
# FROM COLLISIONS TO PHYSICS
# - And how to get to publication -

Janina Nicolini

# Goal of the lesson

DST vs mDST files

Trigger stages

Prescaling

Online selection

Generator level simulation

Stripping (campaign)

Analysis flow

?

Offline selection

Restripping

Filtering

Dataflow

Run 1 vs Run2 dataflow

Fullstreams vs Streams

Reconstruction level simulation

Analysis preservation

# Goal of the lesson

DST vs mDST files

Trigger stages

Prescaling

Generator level simulation

Online selection

Stripping (campaign)

Will it be an awful lot of information?
Yes, but lots of links on the slides

Analysis flow

Offline selection

Restripping

Dataflow

Filtering

Run 1 vs Run2 dataflow

Analysis preservation

Fullstreams vs Streams

Reconstruction level simulation

# How we perform analyses

Software designed to make common analyses as easy as possible.

So how and what do we do in analyses?

# How we perform analyses

We usually measure:

- Production
- Decay properties

Of heavy flavour hadrons

→ short lifetime → what can we do?

# How we perform analyses

We usually measure:

- Production
- Decay properties

Of heavy flavour hadrons

→ short lifetime → what can we do?

Use "stable" particles

Protons $p/\bar{p}$

Photons $\gamma$

Electrons $e^{\pm}$

Deuterons

Charged pions $\pi^{\pm}$

Charged kaons $K^{\pm}$

Muons $\mu^{\pm}$

# How we perform analyses

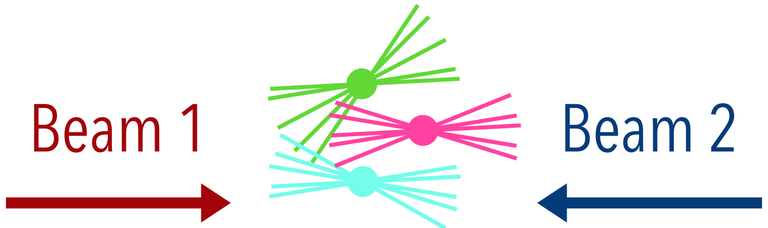
First step reconstruction of properties.

Not single particle, but all charged tracks at the same time

# How we perform analyses

First step reconstruction of properties.

Not single particle, but all charged tracks at the same time



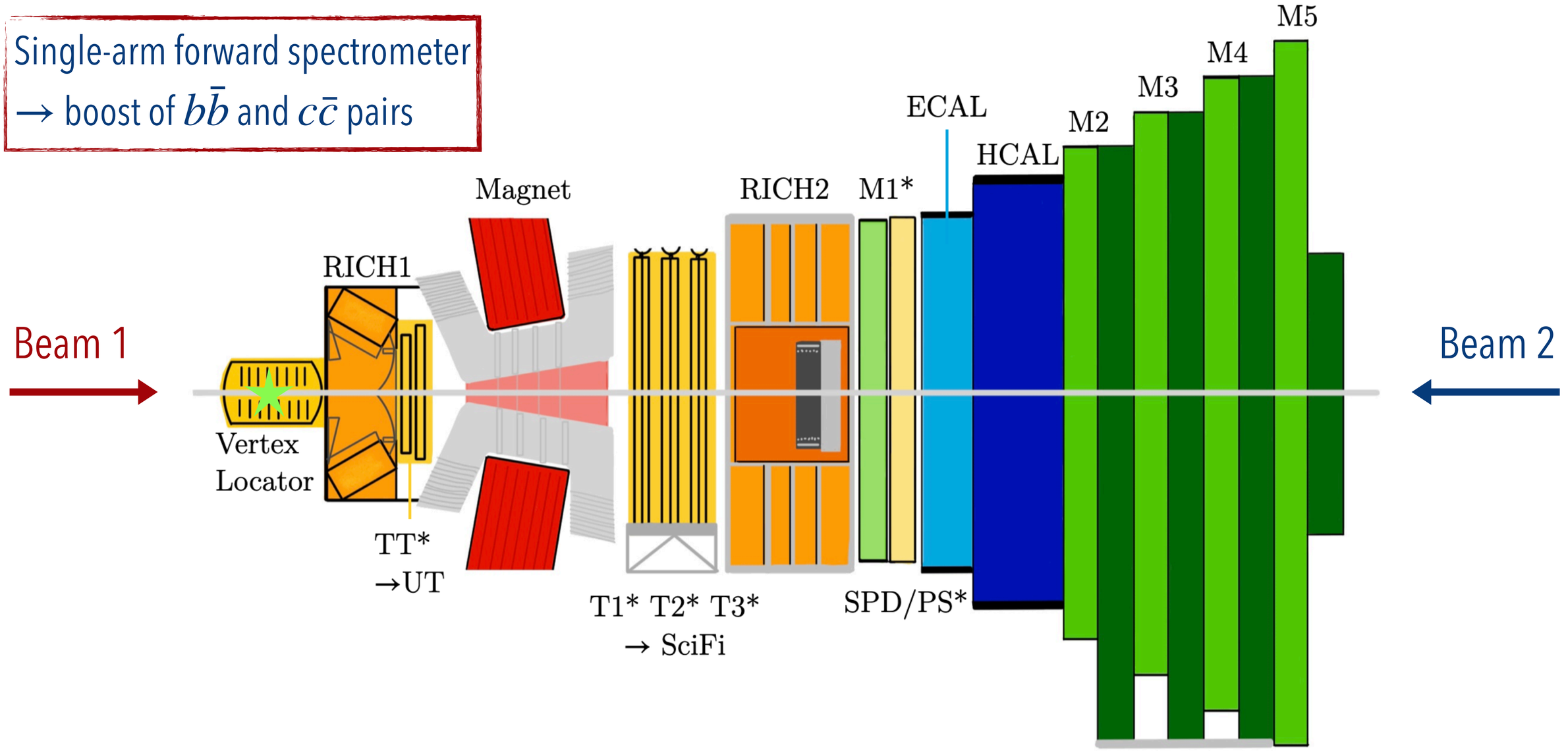

(Several) primary vertices

→ where protons collide

→ large number of tracks intersecting

# Recap: The LHCb detector

Single-arm forward spectrometer
→ boost of $b\bar{b}$ and $c\bar{c}$ pairs

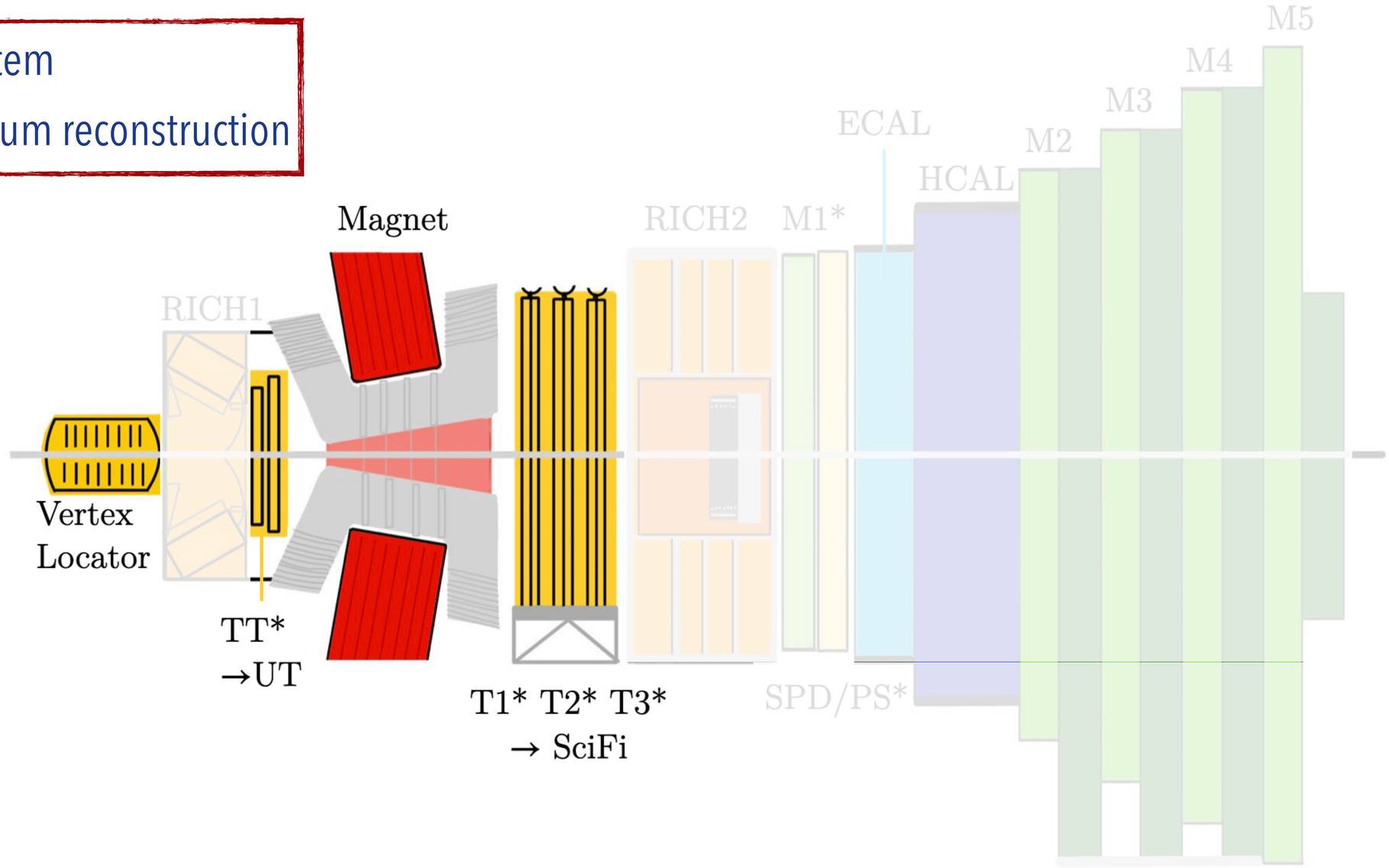



* replaced or removed
during Upgrade for Run 3

# Recap: The LHCb detector

Tracking system

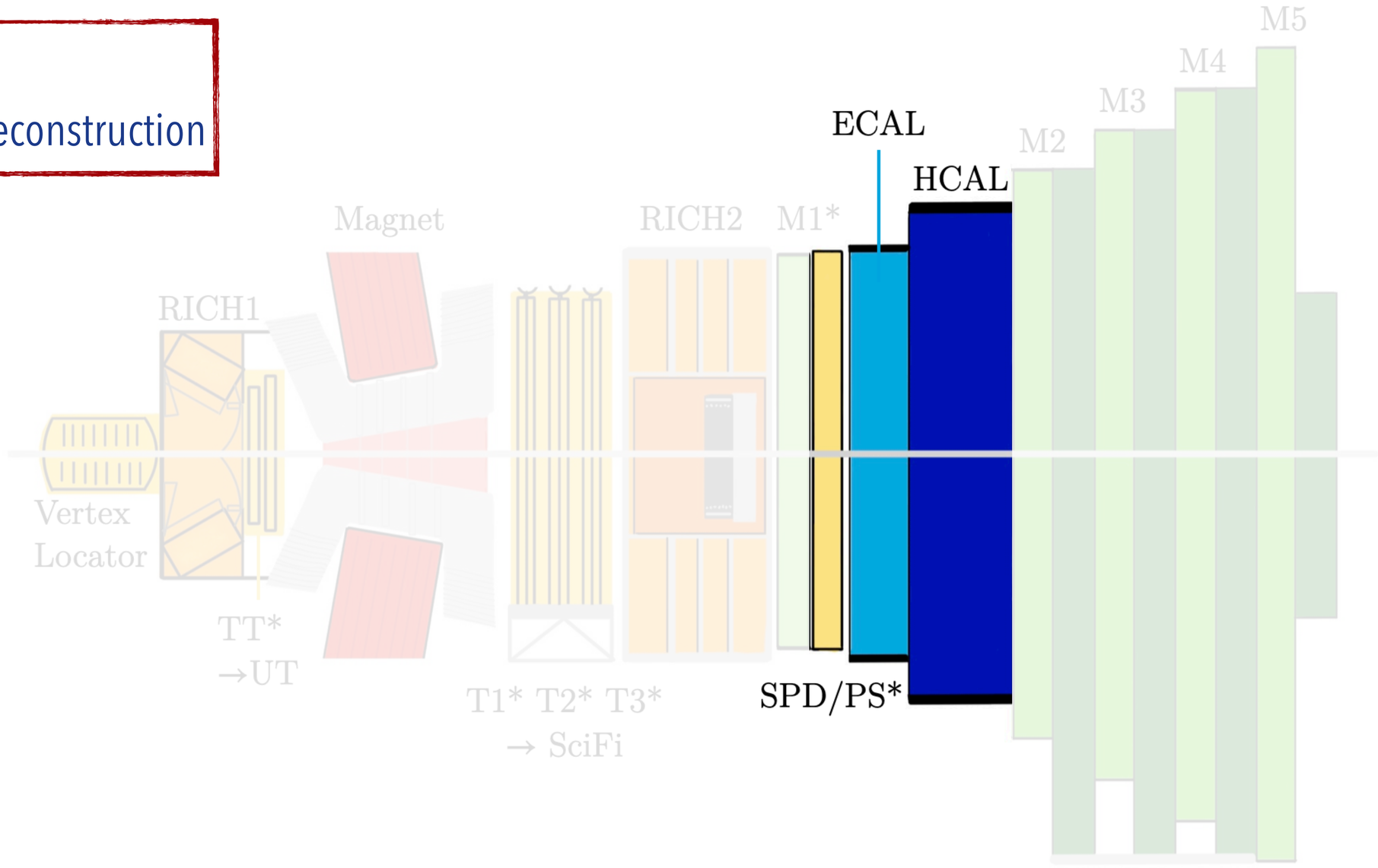
→ Momentum reconstruction



# Recap: The LHCb detector

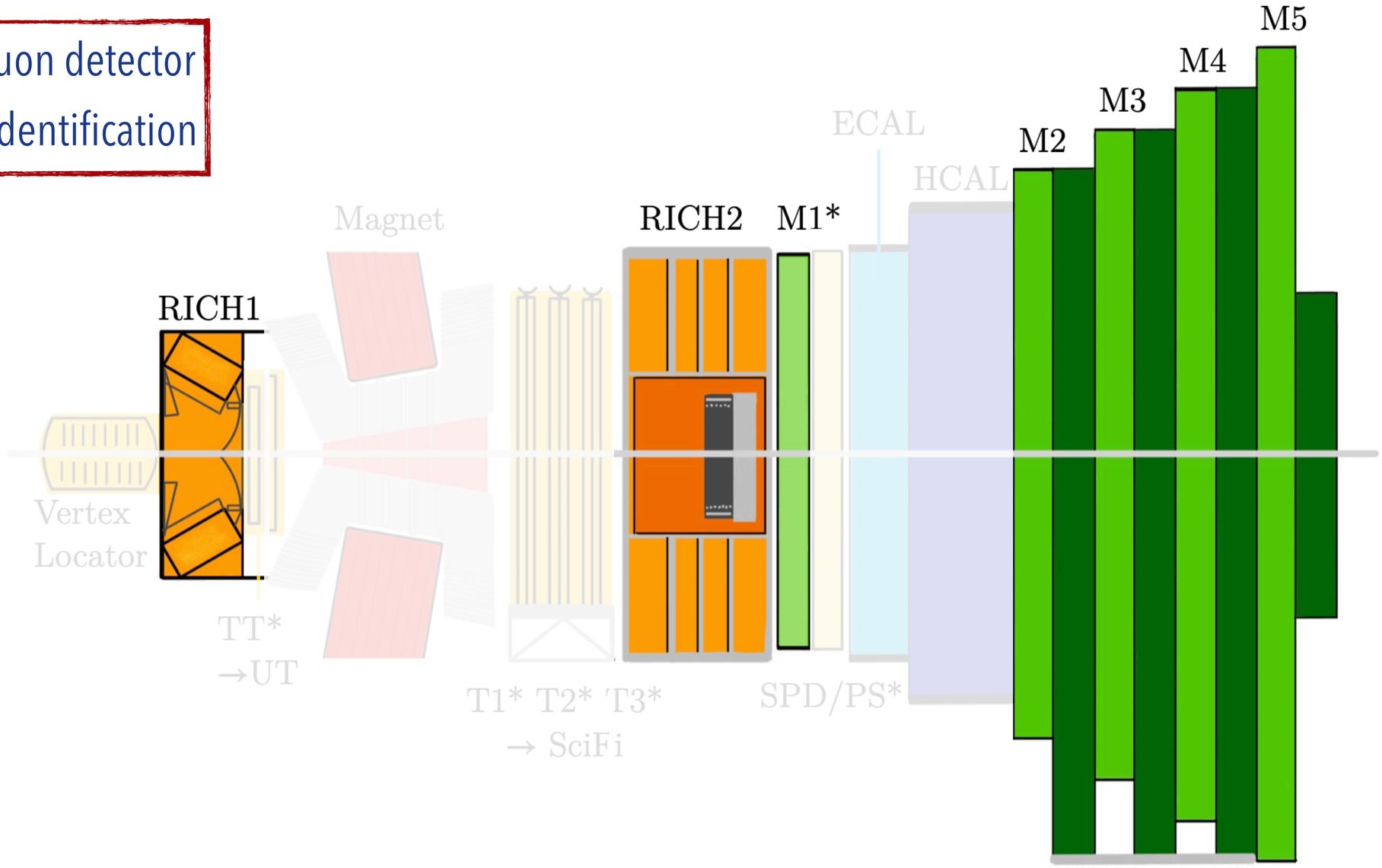
Calorimeter

→ Energy reconstruction



# Recap: The LHCb detector

RICH and muon detector
→ particle identification



# How we perform analyses

## Building decay candidates

Difficulties we encounter

- Contributions from detector effects
- "Ghost" tracks

  → combination of random hits
- Typical hundreds of tracks

→ statistical analysis of events

# How we perform analyses

## Building decay candidates

Reconstructing a $J/\psi \rightarrow \mu^+\mu^-$ decay

- Select suitable tracks created by reconstruction
- Create pairs of oppositely-charged tracks
- Fit each pair under the hypothesis the originate from a common point in space

# How we perform analyses

Building decay candidates

Reconstructing a $J/\psi \rightarrow \mu^+\mu^-$ decay

- Select suitable tracks created by reconstruction
- Create pairs of oppositely-charged tracks
- Fit each pair under the hypothesis the originate from a common point in space

High momentum cut

Use probability of true muon

→called ParticleIDentification variable

# How we perform analyses

Building decay candidates

Reconstructing a $J/\psi \rightarrow \mu^+\mu^-$ decay

- Select suitable tracks created by reconstruction
- Create pairs of oppositely-charged tracks
- Fit each pair under the hypothesis the originate from a common point in space

High momentum cut

Use probability of true muon

→called ParticleIDentification variable

Distance of closest approach between two muons not exceed max value

Invariant mass required to be close to $J/\psi$

# How we perform analyses

Building decay candidates

Reconstructing a $J/\psi \rightarrow \mu^+\mu^-$ decay

- Select suitable tracks created by reconstruction
- Create pairs of oppositely-charged tracks
- Fit each pair under the hypothesis the originate from a common point in space

High momentum cut

Use probability of true muon

→called ParticleIDentification variable

Distance of closest approach between two muons not exceed max value

Invariant mass required to be close to $J/\psi$

DecayTreeFitter tool allows to perform fit in software

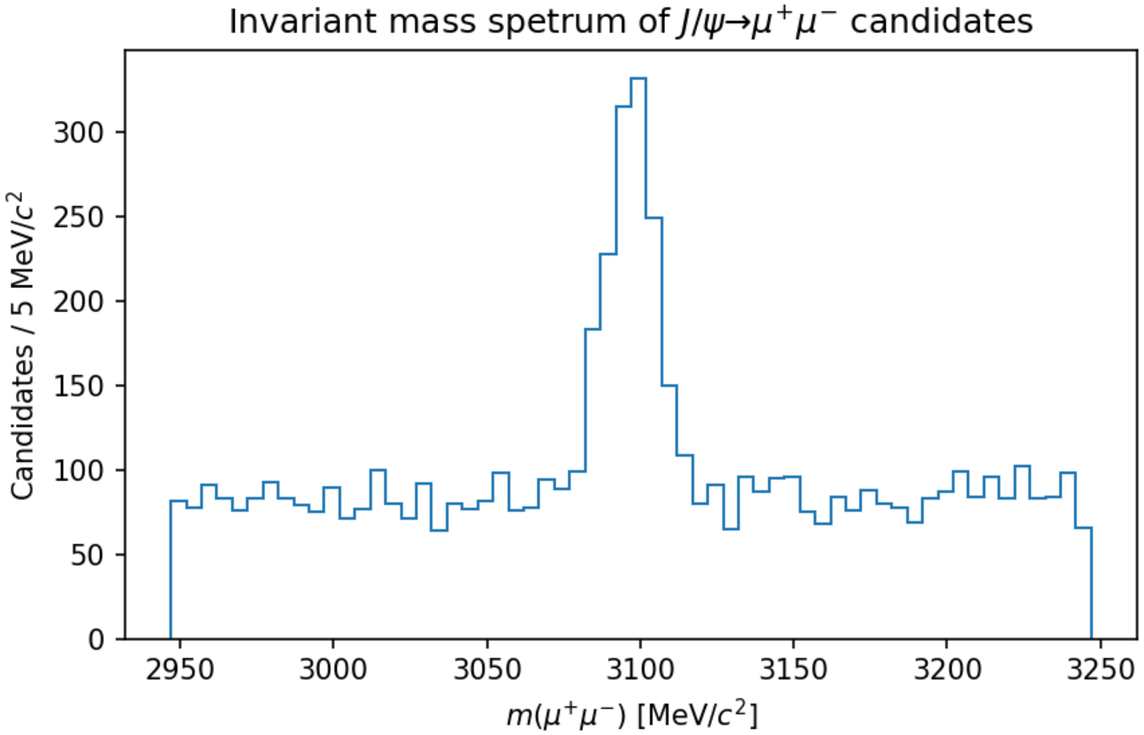
$\rightarrow \chi^2$ of fit can be used in selection

# How we perform analyses

Creating $J/\psi$ candidate from muon four-momenta

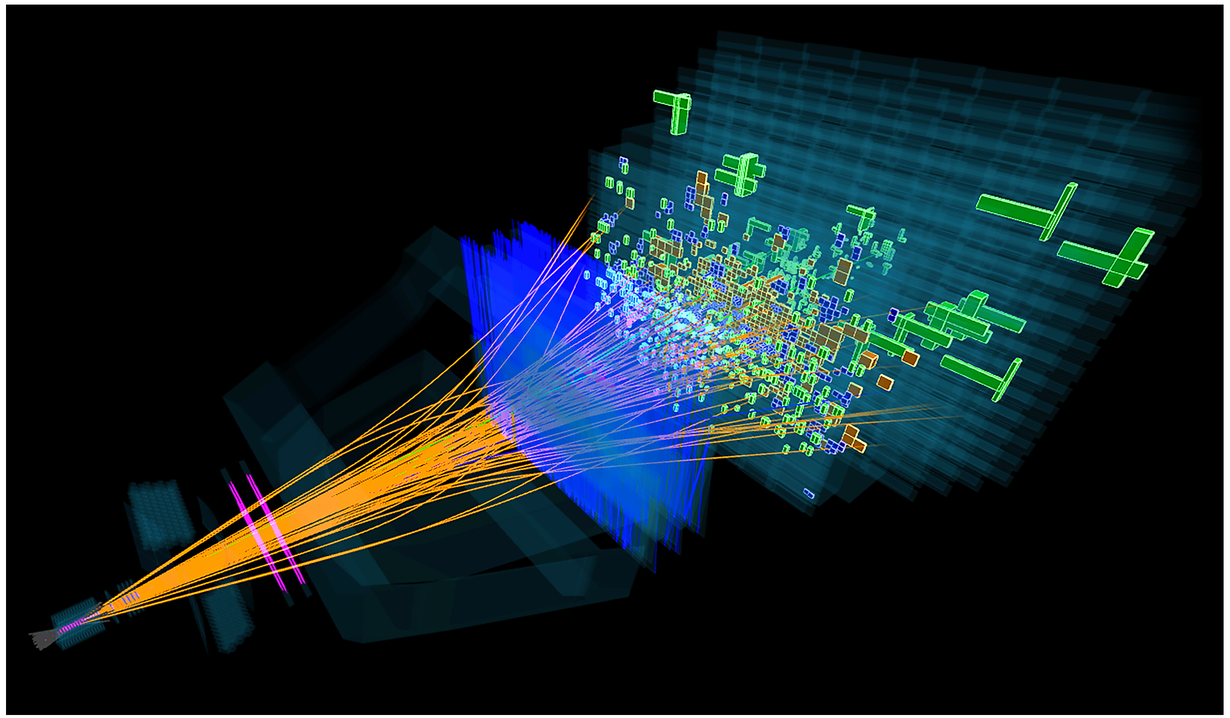
Reconstructing a $J/\psi \rightarrow \mu^+\mu^-$ decay

- Select suitable tracks created by reconstruction
- Create pairs of oppositely-charged tracks
- Fit each pair under the hypothesis the originate from a common point in space



# How do we decide what to save?

The LHC can provide a bunch crossing every 25ns.



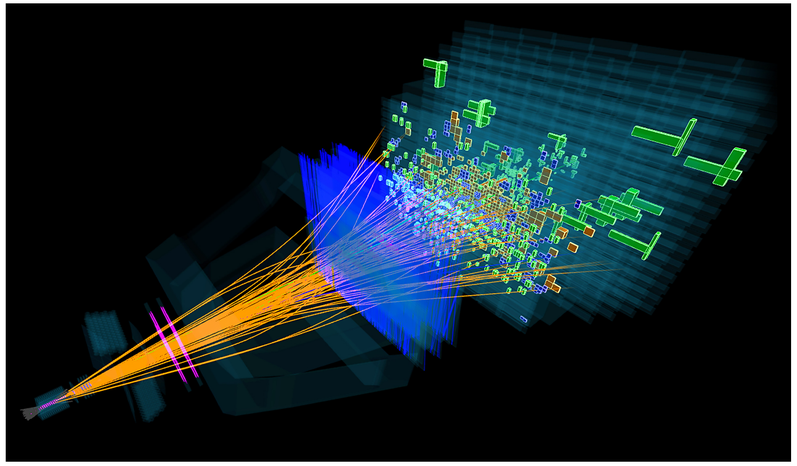
Data rate is 40MHz, but :

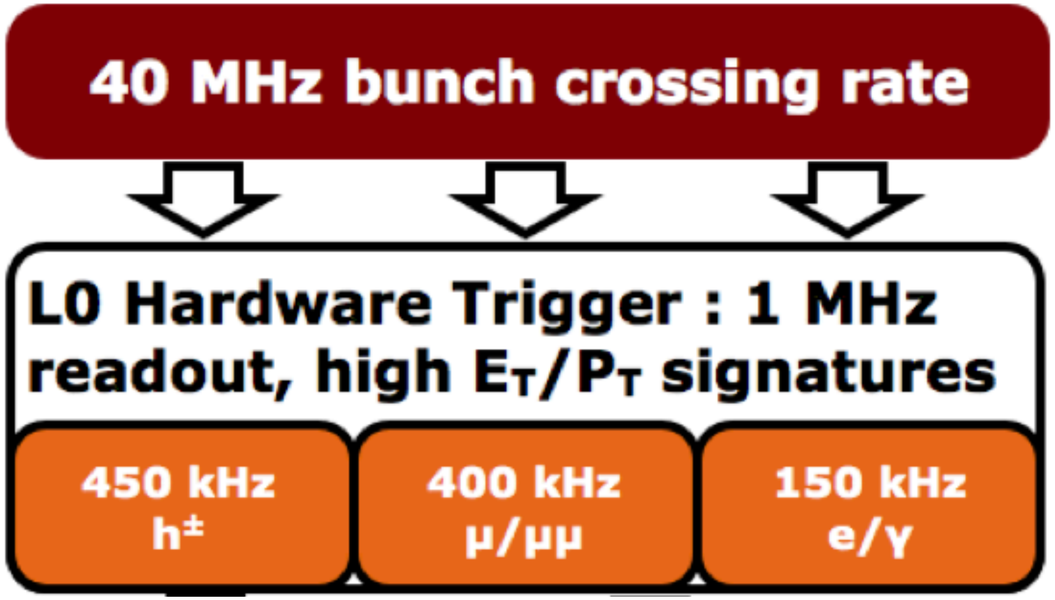
- FPGA readout max. 1MHz
- Run 1: rate to storage max. 5kHz
- Run 2: rate to storage max. 12.5kHz

So we cannot save 1TB/s.

# How do we decide what to save?

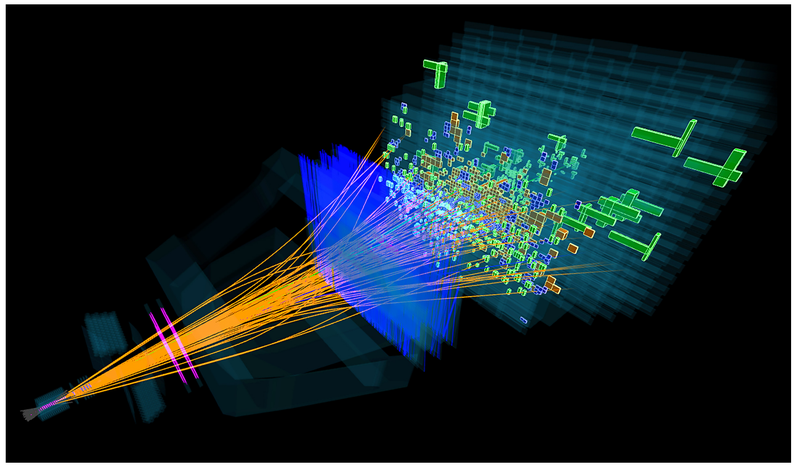
First a hardware trigger stage called L0.

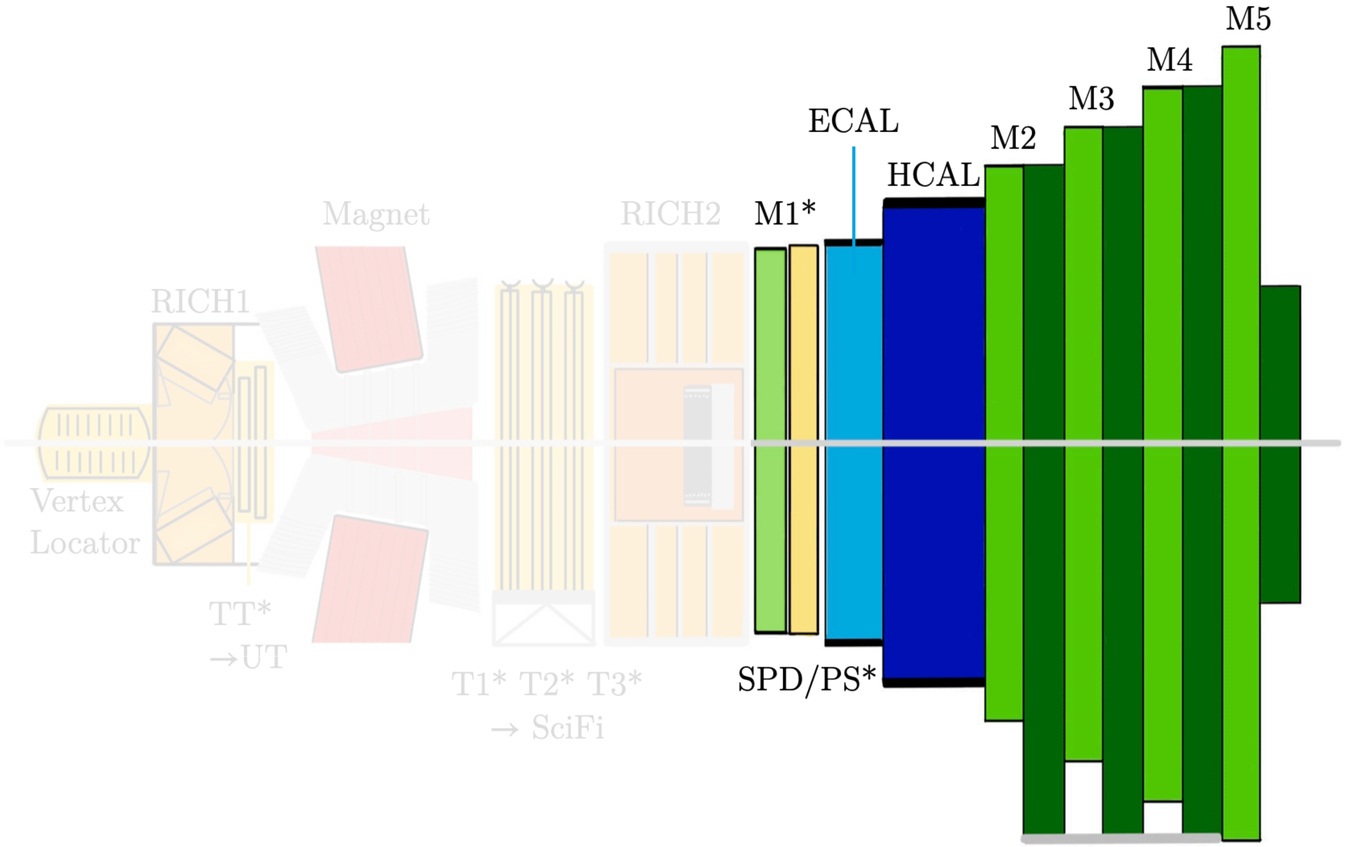




# How do we decide what to save?

First a hardware trigger stage called L0.






**40 MHz bunch crossing rate**

**L0 Hardware Trigger : 1 MHz readout, high $E_T/P_T$ signatures**

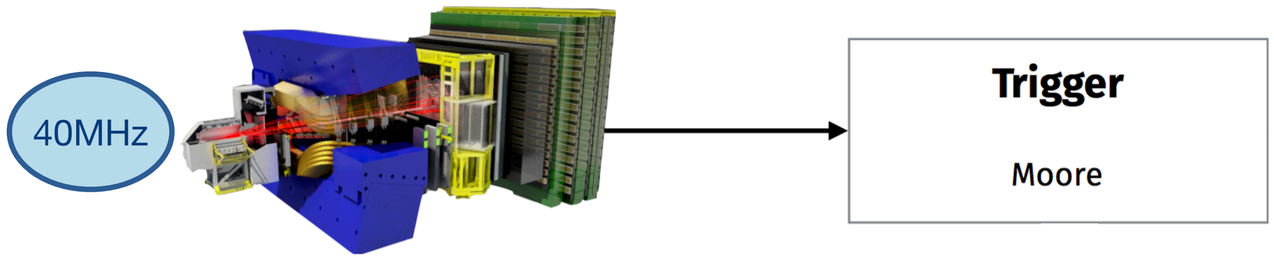
| 450 kHz | 400 kHz | 150 kHz |
|---|---|---|
| $h^{\pm}$ | $\mu/\mu\mu$ | e/γ |

Mainly two detector types firing:

- Hits in the muon stations
- Energy deposit in the ECAL and HCAL

# Collision dataflow during Run 1

Second step of the online reconstruction.

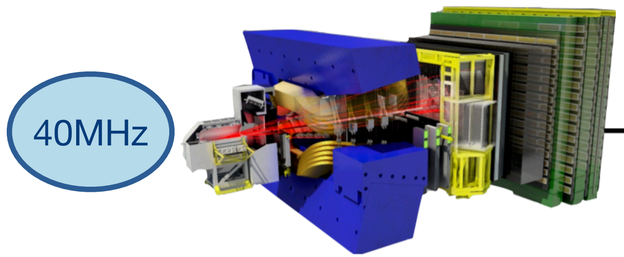



Next software stage called High Level Trigger

- HLT1: Adding tracking information
- HLT2: Adding RICH information
- Both run in Moore framework

# Collision dataflow during Run 1

Online reconstruction done, but..

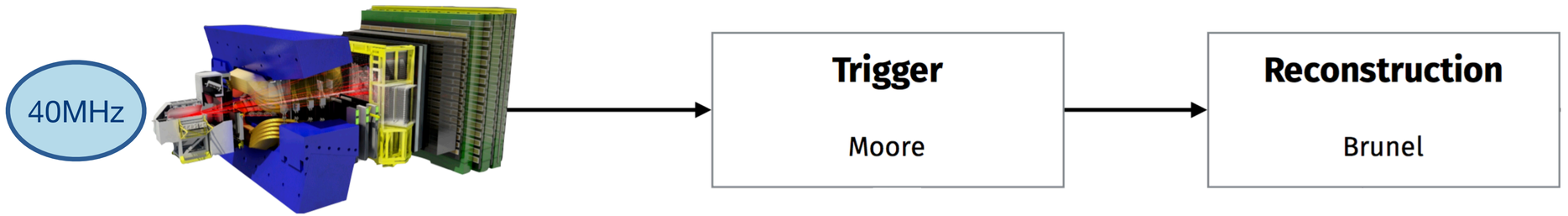


40MHz → **Trigger** (Moore) → 5kHz

Fast trigger lead to worse resolution.
What do we do?

# Collision dataflow during Run 1

## The offline reconstruction

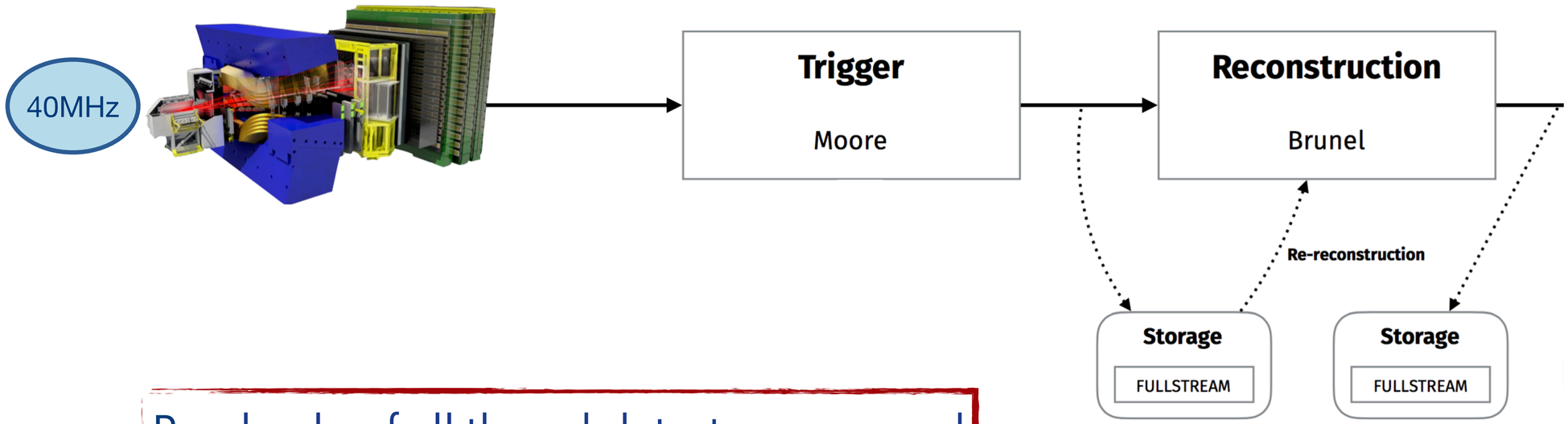



Improve reconstruction of:

- Tracks,
- Clusters,…
- Run in Brunel framework

# Collision dataflow during Run 1

And storage of the data to tape



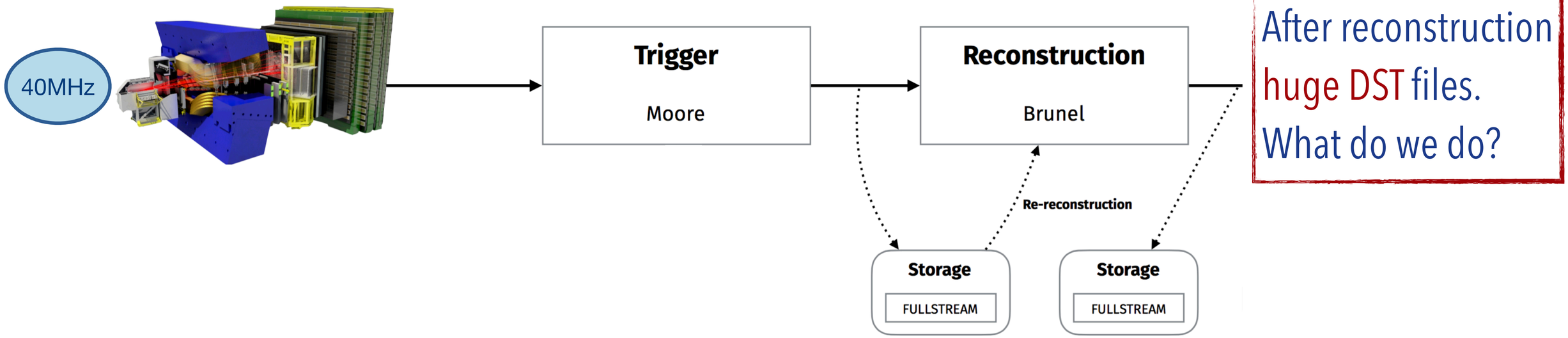

Raw banks of all the subdetectors are saved in FULLSTREAM

→ no further selection

→ Information stored in DST Files

# Collision dataflow during Run 1

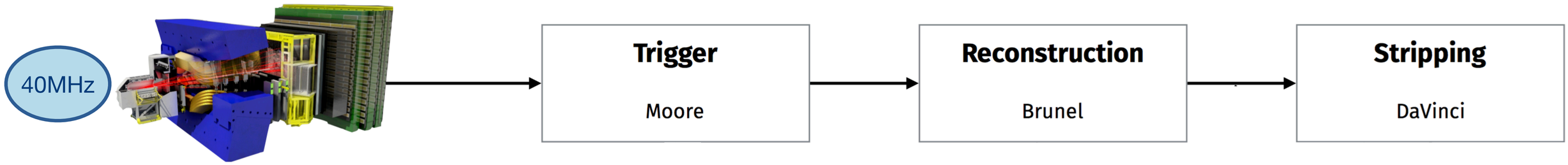
And storage of the data to tape




After reconstruction huge DST files.
What do we do?

# Collision dataflow during Run 1

## Reducing the file sizes

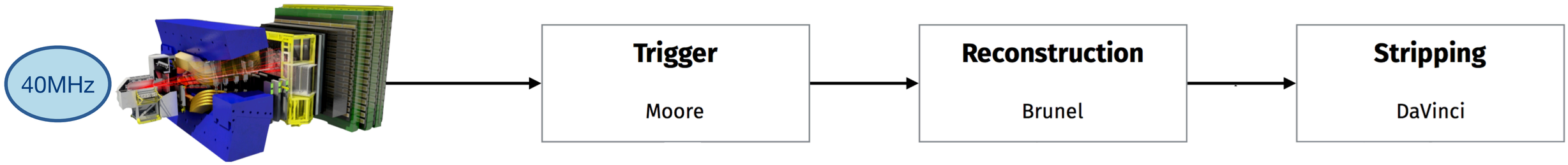



Selecting further with stripping lines

- Select certain decays based on signatures
- Exclusive: only one decay
- Inclusive: several decays combined
- Special lines: minimum-bias, BKG studies, …

# Collision dataflow during Run 1

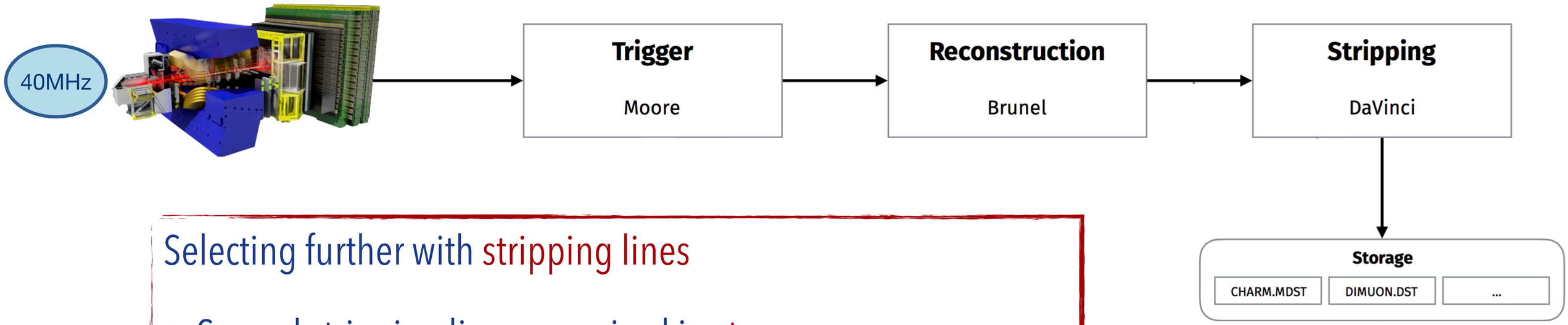
## Reducing the file sizes




### Special stripping lines

- Minimum-bias, BKG study
- Have very little selection → high rates
- Prescaling: save randomly only 0.1 [0.01] of the events
- Never use for other lines! Signal events are also lost!

# Collision dataflow during Run 1

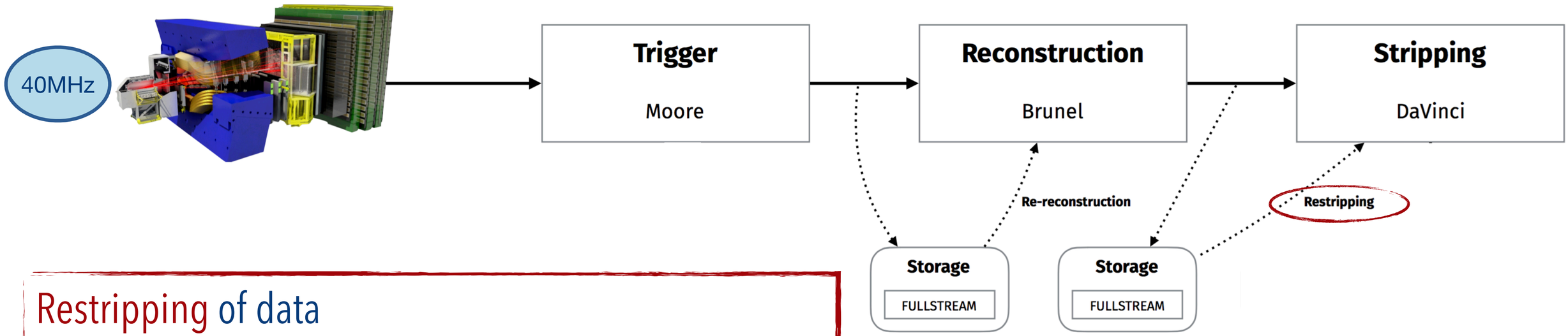



Selecting further with stripping lines

- Several stripping lines organised in streams
- DST or mDST files (150 vs 50 kB/event)
- mDST only store tracks that passed selection, not whole event
- Run in DaVinci framework

# Collision data Flow during Run1

What about new ideas?

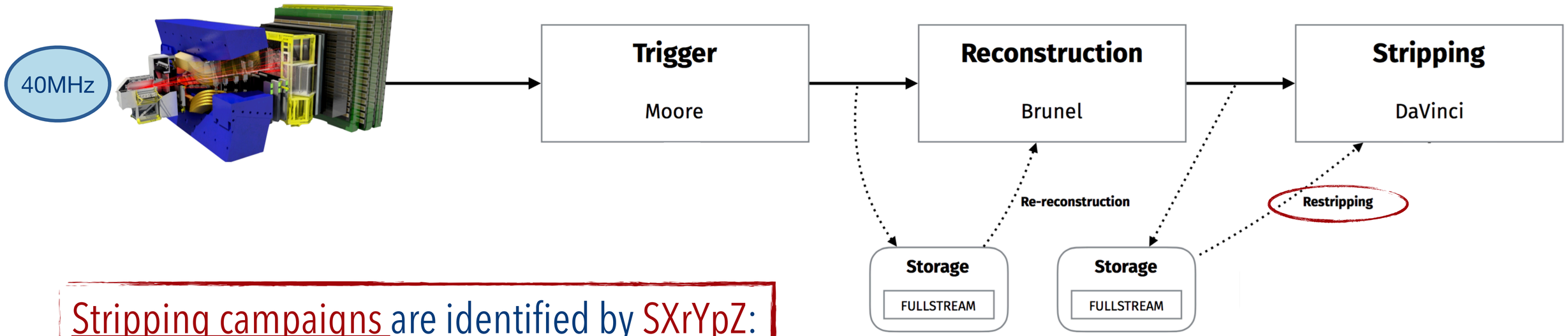



Restripping of data

- Allows to access new ideas with new lines
- Incremental restripping: only new/updated lines
- Full restripping: only done if bugs fixed e.g. in reco

# Collision dataflow during Run 1

What about new ideas?

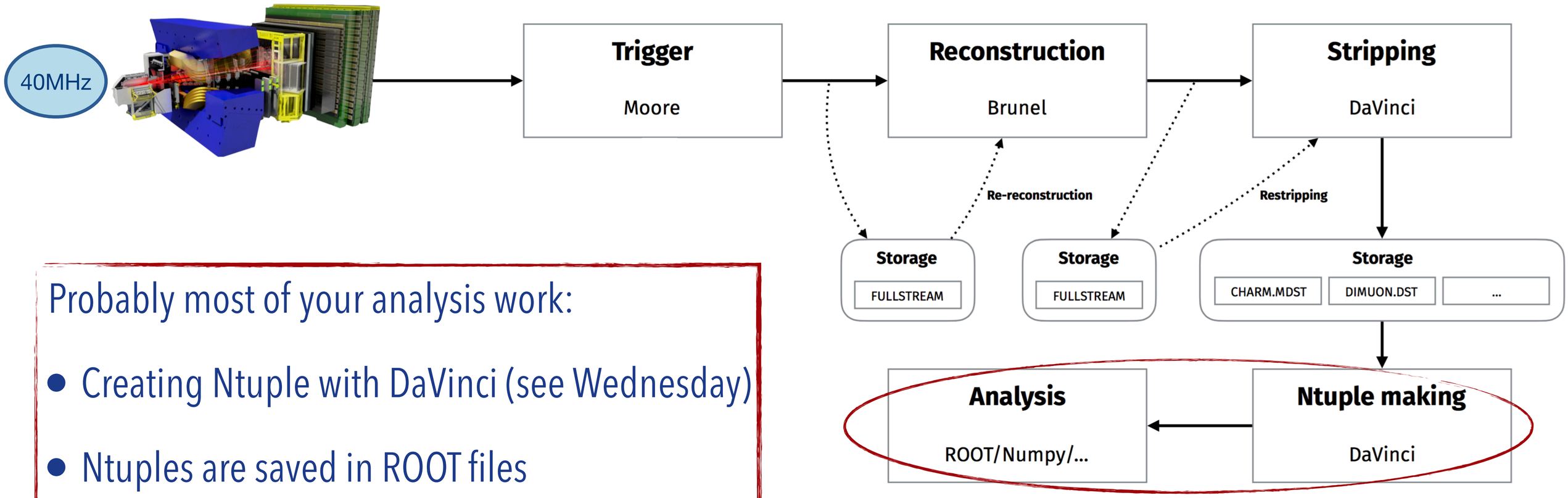



Stripping campaigns are identified by SXrYpZ:

- X: Full restripping campaign
- Y: data taking Year
- Z: incremental restripping

# Collision dataflow during Run 1

And finally:

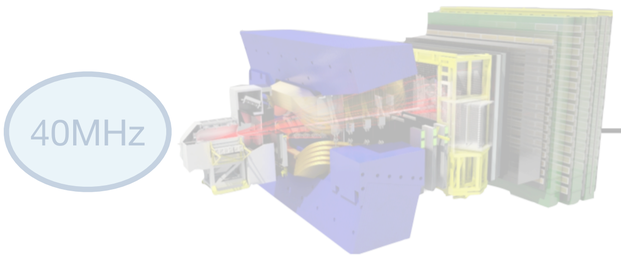


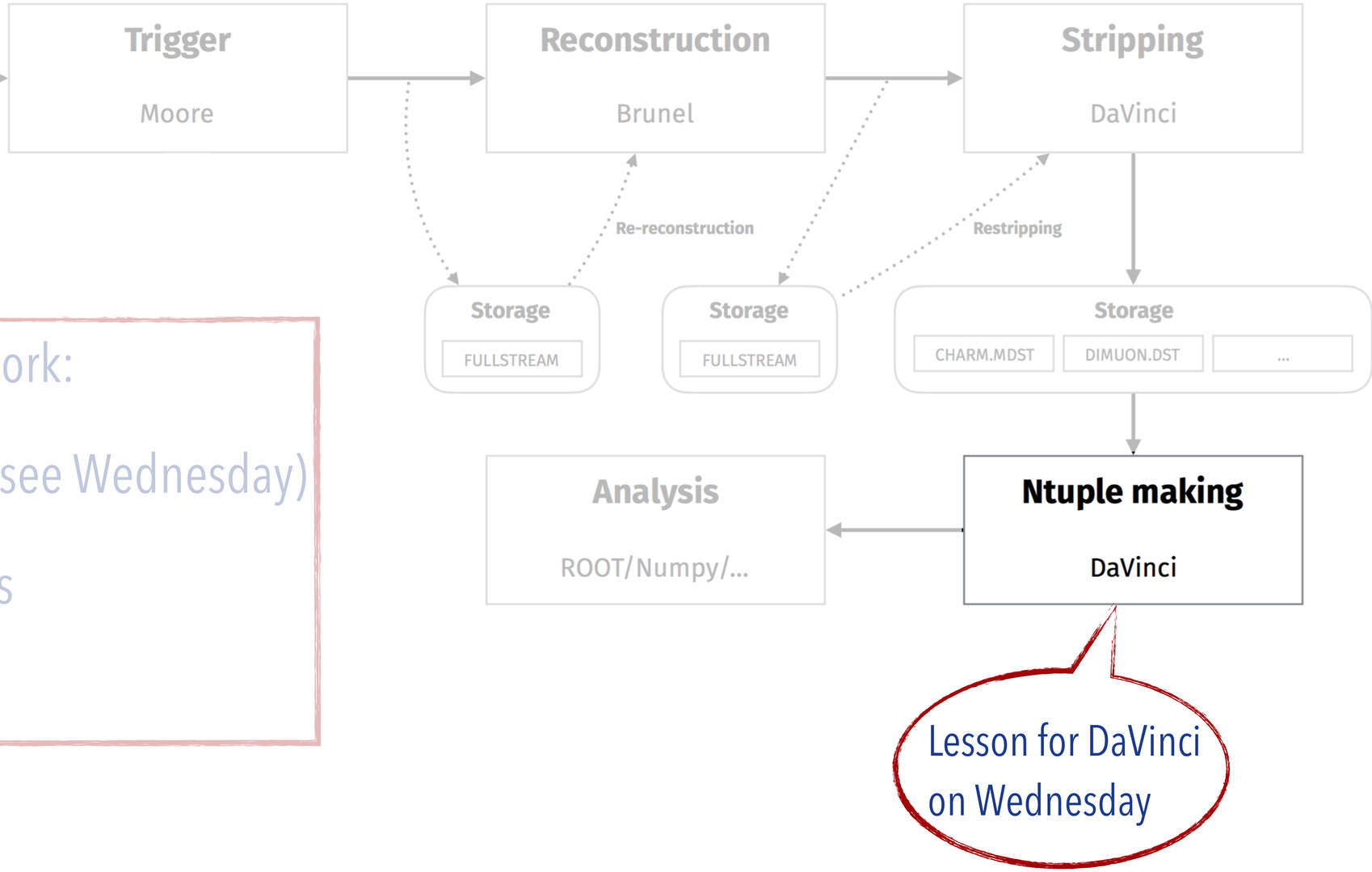

Probably most of your analysis work:

- Creating Ntuple with DaVinci (see Wednesday)
- Ntuples are saved in ROOT files
- Your personal analysis

# Collision data Flow during Run1

And finally:



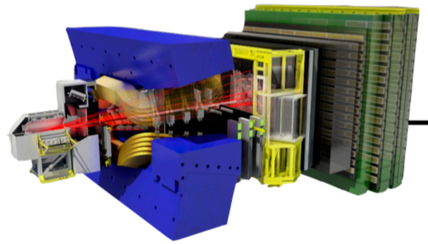



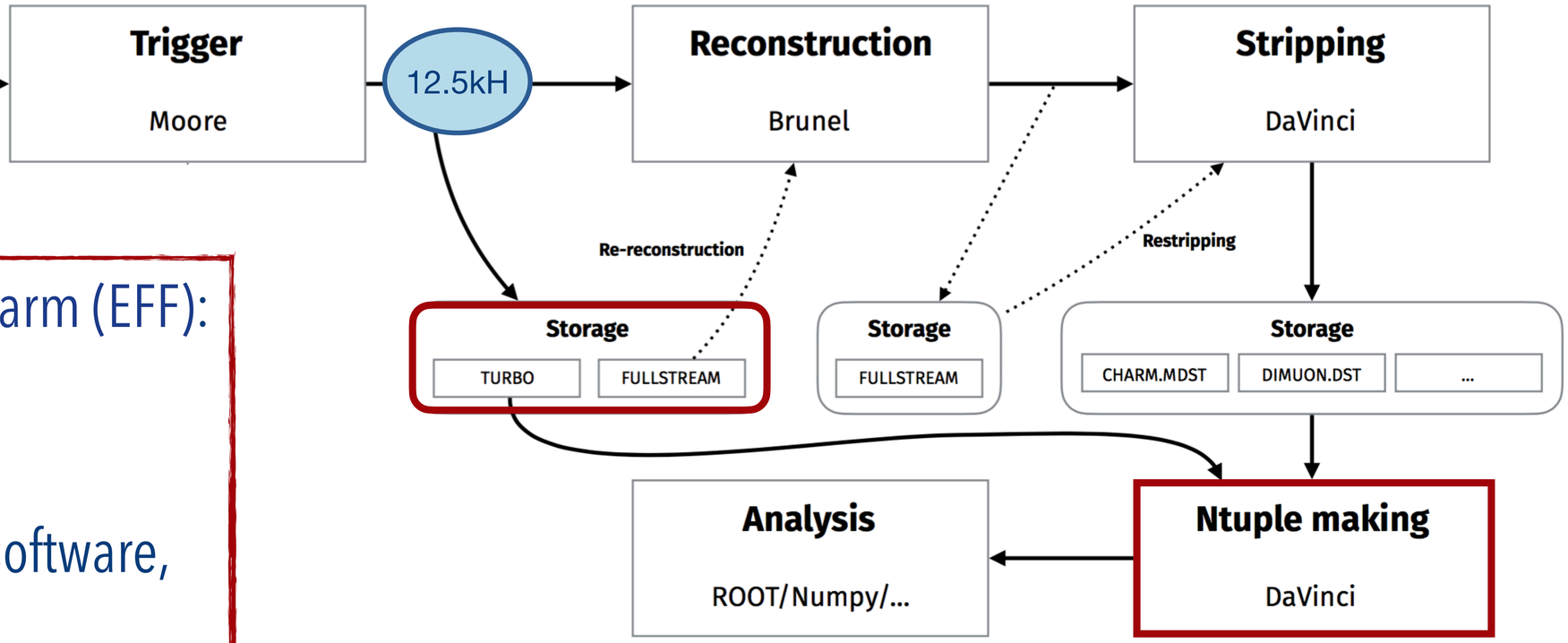

Probably most of your analysis work:

- Creating Ntuple with DaVinci (see Wednesday)
- Ntuples are saved in ROOT files
- Your personal analysis

# Collision dataflow during Run 2

## What has changed?



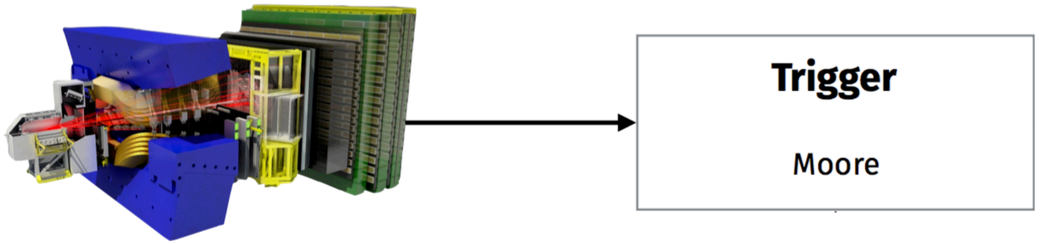



Upgrade of the event filter farm (EFF):

- HLT execution on EFF
- Improved hardware and software, automated calibration
- Full reconstruction now on HLT2!
- No offline reconstruction needed

# Collision dataflow during Run 2

Some details about the time for high level trigger

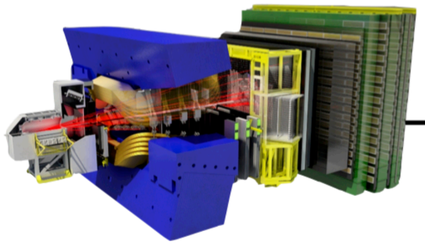


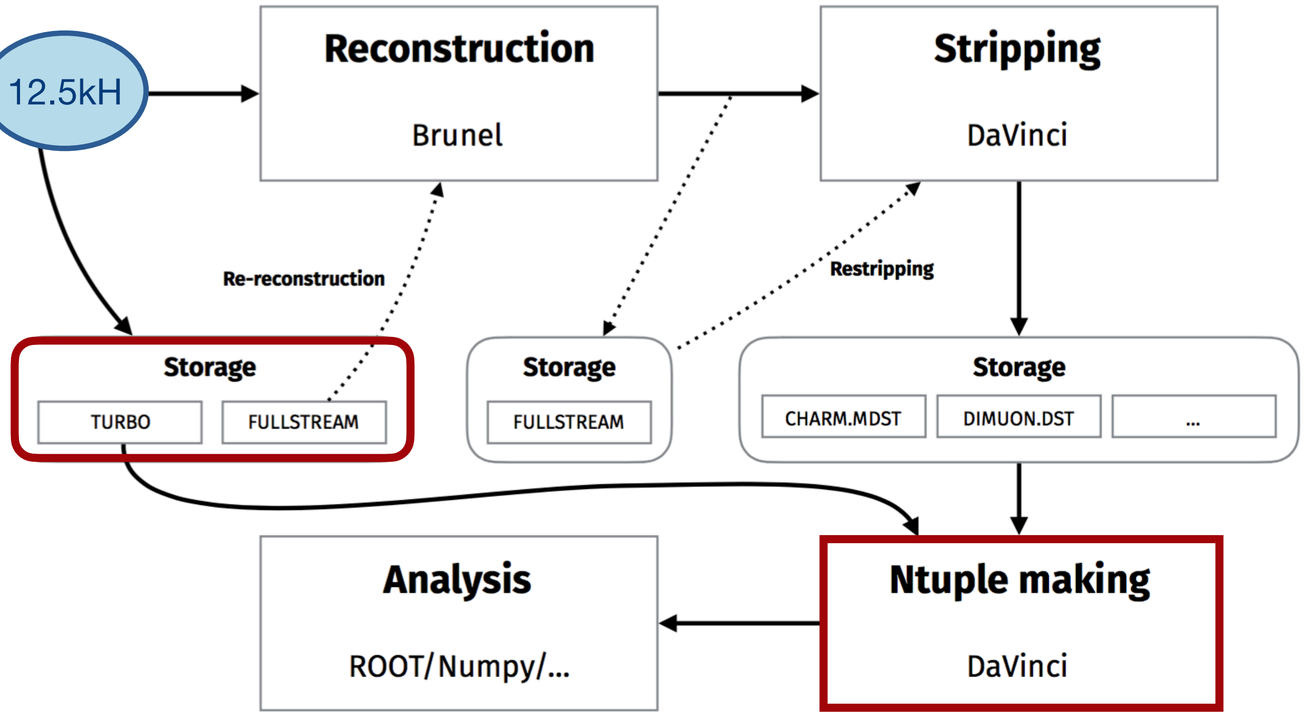

Next software stage called High Level Trigger

- HLT1: ms per event
- Alignment and calibration on 10Pb of buffer: mins/hours
- HLT2: full reconstruction takes hours

# Collision dataflow during Run 2

Opens the option to bypass with Turbo



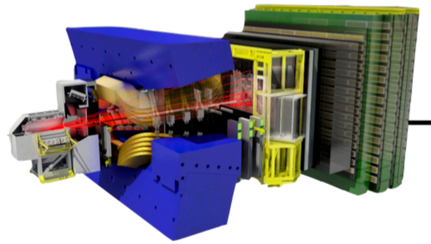


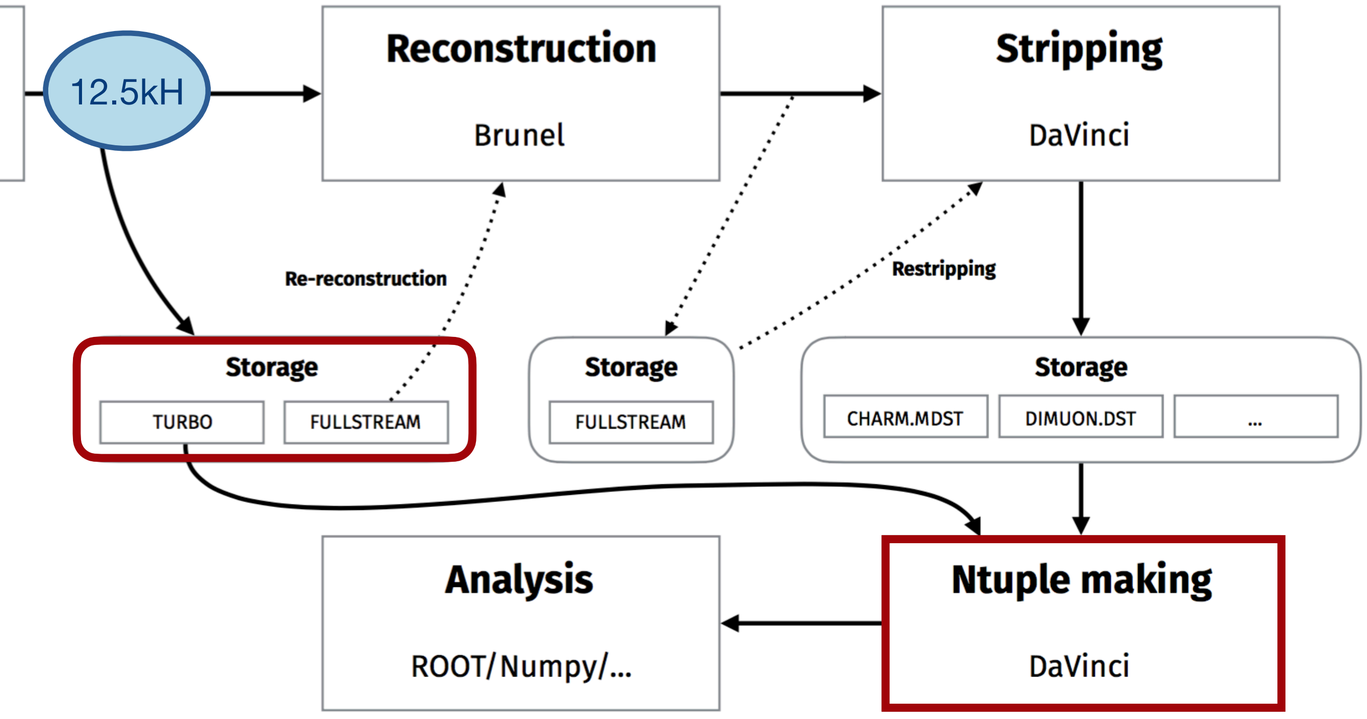

Still more events to store:

- Turbo stream: Saves only triggered tracks, rest of the event deleted
- Cannot be re-reconstructed
- For available lines ask trigger liaisons

# Collision data Flow during Run2

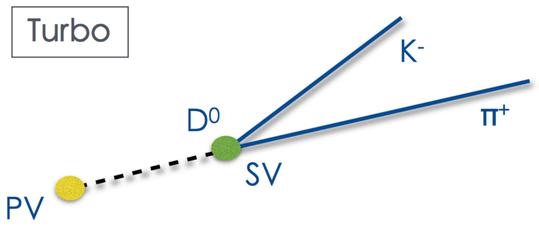
## Opens the option to bypass with Turbo

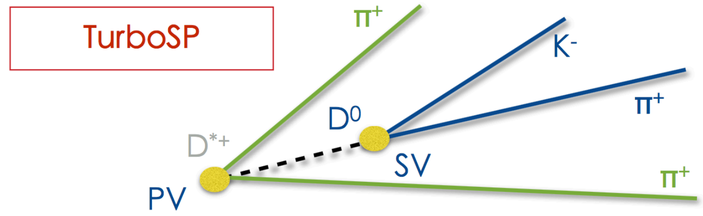





Three different Turbo definitons:

- Turbo: Saves only triggered tracks
- Turbo++: additional track information
- TurboSB: Free selection of additional information to save



# Summary collision dataflow

Run 1

- HLT not accurate enough
- Offline Reconstruction always needed
- Maximum speed to disk 5kHz

# Summary collision data flow

## Run 1

- HLT not accurate enough
- Offline Reconstruction always needed
- Maximum speed to disk 5kHz

## Run 2

- Higher data rate
- HLT same accuracy without offline reconstruction
- Turbo Stream as bypass option
- Maximum speed to disk 12.5kHz

# Summary collision dataflow

Run 1

- HLT not accurate enough
- Offline Reconstruction always needed
- Maximum sp

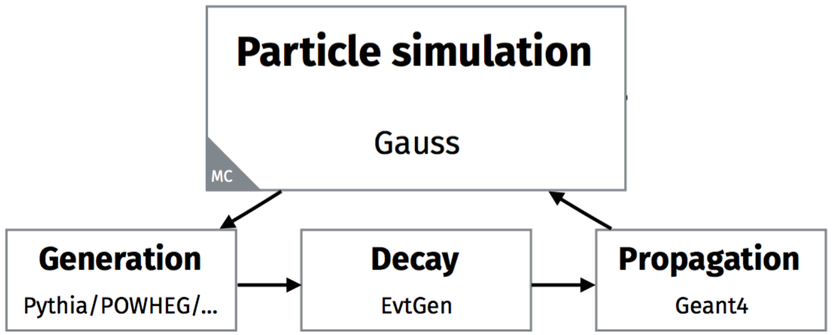
Run 2

- Higher data rate
- HLT same accuracy without offline reconstruction
- Maximum speed to disk 12.5kHz

Dataflow in Run 3 and related topics on Thursday

# Simulation dataflow

But what about simulation?




Creating particle simulation:

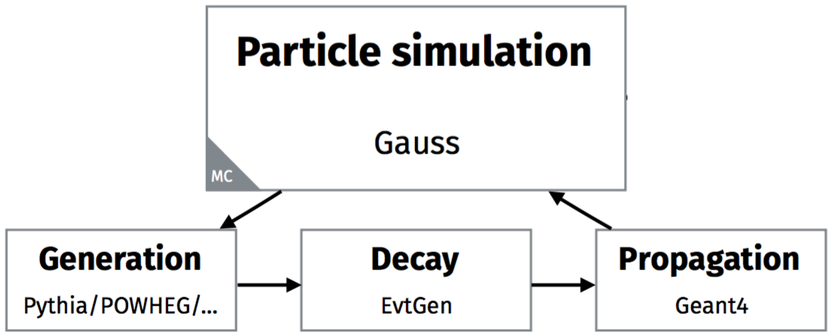
- Generation of the hard process e.g. Pythia
- Decay processed with DecFiles in EvtGen
- Propagation through detector: Geant4
- All executed in Gauss framework

# Simulation dataflow

But what about simulation?

Generator level MC

→ TRUE variables

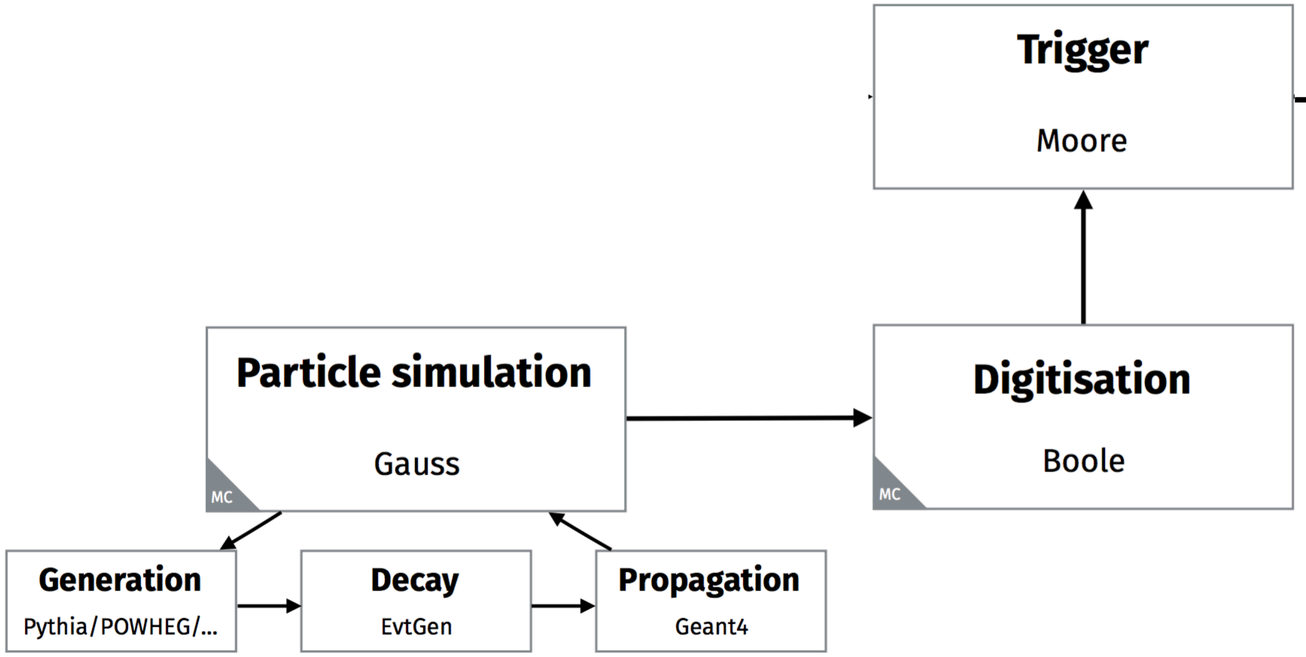



Creating particle simulation:

- Generation of the hard process e.g. Pythia
- Decay processed with DecFiles in EvtGen
- Propagation through detector: Geant4
- All executed in Gauss framework

# Simulation dataflow

More details on the starterkit webpage

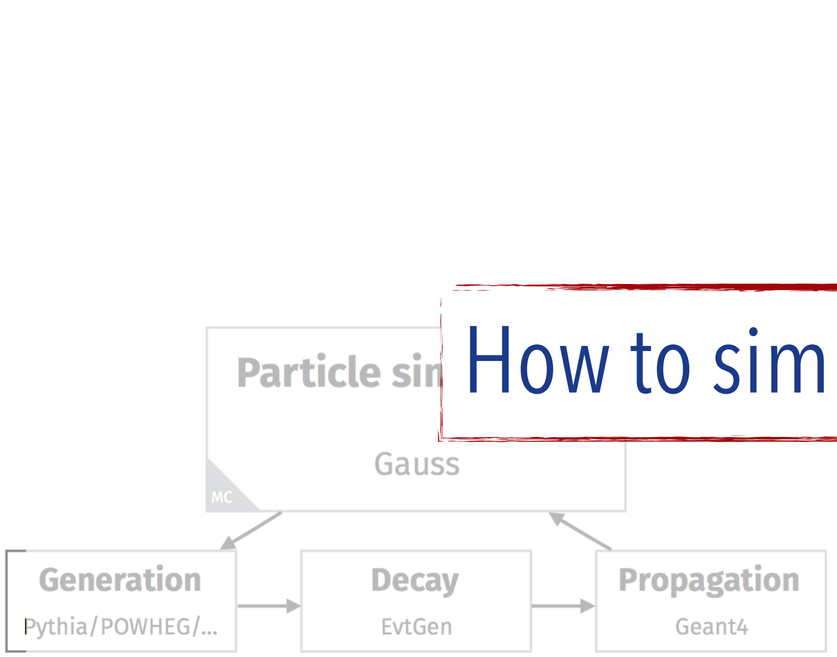



After generating the particle simulation:

- Digitalisation of detector response with Boole
- Simulation can be processed like data
- Access to efficiencies

# Simulation dataflow

More details on the starterkit webpage




... Reconstruction level MC

How to simulation and MC request on Wednesday

After generation, the particle simulation:

Digitisation of detector response with Boole

- Simulation can be processed like data
- Access to efficiencies

# Analysis flow

First accessing Ntuples

Step 1

Testing scripts

Run locally

Starterkit lesson about it

# Analysis flow

First accessing Ntuples

Step 1

Testing scripts

Run locally

Starterkit lesson about it

Step 2a

Running scripts on grid

via ganga

Starterkit lesson about it

# Analysis flow

First accessing Ntuples

Step 1

Testing scr

Run locally

Starterkit lesson about it

Step 2a

Running scripts

via ganga

Starterkit lesson about it

Introduction to this first session on Wednesday

# Analysis flow

First accessing Ntuples

## Step 1

Testing scripts

Run locally

Starterkit lesson about it

## Step 2a

Running scripts on grid

via ganga

Starterkit lesson about it

## Step 2b

Running scripts centralised

via Analysis Production

Starterkit lesson about it

# Analysis flow

First accessing Ntuples

Step 1

Testing

Run locally

Starterkit lesson about it

Step 2a

Running scripts

via ganga

Starterkit lesson about it

Step 2b

Running scripts

via Analysis Production

Starterkit lesson about it

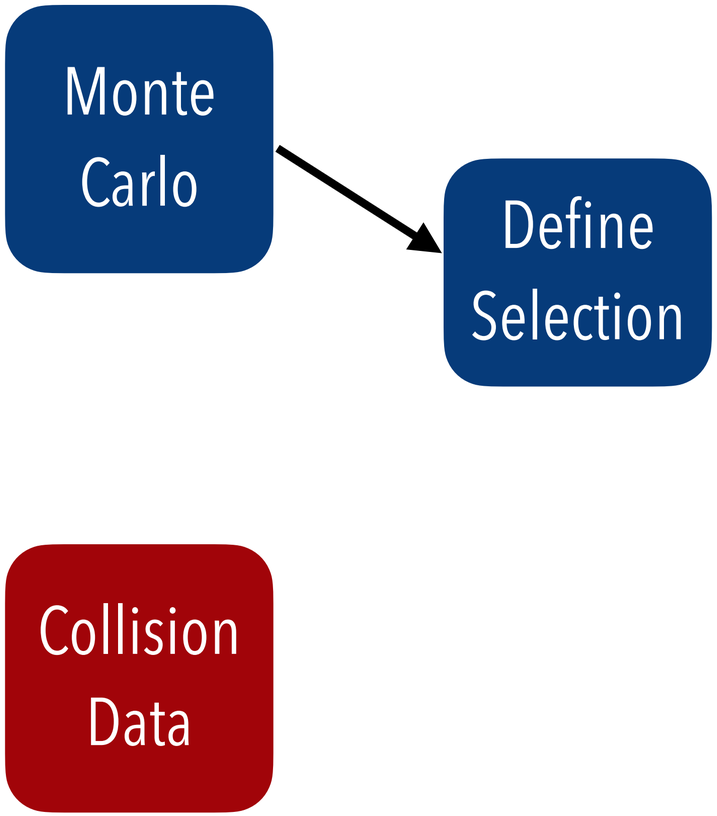
Introduction to Analysis Production on Wednesday

# Analysis flow

Now the actual analysis flow

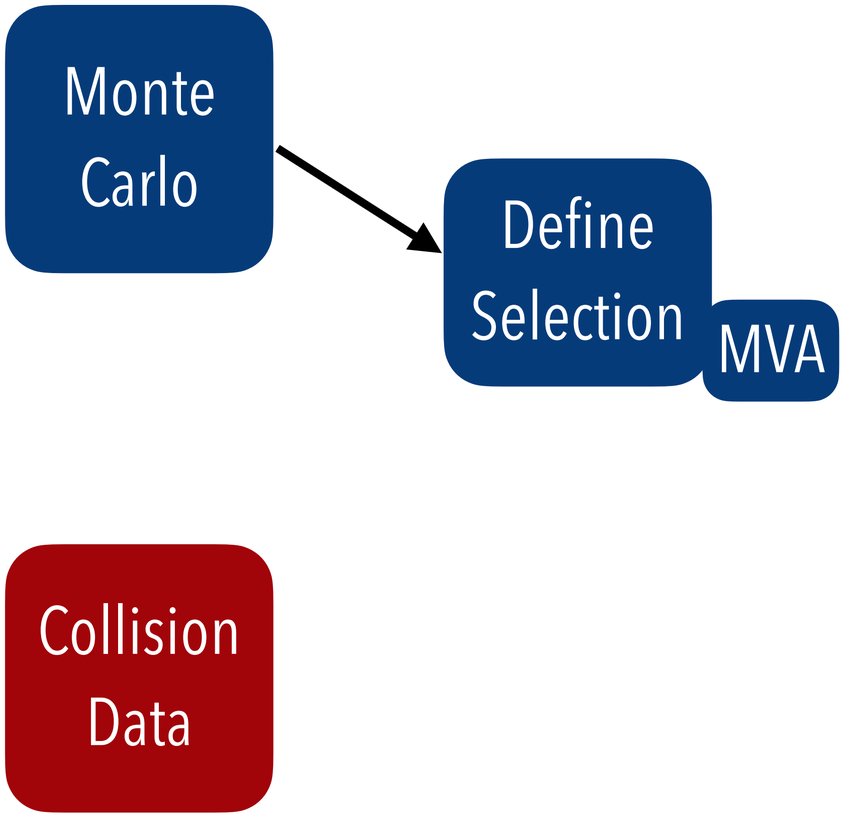
# Analysis flow

Now the actual analysis flow



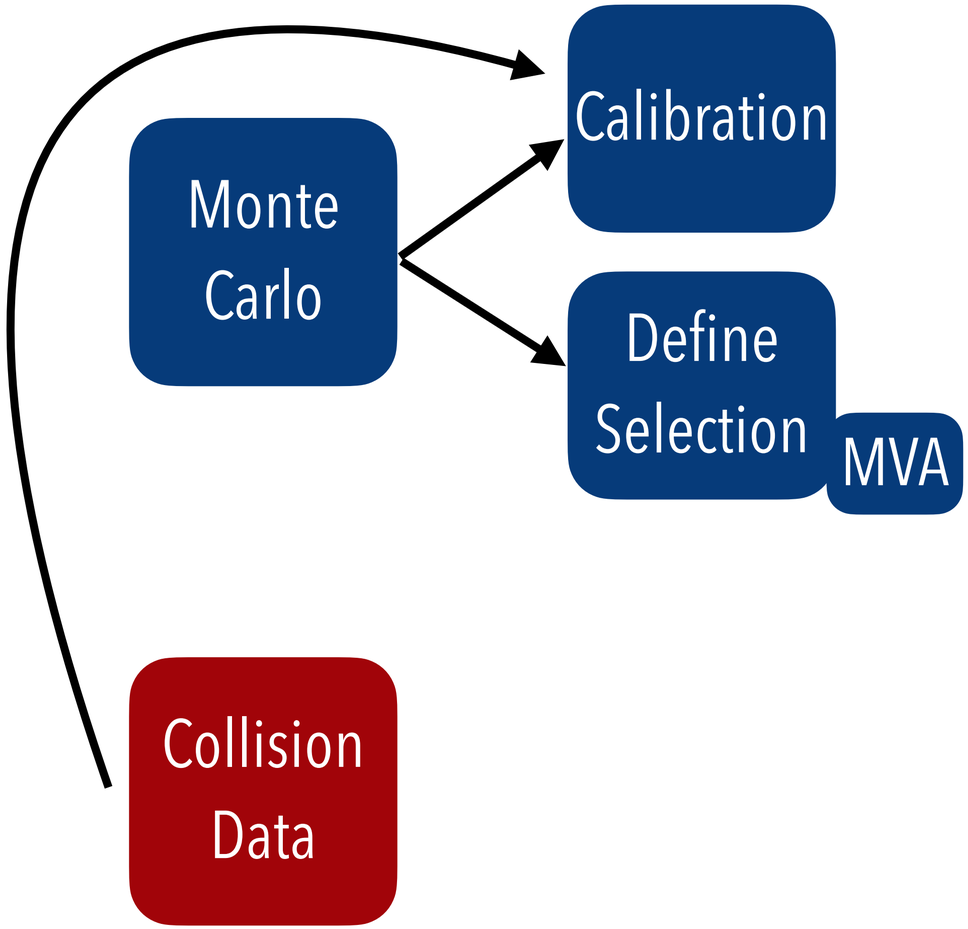
# Analysis flow

Now the actual analysis flow



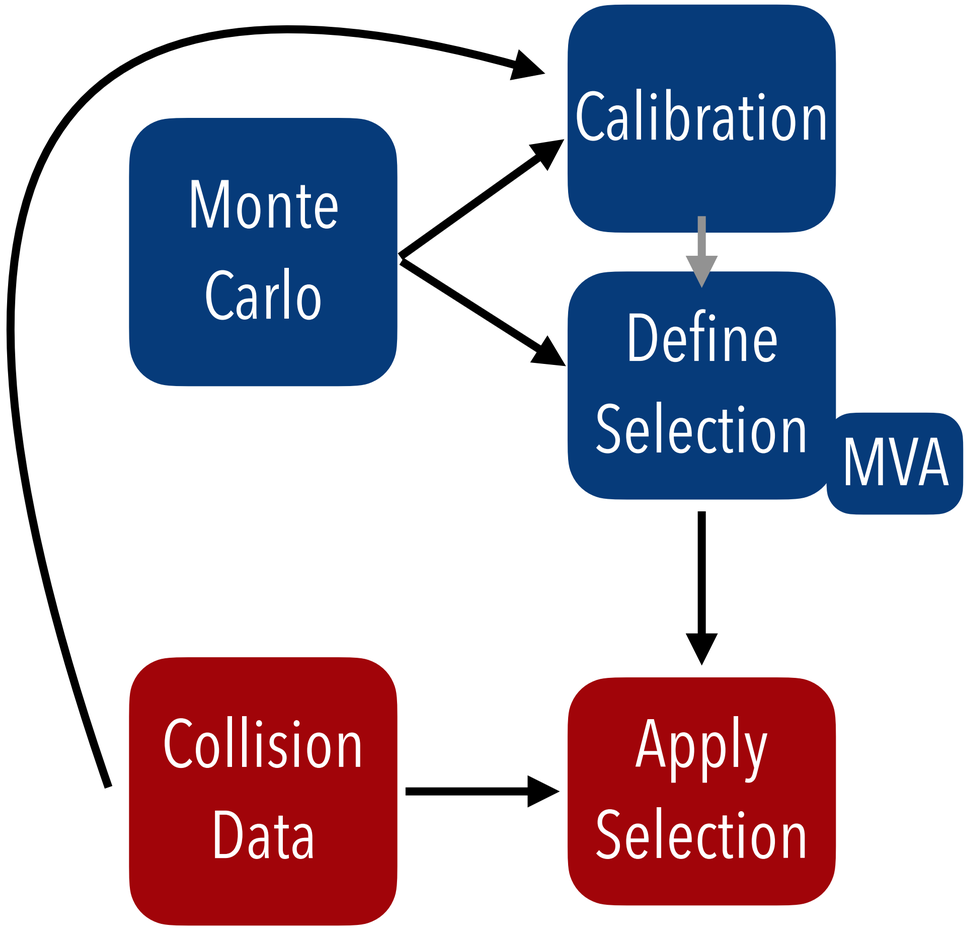
# Analysis flow

LHCb software: PIDCalib, TrackCalib,…



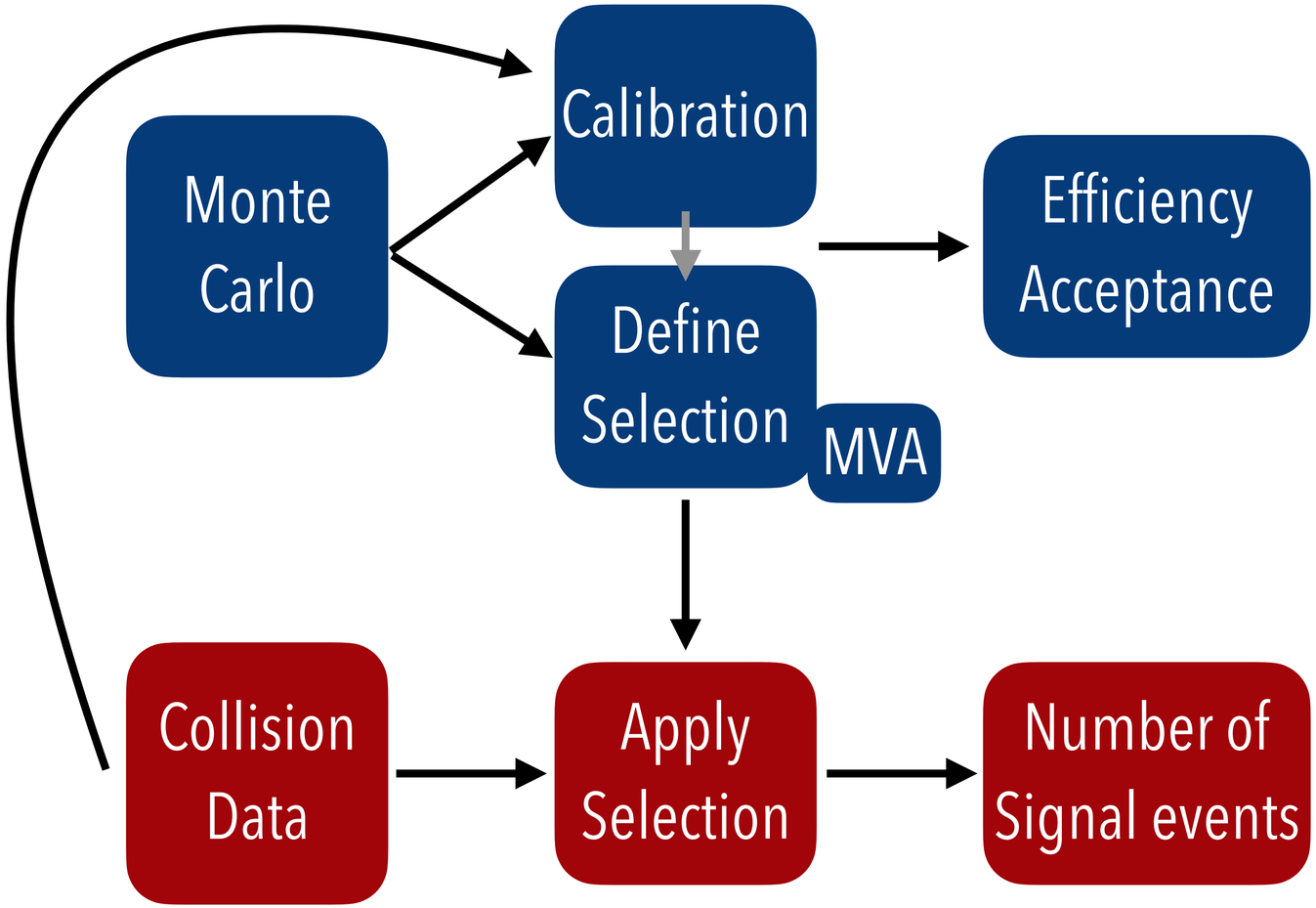
# Analysis flow

Now the actual analysis flow



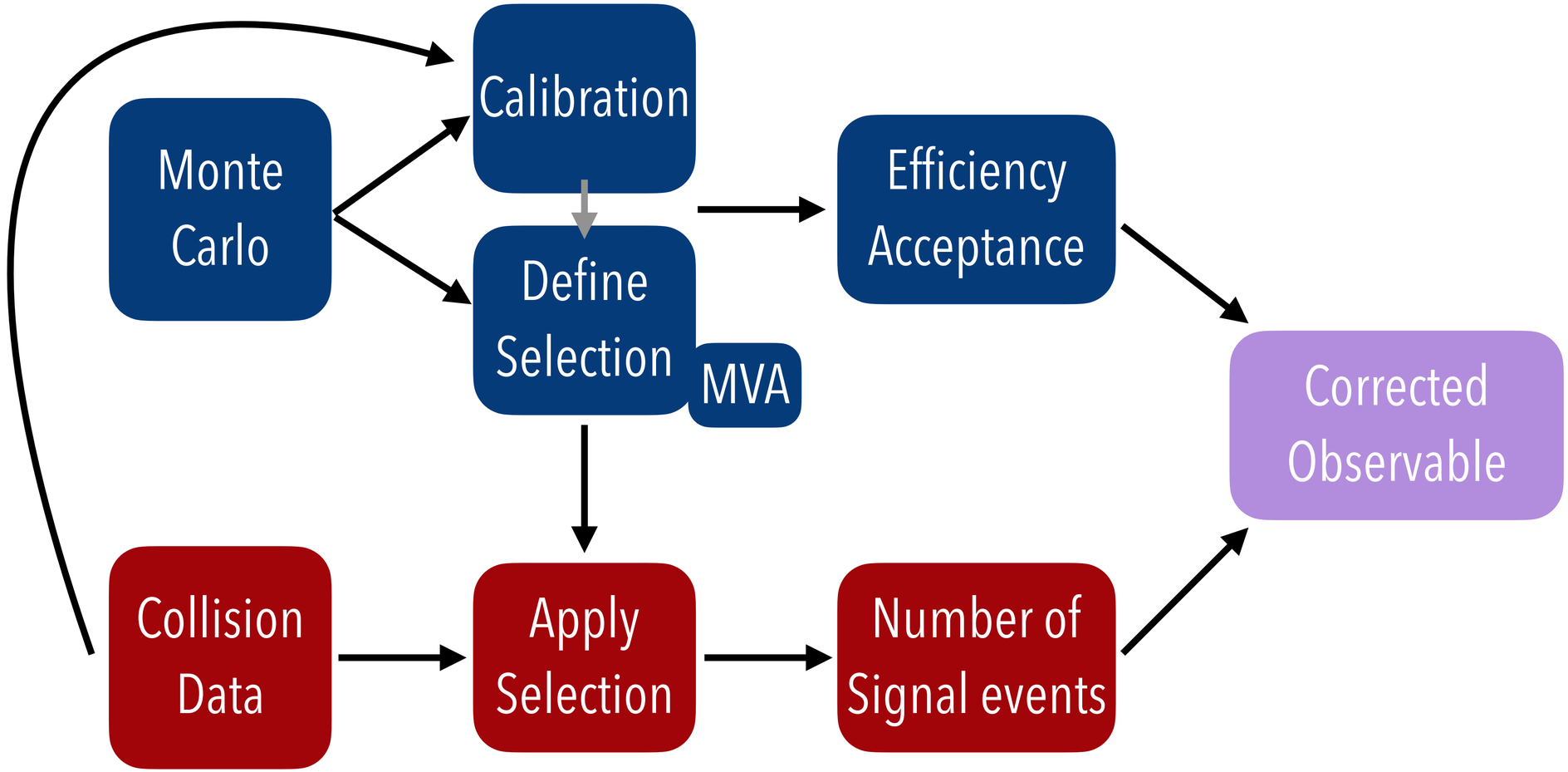
# Analysis flow

Now the actual analysis flow



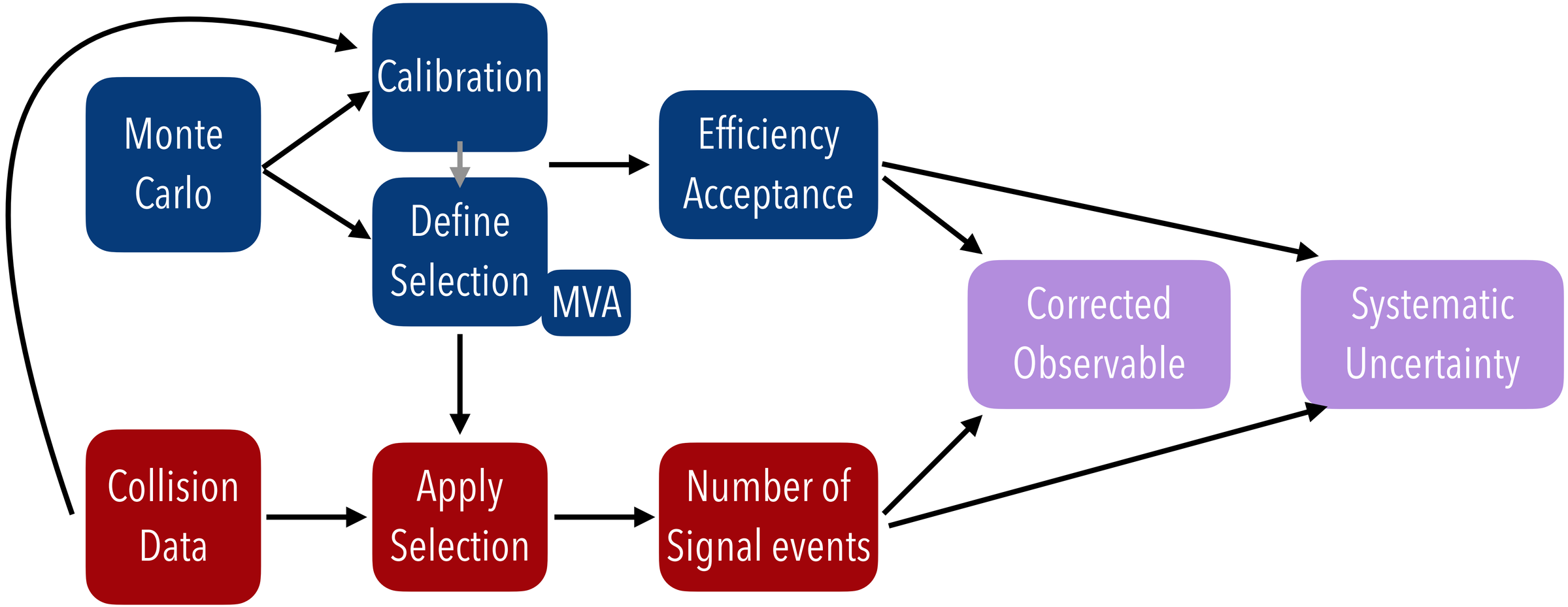
# Analysis flow

Now the actual analysis flow



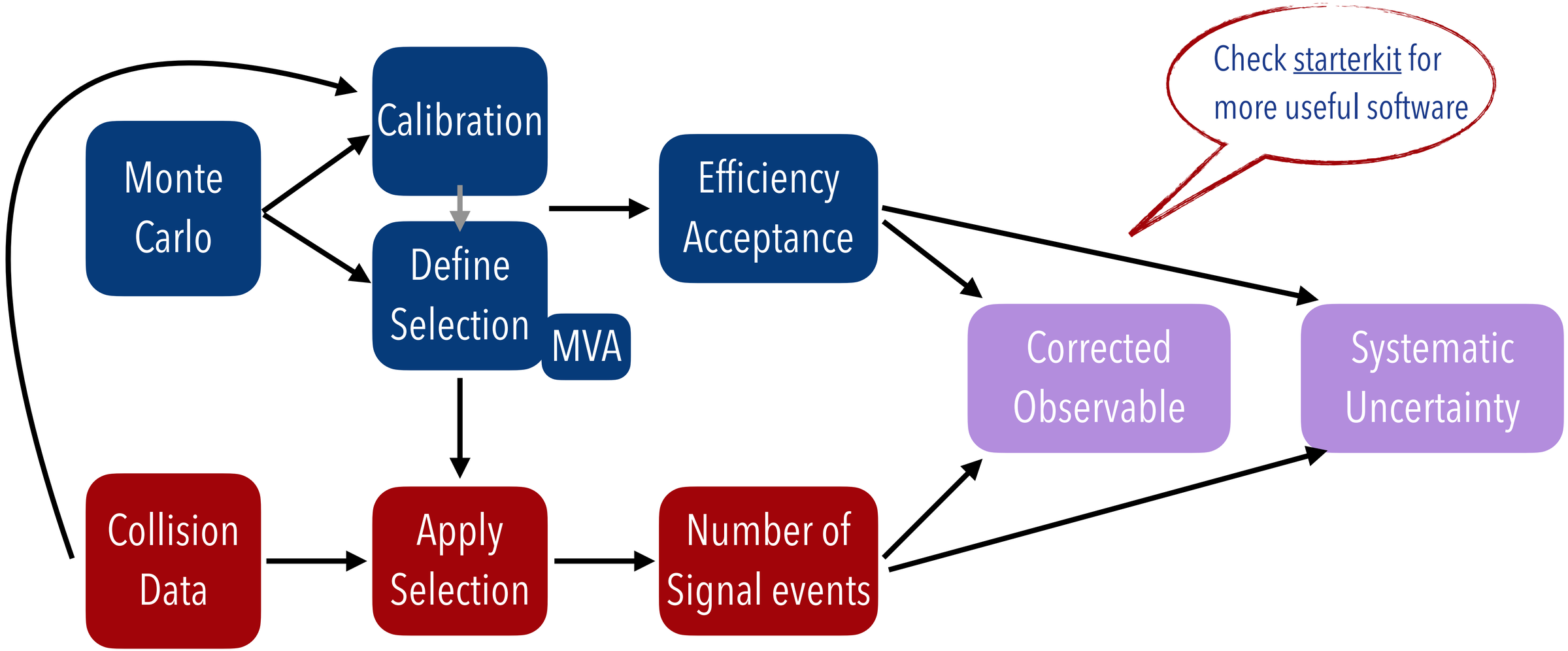
# Analysis flow

Now the actual analysis flow



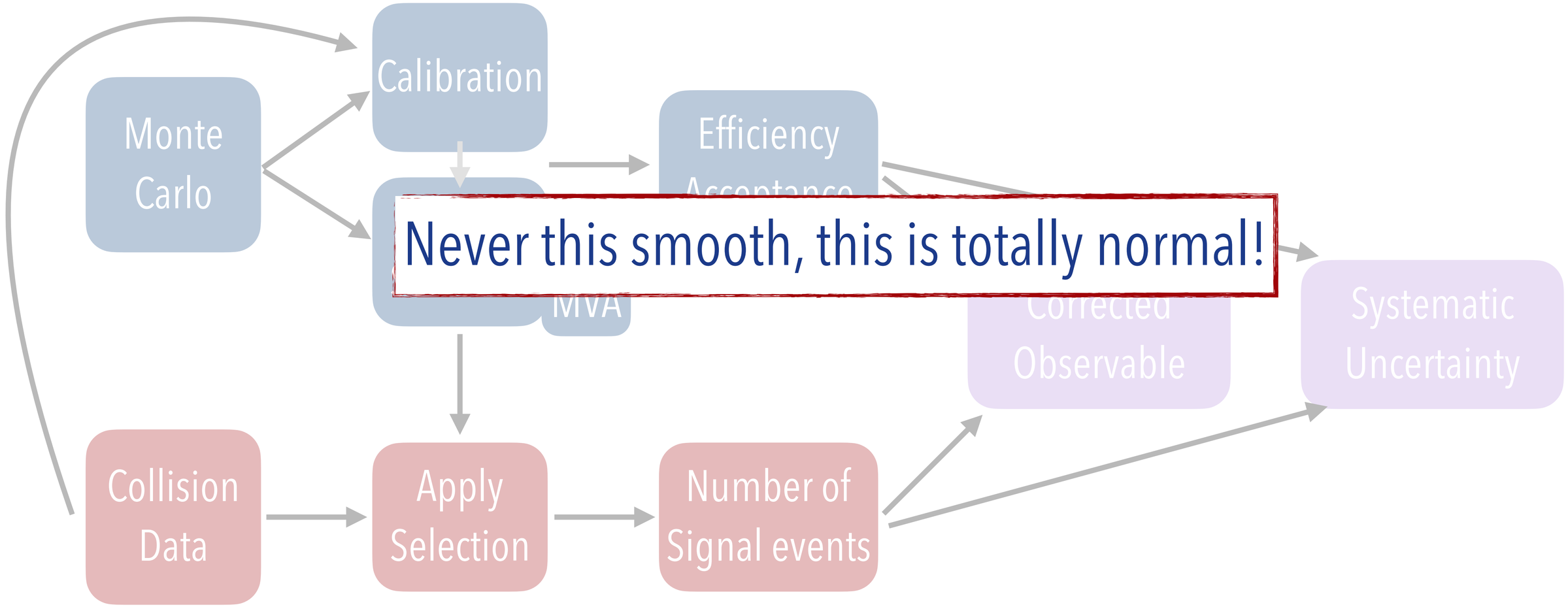
# Analysis flow

Now the actual analysis flow



# Analysis flow

Now the actual analysis flow



# Towards publication

Analysis now in principle done:

- Analysis note: Contains all studies, documentation of your analysis, published on CDS at the end
- Working group review: AnaNote in good state, working group reader, WG pre-approval talk, after answering all questions (from talks and readers) WG approval talk

# Towards publication

Analysis now in principle done:

- Analysis note: Contains all studies, documentation of your analysis, published on CDS at the end

# Towards publication

Analysis now in principle done:

- Analysis note: Contains all studies, documentation of your analysis, published on CDS at the end
- Working group review: AnaNote in good state, working group reader, WG pre-approval talk, after answering all questions (from talks and readers) WG approval talk
- Physics reviewer: Will be assigned upon request of WG convener, same procedure as with WG, update of AnaNote, give ok to unblind with WG convener (if necessary)

# Towards publication

Analysis now in principle done:

- Analysis note: Contains all studies, documentation of your analysis, published on CDS at the end
- Working group review: AnaNote in good state, working group reader, WG pre-approval talk, after answering all questions (from talks and readers) WG approval talk
- Physics reviewer: Will be assigned upon request of WG convener, same procedure as with WG, update of AnaNote, give ok to unblind with WG convener (if necessary)
    - Analysis code need to be accessible on Gitlab, including clear instruction how to reproduce your results, necessary tuples stored on EOS

# Towards publication

Analysis now in principle done:

- Analysis note: Contains all studies, documentation of your analysis, published on CDS at the end
- Working group review: AnaNote in good state, working group reader, WG pre-approval talk, after answering all questions (from talks and readers) WG approval talk
- Physics reviewer: Will be assigned upon request of WG convener, same procedure as with WG, update of AnaNote, give ok to unblind with WG convener (if necessary)
    - Analysis code need to be accessible on Gitlab, including clear instruction how to reproduce your results, necessary tuples stored on EOS
    - Reviewing paper draft, if happy request Editorial Board Reviewer

# Towards publication

Analysis now in principle done:

- Analysis note: Contains all studies, documentation of your analysis, published on CDS at the end
- Working group review: AnaNote in good state, working group reader, WG pre-approval talk, after answering all questions (from talks and readers) WG approval talk
- Physics reviewer: Will be assigned upon request of WG convener, same procedure as with WG, update of AnaNote, give ok to unblind with WG convener (if necessary)
    - Analysis code need to be accessible on Gitlab, including clear instruction how to reproduce your results, necessary tuples stored on EOS
    - Reviewing paper draft, if happy request Editorial Board Reviewer
- Collaboration wide review process: Approval to go to paper talk, institute reviewers, two rounds, all comments need to be addressed, followed by all reviewer and physics coordinator

# Towards publication

Analysis now in principle done:

- Analysis note: Contains all studies, documentation of your analysis, published on CDS at the end
- Working
  - after answ
- Physics re
  - update o
    - Reviewing paper draft, if happy request Editorial Board Reviewer
- Collaboration wide review process: Approval to go to paper talk, institute reviewers, two rounds, all comments need to be addressed, followed by all reviewer and physics coordinator

al talk,

with WG,

ow to

Helpful are the LHCb guidelines for the preservation,

the flowchart for review steps, the Publishing FAQ

And ask your colleagues.

In the end we are a collaboration! :)

# Analysis preservation

Analysis preservation needs to things

- If you don't use the lb-conda environment , preserve the package versions of the software e.g. with your own conda environment or a docker container
- Analysis code need to be accessible on Gitlab, the use of snakemake can make it easier to make your workflow reproducible

# Analysis preservation

Analysis preservation needs to things

- If you don't us ns of the software
  e.g. with your
- Analysis code make it easier
  to make your workflow reproducible

Never heard of snakemake?

No problem, next up a tutorial about it :)