

Data Preservation and Reinterpretation

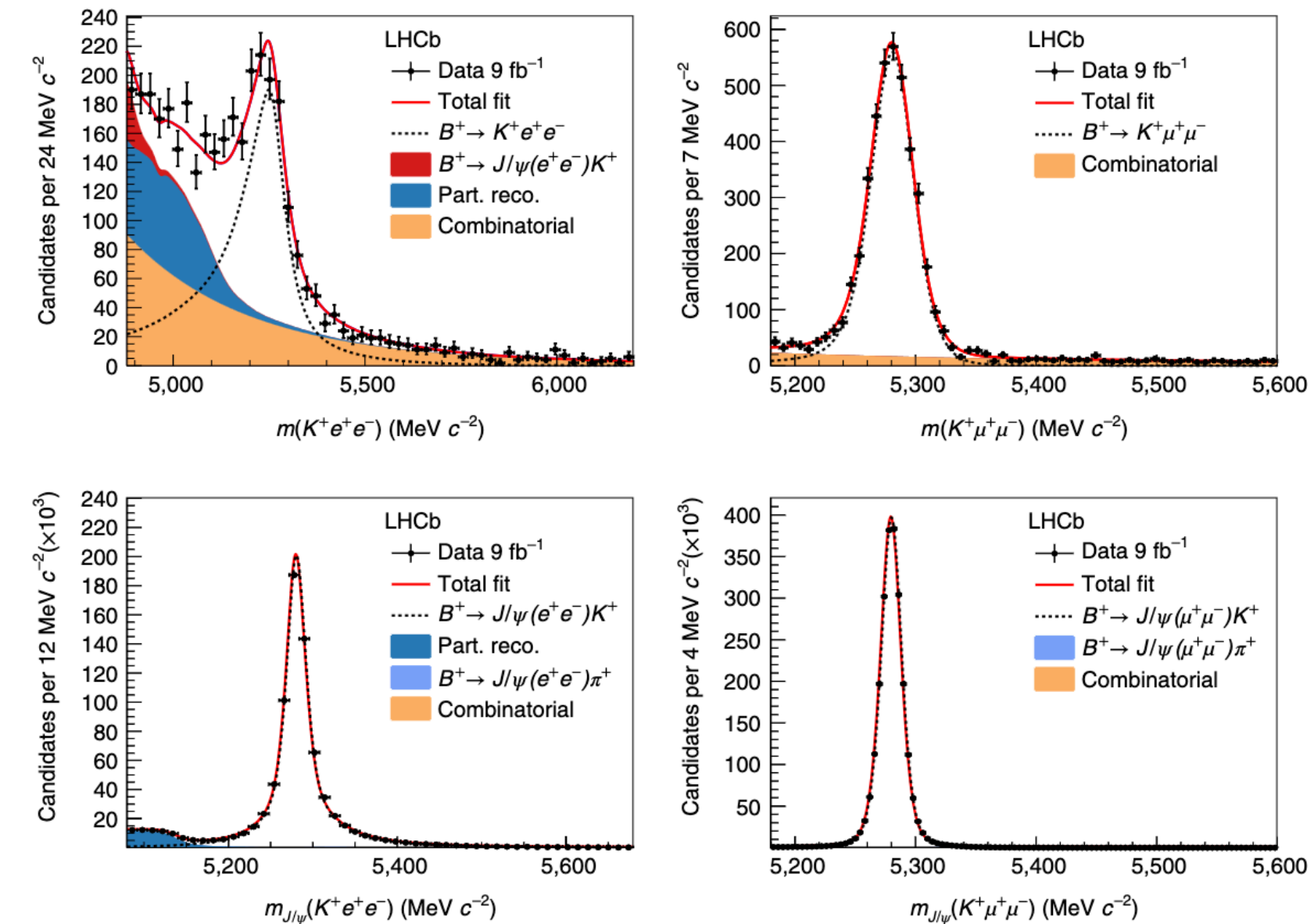
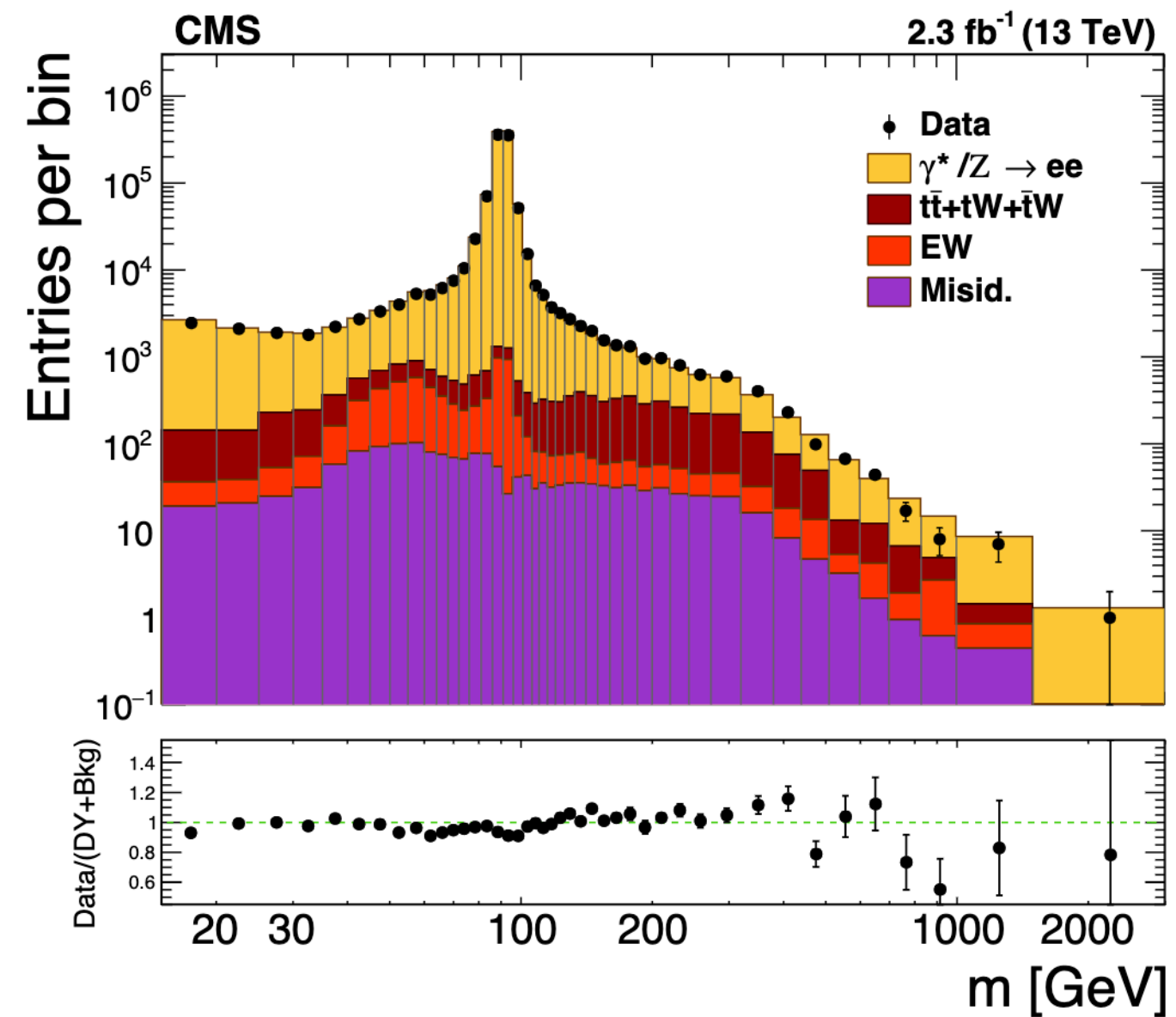
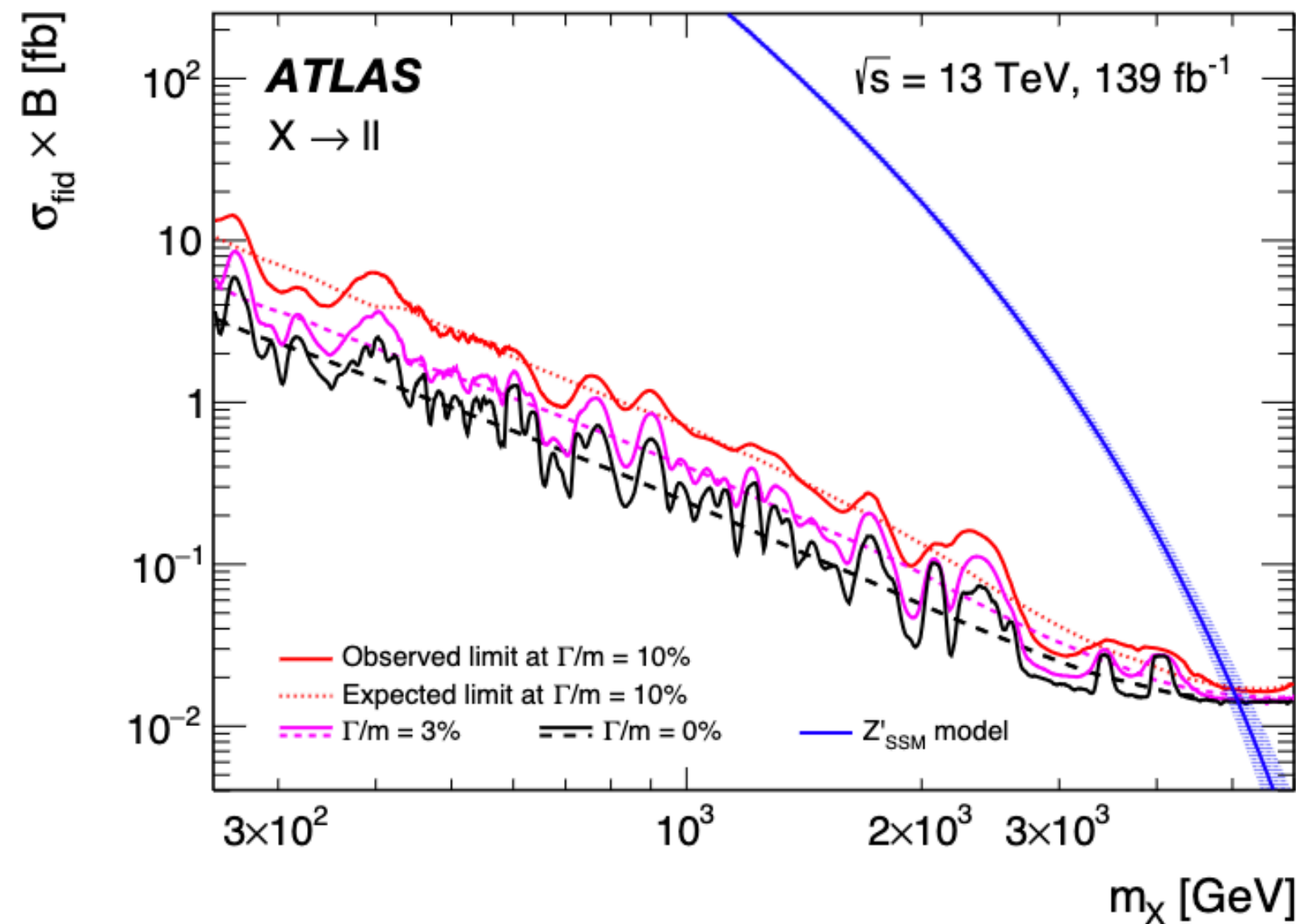
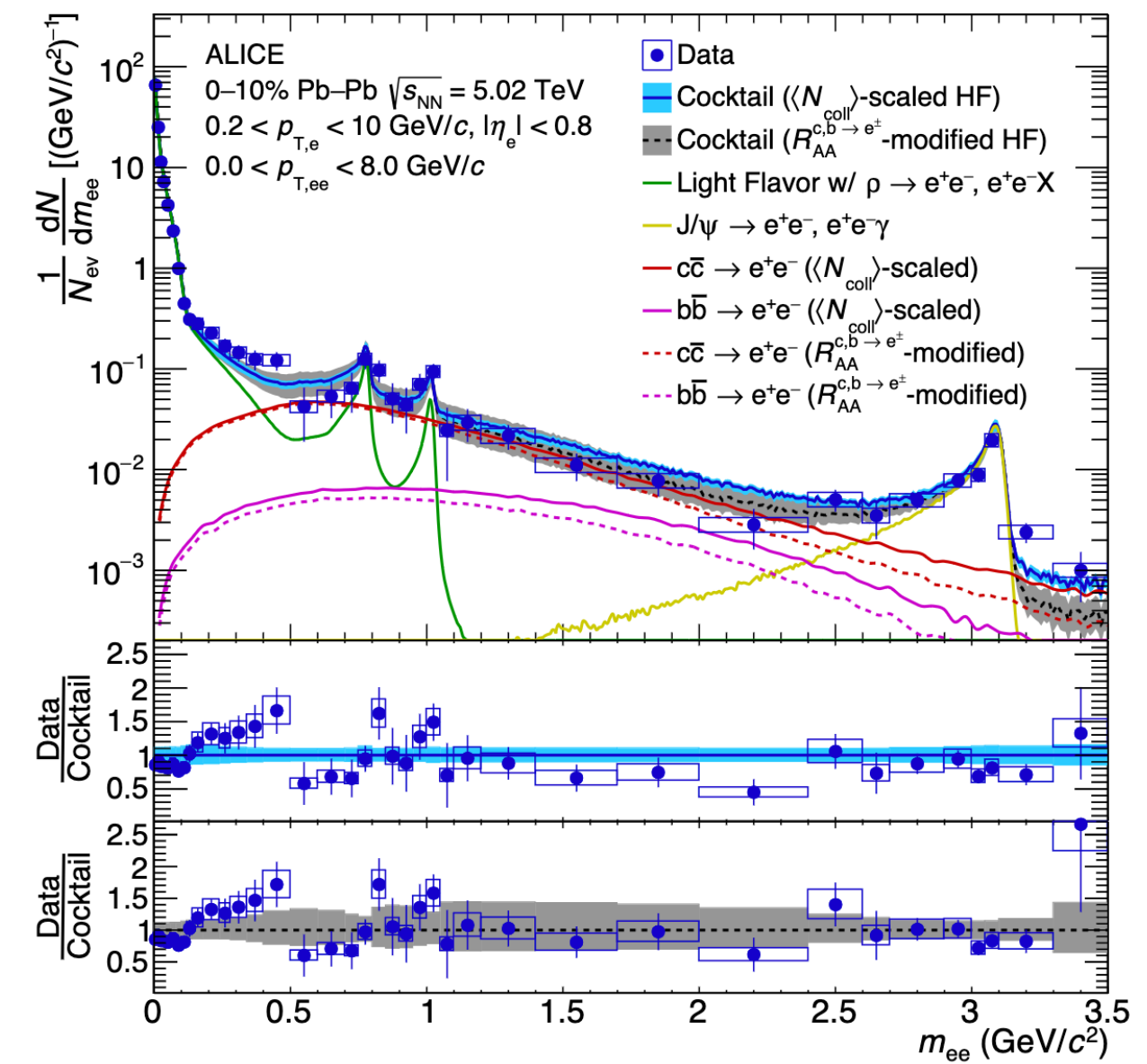
LHCP2024
June 3-7, 2024

Si Hyun Jeon on behalf of LHC collaboration



Interpretations from the LHC

- Can't spell "re" interpretation without interpretation
- Various interpretations, e.g. dilepton final states
 - Z' search, DY cross-section, lepton universality, ...
- Impossible to cover all interpretations carried out at the LHC → Today "BSM reinterpretation"



Preservations at the LHC : HEPData

- HEPData : Electronically tabularizing histograms/distributions
- Additional information relevant to the analysis (signal MC sample configuration, statistical models, ...)

HEPData [About](#) [Submission Help](#) [File Formats](#) [Sign in](#)

Hayrapetyan, Aram et al. Last updated on 2024-05-16 18:52 [URL](#) Accessed 74 times

Hide Publication Information

Search for new physics in high-mass diphoton events from proton-proton collisions at $\sqrt{s} = 13$ TeV

The CMS collaboration

Hayrapetyan, Aram , Tumasyan, Armen , Adam, Wolfgang , Andrejkovic, Janik Walter , Bergauer, Thomas , Chatterjee, Suman , Damanakis, Konstantinos , Dragicevic, Marko , Hussain, Priya Sajid , Jeitler, Manfred

CMS-EXO-22-024, 2024.

<https://doi.org/10.17182/hepdata.150677>

Abstract

Results are presented from a search for new physics in high-mass diphoton events from proton-proton collisions at $\sqrt{s} = 13$ TeV. The data set was collected in 2016-2018 with the CMS detector at the LHC and corresponds to an integrated luminosity of 138 fb^{-1} . Events with a diphoton invariant mass greater than 500 GeV are considered. Two different techniques are used to predict the standard model backgrounds: parametric fits to the smoothly-falling background and a first-principles calculation of the standard model diphoton spectrum at next-to-next-to-leading order in perturbative quantum chromodynamics calculations. The first technique is

Filter 16 data tables

Data from Figure 4 (bottom right) located on page 10.

[10.17182/hepdata.150677.v1/t13](https://doi.org/10.17182/hepdata.150677.v1/t13)

Figure 4 bottom right. Observed 95% CL upper limits on the product of the cross section and branching fraction as...

Fig. 6 Mgg EBEB >

Data from Figure 6 located on page 11.

[10.17182/hepdata.150677.v1/t14](https://doi.org/10.17182/hepdata.150677.v1/t14)

Figure 6. The $m_{\gamma\gamma}$ spectra and background prediction after nuisance parameter marginalization (post-fit) due to SM diphoton production ($\gamma\gamma$) and...

Fig. 6 Mgg EBEE >

Data from Figure 6 located on page 11.

[10.17182/hepdata.150677.v1/t15](https://doi.org/10.17182/hepdata.150677.v1/t15)

Figure 6. The $m_{\gamma\gamma}$ spectra and background prediction after nuisance parameter marginalization (post-fit) due to SM diphoton production ($\gamma\gamma$) and...

Fig. 7 CWk limits >

Data from Figure 7 located on page 12.

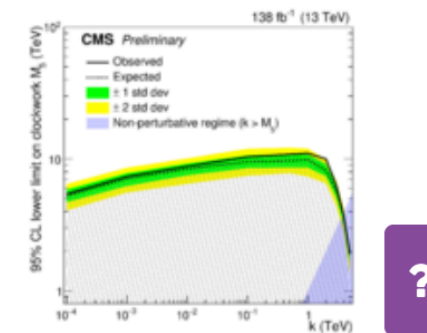
[10.17182/hepdata.150677.v1/t16](https://doi.org/10.17182/hepdata.150677.v1/t16)

Figure 7. The exclusion limit for the

Fig. 7 CWk limits [10.17182/hepdata.150677.v1/t16](https://doi.org/10.17182/hepdata.150677.v1/t16)

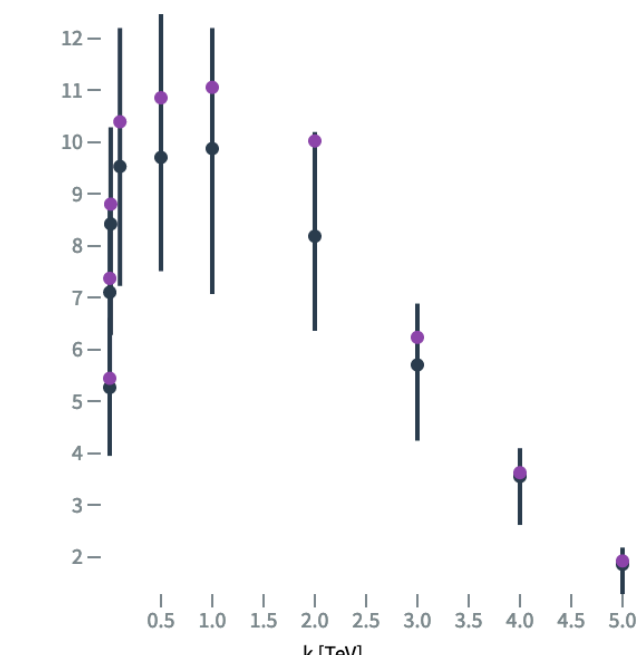
Data from Figure 7 located on page 12.

Figure 7. The exclusion limit for the clockwork framework over the k - M_5 parameter space. The shaded region denotes where the theory becomes nonperturbative. The region below and to the left of the solid line constitutes the excluded region. Expected 1std and 2std limit bands are shown in green and yellow, respectively.



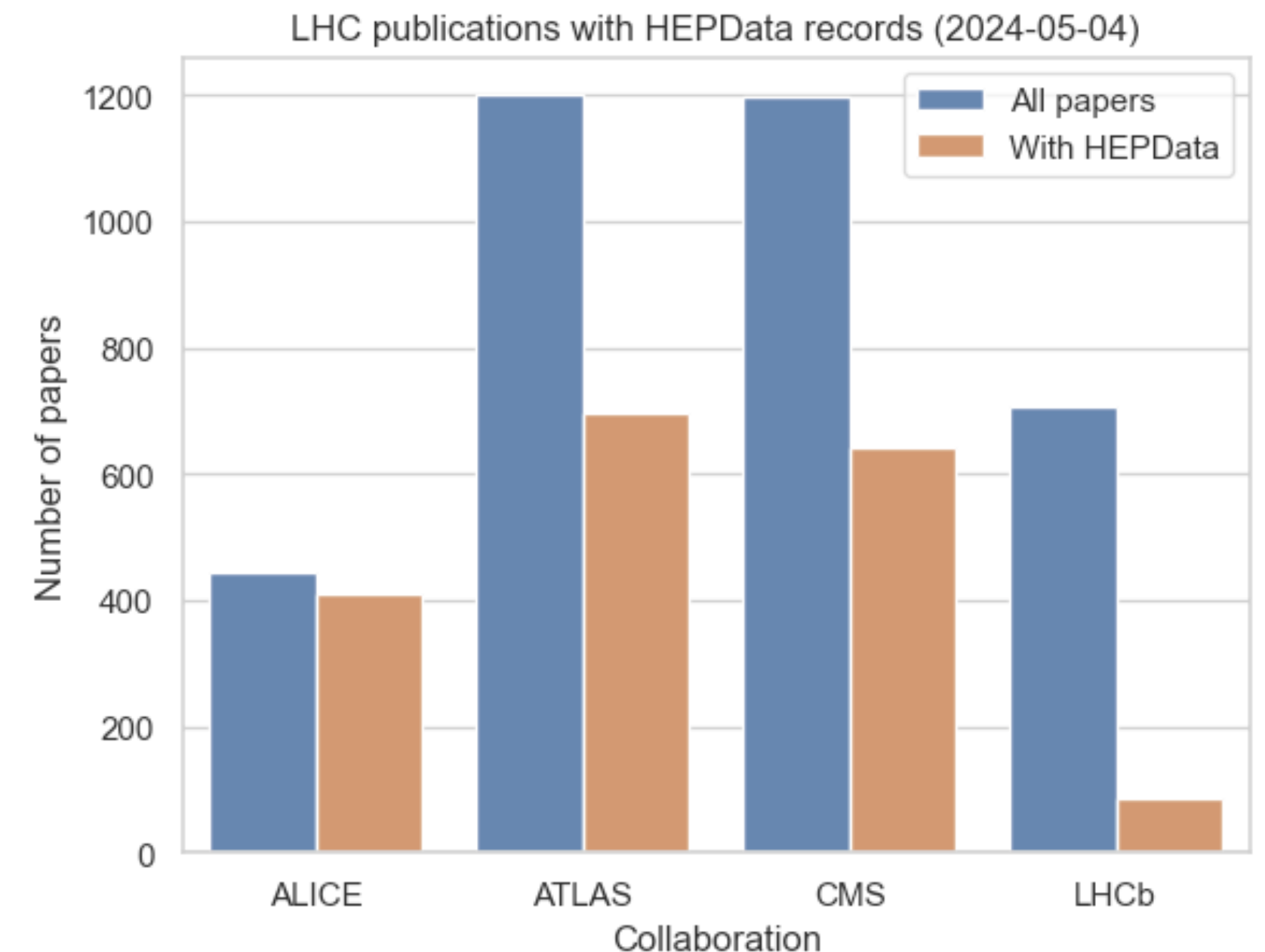
Limit	Observed	Expected
k [TeV]	Lower limit on M_5 [TeV]	
0.0001	5.4438	5.2668 $+0.6125$ -0.55589 $\pm 1\sigma$ $+1.1903$ -1.1947 $\pm 2\sigma$
0.001	7.3723	7.1034 $+0.69372$ -0.85592 $\pm 1\sigma$ $+1.4774$ -1.5381 $\pm 2\sigma$
0.01	8.8034	8.421 $+0.81854$ -0.89714 $\pm 1\sigma$ $+1.6778$ -1.944 $\pm 2\sigma$
0.1	10.392	9.5296 $+1.112$ -0.99591 $\pm 1\sigma$ $+2.4245$ -2.0757 $\pm 2\sigma$
0.5	10.854	9.7041 $+1.2405$ -1.0751 $\pm 1\sigma$ $+2.4702$ -1.9104 $\pm 2\sigma$
1.0	11.055	9.8742 $+0.96313$ -1.2569 $\pm 1\sigma$ $+2.1151$ -2.5087 $\pm 2\sigma$
2.0	10.022	8.1846 $+0.97421$ -0.78477 $\pm 1\sigma$ $+1.7554$ -1.6476 $\pm 2\sigma$

Visualize



HEPData [link](#)

Repository for publication-related High-Energy Physics data



Preservations at the LHC : Rivet

- Rivet : One of the most widely used tools to preserve and recast unfolded measurements at the LHC
- Agnostic to both MC generator and detector : Validation tool

Rivet



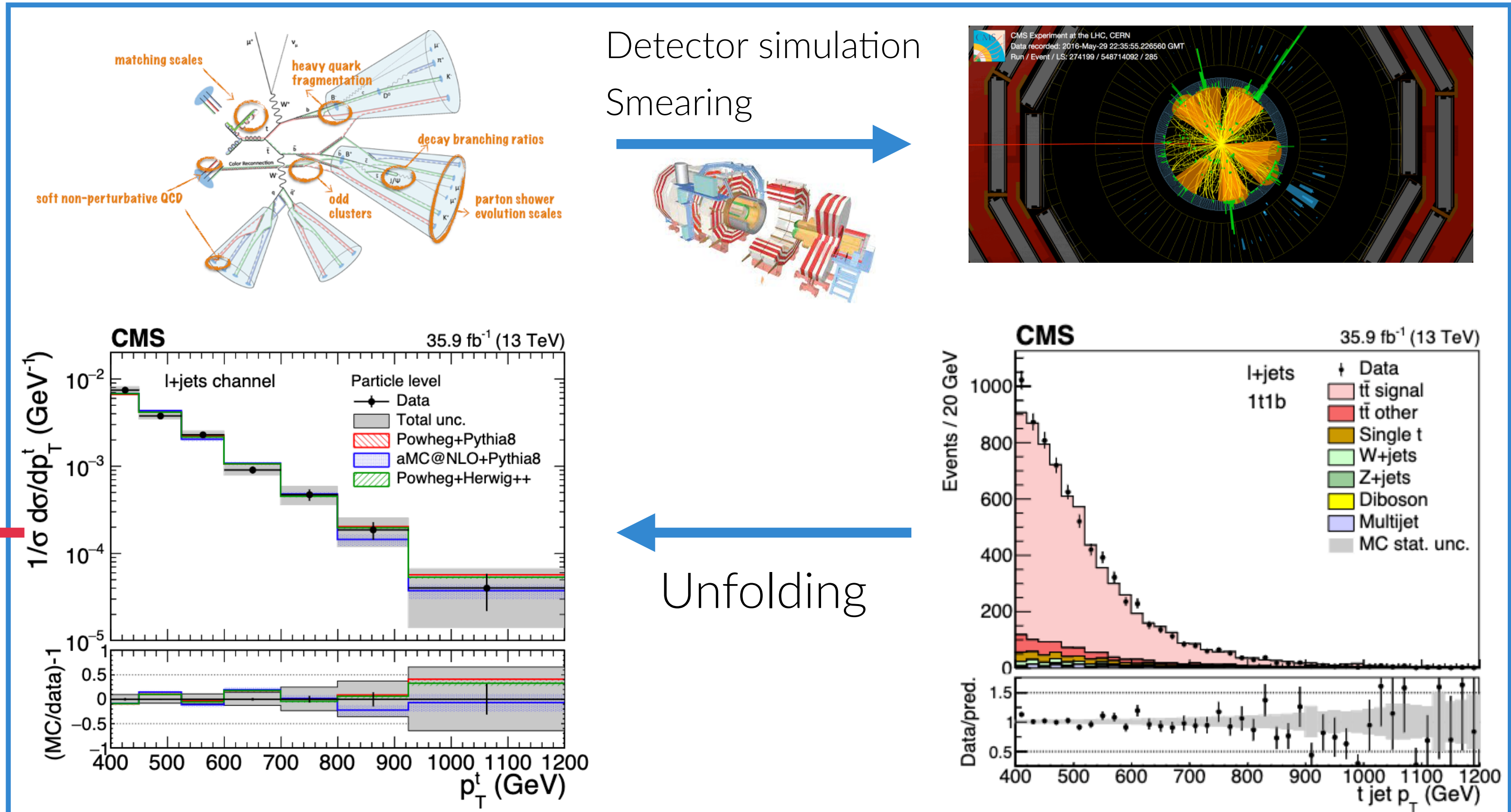
[link](#)

Key	ALICE	ATLAS	CMS	LHCb	Forward	HERA	$e^+e^- (\geq 12 \text{ GeV})$	$e^+e^- (\leq 12 \text{ GeV})$	Tevatron	RHIC	SPS	Other
Rivet wanted (total):	327	401	504	198	16	475	660	364	1118	509	57	97
Rivet REALLY wanted:	55	50	99	16	0	15	1	1	9	2	5	1
Rivet provided:	32/359 = 9%	202/603 = 33%	114/618 = 18%	31/229 = 14%	9/25 = 36%	35/510 = 7%	224/884 = 25%	692/1056 = 66%	59/1177 = 5%	9/518 = 2%	5/62 = 8%	32/129 = 25%

```

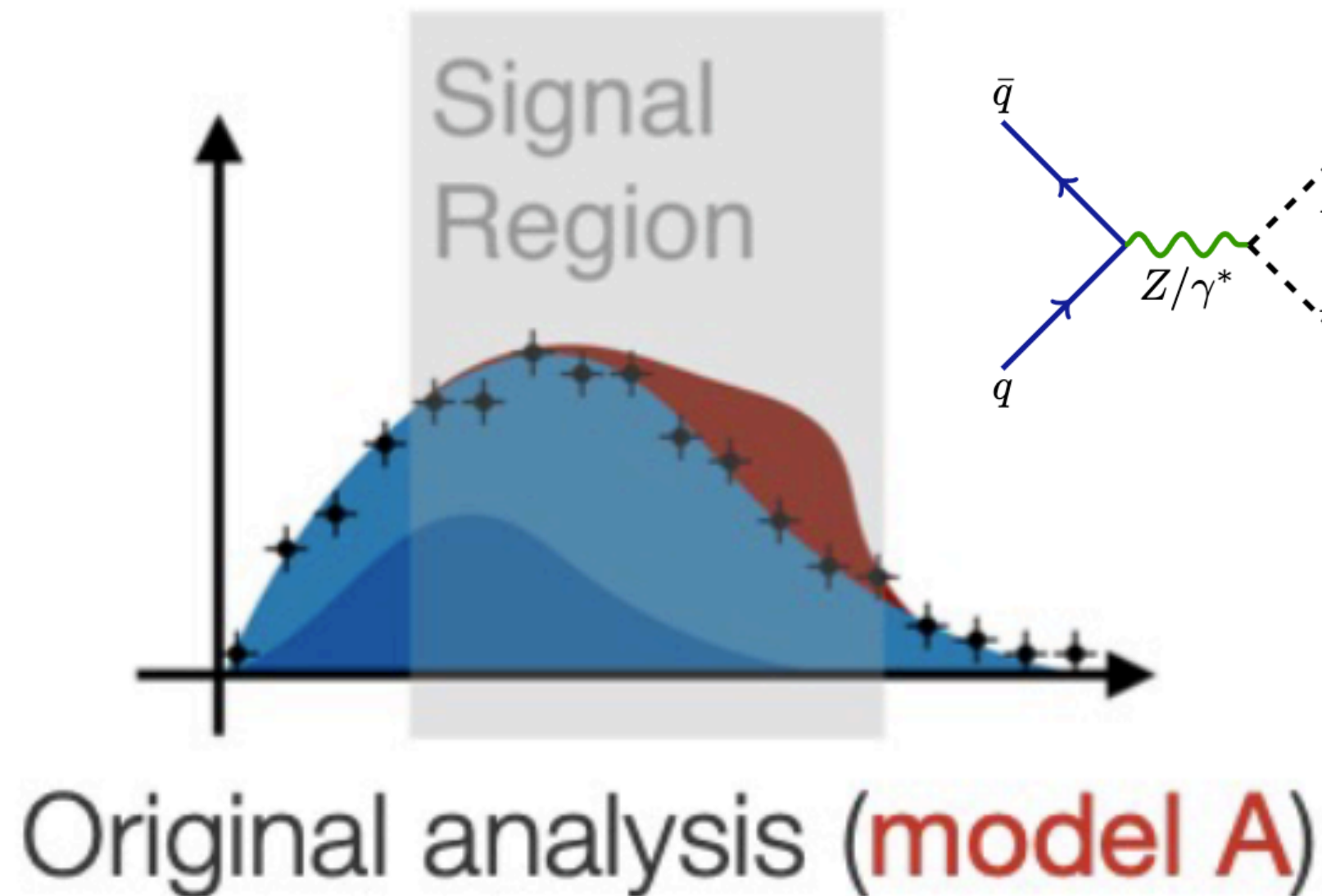
47 void analyze(const Event& event) {
48
49     const Jets jets = apply<JetFinder>(event, "Jets").jetsByPt(Cuts::pT > 300*GeV && Cuts::abseta < 2.1);
50
51     if (jets.size() < 2) vetoEvent;
52     if (jets[0].pT() < 675*GeV) vetoEvent;
53
54     if ( (jets[0].pT()/jets[1].pT()) > 1.5 ) vetoEvent;
55
56     _njets->fill(2);
57
58     const Particles& tracks = apply<ChargedFinalState>(event, "tracks").particlesByPt();
59
60     Particles intracks1;
61     Particles intracks2;
62
63     const Jet& j1 = jets[0];
64     const Jet& j2 = jets[1];
65
66
67     for (const Particle& p : tracks) {
68         const double dr = deltaR(j1, p, PSEUDORAPIDITY);
69         if (dr > 0.4) continue;
70         if (abs(p.pid()) == 13) continue;
71         intracks1.push_back(p);
72     }
73

```

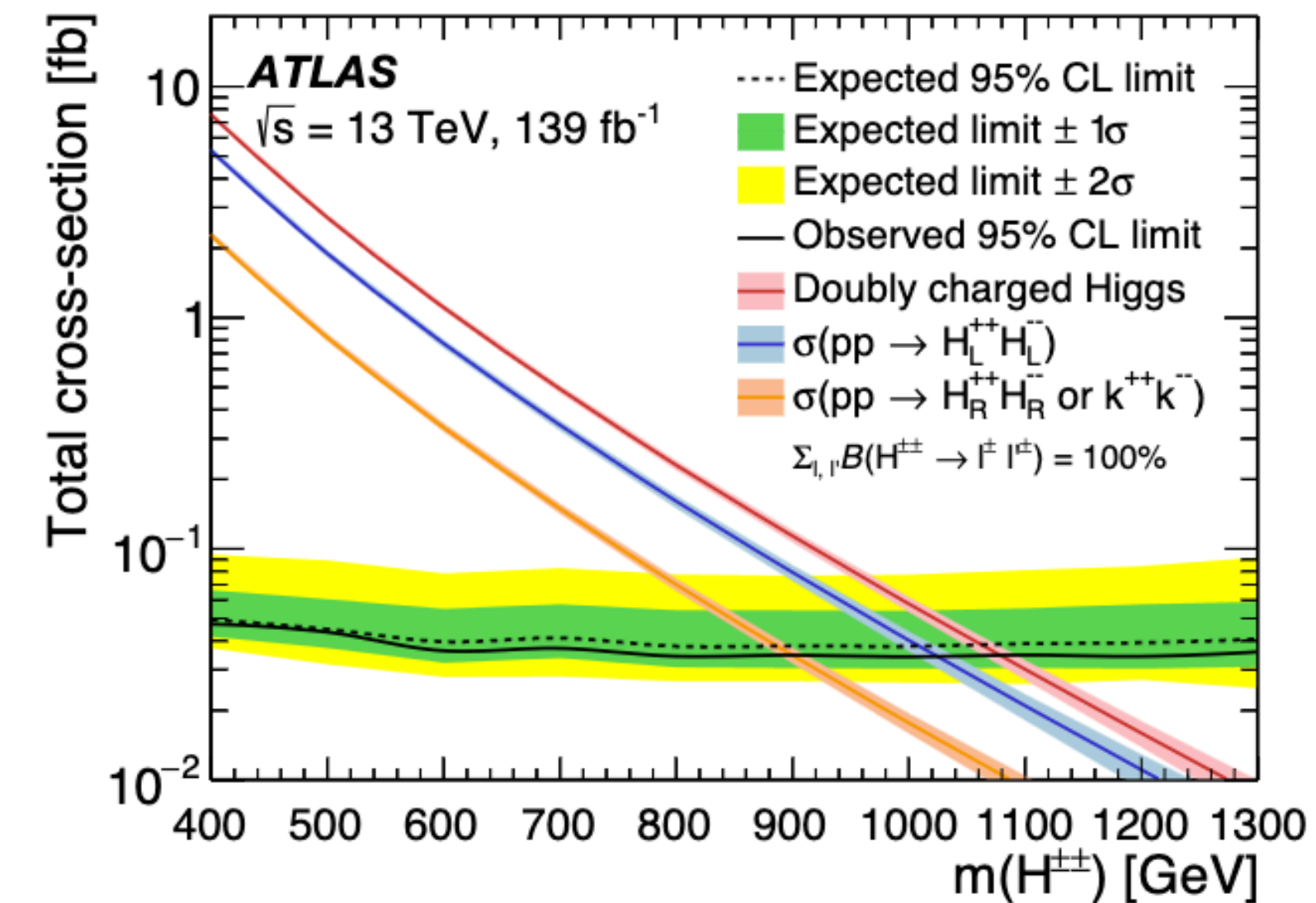
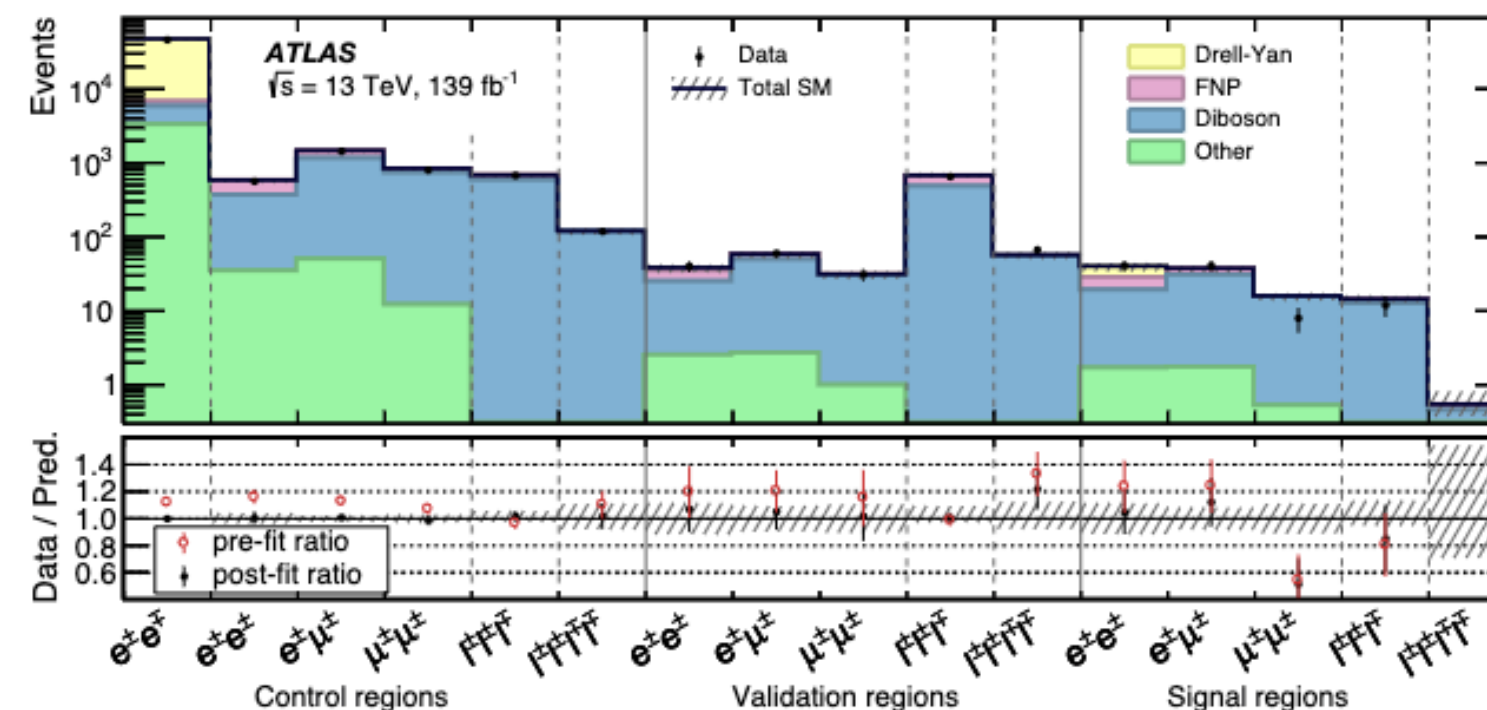


Reinterpretations of the LHC

- BSM interpretations at the LHC mostly consider handful of models
 - Test data against the SM predictions in various channels
 - Try to look for unexplained deviations at a good confidence level
- **Model A** excluded by testing data against **SM predictions**
 - e.g. $m(H^{++})$ 400-1000GeV excluded assuming $BR(H^{++} \rightarrow \ell\ell) = 100\%$

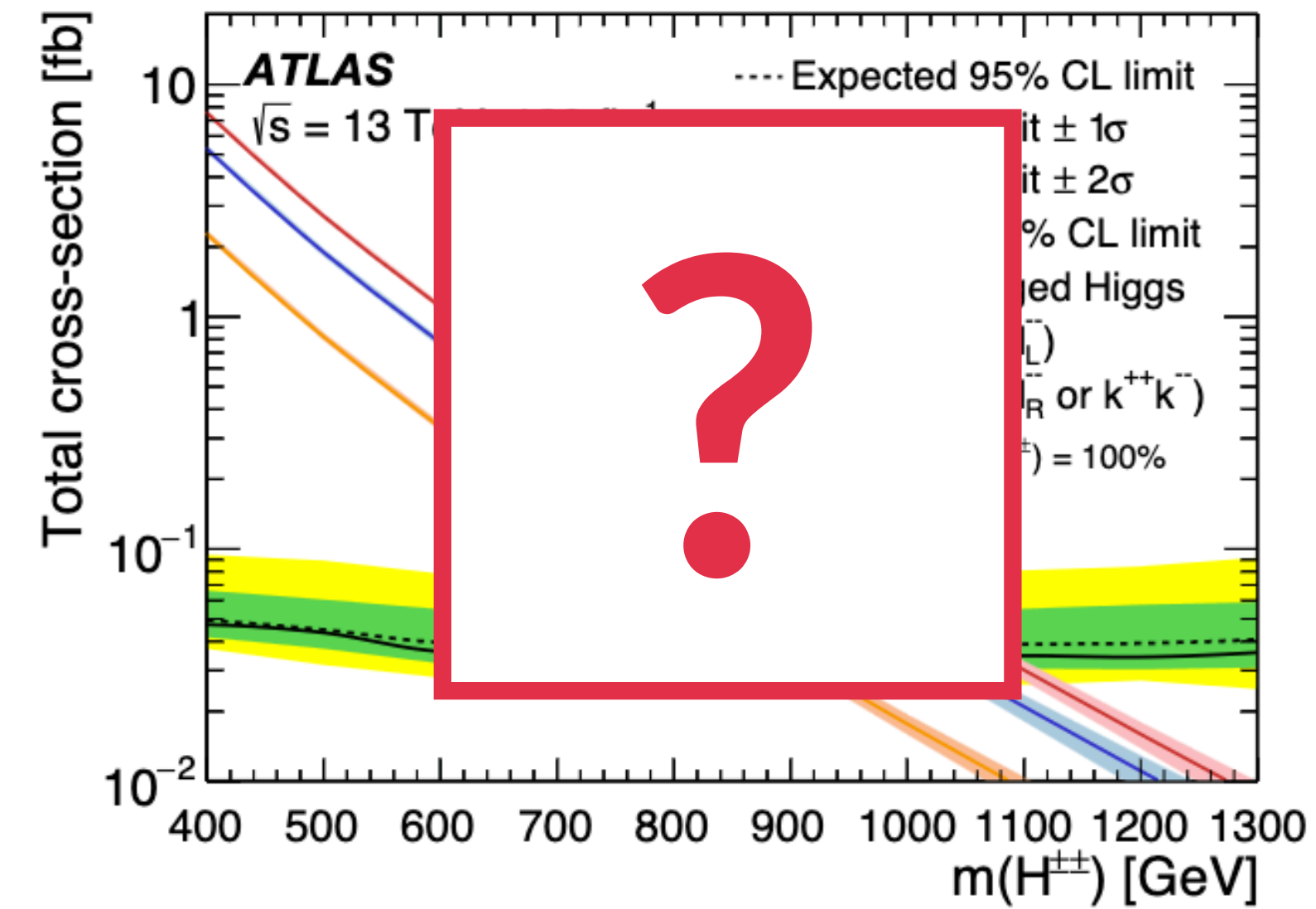
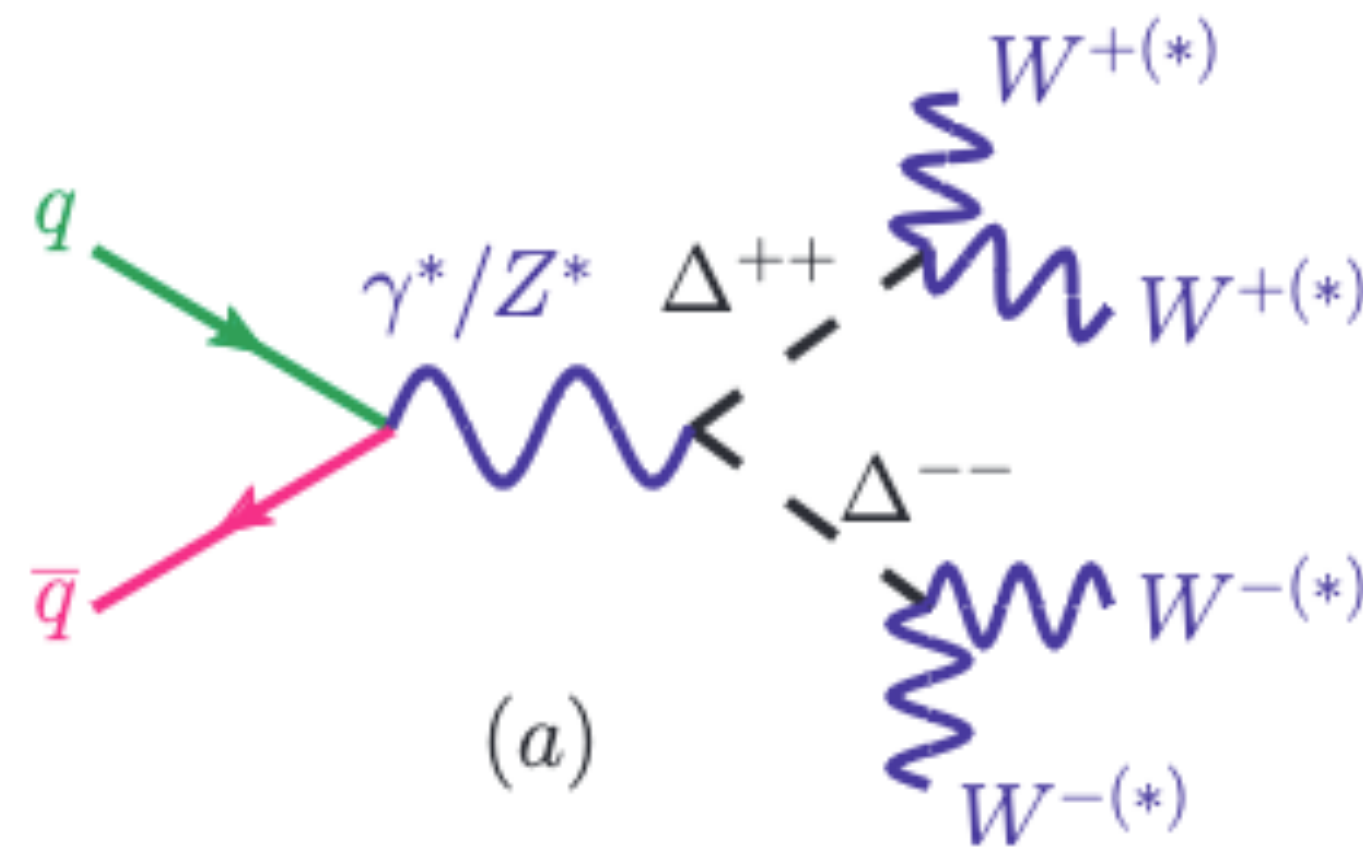
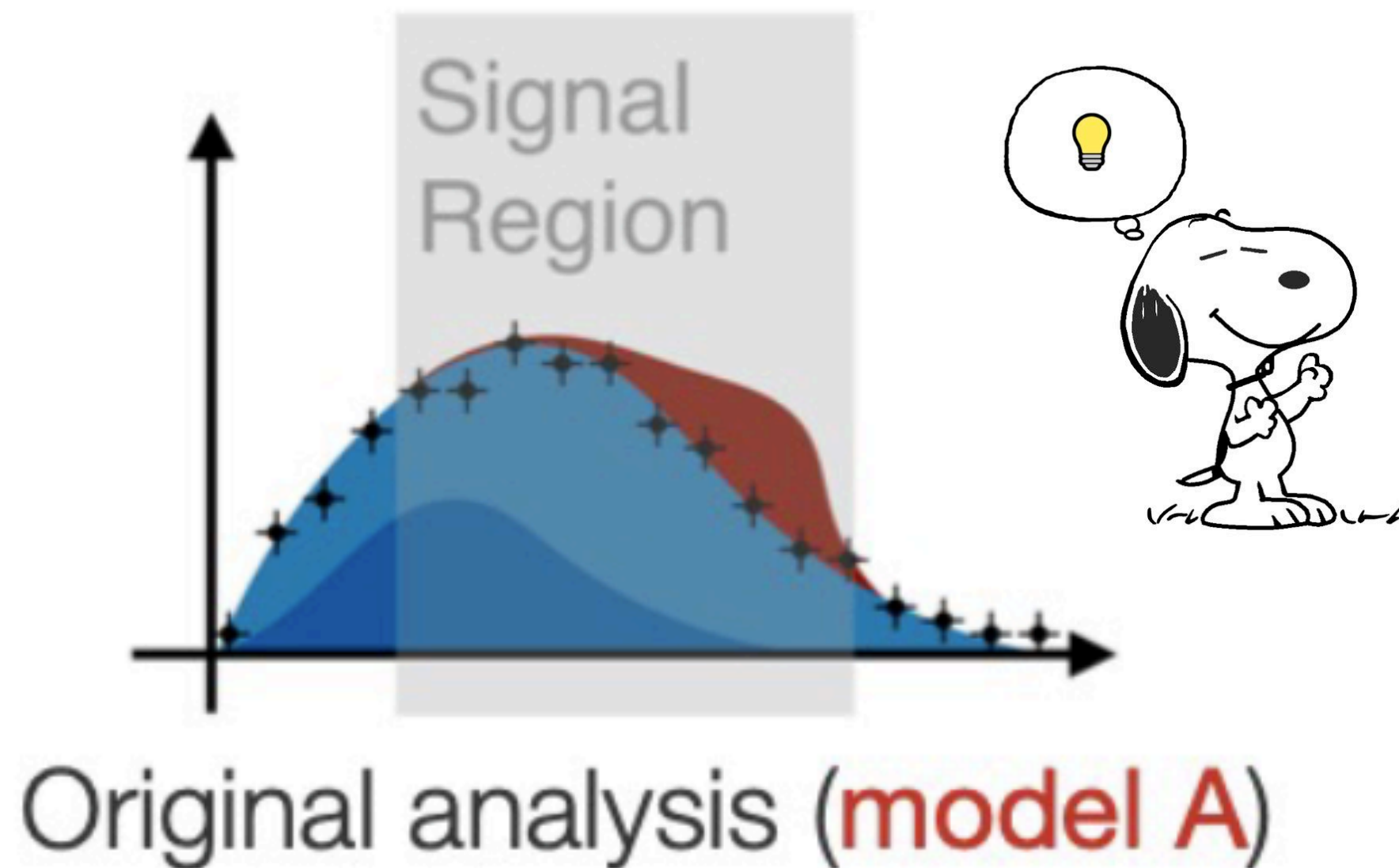


Doubly charged Higgs search
Eur.Phys.J.C83(2023)605 [link](#)



Reinterpretations of the LHC

- What if there are new BSM ideas that share signal features from old publications such as final state particles or event kinematics?
- e.g. With different vev, H^{++} can decay into pair of W bosons
 - 4ℓ final state is shared with the dedicated analysis but less clear resonance structure due to MET energy loss
 - Cannot directly translate one exclusion result to another



Reinterpretations of the LHC

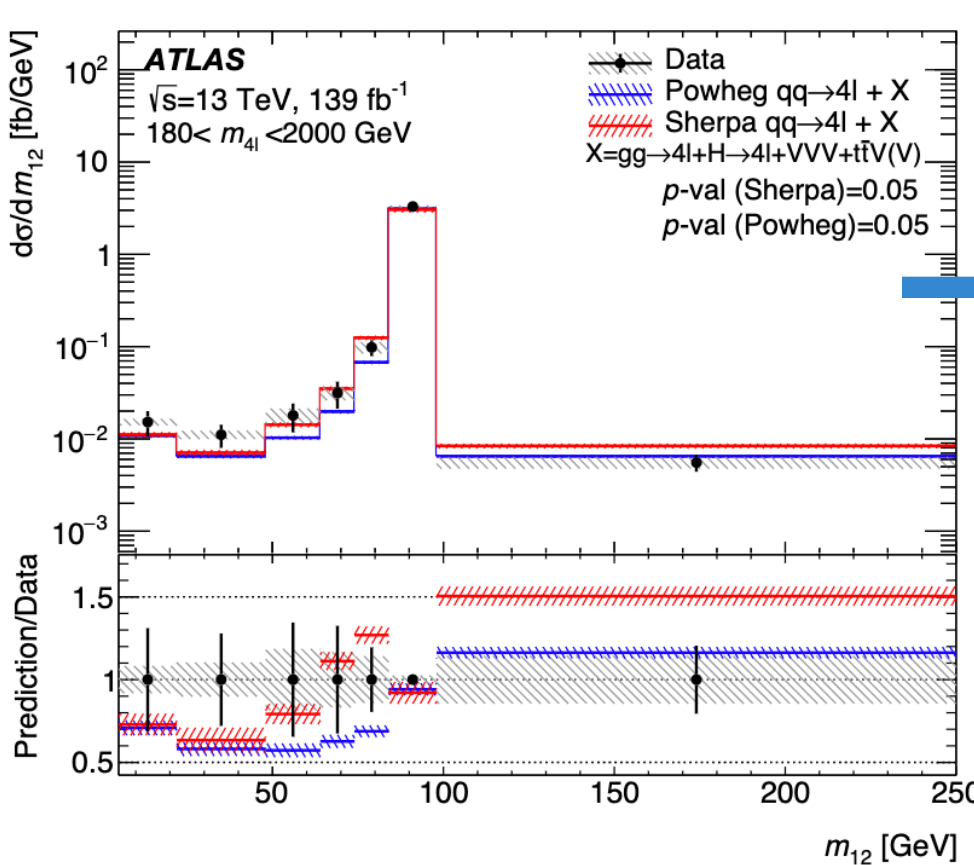
- Takes long time and a lot of effort is needed again to perform a dedicated analysis to scan new BSM ideas
- Assuming the analyses at the LHC are preserved in a recastable format, **model A** can be patched to **model B**
- Reinterpretation : Alternative solution that allows you to perform new interpretations based on existing results



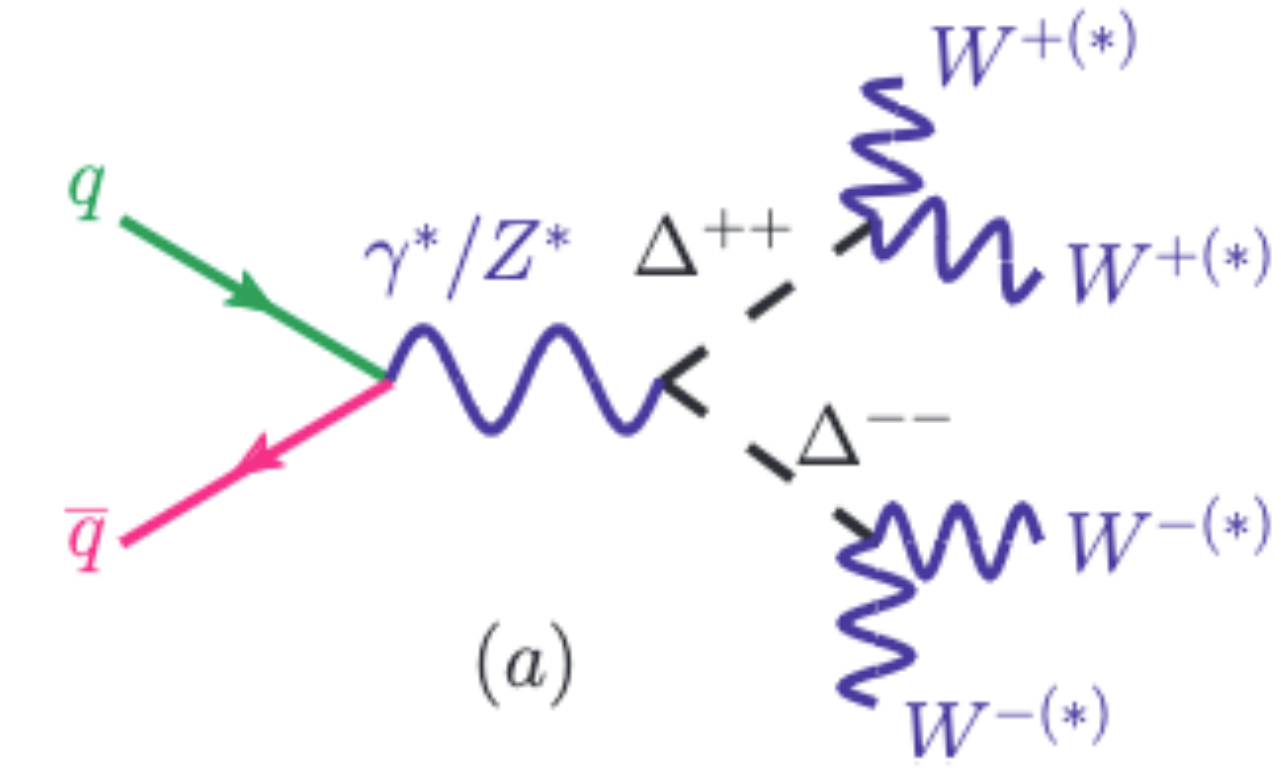
Reinterpretations of the LHC

- By preserving our results early on and making them recastable, new BSM theories can be explored with much lesser effort

PRESERVE in Rivet
4 lepton event cross section
[JHEP07\(2021\)005 link](#)

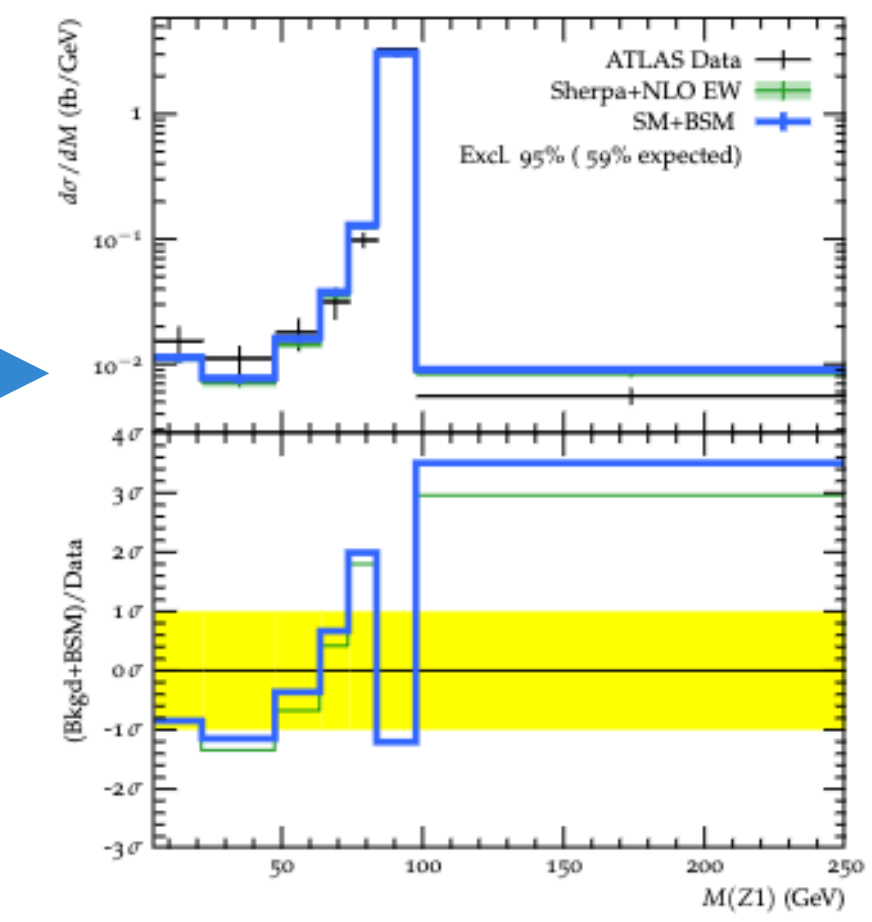


GENERATE new MC samples
New BSM idea to explore

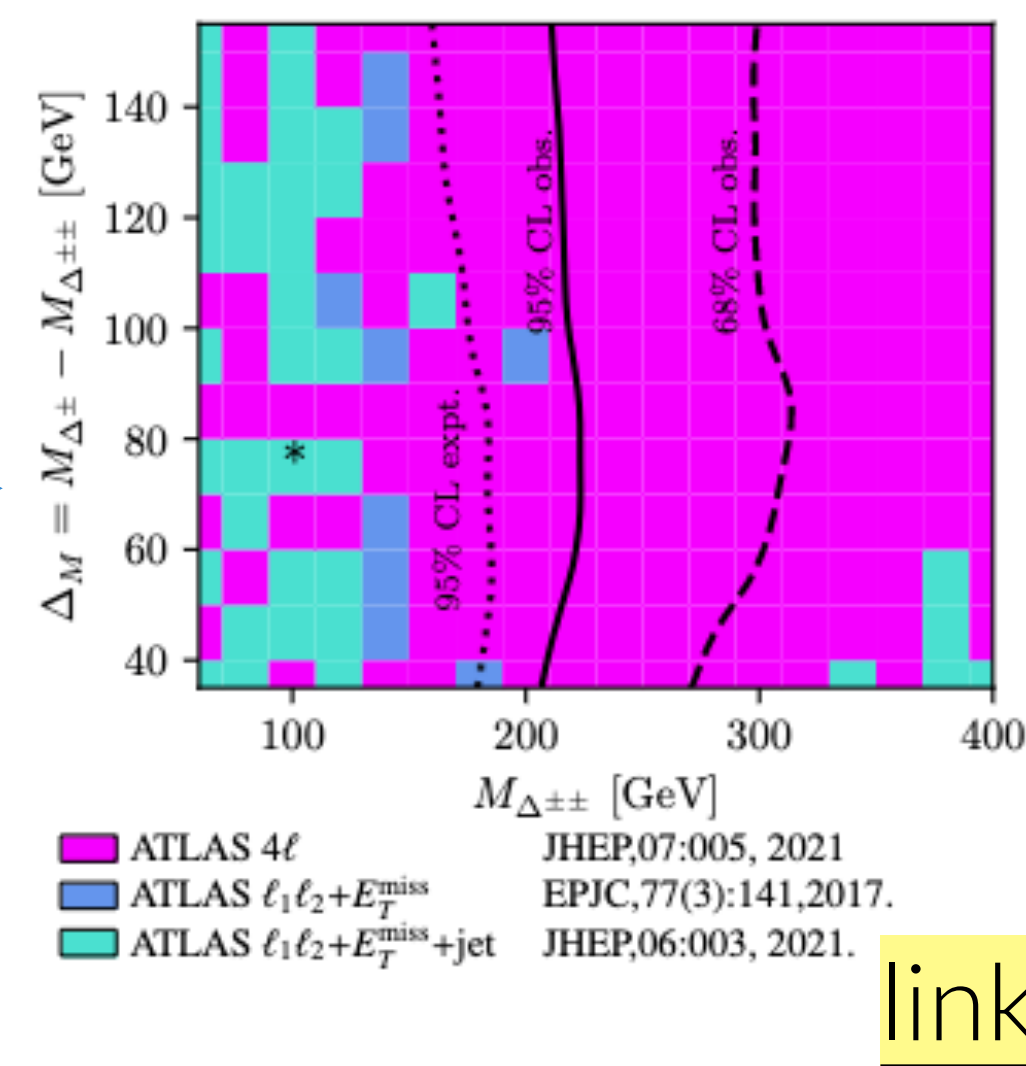


(a)

RECAST the preserved analysis
But with new BSM MC sample



REINTERPRET new BSM idea



link

- Best time to consider preservation is while performing the analysis (or right after it finishes) before our memories fade away and gets busy with other new tasks



The LHC BSM Reinterpretation Forum

- Suggestions from *the LHC BSM Reinterpretation Forum* on what/how to preserve
- Basic items that experiment can provide : Clear analysis descriptions in paper, numerical values of plots, ...
- Some more complicated stuffs were also suggested → Did we (experiments) follow up on such requests?

SciPost

SciPost Phys. 9, 022 (2020)

[link](#)

Reinterpretation of LHC results for new physics: status and recommendations after run 2

The LHC BSM Reinterpretation Forum

Abstract

We report on the status of efforts to improve the reinterpretation of searches and measurements at the LHC in terms of models for new physics, in the context of the LHC Reinterpretation Forum. We detail current experimental offerings in direct searches for new particles, measurements, technical implementations and Open Data, and provide a set of recommendations for further improving the presentation of LHC results in order to better enable reinterpretation in the future. We also provide a brief description of existing software reinterpretation frameworks and recent global analyses of new physics that make use of the current data.



Copyright W. Abdallah *et al.*

This work is licensed under the Creative Commons

[Attribution 4.0 International License](#).

Published by the SciPost Foundation.

Received 02-04-2020

Accepted 06-08-2020

Published 21-08-2020

doi:[10.21468/SciPostPhys.9.2.022](https://doi.org/10.21468/SciPostPhys.9.2.022)

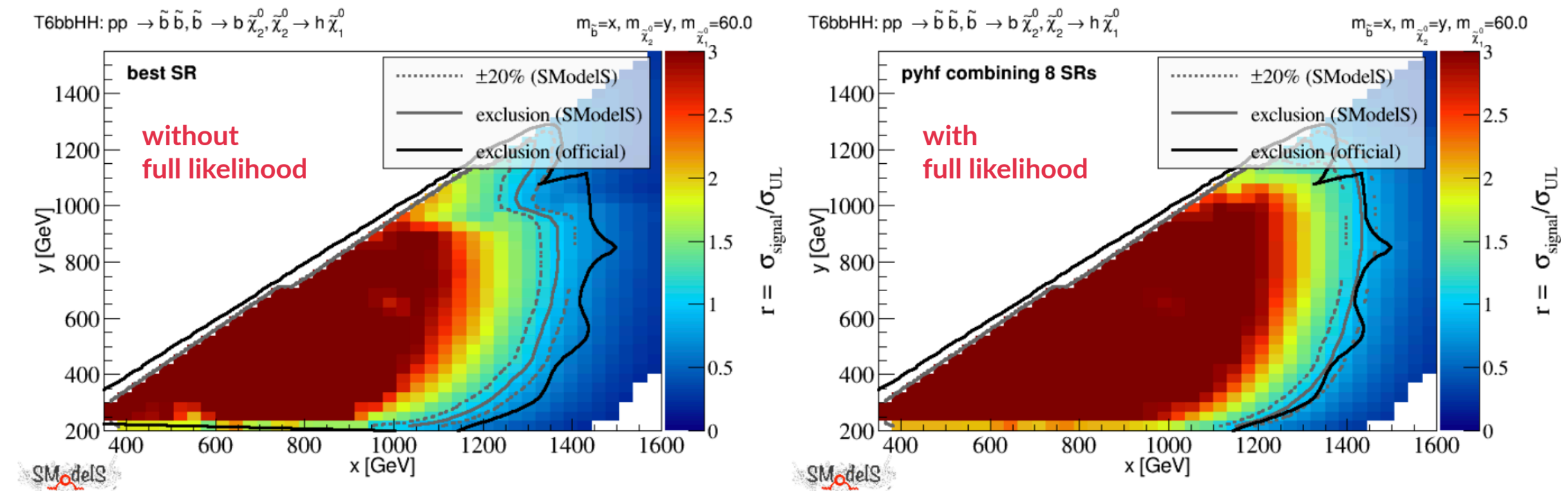


Check for updates

Suggestions and Feedbacks : Statistical Models

- Statistical models : Making use of all possible experimental sources as much as possible → Uncertainty
- Multiple reports on the usefulness of having the full statistical model → Able to recast official results from experiments with much better precision
- ATLAS put a lot of effort to release full likelihood through [pyhf](#) (python implemented HistFactory), already available in HEPData
- CMS recently released the [COMBINE](#) package, used for statistical analysis (submitted to *Computing and Software for Big Science*)

Comput.Phys.Comm.264,July2021,107909 [link](#)



Suggestions and Feedbacks : Reproduction Metadata

- Reproduction metadata : Cutflow tables for validation of fast simulations in recasting, analysis pseudocode with analysis logics
 - No general policy yet so pretty much vary in content and format but started having some concrete examples
 - One full example : CMS EXO-20-004 (monojet search)

1 to 4 of 4 Files

Analysis code in recasting tool (MadAnalysis) format

- `cms_exo_20_004.tcl`
C++ Source - 21,918 Downloads
MD5: 415...f31
Source code for the implementation
- `cms_exo_20_004.h`
C++ Source - 805 B
Published Jun 17, 2021
21,895 Downloads
MD5: 69f...4d7
Header file of the implementation
- `cms_exo_20_004.info`
Unknown - 369.1 KB
Published Oct 7, 2021
21,712 Downloads
MD5: c03...664
Info file with the experimental results / background expectation (including covariance matrix information)
- `delphes_card_cms_exo_20_004.tcl`
application/x-tcl - 22.0 KB
Published Jun 17, 2021
21,897 Downloads
MD5: 816...2f0
Detector parametrisation

Showing 50 of 10201 values

Covariance matrix to be used for simplified likelihood

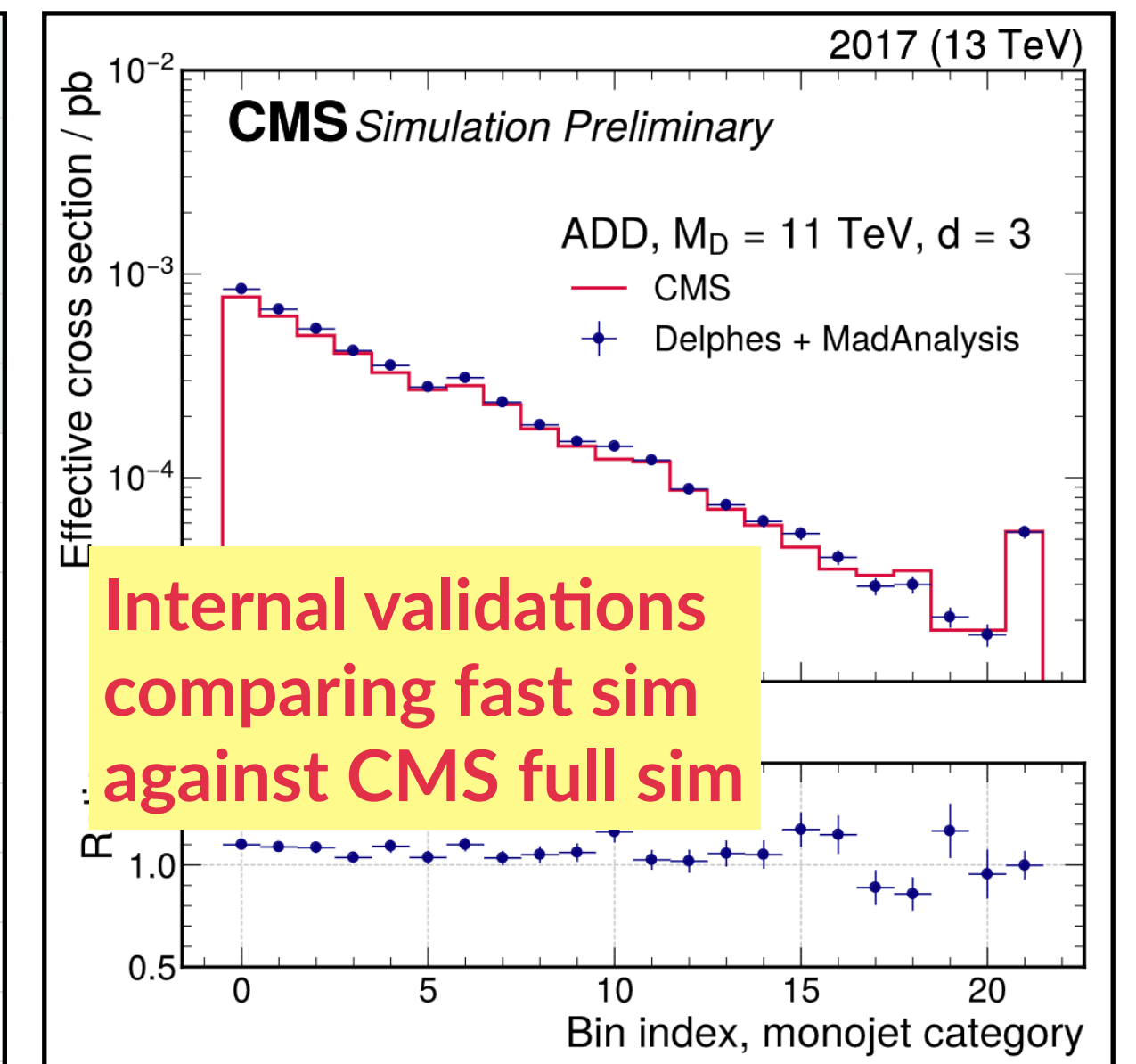
monojet_2016_bin0	monojet_2016_bin6	377857.6261
monojet_2016_bin0	monojet_2016_bin7	216898.8756
monojet_2016_bin0	monojet_2016_bin8	100131.7426
monojet_2016_bin0	monojet_2016_bin9	75505.18217

Yield table for distributions in SRs and CRs

Y	D	C	R	F	P	R
250.0 - 280.0	117100.0	7876.3	5783.5	98107.0	1272.7	273.97
280.0 - 310.0	67932.0	4447.2	3457.0	57258.0	773.44	167.69

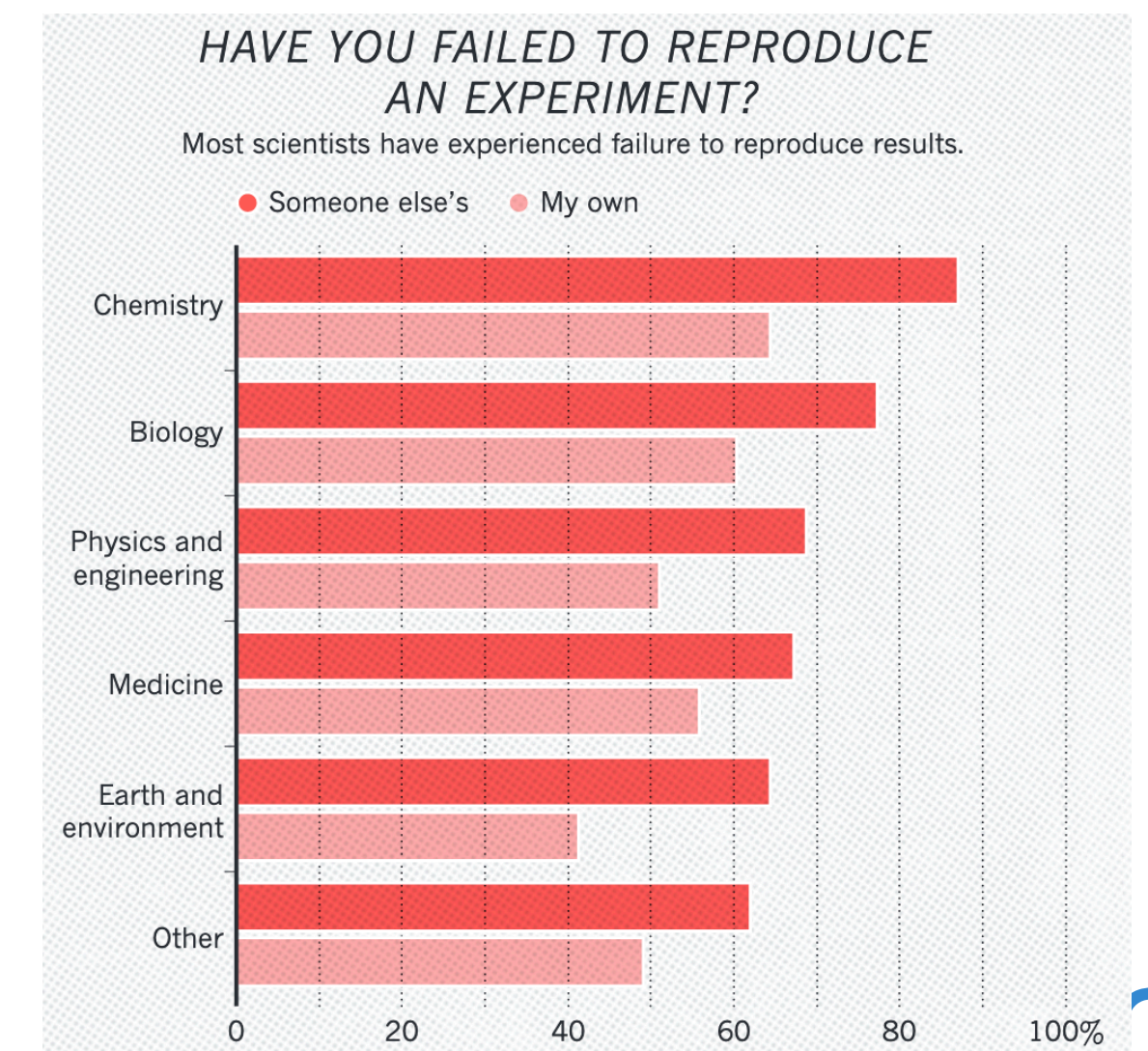
Cutflow table

$m_{\Phi\Phi}$	1000.0 GeV	300.0 GeV	1000.0 GeV
λ	1.0	1.0	1.0
Mode	$\Phi\Phi \rightarrow \chi\chi jj$		
Data-taking period	2017	2018	2018
Cut stage	Fraction of passing events		
Full sample	1.0	1.0	1.0
Trigger emulation	0.90814	0.5588	0.90293
$p_T^{miss} > 250$ GeV	0.72881	0.16807	0.72742
p_T^{miss} quality filters	0.72671	0.16768	0.72562
Electron veto	0.72431	0.1672	0.72357
Muon veto	0.72204	0.16673	0.72145
Tau veto	0.70581	0.16273	0.70537
B jet veto	0.70581	0.16273	0.70537



Legacy Analyses

- One of the main difficulties to conduct a full combination report comes from reproducibility
- e.g. Legacy analysis such as pMSSM studies (scan of 19 parameters) is basically a collection of multiple BSM searches, combining the results and scanning unexplored parameter sets
- This can only be done after accumulating some publications
- It is becoming more important to prepare ourselves for people who will be recasting/reproducing the analysis in the future for such legacy studies
- What do we need to do?



LHC Internal Improvements

- Reana : Reproducible research data analysis platform

Flexible

Run many computational workflow engines.



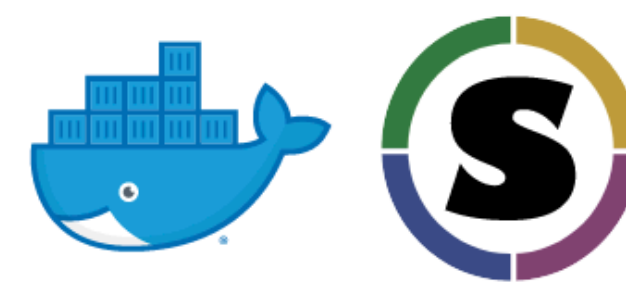
Scalable

Support for remote compute clouds.



Reusable

Containerise once, reuse elsewhere. Cloud-native.



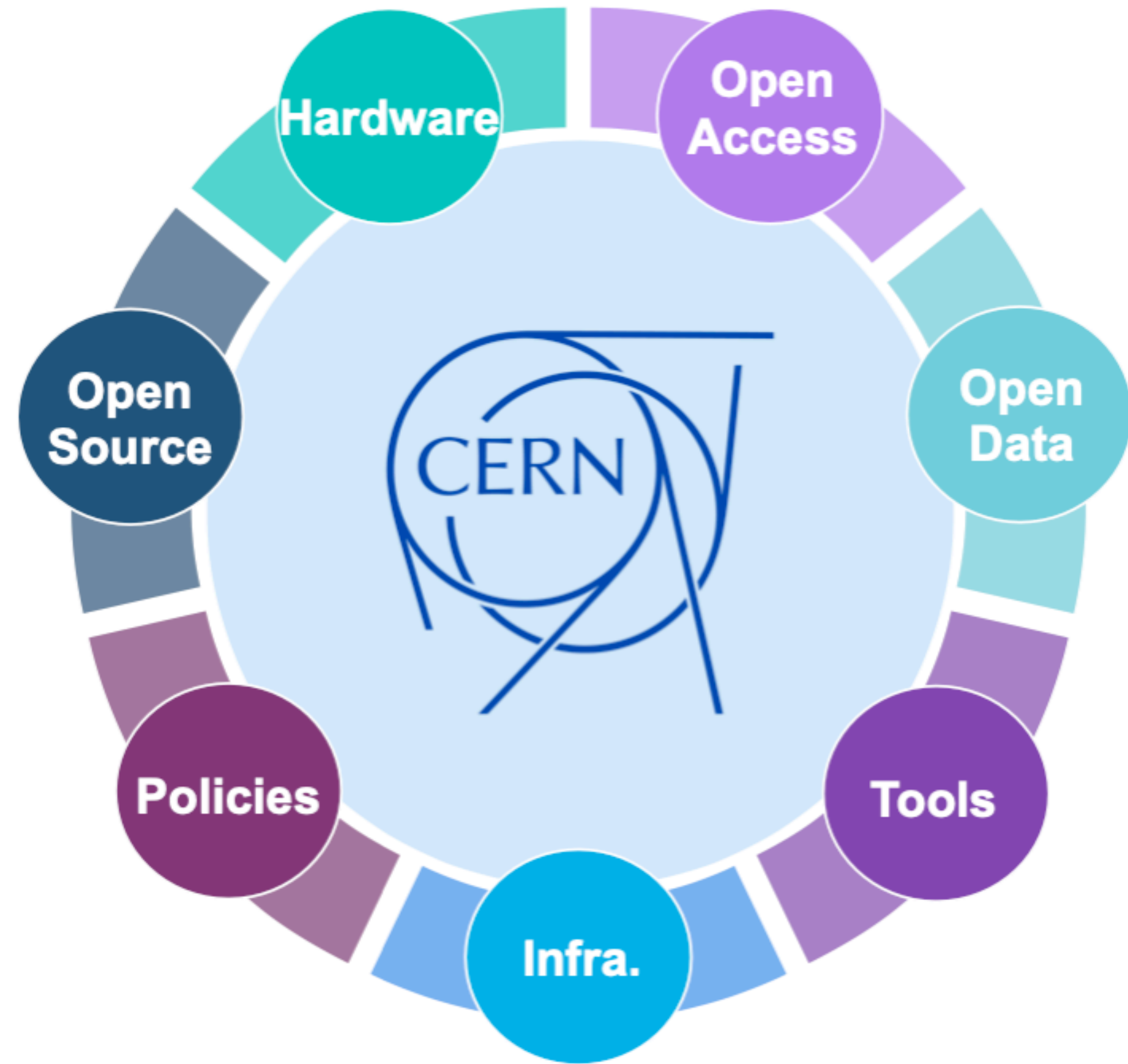
Free

Free Software. MIT licence. Made with ❤️ at CERN.



- Packaging codes that were used for the analysis with all the technical details (OS architecture, compilation environment, ...) containerised so that it can be reusable
- One very recent example is pMSSM studies from ATLAS
 - Running Recast and SimpleAnalysis frameworks (tools developed by ATLAS for reinterpretation) on Reana

CERN Open Science Policy



- Most advanced case of preservation & reinterpretation : Open data
- Release of LHC research quality data for public use

Learn	Visualise	Analyse
Discover the world of open data from particle physics	Explore detector events and run basic histogramming	Run your own physics analyses, start virtual machines
Welcome to our updated portal CMS Guide to education use of CMS Open Data Improving educational content with high school teachers: A field report from our summer students Glossary	CMS Event Display OPERA Event Display CMS Histograms	CMS Guide to research use of CMS Open Data ATLAS Higgs Machine Learning Challenge Getting Started with LHCb Open Data Getting Started with ALICE Open Data
more		more

CERN is committed to the advancement of science and the wide dissemination of knowledge by embracing and promoting practices making scientific research more open, collaborative, and responsive to societal changes

Summary

- Reinterpretation can only be done when LHC experiments talk and listen to outside community
- We also need feedbacks from outside community to understand what is needed
- Allows us (experiments) to focus on unexplored realms of the LHC physics and exploit its full potential
- Starting from very basics (HEPData and Rivet) we are ramping up with more cool stuffs (statistical model, metadata, internal resources such as Reana, and ultimately open data)

Note that there are many more recast/reinterpretation tools that were not mentioned today : Rivet/Contur, MadAnalysis, SModelS, Gambit, CheckMATE, ADL/CutLang, ... all has its unique philosophy and approaches!

Backup Slides

Slides

- Most of the slides inspired by Nick Wardle's seminar link
- And a lot of histograms and also ideas from pyhf authors slides