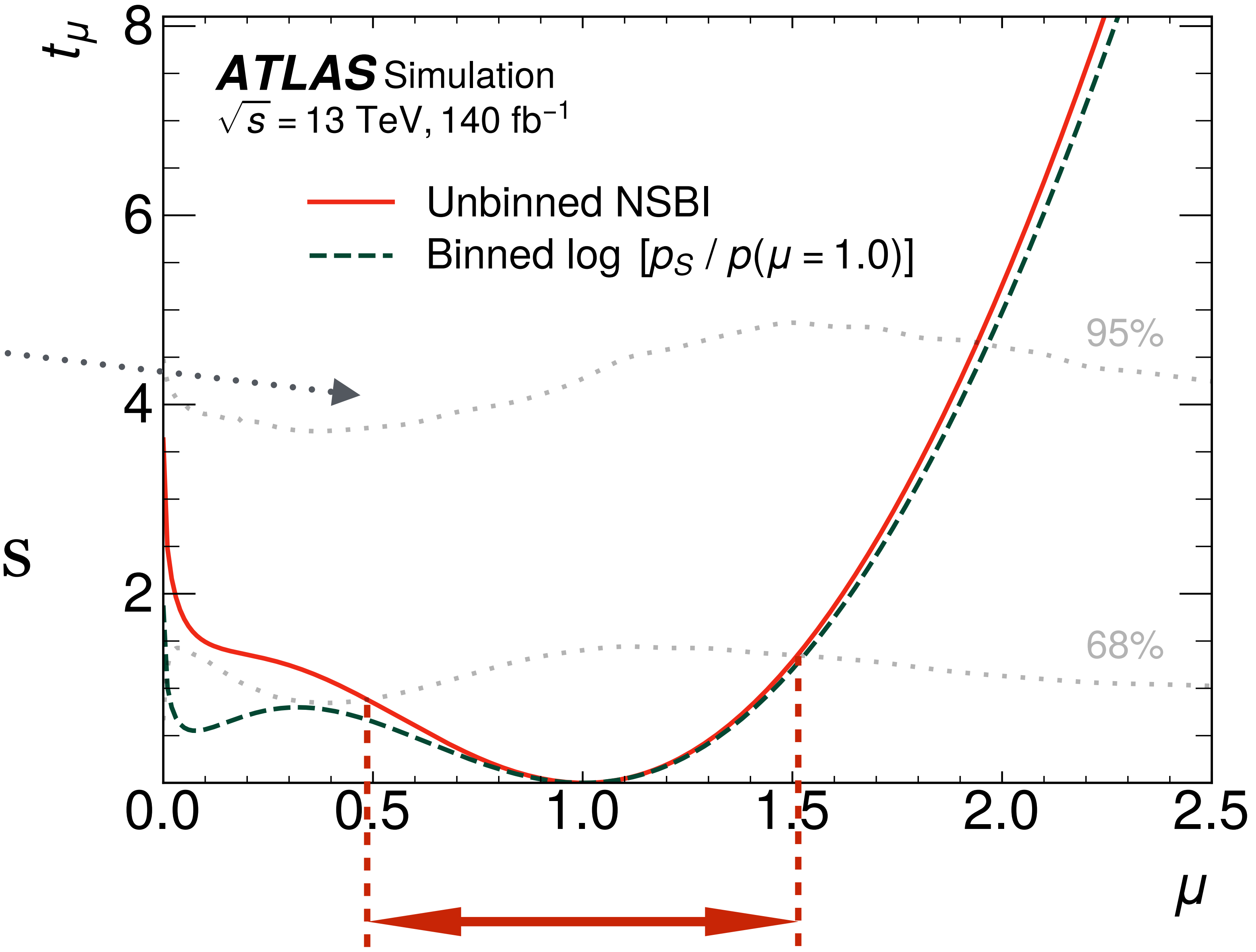


t_μ

ATLAS NSBI paper: arXiv:2412.01600

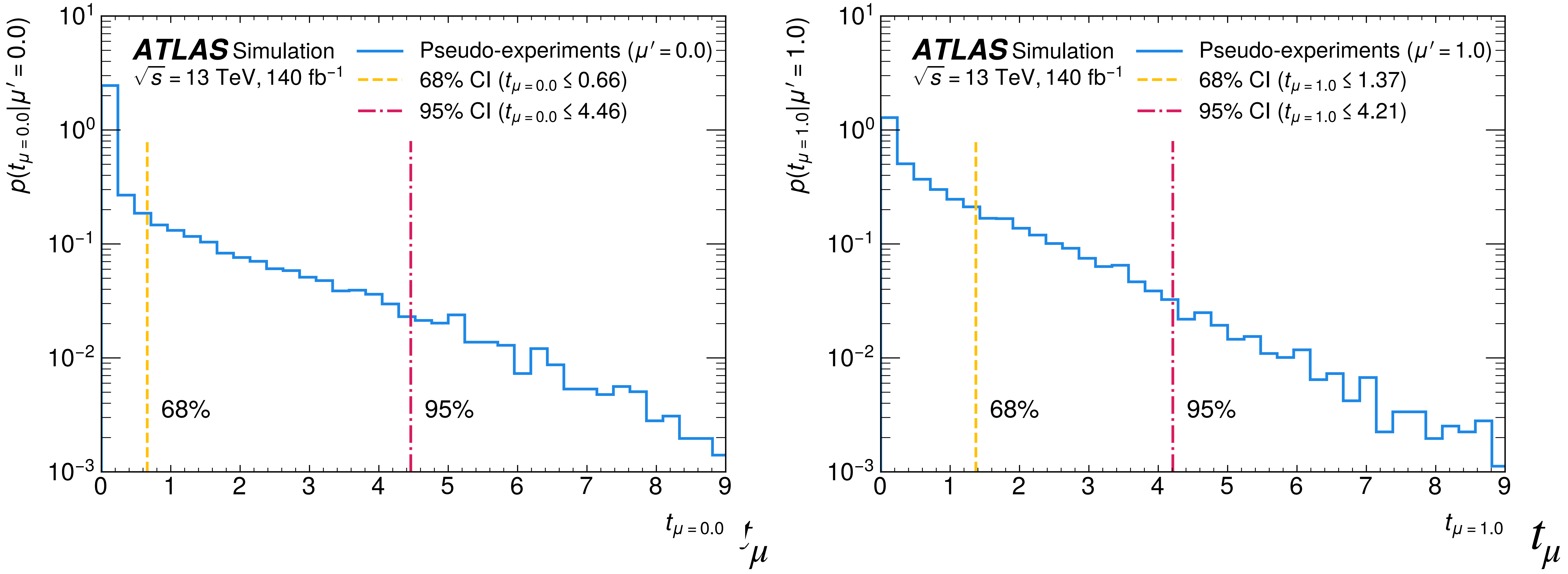
Calibrating confidence intervals

- Determine 68 % & 95 % CI empirically from this distribution
- Do it for each value of μ



Calibrated confidence interval

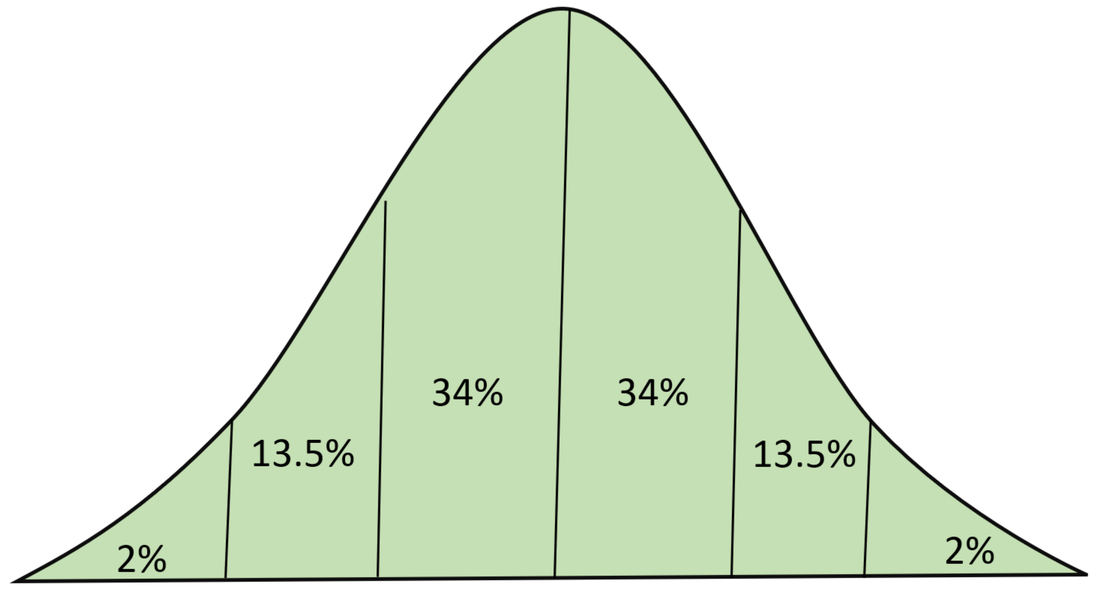
Distribution of test statistic t_μ over thousands of simulated pseudo-experiments



True $\mu = 0$

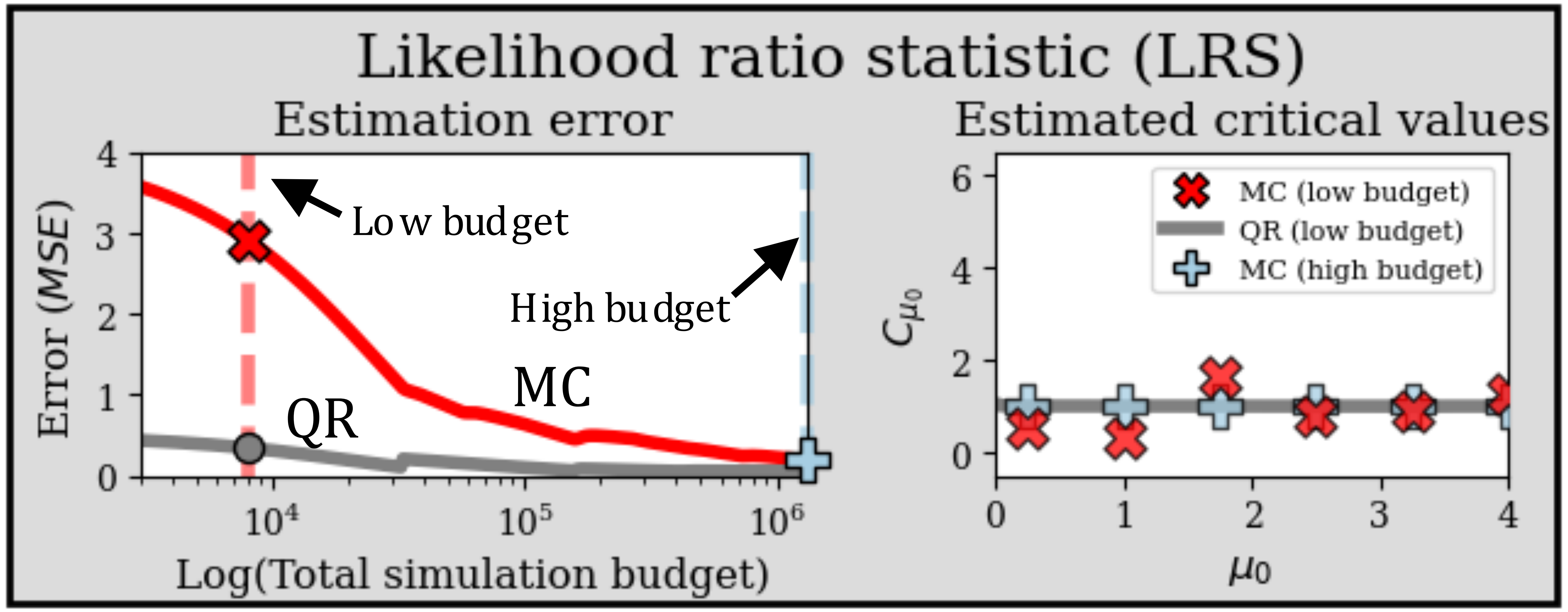
True $\mu = 1$

Task is to determine quantiles



Perform efficiently with quantile regression!

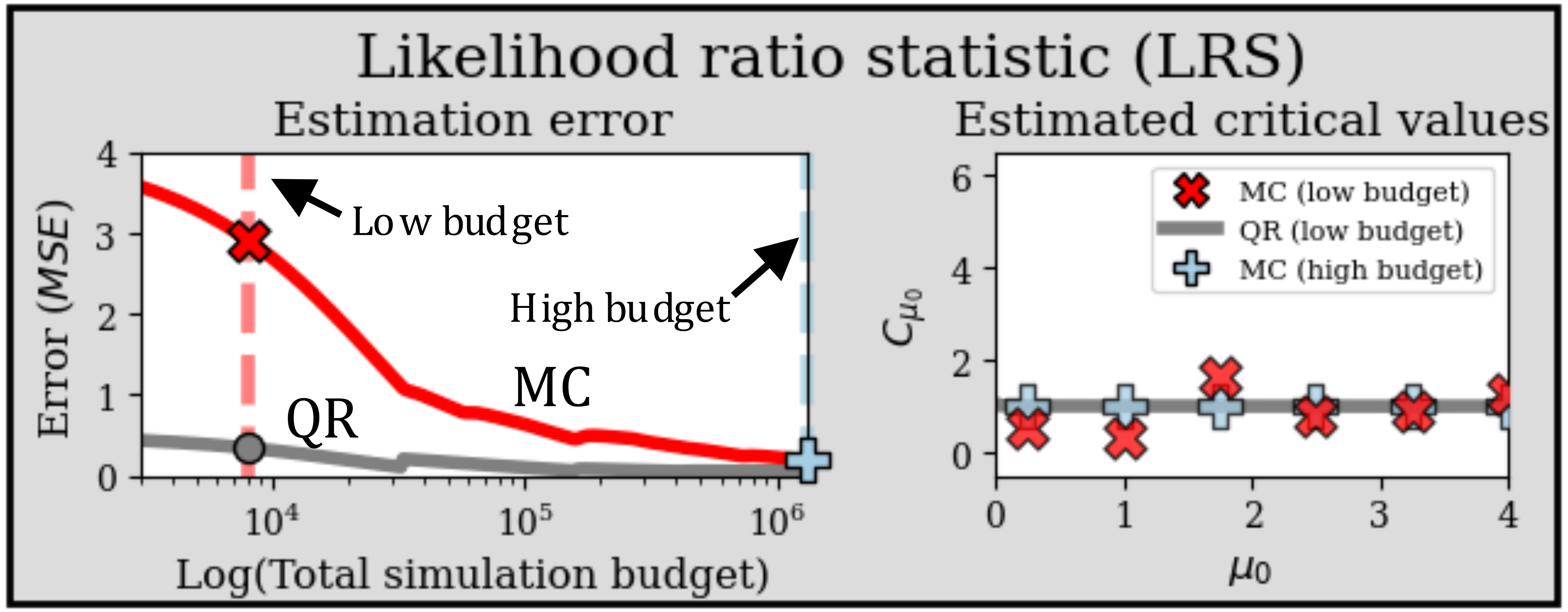
Can do it faster and more accurately with ML technique known as ‘quantile regression’ (QR)



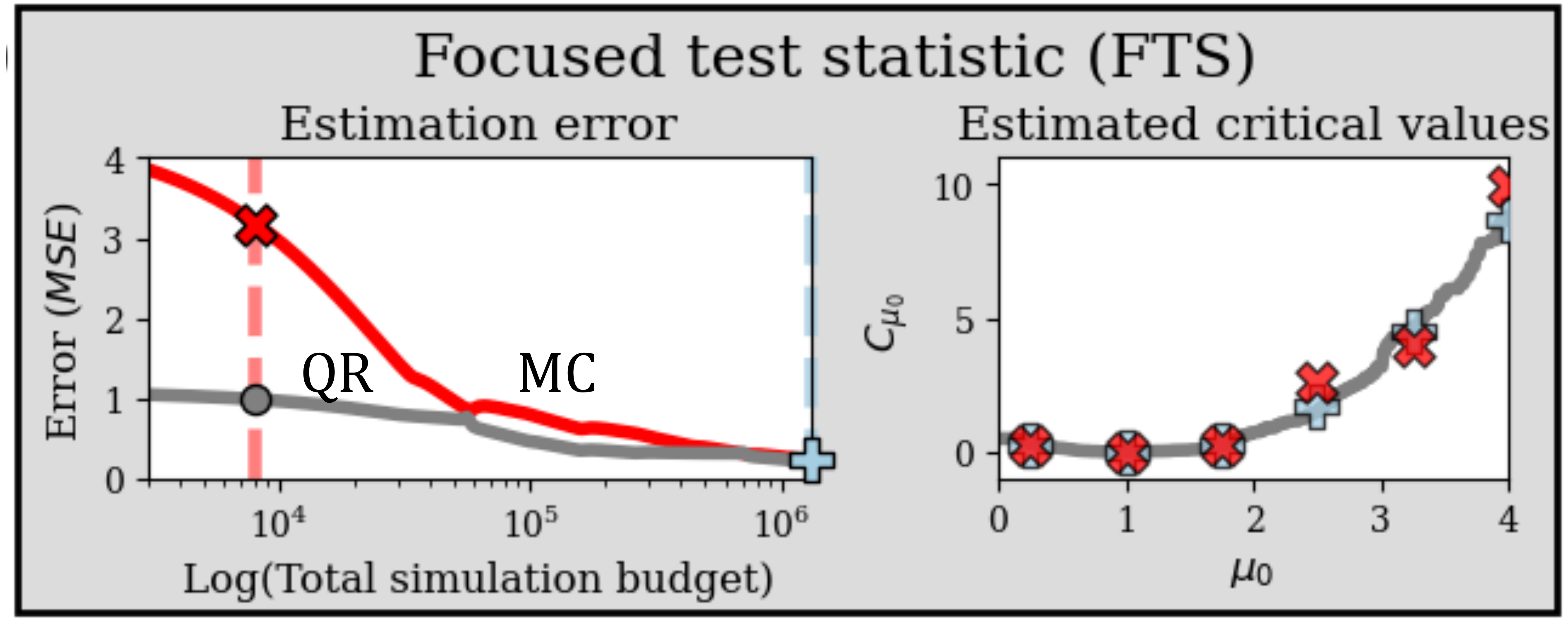
arXiv:2507.17831: Carzon, Ghosh et al

Perform efficiently with quantile regression!

Can do it faster and more accurately with ML technique known as ‘quantile regression’ (QR)



Likelihood ratio statistic (LRS)

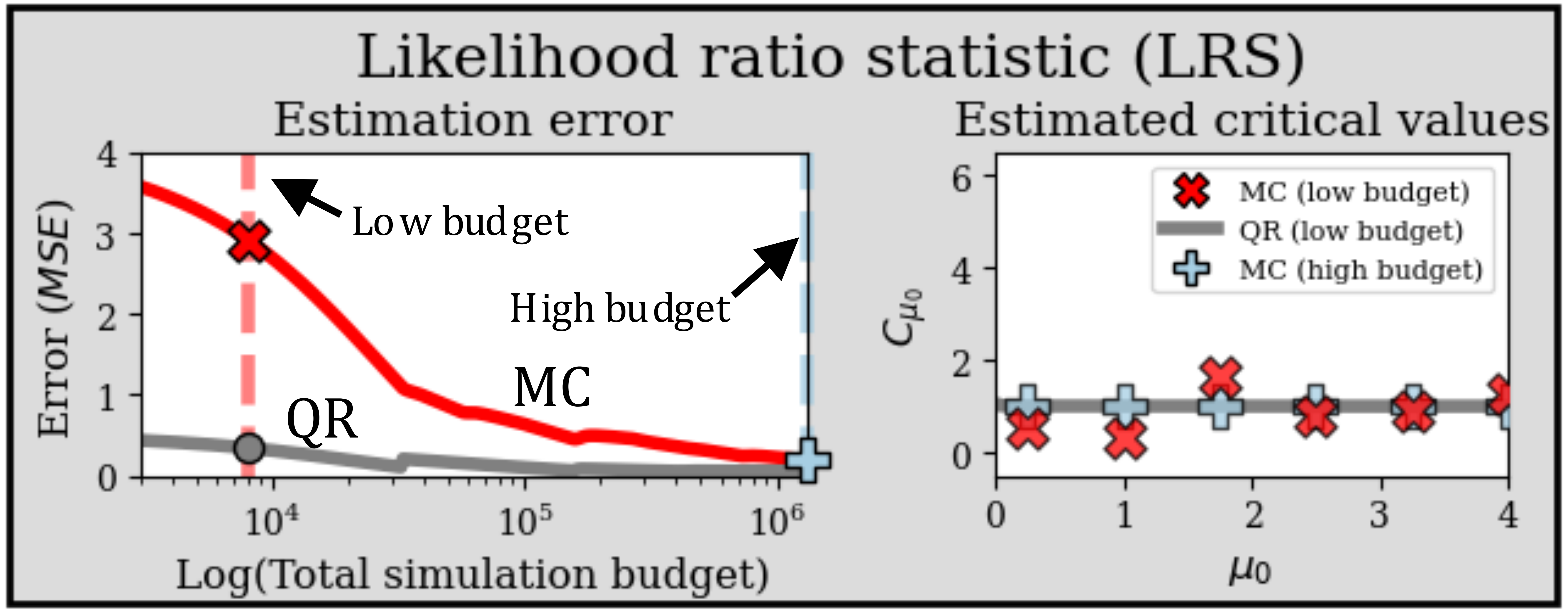


Focused test statistic (FTS)

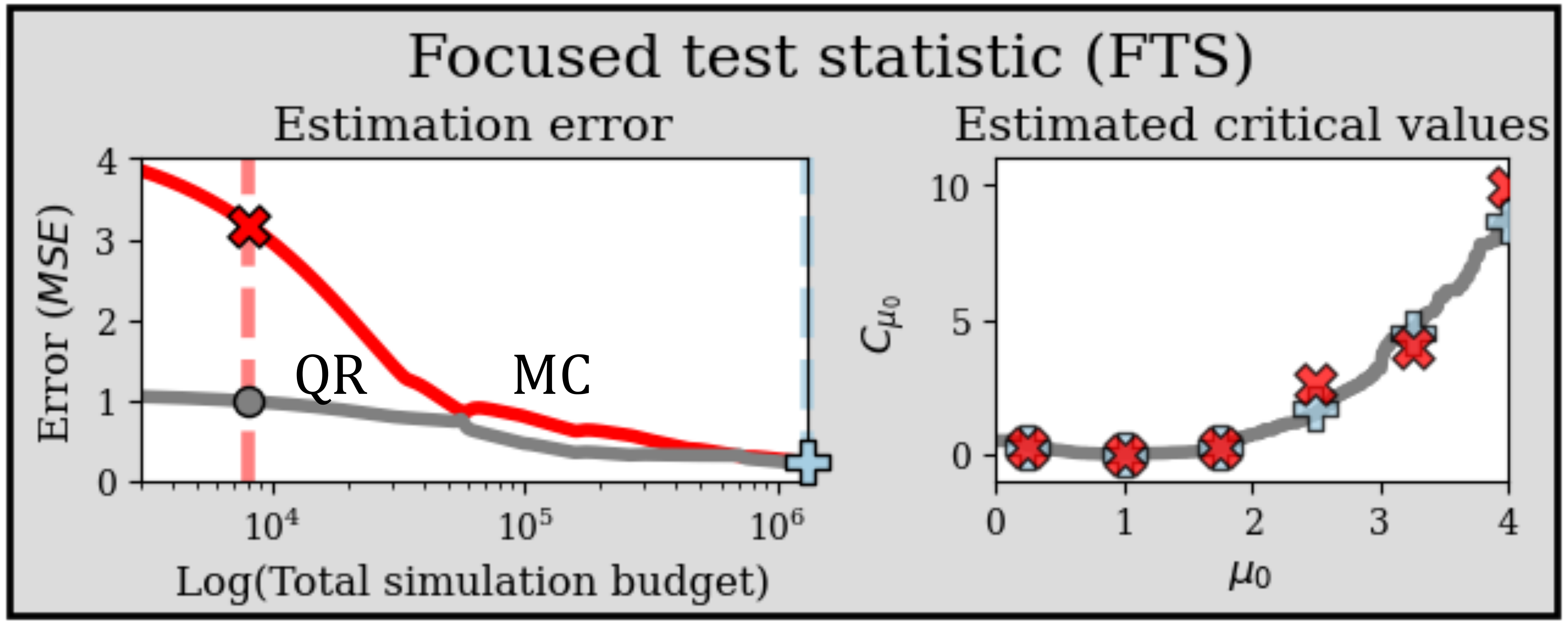
arXiv:2507.17831: Carzon, Ghosh et al

Perform efficiently with quantile regression!

Can do it faster and more accurately with ML technique known as ‘quantile regression’ (QR)



Applications in neutrino or dark matter experiments that routinely employ expensive MC method for confidence belts

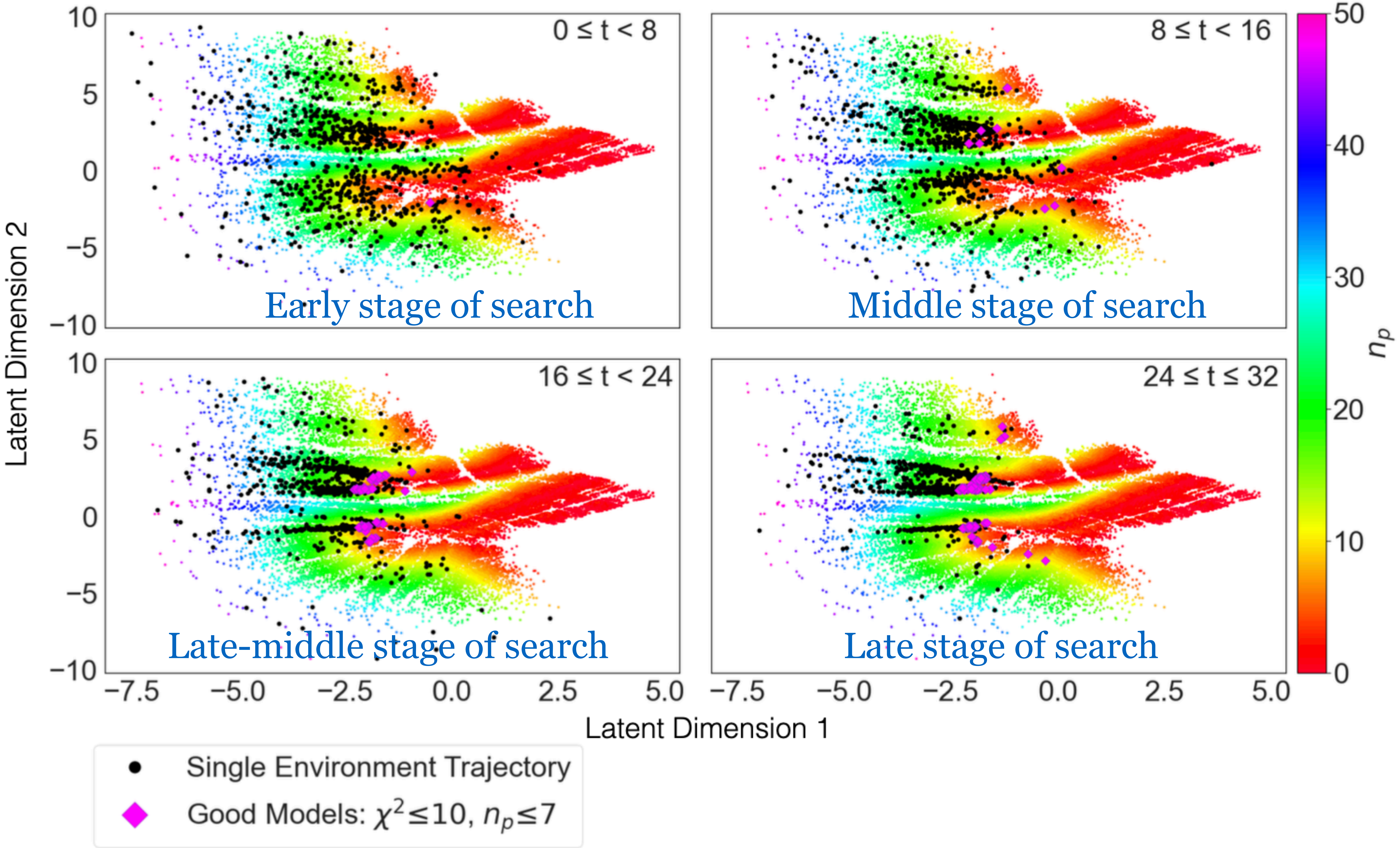


Neutrino model building backup slides

Baretz et al (inc. Ghosh): arXiv:2506.08080

Deep dive into what the agent is doing over time

Visualised in 2D space created by an auto-encoder



Black dots are search trajectory for one agent in one environment

Pink are ‘good models’

Explores widely in the beginning, but realises that zero-parameter models are useless.

Focuses on few parameter models towards the end

Neutrino model building details

Data from experiments:

observables	best-fit values
m_e/m_μ	0.0048 ± 0.0002
m_μ/m_τ	0.0565 ± 0.0045
$\Delta m^2_{21}/\Delta m^2_{31}$	$0.0295^{+0.0012}_{-0.0010}$
$\sin^2\theta_{12}$	$0.307^{+0.012}_{-0.011}$
$\sin^2\theta_{13}$	$0.02224^{+0.00056}_{-0.00057}$
$\sin^2\theta_{23}$	$0.454^{+0.019}_{-0.016}$
δ^ℓ_{CP}/π	$1.289^{+0.217}_{-0.139}$

Search efficiency over random scan

Space	Number	Random scan	AMBer
$A_4 \times \mathbb{Z}_4$	4×10^6	0	1237
$A_4 \times \mathbb{Z}_N$	4×10^6	30	4681
$T_{19} \times \mathbb{Z}_4$	8×10^6	131	423

$$
\begin{aligned}
\mathscr{W} &= \alpha^{(1)}_{ij}\left(L_iE_jH_d\right) + \frac{\alpha^{(2)}_{ijk}}{\Lambda}\left(L_iE_j\phi_k\overset{\substack{\text{Higgs doublet}\\ \downarrow}}{H_d}\right) && \text{Charged lepton masses}\\
\text{Free coupling parameters } \alpha_i \quad & + \alpha^{(3)}_{ij}\left(L_iN_jH_u\right) + \frac{\alpha^{(4)}_{ijk}}{\Lambda}\left(L_iN_j\phi_kH_u\right) && \text{Direct masses}\\
& + \Lambda\alpha^{(5)}_{ij}\left(N_iN_j\right) + \alpha^{(6)}_{ijk}\left(N_iN_j\phi_k\right) + \frac{\alpha_{ijkl^{(7)}}}{\Lambda}\left(N_iN_j\phi_k\phi_l\right) && \text{Majorana masses}
\end{aligned}
$$

Action Space

Leptons and Higgses	Flavons	Global
Change non-Abelian representation Change Abelian charge	Change non-Abelian representation Remove or add a particle Change Abelian charge Change VEV configuration	Change Abelian symmetry order

T_{19} Group

$$a^3 = \mathbb{1}\,, \qquad b^{19} = \mathbb{1}\,, \qquad \text{and} \qquad ba = ab^7$$

$$\mathbf{3_1}: \quad b = \begin{pmatrix} \mathrm{e}^{\frac{2\pi\mathrm{i}}{19}} & 0 & 0 \\ 0 & \mathrm{e}^{-\frac{16\pi\mathrm{i}}{19}} & 0 \\ 0 & 0 & \mathrm{e}^{\frac{14\pi\mathrm{i}}{19}} \end{pmatrix}\,,$$

$$\mathbf{3_2}: \quad b = \begin{pmatrix} \mathrm{e}^{\frac{4\pi\mathrm{i}}{19}} & 0 & 0 \\ 0 & \mathrm{e}^{\frac{6\pi\mathrm{i}}{19}} & 0 \\ 0 & 0 & \mathrm{e}^{-\frac{10\pi\mathrm{i}}{19}} \end{pmatrix}\,,$$

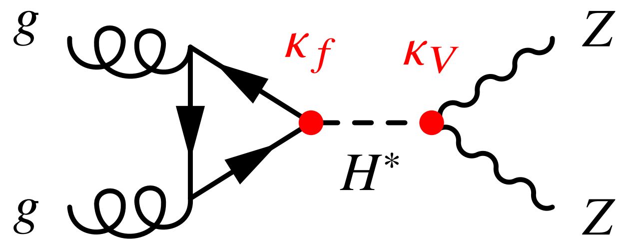
$$\mathbf{3_3}: \quad b = \begin{pmatrix} \mathrm{e}^{-\frac{12\pi}{19}} & 0 & 0 \\ 0 & \mathrm{e}^{-\frac{18\pi}{19}} & 0 \\ 0 & 0 & \mathrm{e}^{-\frac{8\pi}{19}} \end{pmatrix}\,,$$

$$a = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

NSBI backup slides

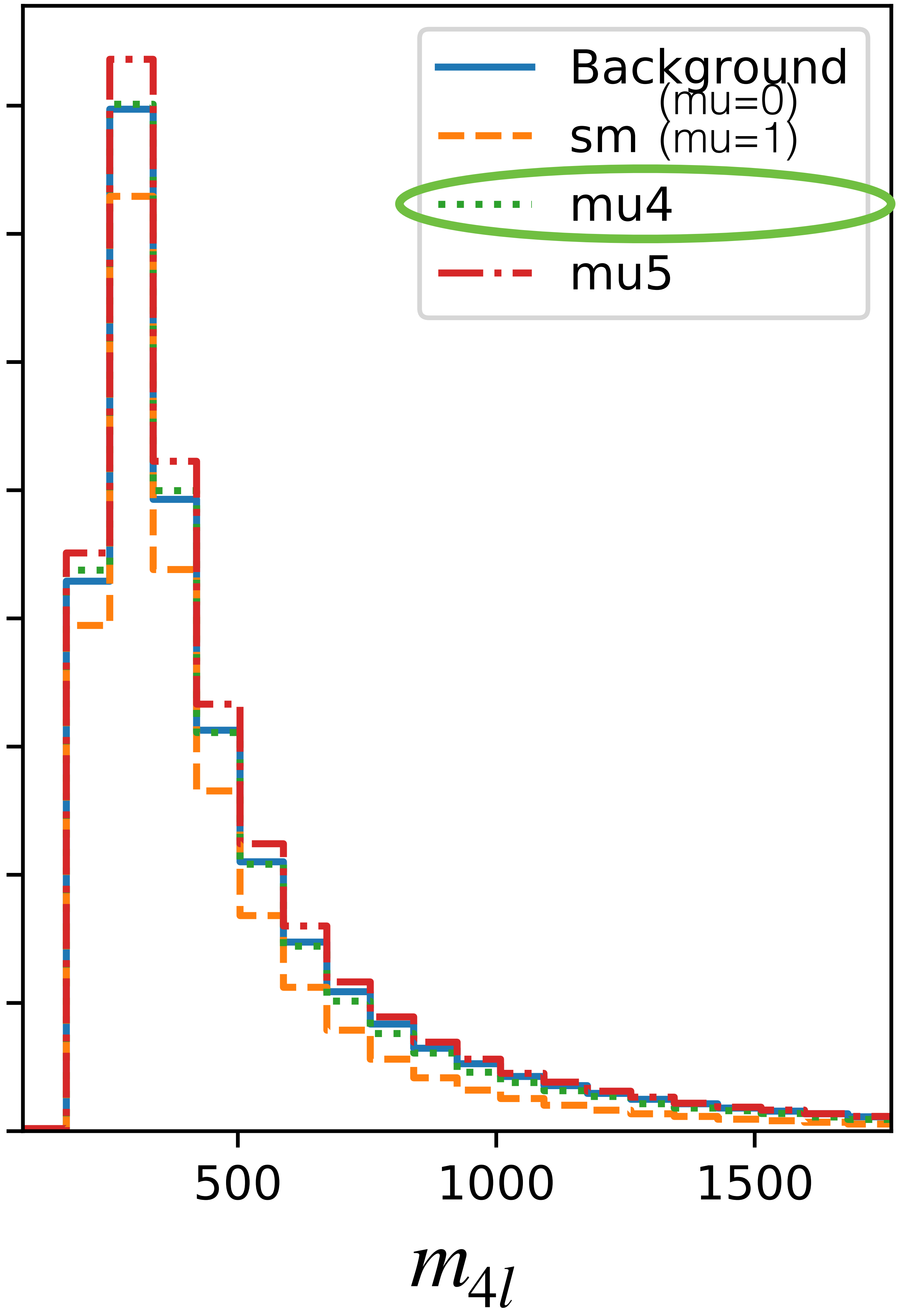
No single observable captures all information in Higgs width study

Signal-background-inference simulations (VBF processes): MG + Pythia+ Delphes

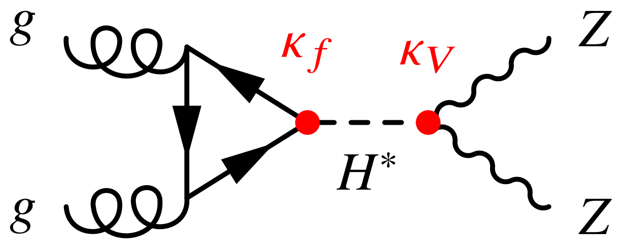


No single observable captures all information in Higgs width study

Signal-background-inference simulations (VBF processes): MG + Pythia+ Delphes

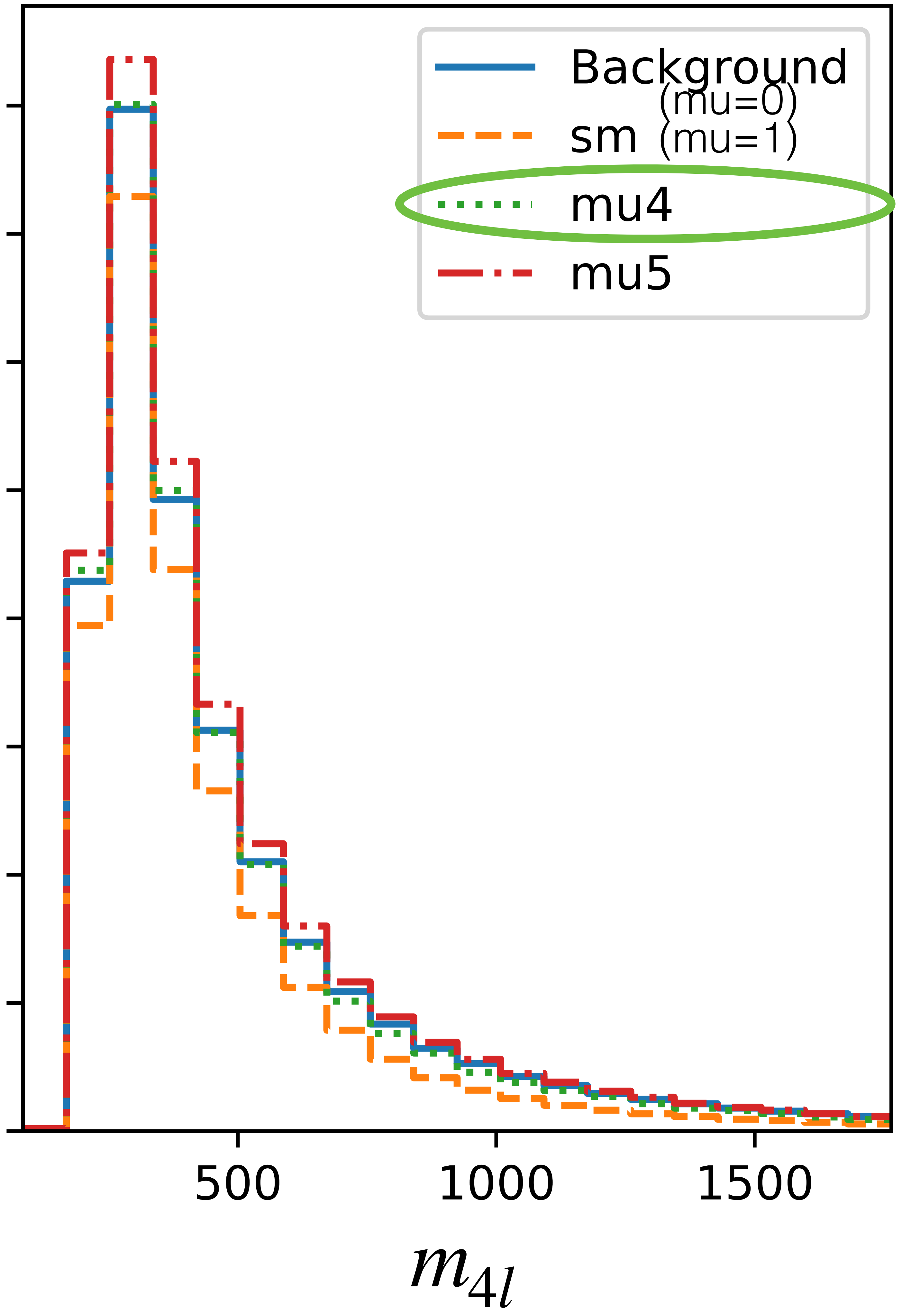


Can you spot the green plot?



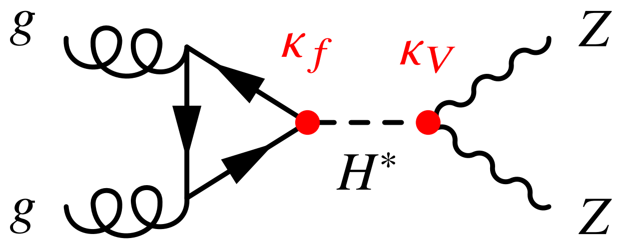
No single observable captures all information in Higgs width study

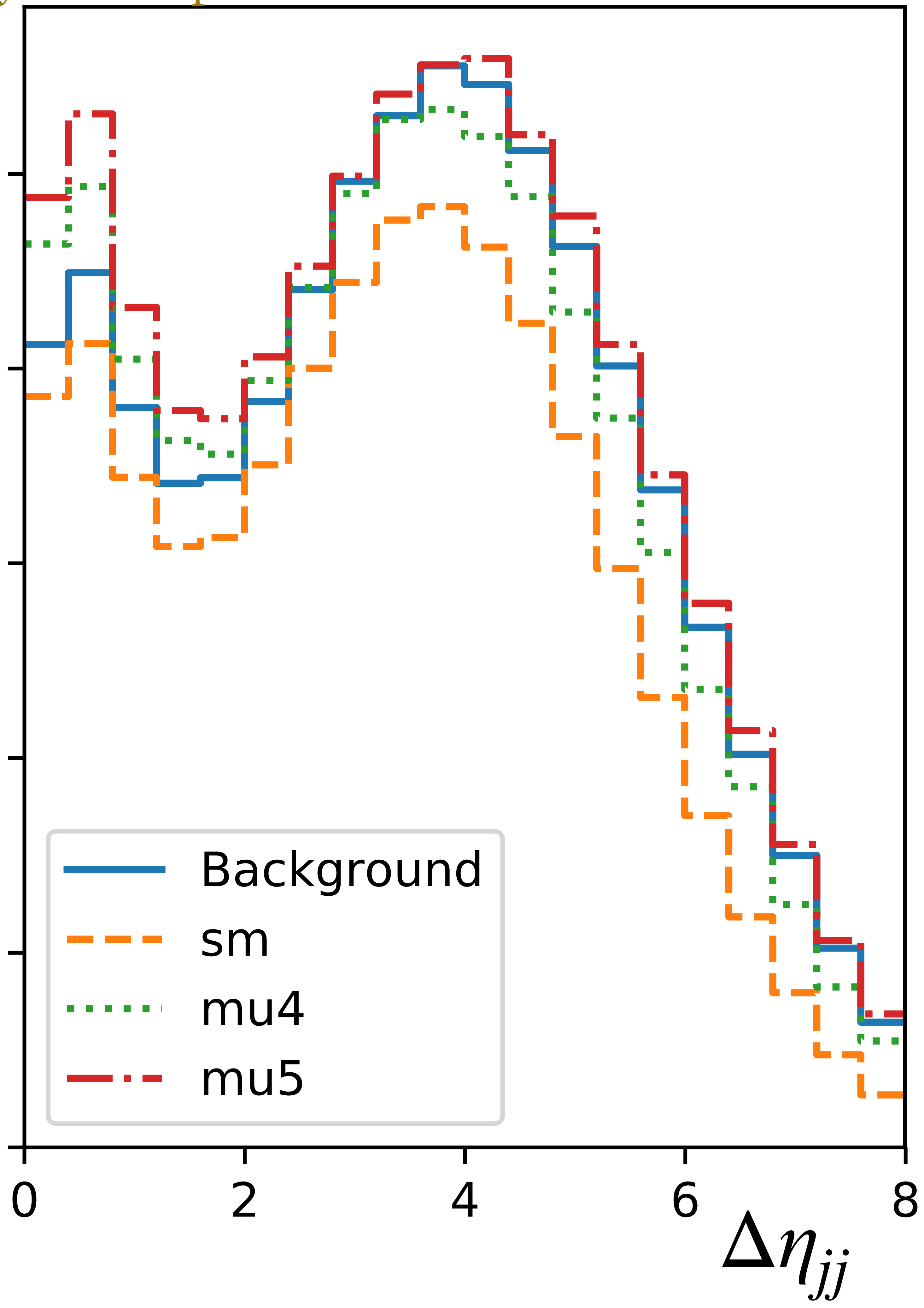
Signal-background-inference simulations (VBF processes): MG + Pythia+ Delphes



Can you spot the green plot?

μ=4 indistinguishable from μ=0 but other observables can break the degeneracy



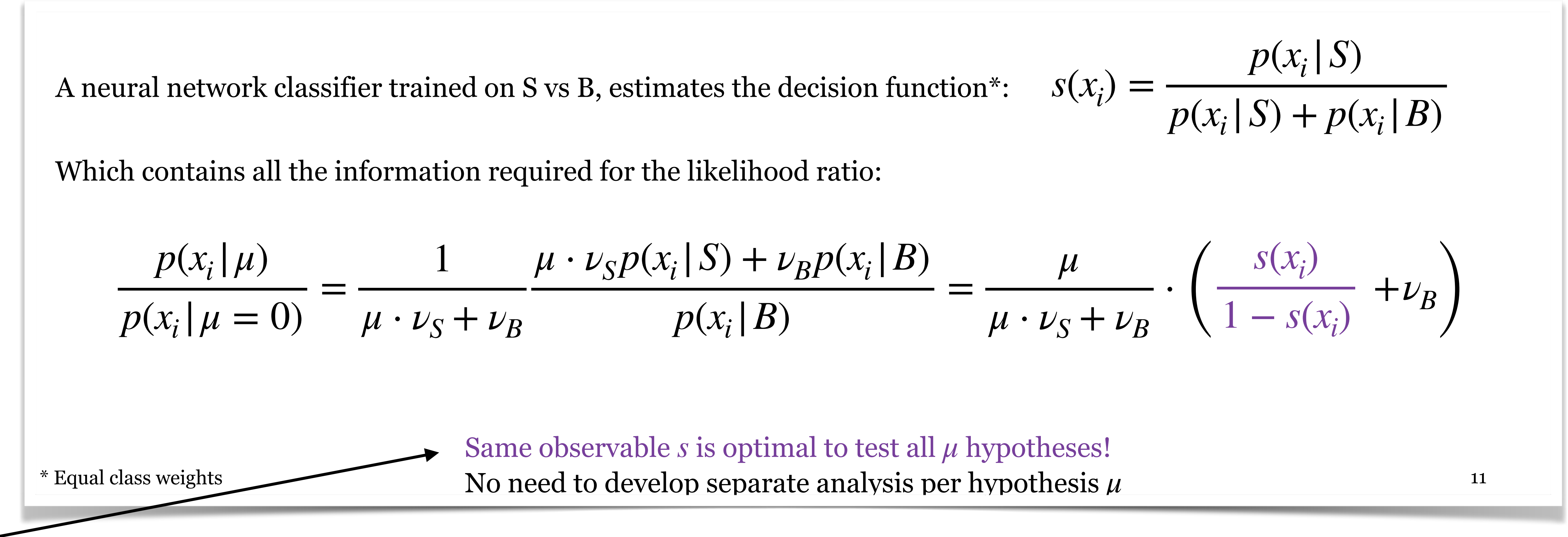


Optimal observable now changes as a function of **μ**: Cannot collapse problem to 1 dimension

What breaks down?

$$P(X) = |M_s(X) + M_b(X)|^2 = \underbrace{|M_s(X)|^2}_{P_s(X)} + \underbrace{|M_b(X)|^2}_{P_b(X)} + 2\,\underbrace{\texttt{Re}(\overline{M_s(X)}M_b(X))}_{P_i(X)}$$

$$N_{exp} = \mu \cdot S + B + \sqrt{\mu} \cdot I$$



A neural network classifier trained on S vs B, estimates the decision function*: $s(x_i) = \frac{p(x_i|S)}{p(x_i|S) + p(x_i|B)}$

Which contains all the information required for the likelihood ratio:

$$\frac{p(x_i|\mu)}{p(x_i|\mu=0)} = \frac{1}{\mu \cdot \nu_S + \nu_B} \frac{\mu \cdot \nu_S p(x_i|S) + \nu_B p(x_i|B)}{p(x_i|B)} = \frac{\mu}{\mu \cdot \nu_S + \nu_B} \cdot \left(\frac{s(x_i)}{1 - s(x_i)} + \nu_B \right)$$

* Equal class weights

Same observable s is optimal to test all μ hypotheses!
No need to develop separate analysis per hypothesis μ

11

No longer in this convenient special case: The same observable no longer optimal due to non-linear effects coming from quantum interference

Also does not generalise to an arbitrary theory parameter θ, (eg. Effective Field Theory parameters), or systematics

Can we modify the HEP statistical methodology to design near-optimal analyse for the general case?

Eg. arXiv:1805.00020: Brehmer et al, arXiv:1506.02169: Cranmer et al, PRD 104.056026: Ghosh et al

Cranmer et al: arXiv:1506.02169

High-dimensional density estimation is possible with NNs

$$\mathscr{L}(\mu \mid \mathscr{D}) = p(\mathscr{D} \mid \mu)$$

Neyman–Pearson lemma: Likelihood ratio is the most powerful test statistic

We want to compare likelihoods:

$$\frac{p(\mathscr{D} \mid \mu)}{p(\mathscr{D} \mid ref)}$$

A neural network classifier trained on **simulated samples from** μ_1 **vs simulated samples from** *ref*, estimates the decision function:

$$s(x_i) = \frac{p(x_i \mid \mu_1)}{p(x_i \mid \mu_1) + p(x_i \mid ref)}$$

Which contains all the information required for the likelihood ratio:

$$\frac{p(x_i \mid \mu_1)}{p(x_i \mid ref)} = \frac{s(x_i)}{1 - s(x_i)}$$

- ✴ Optimal statistic to test each value of μ
- ✴ We get the LR *per event* (unbinned)

Final test statistic

x_i is one individual event

$$\frac{L_{\text{full}}(\mu, \alpha | \mathcal{D})}{L_{\text{ref}}(\mathcal{D})} = \text{Pois}(N_{\text{data}} | \nu(\mu, \alpha)) \prod_{i}^{N_{\text{data}}} \frac{p(x_i | \mu, \alpha)}{p_{\text{ref}}(x_i)} \prod_{k} \text{Gaus}(a_k | \alpha_k, \delta_k)$$

ATLAS methods paper: arXiv:2412.01600

Final test statistic

x_i is one individual event

Now includes nuisances α

$$\frac{L_{\text{full}}(\mu, \alpha | \mathcal{D})}{L_{\text{ref}}(\mathcal{D})} = \text{Pois}(N_{\text{data}} | \nu(\mu, \alpha)) \prod_{i}^{N_{\text{data}}} \boxed{\frac{p(x_i | \mu, \alpha)}{p_{\text{ref}}(x_i)}} \prod_{k} \text{Gaus}(a_k | \alpha_k, \delta_k)$$

ATLAS methods paper: arXiv:2412.01600

Final test statistic

x_i is one individual event

Now includes nuisances α

$$\frac{L_{\text{full}}(\mu,\alpha|\mathcal{D})}{L_{\text{ref}}(\mathcal{D})} = \text{Pois}(N_{\text{data}}|\nu(\mu,\alpha)) \prod_{i}^{N_{\text{data}}} \frac{p(x_i|\mu,\alpha)}{p_{\text{ref}}(x_i)} \prod_{k} \text{Gaus}(a_k|\alpha_k,\delta_k)$$

Prod over events

ATLAS methods paper: arXiv:2412.01600

Final test statistic

x_i is one individual event

Now includes nuisances α

$$\frac{L_{\text{full}}(\mu,\alpha|\mathcal{D})}{L_{\text{ref}}(\mathcal{D})} = \text{Pois}(N_{\text{data}}|\nu(\mu,\alpha)) \prod_{i}^{N_{\text{data}}} \frac{p(x_i|\mu,\alpha)}{p_{\text{ref}}(x_i)} \prod_{k} \text{Gaus}(a_k|\alpha_k,\delta_k)$$

Rate term

Prod over events

ATLAS methods paper: [arXiv:2412.01600](https://arxiv.org/abs/2412.01600)

Final test statistic

x_i is one individual event

Now includes nuisances α

$$\frac{L_{\text{full}}(\mu, \alpha | \mathcal{D})}{L_{\text{ref}}(\mathcal{D})} = \text{Pois}(N_{\text{data}} | \nu(\mu, \alpha)) \prod_{i}^{N_{\text{data}}} \frac{p(x_i | \mu, \alpha)}{p_{\text{ref}}(x_i)} \prod_{k} \text{Gaus}(a_k | \alpha_k, \delta_k)$$

Rate term

Constrain term

Prod over events

ATLAS methods paper: arXiv:2412.01600

Final test statistic

x_i is one individual event

Now includes nuisances α

$$\frac{L_{\text{full}}(\mu,\alpha|\mathcal{D})}{L_{\text{ref}}(\mathcal{D})} = \text{Pois}(N_{\text{data}}|\nu(\mu,\alpha)) \prod_{i}^{N_{\text{data}}} \frac{p(x_i|\mu,\alpha)}{p_{\text{ref}}(x_i)} \prod_k \text{Gaus}(a_k|\alpha_k,\delta_k)$$

Rate term

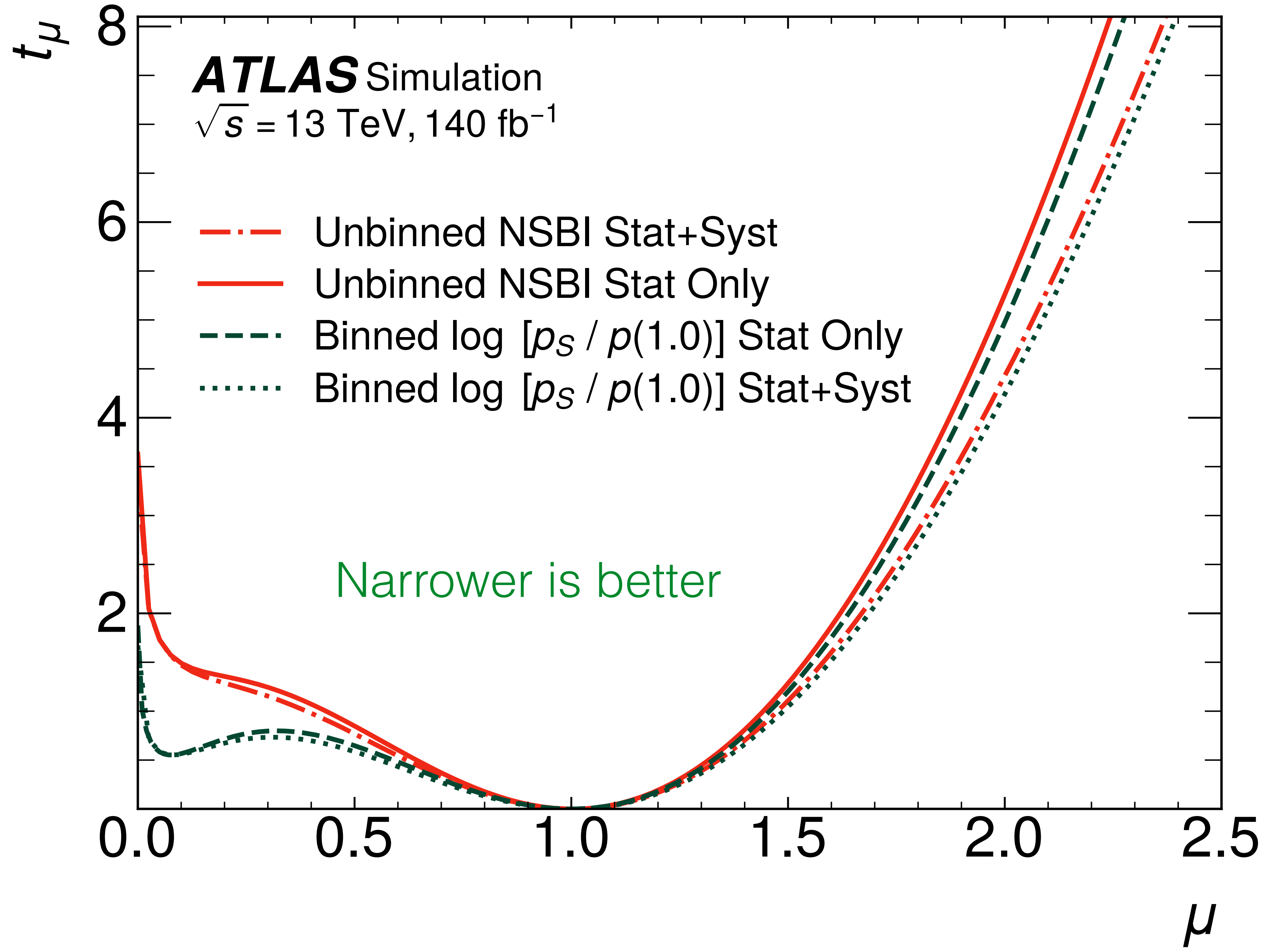
Constrain term

Prod over events

Profiling:

$$t_\mu = -2\ln\left(\frac{L_{\text{full}}(\mu,\hat{\hat{\alpha}})/\cancel{L_{\text{ref}}}}{L_{\text{full}}(\hat{\mu},\hat{\alpha})/\cancel{L_{\text{ref}}}}\right)$$

This is why we define p_{ref} to be independent of μ



ATLAS methods paper: arXiv:2412.01600

Final test statistic

x_i is one individual event

Now includes nuisances α

$$\frac{L_\text{full}(\mu, \alpha | \mathcal{D})}{L_\text{ref}(\mathcal{D})} = \text{Pois}(N_\text{data} | \nu(\mu, \alpha)) \prod_i^{N_\text{data}} \frac{p(x_i | \mu, \alpha)}{p_\text{ref}(x_i)} \prod_k \text{Gaus}(a_k | \alpha_k, \delta_k)$$

Rate term

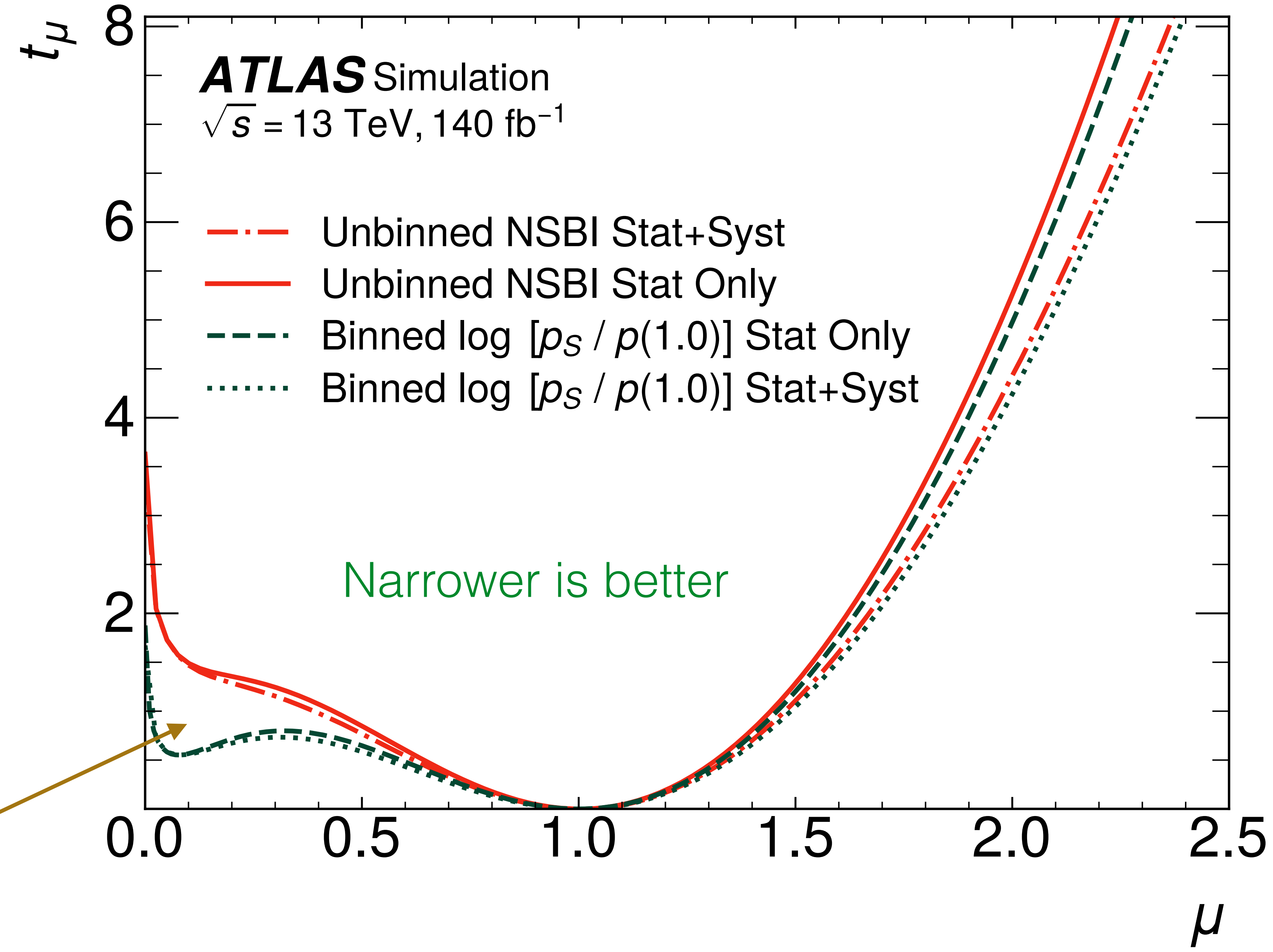
Prod over events

Constrain term

Profiling:

$$t_\mu = -2 \ln \left(\frac{L_\text{full}(\mu, \hat{\hat{\alpha}}) / \cancel{L_\text{ref}}}{L_\text{full}(\hat{\mu}, \hat{\alpha}) / \cancel{L_\text{ref}}} \right)$$

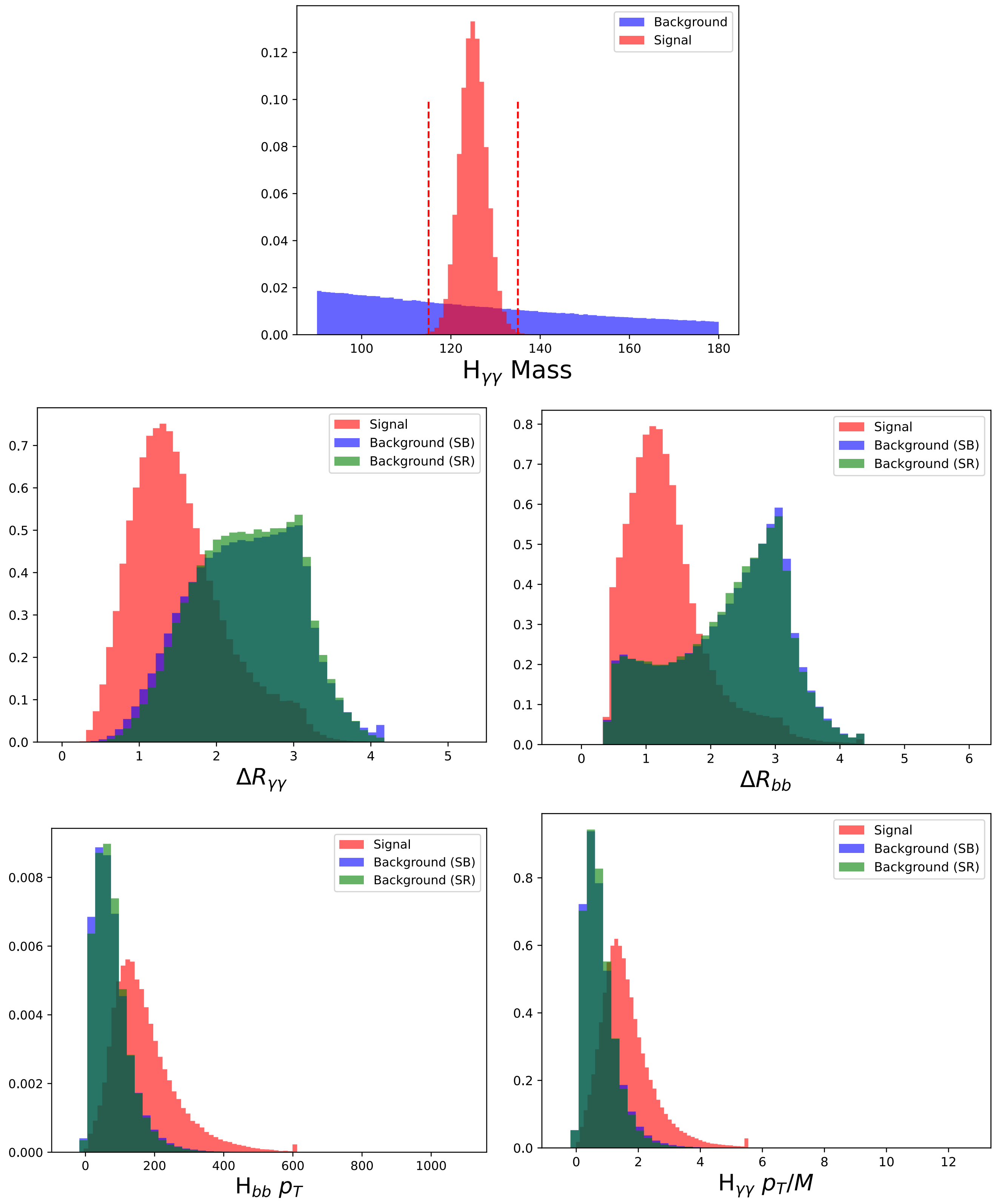
This is why we define p_{ref} to be independent of μ



Non-parabolic shape due to non-linear effects from quantum interference

Unreliable simulator?

- Training data can come from control regions in ‘likelihood-free inference’ !
- Port the usual data-driven background estimation techniques eg. in any HH analyses

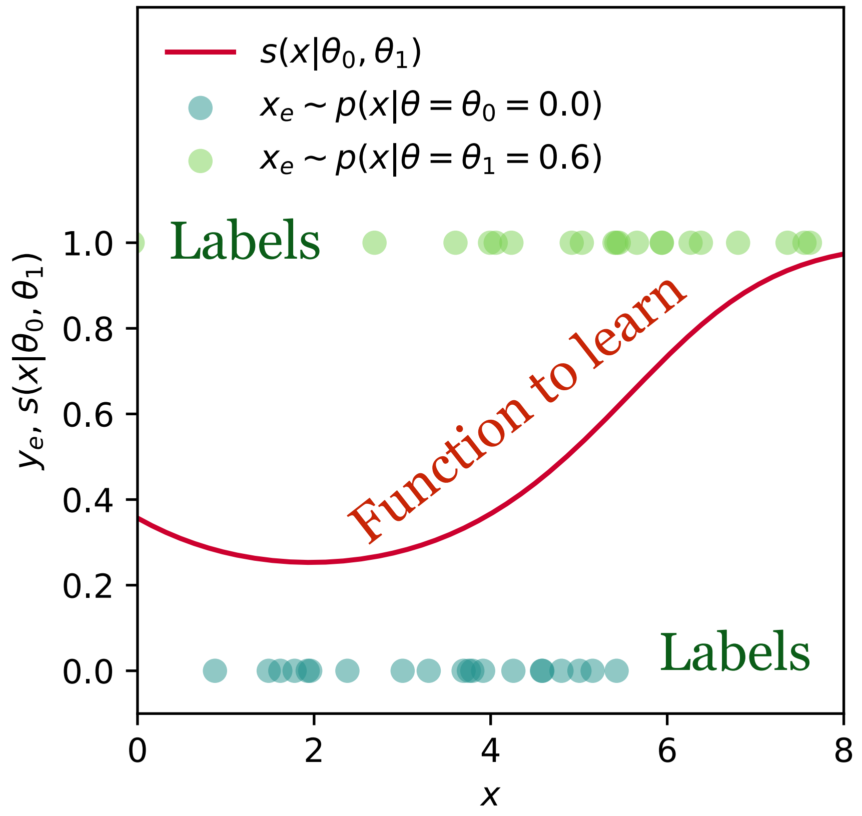


arXiv:2506.06438: Amram & Szewc

Where would you want to use NSBI?

- If you're annoyed by non-linear effects due to quantum interference

- **Effective field theory**:
  - Highly non-linear with parameters of interest
  - But **challenging to train networks on subtle differences** in phase space!
  - Brehmer et al (PRD 98.052004) show that you can use matrix-element information to enhance training. But ME typically not available for background processes
  - Tae Park designed a way to combine both (arXiv:2507.02032), shown for Higgs self-coupling

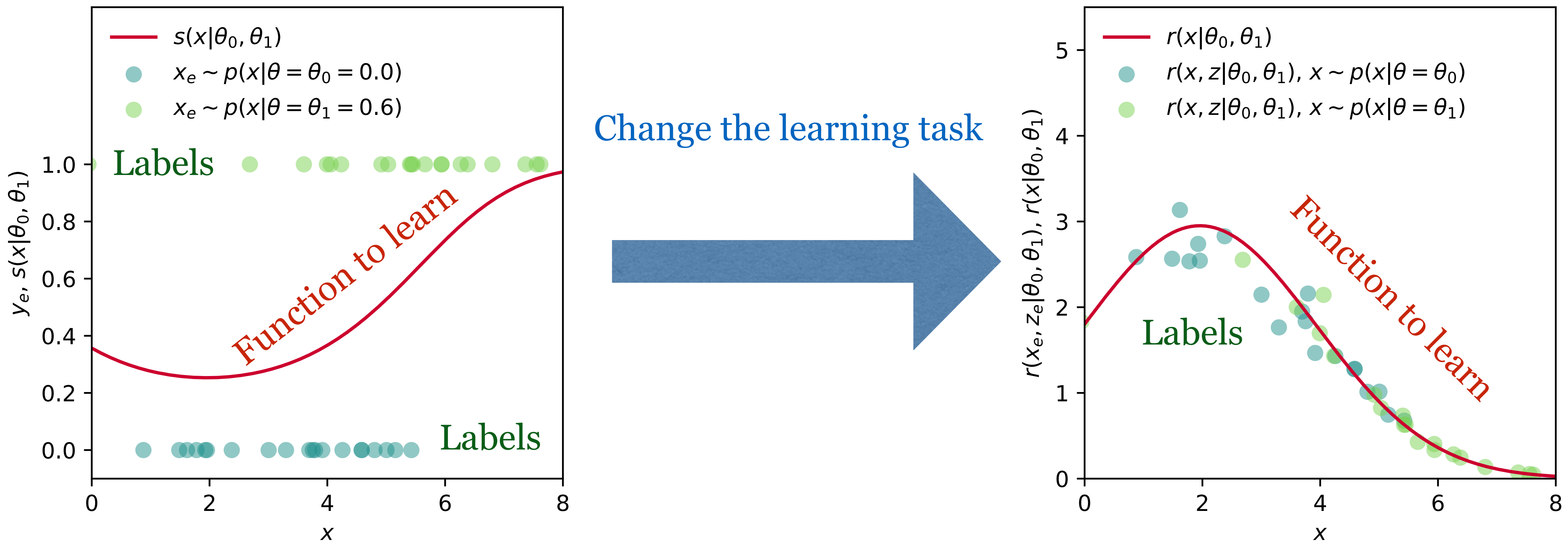


Classification

Where would you want to use NSBI?

- If you're annoyed by non-linear effects due to quantum interference

- **Effective field theory**:
  - Highly non-linear with parameters of interest
  - But **challenging to train networks on subtle differences** in phase space!
  - Brehmer et al (PRD 98.052004) show that you can use matrix-element information to enhance training. But ME typically not available for background processes
  - Tae Park designed a way to combine both (arXiv:2507.02032), shown for Higgs self-coupling



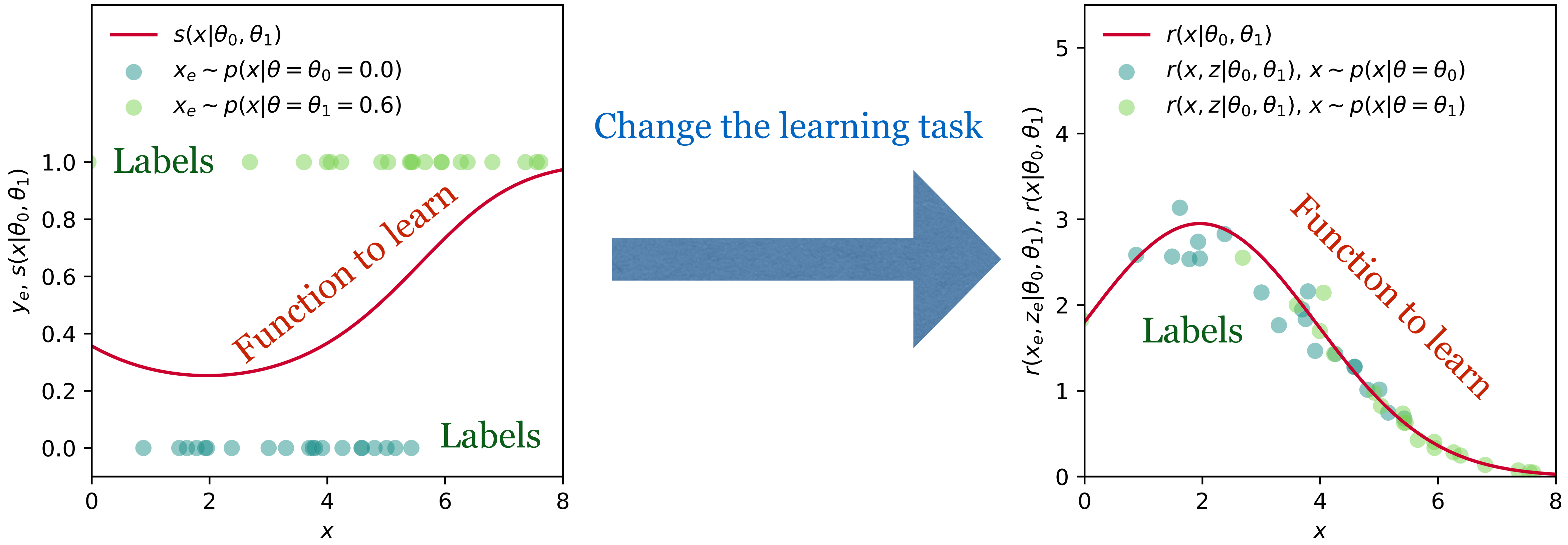
Classification

ME regression

Where would you want to use NSBI?

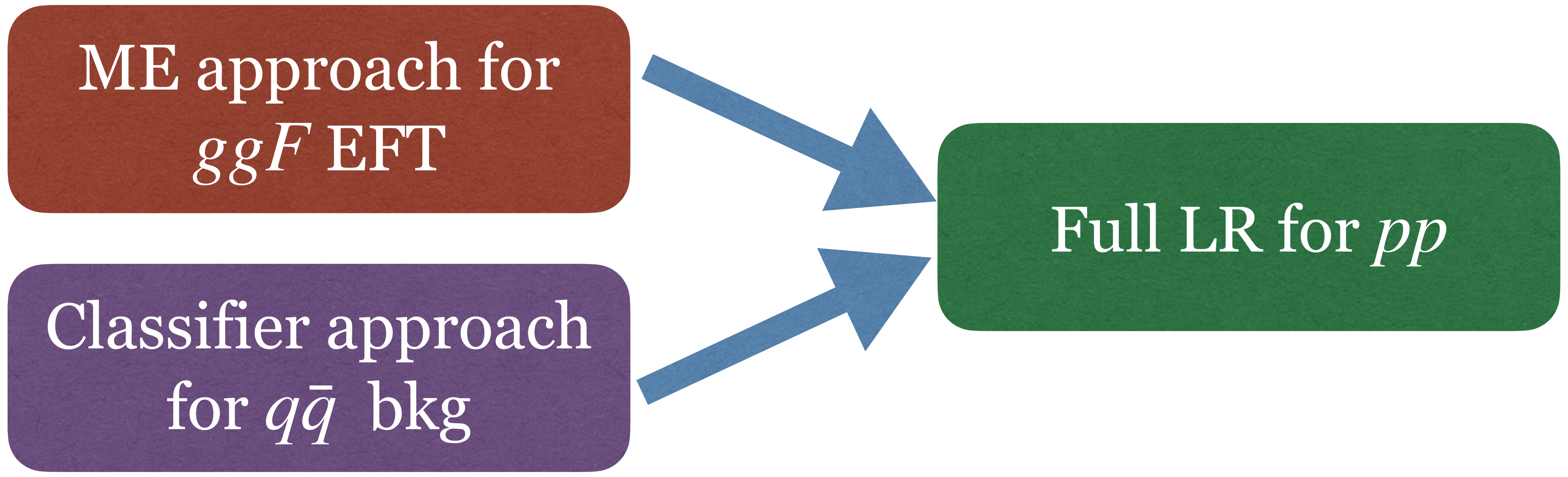
- If you’re annoyed by non-linear effects due to quantum interference

- **Effective field theory**:
  - Highly non-linear with parameters of interest
  - But **challenging to train networks on subtle differences** in phase space!
  - Brehmer et al (PRD 98.052004) show that you can use matrix-element information to enhance training. But ME typically not available for background processes
  - Tae Park designed a way to combine both (arXiv:2507.02032), shown for Higgs self-coupling



Classification

ME regression



ME approach for ggF EFT

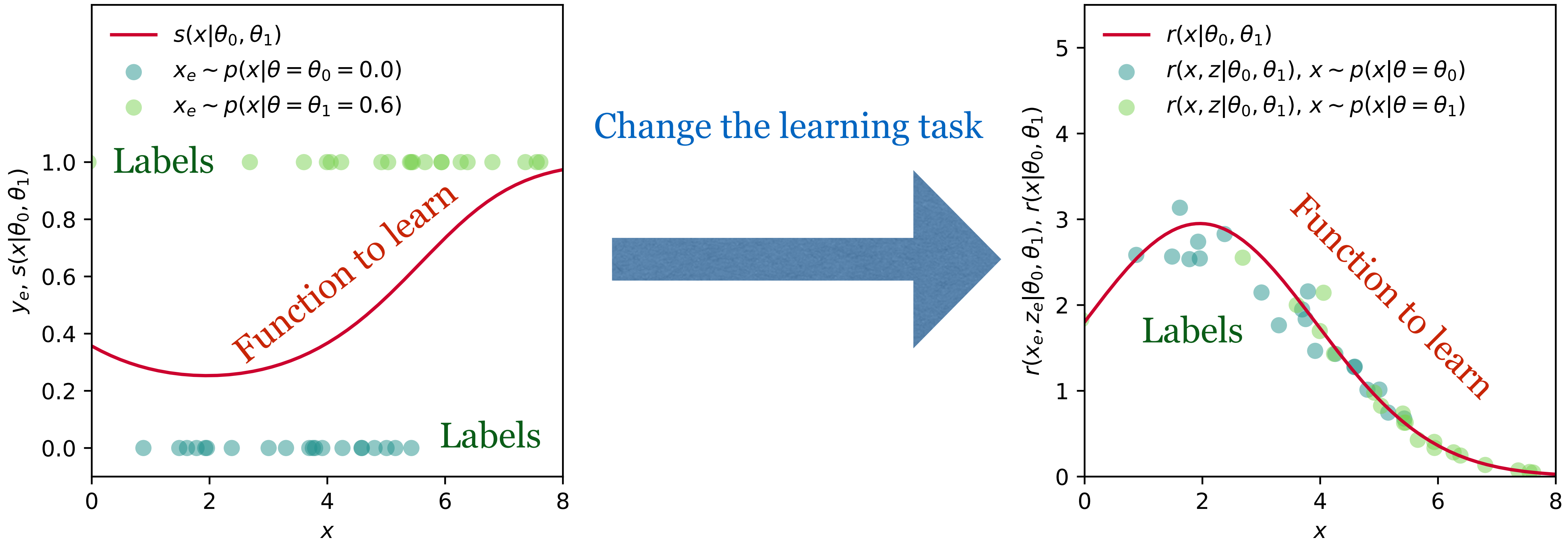
Classifier approach for $q\bar{q}$ bkg

Full LR for pp

Where would you want to use NSBI?

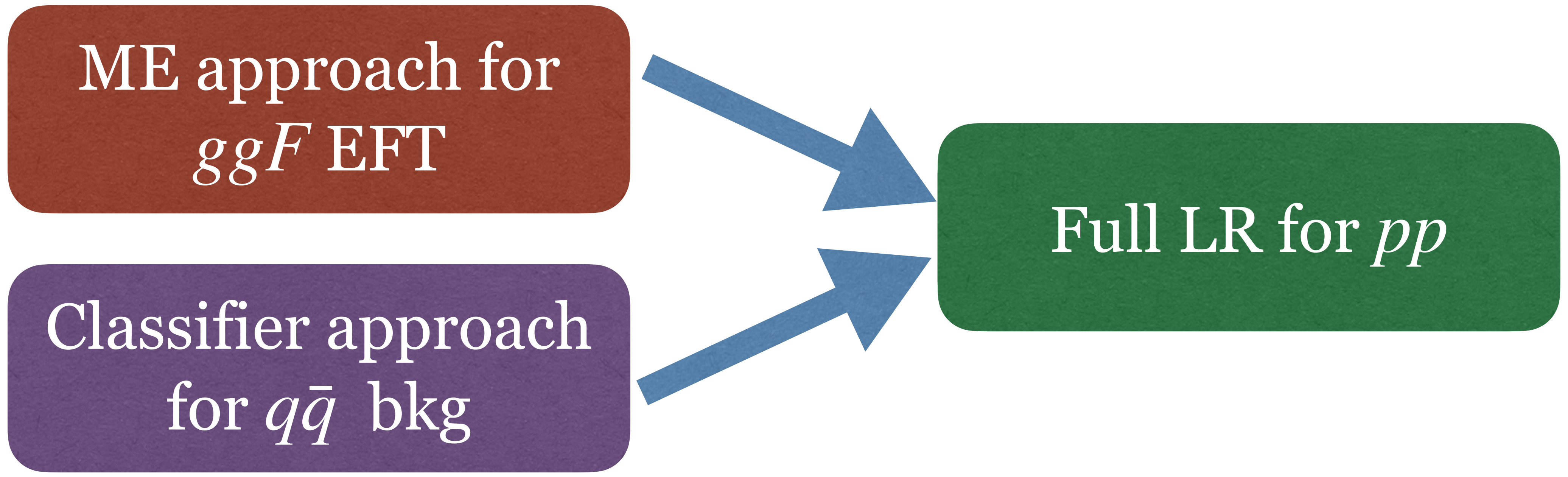
- If you're annoyed by non-linear effects due to quantum interference

- **Effective field theory**:
  - Highly non-linear with parameters of interest
  - But **challenging to train networks on subtle differences** in phase space!
  - Brehmer et al (PRD 98.052004) show that you can use matrix-element information to enhance training. But ME typically not available for background processes
  - Tae Park designed a way to combine both (arXiv:2507.02032), shown for Higgs self-coupling



Classification

ME regression



ME approach for ggF EFT

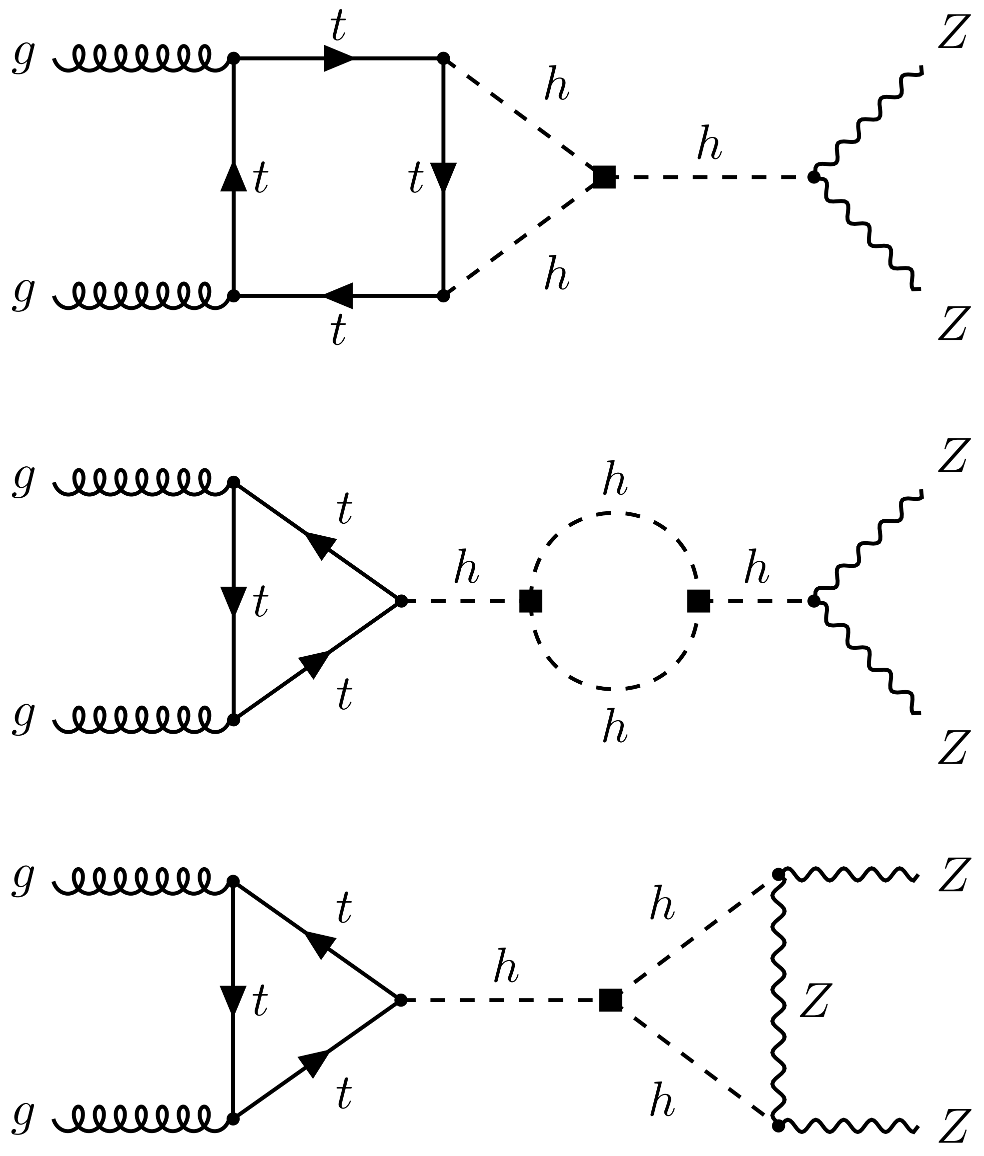
Classifier approach for $q\bar{q}$ bkg

Full LR for pp

Also see generative-model-based NSBI in Oz Amram's talk (for HH), Sascha Diefenbacher's talk

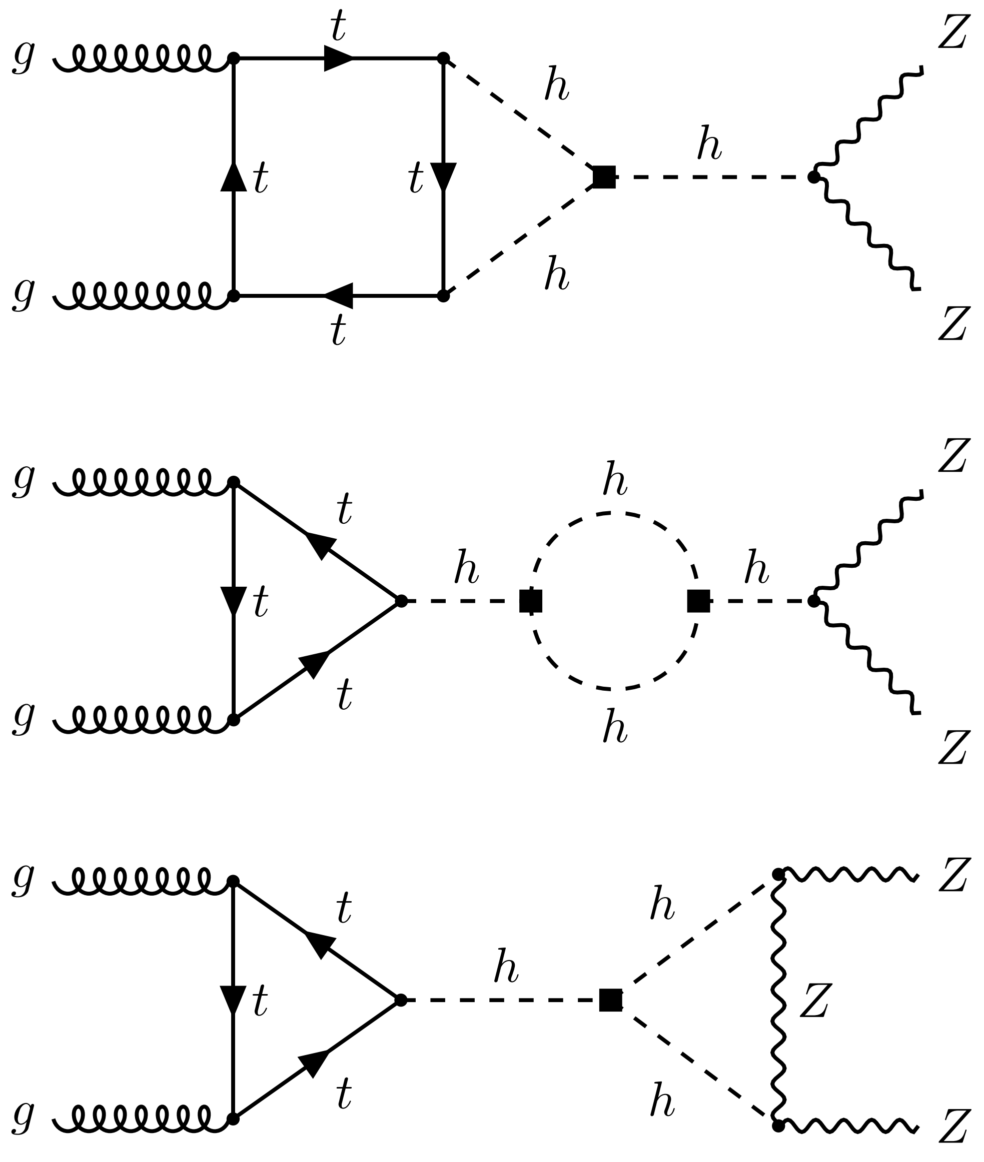
Comparison to applying NSBI on unfolded data in Kevin Grief's talk arXiv:2507.02032: Ghosh, Griese, Haisch & Park

NSBI for Higgs self-coupling

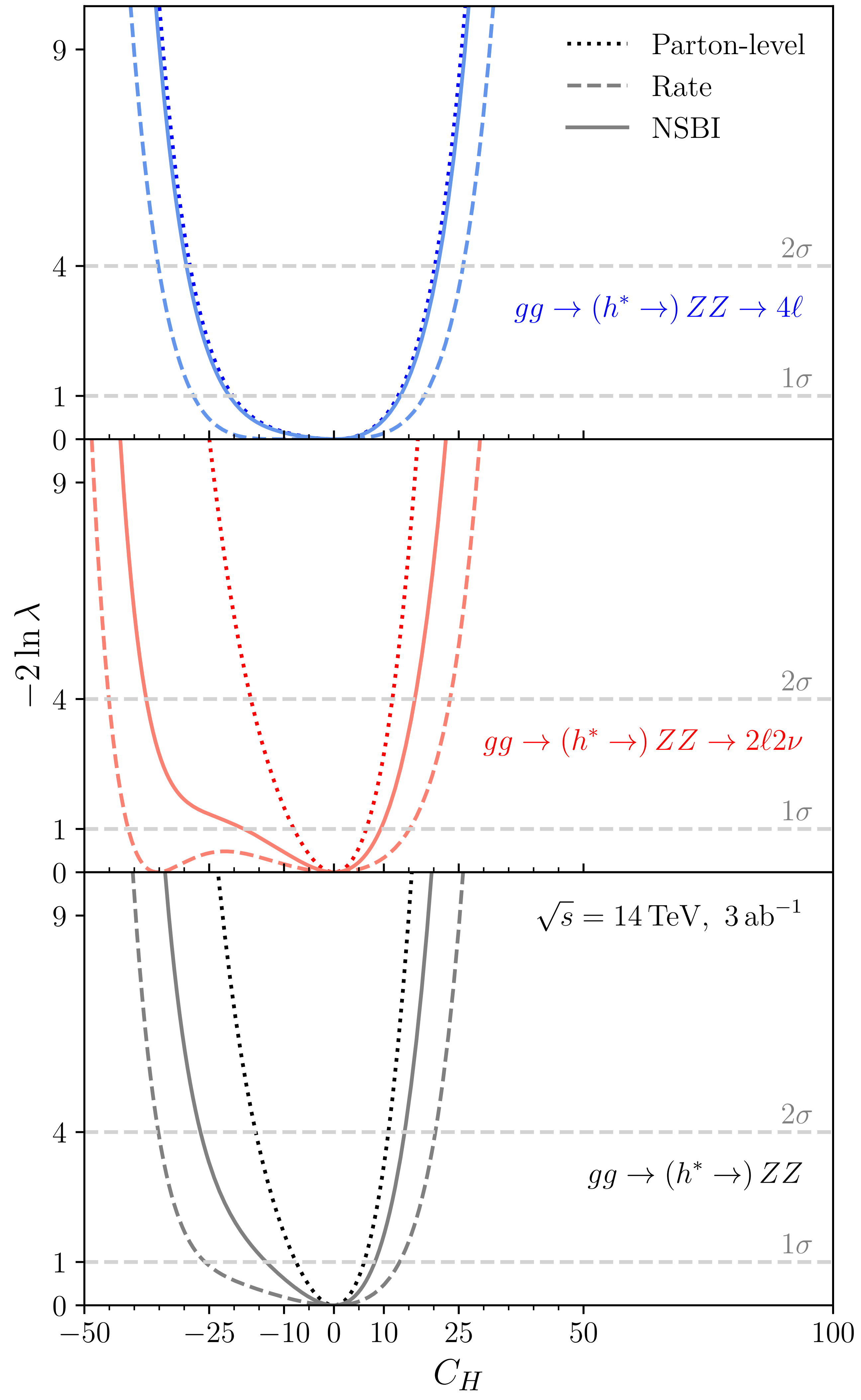


Can be probed in off-shell single Higgs channel arXiv:2507.02032: Ghosh, Griese, Haisch & Park

NSBI for Higgs self-coupling



Can be probed in off-shell single Higgs channel



Learning subtle differences required matrix-element-enhanced network training + factorisation tricks

llνν: NSBI stable despite missing neutrinos!

Search-Oriented Mixture Model

x_i vector representing one individual event

General Formula

$$p(x_i|\mu) = \frac{1}{\nu(\mu)} \sum_{j}^{C} f_j(\mu) \cdot \nu_j \; p_j(x_i)$$

j runs over different physics process
(Eg. $gg \rightarrow H^* \rightarrow 4l$, $gg \rightarrow ZZ \rightarrow 4l$)

Example use case

Search-Oriented Mixture Model

x_i vector representing one individual event

General Formula

$$p(x_i|\mu) = \frac{1}{\nu(\mu)} \sum_{j}^{C} f_j(\mu) \cdot \nu_j \; p_j(x_i)$$

j runs over different physics process
(Eg. $gg \to H^* \to 4l$, $gg \to ZZ \to 4l$)

Example use case

$$p_{\mathrm{ggF}}(x|\mu) = \frac{1}{\nu_{\mathrm{ggF}}(\mu)} \left[(\mu - \sqrt{\mu})\, \nu_S \; p_{\mathrm{S}}(x) + \sqrt{\mu}\, \nu_{\mathrm{SBI}_1} \; p_{\mathrm{SBI}_1}(x) + (1 - \sqrt{\mu}) \nu_{\mathrm{B}} \; p_{\mathrm{B}}(x) \right]$$

ATLAS methods paper: arXiv:2412.01600

Search-Oriented Mixture Model

x_i vector representing one individual event

General Formula

$$p(x_i|\mu) = \frac{1}{\nu(\mu)} \sum_j^C \boxed{f_j(\mu)} \cdot \nu_j \; p_j(x_i)$$

j runs over different physics process
(Eg. $gg \rightarrow H^* \rightarrow 4l$, $gg \rightarrow ZZ \rightarrow 4l$)

Comes from theory model chosen to interpret data

Example use case

$$p_{\mathrm{ggF}}(x|\mu) = \frac{1}{\nu_{\mathrm{ggF}}(\mu)} \left[\underline{(\mu - \sqrt{\mu})} \, \nu_S \, p_S(x) + \underline{\sqrt{\mu}} \, \nu_{\mathrm{SBI}_1} \, p_{\mathrm{SBI}_1}(x) + \underline{(1 - \sqrt{\mu})} \nu_{\mathrm{B}} \, p_{\mathrm{B}}(x) \right]$$

$f_i(\mu)$ will depend on morphing bases points (which values of μ were used to simulate samples)

ATLAS methods paper: arXiv:2412.01600

Search-Oriented Mixture Model

x_i vector representing one individual event

General Formula

$$p(x_i|\mu) = \frac{1}{\nu(\mu)} \sum_j^C f_j(\mu) \cdot \nu_j \, p_j(x_i)$$

Event rates estimated from simulations

Comes from theory model chosen to interpret data

j runs over different physics process
(Eg. $gg \to H^* \to 4l$, $gg \to ZZ \to 4l$)

Example use case

$$p_{\mathrm{ggF}}(x|\mu) = \frac{1}{\underline{\nu_{\mathrm{ggF}}(\mu)}} \left[\underline{(\mu - \sqrt{\mu})}\, \underline{\nu_S}\, p_S(x) + \underline{\sqrt{\mu}}\, \underline{\nu_{\mathrm{SBI}_1}}\, p_{\mathrm{SBI}_1}(x) + \underline{(1 - \sqrt{\mu})}\, \underline{\nu_{\mathrm{B}}}\, p_{\mathrm{B}}(x) \right]$$

$f_i(\mu)$ will depend on morphing bases points (which values of μ were used to simulate samples)

ATLAS methods paper: [arXiv:2412.01600](https://arxiv.org/abs/2412.01600)

Search-Oriented Mixture Model

x_i vector representing one individual event

General Formula

$$p(x_i|\mu) = \frac{1}{\nu(\mu)} \sum_{j}^{C} f_j(\mu) \cdot \nu_j \, p_j(x_i)$$

?

Event rates estimated from simulations

Comes from theory model chosen to interpret data

j runs over different physics process
(Eg. $gg \rightarrow H^* \rightarrow 4l$, $gg \rightarrow ZZ \rightarrow 4l$)

Example use case

$$p_{\mathrm{ggF}}(x|\mu) = \frac{1}{\underline{\nu_{\mathrm{ggF}}(\mu)}} \left[\underline{(\mu - \sqrt{\mu})}\,\underline{\nu_S}\, p_S(x) + \underline{\sqrt{\mu}}\,\underline{\nu_{\mathrm{SBI}_1}}\, p_{\mathrm{SBI}_1}(x) + \underline{(1 - \sqrt{\mu})}\,\underline{\nu_{\mathrm{B}}}\, p_{\mathrm{B}}(x) \right]$$

$f_i(\mu)$ will depend on morphing bases points (which values of μ were used to simulate samples)

ATLAS methods paper: arXiv:2412.01600

Search-Oriented Mixture Model

x_i vector representing one individual event

General Formula

$$p(x_i|\mu) = \frac{1}{\nu(\mu)} \sum_j^C f_j(\mu) \cdot \nu_j \, p_j(x_i)$$

?

$$\frac{p(x_i|\mu)}{p_{\text{ref}}(x_i)} = \frac{1}{\nu(\mu)} \sum_j^C f_j(\mu) \cdot \nu_j \, \frac{p_j(x_i)}{p_{\text{ref}}(x_i)}$$

Reference hypothesis

j runs over different physics process (Eg. $gg \to H^* \to 4l$, $gg \to ZZ \to 4l$)

Event rates estimated from simulations

Comes from theory model chosen to interpret data

Example use case

$$p_{\text{ggF}}(x|\mu) = \frac{1}{\underline{\nu_{\text{ggF}}(\mu)}} \left[\underline{(\mu - \sqrt{\mu})}\, \underline{\nu_S}\, p_S(x) + \underline{\sqrt{\mu}}\, \underline{\nu_{\text{SBI}_1}}\, p_{\text{SBI}_1}(x) + \underline{(1 - \sqrt{\mu})}\underline{\nu_B}\, p_B(x) \right]$$

$f_i(\mu)$ will depend on morphing bases points (which values of μ were used to simulate samples)

ATLAS methods paper: arXiv:2412.01600

Search-Oriented Mixture Model

x_i vector representing one individual event

General Formula

$$p(x_i|\mu) = \frac{1}{\nu(\mu)} \sum_j^C f_j(\mu) \cdot \nu_j \, p_j(x_i) \; ?$$

$$\frac{p(x_i|\mu)}{p_{\text{ref}}(x_i)} = \frac{1}{\nu(\mu)} \sum_j^C f_j(\mu) \cdot \nu_j \, \frac{p_j(x_i)}{p_{\text{ref}}(x_i)}$$

Reference hypothesis

j runs over different physics process (Eg. $gg \to H^* \to 4l$, $gg \to ZZ \to 4l$)

Event rates estimated from simulations

Comes from theory model chosen to interpret data

Example use case

$$p_{\text{ggF}}(x|\mu) = \frac{1}{\underline{\nu_{\text{ggF}}(\mu)}} \left[\underline{(\mu - \sqrt{\mu})} \, \underline{\nu_S} \, p_S(x) + \underline{\sqrt{\mu}} \, \underline{\nu_{\text{SBI}_1}} \, p_{\text{SBI}_1}(x) + \underline{(1 - \sqrt{\mu})} \underline{\nu_B} \, p_B(x) \right]$$

$$\frac{p(x|\mu)}{p_S(x)} = \frac{1}{\nu(\mu)} \left[(\mu - \sqrt{\mu}) \, \nu_S + \sqrt{\mu} \, \nu_{\text{SBI}_1} \frac{p_{\text{SBI}_1}(x)}{p_S(x)} + (1 - \sqrt{\mu}) \nu_B \frac{p_B(x)}{p_S(x)} \right]$$

$f_i(\mu)$ will depend on morphing bases points (which values of μ were used to simulate samples)

ATLAS methods paper: [arXiv:2412.01600](https://arxiv.org/abs/2412.01600)

Search-Oriented Mixture Model

---

x_i vector representing one individual event

General Formula

Estimated using an ensemble of networks

$$p(x_i|\mu) = \frac{1}{\nu(\mu)} \sum_j^C f_j(\mu) \cdot \nu_j \, p_j(x_i) \; ? \quad \Rightarrow \quad \frac{p(x_i|\mu)}{p_{\text{ref}}(x_i)} = \frac{1}{\nu(\mu)} \sum_j^C f_j(\mu) \cdot \nu_j \frac{p_j(x_i)}{p_{\text{ref}}(x_i)}$$

Reference hypothesis

j runs over different physics process (Eg. $gg \to H^* \to 4l$, $gg \to ZZ \to 4l$)

Event rates estimated from simulations

Comes from theory model chosen to interpret data

Example use case

$$p_{\text{ggF}}(x|\mu) = \frac{1}{\underline{\nu_{\text{ggF}}(\mu)}} \left[\underline{(\mu - \sqrt{\mu})} \, \underline{\nu_S} \, p_S(x) + \underline{\sqrt{\mu}} \, \underline{\nu_{\text{SBI}_1}} \, p_{\text{SBI}_1}(x) + \underline{(1 - \sqrt{\mu})} \underline{\nu_B} \, p_B(x) \right]$$

$$\Rightarrow \quad \frac{p(x|\mu)}{p_S(x)} = \frac{1}{\nu(\mu)} \left[(\mu - \sqrt{\mu}) \, \nu_S + \sqrt{\mu} \, \nu_{\text{SBI}_1} \frac{p_{\text{SBI}_1}(x)}{\underline{p_S(x)}} + (1 - \sqrt{\mu}) \nu_B \frac{p_B(x)}{\underline{p_S(x)}} \right]$$

$f_i(\mu)$ will depend on morphing bases points (which values of μ were used to simulate samples)

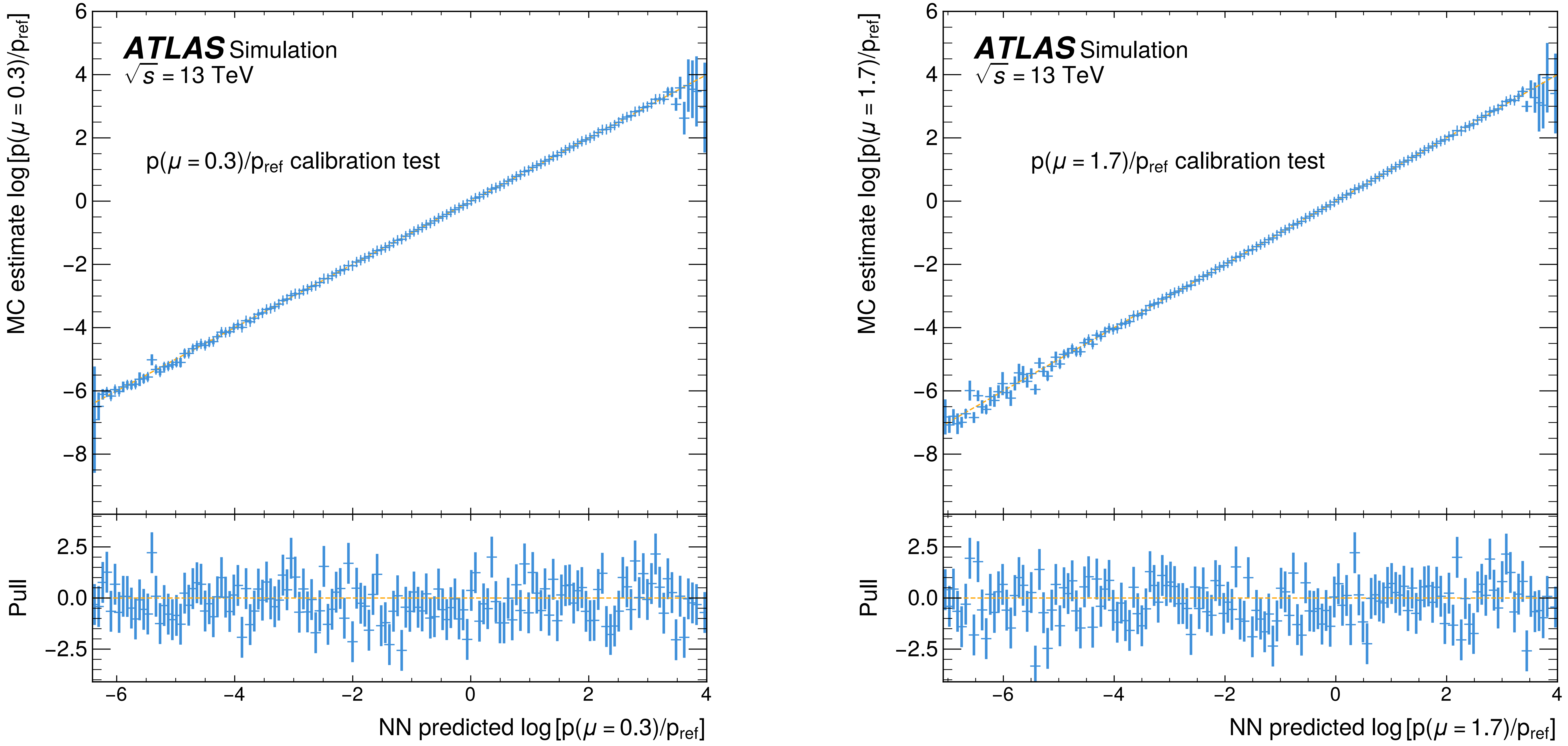
More diagnostics

Calibration curves of probability density ratios

$$\frac{p_{\mu=0.3}(x_i)}{p_{ref}(x_i)}$$

$$\frac{p_{\mu=1.7}(x_i)}{p_{ref}(x_i)}$$

Binned estimate



Perfect calibration would give $y = x$

Ensemble prediction

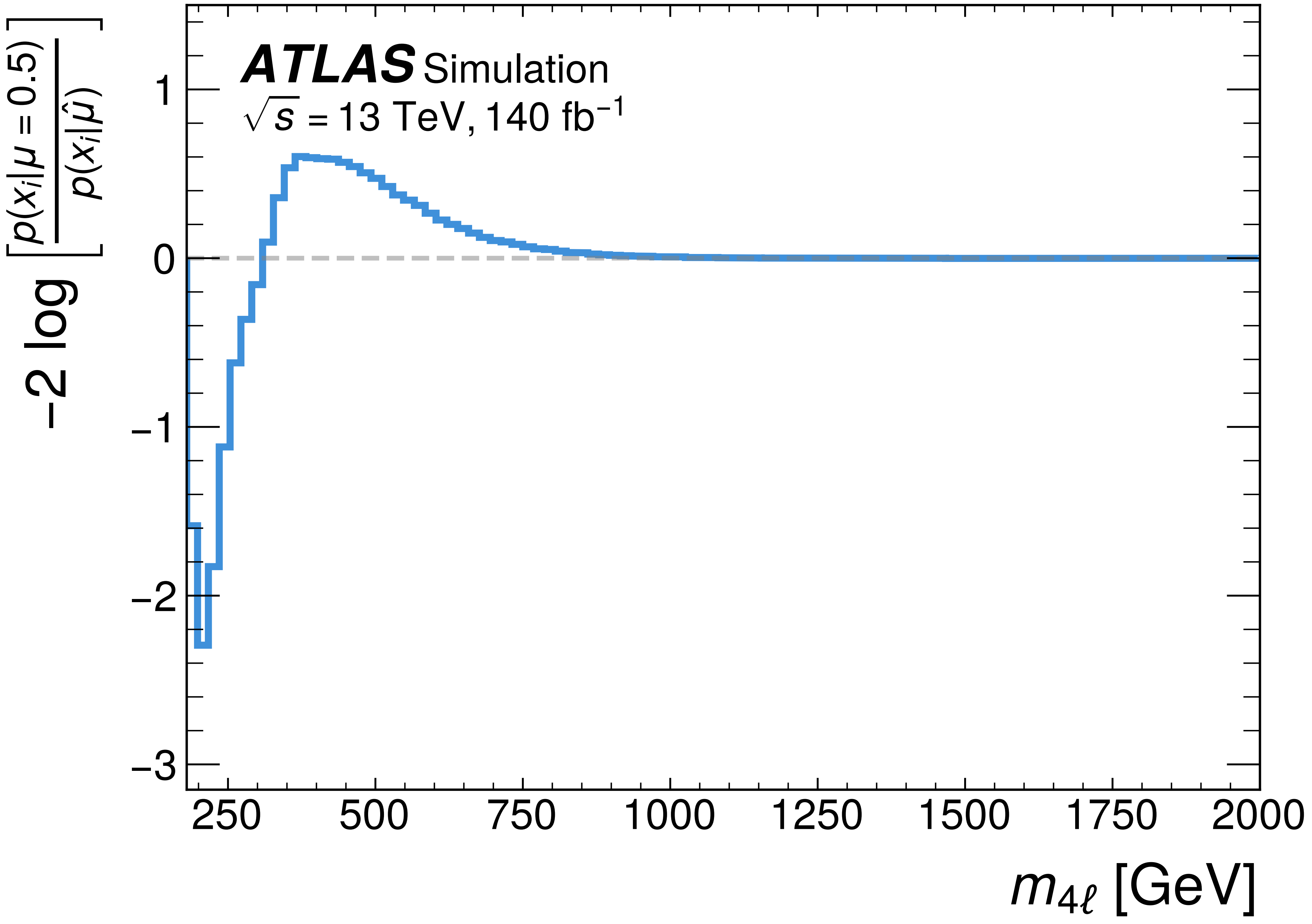
Ensemble prediction

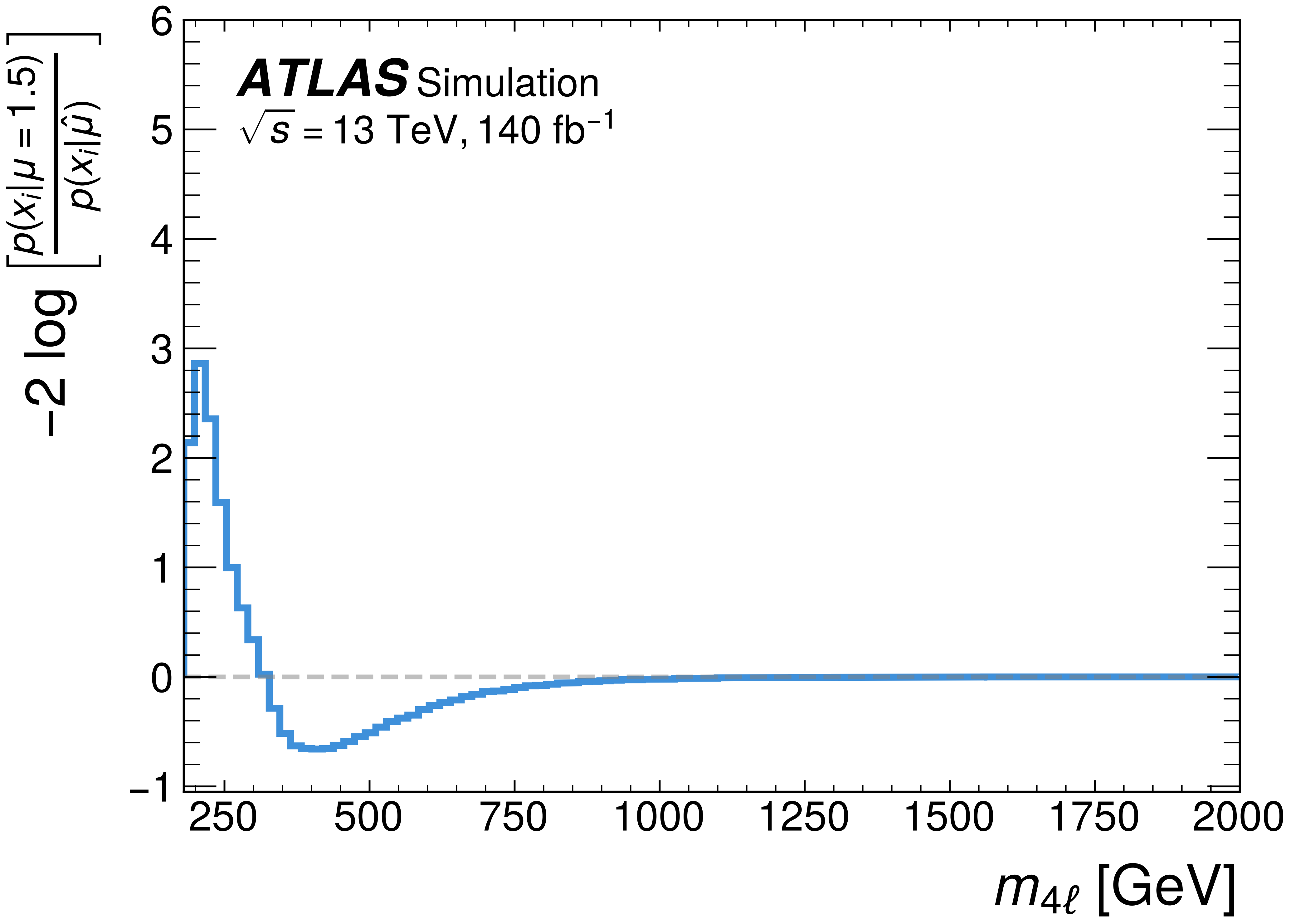
ATLAS methods paper: arXiv:2412.01600

Interpretability:
Which phase space favours one hypothesis over another?

$$-2 \cdot log \frac{P(x_i \,|\, \mu = 0.5)}{P(x_i \,|\, \mu = 1)}$$

$$-2 \cdot log \frac{P(x_i \,|\, \mu = 1.5)}{P(x_i \,|\, \mu = 1)}$$



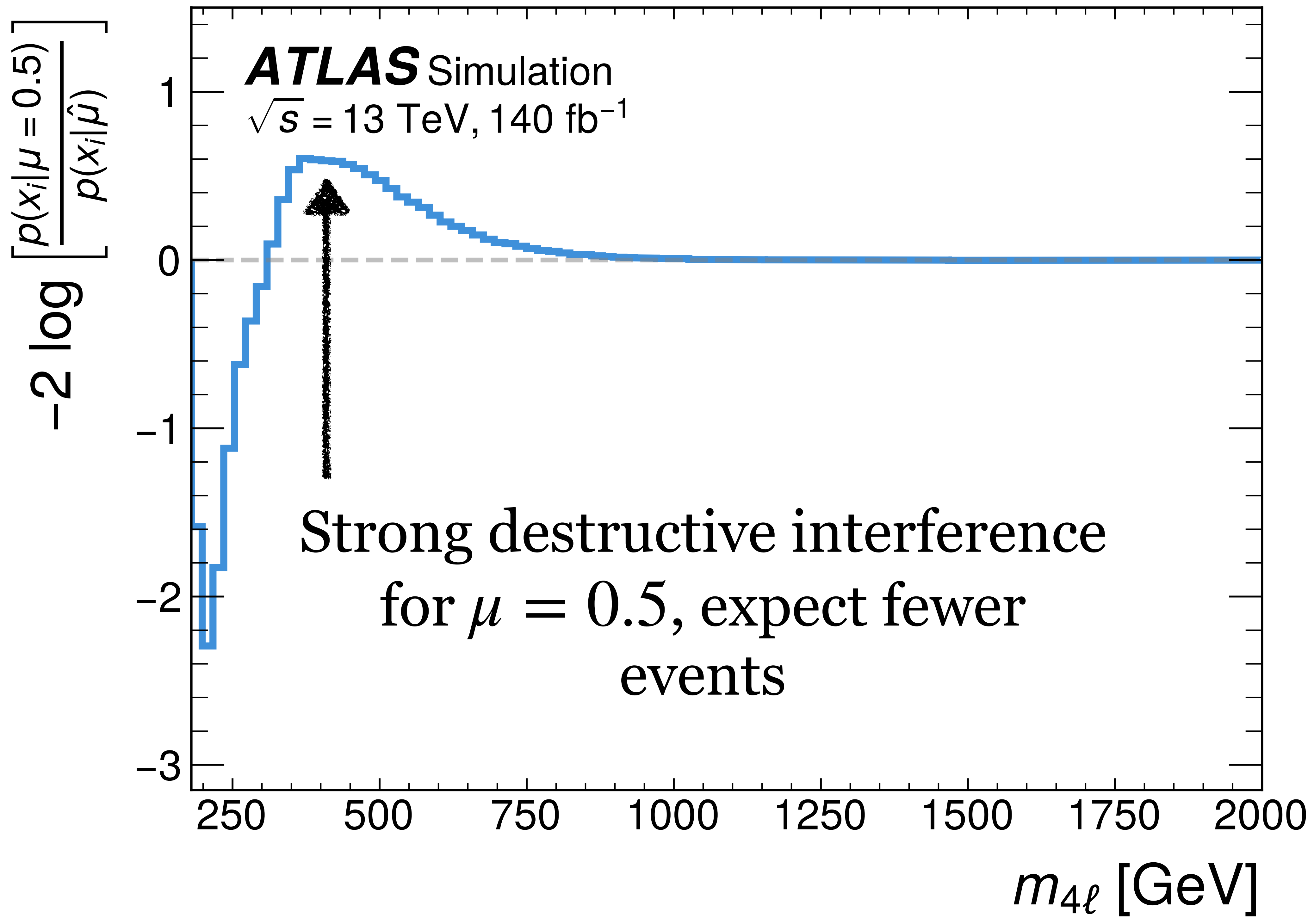


ATLAS methods paper: arXiv:2412.01600

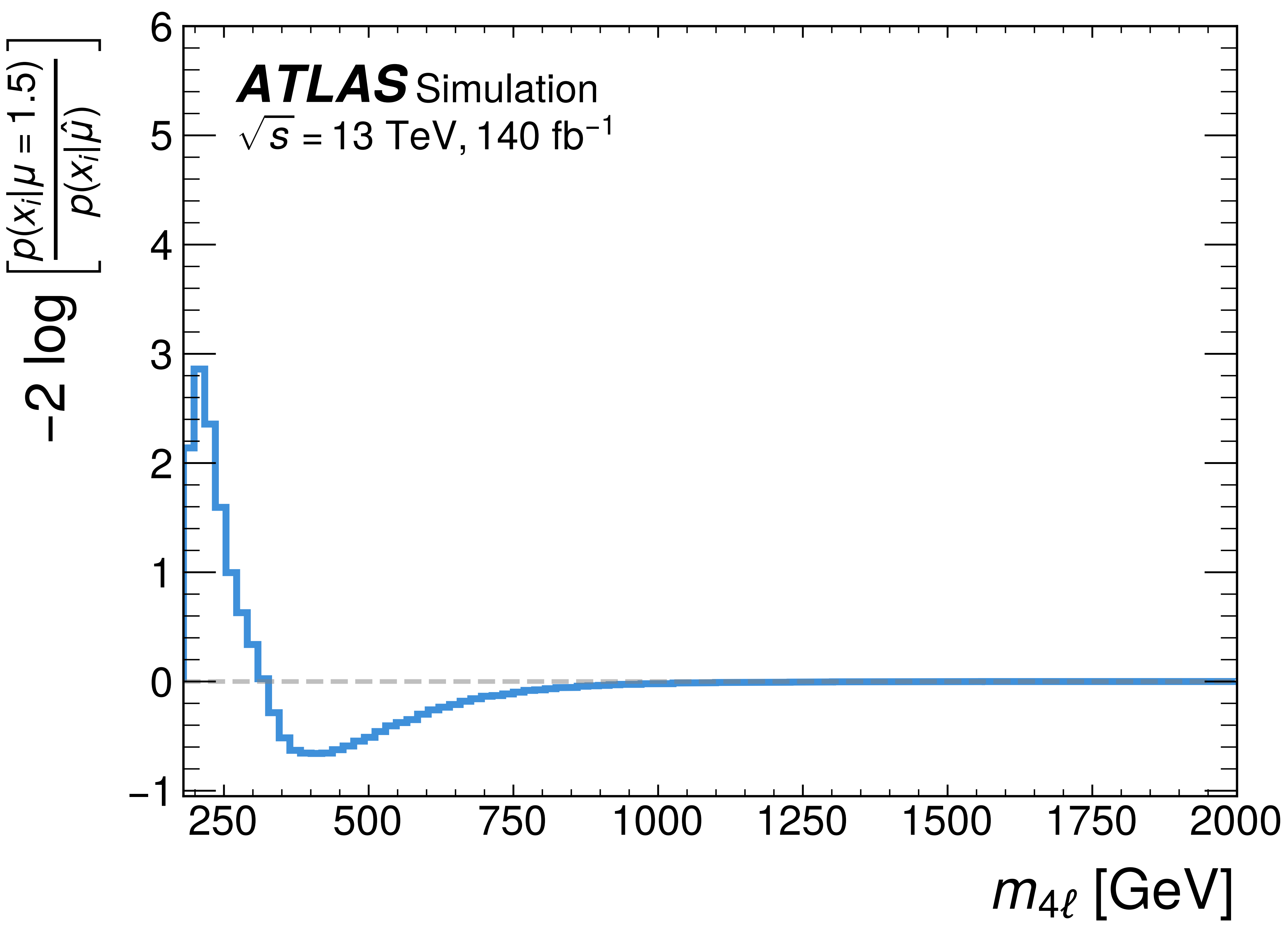
Interpretability: Which phase space favours one hypothesis over another?

$$-2 \cdot log \frac{P(x_i | \mu = 0.5)}{P(x_i | \mu = 1)}$$

$$-2 \cdot log \frac{P(x_i | \mu = 1.5)}{P(x_i | \mu = 1)}$$



Strong destructive interference for $\mu = 0.5$, expect fewer events

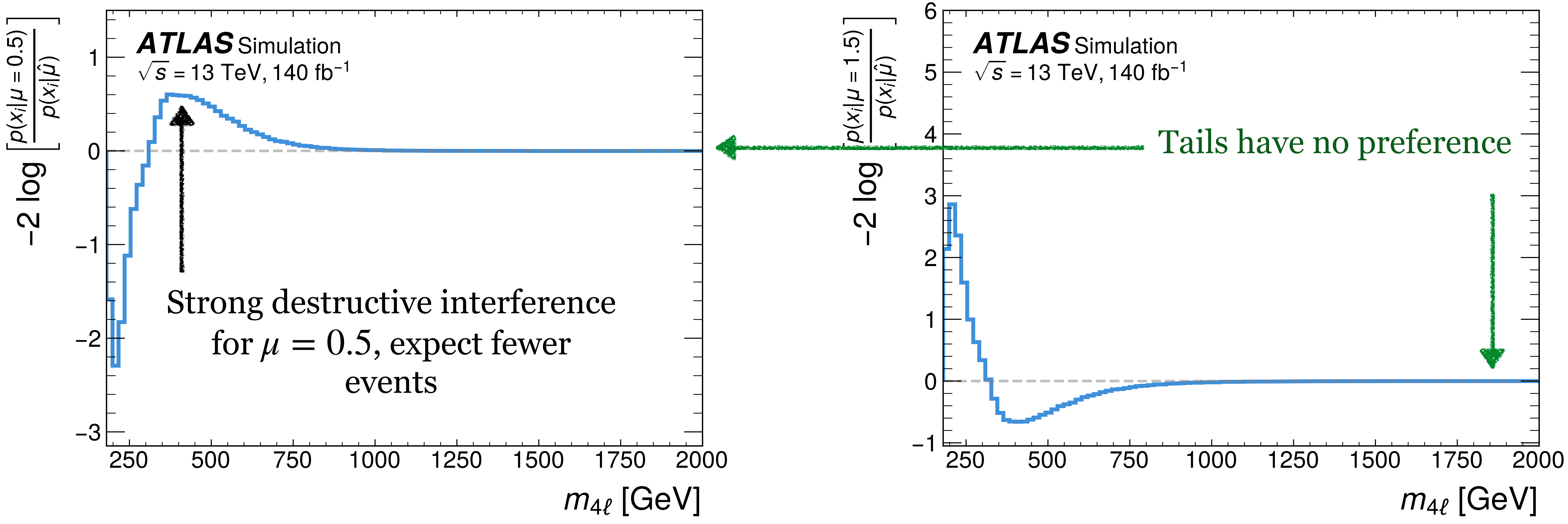


ATLAS methods paper: arXiv:2412.01600

Interpretability:
Which phase space favours one hypothesis over another?

$$-2 \cdot log \frac{P(x_i | \mu = 0.5)}{P(x_i | \mu = 1)}$$

$$-2 \cdot log \frac{P(x_i | \mu = 1.5)}{P(x_i | \mu = 1)}$$



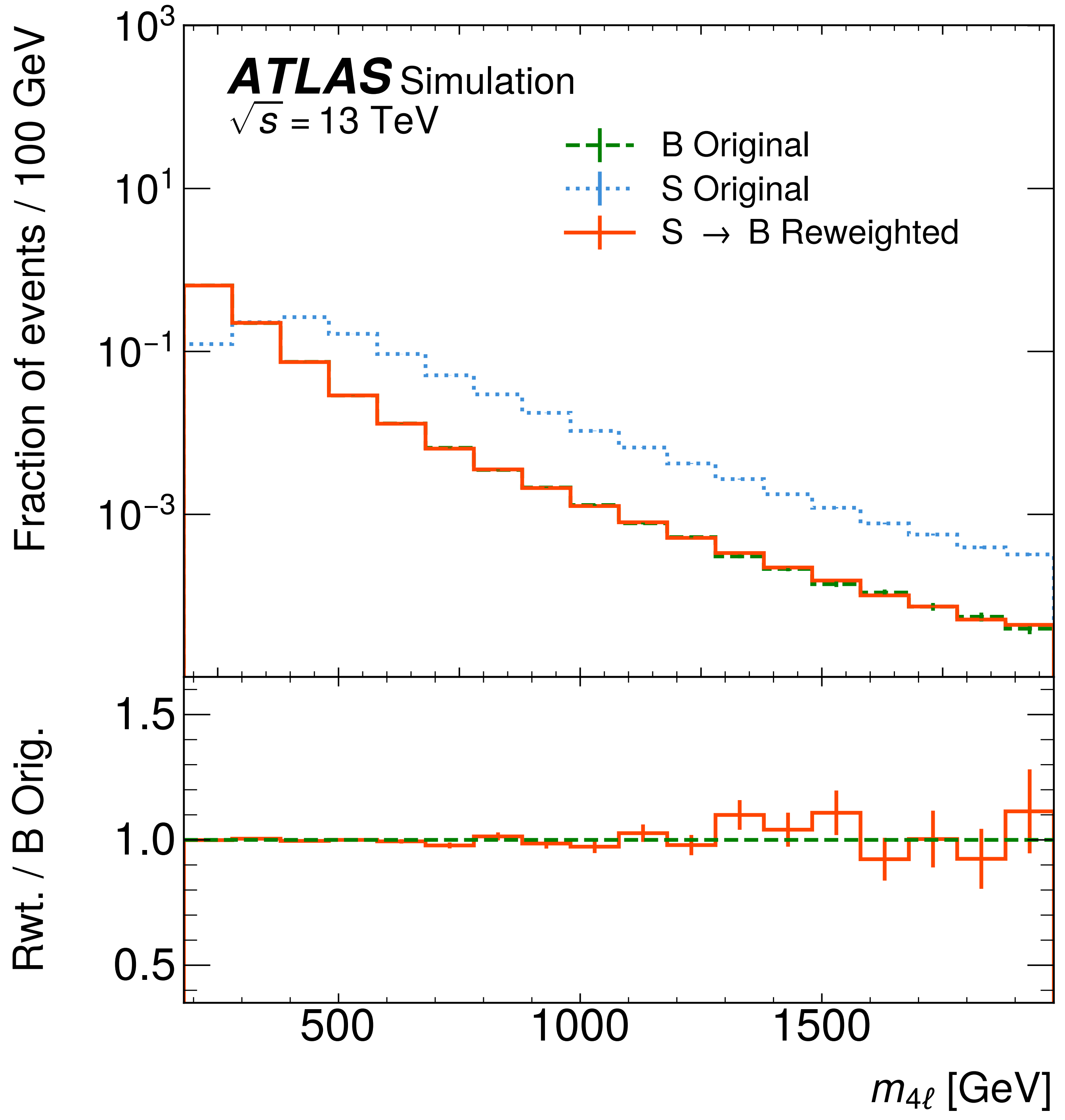
Strong destructive interference for $\mu = 0.5$, expect fewer events

Tails have no preference

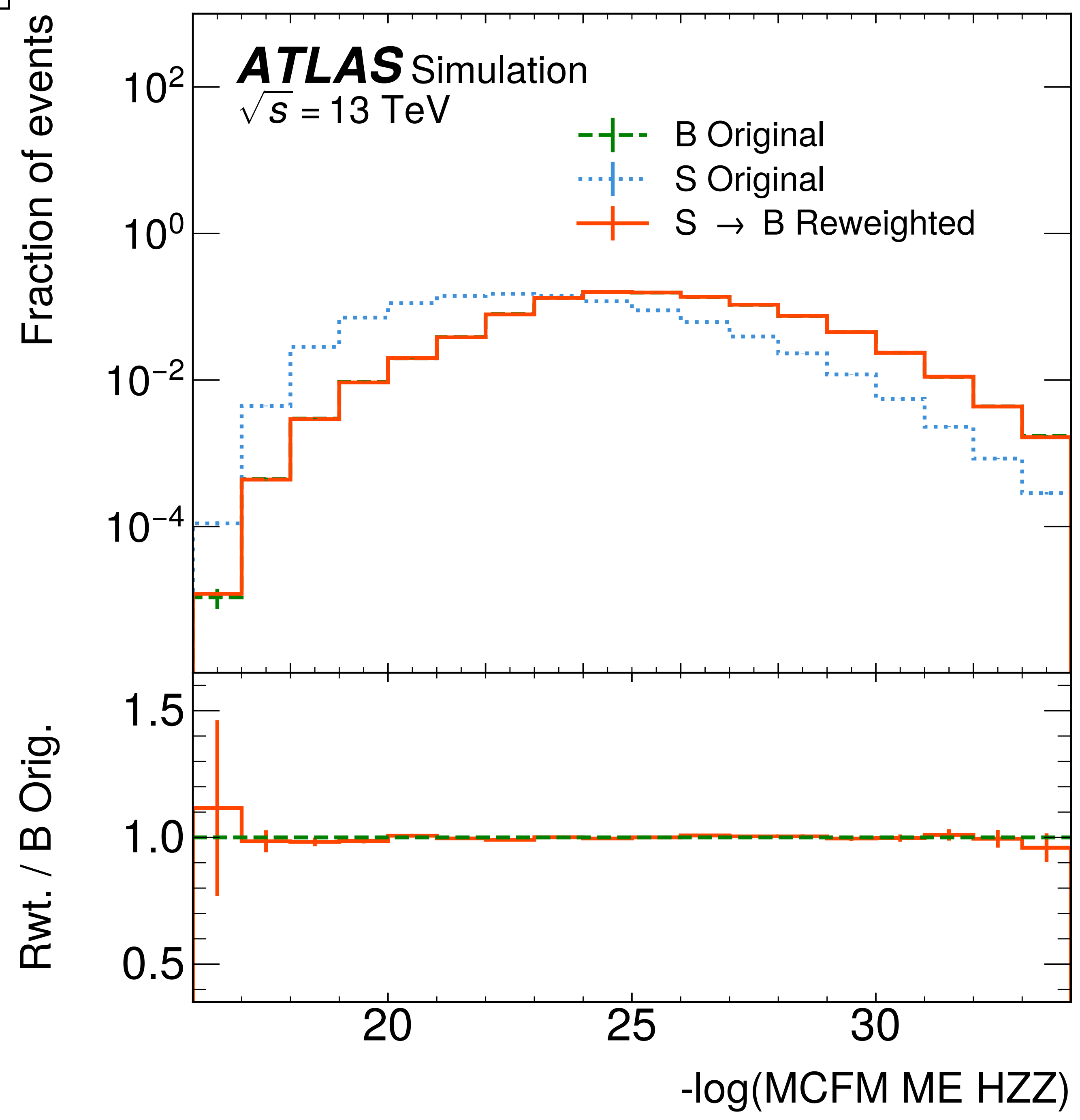
ATLAS methods paper: arXiv:2412.01600

Re-weight closures for B

Source
Target
RW



m_{4l}

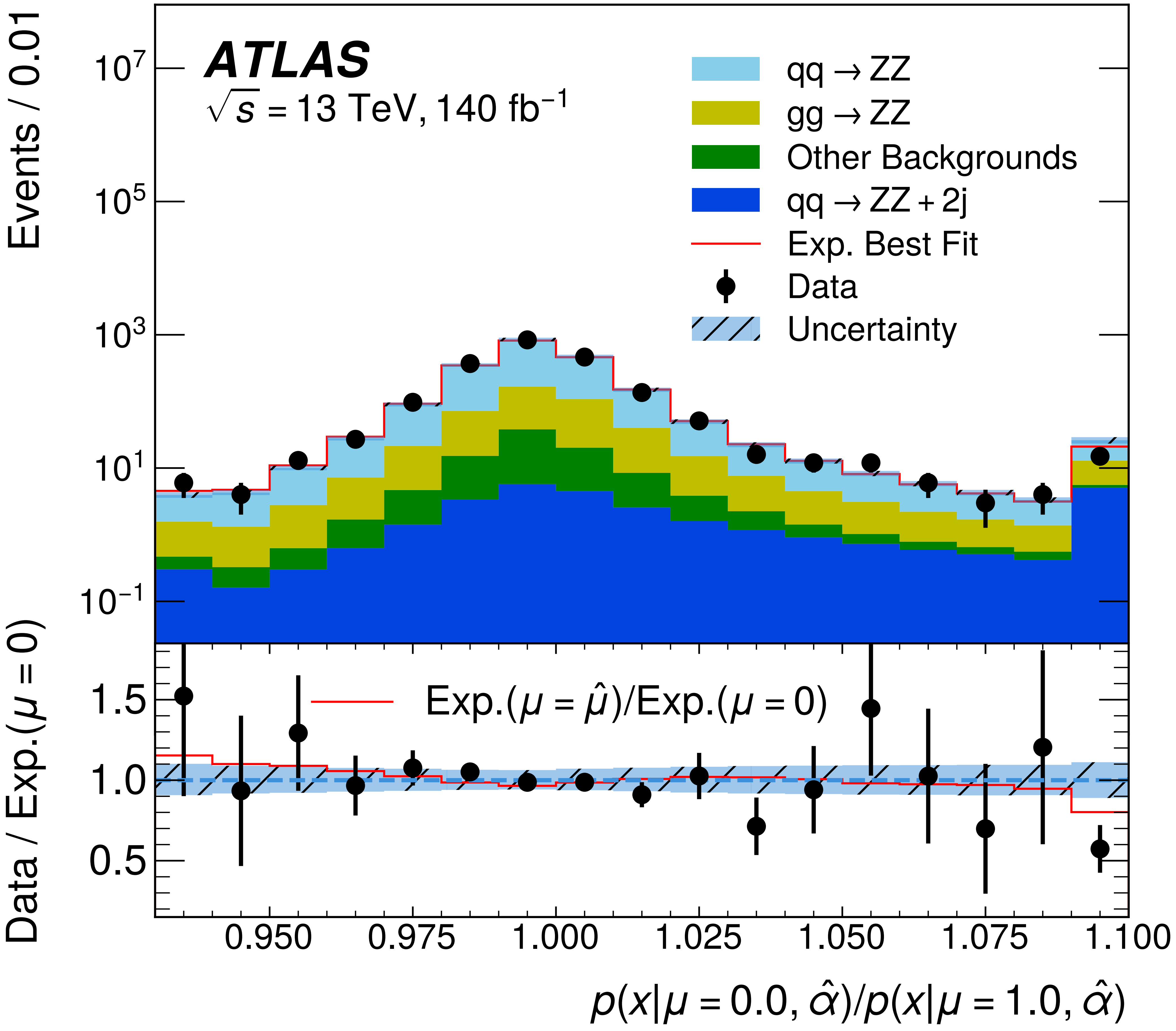


Matrix-Element-based Observable
(ggF from MCFM)

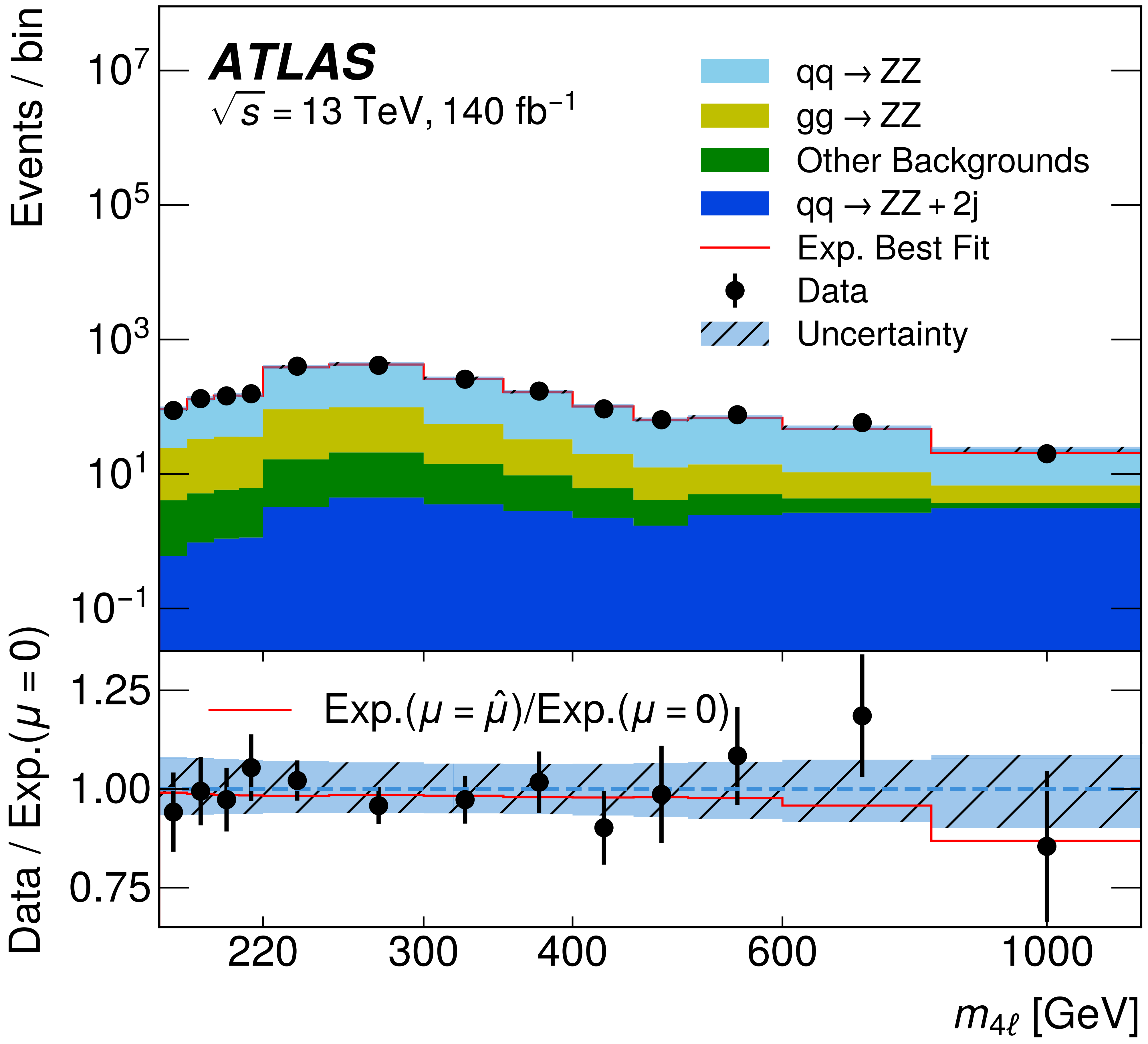
ATLAS Higgs width analysis paper: arXiv:2412.01548

Data-MC validation

NN observable

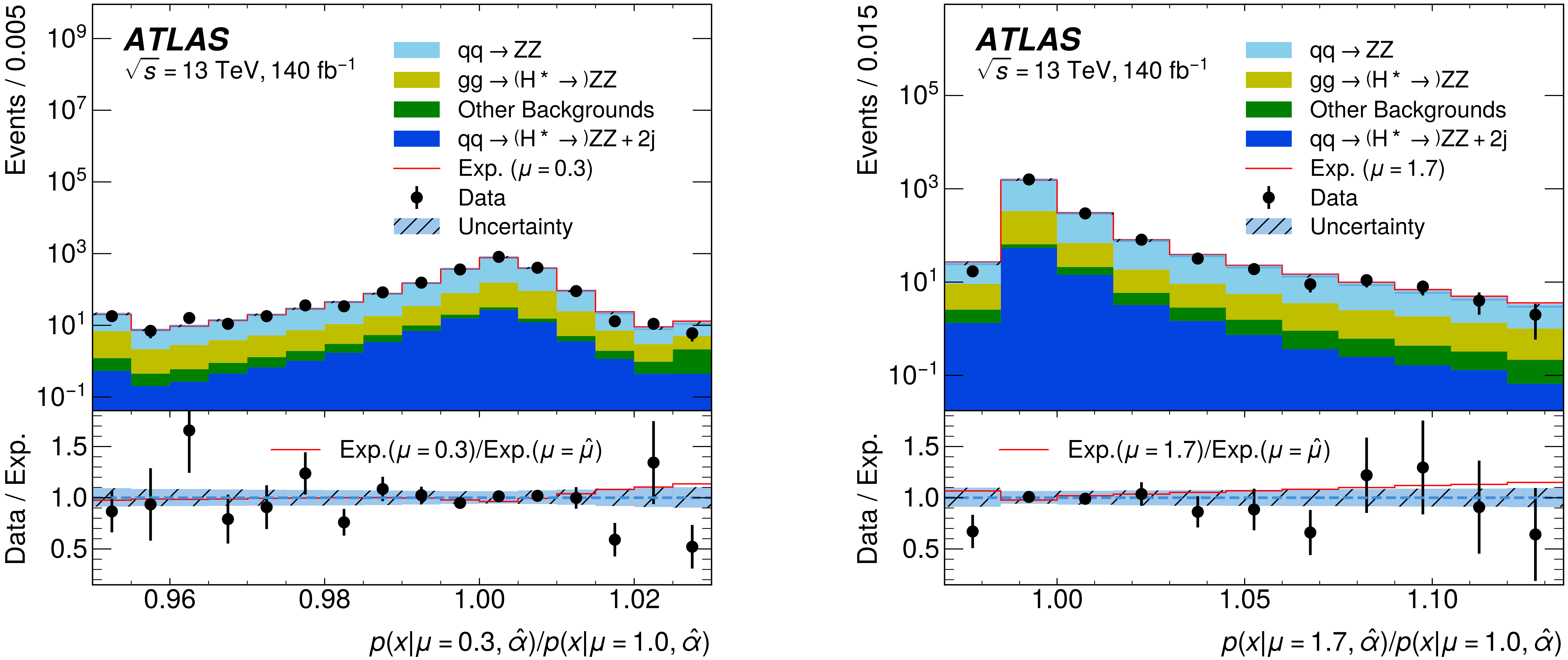


m_{4l}



Data-MC validation

Different NN observables



Physics & statistics motivation

A measurement of the Higgs width via off-shell couplings

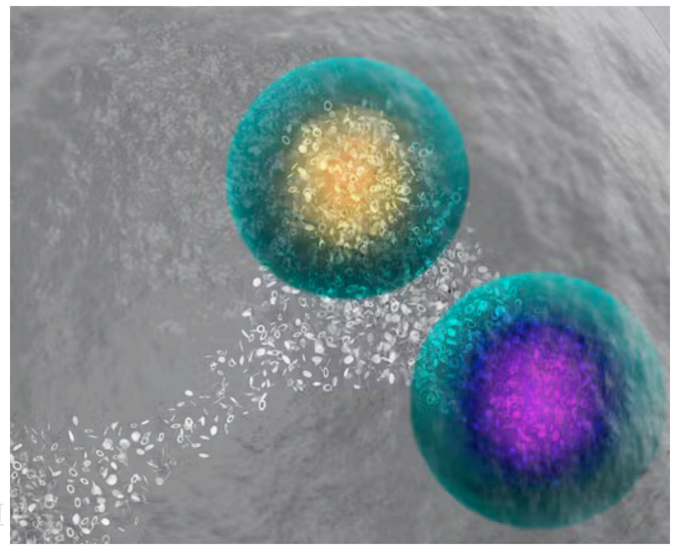
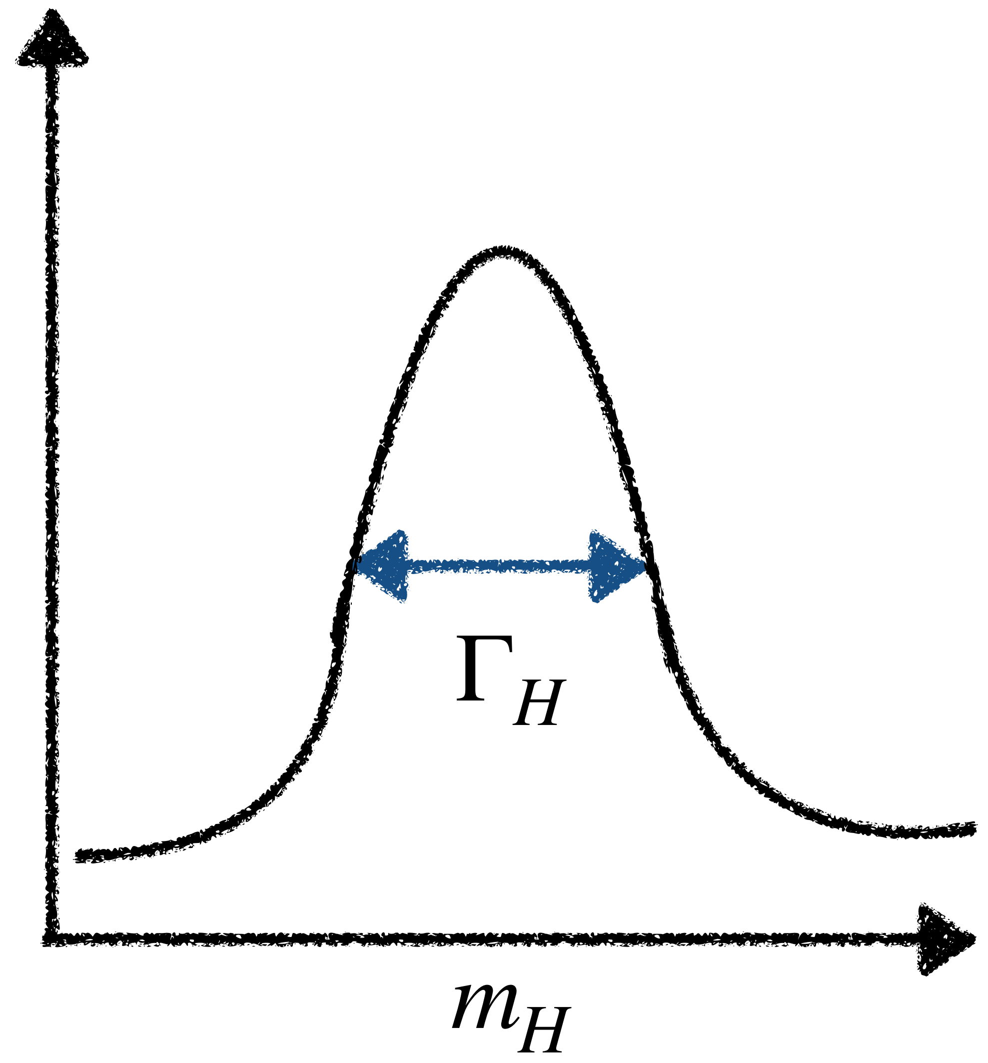


Image: CERN

Undiscovered massive particles

$$\longrightarrow H \longrightarrow \Gamma_H$$

A measurement of the Higgs width via off-shell couplings



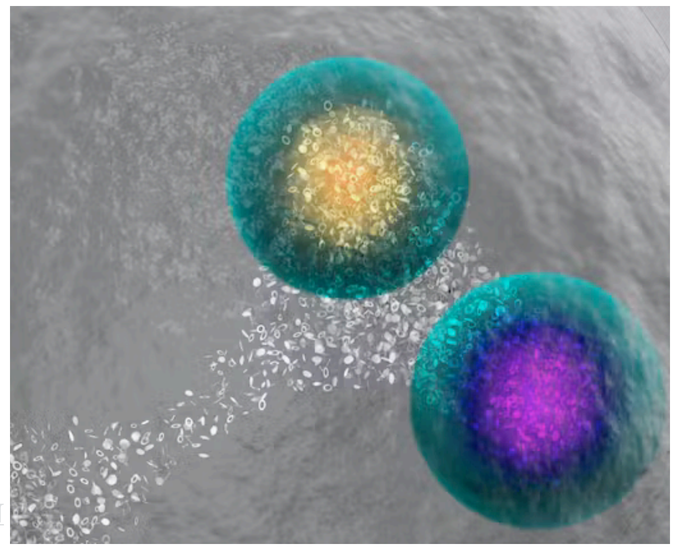
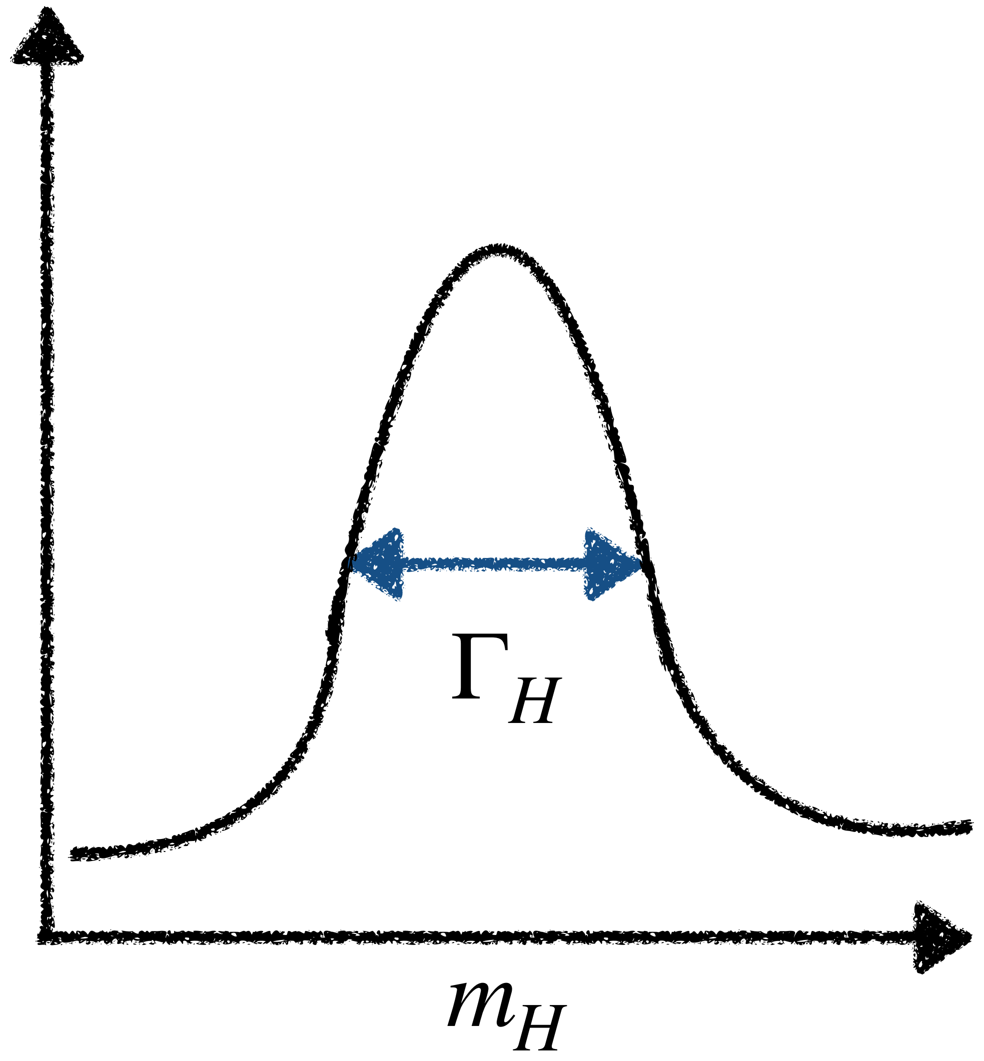


Image: CERN

Undiscovered massive particles → H → Γ_H

A measurement of the Higgs width via off-shell couplings

- Enables the probe of a wide variety of new massive particles, other new physics
- Can't measure directly: SM Higgs width ~4 MeV, resolution of detector ~1 GeV
- Can be probed indirectly by measuring off-shell couplings (and combining with on-shell measurements)



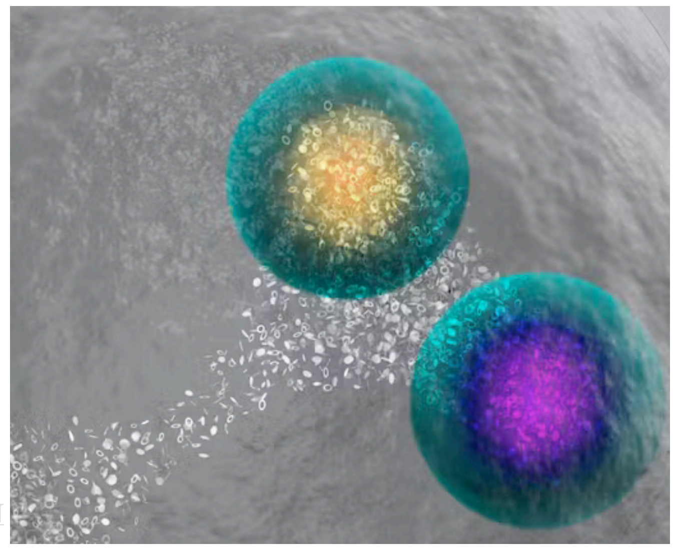


Image: CERN

Undiscovered massive particles $\longrightarrow H \longrightarrow \Gamma_H$

Higgs Width from off-shell Higgs production

arXiv:1405.0285 arXiv:1406.1757

- Off-shell production helps probe Higgs width

$$\boxed{\frac{\mu_{\text{off-shell}}}{\mu_{\text{on-shell}}} = \frac{\Gamma_H}{\Gamma_H^{SM}}.}$$

$$\sigma_{on-shell}^{gg\to H\to ZZ} \sim \frac{g_f^2 g_V^2}{m_H \Gamma_H}$$

$$\frac{d\sigma_{\text{off-shell}}^{gg\to H\to VV}}{dm_{VV}} \propto \frac{g_f^2 g_V^2}{(m_{VV}^2 - m_H^2)^2 + m_H^2 \Gamma_H^2} \simeq \frac{g_f^2 g_V^2}{m_{VV}^4}$$

Higgs Width from off-shell Higgs production

arXiv:1405.0285 arXiv:1406.1757

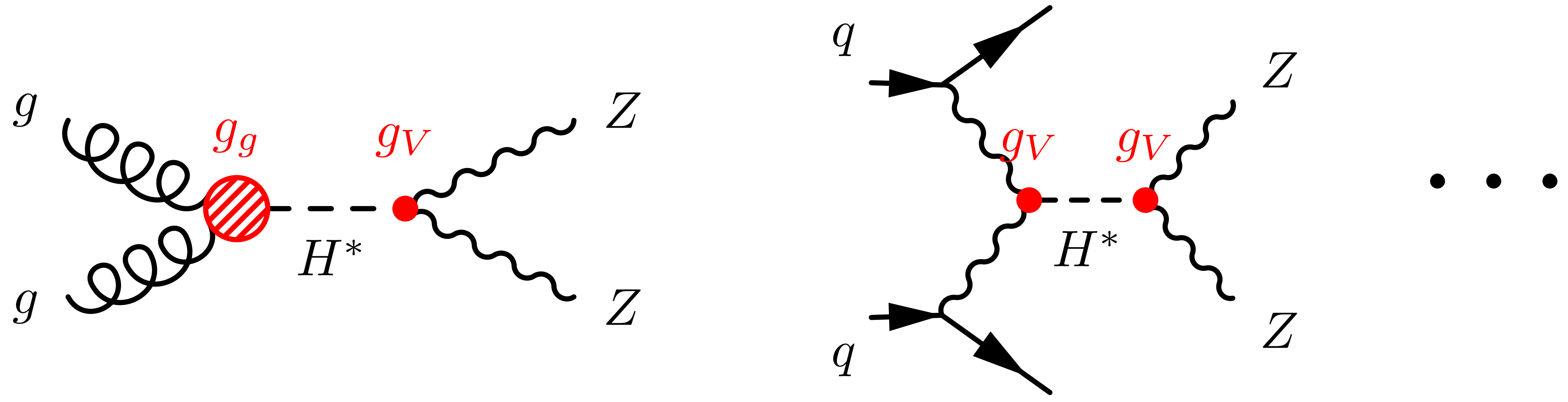
- Off-shell production helps probe Higgs width

$$\boxed{\frac{\mu_{\text{off-shell}}}{\mu_{\text{on-shell}}} = \frac{\Gamma_H}{\Gamma_H^{SM}}.}$$

$$\sigma^{gg\to H\to ZZ}_{on-shell} \sim \frac{g_f^2 g_V^2}{m_H \Gamma_H}$$

$$\frac{d\sigma^{gg\to H\to VV}_{\text{off-shell}}}{dm_{VV}} \propto \frac{g_f^2 g_V^2}{(m_{VV}^2 - m_H^2)^2 + m_H^2\Gamma_H^2} \simeq \frac{g_f^2 g_V^2}{m_{VV}^4}$$

- Interpretation assumes no new physics
- Essential to measure independently in multiple production modes (ggF, VBF) and final states to verify consistent results



NSBI for Higgs width in proof-of-concept study

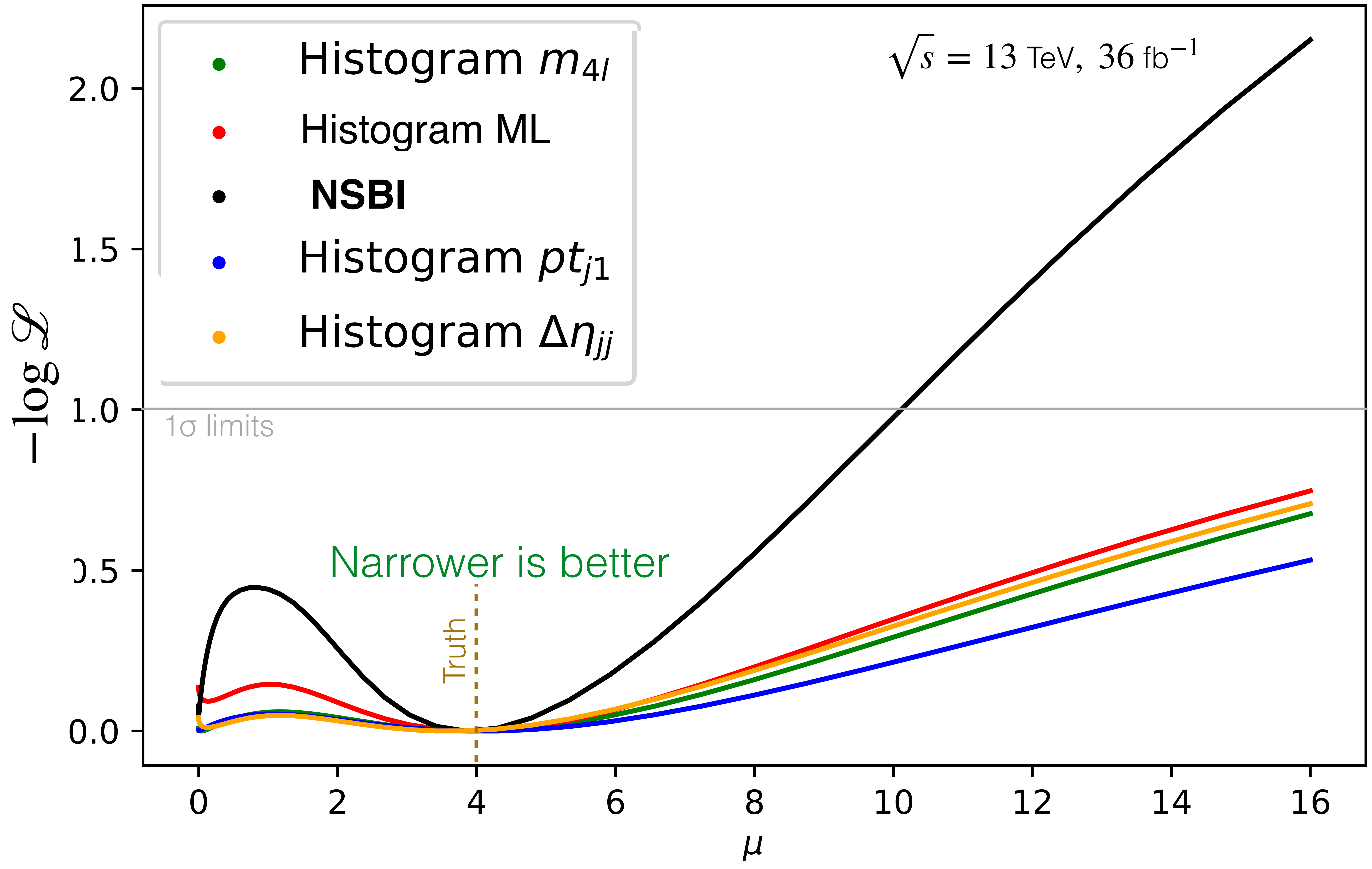
Expected sensitivity

hal-02971995v3 (p172): Ghosh & Rousseau, Thesis: Ghosh

NSBI for Higgs width in proof-of-concept study

Expected sensitivity

hal-02971995v3 (p172): Ghosh & Rousseau, Thesis: Ghosh

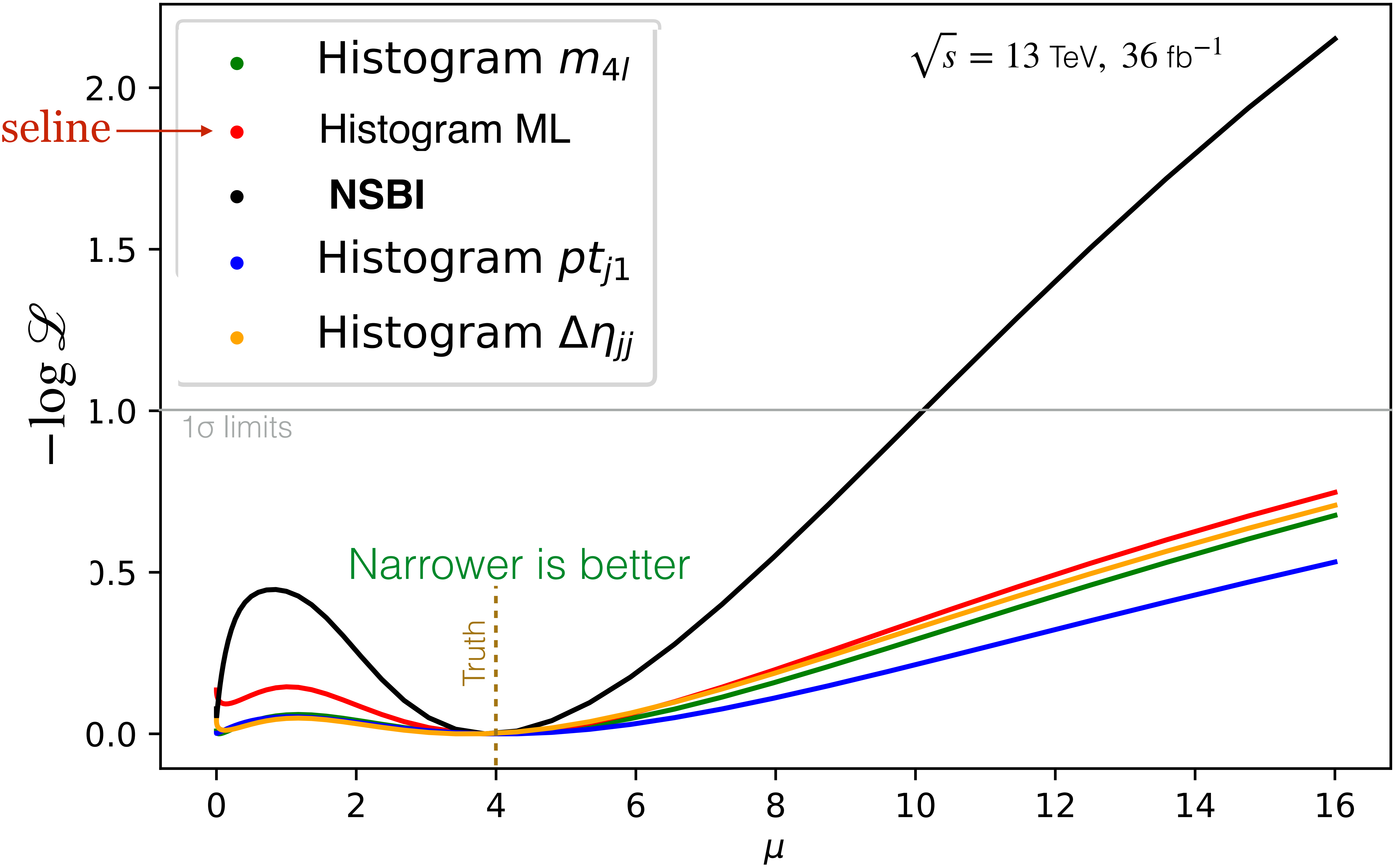


Simulations with $\mu = 4$

NSBI for Higgs width in proof-of-concept study

Expected sensitivity

hal-02971995v3 (p172): Ghosh & Rousseau, Thesis: Ghosh

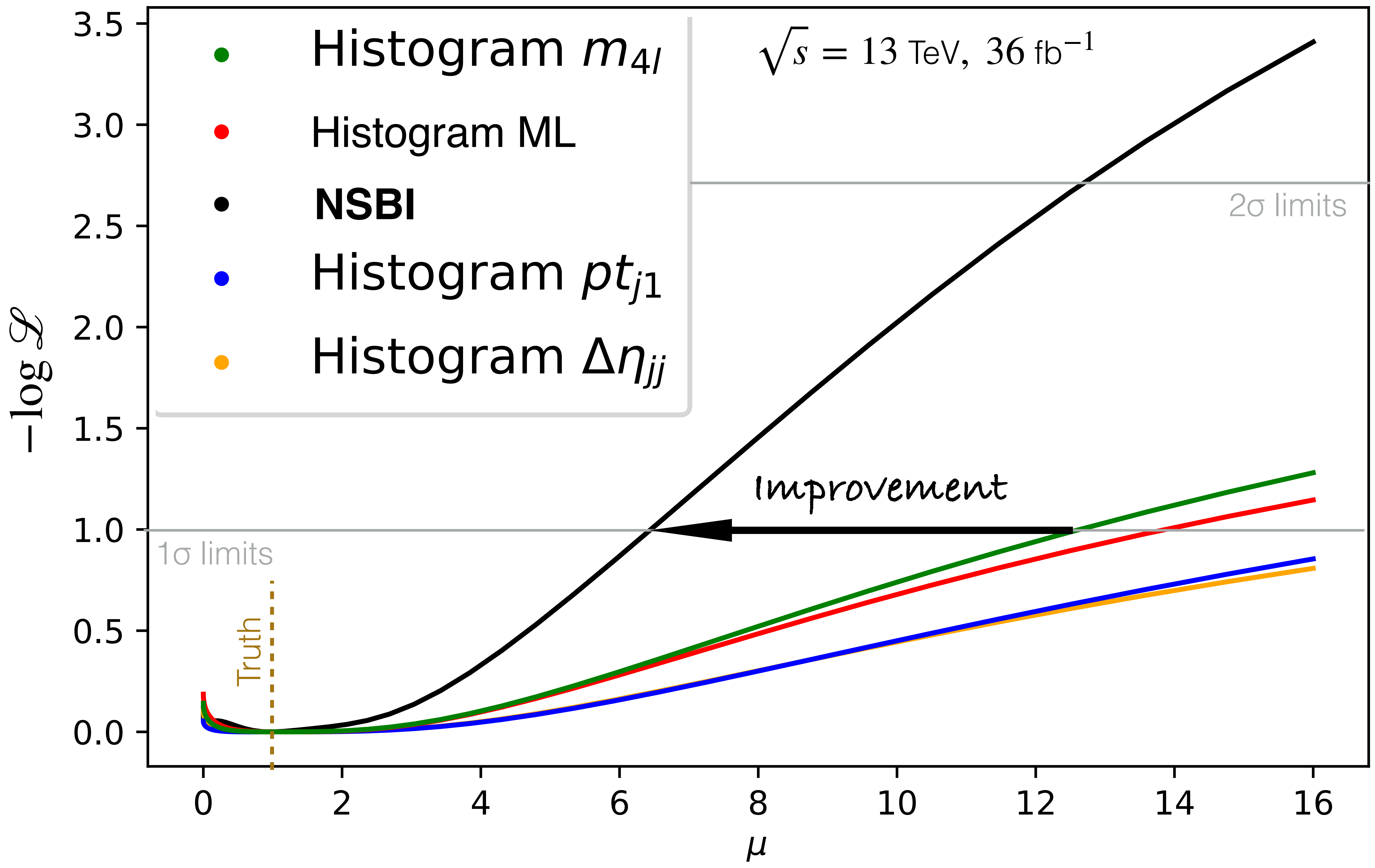


"Traditional ML" baseline

Simulations with $\mu = 4$

Expected improvement for $\mu = 1$ (Standard Model point)

hal-02971995v3 (p172): Ghosh & Rousseau, Thesis: Ghosh



Exciting gains promised!

Simulations with $\mu = 1$

More on the method

Reference Sample

A combination of signal samples, to ensure non-zero probability in entire region of analysis
Does not have to be physical!

$$p_{\text{ref}}(x_i) = \frac{1}{\sum_k \nu_k} \sum_k^{C_{\text{signals}}} \nu_k \cdot p_k(x_i)$$

⇒ In our dataset, $p_{ref}(\,\cdot\,) = p_S(\,\cdot\,)$

Choice of $p_{ref}(\,\cdot\,)$ can be made purely on numerical stability of training, as it drops out in profile step

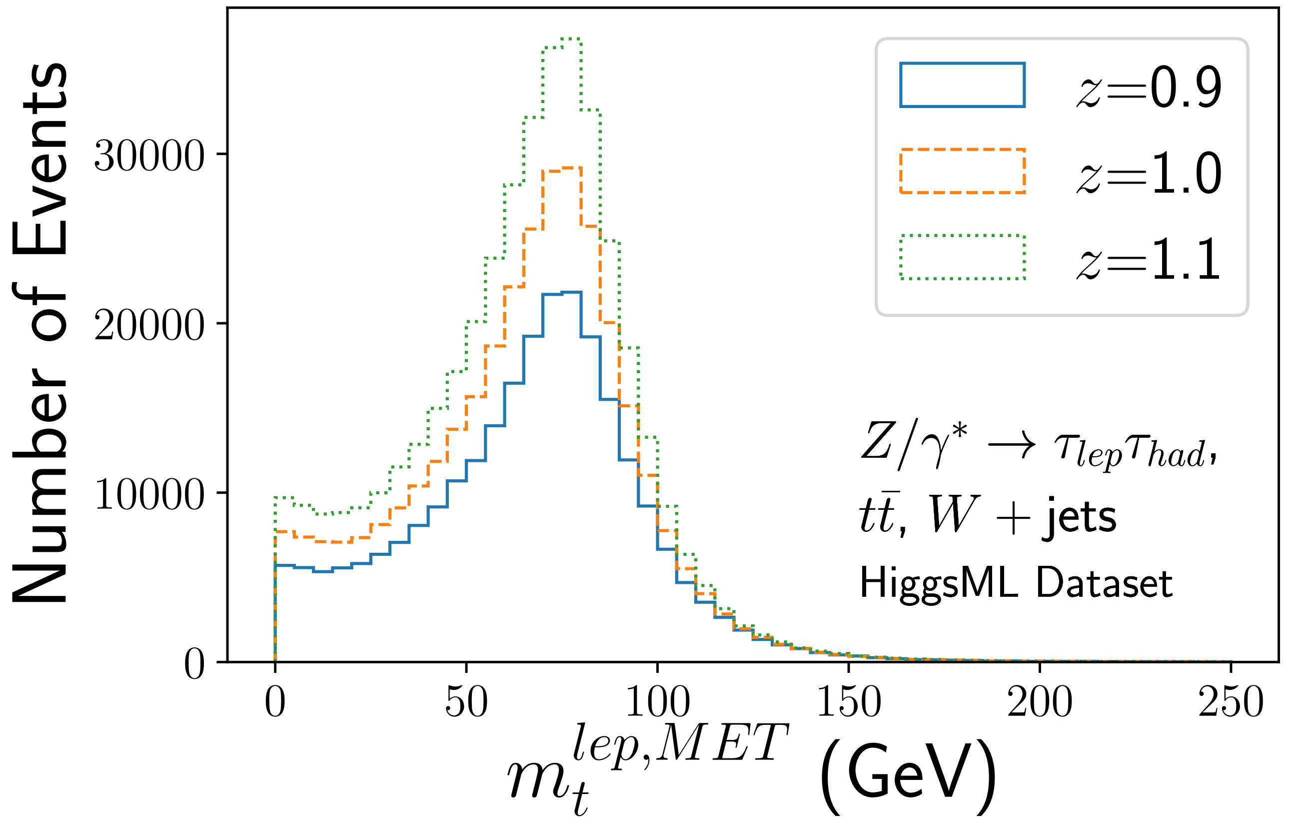
$$t_\mu = -2\ln\left(\frac{L_{\text{full}}(\mu,\widehat{\widehat{\alpha}})/\cancel{L_{\text{ref}}}}{L_{\text{full}}(\widehat{\mu},\widehat{\alpha})/\cancel{L_{\text{ref}}}}\right)$$

MC stat uncertainty? See backup

Systematic uncertainties

Experimental uncertainties:

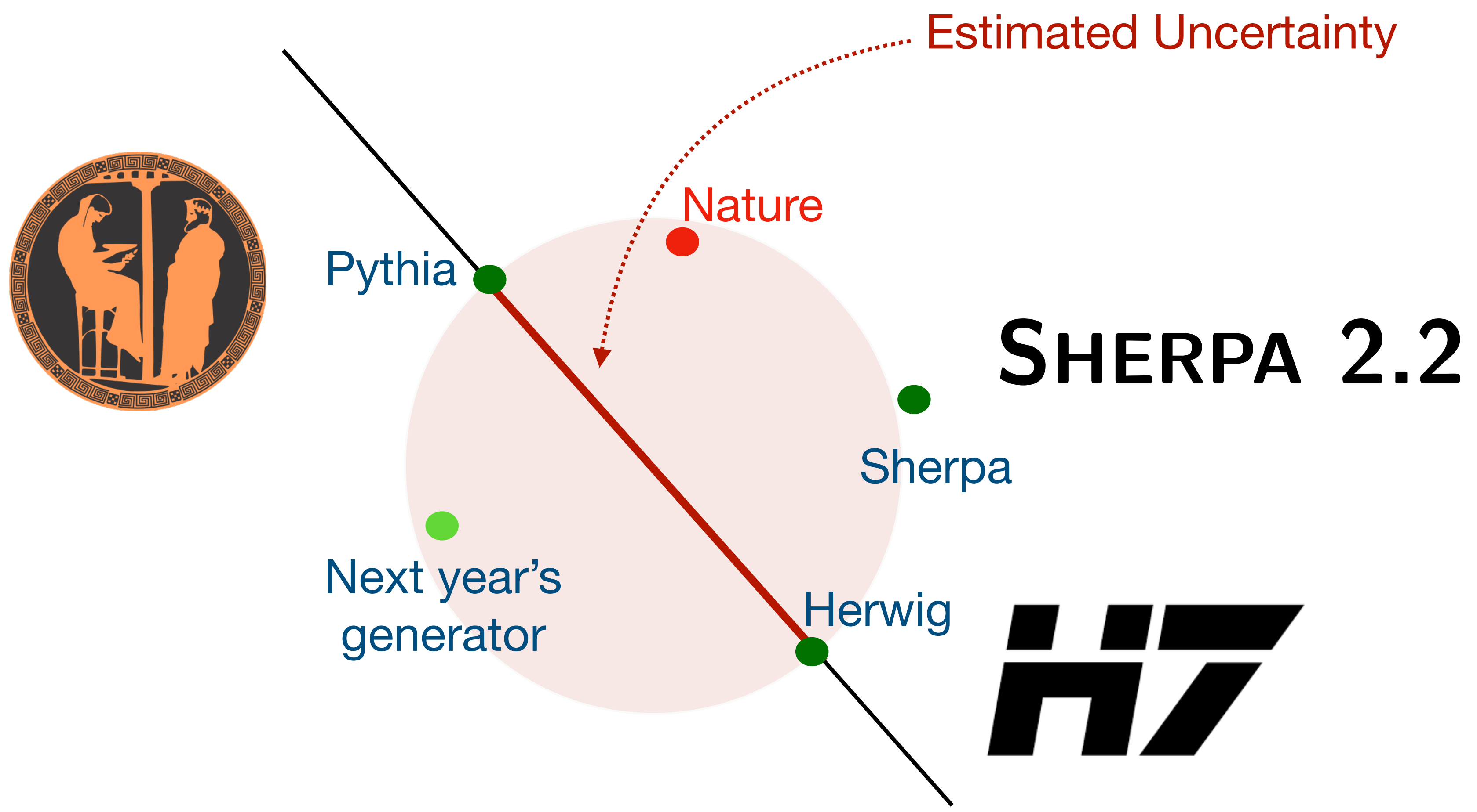
Eg. Inaccuracies in the calibration of our detector



Ghosh et al: PhysRevD.104.056026

Theory uncertainties:

Eg. Inability to compute QFT to infinite order



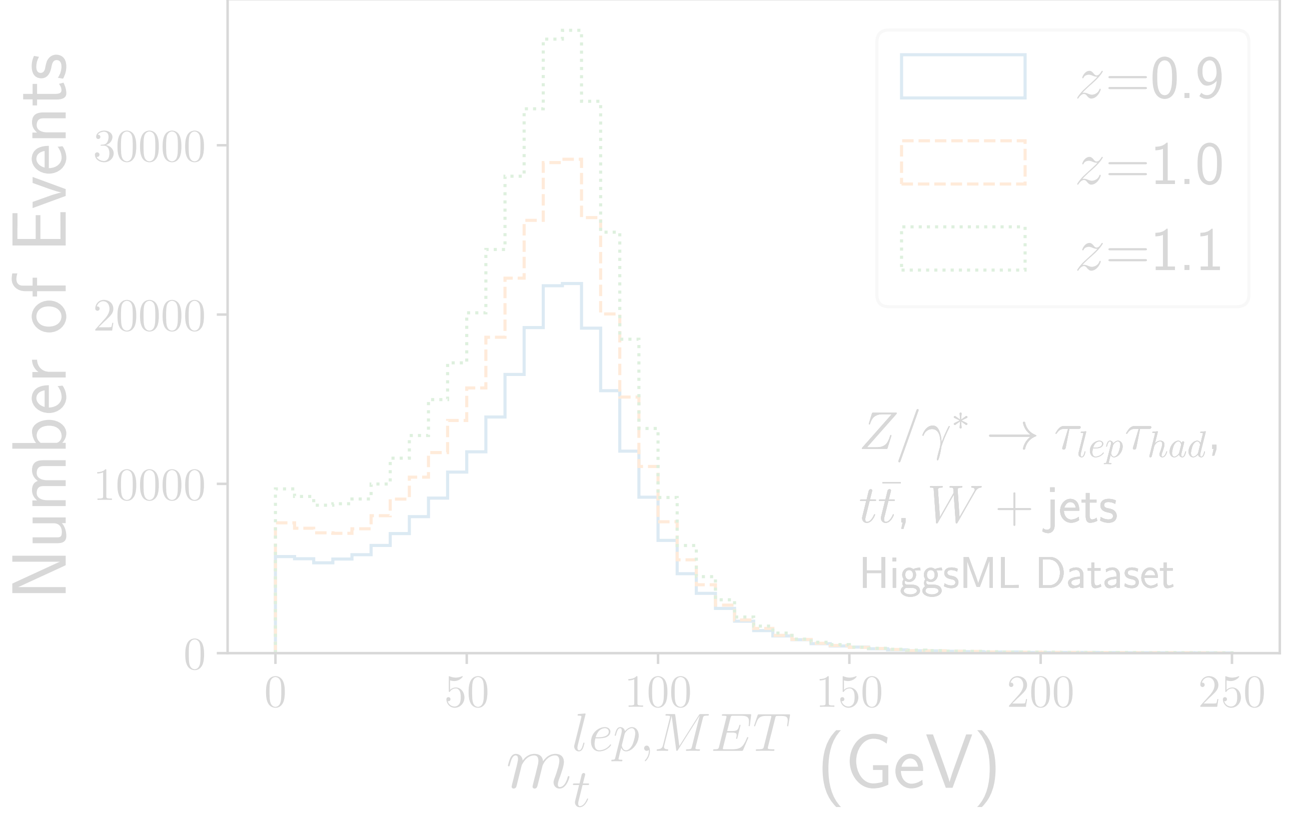
Ghosh and Nachman: epjc.s10052.022.10012.w

MC stat uncertainty? See backup

Systematic uncertainties

Experimental uncertainties:

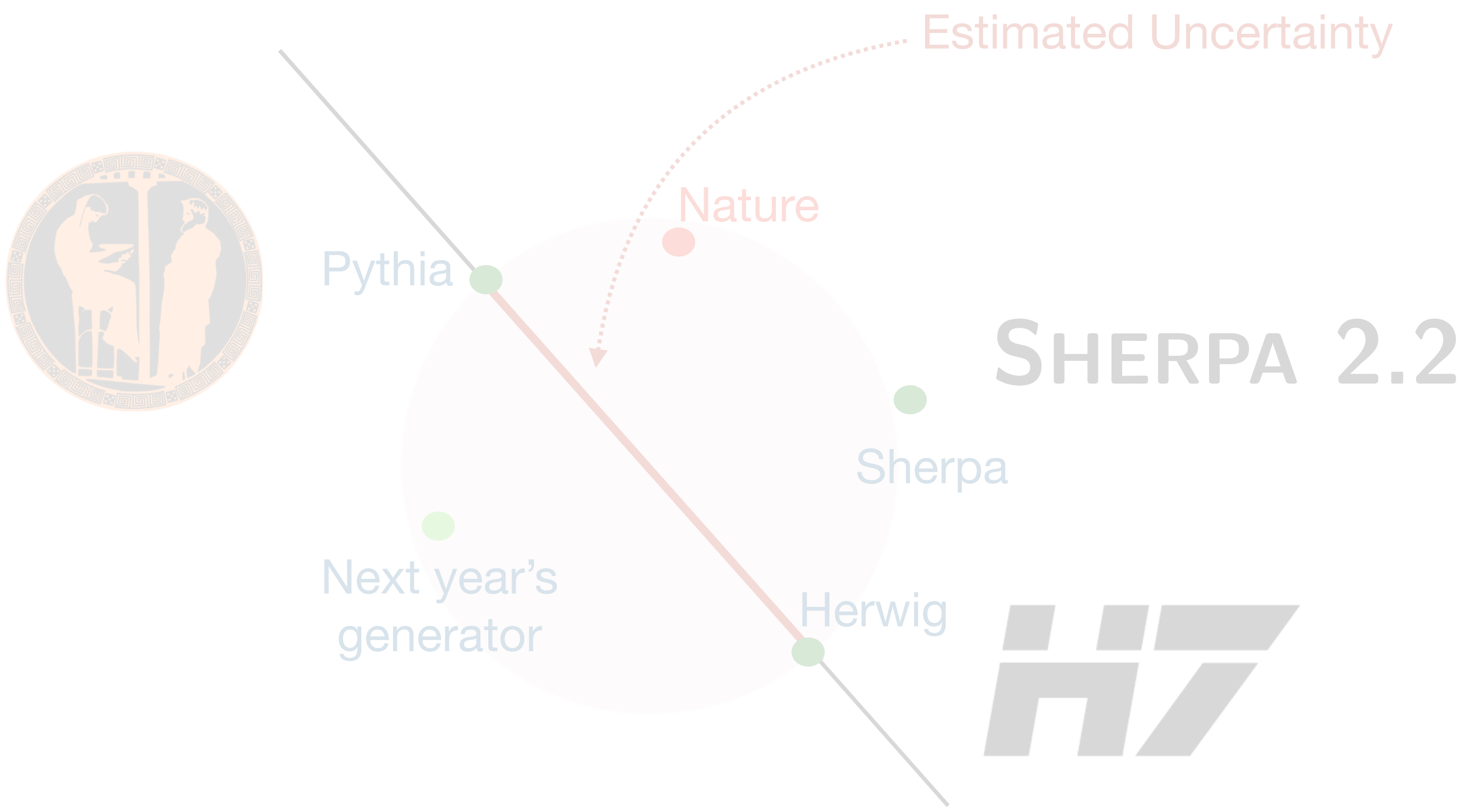
Eg. Inaccuracies in the calibration of our detector



Ghosh et al: PhysRevD.104.056026

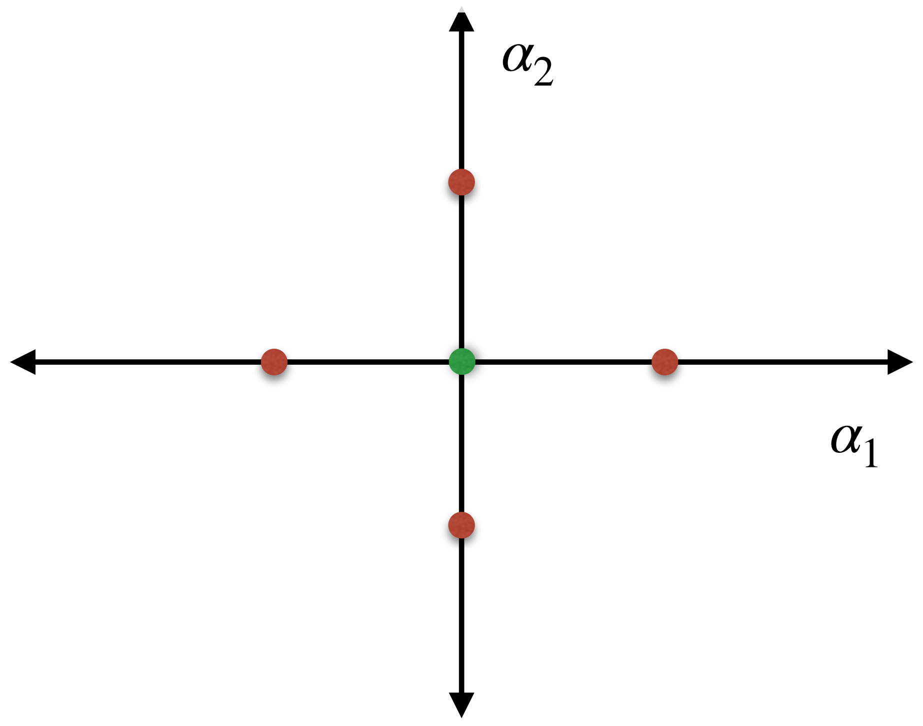
Theory uncertainties:

Eg. Inability to compute QFT to infinite order



Ghosh and Nachman: epjc.s10052.022.10012.w

- We only have simulations at 3 variations of each nuisance parameter α_k



Known interpolation strategies

See formula used in backup

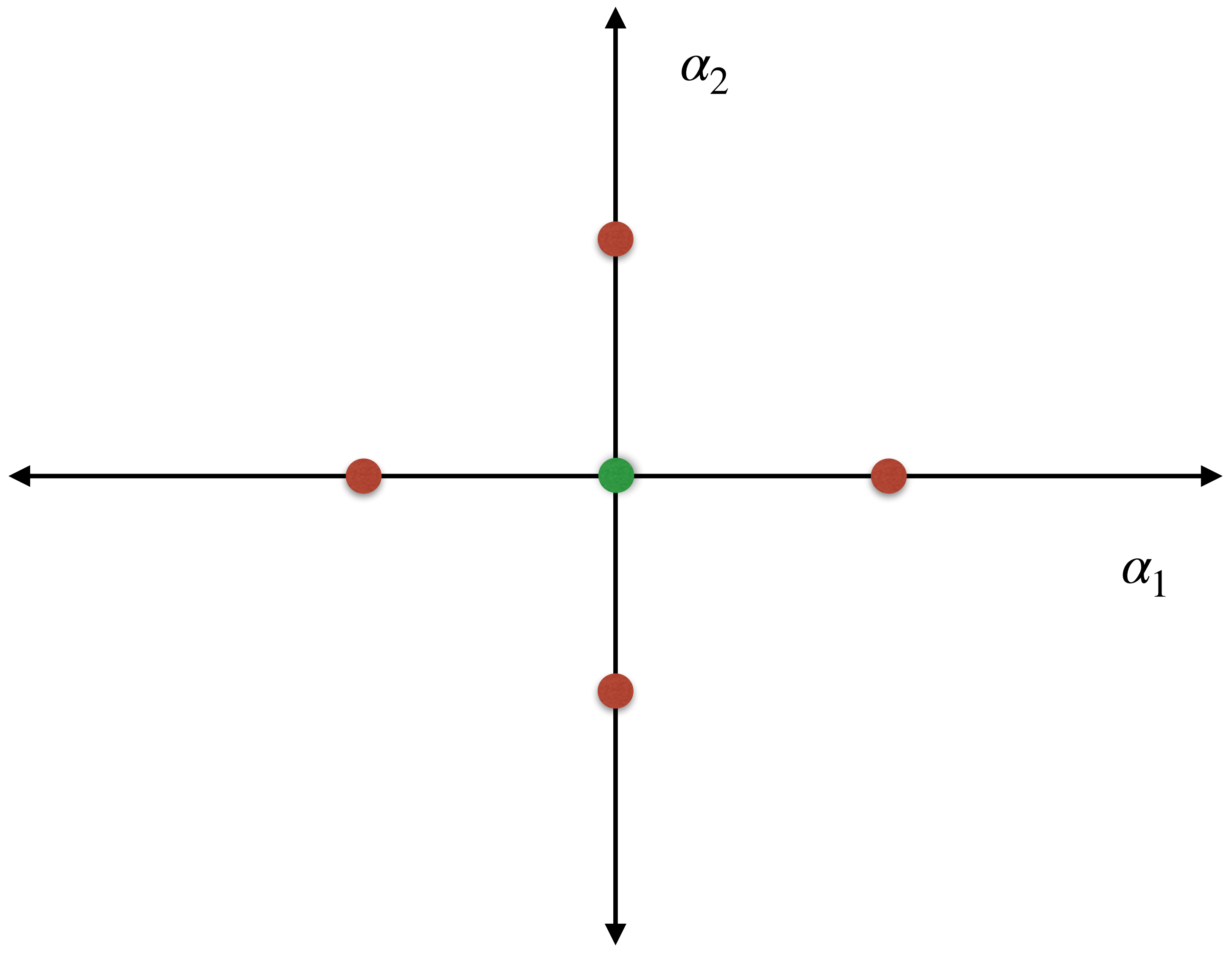
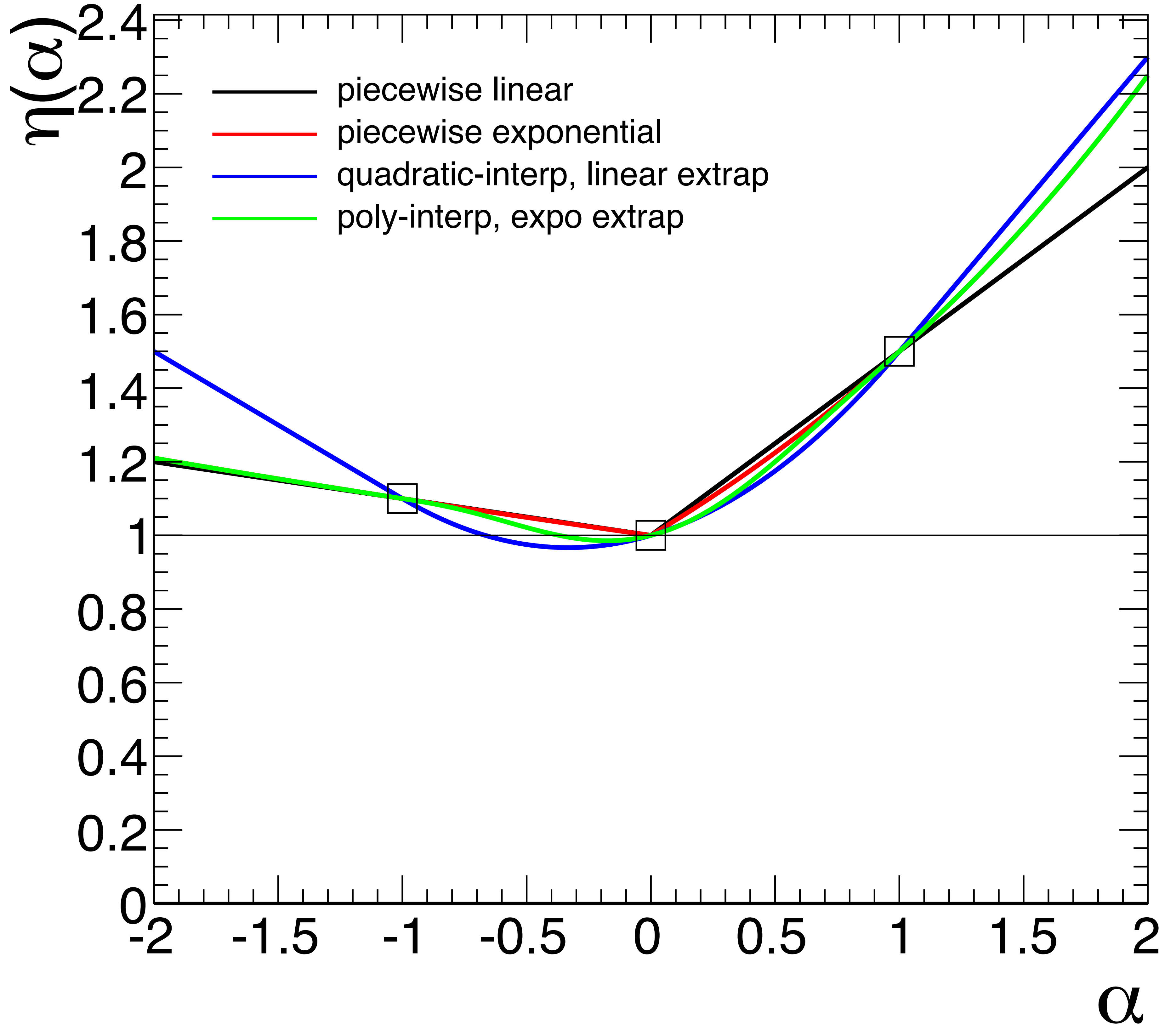


Image: arXiv:1503.07622

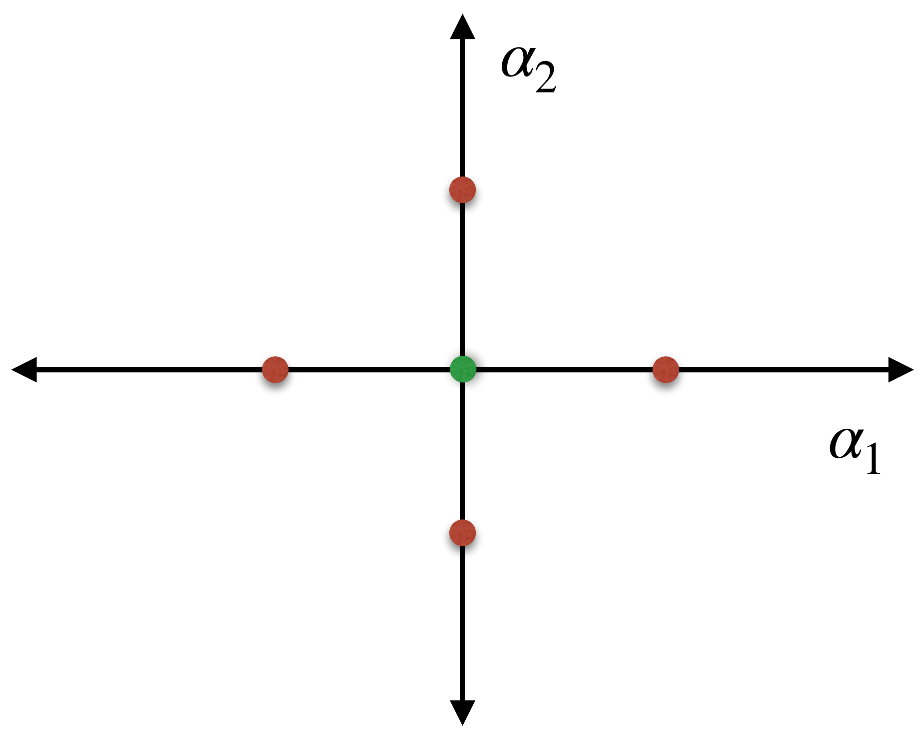


⇒ Combine these traditional interpolation with neural network estimation of per-event likelihood ratios

Probability density ratio including nuisance parameters (α)

x_i is one individual event

$$\frac{p(x_i \,|\, \mu, \underline{\alpha})}{p_{ref}(x_i)} =$$

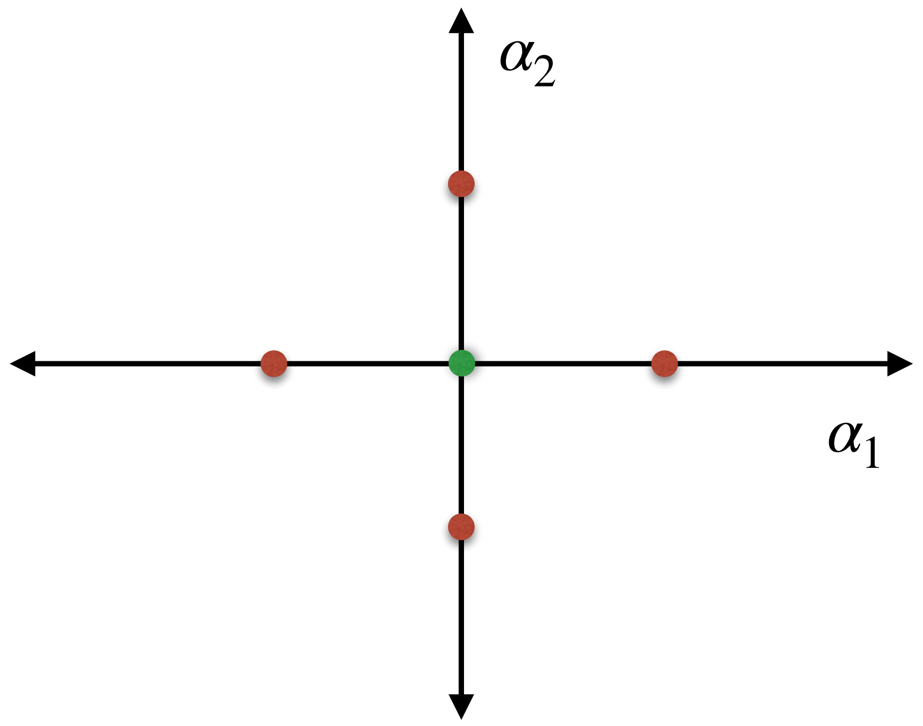


See details of vertical interpolation for $G_j(\alpha_k)$, $g_j(x_i, \alpha_k)$

Probability density ratio including nuisance parameters (α)

x_i is one individual event

$$\frac{p(x_i \mid \mu, \alpha)}{p_{ref}(x_i)} = \frac{1}{\nu(\mu, \alpha)} \sum_{j}^{C} f_j(\mu) \cdot \nu_j \cdot \frac{p_j(x_i)}{p_{ref}(x_i)} \cdot \prod_{k}^{N_{syst}} G_j(\alpha_k) \cdot g_j(x_i, \alpha_k)$$



$$g_j(x_i, \alpha_k) = \frac{p_j(x_i, \alpha_k)}{p_j(x_i)}$$

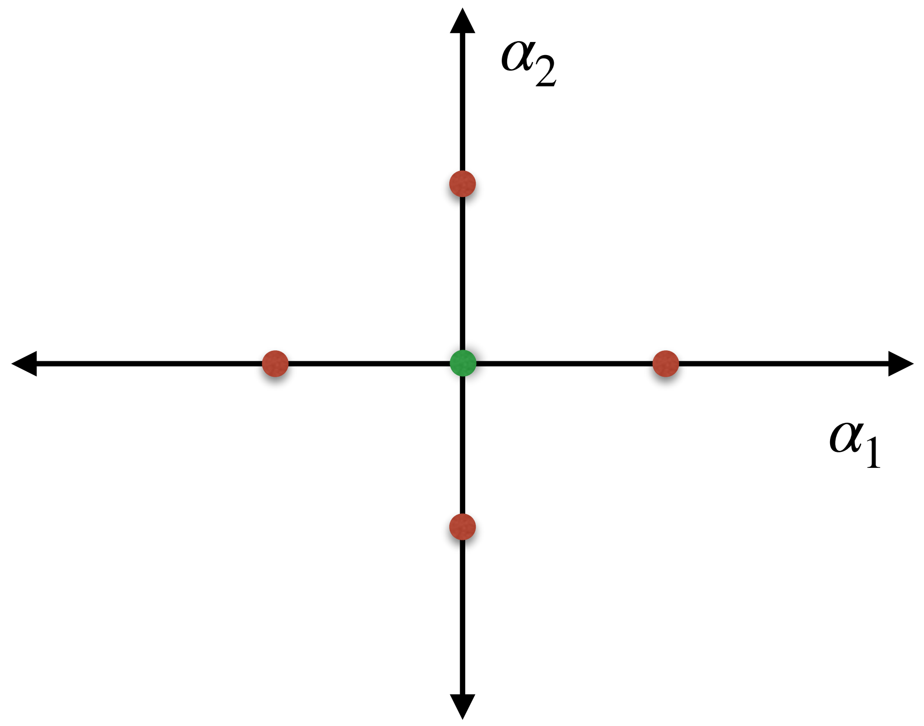
See details of vertical interpolation for $G_j(\alpha_k)$, $g_j(x_i, \alpha_k)$

Probability density ratio including nuisance parameters (α)

x_i is one individual event

$$\frac{p(x_i \,|\, \mu, \underline{\alpha})}{p_{ref}(x_i)} = \frac{1}{\nu(\mu, \alpha)} \sum_{j}^{C} f_j(\mu) \cdot \nu_j \cdot \frac{p_j(x_i)}{p_{ref}(x_i)} \cdot \prod_{k}^{N_{syst}} G_j(\alpha_k) \cdot g_j(x_i, \alpha_k)$$

We have this already



$$g_j(x_i, \alpha_k) = \frac{p_j(x_i, \alpha_k)}{p_j(x_i)}$$

See details of vertical interpolation for $G_j(\alpha_k)$, $g_j(x_i, \alpha_k)$

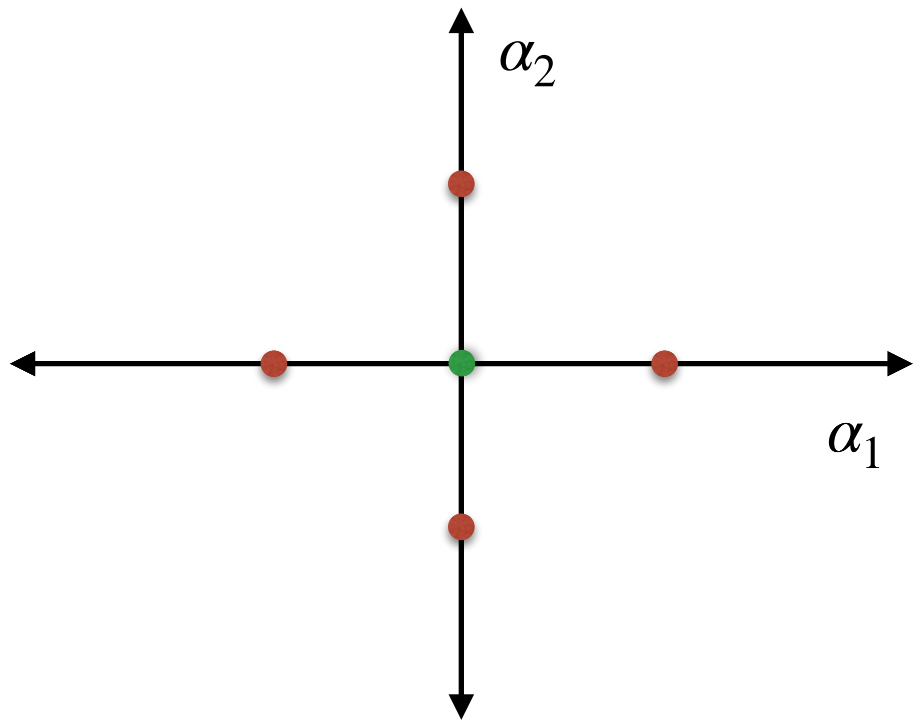
Probability density ratio including nuisance parameters (α)

x_i is one individual event

$$\frac{p(x_i\,|\,\mu,\alpha)}{p_{ref}(x_i)} = \frac{1}{\nu(\mu,\alpha)} \sum_{j}^{C} f_j(\mu)\cdot\nu_j\cdot\frac{p_j(x_i)}{p_{ref}(x_i)}\cdot\prod_{k}^{N_{syst}} G_j(\alpha_k)\cdot g_j(x_i,\alpha_k)$$

We have this already

Estimate from simulations and existing interpolation methods



$$g_j(x_i,\alpha_k) = \frac{p_j(x_i,\alpha_k)}{p_j(x_i)}$$

See details of vertical interpolation for $G_j(\alpha_k)$, $g_j(x_i,\alpha_k)$

Probability density ratio including nuisance parameters (α)

x_i is one individual event

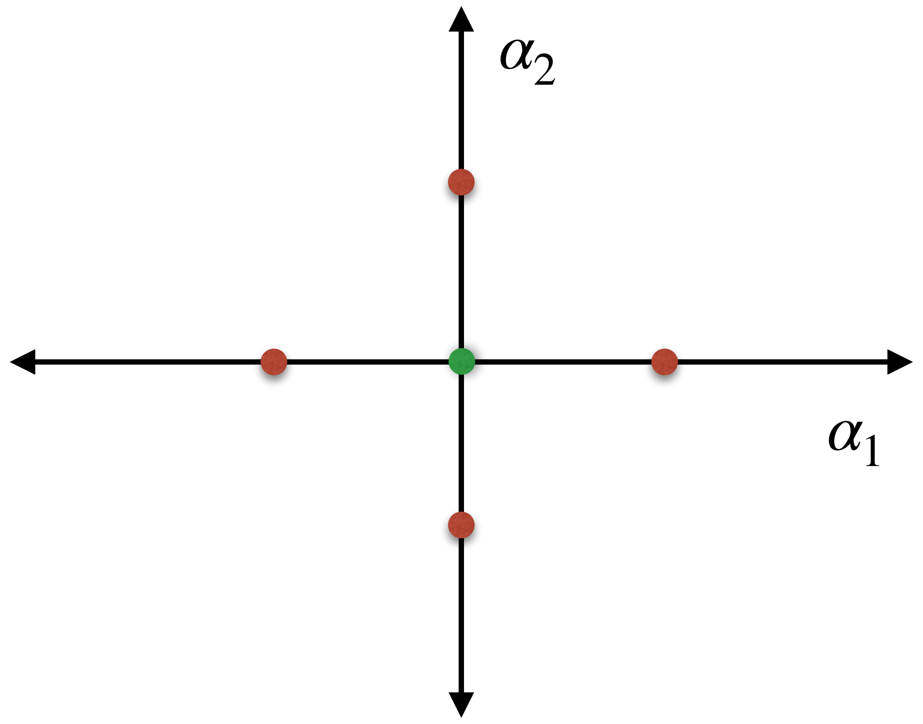
$$\frac{p(x_i\,|\,\mu,\alpha)}{p_{ref}(x_i)} = \frac{1}{\nu(\mu,\alpha)} \sum_{j}^{C} f_j(\mu)\cdot\nu_j\cdot\frac{p_j(x_i)}{p_{ref}(x_i)} \cdot \prod_{k}^{N_{syst}} G_j(\alpha_k)\cdot g_j(x_i,\alpha_k)$$

We have this already

Per-event terms estimated using another ensemble of networks and interpolation methods

$$g_j(x_i,\alpha_k) = \frac{p_j(x_i,\alpha_k)}{p_j(x_i)}$$

Estimate from simulations and existing interpolation methods



See details of vertical interpolation for $G_j(\alpha_k)$, $g_j(x_i,\alpha_k)$

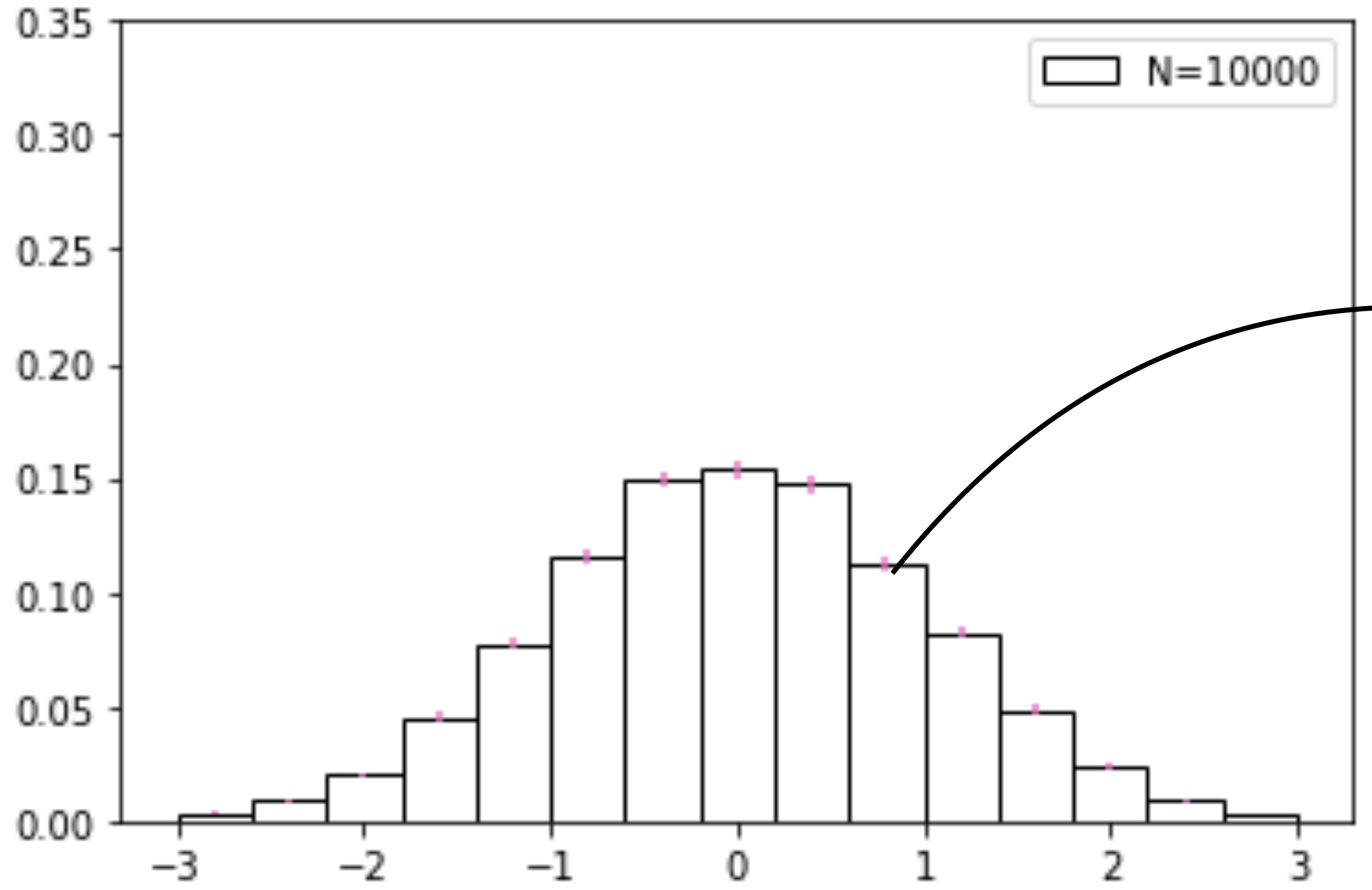
Open problems to extend to full ATLAS analysis:

- ✓ Robustness: Design and validation
- ✓ Systematic Uncertainties: Incorporate them in likelihood (ratio) model
- ‣ Neyman Construction: Throwing toys in high dimensions

Generating event-level pseudo-experiments

Need to generate random possible datasets we could collect at the LHC

Traditionally:

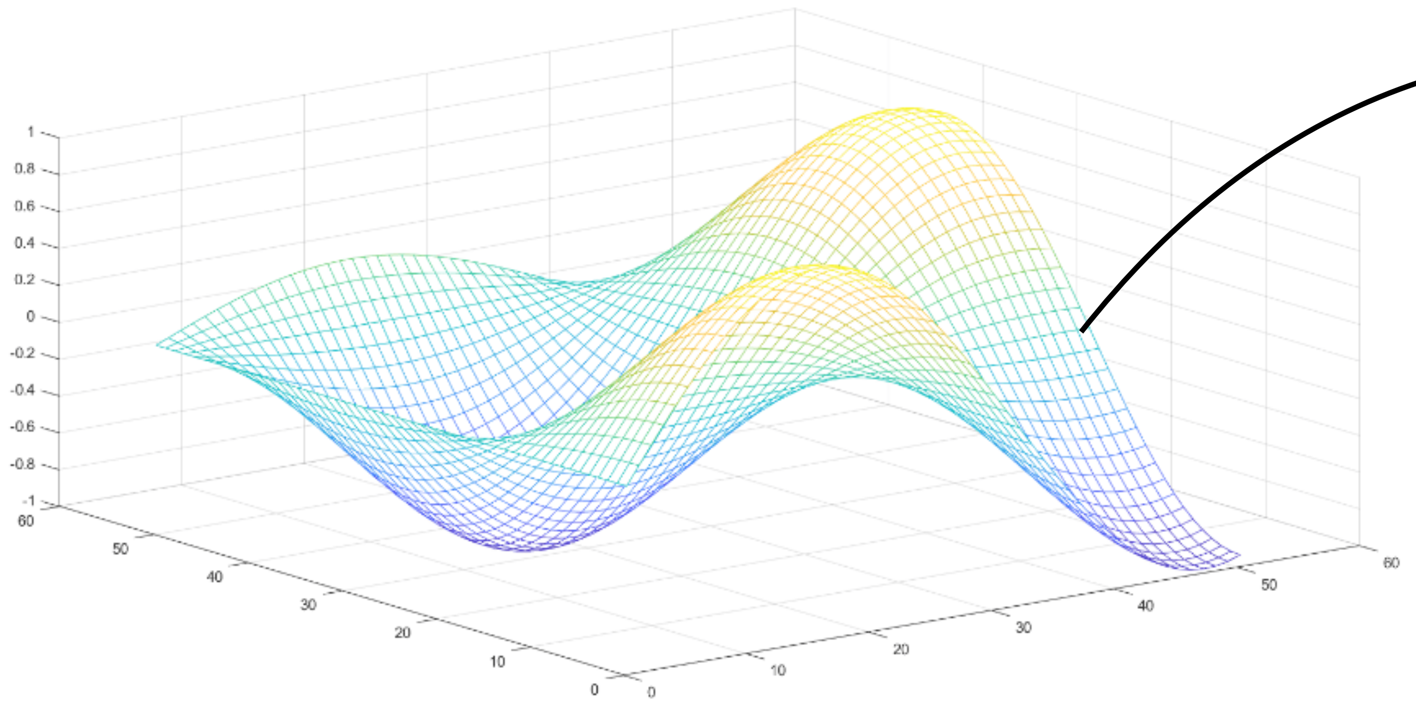


Asimov Histogram

Poisson per bin

$$N_i^{toy} = Poisson(N_i^{Asimov})$$

NSBI:



Poisson per event

$$w_i^{toy} = Poisson(w_i^{Asimov})$$

‘Unweighted’ events, i.e. integer weights

Negative weights? See backup

Dealing with negative weighted events

$$w_i^{toy} = Poisson(w_i^{Asimov})$$

Simulated samples include events with negative weights due to the way we calculate QFT higher order effects

Use a positive weighted sample instead:

1. Start from a positive weighted reference sample

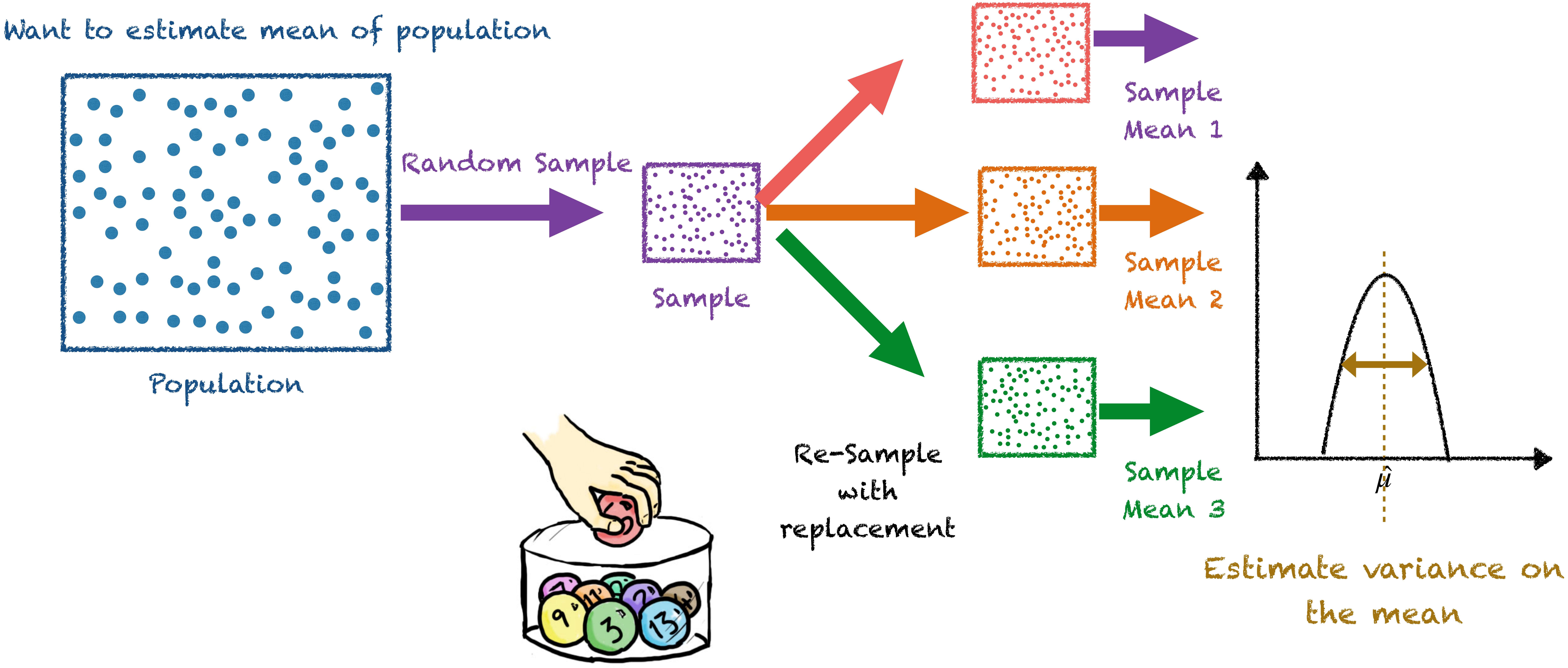
2. Re-weight it to intended parameter point in μ, α

3. Throw toys from this sample

$$w_i^{\text{rwt-ref}} \to w_i^{\text{Asimov}}(\mu, \alpha) = \frac{\nu(\mu, \alpha)}{\nu_{\text{rwt-ref}}} \cdot \frac{p(x_i|\mu, \alpha)}{p_{\text{rwt-ref}}(x_i)} \cdot w_i^{\text{rwt-ref}}$$

Uncertainty from finite training samples

Estimating the variance on mean: Bootstrapping



Want to estimate mean of population

Population

Random Sample

Sample

Re-Sample with replacement

Sample Mean 1

Sample Mean 2

Sample Mean 3

$\hat{\mu}$

Estimate variance on the mean

Image: Source

ATLAS methods paper: arXiv:2412.01600

Quantifying uncertainty on estimated density ratio

$$w_i \rightarrow w_i \cdot Pois(1)$$

- Train an ensemble of networks, each on a Poisson fluctuated version of the training dataset

- Use variance to estimate uncertainty

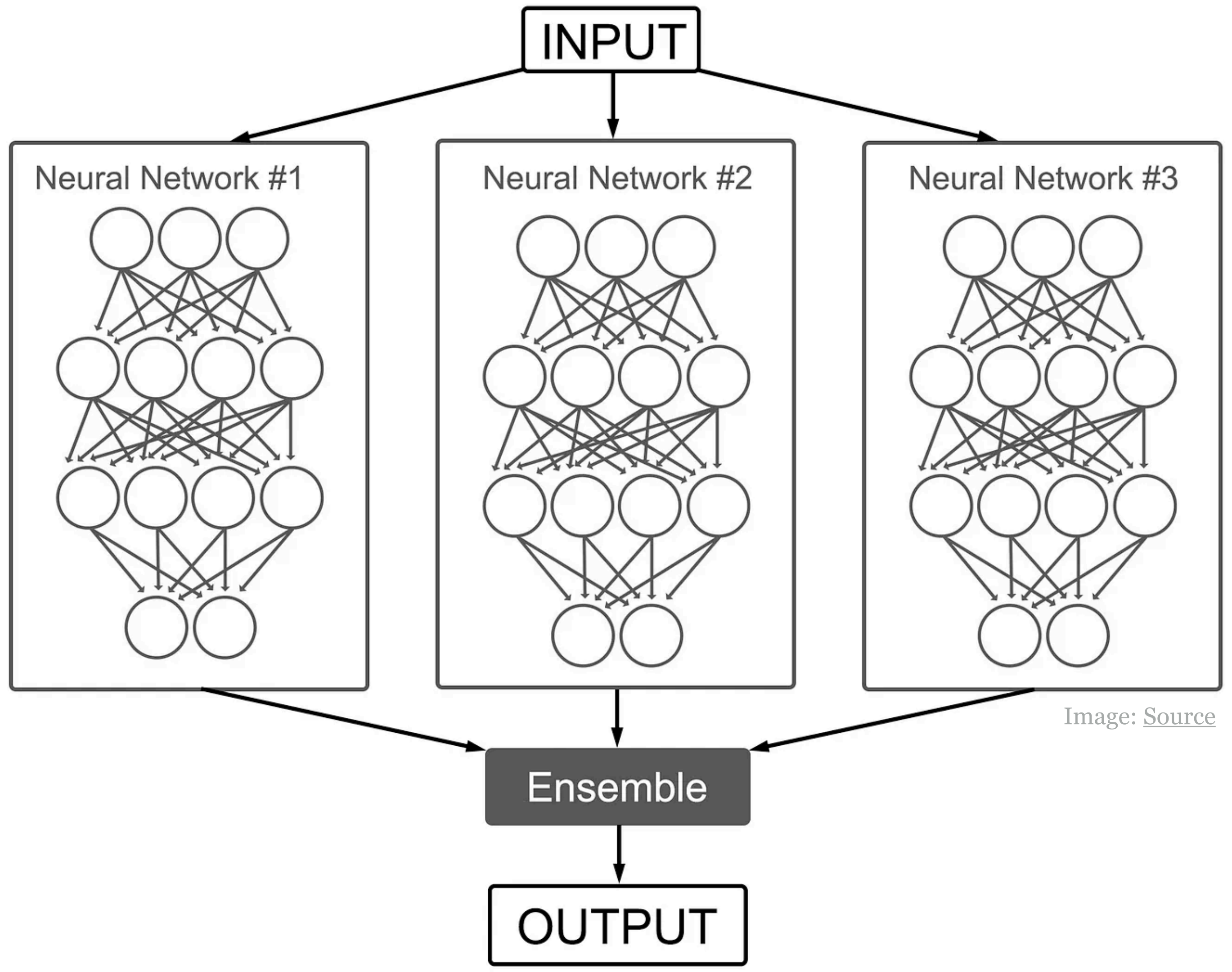


Image: Source

Quantifying uncertainty on estimated density ratio

$$w_i \rightarrow w_i \cdot Pois(1)$$

- Train an ensemble of networks, each on a Poisson fluctuated version of the training dataset
- Use variance to estimate uncertainty

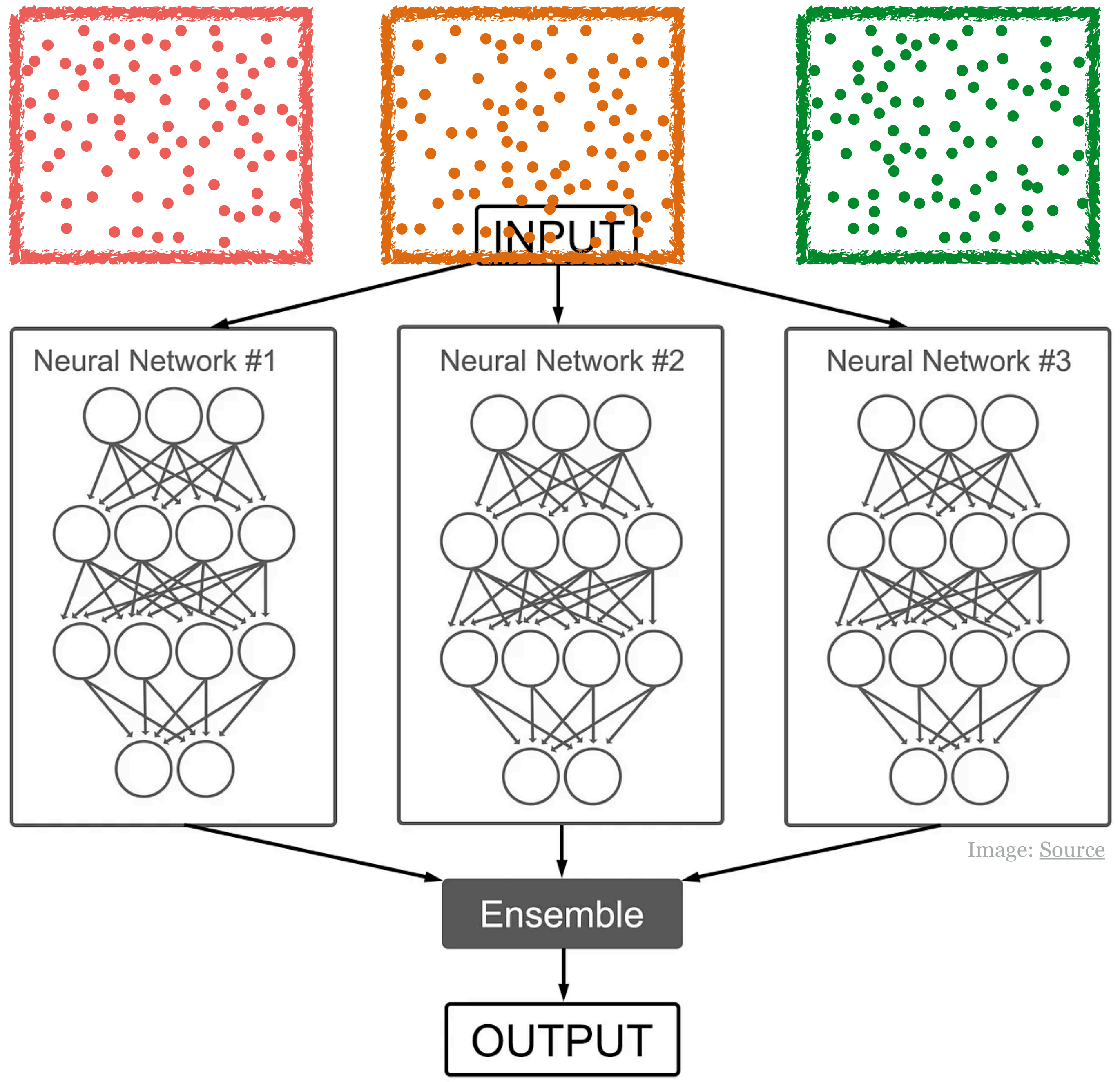


Image: Source

ATLAS methods paper: arXiv:2412.01600

Quantifying uncertainty on estimated density ratio

$$w_i \rightarrow w_i \cdot Pois(1)$$

- Train an ensemble of networks, each on a Poisson fluctuated version of the training dataset
- Use variance to estimate uncertainty

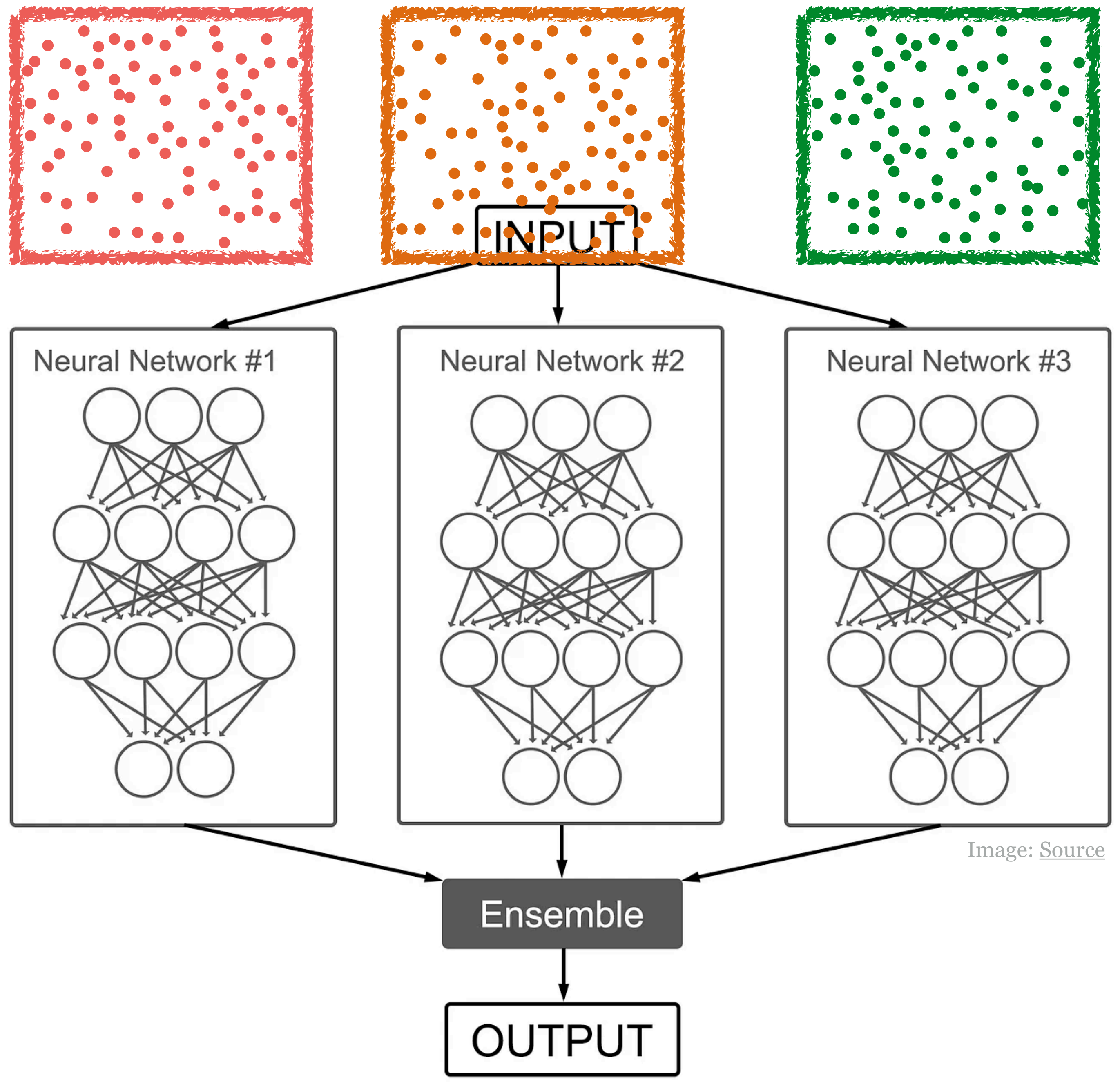
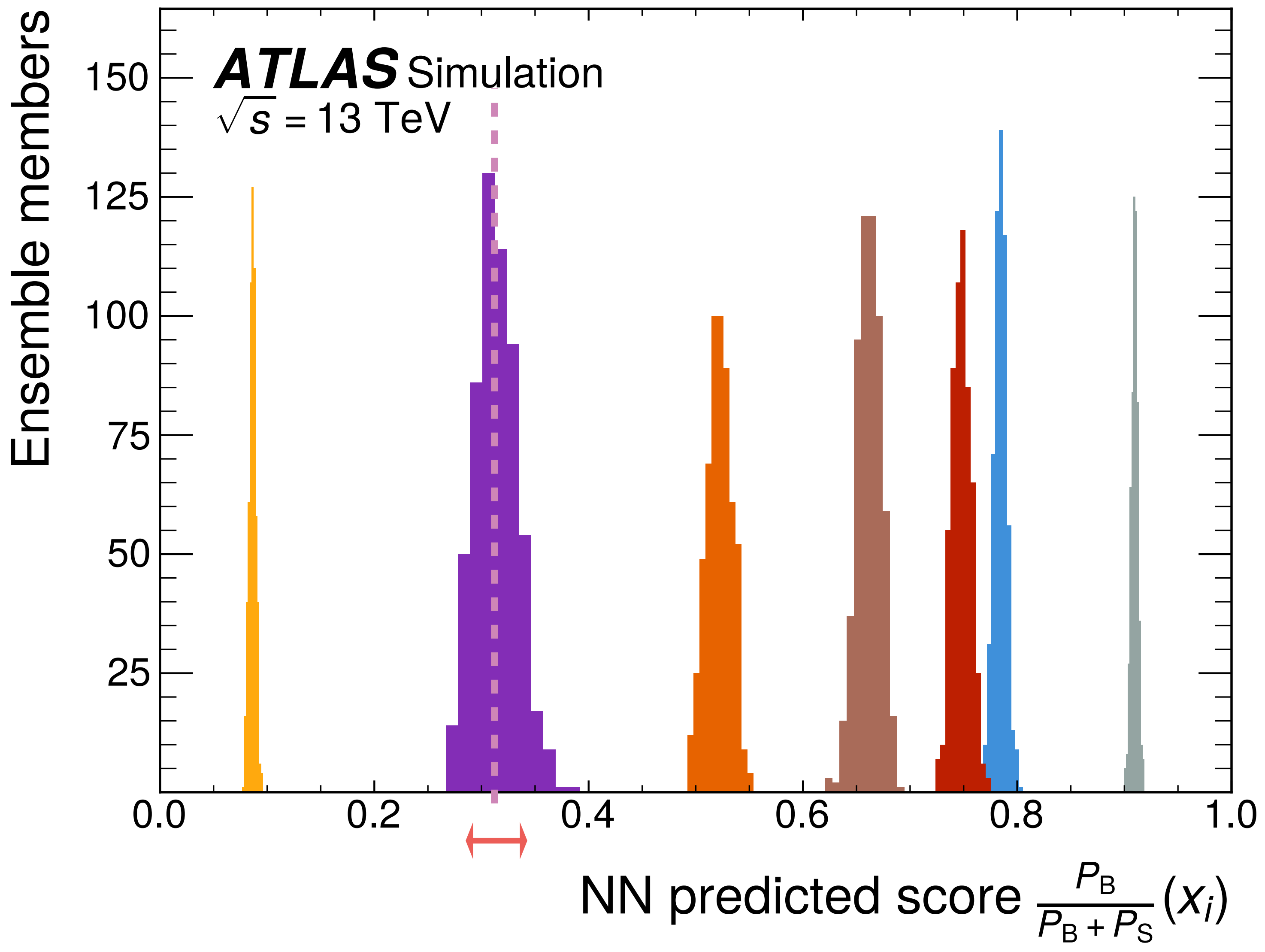


Image: Source

Distribution of NN predictions for example events



ATLAS methods paper: arXiv:2412.01600

Quantifying uncertainty on estimated density ratio

$$w_i \to w_i \cdot Pois(1)$$

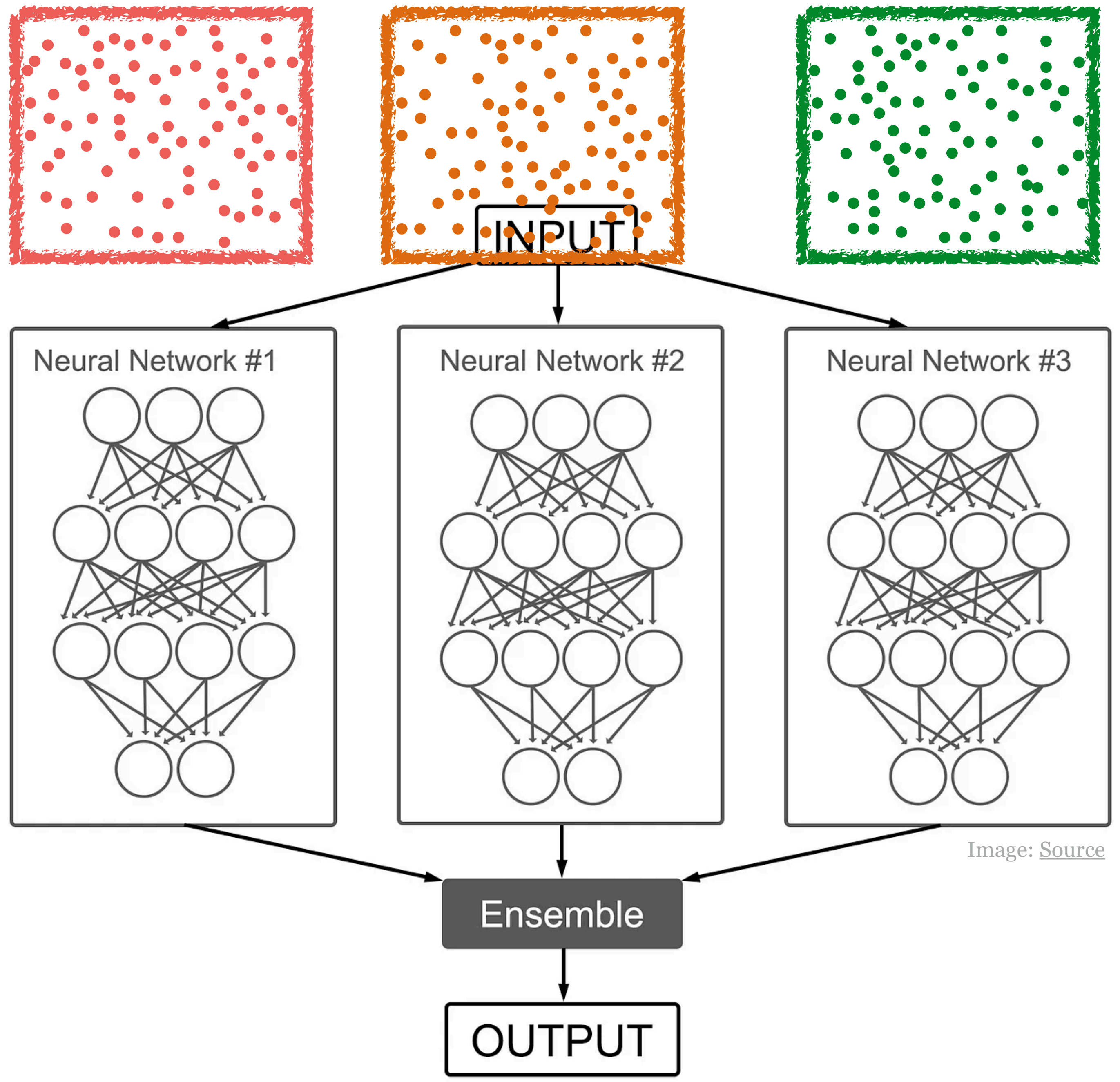


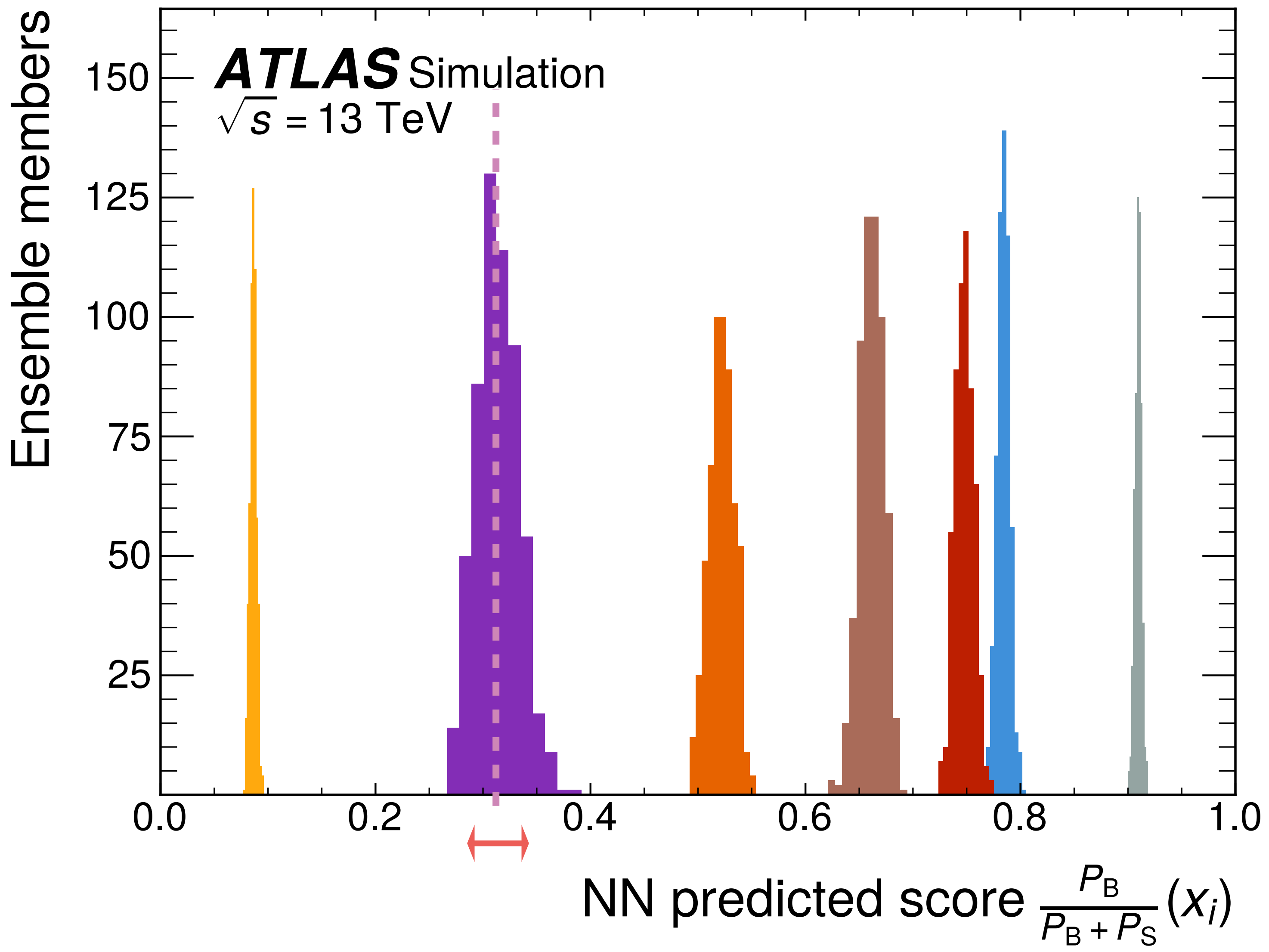
Image: Source

- Train an ensemble of networks, each on a Poisson fluctuated version of the training dataset
- Use variance to estimate uncertainty
- Propagate correlated impact

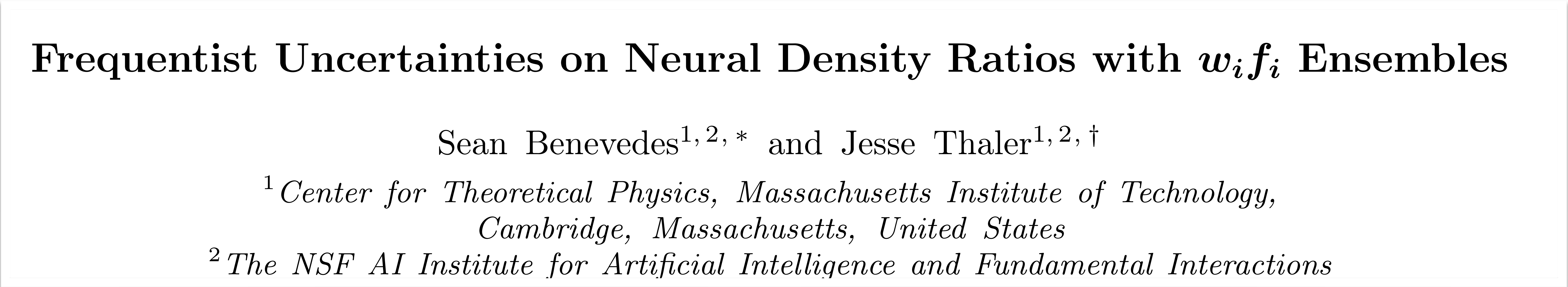
$$f_j(\mu) \to f_j(\mu + \alpha \cdot \Delta\hat{\mu}(\mu))$$

Constraint term: $Gauss(0,1)$

Distribution of NN predictions for example events



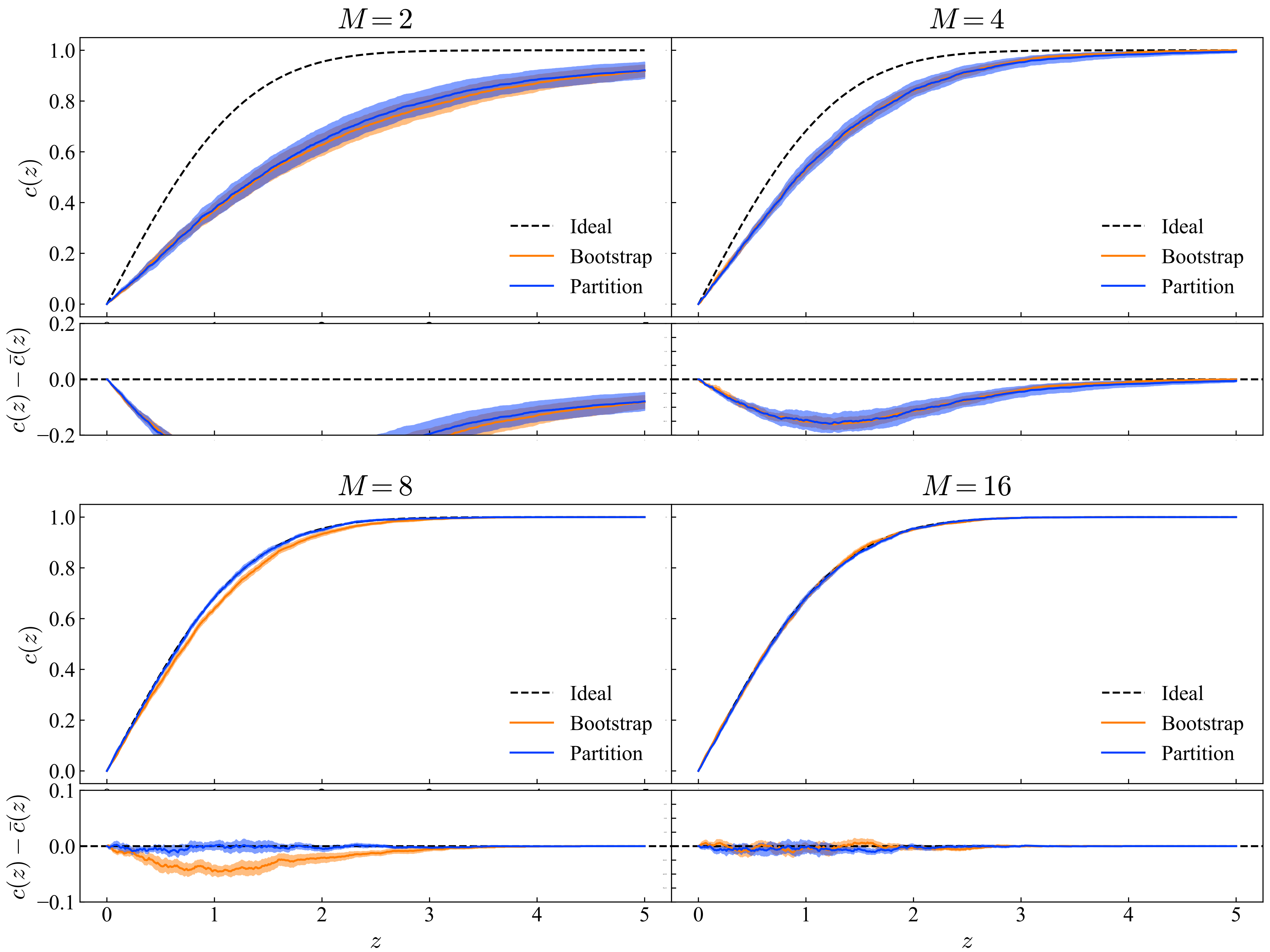
New innovations in efficient MC stat uncertainties



Frequentist Uncertainties on Neural Density Ratios with $w_i f_i$ Ensembles

Sean Benevedes[1, 2, *] and Jesse Thaler[1, 2, †]
[1] *Center for Theoretical Physics, Massachusetts Institute of Technology,*
Cambridge, Massachusetts, United States
[2] *The NSF AI Institute for Artificial Intelligence and Fundamental Interactions*

Gaussian Case Study, Coverage on Log Likelihood Ratio



- Requires fewer networks to ensure coverage
- Elegant, mathematically motivated method to estimate uncertainties

arXiv:2506.00113

Combination with histogram analyses

$$\frac{L_{\text{comb}}(\mu, \alpha)}{L_{\text{ref}}} = \frac{L_{\text{full}}(\mu, \alpha)}{L_{\text{ref}}} \; L_{\text{hist}}(\mu, \alpha)$$

Calculating pulls and impacts in JAX

$$\lambda(\mu, \alpha) = -2\ln(L_{full}(\mu, \alpha)/L_{ref})$$

Hessian:

$$C_{nm} = \left[\frac{1}{2}\frac{\partial^2 \lambda}{\partial \alpha_n \partial \alpha_m}(\hat{\mu}, \hat{\alpha})\right]^{-1}$$

Pulls:

$$\frac{\widehat{\alpha}_k - \alpha_k^0}{\sqrt{C_{kk}}}.$$

Post-fit Impact:

$$\begin{aligned}\Gamma_k &= \frac{\partial \widehat{\mu}}{\partial \alpha_k} \times \sqrt{C_{kk}} \\ &= -\left[\frac{\partial^2 \lambda}{\partial^2 \mu}(\widehat{\mu}, \widehat{\alpha})\right]^{-1} \frac{\partial^2 \lambda}{\partial \mu \partial \alpha_k}(\widehat{\mu}, \widehat{\alpha}) \times \sqrt{C_{kk}},\end{aligned}$$

Vertical interpolation

$$G_j(\alpha_k) = \begin{cases} \left(\frac{\nu_j(\alpha_k^+)}{\nu_j(\alpha_k^0)}\right)^{\alpha_k} & \alpha_k > 1 \\ 1 + \sum_{n=1}^{6} c_n \alpha_k^n & -1 \le \alpha_k \le 1 \\ \left(\frac{\nu_j(\alpha_k^-)}{\nu_j(\alpha_k^0)}\right)^{-\alpha_k} & \alpha_k < -1 \end{cases}$$

$$g_j(x_i, \alpha_k) = \begin{cases} \left(g_j(x_i, \alpha_k^+)\right)^{\alpha_k} & \alpha_k > 1 \\ 1 + \sum_{n=1}^{6} c_n \alpha_k^n & -1 \le \alpha_k \le 1 \\ \left(g_j(x_i, \alpha_k^-)\right)^{-\alpha_k} & \alpha_k < -1 \end{cases}$$

With some continuity requirements

Physics analysis results

ATLAS Higgs width analysis paper: arXiv:2412.01548

Comparison to previous result on same data

Parameter	Value	68% CL interval		95% CL interval	
		Observed	**Expected**	**Observed**	**Expected**
		NSBI analysis			
$\mu_{\text{off-shell}}$ (4ℓ only)	0.87	[0.33, 1.62]	[0.05, 2.04]	[0.05, 2.38]	< 2.38
$\mu_{\text{off-shell}}$	1.06	[0.61, 1.67]	[0.17, 1.83]	[0.21, 2.24]	[0.01, 2.42]
Γ_H [MeV] (4ℓ only)	3.43	[1.37, 6.71]	[0.20, 8.25]	[0.18, 9.98]	< 12.09
Γ_H [MeV]	4.29	[2.41, 6.95]	[0.66, 7.61]	[0.76, 9.66]	[0.12, 10.50]
R_{gg}	1.19	[0.53, 2.07]	[0.02, 1.92]	< 2.96	< 2.73
R_{VV}	0.95	[0.61, 1.39]	[0.31, 1.70]	[0.30, 1.86]	[0.06, 2.14]
		Histogram-based analysis			
$\mu_{\text{off-shell}}$ (4ℓ only)	0.79	[0.02, 2.00]	< 2.14	< 2.97	< 3.10
$\mu_{\text{off-shell}}$	1.09	[0.54, 1.81]	[0.08, 1.90]	[0.10, 2.41]	[0.01, 2.52]
Γ_H [MeV] (4ℓ only)	3.43	[0.10, 8.42]	< 8.89	< 12.48	< 12.89
Γ_H [MeV]	4.37	[2.13, 7.43]	[0.35, 7.94]	[0.39, 10.14]	< 10.79
R_{gg}	1.23	[0.00, 2.20]	< 1.98	< 3.15	< 2.84
R_{VV}	0.95	[0.60, 1.43]	[0.27, 1.74]	[0.26, 1.90]	[0.02, 2.18]

ATLAS Higgs width analysis paper: arXiv:2412.01548

Impact of nuisance parameters

Systematic Uncertainty Fixed	$\mu_{\text{off-shell}}$ **Value at which** $t_{\mu_{\text{off-shell}}} = 4$	
	NSBI analysis	**Histogram-based**
All (stat-only)	1.96	2.13
Parton shower uncertainty for $gg \to ZZ$ (normalization)	2.07	2.26
Parton shower uncertainty for $gg \to ZZ$ (shape)	2.12	2.29
NLO EW uncertainty for $q\bar{q} \to ZZ$	2.10	2.27
NLO QCD uncertainty for $gg \to ZZ$	2.09	2.29
Parton shower uncertainty for $q\bar{q} \to ZZ$ (shape)	2.12	2.29
Jet energy scale and resolution uncertainty	2.11	2.26
None (full result)	2.12	2.30

ATLAS Higgs width analysis paper: arXiv:2412.01548

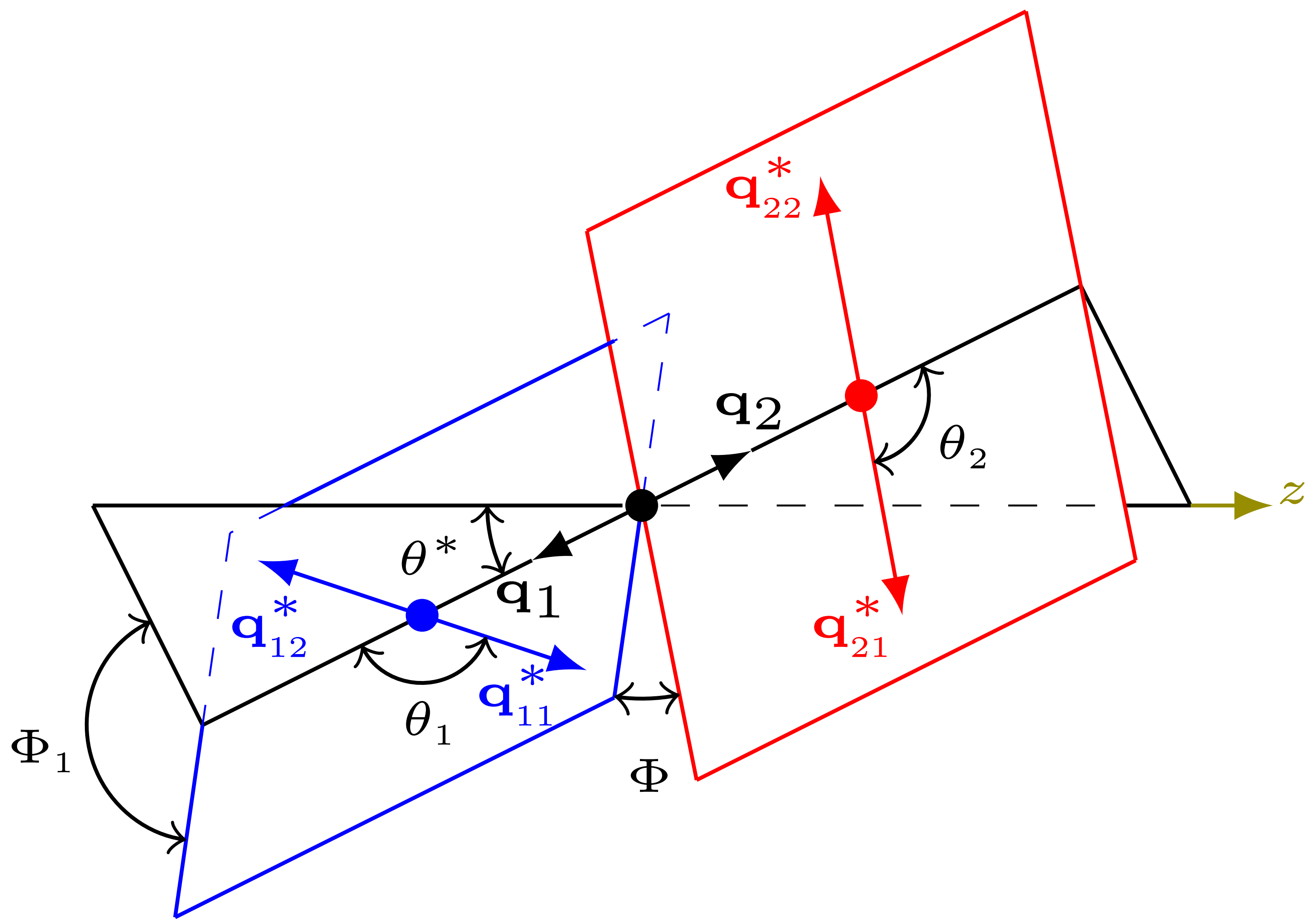
Full probability model, input variables

$$p(x|\mu^{\text{ggF}}_{\text{off-shell}},\mu^{\text{EW}}_{\text{off-shell}}) = \frac{1}{\nu(\mu^{\text{ggF}}_{\text{off-shell}},\mu^{\text{EW}}_{\text{off-shell}})}\times$$

$$\Bigg[\mu^{\text{ggF}}_{\text{off-shell}}\nu^{\text{ggF}}_{\text{S}}p^{\text{ggF}}_{\text{S}}(x) + \sqrt{\mu^{\text{ggF}}_{\text{off-shell}}}\nu^{\text{ggF}}_{\text{I}}p^{\text{ggF}}_{\text{I}}(x) + \nu^{\text{ggF}}_{\text{B}}p^{\text{ggF}}_{\text{B}}(x) +$$

$$\mu^{\text{EW}}_{\text{off-shell}}\nu^{\text{EW}}_{\text{S}}p^{\text{EW}}_{\text{S}}(x) + \sqrt{\mu^{\text{EW}}_{\text{off-shell}}}\nu^{\text{EW}}_{\text{I}}p^{\text{EW}}_{\text{I}}(x) + \nu^{\text{EW}}_{\text{B}}p^{\text{EW}}_{\text{B}}(x) + \nu_{\text{NI}}p_{\text{NI}}(x)\Bigg],$$

Variable	Definition
$m_{4\ell}$	quadruplet mass
m_{Z1}	Z_1 mass
m_{Z2}	Z_2 mass
$\cos\theta^*$	cosine of the Higgs boson decay angle $[\mathbf{q}_1 \cdot \mathbf{n}_z/|\mathbf{q}_1|]$
$\cos\theta_1$	cosine of the Z_1 decay angle $[-(\mathbf{q}_2)\cdot\mathbf{q}_{11}/(|\mathbf{q}_2|\cdot|\mathbf{q}_{11}|)]$
$\cos\theta_2$	cosine of the Z_2 decay angle $[-(\mathbf{q}_1)\cdot\mathbf{q}_{21}/(|\mathbf{q}_1|\cdot|\mathbf{q}_{21}|)]$
Φ_1	Z_1 decay plane angle $[\cos^{-1}(\mathbf{n}_1\cdot\mathbf{n}_{\text{sc}})\,(\mathbf{q}_1\cdot(\mathbf{n}_1\times\mathbf{n}_{\text{sc}})/(|\mathbf{q}_1|\cdot|\mathbf{n}_1\times\mathbf{n}_{\text{sc}}|)]$
Φ	angle between Z_1, Z_2 decay planes $[\cos^{-1}(\mathbf{n}_1\cdot\mathbf{n}_2)\,(\mathbf{q}_1\cdot(\mathbf{n}_1\times\mathbf{n}_2)/(|\mathbf{q}_1|\cdot|\mathbf{n}_1\times\mathbf{n}_2|)]$
$p_T^{4\ell}$	quadruplet transverse momentum
$y^{4\ell}$	quadruplet rapidity
n_{jets}	number of jets in the event
m_{jj}	leading dijet system mass
$\Delta\eta_{jj}$	leading dijet system pseudorapidity
$\Delta\phi_{jj}$	leading dijet system azimuthal angle difference



ATLAS methods paper: arXiv:2412.01600

Network architecture

Feed-forward dense networks

- 5 hidden layers with 1000 nodes
- Swish activation
- Single node output layer with sigmoid activation

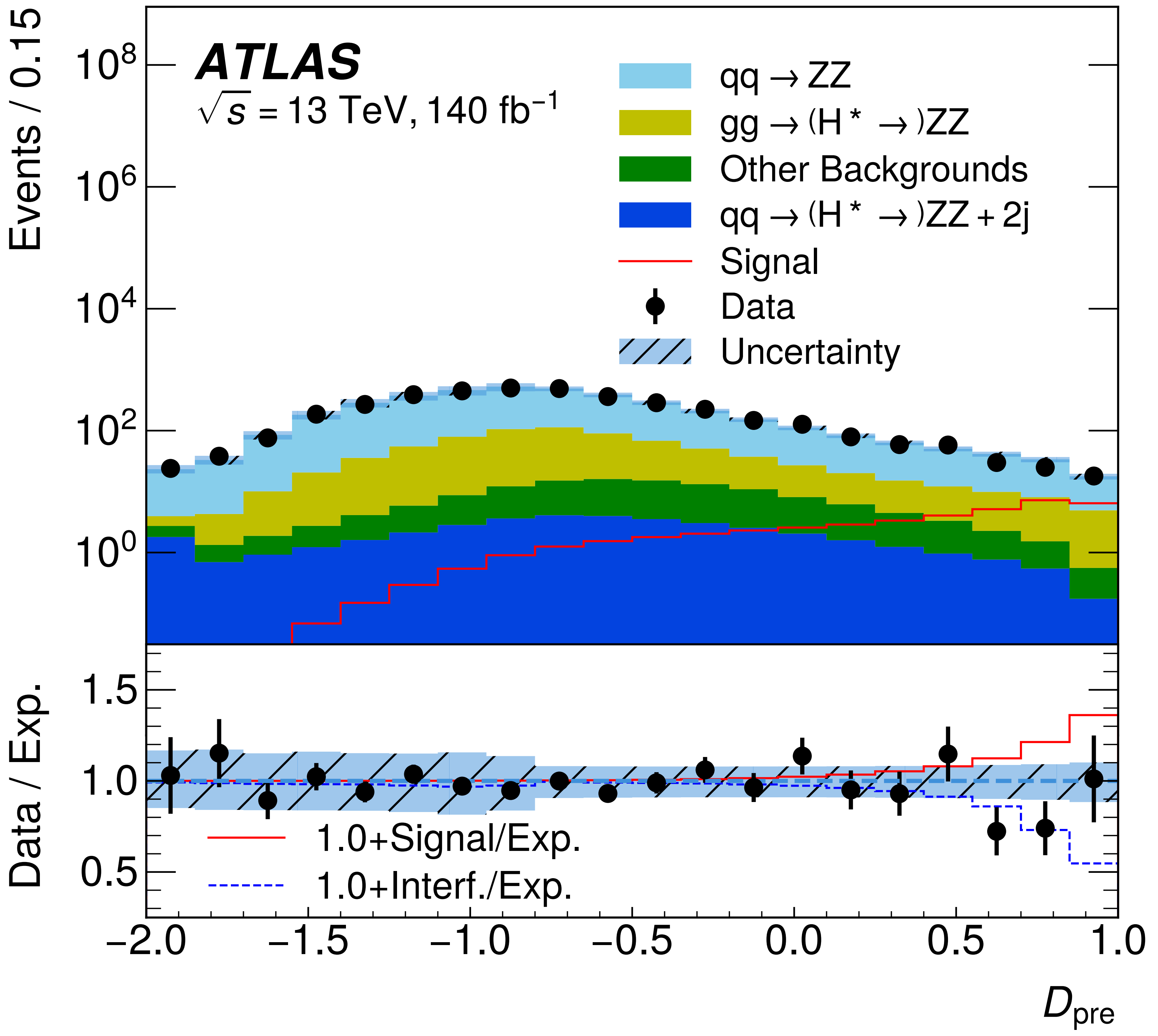
Loss: Weighted binary cross-entropy

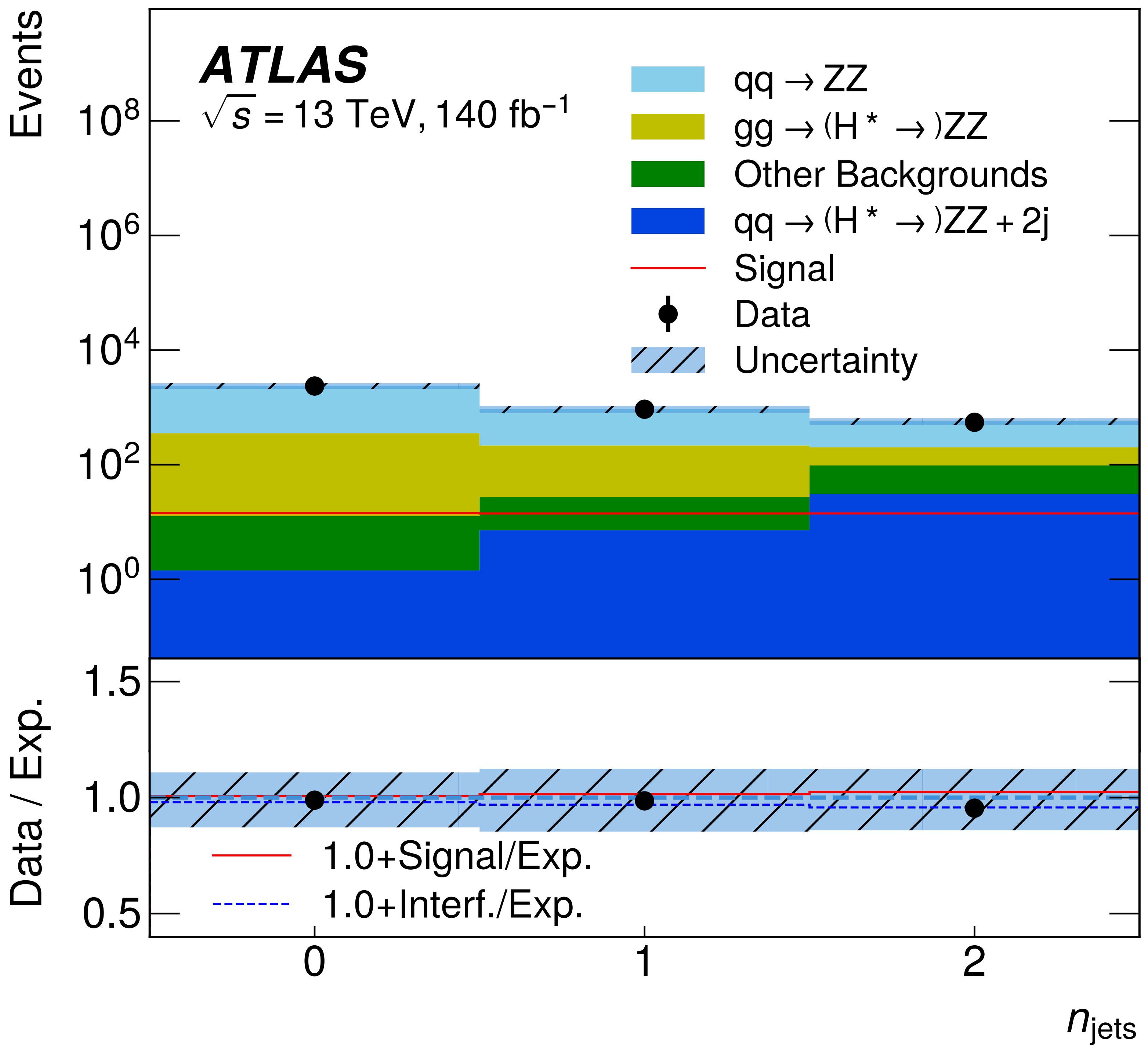
$O(10^4)$ networks takes approx 4000 GPU hours to train

ATLAS Higgs width analysis paper: arXiv:2412.01548

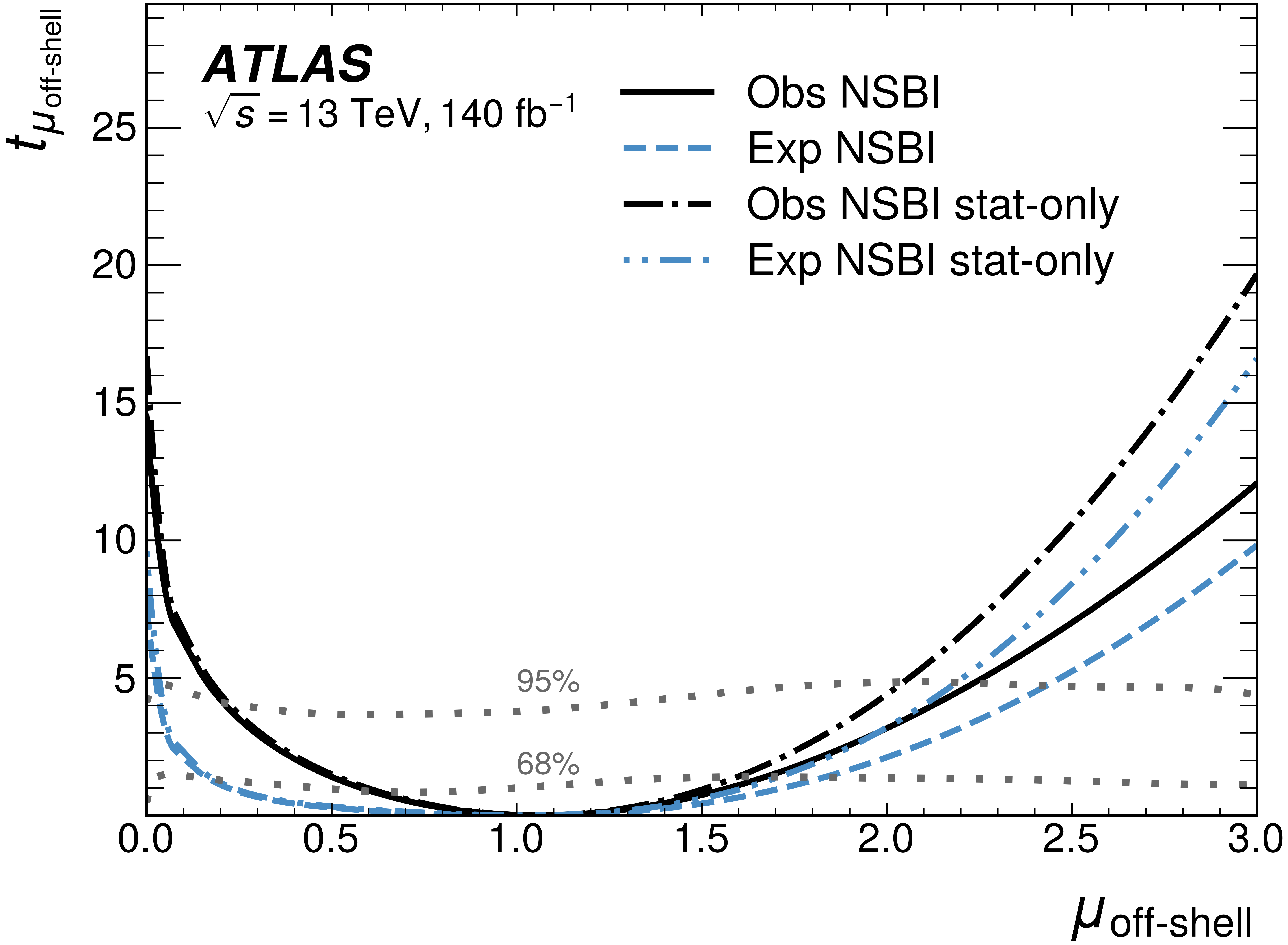
Pre-selection region definition

$$D_{\text{pre}}(x) = \log \frac{s^{\text{ggF}}_{\text{pre, S}}(x) + s^{\text{EW}}_{\text{pre, S}}(x)}{s^{\text{ggF}}_{\text{pre, B}}(x) + s^{\text{EW}}_{\text{pre, B}}(x) + s_{\text{pre, }q\bar{q}ZZ}(x)},$$



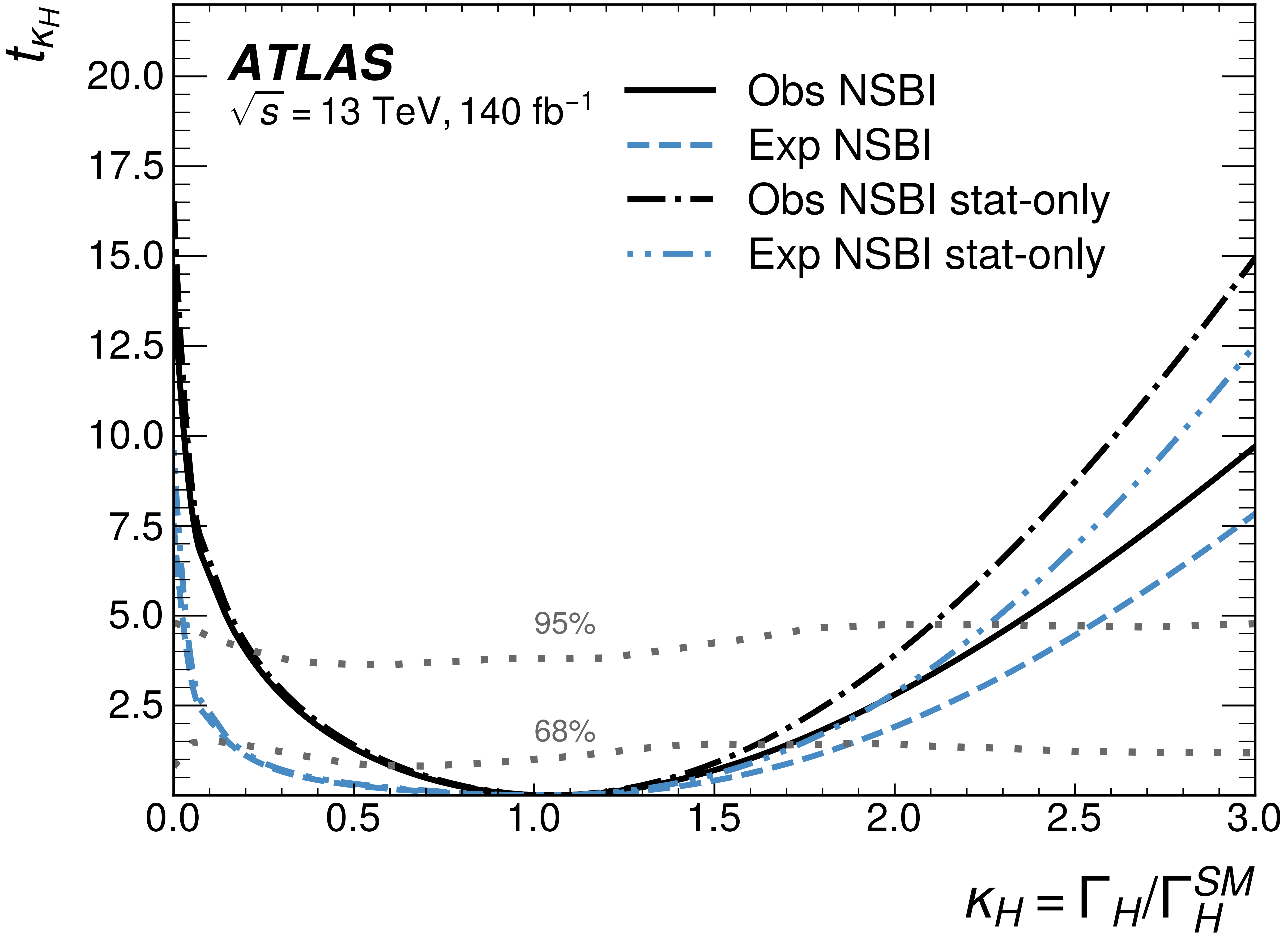


Result after combination with $ll\nu\nu$



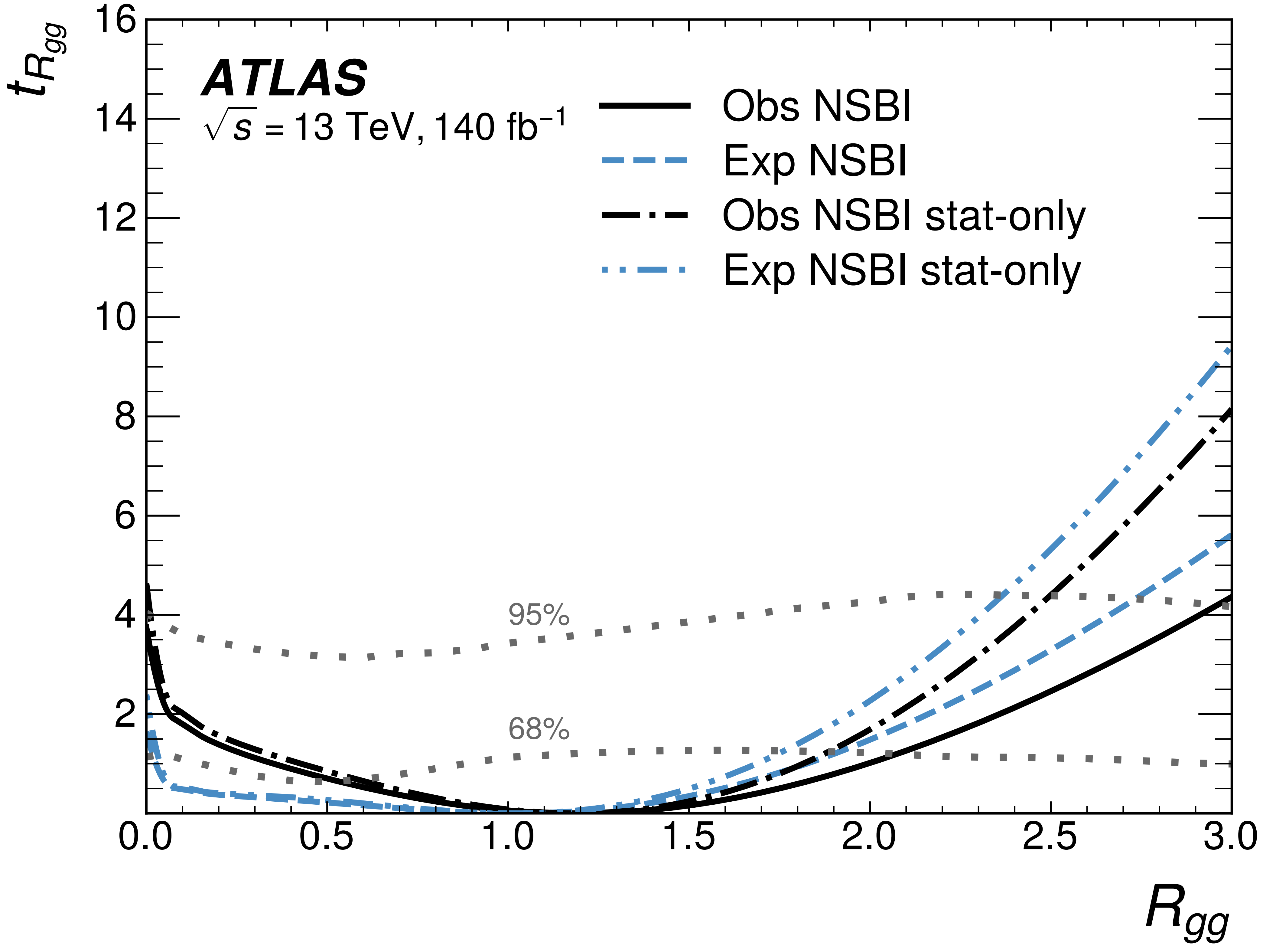
$ll\nu\nu$ dominates sensitivity

Width interpretation

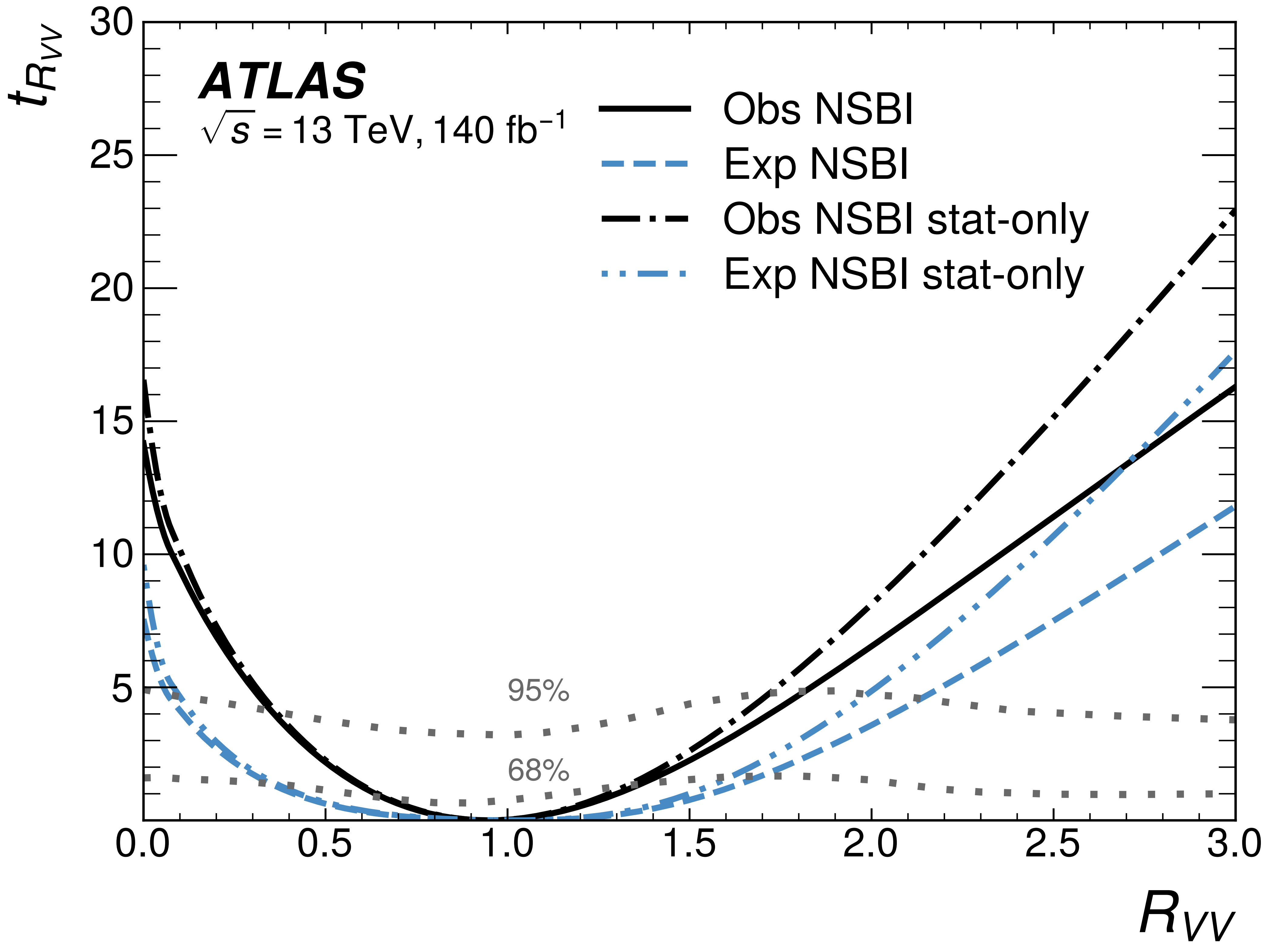


CI obtained from Neyman construction

Width sensitivity in ggF and VBF



$$R_{gg} = \kappa^2_{g,\text{on-shell}}/\kappa^2_{g,\text{off-shell}}$$



$$R_{VV} = \kappa^2_{V,\text{on-shell}}/\kappa^2_{V,\text{off-shell}}$$