A global analysis approach to measure Higgs Br's

Y. Bai, L. Liao, G. Li, W. Song, J. Sun, S. Wang, H. Zhang, Z. Zhang, B. Zhou

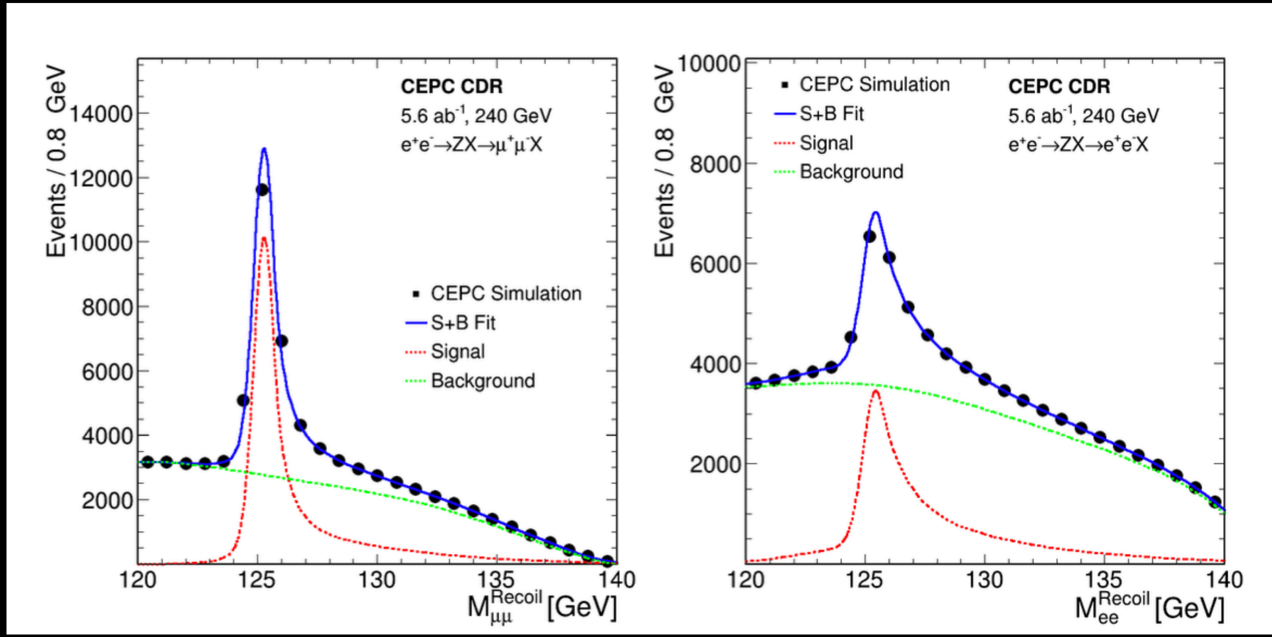
CEPC physics workshop,
Yangzhou, April 14-17, 2021

Outline

- Introduction
- Methods
- Preliminary results of the matrix
- Summary and plan

A global analysis approach, talk at PKU workshop, 2019

- CEPC Higgs Br measurement
  - 9 Higgs decays accessible at CEPC
  - Typical S/N ratio ~ $1/O(1)$
  - Some Br's very small ($\mu\mu, \gamma\gamma, \gamma Z$)
  - Contaminations among them are an issue
    - cc/bb/gg/WW/ZZ/$\tau\tau$: hadronic FS
- Global analysis shows promising results



		MLT		POS
Bcc	2.713%	0.773%	(0.779%	0.790%)
Bbb	57.799%	0.102%	(0.111%	0.171%)
Bmm	0.023%	8.547%	(8.559%	8.560%)
Btt	6.319%	0.492%	(0.501%	0.518%)
Bgg	8.619%	0.413%	(0.424%	0.443%)
Baa	0.227%	2.728%	(2.731%	2.734%)
BaZ	0.150%	3.350%	(3.353%	3.356%)
BZZ	2.647%	0.783%	(0.789%	0.800%)
BWW	21.496%	0.235%	(0.249%	0.281%)

Very rough estimation

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	–	$< 0.30\%$

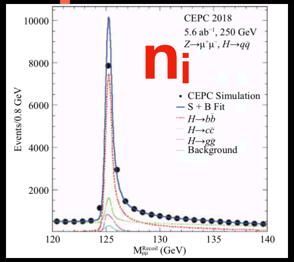
Global analysis approach

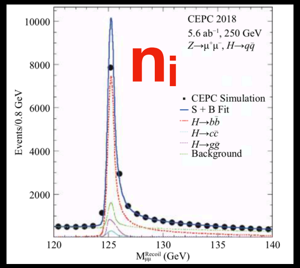
- Measuring all Br's simultaneously
  - **Complete confusion matrix**
  - **Multinomial distribution: smaller stat. uncertainty**
  - **Constraint: improve precision**

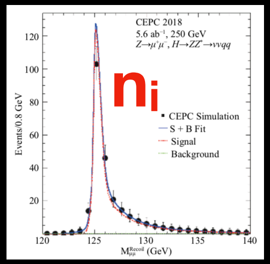
$\sigma_N = \sqrt{N \times p}$ **for Poisson**

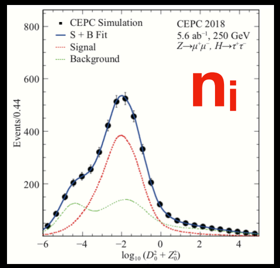
$\sigma_N = \sqrt{N \times p \times (1-p)}$ **for multinomial**

B_i =



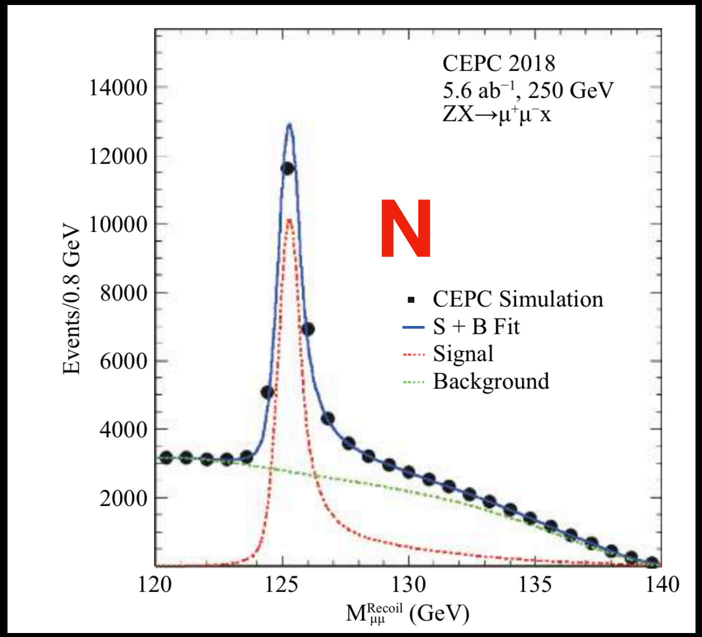






+ … …

―――――――――――――――――――



non-Higgs background

— subtracted with fitting for other method

— worsen σ_{n_i}'s

The simplest example
—2 decays only

Efficiency matrix

From MC, no dependence on Br's

MODULATION Matrix

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

Observation

Production

A produced final state j reconstructed as final state i

Demodulation

Measurement: $N = E^{-1} n$

N→Branching fractions
Formula shows very simple features

- Variance of B's proportional to $1/(N^4|E|^2)$
- N^4 : statistical power
- $|E|^2$: performances of Detector x Reconstruction x Analysis
- 2 Br's have same uncertainties
- Always valid for # of Br's > 2

This formula still valid for more decay modes

Σ_B: **Variance of Br's**

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

If someone wants complete formula …

$$\Sigma^B = \begin{pmatrix} \sigma^2_{B_1} & \sigma_{B_{12}} \\ \sigma_{B_{21}} & \sigma^2_{B_2} \end{pmatrix} = \frac{(n_2\vec{\sigma}_{n_1} - n_1\vec{\sigma}_{n_2})^2}{N^4|E|^2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

Efficiency matrix

Individual analysis

- Signal: high efficiency
- Backgrounds: lower fake rate
  - Higgs backgrounds: contaminations among Higgs decays
  - Non-Higgs backgrounds: SM backgrounds
- $(\epsilon_s, \epsilon_{hb1}, \epsilon_{hb2}, \epsilon_{hb3}, \ldots, \epsilon_{SM})$
- Usually done by a person or a small group
- Different analyses lack of information sharing

Global analysis: Solve all Br's simultaneously

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

1 2 3 4 5 6 7 8 9

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

SM backgrounds not shown here

By minimizing the χ^2

$$\chi^2_{\mu\mu} = \sum_{ij} \left[\sum_k \epsilon_{ik} N_k - n_i \right] V_{ij} \left[\sum_k \epsilon_{jk} N_k - n_j \right] + \frac{\left[\sum_k N_k - N^t \right]^2}{\sigma^2_{N_t}}$$

Global analysis vs. individual analysis

- More information used
- Information sharing among different decay modes via efficiency matrix, one group work
- Better precisions
- Systematics is an issue, open question
- Very convenient for machine learning, …

Efficiency matrix from machine learning

Different approaches

- Usually the DBT used for event classification with hand-engineering variables

- Recently various approaches: DNN, CNN, RNN, …

- Graph Neural Network (GNN) will be used

  - Better inductive bias

  - Using particle level information as input

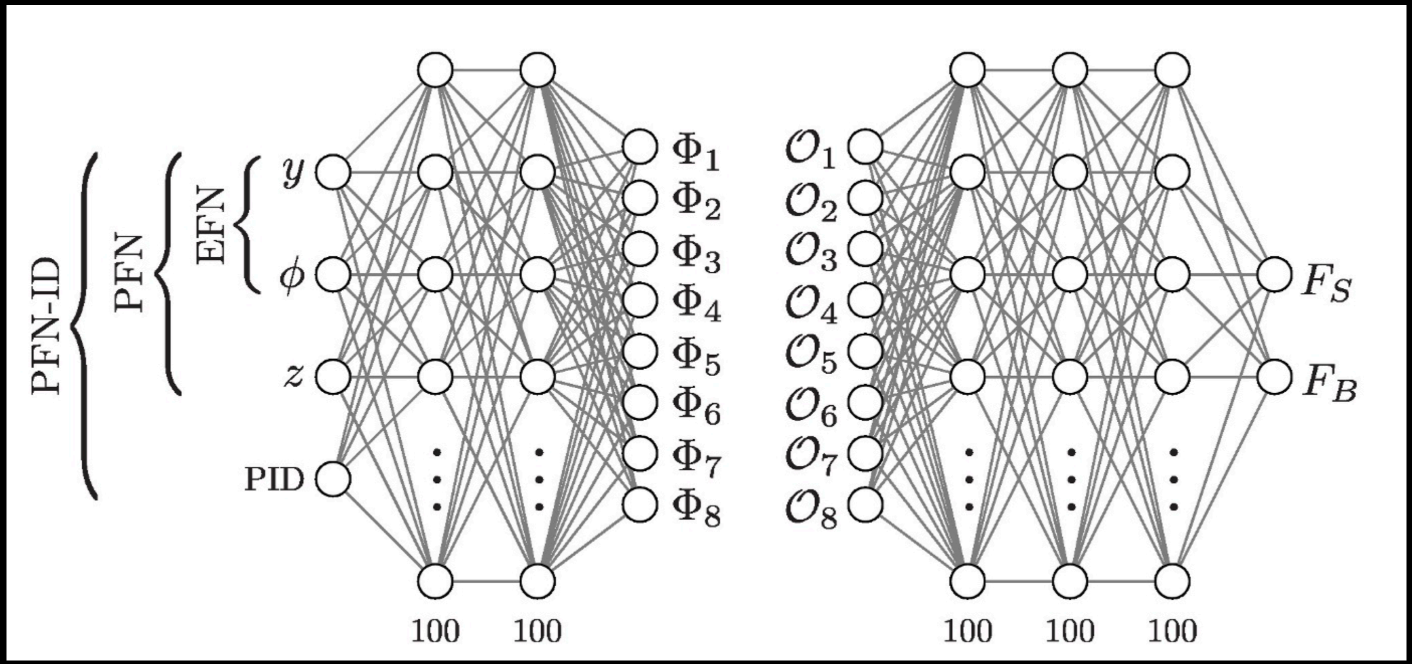
    - No jet-clustering

    - No lepton/photon isolation

Energy Flow Network

JHEP01(2019)121

- **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**
- **No explicitly dependence on the ordering of the inputs**
- **Respect the permutation symmetry**
- **Infra-red and collinear safety naturally achieved**



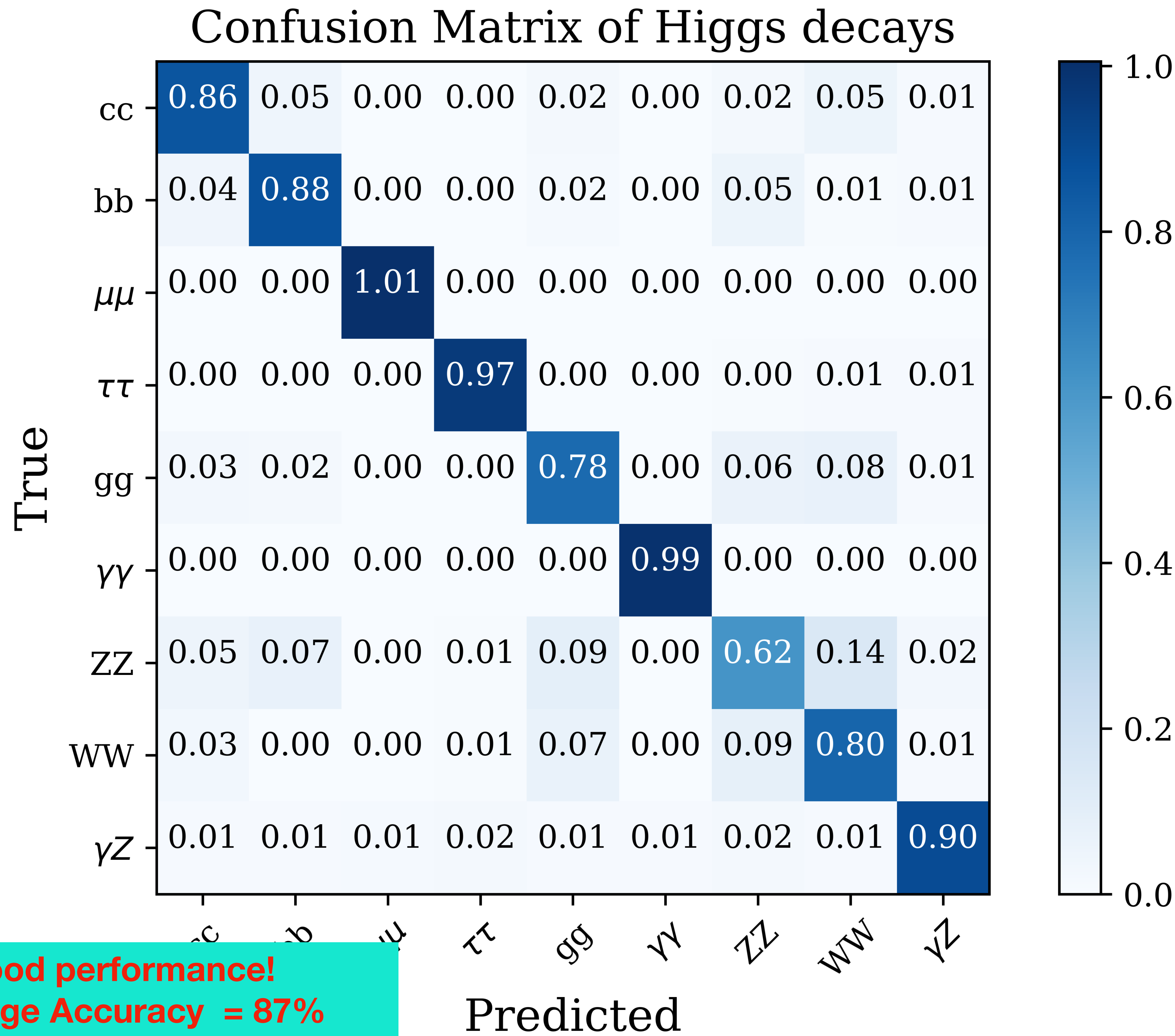
The key mathematical fact: A generic function of a set of objects can be decomposed to arbitrarily good approximation in a practical and intuitive way [arXiv:1703.06114]

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

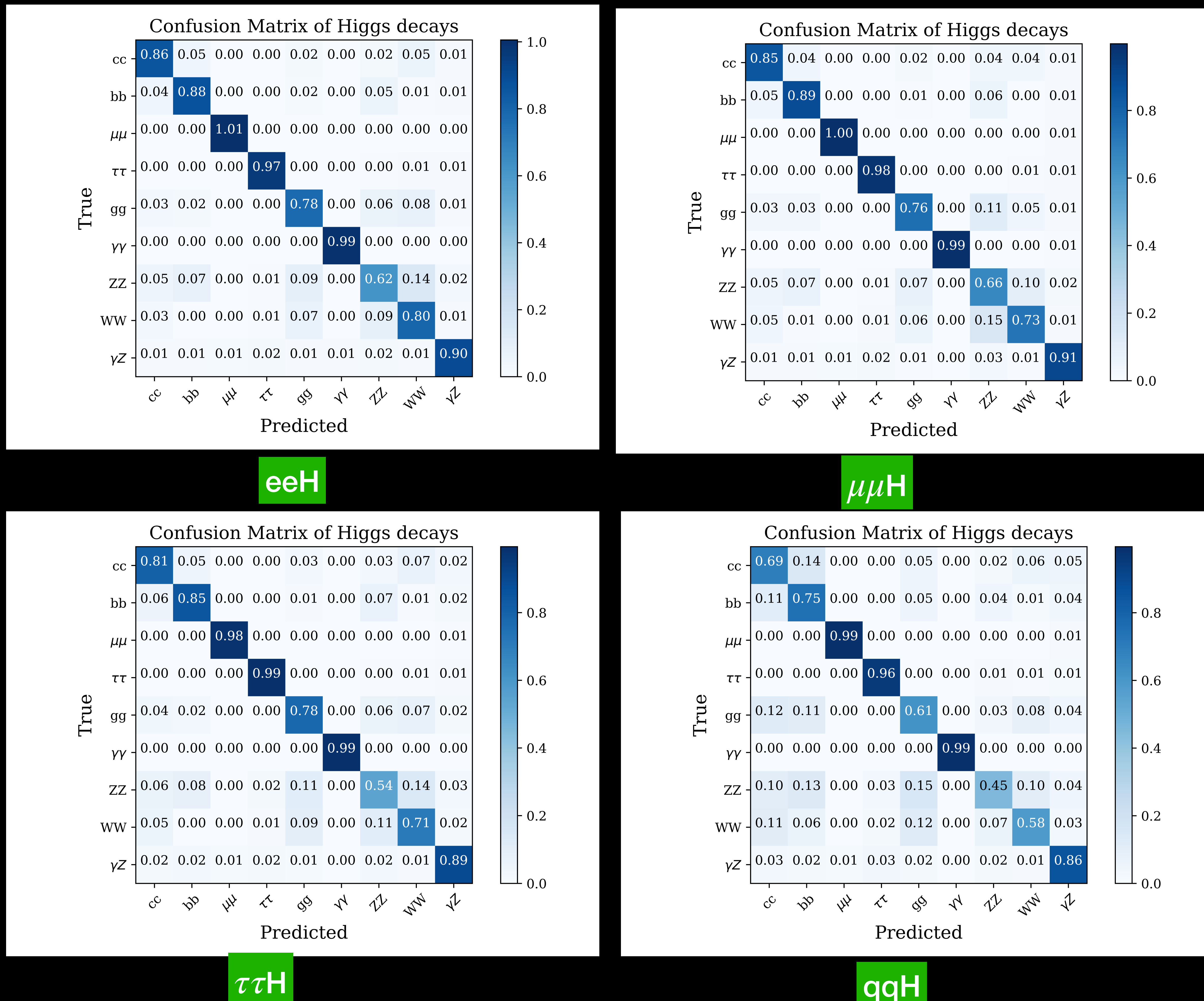
Inputs: $e^+e^- \rightarrow ZH, Z \rightarrow l^+l^-, qq$

- 100 k evts for each 9 Higgs decays:
  $cc, bb, \mu\mu, \tau\tau, gg, \gamma\gamma, ZZ, WW, \gamma Z$
- Train: validation: test = 8:1:1
- Fast simulation
  - momenta of tracks and energies of neutrals smeared according to CEPC_v4
  - Ideal PID and Impact parameters
- No SM backgrounds taken into account

Some preliminary results



Good performance!
Average Accuracy = 87%
(11.1% of random guess)



Confusion Matrix of Higgs decays
cc 0.86 0.05 0.00 0.00 0.02 0.00 0.02 0.05 0.01
bb 0.04 0.88 0.00 0.00 0.02 0.00 0.05 0.01 0.01
μμ 0.00 0.00 1.01 0.00 0.00 0.00 0.00 0.00 0.00
ττ 0.00 0.00 0.00 0.97 0.00 0.00 0.00 0.01 0.01
gg 0.03 0.02 0.00 0.00 0.78 0.00 0.06 0.08 0.01
γγ 0.00 0.00 0.00 0.00 0.00 0.99 0.00 0.00 0.00
ZZ 0.05 0.07 0.00 0.01 0.09 0.00 0.62 0.14 0.02
WW 0.03 0.00 0.00 0.01 0.07 0.00 0.09 0.80 0.01
γZ 0.01 0.01 0.01 0.02 0.01 0.01 0.02 0.01 0.90
cc bb μμ ττ gg γγ ZZ WW γZ
True
Predicted
eeH
Confusion Matrix of Higgs decays
cc 0.85 0.04 0.00 0.00 0.02 0.00 0.04 0.04 0.01
bb 0.05 0.89 0.00 0.00 0.01 0.00 0.06 0.00 0.01
μμ 0.00 0.00 1.00 0.00 0.00 0.00 0.00 0.00 0.01
ττ 0.00 0.00 0.00 0.98 0.00 0.00 0.00 0.01 0.01
gg 0.03 0.03 0.00 0.00 0.76 0.00 0.11 0.05 0.01
γγ 0.00 0.00 0.00 0.00 0.00 0.99 0.00 0.00 0.01
ZZ 0.05 0.07 0.00 0.01 0.07 0.00 0.66 0.10 0.02
WW 0.05 0.01 0.00 0.01 0.06 0.00 0.15 0.73 0.01
γZ 0.01 0.01 0.01 0.02 0.01 0.00 0.03 0.01 0.91
cc bb μμ ττ gg γγ ZZ WW γZ
True
Predicted
μμH
Confusion Matrix of Higgs decays
cc 0.81 0.05 0.00 0.00 0.03 0.00 0.03 0.07 0.02
bb 0.06 0.85 0.00 0.00 0.01 0.00 0.07 0.01 0.02
μμ 0.00 0.00 0.98 0.00 0.00 0.00 0.00 0.00 0.01
ττ 0.00 0.00 0.00 0.99 0.00 0.00 0.00 0.01 0.01
gg 0.04 0.02 0.00 0.00 0.78 0.00 0.06 0.07 0.02
γγ 0.00 0.00 0.00 0.00 0.00 0.99 0.00 0.00 0.00
ZZ 0.06 0.08 0.00 0.02 0.11 0.00 0.54 0.14 0.03
WW 0.05 0.00 0.00 0.01 0.09 0.00 0.11 0.71 0.02
γZ 0.02 0.02 0.01 0.02 0.01 0.00 0.02 0.01 0.89
cc bb μμ ττ gg γγ ZZ WW γZ
True
Predicted
ττH
Confusion Matrix of Higgs decays
cc 0.69 0.14 0.00 0.00 0.05 0.00 0.02 0.06 0.05
bb 0.11 0.75 0.00 0.00 0.05 0.00 0.04 0.01 0.04
μμ 0.00 0.00 0.99 0.00 0.00 0.00 0.00 0.00 0.01
ττ 0.00 0.00 0.00 0.96 0.00 0.00 0.01 0.01 0.01
gg 0.12 0.11 0.00 0.00 0.61 0.00 0.03 0.08 0.04
γγ 0.00 0.00 0.00 0.00 0.00 0.99 0.00 0.00 0.00
ZZ 0.10 0.13 0.00 0.03 0.15 0.00 0.45 0.10 0.04
WW 0.11 0.06 0.00 0.02 0.12 0.00 0.07 0.58 0.03
γZ 0.03 0.02 0.01 0.03 0.02 0.00 0.02 0.01 0.86
cc bb μμ ττ gg γγ ZZ WW γZ
True
Predicted
qqH

With the efficiency matrix and n_i

- The $\chi^2 = \chi^2_{ee} + \chi^2_{\mu\mu} + \chi^2_{\tau\tau} + \chi^2_{q\bar{q}}$ is going to be minimized to extract 9 branching ratios
- This part is ongoing …

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

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

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

Problem successfully becomes how to Maximize |E|

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

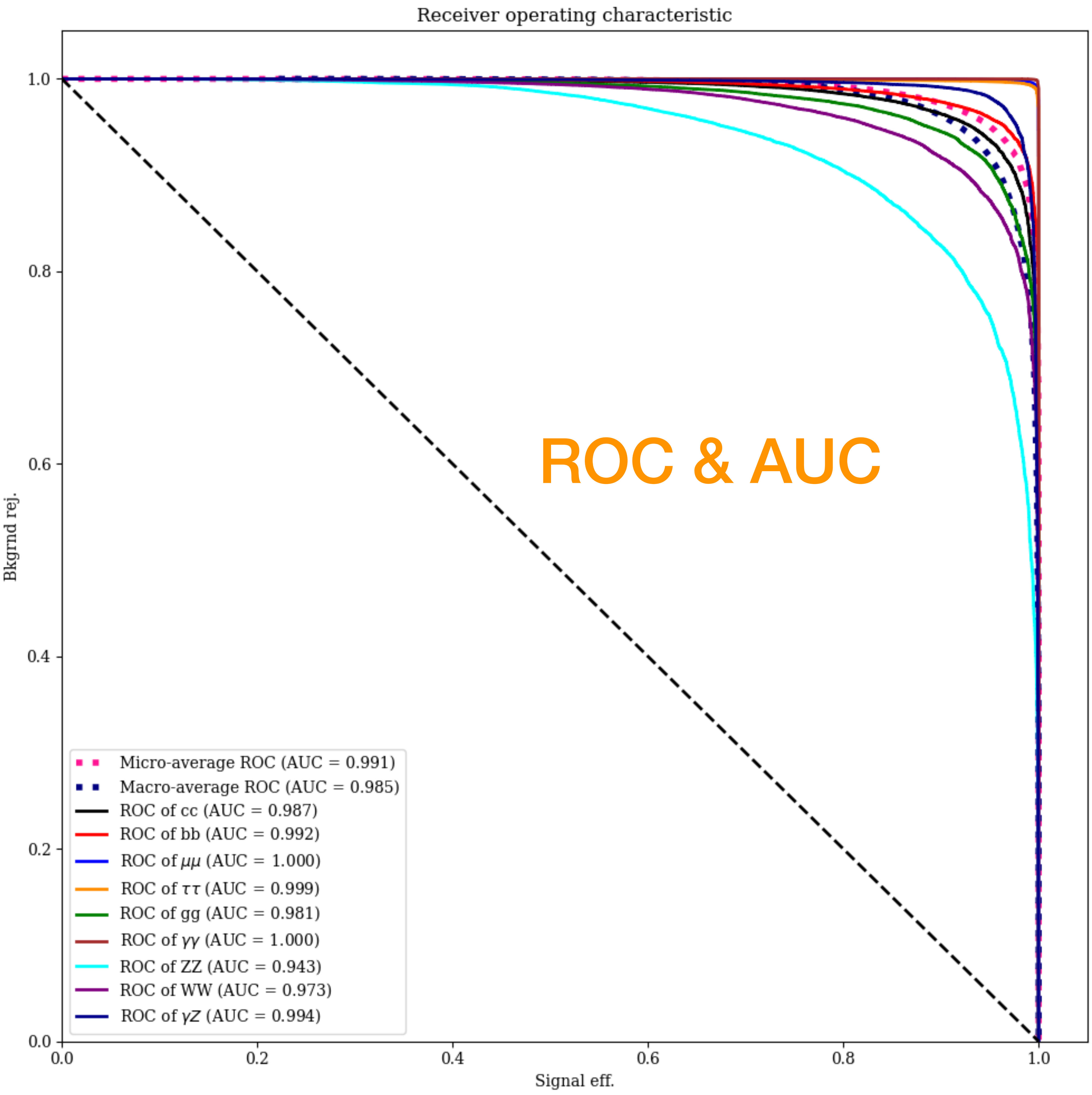
$$\Sigma_B \propto \frac{1}{|E|^2}$$

Summary and plan

- A proof-of-principle study shows
  - ☑ Global analysis improves the precision
  - ☑ ML method used to get the efficiency matrix, using particle level information as input
  - ☑ Confusion matrix as a metric for optimization
    - ☑ Its determinant is a single parameter and easy for optimization
- Plan
  - ✓ More validations
  - ✓ Realize the minimization of 4 combined process
  - ✓ Taking SM backgrounds into account
  - ✓ Apply it in detector optimization

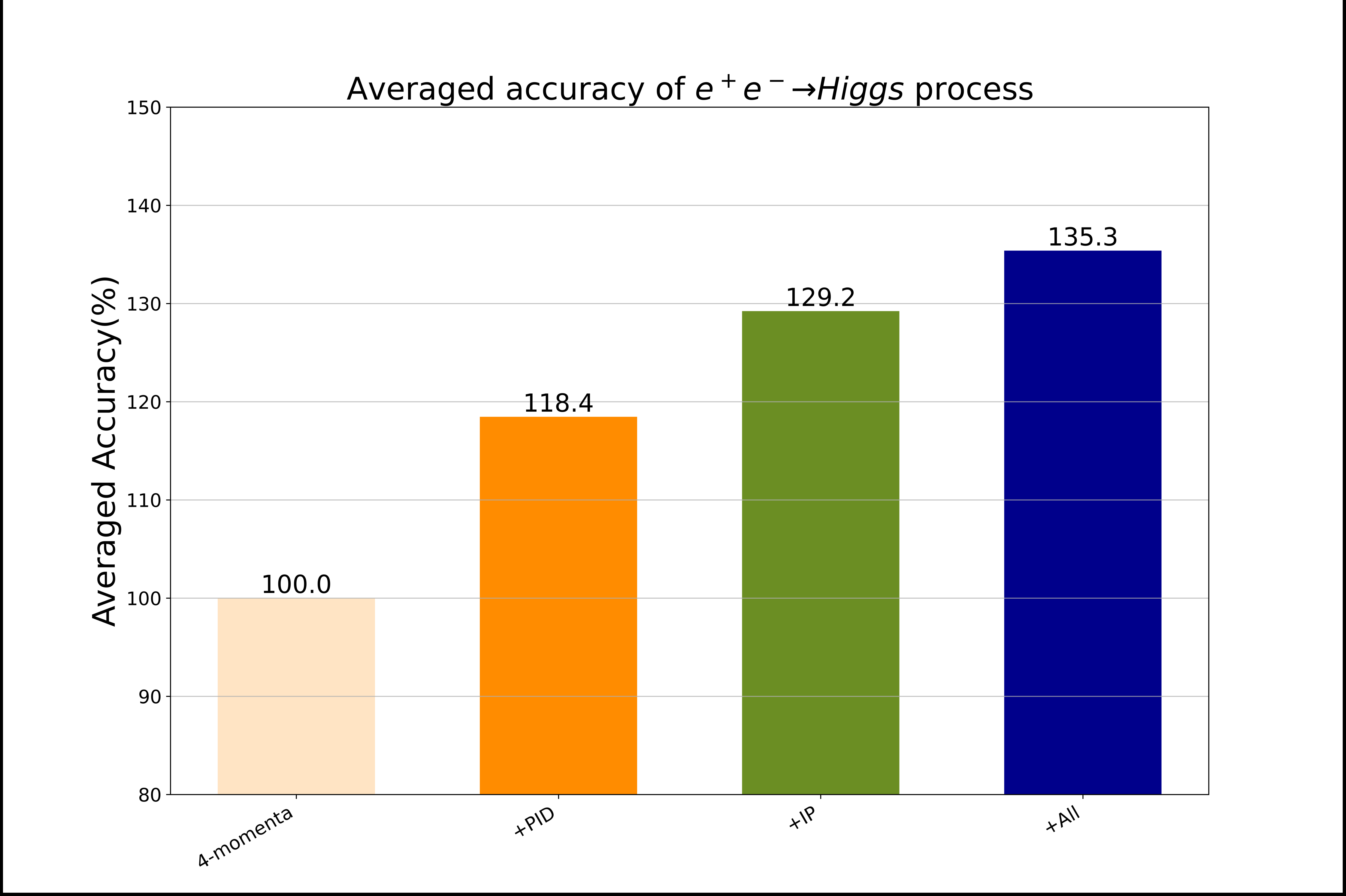
The end

Thanks a lot



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 μμ (AUC = 1.000)
ROC of ττ (AUC = 0.999)
ROC of gg (AUC = 0.981)
ROC of γγ (AUC = 1.000)
ROC of ZZ (AUC = 0.943)
ROC of WW (AUC = 0.973)
ROC of γZ (AUC = 0.994)

The detector performance dependence on E/p/PID/impact parameters

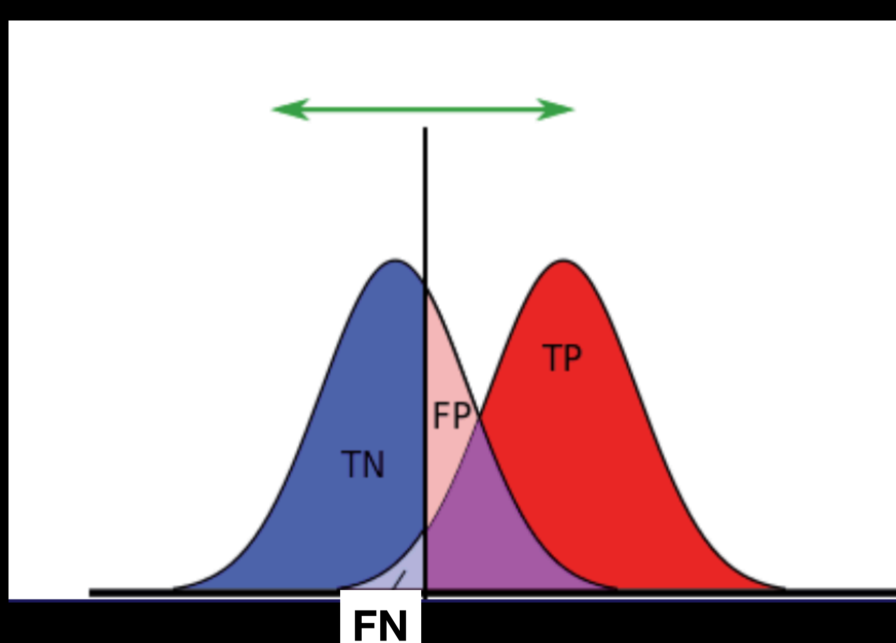


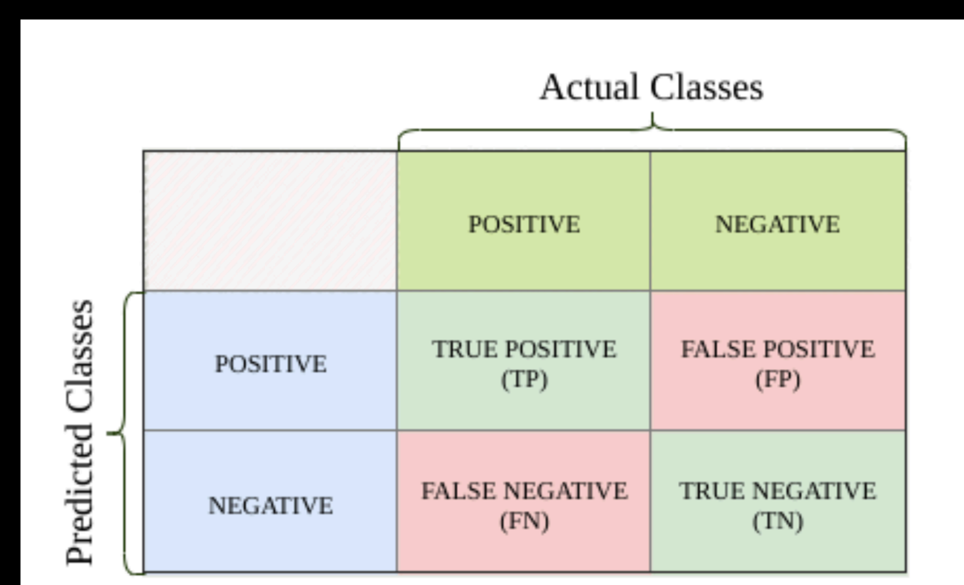
Discussion on the results

- $\mu\mu, \tau\tau$, 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
- γZ rather good

Termnologies

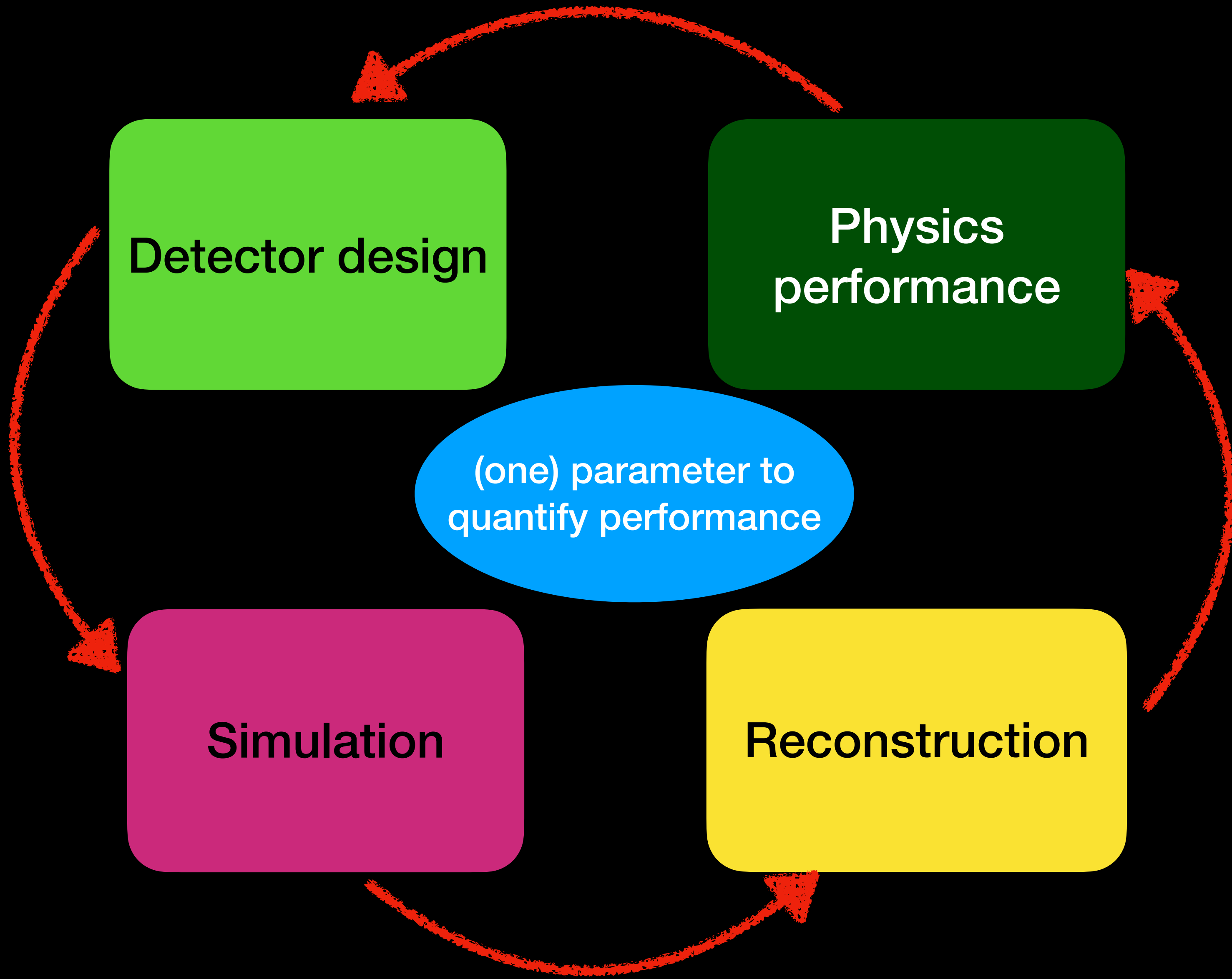
- Tagging efficiency: accuracy in ML
- ROC : Receiver Operating Characteristics Curve, mainly for binary classification,
  - In HEP it is Rejection rate vs. Tagging efficiency (FN rate vs. TP rate)
- AUC : Area under the ROC
- Confusion matrix
  - it is the efficiency matrix when neglecting SM backgrounds





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

Observable $\mathcal{O}$		Map Φ	Function F
Mass	m	p^μ	$F(x^\mu) = \sqrt{x^\mu x_\mu}$
Multiplicity	M	1	$F(x) = x$
Track Mass	m_{track}	$p^\mu \mathbb{I}_{\text{track}}$	$F(x^\mu) = \sqrt{x^\mu x_\mu}$
Track Multiplicity	M_{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