Flavour anomalies and how to solve them

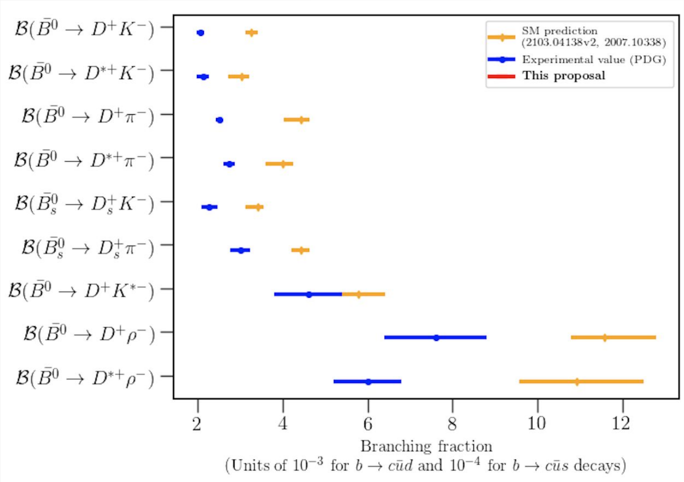
Big anomalies vs. big data

Nicole

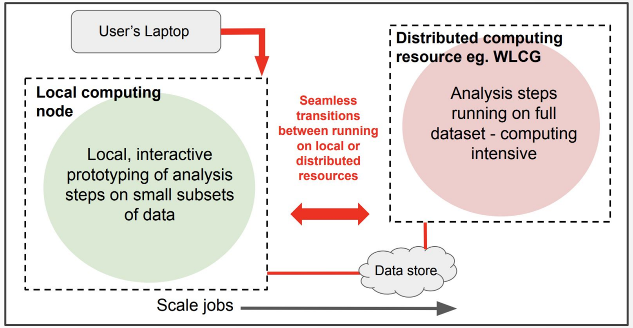
MWAPP
Oct 2023

My research

Flavour anomalies in B2OC

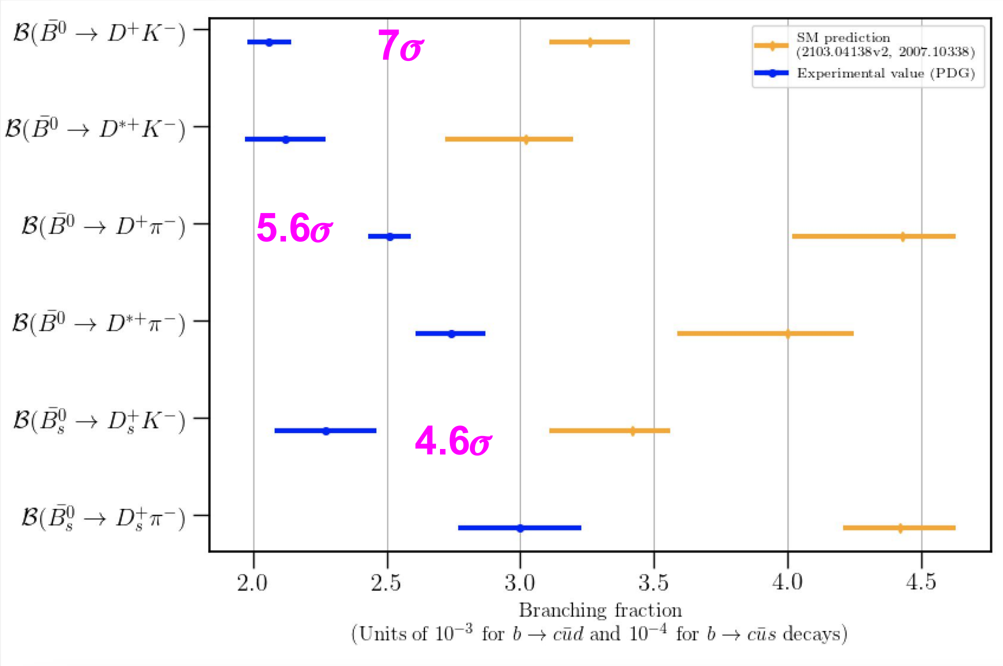


Data processing and analysis



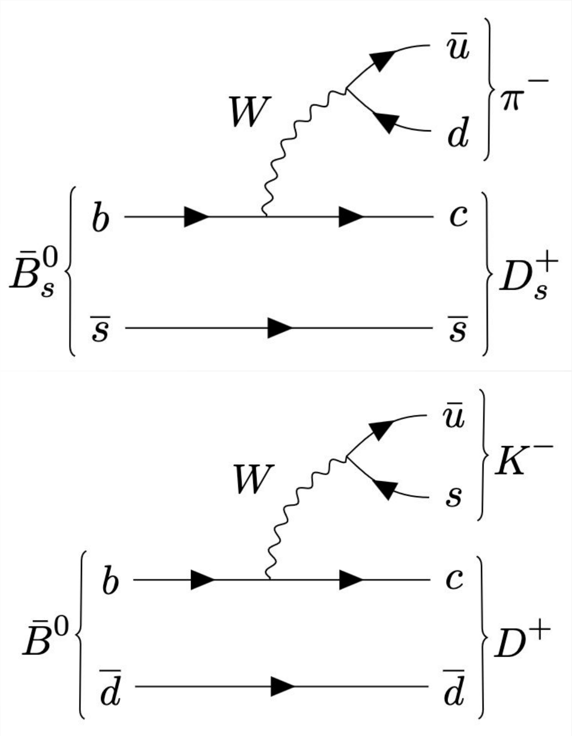
The b-decay rate anomaly

Rates of hadronic b-decays - 2023



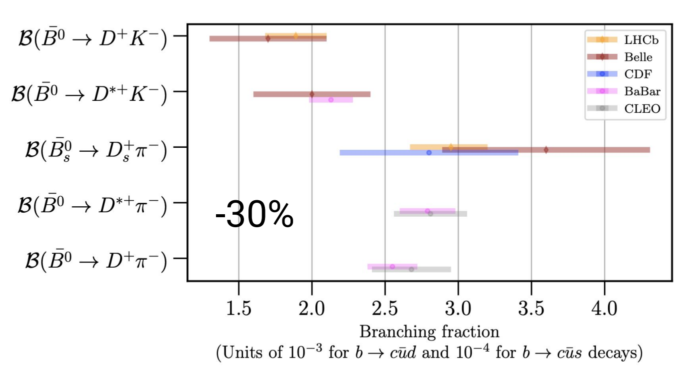
Reliable theory predictions - used extensively for matter-antimatter asymmetry studies

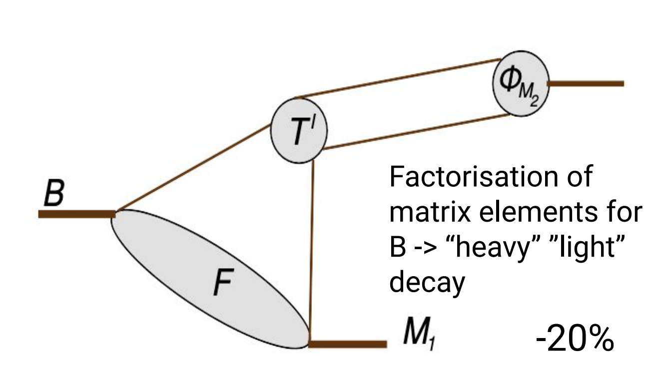
Golden modes



No penguin/annihilation process

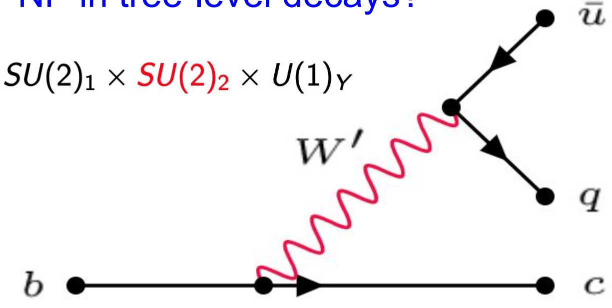
The b-decay rate anomaly



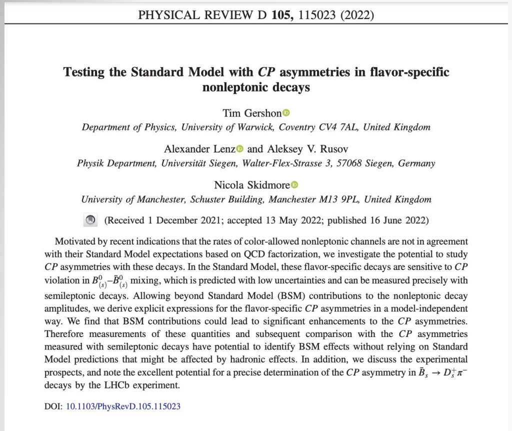


NP in tree-level decays?

$SU(2)_1 \times SU(2)_2 \times U(1)_Y$



CP asymmetries


PHYSICAL REVIEW D 105, 115023 (2022)

Testing the Standard Model with *CP* asymmetries in flavor-specific nonleptonic decays

Tim Gershon
Department of Physics, University of Warwick, Coventry CV4 7AL, United Kingdom

Alexander Lenz and Aleksey V. Rusov
Physik Department, Universität Siegen, Walter-Flex-Strasse 3, 57068 Siegen, Germany

Nicola Skidmore
University of Manchester, Schuster Building, Manchester M13 9PL, United Kingdom

(Received 1 December 2021; accepted 13 May 2022; published 16 June 2022)

Motivated by recent indications that the rates of color-allowed nonleptonic channels are not in agreement with their Standard Model expectations based on QCD factorization, we investigate the potential to study *CP* asymmetries with these decays. In the Standard Model, these flavor-specific decays are sensitive to *CP* violation in $B^0_{(s)}$–$\bar{B}^0_{(s)}$ mixing, which is predicted with low uncertainties and can be measured precisely with semileptonic decays. Allowing beyond Standard Model (BSM) contributions to the nonleptonic decay amplitudes, we derive explicit expressions for the flavor-specific *CP* asymmetries in a model-independent way. We find that BSM contributions could lead to significant enhancements to the *CP* asymmetries. Therefore measurements of these quantities and subsequent comparison with the *CP* asymmetries measured with semileptonic decays have potential to identify BSM effects without relying on Standard Model predictions that might be affected by hadronic effects. In addition, we discuss the experimental prospects, and note the excellent potential for a precise determination of the *CP* asymmetry in $\bar{B}_s \to D_s^+\pi^-$ decays by the LHCb experiment.

DOI: 10.1103/PhysRevD.105.115023

CP asymmetry measurements in the golden modes can probe BSM effects independent of theoretical uncertainties

$$A^q_{\rm fs}(D_q^+h^-)=\frac{\Gamma\left(B_q^0(t)\to D_q^+h^-\right)-\Gamma\left(\bar{B}_q^0(t)\to D_q^-h^+\right)}{\Gamma\left(B_q^0(t)\to D_q^+h^-\right)+\Gamma\left(\bar{B}_q^0(t)\to D_q^-h^+\right)}.$$

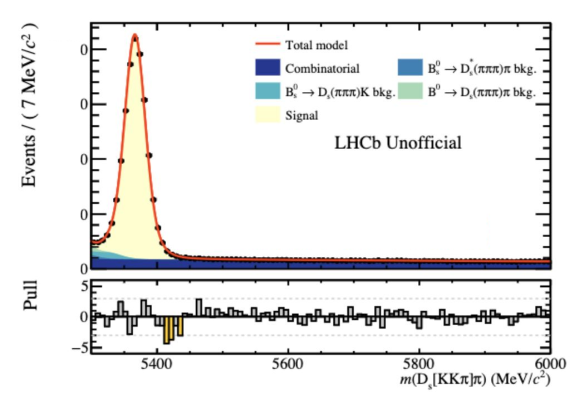
$\bar{B}_q \not\to D_q^- h^+$

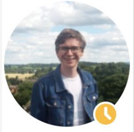
$B_q \not\to D_q^+ h^-$

$\bar{B}_q \to D_q^+ h^-$
$=$
$B_q \to D_q^- h^+$

LHCb run 3 has the best prospects for this BSM search

- These CP asymmetries have never been measured before
- Leading team of PhDs in Run 2 measurement with Bs->Dspi
  - Experimentally favourable
  - Mature state
- Run 3 + will provide unprecedented samples of these decays providing sensitivities to BSM effects at 6x10^{-4}





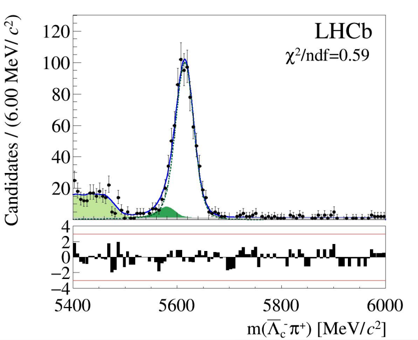
Early measurements

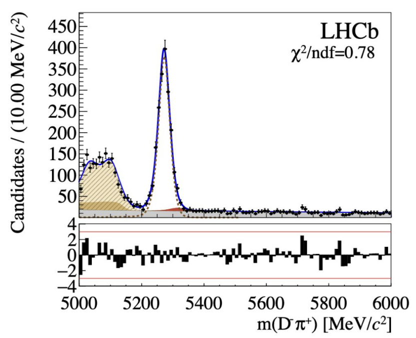
"LHCb Upgrade is a new detector"

Verify LHCb performance by measuring known observables with early Run 3 data

- ERC team measuring b-hadron production cross sections using hadronic decays
  - First hadronic measurements of these observables thanks to Upgrade trigger strategy
- Competitive with muon modes with 100pb^{-1}

Automated GitLab analysis pipelines run periodically - important when working with evolving LHCb projects

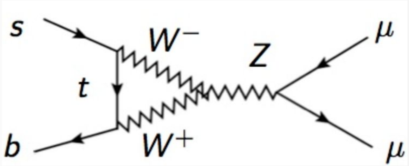


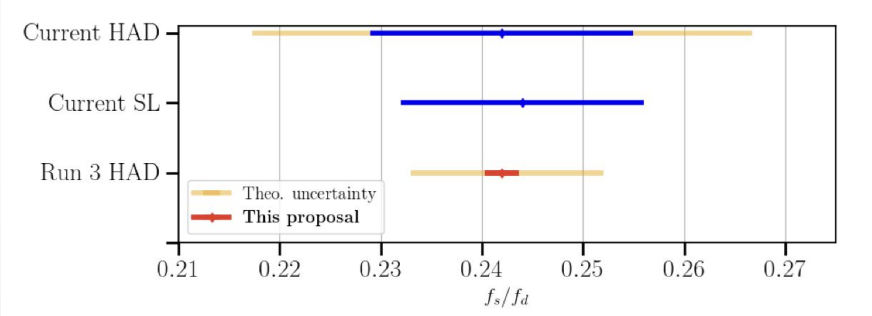


Early measurements

Lead an extension to the project - measurement of fs/fd using hadronic decays

- Both early measurement and full Run 3 analysis
- Key input parameter and limiting systematic to BF(B_s->μμ)
- Current value dominated by SL decays
    - Hadronic measurement sensitive to low pT

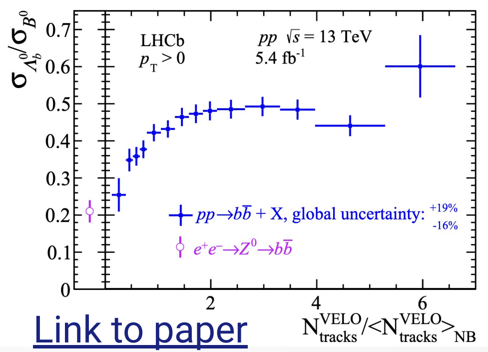




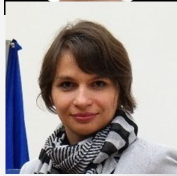
Factor 8 increase in experimental sensitivity and theoretical uncertainty reduction to 5% will allow a comparison with the SL measurement for the first time a which is itself a test of BSM in hadronic b decays

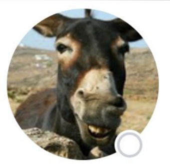
Valeriia has joined the team :)

- Study fLb/fd and its dependence on multiplicity



Link to paper





Data Processing and Analysis

Run 1
0.3GB/s to storage

Run 2
0.65GB/s to storage

Stripping

LS2 → U1

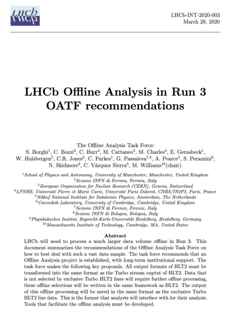
Run 3
10GB/s to storage

Sprucing

Run 3 increase in data rate necessitated re-design of LHCb **Offline** (post trigger) activities

Offline Analysis Task Force

Assess offline requirements for Run 3 and beyond



LHCb-INT-2020-003
March 28, 2020

LHCb Offline Analysis in Run 3 OATF recommendations

The Offline Analysis Task Force:
S. Borghi[1], C. Bozzi[2], C. Burr[3], M. Cattaneo[3], M. Charles[4], E. Gersabeck[1], W. Hulsbergen[5], C.R. Jones[6], C. Parkes[1], G. Passaleva[7,3], A. Pearce[3], S. Perazzini[8], N. Skidmore[9], C. Vázquez Sierra[5], M. Williams[10](chair).

[1] *School of Physics and Astronomy, University of Manchester, Manchester, United Kingdom*
[2] *Sezione INFN di Ferrara, Ferrara, Italy*
[3] *European Organization for Nuclear Research (CERN), Geneva, Switzerland*
[4] *LPNHE, Université Pierre et Marie Curie, Université Paris Diderot, CNRS/IN2P3, Paris, France*
[5] *Nikhef National Institute for Subatomic Physics, Amsterdam, The Netherlands*
[6] *Cavendish Laboratory, University of Cambridge, Cambridge, United Kingdom*
[7] *Sezione INFN di Firenze, Firenze, Italy*
[8] *Sezione INFN di Bologna, Bologna, Italy*
[9] *Physikalisches Institut, Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany*
[10] *Massachusetts Institute of Technology, Cambridge, MA, United States*

Abstract

LHCb will need to process a much larger data volume offline in Run 3. This document summarizes the recommendations of the Offline Analysis Task Force on how to best deal with such a vast data sample. The task force recommends that an Offline Analysis project is established, with long-term institutional support. The task force makes the following key proposals. All output formats of HLT2 must be transformed into the same format as the Turbo stream ouptut of HLT2. Data that is not selected by exclusive Turbo HLT2 lines will require further offline processing, these offline selections will be written in the same framework as HLT2. The output of this offline processing will be saved in the same format as the exclusive Turbo HLT2 line data. This is the format that analysts will interface with for their analysis. Tools that facilitate the offline analysis must be developed.

LHCb-INT-2020-009
May 21, 2020

Data Processing & Analysis (DPA) Project

C. Burr[1], P. Koppenburg[2], D. Lucchesi[3], E. Rodrigues[4], N. Skidmore[5], A. Tully[6]

[1] *European Organization for Nuclear Research (CERN), Geneva, Switzerland*
[2] *Nikhef National Institute for Subatomic Physics, Amsterdam, Netherlands*
[3] *University and INFN Padova, Padova, Italy*
[4] *Oliver Lodge Laboratory, University of Liverpool, Liverpool, United Kingdom*
[5] *Universität Bonn, Bonn, Germany*
[6] *Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland*

Abstract

This note introduces the Data Processing & Analysis (DPA) project recommended by the Offline Analysis Task Force [1]. It presents the proposed organisation of the project in terms of work packages, prioritised deliverables, and person-power required to carry out the programme.

Remit post-trigger to papers

Third year of operation

LHCb DPA deputy PL

40 members from 20 institutes

DPA has a huge remit and person-power is always required on some very cool projects

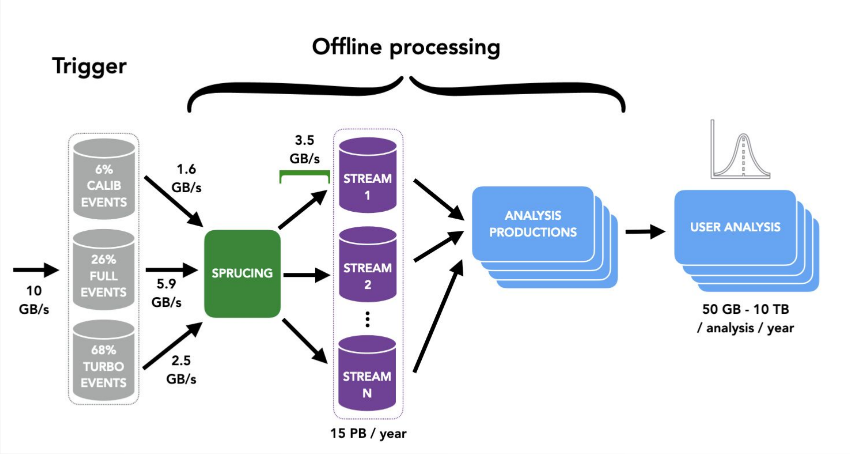
- Exploiting GPU farm for end-user analysis (analysis facilities)
- Open data - nTuple Wizard (UI design)
- LHCb Grand Analysis Challenge
- Unifying frameworks of Allen and Gaudi for HL-LHC - resource independent API (shared with RTA)

DPA will be where the next advancements will take place

Data Processing and Analysis

Fundamental component of re-design is Offline data processing and selection model

Developed the Sprucing for LHCb Run 3 and beyond



Tape **Disk**

Sprucing successfully running on 2023 data

- Bridge between Online and Offline Sectors
- Reduces the data bandwidth from 10 GB/s to 3.5 GB/s through a mix of event slimming, event skimming, creation of metadata stores summarising file content and data compression depending on stream
- Balance cheaper tape against instantly accessible disk storage
- Allows preservation of full event reconstruction despite increase in data rate
    - Essential for data mining in future
- Unify codebase of Online and Offline sectors

Sprucing

Spurcing status

Sprucing campaign status tracker

Link to documentation

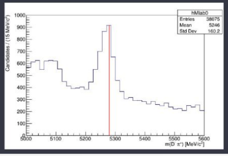
Passthrough Sprucing on TURBO and TURCAL streams

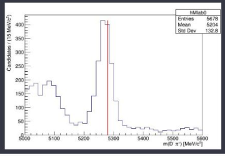
- **Commissioning23** "7-8th July data" ✅
  /LHCb/Commissioning23/Beam6800GeV-VeloClosed-MagDown-Excl-UT/Real Data/SprucingPass23
- **Collision23** ✅
  /LHCb/Collision23/Beam6800GeV-VeloClosed-MagDown-Excl-UT/Real Data/SprucingPass23r1/

Exclusive Sprucing on FULL stream

- **Commissioning23** "7-8th July data" ✅
  /LHCb/Commissioning23/Beam6800GeV-VeloClosed-MagDown-Excl-UT/Real Data/Sprucing23
- **Collision23** ✅
  /LHCb/Collision23/Beam6800GeV-VeloClosed-MagDown-Excl-UT/Real Data/Sprucing23r1/

Alessandro Bertolin 2:03 PM
good afternoon all, this is that I get from an AP running on the B2OC Bd2DPi exclusive spruce line (copy/paste of candidates):





For exclusive Sprucing on FULL streams

Data processed	Moore MR label	HLT2 Moore	Sprucing Moore	processing_pass for exclusive (full/turcal)	Status	BKK path
Collision22			vxyrz	Spurcing22	In preparation	
7-8th July	Spruce_June23	v54r12	v54r19	Spurcing23	Delivered	/LHCb/Commissioning23/Beam6800GeV-VeloClosed-MagDown-Excl-UT/Real Data/Spurcing23/
Rest of 2023 (not 7-8th, COLLISION23)		v54r16	v54r19	Spurcing23r1	Delivered	/LHCb/Collision23/Beam6800GeV-VeloClosed-MagDown-Excl-UT/Real Data/Spurcing23r1/
SMOG 23 data	Spruce_SMOG	vxyrz	vxyrz			

For Passthrough Sprucing on TURBO/TURCAL streams

Data processed	Moore MR label	HLT2 Moore	Sprucing Moore	processing_pass for passthrough (turbo/turcal)	Status	BKK path
Collision22		vxyrz	vxyrz	SprucingPass22	In preparation	
7-8th July		v54r12	v54r15	SprucingPass23	Delivered	/LHCb/Commissioning23/Beam6800GeV-VeloClosed-MagDown-Excl-UT/Real Data/SprucingPass23
Rest of 2023 (not 7-8th, COLLISION23)		v54r16	v54r18	SprucingPass23r1	Delivered	/LHCb/Collision23/Beam6800GeV-VeloClosed-MagDown-Excl-UT/Real Data/SprucingPass23r1/
SMOG 23 data	Spruce_SMOG	vxyrz	vxyrz			

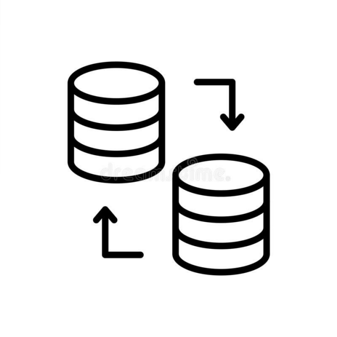
HEP Software Foundation (HSF)

hepsoftwarefoundation.org/

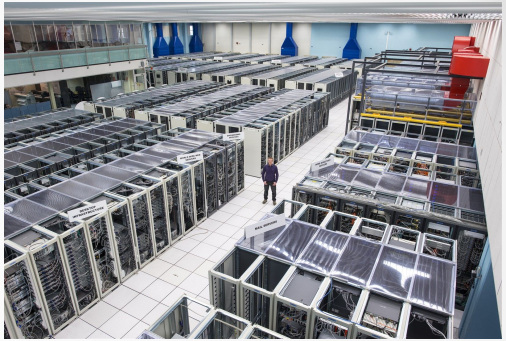
High Energy Physics has a vast investment in software

50M lines of C++
Worth 500M$

1M CPU cores every hour



100PB of data
transfers per year



1000PB of data

LHC and non-LHC experiments face the same software challenges

- Evolve to meet these challenges and overcome limitations
- Exploit expertise inside <u>and</u> outside our community
- Cannot afford duplicated efforts

HEP Software Foundation (HSF)

HEP Software Foundation

hepsoftwarefoundation.org/

Data Analysis

Reconstruction and Software Triggers

Training

"The HEP Software Foundation (HSF) facilitates collaborations in High Energy Physics software and computing internationally and consists of over 700 scientists, software experts and technicians."

Detector simulation

Frameworks

Software Developer tools and packaging

Physics generators

Analysis facilities forum

HEP Software Foundation (HSF)

Aim to make the publication of physics results efficient both in terms of human and computing resources

- 150+ members
- Bridge between physics analysts and technical innovators inside and outside of HEP
- Foster optimal big-data solutions to HEP challenges

Computing and Software for Big Science (2022) 6:13
https://doi.org/10.1007/s41781-022-00086-2

ORIGINAL ARTICLE

Constraints on Future Analysis Metadata Systems in High Energy Physics

T. J. Khoo[5] · A. Reinsvold Hall[10] · N. Skidmore[16] · S. Alderweireldt[15] · J. Anders[13] · C. Burr[3] · W. Buttinger[9] · P. David[11] · L. Gouskos[3] · L. Gray[4] · S. Hageböck[3] · A. Krasznahorkay[3] · P. Laycock[1] · A. Lister[14] · Z. Marshall[6] · A. B. Meyer[2] · T. Novak[2] · S. Rappoccio[12] · M. Ritter[7] · E. Rodrigues[8] · J. Rumsevicius[3] · L. Sexton-Kennedy[4] · N. Smith[4] · G. A. Stewart[3] · S. Wertz[11]

Received: 15 December 2021 / Accepted: 14 June 2022 / Published online: 27 July 2022

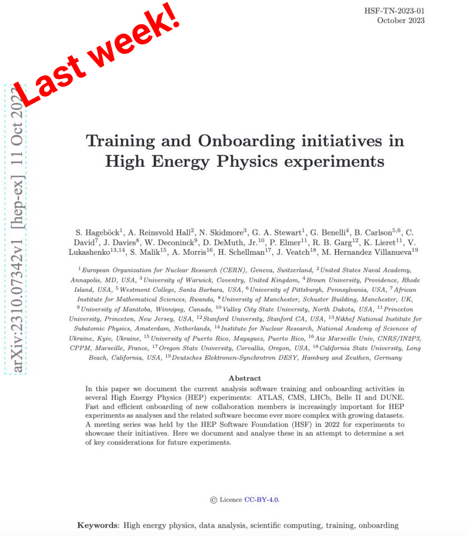
Abstract
In high energy physics (HEP), analysis metadata comes in many forms—from theoretical cross-sections, to calibration corrections, to details about file processing. Correctly applying metadata is a crucial and often time-consuming step in an analysis, but designing analysis metadata systems has historically received little direct attention. Among other considerations, an ideal metadata tool should be easy to use by new analysers, should scale to large data volumes and diverse processing paradigms, and should enable future analysis reinterpretation. This document, which is the product of community discussions organised by the HEP Software Foundation, categorises types of metadata by scope and format and gives examples of current metadata solutions. Important design considerations for metadata systems, including sociological factors, analysis preservation efforts, and technical factors, are discussed. A list of best practices and technical requirements for future analysis metadata systems is presented. These best practices could guide the development of a future cross-experimental effort for analysis metadata tools.

Keywords High energy physics · Data analysis · Metadata · Scientific computing · Databases

✉ S. Hageböck
stephan.hageboeck@cern.ch

A. Reinsvold Hall
achall@usna.edu

1 Physics Department, Brookhaven National Laboratory, Upton, NY, USA
2 Deutsches Elektronen-Synchrotron DESY, Hamburg, Germany
3 European Organization for Nuclear Research (CERN), Geneva, Switzerland
4 Fermi National Accelerator Laboratory, Batavia, IL, USA
5 Humboldt-Universität zu Berlin, Institut für Physik, 12489 Berlin, Germany
6 Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
7 Ludwig-Maximilians-Universität, Munich, Germany
8 Oliver Lodge Laboratory, University of Liverpool, Liverpool, UK
9 Rutherford Appleton Laboratory, Didcot OX11 0DE, UK
10 United States Naval Academy, Annapolis, MD, USA
11 Centre for Cosmology, Particle Physics and Phenomenology (CP3), Université Catholique de Louvain, Louvain-la-Neuve, Belgium
12 University at Buffalo, State University of New York, Amherst, NY, USA
13 Laboratory for High Energy Physics, University of Bern, Bern, Switzerland
14 University of British Columbia, Vancouver, BC, Canada
15 The University of Edinburgh, Edinburgh, UK
16 University of Manchester, Schuster Building, Manchester M13 9PL, UK



HSF-TN-2023-01
October 2023

arXiv:2310.07342v1 [hep-ex] 11 Oct 2023

Training and Onboarding initiatives in High Energy Physics experiments

S. Hageböck[1], A. Reinsvold Hall[2], N. Skidmore[3], G. A. Stewart[1], G. Benelli[4], B. Carlson[5,6], C. David[7], J. Davies[8], W. Deconinck[9], D. DeMuth, Jr.[10], P. Elmer[11], R. B. Garg[12], K. Lieret[11], V. Lukashenko[13,14], S. Malik[15], A. Morris[16], H. Schellman[17], J. Veatch[18], M. Hernandez Villanueva[19]

[1] *European Organization for Nuclear Research (CERN), Geneva, Switzerland,* [2] *United States Naval Academy, Annapolis, MD, USA,* [3] *University of Warwick, Coventry, United Kingdom,* [4] *Brown University, Providence, Rhode Island, USA,* [5] *Westmont College, Santa Barbara, USA,* [6] *University of Pittsburgh, Pennsylvania, USA,* [7] *African Institute for Mathematical Sciences, Rwanda,* [8] *University of Manchester, Schuster Building, Manchester, UK,* [9] *University of Manitoba, Winnipeg, Canada,* [10] *Valley City State University, North Dakota, USA,* [11] *Princeton University, Princeton, New Jersey, USA,* [12] *Stanford University, Stanford CA, USA,* [13] *Nikhef National Institute for Subatomic Physics, Amsterdam, Netherlands,* [14] *Institute for Nuclear Research, National Academy of Sciences of Ukraine, Kyiv, Ukraine,* [15] *University of Puerto Rico, Mayaguez, Puerto Rico,* [16] *Aix Marseille Univ, CNRS/IN2P3, CPPM, Marseille, France,* [17] *Oregon State University, Corvallis, Oregon, USA,* [18] *California State University, Long Beach, California, USA,* [19] *Deutsches Elektronen-Synchrotron DESY, Hamburg and Zeuthen, Germany*

Abstract
In this paper we document the current analysis software training and onboarding activities in several High Energy Physics (HEP) experiments: ATLAS, CMS, LHCb, Belle II and DUNE. Fast and efficient onboarding of new collaboration members is increasingly important for HEP experiments as analyses and the related software become ever more complex with growing datasets. A meeting series was held by the HEP Software Foundation (HSF) in 2022 for experiments to showcase their initiatives. Here we document and analyse these in an attempt to determine a set of key considerations for future experiments.

Keywords: High energy physics, data analysis, scientific computing, training, onboarding

Next meeting on HS3

HS3 standardizes the machine- and human-readable serialization of all components involved in model fitting as used in High Energy Physics. This includes the ***definition of the model, the data as well as the loss function.***

Particularly important for LHCb!

Analysis Facilities

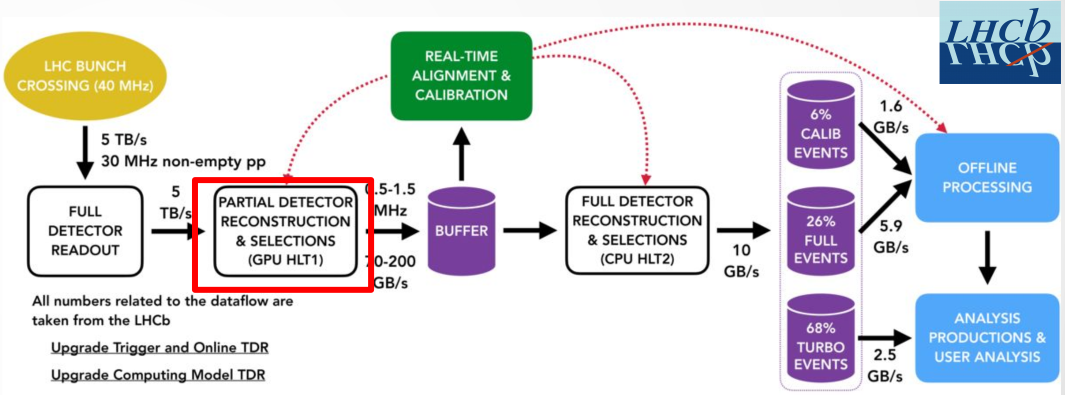
HL-LHC era (2029---) will produce unprecedented scientific data volumes at multi-exabyte scale to be analysed

- Other big-data scientific experiments such as DUNE and SKA also come online in next decade
- Current methods will not scale with event count

Significant R&D has taken place for detector technologies and computing for online data processing

Almost no R&D into end-user analysis

Urgent R&D into end-user analysis is required so that this does not become the bottleneck that limits the return of future scientific endeavours



HL-LHC Software and Computing Review Panel Report 2022

For the HL-LHC, the analysis model will continue to evolve, driven by significantly more data, and the experiments should consider becoming more involved in strategies for end-user analysis including the development of tools.

Analysis Facilities

Integrated Analysis Facilities (AF) can be the way forward

"Infrastructure providing the data, software and computational resources to execute an analysis workflow. … shared and supported through virtual organization."

AFs are a focus at the highest levels

Norfolk, Virginia, USA • May 8-12, 2023 CHEP 2023 Computing in High Energy & Nuclear Physics

154th LHCC Meeting

"The LHCC recommends that experiments engage in the process of developing and defining the structure of the future Analysis Facilities"


1

Analysis Facilities White Paper

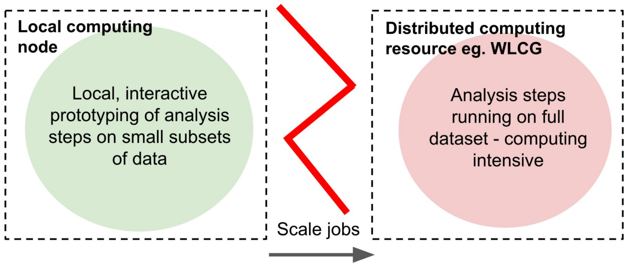
Diego Ciangottini (INFN Perugia, Editor), Alessandra Forti (University of Manchester, Editor), Lukas Heinrich (Technische Universität München, Editor), Nicola Skidmore (University of Manchester, Editor),

Abstract

This whitepaper presents the current status of the R&D for analysis facilities and attempts to summarize the current views on the future direction of these facilities. These views have been collected through the High Energy Physics (HEP) Software Foundation's (HSF) Analysis Facilities forum, established in March 2022, and through the Analysis Ecosystems II workshop that took place in May 2022.

Introduction

In the HL-LHC era (to begin in 2029) LHC analysts will have over an order of magnitude more data to analyze than currently. This will require significant changes to analyses techniques and workflows some of which may be better suited to run using cloud technologies and heterogeneous resources. As a consequence the infrastructure used to carry out analysis must



Local and distributed analysis worlds fundamentally disjoint

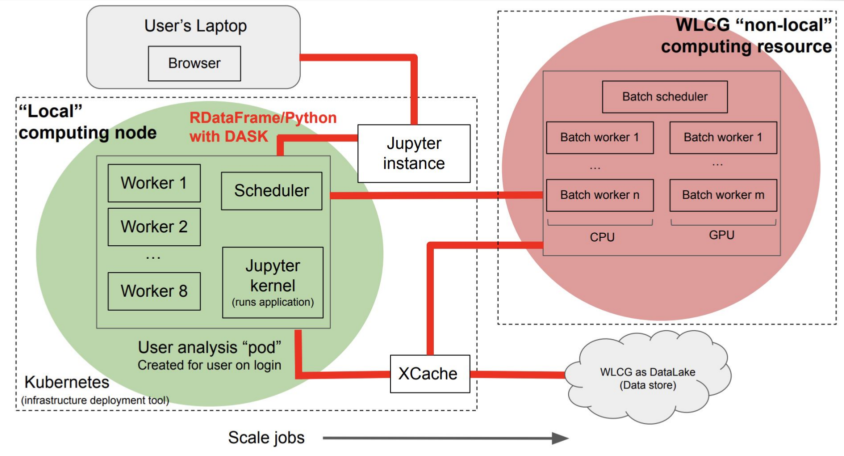
- Disrupt analysis workflows - inefficient analysis
- Submission to distributed resources unnecessarily complex - users do not exploit resources fully

AFs can create the required link between local and distributed worlds allowing seamless scaling

Analysis Facilities

- Instances of facilities exist but have fundamental limitations
  - US ATLAS and CMS groups - LHC GPD specific
  - Scale to on-site resources at AF host labs - have to be member of US ATLAS VO

- ERC proposal to create cross-experiment, packaged analysis facility that uses the **WLCG*** as a datalake and HTC resource
  - WLCG available to everyone in a LHC-recognised VO
- Integration with WLCG is crazy challenging but not impossible
- **Needs dedicated FTEs and RSEs** working on this
  - The R&D by itself would push us forward



Builds on well established open-source projects

* Worldwide LHC Computing Grid

Baler project

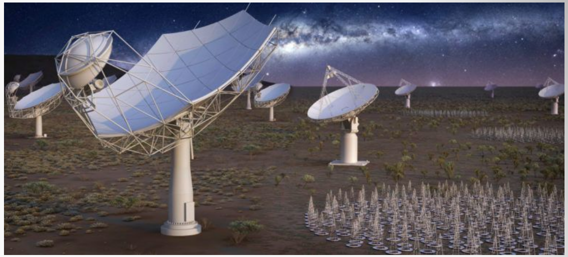
Storing and sharing ever-increasing datasets is a cross-discipline issue affecting scientific research and industry

- High demand for data compression ratios higher than loss-less methods can achieve

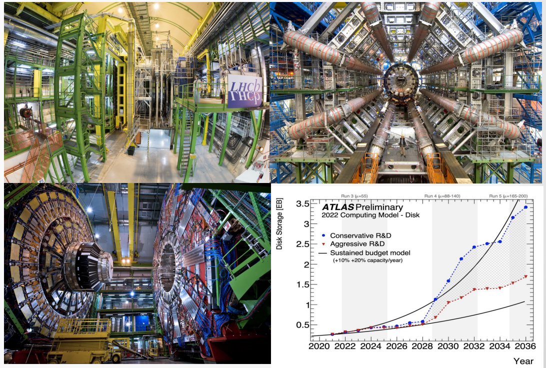
Lossy compression results in the permanent discarding of data but when compression techniques are tailored to the datatype the data loss is inconsequential

- MP3 compression uses psychoacoustics exploiting the limitations of human hearing to discard data
- MP3 will not work for HEP data

Is there a way to tailor compression techniques to the input dataset?



8.5EB of data over 15-year lifespan



Order of magnitude more analysis data by Run 5

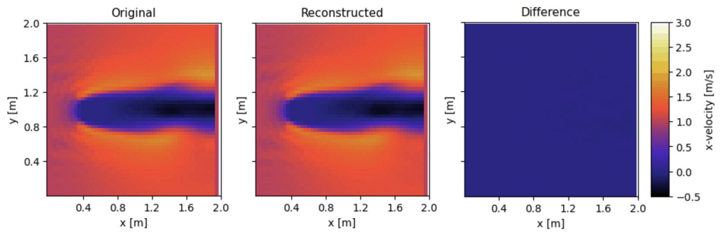
Baler project

Baler is tool for Machine Learning based lossy data compression for use across scientific disciplines and industry

- Trains Autoencoders* to compress input data based on chosen hyper-parameters
- Cross-discipline collaboration between HEP/CFD researchers, Computer Scientists and ML experts

CFD researchers have to store and share TB-sized simulation files that often have to be deleted after results are published

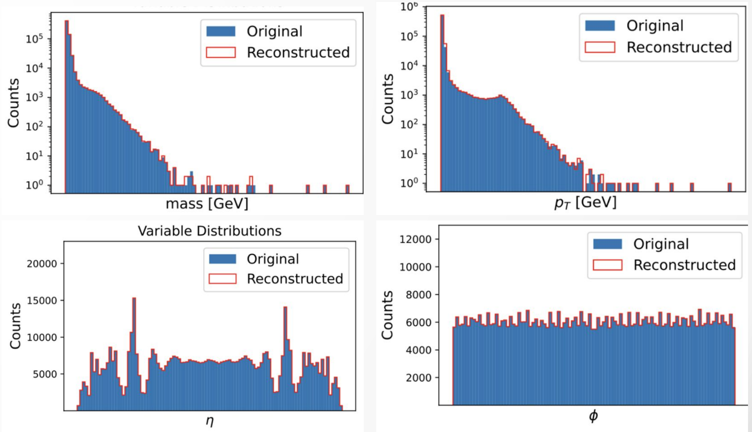
Compression ratio 200



Baler performance on CFD data

In HEP more data is better - higher statistics compensate for loss in precision

Compression ratio 1.6

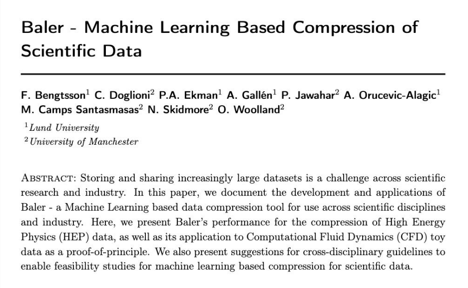


Baler performance on CMS open data

*Also used in noise reduction and anomaly finding

Baler project

github.com/baler-compressor



Baler - Machine Learning Based Compression of Scientific Data

F. Bengtsson[1] C. Doglioni[2] P.A. Ekman[1] A. Gallén[1] P. Jawahar[2] A. Orucevic-Alagic[1] M. Camps Santasmasas[2] N. Skidmore[2] O. Woolland[2]

[1] *Lund University*

[2] *University of Manchester*

Abstract: Storing and sharing increasingly large datasets is a challenge across scientific research and industry. In this paper, we document the development and applications of Baler - a Machine Learning based data compression tool for use across scientific disciplines and industry. Here, we present Baler's performance for the compression of High Energy Physics (HEP) data, as well as its application to Computational Fluid Dynamics (CFD) toy data as a proof-of-principle. We also present suggestions for cross-disciplinary guidelines to enable feasibility studies for machine learning based compression for scientific data.

Arxiv: 2305.02283

CHEP talk

Working Example

Create New Project Directory

```
poetry run python baler --mode=newProject --project=firstProject
```

Training

```
poetry run python baler --project=firstProject --mode=train
```

Compressing

```
poetry run python baler --project=firstProject --mode=compress
```

Decompressing

```
poetry run python baler --project=firstProject --mode=decompress
```

Plotting

```
poetry run python baler --project=firstProject --mode=plot
```

Baler has extensive applications in scientific research

- Huge potential for HEP **simulation samples**
- Preservation of derived datasets for **analysis reproducibility** (average LHCb analysis now has 5TB of derived data)

…But also in **industry**

Heavy focus now on offloading of training to **GPUs** and **FPGA** implementation

- Particularly seeking **FPGA expertise** for bandwidth compression - telecommunications, autonomous vehicles, and real-time analysis…

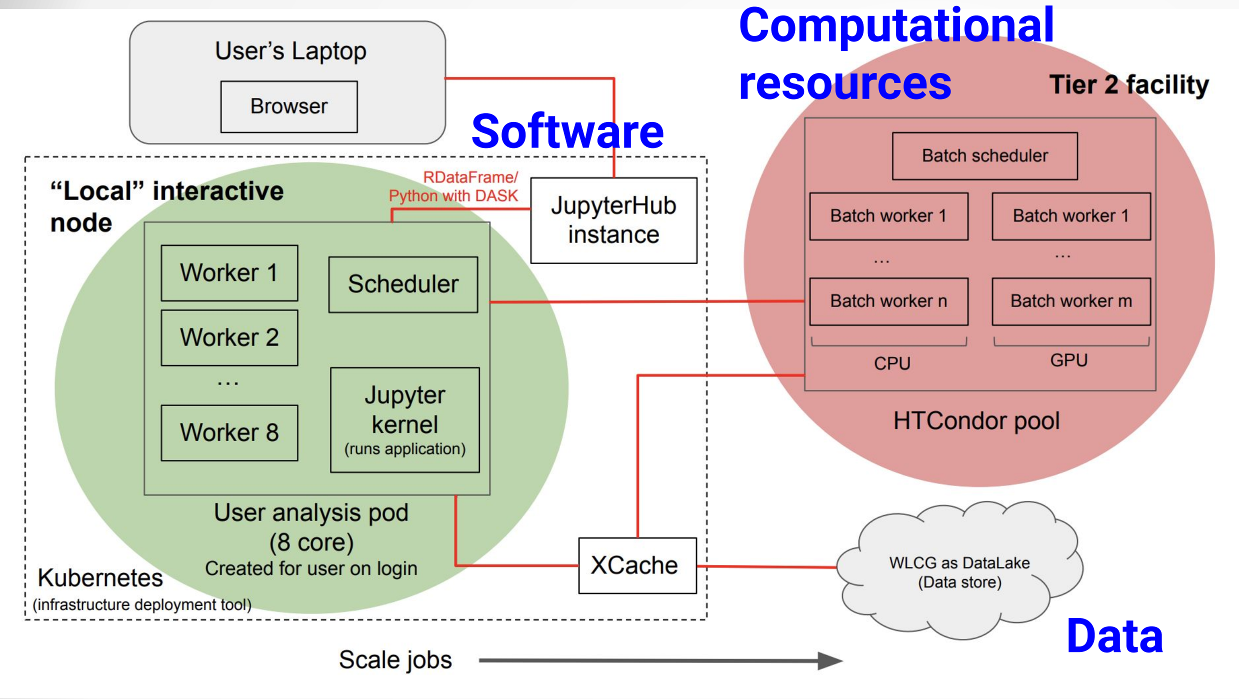
Thank you

Back Up

What makes this possible?

Kubernetes - infrastructure deployment tool

Adopting industry solutions and new DOMA advances



Virtual organisation = your institute

Analyst logs into "local to institute" interactive node

Opens Jupyter Notebook with web-browser interface

Uses Python columnar analysis libraries (or RDataFrame) with Dask to analyse data locally

WLCG is DataLake - data is cached at local resources (XCache)

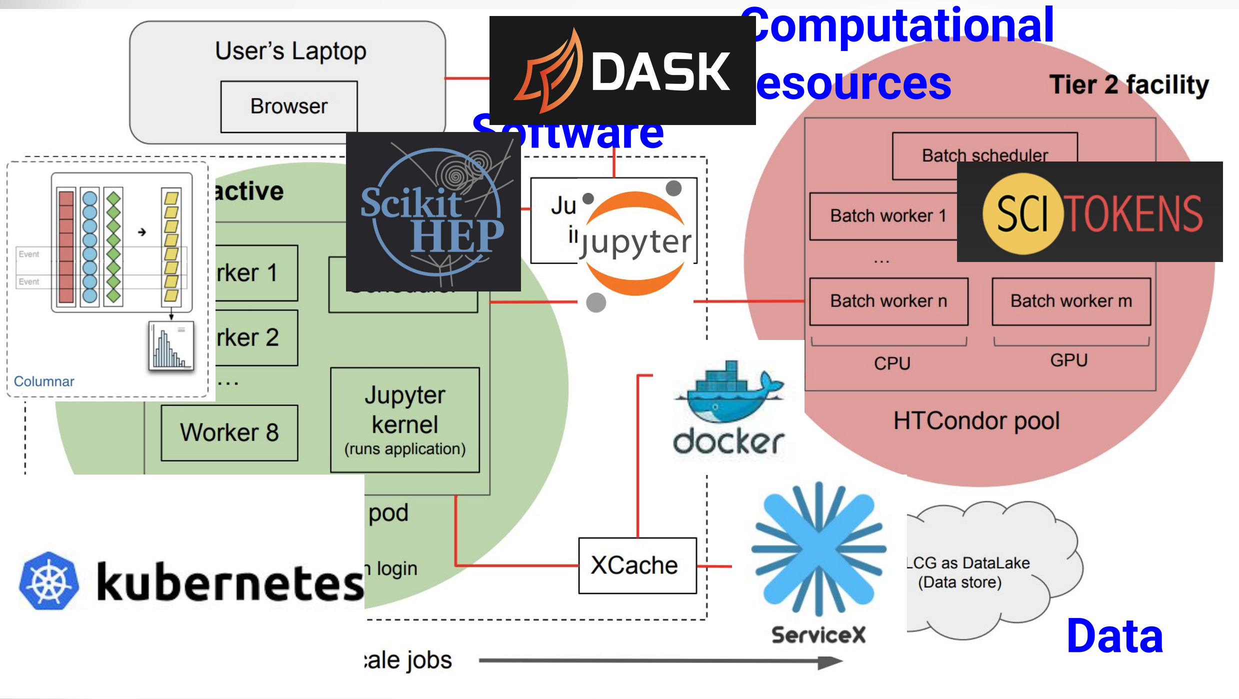
Once node resources no longer sufficient job is automatically scaled out to WLCG tier 2 sites

Results stored on local resources and returned to the interactive Jupyter session

What makes this possible?

Kubernetes - infrastructure deployment tool

Adopting industry solutions and new DOMA advances



Virtual organisation = your institute

Analyst logs into "local to institute" interactive node

Opens Jupyter Notebook with web-browser interface

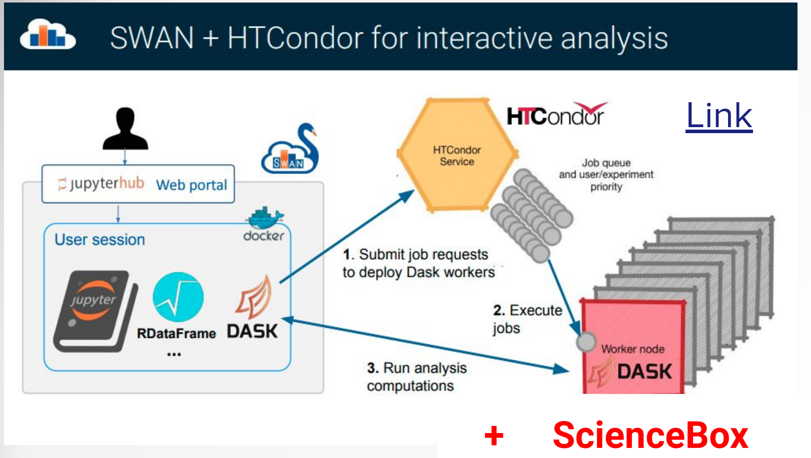
Uses Python columnar analysis libraries (or RDataFrame) with Dask to analyse data locally

WLCG is DataLake - data is cached at local resources (XCache)

Once node resources no longer sufficient job is automatically scaled out to WLCG tier 2 sites

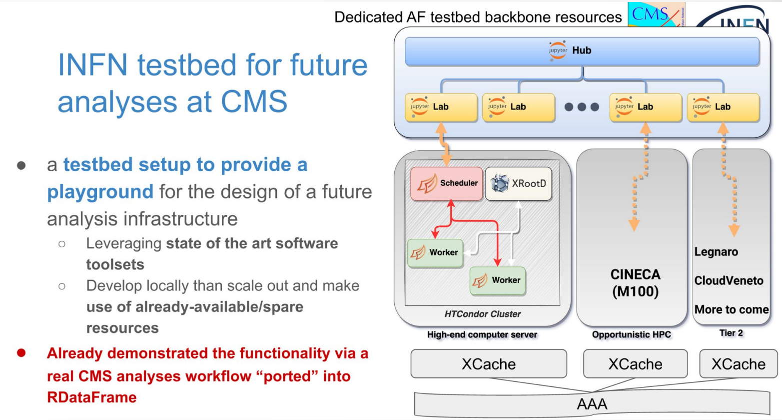
Results stored on local resources and returned to the interactive Jupyter session

Doesn’t something similar exist already?



SWAN: Service for Web-based Analysis

- Interface: Jupyter notebook
- Storage: EOS/CERNbox
- Resources: batch “local” to CERN
- For: LHC experiment agnostic
- Runs: RDataFrame/Coffea with DASK
  VERY POPULAR DUE TO EOS



INFN analysis facility

- Interface: Jupyter notebook
- Storage: Local AF area (CEPH)
- Resources: T2 sites + CINECA (HPC)
- For: CMS, expanding to other LHC experiments incl. LHCb
  Runs: RDataFrame with DASK
- Services: XCache

Yes, but...

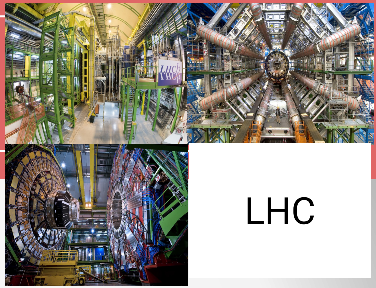
There is nothing like this in the UK (most analysis facilities are US based)

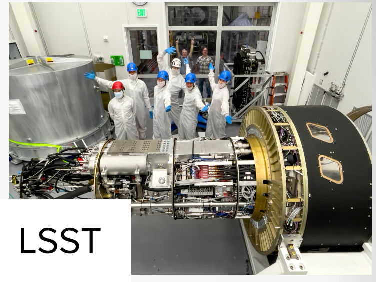
Almost all are CMS/ATLAS targeted. In this proposal there is nothing experiment or even HEP specific (dont have to use WLCG, can use local or national Tier 2 sites)

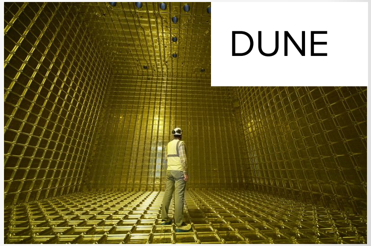
The UK has investment in many big-data HEP/Astro/Neutrino experiments as well as large Tier 2 sites with heterogeneous resources

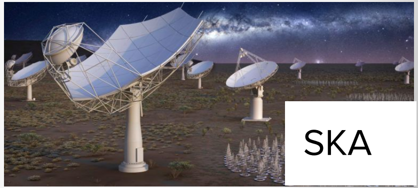
- Sharing and optimising the efficient use of specialized infrastructure (increasingly used in end-user data analysis)
- Common solution to cross-discipline issue of current, “local” end-user data analysis methods and corresponding tools not scaling with event count

If UK institutes adopt instances of this AF they will have the most powerful, cross experiment analysis facilities









$b \to c\, \bar{u}\, q$ anomaly - potential causes

BSM physics in SM tree-level decays?

Universal destructive interference with the SM predictions at the level of 17% is favored

$$\mathcal{L} = -\frac{4G_F}{\sqrt{2}} \sum_q V_{cb} V_{uq}^* \sum_{i=1,2} C_i^q(\mu) \mathcal{Q}_i^q(\mu)\,,$$

$$\frac{C_2^{\mathrm{NP}}(m_b)}{C_2^{\mathrm{SM}}(m_b)} = -0.17 \pm 0.03\,.$$

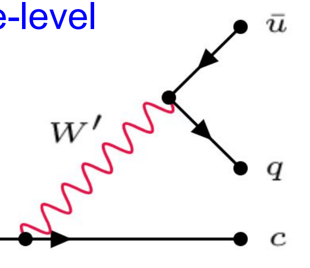
Remains compatible with other observables

Such a shift can be achieved by an extended electroweak gauge group

$$SU(2)_1 \times SU(2)_2 \times U(1)_Y$$

Which gives heavy gauge bosons **W±’** and Z’

NP in tree-level decays?

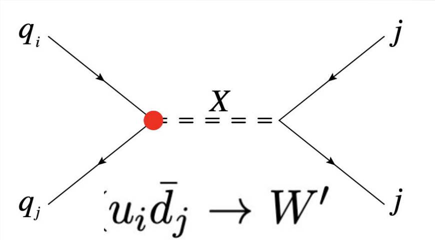


arXiv:2008.01086v3

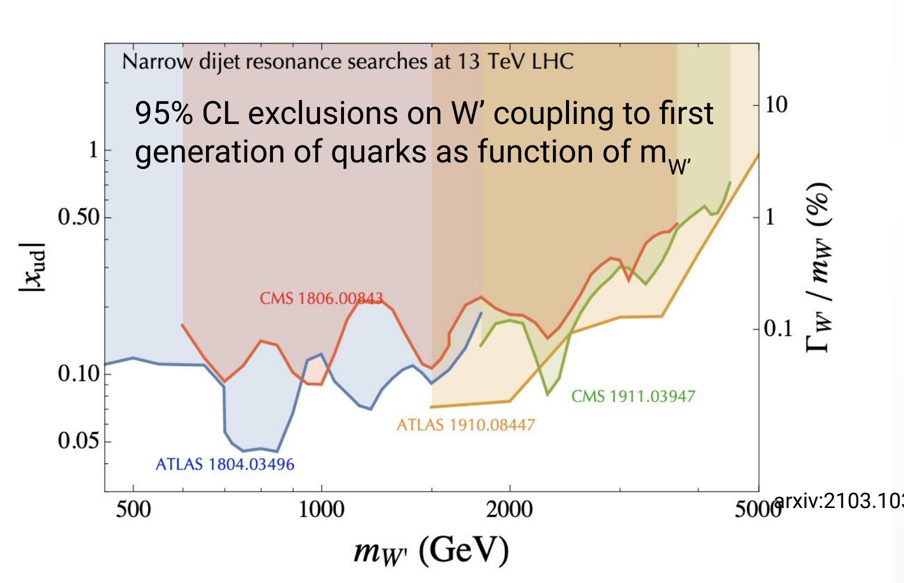
$b \to c \bar{u} q$ anomaly - potential causes

Would we not have seen a new boson at the LHC?

Coupling of a W’ to quarks implies resonant production in pp collisions followed by dijet



ATLAS and CMS dijet searches



arxiv:2103.10332

arXiv:2008.01086v3

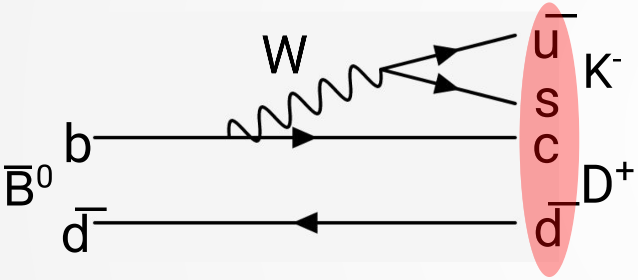
ATLAS and CMS have set robust constraints on hypothetical tree-level mediators with masses up to 5 TeV

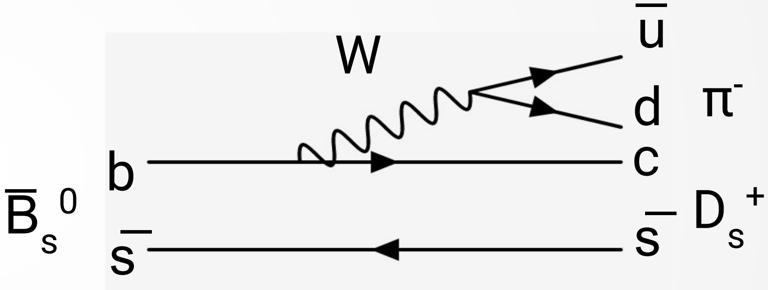
Still allowed parameter space for a high mass W’ from di-jet searches

A closer look at these key decays...

Predictions for $\overline{B}_s \rightarrow D_s^{(*)+} \pi^-$ and $\overline{B}^0 \to D^{(*)+}K^-$ in particular are among the most reliable for non-leptonic decays

- No penguin or weak annihilation contributions
- Dominated by colour allowed trees





CP-conserving	Flavour specific
$\overline{B}_s \rightarrow D_s^+ \pi^- = B_s \rightarrow D_s^- \pi^+$	$\overline{B}_s \not\rightarrow D_s^- \pi^+ \quad B_s \not\rightarrow D_s^+ \pi^-$

CP asymmetries as a test of the SM

Can use CP asymmetry of $\overline{B}_S \to D_S^+ \pi^-$ to probe NP effects

$\overline{B}_S \not\to D_S^- \pi^+$

$B_S \not\to D_S^+ \pi^-$

$\overline{B}_S \to D_S^+ \pi^-$ **=** $B_S \to D_S^- \pi^+$

In the SM $A_{CP}(D_S\pi) = a^s_{fs} \to$ CPV in mixing

$A_{CP}(D_S\pi)$ in the SM only sensitive to CPV in mixing

$= a^s_{SL} = (-60\pm280)\text{x}10^{-5}$

Consistent with SM prediction $a^s_{fs}=(2.06\pm0.18)\text{x}10^{-5}$

With most general BSM contribution

$$A_{CP}(D_S\pi) = a^s_{fs} + \underbrace{A^s_{dir}}$$

Direct CPV term manifestly NP

$A_{CP}(D_S\pi)$ may be enhanced up to a level of 10^{-2} by BSM direct CPV effects

$A_{CP}(D_S\pi)$ can probe BSM effects in tree-level hadronic decays, independent of theoretical assumptions

$A_{CP}(D_S\pi)$ has never been explicitly measured

Inputs to theory

arxiv:2007.10338

measurement	value	source	reference(s)
$\mathcal{B}(B_s^0 \to D_s^- \pi^+)$	$(3.6 \pm 0.5 \pm 0.5)\ 10^{-3}$	Belle	[15, 21]
$\frac{f_s}{f_d} \frac{\mathcal{B}(B_s^0 \to D_s^-(\to \phi(\to K^+K^-)\pi^-)\pi^+)}{\mathcal{B}(B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+)}$	$(6.7 \pm 0.5)\%$	CDF	[41]*
$\frac{f_s}{f_d} \frac{\mathcal{B}(B_s^0 \to D_s^-(\to K^+K^-\pi^-)\pi^+)}{\mathcal{B}(B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+)}$	0.174 ± 0.007	LHCb	[42]
$\frac{f_s}{f_d} \frac{\mathcal{B}(B_s^0 \to D_s^-(\to K^+K^-\pi^-)\pi^+)}{\mathcal{B}(B^0 \to D^-(\to K^+\pi^-\pi^-)K^+)}$	2.08 ± 0.08	LHCb	[25]†
$\frac{\mathcal{B}(B^0 \to D^-K^+)}{\mathcal{B}(B^0 \to D^-\pi^+)}$	$(8.22 \pm 0.28)\%$	LHCb	[25]†
$\frac{\mathcal{B}(B^0 \to D^-K^+)}{\mathcal{B}(B^0 \to D^-\pi^+)}$	$(6.8 \pm 1.7)\%$	Belle	[43]
$f_{00}\mathcal{B}(B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+)$	$(1.21 \pm 0.05)\ 10^{-4}$	BaBar/CLEO	[24, 44]
$\mathcal{B}(B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+)$	$(2.88 \pm 0.29)\ 10^{-4}$	BaBar	[45]§
$\frac{\mathcal{B}(B_s^0 \to D_s^{*-}\pi^+)}{\mathcal{B}(B_s^0 \to D_s^-\pi^+)}$	0.66 ± 0.16	Belle	[46]
$\frac{\mathcal{B}(B^0 \to D^{*-}K^+)}{\mathcal{B}(B^0 \to D^{*-}\pi^+)}$	$(7.75 \pm 0.30)\%$	LHCb/BaBar/Belle	[43, 47, 48]
$f_{00}\mathcal{B}(B^0 \to D^{*-}\pi^+)$	$(2.72 \pm 0.14)\ 10^{-3}$	BaBar/CLEO	[24, 49]
$\frac{\mathcal{B}(B^0 \to D^{*-}\pi^+)}{\mathcal{B}(B^0 \to D^-\pi^+)}$	0.99 ± 0.14	BaBar	[45]
$\mathcal{B}(D_s^- \to \phi(\to K^+K^-)\pi^-)$	$(2.27 \pm 0.08)\%$	PDG average	[15]
$\mathcal{B}(D_s^- \to K^+K^-\pi^-)$	$(5.45 \pm 0.17)\%$	PDG average	[15]
$\mathcal{B}(D^- \to K^+\pi^-\pi^-)$	$(9.38 \pm 0.16)\%$	PDG average	[15]
$\mathcal{B}(D_s^- \to K^+K^-\pi^-)(f_s/f_d)^{7\text{TeV}}_{\text{LHCb,sl}}$	0.0144 ± 0.0010	LHCb	[22, 23]
$\mathcal{B}(D_s^- \to K^+K^-\pi^-)(f_s/f_d)^{13\text{TeV}}_{\text{LHCb,sl}}$	0.0133 ± 0.0005	LHCb	[50]
$(f_s/f_d)_{\text{Tev}}$	0.334 ± 0.040	HFLAV average	[29]
f_{00}	0.488 ± 0.010	pheno comb. of BaBar/Belle	[40, 51, 52]

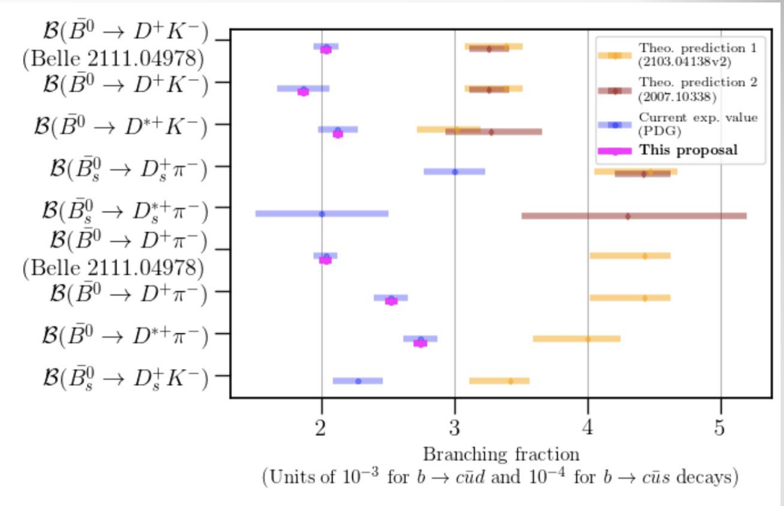
TABLE IV. Relevant experimental measurements entering our determination of branching fractions and ratios for the considered modes. A * indicates that here f_s/f_d corresponds to the value at Tevatron, see text. The measurements marked by a † from ref. [25] have a correlation coefficient of −56%. The reference marked with § uses both $D^- \to K^+\pi^-\pi^-$ and $D^- \to K_S\pi^-$ decays, however, the former is dominating.

Measurements at LHCb -prospects

Improving experimental precision on branching fractions is key to further understand anomaly

Measure relative branching fractions

$$\mathcal{B}(B^0 \to D^+K^-) = \mathcal{B}(B^0 \to D^+\pi^-) \cdot \frac{N_{B^0\to D^+K^-}}{N_{B^0\to D^+\pi^-}} \frac{\epsilon_{B^0\to D^+\pi^-}}{\epsilon_{B^0\to D^+K^-}}$$



In Runs 1 and 2 normalisation channels limited to those o other hadronic modes such that L0 trigger efficiency cancelled

- Normalisation channel was limiting systematic

Exploit Run 3 ability to precisely determine hadronic trigger efficiencies

- Use $B^+ \to J/\Psi K^+$ as normalisation channel ($f_u/f_d \sim 1$) - branching fraction is far more precisely known (1.9%)
- Measure branching fractions to unprecedented precision - enhancing global experimental sensitivities by factor 4

Autoencoders

[Autoencoders](#)

An autoencoder is an unsupervised artificial neural network that learns how to efficiently compress and encode data.

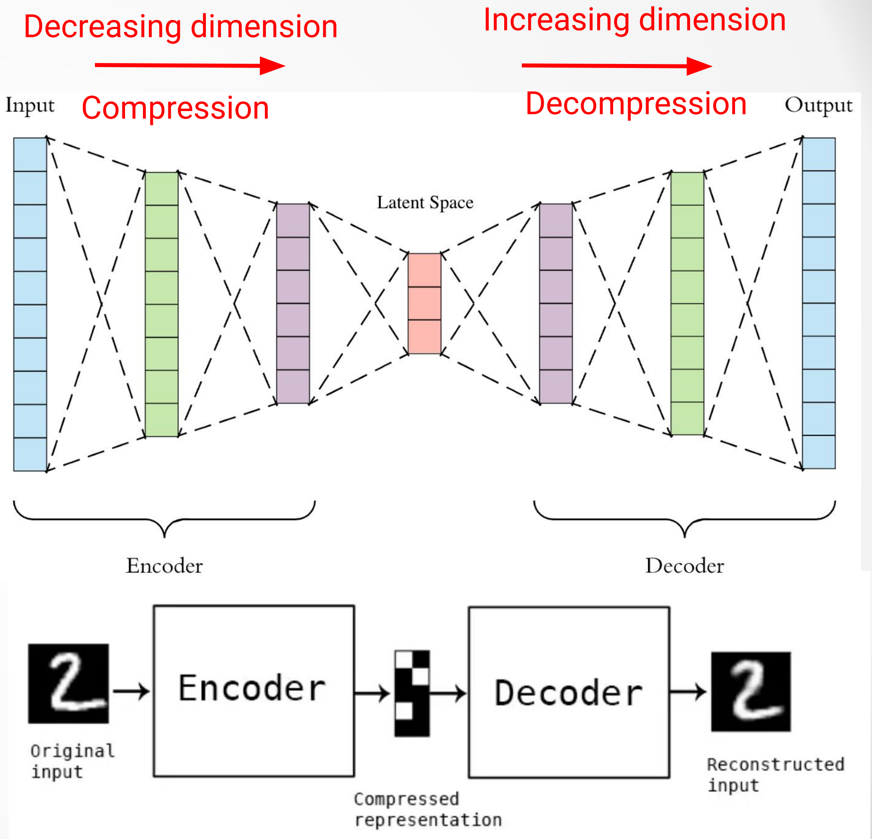
Steps

1- Encoder: Model learns how to reduce the input dimensions and compress the input data into an encoded representation.

2- Bottleneck/latent space: layer that contains the compressed representation of the input data

3- Decoder: Model learns how to reconstruct the data from the encoded representation

4- Reconstruction Loss: measure how close the output is to the original input



Used in ***noise reduction and anomaly finding***

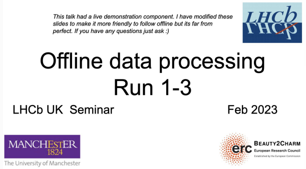
Student teaching/engagement and support

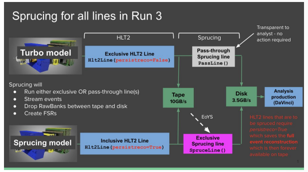
I have a strong record of lecturing/tutoring/demonstrating for my position

Volunteered to give “Physics at the LHC” courses - including weekly journal club

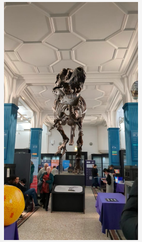
- Would be happy to implement such a course at Warwick

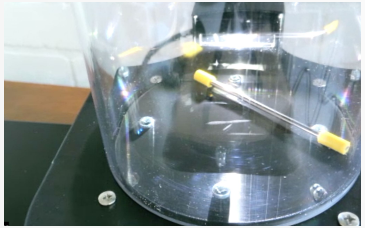
Significant, longstanding support for ECRs on LHCb



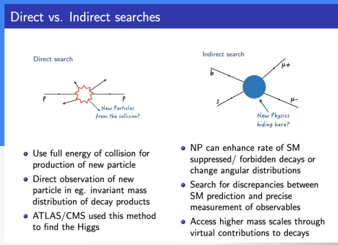


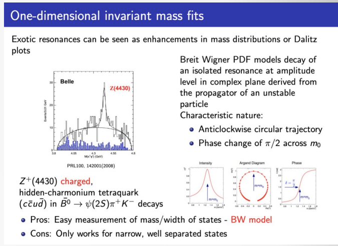
Recent and upcoming outreach at British Science Week and BlueDot

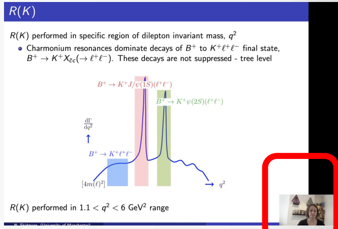


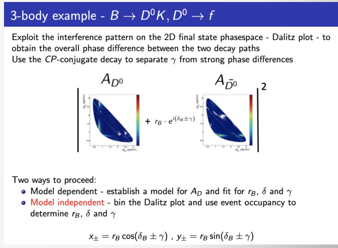


THE BIG BANG

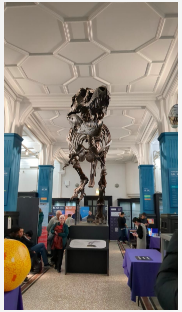


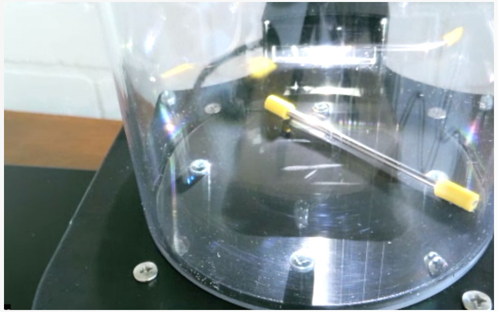






Student teaching/engagement and support





Cloud chamber exhibit at Manchester Museum for British Science week - attended by 700 secondary school students

THE BIG BANG

History of particle detectors exhibit at Bluedot 2023

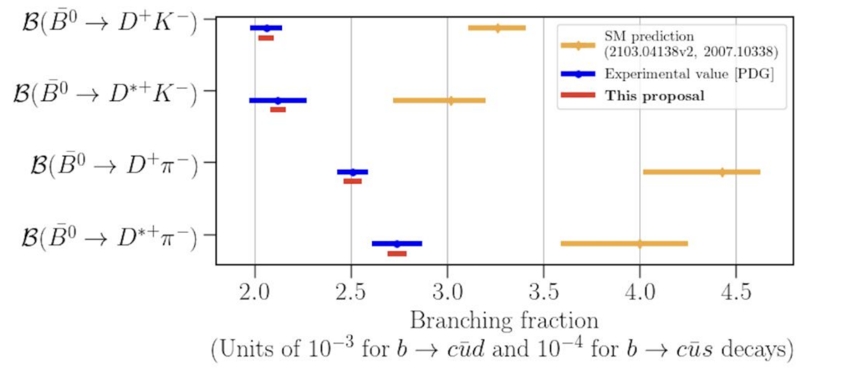
LHCb virtual visit guide

brew monday
SAMARITANS

"Brew Monday" Samaritans event

- Due to popularity will make a regular feature in the Natural Sciences calendar particularly in exam periods
- Would be extremely proud to implement at Warwick

Solving big and small problems



- Combinations from ~6 experiments with different collision environments - conservative assumptions made on fragmentation fractions
- Poorly known normalisation channel BF - limiting systematic
- Normalisation channel suffers from anomaly

Upgrade LHCb trigger strategy allows use of “cleaner” leptonic normalisation channels

For the first time measure decays at same experiment with consistent methodology enhancing sensitivity up to **factor 4**

Sensitivity to π^0 in Upgrade 2 will allow measurements of vector bachelor modes that also suffer from anomaly