



# Run Dependent Monte Carlo at Belle II

[Alberto Martini](#)

On behalf of the Belle II Data Production group

21<sup>st</sup> International Workshop - ACAT 2022, 26 October 2022



# The Belle II experiment

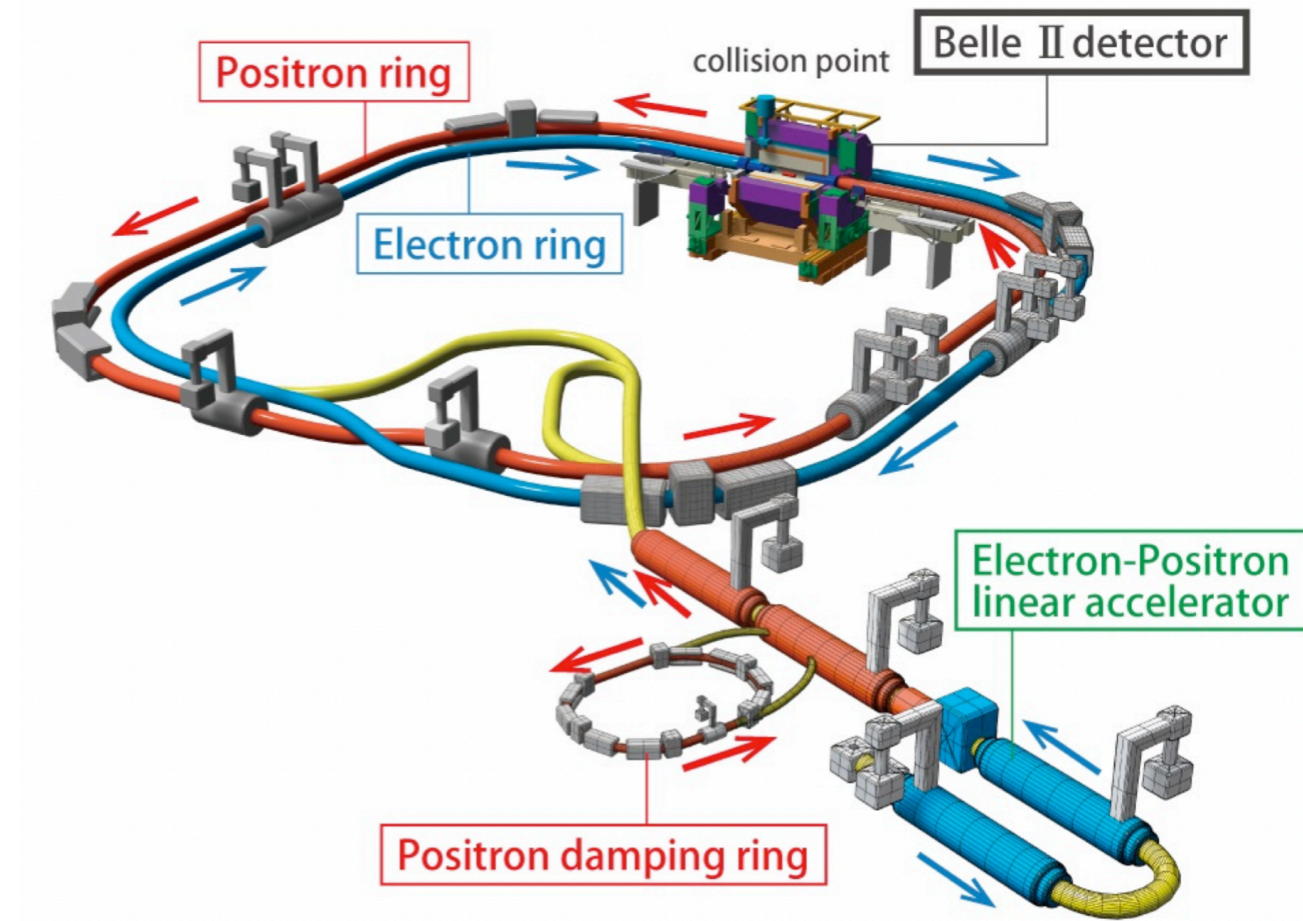
Belle II is an experiment at SuperKEKB, a second generation B-factory

Collisions at a specific energy value:

10.58 GeV



Production of a lot of events of the same type

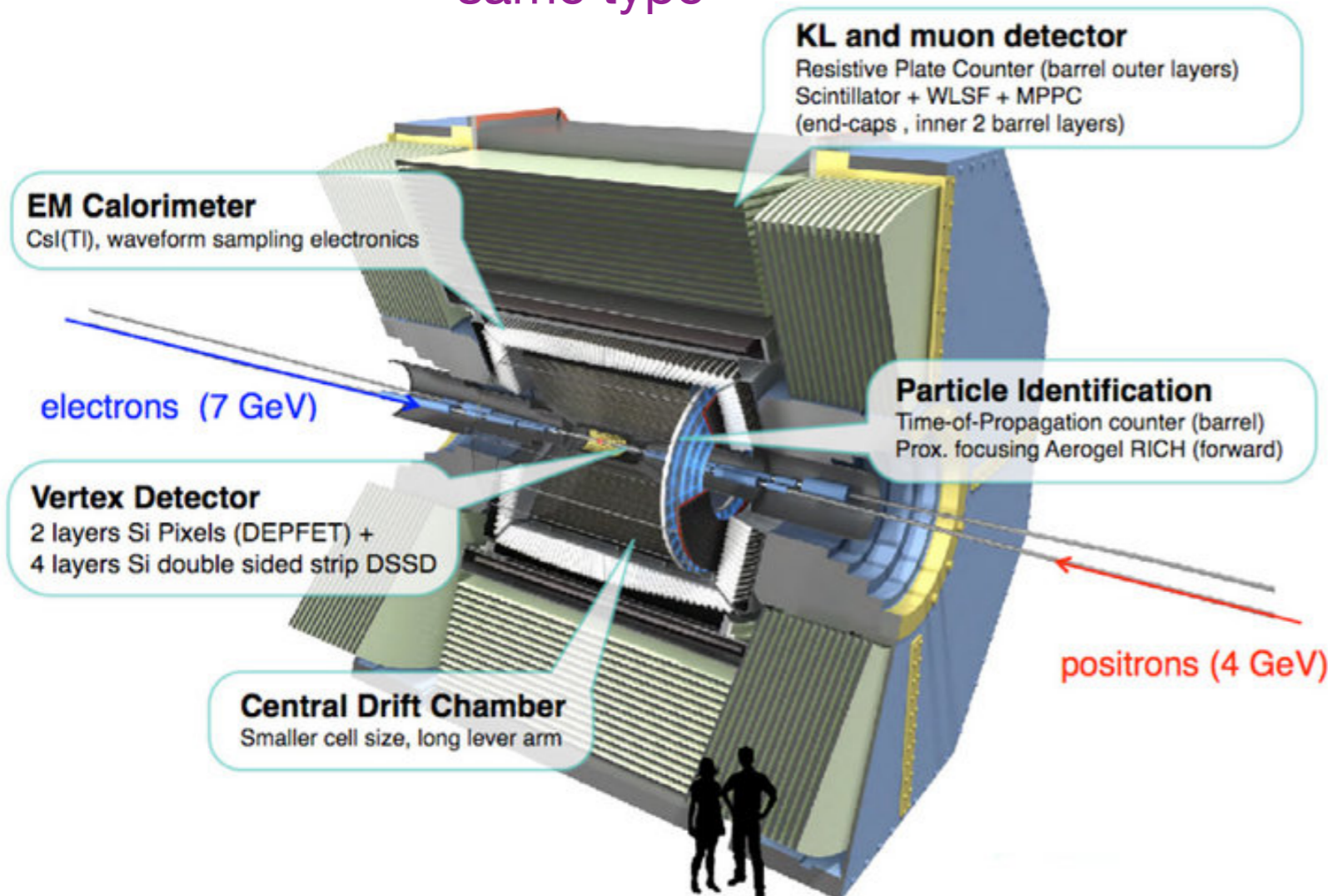
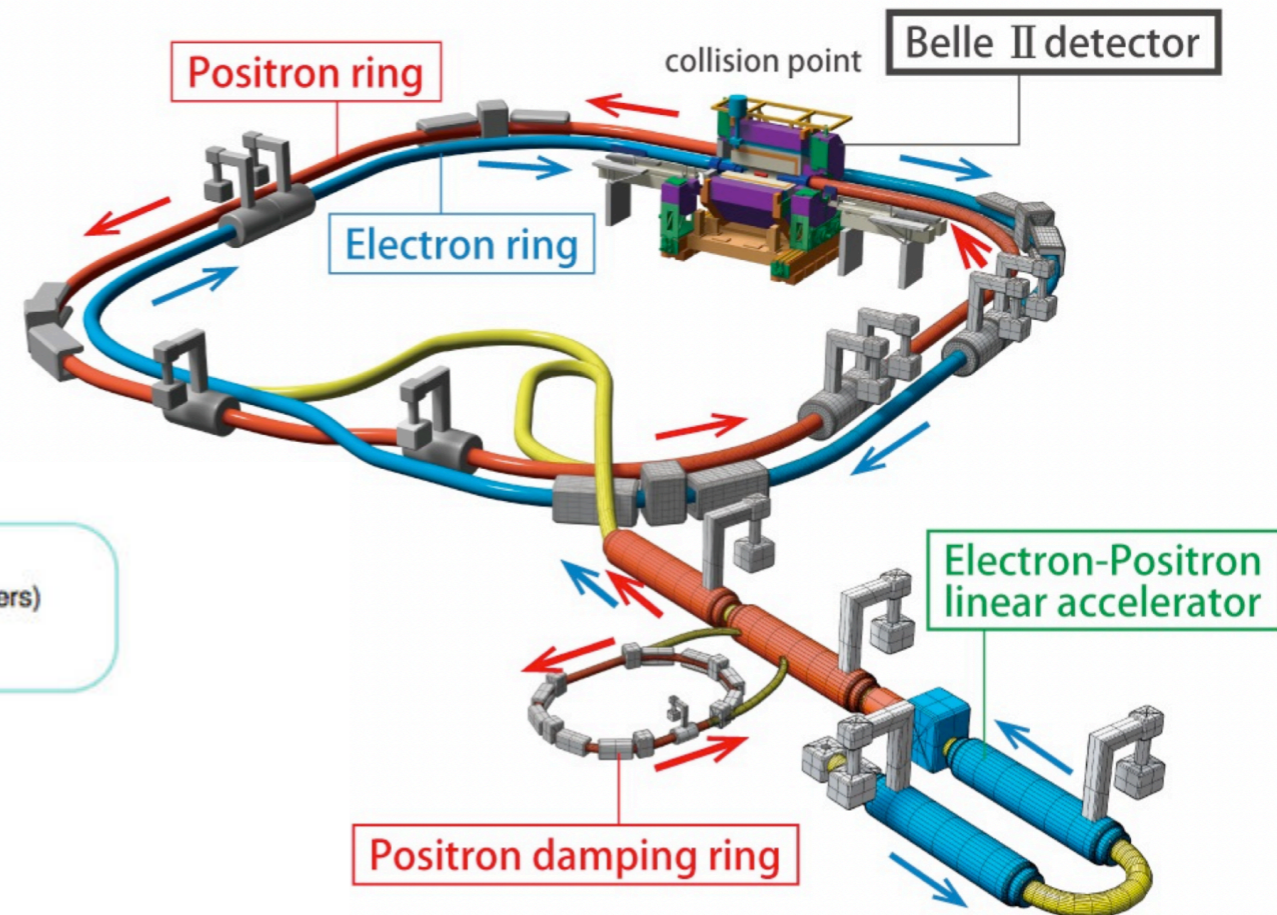


# The Belle II experiment

Belle II is an experiment at SuperKEKB, a second generation B-factory

Collisions at a specific energy value:  
10.58 GeV

Production of a lot of events of the  
same type



**Each sub-detector** is used to  
measure specific quantities and  
**is calibrated individually**

# Data taking status

Belle II is currently going through the first long shutdown since July 2022

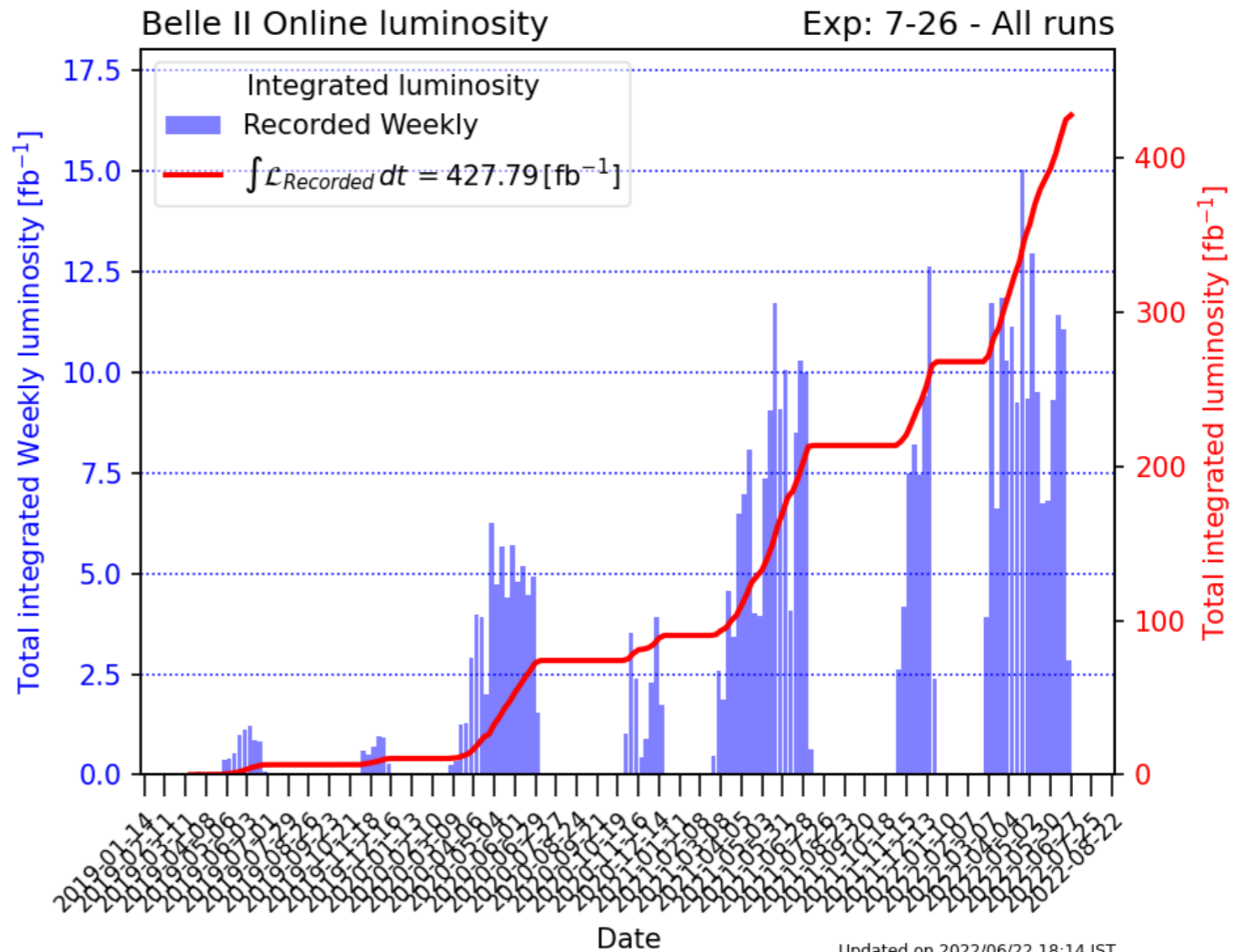


Between 2019-2022 Belle II collected a huge amount of data: >2PB of raw data so far

A lot of more data will be collected before 2033



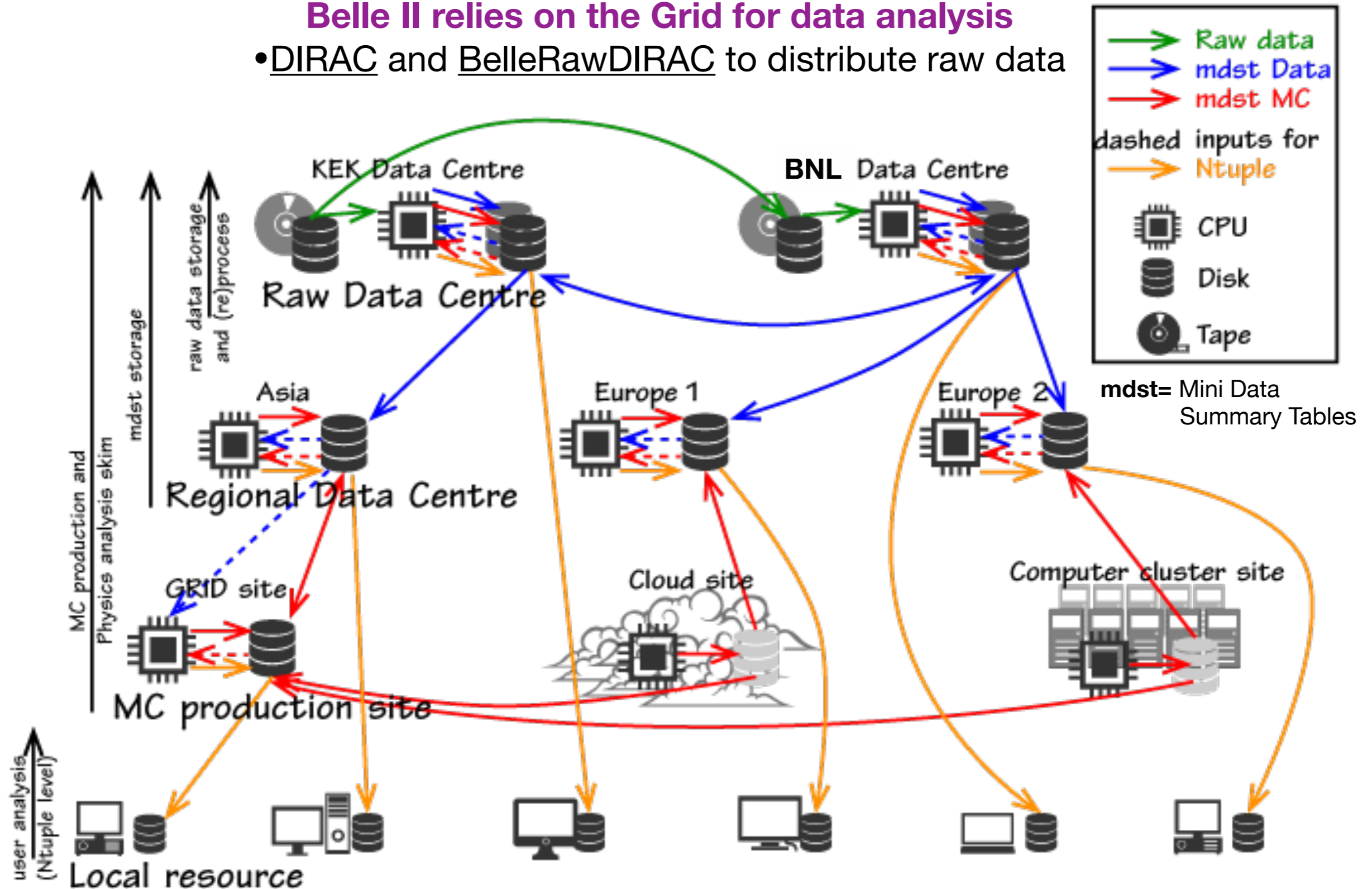
Computing resources are a key aspect to properly store and analyse the data



# Grid computing system

## Belle II relies on the Grid for data analysis

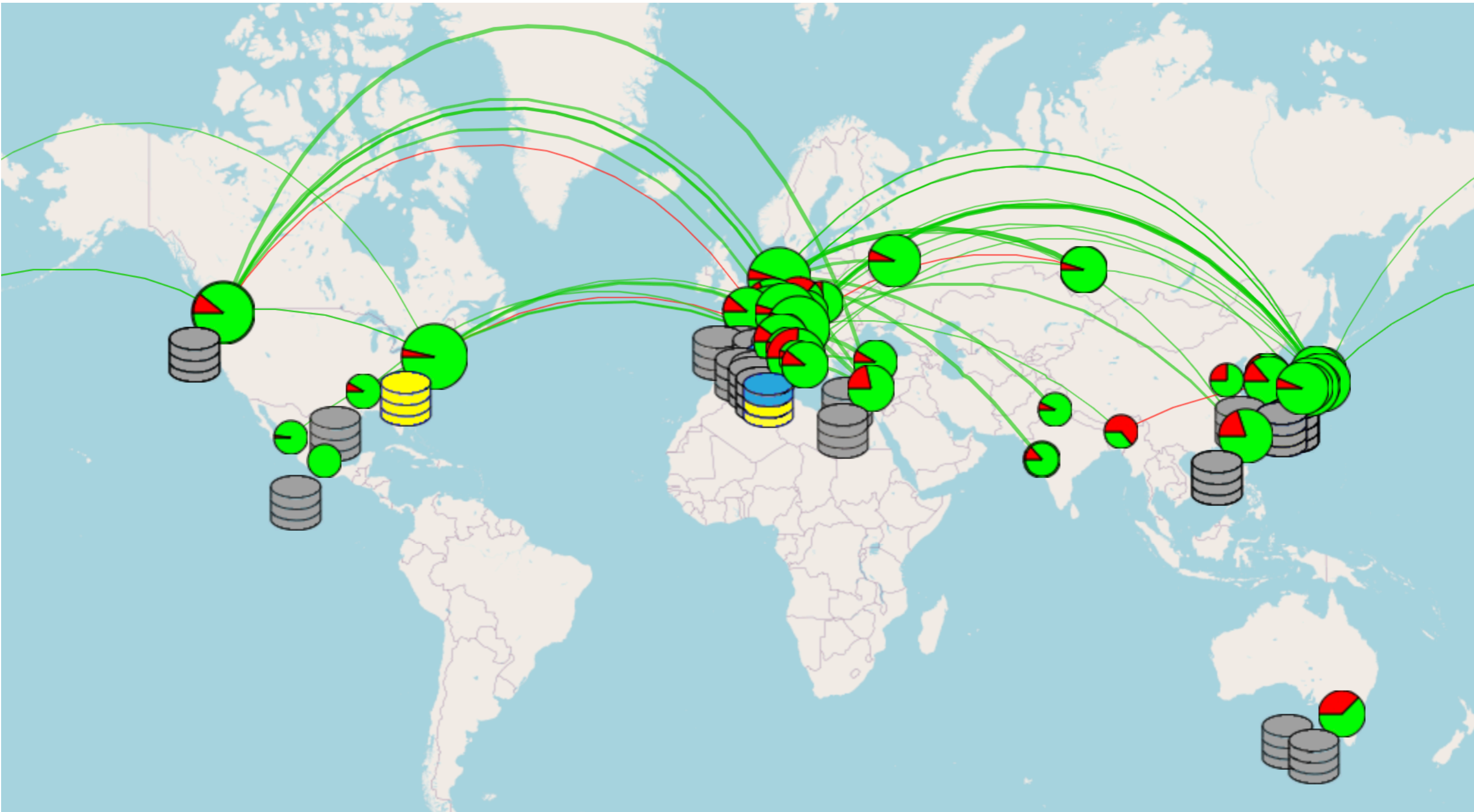
- DIRAC and BelleRawDIRAC to distribute raw data



ref: <https://iopscience.iop.org/article/10.1088/1742-6596/119/6/062048/pdf>

# GRID computing system

Data centres are displaced all over the world → major sites in USA, Canada, Japan, Europe



# Data production schema

**payloads**= set of specific sub-detector calibrations (i.e. KLM dead channels)

## *Production steps at Belle II after raw data*

### Calibration

### Data production

Collection of **payloads** from each sub-group

Production starts just after calibration is over



Extract **proper detector configuration** used for data processing

**Skimmed and not skimmed datasets are provided**



# Data production schema

**Skim**= data and MC produced according to specific analysis requirements  
→ reduced events/size

## *Production steps at Belle II after raw data*

### Calibration

Collection of payloads from each sub-group



Extract proper detector configuration used for data processing

### Data production

Production starts just after calibration is over



Skimmed and not skimmed datasets are provided

### Full simulation-Monte Carlo (MC) production

Two types of MC being produced:

- MC run-independent (MCri)
- **MC run-dependent (MCrd)**



Focus of this talk

### Skim production

**Skim** production starts just after data & MC are being produced



Skims should cover each physics analysis → very convenient due to reduced sizes





# Info on data-taking nomenclature



**INFO:** Data taking periods at Belle II are identified via **experiment and run numbers**.  
Each **experiment** corresponds to a specific configuration of the detector or accelerator → changing experimental conditions implies a change of experiment number  
Each experiment consists of a set of **runs** with different size and time length



# Why MC run-dependent?



**INFO:** Data taking periods at Belle II are identified via **experiment and run numbers**.  
Each **experiment** corresponds to a specific configuration of the detector or accelerator → changing experimental conditions implies a change of experiment number  
Each experiment consists of a set of **runs** with different size and time length

## What is MCrd and why it is important?

Two main aspects that distinguishes MCrd from MCri:

- **Background type:** data driven background events → discussed later
- **Detector geometry and configuration:** data driven calibration constants extracted → discussed later
- **Run per run evolving detector conditions**
- **The weights of different runs in an experiment is preserved**
- **More realistic detector description**



# MC run-dependent production steps

---

- 1 Background production (BGOverlay) → see next slides
- 2 Preparation of detector configuration → see next slides



# MC run-dependent production steps

1 Background production (BGOverlay) → see next slides

2 Preparation of detector configuration → see next slides

3 Preparation of the production:

- Scripts are prepared and tested on a dedicated server by the manager
- Productions are registered on the grid by the manager
- Each production is launched by the Computing manager
- Automatic preparation and submission of productions is in preparation

4 Bookkeeping:

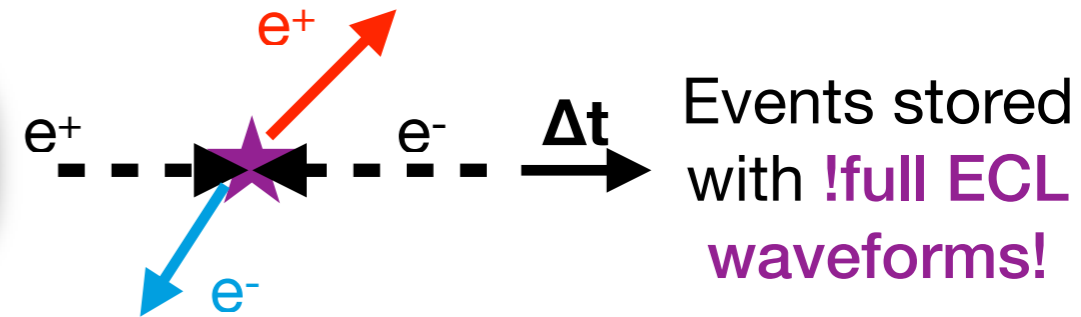
- Usage of a dedicated database to search for the MC samples
- Status and readiness are constantly uploaded on confluence



# MC run-dependent BGOOverlay files

Uses the random triggers events to extract the background information

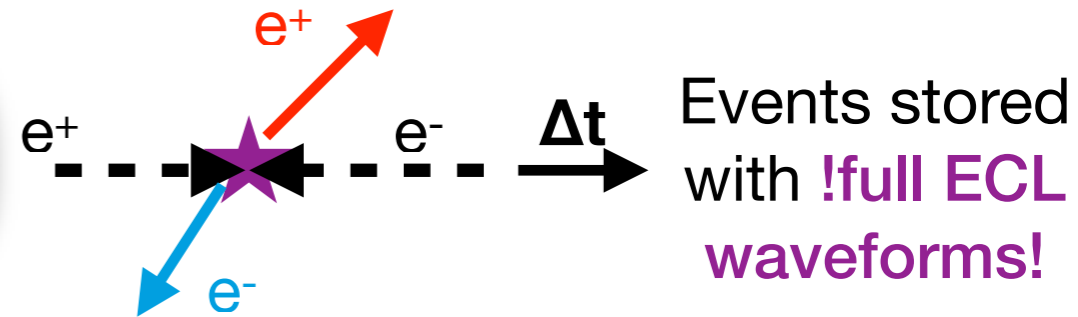
Events collected after a specific time from bhabha trigger



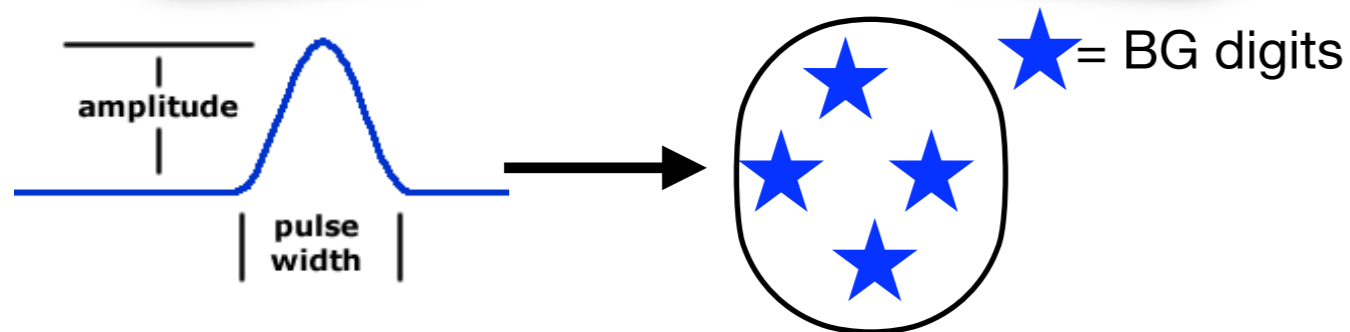
# MC run-dependent BGOOverlay files

Uses the random triggers events to extract the background information

Events collected after a specific time from bhabha trigger



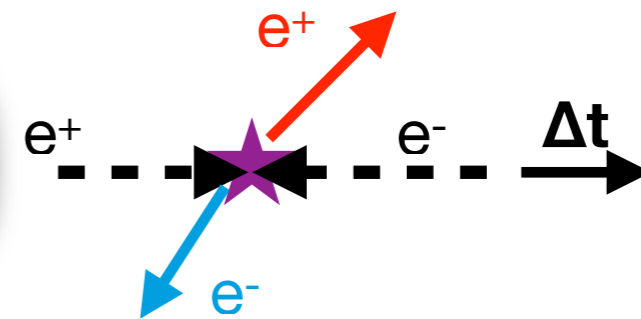
Waveform signals being unpacked to get the digit information



# MC run-dependent BGOOverlay files

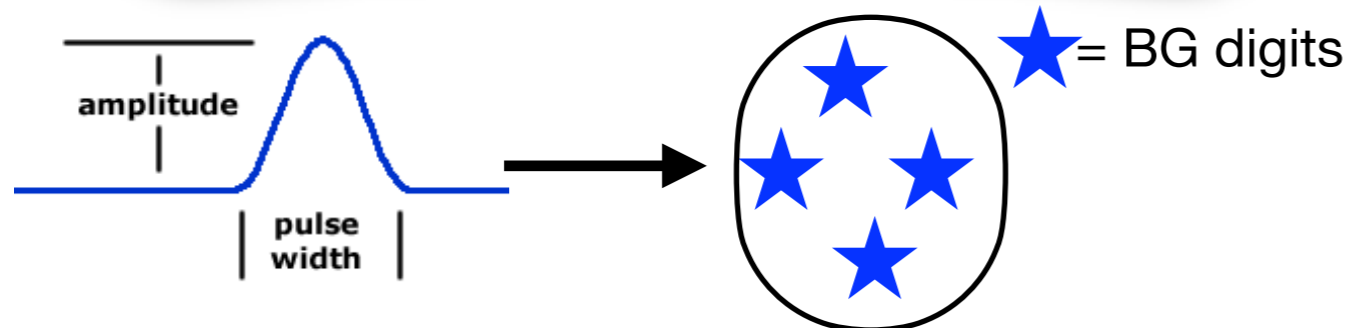
Uses the random triggers events to extract the background information

Events collected after a specific time from bhabha trigger



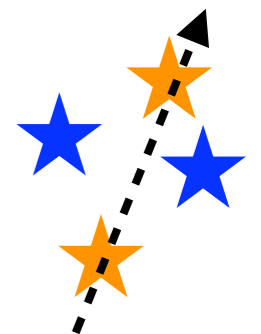
Events stored with **!full ECL waveforms!**

Waveform signals being unpacked to get the digit information



Digit information are overlaid to the MC events to be produced

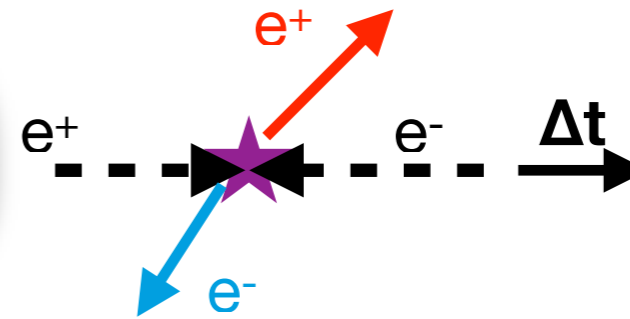
★ MC signal  
★ BGOOverlay



# MC run-dependent BGOOverlay files

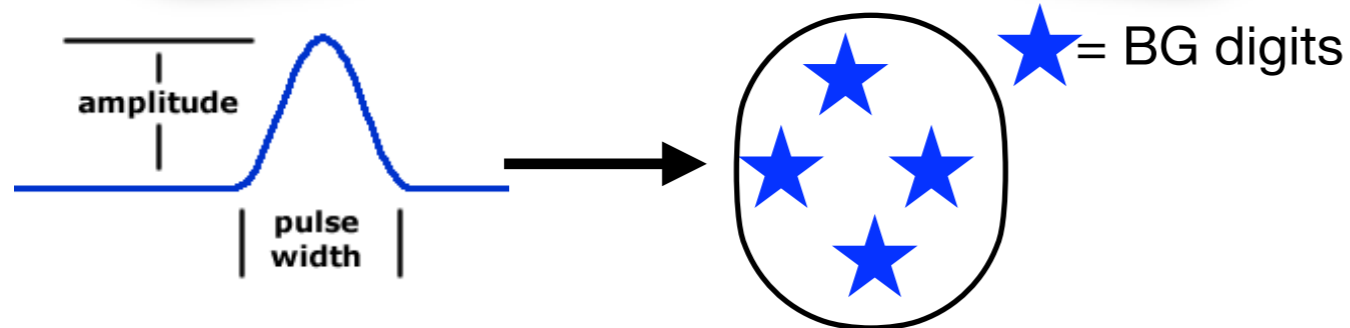
Uses the random triggers events to extract the background information

Events collected after a specific time from bhabha trigger



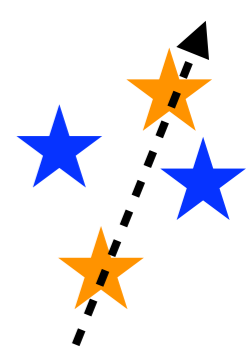
Events stored with **!full ECL waveforms!**

Waveform signals being unpacked to get the digit information



Digit information are overlaid to the MC events to be produced

★ MC signal  
★ BGOOverlay



MC samples with data-driven background overlaid

3





# Preparation of detector configuration

MCrd should have data-driven detector configurations



Manipulate the data calibration constants to provide MC payloads → gather together detector related quantities from each sub-detector



# Preparation of detector configuration

MCrd should have data-driven detector configurations



Manipulate the data calibration constants to provide MC payloads → gather together detector related quantities from each sub-detector



Alignment constants



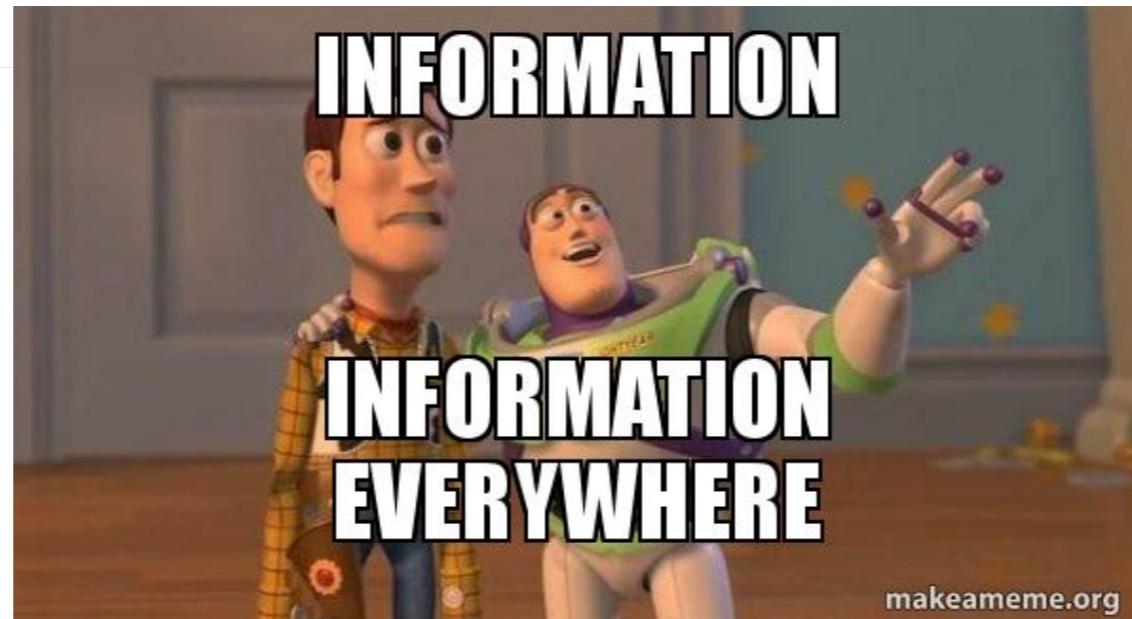
Dead channel mapping



Hit times



Trigger map



Detector geometry



PDFs of the energy depositions



# Timeline and complexity

The MCrD productions are very complex w.r.t. the run-independent version for which a simplified and constant detector configuration is used.

Specific signal samples from the entire collaboration → a lot of computing resources are needed to accommodate every request

Productions are submitted for specific data-taking periods → each data taking run corresponds to a job submitted to the grid



# Timeline and complexity

The MCrD productions are very complex w.r.t. the run-independent version for which a simplified and constant detector configuration is used.

Specific signal samples from the entire collaboration → a lot of computing resources are needed to accommodate every request

Productions are submitted for specific data-taking periods → each data taking run corresponds to a job submitted to the grid

Main samples are being produced by default  
+  
Additional samples can be requested by analysts



1 major MC processing per year



Each MC sample is ~4 times larger than the data-taking statistics  
This factor changes as a function of the nature of the samples



# Conclusions

---

- The Belle II experiment is collecting a huge amount of data
- Production of MC samples that well describe the data is a crucial aspect
- A run-dependent MC production system is in place and will be used for physics analysis
- MCrd productions are complex and computationally demanding
- Run-by-run evolving detector conditions → Belle II produced MCrd samples for each collected run

# Conclusions

- The Belle II experiment is collecting a huge amount of data
- Production of MC samples that well describe the data is a crucial aspect
- A run-dependent MC production system is in place and will be used for physics analysis
- MCrd productions are complex and computationally demanding
- Run-by-run evolving detector conditions → Belle II produced MCrd samples for each collected run

With  $\sim 400 \text{ fb}^{-1}$  of data collected as of today, we already produced MCrd samples corresponding to:

$\sim 1600 \text{ fb}^{-1}$  for the high multiplicity samples (largest size)

$\sim 80 \text{ fb}^{-1}$  of bhabha (smallest size)

*Thank  
you*



# Emergency slides!!

