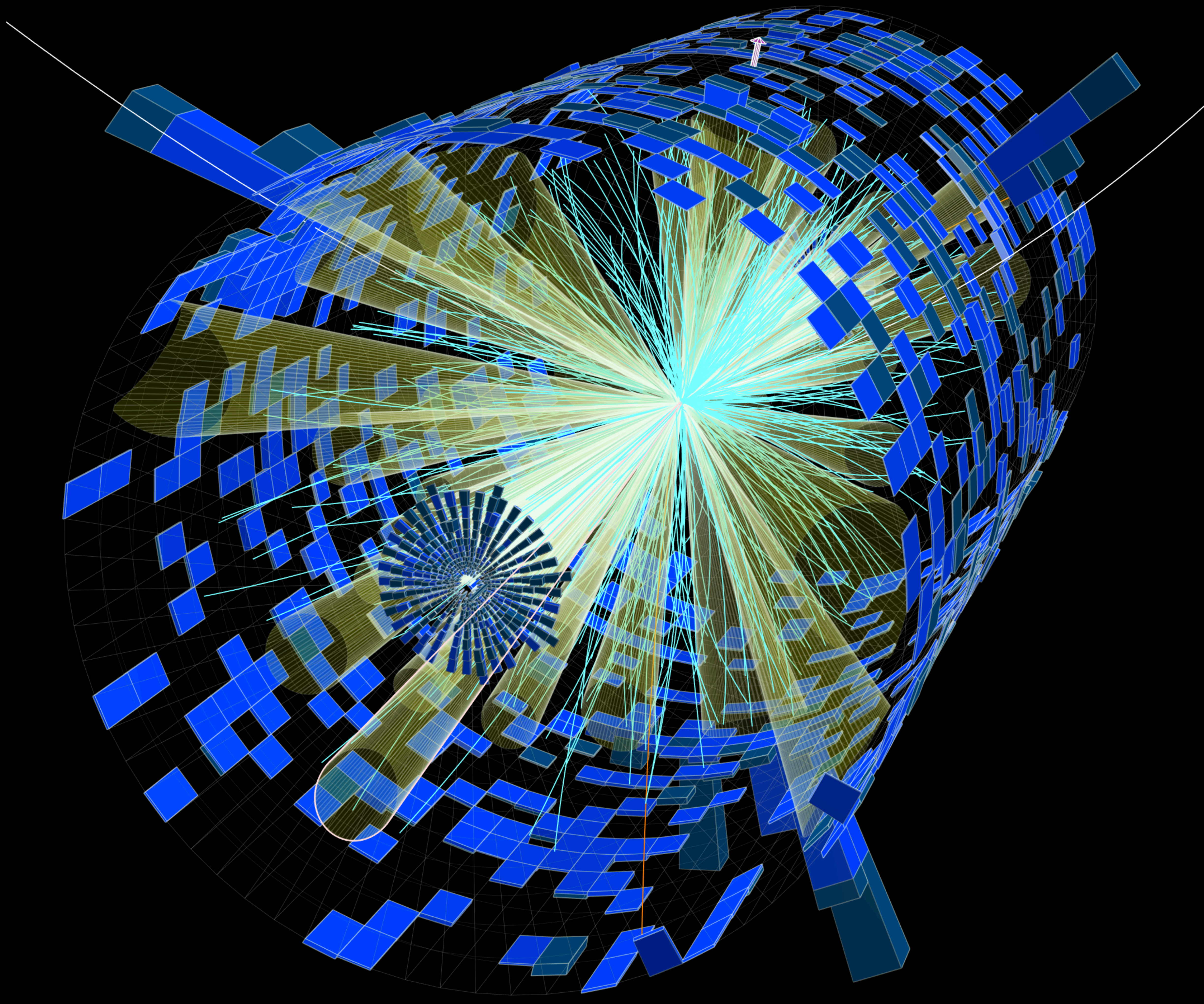
~~A decade~~ 14 years of HistFactory



@KyleCranmer
University of Wisconsin-Madison
Data Science Institute
Physics, Computer Science, Statistics

Prompt

Create a light-hearted talk of a middle-aged physicist walking down memory lane.

It should incorporate:

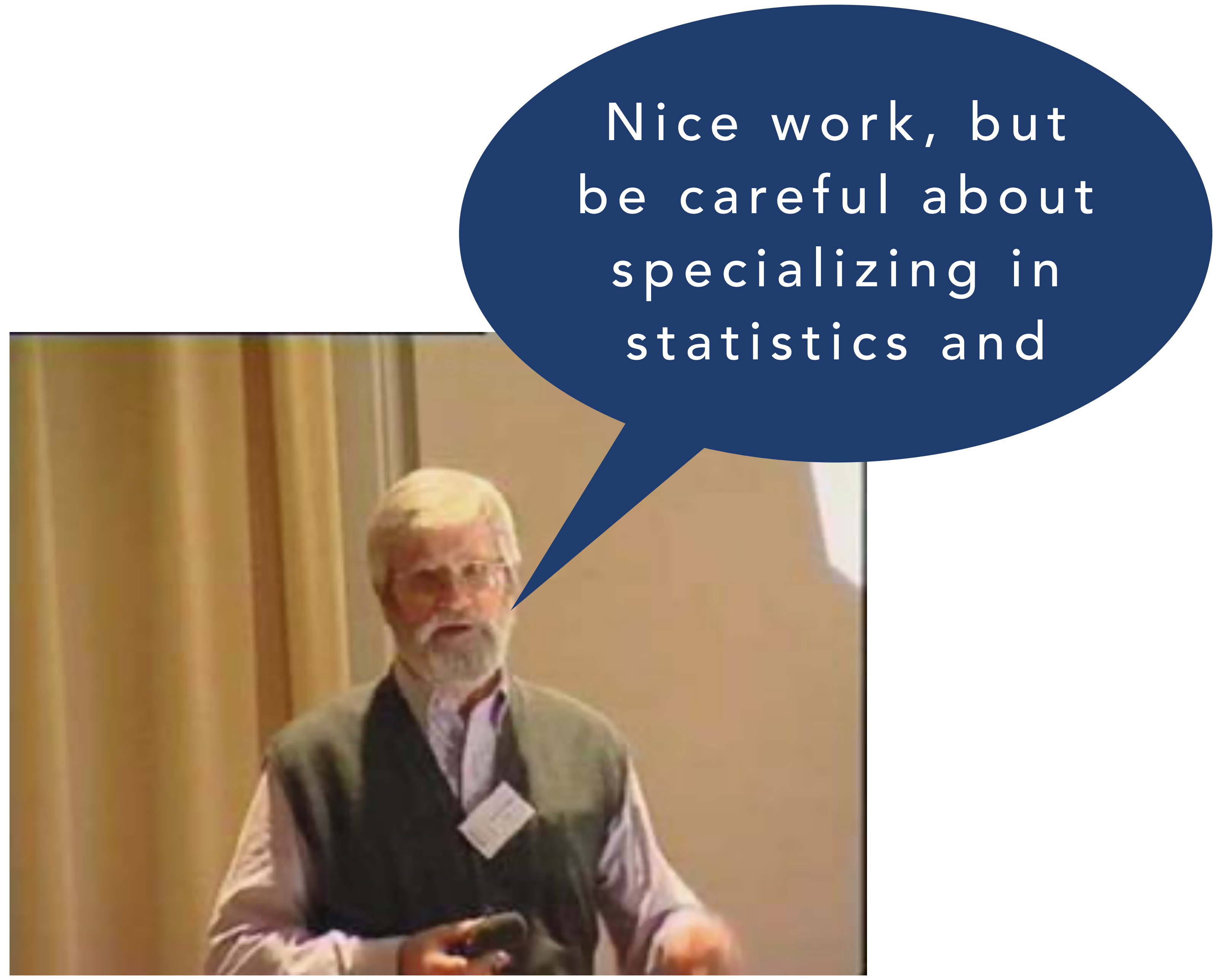
- computers
- equations, and
- the large hadron collider.

Make a few observations or insightful points and end with something inspiring

KEYS

I joined ALEPH in 1999. My first paper was about modeling distributions so they would be smoother than histograms. Also my first community software project (in perl and PAW)

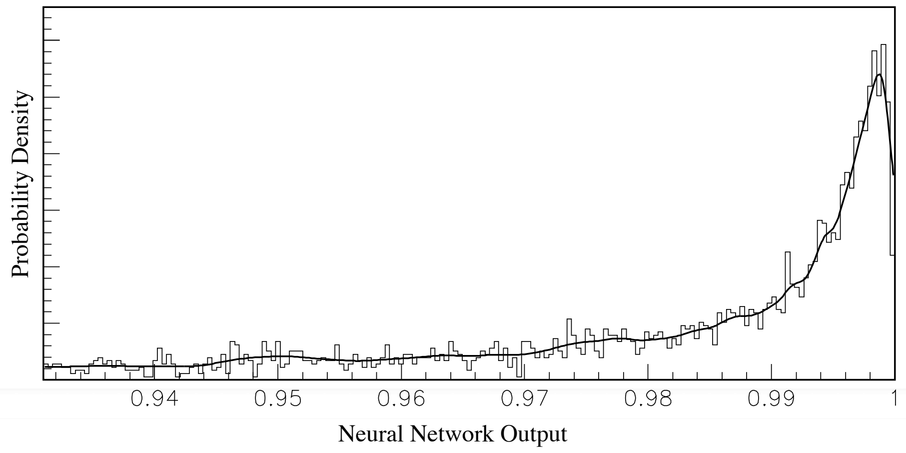
- Similarities to HistFactory: simple configuration file, users provide ntuples as input

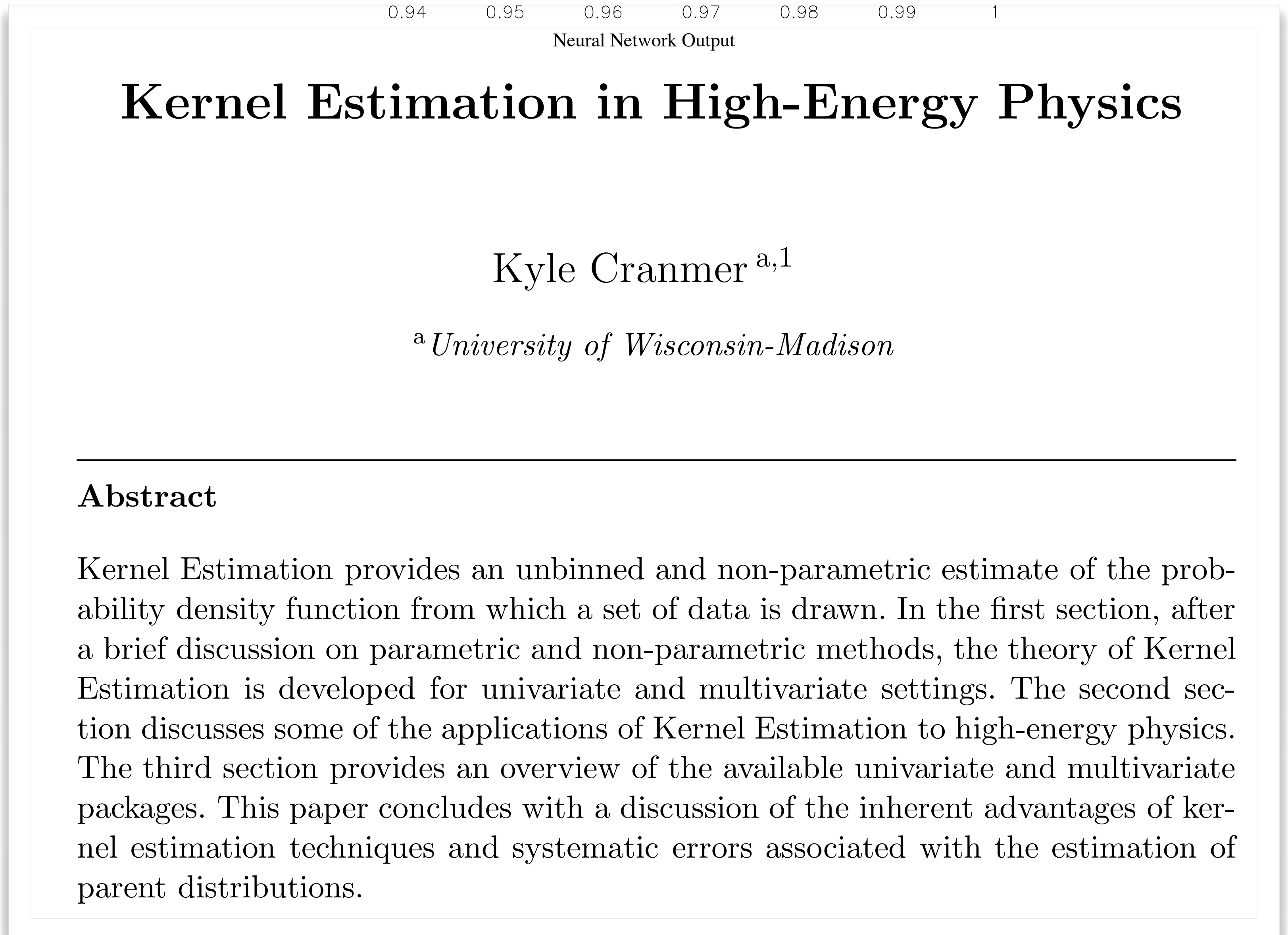


Fred James

Author of MINUIT / MINOS

Editor of Computer Physics Communications for many years





Kernel Estimation in High-Energy Physics

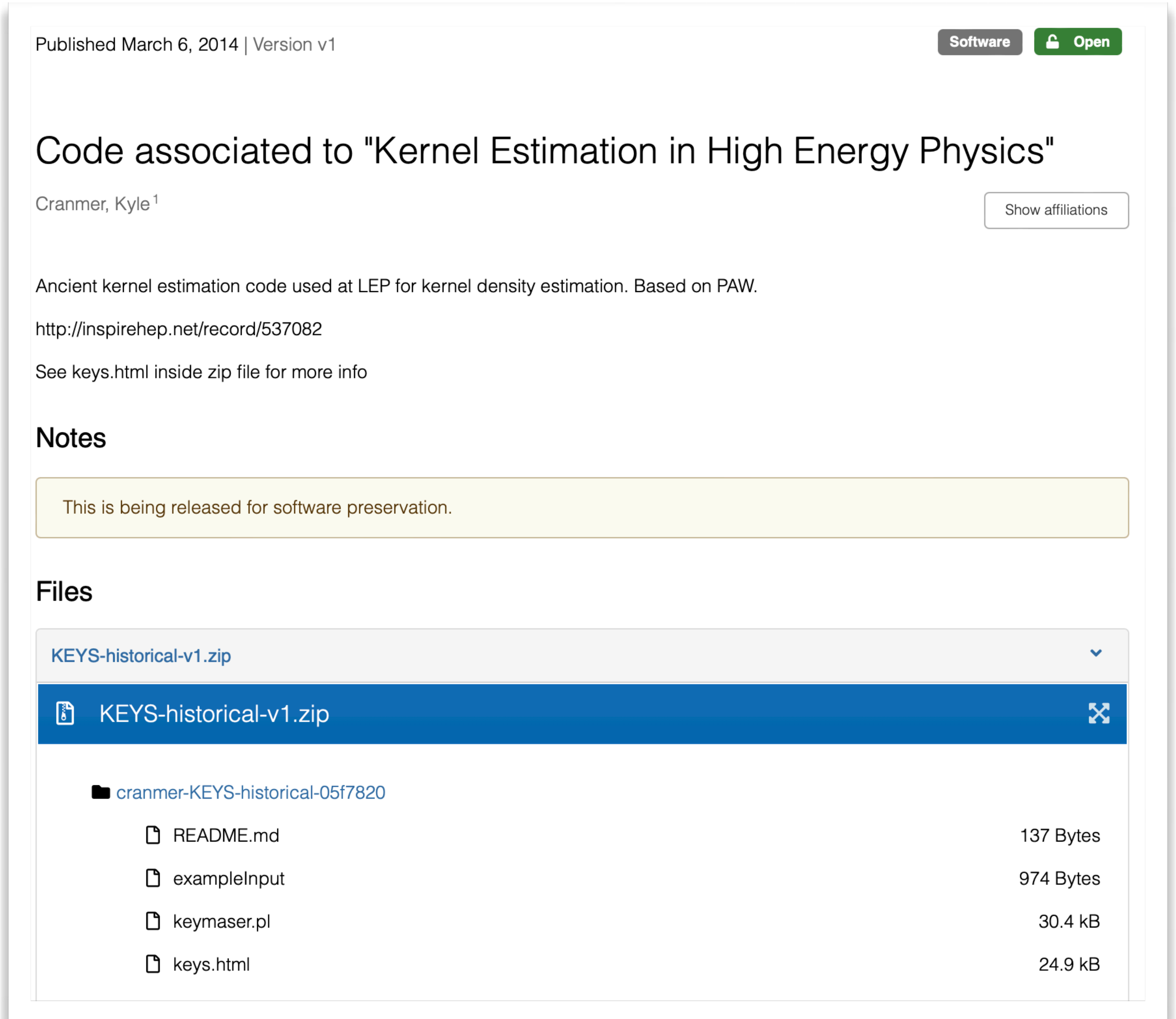
Kyle Cranmer [a,1]

[a] *University of Wisconsin-Madison*

Abstract

Kernel Estimation provides an unbinned and non-parametric estimate of the probability density function from which a set of data is drawn. In the first section, after a brief discussion on parametric and non-parametric methods, the theory of Kernel Estimation is developed for univariate and multivariate settings. The second section discusses some of the applications of Kernel Estimation to high-energy physics. The third section provides an overview of the available univariate and multivariate packages. This paper concludes with a discussion of the inherent advantages of kernel estimation techniques and systematic errors associated with the estimation of parent distributions.

Comput.Phys.Commun. 136 (2001) 198-207
e-Print: hep-ex/0011057



Published March 6, 2014 | Version v1 Software Open

Code associated to "Kernel Estimation in High Energy Physics"

Cranmer, Kyle [1] Show affiliations

Ancient kernel estimation code used at LEP for kernel density estimation. Based on PAW.

http://inspirehep.net/record/537082

See keys.html inside zip file for more info

Notes

This is being released for software preservation.

Files

KEYS-historical-v1.zip

KEYS-historical-v1.zip

cranmer-KEYS-historical-05f7820

README.md	137 Bytes
exampleInput	974 Bytes
keymaser.pl	30.4 kB
keys.html	24.9 kB

One of the first zenodo software DOIs
https://doi.org/10.5281/zenodo.8468

The first PhyStat

It was 23 years ago!

- I was just starting as a graduate student

Louis Lyons of Oxford, co-convenor of the workshop on confidence limits.

https://cds.cern.ch/record/411537?ln=en

CERN 2000–005
30 May 2000

ORGANISATION EUROPÉENNE POUR LA RECHERCHE NUCLÉAIRE
CERN EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH

WORKSHOP ON CONFIDENCE LIMITS

CERN, Geneva, Switzerland
17–18 January 2000

PROCEEDINGS

Editors: F. James, L. Lyons, Y. Perrin

GENEVA
2000

MASSIMO CORRADI:

Does everyone agree on this statement,
to publish likelihoods?

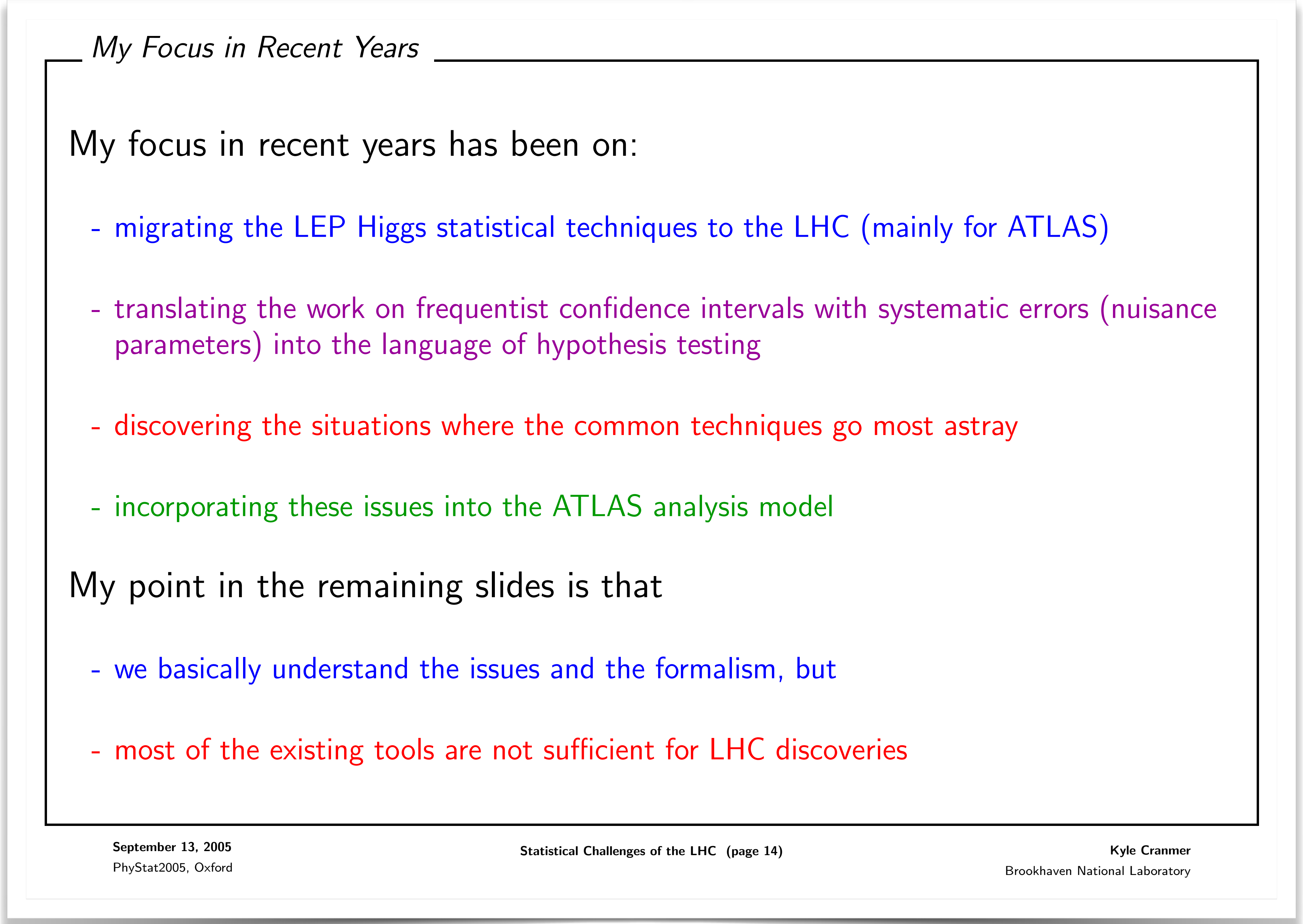
LOUIS LYONS:

Any disagreement? Carried unanimously.
That's actually quite an achievement for this workshop.

Birth of RooStats

From 2003-2005 I was working on understanding the statistical challenges of the LHC and how it was different than LEP (and Tevatron)

- After PhyStat2005, Rene Brun (ROOT) asked me to propose a statistical framework for ROOT
- Decided to build it on top of RooFit developed at BaBar



My Focus in Recent Years

My focus in recent years has been on:

- migrating the LEP Higgs statistical techniques to the LHC (mainly for ATLAS)
- translating the work on frequentist confidence intervals with systematic errors (nuisance parameters) into the language of hypothesis testing
- discovering the situations where the common techniques go most astray
- incorporating these issues into the ATLAS analysis model

My point in the remaining slides is that

- we basically understand the issues and the formalism, but
- most of the existing tools are not sufficient for LHC discoveries

September 13, 2005
PhyStat2005, Oxford

Statistical Challenges of the LHC (page 14)

Kyle Cranmer
Brookhaven National Laboratory

https://confs.physics.ox.ac.uk/phystat05/Talks/PhyStat2005_LHC_Challenges.pdf

Birth of RooStats

From 2003-2005 I was working on understanding the statistical challenges of the LHC and how it was different than LEP (and Tevatron)

- After PhyStat2005, Rene Brun (ROOT) asked me to propose a statistical framework for ROOT
- Decided to build it on top of RooFit developed at BaBar

My Focus in Recent Years

My focus in recent years has been on:

- migrating the LEP Higgs statistical techniques to the LHC (mainly for ATLAS)
- translating the work on frequentist confidence intervals with systematic errors (nuisance parameters) into the language of hypothesis testing
- discovering the situations where the common techniques go most astray
- incorporating these issues into the ATLAS analysis model

My point in the remaining slides is that

- we basically understand the issues and the formalism, but
- most of the existing tools are not sufficient for LHC discoveries

September 13, 2005
PhyStat2005, Oxford

Statistical Challenges of the LHC (page 14)

Kyle Cranmer
Brookhaven National Laboratory

https://confs.physics.ox.ac.uk/phystat05/Talks/PhyStat2005_LHC_Challenges.pdf

Design principles for RooStats

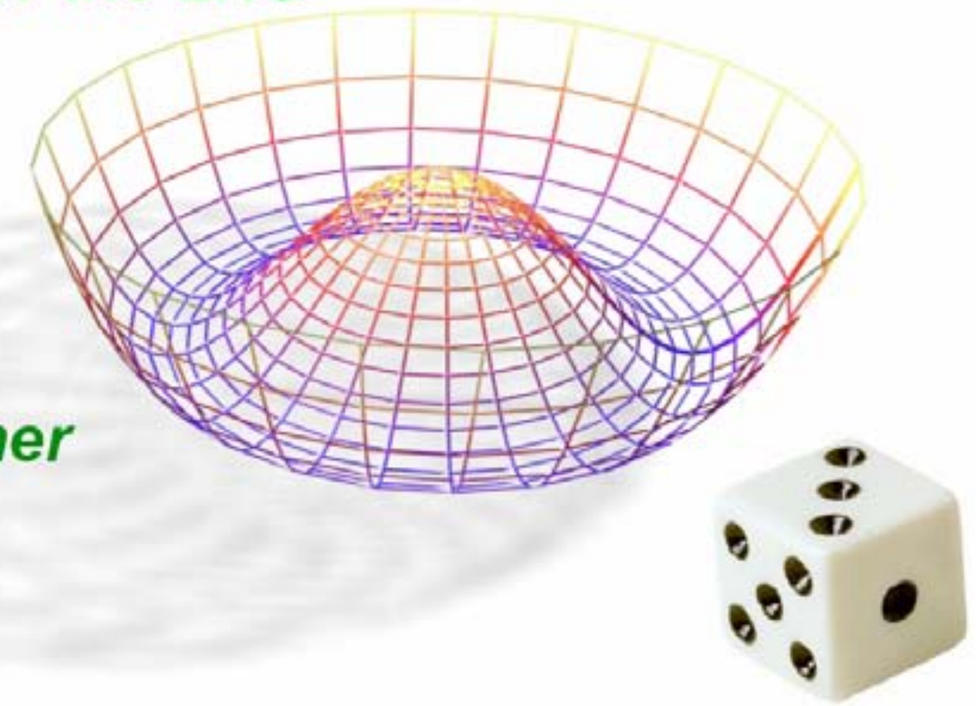
- At LEP and TeVatron the statistical tools mixed everything together
  - Bayesian vs Frequentist
  - What test statistic to use
  - How to model distributions
  - How to model systematics

With RooStats I wanted to **factorize** the **statistical modeling** (RooFit & later HistFactory) from **statistical procedures** (RooStats).

PhyStat 2007

BROOKHAVEN
NATIONAL LABORATORY

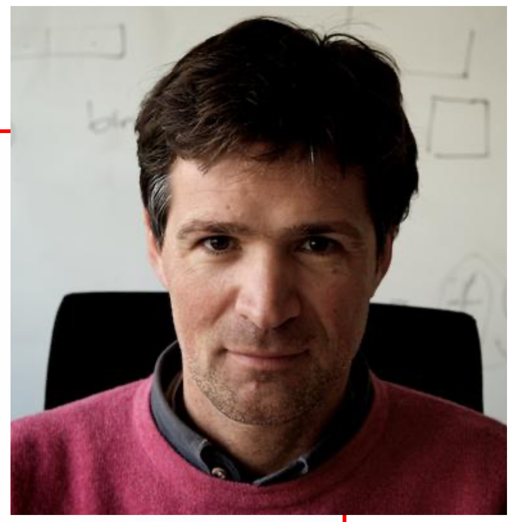
Progress, Challenges, and Future of Statistics for the LHC



Kyle Cranmer (BNL)

Kyle Cranmer (BNL) PhyStat 2007, CERN, June 26, 2007 1

ROOT Statistical Software



Lorenzo Moneta (CERN, PH-SFT)

on behalf of the ROOT Math Work Package

(R. Brun, A. Kreshuk, E. Offermann + many others contributors)

Wouter Verkerke

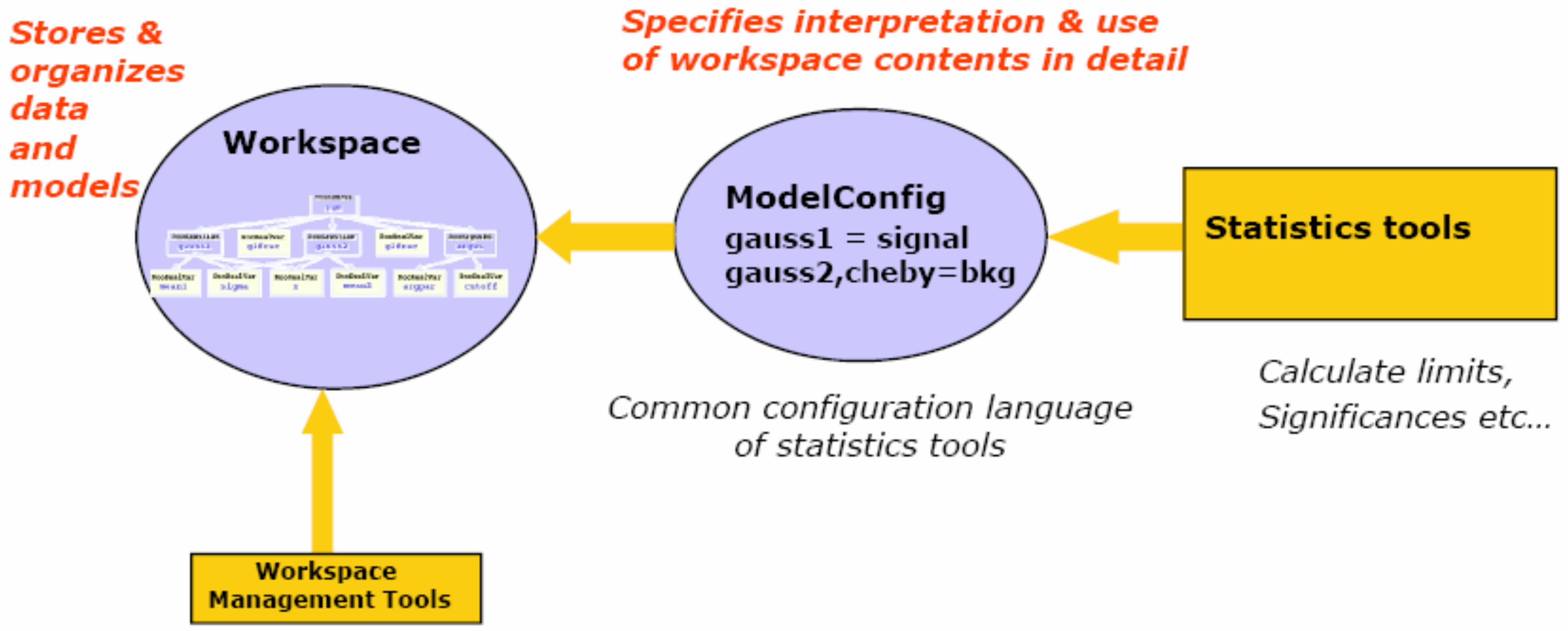
Statistics software for the LHC

The Workspace as publication

- Now have functional RooWorkspace class that can contain
  - Probability density functions and its components
  - (Multiple) Datasets
  - Supporting interpretation information (RooModelConfig)
  - Can be stored in file with regular ROOT persistence
- **Ultimate publication of analysis...**
  - Full likelihood available for Bayesian analysis
  - Probability density function available for Frequentist analysis
  - Information can be easily extracted, combined etc...
  - Common format for sharing, combining of various physics results

Framework design & RooFit adaptations

- Have had more meetings last 3 months to review RooFit lessons from BaBar
  - Kyle, Amir Farbin (ex-Babar), Frank Wrinklmeyer (ex-Babar), WV
  - Design for *WorkSpace* and *ModelConfig* concept in RooFit to interface with statistics tools



ROOT Team Meeting

Friday 18 Apr 2008, 11:00 → 13:00 Europe/Zurich

32/1-A24 (CERN)

11:00 → 11:05 **News**

-New schema Evolution document will be ready early next week (Lukasz)
-Bertrand poster
-Document by Axel/Philippe on new multi-threaded CINT C++ design
-Progress report next meeting
-David/Lorenzo,
-Axel
-Olivier: progress with TGraph restructure

11:05 → 11:30 **SVN restructuring status**

-still to be done
-/doc and ReleaNotes.html in each top level dir
-$ROOTSYS/doc pointing to the top chapters and release notes

-misc/table
/minicern (still to be ported on Windows and Solaris)
removal of references to cernlib/shift from configure

11:30 → 11:50 **5.19/04 dev release scheduled for May 7**

-expect new roofit/roostats package from Kyle Cranmer and Wouter Verkerke
-material in branches must be moved before middle of next week
-developments by Ilka/Roj will be introduced after the release
-fixes to get "make static" and the code checker working again
-THtml must be modified to support the new dir structure (urgent)

Bob Cousins Slide: http://indico.cern.ch/conferenceDisplay.py?confId=100458

Not just the likelihood, the full model

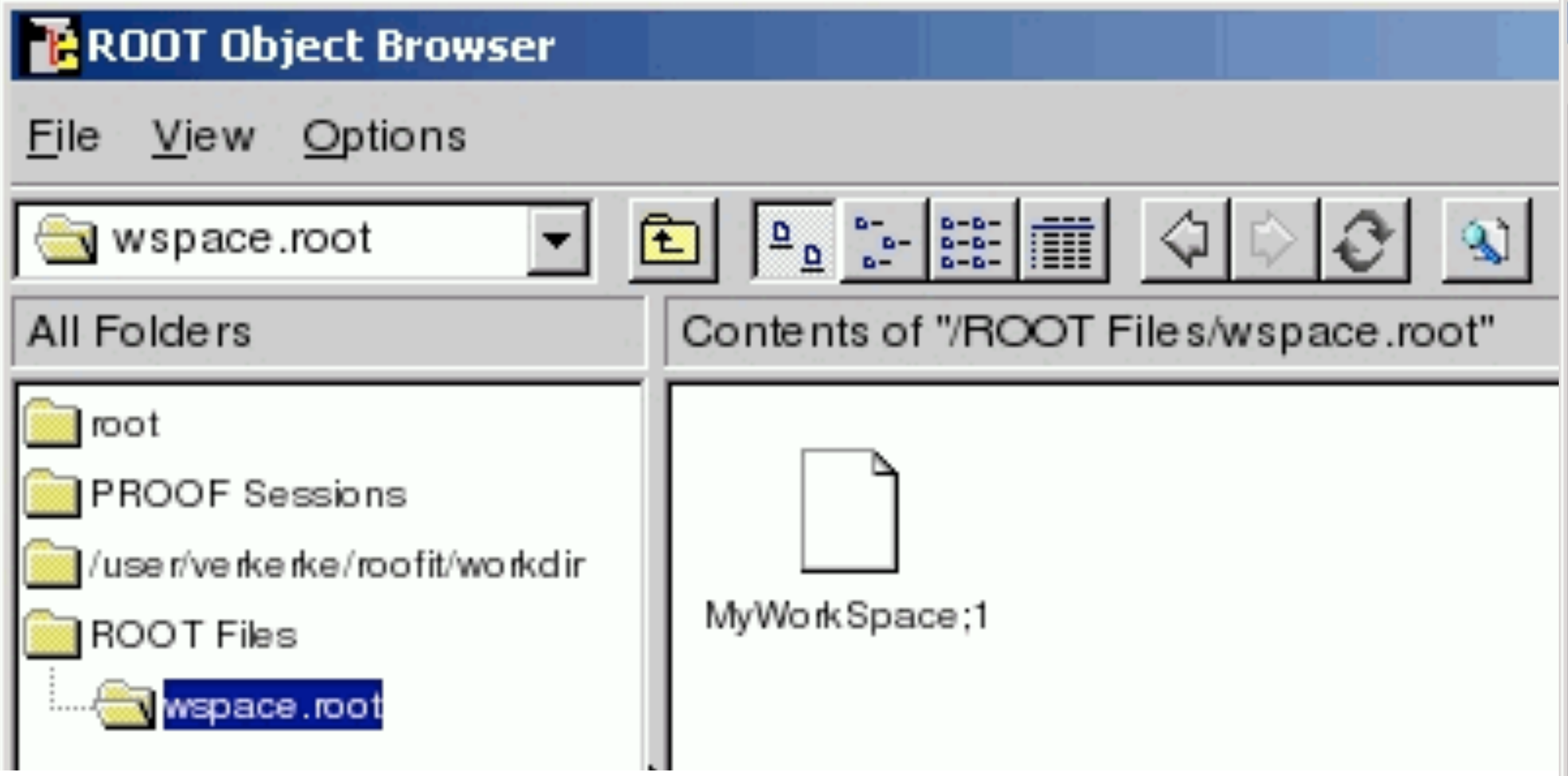
Full statistical model

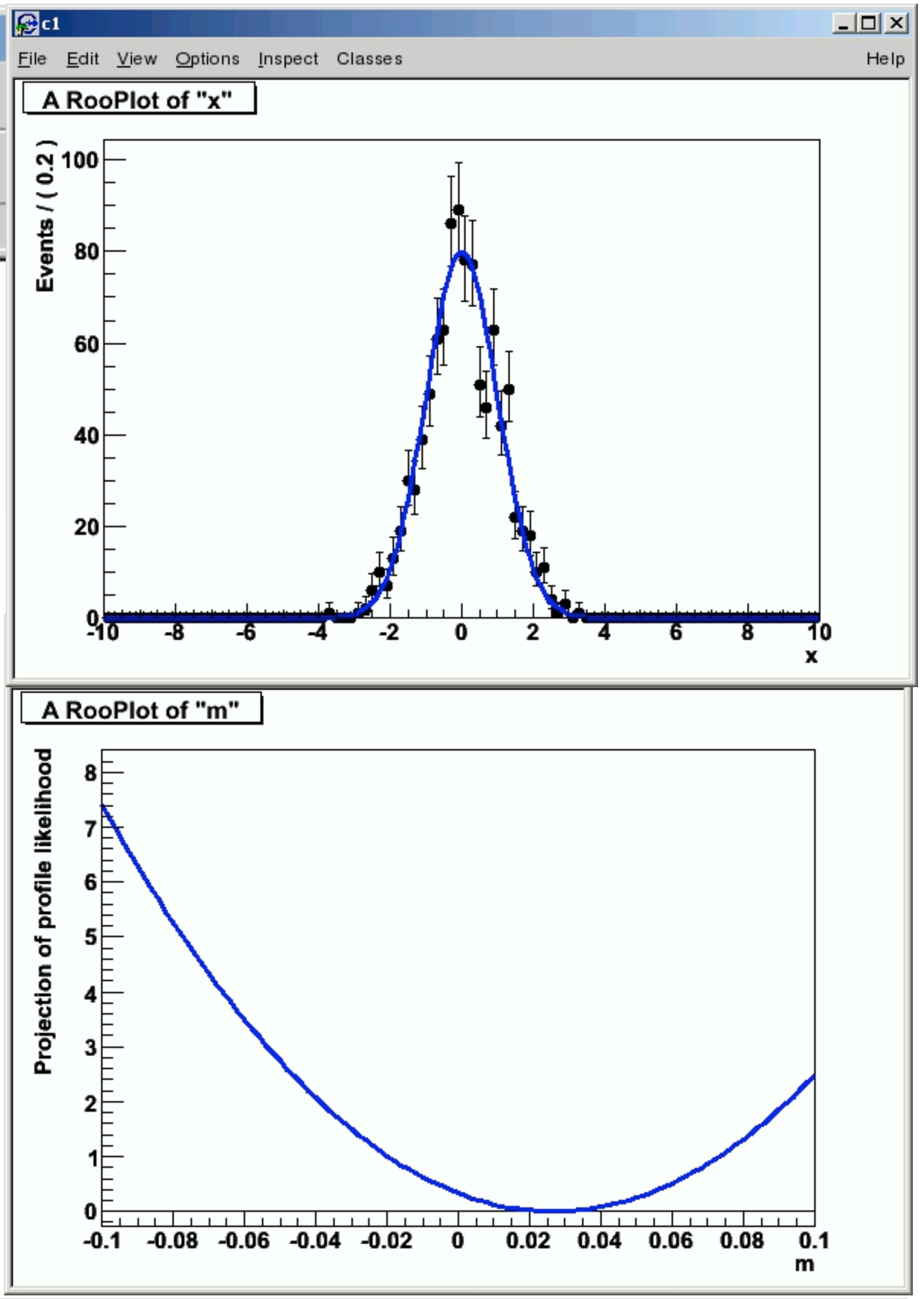
$p(X|\theta)$ →

can generate toys

Example of Digital Publishing

BROOKHAVEN NATIONAL LABORATORY



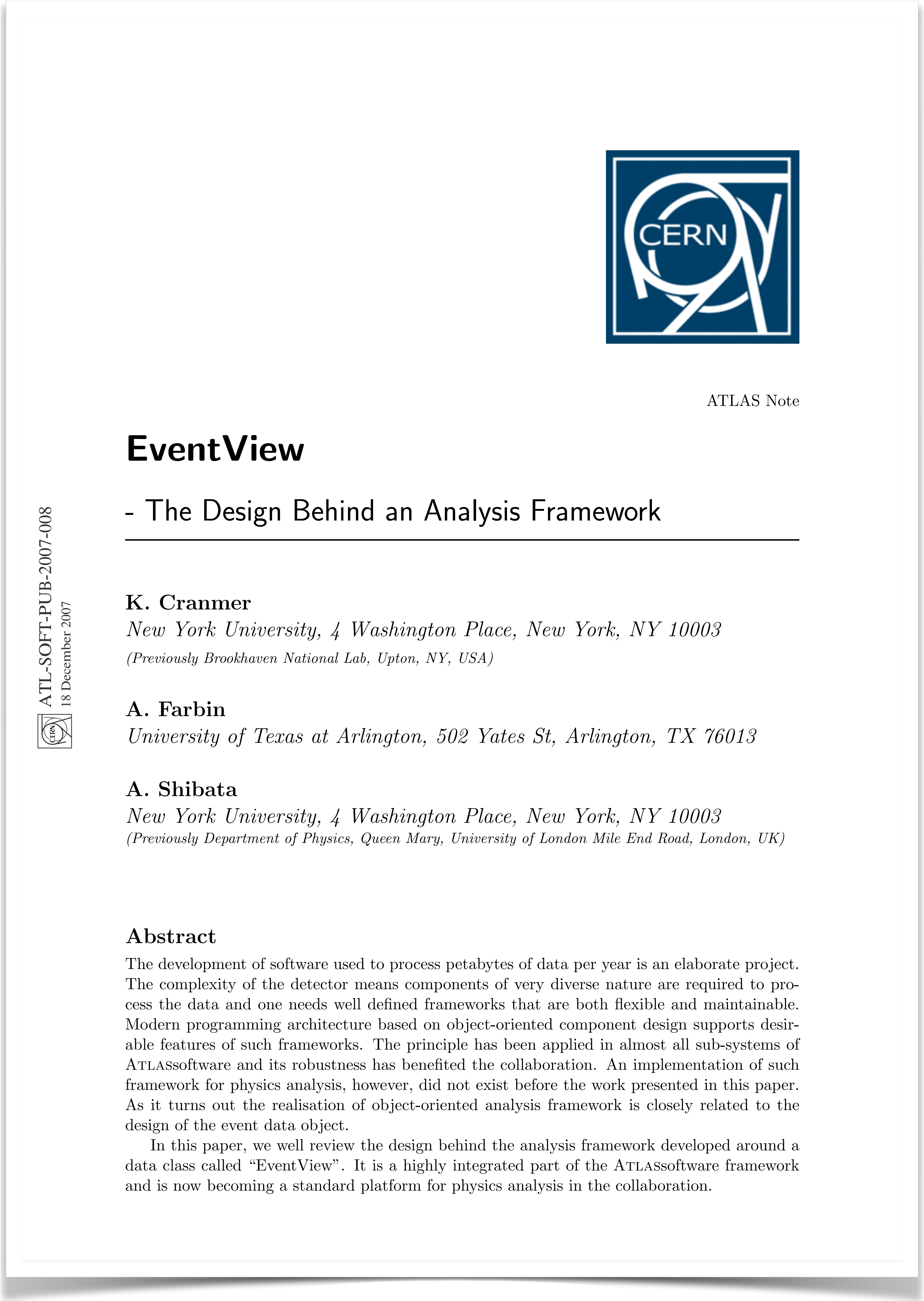


Wouter recently demonstrated the ability to save the function $L(x|\theta_r, \theta_s)$ in a Root file with minimal data necessary to reproduce likelihood function.

Can also evaluate integrals over x necessary for Neyman construction!

Need this for combinations, we should publish them to some repository!

EventView: A learning lesson



ATLAS Note

EventView

- The Design Behind an Analysis Framework

ATL-SOFT-PUB-2007-008
18 December 2007

K. Cranmer
New York University, 4 Washington Place, New York, NY 10003
(Previously Brookhaven National Lab, Upton, NY, USA)

A. Farbin
University of Texas at Arlington, 502 Yates St, Arlington, TX 76013

A. Shibata
New York University, 4 Washington Place, New York, NY 10003
(Previously Department of Physics, Queen Mary, University of London Mile End Road, London, UK)

Abstract

The development of software used to process petabytes of data per year is an elaborate project. The complexity of the detector means components of very diverse nature are required to process the data and one needs well defined frameworks that are both flexible and maintainable. Modern programming architecture based on object-oriented component design supports desirable features of such frameworks. The principle has been applied in almost all sub-systems of ATLASsoftware and its robustness has benefited the collaboration. An implementation of such framework for physics analysis, however, did not exist before the work presented in this paper. As it turns out the realisation of object-oriented analysis framework is closely related to the design of the event data object.

In this paper, we well review the design behind the analysis framework developed around a data class called "EventView". It is a highly integrated part of the ATLASsoftware framework and is now becoming a standard platform for physics analysis in the collaboration.

Shortly before the LHC started, EventView was the leading Analysis Framework within ATLAS

- I was just starting as a professor at NYU

There was serious resistance from some for any sort of "framework," about 1/3 - 1/2 of the collaboration wanted to roll their own

- EventView was abandoned for vaporware right before data taking 🤦‍♂️

Personally it was a terrible experience, but also a learning lesson

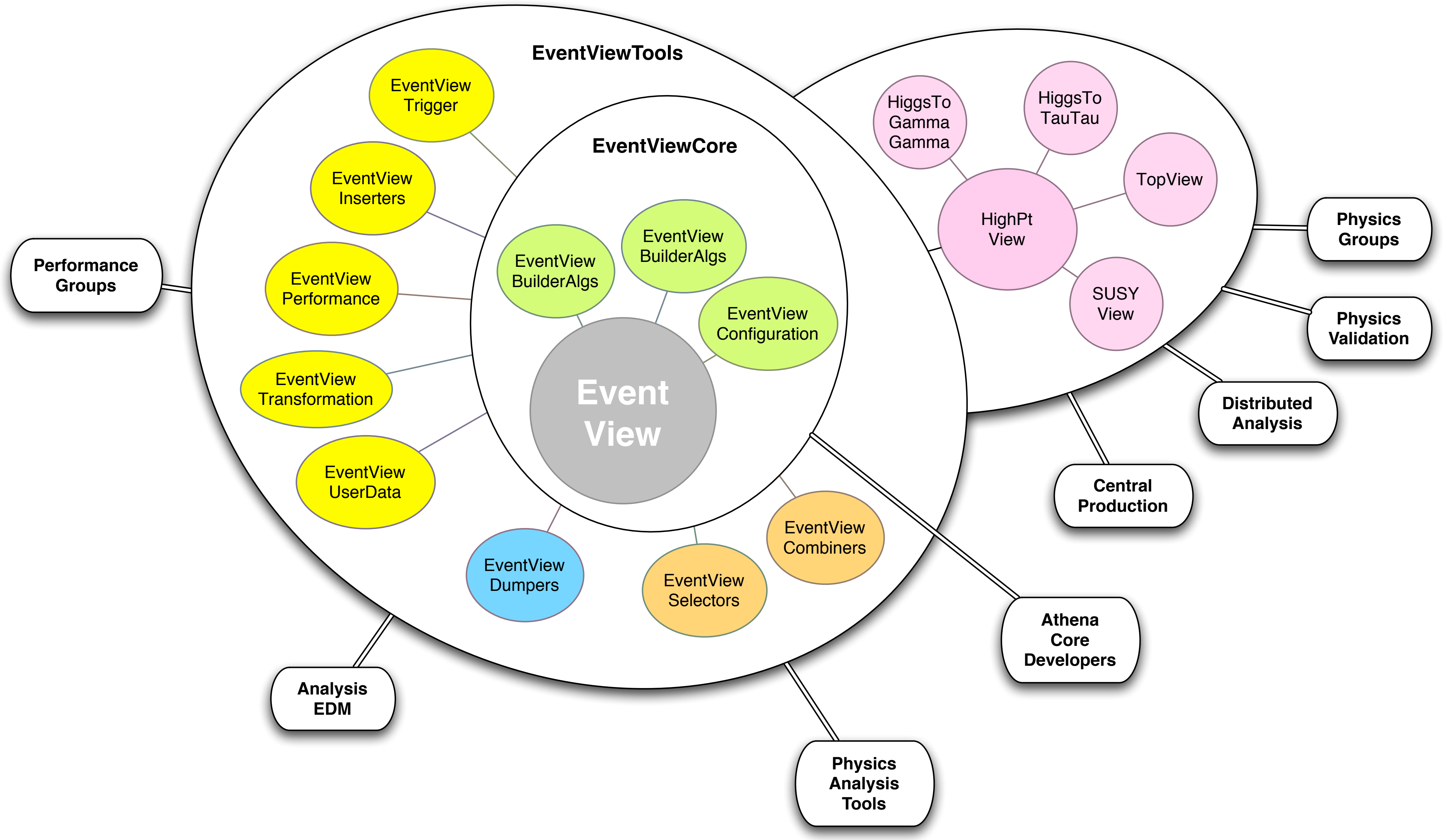


Figure 16: Package organisation of EventView and relationship with working groups.

The birth of HistFactory (2008 - 2009)

The philosophy for the initial design of HistFactory was largely in response to the experience with EventView

- Avoid a framework that does everything

Minimal assumptions on users:

- A distributed group of people using any tool they want come to the table with histograms stored in ROOT files
- Similar to KEYS design
- Use a configuration file with clear semantics

Later recognized this as embracing a **declarative specification.**

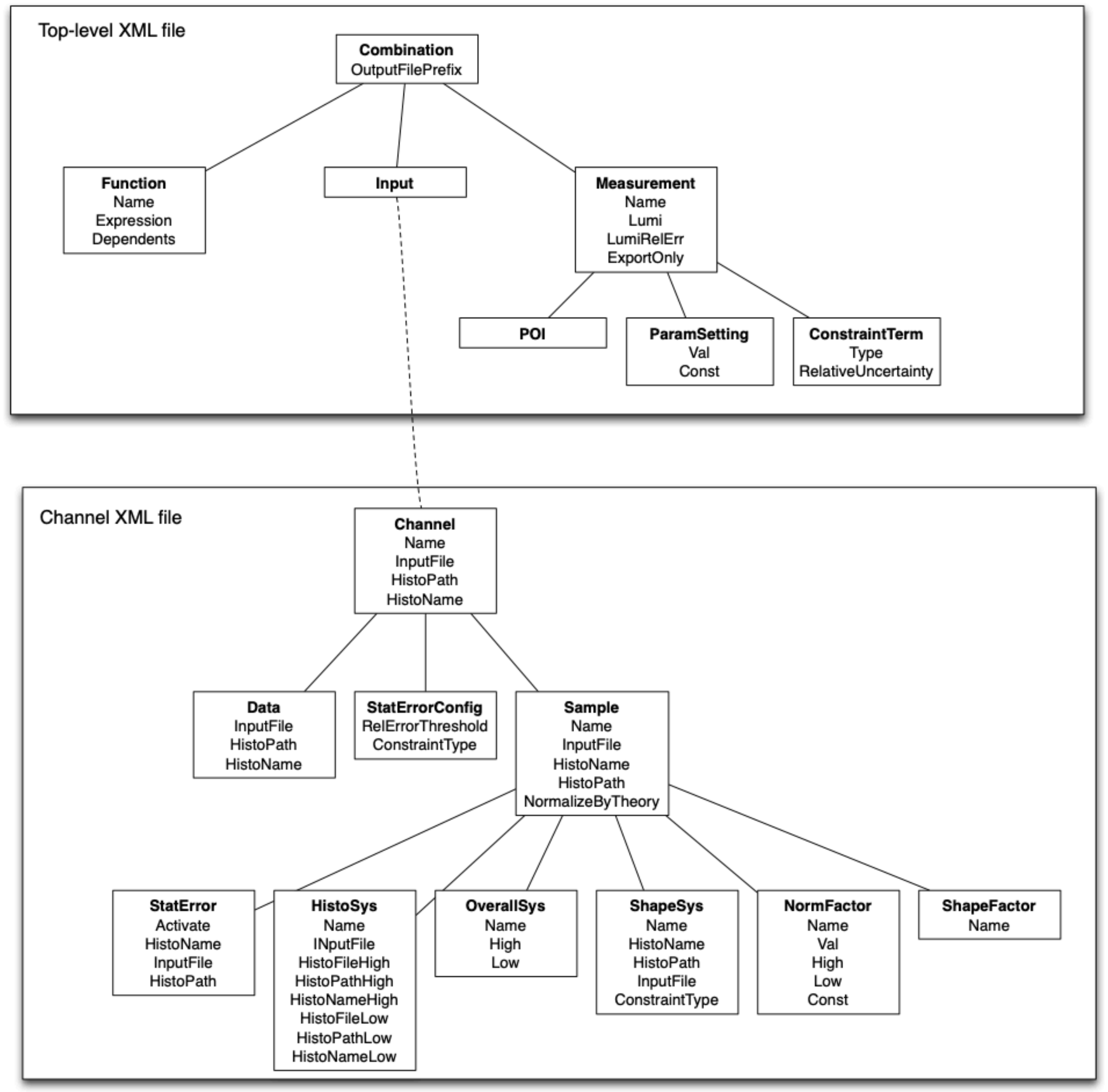
HistFactory

CENTER FOR COSMOLOGY AND PARTICLE PHYSICS

HistFactory tool that ships with ROOT targeting binned analyses

- XML files organize the histograms
- conventions define model exactly
  - CERN-OPEN-2012-016
- command line tool creates likelihood

$$\mathbf{f}_{\rm tot}(\mathcal{D}_{\rm sim},\mathcal{G}|\boldsymbol{\alpha}) = \prod_{c\in{\rm channels}}\left[{\rm Pois}(n_c|\nu_c(\boldsymbol{\alpha}))\prod_{e=1}^{n_c} f_c(x_{ce}|\boldsymbol{\alpha})\right]\cdot\prod_{p\in\mathbb{S}} f_p(a_p|\alpha_p)$$




```
<!DOCTYPE Channel  SYSTEM 'HistFactorySchema.dtd'>

  <Channel Name="A" InputFile="./data/ABCD.root"  >
    <Data HistoName="A_data" HistoPath="" />

    <!-- This is the signal (eg. mu)-->
    <Sample Name="A_signal" HistoPath="" HistoName="unit_histogram">
      <!-- now mu is number of events-->
      <NormFactor Name="mu" Val="1" Low="0" High="200" />
      <OverallSys Name="syst1" High="1.01" Low="0.99" />
    </Sample>

    <!-- This bkg is estimated from MC (eg. mu_A^K) -->
    <Sample Name="A_backgroundMC" HistoPath="" NormalizeByTheory="True" HistoName="unit_histogram"  >
      <NormFactor Name="mu_K_A" Val="100" Low="0" High="200" />
    </Sample>

    <!-- Background 2 is completely Data-Driven -->
    <Sample Name="A_backgroundDD" HistoPath="" NormalizeByTheory="False" HistoName="unit_histogram" >
      <NormFactor Name="mu_D_U" Val="100" Low="24500" High="26000" />
      <NormFactor Name="etaB" Val="1" Low="0." High="0.02" Const="False" />
      <NormFactor Name="etaC" Val="1" Low="0." High="0.3" Const="False" />
      <!-- NormFactor and ShapeFactor same for a 1-bin histogram. But we can name NormFactor-->
    </Sample>

  </Channel>
```

Kyle Cranmer (NYU) Tools, Stockholm, June 18, 2012 10

Akira (my first postdoc) & the t-tbar cross-section

ATLAS NOTE

August 25, 2009

ATL-PHYS-PUB-2009-086
25/08/2009

Prospects for measuring top pair production in the dilepton channel with early ATLAS data at $\sqrt{s} = 10$ TeV

To incorporate uncertainty on the luminosity, the likelihood function is extended to

$$L(\sigma_{sig}, \mathcal{L}) = Pois(N^{obs}|N^{exp}_{tot}) \times Gaus(\tilde{\mathcal{L}}|\mathcal{L}, \Delta_{\mathcal{L}}), \quad (9)$$

where $\mathcal{L}$ is interpreted as the true, unknown integrated luminosity, $\tilde{\mathcal{L}}$, is the nominal estimate of the luminosity, and $\Delta_{\mathcal{L}}$ is the uncertainty on that estimate.

Similarly, the likelihood function can be extended to incorporate uncertainties on the efficiencies ε_{jk}. This is achieved in two steps. First, we group the sources of systematics, α, such that the corresponding variations in ε are expected to be uncorrelated. Next, we vary the sources of the systematics (e. g. jet energy scale, trigger efficiencies, etc.) by the $\pm 1\sigma$ variations and determine the $\varepsilon_{jk}(\alpha_j^{\pm})$ for each signal and background. A change in the source of the j^{th} systematic will cause a totally correlated variation among the contributions (indexed by k). Thus we describe the efficiency as a piece-wise linear function $\varepsilon_{jk}(\alpha_j)$, and this parametrized efficiency is used to in place of the nominal efficiency. Additional Gaussian terms are added to the likelihood function to represent the constraints on the α_j derived from auxiliary measurements or our assumptions about the uncertainty in the Monte Carlo modeling. This modeling of the likelihood function is repeated for each of the channels, and the total likelihood function is simply the product of the individual likelihood functions. Several of the α_j are shared between the channels (for example, the jet energy scale uncertainty), which explicitly introduces correlations between the channels. Thus, we arrive at the final likelihood function:

$$L(\sigma_{sig}, \mathcal{L}, \alpha_j) = \prod_{l \in \{ee,\mu\mu,e\mu\}} \left\{ \prod_{i \in bins} \left[Pois(N_i^{obs}|N_{i,tot}^{exp})\, Gaus(\tilde{\mathcal{L}}|\mathcal{L}, \sigma_{\mathcal{L}}) \prod_{j \in syst} Gaus(0|\alpha_j, 1) \right] \right\}. \quad (10)$$

4.2 Extracting Measurements from the Profile Likelihood Ratio

The likelihood function can be maximized to determine the maximum likelihood estimate of all the parameters $\hat{\sigma}_{sig}, \hat{\mathcal{L}}, \hat{\alpha}_j$. We then consider the likelihood ratio

$$r(\sigma_{sig}) = \frac{L(\sigma_{sig}, \hat{\hat{\mathcal{L}}}, \hat{\hat{\alpha}}_j)}{L(\hat{\sigma}_{sig}, \hat{\mathcal{L}}, \hat{\alpha}_j)} \quad (11)$$

1) Here we include geometrical acceptances among the ε_{jk}.

http://cdsweb.cern.ch/record/1200287

Akira Shibata (He/Him) · 1st
Head of Japan/Korea at W&B - Data x AI Entrepreneur

https://www.linkedin.com/feed/update/urn:li:activity:7136930227594698753/

Akira (my first postdoc) & the t-tbar cross-section

ATLAS NOTE

August 25, 2009

ATL-PHYS-PUB-2009-086
25/08/2009

Prospects for measuring top pair production in the dilepton channel with early ATLAS data at $\sqrt{s} = 10$ TeV

To incorporate uncertainty on the luminosity, the likelihood function is extended to

$$L(\sigma_{sig}, \mathcal{L}) = Pois(N^{obs}|N^{exp}_{tot}) \times Gaus(\tilde{\mathcal{L}}|\mathcal{L}, \Delta_{\mathcal{L}}), \quad (9)$$

where $\mathcal{L}$ is interpreted as the true, unknown integrated luminosity, $\tilde{\mathcal{L}}$, is the nominal estimate of the luminosity, and $\Delta_{\mathcal{L}}$ is the uncertainty on that estimate.

Similarly, the likelihood function can be extended to incorporate uncertainties on the efficiencies ε_{jk}. This is achieved in two steps. First, we group the sources of systematics, α, such that the corresponding variations in ε are expected to be uncorrelated. Next, we vary the sources of the systematics (e. g. jet energy scale, trigger efficiencies, etc.) by the $\pm 1\sigma$ variations and determine the $\varepsilon_{jk}(\alpha_j^{\pm})$ for each signal and background. A change in the source of the j^{th} systematic will cause a totally correlated variation among the contributions (indexed by k). Thus we describe the efficiency as a piece-wise linear function $\varepsilon_{jk}(\alpha_j)$, and this parametrized efficiency is used to in place of the nominal efficiency. Additional Gaussian terms are added to the likelihood function to represent the constraints on the α_j derived from auxiliary measurements or our assumptions about the uncertainty in the Monte Carlo modeling. This modeling of the likelihood function is repeated for each of the channels, and the total likelihood function is simply the product of the individual likelihood functions. Several of the α_j are shared between the channels (for example, the jet energy scale uncertainty), which explicitly introduces correlations between the channels. Thus, we arrive at the final likelihood function:

$$L(\sigma_{sig}, \mathcal{L}, \alpha_j) = \prod_{l \in \{ee,\mu\mu,e\mu\}} \left\{ \prod_{i \in bins} \left[Pois(N_i^{obs}|N_{i,tot}^{exp})\, Gaus(\tilde{\mathcal{L}}|\mathcal{L}, \sigma_{\mathcal{L}}) \prod_{j \in syst} Gaus(0|\alpha_j, 1) \right] \right\}. \quad (10)$$

4.2 Extracting Measurements from the Profile Likelihood Ratio

The likelihood function can be maximized to determine the maximum likelihood estimate of all the parameters $\hat{\sigma}_{sig}, \hat{\mathcal{L}}, \hat{\alpha}_j$. We then consider the likelihood ratio

$$r(\sigma_{sig}) = \frac{L(\sigma_{sig}, \hat{\mathcal{L}}, \hat{\alpha}_j)}{L(\hat{\sigma}_{sig}, \hat{\mathcal{L}}, \hat{\alpha}_j)} \quad (11)$$

[1] Here we include geometrical acceptances among the ε_{jk}.

http://cdsweb.cern.ch/record/1200287

Akira Shibata (He/Him) · 1st
Head of Japan/Korea at W&B - Data x AI Entrepreneur

Yesterday! →

https://www.linkedin.com/feed/update/urn:li:activity:7136930227594698753/

Akira Shibata (He/Him) • 1st
Head of Japan/Korea at W&B - Data x AI Entrepreneur
1d •

It was an honor to moderate the Q&A sessions for **Guido van Rossum** the creator of Python language. When I was coding my analysis for the particle physics experiment around 2005, I met Python and fell in love. It actually helped open up my career to have that expertise early in the time. I was glad I was able to communicate my gratitude to Guido-san!

Python言語のクリエーターであるGuidoさんへのQ&Aセッションをモデレートしました。2005年ごろ研究にPythonを使い始め、この素晴らしい言語にハマりました。実際このタイミングでPythonエクスパートになれたことは自分のキャリア形成にも大きな影響がありました、とGuidoさんにお伝えできました。

You and 82 others · 5 comments · 2 reposts

The first paper using HistFactory with 3.1 pb^{-1}

CERN-PH-EP-2010-064
(Submitted to EPJC)

December 8, 2010

Measurement of the top quark-pair production cross section with ATLAS in pp collisions at $\sqrt{s} = 7$ TeV

The ATLAS Collaboration

8 Summary

Measurements of the $t\bar{t}$ production cross-section in the single-lepton and dilepton channels using the ATLAS detector are reported. In a sample of 2.9 pb^{-1}, 37 $t\bar{t}$ candidate events are observed in the single-lepton topology, as well as 9 candidate events in the dilepton topology, resulting in a measurement of the inclusive $t\bar{t}$ cross-section of

$$\sigma_{t\bar{t}} = 145 \pm 31\ ^{+42}_{-27}\ \text{pb}.$$

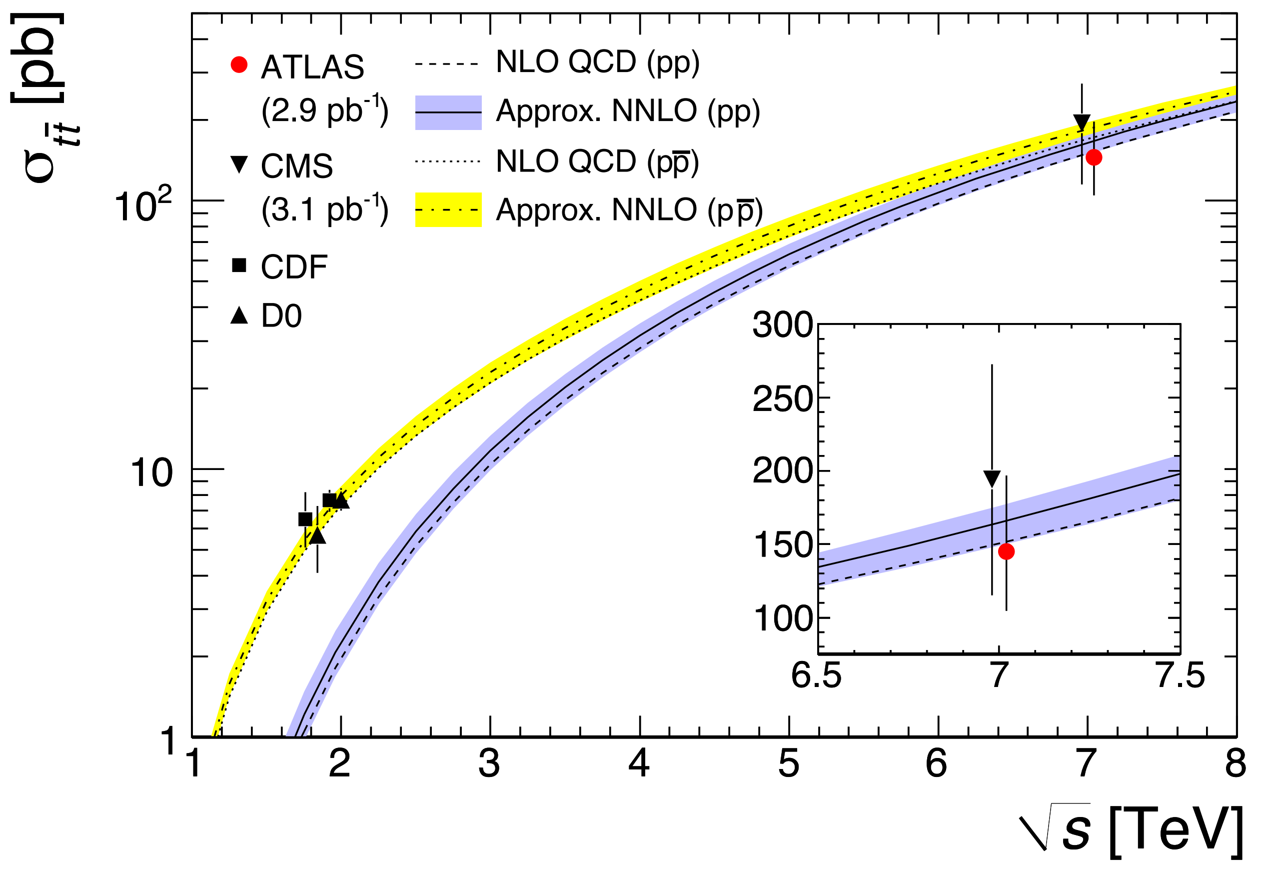
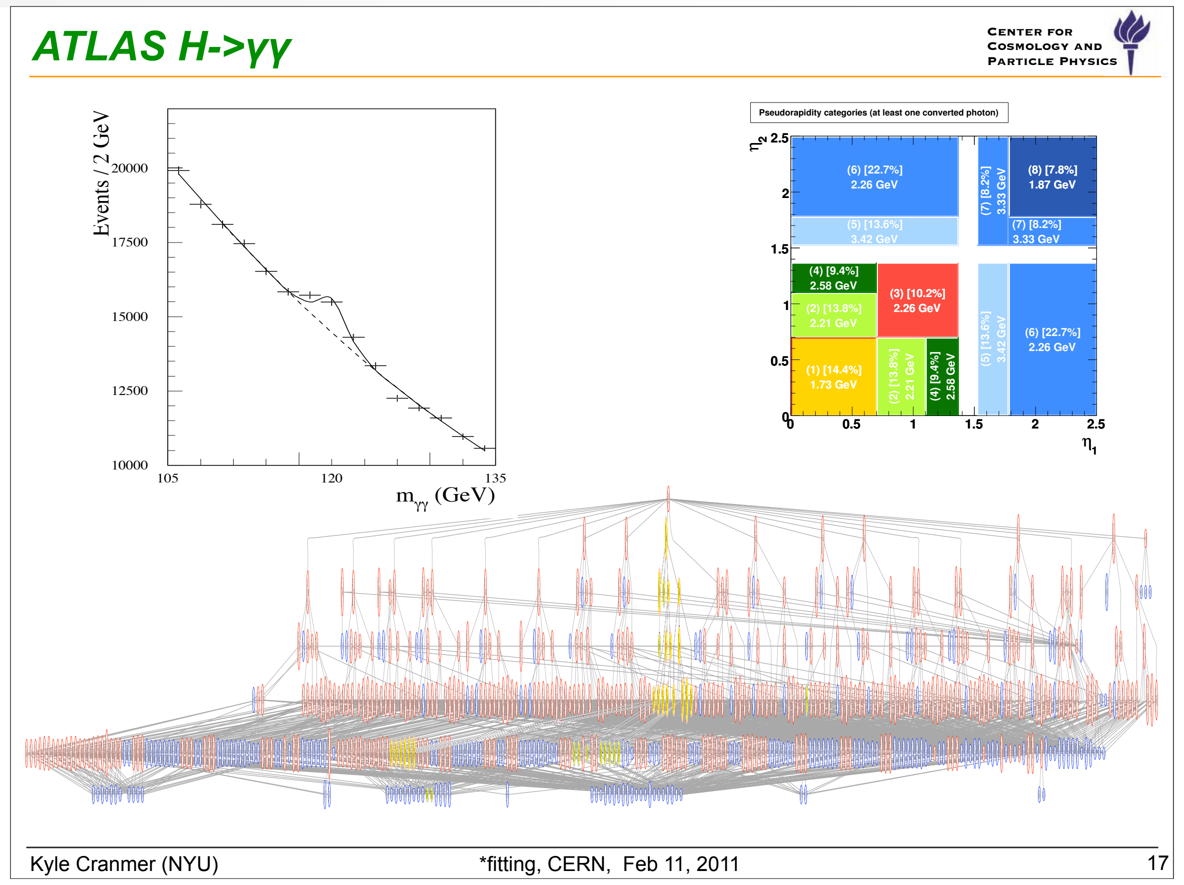


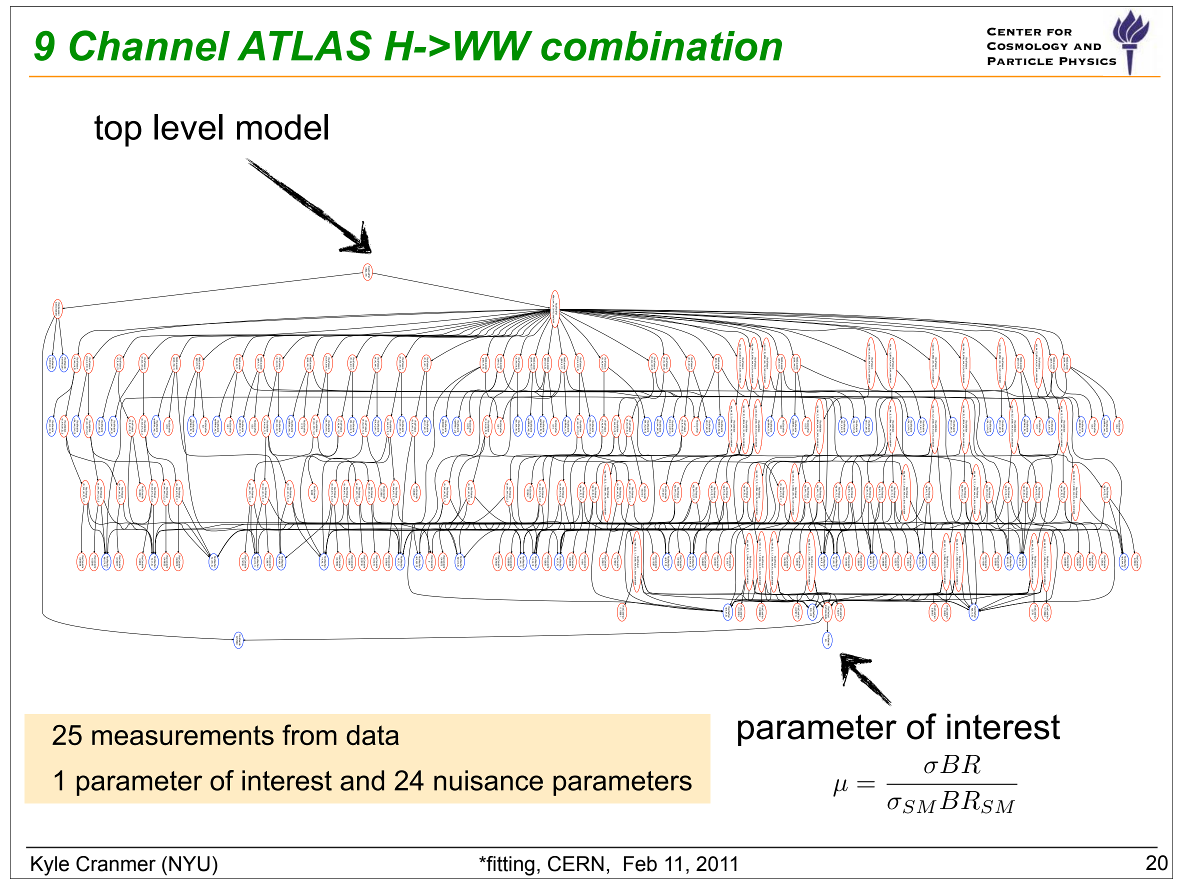
Figure 9: Top quark pair-production cross-section at hadron colliders as measured by CDF and D0 at Tevatron [3], CMS [4] and ATLAS (this measurement). The theoretical predictions for pp and $p\bar{p}$ collisions [33] include the scale and PDF uncertainties, obtained using the HATHOR tool with the CTEQ6.6 PDFs [34] and assume a top-quark mass of 172.5 GeV.

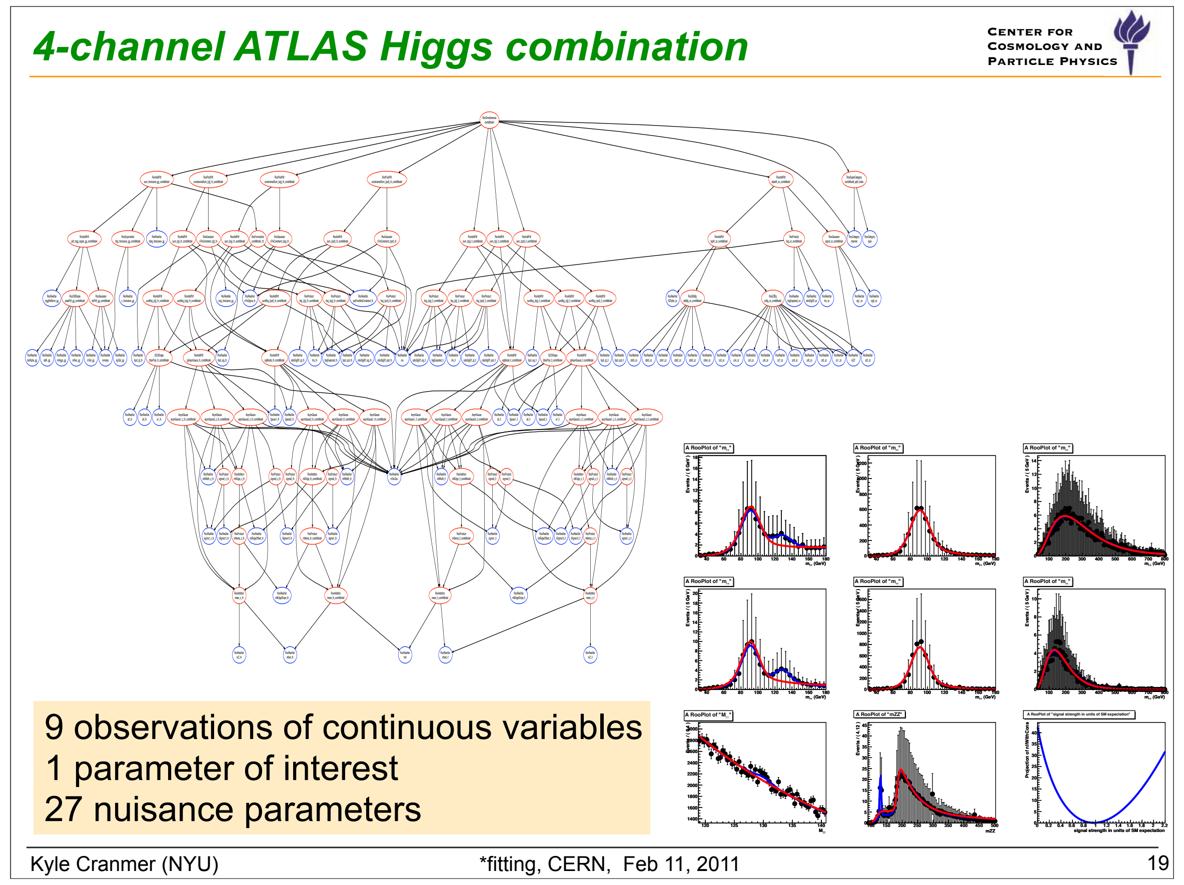
It was not a relaxing Thanksgiving in 2010

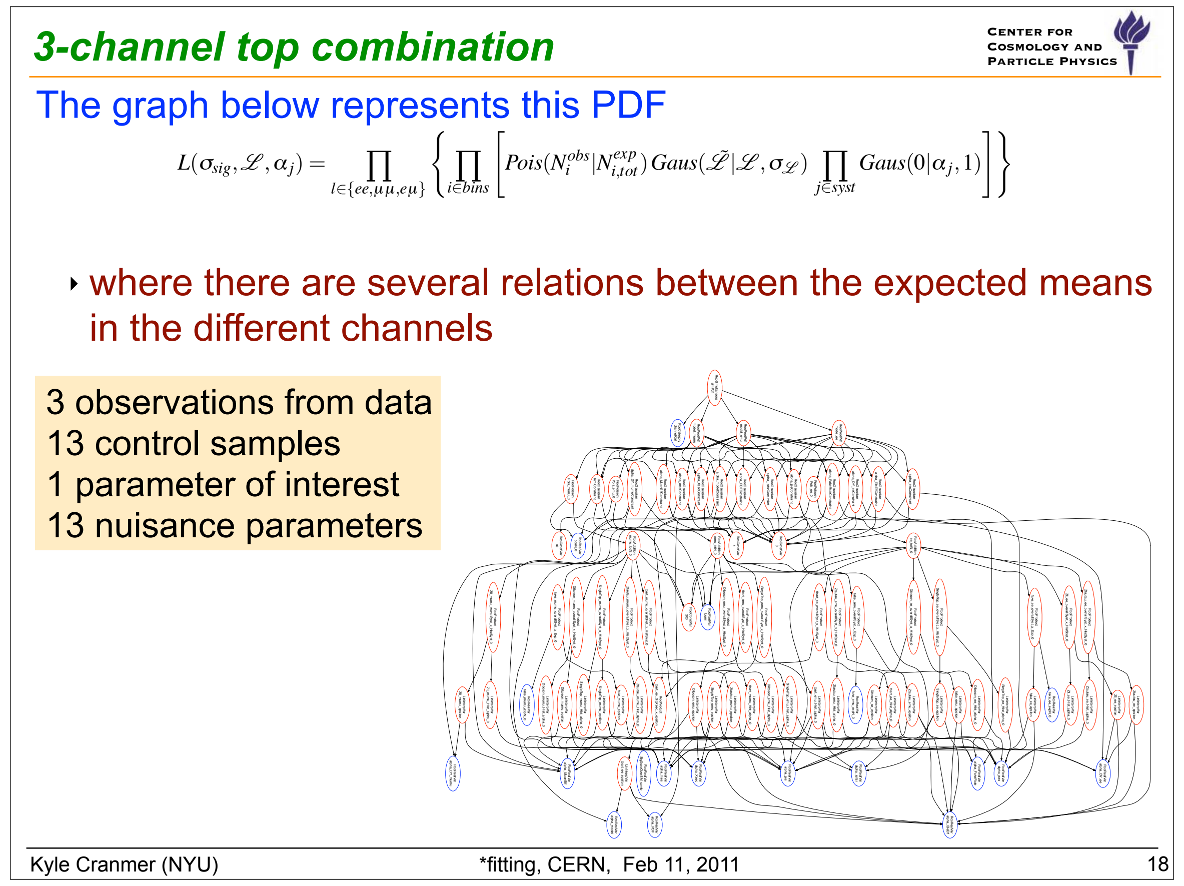
Early examples for Higgs (2011)

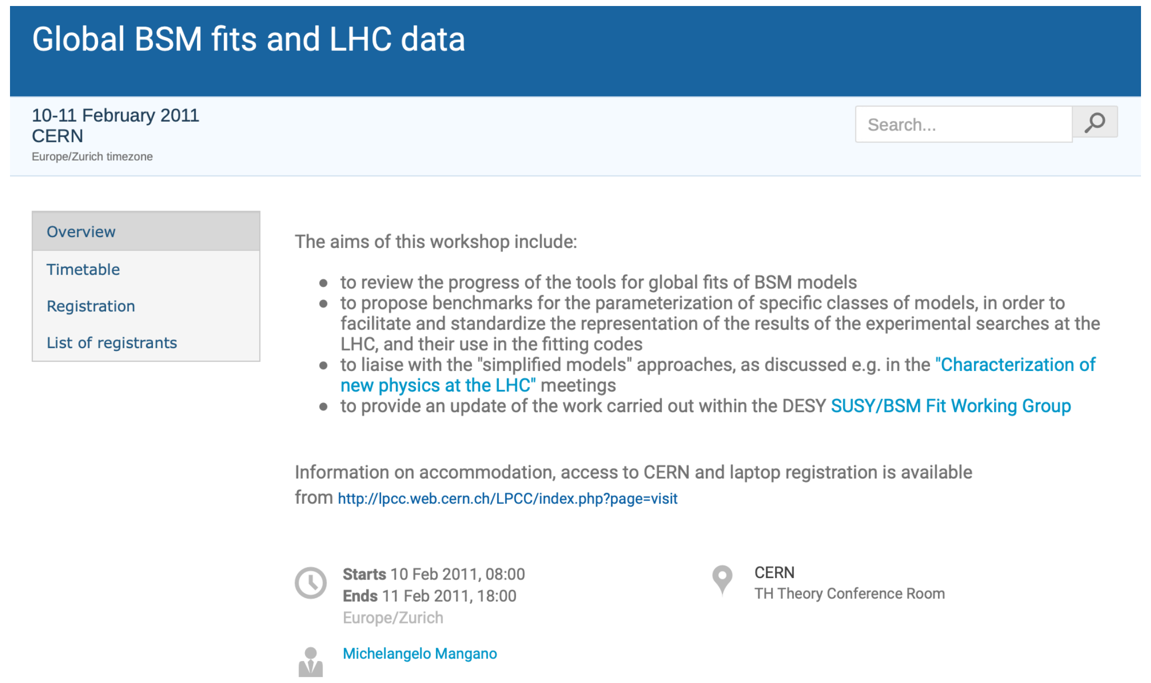
https://indico.cern.ch/event/118137/overview

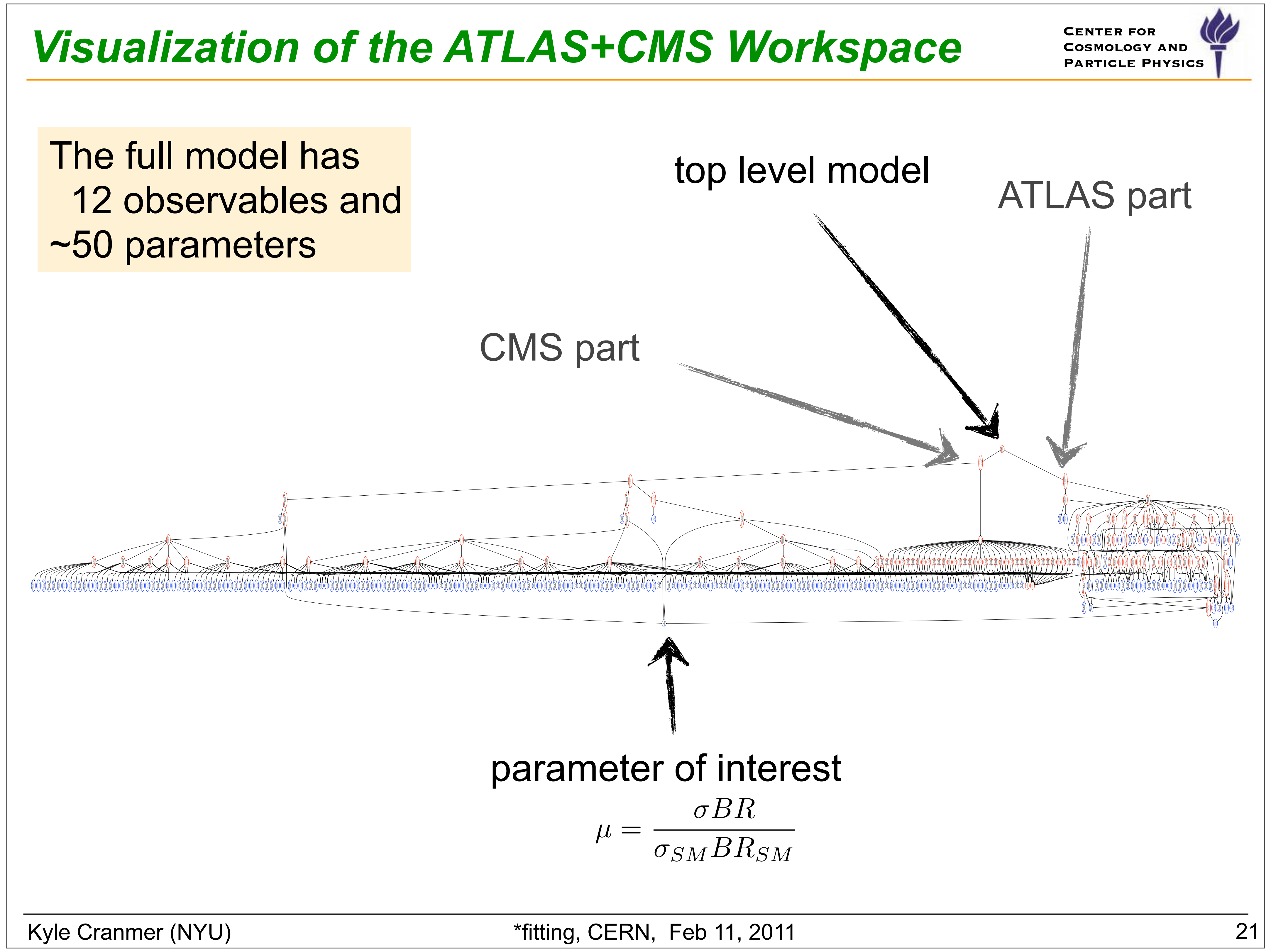




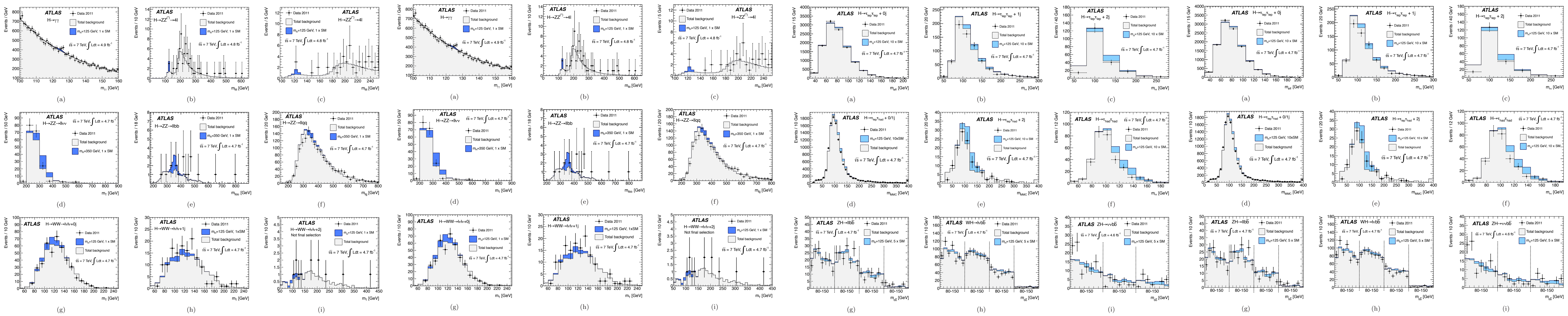


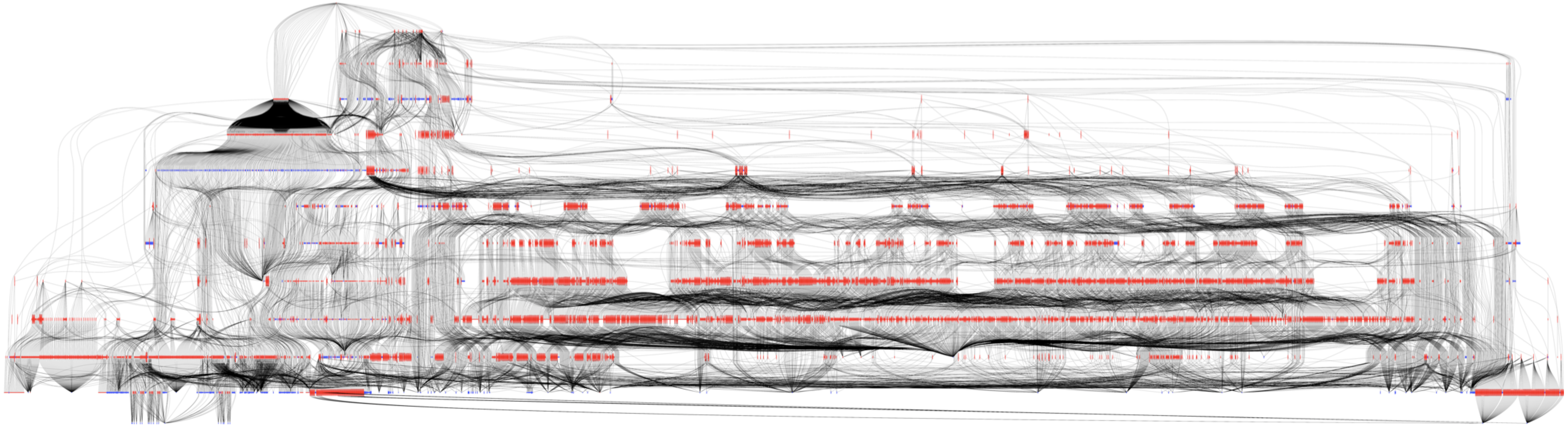






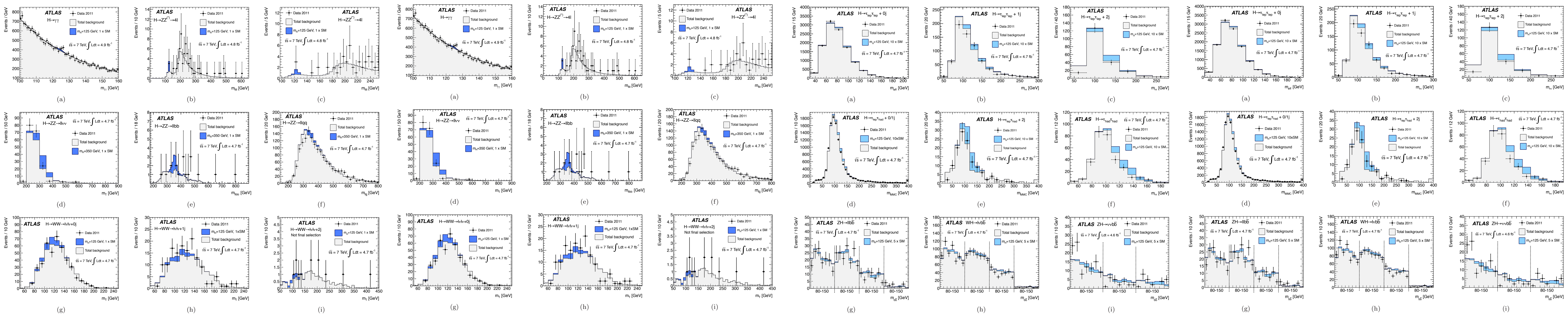
Collaborative Statistical Modeling

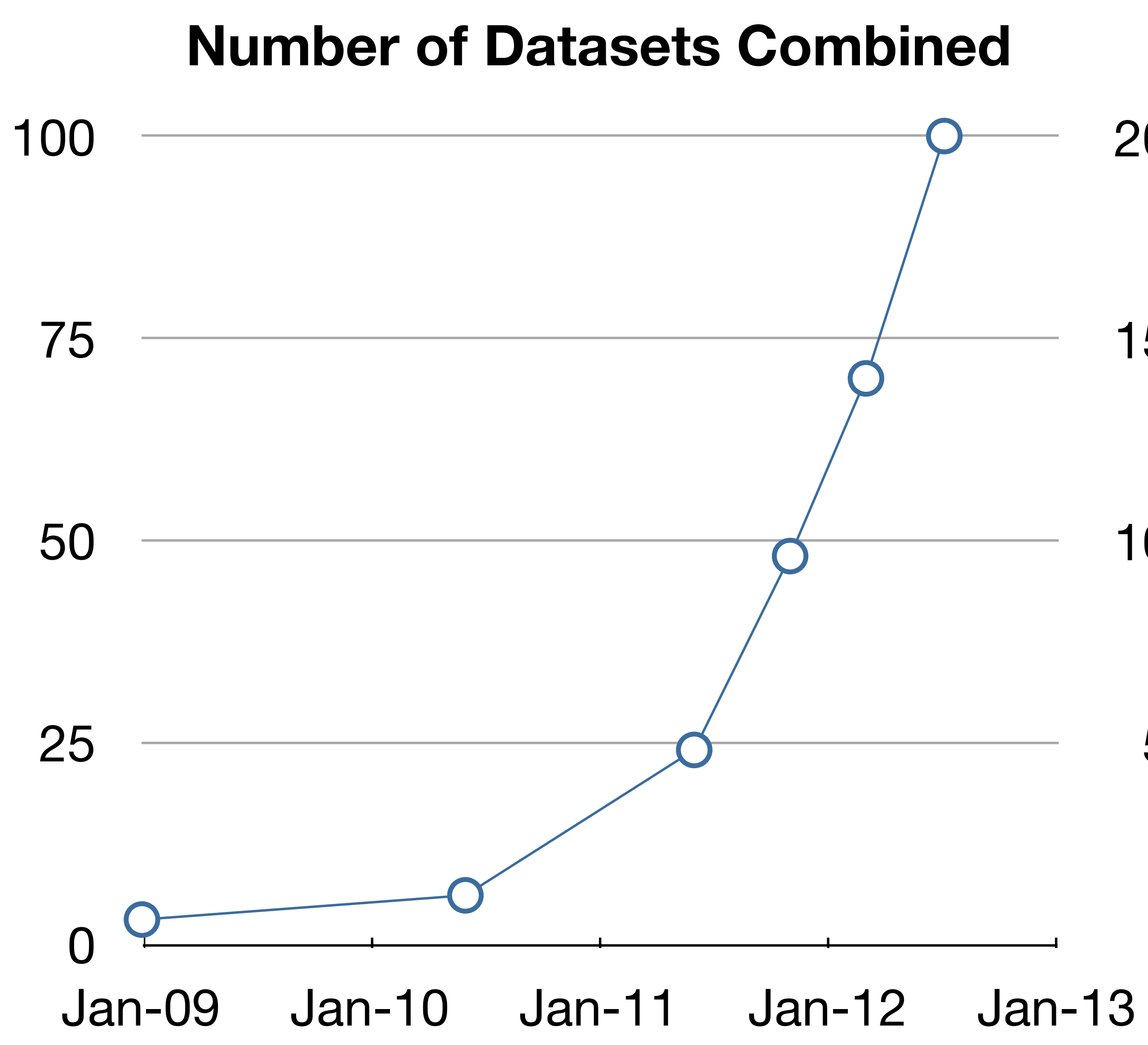


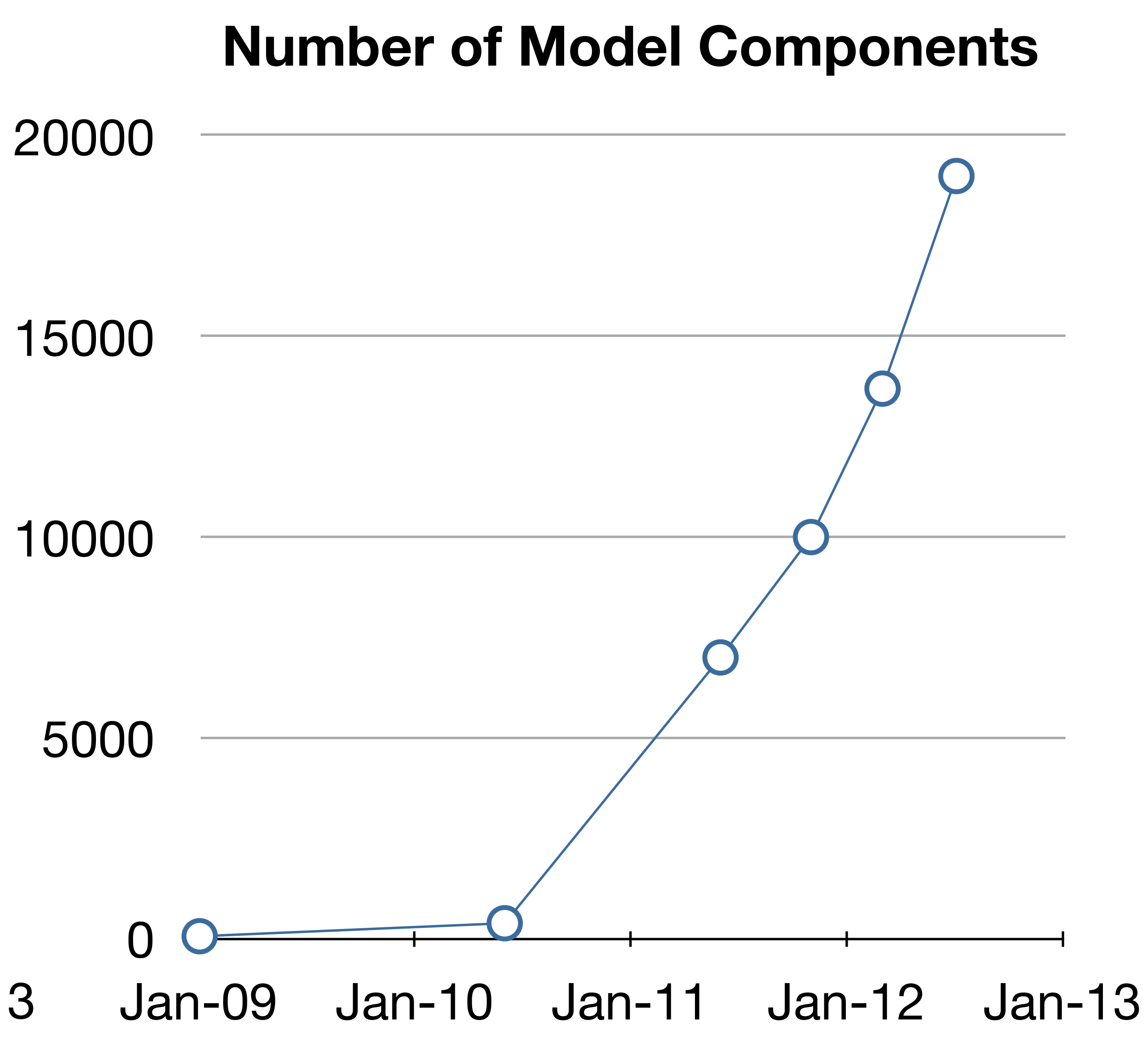


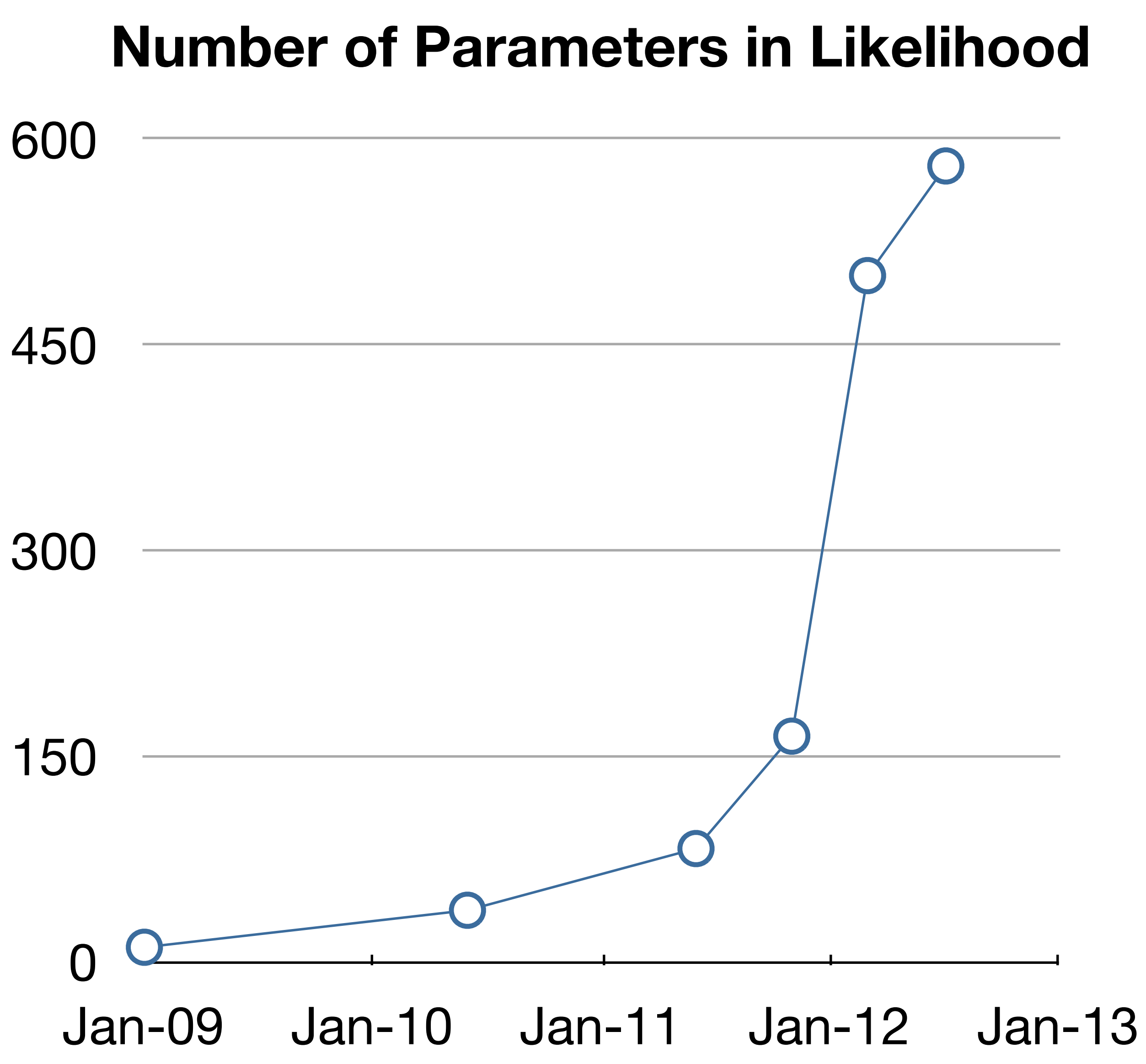
$$\mathbf{f}_{\mathrm{tot}}(\mathcal{D}_{\mathrm{sim}}, \mathcal{G}|\boldsymbol{\alpha}) = \prod_{c\in\mathrm{channels}} \left[\mathrm{Pois}(n_c|\nu_c(\boldsymbol{\alpha})) \prod_{e=1}^{n_c} f_c(x_{ce}|\boldsymbol{\alpha})\right] \cdot \prod_{p\in\mathbb{S}} f_p(a_p|\alpha_p)$$

Collaborative Statistical Modeling









HistFitter origins

While I was still recovering from EventView experience, the RooStats + HistFactory user community did want a more integrated user experience that integrated late-stage analysis, statistical modeling, and statistical inference

- Combine in CMS was a somewhat parallel development

Configurable analysis with HistFitter

Dan Short[1]
On behalf of the HistFitter group

[1]Oxford

12th April 2012

UNIVERSITY OF OXFORD

ATLAS

Introduction

- Attempt to build on HistFactory framework
- Provide inputs from TTrees or TH1s and generate XML
  - Flexibility in terms of selections used
  - Possible to quickly switch systematics and change PDF
  - Optimized for iteration over grids
  - Hierarchical TopLevel → Channel → Sample system
- Can run interactively
  - Displays 'before-and-after' plots
  - Instant access to results and workspaces
- Notion of control, validation and signal regions
  - Plotting functions breakdown PDF into components
  - Ability to project errors from one channel to another
  - Draw ratio plots with correct error visualization
- Ability to run fits and hypothesis tests in one line
- Used in 1-lepton ATLAS SUSY analysis for SEARCH 2012

Autodiff in 2012?

I remember being very excited about automatic differentiation and potential for it in our stats software

- Tools were lacking

First discussions of autodiff in 2012!

Minutes of our April 26 meeting

KC Kyle Cranmer <kyle.cranmer@nyu.edu>
To: **roostats-development (A mailing list for RooStats development. RooStats is a p...)**
Thu 4/26/2012 09:45

RooStats meeting April 26
Lorenzo, Kyle, George, Gabriel, Gregory, Alfio

Candidate release next week May 2

Lorenzo:
log scan for inverter
some patches etc.
tutorials now linked to Twiki page

George:
HistFactory tutorials to be visible on web page
Steamer fix from Kyle
George to migrate HistFactory changes to RooStats dev branch

Gabriel:
moving tests into trunk
Simple likelihood ratio not working with Frequentist calculator
George to send some tests for HistFactory

Kyle:
We will have Google Summer of Code student working on Graphical Models: Shruti Gupta
Automatic Differentiation could be useful for us
http://en.wikipedia.org/wiki/Automatic_differentiation

Alfio:
80% of time dedicated to vectorization of RooFit core
leaving CERN in ~July :-(

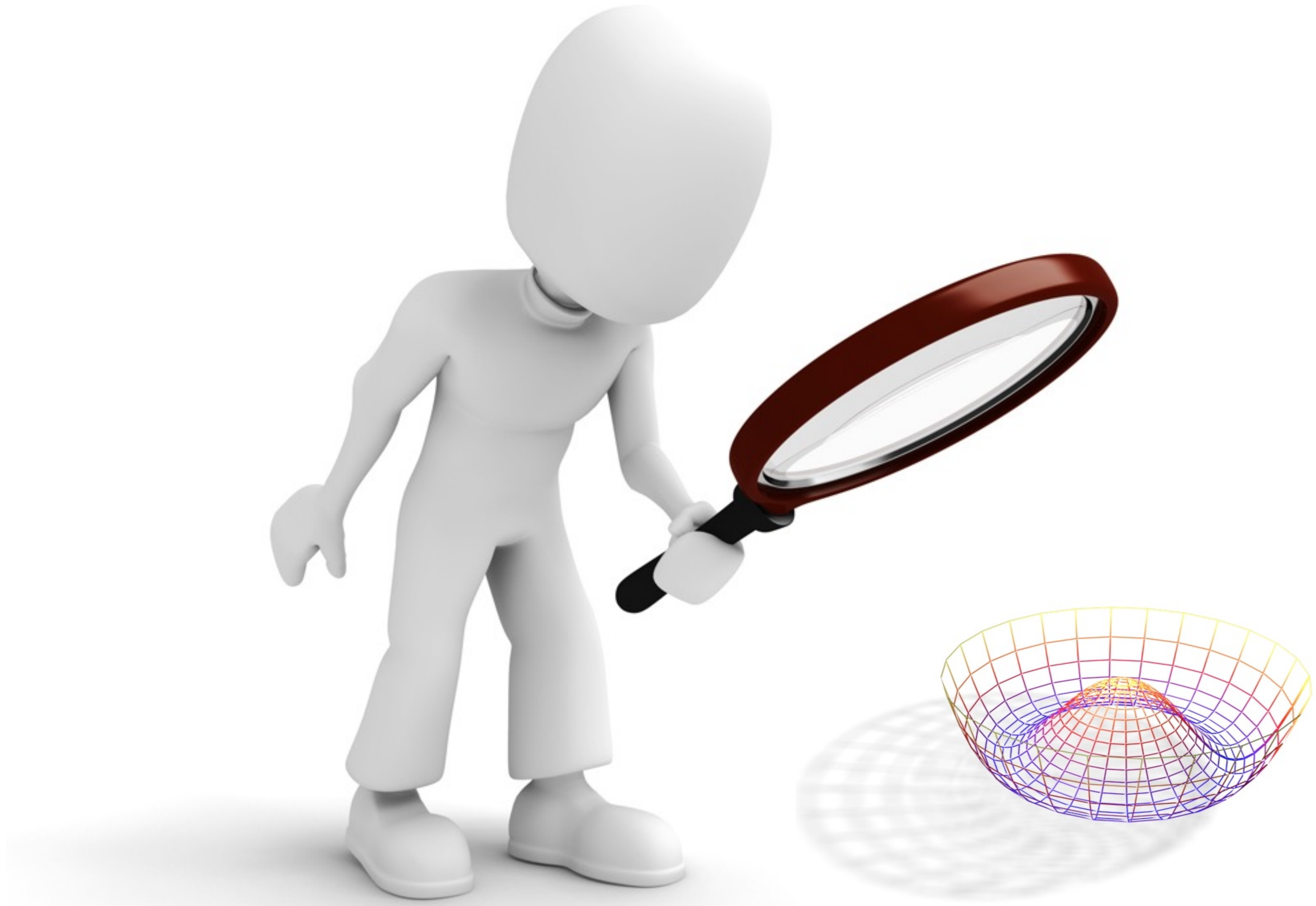
Sven (in Dresden):
--configure multinest in build system
- other stuff, please report

Wouter:
please send some update on RooFit

next meeting: May 10

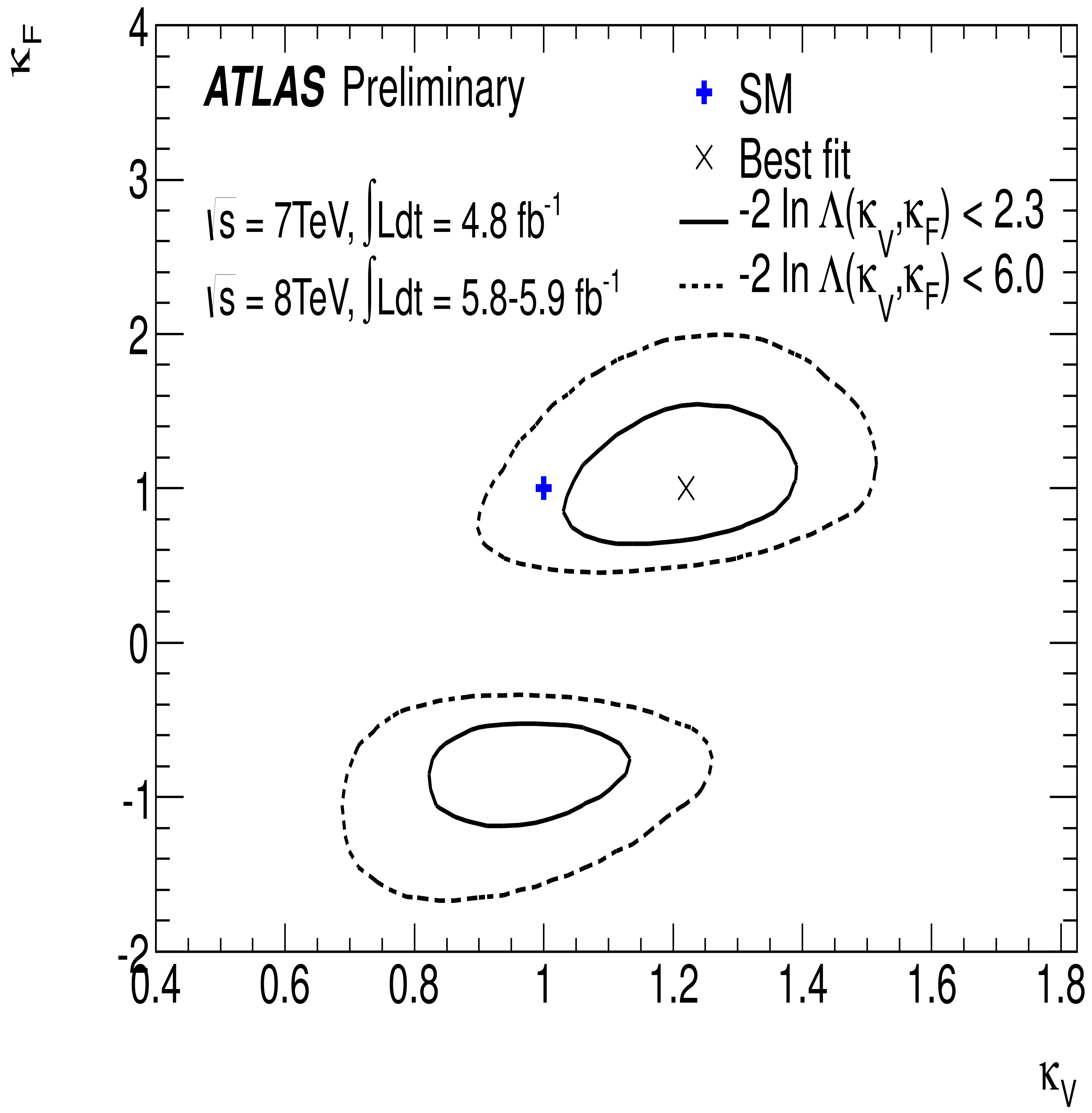
All the best,

Kyle



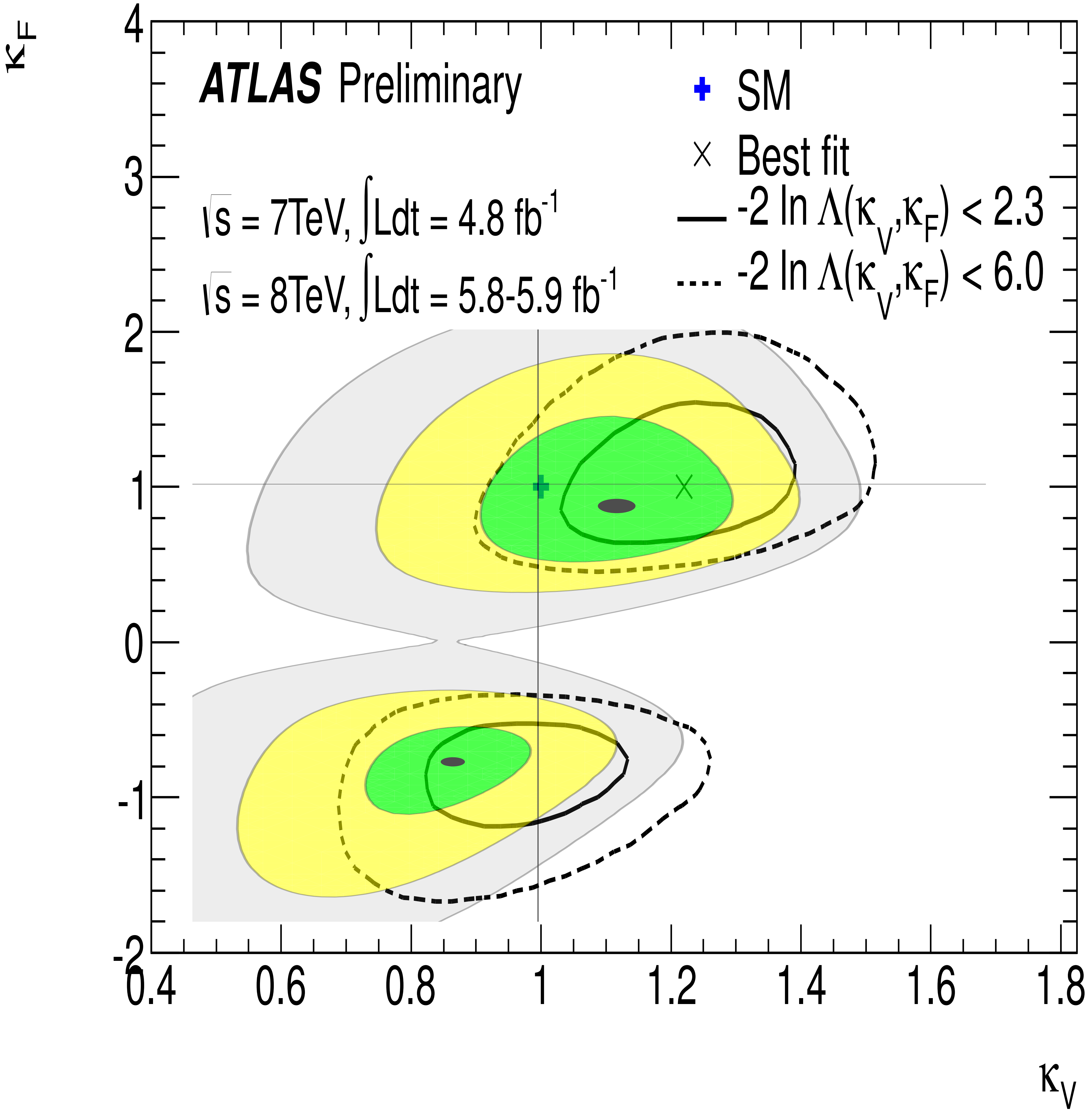
REPRODUCIBILITY PROBLEM

Not possible for others to reproduce results from paper.



REPRODUCIBILITY PROBLEM

Not possible for others to reproduce results from paper.



2012 recommendations

Searches for New Physics: Les Houches Recommendations for the Presentation of LHC Results

S. Kraml[1], B.C. Allanach[2], M. Mangano[3], H.B. Prosper[4], S. Sekmen[3,4] (editors), C. Balazs[5], A. Barr[6], P. Bechtle[7], G. Belanger[8], A. Belyaev[9,10], K. Benslama[11], M. Campanelli[12], K. Cranmer[13], A. De Roeck[3], M.J. Dolan[14], T. Eifert[15], J.R. Ellis[16,3], M. Felcini[17], B. Fuks[18], D. Guadagnoli[8,19], J.F. Gunion[20], S. Heinemeyer[17], J. Hewett[15], A. Ismail[15], M. Kadastik[21], M. Krämer[22], J. Lykken[23] F. Mahmoudi[3,24], S.P. Martin[25,26,27], T. Rizzo[15], T. Robens[28], M. Tytgat[29], A. Weiler[30]

Why public likelihoods

- The statistical model of an experimental analysis provides the complete mathematical description of that analysis

  $p(o|\alpha)$ relating the observed quantities o to the parameters α

- Given the likelihood, all the standard statistical approaches are available for extracting information from it

- Essential information for any detailed interpretation of experimental results

  = determining the compatibility of the observations with theoretical predictions

Les Houches Recommandations (2012)

3b: When feasible, provide a mathematical description of the final likelihood function in which experimental data and parameters are clearly distinguished, either in the publication or the auxiliary information. Limits of validity should always be clearly specified.

3c: Additionally provide a digitized implementation of the likelihood that is consistent with the mathematical description.

arXiv:1203.2489

https://indico.cern.ch/event/957797/contributions/4026032/

2012

Searches for new physics: recommendations for the presentation of LHC results

13 February 2012
CERN
Europe/Zurich timezone

Search...

- Overview
- Timetable
- Registration
- Participant List

During the Les Houches 2011 workshop, discussions started to define a set of recommendations for the presentation of LHC results on searches for new physics, aimed at providing a more efficient flow of scientific information between the experimental collaborations and the rest of the high energy physics community, and facilitating the interpretation of the results in a wide class of models. This discussion evolved into a draft document, available for download from this page. The goal of this meeting is to review this draft and present it to the experiments for discussion and eventual endorsement.
EVO connection will be available: the link will appear on the agenda page 30' before the start of the meeting. **Please register even if you participate only by EVO**

Starts 13 Feb 2012, 09:00
Ends 13 Feb 2012, 19:00
Europe/Zurich

CERN
TH Theory Conference Room

Michelangelo Mangano
Sabine Kraml

Paper
- Final paper, arXiv:1203.2489
- LH proceedings contribution

https://indico.cern.ch/event/173341/

Likelihoods for the LHC Searches

21-23 January 2013
CERN
Europe/Zurich timezone

Search...

- Overview
- Timetable
- Registration
- Participant List

The primary goal of this 3-day workshop is to educate the LHC community about the scientific utility of likelihoods. We shall do so by describing and discussing several real-world examples of the use of likelihoods, including a one-day in-depth examination of likelihoods in the Higgs boson studies by ATLAS and CMS.
The workshop will start with two pedagogical lectures that introduce likelihood concepts and terminology. These lectures are followed, in the afternoon of Day 1, by a moderated discussion that focuses on the concepts and issues raised in the lectures. Day 1 ends with several presentations that illustrate the use of likelihoods in Higgs and Beyond the Standard Model (BSM) research. The goal here is to get feedback from researchers who have used Higgs and BSM results in their work.
Given the importance of the work on the Higgs boson, we shall devote the second day of the workshop to the thorough deconstruction of likelihood usage in the Higgs boson work by ATLAS and CMS. The goal is to shed a bright light on the many details, and assumptions, that underlie the likelihoods used in the recently published results.
The final day of the workshop covers the use of likelihoods in BSM work and ends with an examination and discussion of the concrete steps needed to make the publication of likelihoods by the LHC community systematic and routine.

Starts 21 Jan 2013, 09:00
Ends 23 Jan 2013, 18:00
Europe/Zurich

CERN
4/3-006 - TH Conference Room
Go to map

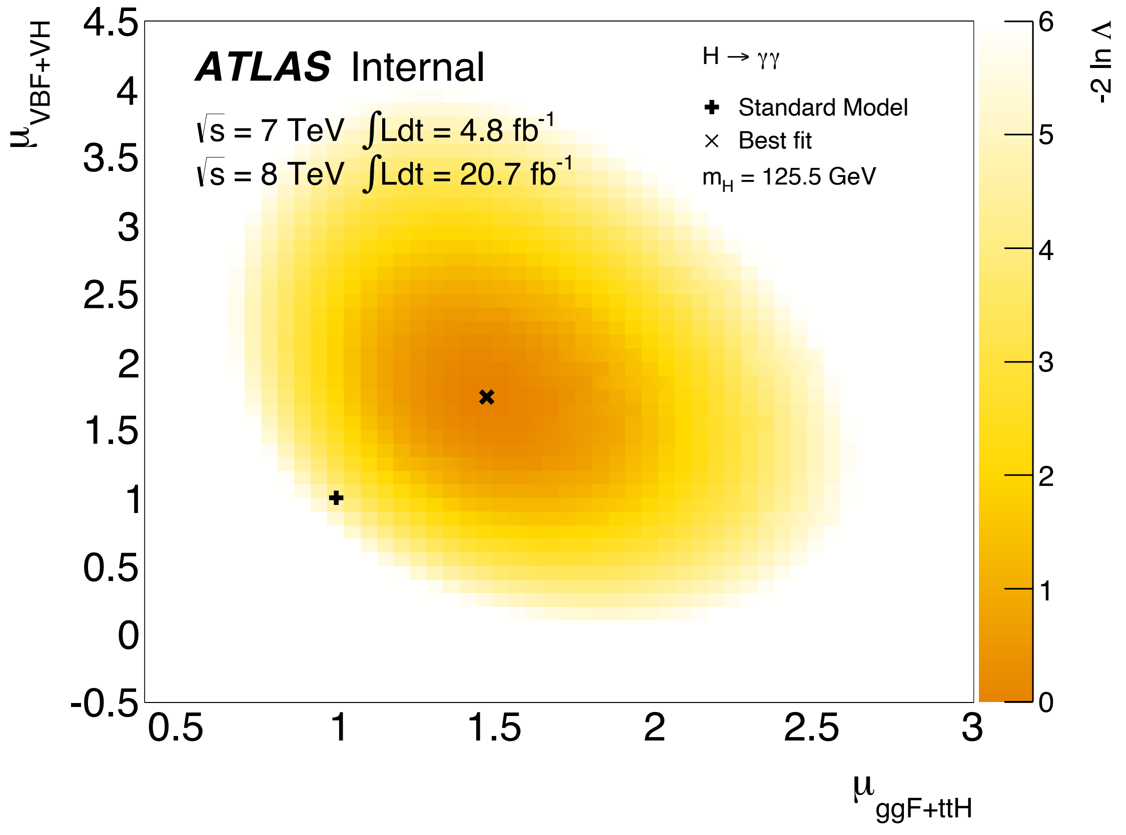
Harrison Prosper
Kyle Stuart Cranmer
Sezen Sekmen

https://indico.cern.ch/event/218693/overview

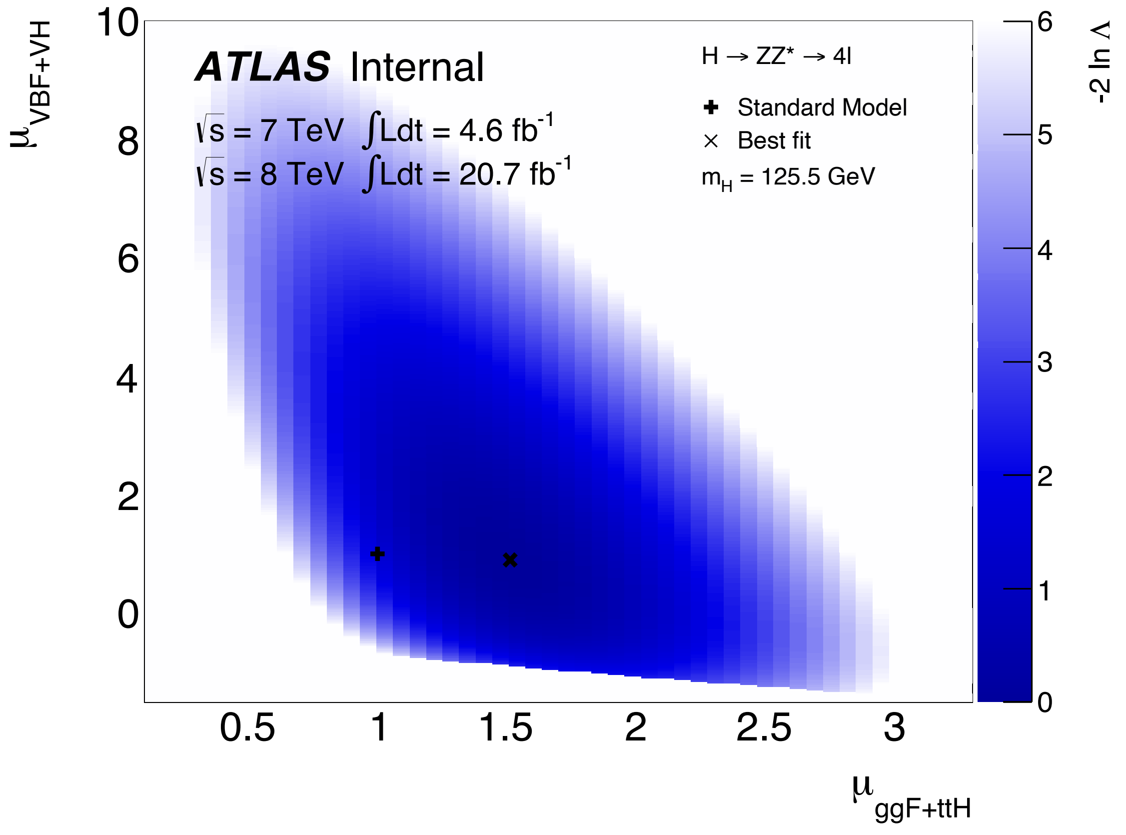
LIKELIHOOD SCANS

First step: publish likelihood scans for communicating LHC Higgs results.

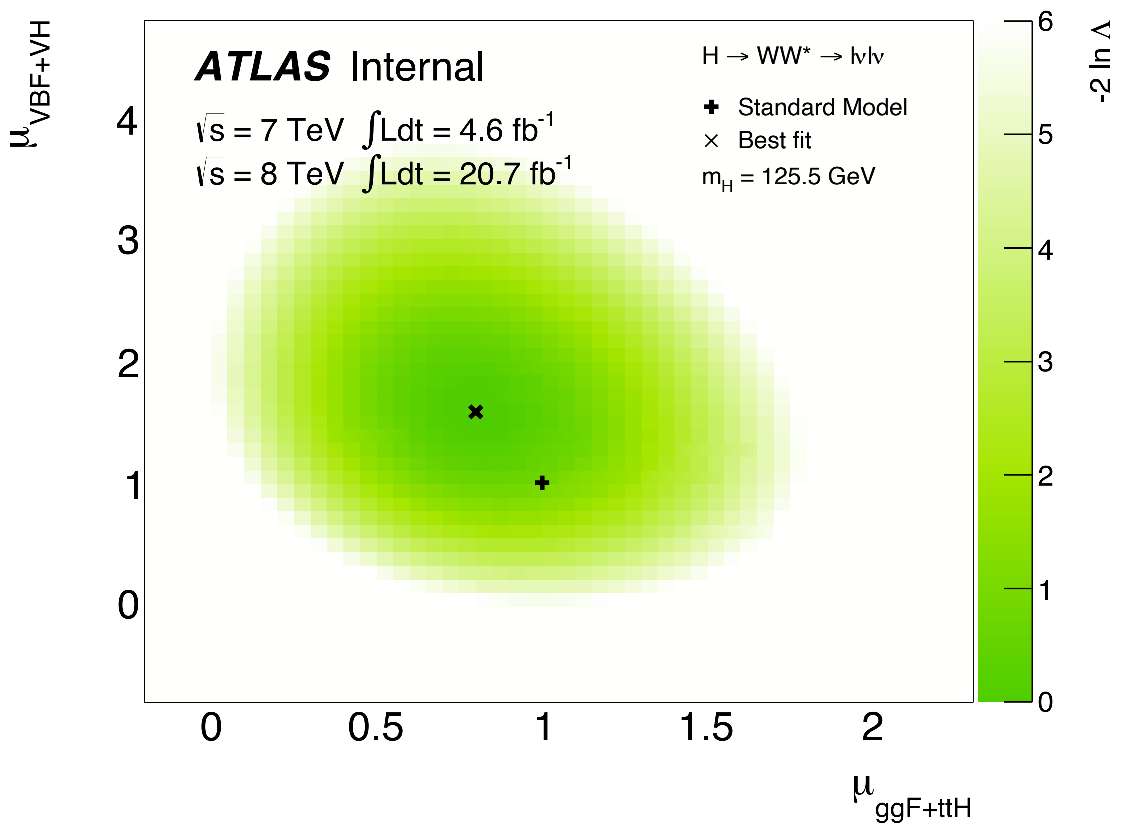
http://doi.org/10.7484/INSPIREHEP.DATA.A78C.HK44



http://doi.org/10.7484/INSPIREHEP.DATA.RF5P.6M3K



http://doi.org/10.7484/INSPIREHEP.DATA.26B4.TY5F



Information | References (121) | Citations (128) | Files | Plots | HepData

Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC

ATLAS Collaboration (Georges Aad (Freiburg U.) *et al.*) *Show all 2923 authors*

Jul 4, 2013 - 32 pages

Phys.Lett. B726 (2013) 88-119

Blogged by **3**
Tweeted by **6**
Click for more details

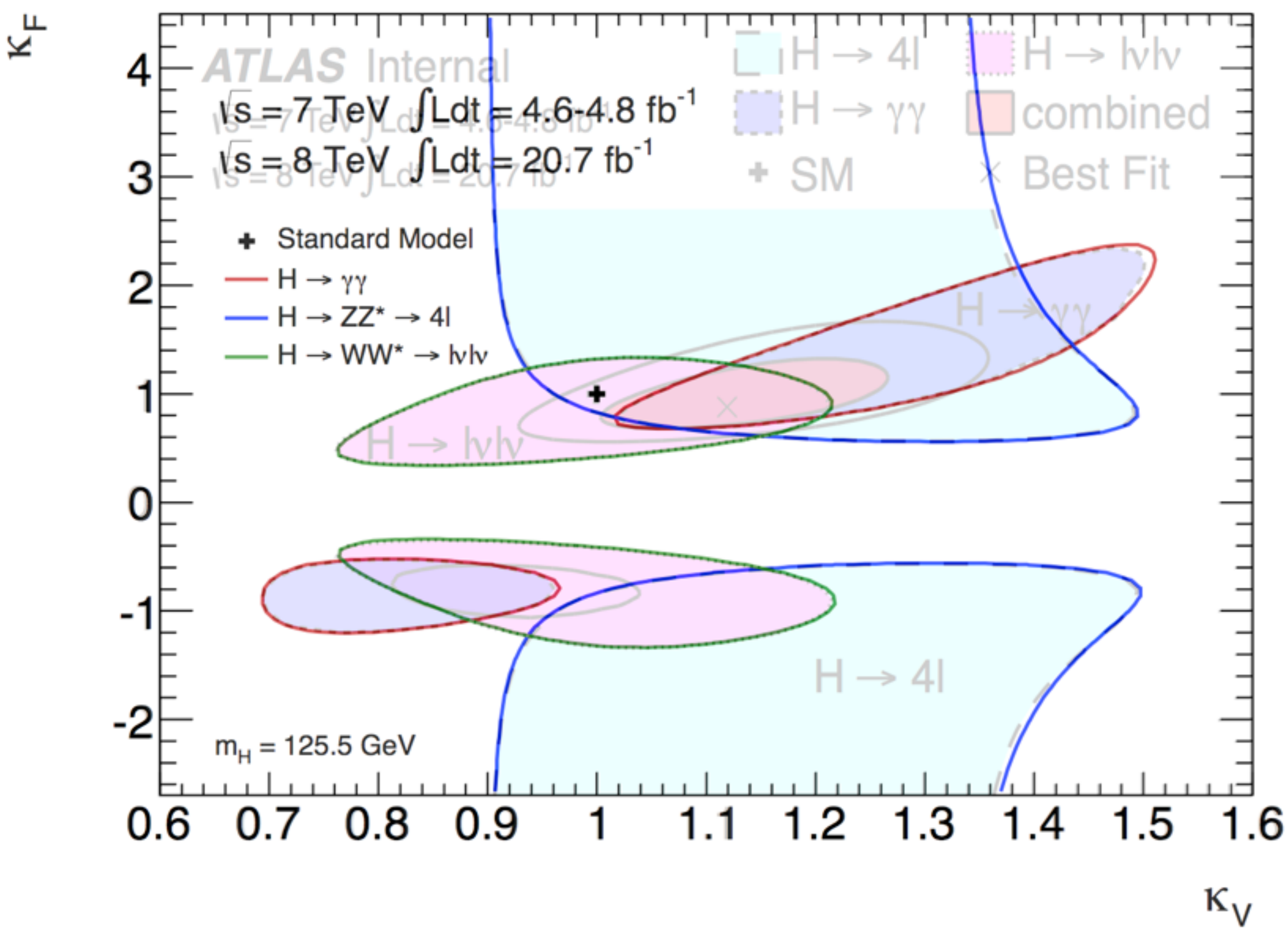
Information | Citations (7) | Files

Data from Figure 7 from: Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC

ATLAS Collaboration (Aad, Georges (Freiburg U.) [...]) *Show all 2923 authors*

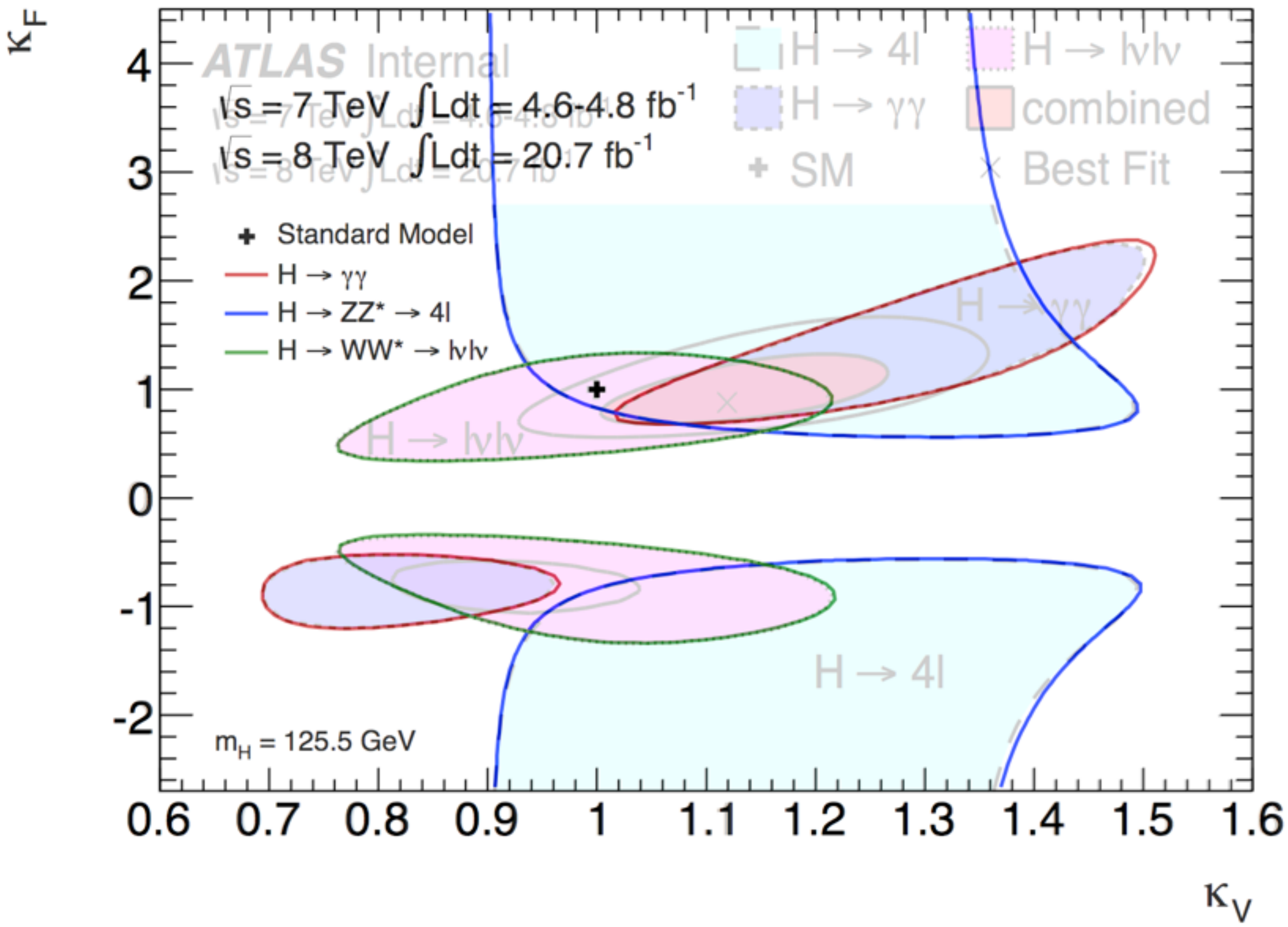
Cite as: ATLAS Collaboration (2013) HepData, http://doi.org/10.7484/INSPIREHEP.DATA.A78C.HK44

Reproducing derived results from original paper!



LIKELIHOODS SCANS

Reproducing derived results from original paper!

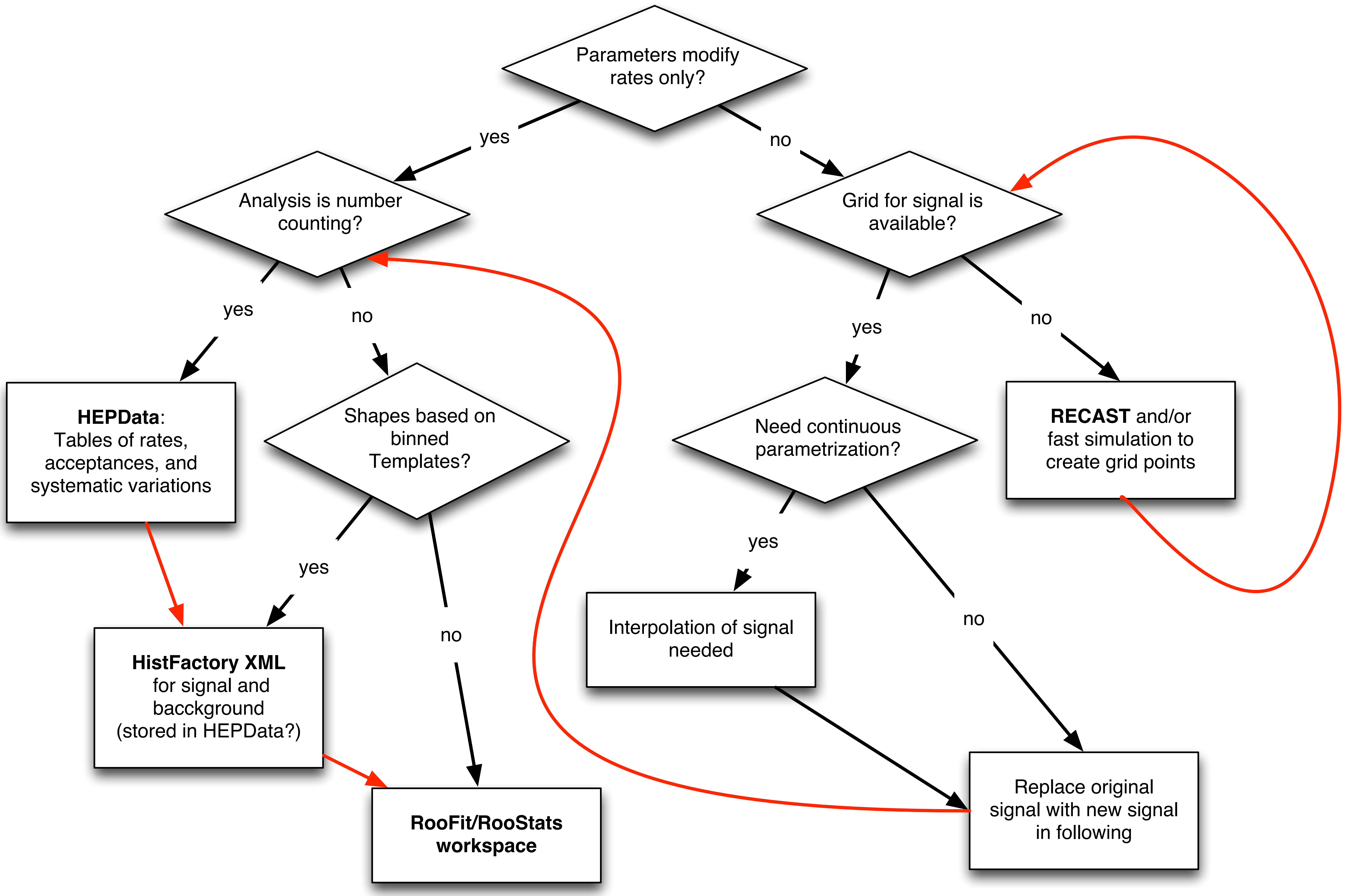


But still simplified likelihood scans, not the full statistical model

Likelihood Publishing + RECAST **=**

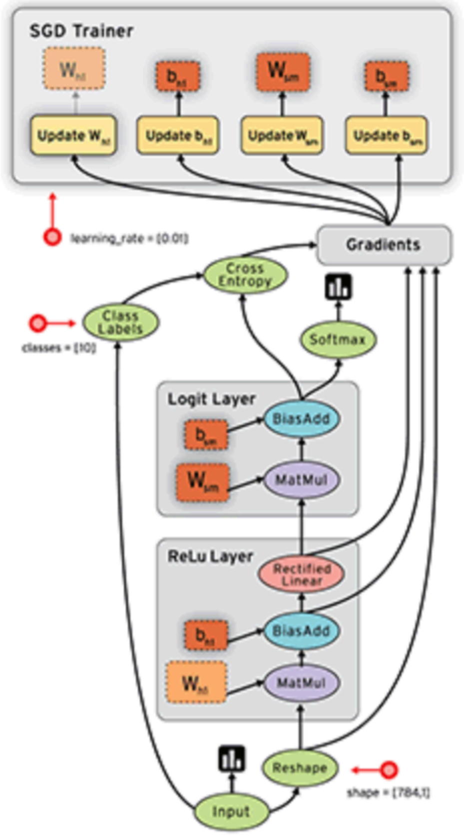
Roadmap (2012)

CENTER FOR
COSMOLOGY AND
PARTICLE PHYSICS



2013-2015 I was mainly focusing on the rise of deep learning, SBI, and RECAST

- TensorFlow was released during DS@LHC at CERN in 2015!



What is a Data Flow Graph?

Data flow graphs describe mathematical computation with a directed graph of nodes & edges. Nodes typically implement mathematical operations, but can also represent endpoints to feed in data, push out results, or read/write persistent variables. Edges describe the input/output relationships between nodes. These data edges carry dynamically-sized multidimensional data arrays, or tensors. The flow of tensors through the graph is where TensorFlow gets its name. Nodes are assigned to computational devices and execute asynchronously and in parallel once all the tensors on their incoming edges becomes available.

TensorFlow was developed by the Google Brain team for internal Google use in research and production.[5][6][7] The initial version was released under the Apache License 2.0 in 2015.[1][8] Google released the updated version of TensorFlow, named TensorFlow 2.0, in September 2019.[9]

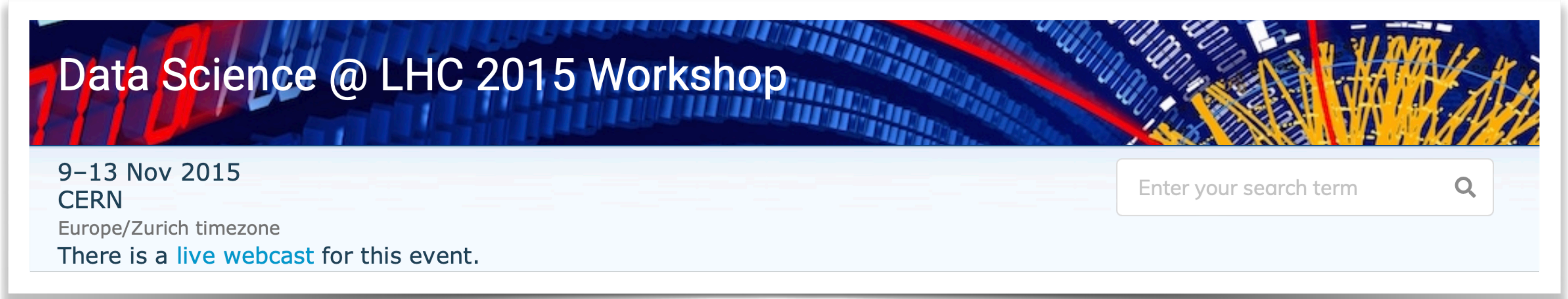
References [edit]

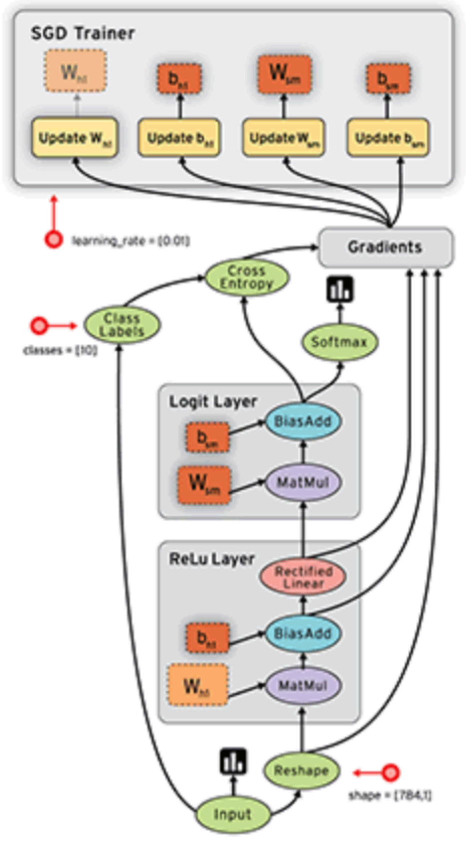
1. ^ a b "Credits". *TensorFlow.org*. Retrieved November 10, 2015.

Data Science / TensorFlow / DIANA-HEP

2013-2015 I was mainly focusing on the rise of deep learning, SBI, and RECAST

- TensorFlow was released during DS@LHC at CERN in 2015!





What is a Data Flow Graph?

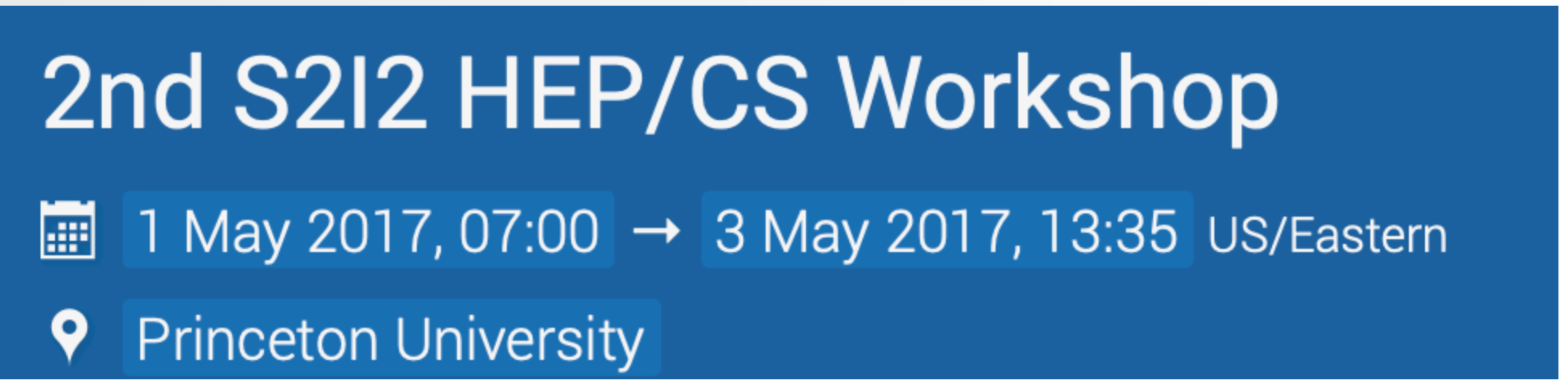
Data flow graphs describe mathematical computation with a directed graph of nodes & edges. Nodes typically implement mathematical operations, but can also represent endpoints to feed in data, push out results, or read/write persistent variables. Edges describe the input/output relationships between nodes. These data edges carry dynamically-sized multidimensional data arrays, or tensors. The flow of tensors through the graph is where TensorFlow gets its name. Nodes are assigned to computational devices and execute asynchronously and in parallel once all the tensors on their incoming edges becomes available.

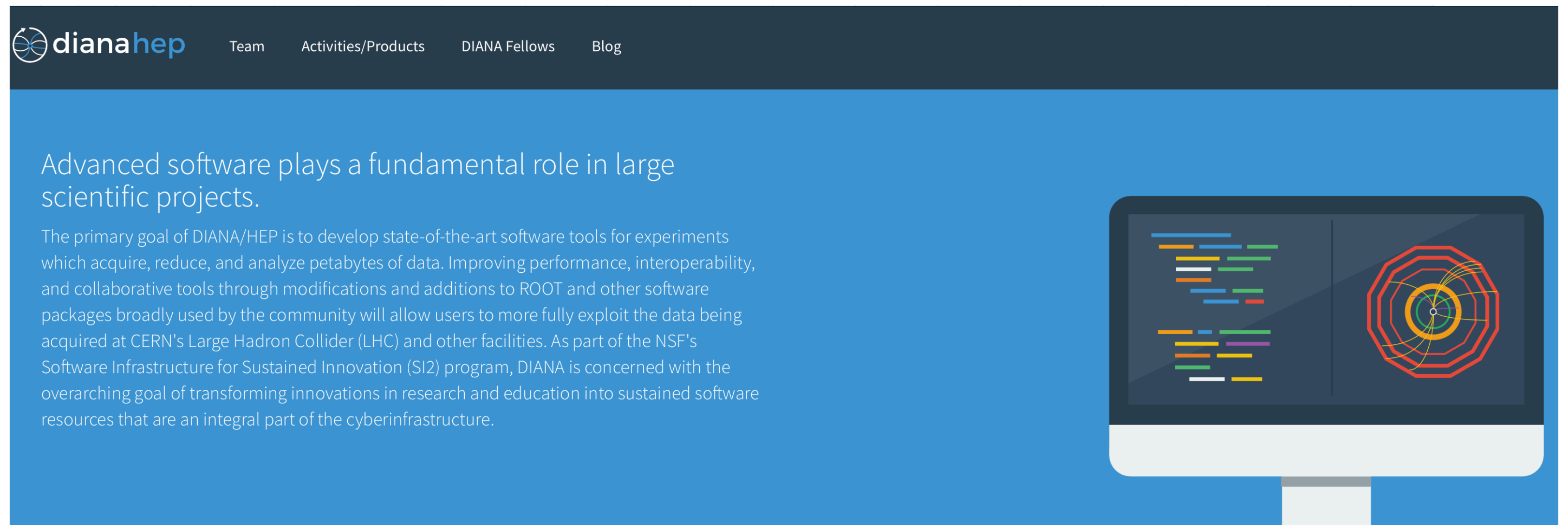
TensorFlow was developed by the Google Brain team for internal Google use in research and production.[5][6][7] The initial version was released under the Apache License 2.0 in 2015.[1][8] Google released the updated version of TensorFlow, named TensorFlow 2.0, in September 2019.[9]

References [edit]

1. ^ a b "Credits". TensorFlow.org. Retrieved November 10, 2015.

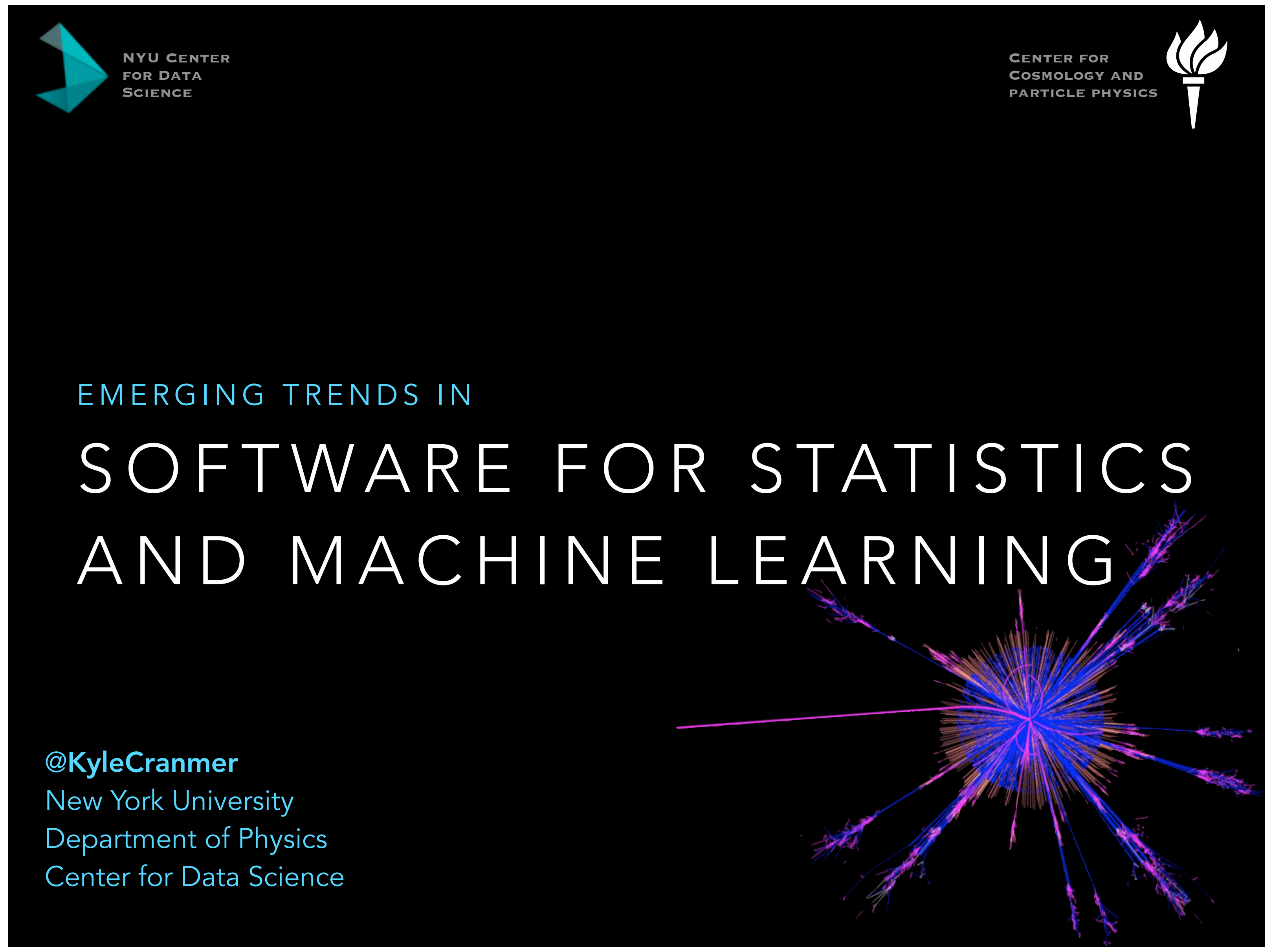
The birth of pyhf

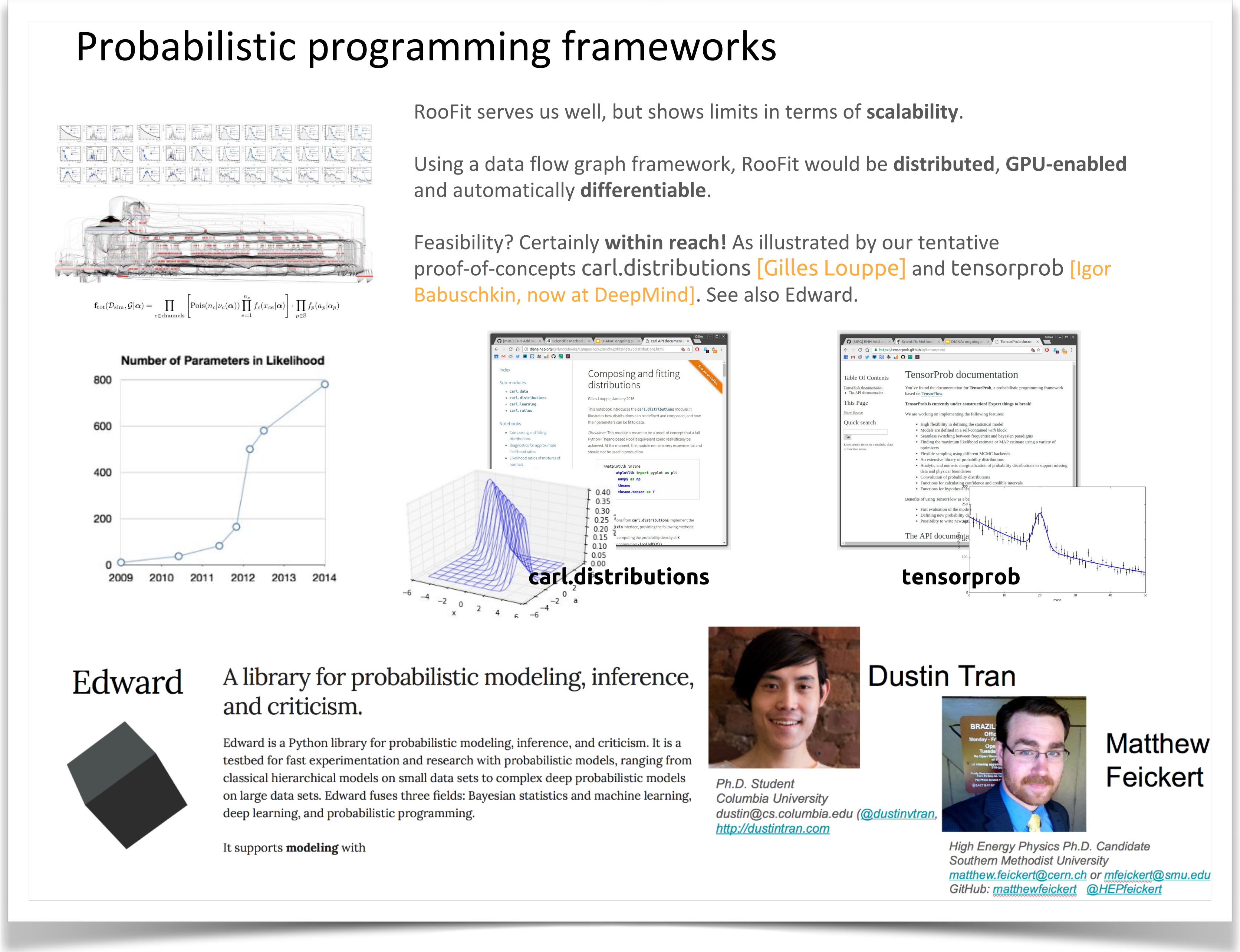


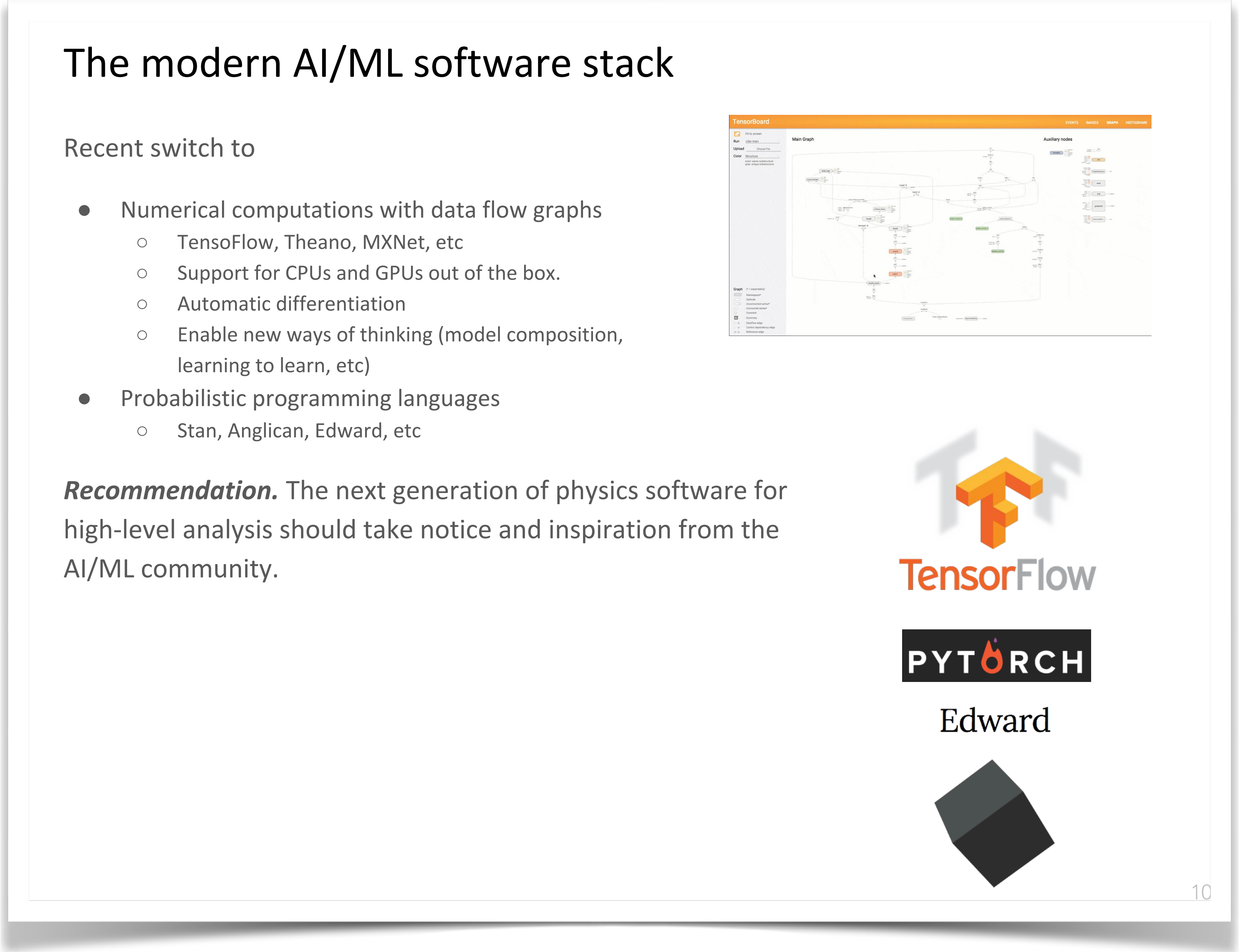


Matthew Feickert, Southern Methodist University [Grad]

- Topic: Investigation of use of Tensorflow/Theano for realistic physics statistics models - proposal, final report
- Mentor: Gilles Louppe/Vincent Croft, NYU
- Dates/Location:







https://indico.cern.ch/event/622920/timetable/#sc-5-2-lightning-talk-emerging

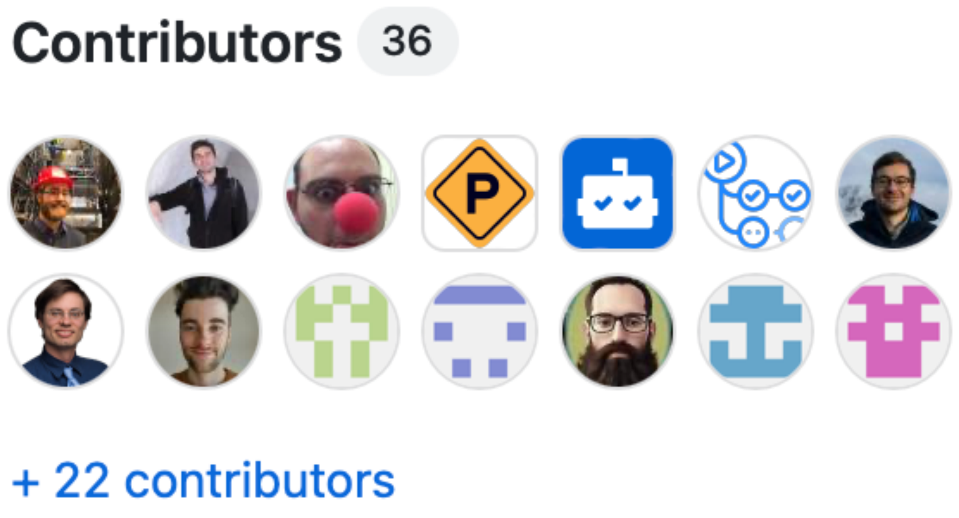
pyhf matures

Thanks to the amazing developer team, pyhf quickly matures along with scikit-hep &PyHEP community

Matthew Feickert

Giordon Stark

Lukas Heinrich



iris hep

Scikit HEP

pyhf ∂ifferentiable ℒikelihoods


JOSS The Journal of Open Source Software

pyhf: pure-Python implementation of HistFactory statistical models

Lukas Heinrich[1], Matthew Feickert*[2], Giordon Stark[3], and Kyle Cranmer[4]

1 CERN 2 University of Illinois at Urbana-Champaign 3 SCIPP, University of California, Santa Cruz 4 New York University

DOI: 10.21105/joss.02823

Software

- Review
- Repository
- Archive

Editor: Eloisa Bentivegna

Reviewers:

- @suchitakulkarni
- @bradkav

Submitted: 07 October 2020
Published: 04 February 2021

Summary

Statistical analysis of High Energy Physics (HEP) data relies on quantifying the compatibility of observed collision events with theoretical predictions. The relationship between them is often formalised in a statistical model $f(\mathbf{x}|\phi)$ describing the probability of data $\mathbf{x}$ given model parameters ϕ. Given observed data, the likelihood $\mathcal{L}(\phi)$ then serves as the basis for inference on the parameters ϕ. For measurements based on binned data (histograms), the `HistFactory` family of statistical models (Cranmer et al., 2012) has been widely used in both Standard Model measurements (ATLAS Collaboration, 2013) as well as searches for new physics (ATLAS Collaboration, 2018). `pyhf` is a pure-Python implementation of the `HistFactory` model specification and implements a declarative, plain-text format for describing `HistFactory`-based likelihoods that is targeted for reinterpretation and long-term preservation in analysis data repositories such as HEPData (Maguire et al., 2017). The source code for `pyhf` has been archived on Zenodo with the linked DOI: (Heinrich, Lukas and Feickert, Matthew and Stark, Giordon, 2020). At the time of writing this paper, the most recent release of `pyhf` is `v0.5.4`.

Statement of Need

Through adoption of open source "tensor" computational Python libraries, `pyhf` decreases the abstractions between a physicist performing an analysis and the statistical modeling without sacrificing computational speed. By taking advantage of tensor calculations, `pyhf` outperforms the traditional `C++` implementation of `HistFactory` on data from real LHC analyses. `pyhf`'s default computational backend is built from NumPy and SciPy, and supports TensorFlow, PyTorch, and JAX as alternative backend choices. These alternative backends support hardware acceleration on GPUs, and in the case of JAX JIT compilation, as well as auto-differentiation allowing for calculating the full gradient of the likelihood function — all contributing to speeding up fits.

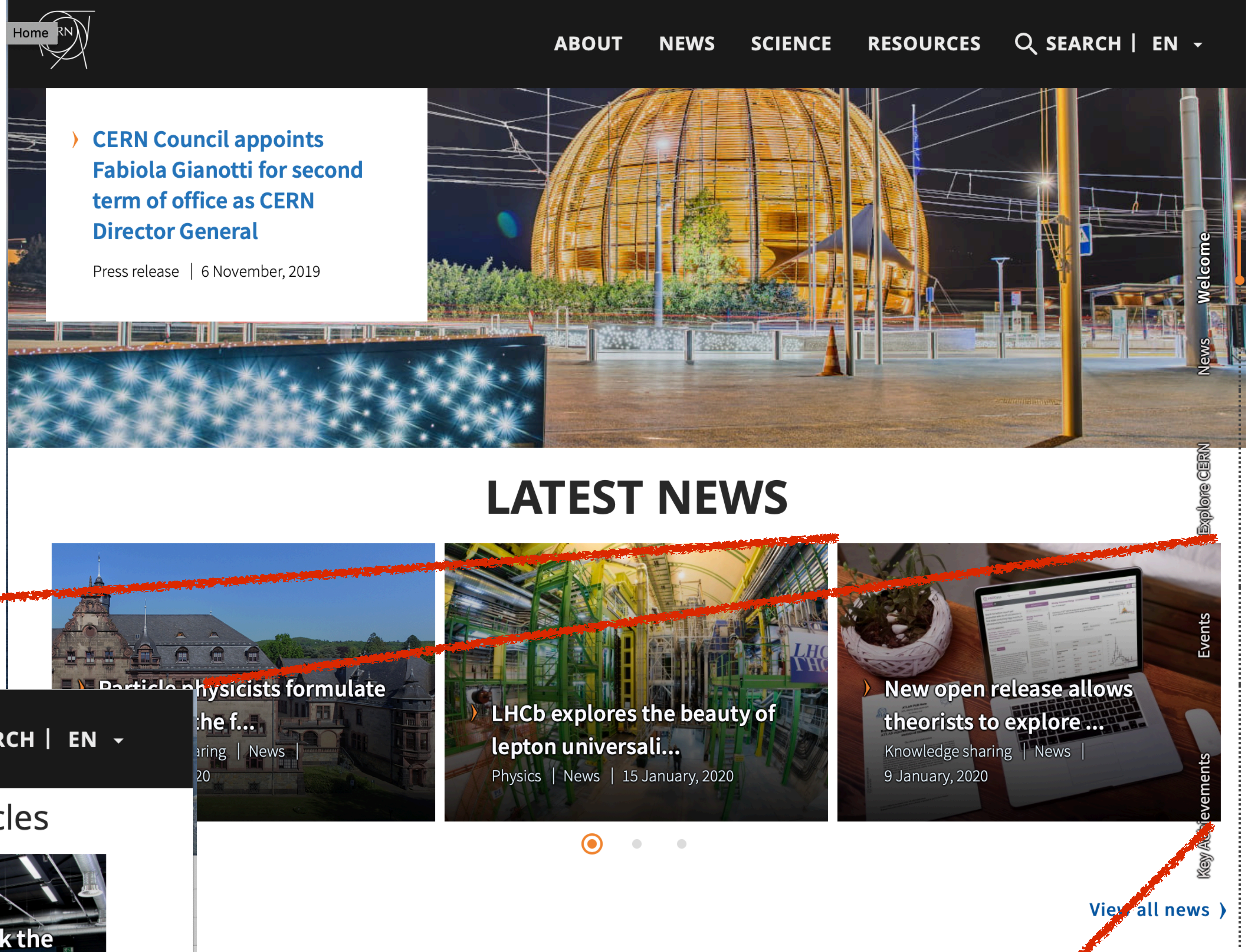
Impact on Physics

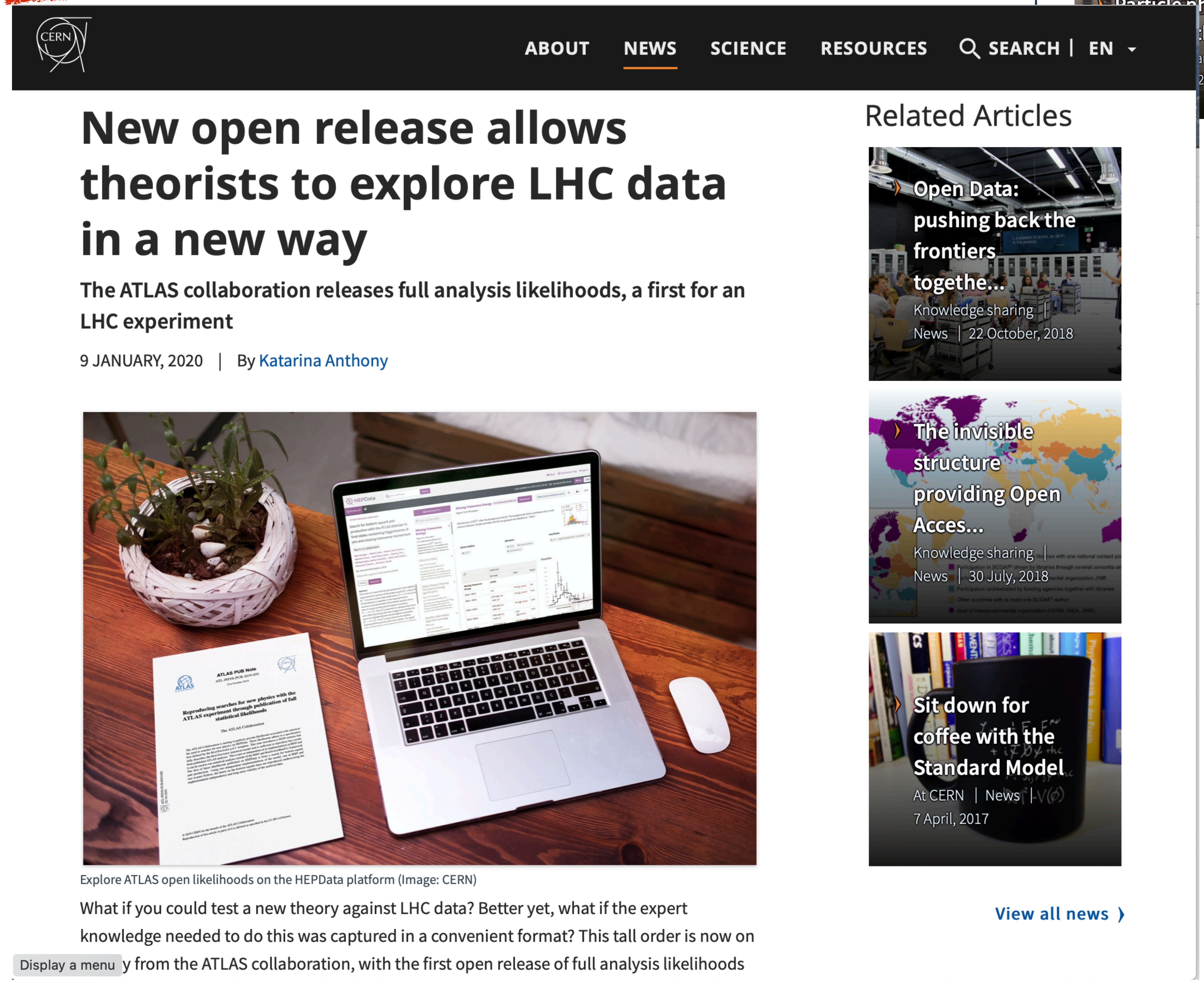
In addition to enabling the first publication of full likelihoods by an LHC experiment (ATLAS Collaboration, 2019), `pyhf` has been used by the `SModelS` library to improve the reinterpretation of results of searches for new physics at LHC experiments (Abdallah & others, 2020; Alguero et al., 2020; Khosa et al., 2020).

*Corresponding author.

Recent progress

ATLAS has started publishing full likelihoods to HEPData!





pyhf
differentiable likelihoods

https://scikit-hep.org/pyhf/

SciPost Physics [arxiv:2109.04981] Submission

Publishing statistical models: Getting the most out of particle physics experiments

Kyle Cranmer[1*], Sabine Kraml[2‡], Harrison B. Prosper[3§] (editors), Philip Bechtle[4], Florian U. Bernlochner[4], Itay M. Bloch[5], Enzo Canonero[6], Marcin Chrzaszcz[7], Andrea Coccaro[8], Jan Conrad[9], Glen Cowan[10], Matthew Feickert[11], Nahuel Ferreiro Iachellini[12,13] Andrew Fowlie[14], Lukas Heinrich[15], Alexander Held[1], Thomas Kuhr[13,16], Anders Kvellestad[17], Maeve Madigan[18], Farvah Mahmoudi[15,19], Knut Dundas Morå[20], Mark S. Neubauer[11], Maurizio Pierini[15], Juan Rojo[8], Sezen Sekmen[22], Luca Silvestrini[23], Veronica Sanz[24,25], Giordon Stark[26], Riccardo Torre[8], Robert Thorne[27], Wolfgang Waltenberger[28], Nicholas Wardle[29], Jonas Wittbrodt[30]

It's a reality

twiki.cern.ch

Papers, Conference notes, Public notes

ATLAS Public Results Page

The buttons below provide document filters along predefined keywords. Selecting a "Physics theme" gives access to specific additional keywords allowing to refine the selection.

Global Selections	Show All · Deselect All · Show Latest 10
CM Energy	14 TeV · 13.6 TeV · 13 TeV · 8 TeV · 7 TeV · 5 TeV · 2.36 TeV · 2.76 GeV · 900 GeV
	8.16 TeV/NN · 5.44 TeV/N · 5.02 TeV/N · 2.76 TeV/NN
Physics theme	B-physics and light states · Standard Model · Top · Higgs · BSM Searches · Heavy Ion · Upgrade Studies · Outreach · Statistical methods
	Tracking · Egamma · Muon · Tau · Jet/Etmiss · Flavour tagging · Physics Modelling
Signature *'OR' logic within row* *'AND' between rows*	W · Z · Photon · H · WW · WZ · ZZ · Di-photon · Vphoton · HH · VVV · Single top · Top pair · ≥3 tops
	Charged tracks
	0 lepton · 1 lepton · 2 leptons · 2 leptons (same charge) · ≥3 leptons
	Taus · Photons
	0 jets · 1 jet · 2 jets · ≥3 jets · All hadronic · c-jets · b-jets · Boosted
	MET
	Long-lived massive particle
	Forward Proton
Analysis characteristics	Cross-section measurement · Mass measurement · Statistical combination · ISR · Gluon fusion · VBF · VBS · PDF fits
	Double parton scattering · BSM search · BSM reinterpretation · LFV · FCNC · Particle flow · MVA / machine learning · EFT interpretation
	Differential measurement · Displaced vertex · Lepton-jets · Trigger-level analysis · High luminosity upgrade studies · Photon-induced
	Likelihood available
Min luminosity :	0 fb^{-1} · Filter by minimum integrated luminosity
Date :	Min: 11/14/2023 Max: 11/14/2023 · ArXiv relase · Publication

Quick links: Papers Confnotes Pubnotes

It's a reality

https://www.hepdata.net/search/?q=analysis:HistFactory

Find all papers which include specific types of **analysis**.

- analysis:rivet (Rivet analysis)
- analysis:MadAnalysis (MadAnalysis 5 analysis)
- analysis:HistFactory (likelihoods in HistFactory format)

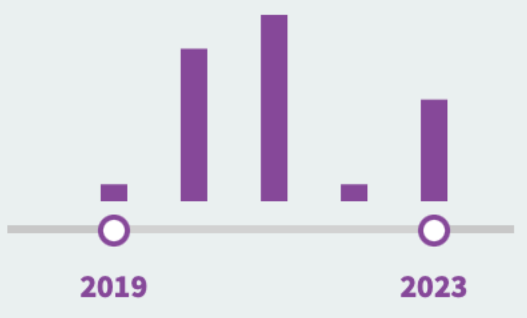
HEPData

analysis:HistFactory | Search | Rese

Max results | Sort by | Reverse order | Showing **10** of **28** results

« ‹ 1 2 › »

Date



Collaboration

ATLAS	28

Subject_areas

hep-ex	28

Phrases

Proton-Proton Scattering	3
Cross Section	2
SUSY	2
Supersymmetry	2
Top	2

Next 5 Show All

Reactions

P P --> CHARGINO+ CHARGINO-	1
P P --> CHARGINO+ NEUTRALINO	1
P P --> CHARGINO+- NEUTRALINO	1

HistFactory HistFactory HistFactory HistFactory **Search for flavour-changing neutral-current couplings between the top quark and the photon with the ATLAS detector at $\sqrt{s} = 13$ TeV**

The ATLAS collaboration Aad, Georges ; Abbott, Braden Keim ; Abbott, Dale ; *et al.*

Phys.Lett.B 842 (2023) 137379, 2023.

Inspire Record 2077557 DOI 10.17182/hepdata.129959

This letter documents a search for flavour-changing neutral currents (FCNCs), which are strongly suppressed in the Standard Model, in events with a photon and a top quark with the ATLAS detector. The analysis uses data collected in pp collisions at $\sqrt{s} = 13$ TeV during Run 2 of the LHC, corresponding to an integrated luminosity of 139 fb^{-1}. Both FCNC top-quark production and decay are considered. The final state consists of a charged lepton, missing transverse momentum, a b-tagged jet, on…

0 data tables match query

HistFactory HistFactory **Measurement of the $t\bar{t}t\bar{t}$ production cross section in pp collisions at $\sqrt{s}$=13 TeV with the ATLAS detector**

The ATLAS collaboration Aad, Georges ; Abbott, Braden Keim ; Abbott, Dale ; *et al.*

JHEP 11 (2021) 118, 2021.

Inspire Record 1869695 DOI 10.17182/hepdata.105039

A measurement of four-top-quark production using proton-proton collision data at a centre-of-mass energy of 13 TeV collected by the ATLAS detector at the Large Hadron Collider corresponding to an integrated luminosity of 139 fb^{-1} is presented. Events are selected if they contain a single lepton (electron or muon) or an opposite-sign lepton pair, in association with multiple jets. The events are categorised according to the number of jets and how likely these are to contain b-hadrons. A…

0 data tables match query

HistFactory HistFactory **Observation of single-top-quark production in association with a photon using the ATLAS detector**

Using published likelihoods

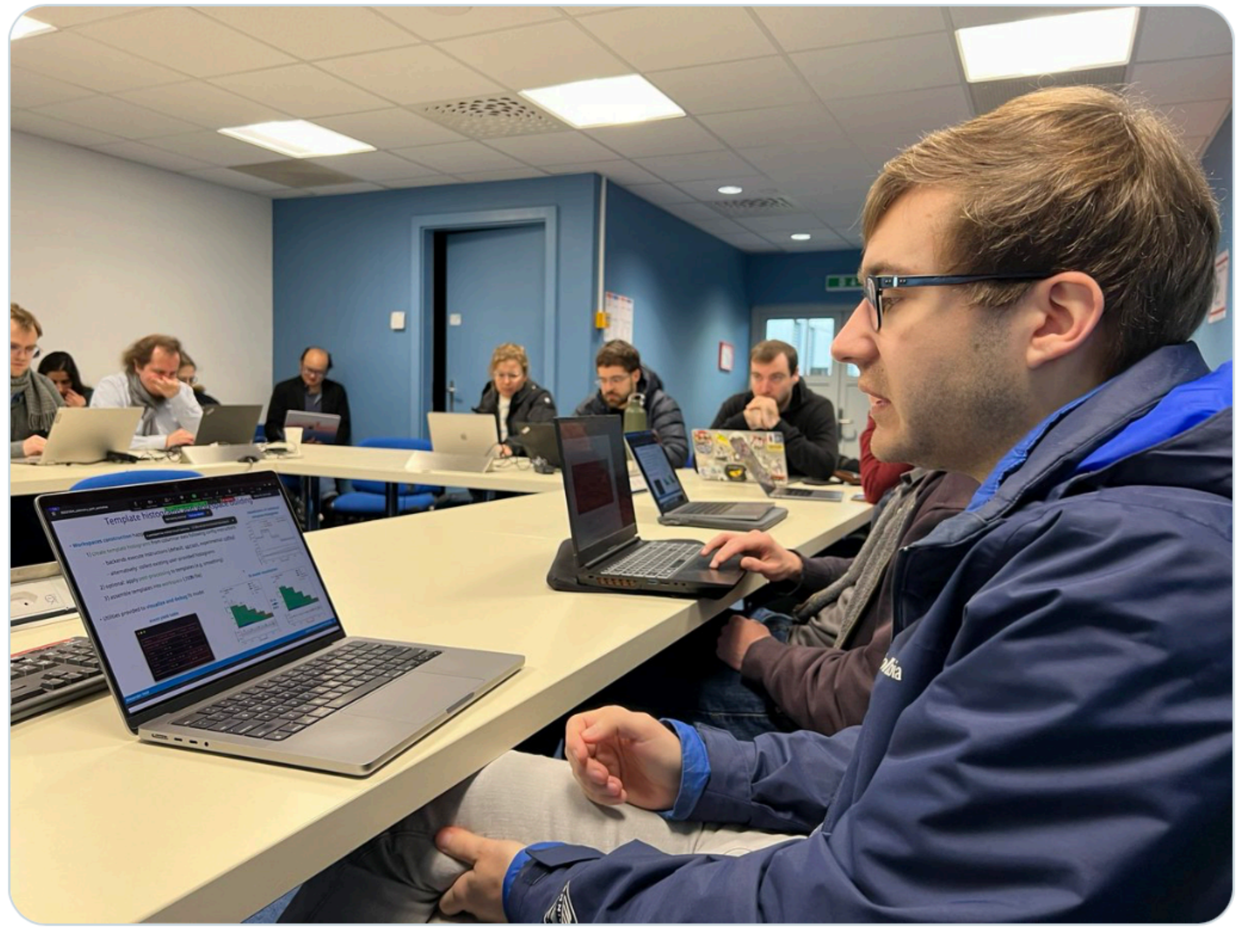
cabinetry

https://github.com/scikit-hep/cabinetry

```python
import json
import cabinetry
import pyhf
from cabinetry.model_utils import prediction
from pyhf.contrib.utils import download

# download the ATLAS bottom-squarks analysis probability models from HEPData
download("https://www.hepdata.net/record/resource/1935437?view=true", "bottom-squarks")

# construct a workspace from a background-only model and a signal hypothesis
bkg_only_workspace = pyhf.Workspace(json.load(open("bottom-squarks/RegionC/BkgOnly.json")))
patchset = pyhf.PatchSet(json.load(open("bottom-squarks/RegionC/patchset.json")))
workspace = patchset.apply(bkg_only_workspace, "sbottom_600_280_150")

# construct the probability model and observations
model, data = cabinetry.model_utils.model_and_data(workspace)

# produce visualizations of the pre-fit model and observed data
prefit_model = prediction(model)
cabinetry.visualize.data_mc(prefit_model, data)

# fit the model to the observed data
fit_results = cabinetry.fit.fit(model, data)

# produce visualizations of the post-fit model and observed data
postfit_model = prediction(model, fit_results=fit_results)
cabinetry.visualize.data_mc(postfit_model, data)
```

pyhf - python HistFactory @pyhf_ · 17h

Now we're moving on to an overview of cabinetry from @alheld_!
indico.cern.ch/event/1294577/...



Alexander Held

Using published likelihoods

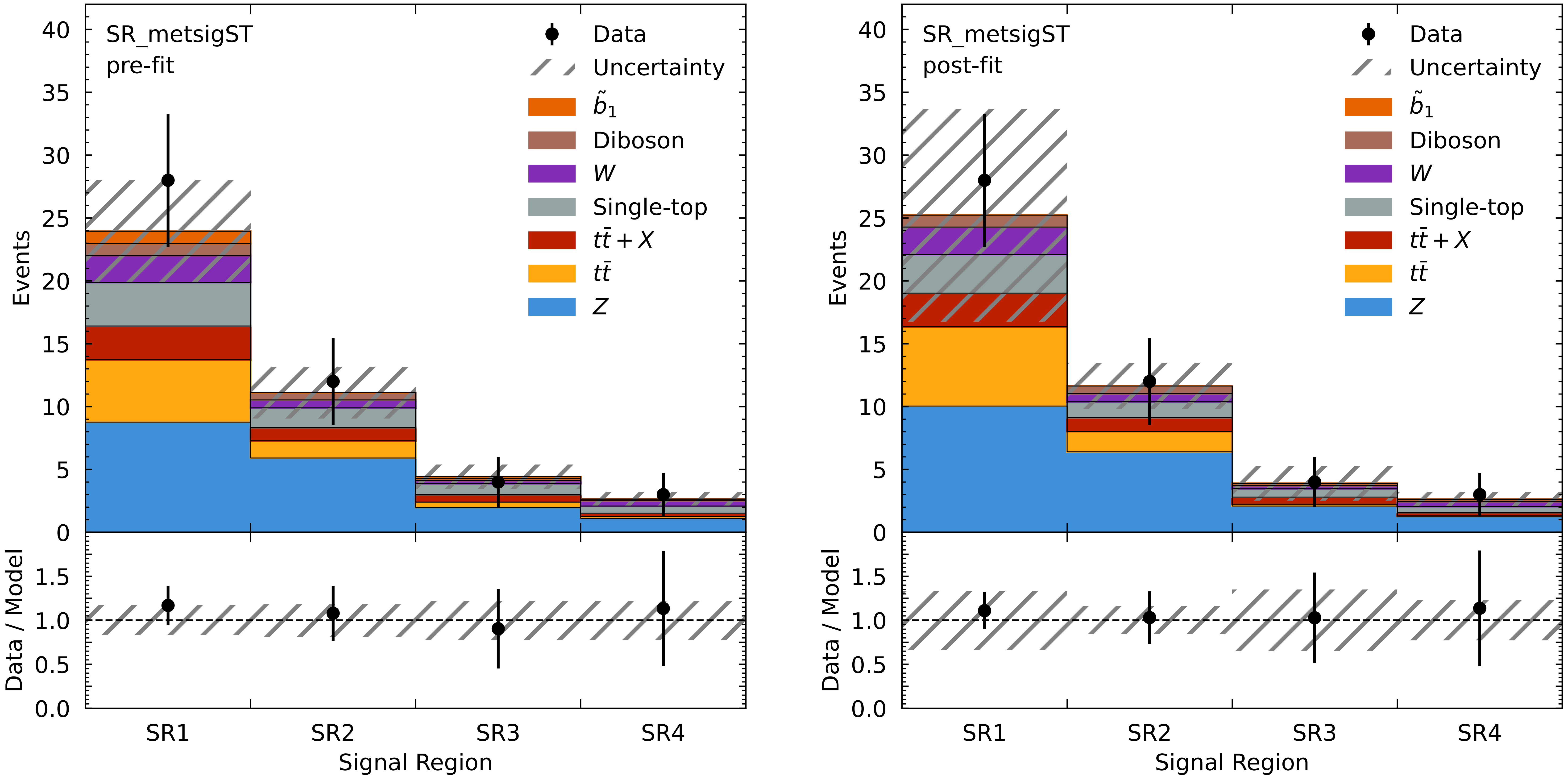
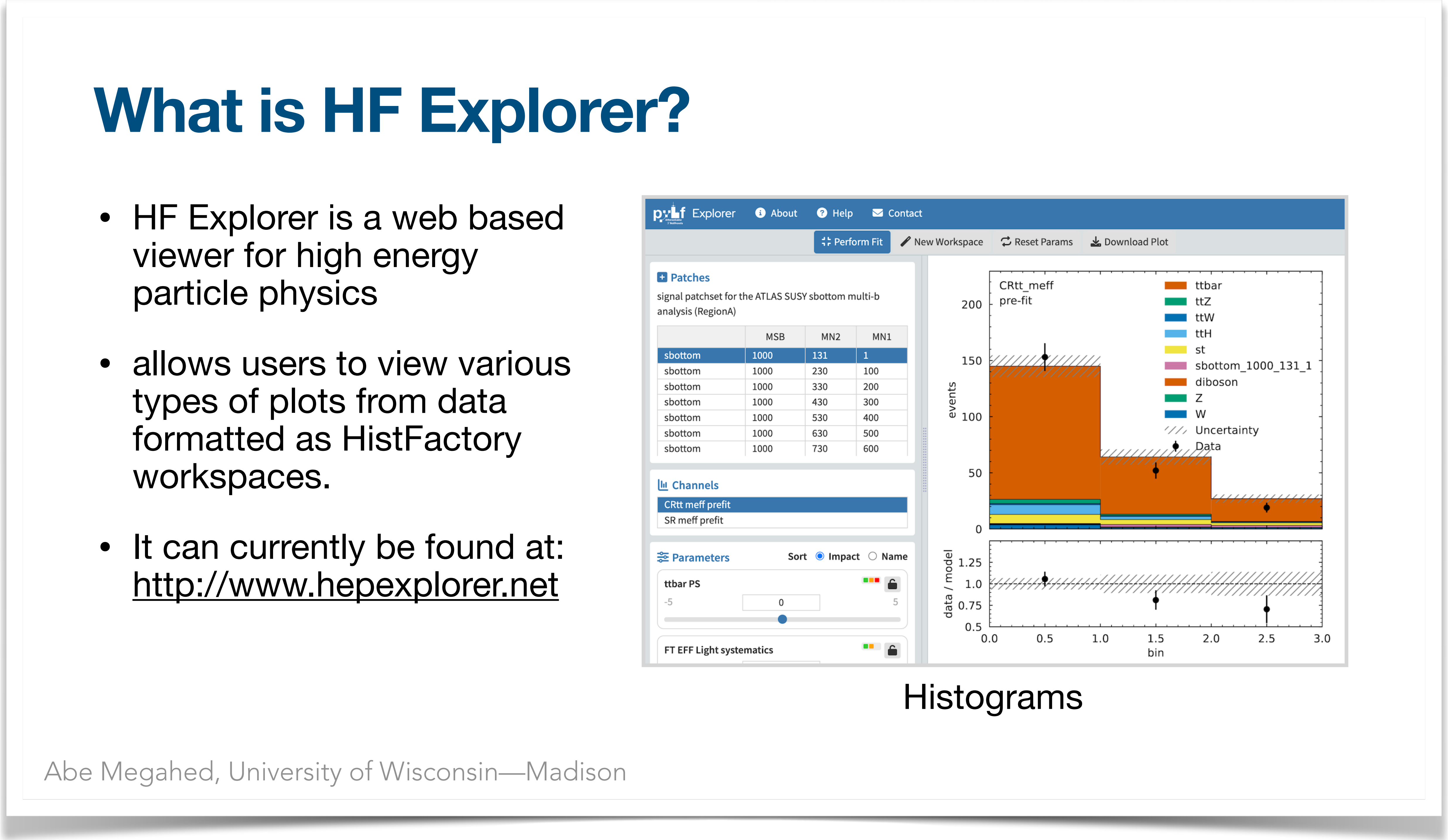


Figure 3: Pre-fit (left) and post-fit (right) visualizations of a selected signal hypothesis for four signal regions of the ATLAS search [41] of a bottom-squark of mass 600 GeV with a second-lightest neutralino of mass 280 GeV and lightest supersymmetric particle of mass 150 GeV generated from the full statistical models published in Ref. [20] using code from Ref. [40].

A growing ecosystem

It's great to see the emergence of an ecosystem, including use by theorist and user-friendly, zero-install, web-based tools!



https://indico.cern.ch/event/1294577/timetable/#20231204.detailed

Declarative specifications FTW!

LiteHF in Julia!

J. Ling 🦑 @l_II_llI · May 3, 2022

happy to report that after a few days of wrestling with channel-sample-bin structure tunneling/masking, as well as trying not to blow up compilation time, **LiteHF**.jl is slightly faster than even Jax+pyhf. It's AD friendly out of the box (trade-off being increased allocation).

```
julia> R = build_pyhf(load_pyhfjson("./test/pyhfjson/single_channel_big.json"));

julia> f() = R.LogLikelihood(R.prior_inits)
f (generic function with 1 method)

julia> @benchmark f()
BenchmarkTools.Trial: 10000 samples with 1 evaluation.
 Range (min … max):  14.548 μs …   4.274 ms  ┊ GC (min … max):  0.00% … 90.84%
 Time  (median):     15.870 μs               ┊ GC (median):     0.00%
 Time  (mean ± σ):   19.989 μs ± 121.307 μs  ┊ GC (mean ± σ):  17.45% ±  2.88%

 14.5 μs         Histogram: frequency by time         22.8 μs <

 Memory estimate: 48.33 KiB, allocs estimate: 445.
```

JuliaPackages @JuliaPackages · Dec 16, 2022
Automated

New package: LiteHF v0.1.0 announced #JuliaLang

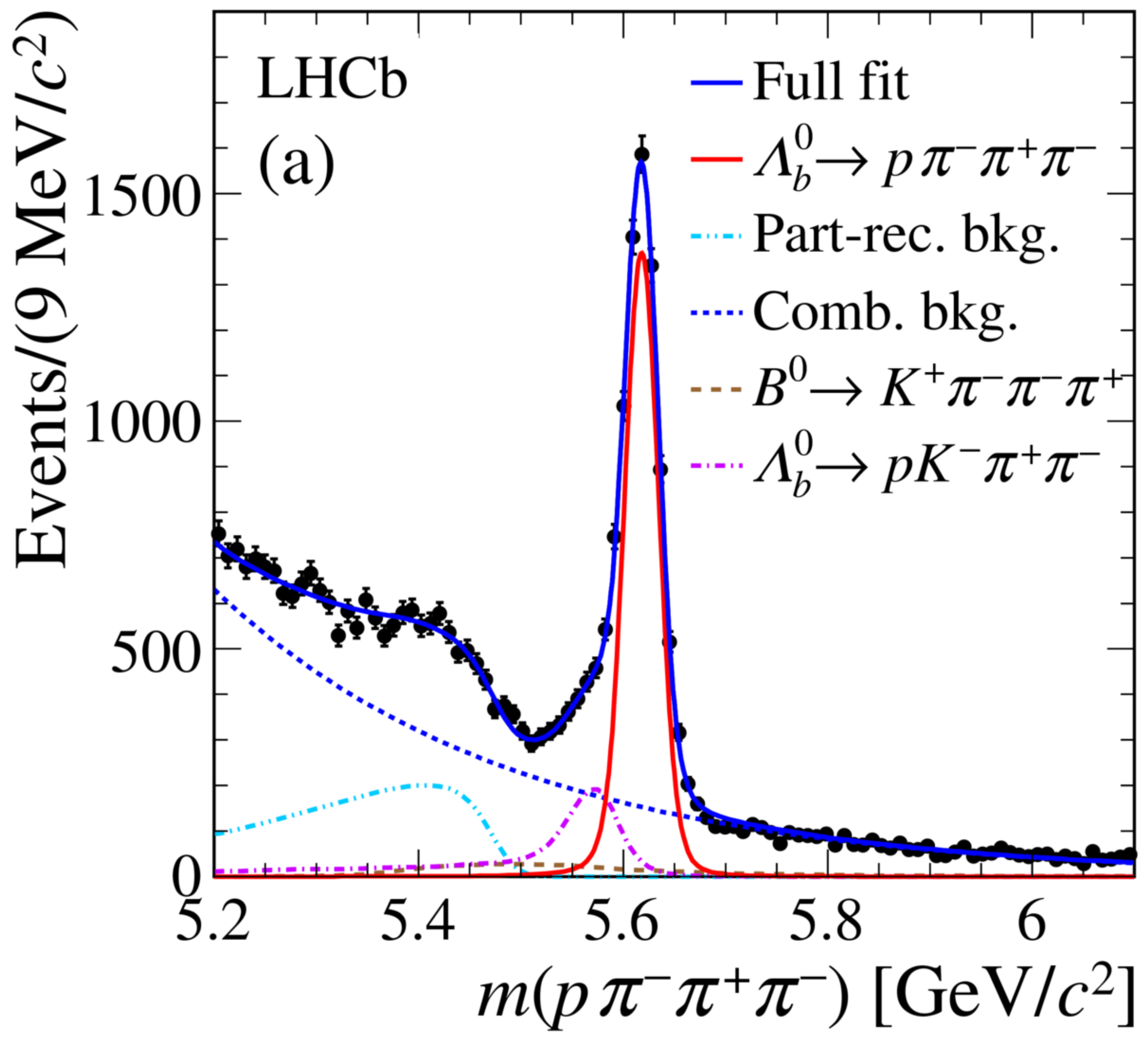
LiteHF: Light-weight **HistFactory** in pure **Julia**, attempts to be compatible with `pyhf` json format

Registration: github.com/JuliaRegistrie...
Repository: github.com/JuliaHEP/LiteH...

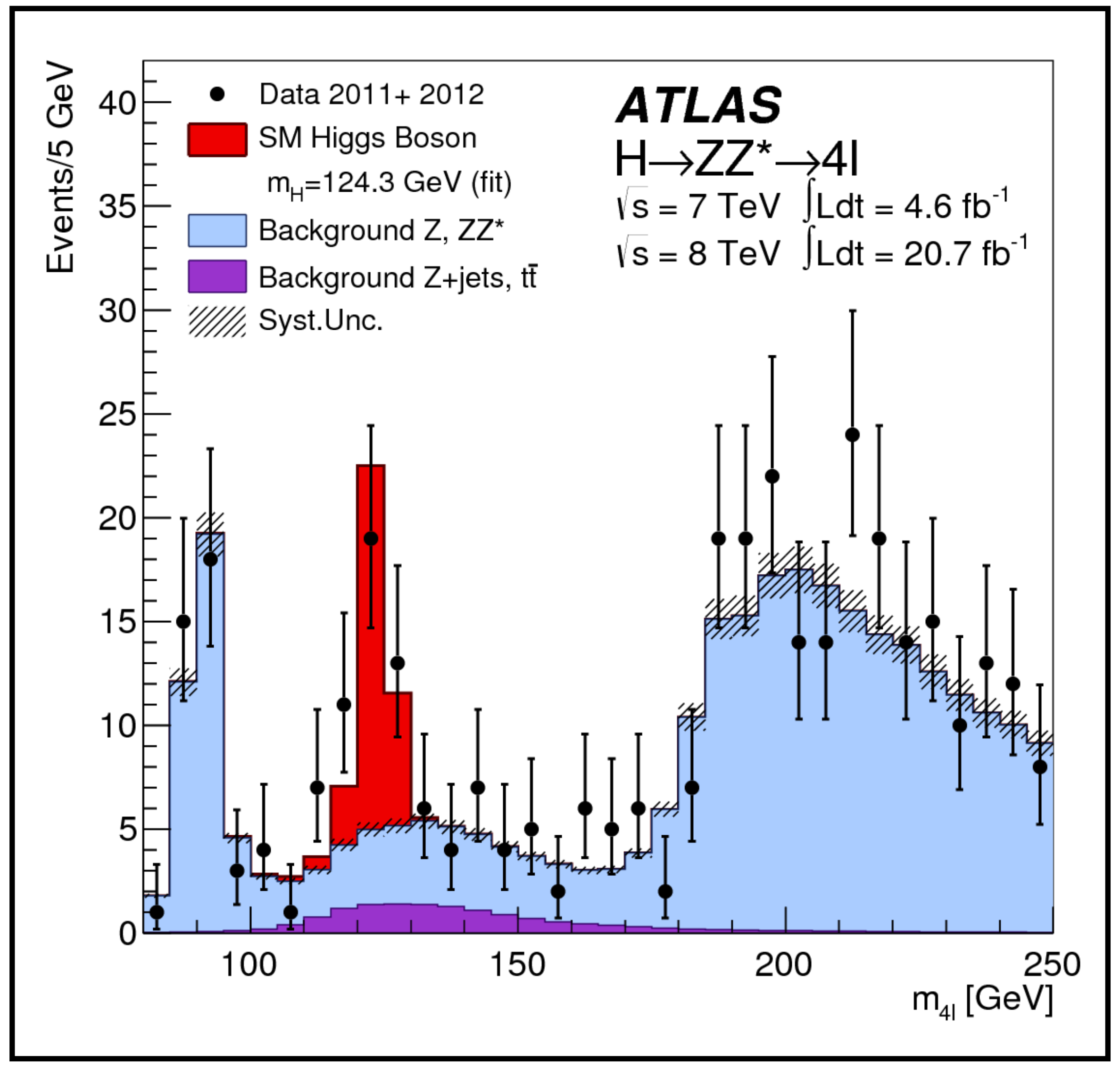
The RooWorkspace was designed to be able to store any type of statistical model → source of many complications

broadly we have two classes of analyses: **binned** and **unbinned**

unbinned



binned



[Slide from Lukas Heinrich]

An exciting initiative

HS³ - A serialization standard for statistical models in high energy physics

Carsten Burgard[1] , Robin Pelkner[1]

Many people involved: Matthew Feickert, Lukas Heinrich, Alexander Held, Cornelius Grunwald, Oliver Schulz, Mikhail Mikhasenko, Jerry Ling, Wouter Verkerke, Jonas Eschle, Lorenzo Moneta, Louis Moureaux, Tomas Dado and many others

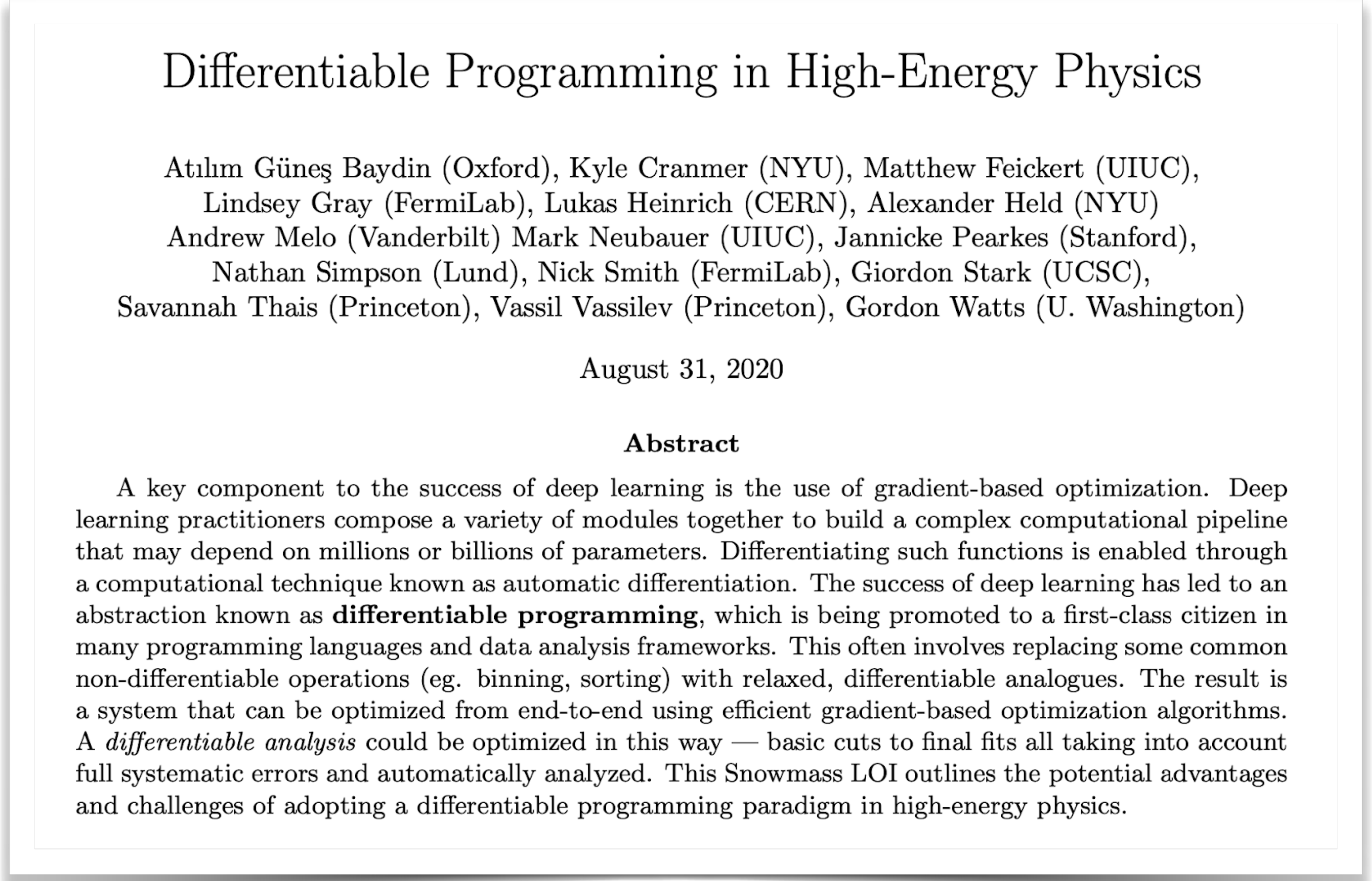
pyhf Workshop 2023 - 04.12.2023

[1] TU Dortmund University

technische universität dortmund

Differentiating through the fit

Another highlight of the last few years was the idea & demonstration of differentiating through the entire analyses, including the fit itself



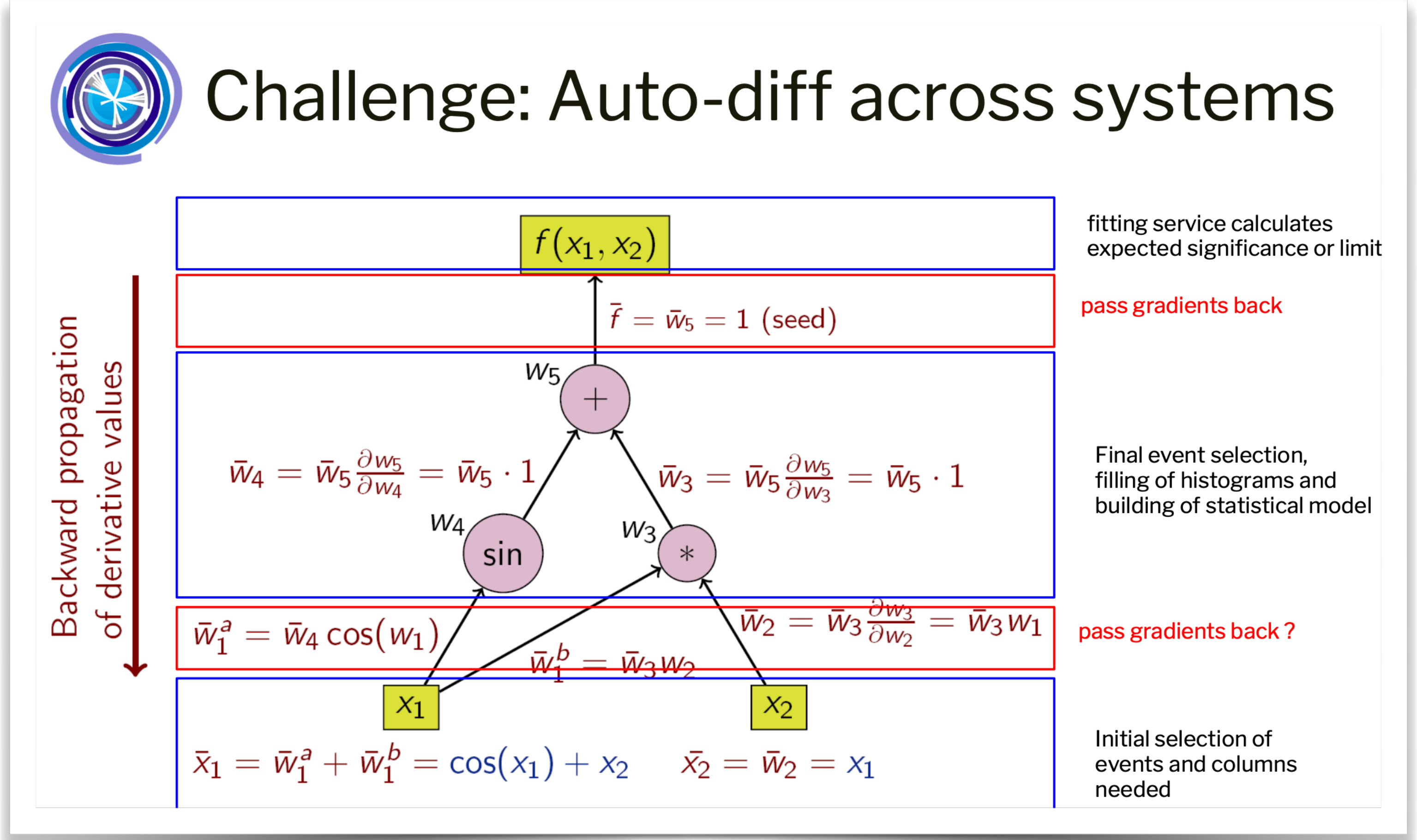
Differentiable Programming in High-Energy Physics

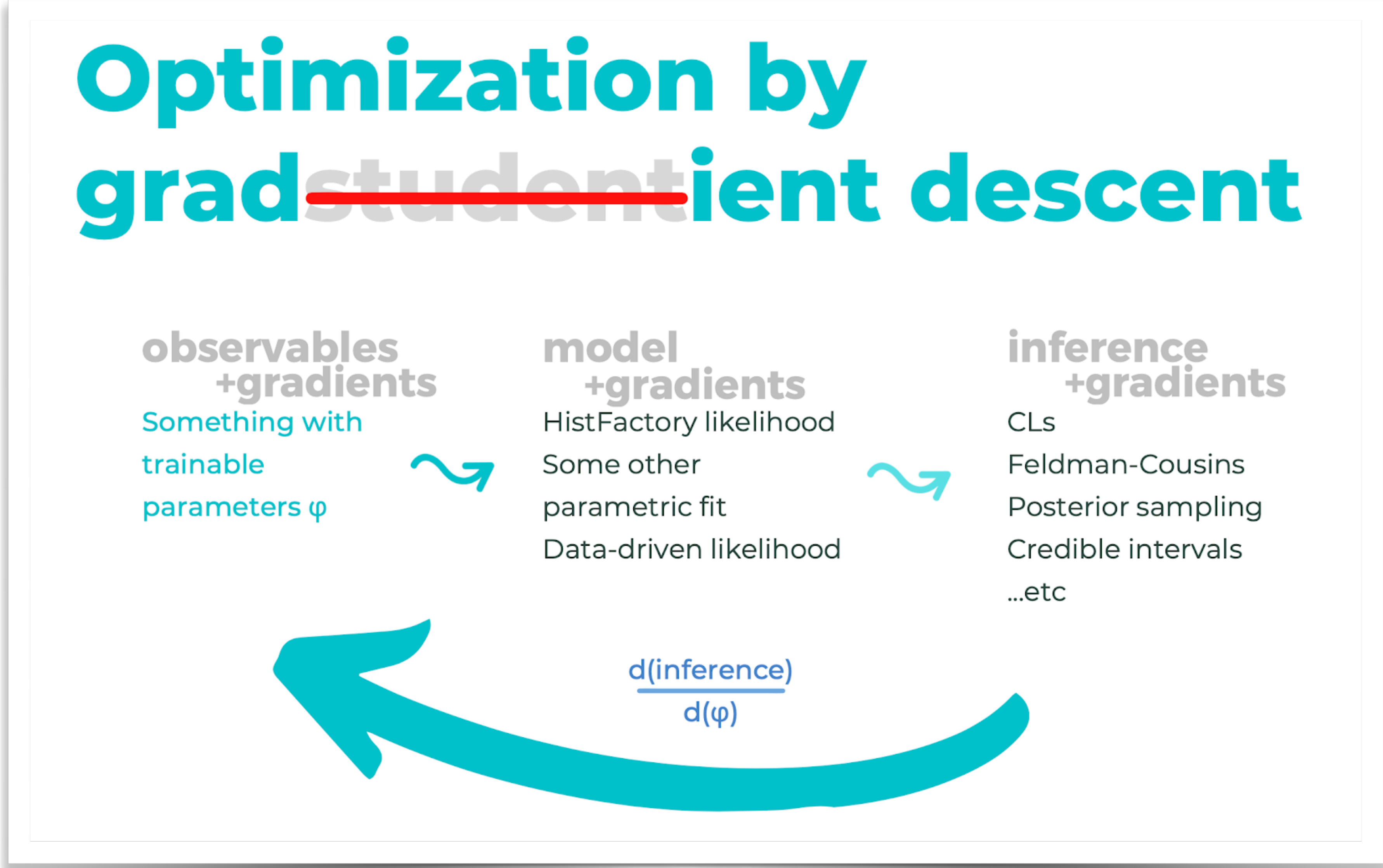
Atılım Güneş Baydin (Oxford), Kyle Cranmer (NYU), Matthew Feickert (UIUC), Lindsey Gray (FermiLab), Lukas Heinrich (CERN), Alexander Held (NYU) Andrew Melo (Vanderbilt) Mark Neubauer (UIUC), Jannicke Pearkes (Stanford), Nathan Simpson (Lund), Nick Smith (FermiLab), Giordon Stark (UCSC), Savannah Thais (Princeton), Vassil Vassilev (Princeton), Gordon Watts (U. Washington)

August 31, 2020

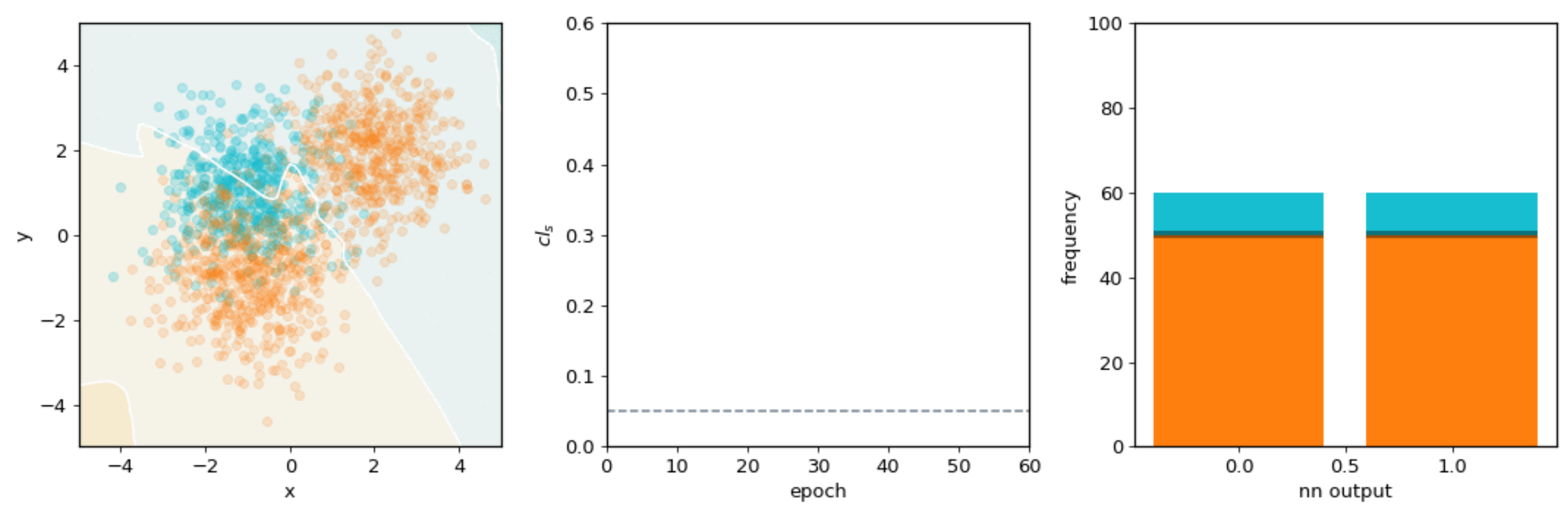
Abstract

A key component to the success of deep learning is the use of gradient-based optimization. Deep learning practitioners compose a variety of modules together to build a complex computational pipeline that may depend on millions or billions of parameters. Differentiating such functions is enabled through a computational technique known as automatic differentiation. The success of deep learning has led to an abstraction known as **differentiable programming**, which is being promoted to a first-class citizen in many programming languages and data analysis frameworks. This often involves replacing some common non-differentiable operations (eg. binning, sorting) with relaxed, differentiable analogues. The result is a system that can be optimized from end-to-end using efficient gradient-based optimization algorithms. A *differentiable analysis* could be optimized in this way — basic cuts to final fits all taking into account full systematic errors and automatically analyzed. This Snowmass LOI outlines the potential advantages and challenges of adopting a differentiable programming paradigm in high-energy physics.



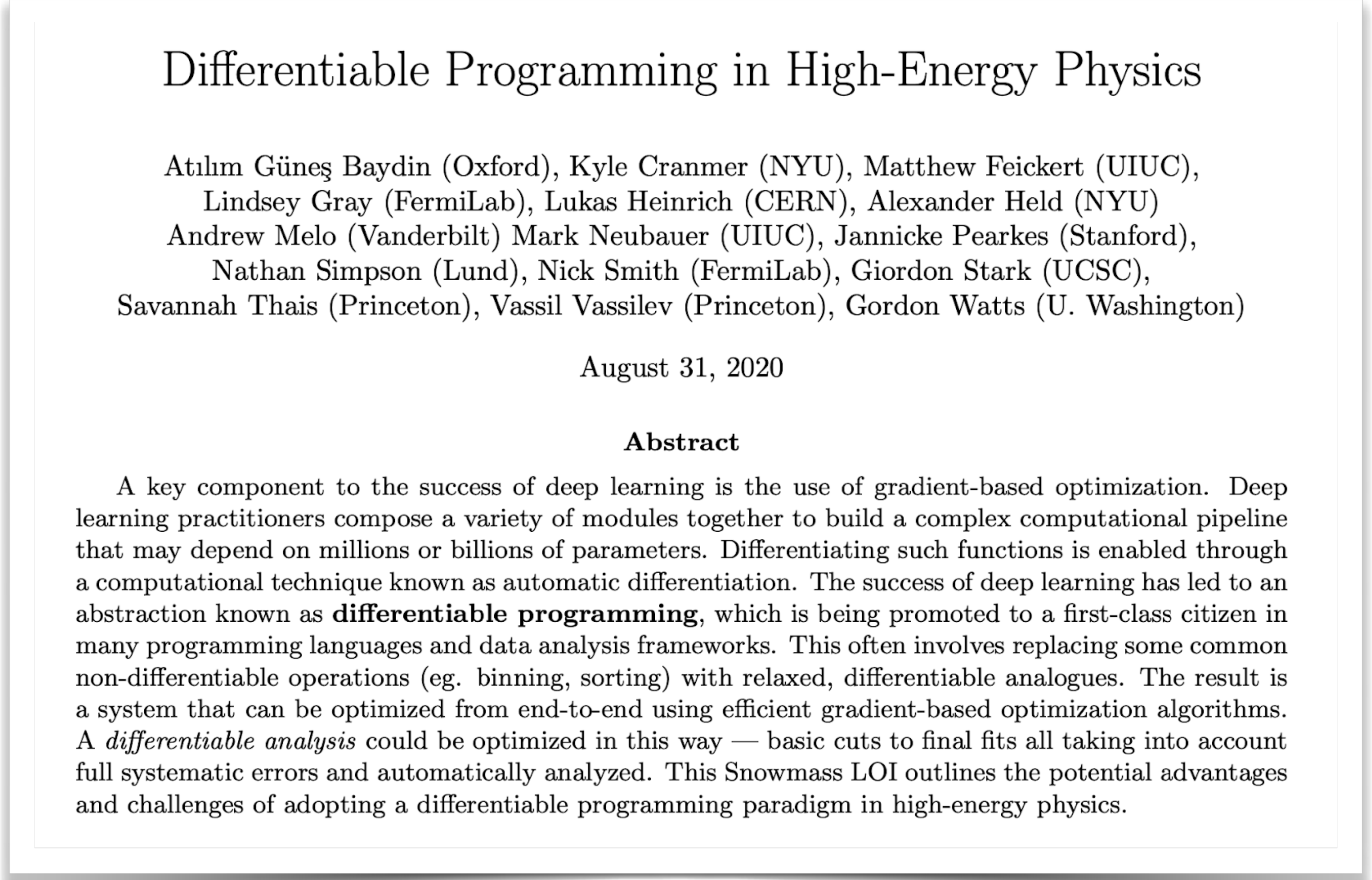


slide from Nathan Simpson: [link to talk]



Differentiating through the fit

Another highlight of the last few years was the idea & demonstration of differentiating through the entire analyses, including the fit itself

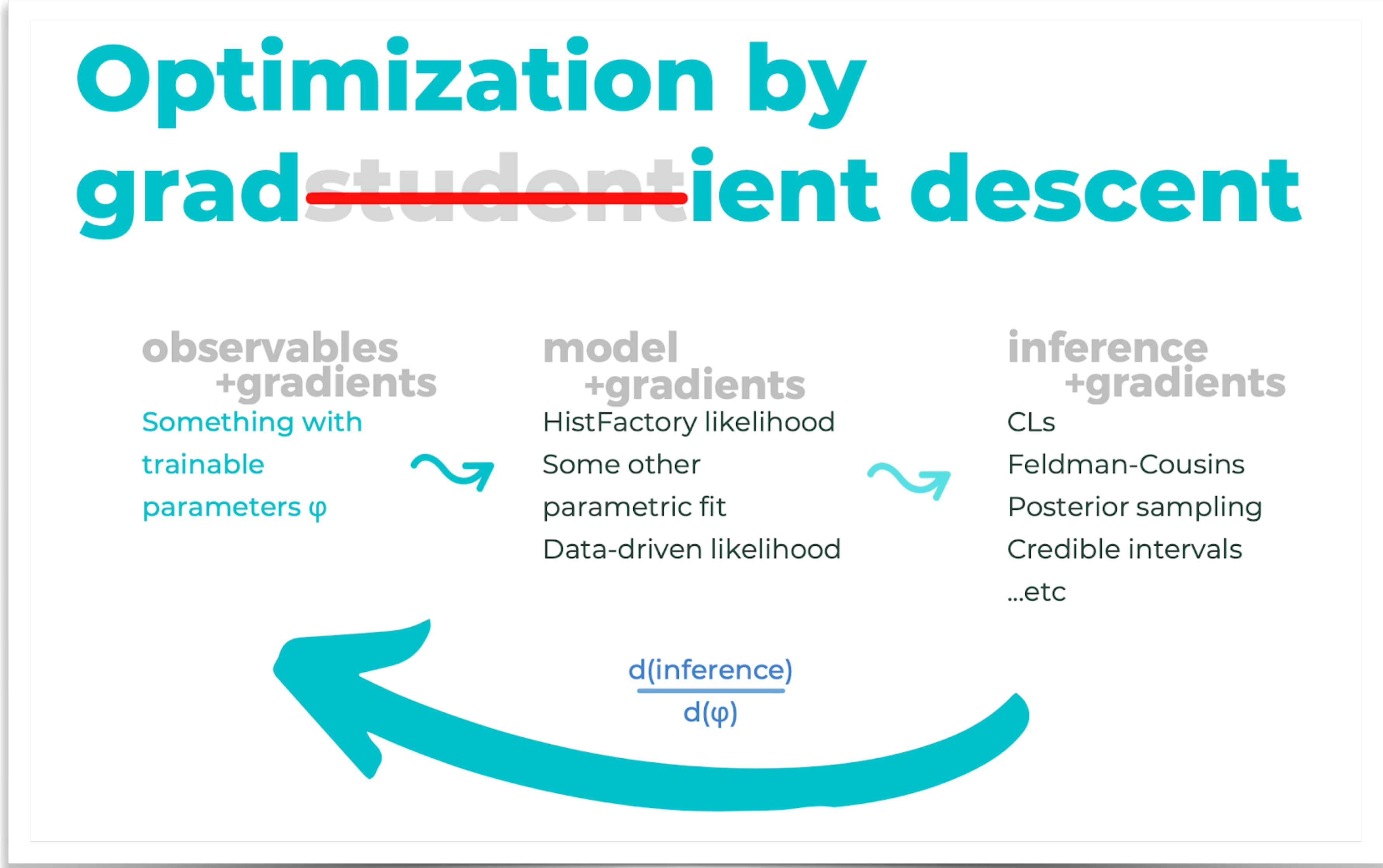

Differentiable Programming in High-Energy Physics

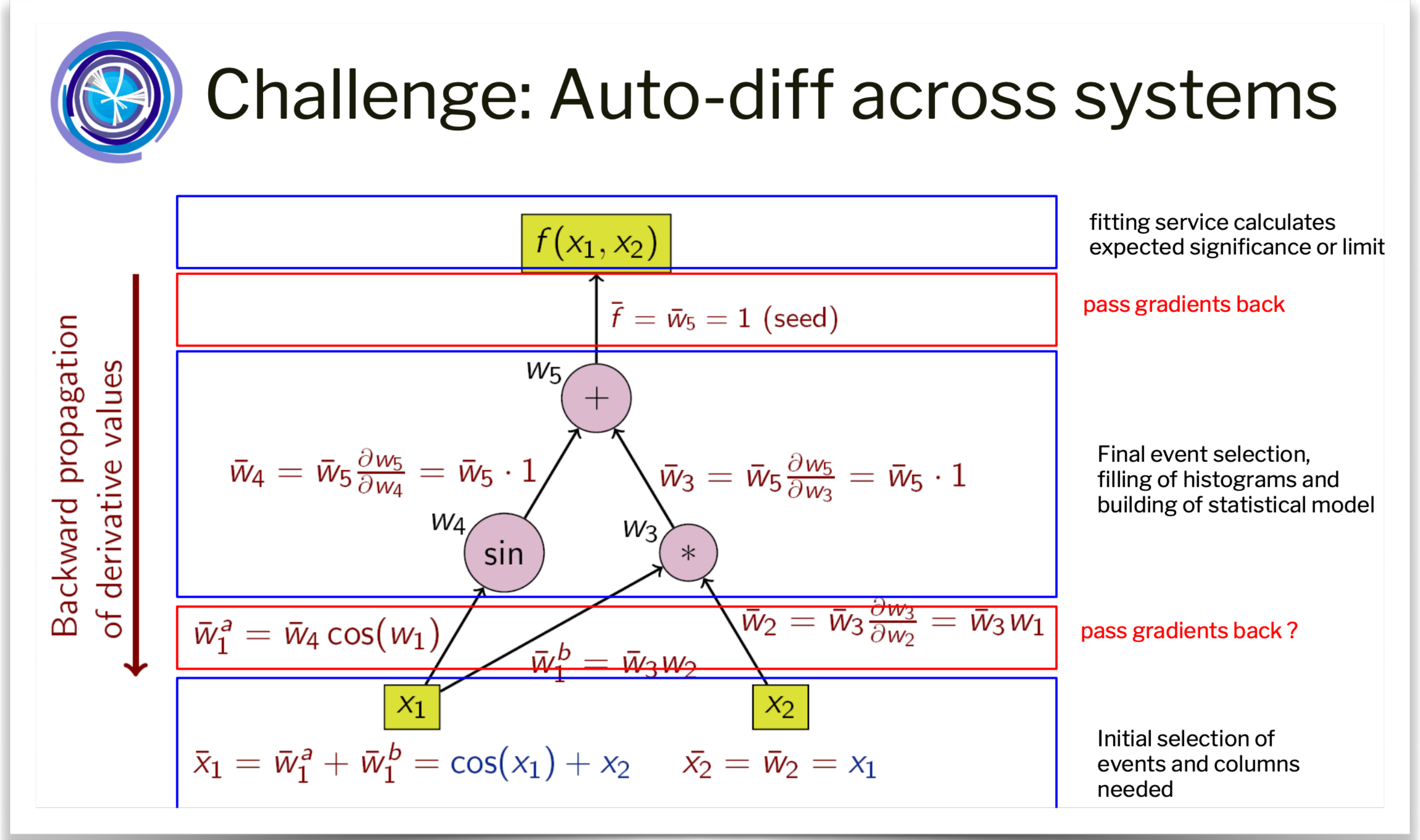
Atılım Güneş Baydin (Oxford), Kyle Cranmer (NYU), Matthew Feickert (UIUC), Lindsey Gray (FermiLab), Lukas Heinrich (CERN), Alexander Held (NYU) Andrew Melo (Vanderbilt) Mark Neubauer (UIUC), Jannicke Pearkes (Stanford), Nathan Simpson (Lund), Nick Smith (FermiLab), Giordon Stark (UCSC), Savannah Thais (Princeton), Vassil Vassilev (Princeton), Gordon Watts (U. Washington)

August 31, 2020

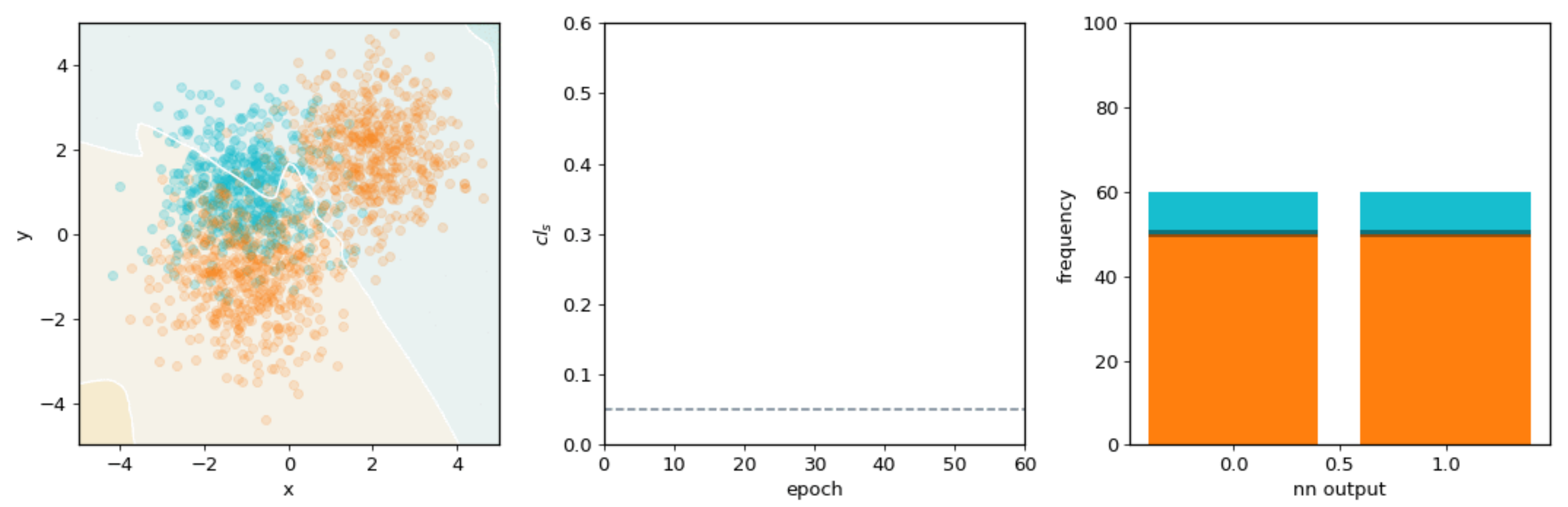
Abstract

A key component to the success of deep learning is the use of gradient-based optimization. Deep learning practitioners compose a variety of modules together to build a complex computational pipeline that may depend on millions or billions of parameters. Differentiating such functions is enabled through a computational technique known as automatic differentiation. The success of deep learning has led to an abstraction known as **differentiable programming**, which is being promoted to a first-class citizen in many programming languages and data analysis frameworks. This often involves replacing some common non-differentiable operations (eg. binning, sorting) with relaxed, differentiable analogues. The result is a system that can be optimized from end-to-end using efficient gradient-based optimization algorithms. A *differentiable analysis* could be optimized in this way — basic cuts to final fits all taking into account full systematic errors and automatically analyzed. This Snowmass LOI outlines the potential advantages and challenges of adopting a differentiable programming paradigm in high-energy physics.

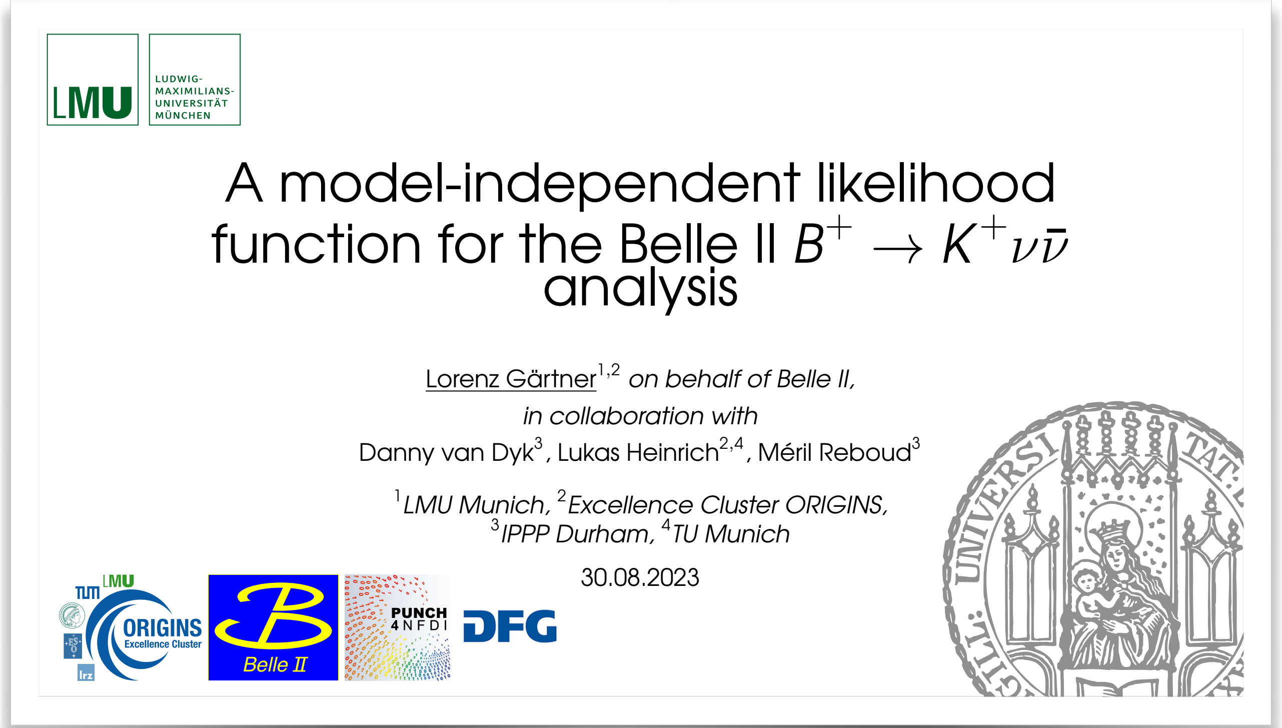


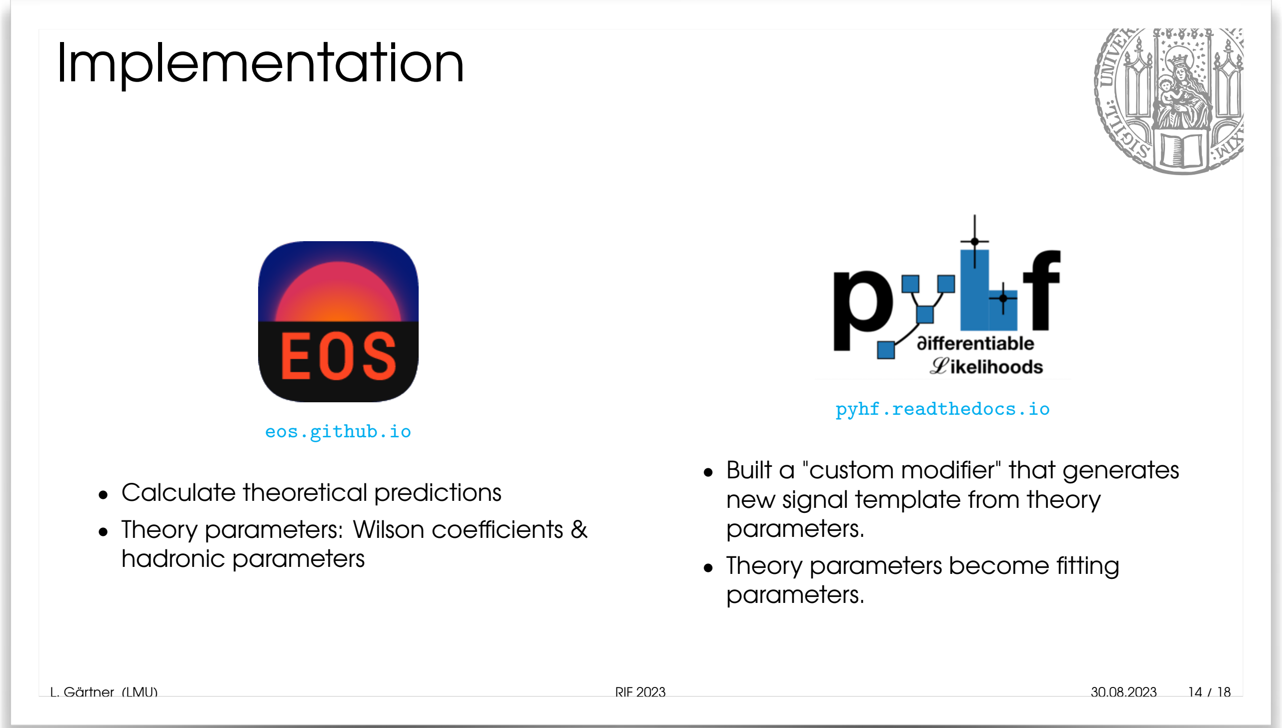


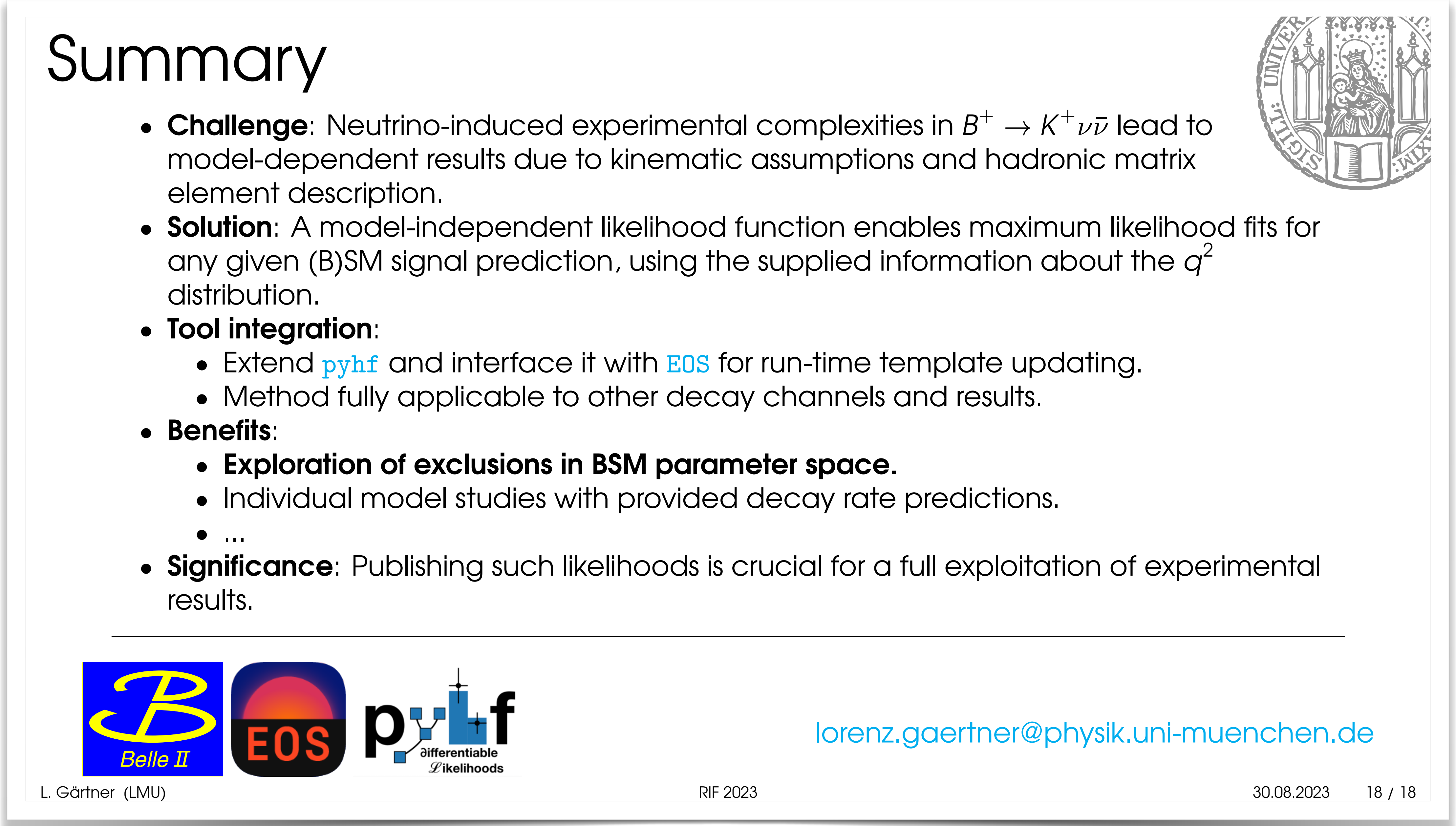
slide from Nathan Simpson: [link to talk]



Extending pyhf for EFT fits

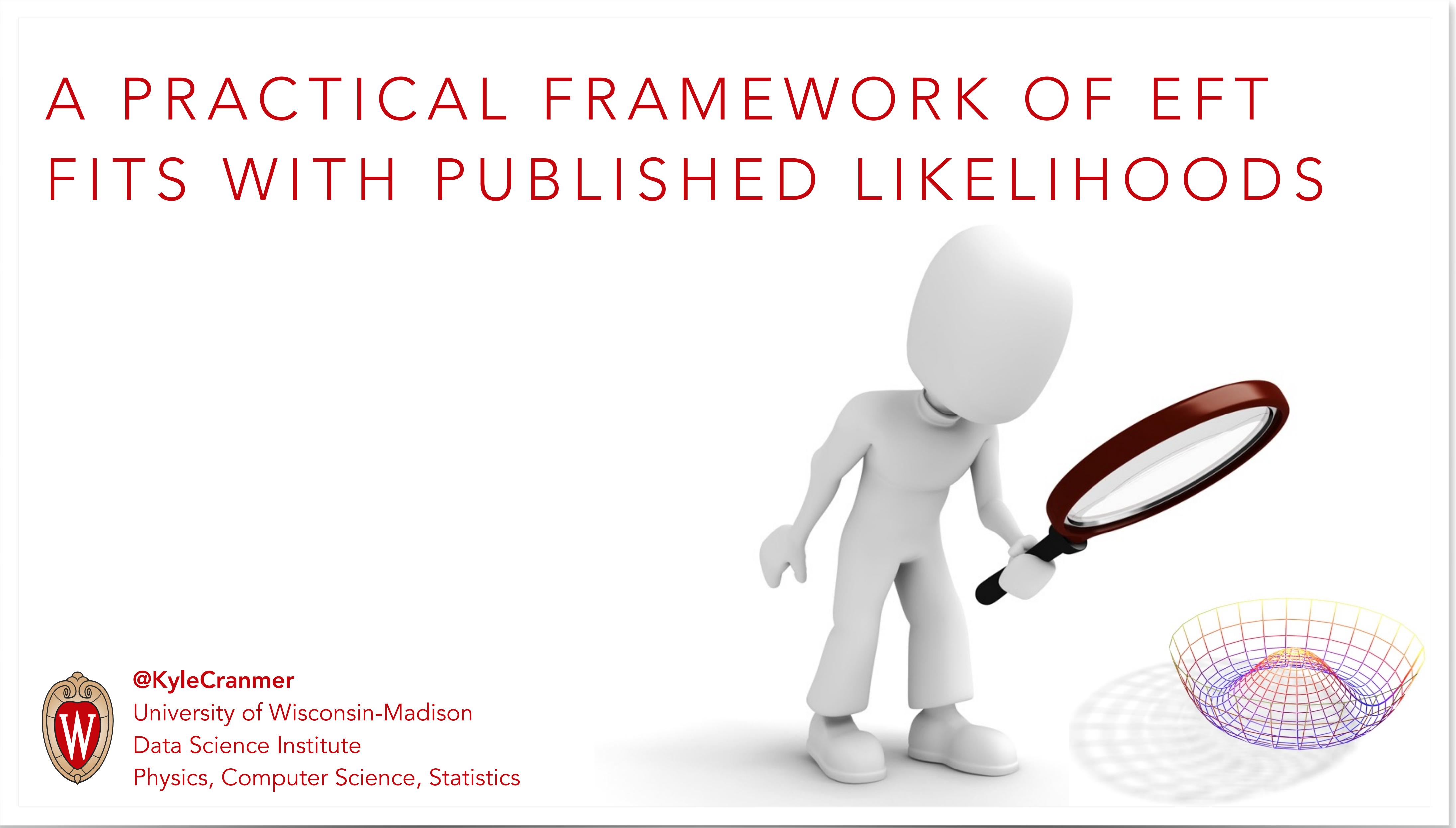






Happy to see pyhf getting attention beyond the LHC Reinterpretation Working Group

- Specifically the EFT working group
- Nice follow-up of BelleII result



https://conference.ippp.dur.ac.uk/event/1178/contributions/6443/
https://indico.cern.ch/event/1296757/timetable/#3-area-4-a-practical-framework

This workshop

pyhf Users and Developers Workshop 2023

4–8 Dec 2023
CERN
Europe/Zurich timezone

Enter your search term

- Overview
- Timetable
- Contribution List
- My Conference
  - My Contributions
- Registration
- Call for Abstracts
- Participant List
- Videoconference
- Code of Conduct
- Practical Information
  - Getting to/from CERN
  - Accommodation
  - Map
  - Internet/WiFi Access
  - CERN Restaurant Schedule

Contact

pyhf-workshop-organise…

pyhf ∂ifferentiable ℒikelihoods

What is this workshop?

This is the inaugural pyhf workshop for users and developers across all of physics (following the 2023 Belle II pyhf workshop). The goals of the workshop are twofold:

- Gather the pyhf user community together to learn from their experience and better understand what changes will improve things for the most number of people.

  We expect to have contributions from:
  - Scikit-HEP
  - IRIS-HEP
  - ATLAS
  - CMS (sorry this conflicts with CMS week)
  - Belle II
  - MicroBooNE
  - SModelS
  - MadAnalysis 5
  - and more

NUMFOCUS
OPEN CODE = BETTER SCIENCE

What’s next?

The big picture

Community white paper

Submitted to the Proceedings of the US Community Study
on the Future of Particle Physics (Snowmass 2021)

Data and Analysis Preservation, Recasting, and Reinterpretation

TF07 (Collider Phenomenology in the Theory Frontier)
COMPF7 (Reinterpretation and long-term preservation of data and code)

Stephen Bailey [1], Christian Bierlich [2], Andy Buckley [3], Jon Butterworth [4], Kyle Cranmer [5], Matthew Feickert [6*], Lukas Heinrich [7], Axel Huebl [1], Sabine Kraml [8‡], Anders Kvellestad [9], Clemens Lange [10], Andre Lessa [11], Kati Lassila-Perini [12], Christine Nattrass [13], Mark S. Neubauer [6], Sezen Sekmen [14], Giordon Stark [15], Graeme Watt [16]

1 Lawrence Berkeley National Laboratory, USA **2** Lund University, Lund, Sweden **3** University of Glasgow, UK **4** University College London, UK **5** New York University, USA **6** University of Illinois at Urbana-Champaign, USA **7** Technische Universität München, Germany **8** Univ. Grenoble Alpes, CNRS, Grenoble INP, LPSC-IN2P3, Grenoble, France **9** University of Oslo, Norway **10** Paul Scherrer Institute, Villigen, Switzerland **11** Universidade Federal do ABC, Brazil **12** Helsinki Institute of Physics, Finland **13** University of Tennessee, Knoxville, USA **14** Kyungpook National University, Korea **15** SCIPP, UC Santa Cruz, CA, USA **16** IPPP, Durham University, UK

Corresponding authors:
* matthew.feickert@cern.ch, ‡ sabine.kraml@lpsc.in2p3.fr

Snowmass Recommendations

7

Reinterpretation and Long-Term Preservation of Data and Code

Stephen Bailey [1], K. S. Cranmer [2], Matthew Feickert [2], Rob Fine [3], Sabine Kraml [4], Clemens Lange [5]

(and contributors from the community)

[1] Lawrence Berkeley National Lab, USA
[2] University of Wisconsin-Madison, USA
[3] Los Alamos National Lab, USA
[4] LPSC Grenoble, CNRS/IN2P3, France
[5] Paul Scherrer Institute, Villigen, Switzerland

Note: This report is based upon contributed white papers, letters of interest, and discussions with members of the community. The authors listed above made specific text contributions to this report or otherwise contributed white papers cited here and specifically opted-in to authorship. The full community that participated in discussions leading to this report is broader than just this author list.

P5 Report

High Energy Physics Advisory Panel

December 7-8, 2023
Westin Washington DC Downtown
999 9th St. NW, Washington DC 20001

	Thursday, December 7, 2023	
9:00 am	Convene	
9:00 - 9:30	Report from the DOE	Regina Rameika
9:30 - 9:45	Discussion	
9:45 - 10:15	Report from the NSF	C. Denise Caldwell
10:15 - 10:30	Discussion	
10:30 - 10:40	Break	
10:40 - 10:55	A Committee of Visitors for the HEP Facilities Division	Michael Procario, Young-Kee Kim
10:55 - 11:05	Discussion	
11:05 - 11:20	A Coordinating Panel for Software and Computing	Joel Butler
11:20 - 11:25	Discussion	
11:25 - 12:35	Lunch	
12:35 - 12:50	Remarks by the Director of the Office of Science	Asmeret Asefaw Berhe
12:50- 2:00	Presentation of the P5 Report	Hitoshi Murayama, Karsten Heeger
2:00 - 2:10	Break	
2:10 - 5:00	Discussion	
	Friday, December 8, 2023	
9:00 am	Convene	
9:00 - 10:30	Continued discussion of the P5 Report	
10:30 - 10:40	Break	
10:40 - 12:00	Continued discussion of the P5 Report	
12:00 - 13:30	lunch	
13:30 - 13:45	Panel vote on the P5 report	Sally Seidel
13:45 - 13:55	Communications with the Particle Physics Community	Sekhar Chivukula
13:55 - 14:00	General discussion	
14:00	Adjourn	

(Tune in Thursday!)

https://arxiv.org/abs/2203.10057
https://arxiv.org/abs/2209.08054
https://science.osti.gov/hep/hepap/Meetings

Conclusion:

It is amazing to see HistFactory / pyhf,
likelihood publishing, and RECAST thrive.

It's only possible due to the shared vision
and hard work of the community.

Keep up the great work!