2. Application of the Machine Learning to the Collider Experiments

2-1 Accelerator tuning using ML
2-2 DNN with low-level data
2-3 Data reduction based on the sparse sampling

Machine Learning in Collider Experiments

High Energy Physics Collider experiment
→ based on the "Big Data" processing

There are many layers of data processing

Challenges

- Huge amount of Accelerator components/Detector sensors
# of signals to control **~200,000** in SuperKEKB Accelerator

- Huge amount of Experimental data
~30 Peta Byte / year in Belle II experiment

- Signal vs BG
B-meson signal BR **~$O(10^{-7})$,** Extract signal from huge BG

Novel technology to treat the big data is required

Machine Learning in Collider Experiments

High Energy Physics Collider experiment
→ based on the "Big Data" processing

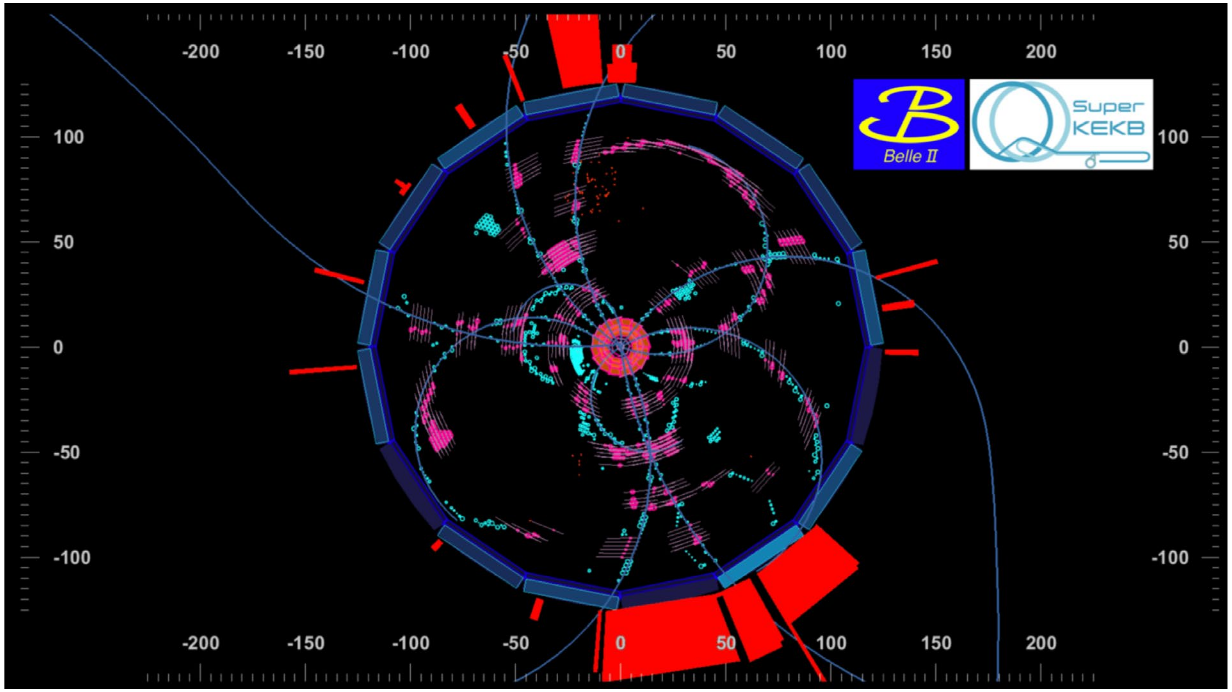
There are many layers of data processing

Solution

Modern Machine Learning techniques, such as Deep Neural Network developed in data science, are expected to be **powerful tools** to provide **more efficient / precise big-data processing** for the high energy physics collider experiments

HEP collider exp. x Da

M.L. Application



Beam Colliding exp.
DAQ

Data calibration
Reconstruction

Physics Analysis

HEP collider exp.
Large amount of data
High operating costs

e. g. Belle II exp.
~30 Peta Byte / year
Several Billion yen / year

Improvement of the data processing performance by ML application, affects the great cost effectiveness

Machine Learning

Supervised Learning

Task driven

- **Classification**
- **Regression**

Unsupervised Learning

Data driven

- **Dimensionality Reduction**
- **Clustering**

Reinforcement Learning

Environment driven

Algorithm learns to react to the environment

- Real-time decisions
- Game AI
- Leaning Tasks
- Robot Navigator
- ....

There are several Machine Learning types

Ma

In the past high-energy experiments
we mainly used
Supervised Learning ML
for the **Classification**
based on the **High-level feature data**

Supervised Learning

Task driven

Classification

Regression

Dimensionality Reduction

Clustering

Algorithm learns to react to the environment

Real-time decisions
Game AI
Leaning Tasks
Robot Navigator
....

There are several
Machine Learning types

Recently, it is possible to apply

- Various **modern ML methods**
- ML based on the **Low-level feature data**

Reasonably low-cost / fast computing resources

ML algorithm improvement

Classification

Regression

Dimensionality Reduction

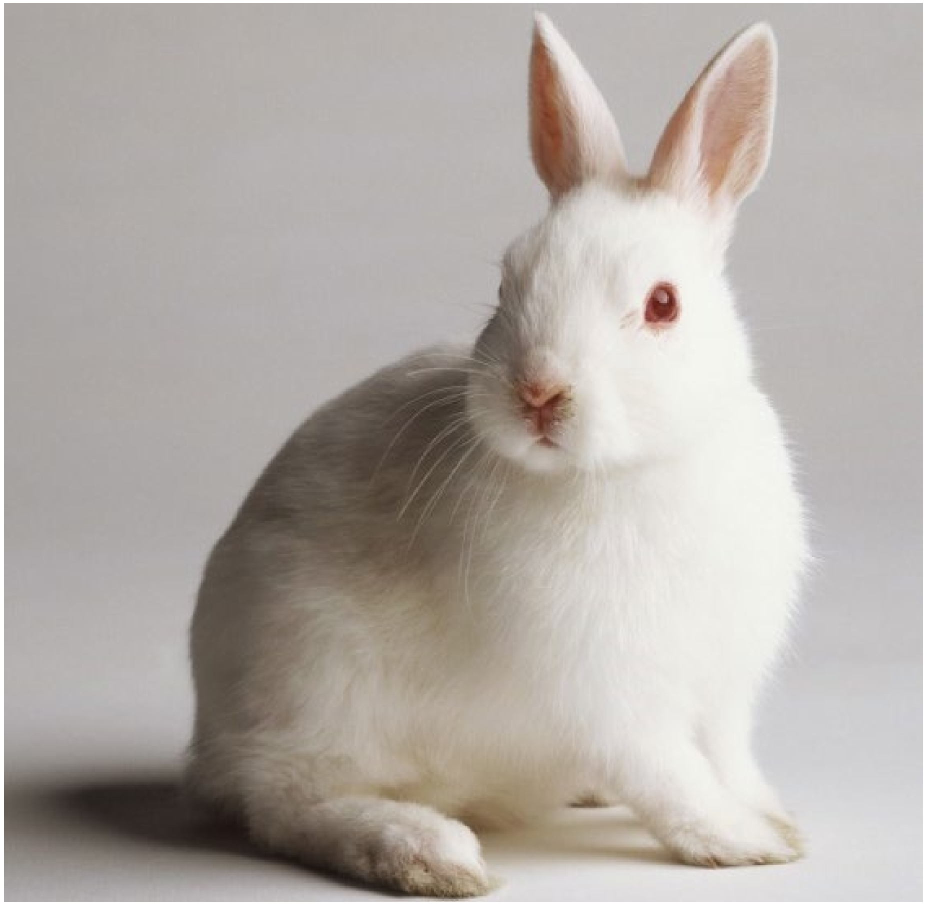
Clustering

Real-time decisions
Game AI
Leaning Tasks
Robot Navigator
....

There are several
Machine Learning types

Low-level data vs High-level data

Low-level data
(Image data)



High-level data
(Feature variables)

Body Color	White
Eye Color	Red
Ear length	8cm
# of legs	4
Weight	800g

Low-level data vs High-level data

Low-level data
(Image data)

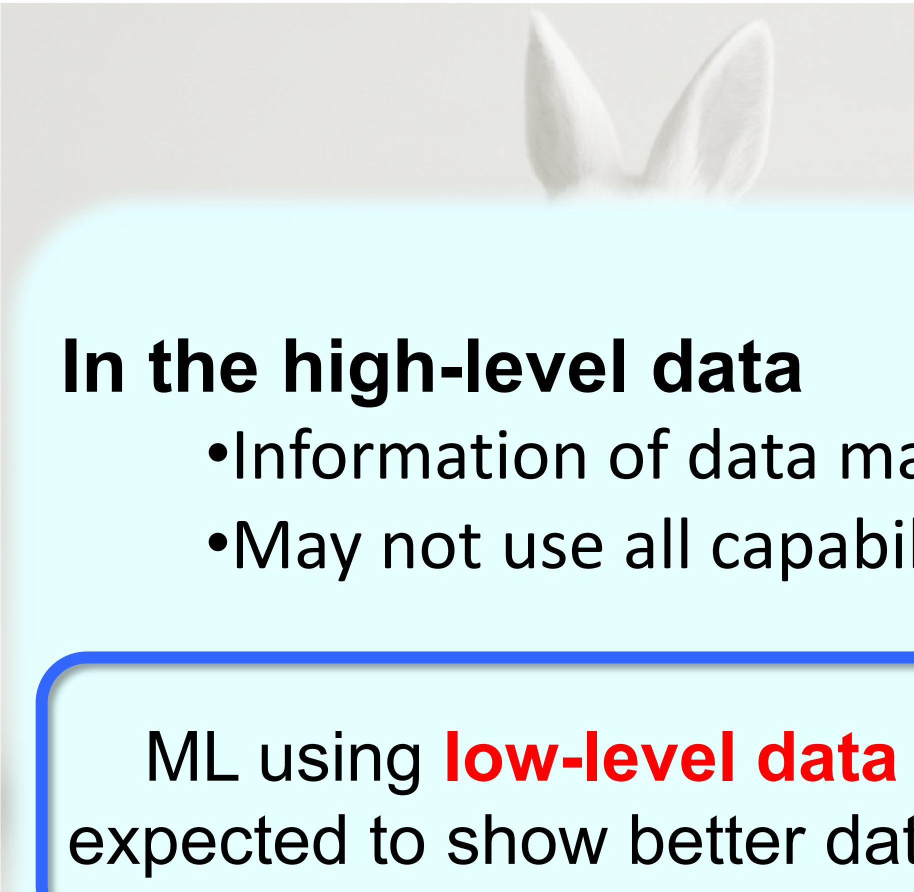
High-level data
(Feature variables)

Body Color	White
Eye Color	Red
Ear length	8cm
# of legs	4
Weight	800g

Information disappears ?!

Low-level data vs High-level data

Low-level data
(Image data)



High-level data
(Feature variables)

Body Color	White

In the high-level data

- Information of data may be degraded
- May not use all capabilities of data

ML using **low-level data** (pre-processing data) is expected to show better data processing performance

AI : Artificial Intelligence

Any technique that enables computers to mimic the intelligence or behavioral pattern of humans

ML : Machine Learning

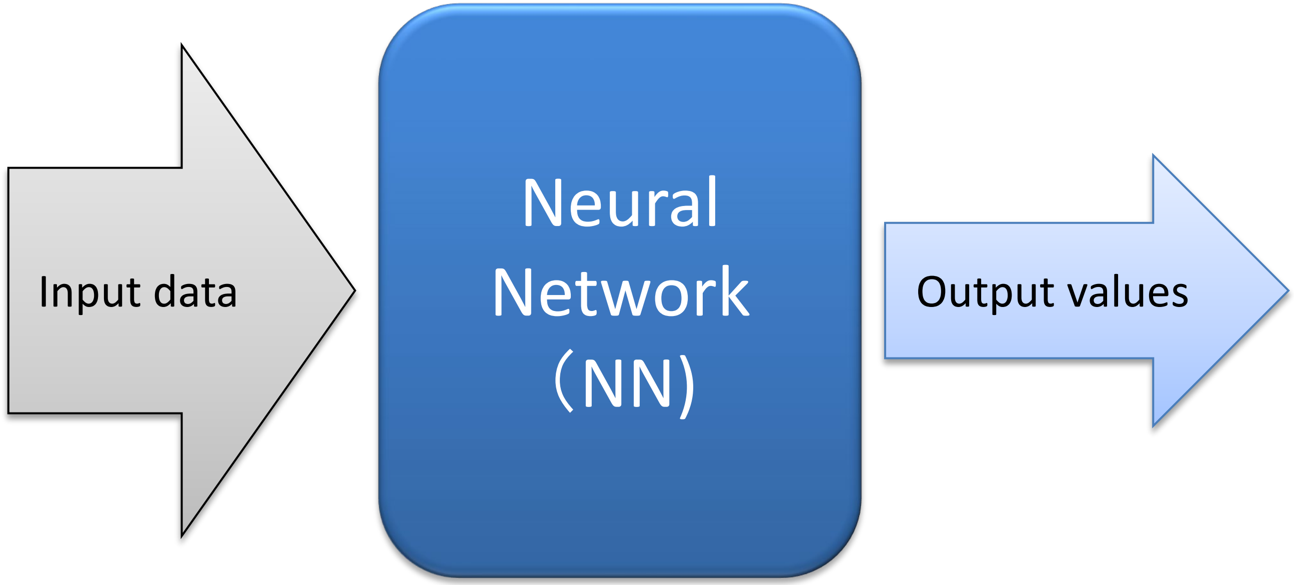
A subset of AI. Techniques that enable computers to learn the best rules to solve tasks.

Algorithms of ML

- Linear Regression
- Logistic Regression
- Decision Trees
- Naive Bayes
- Neural Network
  - Deep Neural Network
- Support Vector Machine
- ……

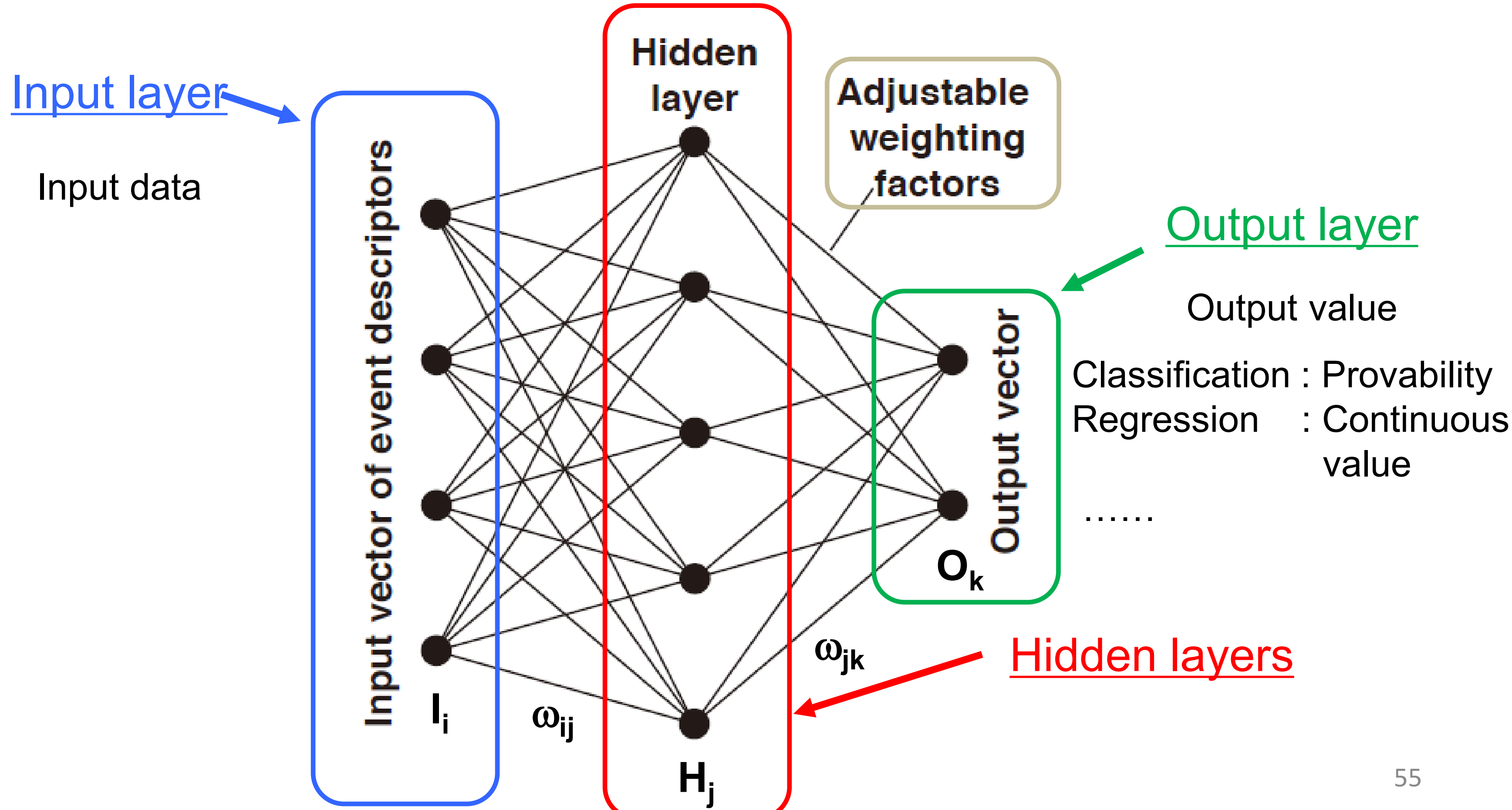
ML with Neural Network

Trained Neural Network (NN) Predict the values



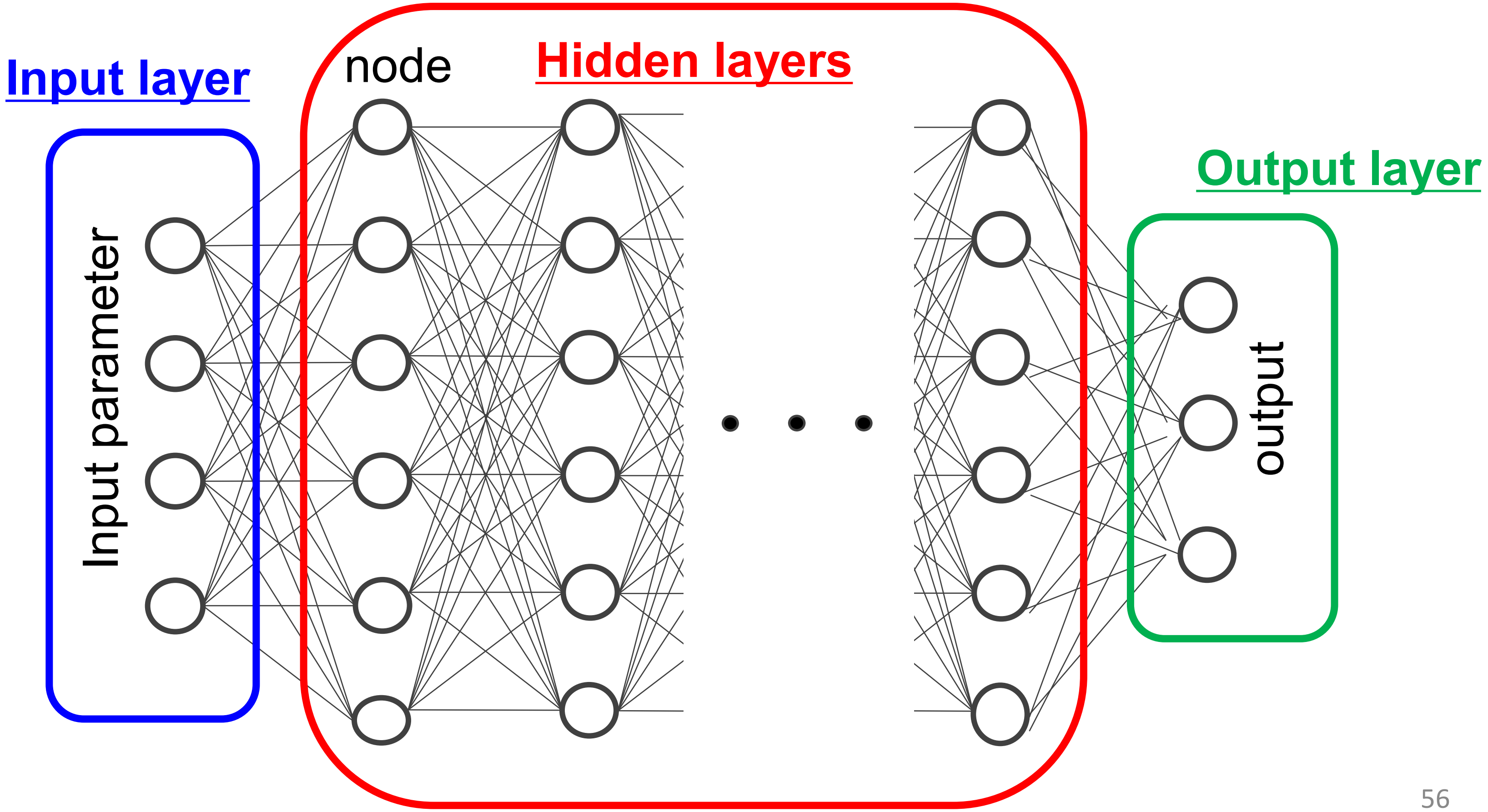
Neural Network (NN)

ML using a layered network structure, based on neuron (node) and its connection expressed in a mathematical model.
NN is consist with input, hidden, and output layers.



Deep Neural Network (DNN)

NN with many hidden layers and/or many input param.s is called DNN (Deep NN).



Deep Neural Network (DNN)

Deep learning (Deep Neural Network) enables ML using low-level data.

DNN

Many input parameters

and / or

Many hidden layers

Complicated structure NN

Expressiveness of NN is improved

Reasonably low-cost / fast computing resources

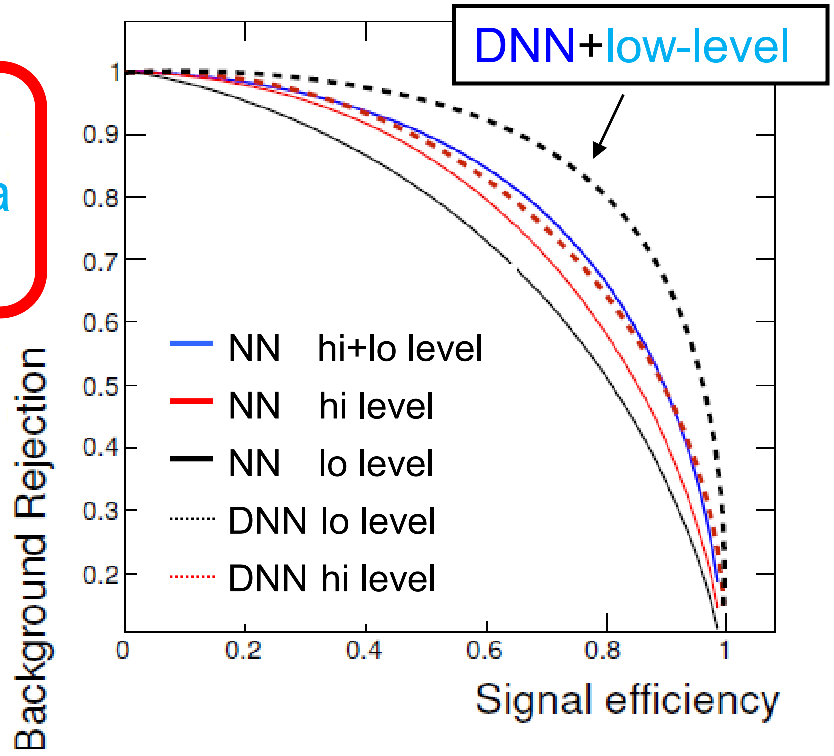
NN algorithm improvement

Deep Neural Network (DNN)

Deep learning (Deep Neural Network) enables ML using low-level data.

Recent studies show the identification performance for the DNN with low-level data is better than (shallow) NN

b-jet identification in LHC



D. Guest, K. Cranmer, D. Whiteson, Annu. Rev. Nucl. Part. Sci. **68** 1-22 (2018)

Recent ML applications in Acc. experiment

【Keywords】

- <u>Low-level feature data</u>
- <u>Various modern ML methods</u>
  Regression, Dimensionality Reduction, Reinforcement ML, RNN, CNN, …

Results shown here are preliminary.

In 2018, we form a group to proceed

"Application of Deep Learning for Accelerator Experiments"

→ *<u>As RCNP project / IDS project</u>*

The group is formed with particle physicists and data scientists

Particle Physics

Data Science

ML applications in our project

- BG suppression in Belle & Jet flavor-tag in ILC (Osaka-City U., Osaka U. IDS, RCNP)
- ILC SiD ECL energy calibration
  (Osaka-City U., Osaka U. IDS, U. Oregon, SLAC)
- Jet Clustering (U. Tokyo, Kyushu U.)
- Vertex Finding using Recurrent Neural Network (Kyushu U.)
- Jet flavor-tag in ILC (Kyushu U., Osaka U. IDS)
- Machine tuning for KEK Linac (KEK, Osaka-City U., IDS, RCNP)
- Machine tuning for RCNP Cyclotron (RCNP, IDS)
- RI Beam particle ID (Kyushu U., U. Tokyo)
- Beam size measurement in ILC (Tohoku U.)
- Lattice-QCD application (RCNP, IDS)

  …….

ML applications in our project

ML based on the Low-level feature data

- Jet flavor-tag,
  BG suppression
- Detector calibration
- Vertexing
- Clustering
- …

Accelerator tuning using ML

KEK Linac
RCNP
Kyushu U. RI beam
…

Others

(Other new R&D's are on going)

Contents

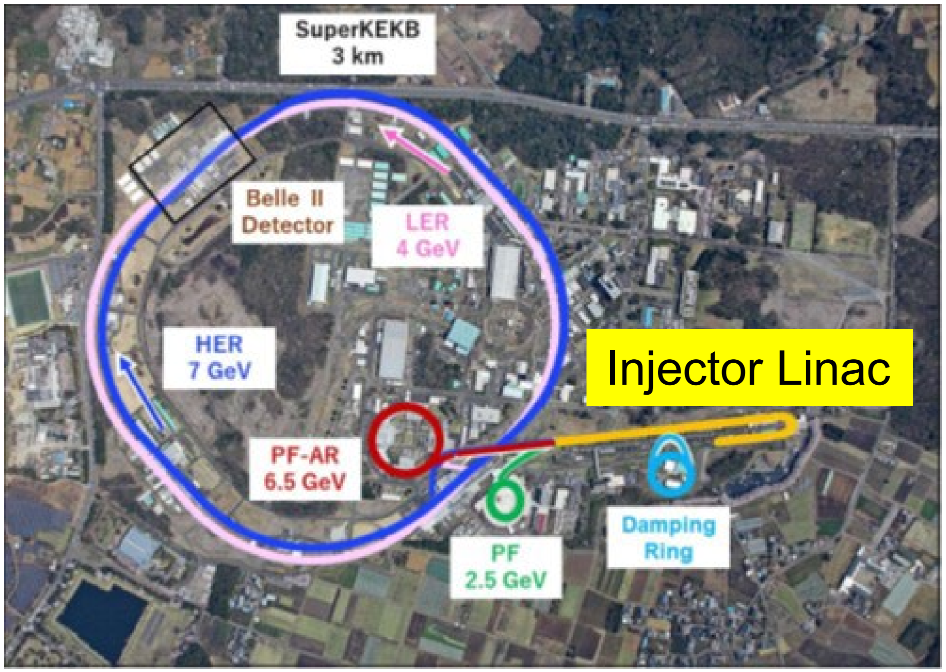
Skip if time is short

2-1: Accelerator tuning using ML

R&D of the
KEK Linac operation tuning

KEK Injector Linac Operation Tuning using ML

KEK, Osaka-City U., Osaka U. IDS, RCNP

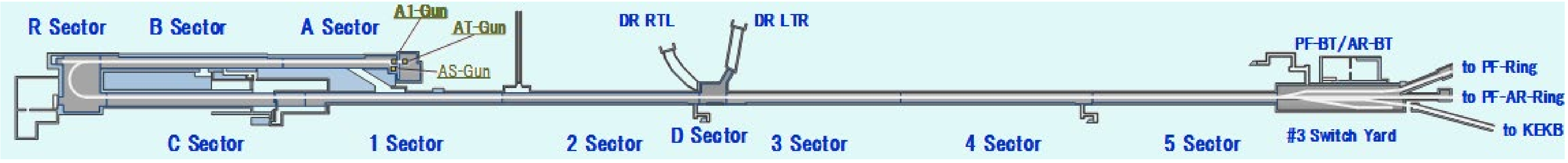


To achieve the high luminosity, operation tuning for the higher injection efficiency is important

R&D of operation tuning for the KEK injector Linac using ML is ongoing

100 Beam Position Monitors (BPM)
200 Steering Magnets, 60 RF monitors

We use the Linac operation data accumulated in 2018 Nov. - 2021 June



Linac Operation Tuning using ML

Problems on the operation tuning

1. **Operation system becomes much complicated**
   due to the large number of the control points **O(~1000)**

   → **Visualization of the accelerator status**
   by **Dimensionality Reduction**
   to model and monitor the accelerator status

2. **Accelerator condition (environment) vary**
   due to ground motion, tidal force, temperature, etc.
   Then the operation tuning is continuously done by operators.

   → **Operation tuning based on Reinforcement ML**
   to continuously optimize the parameters
   to maximize the injection efficiency

Contents

I. Visualization of the Accelerator Status

Dimensionality Reduction using unsupervised ML

II. Operation tuning based on Reinforcement ML

- Preparation Study 1

  Prediction of the injection efficiency with ML

- Preparation Study 2

  Acc. Simulator using ML

I. Visualization of the Acc. status

**To visualize the acc. Status,
we use Dimensionality Reduction
based on the unsupervised learning**

NN (VAE, Variational Auto Encoder)

Input：Acc. data

Operation param.：500
Env. param：732

NN
Dim. Reduction

Output：
Latent Variable **Z(2)**

Acc. Data
Environment monitoring data
(Temperature etc.)
~O(1000) parameters

Unsupervised Learning

Latent Variables show the feature of data

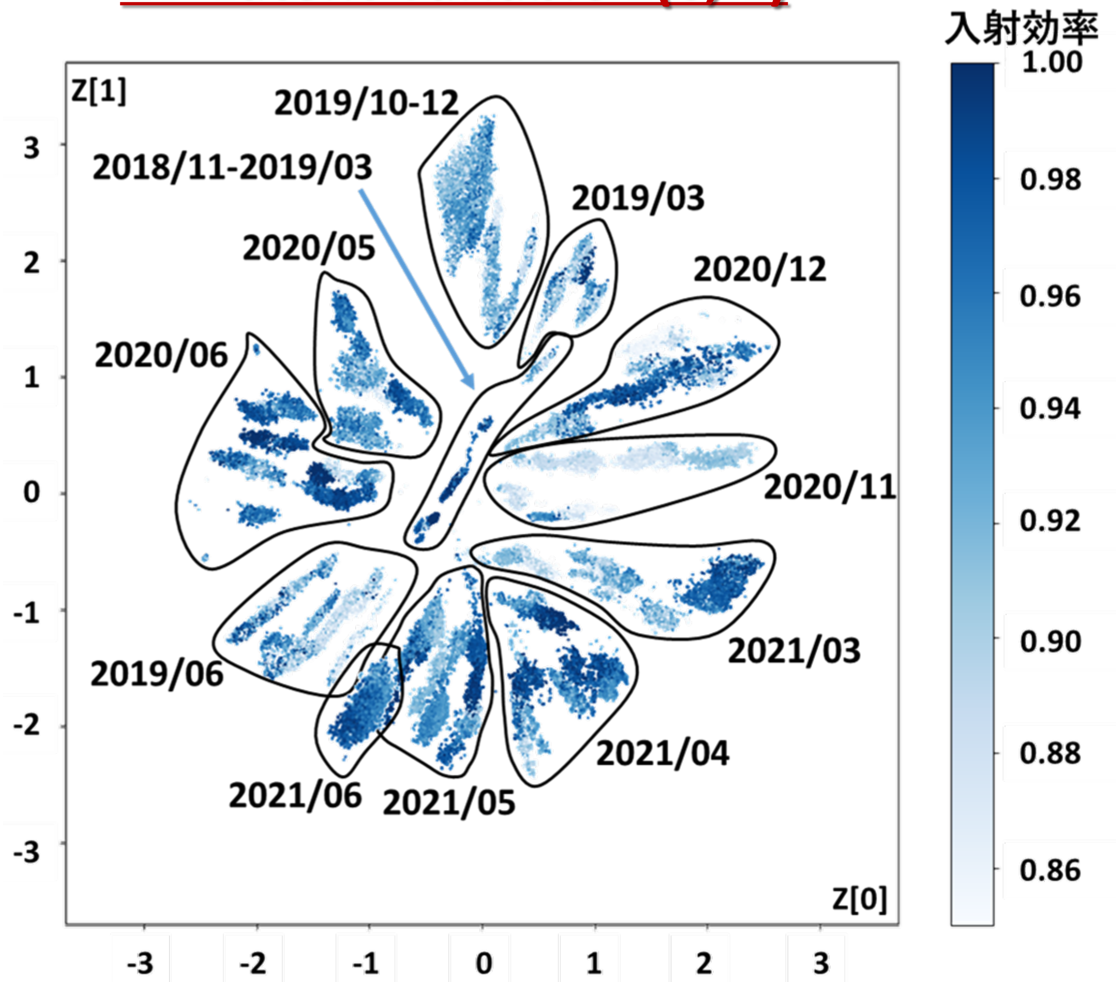
Feature Value
2D z[0],z[1]

I. Visualization of the Acc. status

We did dimensionality reduction
1295 accelerator parameters → 2 parameter Latent Variable

2D Latent Variable Z(0, 1)

A.Hisano (Osaka-City U.)



Visualize the accelerator status with 2 parameters

In short term (~1month) acc. status does not drastically change

In long term (> 6 month) acc. parameters vary over a wide range

II. Operation Tuning based on Reinforcement ML

To continuously adapt to the environmental changes
to get the high injection efficiency,
Reinforcement ML is a good candidate

To apply the Reinforcement ML to the operation tuning, we need

- DNN which can express the relationship between accelerator parameters and injection efficiency, and
- Tools for the Pre-training of DNN

We have carried out the preparation studies of

1. **Injection efficiency prediction using ML**
   To see if NN can express the injection eff. from acc. parameters
2. **Accelerator Simulator using ML**
   For the “pre-training” ,
   so that no actual operation required for optimization

Operation Tuning using ML：Injection eff. prediction

KEK, Osaka-City U., Osaka U. IDS, RCNP

We use DNN regression to predict the injection eff. from acc. Parameters.
The DNN is trained with the past acc. Operation data.

Input data → NN (Regression) → **Injection eff. (Q_{ratio})**

Accelerator Operation parameters /
Environment monitoring data
(Temperature etc.)
~O(1000) Parameters

Hidden Layers : 4
Optimizer: Adam
Activation Function: Relu

We use the Linac operation data accumulated in
2018 Nov. - 2021 June

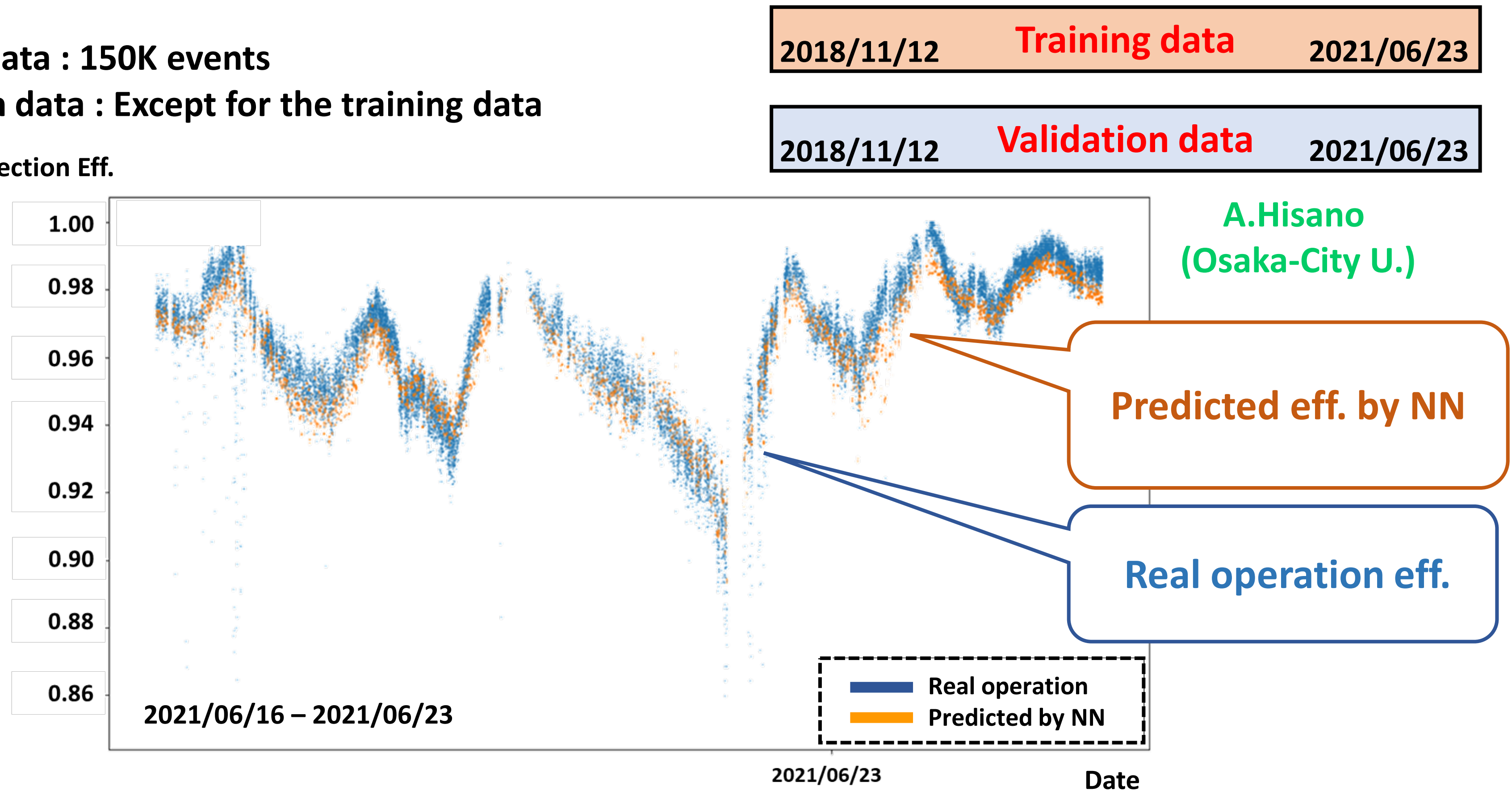
Operation Tuning using ML：Injection eff. prediction

KEK, Osaka-City U., Osaka U. IDS, RCNP

Result : 1

Training / Validation data → Accumulated in 2018/11 to 2021/06

Training data : 150K events
Validation data : Except for the training data



DNN can predict the injection eff.
→ DNN can express the relationship btw Acc. parameters and injection eff.

Operation Tuning using ML：Injection eff. prediction

KEK, Osaka-City U., Osaka U. IDS, RCNP

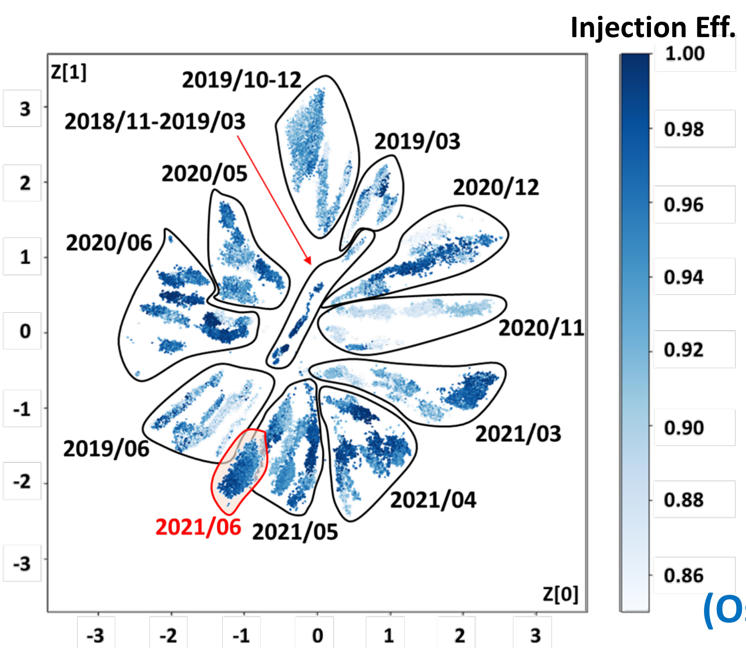
Result : 2

DNN is trained with the "past data"

2018/11/12 **2021/05/26** **2021/06/16** **6/23**

Data ①	**Training data** (2018/11/12 – 2021/05/26)	**Val. data** (2021/06/16 – 6/23)
Data ②	**Training data** (2021/04/28 – 2021/05/26)	**Val. data**
Data ③	**Training data** (2018/11/12 – 2021/04/28)	**Val. data**

2021/04/28



2D Latent value for 2021, June is close to 2021, May

→ 2021, June injection efficiency might be predicted using the DNN trained with 2021, May data.

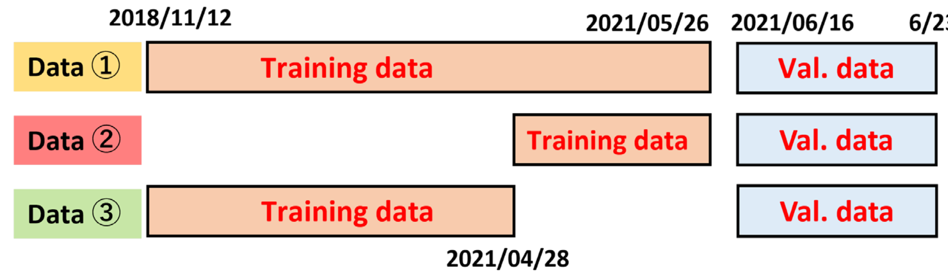
A.Hisano
(Osaka-City U.)

Operation Tuning using ML：Injection eff. prediction

A.Hisano (Osaka-City U.)

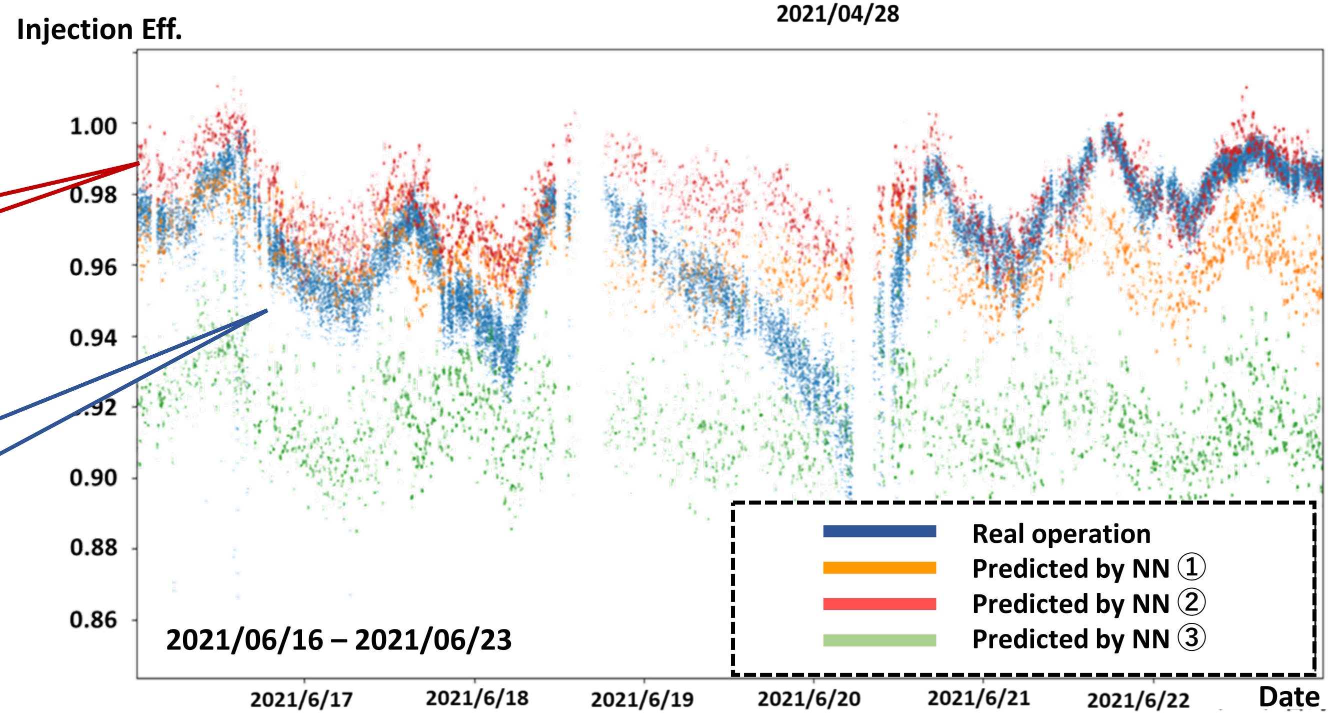
Result : 2

Validation data → 2021, June
Training data : 150K events each



Predicted eff. by NN

Real operation eff.



DNN trained with 2021, May data (①②) can predict 2021, June injection eff.

Training based on the current recent data is important

II. Operation Tuning based on Reinforcement ML

Preparation study 2 : Acc. Simulator using ML

For the **pre-training** of the reinforcement ML,
It is necessary to use the **acc. simulator**

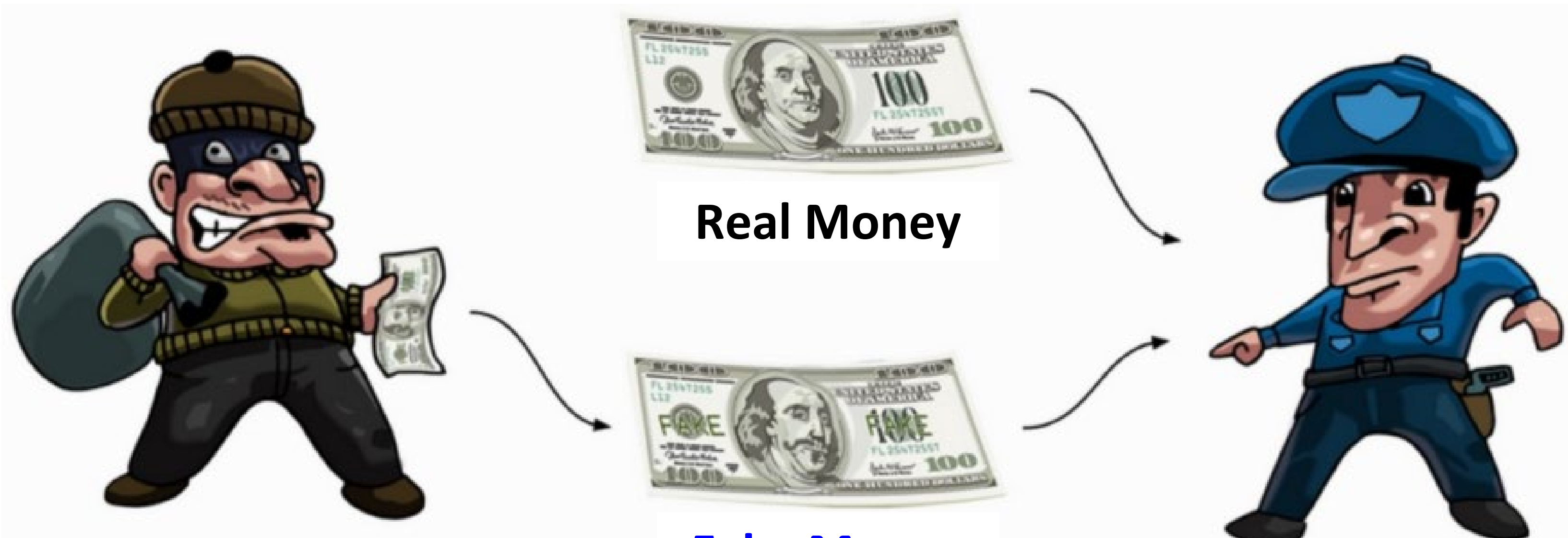
We have developed the acc. simulator using ML (GAN)

Acc. Simulator using GAN

GAN : Generative Adversarial Network

Learns to generate "fake data" that has the similar function as "real data"

Consists of 2 networks, Generator and Discriminator that compete with each other.



Fake Money

Generator
is trained to generate data which can fool the Discriminator.

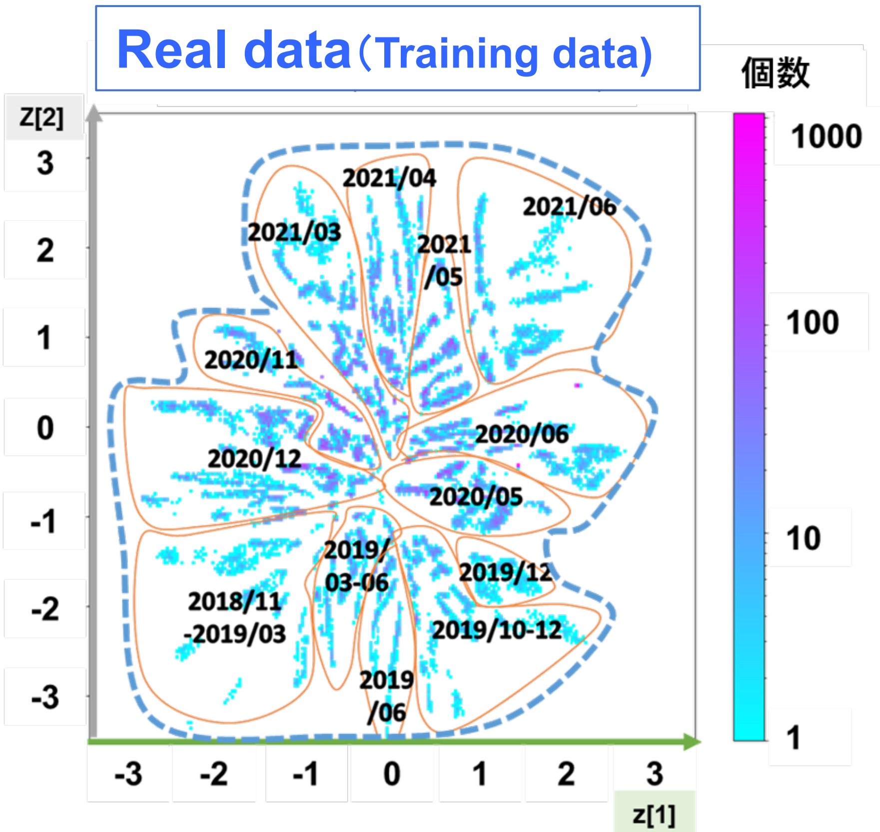
Discriminator
is trained to distinguish between fake and real data.

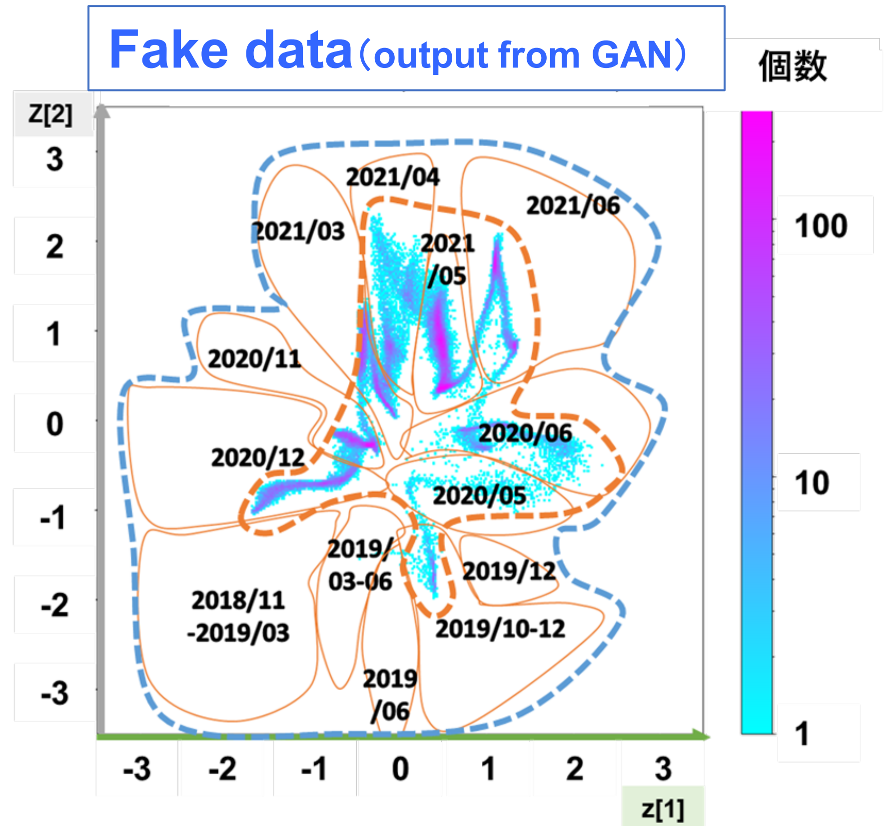
https://dzone.com/articles/working-principles-of-generative-adversarial-netwo

Acc. Simulator using GAN

A.Hisano
(Osaka-City U.)

**Generate the fake data
from the real Acc. Operation data (1233 params.)**



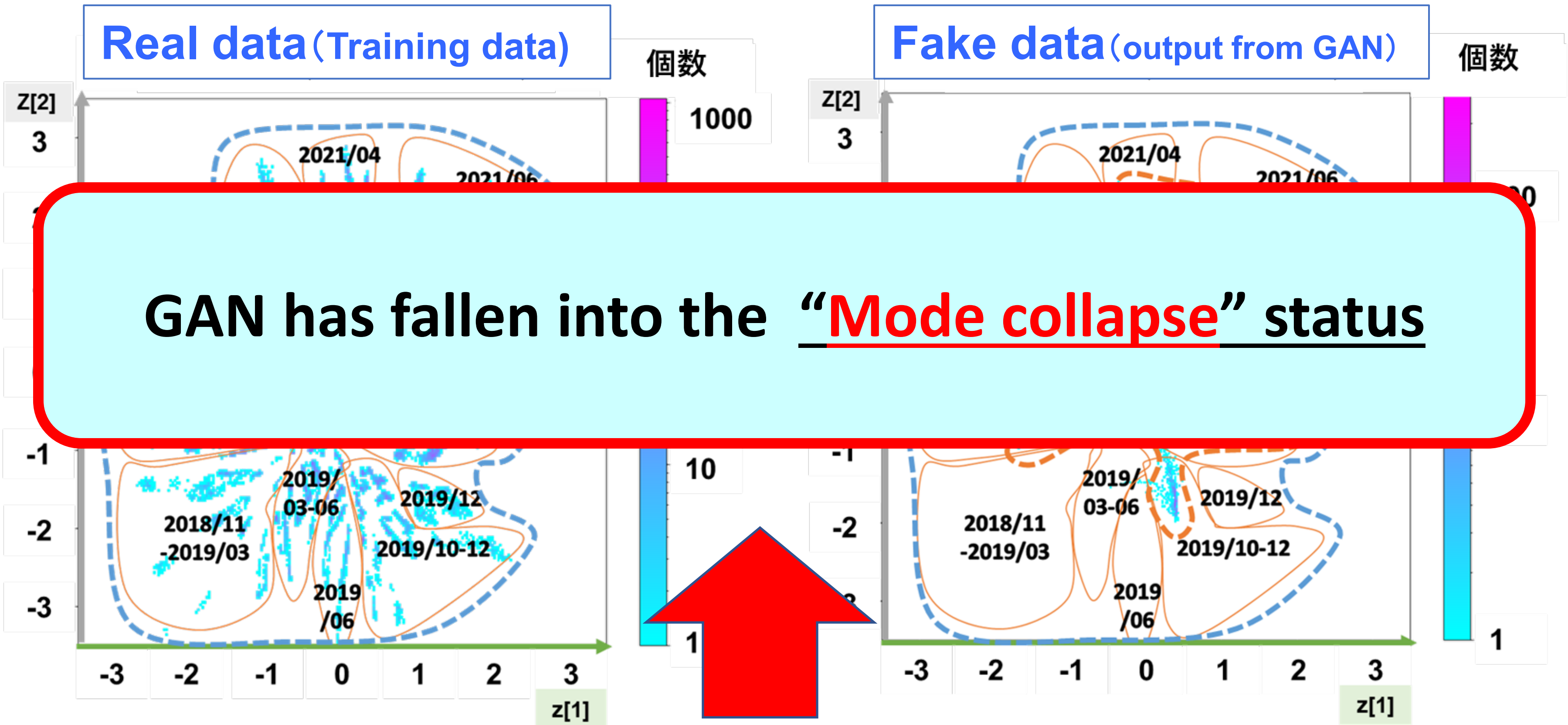


**GAN output data don't cover
the full parameter ranges of the input data**

Acc. Simulator using GAN

A.Hisano
(Osaka-City U.)

Generate the fake data
from the real Acc. Operation data (1233 params.)



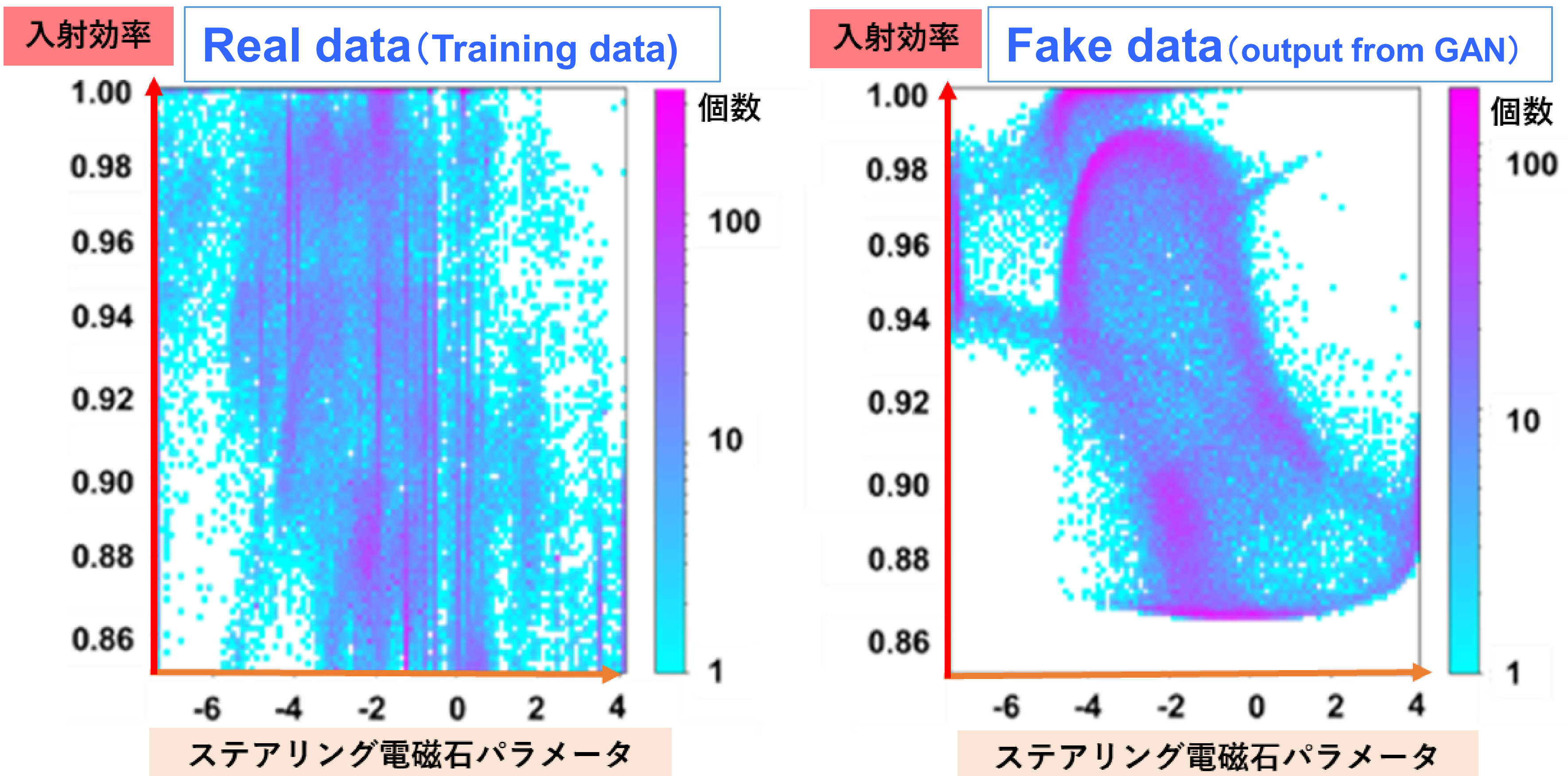
GAN has fallen into the <u>“Mode collapse” status</u>

GAN output data don’t cover
the full parameter ranges of the input data

Acc. Simulator using GAN

A.Hisano
(Osaka-City U.)

Generate the fake data
from the 2 params. Acc. Operation data



"Mode collapse" has occurred

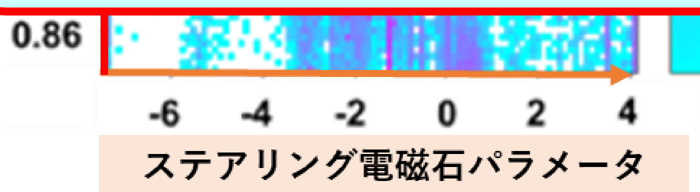
Acc. Simulator using GAN

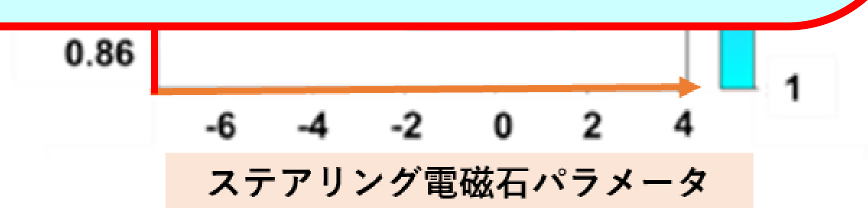
A.Hisano
(Osaka-City U.)

Generate the fake data
from the 2 params. Acc. Operation data

Future Plan

1. **Develop the acc. Simulator which can work without mode collapse**
2. **Develop the Pre-training method without the acc. simurator**





"Mode collapse" has occurred

2-2: R&D of the analysis method using DNN with low-level data

- BG suppression
- Jet flavor-tag

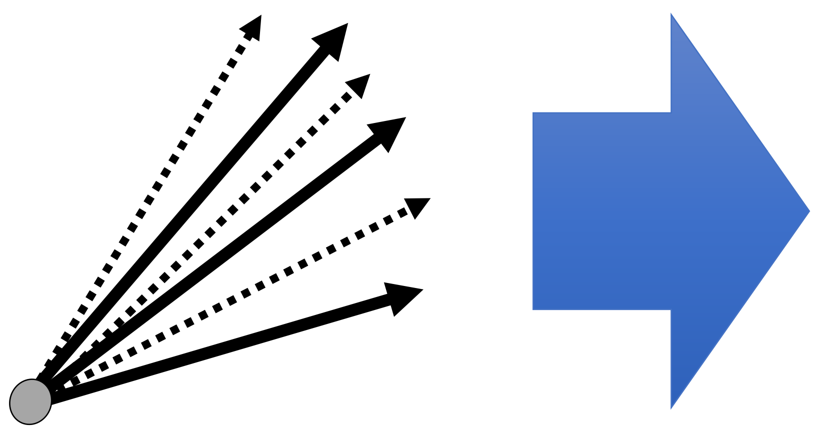
DNN application to the HEP Experiments

In data science, the modern ML technology such as DNN has been developed for analysis of image, natural language, etc.

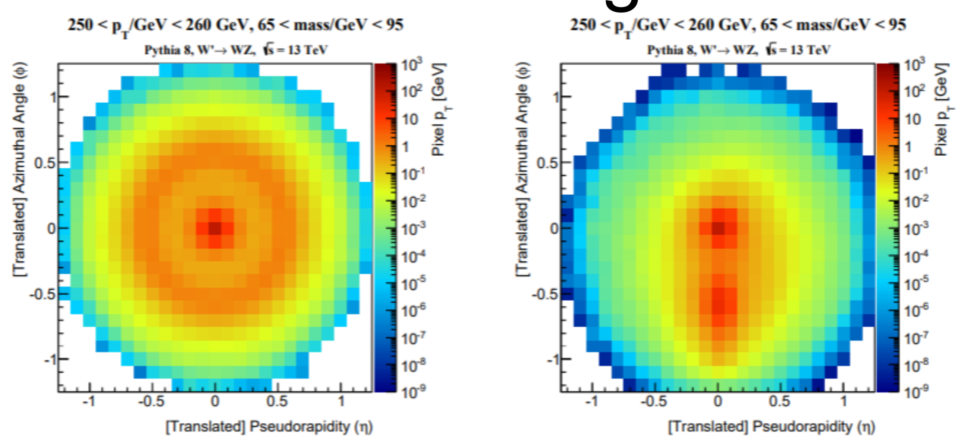
→ Many tools exist for image processing, natural language analysis..

In many HEP experiment cases,
the HEP experimental data is converted to **"image data"**
to use the modern ML tools with low-level data, e.g. CNN.

Particle Jet



Jet as image



D. Guest, K. Cranmer, D. Whiteson, Annu. Rev. Nucl. Part. Sci. **68** 1-22 (2018)

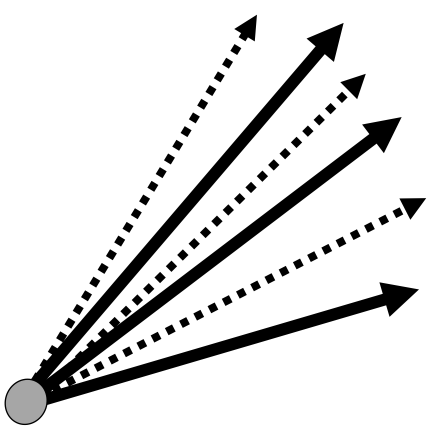
DNN application to the HEP Experiments

Problems

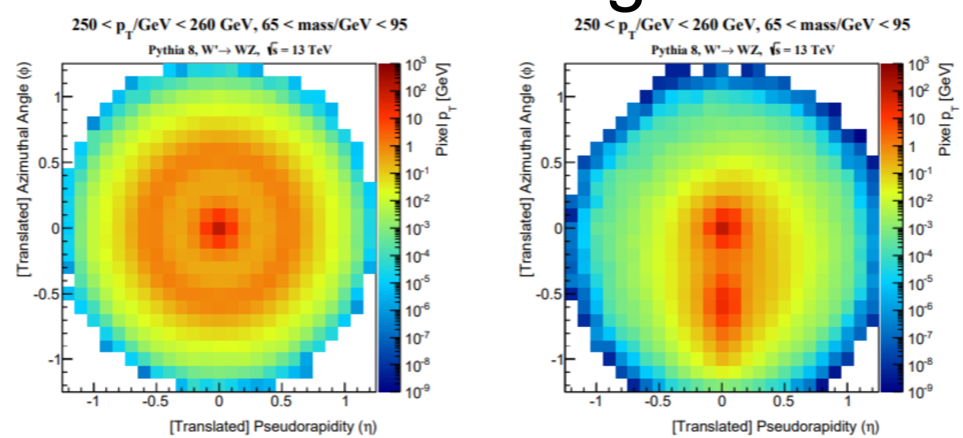
If we convert the HEP exp. data to the image data,
- **Lose position resolution**
- In the HEP colliding experiment, arbitrary number of particles, with arbitrary direction, energy... are produced

→ **Huge amount of training data is necessary**

Particle Jet



Jet as image



D. Guest, K. Cranmer, D. Whiteson, Annu. Rev. Nucl. Part. Sci. **68** 1-22 (2018)

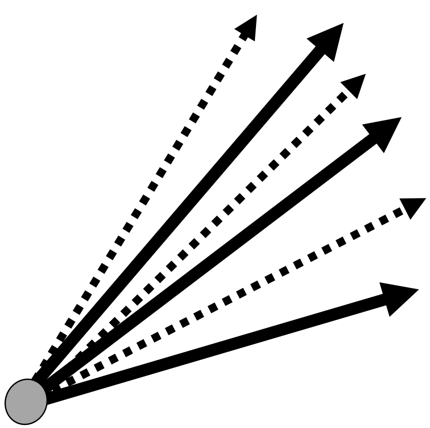
New Analysis Method using DNN with Low-level Data

Osaka-City U., Osaka U. IDS, RCNP, Kyushu U.

We have developed a new analysis method using DNN with low-level data for HEP experiments

We directly input the 4-momentum and position of particles, as low-level data, to DNN

Particle Jet



Jet as low-level data

Part1 (E, Px, Py, Pz, x, y, z)
Part2 (E, Px, Py, Pz, x, y, z)
Part3 (E, Px, Py, Pz, x, y, z)
……

New Analysis Method using DNN with Low-level Data

Osaka-City U., Osaka U. IDS, RCNP, Kyushu U.

We have developed a new analysis method using DNN with low-level data for HEP experiments

We directly input the 4-momentum and position of particles, as low-level data, to DNN

We apply the method to

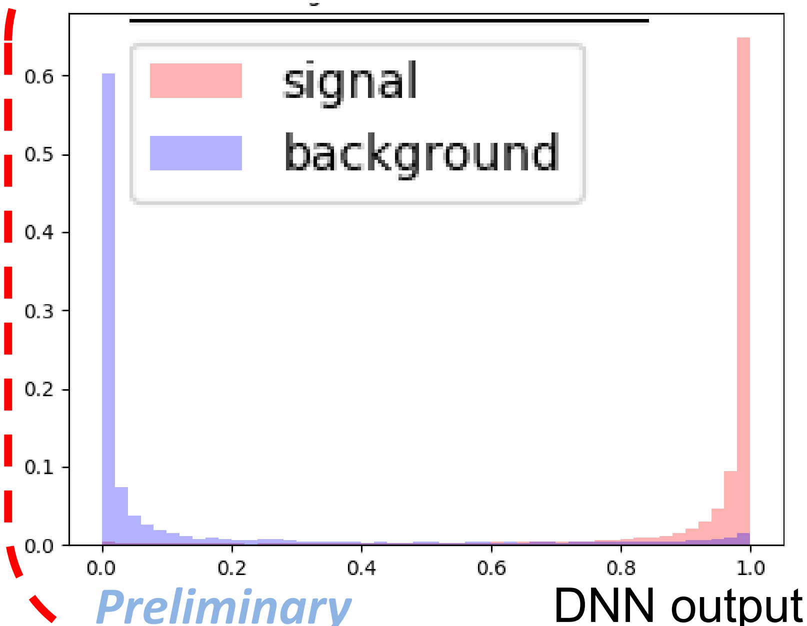
1. BG suppression in Belle, and
2. Jet Flavor-tag in ILC

Result1: BG Suppression in Belle

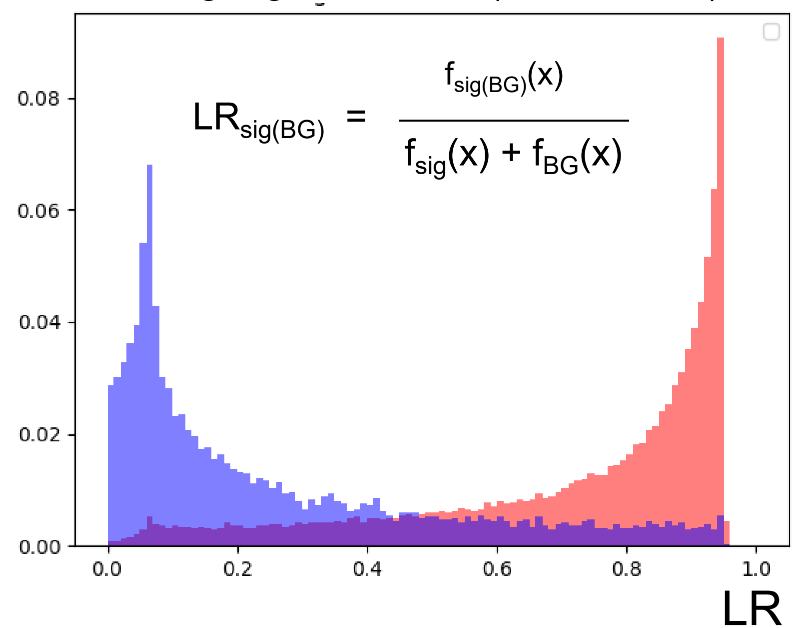
N.Kishida (Osaka City U.)

We apply DNN with low-level data input to the Signal / BG classification

DNN Output



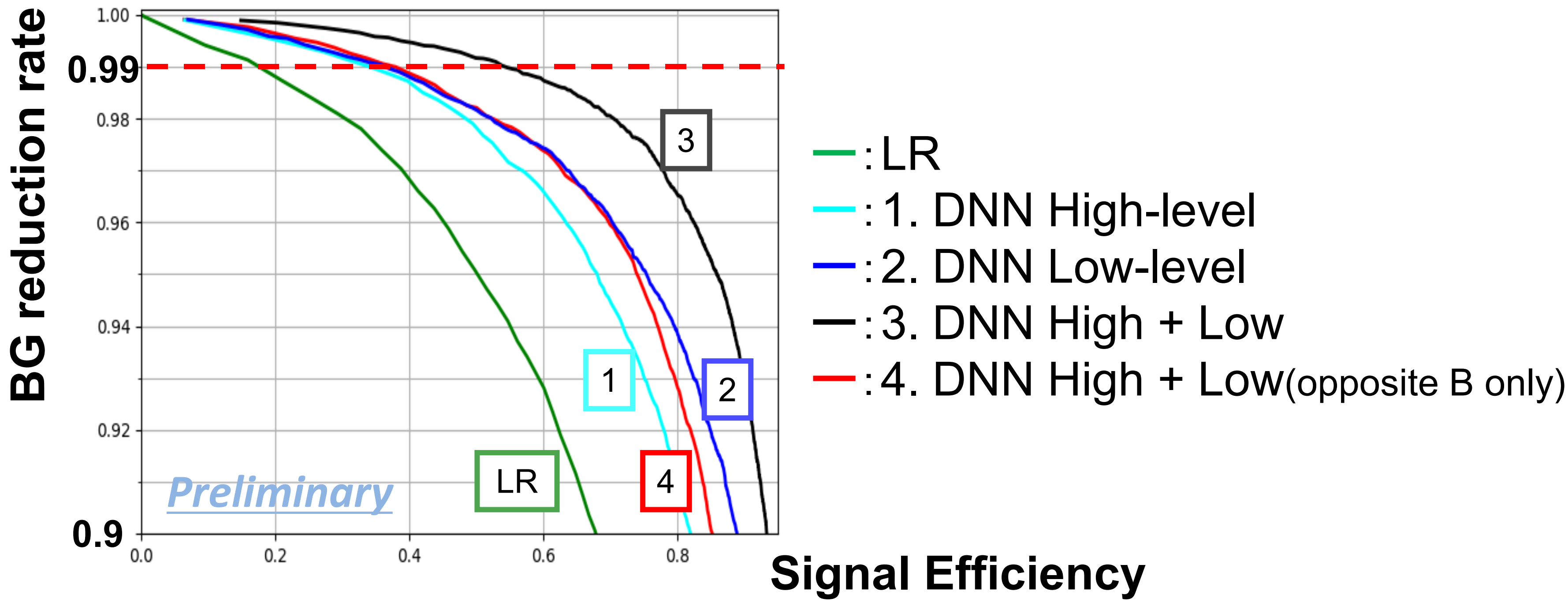
Likelihood Ratio (LR)
using High-level data (cosθ, KSFW)



DNN provides better performance than the conventional method (LR)

Result1: BG Suppression in Belle

N.Kishida (Osaka City U.)



- DNN provides better S/BG classification performance than LR
- DNN with Low-level data is better than DNN with High-level data
- DNN with Low- + High–level data provides the best performance
- We obtain x2 Signal efficiency at BG reduction rate of 0.99

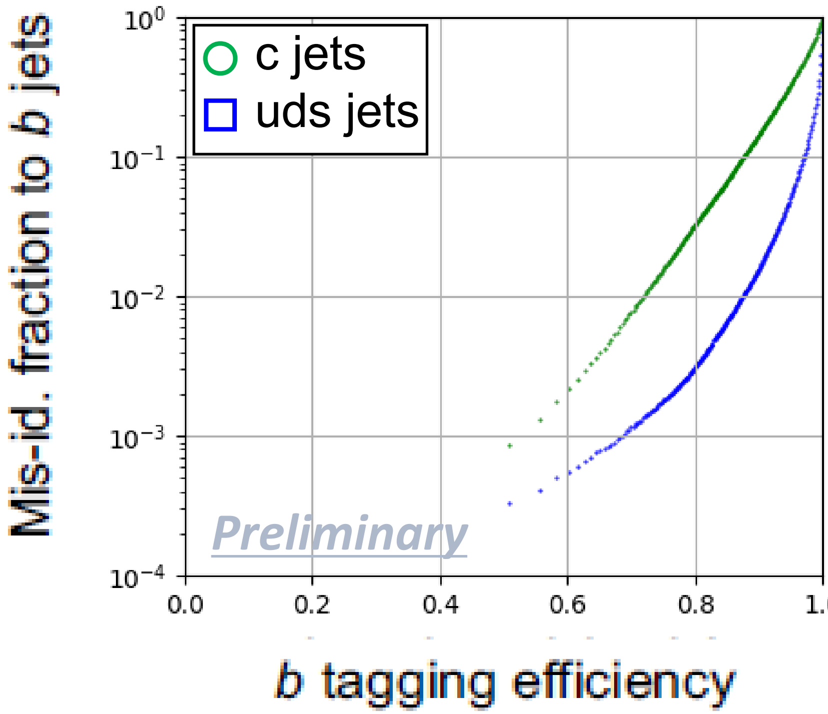
Result2: Jet flavor-tag in ILC

N.Kishida (Osaka City U.)

We apply DNN with low-level data input to the Flavor classification (uds/c/b)

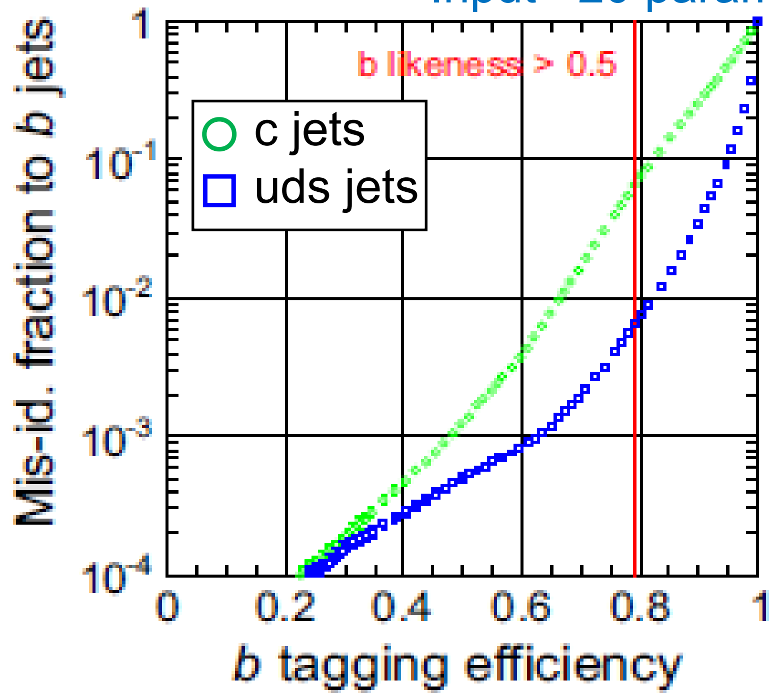
- **b-jet** identification performance

DNN with Low + High level data



BDT with High-level data※

Input ~20 params



※ : T.Suehara T. Tanabe NIM A808(2016) 109

~1/3 mis-id fraction
with the same b-tagging efficiency

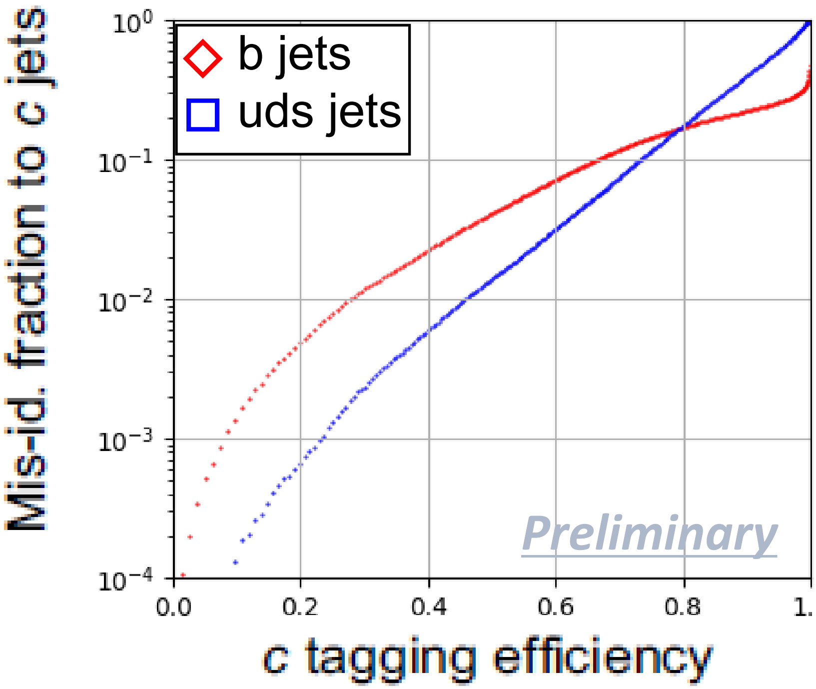
Result2: Jet flavor-tag in ILC

N.Kishida (Osaka City U.)

We apply DNN with low-level data input to the Flavor classification (uds/c/b)

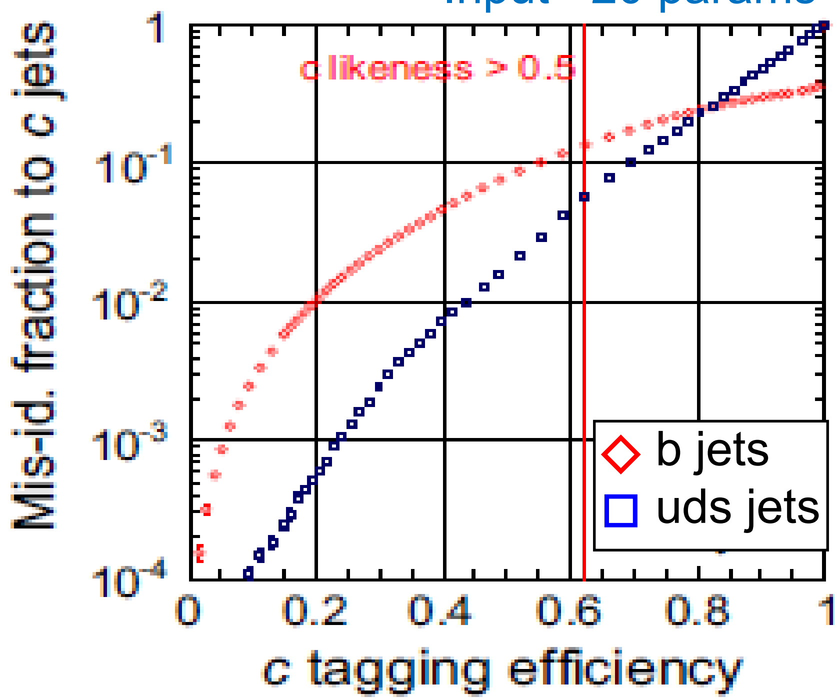
■ c-jet identification performance

DNN with Low + High level data



BDT with High-level data※

Input ~20 params

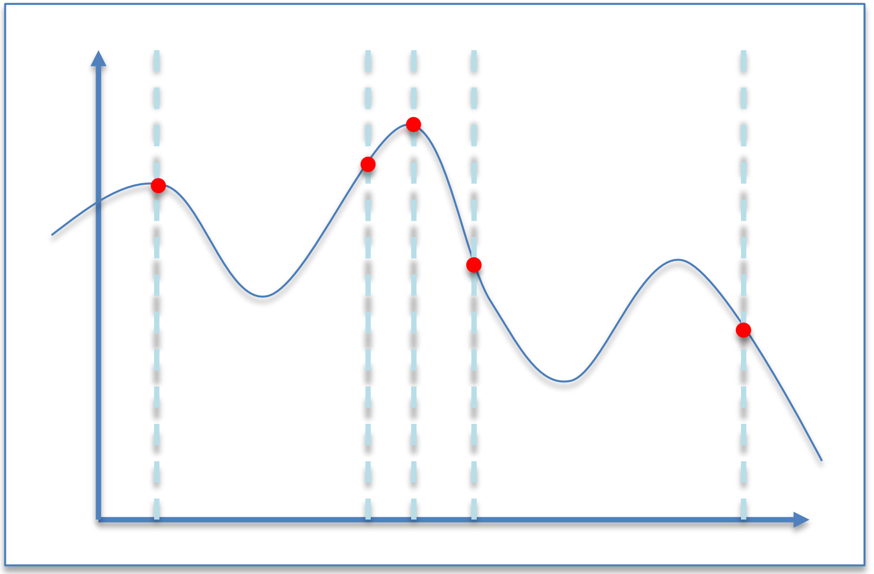


※ : T.Suehara T. Tanabe NIM A808(2016) 109

~1/2 mis-id fraction
with the same c-tagging efficiency

Skip if time is short

2-3: Data Reduction based on the Sparse Sampling



DAQ in Collider experiment

Challenges

High Trigger rate e.g. up to **30kHz** in Belle II

Big data size # of detector sensors **O(10^3) ~O(10^4) or more**

We cannot record the data exceed the DAQ bandwidth

e.g. **~30 GB / sec** in Belle II experiment

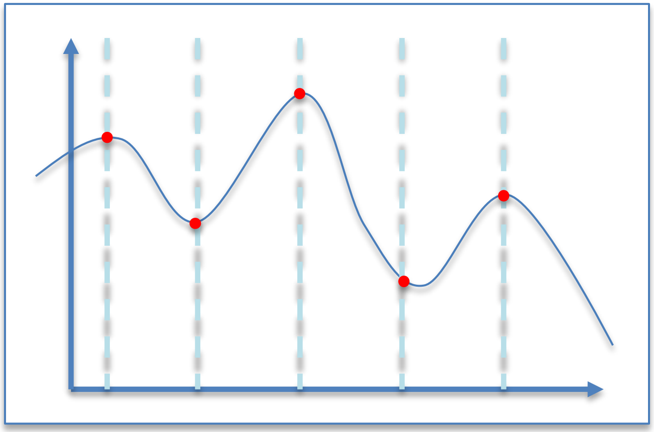
If the data size can be reduced in real time, we can accumulate more signal events

We have developed the real time data reduction based on the Sparse Sampling

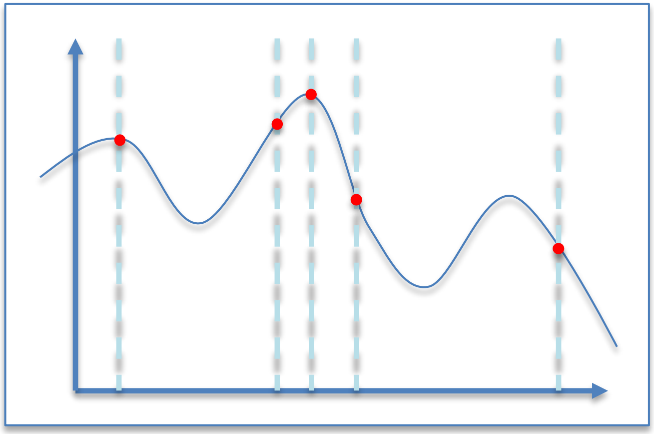
Sparce Sampling and Reconstruction

Sparce Sampling + Reconstruction

← Method to "measure" the Blackhole



Normal sampling



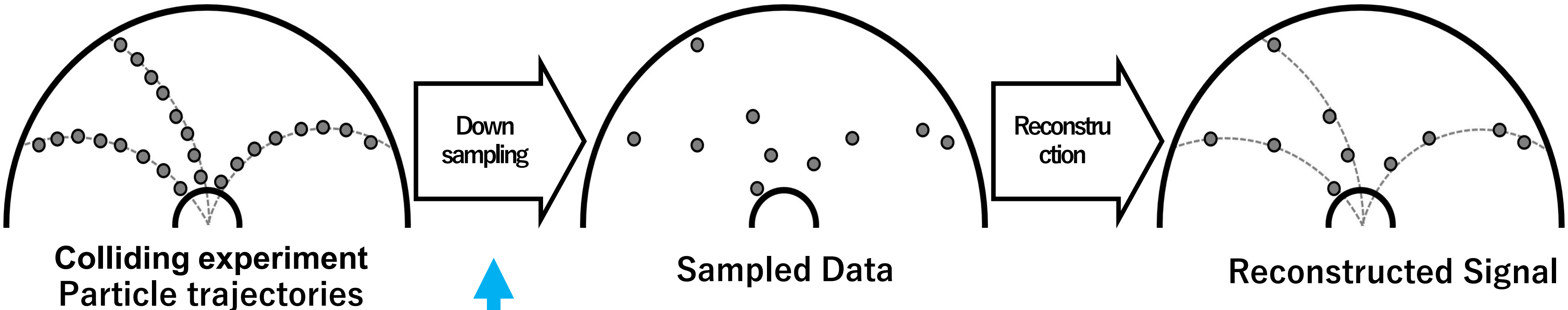
Down sampling (Sparce Sampling)

If sparsity can be assumed,
the sparce sampled data can be reconstructed

Sparce Sampling and Reconstruction

Application to the collider experiments

We currently try to apply the sparce sampling + reconstruction method to the colliding experiment DAQ

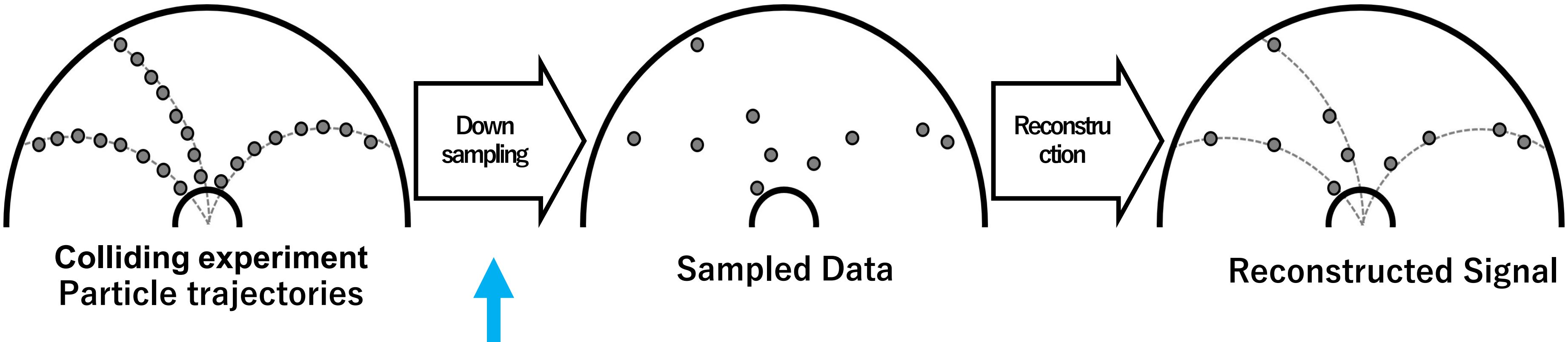


In DAQ, the necessary process is only
“Sampling according to a predetermined pattern”
→ We can reduce the data size **in real time**

Sparce Sampling and Reconstruction

Application to the collider experiments

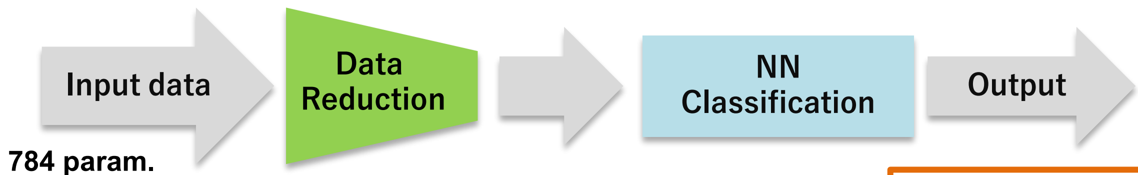
We currently try to apply the sparce sampling + reconstruction method to the colliding experiment DAQ

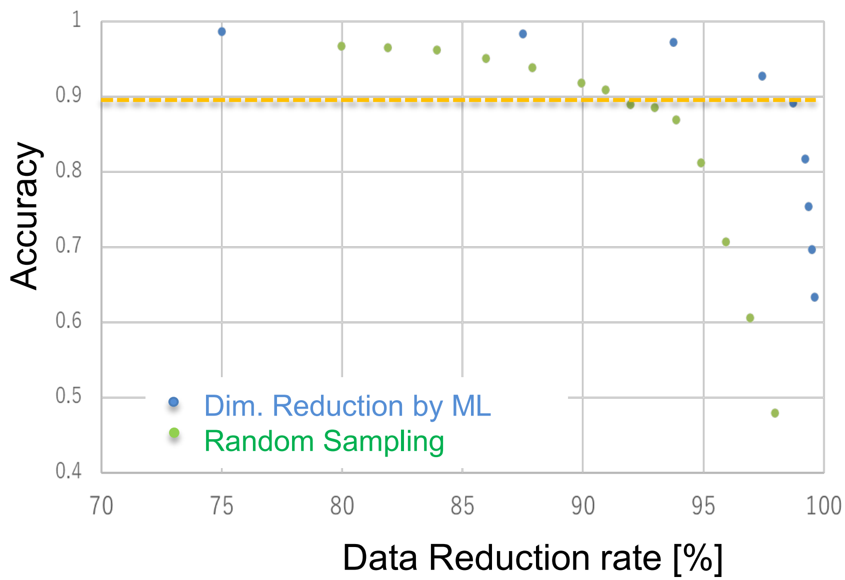


Using ML, we try to optimize the sampling pattern which does not loose / distort the essential information

Character Classification Using Dimensionality Reduction

As a basic study of the Space Sampling + Reconstruction, we firstly try the Character Classification using Data reduction (dim. reduction / random sampling)





It is possible to keep 90% of the ID performance, after data compression to 1/10 (random sampling) 1/50 (Dim. Reduction by ML)

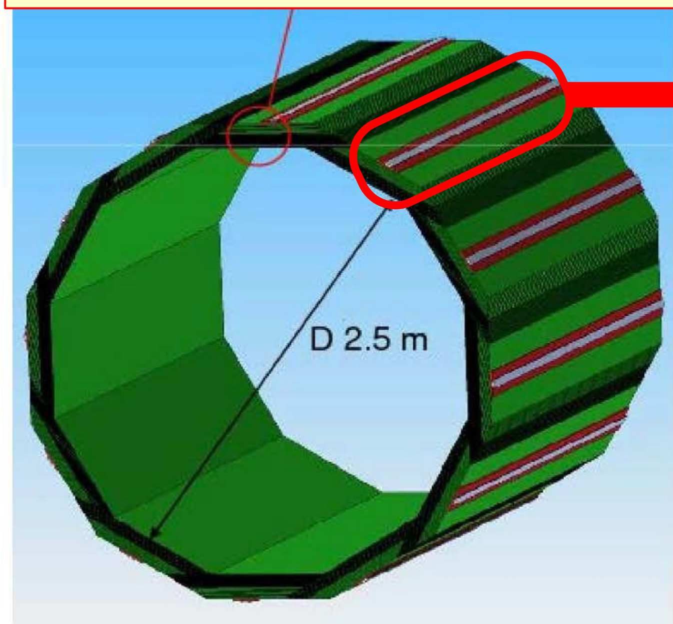
Nest Step: Apply to the Collider exp.

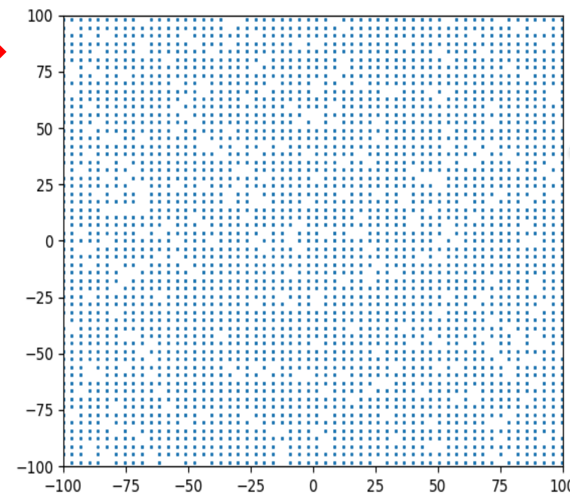
C.Kato (Osaka City U.)

First Trial : ILC SiD EM Calorimeter Data reduction

C.Kato (Osaka City U.)

- **30 Layer Si + W sampling calorimeter**
- ~26X_0 in total
- Energy resolution(design value) $(17/\sqrt{E}\oplus 1)\%$





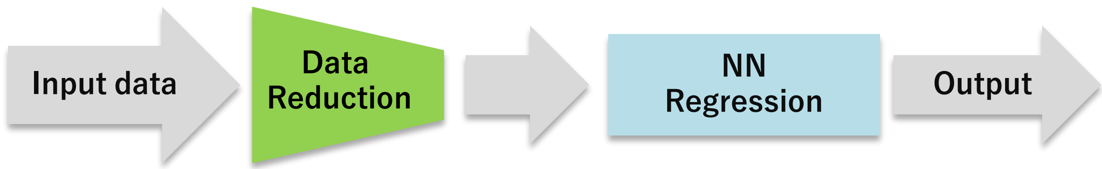
Randomly reduced 10% sensor

ILC TDR, vol.4,Page 89
arXiv:1306.6329[physics.ins-det]

First Trial : ILC SiD EM Calorimeter Data reduction

C.Kato (Osaka City U.)

Compare the Energy calibration performance
With and without the Data Reduction (random sampling)



Resolution of the Energy Calibration (for electrons)

Electron Energy	2GeV	5GeV
w/o reduction	20.2%	20.5%
10% reduction	19.2%	20.1%

Energy Calibration performance does not change,
but we need more data statistics to evaluate it.

Summary

- Belle II produces interesting and competitive results with a little data already
- We will have more data by this summer. PXD2 will be installed in 2022-2023.
- Modern ML methods are powerful tools for physics experiment big data processing
  → Several R&Ds are on going

Stay Tuned!

Thank you!!

ありがとう
ございました