Application of artificial neural networks for problems of gamma astronomy

A.Kryukov (SINP MSU), E.Gres (IGU)

Outline

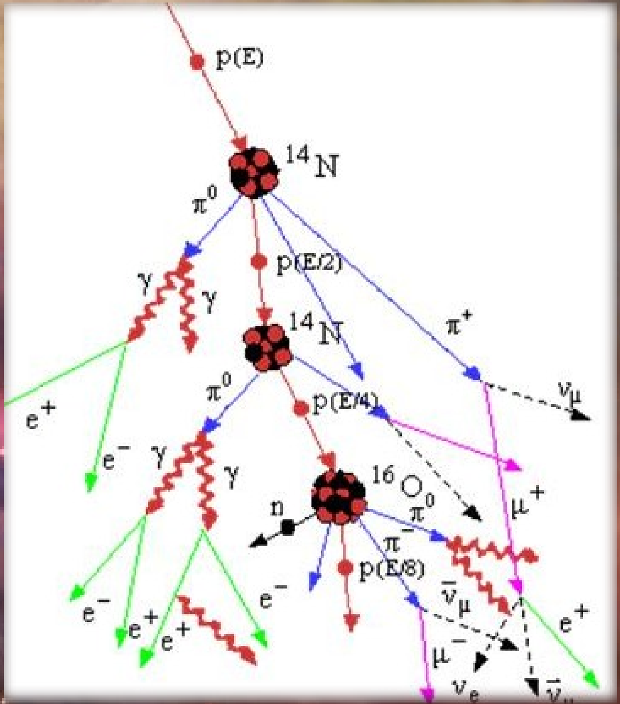
- Introduction to ANN
- What is an Artificial neuron network
  - Convolutional NN (CNN)
- Problems of data processing in gamma astronomy
- Energy identification by IACT in TAIGA
- Stereo mode in TAIGA
- Conclusion

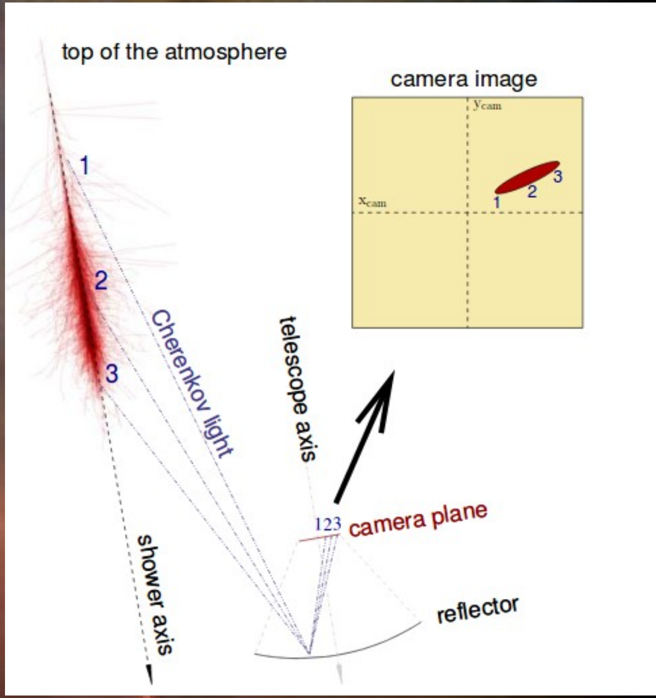
DL Particle Astrophysics Team

- Team members:
  - A. Demichev (SINP MSU),
  - Ju Dubenskaya (SINP MSU),
  - E. Gres (IGU, Irkutsk),
  - S. Polyakov (SINP MSU),
  - A. Vlaskina (Phys. Faculty of Moscow State University)
- Tasks (not a closed list):
  - Particle identification (gamma, protons, ...
  - Determination of EAS parameters (axis direction, EAS coordinators)
  - Determination of particle parameters (energy, spectra)
  - Image generation (fast generator as a MC replacement)

Introduction: What is we investigate?

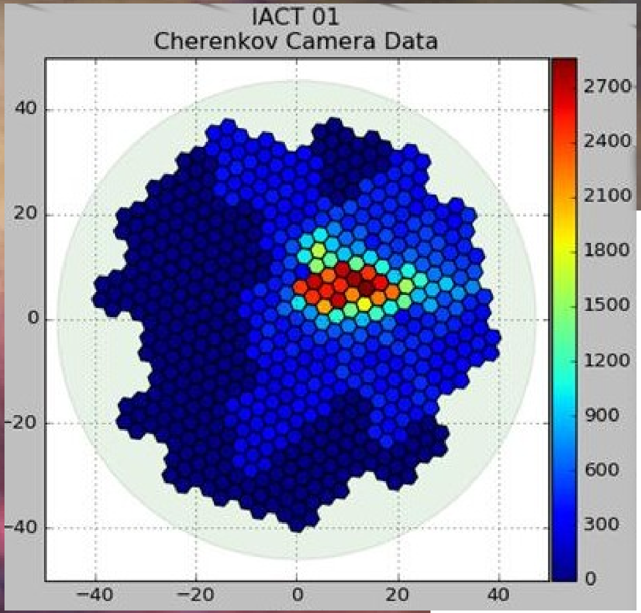
- Object: Universe
- How: Cosmic rays
- With: Extensive Air Shower (EAS)
- Methods:
  - IACT (TAIGA-IACT),
  - Array of Cherenkov detectors (HiSCORE)

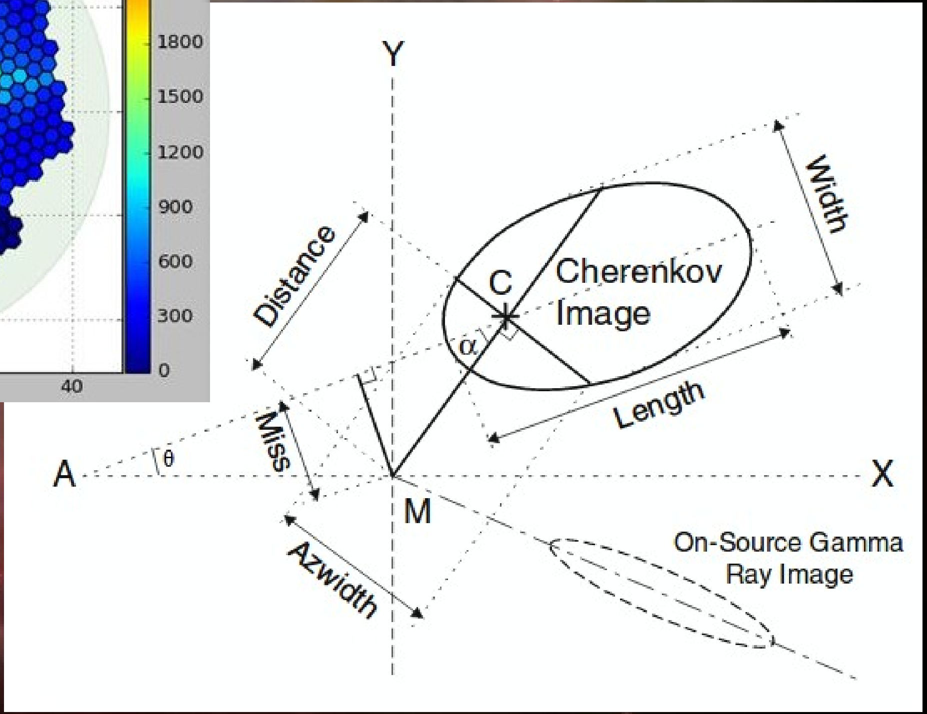




Introduction: Traditional approach

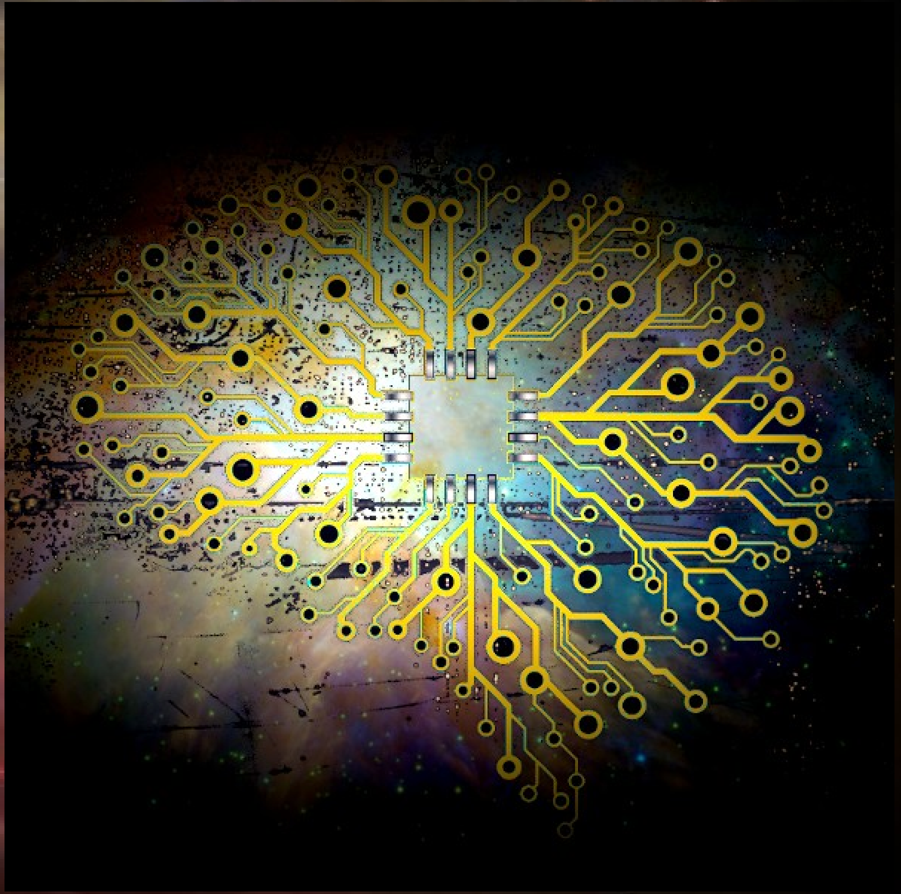
- Model: Ellipse image
- Methods: Hillas parameters
  - Length, width
  - Distance, Θ, α
  - Size ☺ (= intensity)

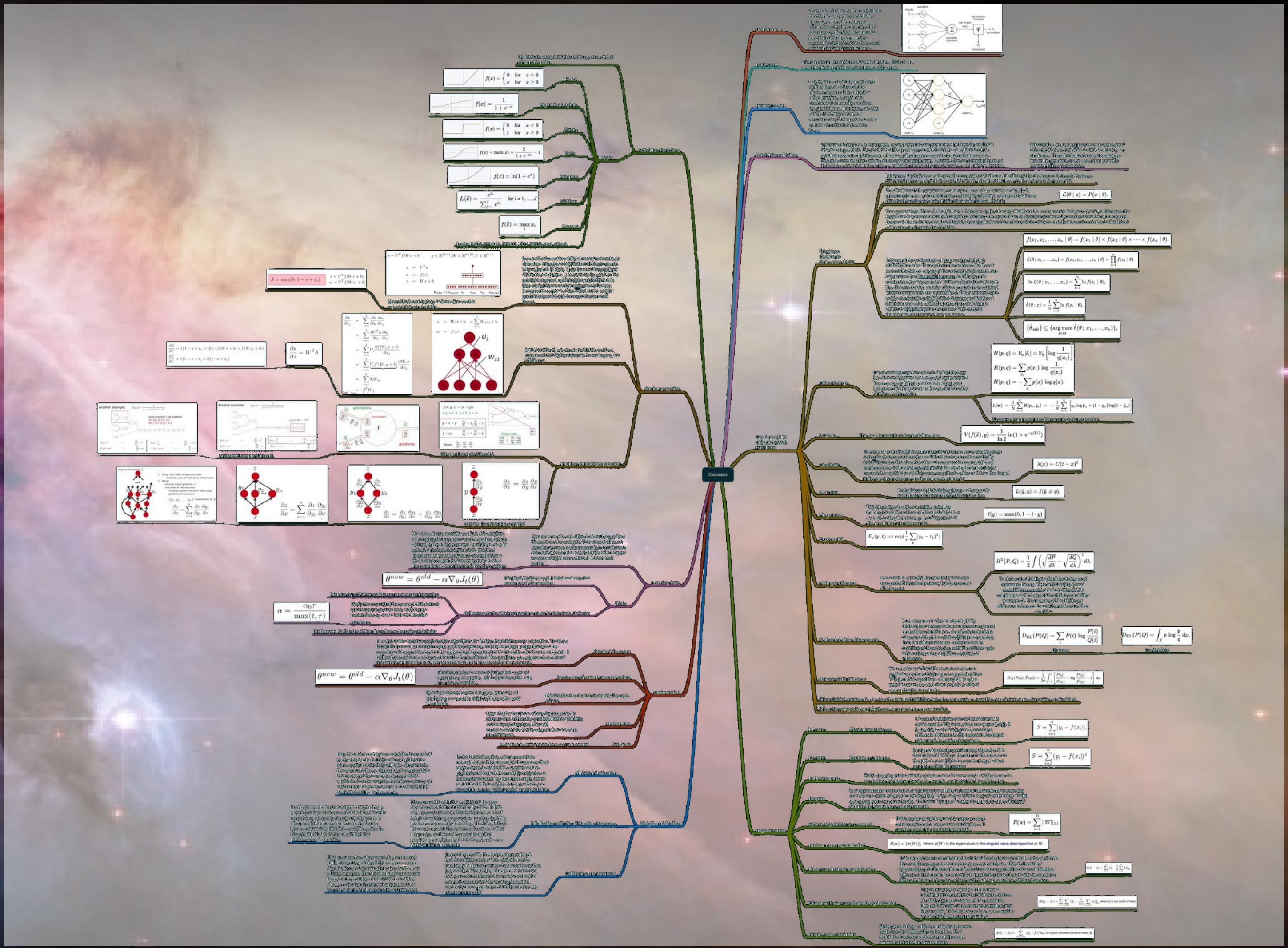




Introduction: New approach is AI / ML

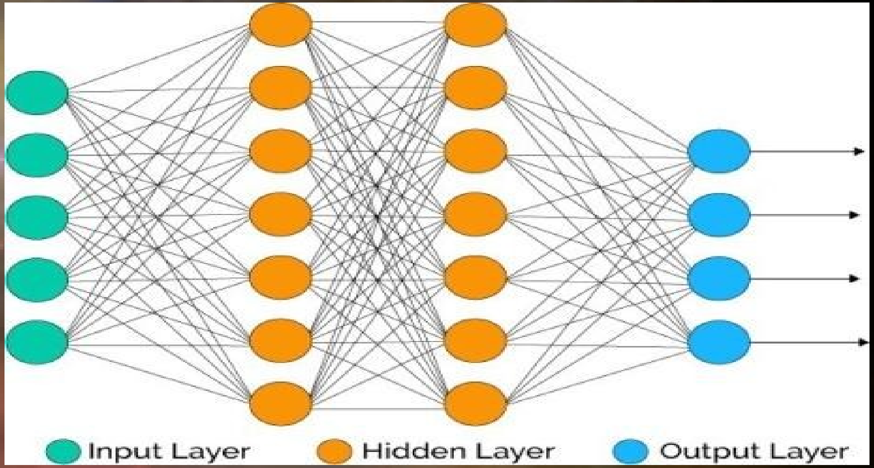
- Artificial intelligence
  - Machine learning
    - Deep learning
      - Convolutional neural network
        - …
- Success story:
  - Pattern recognition
  - Natural language processing
  - Social media analytics
  - Robotics
  - Natural sciences

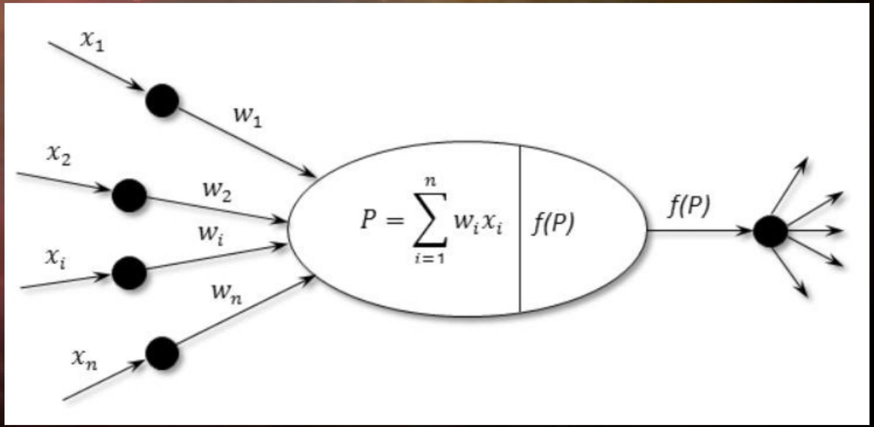



Concepts

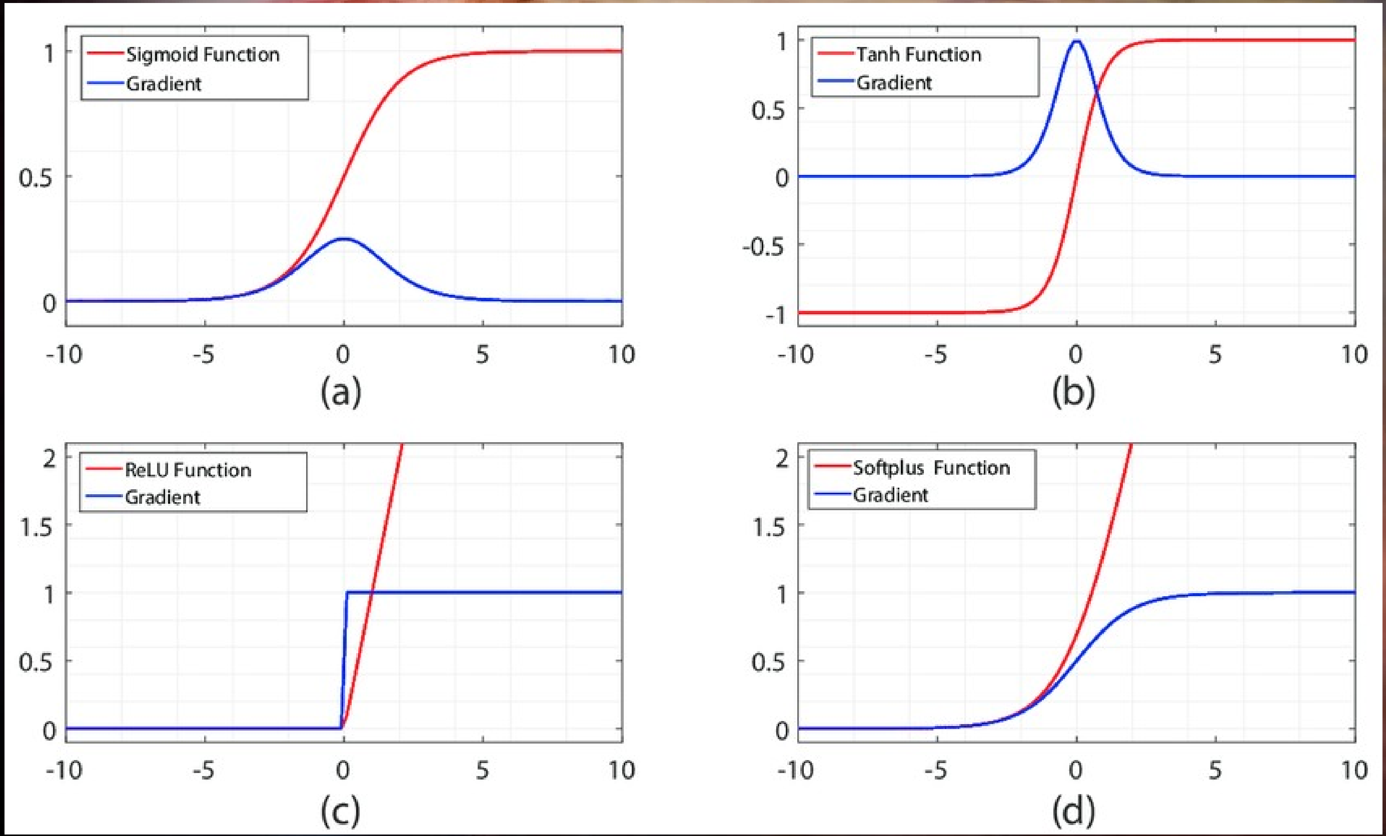
What is ANN

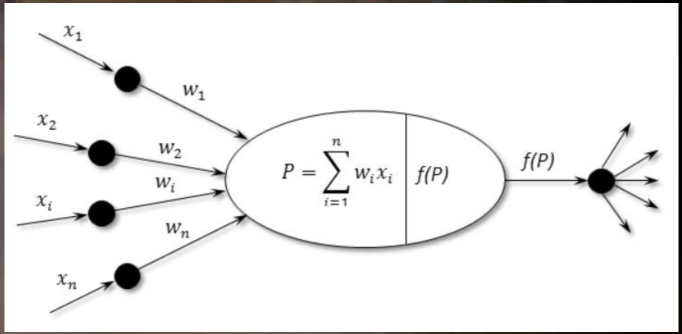
- ANN is a directed graph where vertexes are artifical neurons and edges are links between them
- Each AN has multiple inputs and one output.





ANN: Activation function





- Most popular activation function right now is ReLU

ANN: Learning

- The loss function is a measure between the desired value of the output vector and the resulting value.

$$L(W) = \underbrace{\frac{1}{N}\sum_{i=1}^{N} L_i(f(x_i, W), y_i)} + \underbrace{\lambda R(W)}$$

Потери данных: прогнозы модели должны работать на обучающей выборке

Регуляризация: модель должна работать на тестовой выборке

ANN: Lost functions

- The process of adjusting the weight vector is called training the ANN.
- Popular lost functions
  - MSE (Mean Square Error);
  - Cross enthropy;
  - Kulbaka-Leiblera mesure

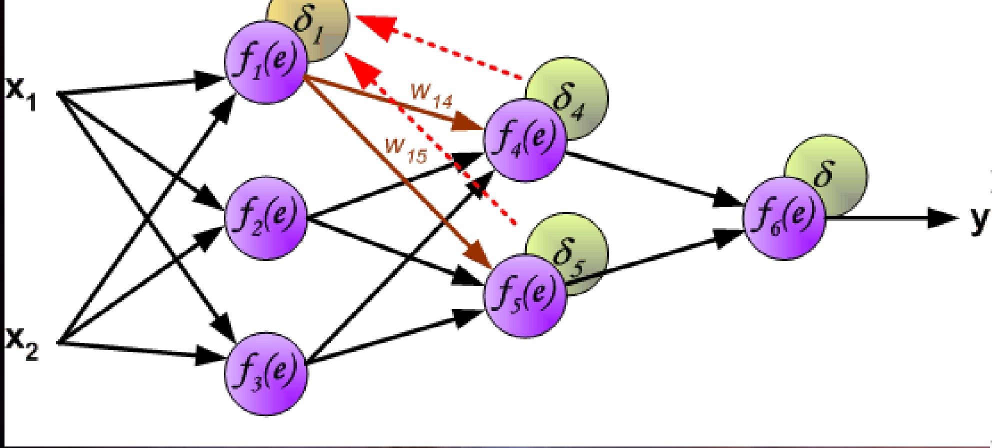
$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

$$H(p,q) = -\sum_x p(x)\,\log q(x).$$

$$KL(P||Q) = \sum_i^n P_i \ln\frac{P_i}{Q_i}$$

ANN: Error back propagation

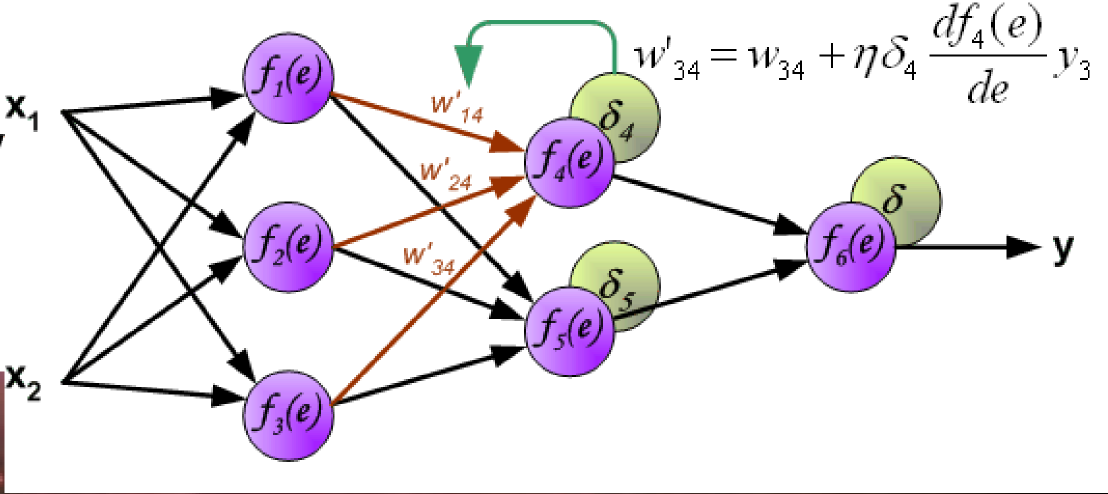
$$\delta_1 = w_{14}\delta_4 + w_{15}\delta_5$$



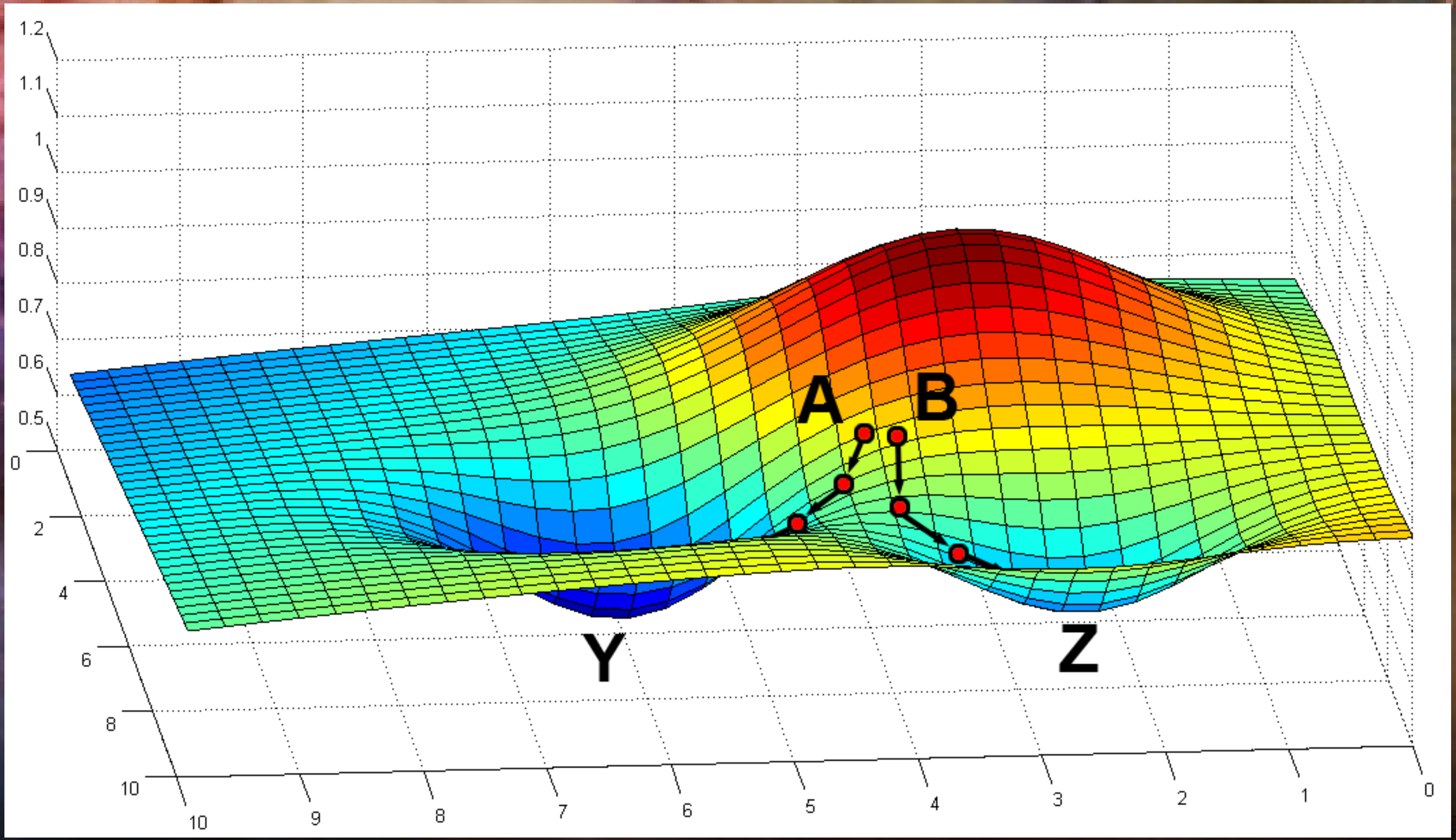
$$w'_{14} = w_{14} + \eta\delta_4 \frac{df_4(e)}{de} y_1$$

$$w'_{24} = w_{24} + \eta\delta_4 \frac{df_4(e)}{de} y_2$$

$$w'_{34} = w_{34} + \eta\delta_4 \frac{df_4(e)}{de} y_3$$

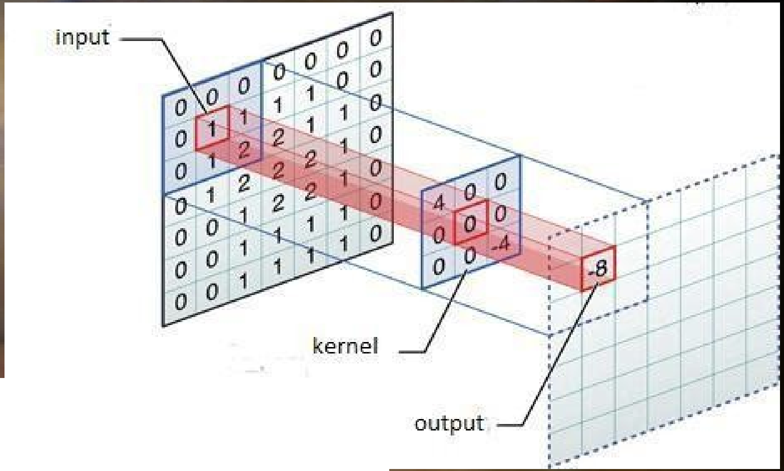


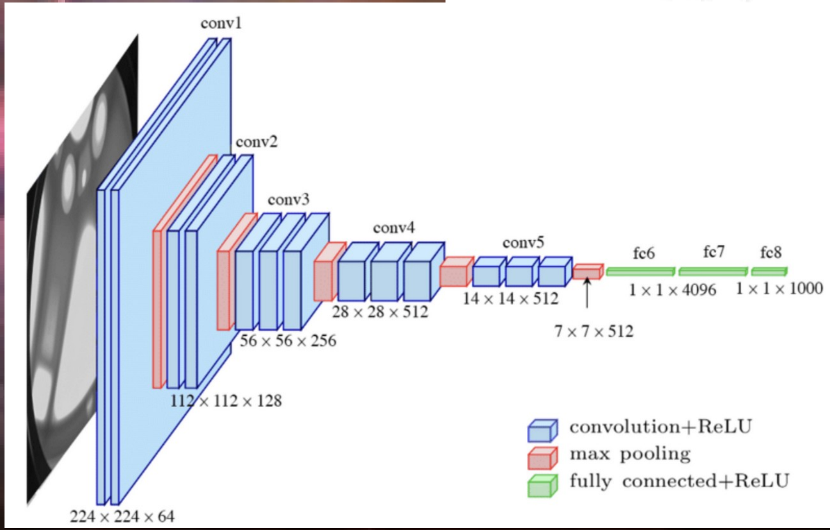
ANN: Gradient descent



Convolutional NN

- CNN is a special case of ANN which have one or more so called convolutional layers in the beginning.
- CNN is widly use for image analysis.





ANN: Conclusion

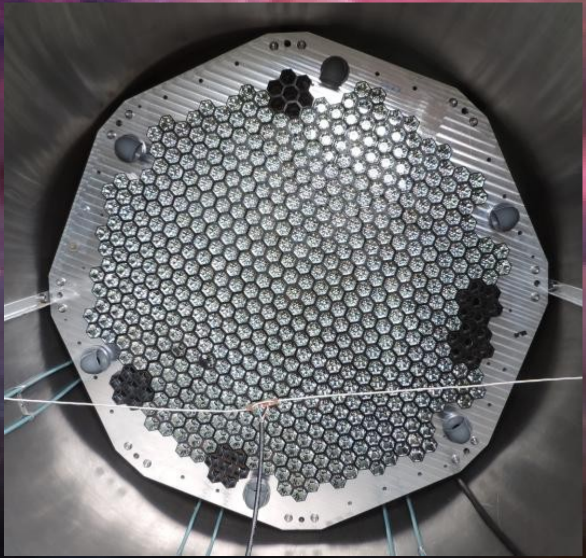
- ANN is a universal approximator, the minimum of the loss function of which is achieved by means of error back propagation (gradient descent) using training examples.
- In fact, the training set is a function defined by the table, and the weights are optimization parameters in the class of functions defined by the ANN.

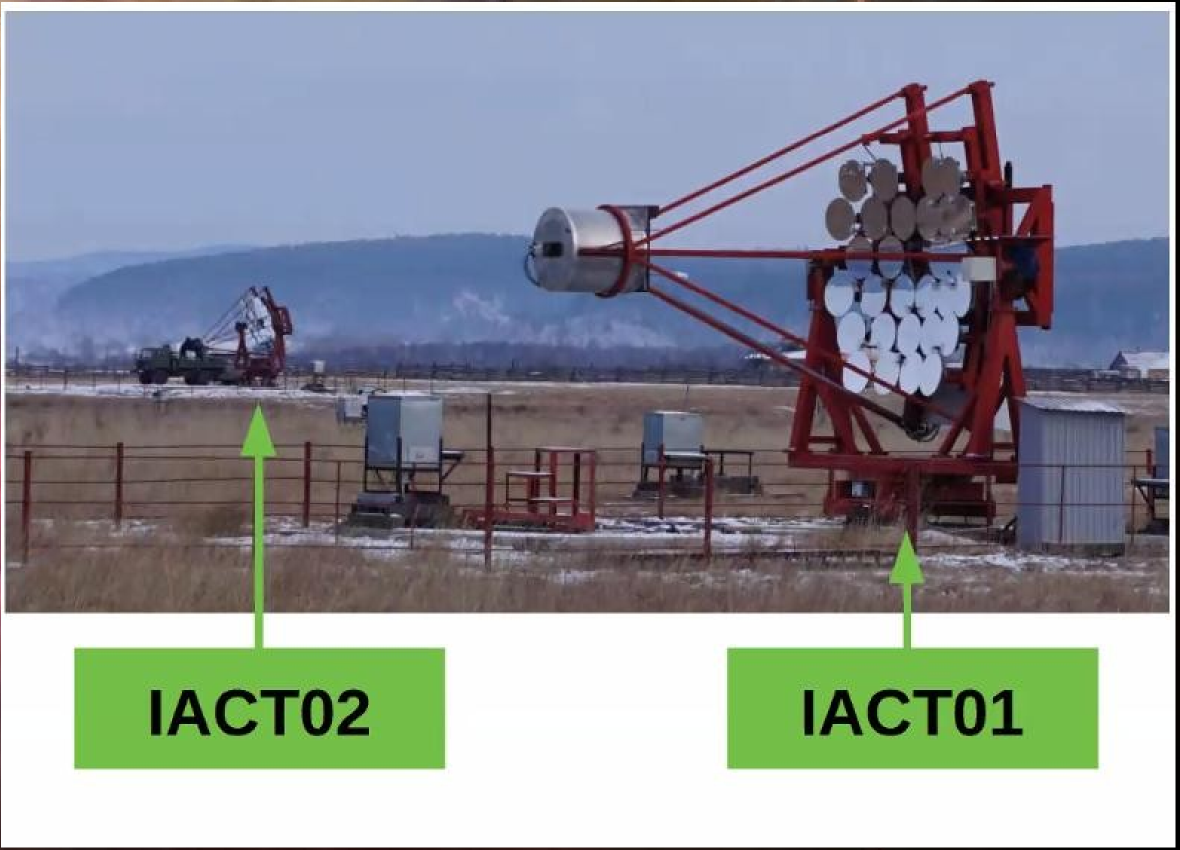
Data processing in gamma astronomy

- Main problems of data processing
  - Primary particle type identification
  - Determination of the direction of the EAS axis and its coordinates
  - Determination of the parameters of primary particles
- Here we will consider the third type of problem - the determination of the energy of gamma quanta from the IACT data in the TAIGA experiment.

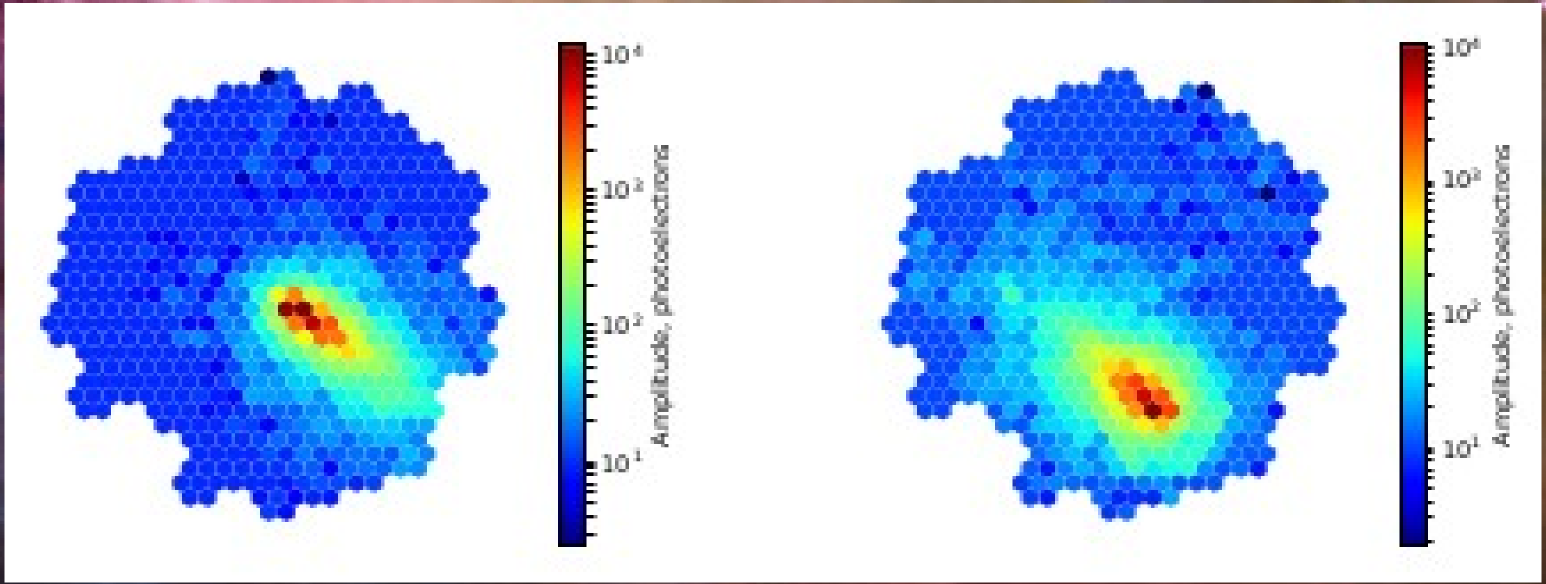
IACT in TAIGA

- Currently, the TAIGA experiment operates three IACTs.

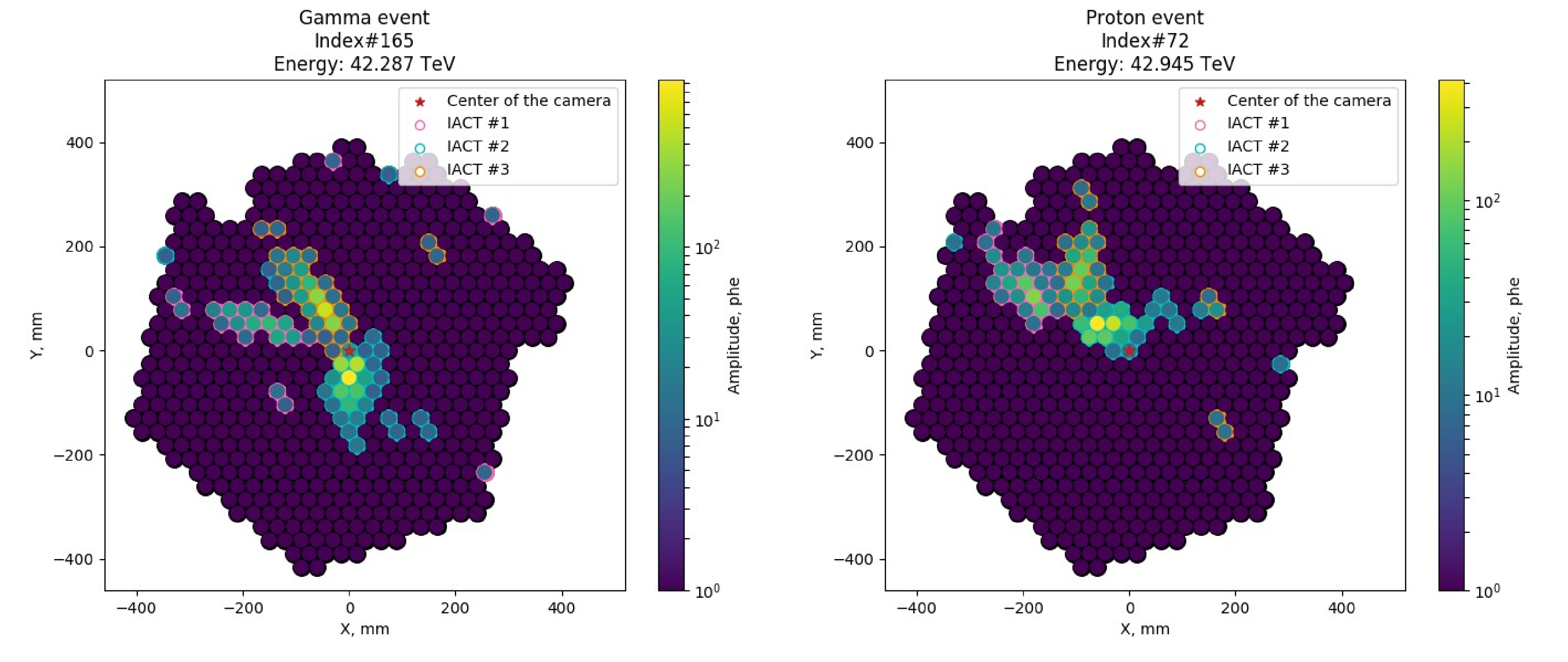




IACT: Image example (MC)

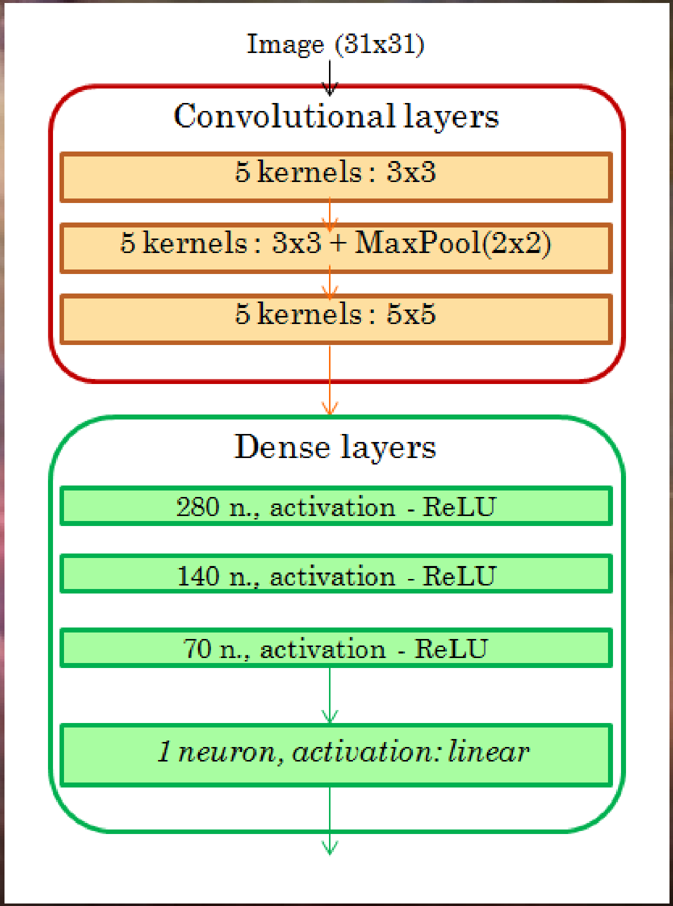


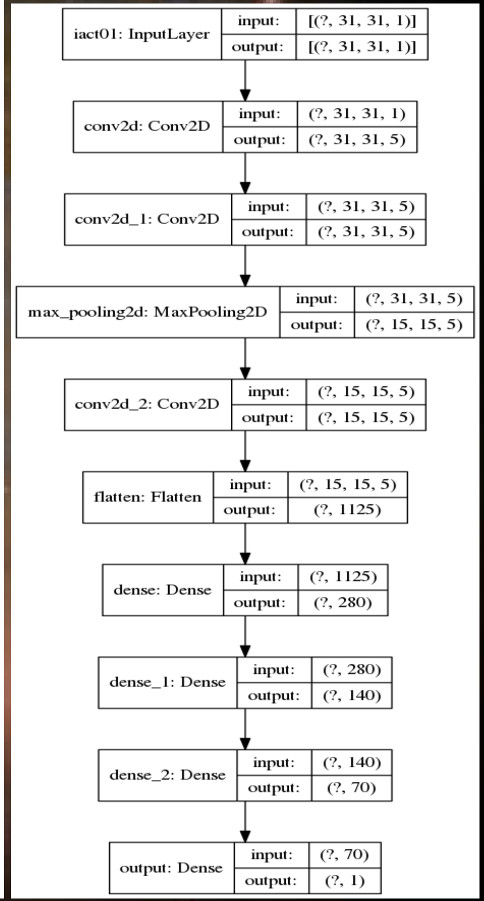
IACT: Stereo mode (MC)



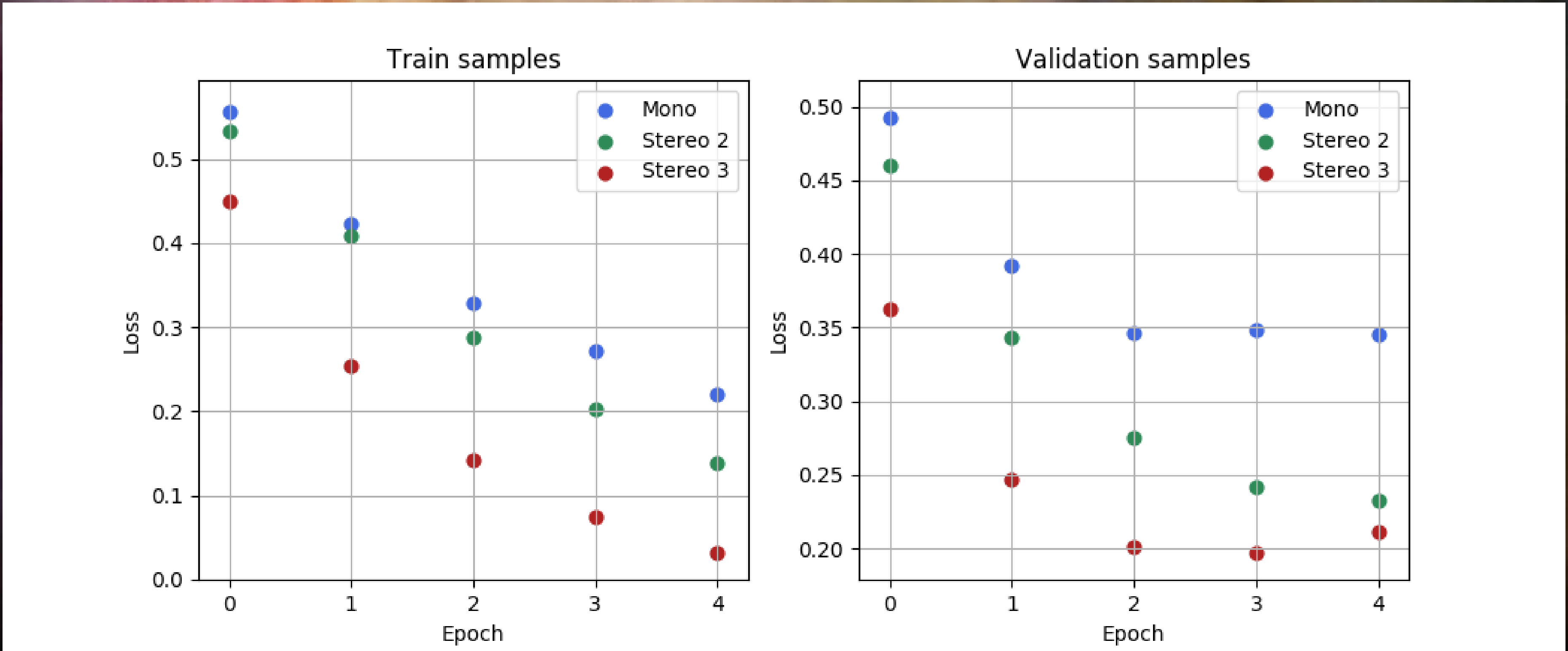
IACT: CNN structure

- CNN Structure for mono mode
  - 3 convolutional layers
  - 3 dense layers
  - 1 output neuron
  - total number of weights is 365 471



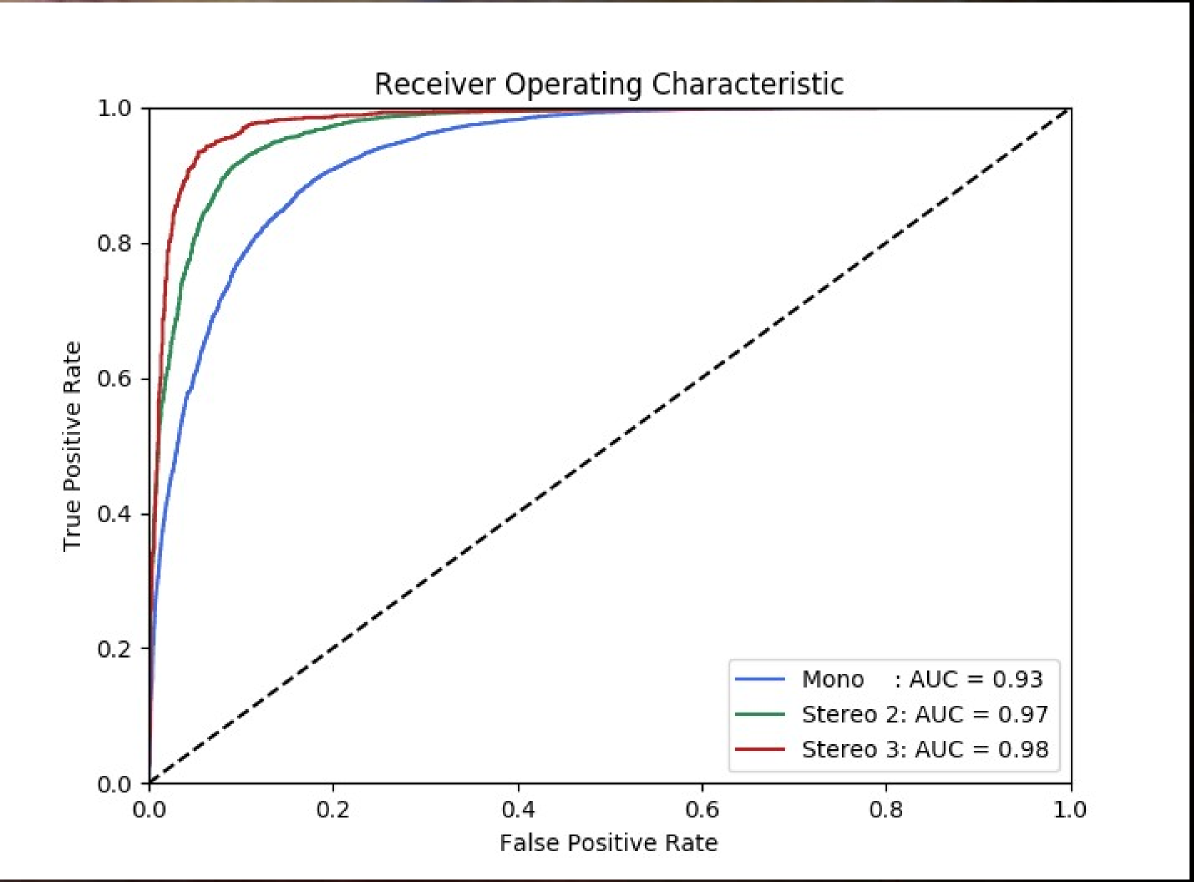


Learning process



IACT: ROC

- Receiver Operator Characteristic (ROC) id a dependence TPR from FPR
- The ROC curve is the curve most often used to represent the results of binary classification in machine learning.
- We have two classes: gammas and hadrons



IACT: Particle type identification

	Train	N_{tot}	N_g	$N_{g,g}$ (TP)	$N_{h,g}$ (FP)	S_{before}	S_{after}	Q	Precisi on
Mono	44246	14748	7800	6948	213	64.23	476.07	7.44	0.95
Stereo2	44246	7374	3887	2090	45	63.94	439.30	6.86	0.98
Stereo3	44246	3246	1927	1014	14	72.09	577.65	8.01	0.99

- Signal/Noise relation
- Quality factor
- Precision

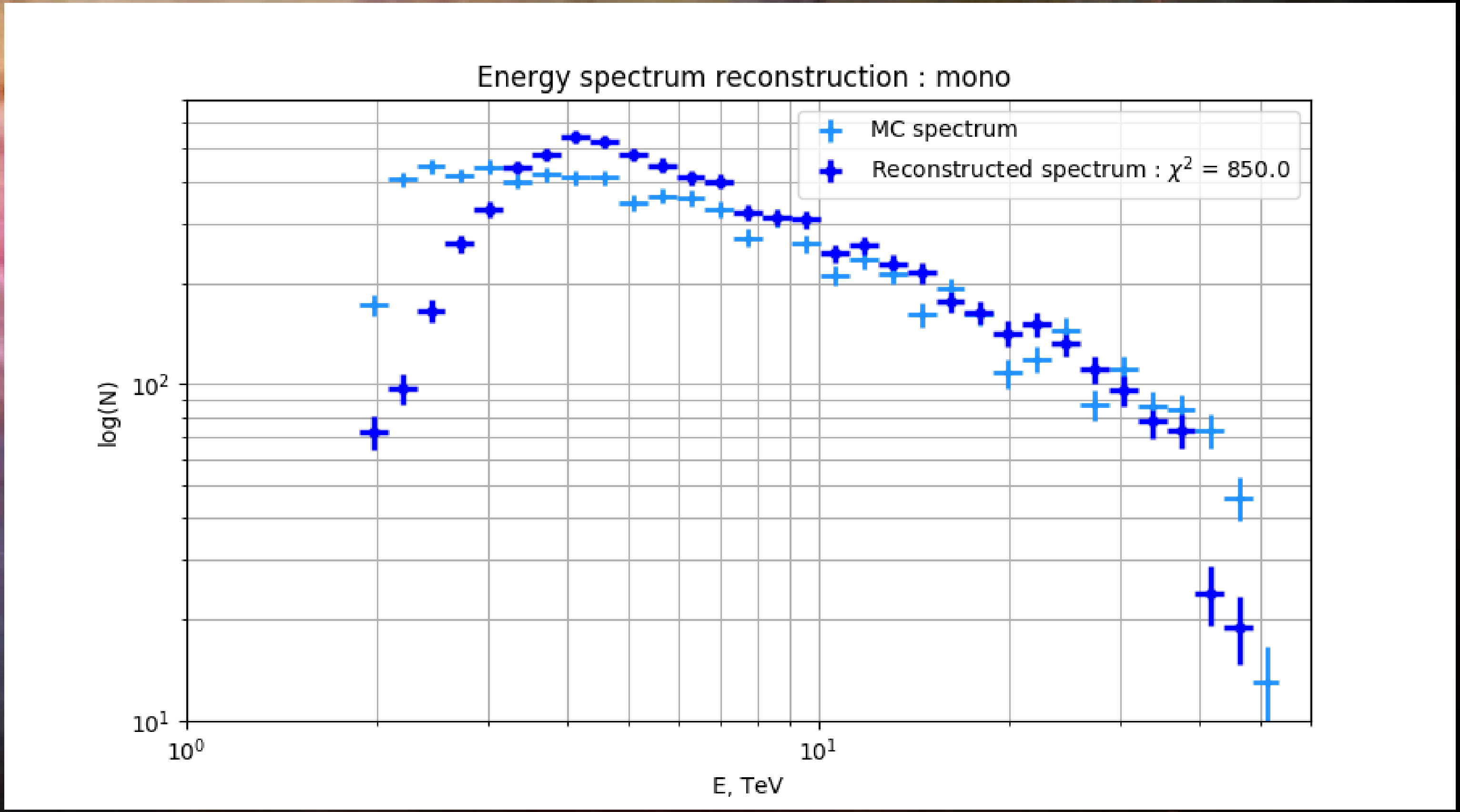
$S_{bef}=N_g/\sqrt{N_h}$
$S_{aft}=N_{g,g}/\sqrt{N_{h,g}}$

$Q=S_{aft}/S_{bef}=(N_{g,g}/\sqrt{N_{h,g}})/(N_g/\sqrt{N_h})$

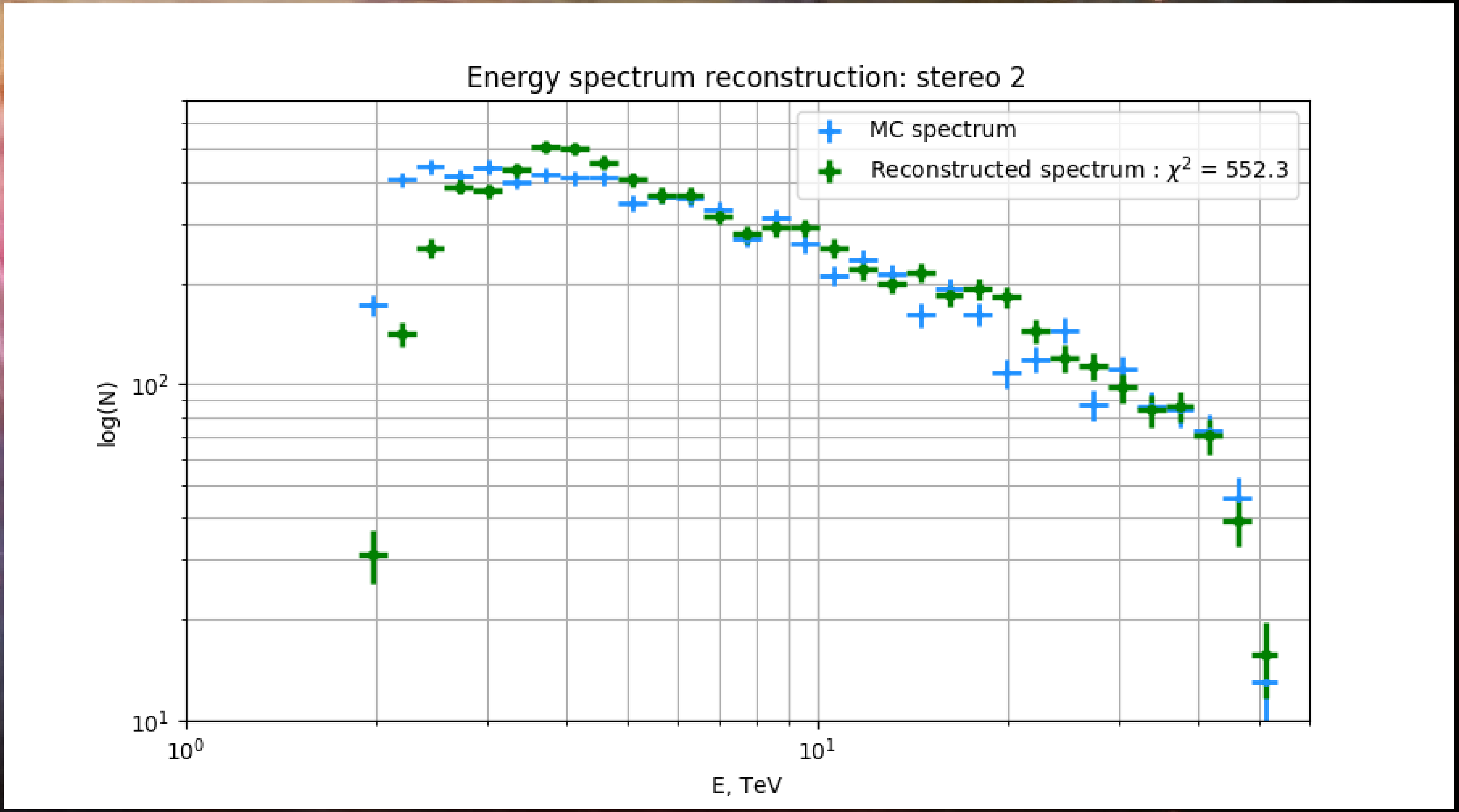
Precision=TP/(TP+FP)

IACT: Gamma spectrum (Mono)

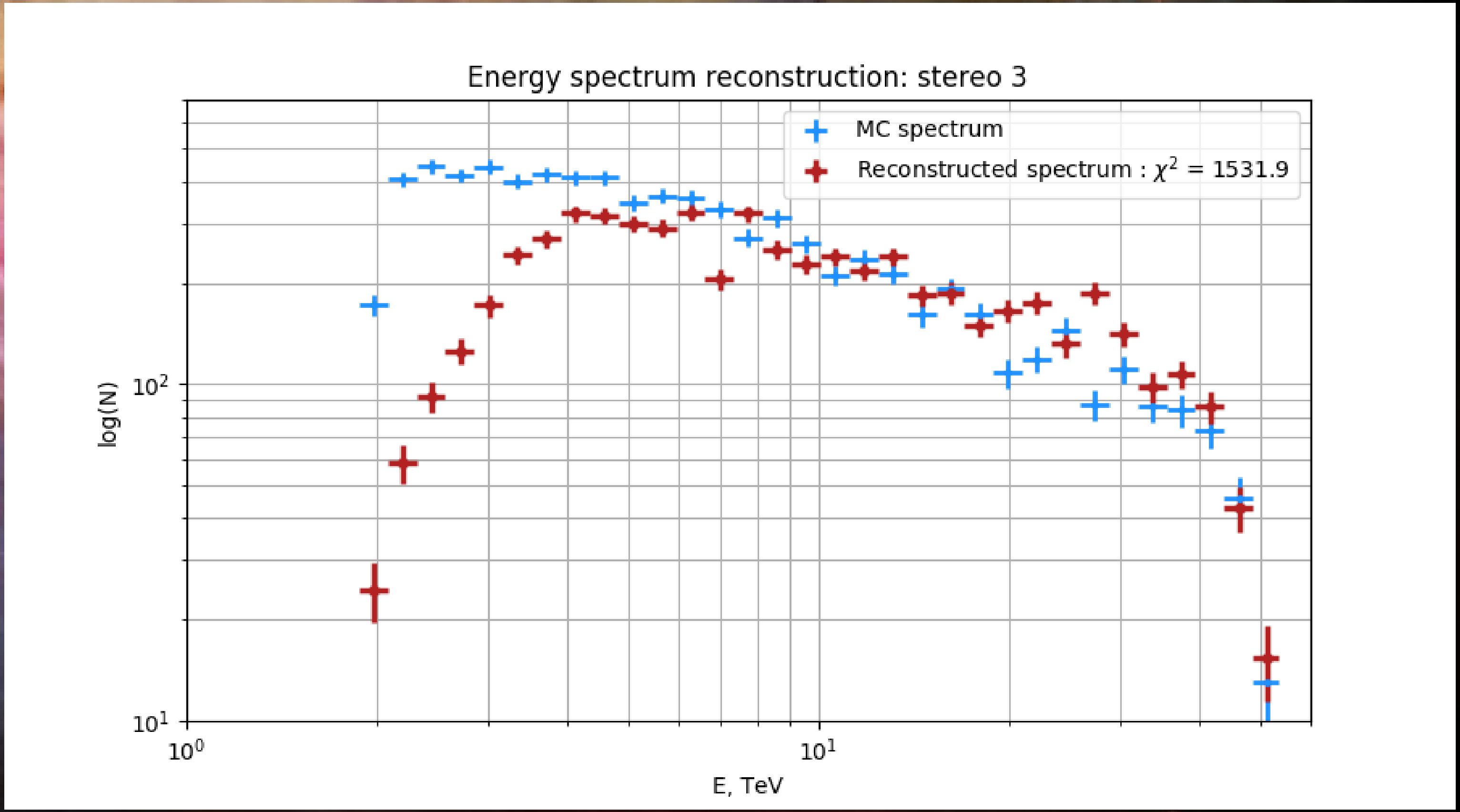
- Rel. err = 0.32



IACT: Gamma spectrum (Stereo2)



IACT: Gamma spectrum (Stereo3)



Conclusion

- ML is a flexible and universal data analysis tool
- CNN gives good results for particle identification, energy and other parameter values from IACT images.
- To get much better results, you need to use a more complex ANN and much more MC events for both training and testing.

Thank you for attention

QUESTIONS?

Back slides

Обозначения:

N — номер прогона

Eps — средняя относительная ошибка: Eps = 1/K * sum(1..K, |E_mc-E_rec| / E_mc), K — количество событий

<Eps> — средняя по прогонам относительная ошибка: <Eps> = 1/N * sum(1..N, Eps)

Sig^2 — среднеквадратичная ошибка: Sig^2 = 1/K * sum(1..K, (E_mc-E_rec)^2 / K), K — количество событий

<Sig^2> — средняя по прогонам среднеквадратичная ошибка: <Sig^2> = 1/N * sum(1..N, Sig^2)

Hi^2 — критерий хи-квадрат: Hi^2 = 1/K * sum(1..M, (N_mc-N_rec)^2 / N_mc), N_** - число событий в бине спектра, M — количество бинов (40)

<Hi^2> — средний по прогонам критерий хи-квадрат: <Hi^2> = 1/N * sum(1..N, Hi^2)

N	Eps	<Eps>	Sig^2	<Sig^2>	Hi^2	<Hi^2>	Value of loss function in train set	Value of loss function in test set
1	0.339	0.327	5.86	5.85	0.163	0.166	4.46	5.86
2	0.317		5.89		0.131		4.29	5.89
3	0.316		5.87		0.136		3.82	5.87
4	0.407		6.78		0.267		4.44	6.78
5	0.301		5.73		0.132		4.26	5.73
6	0.321		5.81		0.217		4.23	5.82
7	0.315		5.65		0.123		4.21	5.65
8	0.315		5.63		0.140		4.42	5.63
9	0.308		5.66		0.180		4.19	5.66
10	0.325		5.60		0.168		4.53	5.60