July 15, 2019

Grid Computing and particle-ID performance study with untagged $\Lambda \rightarrow p^+\pi^-$

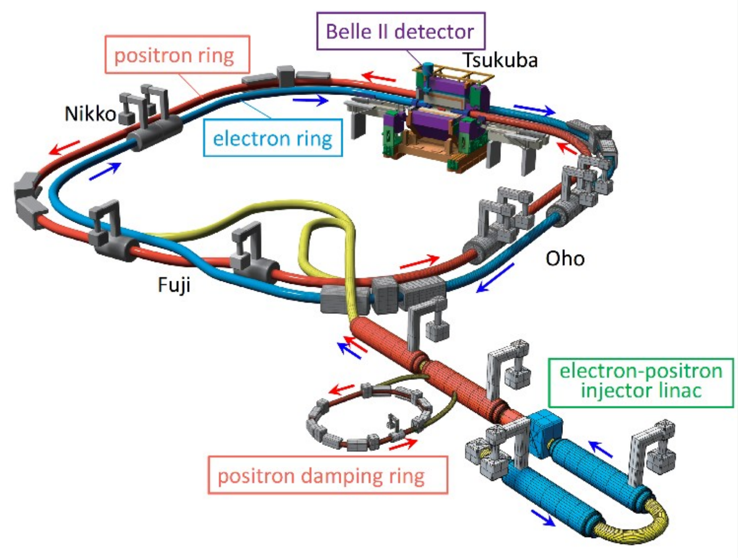
Presented by: Tobias Jenegger

Belle II experiment

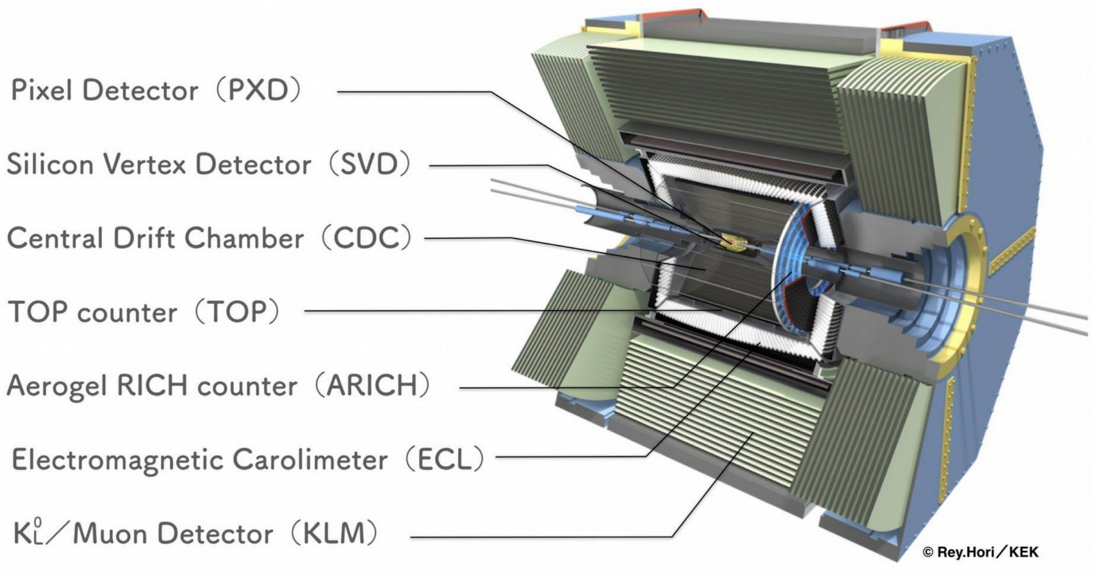
- Electron-positron collider, SuperKEKB in Japan
- Center-of-momentum energy (11 GeV) → production of B mesons
- Flavour physics, new physics → high precision needed

  → huge amount of data

SuperKEKB

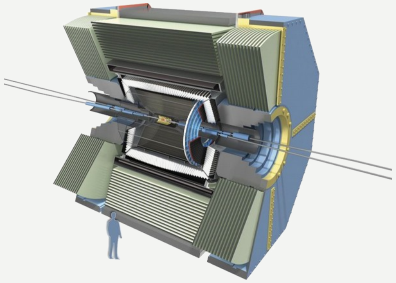


Belle II detector



1. Grid Computing

Experiment

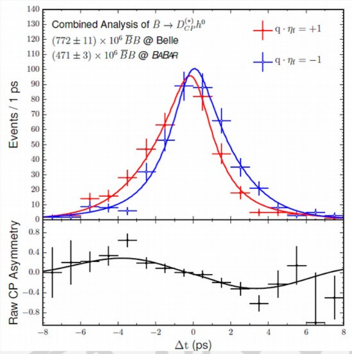


Belle II detector
(raw data from collisions)

Grid

heterogeneous resources
(cpu, memory, storage)

Physics analysis



Grid computing is the use of **widely distributed computer resources** to reach a common goal

Requirement: HW & SW infrastructure that provides dependable and consistent access to high end computational capabilities

Grid Software Infrastructure

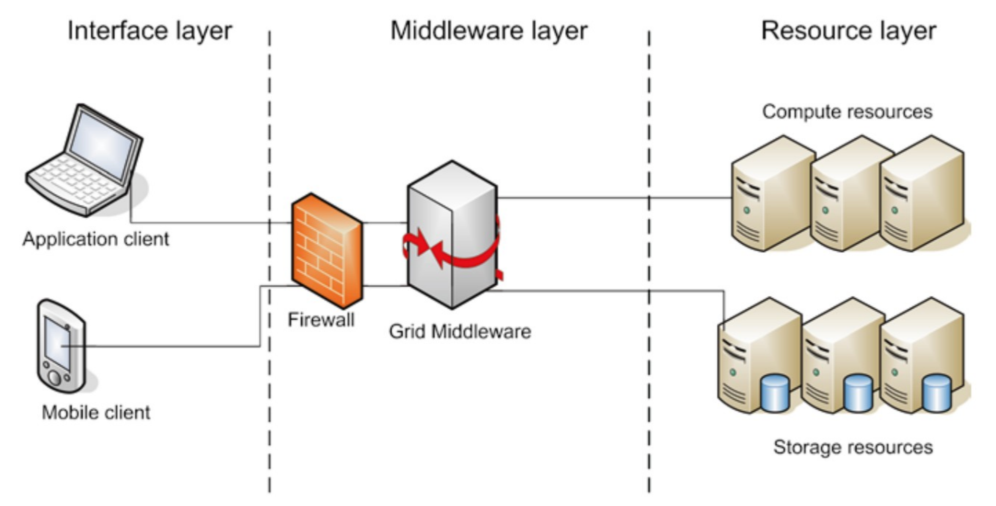
- Middleware: manages interconnection between resources and user clients
- Software framework that provides unified interface and access to distributed resources
- For Belle II: **DIRAC**

  →Developed at LHCb

  → Designed as modular system that can be extended

  ↓

  BelleDIRAC



Minimal BelleDIRAC Grid System (MBGS)

Motivation:

- Software under continuous development
- We have to test the new software
- Requirement: separation of production and development system
- Belle II computing group doesn’t provide grid environment for development

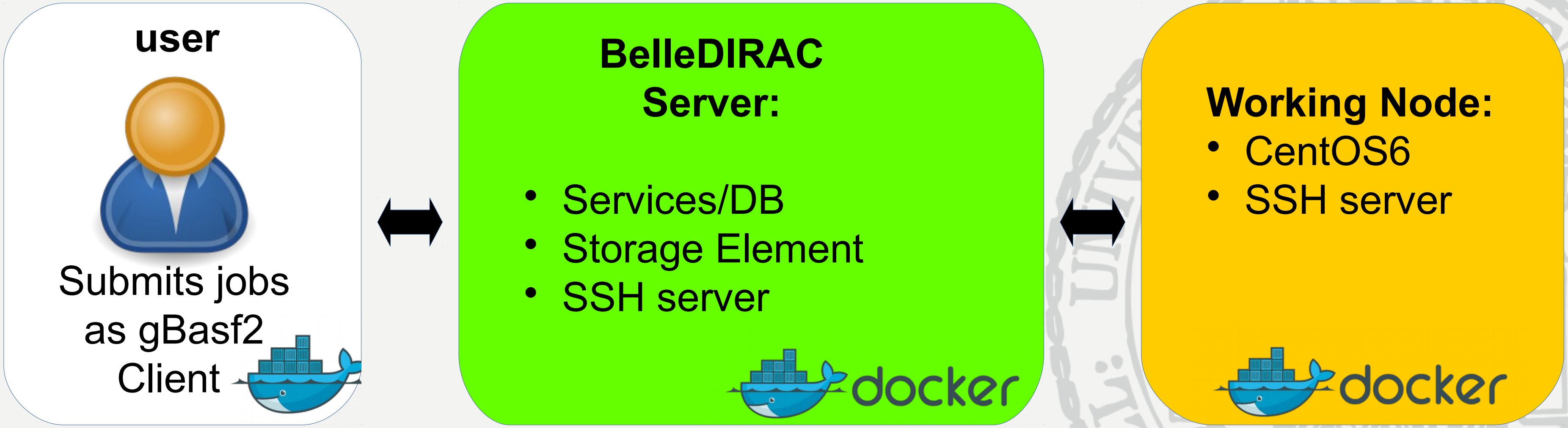
Solution:

→ **Minimal BelleDIRAC Grid System**

Minimal BelleDIRAC Grid System (MBGS)

My own developed minimal grid infrastructure based on docker containers

Advantages:
- It runs locally
- User cannot corrupt anything, it’s an isolated system
- No fear of break down of the large grid system



user

Submits jobs as gBasf2 Client

BelleDIRAC Server:
- Services/DB
- Storage Element
- SSH server

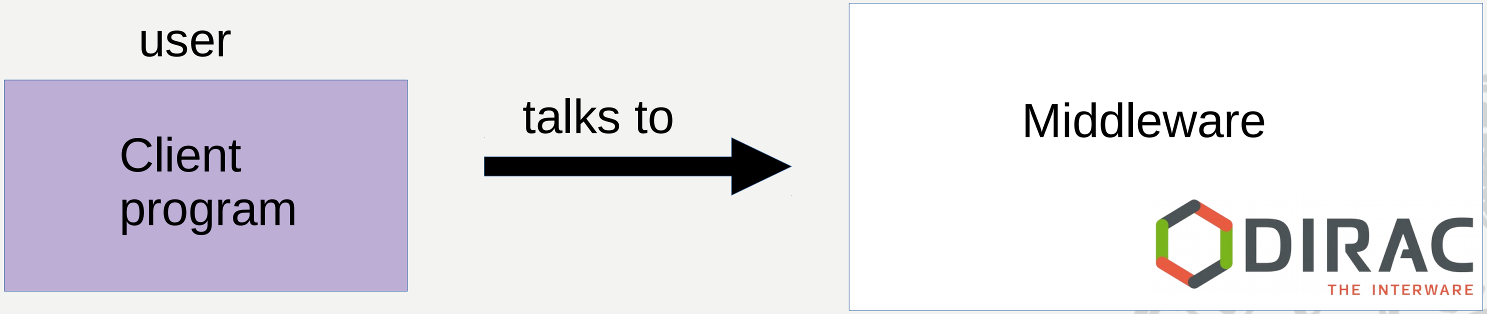
Working Node:
- CentOS6
- SSH server

Gbasf2ManagementSystem (GMS)

Motivation

For job submission:

- User needs to install a client program
- The client program demands special system environment
- Logic is on the user side



user

Client program

talks to

Middleware

DIRAC THE INTERWARE

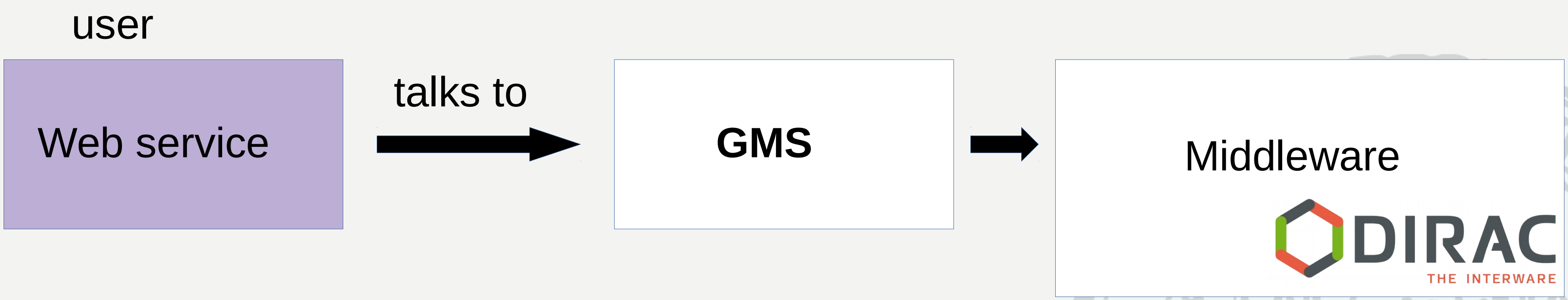
Solution

→ **Gbasf2ManagementSystem**

Gbasf2ManagementSystem (GMS)

My own developed BelleDIRAC service block

- Transfers the logic of the client program to server side
- Points towards **system independent job submission**



user

Web service

talks to

GMS

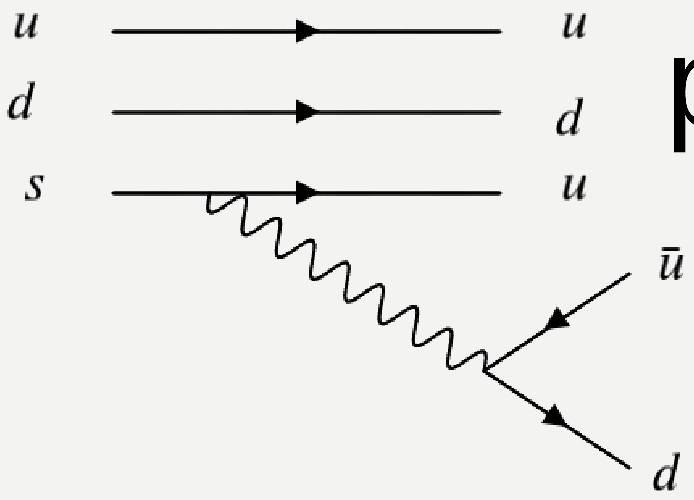
Middleware

DIRAC

THE INTERWARE

Note: GMS was tested on my own Minimal BelleDIRAC Grid System

2. Particle-ID performance study with untagged $\Lambda \rightarrow p^{+}\pi^{-}$



$\Lambda \left\{ \right.$

p^{+}

π^{-}

Particle Identification (PID)

- **Central role for particle detectors, especially for flavor physics:** Identification of charged particles stable enough to be detected

  → e, μ, π, K, p, d

- Proton identification probability = $\mathcal{L}_p/(\mathcal{L}_e + \mathcal{L}_\mu + \mathcal{L}_\pi + \mathcal{L}_K + \mathcal{L}_p + \mathcal{L}_d)$
  where $\mathcal{L}_i$ (i = e, π , K, p, d) is likelihood with the mass hypothesis of particle *i* from each sub-deterctors

- **Goal of the study: compare proton-ID performance on MC and data**

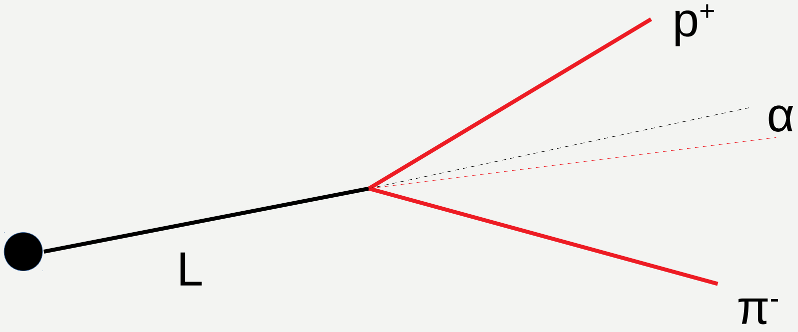
- Performance of PID selector characterized by **efficiency** and (a set) of **mis-identification probabilities**

$$\varepsilon = \frac{Number\ of\ protons\ with\ proton\ ID\ requirement}{Number\ of\ protons\ without\ proton\ ID\ requirement}$$

$$f_\pi = \frac{Number\ of\ \pi\ with\ proton\ ID\ requirement}{Number\ of\ \pi\ without\ proton\ ID\ requirement}$$

Λ selection criteria and binning

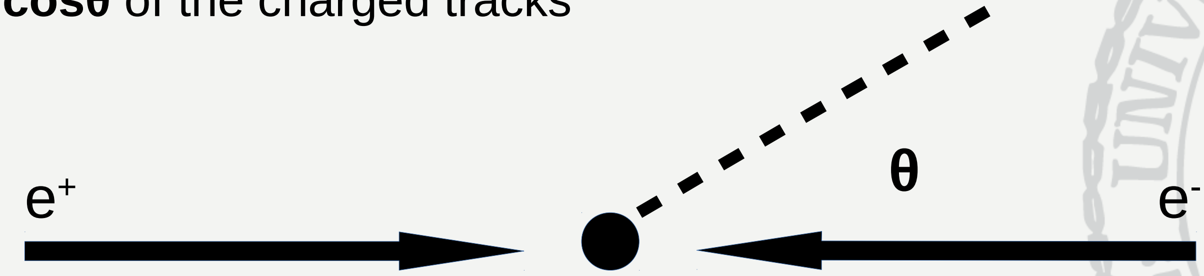
Λ selection based on kinematics and vertex information (after vertex fit):



Vertex information
→ $L > 0.5$ [cm]
→ $\cos\alpha > 0.995$

Binning:

PID performance depends on **momentum** and **cosθ** of the charged tracks



Point of interaction

Fitting Procedure

RooFit Toolkit for Data Modeling

Signal p.d.f.:

Gaussian + Crystal Ball function

Background p.d.f.:

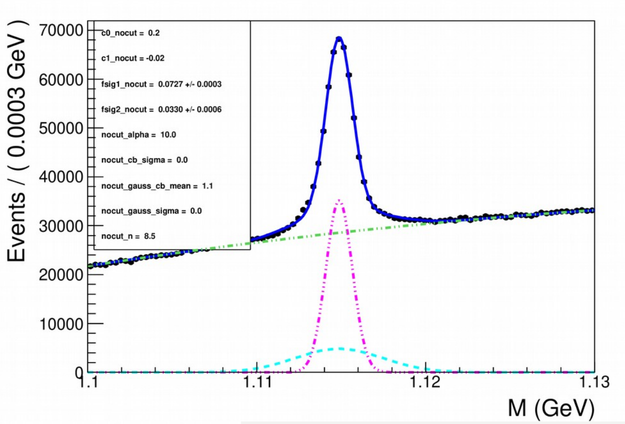
Chebychev polynomial of the first kind

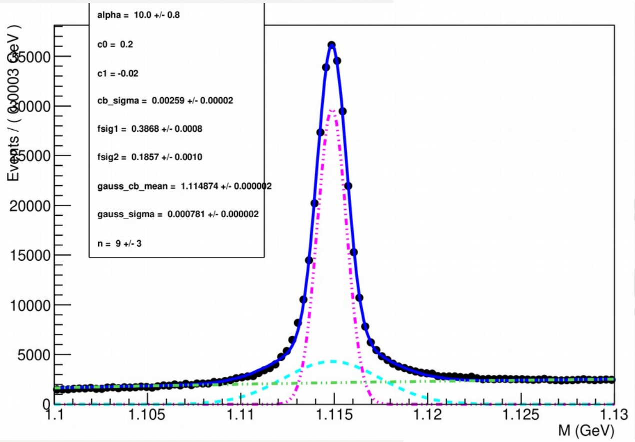
Fit $M_{p\pi}$ of MC/data sets with

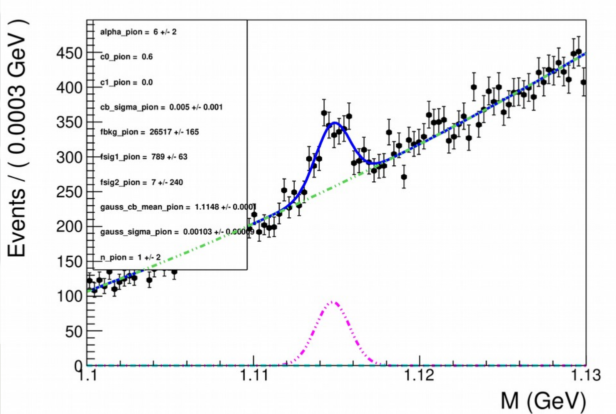
Signal and background p.d.f.:

$$efficiency = \frac{signal\ yield\ for\ protonID\, req.\, on\ protons}{signal\ yield\ without\ protonID\, req.}$$

$$pion\, fake\, rate = \frac{signal\ yield\ for\ protonID\, req.\, on\ pions}{signal\ yield\ without\ protonID\, req.}$$

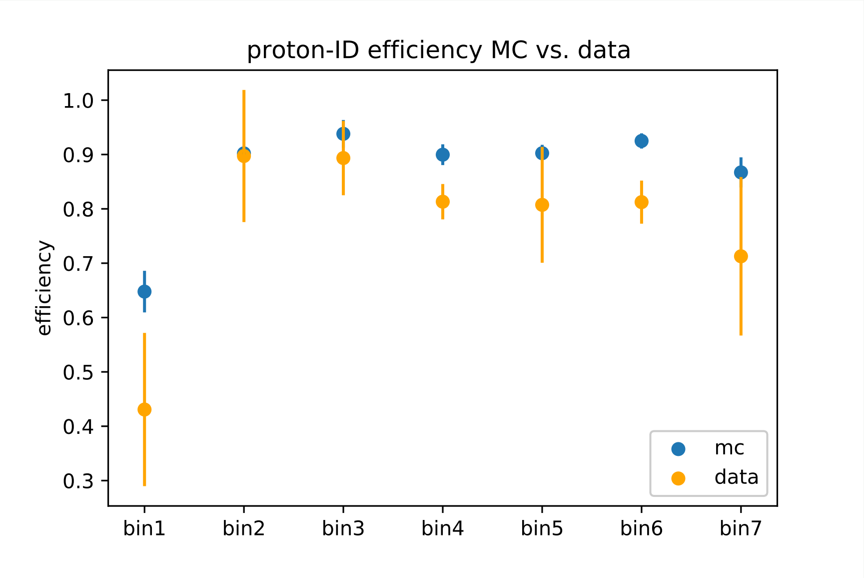


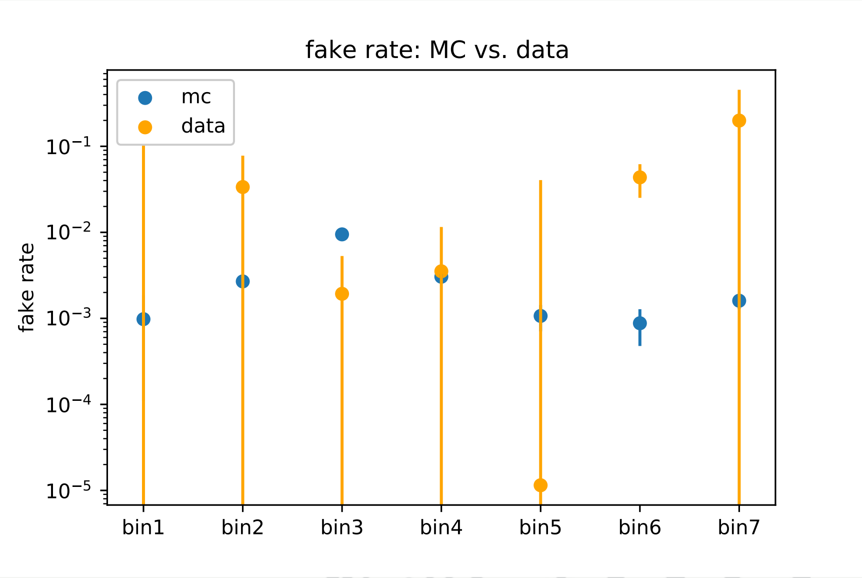




Results

Proton-ID efficiency vs. pion fake rate in **cosθ bins**:





Efficiency:

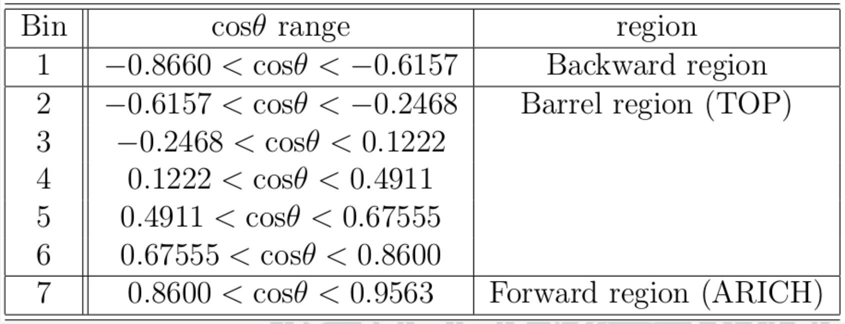
→ MC average: 0.87

→ data average: 0.77

Fake rate:

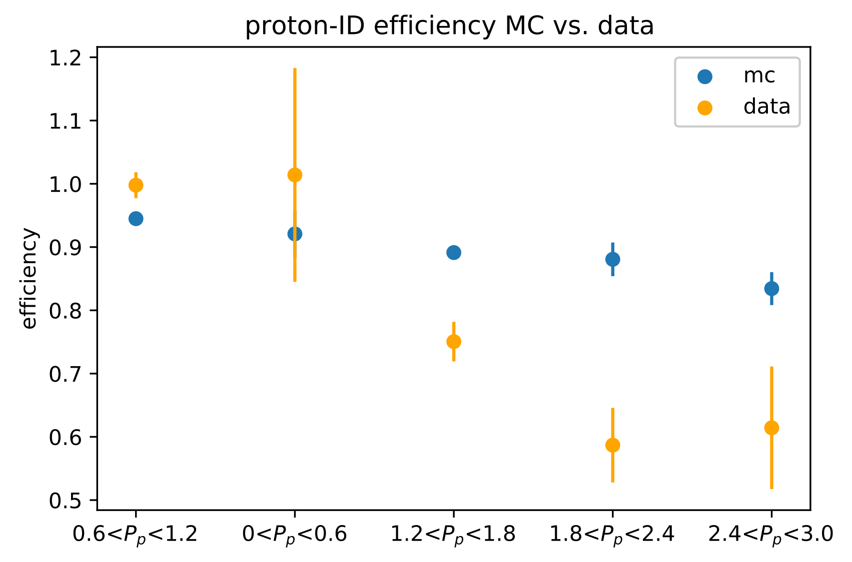
→ MC average: 0.0028

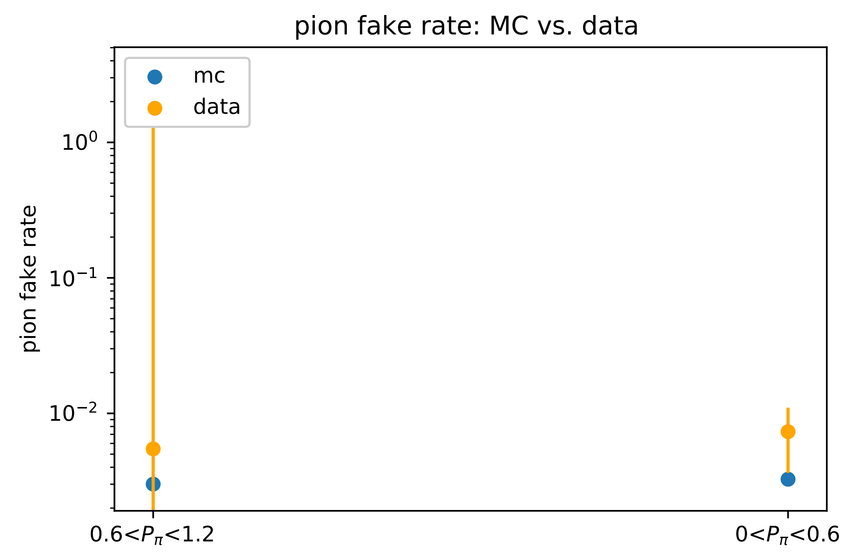
→ data average: 0.0661



Bin	$\cos\theta$ range	region
1	$-0.8660 < \cos\theta < -0.6157$	Backward region
2	$-0.6157 < \cos\theta < -0.2468$	Barrel region (TOP)
3	$-0.2468 < \cos\theta < 0.1222$	
4	$0.1222 < \cos\theta < 0.4911$	
5	$0.4911 < \cos\theta < 0.67555$	
6	$0.67555 < \cos\theta < 0.8600$	
7	$0.8600 < \cos\theta < 0.9563$	Forward region (ARICH)

Proton-ID efficiency vs. pion fake rate in **momentum bins**:





Efficiency:
→ mc average: 0.89
→ data average: 0.79

Fake rate:
→ mc average: 0.0031
→ data average: 0.0063

Summary and Outlook

- **Minimal BelleDIRAC Grid System**

  → isolated development environment

- **GBasf2ManagementSystem**

  →towards system independent job submission (Web application)

- **Particle-ID performance study**

  → further studies on sub-detector performances

  → explanation for data/MC discrepancies