Strange Jet Tagging

Yuichiro Nakai

T. D. Lee Institute & Shanghai Jiao Tong U.

Based on YN, D. Shih and S. Thomas, arXiv:2003.09517.

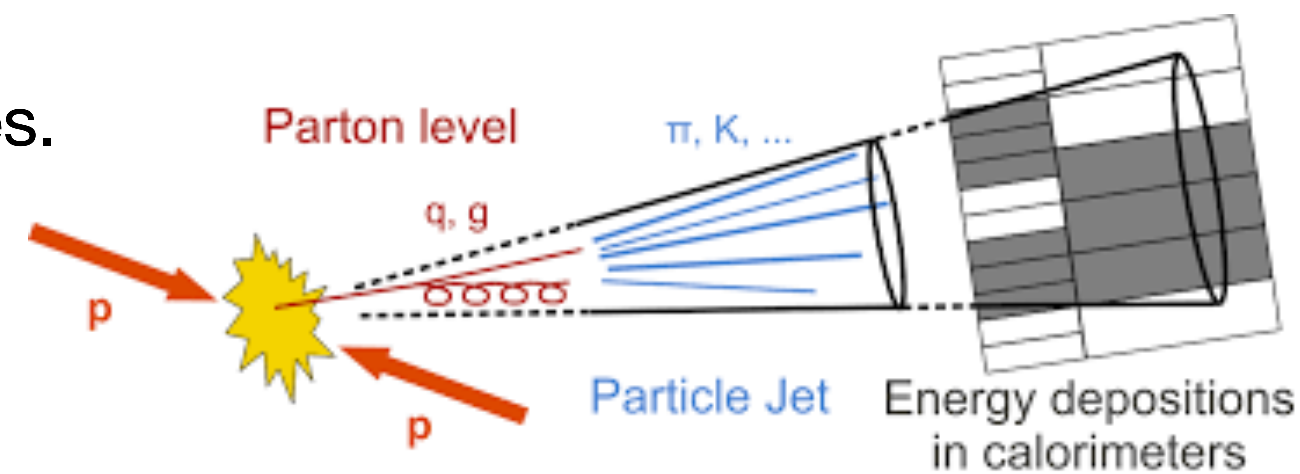
Jets at colliders

Jet : collimated bunch of hadrons as the signatures of quarks and gluons produced in high-energy collisions

✓ QCD partons are never observed isolated due to confinement.

✓ They give cascades of radiation (parton shower) by QCD processes.

✓ Hadrons are formed at $\sim \Lambda_{\mathrm{QCD}}$

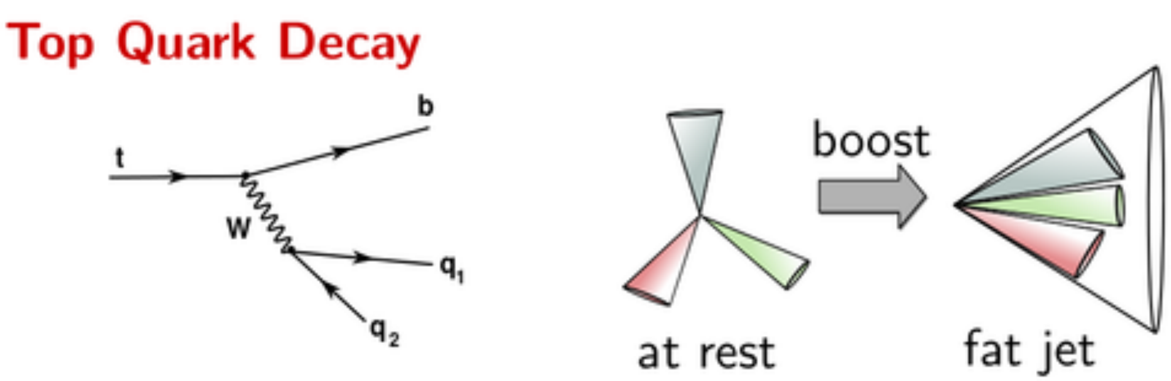


Understanding jets is a key ingredient of physics measurements and new physics searches at colliders.

What initial parton produces a jet ?

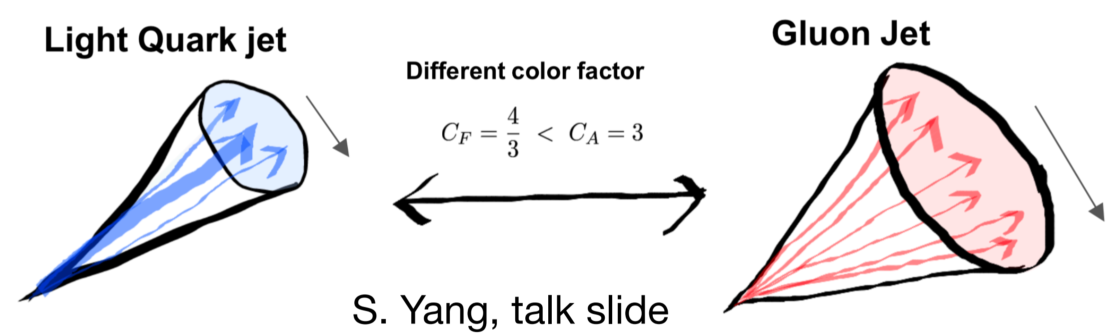
Quark and Gluon Tagging

☑ Top quark



Jet mass, N-subjettiness, …

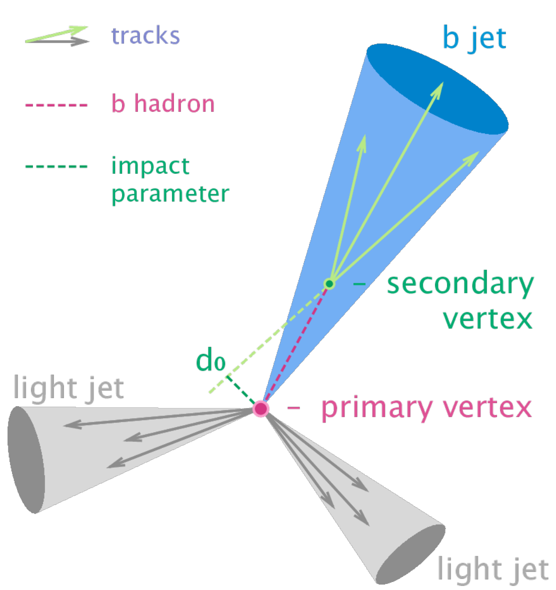
☑ Gluon



S. Yang, talk slide

More constituents with more uniform energy fragmentation and wider.

☑ Bottom/Charm



Look for a displaced (secondary) vertex.

Wikipedia

☑ Up-type vs Down-type

p_T-weighted jet charge

$$Q_\kappa^i = \frac{1}{(p_T^{\text{jet}})^\kappa} \sum_{j \in \text{jet}} Q_j (p_T^j)^\kappa$$

The last missing piece :

Strange quark tagging?

Applications of Strange Tagging

• CKM mixings

The CKM matrix elements are fundamental parameters of the SM and their precise determination is important.

However… **The values for $|V_{cs}|$ and $|V_{cd}|$ are not measured very well.**

Because the charm quark mass is too heavy to be considered light but not heavy enough to treat in the heavy quark limit.

One process to probe $|V_{cs}|$ is through **semileptonic decays $D \to Ke\nu$**.

Our best effort is to use lattice QCD :

$$V_{cs} = 0.98 \pm 0.01_{exp} \pm 0.10_{th}$$

The experimental error is small but the theoretical error is huge !

W boson decay $W \to cs$ gives the most direct measurement of $|V_{cs}|$ if strange tagging is possible.

Applications of Strange Tagging

- **Top quark reconstruction**

  ✓ **All-hadronic channel**

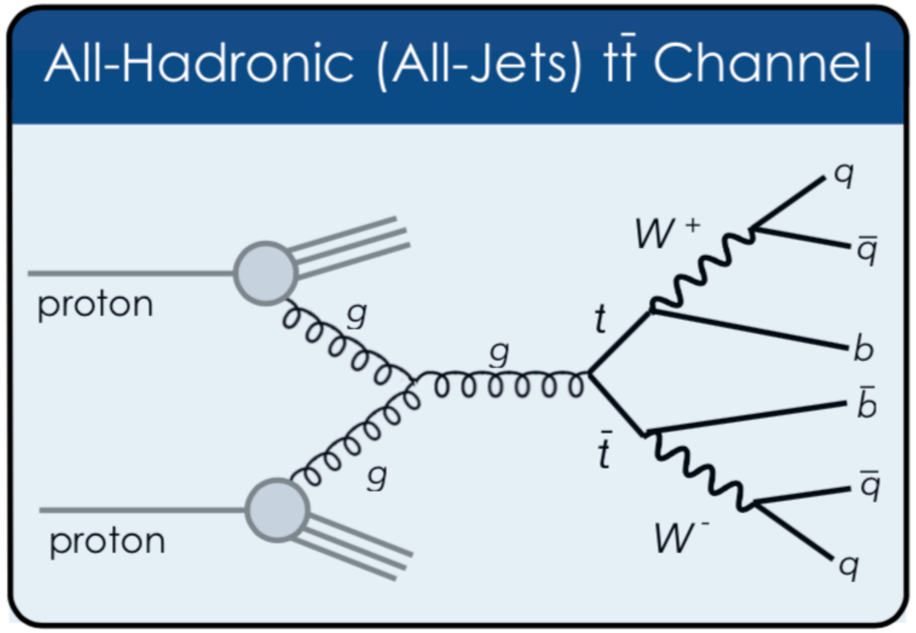
  Full event reconstruction is possible.

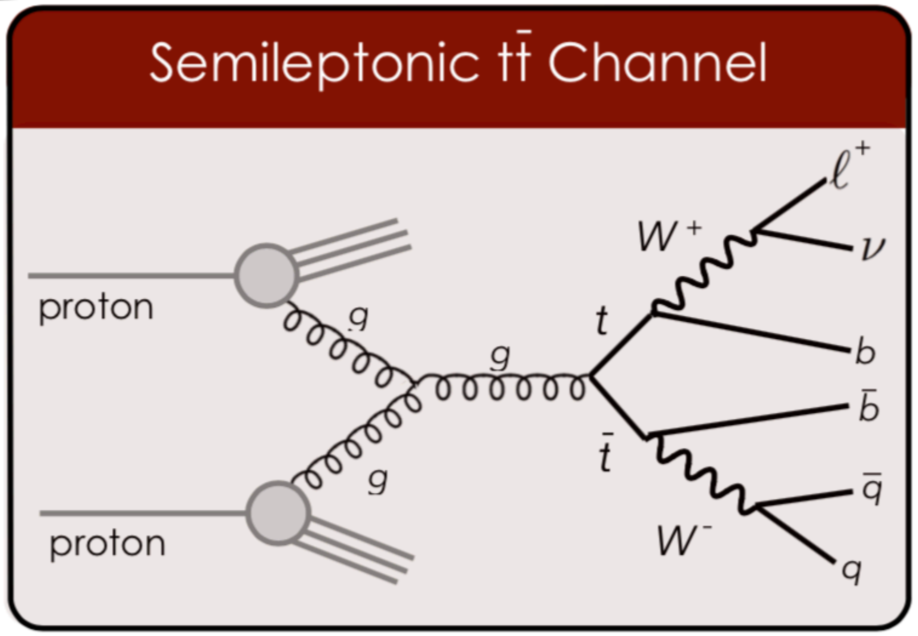
  Jet combinatorics and large multi-jet background are problematic.

  ✓ **Semileptonic channel**

  Leptonic top identifies event and hadronic top can be reconstructed.

  Jet combinatorics and multi-jet background are still issues.





T. McCarthy, talk slide

Which jets are W → cs, us, cd, ud decay products?

Identification of strange jet may give some help.

Applications of Strange Tagging

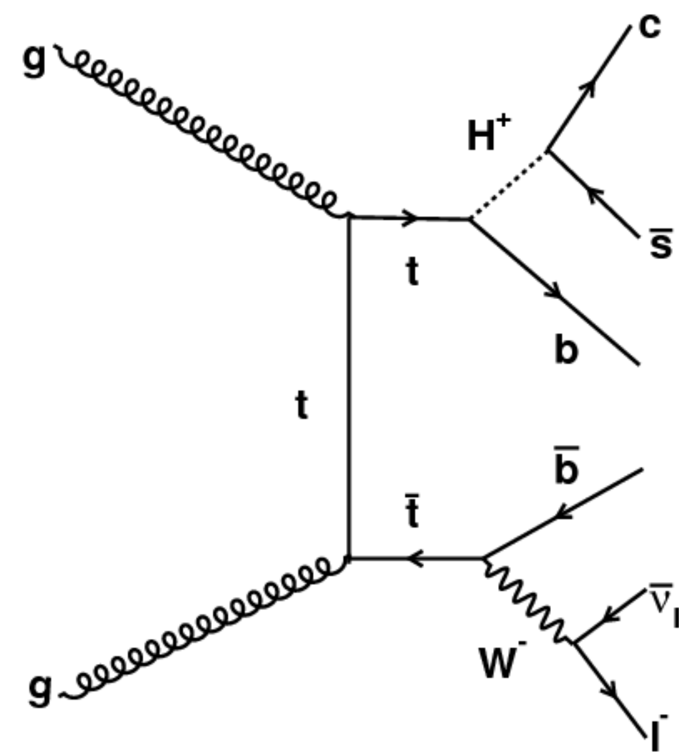
- **Light charged Higgs search**

Production : $t\overline{t} \rightarrow W^{\pm}bH^{\mp}\overline{b}$

Decay : $H^{+} \rightarrow c\overline{s}$

😰 The same issue as top quark reconstruction is applied.

😱 We do not know the charged Higgs mass!



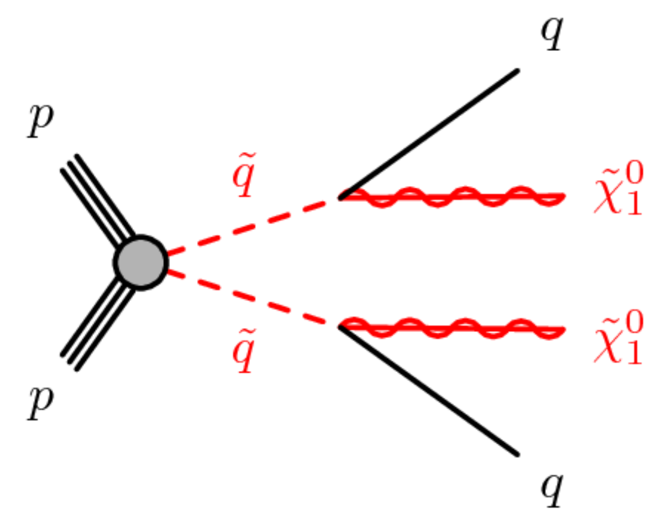
CMS

- **Squark search**

Identification of strange jet can …

✓ reduce the background $Z(\rightarrow \nu\nu) +$ jets

✓ identify squark flavor after the discovery



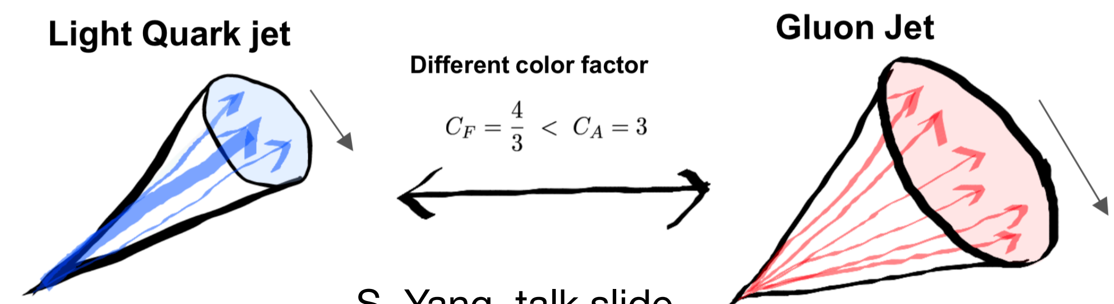
CMS

Strange tagging may be important !

Tagging Strategy

- **Strange vs Gluon**

  We can expect the same thing as quark/gluon discrimination.



S. Yang, talk slide

More constituents with more uniform energy fragmentation and wider.

- **Strange vs Up**

  We can expect the same thing as up/down discrimination.

p_T-weighted jet charge

$$Q_\kappa^i = \frac{1}{(p_T^{\text{jet}})^\kappa} \sum_{j \in \text{jet}} Q_j (p_T^j)^\kappa$$

- **Strange vs Down**

  Possible ??

  Both are quarks with the same charge.

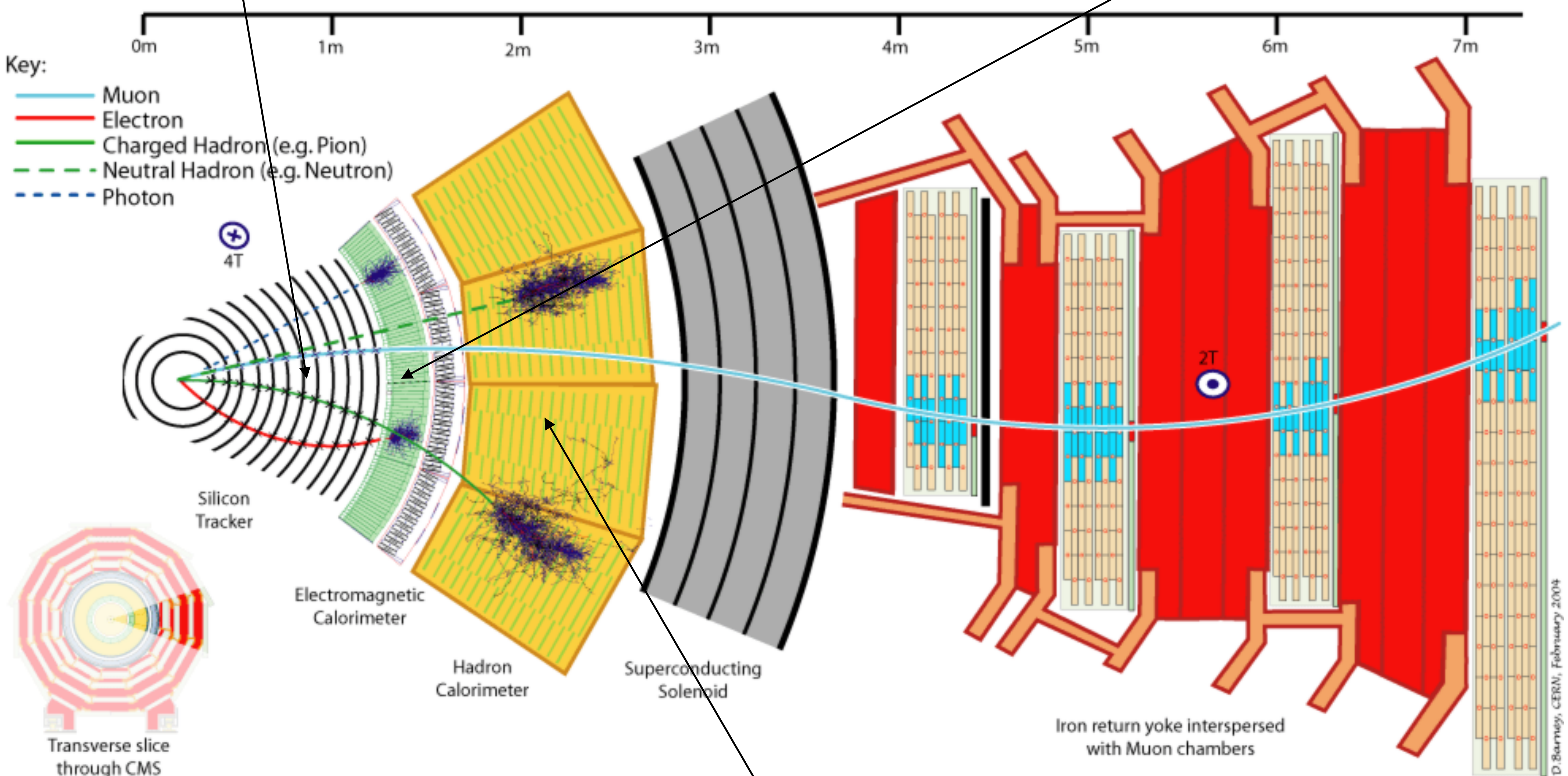
Main theme of this talk

Tagging Strategy

CMS experiment at the LHC

Tracker : trajectories of charged particles

ECAL : energy of electrons and photons



HCAL : energy deposits of hadrons

Tagging Strategy

After hadronization, strange quarks form Kaons :

$$K^- = s\bar{u}, \quad K^+ = \bar{s}u, \quad K_L \approx \frac{s\bar{d} - d\bar{s}}{\sqrt{2}}, \quad K_S \approx \frac{s\bar{d} + d\bar{s}}{\sqrt{2}}$$

No difference

$\boxed{K_L,\ K^\pm}$ $\quad \gamma c\tau \sim 3$ m

No decay inside detectors

$\boxed{K_S}$ $\quad \gamma c\tau \sim 3$ cm

Decay inside detectors

$$K_S \to \pi^+\pi^-(\sim 70\%), \quad \pi^0\pi^0(\sim 30\%)$$

Detector responses to hadrons :

Down jets

	K_L	K_S	$K^\pm$	π^0	$\pi^\pm$
HN	○	△			
ECAL		△		○	
Tracker		△	○		○

Strange jets

Hadronic Neutral (HN) = HCAL - Tracker

K-long (and K-short) can be used for tagging !

Tagging Strategy

K-short behaves very differently in detectors depending on decay length and decay mode.

$\boxed{K_S}$ $\gamma c\tau \sim 3$ cm

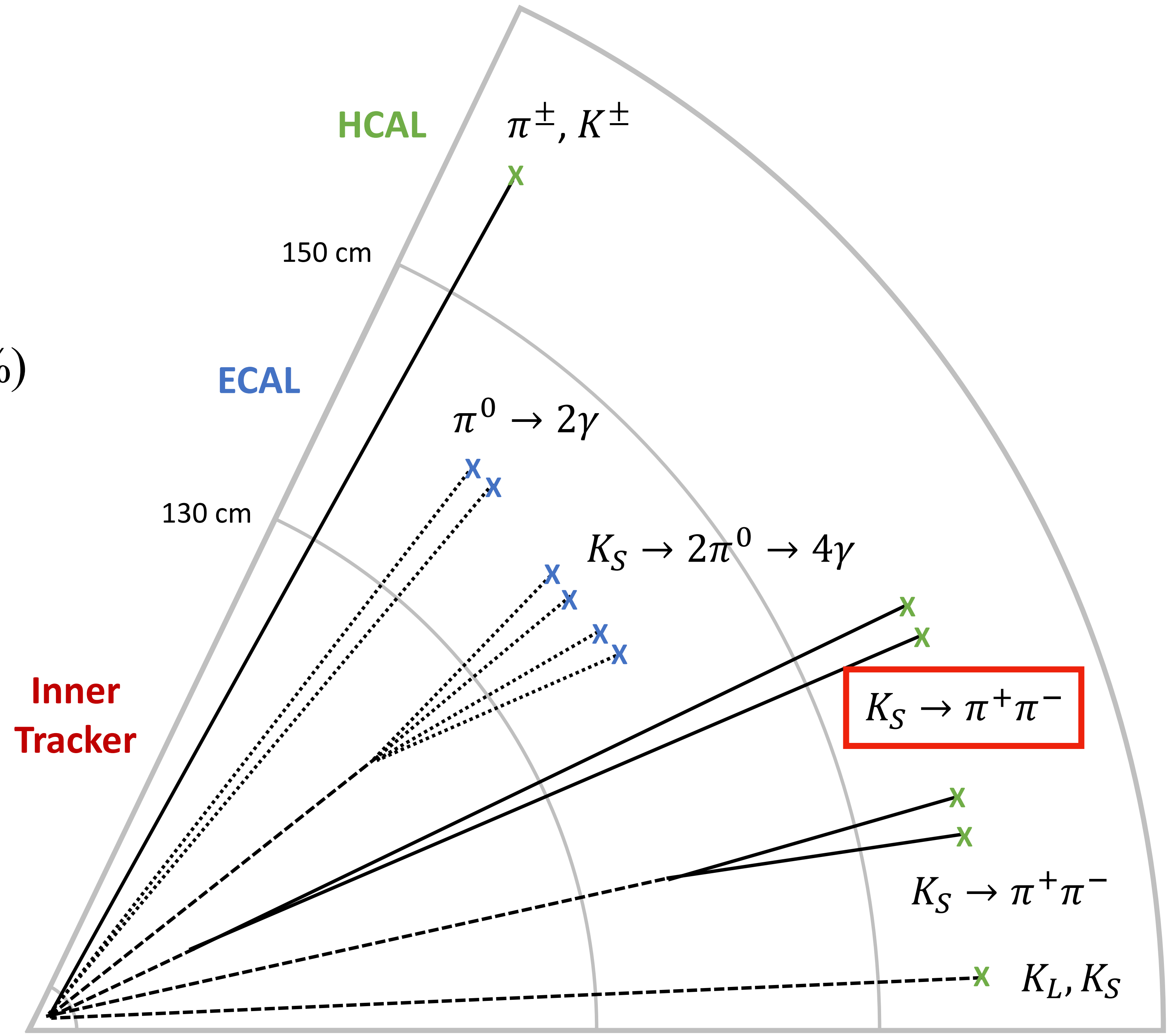
Decay inside detectors

$$K_S \to \pi^+\pi^-(\sim 70\%),\quad \pi^0\pi^0(\sim 30\%)$$

$K_S \to \pi^+\pi^-$ inside tracker leaves charged track pairs from the secondary vertex.

↓

We can identify such K-shorts by reconstructing invariant mass !



Jet Samples

Generate strange/down jet samples by using MadGraph, PYTHIA and Delphes.

1M events for each case of :

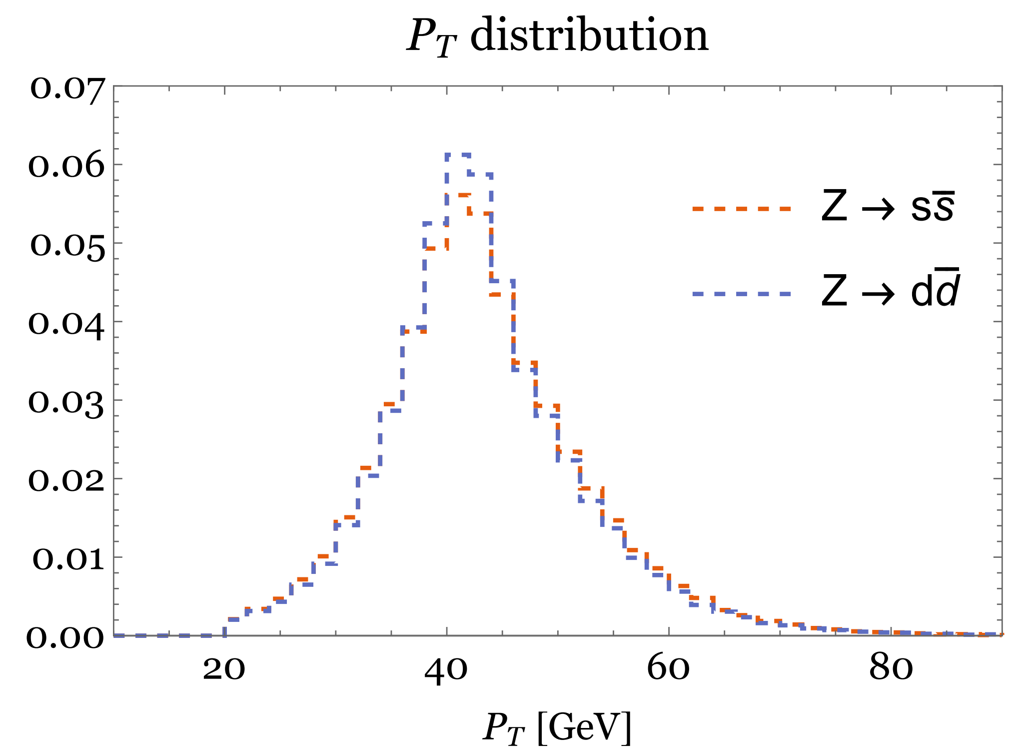
$$Z \to s\bar{s} \quad (p_T > 20\ \mathrm{GeV})$$
$$Z \to d\bar{d} \quad (p_T > 20\ \mathrm{GeV})$$

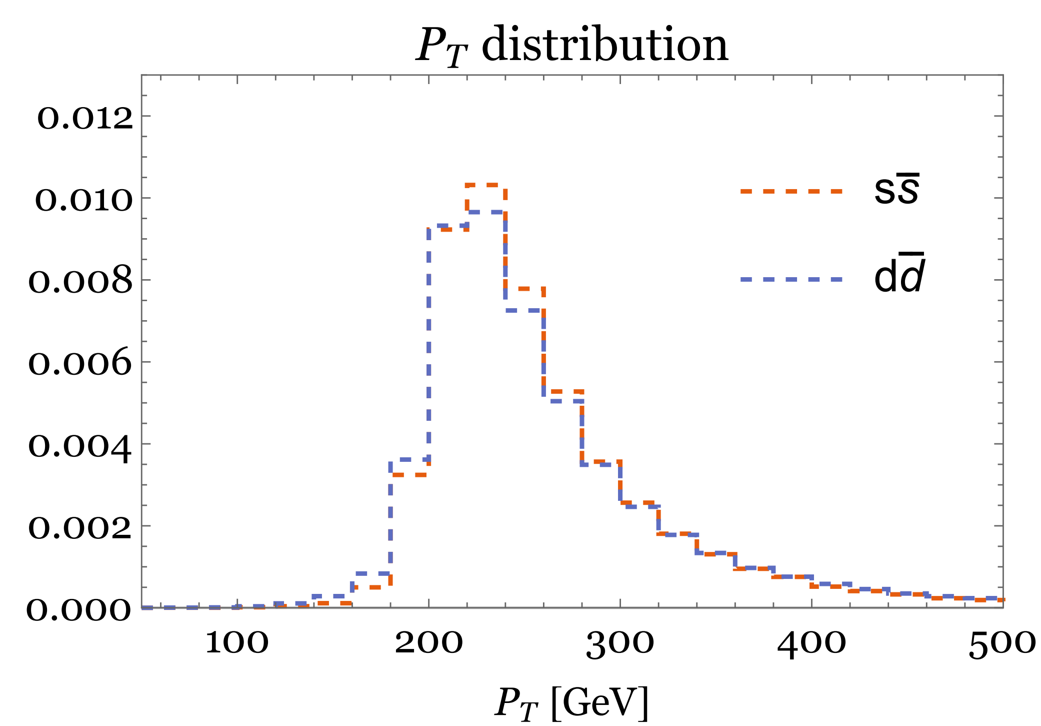
$$s\bar{s} \quad (p_T > 200\ \mathrm{GeV})$$
$$d\bar{d} \quad (p_T > 200\ \mathrm{GeV})$$

$$(|\eta| < 0.05)$$

Initial parton is required to be inside the leading jet : $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 0.4$

Herwig gives the similar results.



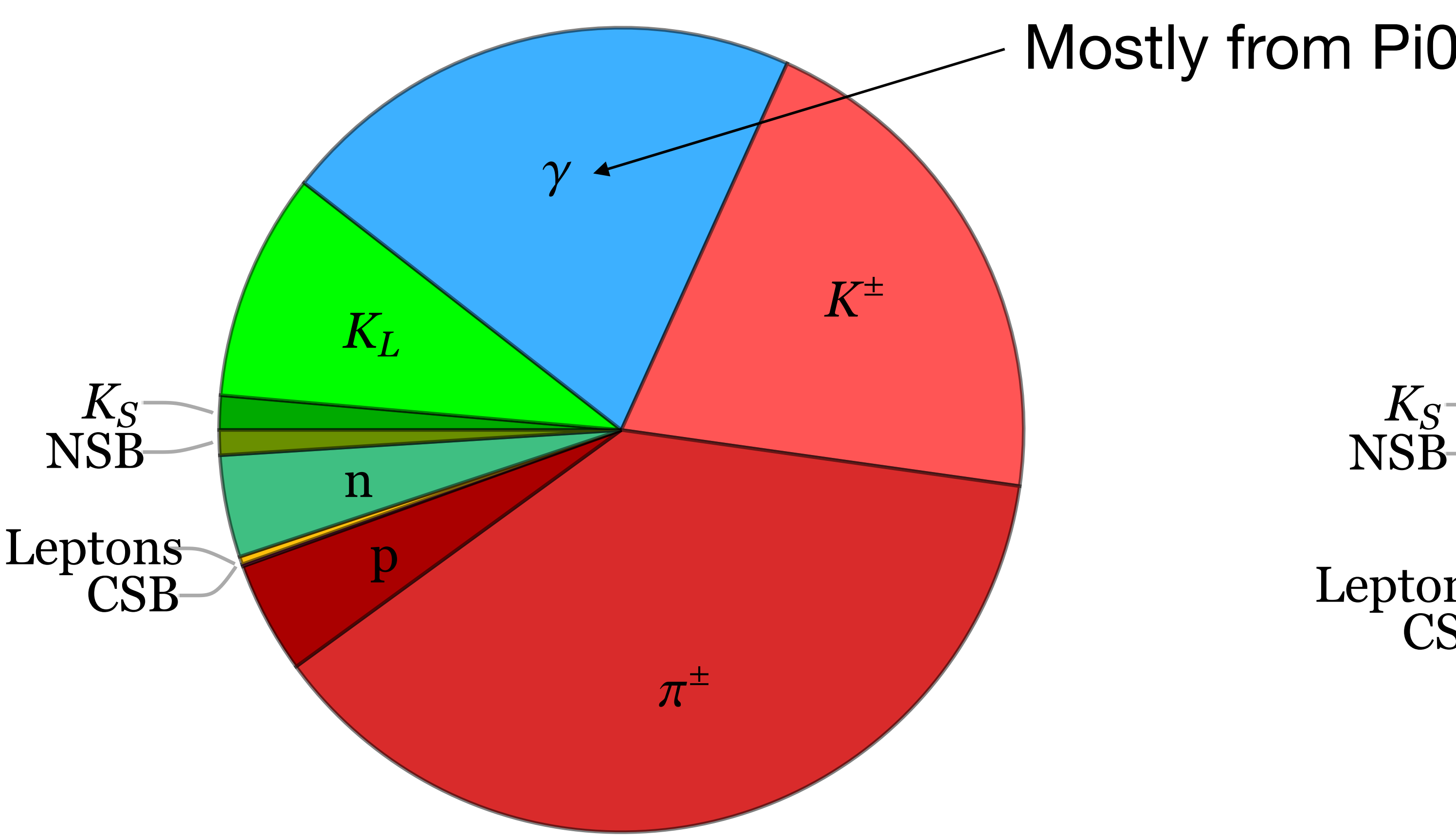


Ingredients of strange/down jets

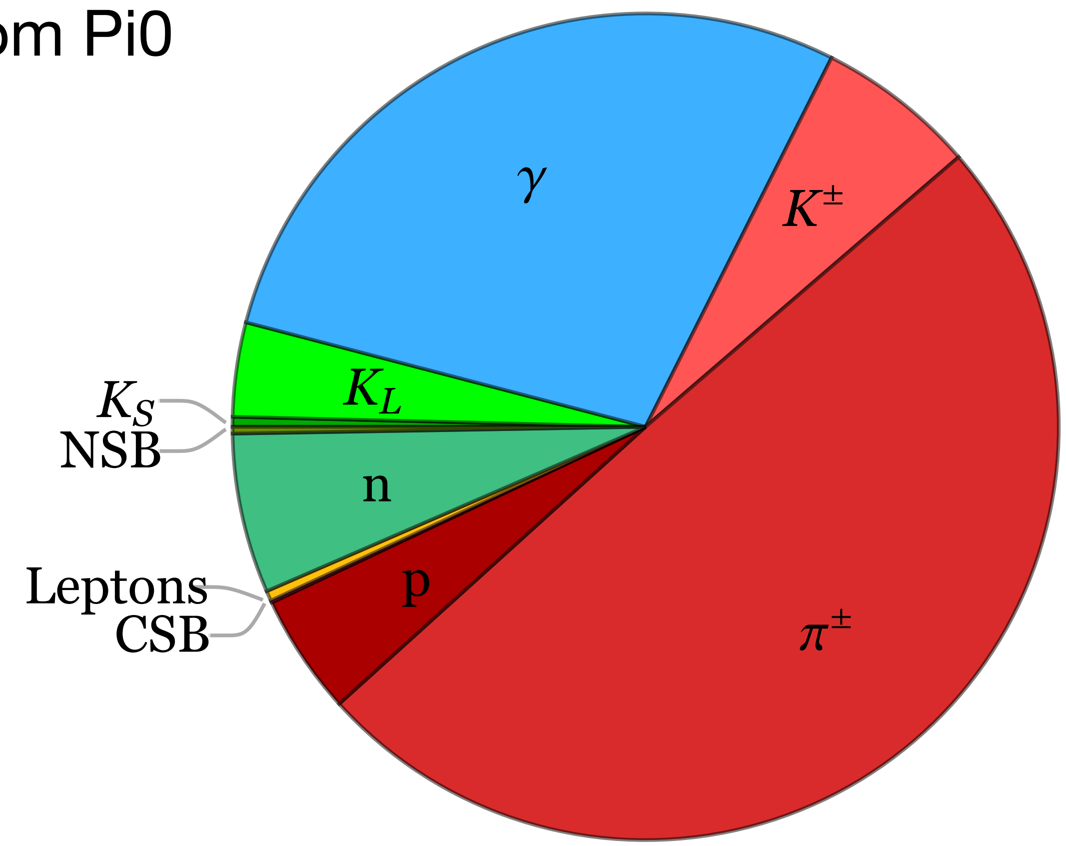
Strange jets contain more energetic Kaons than down jets.

The pT fraction of a detector-stable particle averaged over jet samples :

$Z \to s\bar{s} \quad (p_T > 20\ \text{GeV})$



$Z \to d\bar{d} \quad (p_T > 20\ \text{GeV})$



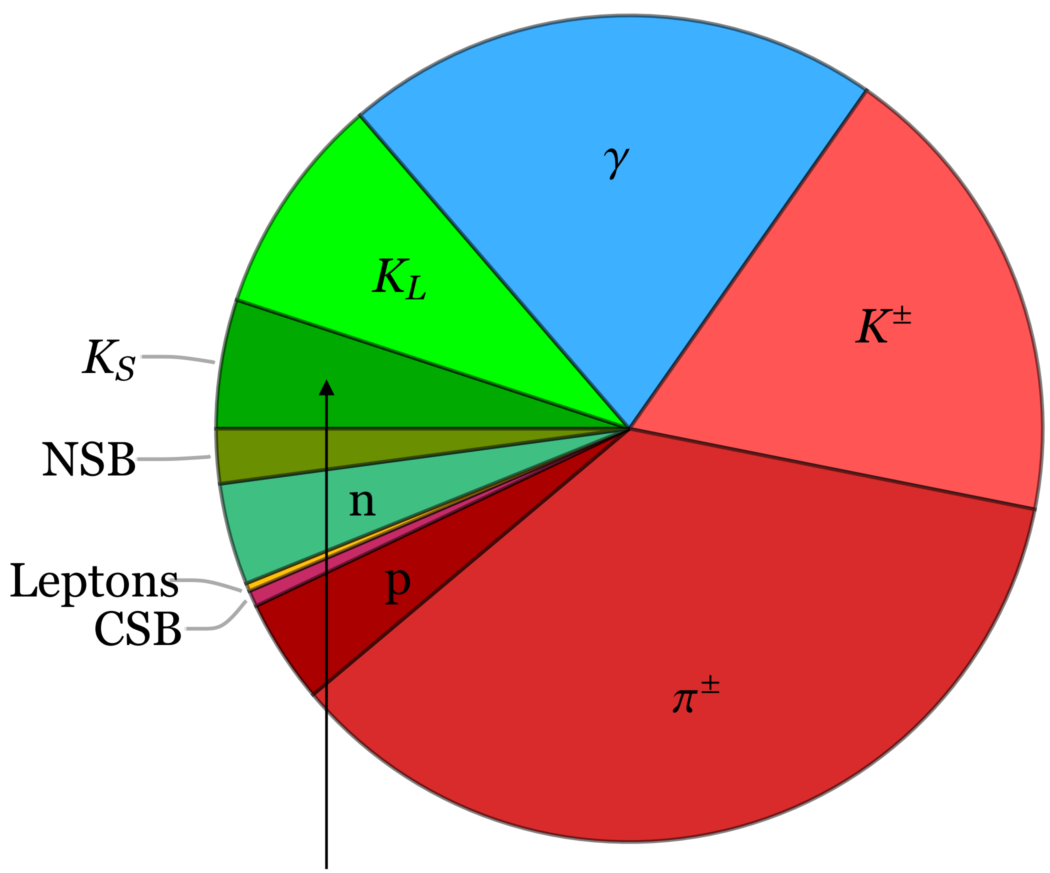
NSB: neutral strange baryons, CSB: charged strange baryons

Ingredients of strange/down jets

Strange jets contain more energetic Kaons than down jets.

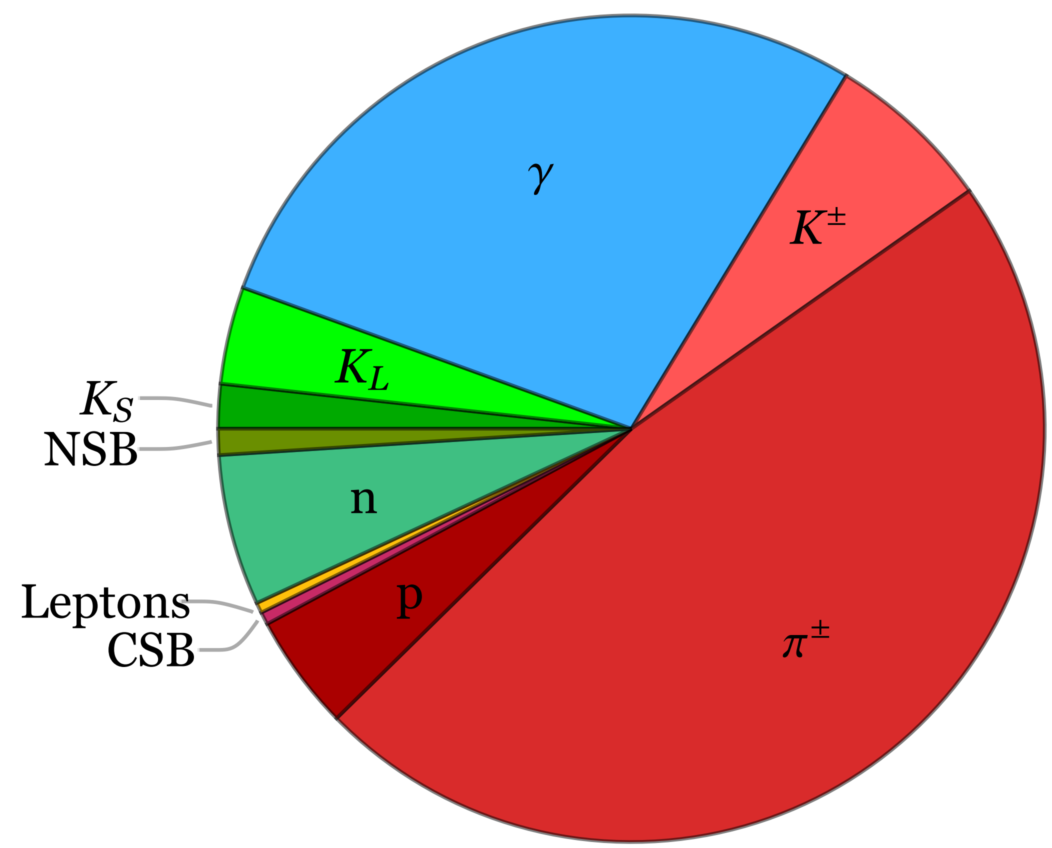
The pT fraction of a detector-stable particle averaged over jet samples :

Strange $P_T > 200$ GeV



More K-shorts due to boost factor

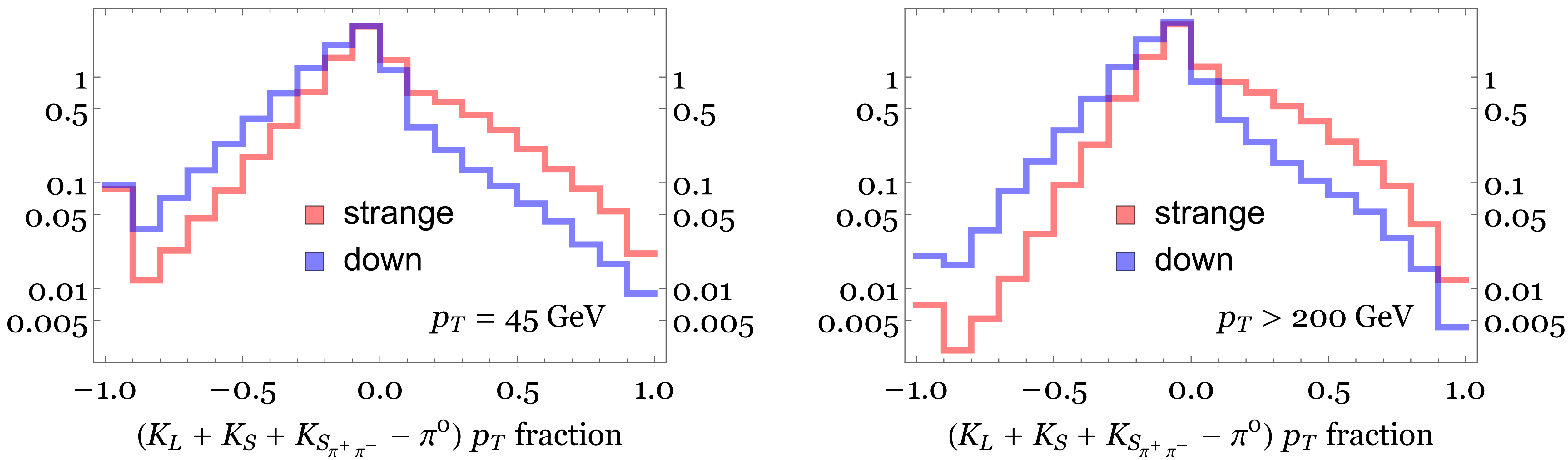
Down $P_T > 200$ GeV



NSB: neutral strange baryons, CSB: charged strange baryons

Truth-level classifier

Classification problem : **Strange jet** vs **Down jet**



Use **Boosted Decision Tree (BDT)** for classification.

Inputs have 3 dimensions: K_L , K_S , π^0 p_T

Approximately set the maximal performance we can achieve.

Tagging Strategy

After hadronization, strange quarks form Kaons :

$$K^- = s\bar{u}, \quad K^+ = \bar{s}u, \quad K_L \approx \frac{s\bar{d} - d\bar{s}}{\sqrt{2}}, \quad K_S \approx \frac{s\bar{d} + d\bar{s}}{\sqrt{2}}$$

No difference

$\boxed{K_L,\ K^\pm}$ $\gamma c\tau \sim 3\ \mathrm{m}$

No decay inside the detectors

$\boxed{K_S}$ $\gamma c\tau \sim 3\ \mathrm{cm}$

Decay inside the detectors

$$K_S \to \pi^+\pi^-(\sim 70\%),\ \ \pi^0\pi^0(\sim 30\%)$$

Detector responses to hadrons :

Down jets

	K_L	K_S	$K^\pm$	π^0	$\pi^\pm$
HN	○	△			
ECAL		△		○	
Tracker		△	○		○

Strange jets

Hadronic Neutral (HN) = HCAL - Tracker

K-long (and K-short) can be used for tagging !

Cut-Based Tagging

Classify each jet into strange jet (signal) or down jet (background).

Put a cut in distribution of **HN − E**.

Measures to estimate efficiency and accuracy of taggers

$$\varepsilon_S = \frac{\text{(Correctly classified into signals)}}{\text{(Total number of signal jets)}}$$

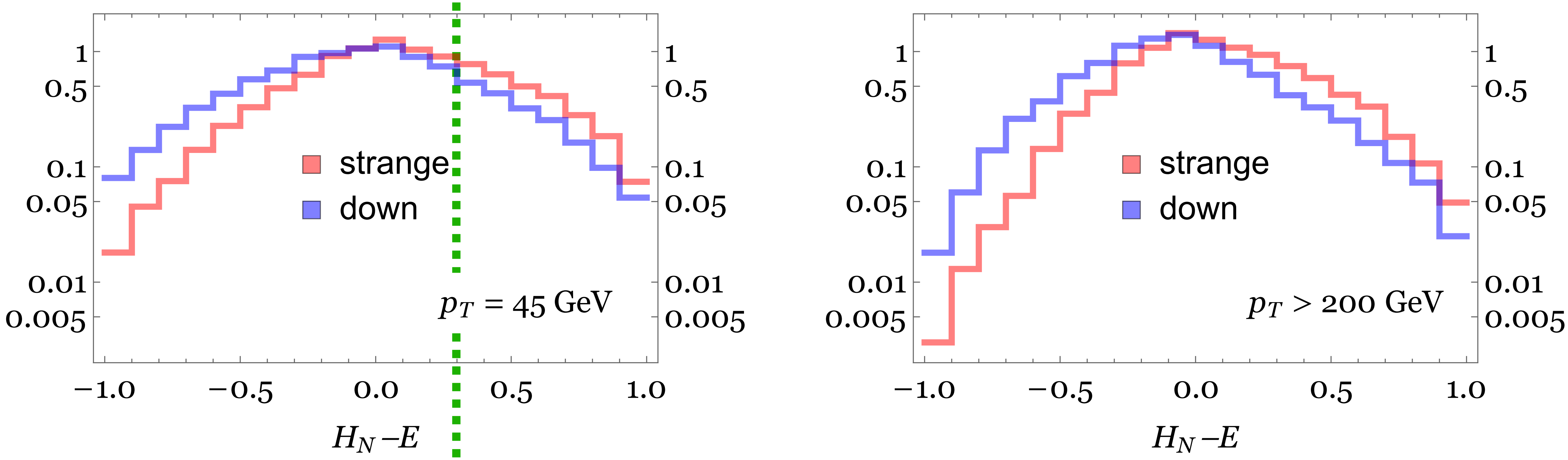
$$\varepsilon_B = \frac{\text{(Misclassified into signals)}}{\text{(Total number of backgrounds)}}$$

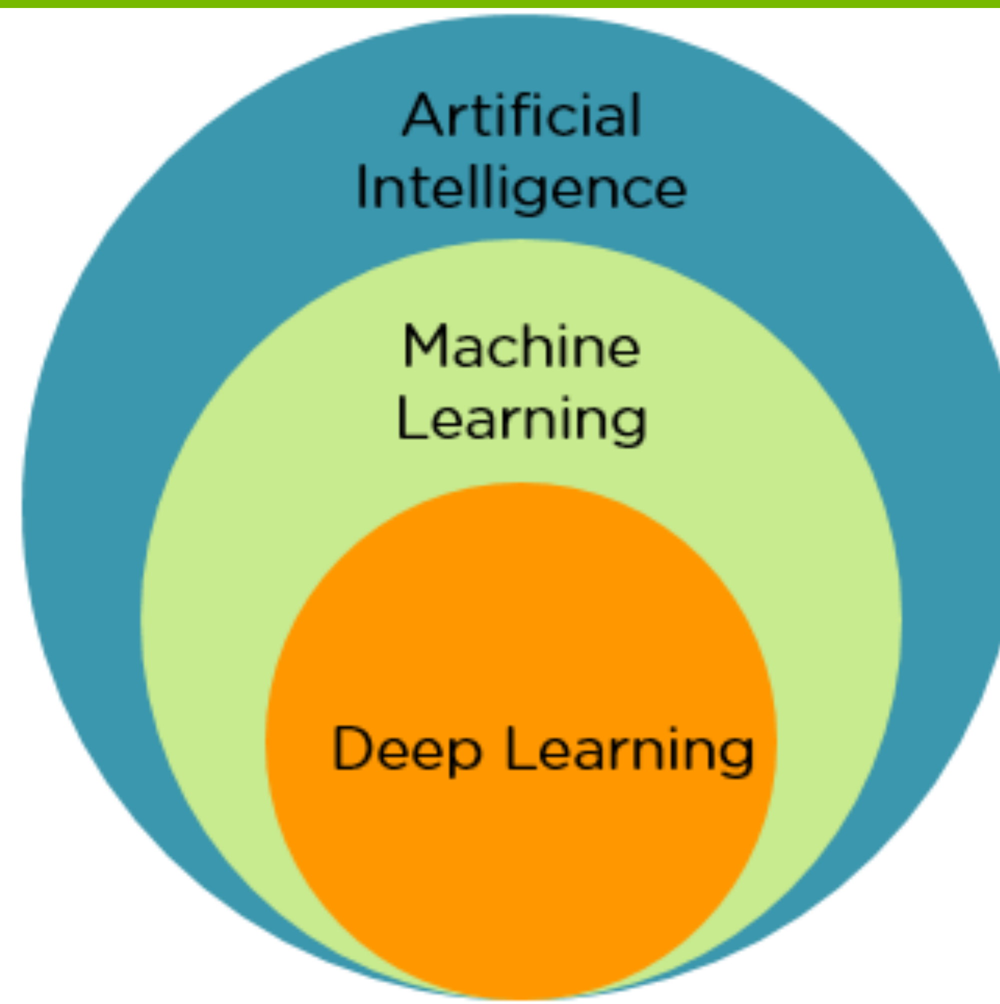
Larger ε_S

Larger ε_B

Smaller ε_S

Smaller ε_B





Recently, machine learning has shown tremendous progress.

Much of the current excitement concerns a subfield called deep learning.

What is Machine Learning?

Machine learning : technique to give computer systems the ability to learn with data <u>without being explicitly programmed</u>.

Machine can learn the feature of data which human has not realized !

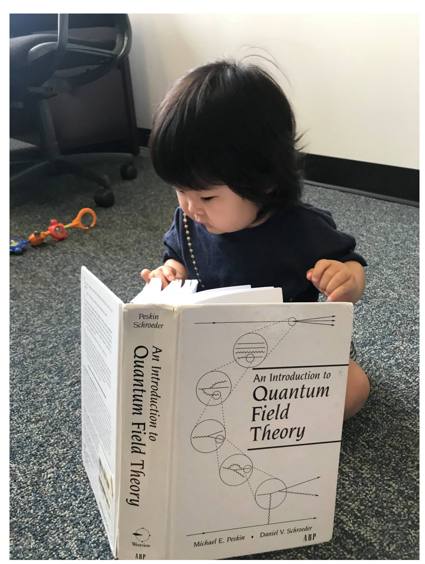
How our jobs are changed in collider searches?

Traditional approach

Think about physics

Design observables

Use on data



Machine learning

Think about input and model

Machine learns best observables from data (or simulations)

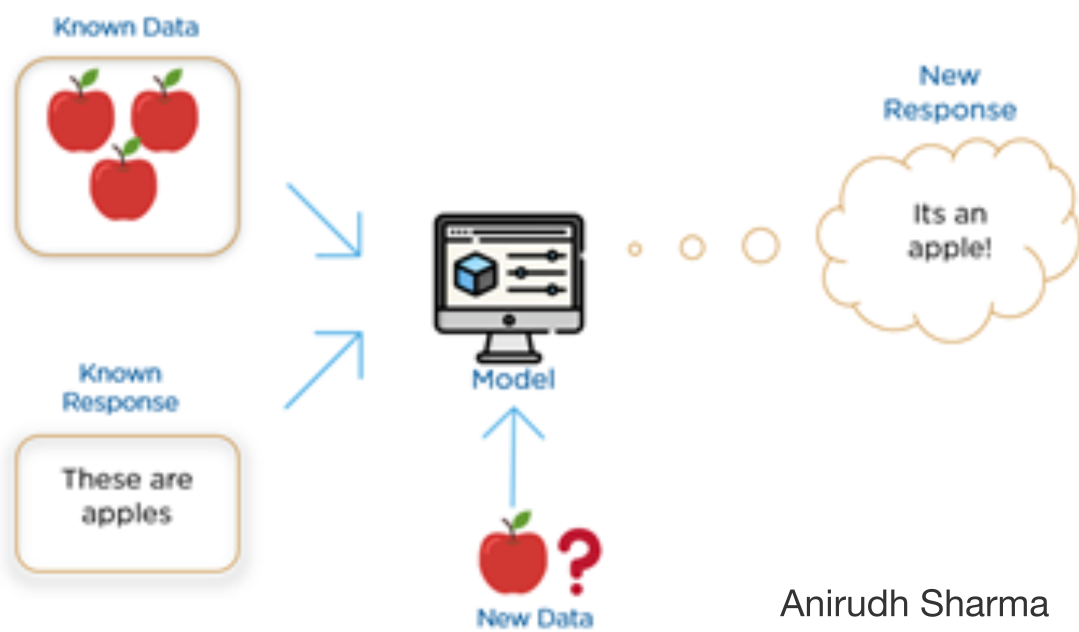
Use on data

What is Machine Learning?

Machine learning algorithms can be classified into:

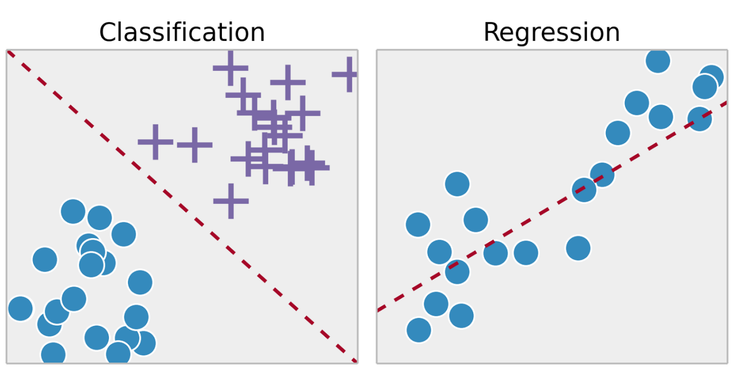
Supervised learning

Learn from labeled data



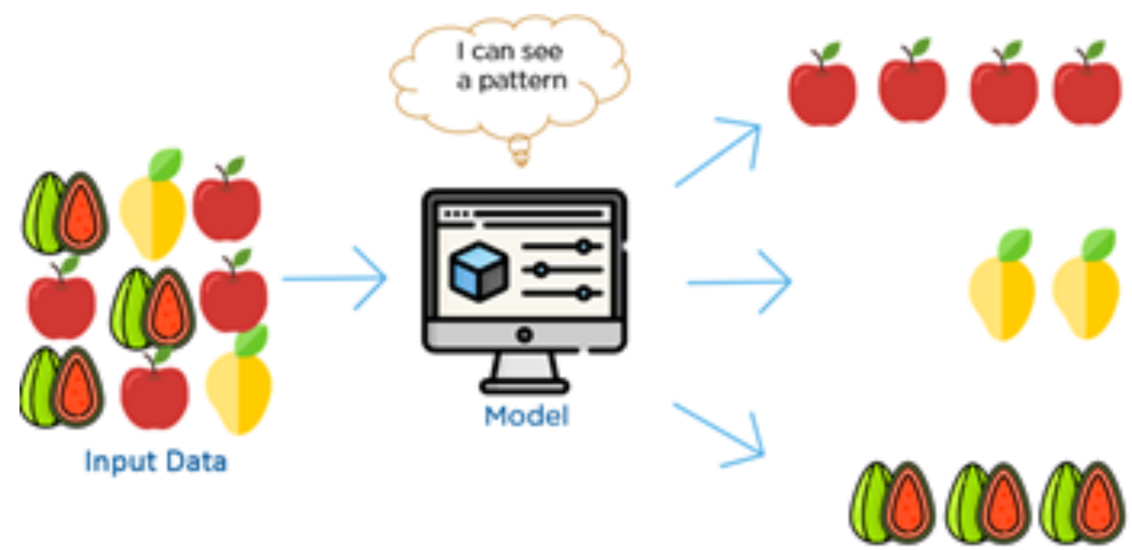
Anirudh Sharma

Applications)



Unsupervised learning

Learn from unlabeled data



The system looks for patterns and extracts features in data.

Applications) Clustering

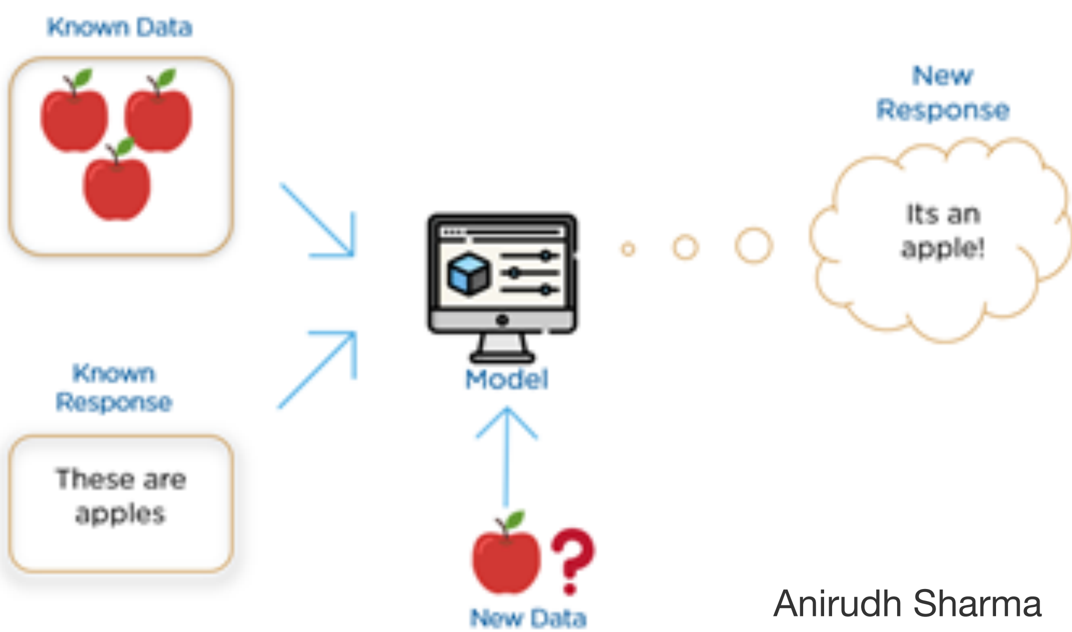
Anomaly detection

What is Machine Learning?

Machine learning algorithms can be classified into:

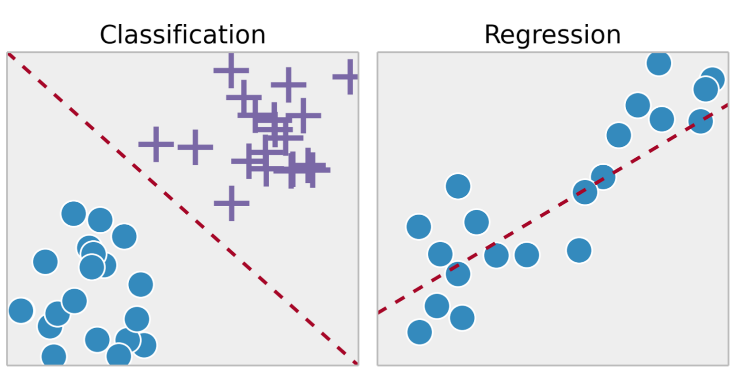
Supervised learning

Learn from labeled data



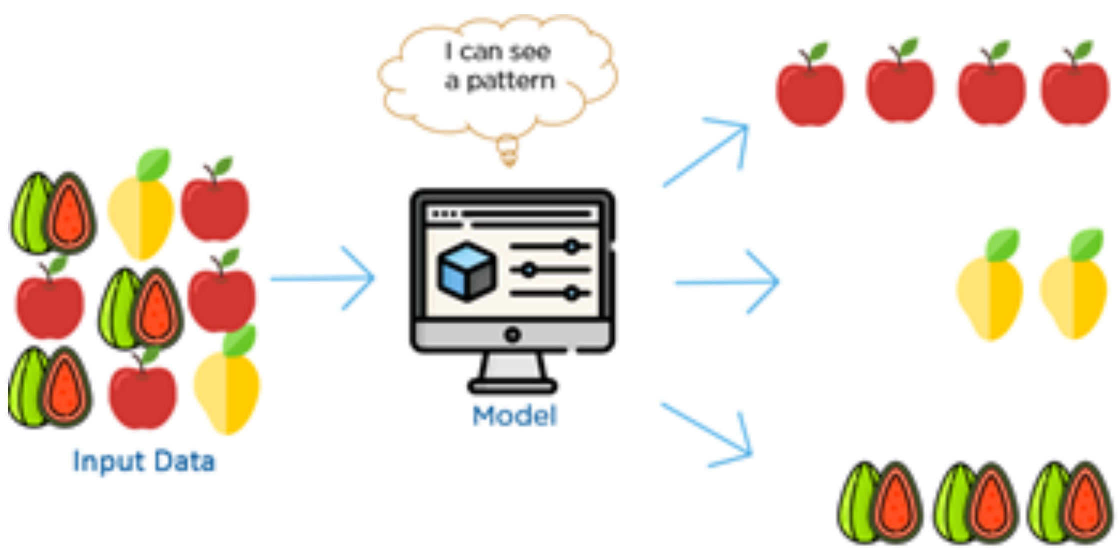
Anirudh Sharma

Applications)



Unsupervised learning

Learn from unlabeled data



The system looks for patterns and extracts features in data.

Applications) Clustering

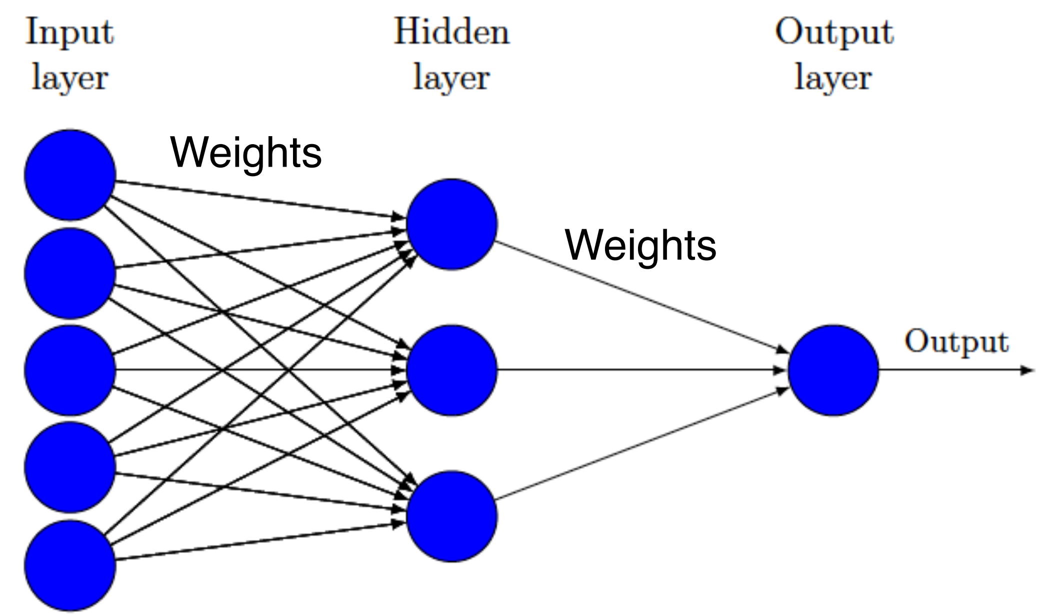
Anomaly detection

Neural Networks

✓ Powerful machine learning-based techniques used to solve many real-world problems

✓ Modeled loosely after the human brain and designed to recognize patterns

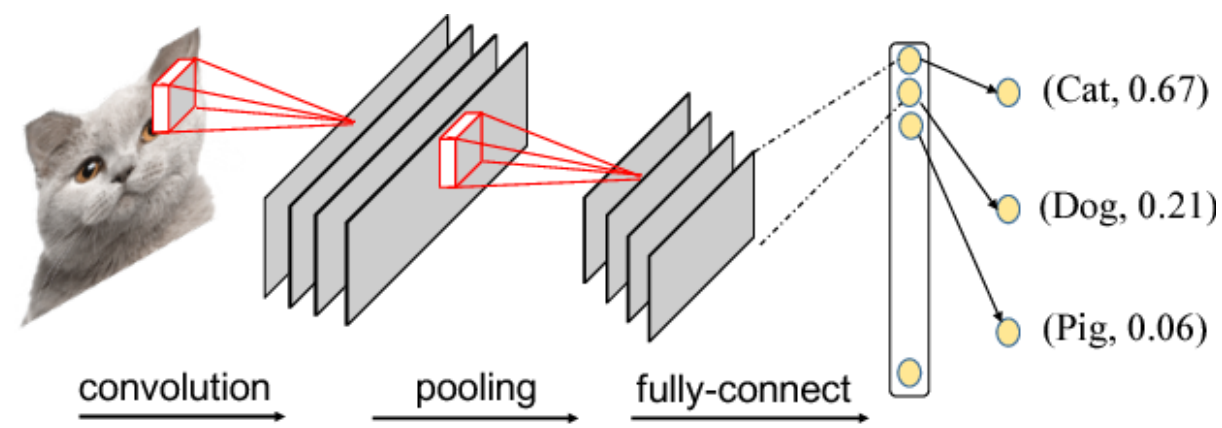
✓ Containing weights between neurons that are tuned by learning from data



Convolutional Neural Network (CNN)

✓ Show high performance for image recognitions

✓ Maintain the spacial information of images



arXiv:1712.01670

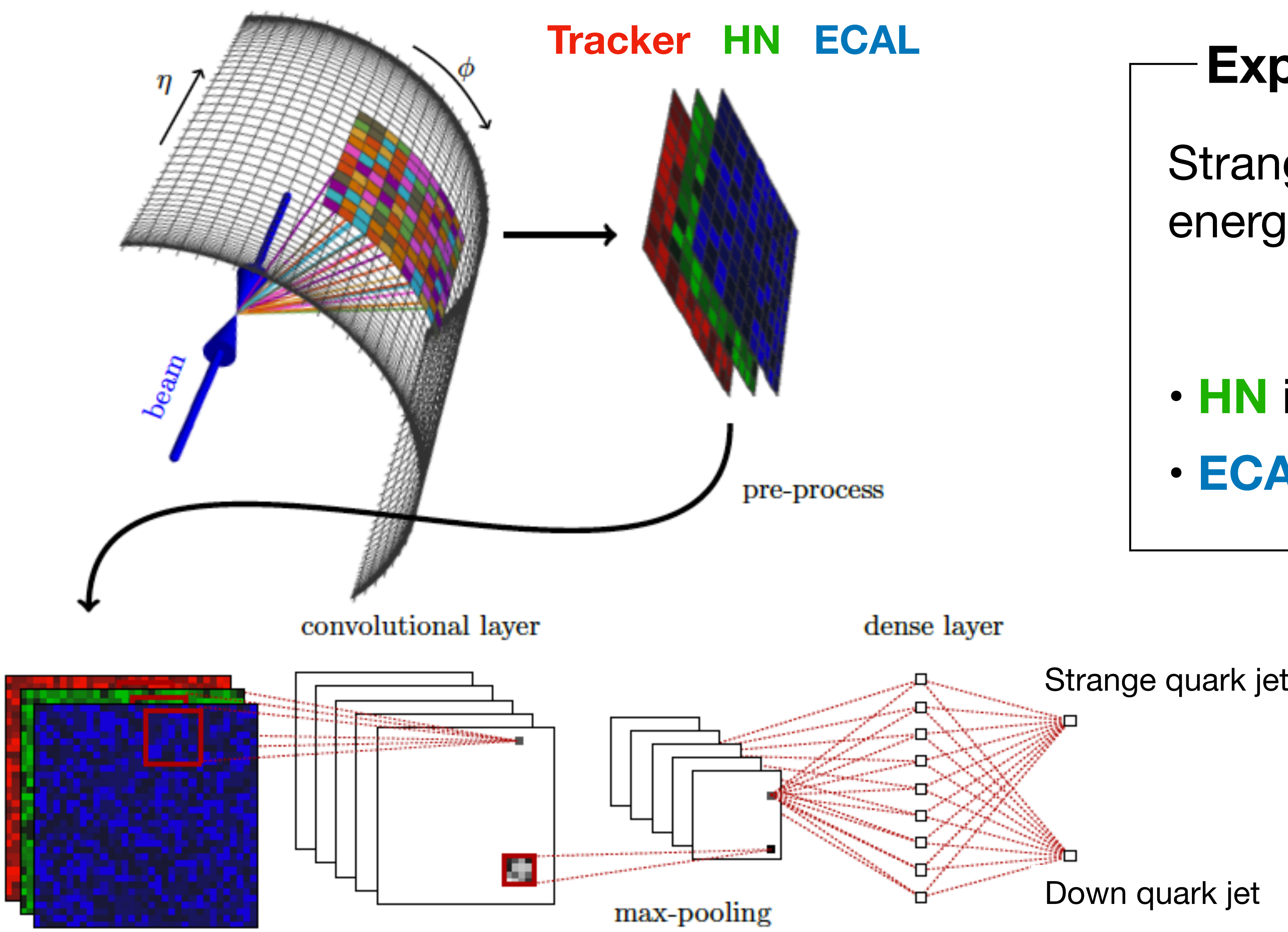
Apply a convolution operation to the input, passing the result to the next layer

Reduce the image size

Jet Images and CNN

Classification problem : **Strange jet** vs **Down jet**

Create jet images with colors (Tracker, HN, ECAL) and feed them into CNN.



Komiske, Metodiev, Schwartz (2016)

Expectation

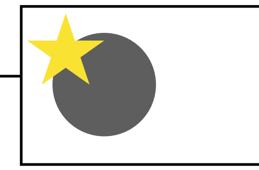
Strange jet contains more energetic neutral Kaons.

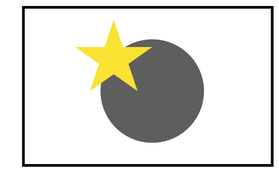
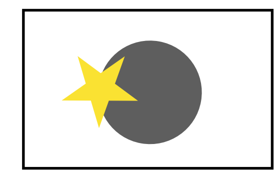
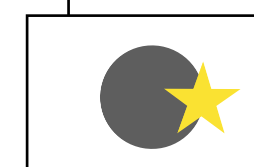
- **HN** image : brighter
- **ECAL** image : darker

Jet Images

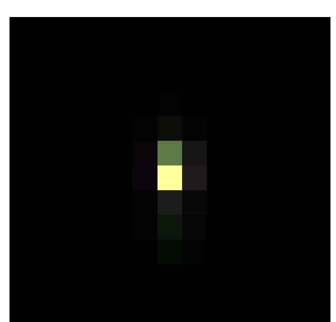
Create jet images with colors (Tracker, HN = HCAL - Tracker, ECAL).

Image pre-processing

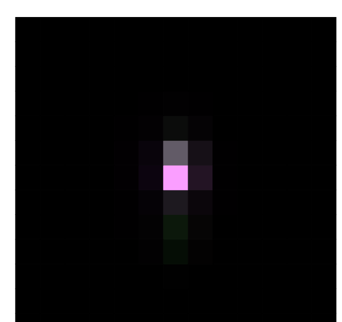


1. Shift an image so that the centroid is at the origin 
2. Rotate the image so that the major principal axis is vertical 
3. Flip the image so that the maximum intensity is in the upper right region 
4. Normalize the image to unit total intensity : $\sum_{jet}(\hat{p}_T^{track} + \hat{E}_{had} + \hat{E}_{em}) = 1$
5. Pixelate the image : $\Delta\eta = \Delta\phi = 1.2$ 13 x 13 pixels

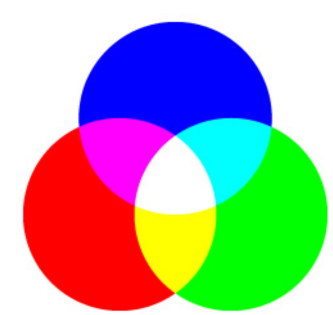
Average images :



$Z \to s\bar{s} \quad (p_T > 20 \text{ GeV})$

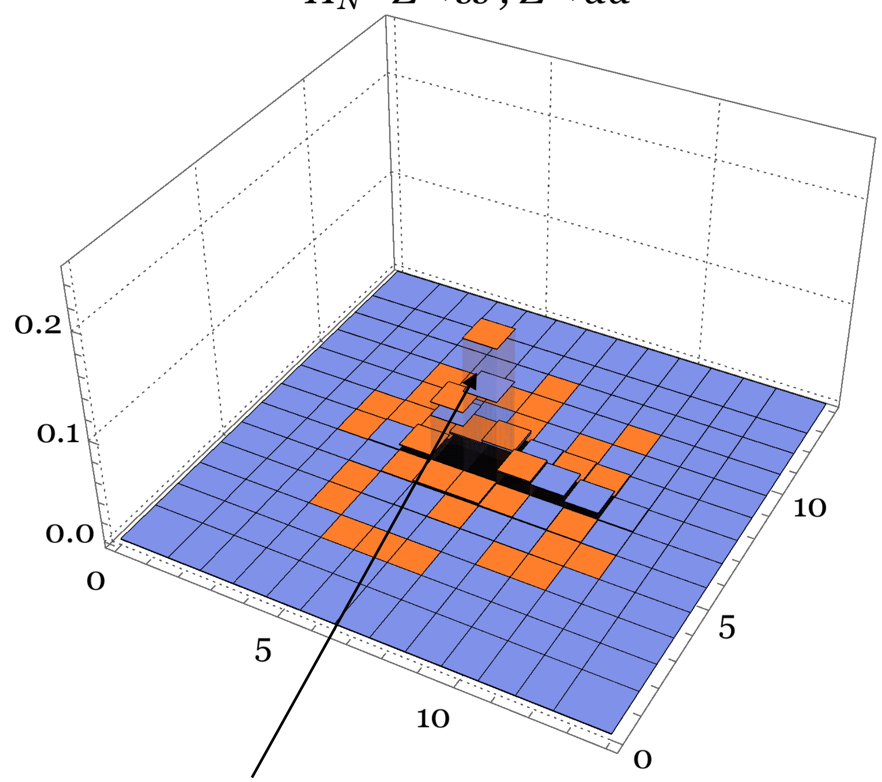


$Z \to d\bar{d} \quad (p_T > 20 \text{ GeV})$

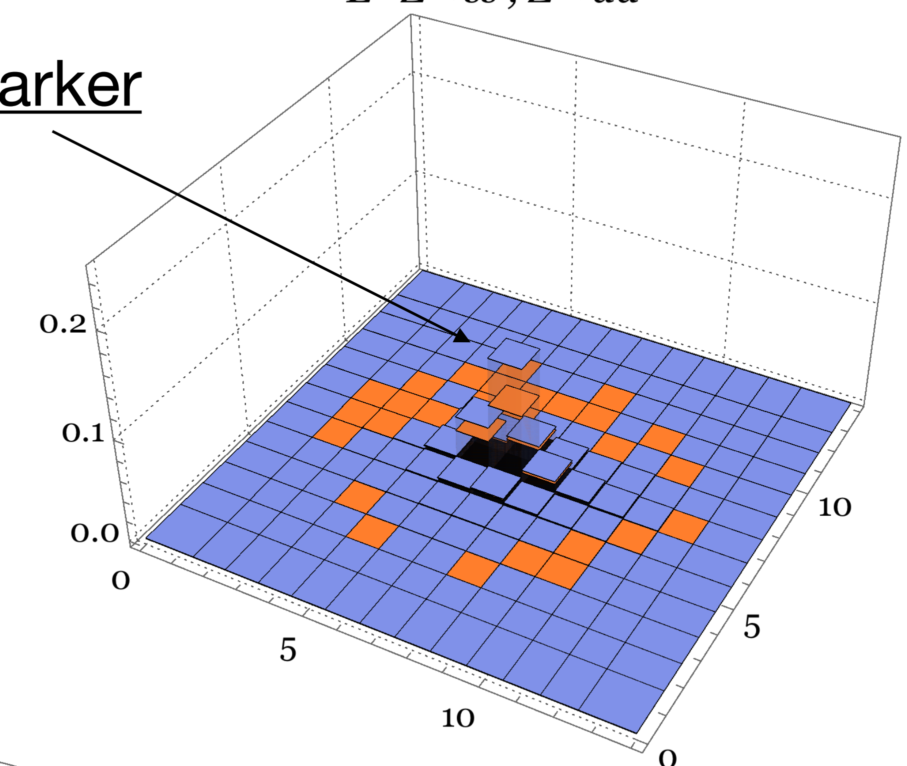


Average Images

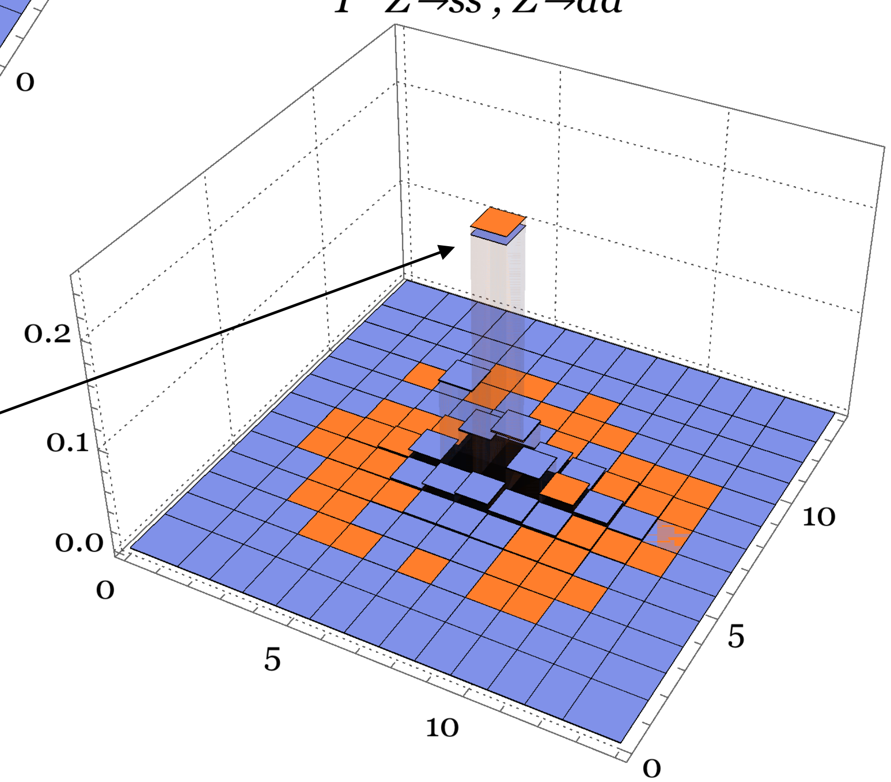
Strange jet (average) image is brighter in HN and darker in ECAL.



Strange is darker



Strange is brighter



No difference

$Z \to s\bar{s}$ $(p_T > 20\ \mathrm{GeV})$

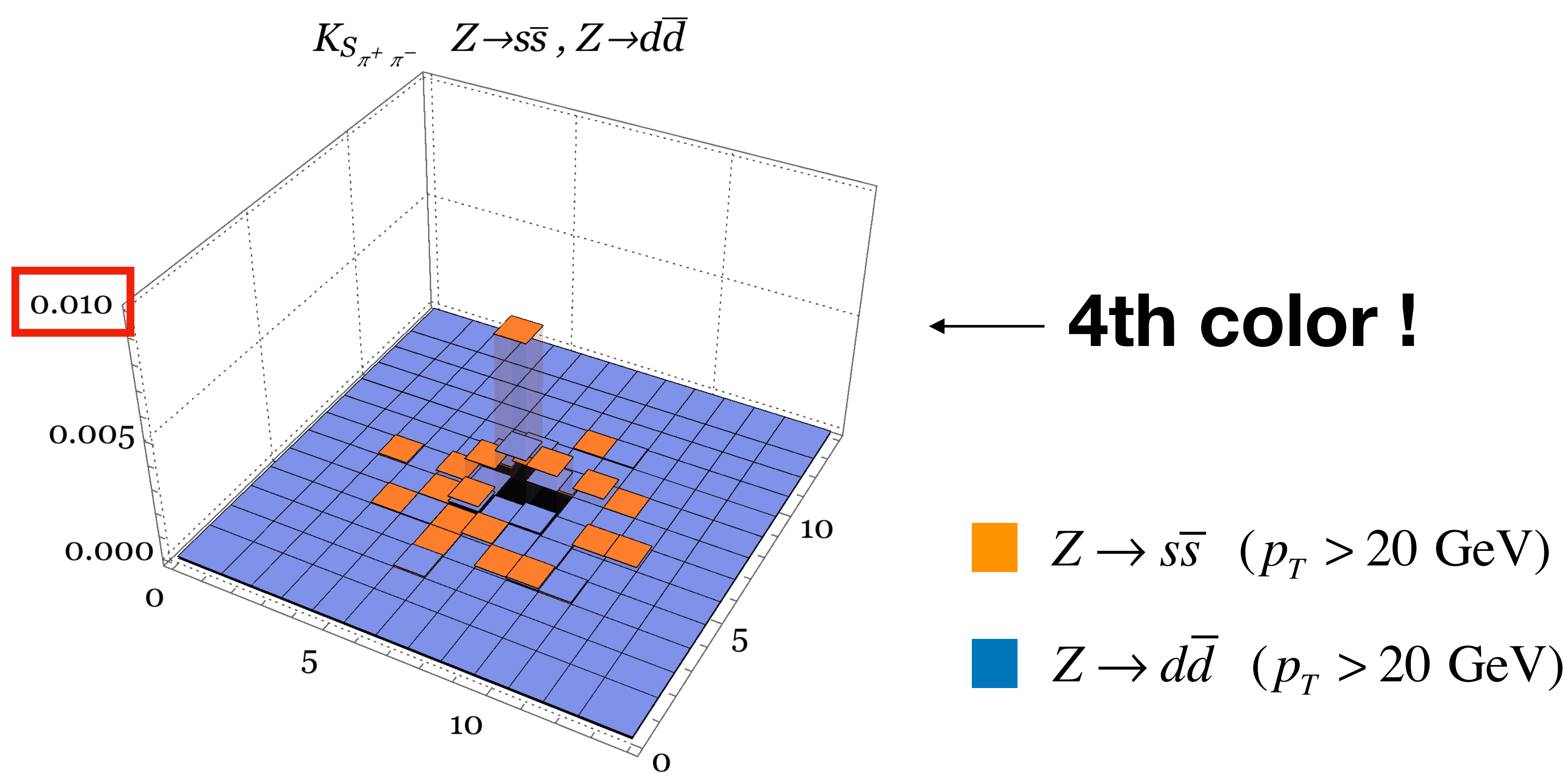
$Z \to d\bar{d}$ $(p_T > 20\ \mathrm{GeV})$

Average Images

We add **the fourth color of the reconstructable K_S p_T**.

The intensity is normalized by the sum of the track p_T , ECAL and HN in the whole image.

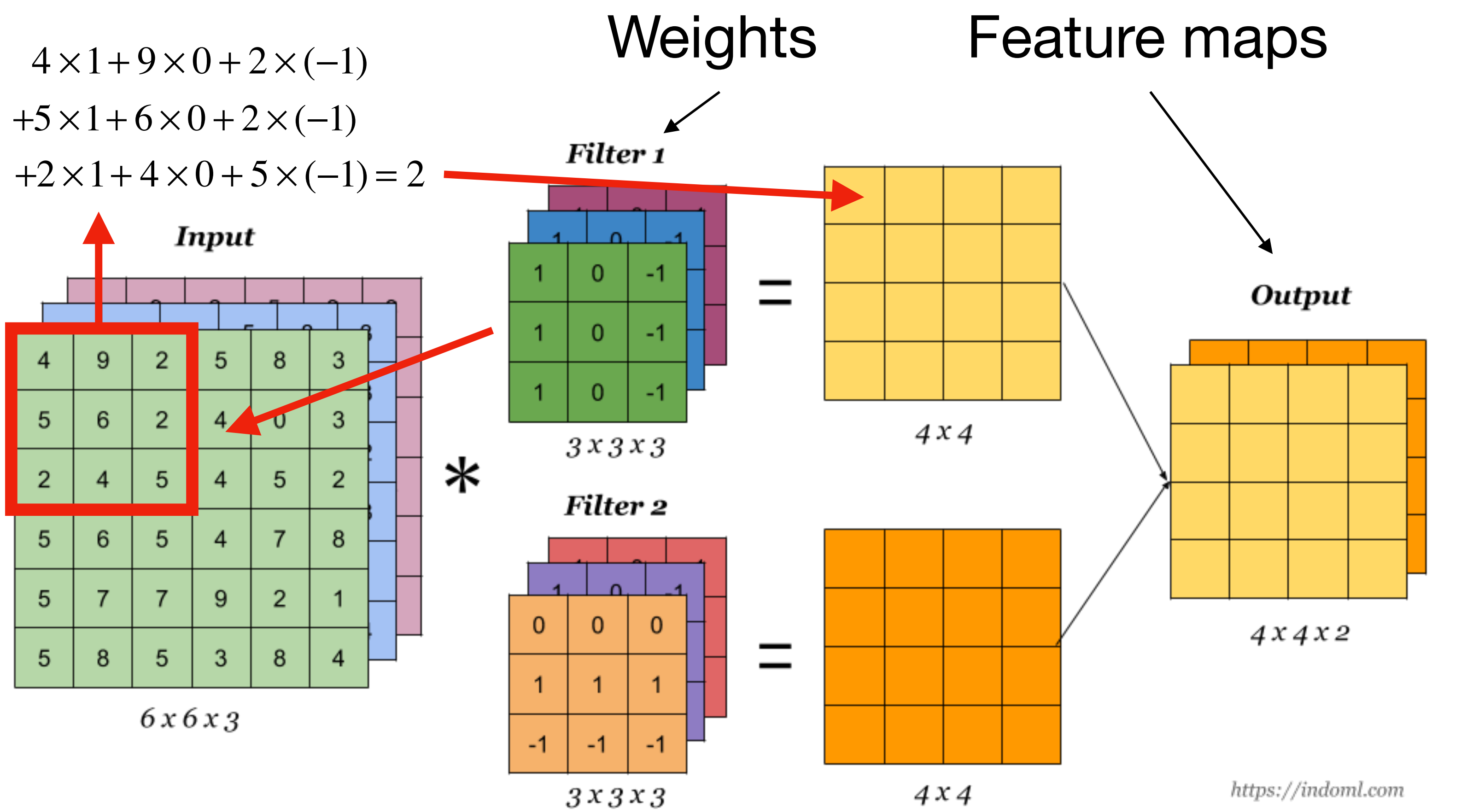
Strange jets contain more energetic K-shorts than down jets.



★ The intensity is much small compared to the other colors because the number of images including the reconstructable K_S is less than 8% (5%) of the total number of images for strange (down) jets.

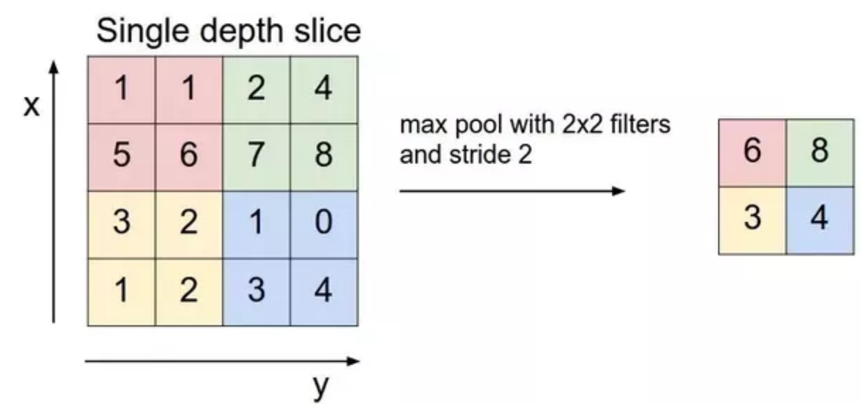
Network Architecture

Convolutional layer

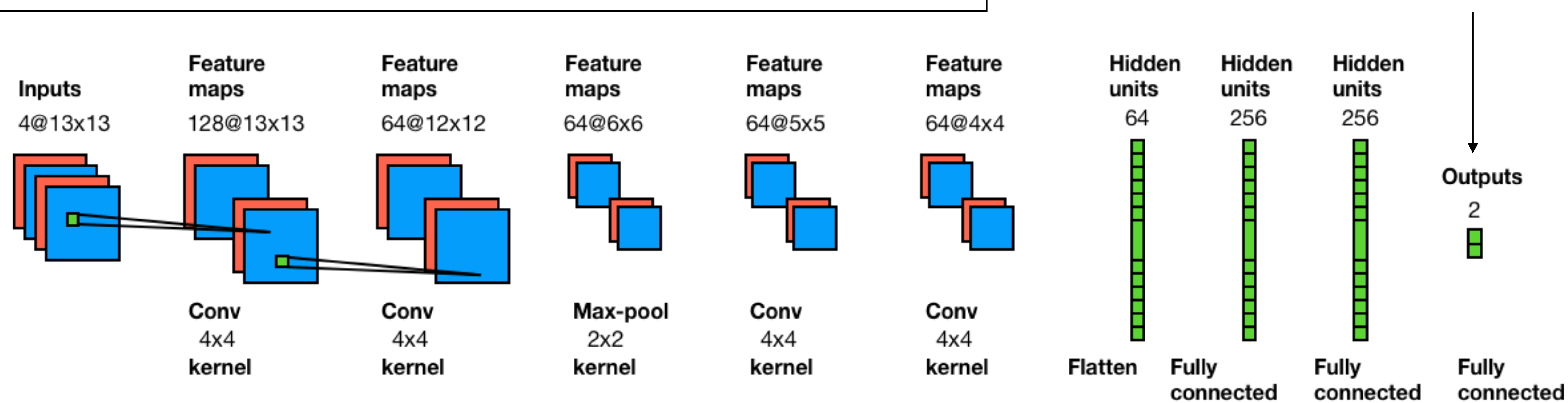


Max pooling

Reduce the image size



Probabilities of signal and background



Training

The goal of training is to minimize loss function :

Network prediction

$$L = \sum_i f(p(\theta, x_i), y_i)$$

$p(\theta, x_i)$: Signal probability

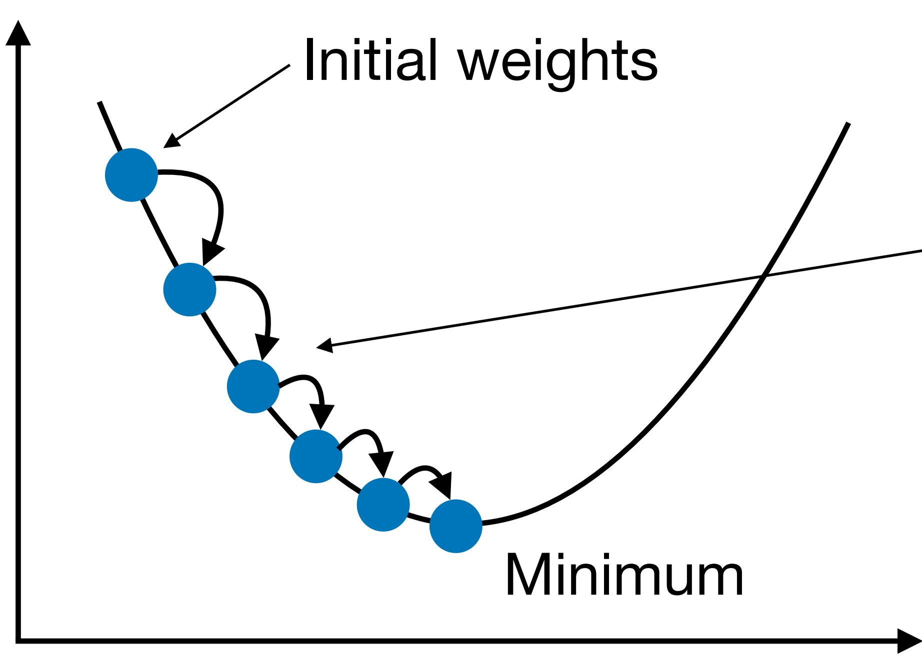
θ : Weights

x_i : Input

y_i : Truth label of example i
$\left(\begin{array}{l} y_i = 0 \text{ : Signal} \\ y_i = 1 \text{ : Background} \end{array} \right)$

We use cross entropy : $f(p, y) = -(y \log(1-p) + (1-y) \log p)$

Loss function L



Weights θ

Weights are updated according to derivative of loss function :

$$\Delta\theta = -\eta \nabla L$$

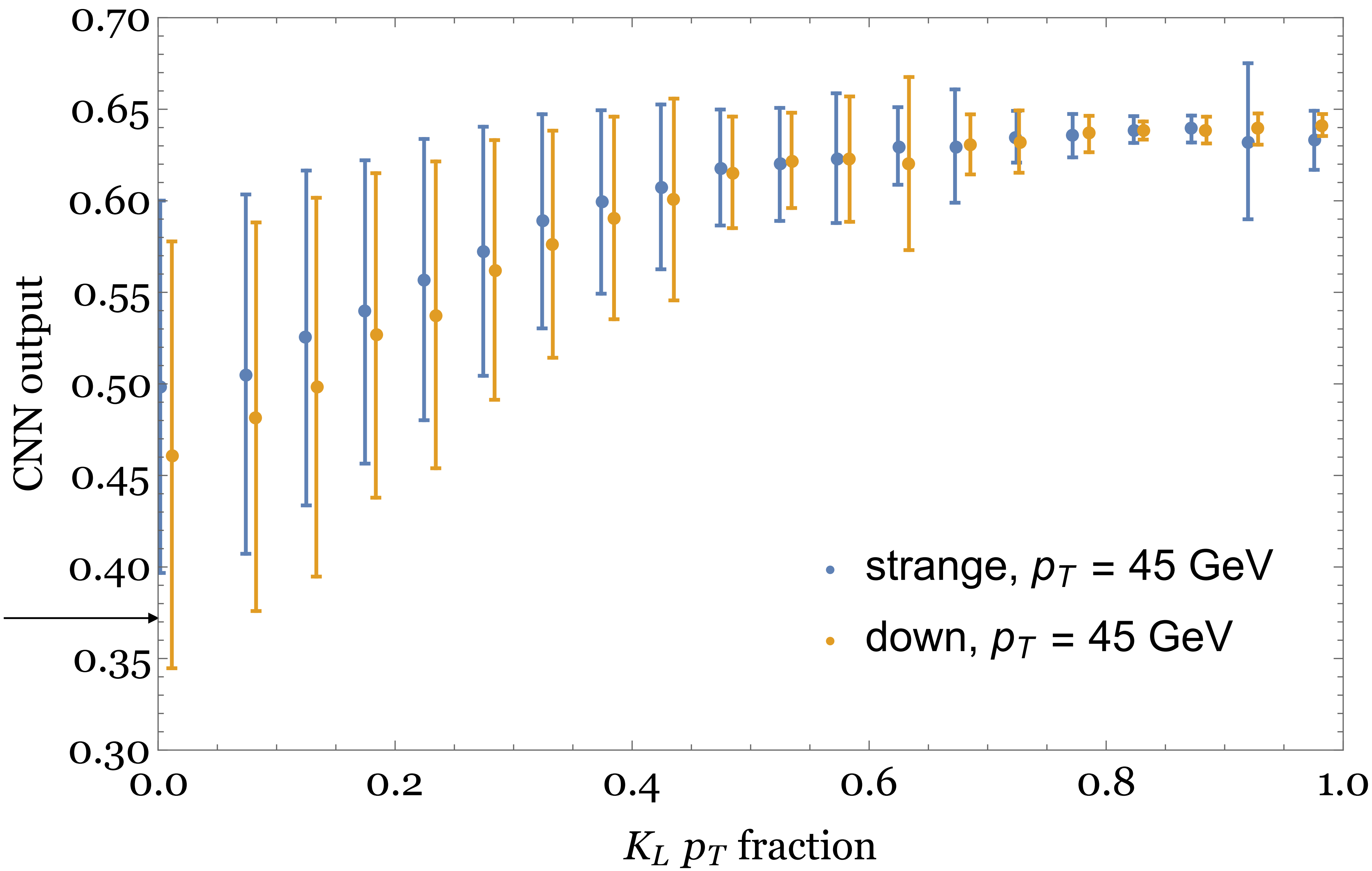
Learning rate

Neural Network Output

The correlation between the K_L pT fraction of input images and CNN (with 3 colors of tracker, HN and ECAL) outputs.

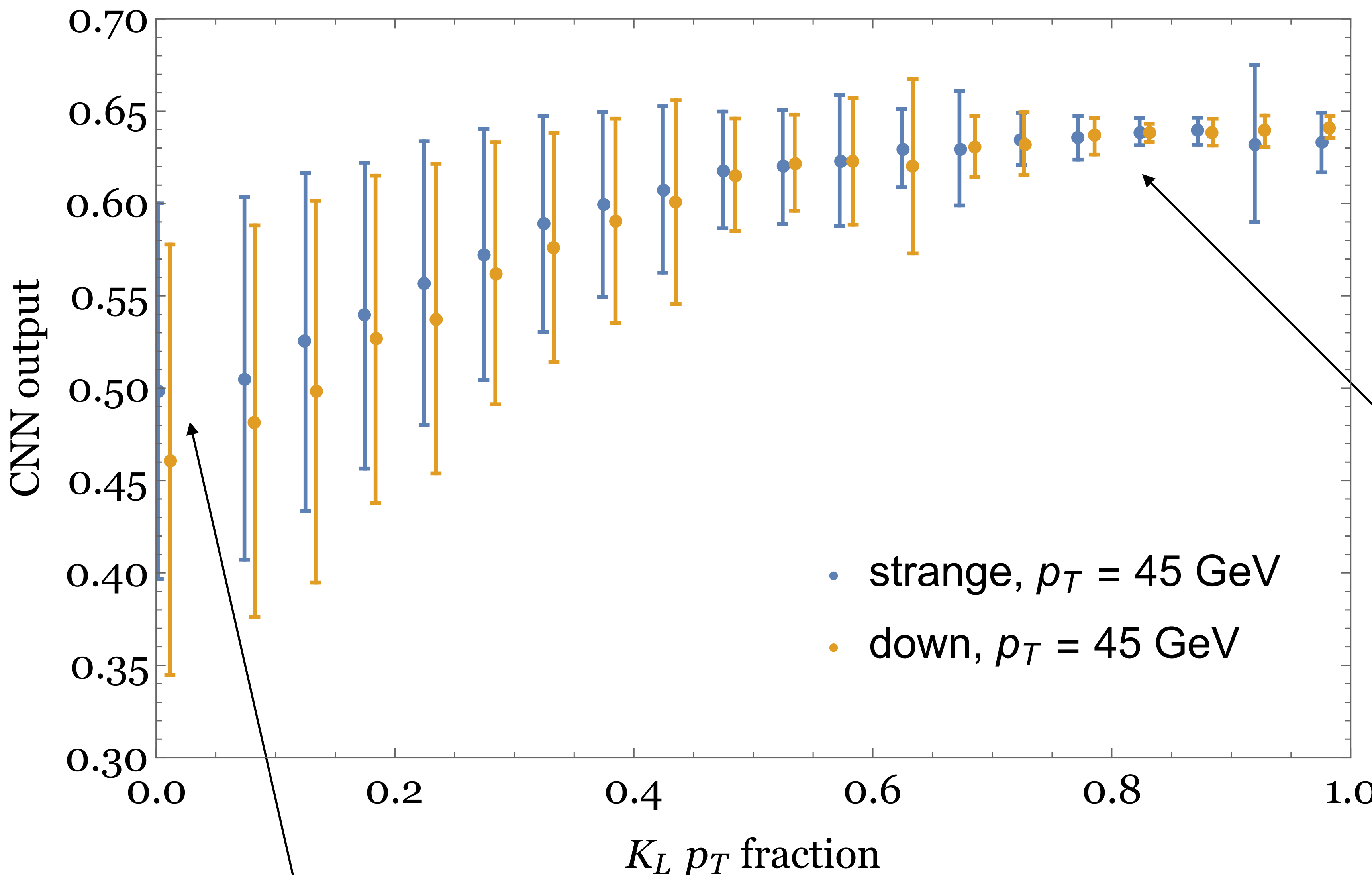
Take the average in each bin.

Standard deviation



Neural Network Output

A clear correlation: **Signal probability increases as K-long pT ratio increases**

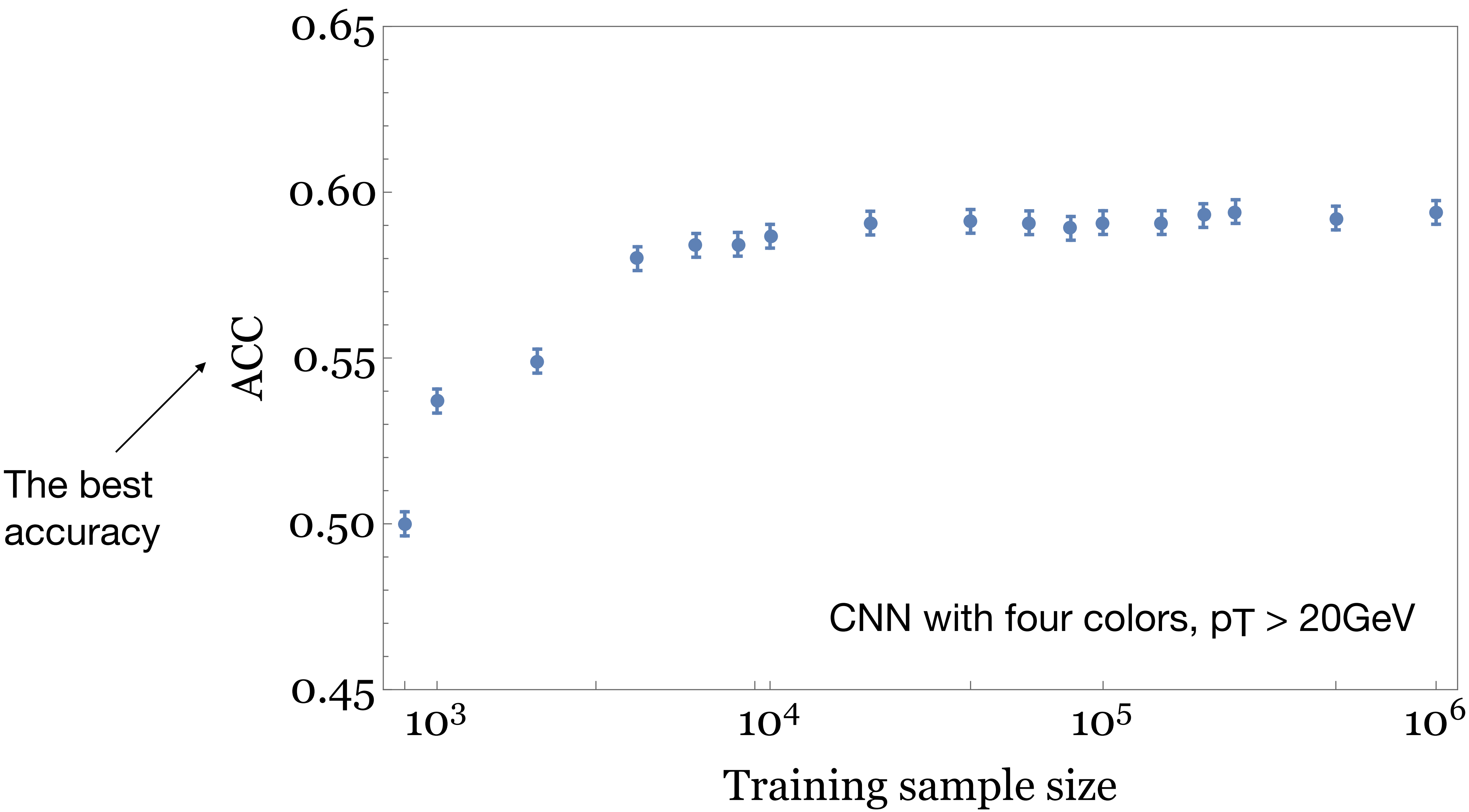


Standard deviation decreases.

The signal probability of strange jets is larger than that of down jets due to the K_S component.

Training Curve

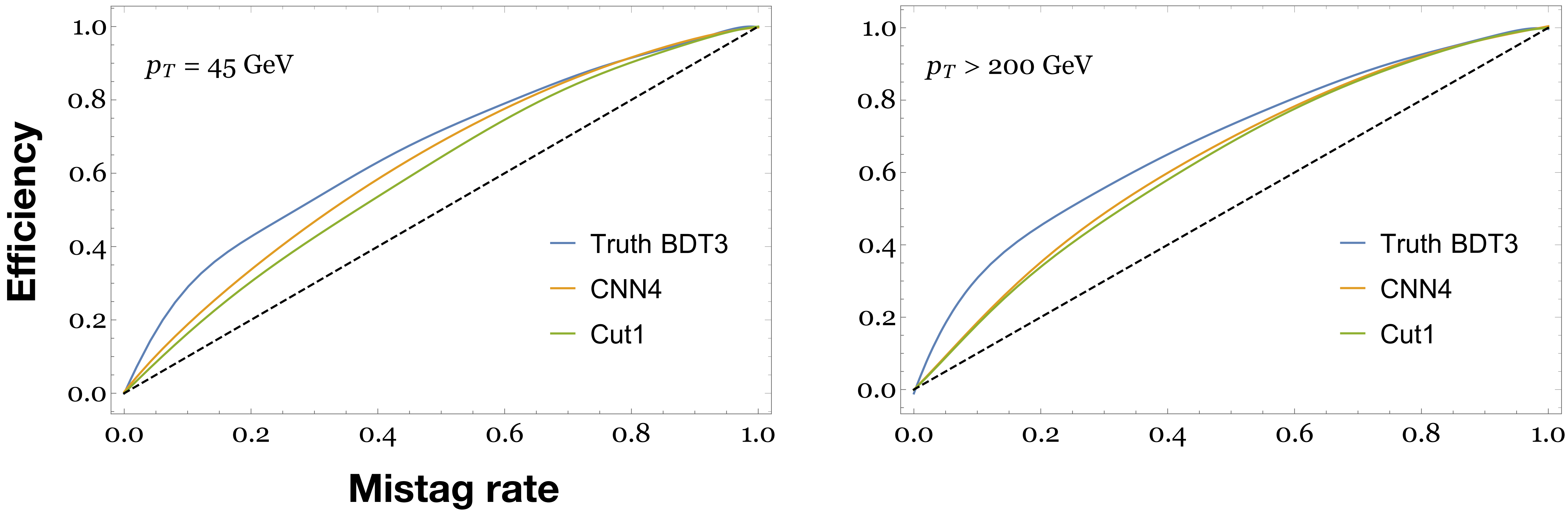
How the performance of the CNN is affected by the number of training samples.



The performance saturates immediately for more than 10000 training samples.

Results

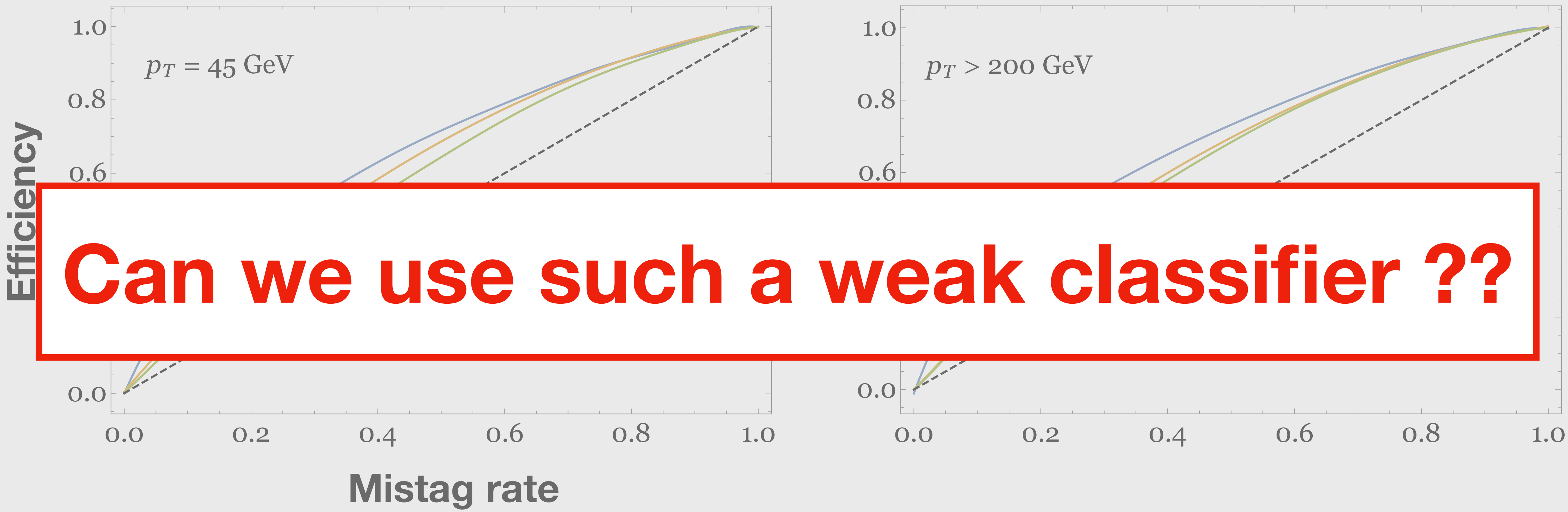
ROC curves :



The performance is worse than other classification problems.
(quark vs. gluon, …)

Results

ROC curves :



The performance is worse than other classification problems.
(quark vs. gluon, ...)

Significance Improvement

Consider a binary classifier with efficiency **ε_S** and mistag rate **ε_B.**

Before a cut on the classifier…

Statistical significance of the signal : $S/\sqrt{B}$ (S « B)

After a cut on the classifier…

If we throw away the events that fail the cut…

Statistical significance of the signal : $$q = \boxed{\frac{\epsilon_S}{\sqrt{\epsilon_B}}}\frac{S}{\sqrt{B}}$$

Significance improvement factor

If a weak classifier gives <u>a significance improvement factor smaller than 1</u>, the classifier reduces our significance ??

Significance Improvement

If we view the classifier as defining two categories (pass vs. fail)…

Combined significance of two categories :

$$q = \sqrt{\left(\frac{\epsilon_S}{\sqrt{\epsilon_B}}\frac{S}{\sqrt{B}}\right)^2 + \left(\frac{(1-\epsilon_S)}{\sqrt{(1-\epsilon_B)}}\frac{S}{\sqrt{B}}\right)^2}$$

$$= \sqrt{1 + \frac{(\epsilon_S - \epsilon_B)^2}{\epsilon_B(1-\epsilon_B)}}\frac{S}{\sqrt{B}}$$

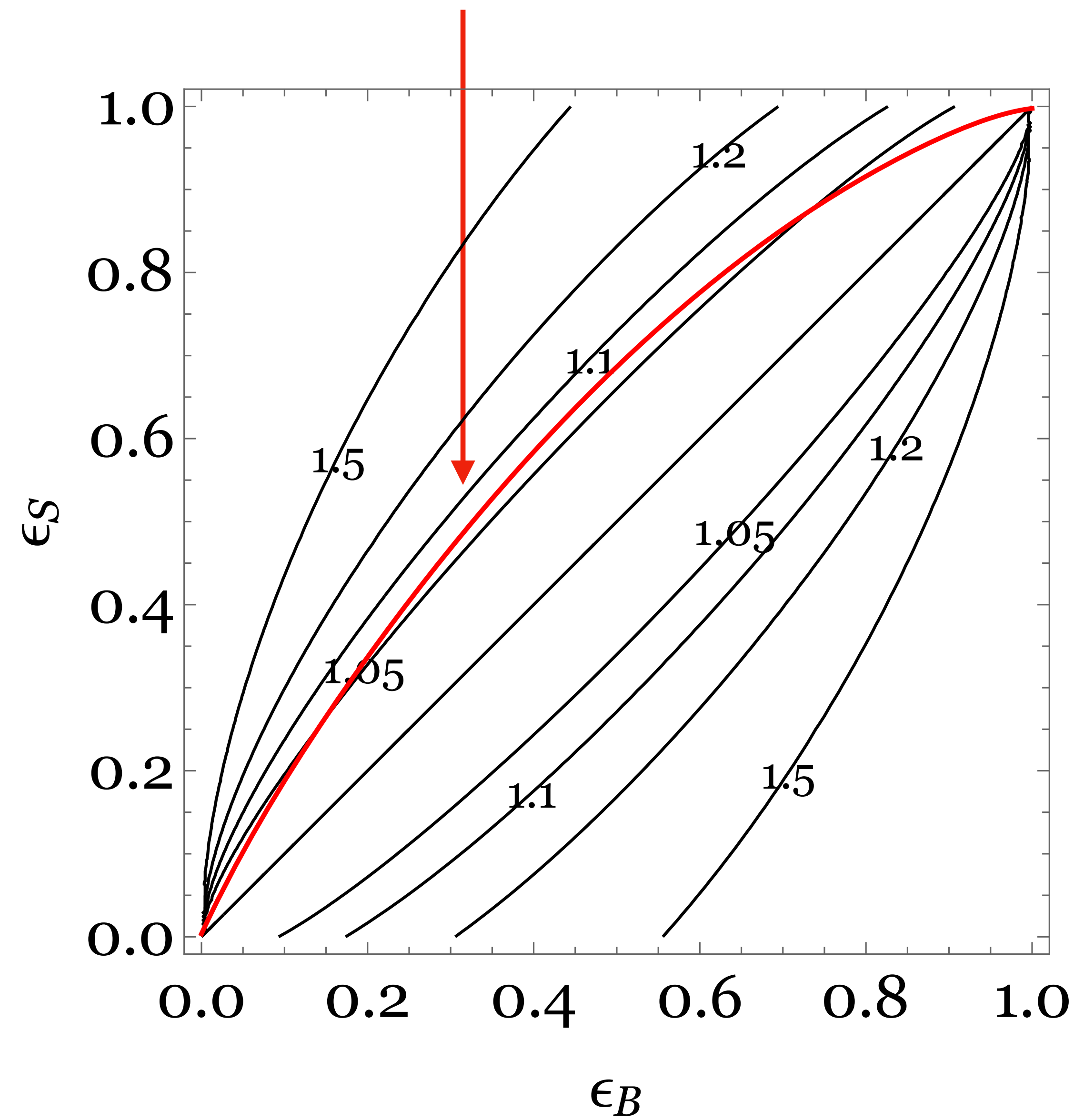
The significance can only increase.

Our best strange jet tagger (CNN)

Significance improvement is <u>only 5-10%</u>.

The importance of our strange tagger is limited…

Our best strange jet tagger (CNN)



CKM Mixings

A better use for the strange tagger is to measure the ratio of strange to down jets in some setting.

In this case…

Any amount of discrimination power will make the measurement possible with enough data.

Remember…

The values for $|V_{cs}|$ and $|V_{cd}|$ are not measured very well.

- ✓ Because the charm quark mass is too heavy to be considered light but not heavy enough to treat in the heavy quark limit.
- ✓ W boson decay $W \to cs$ gives the most direct measurement if strange tagging is possible.

Let's consider the measurement of ratio : $\dfrac{|V_{cs}|^2}{|V_{cd}|^2}$

CKM Mixings

A simple estimate

\# of data

A fraction of strange to down

\# of events passing the cut : $N_{pass}(f_S) = N(f_S\epsilon_S + (1-f_S)\epsilon_B)$

\# of events failing the cut : $N_{fail}(f_S) = N(f_S(1-\epsilon_S) + (1-f_S)(1-\epsilon_B))$

$$\chi^2 = \frac{(N_{pass}(f_S) - N_{pass}(\hat{f}_S))^2}{N_{pass}(\hat{f}_S)} + \frac{(N_{fail}(f_S) - N_{fail}(\hat{f}_S))^2}{N_{fail}(\hat{f}_S)}$$

$$= \frac{N(\epsilon_B - \epsilon_S)^2}{f_{eff}(1-f_{eff})}(f_S - \hat{f}_S)^2$$

$$f_{eff} = \hat{f}_S\epsilon_S + (1-\hat{f}_S)\epsilon_B$$

True fraction

$$\boxed{f_S = \hat{f}_S \pm \delta f_S \qquad \delta f_S = \frac{1}{|\epsilon_B - \epsilon_S|}\sqrt{\frac{f_{eff}(1-f_{eff})}{N}}}$$

As long as $\varepsilon_B \neq \varepsilon_S$, a sufficiently large N gives an accurate measurement.

CKM Mixings

W boson decay

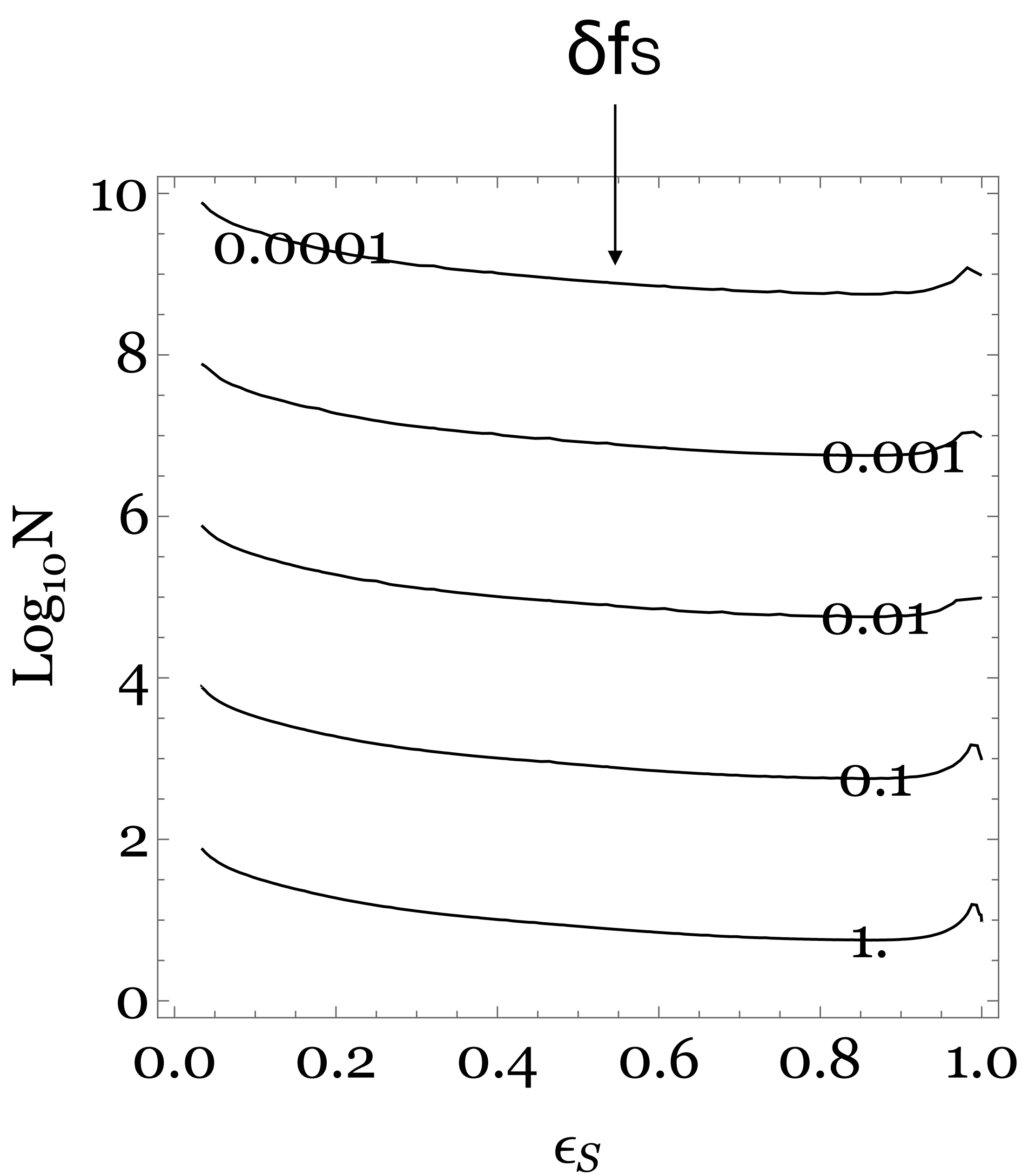
$$\Gamma(W^- \to s\bar{c}) = \Gamma(W^- \to e^-\bar{\nu}) \times 3|V_{cs}|^2$$
$$\Gamma(W^- \to d\bar{c}) = \Gamma(W^- \to e^-\bar{\nu}) \times 3|V_{cd}|^2$$

$$f_S = \hat{f}_S \pm \delta f_S$$

$$\delta f_S = \frac{1}{|\epsilon_B - \epsilon_S|}\sqrt{\frac{f_{eff}(1-f_{eff})}{N}}$$

$$f_{eff} = \hat{f}_S \epsilon_S + (1-\hat{f}_S)\epsilon_B$$

$$\boxed{\hat{f}_s = \frac{\frac{|V_{cs}|^2}{|V_{cd}|^2}}{1+\frac{|V_{cs}|^2}{|V_{cd}|^2}}} \qquad \frac{|V_{cs}|^2}{|V_{cd}|^2} \sim 20$$



Our best strange jet tagger (CNN)

Since the LHC generates a lot of W bosons, a precise measurement is possible !

Summary

✓ Strange tagging is the last missing piece of quark/gluon tagging.

✓ Neutral Kaons can be used for strange tagging.

✓ We create jet images with colors (**Tracker**, **Hadronic Neutral**, **ECAL**, **Ks pT**).
(= HCAL - Tracker)

✓ Average images of strange jets can be distinguished from down images.

✓ Convolutional Neural Network outperforms cut-based tagger.

✓ Strange jet tagger may be important for a measurement of CKM mixings.

Thank you.

Backup Material

Various taggers

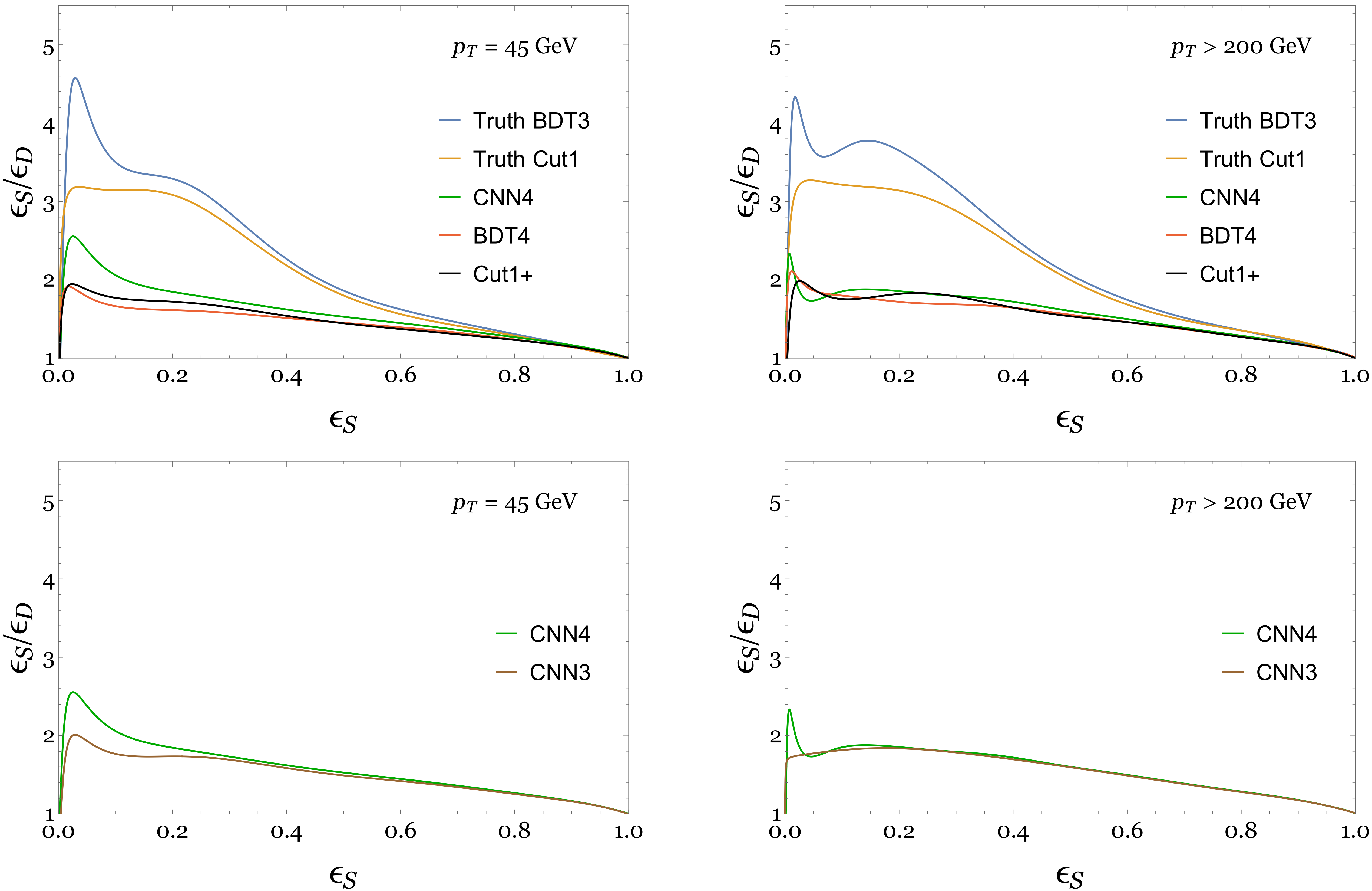
Algorithm	Input Source	Input Variable(s)	
Truth Cut1	Pythia 8	$-\pi^0 + K_L + K_S + K_{S_{\pi^+\pi^-}}$	
Truth BDT3	Pythia 8	$\pi^0\,,\ K_L\,,\ K_S + K_{S_{\pi^+\pi^-}}$	
Cut1	Delphes	$H_N - E$	
Cut1+	Delphes	$H_N - E + K_{S_{\pi^+\pi^-}}$	
BDT3	Delphes	$H_N\,,\ E\,,\ T$	
BDT4	Delphes	$H_N\,,\ E\,,\ T\,,\ K_{S_{\pi^+\pi^-}}$	
CNN3	Delphes	$H_N\,,\ E\,,\ T$	13×13 Jet Image
CNN4	Delphes	$H_N\,,\ E\,,\ T\,,\ K_{S_{\pi^+\pi^-}}$	13×13 Jet Image

Results

	AUC	ACC	R10	R50
Truth Cut1	0.65 (0.68)	0.61 (0.62)	31.9 (32.3)	3.6 (3.9)
Truth BDT3	0.66 (0.68)	0.62 (0.63)	36.4 (37.0)	3.7 (4.2)
Cut1	0.60 (0.63)	0.57 (0.59)	16.8 (17.9)	2.7 (3.0)
Cut1+	0.62 (0.63)	0.58 (0.60)	17.8 (17.8)	2.9 (3.1)
BDT3	0.61 (0.63)	0.58 (0.60)	16.0 (18.0)	2.8 (3.1)
BDT4	0.61 (0.63)	0.59 (0.60)	17.0 (18.0)	2.9 (3.1)
CNN3	0.62 (0.63)	0.59 (0.60)	17.7 (18.2)	3.0 (3.2)
CNN4	0.63 (0.64)	0.59 (0.60)	20.9 (18.3)	3.1 (3.2)

$\mathrm{R10} = 1/\epsilon_D$ for $\epsilon_S = 0.1$ $\quad$ $\mathrm{R50} = 1/\epsilon_D$ for $\epsilon_S = 0.5$

Results


p_T = 45 GeV
Truth BDT3
Truth Cut1
CNN4
BDT4
Cut1+
ϵ_S/ϵ_D
ϵ_S
1
2
3
4
5
0.0
0.2
0.4
0.6
0.8
1.0
p_T > 200 GeV
Truth BDT3
Truth Cut1
CNN4
BDT4
Cut1+
ϵ_S/ϵ_D
ϵ_S
p_T = 45 GeV
CNN4
CNN3
ϵ_S/ϵ_D
ϵ_S
p_T > 200 GeV
CNN4
CNN3
ϵ_S/ϵ_D
ϵ_S