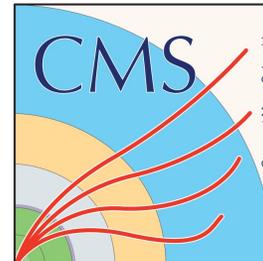


Open Data Policy for large LHC Experiments

J. Boyd

21/6/2021



Introduction

- There is an increasing interest in making experimental data Open
- Becoming more relevant for CERN
 - For example related to EU funding, but also from other funders
- Important for CERN to be play a leading role in Open Science
- The latest European Strategy report encouraged the development of internal policies on Open Data
- Somewhat triggered by a proposal made at the last WLCG preservation workshop in 2019
 - Suggesting that a common policy might be desirable & achievable

D. Large-scale data-intensive software and computing infrastructures are an essential ingredient to particle physics research programmes. The community faces major challenges in this area, notably with a view to the HL-LHC. As a result, the software and computing models used in particle physics research must evolve to meet the future needs of the field. ***The community must vigorously pursue common, coordinated R&D efforts in collaboration with other fields of science and industry, to develop software and computing infrastructures that exploit recent advances in information technology and data science. Further development of internal policies on open data and data preservation should be encouraged, and an adequate level of resources invested in their implementation.***

Open Data Policy Working Group for the LHC expts

- Initiated by Eckhard Elsen (then CERN DRC) in Feb 2020
- Working Group consists of 2 representatives each from the large LHC experiments (1 from TOTEM), and a representative from IT and SIS
 - RCS: J. Boyd (convener), E. Elsen (ex-officio)
 - SIS: K. Naim
 - IT: D. Duellmann
 - ALICE: J. Klein, S. Piano
 - ATLAS: L. Heinrich, A. Hoecker
 - CMS: K. Lassila-Perini, M. Pieri
 - LHCb: C. Bozzi, S. Neubert
 - TOTEM: L. Grzanka
- Mandated to draft a policy document relevant to the LHC experiments Open Data
- The working group has met 6 times to date to discuss and refine the policy and documents

WG Mandate

The European Commission is driving an Open Data Policy. CERN Council have encouraged the CERN experiments to make their data openly available for external analysis. The CERN Open Data portal is an initial attempt at making the data of the four experiments available through a common interface. So far this has largely been an experimental effort and a common strategy for Open Data has not been established.

This memo calls for forming a small **working group** to explore the conditions for a common Open Data policy from the LHC experiments. The **LHC collaborations are asked to nominate two representatives** each who will be complemented by one representative of RCS-SIS and IT each. The meetings will be convened by Jamie Boyd who reports to the Director of Research and Computing. The charge of the group is to draft a concise document on a common approach by the start of the summer addressing the following open points:

- Who are the expected users of the Open Data and what are their use cases?
 - o What would be the target typical event size for the data that is made available and what would a rough conceptual breakdown of the event content look like? What operations (tightening selections, re-calibrating objects, applying systematic uncertainties etc.) would be possible / not-possible with this event format?
 - o Would simulated data samples be made available with the data? If so which samples, and how would the total simulated data event size compare with the real data? Would there be a mechanism and/or recommendation for users to be able to obtain new simulated samples (e.g. of hypothetical signals)?
 - o What would be the typical total annual data size on the Open Data portal? (note for this to be useful this should be something that can be downloaded to, and processed at a typical university computing installation);
- After what latency period would the collaborations relinquish their exclusive access to experimental data?
- Should there be any rules or recommendations related to who can use the Open data, for example related to members of the experiment that the data is from, or competing experiments?

Strategy

- The policy has been broken down into the 4 levels of data as defined in the DPHEP study on data preservation:
 - Level 1 – Scientific publications, and associated additional data
 - Level 2 – Data useful for Education and Outreach
 - Level 3 – Reconstruction level data useful for general physics analysis
 - Level 4 – RAW data
- All experiments already release data for L1 and L2 in broadly similar ways, and all agree that L4 is not practically useful
 - The WG discussion therefore strongly focussed on the policy for L3 data
- Decided to write a short general policy document (to be made public) outlining the general principles, and in addition an internal document outlining the implementation of the policy by each experiment
 - Implementation strategy for each experiment cannot be changed without discussing in the WG

Public Policy Document

- 2 page document:
- Introduction:
 - Motivation and scope
 - Outline DPHEP levels of data
- L1: Published results policy:
 - Publish results in Open Access journals
 - Provide additional information through HepData / RIVET etc..
 - No change in current policy
- L2: Outreach and Education
 - Provide rich data samples tailored for outreach/education (easy to use)
- L3: Reconstructed data
 - See next slides...
 - Real change to Collaborations policy
- L4: Raw Data
 - Not useful for external people

The CERN Open Data Policy reflects values that have been enshrined in the CERN Convention for more than sixty years that were reaffirmed in the European Strategy for Particle Physics (2020)¹, and aims to empower the LHC experiments to adopt a consistent approach towards the openness and preservation of experimental data. Making data available responsibly (applying FAIR standards²), at different levels of abstraction and at different points in time, allows the maximum realisation of their scientific potential and the fulfillment of the collective moral and fiduciary responsibility to member states and the broader global scientific community. CERN understands that in order to optimise reuse opportunities, immediate and continued resources are needed. The level of support that CERN and the experiments will be able to provide to external users will depend on available resources.

This policy relates to the data collected by the LHC experiments, for the main physics programme of the LHC — high-energy proton–proton and heavy-ion collision data. The foreseen use cases of the Open Data include reinterpretation and reanalysis of physics results, education and outreach, data analysis for technical and algorithmic developments and physics research. The Open Data will be released through the CERN Open Data Portal which will be supported by CERN for the lifetime of the data. The data will be tailored to the different uses, and will be made available in formats defined by each experiment that afford a range of opportunities for long-term use, reuse and preservation. In general, four levels of complexity of HEP data have been identified by the Data Preservation and Long Term Analysis in High Energy Physics (DPHEP) Study Group³, which serve varying audiences and imply a diversity of openness solutions and practices.

Published Results (Level 1) Policy: Peer-reviewed publications represent the primary scientific output from the experiments. In compliance with the CERN Open Access Policy, all such publications are available with Open Access, and so are available to the public. To maximise the scientific value of their publications, the experiments will make public additional information and data at the time of publication, stored in collaboration with portals such as HEPData,⁴ with selection routines stored in specialised tools. The data made available may include simplified or full binned likelihoods, as well as unbinned likelihoods based on datasets of event-level observables extracted by the analyses. Reinterpretation of published results is also made possible through analysis preservation and direct collaboration with external researchers.

Outreach and Education (Level 2) Policy: For the purposes of education and outreach, dedicated subsets of data are used, selected and formatted to provide rich samples to maximise their educational impact, and to facilitate the easy use of the data. These data are released with a schedule and scope determined by each experiment. The data are provided in simplified, portable and self-contained formats suitable for educational and public understanding purposes; but are not intended nor adequate for the publication of scientific results. Lightweight environments to allow the easy

also be provided. CERN experiments will make data of such high level of e through the CERN Open Data Portal.⁵

Policy: The LHC experiments will release calibrated reconstructed data for algorithmic, performance and physics studies. The release of these provenance metadata, and by a concurrent release of appropriate ware, reproducible example analysis workflows, and documentation. nts that are compatible with the data and software will be made ovided will be sufficient to allow high-quality analysis of the data llication of the main correction factors and corresponding systematic ations, detector reconstruction and identification. A limited level of vel 3 Open Data will be provided on a best-effort basis by the

periodically, following an appropriate latency period to allow thorough re reconstruction and calibrations, as well as to allow time for the data by the collaboration. The size of the released datasets will be amount of data collected of similar type, with the aim to commence s of the conclusion of the run period. Data may be withheld by an analyses ongoing. Full datasets will be made available at the close of the

¹ European Strategy Group (2020), '2020 Update of the European Strategy for Particle Physics'.

² FAIR Guiding Principles for scientific data management and stewardship. Available at: <https://www.go-fair.org/fair-principles/>.

³ Data management plans are defined by the LHC experiments to address the long-term preservation of internal data products. See: Akopov et al., Status report of the DPHEP Study Group: Towards a global effort for sustainable data preservation in high energy physics. arXiv preprint arXiv:1205.4667 (2012).

⁴ Repository for publication-related High-Energy Physics data: <http://www.hepdata.net>.

The data will be released from the CERN Open Data Portal under the Creative Commons CC0 waiver, and will be identified with persistent data identifiers, and the data must be cited through these identifiers. Similarly, appropriate acknowledgements of the experiment(s) should be included in publications released using such data, and the publications made clearly distinguishable from those released by the collaboration. Any scientific claims in such publications are the responsibility of their authors and not of the experiments. It is expected that scientific results released using Open Data follow best scientific practices. The experiments may impose rules related to the use of the data by members of their respective collaborations.

External authors should be aware that they will not have access to the vast amount of tacit knowledge built up within the LHC collaborations over the decades of design, construction and operation of the experimental apparatus. To allow external scientists to fully benefit from all the data, knowledge and tools, the collaborations may offer appropriate association programmes.

Raw Data (Level 4) Policy: It is not practically possible to make the full raw data-set from the LHC experiments usable in a meaningful way outside the collaborations. This is due to the complexity of the data, metadata and software, the required knowledge of the detector itself and the methods of reconstruction, the extensive computing resources necessary and the access issues for the enormous volume of data stored in archival media. It should be noted that, for these reasons, general direct access to the raw data is not even available to individuals within the collaboration, and that instead the production of reconstructed data (i.e. Level-3 data) is performed centrally. Access to representative subsets of raw data—useful for example for studies in the machine learning domain and beyond—can be released together with Level-3 formats, at the discretion of each experiment.

⁵ CERN Open Data portal: <http://opendata.cern.ch>.

L3 data – discussion - 1

- Any rules relating to publically releasing L3 data need to be approved by each experiments Collaboration Board
- Generally tried to find a good balance between:

1. Making data openly available

And

2. Protecting the collaborations:

- Avoiding collaboration members publishing with open data rather than a Collaboration paper
- Having to deal with wild claims made by external analysts with their data
- Not taking too much resources (both human and computing) from the Collaboration
- After much discussion converged on limiting of the amount of data released after a given time as a tool to make it unattractive for collaboration members to publish with Open Data
 - Here a common approach across experiments is needed since e.g. ATLAS rules could effect a CMS collabotor's behaviour in this regard
 - Exact fraction of data released after what latency experiment specific, but general principle common across experiments

L3 data – discussion - 3

- We should only release L3 Open Data if it can be used for high quality science
 - Release data and simulated samples
 - Release data with best calibrations available
 - Release sufficient information to allow main systematic uncertainties to be applied
 - Release analysis s/w to allow efficient analysis of data
- Data expected to be useable not only for particle physics, but also computing, algorithmic development, big data studies etc..
- Publications using Open Data should:
 - Have appropriate acknowledgements
 - Be clearly identifiable from Collaboration papers
 - Follow best scientific practices
- All experiments to use CERN Open Data Portal to house the released data

Text in public policy document on releasing L3 data

To be released:

- calibrated data
- accompanying Monte Carlo
- Analysis software
- Software environments (VM, containers)

importance of sufficient latency period.

- start 5y after run period
- guideline:
more data → more open data
- full data at close of expt
- escape hatch for ongoing analyses

Reconstructed Data (Level 3) Policy: The LHC experiments will release calibrated reconstructed data with the level of detail useful for algorithmic, performance and physics studies. The release of these data will be accompanied by provenance metadata, and by a concurrent release of appropriate simulated data samples, software, reproducible example analysis workflows, and documentation. Virtual computing environments that are compatible with the data and software will be made available. The information provided will be sufficient to allow high-quality analysis of the data including, where practical, application of the main correction factors and corresponding systematic uncertainties related to calibrations, detector reconstruction and identification. A limited level of support for users of the Level 3 Open Data will be provided on a best-effort basis by the collaborations.

Public data releases will occur periodically, following an appropriate latency period to allow thorough understanding of the data, the reconstruction and calibrations, as well as to allow time for the scientific exploitation of the data by the collaboration. The size of the released datasets will be commensurate with the total amount of data collected of similar type, with the aim to commence data releases within five years of the conclusion of the run period. Data may be withheld by an experiment if there are active analyses ongoing. Full datasets will be made available at the close of the collaboration.

The data will be released from the CERN Open Data Portal under the Creative Commons CC0 waiver, and will be identified with persistent data identifiers, and the data must be cited through these identifiers. Similarly, appropriate acknowledgements of the experiment(s) should be included in publications released using such data, and the publications made clearly distinguishable from those released by the collaboration. Any scientific claims in such publications are the responsibility of their authors and not of the experiments. It is expected that scientific results released using Open Data follow best scientific practices. The experiments may impose rules related to the use of the data by members of their respective collaborations.

External authors should be aware that they will not have access to the vast amount of tacit knowledge built up within the LHC collaborations over the decades of design, construction and operation of the experimental apparatus. To allow external scientists to fully benefit from all the data, knowledge and tools, the collaborations may offer appropriate association programmes.

need for sufficient quality in OD

commit only to best-effort support

data citation (of course) required

but waive responsibility for results

must be clearly marked as external publication

rules for use for own collaborators possible (but not required)

emphasize know-how asymmetry and advertise association programs

Text in public policy document on releasing L3 data

To be released:

- calibrated data
- accompanying Monte Carlo
- Analysis software
- Software environments (VM, containers)

importance of sufficient latency period.

- start 5y after run period
- guideline:
more data →
- full data at clo
- escape hatch
analyses

Reconstructed Data (Level 3) Policy: The LHC experiments will release calibrated reconstructed data with the level of detail useful for algorithmic, performance and physics studies. The release of these data will be accompanied by provenance metadata, and by a concurrent release of appropriate simulated data samples, software, reproducible example analysis workflows, and documentation. Virtual computing environments that are compatible with the data and software will be made available. The information provided will be sufficient to allow high-quality analysis of the data including, where practical, application of the main correction factors and corresponding systematic uncertainties related to calibrations, detector reconstruction and identification. A limited level of support for users of the Level 3 Open Data will be provided on a best-effort basis by the collaborations.

Public data releases will occur periodically, following an appropriate latency period to allow thorough understanding of the data, the reconstruction and calibrations, as well as to allow time for the scientific exploitation of the data by the collaboration. The size of the released datasets will be commensurate

released by the collaboration. Any scientific claims in such publications are the responsibility of their authors and not of the experiments. It is expected that scientific results released using Open Data follow best scientific practices. The experiments may impose rules related to the use of the data by members of their respective collaborations.

External authors should be aware that they will not have access to the vast amount of tacit knowledge built up within the LHC collaborations over the decades of design, construction and operation of the experimental apparatus. To allow external scientists to fully benefit from all the data, knowledge and tools, the collaborations may offer appropriate association programmes.

need for sufficient quality in OD

commit only to best-effort support

The policy and implementation documents have been endorsed by the collaboration board of the 4 large LHC experiments during the last quarter of 2020!

(course) required

responsibility for results

must be clearly marked as external publication

rules for use for own collaborators possible (but not required)

emphasize know-how asymmetry and advertise association programs

Next Steps...

Implementation:

- We are now working with the experiments, CERN IT, and SIS to come up with a detailed implementation strategy for the OD policy
- This needs to limit the needed human resources (both from IT/SIS and within the experiments) and the long term costs of storage media
 - (e.g. using TAPE instead of disk for large storage)
- Some of this will be discussed in the LHC session of this workshop

Expanding the policy to cover all CERN experiments:

- There is a longer term desire to expand the policy to (where possible) cover the full suite of experiments at CERN (including future experiments)
- To start with we are discussing with the small LHC experiments about their ability to sign up to the existing policy document
 - Currently have discussed with TOTEM, LHCf and FASER who seem generally positive
 - Need to start the discussion with MoEDAL, SND@LHC

Summary

- Open Data Working group started working in Feb 2020, and has produced two Open Data policy documents endorsed by the 4 large LHC Collaborations:
 - A public policy document,
 - An internal 'implementation' document
- Many thanks to the WG members for their valuable input in this process
- We believe the documents find a good balance between promoting Open'ness of the experimental data, while protecting the Collaborations
- Now working on the detailed implementation of the policy
 - Need to find a good path to limit the needed human resources (both from CERN IT/SIS) and within the experiments, as well as to limit the media storage costs (e.g. by using TAPE storage)
- The public document could serve as a good starting point for a CERN wide policy, or for other CERN experiments to follow
 - Already starting to see if other smaller LHC experiments can sign up to the principles in the policy document
 - Further in the future would try to get ALL CERN experiments to do this

Backup...

Public Policy Document - Introduction

The CERN Open Data Policy reflects values that have been enshrined in the CERN Convention for more than sixty years that were reaffirmed in the European Strategy for Particle Physics (2020)¹, and aims to empower the LHC experiments to adopt a consistent approach towards the openness and preservation of experimental data. Making data available responsibly (applying FAIR standards²), at different levels of abstraction and at different points in time, allows the maximum realisation of their scientific potential and the fulfillment of the collective moral and fiduciary responsibility to member states and the broader global scientific community. CERN understands that in order to optimise reuse opportunities, immediate and continued resources are needed. The level of support that CERN and the experiments will be able to provide to external users will depend on available resources.

This policy relates to the data collected by the LHC experiments, for the main physics programme of the LHC — high-energy proton–proton and heavy-ion collision data. The foreseen use cases of the Open Data include reinterpretation and reanalysis of physics results, education and outreach, data analysis for technical and algorithmic developments and physics research. The Open Data will be released through the CERN Open Data Portal which will be supported by CERN for the lifetime of the data. The data will be tailored to the different uses, and will be made available in formats defined by each experiment that afford a range of opportunities for long-term use, reuse and preservation. In general, four levels of complexity of HEP data have been identified by the Data Preservation and Long Term Analysis in High Energy Physics (DPHEP) Study Group³, which serve varying audiences and imply a diversity of openness solutions and practices.

L1 data - text

Published Results (Level 1) Policy: Peer-reviewed publications represent the primary scientific output from the experiments. In compliance with the CERN Open Access Policy, all such publications are available with Open Access, and so are available to the public. To maximise the scientific value of their publications, the experiments will make public additional information and data at the time of publication, stored in collaboration with portals such as HEPData,⁴ with selection routines stored in specialised tools. The data made available may include simplified or full binned likelihoods, as well as unbinned likelihoods based on datasets of event-level observables extracted by the analyses. Reinterpretation of published results is also made possible through analysis preservation and direct collaboration with external researchers.

L2 data - text

Outreach and Education (Level 2) Policy: For the purposes of education and outreach, dedicated subsets of data are used, selected and formatted to provide rich samples to maximise their educational impact, and to facilitate the easy use of the data. These data are released with a schedule and scope determined by each experiment. The data are provided in simplified, portable and self-contained formats suitable for educational and public understanding purposes; but are not intended nor adequate for the publication of scientific results. Lightweight environments to allow the easy exploration of these data may also be provided. CERN experiments will make data of such high level of abstraction available, accessible through the CERN Open Data Portal.⁵

L4 data - text

Raw Data (Level 4) Policy: It is not practically possible to make the full raw data-set from the LHC experiments usable in a meaningful way outside the collaborations. This is due to the complexity of the data, metadata and software, the required knowledge of the detector itself and the methods of reconstruction, the extensive computing resources necessary and the access issues for the enormous volume of data stored in archival media. It should be noted that, for these reasons, general direct access to the raw data is not even available to individuals within the collaboration, and that instead the production of reconstructed data (i.e. Level-3 data) is performed centrally. Access to representative subsets of raw data—useful for example for studies in the machine learning domain and beyond—can be released together with Level-3 formats, at the discretion of each experiment.

CMS Open Data Results

- CMS have been a pioneer of releasing L3 Open Data
- There have been a number of publications on this data covering BSM, QCD, jet-reconstruction and machine learning topics
- Also interesting presentations on CMS Open Data at SnowMass:

<https://indico.fnal.gov/event/43829/timetable/?view=standard#b-18364-compf7-reinterpretatio>

Adversarially Learned Anomaly Detection on CMS Open Data: re-discovering the top quark Oliver Knapp (Zurich, ETH), Guenther Dissertori (Zurich, ETH), Olmo Cerri (Caltech), Thong Q. Nguyen (Caltech), Jean-Roch Vlimant (Caltech) et al. (May 4, 2020) e-Print: 2005.01598 [hep-ex] pdf cite 6 citations	#1
End-to-end particle and event identification at the Large Hadron Collider with CMS Open Data John Alison (Carnegie Mellon U.), Sitong An (CERN), Patrick Bryant (Carnegie Mellon U.), Bjorn Burkle (Carnegie Mellon U. and Brown U.), Sergei Gleyzer (Alabama U.) et al. (Oct 15, 2019) Contribution to: DPF2019 • e-Print: 1910.07029 [hep-ex] pdf cite 1 citation	#2
Exploring the Space of Jets with CMS Open Data Patrick T. Komiske (MIT, Cambridge, CTP and Harvard U.), Rajha Mastandrea (MIT, Cambridge, CTP), Eric M. Metodiev (MIT, Cambridge, CTP and Harvard U.), Preksha Naik (MIT, Cambridge, CTP), Jesse Thaler (MIT, Cambridge, CTP and Harvard U.) (Aug 22, 2019) Published in: <i>Phys.Rev.D</i> 101 (2020) 3, 034009 • e-Print: 1908.08542 [hep-ph] pdf DOI cite 9 citations	#3
Explicit jet veto as a tool to purify the underlying event in the Drell-Yan process using CMS Open Data Saeid Pakinat Mehdiabadi (Yazd U. and IPM, Tehran), Ali Fahim (U. Tehran, Coll. Engin.) (Jul 20, 2019) Published in: <i>J.Phys.G</i> 46 (2019) 9, 095003 • e-Print: 1907.08842 [hep-ph] pdf DOI cite 1 citation	#4
Opportunities and challenges of Standard Model production cross section measurements in proton-proton collisions at $\sqrt{s} = 8$ TeV using CMS Open Data Aram Apyan (Fermilab), William Cuzzo (MIT), Markus Klute (MIT), Yoshihiro Saito (MIT), Matthias Schott (MIT and Mainz U.) et al. (Jul 18, 2019) Published in: <i>JINST</i> 15 (2020) 01, P01009 • e-Print: 1907.08197 [hep-ex] pdf links DOI cite 3 citations	#5
Testing non-standard sources of parity violation in jets at the LHC, trialed with CMS Open Data Christopher G. Lester (Cambridge U.), Matthias Schott (MIT and Mainz U.) (Apr 25, 2019) Published in: <i>JHEP</i> 12 (2019) 120 • e-Print: 1904.11195 [hep-ex] pdf DOI cite 5 citations	#6
End-to-end jet classification of quarks and gluons with the CMS Open Data M. Andrews (Carnegie Mellon U.), J. Alison (Carnegie Mellon U.), S. An (Carnegie Mellon U. and CERN), Patrick Bryant , B. Burkle (Brown U.) et al. (Feb 21, 2019) Published in: <i>Nucl.Instrum.Meth.A</i> 977 (2020) 164304 • e-Print: 1902.08276 [hep-ex] pdf DOI cite 10 citations	#7
Searching in CMS Open Data for Dimuon Resonances with Substantial Transverse Momentum Cari Cesarotti (Harvard U.), Yotam Soreq (CERN and Technion), Matthew J. Strassler (Harvard U.), Jesse Thaler (Harvard U. and MIT, Cambridge, CTP), Wei Xue (CERN) (Feb 11, 2019) Published in: <i>Phys.Rev.D</i> 100 (2019) 1, 015021 • e-Print: 1902.04222 [hep-ph] pdf DOI cite 4 citations	#8
End-to-End Physics Event Classification with CMS Open Data: Applying Image-Based Deep Learning to Detector Data for the Direct Classification of Collision Events at the LHC M. Andrews (Carnegie Mellon U.), M. Paulini (Carnegie Mellon U.), S. Gleyzer (Florida U.), B. Poczopko (Carnegie Mellon U.) (Jul 31, 2018) Published in: <i>Comput.Softw.Big Sci.</i> 4 (2020) 1, 6 • e-Print: 1807.11916 [hep-ex] pdf DOI cite 16 citations	#9
Jet Substructure Studies with CMS Open Data Aashish Tripathy (MIT, Cambridge, CTP), Wei Xue (MIT, Cambridge, CTP), Andrew Larkoski (Reed Coll.), Simone Marzani (SUNY, Buffalo), Jesse Thaler (MIT, Cambridge, CTP) (Apr 19, 2017) Published in: <i>Phys.Rev.D</i> 96 (2017) 7, 074003 • e-Print: 1704.05842 [hep-ph] pdf DOI cite 31 citations	#10
Exposing the QCD Splitting Function with CMS Open Data Andrew Larkoski (Reed Coll.), Simone Marzani (SUNY, Buffalo), Jesse Thaler (MIT, Cambridge, CTP), Aashish Tripathy (MIT, Cambridge, CTP), Wei Xue (MIT, Cambridge, CTP) (Apr 17, 2017) Published in: <i>Phys.Rev.Lett.</i> 119 (2017) 13, 132003 • e-Print: 1704.05066 [hep-ph] pdf DOI cite 33 citations	#11