用 Python 來玩粒子物理實驗吧

An Example of Using Convolutional Neural Network to Correct Gamma-ray Energy Measurement

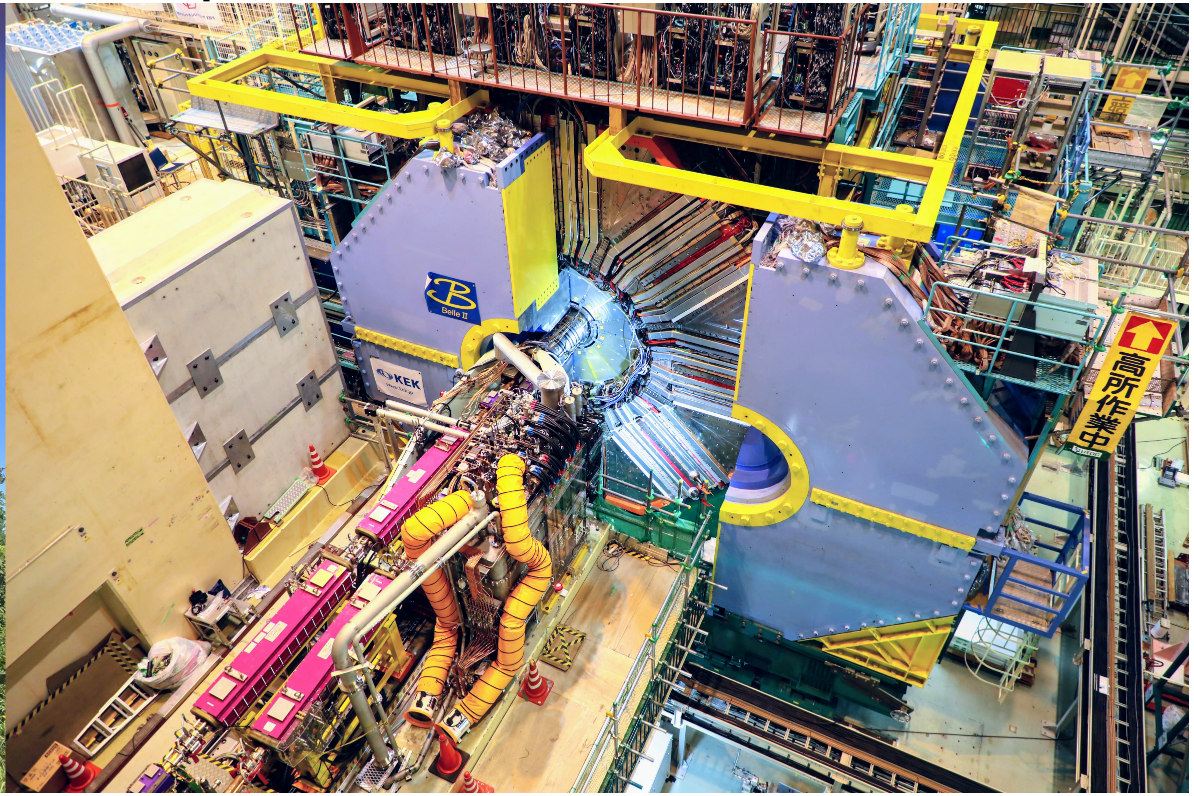
黃坤賢　(K. Huang)
台灣大學　(National Taiwan Univ.)
outreach @PyCon Taiwan 02, Jun,2018

About myself and research

Name: 黃坤賢 (Kunxian Huang)
Position: Postdoctoral Research at NTU
Research field: Particle physics experiment
Also a PYTHON-lover

About my research

What I do at postdoc………

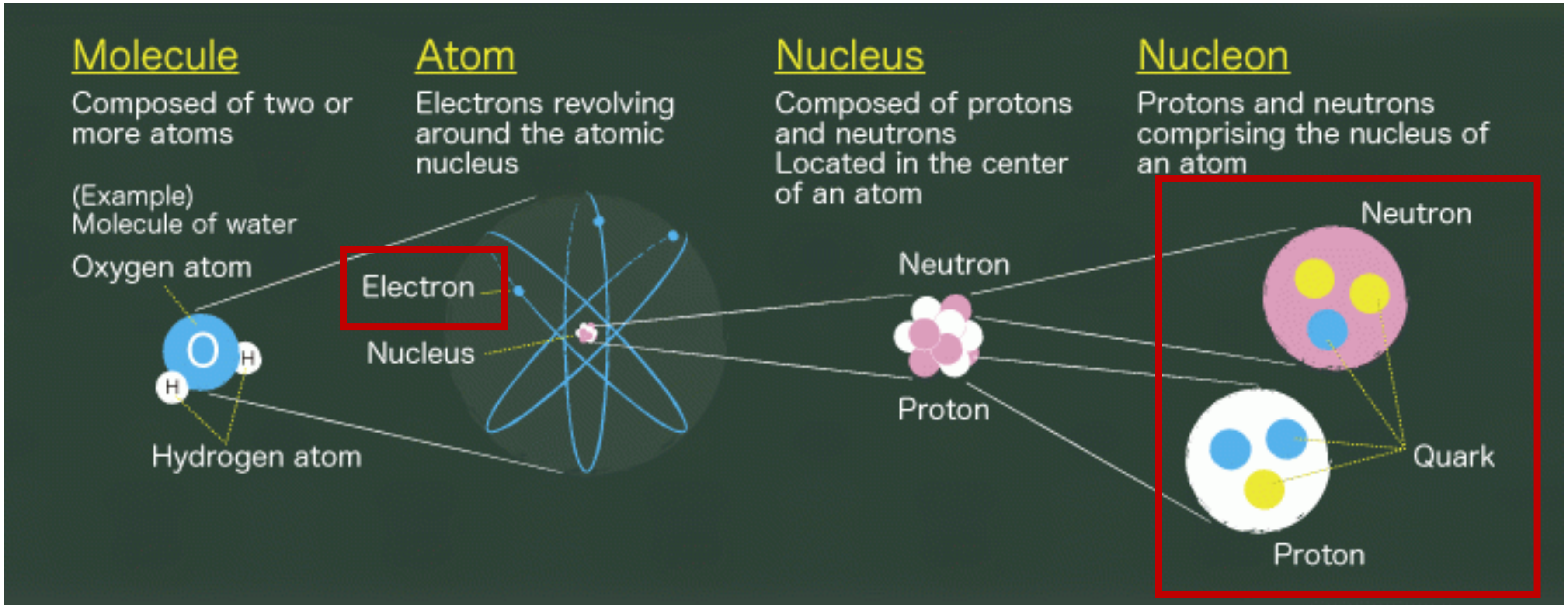


B meson experiment @Tsukuba, Japan: Belle II

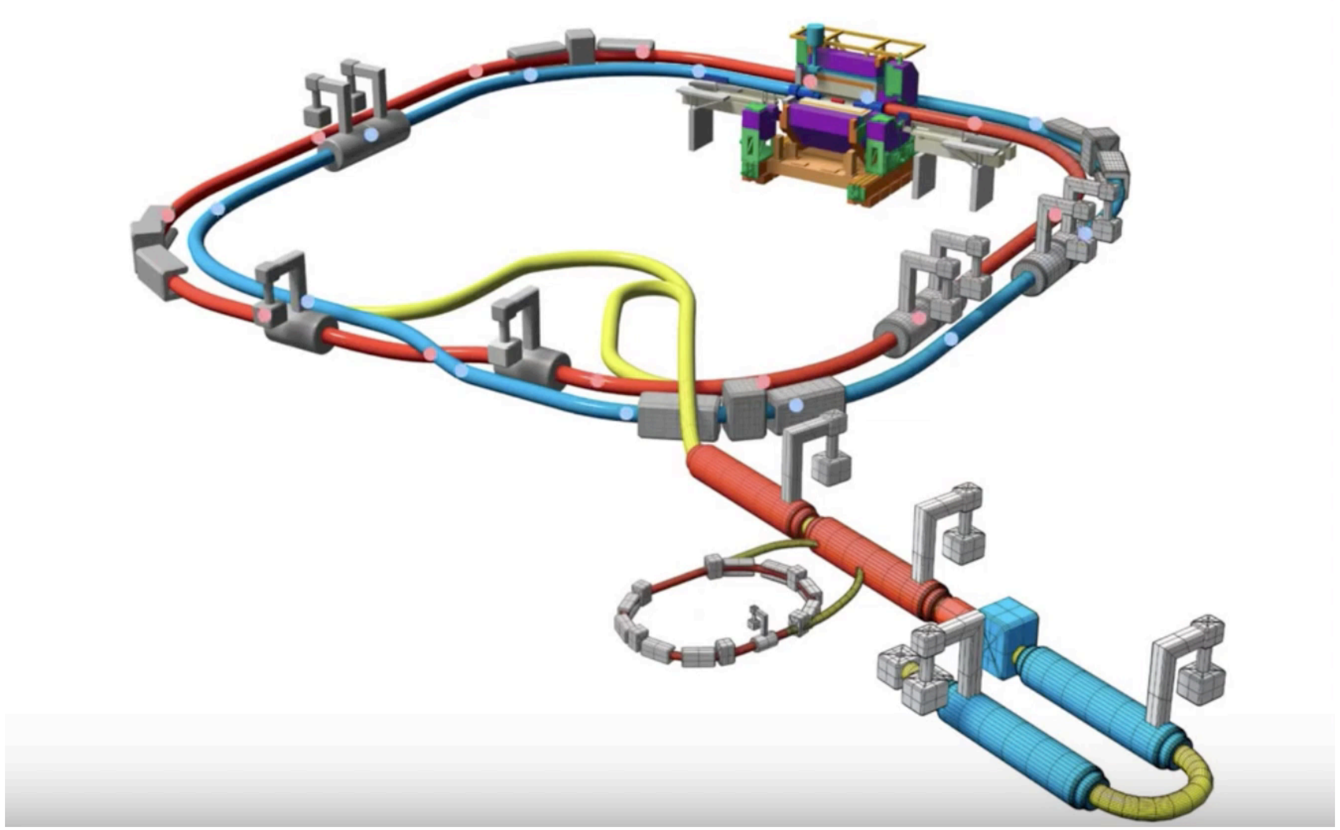
Using Python to <u>simulate, analyze, and storage for data, even distribute</u> data worldwide groups

What is "particle" physics

- Particle physics studies the tiniest object of Nature.
  → "Elementary particles" are smaller than proton or neutron.



Particle physics experiment (or called High energy physics exp.)



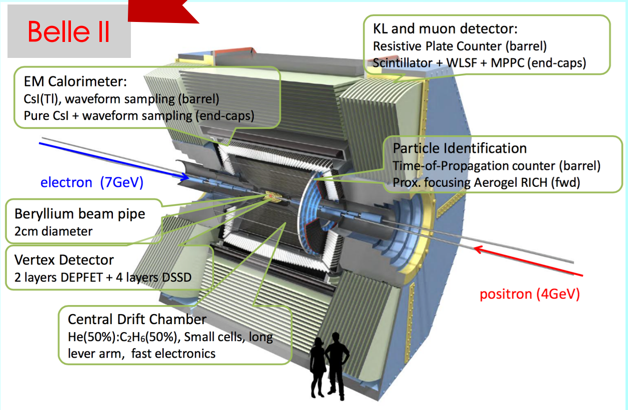
Accelerated particles are used to produce "unfamiliar" particles in our life.

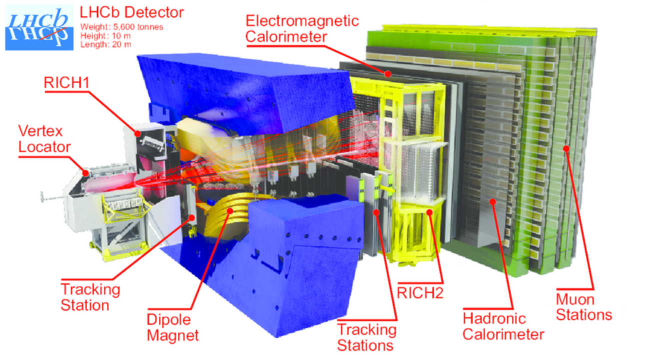
From KEK youtube channel
https://www.youtube.com/watch?v=cEnraaGAqwQ

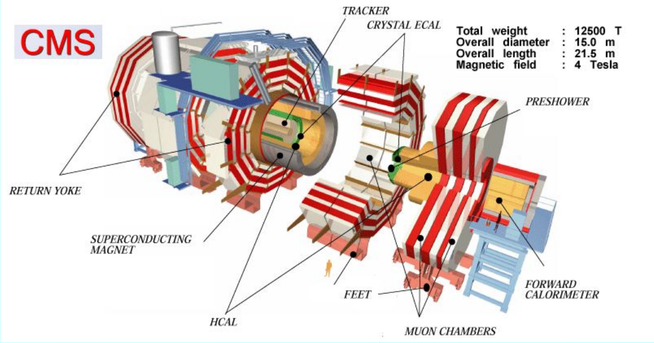
Large particle experiments worldwide

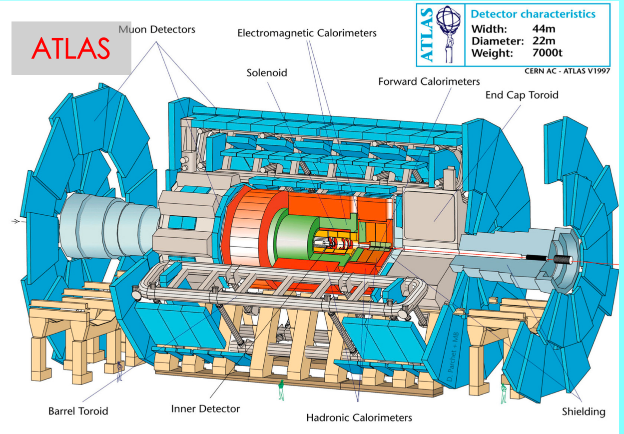
今天的主角

Collider exp.

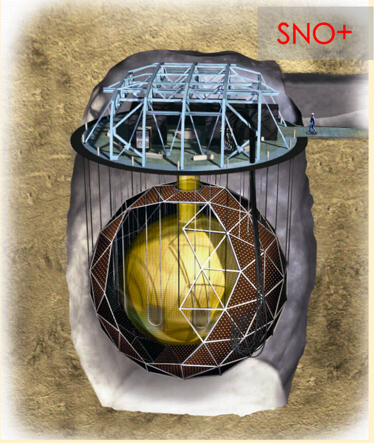


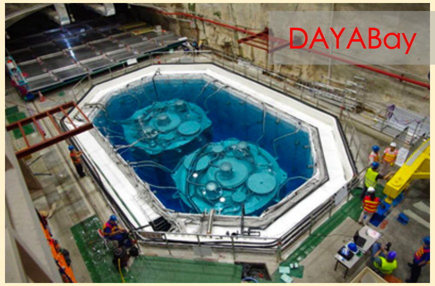




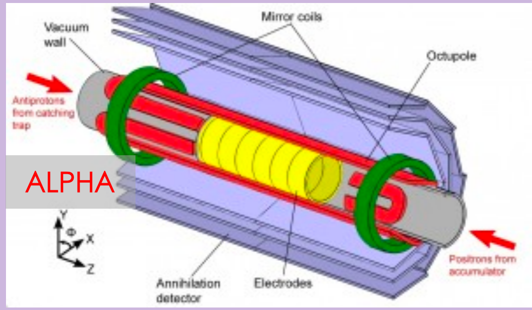


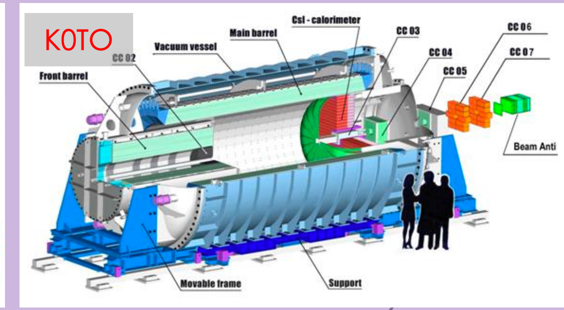
Neutrino exp.





Other exp.





“Why” to do Belle II experiment???

5 mins physics !!

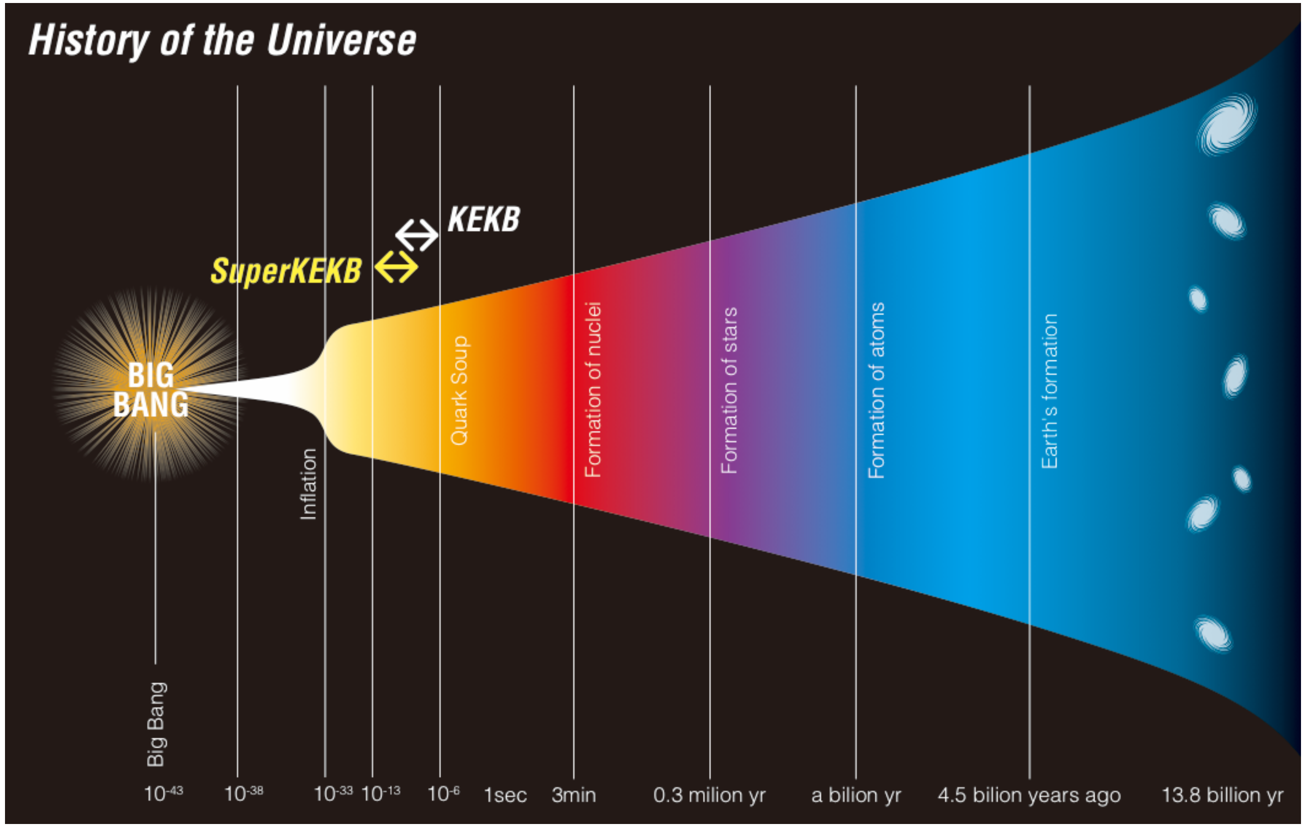
Introduction video

- Belle II detector observes B mesons produced by accelerator to measure CP violation and other new physics.
- Now, we can see matter in our Universe. Why?
  CP violation is the key.
- https://youtu.be/nGCrrgXSEOk?t=110

Universe after Big Bang

Hot Universe

Cold Universe



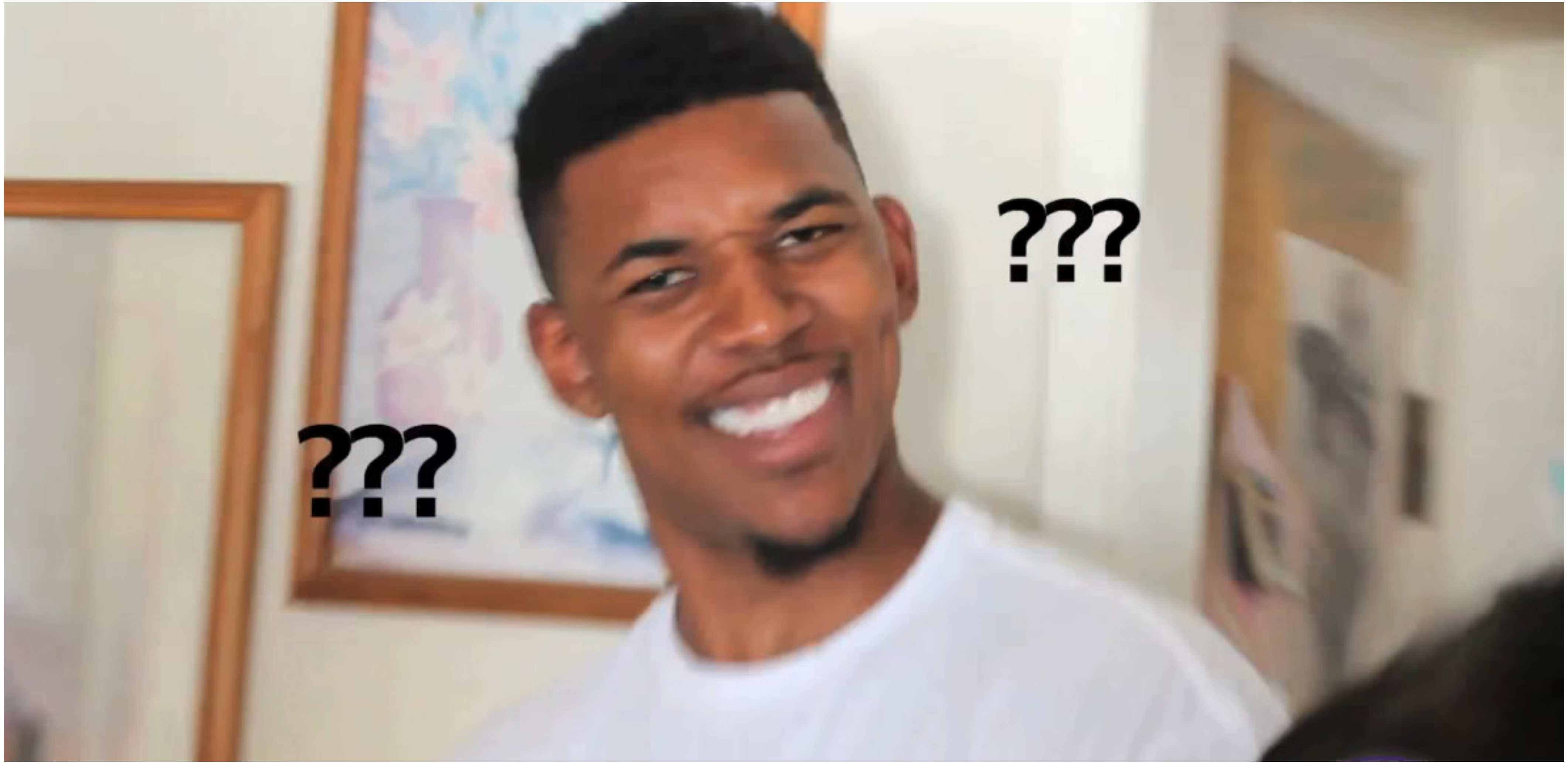
SuperKEKB accelerator produces “un-familiar” particles that are common existing right after Big Bang.

CP violation in B meson

- CP violation (電荷宇稱不守恆) first discovered in kaon meson decay by James Cronin, Val Fitch in 1964.
- M. Kobayashi and T. Maskawa introduced the mix matrix with three generation in standard model, and it also implies CP violation in the third quark decay in 1973.
- Sanda and Carter proposed CP violation in B meson decays in 1980.

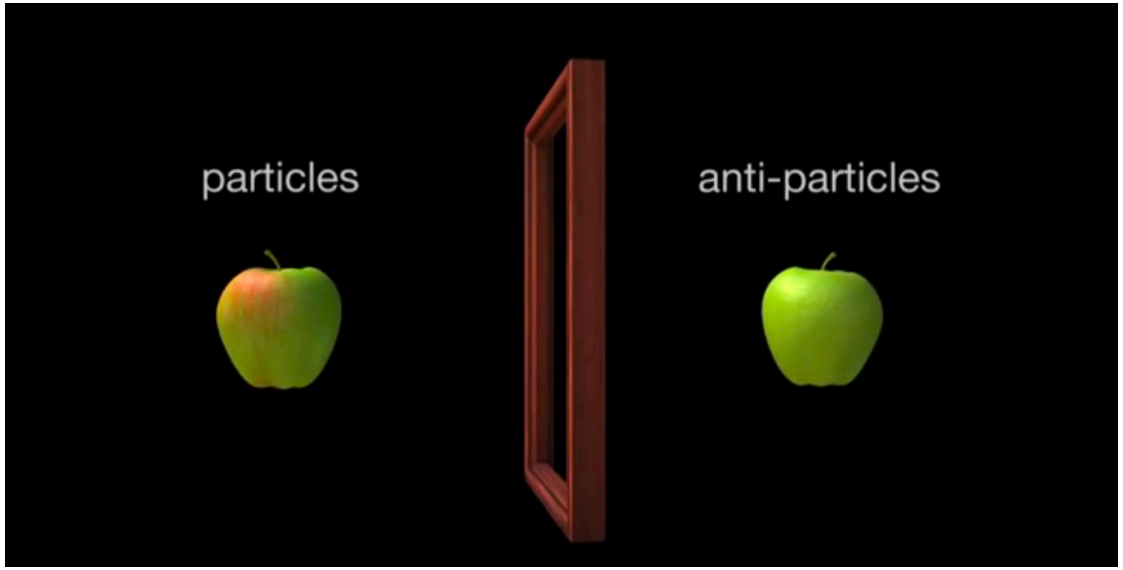
M. Kobayashi and T. Maskawa

CP violation



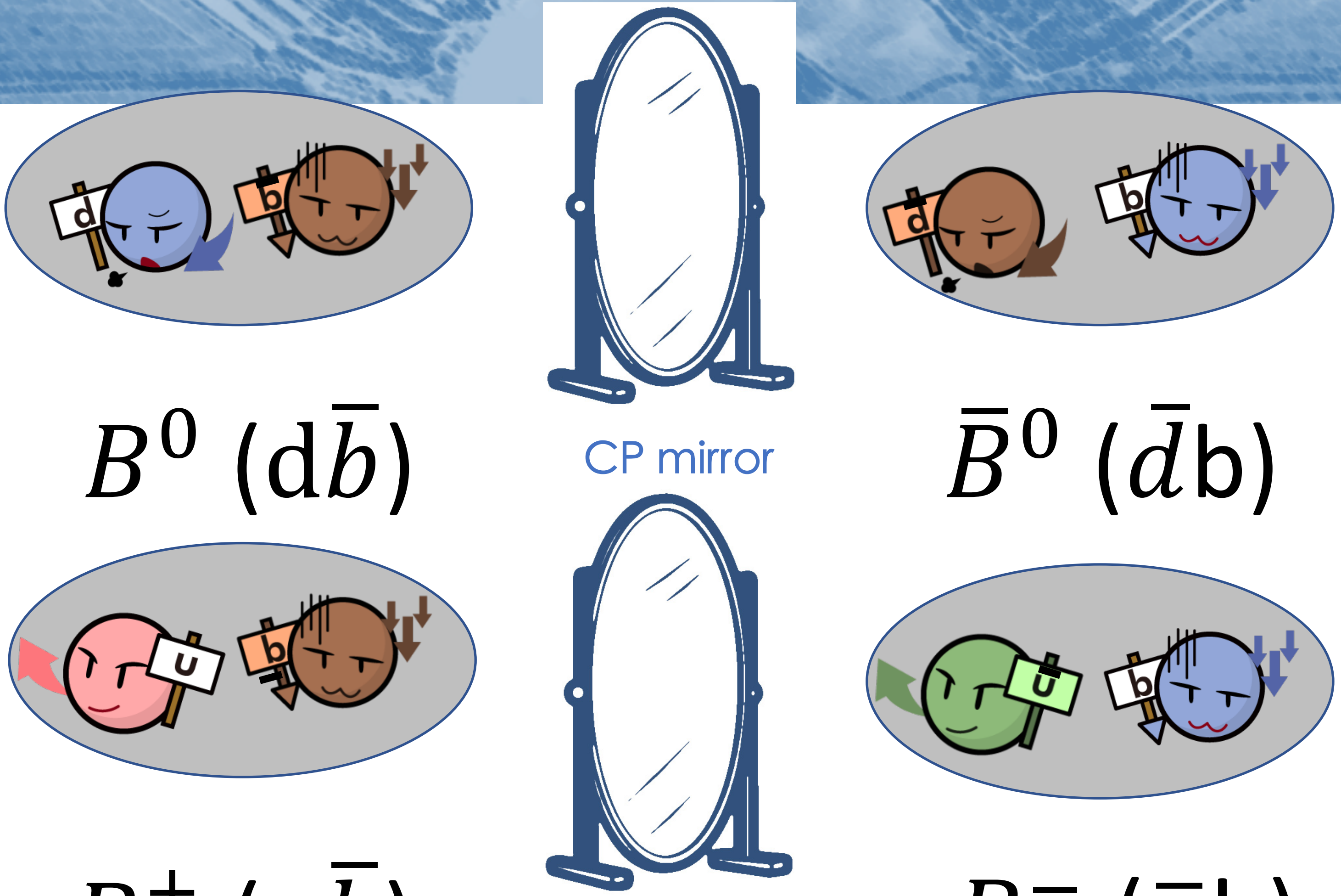
CP violation

- We believed that particle and anti-particle are the same (except charge and parity). But, why in our universe there are only matters (particles)??
- Because of charge-parity operator can be the particle-antiparticle conjugate operator, CP violation means difference of particle-antiparticle.
- Differences between matter-antimatter except C,P: life time, and decay branching ratio.



B mesons (B 介子)

From Higgs-Tan
http://higgstan.com/

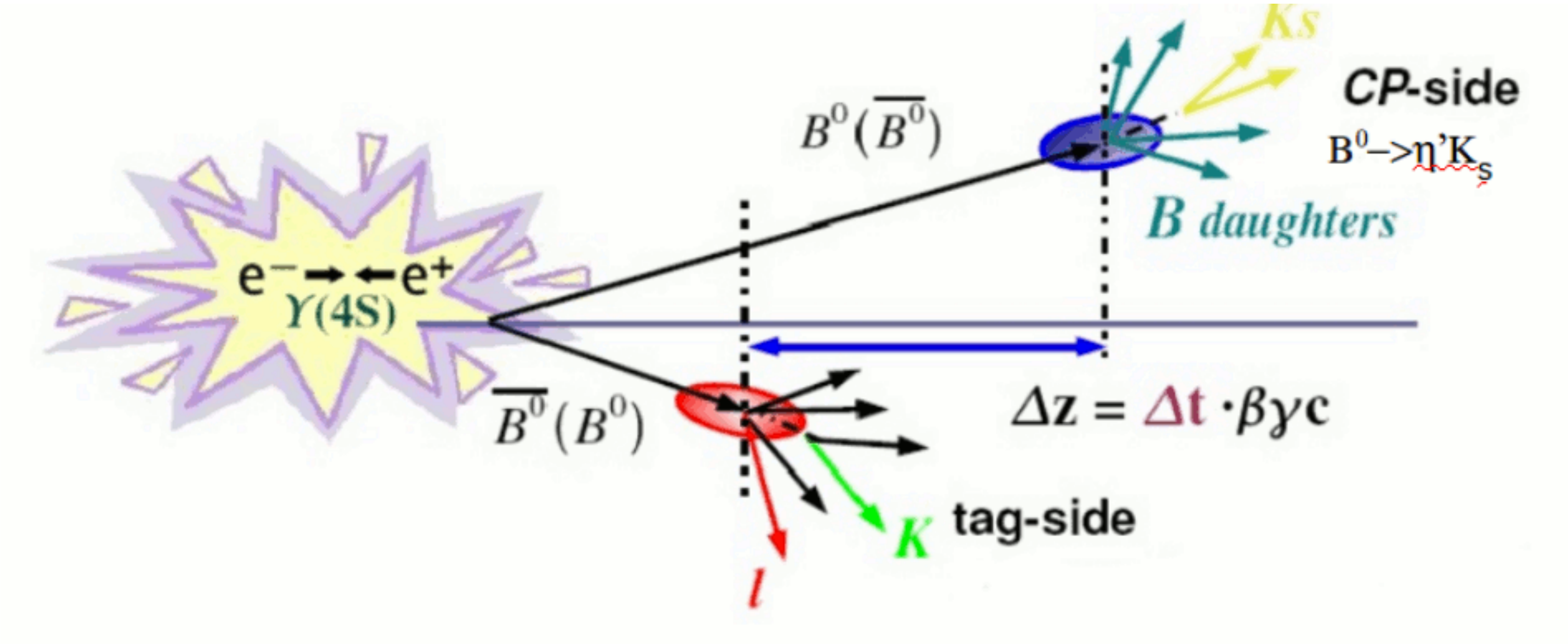


B^0 $(\mathrm{d}\bar{b})$ CP mirror $\bar{B}^0$ $(\bar{d}\mathrm{b})$

B^+ $(\mathrm{u}\bar{b})$ CP mirror B^- $(\bar{u}\mathrm{b})$

Particle: More common in Universe anti-Particle: rare in Universe

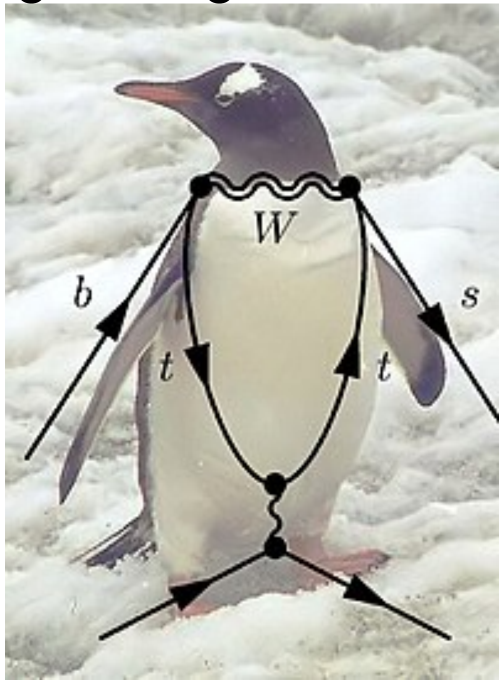
CP violation in B meson



- The longer life time , the farer they goes.
- B meson and anti-B meson have different behavior in some decay modes.

Other new physics in Belle II

From wiki page of
"Penguin diagram"

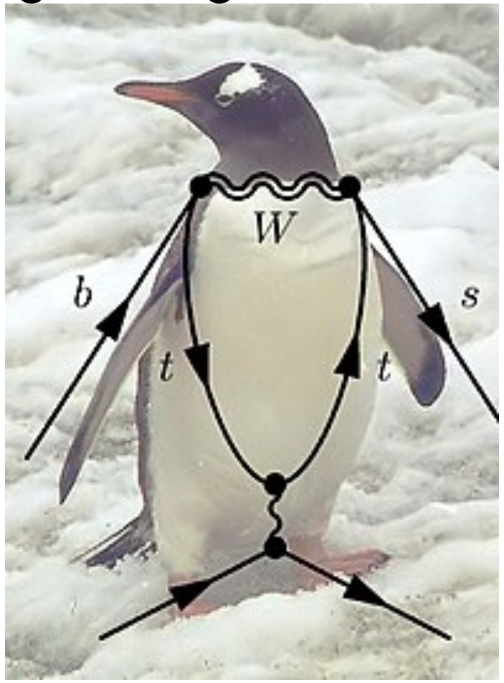


New intermedia particle
in penguin diagram decay

Dark matter search

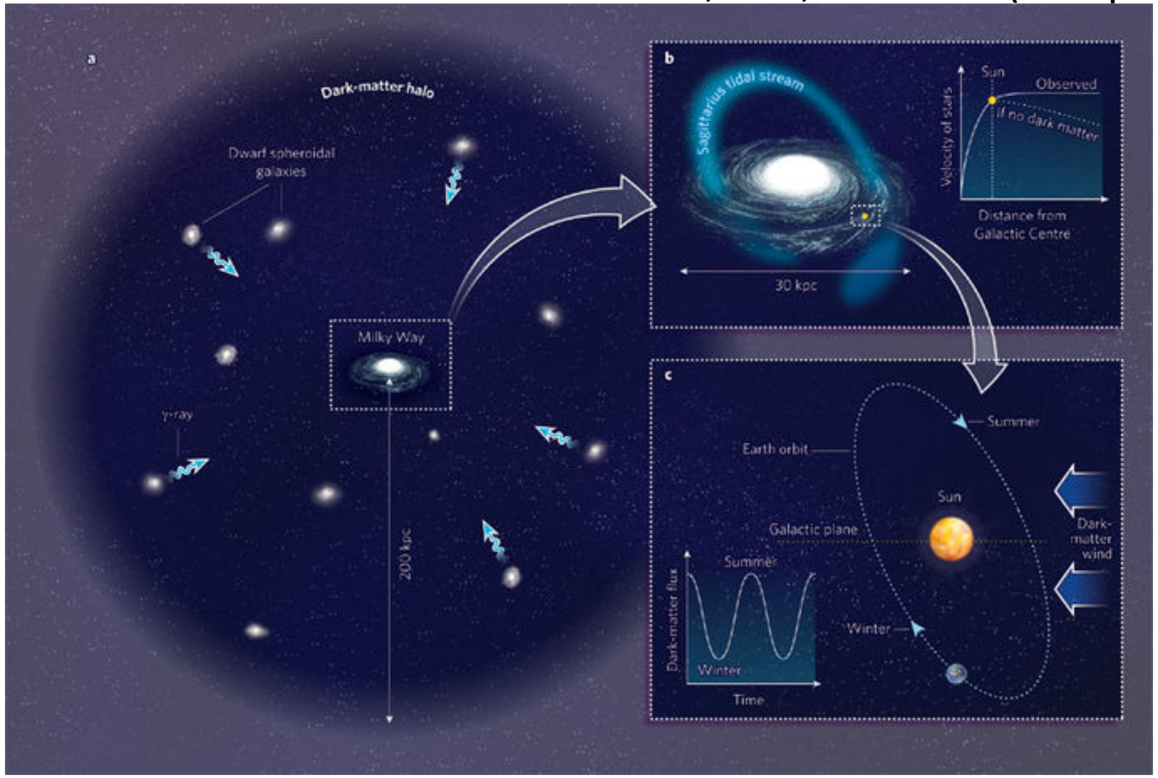
Other new physics in Belle II

From wiki page of
"Penguin diagram"



New intermedia particle
in penguin diagram decay

From Nature, **458**, 587–589 (02 April 2009)



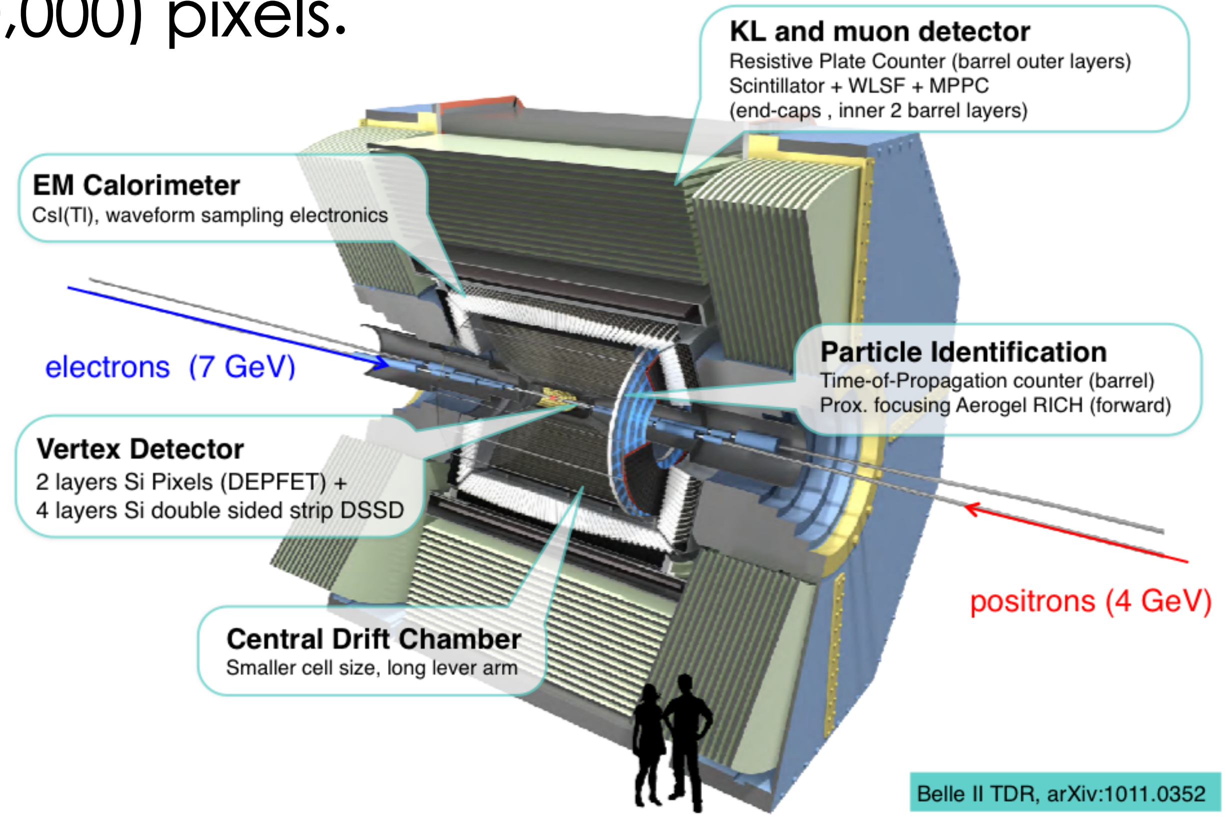
Dark matter search
~~(Gundam)~~

Physics explanation finished!!

Big Data in particle physics exp.

- Large detector use complex of sub-detectors to measure a single event. Parts of sub-detectors can detect specific particles.
- Recent years, each sub-detector are equipped with O(1,000~100,000) pixels. So, the complex detector have O(1,000,000) pixels.

- Big Data techniques are widely used in large particle experiments.
- Example: Belle II detector
- Data size: 1MB (after event builder by firmware)
- Trigger rate (event rate): max 30 kHz.
- Data rate: max 30 GB/sec



Belle II TDR, arXiv:1011.0352

Big Data Science

Belle II

Google internet archive ~1EB

Belle II data+ simulation 30~50 PB

LHC raw data 50 PB

LHC Science data ~200 PB

Google Search ~98PB

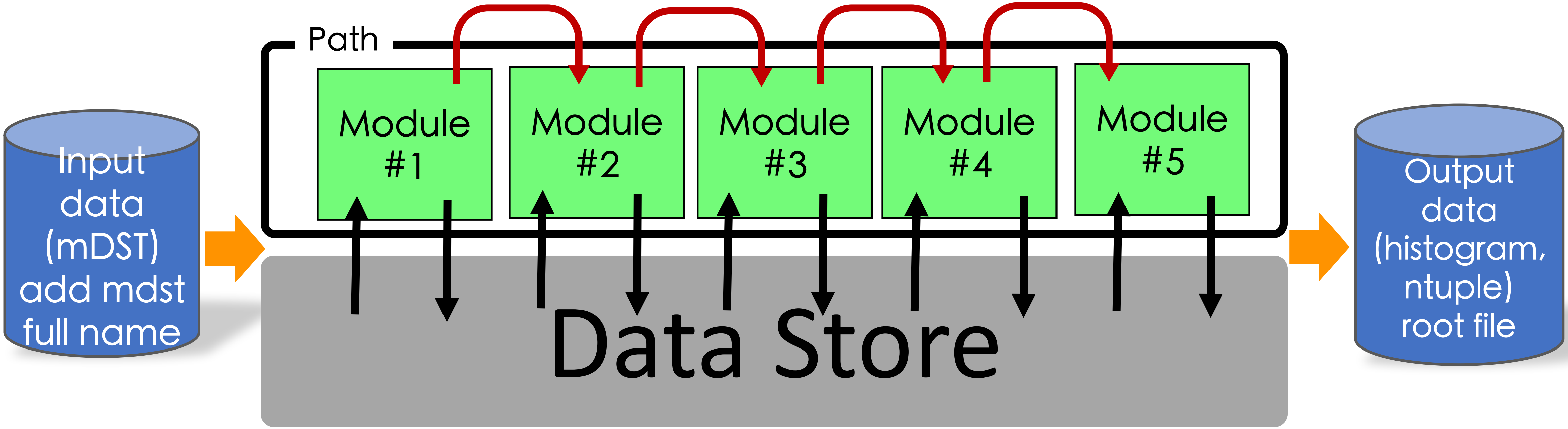
Facebook upload ~180PB

PB=10^{15}Byte
EB=10^{18}Byte

Information from
Elizabeth Sexton-Kennedy (Fermi National Accelerator Lab. (US))
https://indico.cern.ch/event/658060/contributions/2844782/

Structure of Belle II data analysis

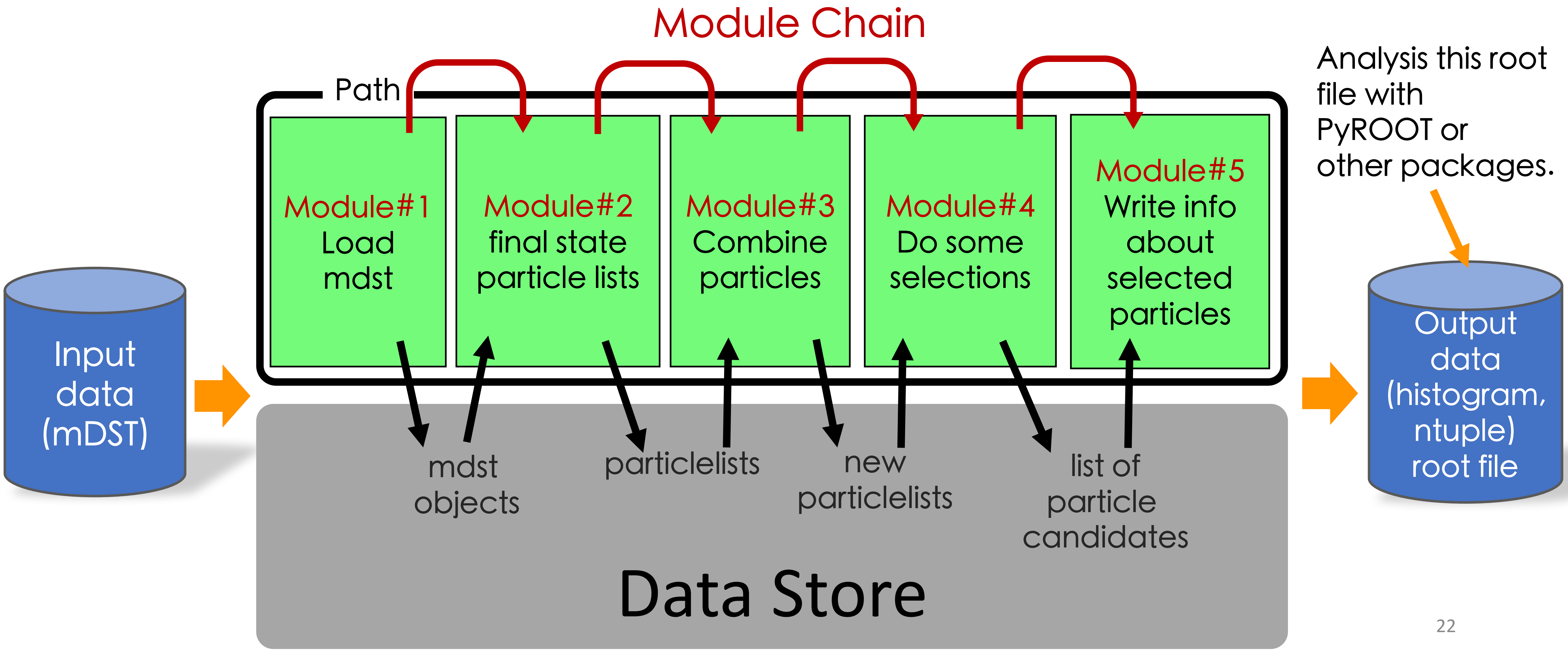
Module Chain



C++14:
- Module
- code to build particles from primitive objects (like tracks and calorimeter clusters)

Python3.6:
- Load and configure C++ modules
- Other high-level analysis tasks.

Example of B meson candidates reconstruction



Numerical Analysis with Python at Belle II

- "ROOT" file format is usually for data storage and analysis format in HEP.
- Website: https://root.cern.ch/

ROOT is ...

A modular scientific software framework. It provides all the functionalities needed to deal with big data processing, statistical analysis, visualisation and storage. It is mainly written in C++ but integrated with other languages such as Python and R.

ROOT
Data Analysis Framework

Installation instruction: (https://root.cern.ch/building-root)

- use of CMake cross-platform build-generator tool to build and install
- download source code: https://root.cern.ch/get-root-sources
- Don't forget to turn on cmake option "python" to install PyROOT

PyROOT
python binding for ROOT

root_numpy
interface between ROOT and NumPy

root_pandas
loading/saving ROOT files as pandas DataFrames

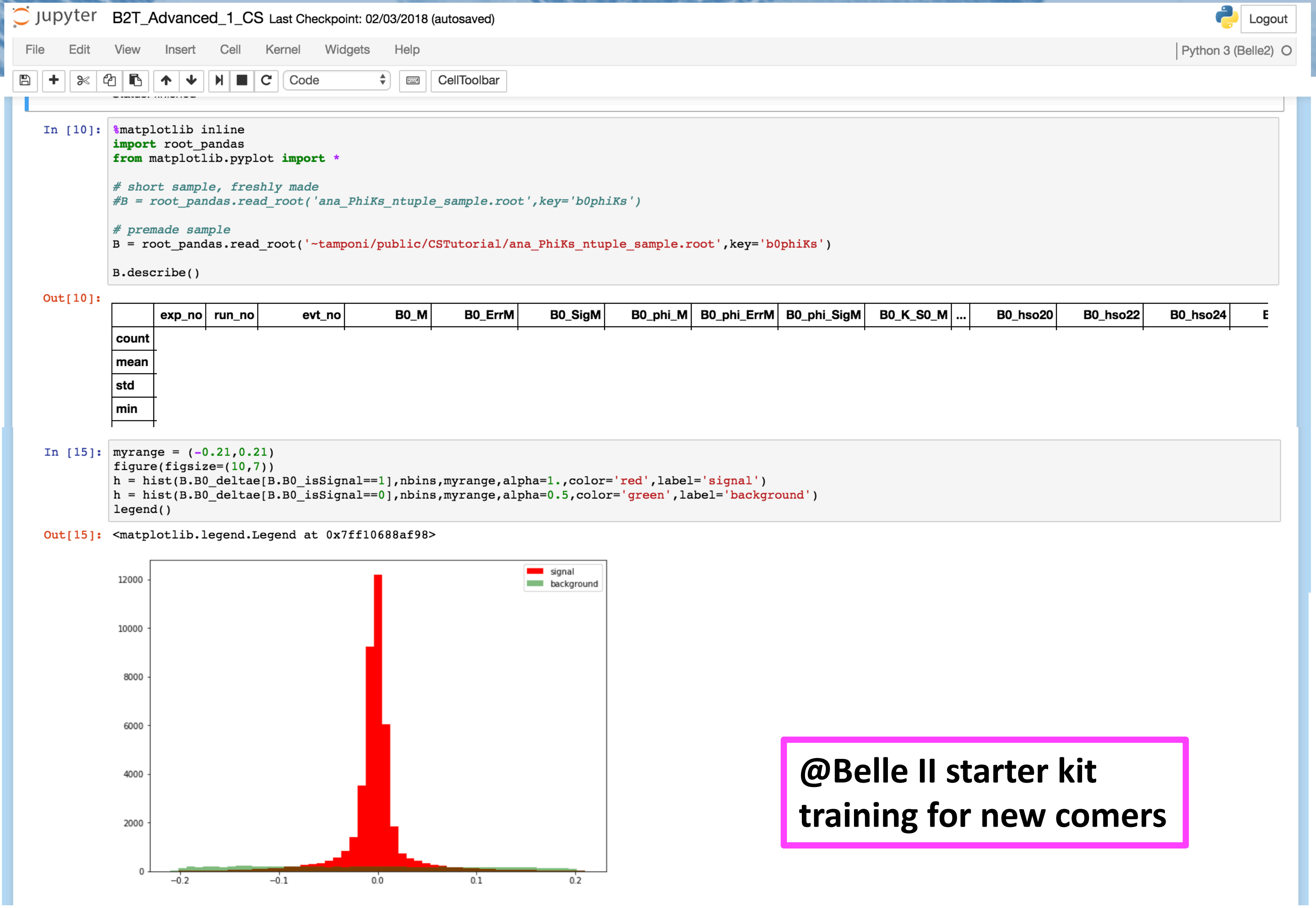
ROOT
written in C++

Installing PyROOT, root_numpy, root_pandas

- Pre-requires: ROOT, Numpy, Pandas
- PyROOT
  - website: https://root.cern.ch/pyroot
  - enable option of "python" when installing ROOT using cmake.
- root_numpy
  - website: http://scikit-hep.org/root_numpy/index.html
  - installation:
    - git clone git://github.com/rootpy/root_numpy.git
    - python setup.py install --user
- root_pandas
  - website: https://github.com/scikit-hep/root_pandas
  - installation:
    - git clone https://github.com/scikit-hep/root_pandas.git
    - python setup.py install --user

Training course with Jupyter Notebook

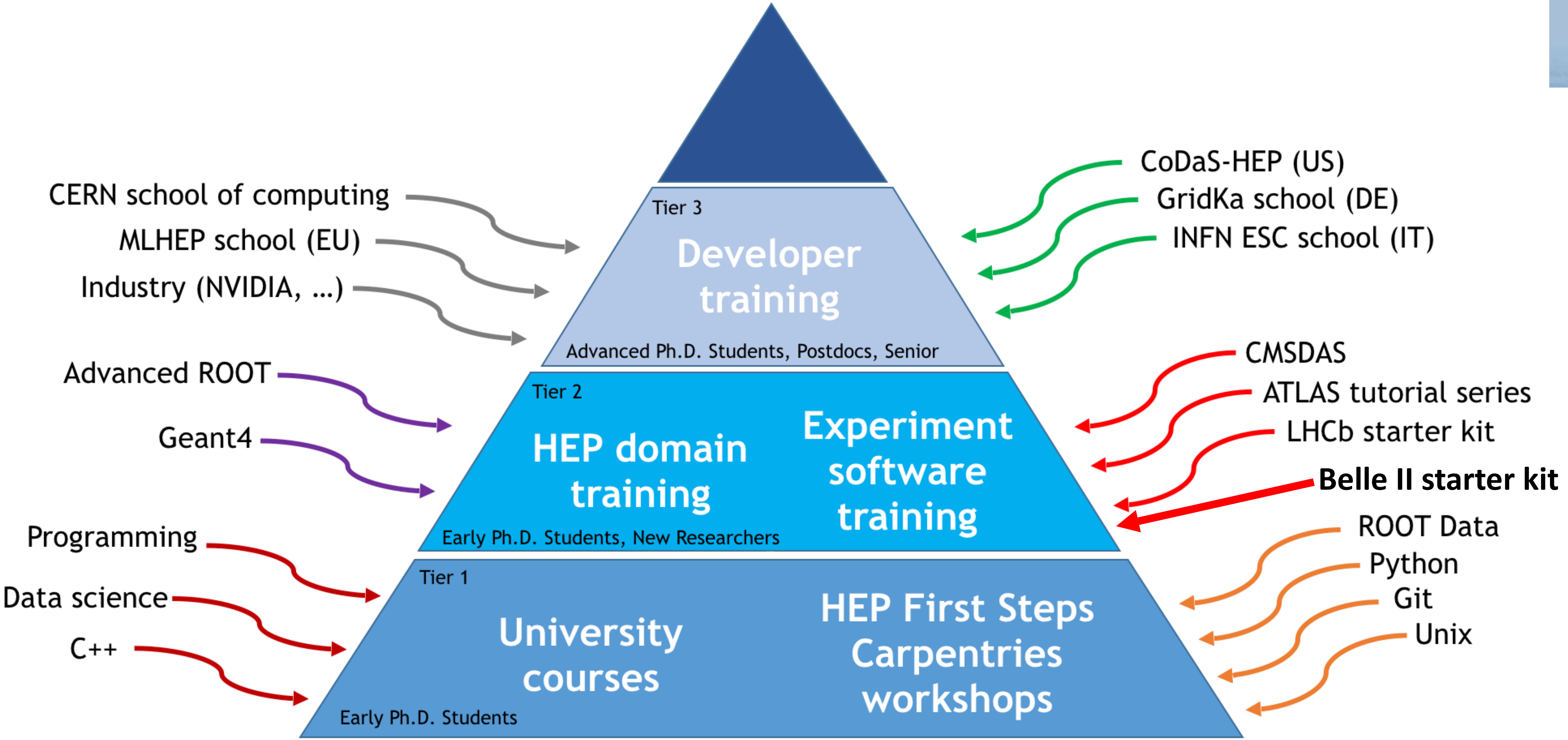
- For new collaborators, Belle II hold a start kit workshop for 3 days per 4 month.
- Jupyter Notebooks on internal server.
- Self-sustained: combine documentation + code
- Aim for large tutor/student ratio.



@Belle II starter kit training for new comers

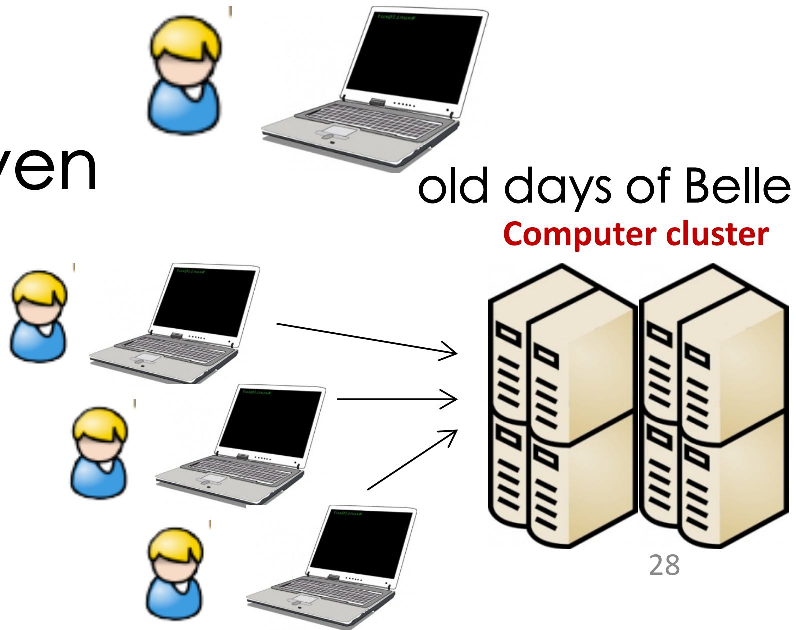
HEP Software Training

From Peter Elmer
@WLCG and HSF workshop 2018



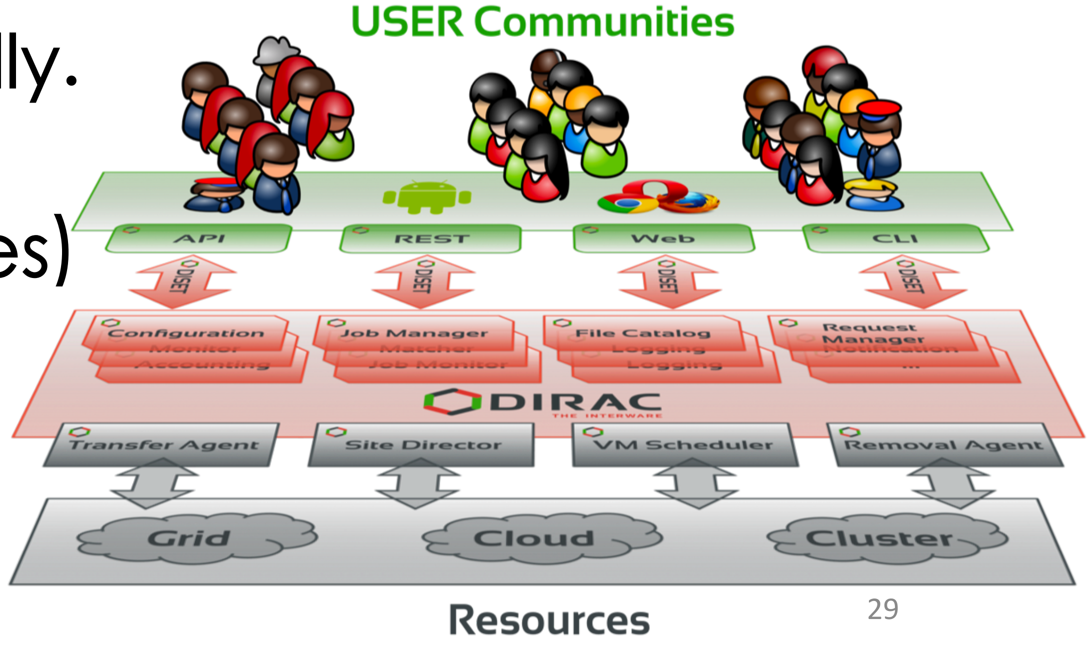
Needs of Grid Computing

- In the age of Belle, KEK constructed central computing cluster for the computing resource.
- Belle II: 50 times luminosity, higher background rate, 1 MB size per event.
- Bigger data, need more CPUs & memories.
  →dramatically larger computing resources are needed.
- Belle II collaboration: 25 countries,
  ~100 institutes world wide.
- Each institute contribute their cluster, grid, or even buy cloud service from company.
- Concept of distributed resource usage:
  - Send Job (to central service)
  - "Heterogeneous" resource distributed over the world.
  - Data distributed over the world.



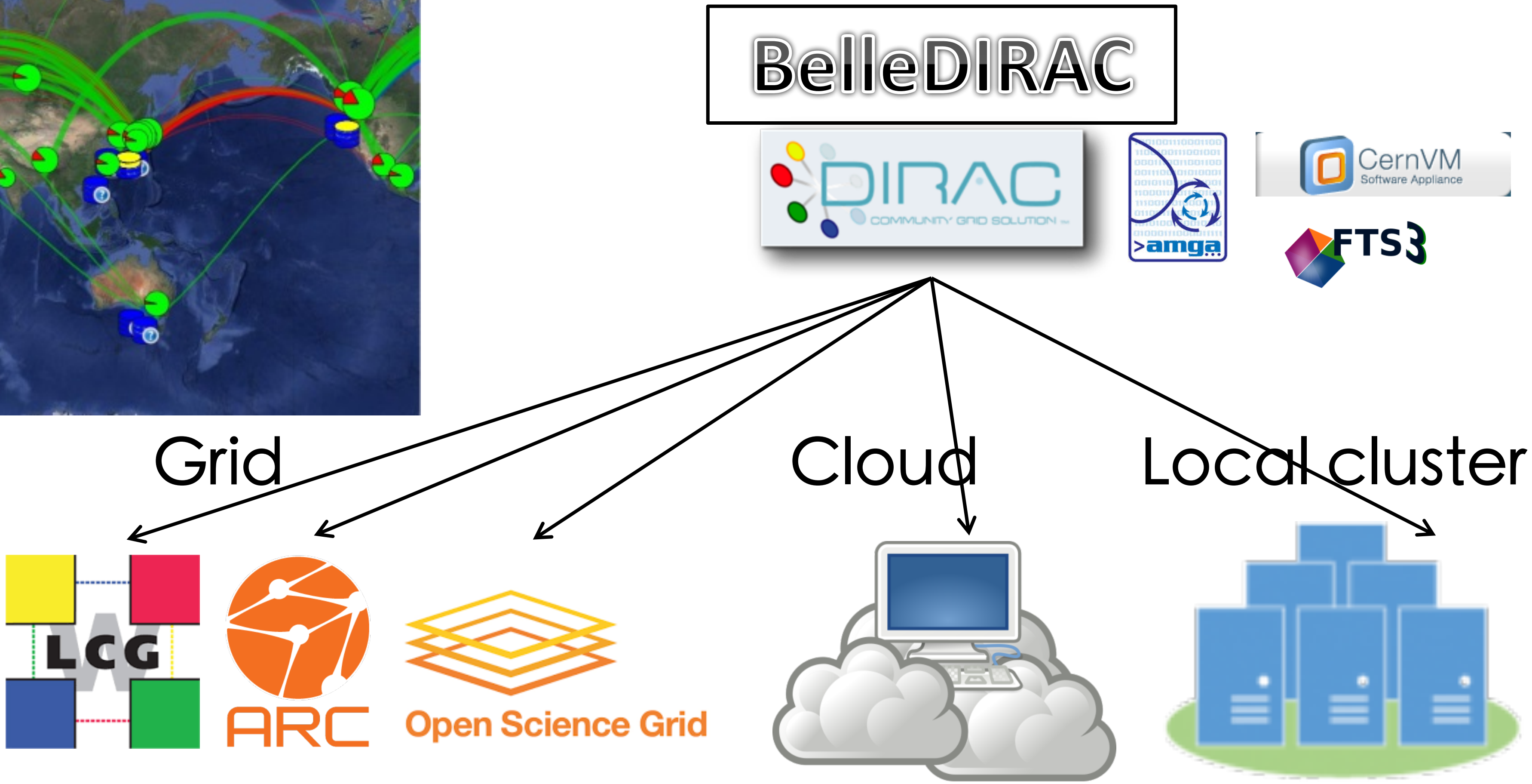
Distributed GRID computing

- "DIRAC" (Distributed Infrastructure with Remote Agent Control)
- A software framework for distributed computing providing a complete solution to one (or more) user community requiring access to distributed resources.
- Developed by LHCb experiment initially. Then, CMS, ATLAS, and Belle II joined.
- Source code @GITHUB: (PYTHON codes)
  https://github.com/DIRACGrid/DIRAC



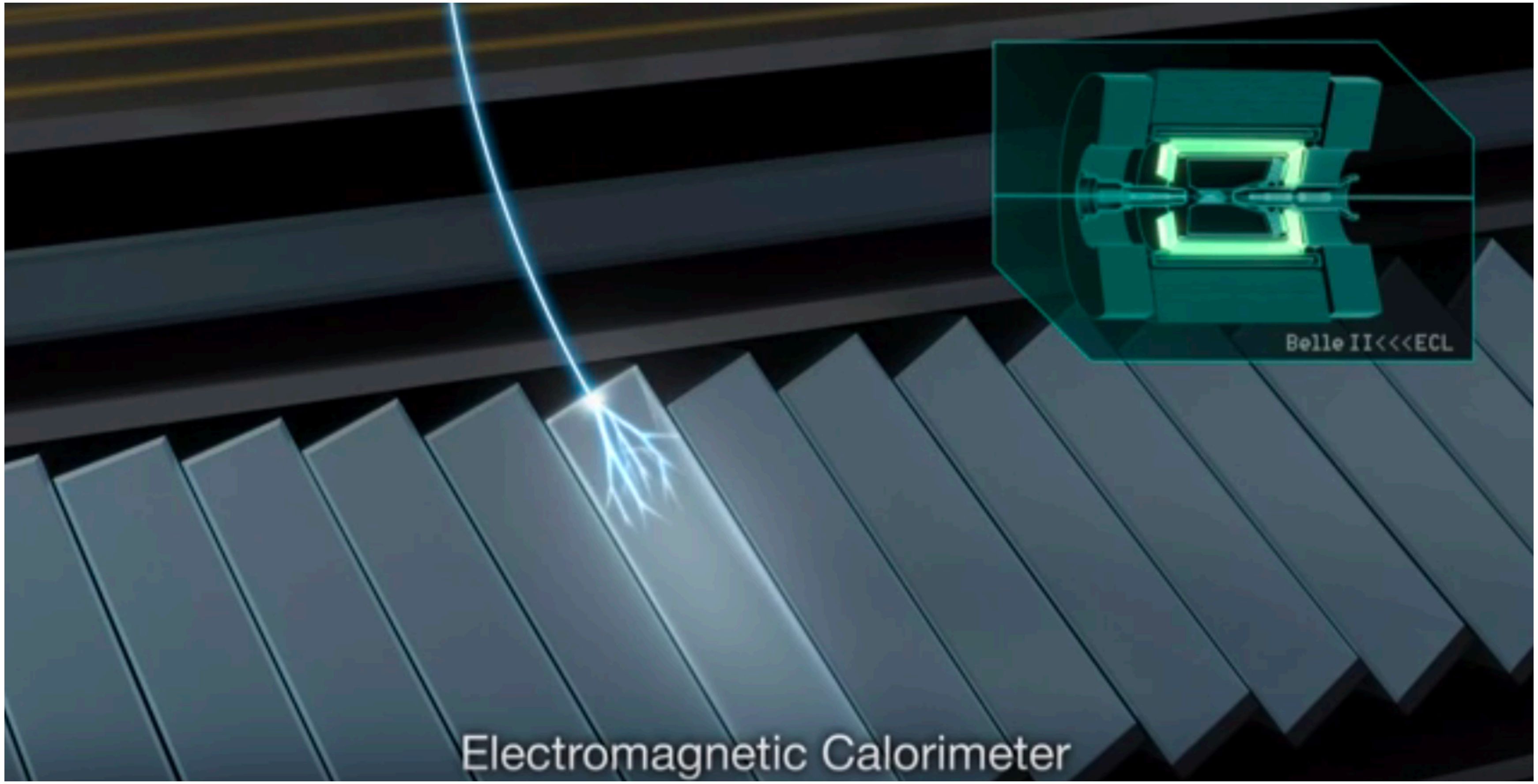
Belle II DIRAC system

BelleDIRAC: system developed by Belle II computing team as an extension of DIRAC.

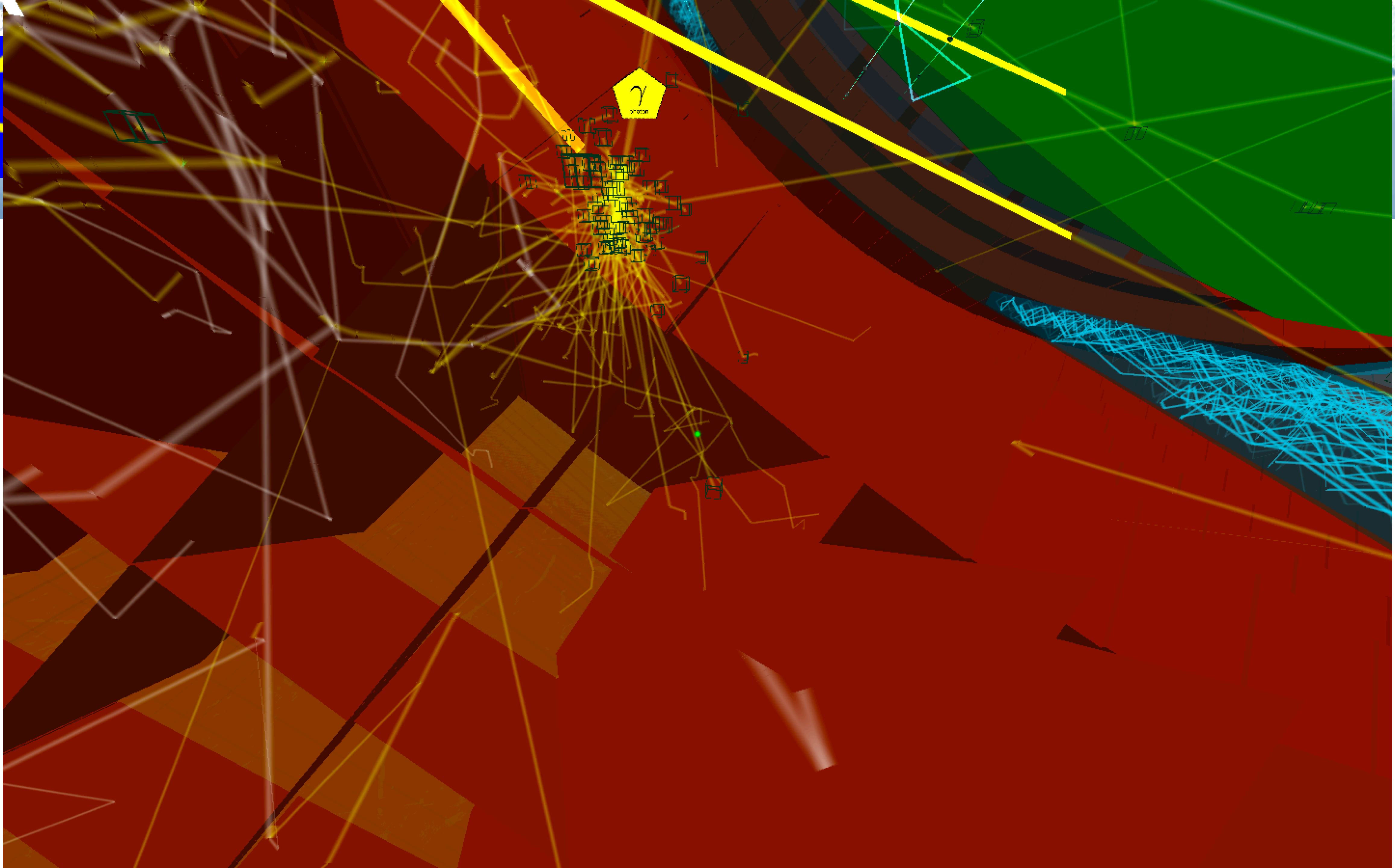


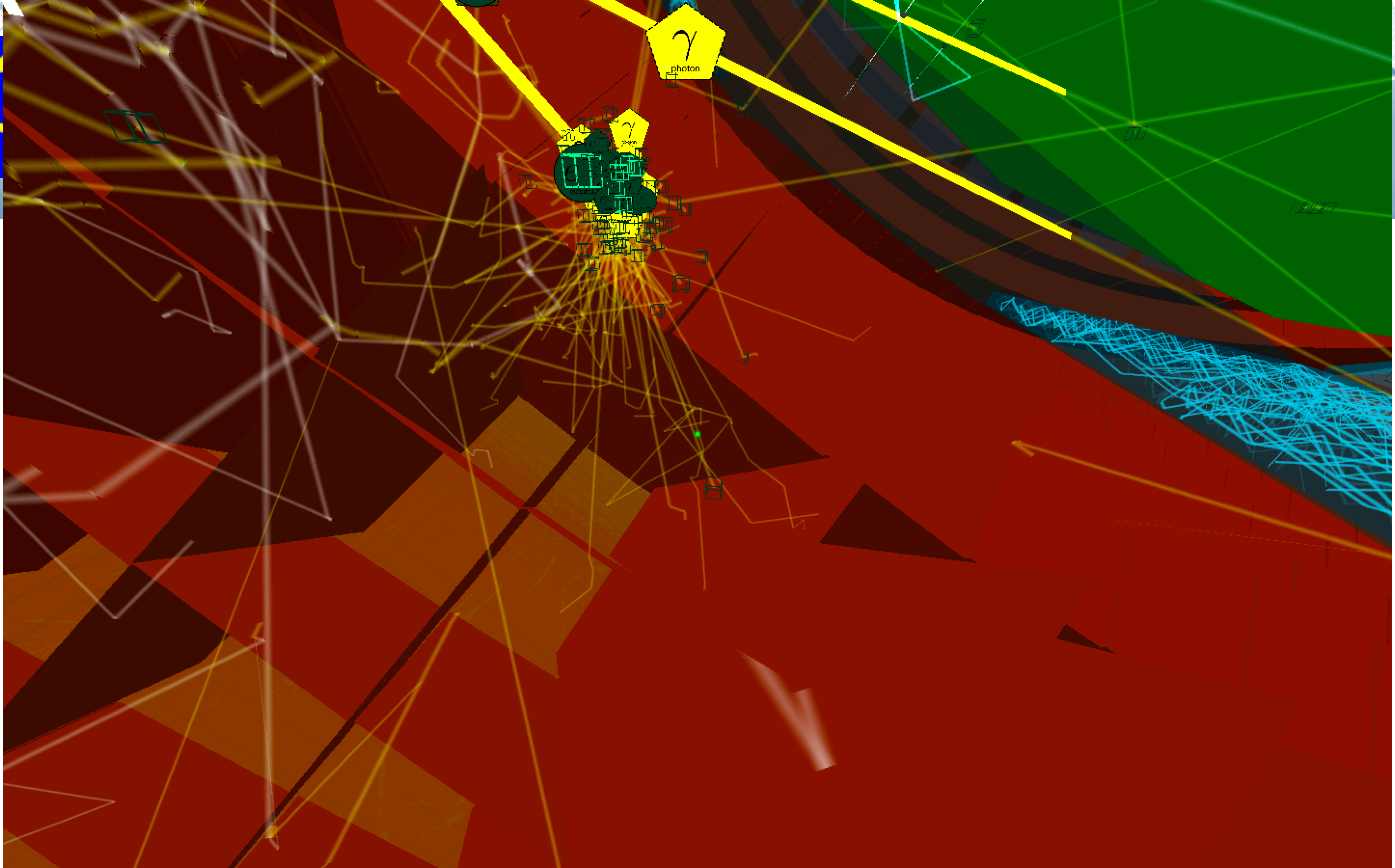
An Example of Using Convolutional Neural Network to Correct Gamma-ray Energy

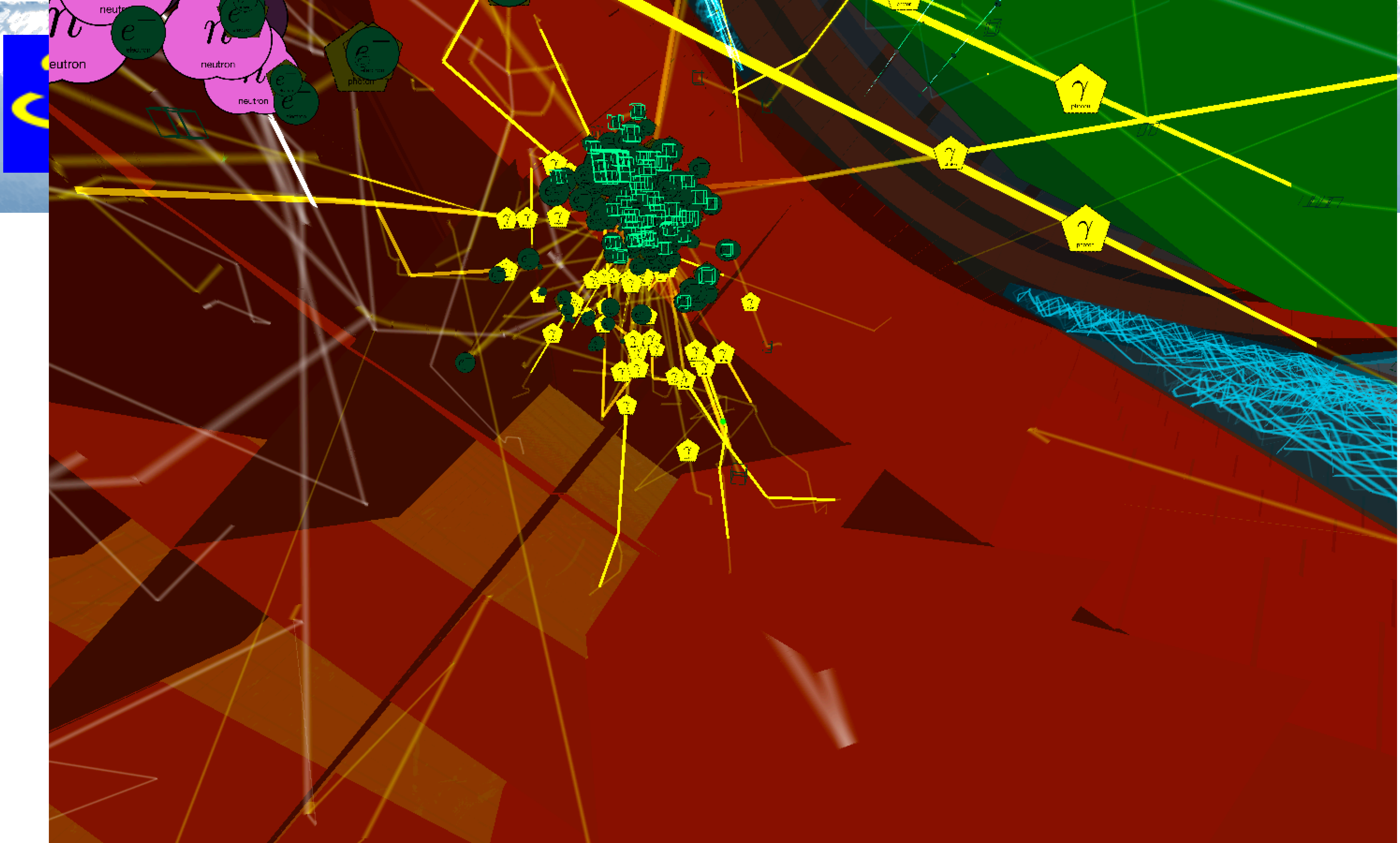
Gamma-ray energy measurement in Crystal array detector

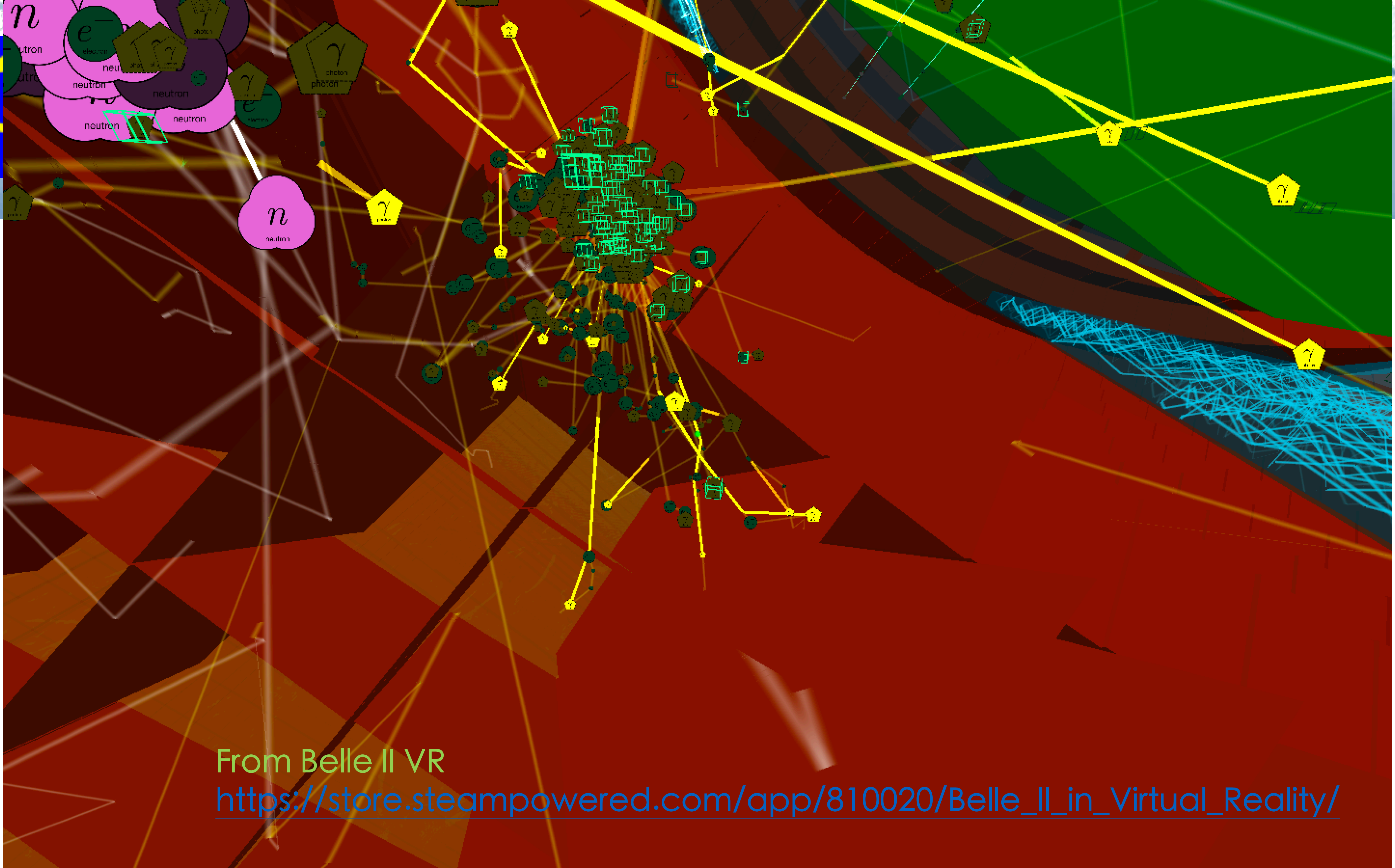


From KEK youtube channel
https://www.youtube.com/watch?v=nGCrrgXSEOk&t=1s


γ

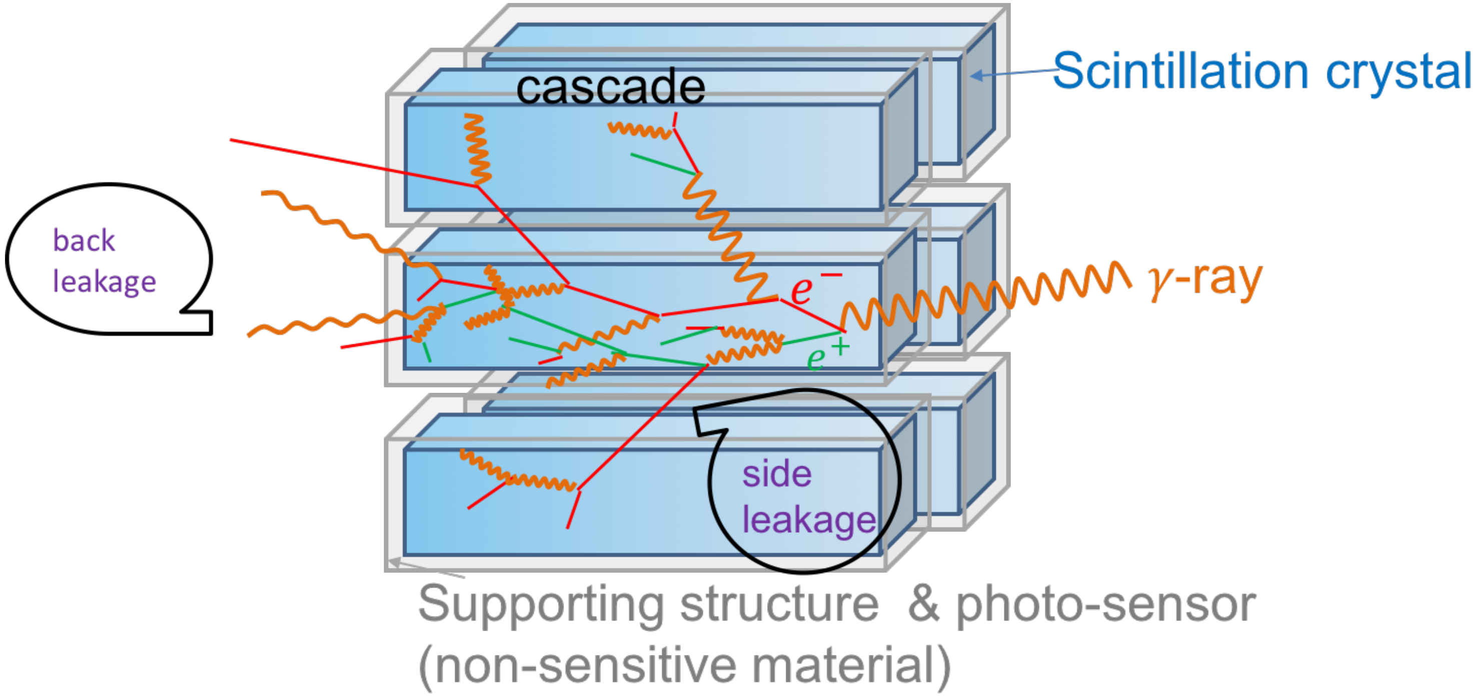

γ
photon


eutron
neutron
neutron
photon



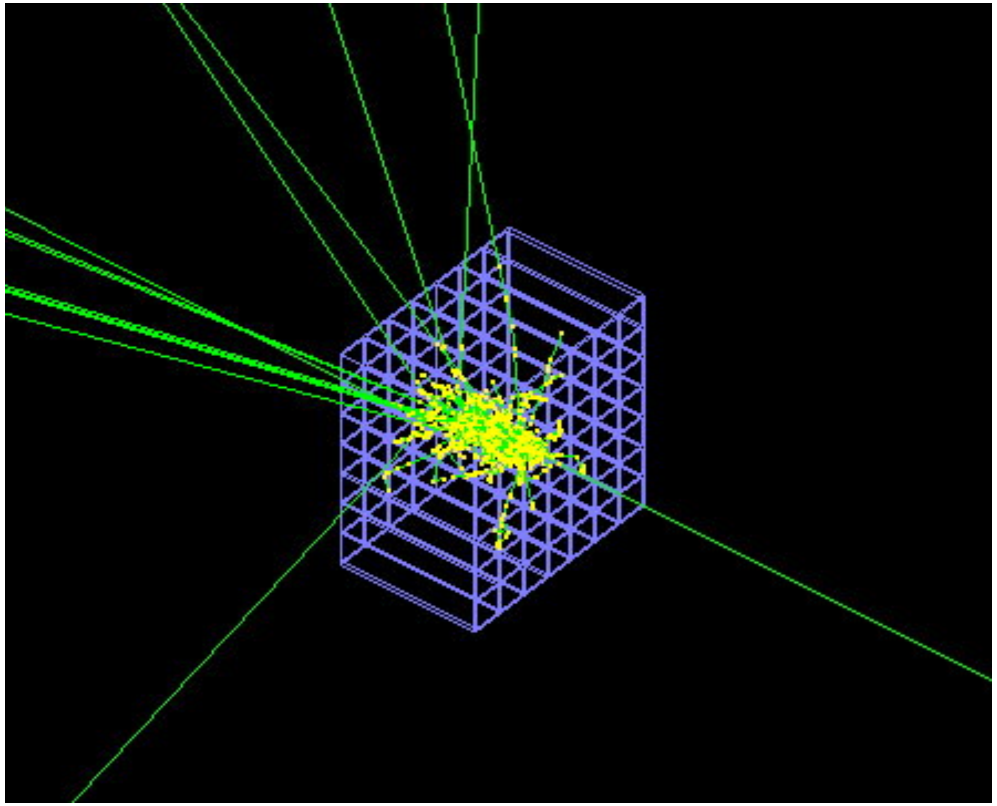
From Belle II VR
https://store.steampowered.com/app/810020/Belle_II_in_Virtual_Reality/

Gamma-ray energy leakage @crystal array



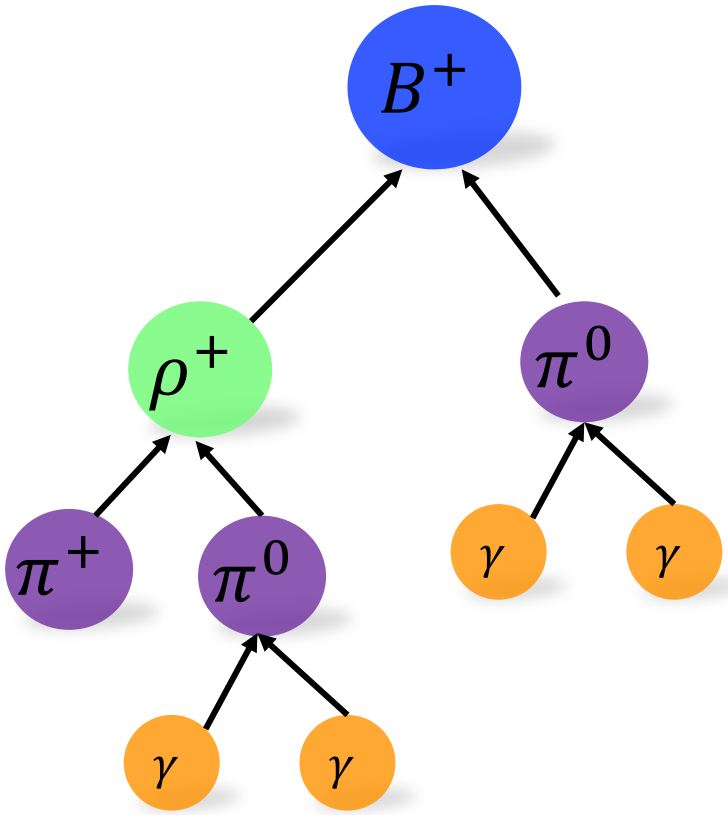
- Gamma-ray measurement at scintillation crystal:
  γ-ray with energy O (>0.1 GeV) transforms to charged electrons (positions).
  The charged particles moving inside crystal will cause light emission. Then, we use photon-sensor to measure light
  → obtain energy of gamma-ray.

Gamma-ray energy leakage @crystal array



A simulated γ-ray event:
side leakage and back leakage can be seen.

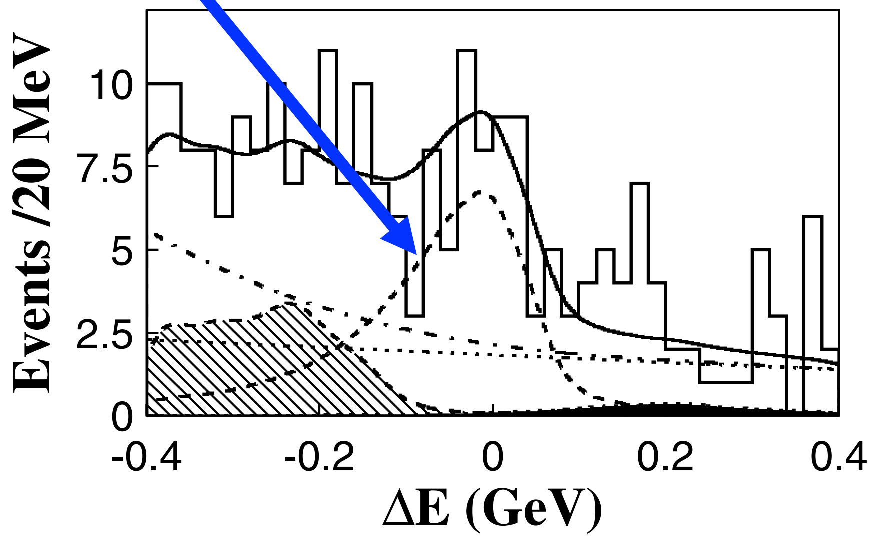
Gamma-ray energy leakage



$$E_B = E_{....} + E_\gamma + E_\gamma + E_\gamma + E_\gamma$$

$B^+ \to \rho^+ \pi^0$

$\pi^0 \to \gamma\gamma$



From PRL 94, 031801 (2005)

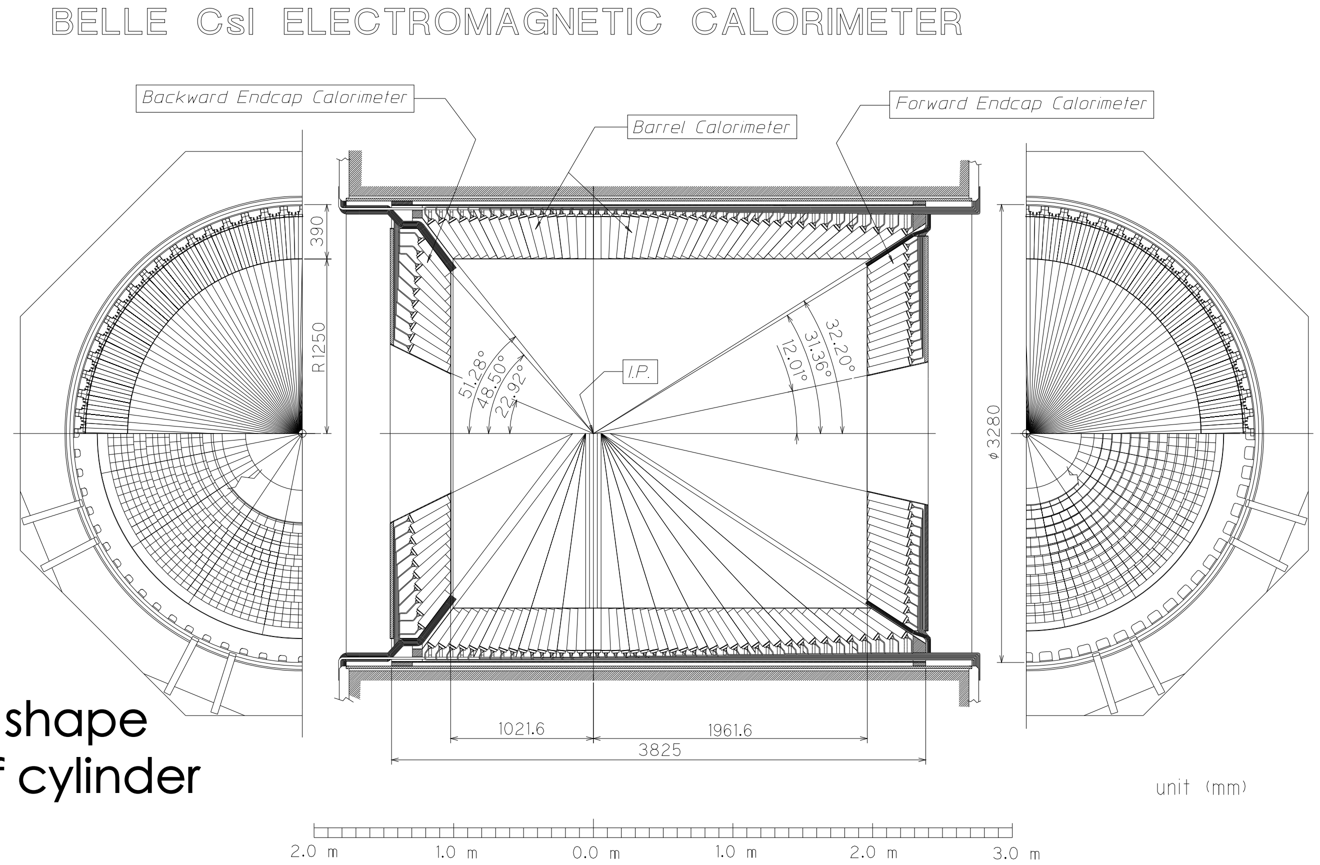
Problem:

Signal (dash line) tail overlap with other backgrounds

→Backgrounds also have uncertainty

→Affect final measurement uncertainty

ECL detector structure



Backward Endcap Barrel Forward Endcap

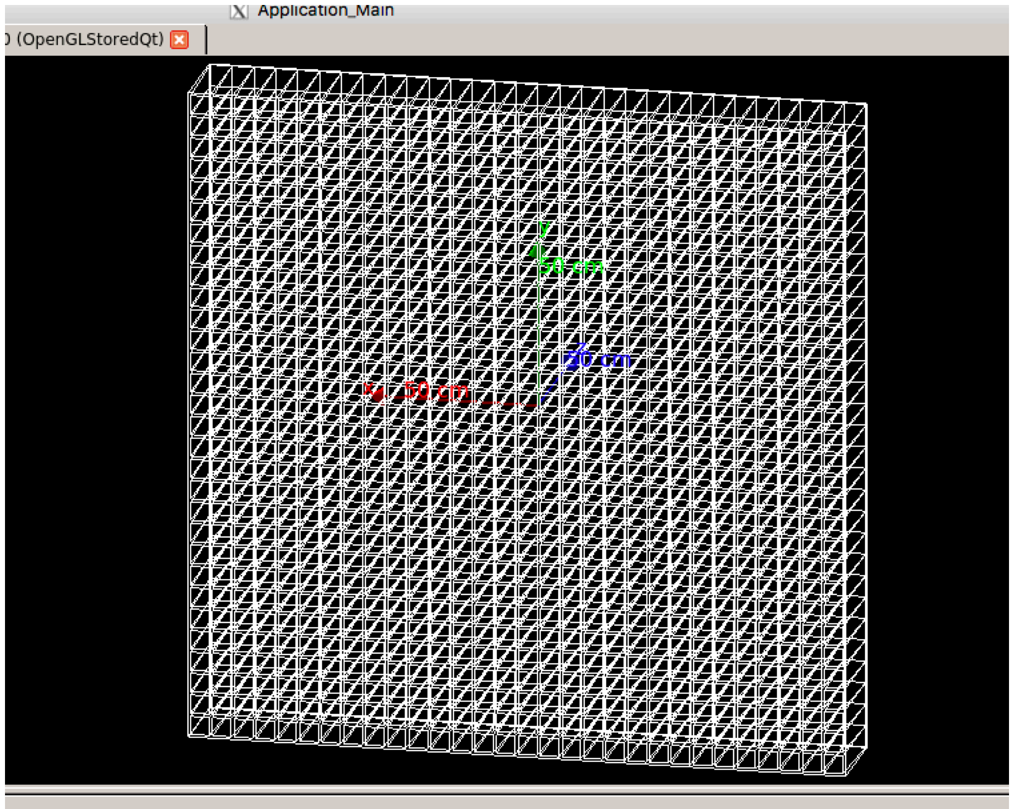
From Ikeda, Hitomi
PhD disserlation,
"Development of the CsI(Tl) Calorimeter for the measurement of CP Violation at KEK B-factory "

- ~8,000 CsI crystals
- ~100 kinds of crystal shape
- Equipped outside of cylinder

CsI crystal wiki:
https://en.wikipedia.org/wiki/Caesium_iodide

CsI Crystal 30X30 Array model

30 by 30 CsI crystal array geometry in Geant4 simulation



- Simulate[1] γ-ray incident into CsI crystal array.
- Each crystal has below geometry:

  55 mm X 55 mm (front face)
  65 mm X 65 mm (rear face)
  length: 30 cm
  The 0.5 cm gap is filled with aluminum.

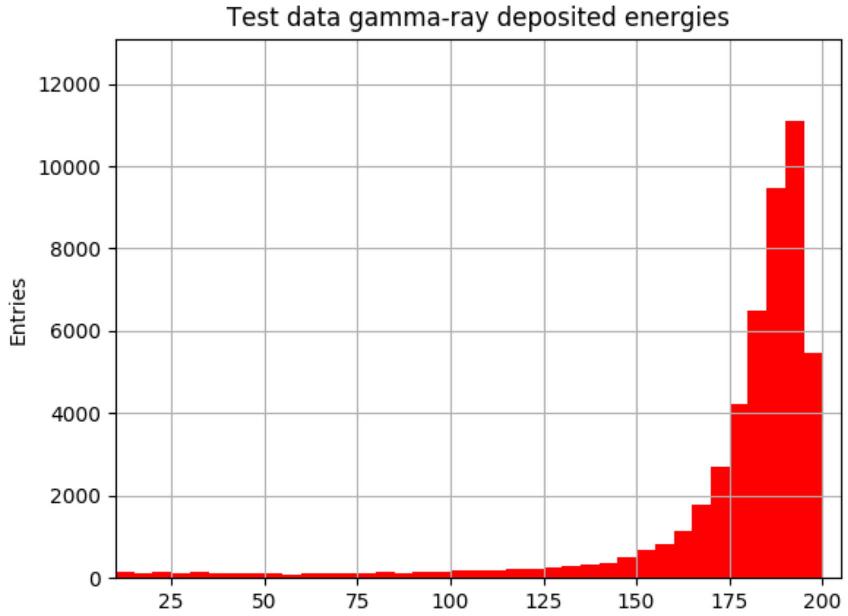
Measured γ-ray energy

$$E_{Total} = \sum_{i=1}^{900} E_{i,deposited}$$

[1]GEANT 4
C++ API for particle and radioactive source simulation
https://geant4.web.cern.ch/

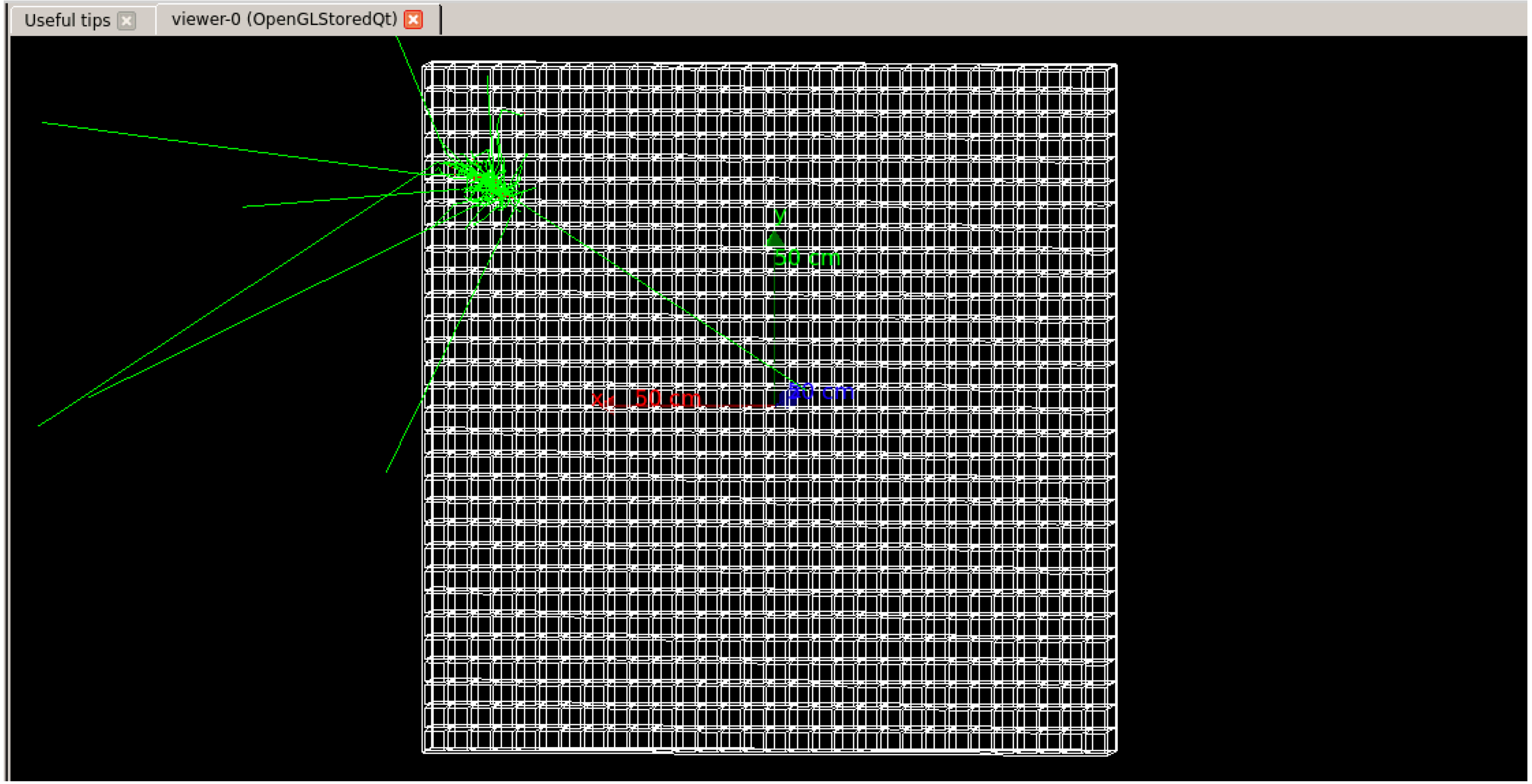
CsI Crystal 30X30 Array model

Incident of 200 MeV, 10,000 times



$\mathbf{E_{Total}}$ (MeV)

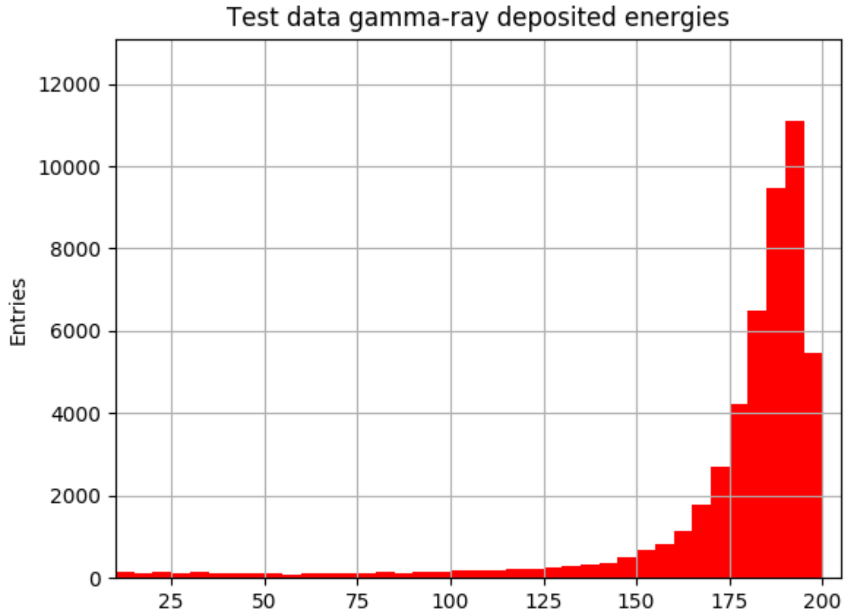
Measured $\boldsymbol{\gamma}$-ray energies ($\mathbf{E_{Total}}$) may not be close to 200 MeV.



A simulated $\boldsymbol{\gamma}$-ray event of 30X30 CsI array: energy leakage can be seen.

Regression problem

Incident of 200 MeV, 10,000 times



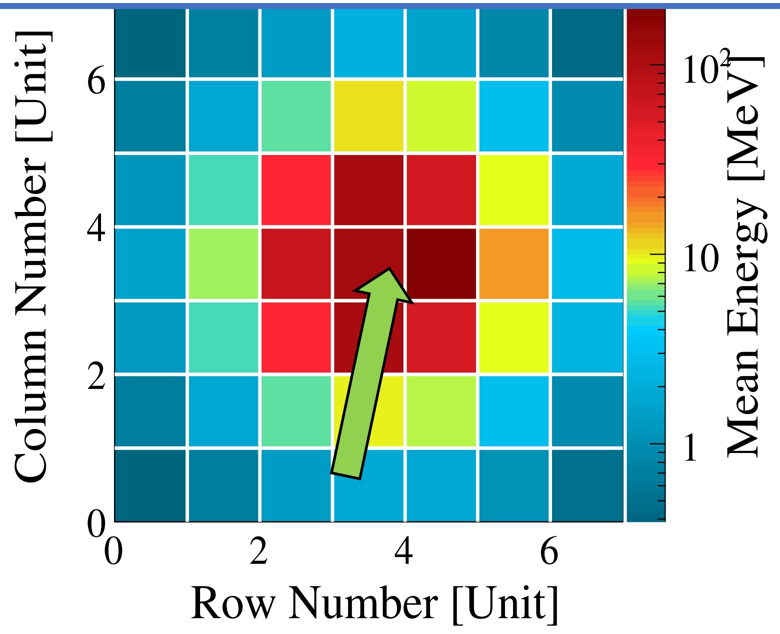
$\mathbf{E_{Total}}$(MeV)

Measured $\boldsymbol{\gamma}$-ray energies ($\mathbf{E_{Total}}$) may not be close to 200 MeV.

Problem:
How to make prediction of energy ($\boldsymbol{E_{pre}}$) that close to "true" gamma-ray energy ($\boldsymbol{E_{True}}$).
→Regression problem

Energy pattern on CsI crystal array

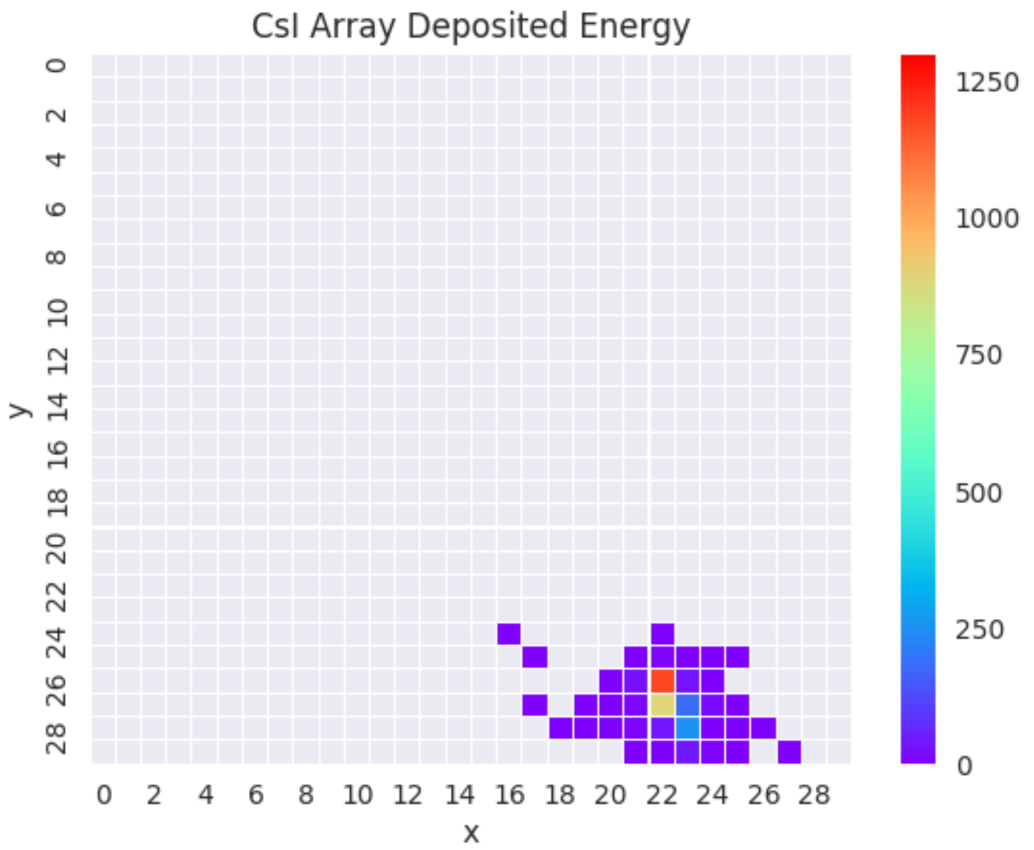
Photon of energy 1 GeV hits 2/3 towards the right of the central crystal. 1000 simulated events



- The deposited energy pattern is supposed to be related to incident angle and position. Also, the energy leakages relate to energy pattern.
- As the crystals are homogeneous, energy pattern can be the same over the crystal array.
- Regression problem.

- Solution: Convolutional Neural Network (CNN)
- Introduction of CNN webpage: http://cs231n.github.io/convolutional-networks/ (page of Cs231n class Standford Univ.)

Input Array, Predict Energy, Target



Input Data:
Deposited Energy of CsI Array

Target: (y_{True})

Ratio of Total deposited energy/True energy:

$$y_{True} = (\mathbf{E}_{\mathbf{Total}}/\mathbf{E}_{\mathbf{True}})$$

Prediction:($y_{Predict}$)

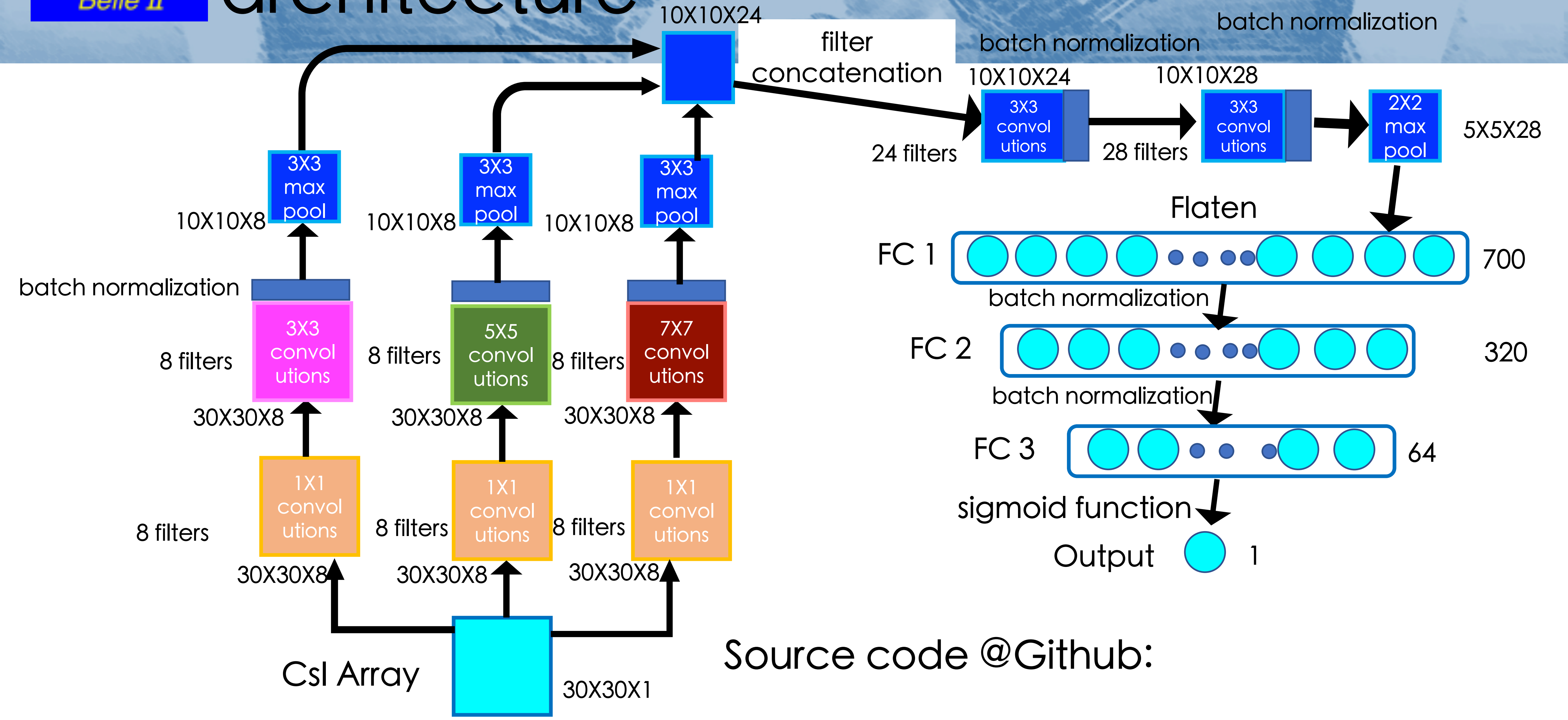
Predict energy by CNN:

$$E_{CNN} = \mathbf{E}_{\mathbf{Total}}/y_{Predict}$$

Loss function:

MSE of ($y_{Predict}$- y_{True})

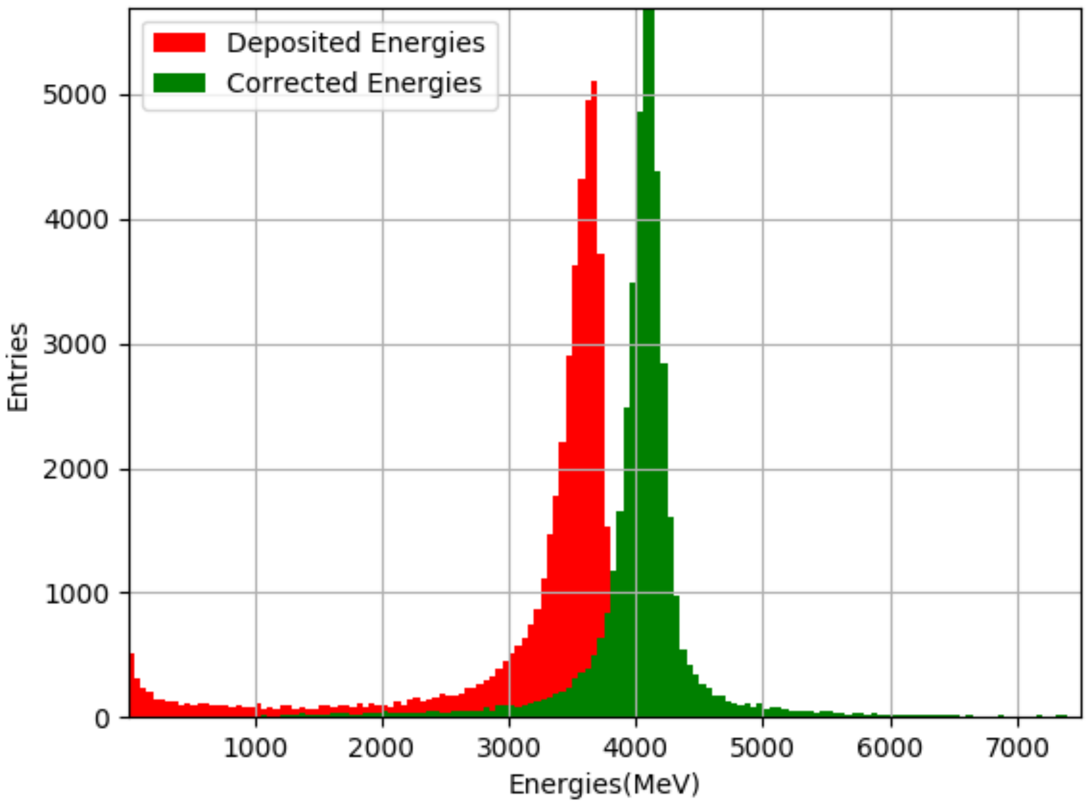
Convolutional Neural Network architecture



Source code @Github:

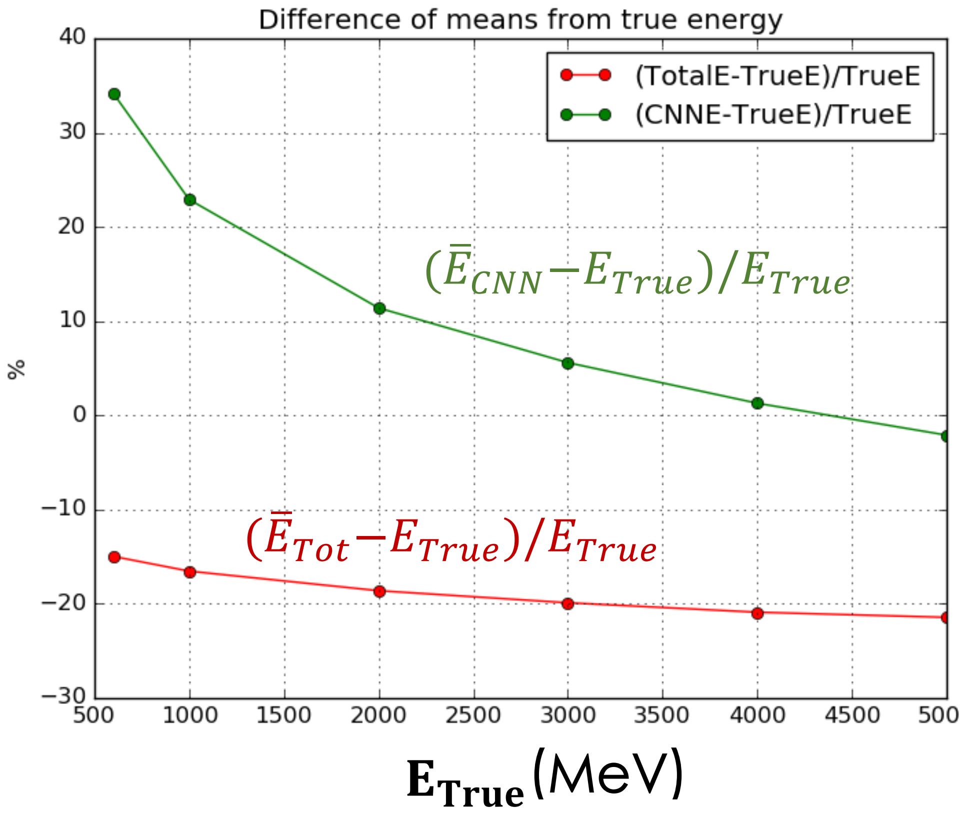
Test data result of mono-energy events

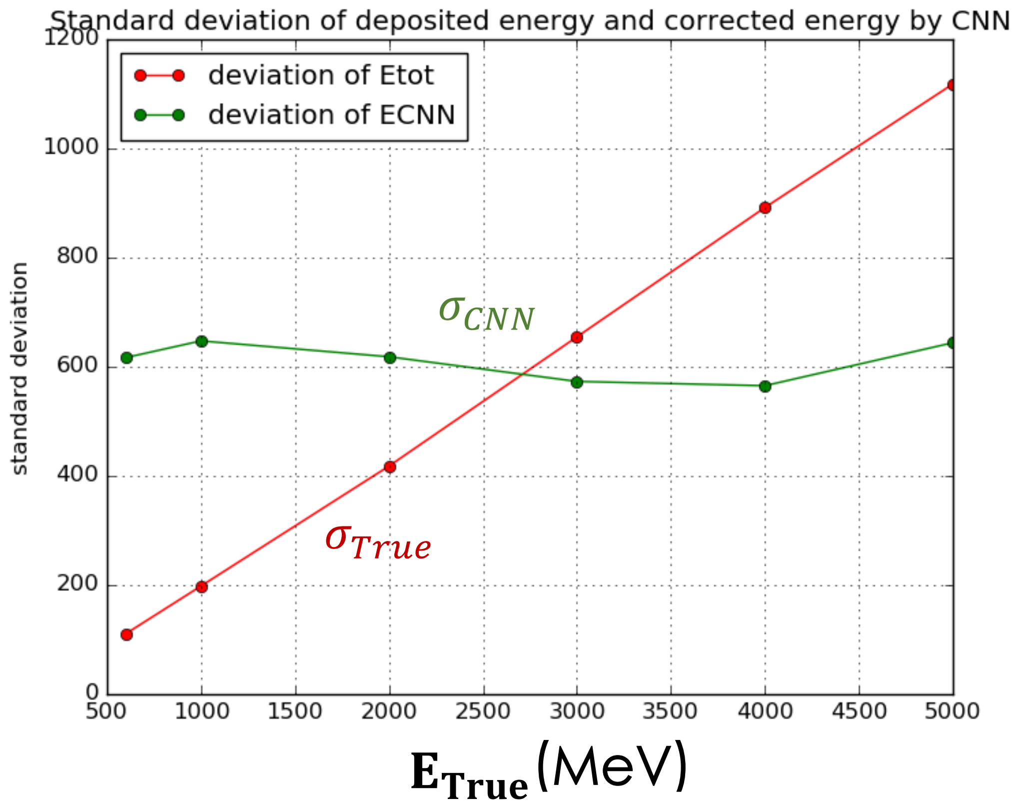
Measured $\boldsymbol{\gamma}$-ray energies ($\mathbf{E_{Total}}$)
Corrected $\boldsymbol{\gamma}$-ray energies ($\mathbf{E}_{CNN}$)



Incident of 4000 MeV $\boldsymbol{\gamma}$-ray, 10,000 times

Test data result of mono-energy events



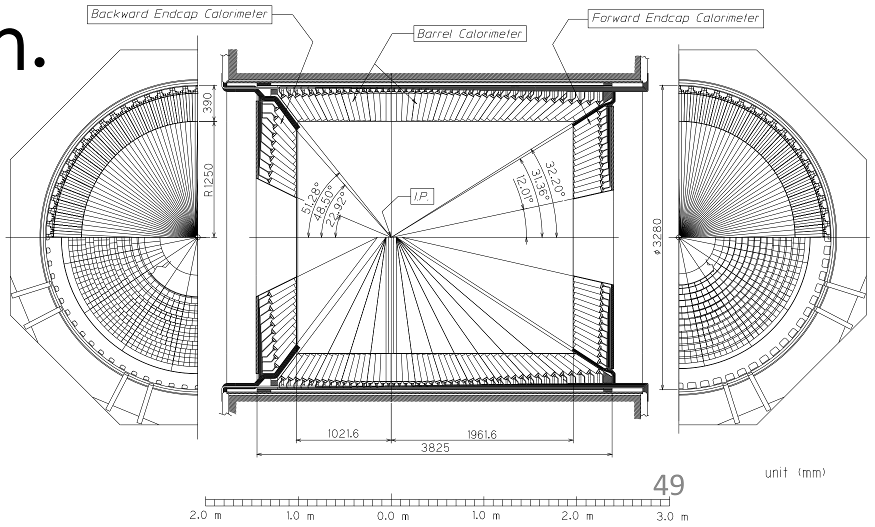


Correction of CNN method performance at high energy (>2000MeV) is good.
However, many gamma-ray having energy less than 2GeV in experiment
→Further improvement is needed.

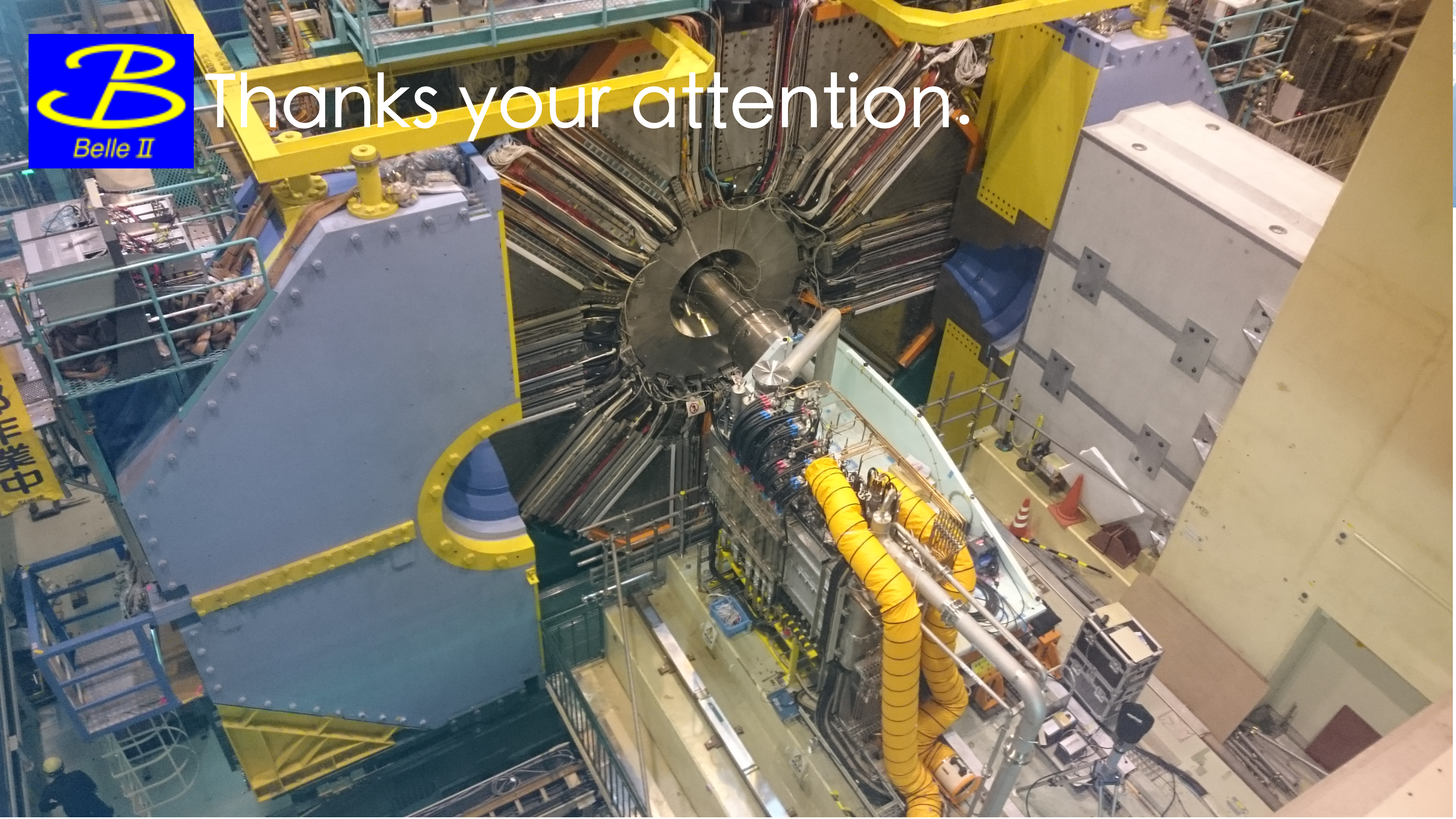
Simple CNN may be not good enough·········

- Proposed improvement:
  - group events by their energy pattern
    -->train multiple CNN models
  - target function:$(\mathbf{E}_{\mathbf{Total}}/\mathbf{E}_{\mathbf{True}}) \rightarrow (\mathbf{E}_{\mathbf{total}}/\mathbf{E}_{\mathbf{peak}})$
  - Others···
- Using CNN model for real “Belle II “ problem.
  - Heterogeneous CsI crystal
  - Crystal arrangement: cylinder geometry
  - Background gamma-ray and electronics noise

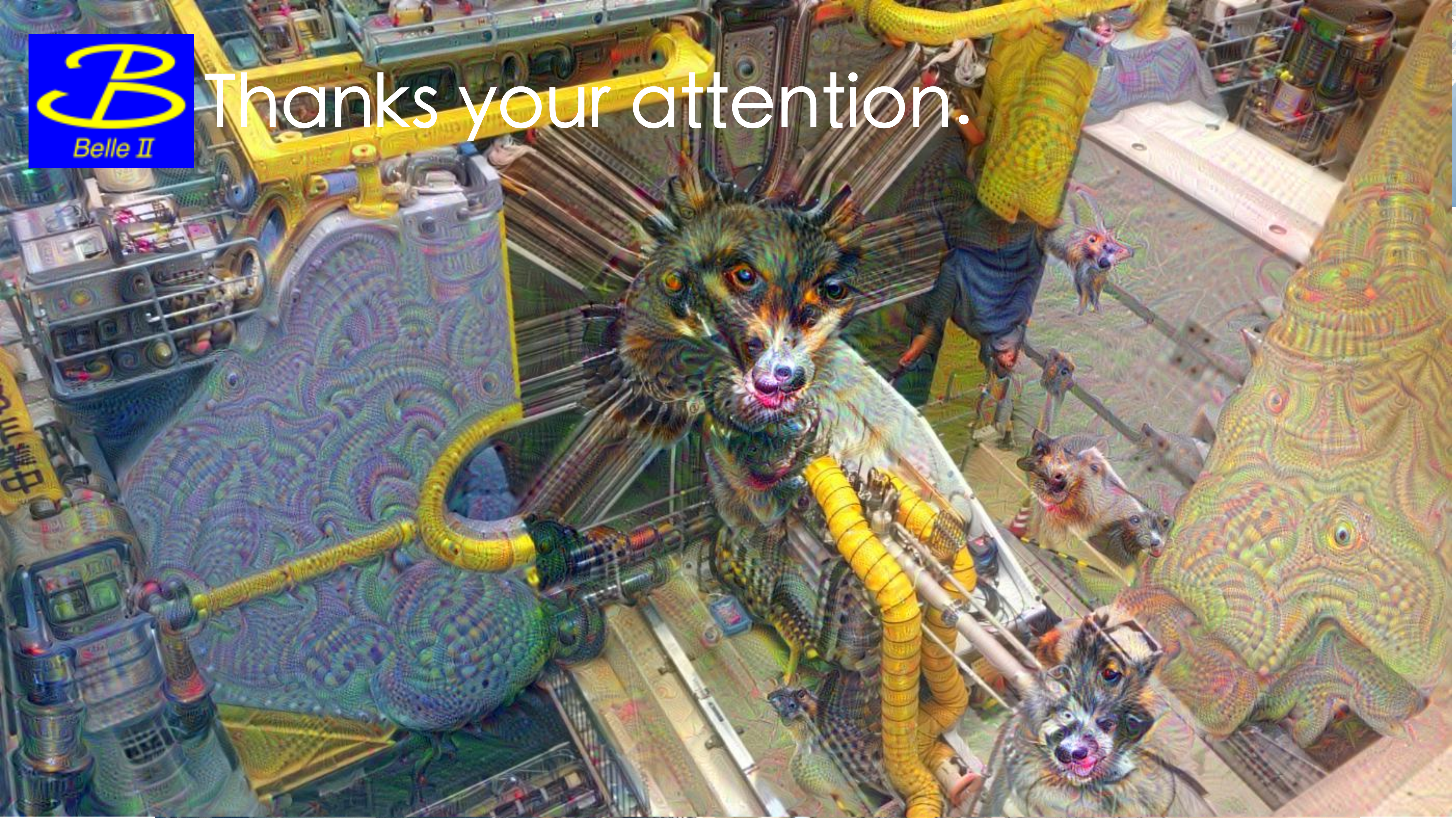
BELLE CsI ELECTROMAGNETIC CALORIMETER



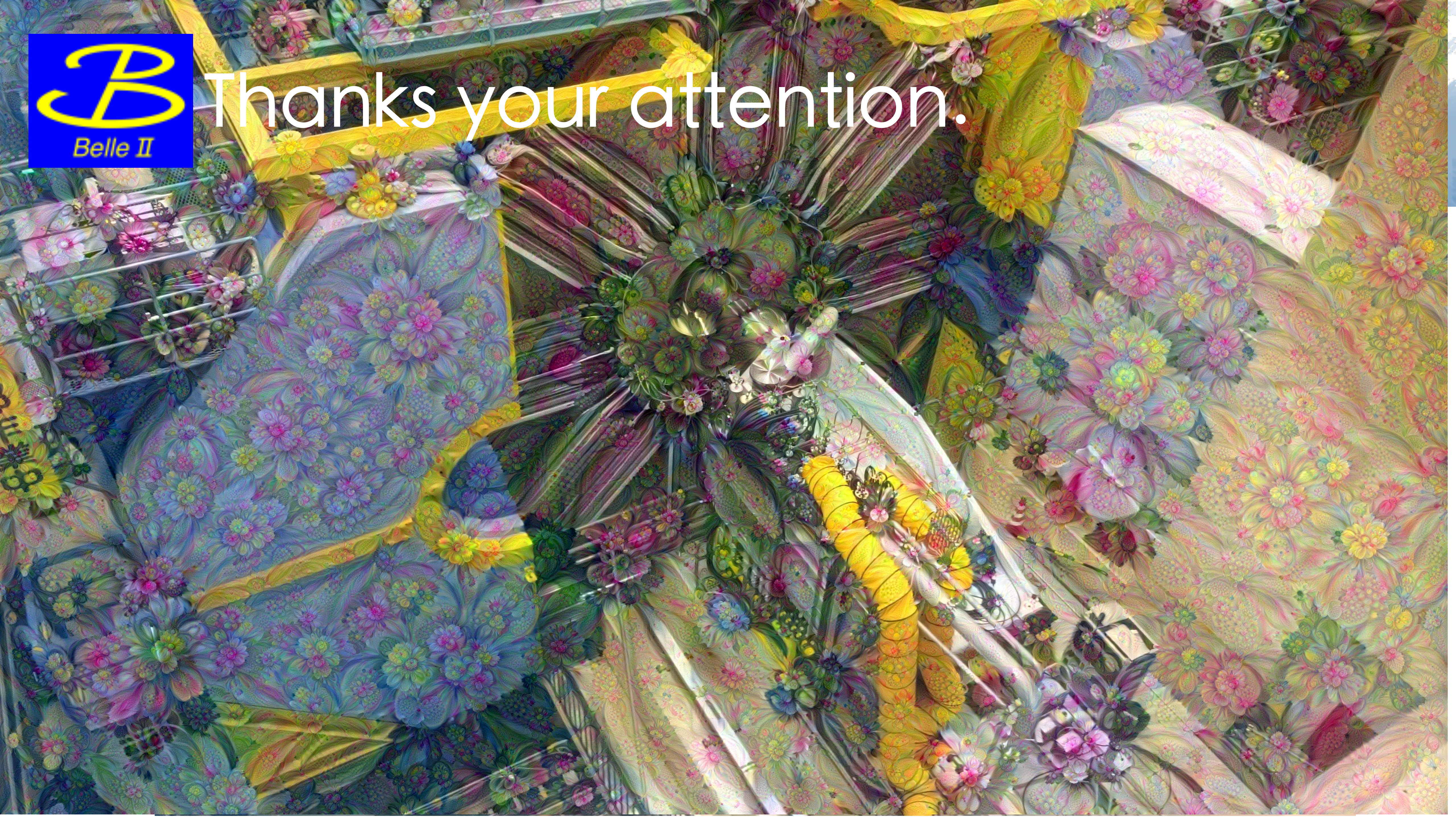
Thanks your attention.



Thanks your attention.



Thanks your attention.

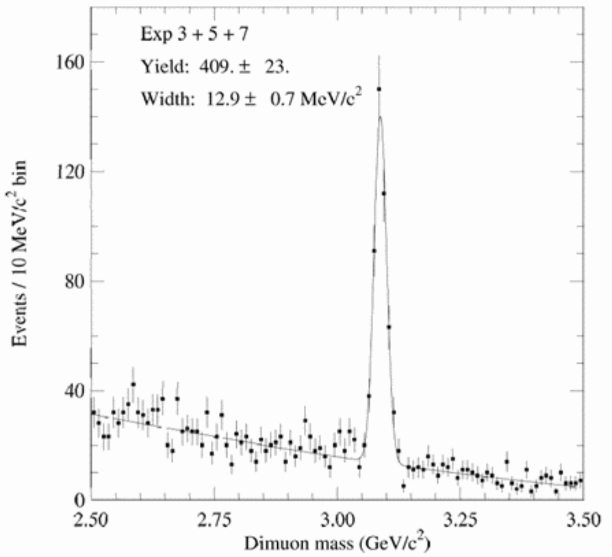


References

- Belle II experiment official website: https://www.belle2.org/
- Belle II facebook page: https://zh-tw.facebook.com/belle2collab/
- KEK Youtube channel: https://www.youtube.com/channel/UCMvGyDHrm820nYLUhwtm2DQ
- niconico生放送 Belle II first collision: http://live.nicovideo.jp/watch/lv312661713#49:24:30

B-Lab
ビー・ラボ：新しい素粒子発見のための公開データ解析プログラム
Open data analysis program to search for new particles
since 2004 copyright @ Belle collaboration

B-Lab: open data project of Belle

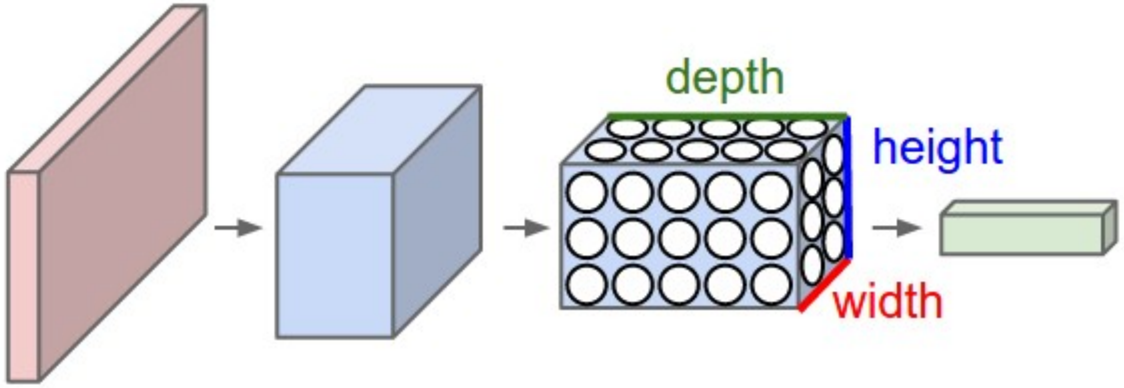


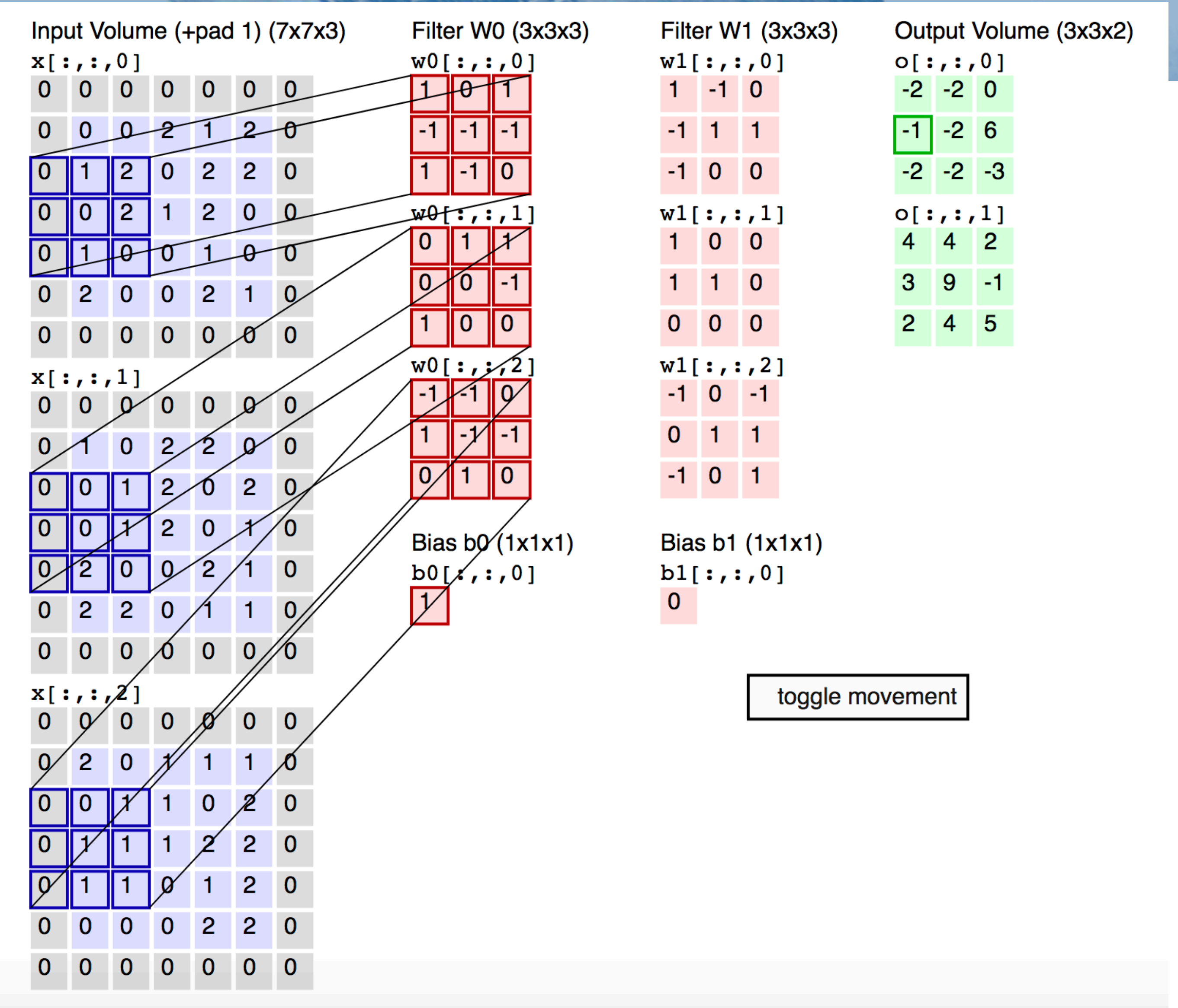
- Open data of "Belle" experiment.
  (only 0.1% of full data, ~1fB^{-1})
- B-Lab program is a project to open part of the Belle experimental data mainly to high school science clubs via the Internet and to give them a chance to "search for new particles."
- website: http://belle.kek.jp/b-lab/b-lab-english/
- Data in format of ROOT files.
  Can be processed by PyROOT, root_numpy, and root_pandas.

Backup

Convolutional Neural Network

- "convolution" layer

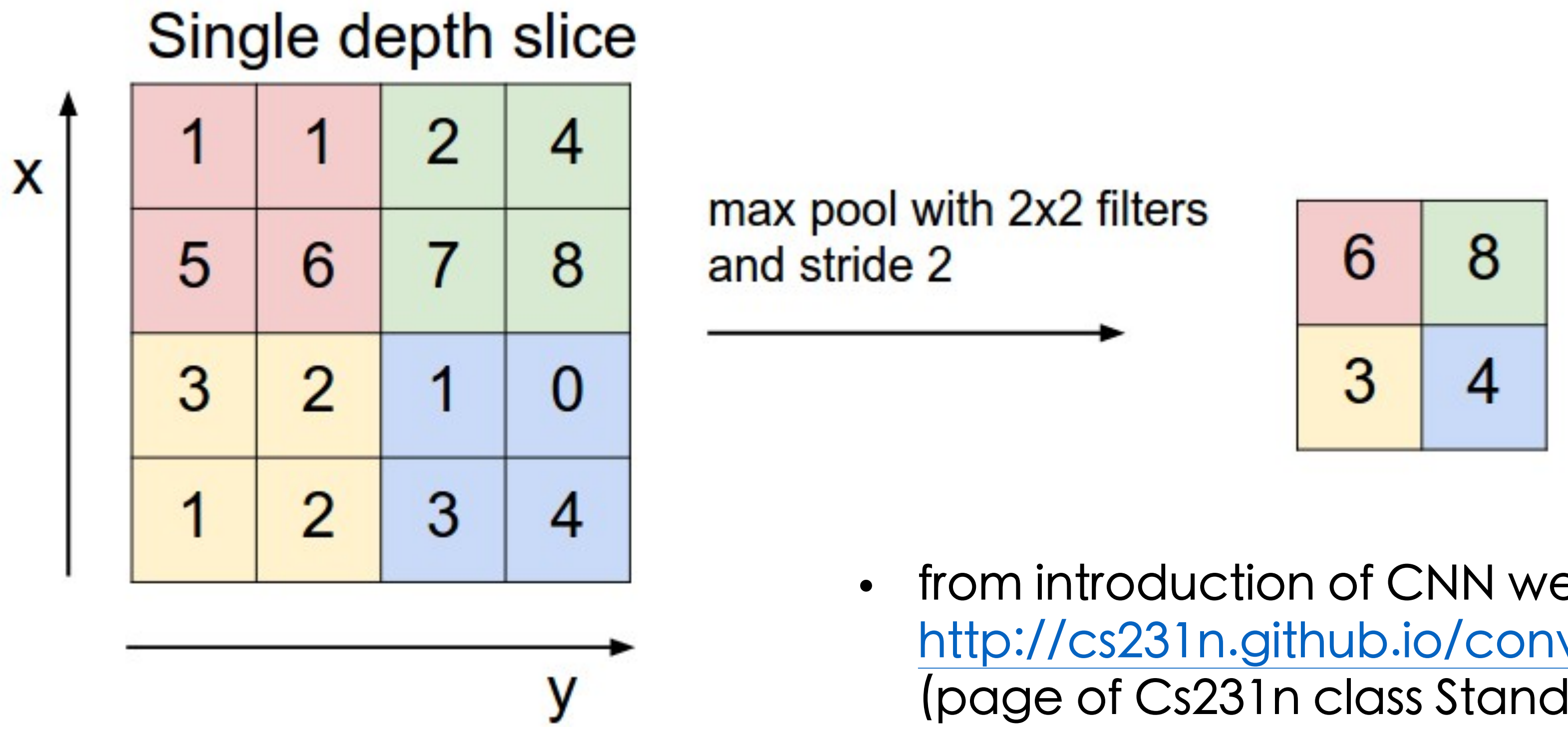




- from introduction of CNN webpage: http://cs231n.github.io/convolutional-networks/ (page of Cs231n class Standford Univ.)

Convolutional Neural Network

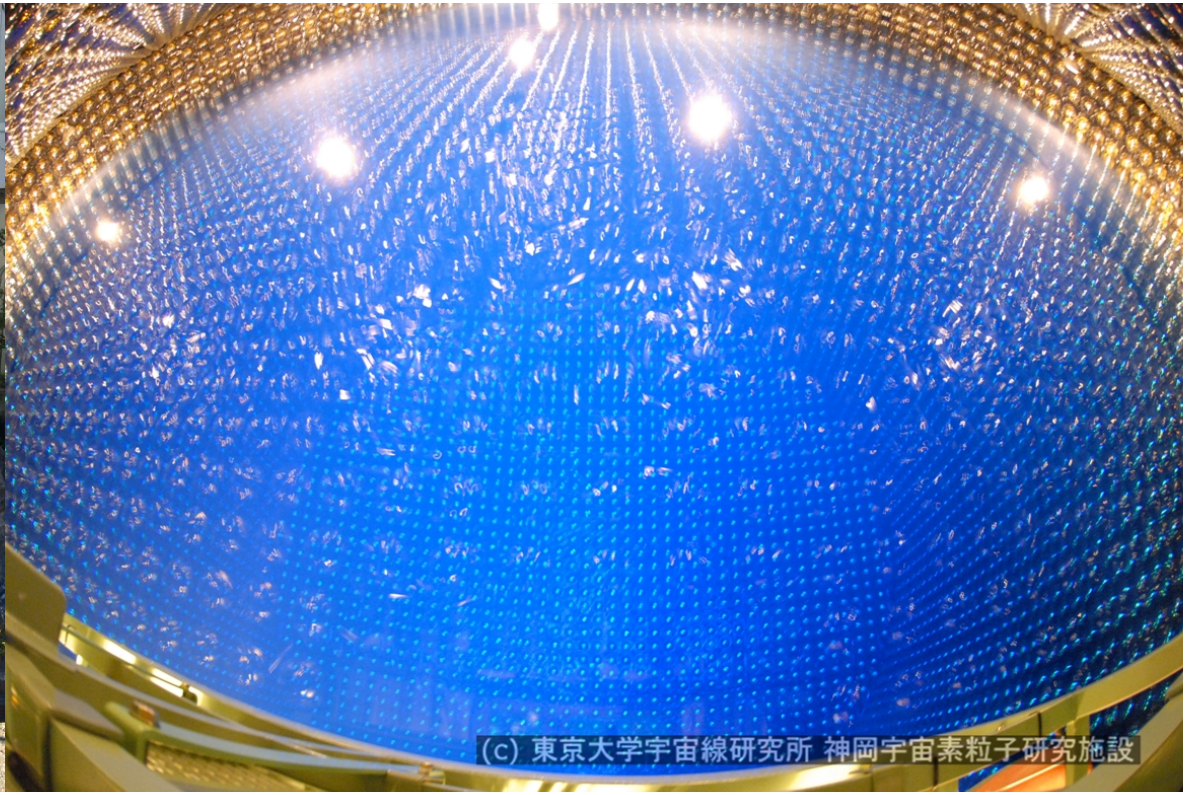
- "pooling" layer



- from introduction of CNN webpage: http://cs231n.github.io/convolutional-networks/ (page of Cs231n class Standford Univ.)

About my research

What I did for my PhD·········

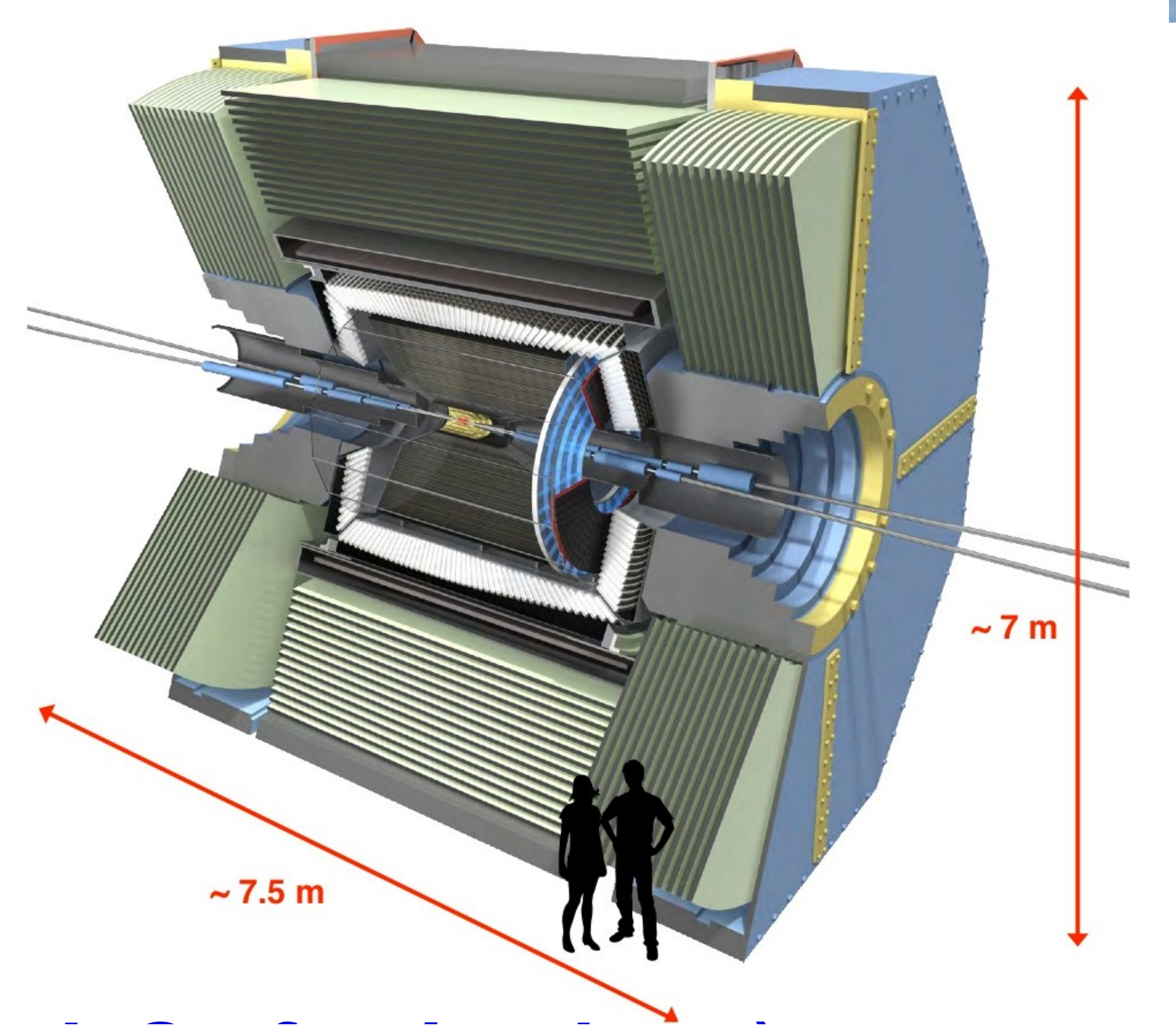


Neutrino experiment@Kamioka, Japan: T2K

Using Python to analyze Super-Kamiokande data

Belle II detector size

- “”

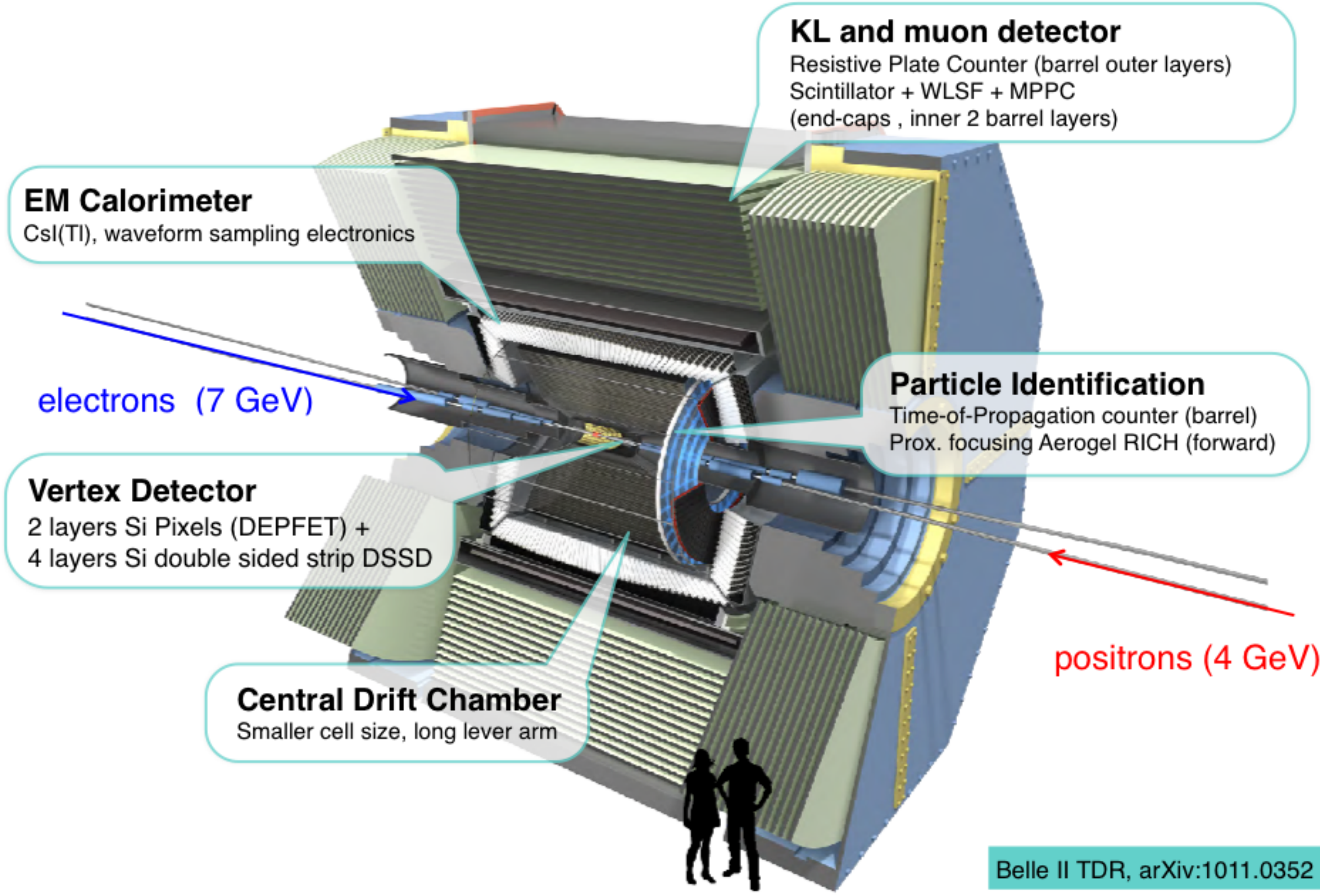


Big Data in particle physics exp.

- Large detector use complex of sub-detectors to measure a single event. Parts of sub-detectors can detect specific particles.
- Recent years, each sub-detector has O(1,000~10,000) pixels. So, the complex detector usually have totally O(1,000,000) pixels.
  → one event has 1 million of "features" raw data.
- "Big Data" techniques are widely used in large particle experiments.

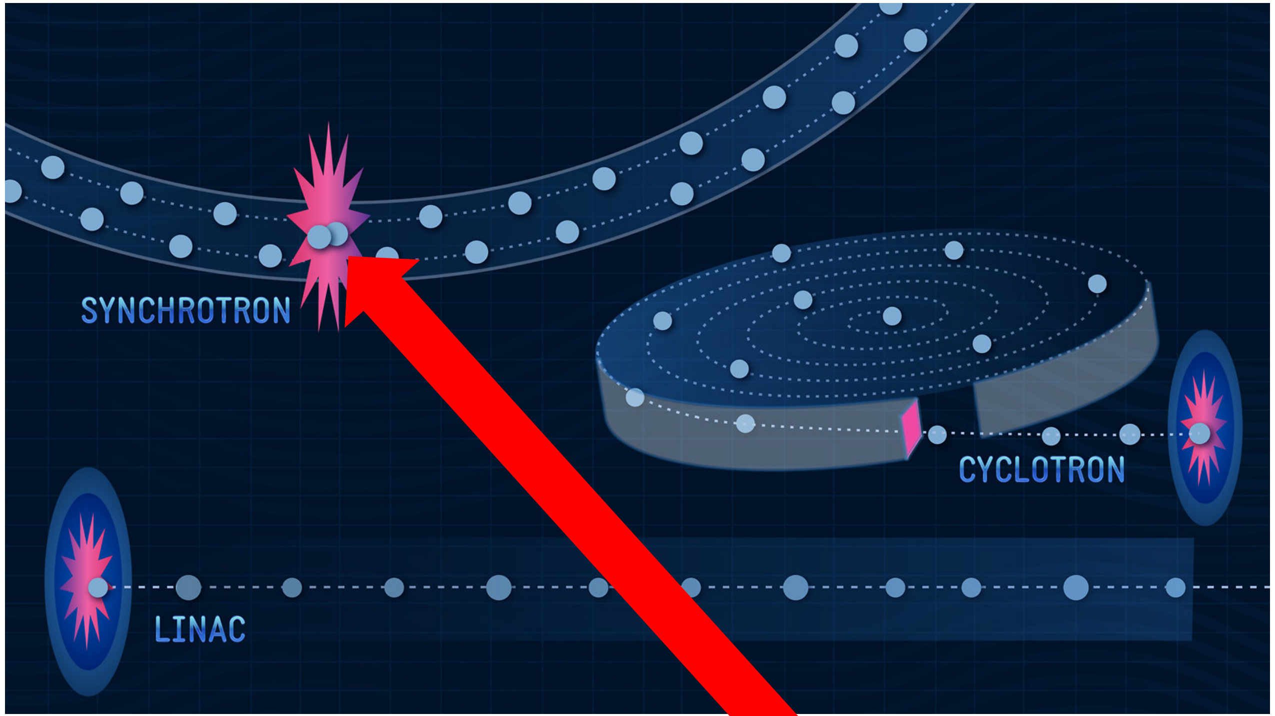
Big Data in particle experiments

- Example: Belle II detector
  - PXD: 1600*250*2 pixels
  - SVD: 187 sensors and O(1,200~1,500)strips/per sensor
  - CDC: 56,576 wires
  - TOP: ~8,000 electronics channels
  - ARICH: 131,760 pixels
  - ECL: 8,736 crystals
  - KLM: 17,000 photo-sensors
- Data size: 1MB (after event builder by firmware)
- Trigger rate: max 30 kHz.
- Data rate: max 30 GB/sec



Belle II TDR, arXiv:1011.0352

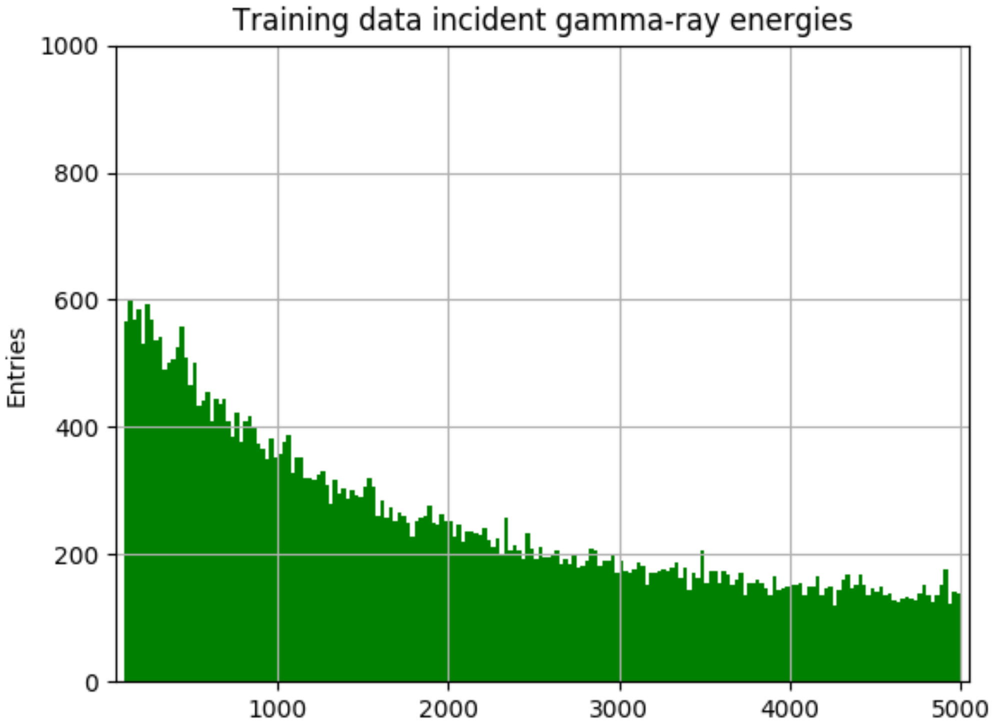
Particle physics experiment (or called High energy physics exp.)



From Symmetry magazine
"**A primer on particle accelerators**"
07/12/16

Accelerated particles are used to produce "unfamiliar" particles in our life.

Training samples & Target



$\mathbf{E_{True}}$ (MeV)

Target: ($y_{\_True}$)

Ratio of Total deposited energy/True energy:

$$(\mathbf{E_{Total}}/\mathbf{E_{True}})$$

Another target option: $(\mathbf{E_{total}}/\mathbf{E_{peak}})$

$\mathbf{E_{peak}}$: peak value of total energy histogram of specific true energies

Test data result of mono-energy events

E_{real} (MeV)	600	1000	2000	3000	4000	5000
mean of E_{total} (MeV)	510.2	834.7	1627.9	2403.4	3164.4	3928.2
mean of E_{CNN} (MeV)	805.0	1229.1	2228.1	3167.9	4051.6	4895.1

E_{real} (MeV)	600	1000	2000	3000	4000	5000
RMS of E_{total} (MeV)	110.6	198.1	417.9	654.7	891.73	1118.1
RMS of E_{CNN} (MeV)	616.73	647.4	618.3	573.1	565.22	643.99

施工中

Backup

E_{real} (MeV)	500	1000	2000	2500	3000	3500	4000
RMS of E_{total} (MeV)	67.3	149.3	334.0	432.1	530.7	629.5	742.4
RMS of E_{NN} (MeV)	62.5	146.2	330.0	428.2	527.1	620.3	736.3