Lepton ID in Jets & Background Estimation for Flavor Physics

International Workshop on the High Energy Circular Electron Positron Collider
2021/11/11

中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

CEPC

By: Dan YU

Plan

- Current status & highlights
  - Jet lepton identification
  - Tau identification
  - Bc analysis
- Bkg estimation
- Summary

CEPC TeraZ

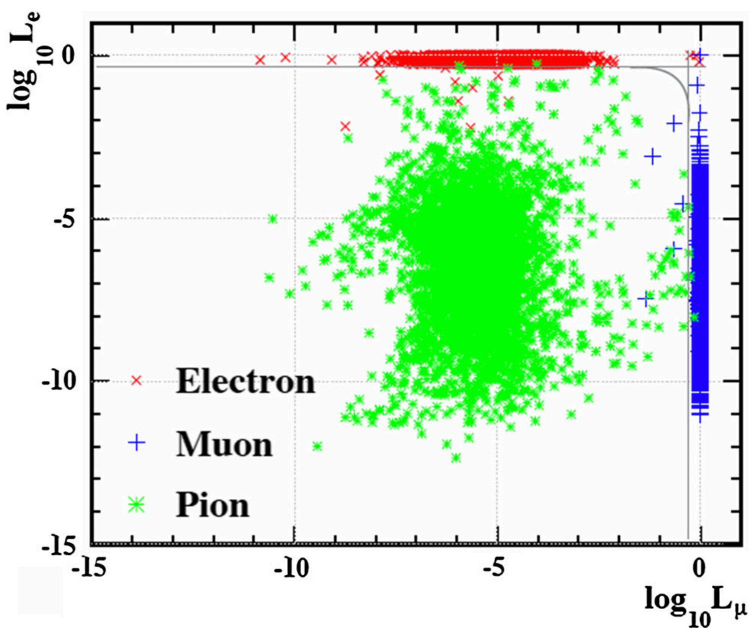
- High lumi: 10^12 Z bosons
- 91.2GeV: Boosted decay products
- Clean environment
  - low QCD bkg
  - no pile-ups
  - fixed Ecm
  - precise detector system & reconstruction algorithm

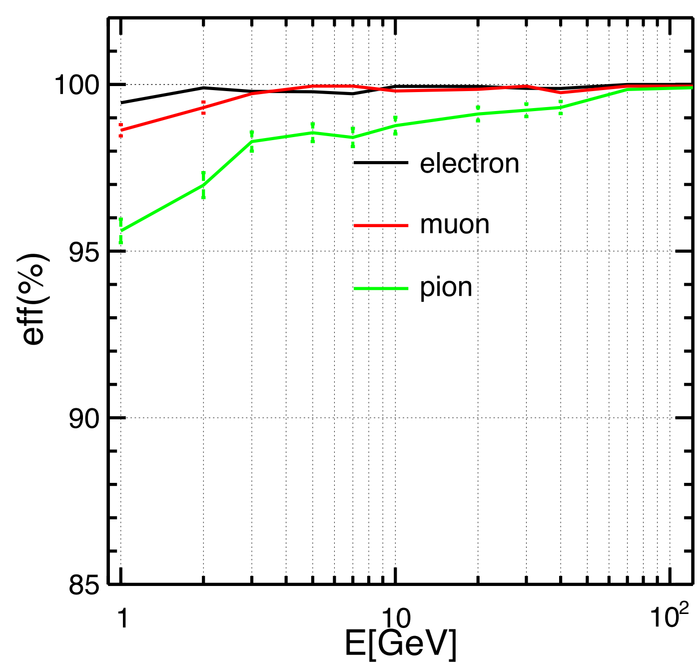
Performances for Physics objects

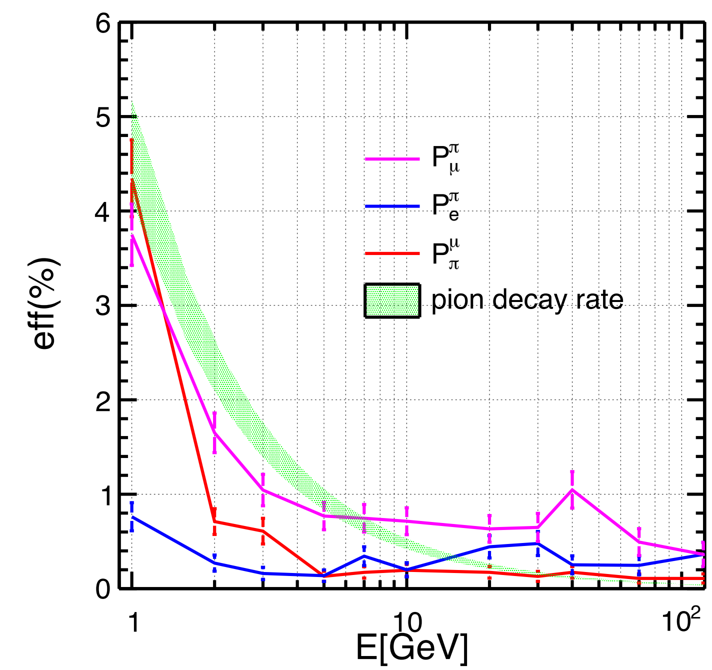
- Acceptance: |cos(θ)| < 099
- Tracks:
  - Pt threshold, ~ 100 MeV
  - δp/p ~ o(0.1%)
- Photons:
  - Energy threshold, ~ 100 MeV
  - δE/E: 3 – 15%/sqrt(E)
- Pi-Kaon separation: 3-sigma
- Pi-0: rec. eff*purity @ Z→qq > 60% @ 5GeV
- Jet charge: eff*(1-2ω)2 ~ 15%/30% @Z→bb/cc
- B-tagging: eff*purity @ Z→qq: 70%
- C-tagging: eff*purity @ Z→qq: 40%
- Lepton inside jets: eff*purity @ Z→qq ~ 90% (energy > 3 GeV)
- Tau: eff*purity @ WW→tauvqq: 70%, mis id from jet fragments ~ o(1%)
- Reconstruction of simple combinations: Ks/Lambda/D with all tracks @ Z→qq: 60/75 – 80/85%
- BMR: 3.7%
- Missing Energy: Consistent with BMR

Isolate Leptons

- LICH uses TMVA methods to summarize 24 input variables into two likelihoods, corresponding to electrons and muons.
- The efficiency for electron and muon is higher than 99.5% (E>2 GeV). Pion efficiency ~ 98%.







Migration Matrix at 40GeV (LICH)

Type	e^- like	μ^- like	π^+ like
e^-	99.71 ± 0.08	< 0.07	0.21 ± 0.07
μ^-	< 0.07	99.87 ± 0.08	0.05 ± 0.05
π^+	0.14 ± 0.05	0.35 ± 0.08	99.26 ± 0.12

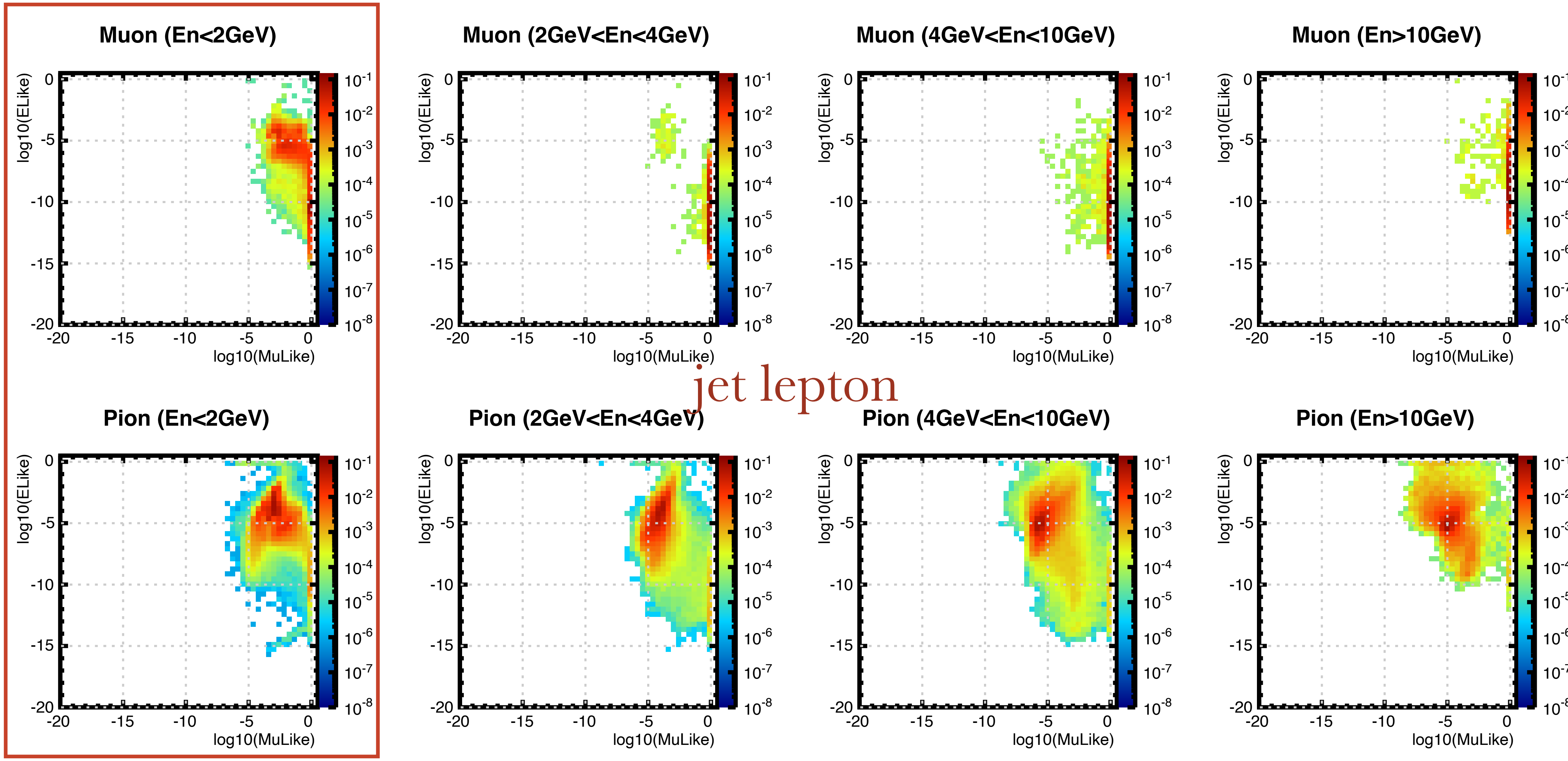
Migration Matrix for ALEPH PID ($> 2GeV$)(*Eur.Phys.J.C20:401-430,2001*)

Type	e^- like	μ^- like	π^+ like	undefined
e^-	99.57 ± 0.07	< 0.01	0.32 ± 0.0	0.09 ± 0.04
μ^-	< 0.01	99.11 ± 0.08	0.88 ± 0.08	0.01 ± 0.01
π^+	0.71 ± 0.04	0.72 ± 0.04	98.45 ± 0.06	0.12 ± 0.03

Eur. Phys. J. C (2017) 77: 591

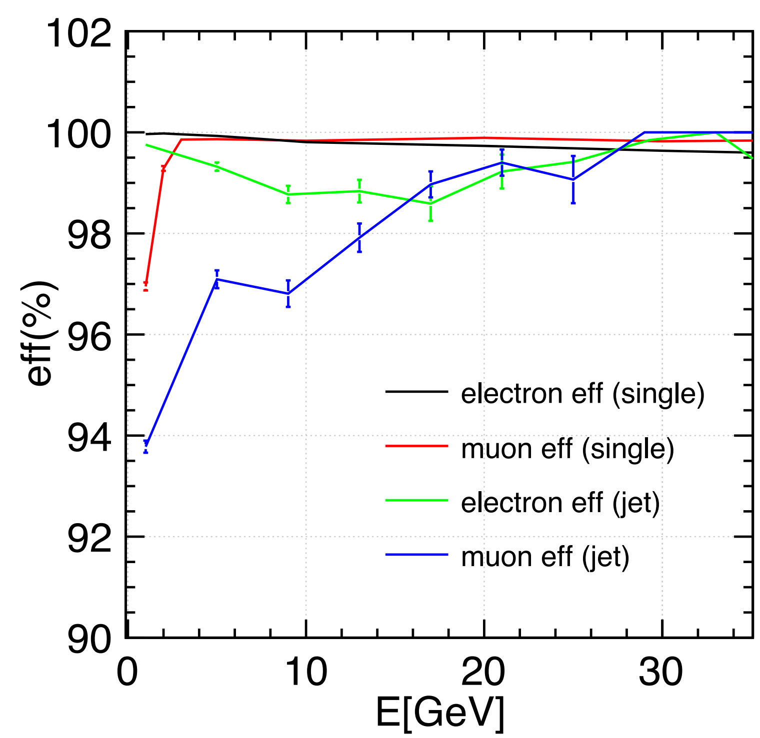
Likelihood vs Energy

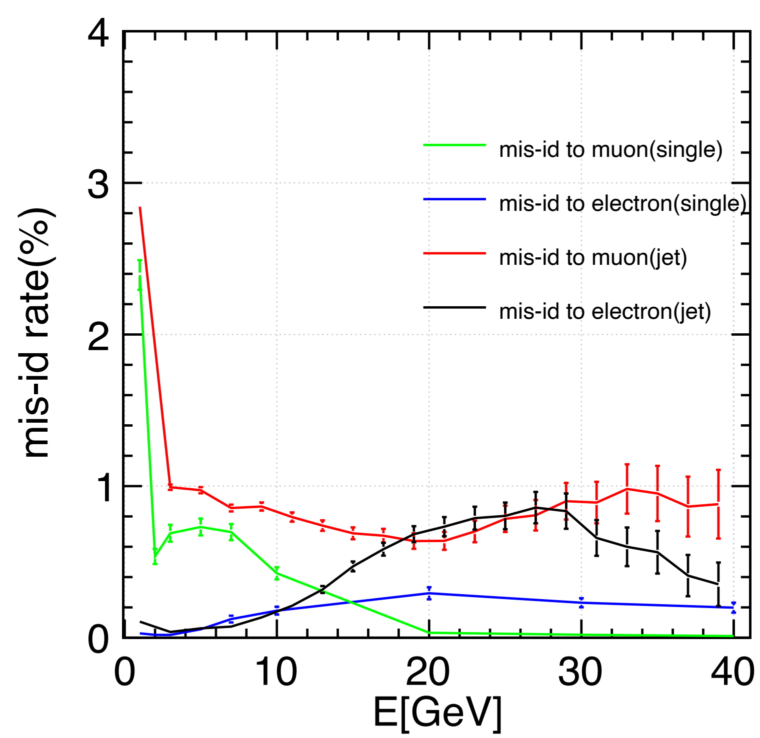
- For higher energy, still nice separation
- For lower energy, pion mixed with muon



Lepton in jets

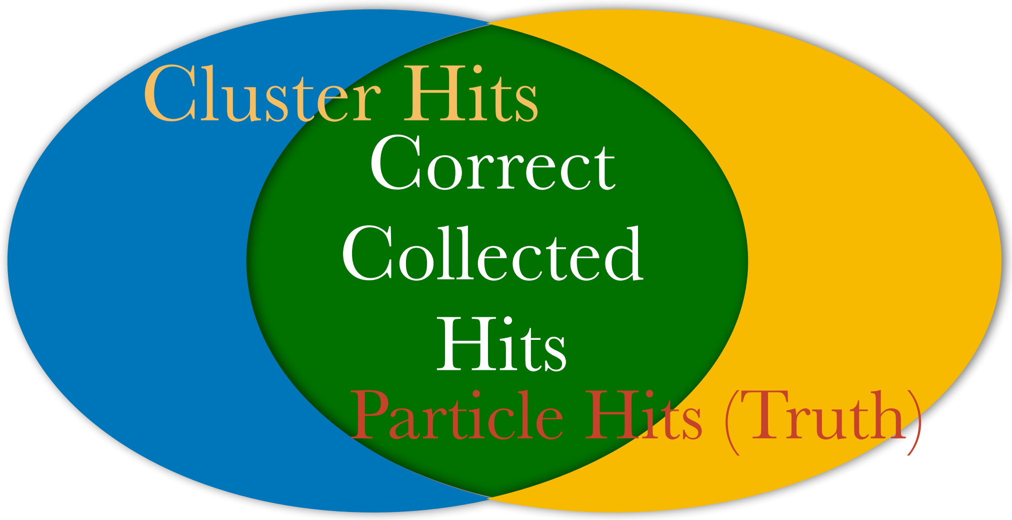
- The performance for lepton in jets degrades comparing to the single particle results
- Application: Bc→τν (arxiv:2007.08234 by Taifan ZHENG)
- Further: more flavor physics such as lepton flavor universality, etc.





Clustering Performance

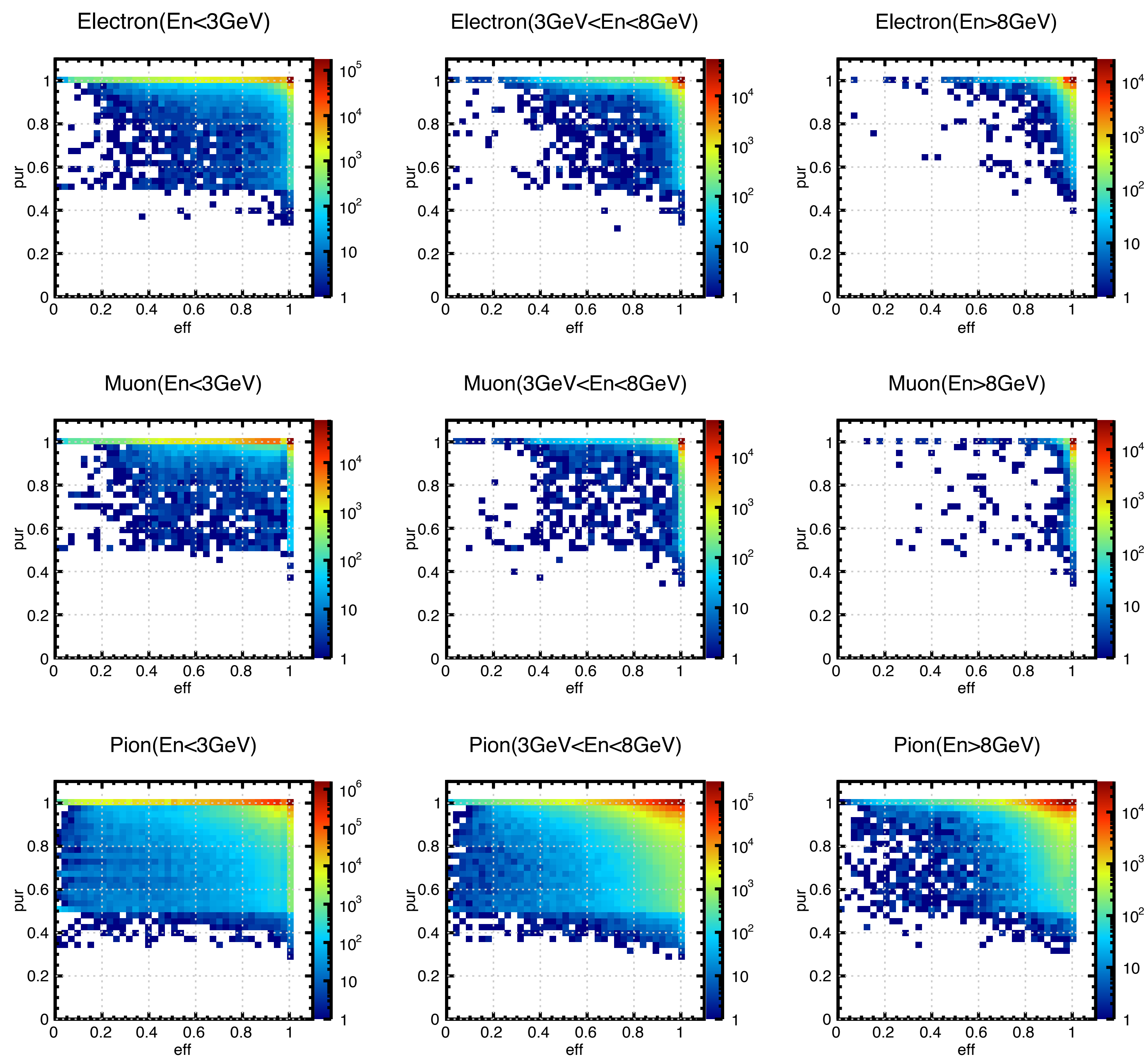
- Use clustering
  - efficiency (correct collected hits/particle hits)
  - purity (correct collected hits/cluster hits)
- to characterize clustering performance



Clustering Performance

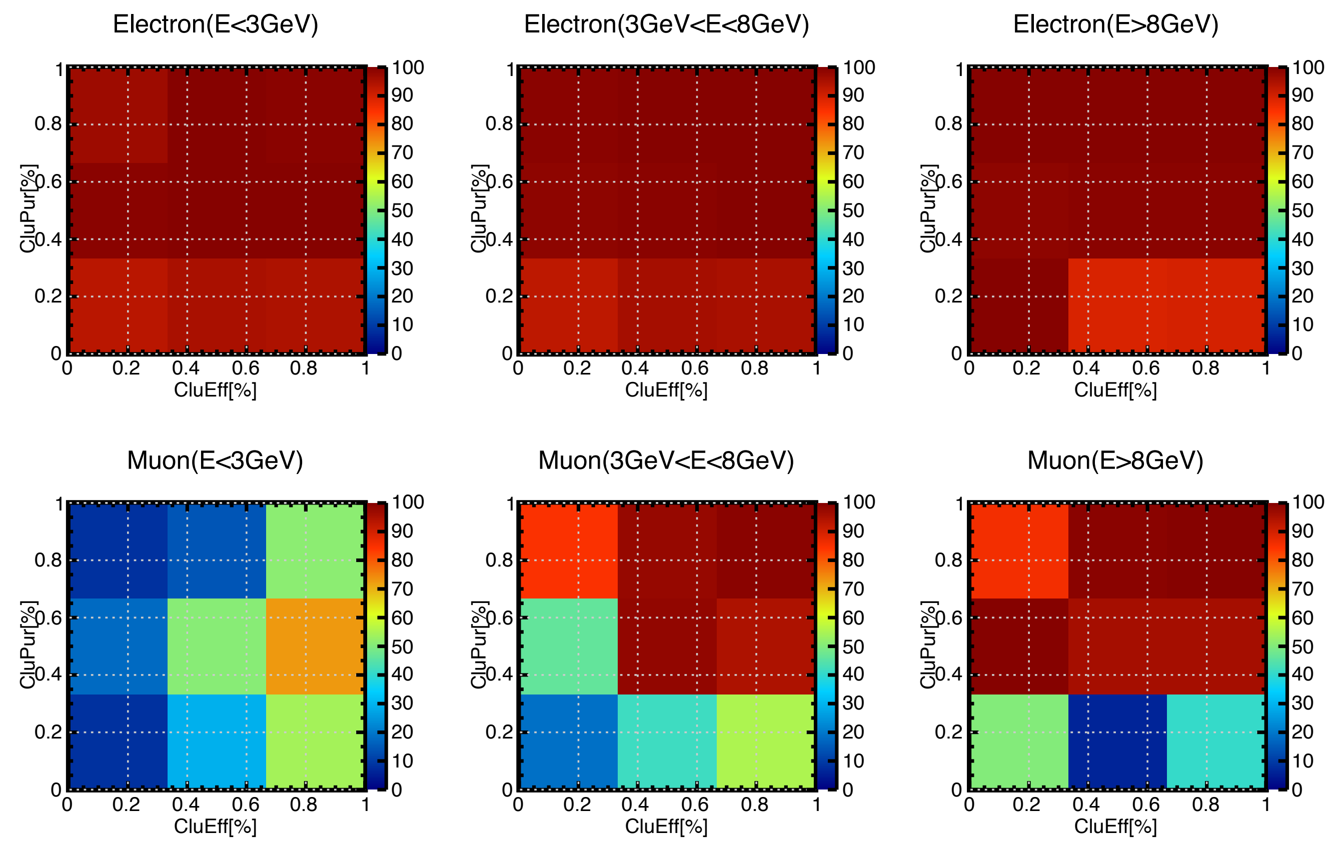
- Higher energy, better clustering performance

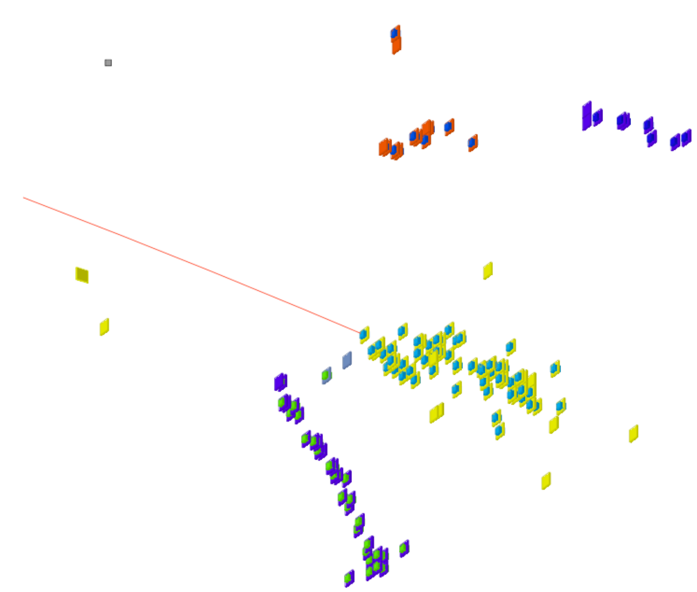
ratio(%)		<3GeV	3-8GeV	>8GeV
e	1	71.06	65.16	50.85
	<0.9	18.87	9.84	6.64
μ	1	54.55	81.68	82.53
	<0.9	31.45	5.75	3.55
π	1	52.84	24.75	13.45
	<0.9	26.77	32.68	30.52



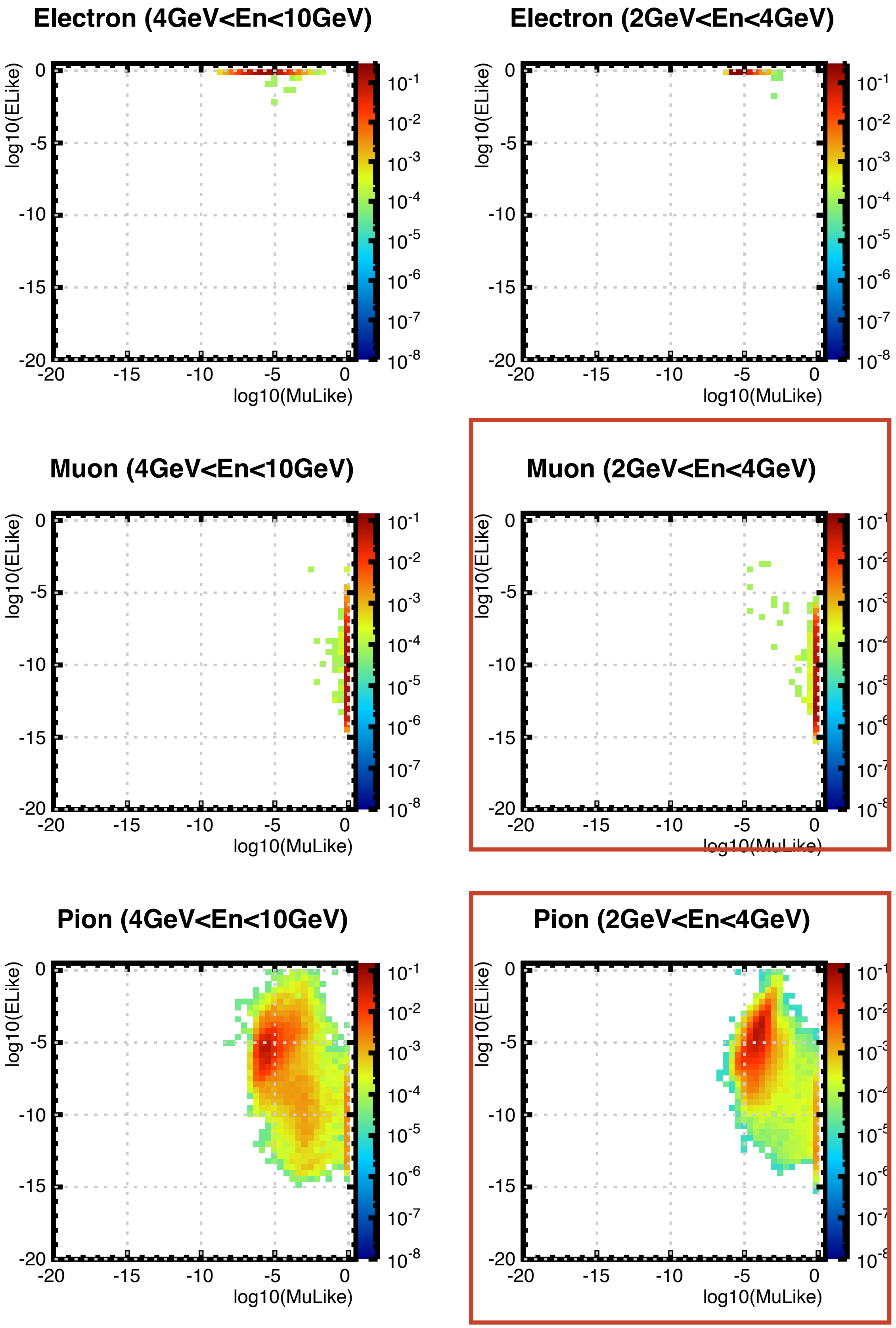
Clustering vs PID

- Electrons:
  - low energy: dE/dx dominate
  - clusters are compact, the splitting clusters still electron-like
- Muon:
  - cluster is not MIP-like if mixed with other hits
  - muon likeliness is lost when the muon cluster splits into small pieces
- Pion:
  - likely to be a EM cluster with some branches
  - more likely to be mis-identified as an electron for lower clustering efficiency

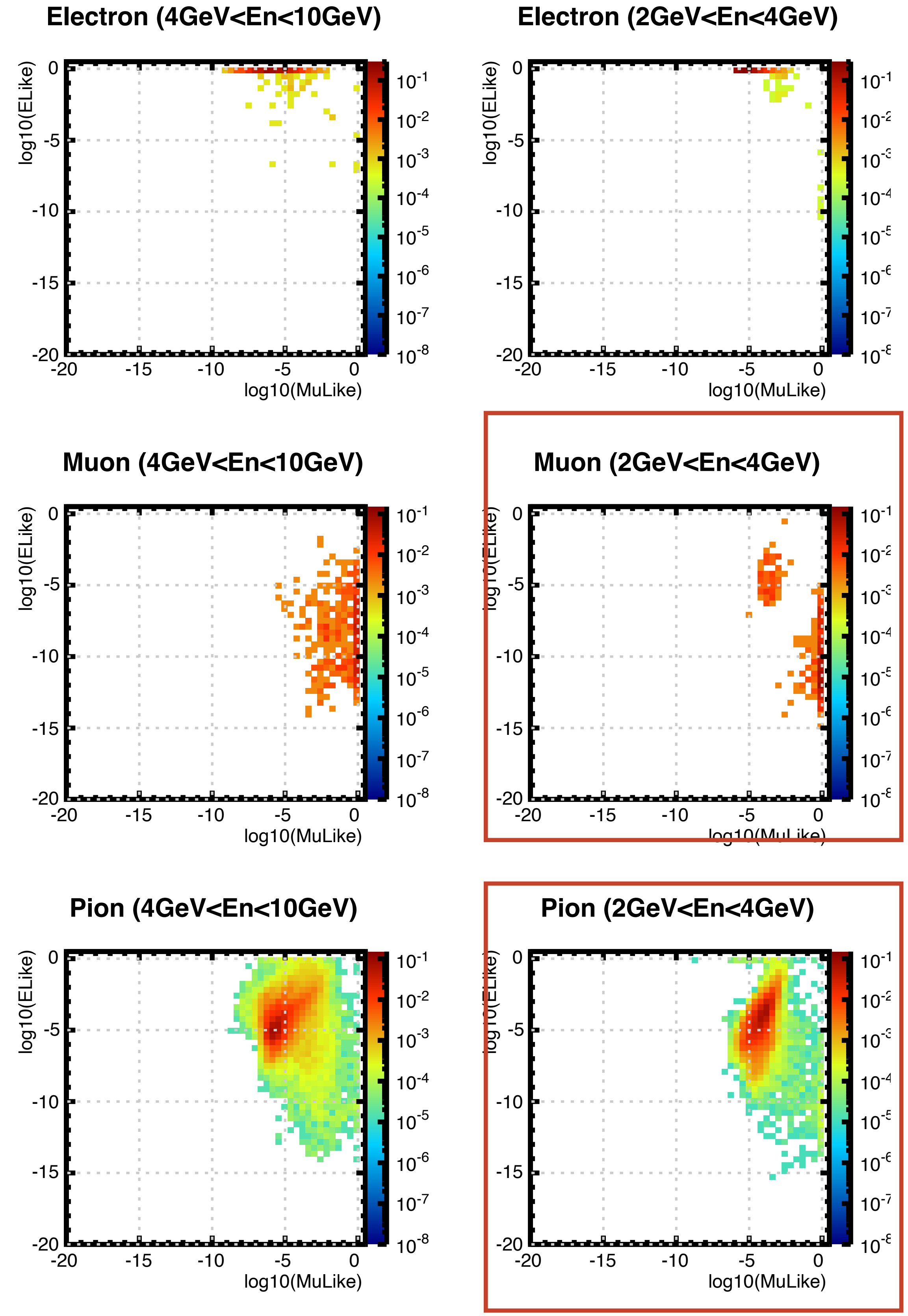




Clustering eff*pur=1

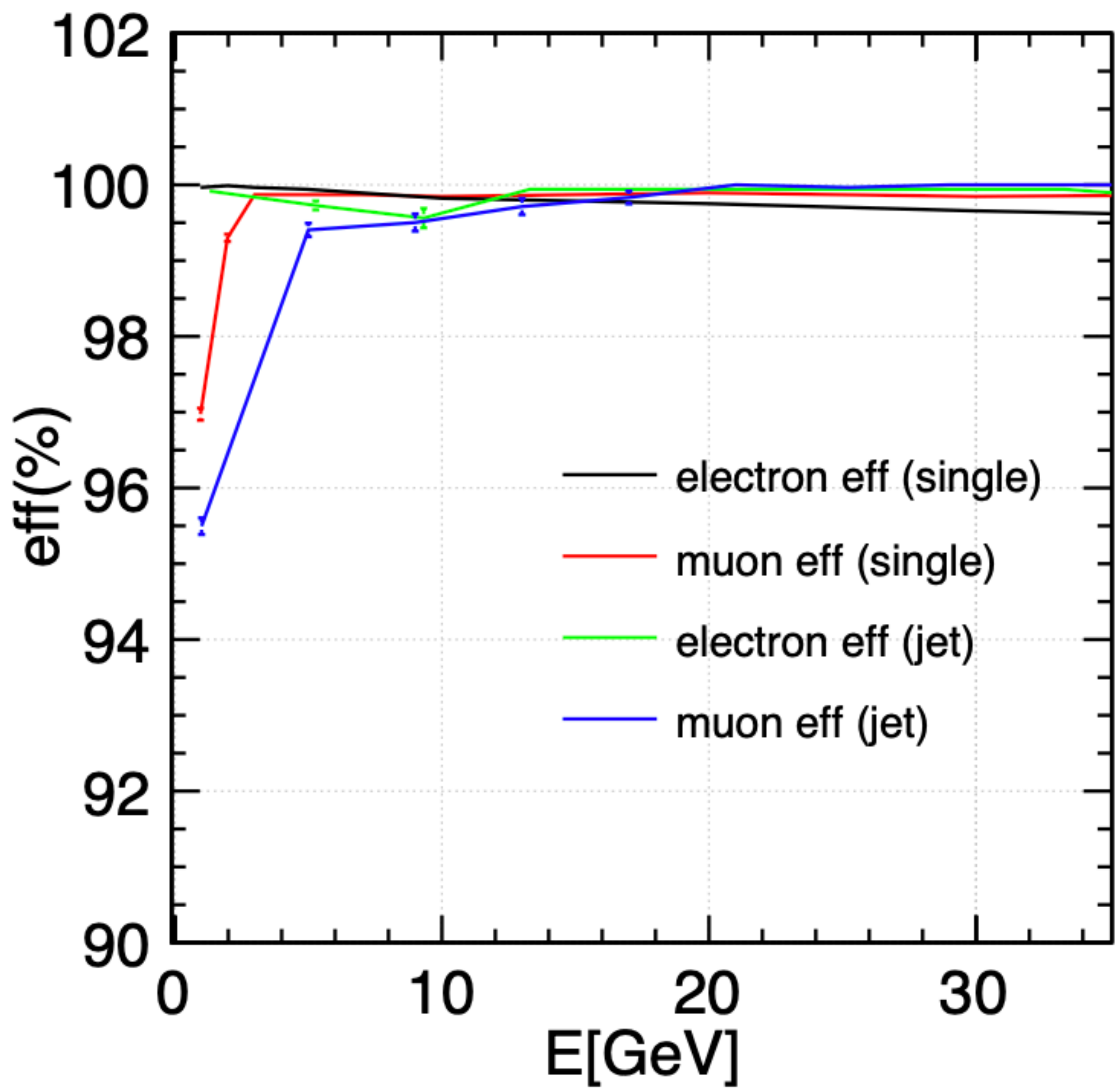


Clustering eff*pur<0.9



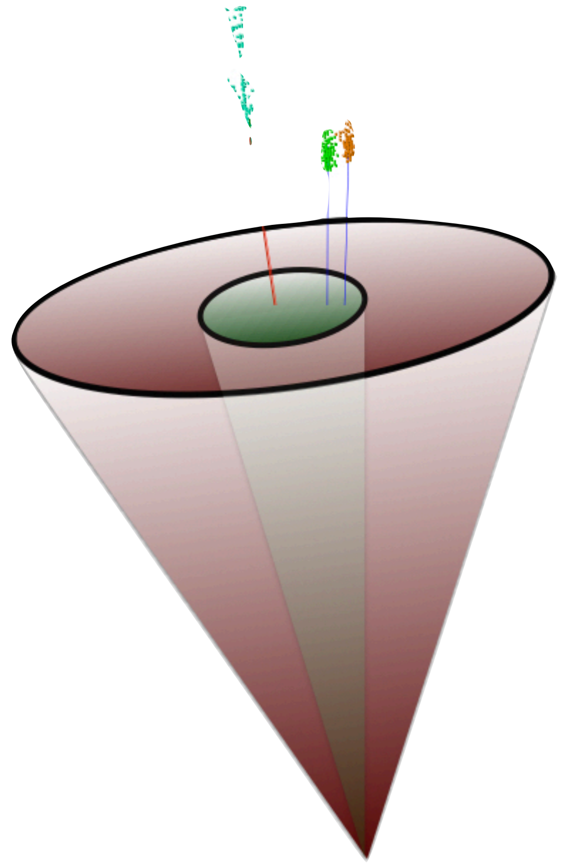
Comparison

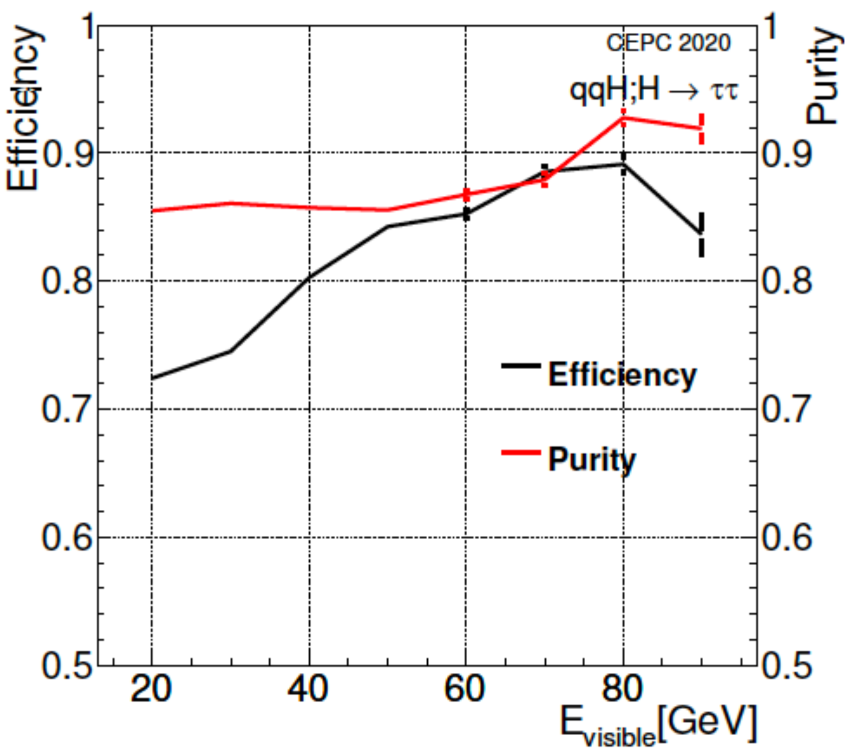
- Comparison of lepton identification performance for perfect clusters and the performance of single particle

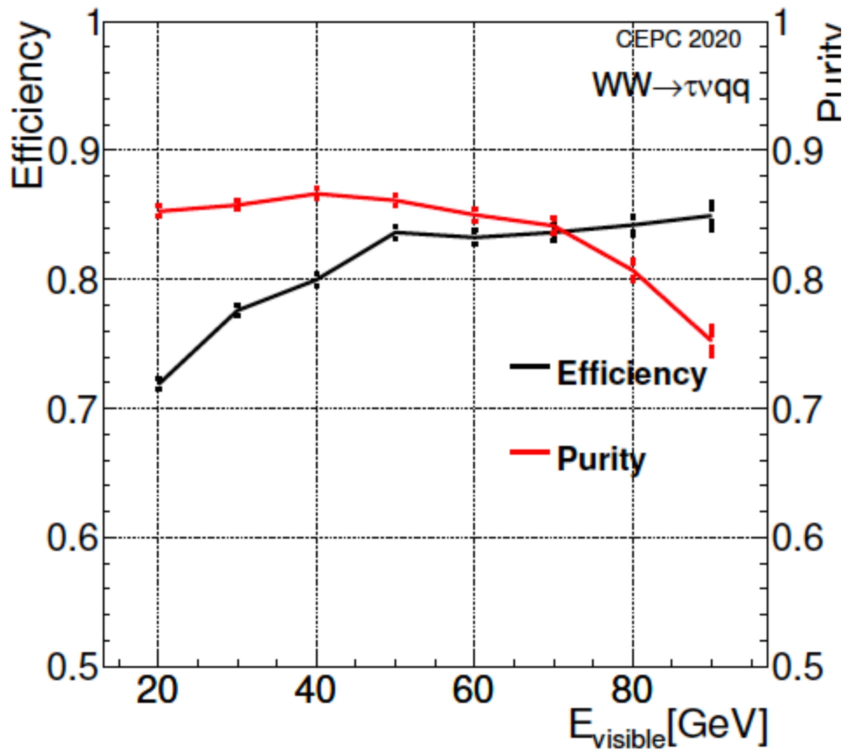


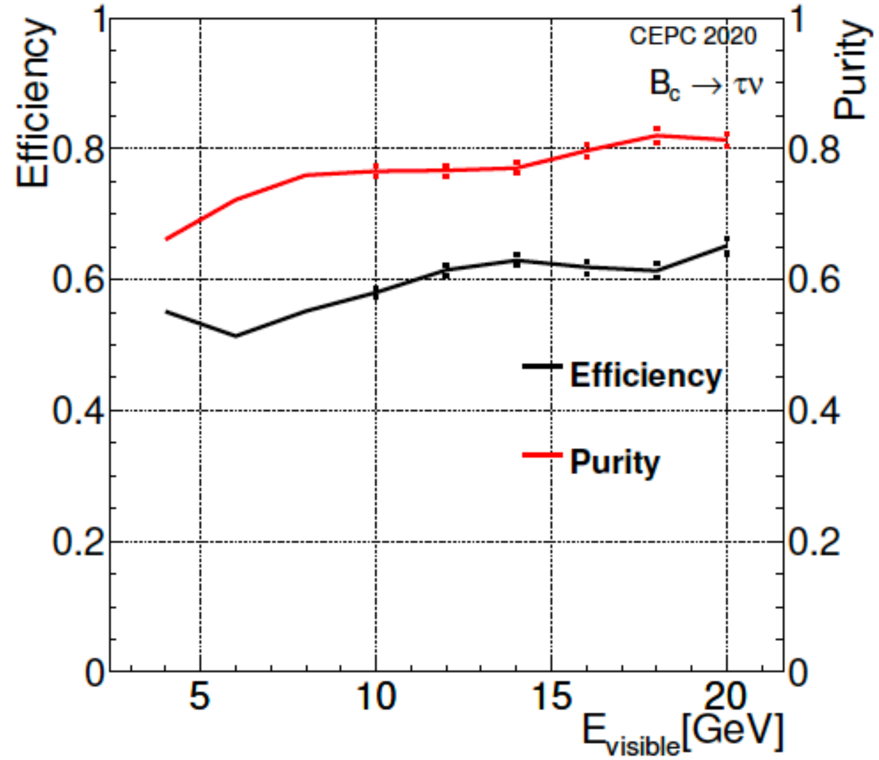
Taurus

- Double cone based algorithm
- Use the multiplicity, energy ratio between two cones, invariant mass for τ tagging



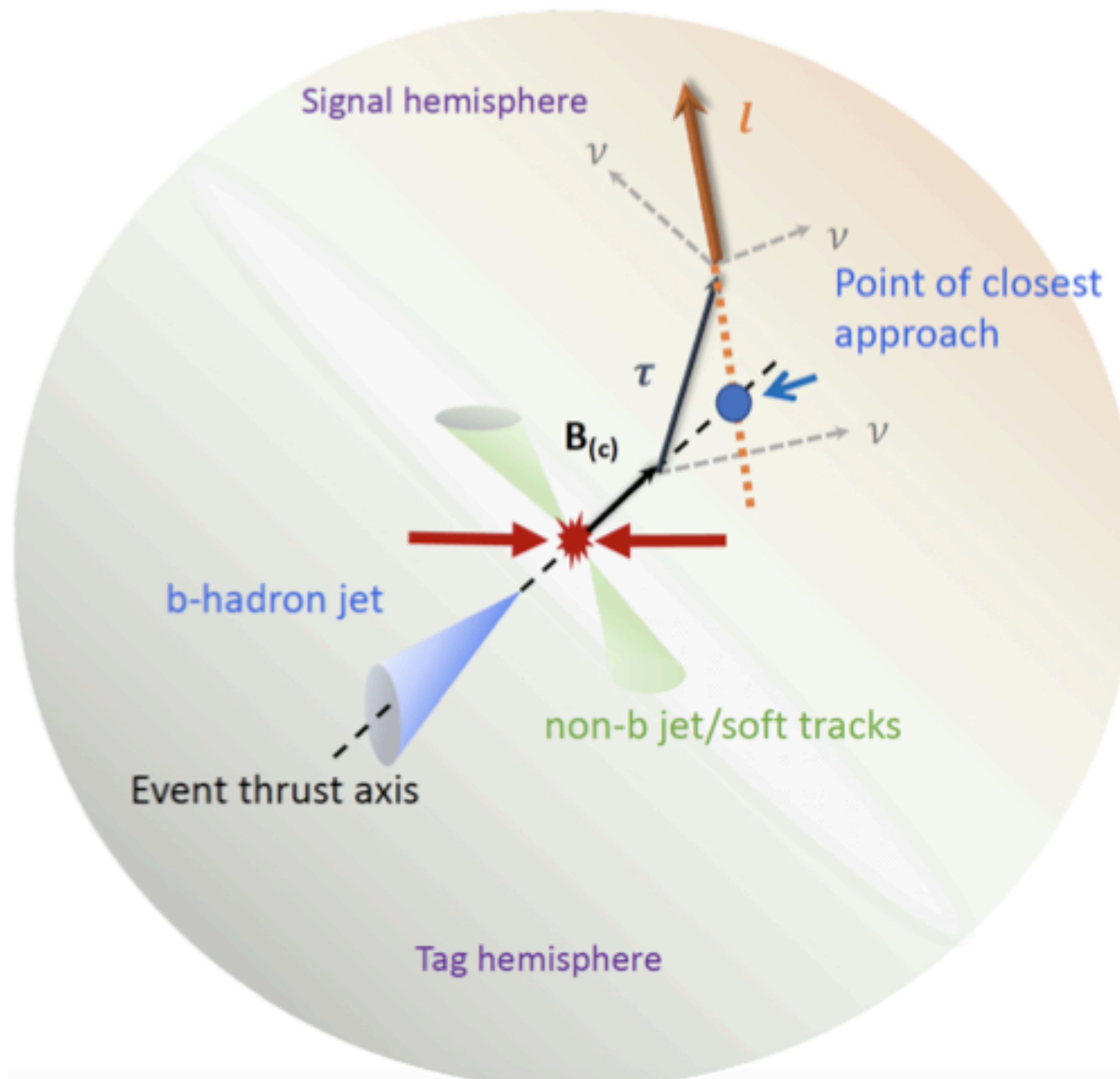






Bc Analysis — Method

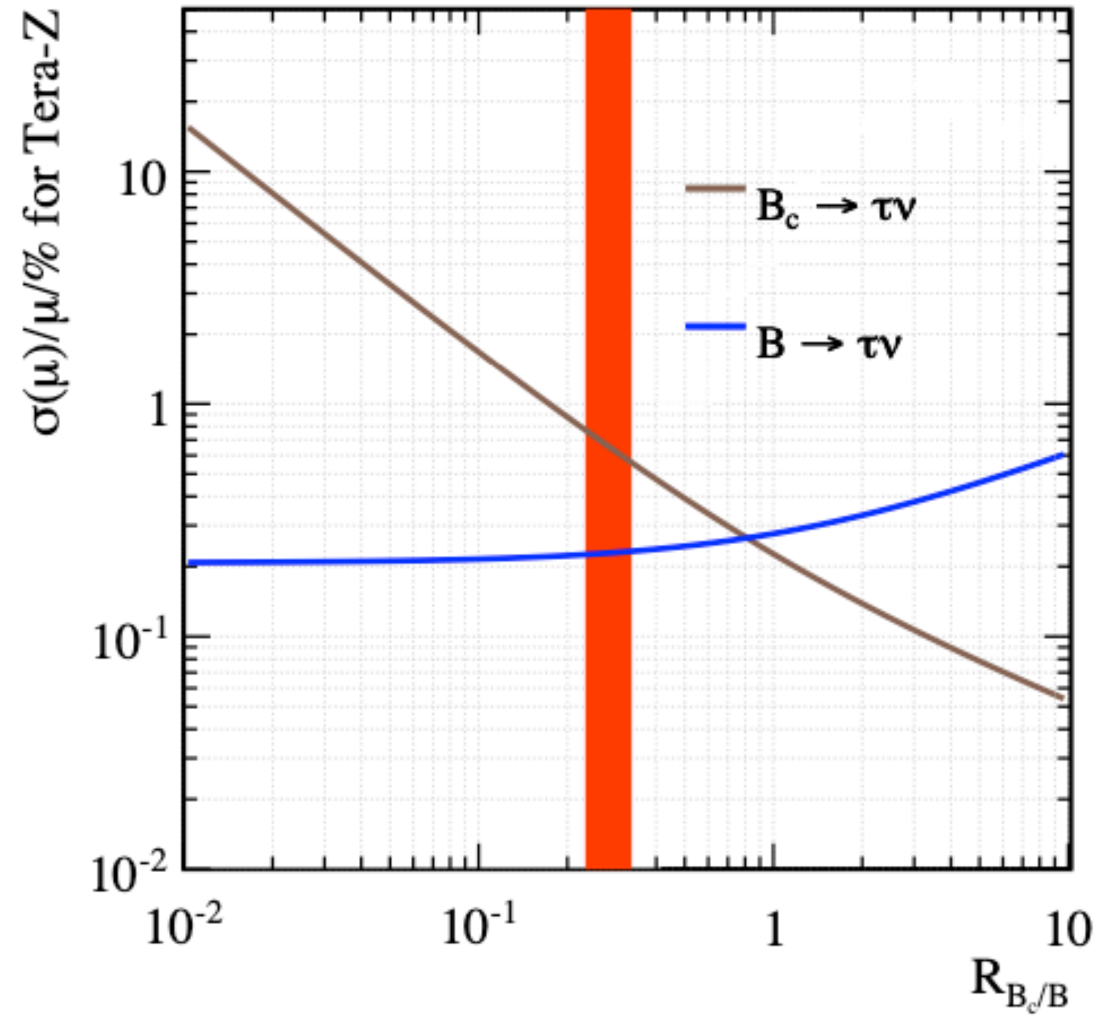
- Reducing light flavor events:
  - B-tagging
  - energy asymmetry
  - Signal hemisphere with electron & electron is the most energetic particle in signal hemisphere
  - Nominal B meson energy > 20GeV
- Reducing bb/cc background (BDT)
- Separate Bc & B (BDT with the electron's impact parameter along the thrust)



by: Taifan Zheng

Result

- Relative accuracy of the signal strength @ 10^9 Z: 7.2% (combined)
- Tera-Z: ~1% level
  - Bc production
  - |Vcb|



	$B_c^{\pm} \to \tau\nu_\tau(0.013)$ $\tau \to e\nu\bar{\nu}$	$B_c^{\pm} \to \tau\nu_\tau(0.013)$ excl. $\tau \to e\nu\bar{\nu}$	$B^{\pm} \to \tau\nu_\tau(0.013)$ $\tau \to e\nu\bar{\nu}$	$B^{\pm} \to \tau\nu_\tau(0.013)$ excl. $\tau \to e\nu\bar{\nu}$	$d\bar{d}(15) + u\bar{u}(12) + s\bar{s}(15)$	$c\bar{c}(4.8)$	$b\bar{b}(3.25)$
All events	2,303	10,691	2,270	10,633	419,928,342	119,954,033	151,286,603
b-tag > 0.6	1,611	7,463	1,547	7,151	2,134,617	7,344,014	116,723,067
Energy asymmetry > 10 GeV	1,425	6,184	1,389	5,801	486,762	1,609,771	30,064,030
Has electron in signal hemisphere	1,273	1,300	1,243	1,132	143,595	625,670	15,905,613
Electron is the most energetic particle	915	116	859	93	8,490	79,190	4,587,248
$E_B > 20$ GeV	909	112	852	88	981	34,147	3,203,073
1^{st} BDT score > 0.99	390	12	259	4	—	48	910
2^{nd} BDT score > 0.4	199	12*	73	4*	—	48*	33

Background Estimation

Earlier Estimation

- LFV: Mogens Dam (arxiv:1811.09408)
- Z hadronic decays: draft by Shan Cheng, Qin Qin and Fu-Sheng YU

Decay mode	Branching ratio	CEPC Uncertainty
$Z \to J/\psi\gamma$	8.02×10^{-8} [29]	$\sim 1.8\%$
$Z \to \Upsilon(1S)\gamma$	5.39×10^{-8} [29]	$\sim 3.4\%$
$Z \to \rho^0\gamma$	4.19×10^{-9} [29]	$\sim 1.8\%$
$Z \to \omega\gamma$	2.82×10^{-8} [29]	$\sim 0.8\%$
$Z \to \phi\gamma$	1.04×10^{-8} [29]	$\sim 1.6\%$
$Z \to \pi^0\gamma$	9.80×10^{-12} [29]	$< 3.4 \times 10^{-8}$
$Z \to \eta\gamma$	$0.1 - 1.7 \times 10^{-10}$ [30]	$\sim 12\% - 50\%$
$Z \to \eta'\gamma$	$3.1 - 4.8 \times 10^{-9}$ [30]	$\sim 2.7 - 3.4\%$

Decay	Present bound	FCC-ee sensitivity
$Z \to \mu e$	0.75×10^{-6}	10^{-10}–10^{-8}
$Z \to \tau\mu$	12×10^{-6}	10^{-9}
$Z \to \tau e$	9.8×10^{-6}	10^{-9}
$\tau \to \mu\gamma$	4.4×10^{-8}	2×10^{-9}
$\tau \to 3\mu$	2.1×10^{-8}	10^{-10}

Decay mode	Branching ratio	CEPC Uncertainty
$Z \to \pi^{\pm}W^{\mp}$	1.5×10^{-10}	$\sim 20\%$
$Z \to \rho^{\pm}W^{\mp}$	4.0×10^{-10}	$\sim 13\%$
$Z \to K^{\pm}W^{\mp}$	1.2×10^{-11}	$\sim 70\%$
$Z \to K^{*\pm}W^{\mp}$	2.0×10^{-11}	$\sim 59\%$
$Z \to D_s^{\pm}W^{\mp}$	6.0×10^{-10}	$\sim 75\%$
$Z \to D^{\pm}W^{\mp}$	2.0×10^{-11}	$< 3 \times 10^{-10}$

Samples

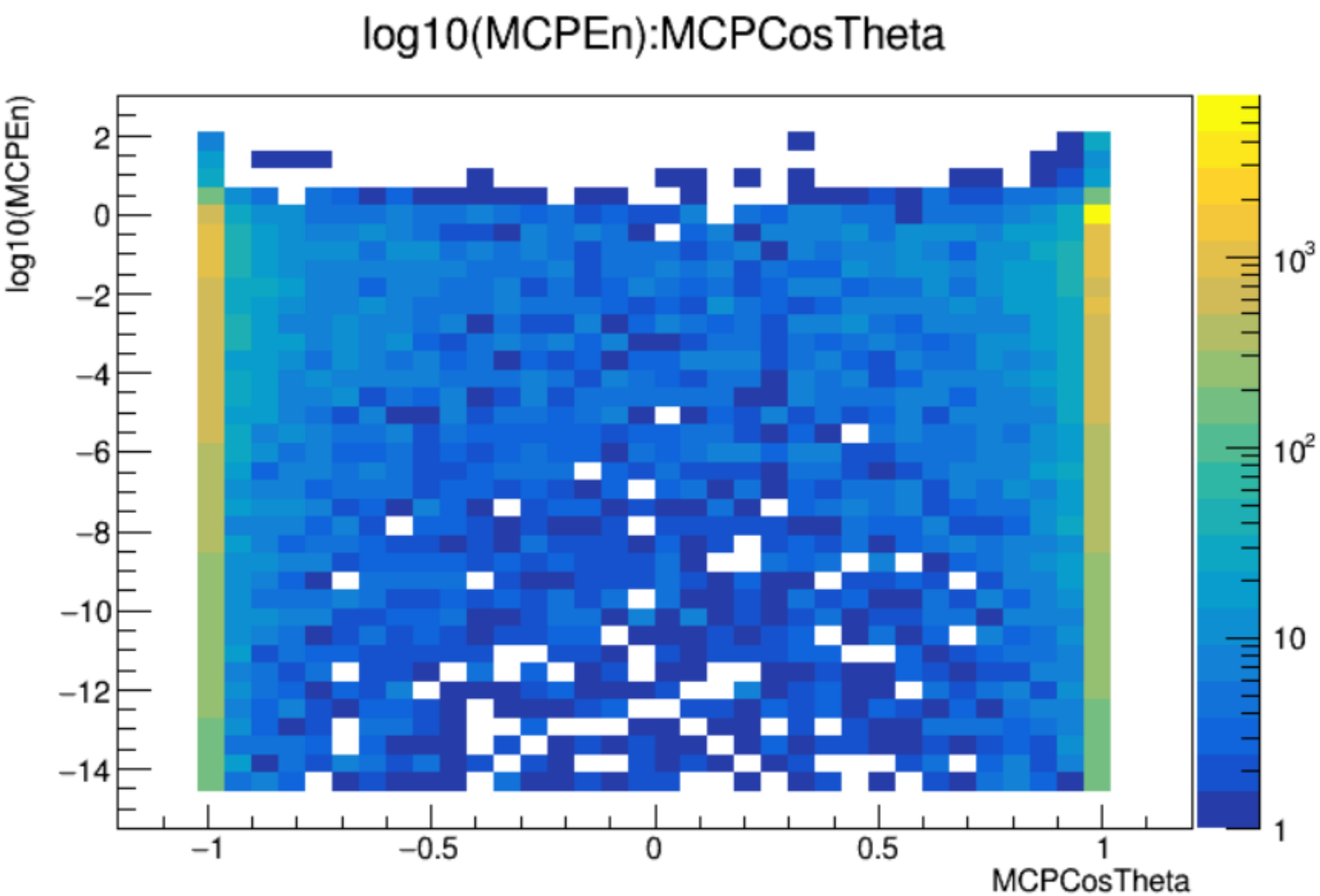
- CEPC Zpole: 10^{12} Z bosons ~ 3.36×10^{10} $\tau\tau$, 6.99×10^{11} qq
- Current samples: Truth ~ 4 $\times10^{-4}$ SM; Reco ~ 4 $\times10^{-5}$ SM
- No Z boson width or beam energy spread

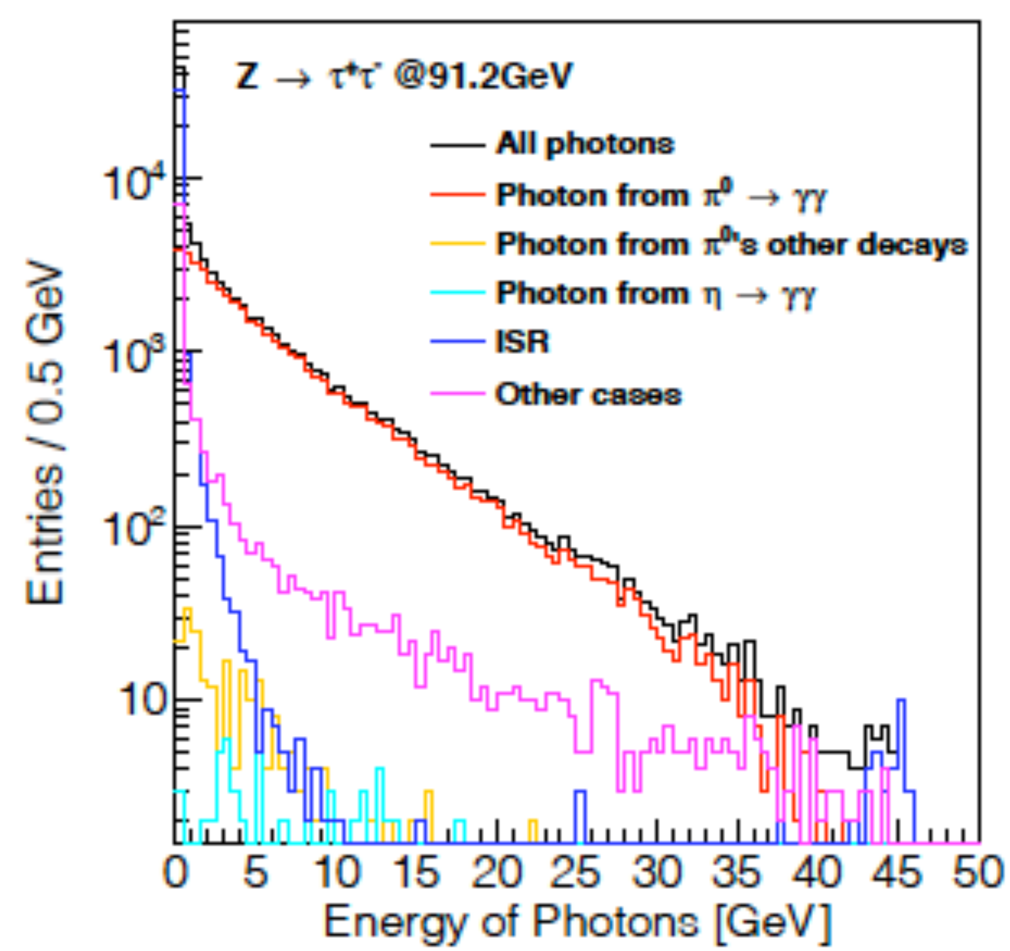
Table 1: CEPC91.2GeV

	Channel	Generator		FullSim		DstDate	
		size (GB)	yield (Million)	size (GB)	yield (Million)	size (GB)	yield (Million)
wo_isr	bb	3713	376	-	-	-	-
	cc	2610	294	-	-	-	-
	uu	2419	295	-	-	-	-
	e3e3	137	851	-	-	-	-
wi_isr	bhabha	49	33	619	1.1	19	1.1
	e2e2	151	120	105	1.6	9	1.6
	e3e3	25	13	355	1.6	25	1.6
	n1n1	22	24	-	-	-	-
	n2n2	23	25	-	-	-	-
	n3n3	22	24	-	-	-	-
	nn	67	73	-	-	-	-
	uu	365	42	5918	8.1	687	8.1
	dd	470	55	5931	8.1	678	8.1
	ss	467	55	5731	8.1	678	8.1
	bb	559	54	6332	8.1	775	8.1
	cc	404	43	6057	8.1	725	8.1
	qq	2234	253	-	-	-	-

ISR

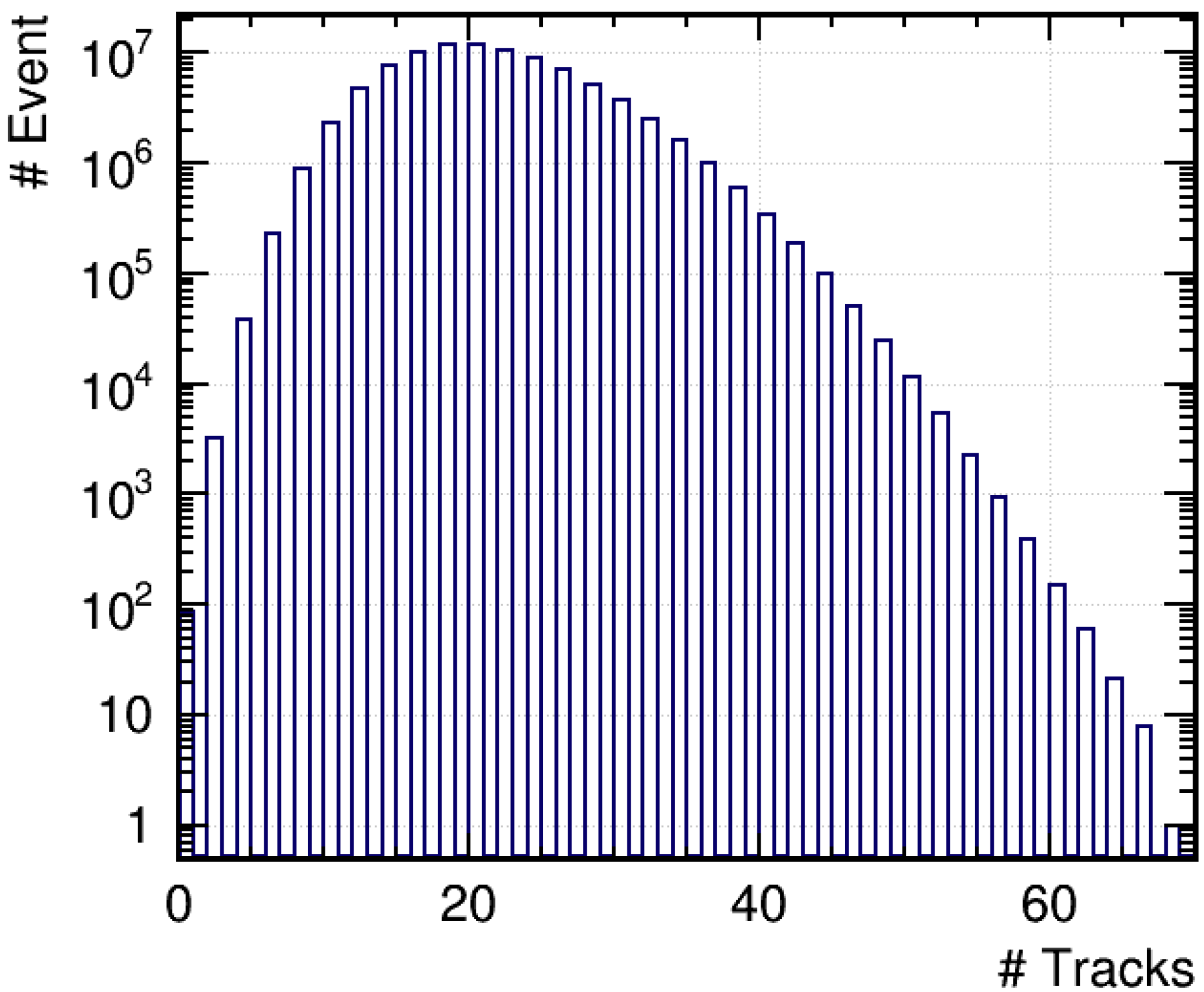
- Mostly to the forward region
- Energy < 1GeV
- Only leading order

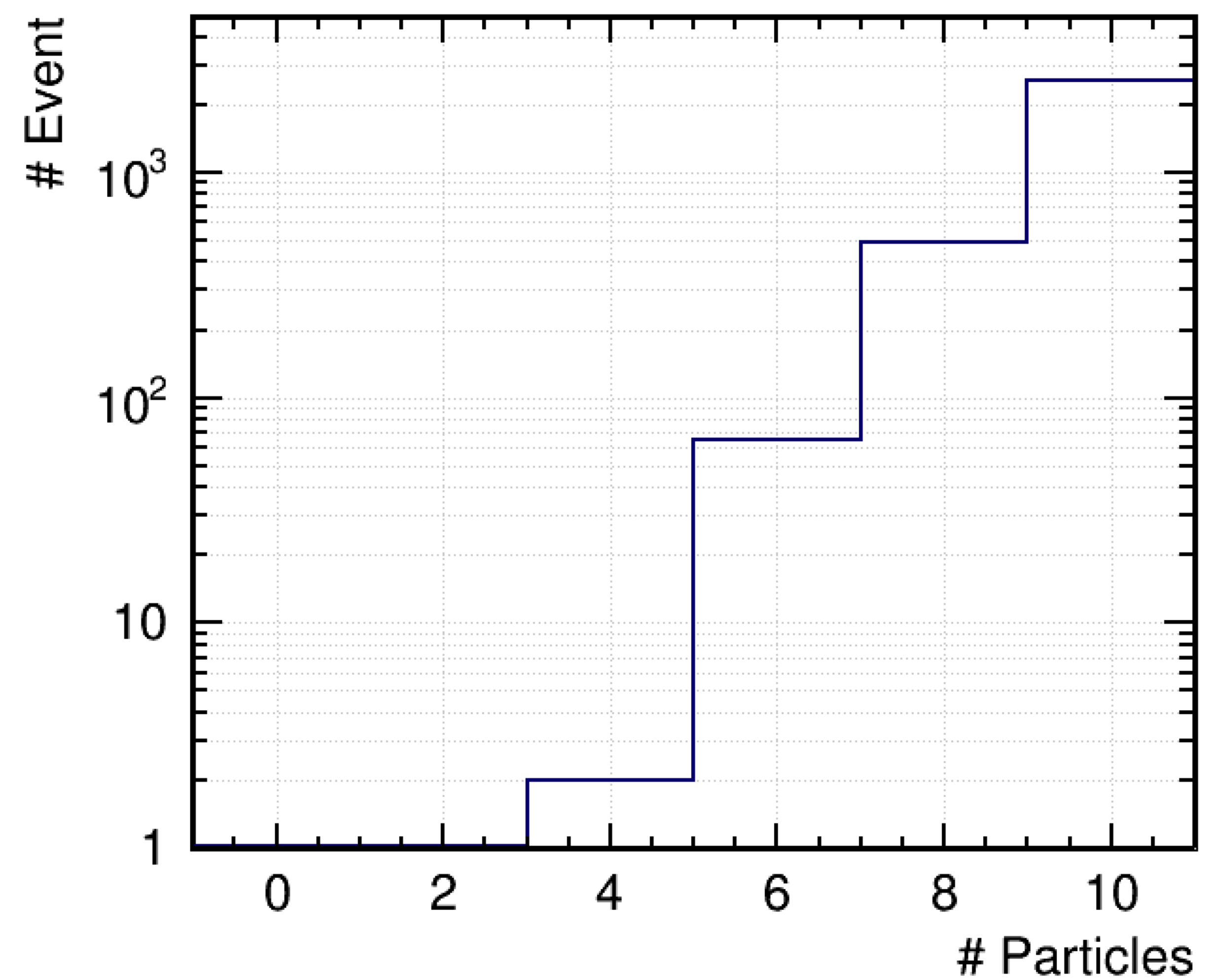




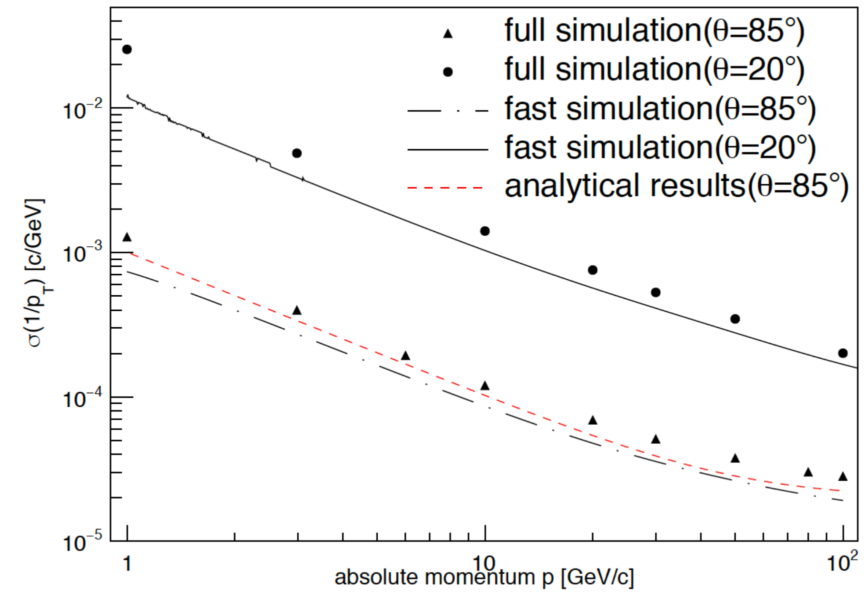
Jet multiplicity

- Jets' multiplicities high, not likely to be mixed with leptonic events
- Multiplicity < 10, need to take care

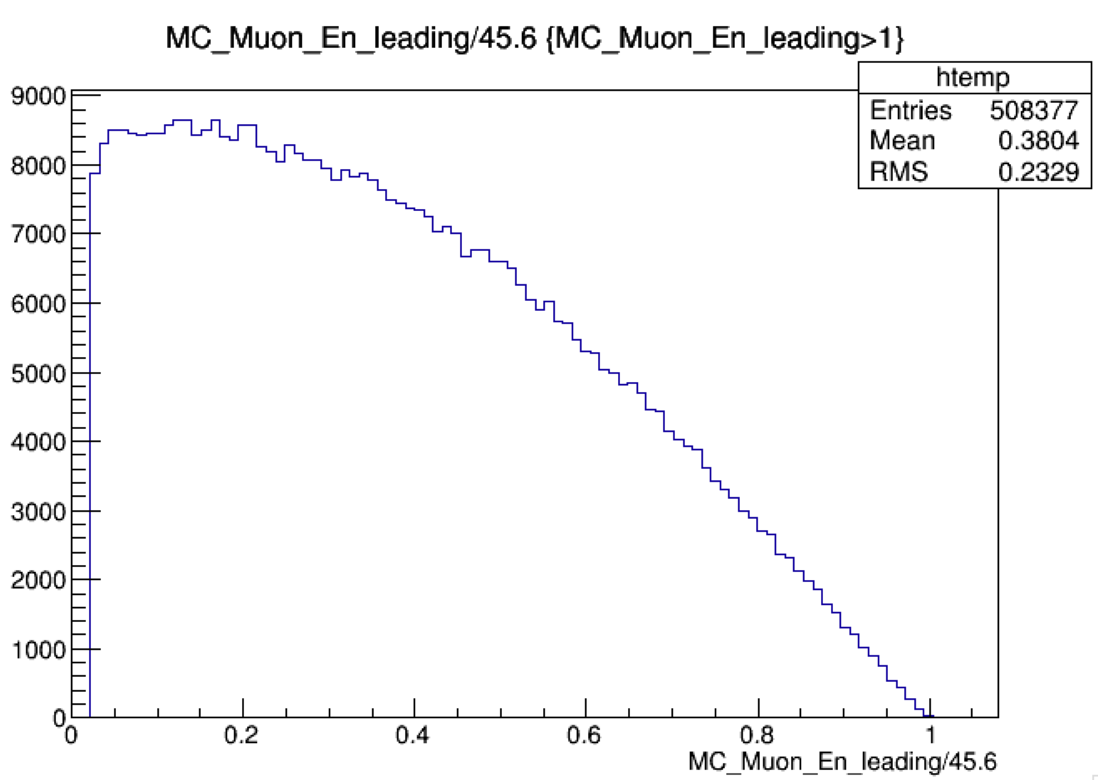


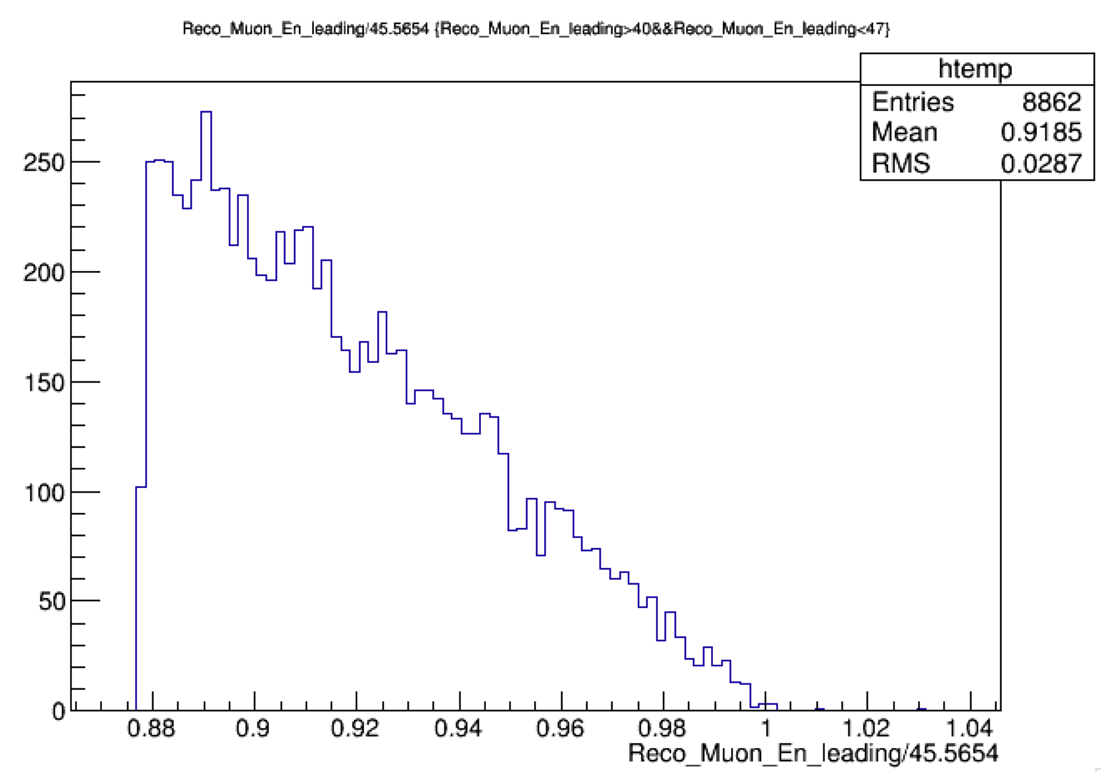


LFV — Z→τμ



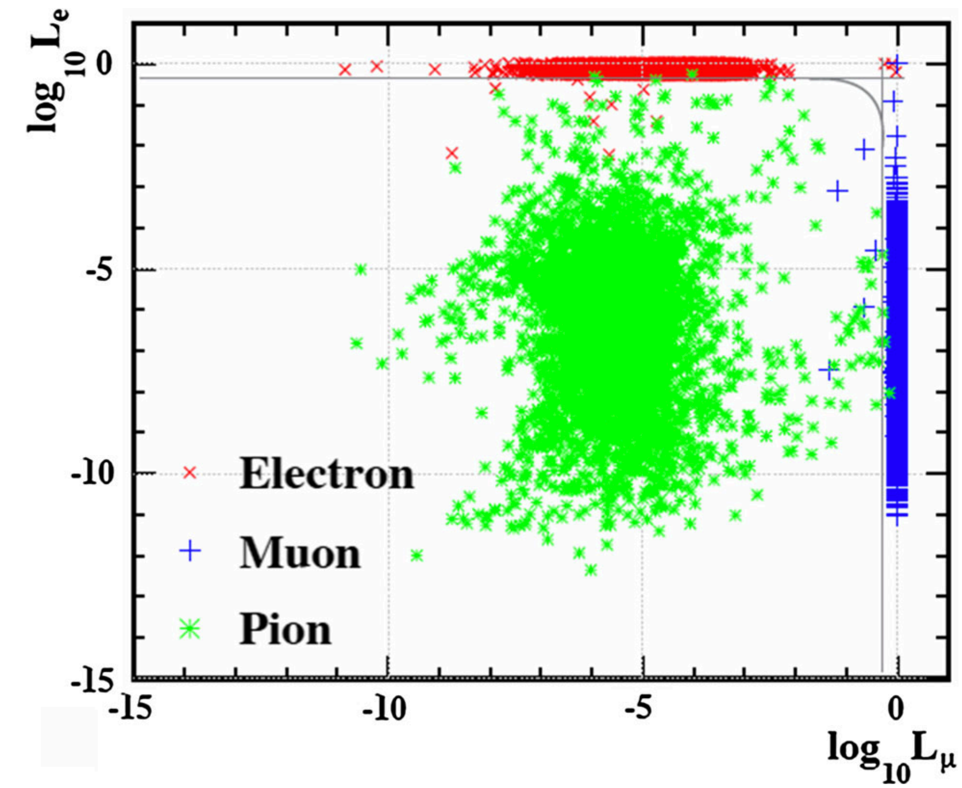
- Main background Z→ττ, τ→μνν
- Current bound: $1.2*10^{-5}$ (LEP) FCC-ee estimation: 10^{-9}
- Key distribution ($P_\mu/P_{beam}>1$):
  - Signal accuracy depend on the momentum resolution ($\delta p/p \sim 10^{-3}$), signal window: (0.998, 1.002)
  - Background surviving Nbkg: 5*ScaleFactor ~ $3.36*10^5$
- Sensitivity estimated: $1.1*10^{-9}$

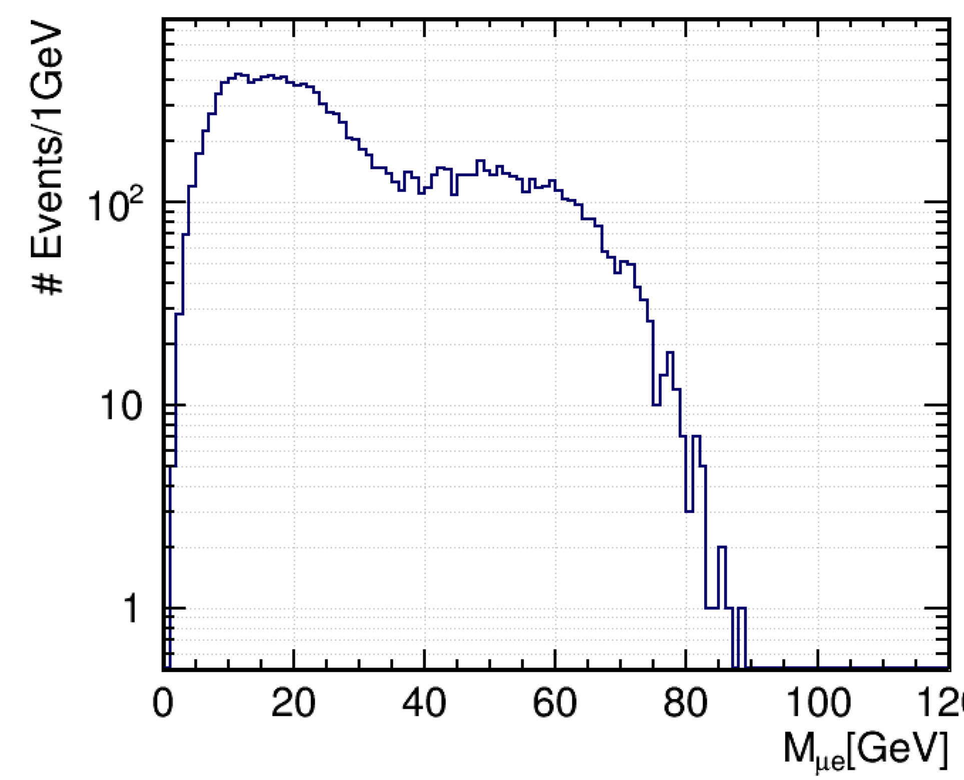




LFV — $Z \to \mu e$

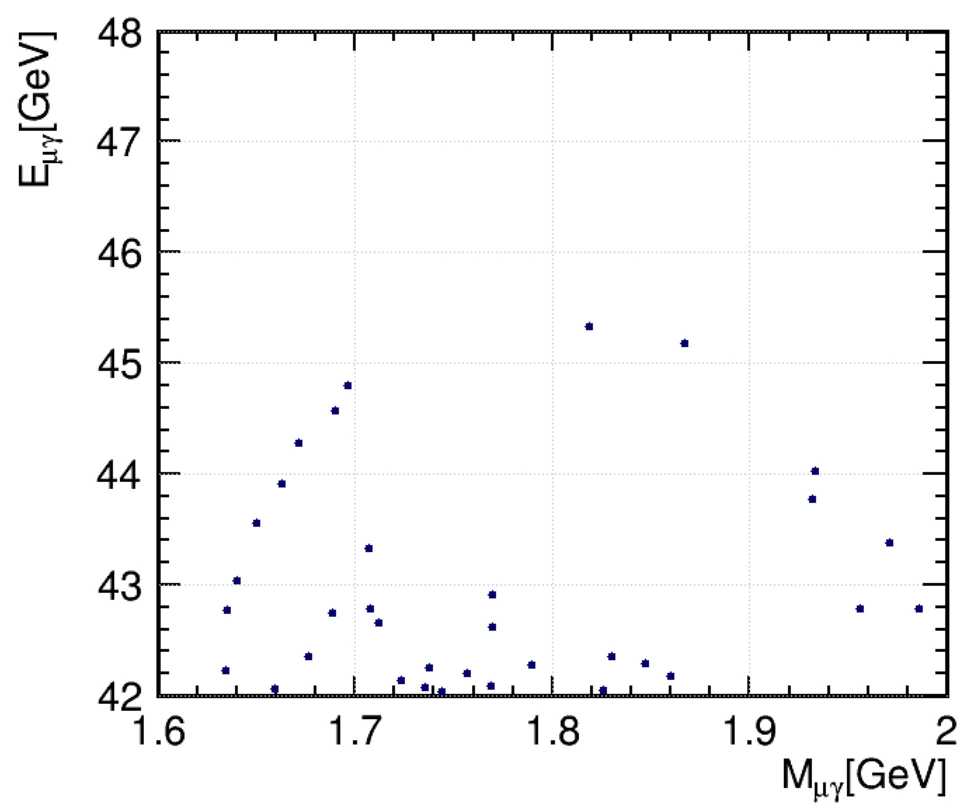
- Physics background: $Z \to$ bhabha/$\mu\mu$/$\tau\tau$
- Current bound: 7.5×10^{-7} (ATLAS)
  - FCC-ee estimation: 10^{-9}
- Key distribution:
  - μ/e mis-id rate: by sacrificing the id efficiency, barely bhabha/$\mu\mu$ surviving (except for muon decay: 10^{-7})
  - Invariant mass: no $\tau\tau$ surviving
- Sensitivity ~ 10^{-10}





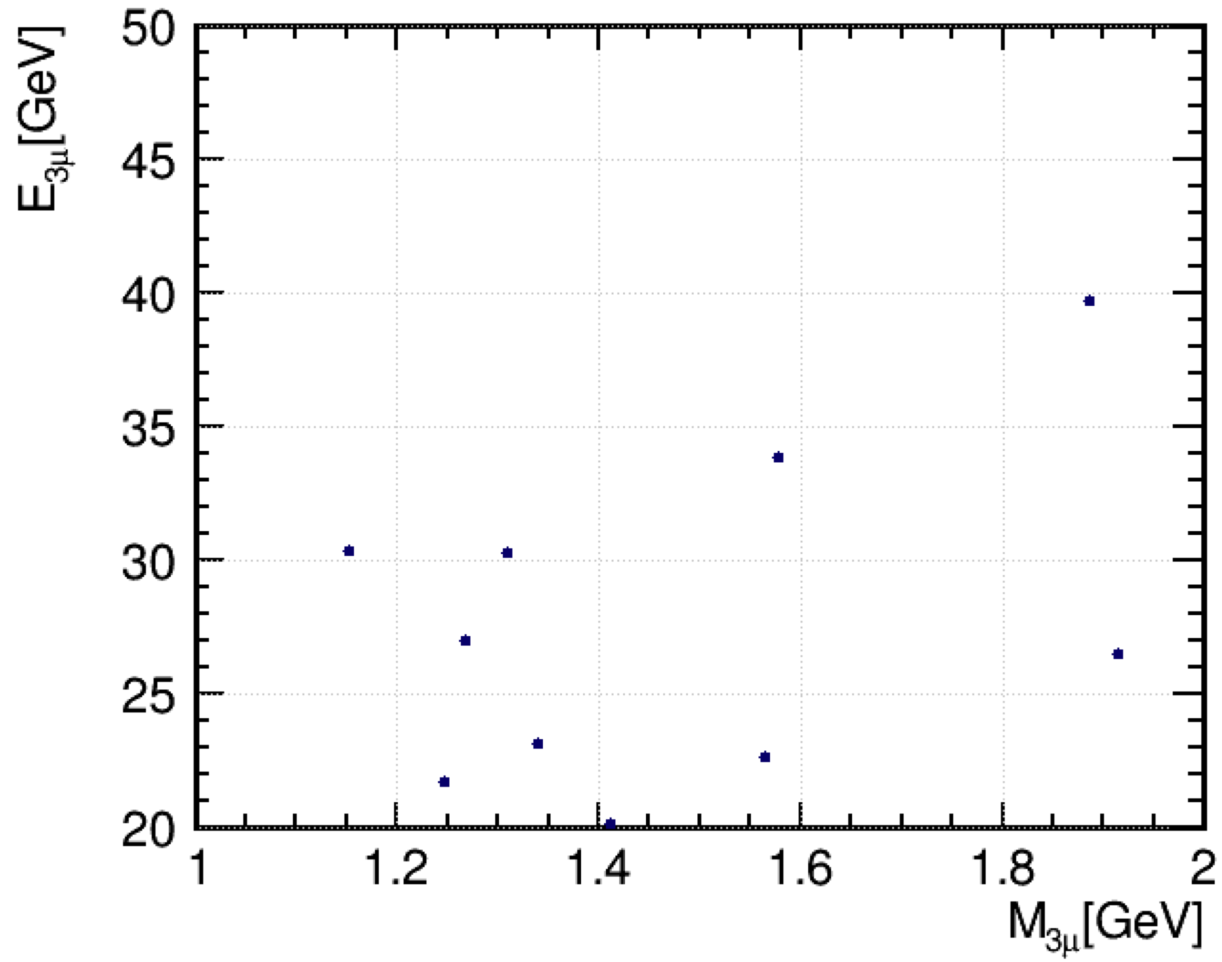
LFV — $\tau \to \mu(e)\gamma$

- Physics background: $Z \to \tau\tau\gamma$, $\tau \to \mu\nu\nu$
- Current bound: $2.7*10^{-8}$ (Babar) FCC-ee estimation: $2*10^{-9}$
- Key distribution: $M(\mu\gamma)$, $E(\mu\gamma)$
  - Signal resolution: $\sigma(m)$ = 26 MeV, $\sigma(E)$=850MeV (Ecal energy resolution $\oplus$ Track momentum resolution $\oplus$ Position resolution, from Mogens' paper)
  - Background surviving: 1*SF ~ 25k
- Sensitivity: 10^{-10}



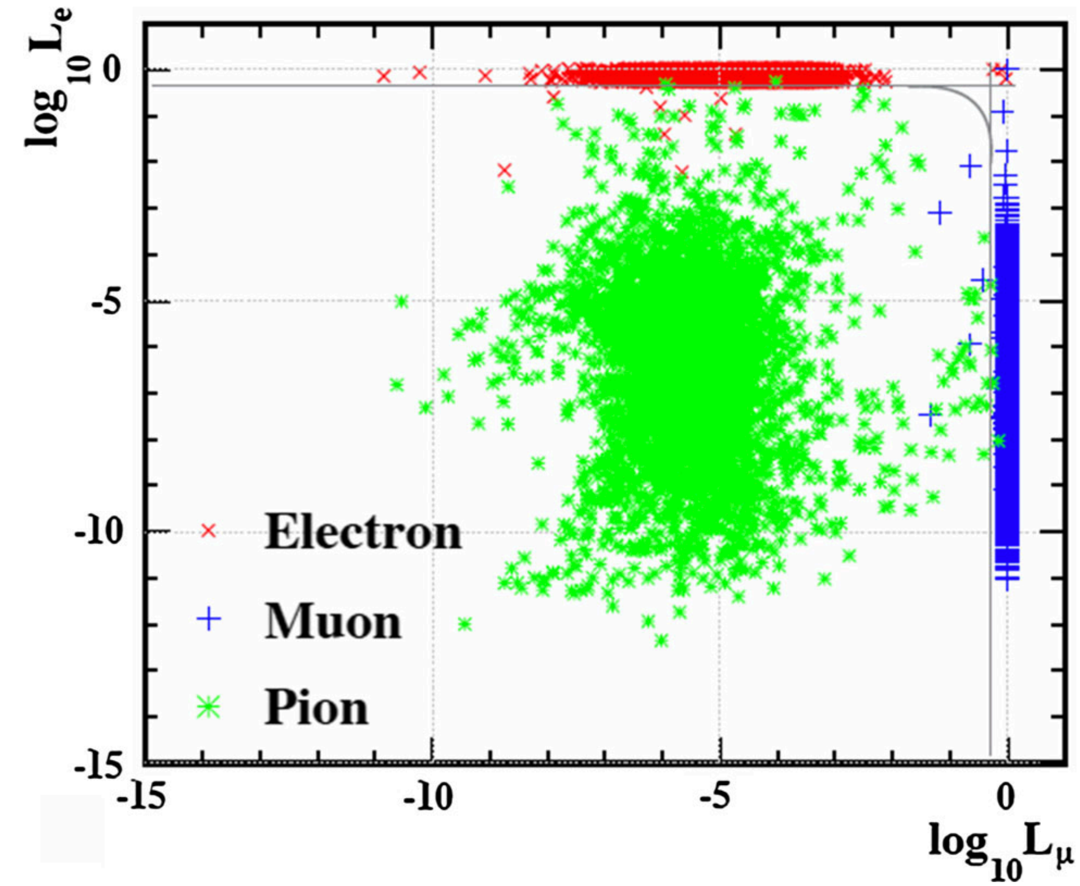
LFV — $\tau \rightarrow 3\mu$

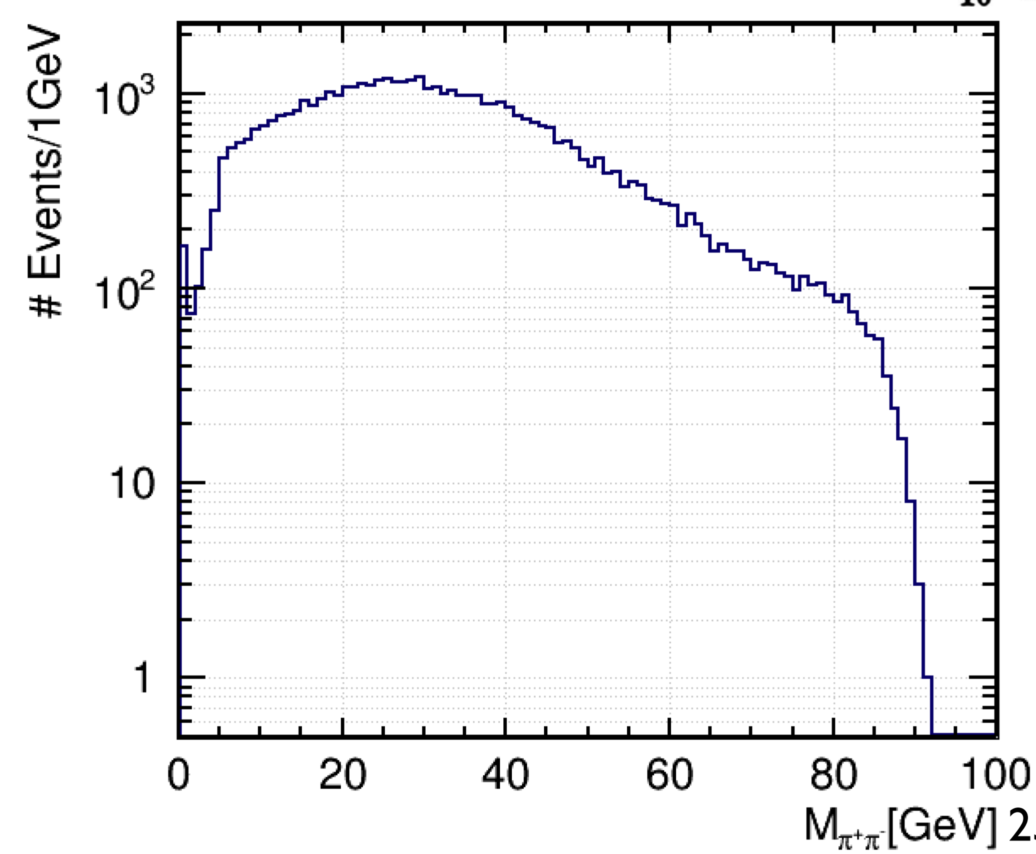
- Main background: free
- Current bound: $2.1*10^{-8}$ (Belle) FCC-ee estimation: 10^{-10}
- Key distribution: M(3μ), E(3μ)
  - Signal resolution: track momentum resolution $\delta p/p \sim 10^{-3}$, a narrow window
  - No background surviving
- Sensitivity: 10^{-10}



Hadronic Z decays — $Z \to \pi\pi$

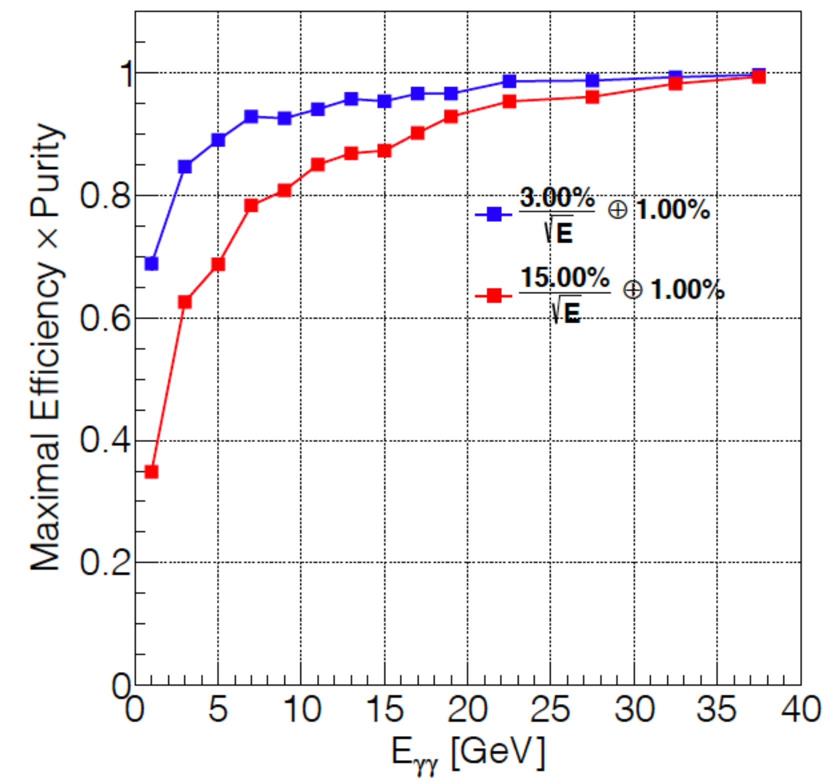
- Physics background: $Z \to \mu\mu$, $Z \to \tau\tau$
- Key distribution:
  - invariant mass
    - Signal resolution: track
    - $Z \to \tau\tau$ surviving: 1*SF
  - mis-id rate:
    - Muon mis-id rate ~ 0
- Sensitivity: 10^{-10}

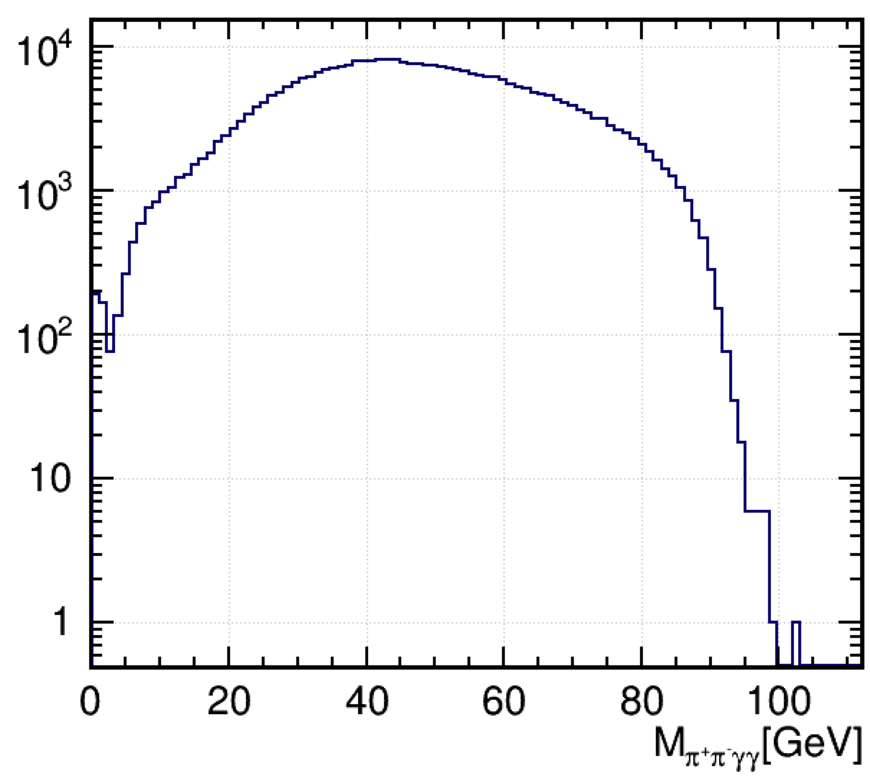




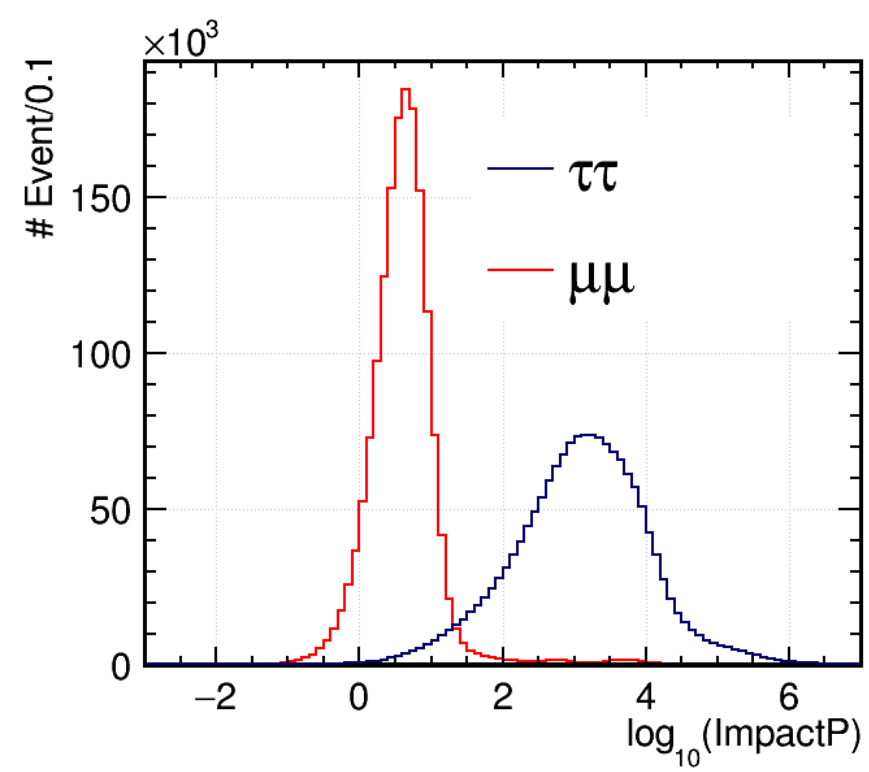
Hadronic Z decay — $Z \to \pi^+\pi^-\pi^0$

- Main background: $Z \to \tau\tau$
- Key distribution:
  - InvM:
    - Signal resolution: $\sigma(m)$ ~ sub MeV
    - Background reduced to 10^{-4}
  - Impact parameter: reduce 10%
  - M_{12} & M_{23} (M>Mtau): if Dalitz plot predicted
    - bkg reduced to 10^{-7} ~ 10k
- Sensitivity: 10^{-9}



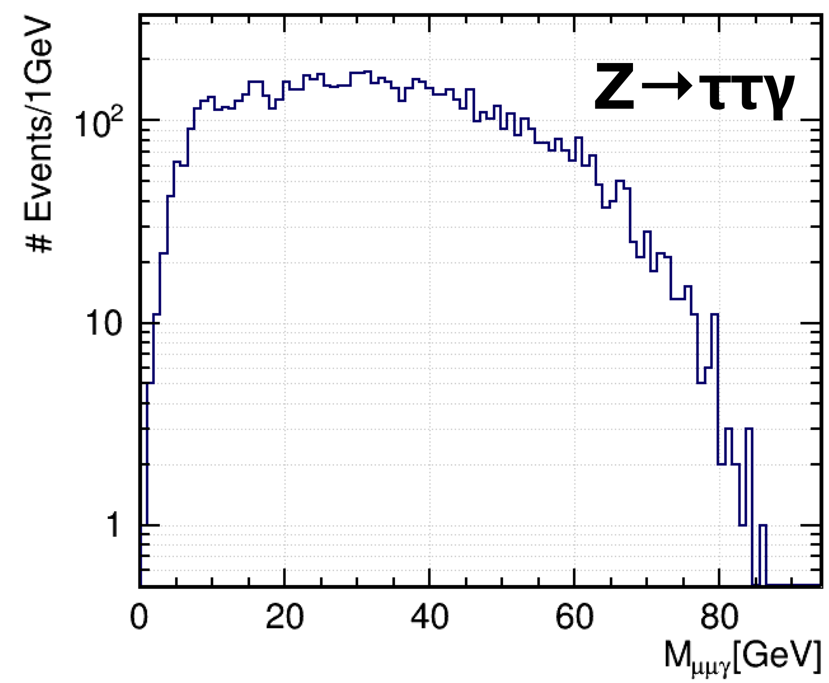


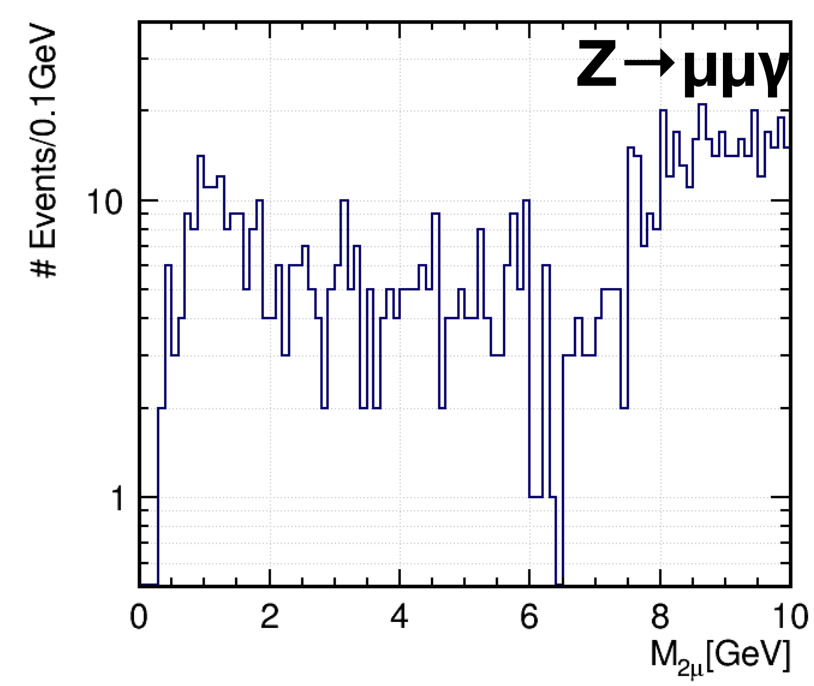
VS2021



Radiative Z decay — $Z \to J/\psi\ \gamma$, $J/\psi \to \mu^+\mu^-$

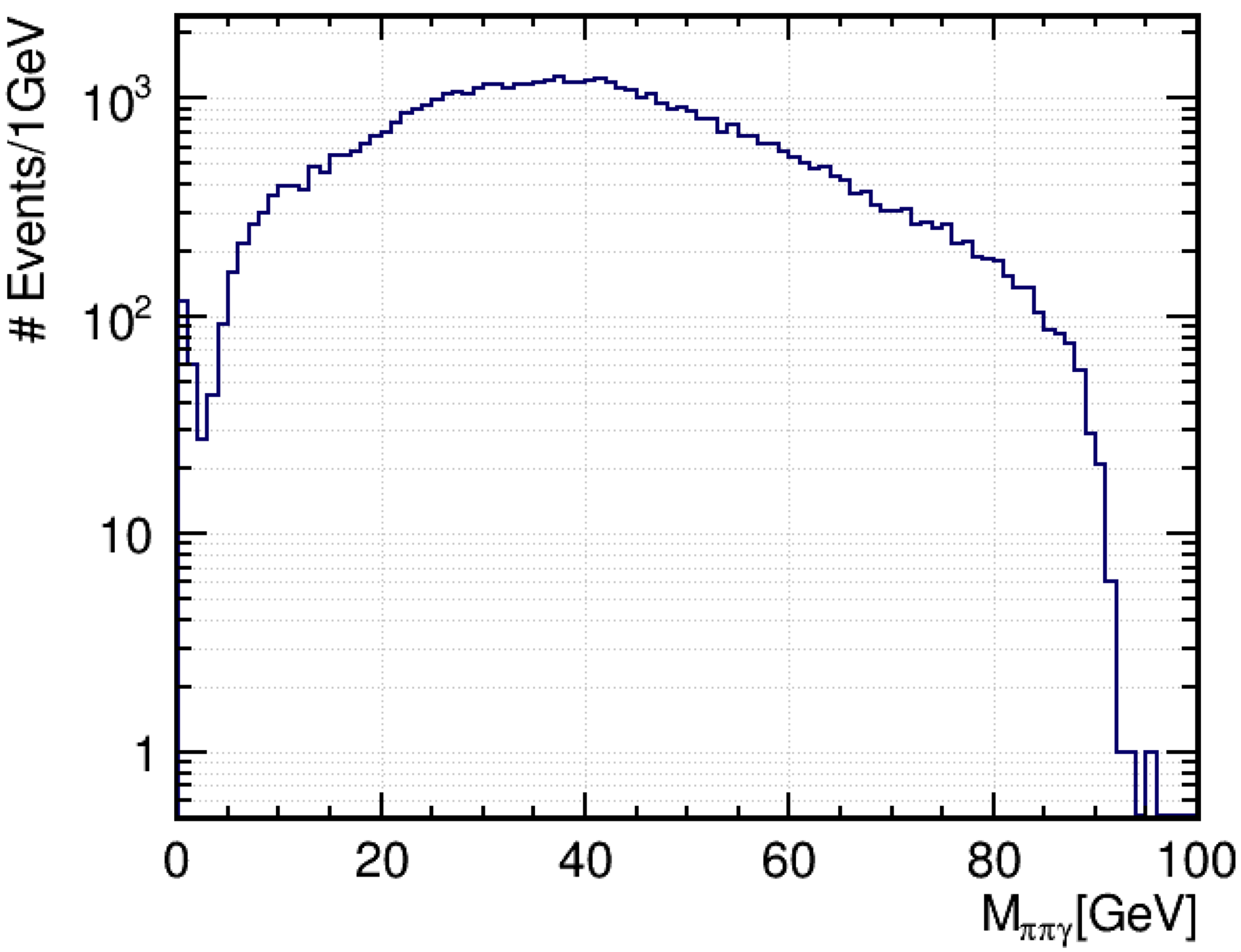
- Main background: $Z \to \tau\tau\gamma$, $\mu\mu\gamma$
- Current bound: $2.6*10^{-6}$(ATLAS)
  White paper prediction: $8*10^{-8}$
- Key distribution:
  - total invariant mass:
    - 1*SF $Z \to \tau\tau\gamma$ surviving
  - di-muon invariant mass:
    - 3*SF $Z \to \mu\mu\gamma$ surviving
  - impact parameter:
    - reduce 10%
- Sensitivity: 10^{-9}~10^{-10}





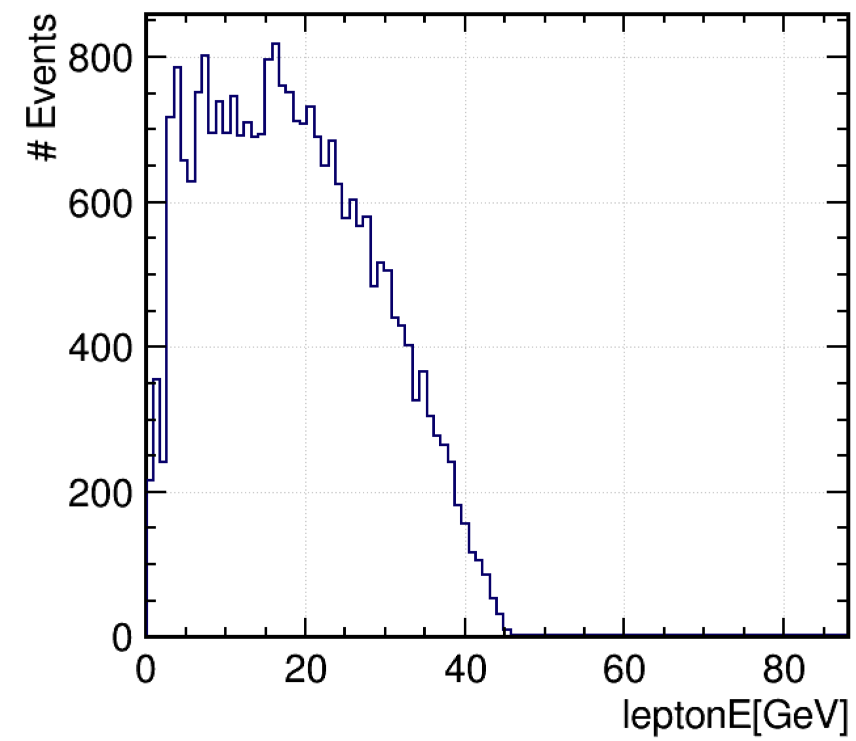
Radiative Z decay — $Z \to \rho\gamma$, $\rho \to \pi^+\pi^-$

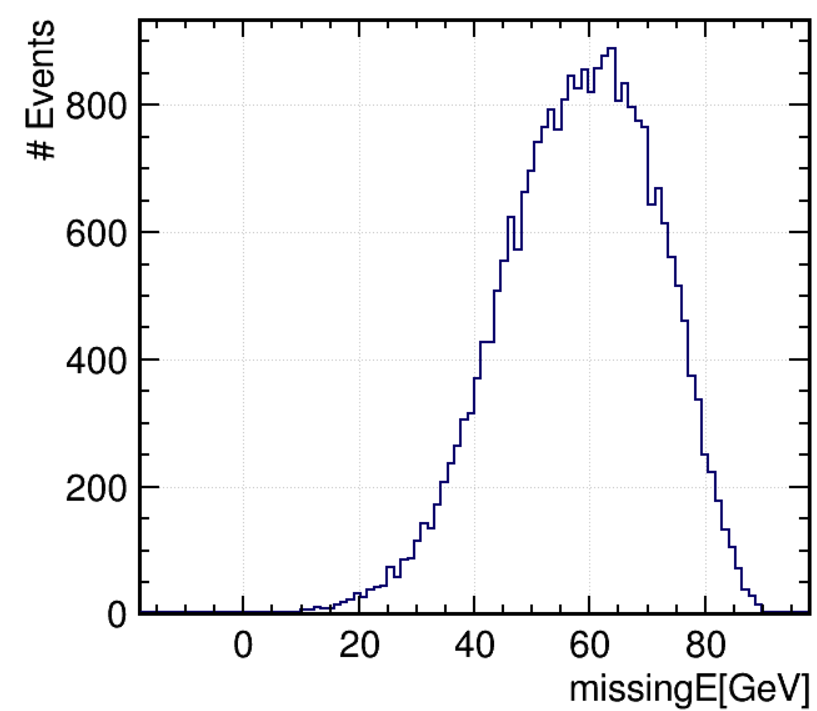
- Main background: $Z \to \tau\tau\gamma$
- White paper prediction: 4×10^{-9}
- Key distribution:
  - total invariant mass:
    - 10*SF $Z \to \tau\tau\gamma$ surviving
  - impact parameter:
    - reduce10%
- Sensitivity: 10^{-9}



Weak Radiative Z decay — $Z \to \pi^{+/-} W^{-/+}$ (leptonic)

- Main background: $Z \to \tau\tau$, one $\tau \to \pi\nu$, the other $\tau \to l\nu\nu$
- Current bound: $7.0 * 10^{-5}$(LEP) White paper prediction: 10^{-10}
- Key distribution:
  - Acoplanarity, Missing E, lepton E: assuming bkg reduce rate same order as LEP
  - impact parameter: reduce 10%, ~10k bkg surviving
- Sensitivity: 10^{-10}





Summary

Channel	$Z\to\tau\mu$	$Z\to\mu e$	$\tau\to\mu\gamma$	$\tau\to 3\mu$	$Z\to\pi\pi$	$Z\to\pi^+\pi^-\pi^0$	$Z\to J/\psi\,\gamma$	$Z\to\rho\gamma$	$Z\to\pi^{+/-}W^{-/+}$
Current Bounds/ BR prediction	$1.2*10^{-5}$	$7.5*10^{-7}$	$4.4*10^{-8}$	$2.1*10^{-8}$	10^{-12}	10^{-8} ~10^{-5}	$2.6*10^{-6}$	10^{-9}	$7.5*10^{-5}$
Earlier Estimation	10^{-9}	10^{-9}	10^{-9}	10^{-10}	-	-	10^{-8}	10^{-9}	10^{-10}
FullSim Estimation	10^{-9}	10^{-9}	10^{-10}	10^{-10}	10^{-10}	10^{-9}	10^{-9} ~10^{-10}	10^{-9}	10^{-10}

Summary

- For lepton identification, the performance of jet leptons is slightly degraded (1-2%) compare to the isolate case, due to the clustering confusion
- Application: FullSim result consistent with the earlier estimation
    - mostly background free, great potential of CEPC Flavor
    - requirements on detector & algorithm
    - need more theorists' interpretations

Thank you!