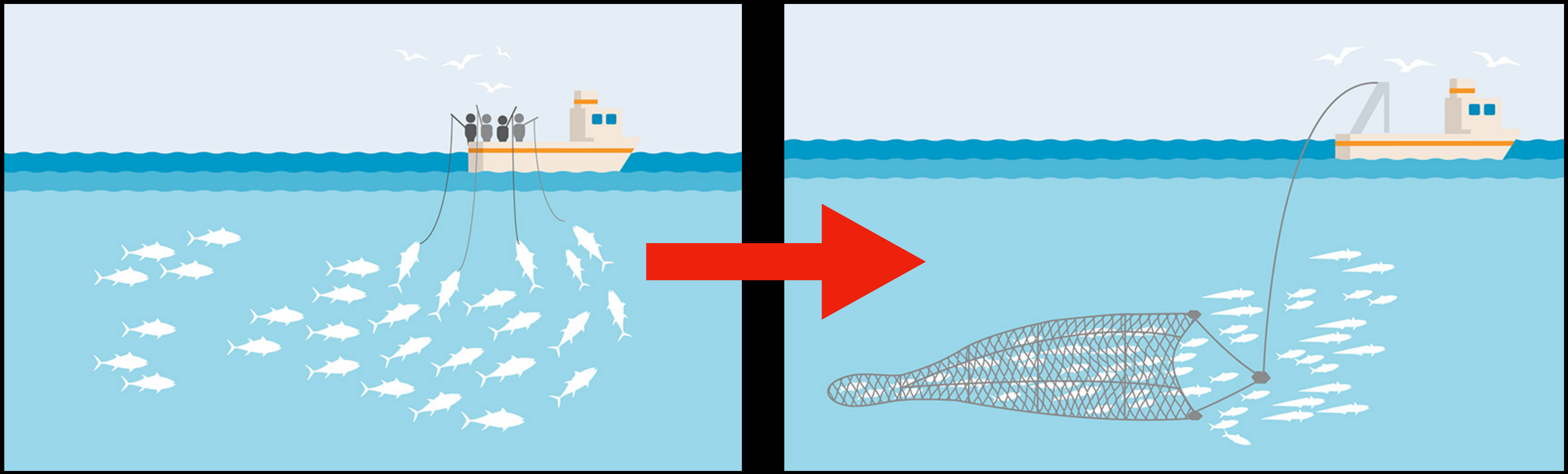
# A global analysis of Higgs branching fraction measurement



Gang Li

CEPC Detector plenary, March 17, 2021

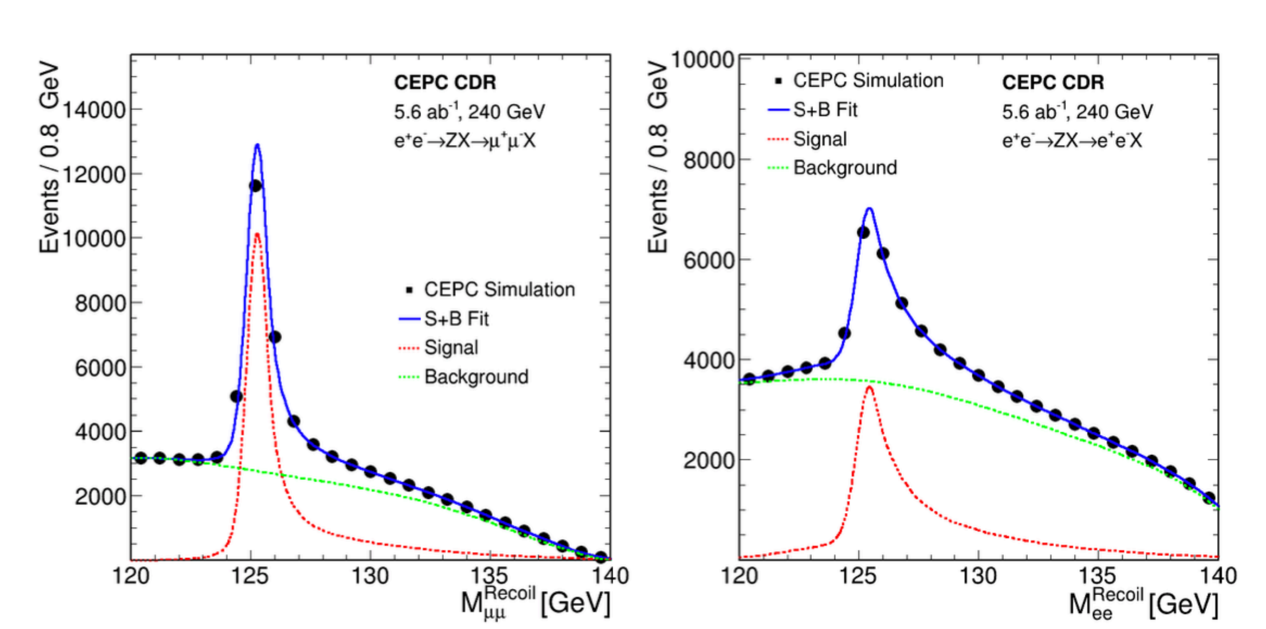
# Outline

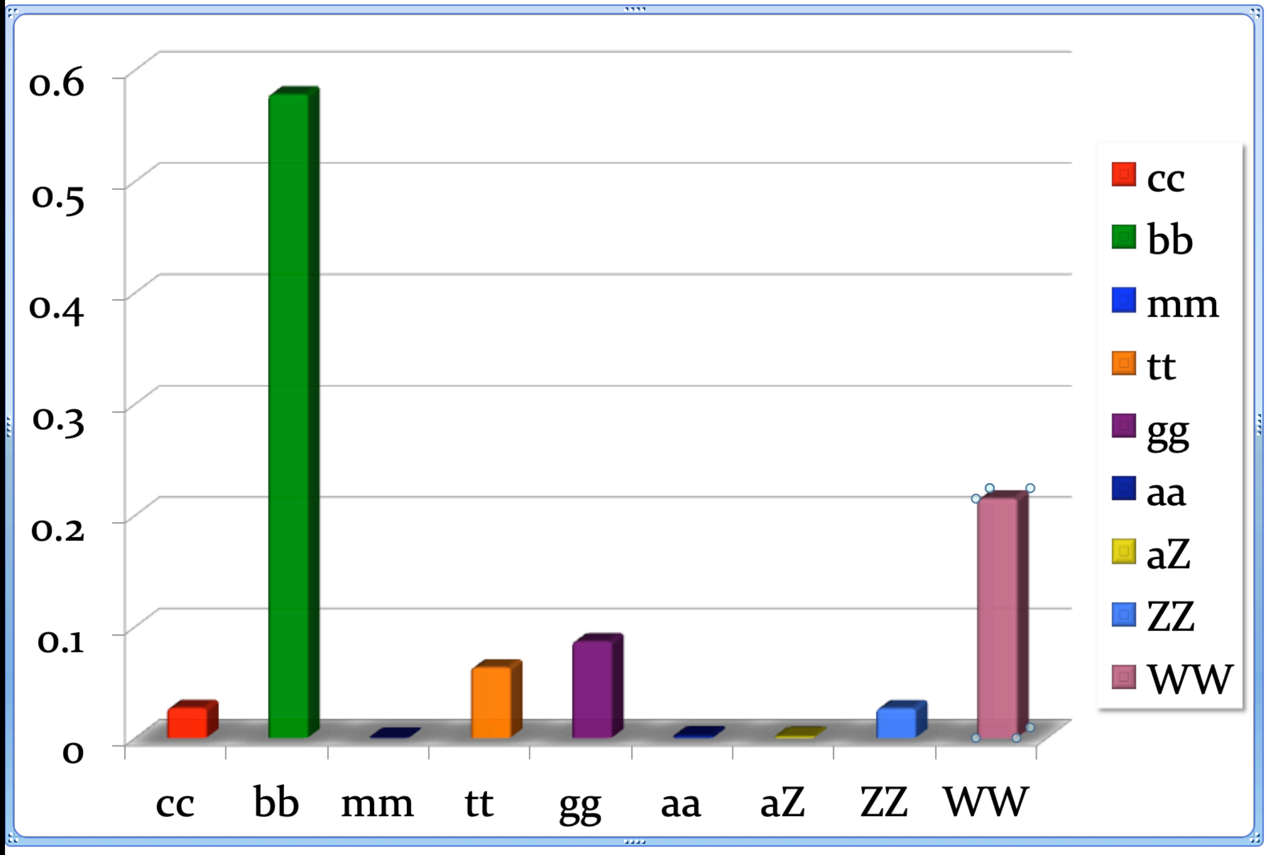
- Introduction
- Methods
- Some preliminary results
- Summary and plan

# Introduction

## A global analysis approach, talk at PKU workshop, 2019

- CEPC Higgs Br measurement
  - 9 Higgs decays accessible at CEPC
  - Typical S/N ratio ~ 1/O(1)
  - Brs very small ( $\mu\mu, \gamma\gamma, \gamma Z$) or large (bb,WW)
  - Contaminations among them is an issue
    - cc/bb/gg/WW/ZZ/$\tau\tau$ hadronic FS




| Decay mode | $\sigma(ZH) \times$ BR | BR |
|---|---|---|
| $H \to b\bar{b}$ | 0.27% | 0.56% |
| $H \to c\bar{c}$ | 3.3% | 3.3% |
| $H \to gg$ | 1.3% | 1.4% |
| $H \to WW^*$ | 1.0% | 1.1% |
| $H \to ZZ^*$ | 5.1% | 5.1% |
| $H \to \gamma\gamma$ | 6.8% | 6.9% |
| $H \to Z\gamma$ | 15% | 15% |
| $H \to \tau^+\tau^-$ | 0.8% | 1.0% |
| $H \to \mu^+\mu^-$ | 17% | 17% |
| $H \to$ inv | – | $< 0.30\%$ |

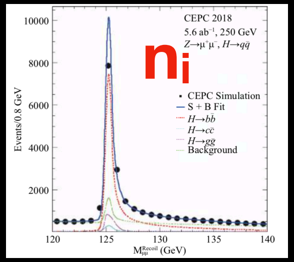
# Global analysis approach and advantages of e+e- collider

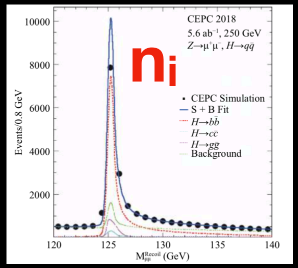
- Absolute/model-independent
- Two types of backgrounds
  - Higgs bkgd (crosstalk, characterized by a confusion matrix )
  - non-Higgs bkgd ( enlarge the stat. unc. of $n_i$ )
- Measuring all Br’s simultaneously
  - **Global confusion has more information**
  - **Multinomial distribution: smaller stat. uncertainty**
  - **Global constraint: improve precision**

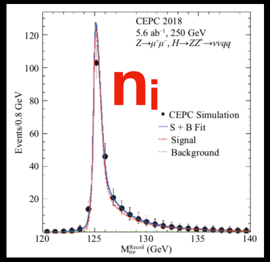
$\sigma_N = \sqrt{N \times p}$ **for Poisson**

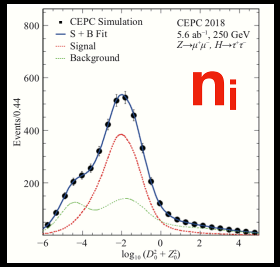
$\sigma_N = \sqrt{N \times p \times (1-p)}$ **for multinomial**

**$B_i$ =**

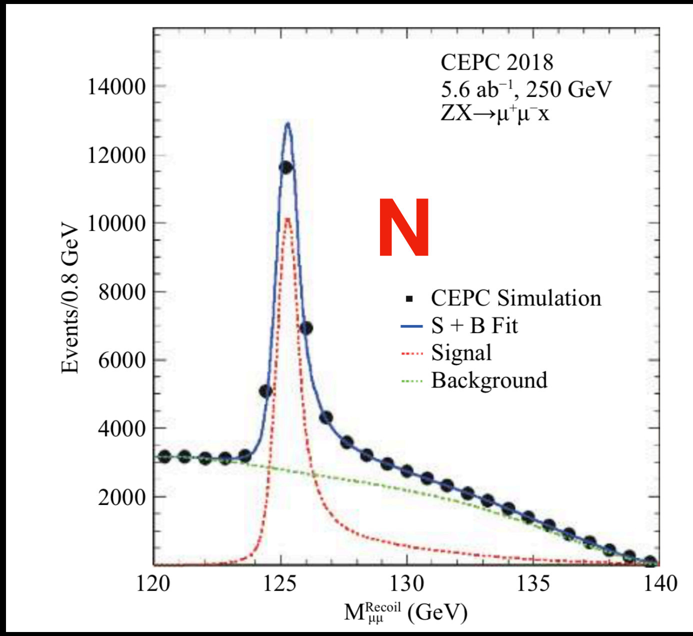



**+ … …**




**non-Higgs background**

**— subtracted with fitting for other method**

**— worsen $\sigma_{n_i}$’s**

Take the simplest case as an example
—2 decays only

# Efficiency matrix

**From MC, no dependence on Br's**

$$\begin{pmatrix} n_1 \\ n_2 \end{pmatrix} = \begin{pmatrix} \epsilon_{11} & \epsilon_{12} \\ \epsilon_{21} & \epsilon_{22} \end{pmatrix} \times \begin{pmatrix} N_1 \\ N_2 \end{pmatrix}$$

Observation

MODULATION Matrix

Production

A produced final state j reconstructed as final state i

**Measurement:**

**Solve N and minimize its uncertainty**

Demodulation

$$\boxed{N = E^{-1} n}$$

# N→B : error propagation shows very simple features

$\Sigma_B$: **Variance of Branching fractions**

$$\boxed{\Sigma_B \propto \frac{1}{N^4 |E|^2}}$$

- Variance of B's proportional to $1/(N^4|E|^2)$
- $N^4$ : statistical power
- $|E|^2$ : the performance of Detector x Reconstruction x Analysis
- 2 Br's  have same uncertainties

# Same for more than 2 decays

$$\Sigma_B = \frac{1}{N^4\,|E|^2}\begin{pmatrix} \left\{\sum_j [N_1 \frac{\partial N}{\partial n_j} - N \frac{\partial N_1}{\partial n_j}]\overrightarrow{\sigma_{n_j}}\right\}^2 & .. & .. \\ .. & \left\{\sum_j [N_2 \frac{\partial N}{\partial n_j} - N \frac{\partial N_2}{\partial n_j}]\overrightarrow{\sigma_{n_j}}\right\}^2 & .. \\ .. & .. & \left\{\sum_j [N_3 \frac{\partial N}{\partial n_j} - N \frac{\partial N_3}{\partial n_j}]\overrightarrow{\sigma_{n_j}}\right\}^2 \end{pmatrix}$$

- Binomial to multinomial distribution
- $N^4$ : statistical power
- $|E|^2$ : the performance of Detector x Reconstruction x Analysis

From the point view of detector design and optimization, a parameterization of performance as simple as possible is desired

The determinant of efficiency matrix |E| is a good candidate

$$
\begin{aligned}
P &= \underline{f(\sigma_p, \sigma_{E_\gamma}, PID, JID, JER, \ldots)} \\
&= |E|^2 \propto \frac{1}{|\Sigma_B|^2}
\end{aligned}
$$

Now problem successfully becomes how to Maximize |E|

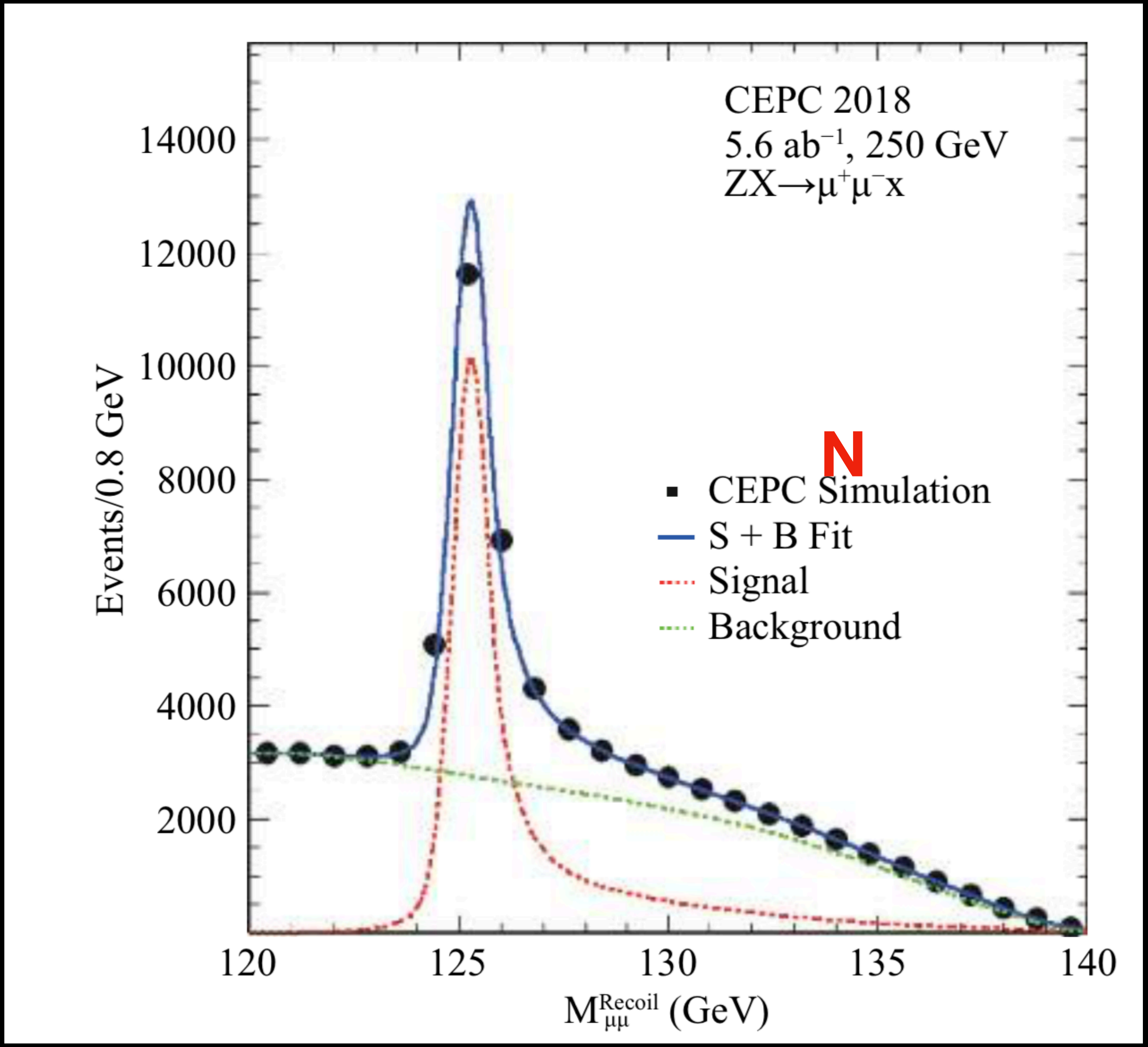
# Methods to get efficiency matrix

# On backgrounds

- Two type of backgrounds

☑ Non-uuH backgrounds: subtracted by fitting, enlarging statistical uncertainty of $n_i$

☑ uuH backgrounds (cross talks): the efficiency matrix dealing with them



# Solve B or N by minimizing the $\chi^2$

$$\chi^2 = \sum_i \frac{(\sum \epsilon_{ij} N_j - n_i)^2}{\sigma_{n_i}^2} + \frac{(\sum_l N_l - N)^2}{\sigma_N^2}$$

Higgs -> cc, bb, mm, $\tau\tau$ gg, $\gamma\gamma$, ZZ, WW, $\gamma$Z

1 2 3 4 5 6 7 8 9

$$\begin{pmatrix} n_1 \\ n_2 \\ n_3 \\ n_4 \\ n_5 \\ n_6 \\ n_7 \\ n_8 \\ n_9 \end{pmatrix} = \begin{pmatrix} \epsilon_{11} & \epsilon_{12} & \epsilon_{13} & \epsilon_{14} & \epsilon_{15} & \epsilon_{16} & \epsilon_{17} & \epsilon_{18} & \epsilon_{19} \\ \epsilon_{21} & \epsilon_{22} & \epsilon_{23} & \epsilon_{24} & \epsilon_{25} & \epsilon_{26} & \epsilon_{27} & \epsilon_{28} & \epsilon_{29} \\ \epsilon_{31} & \epsilon_{32} & \epsilon_{23} & \epsilon_{34} & \epsilon_{35} & \epsilon_{36} & \epsilon_{37} & \epsilon_{38} & \epsilon_{39} \\ \epsilon_{41} & \epsilon_{42} & \epsilon_{33} & \epsilon_{44} & \epsilon_{45} & \epsilon_{46} & \epsilon_{47} & \epsilon_{48} & \epsilon_{49} \\ \epsilon_{51} & \epsilon_{52} & \epsilon_{43} & \epsilon_{54} & \epsilon_{55} & \epsilon_{56} & \epsilon_{57} & \epsilon_{58} & \epsilon_{59} \\ \epsilon_{61} & \epsilon_{62} & \epsilon_{53} & \epsilon_{64} & \epsilon_{65} & \epsilon_{66} & \epsilon_{67} & \epsilon_{68} & \epsilon_{69} \\ \epsilon_{71} & \epsilon_{72} & \epsilon_{63} & \epsilon_{74} & \epsilon_{75} & \epsilon_{76} & \epsilon_{77} & \epsilon_{78} & \epsilon_{79} \\ \epsilon_{81} & \epsilon_{82} & \epsilon_{73} & \epsilon_{84} & \epsilon_{85} & \epsilon_{86} & \epsilon_{87} & \epsilon_{88} & \epsilon_{89} \\ \epsilon_{91} & \epsilon_{92} & \epsilon_{83} & \epsilon_{94} & \epsilon_{95} & \epsilon_{96} & \epsilon_{97} & \epsilon_{98} & \epsilon_{99} \end{pmatrix} \begin{pmatrix} N_1 \\ N_2 \\ N_3 \\ N_4 \\ N_5 \\ N_6 \\ N_7 \\ N_8 \\ N_9 \end{pmatrix}$$

$$\sum_i N_i = N^{tag} \text{ or } \sum_i B_i = 1$$

# Now we focus on the efficiency matrix

Traditionally only one row can be obtained for a individual analysis

$$
\begin{pmatrix}
\epsilon_{11} & \epsilon_{12} & \epsilon_{13} & \epsilon_{14} & \epsilon_{15} & \epsilon_{16} & \epsilon_{17} & \epsilon_{18} & \epsilon_{19} \\
\epsilon_{21} & \epsilon_{22} & \epsilon_{23} & \epsilon_{24} & \epsilon_{25} & \epsilon_{26} & \epsilon_{27} & \epsilon_{28} & \epsilon_{29} \\
\epsilon_{31} & \epsilon_{32} & \epsilon_{23} & \epsilon_{34} & \epsilon_{35} & \epsilon_{36} & \epsilon_{37} & \epsilon_{38} & \epsilon_{39} \\
\epsilon_{41} & \epsilon_{42} & \epsilon_{33} & \epsilon_{44} & \epsilon_{45} & \epsilon_{46} & \epsilon_{47} & \epsilon_{48} & \epsilon_{49} \\
\epsilon_{51} & \epsilon_{52} & \epsilon_{43} & \epsilon_{54} & \epsilon_{55} & \epsilon_{56} & \epsilon_{57} & \epsilon_{58} & \epsilon_{59} \\
\epsilon_{61} & \epsilon_{62} & \epsilon_{53} & \epsilon_{64} & \epsilon_{65} & \epsilon_{66} & \epsilon_{67} & \epsilon_{68} & \epsilon_{69} \\
\epsilon_{71} & \epsilon_{72} & \epsilon_{63} & \epsilon_{74} & \epsilon_{75} & \epsilon_{76} & \epsilon_{77} & \epsilon_{78} & \epsilon_{79} \\
\epsilon_{81} & \epsilon_{82} & \epsilon_{73} & \epsilon_{84} & \epsilon_{85} & \epsilon_{86} & \epsilon_{87} & \epsilon_{88} & \epsilon_{89} \\
\epsilon_{91} & \epsilon_{92} & \epsilon_{83} & \epsilon_{94} & \epsilon_{95} & \epsilon_{96} & \epsilon_{97} & \epsilon_{98} & \epsilon_{99}
\end{pmatrix}
$$

Now we will try some sophisticated machine learning

- EFN (energy flow network) and
- DGNN(dynamic graph convolution neural network)

# EFN approach (deep sets)

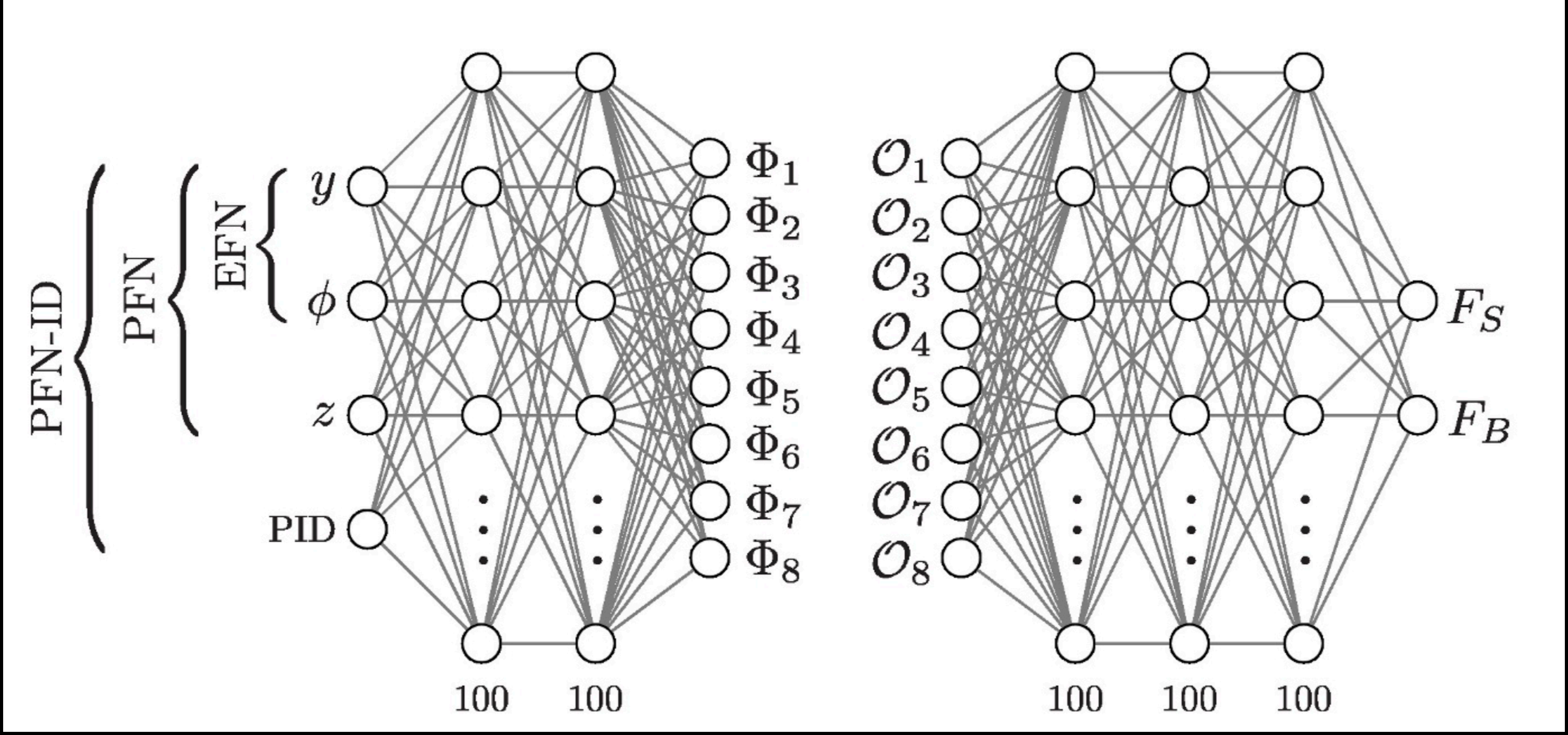
Deep sets for particle (jet): JHEP01(2019)121

- It was proposed to identify jets at LHC
- ALL particle level information as input: 4-momenta, impact parameters, and PID

- No jet-clustering, no isolation of leptons and photons
- No requirement on input size such as DNN
- No explicitly dependence on the ordering of the inputs such as RNN
- Respect the permutation symmetry
- Able to handle variable-length inputs
- Infra-red and collinear safety naturally achieved

The key mathematical fact:

A generic function of a set of objects can be decomposed to arbitrarily good approximation in a practical and intuitive way [arXiv:1703.06114]

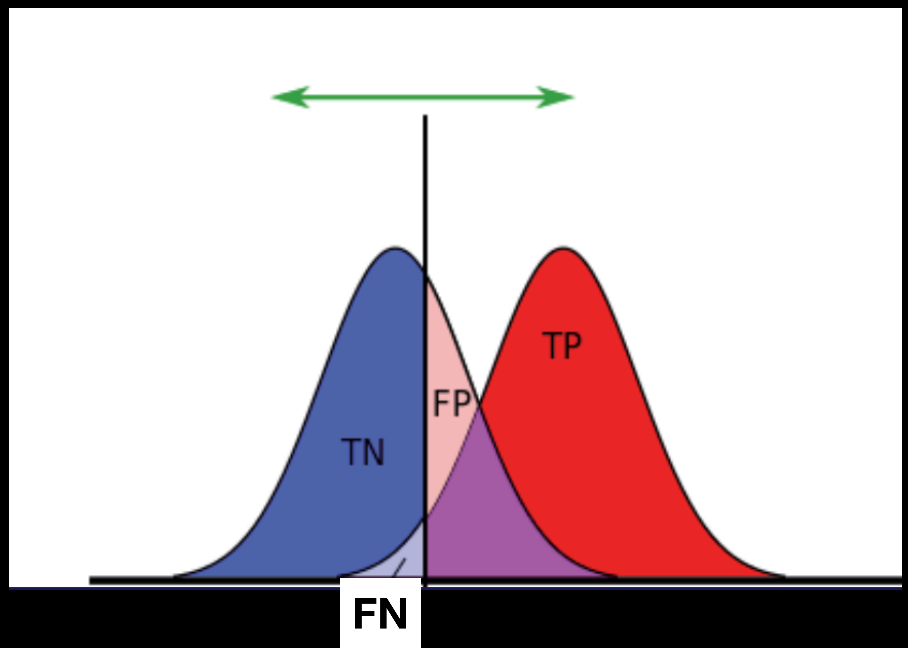


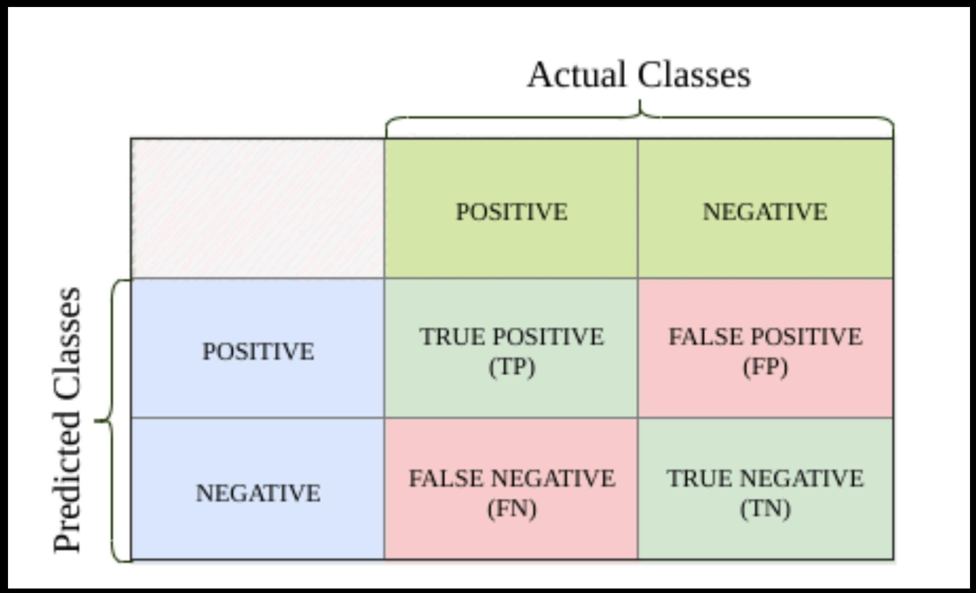

$$\mathcal{O}(\{p_1, \ldots, p_M\}) = F\left(\sum_{i=1}^{M} \Phi(p_i)\right)$$

# Some preliminary results

# Termnologies

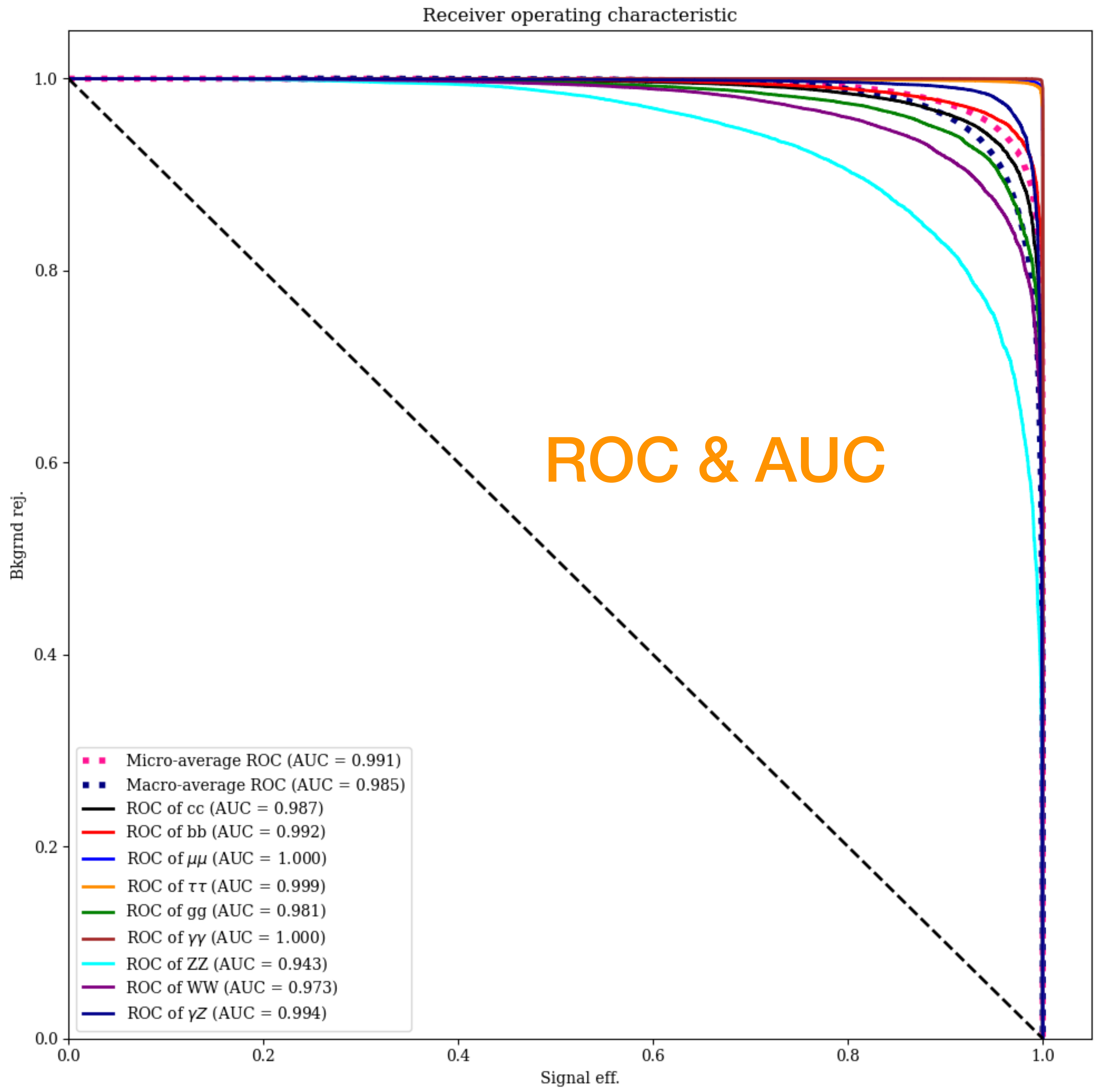
- Tagging efficiency: accuracy in ML
- ROC : Receiver Operating Characteristics Curve, mainly for binary classification,
  - In HEP it is  Rejection rate vs. Tagging efficiency  (FN rate vs. TP rate )
- AUC : Area under the ROC
- Confusion matrix
  - it is the efficiency matrix  when neglecting SM backgrounds



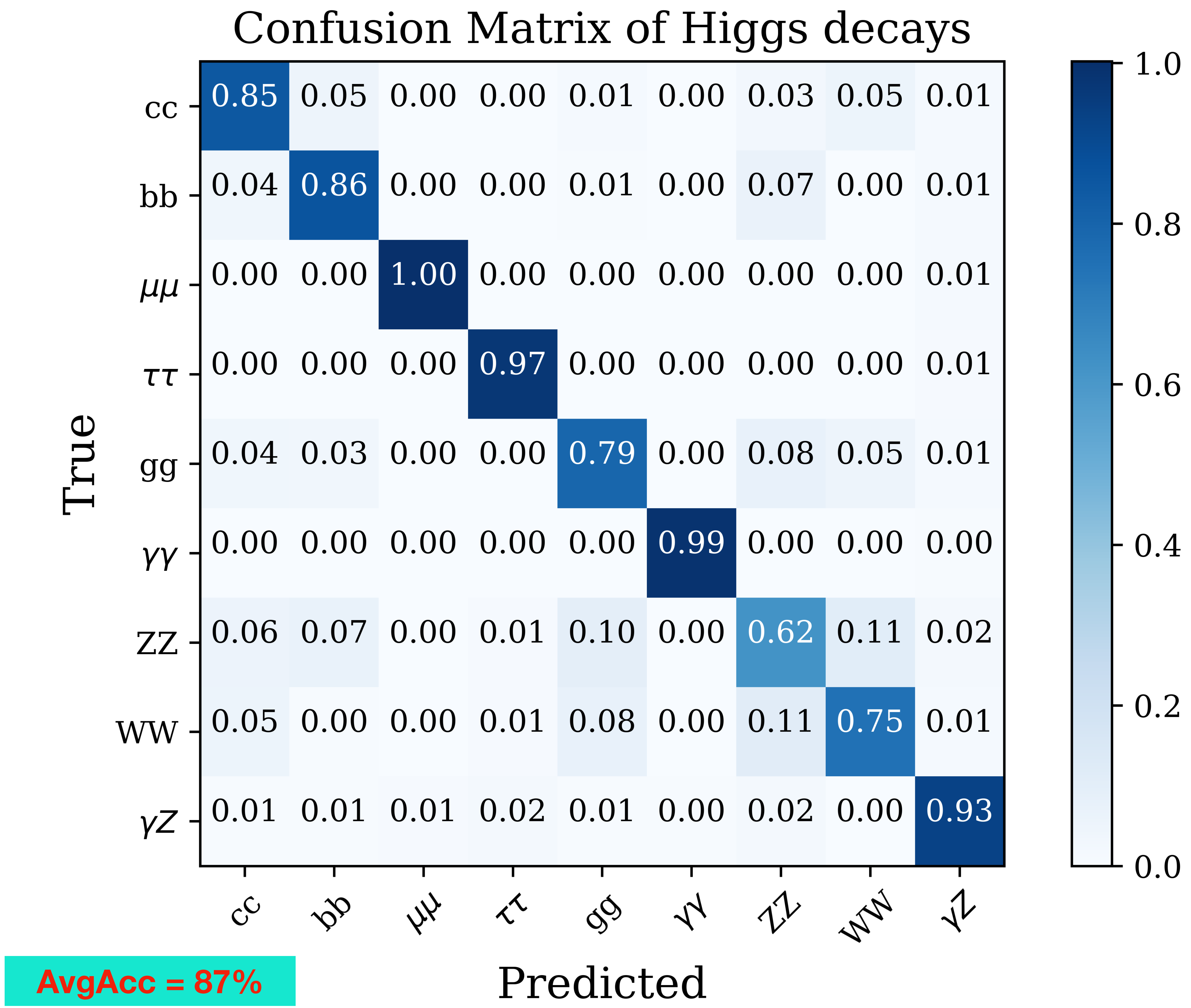



# Samples: $e^+e^- \to ZH, Z \to \mu^+\mu^-, e^+e^-$

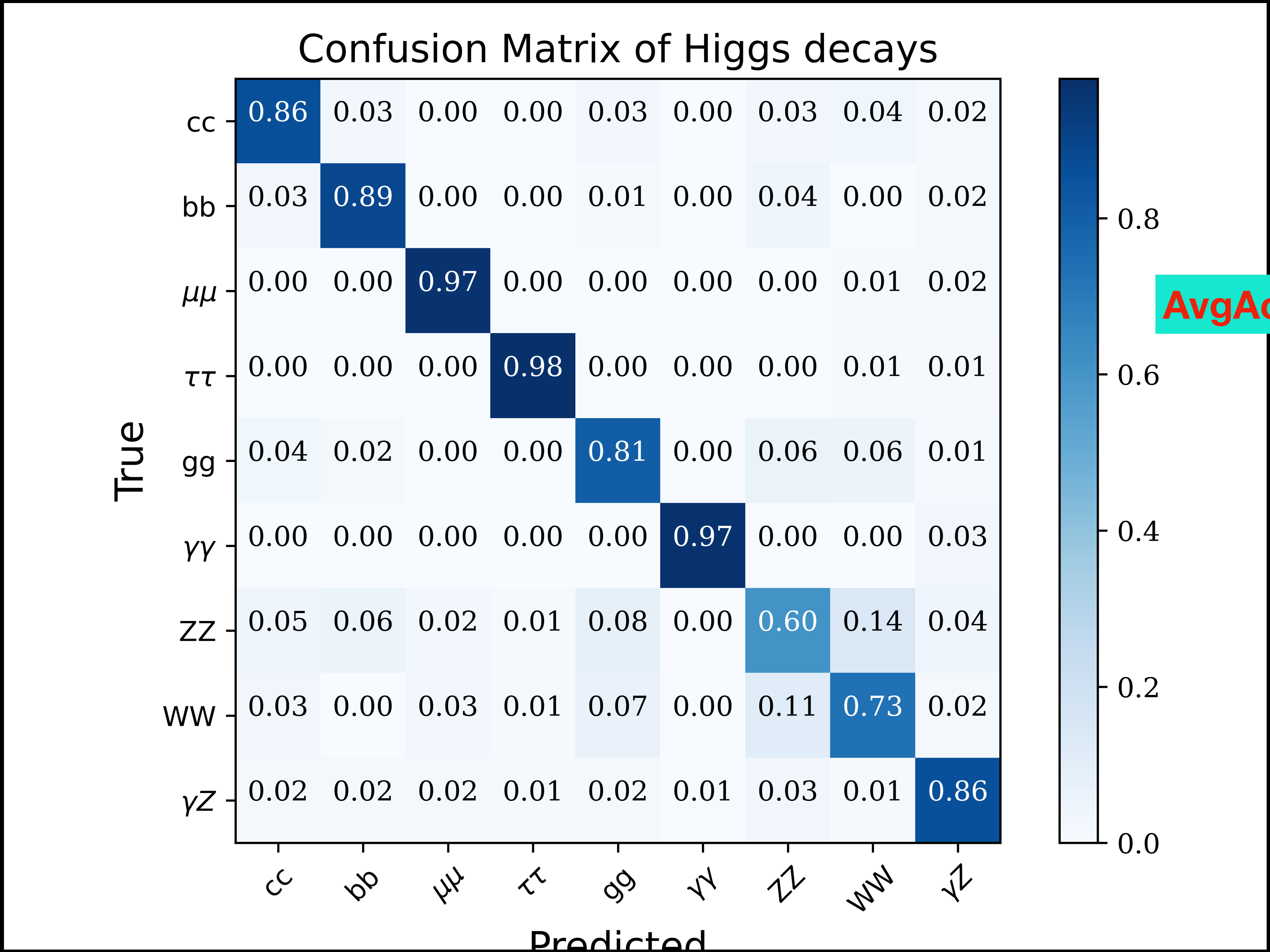
- 100 k events for each of 9 Higgs decay modes:
  $cc, bb, \mu\mu, \tau\tau, gg, \gamma\gamma, ZZ, WW, \gamma Z$
- Train: validation: test = 8:1:1
- Fast simulation
  - momenta of charged tracks and energies of neutral particle smeared according to the performance in CDR
  - Impact parameters of charged track using ideal resolution
- No SM backgrounds taken into account



Receiver operating characteristic
ROC & AUC
Bkgrnd rej.
Signal eff.
0.0
0.2
0.4
0.6
0.8
1.0
Micro-average ROC (AUC = 0.991)
Macro-average ROC (AUC = 0.985)
ROC of cc (AUC = 0.987)
ROC of bb (AUC = 0.992)
ROC of $\mu\mu$ (AUC = 1.000)
ROC of $\tau\tau$ (AUC = 0.999)
ROC of gg (AUC = 0.981)
ROC of $\gamma\gamma$ (AUC = 1.000)
ROC of ZZ (AUC = 0.943)
ROC of WW (AUC = 0.973)
ROC of $\gamma Z$ (AUC = 0.994)


Confusion Matrix of Higgs decays
True
Predicted
cc
bb
μμ
ττ
gg
γγ
ZZ
WW
γZ
0.85 0.05 0.00 0.00 0.01 0.00 0.03 0.05 0.01
0.04 0.86 0.00 0.00 0.01 0.00 0.07 0.00 0.01
0.00 0.00 1.00 0.00 0.00 0.00 0.00 0.00 0.01
0.00 0.00 0.00 0.97 0.00 0.00 0.00 0.00 0.01
0.04 0.03 0.00 0.00 0.79 0.00 0.08 0.05 0.01
0.00 0.00 0.00 0.00 0.00 0.99 0.00 0.00 0.00
0.06 0.07 0.00 0.01 0.10 0.00 0.62 0.11 0.02
0.05 0.00 0.00 0.01 0.08 0.00 0.11 0.75 0.01
0.01 0.01 0.01 0.02 0.01 0.00 0.02 0.00 0.93
1.0
0.8
0.6
0.4
0.2
0.0
AvgAcc = 87%

# DGCNN on same samples achieves similar performance




AvgAcc = 85%

# Discussion on the results

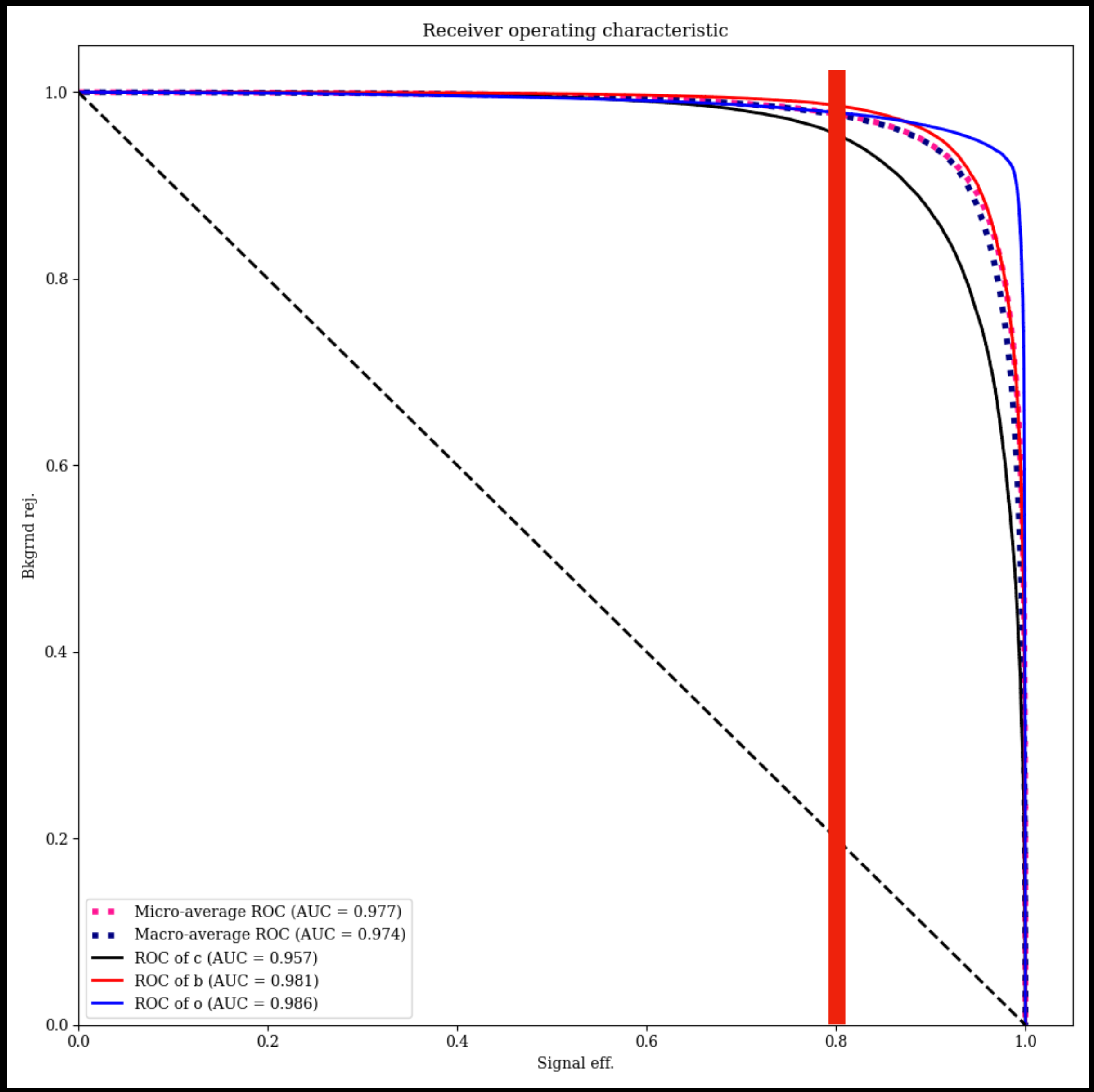
- $\mu\mu, \tau\tau$, and $\gamma\gamma$ best performance
  - Bonus: $\tau\tau$ as good as $\mu\mu$
- ZZ not as good as others as expected
- Confusion among di-jets, WW,  and ZZ
  - gg fakes cc since gluon more likely splits into cc than bb
  - gg also  fakes WW, ZZ hadronic decays
- $\gamma Z$ rather good

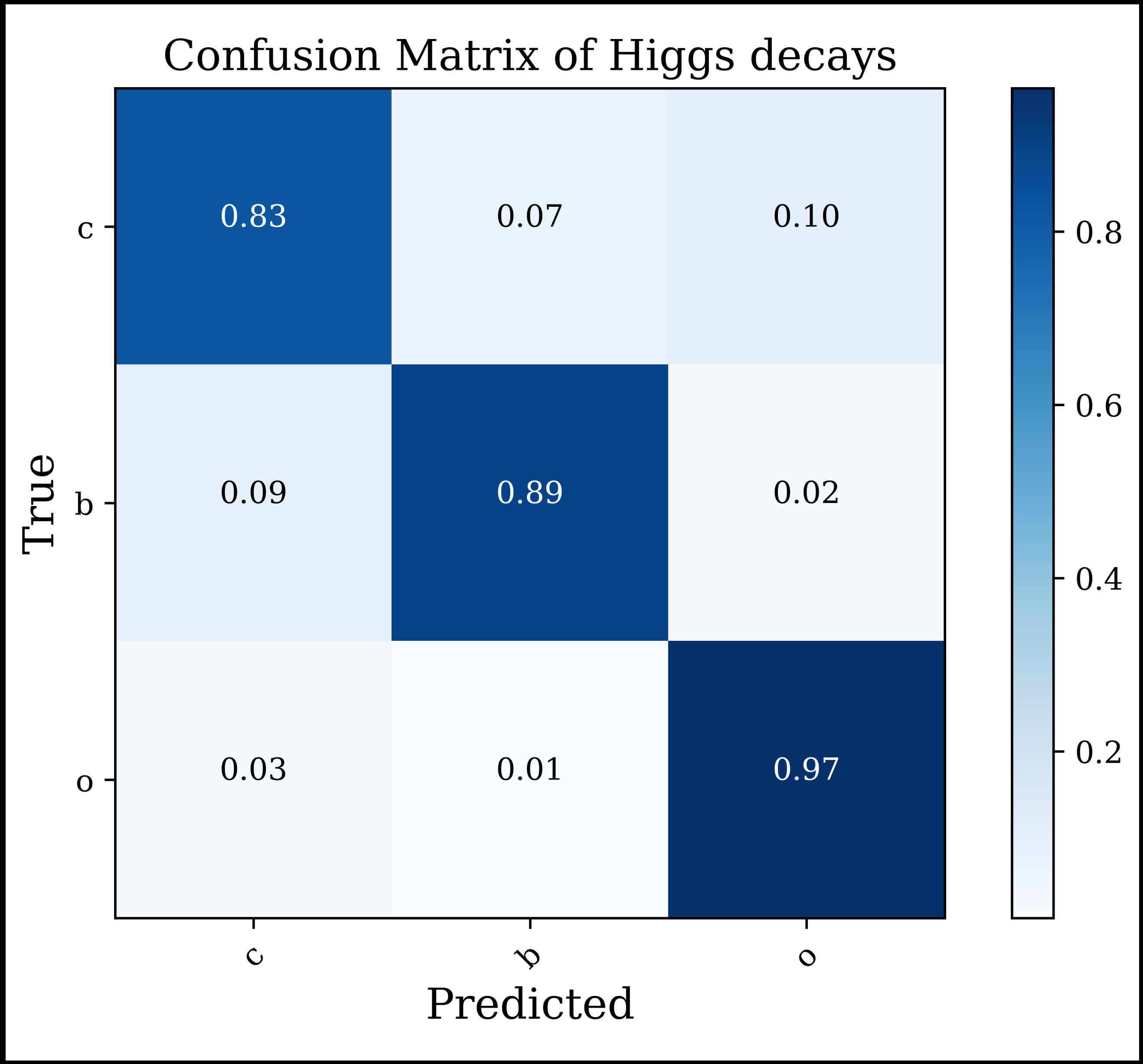
# EFN for flavor tagging

- Input: 300 k Z—> bb, cc, and other jets events
- Same fast simulation configuration
- Using fastjet/ee-kt algorithm to force all particles to 2 jets
- Train: validation: test  = 8:1:1

# Single jet tagging

**Averaged accuracy is 89%, CDR: 80%**

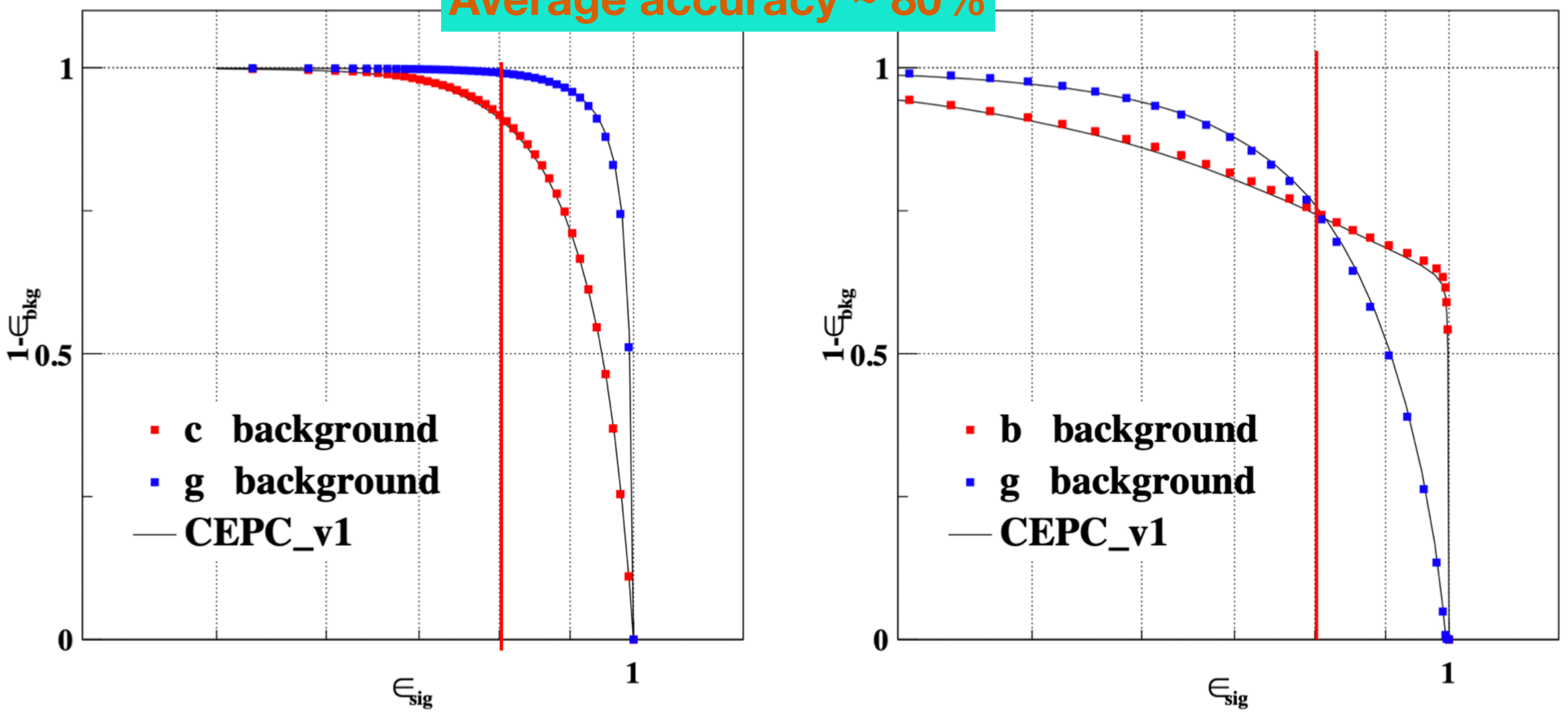



**In the worst case of c-tagging rejection power still > 90% for 80% efficiency**

# CDR referene

Average accuracy ~ 80%



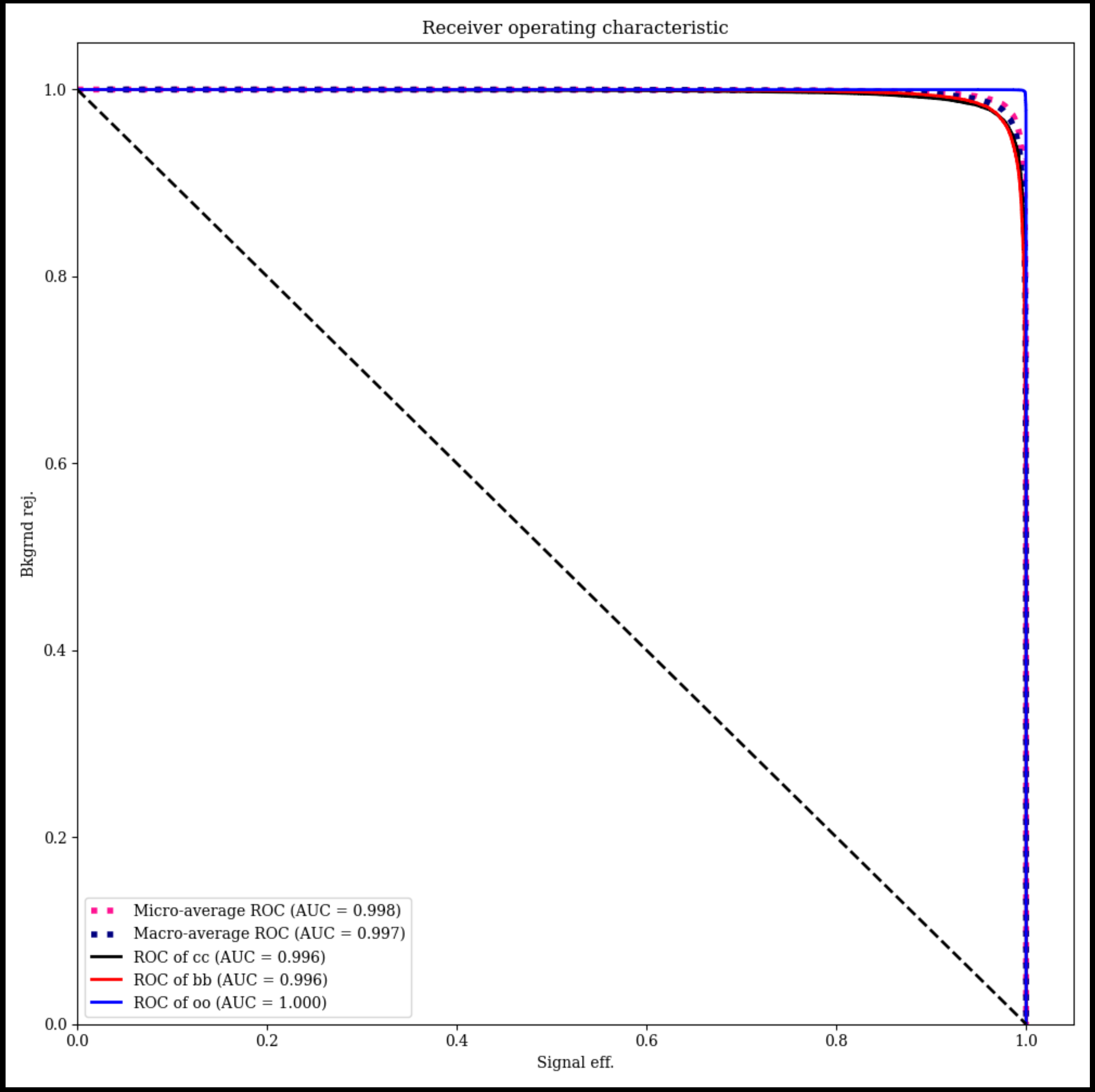

Receiver Operating Characteristic Curve (ROC)

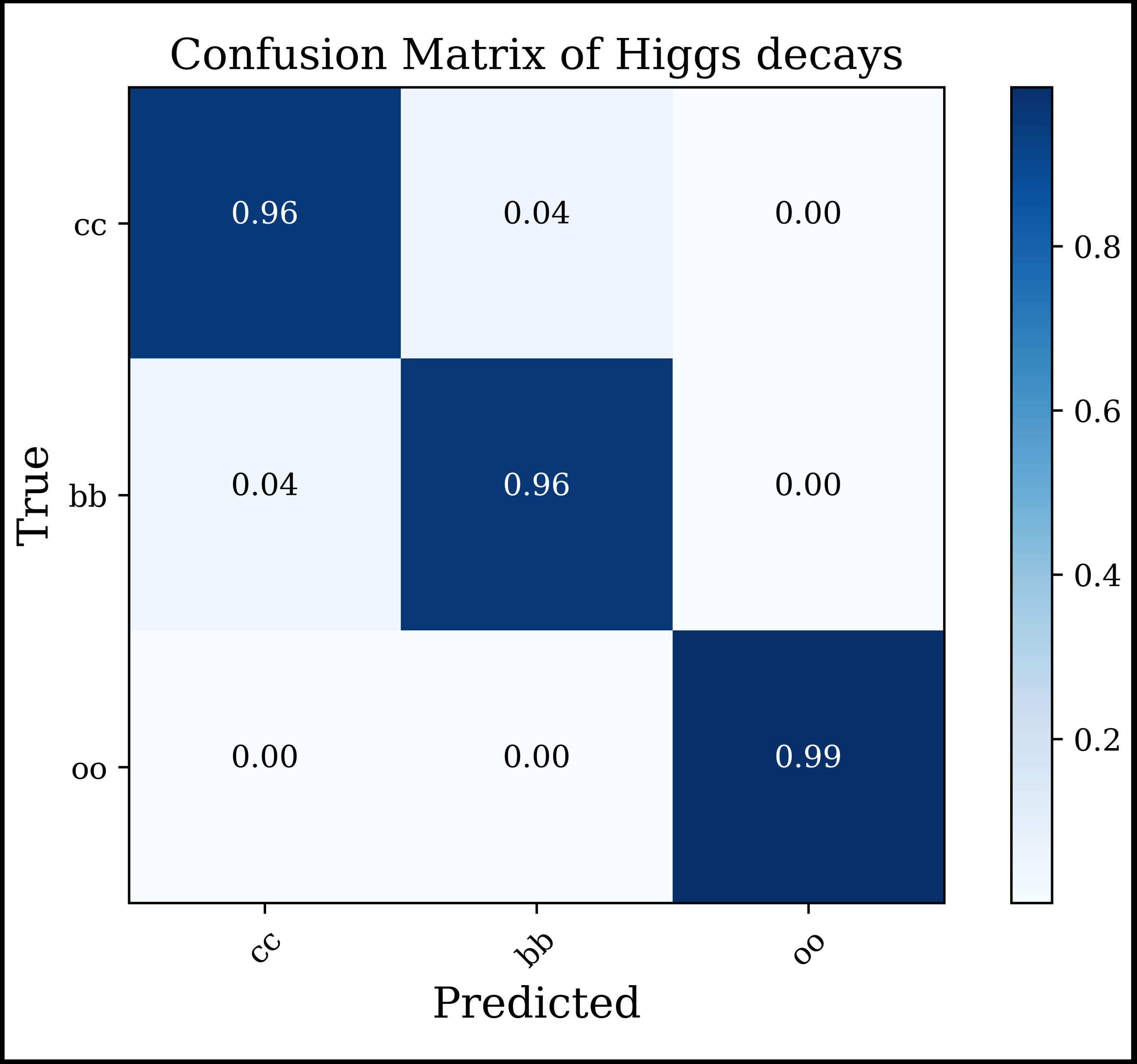
80% b-tagging eff. : Reject 90% c and 99% o jets

80% c-tagging eff. : Reject 75% b and 75% o jets

# Jet pair tagging

**Averaged accuracy is 97%, good news for Rb & Rc measurement**



# Back to the efficiency matrix

- *No dependence on the branching fractions of Higgs decays in MC*
- *Make use of full confusion matrix information*
- one single parameter, det |E|, *quantifies the detector performance*

Single purpose optimization instead of a bunch of benchmarks

Useful for detector optimization

$$\boxed{\Sigma_B \propto \frac{1}{|E|^2}}$$

# Summary and plan

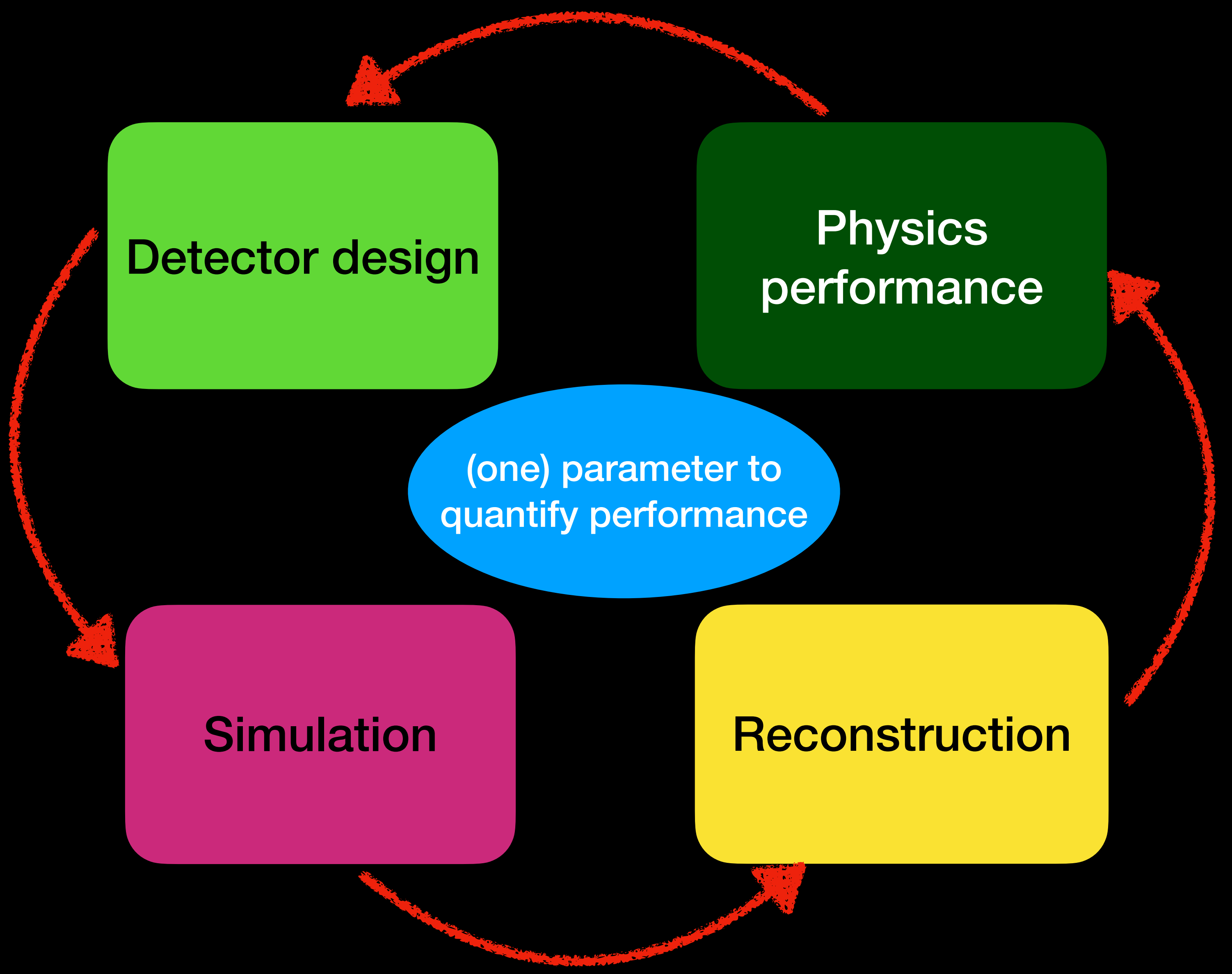
- Summary
  - Global analysis of Higgs Br's could improve the precision simultaneously
  - Global analysis of the all 9 decays can serve as a metric for optimization: advantage of one single parameter
  - Byproducts: same ML approach improves jet-tagging performance
- Plan
  - More validations for the procedure
  - Use GCNN algorithm to cross check EFN
  - Add more decay modes, such as invisible decay, ….
  - Take the main backgrounds into account
  - Quantify the impacts of key detector performances on |E|, such as P, E, impact parameters, …
  - Setup a framework based ML for fast iteration of detector optimization
  - Move to full simulation in CEPCSW for detector optimization

The end

Thanks a lot

# Common collider observables decomposed into per-particle maps Φ and functions F

| **Observable $\mathcal{O}$** | | **Map $\Phi$** | **Function $F$** |
|---|---|---|---|
| Mass | $m$ | $p^\mu$ | $F(x^\mu) = \sqrt{x^\mu x_\mu}$ |
| Multiplicity | $M$ | $1$ | $F(x) = x$ |
| Track Mass | $m_{\text{track}}$ | $p^\mu \mathbb{I}_{\text{track}}$ | $F(x^\mu) = \sqrt{x^\mu x_\mu}$ |
| Track Multiplicity | $M_{\text{track}}$ | $\mathbb{I}_{\text{track}}$ | $F(x) = x$ |
| Jet Charge [72] | $\mathcal{Q}_\kappa$ | $(p_T, Q\, p_T^\kappa)$ | $F(x, y) = y/x^\kappa$ |
| Eventropy [74] | $z \ln z$ | $(p_T, p_T \ln p_T)$ | $F(x, y) = y/x - \ln x$ |
| Momentum Dispersion [93] | $p_T^D$ | $(p_T, p_T^2)$ | $F(x, y) = \sqrt{y/x^2}$ |
| $C$ parameter [94] | $C$ | $(\lvert\vec{p}\rvert, \vec{p} \otimes \vec{p}/\lvert\vec{p}\rvert)$ | $F(x, Y) = \frac{3}{2x^2}[(\operatorname{Tr} Y)^2 - \operatorname{Tr} Y^2]$ |




# Detector design & Optimization

Multi-purpose optimization: a bunch of benchmarks —

A single parameter is favored, single-purpose optimization