Tagging each jet flavour at FCC-ee with a Transformer-based Neural Network

Freya Blekman (DESY+UHH), Florencia Canelli (UZH), Alexandre De Moor (VUB),
Kunal Gautam (VUB+UZH), Armin Ilg (UZH), Anna Macchiolo (UZH), Eduardo Ploerer (VUB+UZH)

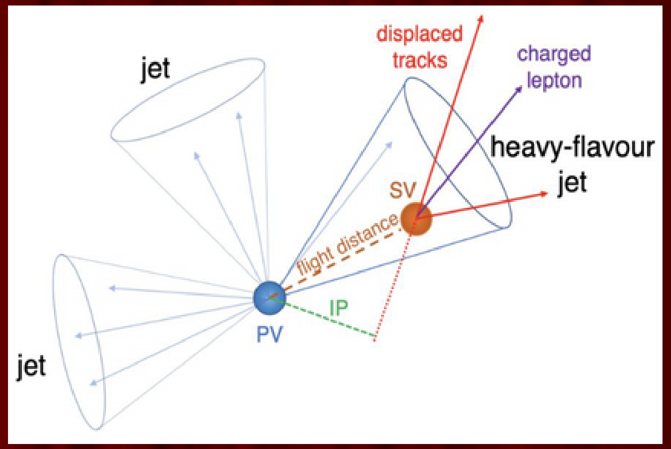
reference: arXiv:2406.08590 (under EPJC review)

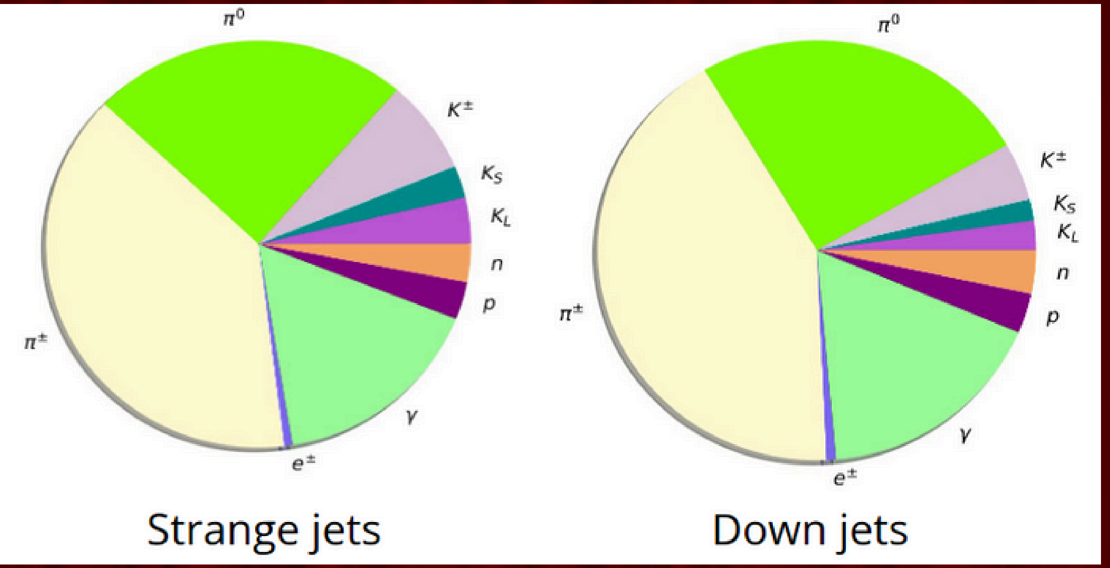
Setting the Scene

- Identification of hadronic final states is an essential to collider experiments
- Future lepton collider such as FCC-ee offer much cleaner environment than hadronic collisions (Initial state kinematics known, no PDFs, no QCD ISR, …)
- Distinguishing features:
    - Differing colour factors for q vs g
    - Displaced SVs for b/c's
    - Kaon excess for s
    - Jet charge for up/down

$$C_A/C_F = \frac{9}{4}$$

$$Q_\kappa = \frac{1}{p_{T,jet}^\kappa}\Sigma_j q_j (p_T^j)^\kappa$$





Simulation samples

- Pythia8.303 for event generation
- Assuming FCC-ee style scenario on Z resonance (6 10e12 decays expected) 125 ab-1

- Delphes + IDEA detector

- FastJet-3.3.4 exclusive (N=2) e+e- kT algorithm
- Physics process:
  - Z->qqbar
  - also use Z(->vv)H(->qqbar) for cross check/comparison to H performance
- Using 'MC truth' Z->qqbar quark flavour
  - Allows comparison to other taggers
  - Paper also discusses pros/cons ghost MC matching)

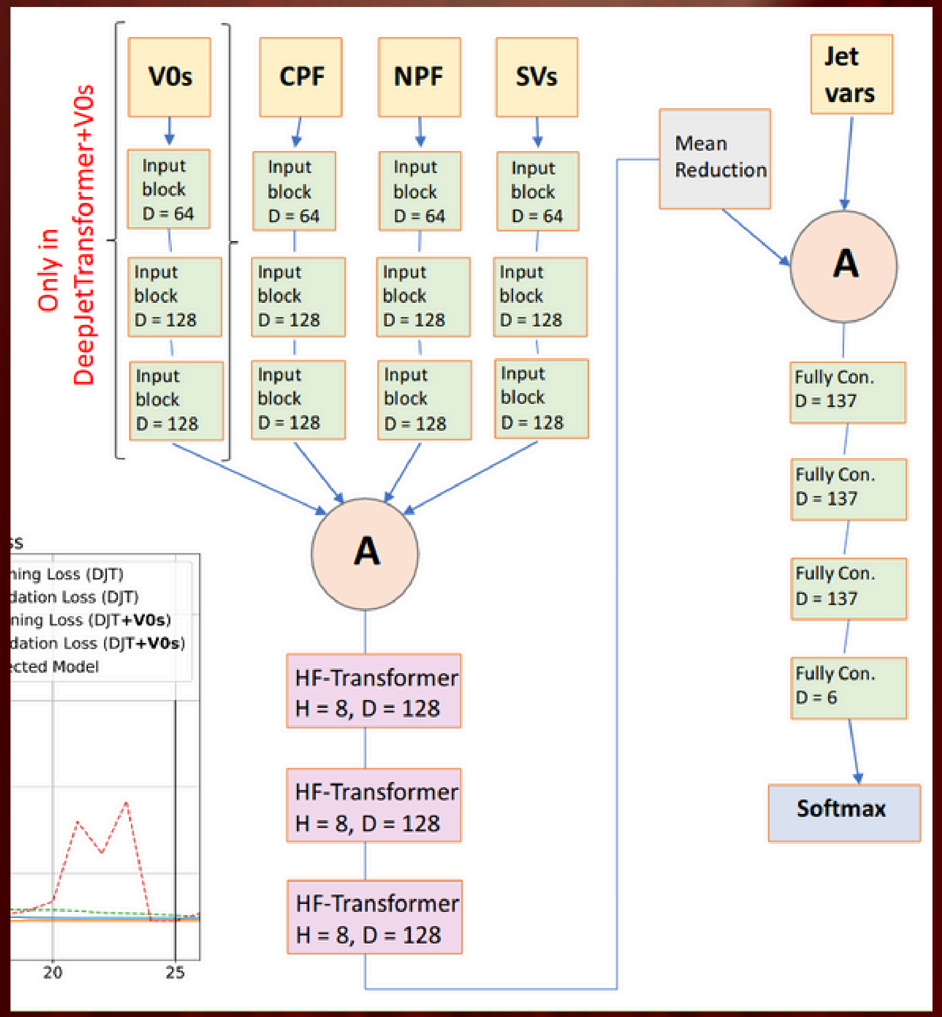
Episode 1

The Neural Network

DeepJetTransformer

DeepJetTransformer is:

- Transformer-based architecture achieving state-of-the-art performance using an encoder-decoder architecture (=it’s fast to train!)
  - Only 2 hours of training to converge after approximately 50 epochs on an NVIDIA Tesla V100s GPU
- Self-attention allows dynamic assignment of weights to individual elements within the jet capturing intricate dependencies across the entirety of the jet structure (=it’s efficient)
- More lightweight/still performant,˜1M trainable weights, only 65k per encoder layer



Training

Pick epoch before overtraining (normal)

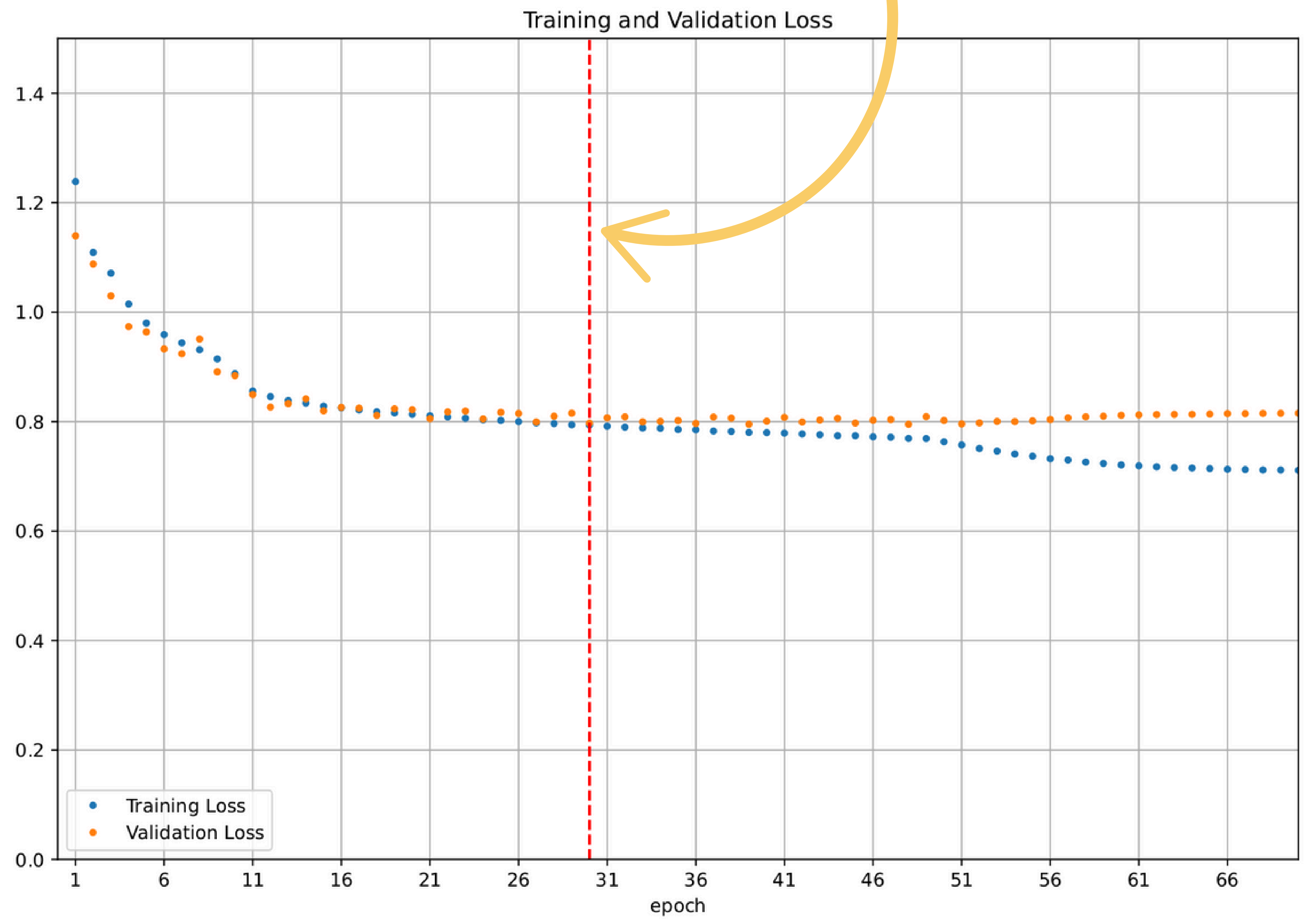
-

Trained network with 10^6 Z ->qqbar jets (80%/20% train/validation), evenly split into b, c, s, u, d

Implemented in Pytorch (v1.10.1)

- 70 epochs w/ batch size of 4000 trained in ~2 hours
- No obvious overfitting/overtraining
- Categorical cross entropy as loss function

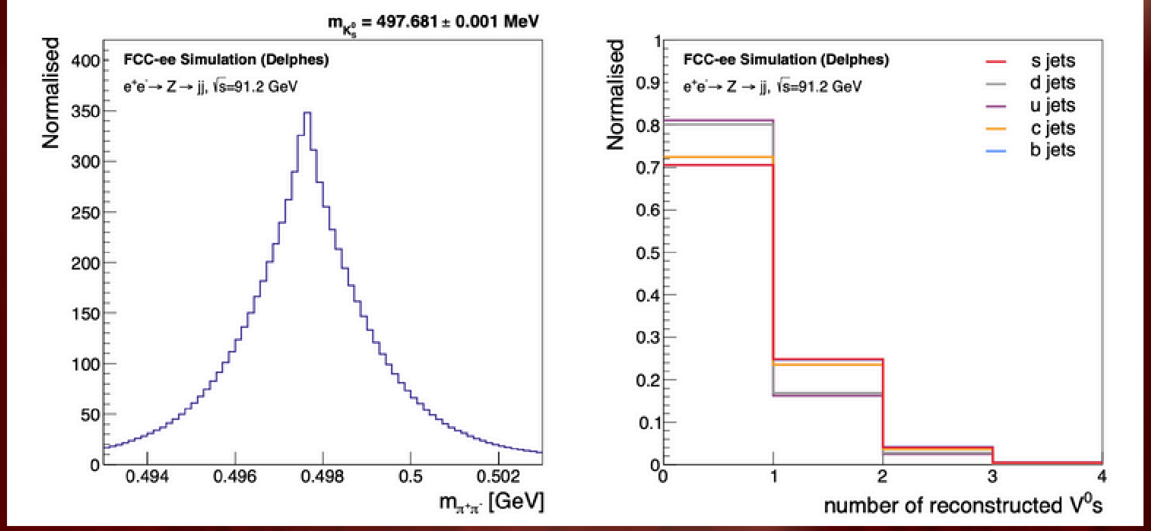
$$L(\mathbf{y}, \mathbf{p}) = -\Sigma_i^C y_i log(p_i)$$

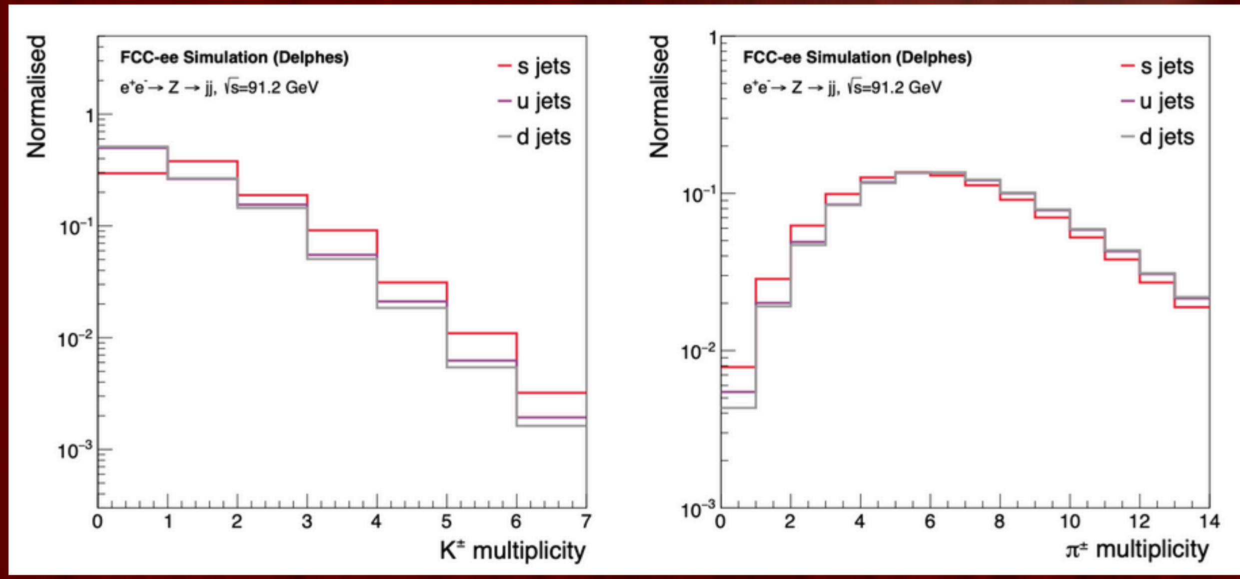


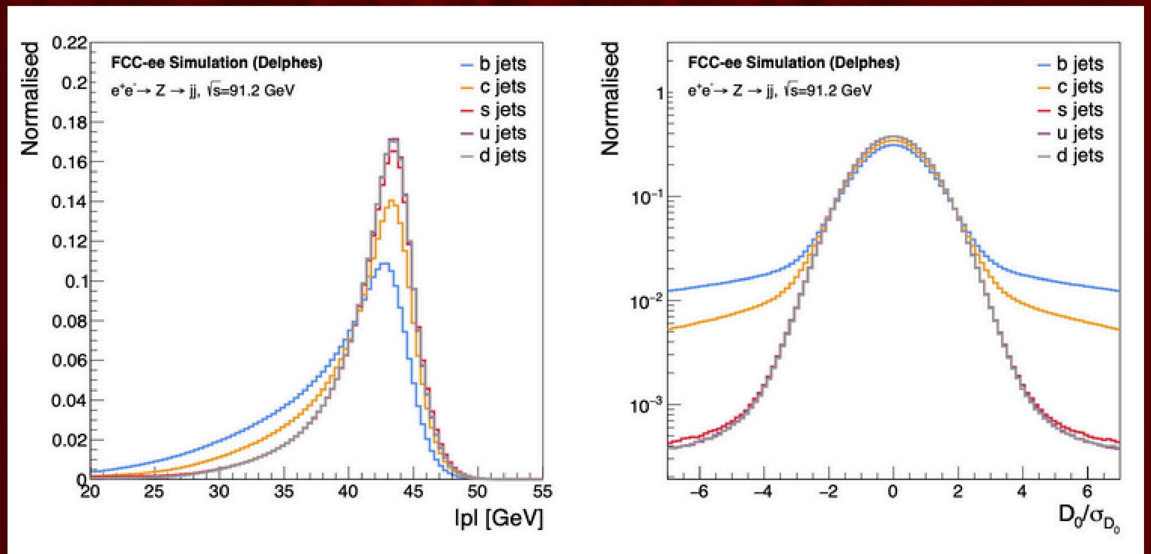
Physics inputs

- Using low-level information (particle flow four vectors, track charges, PID flags) as is nowadays the norm for ML at LHC
  - NN can make vector sums, etc, and decide which are good discriminators internally
- plus still includes secondary vertex information
- reconstructed with LCFIPlus (implemented in FCCSW)

- **Includes V0s** and SVs





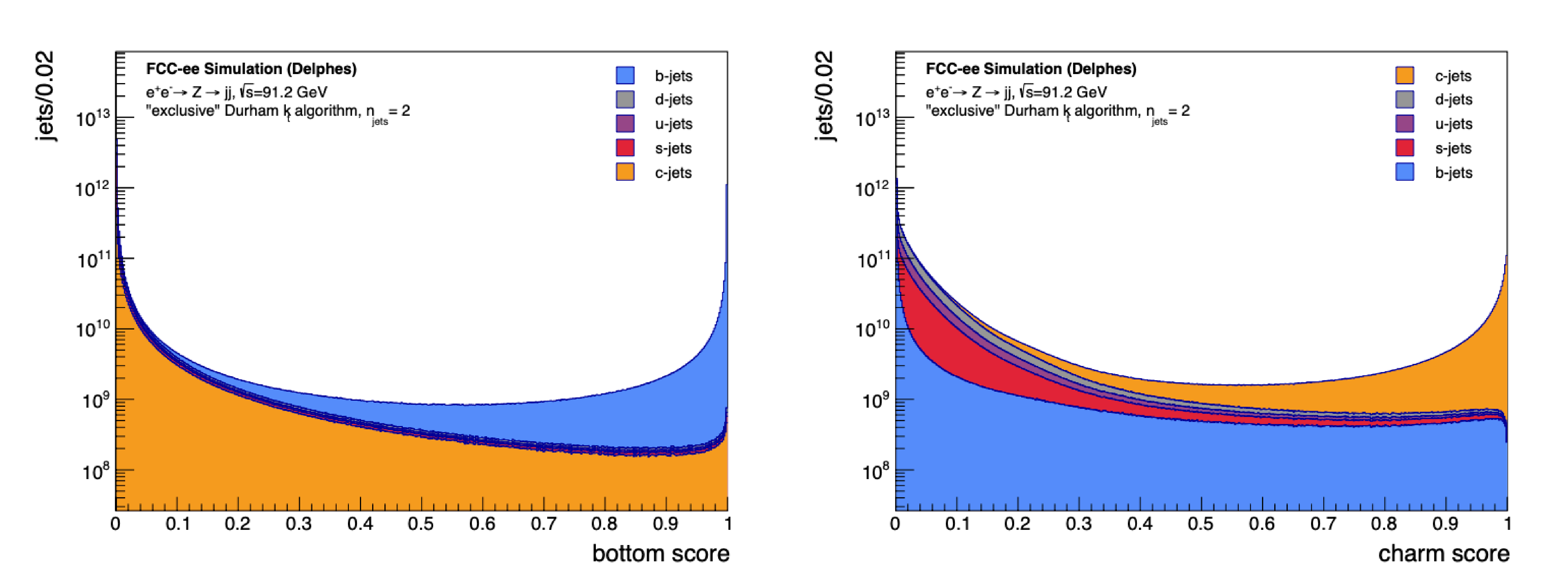


Episode 2

The performance

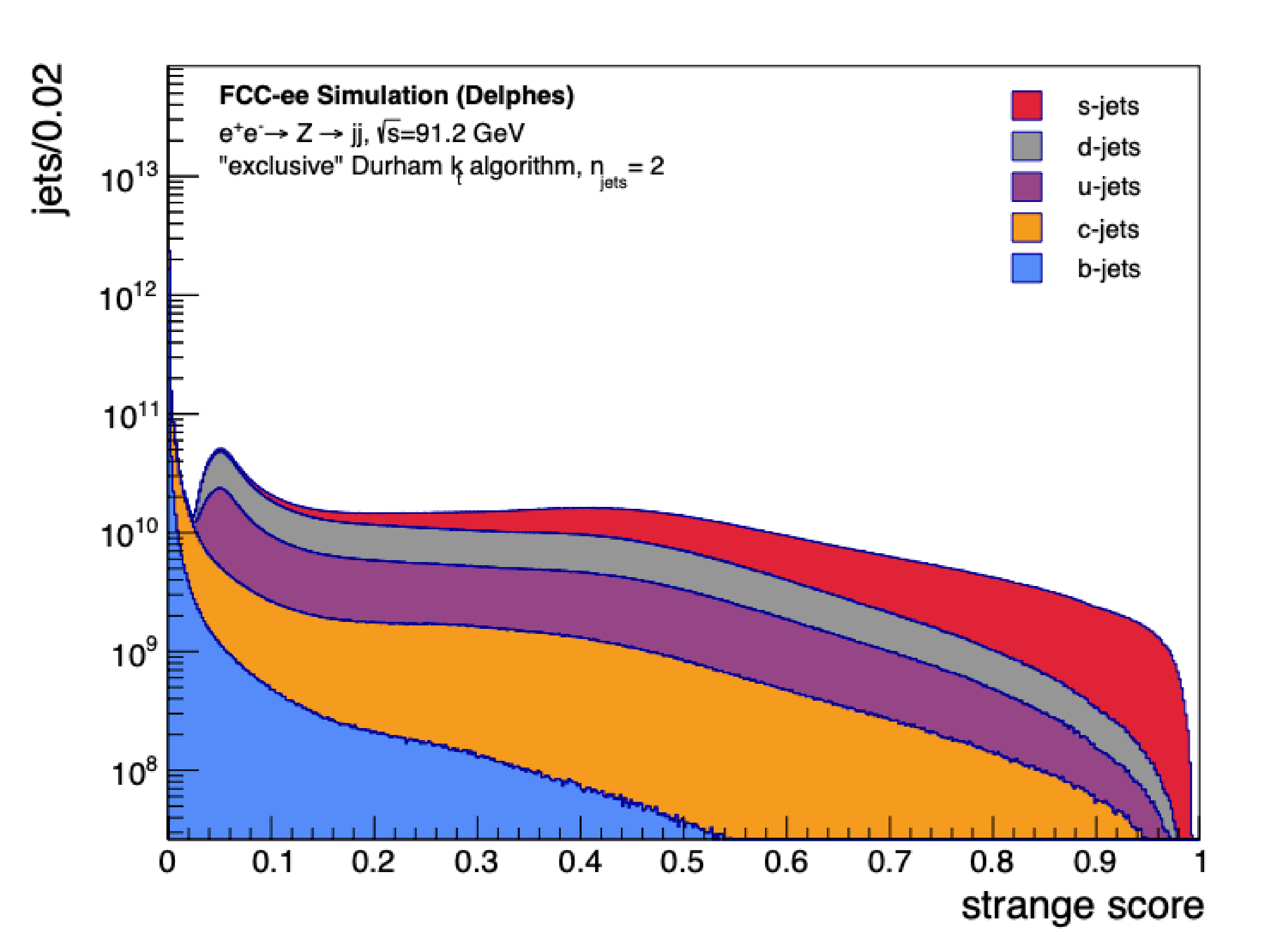
Performance: b & c

- Already known that LHC tools are excellent at b and c identification
- Great impact parameter resolution from IDEA helps
- For LHC taggers, V0s are traditionally rejected
    - reason: no interest in lighter quarks typically
    - but we do see they help a little bit for b/c too!



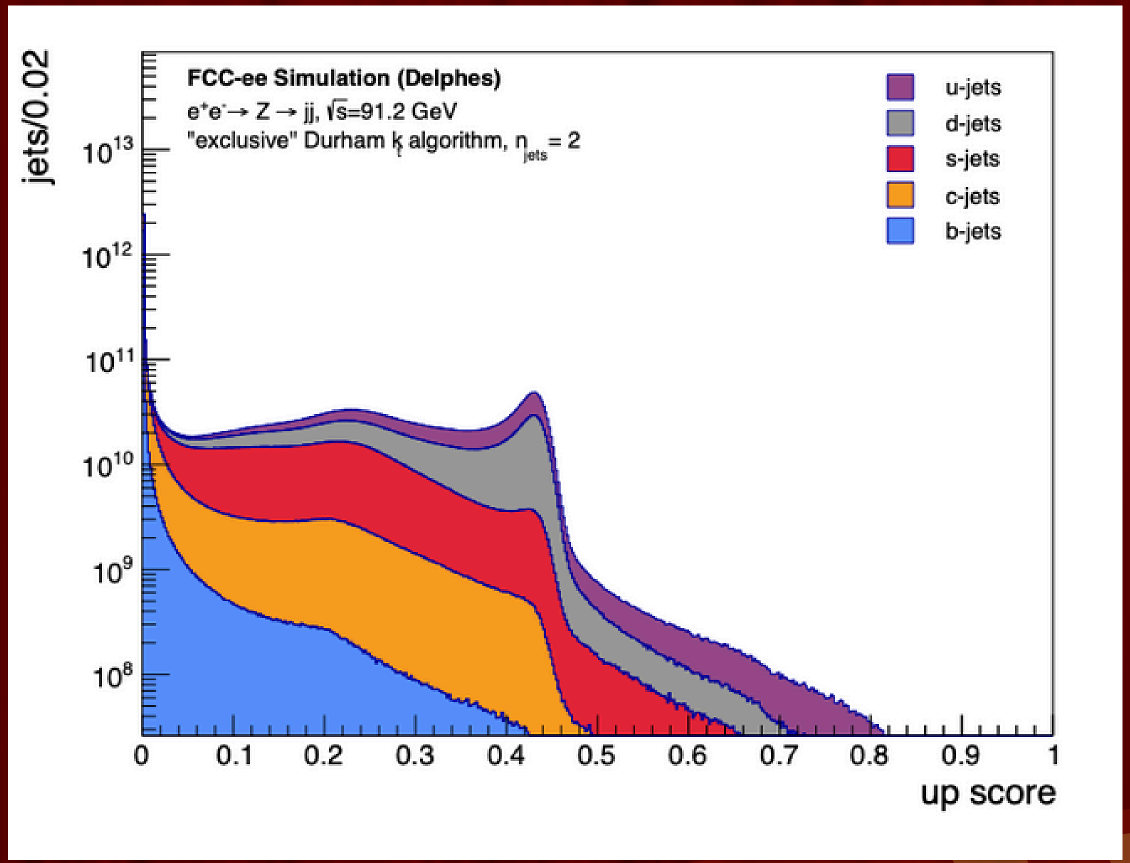
Performance: strange

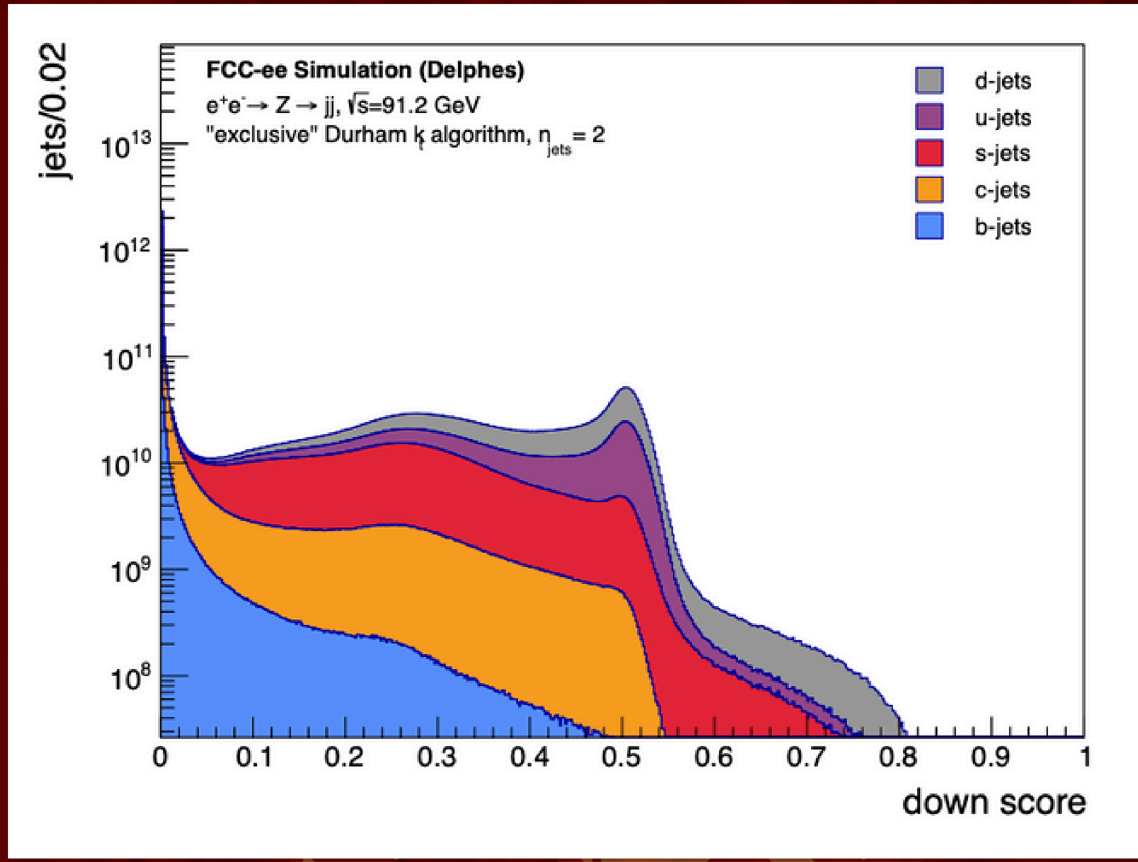
- Strange tagging relevant for H->ssbar
  - also interesting for FCNC, BSM searches and the rest of SM precision program
- Clear separation of s jets
  - main background: u, d
  - separation from b,c (relatively) straightforward
- We focus on the Z resonance and are interested in exploring physics opportunities there
- Also likely Z runs will be first runs at any future e+e- machine, so will be the benchmark eventually



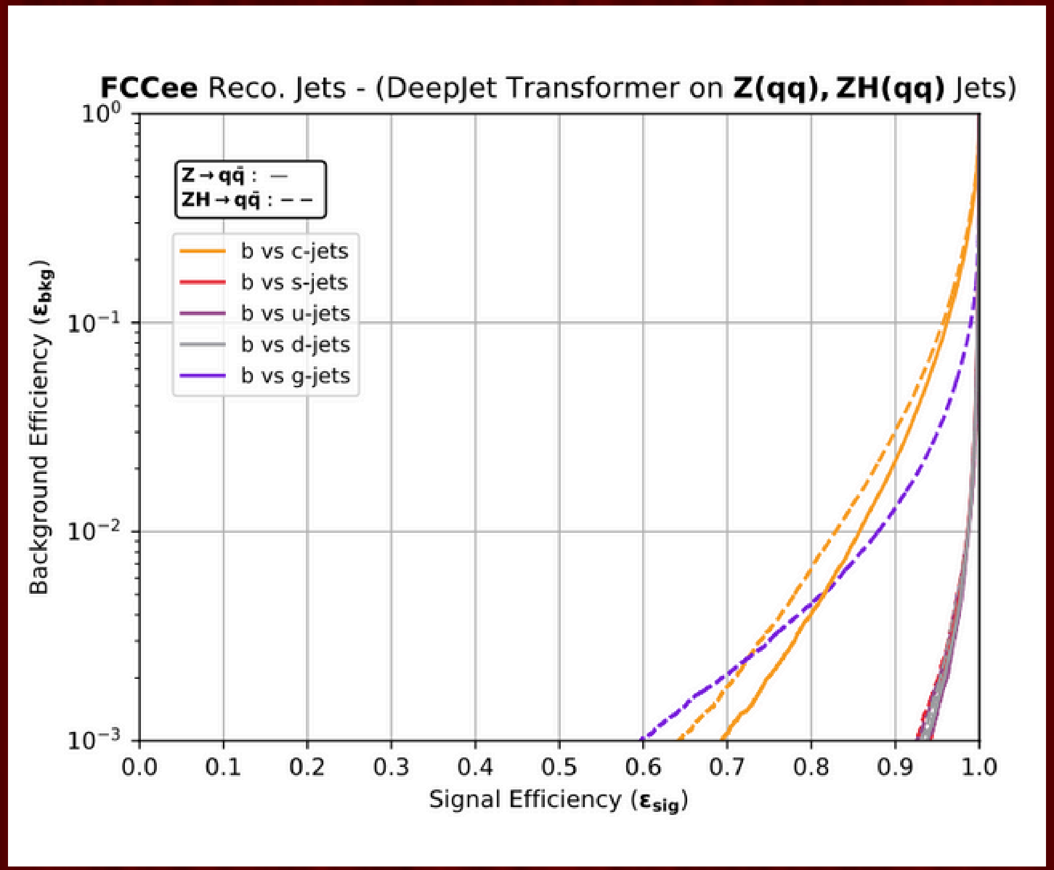
Performance: up,down

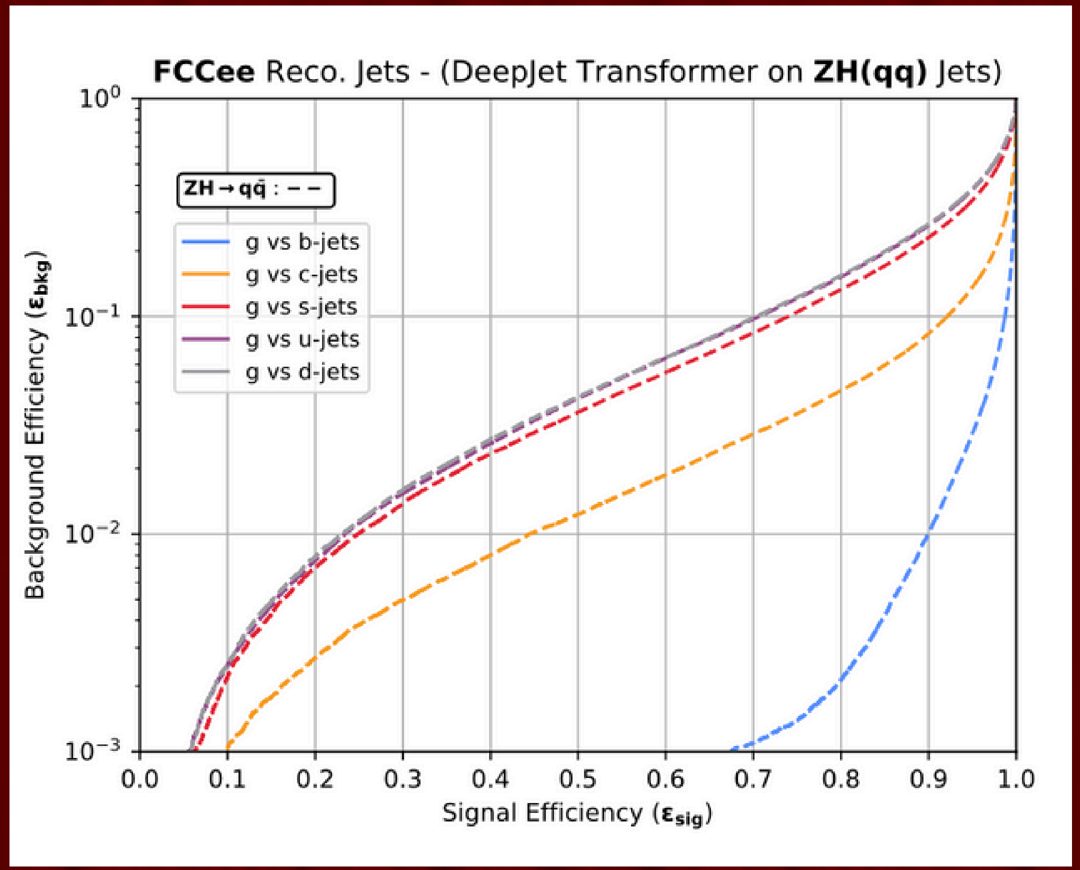
- Important to note that there also is separation power between up and down!
- Previous colliders full separation of each quark not available
    - physics potential/opportunity?
- gluon separation not really considered as we work on Z pole
    - with exclusive eekT, so gluons included in jet

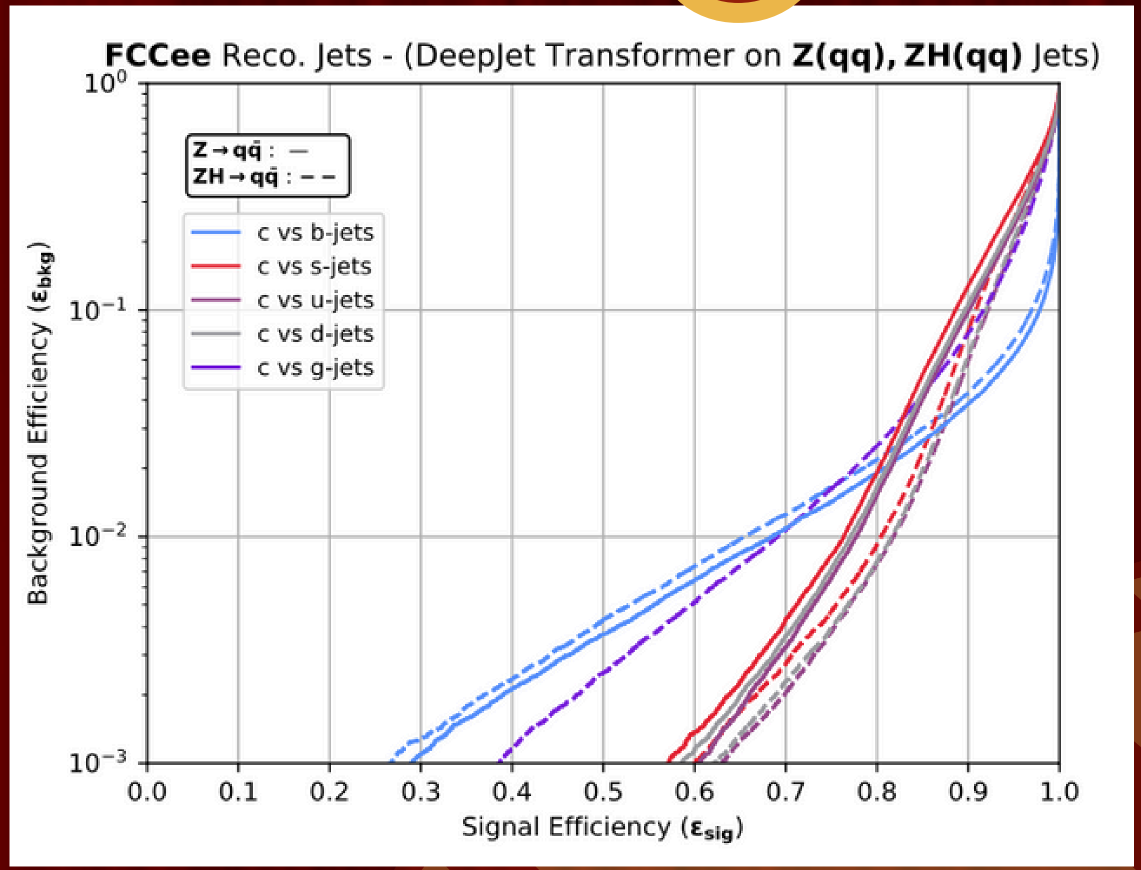




ROC Curves: b, c, g

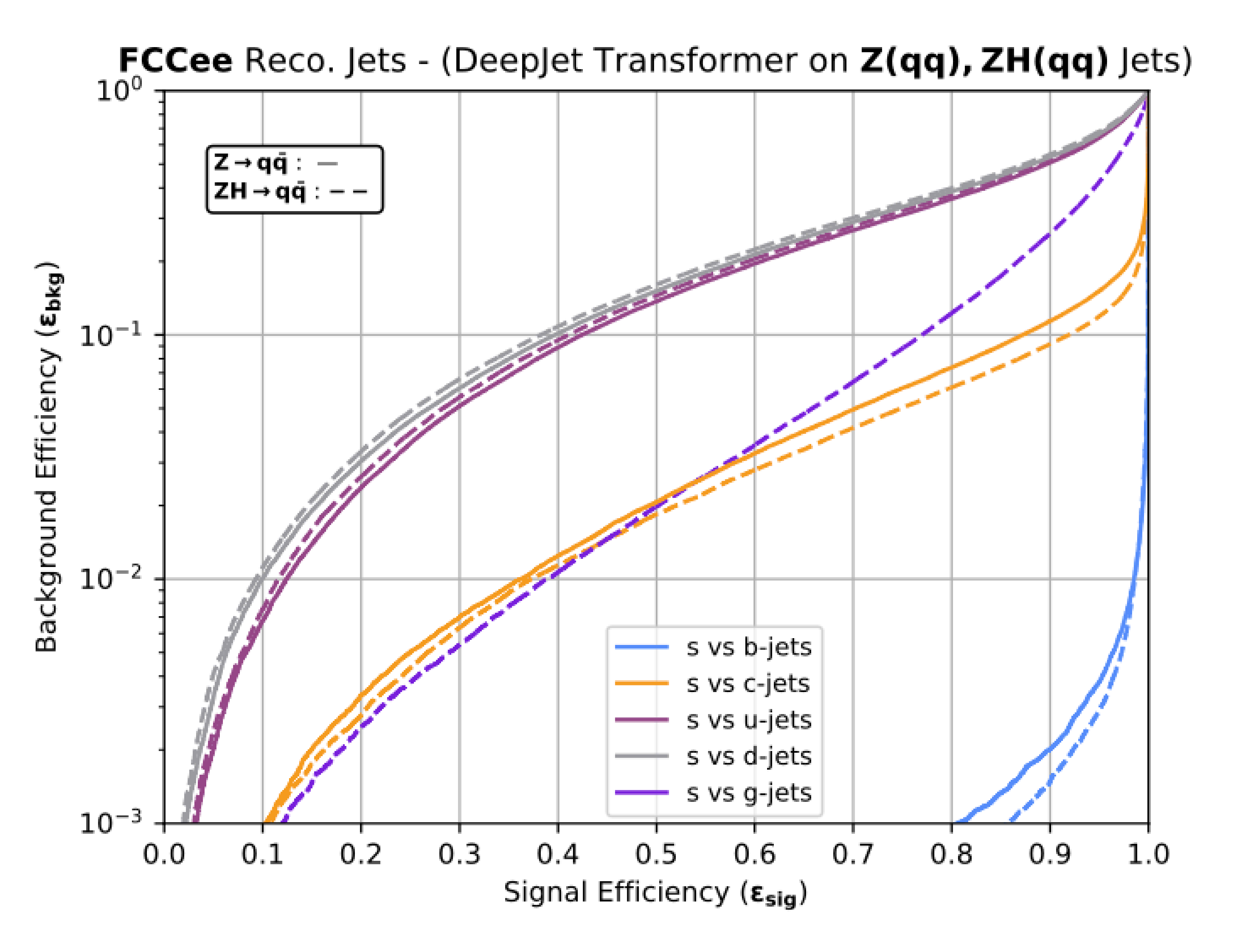






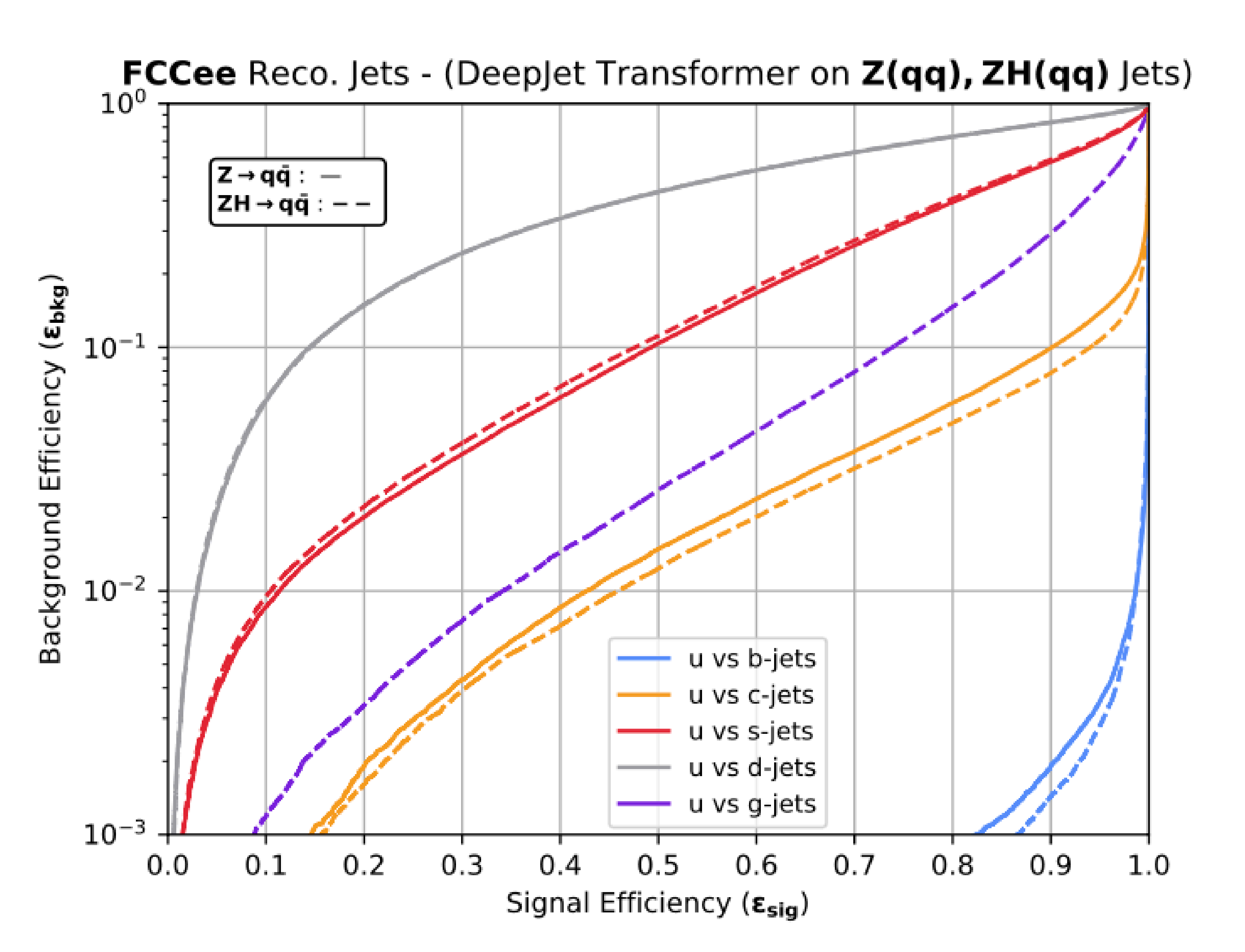
- As expected (and confirmed by many other tagger studies):
  - Excellent performance: discrimination of b jets wrt light jets w/ 90%+ at bkg eff 0.1%
  - gluon splitting to bbbar will be interesting challenge for H
- How excellent is excellent enough?
- start genuine link to detector performance (previous talk) really great!

ROC Curves: strange



- For s-tagging, up and down jets present by far most challenging background
- PID is central to this type of discrimination
- Charm and gluon jets present second most challenging, likely due to
- Charm hadron decay to strange hadron
- g->ss gluon splitting discrimination becomes important at high efficiencies

ROC Curves: u,d



- most difficult challenge to disentangle
- still some separation: up jet vs down jet discrimination with sig eff ~15% and bkg eff 10%

Episode 3

What detector effects are important?

Variable importance

Variable Class		Jet-level	Charged	Neutral	SV	V^0
$\epsilon_{bkg} = 10\%$	b vs c	2.4%	62.4%	2.2%	13.9%	0.1%
	c vs s	1.2%	65.7%	2.9%	29.6%	0.2%
	s vs ud	7.6%	59.4%	21.8%	5.0%	16.4%
$\epsilon_{bkg} = 0.1\%$	b vs c	6.6%	97.0%	8.0%	89.9%	0.6%
	c vs s	9.3%	96.1%	11.0%	77.9%	0.2%
	s vs ud	35.9%	91.0%	57.3%	7.4%	43.8%

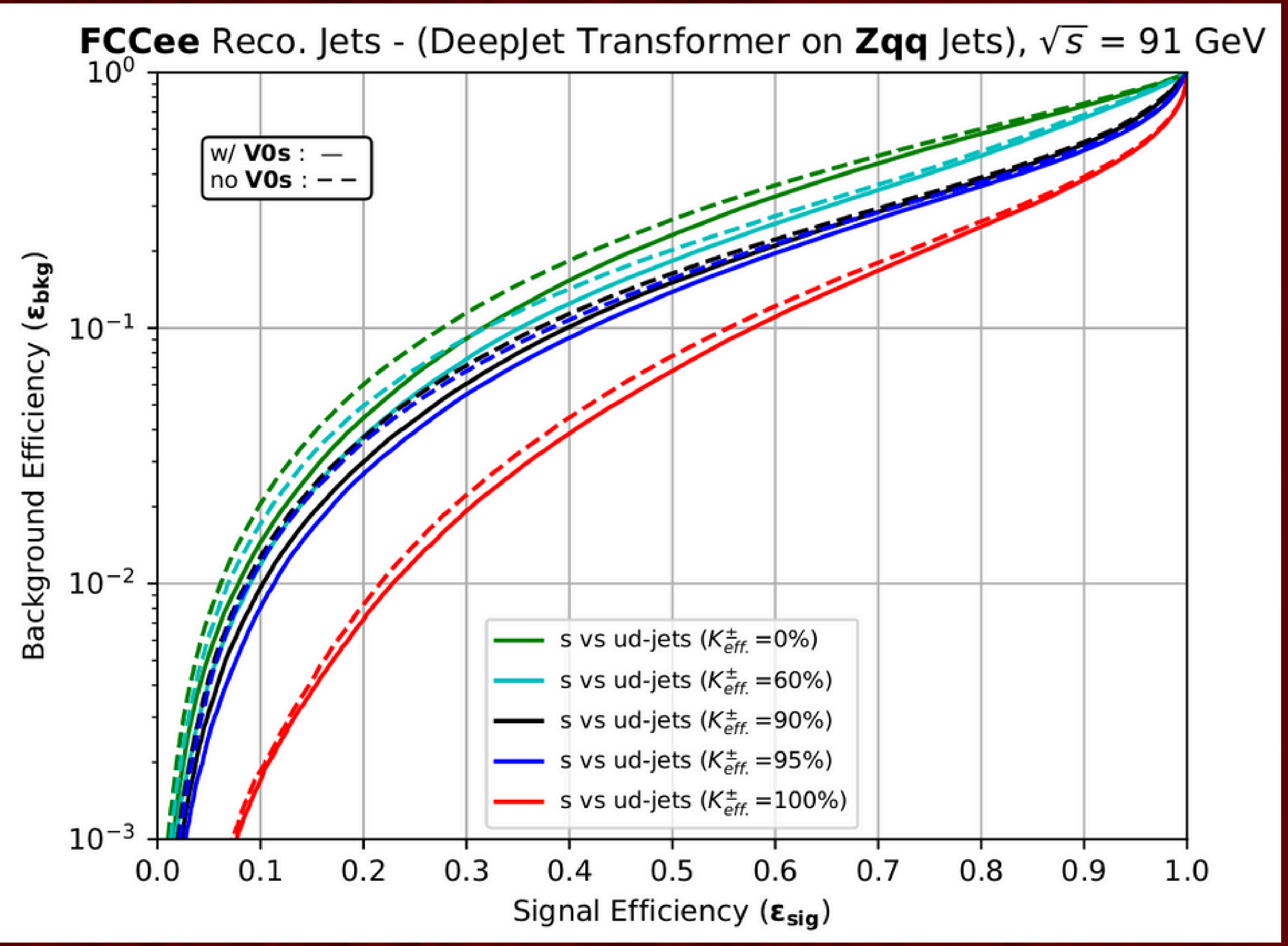
- Shuffle entire group of variables (e.g. Neutral RP variables) amongst different jets to estimate importance
- Consider % change in signal efficiency at fixed background efficiency of 10% for s vs ud, c vs s, b vs c

Detector performance importance

Variable		$\ln(E_{ch.})$	isPhoton	$K^{\pm}$ID	m^{SV}	p^{V^0}	z_0	D_0/σ_{D_0}
$\epsilon_{bkg} = 10\%$	b vs c	3.5%	0.3%	0.2%	3.0%	0.1%	7.8%	11.6%
	c vs s	23.8%	0.7%	0.5%	0.3%	0.2%	20.9%	39.1%
	s vs ud	12.8%	16.6%	38.8%	0.0%	9.2%	23.3%	26.7%
$\epsilon_{bkg} = 0.1\%$	b vs c	13.8%	1.3%	0.9%	67.2%	0.8%	34.1%	45.0%
	c vs s	57.6%	0.9%	4.8%	7.0%	0.3%	56.2%	79.5%
	s vs ud	35.0%	28.0%	59.0%	0.4%	34.7%	60.5%	80.1%

- Photon identificationon (pi0s) important for s vs ud
- charged Kaon identification important for s vs ud
- V0 measurements affect s vs ud
- impact parameter overall important (reason for impact s vs ud related to b vs s vs ud separation)

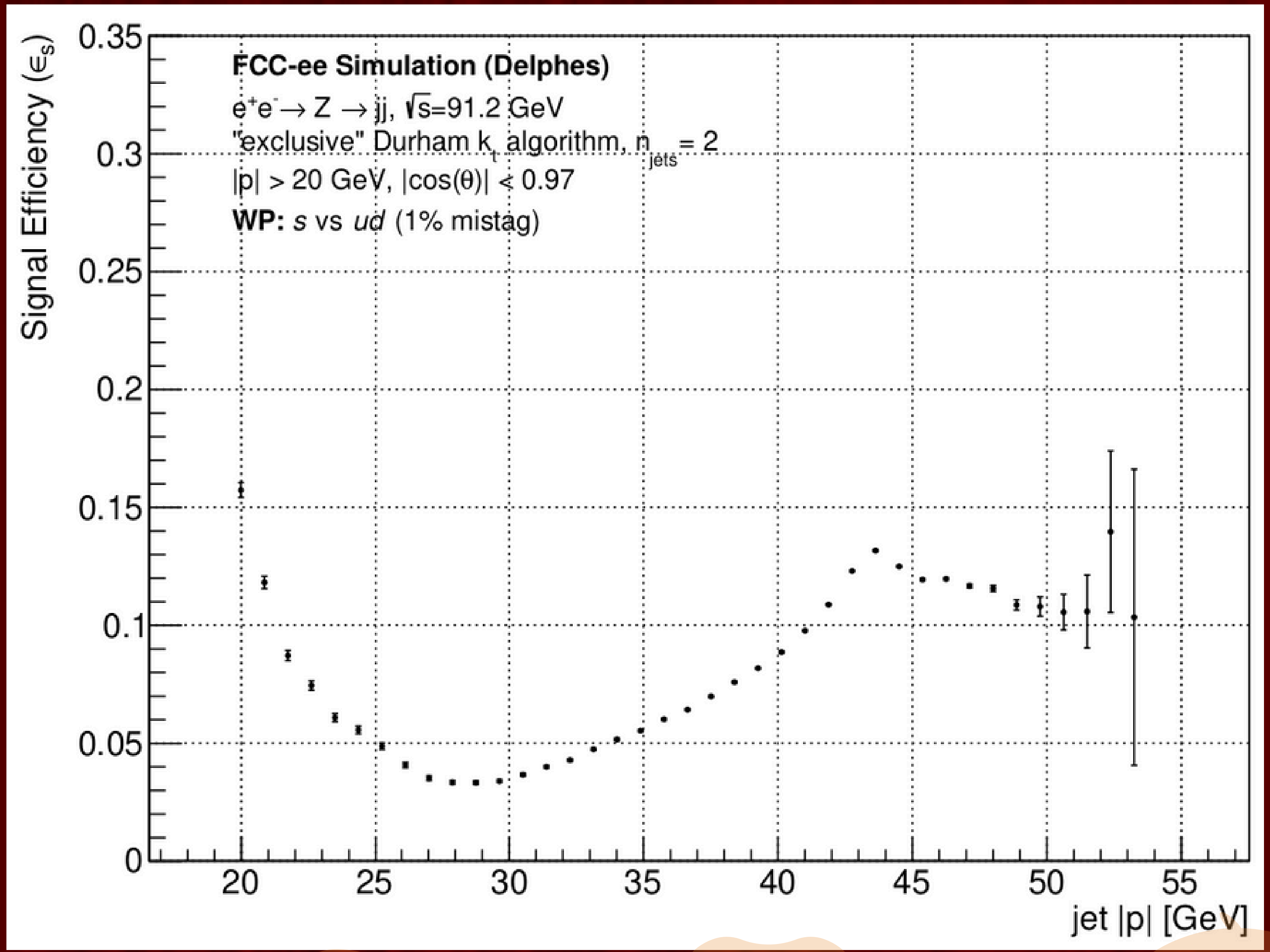
s-tagging: Particle ID Impact

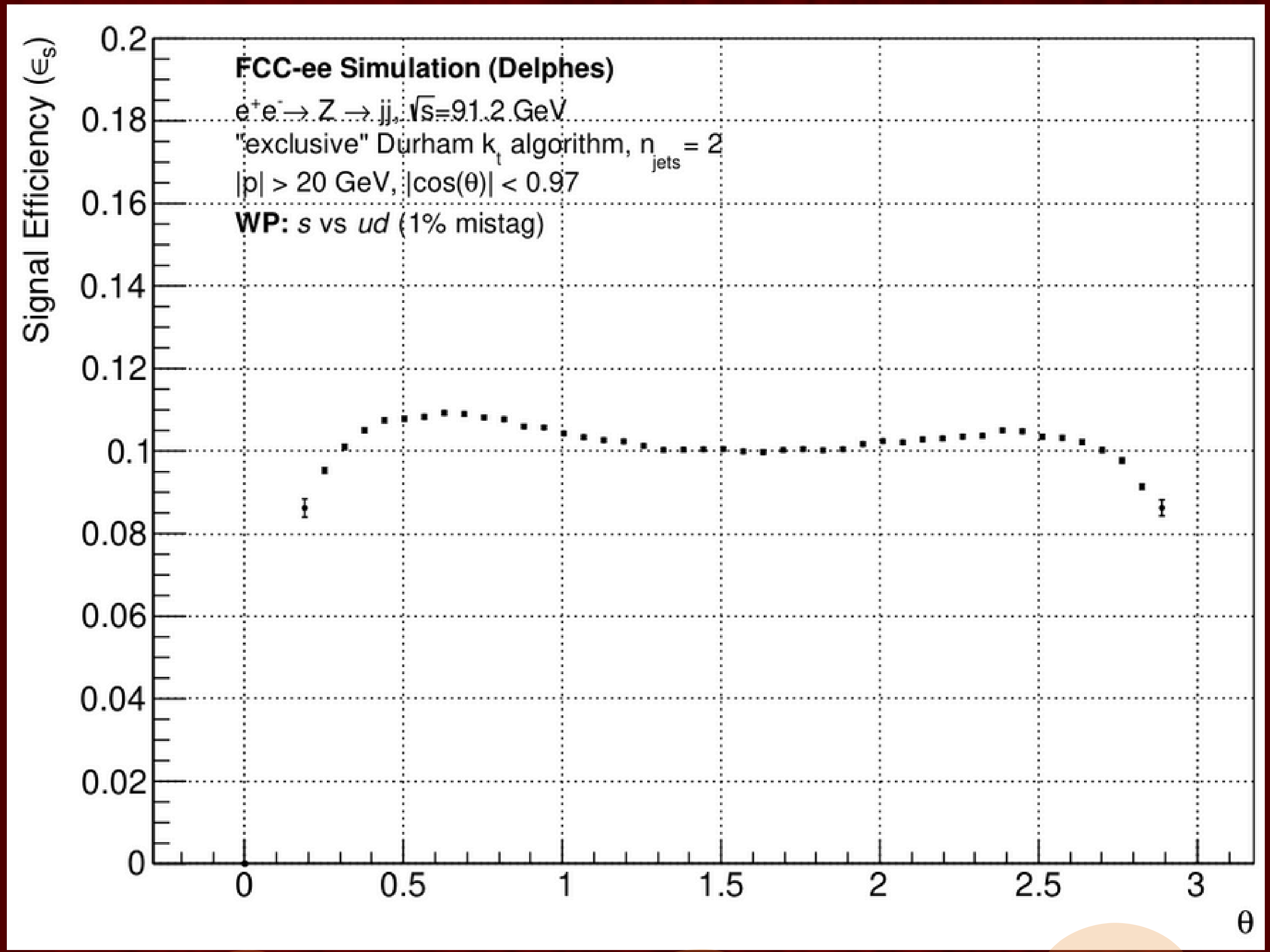


$K^{\pm}$ ID efficiency	0%	20%	40%	60%	80%	90%	95%	100%
$\pi^{\pm}$ misID efficiency	0%	10%	10%	10%	10%	10%	10%	0%

- s-tagging performance versus ud-jets is extremely sensitive to K^{+-} ID
- Further gains are possible through inclusion of V^0 variables

s-tagging: low momentum



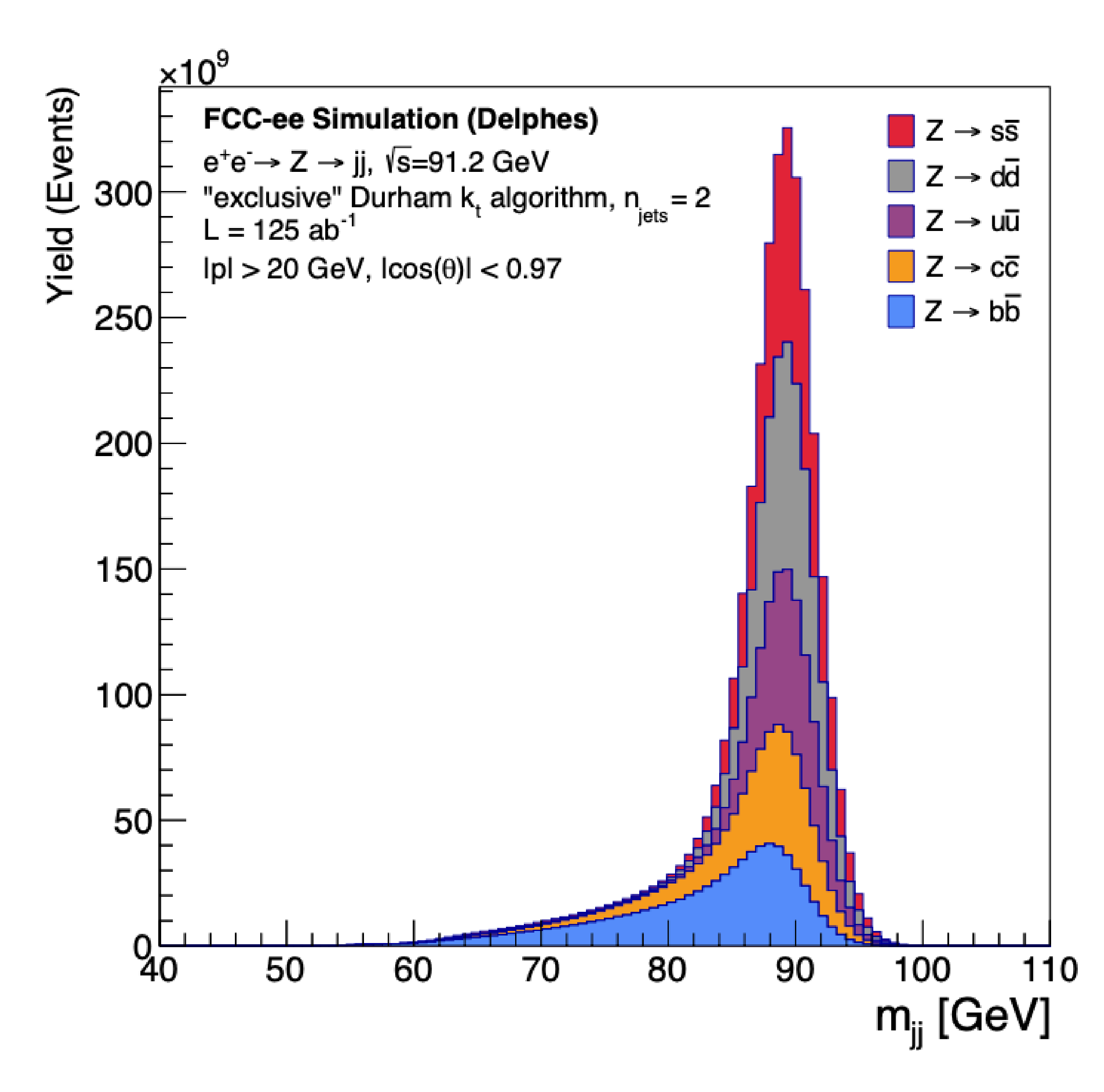


- Very low momentum strange jets have low particle multiplicities overall, where a single reconstructed V^0 becomes a distinguishing feature and can increase efficiency
    - also impactful: lower K^{+-} multiplicities lead to reduced tagging efficiency

Episode 4

What is possible at the Z peak?

Performance at the Z peak



note: no background included but was also reducible to percent level at LEP...

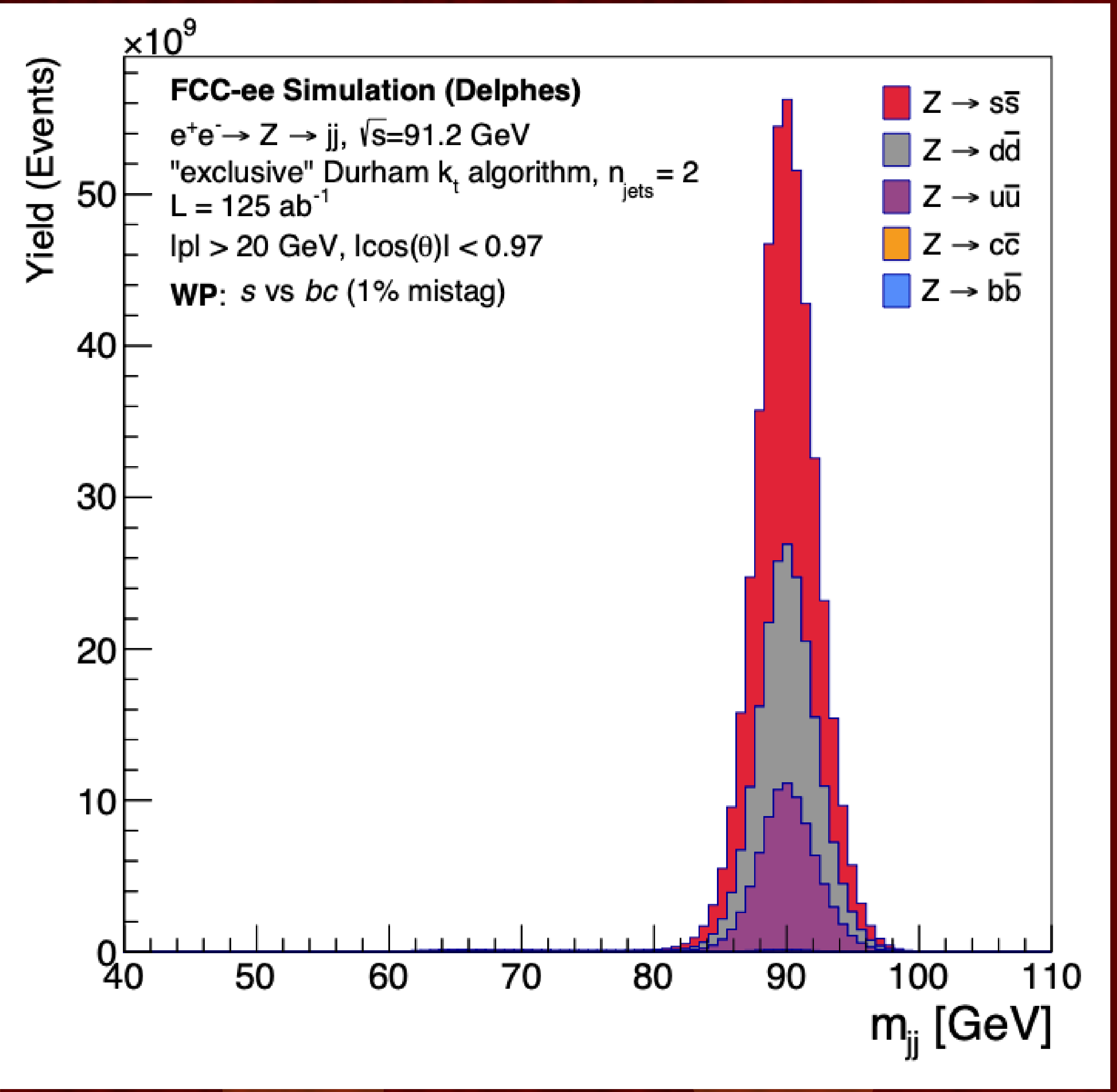
low mass tails for heavy flavour more pronounced as neutrinos in b/c jets not measured

Performance at the Z peak

Set some typical working points definitely enough events available no matter what

		Mistag Rate [%]	Efficiency [%]	N_{sig}	N_{bkg}
WP1	s vs bc	10	98.93 ± 0.03	7.35×10^{11}	1.35×10^{12}
	s vs ud	10	40.03 ± 0.04	1.45×10^{11}	3.25×10^{10}
WP2	s vs bc	1	54.18 ± 0.04	2.38×10^{11}	2.06×10^{11}
	s vs ud	10	39.28 ± 0.06	5.10×10^{10}	5.57×10^{9}
WP3	s vs bc	1	54.18 ± 0.04	2.38×10^{11}	2.06×10^{11}
	s vs ud	1	10.05 ± 0.11	1.12×10^{10}	4.77×10^{8}
WP4	s vs bc	0.1	17.96 ± 0.06	3.23×10^{10}	6.98×10^{9}
	s vs ud	0.1	1.98 ± 0.33	3.56×10^{8}	3.38×10^{6}

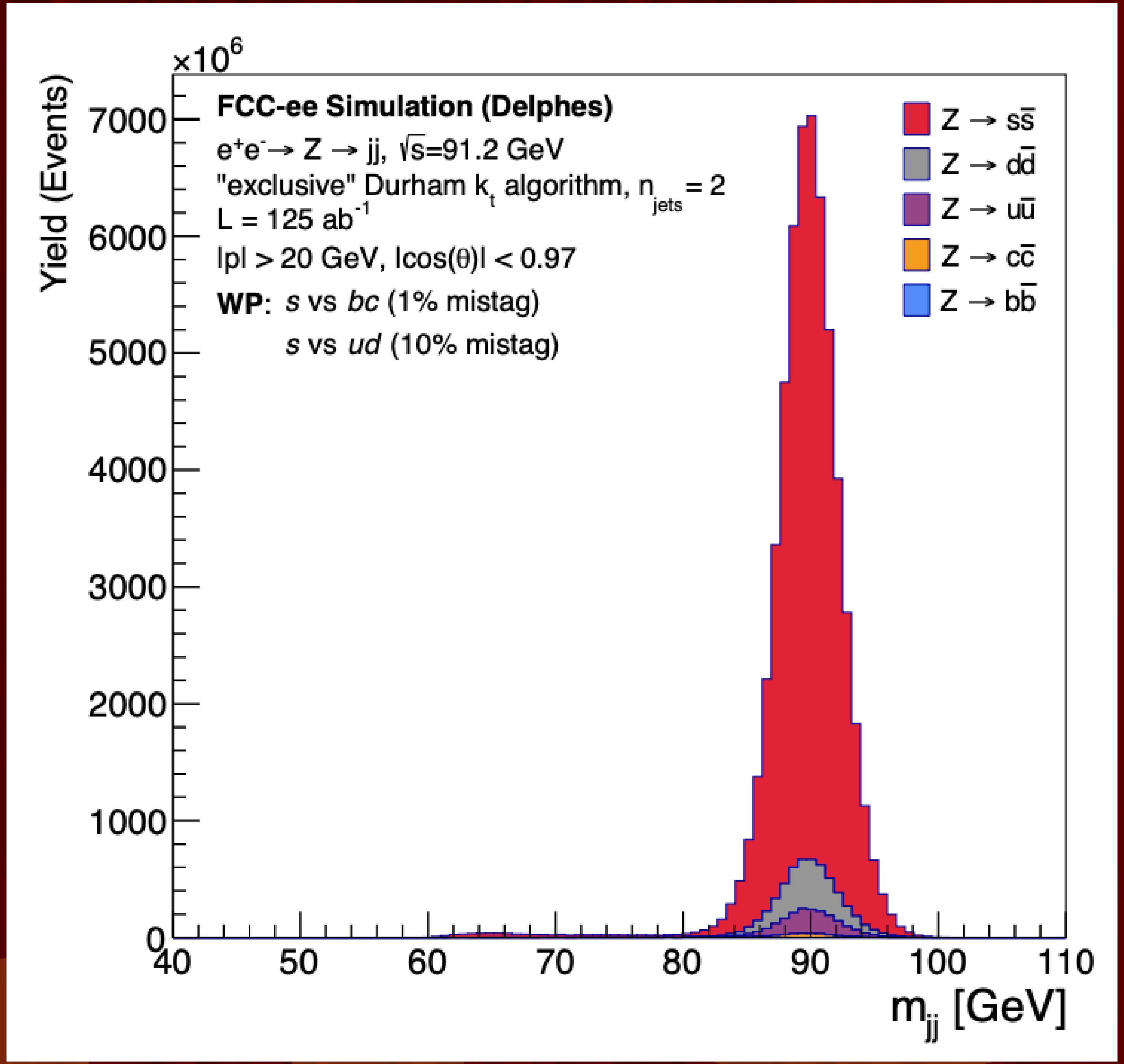
Performance at the Z peak



Only b and c anti-tagging very strongly suppresses those flavours

Expected: b,c tagging extremely efficient at separation u,d,s

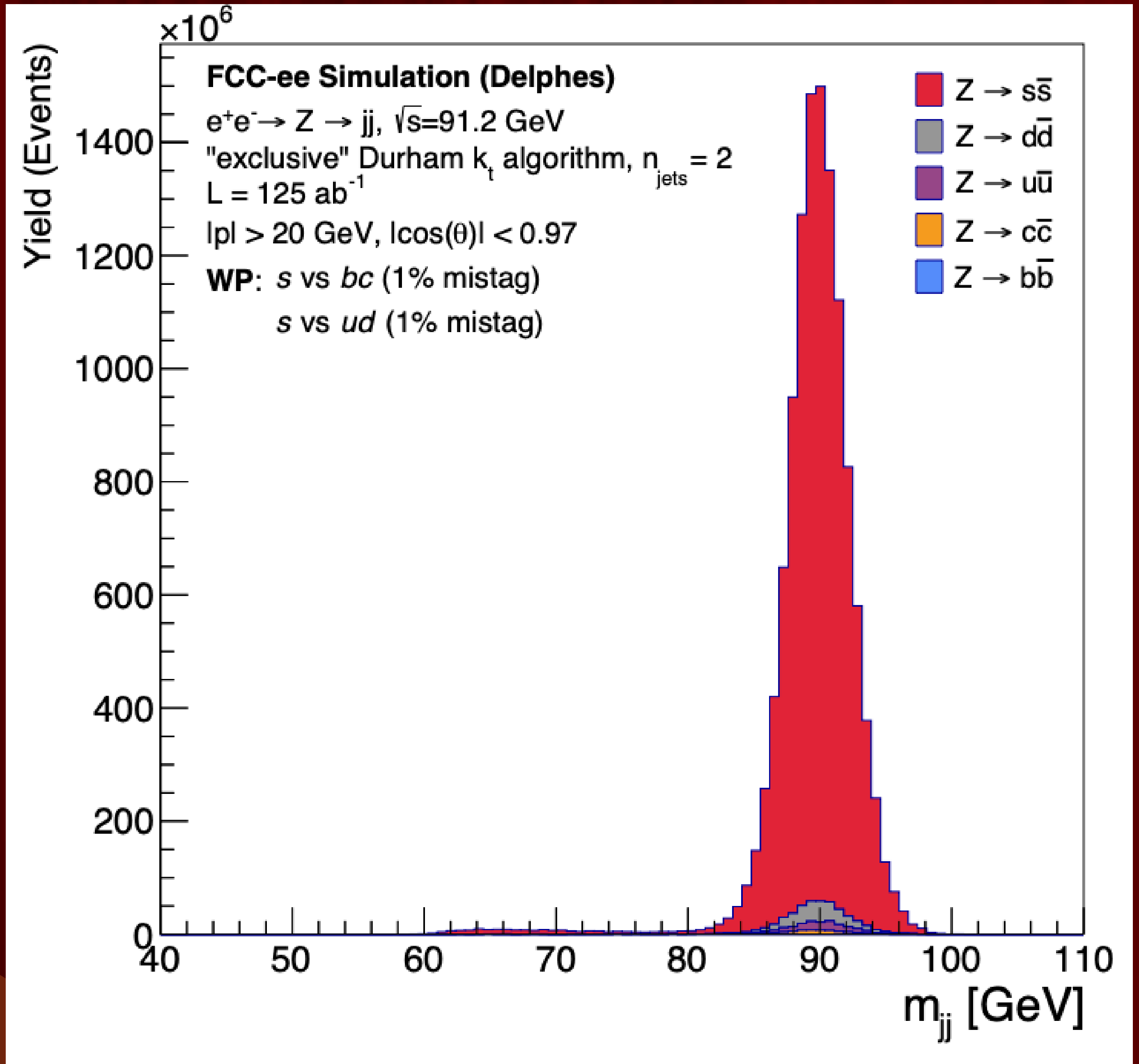
Performance at the Z peak



After b,c tagging (1% mistag) and s vs u,d tagging (loose wp, 10% mistag)

possibility to have very pure peak for physics studies

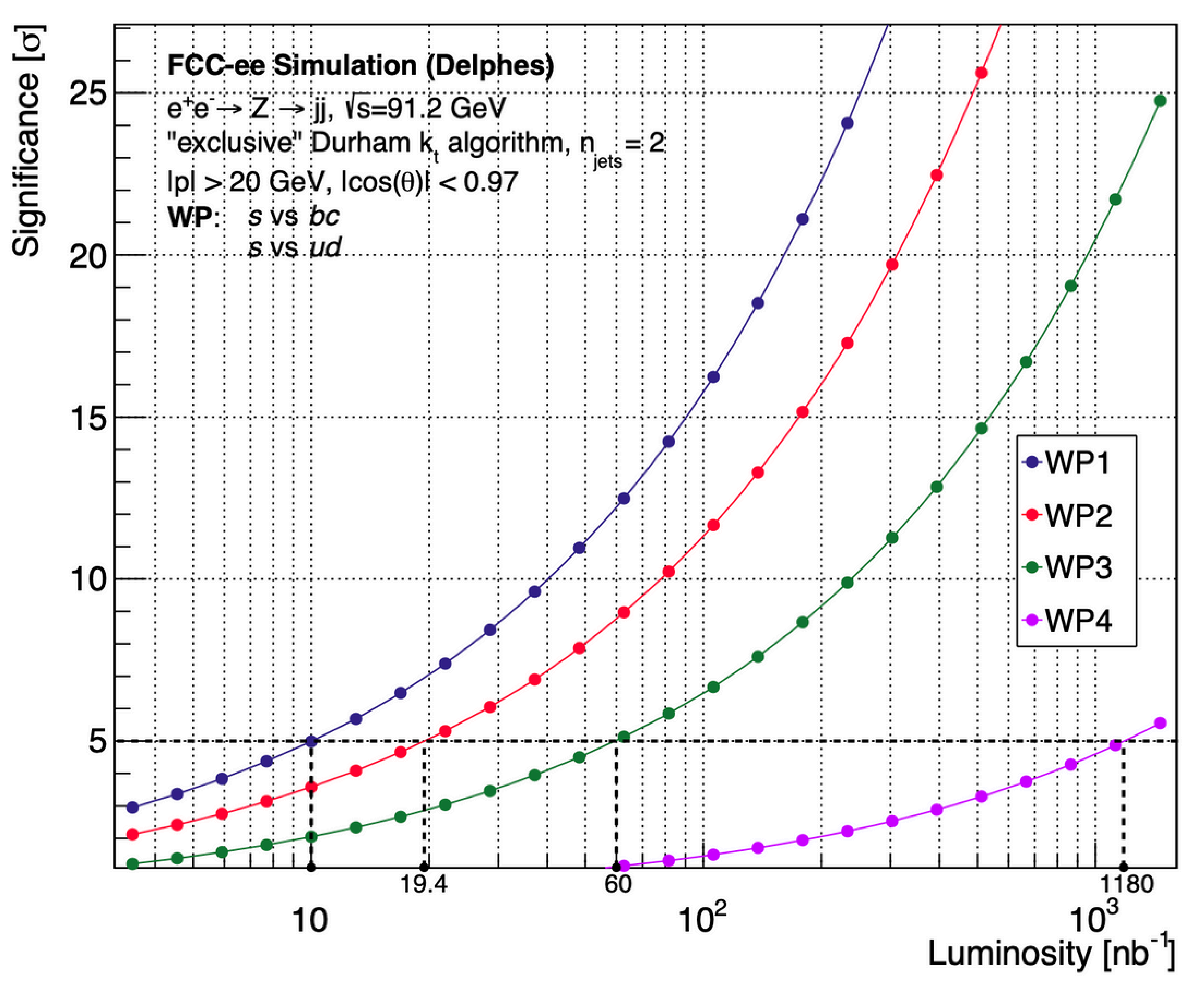
Performance at the Z peak



After b,c tagging (1% mistag) and s vs u,d tagging (1% mistag)

possibility to have very pure peak for physics studies

Performance at the Z peak



Q: How much luminosity is needed to reach canonical significance Z → ssbar ?

A: Should be easy: 5σ significance can be reached with a luminosity of 60 nb-1, equivalent to ***less than a second*** of the FCC-ee run at the Z resonance!

Conclusion

- s tagging definitely possible – and maybe even u,d separation?
  - Physics potential at Z peak beyond partial BRs?
- DeepJetTransformer is fast to train
  - Good for detector performance studies
  - Comparable performance to other tools
  - Still some room for improvement (simplified PID, DELPHES does not have fakes, older IDEA scenario)
- to go beyond b,c tagging: V0s matter, K+-/pi+- separation matters
- reference: arXiv:2406.08590
  - also follow full open science policies, so code already available, all performance curves will be to at publication
  - https://github.com/Edler1/DeepJetFCC/tree/master/docs

THANK YOU

Special thanks to F. Gaede, L. Gouskos, M. Selvaggi for their pre-submission feedback

reference: arXiv:2406.08590 (under EPJC review, new version very soon)