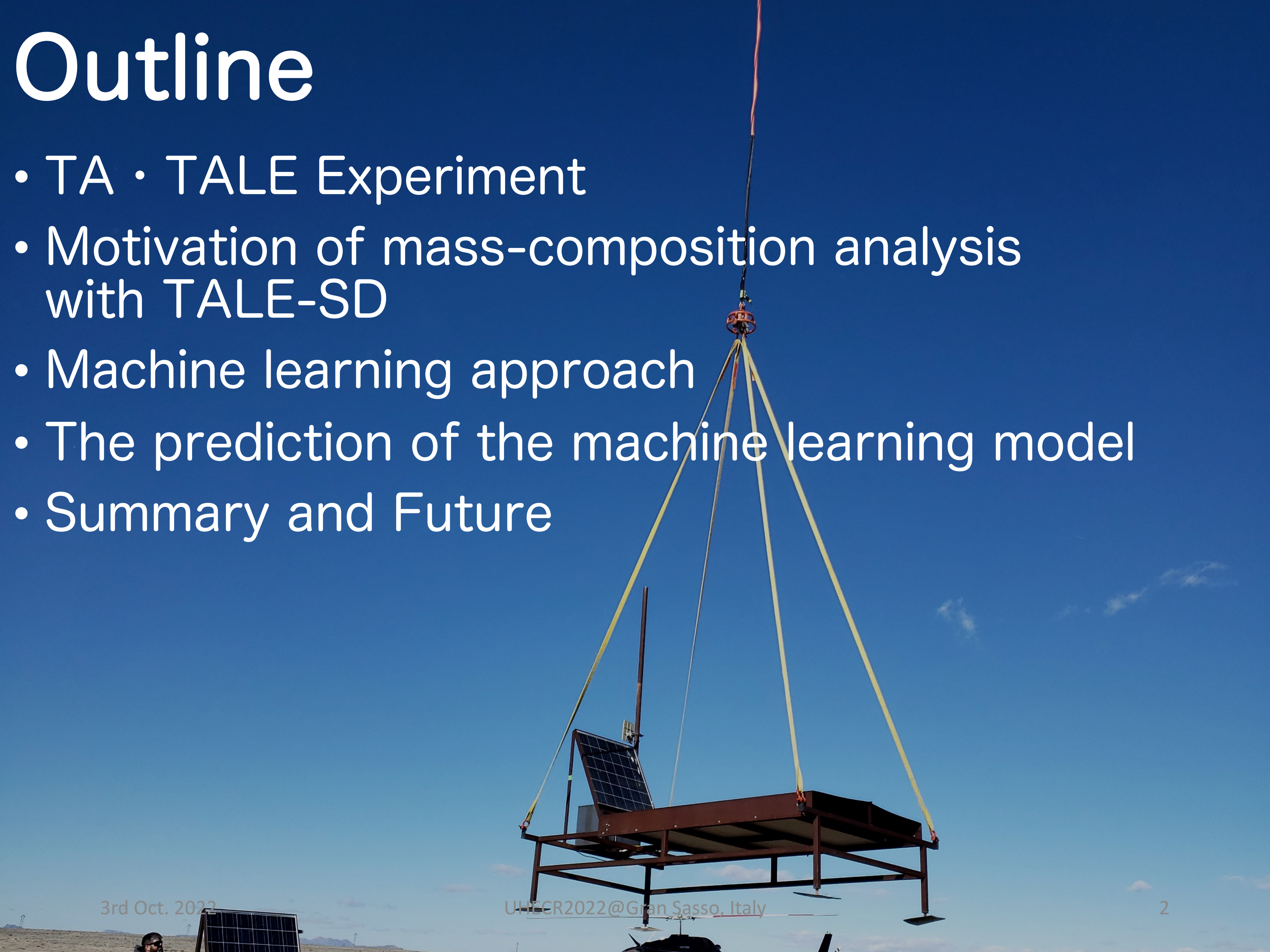
A machine learning approach for mass composition analysis with TALE-SD data

Ryuhei Arimura
Osaka Metropolitan University
for the Telescope Array Collaboration

October 3rd, 2022
UHECR 2022

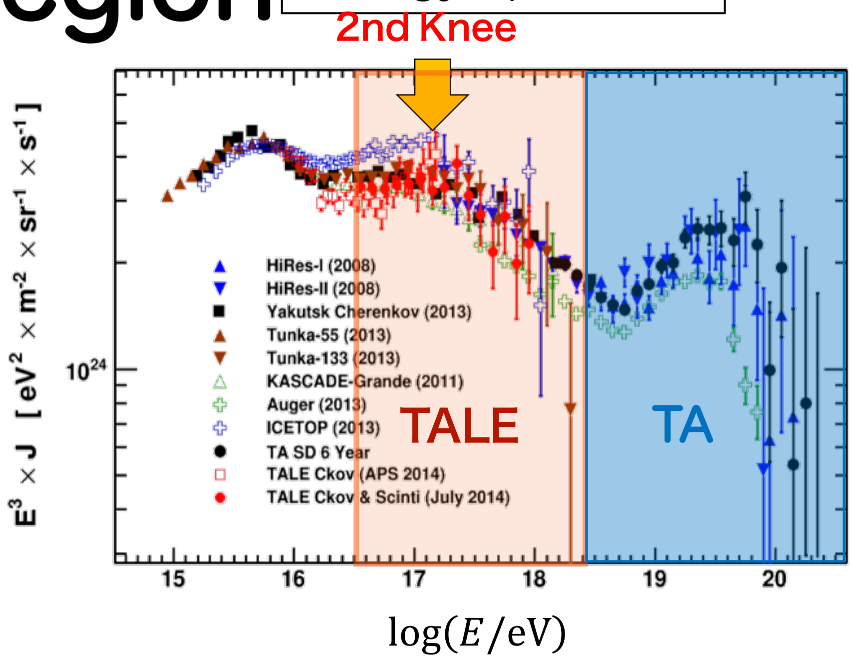
Outline

- TA・TALE Experiment
- Motivation of mass-composition analysis with TALE-SD
- Machine learning approach
- The prediction of the machine learning model
- Summary and Future

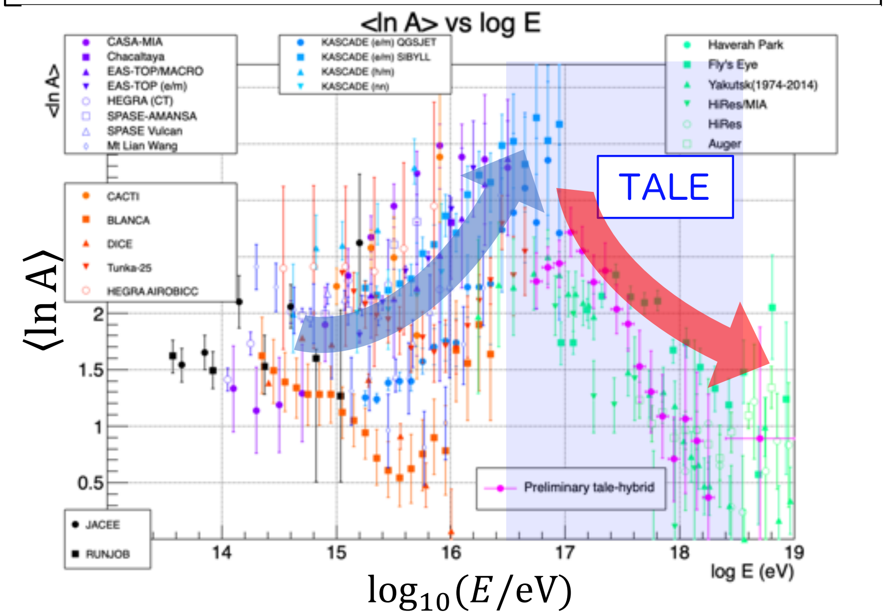


Cosmic ray around the 2nd knee region

Energy spectrum



Transition of mass-composition

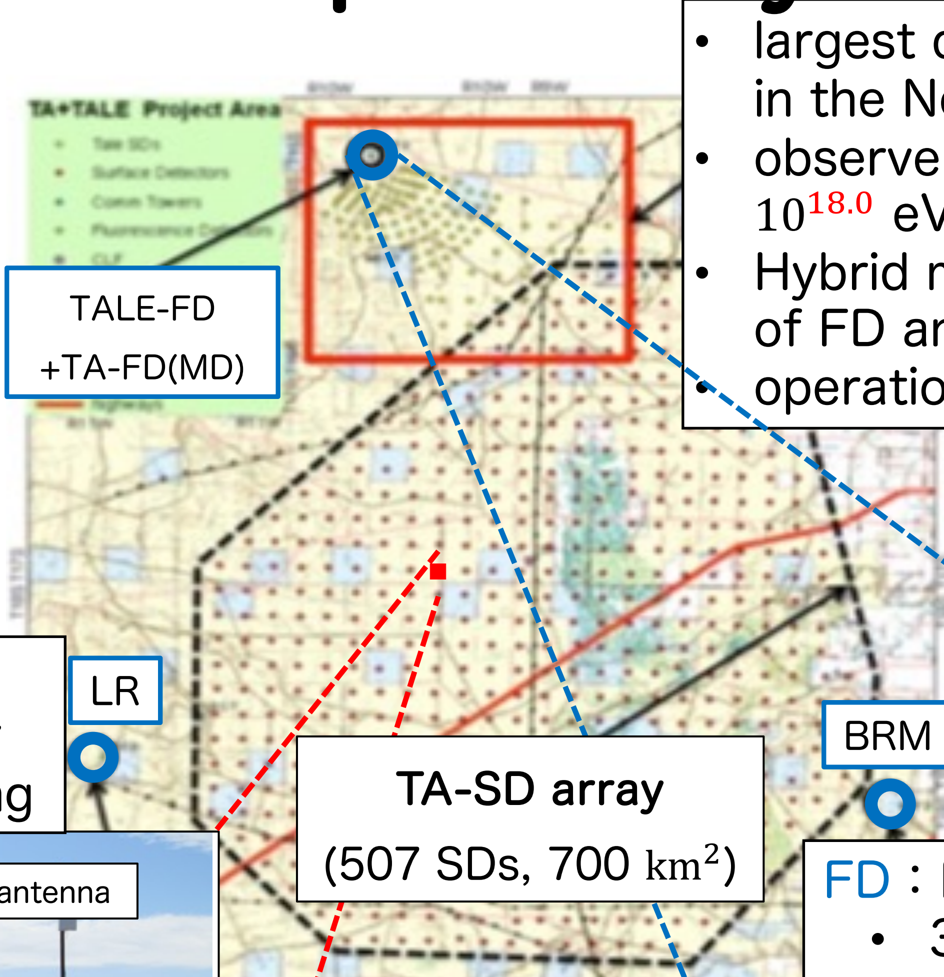


Origin of cosmic rays around the 2nd knee region($\sim 10^{17}$ eV)
➡ galaxtic ↔ extragalactic transition?

Precise measurement of the energy spectrum and mass-composition of cosmic rays in the 2nd knee region by fluorescence detector and surface detector

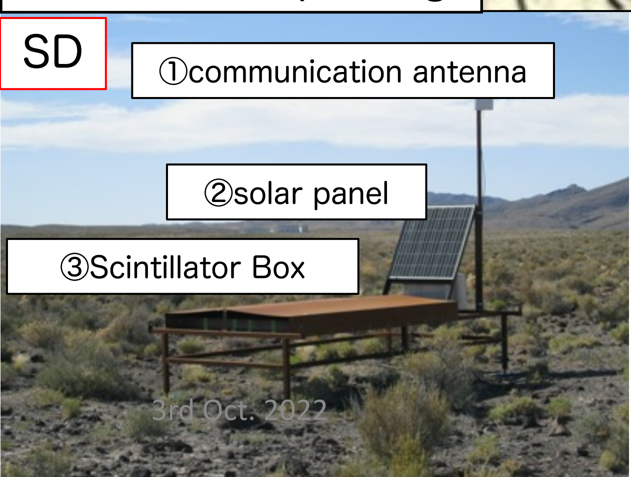
TA→Telescope Array Experiment

- largest detected area in the Northern Hemisphere
- observe cosmic rays above $10^{18.0}$ eV in Utah (U.S.A.)
- Hybrid measurement of FD and SD
- operation since May 2008



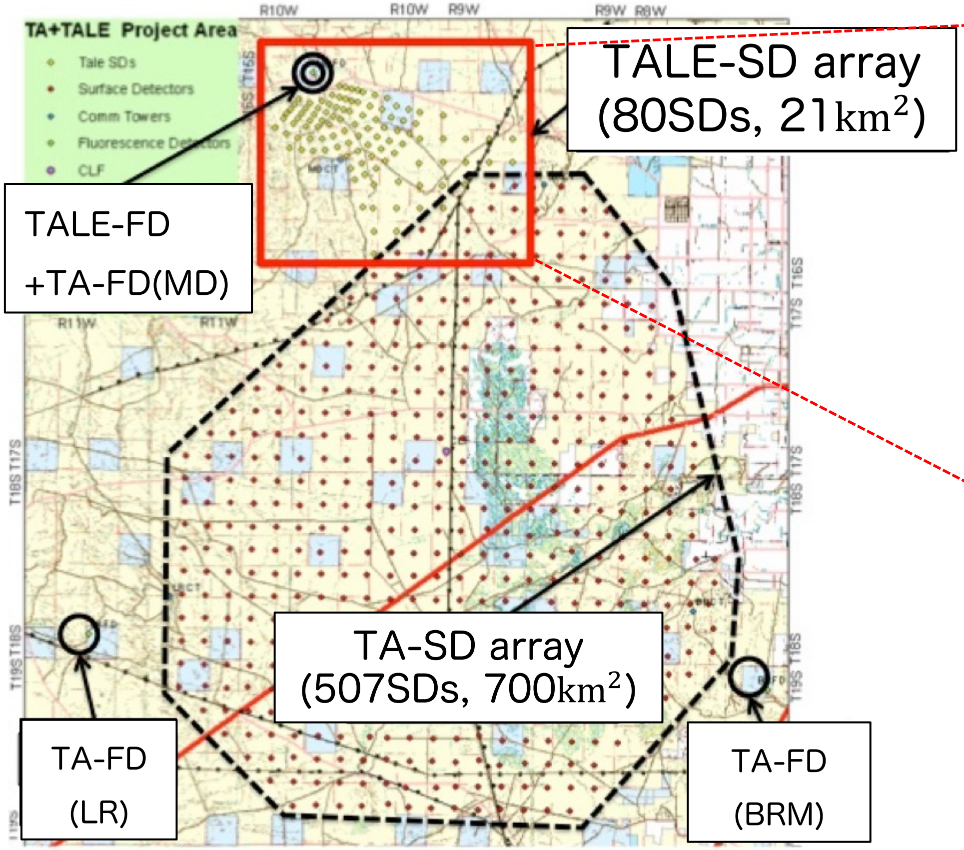
↑ TA general map

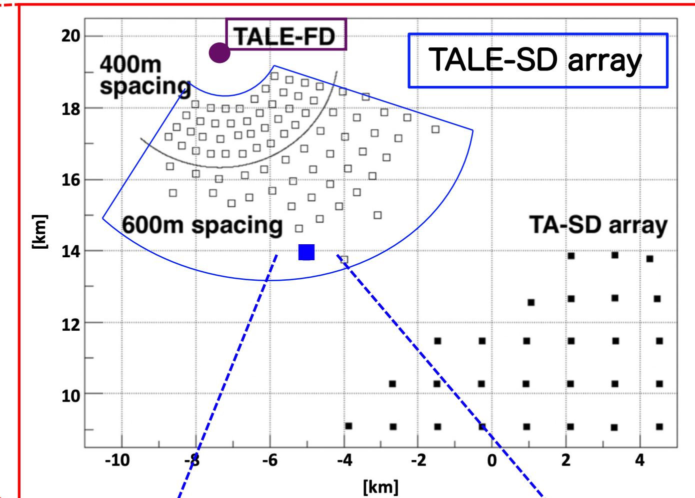
SD : Surface Detector
- 1200 m spacing



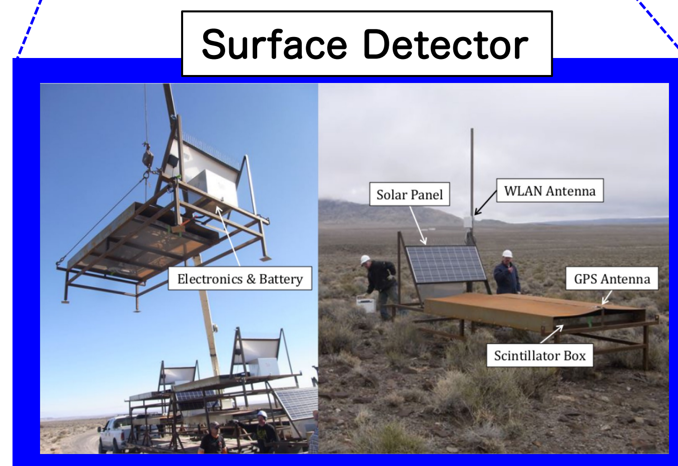
FD : Fluorescence Detector
- 38 units in 3 stations

TALE→TA Low energy Extension





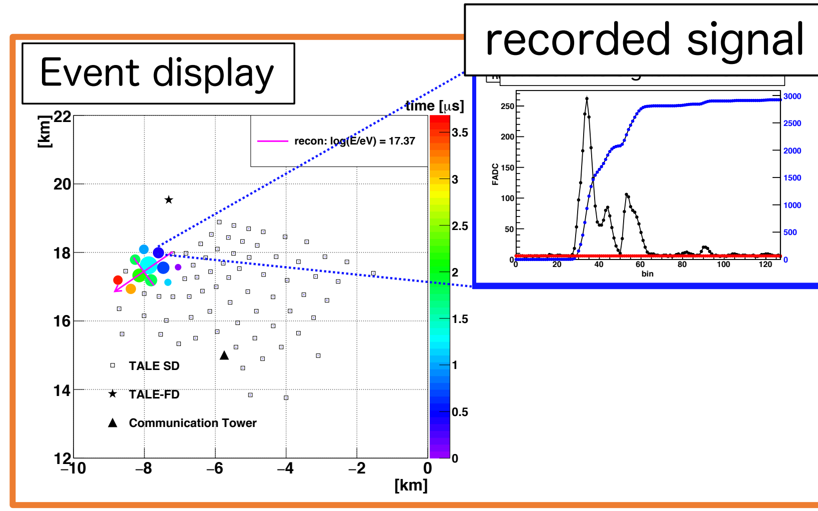
80SDs →
40SDs (400m spacing)
+ 40SDs (600m spacing)



TALE-SD array
- are located in the northwest of TA-SD
- observe cosmic rays above $10^{16.5}$ eV
- have **high statistics** and uniform sensitivity

Motivation of Mass composition analysis with TALE-SD

- <u>Advantage of SD array</u>
  - over 90% duty cycle, **high statistics** (FD → 10% duty cycle)
  - Application to cosmic ray anisotropy per composition



- <u>Approach</u>
  - machine learning

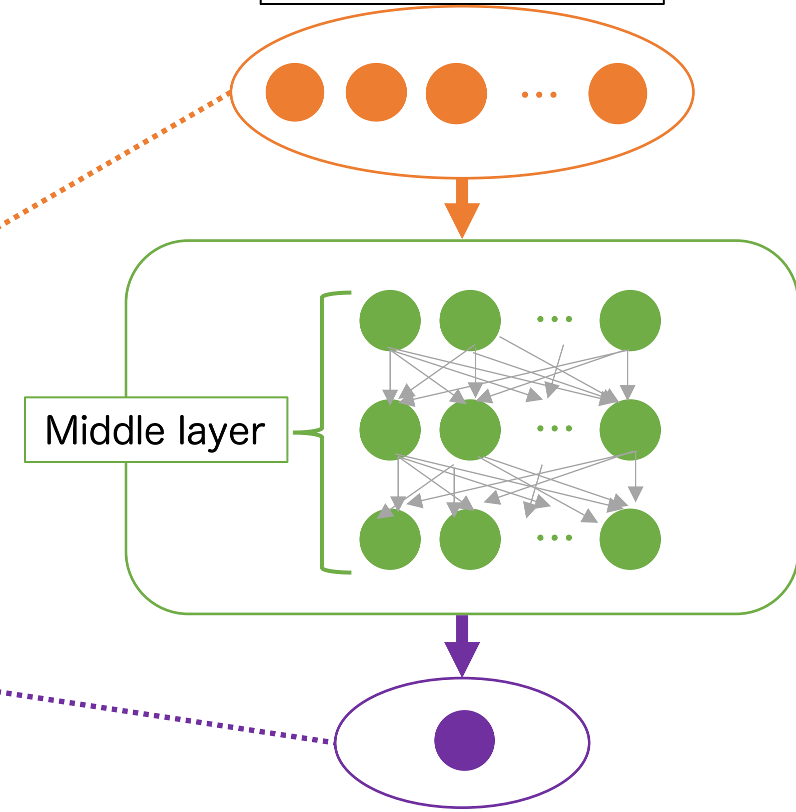
Composition-sensitive parameters
$(\alpha_1, \alpha_2, \alpha_3, \ldots)$

Machine learning approach

Machine learning model Outline

- proton/iron cosmic rays by MC simulation discrimination model (binary classification)

machine learning model Image



$(\alpha_1, \alpha_2, \ldots) \rightarrow \beta$

Input : α

→ **22 composition-sensitive parameters**

- parameters of geometry fit function
- parameter of lateral distribution
- the thickness of an air shower

Output : β

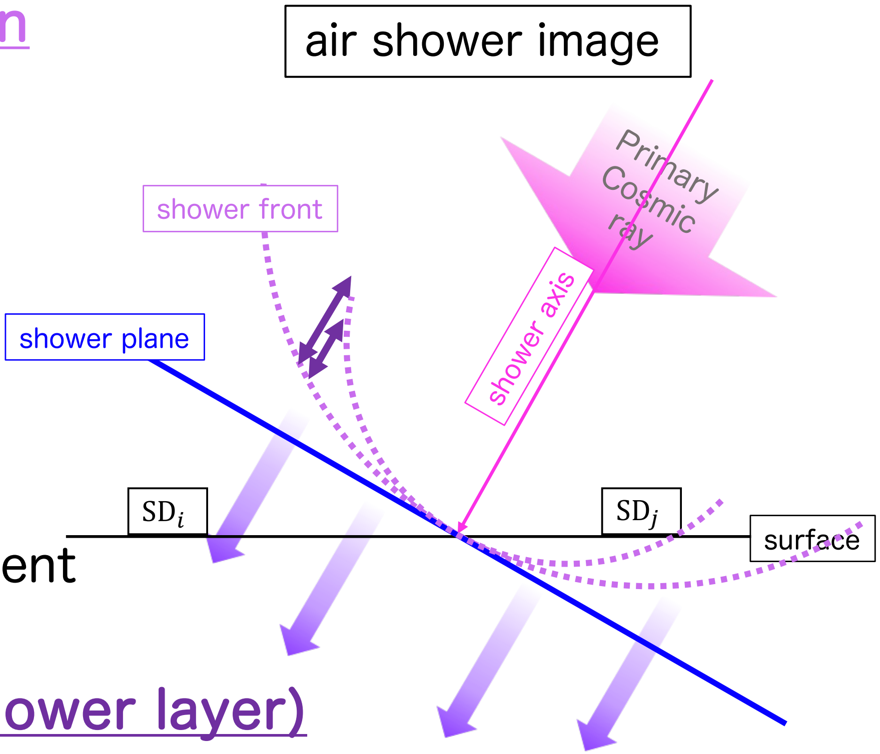
→ **Real number from 0 to 1**

- labeled proton to 0, iron to 1
- $\beta < 0.5$ → consider as proton
- $\beta \geq 0.5$ → consider as iron

22 Composition-sensitive parameters with TALE-SD

22 mass-composition sensitive parameters from the characteristics of air showers →

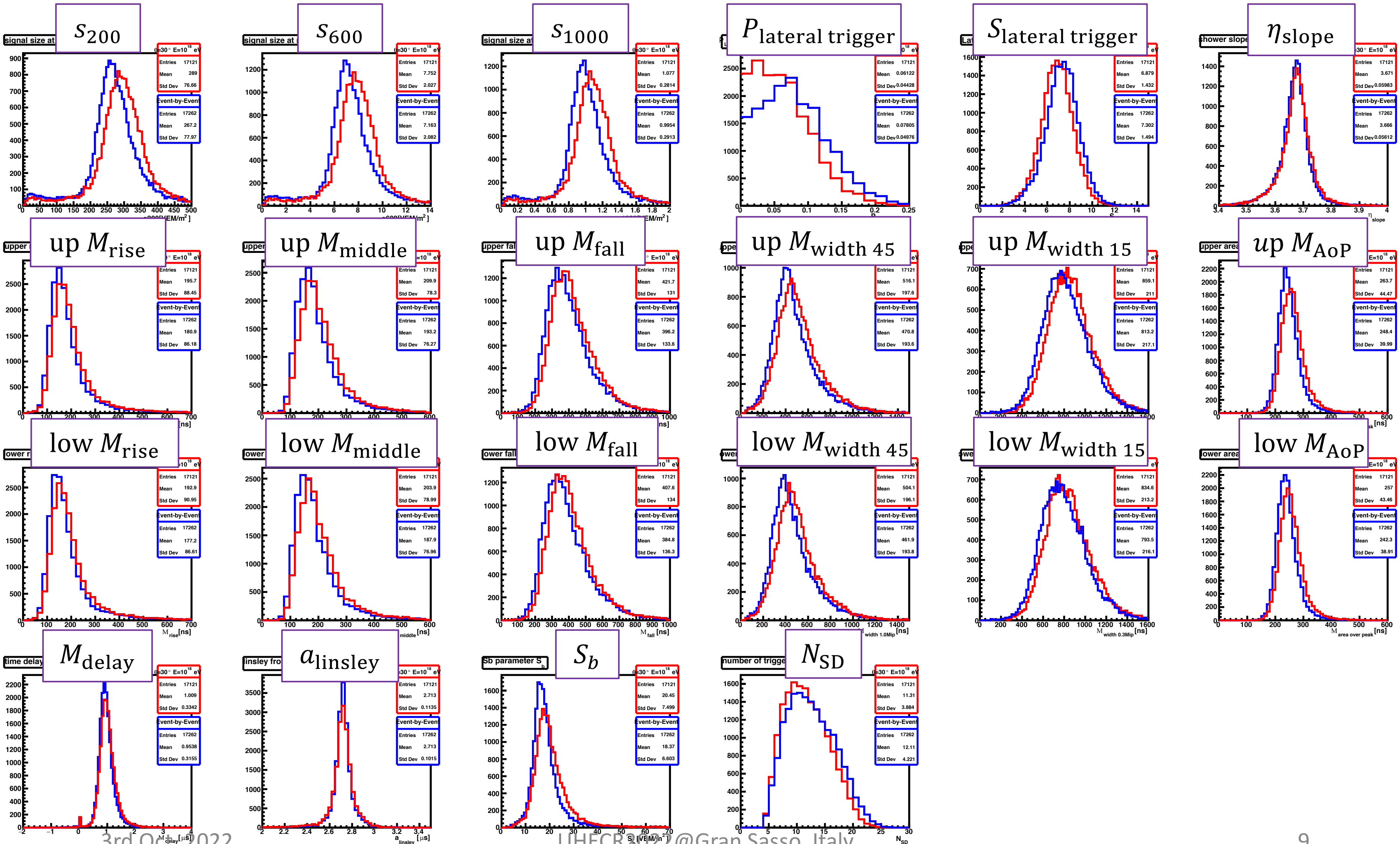
- parameters of geometry fit function
  - linsley curvature parameter
  - time delay average



- parameters of lateral distribution
  - lateral trigger probability
  - energy estimator
  - LDF function's parameter
  - number of triggered stations in 1 event
- thickness of an air shower (upper/lower layer)
  - rise・middle・fall time stations' average
  - time width stations' average
  - area over peak stations' average

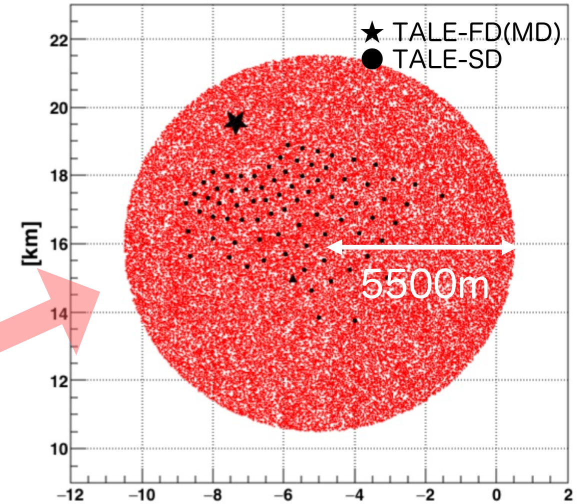
Composition-sensitive parameters with TALE-SD

- proton
- iron



MC simulation Data set and Cut

Data Set	proton	iron
MC Simulation	CORSIKA	
Interaction model	QGSJET II-04	
Energy	10^{18}eV (fixed)	
Zenith Angle	30° (fixed)	
Azimuth Angle	$0^\circ - 360^\circ$	
Core position	Uniform in the red circle	
$N_{\mathrm{TALE-SDs}}$	78 (trigger : any4)	
N_{generate}	100,000	100,000
$N_{\mathrm{reconstructed+after\ Event\ Cut}}$	17,121	17,262
N_{used}	17,121	17,121



Shower Core Position

Event Cut

- $N_{\mathrm{SD}} \geq 5$
- $\chi^2_{\mathrm{G}}/\mathrm{d.o.f.} \leq 4, \chi^2_{\mathrm{L}}/\mathrm{d.o.f.} \leq 2$
- $\left(\sigma_\theta^2 + \sin^2\theta\,\sigma_\varphi^2\right)^{0.5} \leq 2.5\ \mathrm{deg}$
- $\sigma_{S_{600}}/S_{600} \leq 0.25$
- $N_{\mathrm{useSD}} \geq 1$ (in 1 event)

N_{useSD} condition (SD selection)

- $t_{\mathrm{record\ WF}} \leq 2.56\ \mu\mathrm{s}$ (128 bin)
- N_{bin} (>15 FADC count) ≥ 2
- N_{bin} (>45 FADC count) ≥ 1
- Not saturate
- $400\ \mathrm{m} \leq r \leq 700\ \mathrm{m}$

Machine learning model Details

Hidden layer

Input : $\alpha \rightarrow$
22 cosmic ray event parameters

Dimension : 64
Activation function : "tanh"

Dimension : 32
Activation function : "tanh"

Dimension : 16
Activation function : "tanh"

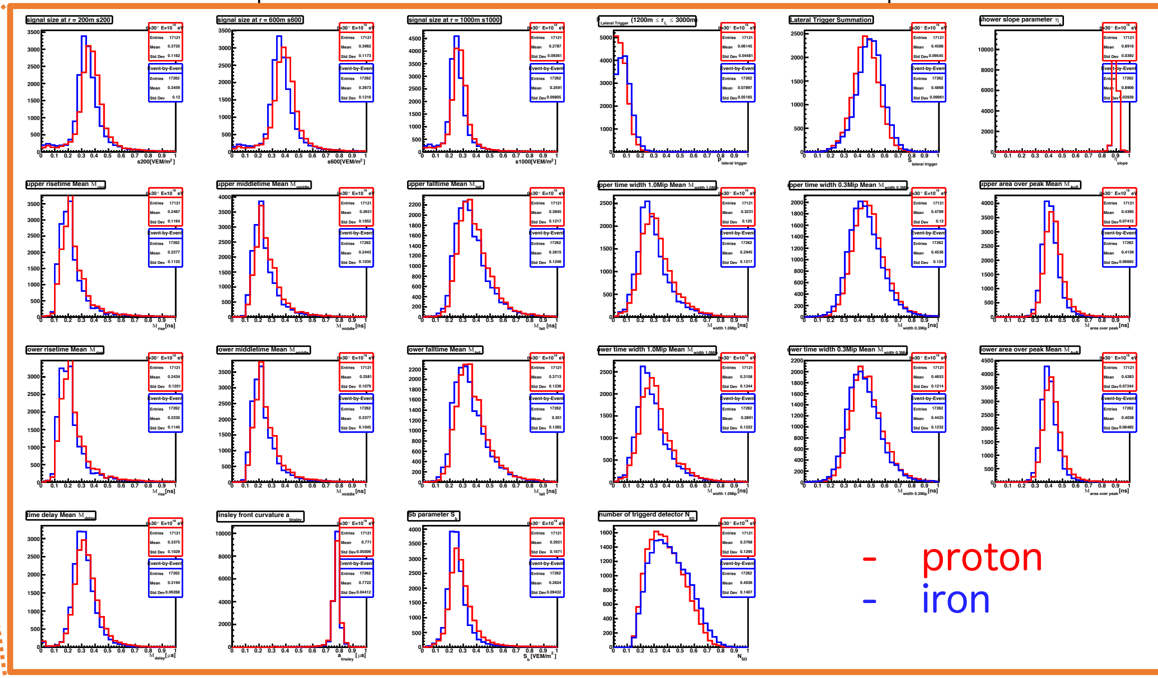
Dimension : 16
Activation function : "tanh"

Dimension : 8
Activation function : "tanh"

Output : $\beta \rightarrow 0 \leq \beta \leq 1$
Activation function : "sigmoid"

- labeled proton to 0, iron to 1
- $\beta < 0.5 \rightarrow$ consider as proton
- $\beta \geq 0.5 \rightarrow$ consider as iron

Input parameters (normalized) α



Model information	
All data	**34242** events
Training data	27392 events (**80**% of All data)
Validation data	3424 events (**10**% of All data)
Test data	3424 events (**10**% of All data)
Optimizer	Adam
Loss function	Cross-Entropy Error function

The prediction of the machine learning model

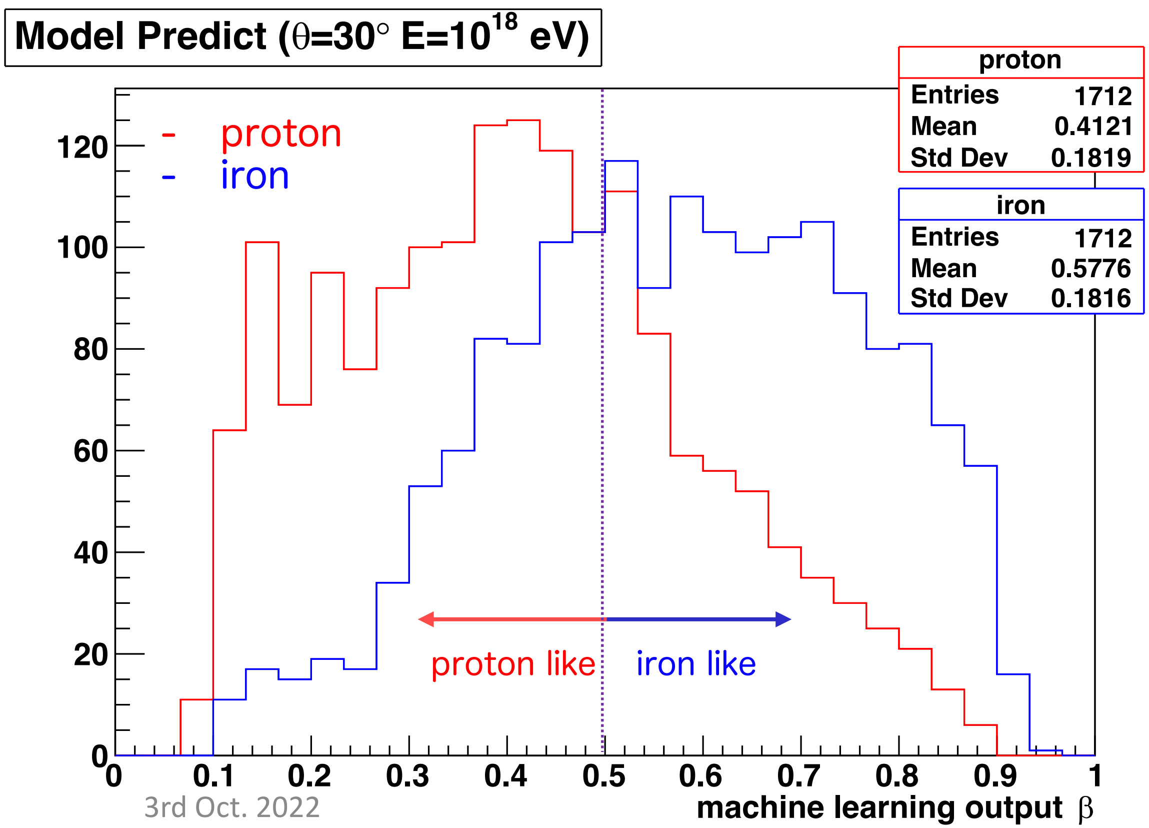
Input α :

→ 22 composition-sensitive parameters

- parameters of geometry fit function
- parameter of lateral distribution
- the thickness of an air shower

output β : $0 \leq \beta \leq 1$

- ➢ proton → 0
- ➢ iron → 1



predict \ answer	Proton	Iron
Proton	1180 (68.93%)	532 (31.07%)
Iron	593 (34.64%)	1119 (65.36%)

図 :

classification results for all test events

→ **accuracy = 67.14%**

Summary and Future

- **Summary**
  - Search composition-sensitive parameters
  - Input 22 Event-by-Event parameters
    to machine learning
  - Current Accuracy : 67.14% $(\boldsymbol{\theta = 30^{\circ}, E = 10^{18}}\ \mathbf{eV})$

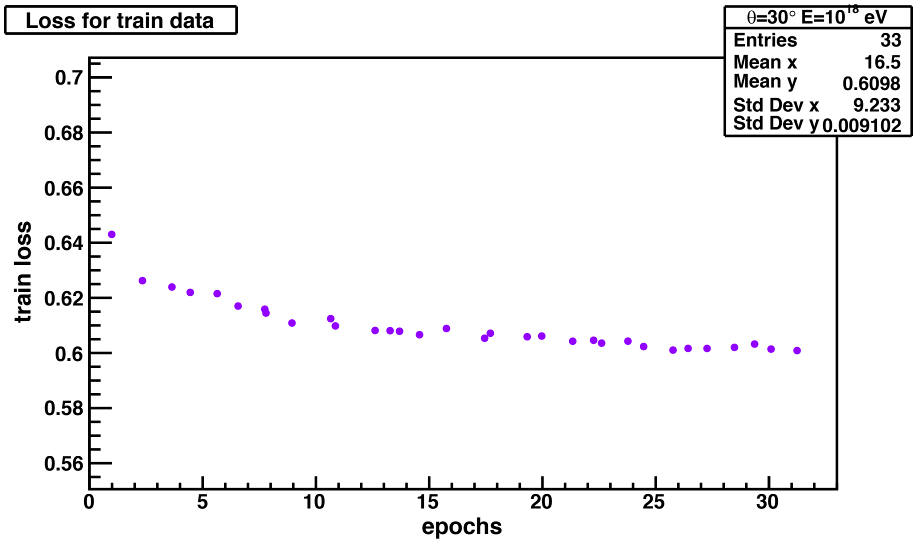
- **Future**
  - To improve the accuracy of machine learning model,

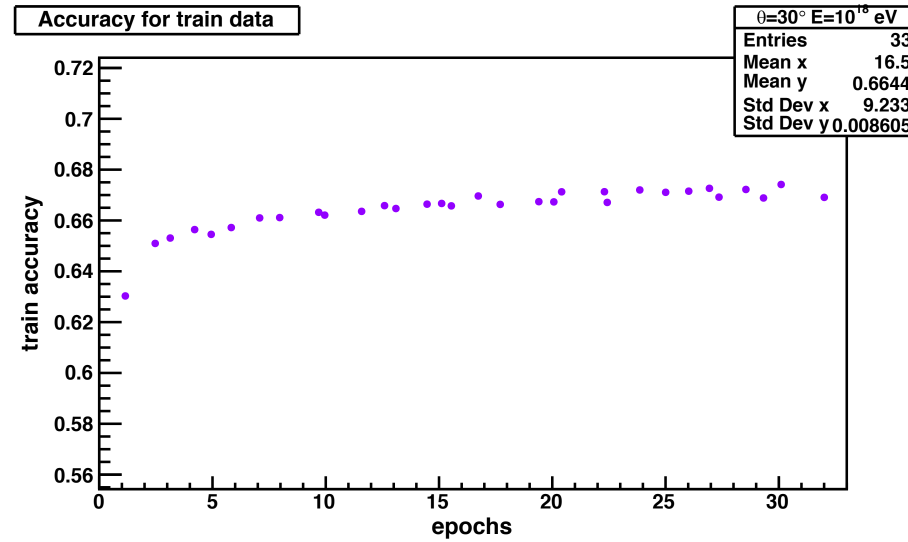
  - ➢Addiction of Detectors’ parameters as input
  - ➢Search new composition-sensitive parameters
  - ➢Using Graph Neural Network

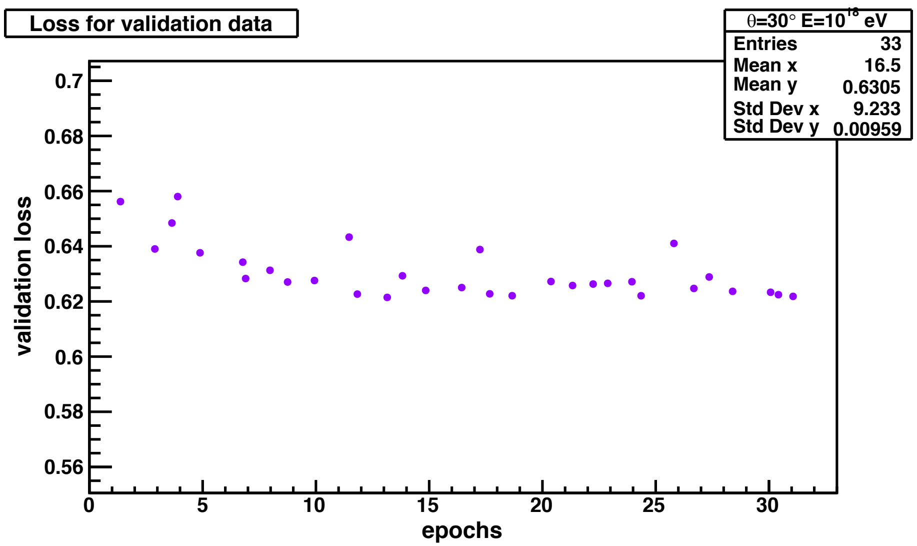
Backup

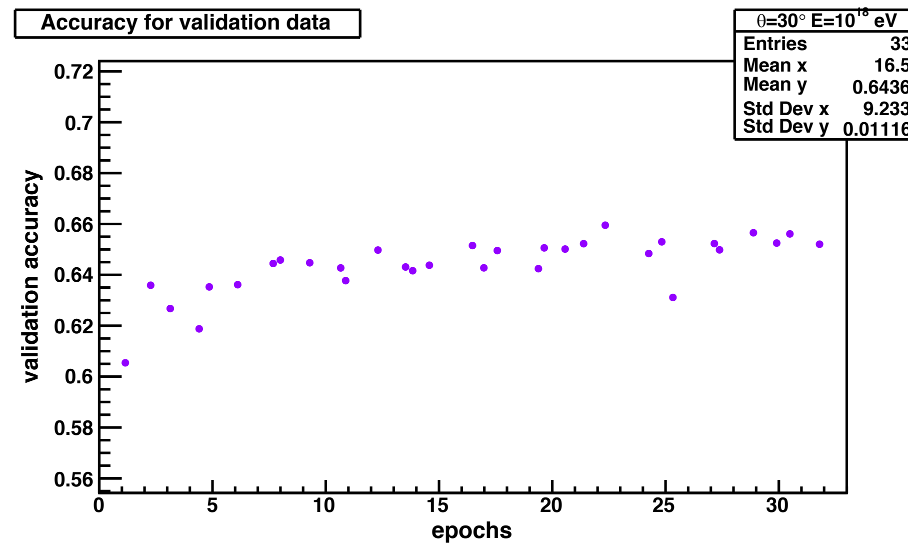
- Accuracy and loss transition
- Zenith angle and energy distribution
- Parameter definition
  - Parameter of geometry fit function
  - Parameter of lateral distribution
  - Thickness of an air shower (upper/lower layer)

Accuracy and loss transition



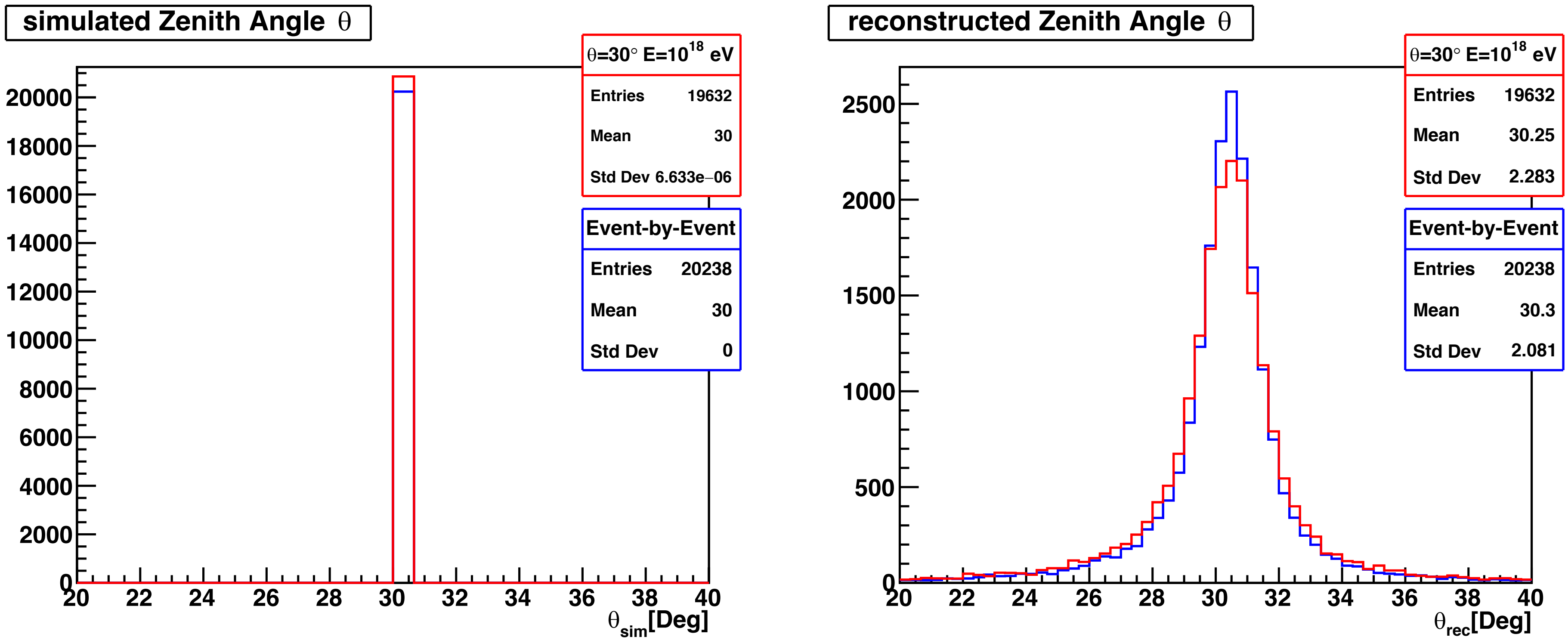




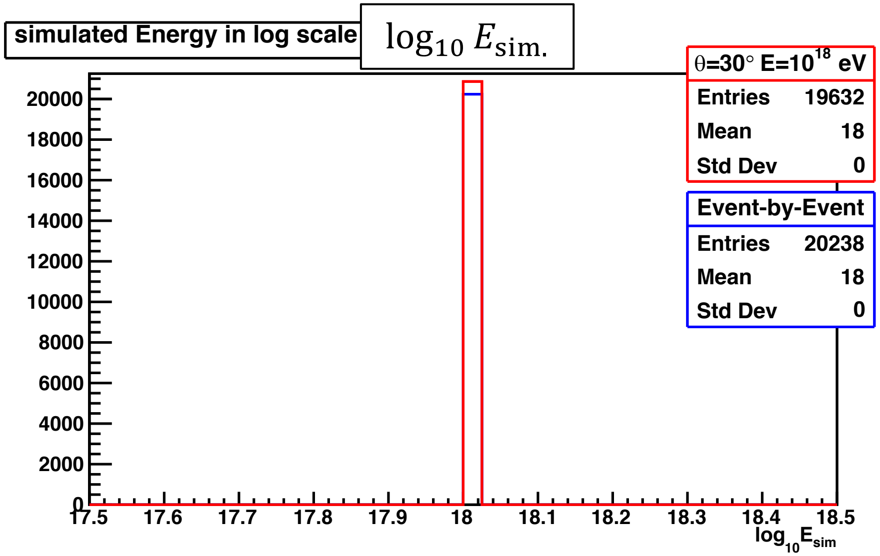


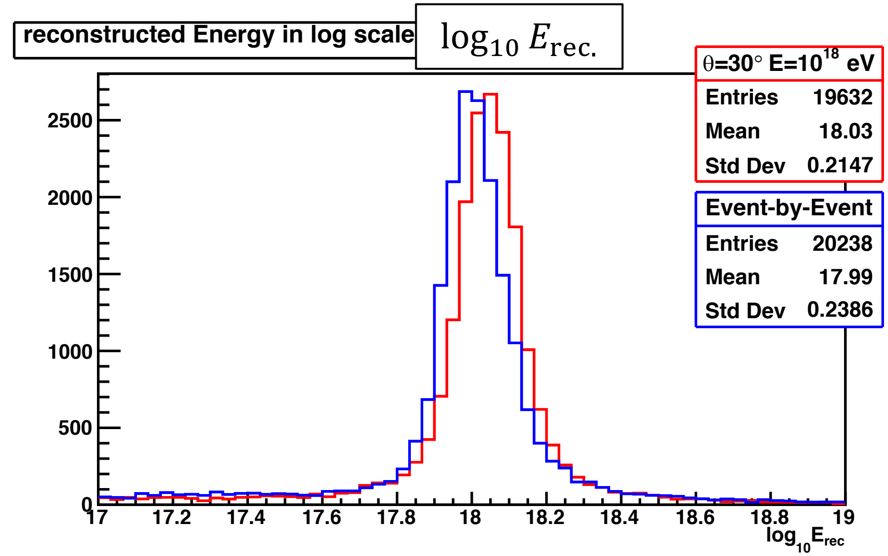
Zenith angle

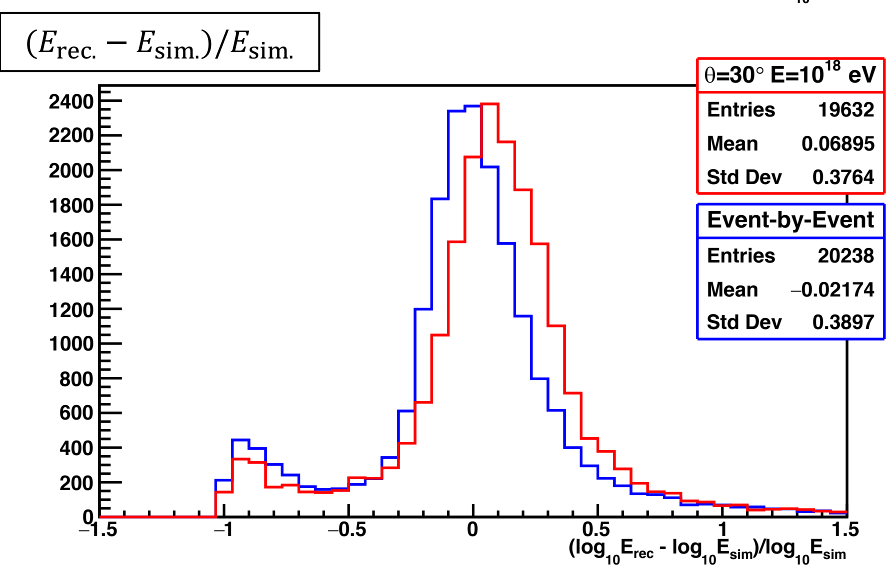
- proton
- iron



Primary Energy





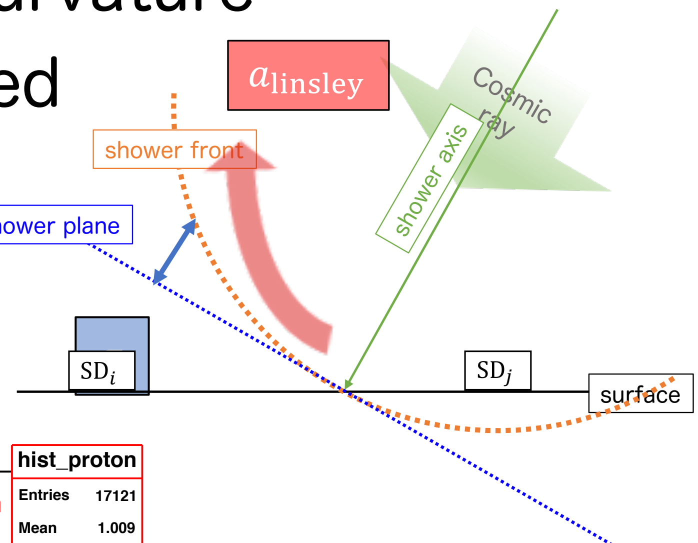


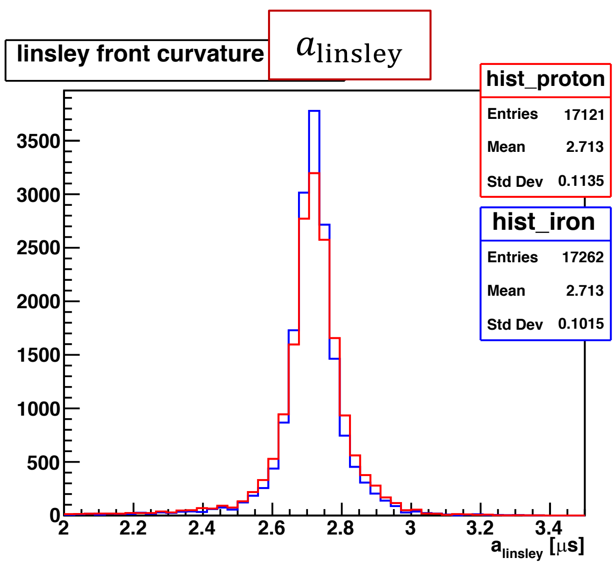
- proton
- iron

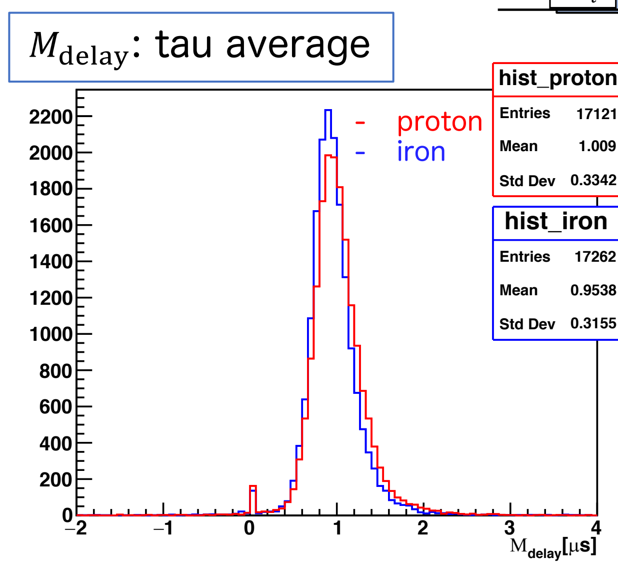
Parameters of geometry fit function

$$\tau = a_{\mathrm{linsley}} \left(1 - \frac{l}{12000\ \mathrm{m}}\right)^{1.05} \left(1 + \frac{r}{30\ \mathrm{m}}\right)^{1.35} \rho^{-0.3}$$

- the parameter of air shower front curvature
- As a_{linsley} is free parameter, τ is fitted using χ^2 test, r, and ρ







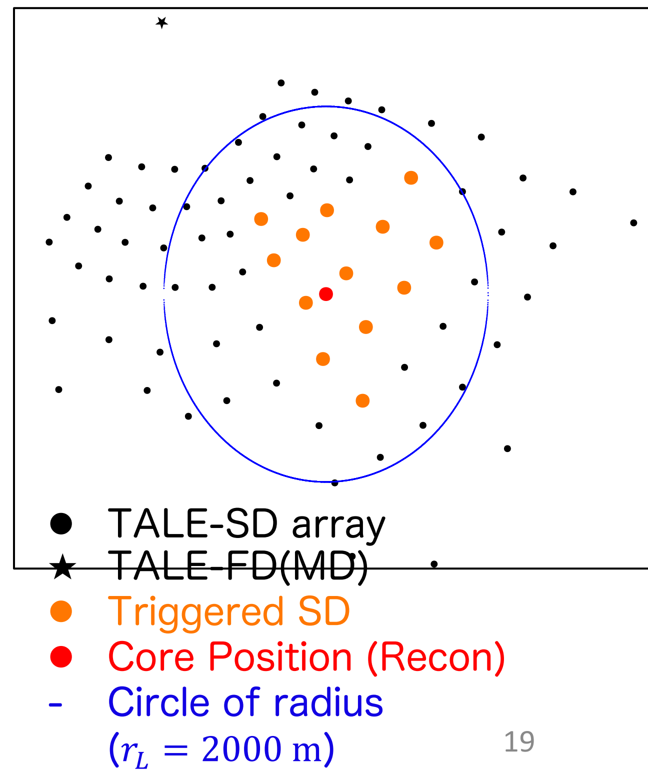
Lateral Trigger Probability

$\rightarrow P_{\text{Lateral Trigger}}$

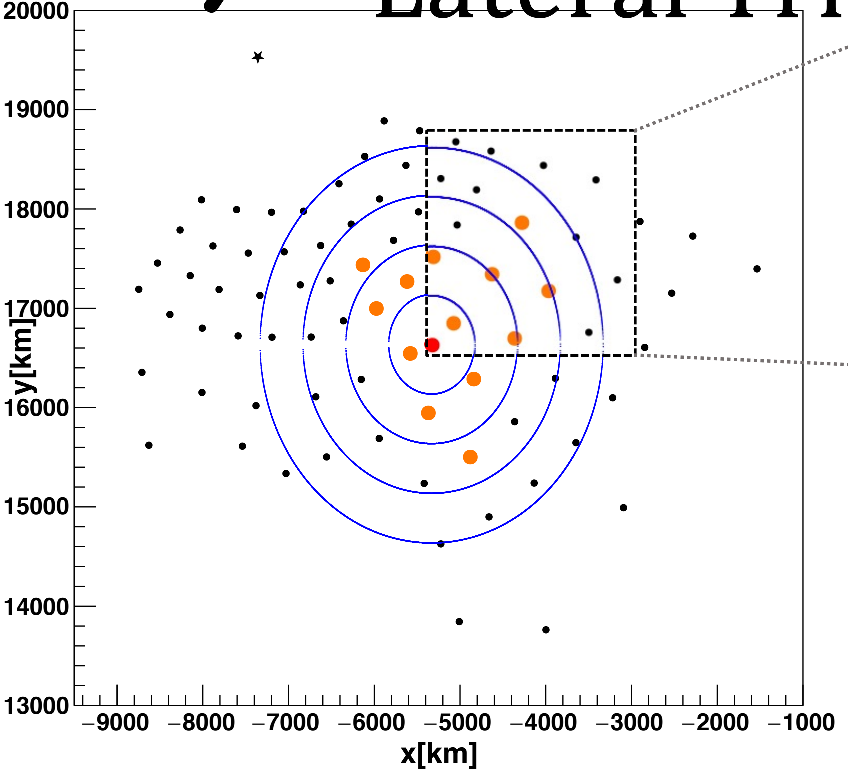
$$P_{\text{Lateral Trigger}}(r_L) = \frac{N_{\text{Trigger}}}{N_{\text{Trigger}} + N_{\text{non-Trigger}}}$$

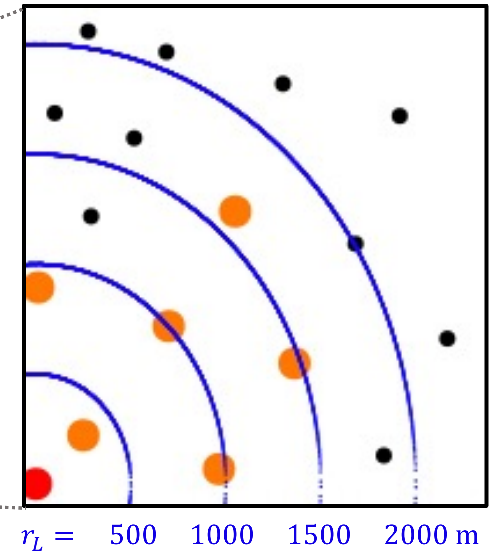
parameter	
r_L	Lateral distance between shower core and detector position
N_{trigger}	number of triggered stations contained within radius $= r_L$
$N_{\text{non-trigger}}$	number of untriggered stations contained within radius $= r_L$

Ex) $P(2000\text{m}) = 13/40 = 0.325$



Ex.) $P_{\mathrm{Lateral\ Trigger}}$





Left Fig. : Event display of MC(proton)

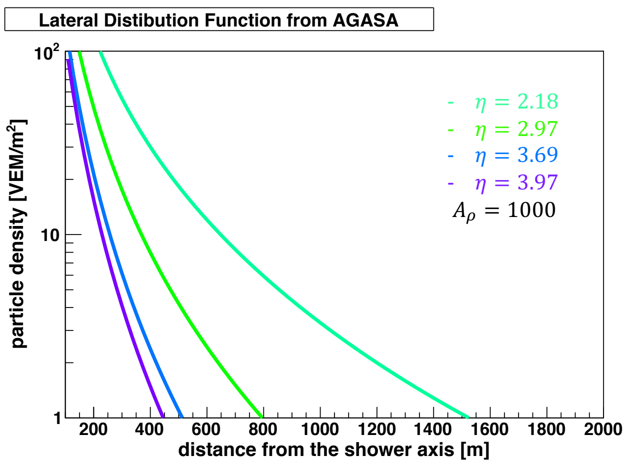
- ● TALE-SD array
- ★ TALE-FD(MD)
- ● Triggered SD
- ● Core Position (Recon)
- \- Circle of radius "r_L [m]"

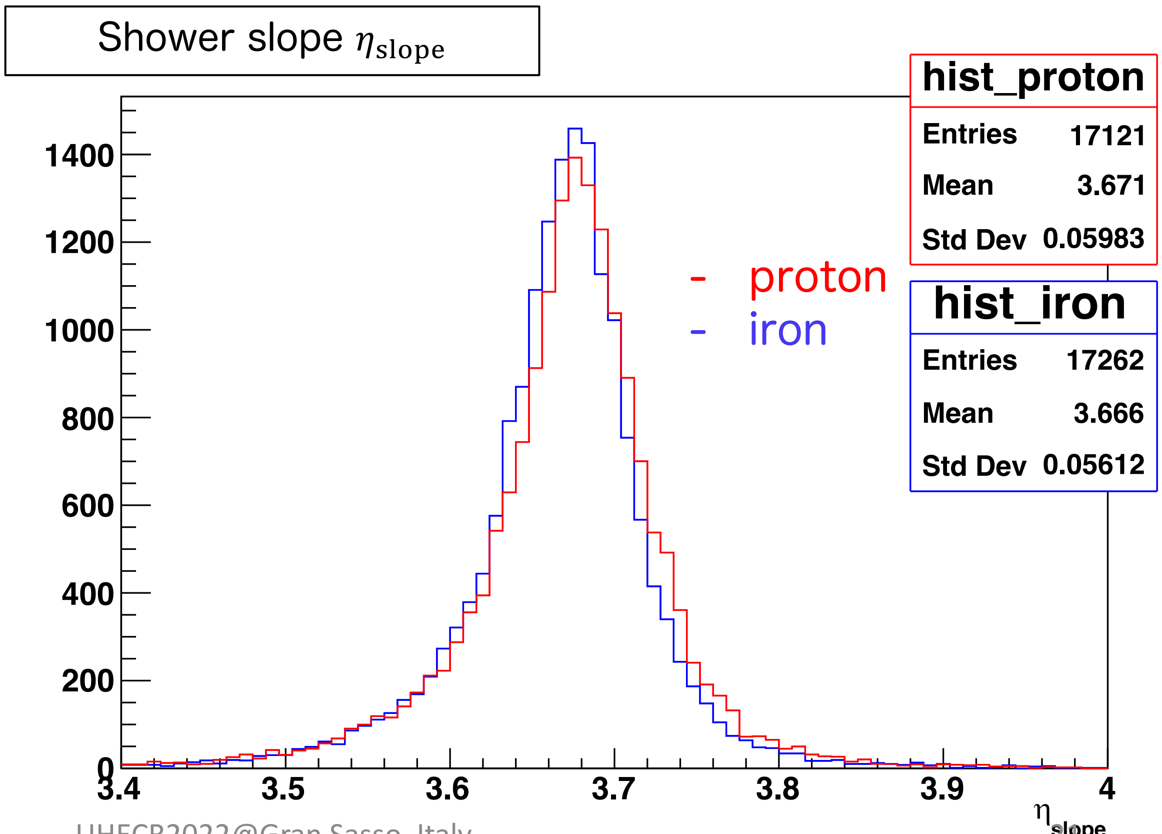
計算例	N_{trigger}	$N_{\mathrm{non-trigger}}$	N_{all}	$P_{\mathrm{Lateral\ Trigger}} = N_{\mathrm{trigger}}/N_{\mathrm{all}}$
$r_L \leq 500$ m	2	0	2	1.000
$r_L \leq 1000$ m	8	1	9	0.889
$r_L \leq 1500$ m	12	12	24	0.500
$r_L \leq 2000$ m	13	27	40	0.325

LDF function slope

$$\rho(r) = A\left(\frac{r}{91.6\ \mathrm{m}}\right)^{-1.2}\left(1+\frac{r}{91.6\ \mathrm{m}}\right)^{-\eta_s+1.2}\left(1+\left[\frac{r}{1000\ \mathrm{m}}\right]^2\right)^{-0.6}$$

- the parameter of air shower longitudinal development
- As η_s is a free parameter, $\rho(r)$ is fitted using χ^2 test and r





S_b parameter

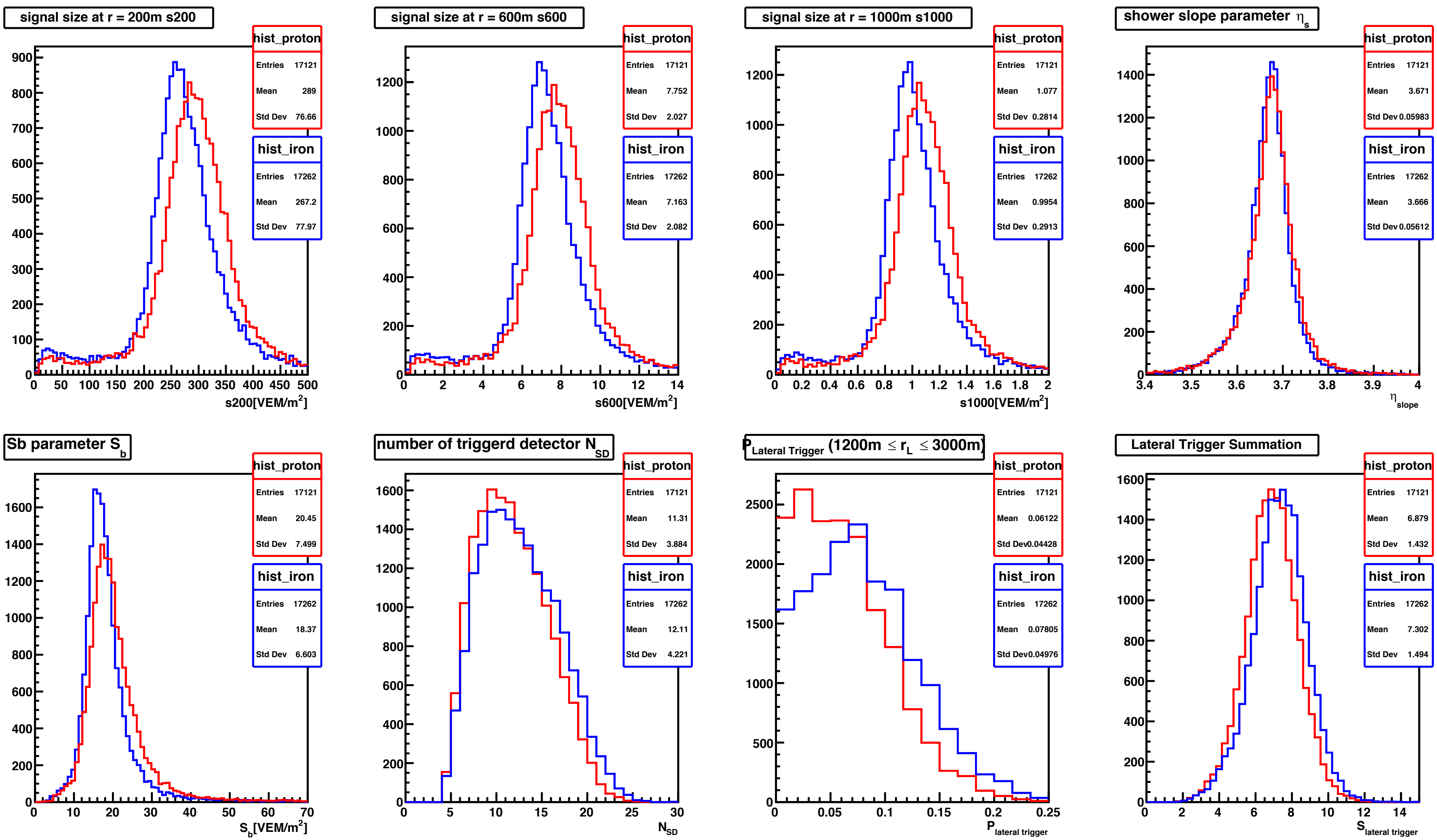
- Definition

$$S_b = \frac{1}{N_{\mathrm{SD}}} \sum_{i}^{N_{\mathrm{SD}}} \left[\rho_i \times \left(\frac{r_i}{r_0} \right)^b \right]$$

ρ_i	Particle density of i-th detector
r_i	Distance from the shower axis [m]
r_0	Reference distance [m] → 400 m
b	Separation parameter → b = 2

Parameters of lateral distribution

- proton
- iron



The thickness of an air shower

→ t_{rise} · t_{middle} · t_{fall} · t^{45}_{width} · t^{15}_{width}

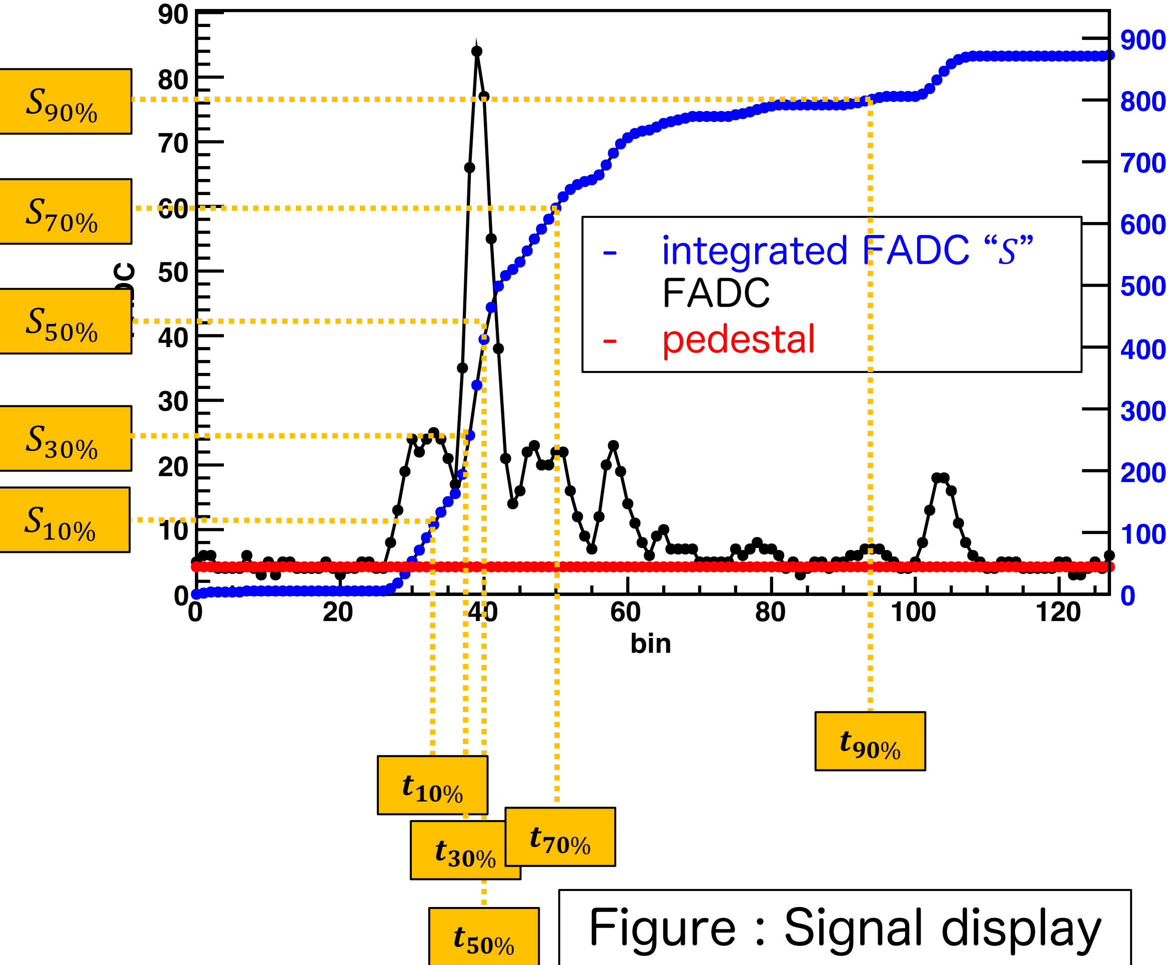


Figure : Signal display
MC iron SD6010

Represents the amount of increase in the S

- $t_{\mathrm{rise}} = t_{50\%} - t_{10\%}$
- $t_{\mathrm{middle}} = t_{70\%} - t_{30\%}$
- $t_{\mathrm{fall}} = t_{90\%} - t_{50\%}$

➢ $t_{\mathrm{xx}\%}$:
time bin when the S is xx% of the whole

The thickness of an air shower

$\rightarrow t_{\mathrm{rise}} \cdot t_{\mathrm{middle}} \cdot t_{\mathrm{fall}} \cdot t^{45}_{\mathrm{width}} \cdot t^{15}_{\mathrm{width}}$

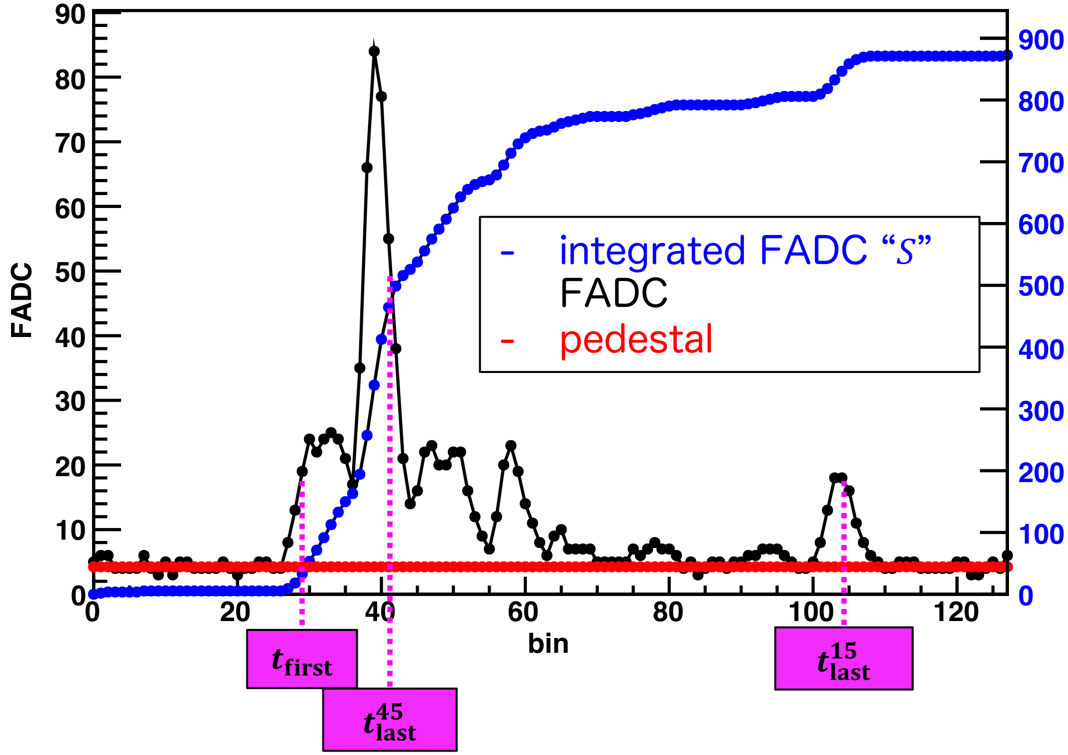


Figure : Signal display
MC iron SD6010

Represents the signal time difference

- $t^{45}_{\mathrm{width}} = t^{45}_{\mathrm{last}} - t_{\mathrm{first}}$
- $t^{15}_{\mathrm{width}} = t^{15}_{\mathrm{last}} - t_{\mathrm{first}}$

- t_{first} :
  time bin when the 0.3mip (=15fadc) signal is first input
- $t^{\mathrm{xx}}_{\mathrm{last}}$:
  time bin when the xx fadc signal is last input

Detectors' average

- Definition of Mean for thickness parameters
  → "$\mathrm{M}_{\mathrm{thickness}}$"

$$\mathbf{M}_{\mathbf{thickness}} = \frac{1}{N_{\mathbf{good\ SD}}} \sum_{i}^{N_{\mathbf{good\ SD}}} t_{\mathbf{thickness}}$$

$t_{\mathrm{thickness}}$ →	$t_{\mathrm{rise}}, t_{\mathrm{middle}}, t_{\mathrm{fall}}, t_{\mathrm{width\ 1.0Mip}}, t_{\mathrm{width\ 0.3Mip}}$
$N_{\mathrm{good\ SD}}$	# of SDs that meet the following Signal and r Cut

Signal Cut	r Cut	
➢ WF recording time is within 128 bin ➢ 2 or more bins higher than 15FADC ➢ 1 or more bins higher than 45FADC ➢ Not saturate	t_{rise}	$400\ \mathrm{m} \le r \le 1000\ \mathrm{m}$
	t_{middle}	$0\ \mathrm{m} \le r \le 800\ \mathrm{m}$
	t_{fall}	$0\ \mathrm{m} \le r \le 800\ \mathrm{m}$
	$t_{\mathrm{width}}^{1.0\mathrm{Mip}}$	$100\ \mathrm{m} \le r \le 700\ \mathrm{m}$
	$t_{\mathrm{width}}^{0.3\mathrm{Mip}}$	$100\ \mathrm{m} \le r \le 800\ \mathrm{m}$

Thickness of an air shower (upper/lower layer)

- proton
- iron