Global analysis of Higgs aided by machine learning

Gang Li

IAS Program on High Energy Physic

Feb 11-17, Hongkong

Outline

- Introduction
- Determining efficiency (confusion) matrix with ML
- Preliminary results
- Summary

Personal rank the difficultness of Higgs analysis at ee colliders

4 x 9 modes in this study, [5 production and 13 (9) decays modes in SM]

Prod/decay	cc	bb	$\mu\mu$	$\tau\tau$	$\gamma\gamma$	gg	WW	ZZ	γZ	ee, uu,dd,ss
eeH (incl. Z fusion)	3	1	5	2	4	1	2	3	5	Not covered yet
$\mu\mu$H	3	1	5	2	4	1	2	3	5	
$\tau\tau$H	3	1	5	2	4	1	2	3	5	
qqH	4	1	2	1	2	5	5	5	3	
$\nu\nu$H (incl. W fusion)	5	1	3	2	3	5	4	2	4	

According to production rate, signal signature, backgrounds, complication of analysis, …

Current estimation of Higgs precision

CEPC: 2205.08553

	240 GeV, 20 ab^{-1}		360 GeV, 1 ab^{-1}		
	ZH	vvH	ZH	vvH	eeH
any	0.26%		1.40%	\	\
H→bb	0.14%	1.59%	0.90%	1.10%	4.30%
H→cc	2.02%		8.80%	16%	20%
H→gg	0.81%		3.40%	4.50%	12%
H→WW	0.53%		2.80%	4.40%	6.50%
H→ZZ	4.17%		20%	21%	
$H \to \tau\tau$	0.42%		2.10%	4.20%	7.50%
$H \to \gamma\gamma$	3.02%		11%	16%	
$H \to \mu\mu$	6.36%		41%	57%	
$Br_{upper}(H \to inv.)$	0.07%		\	\	
$H \to Z\gamma$	8.50%		35%	\	
Width	1.65%		1.10%		

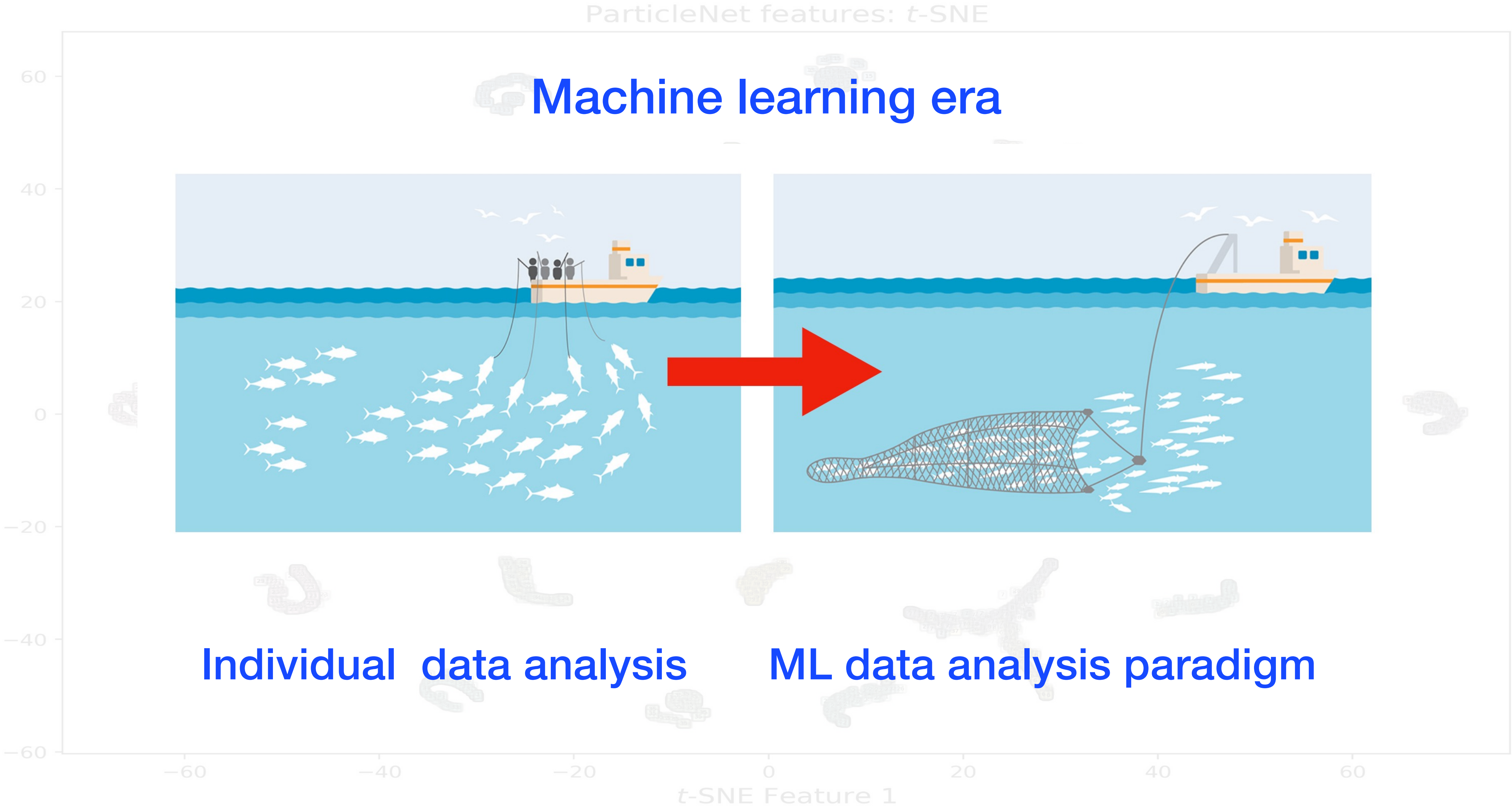
FCC-ee

$\sqrt{s}$ (GeV)	240		365	
Luminosity (ab^{-1})	5		1.5	
$\delta(\sigma BR)/\sigma BR$ (%)	HZ	$\nu\bar{\nu}$ H	HZ	$\nu\bar{\nu}$ H
$H \to$ any	±0.5		±0.9	
$H \to b\bar{b}$	±0.3	±3.1	±0.5	±0.9
$H \to c\bar{c}$	±2.2		±6.5	±10
$H \to gg$	±1.9		±3.5	±4.5
$H \to W^+W^-$	±1.2		±2.6	±3.0
$H \to ZZ$	±4.4		±12	±10
$H \to \tau\tau$	±0.9		±1.8	±8
$H \to \gamma\gamma$	±9.0		±18	±22
$H \to \mu^+\mu^-$	±19		±40	
$H \to$ invisible	< 0.3		< 0.6	

- **Results of CEPC and FCC-ee based individual analysis**
- **Comparable precision**

A lots of efforts

Machine learning era



Individual data analysis

ML data analysis paradigm

Individual data analysis

- **Many types of events produced (N_s, N_b=N_{b1}+N_{b2} + …)**
- **B_s to be measured**
- **N_s : signal and**
- **N_b: backgrounds, could be different types**
- **Event selection:**

$$B = \frac{N_s}{N_s + N_B}$$

$$\begin{pmatrix} n_s \\ n_b \end{pmatrix} = \begin{pmatrix} \boxed{\epsilon_{ss} \quad \epsilon_{sb}} \\ \epsilon_{bs} \quad \epsilon_{bb} \end{pmatrix} \times \begin{pmatrix} N_s \\ N_b \end{pmatrix}$$

What if we measure all branching ratios simultaneously, …

It could be done like ...

$$\begin{pmatrix} n_1 \\ n_2 \\ \vdots \\ n_9 \end{pmatrix} = \begin{pmatrix} \epsilon_{11} & \epsilon_{12} & \cdots & \epsilon_{19} \\ \epsilon_{21} & \epsilon_{22} & \cdots & \epsilon_{29} \\ \vdots & \vdots & \ddots & \vdots \\ \epsilon_{91} & \epsilon_{92} & \cdots & \epsilon_{99} \end{pmatrix} \begin{pmatrix} N_1 \\ N_2 \\ \vdots \\ N_9 \end{pmatrix}$$

Inverting matrix to calculate N_i

$$B_i = \frac{N_i}{N_1 + N_2 + N_3 + \ldots}$$

“on-stop” analysis: measuring all quantities simultaneously

more efficient and better precision

Some mathematics

Efficiency modulate N → n

$$\mathbf{n} = \mathbf{E}\mathbf{N} \ .$$

Their covariances connected by

$$\mathbf{\Sigma}^n \equiv \left(c^n_{ij}\right) = \mathbf{E}\mathbf{\Sigma}^N\mathbf{E}^T \ ,$$

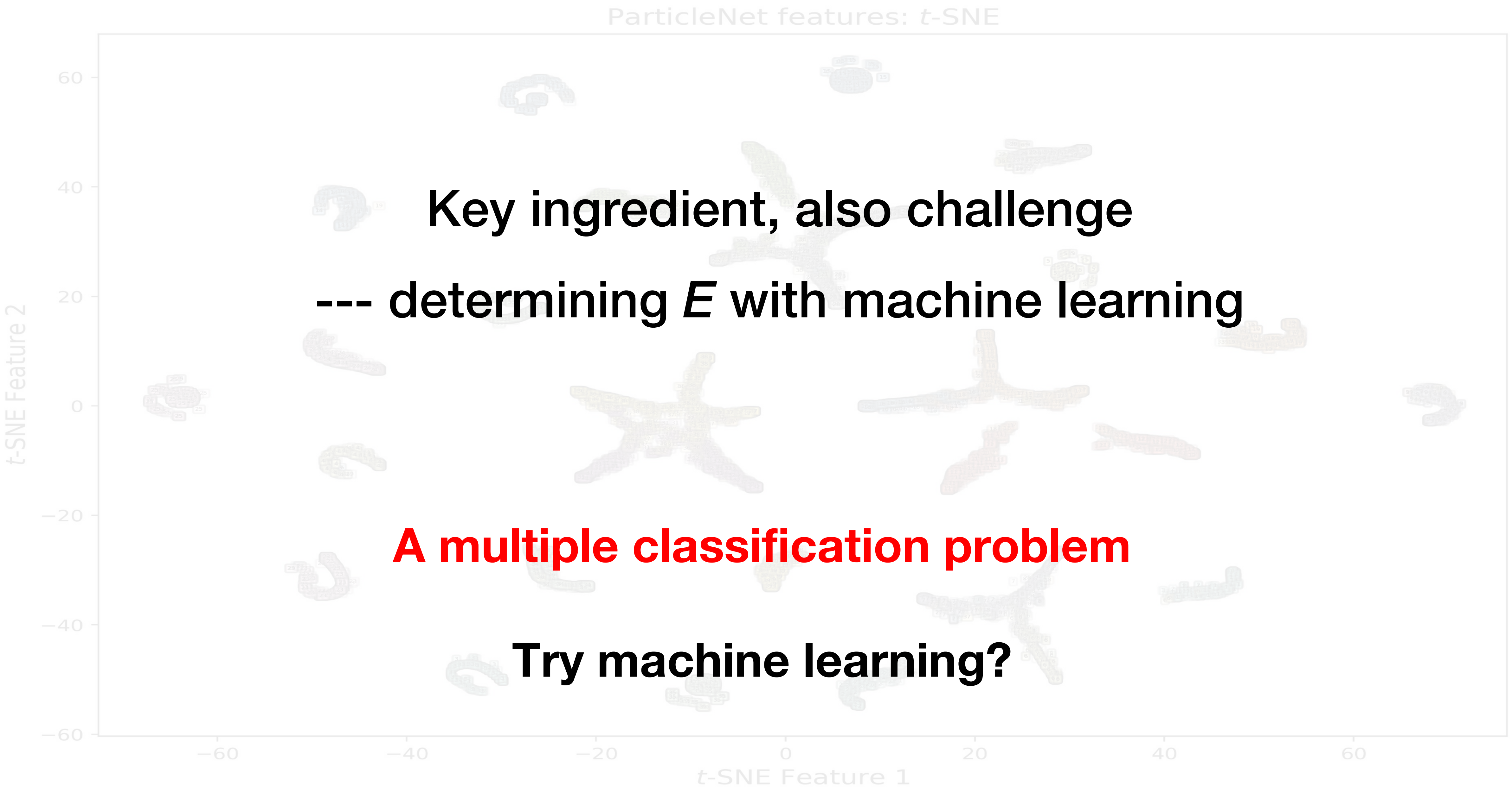
The covariance of N is well known
Σ^N ~ multinomial
so Σ^n is easy

$$\mathbf{\Sigma}^N = N_t^e \begin{pmatrix} B_1(1-B_1) & -B_1B_2 & \dots & -B_1B_m \\ -B_2B_1 & B_2(1-B_2) & \dots & -B_2B_m \\ \vdots & \vdots & \ddots & \vdots \\ -B_mB_1 & -B_mB_2 & \dots & B_m(1-B_m) \end{pmatrix} ,$$

Solve all N_i by minimizing

$$\chi^2_{ee} = \sum_i \frac{\left(\sum_k \epsilon_{ik} N_k - n_i\right)^2}{c_{ii}} + \frac{\left(\sum_k N_k - N_t^e\right)^2}{\sigma^2_{N_t}},$$

or maximizing the likelihood



Key ingredient, also challenge

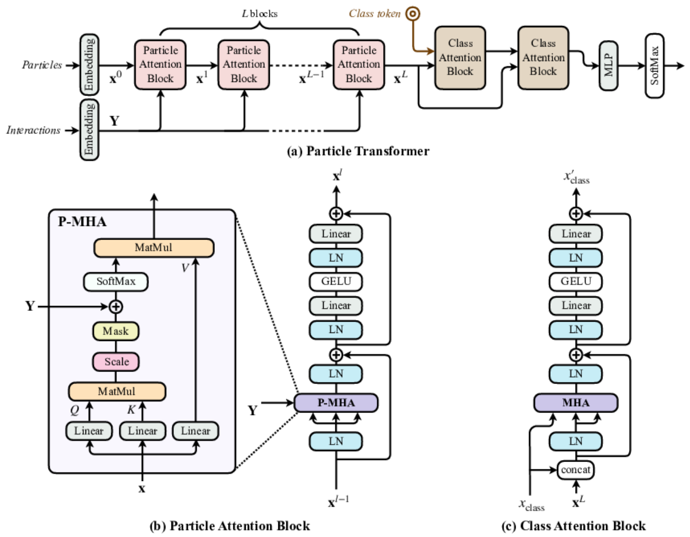
--- determining *E* with machine learning

A multiple classification problem

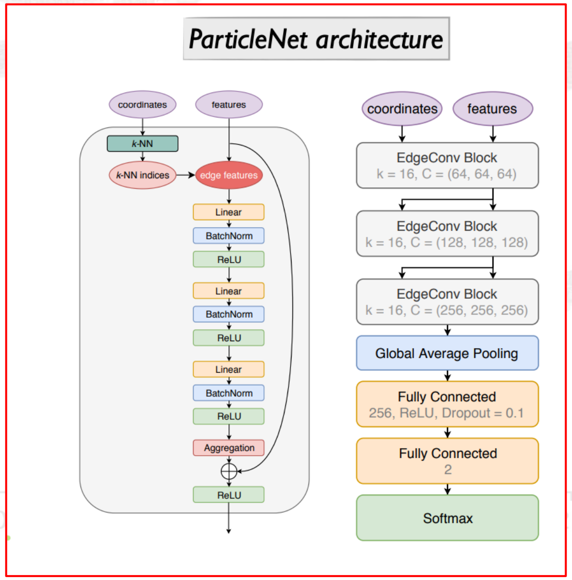
Try machine learning?

The-state-of-the-art ML algorithms developed in recent years

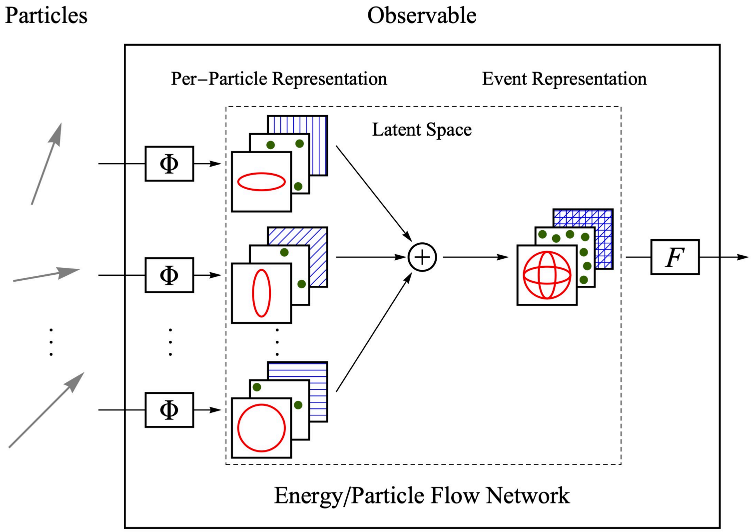
- ParticleNet and ParticleTransformer, see Huilin's talk for more information
- Energy flow networks
- All can achieve excellent performance, ParticleNet used in this study



ParticleTransformer: 2022



ParticleNet: 2019



Energyflow networks: 2018

Data sets: $e^{+}e^{-} \rightarrow ZH, Z \rightarrow l^{+}l^{-}, qq$

- 400 k events for each Higgs decays :

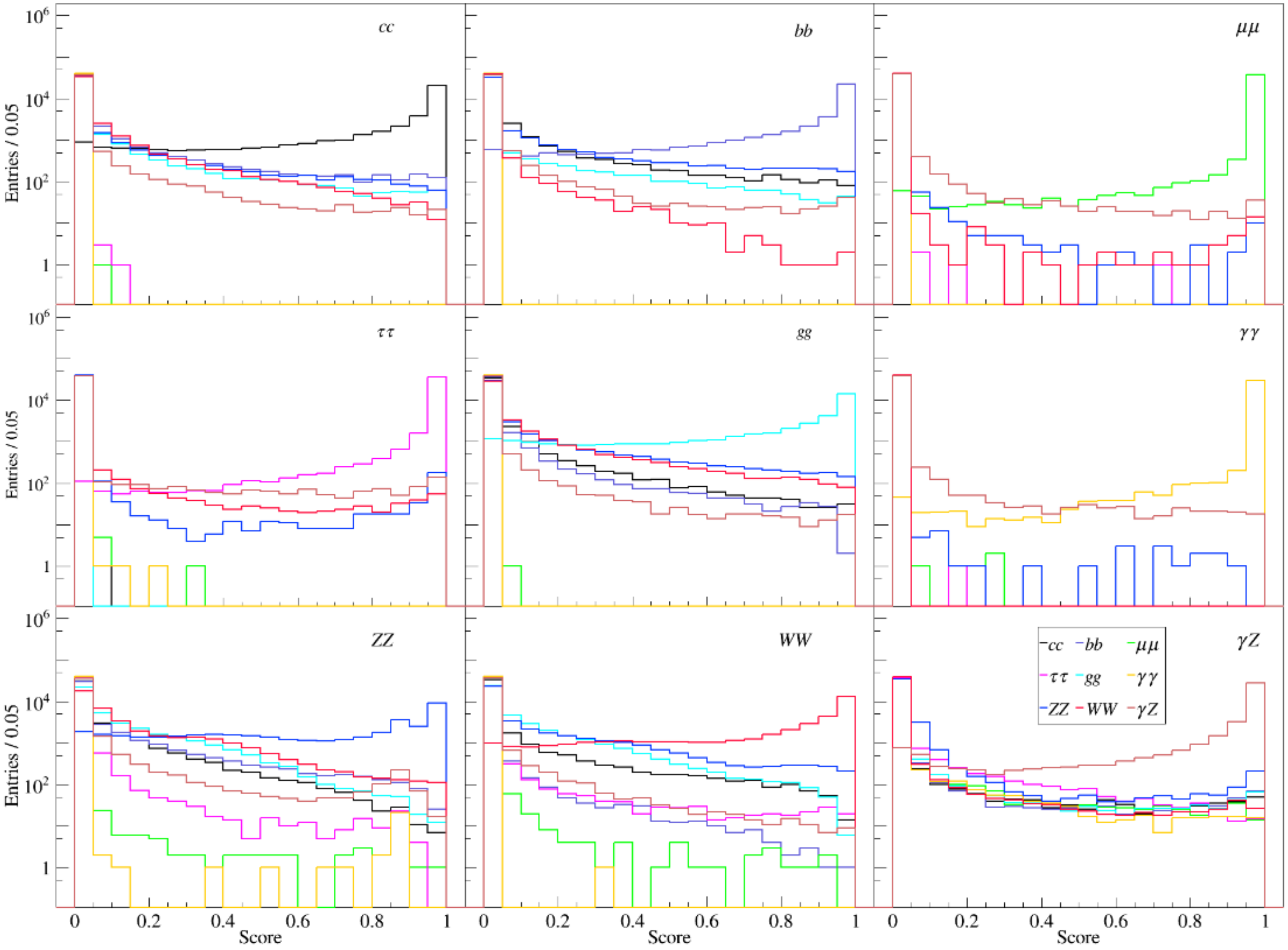
  $cc, bb, \mu\mu, \tau\tau, gg, \gamma\gamma, ZZ, WW, \gamma Z$

- Train: validation: test = 8:1:1

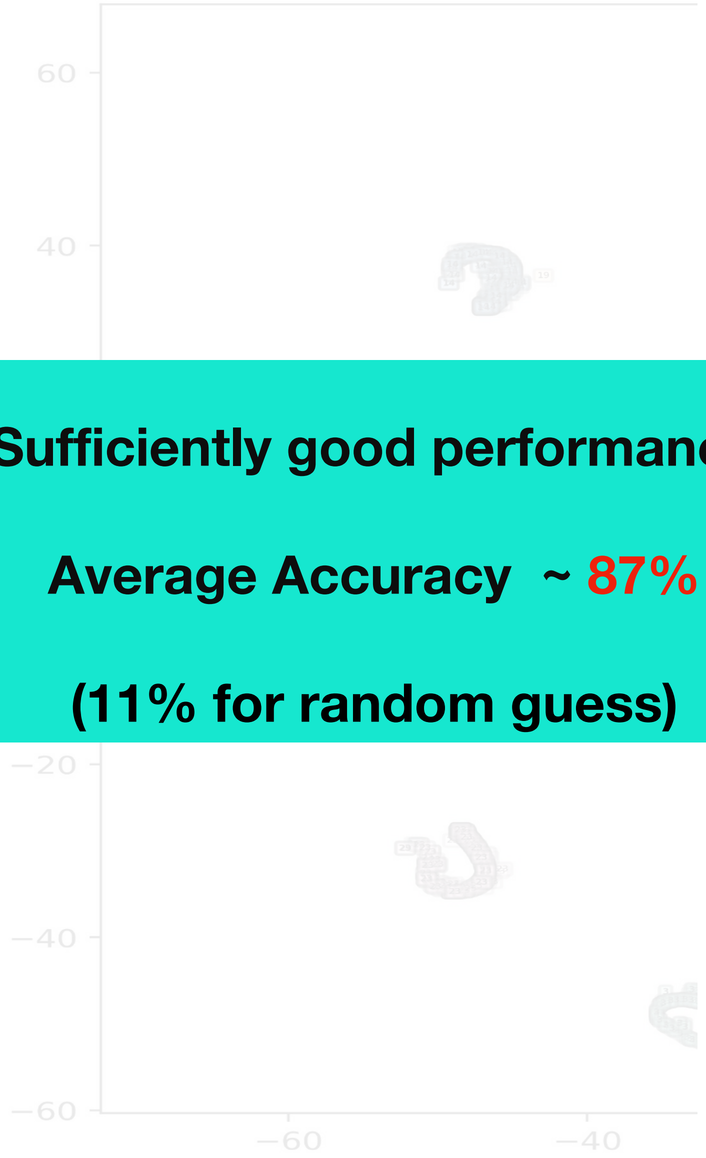
- Simple smearing fast simulation

Try eeH firstly

Probability distributions of each class



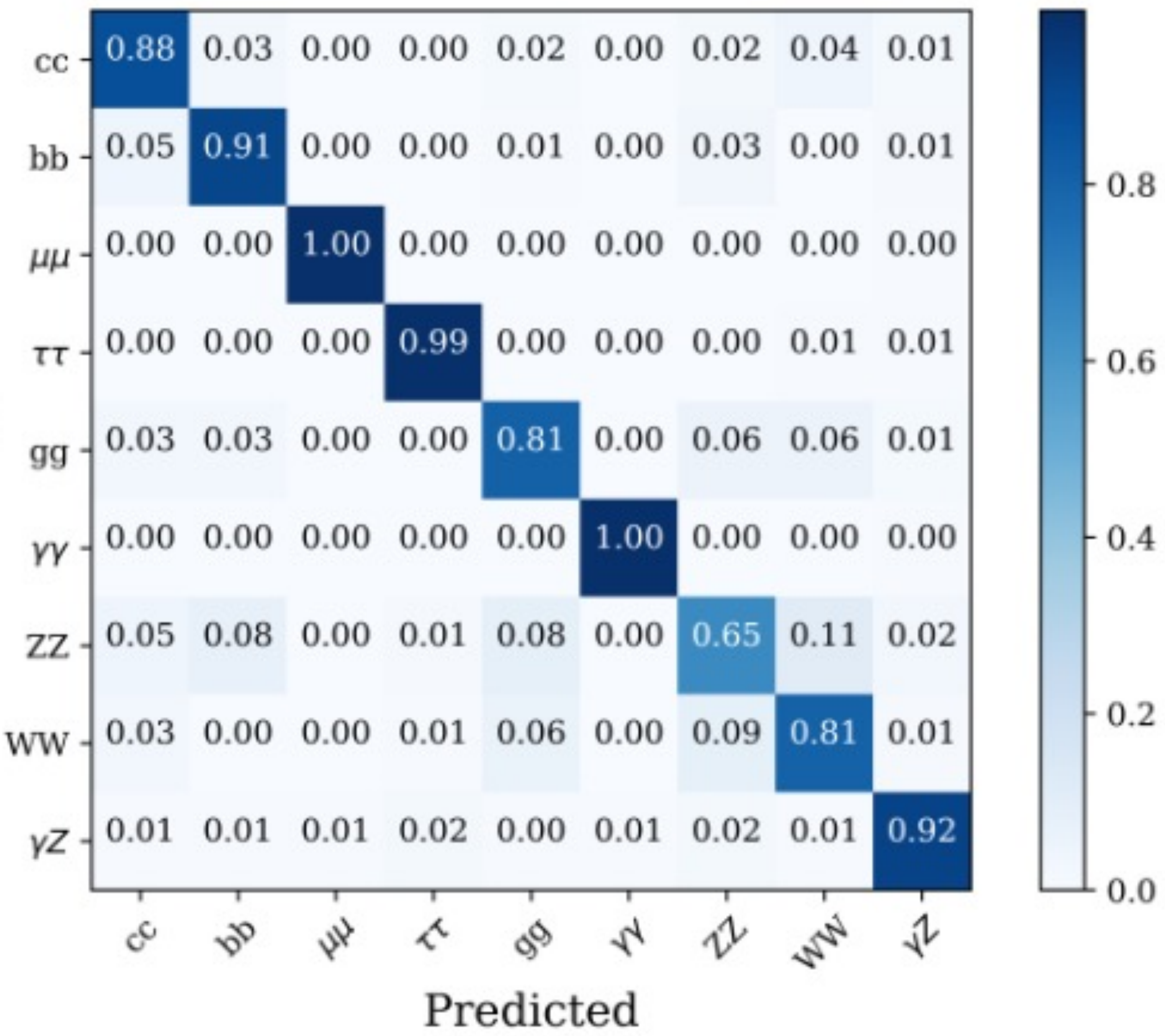
Try eeH firstly



Sufficiently good performance

Average Accuracy ~ 87%

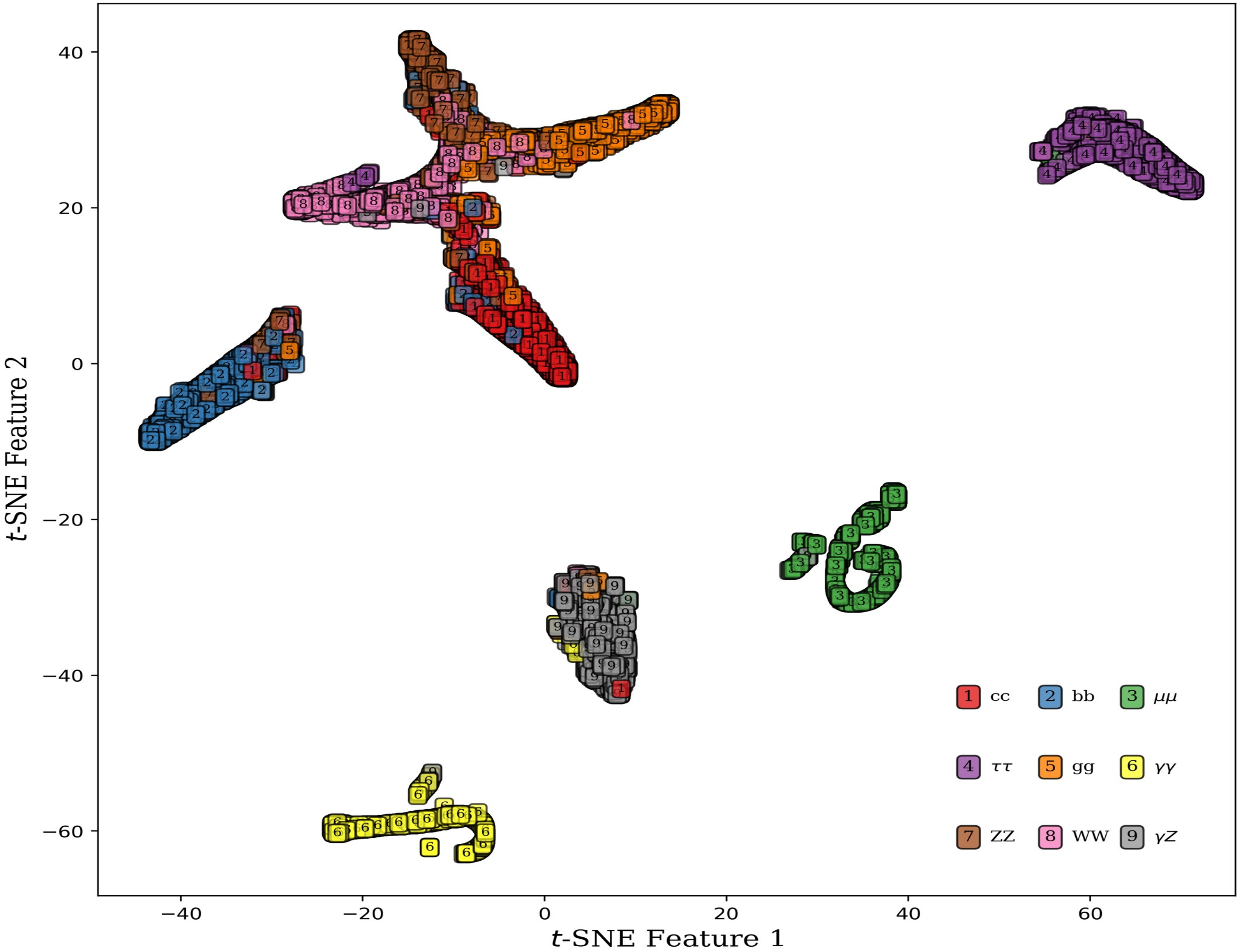
(11% for random guess)



Taking the one of the largest probability (ArgMax)

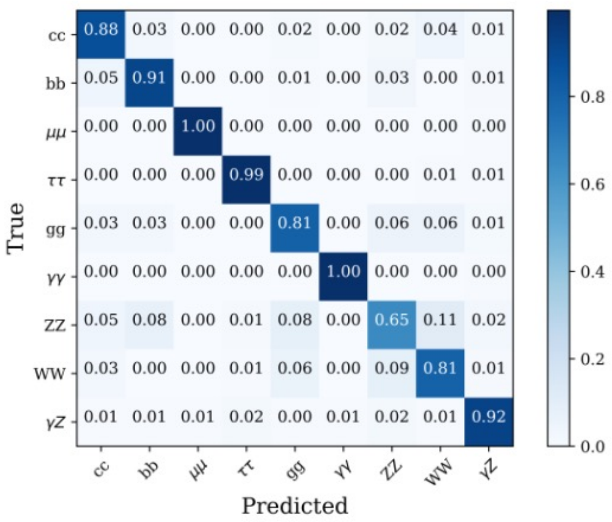
Dimension reduction tells us more

- ✓ μμ, γγ, ττ well classified as expected
- ✓ bb and γZ also good
- ✓ cc, gg, WW, and ZZ fake each other, but under control

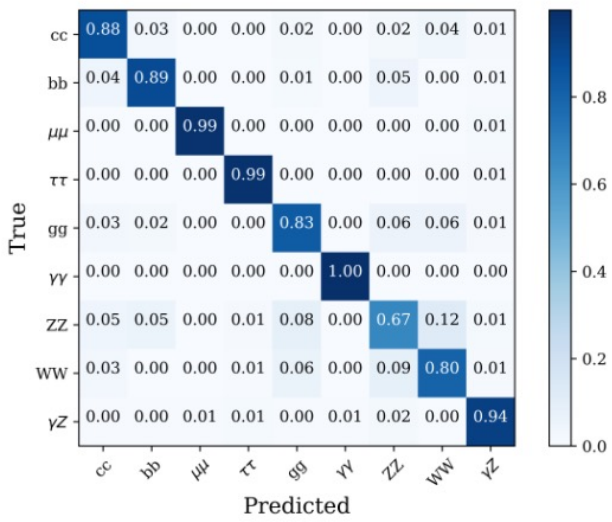


Dimensional reduction (t-SNE)

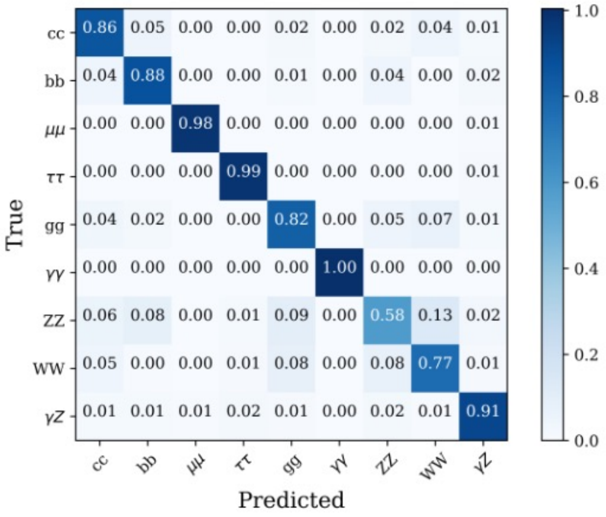
All 4 production modes



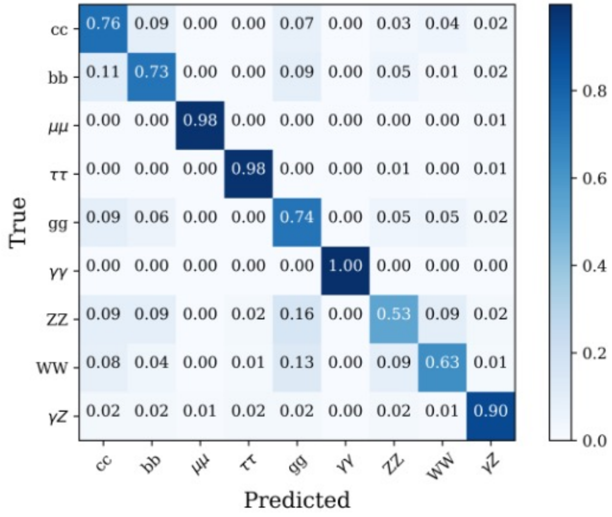
eeH



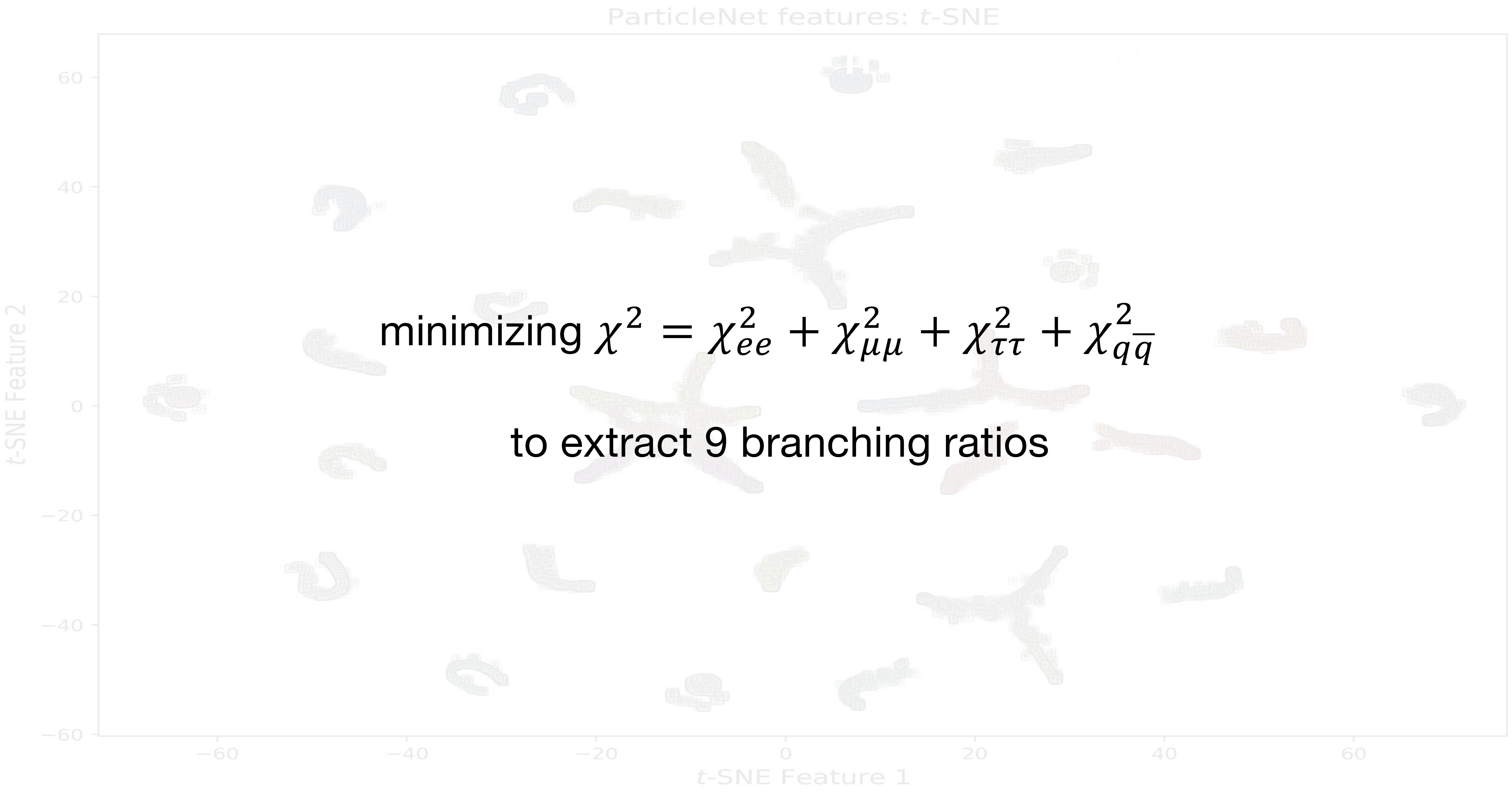
$\mu\mu$H



$\tau\tau$H



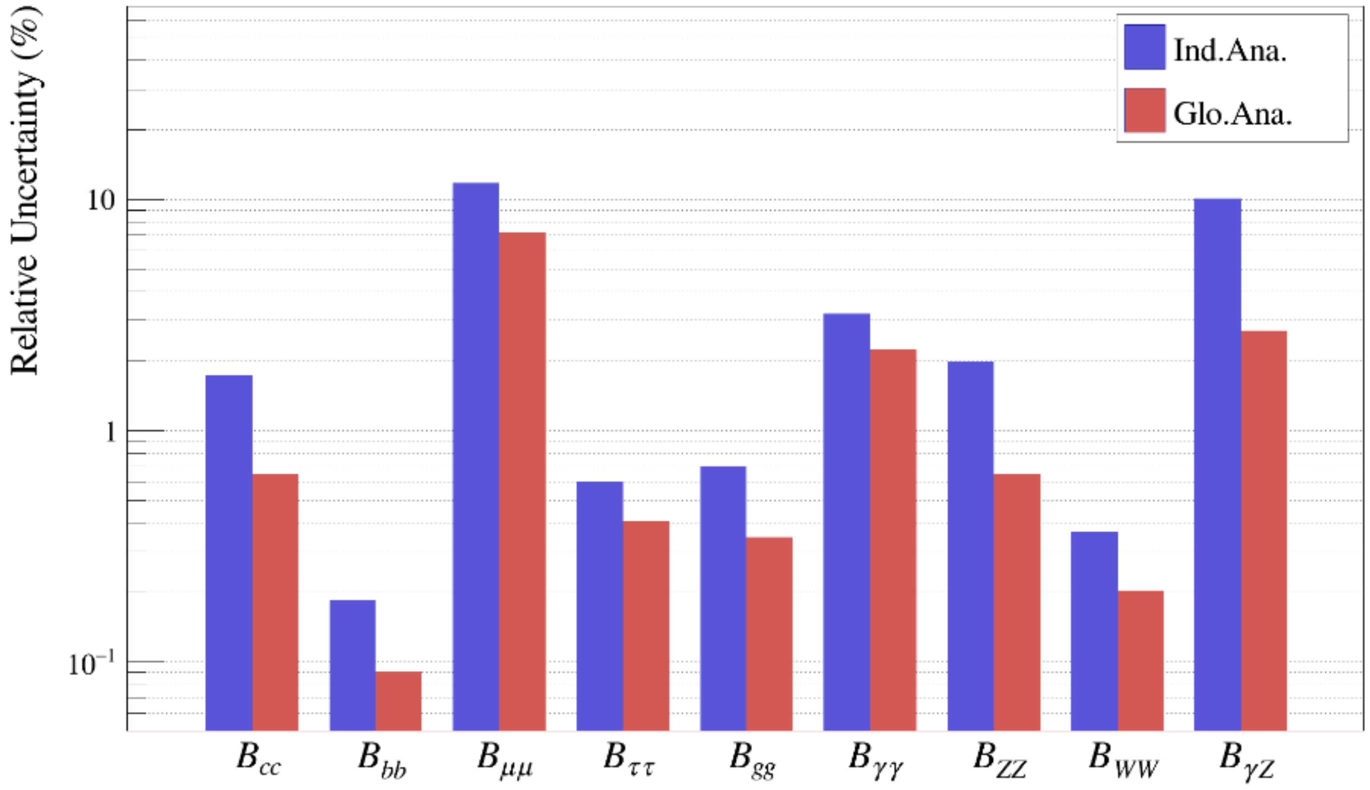
qqH



minimizing $\chi^2 = \chi^2_{ee} + \chi^2_{\mu\mu} + \chi^2_{\tau\tau} + \chi^2_{q\overline{q}}$

to extract 9 branching ratios

Feasibility study results



This is only a comparison to see relative improvement
without backgrounds taken into account

Feasibility study results

Improved roughly by a factor of 2

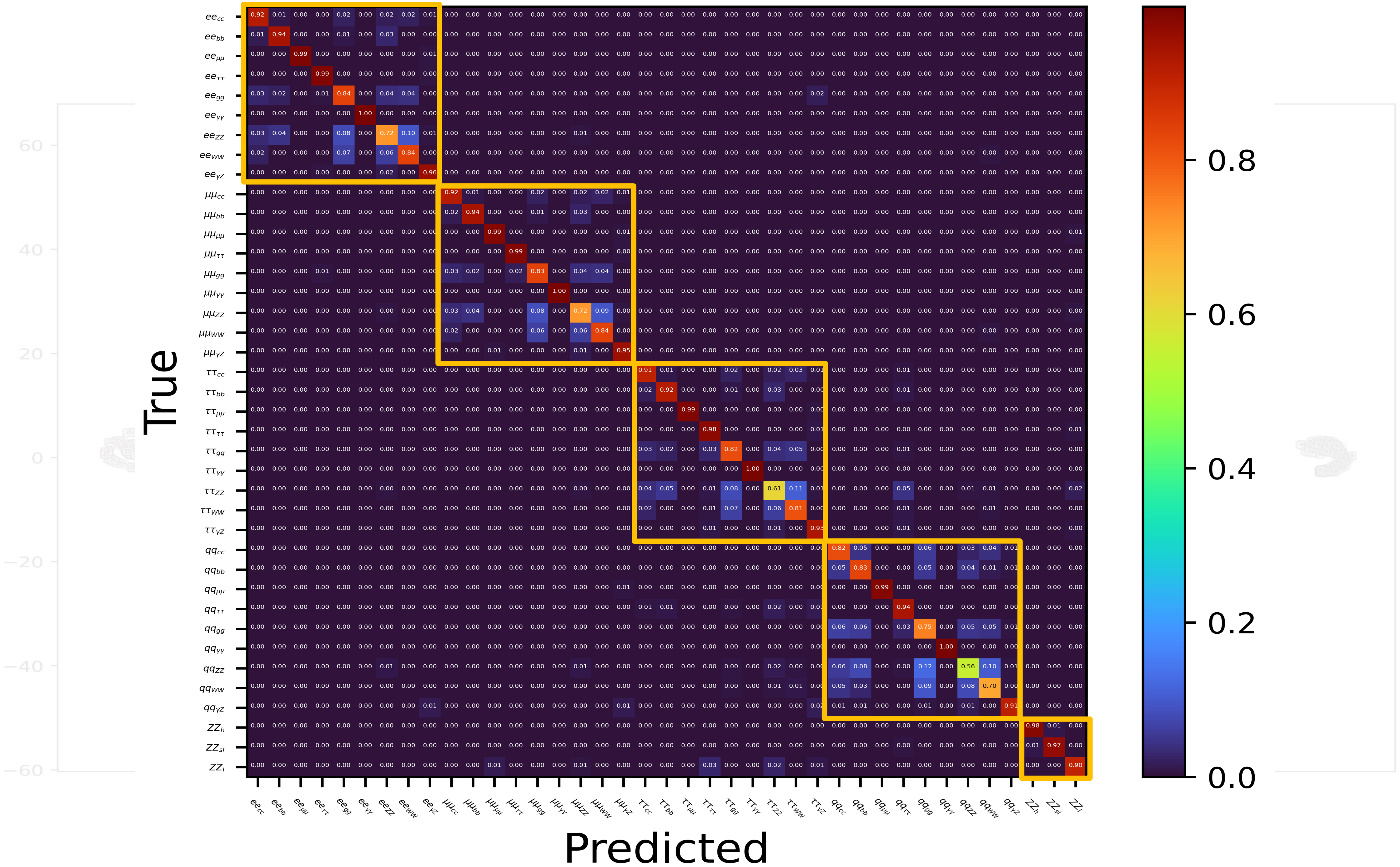
Decay Mode	Ind.Ana.	Glo.Ana.	IP	CEPC CDR
$H \to c\bar{c}$	1.8%	0.65%	2.7	3.3%
$H \to b\bar{b}$	0.19%	0.09%	2.1	0.56%
$H \to \mu^+\mu^-$	12%	7.2%	1.7	17%
$H \to \tau^+\tau^-$	0.61%	0.41%	1.4	1.0%
$H \to gg$	0.7%	0.35%	2.0	1.4%
$H \to \gamma\gamma$	3.3%	2.3%	1.4	6.9%
$H \to ZZ$	2.0%	0.65%	3.0	5.1%
$H \to W^+W^-$	0.37%	0.21%	1.7	1.1%
$H \to \gamma Z$	11%	2.8%	3.9	15%

Improvements

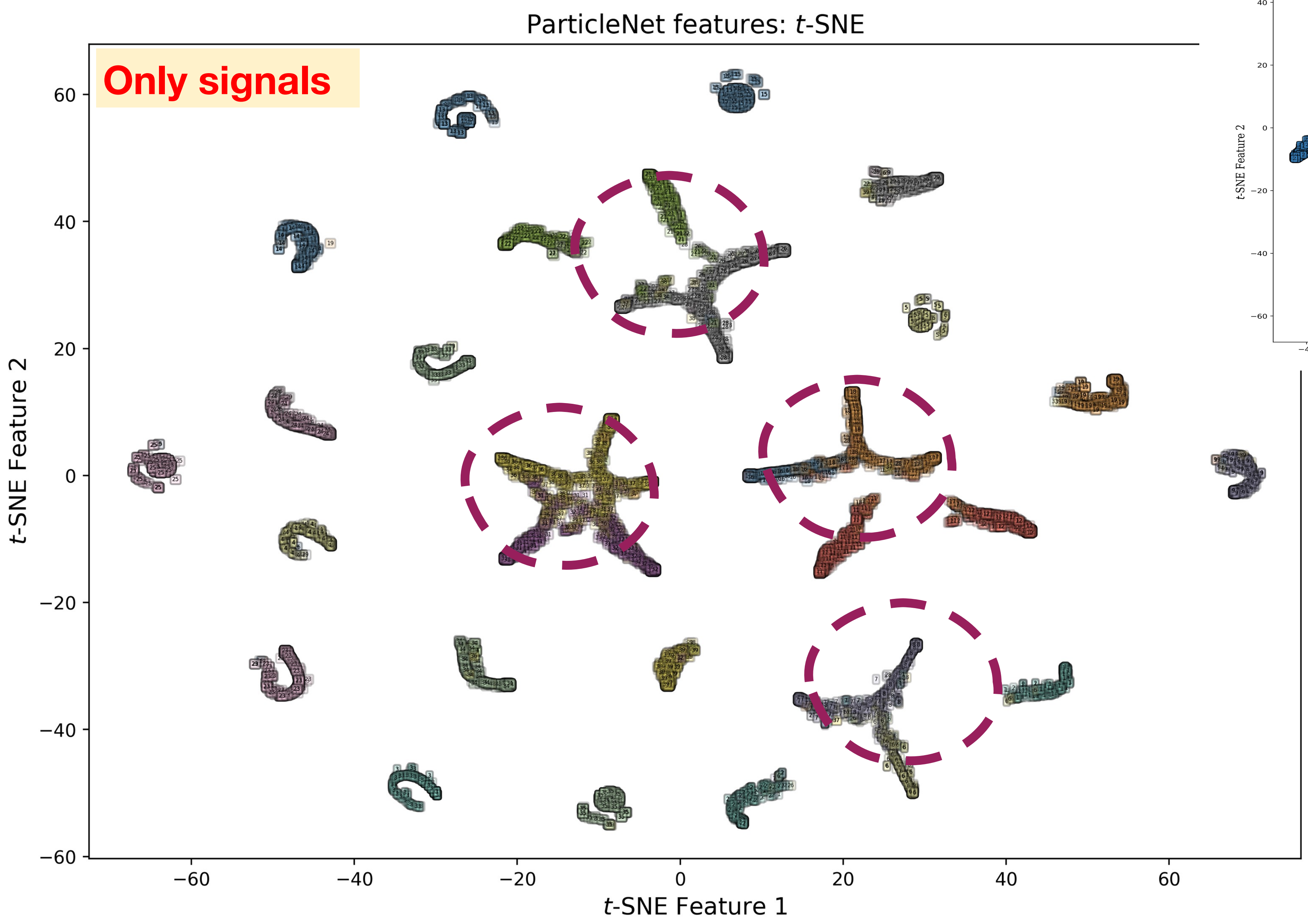
- **Multinomial law: larger Br, smaller stat. uncertainties**
- **Global: the ones with more cross talk benefit from the global constraint of the efficiency matrix**

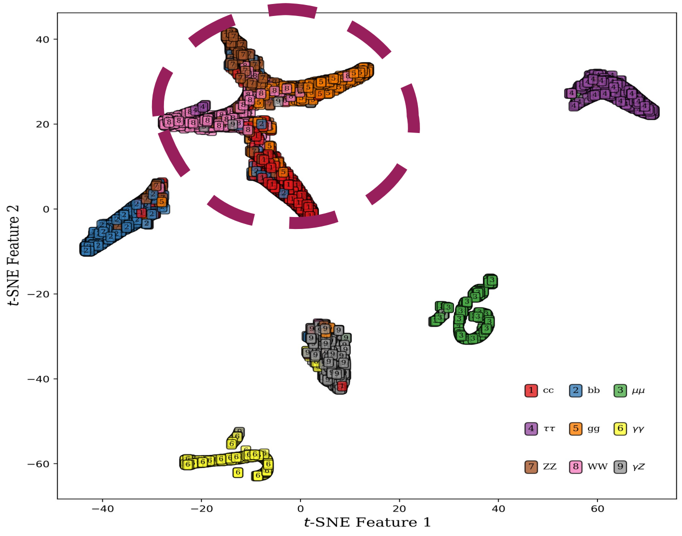
Is it possible to classify all events into even more categories …

Yes!



True
Predicted
0.0
0.2
0.4
0.6
0.8
t-SNE Feature 2

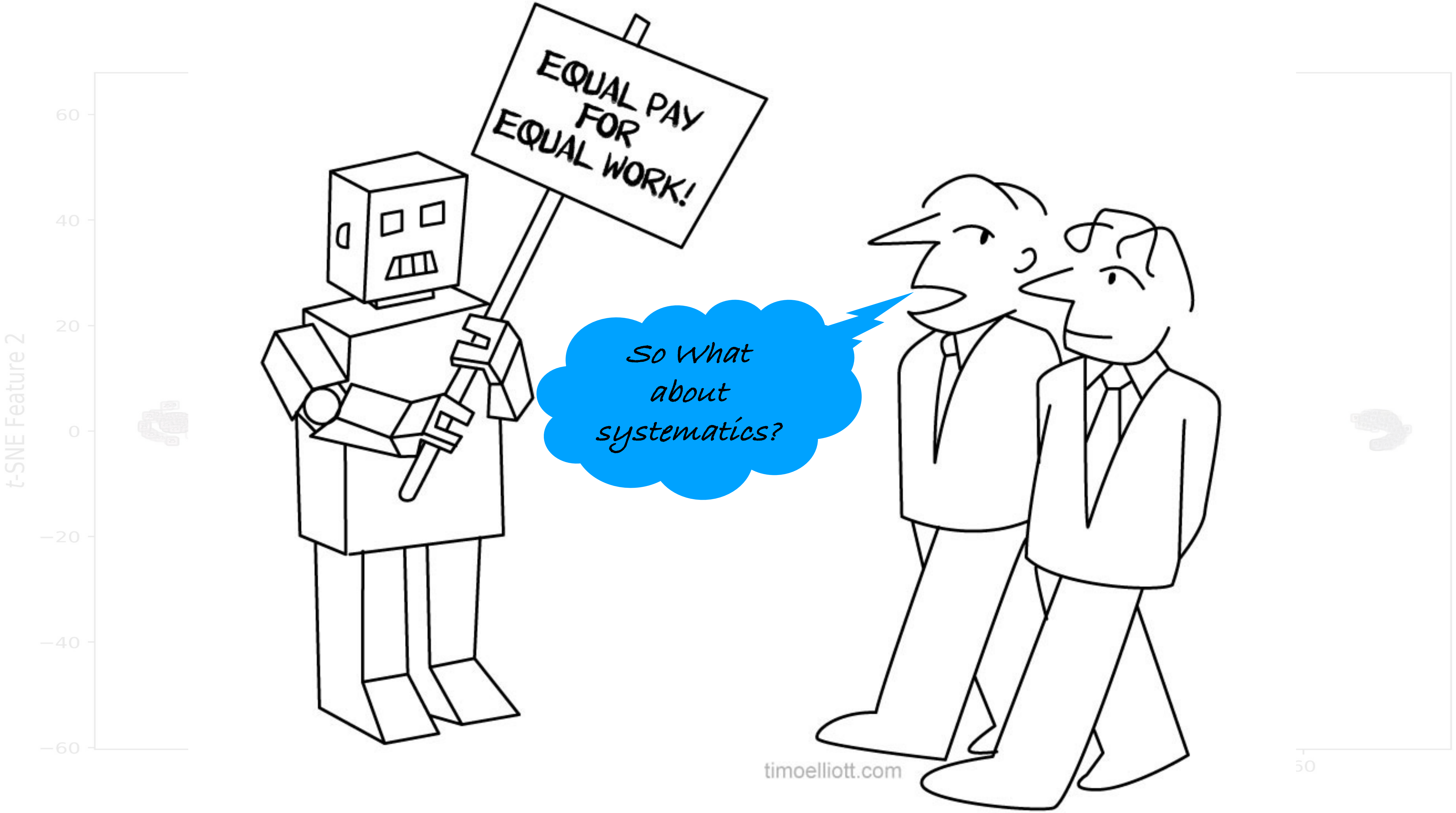




Will add more backgrounds, more statistics, …

Summary

- ➢ Higgs physics is the first prior in future Higgs factories
- ➢ Feasibility study shows that the ParticleNet could support ‘one-stop’ analysis of many physics processes and improve the precision
- ➢ A new analysis paradigm
  - ✓ Multi-classification: accurate enough and fast enough
  - ✓ Only particle level information as input, no dependence on jet algorithm, … less software work
  - ✓ Systematics will be very challenging, need more study


EQUAL PAY FOR EQUAL WORK!
So What about systematics?
timoelliott.com
t-SNE Feature 2
60
40
20
0
−20
−40
−60

Extras

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

Obvious,
the determinant, |E|, is a good one

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

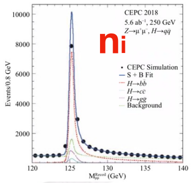
Problem successfully becomes
---- a single purpose optimization

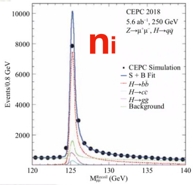
- Look at Higgs decays
  - **13 decays in Standard Model**

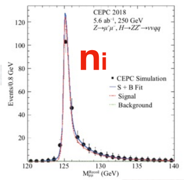
$$\sigma_N = \sqrt{N \times p} \quad \textit{for Poisson}$$

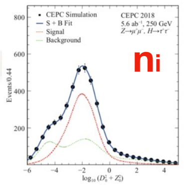
$$\sigma_N = \sqrt{N \times p \times (1-p)} \quad \textit{for multinomial}$$

B_i =

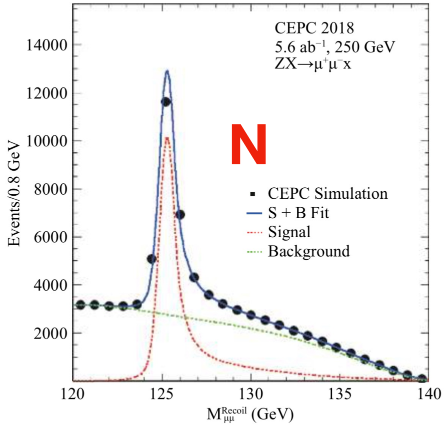








+ … …



non-Higgs background

— subtracted with fitting for other method