The impact of flavor tagging performance on nnH(H to bb, cc) accuracy measurement

Zhu Yongfeng

Motivation & Content :

The g(Hcc), the second generation fermion Yukawa coupling, is one of the most important Higgs property at the CEPC.

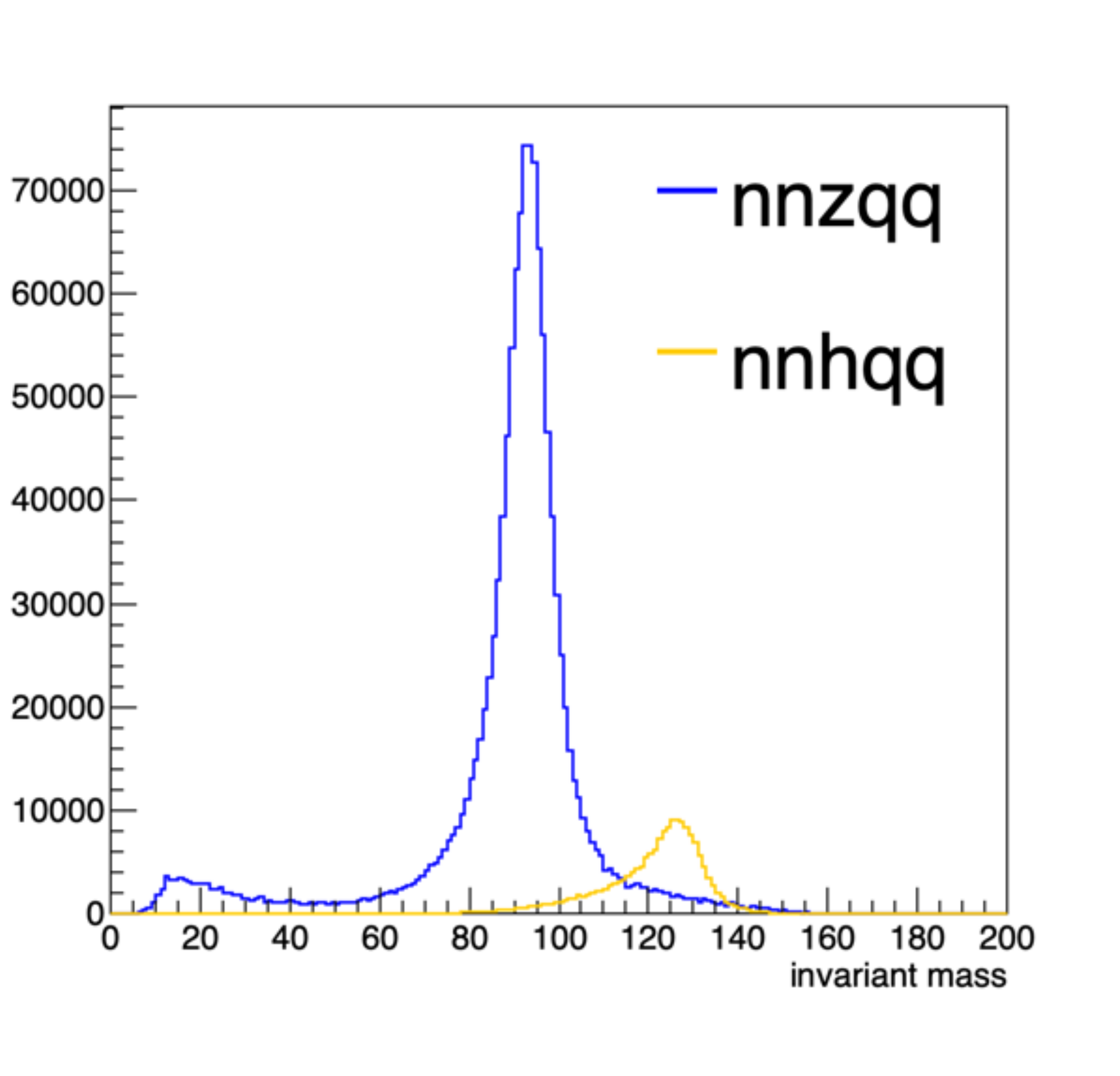
The g(Hcc) measurement highly depends on the detector performance of Flavor tagging.

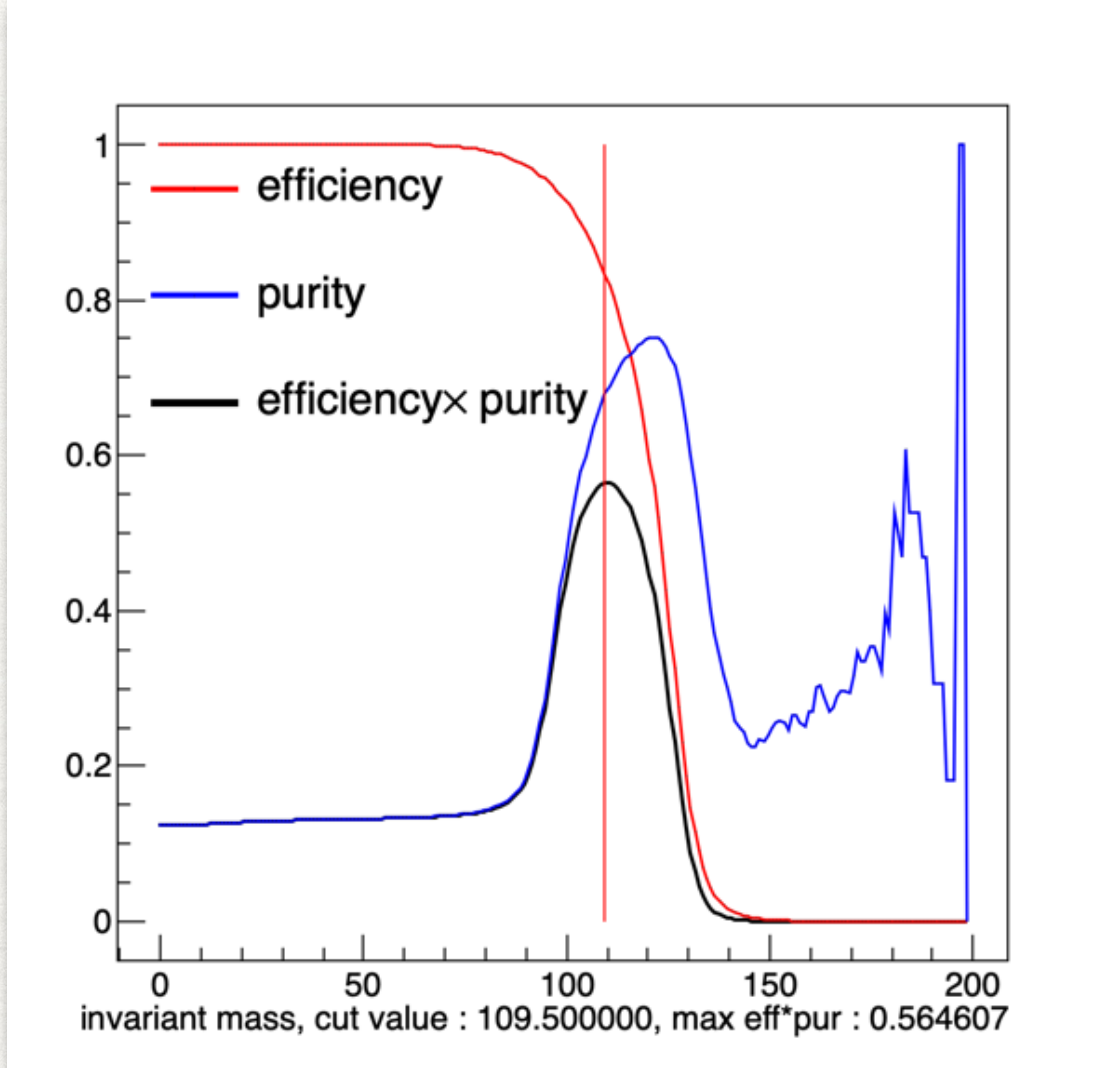
The flavor tagging performance, as part of the detector requirement for the CEPC detector, we study its impact on the g(Hcc) measurement.

decay mode	$H \to b\bar{b}$	$H \to c\bar{c}$	$H \to gg$
branching ratio	57.7%	2.91%	8.57%

Samples : nnH with Higgs to $q\bar{q}$ and gg

nnZ with Z to $q\bar{q}$



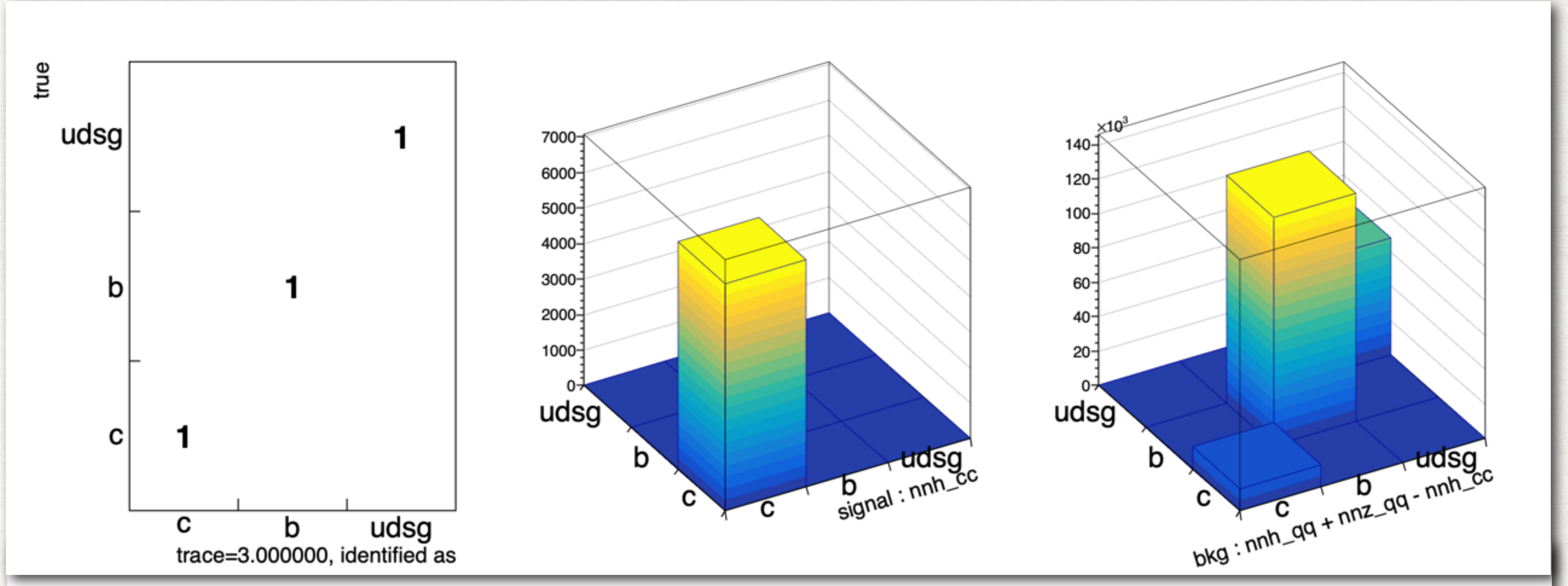


after setting invariant mass >= 110GeV,

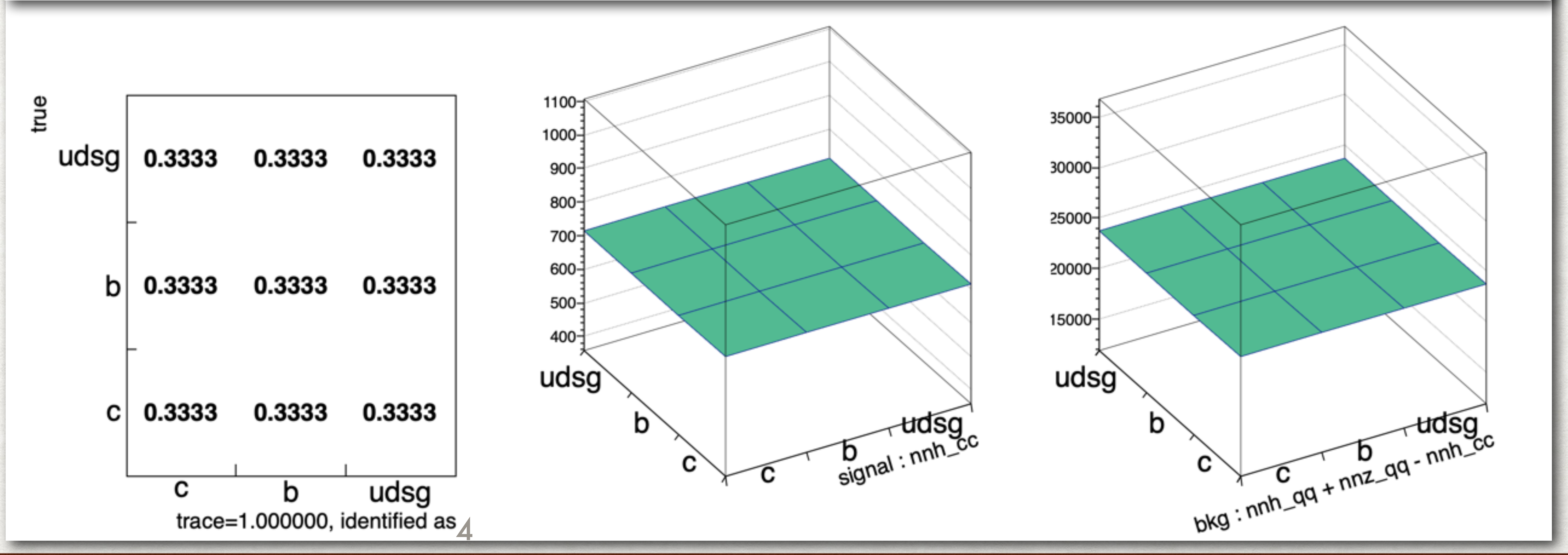
nnH left : 124246/179186 =69.34%, nnZ left : 68426/1.25355×10^6= 5.46%

Flavor Tagging modeling: Migration Matrix

perfect flavor tagging

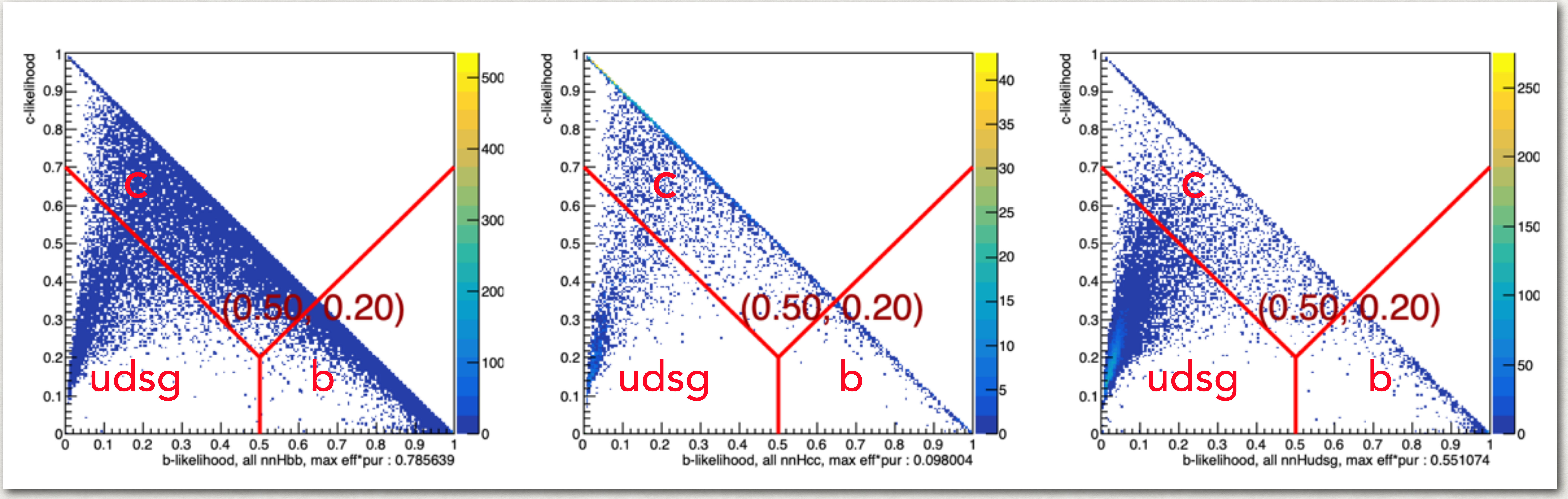


non flavor tagging

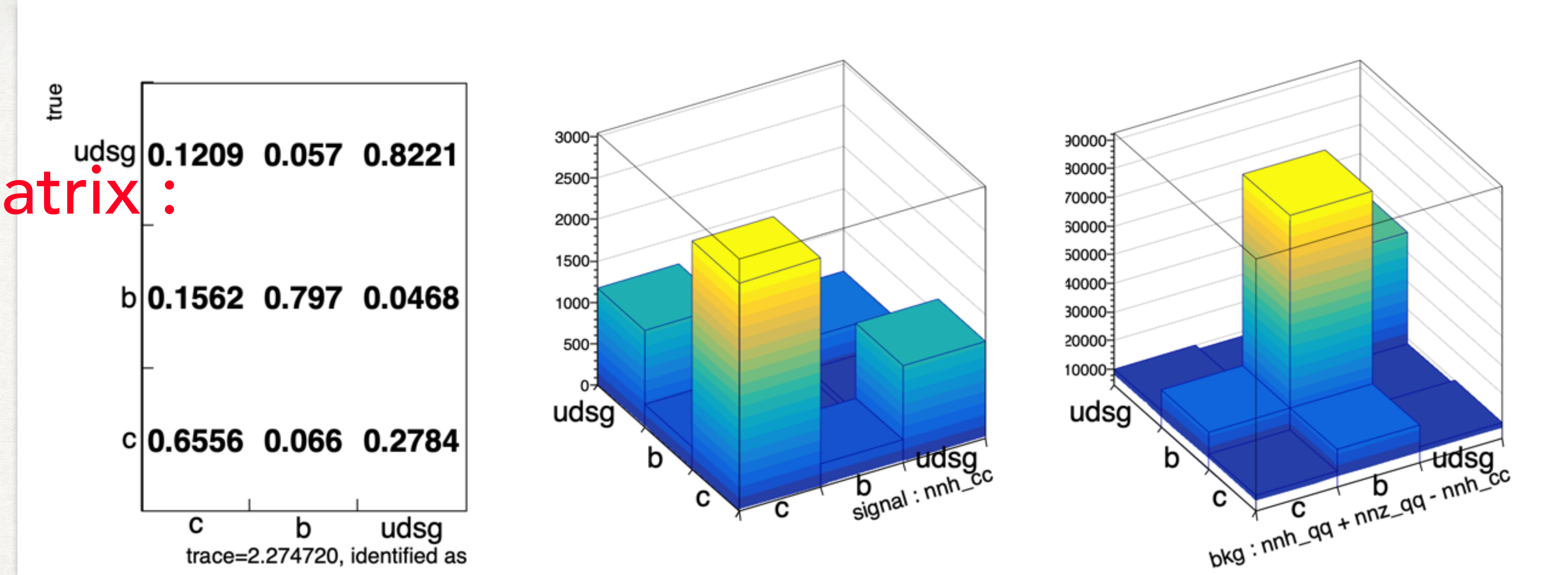


Actual Migration Matrix of the baseline detector :

With all the nnH(H to bb, H to cc, H to udsg) samples, divide flavor space into three parts and find the suitable division way, which can maximize the value of $eff(c \to c) + eff(b \to b) + eff(udsg \to udsg)$.



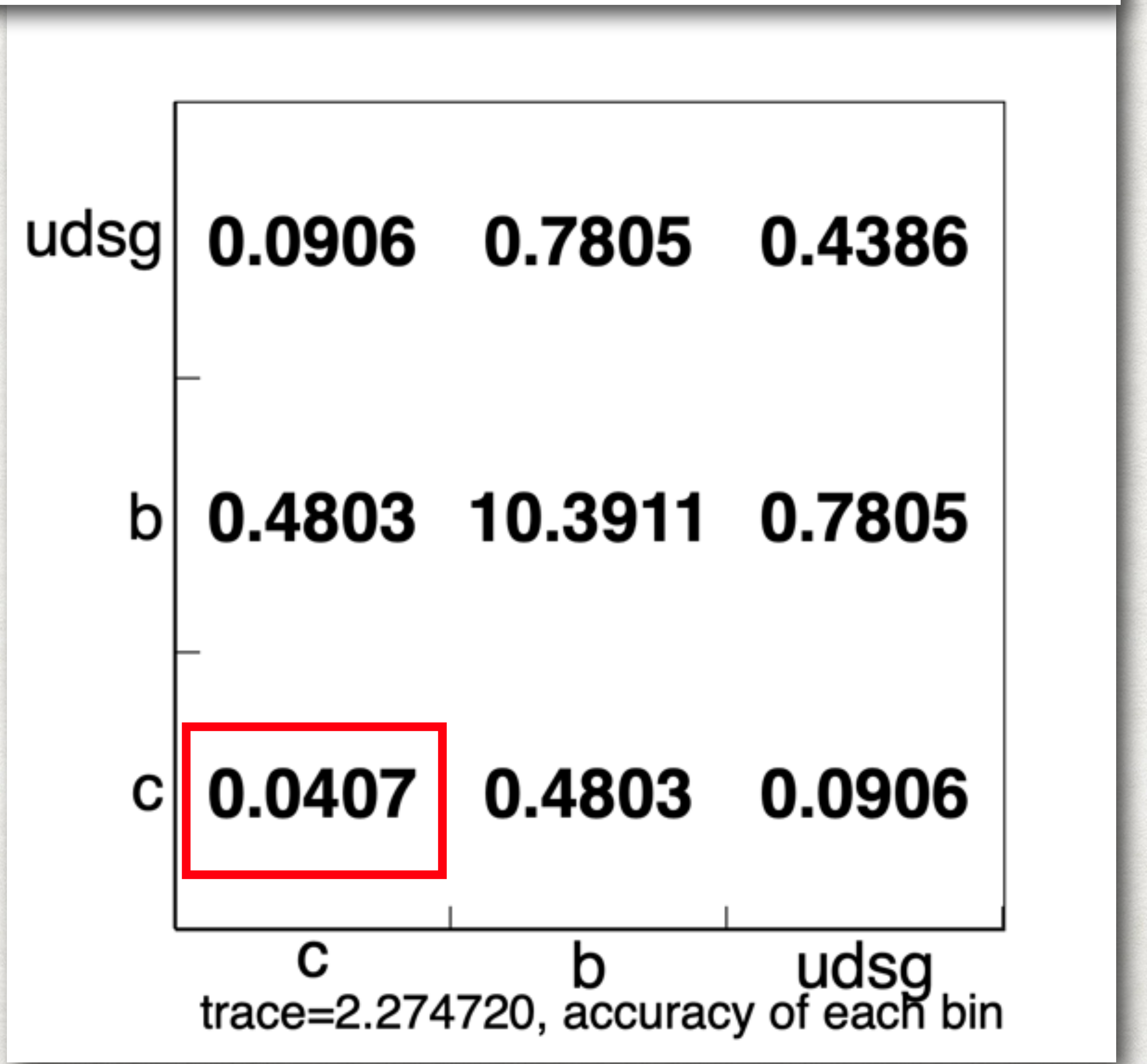
optimized matrix :



accuracy : $\frac{\sqrt{S+B}}{S}$

maximum accuracy : bin with minimum accuracy value

combined accuracy : $\frac{bin_1 \cdot bin_2}{\sqrt{bin_1 \cdot bin_1 + bin_2 \cdot bin_2}}$ iterate for each pair of bins



non flavor tagging $\longrightarrow$ perfect flavor tagging

the changing procedure of flavor tagging performance matrix :

$$temp\ matrix = \frac{x - trace_I}{trace_T - trace_I} \cdot (T - I) + I \qquad (trace_I \leq x \leq trace_T)$$

T : matrix with perfect flavor tagging

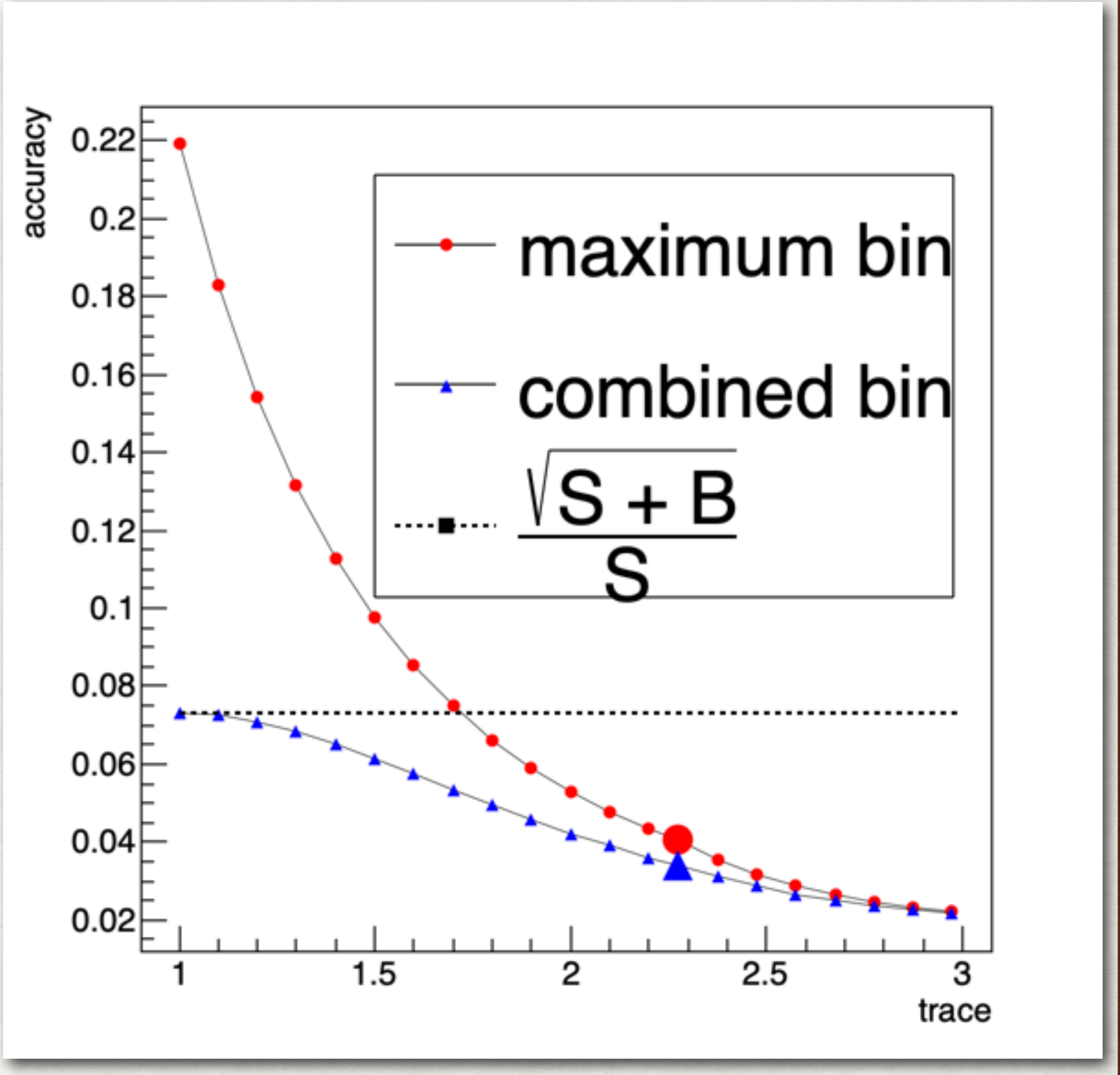
I : matrix with non flavor tagging

$trace_I, trace_T$: the trace of matrix I and T

The impact of flavor tagging performance on nnHcc accuracy measurement, the maximum bin is always cc.

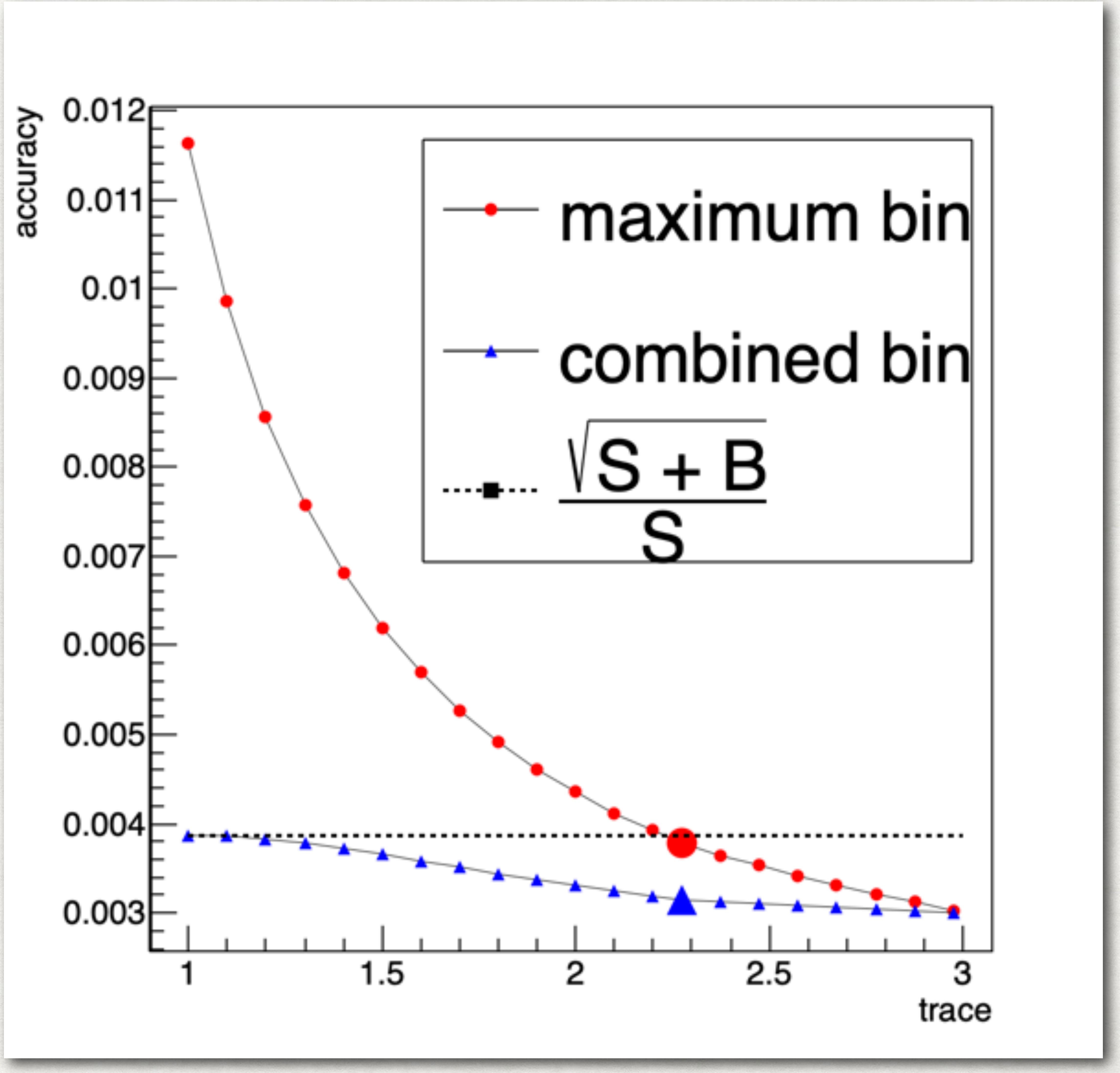
The red circle and blue triangle represents the performance of baseline detector.

trace	2.1	2.2	2.27	2.37	2.47	2.57	2.67	2.77	2.87	2.97
maximum bin	4.78%	4.35%	4.07%	3.56%	3.18%	2.88%	2.65%	2.47%	2.32%	2.20%
combined bin	3.90%	3.60%	3.40%	3.11%	2.86%	2.67%	2.51%	2.38%	2.28%	2.19%



With the similar analysis method, see nnH_bb as signal, the maximum bin is always bb.

trace	2.1	2.27	2.97
maximum bin	0.41%	0.37%	0.30%
combined bin	0.32%	0.31%	0.30%



Conclusion :

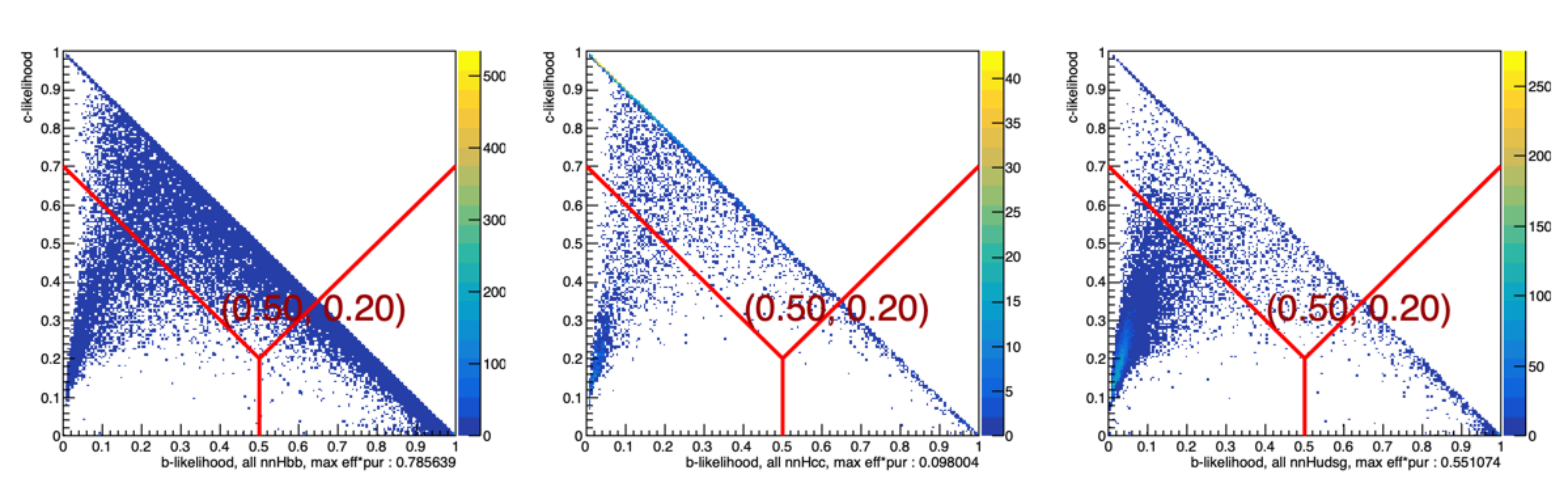
- The flavor tagging performance has great impact on nnH(H to cc, bb) accuracy measurement. At the CEPC baseline detector, the nnH with Higgs to cc measurement accuracy is 4.07%, the nnH with Higgs to bb measurement accuracy is 0.37%.
- The $\mu\mu$H and eeH(H to cc, bb) accuracy measurement also can be done, since the main background is ZZ with one Z to $ee\ or\ \mu\mu$ and the other Z to qq, this background can be efficiently suppressed by the invariant mass and recoil mass of two leptons.(page 11) It's our next step.

LLH(H to bb, cc) accuracy measurement also can be done

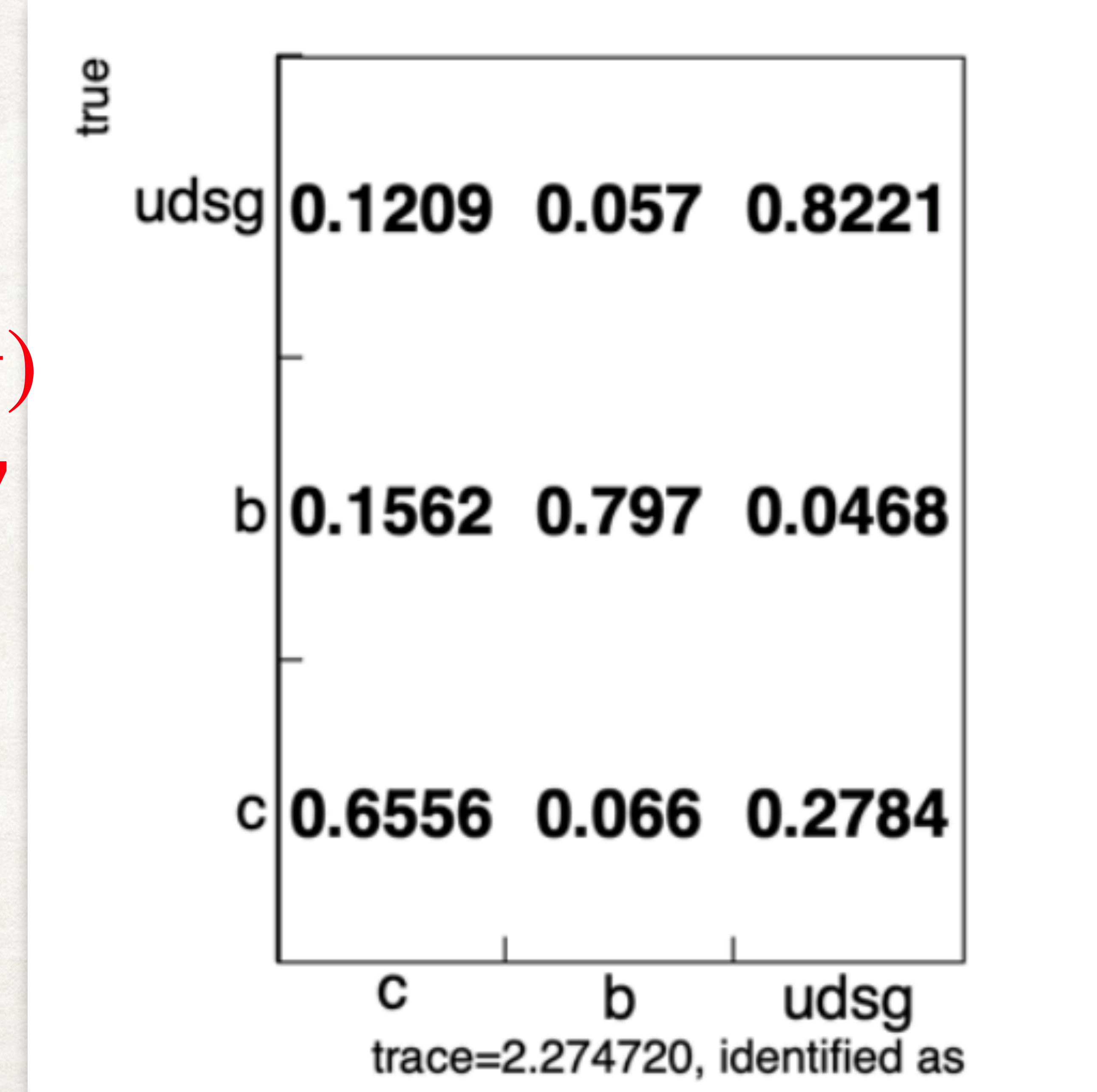
Table 1. Event yields of cut flow. Signal events are $l\bar{l}+H \rightarrow l\bar{l}+b\bar{b}/c\bar{c}/gg$ combined. $\mu^+\mu^- H$ and $e^+e^- H$ background refers to the background which Higgs are produced associated with $\mu^+\mu^-$ and e^+e^-, but decay to final states other than $b\bar{b}/c\bar{c}/gg$. 'Other Higgs background' stands for the Higgs production process different from the signal. 'Irreducible SM background' is the $e^+e^-/\mu^+\mu^-$+jet pair process without Higgs productions. 'Other SM background' includes all the other background processes. The 'fit region' will be described in Section 5.

$\mu^+\mu^- H \rightarrow \mu^+\mu^- + b\bar{b}/c\bar{c}/gg$ channel					
	signals	$\mu^+\mu^- H$ background	other Higgs background	irreducible background	other SM background
original	2.45×10^4	1.10×10^4	1.01×10^6	1.05×10^6	4.96×10^8
lepton pair selection without recoil mass cut	1.51×10^4	6.56×10^3	227	1.09×10^4	2.79×10^4
jets pair selection and lepton pair recoil mass cut for fit region	1.32×10^4	1.80×10^3	108	7.75×10^3	43.6
signal region	1.31×10^4	1.80×10^3	96.1	5.78×10^3	38.4
$e^+e^- H \rightarrow e^+e^- + b\bar{b}/c\bar{c}/gg$ channel					
	signals	$e^+e^- H$ background	other Higgs background	irreducible background	other SM background
original	2.63×10^4	1.17×10^4	1.01×10^6	1.62×10^6	4.95×10^8
lepton pair selection without recoil mass cut	9.17×10^3	3.53×10^3	128	9.00×10^3	7.11×10^4
jets pair selection and recoil lepton pair mass cut of fit region	7.14×10^3	917	56.1	8.63×10^3	69.4
signal region	7.13×10^3	913	36.4	4.14×10^3	67.4

BackUp

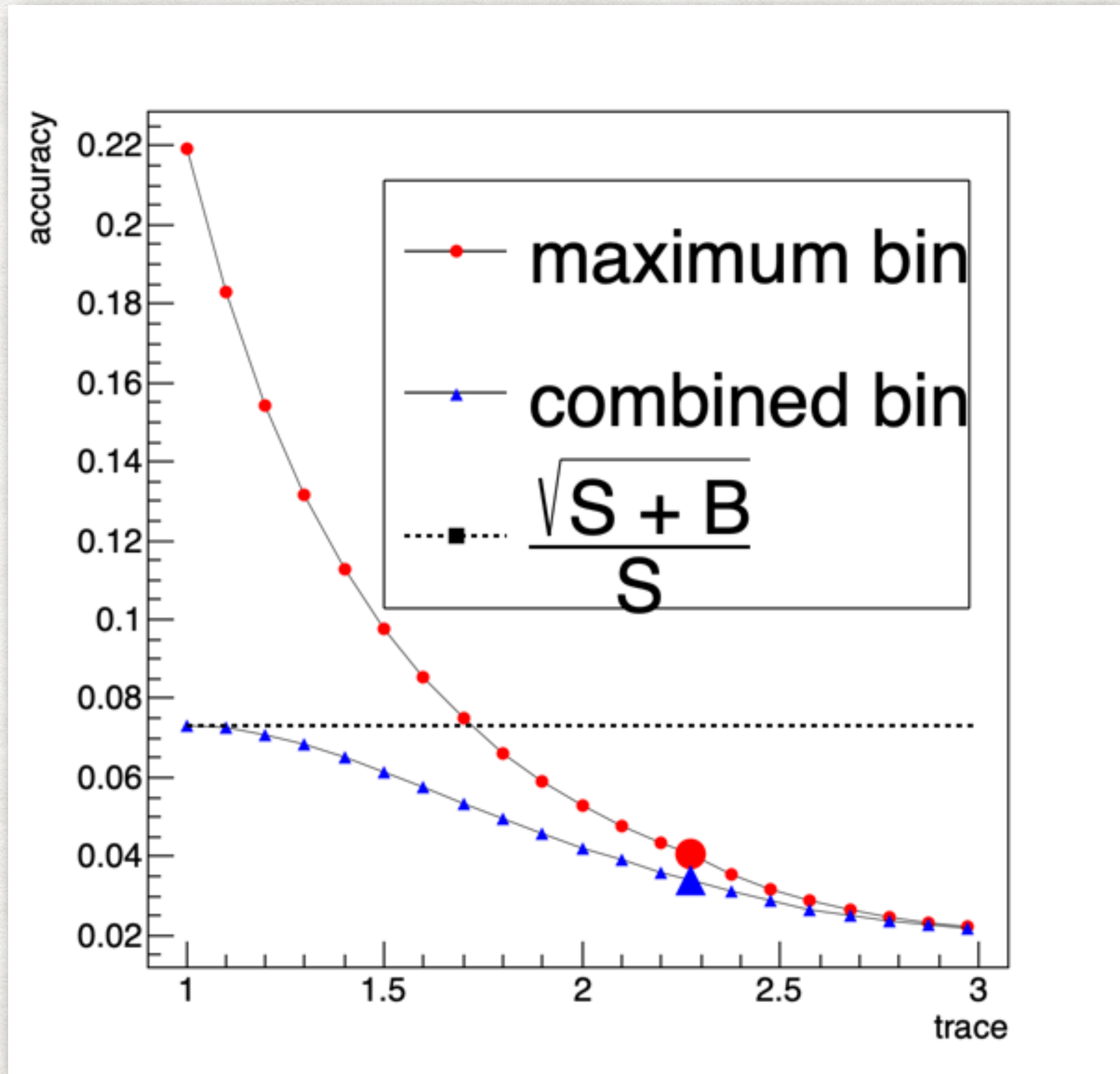


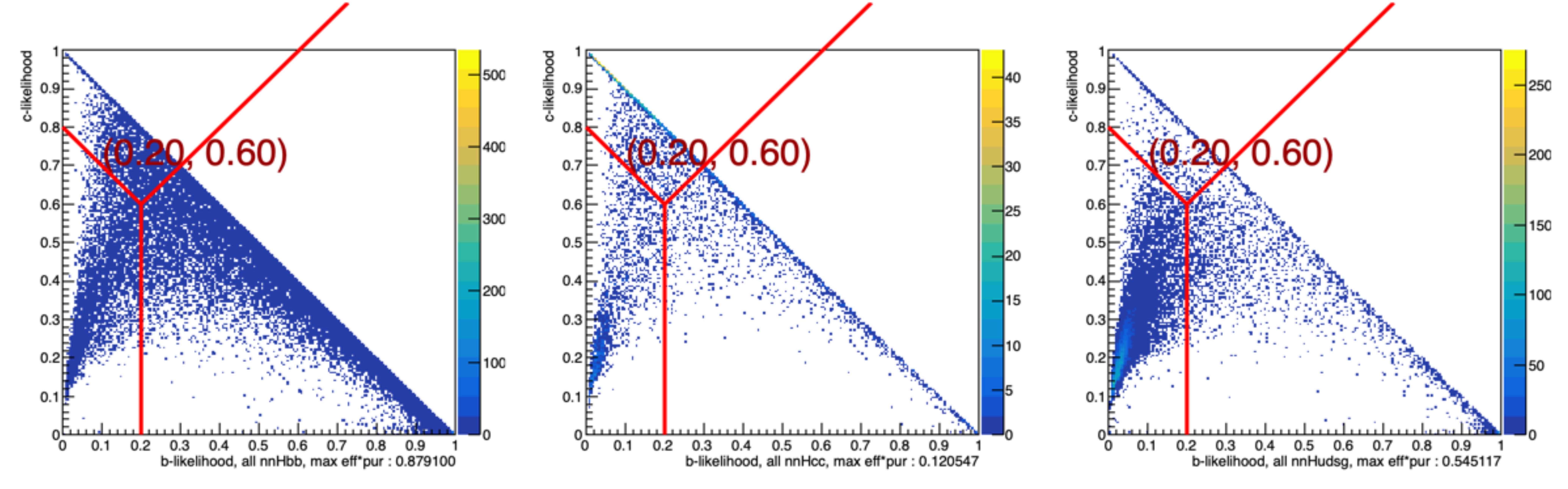
将$eff(c \to c) + eff(b \to b) + eff(udsg \to udsg)$作为优化目标, maximum accuracy is 0.0407



true \ identified as	c	b	udsg
udsg	0.1209	0.057	0.8221
b	0.1562	0.797	0.0468
c	0.6556	0.066	0.2784

trace=2.274720, identified as

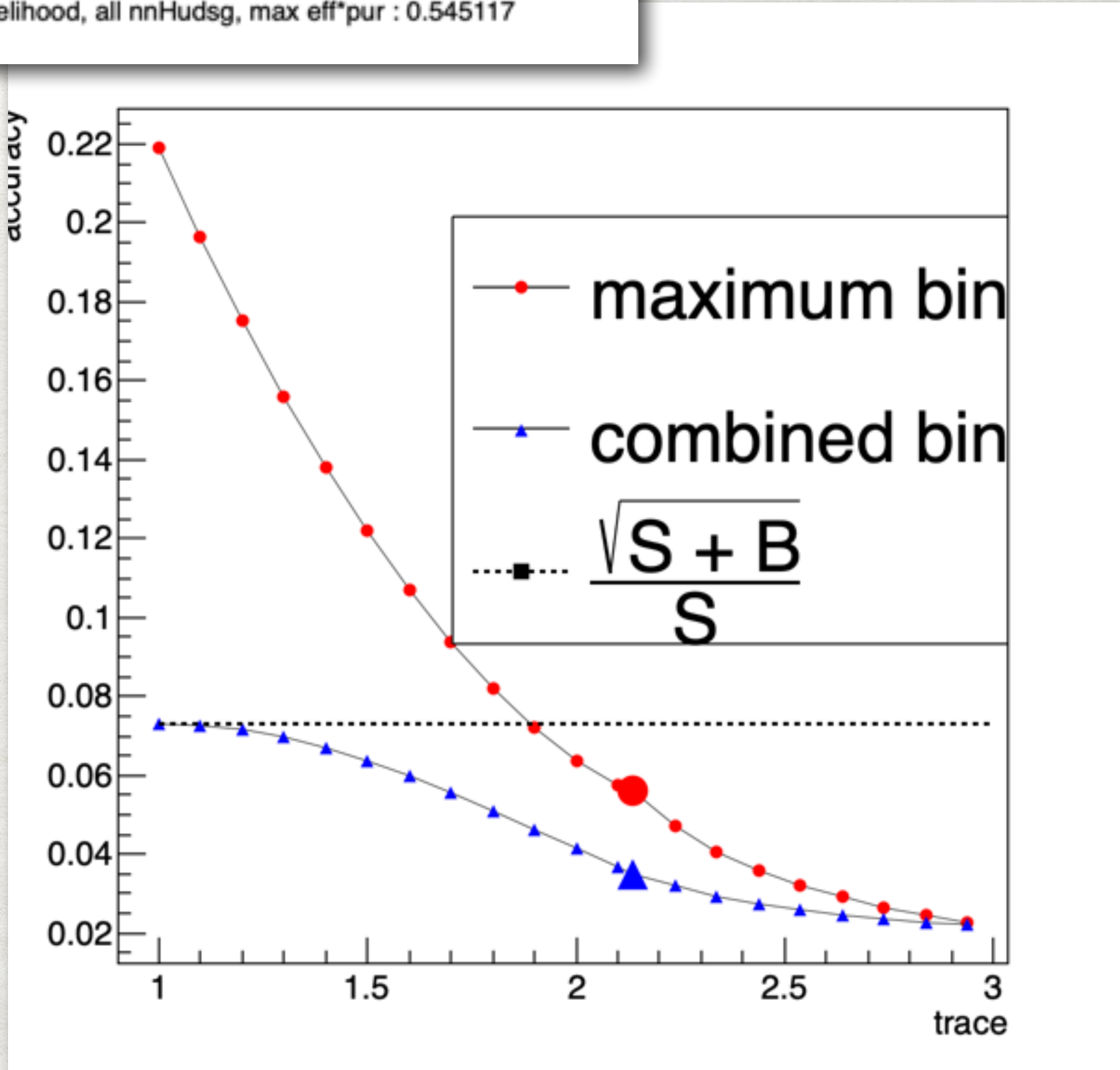


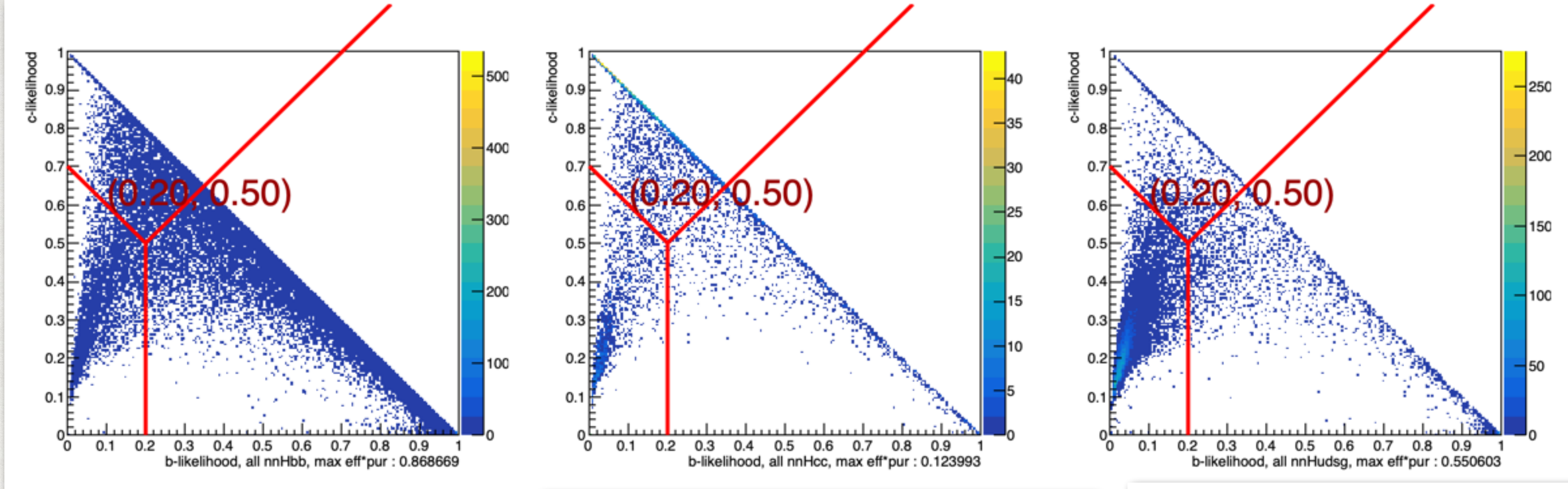


将 $\sum_{b,c,udsg} eff \times purity$ 作为优化目标, maximum accuracy is 0.0561091.

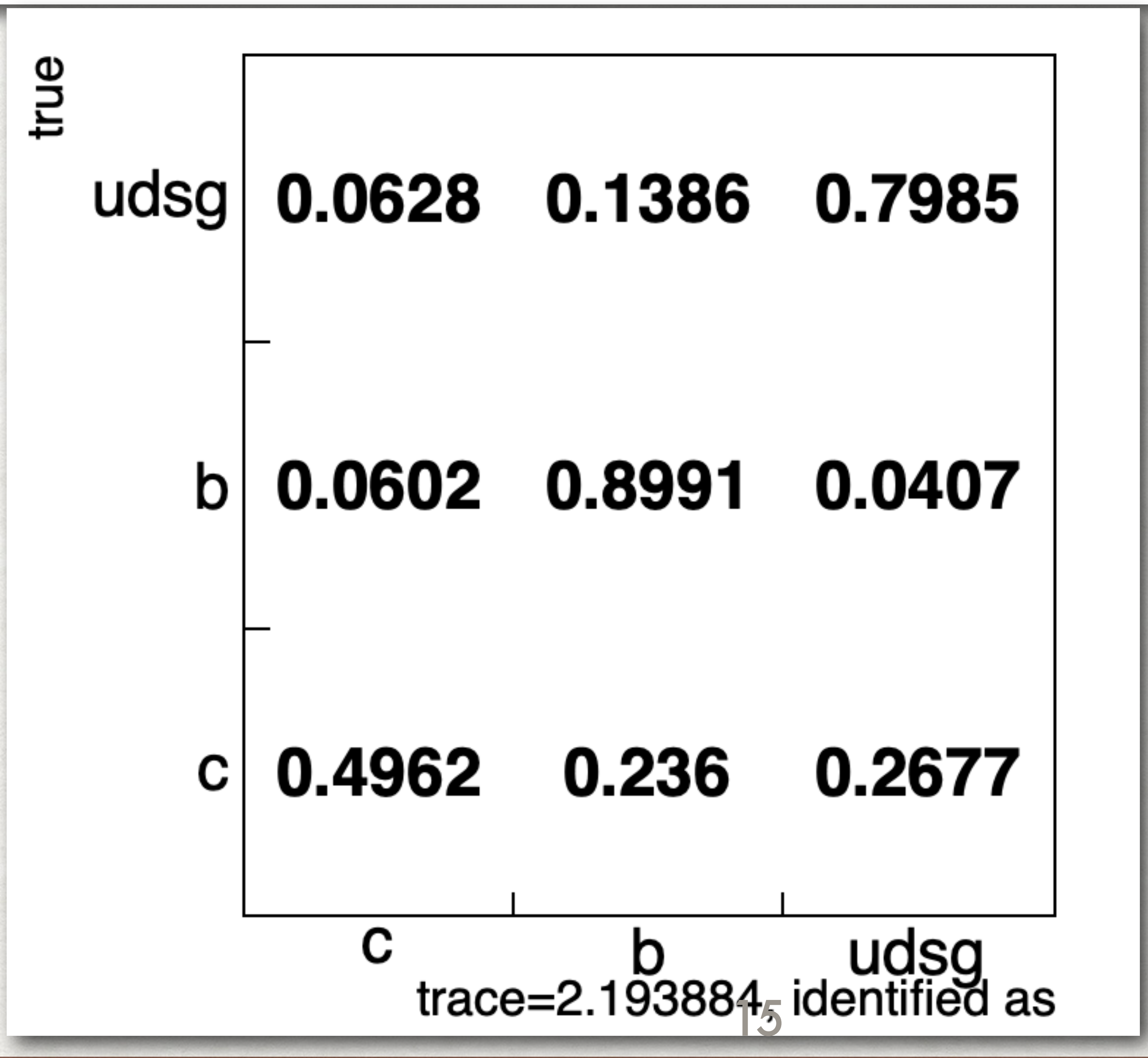
true \ identified as	b	c	udsg
udsg	0.151386	0.0298766	0.818738
c	0.286596	0.404868	0.308536
b	0.913505	0.0401514	0.0463433

tri=2.137111, identified as





将 $\prod_{b,c,udsg} eff \times purity$作为优化目标, maximum accuracy is 0.0472265,



true \ identified as	c	b	udsg
udsg	0.0628	0.1386	0.7985
b	0.0602	0.8991	0.0407
c	0.4962	0.236	0.2677

trace=2.193884, identified as