Data Preservation and Reinterpretation

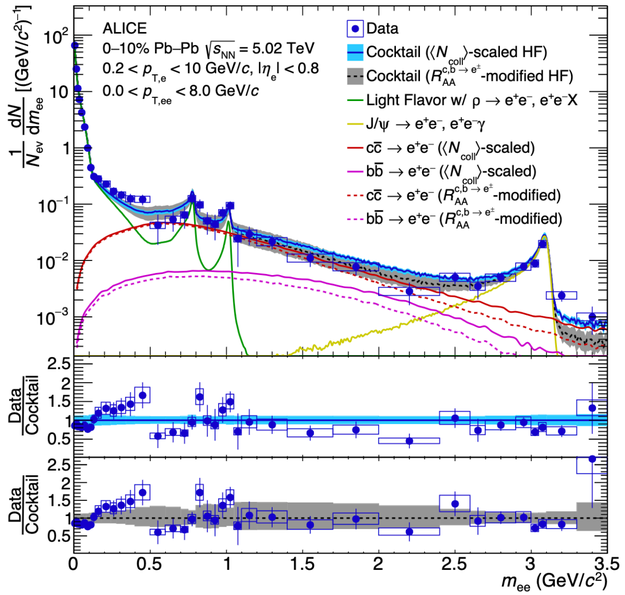
LHCP2024
June 3-7, 2024

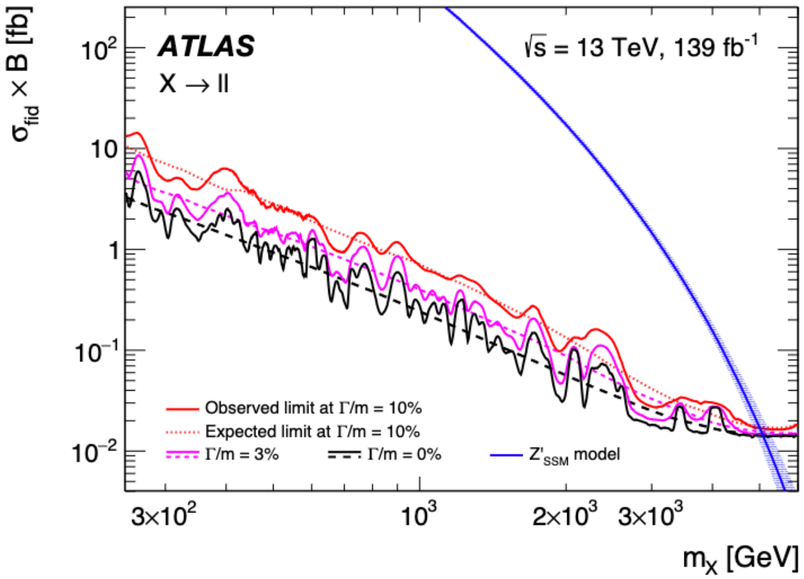
Si Hyun Jeon on behalf of LHC collaboration

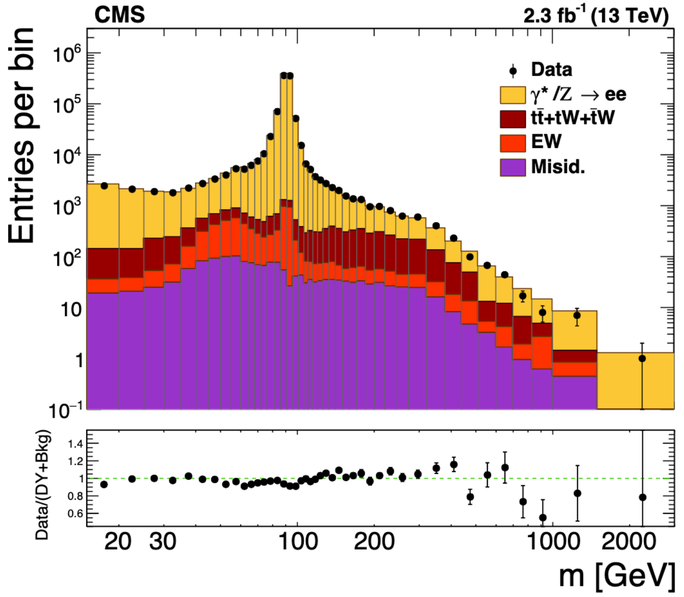
BOSTON UNIVERSITY

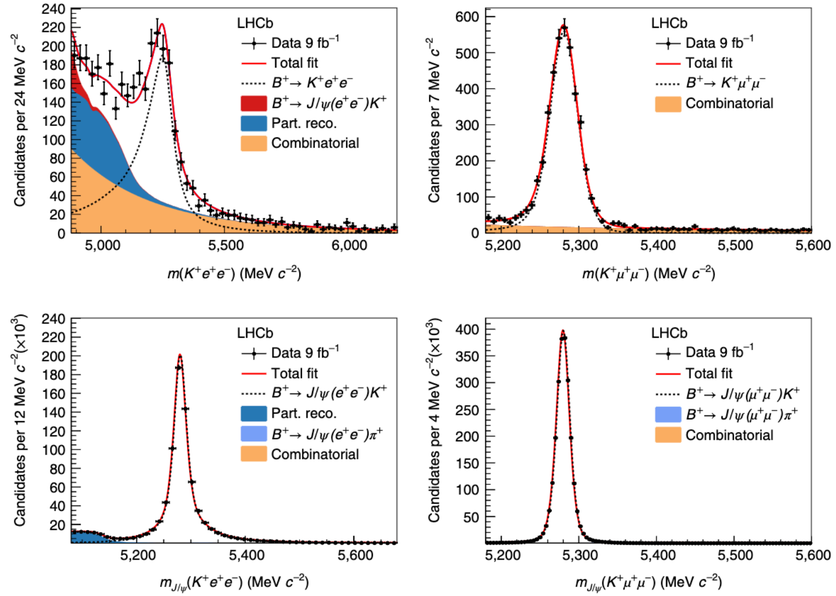
Interpretations from the LHC

- 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"



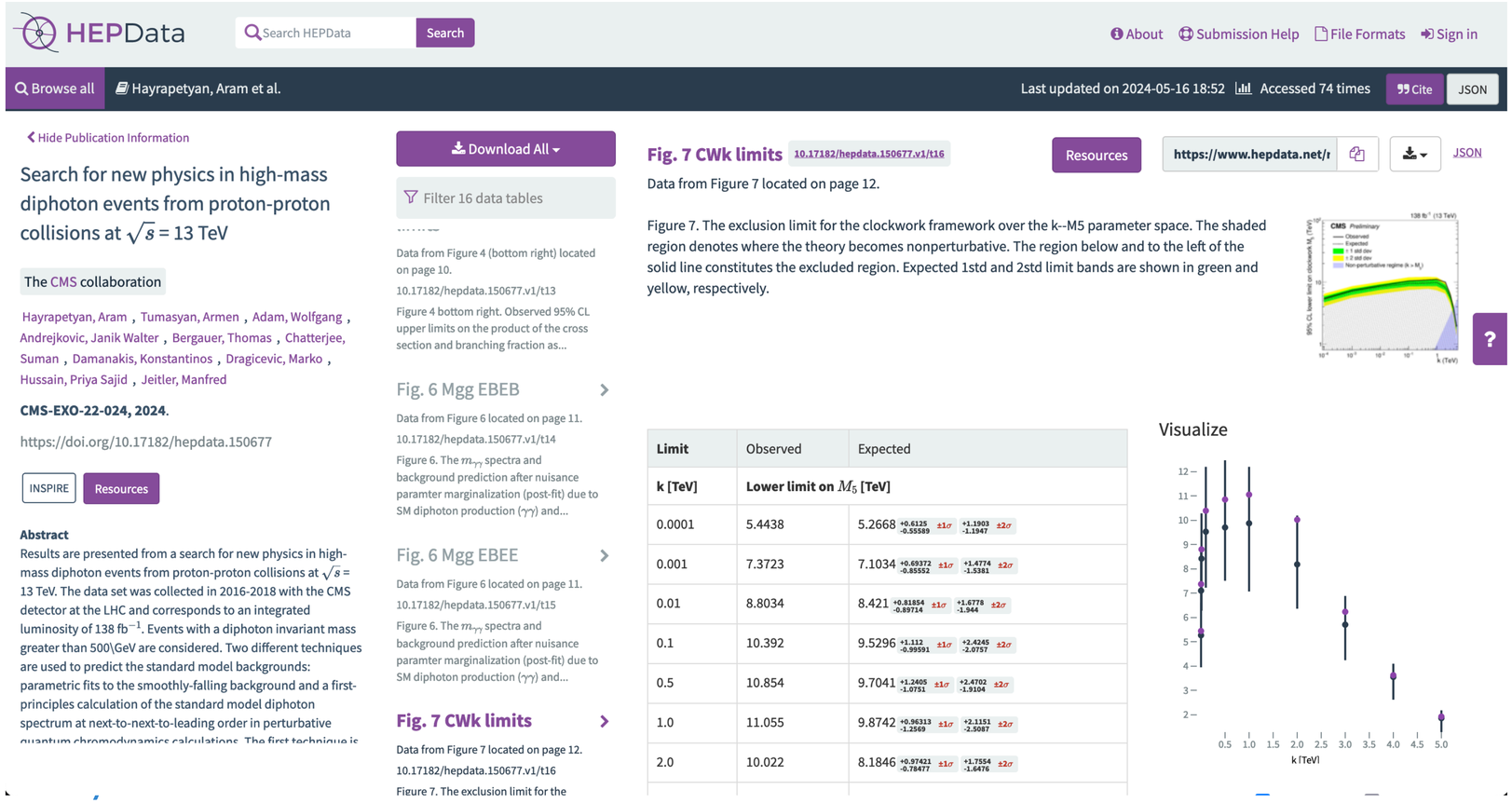


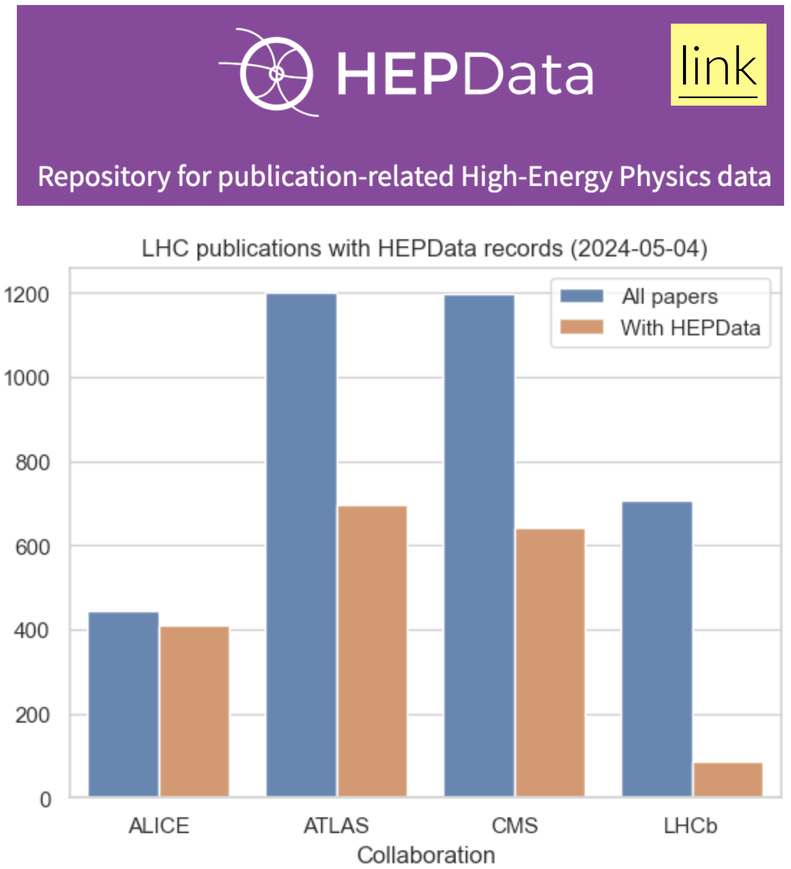




Preservations at the LHC : HEPData

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





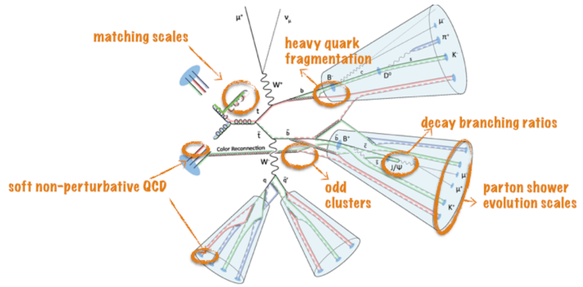
Preservations at the LHC : Rivet

- 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

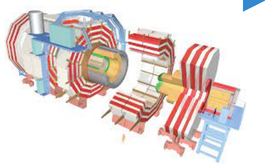
Key	ALICE	ATLAS	CMS	LHCb	Forward	HERA	e^+e^- (≥ 12 GeV)	e^+e^- (≤ 12 GeV)	Tevatron	RHIC	SPS	Other
Rivet wanted (total):	327	401	504	198	16	475	660	364	1118	509	57	97
Rivet REALLY wanted:	55	50	99	16	0	15	1	1	9	2	5	1
Rivet provided:	32/359 = 9%	202/603 = 33%	114/618 = 18%	31/229 = 14%	9/25 = 36%	35/510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

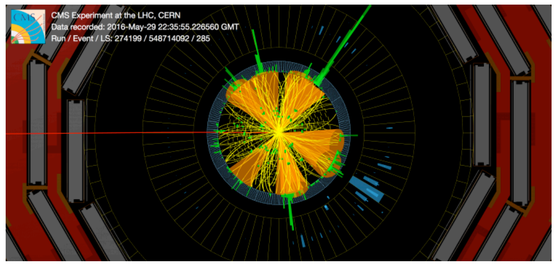
```
47    void analyze(const Event& event) {
48
49      const Jets jets = apply<JetFinder>(event, "Jets").jetsByPt(Cuts::pT > 300*GeV && Cuts::abseta < 2.1);
50
51      if (jets.size() < 2)  vetoEvent;
52      if (jets[0].pT() < 675*GeV)  vetoEvent;
53
54      if ( (jets[0].pT()/jets[1].pT()) > 1.5 ) 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
```

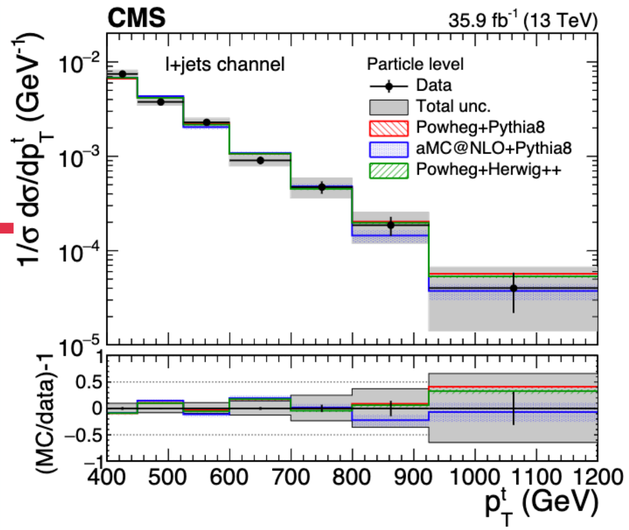


Detector simulation

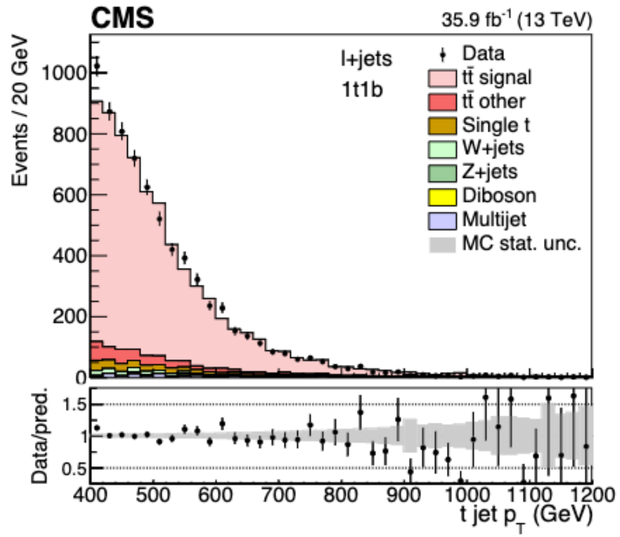
Smearing





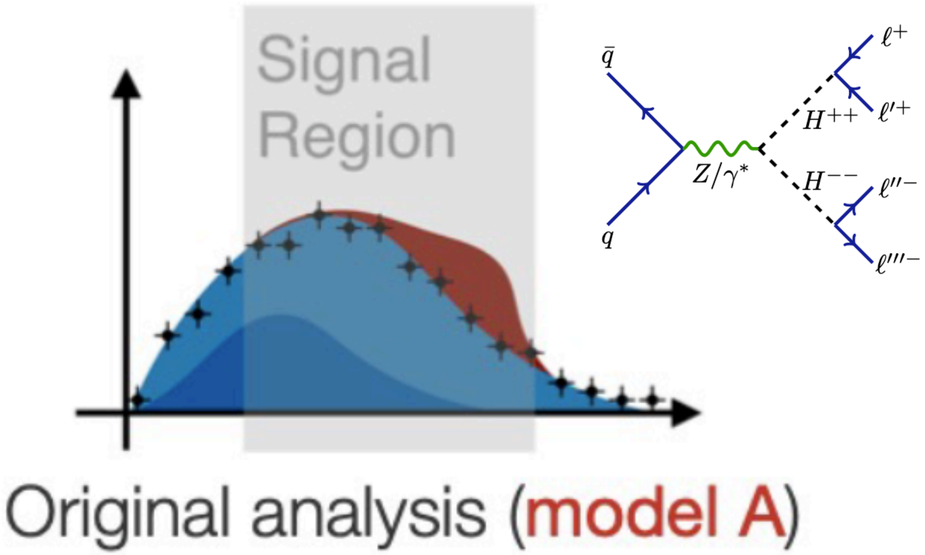


Unfolding



Reinterpretations of the LHC

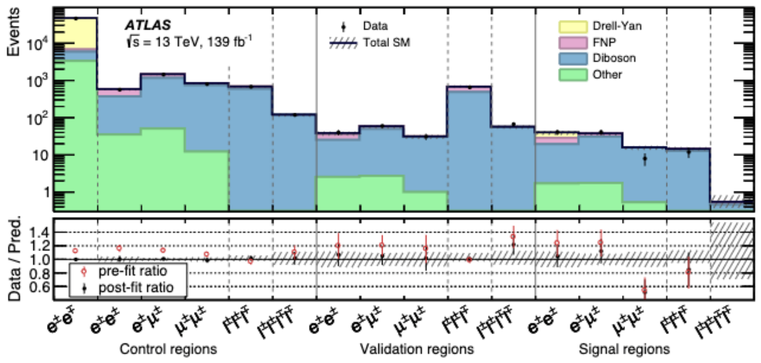
- BSM interpretations at the LHC mostly consider handful of models
  - Test data against the SM predictions in various channels
  - Try to look for unexplained deviations at a good confidence level
- Model A excluded by testing data against SM predictions
  - e.g. m(H^{++}) 400-1000GeV excluded assuming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

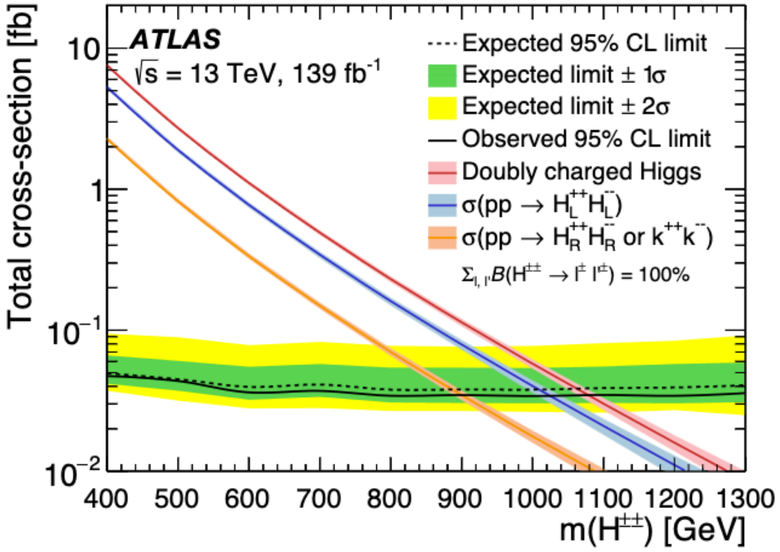


Original analysis (model A)

Doubly charged Higgs search

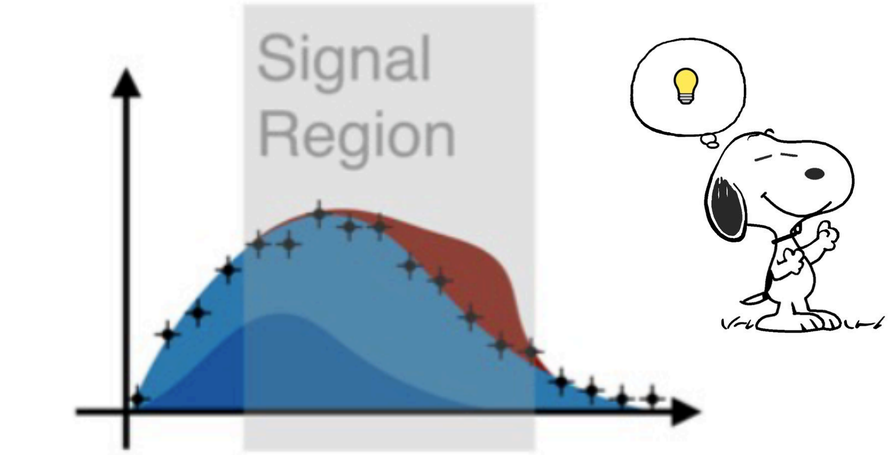
Eur.Phys.J.C83(2023)605 link



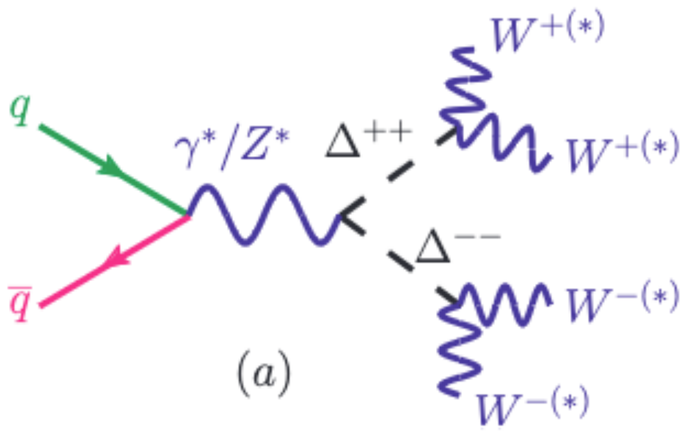


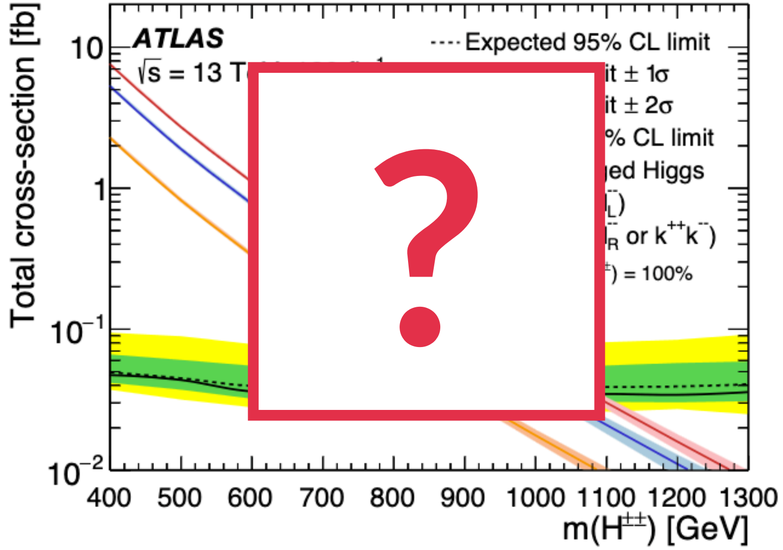
Reinterpretations of the LHC

- 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



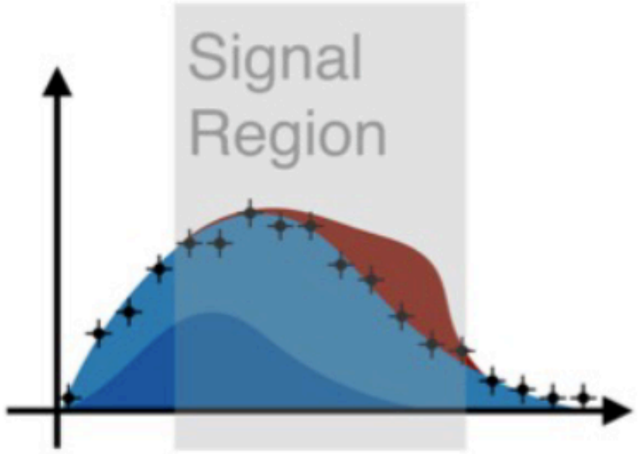
Original analysis (model A)



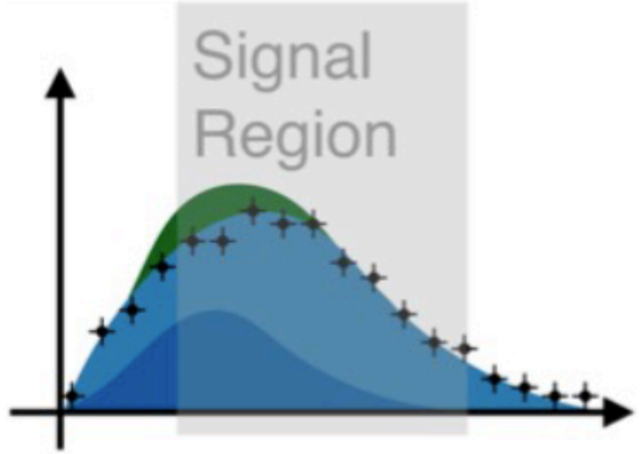


Reinterpretations of the LHC

- 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



Recasting &
Reinterpretation



Original analysis (model A)

Reinterpreted analysis (model B)

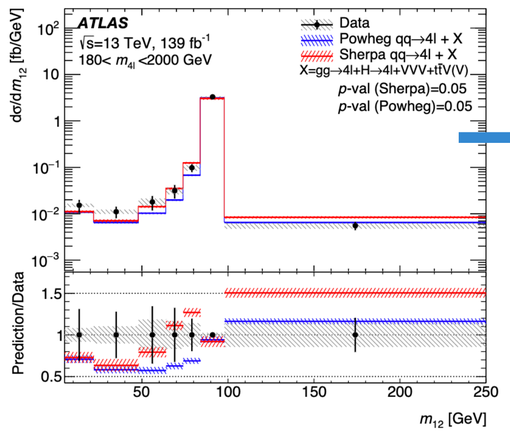
Reinterpretations of the LHC

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

PRESERVE in Rivet

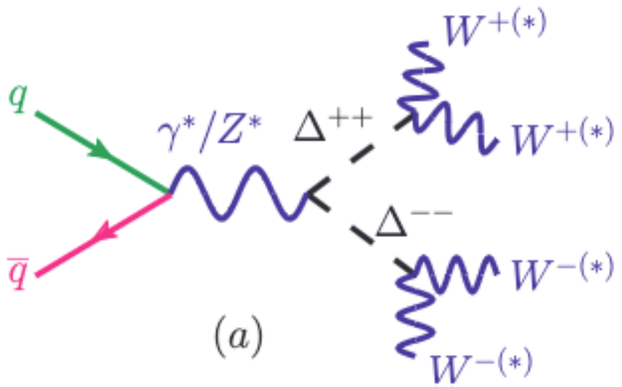
4 lepton event cross section

JHEP07(2021)005 link



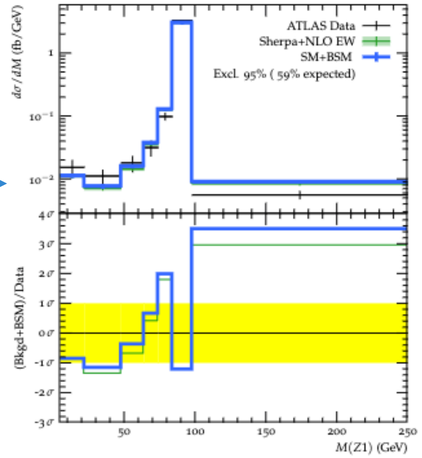
GENERATE new MC samples

New BSM idea to explore

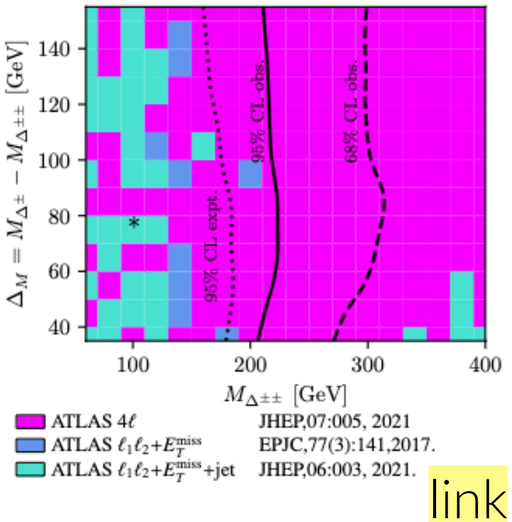


RECAST the preserved analysis

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

Si Hyun Jeon

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.

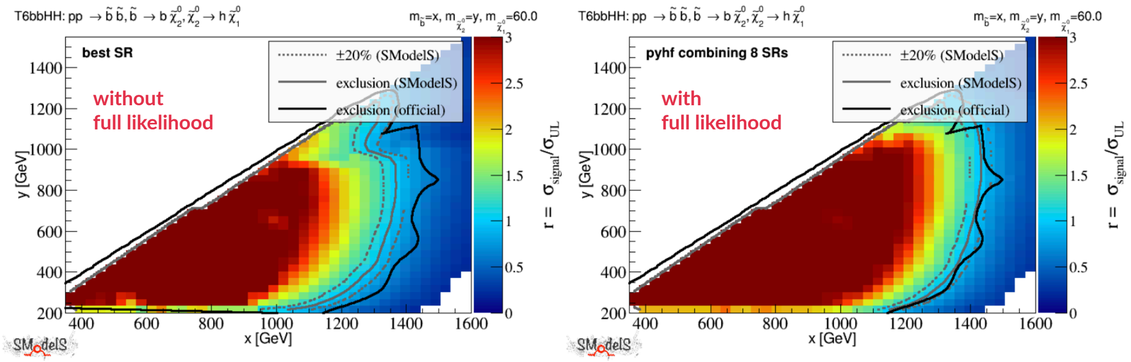
Received 02-04-2020
Accepted 06-08-2020
Published 21-08-2020
doi:10.21468/SciPostPhys.9.2.022

Check for updates

Suggestions and Feedbacks : Statistical Models

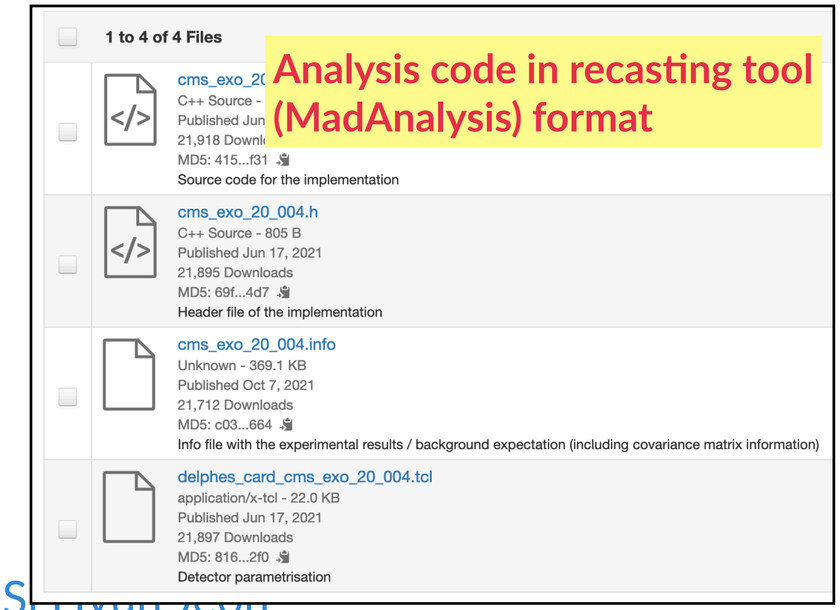
Comput.Phys.Comm.264,July2021,107909 link

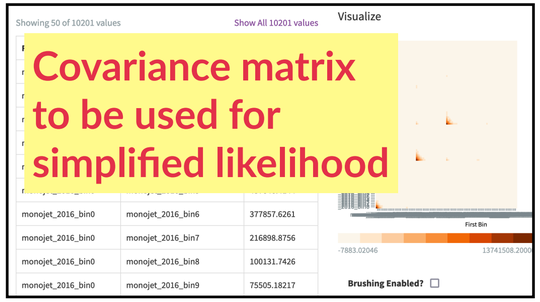
- Statistical models : Making use of all possible experimental sources as much as possible → Uncertainty
  - Multiple reports on the usefulness of having the full statistical model → Able to recast official results from experiments with much better precision
  - ATLAS put a lot of effort to release full 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*)

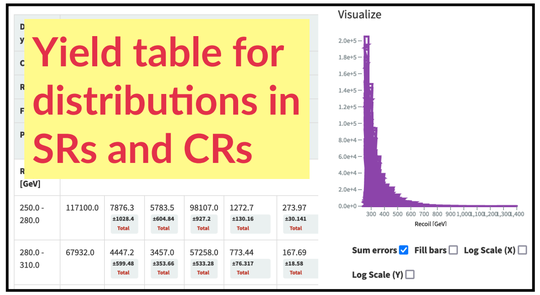


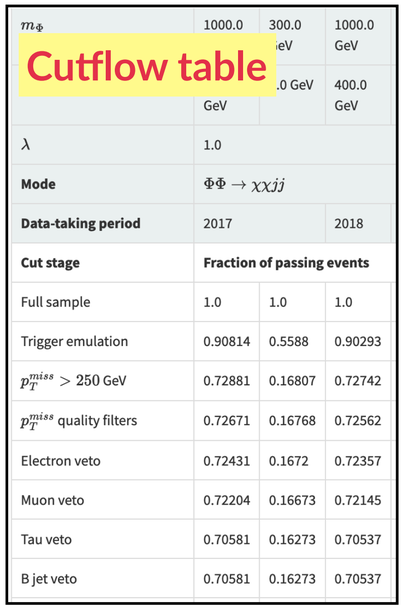
Suggestions and Feedbacks : Reproduction Metadata

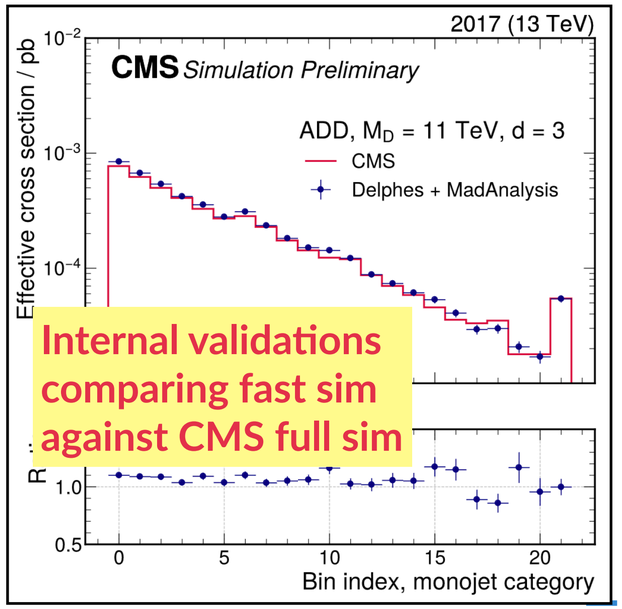
- Reproduction metadata : Cutflow tables for validation of fast simulations in recasting, analysis pseudocode with analysis logics
  - No general policy yet so pretty much vary in content and format but started having some concrete examples
  - One full example : CMS EXO-20-004 (monojet search)





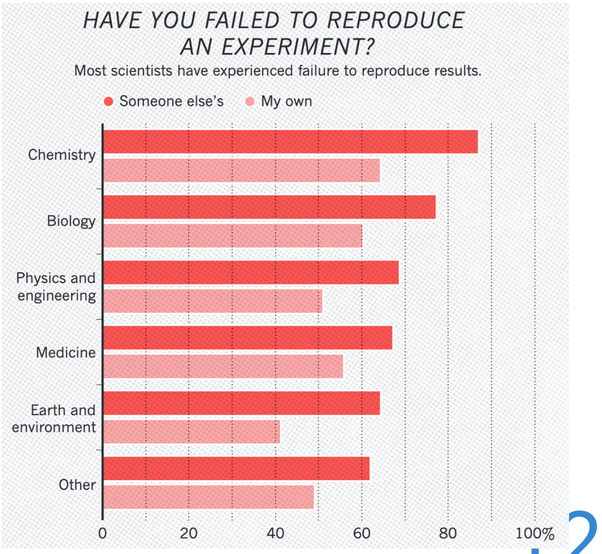






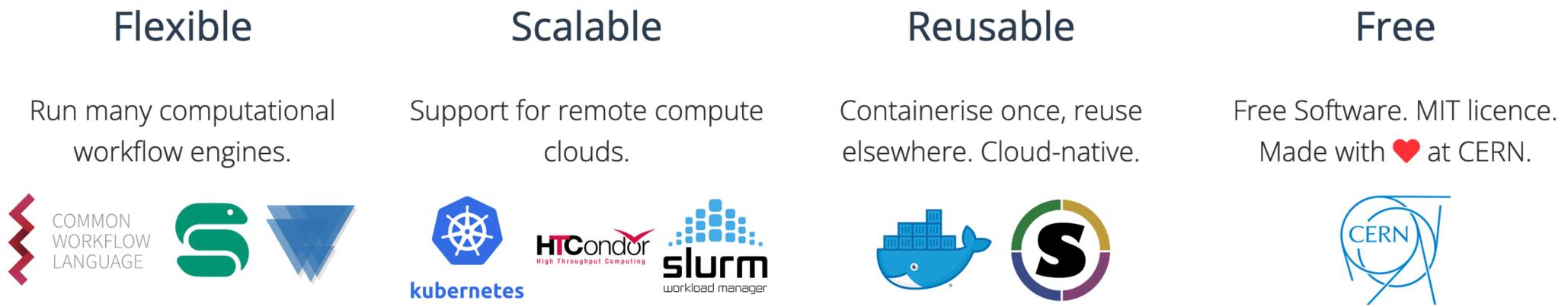
Legacy Analyses

- 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?



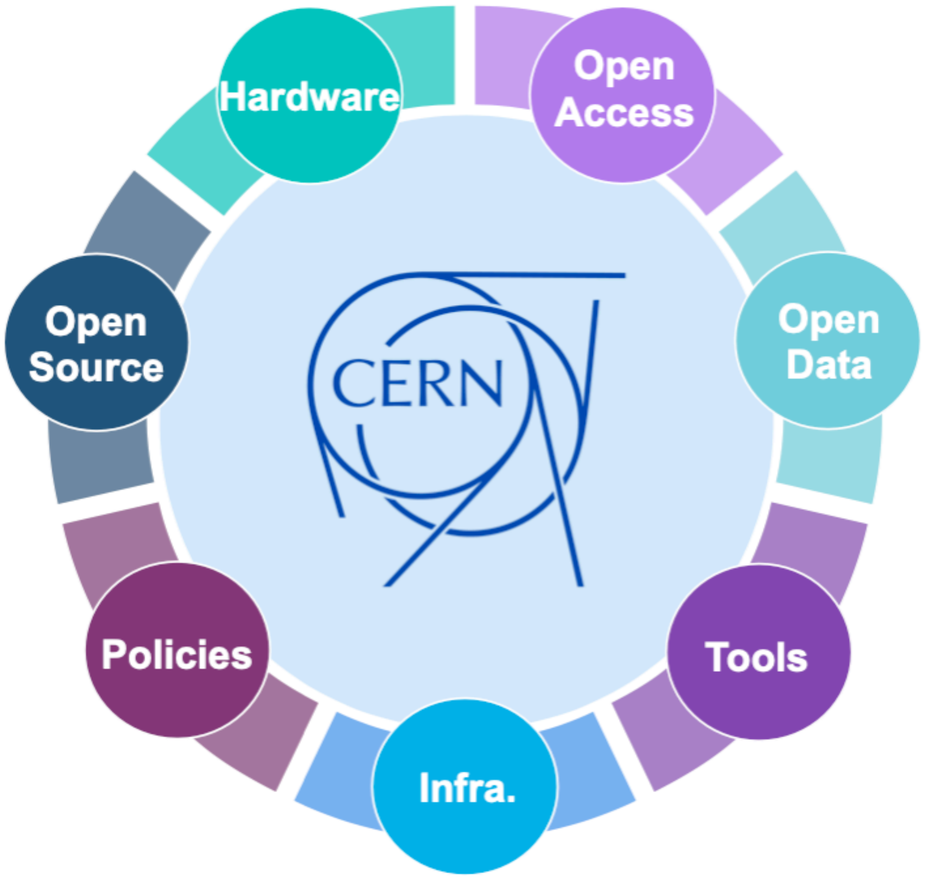
LHC Internal Improvements

- Reana : Reproducible research data analysis platform



- 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

CERN is commited 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