IPUs for HEP

Kostas Petridis & Jonas Rademacker, University of Bristol
members of and potentially biased towards, but not speaking on behalf of, LHCb (both), DUNE (JR), SHiP (KP)

ECHEP, Edinburgh, 17 Feb 2020

IPUs for HEP

University of BRISTOL

The Jean Golding* Institute

link/support/funding ← The Alan Turing Institute

School of Physics

GRAPHCORE

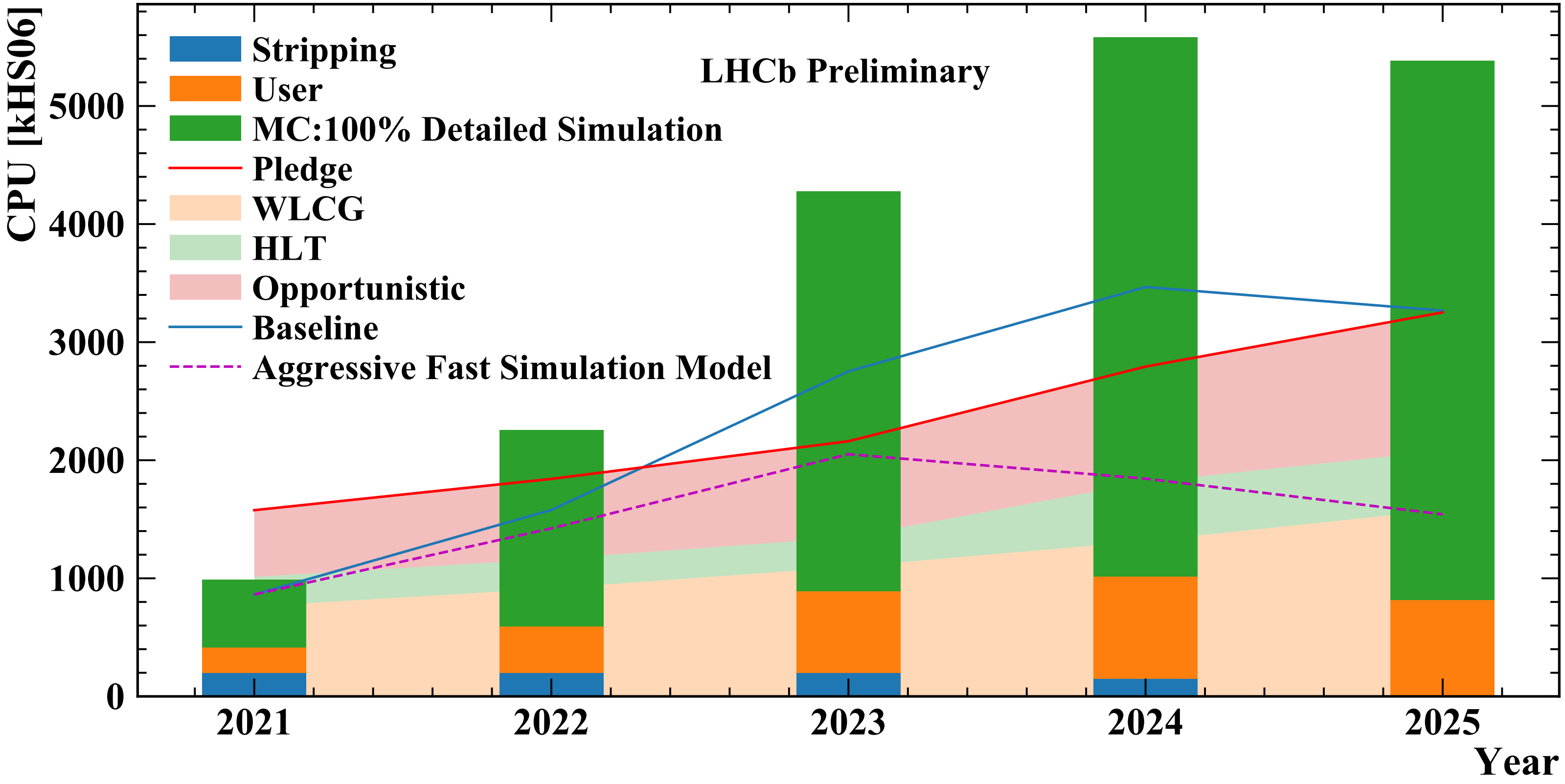
Initial project

Proof or Principle: Use IPUs for

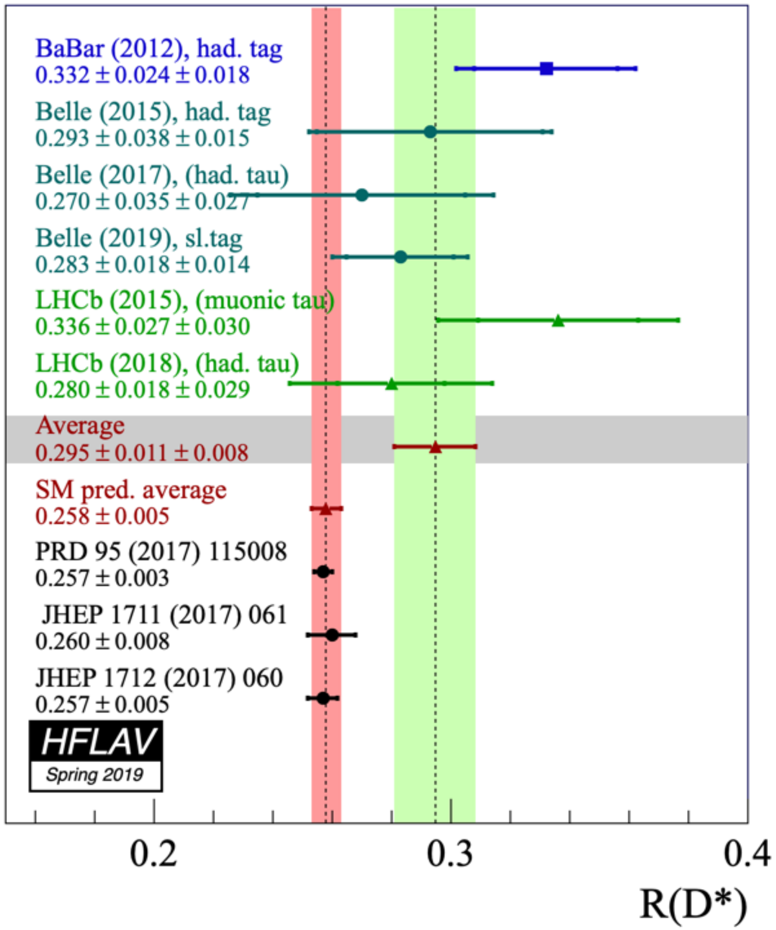
- Fast event generation
- Event reconstruction
- Data analysis (amplitude analysis)
- Trigger/Online computing

*  Professor, Jean Golding, mathematician, epidemiologist and founder of the Children of the 90s cohort study.

Event Generation



R(D*) and MC sample size



HFLAV, arXiv:1909.12524

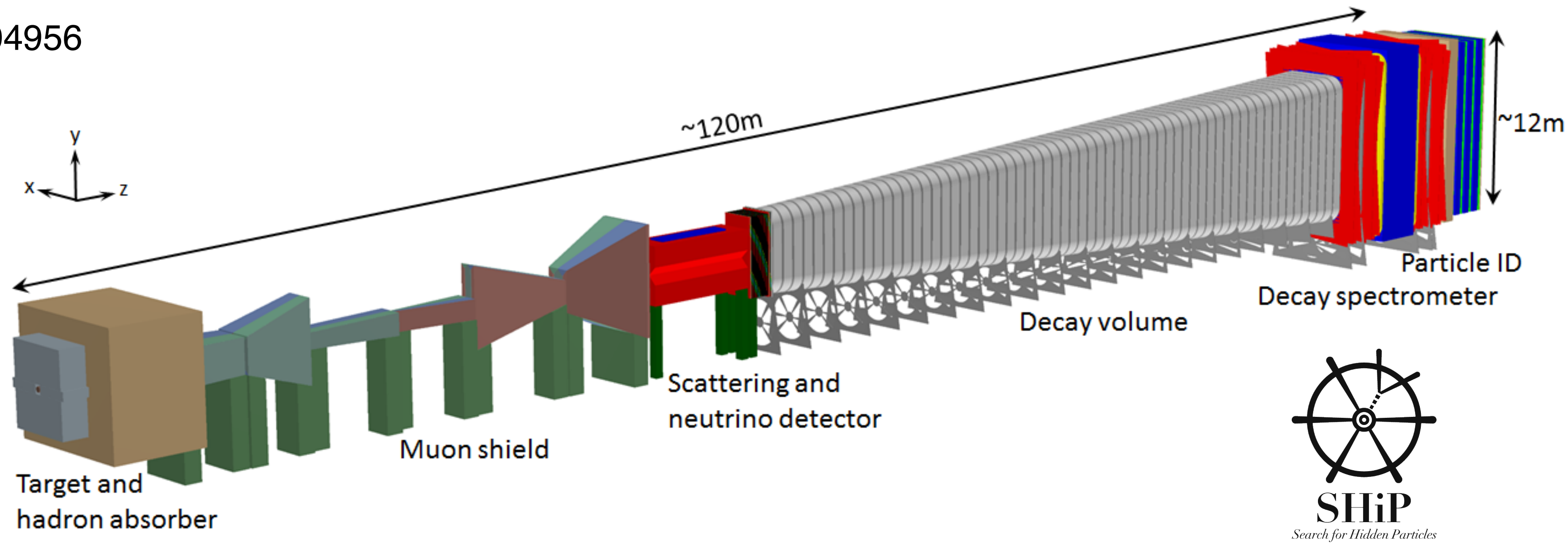
systematic uncertainties in LHCb measurement

Contribution	Value in %
$\mathcal{B}(\tau^+ \to 3\pi\overline{\nu}_\tau)/\mathcal{B}(\tau^+ \to 3\pi(\pi^0)\overline{\nu}_\tau)$	0.7
Form factors (template shapes)	0.7
Form factors (efficiency)	1.0
τ polarization effects	0.4
Other τ decays	1.0
$B \to D^{**}\tau^+\nu_\tau$	2.3
$B_s^0 \to D_s^{**}\tau^+\nu_\tau$ feed-down	1.5
$D_s^+ \to 3\pi X$ decay model	2.5
D_s^+, D^0 and D^+ template shape	2.9
$B \to D^{*-}D_s^+(X)$ and $B \to D^{*-}D^0(X)$ decay model	2.6
$D^{*-}3\pi X$ from B decays	2.8
Combinatorial background (shape + normalization)	0.7
Bias due to empty bins in templates	1.3
Size of simulation samples	4.1
Trigger acceptance	1.2
Trigger efficiency	1.0
Online selection	2.0
Offline selection	2.0
Charged-isolation algorithm	1.0
Particle identification	1.3
Normalization channel	1.0
Signal efficiencies (size of simulation samples)	1.7
Normalization channel efficiency (size of simulation samples)	1.6
Normalization channel efficiency (modeling of $B^0 \to D^{*-}3\pi$)	2.0
Total uncertainty	9.1

LHCb R(D*) (hadronic tau) PRL 120 (2018) no.17, 171802
PD97 (2018) no.7, 072013

Ship: Simulating the tails

arXiv:1504.04956

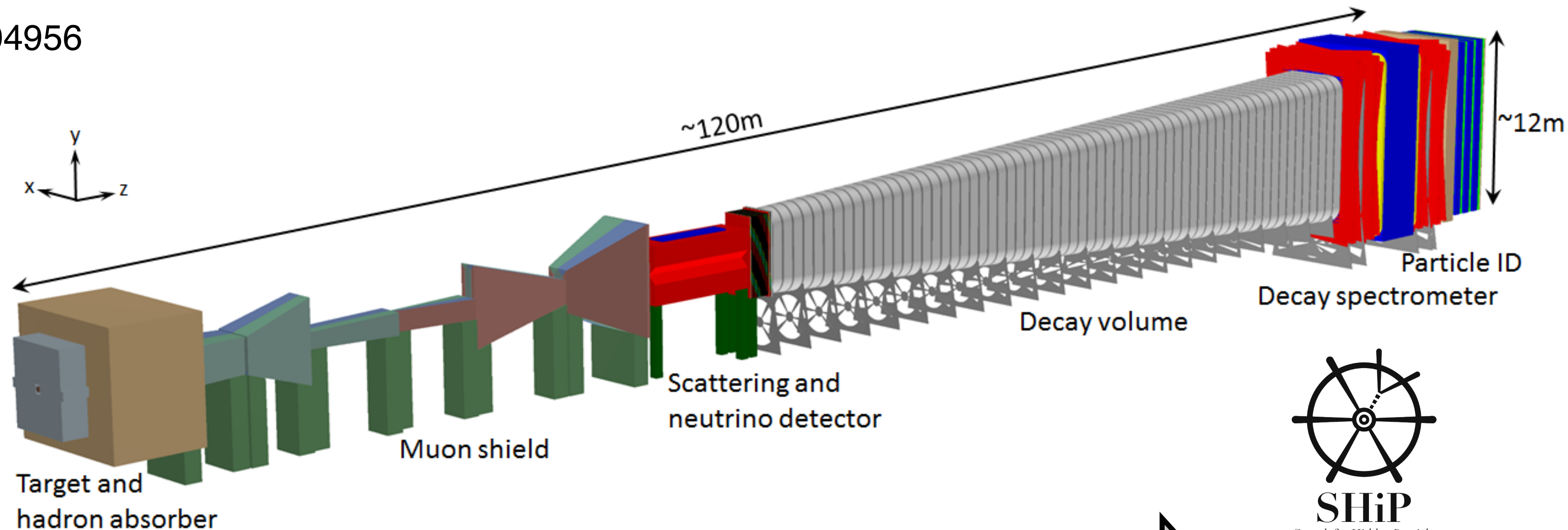


The SHiP experiment will search for new, very weakly interacting long-lived particles produced in a 400 GeV/c SPS proton beam dump.

Start data taking ~2026

Ship: Simulation challenge

arXiv:1504.04956



SHiP
Search for Hidden Particles

$2 \cdot 10^{20}$ protons

$5 \cdot 10^{16}$ muons

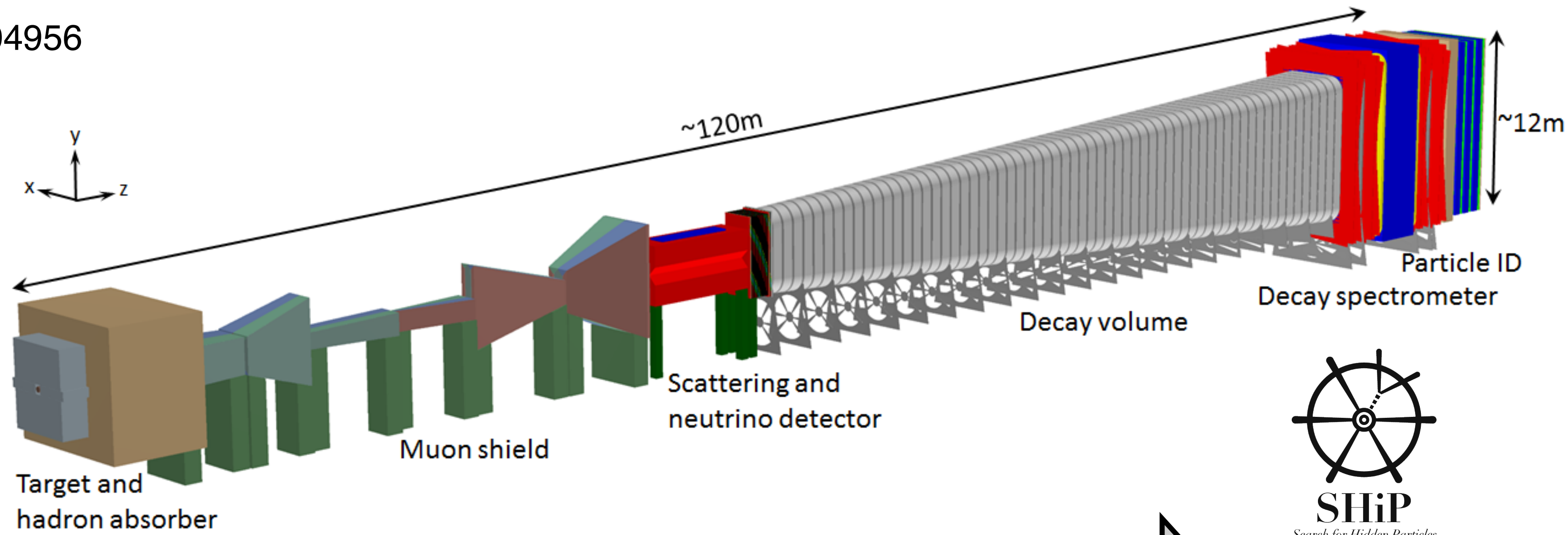
active and passive shields, event selection

<0.1 background event

Large simulated samples of muon induced background processes are required in order to optimise the experimental facility and develop background suppression methods.

Ship: Simulation challenge

arXiv:1504.04956



$2 \cdot 10^{20}$ protons

$5 \cdot 10^{16}$ muons

active and passive shields, event selection

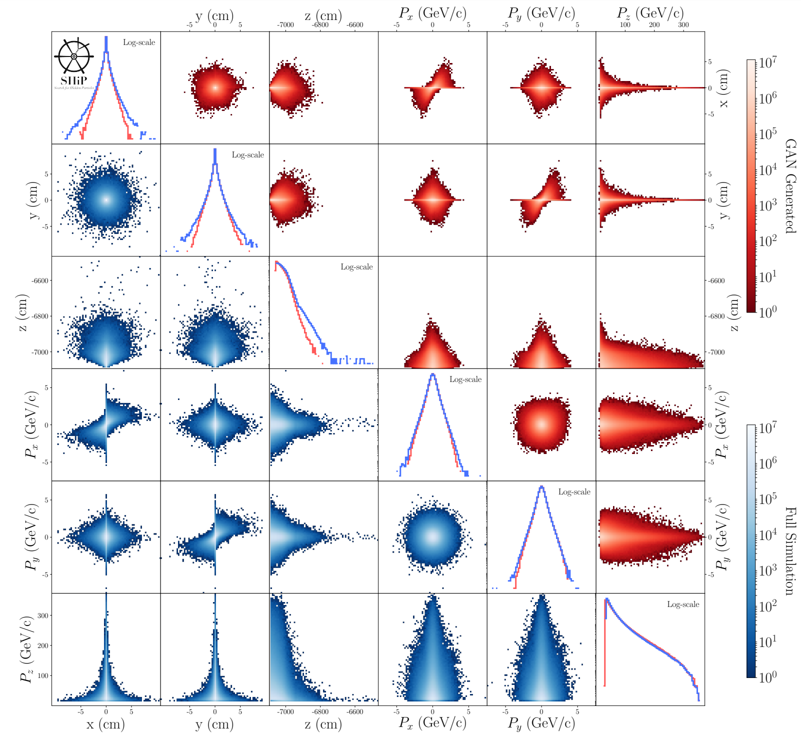
<0.1 background event

This part of the process is computationally very expensive!

Fast simulation of muons produced at the SHiP experiment using GANs

A. Marshall + K. Petridis **JINST 14 (2019) P11028**

- Use GANs to simulate proton interaction with the dense fixed target.
- The residual mis-modelling of the GAN approach was found to have a minimal impact on the muon rate reaching the decay volume
- Using this approach can produce an equivalent sample of fully simulated muons $\mathcal{O}(10^6)$ times faster.

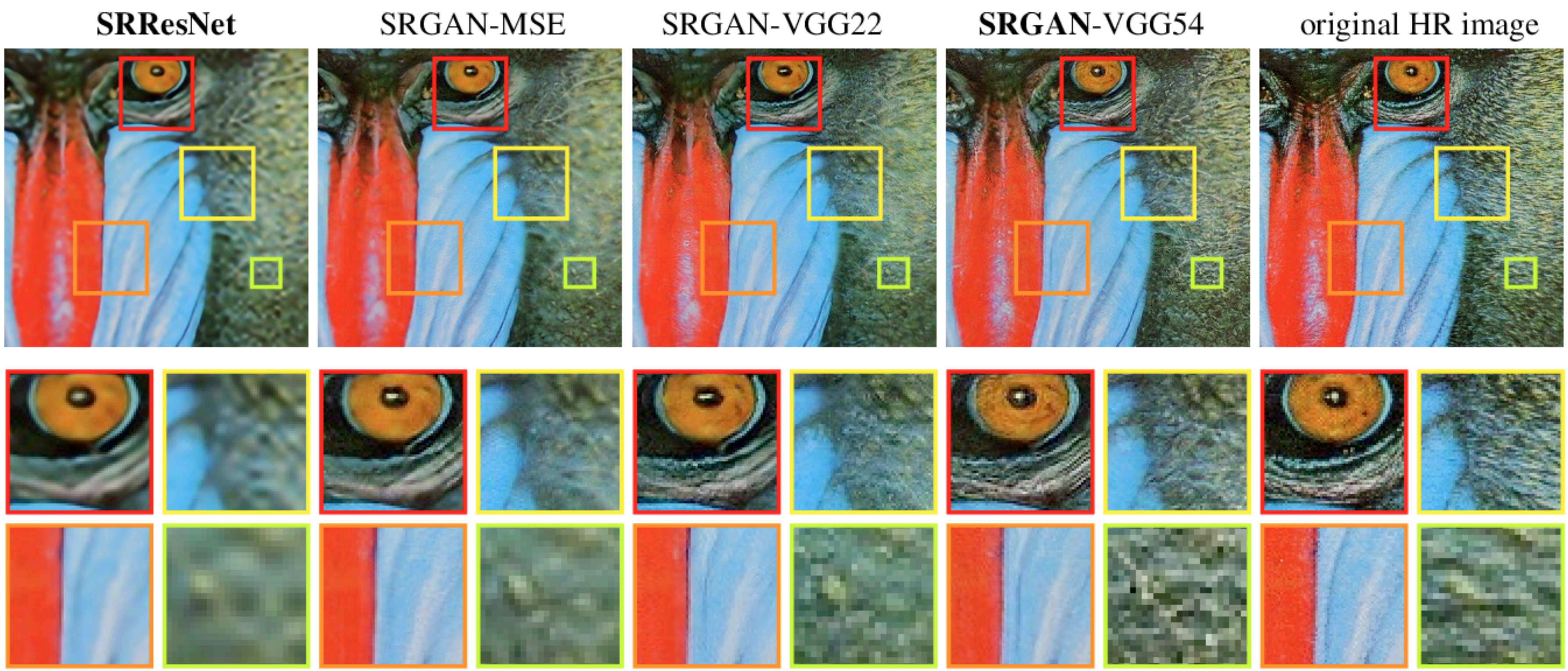


Target simulation method	Muons produced in 5 minutes	Time to simulate single muon (s)
Pythia8 and GEANT4	~ 1	1.1×10^{-1}
GAN (CPU)	7.5×10^{5}	4.0×10^{-4}
GAN (GPU)	3.5×10^{6}	8.6×10^{-5}
GAN(IPU)	?	?

Super-resolution with GANs

Ledig, Christian et al. "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network." 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016): 105-114.

Super-resolution with GANs

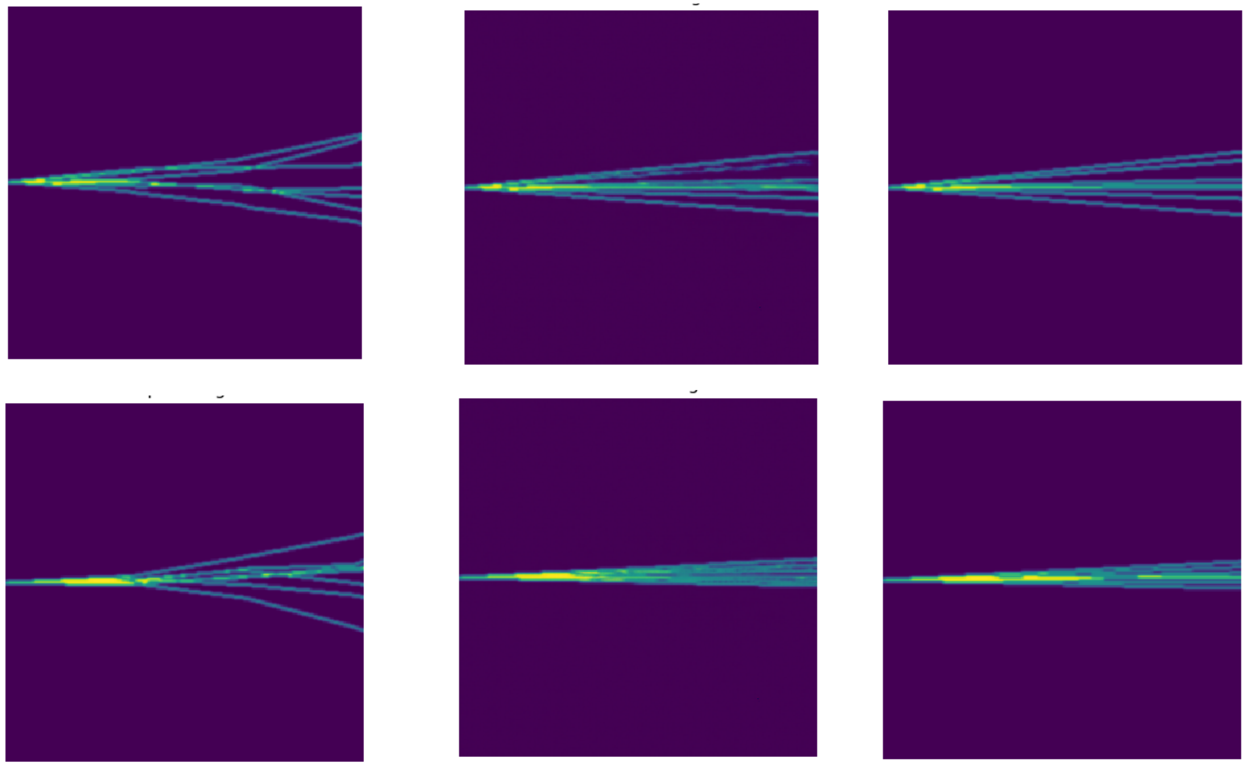


Ledig, Christian et al. "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network." 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016): 105-114.

Recovering track information

S. Maddrell Mander + K. Petridis

Input, after multiple scattering | GAN-reconstructed tracks w/o scattering | “Truth” w/o scattering

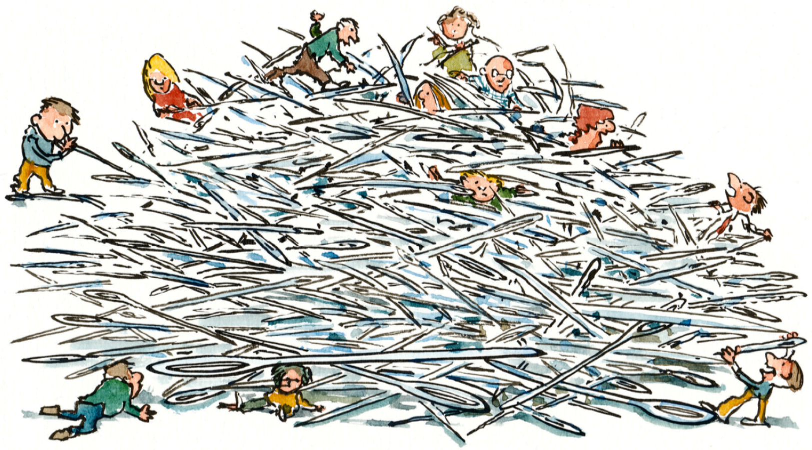


Simplified 2-D model for now, but encouraging initial results - can even deal with overlapping tracks.

Problems we love to have

then

now/future



Fitting

- LHCb has millions, in some cases billions of signal events. E.g. 1M $D^0 \to K^-\pi^+\pi^-\pi^+$ in run 1 alone [Eur.Phys.J. C78 (2018) no.6, 443]. Even “rare” decays like $B^0 \to K^+\pi^-\mu^+\mu^-$ require analysing 1M events (run 2, incl decays via K^*J/ψ).

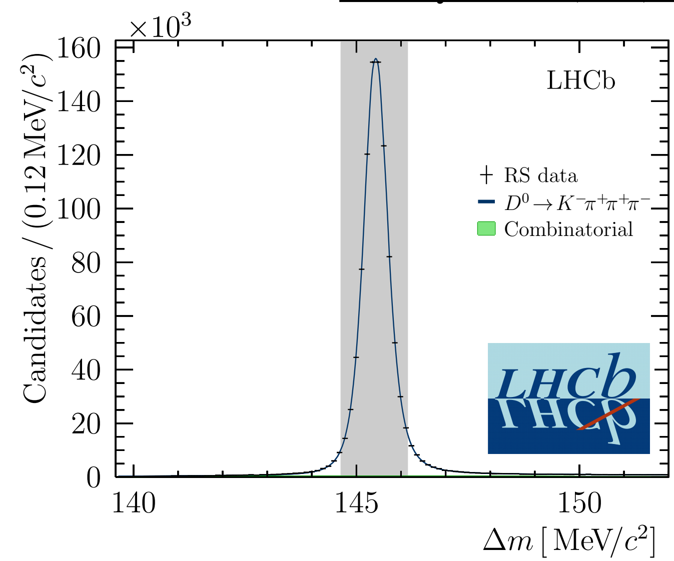
- 5- or more dimensional unbinned fits. Binned fit not an option. MC-based normalisation requires ~10x dataset. Detector resolution requires convolution of PDF.

- Lends itself to parallelisation: 1M events -> 1M processes. Also, detector effects -> FFT

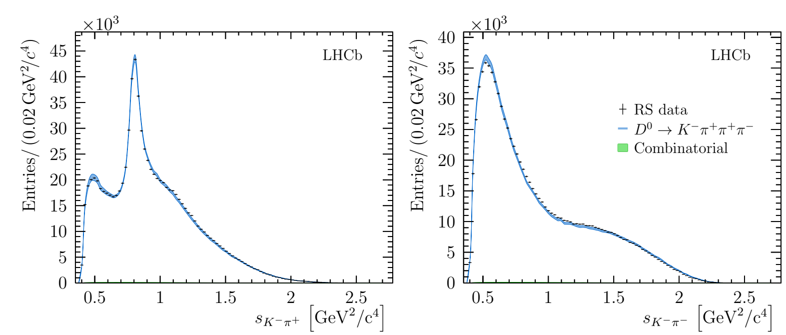
- 5-D, 100 parameter unbinned likelihood fit to 1M $B^0 \to K^+\pi^-\mu^+\mu^-$ events, incl detector resolution [Blake et al EPJC (2018) 78: 453], took ~5hours in a CPU ~5 min in a GPU. 5 min is still rather long.

- See Ben’s talk tomorrow morning for tools such as zfit.

Eur.Phys.J. C78 (2018) no.6, 443

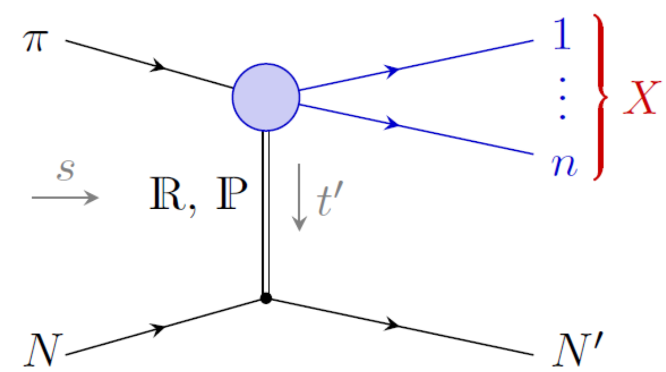


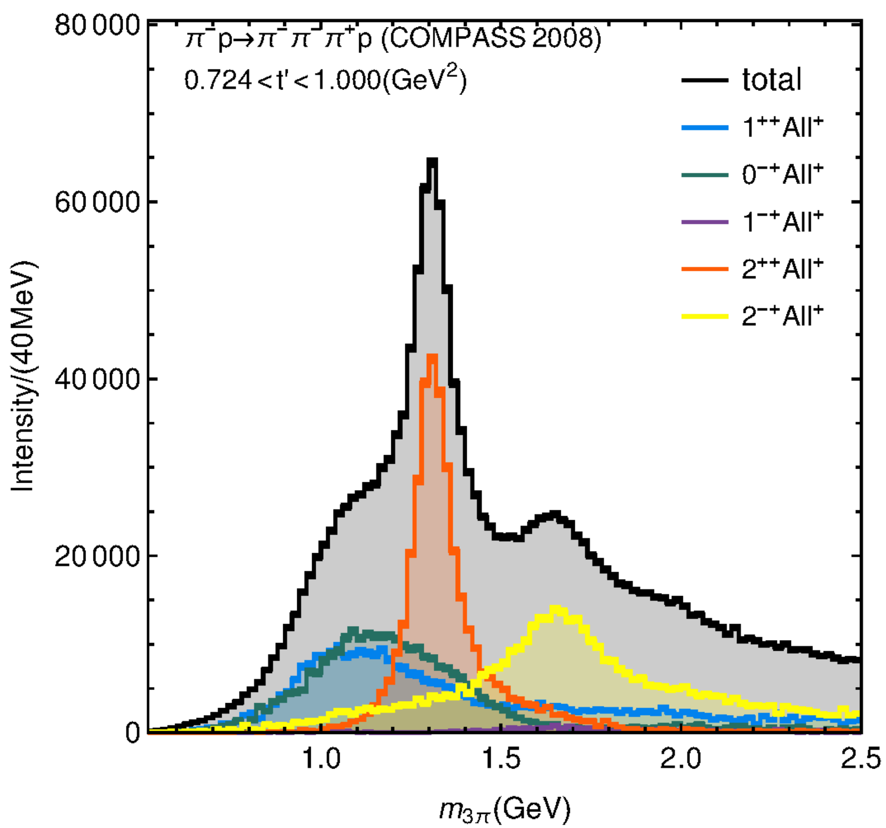
~1M $D^0 \to K^-\pi^+\pi^-\pi^+$ events in Run 1, 5-D amplitude model with 38 fit parameters



Fitting

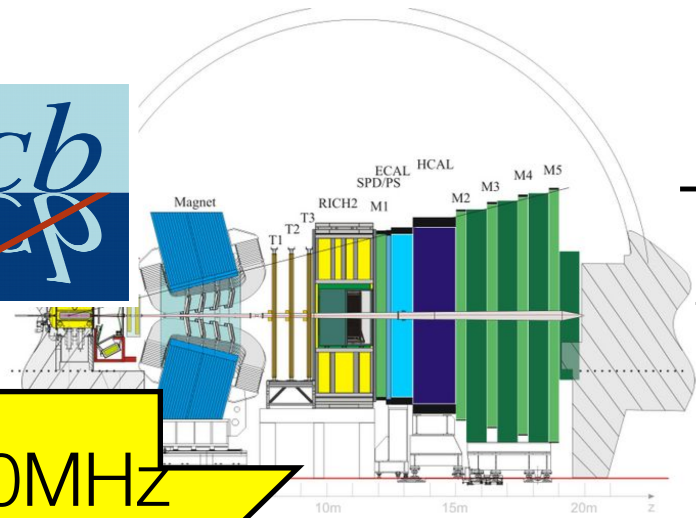
There are other experiments with large signal datasets and complicated analyses. See e.g. COMPASS

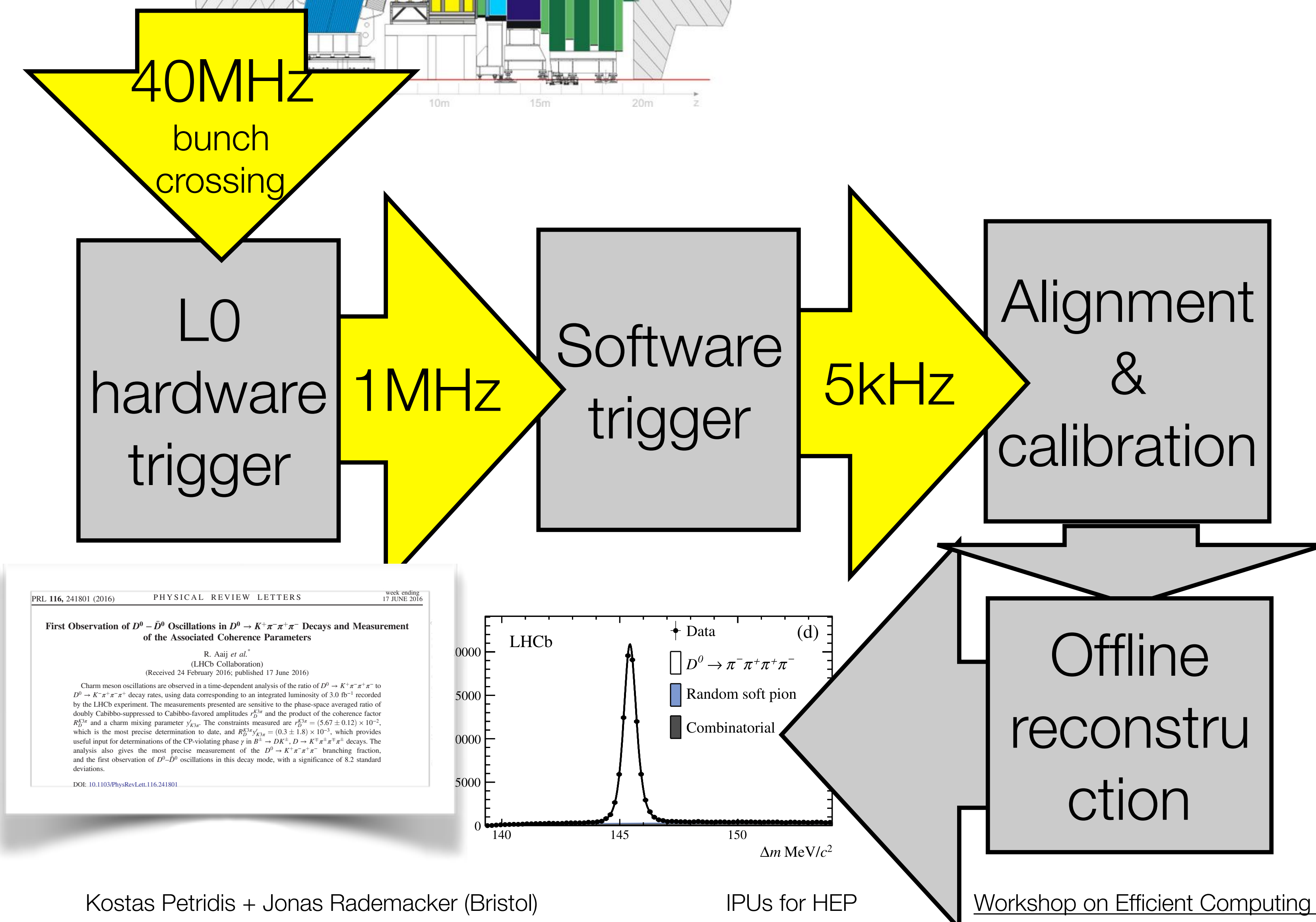




Simultaneous fit of 88 partial waves, each with > 10 parameters in some approaches, in 1100 ($m_{3\pi}$,t') bins (each bin its own Dalitz plot)

Trigger Run 1





PRL **116,** 241801 (2016) PHYSICAL REVIEW LETTERS week ending 17 JUNE 2016

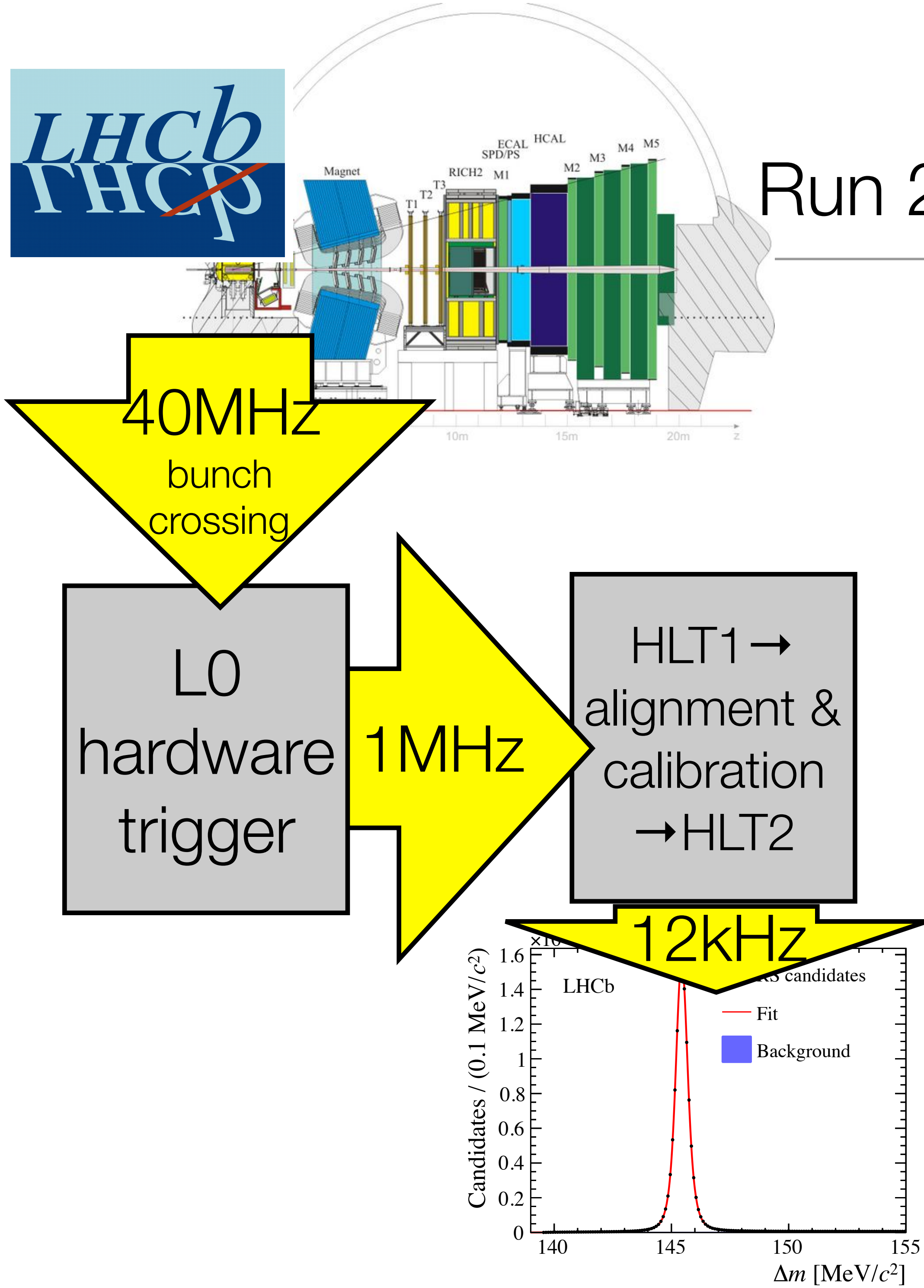
First Observation of $D^0 - \bar{D}^0$ Oscillations in $D^0 \to K^+\pi^-\pi^+\pi^-$ Decays and Measurement of the Associated Coherence Parameters

R. Aaij *et al.**
(LHCb Collaboration)
(Received 24 February 2016; published 17 June 2016)

Charm meson oscillations are observed in a time-dependent analysis of the ratio of $D^0 \to K^+\pi^-\pi^+\pi^-$ to $D^0 \to K^-\pi^+\pi^-\pi^+$ decay rates, using data corresponding to an integrated luminosity of 3.0 fb^{-1} recorded by the LHCb experiment. The measurements presented are sensitive to the phase-space averaged ratio of doubly Cabibbo-suppressed to Cabibbo-favored amplitudes $r_D^{K3\pi}$ and the product of the coherence factor $R_D^{K3\pi}$ and a charm mixing parameter $y'_{K3\pi}$. The constraints measured are $r_D^{K3\pi} = (5.67 \pm 0.12) \times 10^{-2}$, which is the most precise determination to date, and $R_D^{K3\pi} y'_{K3\pi} = (0.3 \pm 1.8) \times 10^{-3}$, which provides useful input for determinations of the CP-violating phase γ in $B^\pm \to DK^\pm$, $D \to K^\mp\pi^\pm\pi^\mp\pi^\pm$ decays. The analysis also gives the most precise measurement of the $D^0 \to K^+\pi^-\pi^+\pi^-$ branching fraction, and the first observation of D^0–$\bar{D}^0$ oscillations in this decay mode, with a significance of 8.2 standard deviations.

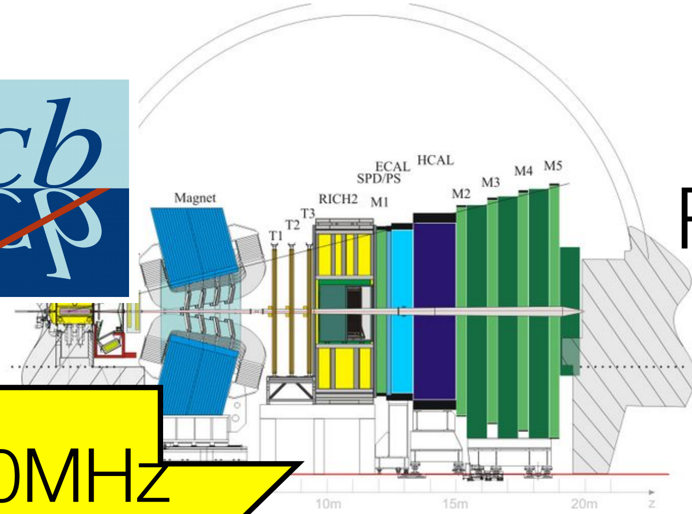
DOI: 10.1103/PhysRevLett.116.241801

Run 2 (2015-2019)



better trigger
& doubled cross section

Run 2 (2015-2019)



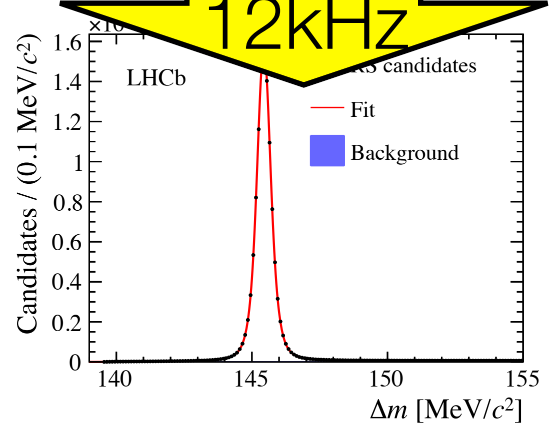
40MHz bunch crossing

L0 hardware trigger

1MHz

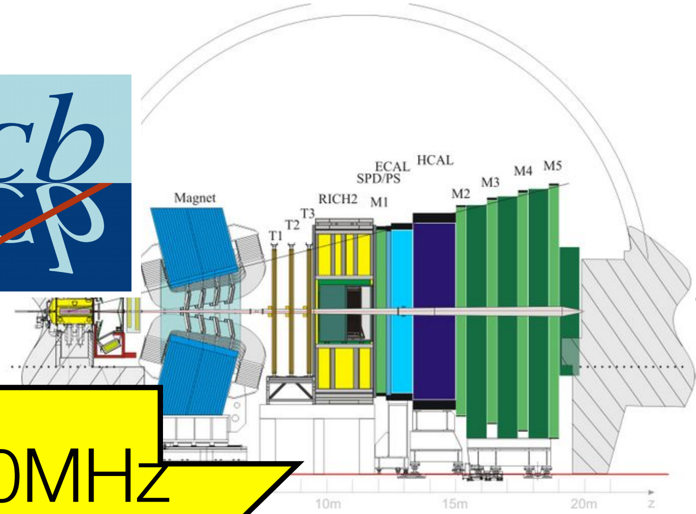
HLT1→ alignment & calibration →HLT2

12kHz



better trigger
& doubled cross section

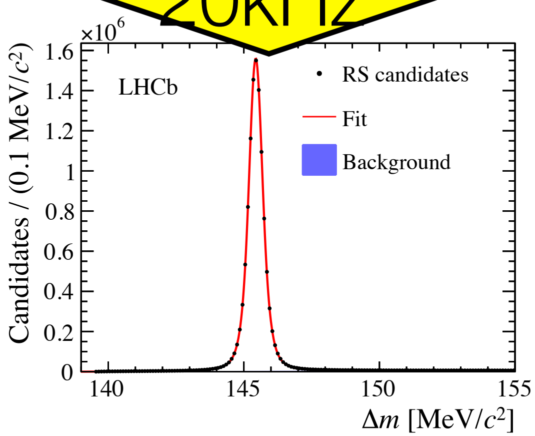
upgrade (data taking in 2021)



40MHz
bunch crossing

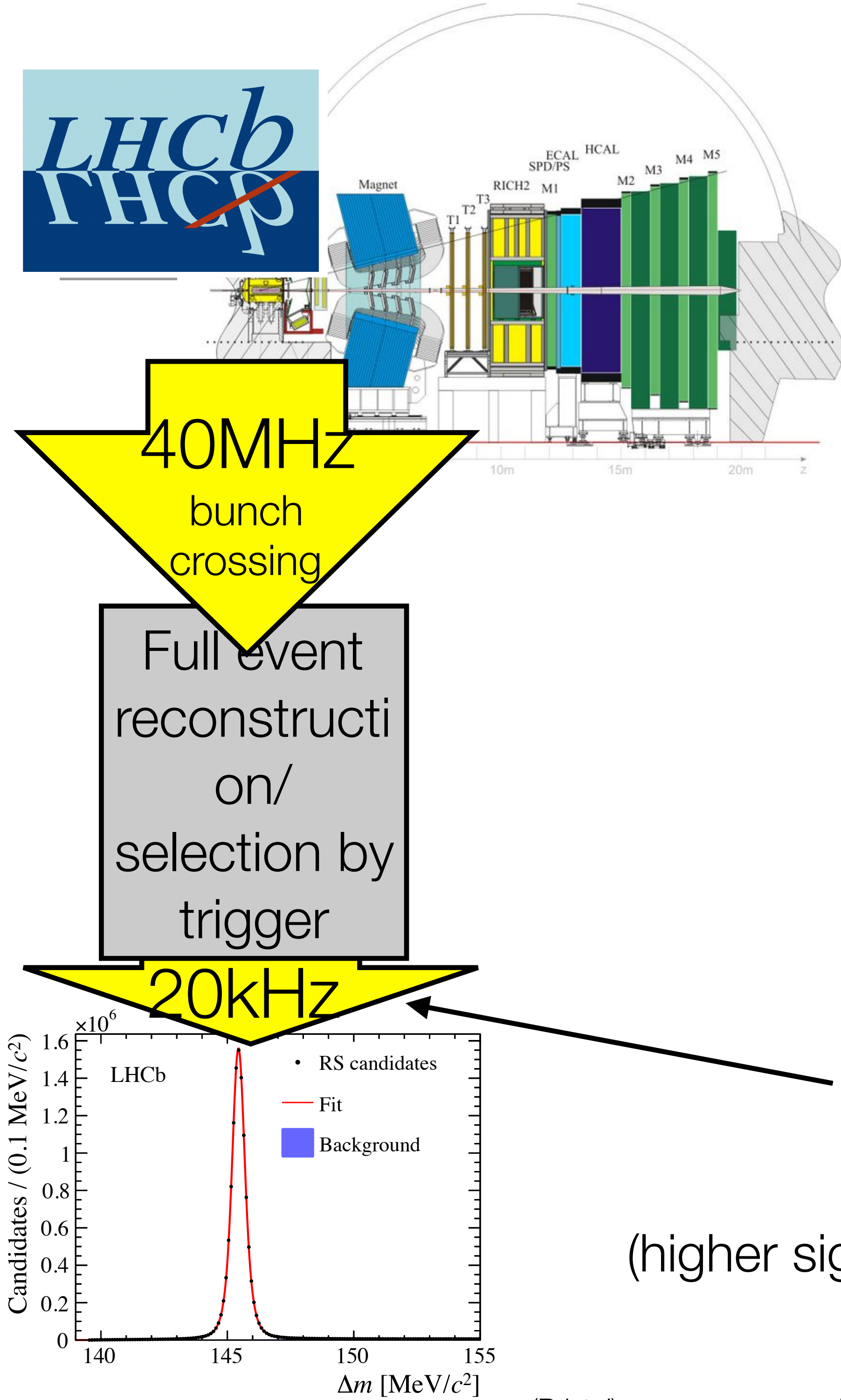
Full event reconstruction/ selection by trigger

20kHz



Full event reconstruction and selection online.

See Conor's talk, tomorrow morning

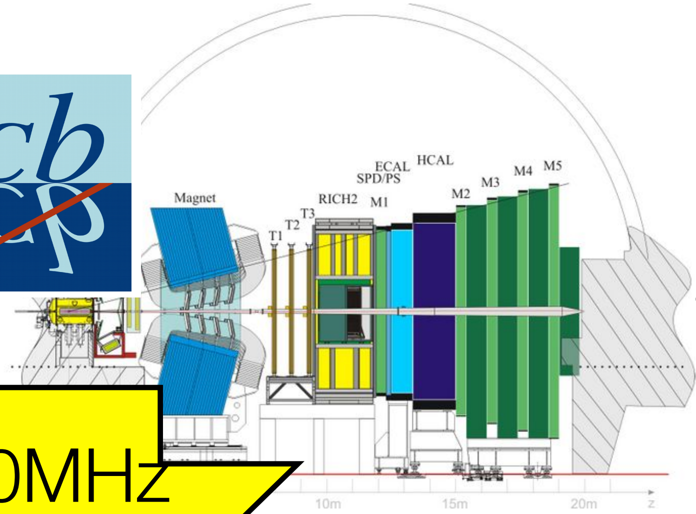


upgrade (data taking in 2021)

Full event reconstruction and selection online.

See Conor's talk, tomorrow morning

Not just *more* kHz.
Also ***better*** kHz
(higher signal fraction / benefit depends on mode)



upgrade (data taking in 2021)

40MHz
bunch crossing

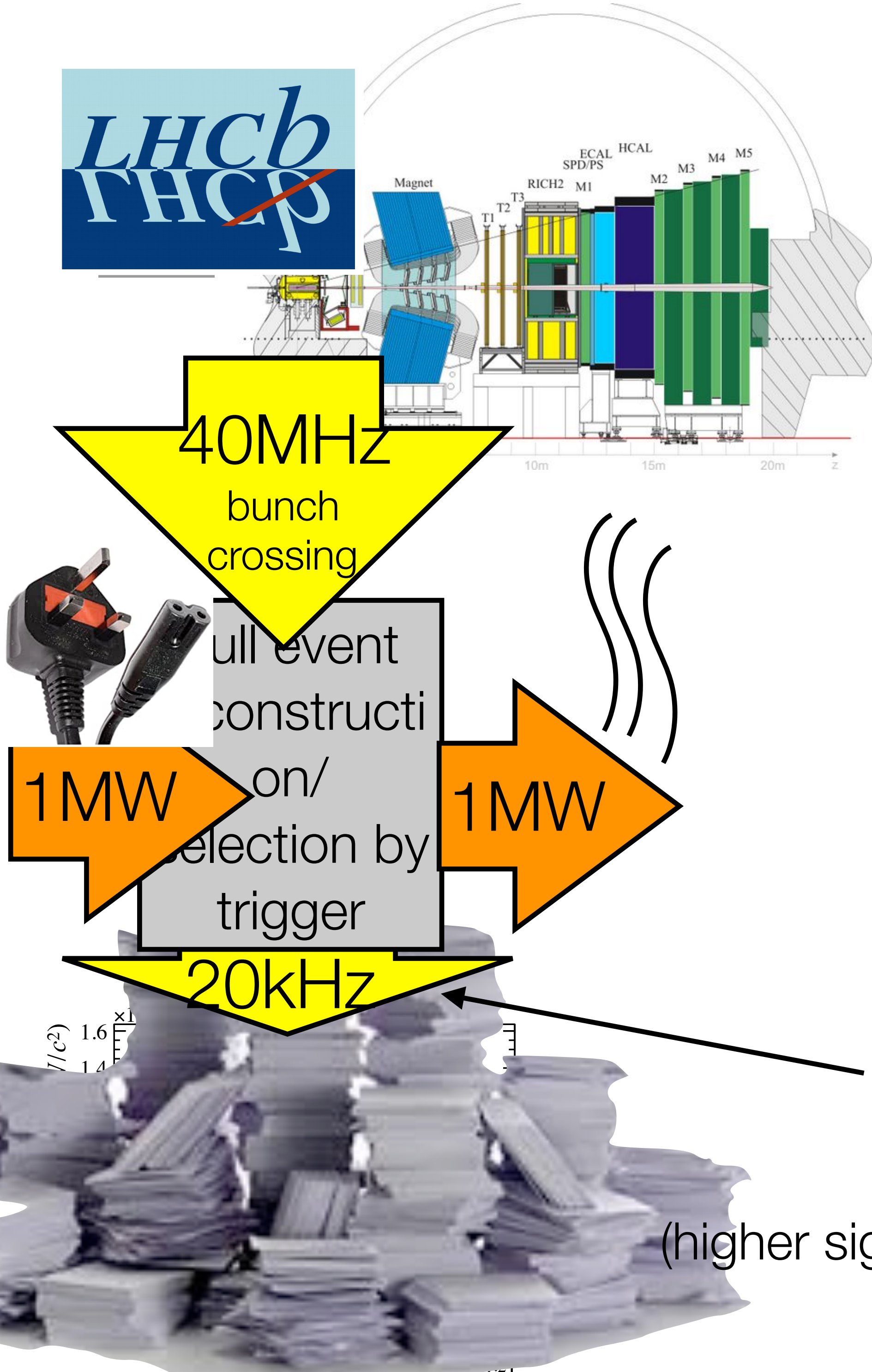
Full event reconstruction/ selection by trigger

20kHz

Full event reconstruction and selection online.

See Conor’s talk, tomorrow morning

Not just *more* kHz.
Also ***better*** kHz
(higher signal fraction / benefit depends on mode)



University of BRISTOL

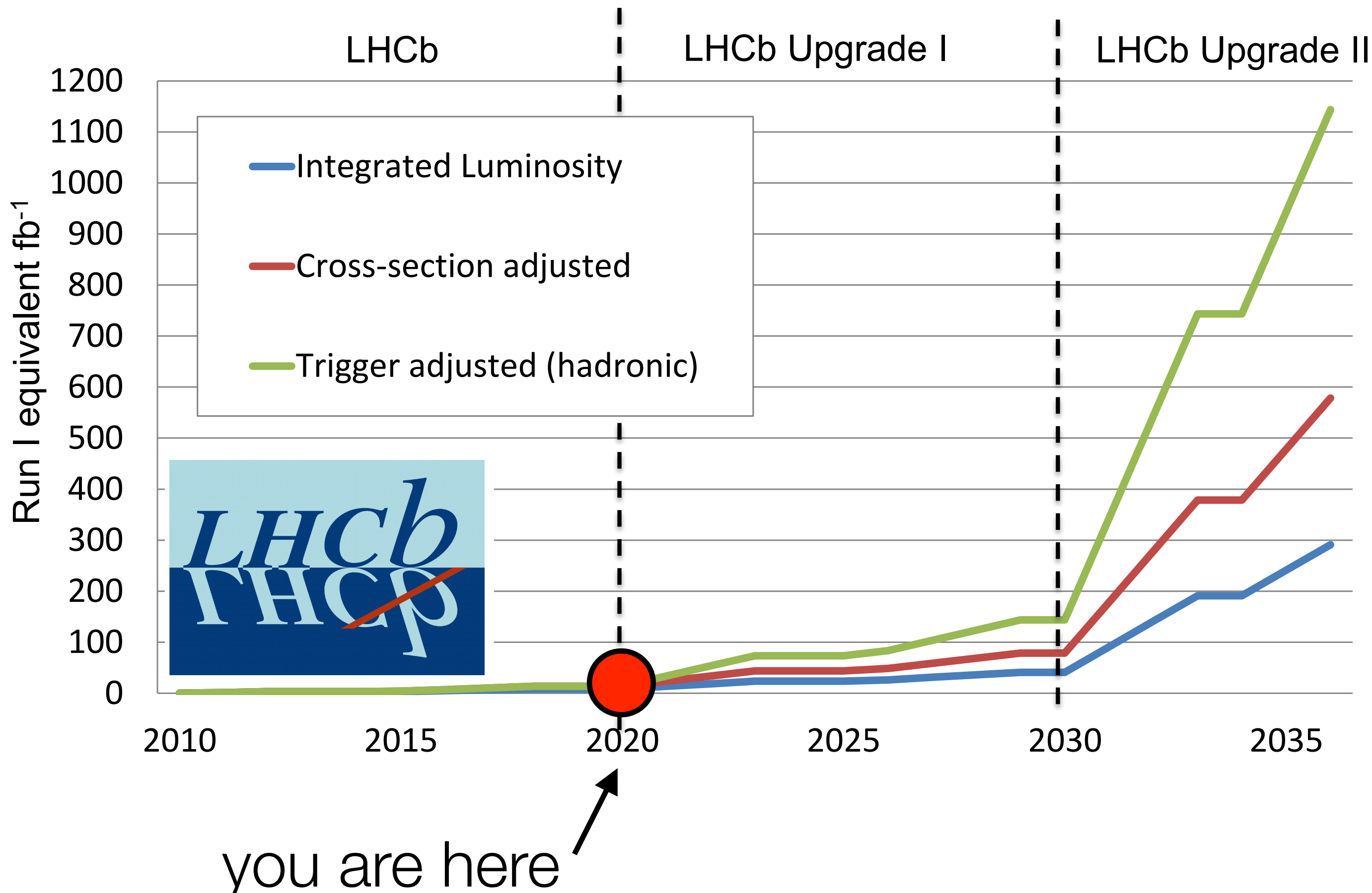
upgrade (data taking in 2021)

Full event reconstruction and selection online.

See Conor's talk, tomorrow morning

Not just *more* kHz.
Also ***better*** kHz
(higher signal fraction / benefit depends on mode)

LHCb upgrade II



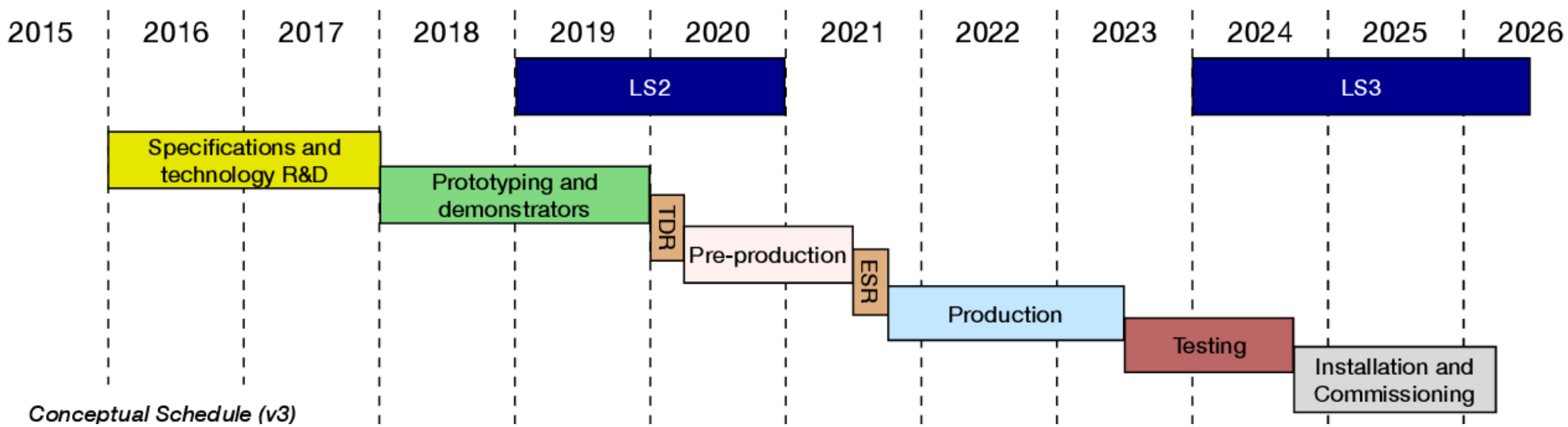
Further upgrades (more data to be processed faster and better) are likely to require new computing architectures for LHCb's trigger. Considerations include CPU, GPU, FPGAs, ASICs.

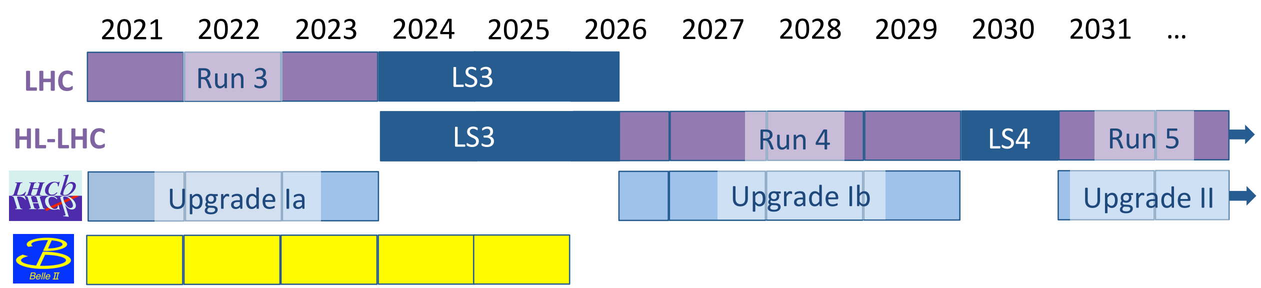
IPUs might occupy a sweet-spot in performance vs flexibility. To be investigated.

Timescales in big HEP projects

CMS L1phase 2 trigger upgrade

CERN-LHCC-2017-013





Current state of project

The Alan Turing Institute

link/support/funding

University of BRISTOL

The Jean Golding* Institute

Research s/w Engineer
(~0.75 FTE for 1 year)

GRAPHCORE

Access to IPUs, POPLAR, expertise

RA-time

DAQ/ Trigger, Computing in HEP expertise

University of BRISTOL

School of Physics

Initial project

Proof or Principle: Use IPUs for

- o fast event generation
- o event reconstruction
- o data analysis (amplitude analysis)
- o trigger / online-reconstruction

* Professor, Jean Golding, mathematician, epidemiologist and founder of the Children of the 90s cohort study.

University of BRISTOL

Current state of project

The Alan Turing Institute

link/support/funding

University of BRISTOL

The Jean Golding* Institute

Research s/w

RA-
time

of
DL
sics

GR

A project to develop and benchmark IPU-based software in HEP that, we hope, can be taken further in the context of ECHEP

POPLAR,
expertise

Initial project

Proof or Principle: Use IPUs for

- o fast event generation
- o event reconstruction
- o data analysis (amplitude analysis)
- o trigger / online-reconstruction

*  Professor, Jean Golding, mathematician, epidemiologist and founder of the Children of the 90s cohort study.

Backup

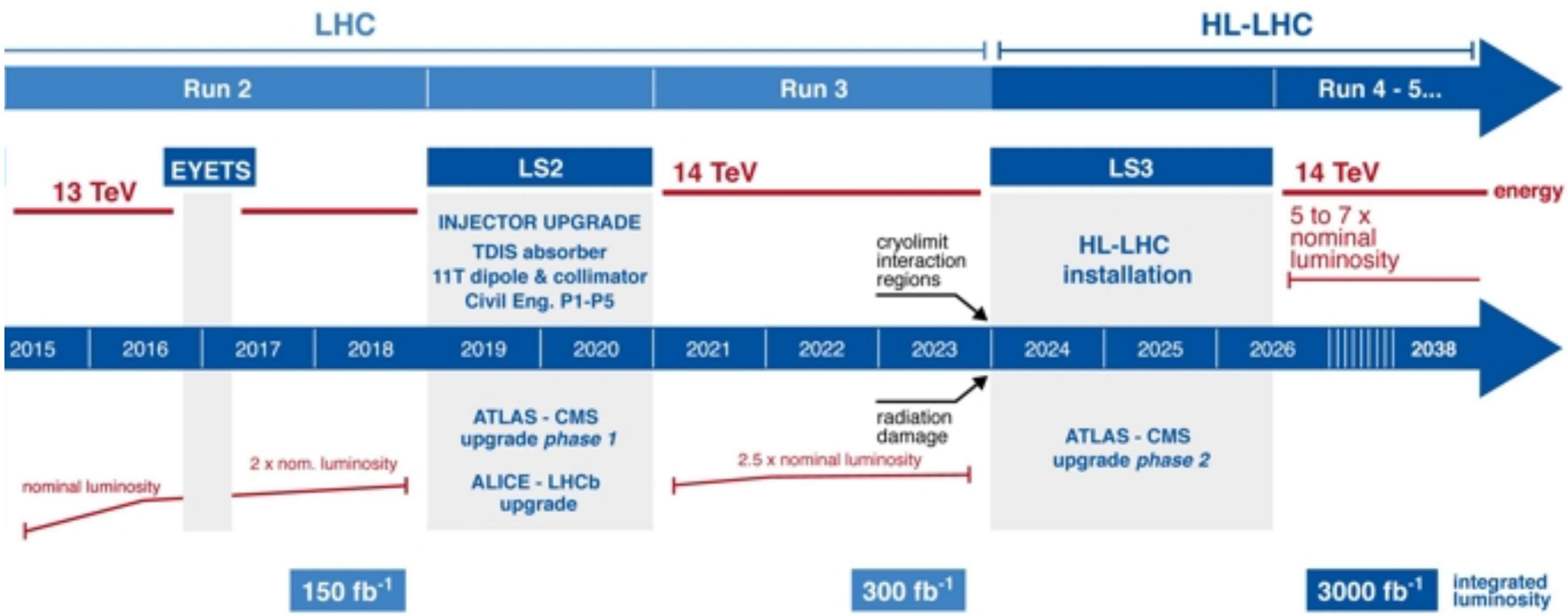
Initial Projects

Fast & Friendly python programming, using tensor-flow interface

Event generator
Culomb-scattering correction
Fitting
and the like

Maximum control and efficiency with POPLAR

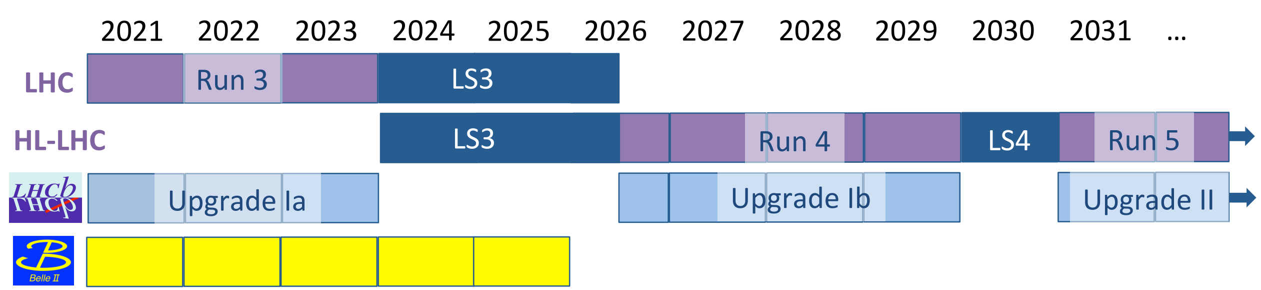
Trigger
Online Computing

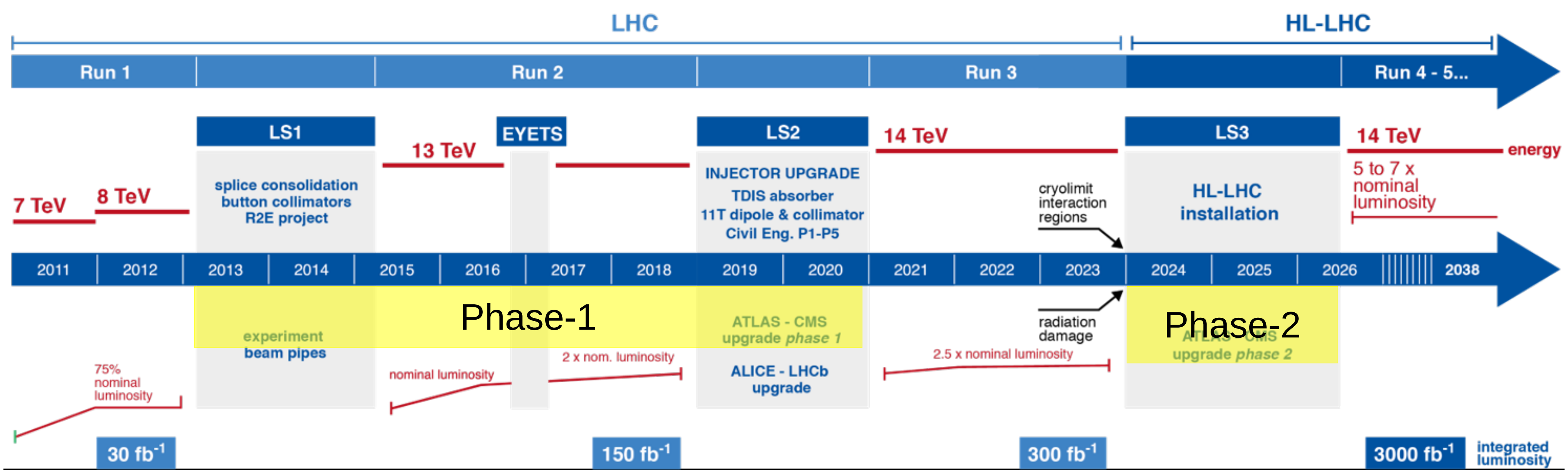


x5 Run1

x2 Run2

x10 Run3




LHC
HL-LHC
Run 1
Run 2
Run 3
Run 4 - 5...
LS1
EYETS
LS2
LS3
7 TeV
8 TeV
13 TeV
14 TeV
14 TeV
energy
splice consolidation
button collimators
R2E project
INJECTOR UPGRADE
TDIS absorber
11T dipole & collimator
Civil Eng. P1-P5
cryolimit
interaction
regions
HL-LHC
installation
5 to 7 x
nominal
luminosity
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2038
Phase-1
experiment
beam pipes
ATLAS - CMS
upgrade phase 1
ALICE - LHCb
upgrade
radiation
damage
Phase-2
upgrade phase 2
75%
nominal
luminosity
nominal luminosity
2 x nom. luminosity
2.5 x nominal luminosity
30 fb-1
150 fb-1
300 fb-1
3000 fb-1
integrated
luminosity

Parameter estimation using multi-dimensional maximum likelihood fits

- Amplitude analyses of beauty and charm decays at LHCb require
fitting 10s to 100s of millions of decays to extract physics parameters

- The large LHCb datasets, coupled with the need to control hadronic nuisance parameters mean that even for rare B-meson decays involve fitting millions of decays in LHCb's Run1+2
  - Decays like $B^0 \to K^+\pi^-\mu^+\mu^-$ involve a 5D unbinned maximum likelihood fit to 1×10^6 decays in order to extract $\mathcal{O}(100)$ parameters [Blake et al EPJC (2018) 78: 453].
  - Calculation of pdf also requires to convolve with detector resolution

Tools like *TensorFlow analysis* allow a fit that takes ~5hours in a CPU to take ~5minutes in a GPU

Enables toy studies and as a result debugging/physics checks in hours rather than days

Larger datasets and analysis complexity means capitalising on intrinsic parallelisation of our fitting methods is paramount

The size of an event therefore depends on the amount of information that is stored, ranging from about a few kB, when only signal candidates are saved, to tens of kB if extra information is added, to roughly 200–250 kB if the full event is persisted in the RDST or RAW formats.