

Pentaquarks and Open Data

LHCb-Germany meeting, Dortmund 2021

Adam Morris¹, with material from: Abhay Mehta¹, Mindaugas Šarpis¹, Marian Stahl², Chris Burr³, Dillon Fitzgerald⁴, Donijor Tropmann⁵

¹University of Bonn, ²University of Cincinnati, ³CERN, ⁴University of Michigan, ⁵RWTH Aachen

8th October 2021



UNIVERSITÄT

BONN

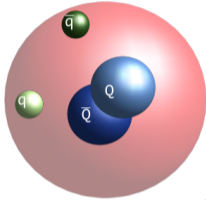


Activities in the Bonn LHCb group

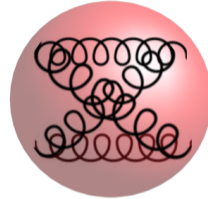
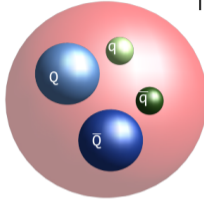
- **Searches for $J/\psi p$ pentaquarks in other channels**
- Mighty Tracker R&D
- Analysis Preservation & **Open Data**
- Simulation production

Pentaquarks

HADRO-
QUARKONIUM

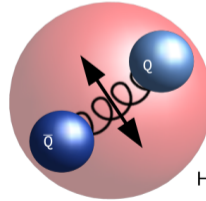
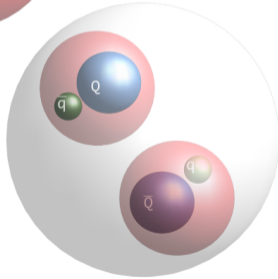


TETRAQUARK



GLUEBALL

HADRONIC
MOLECULE



HYBRID

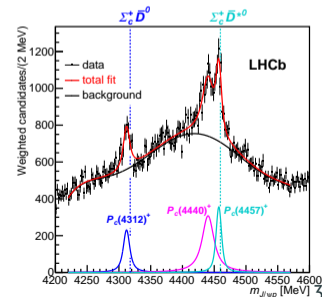
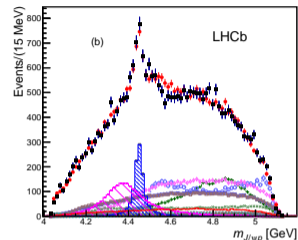
Pentaquarks: some history

Discovery of pentaquarks [\[arXiv:1507.03414\]](#)

- Amplitude analysis of $\Lambda_b^0 \rightarrow J/\psi p K^-$ showed clear structure in $J/\psi p$ spectrum
 - $P_c(4450)^+$, $\Gamma \approx 40$ MeV
 - $P_c(4380)^+$, $\Gamma \approx 200$ MeV
- Proximity to $\Sigma_c^+ \bar{D}^{*0}$ threshold

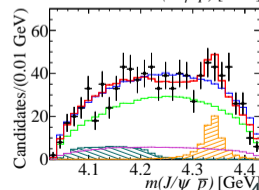
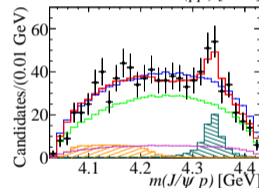
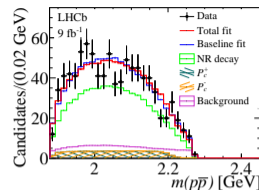
Adding the Run 2 data [\[arXiv:1904.03947\]](#)

- $P_c(4450)^+$ resolved to two-peak structure
 - $P_c(4440)^+$, $\Gamma \approx 21$ MeV
 - $P_c(4457)^+$, $\Gamma \approx 6$ MeV
- New pentaquark right on $\Sigma_c^+ \bar{D}^0$ threshold
 - $P_c(4312)^+$, $\Gamma \approx 10$ MeV



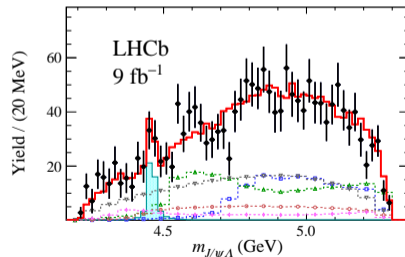
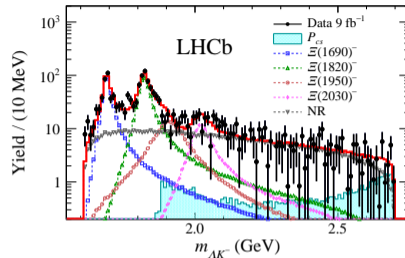
Evidence of a new $J/\psi p$ structure in $B_s^0 \rightarrow J/\psi p \bar{p}$ [\[arXiv:2108.04720\]](https://arxiv.org/abs/2108.04720)

- Amplitude analysis of $B_s^0 \rightarrow J/\psi p \bar{p}$ decays
- Observe structure in $J/\psi p$ and $J/\psi \bar{p}$ spectra
- Potentially a new $P_c(4337)^+$ state
 - $m = 4337_{-4}^{+7}{}_{-2}^{+2}$ MeV, $\Gamma \approx 30$ MeV
- Significance 3.1σ to 3.7σ depending on J^P assignment
- Inconsistent with $P_c(4312)^+$
 - $m = 4311.9 \pm 0.7_{-0.6}^{+6.8}$ MeV



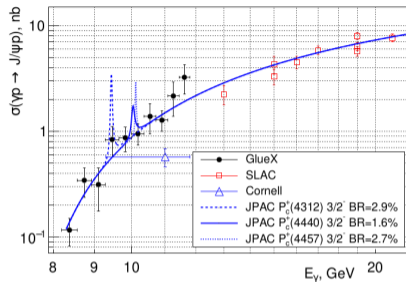
Evidence of a $c\bar{c}sud$ pentaquark [arXiv:2012.10380]

- Amplitude analysis of $\Xi_b^- \rightarrow J/\psi \Lambda K^-$ decays
- Evidence of structure in $J/\psi \Lambda$ spectrum
 - $P_{cs}(4459)^0$
- Significance 3.1σ
- Potentially a first charmonium-pentaquark with strangeness



Where do pentaquarks get their width?

- Photoproduction experiments put upper limit on partial width $\Gamma(P_c \rightarrow J/\psi p) < 2 \sim 5\%$
[Phys. Rev. Lett. 123 (2019) 072001]
- Expect to see P_c in other channels with larger branching fractions
- Molecules should couple strongly to channels with open charm, e.g. $\Lambda_c^+ \bar{D}^0$, $\Sigma_c^+ \bar{D}^0$, $\Lambda_c^+ \bar{D}^{*0}$
- Focus of analyses in the Bonn LHCb group



P_c State	$\mathcal{B}(P_c \rightarrow J/\psi p)$
$P_c^+(4312)$	$< 4.6\%$ (90% CL)
$P_c^+(4440)$	$< 2.3\%$ (90% CL)
$P_c^+(4457)$	$< 3.8\%$ (90% CL)

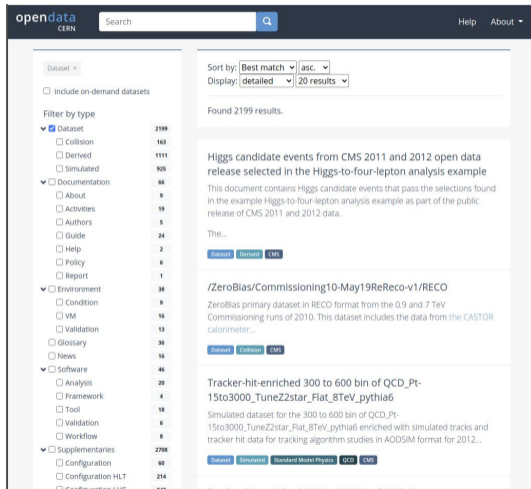
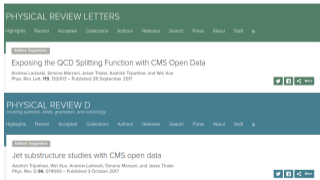
Open data



<https://cern.ch/opendata>

CERN repository for public datasets

- > 2 PB of data from ATLAS, CMS, ALICE, LHCb and OPERA
- Used in education, outreach and real physics research





<https://cern.ch/opendata>

Search or jump to... [Pull requests](#) [Issues](#) [Marketplace](#) [Explore](#)

cms-opendata-analyses / HiggsExample20112012

< Code Issues Pull requests Actions Projects Wiki Security Insights

master 2 branches 0 tags Go to file Add file Code

katlip Merge pull request #1 from cms-opendata-analyses/katlip-gitclone 66648f9 on Feb 27, 2018 117 commits

HiggsDemoAnalyzer	rename to avoid conflict with open data portal version of analyser	3 years ago
Level2	moving the macro to a dedicated L2 directory. Includes modifications ...	3 years ago
Level3	rename to avoid conflict with open data portal version of analyser	3 years ago
Level4	adding one example MC config list	3 years ago
datasets	adding one example MC config list	3 years ago
rootfiles	moving the macro to a dedicated L2 directory. Includes modifications ...	3 years ago
LICENSE	Create LICENSE	3 years ago
README.md	git clone https -> git clone git	3 years ago
mass4_combine.pdf	Added the configuration from outside git	3 years ago
mass4_combine.png	Added the configuration from outside git	3 years ago

Root files for Higgs-to-four-lepton analysis example using 2011-2012 data

Jomhari, Nur Zulaiha ; Geiser, Achim ; Bin Anuar, Afiq Alzuddin

Cite as: Jomhari, Nur Zulaiha; Geiser, Achim; Bin Anuar, Afiq Alzuddin; (2017). Root files for Higgs-to-four-lepton analysis example using 2011-2012 data. CERN Open Data Portal. DOI:10.7483/OPENDATA.CMS.JK88.D634

[Dataset](#) [Dataset](#) [Code](#) [CERN LHC](#)

Description

This record contains root files that were processed for the Higgs analysis example on the CMS 2011-2012 Open Data.

Dataset characteristics

24 files. 1.6 MB in total.

How can you use these data?

These datasets are provided as part of the Higgs-to-four-lepton analysis example using 2011-2012 data. You can download all root files needed for the Level 3 exercise to your working area by downloading first the file list `rootfilelist.txt` and then the root files with `wget -i rootfilelist.txt`.

Please note that the following files are not needed to produce the plot but are only the products of intermittent processing steps:

- DY101jets12.root
- DY50Mag12.root
- DYTo2mu12.root
- TTjets11.root
- TTjets12.root

[Higgs-to-four-lepton analysis example using 2011-2012 data](#)

Files

Filename	Size	
rootfilelist.txt	1.1 kB	Download
DY1011.root	64.7 kB	Download

Example analysis: CMS $H \rightarrow 4\ell$ discovery

Updated policy endorsed by the large LHC experiments [CERN-OPEN-2020-013](#)

Applies FAIR principles

Four levels of data complexity:

Level 1: Public results

e.g. HEPData

Level 2: Education and outreach

e.g. Masterclasses

Level 3: **Reconstructed data**

Open Data Portal

Level 4: Raw data

Not released

Updated policy endorsed by the large LHC experiments [CERN-OPEN-2020-013](#)

Applies FAIR principles

Four levels of data complexity:

Level 1: Public results

e.g. HEPData

Level 2: Education and outreach

e.g. Masterclasses

Level 3: **Reconstructed data**

Open Data Portal

Level 4: Raw data

Not released

Level 3 data

- Data for main physics programme (protons and heavy ions)
- Special runs may not be included
- Software must also be available
- Data formats same as internal
- Collaborations will not review or sign third-party analyses

- Open same datasets that we use internally
Stripping/Turbo/Sprucing streams
- 50% (100%) after 5 (10) years since the end of each LHC Run
Exceptionally withhold individual datasets still being used in ongoing analyses
- Provide tools to perform calibrations & corrections but no further support
PIDCalib, TrackCalib, etc...
- Require proper acknowledgement in papers using LHCb Open Data
Including disclaimer that LHCb has not reviewed the result
- LHCb collaborators are not permitted to sign papers using LHCb Open Data
Possible exceptions if no physics results shown (*e.g.* methodology papers)

Dataset x Collision x Derived x LHCb x

Include on-demand datasets

Filter by type

- Dataset 4
 - Collision 1
 - Derived 3
- Documentation 6
- Environment 1
- Software 1

Filter by experiment

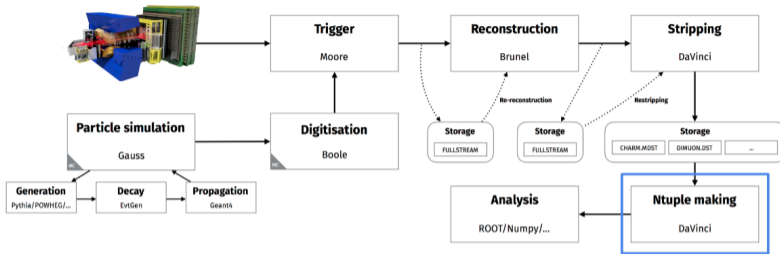
- ALICE 15
- ATLAS 109
- CMS 1167
- LHCb 4
- OPERA 904

Why hasn't the Open Data been released yet?

Significantly larger filesizes than other LHC experiments!

	ALICE	ATLAS	CMS	LHCb
Run-2	2 PB	0.5 PB	2 PB	10 PB (including Run-1)
Run-3	4 PB	1 PB	4 PB	45 PB
Total	6 PB	1.5 PB	6 PB	55 PB

Not scalable: quickly hit storage quota on Open Data Portal.



Ntuple making currently requires knowledge of the LHCb software stack

- Lots of documentation and support required to be successful
 - See [LHCb Starterkit](#)
- High barrier of entry for external analysts
- Complete overhaul of data processing (including ntupling) for Run 3
 - Knowledge required to access Run1+2 data will start fading “soon”

Ntuple Wizard

Make ntuples of LHCb data without knowledge of the software stack

- Render necessary option files from intuitive user inputs
- Create ntuples via Analysis Productions
- Return ntuples to user on the CERN Open Data Portal

Ntuple Wizard

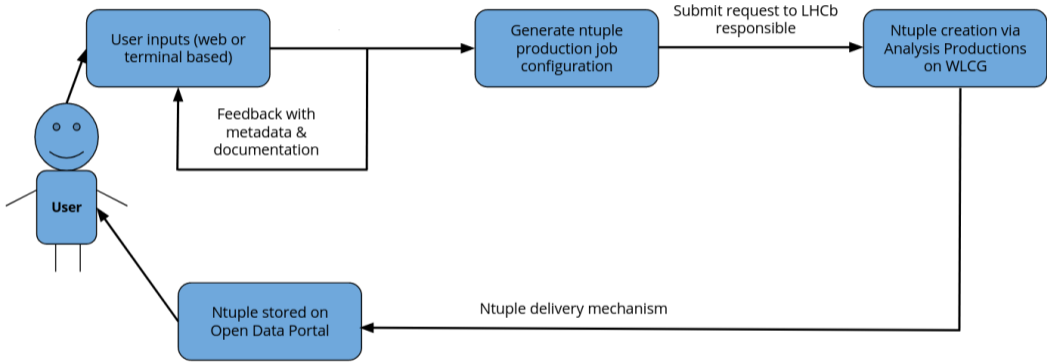
Make ntuples of LHCb data without knowledge of the software stack

- Render necessary option files from intuitive user inputs
- Create ntuples via Analysis Productions
- Return ntuples to user on the CERN Open Data Portal

This solves several problems:

- ✓ Much smaller storage and bandwidth requirements on Open Data Portal
- ✓ Significantly flattened learning curve for accessing LHCb data
- ✓ No need for a computing cluster just to download and process the data

NB: *Also benefits LHCb members*



Backend

- Collect metadata from various sources and serve as static files

 [lhcb-dpa/wp6-analysis-preservation-and-open-data/lhcb-ntuple-wizard](#)

Web frontend

- Guide the user through choosing a dataset and configuring algorithms

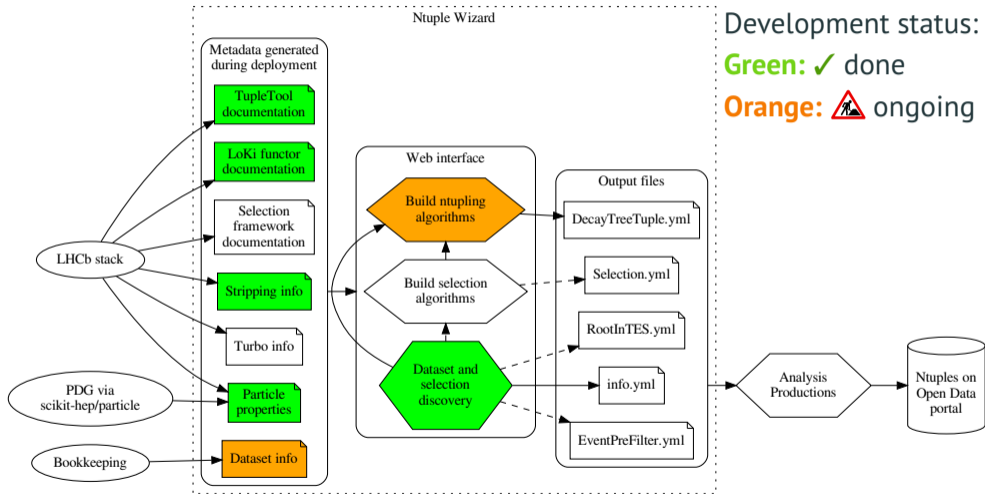
 [lhcb-temp-transfer/tuple-wizard-frontend](#)

Parsers

- Configure DaVinci with the output of the Wizard

 [admorris/analysishelpers](#)

NB: repo locations subject to change



DaVinci

- Access via Gaudi python scripts
- TupleTool interfaces
- Stripping lines
 - Decay descriptors
 - Streams, TES locations

PDG

- [scikit-hep/particle](#)
 - Physical properties
 - LaTeX names
- [ParticleTable.txt](#) from DDDB
 - Gaudi names
 - Custom particles

Doxygen

- Non-trivial URL discovery
- HTML pages parsed with bs4
- TupleTool and LoKi functor docs

Bookkeeping

- Access via LHCbDIRAC API scripts
- Delegated to sister project: [lhcb-opendata-curation](#)
 - Generates the ODP records

Frontend: Decay descriptor search

$B^0 \times D^0 \times \pi^+ \times$ x | v

- π^+
- π^-
- π^0
- π^-
- η

$(\pi^+\pi^-)D^0 \rightarrow (\pi^+\pi^-)$	1 Data Types
$(\pi^+\pi^+\pi^-\pi^-)D^0 \rightarrow (\pi^+\pi^+\pi^-\pi^-)$	1 Data Types
$(\pi^+\pi^-)D^0 \rightarrow (\pi^+\pi^-)K_S^0$	1 Data Types
$B^0 \rightarrow (D^0 \rightarrow (\pi^+\pi^+\pi^-\pi^-)D^0 \rightarrow (\pi^+\pi^+\pi^-\pi^-)K_S^0)$	4 Stripping Lines 1 Data Types
$B^0 \rightarrow (D^0 \rightarrow (\pi^+\pi^-)D^0 \rightarrow (\pi^+\pi^-)K^*(892)^0)$	2 Stripping Lines 1 Data Types
$B^0 \rightarrow (D^0 \rightarrow (\pi^+\pi^-)D^0 \rightarrow (\pi^+\pi^-)\bar{K}^*(892)^0)$	3 Stripping Lines 1 Data Types

Select Lines

$B^0 \times D^0 \times \pi^+ \times \pi^- \times$ x | v

$B^0 \rightarrow (D^0 \rightarrow (\pi^+\pi^-)D^0 \rightarrow (\pi^+\pi^-))$	2 Stripping Lines 1 Data Types
$B^0 \rightarrow (D^0 \rightarrow (\pi^+\pi^+\pi^-\pi^-)D^0 \rightarrow (\pi^+\pi^+\pi^-\pi^-))$	2 Stripping Lines 1 Data Types
$B^0 \rightarrow (D^0 \rightarrow (\pi^+\pi^-)D^0 \rightarrow (\pi^+\pi^-)K_S^0)$	4 Stripping Lines 1 Data Types
$B^0 \rightarrow (D^0 \rightarrow (\pi^+\pi^+\pi^-\pi^-)D^0 \rightarrow (\pi^+\pi^+\pi^-\pi^-)K_S^0)$	4 Stripping Lines 1 Data Types
$B^0 \rightarrow (D^0 \rightarrow (\pi^+\pi^-)D^0 \rightarrow (\pi^+\pi^-)K^*(892)^0)$	2 Stripping Lines 1 Data Types
$B^0 \rightarrow (D^0 \rightarrow (\pi^+\pi^-)D^0 \rightarrow (\pi^+\pi^-)\bar{K}^*(892)^0)$	3 Stripping Lines 1 Data Types



Select Lines

- Filtering of decay descriptors by multiple particle types
 - More advanced filtering options (e.g. decay head)
- Selection of multiple decay descriptors

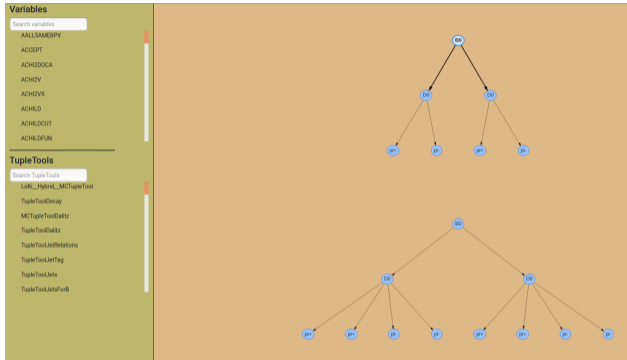
Lines Search


Decay	Line	Datatype	
$B^0 \rightarrow (D^0 \rightarrow (\pi^+\pi^-)D^0 \rightarrow (\pi^+\pi^-))$	StrippingB02D0D0D02HHD02H... ▾	2011 ▾	Remove
$B^0 \rightarrow (D^0 \rightarrow (\pi^+\pi^+\pi^-\pi^-)D^0 \rightarrow (\pi^+\pi^+\pi^-\pi^-))$	StrippingB02D0D0D02HHD02K3... ▾	2011 ▾	Remove

[Build Tree](#)

- List available stripping lines and running years for selected decay descriptors
-  Select multiple running years for the same decay descriptor
-  Provide enough documentation for users to choose the appropriate stripping line for their physics interest

Frontend: DecayTreeTuple configuration



- Visualisation of decay tree as interactive DAGs
- LoKi functor and TupleTool documentation in info blurbs
-  CRUD form for TupleTools on each node

The image shows a configuration interface for TupleTools and Variables. The 'TupleTools' panel contains a search bar with 'LoKi' and a list of tools including LoKi_Hybrid_TupleTool and LoKi_Hybrid_EvtTupleTool. A tooltip points to the first tool with the text 'The simple "hybrid"-based implementation of the abstract interface IParticleTupleTool.' The 'Variables' panel contains a search bar with 'pt|' and a list of variables including PT, ACCEPT, BPVPTFLIGHT, GPTDIR, GPTREL, MCPTDIR, MCPTREL, and PTDIR. A tooltip points to the search bar with the text 'Particle's transverse momentum.'

Problem

- Output of Ntuple Wizard must be data-structures to prevent code injection
- Gaudi apps can be configured natively with JSON or a bespoke `.opts` format
 - **X** Requires explicit configuration of all options incl. default values
- Want to configure only certain parts *à la* python options files → **custom parsers**

AnalysisHelpers

- *Importable* data-package, inspired by a trick used in AnalysisProductions
 - Not the same use-case as AppConfig
- Principal aim: **reduce boilerplate code** within options files
- Contains general helpful code for configuring DaVinci
 - Restripping, MC filtering
 - **DecayTreeTuple from YAML**
 - *etc...* (open for contributions)
- Release cycle decoupled from DaVinci
- Will be made available to AnalysisProductions

Conclusion

Pentaquarks

- Hints of new charmonium-pentaquark species, including one with strangeness

Open Data

- LHCb data format poses a challenge for Open Data release
 - Size and ease-of-use
- Ntuple Wizard offers a solution
 - Development of key features for first release in mature state
 - Our experience will provide input to Run 3 ntupling

Thanks for listening

Appendix

Open Data



<https://go-fair.org>

Findable: Persistent identifiers, rich metadata, indexed and searchable.

Accessible: Retrievable over standard protocols. Metadata accessible even if data isn't.

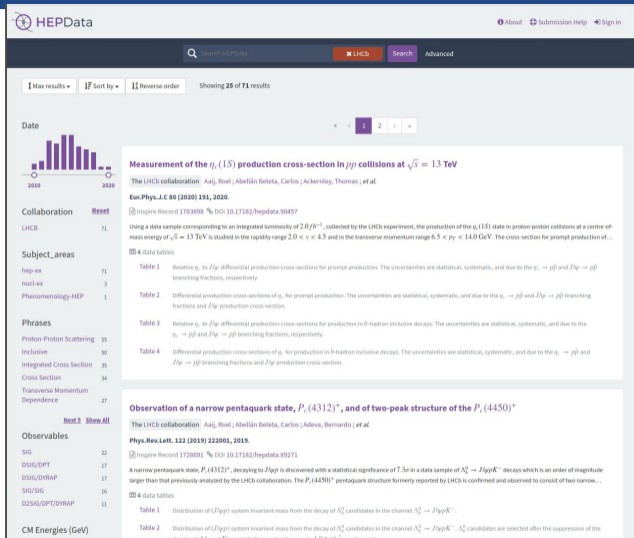
Interoperable: Appropriate language/format. Qualified references to other data.

Reusable: Clear usage licences. Detailed provenance. Meet community standards.



<https://hepdata.net>

- Mainly high-level observables / analysis results:
 - cross-sections
 - particle properties
 - SM parameters
 - limits and confidence intervals
- Machine-readable format



HEPData

Search HEPData

About Submission Help Sign In

Max results: 25 Sort by: Reverse order Showing 25 of 71 results

Date

Collaboration **Reset**

LHCb 71

Subject areas

hep-ex 71

nucl-ex 3

Phenomenology-HEP 1

Phrases

Proton-Proton Scattering 55

Inclusive 50

Integrated Cross Section 35

Cross Section 34

Transverse Momentum Dependence 27

Next 5 Show All

Observables

SIG 22

DSIG/DPT 17

DSIG/DVRAP 17

SIG/SIG 16

DSIG/DPT/DVRAP 11

CM Energies (GeV)

Measurement of the $\eta_c(1S)$ production cross-section in pp collisions at $\sqrt{s} = 13$ TeV

The LHCb collaboration Aaij, Ruel; Abellán Beteta, Carlos; Ackersley, Thomas; et al.
 Eur.Phys.J.C 80 (2020) 191, 2020.
 Inspire Record 1763398 % DOI 10.17182/hepdata.80457

Using a data sample corresponding to an integrated luminosity of $2.0/fb^{-1}$, collected by the LHCb experiment, the production of the $\eta_c(1S)$ state in proton-proton collisions at a centre-of-mass energy of $\sqrt{s} = 13$ TeV is studied in the rapidity range $2.0 < y < 4.5$ and in the transverse momentum range $6.5 < p_T < 14.0$ GeV. The cross-section for prompt production of...

4 data tables

Table 1 Relative η_c to J/ψ differential production cross-sections for prompt production. The uncertainties are statistical, systematic, and due to the $\eta_c \rightarrow \psi\psi$ and $J/\psi \rightarrow \psi\psi$ branching fractions, respectively.

Table 2 Differential production cross-sections of η_c , for prompt production. The uncertainties are statistical, systematic, and due to the $\eta_c \rightarrow \psi\psi$ and $J/\psi \rightarrow \psi\psi$ branching fractions and J/ψ production cross-section.

Table 3 Relative η_c to J/ψ differential production cross-sections for production in b -hadron inclusive decays. The uncertainties are statistical, systematic, and due to the $\eta_c \rightarrow \psi\psi$ and $J/\psi \rightarrow \psi\psi$ branching fractions, respectively.

Table 4 Differential production cross-sections of η_c for production in b -hadron inclusive decays. The uncertainties are statistical, systematic, and due to the $\eta_c \rightarrow \psi\psi$ and $J/\psi \rightarrow \psi\psi$ branching fractions and J/ψ production cross-section.

Observation of a narrow pentaquark state, $P_c(4312)^+$, and of two-peak structure of the $P_c(4450)^+$

The LHCb collaboration Aaij, Ruel; Abellán Beteta, Carlos; Adeva, Bernardo; et al.
 Phys.Rev.Lett. 122 (2019) 222001, 2019.
 Inspire Record 1728091 % DOI 10.17182/hepdata.89271

A narrow pentaquark state, $P_c(4312)^+$, decaying to $J/\psi p$ is discovered with a statistical significance of 7.3 σ in a data sample of $\Lambda_b^0 \rightarrow J/\psi p K^0$ decays which is an order of magnitude larger than that previously analyzed by the LHCb collaboration. The $P_c(4450)^+$ pentaquark structure formerly reported by LHCb is confirmed and observed to consist of two narrow...

4 data tables

Table 1 Distribution of $(J/\psi p)$ system invariant mass from the decay of Λ_b^0 candidates in the channel $\Lambda_b^0 \rightarrow J/\psi p K^0$.

Table 2 Distribution of $(J/\psi p)$ system invariant mass from the decay of Λ_b^0 candidates in the channel $\Lambda_b^0 \rightarrow J/\psi p K^0$. Λ_b^0 candidates are selected after the suppression of the...

Article II

Purposes

1. The Organization shall provide for collaboration among European States in nuclear research of a pure scientific and fundamental character, and in research essentially related thereto. The Organization shall have no concern with work for military requirements and the results of its experimental and theoretical work shall be published or otherwise made generally available.



SCOAP³ – Sponsoring Consortium for Open Access Publishing in Particle Physics

Home What is SCOAP³ SCOAP³ Partners SCOAP³ Journals SCOAP³ Repository FAQs Resources Contact Search

SCOAP³ converts high-quality subscription journals in the field of High-Energy Physics to Open Access through re-direction of subscription funds.

Access the Repository

Articles funded by SCOAP ³ :	40	734	1 248	40 867
	yesterday	last 30 days	in 2021	since 2014

‘International cooperation and dissemination of information have been enshrined in the CERN Convention for sixty years...’

Rolf Heuer, former CERN Director General

Article II Purposes

1. The Organization shall provide for collaboration among European States in nuclear research of a pure scientific and fundamental character, and in research essentially related thereto. The Organization shall have no concern with work for military requirements and the results of its experimental and theoretical work shall be published or otherwise made generally available.

The screenshot shows the Zenodo website interface. At the top, there is a blue header with the Zenodo logo, a search bar, and links for 'Upload' and 'Communities'. Below the header, there is a section for 'Featured communities'. The first featured community is the 'National COVID Cohort Collaborative (N3C)'. It includes a circular logo with a globe and a gear, a 'Browse' button, and a 'New upload' button. The text describes N3C as a partnership between various research hubs and data networks. Below this, there are navigation arrows and a link for 'Need help uploading? Contact us'.

Recent uploads

March 3, 2021 (v0.4.0) [Fullstack](#) [Open Access](#)
dfm/python-fsps: python-fsps v0.4.0

Ben Johnson, Dan Foreman-Mackey, Jonathan Sick, Joel Leja, Neil Byler, Mike Wainsley, Erik Tollerut, Henry Leung, Spencer Scott

A major refactor of the build infrastructure.

Uploaded on March 3, 2021

3 new version(s) exist for this record

[View](#)

Need help?

Contact us

Zenodo prioritizes all requested related to the COVID-19 outbreak.

We can help with:

- Uploading your research data, software, preprints, etc.
- One-on-one with Zenodo supporters.
- Quick responses to your specific needs.

General-purpose open-access repository developed under the European OpenAIRE program and operated by CERN

Article II

Purposes

1. The Organization shall provide for collaboration among European States in nuclear research of a pure scientific and fundamental character, and in research essentially related thereto. The Organization shall have no concern with work for military requirements and the results of its experimental and theoretical work shall be published or otherwise made generally available.

Explore more than **two petabytes** of open data from particle physics!

Start typing... Search

search examples: collides, datasets, keywords, education, energy, lhc

Explore

- [datasets](#)
- [software](#)
- [experiments](#)
- [documentation](#)

Focus on

- [ATLAS](#)
- [AULC](#)
- [CMS](#)
- [LHCb](#)
- [OESDA](#)
- [Data Science](#)

Get started

‘ The open data policy reflects CERN’s commitment to open science, which was already asserted in the CERN Convention over 60 years ago.’

Eckhard Elsen, former CERN Director for Research and Computing

Ntuple Wizard

Security & permissions

- **!** DaVinci typically configured with python scripts
 - **X Running arbitrary code from outside is a bad idea**
 - Wizard saves configuration as data structures, to be interpreted by our parsers
- **!** Dataset discovery requires metadata from Bookkeeping and LHCb stack
 - Pull all metadata at “deployment time”
 - **Read static files at runtime**, no interaction with DIRAC
- **!** Locations of “unreleased” datasets easily guessable
 - Check input data against allowed list
- **!** LHCb policy reserves right to withhold part of a dataset (e.g. ongoing analysis)
 - Require **fine-grained control** over:
 - building/accessing decay candidates
 - Stripping lines or equivalent selections
 - No elegant/agreed solution yet

Ntuple Wizard backend

URL discovery in two parts:

- Prefix for a specific application and version found from docs/db.json
 - e.g. DaVinci/v45r8 → DOC_008155
- Pages for all classes and namespaces extracted from {prefix}/search/search.idx into a JSON file, e.g.:

```
{  
  ...  
  "class TupleToolPid": "d8/ddd/class_tuple_tool_pid.html",  
  "class TupleToolPIDCalib": "d9/db9/class_tuple_tool_p_i_d_calib.html",  
  "Namespace LoKi__Cuts": "d7/dae/namespace_lo_ki_1_1_cuts.html",  
  ...  
}
```

- Pages parsed with [BeautifulSoup](#)

- Extract URL for `LoKi::Cuts` namespace:

https://lhcb-doxygen.web.cern.ch/lhcb-doxygen/davinci/v45r8/d7/dae/namespace_lo_ki_1_1_cuts.html

- Iterate through all `<div class="memitem">`
 - Name from `<td class="memname">`
 - Description from `<div class="memdoc">`
- Save `{name: description}` pairs to JSON file

- Extract URL for specific TupleTool class, e.g.:

https://lhcb-doxygen.web.cern.ch/lhcb-doxygen/davinci/v45r8/d8/ddd/class_tuple_tool_pid.html

- Take description from the first `<div class="textblock">`

Complete list of TupleTools discoverable from python:

```
from Configurables import __all__ as all_configurables
all_tupletools = list(filter(lambda t: "TupleTool" in t, all_configurables))
```

Interfaces extracted using `getPropertiesWithDescription`:

```
interfaces = {t: {} for t in all_tupletools}
for tool_name in all_tupletools:
    tool = getattr(Configurables, tool_name)()
    interfaces[tool_name] = tool.getPropertiesWithDescription()
```

Output contains option names, default values and descriptions:

```
"LoKi__Hybrid__MCTupleTool": {
  "Variables": [{}], "The {'name': 'functor'}-map of columns for N-tuple",
  "Preambulo": [{}], "The preambulo to be used for Bender/Python script",
  ...
},
```

Description (from Doxygen) and interface (from DaVinci) are combined into JSON entry:

```
"LoKi_Hybrid_MCTupleTool": {
  "description": "Simple implementation of IMCParticleTupleTool interface. ",
  "interface": [
    {
      "name": "Variables",
      "default": {},
      "description": "The {'name':'functor'}-map of columns for N-tuple",
      "type": "dict"
    },
    {
      "name": "Preamble",
      "default": [],
      "description": "The preamble to be used for Bender/Python script",
      "type": "list"
    },
    ...
  ]
},
```

Stripping lines: step 1

- Streams and lines for a specific Stripping version from `strippingArchive`
 - c.f. [any restripping example](#)
- Record output location, decay descriptors, stream, MDST flag
- Iterate recursively through dependency tree of algorithms
 - decay descriptors
 - PID substitutions
 - input locations
- Prune empty nodes, consolidate and save to JSON:

```
"StrippingBsPhiRhoLine": {  
  "MDST": false,  
  "decay": ["B0 -> phi(1020) rho(770)0"],  
  "inputs": [{  
    "decay": ["phi(1020) -> K+ K-", "rho(770)0 -> pi+ pi-"],  
    "inputs": []  
  }],  
  "output": "Phys/BsPhiRhoLine/Particles",  
  "stream": ["BhadronCompleteEvent"]  
},
```

Stripping lines: step 2

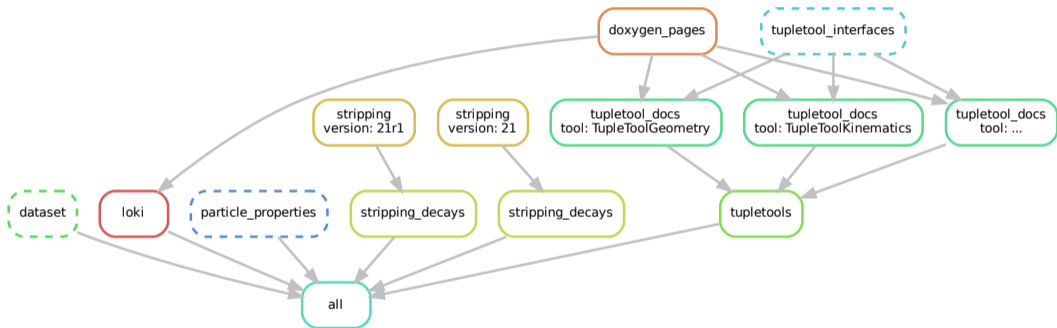
- Recursively substitute decaying particles with their descriptors until all combinations of decays are built
- Apply substitutions extracted from `SubstitutePID` and `SubPIDMMFilter`
- Parsing and substitution code from scratch with `pyparsing`
 - Tools within `Phys.SelPy.utils` broken and incomplete

```
"StrippingBsPhiRhoLine": {  
  "MDST": false,  
  "decay": ["B0 -> phi(1020) rho(770)0"],  
  "output": "Phys/BsPhiRhoLine/Particles",  
  "stream": ["BhadronCompleteEvent"],  
  "full_decays": ["B0 -> (phi(1020) -> K+ K-) (rho(770)0 -> pi+ pi-)"]  
},
```

- Gaudi names from `ParticleTable.txt`
- Properties, LaTeX string and antiparticle relation from `scikit-hep/particle`
- Save to JSON:

```
"B0": {  
  "mass": 5279.65,  
  "charge": 0,  
  "stable": false,  
  "pid": 511,  
  "latex": "B^{0}",  
  "antiparticle": "B~0",  
  "categories": ["Hadron", "Meson", "Down", "Beauty", "Bottom", "Scalar", "LongLived"]  
},
```


- Snakemake used to execute the workflow
 - Adopting suggested practices from [snakemake-workflows repo](#)



- GitLab CI executes the Snakemake workflow, pytest tests and linting

For those interested in developing this package:

```
od-ntuple-wizard/
|-- {COPYING, README.md}           # documentation
|-- {pyproject.toml, setup.cfg, setup.py} # standard python packaging stuff
|-- {deployment/, data/, logs/}    # output directories (created by snakemake)
|-- config.yml                     # versions, urls, paths etc
|-- scripts/
|   |-- {dump_*.sh} {dump_*.py}    # shell and Gaudi python scripts
|-- src/
|   |-- lhcb_ntuple_wizard/ # the main python module
|       |-- __init__.py
|       |-- __main__.py      # command line entry point (argparse)
|       |-- deployment.py    # metadata acquisition, parsing and formatting
|       |-- decays.py        # decay descriptor manipulation
|-- tests/
|   |-- test_*.py            # tests of all functions and classes (pytest)
|-- workflow/
|   |-- __init__.py # makes utils.py importable from snakemake
|   |-- Snakefile  # workflow automation rules
|   |-- utils.py   # helper functions for the snakemake rules
```

AnalysisHelpers

```
<?xml version="1.0" encoding="UTF-8"?>
<env:config xmlns:env="EnvSchema"
            xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
            xsi:schemaLocation="EnvSchema EnvSchema.xsd ">
  <env:set variable="ANALYSIS_HELPERS_BASE">${.}</env:set>
  <!-- Allow data package to be imported in python -->
  <env:prepend variable="PYTHONPATH">${.}</env:prepend>
</env:config>
```

AnalysisHelpers.xenv

```
from Configurables import DaVinci
from AnalysisHelpers.stripping import line_name_contains, restripping
from AnalysisHelpers.dst_writer import write_dst

restrip_sequence = restripping(
    version = "34",
    line_filter = line_name_contains("B2XTauNu"),
    stream_name = "AllStreams"
)
dst_writer = write_dst(
    selection_sequences = restrip_sequence.activeStreams(),
    extra_items = ["/Event/pRec/ProtoP#99", "/Event/pRec/Calo#99"]
)

DaVinci.appendToMainSequence([restrip_sequence, dst_writer.sequence()])
```

examples/Restrip_RXcHad_s34.py

Drastic reduction in amount of boilerplate code compared to StarterKit example.

```
from Configurables import DaVinci
from AnalysisHelpers.stripping import line_name_contains, stripping_mc_filter
from AnalysisHelpers.dst_writer import write_dst

filter_sequence = stripping_mc_filter(
    version = "34",
    line_filter = line_name_contains("B2XTauNu"),
    stream_name = "RXcHad.Strip"
)
dst_writer = write_dst(
    selection_sequences = filter_sequence.activeStreams(),
    extra_items = ["/Event/pRec/ProtoP#99", "/Event/pRec/Calo#99"]
)

DaVinci.appendToMainSequence([filter_sequence, dst_writer.sequence()])
```

examples/Filter_RXcHad_s34.py

Very similar interface to
restripping, with
sensible defaults for
MCFiltering production.

Use case: DecayTreeTuple from YAML

```
from Configurables import DaVinci
from AnalysisHelpers.decaytreetuple import configure_dtt
import yaml
config = yaml.safe_load(open("dtt_options.yml", "rb"))
for key in config:
    dtt = configure_dtt(key, config[key], verbose = True)
    DaVinci().UserAlgorithms += [dtt]
```

examples/dtt_from_yaml.py

Crucial component of
Ntuple Wizard