Strange-quark Tagging for Higgs and EW Physics

Caterina Vernieri on behalf of the HtoSS group (twiki)

Valentina Cairo, Taikan Suehara, Loukas Gouskos, Matt Basso, John Alison, Yotam Soreq, Valerio Dao, Karsten Koeneke

SLAC NATIONAL ACCELERATOR LABORATORY

October 17, 2024

Stanford University | U.S. DEPARTMENT OF ENERGY

1811.00017
1908.11376
2101.04119

SM, EFT, and beyond

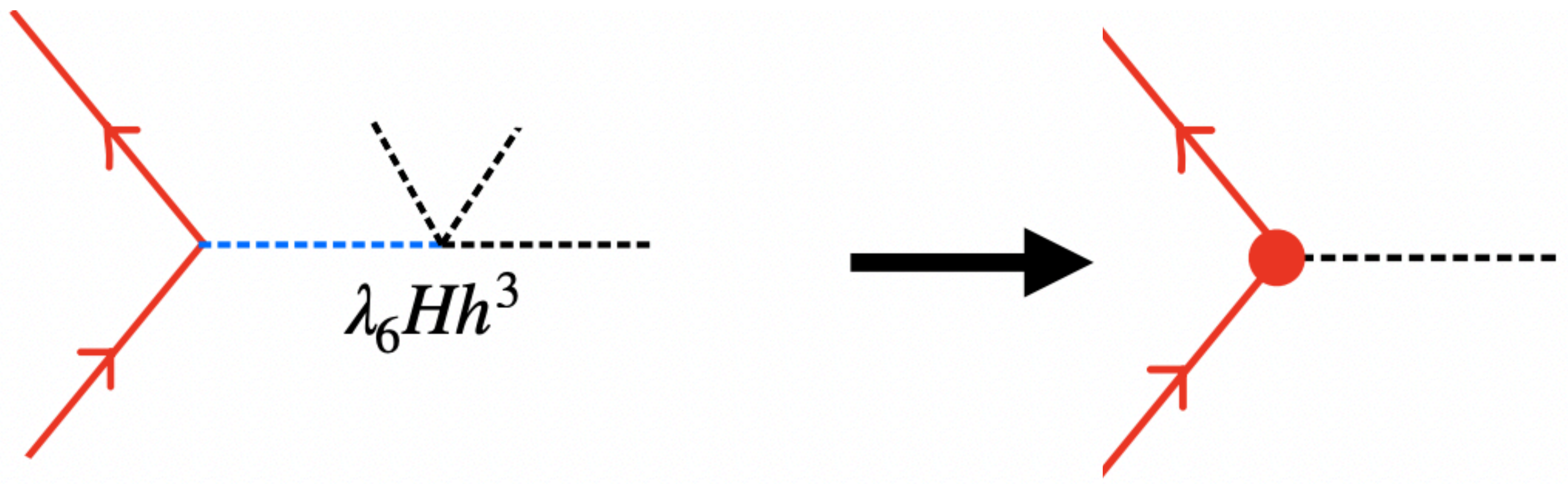
Higgs to strange coupling is an appealing signature to probe new physics

SM measurement, necessity to define BR(H →ss) on the TH side ⇒ LHCHWG M. Spira

Is the Higgs the source for all flavor?

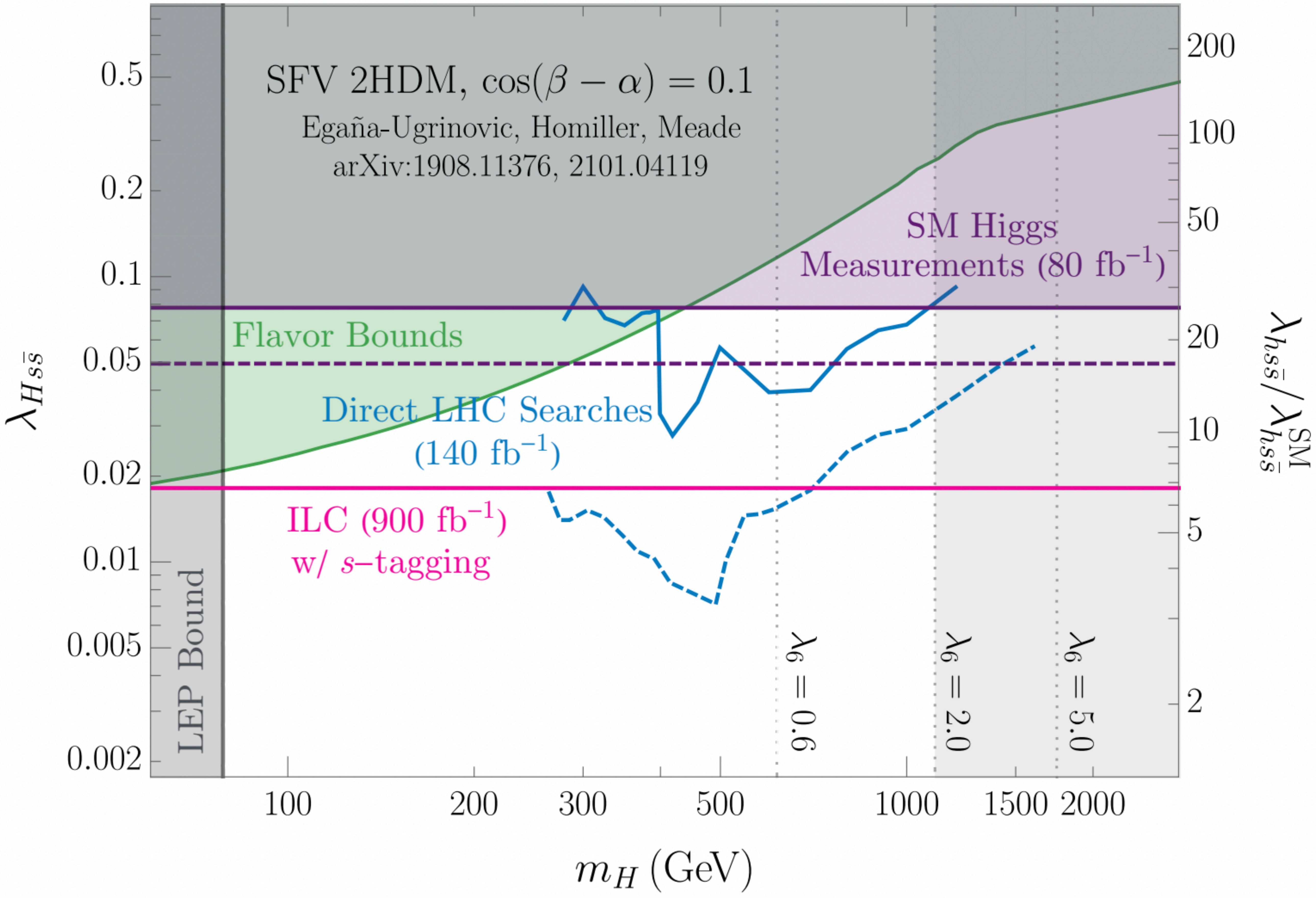
In a **Spontaneous Flavor Violation** model new physics can couple in a strongly flavor dependent way if it is aligned in the down-type quark or up-type quark sectors

- It allows for large couplings of additional Higgs to strange/light quarks
- No flavor-changing neutral currents



$\sim \frac{1}{\Lambda^2}(sh\bar{s})h^2$

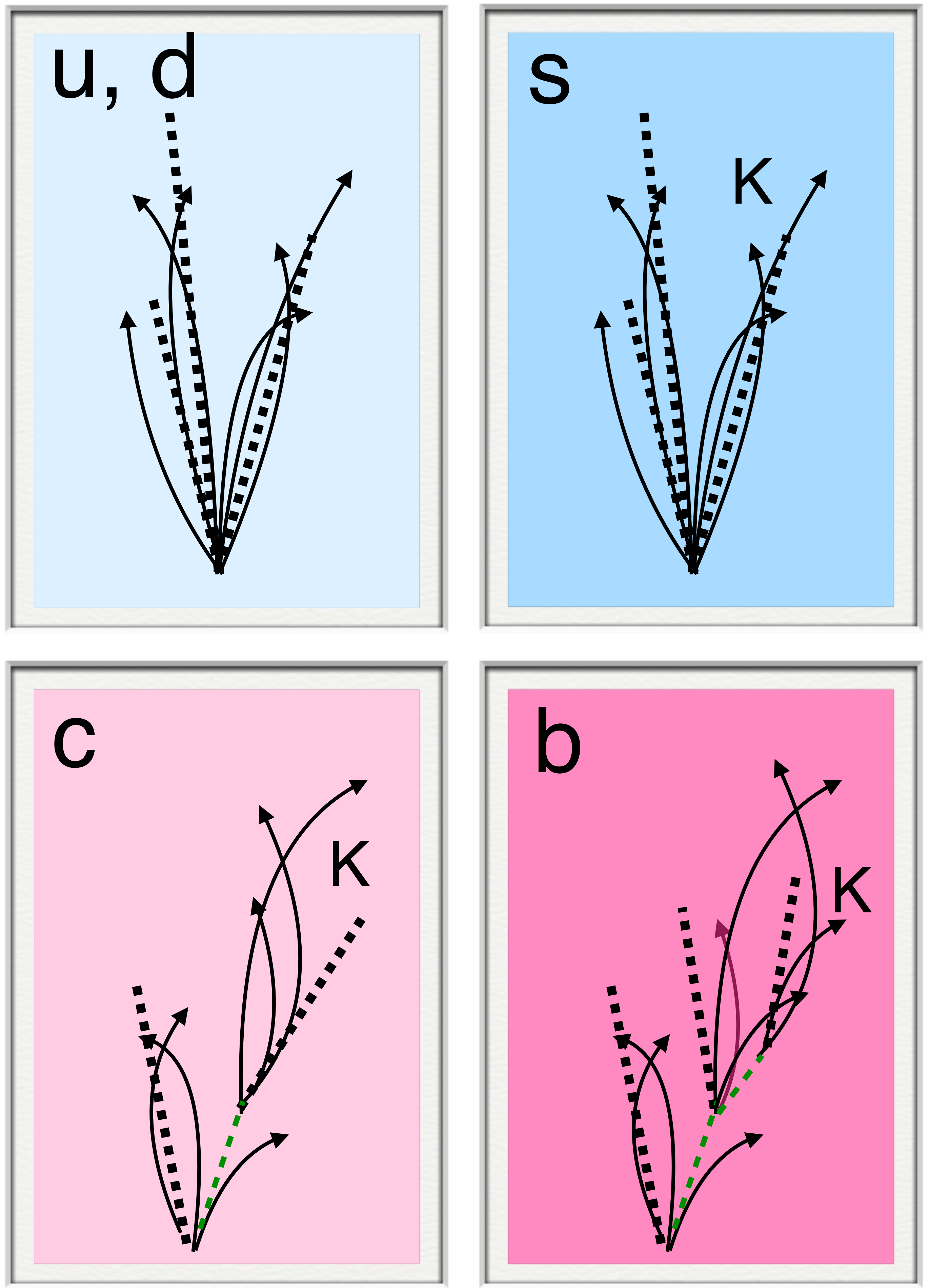
P. Meade



2101.04119
2203.07535

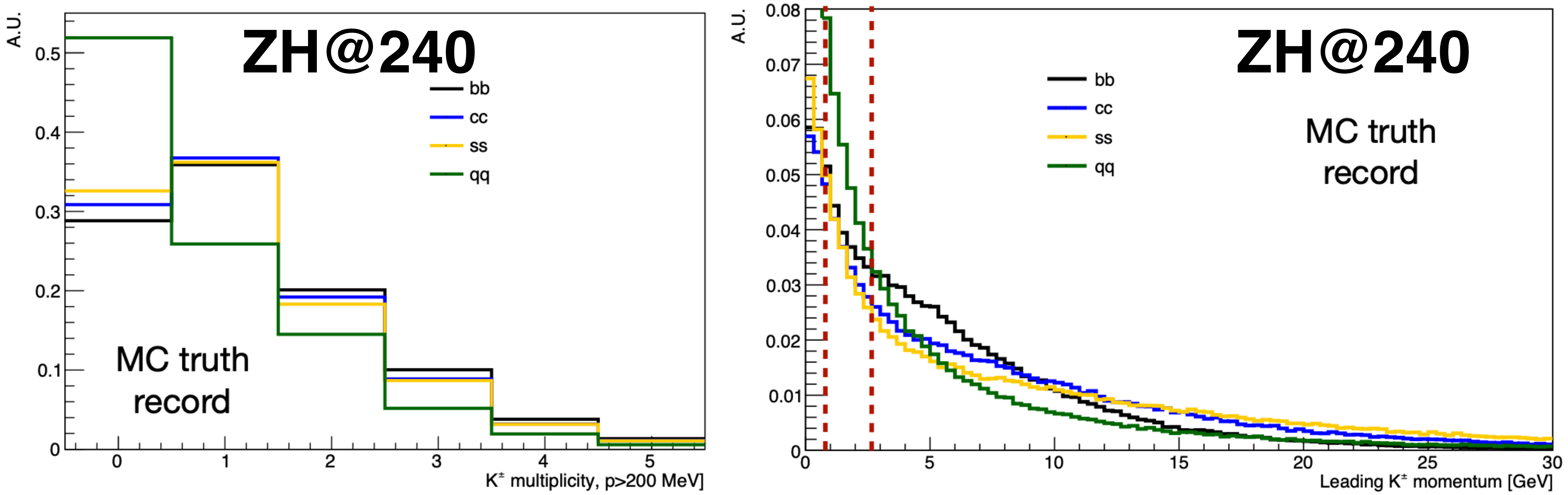
s-tagging

Tagging strange is a challenging but not impossible task for future detectors at e^+e^-



- As b,c, and s jets contain at least one strange hadron
- Strange quarks mostly hadronize to prompt kaons which carry a large fraction of the jet momentum
- Strange hadron reconstruction:
  - $K^{\pm}$ PID
  - K^0_L PF (neutral)
  - $K^0_S \to \pi^+\pi^-$ (~70%) / $\pi^0\pi^0$ (~30%)
  - $\Lambda^0 \to p\pi^-$ (~65%)

Distinctive two-prong vertices topology



A. Sciandra ECFA 2024

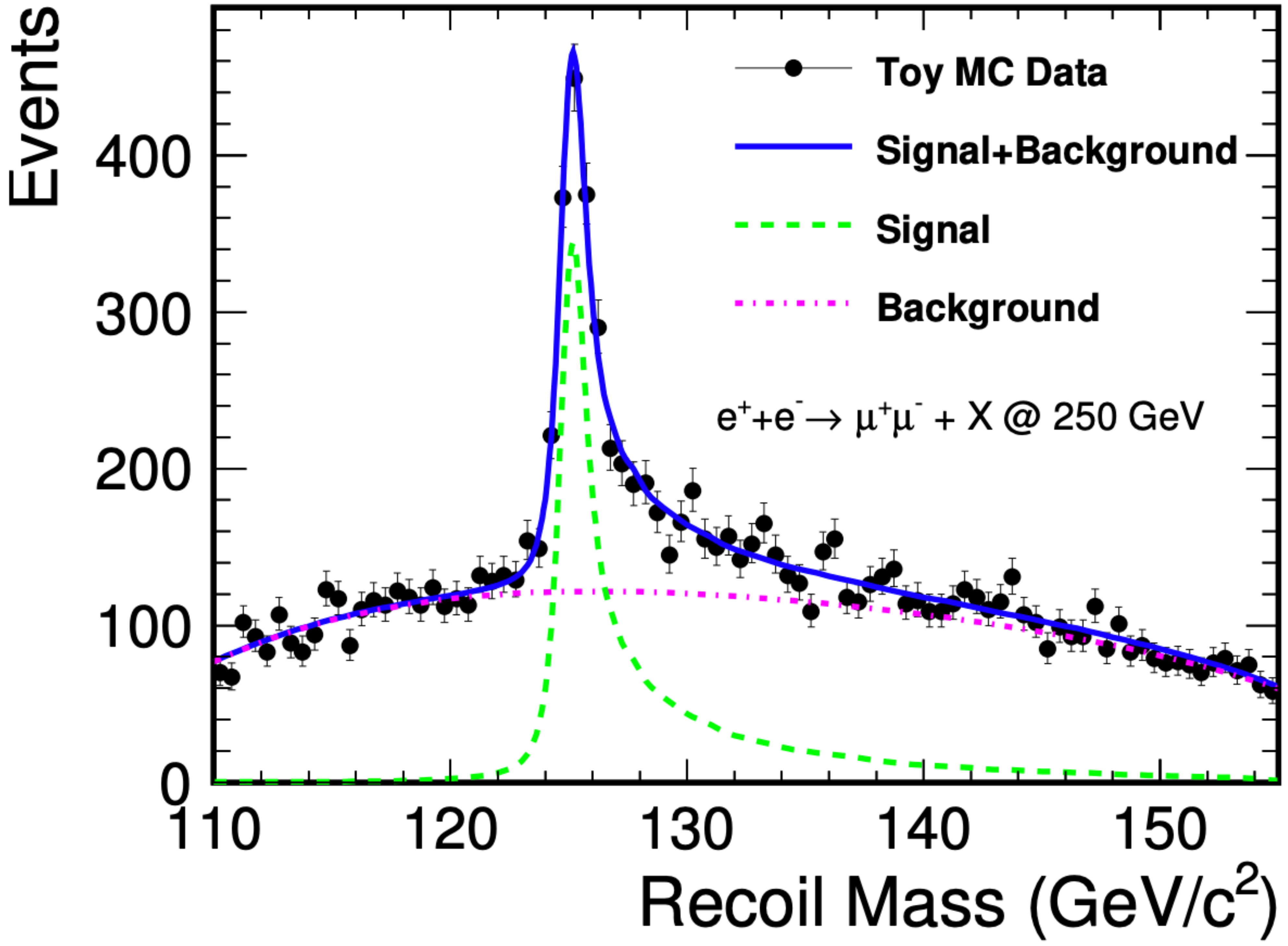
arXiv:2003.01116

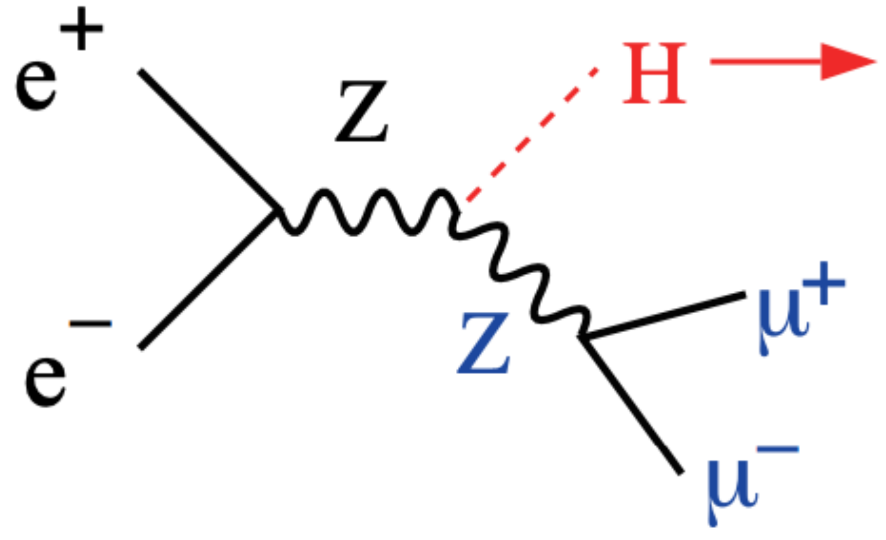
Detectors at future e^+e^-

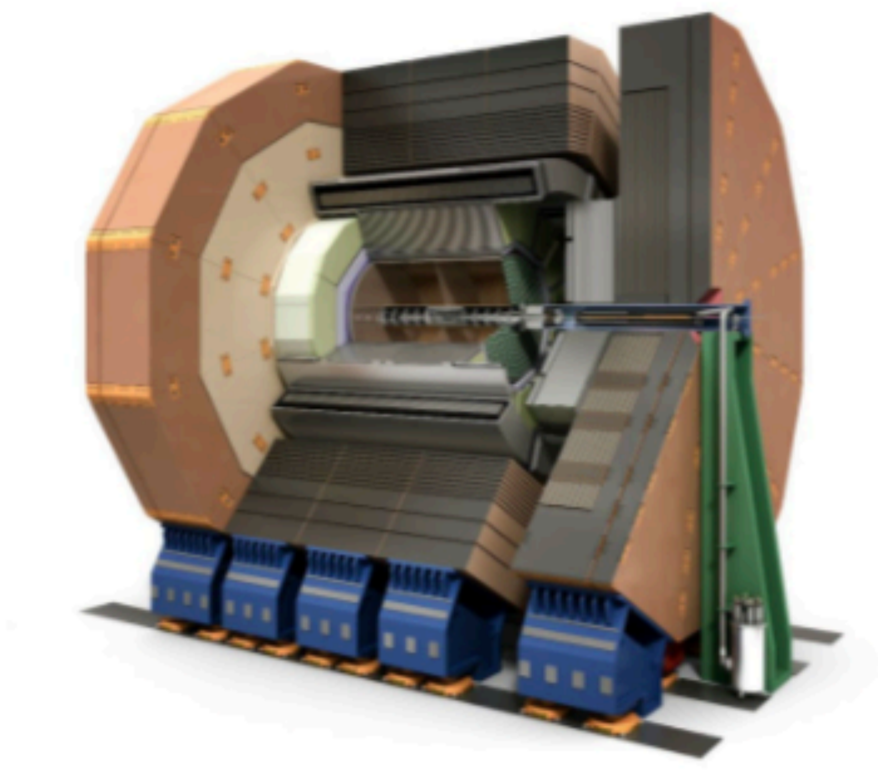
Stringent detector requirements from ZH reconstruction

Detector designs at e^+e^- colliders are converging to very similar strategies

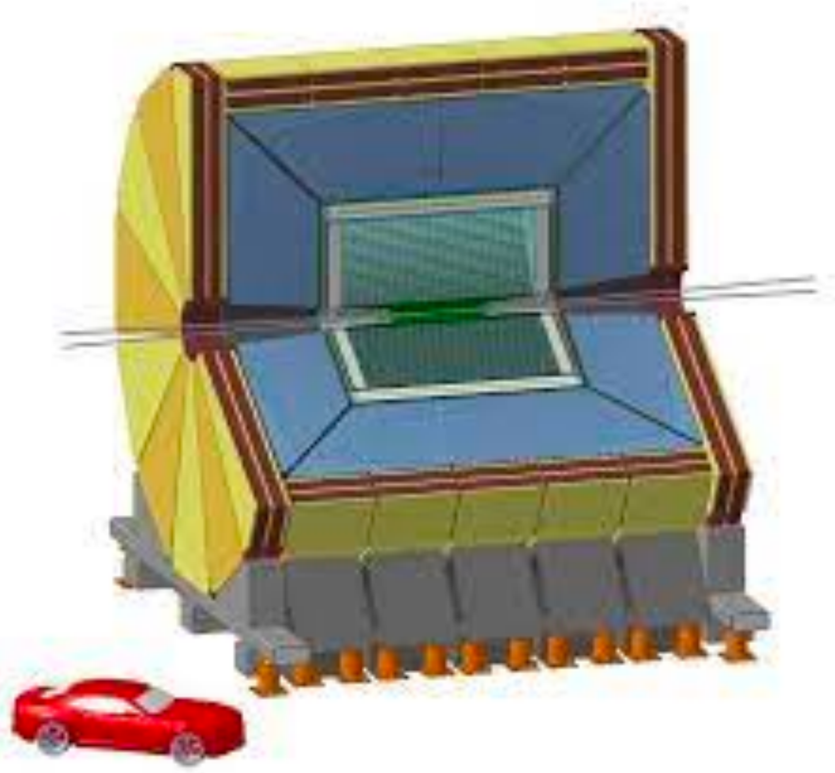
- Strong magnetic field 2-5 T
- (Ultra) low material budget tracker (<0.3% X0)
    - Close to the interaction region (10-25 mm)
- High granularity calorimetry
    - Particle Flow reconstruction → plays a big part in many designs



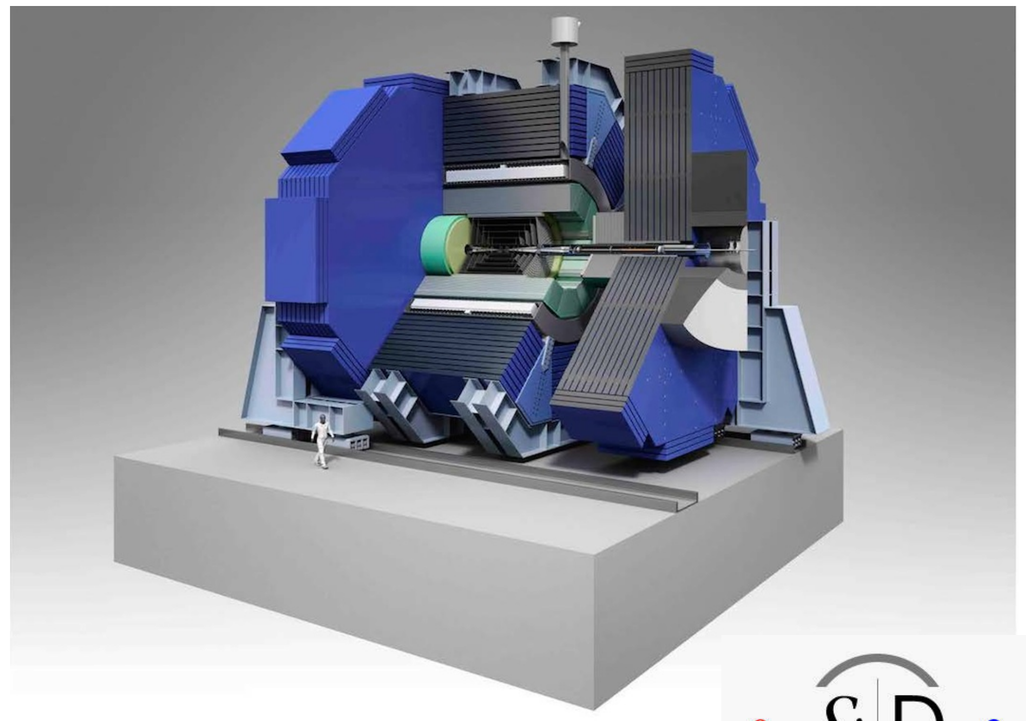




ILD



IDEA

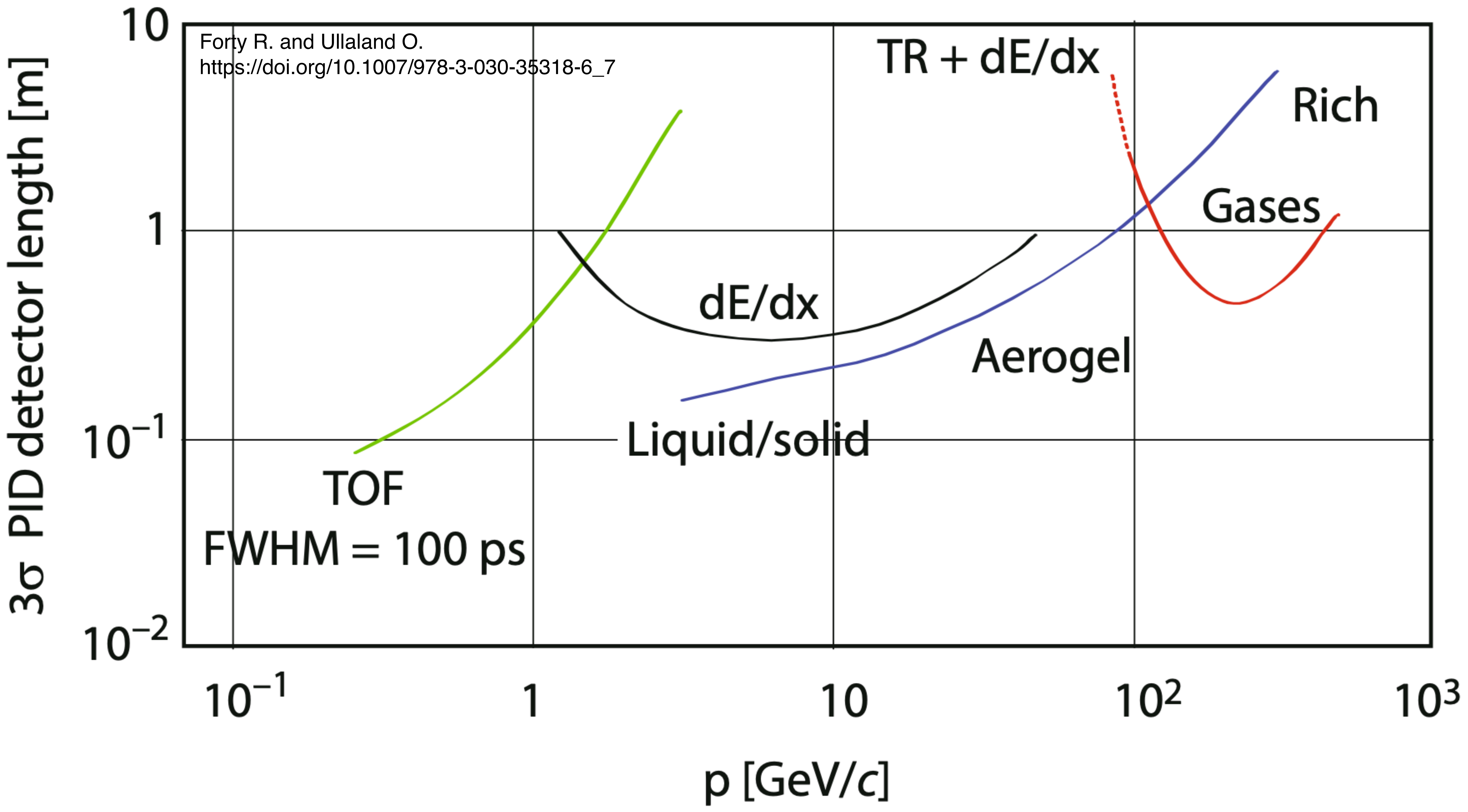


SiD

1912.04601
e2019-900045-4

Particle ID for s-tagging

Combining different strategies for optimal PID performance across a wide p_T range

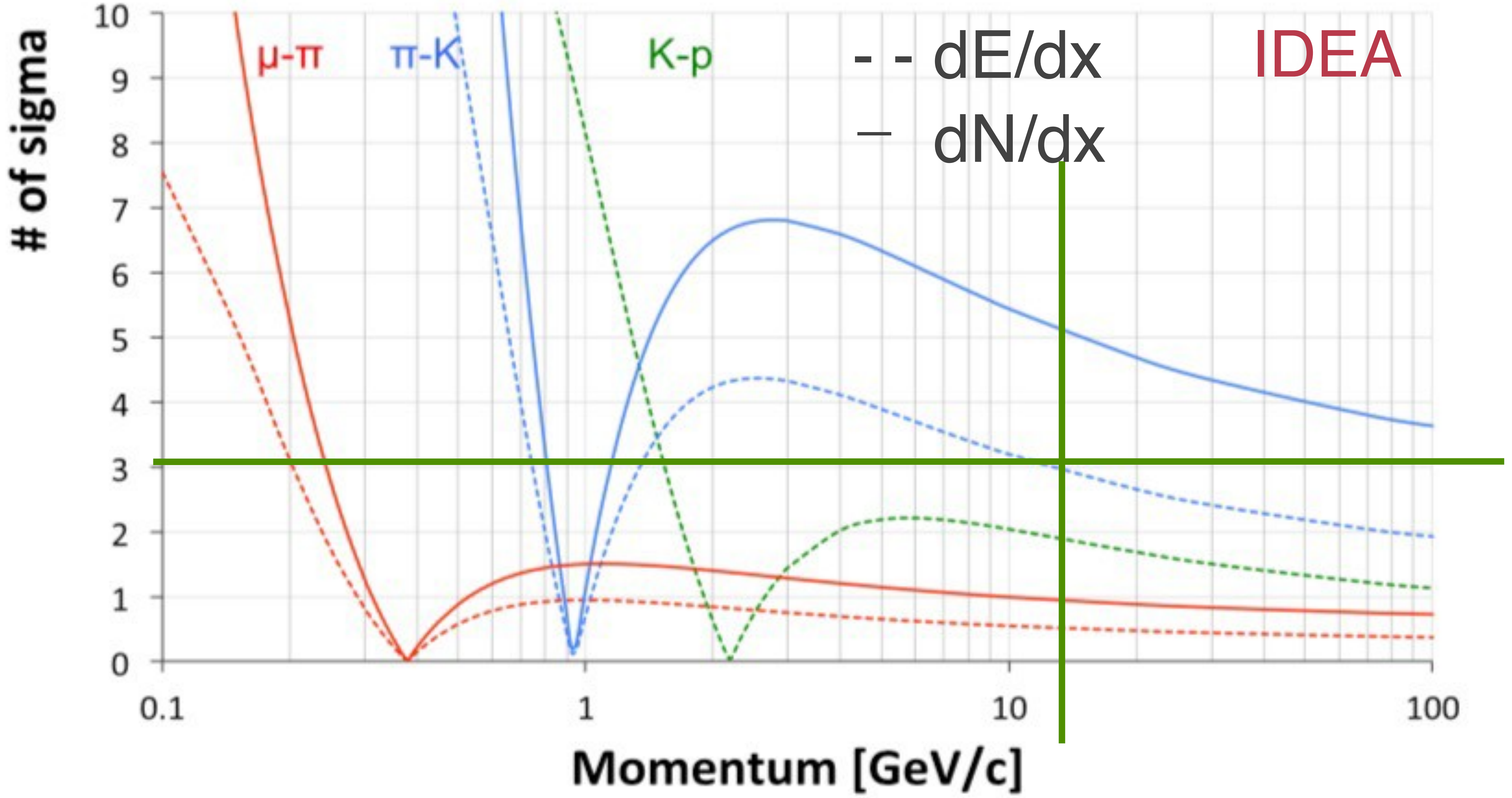


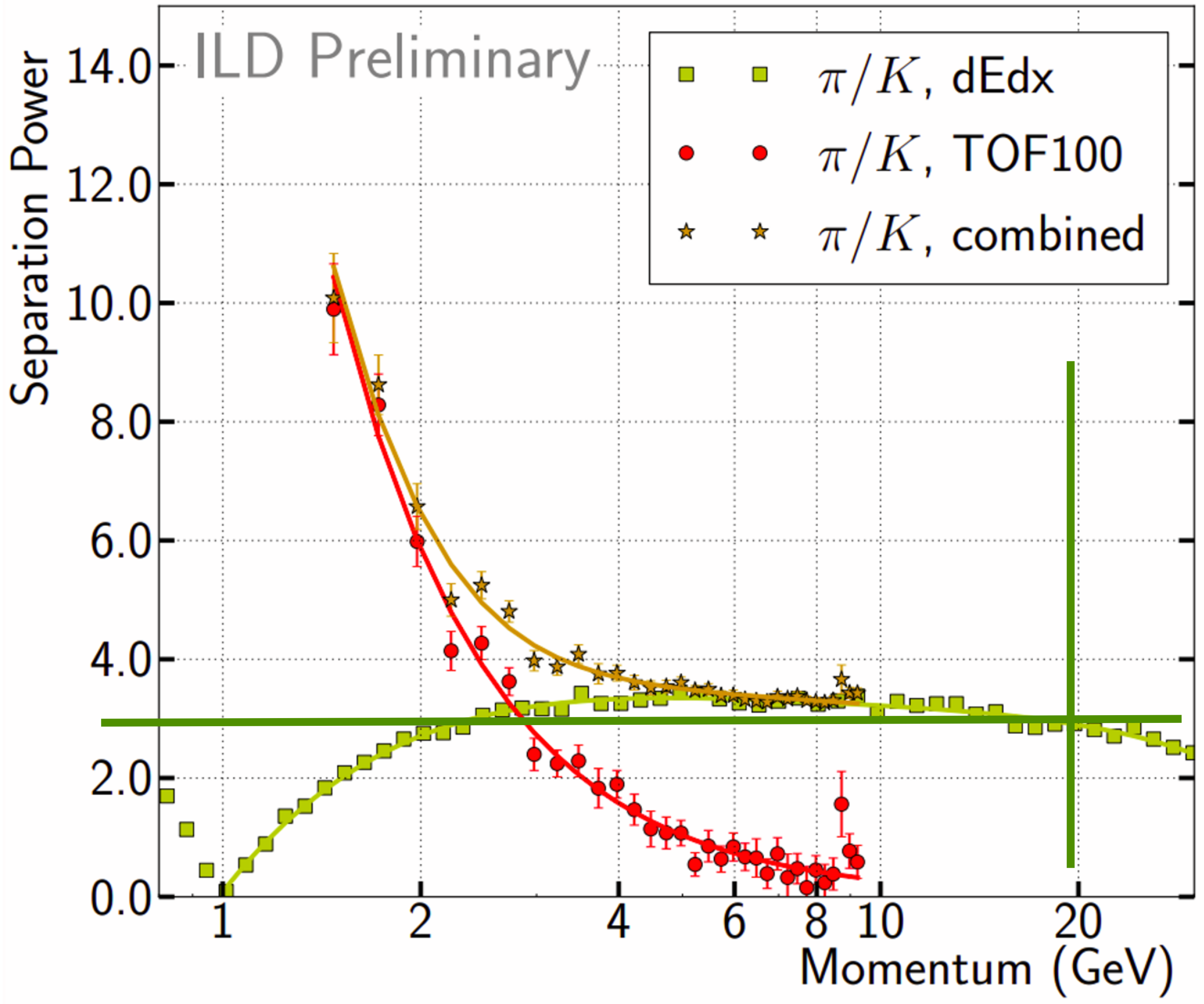
1912.04601
e2019-900045-4

Particle ID for s-tagging

Combining different strategies for optimal PID performance across a wide p_T range

- dE/dx from silicon (< 5 GeV) and large gaseous tracking detectors (< 30 GeV)
- < 5 GeV, time-of-flight (i.e. 100 ps from ECAL)



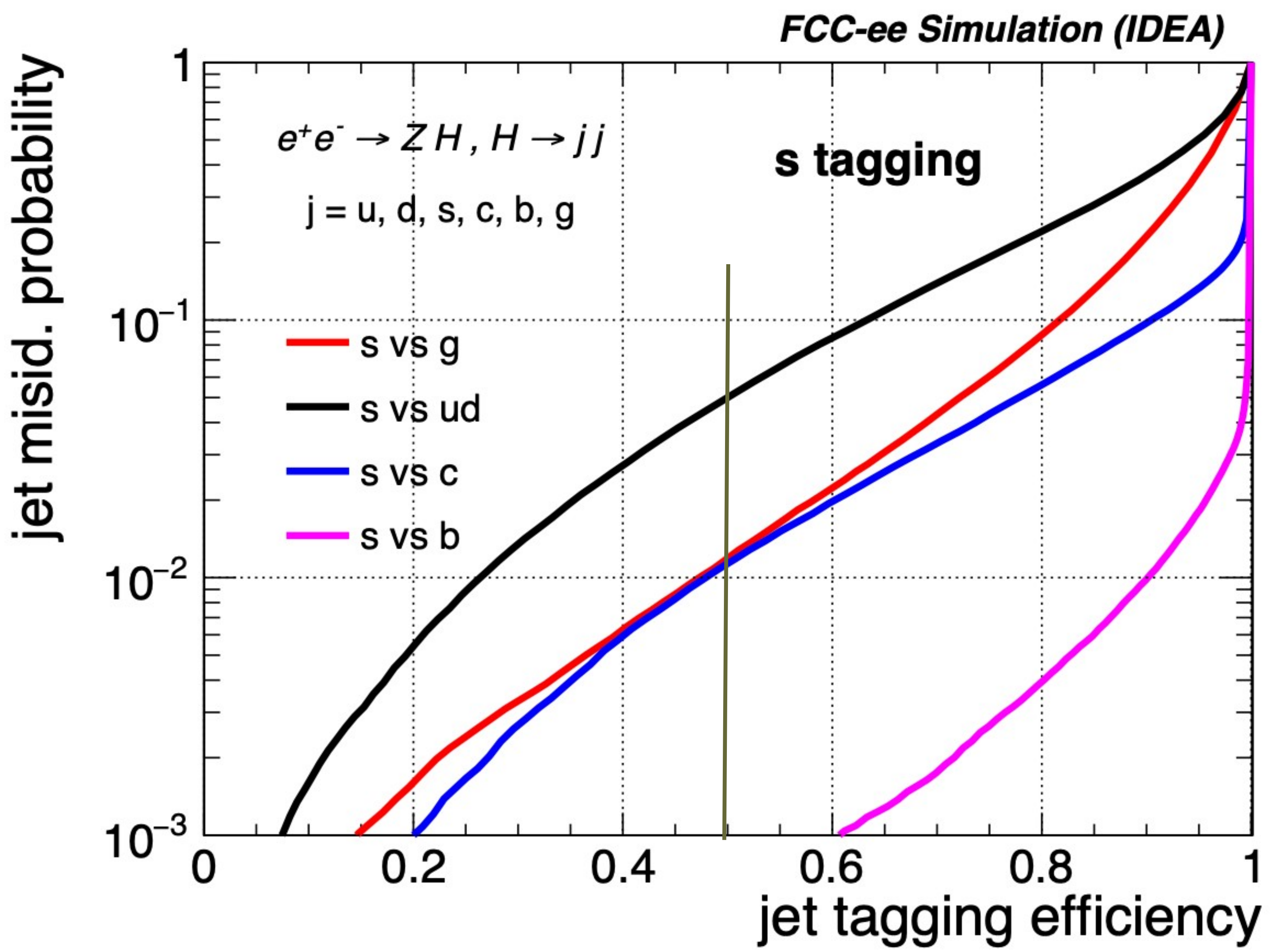


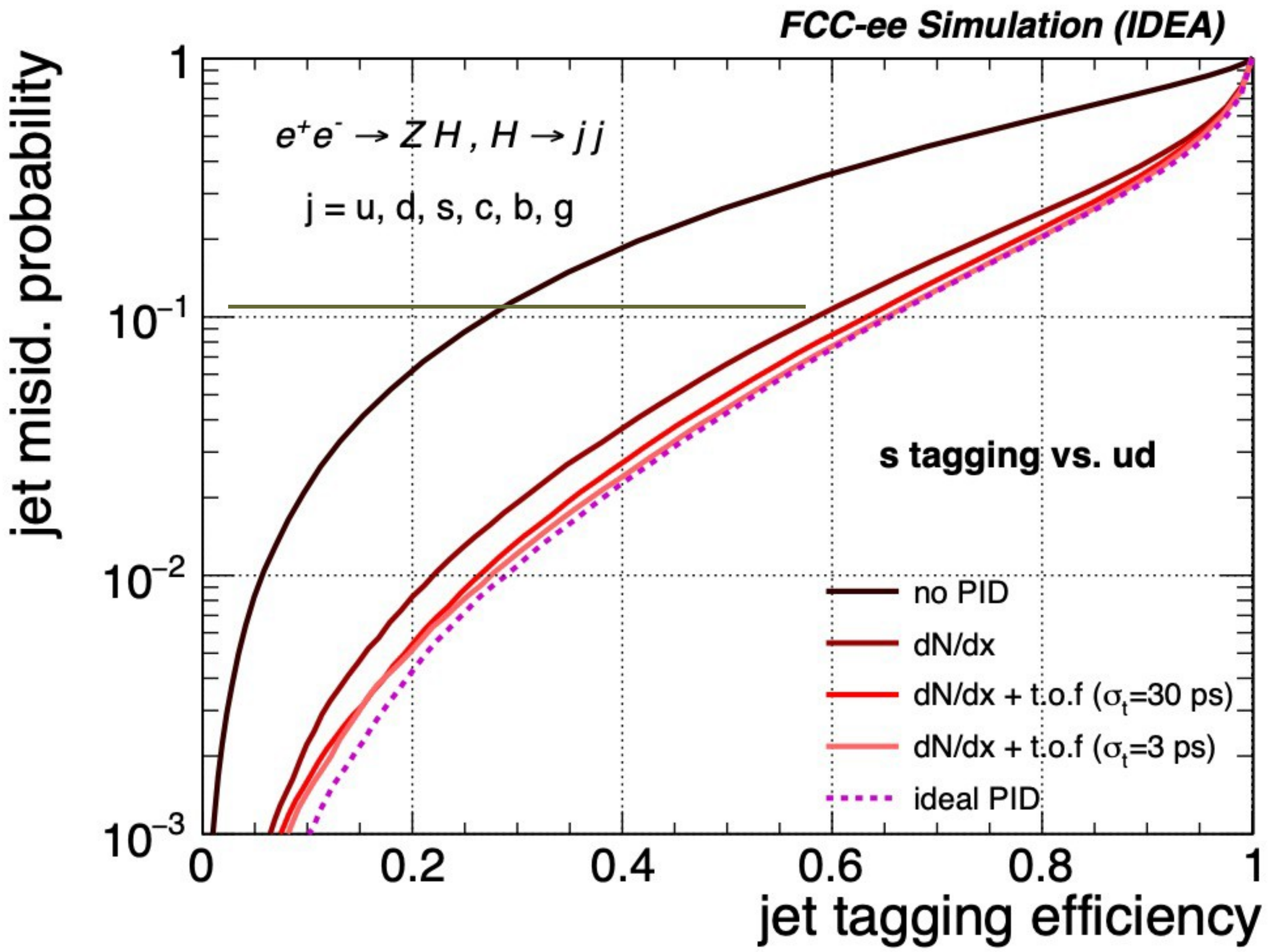
PRD 101 056019 (2020)
EPJ C 82 646 (2022)
L. Gouskos @FCC week

Strange tagging performance

IDEA-like detector and Particle cloud graph neural network (fast sim)

- Both TOF and dN/dx ($3\sigma < 30$ GeV) included as inputs
- No PID to PID with dN/dx → at fixed mistag, efficiency doubles





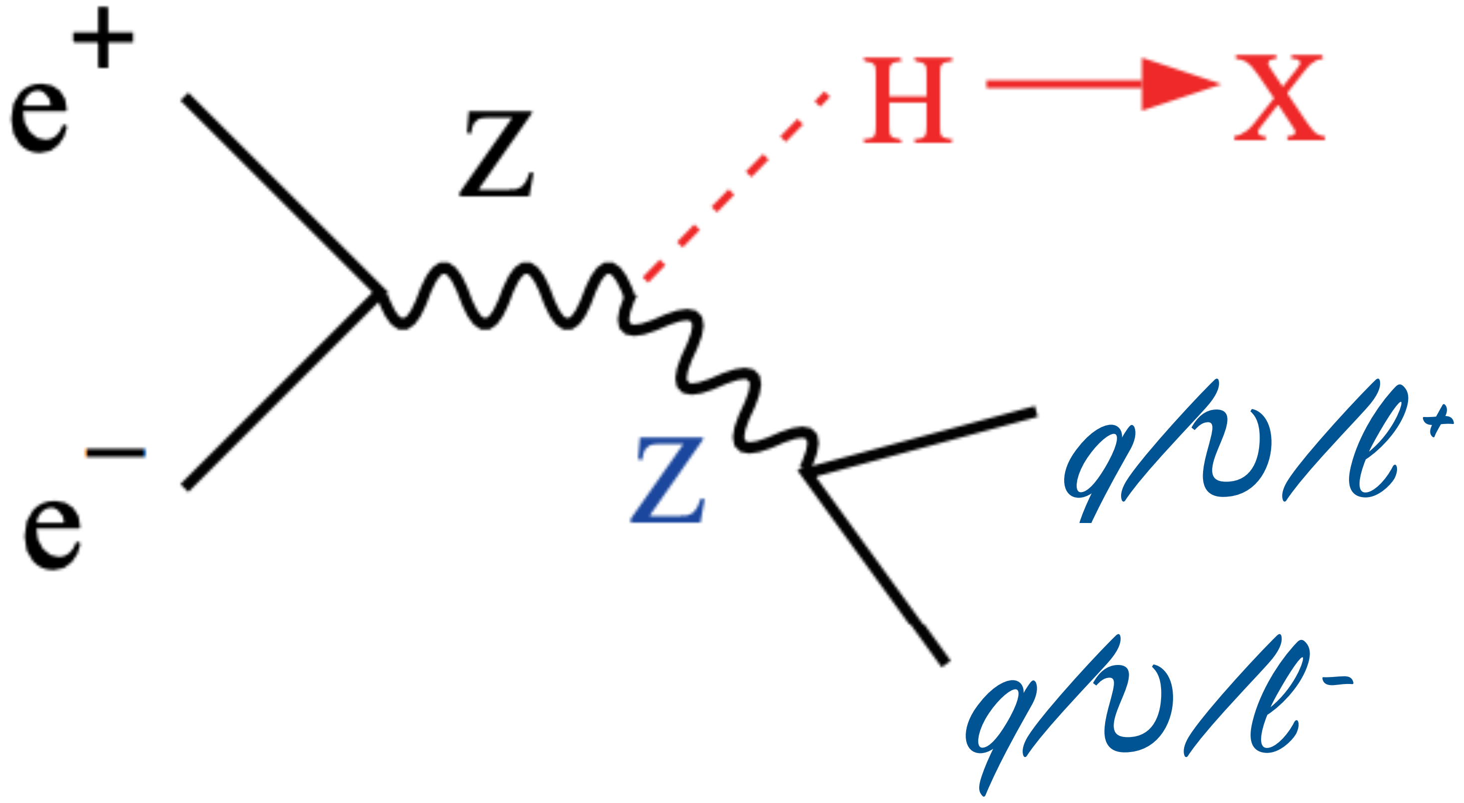
WP	Eff (s)	Mistag (g)	Mistag (ud)	Mistag (c)	Mistag (b)
Loose	90%	20%	40%	10%	1%
Medium	80%	9%	20%	6%	0.4%

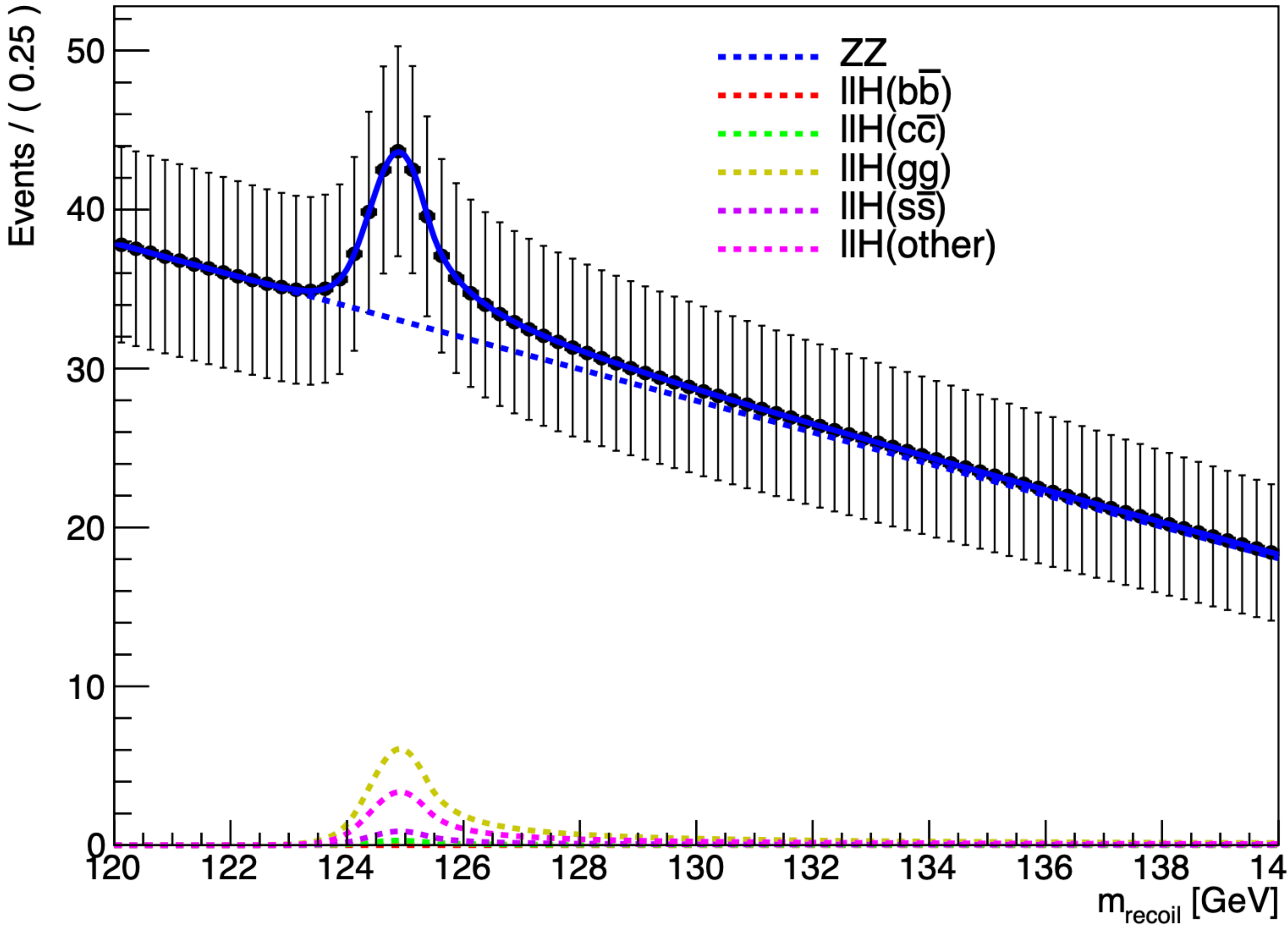
arXiv:2203.07622
L. Gouskos @FCC week

Analysis strategy to target H→ss

Exploit Z boson reconstruction in the ZH associated mode

- At 240/250 GeV the total Zh cross section can be extracted independently of the Higgs boson's detailed properties by counting events with an identified Z boson
- Looking at 0 or 2 leptons Z decay modes
  - **New** fully hadronic final state





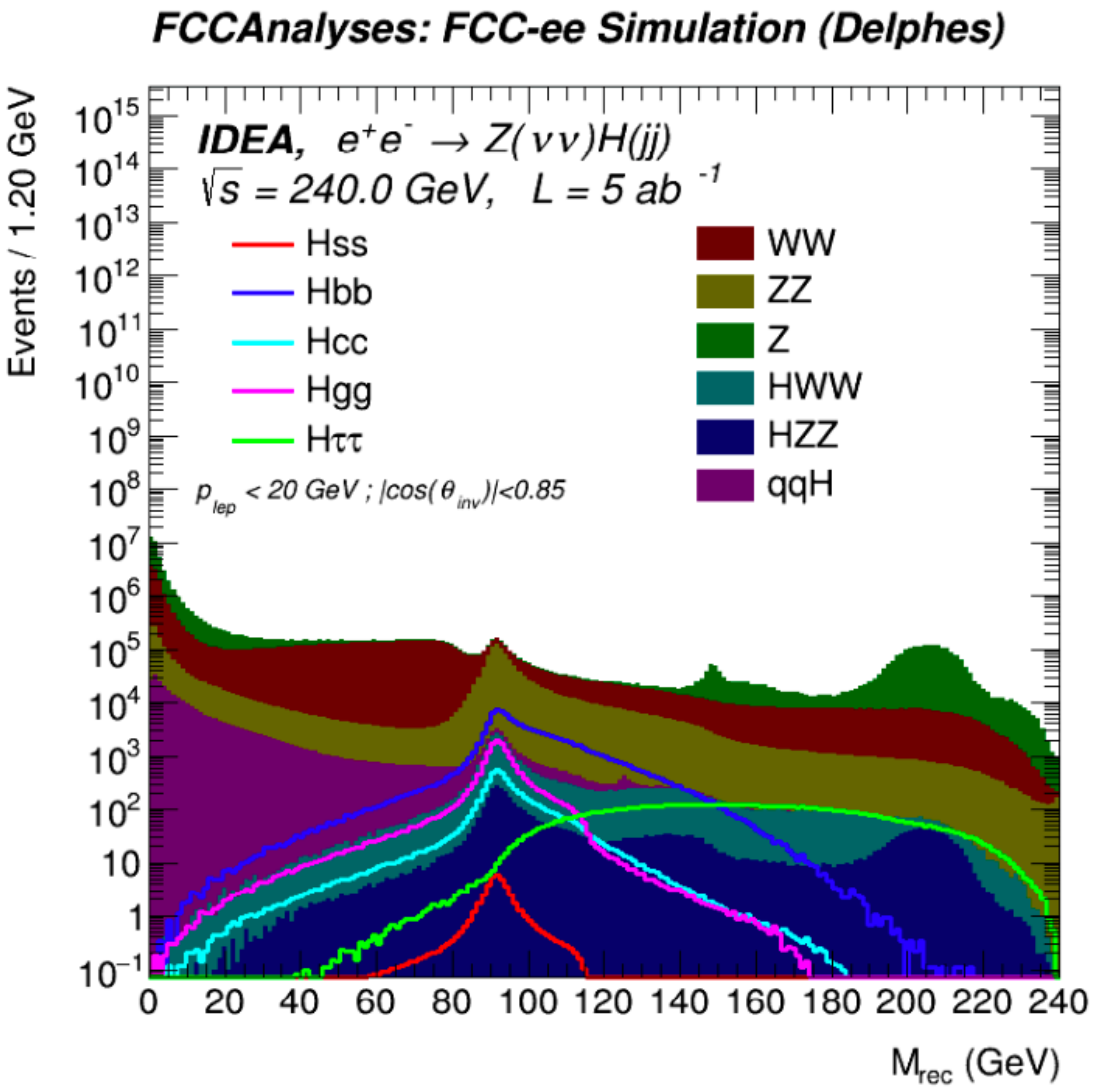
$$M_{\mathrm{rec}} = \sqrt{(\sqrt{s} - E_Z)^2 - \vec{p}_Z^{\,2}}$$

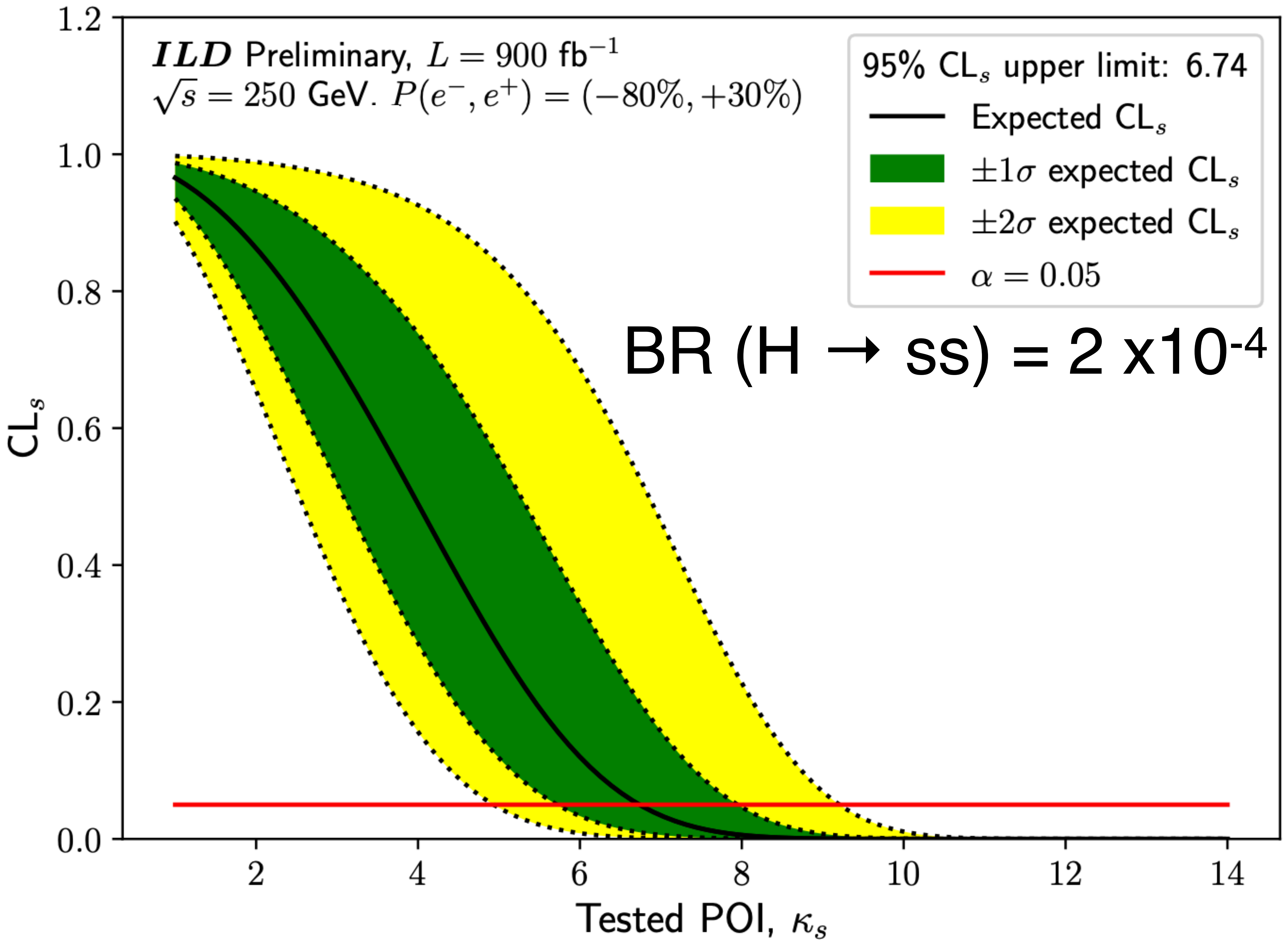
arXiv:2203.07535
L. Gouskos @FCC week

Constraints on s-coupling

Compatible results for both FCC and ILC like analyses

- ILD combined limit of $\kappa_s < 6.74$ at 95% CL with 900/fb at 250 GeV (i.e. half dataset)
  - No PID worsen the results by 8%
- FCC for Z(vv) only sets a limit of $\kappa_s < 1.3$ at 95% CL with 10/ab at 240 GeV and 2 IPs





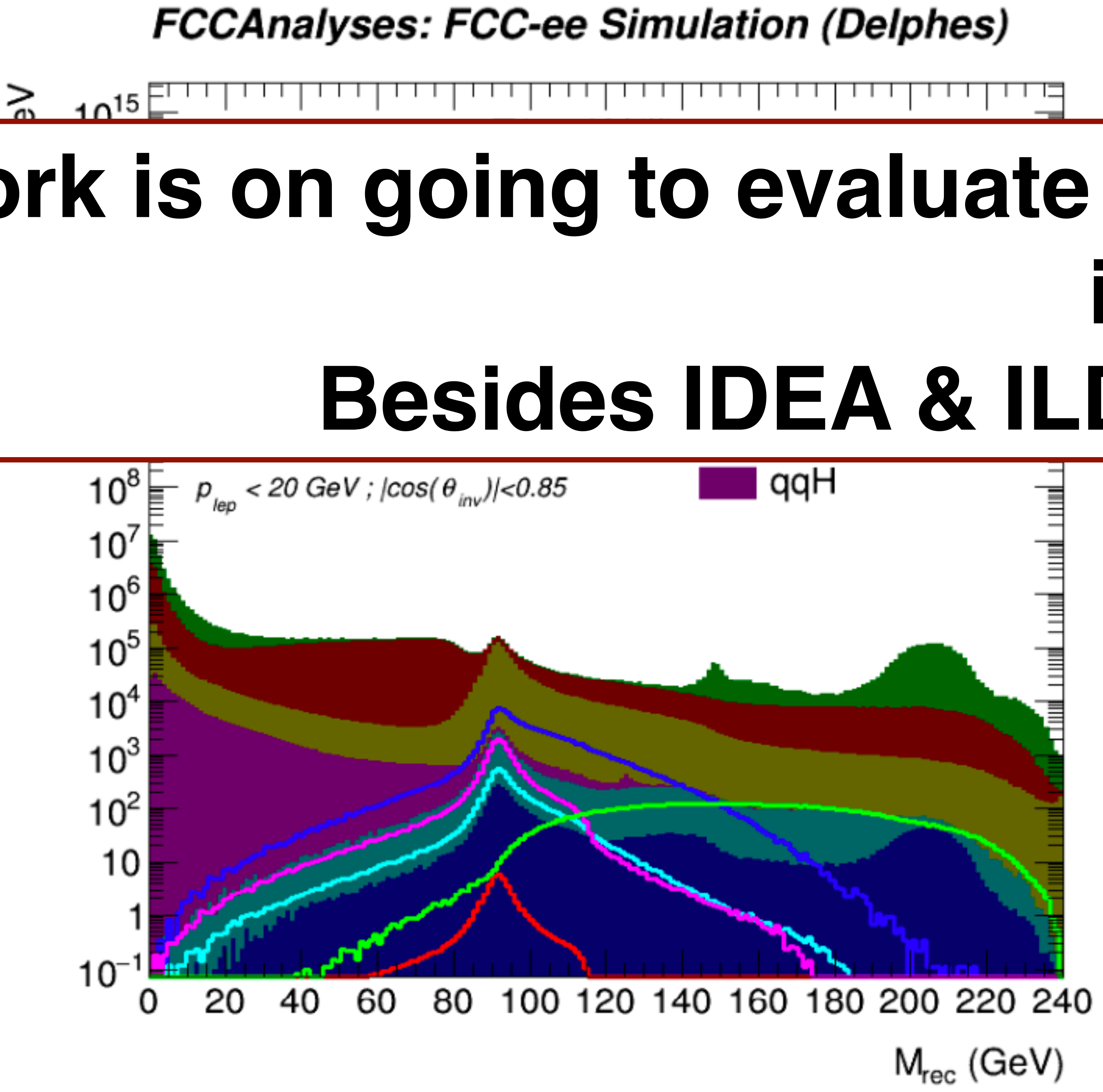
BR (H → ss) = 2 x10^{-4}

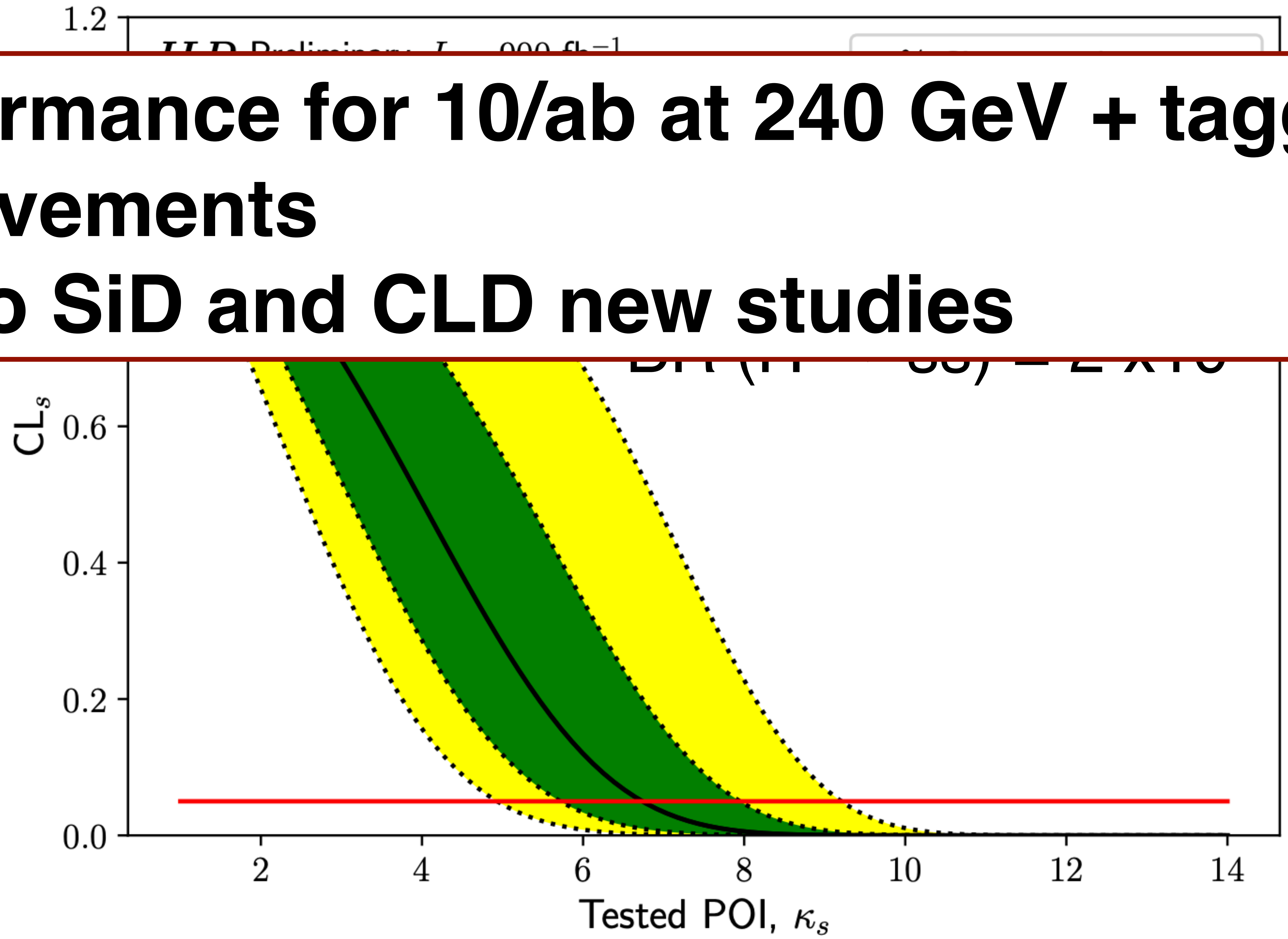
arXiv:2203.07535
L. Gouskos @FCC week

Constraints on s-coupling

Compatible results for both FCC and ILC like analyses

- ILD combined limit of $\kappa_s < 6.74$ at 95% CL with 900/fb at 250 GeV (i.e. half dataset)
  - No PID worsen the results by 8%
- FCC for Z(vv) only sets a limit of $\kappa_s < 1.3$ at 95% CL with 10/ab at 240 GeV and 2 IPs





Work is on going to evaluate performance for 10/ab at 240 GeV + tagger improvements
Besides IDEA & ILD, also SiD and CLD new studies

NEW - Constraints on s-coupling

Evaluating simultaneously all hadronic Higgs final states and impact of PID

Higgs decay	**H⟶bb**	**H⟶WW/ZZ**	**H⟶gg**	**H⟶cc**	? **H⟶ss**
BR	57.7%	11%	8.6%	2.9%	0.024%

only one observed to this day

Observable at FCC-ee

240 GeV	**H⟶bb**	**H⟶cc**	**H⟶gg**	**H⟶ss**
Z⟶ll	0.68	4.02	2.18	234
Z⟶qq	0.32	3.52	3.07	408.55
Z⟶$\nu\nu$ (BNL)	0.33	2.27	0.94	137
Z⟶$\nu\nu$ (APC)	0.36	2.18	1.10	151
Combined (BNL)	0.21	1.66	0.8	104.99
Combined (APC)	0.22	1.65	0.93	121

A. Maloizel ECFA 2024

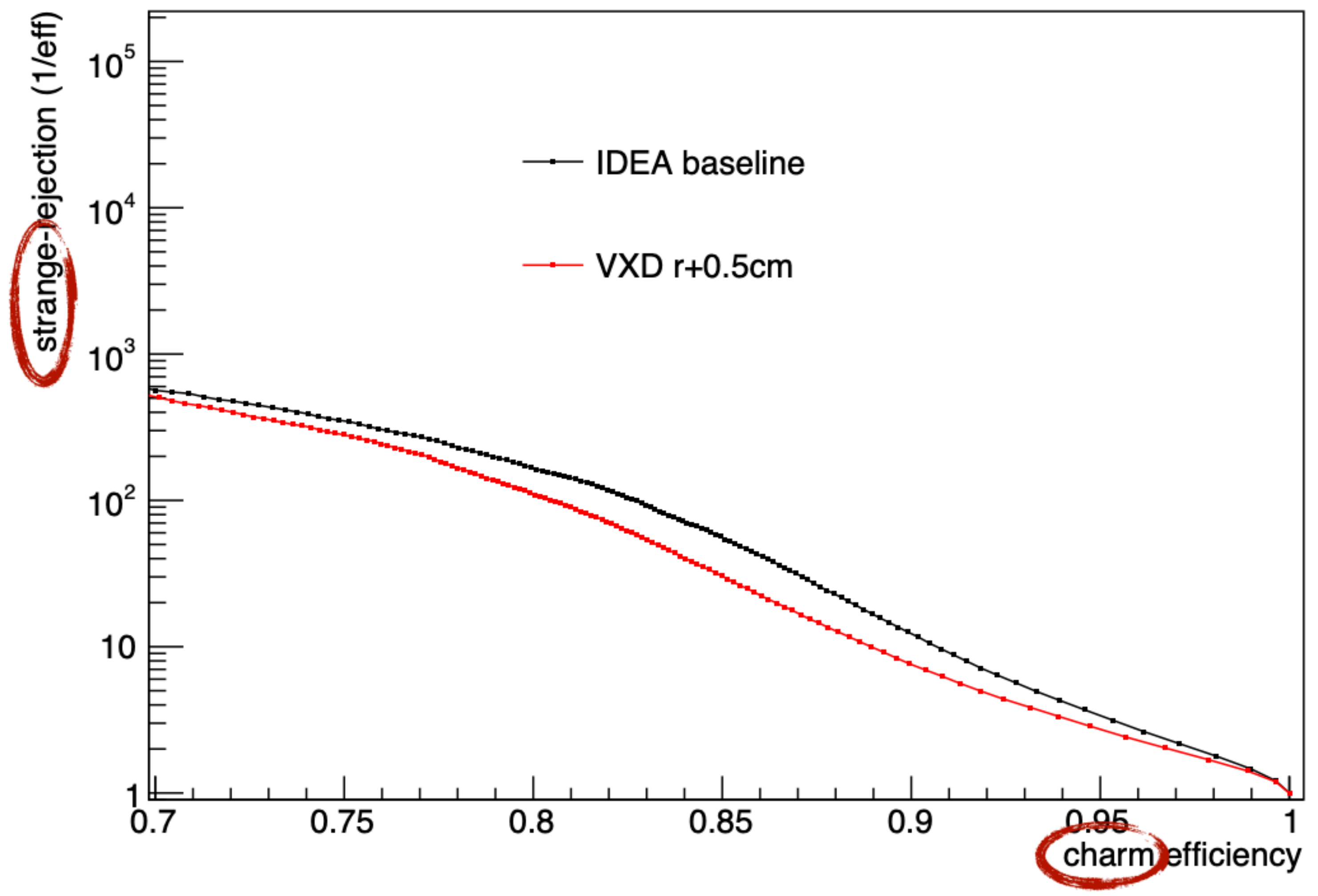
VARIATION	**68% CL precision**			
	μ_{Hbb}	μ_{Hcc}	μ_{Hgg}	μ_{Hss}
Baseline	**±0.3%**	**±4.2%**	**±2.8%**	**+674% -669%**
Relative change compared to baseline ($\mu_{variation}/\mu_{baseline}$)				
No TOF	x1.3	x1.02 (upper limit only)	x1	x1.03
No dNdX	x1.3	x1.07	x1.07	**x1.6**
VXDR +500µm	x1.3	x0.98 (lower limit only)	x1.04	x1

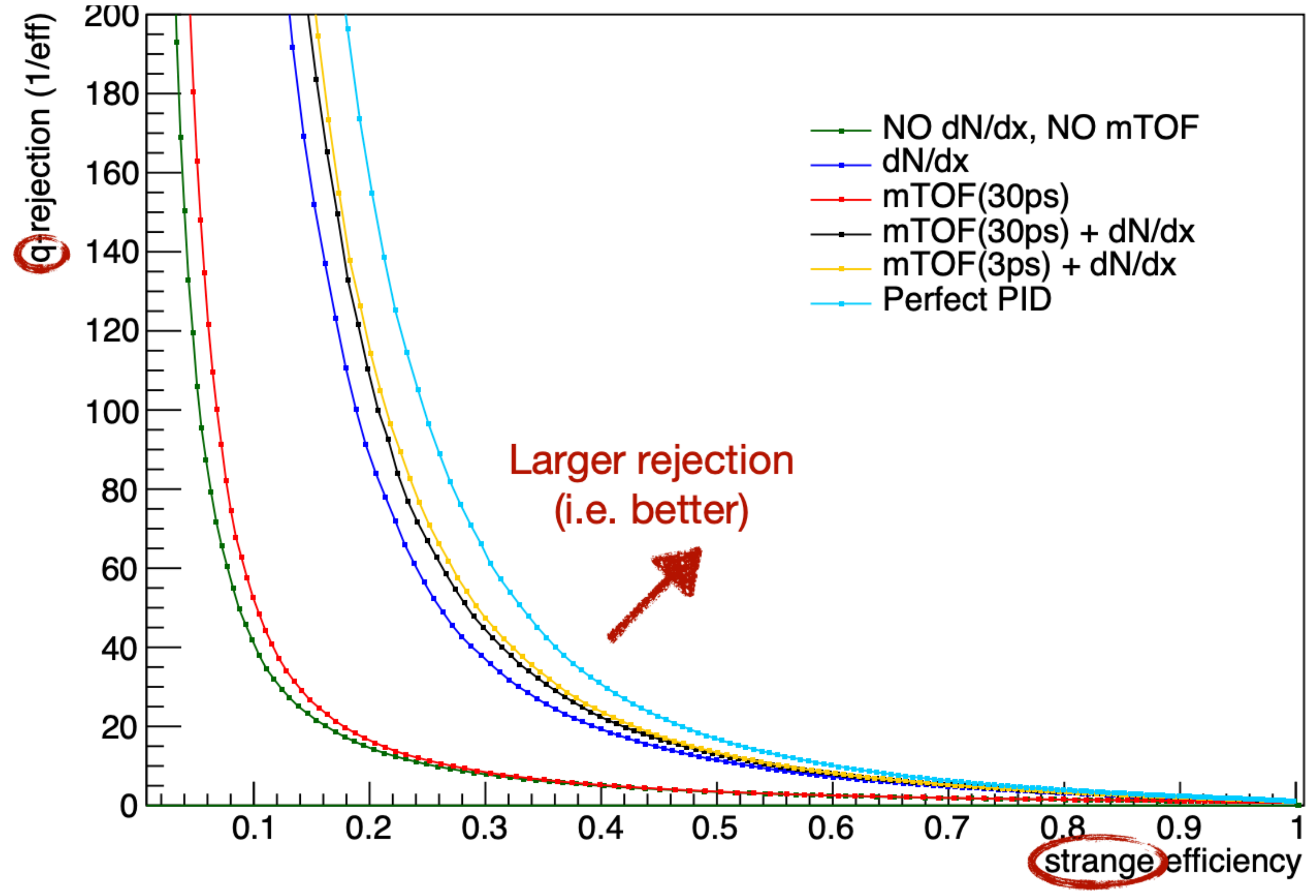
I. Veliscek ECFA 2024

IDEA: tagging performance vs detector configurations

Evaluating impact of different silicon and particle-identification detector configurations

- Current IDEA pixel/tracking system:
  - beam pipe at 1cm and 3 innermost silicon barrel layers: 1.2cm, 2cm, 3.15cm
  - PID: cluster-counting (dN/dx) + 30ps ToF system
- Number of pixel layers and various configurations have been tested
  - Hit resolution & position first layer
  - PID capabilities: timing, energy loss (gas/silicon)
- ***Findings***:
  - Very limited impact of TOF mass measurement on strange tagging
  - Ideal PID shows visible enhancement, especially at low efficiency



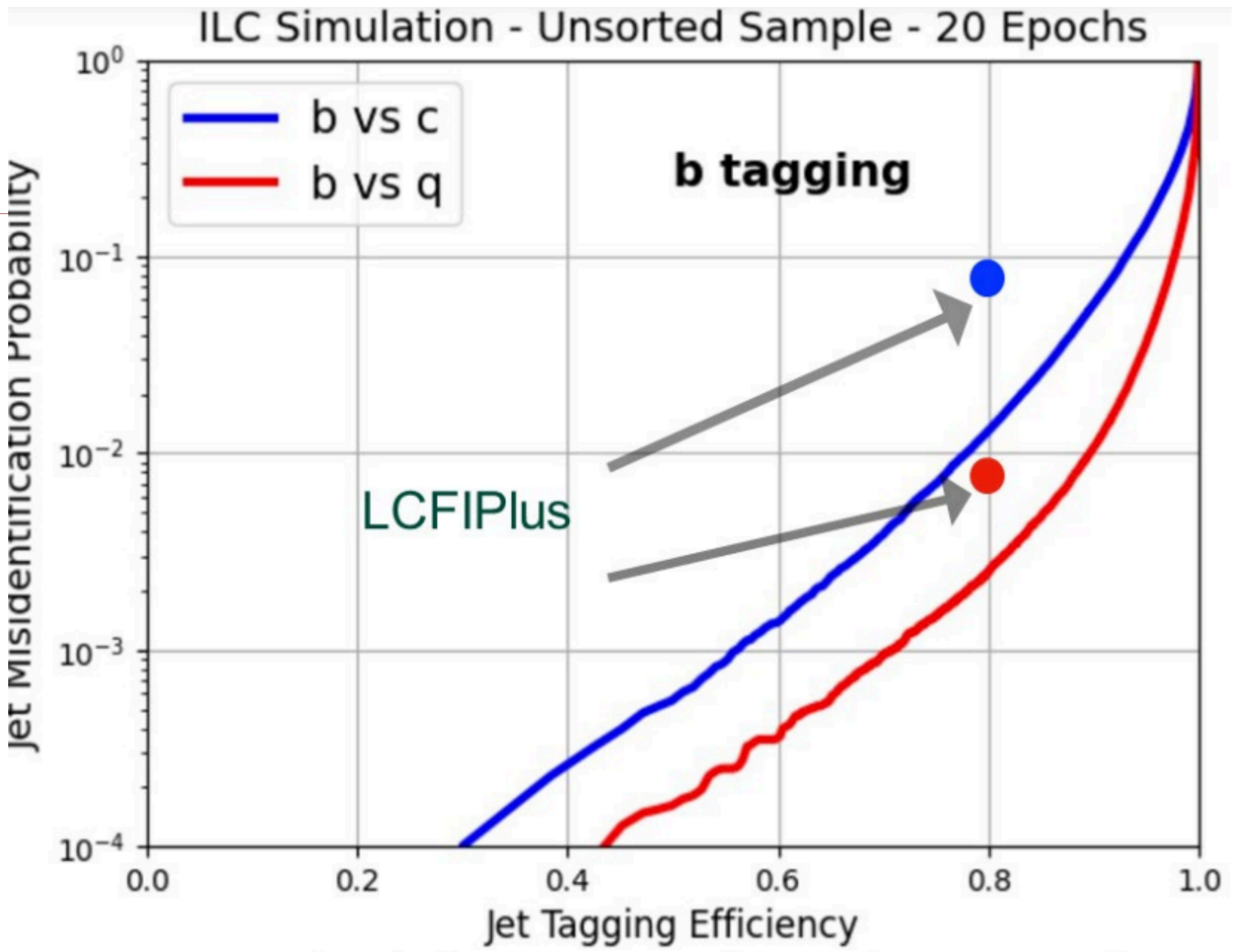


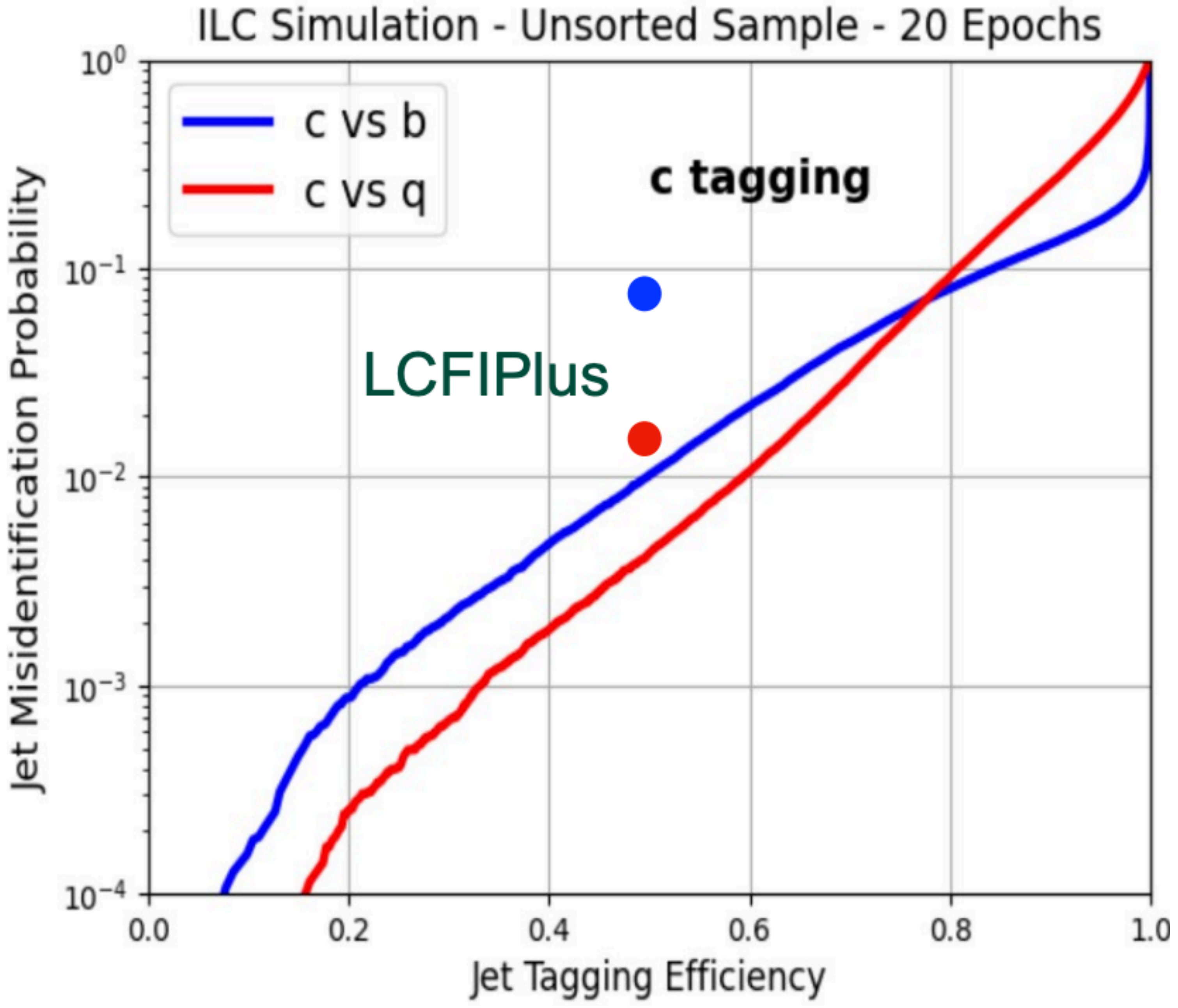
A. Sciandra ECFA 2024

ILD new tagger with Particle Transformers

Expected a large improvement wrt LCFIPlus based on full sim

- Factor (3-9) improvement using ParT from LCFIPlus
  - More work on going to tune the network, optimize input variables and separate embedding for tracks/neutrals
- ***Findings***:
  - dE/dx is essential for Particle ID in ILD – As well as ToF, but only effective in low energy tracks (which are less important in strange tagging)
- Still investigating inconsistencies with FCC/IDEA (much better) results based on Delphes samples
- Strange tagging dependence on PID performance to be investigated
  - Being studied with various detector configurations
- **GNN based PFA** for track-cluster matching yields to better performance – to replace PandoraPFA in ~a few years





T. Suehara ECFA 2024

NEW - SiD detector studies

Using FCC/IDEA results as a benchmark

	ILD	SID	IDEA	CLD	ALLEGRO
Vertex Inner Radius (cm)	1.6	1.4	1.2	1.2	1.2
Tracker technology	TPC+Silicon	Silicon	Si+Drift Chamber	Si	Si+Drift Chamber
Outer Tracker Radius (m)	1.77	1.22	2	3.3	2
ECal thickness	24 X_0	26 X_0	Dual RO	22 X_0	22 X_0
HCal thickness	5.9 λ_0	4.5 λ_0	7 λ_0	6.5 λ_0	9.5 λ_0
HCal Outer Radius (m)	3.3	2.5	4.5	3.5	4.5
Solenoid field (T)	3.5	5	2	2	2
Solenoid length (m)	7.9	6.1	6	7.4	6
Solenoid Radius (m)	3.4	2.6	2.1	4	2.7

$\frac{\sigma(\mathbf{E})}{\mathbf{E}}$	**SiD**	**IDEA**
ECAL	$\frac{17\%}{\sqrt{E}} \oplus 1\%$	$\frac{3\%}{\sqrt{E}} \oplus \frac{0.2\%}{E} \oplus 0.5\%$
HCAL	$\frac{55.9\%}{\sqrt{E}} \oplus 9.4\%$	$\frac{30\%}{\sqrt{E}} \oplus \frac{5\%}{E} \oplus 1\%$

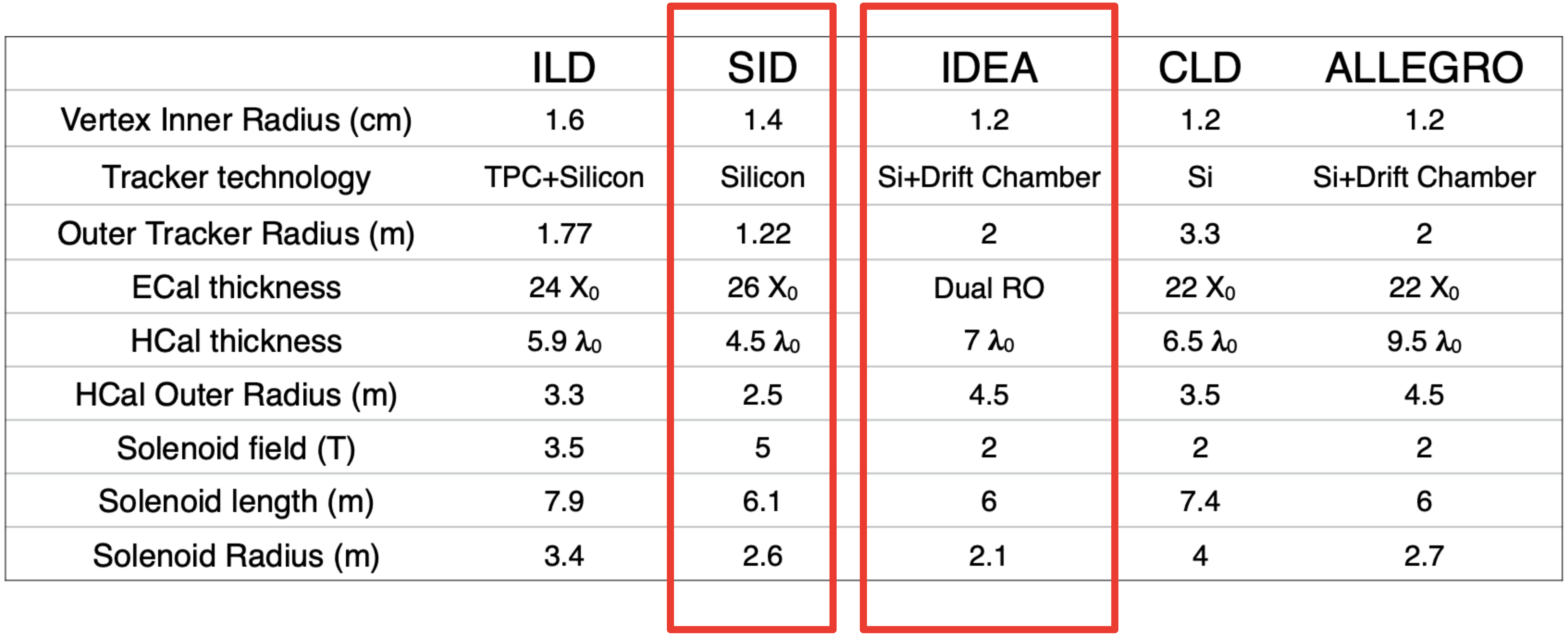
ILC TDR

2008.00338

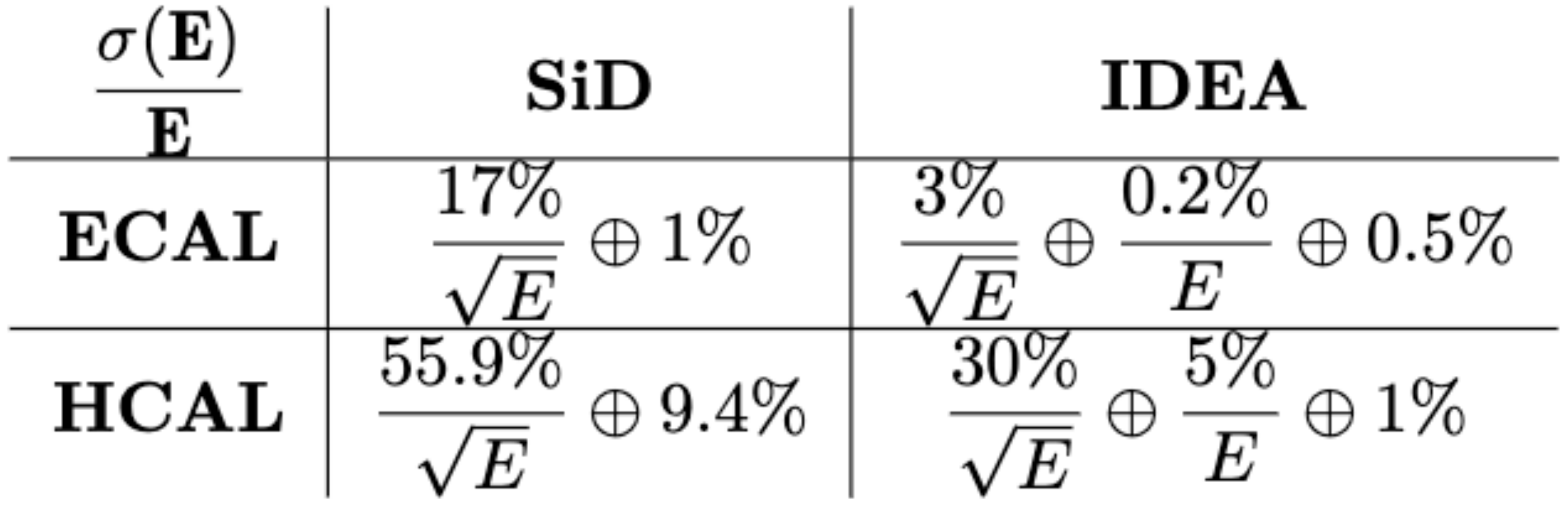
NEW - SiD detector studies

Using FCC/IDEA results as a benchmark

- ***SiD:*** all silicon vtx and tracker, sampling ECAL and HCAL, 5T B-field
- ***IDEA:*** silicon vtx, DCH + Si wrapper, DRO calorimeter, 2 T B-field → PID with dN/dx, TOF, supreme JER



	ILD	SID	IDEA	CLD	ALLEGRO
Vertex Inner Radius (cm)	1.6	1.4	1.2	1.2	1.2
Tracker technology	TPC+Silicon	Silicon	Si+Drift Chamber	Si	Si+Drift Chamber
Outer Tracker Radius (m)	1.77	1.22	2	3.3	2
ECal thickness	24 X_0	26 X_0	Dual RO	22 X_0	22 X_0
HCal thickness	5.9 λ_0	4.5 λ_0	7 λ_0	6.5 λ_0	9.5 λ_0
HCal Outer Radius (m)	3.3	2.5	4.5	3.5	4.5
Solenoid field (T)	3.5	5	2	2	2
Solenoid length (m)	7.9	6.1	6	7.4	6
Solenoid Radius (m)	3.4	2.6	2.1	4	2.7



$\frac{\sigma(\mathbf{E})}{\mathbf{E}}$	**SiD**	**IDEA**
ECAL	$\frac{17\%}{\sqrt{E}} \oplus 1\%$	$\frac{3\%}{\sqrt{E}} \oplus \frac{0.2\%}{E} \oplus 0.5\%$
HCAL	$\frac{55.9\%}{\sqrt{E}} \oplus 9.4\%$	$\frac{30\%}{\sqrt{E}} \oplus \frac{5\%}{E} \oplus 1\%$

ILC TDR

2008.00338

Towards a flexible framework for detector studies

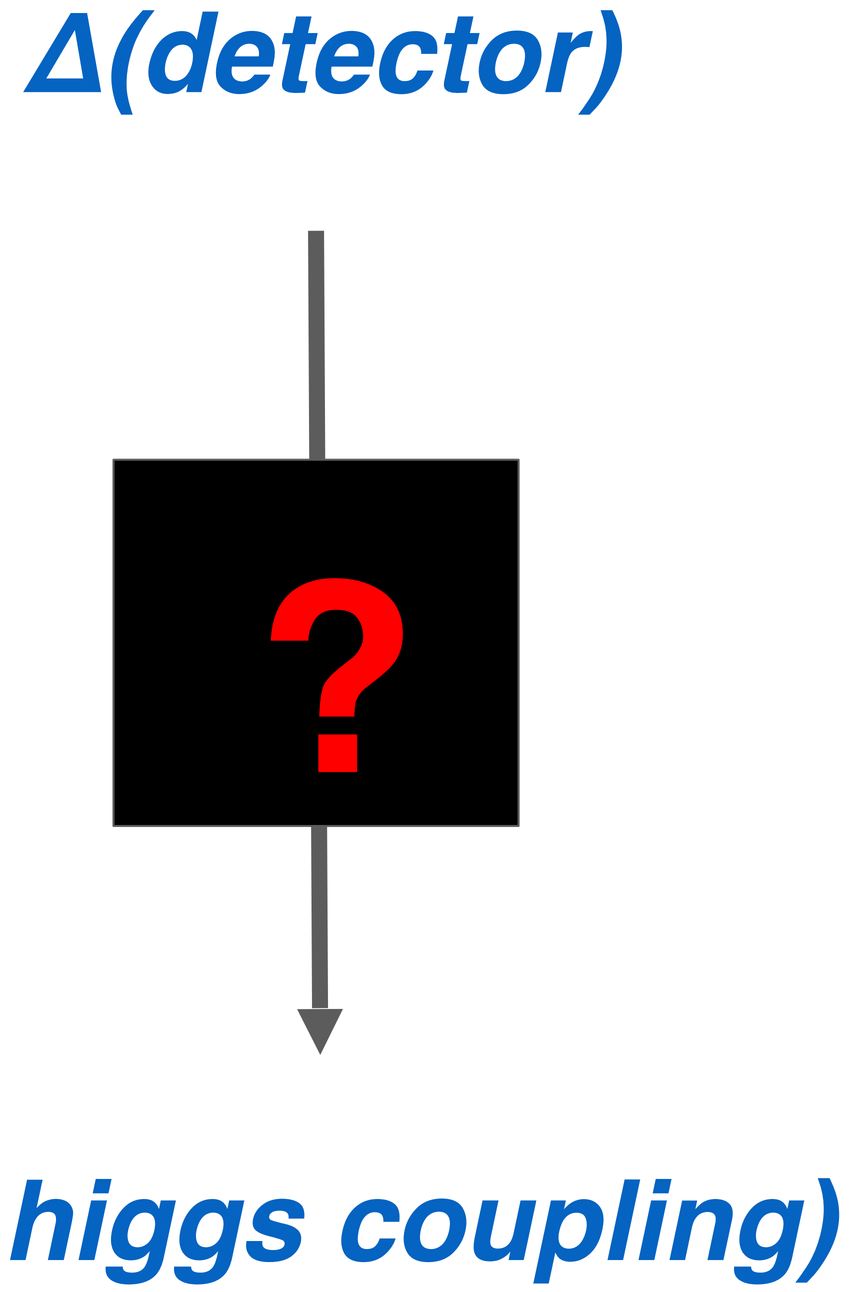
Determining the physics impact of detector choices is paramount for detector design

Systematic approach to evaluate different detector/sub-detector configurations based on the impact on jet identification

- Compare different detector concepts (SiD and IDEA) with FastSim
  - Reproduced IDEA published results based on [IDEA Delphes card]
  - Compare FastSim against FullSim

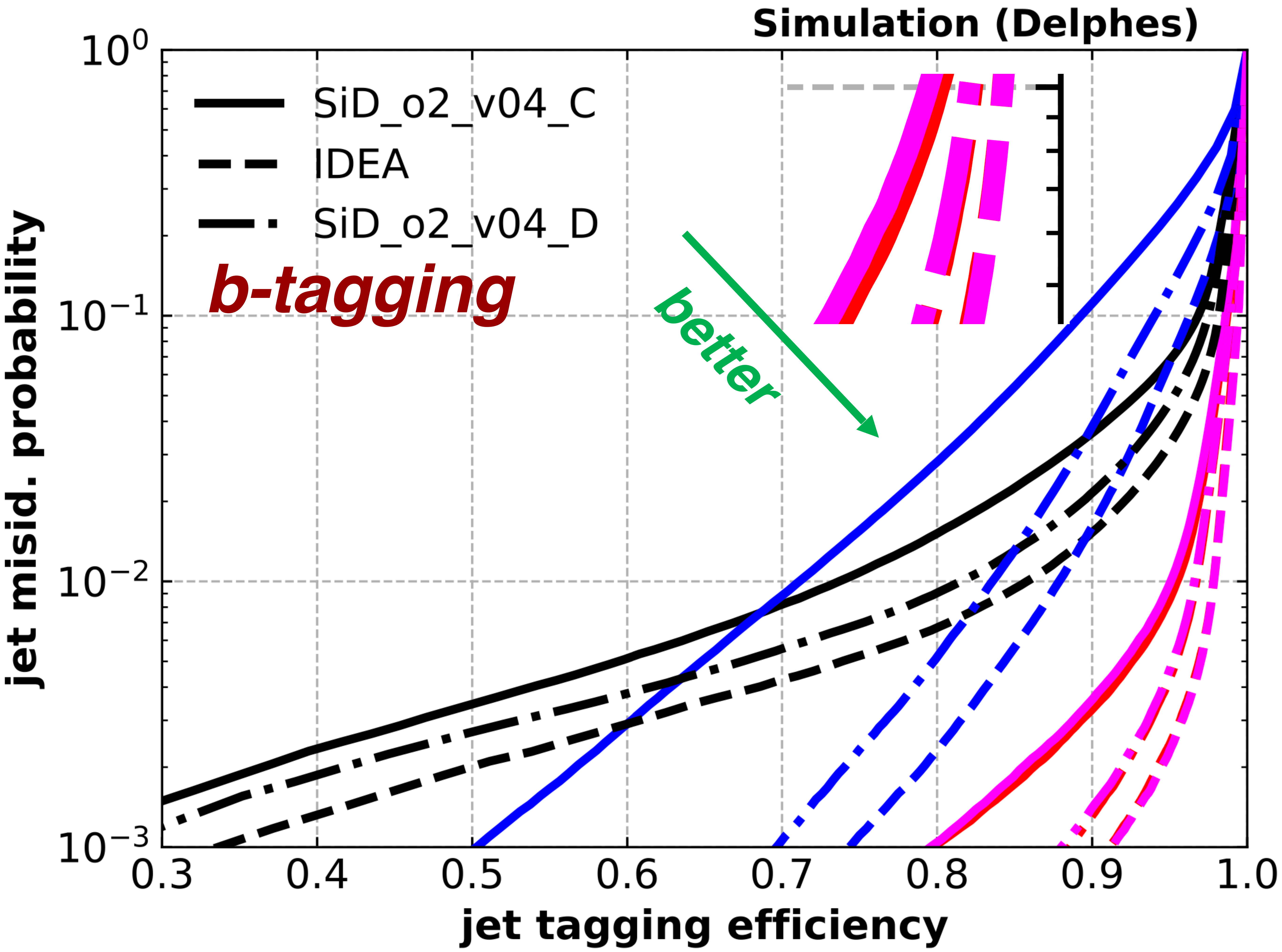
End-product:

- A versatile framework, building on existing tools, critical for R&D exploration
- Answer ***how detector variations impact precision on Higgs couplings.***



Jet flavor tagging for SiD using Particle Net

- Existing Delphes card for SiD, based on ILC TDR performance: https://dsid.hepforge.org/ (~9 years old).
- A new Delphes card to include newer TrackCovariance, ClusterCounting modules, ***assuming same tracking performance as for IDEA:*** **SiD_o2_v04_C**
- Also considered a modified scenario for SiD, with the resolution of the ECAL and the HCAL matching that of the IDEA dual calorimeter → **SiD_o2_v04_D**.
  - Improvement in b-tagging driven by calorimeter resolution → better reconstruction of PFOs



SiD different detector configurations

Two configurations for SiD detector concept vs IDEA

- ***Significant gains in strange vs udg and b discrimination from PID.***
- Improved calorimeter resolution only brings marginal gain

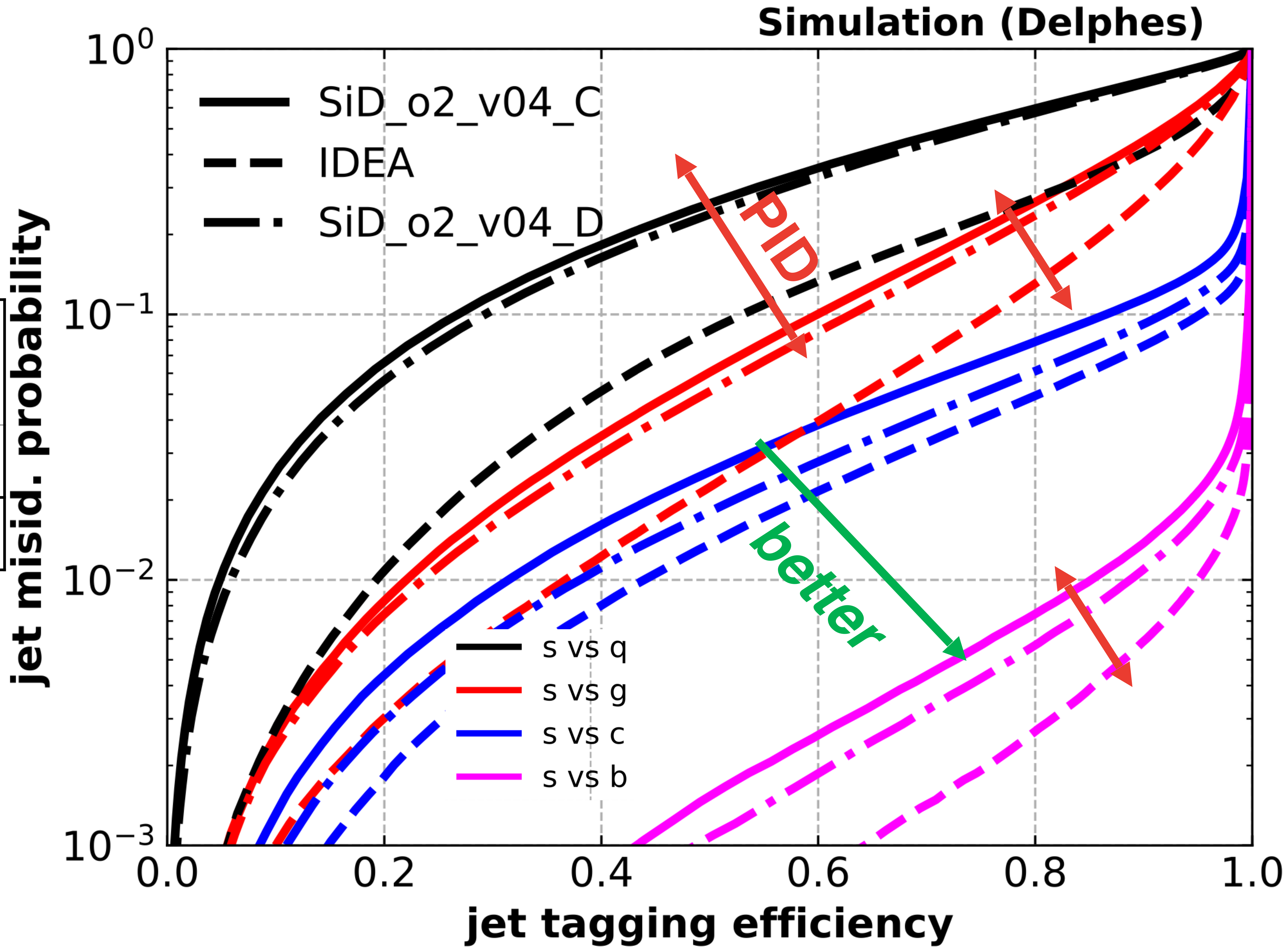
SiD_o2_v04_C (_D)

WP	Eff (s)	Mistag (g)	Mistag (q)	Mistag (c)	Mistag (b)
Loose	90%	45 (42)%	75 (75)%	12 (9)%	1.3 (1.1)%
Medium	80%	27 (24)%	55 (55)%	8 (6)%	0.7 (0.6)%

IDEA

WP	Eff (s)	Mistag (g)	Mistag (q)	Mistag (c)	Mistag (b)
Loose	90%	27%	41%	7.5%	0.6%
Medium	80%	13%	27%	5%	0.3%

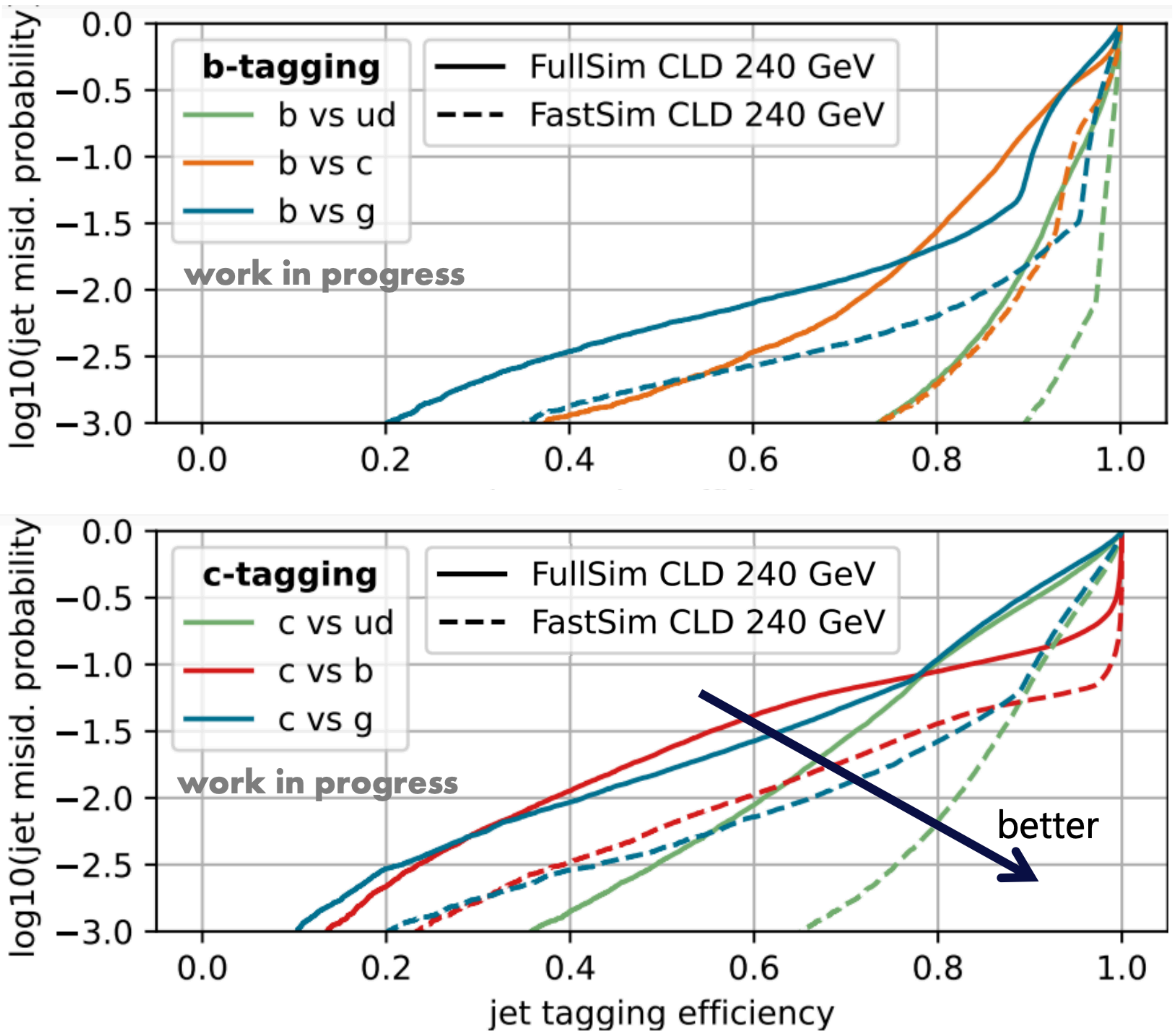
s-tagging



NEW - CLD full vs fast studies

Using tagging performance as a benchmark

- Loss in performance when moving from fast to full simulation
    - Fake neutrals and lost tracks in full sim
    - misidentification probability of 10–2 for c vs. ud: 82% (fast sim) / 61% (full sim)
    - misidentification probability of 10–2 for b vs. ud: 97% (fast sim) / 88% (full sim)
- Tracks for charged particles - PFOs for neutral particles - helps recovering performance
- Adding vertex information does not improve tagging performance



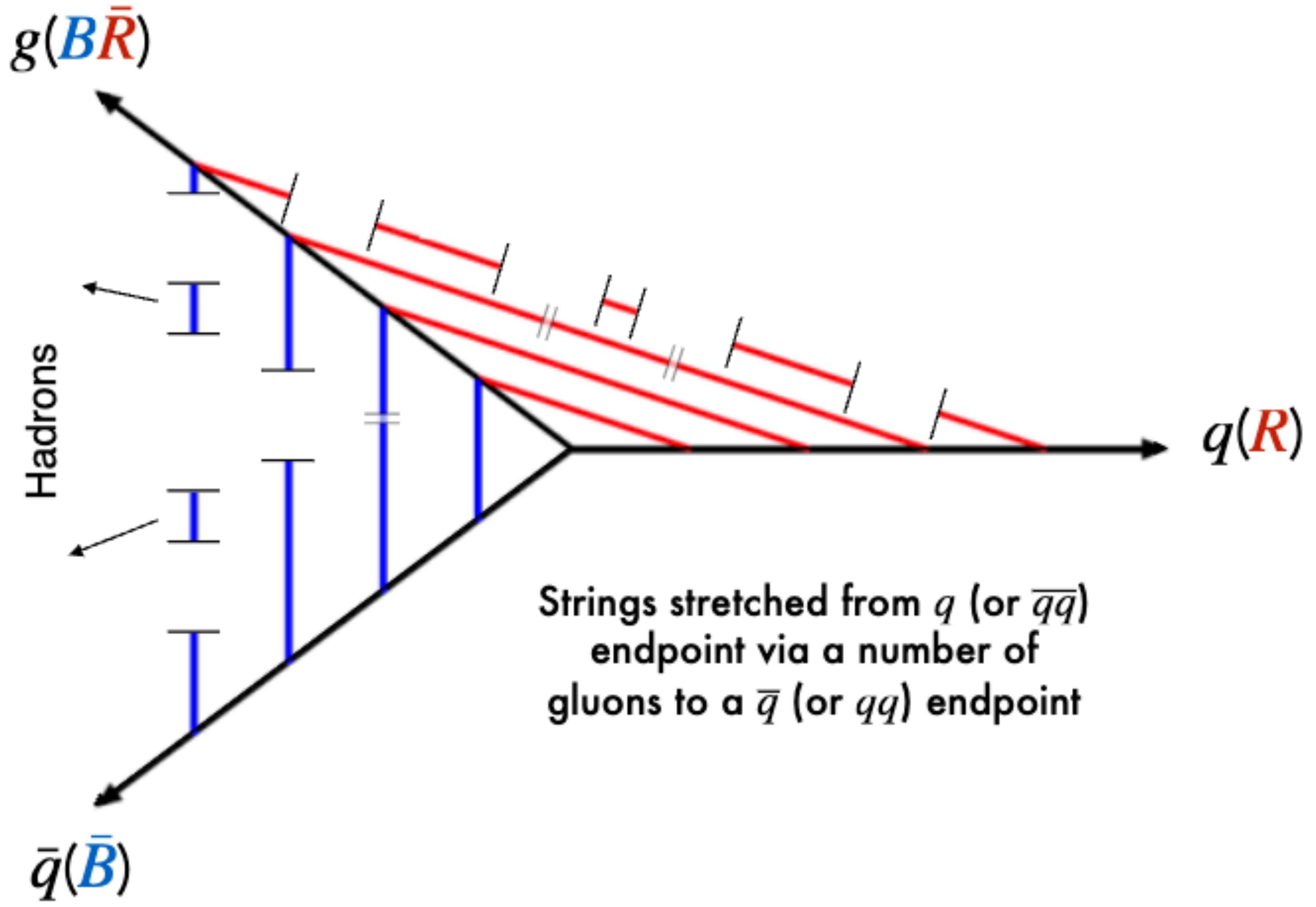
S. Aumiller ECFA 2024

NEW Fragmentation modelling

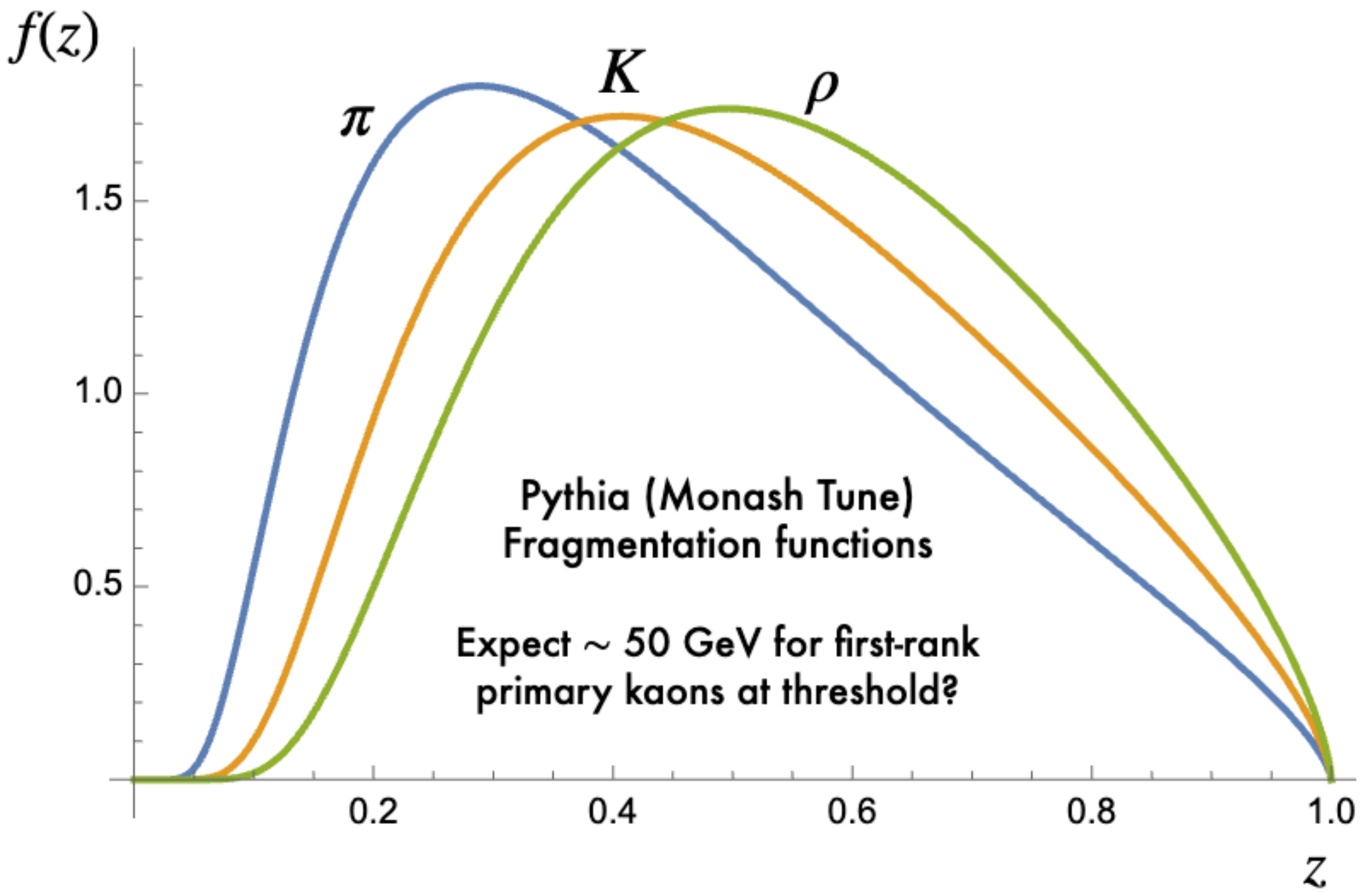
Peter Skands and Javira Altmann
L.Gouskos ECFA 2024

Review of the state of the art and challenges

- Hadronization in PYTHIA is modelled using Lund ***string fragmentation***
  - partons move apart at high energies, which in turn stretches the confining potentials and leads to string breaking
- This leads to the two resulting strange hadrons also forming adjacent to each other, and a local strangeness conservation
- Further qualitative modelling variations could include:
  - thermodynamical string fragmentation
  - explicit modelling of hyperfine splitting effects
  - effects of a fluctuating or time-dependent string tension
  - and/or effects of string excitations
- Ultimately in-situ calibrations will validate the model assumptions
  - PID capabilities will be critical to provide enough information to test hadronization processes
  - Tera-Z will provide $> 10^5$ M bottom jets and $>10^4$ M strange jets



Strings stretched from q (or $\overline{qq}$) endpoint via a number of gluons to a $\bar{q}$ (or qq) endpoint



Pythia (Monash Tune) Fragmentation functions

Expect ~ 50 GeV for first-rank primary kaons at threshold?

NEW Fragmentation modelling

Peter Skands and Javira Altmann
L.Gouskos ECFA 2024

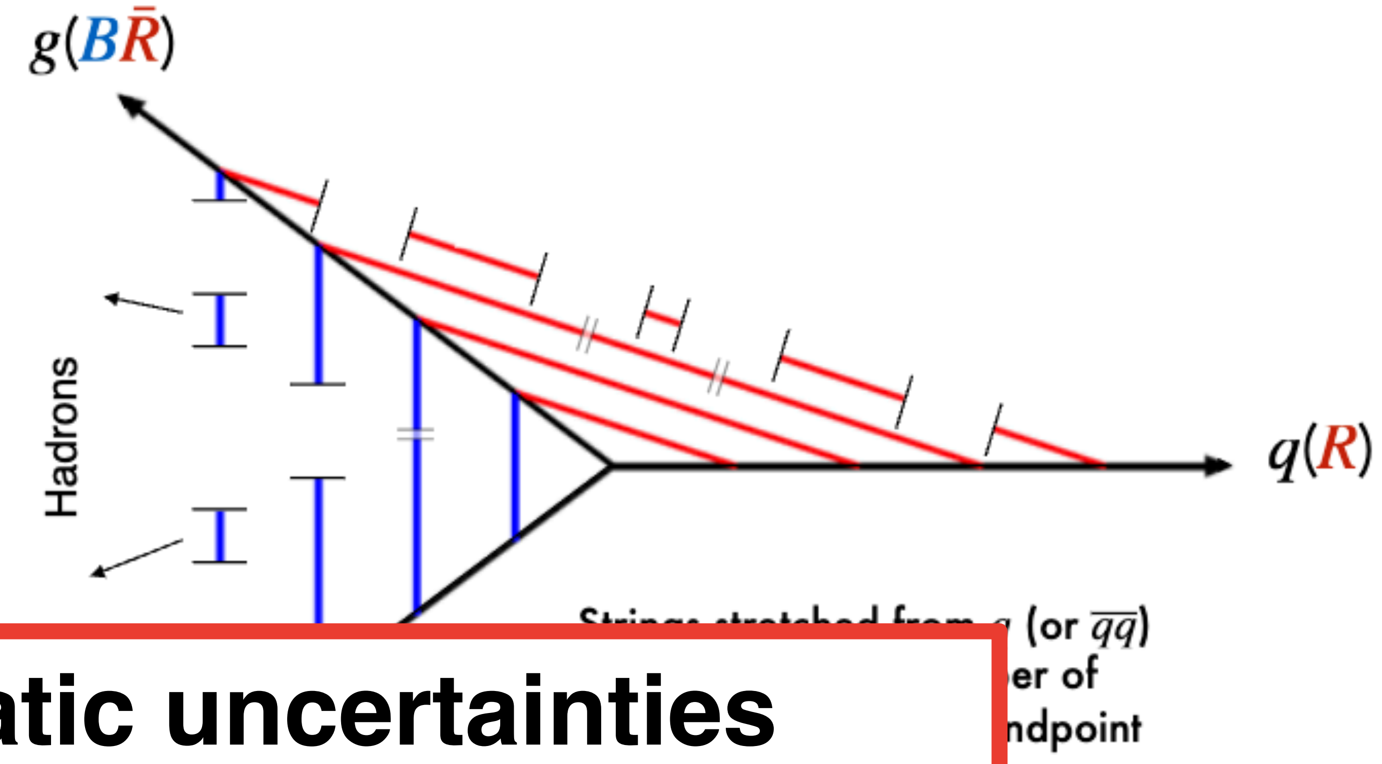
Review of the state of the art and challenges

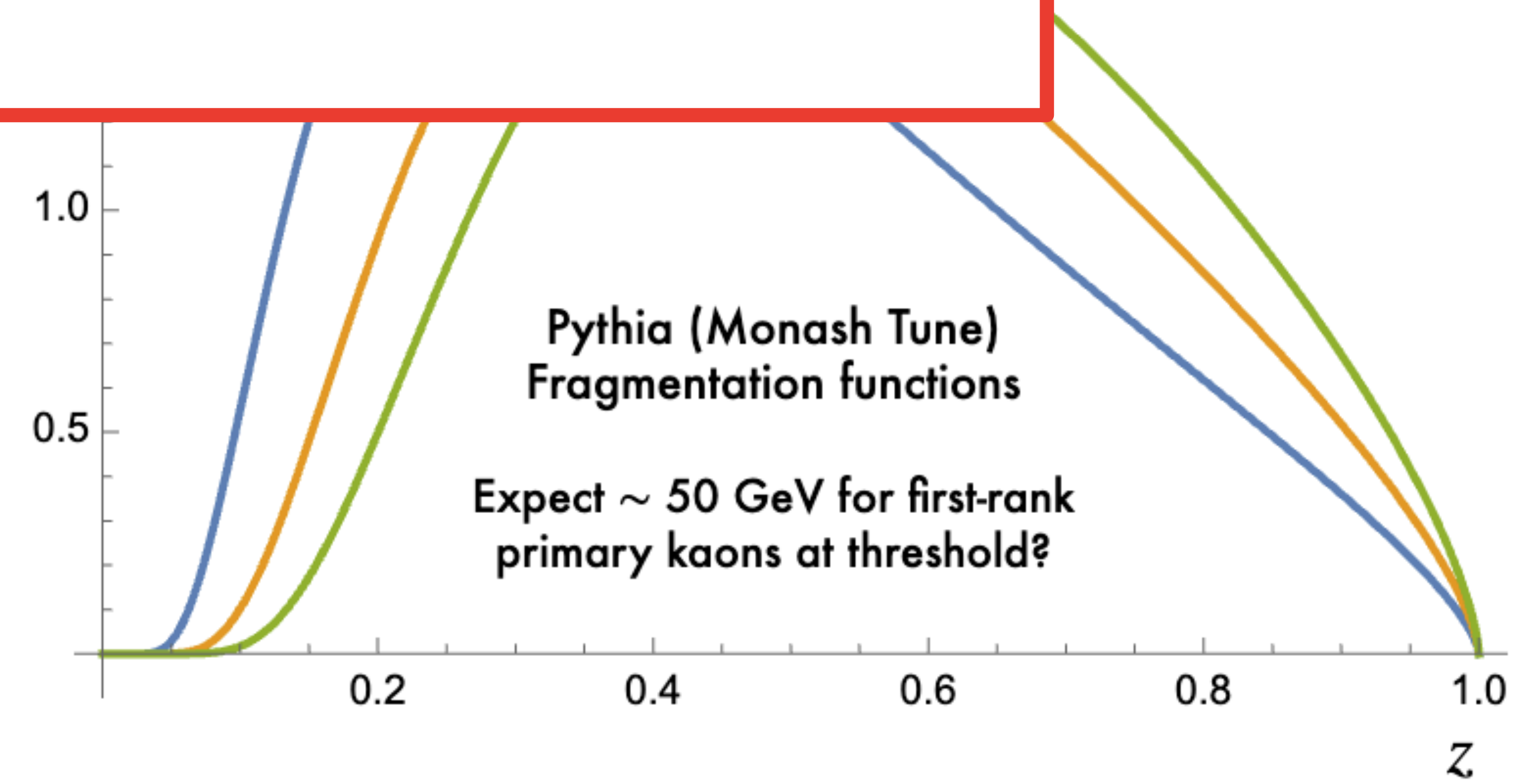
- Hadronization in PYTHIA is modelled using Lund ***string fragmentation***
  - partons move apart at high energies, which in turn stretches the confining potentials and leads to string breaking
- This leads to the two resulting strange hadrons also forming adjacent to each oth
- Further
  - therr
  - expli
  - effec
  - and/
- Ultimate
  - PID capabilities will be critical to provide enough information to test hadronization processes
  - Tera-Z will provide > 10^5 M bottom jets and >10^4 M strange jets

current knowledge of gbb/cc results in large systematic uncertainties

Within WG1: Dedicated “focus team” to address the challenge of Heavy Quark Fragmentation & Gluon Splitting (BCFrag & GSplit)

Dedicated Twiki: https://gitlab.in2p3.fr/ecfa-study/ECFAHiggsTopEW-Factories/-/wikis/FocusTopics/BCfrag





Pythia (Monash Tune) Fragmentation functions

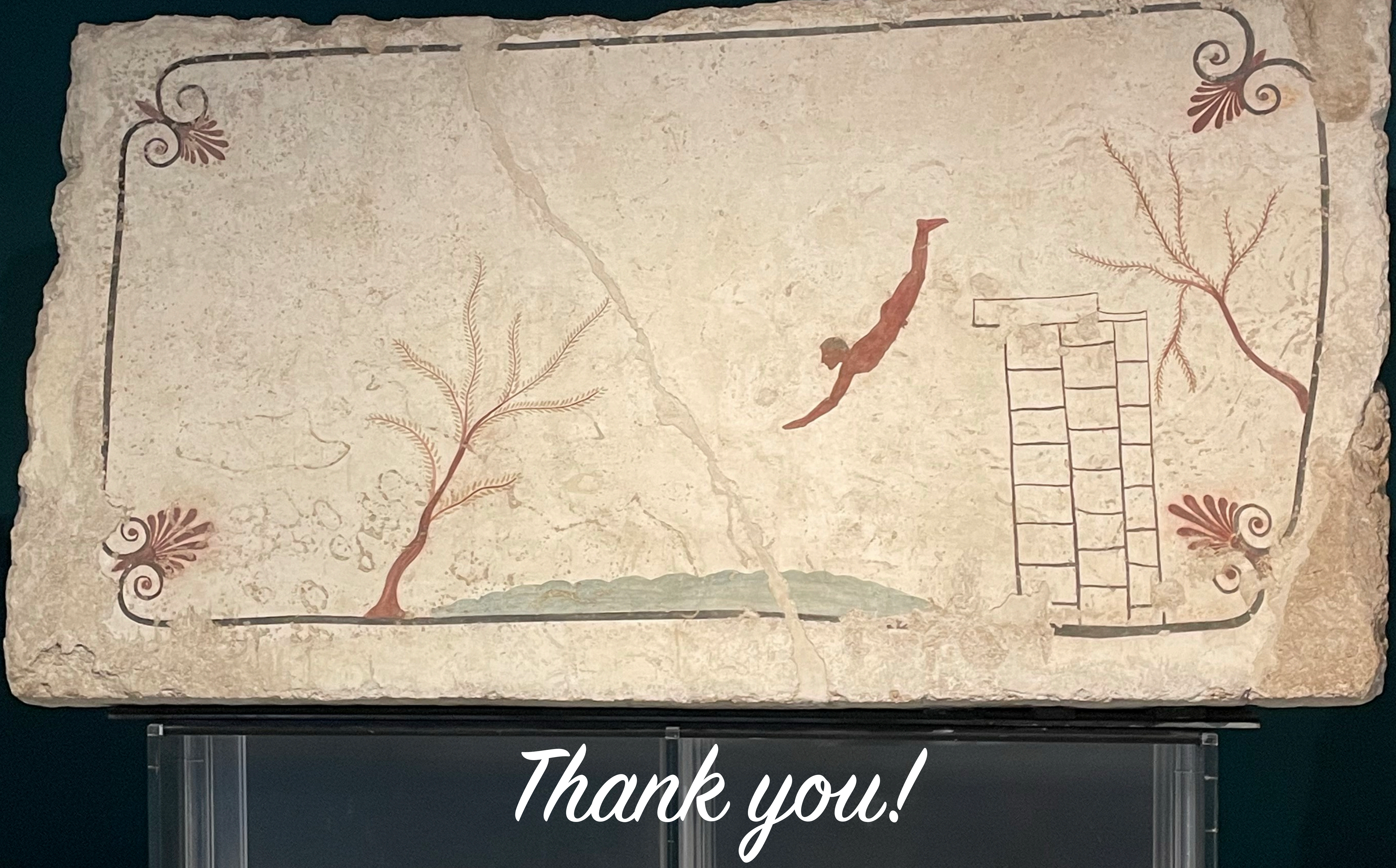
Expect ~ 50 GeV for first-rank primary kaons at threshold?

Towards the end of the year report

- Work on going to evaluate impact of (sub)detector resolutions on tagging performance :
    - the complementarity in momentum reach of charged hadron ID from dN/dx, dE/dx, ToF, RICH
    - ***Important to evaluate simultaneously other Higgs benchmarks***
- A new framework is available (based on Delphes/Weaver/FCCAnalyses) to evaluate and compare jet flavor tagging performance of different detector concepts on an equal footing**.**
- ***Important caveat*** in these studies: relying on fast-simulation!
    - We need to benchmark our results against realistic, full-simulation
        - Work on going for CLD and SiD

Contents

◆ 20/10	Deadline for physics studies to submit 2-page summary
◆ 20/10 – 10/11	Compilation and editing by WG1 subgroup conveners / nominated editors, and WG2/3 editors (as well as coordinators & chief editors) ***10/11 is the deadline for WG1 subgroup conveners finish their part!***
◆ 10/11 – 27/11	Editing by WG1 coordinators, WG2/3 editors & coordinators, and chief editors. ***27/11 is deadline for complete draft to be handed over to chief editors.***
◆ 27/11 – 18/12	Editing by chief editors only
◆ 18/12	Circulation of version 1 to contributors and R-ECFA
◆ 17/1	Deadline to receive comments on version 1
◆ 24/1	Deadline to receive final results/plots from contributors
◆ February	Incorporation of comments, final results, and references
◆ 21/2	Final version to R-ECFA
◆ 7–8/3	R-ECFA approval during country visit followed by submission to arXiv



Thank you!

Analysis pipeline for jet flavor tagging studies

1. Sample generation:
- Whizard 3.1.4 + Pythia6
- Z(→vv)H(→uu/dd/cc/ss/bb/gg)
- 1.5M events (3M jets) per flavor

2. Fast simulation:
- Delphes with edm4hep output using k4SimDelphes

3. Preprocessing:
- Jet Clustering on PFCandidates and tree flattening using FCCAnalyses

4. ParticleNet training:
- Using the weaver framework
- Use 1.8M jets/per flavor with 80%/20% train-val split

5. Infererce
- Within FCCAnalyses
- Using 1.2M jets/flavor

DELPHES fast simulation

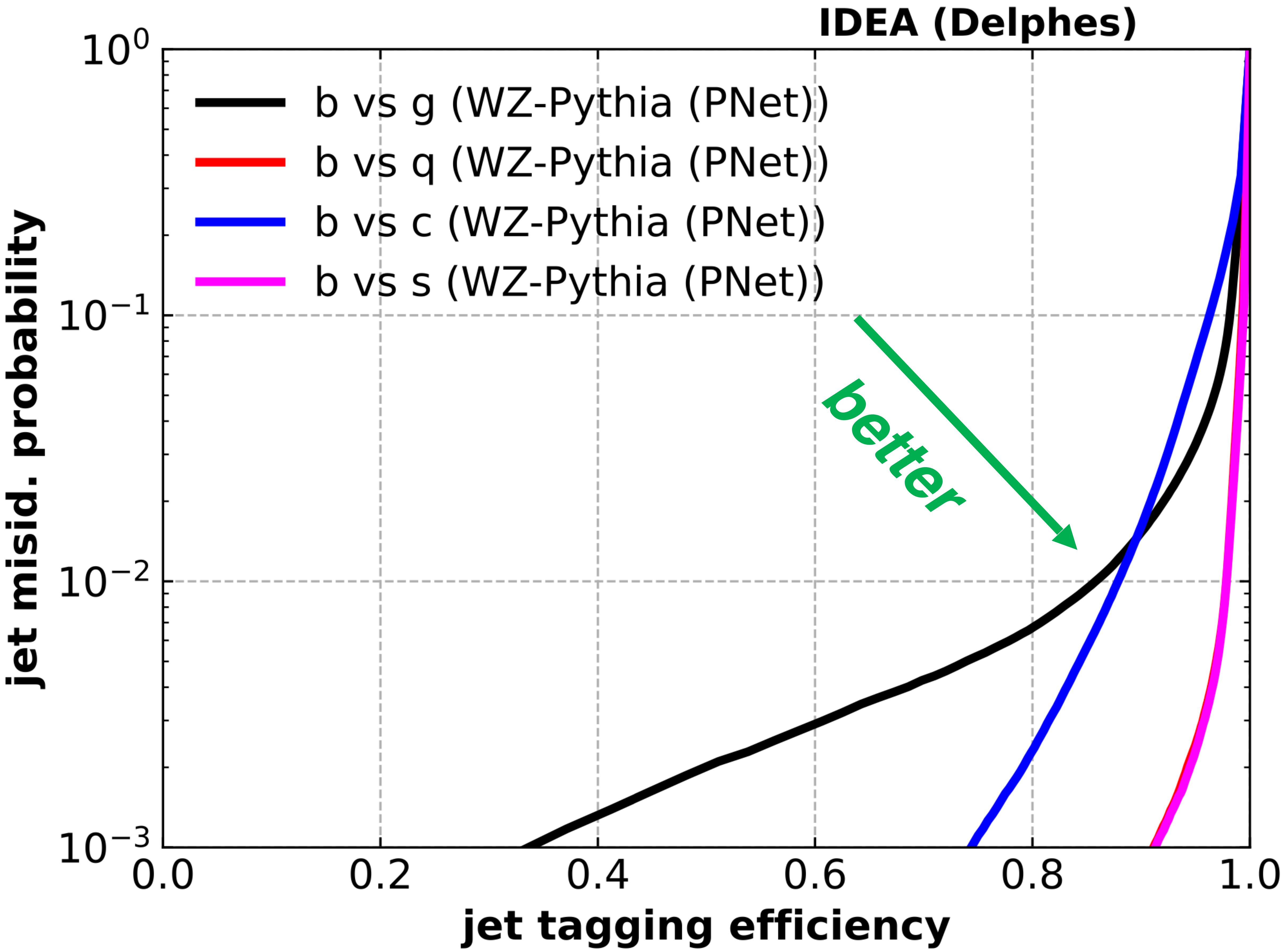
KEY4HEP

FCCAnalyses

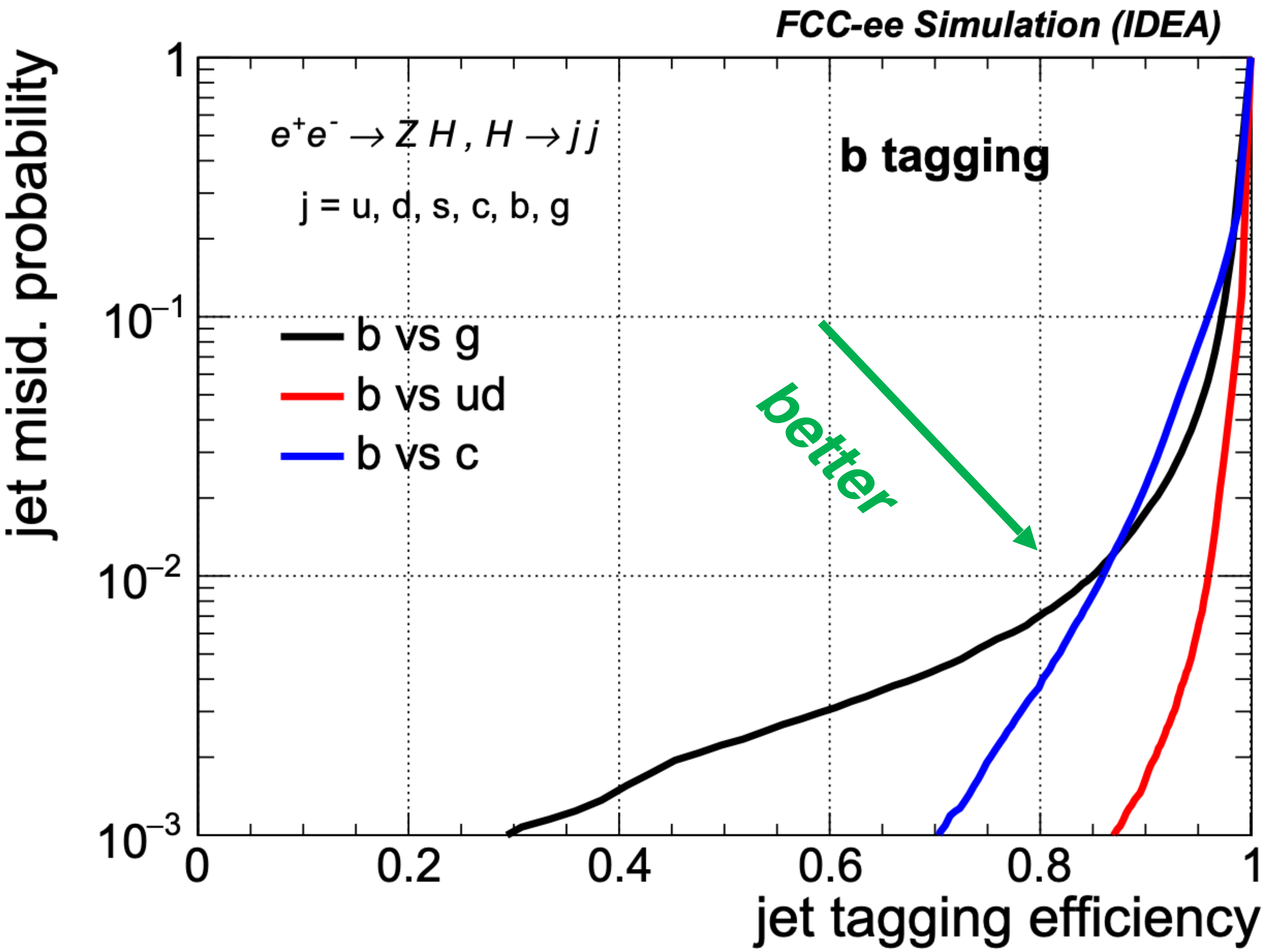
weaver-core

SiD - validation

Eur. Phys. J. C **82**, 646 (2022) : Jet Flavour Tagging for Future Colliders with Fast Simulation



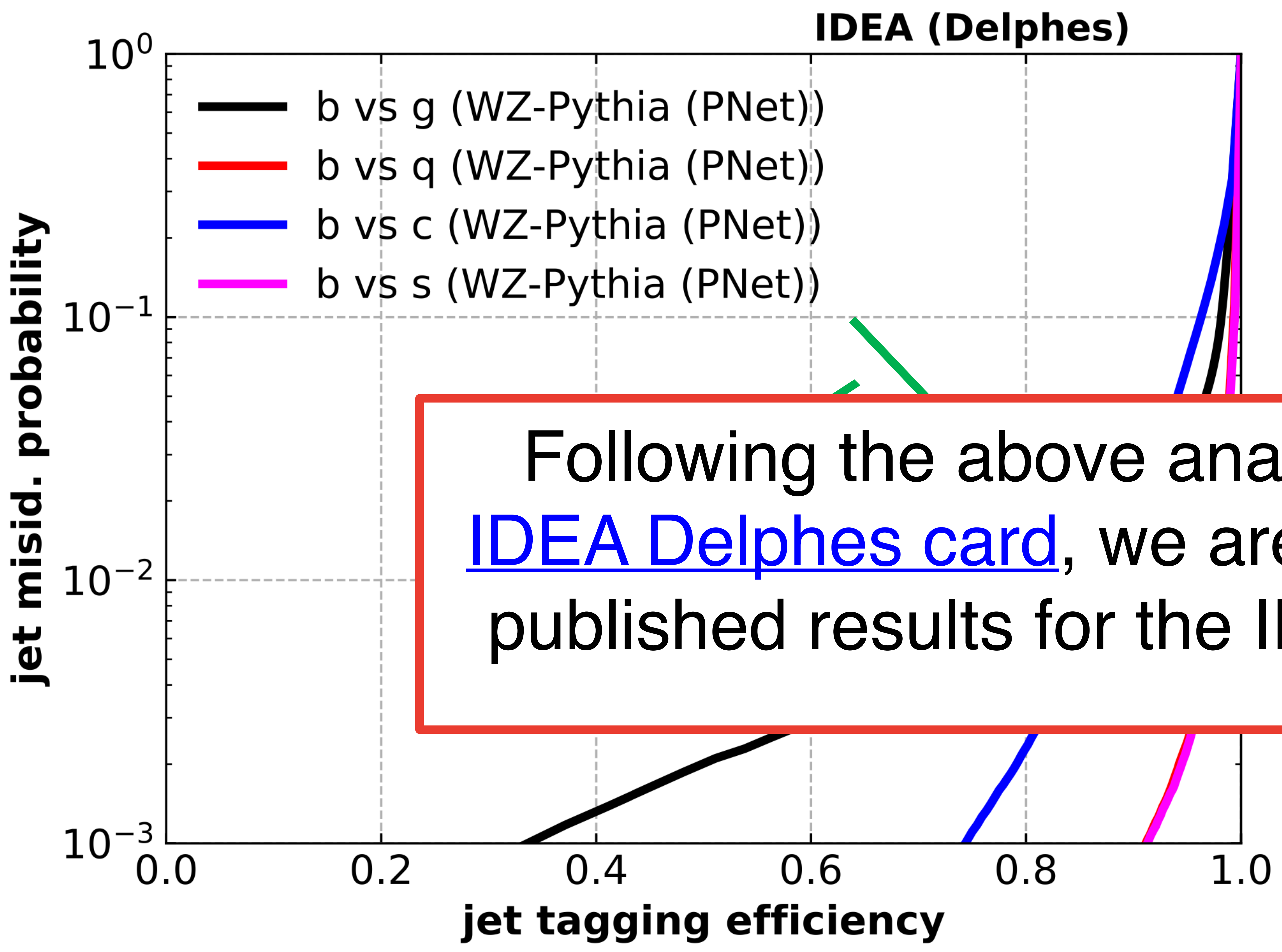
Our results

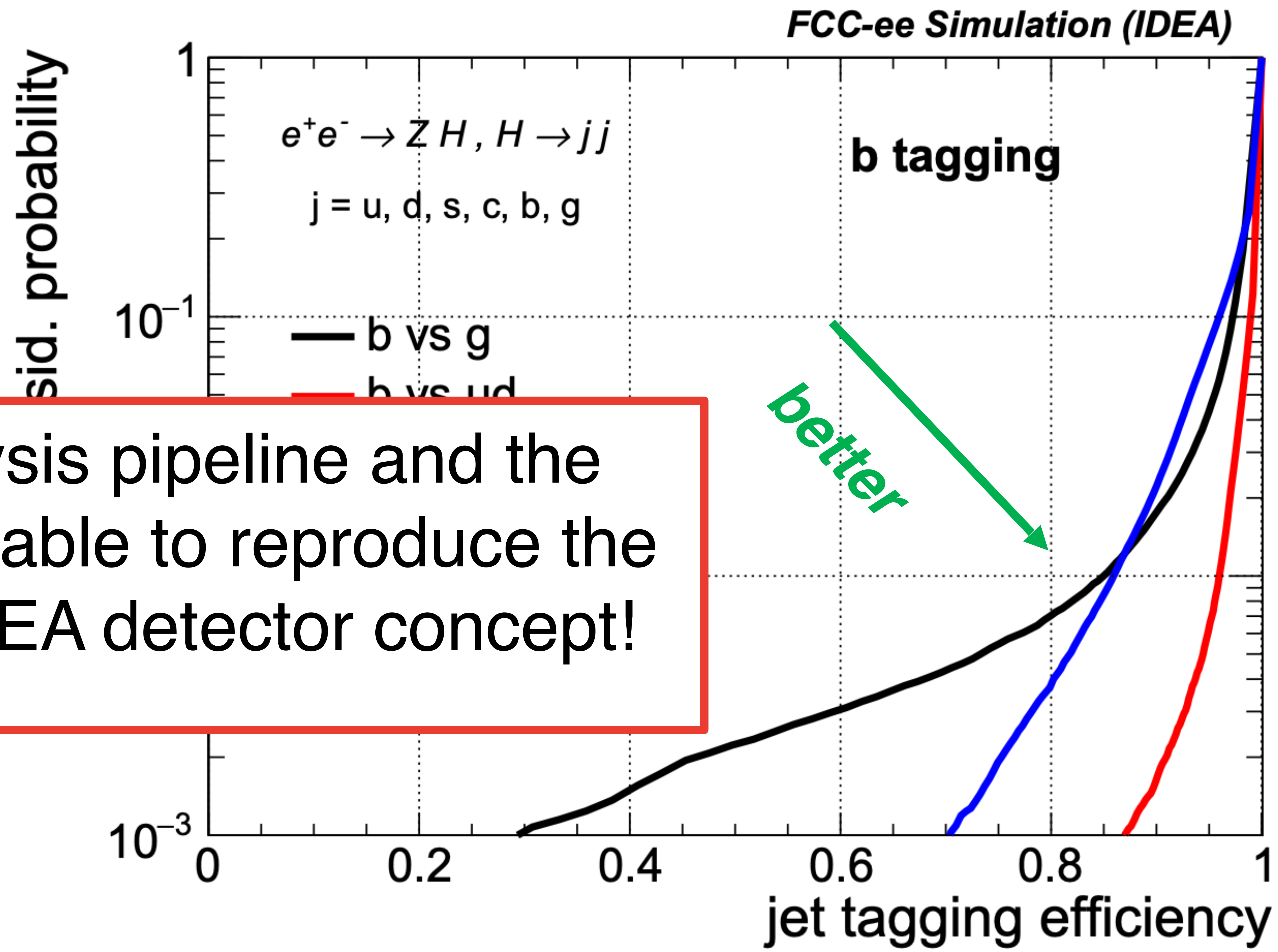


Eur. Phys. J. C **82**, 646 (2022)

SiD - validation

Eur. Phys. J. C **82**, 646 (2022) : Jet Flavour Tagging for Future Colliders with Fast Simulation





Following the above analysis pipeline and the IDEA Delphes card, we are able to reproduce the published results for the IDEA detector concept!

Our results

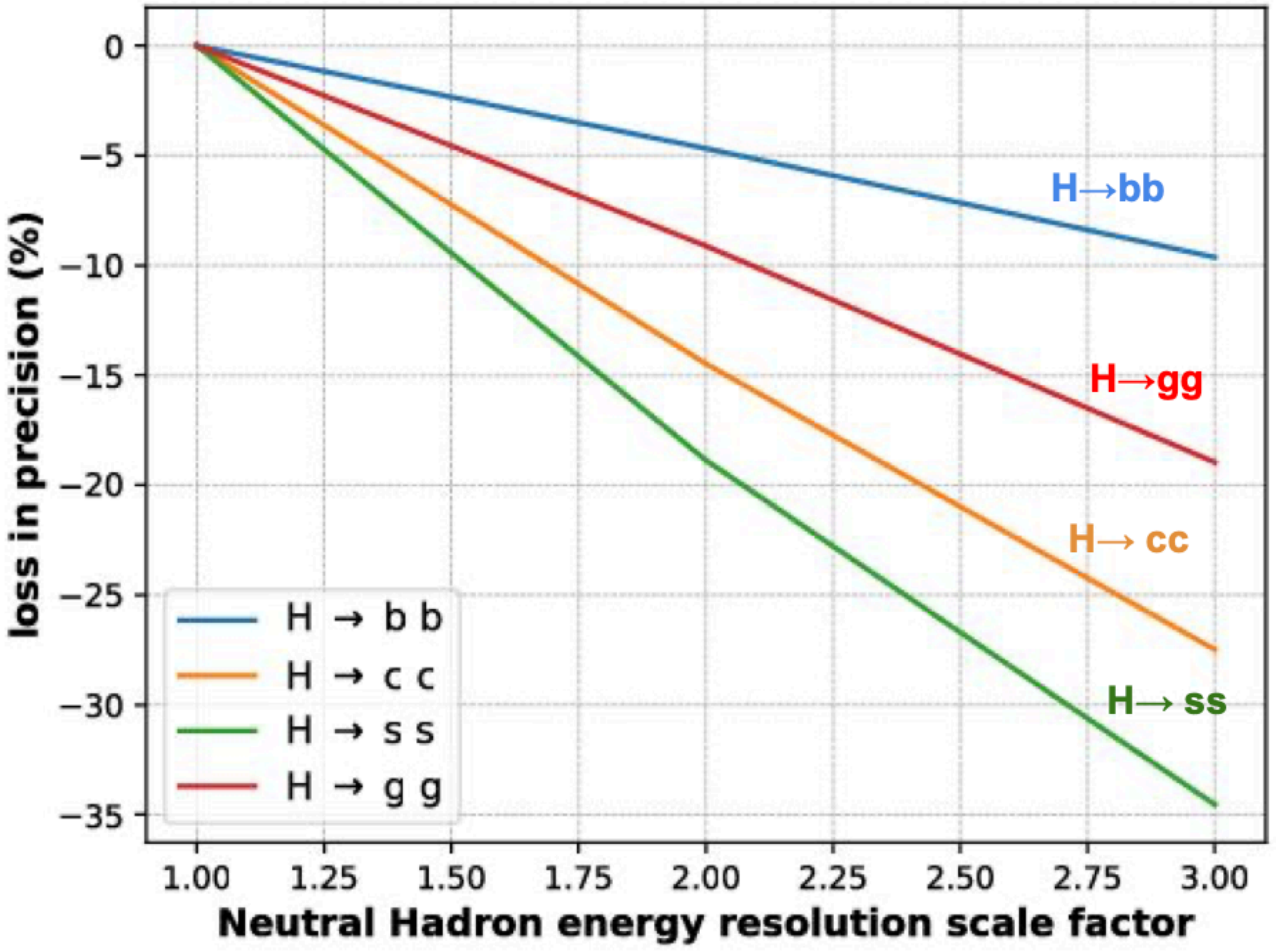
Eur. Phys. J. C **82**, 646 (2022)

arXiv:2203.07535
L. Gouskos @FCC week

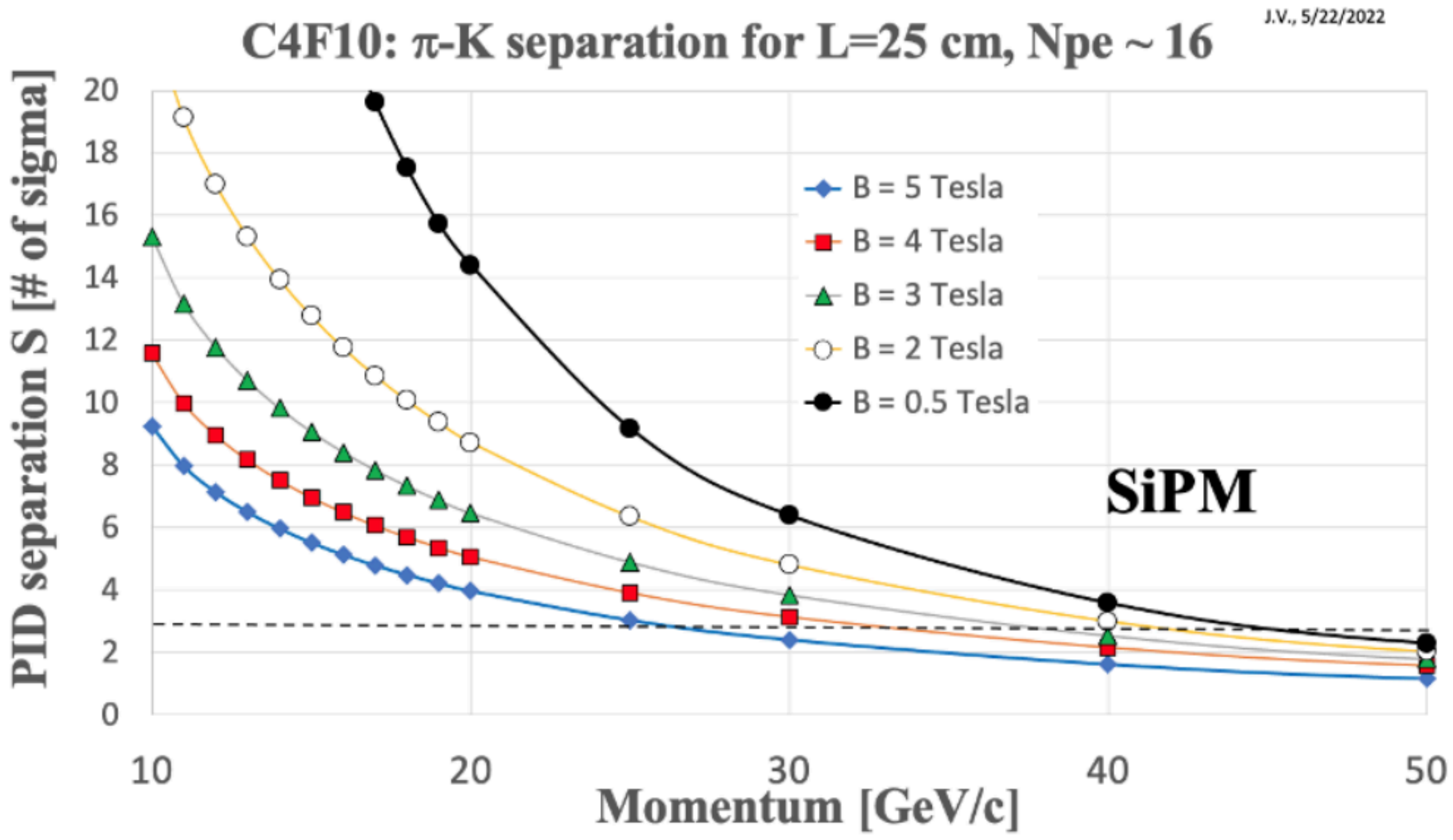
Lesson learned and moving forward

Use H → ss to inform detector design, while monitoring other benchmarks’ performance

- Neutral Hadron energy resolution
- dE/dx and dN/dx: powerful PID essential for H-strange coupling
- Timing resolution to be further investigated but less critical for s-tagging
- RHIC for improved reconstruction of $K^{+/-}$ at high momentum (< 30 GeV)



Compact RHIC

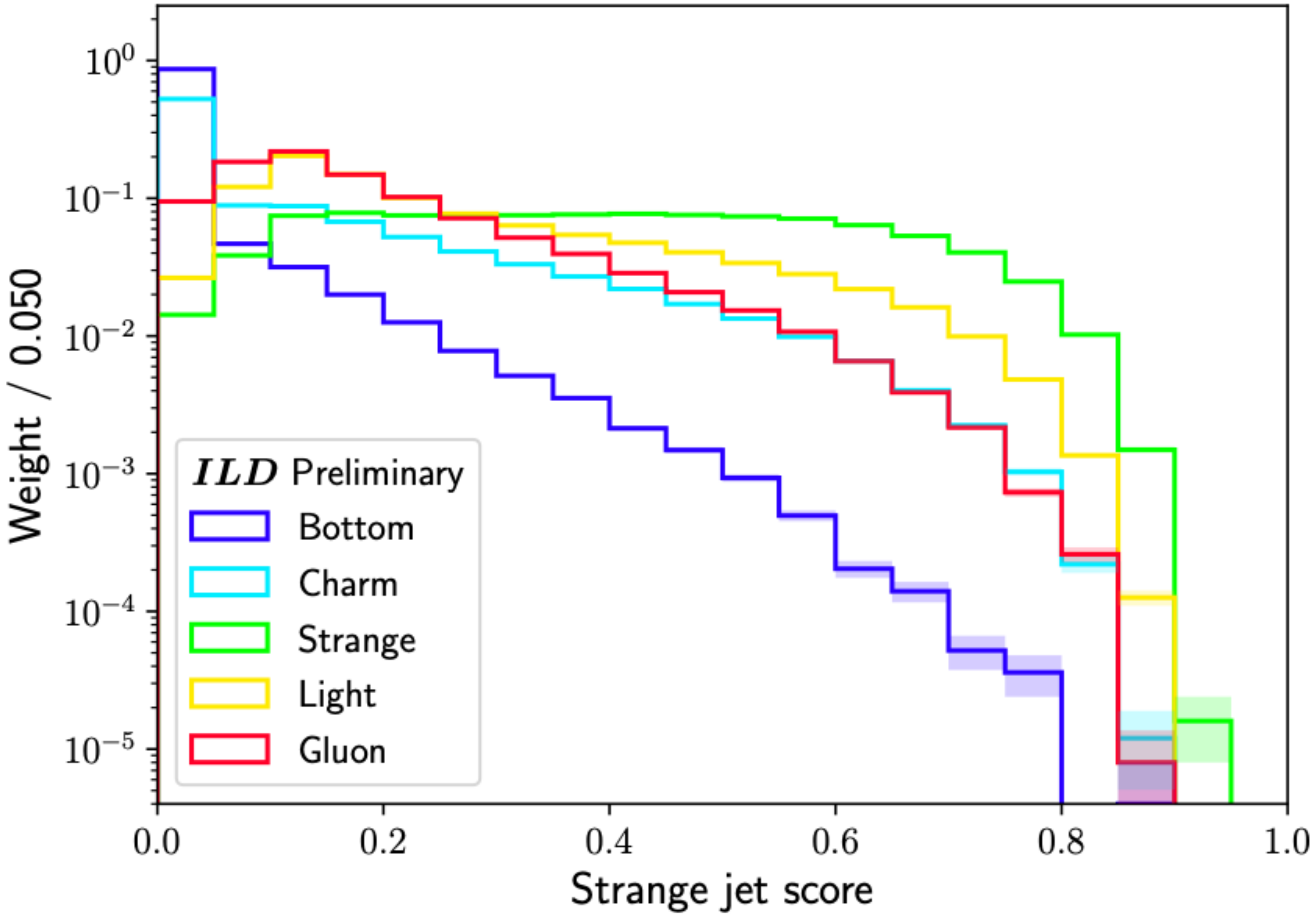


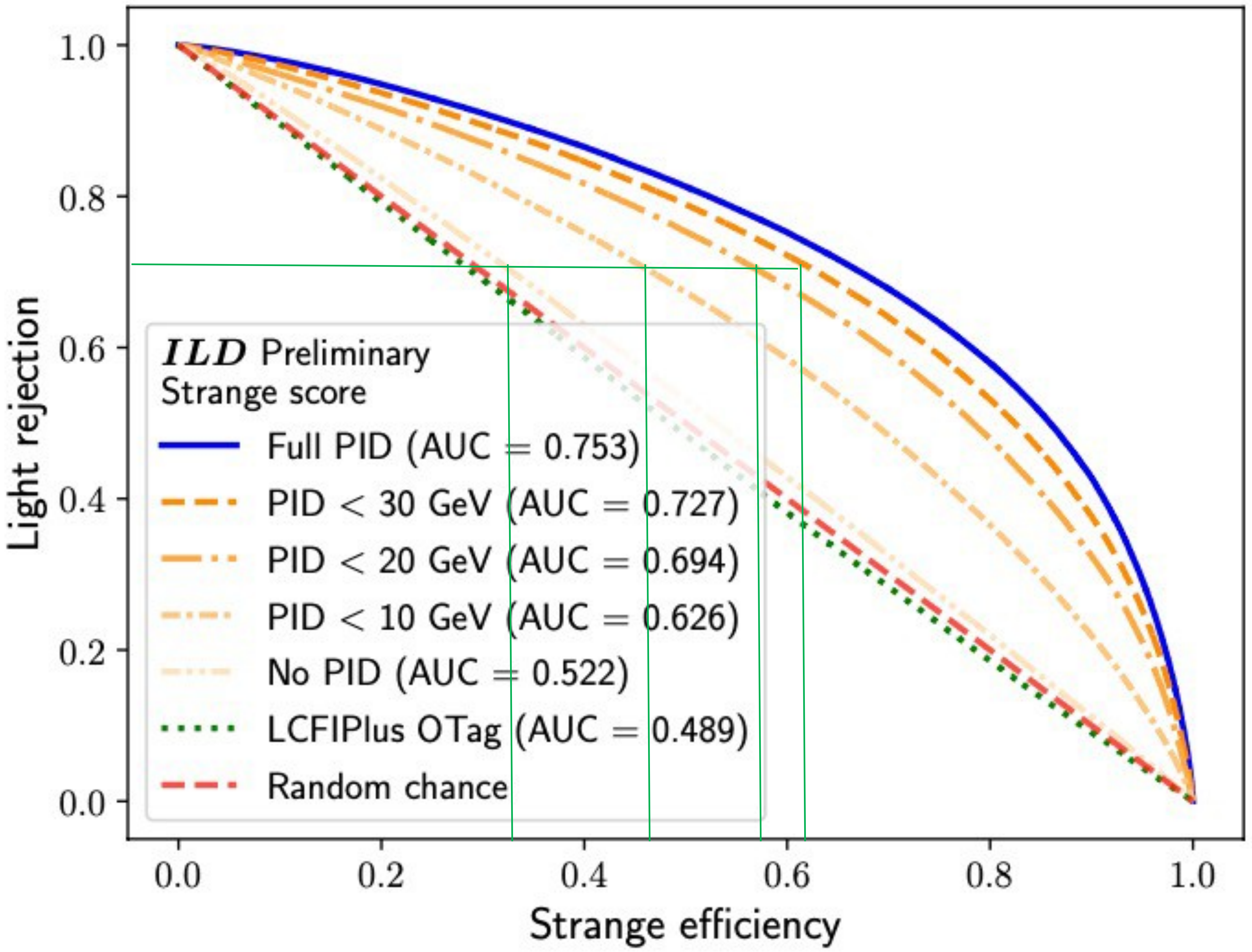
2203.07535

Strange tagging performance 2/2

ILD-like detector with full simulation and Recurrent NN

- Includes PDG-based PID → assuming perfect detector capability
- At 50% s-tag efficiency, 90% background rejection
- No PID to PID < 10 (30) GeV → at fixed mistag, 1.5x (2x) efficiency





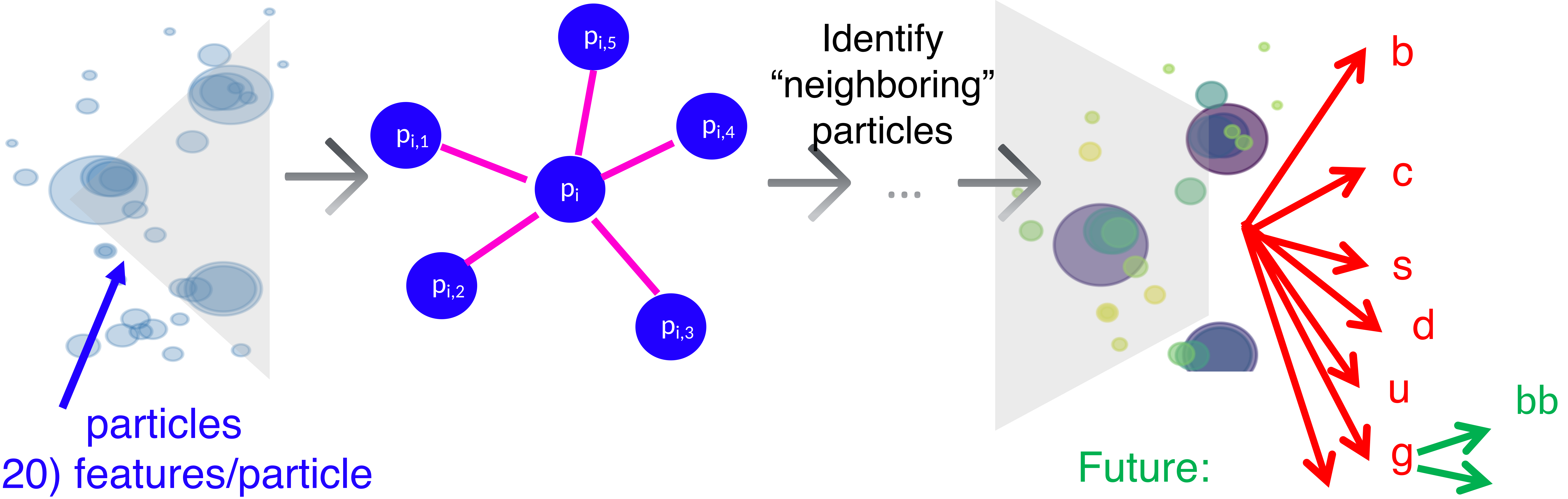
Particle cloud represented as a graph

Jet representation: Particle cloud i.e. unordered set of particles

Network architecture: Graph Neural Networks

Particles: vertices of graph; interactions b/w particles: edges of graph

Hierarchical learning approach: local → global structures



Identify “neighboring” particles

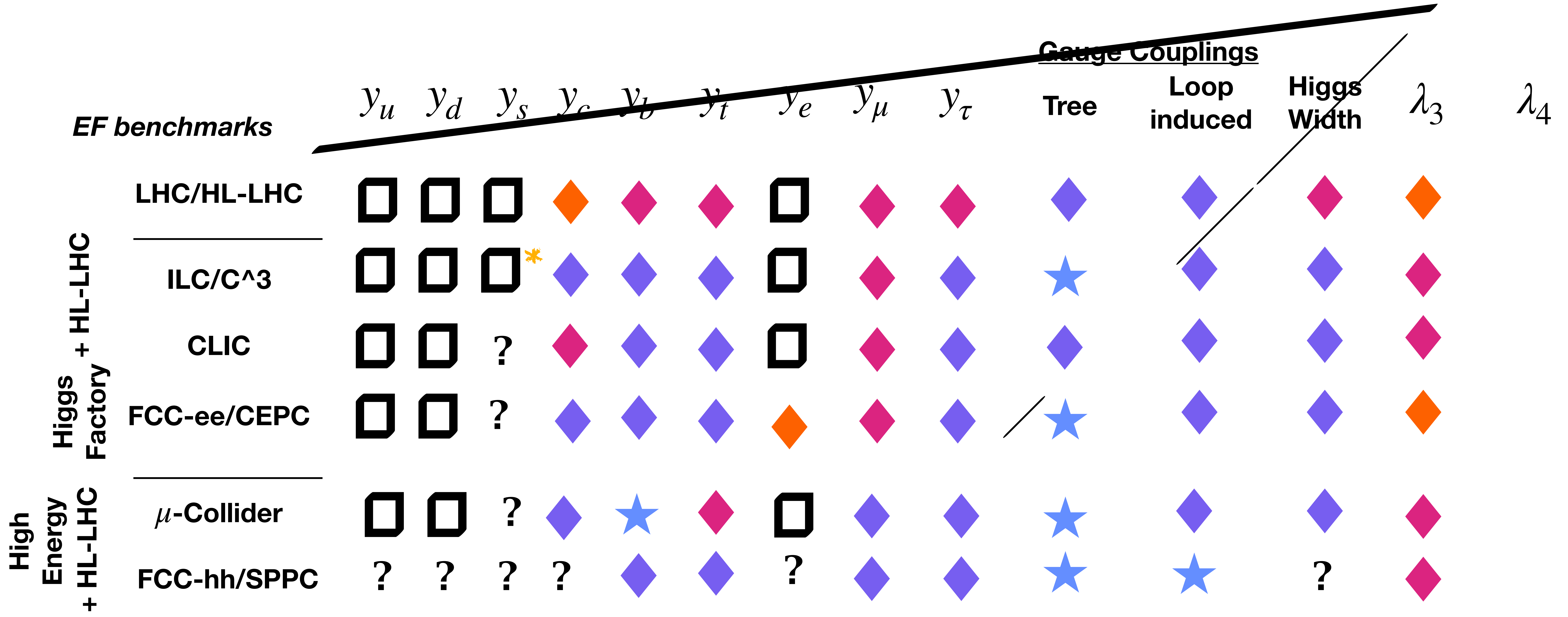
particles
O(20) features/particle

Future:
add LLP, split gluon class …

L. Gouskos

The Energy Frontier 2021 Snowmass Report

Moving forward



EF benchmarks		y_u	y_d	y_s	y_c	y_b	y_t	y_e	y_μ	y_τ	Gauge Couplings Tree	Gauge Couplings Loop induced	Higgs Width	λ_3	λ_4
	LHC/HL-LHC	☐	☐	☐	🔶	♦	♦	☐	♦	♦	🔷	🔷	♦	🔶	
Higgs Factory + HL-LHC	**ILC/C^3**	☐	☐	☐*	🔷	🔷	🔷	☐	♦	🔷	★	🔷	🔷	♦	
	CLIC	☐	☐	?	♦	🔷	🔷	☐	♦	🔷	🔷	🔷	🔷	♦	
	FCC-ee/CEPC	☐	☐	?	🔷	🔷	🔷	🔶	♦	🔷	★	🔷	🔷	🔶	
High Energy + HL-LHC	**μ-Collider**	☐	☐	?	🔷	★	♦	☐	🔷	🔷	★	🔷	🔷	♦	
	FCC-hh/SPPC	?	?	?	?	🔷	🔷	?	🔷	🔷	★	★	?	♦	

Order of Magnitude for Fractional Uncertainty ★ $\lesssim \mathcal{O}(10^{-3})$ 🔷 $\mathcal{O}(.01)$ ♦ $\mathcal{O}(.1)$ 🔶 $\mathcal{O}(1)$ ☐ $> \mathcal{O}(1)$? **No study Beyond HL-LHC**

ATLAS Twiki

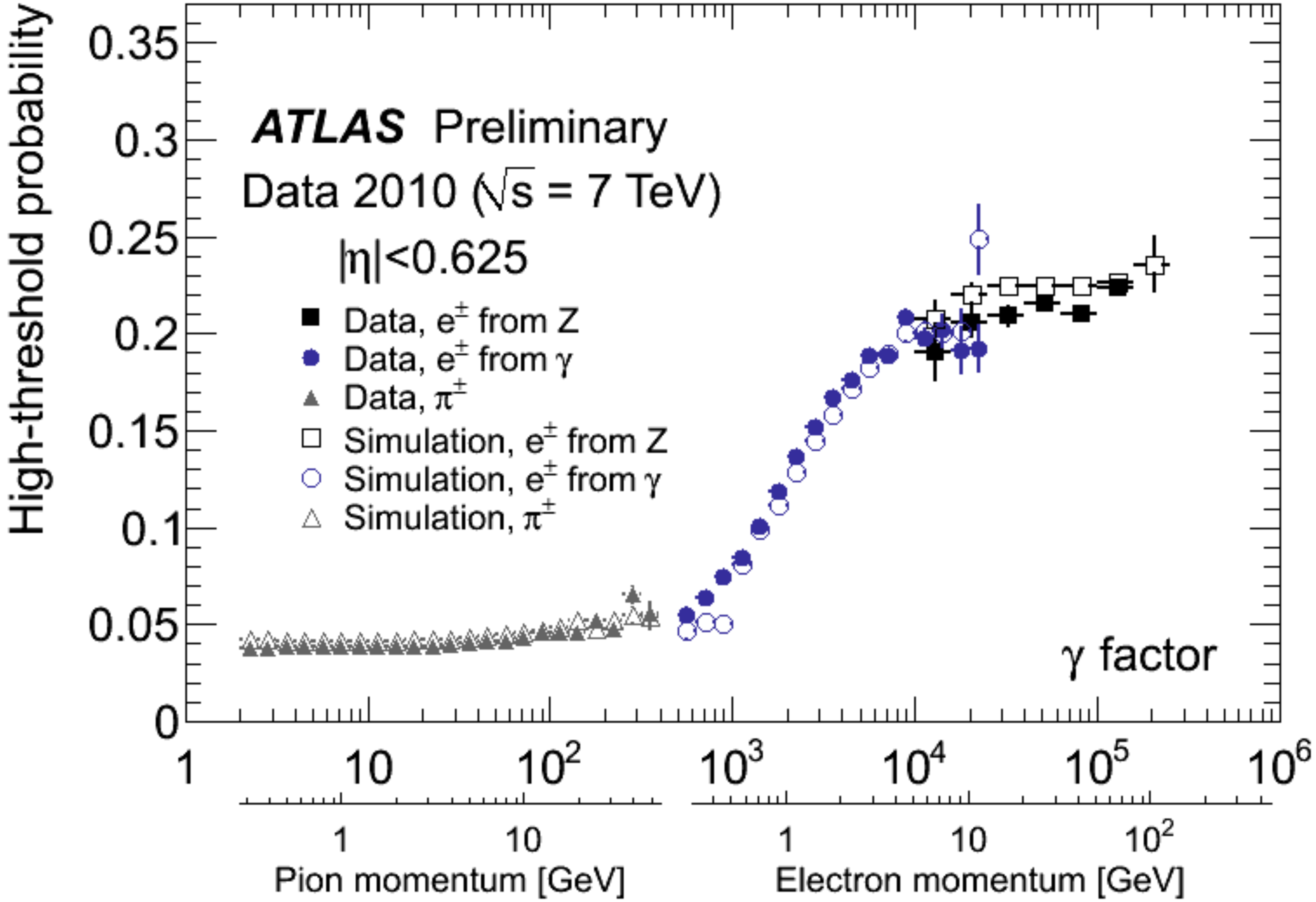
e/π separation with TR+dE/dx

e/π separation via detection of transition radiation photons

Transition radiation is emitted when a highly relativistic charged particle with a Lorentz factor γ > 10^3 traverses boundaries between materials of different dielectric constants.

To achieve the best e/π separation, TR and dE/dx-based measurements are combined in a single likelihood function for a particle type.

The HT fraction is defined as the fraction of hits on track that exceed the high threshold (6-7 KeV)



Light Yukawa ?