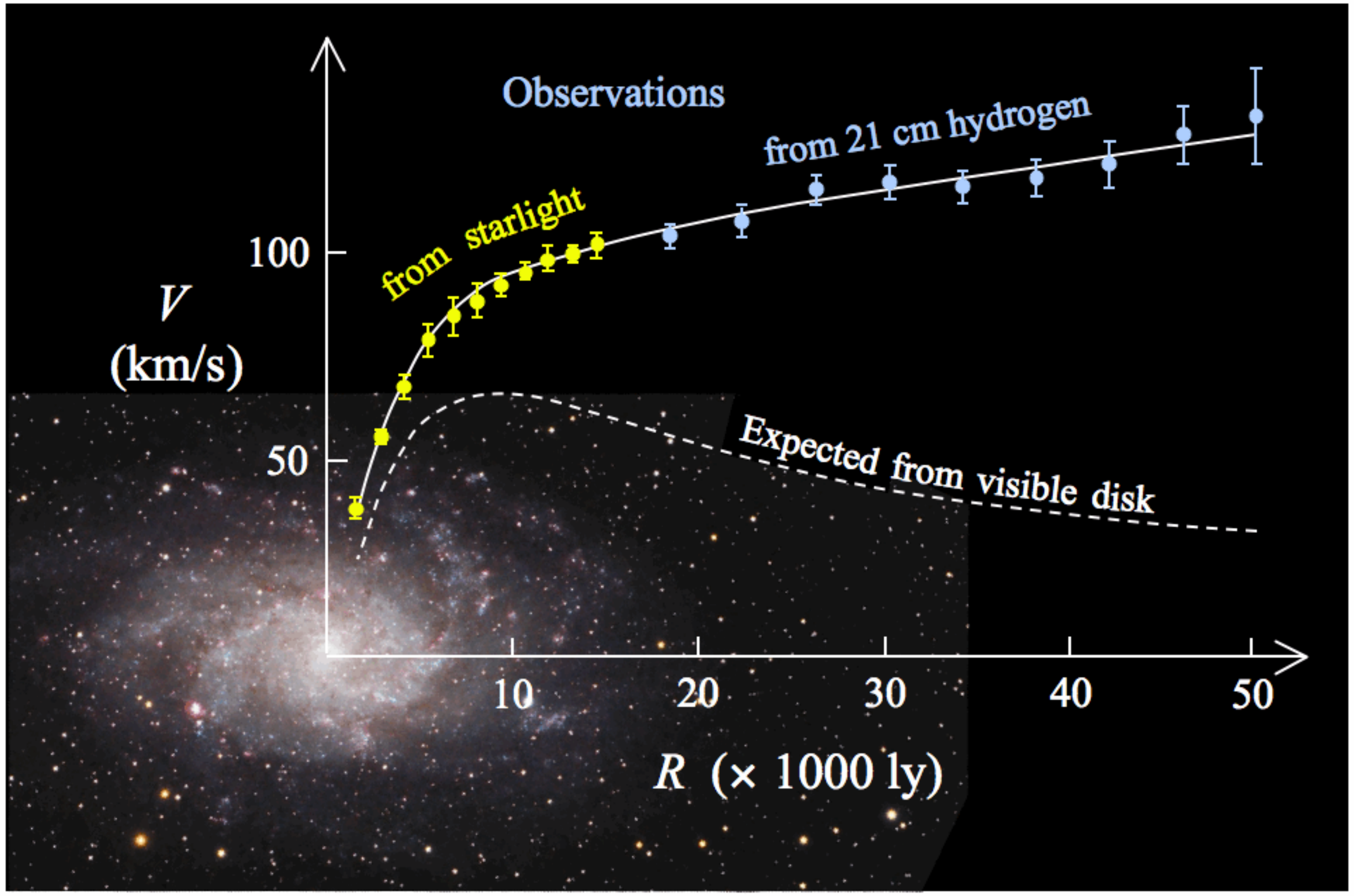
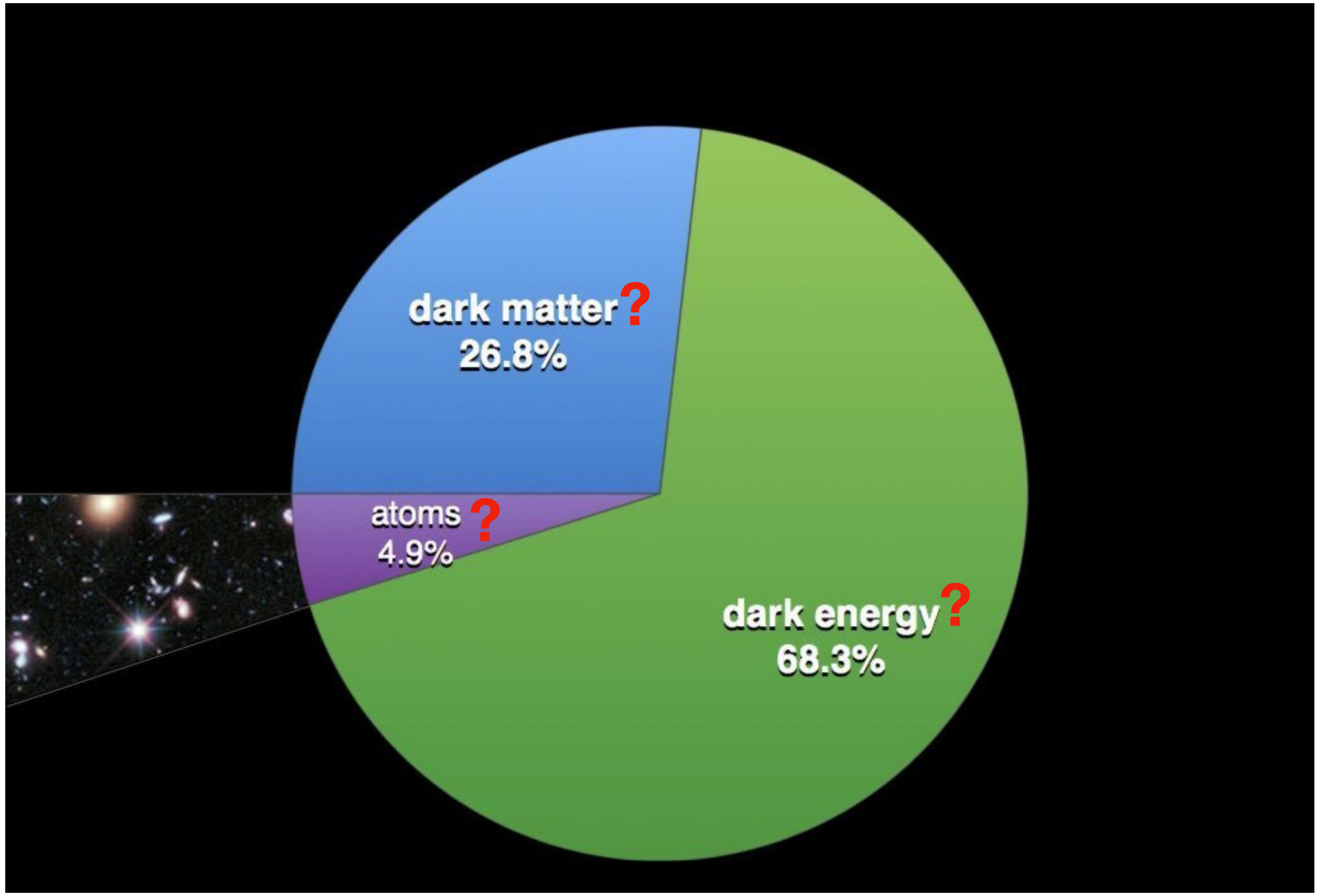
Why stop at two?
The search for HHH production

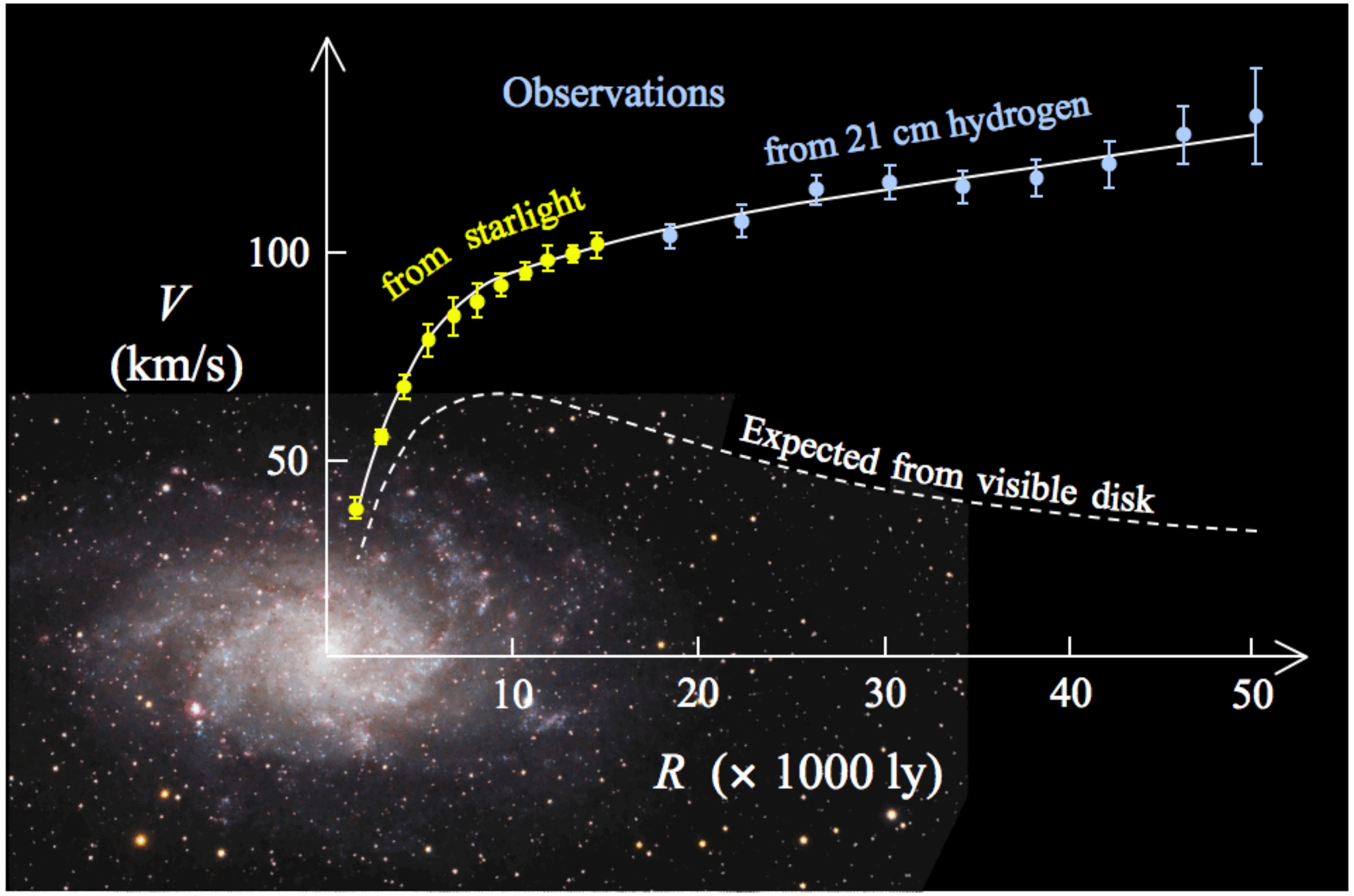
Kevin Nelson
SLAC Seminar
21 January, 2025

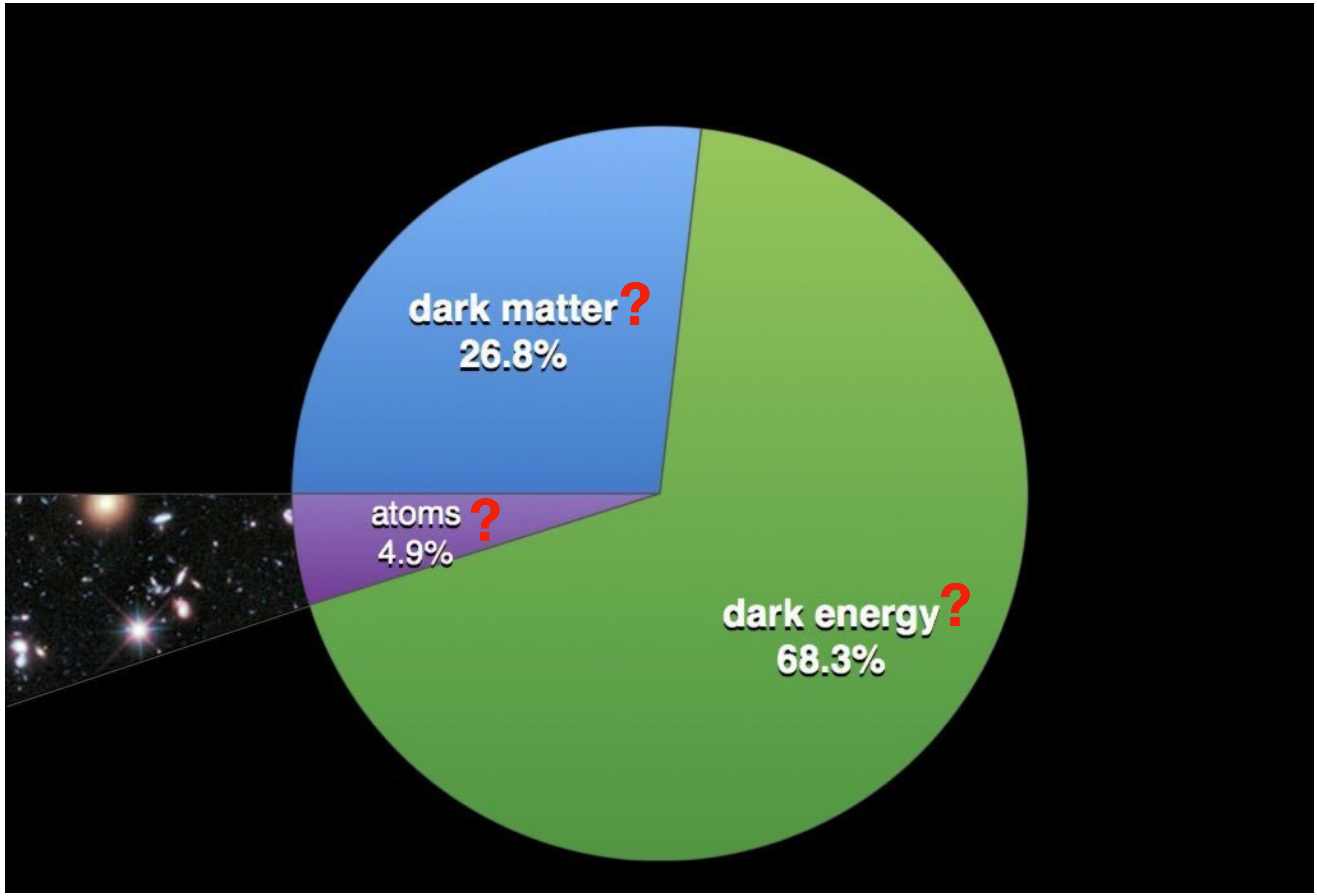
We've already seen "new physics"





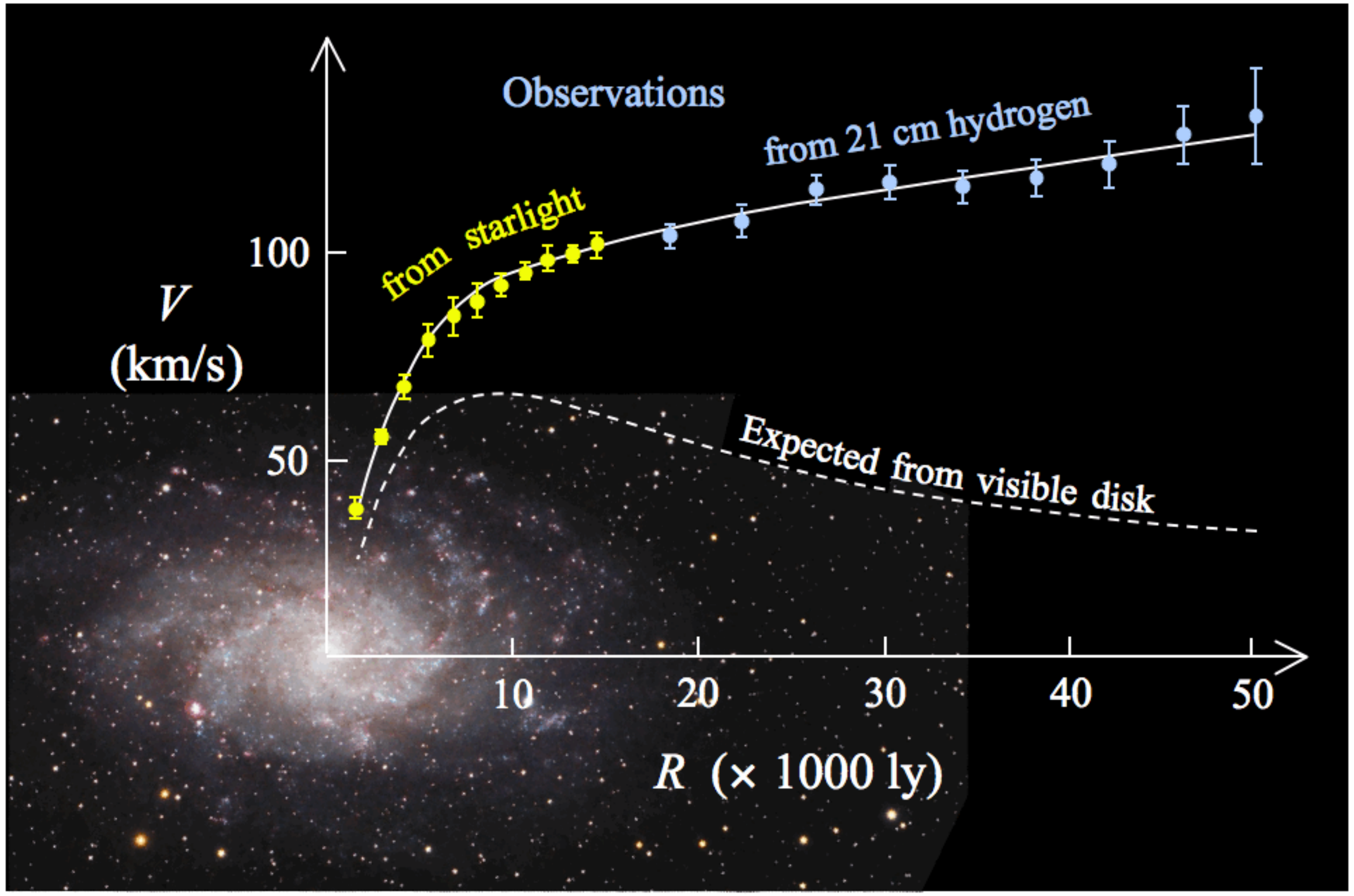
We've already seen "new physics"

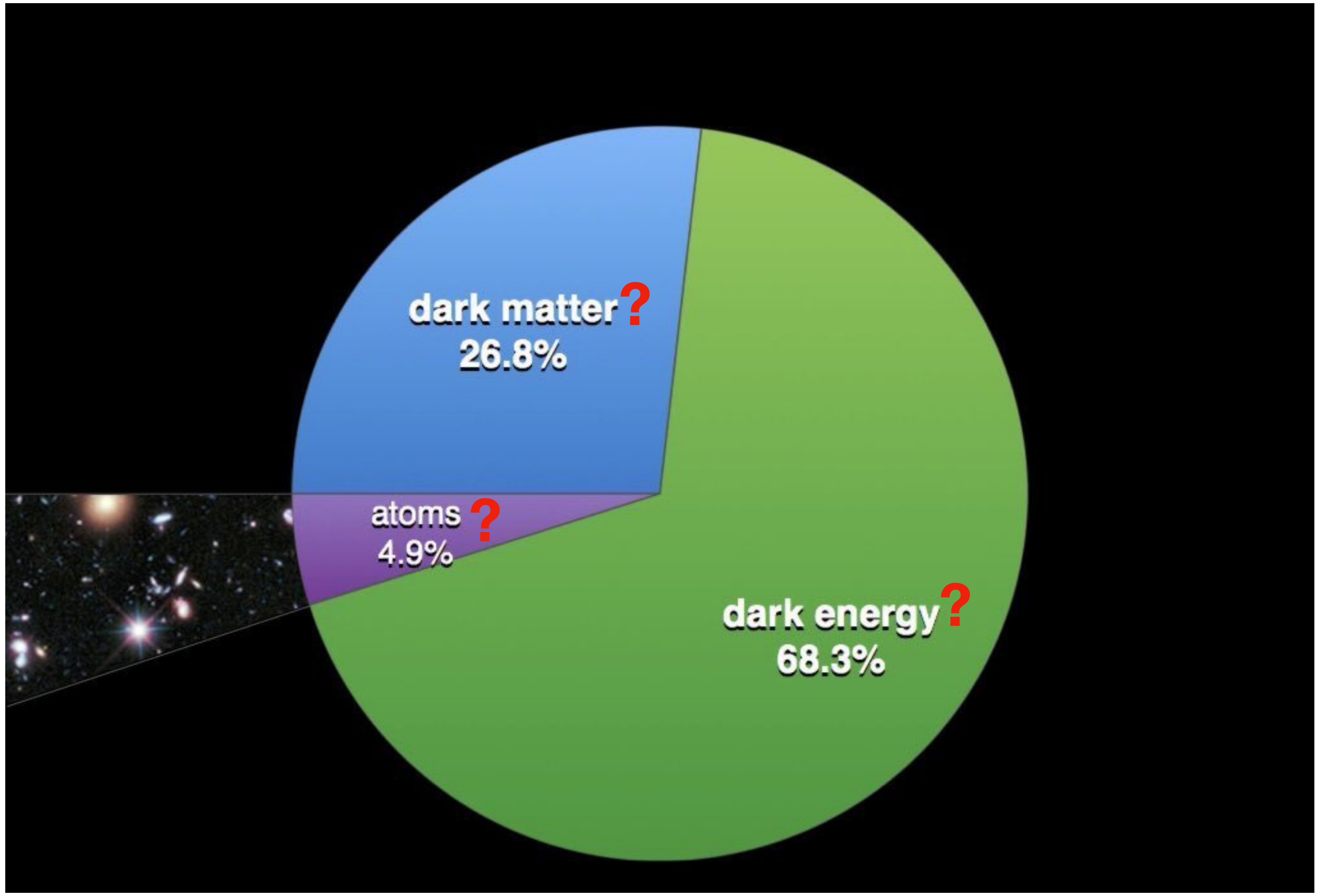




- **But we are left with some questions:**
  - Can we describe the particle content of dark matter?
  - Are there mediators between dark and luminous matter?
  - Can we explain the matter-anti-matter asymmetry and baryogenesis?

We've already seen "new physics"





- **But we are left with some questions:**
  - Can we describe the particle content of dark matter?
  - Are there mediators between dark and luminous matter?
  - Can we explain the matter-anti-matter asymmetry and baryogenesis?

All of these questions are related to the search for HHH!

Can we use modern AI methods to find it?

Is there physics beyond the standard model?

Can we use modern AI methods to find it?

Is there physics beyond the standard model?

Yes, there is physics beyond the Standard Model.

Can we use modern AI methods to find it?

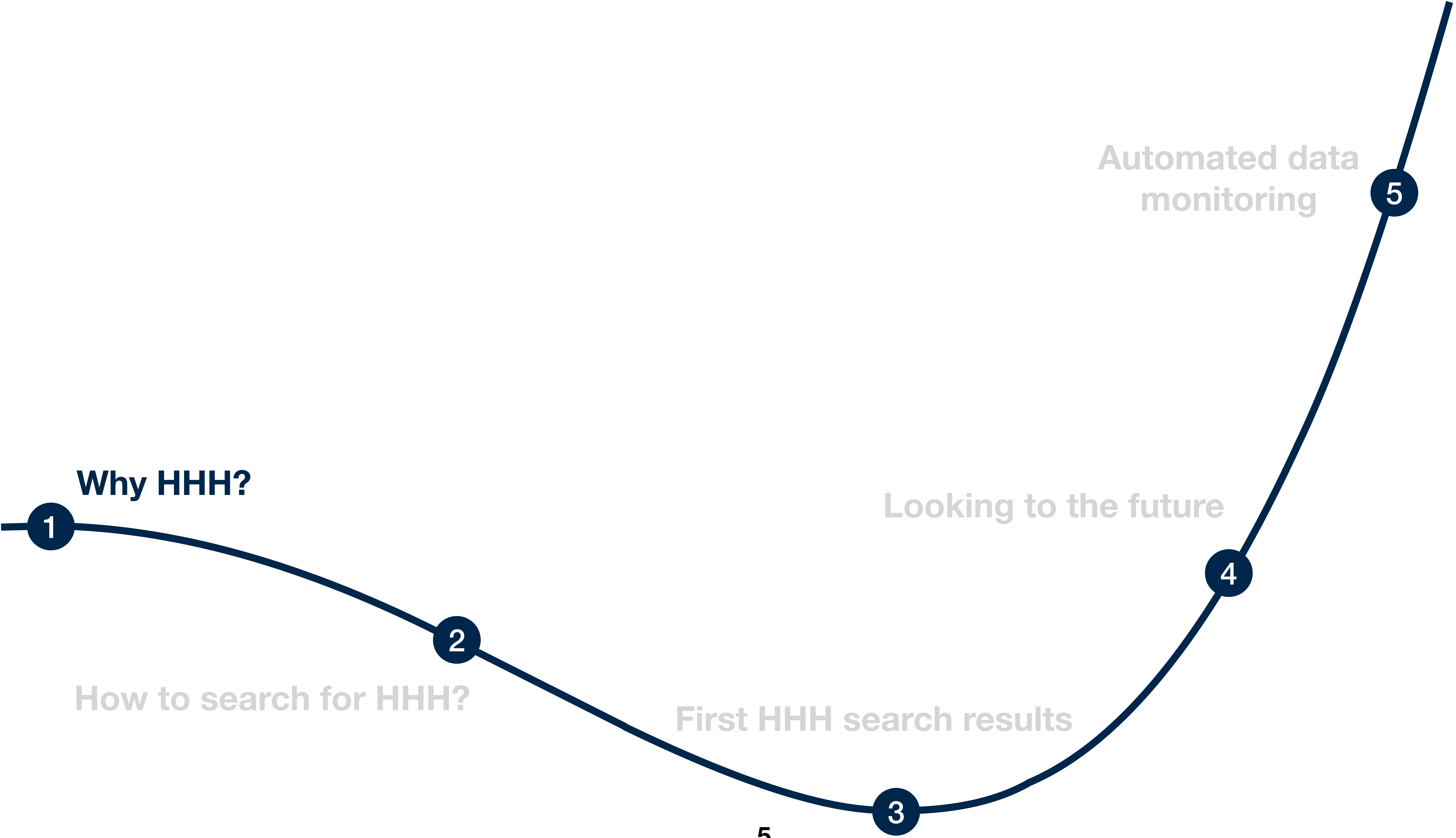
Is there physics beyond the standard model?

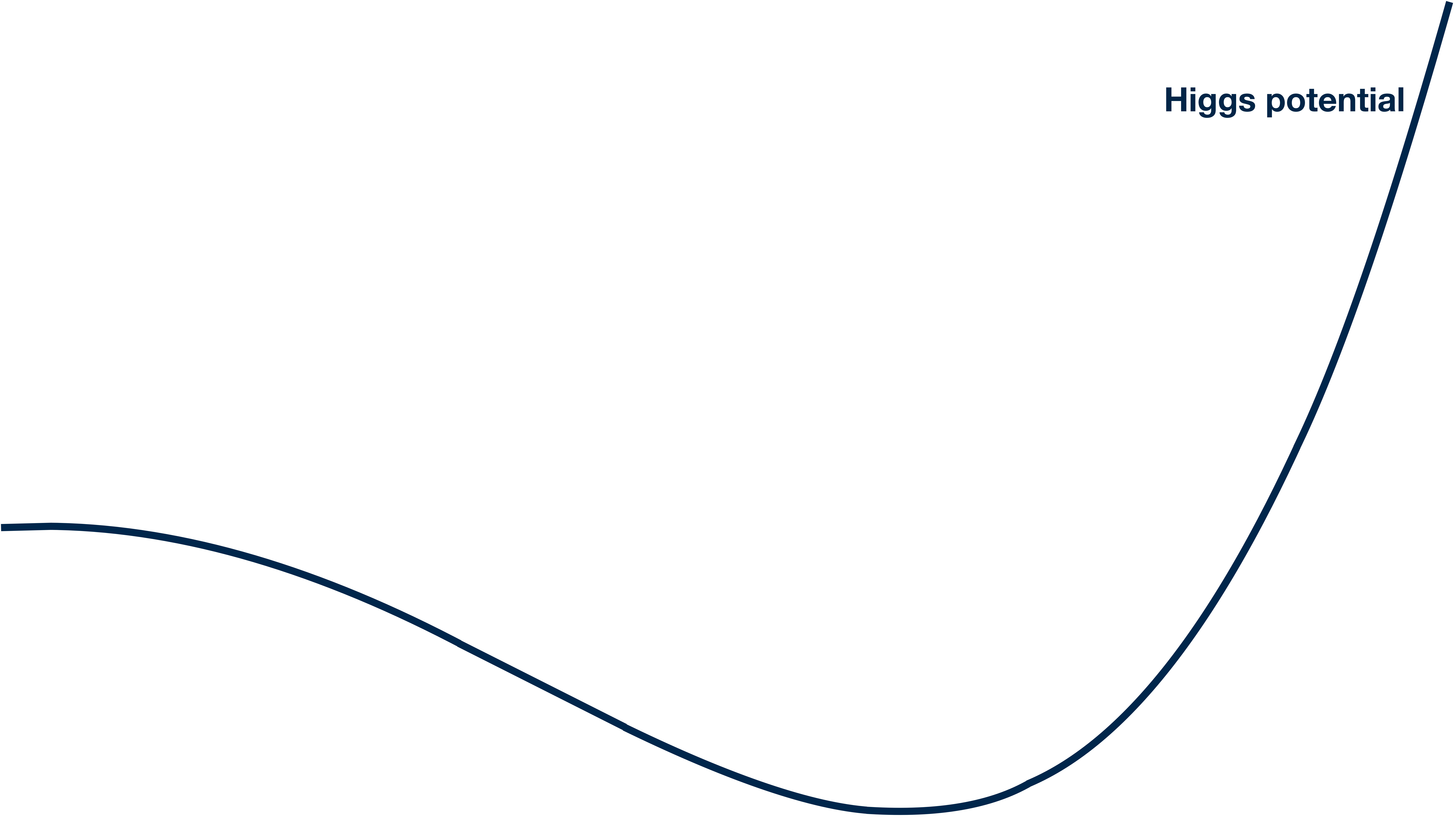
Yes, there is physics beyond the Standard Model. The Standard Model of particle physics is a highly successful theory that describes the electromagnetic, weak, and strong nuclear forces, along with the particles that make up matter, such as quarks, leptons, and bosons. However, it is known to be incomplete and has several limitations. Some of the major areas of physics beyond the Standard Model (BSM) include:

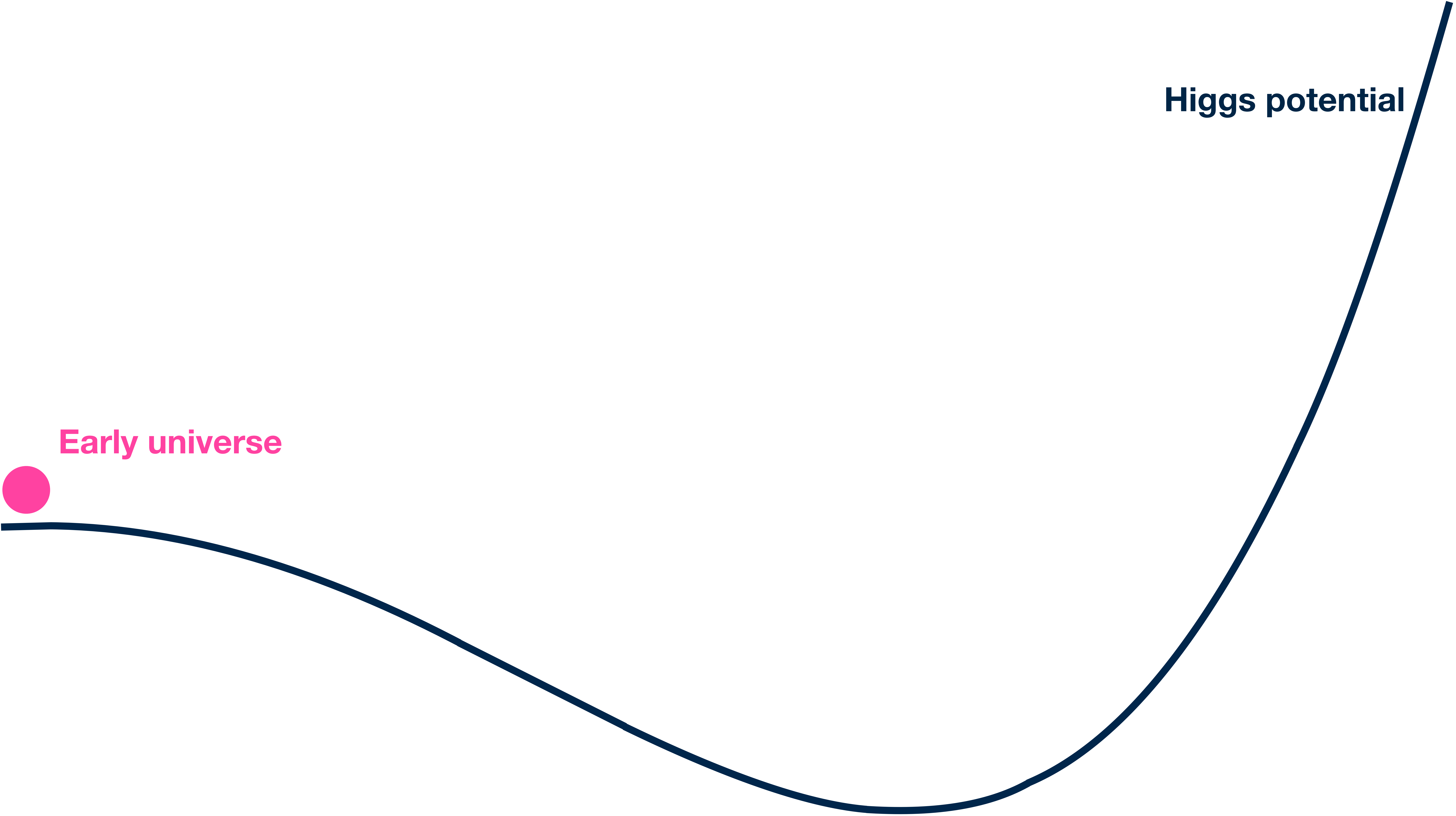
1. Gravity

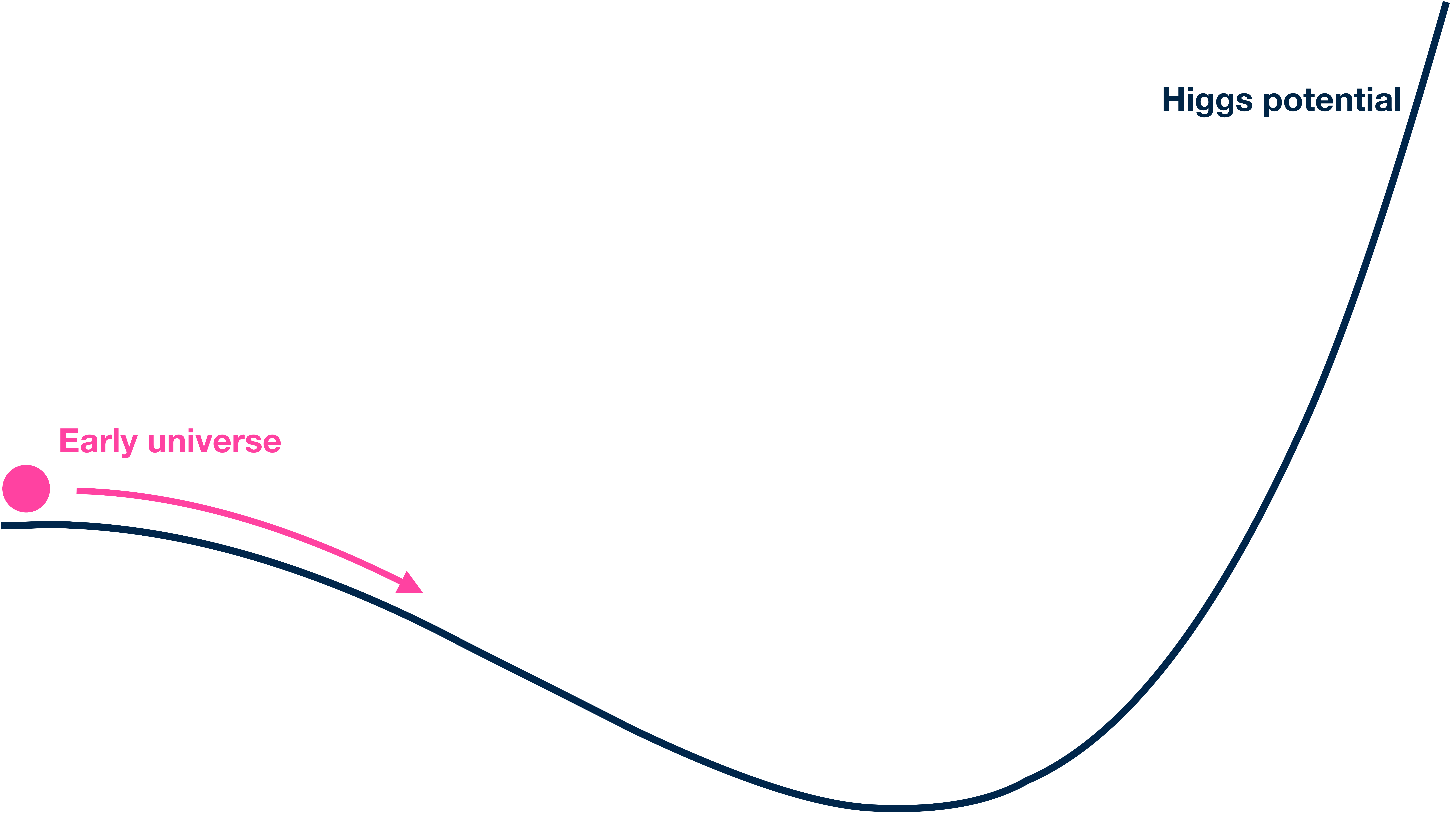
The Standard Model does not include gravity. General relativity describes gravity on a large scale, but it is not compatible with quantum mechanics, which underlies the Standard Model. A theory of

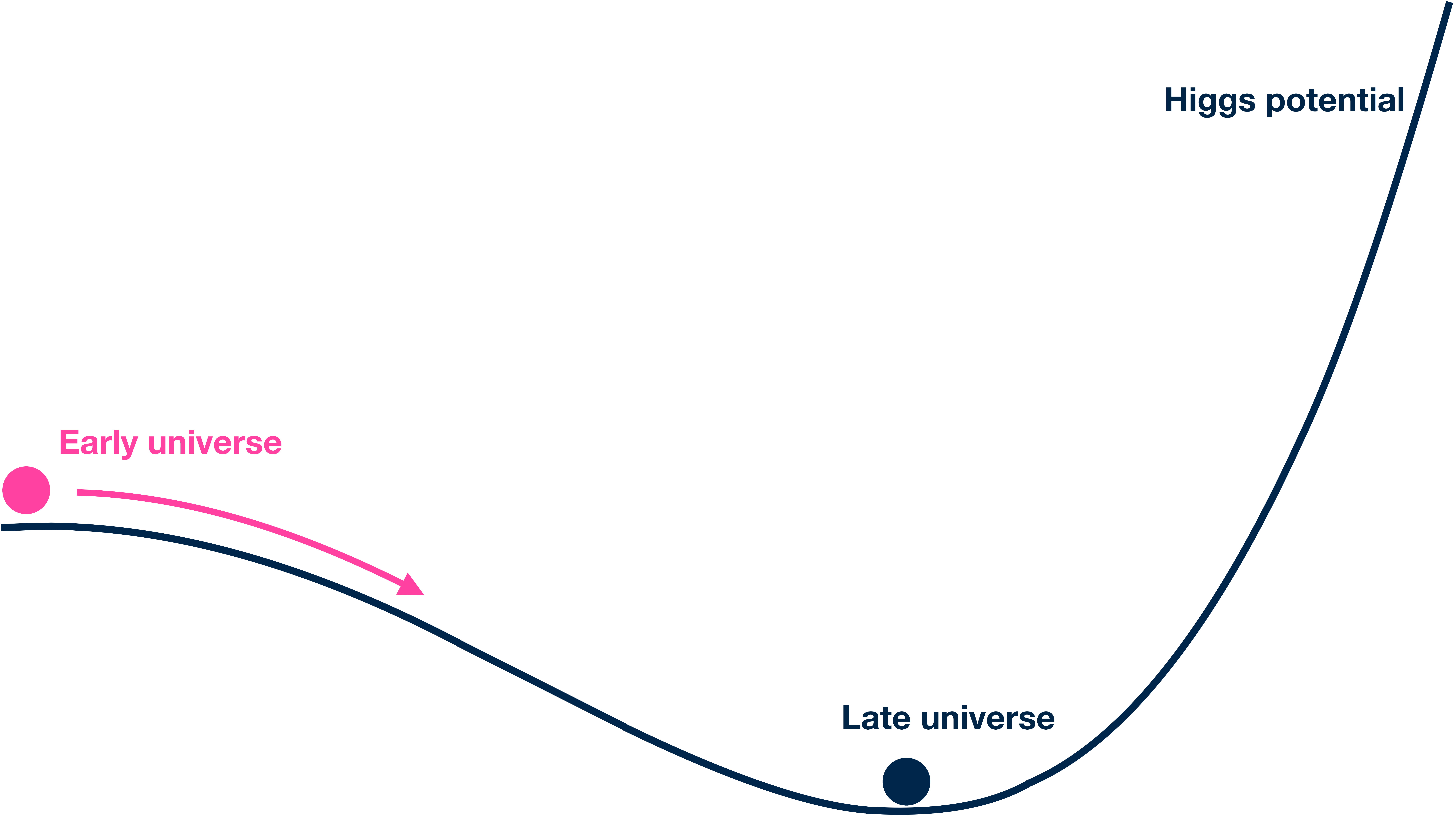

1
Why HHH?
2
How to search for HHH?
3
First HHH search results
4
Looking to the future
5
Automated data
monitoring

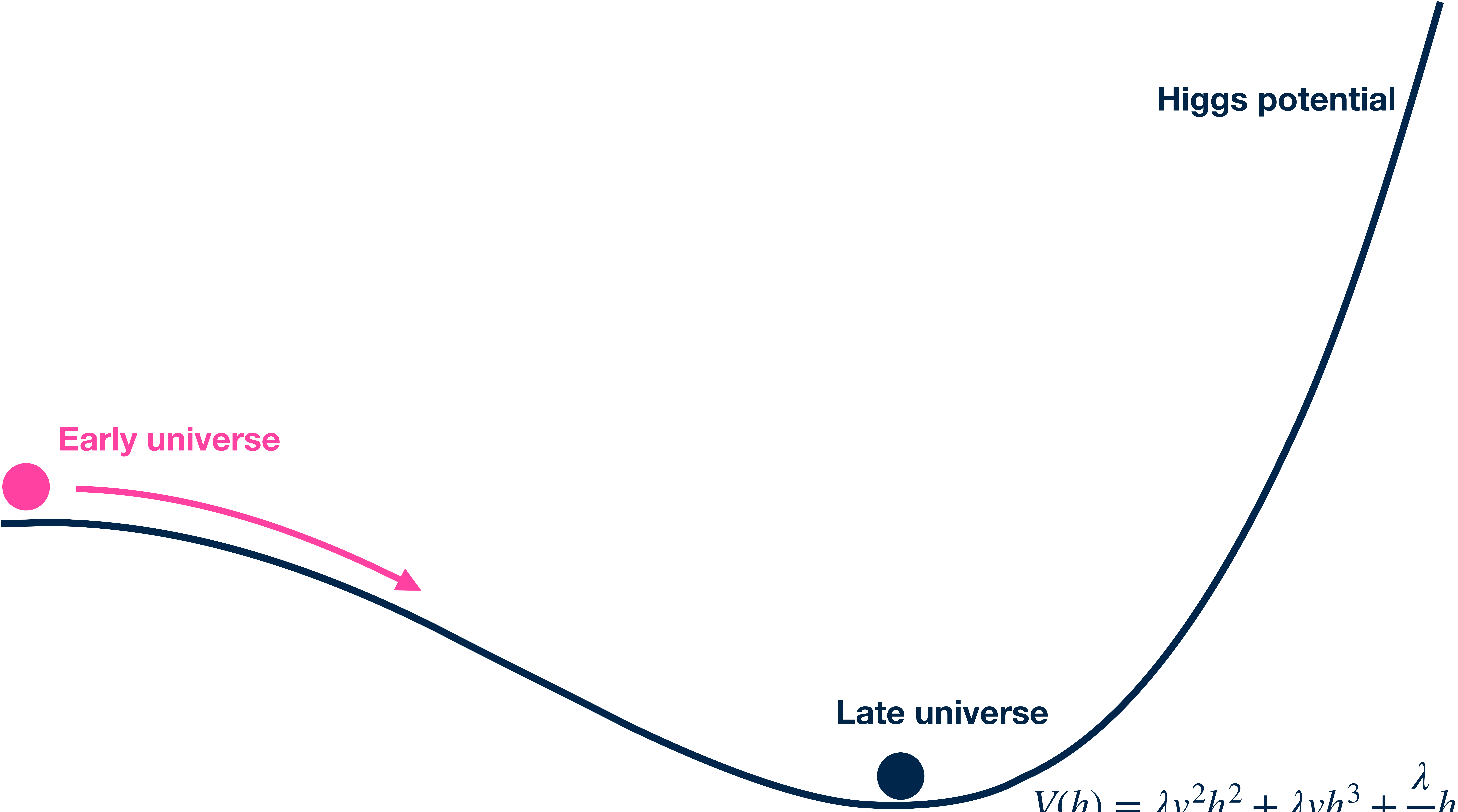

Automated data monitoring
5
Why HHH?
Looking to the future
1
4
2
How to search for HHH?
First HHH search results
3


Higgs potential

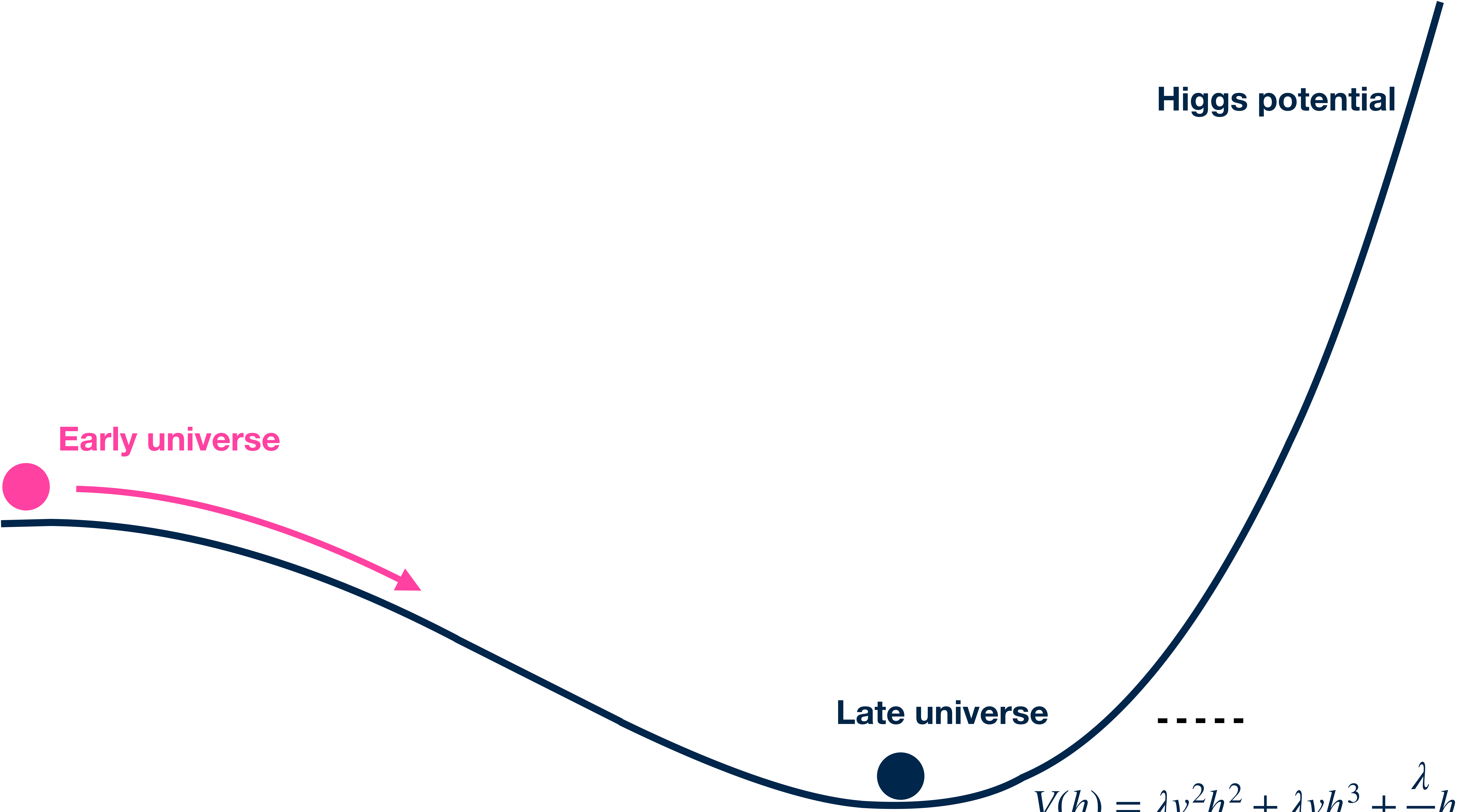

Early universe
Higgs potential


Higgs potential
Early universe

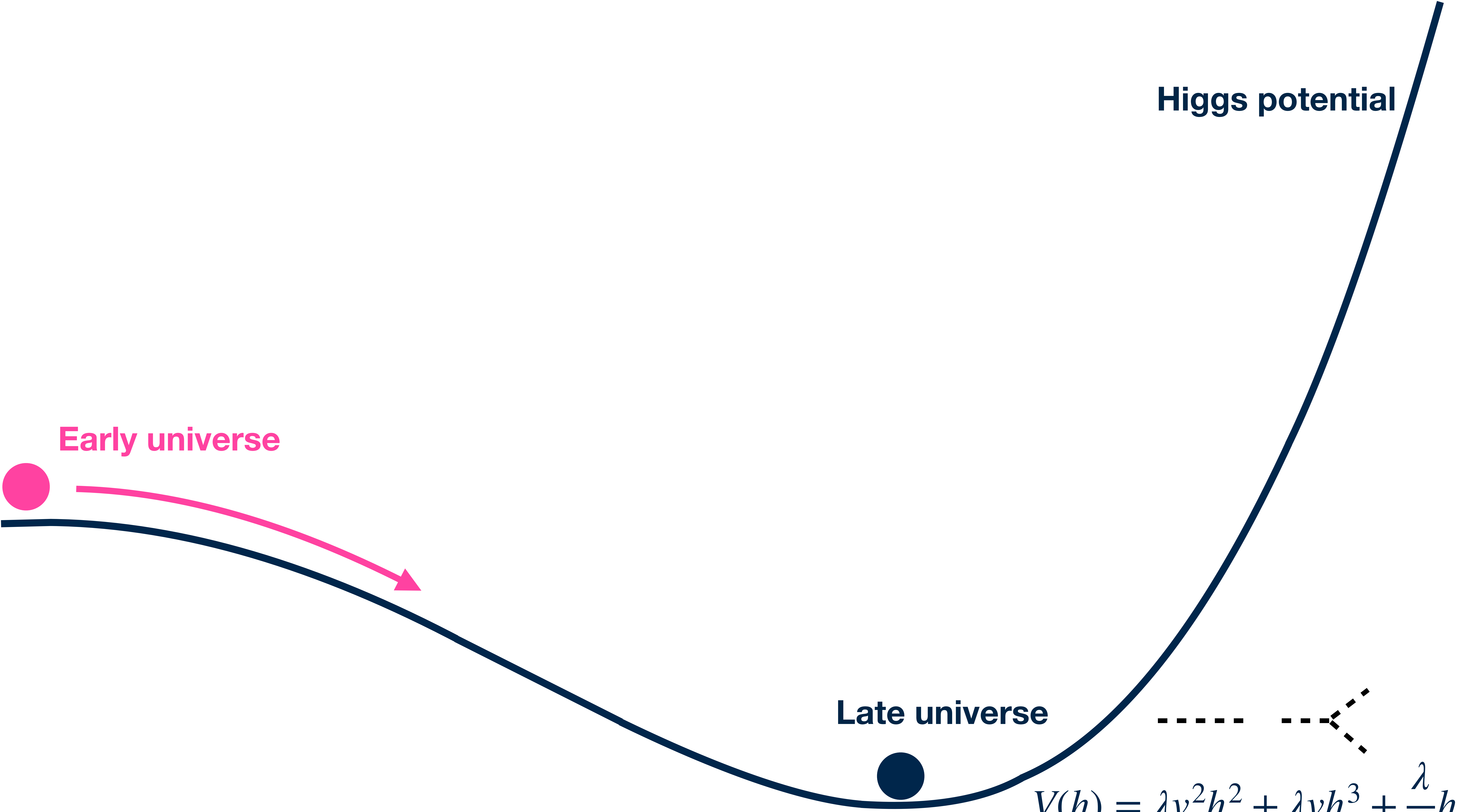

Higgs potential
Early universe
Late universe



$$V(h) = \lambda v^2 h^2 + \lambda v h^3 + \frac{\lambda}{4} h^4$$



$$V(h) = \lambda v^2 h^2 + \lambda v h^3 + \frac{\lambda}{4} h^4$$



$$V(h) = \lambda v^2 h^2 + \lambda v h^3 + \frac{\lambda}{4} h^4$$

$$V(h) = \lambda v^2 h^2 + \lambda v h^3 + \frac{\lambda}{4} h^4$$


Higgs potential
Early universe
Late universe
$V(h) = \lambda v^2 h^2 + \lambda v h^3 + \frac{\lambda}{4} h^4$


Higgs potential
Early universe
Late universe
$V(h) = \lambda v^2 h^2 + \lambda v h^3 + \frac{\lambda}{4} h^4$


Higgs potential
Early universe
$V(h) = \lambda v^2 h^2 + \kappa_3 \lambda v h^3 + \kappa_4 \frac{\lambda}{4} h^4$
Late universe
$V(h) = \lambda v^2 h^2 + \lambda v h^3 + \frac{\lambda}{4} h^4$

- **First order phase transitions cause more baryon production**

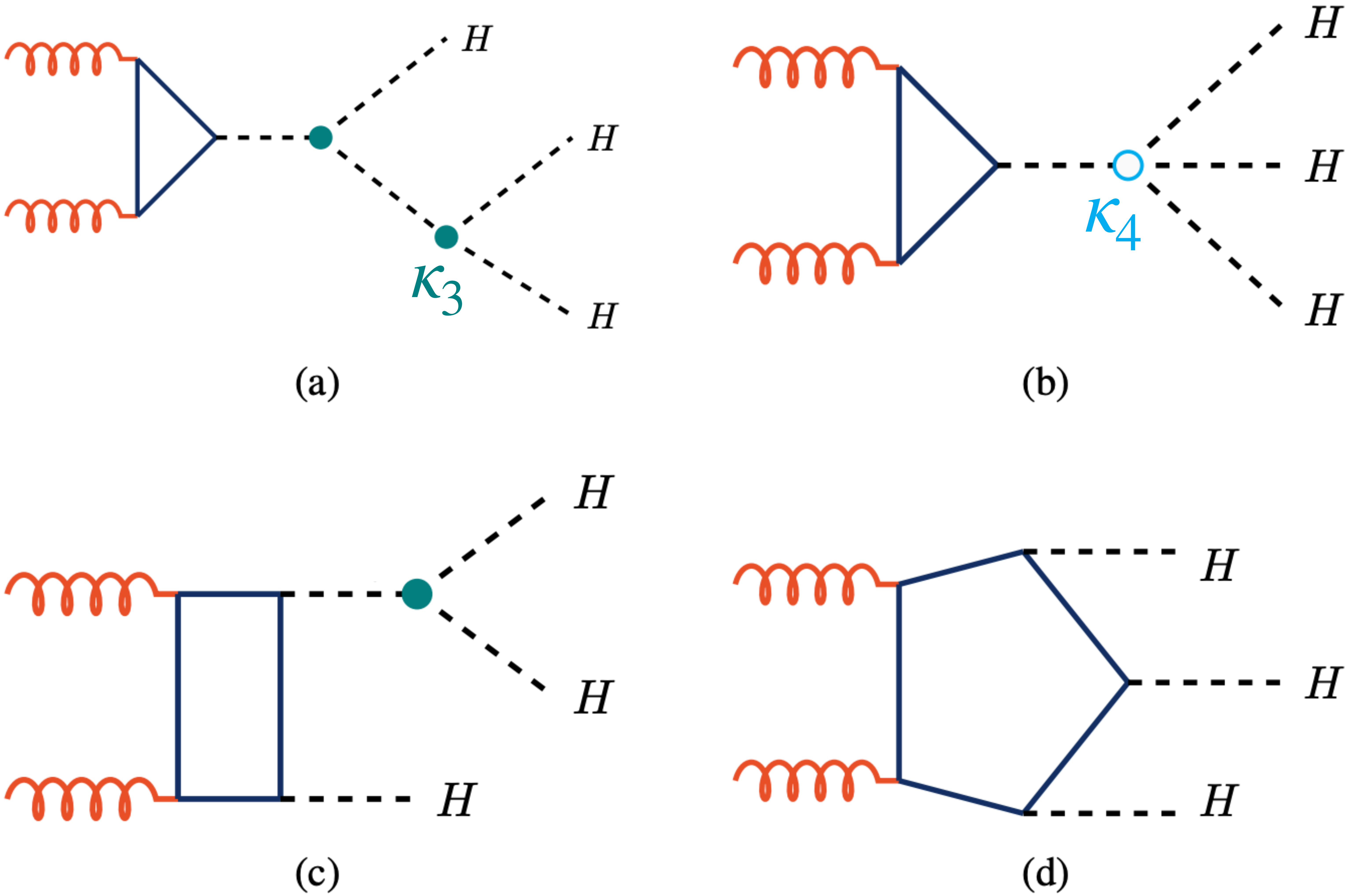
- First order phase transitions cause more baryon production
- CP violation can be introduced in the scalar potential

- First order phase transitions cause more baryon production
- CP violation can be introduced in the scalar potential
- New scalar particles could mediate to dark sector

- First order phase transitions cause more baryon production
- CP violation can be introduced in the scalar potential
- New scalar particles could mediate to dark sector

- Effect on the late universe: modifications to Higgs self-coupling κ_3, κ_4

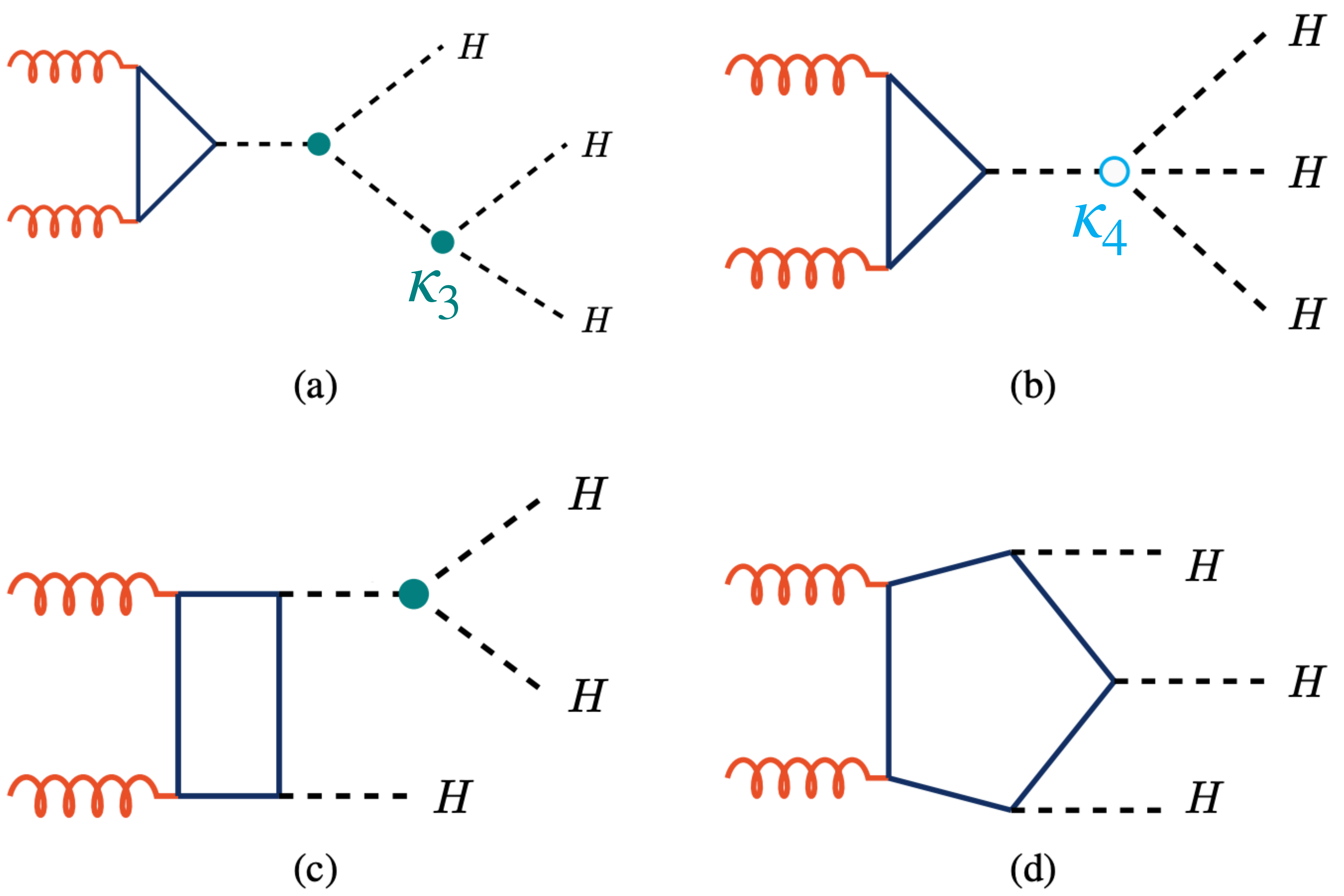
Why HHH?



$\sigma_{SM} = 0.079$ **fb**

Why HHH?

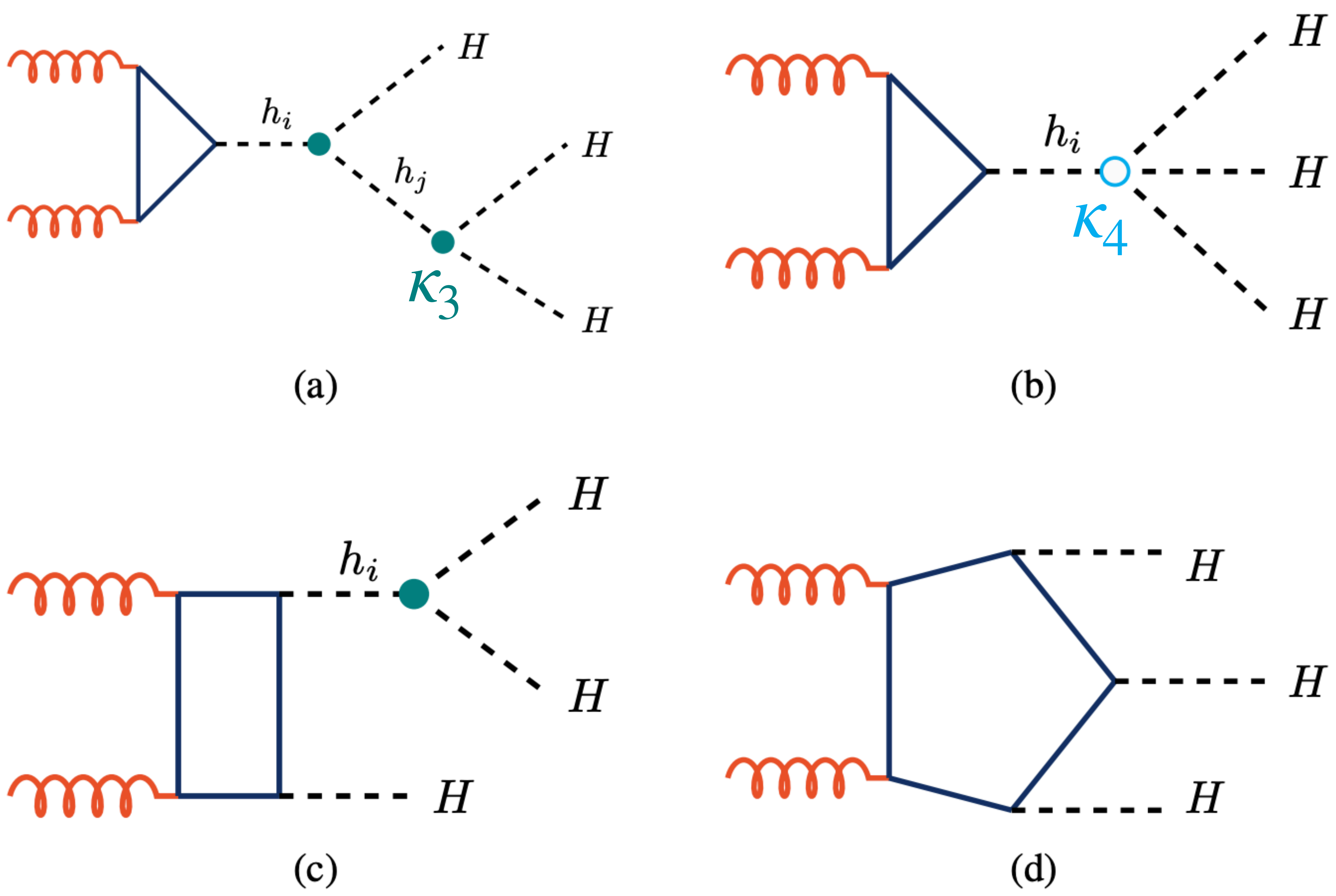
- **Measuring the shape of the Higgs potential probes electroweak baryogenesis**
  - Provides unique access to the quartic coupling κ_4
  - And can simultaneously limit κ_3



$\sigma_{SM} = 0.079$ fb

Why HHH?

- **Measuring the shape of the Higgs potential probes electroweak baryogenesis**
  - Provides unique access to the quartic coupling κ_4
  - And can simultaneously limit κ_3
- **BSM theories can predict enhanced HHH, even above HH production**
  - h_i, h_j can be new scalars
  - HHH production can be 100s fb when new scalars are present: [1908.08554], [2211.10557]
  - Additional scalars can provide sources of dark matter, CP violation: [2202.02954]
    - CP violation in the scalar potential and/or between scalars and DM.



$\sigma_{BSM} \approx 100$ **fb**

TRSM model

TRSM model

- **Two Real Singlet Model: [1908.08554]**

TRSM model

- **Two Real Singlet Model: [1908.08554]**

- **Add two new real singlets to SM:**
  - Both can have a VeV, which leads to HHH

$$\Phi = \begin{pmatrix} 0 \\ \frac{\phi_h + v}{\sqrt{2}} \end{pmatrix}, \quad S = \frac{\phi_S + v_S}{\sqrt{2}}, \quad X = \frac{\phi_X + v_X}{\sqrt{2}}$$

TRSM model

- **Two Real Singlet Model: [1908.08554]**

$$\Phi = \begin{pmatrix} 0 \\ \frac{\phi_h + v}{\sqrt{2}} \end{pmatrix}, \quad S = \frac{\phi_S + v_S}{\sqrt{2}}, \quad X = \frac{\phi_X + v_X}{\sqrt{2}}$$

- **Add two new real singlets to SM:**
  - Both can have a VeV, which leads to HHH

- **Discrete symmetries introduced:**
  - Note: the new particles are SM singlets, so Z_2 symmetries have nothing to do with FCNCs here!
  - More relevant for vacuum stability: bounding scalar potential from below.

$$\mathbb{Z}_2{}^S : \quad S \to -S\,, \; X \to X\,, \; \mathrm{SM} \to \mathrm{SM}\,,$$
$$\mathbb{Z}_2{}^X : \quad X \to -X\,, \; S \to S\,, \; \mathrm{SM} \to \mathrm{SM}$$

TRSM model

- **Two Real Singlet Model: [1908.08554]**

$$\Phi = \begin{pmatrix} 0 \\ \frac{\phi_h + v}{\sqrt{2}} \end{pmatrix}, \quad S = \frac{\phi_S + v_S}{\sqrt{2}}, \quad X = \frac{\phi_X + v_X}{\sqrt{2}}$$

- **Add two new real singlets to SM:**
  - Both can have a VeV, which leads to HHH

- **Discrete symmetries introduced:**
  - Note: the new particles are SM singlets, so Z_2 symmetries have nothing to do with FCNCs here!
  - More relevant for vacuum stability: bounding scalar potential from below.

$$\mathbb{Z}_2{}^S : \quad S \to -S\,, \; X \to X\,, \; \mathrm{SM} \to \mathrm{SM}\,,$$
$$\mathbb{Z}_2{}^X : \quad X \to -X\,, \; S \to S\,, \; \mathrm{SM} \to \mathrm{SM}$$

- **Seven free input parameters:**
  - 3 mixing angles
  - 2 unkown masses
  - 2 VeVs for new fields

TRSM model

- **Two Real Singlet Model: [1908.08554]**

$$\Phi = \begin{pmatrix} 0 \\ \frac{\phi_h + v}{\sqrt{2}} \end{pmatrix}, \quad S = \frac{\phi_S + v_S}{\sqrt{2}}, \quad X = \frac{\phi_X + v_X}{\sqrt{2}}$$

- **Add two new real singlets to SM:**
  - Both can have a VeV, which leads to HHH

- **Discrete symmetries introduced:**
  - Note: the new particles are SM singlets, so Z_2 symmetries have nothing to do with FCNCs here!
  - More relevant for vacuum stability: bounding scalar potential from below.

$$\mathbb{Z}_2{}^S: \quad S \to -S\,, \ X \to X\,, \ \mathrm{SM} \to \mathrm{SM}\,,$$
$$\mathbb{Z}_2{}^X: \quad X \to -X\,, \ S \to S\,, \ \mathrm{SM} \to \mathrm{SM}$$

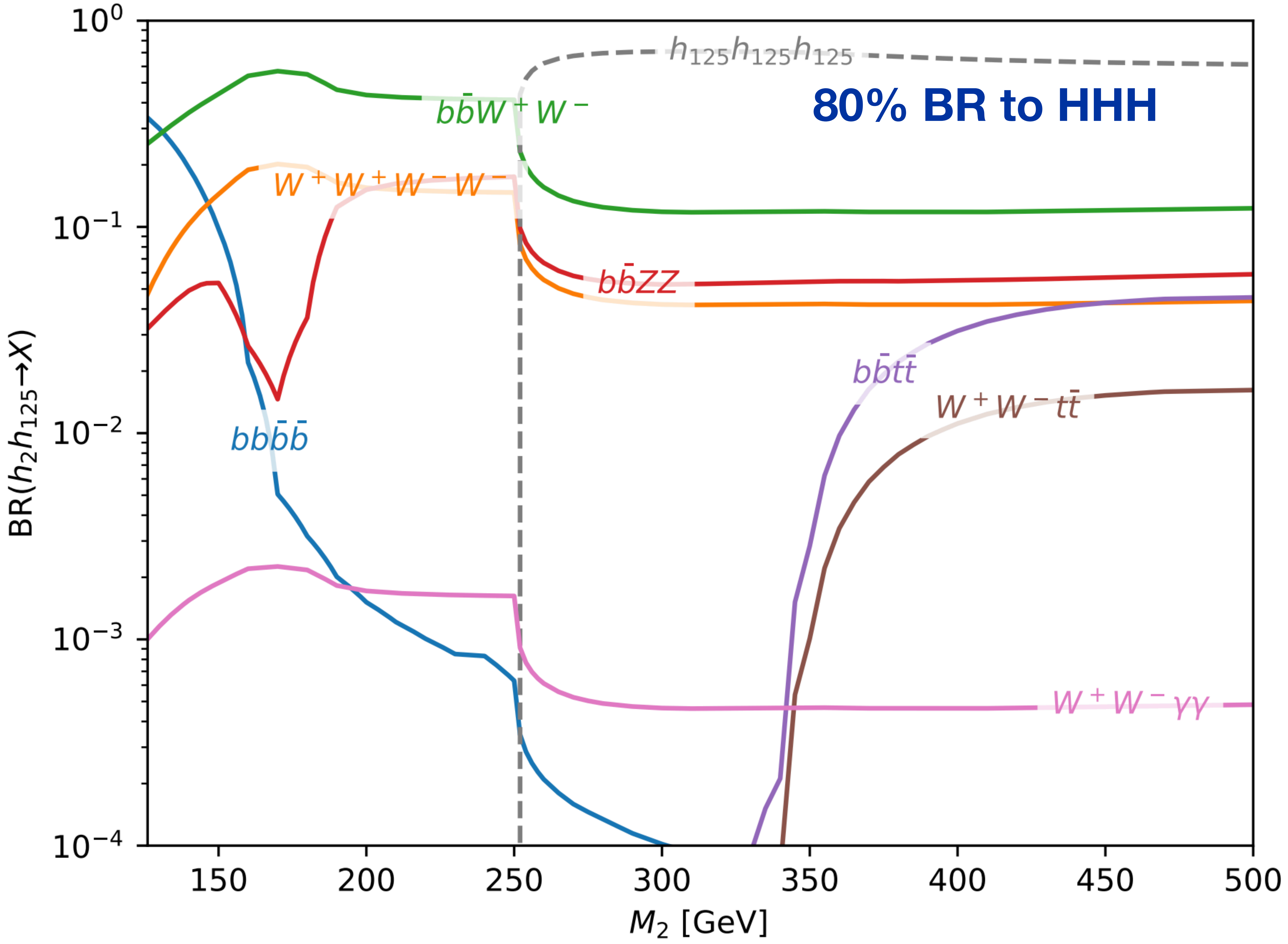
- **Seven free input parameters:**
  - 3 mixing angles
  - 2 unkown masses
  - 2 VeVs for new fields

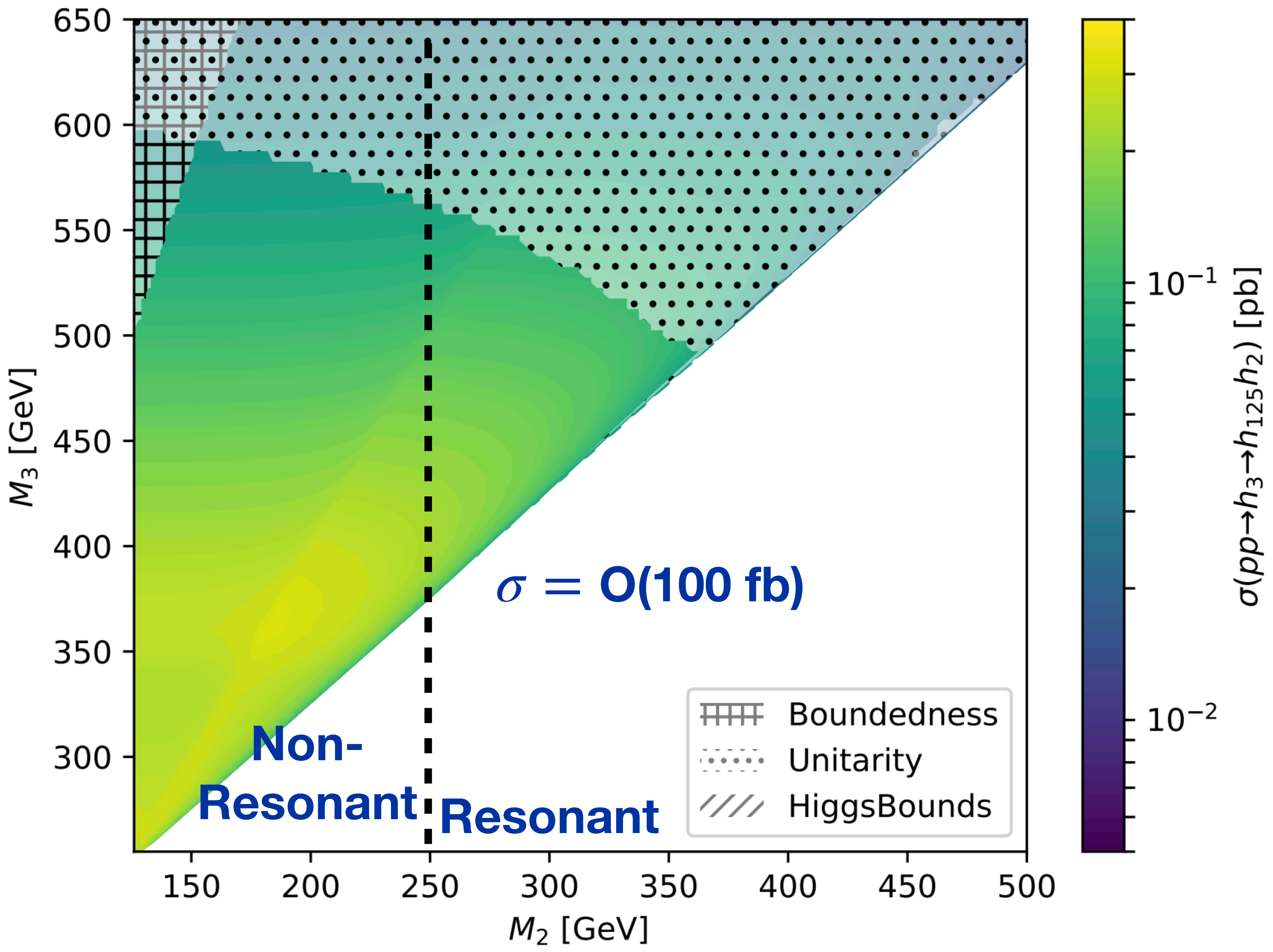
- **Scalar couplings, mixing from V:**
  - Remember, any can have VeV

$$V = \mu_\Phi^2 \Phi^\dagger \Phi + \lambda_\Phi (\Phi^\dagger \Phi)^2 + \mu_S^2 S^2 + \lambda_S S^4 + \mu_X^2 X^2 + \lambda_X X^4 + \lambda_{\Phi S} \Phi^\dagger \Phi S^2 + \lambda_{\Phi X} \Phi^\dagger \Phi X^2 + \lambda_{SX} S^2 X^2\,.$$

BSM searches

- In a **most favorable** scenario, $X \to SH$ is up to ~150 of fb and $Br(SH \to HHH)$ is up to 80%
- A large HHH signal which could be visible in Run 2 at the LHC





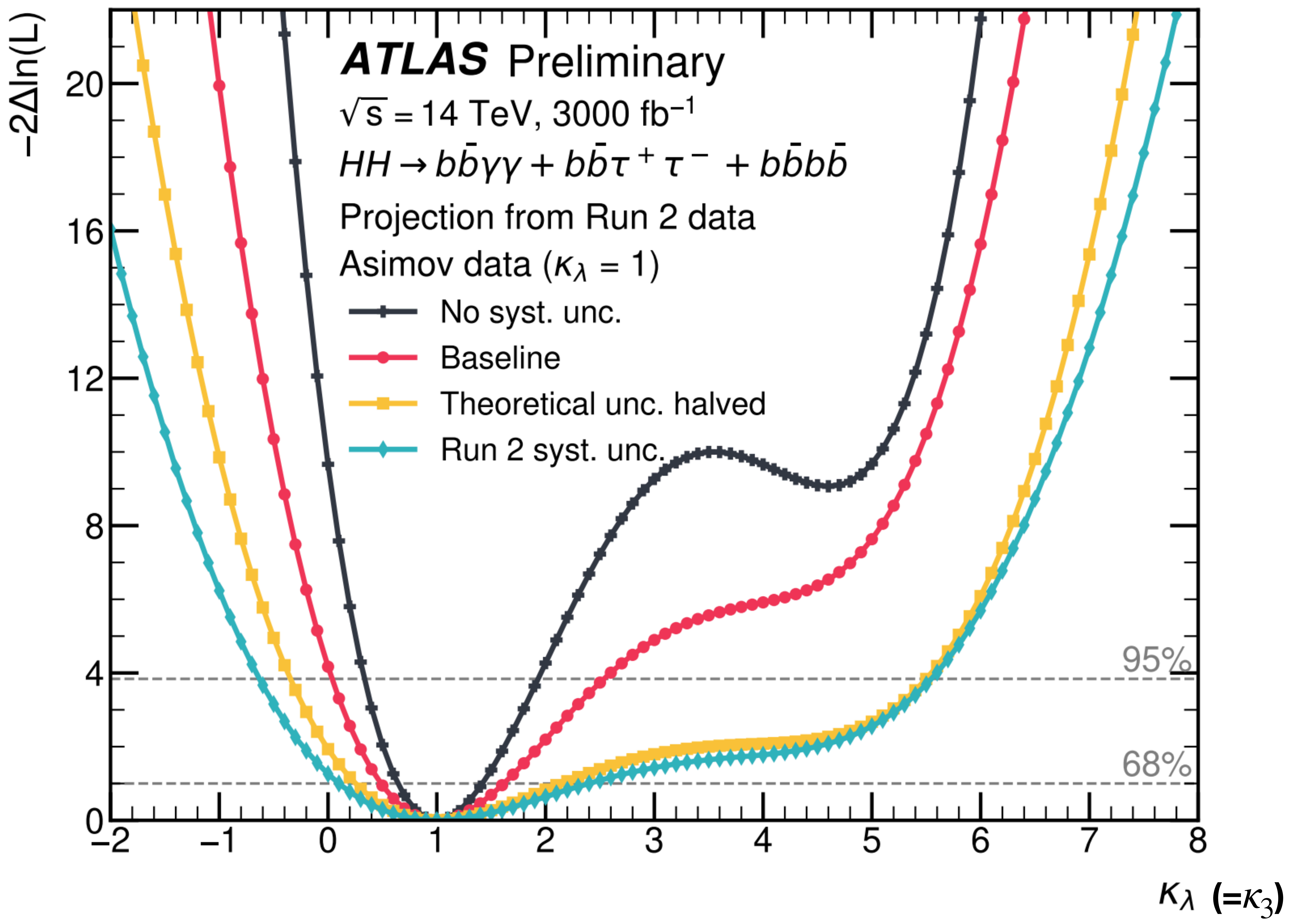
[1908.08554]

DM-CPV model

- **A simple model of dark matter and CP violation: [2202.02954]**
- **Similar phenomenology due to a complex scalar instead of 2 real scalars**
  - Also introduces a dark vector-like fermion χ
- **More free parameters and Z_2 symmetries are not imposed**
- **Produces cross sections for HHH up to 55fb. In most favorable scenarios kinematics are nearly identical to TRSM model.**
  - ATLAS tested generation of TRSM and DM-CPV signals and differences are negligible.

Measuring the Higgs potential

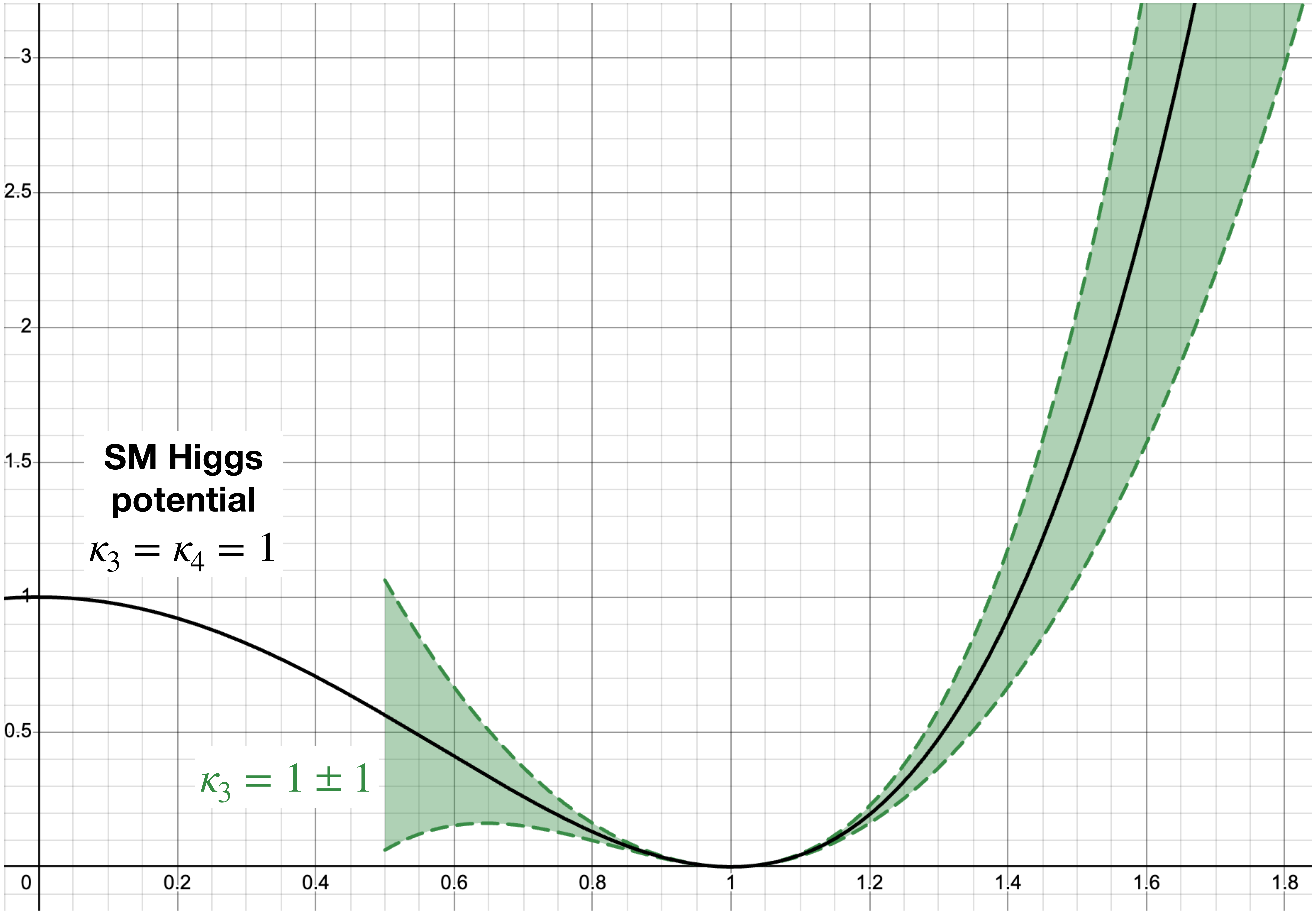
- **By 2040, a realistic goal of the HL-LHC program is to measure the tri-linear Higgs self-coupling to a precision $\kappa_3 = 1 \pm 1$**
  - ATLAS projections



Measuring the Higgs potential

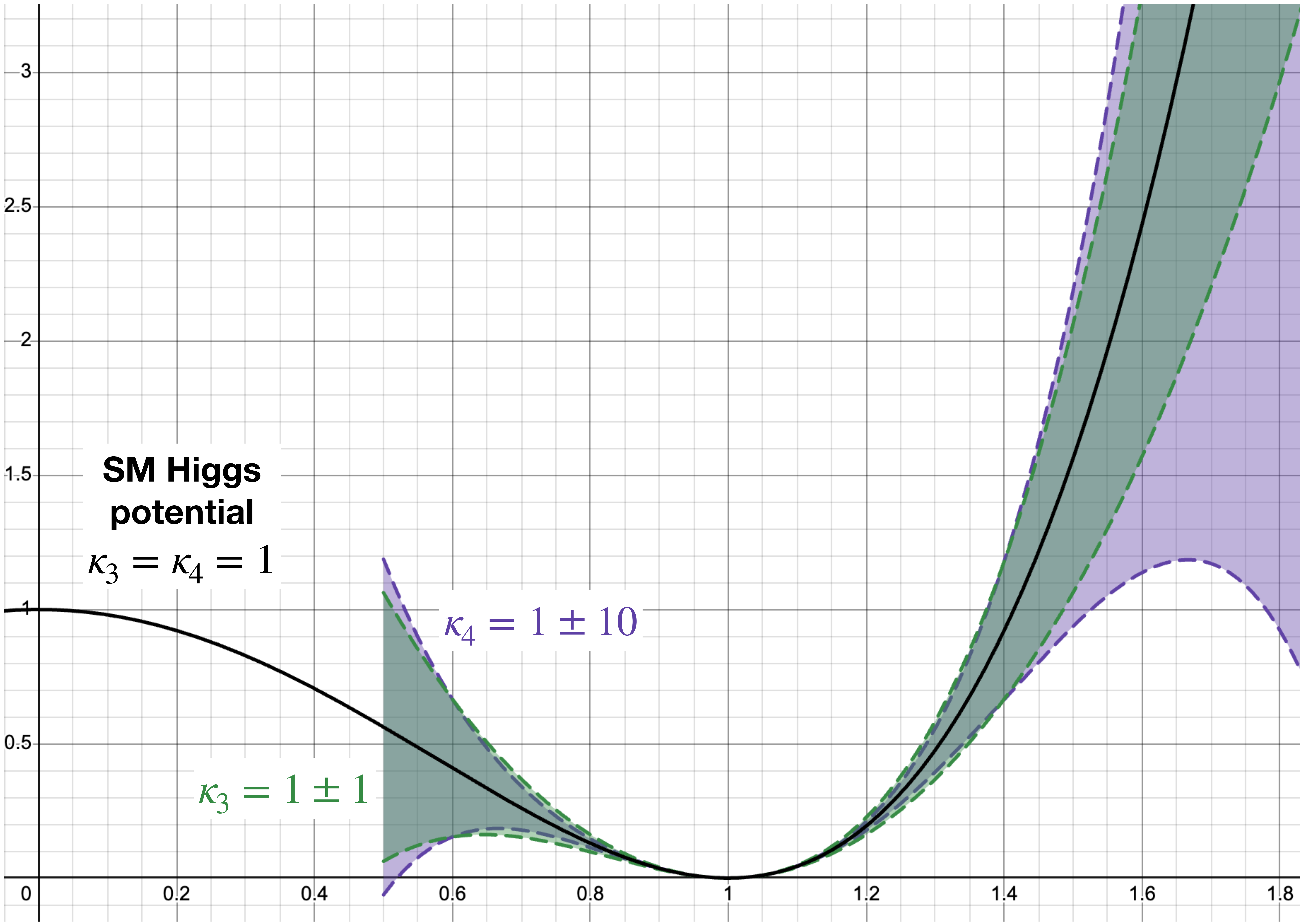
- **By 2040, a realistic goal of the HL-LHC program is to measure the tri-linear Higgs self-coupling to a precision $\kappa_3 = 1 \pm 1$**
  - ATLAS projections

- **We (experimentalists) claim we are measuring the shape of the Higgs potential! This is true, but…**
  - To *fully* measure the Higgs potential independent measurement of κ_4 is needed



Measuring the Higgs potential

- **By 2040, a realistic goal of the HL-LHC program is to measure the tri-linear Higgs self-coupling to a precision $\kappa_3 = 1 \pm 1$**
  - ATLAS projections
- **We (experimentalists) claim we are measuring the shape of the Higgs potential! This is true, but…**
  - To *fully* measure the Higgs potential independent measurement of κ_4 is needed
- **Luckily, the shape of the potential is less sensitive to κ_4, at least near the minimum.**
- **To obtain similar constraints on the potential shape, need $\kappa_4 = 1 \pm 10$**



Kevin Nelson

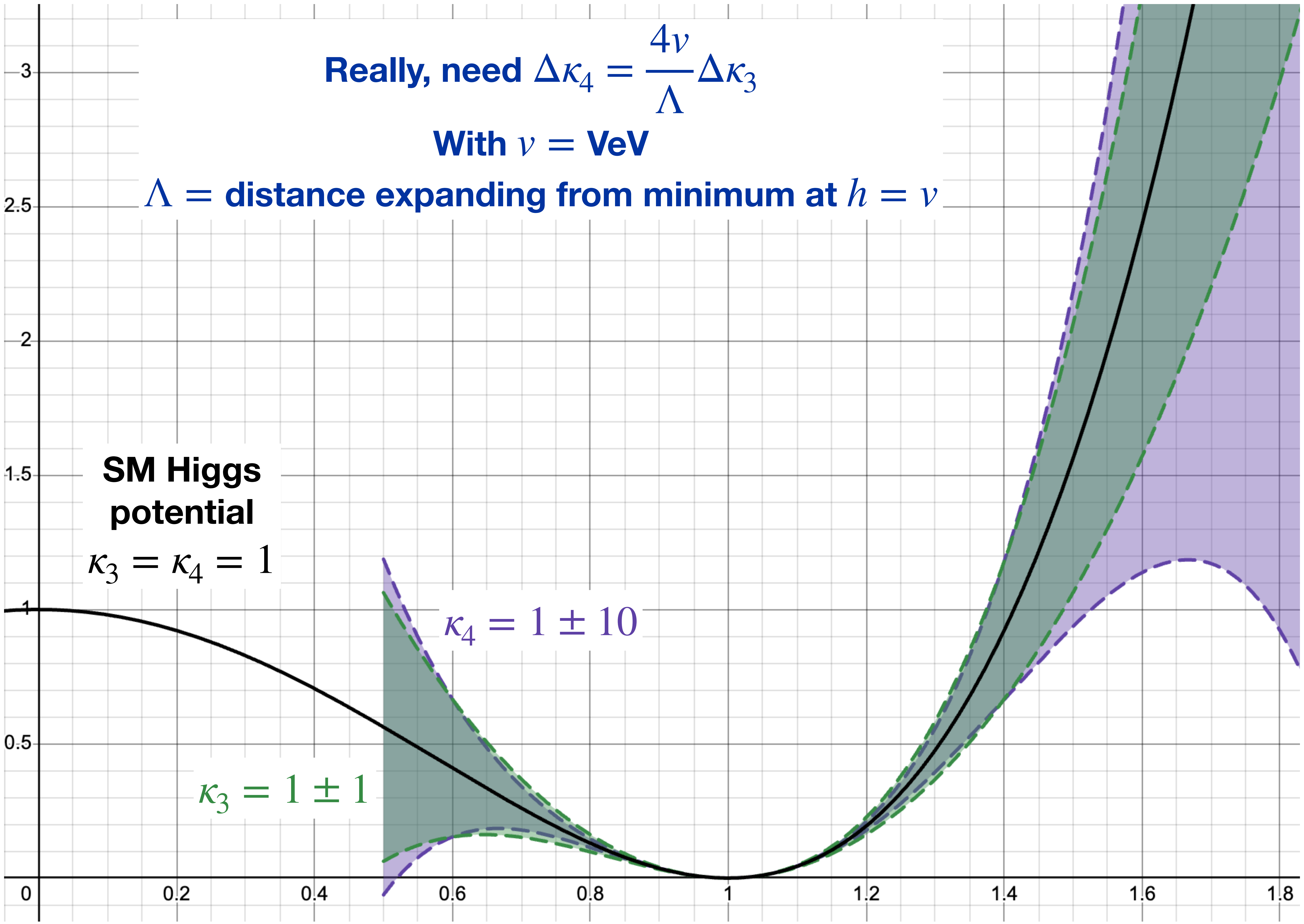
Measuring the Higgs potential

- **By 2040, a realistic goal of the HL-LHC program is to measure the tri-linear Higgs self-coupling to a precision $\kappa_3 = 1 \pm 1$**
  - [ATLAS projections]

- **We (experimentalists) claim we are measuring the shape of the Higgs potential! This is true, but…**
  - To *fully* measure the Higgs potential independent measurement of κ_4 is needed

- **Luckily, the shape of the potential is less sensitive to κ_4, at least near the minimum.**

- **To obtain similar constraints on the potential shape, need $\kappa_4 = 1 \pm 10$**



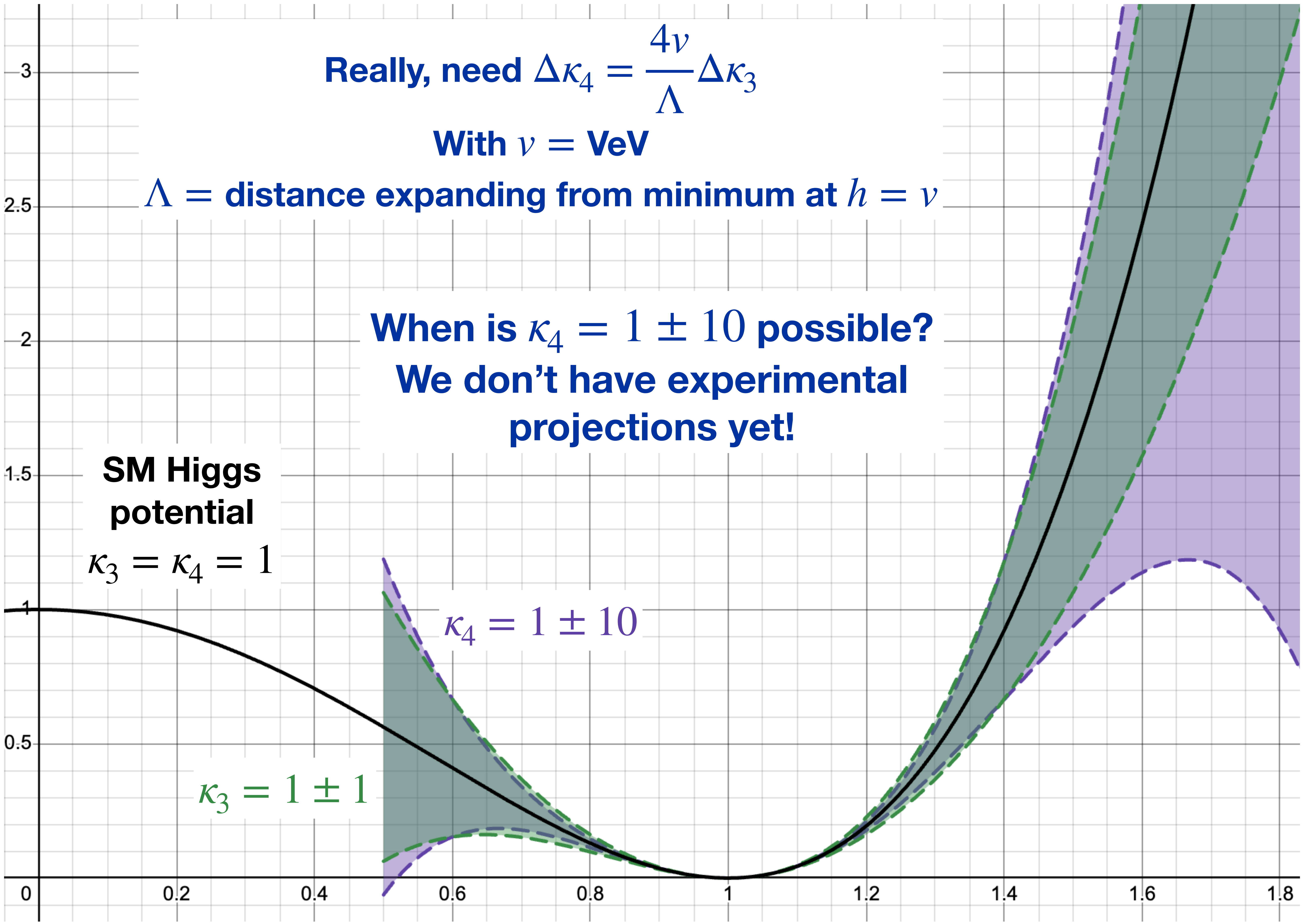
Measuring the Higgs potential

- **By 2040, a realistic goal of the HL-LHC program is to measure the tri-linear Higgs self-coupling to a precision $\kappa_3 = 1 \pm 1$**
  - ATLAS projections

- **We (experimentalists) claim we are measuring the shape of the Higgs potential! This is true, but…**
  - To *fully* measure the Higgs potential independent measurement of κ_4 is needed

- **Luckily, the shape of the potential is less sensitive to κ_4, at least near the minimum.**

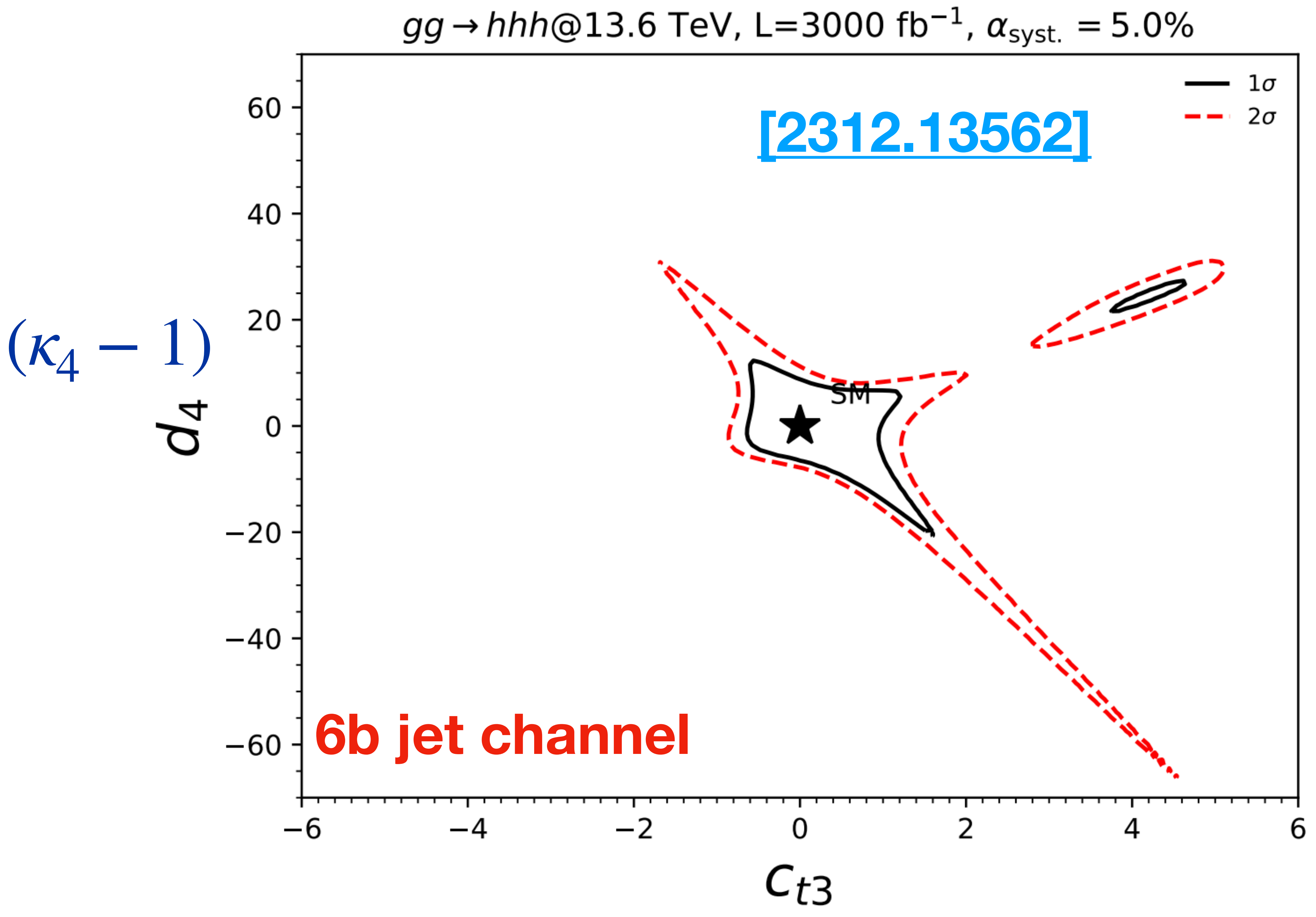
- **To obtain similar constraints on the potential shape, need $\kappa_4 = 1 \pm 10$**

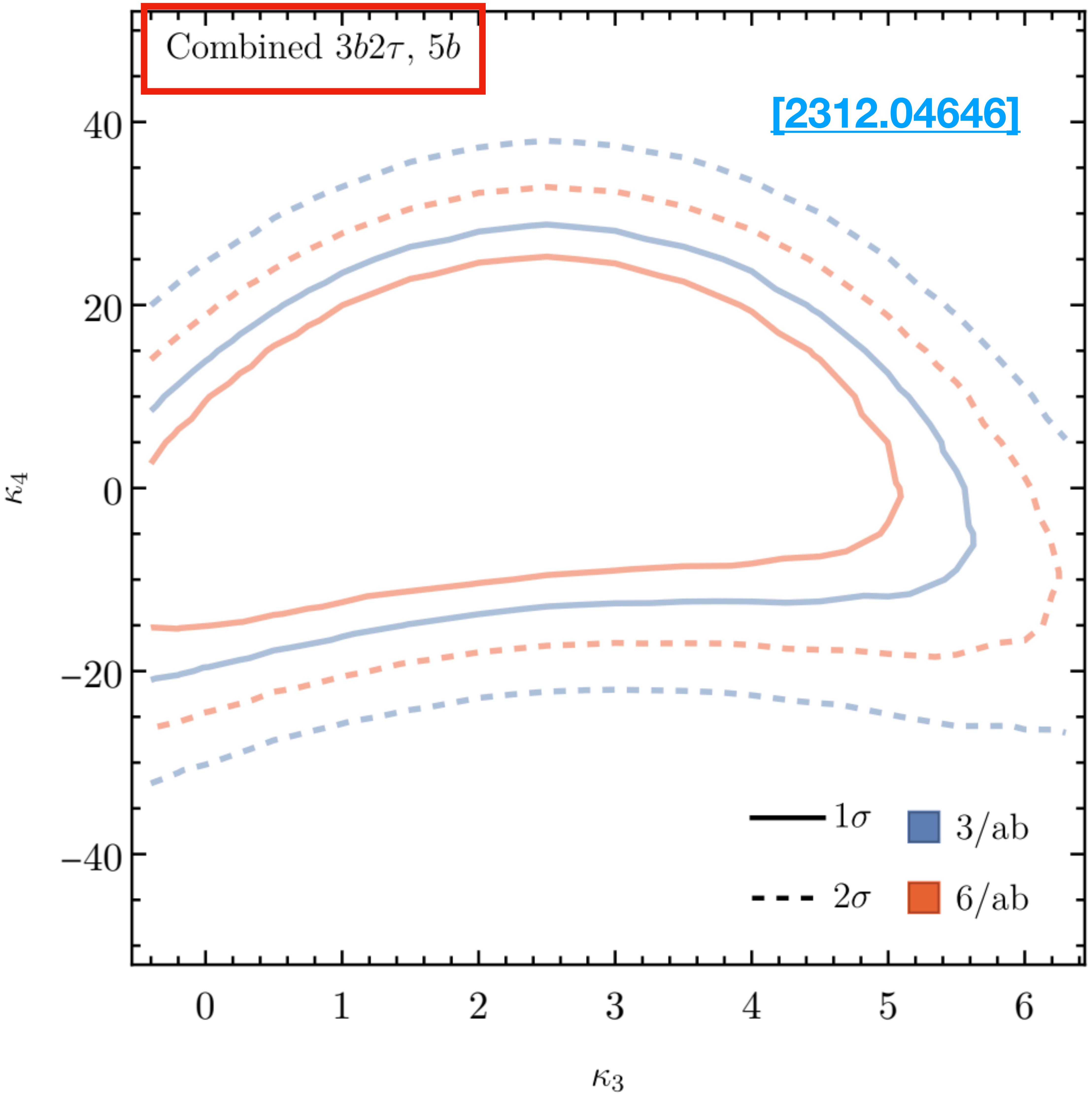


Kevin Nelson

(Phenomenological) HL-LHC Projections

- Some projections from the HHH whitepaper, [2312.13562], and [2312.04646] suggest $\kappa_4 = 1 \pm 30$ is possible in HL-LHC. 95% CL.
- Take them with a grain of salt, as projections are not yet based in a published analysis.

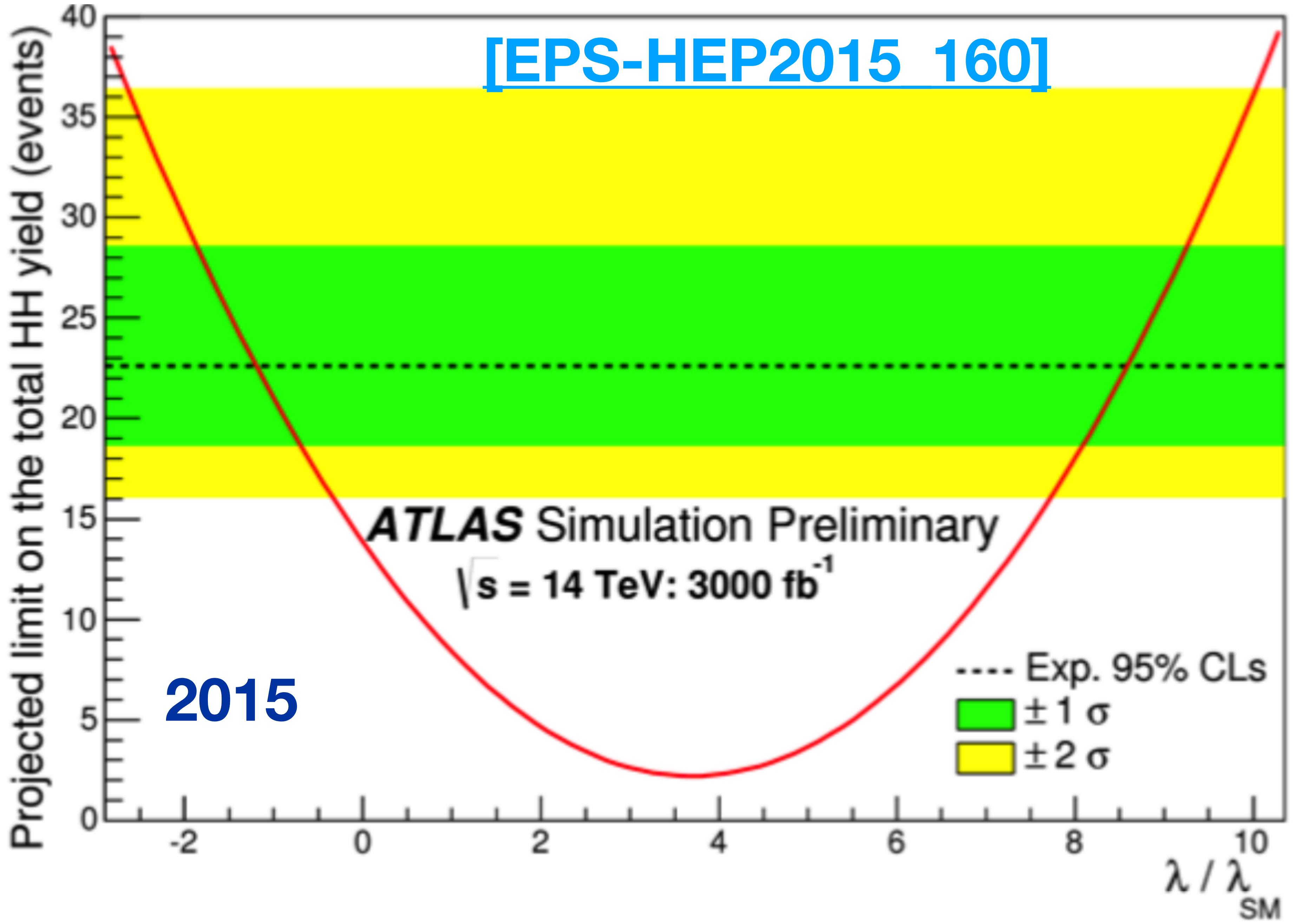




Performance of projections: HHistory

Performance of projections: HHistory

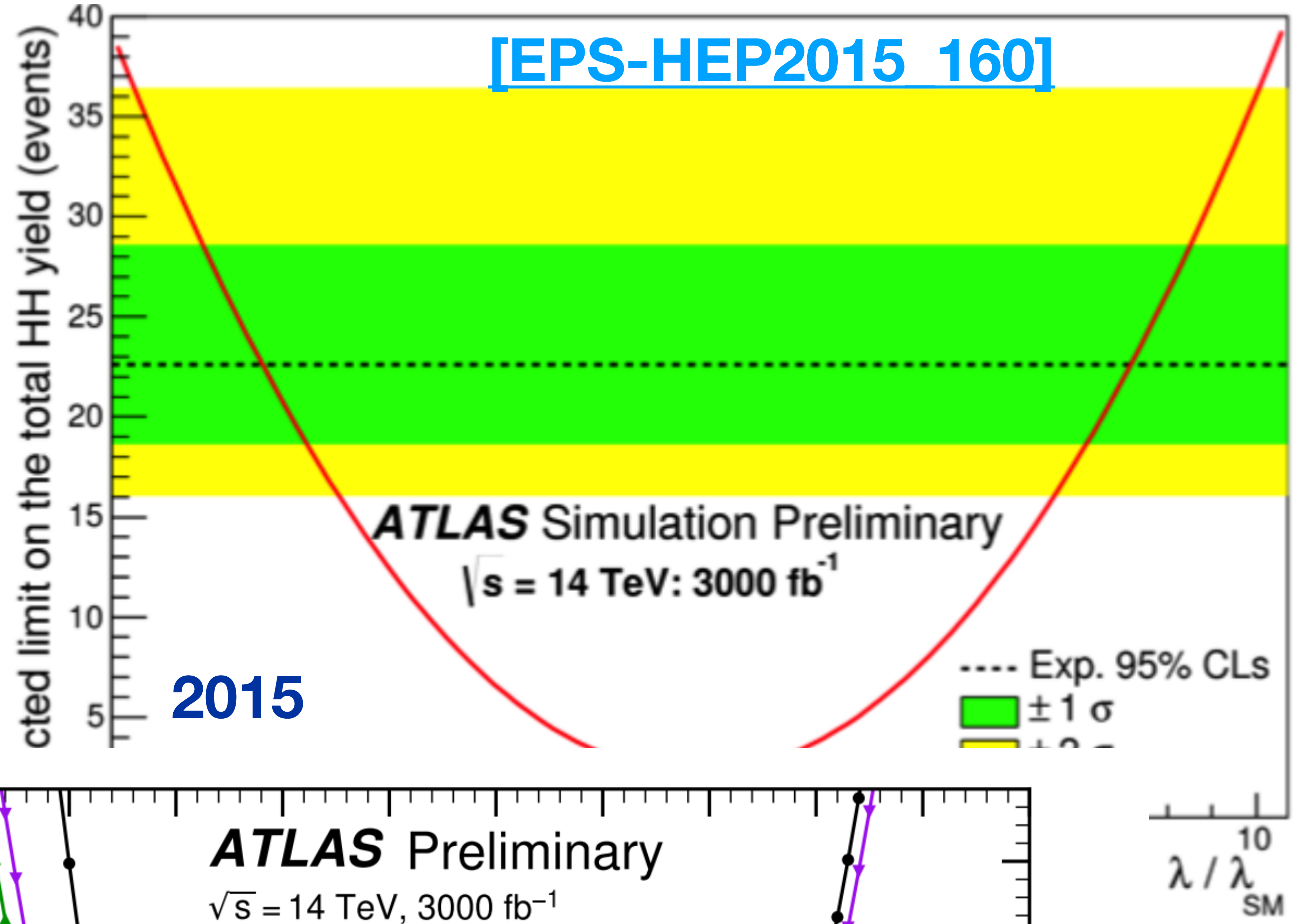
- In 2015, ATLAS projected that the $bb\gamma\gamma$ channel would be the most sensitive, and estimated limits on κ_3: $-1.3 < \kappa_3 < 8.7$ in HL-LHC

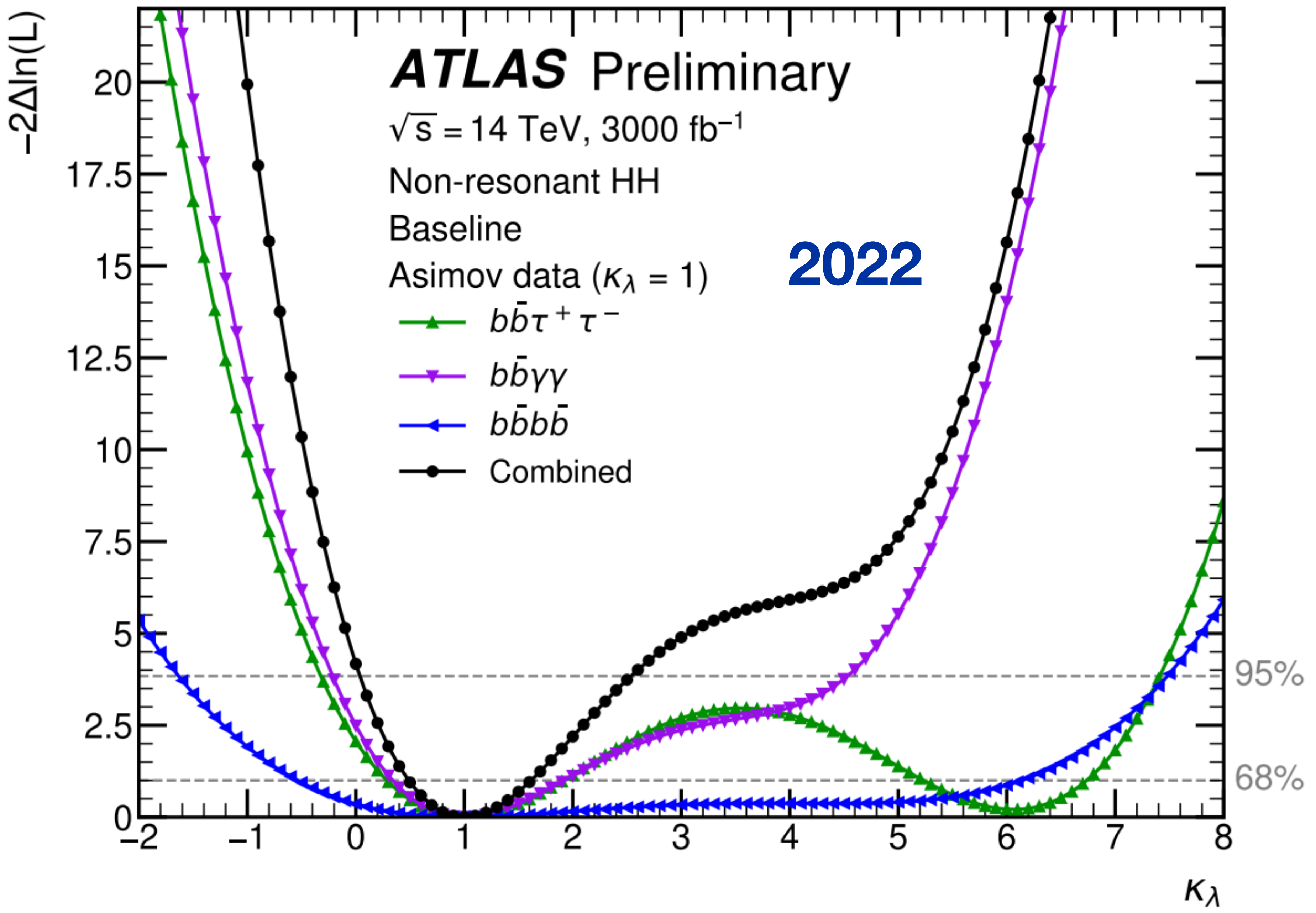


Performance of projections: HHistory

- **In 2015, ATLAS projected that the $bb\gamma\gamma$ channel would be the most sensitive, and estimated limits on κ_3: $-1.3 < \kappa_3 < 8.7$ in HL-LHC**

- **By 2022, the $bb\gamma\gamma$ projection had improved to $-0.2 < \kappa_3 < 4.6$**



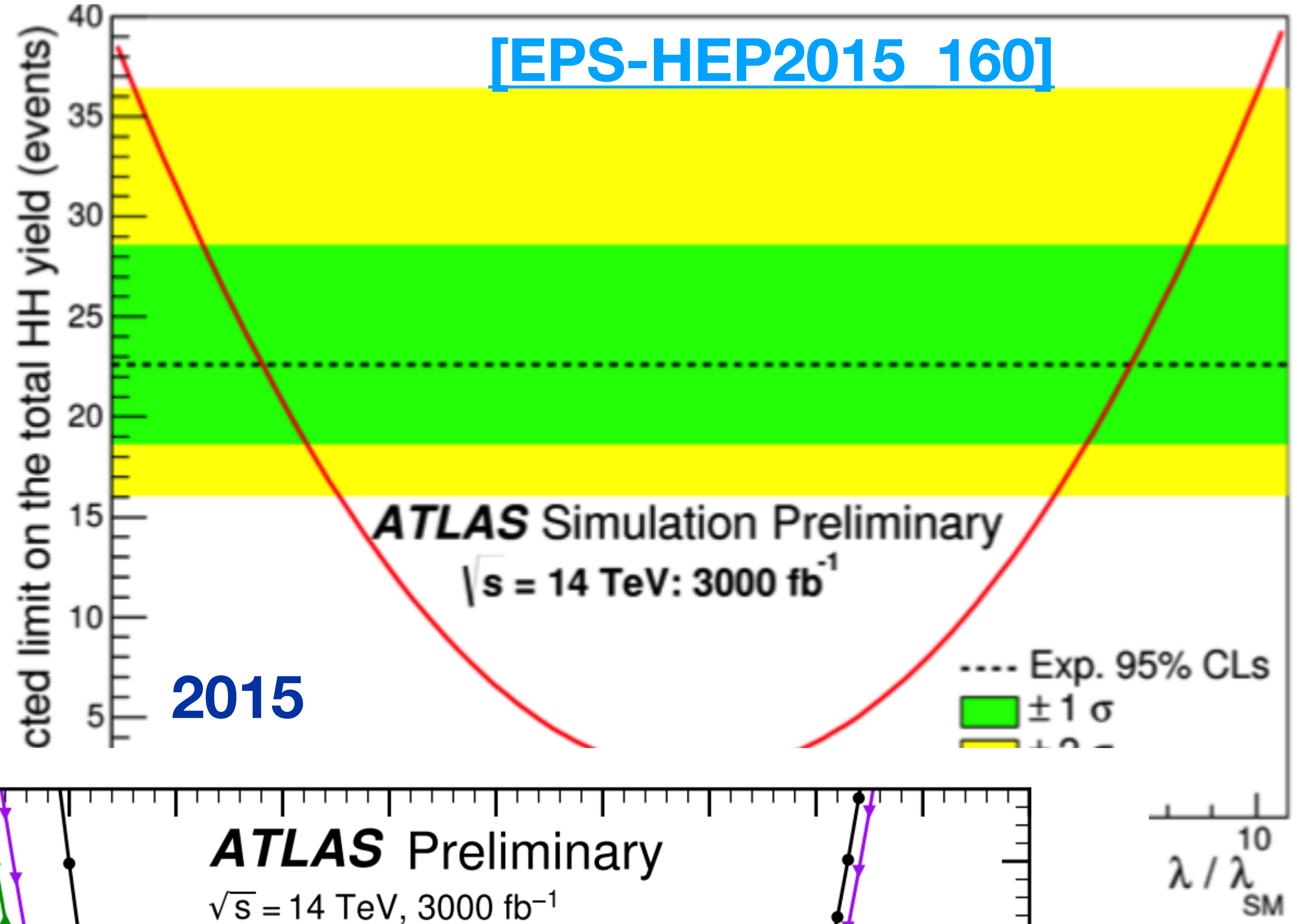


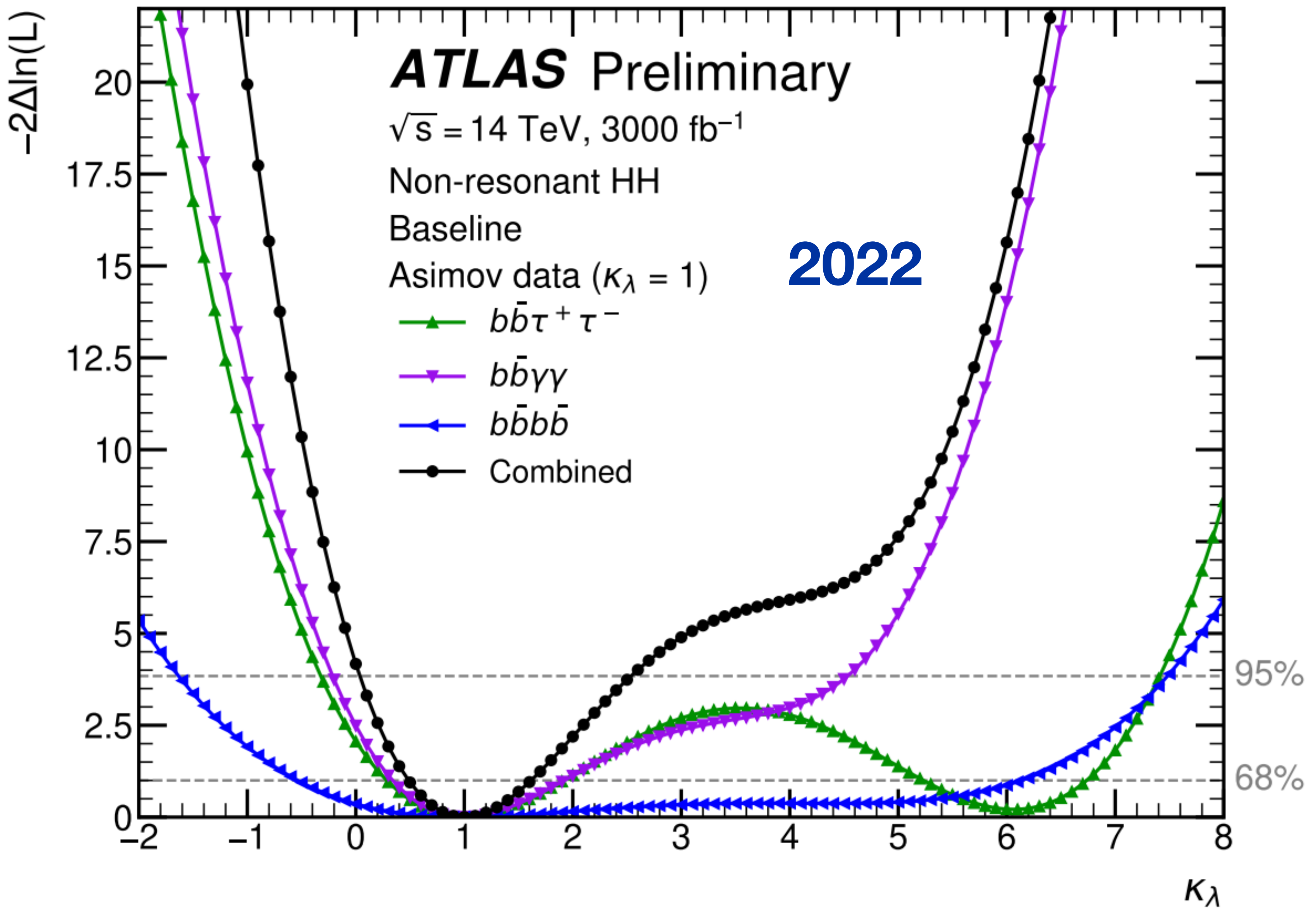
Performance of projections: HHistory

- **In 2015, ATLAS projected that the $bb\gamma\gamma$ channel would be the most sensitive, and estimated limits on κ_3: $-1.3 < \kappa_3 < 8.7$ in HL-LHC**

- **By 2022, the $bb\gamma\gamma$ projection had improved to $-0.2 < \kappa_3 < 4.6$**

- **Other channels in combination improved projection to** $0.1 < \kappa_3 < 2.5$





Performance of projections: HHistory

- **In 2015, ATLAS projected that the $bb\gamma\gamma$ channel would be the most sensitive, and estimated limits on κ_3: $-1.3 < \kappa_3 < 8.7$ in HL-LHC**

- **By 2022, the $bb\gamma\gamma$ projection had improved to $-0.2 < \kappa_3 < 4.6$**

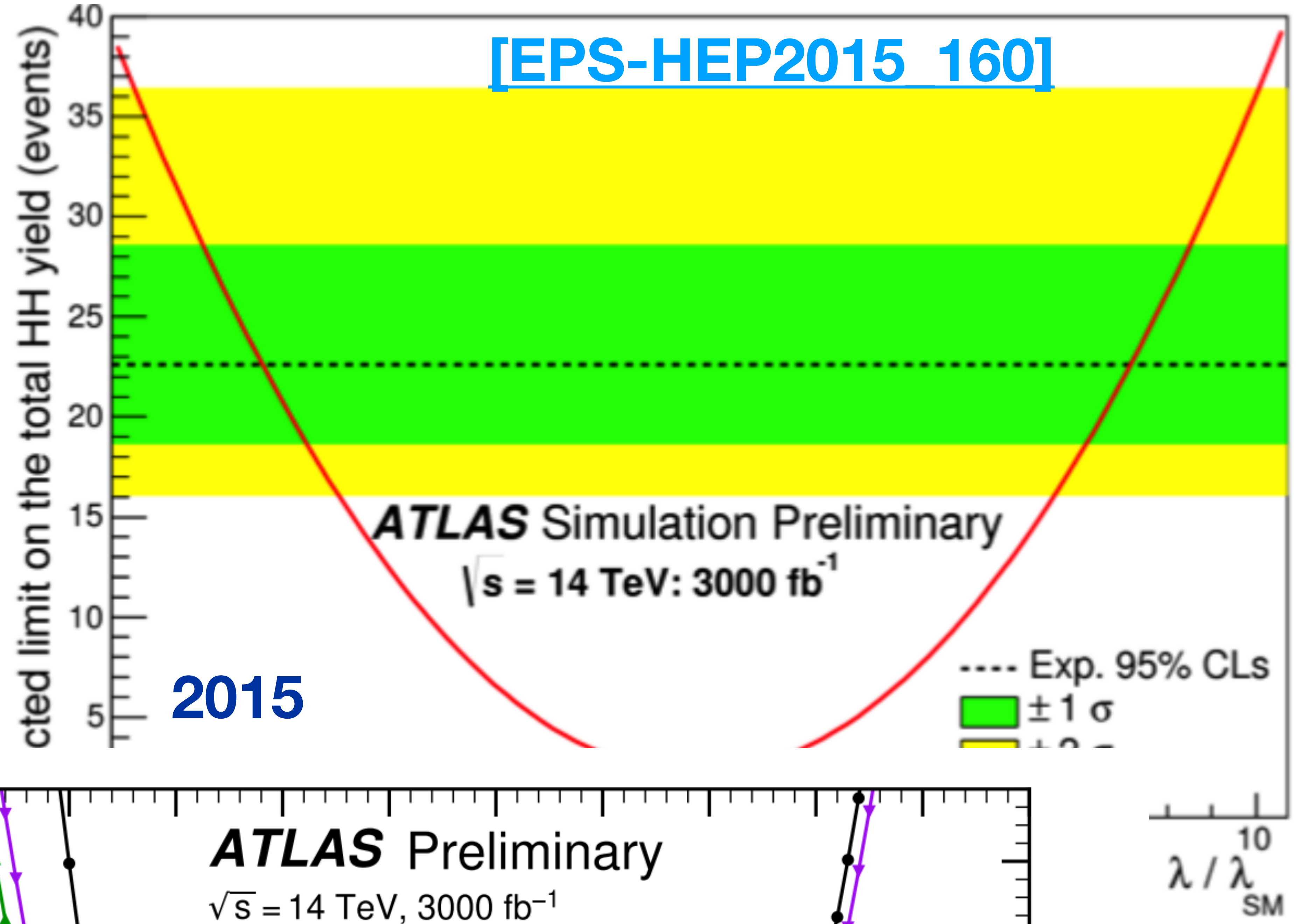
- **Other channels in combination improved projection to $0.1 < \kappa_3 < 2.5$**

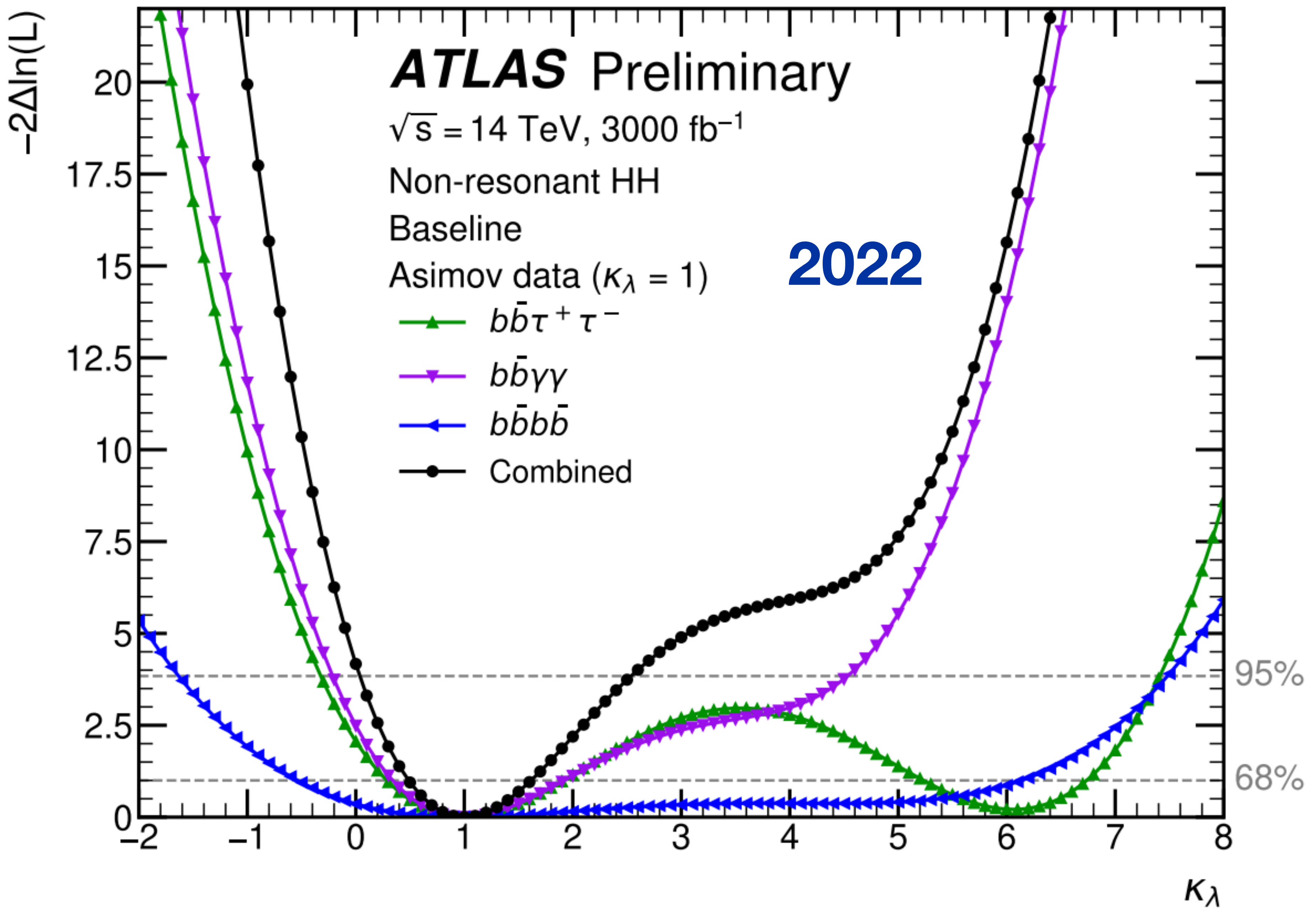
- **In just 7 years, phase space for projection reduces by 4x:**
  - Detector performance improvements
  - Improvements in b-tagging
  - New analysis ideas: AI/ML, categorization
  - Reduction of theoretical uncertainties
  - Exploration of more channels

I argue both can be improved

HHH is behind in these aspects

[EPS-HEP2015_160]





Performance of projections: HHistory

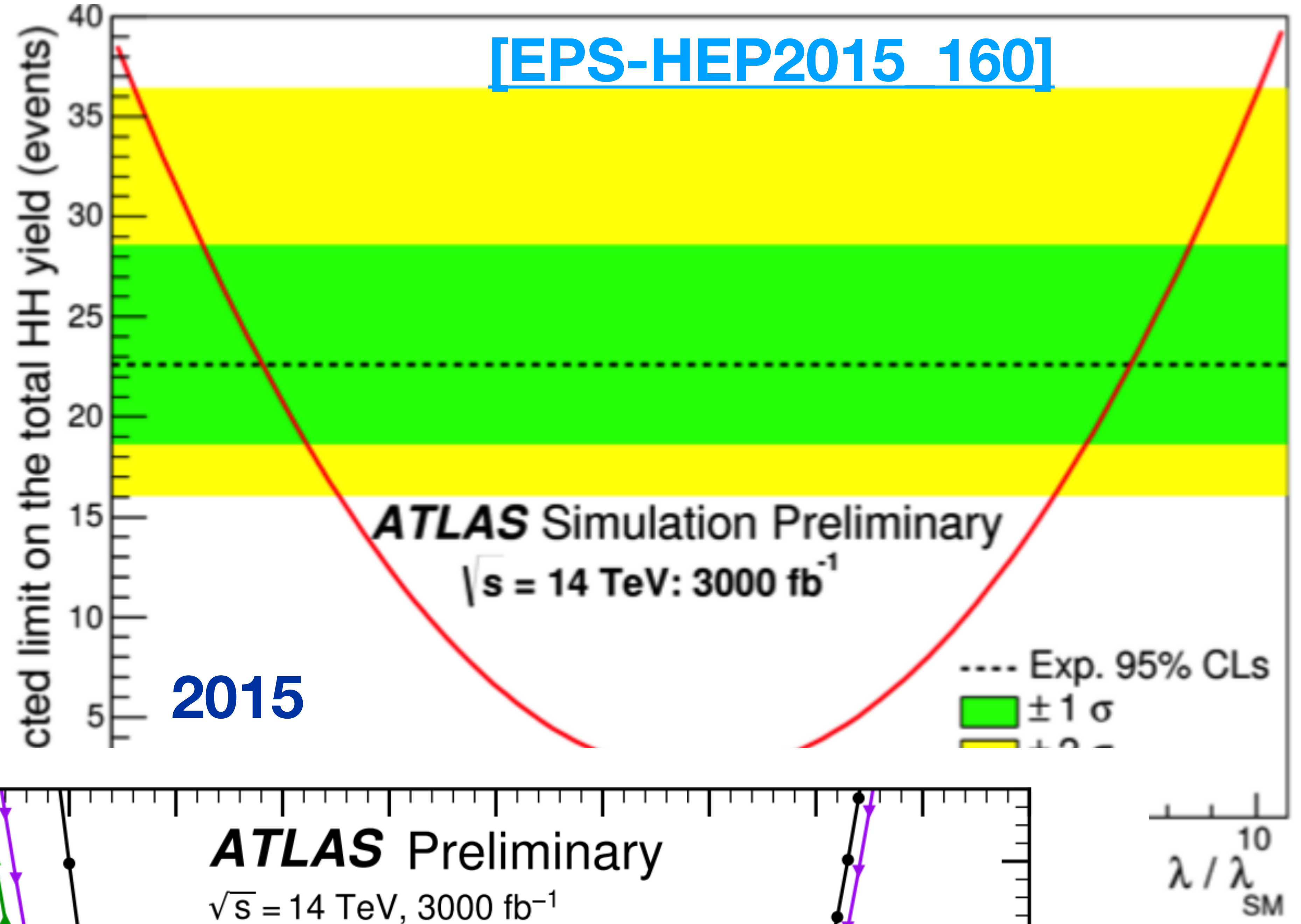
- **In 2015, ATLAS projected that the $bb\gamma\gamma$ channel would be the most sensitive, and estimated limits on κ_3: $-1.3 < \kappa_3 < 8.7$ in HL-LHC**
- **By 2022, the $bb\gamma\gamma$ projection had improved to $-0.2 < \kappa_3 < 4.6$**
- **Other channels in combination improved projection to $0.1 < \kappa_3 < 2.5$**
- **In just 7 years, phase space for projection reduces by 4x:**
  - Detector performance improvements
  - Improvements in b-tagging
  - New analysis ideas: AI/ML, categorization
  - Reduction of theoretical uncertainties
  - Exploration of more channels

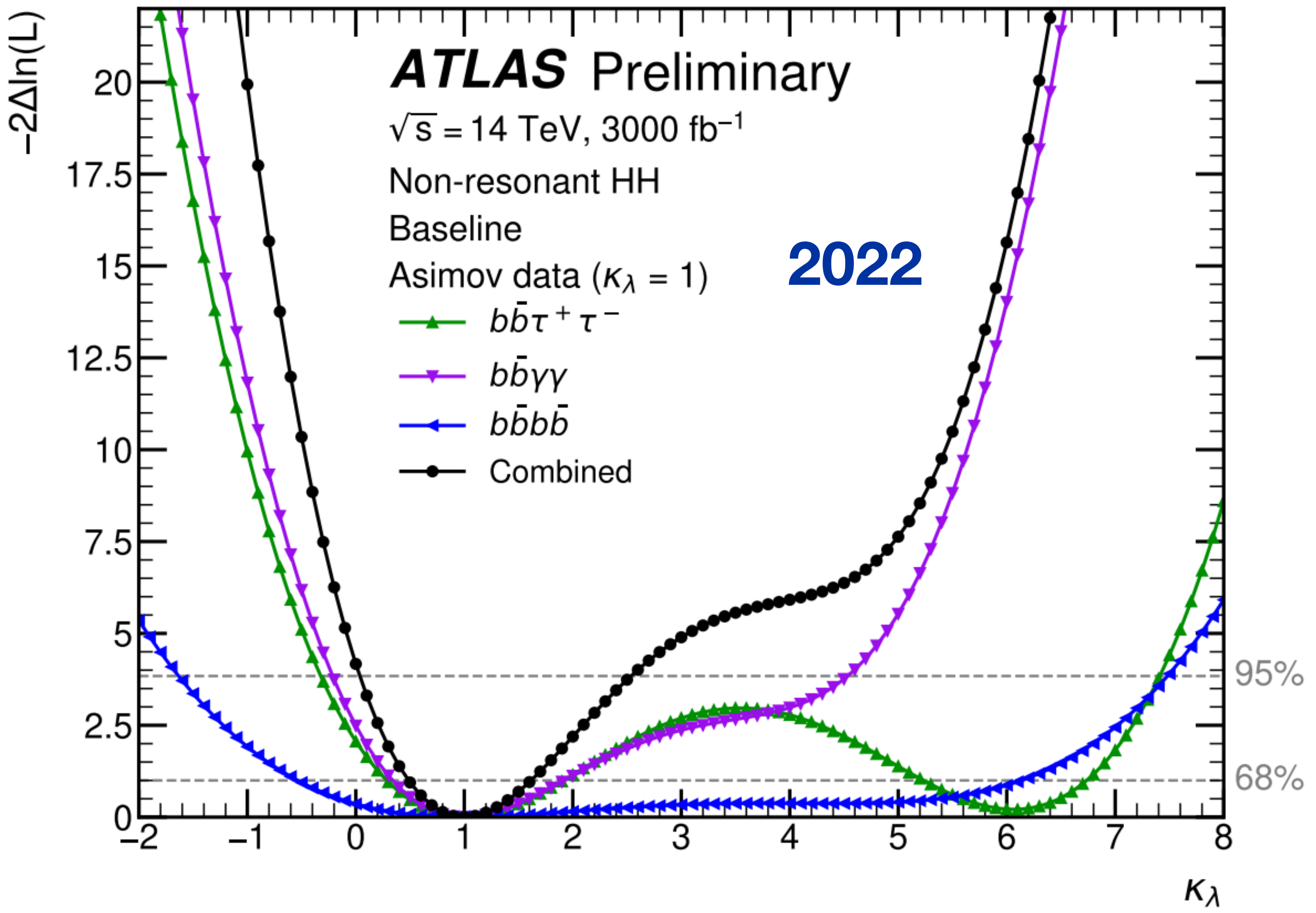
I argue both can be improved

HHH is behind in these aspects

- **Is $\kappa_4 = 1 \pm 10$ within the realm of possibility? Probably not, but we don't know yet. HHH not guaranteed to evolve like of HH projections**

[EPS-HEP2015_160]





Kevin Nelson


Automated data monitoring
5
Why HHH?
Looking to the future
1
4
2
How to search for HHH?
First HHH search results
3

Decay modes

Decay modes

- **The first question is what decay mode to search for!**

Decay modes

- **The first question is what decay mode to search for!**
- **There are more options for HHH than HH**

Reminder: HH branching fractions

	bb	WW	$\tau\tau$	ZZ	$\gamma\gamma$
bb	34%				
WW	25%	4.6%			
$\tau\tau$	7.4%	2.5%	0.39%		
ZZ	3.1%	1.2%	0.34%	0.076%	
$\gamma\gamma$	0.26%	0.10%	0.029%	0.013%	0.0005%

Decay modes

- **The first question is what decay mode to search for!**
- **There are more options for HHH than HH**
- **Combinatorics can make things interesting:**
  - For example, bbbbWW is *technically* dominant, but only some of those decays will have the utility of a lepton.
  - In terms of experimental signature 6*b* is dominant.

Reminder: HH branching fractions

	bb	WW	$\tau\tau$	ZZ	$\gamma\gamma$
bb	34%				
WW	25%	4.6%			
$\tau\tau$	7.4%	2.5%	0.39%		
ZZ	3.1%	1.2%	0.34%	0.076%	
$\gamma\gamma$	0.26%	0.10%	0.029%	0.013%	0.0005%

Decay modes

- **The first question is what decay mode to search for!**
- **There are more options for HHH than HH**
- **Combinatorics can make things interesting:**
  - For example, bbbbWW is *technically* dominant, but only some of those decays will have the utility of a lepton.
  - In terms of experimental signature 6*b* is dominant.
- **Selected HHH decay modes and branching fractions:**

Decay Mode	Branching Fraction
bbbbbb	19.5%
bbbb+1 lepton*	15.2%
bbbb+yy	0.2%
bbbb+taus(had)	2.5%
bb+SS leptons	1.9%

Reminder: HH branching fractions

	bb	WW	$\tau\tau$	ZZ	$\gamma\gamma$
bb	34%				
WW	25%	4.6%			
$\tau\tau$	7.4%	2.5%	0.39%		
ZZ	3.1%	1.2%	0.34%	0.076%	
$\gamma\gamma$	0.26%	0.10%	0.029%	0.013%	0.0005%

Decay modes

- **The first question is what decay mode to search for!**
- **There are more options for HHH than HH**
- **Combinatorics can make things interesting:**
  - For example, bbbbWW is *technically* dominant, but only some of those decays will have the utility of a lepton.
  - In terms of experimental signature 6*b* is dominant.
- **Selected HHH decay modes and branching fractions:**

Decay Mode	Branching Fraction
bbbbbb	19.5%
bbbb+1 lepton*	15.2%
bbbb+yy	0.2%
bbbb+taus(had)	2.5%
bb+SS leptons	1.9%

Ordered by likely sensitivity to SM HHH

Reminder: HH branching fractions

	bb	WW	$\tau\tau$	ZZ	$\gamma\gamma$
bb	34%				
WW	25%	4.6%			
$\tau\tau$	7.4%	2.5%	0.39%		
ZZ	3.1%	1.2%	0.34%	0.076%	
$\gamma\gamma$	0.26%	0.10%	0.029%	0.013%	0.0005%

Decay modes

- **The first question is what decay mode to search for!**
- **There are more options for HHH than HH**
- **Combinatorics can make things interesting:**
  - For example, bbbbWW is *technically* dominant, but only some of those decays will have the utility of a lepton.
  - In terms of experimental signature 6*b* is dominant.
- **Selected HHH decay modes and branching fractions:**

Decay Mode	Branching Fraction
bbbbbb	19.5%
bbbb+1 lepton*	15.2%
bbbb+yy	0.2%
bbbb+taus(had)	2.5%
bb+SS leptons	1.9%

***An understudied mode**

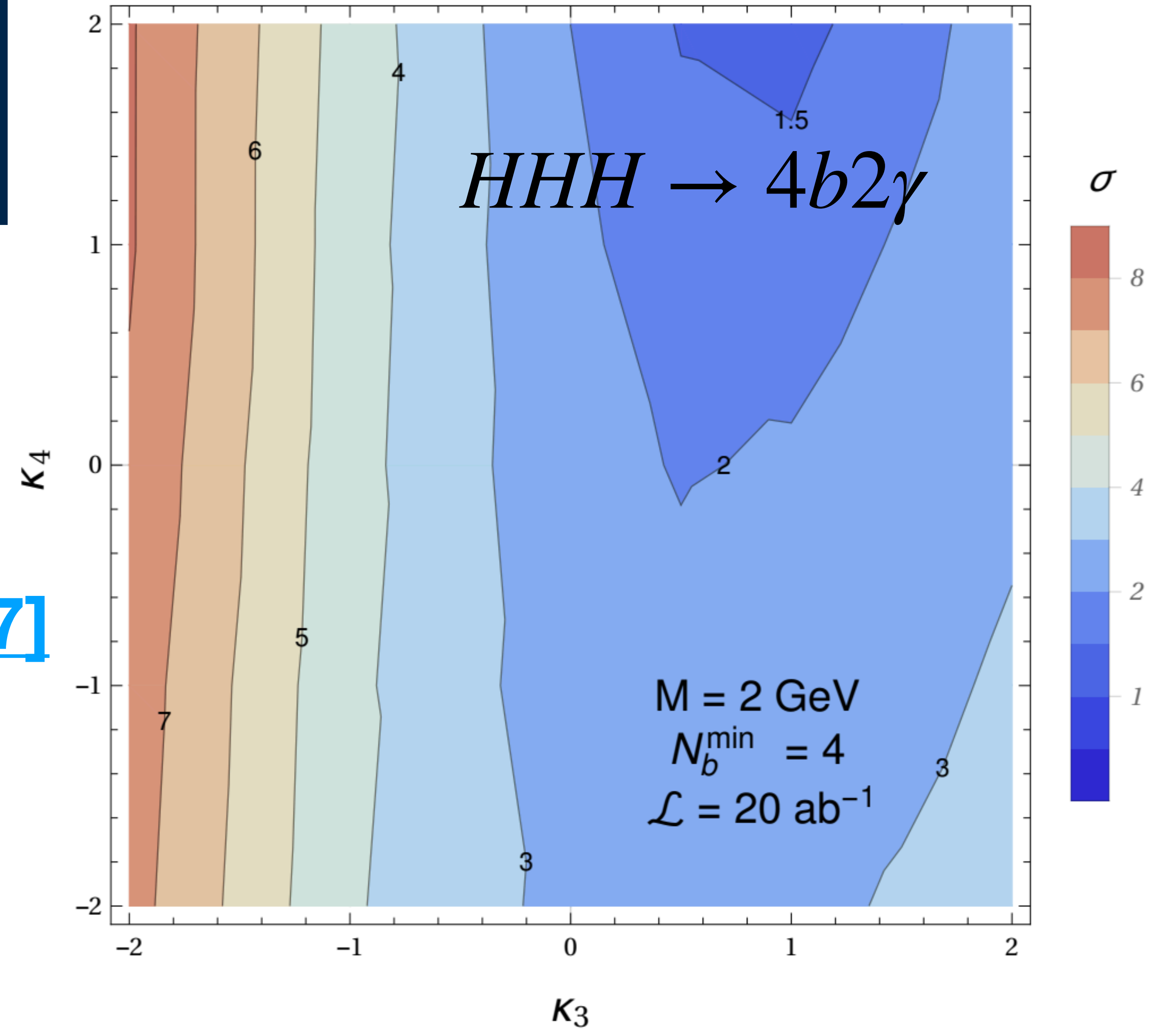
Ordered by likely sensitivity to SM HHH

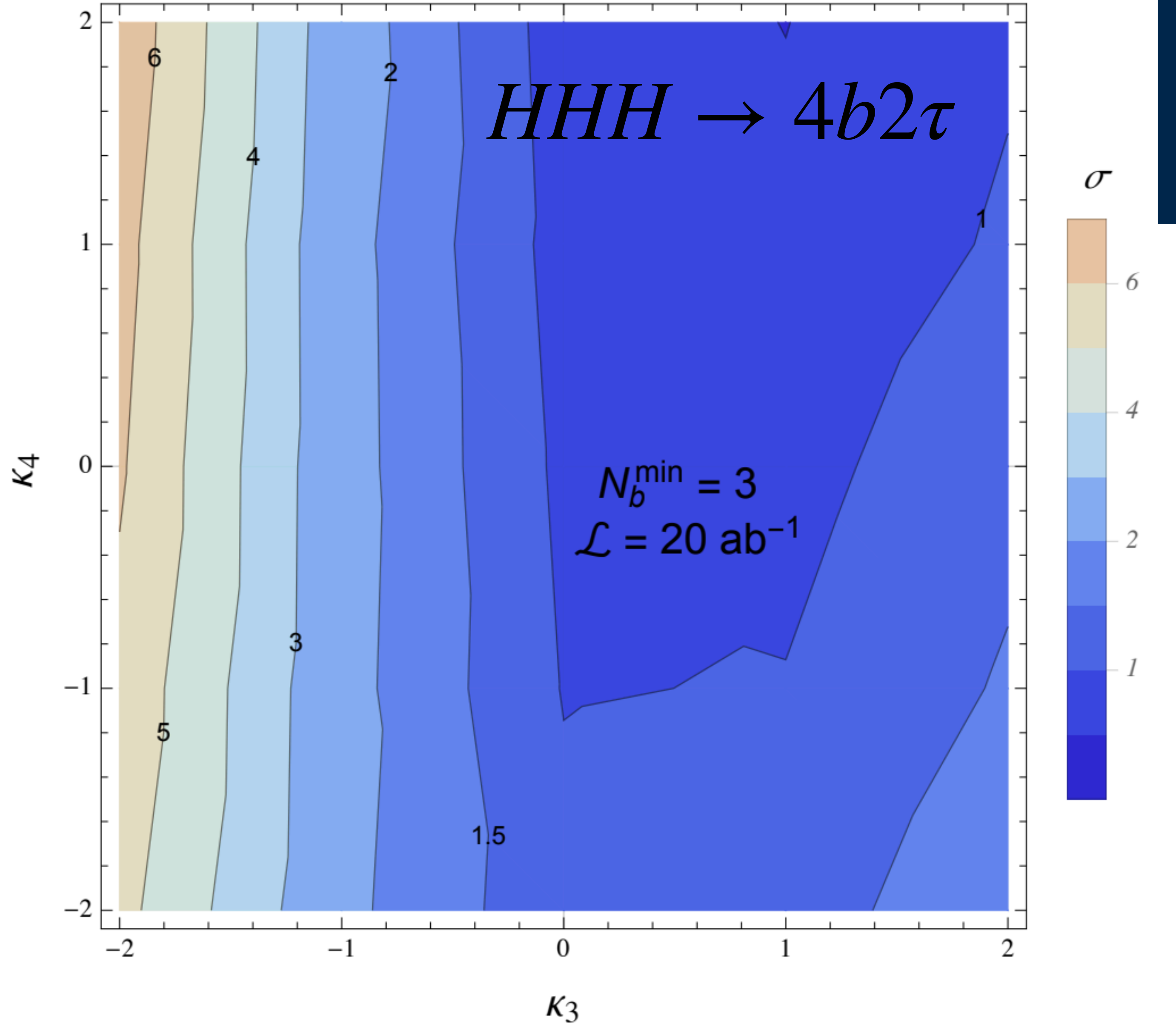
Reminder: HH branching fractions

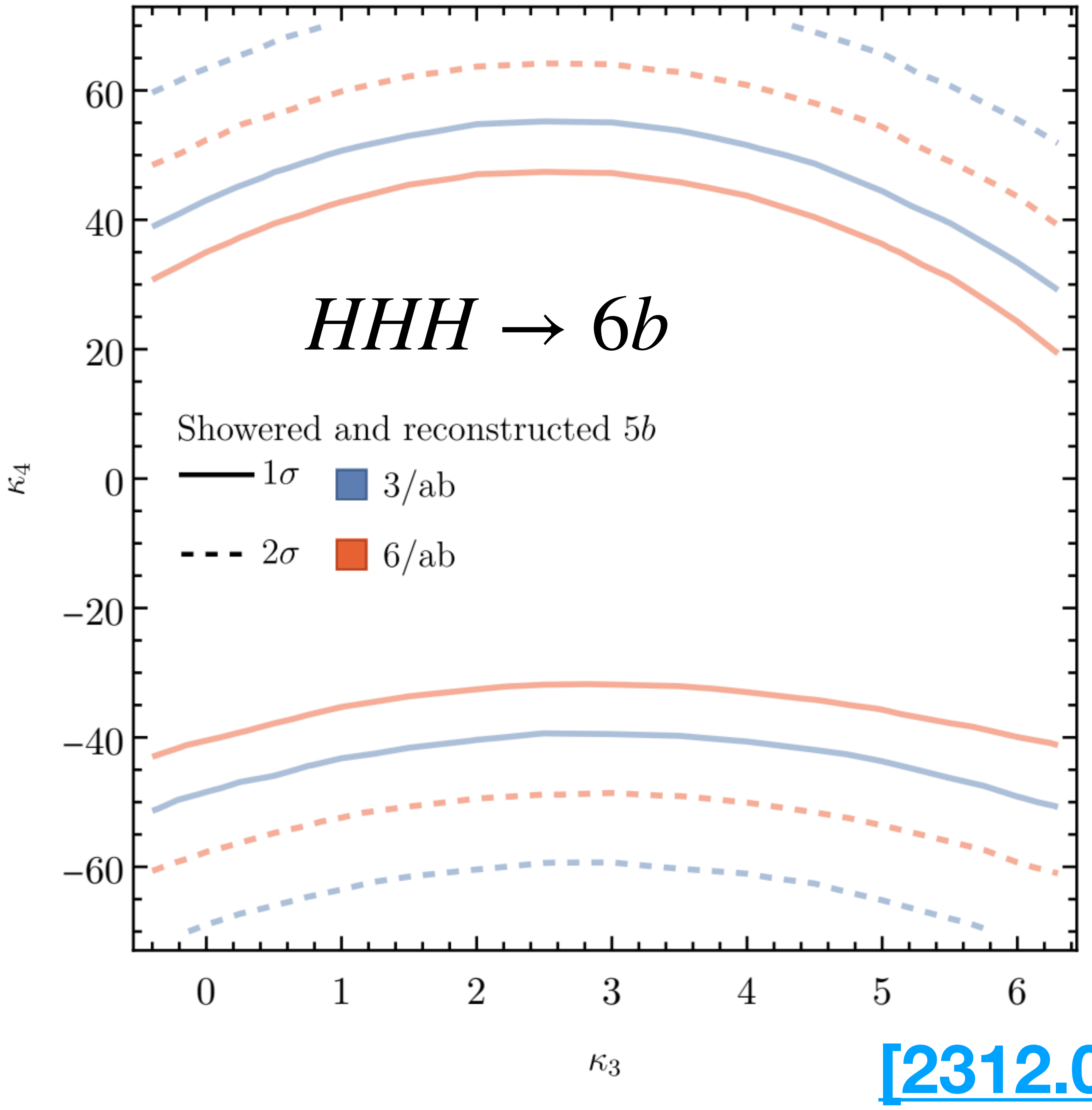
	bb	WW	$\tau\tau$	ZZ	$\gamma\gamma$
bb	34%				
WW	25%	4.6%			
$\tau\tau$	7.4%	2.5%	0.39%		
ZZ	3.1%	1.2%	0.34%	0.076%	
$\gamma\gamma$	0.26%	0.10%	0.029%	0.013%	0.0005%

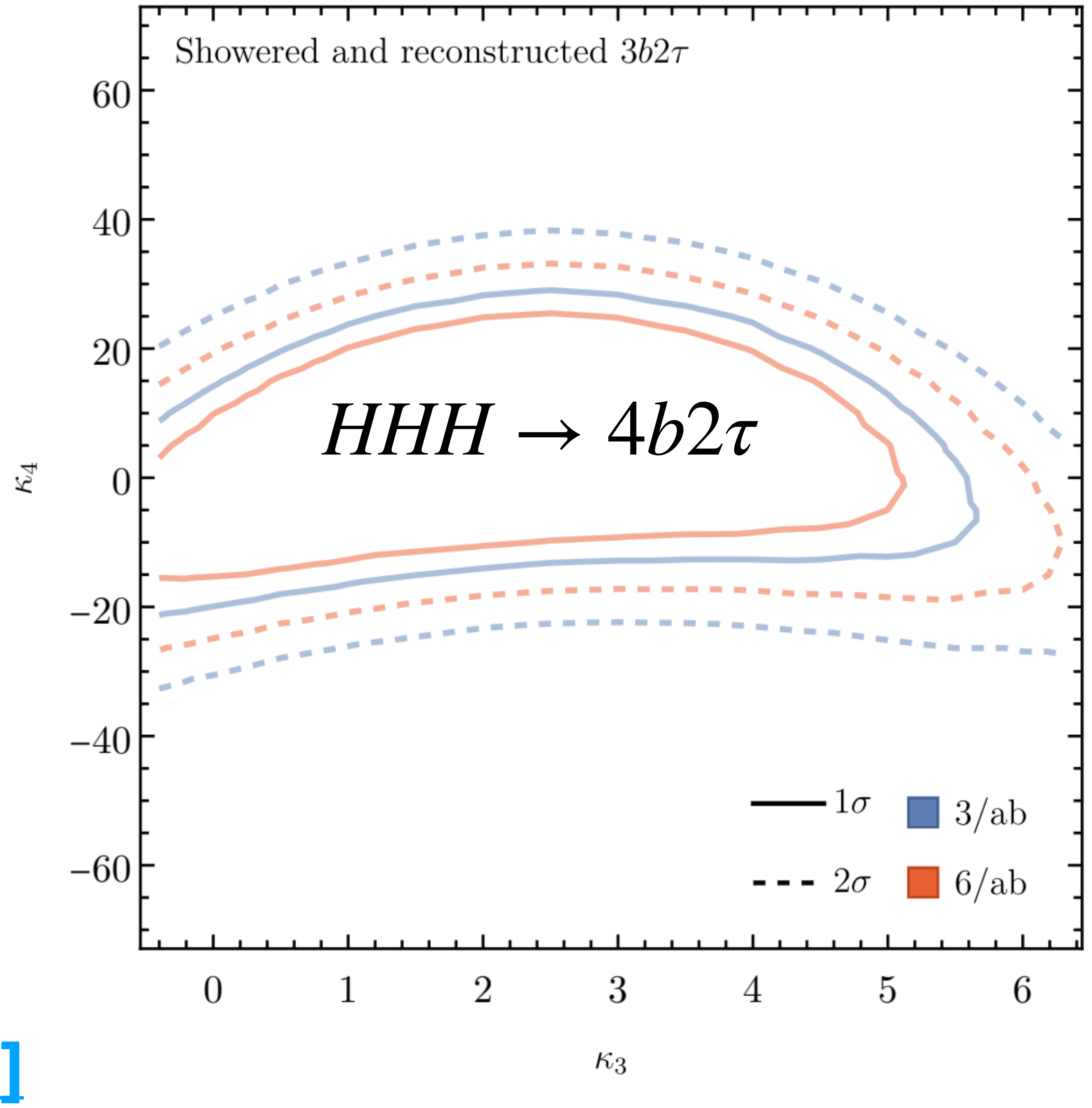
Decay modes

[1510.07697]







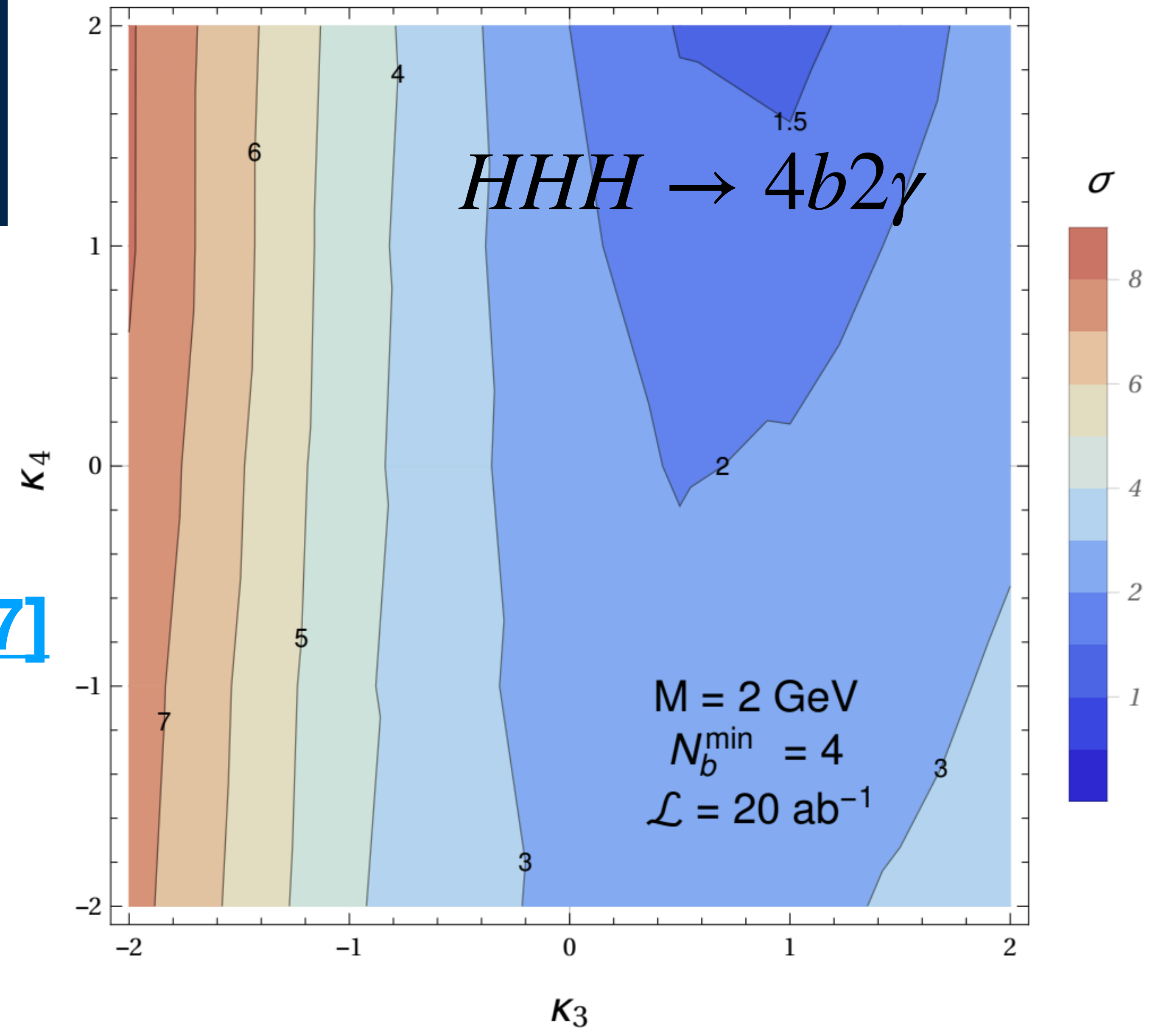


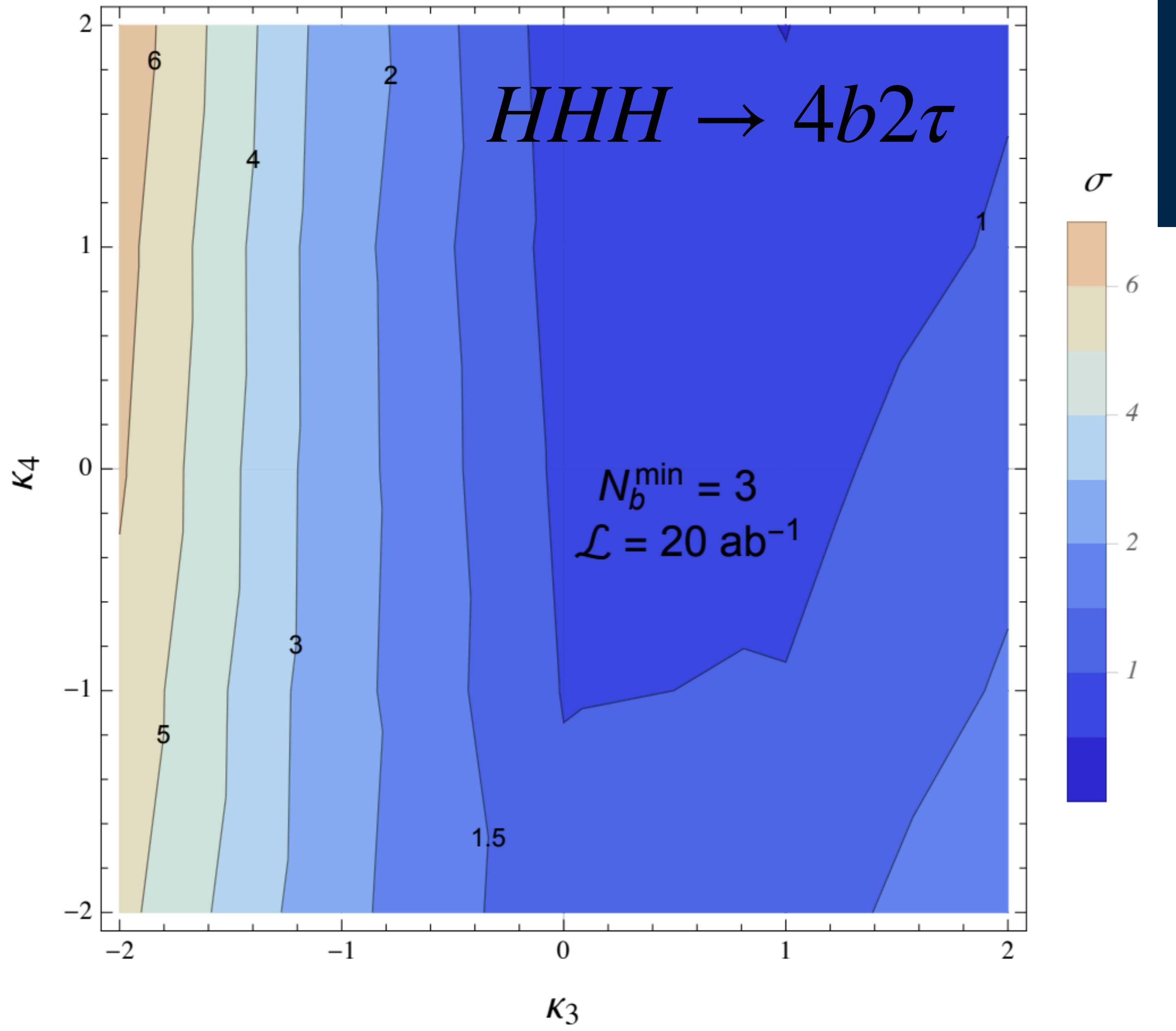
[2312.04646]

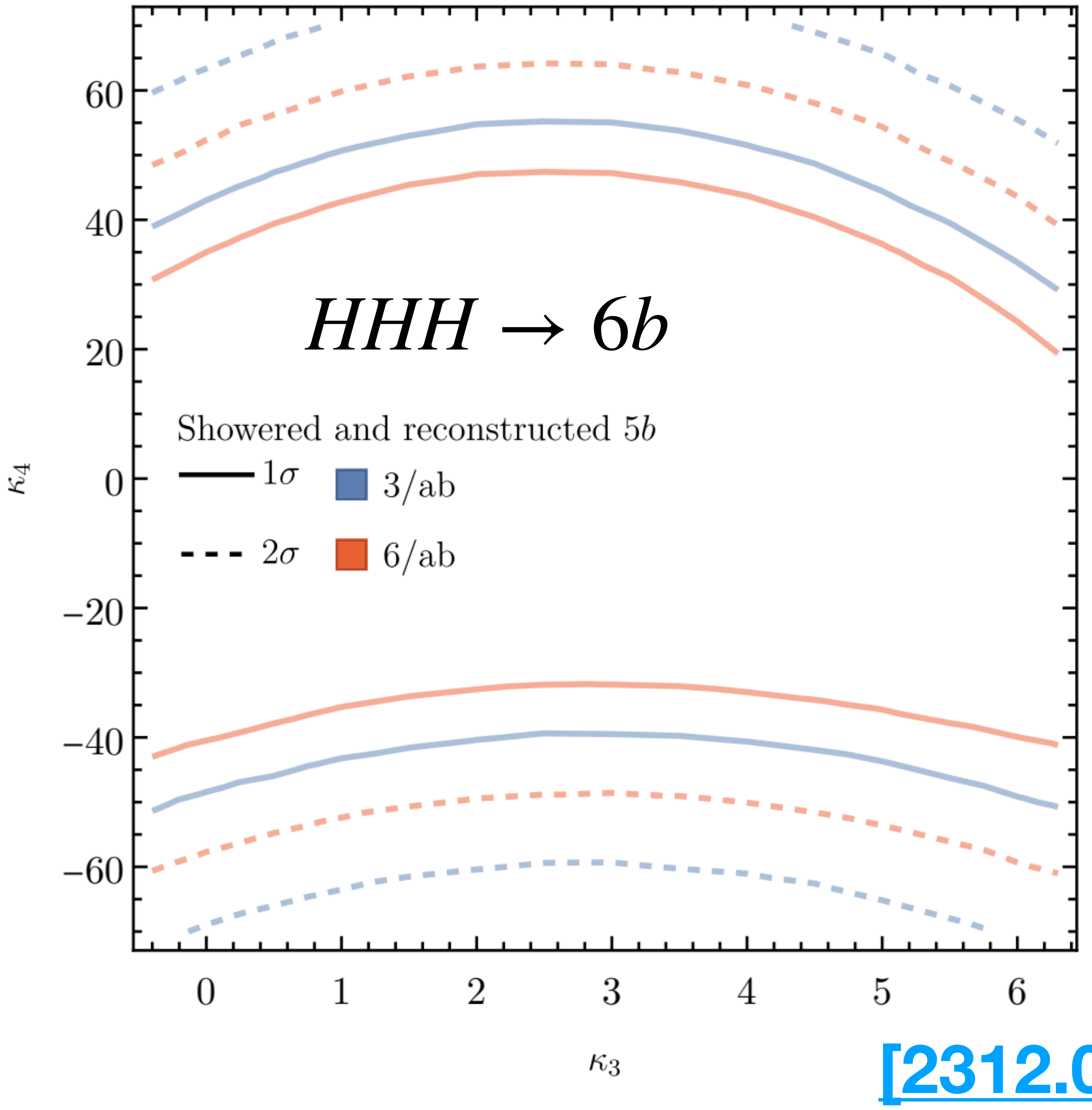
Decay modes

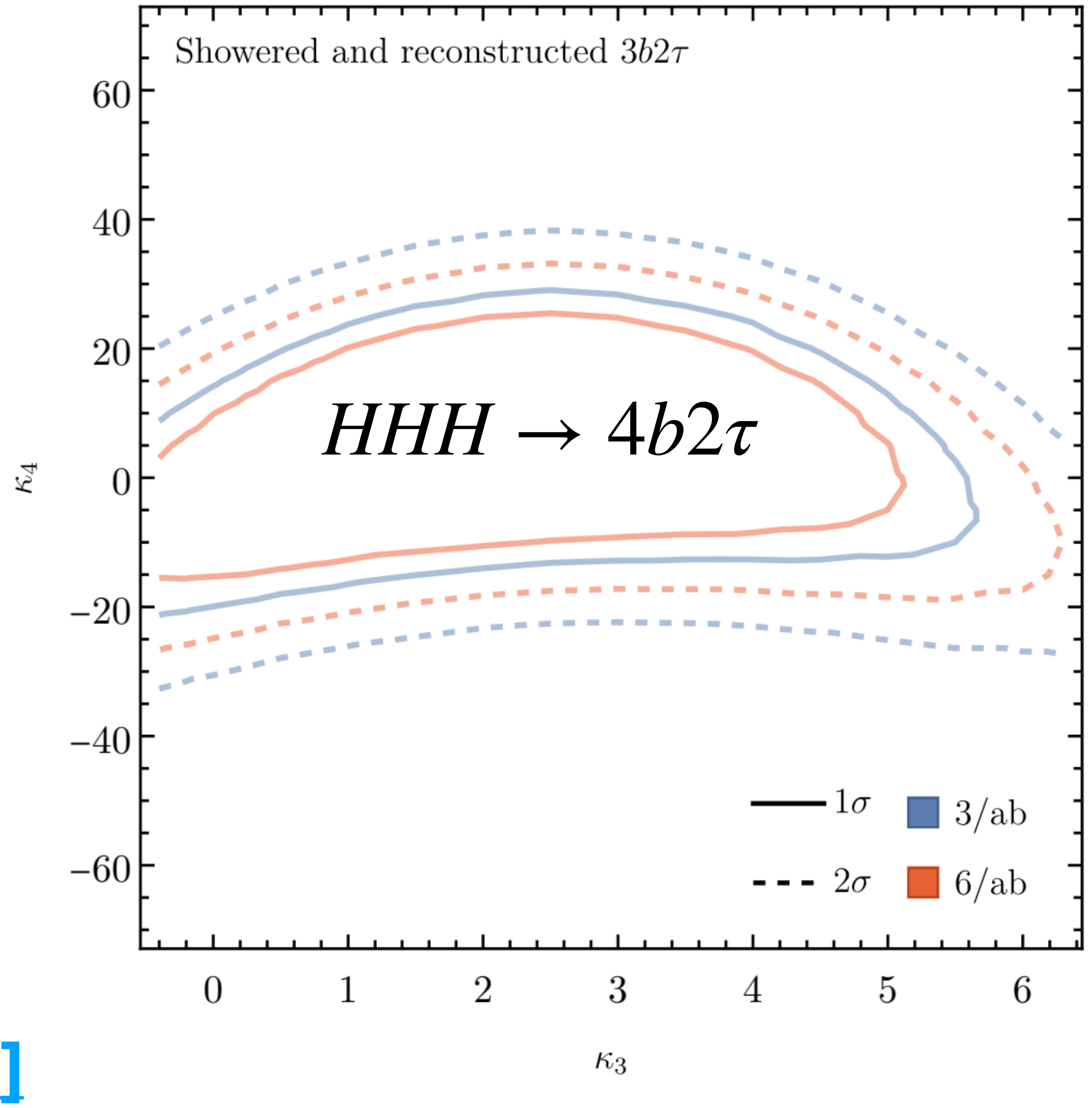
- **Studies on sensitivity of decay modes are in (slight) conflict**

[1510.07697]





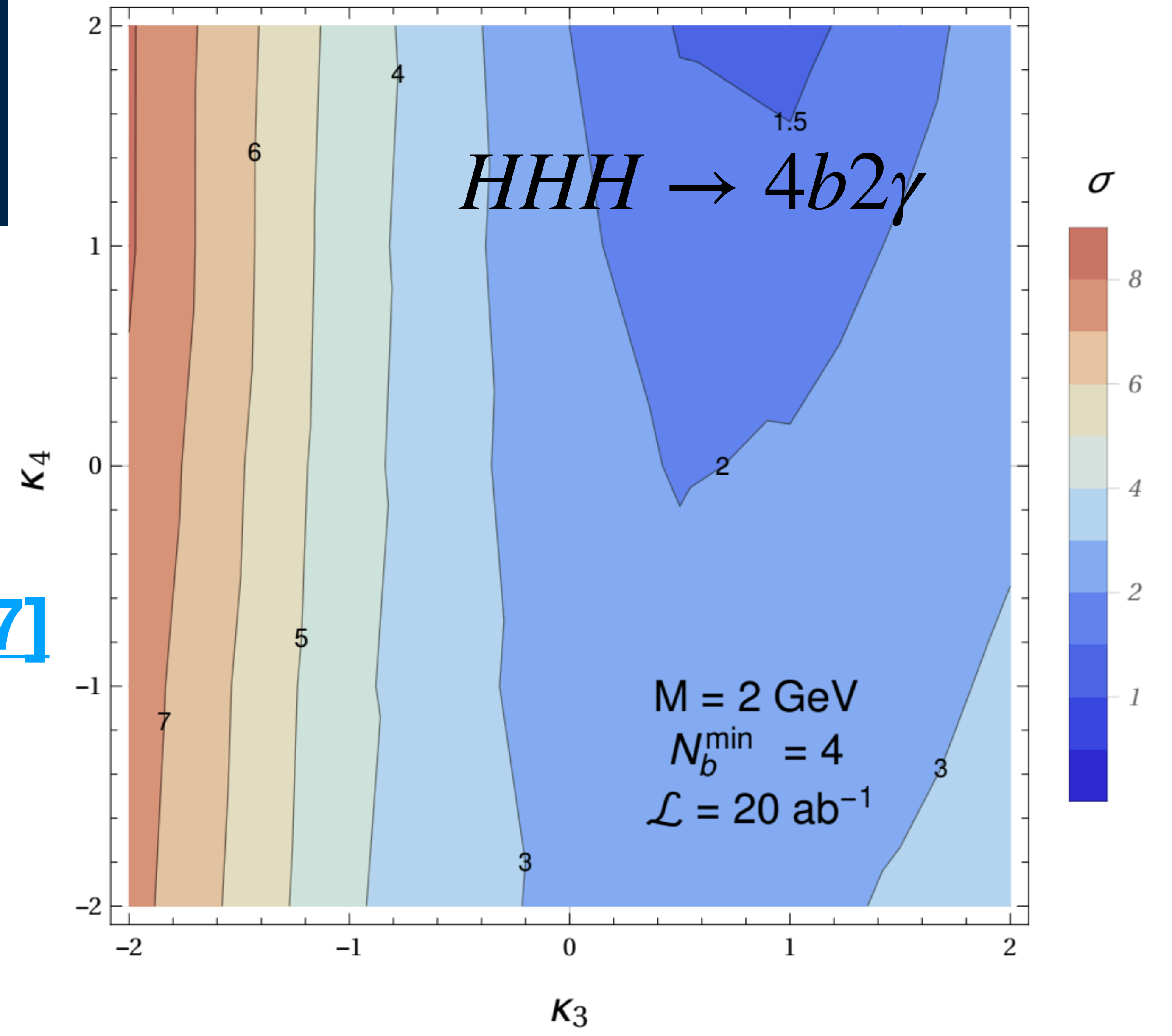


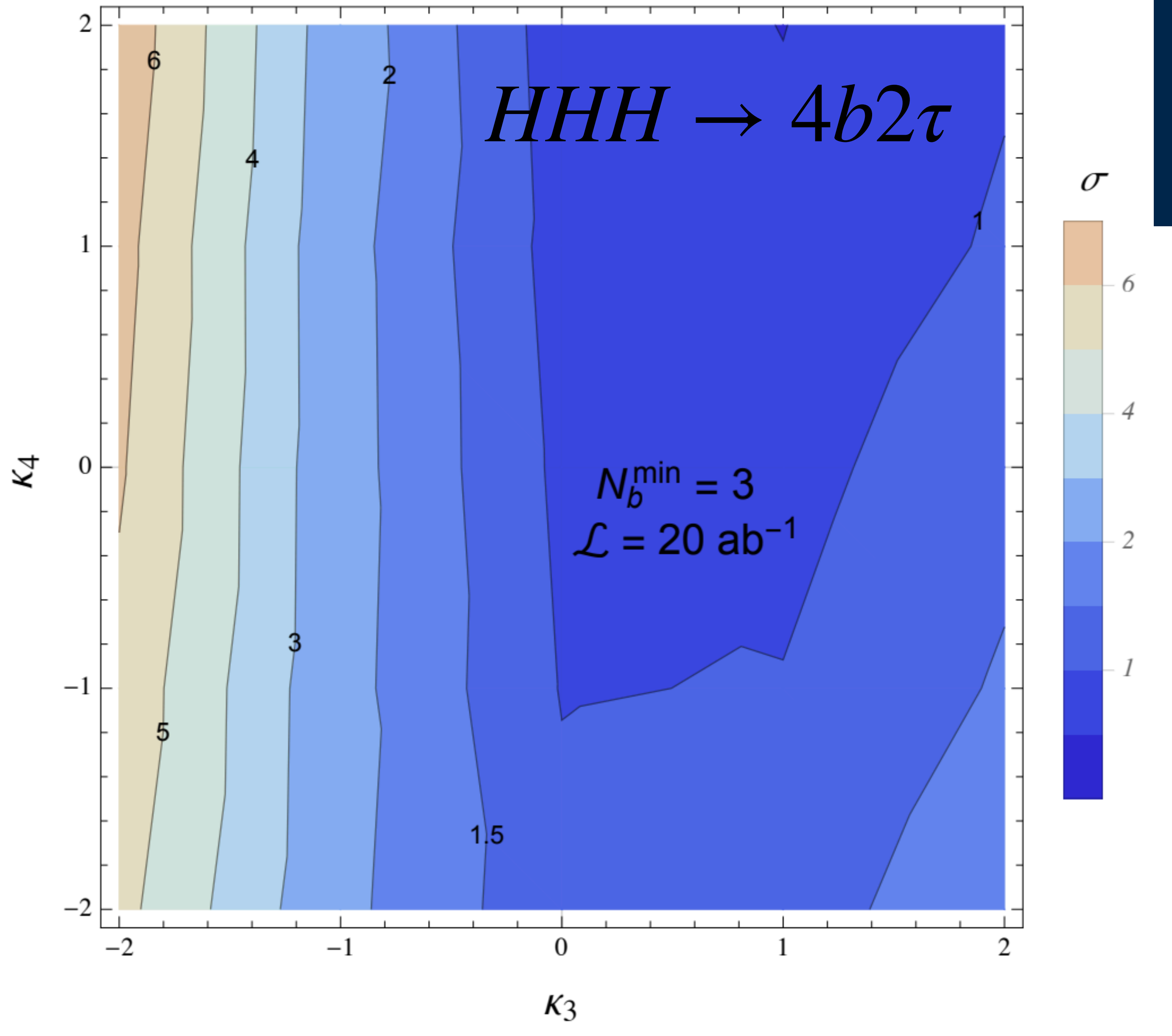


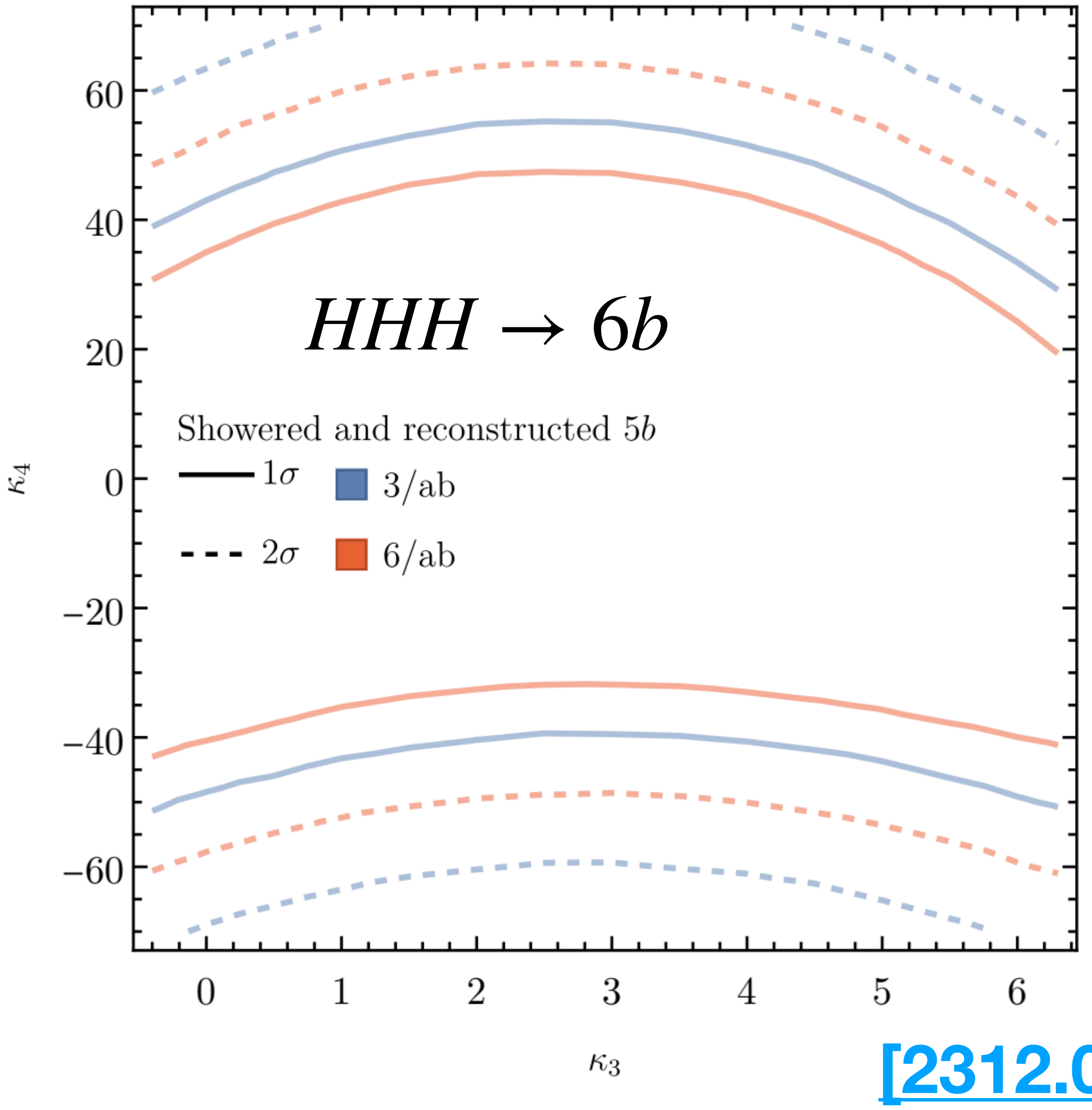
[2312.04646]

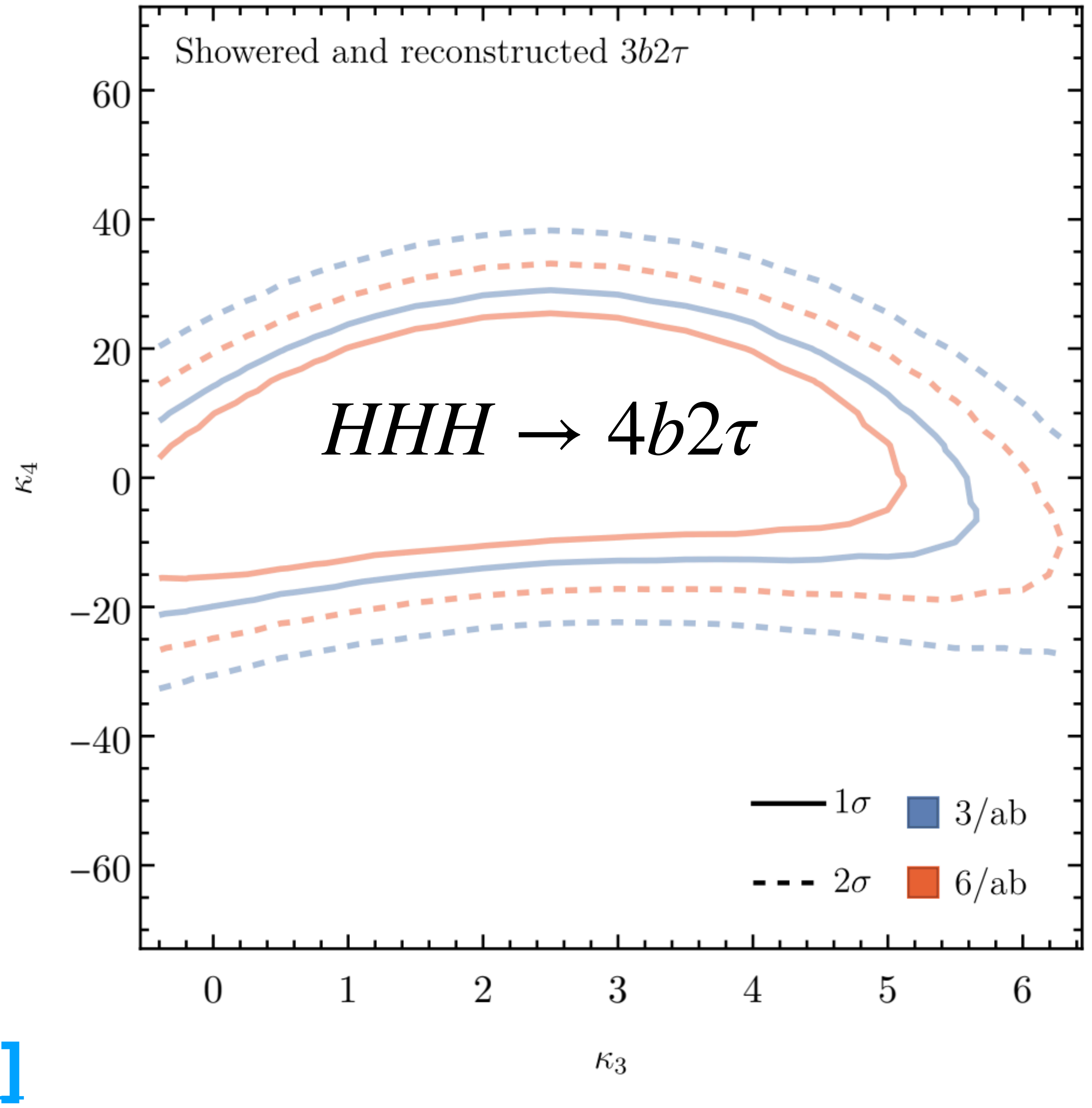
Decay modes

- **Studies on sensitivity of decay modes are in (slight) conflict**
- **Some (top) show $4b2\gamma$ is best** [1510.07697]





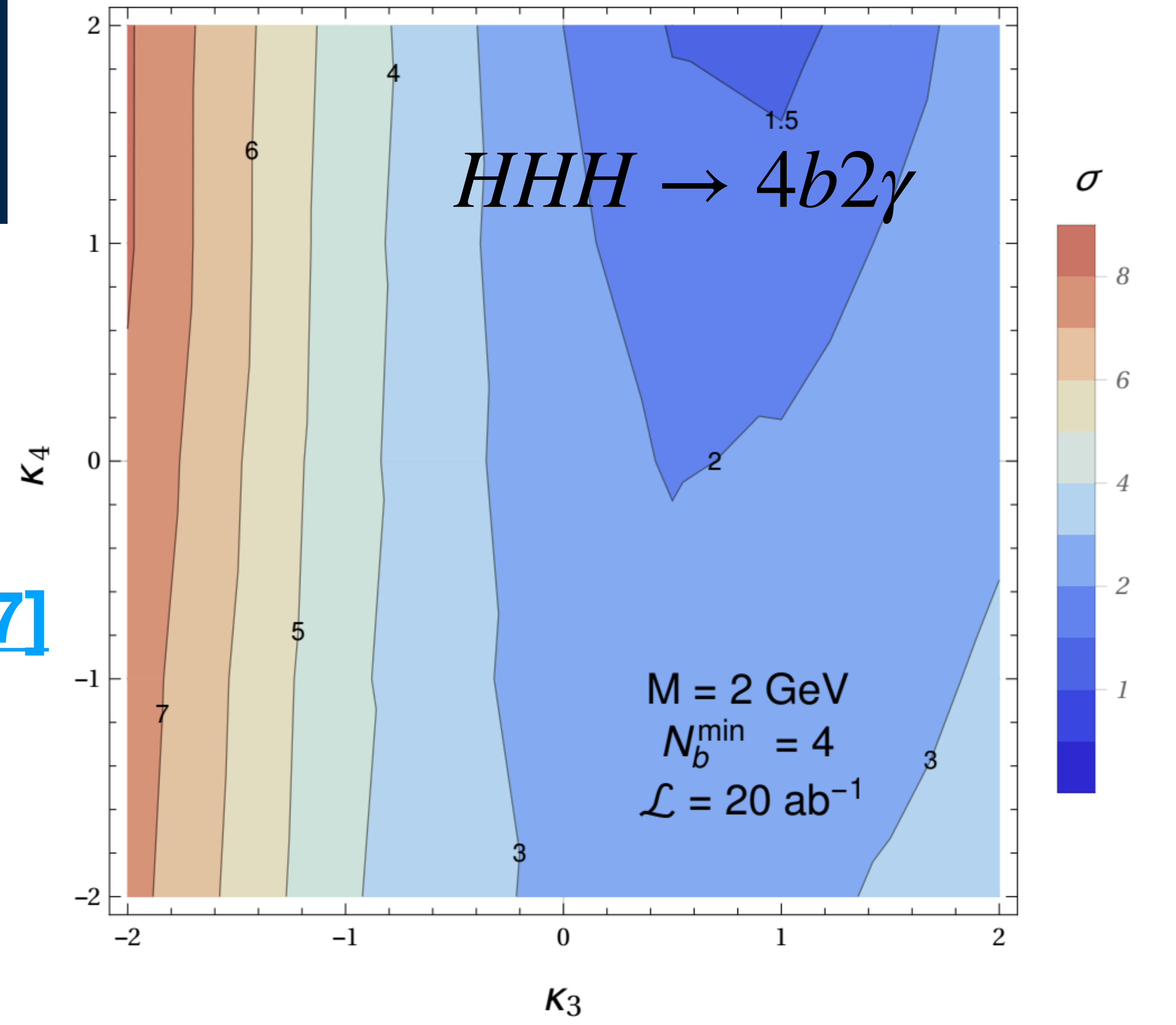


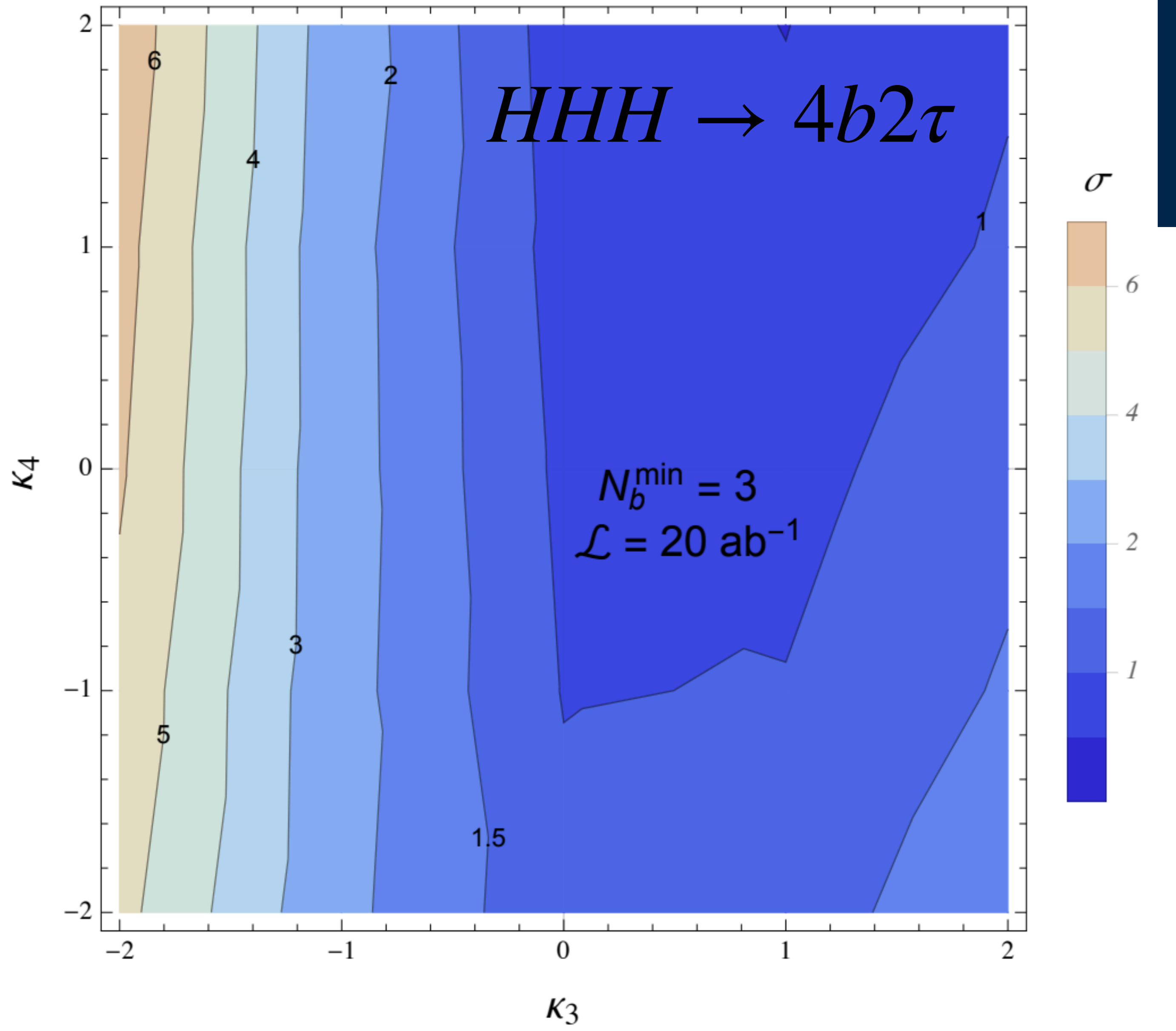


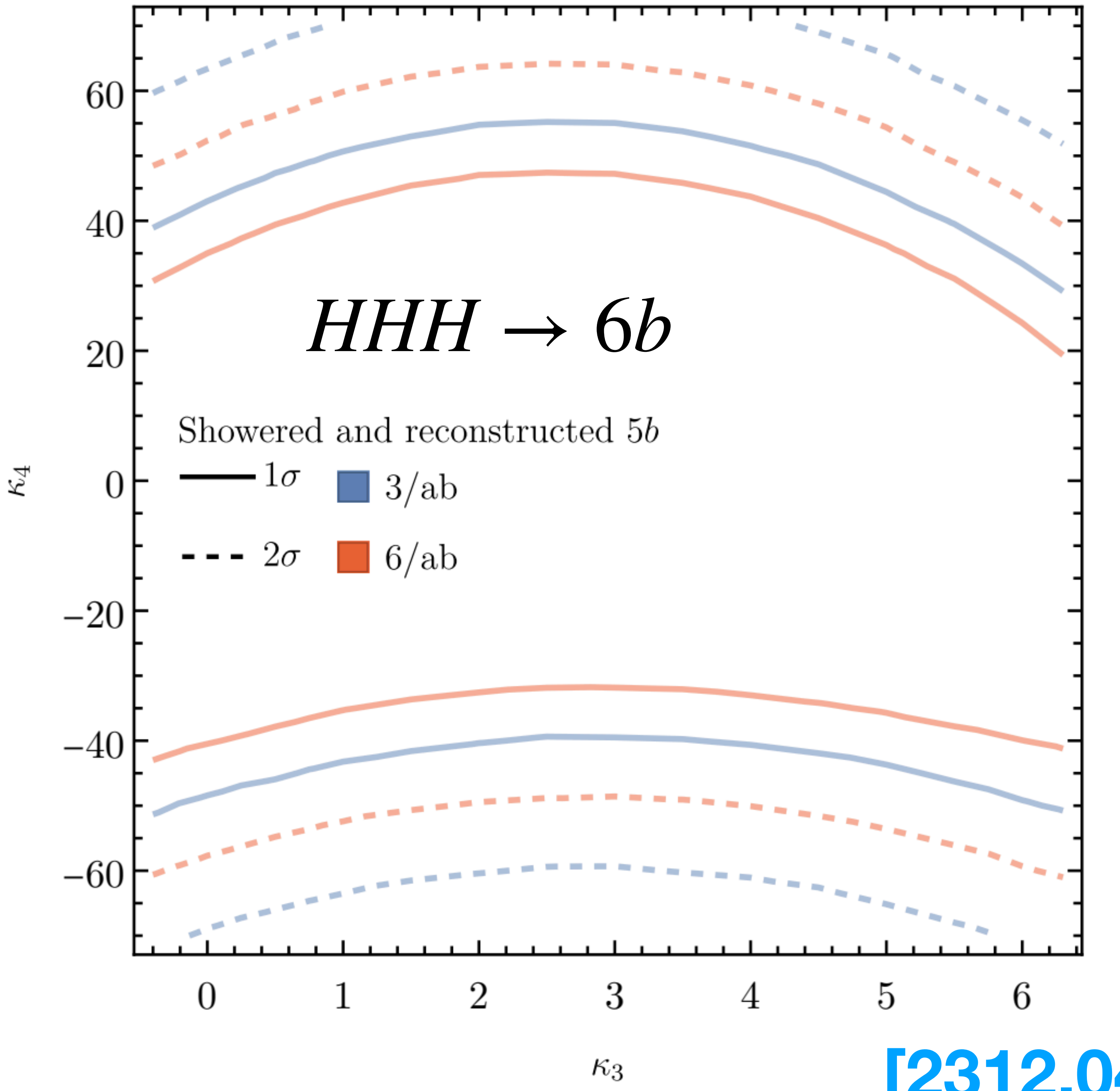
[2312.04646]

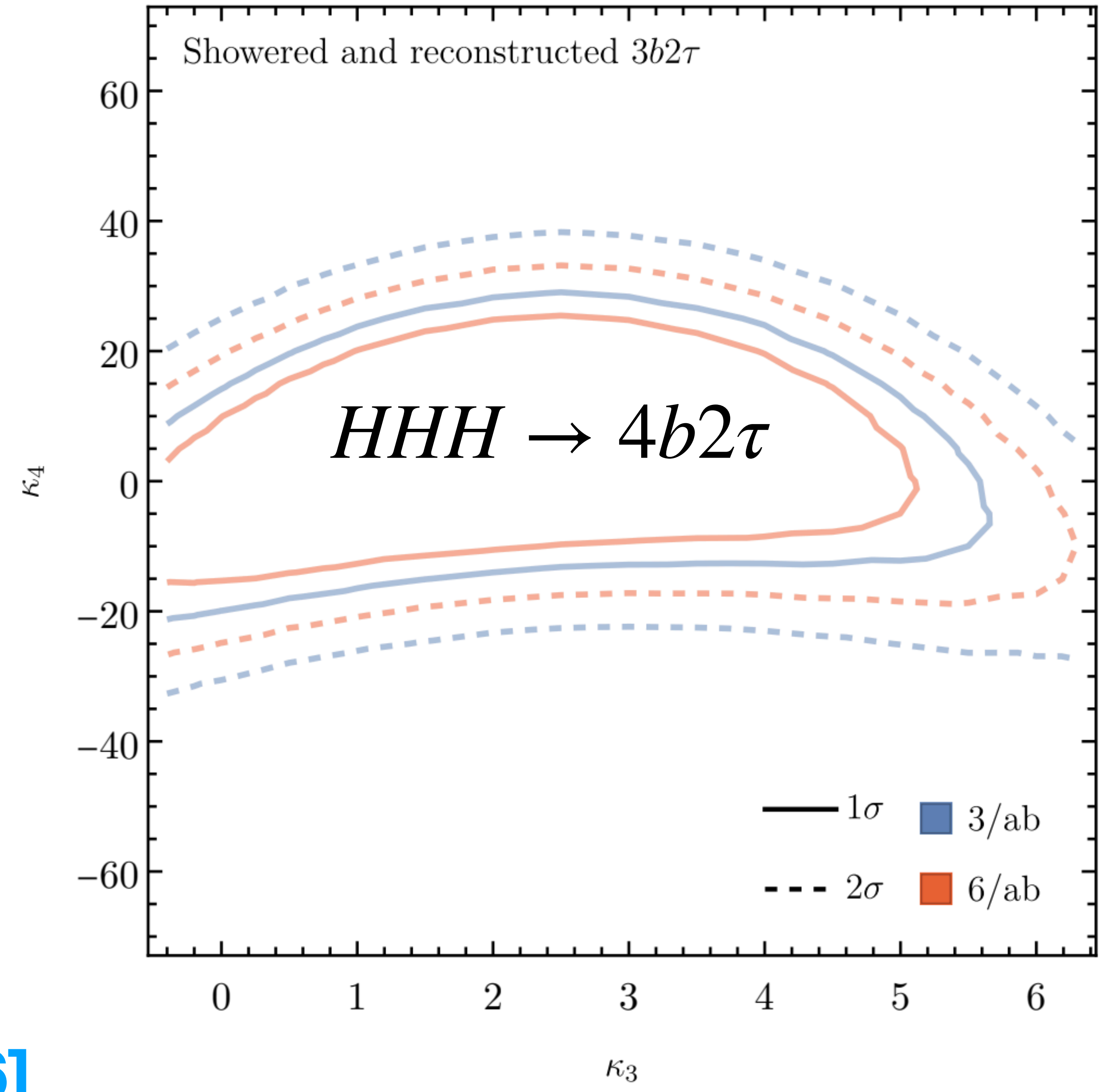
Decay modes

- **Studies on sensitivity of decay modes are in (slight) conflict**
- **Some (top) show $4b2\gamma$ is best** [[1510.07697]]
- **Others (bottom) show $4b2\tau$ is best**





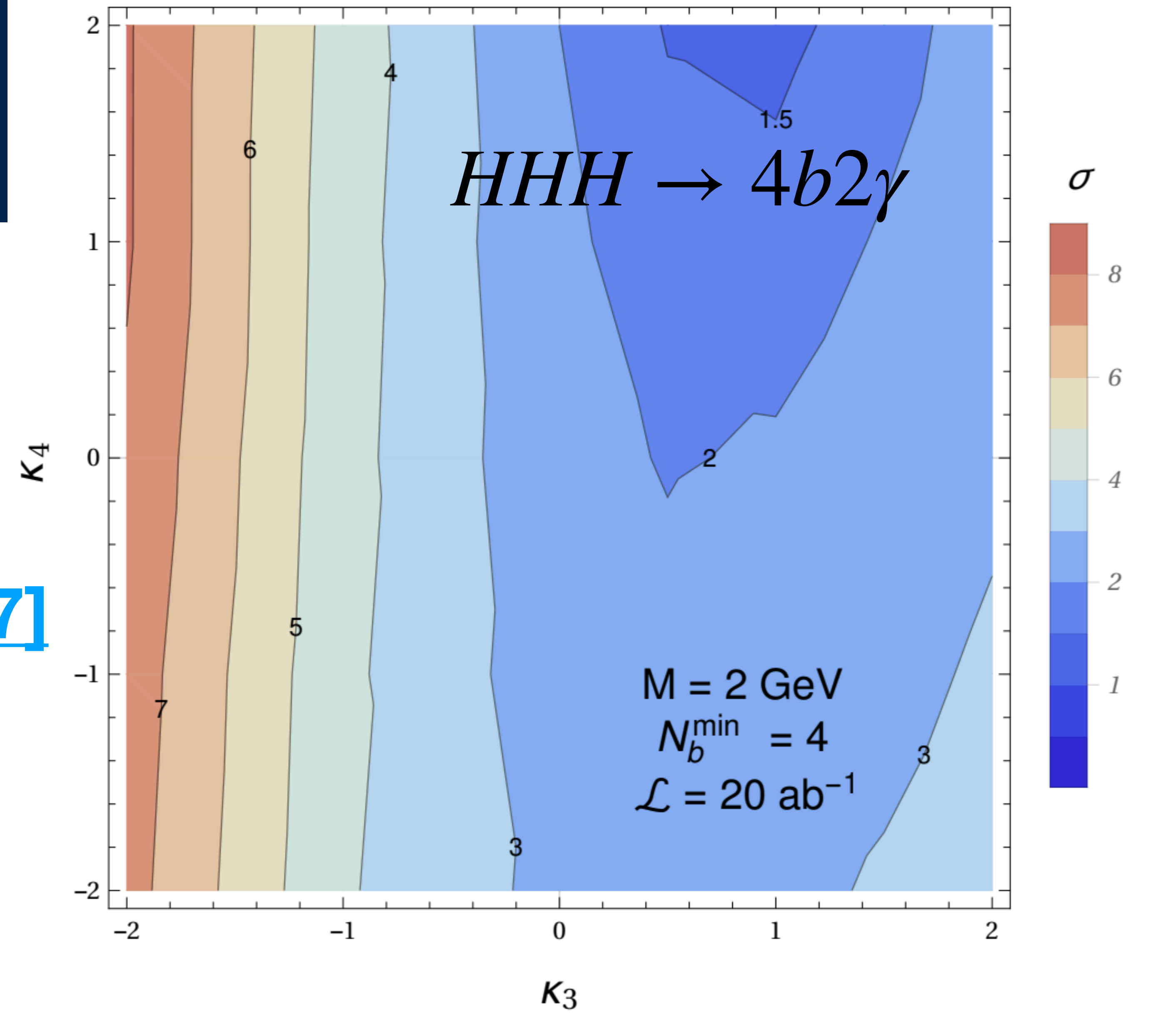


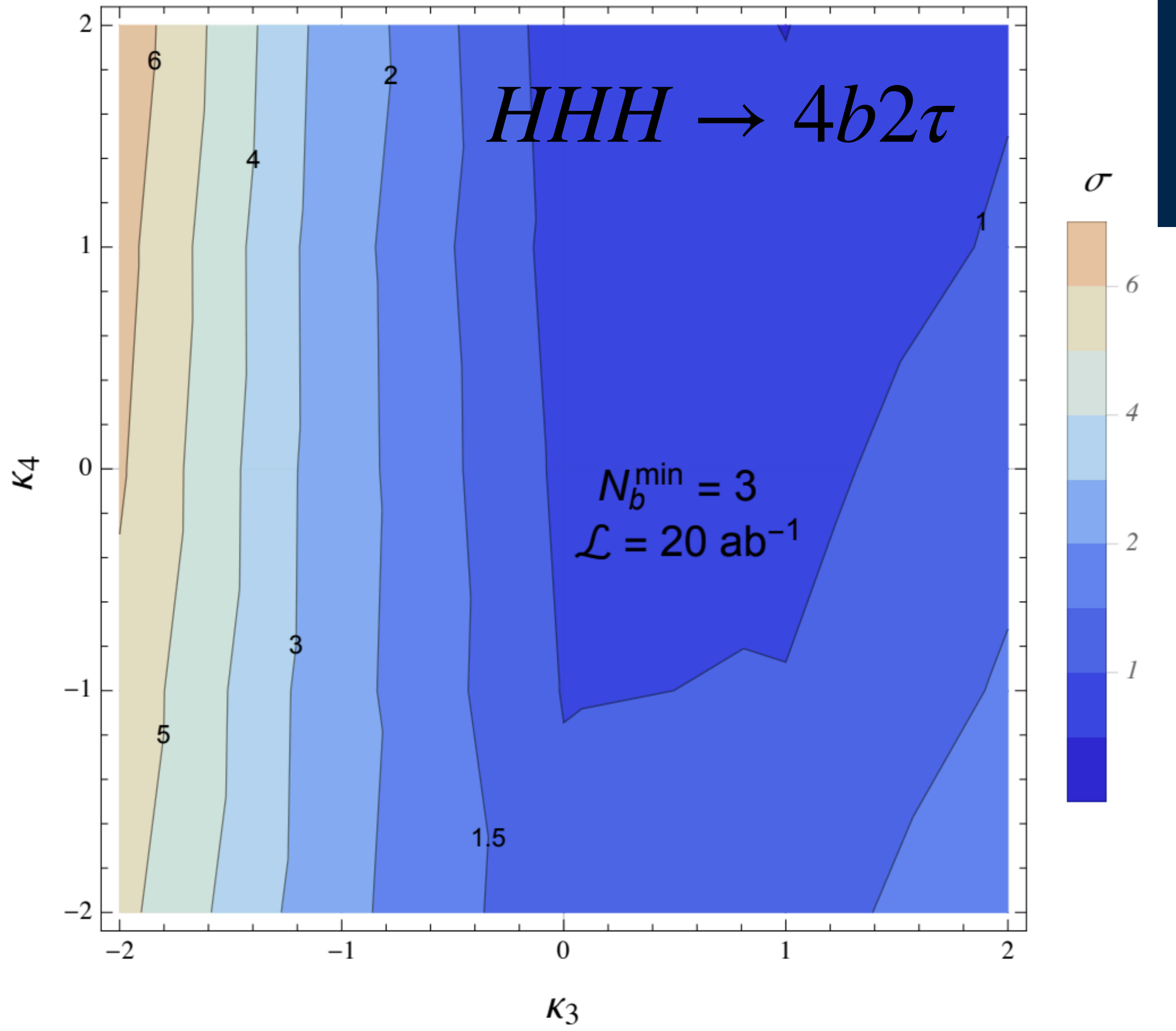


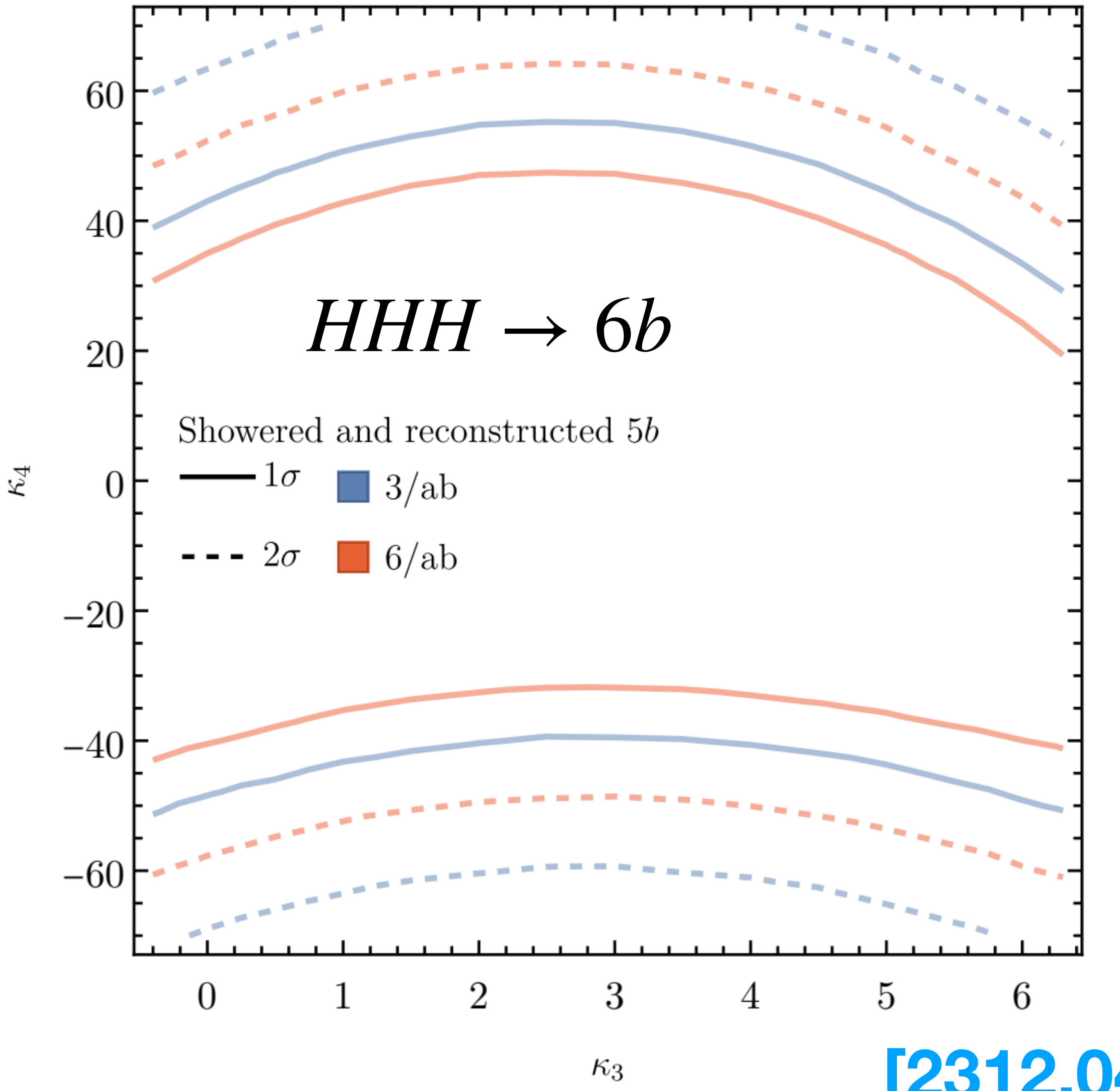
[[2312.04646]]

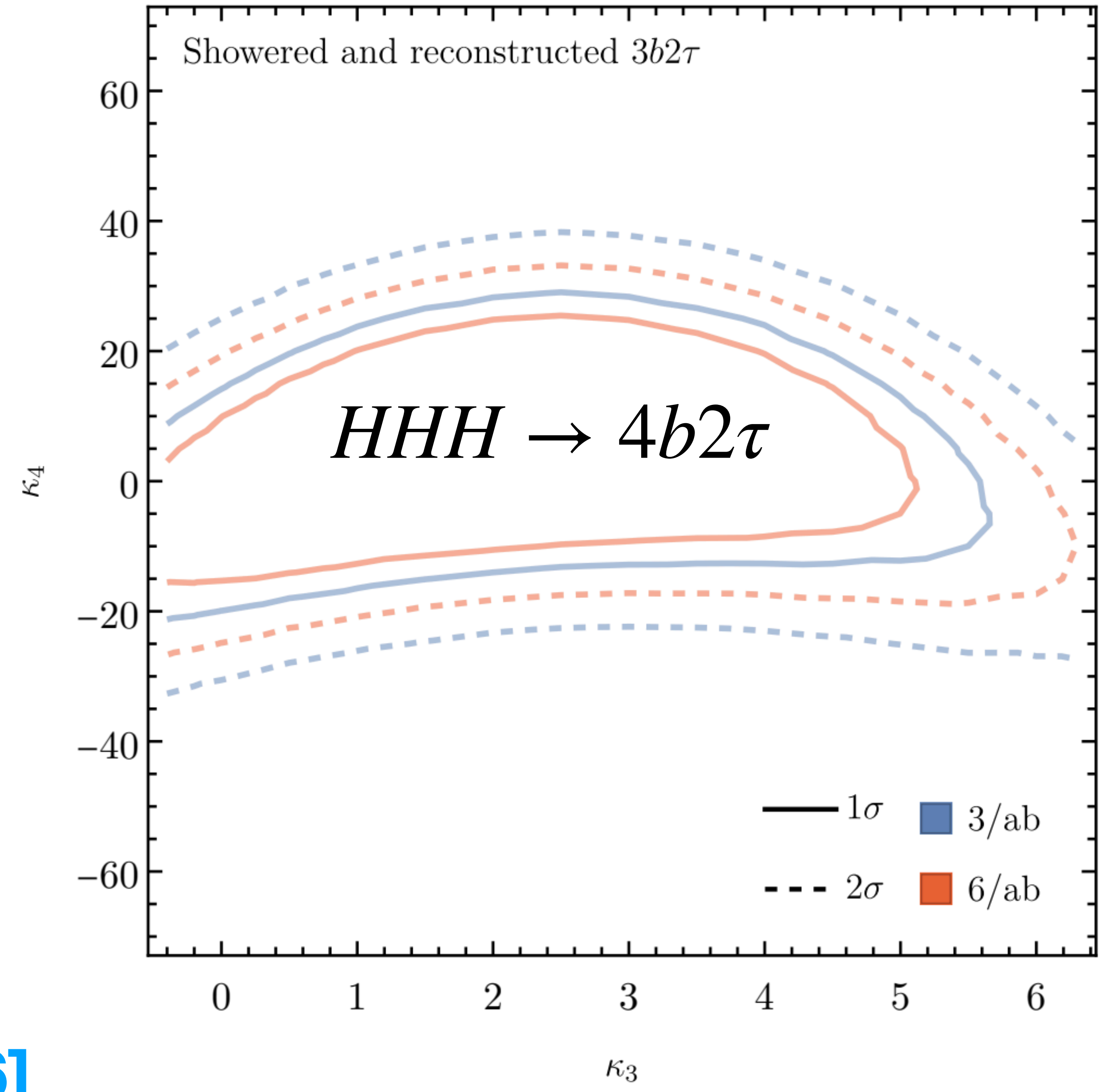
Decay modes

- **Studies on sensitivity of decay modes are in (slight) conflict**
- **Some (top) show $4b2\gamma$ is best** [1510.07697]
- **Others (bottom) show $4b2\tau$ is best**
- **ATLAS internal studies showed $6b$ was best**
  - May be because other studies used $5b$ SR





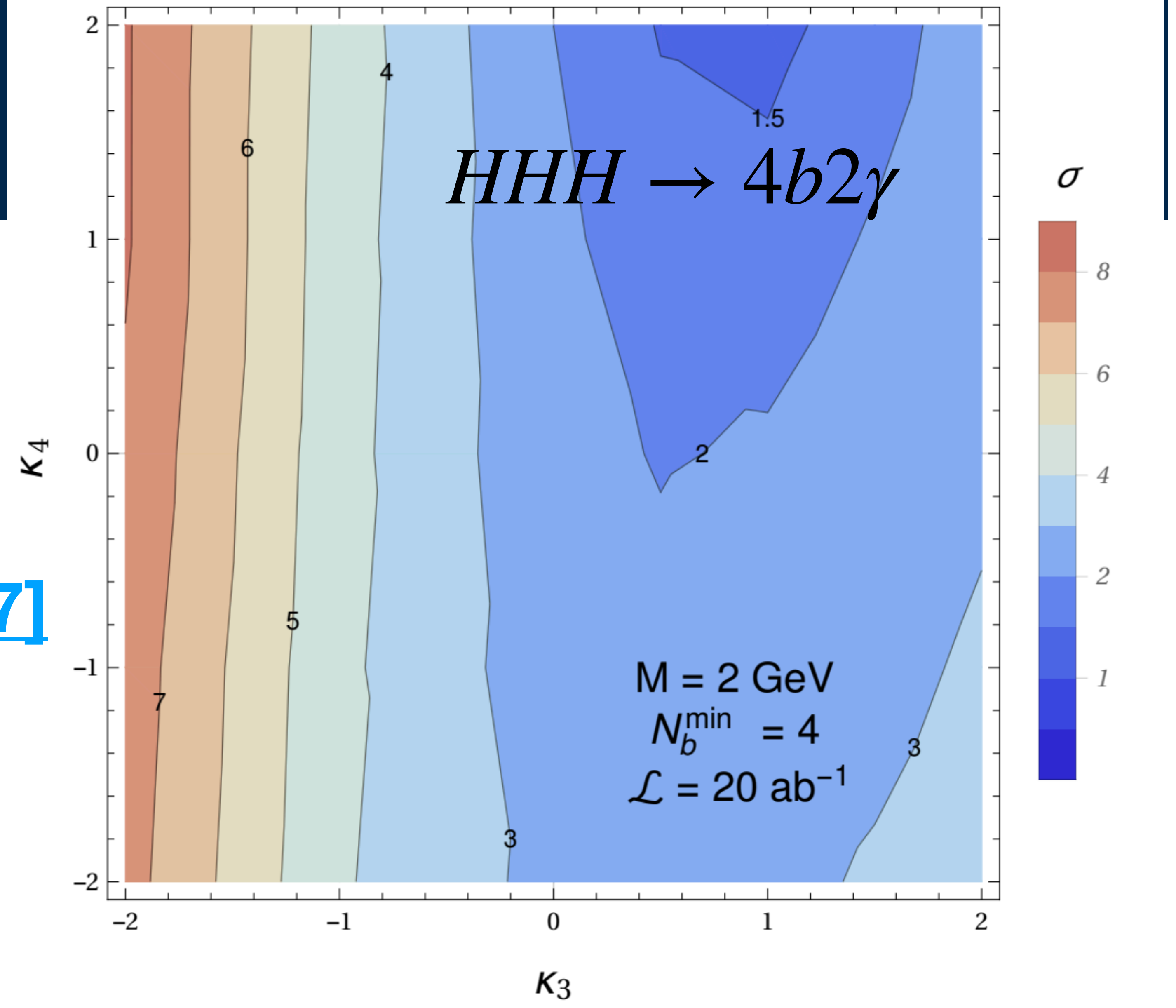


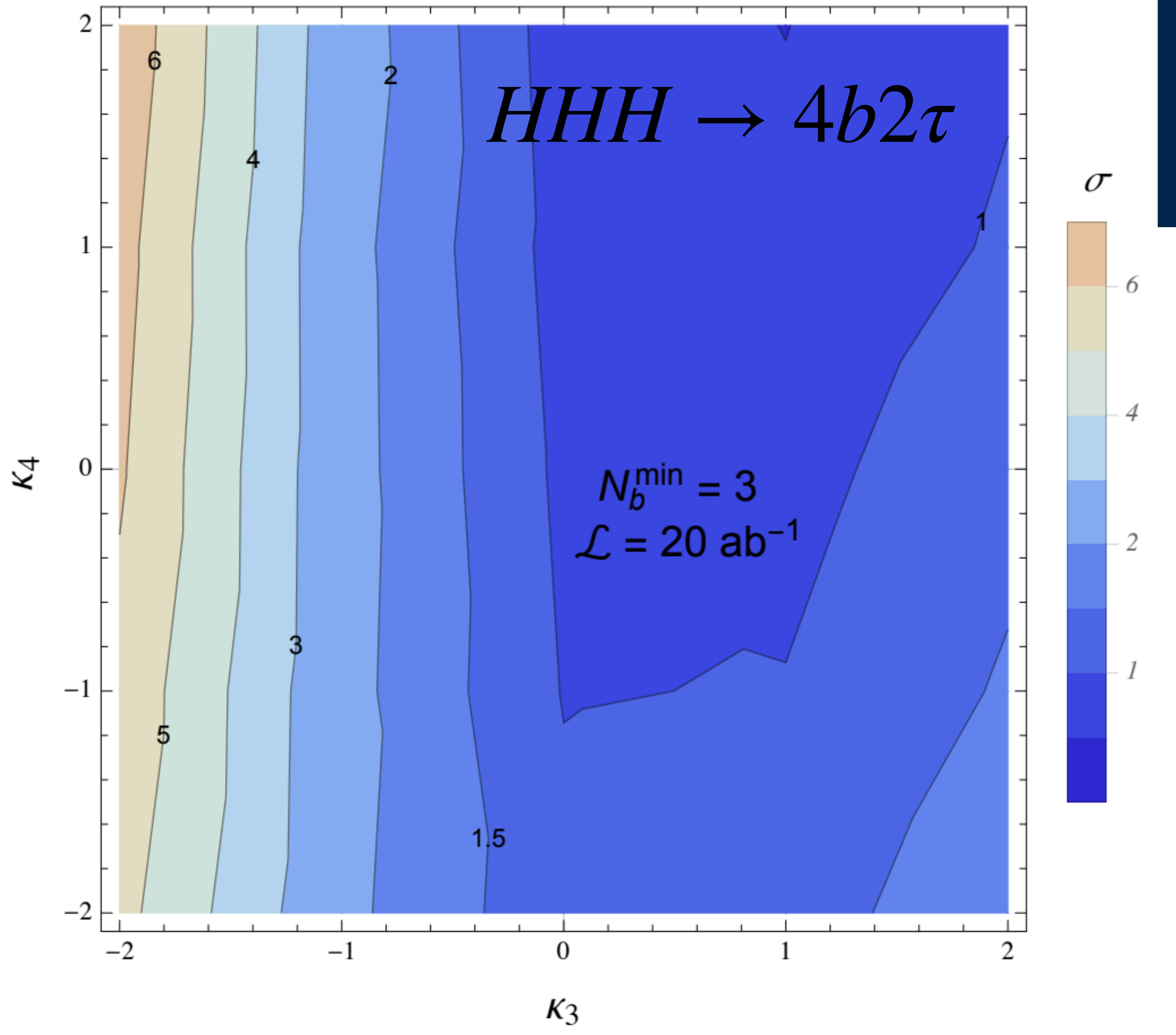


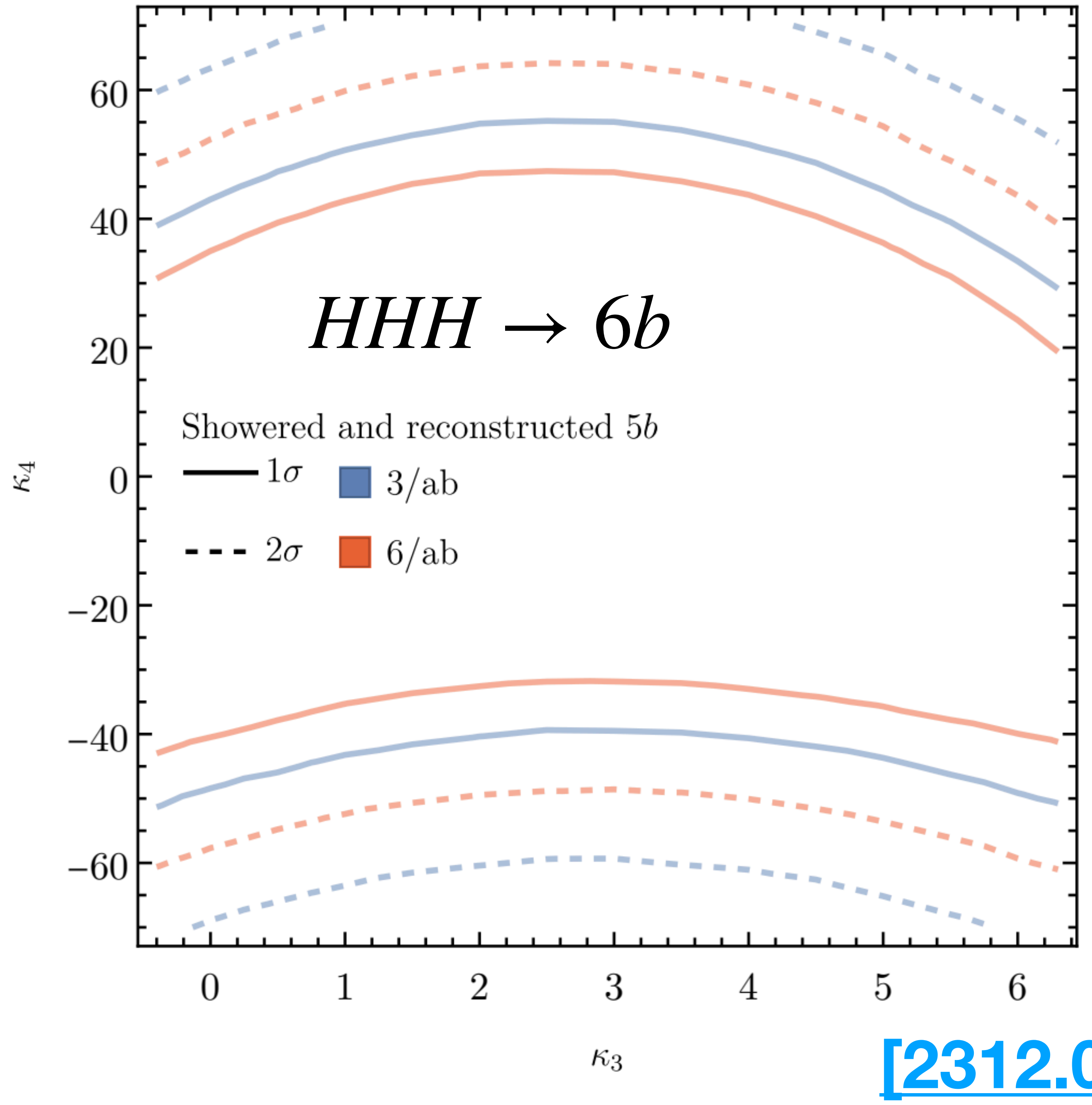
[2312.04646]

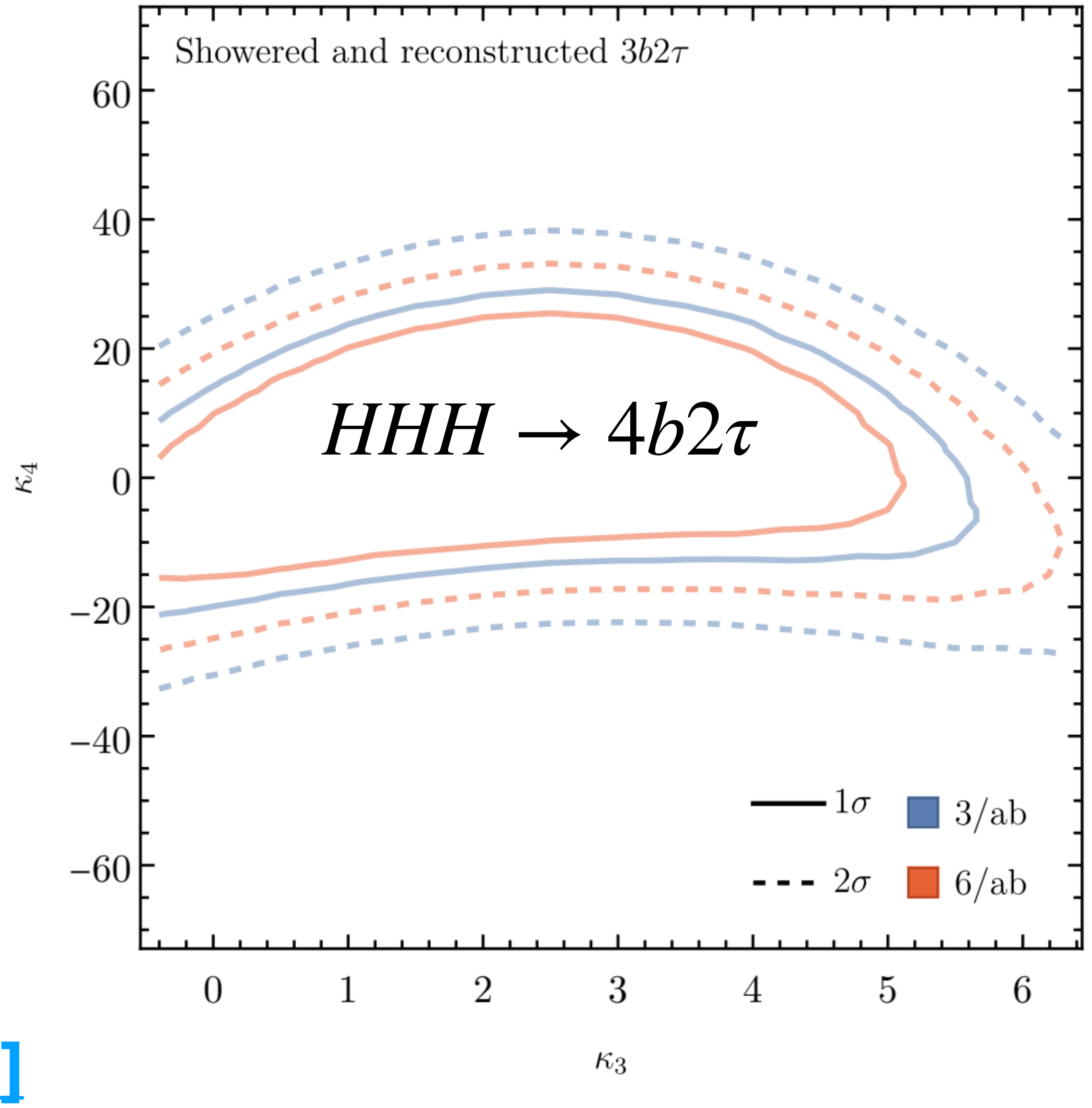
Decay modes

- **Studies on sensitivity of decay modes are in (slight) conflict**
- **Some (top) show $4b2\gamma$ is best** [1510.07697]
- **Others (bottom) show $4b2\tau$ is best**
- **ATLAS internal studies showed $6b$ was best**
  - May be because other studies used $5b$ SR
- **$4b1\ell$ is in my opinion under studied!**
  - Benefit from at least one lepton from many final states: $WW*$, $\tau\tau$, $ZZ*$





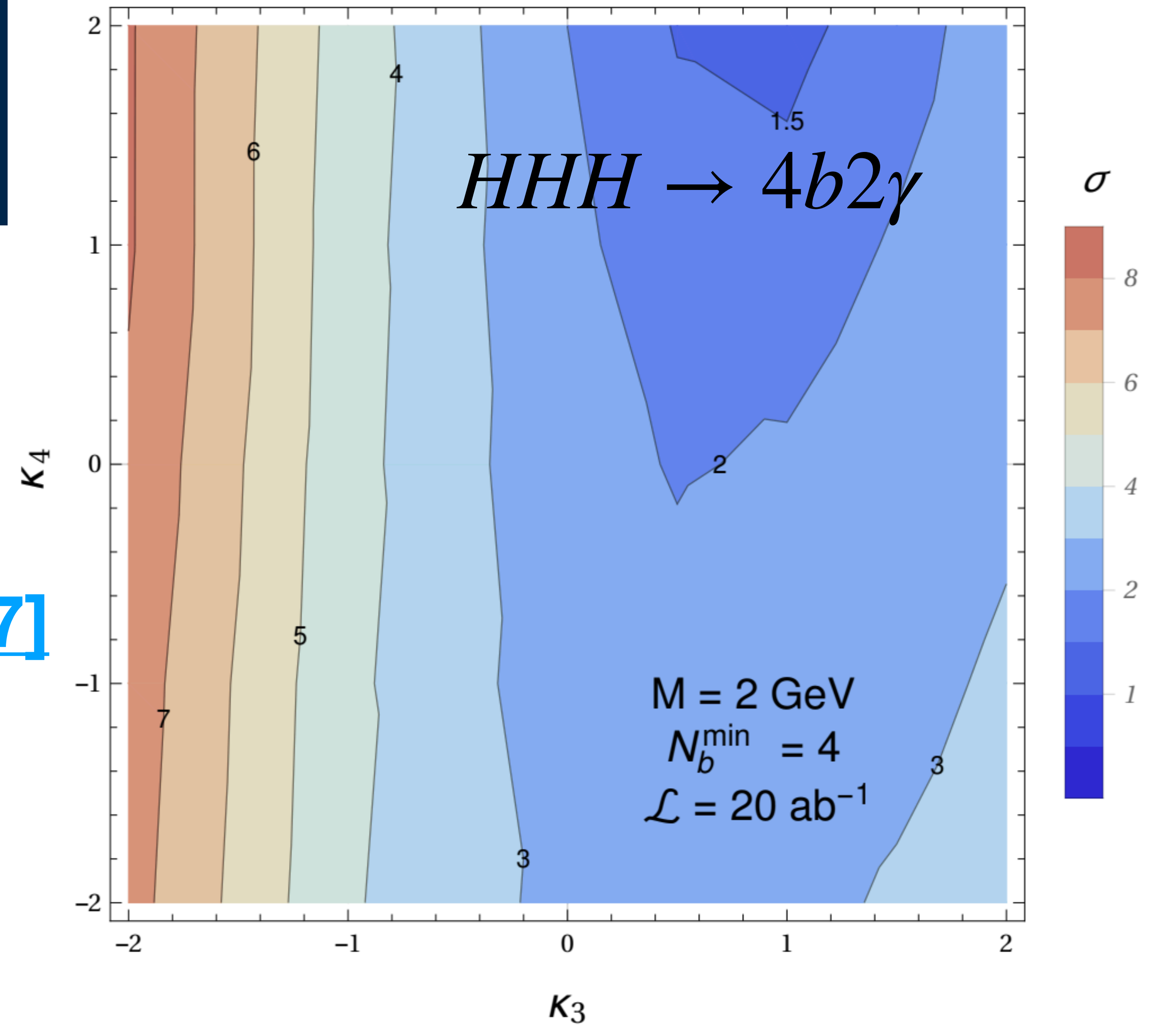


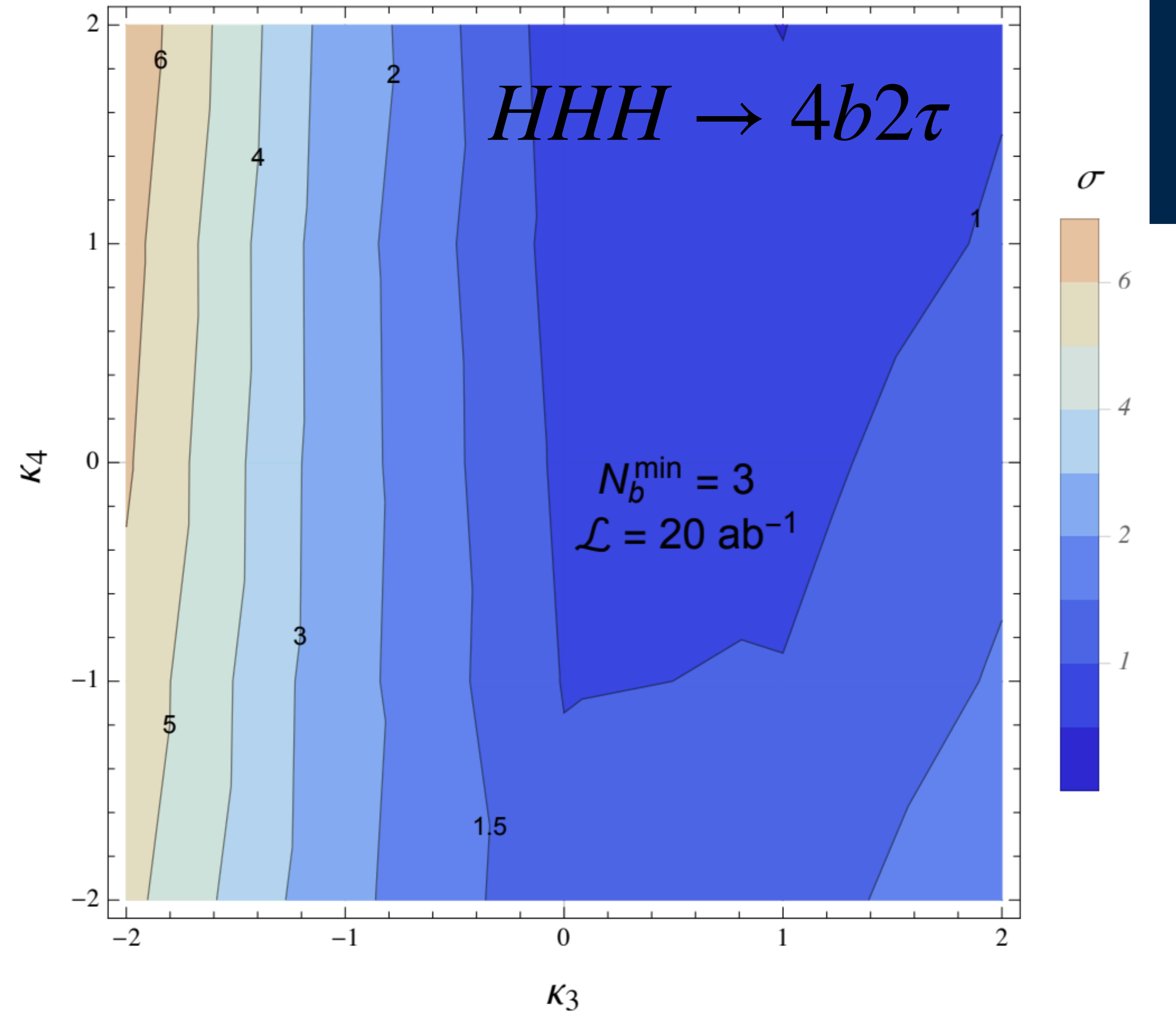


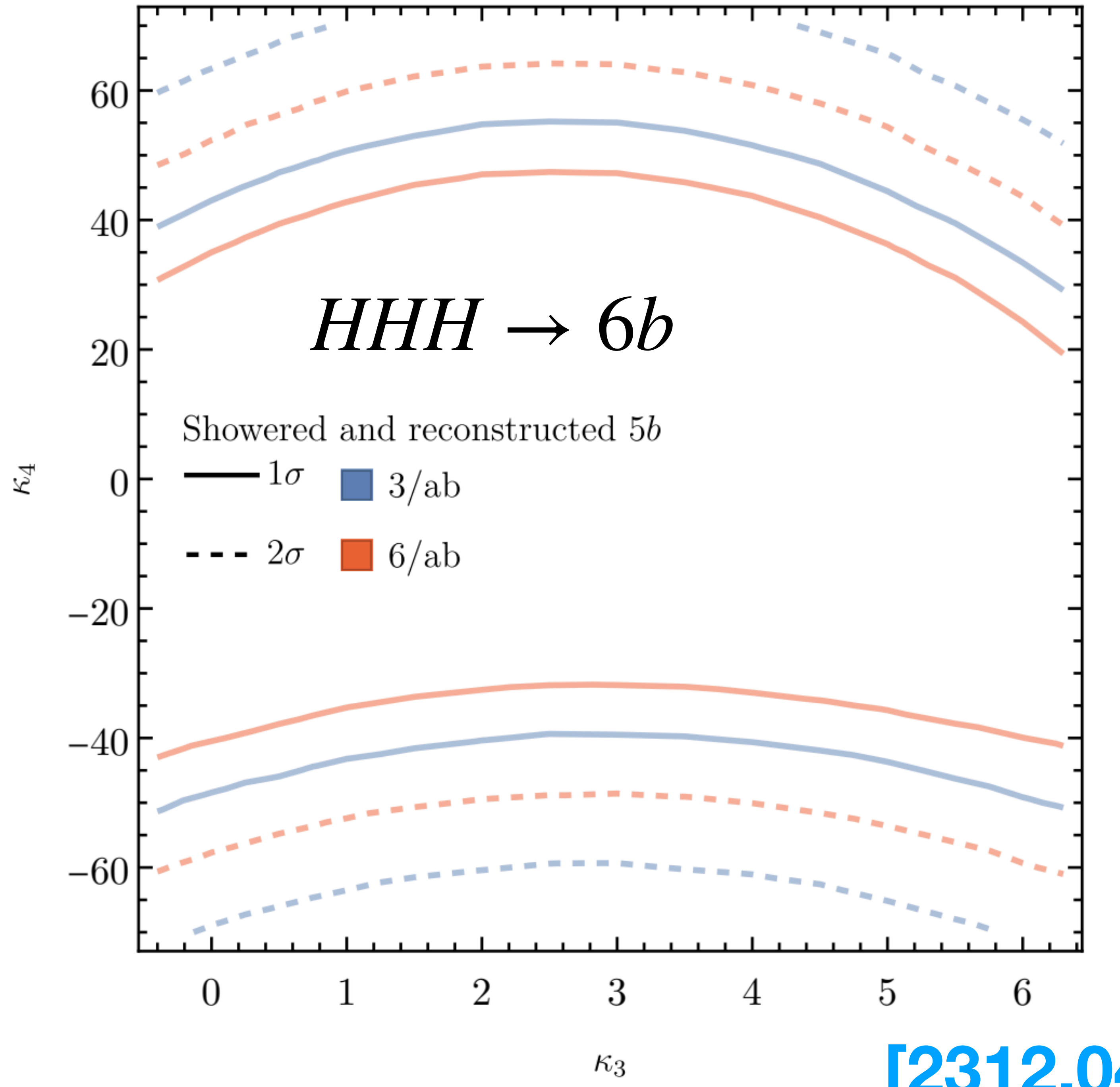
[2312.04646]

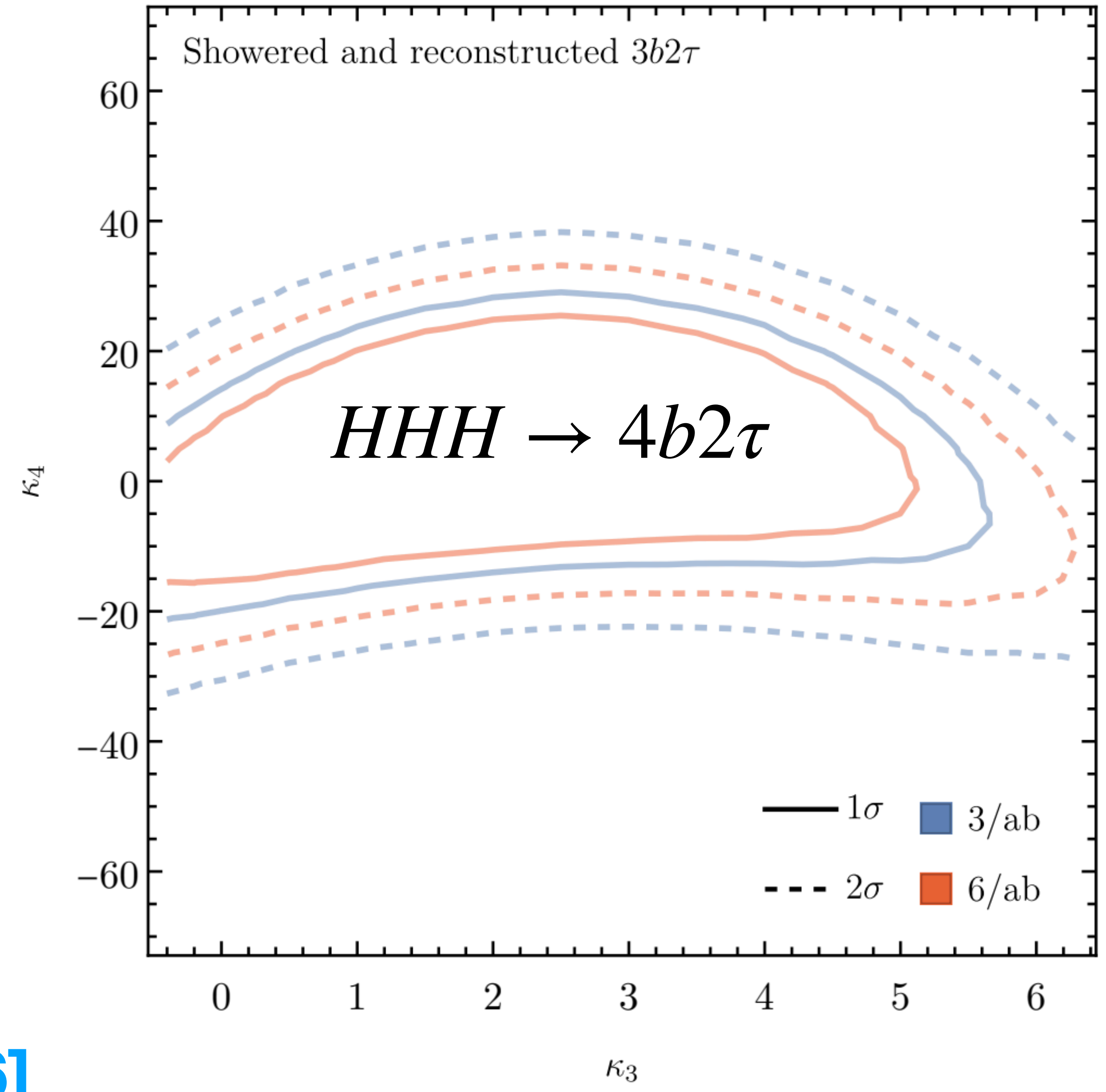
Decay modes

- **Studies on sensitivity of decay modes are in (slight) conflict**
- **Some (top) show $4b2\gamma$ is best** [1510.07697]
- **Others (bottom) show $4b2\tau$ is best**
- **ATLAS internal studies showed $6b$ was best**
  - May be because other studies used $5b$ SR
- $4b1\ell$ **is in my opinion under studied!**
  - Benefit from at least one lepton from many final states: $WW*$, $\tau\tau$, $ZZ*$
- **ATLAS chose $6b$ for 2 reasons:**
  1. Our projections showed this was best
  2. The team had expertise from $HH \rightarrow 4b$





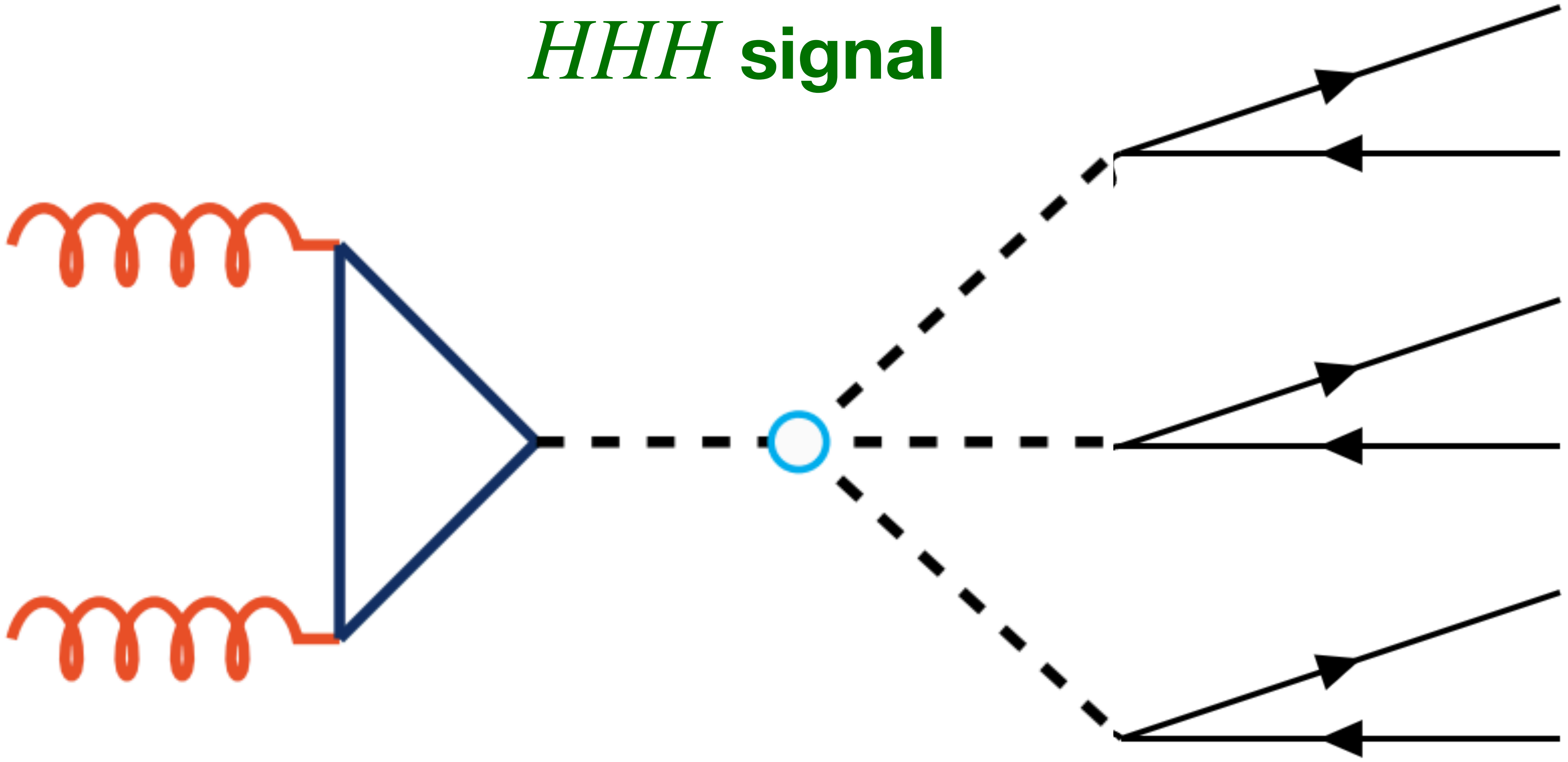




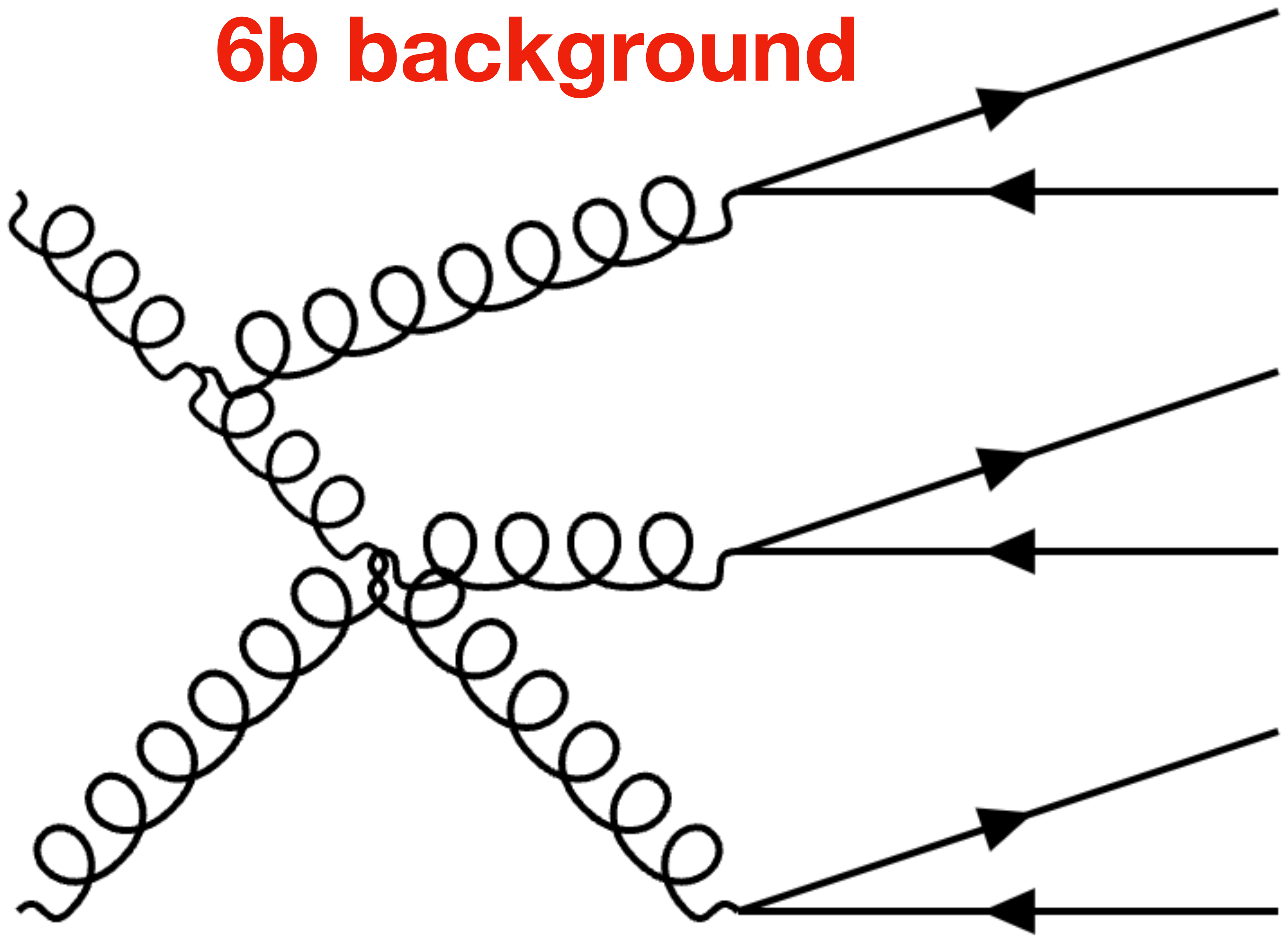
[2312.04646]

Signal and background

HHH **signal**



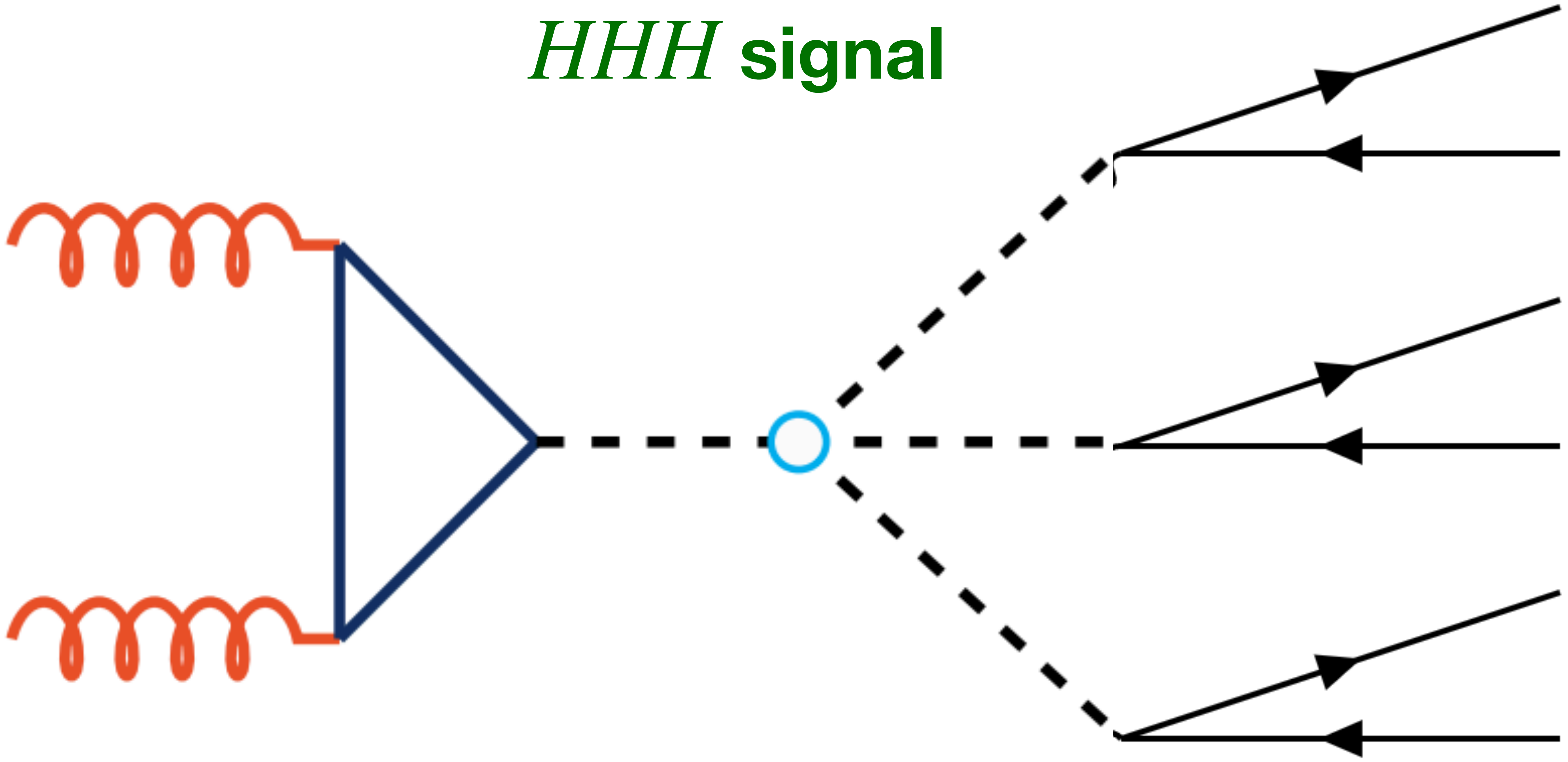
6b background



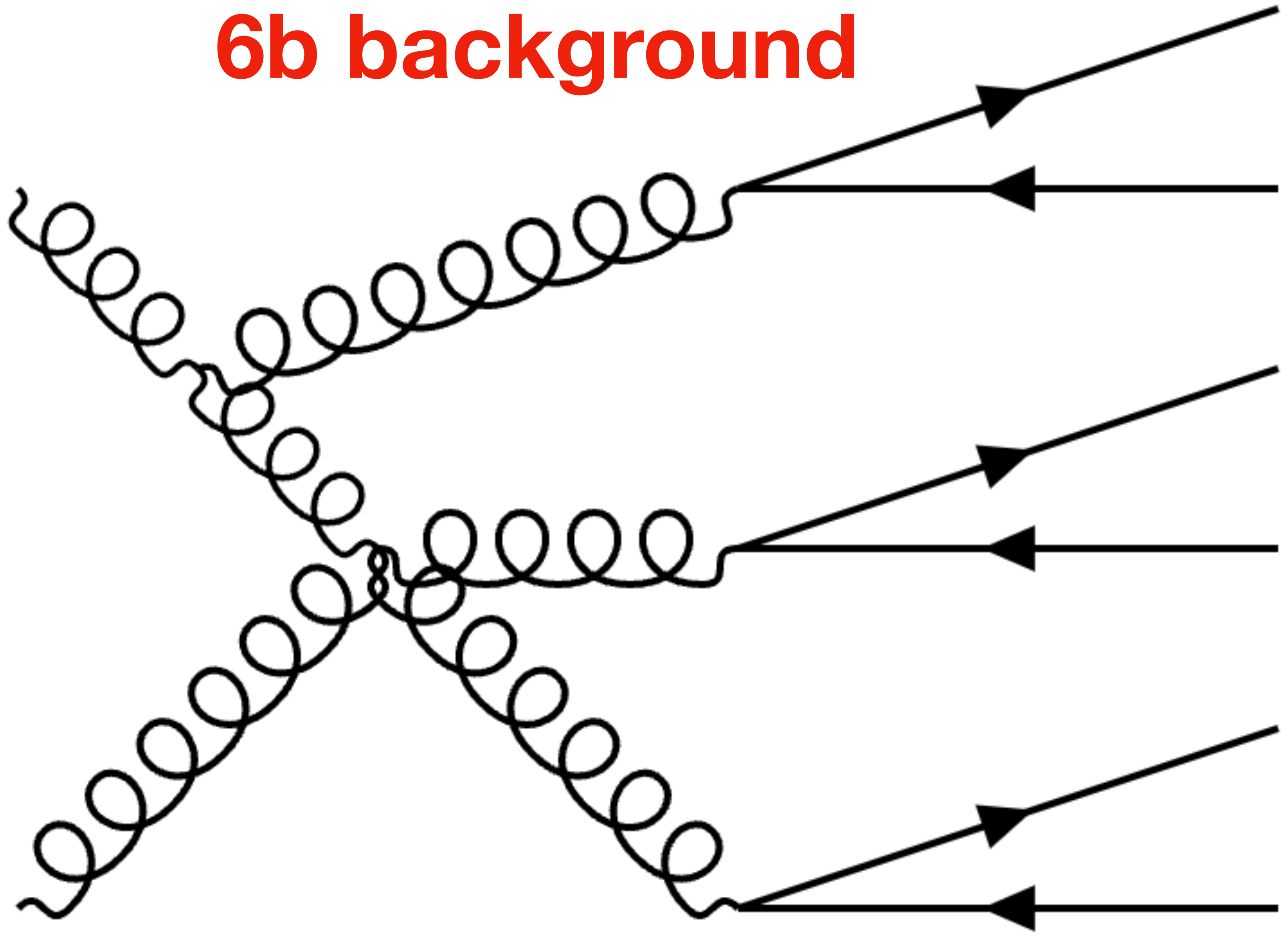
Signal and background

- We search in the 6b final state
- Primary background is from QCD multi-jet production

HHH signal

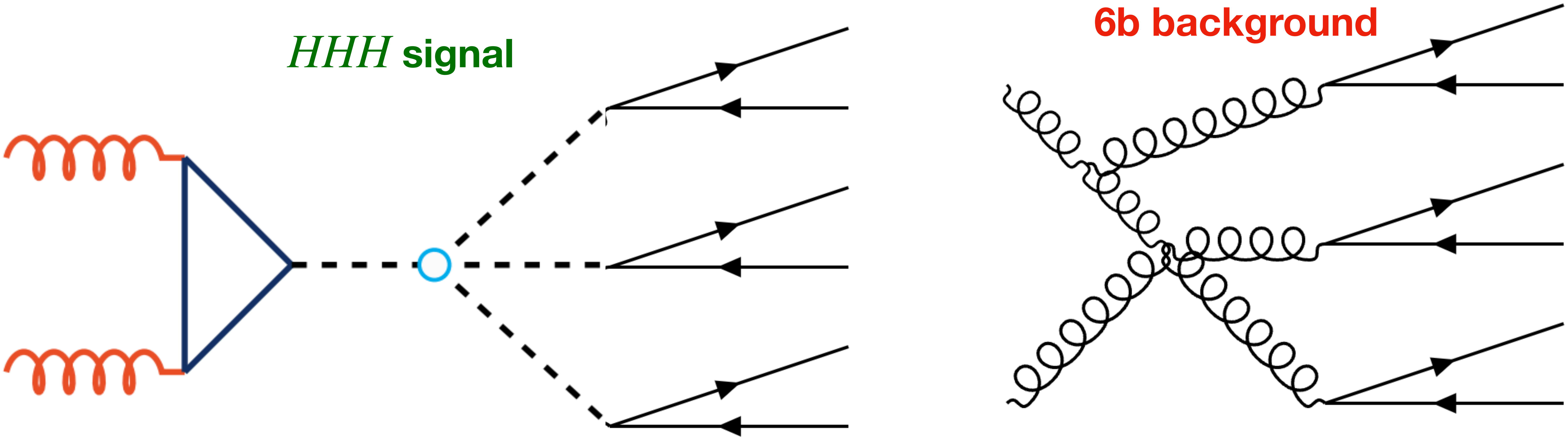


6b background



Signal and background

- We search in the 6b final state
- Primary background is from QCD multi-jet production
- Other backgrounds like $t\bar{t} + HF$ still have many additional b-jets and are difficult to model with MC
- Motivates a data-driven background estimate



Event selection

- Trigger on the jets/b-jets (some online b-tagging is used in the trigger)
- Select all events to have at least 6 jets, categorize into the 4b, 5b, 6b regions

Events

- $\geqslant$ 6 jets
- $\geqslant$ 4 jets with p_T > 40 GeV
- $\geqslant$ 4 jets passing DL1dv01 77% working point

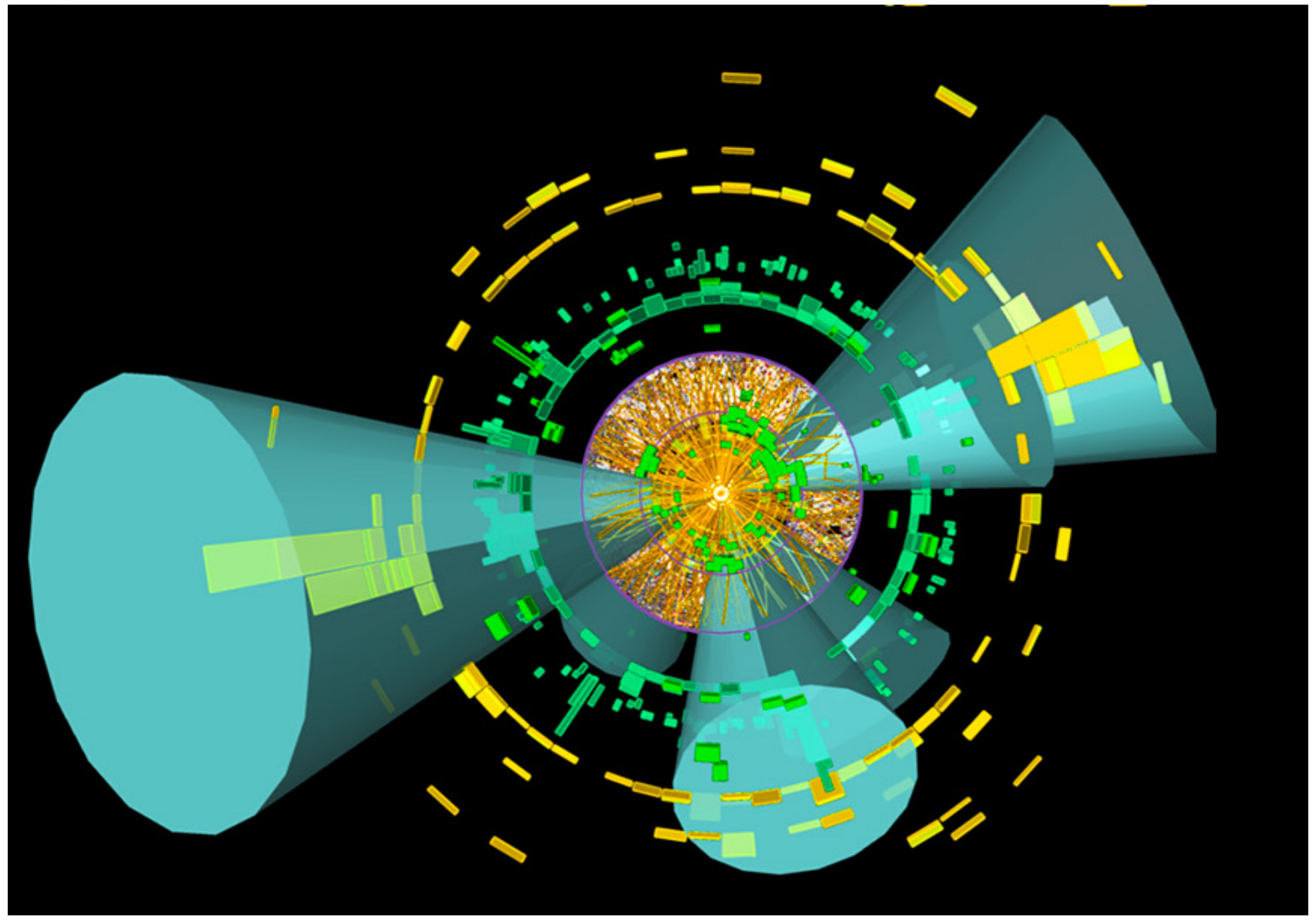
$\geqslant$ 6 *b*-tags
Signal Region

== 5 *b*-tags
Background Estimate

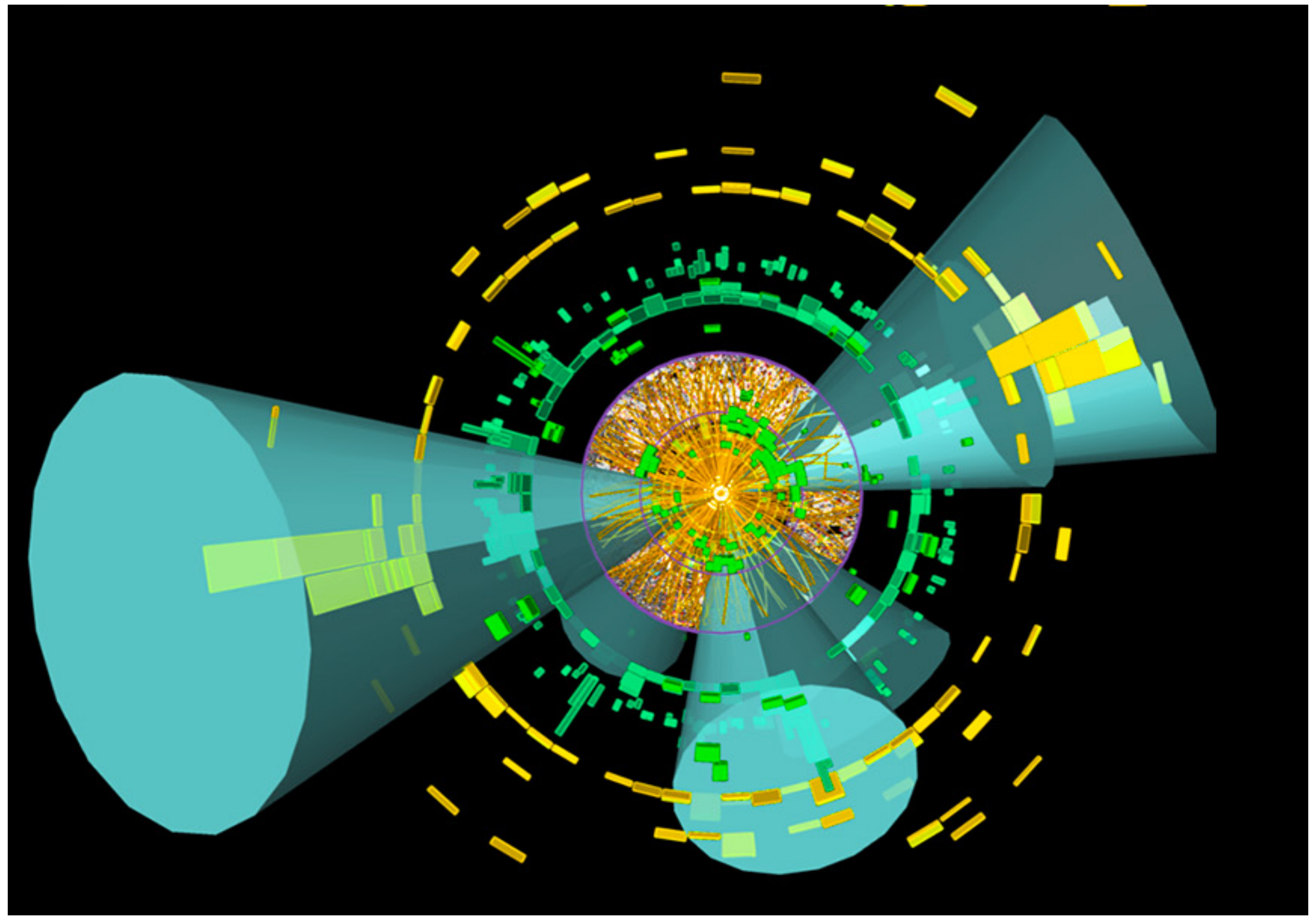
== 4 *b*-tags
Background Estimate/Validation


Example 6b event
Run: 349592
Event: 793710422
2018-05-06 03:37:28 UTC

Analysis strategy

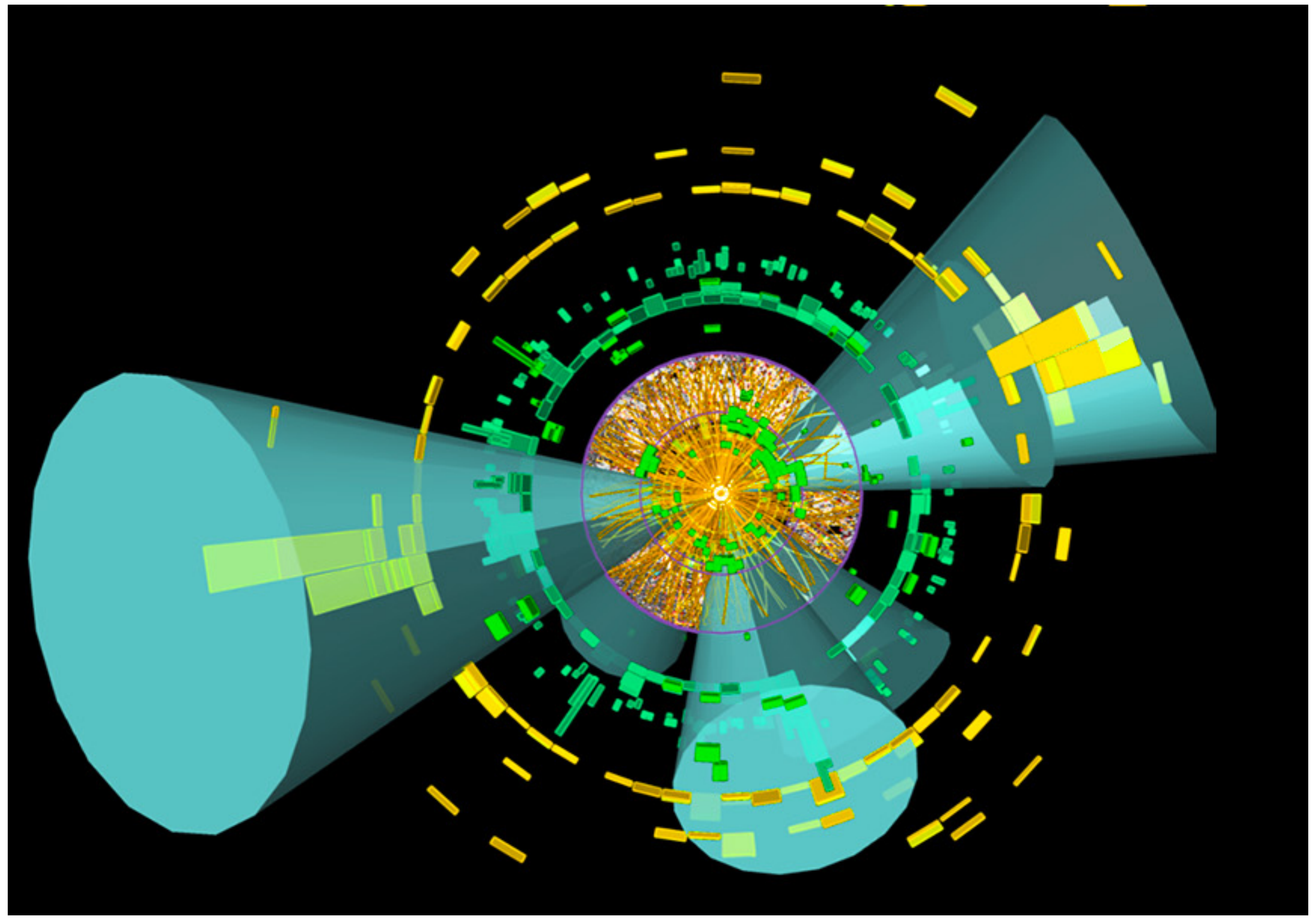


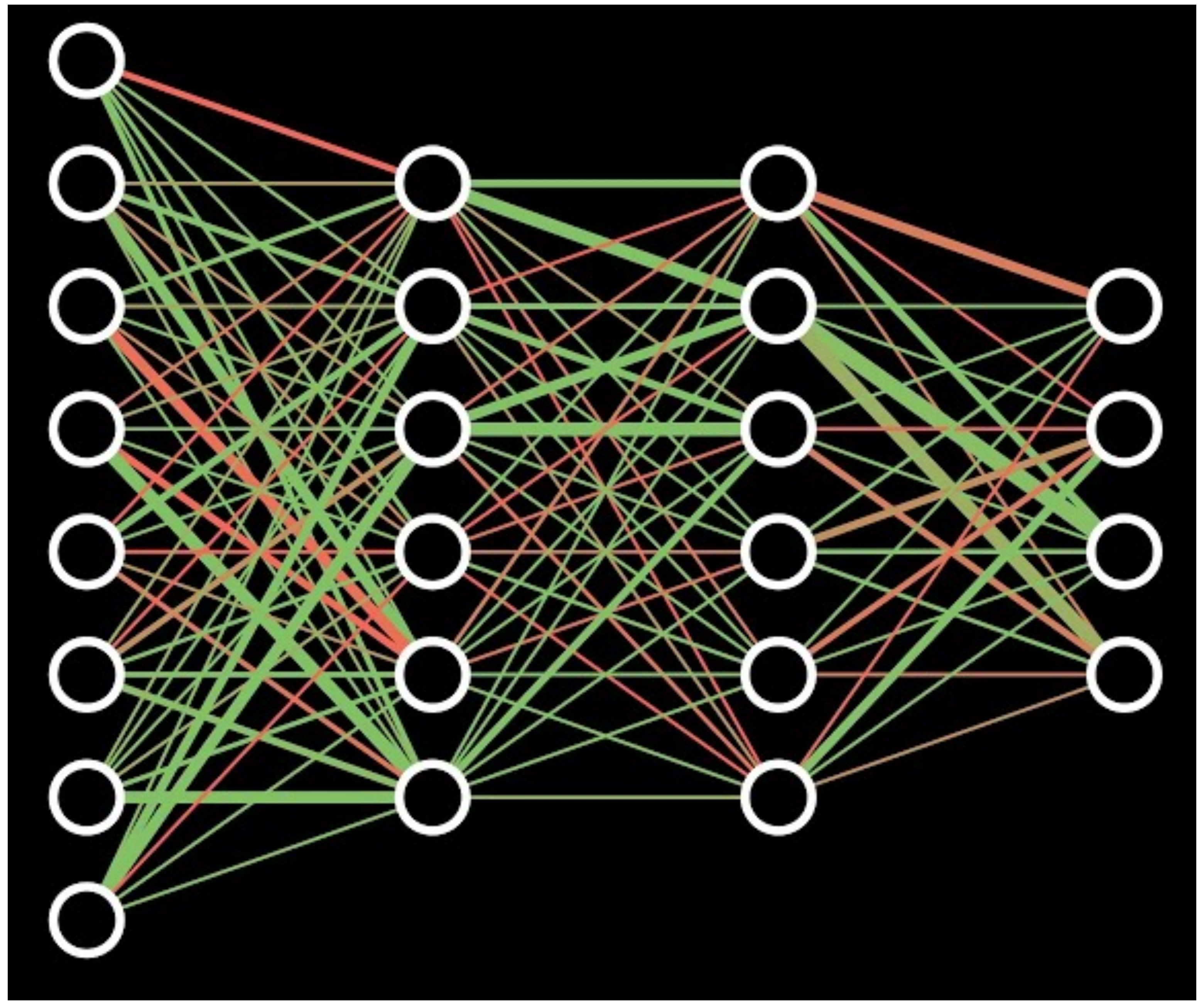
Analysis strategy



- All events have 6 reconstructed jets. Some have 4, 5, or 6 b-tagged jets.

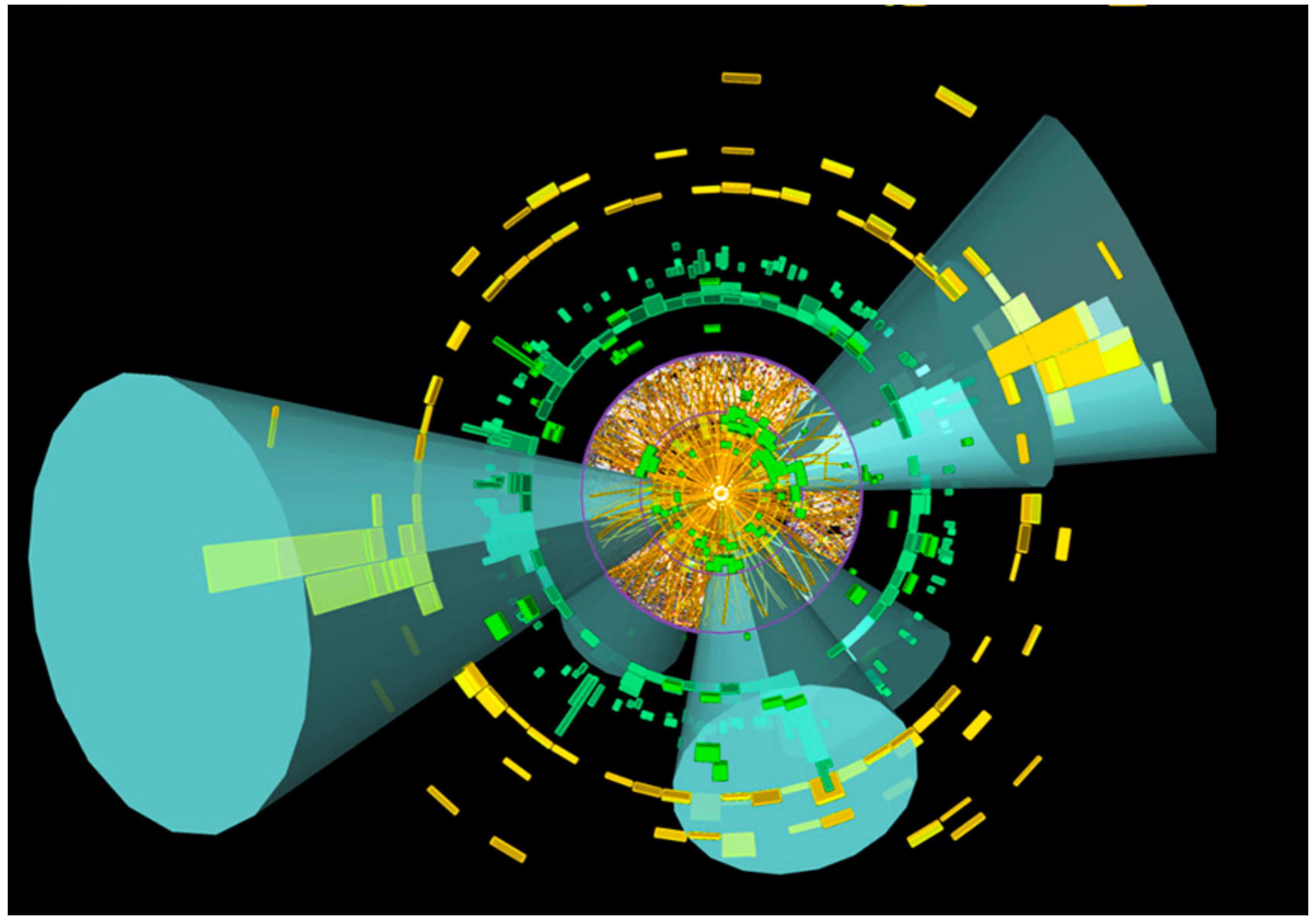
Analysis strategy

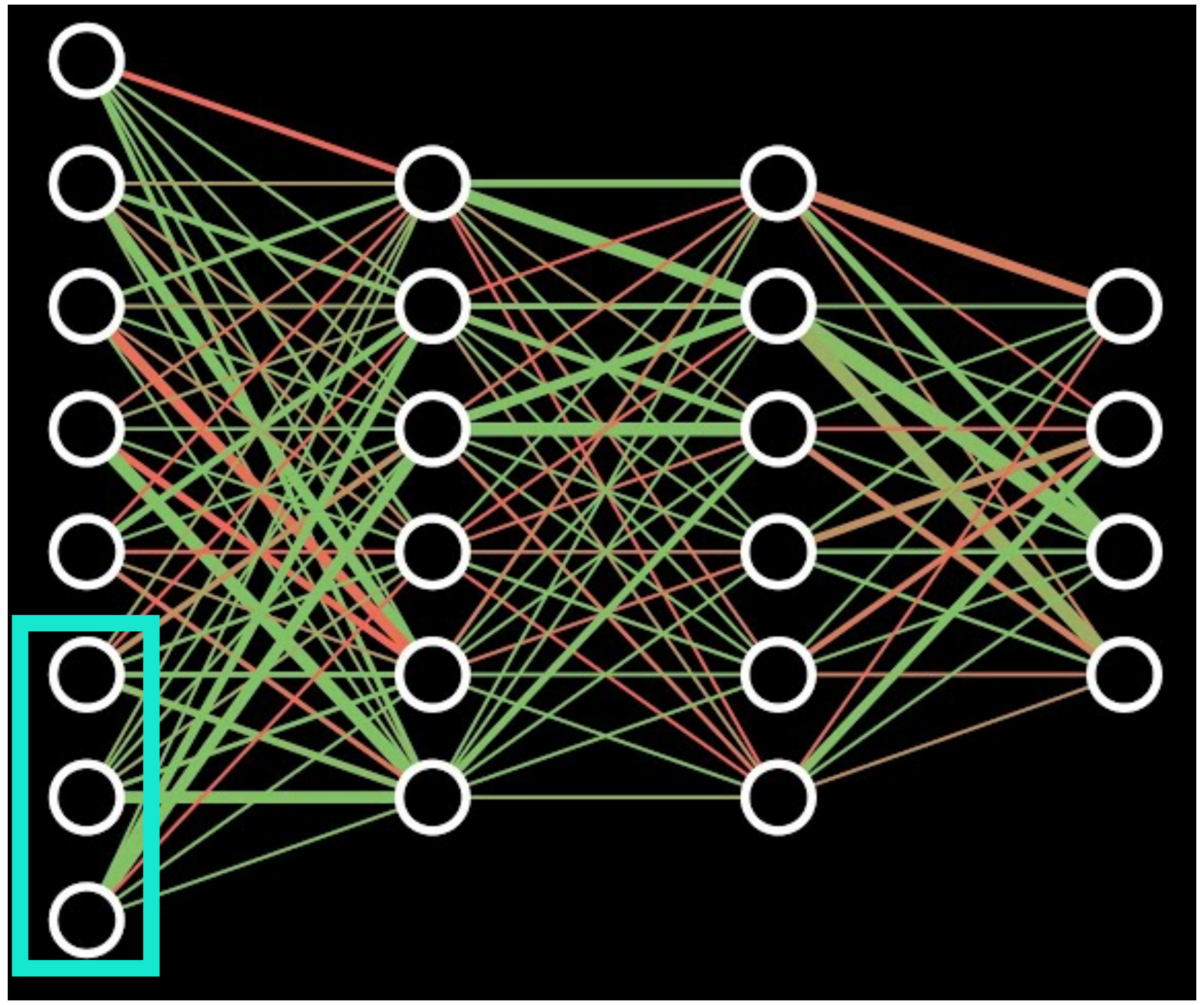




- All events have 6 reconstructed jets. Some have 4, 5, or 6 b-tagged jets.
- Train a DNN to discriminate between HHH signal and 5b background.

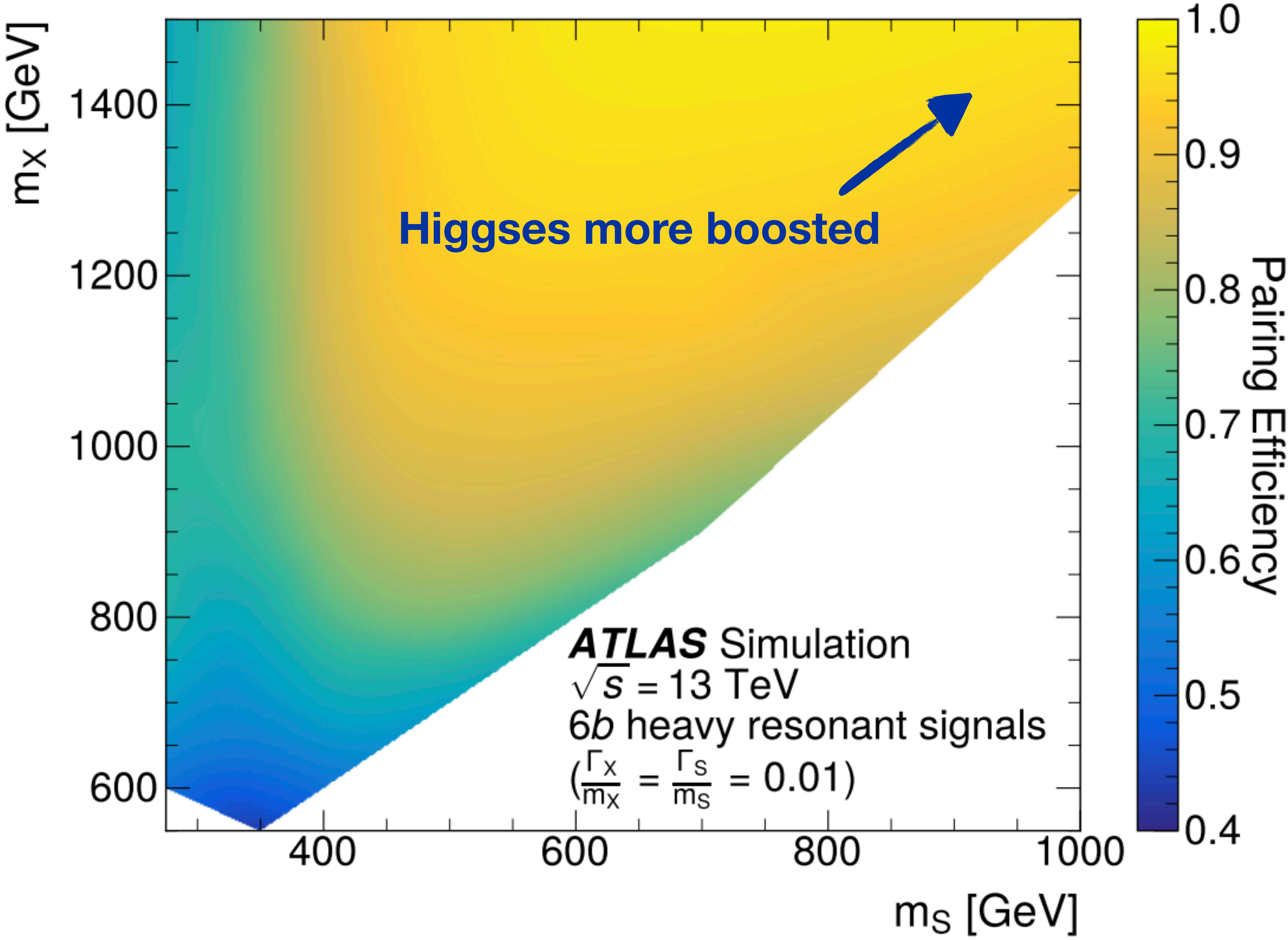
Analysis strategy





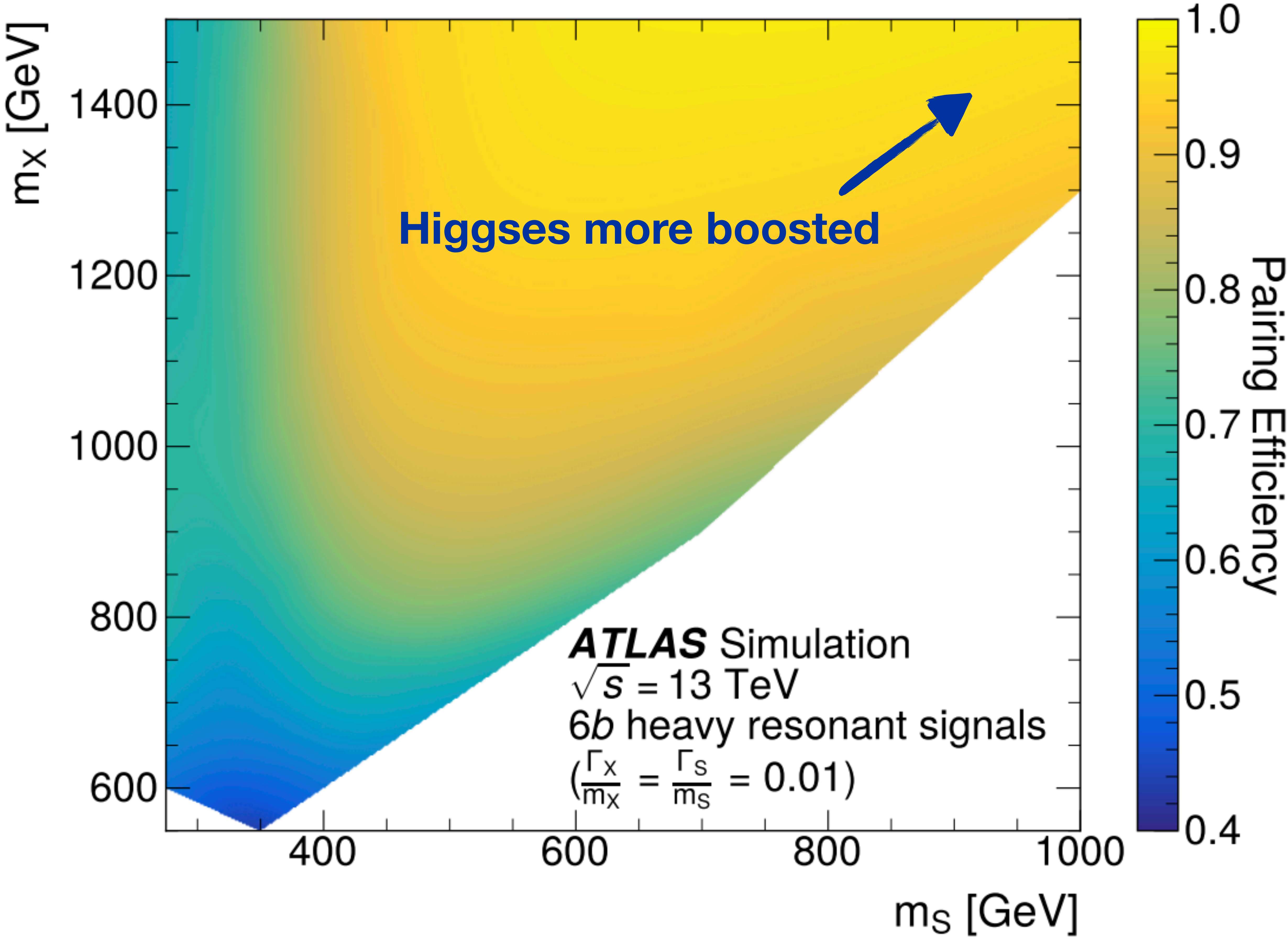
- **All events have 6 reconstructed jets. Some have 4, 5, or 6 b-tagged jets.**
- **Train a DNN to discriminate between HHH signal and 5b background.**
  - Some of the DNN inputs are reliant on paired jets, for example to compute m(bb) for the three pairs of jets.

b-jet pairing



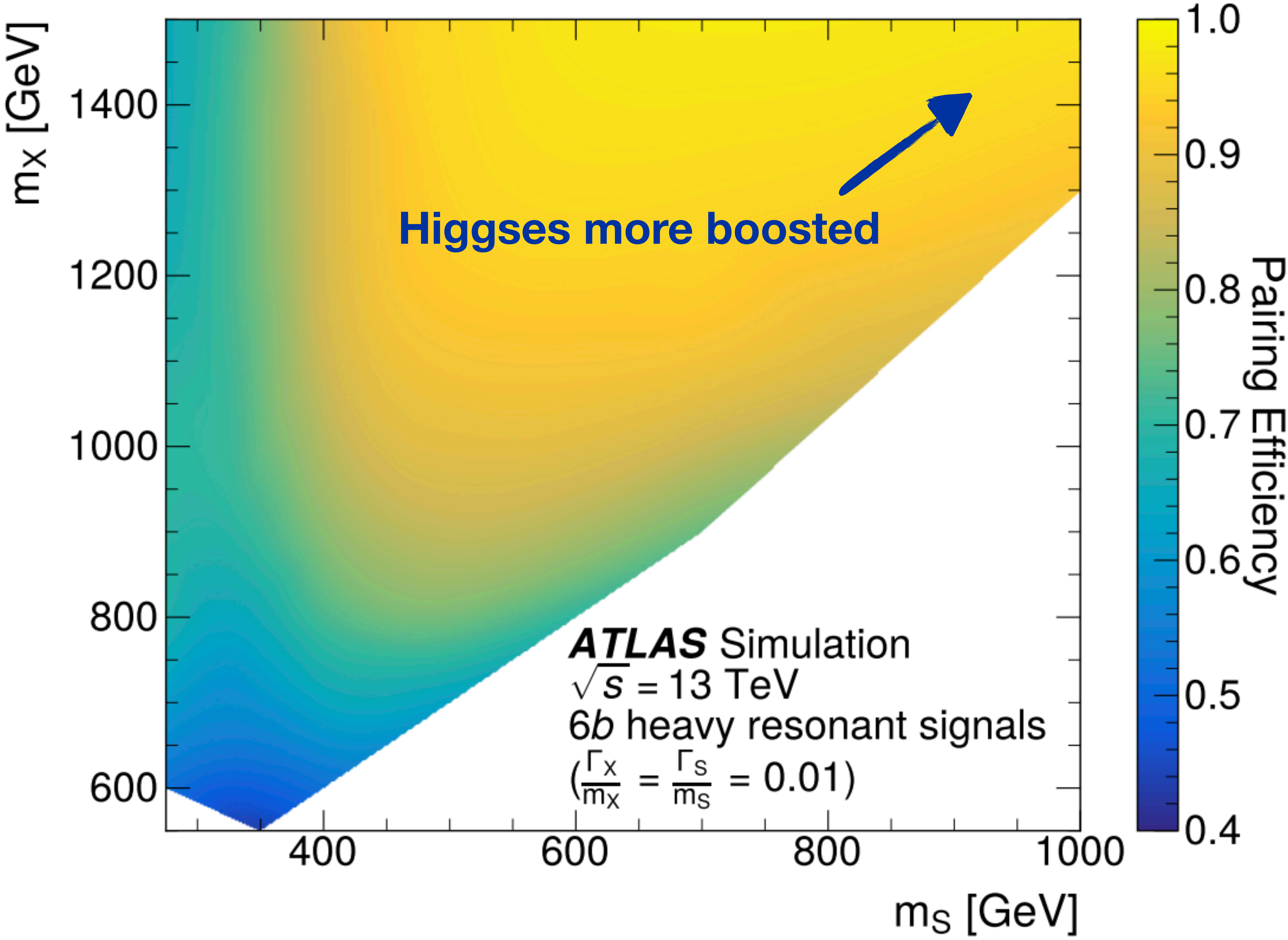
b-jet pairing

- **Some variables in the DNN will be based on paired jets**



b-jet pairing

- **Some variables in the DNN will be based on paired jets**
- **There are 15 ways to pair 6 jets**

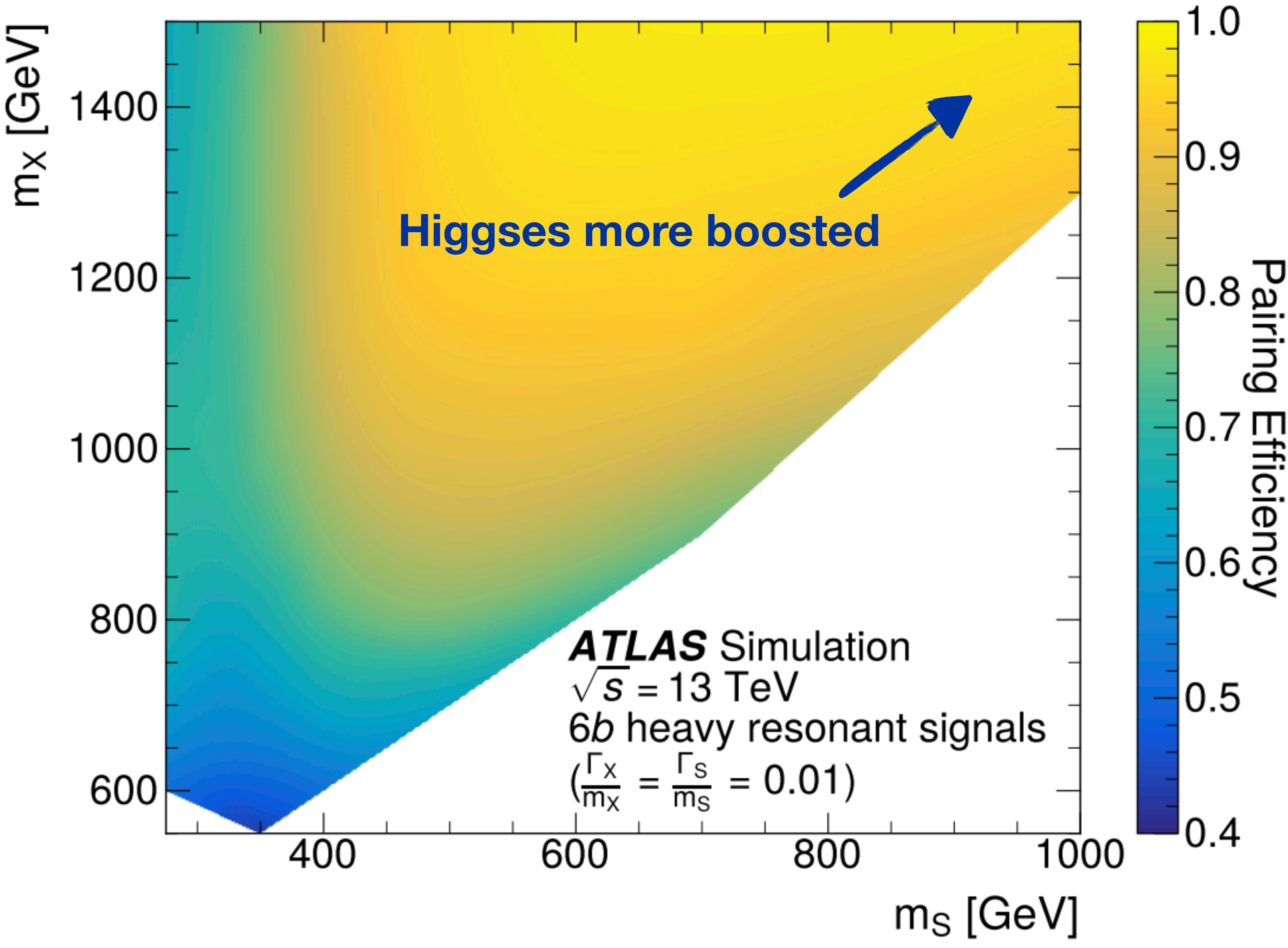


b-jet pairing

- **Some variables in the DNN will be based on paired jets**
- **There are 15 ways to pair 6 jets**
- **We use a mass-based pairing with additional constraint on p_T-ordering, reducing from 15 possibilities**

$$|m_{H1} - 120\,\mathrm{GeV}| + |m_{H2} - 115\,\mathrm{GeV}| + |m_{H3} - 110\,\mathrm{GeV}|,$$

Leading **Subleading** **Third**



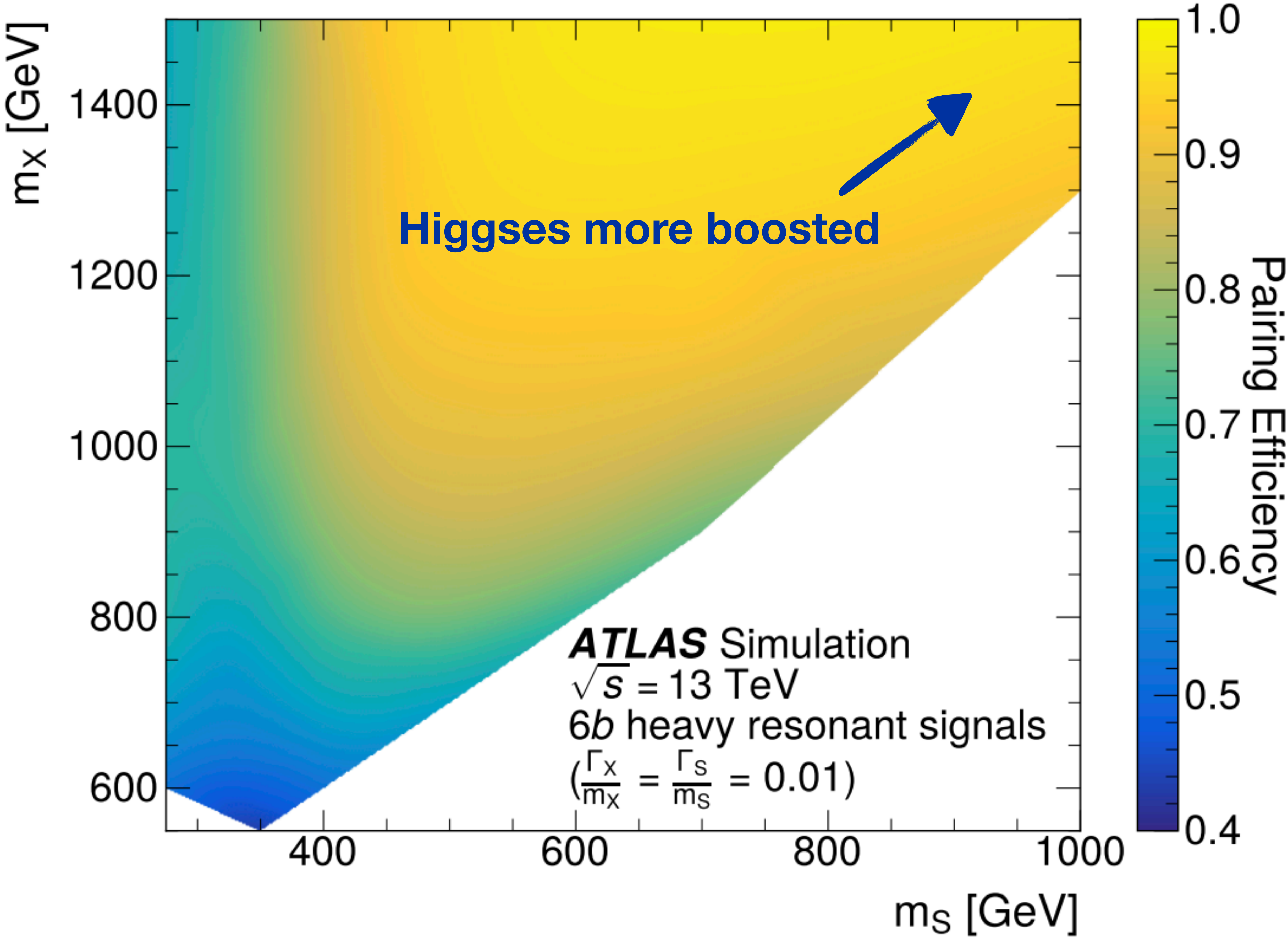
b-jet pairing

- **Some variables in the DNN will be based on paired jets**
- **There are 15 ways to pair 6 jets**
- **We use a mass-based pairing with additional constraint on p_T-ordering, reducing from 15 possibilities**

$$|m_{H1} - 120\,\mathrm{GeV}| + |m_{H2} - 115\,\mathrm{GeV}| + |m_{H3} - 110\,\mathrm{GeV}|,$$

Leading **Subleading** **Third**

- **This is empirical!**



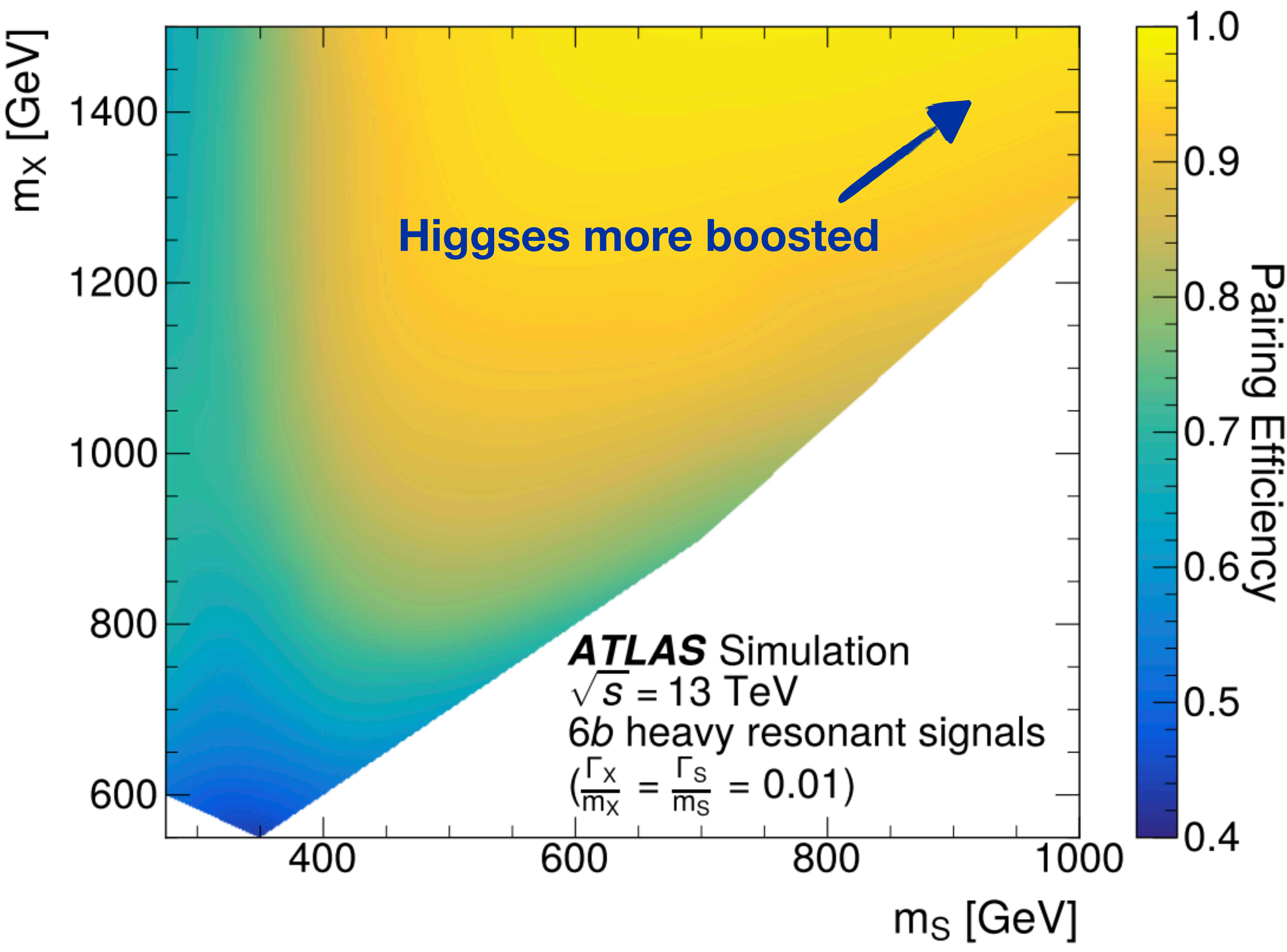
b-jet pairing

- **Some variables in the DNN will be based on paired jets**
- **There are 15 ways to pair 6 jets**
- **We use a mass-based pairing with additional constraint on p_T-ordering, reducing from 15 possibilities**

$$|m_{H1} - 120\,\mathrm{GeV}| + |m_{H2} - 115\,\mathrm{GeV}| + |m_{H3} - 110\,\mathrm{GeV}|,$$

Leading **Subleading** **Third**

- **This is empirical!**
- **Performs very well in the boosted regime.**
  - Still far better than random chance when not boosted, but that is a low bar.



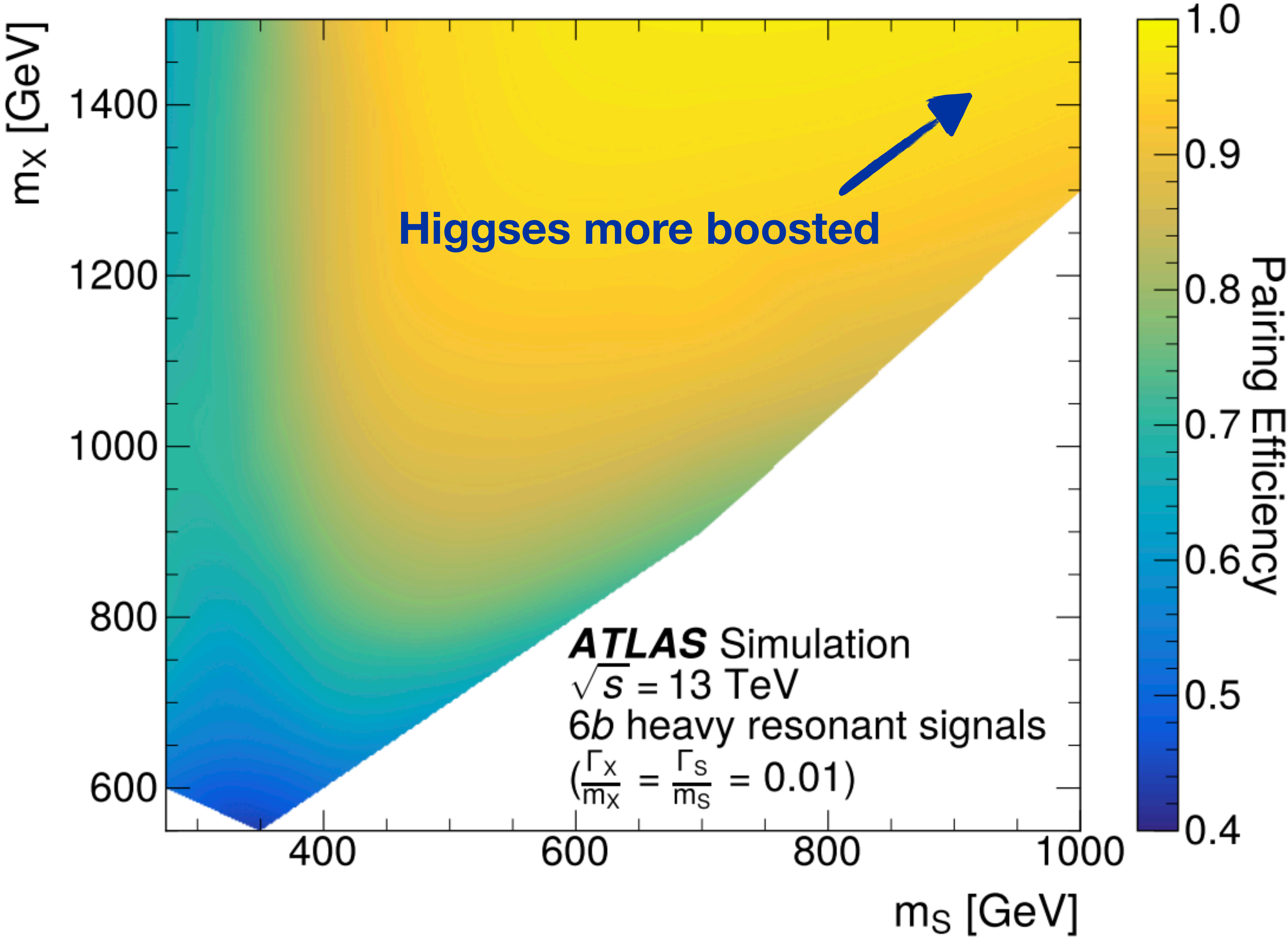
b-jet pairing

- **Some variables in the DNN will be based on paired jets**
- **There are 15 ways to pair 6 jets**
- **We use a mass-based pairing with additional constraint on p_T-ordering, reducing from 15 possibilities**

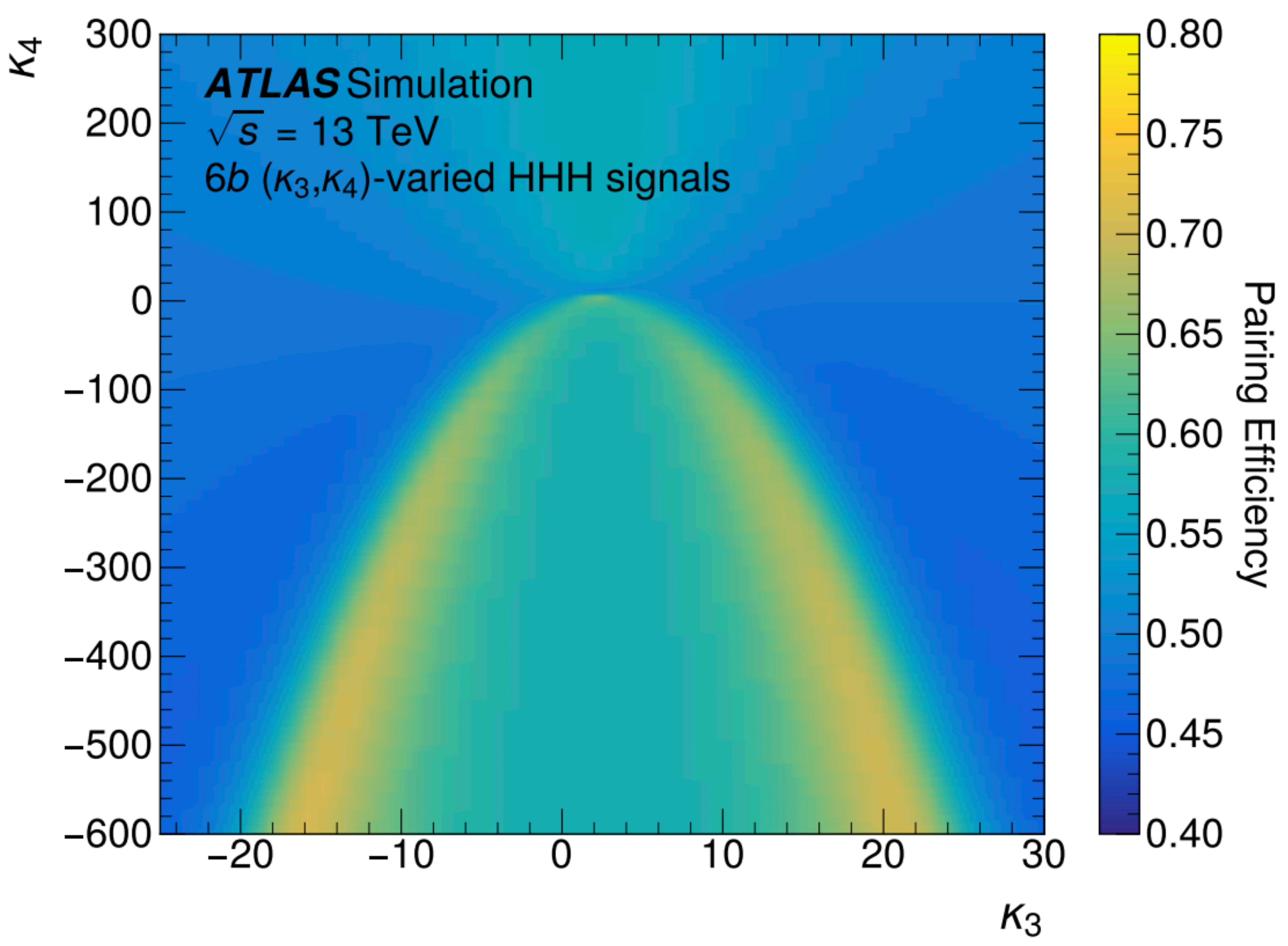
$$|m_{H1} - 120\,\text{GeV}| + |m_{H2} - 115\,\text{GeV}| + |m_{H3} - 110\,\text{GeV}|,$$

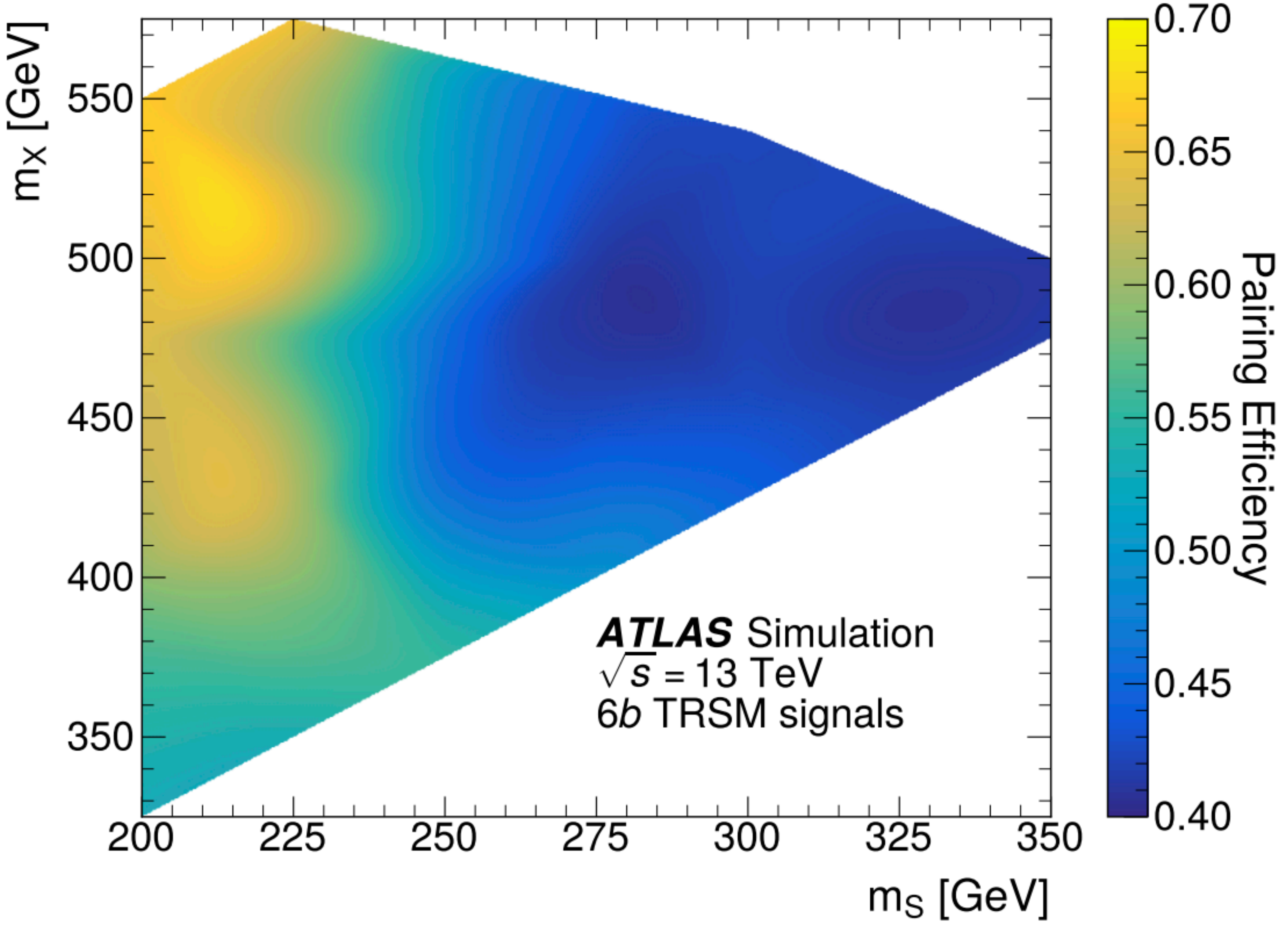
Leading Subleading Third

- **This is empirical!**
- **Performs very well in the boosted regime.**
  - Still far better than random chance when not boosted, but that is a low bar.
- **Pairing algorithms are one area for future R&D on HHH**



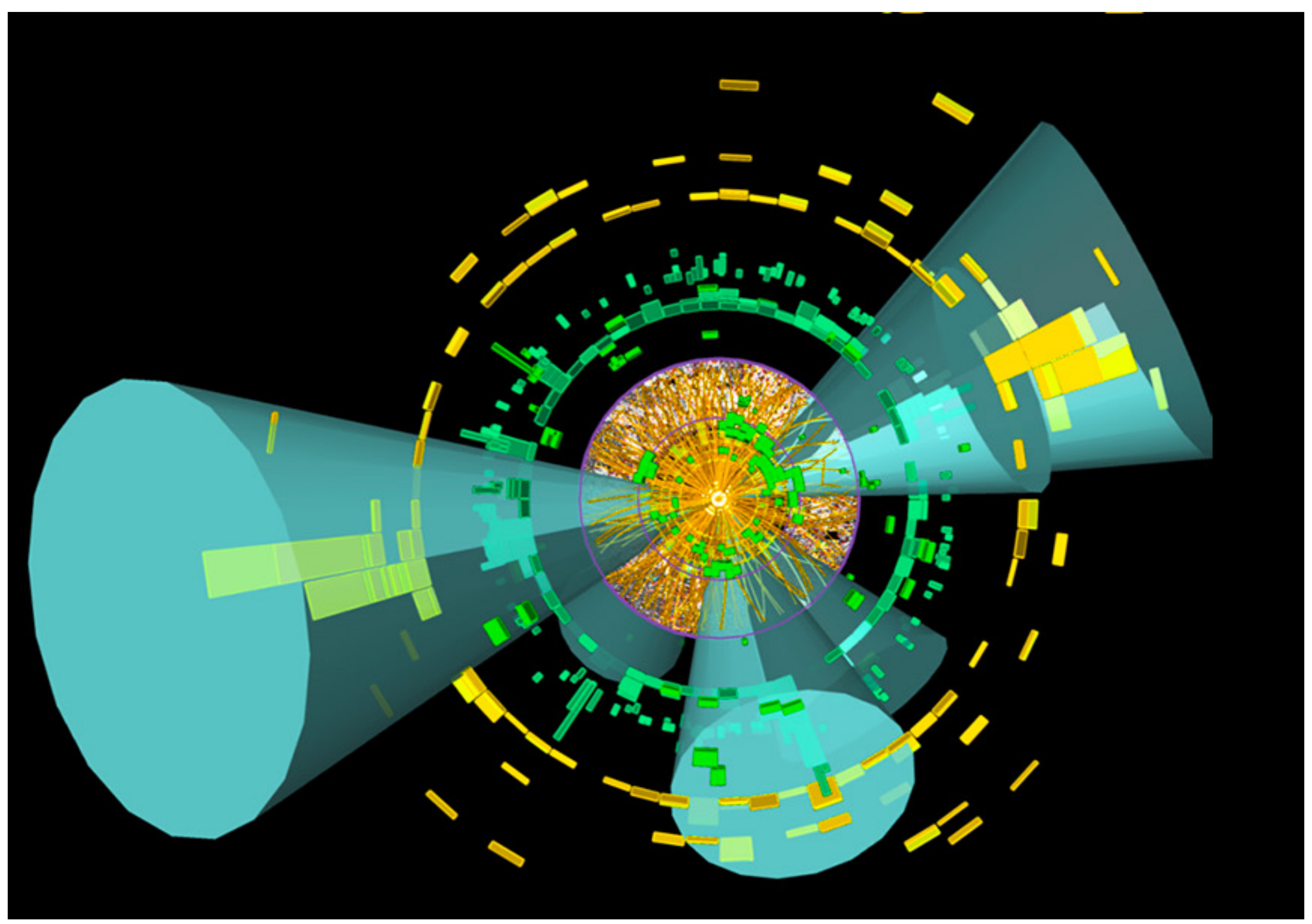
b-jet pairing

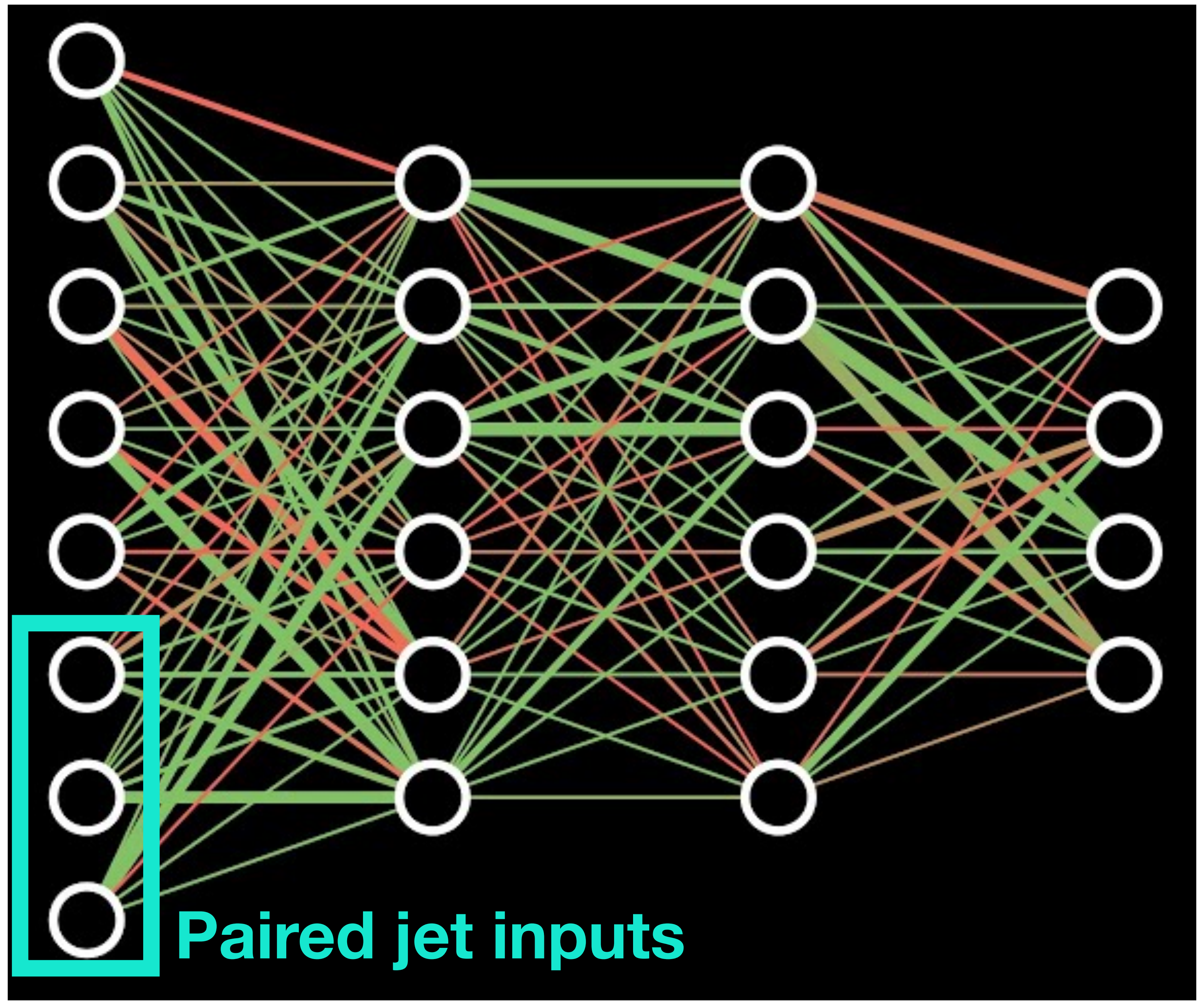




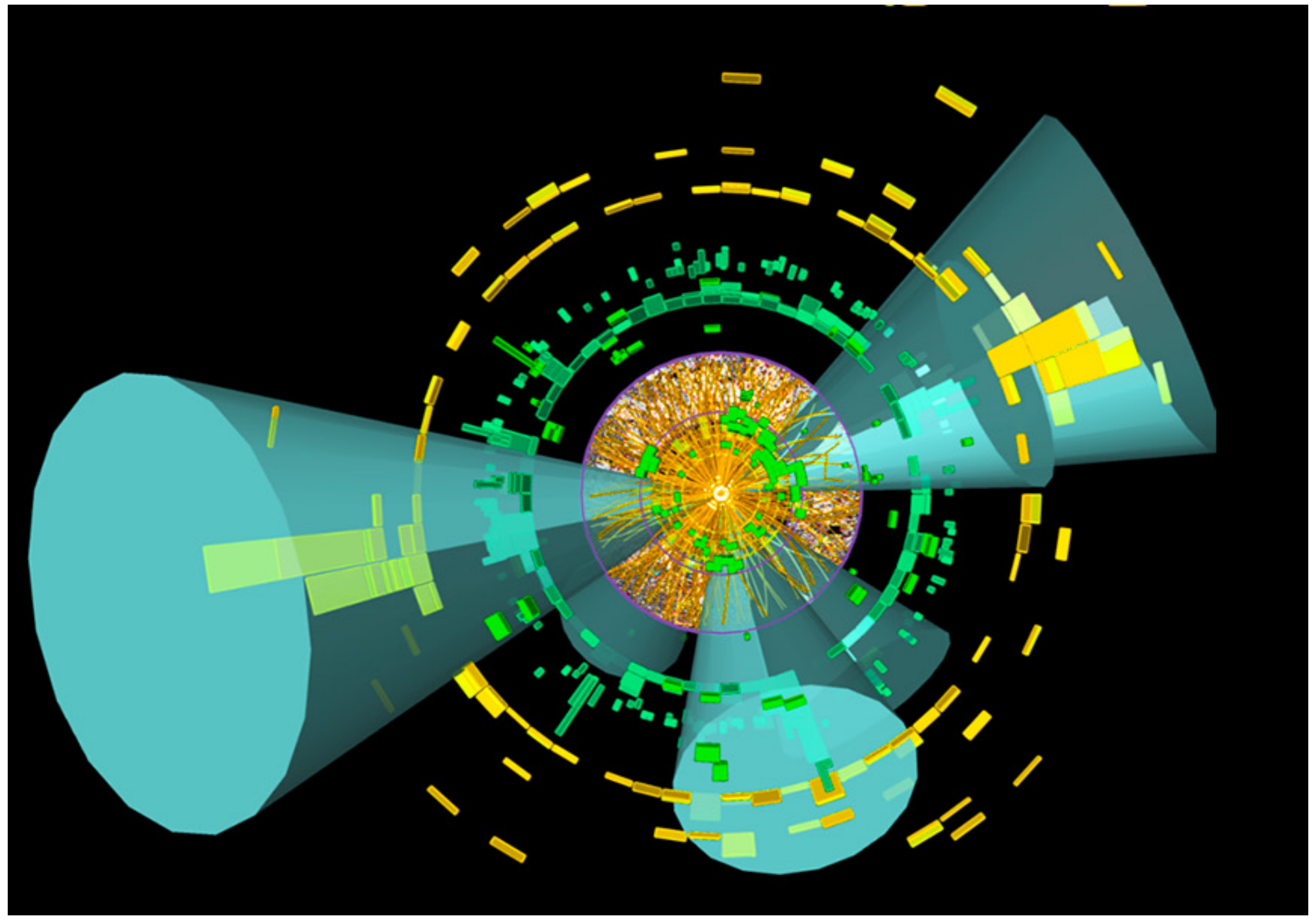
- **Without large boost, pairing efficiency is as low as 40%**
- **SM pairing efficiency: 51%**

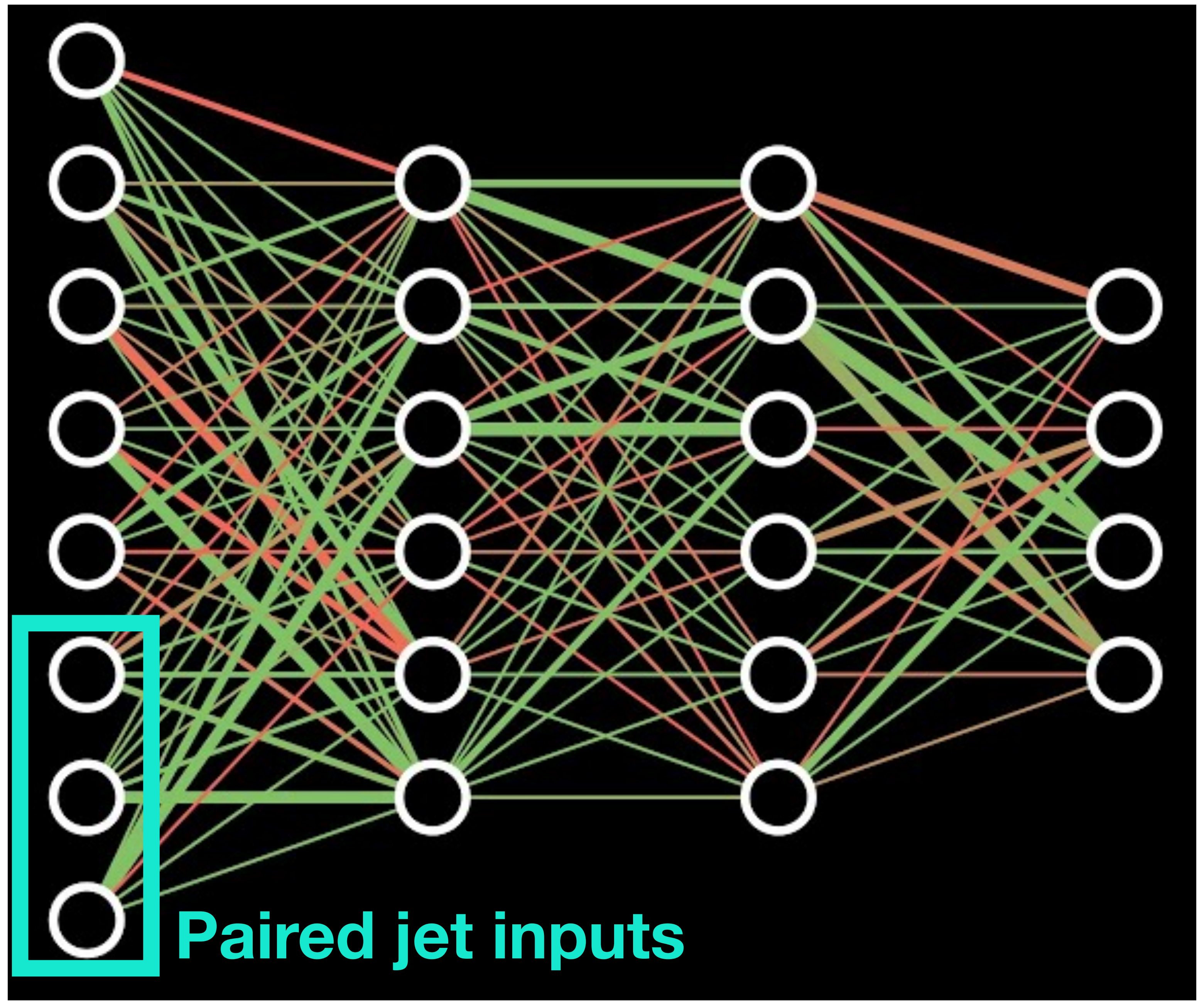
Analysis strategy

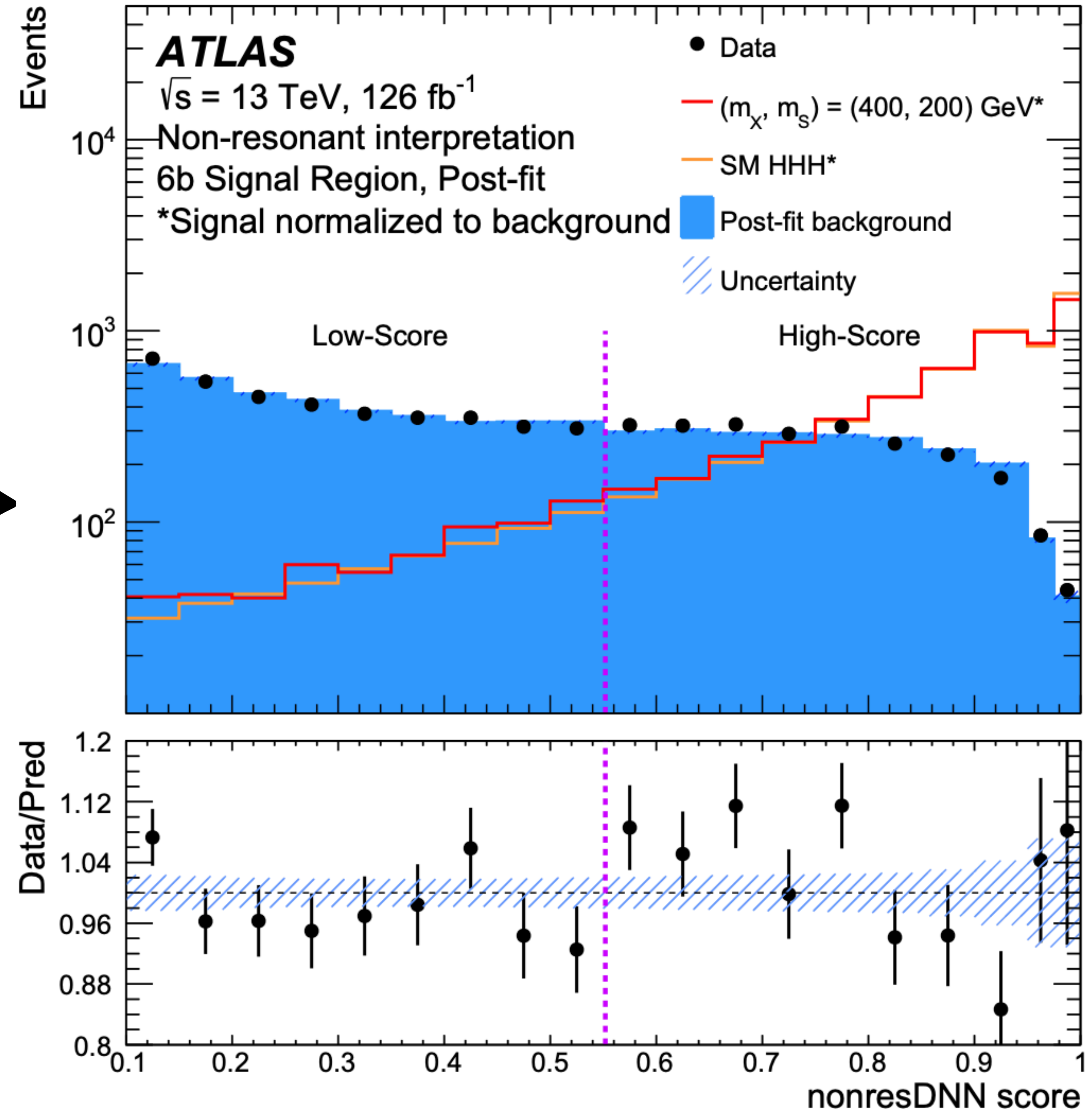




Analysis strategy

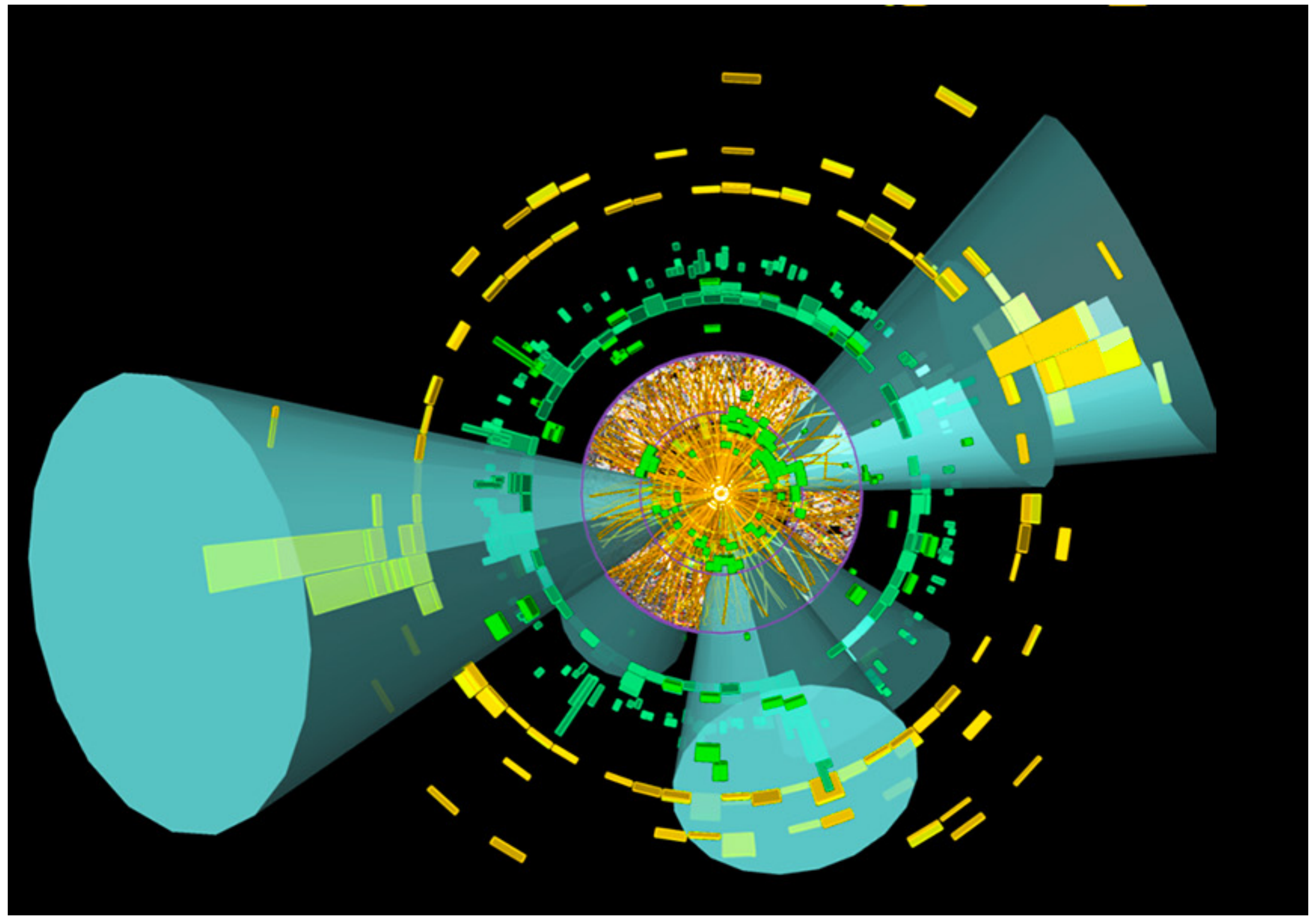


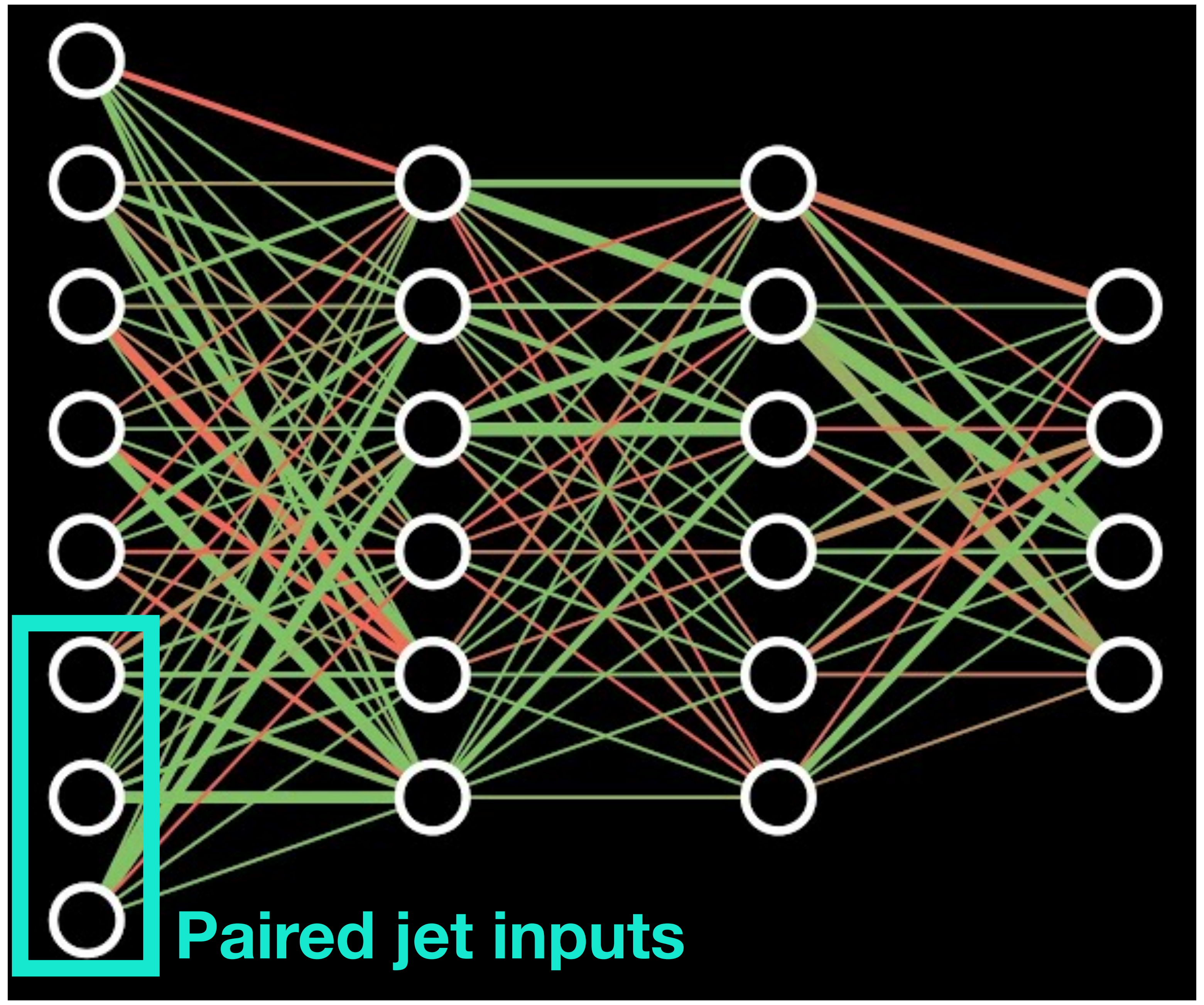


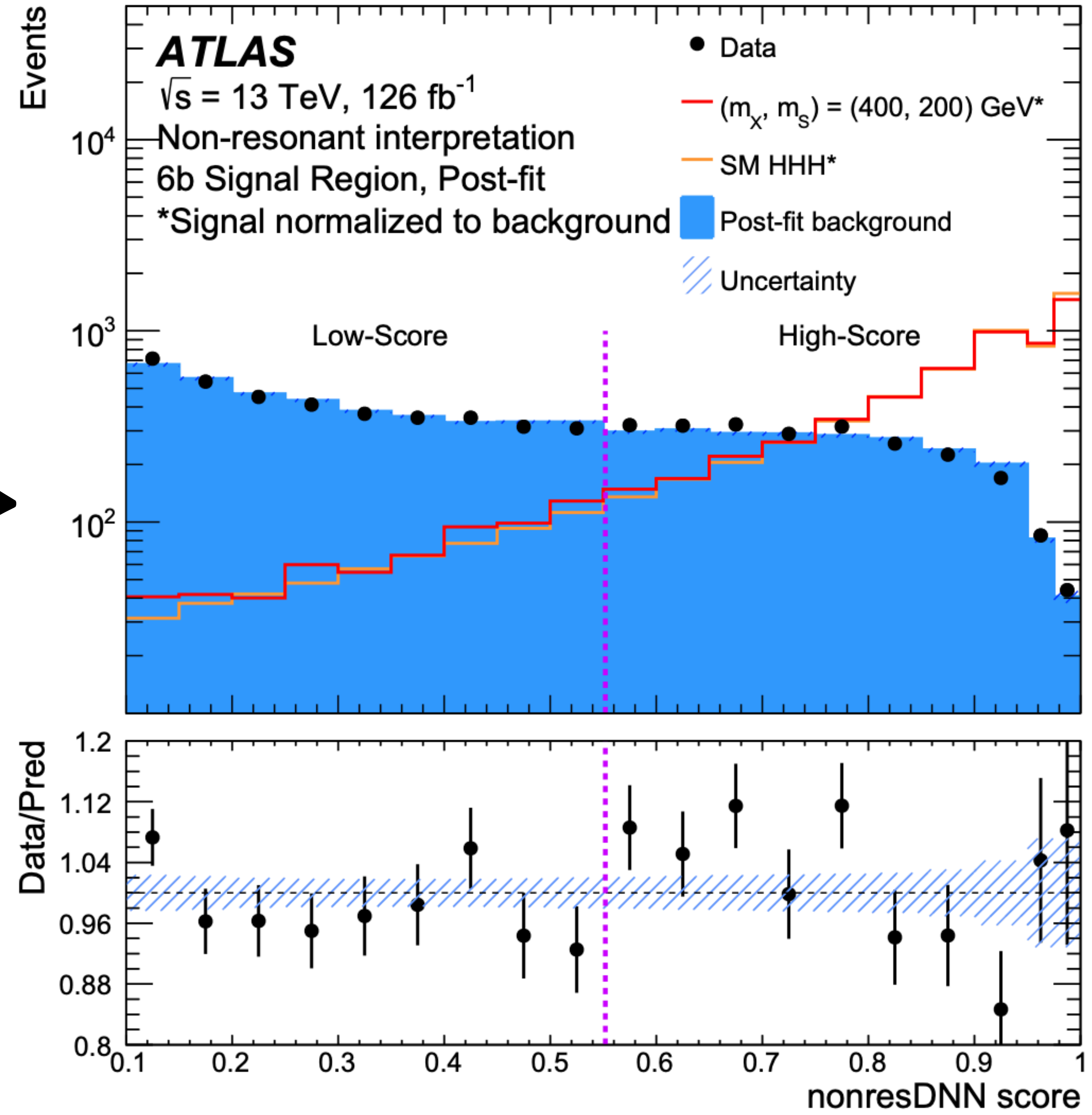


- **The DNN output score is a discriminant we fit and compute p-values/limits.**

Analysis strategy





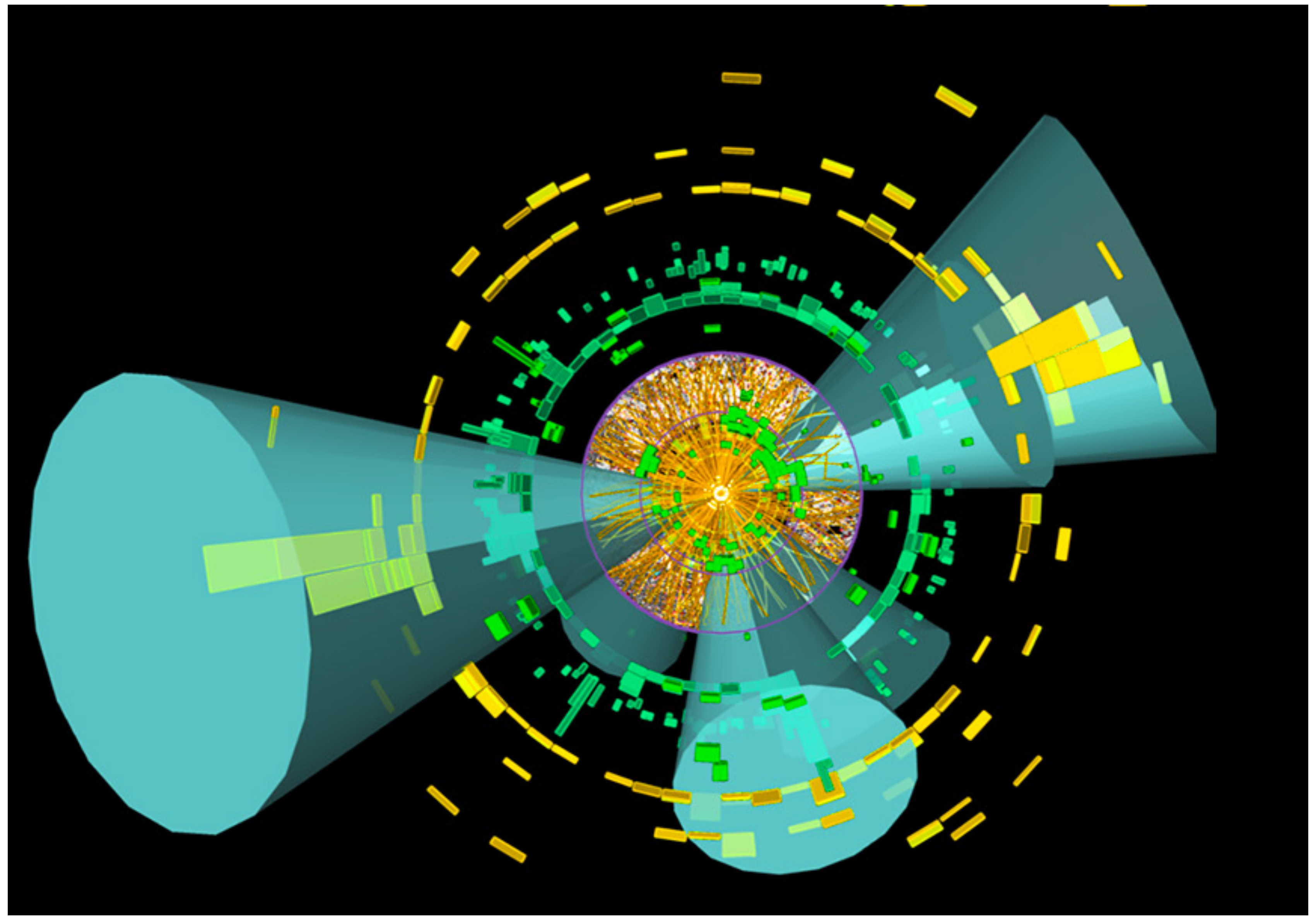


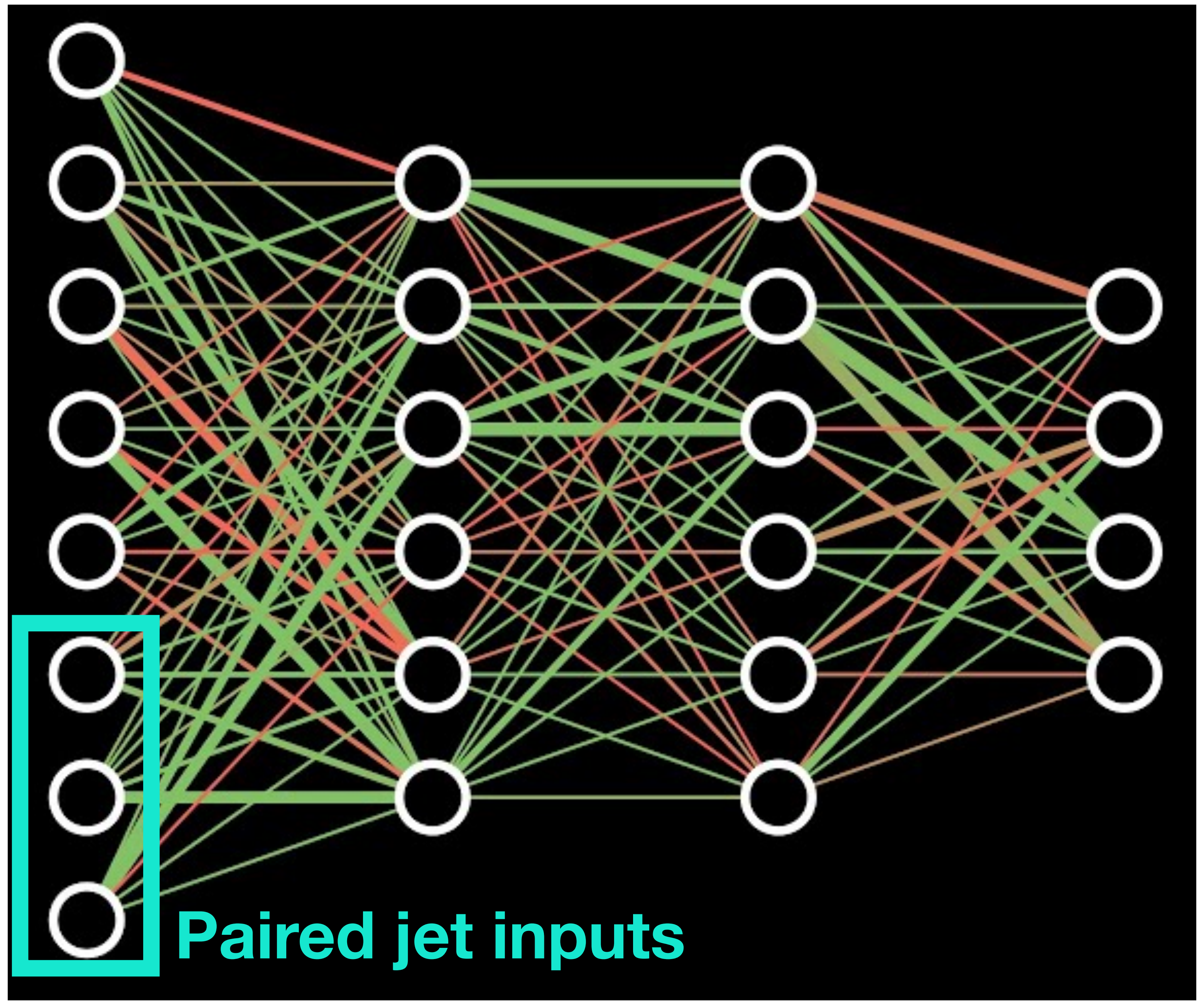
- **The DNN output score is a discriminant we fit and compute p-values/limits.**

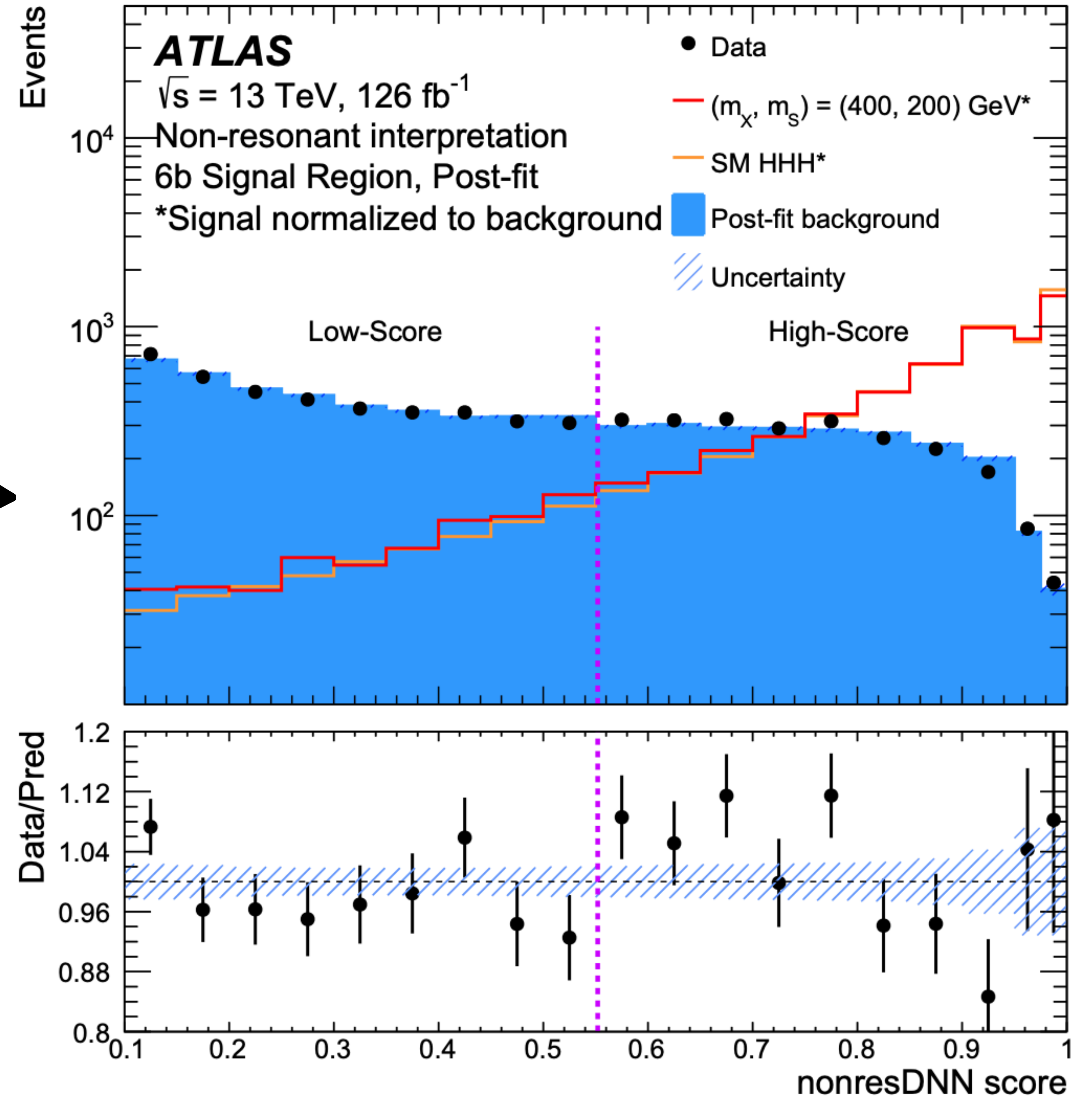
- **The predicted background is the DNN response using 5b data, reweighted by an extrapolation factor**
  - Corrects for small differences between 4b, 5b, 6b data.

$$N_i^{6b,\textbf{predicted}} = \mu_{NF} \times N_i^{5b,\textbf{obs}} \times \left(\frac{N_i^{5b,\textbf{obs}}}{N_i^{4b,\textbf{obs}}}\right)$$

Analysis strategy





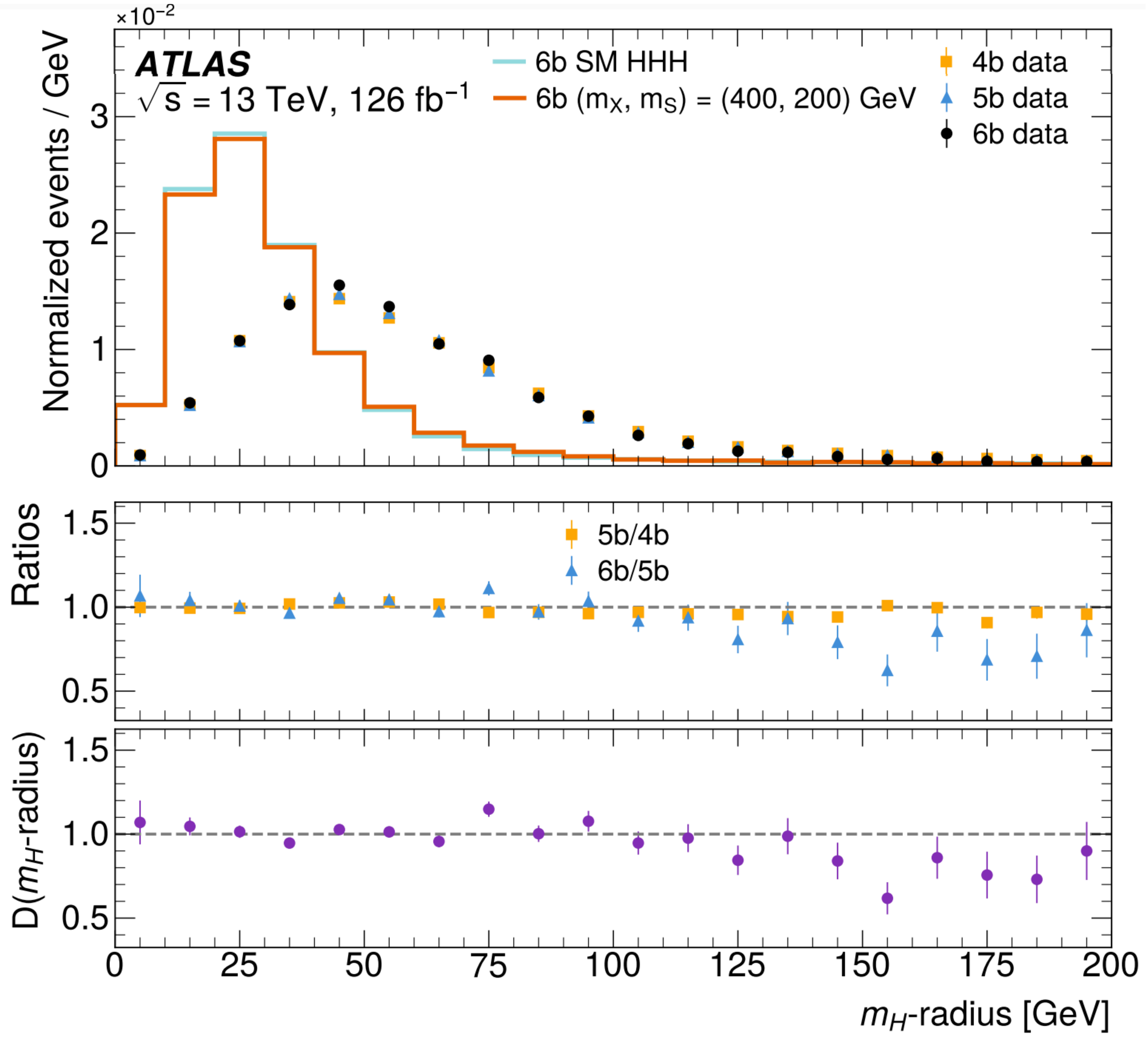


- **The DNN output score is a discriminant we fit and compute p-values/limits.**

- **The predicted background is the DNN response using 5b data, reweighted by an extrapolation factor**
  - Corrects for small differences between 4b, 5b, 6b data.
  - Larger difference between 4b, 5b, 6b means larger systematic uncertainties
  - Need to choose DNN inputs carefully!

$$N_i^{6b,\textbf{predicted}} = \mu_{NF} \times N_i^{5b,\textbf{obs}} \times \left(\frac{N_i^{5b,\textbf{obs}}}{N_i^{4b,\textbf{obs}}} \right)$$

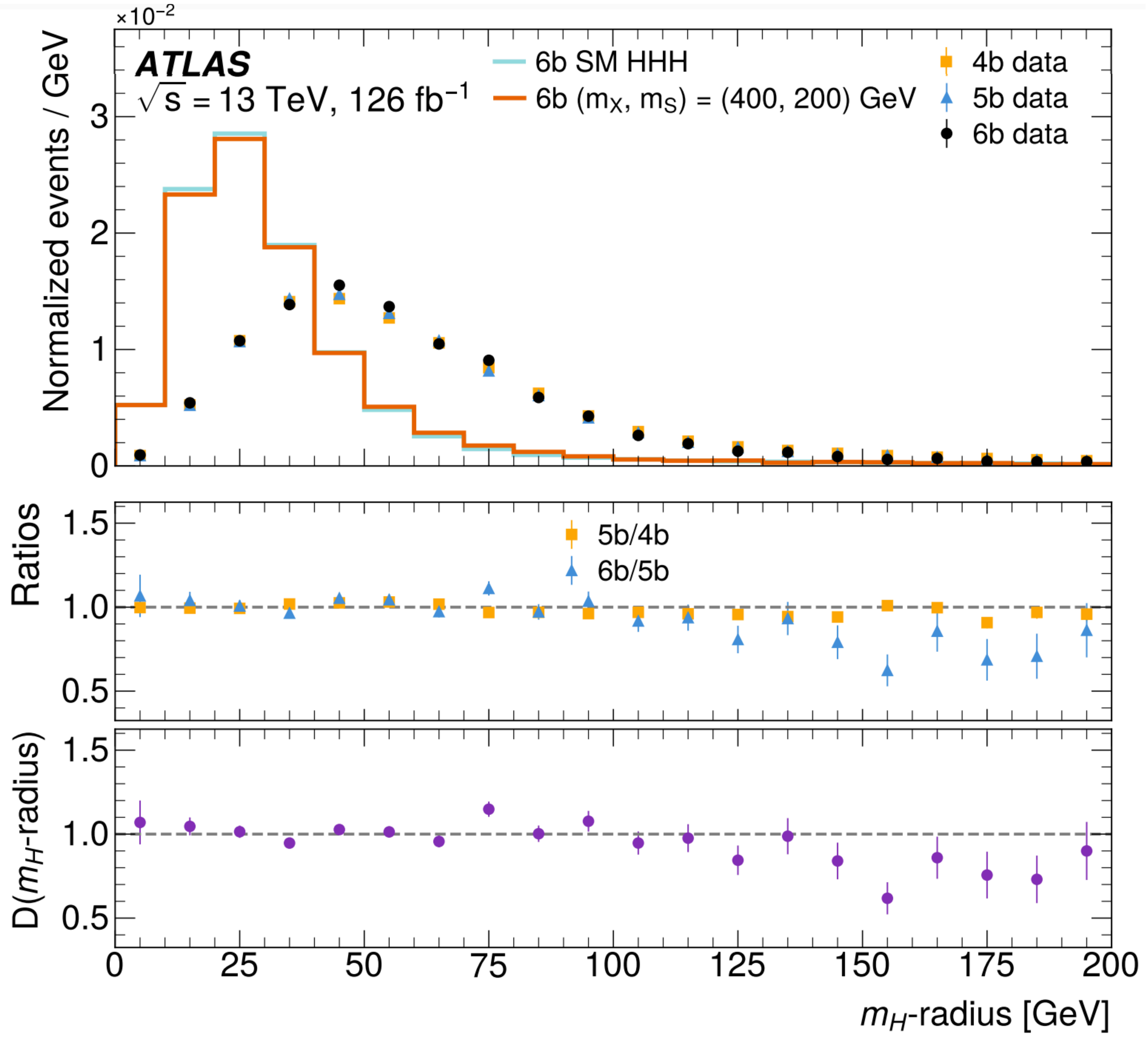
DNN inputs



×10^{-2}
ATLAS
$\sqrt{s}$ = 13 TeV, 126 fb^{-1}
6b SM HHH
6b (m_X, m_S) = (400, 200) GeV
4b data
5b data
6b data
Normalized events / GeV
5b/4b
6b/5b
Ratios
D(m_H-radius)
m_H-radius [GeV]

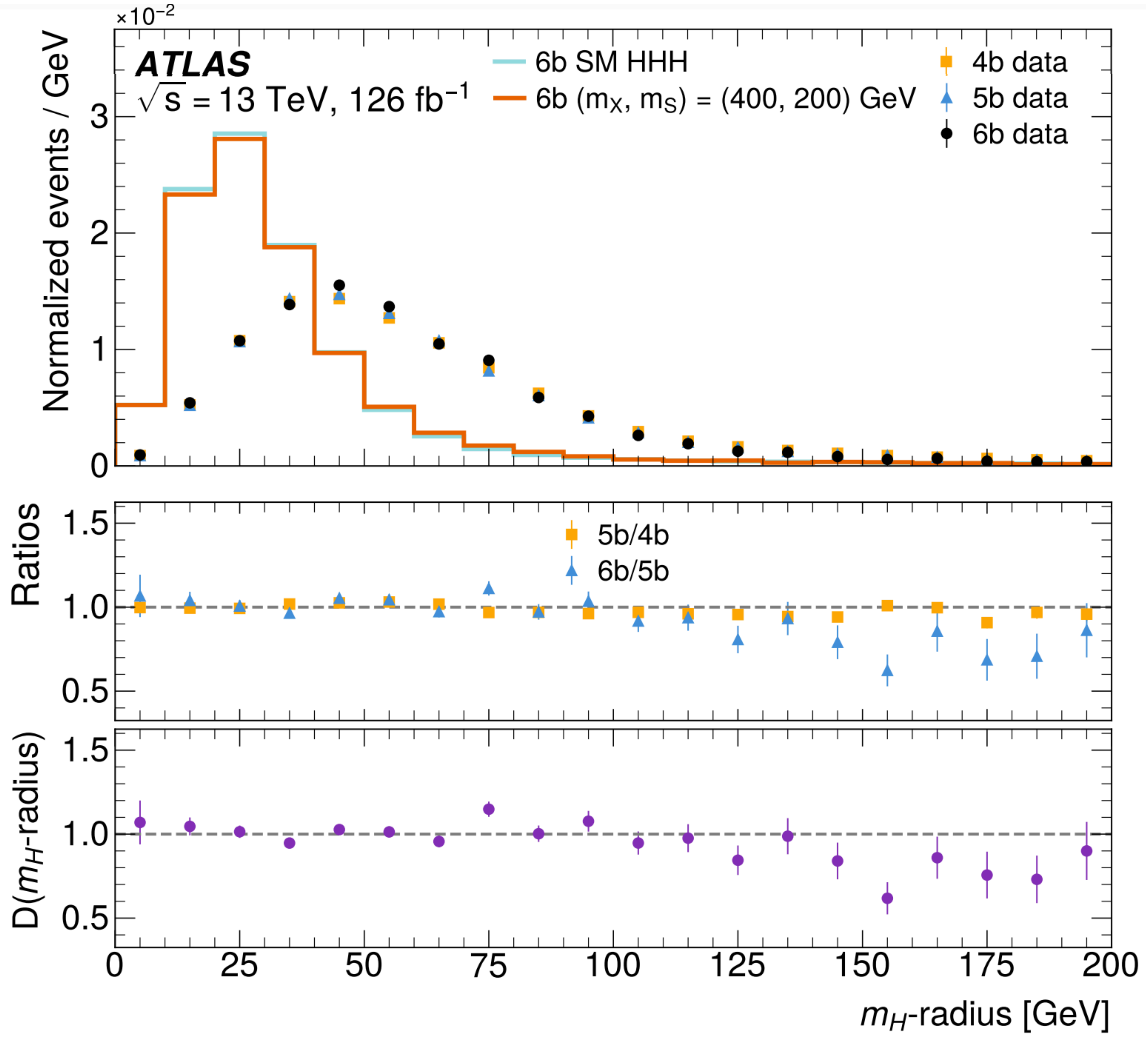
DNN inputs

- **Example of one DNN input shown.**
  - Distance from pairing center: (120, 115, 110) GeV
  - This is an example of a variable for which pairing matters

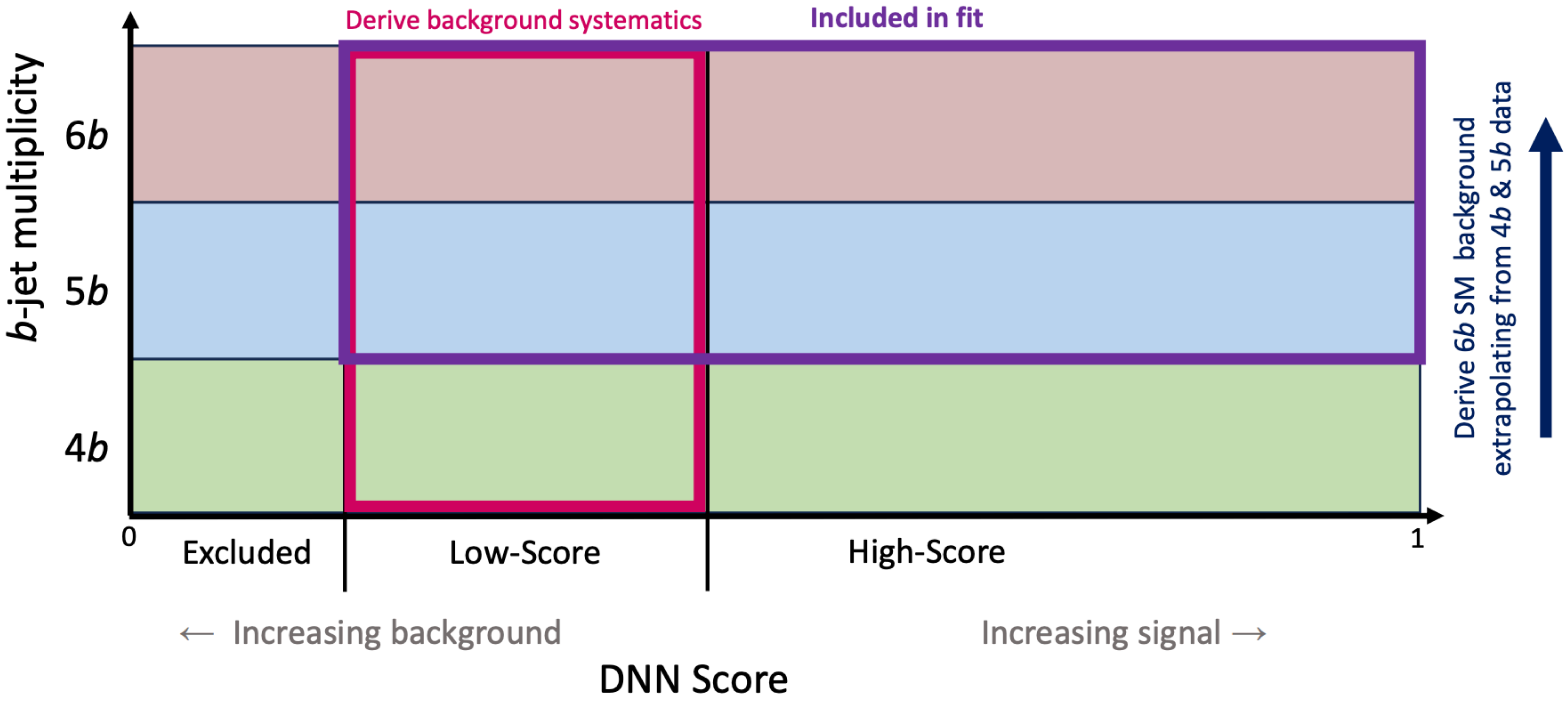


DNN inputs

- **Example of one DNN input shown.**
  - Distance from pairing center: (120, 115, 110) GeV
  - This is an example of a variable for which pairing matters
- **Also shown are 4b, 5b, 6b data and their ratios**
  - We limit ourselves to variables which have small shape difference vs. b-jet multiplicity.
  - This is a trade-off between using more discriminating variables vs. having larger systematic uncertainty on the b-jet extrapolation.

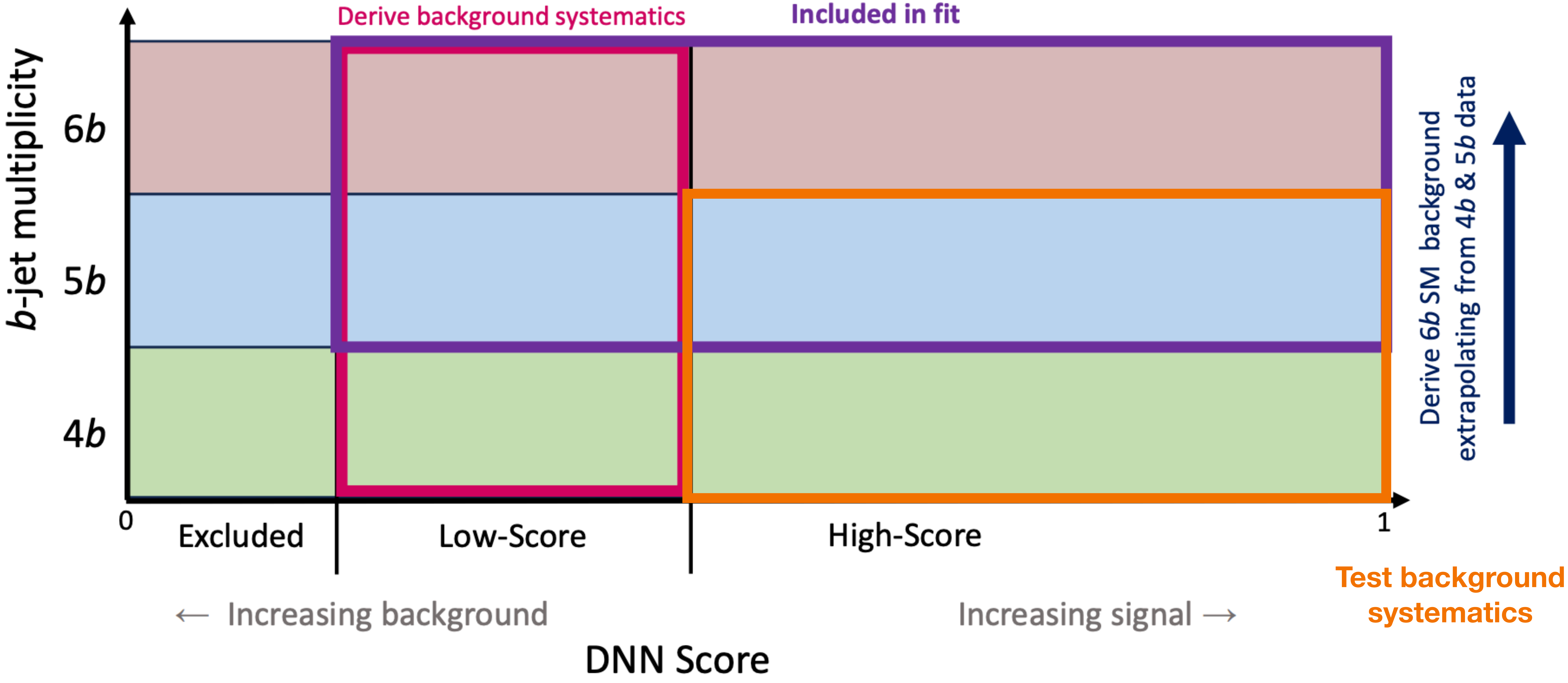


Analysis strategy



- **3 DNNs for the 3 searches: Non-resonant/SM-like, resonant TRSM, heavy resonant**

Analysis strategy



- **3 DNNs for the 3 searches: Non-resonant/SM-like, resonant TRSM, heavy resonant**

Background shape uncertainties

Background shape uncertainties

- The nominal background model reweights the 5b data to extrapolate to 6b data, as a function of DNN score

Background shape uncertainties

- The nominal background model reweights the 5b data to extrapolate to 6b data, as a function of DNN score

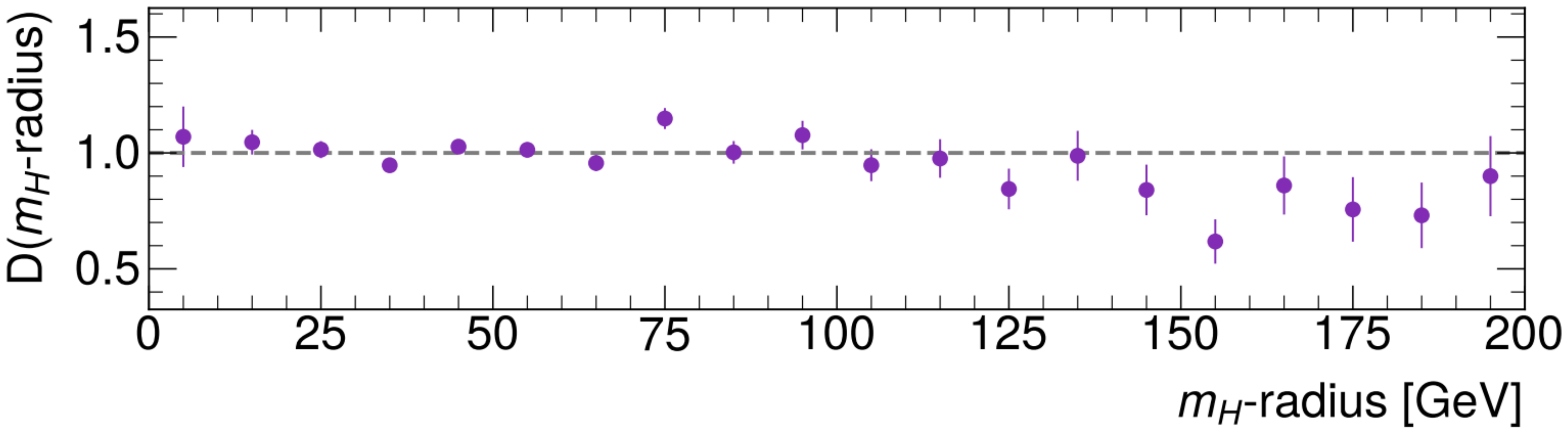
- Create alternative templates performing a similar reweighing but as a function of other variables

Background shape uncertainties

- The nominal background model reweights the 5b data to extrapolate to 6b data, as a function of DNN score

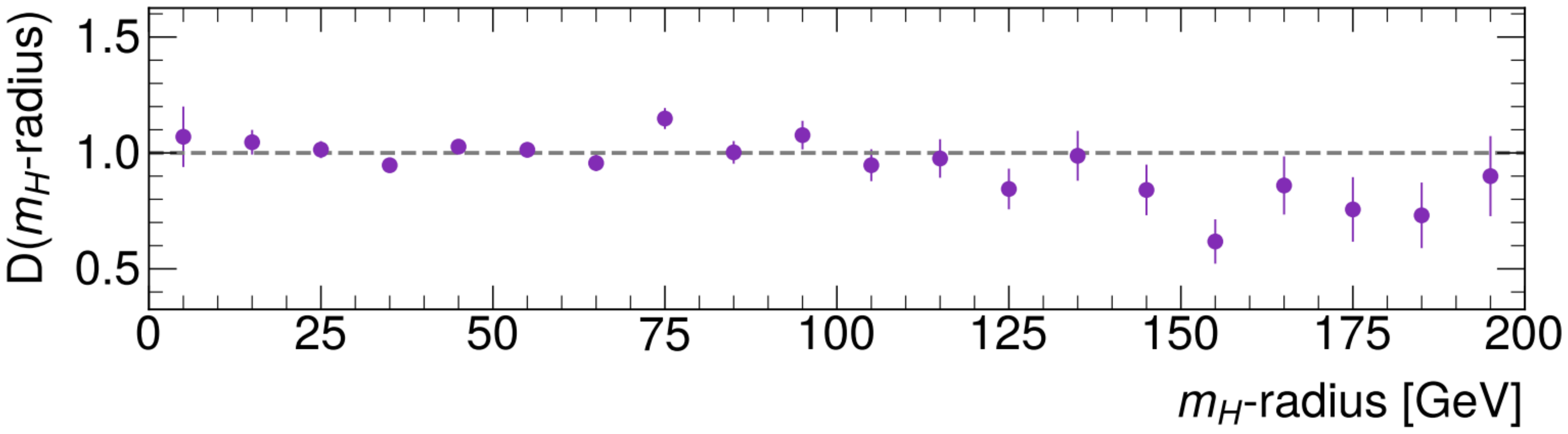
- Create alternative templates performing a similar reweighing but as a function of other variables

- In the “Low-Score”, signal depleted region, compute the extrapolation factor between 4b-5b-6b as a function of each DNN input variable

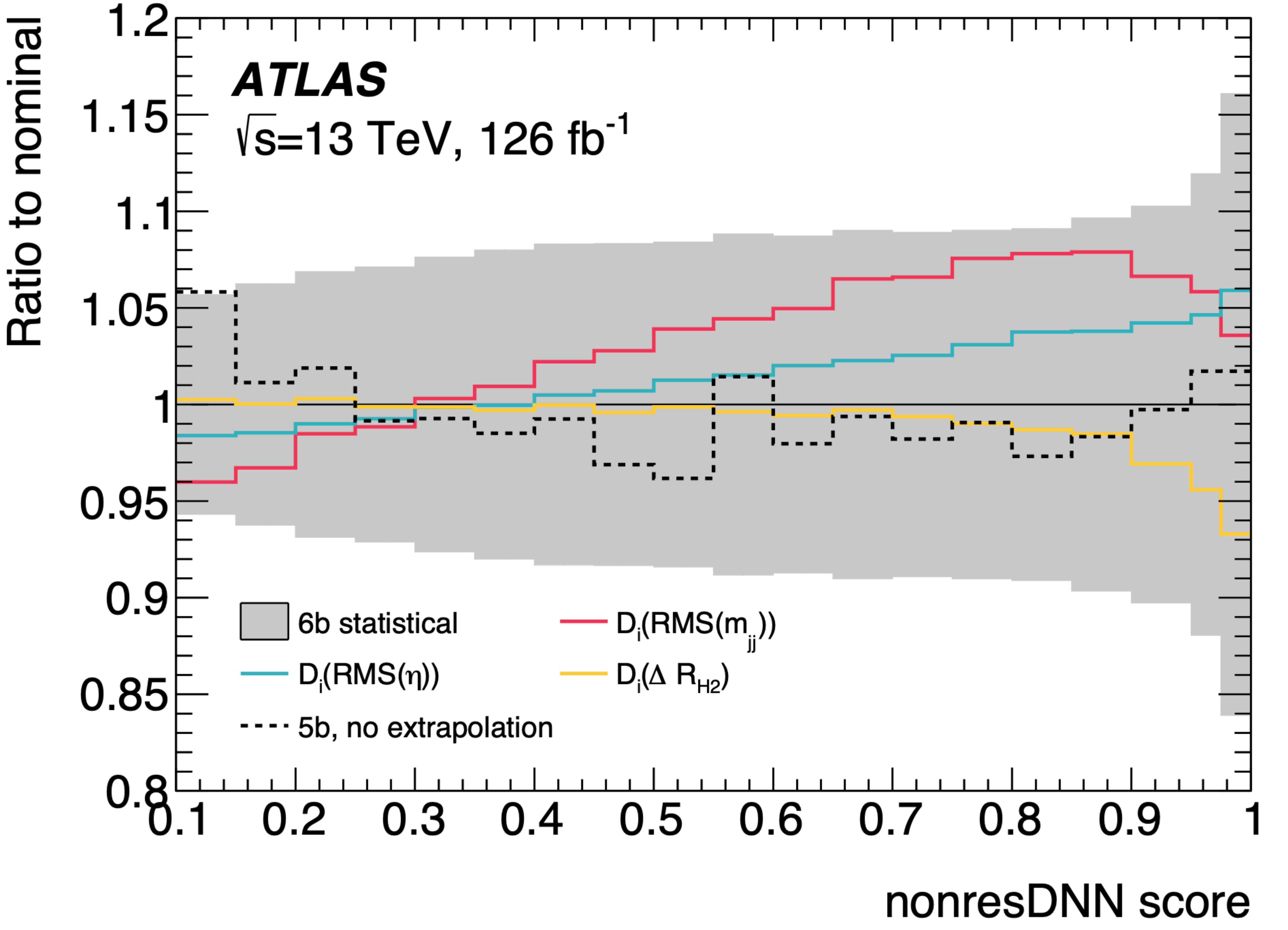


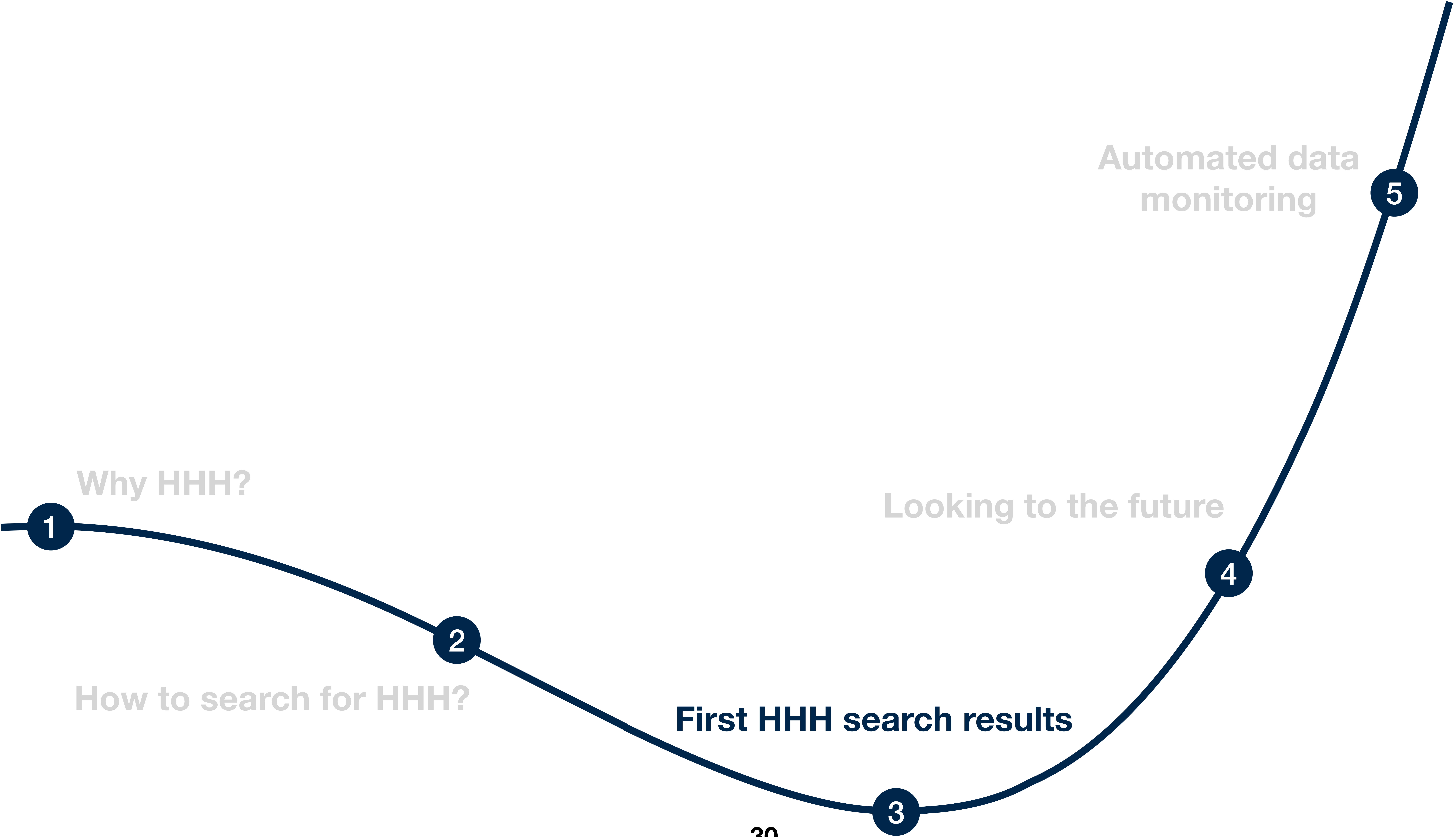
Background shape uncertainties

- **The nominal background model reweights the 5b data to extrapolate to 6b data, as a function of DNN score**
- **Create alternative templates performing a similar reweighing but as a function of other variables**
- **In the "Low-Score", signal depleted region, compute the extrapolation factor between 4b-5b-6b as a function of each DNN input variable**



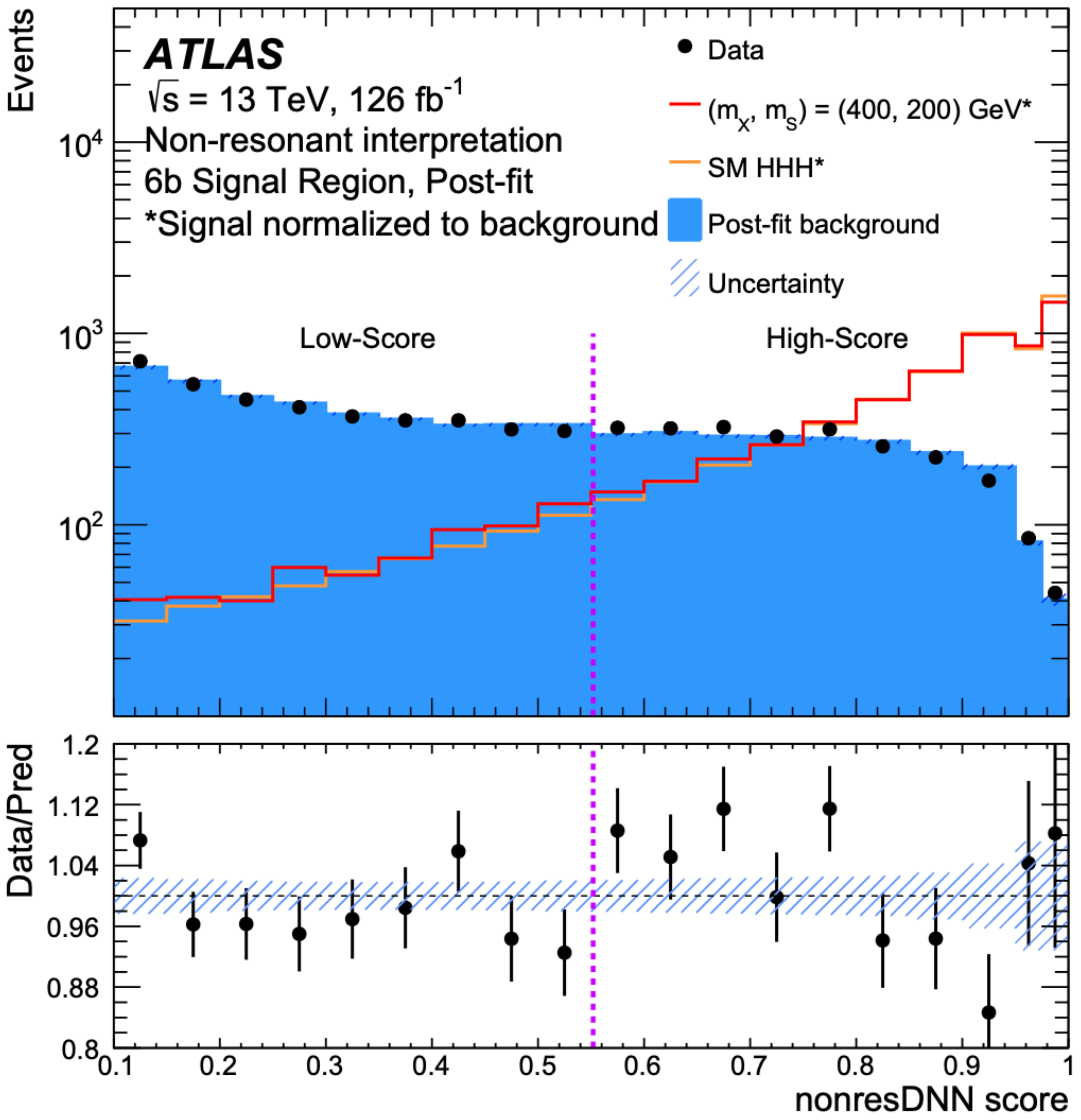
- **Reweight each event by this histogram. Select a linearly independent subset of alternate shapes.**



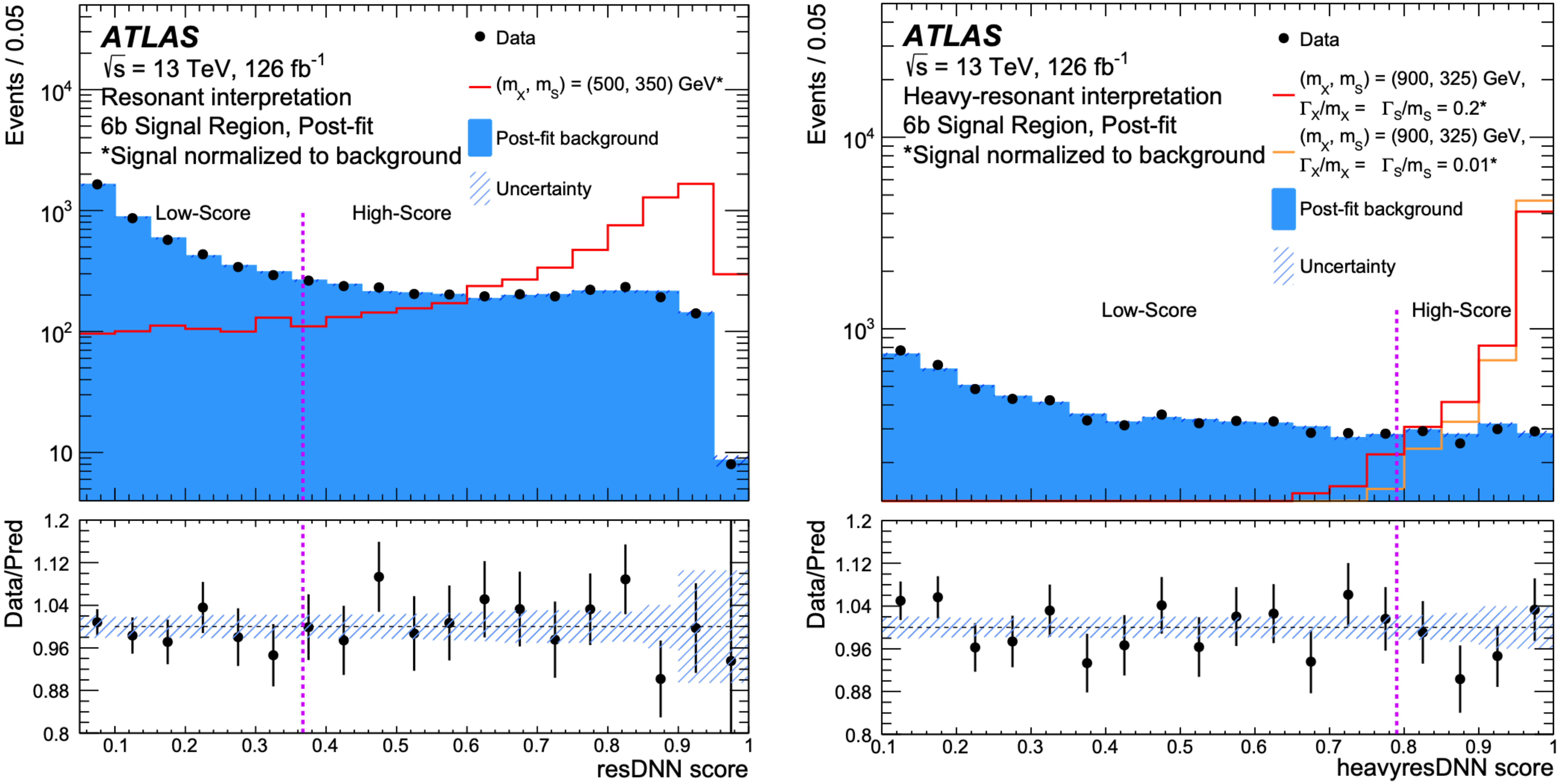

1
Why HHH?
2
How to search for HHH?
3
First HHH search results
4
Looking to the future
5
Automated data monitoring

Results

- **No evidence for non-resonant HHH production**
- **Set limit at $\mu = \sigma/\sigma_{SM} < 750$, or 59 fb**
- **Non-resonant BSM modes can have nearly identical signature to SM HHH.**



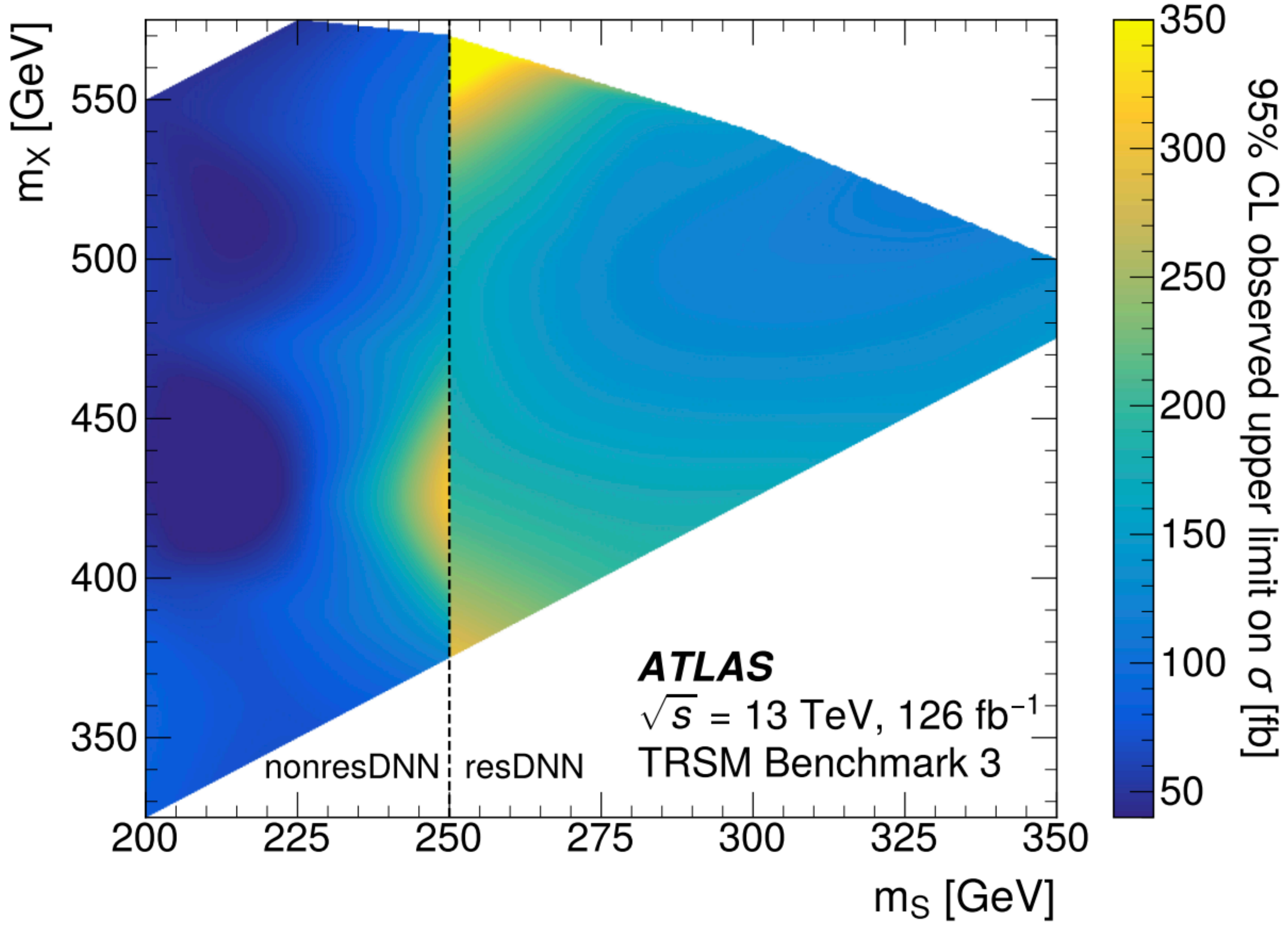
Results



- **No evidence for resonant HHH production**

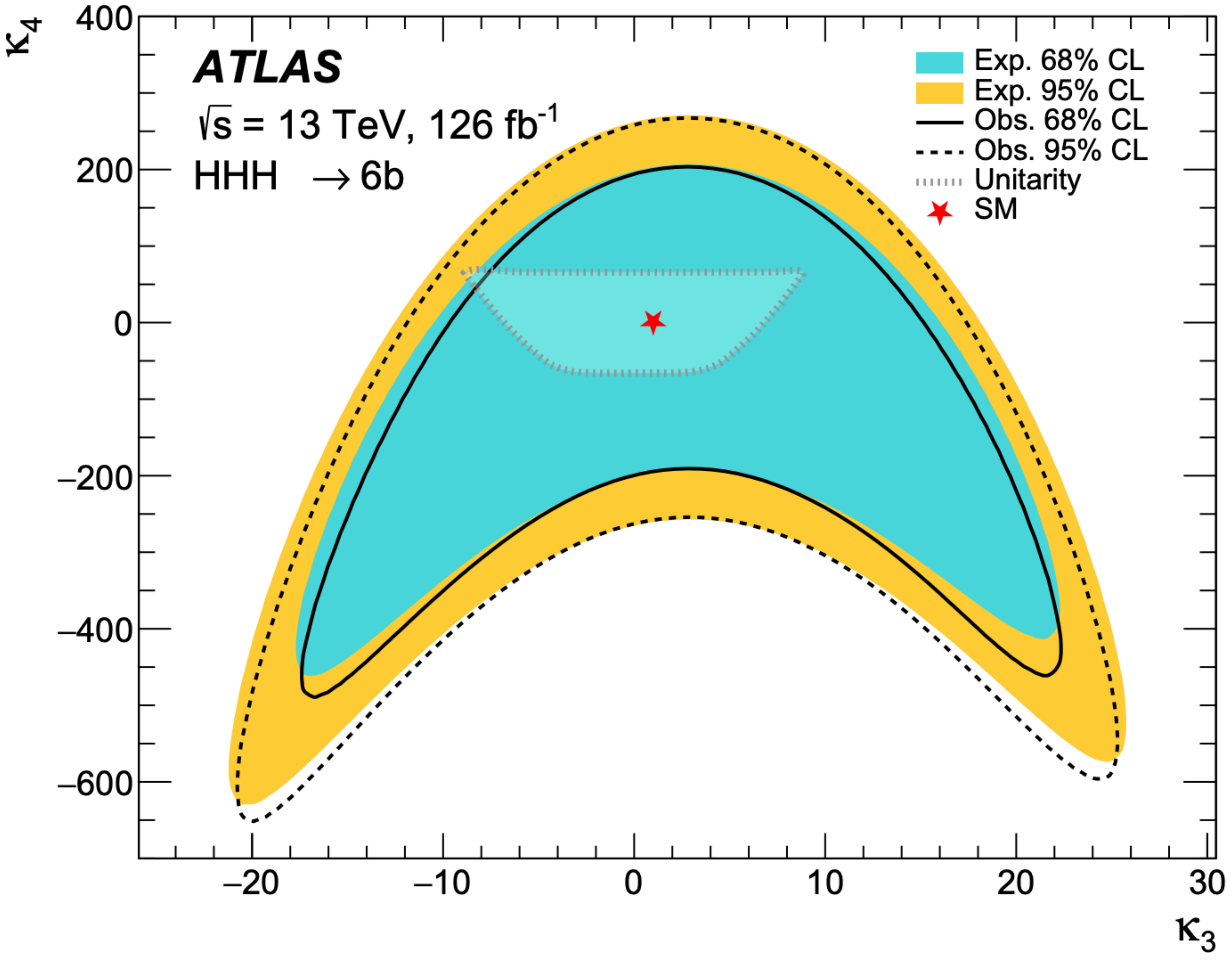
Conclusions: BSM searches

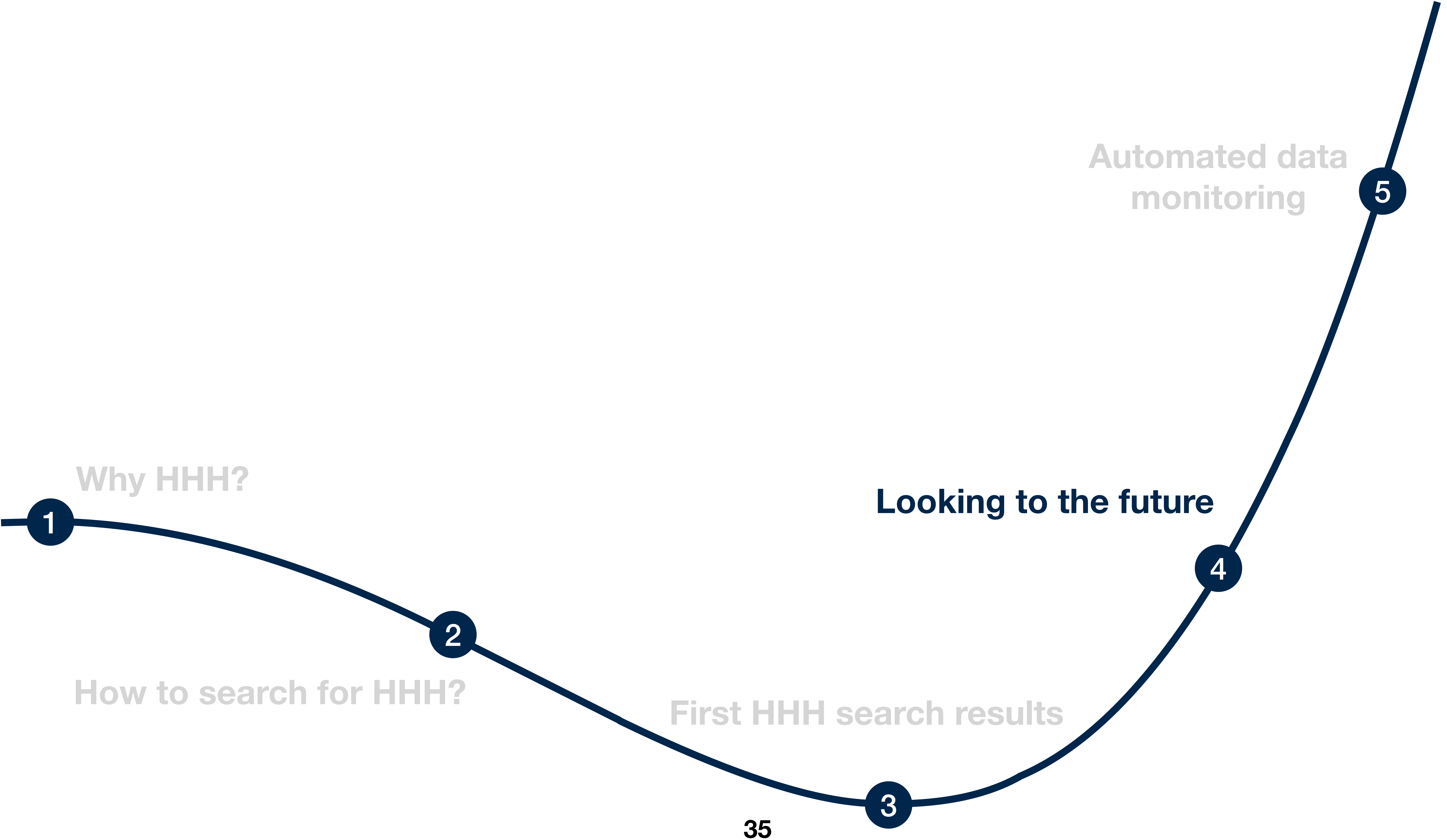
- **The first search for HHH production was recently published**
- **TRSM model predicts HHH production cross sections up to ~150 fb**
- **HHH production limits are on the same order as predicted cross section (50-350 fb as a function of m_X, m_S)**
- **Searches for BSM physics in HHH is relevant TODAY!**



Conclusions: SM-like interpretation

- **The first search for HHH production was recently published**
- **Results include first ever limits on κ_4, the quartic Higgs self-coupling**
- **Current results suggest $\kappa_4 = 1 \pm 30$ is an ambitious target at the HL-LHC.**
  - ATLAS projects in progress. This is back of envelope math
  - Need 10x improvement on top of lumi (6 ab^{-1}).
    - With no improvement other than lumi, more like $\kappa_4 = 1 \pm 100$
  - Some improvement expected from:
    - Multiple channels
    - Reduction of theoretical uncertainties
- **HHH may be especially suited to AI/ML due to the complicated final state.**




Why HHH?
1
2
How to search for HHH?
First HHH search results
3
Looking to the future
4
Automated data monitoring
5

Low hanging fruit

Low hanging fruit

- **Theoretical uncertainties**
  - A HHH cross-section at 13 TeV was not available in the literature.
  - We computed a quadratic fit to the 14 TeV, 27 TeV, 100 TeV cross-sections and added an additional 20% uncertainty.
  - This only affects the mu-value limit, not the cross section upper limit, but is a leading systematic uncertainty on the mu limit.

Low hanging fruit

- **Theoretical uncertainties**
  - A HHH cross-section at 13 TeV was not available in the literature.
  - We computed a quadratic fit to the 14 TeV, 27 TeV, 100 TeV cross-sections and added an additional 20% uncertainty.
  - This only affects the mu-value limit, not the cross section upper limit, but is a leading systematic uncertainty on the mu limit.

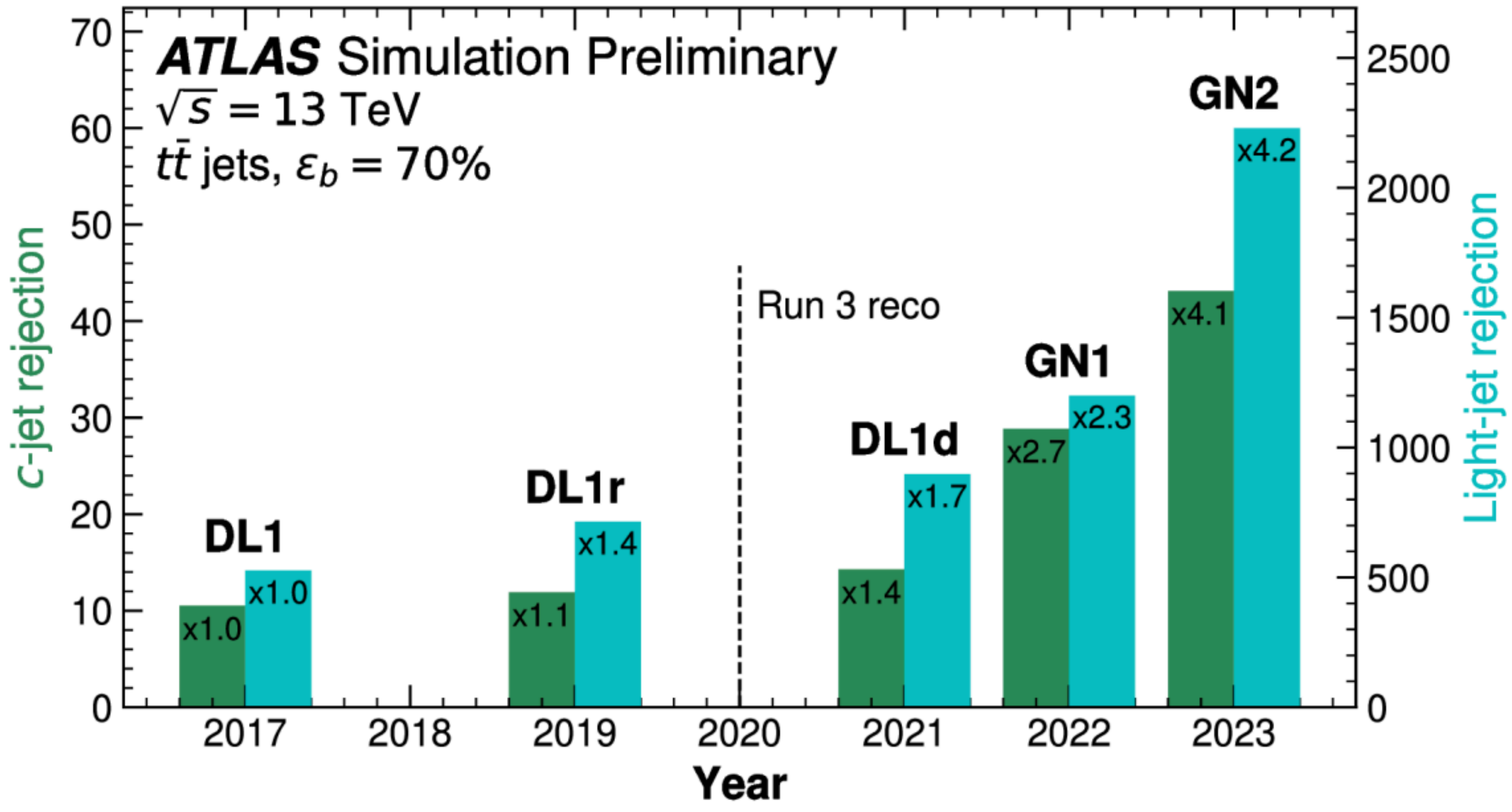
- **Exploration of more channels**
  - There is slight tension in phenomenological studies with respect to which channel is best
  - This is suggestive that in a HHH combination many channels would contribute

Decay Mode	Branching Fraction
bbbbbb	19.5%
bbbb+1 lepton*	15.2%
bbbb+yy	0.2%
bbbb+taus(had)	2.5%
bb+SS leptons	1.9%

HHH as an AI laboratory

HHH as an AI laboratory

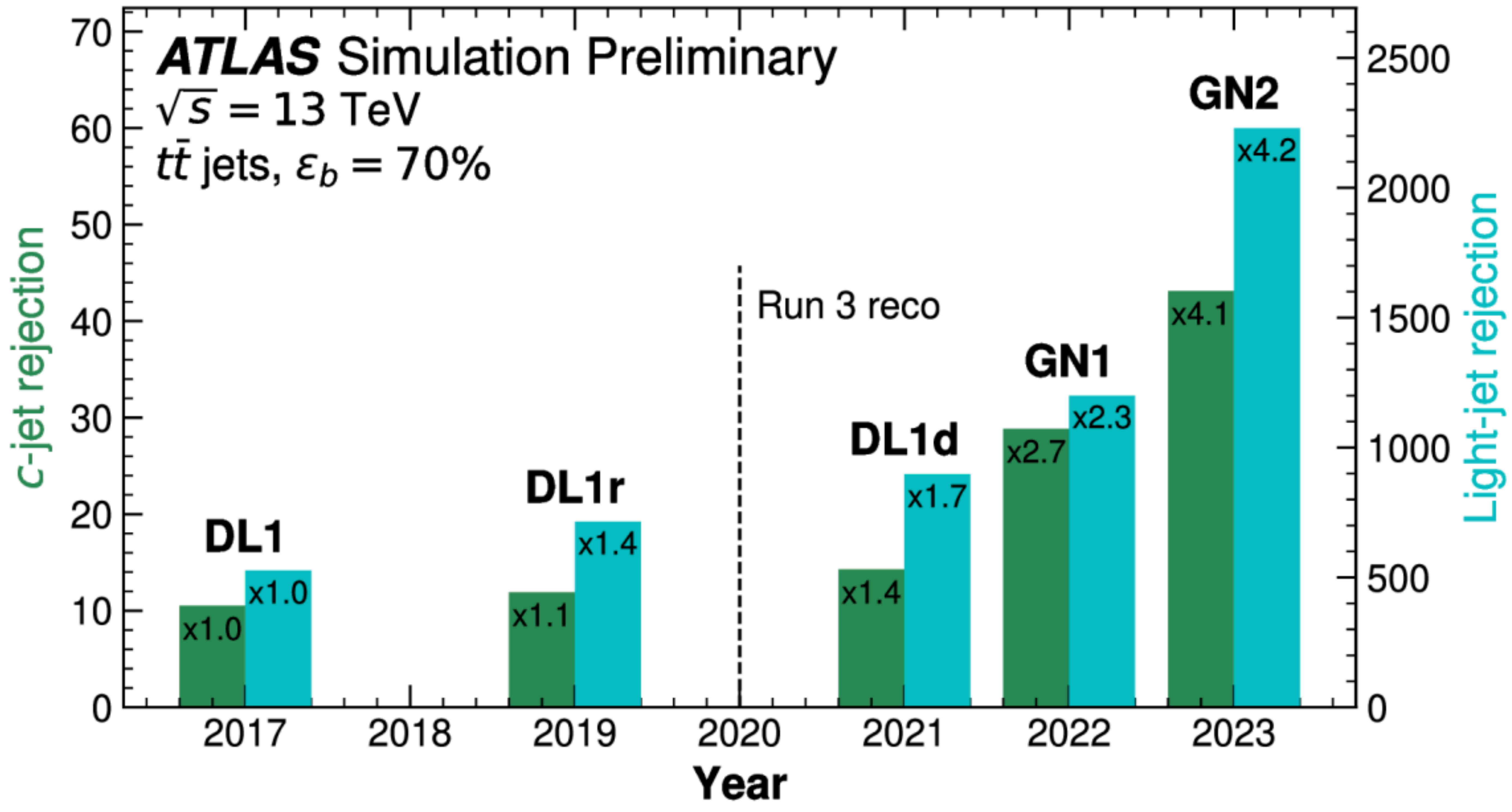
- **How can we learn from the success of the flavor tagging group in ATLAS?**



HHH as an AI laboratory

- **How can we learn from the success of the flavor tagging group in ATLAS?**

- **Speaking from personal experience, ML strategy is often trimmed to meet deadlines:**
  - Conference deadlines
  - Graduation timelines
  - Individual analysis teams are often re-inventing the wheel

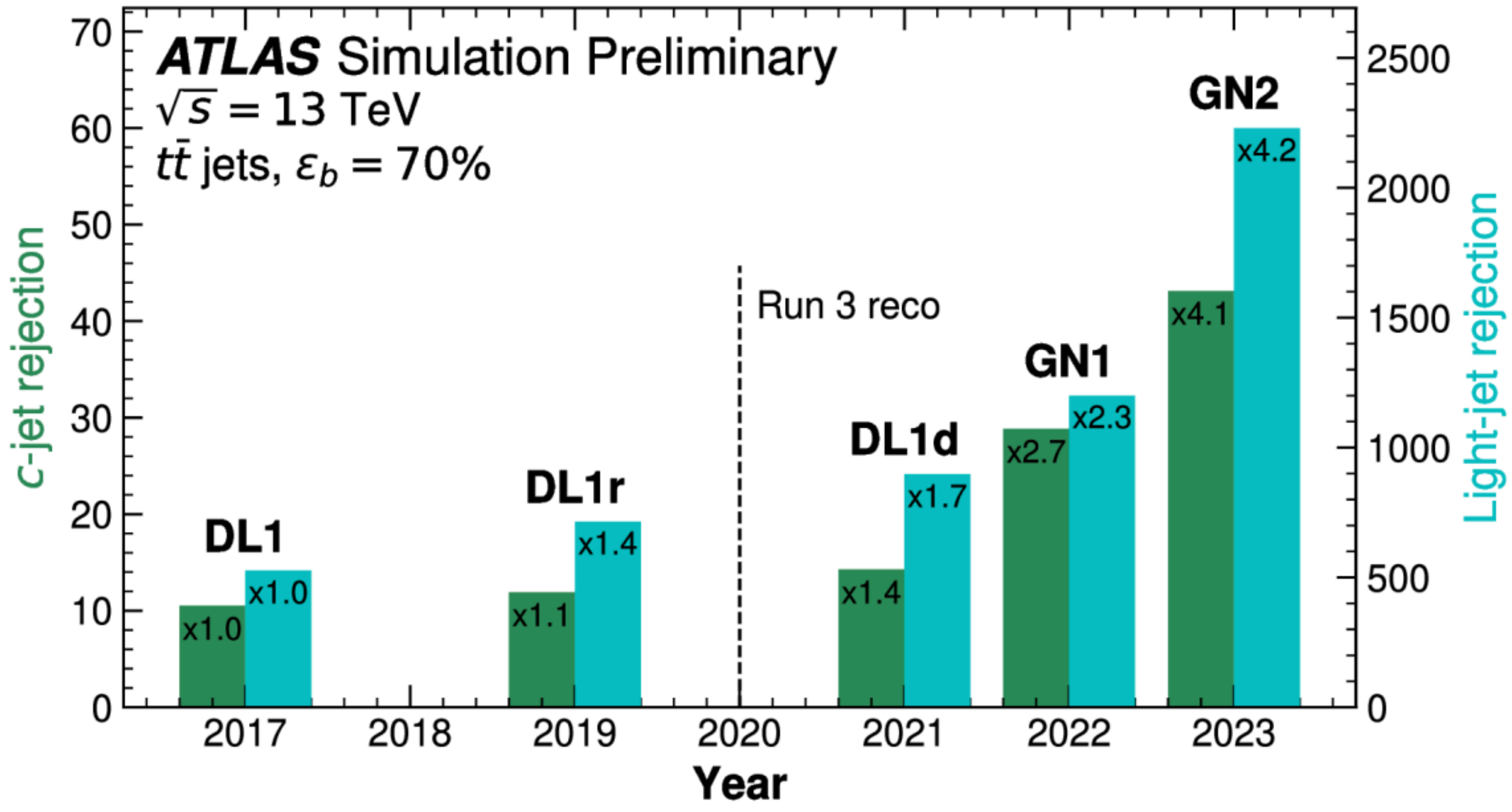


HHH as an AI laboratory

- **How can we learn from the success of the flavor tagging group in ATLAS?**

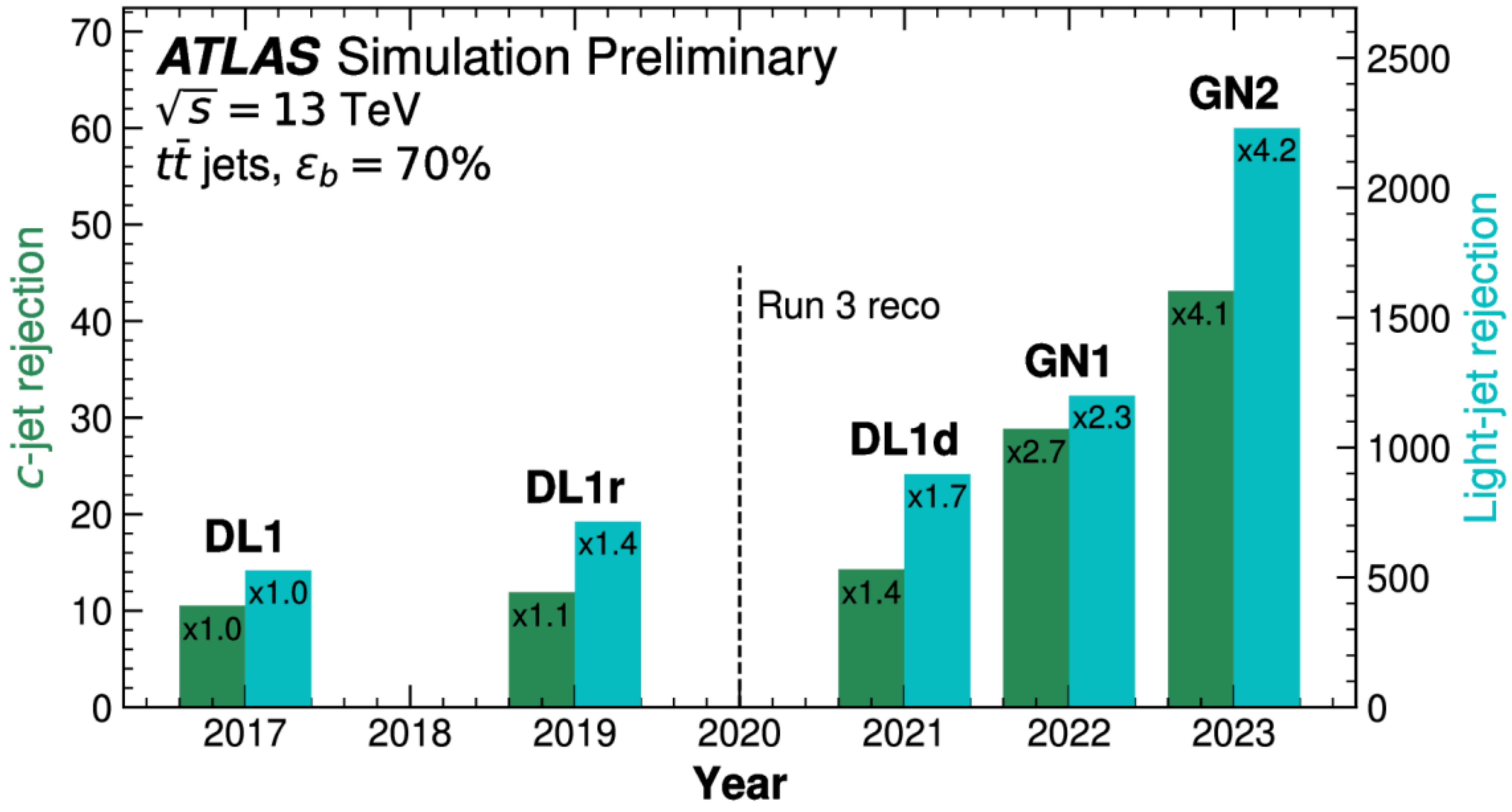
- **Speaking from personal experience, ML strategy is often trimmed to meet deadlines:**
  - Conference deadlines
  - Graduation timelines
  - Individual analysis teams are often re-inventing the wheel

- **HHH has some nice properties:**
  - Fully data-driven background model
  - Stats limited, simple signal systematic uncertainties
  - Complex final state with non-trivial pairing tasks



HHH as an AI laboratory

- **How can we learn from the success of the flavor tagging group in ATLAS?**
- **Speaking from personal experience, ML strategy is often trimmed to meet deadlines:**
  - Conference deadlines
  - Graduation timelines
  - Individual analysis teams are often re-inventing the wheel
- **HHH has some nice properties:**
  - Fully data-driven background model
  - Stats limited, simple signal systematic uncertainties
  - Complex final state with non-trivial pairing tasks



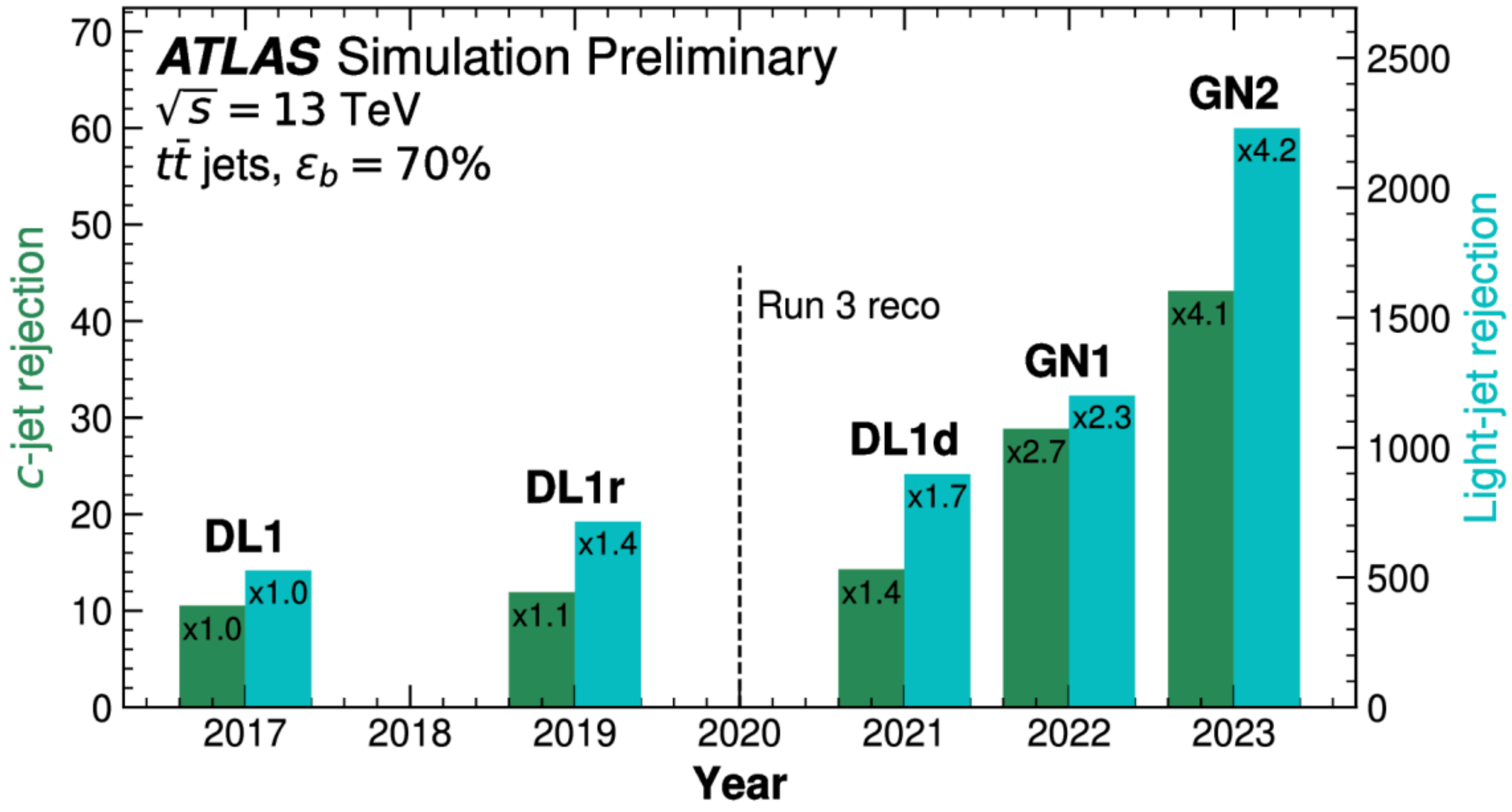
- **In my opinion, there is much to be gained in the application of AI/ML in analyzing the final observables in individual searches.**

HHH as an AI laboratory

- **How can we learn from the success of the flavor tagging group in ATLAS?**

- **Speaking from personal experience, ML strategy is often trimmed to meet deadlines:**
  - Conference deadlines
  - Graduation timelines
  - Individual analysis teams are often re-inventing the wheel

- **HHH has some nice properties:**
  - Fully data-driven background model
  - Stats limited, simple signal systematic uncertainties
  - Complex final state with non-trivial pairing tasks



- **In my opinion, there is much to be gained in the application of AI/ML in analyzing the final observables in individual searches.**

- **Need some classic datasets "MNIST of ATLAS/LHC" for ML challenges**

Pairing b-jets

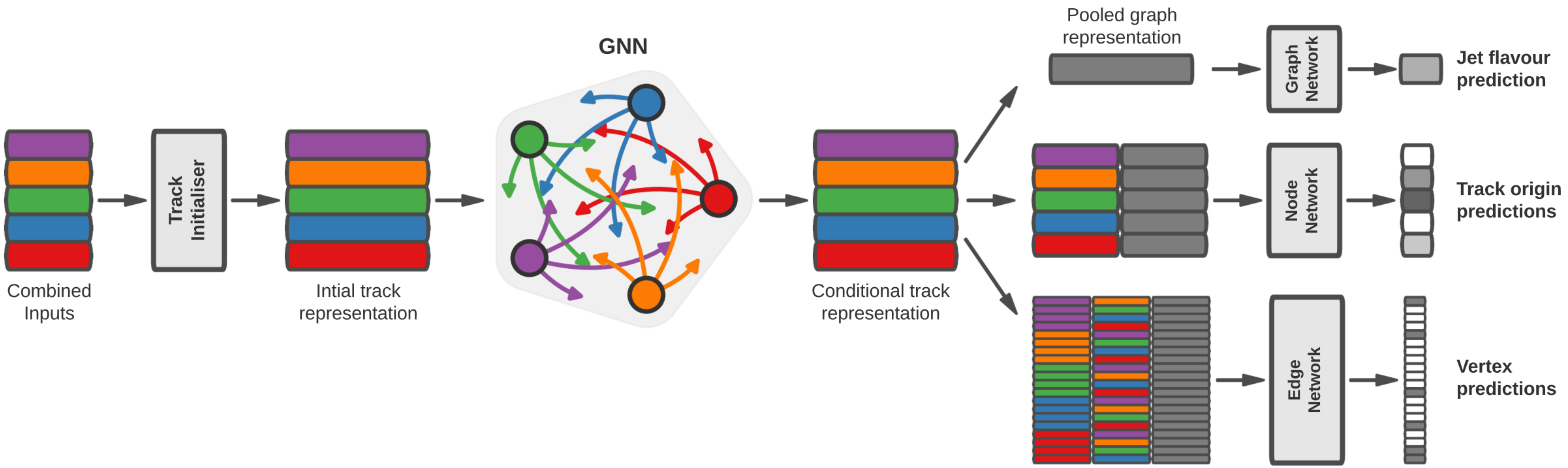
Pairing b-jets

- **Pairing currently uses simplest-possible mass-based method.**

Pairing b-jets

- **Pairing currently uses simplest-possible mass-based method.**

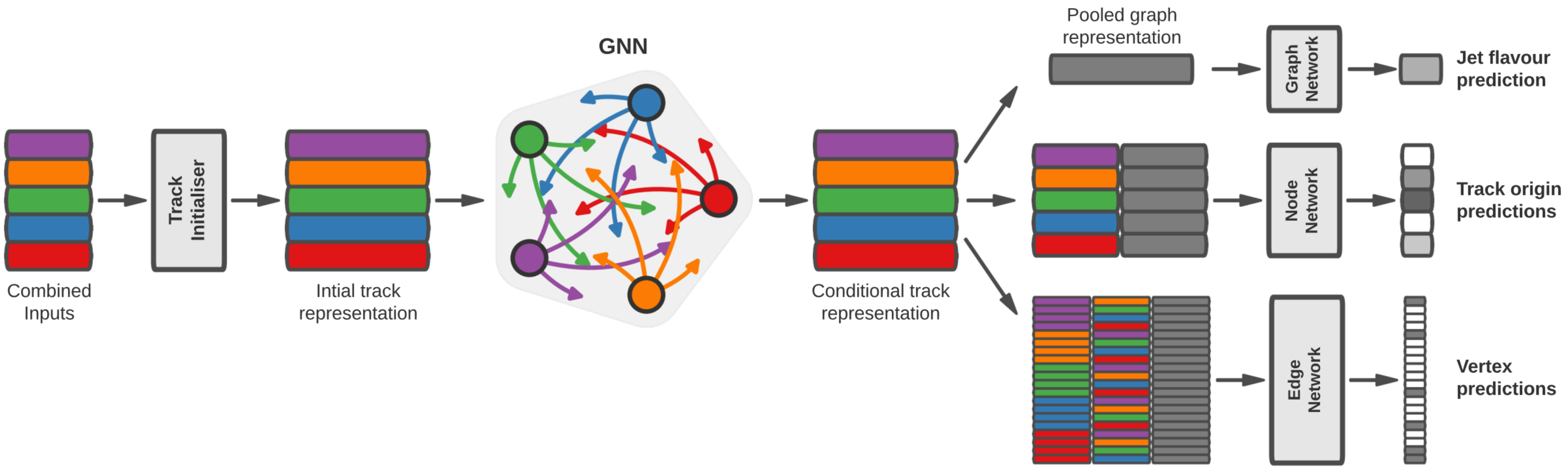
- **Let's learn from experience with ATLAS GN2 tagger:**
  - Auxiliary tasks predict which tracks within the jet come from the same secondary vertex.



Pairing b-jets

- **Pairing currently uses simplest-possible mass-based method.**

- **Let's learn from experience with ATLAS GN2 tagger:**
  - Auxiliary tasks predict which tracks within the jet come from the same secondary vertex.



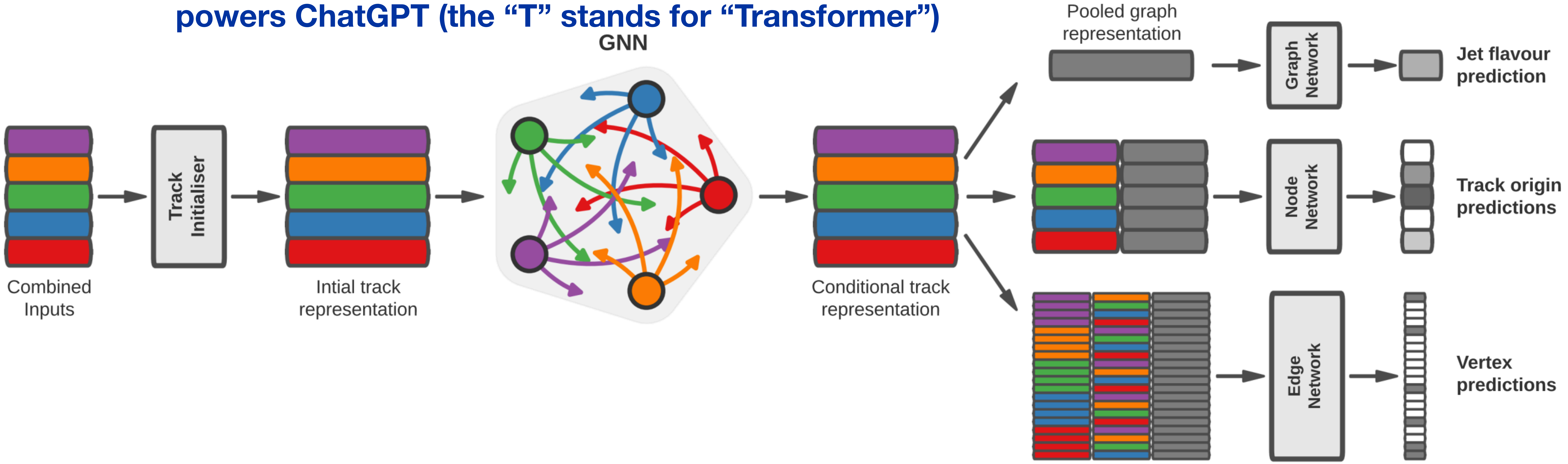
- **Use a similar architecture with tracks replaced by jets. Classify not only global event prediction, but the origin and pairing of jets within the event. Makes use of more MC truth information to guide training.**

Pairing b-jets

- **Pairing currently uses simplest-possible mass-based method.**

- **Let's learn from experience with ATLAS GN2 tagger:**
  - Auxiliary tasks predict which tracks within the jet come from the same secondary vertex.

This is a transformer network, the same technology that powers ChatGPT (the "T" stands for "Transformer")



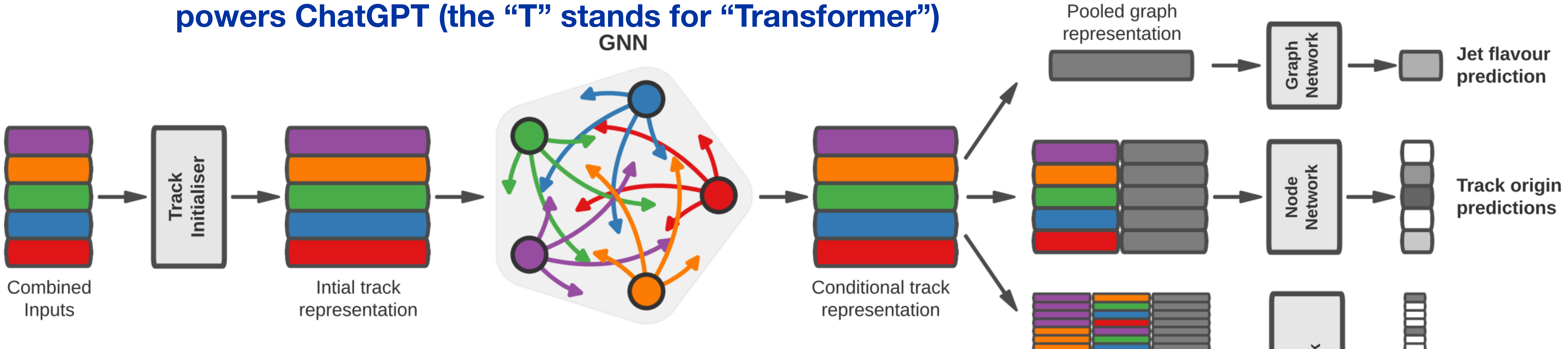
- **Use a similar architecture with tracks replaced by jets. Classify not only global event prediction, but the origin and pairing of jets within the event. Makes use of more MC truth information to guide training.**

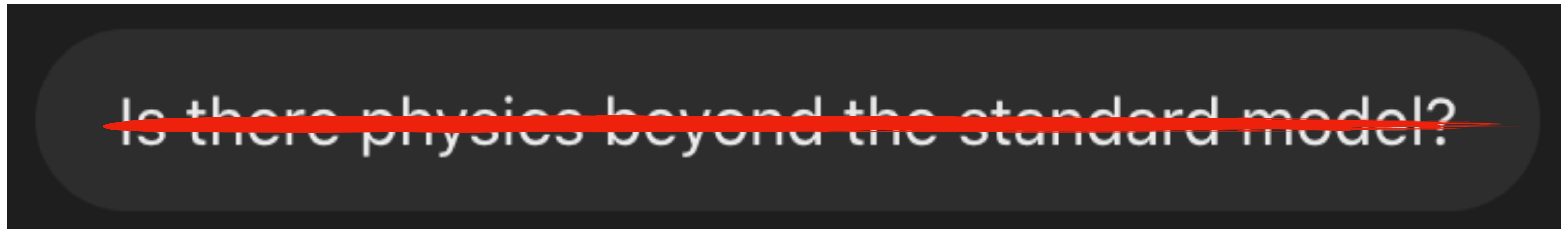
Pairing b-jets

- **Pairing currently uses simplest-possible mass-based method.**

- **Let's learn from experience with ATLAS GN2 tagger:**
  - Auxiliary tasks predict which tracks within the jet come from the same secondary vertex.

This is a transformer network, the same technology that powers ChatGPT (the "T" stands for "Transformer")





- **Use a similar architecture with tracks replaced by jets. Classify not only global event prediction, but the origin and pairing of jets within the event. Makes use of more MC truth information to guide training.**

Why would transformers help pairing?

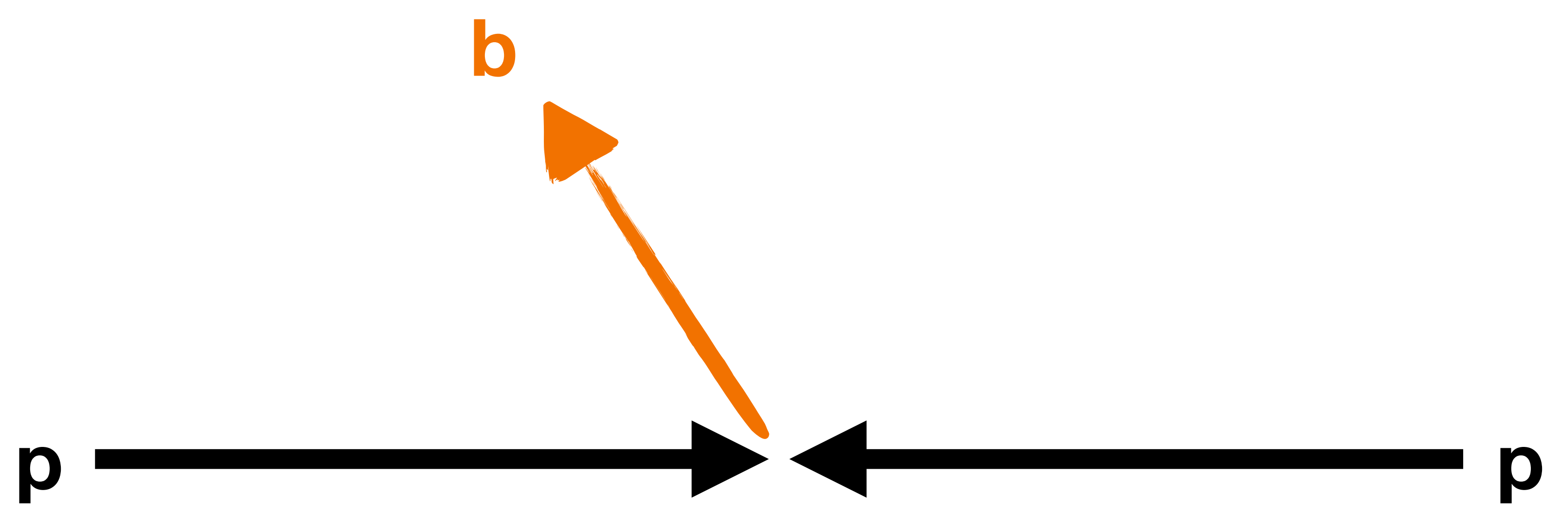
Why would transformers help pairing?

- **Transformers allows for context-dependent encoding of particles**
  - They are “transformed” by the context of the other particles in the event

Why would transformers help pairing?

- **Transformers allows for context-dependent encoding of particles**
  - They are “transformed” by the context of the other particles in the event

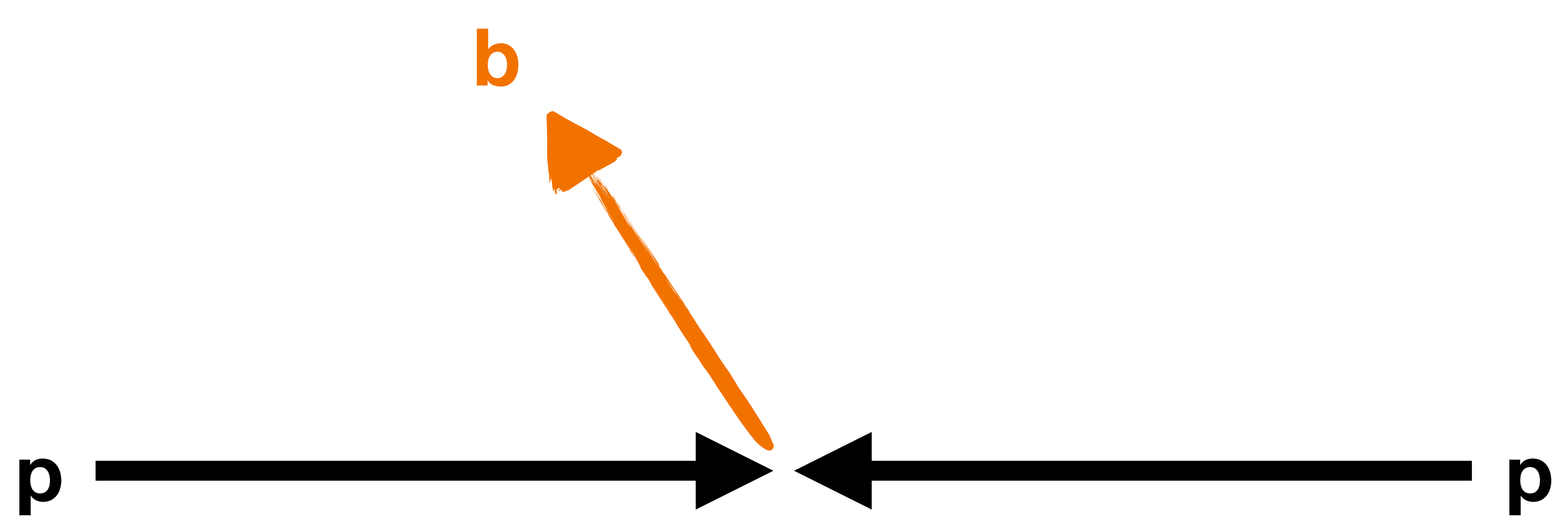
- **If you see a b-jet in the event, what is the source?**
  - Imagine we have some embedding space where, depending on the kinematics of the b-jet, it is more top-like or Higgs-like.
  - The embedding is defined by some neural network.

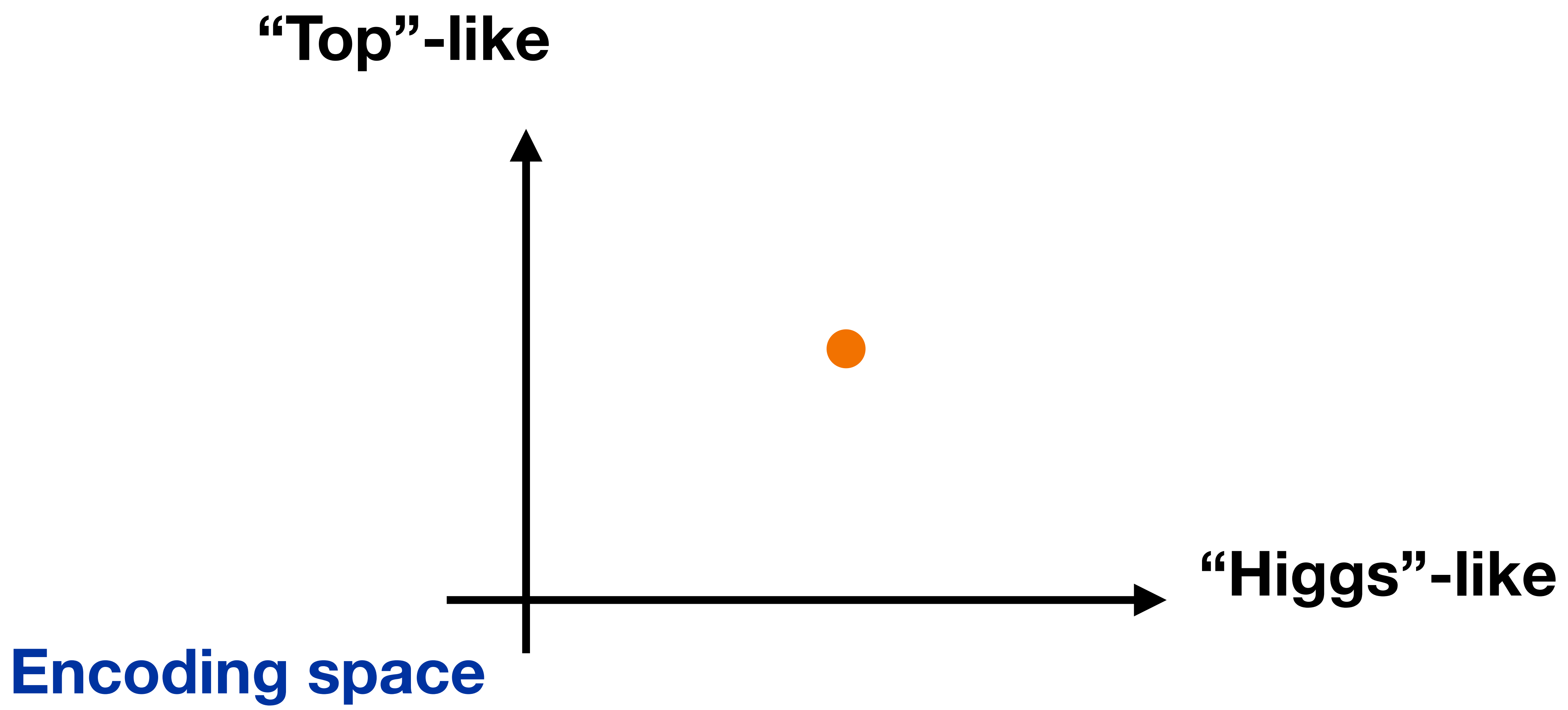


Why would transformers help pairing?

- **Transformers allows for context-dependent encoding of particles**
  - They are "transformed" by the context of the other particles in the event

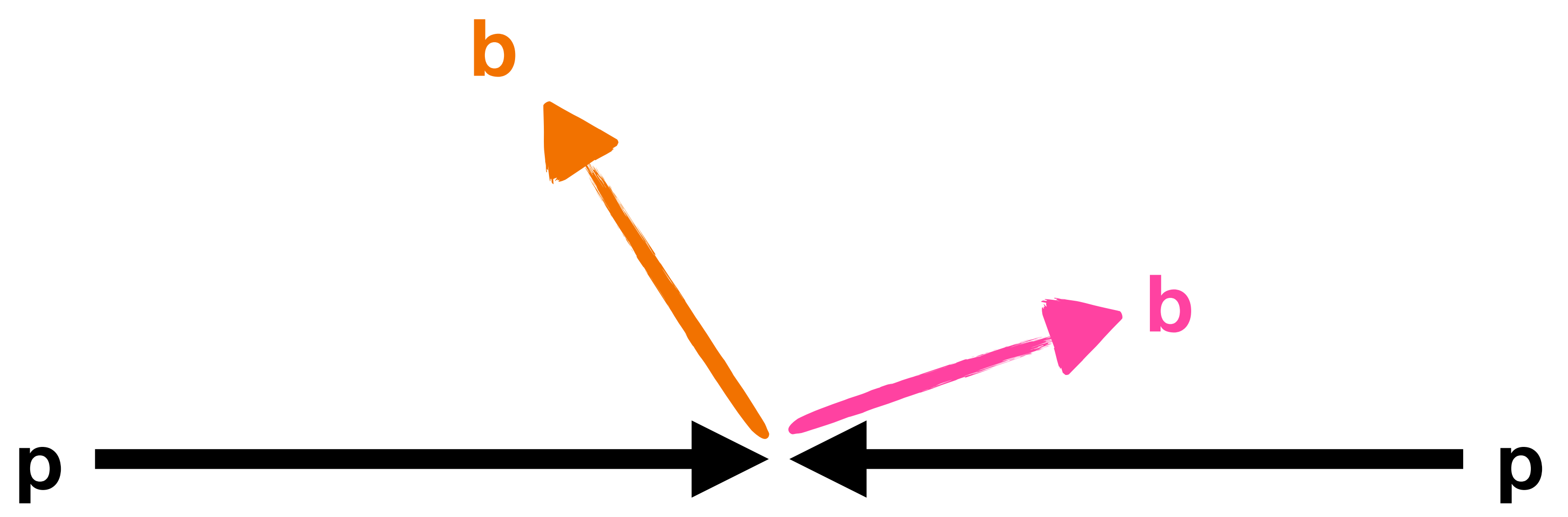
- **If you see a b-jet in the event, what is the source?**
  - Imagine we have some embedding space where, depending on the kinematics of the b-jet, it is more top-like or Higgs-like.
  - The embedding is defined by some neural network.

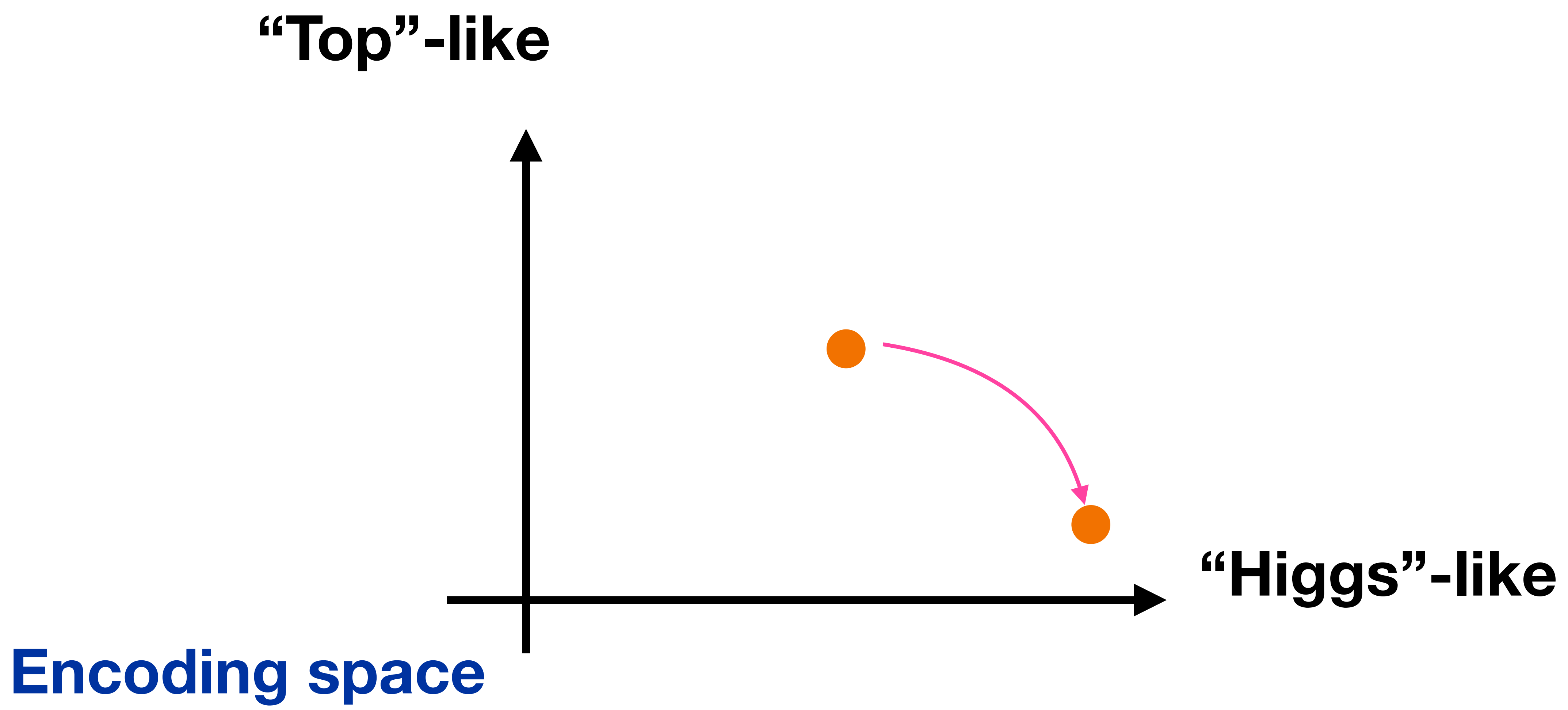




Why would transformers help pairing?

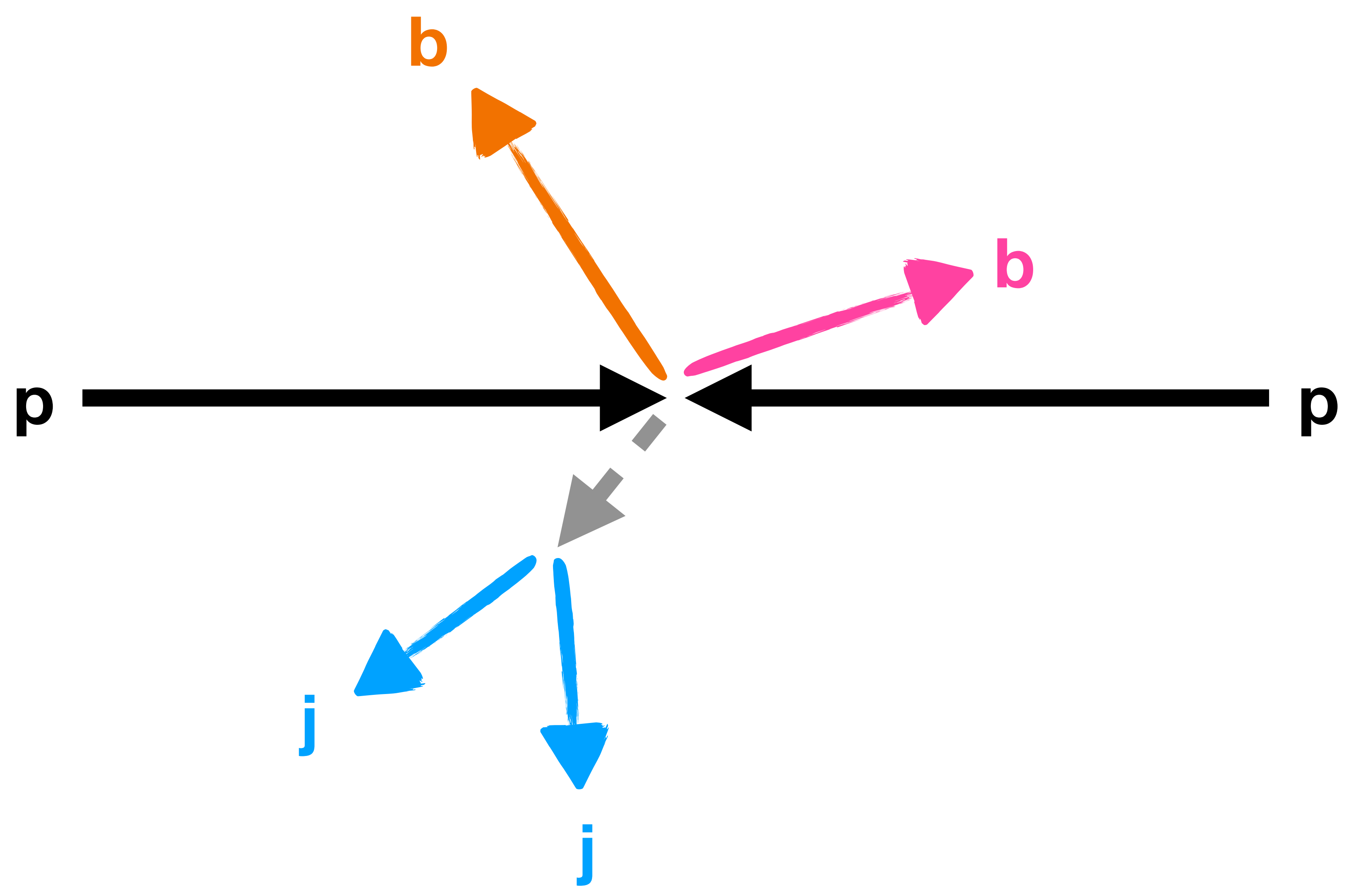
- **Transformers allows for context-dependent encoding of particles**
  - They are “transformed” by the context of the other particles in the event
- **If you see a b-jet in the event, what is the source?**
  - Imagine we have some embedding space where, depending on the kinematics of the b-jet, it is more top-like or Higgs-like.
  - The embedding is defined by some neural network.
- **If there is a second b-jet (and $M(bb)$~125 GeV), would you update your prior?**

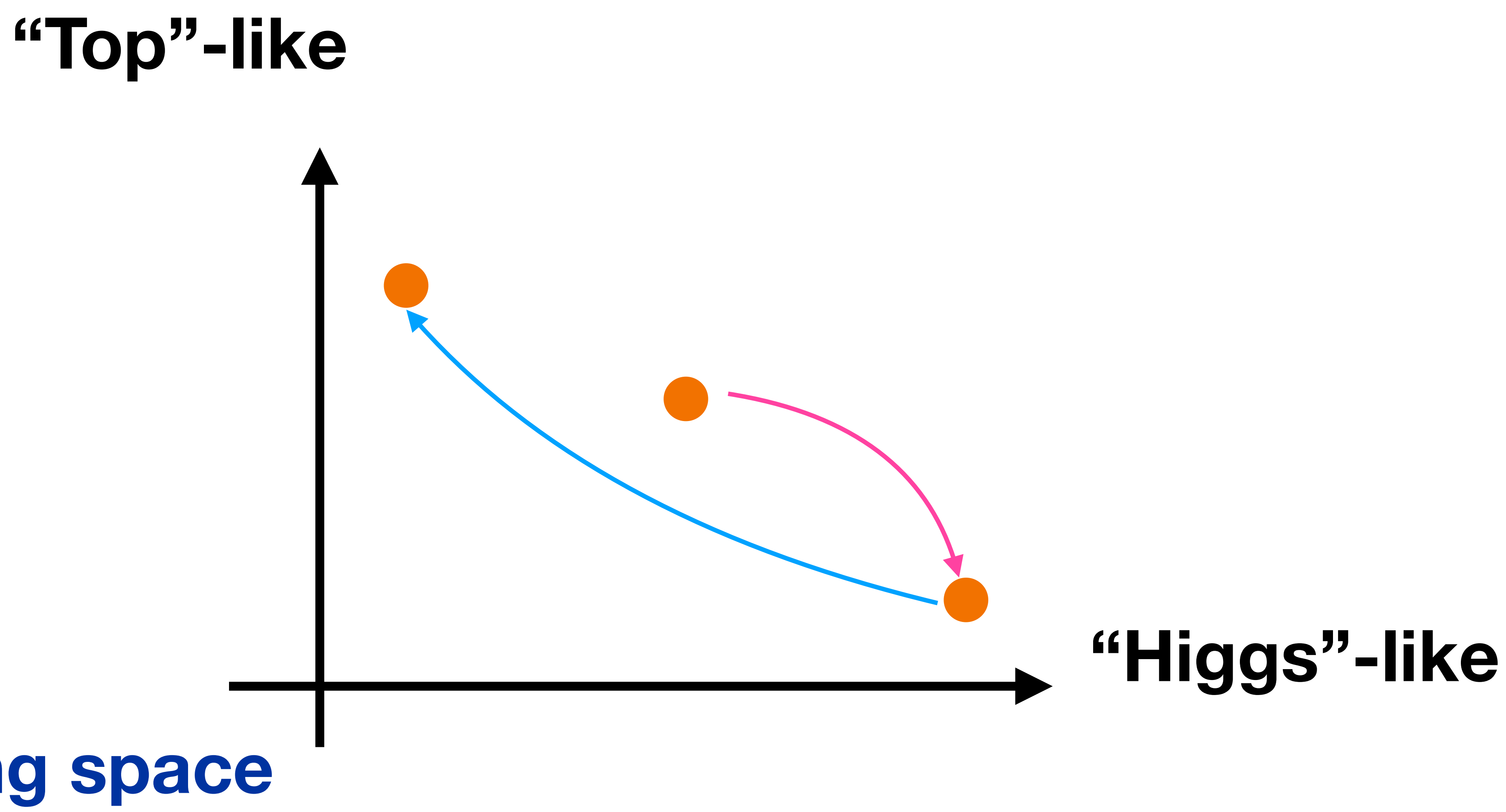




Why would transformers help pairing?

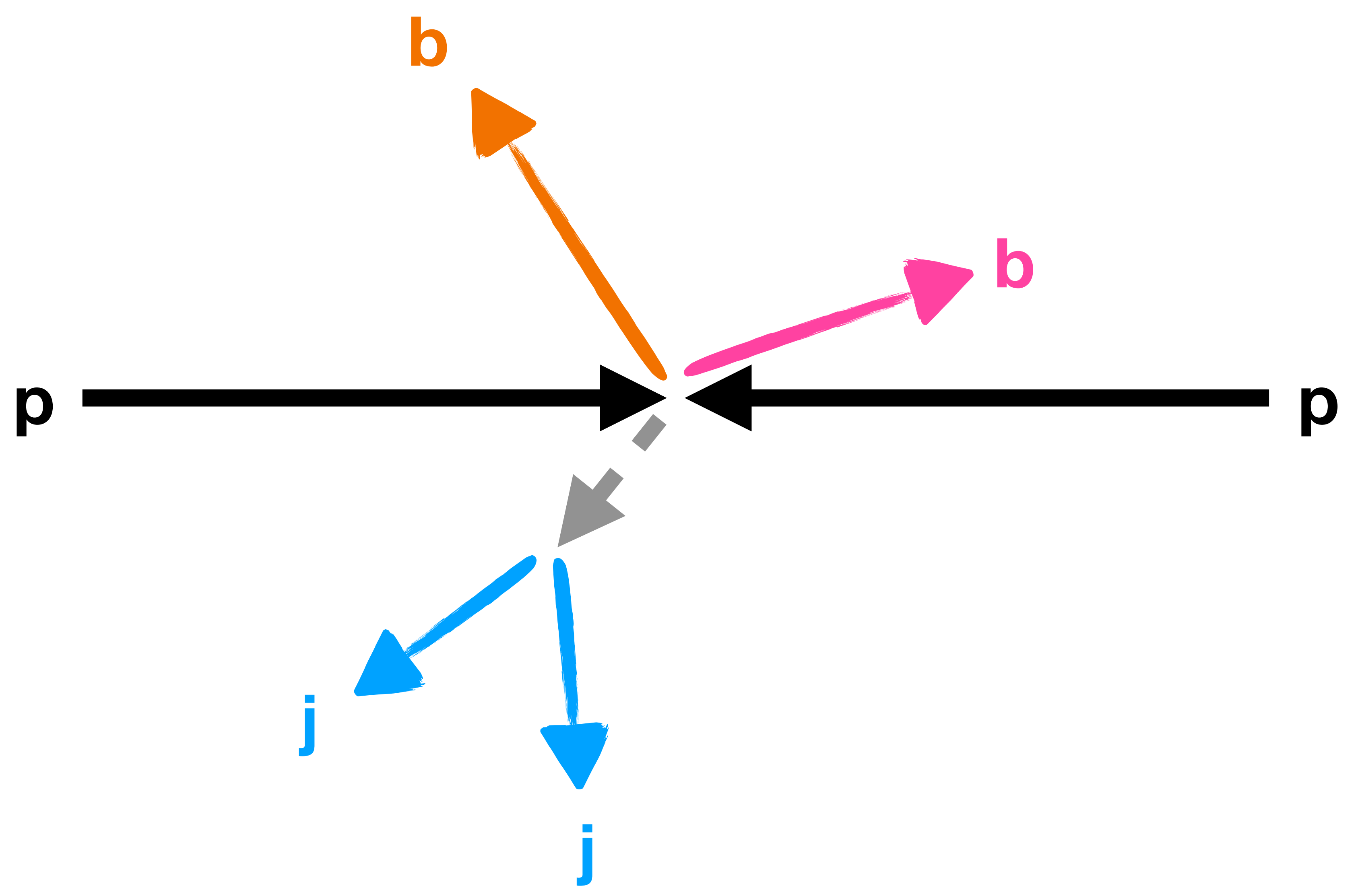
- **Transformers allows for context-dependent encoding of particles**
  - They are "transformed" by the context of the other particles in the event
- **If you see a b-jet in the event, what is the source?**
  - Imagine we have some embedding space where, depending on the kinematics of the b-jet, it is more top-like or Higgs-like.
  - The embedding is defined by some neural network.
- **If there is a second b-jet (and $M(bb)$~125 GeV), would you update your prior?**
- **What if there was also a *jj* pair with $M(bjj)$ ~173 GeV?**

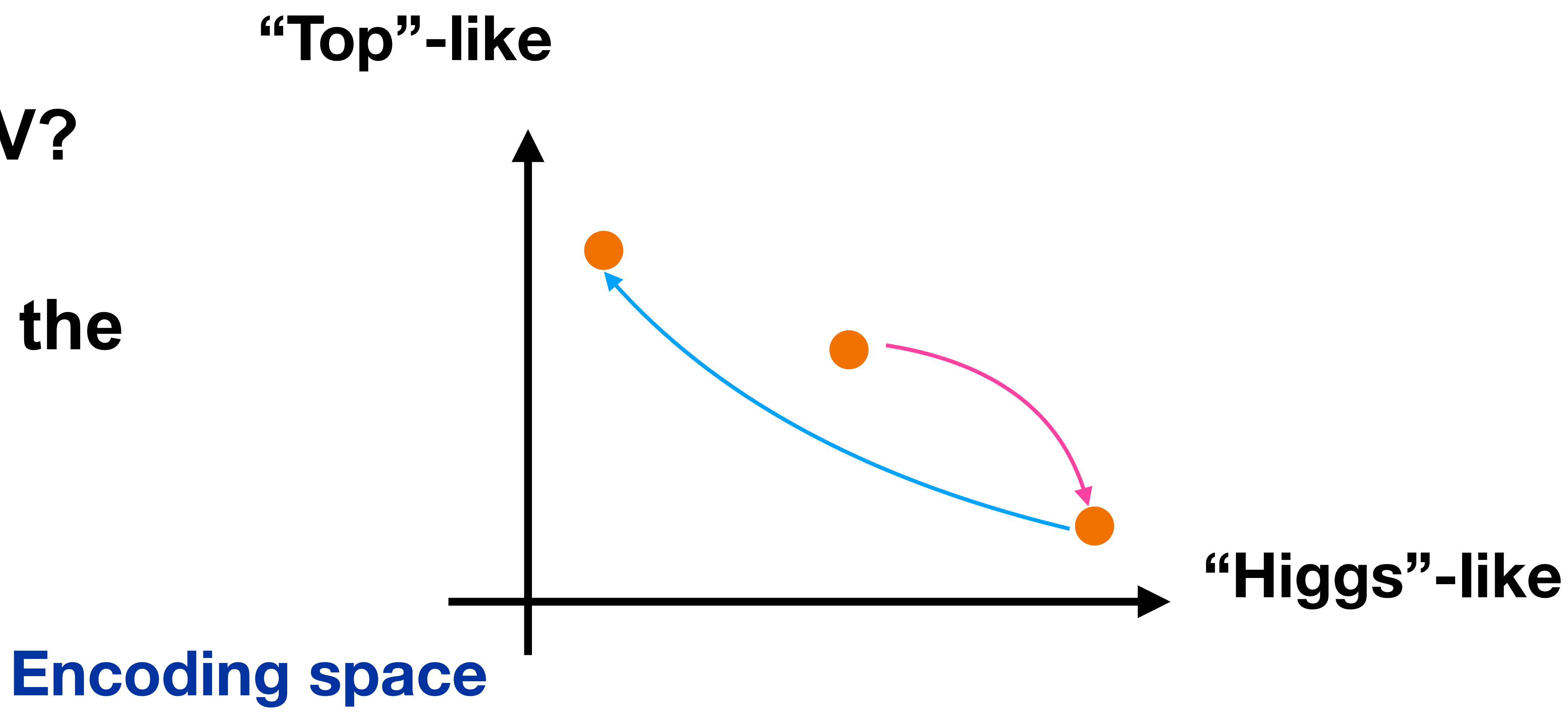




Why would transformers help pairing?

- **Transformers allows for context-dependent encoding of particles**
  - They are "transformed" by the context of the other particles in the event
- **If you see a b-jet in the event, what is the source?**
  - Imagine we have some embedding space where, depending on the kinematics of the b-jet, it is more top-like or Higgs-like.
  - The embedding is defined by some neural network.
- **If there is a second b-jet (and $M(bb)$~125 GeV), would you update your prior?**
- **What if there was also a *jj* pair with $M(bjj)$ ~173 GeV?**
- **Transformers can add these conflicting contexts in the embedding space and pair in auxiliary tasks**

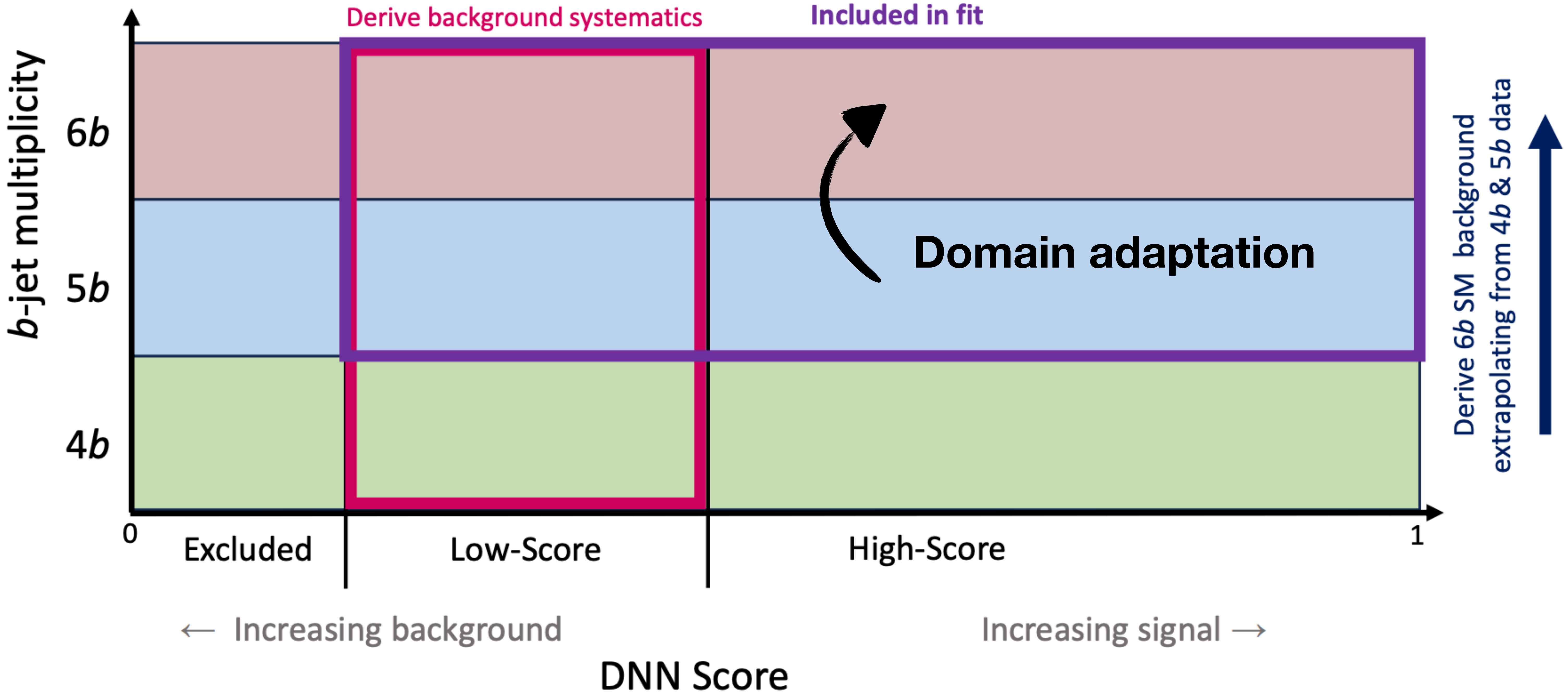




Domain adaptation

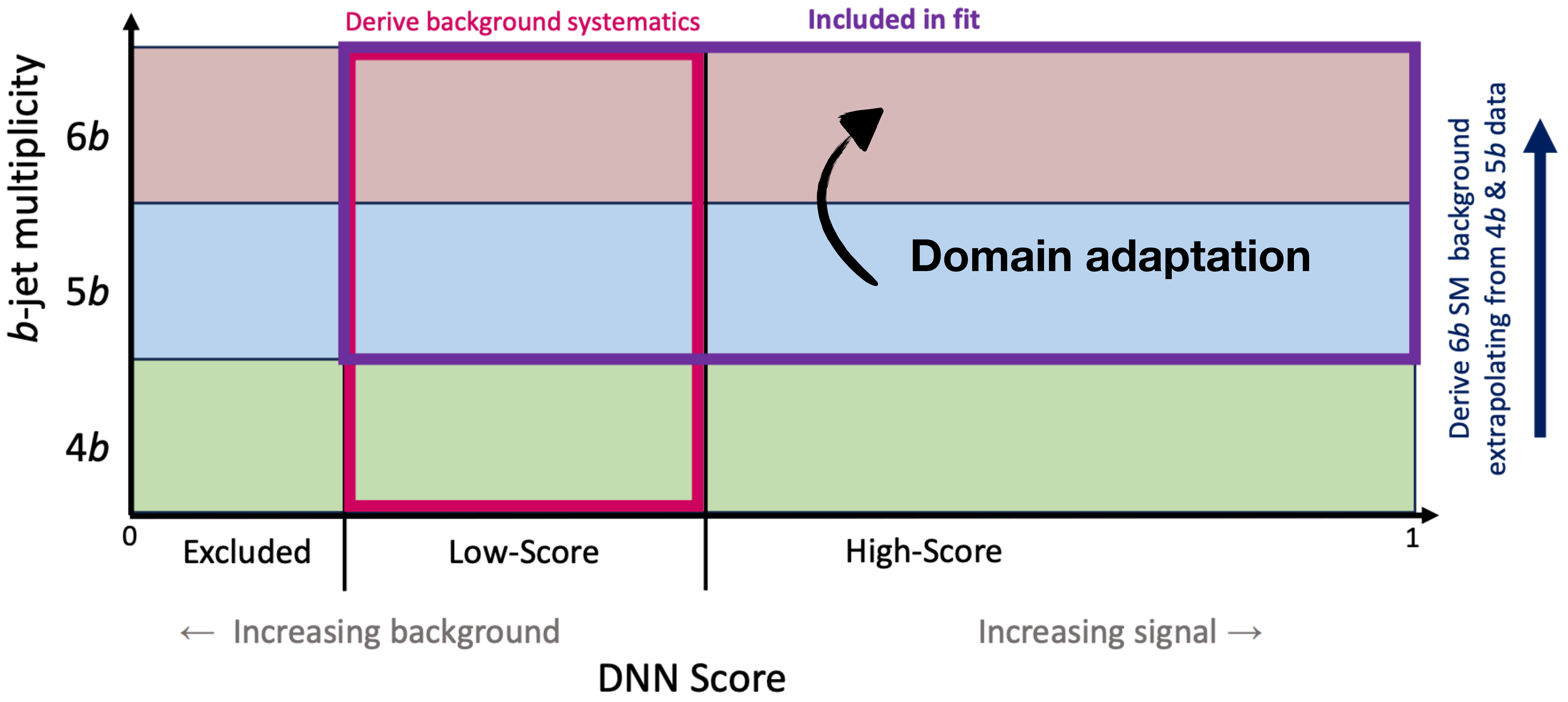
Domain adaptation

- **The problem of training on 5b and applying the NN on 6b is referred to as "domain adaptation"**



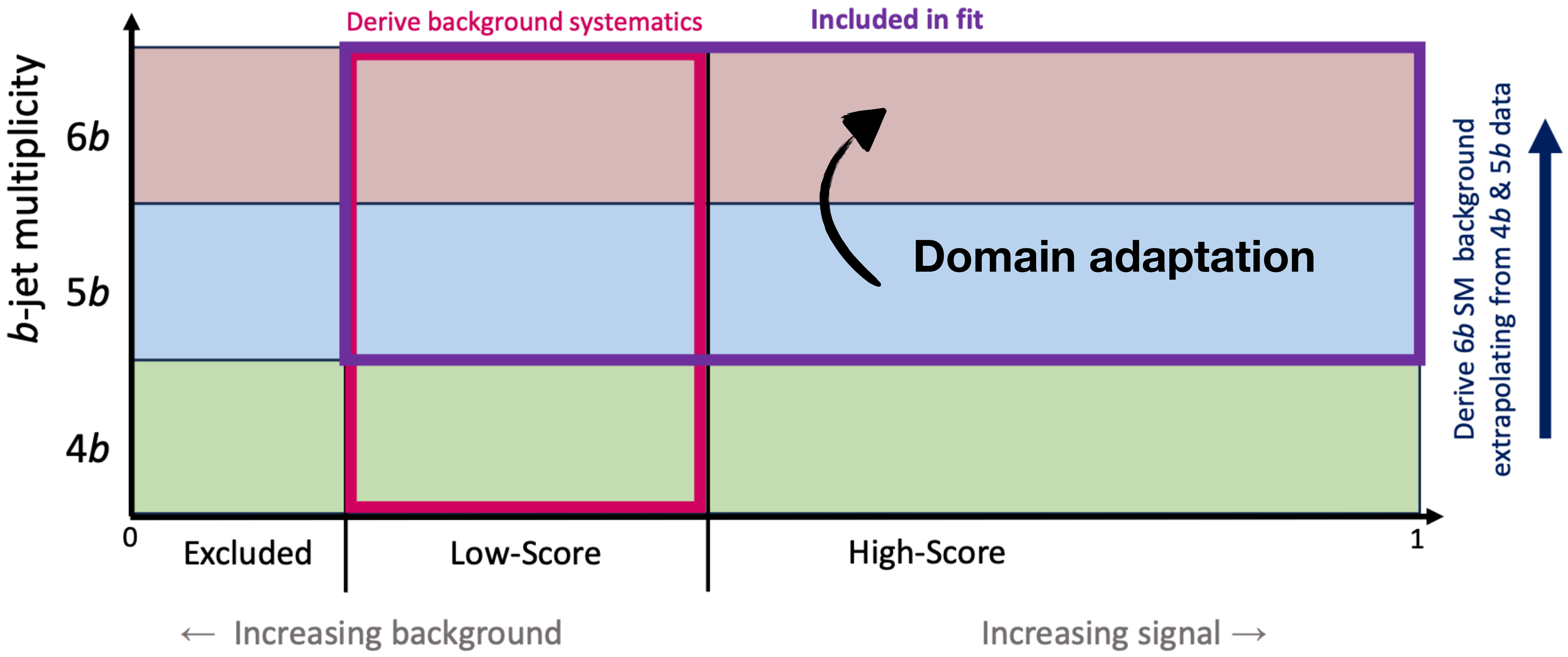
Domain adaptation

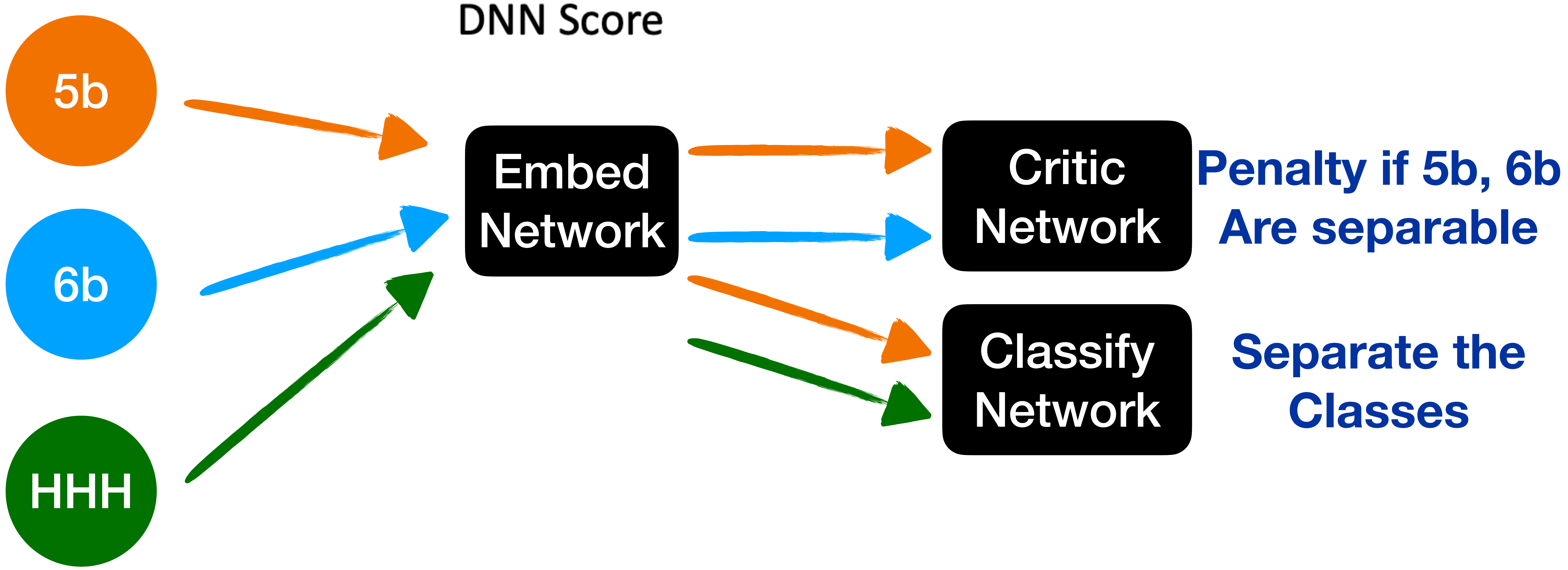
- **The problem of training on 5b and applying the NN on 6b is referred to as "domain adaptation"**
- **We selected input features which minimized difference between 5b and 6b data**
  - But this may have left us with a non-optimal DNN



Domain adaptation

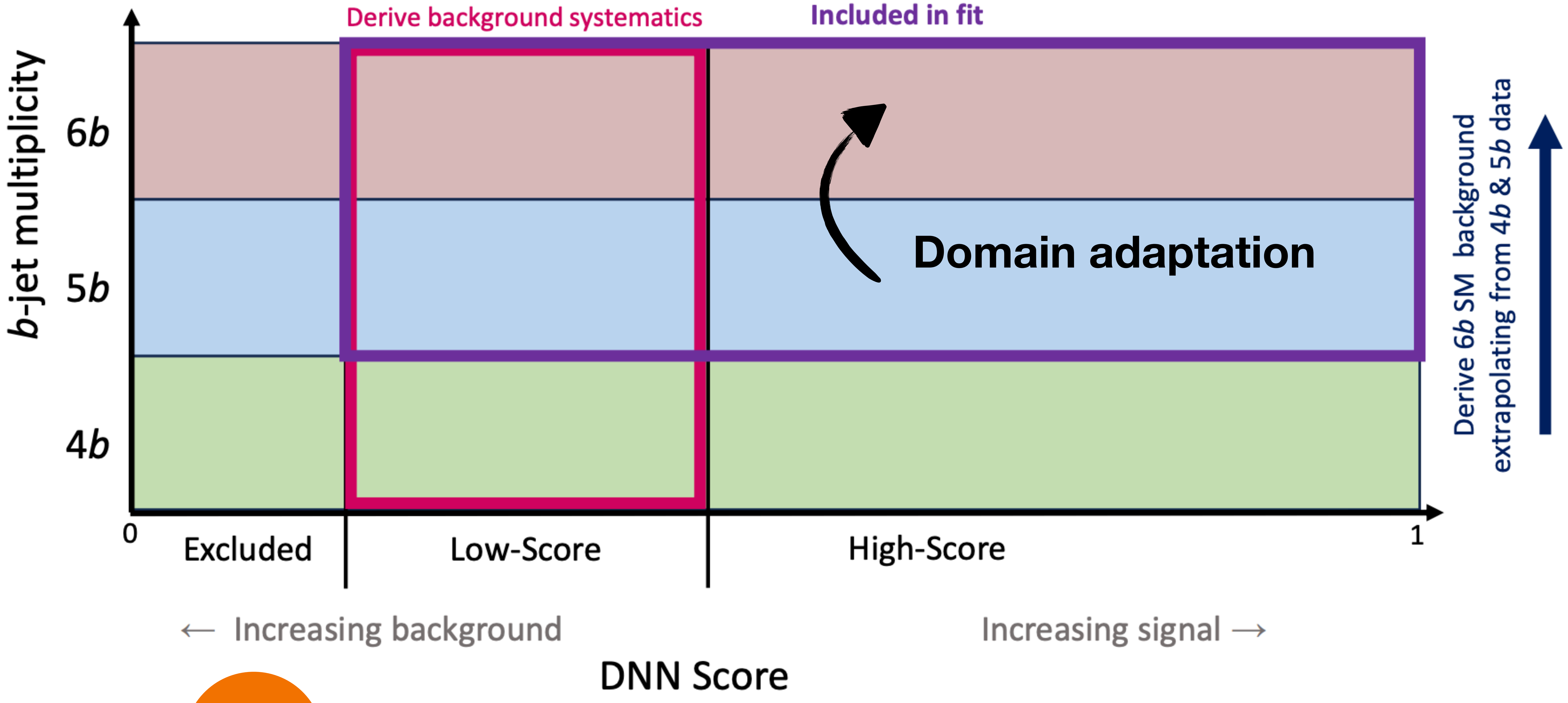
- **The problem of training on 5b and applying the NN on 6b is referred to as "domain adaptation"**
- **We selected input features which minimized difference between 5b and 6b data**
  - But this may have left us with a non-optimal DNN
- **One possible solution is to use an adversarial training, in which a critic network penalizes the training if it can learn any difference between 5b and 6b.**

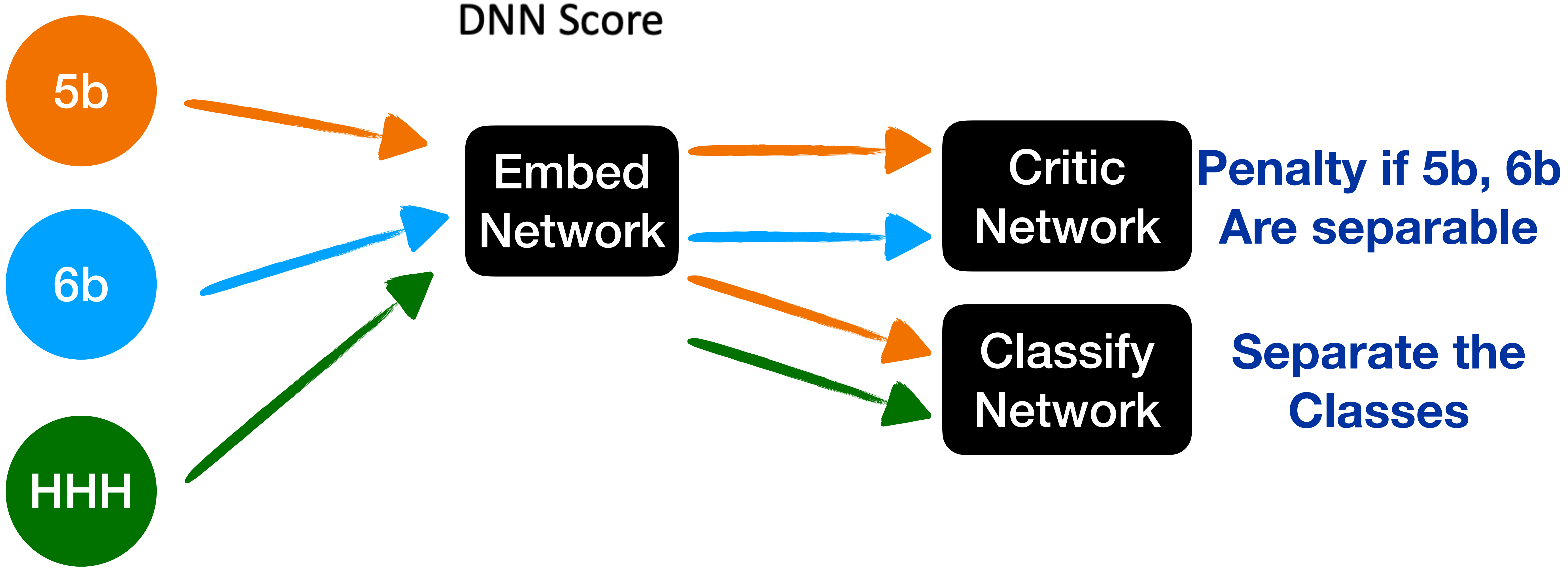




Domain adaptation

- **The problem of training on 5b and applying the NN on 6b is referred to as "domain adaptation"**
- **We selected input features which minimized difference between 5b and 6b data**
  - But this may have left us with a non-optimal DNN
- **One possible solution is to use an adversarial training, in which a critic network penalizes the training if it can learn any difference between 5b and 6b.**
- **The loss function has some "pressure" to learn difference between 5b and HHH signal, but also a competing "pressure" to learn no difference between 5b and 6b.**





Functional decomposition

Functional decomposition

- **Background modeling uncertainties are currently dominant systematic uncertainties.**
  - Generally 2x bigger than FTAG, JET/MET, etc *combined*.

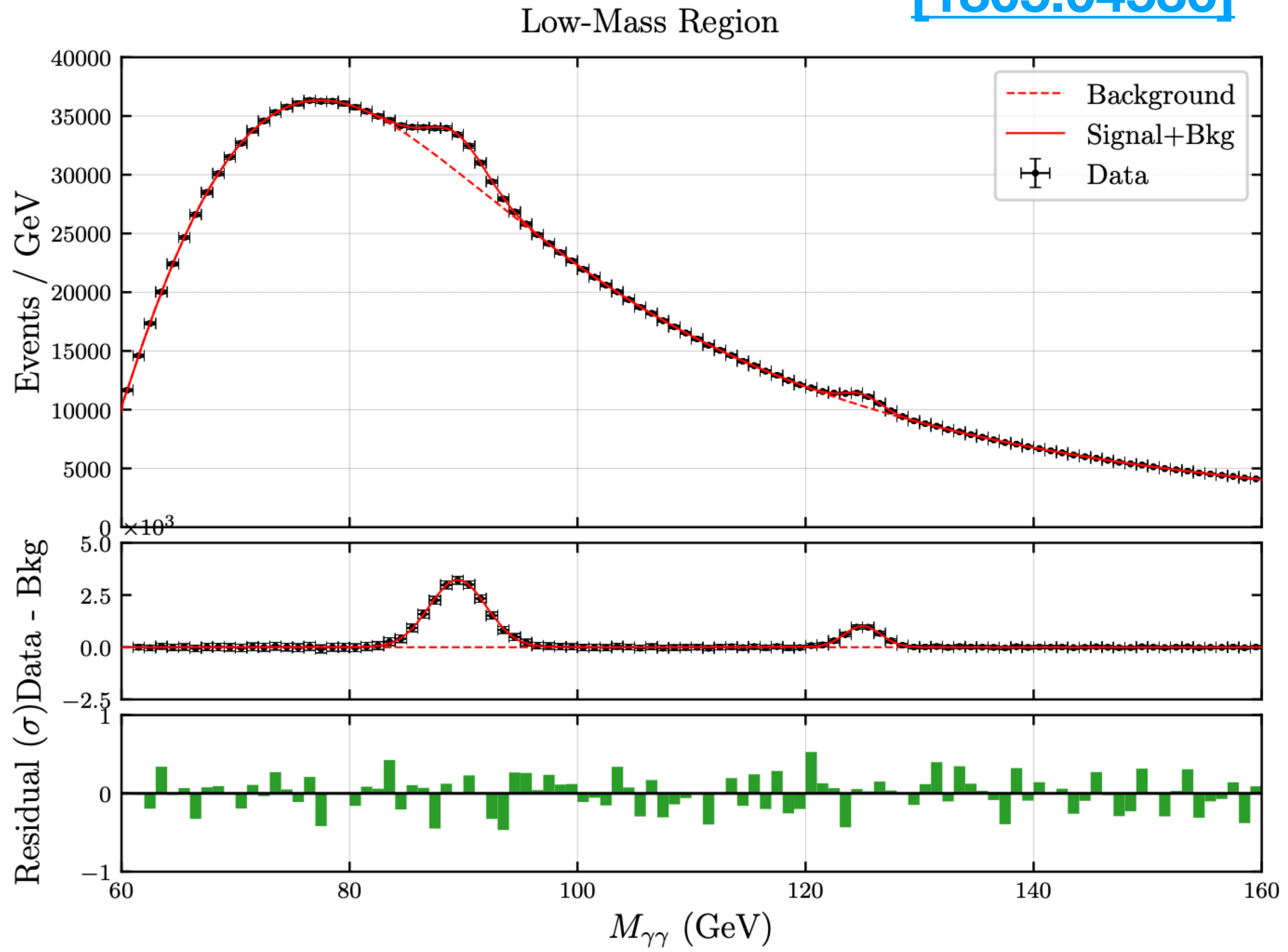
Uncertainty source	SM-like
All uncertainties	24
Experimental	22
Detector response	7.4
Background modeling	16

Functional decomposition

- **Background modeling uncertainties are currently dominant systematic uncertainties.**
  - Generally 2x bigger than FTAG, JET/MET, etc *combined*.
- **In diphoton searches, Michigan developed a new method to model data-driven backgrounds using an orthonormal basis.**
  - Background is by definition the non-resonant component (low "frequency")
  - Signal is resonant (high "frequency")

Uncertainty source	SM-like
All uncertainties	24
Experimental	22
Detector response	7.4
Background modeling	16

[1805.04536]

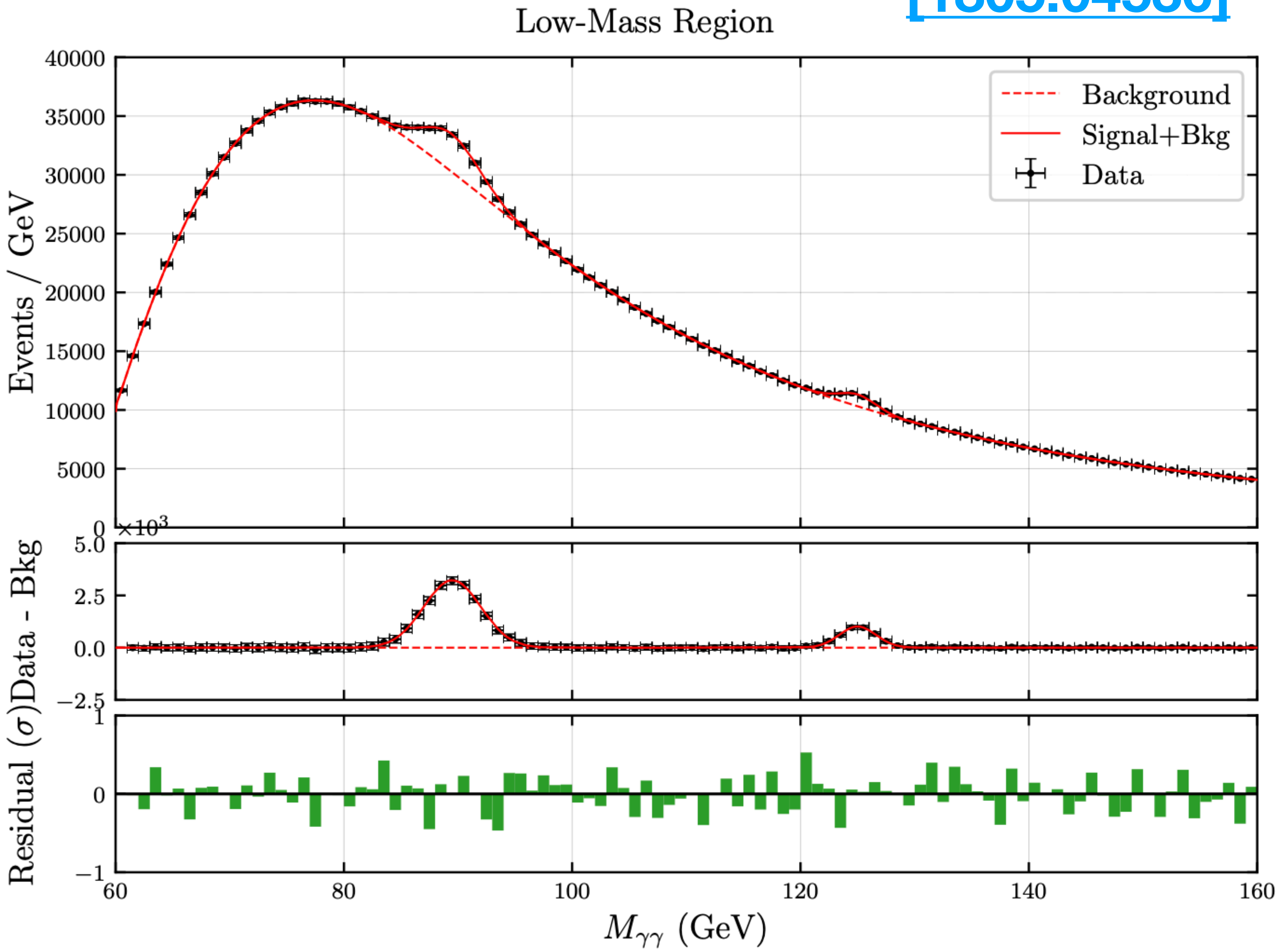


Functional decomposition

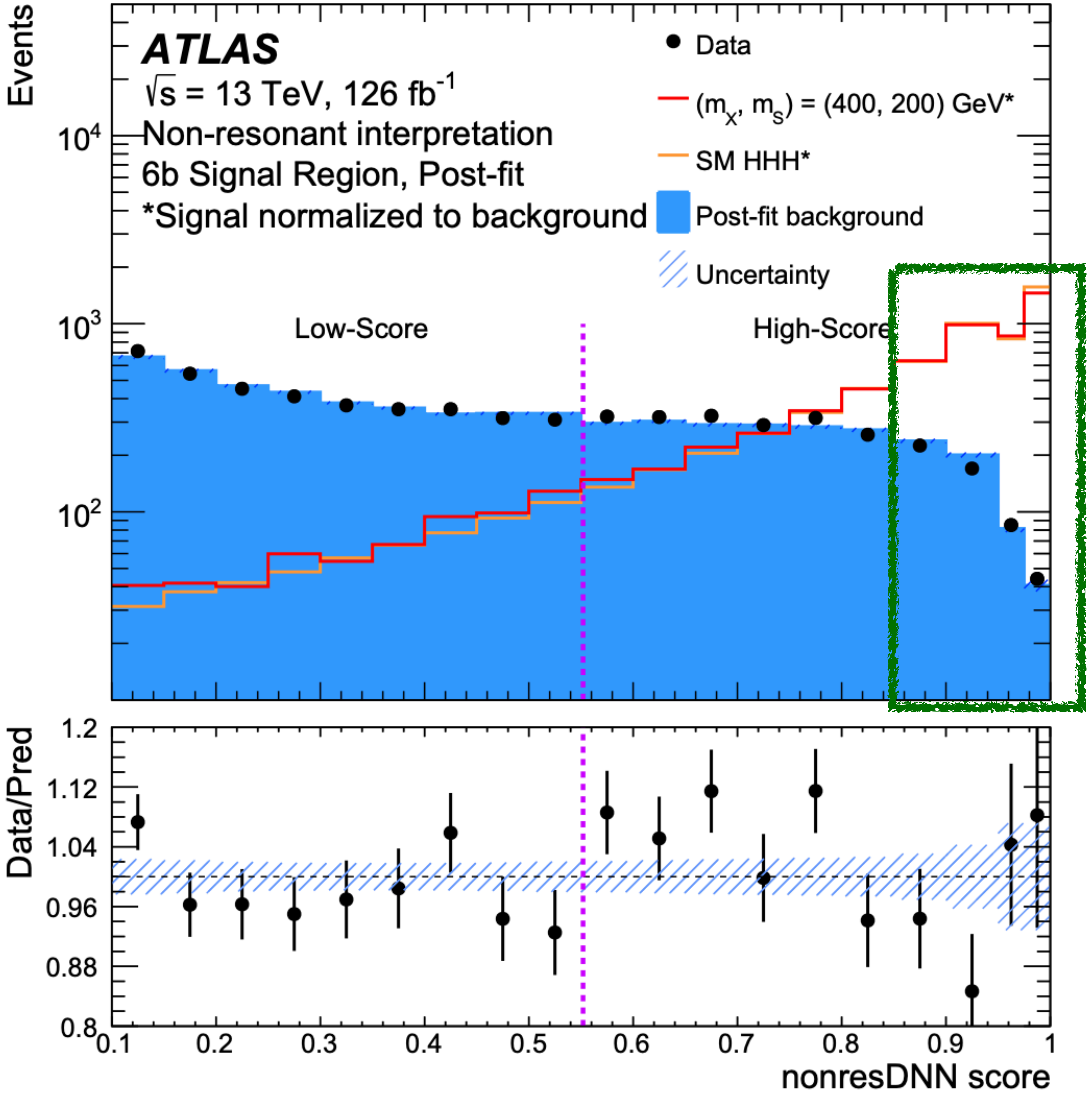
- **Background modeling uncertainties are currently dominant systematic uncertainties.**
  - Generally 2x bigger than FTAG, JET/MET, etc *combined*.
- **In diphoton searches, Michigan developed a new method to model data-driven backgrounds using an orthonormal basis.**
  - Background is by definition the non-resonant component (low "frequency")
  - Signal is resonant (high "frequency")
- **FD reduced systematic uncertainty on data-driven shape by 3x in $\gamma\gamma$ searches, 30% improvement in limit [2102.13405]**

Uncertainty source	SM-like
All uncertainties	24
Experimental	22
Detector response	7.4
Background modeling	16

[1805.04536]

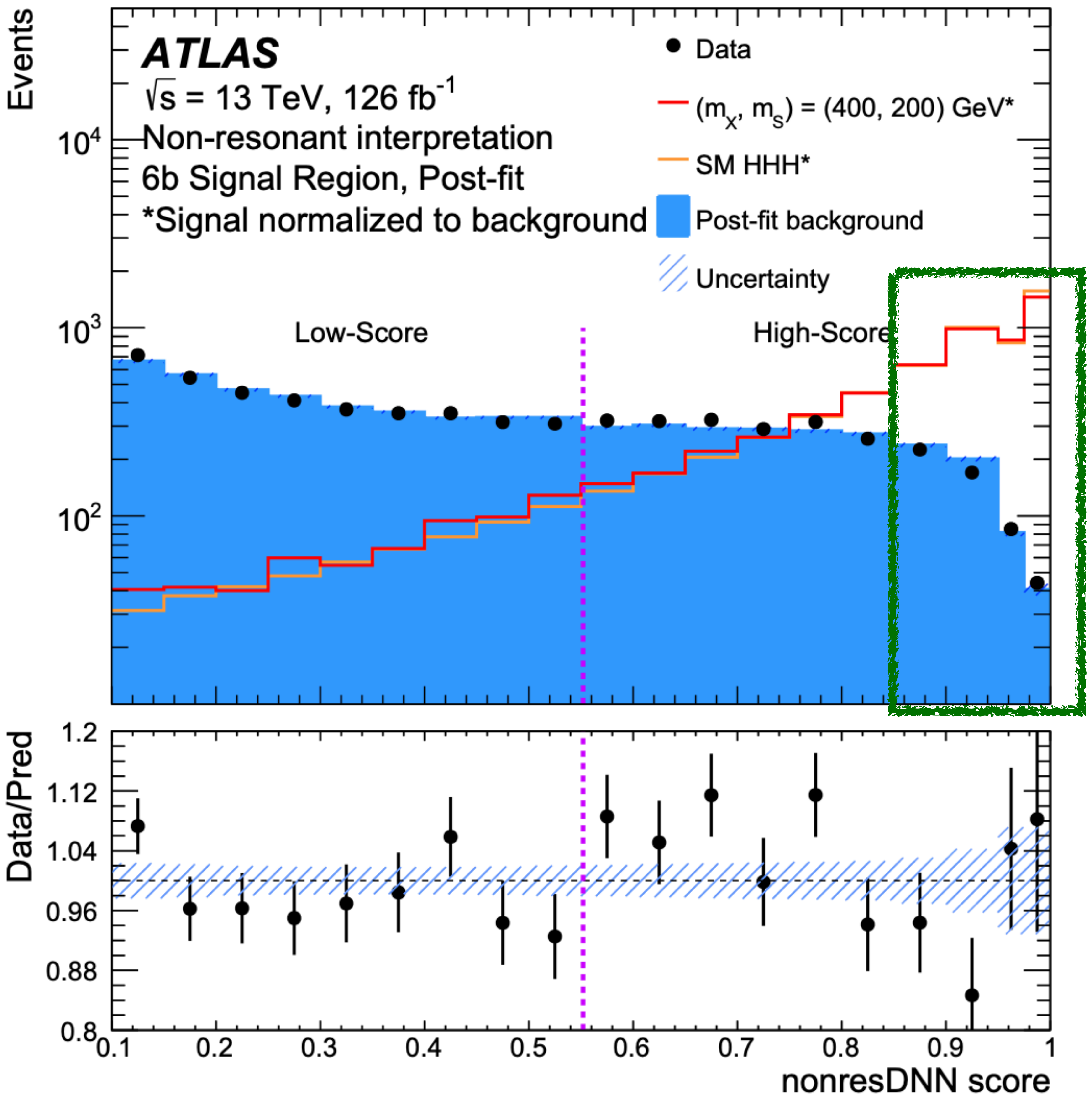


Differentiable p-values with FD



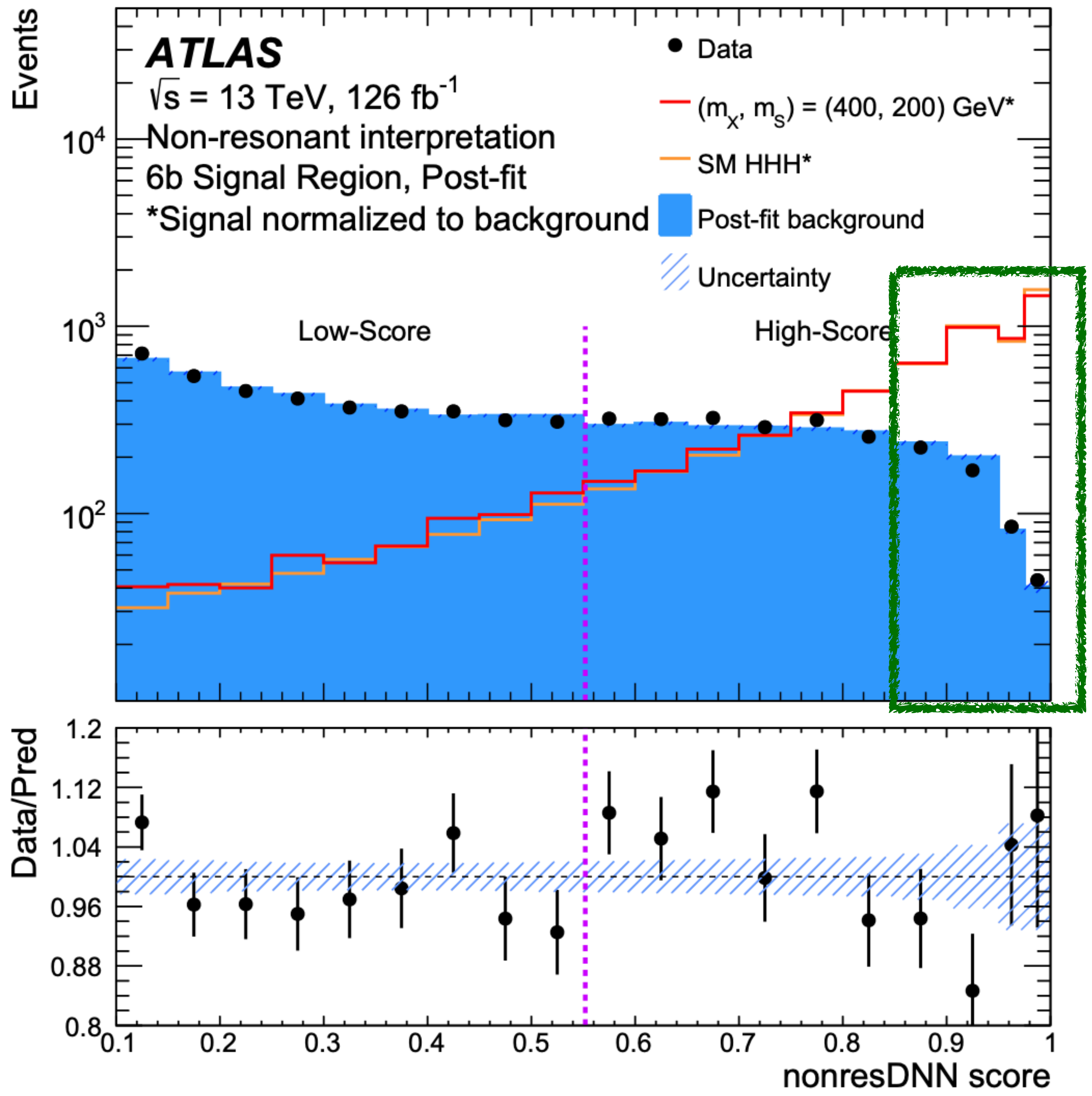
Differentiable p-values with FD

- **A minority of background events have an outsized impact on the sensitivity of the search**



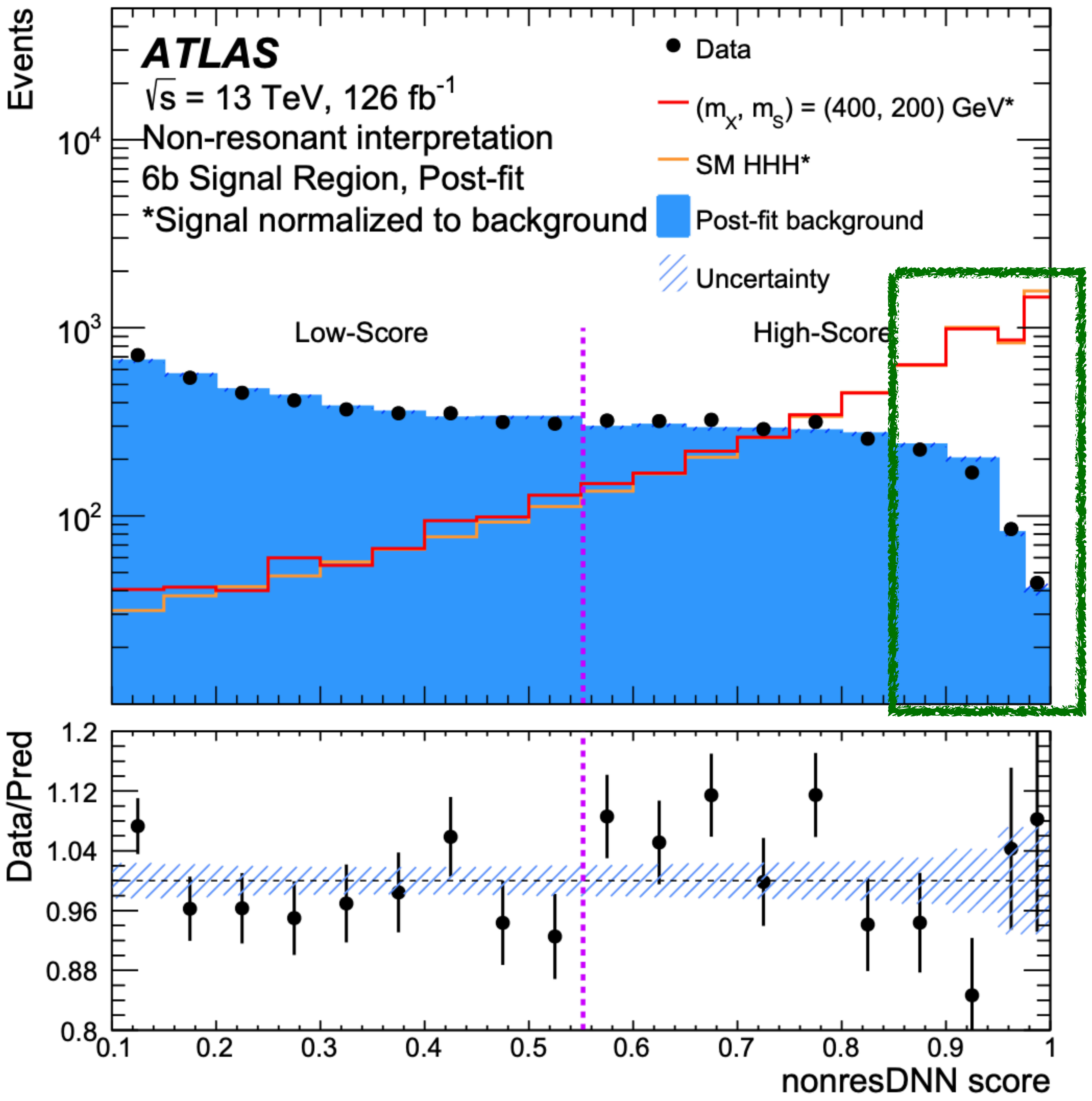
Differentiable p-values with FD

- **A minority of background events have an outsized impact on the sensitivity of the search**
- **Standard cross-entropy loss can become de-coupled from sensitivity in a regime where the bulk of the background is far from the bulk of the signal**



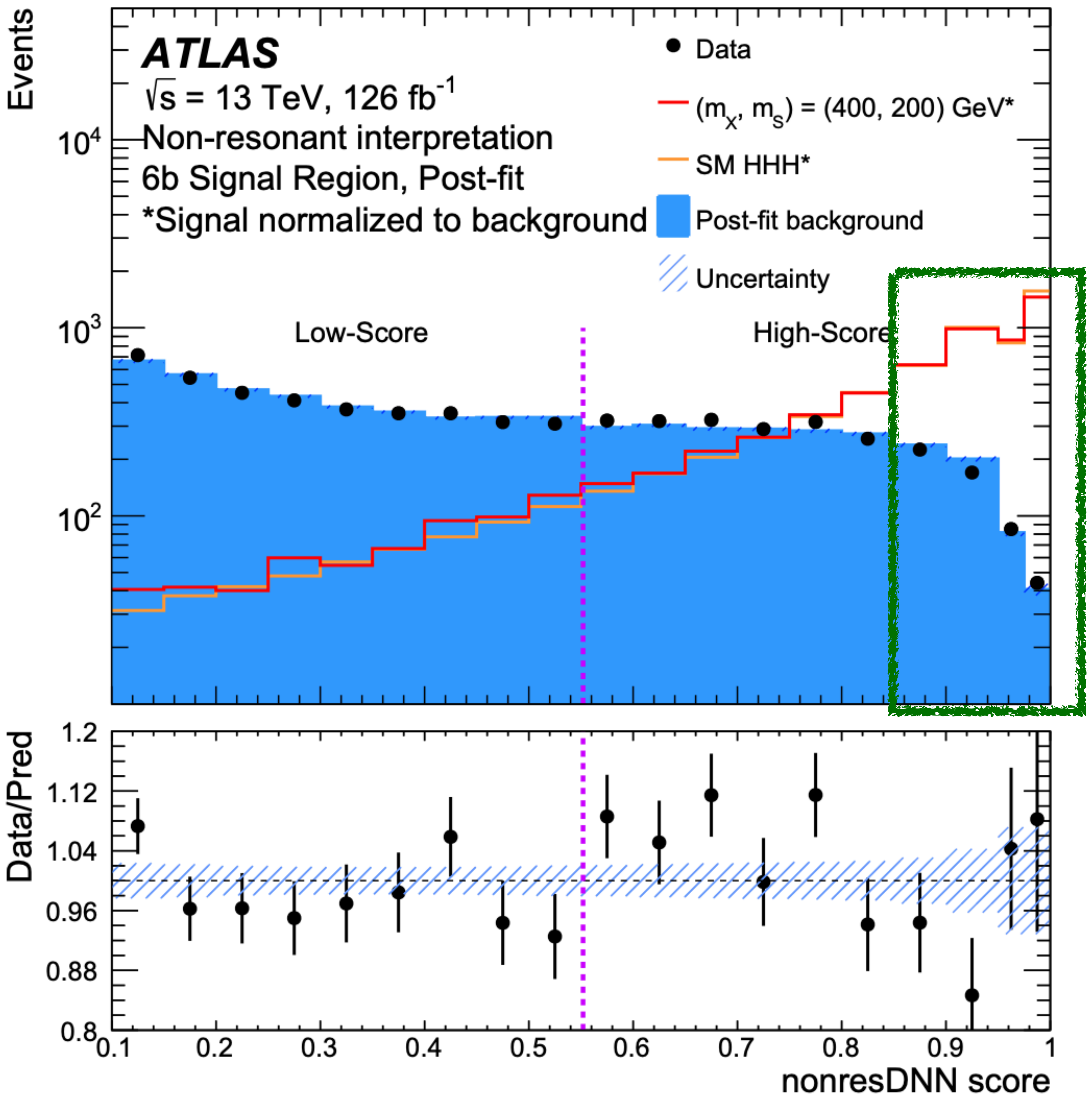
Differentiable p-values with FD

- **A minority of background events have an outsized impact on the sensitivity of the search**
- **Standard cross-entropy loss can become de-coupled from sensitivity in a regime where the bulk of the background is far from the bulk of the signal**
- **Wouldn't it be nice to simply optimize the p-value or limit directly?**



Differentiable p-values with FD

- **A minority of background events have an outsized impact on the sensitivity of the search**
- **Standard cross-entropy loss can become de-coupled from sensitivity in a regime where the bulk of the background is far from the bulk of the signal**
- **Wouldn't it be nice to simply optimize the p-value or limit directly?**
- **Major caveat: of course we need to be careful not to over-fit! This is active research…**



Differentiable p-values with FD

Differentiable p-values with FD

- **FD is a computationally inexpensive way to compute p-values directly from unbinned data**
  - The covariance of the Hilbert space coefficients is directly calculable from the coefficients themselves
  - It takes virtually no extra effort to compute the covariance of the normally-distributed Hilbert space coefficients
  - Once we know the covariance, we can compute a p-value in the frequency domain

$$\begin{aligned}\hat{\Sigma}_{\tilde{\boldsymbol{f}}nm} &= \frac{1}{M}\sum_{i=1}^{M} E_n\left(z_i\right) E_m\left(z_i\right) - \tilde{\boldsymbol{f}}_n \tilde{\boldsymbol{f}}_m \\ &= \tilde{\boldsymbol{f}}^i \hat{\boldsymbol{I}}_{inm} - \tilde{\boldsymbol{f}}_n \tilde{\boldsymbol{f}}_m\end{aligned}$$

  - The coefficients are distributed as a multi-variable normal distribution (and skewness is taken in to account for p-value)

Differentiable p-values with FD

- **FD is a computationally inexpensive way to compute p-values directly from unbinned data**
  - The covariance of the Hilbert space coefficients is directly calculable from the coefficients themselves
  - It takes virtually no extra effort to compute the covariance of the normally-distributed Hilbert space coefficients
  - Once we know the covariance, we can compute a p-value in the frequency domain

$$\hat{\Sigma}_{\tilde{\boldsymbol{f}}nm} = \frac{1}{M}\sum_{i=1}^{M} E_n(z_i)\, E_m(z_i) - \tilde{\boldsymbol{f}}_n \tilde{\boldsymbol{f}}_m$$
$$= \tilde{\boldsymbol{f}}^i \hat{\boldsymbol{I}}_{inm} - \tilde{\boldsymbol{f}}_n \tilde{\boldsymbol{f}}_m$$

  - The coefficients are distributed as a multi-variable normal distribution (and skewness is taken in to account for p-value)

Unbinned data
NN output score

Differentiable p-values with FD

- **FD is a computationally inexpensive way to compute p-values directly from unbinned data**
  - The covariance of the Hilbert space coefficients is directly calculable from the coefficients themselves
  - It takes virtually no extra effort to compute the covariance of the normally-distributed Hilbert space coefficients
  - Once we know the covariance, we can compute a p-value in the frequency domain

$$\begin{aligned}\hat{\boldsymbol{\Sigma}}_{\tilde{\boldsymbol{f}}nm} &= \frac{1}{M}\sum_{i=1}^{M} E_n\left(z_i\right) E_m\left(z_i\right) - \tilde{\boldsymbol{f}}_n \tilde{\boldsymbol{f}}_m \\ &= \tilde{\boldsymbol{f}}^i \hat{\boldsymbol{I}}_{inm} - \tilde{\boldsymbol{f}}_n \tilde{\boldsymbol{f}}_m\end{aligned}$$

  - The coefficients are distributed as a multi-variable normal distribution (and skewness is taken in to account for p-value)

Unbinned data
NN output score → Histogram

Differentiable p-values with FD

- **FD is a computationally inexpensive way to compute p-values directly from unbinned data**
  - The covariance of the Hilbert space coefficients is directly calculable from the coefficients themselves
  - It takes virtually no extra effort to compute the covariance of the normally-distributed Hilbert space coefficients
  - Once we know the covariance, we can compute a p-value in the frequency domain

$$
\begin{aligned}
\hat{\Sigma}_{\tilde{\boldsymbol{f}}nm} &= \frac{1}{M}\sum_{i=1}^{M} E_n\left(z_i\right) E_m\left(z_i\right) - \tilde{\boldsymbol{f}}_n \tilde{\boldsymbol{f}}_m \\
&= \tilde{\boldsymbol{f}}^i \hat{\boldsymbol{I}}_{inm} - \tilde{\boldsymbol{f}}_n \tilde{\boldsymbol{f}}_m
\end{aligned}
$$

  - The coefficients are distributed as a multi-variable normal distribution (and skewness is taken in to account for p-value)

Unbinned data NN output score → Histogram → P-value

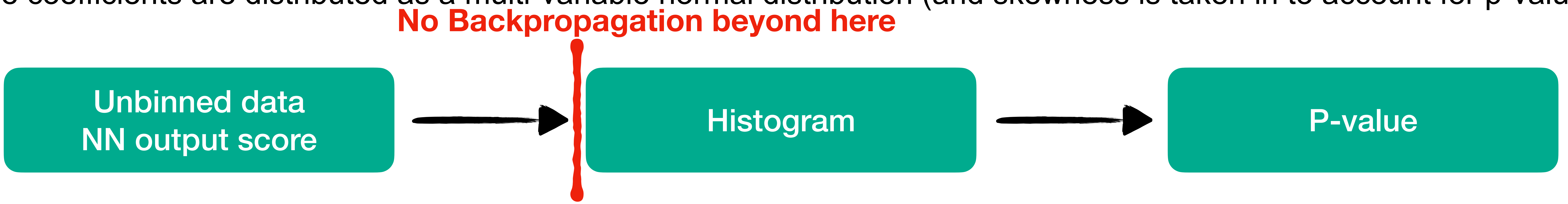
Kevin Nelson

Differentiable p-values with FD

- **FD is a computationally inexpensive way to compute p-values directly from unbinned data**
  - The covariance of the Hilbert space coefficients is directly calculable from the coefficients themselves
  - It takes virtually no extra effort to compute the covariance of the normally-distributed Hilbert space coefficients
  - Once we know the covariance, we can compute a p-value in the frequency domain

$$\begin{aligned}\hat{\Sigma}_{\tilde{f}nm} &= \frac{1}{M}\sum_{i=1}^{M} E_n\left(z_i\right) E_m\left(z_i\right) - \tilde{\boldsymbol{f}}_n \tilde{\boldsymbol{f}}_m \\ &= \tilde{\boldsymbol{f}}^i \hat{\boldsymbol{I}}_{inm} - \tilde{\boldsymbol{f}}_n \tilde{\boldsymbol{f}}_m\end{aligned}$$

  - The coefficients are distributed as a multi-variable normal distribution (and skewness is taken in to account for p-value)

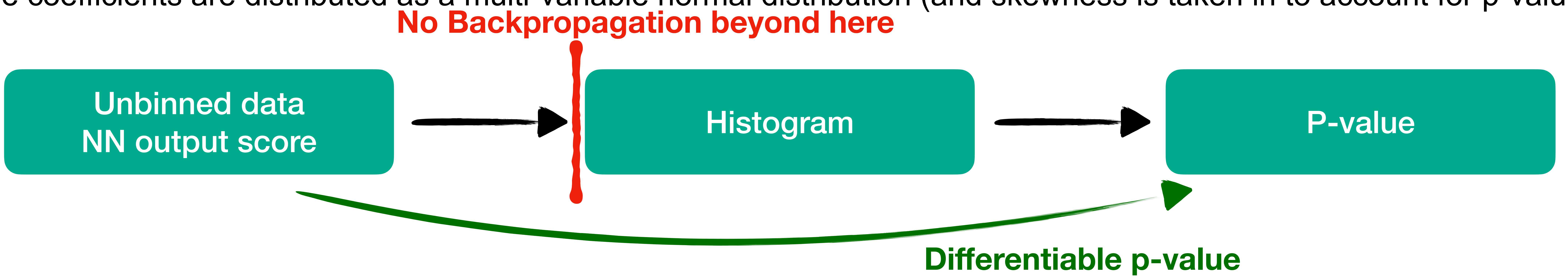


Differentiable p-values with FD

- **FD is a computationally inexpensive way to compute p-values directly from unbinned data**
  - The covariance of the Hilbert space coefficients is directly calculable from the coefficients themselves
  - It takes virtually no extra effort to compute the covariance of the normally-distributed Hilbert space coefficients
  - Once we know the covariance, we can compute a p-value in the frequency domain

$$\begin{aligned}\hat{\boldsymbol{\Sigma}}_{\tilde{\boldsymbol{f}}nm} &= \frac{1}{M}\sum_{i=1}^{M} E_n\left(z_i\right) E_m\left(z_i\right) - \tilde{\boldsymbol{f}}_n \tilde{\boldsymbol{f}}_m \\ &= \tilde{\boldsymbol{f}}^i \hat{\boldsymbol{I}}_{inm} - \tilde{\boldsymbol{f}}_n \tilde{\boldsymbol{f}}_m\end{aligned}$$

  - The coefficients are distributed as a multi-variable normal distribution (and skewness is taken in to account for p-value)

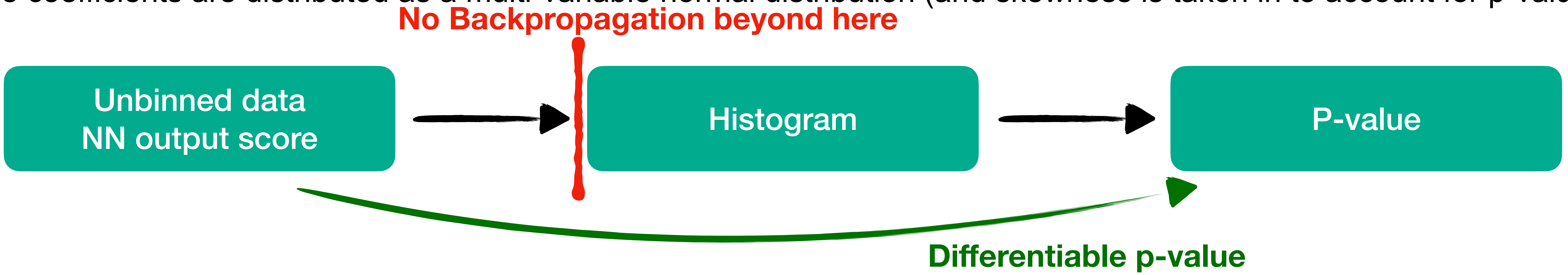


Differentiable p-values with FD

- **FD is a computationally inexpensive way to compute p-values directly from unbinned data**
  - The covariance of the Hilbert space coefficients is directly calculable from the coefficients themselves
  - It takes virtually no extra effort to compute the covariance of the normally-distributed Hilbert space coefficients
  - Once we know the covariance, we can compute a p-value in the frequency domain

$$\begin{aligned}\hat{\Sigma}_{\tilde{\boldsymbol{f}}nm} &= \frac{1}{M}\sum_{i=1}^{M} E_n\left(z_i\right) E_m\left(z_i\right) - \tilde{\boldsymbol{f}}_n \tilde{\boldsymbol{f}}_m \\ &= \tilde{\boldsymbol{f}}^i \hat{\boldsymbol{I}}_{inm} - \tilde{\boldsymbol{f}}_n \tilde{\boldsymbol{f}}_m\end{aligned}$$

  - The coefficients are distributed as a multi-variable normal distribution (and skewness is taken in to account for p-value)



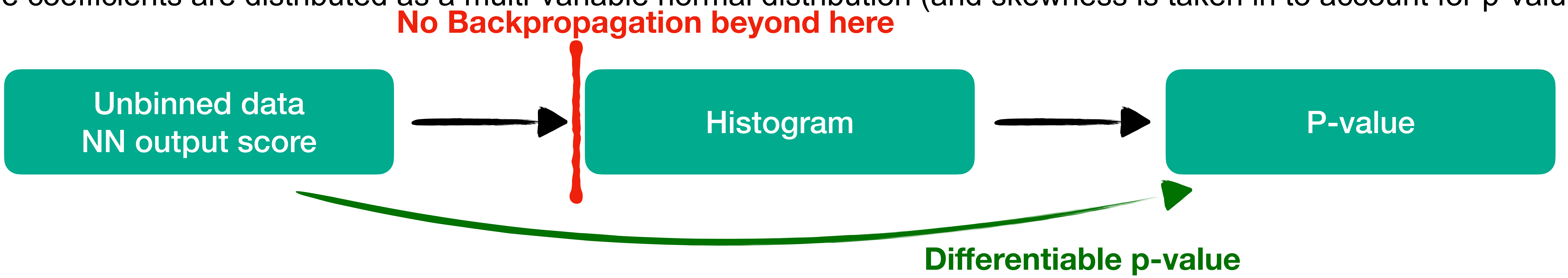
- **The idea is to put an “FD p-value layer” in a neural network so that we can optimize the sensitivity of the analysis directly, rather than optimize the binary cross-entropy loss.**

Differentiable p-values with FD

- **FD is a computationally inexpensive way to compute p-values directly from unbinned data**
  - The covariance of the Hilbert space coefficients is directly calculable from the coefficients themselves
  - It takes virtually no extra effort to compute the covariance of the normally-distributed Hilbert space coefficients
  - Once we know the covariance, we can compute a p-value in the frequency domain

$$\begin{aligned}\hat{\Sigma}_{\tilde{\boldsymbol{f}}nm} &= \frac{1}{M}\sum_{i=1}^{M} E_n\left(z_i\right) E_m\left(z_i\right) - \tilde{\boldsymbol{f}}_n \tilde{\boldsymbol{f}}_m \\ &= \tilde{\boldsymbol{f}}^i \hat{\boldsymbol{I}}_{inm} - \tilde{\boldsymbol{f}}_n \tilde{\boldsymbol{f}}_m\end{aligned}$$

  - The coefficients are distributed as a multi-variable normal distribution (and skewness is taken in to account for p-value)



- **The idea is to put an "FD p-value layer" in a neural network so that we can optimize the sensitivity of the analysis directly, rather than optimize the binary cross-entropy loss.**
- **Another benefit: with FD background model is learned in a semi-supervised manner**

Summary

Summary

- Is there reason to be interested in HHH, when we haven't even found HH?

Summary

- **Is there reason to be interested in HHH, when we haven't even found HH?**

- **Yes!**
  - Right now, in LHC Run 2/3, HHH is constraining phase space of BSM models with additional scalars
  - In some senses, HHH is "behind" HH, and we need to develop our strategies over time:
    - Reduction of theoretical uncertainties
    - Explore multiple channels
  - Novel machine learning ideas: domain adaptation, functional decomposition, pairing tasks (GNN/transformer), etc.

Summary

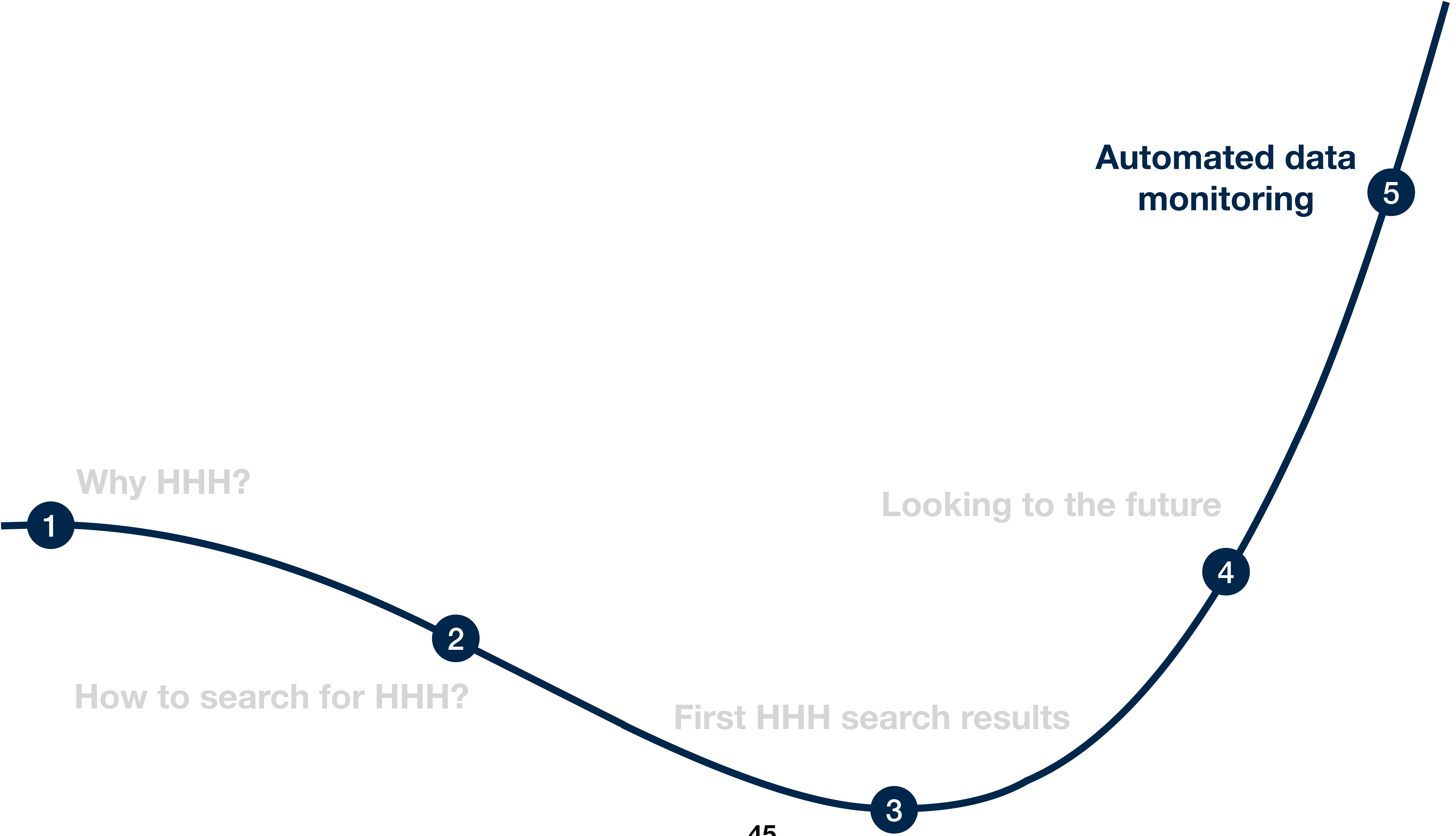
- **Is there reason to be interested in HHH, when we haven't even found HH?**

- **Yes!**
  - Right now, in LHC Run 2/3, HHH is constraining phase space of BSM models with additional scalars
  - In some senses, HHH is "behind" HH, and we need to develop our strategies over time:
    - Reduction of theoretical uncertainties
    - Explore multiple channels
  - Novel machine learning ideas: domain adaptation, functional decomposition, pairing tasks (GNN/transformer), etc.

- **Even if you aren't excited about HHH, is there reason to be excited about this talk?**

Summary

- **Is there reason to be interested in HHH, when we haven't even found HH?**
- **Yes!**
  - Right now, in LHC Run 2/3, HHH is constraining phase space of BSM models with additional scalars
  - In some senses, HHH is "behind" HH, and we need to develop our strategies over time:
    - Reduction of theoretical uncertainties
    - Explore multiple channels
  - Novel machine learning ideas: domain adaptation, functional decomposition, pairing tasks (GNN/transformer), etc.
- **Even if you aren't excited about HHH, is there reason to be excited about this talk?**
- **Yes!**
  - Many of the ML/AI ideas discussed at the end are entirely general!


1
Why HHH?
2
How to search for HHH?
3
First HHH search results
4
Looking to the future
5
Automated data monitoring

Muon performance studies

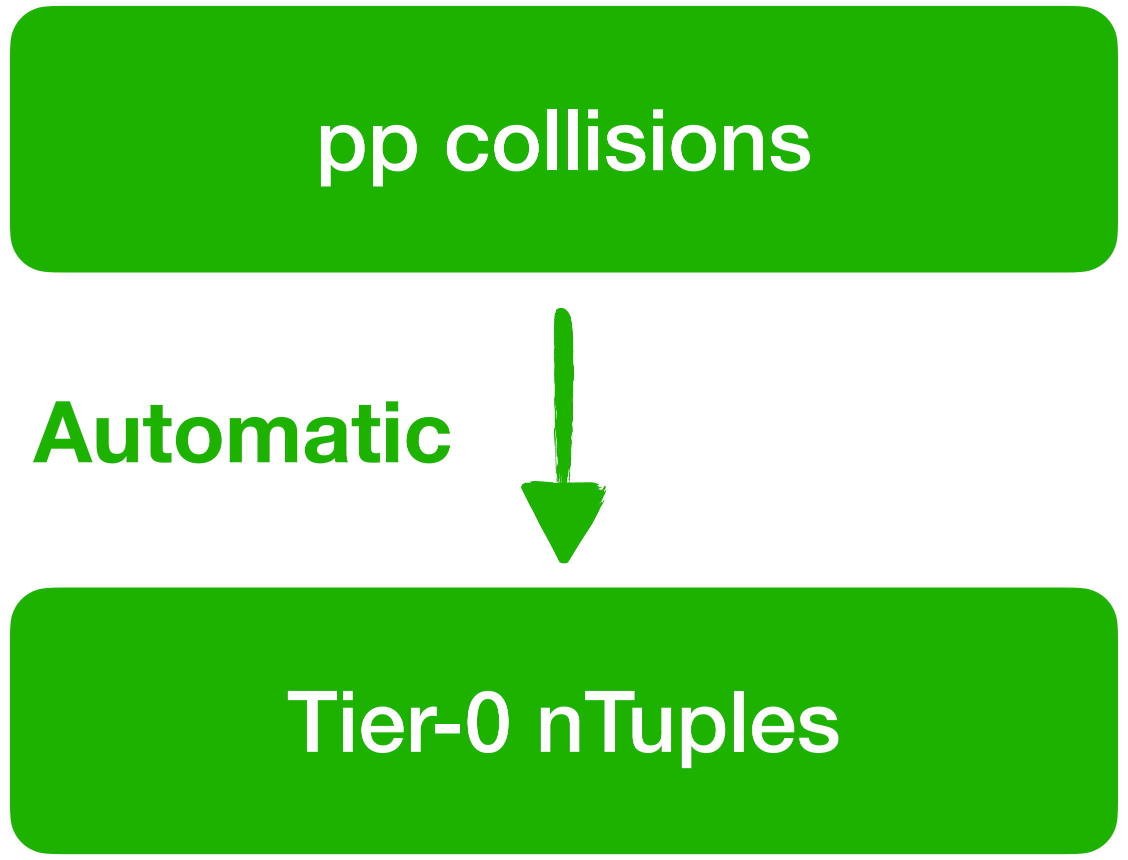
Muon performance studies

- **In the muon performance group we analyze data and MC to:**
  - Measure SFs
  - Calibrate muon p_T
  - Apply corrections to data

Muon performance studies

- **In the muon performance group we analyze data and MC to:**
  - Measure SFs
  - Calibrate muon p_T
  - Apply corrections to data

- **When data are collected, the muon performance group has (for a while) automatically produced nTuples for analysis**



Muon performance studies

- **In the muon performance group we analyze data and MC to:**
  - Measure SFs
  - Calibrate muon p_T
  - Apply corrections to data

- **When data are collected, the muon performance group has (for a while) automatically produced nTuples for analysis**

- **The nTuples are not automatically analyzed (no plots, histograms produced).**
  - Analysis of nTuples is “self-service”, like the kitchens in the CERN hotel
  - Can we make it automatic (to the user), like R1?
  - There is simply delicious food when you arrive

pp collisions

Automatic ↓

Tier-0 nTuples

Self-service ↓

Muon performance analysis toolkit

Muon performance studies

- **In the muon performance group we analyze data and MC to:**
  - Measure SFs
  - Calibrate muon p_T
  - Apply corrections to data
- **When data are collected, the muon performance group has (for a while) automatically produced nTuples for analysis**
- **The nTuples are not automatically analyzed (no plots, histograms produced).**
  - Analysis of nTuples is “self-service”, like the kitchens in the CERN hotel
  - Can we make it automatic (to the user), like R1?
  - There is simply delicious food when you arrive
- **The lack of automation has real consequences**
  - This year, some bugs in muon reconstruction were found quite late
  - We had to re-run a significant amount of computation

pp collisions

Automatic ↓

Tier-0 nTuples

Self-service ↓

Muon performance analysis toolkit

Muon performance studies

- **In the muon performance group we analyze data and MC to:**
  - Measure SFs
  - Calibrate muon p_T
  - Apply corrections to data

- **When data are collected, the muon performance group has (for a while) automatically produced nTuples for analysis**

- **The nTuples are not automatically analyzed (no plots, histograms produced).**
  - Analysis of nTuples is “self-service”, like the kitchens in the CERN hotel
  - Can we make it automatic (to the user), like R1?
  - There is simply delicious food when you arrive

- **The lack of automation has real consequences**
  - This year, some bugs in muon reconstruction were found quite late
  - We had to re-run a significant amount of computation

- **Even worse, sometimes we lose “personal” data analysis “recipes”**

pp collisions

Automatic ↓

Tier-0 nTuples

Self-service ↓

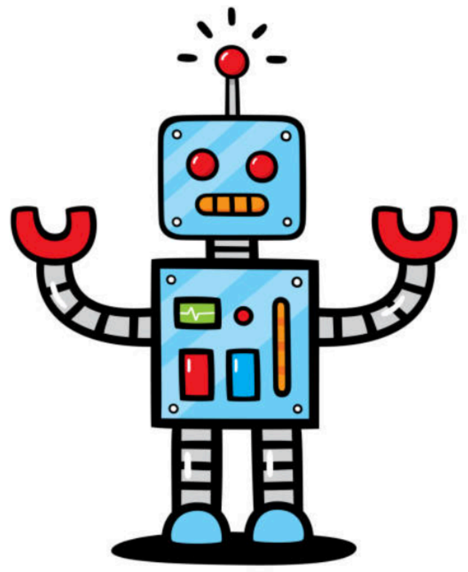
Muon performance analysis toolkit

New pipeline framework

- The goal is to **accelerate muon performance checks** via **preservation** and **automation**

New pipeline framework

- The goal is to **accelerate muon performance checks** via **preservation** and **automation**

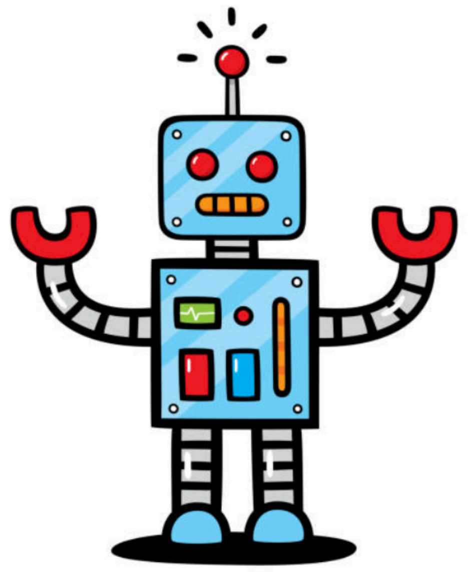


Automated trigger: for Example when a new nTuple is published at Tier0

On demand user trigger: for example to **check a new set of recommendations**

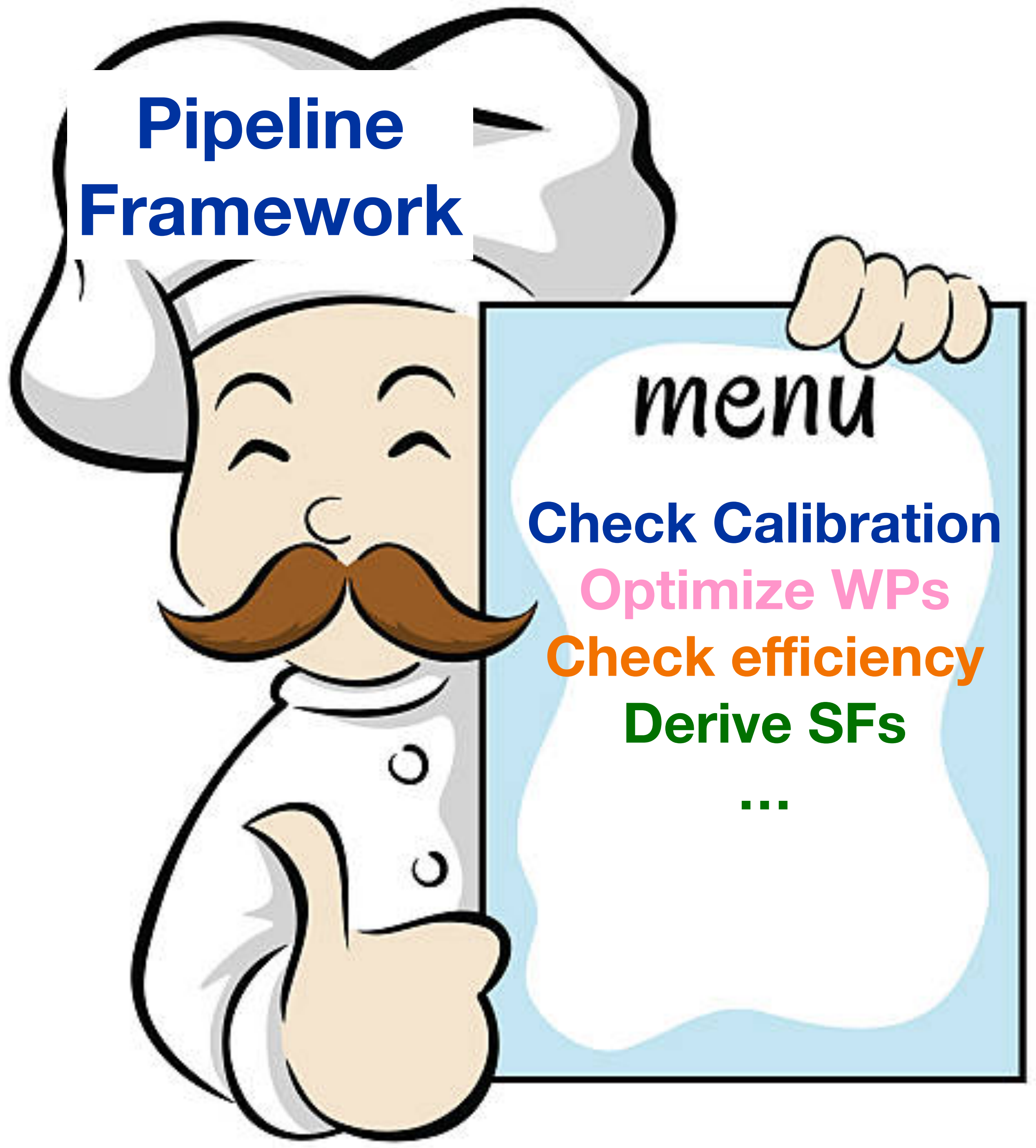
New pipeline framework

- The goal is to **accelerate muon performance checks** via **preservation** and **automation**



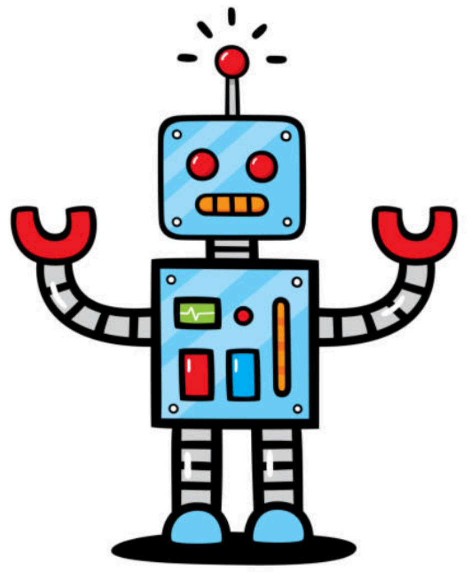
Automated trigger: for Example when a new nTuple is published at Tier0

On demand user trigger: for example to **check a new set of recommendations**



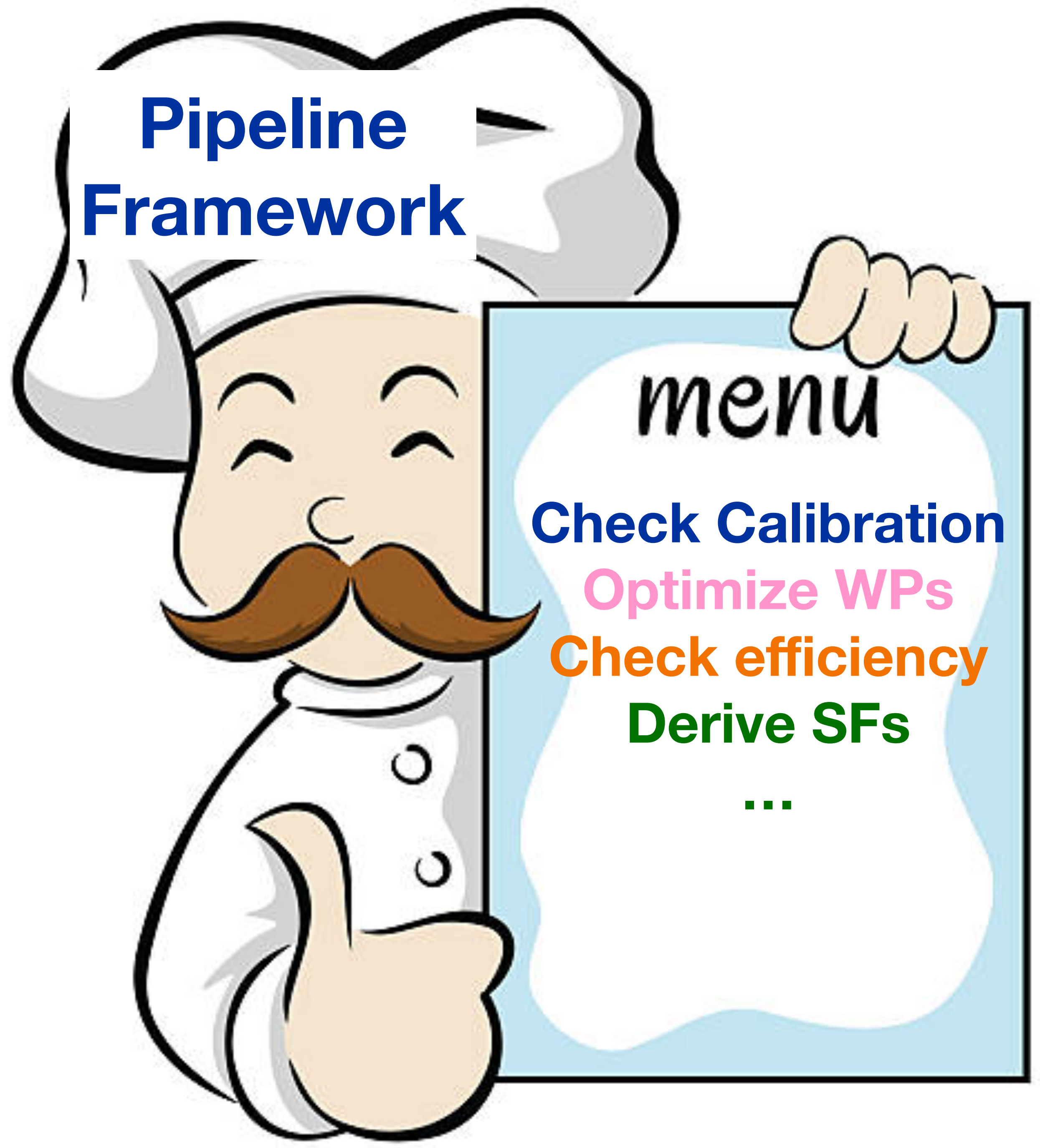
New pipeline framework

- The goal is to **accelerate muon performance checks** via **preservation** and **automation**



Automated trigger: for Example when a new nTuple is published at Tier0

On demand user trigger: for example to **check a new set of recommendations**

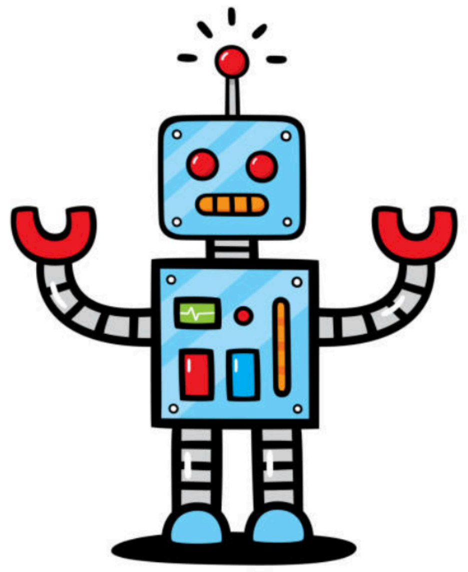


Trigger orders a specific "recipe" from the "menu"
Then the pipeline workflow handles interaction with computing grid
All code executed is in docker image
Naturally requires all code to be committed in git before execution!!
Anyone can run *any* code, forever. No compiling, no loss of "personal" code to time

New pipeline framework

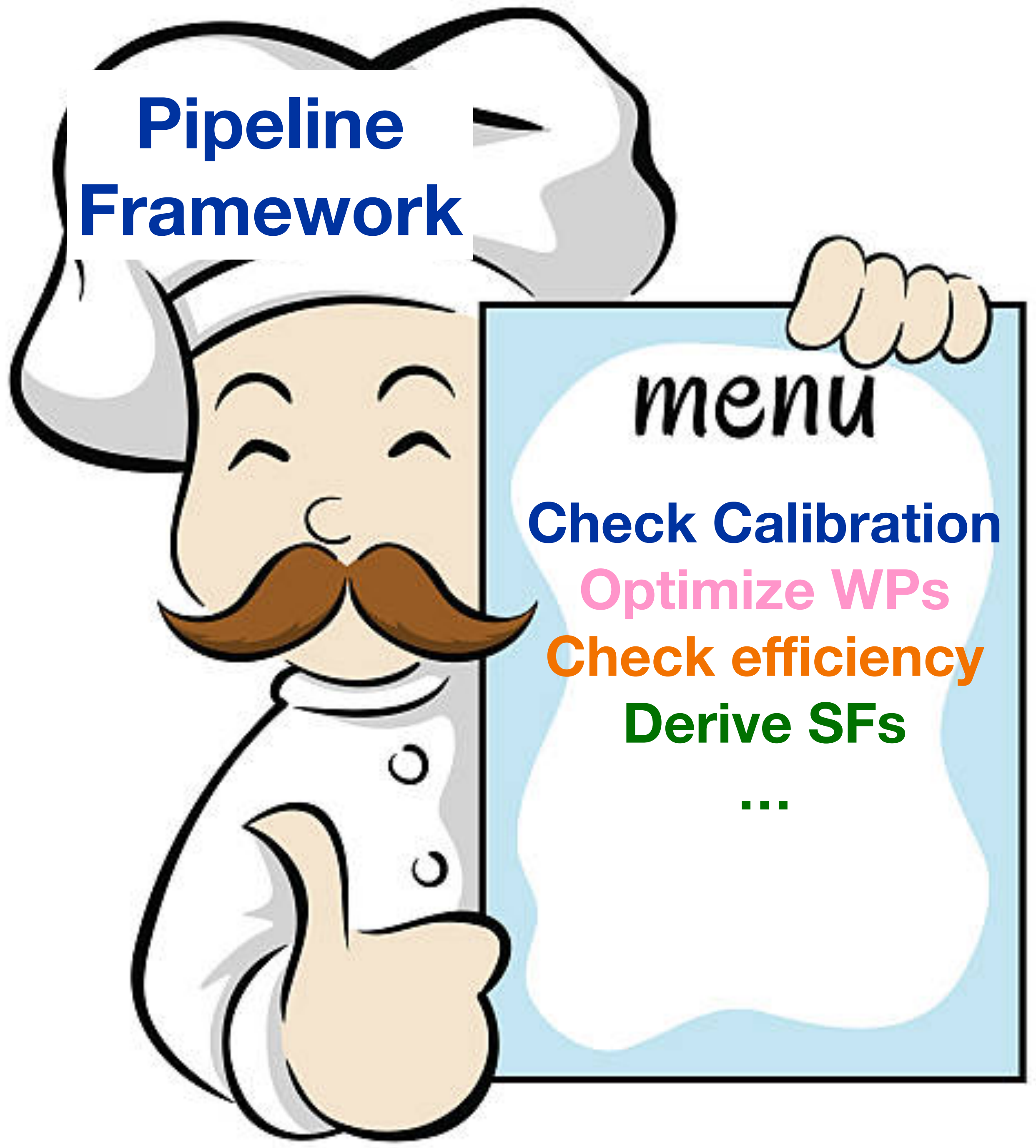
- The goal is to **accelerate muon performance checks** via **preservation** and **automation**

Muon performance analysis toolkit
Packaged into docker images



Automated trigger: for Example when a new nTuple is published at Tier0

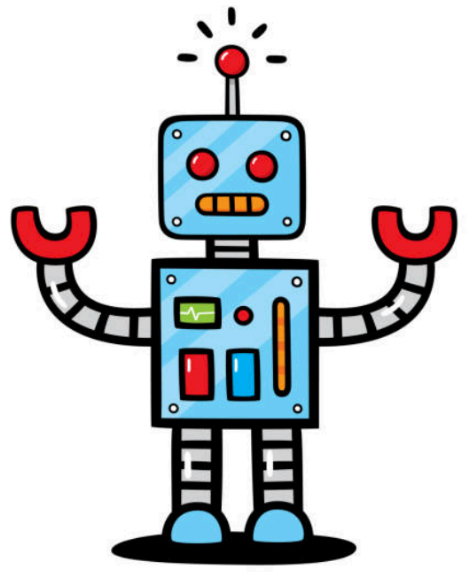
On demand user trigger: for example to **check a new set of recommendations**



Trigger orders a specific "recipe" from the "menu"
Then the pipeline workflow handles interaction with computing grid
All code executed is in docker image
Naturally requires all code to be committed in git before execution!!
Anyone can run *any* code, forever. No compiling, no loss of "personal" code to time

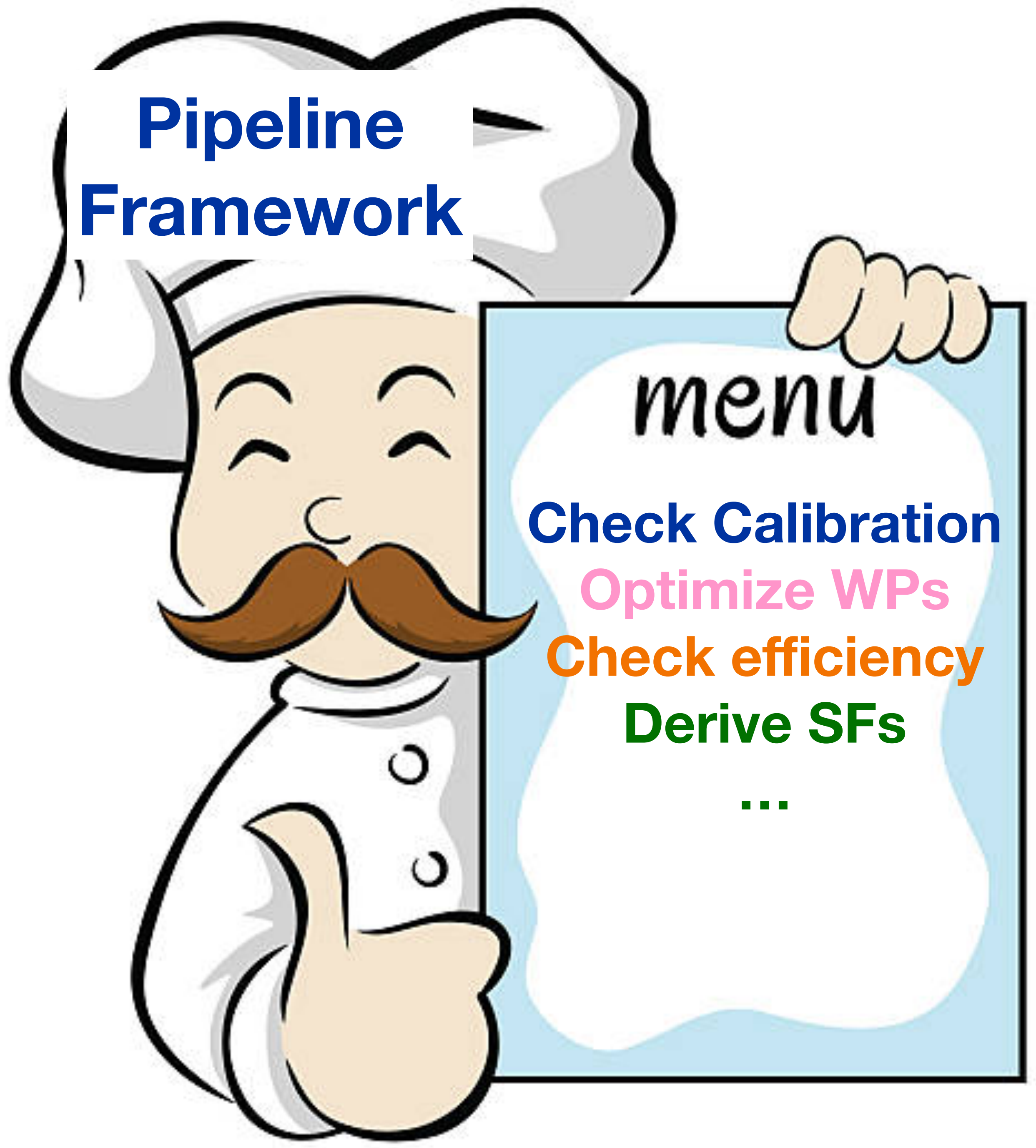
New pipeline framework

- The goal is to **accelerate muon performance checks** via **preservation** and **automation**



Automated trigger: for
Example when a new nTuple
is published at Tier0

On demand user trigger:
for example to **check a new set of recommendations**



Trigger orders a specific "recipe" from the "menu"
Then the pipeline workflow handles interaction with computing grid
All code executed is in docker image
Naturally requires all code to be committed in git before execution!!
Anyone can run *any* code, forever. No compiling, no loss of "personal" code to time

Muon performance analysis toolkit
Packaged into docker images

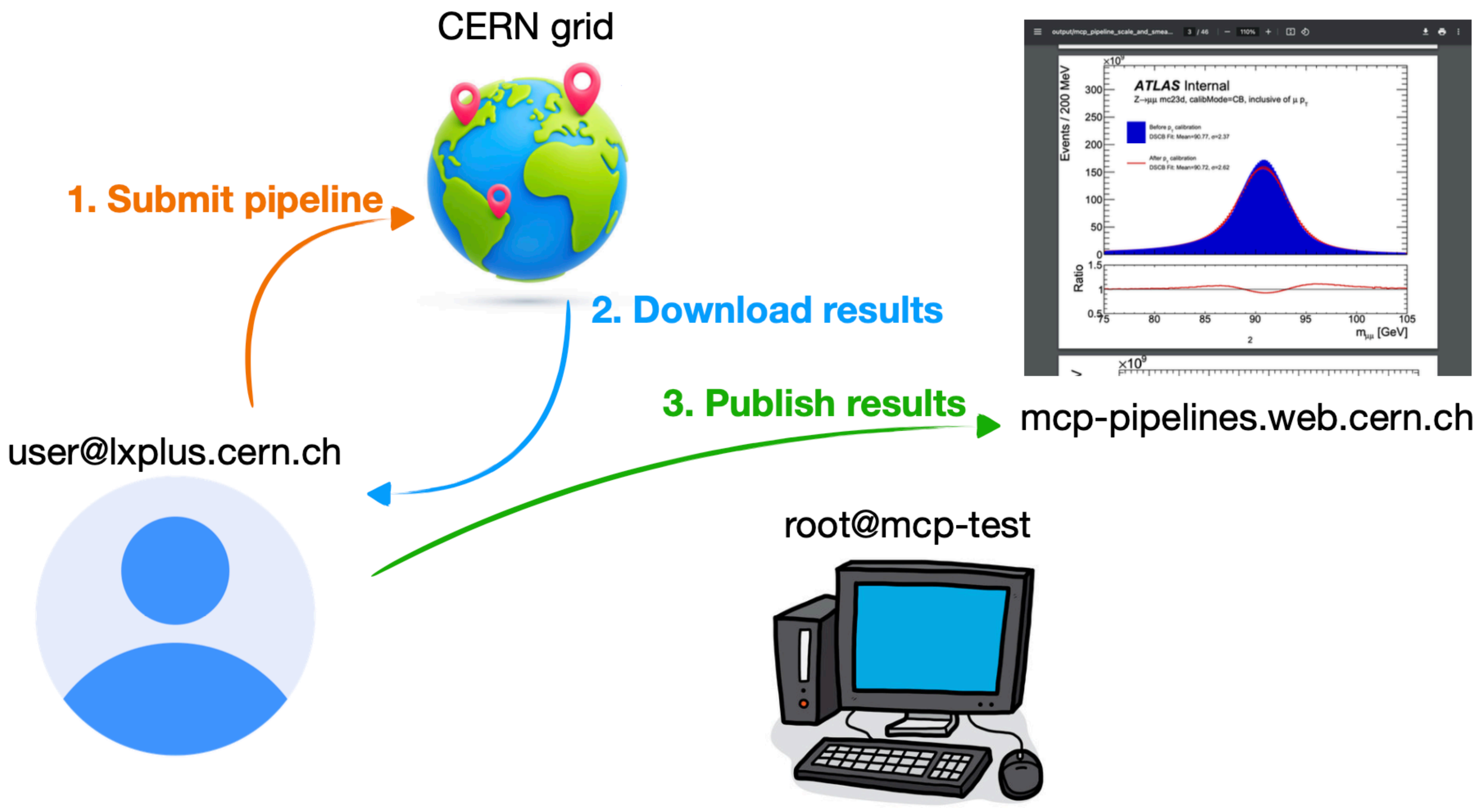
.pdf, .root output files
Published to the web
Can compare previous runs of the same workflow and check consistency

Automation

- **The typical workflow in the pipeline framework**

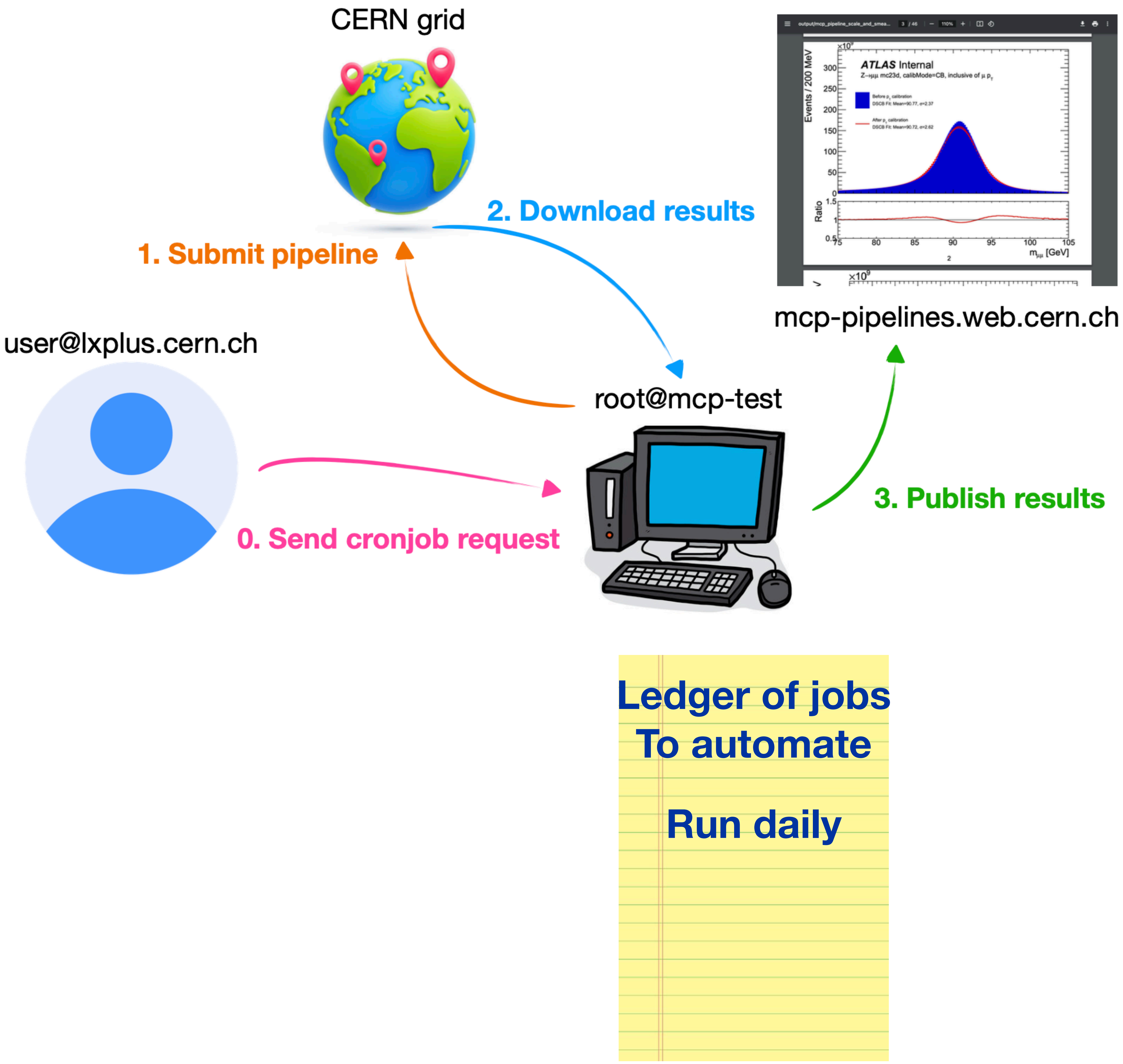
- **A lot is already automatic, as a full workflow is encapsulated into a single interaction with the grid**
  - Ntuple creation, histogram creation, plot creation all happens in a single “chain” of grid jobs.

- **But, what if we could remove the user entirely??**



Automation

- The workflow of an automated job is similar
- The user can modify the jobs.txt file on the remote server at any time with a command line interface
- Whenever new data is taken, nTuples are automatically created at Tier0 computing center
- Previously, we did very little with these nTuples!
- Now, we analyze them automatically and check for changes in muon performance on a regular basis.
- This automated method will be deployed in 2025!



Conclusion

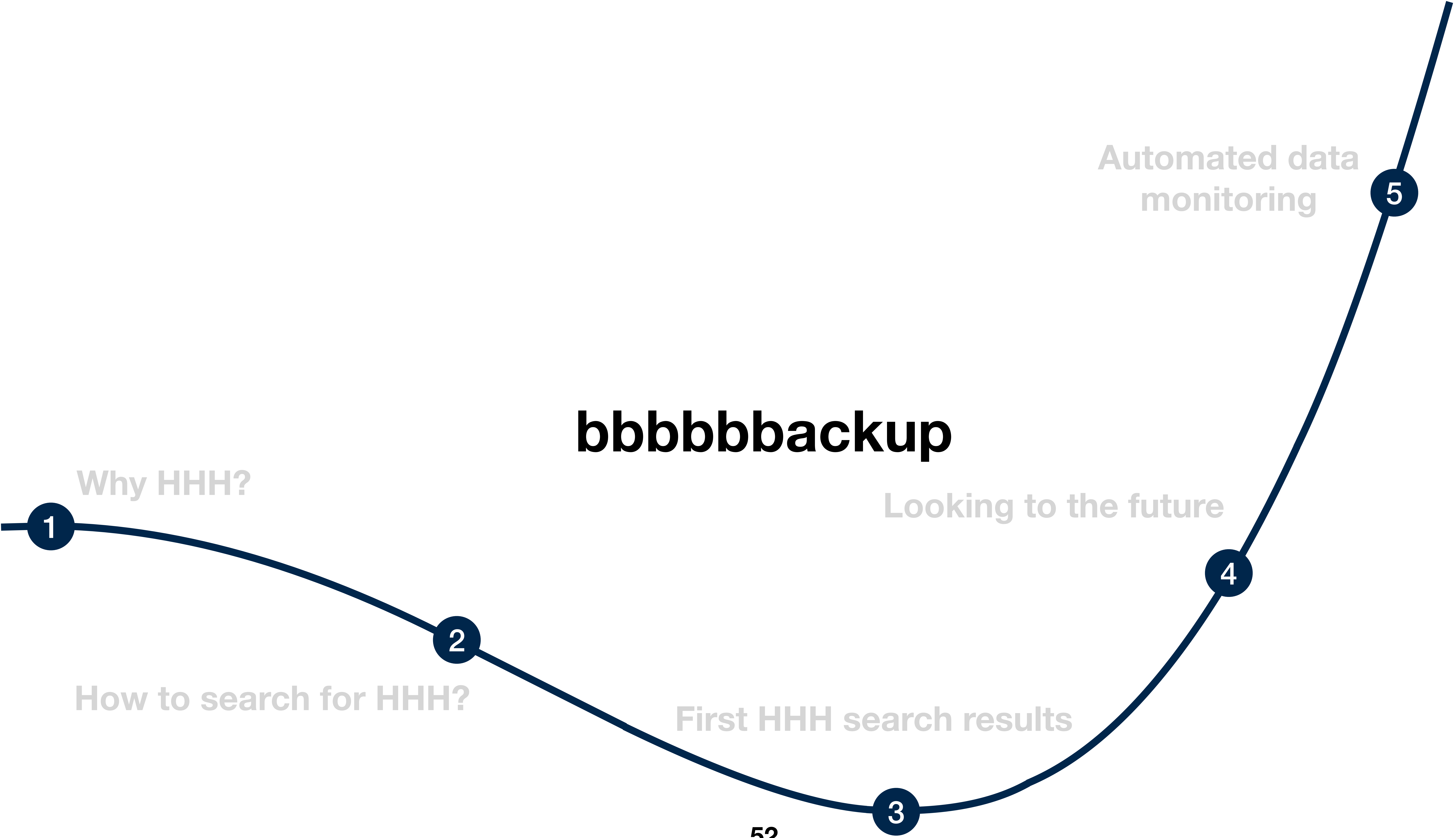
- **The first search for HHH has been performed**
  - Unique sensitivity to κ_4, which is the limiting factor on our knowledge of the Higgs potential shape for large deviations from VeV
  - Relevant *today* because of possible HHH enhancement in models with extra scalars (TRSM, DM-CPV)

- **HHH is a suitable laboratory for AI studies**
  - Statistics limited, simple data-driven method.
  - Applications of modern AI methods can be studied:
    - Domain adaptation
    - Functional decomposition
    - Transformers and object pairing

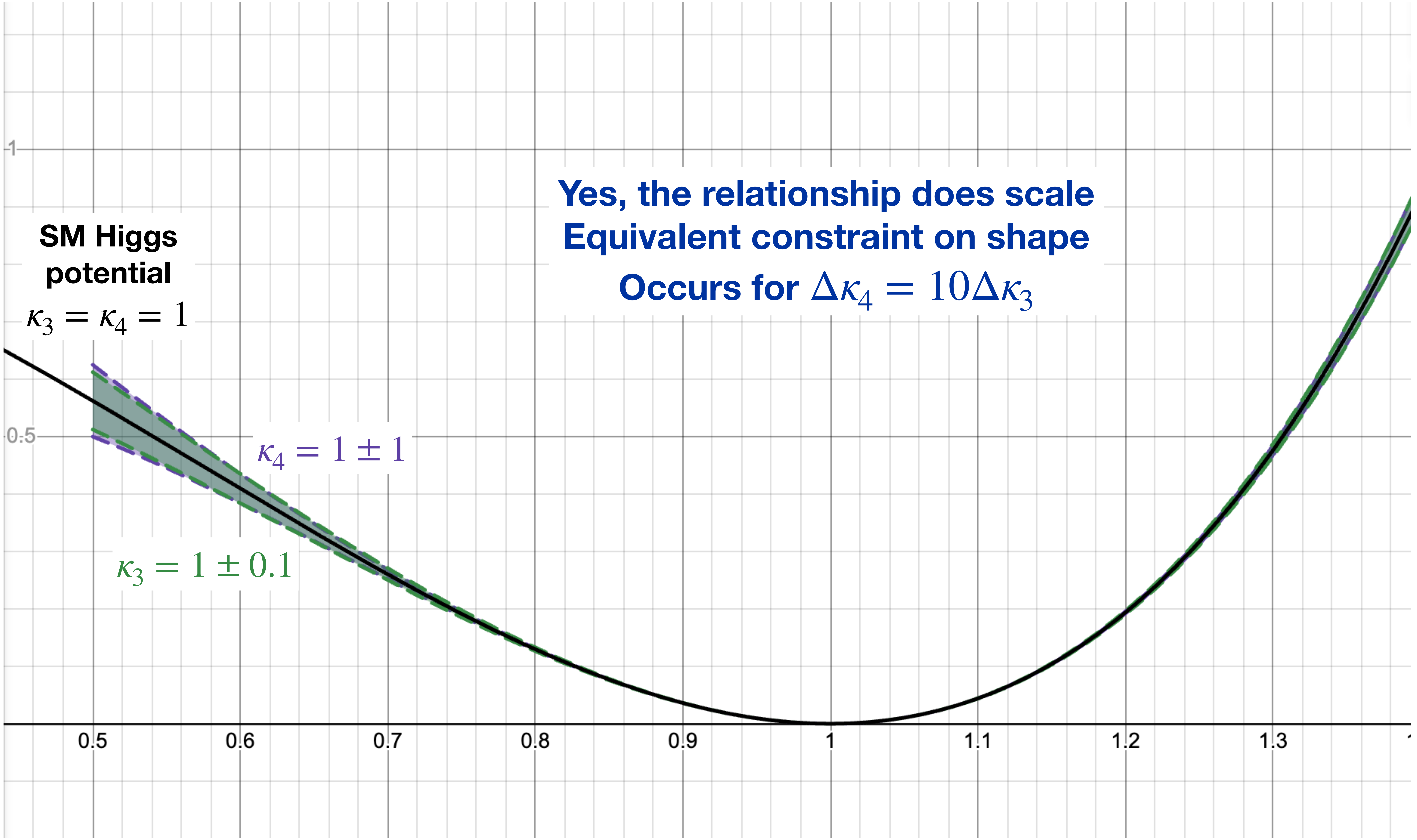
- **Muon performance studies are *accelerated* via *preservation* and *automation* in a new framework**

Google image FX’s interpretation of “discovering new physics”

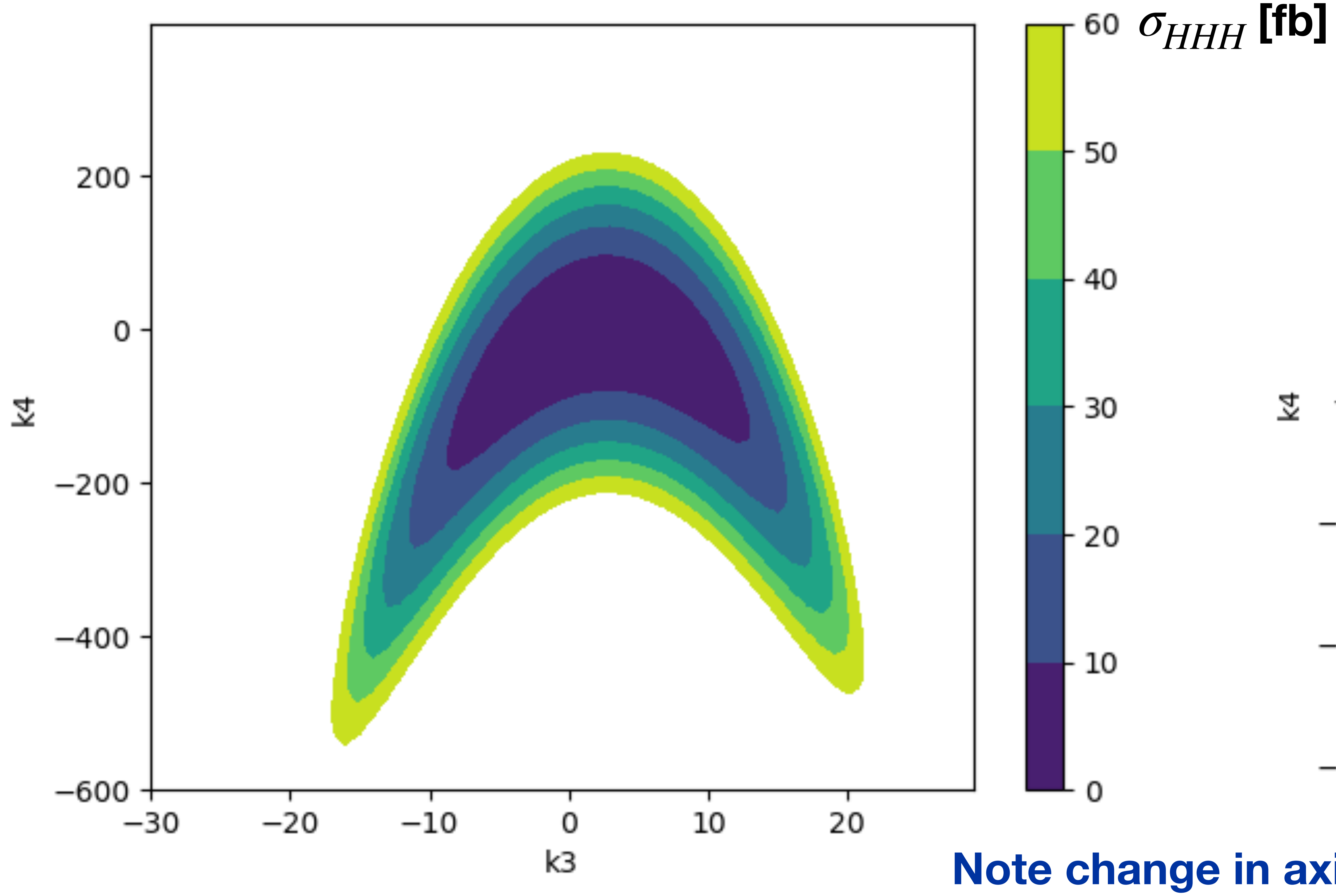
bbbbbbackup

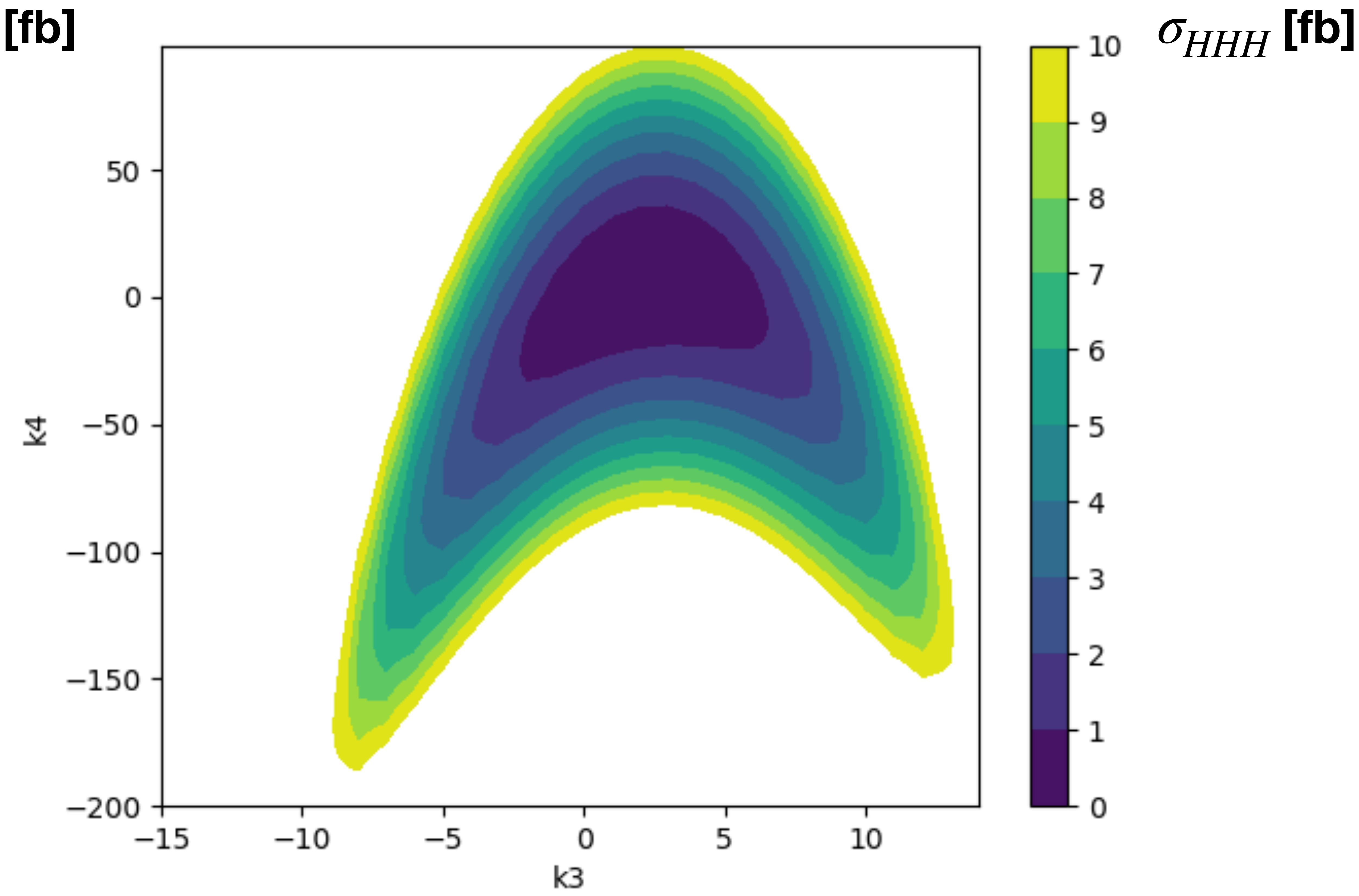


Scaling of κ_3 vs κ_4



HHH production cross section



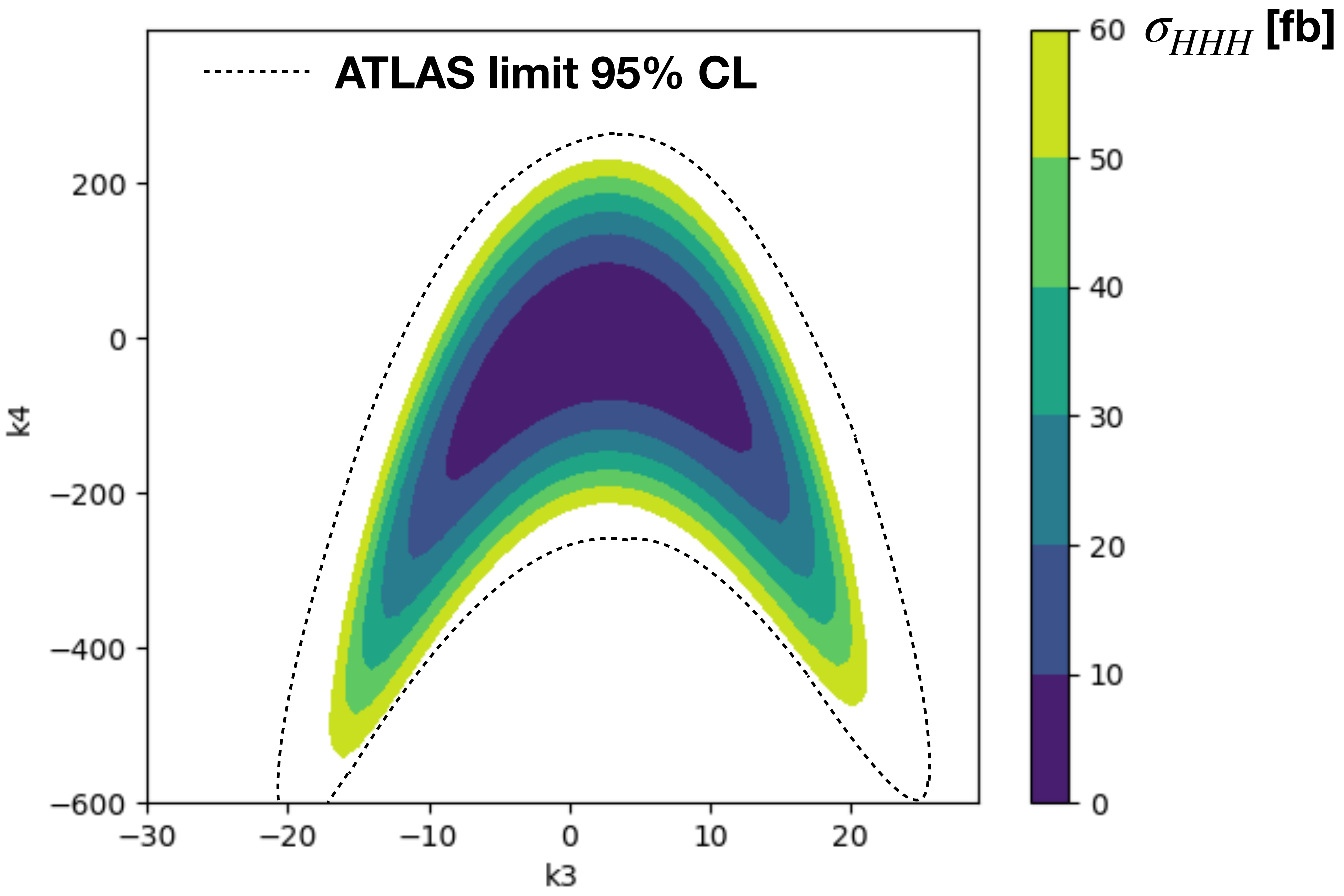


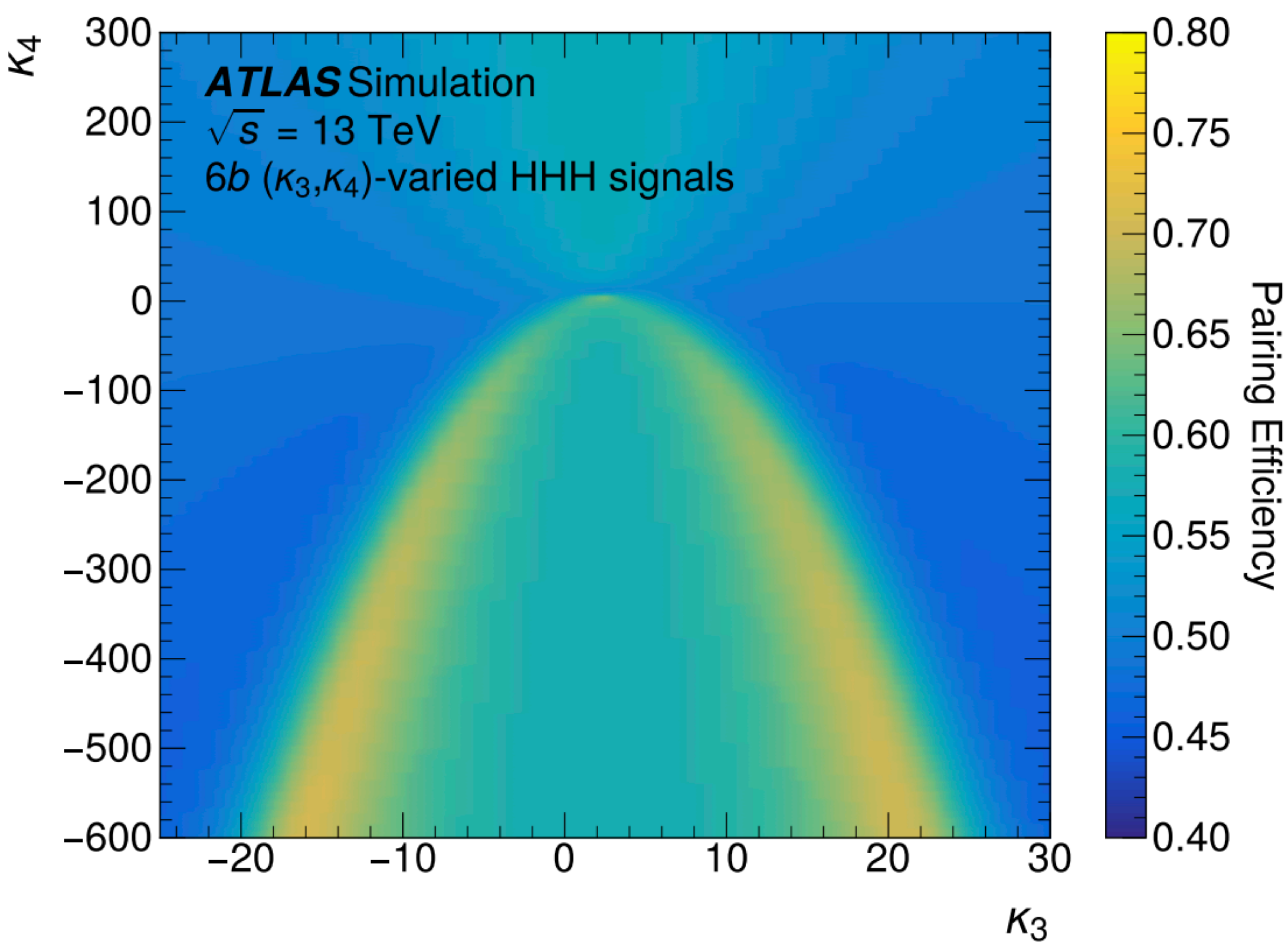
Note change in axis scales

- **Constraint of $\kappa_4 = 1 \pm 30$ roughly translates to a constraint $\sigma_{HHH} \leq$ 1 fb**
- **Luminosity-only scaling for HL-LHC should get us to around $\sigma_{HHH} \leq$ 10 fb**

Kevin Nelson

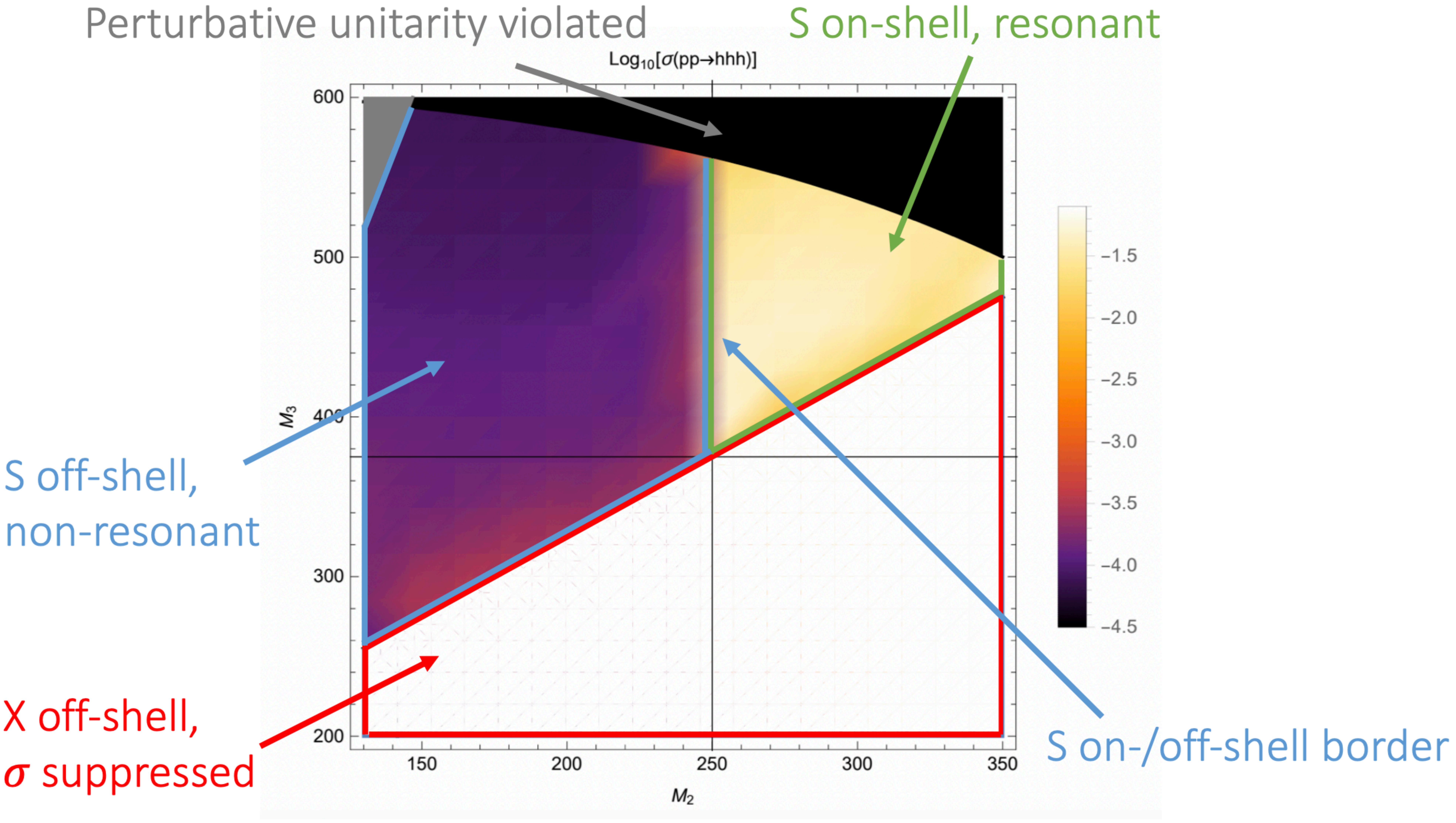
Shape of ATLAS limits



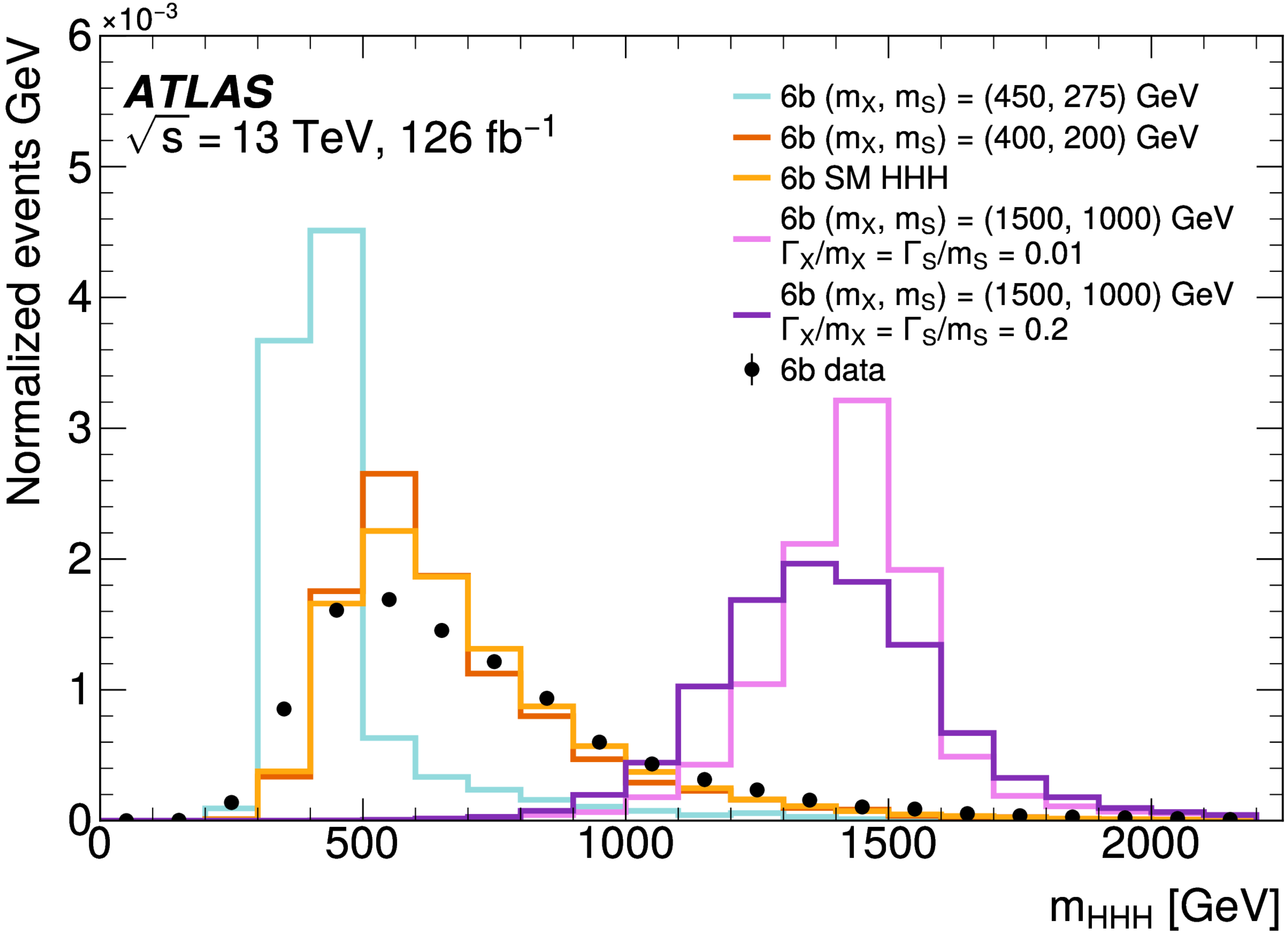


- **ATLAS limit inherits its shape from the contours of constant σ_{HHH}**
- **Slight modulations in shape come from how boosted the Higgses are, and therefore how well we pair the jets.**
- **In regions of higher pairing efficiency, we set a more strict limit**

TRSM perturbative unitarity



Signal kinematics



Trigger strategy

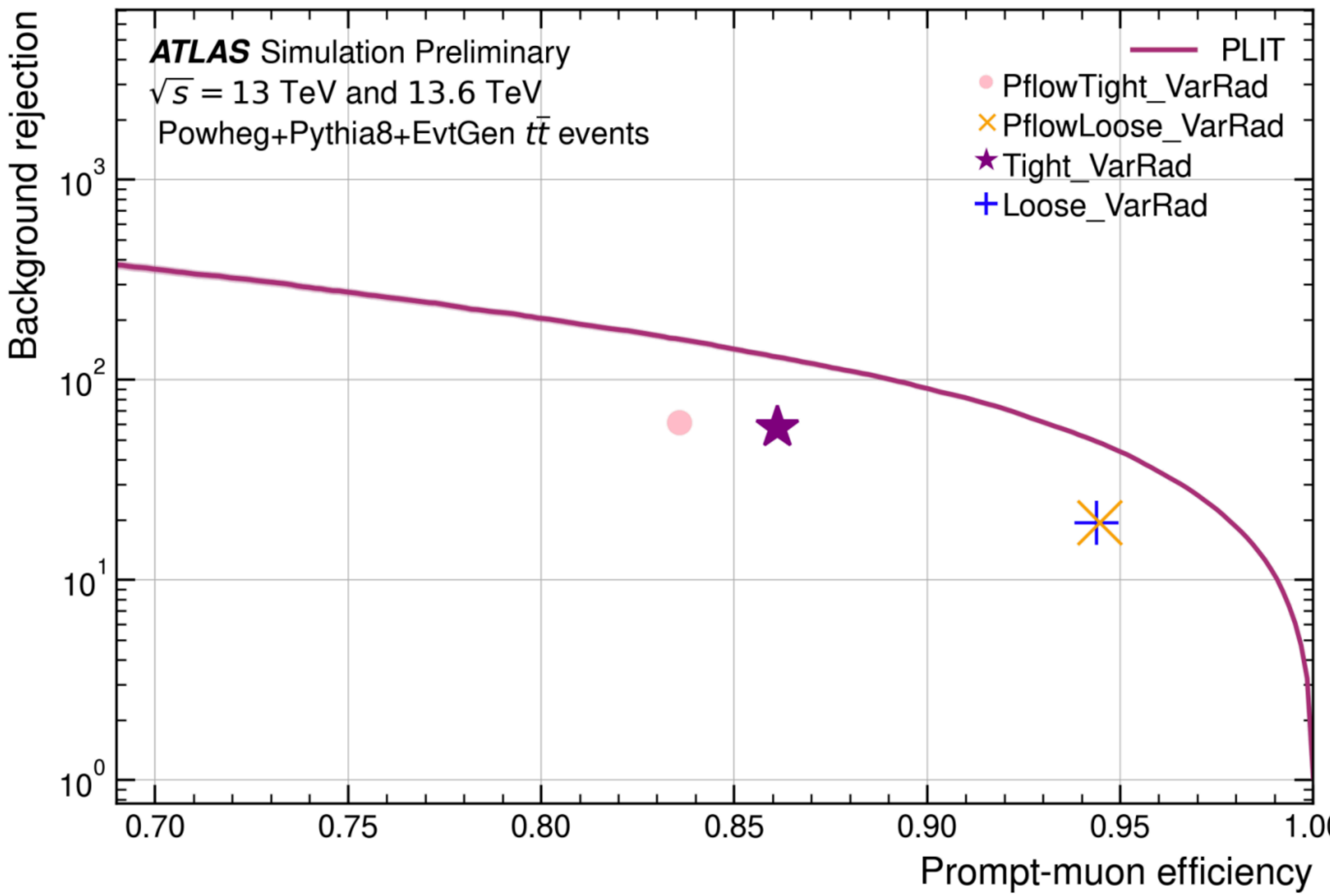
- Using multi-*b*-jet triggers for a 6b final state – similar to the resolved HH4b analysis
  - Omitting 2015 data due to lack of jet trigger-matching information
- 2b2j triggers in 2016, 2018
- 3b1j triggers in 2017
  - Due to very tight online *b*-tagging working point in 2017 2b2j triggers

Table 5.1: List of *b*-jet triggers selected for efficiency study.

Year	Type	Trigger Name
2016	2b+2j	HLT_2j35_bmv2c2060_split_2j35_L14J15.0ETA25
2017	3b+1j	HLT_3j15_gsc35_bmv2c1070_split_j15_gsc35_boffperf_split_L14J15.0ETA25
2018	2b+2j	HLT_2j35_bmv2c1060_split_2j35_L14J15.0ETA25

Other applications of transformers

- **Factor of 2 improvement in lepton isolation from transformer architecture**



Differentiable p-values with FD

- **Transformation from physical to frequency domain is fast:**
  - Recursion relationships for the orthonormal functions
  - Use unbinned data: inner product with Dirac delta functions is simple to compute

- **In the frequency domain background and signal are separable by a cut**