Jet reconstruction at future e^+e^- colliders

Rosa Simoniello (CERN)

On behalf of the CLICdp collaboration

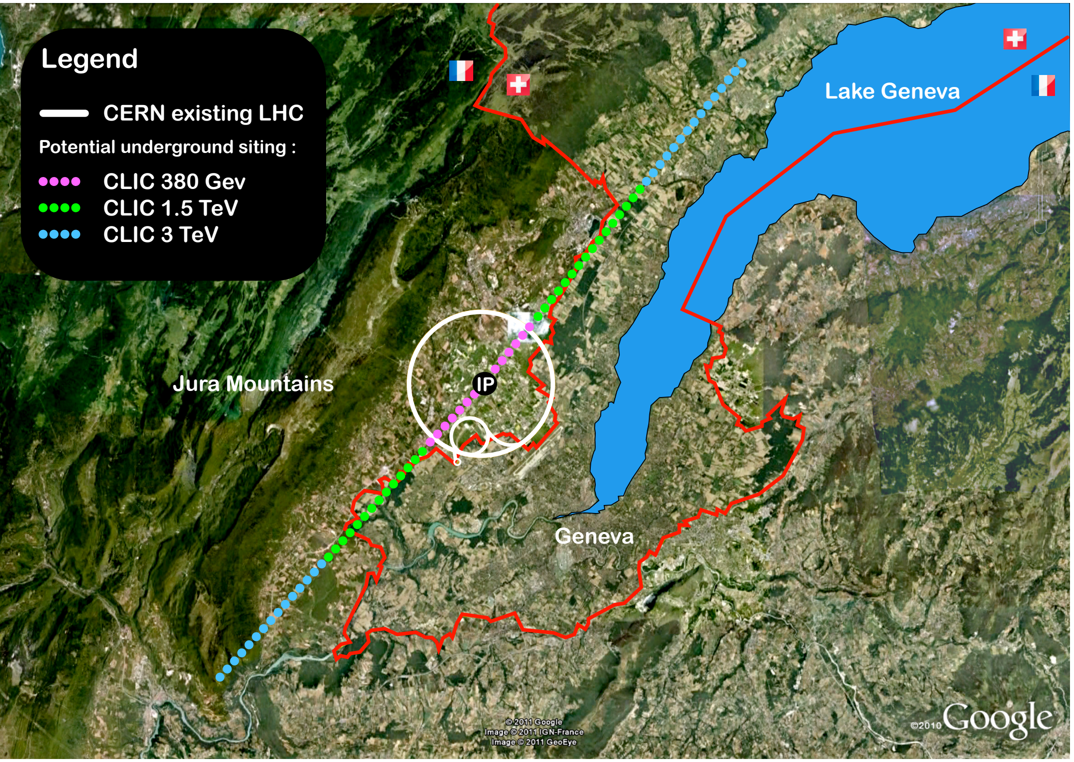
BOOST 2016, Zurich, 18-22 July

Outline

*Observations on jet reconstruction in this talk are valid for a general lepton collider. Most challenging beam background environment for high energies → **CLIC at 3 TeV** (info about other projects in backup)*

- Introduction to CLIC
- Physics case for a linear high energy collider (CLIC)
- CLIC experimental conditions
  - ❑ Beam background
- Calorimeters R&D
  - ❑ High granularity
- Jet clustering at lepton colliders
- Plans for jet substructure techniques
- Conclusion

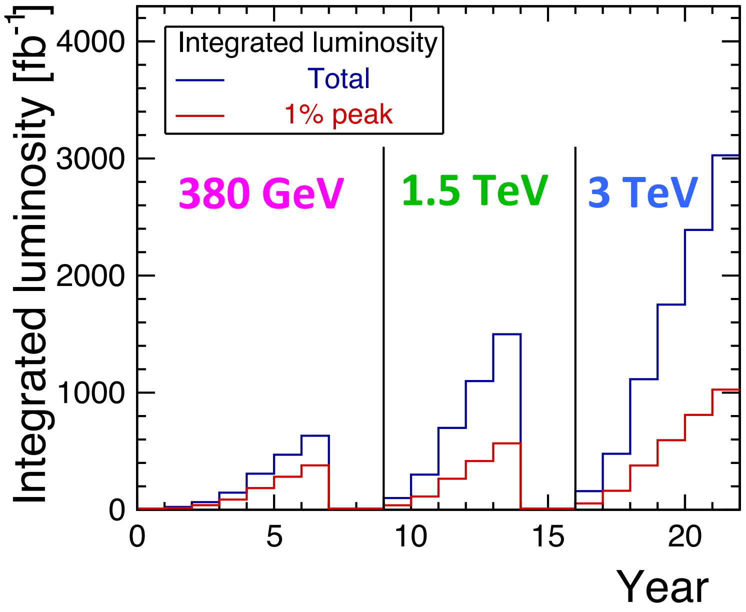
CLIC in a nutshell



- 2-beam acceleration scheme
- Gradient 100 MV/m at room temperature
- Proof of concept in CTF3
- Built in energy stages: 380 GeV – 3 TeV
- Physics program over 20 years: *H*, *t*, BSM

"Updated baseline for a staged Compact Linear Collider" → soon available as CERN Yellow Report

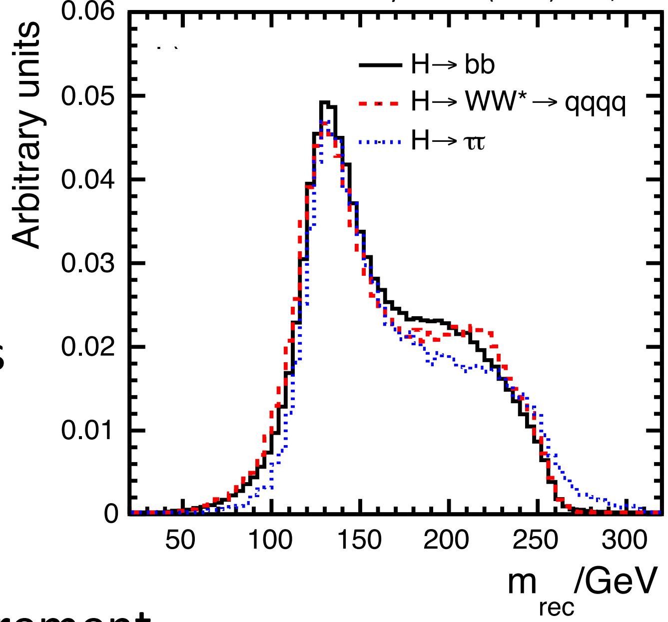
Stage	$\sqrt{s}$ (GeV)	$\mathscr{L}_{\text{int}}$ (fb^{-1})
1	380	500
	350	100
2	1500	1500
3	3000	3000

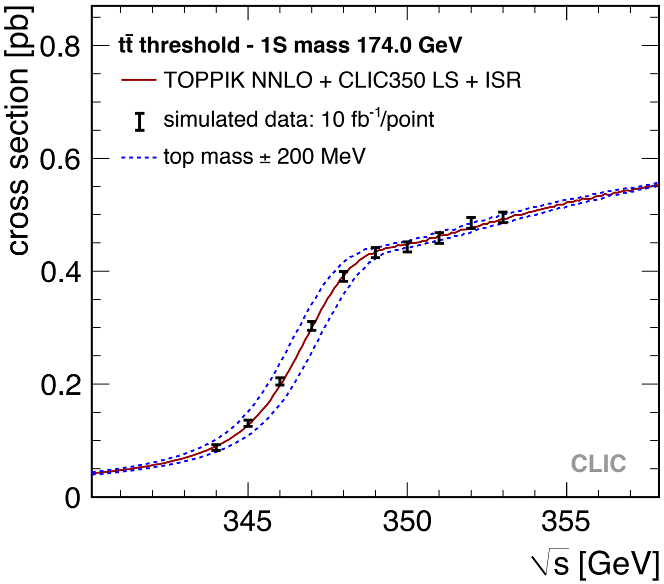


Physics program

- High precision measurement of Higgs properties
  - Couplings: mostly sub-% level (% level for rare decays)
  - Precision on Higgs width: 3.4% (24 MeV)
- Model independent couplings determination
  - Mass recoil method in *ee->Z(μμ/ee/qq)H* events
- *"Higgs Physics at the CLIC Electron-Positron Linear Collider"* → *to be published soon*

Eur.Phys.J. C76 (2016) no.2, 72

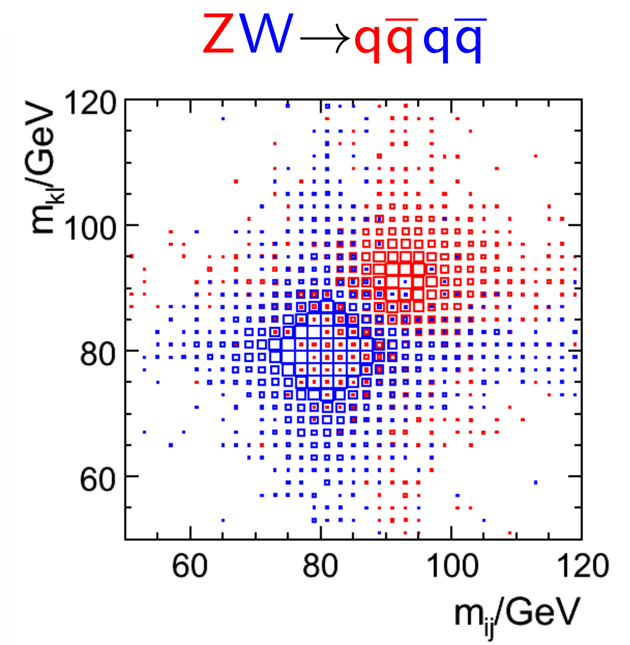


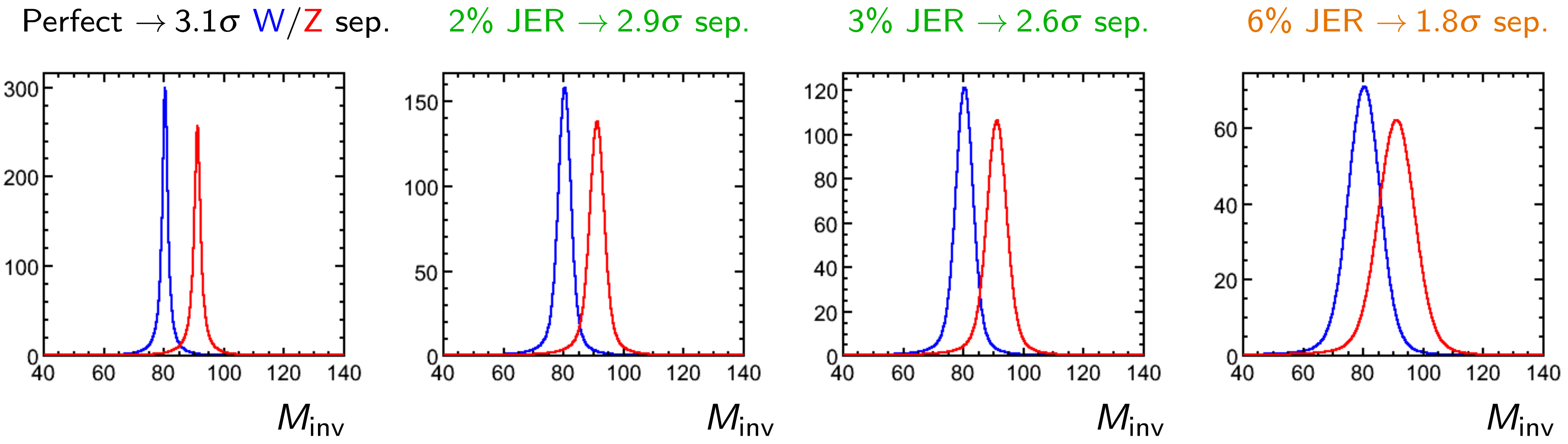


- Top mass measurement
  - Precision for 1S mass scheme: 50 MeV
- Top quark form factors looking at A_{fb}
  - Sub-percent level (1 order of magnitude better than HL-LHC projection)
- Direct and indirect BSM searches (including top)
  - Dedicated program at 1.5 TeV and 3 TeV
  - High multiplicity final states

Requirement on jet energy resolution

- Excellent jet performance is crucial in the full detector
    - At high energies → forward physics objects
- Jet energy resolution important to separate final state decay products
- In this example: hadronic W and Z decays
- **3%–5% jet energy resolution** gives ~2.6σ – 2.1σ W/Z separation


$ZW \to q\bar{q}q\bar{q}$


Perfect → 3.1σ W/Z sep.

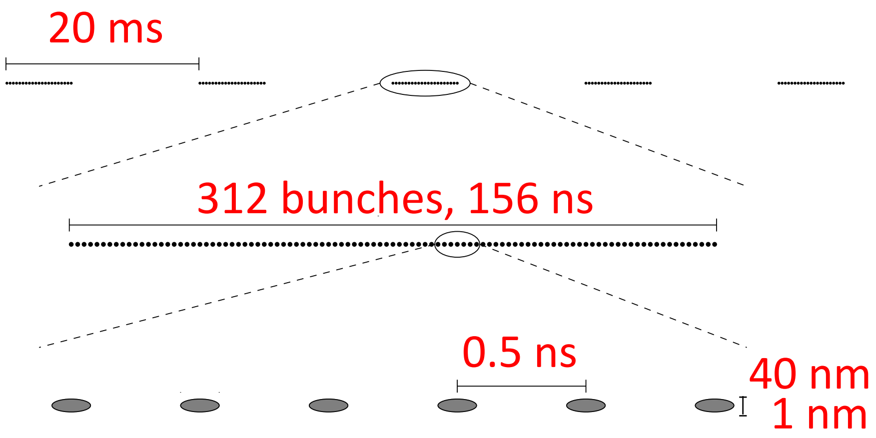
2% JER → 2.9σ sep.

3% JER → 2.6σ sep.

6% JER → 1.8σ sep.

CLIC machine environment at 3 TeV

Total luminosity: $5.9\times10^{34}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}$



- Small beams at IP lead very high E field

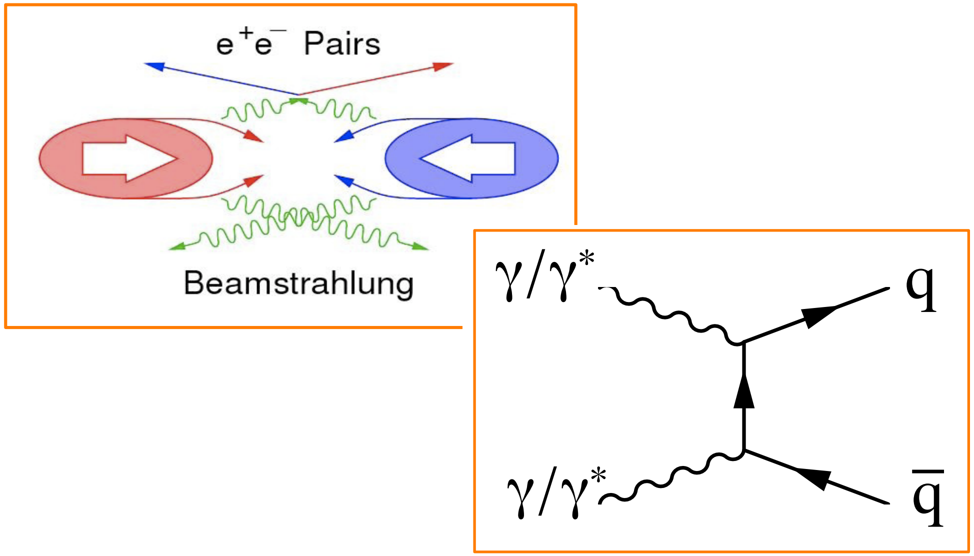
→ Beamstrahlung

- ❑ Pair production and γγ->hadrons

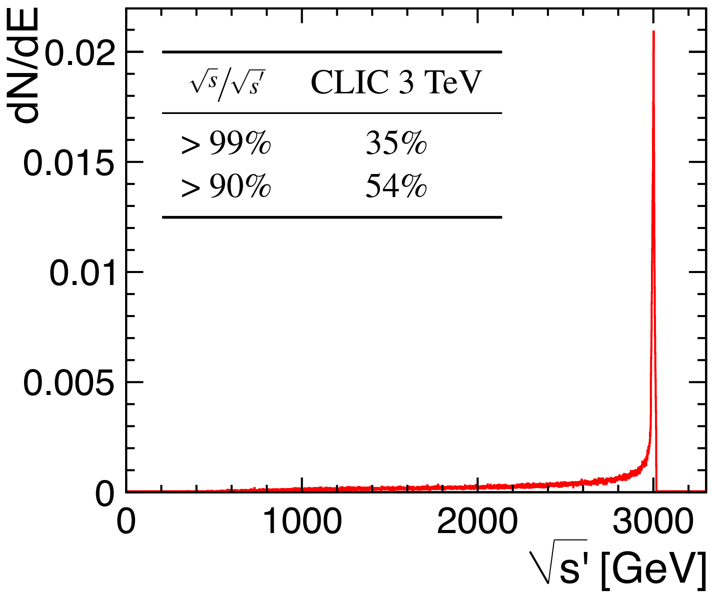
  ~ 19 TeV energy per 156 ns train

- ❑ Luminosity spectrum → System can be boosted along beam axis

→ Jet finding algorithms from LEP no longer appropriate



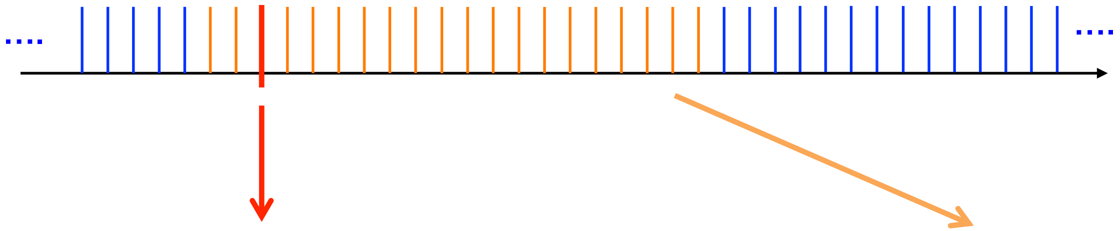
CERN–2012–003



$\sqrt{s}/\sqrt{s'}$	CLIC 3 TeV
> 99%	35%
> 90%	54%

Background rejection

- Triggerless readout



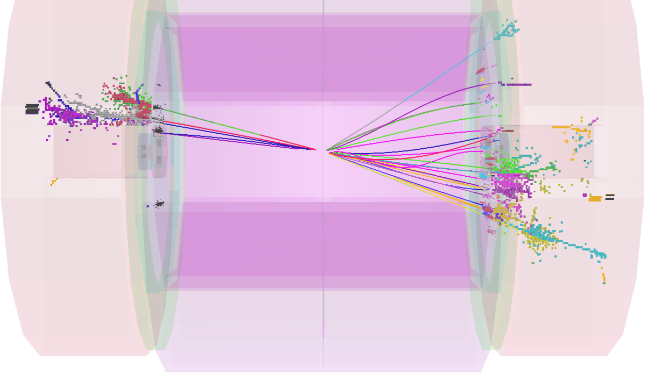
- Find the t_0 physics event offline and pass a window around t_0 to the reconstruction
  - Compromise between calorimeter integration time and bkg minimisation → 10 ns (100 ns for W)

  → *Energy in calorimeters: from ~19 TeV to ~1.2 TeV*

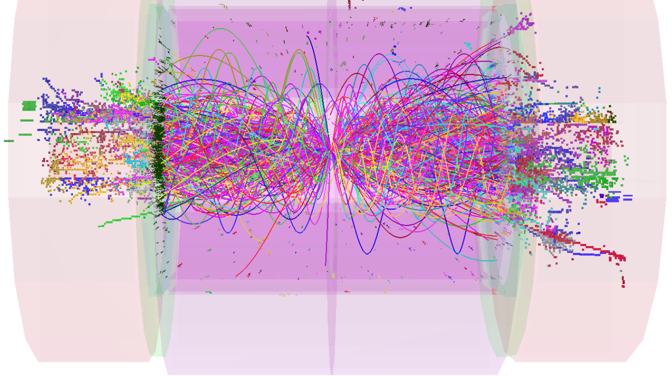
- Apply p_T and timing cuts on PFOs (loose, default and tight selections available)
  - Calorimeter time stamp resolution: 1 ns
  - Time corrected for shower development and TOF
  - Cuts depend on particle type, p_T and θ
    - Allow to protect high p_T object

  → *Energy in calorimeters: from ~1.2 TeV to ~100 GeV*

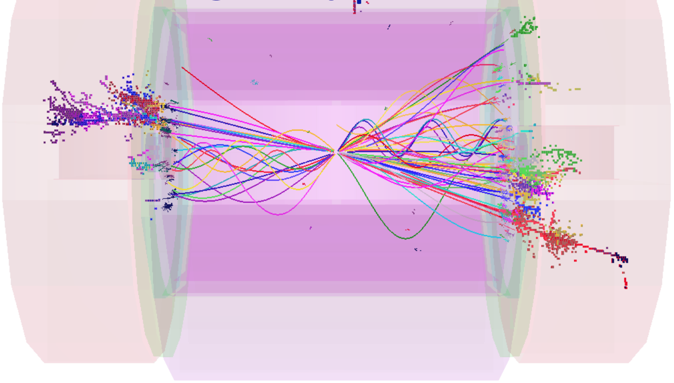
Forward WW, no background



After time window: ~1.2 TeV

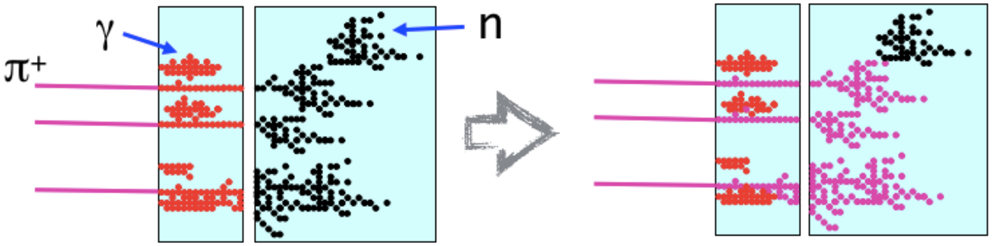


After timing and p_T cuts: ~100 GeV



Imaging calorimeters

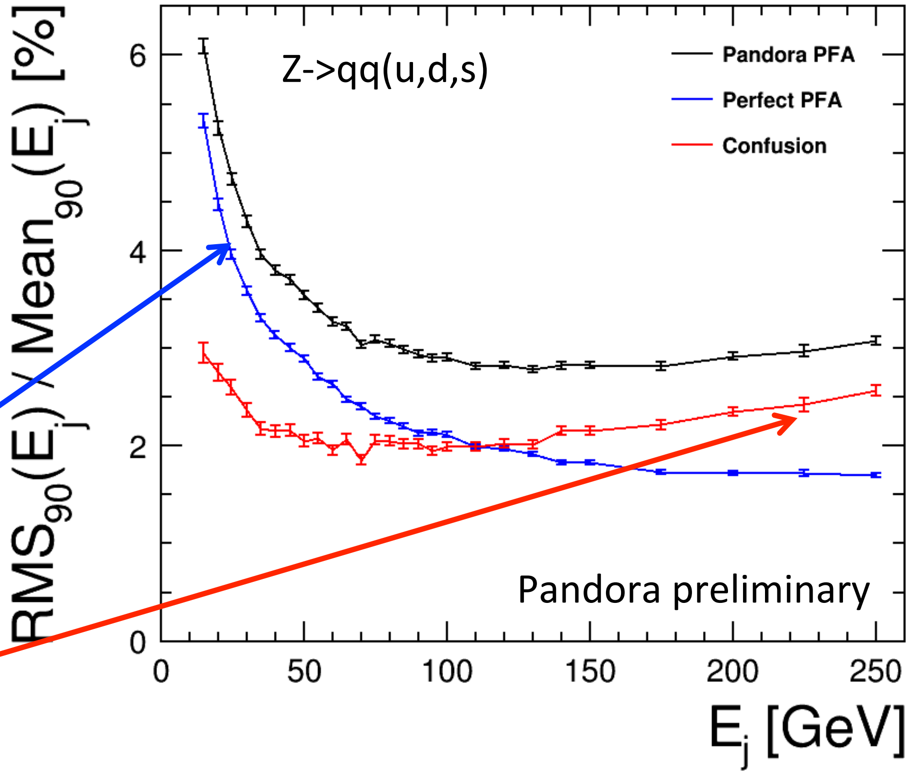
- Achieve JER 3%–5% and cope with bkg occupancy
- Particle flow (PFA): always use best info available
- High granularity calorimeters designed for PFA: resolve energy deposits from different particles
  - W-Si ECAL, 5×5mm^2 cell size $\Delta\eta\times\Delta\phi = 0.003\times0.003$
  - Fe-Sc HCAL, 30×30mm^2 cell size $\Delta\eta\times\Delta\phi = 0.0015\times0.0015$
- Software (PandoraPFA, *EPJC.75.439*): identify energy deposits from each particle



$E_{JET} = E_{ECAL} + E_{HCAL}$

$E_{JET} = E_{TRACK} + E_{\gamma} + E_{n}$

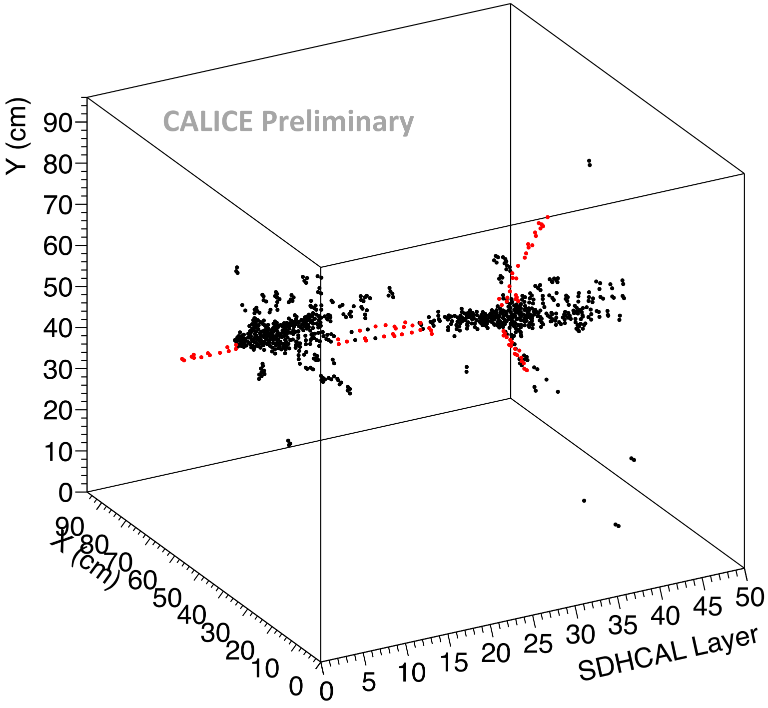
jet composition 60% 30% 10%

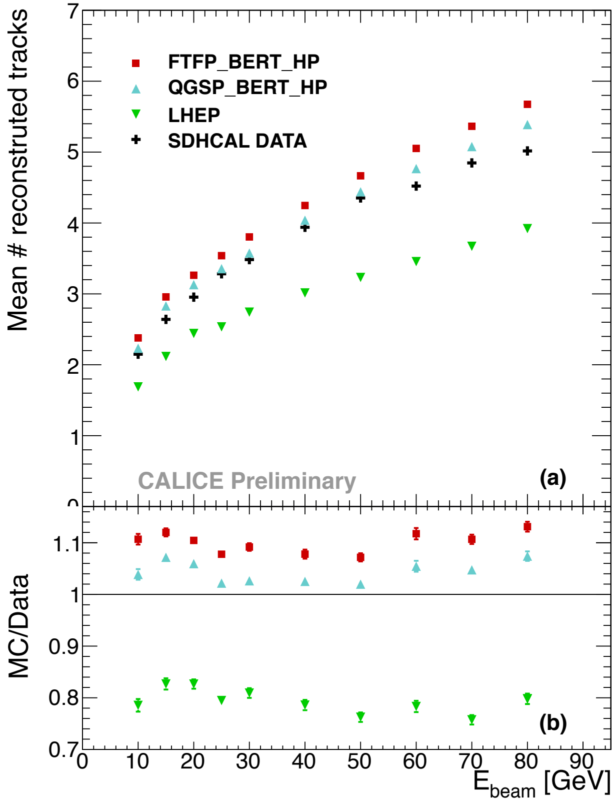


$$\sigma_{\rm jet} = \sqrt{\boxed{\sigma^2_{\rm trk} + \sigma^2_{\rm ECAL} + \sigma^2_{\rm HCAL}} + \boxed{\sigma^2_{\rm conf}}}$$

Tracking within hadronic showers

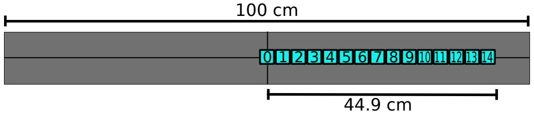
- Fe-SDHCAL prototype, 1x1cm^2, pions 10-80 GeV, *CaliceAnalysisNote-047*
- Study of substructure of hadronic shower → tracks from charged particles
  - Understand the calo response and estimate optimally the hadronic energy
  - Connecting clusters produced by hadronic interaction of secondary charged particles to the main one → it helps PFA
- Best agreement between data and **QGSP_BERT** (OK also **FTFP_BERT**)

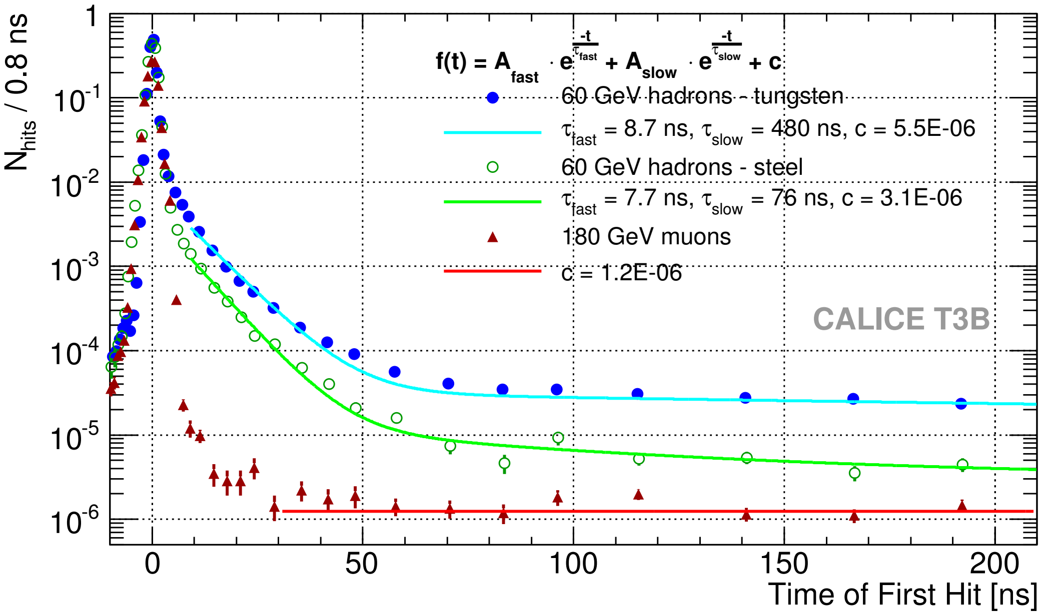


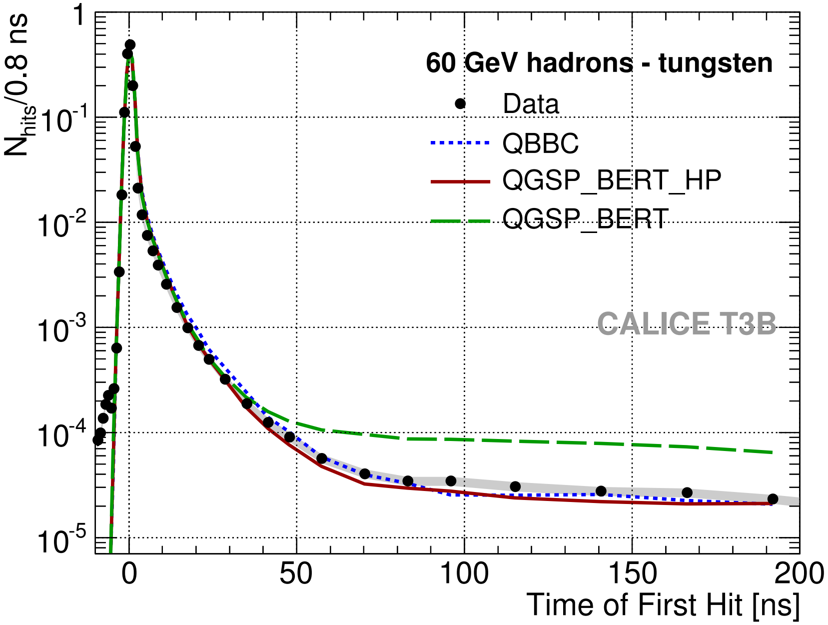


Time structure of the hadronic shower

- Time structure of hadronic showers affects timing capability and integration time of calo
- Scintillator cells placed behind HCAL
- Fe and W HCAL tested, *arXiv:1404.6454*
- Slower response for W
- For W, High Precision (HP) neutron needed → neutron tracking down to thermal energies







Jet reconstruction at lepton colliders

Durham or e^+e^- k_t algorithm
(LEP and SLC)

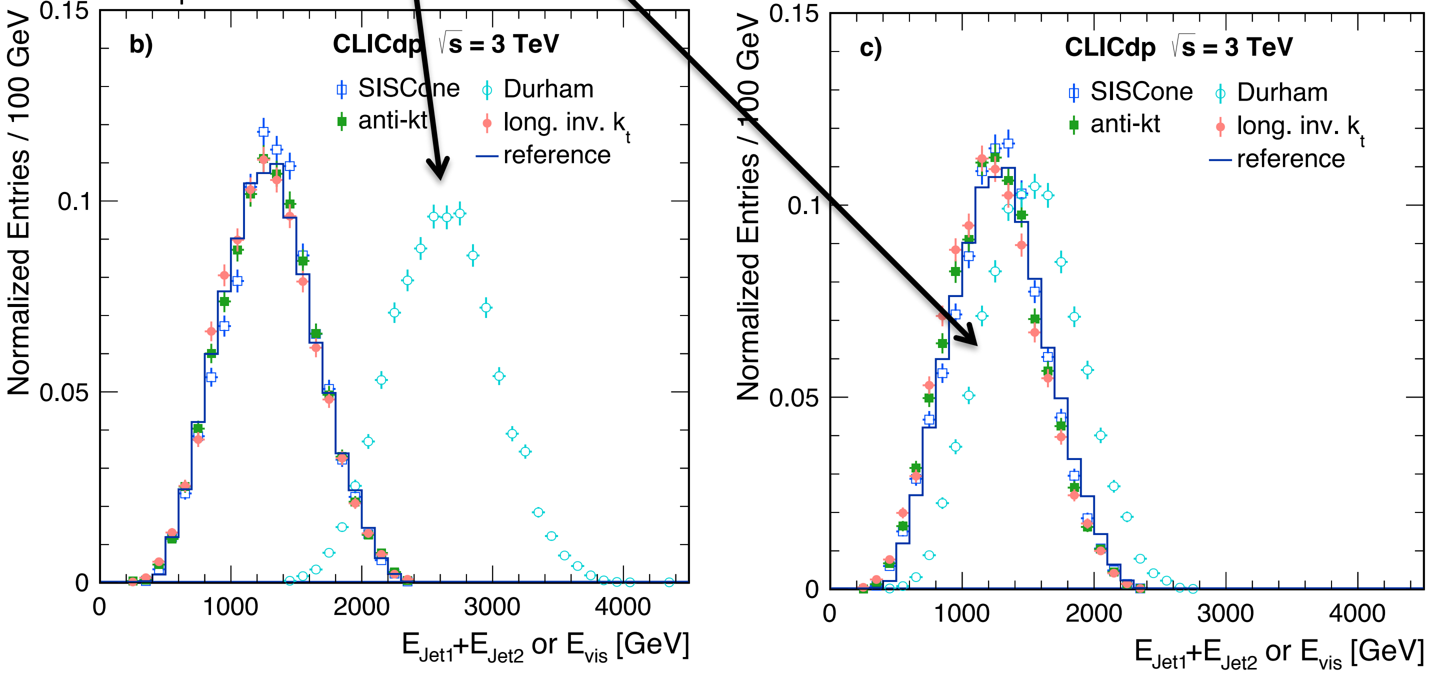
$$d_{ij} = 2min(E_i^2, E_j^2)(1 - \cos\theta_{ij})$$

$$d_{ij} = 2min(p_{Ti}^{2n}, p_{Tj}^{2n})\Delta R_{ij}^2/R^2$$
$$d_{iB} = p_{Ti}^{2n}$$

n=0: Cambridge-Aachen
n=1: Longitudinally invariant k_t
n=-1: Anti-k_t (LHC default)

After time window

After time and p_T cuts



CLICdp-Pub-2015-002

→ Cluster all particles not adequate
→ Level of background still significant even after tight cuts

Definitions of algorithms

Durham or e^+e^- k_t algorithm
(LEP and SLC)

$$d_{ij} = 2min(E_i^2, E_j^2)(1 - \cos\theta_{ij})$$

$$d_{ij} = 2min(p_{Ti}^{2n}, p_{Tj}^{2n})\Delta R_{ij}^2/R^2$$
$$d_{iB} = p_{Ti}^{2n}$$

n=0: Cambridge-Aachen
n=1: Longitudinally invariant k_t
n=-1: Anti-k_t (LHC default)

- Time to re-think the lepton collider jet algorithms
- Can we import what we learnt from hadron colliders experience in an algorithm optimised for lepton collider needs?

Definitions of algorithms

Durham or e$^+$e$^-$ k$_t$ algorithm
(LEP and SLC)

$$d_{ij} = 2min(E_i^2, E_j^2)(1 - \cos\theta_{ij})$$

$$d_{ij} = 2min(p_{Ti}^{2n}, p_{Tj}^{2n})\Delta R_{ij}^2/R^2$$
$$d_{iB} = p_{Ti}^{2n}$$

n=0: Cambridge-Aachen
n=1: Longitudinally invariant k$_t$
n=-1: Anti-k$_t$ (LHC default)

Simply add a beam distance concept
→ *not enough (see next results)*

Durham inspired distance
Resilience against bkg from long inv k$_t$

Generalised e$^+$e$^-$ k$_t$ algorithm

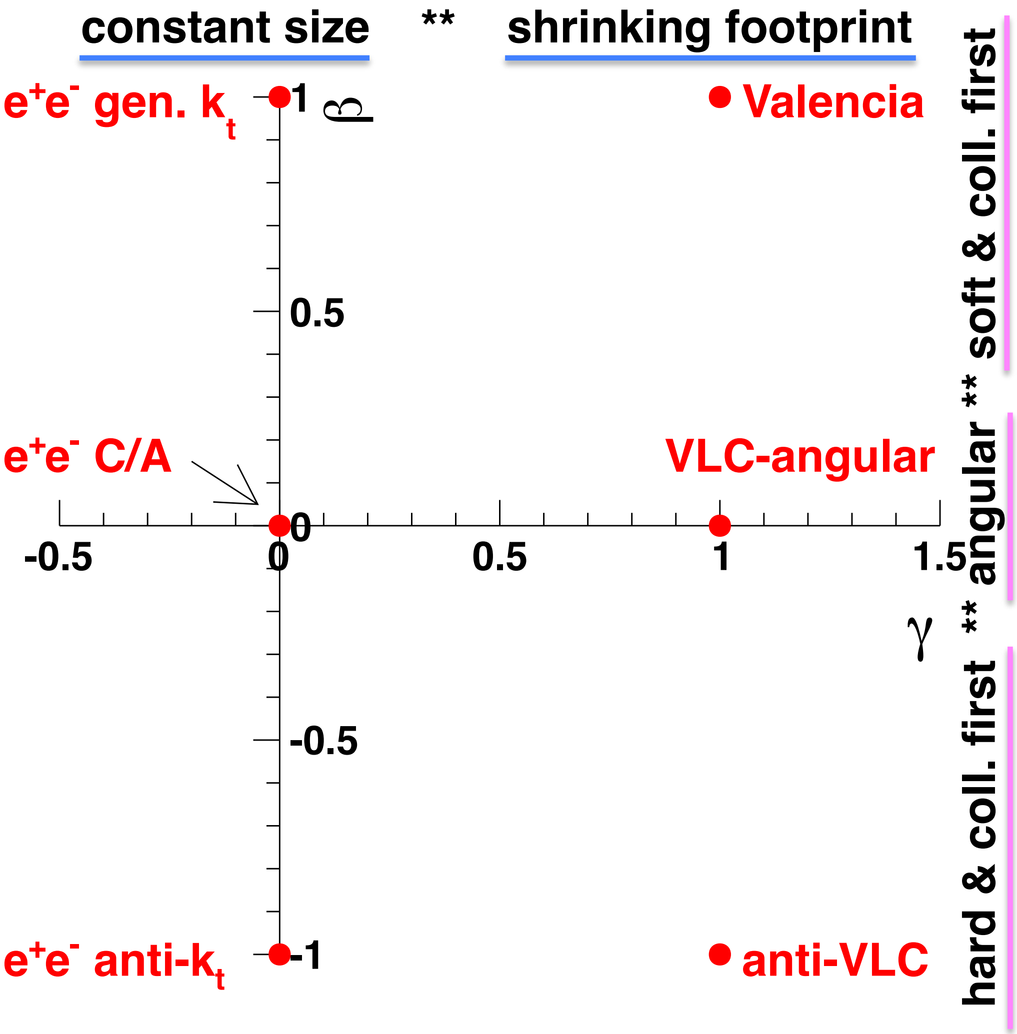
$$d_{ij} = 2min(E_i^2, E_j^2)(1 - \cos\theta_{ij})/(1 - \cos R)$$
$$d_{iB} = E_i^2$$

VLC algorithm

$$d_{ij} = 2min(E_i^{2\beta}, E_j^{2\beta})(1 - \cos\theta_{ij})/R^2$$
$$d_{iB} = E_i^{2\beta}\sin^{2\gamma}\theta_{iB}$$

arXiv:1607.05039 and arXiv:1404.4294

VLC parameter space (β,γ)



Parameters can assume real values – smooth transition

- *γ gives the evolution of the jet area with polar angle*
- *β allows to change the clustering order*

$$d_{ij} = 2min(E_i^{2\beta}, E_j^{2\beta})(1-\cos\theta_{ij})/R^2$$

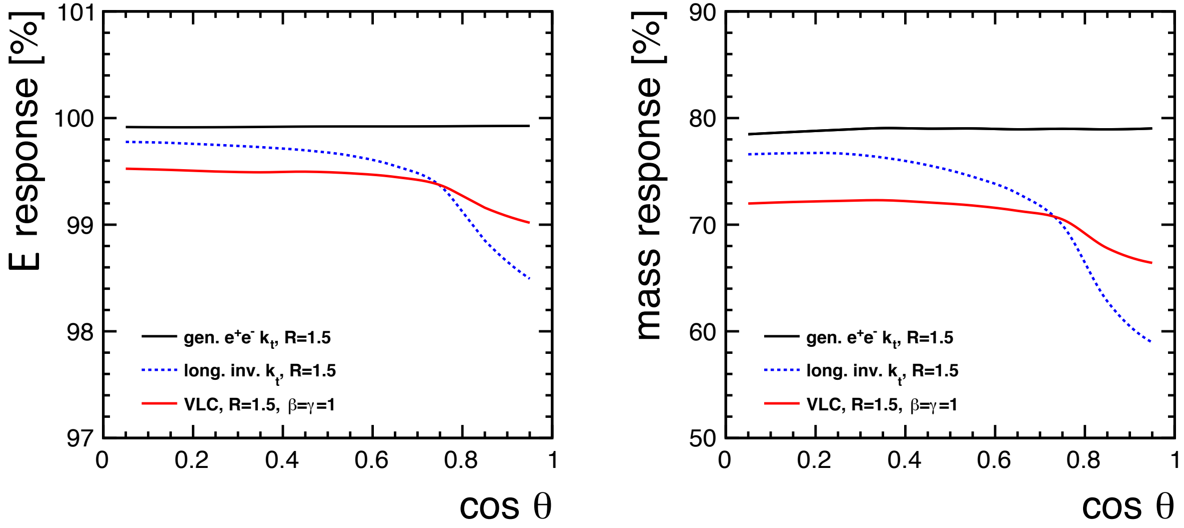
$$d_{iB} = E_i^{2\beta}\sin^{2\gamma}\theta_{iB}$$

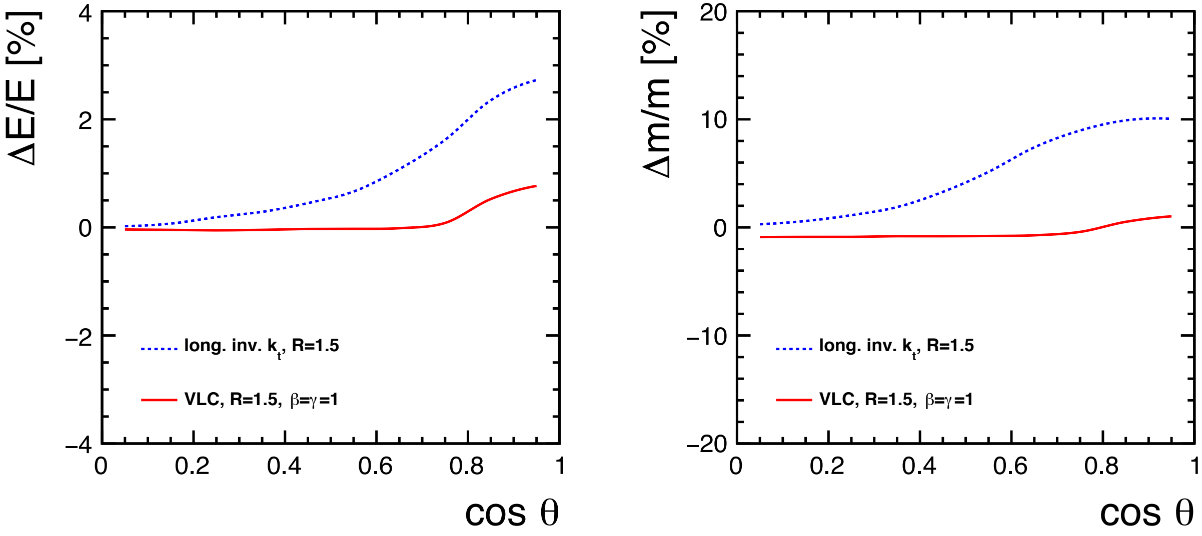
$$\text{for } \beta = \gamma = 1 \ \rightarrow d_{iB} = p_{T,i}^2$$

At lepton colliders exclusive clustering is usually used

Jet response w/wo bkg

- Jet energy and mass response to 1.5 TeV top jets *without bkg*
- cosθ>0.6: lower response for long inv kt and VLC due to the shrinking jet area

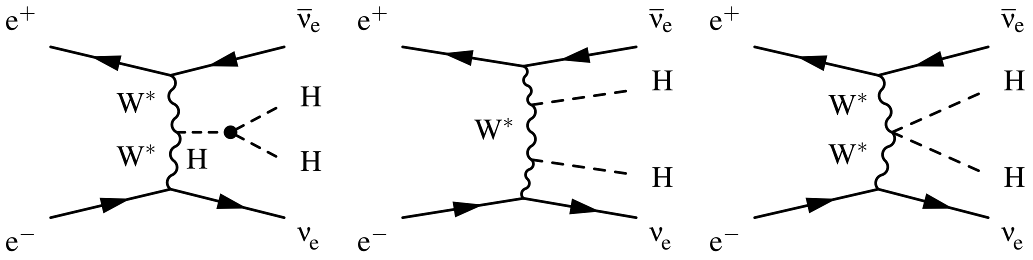


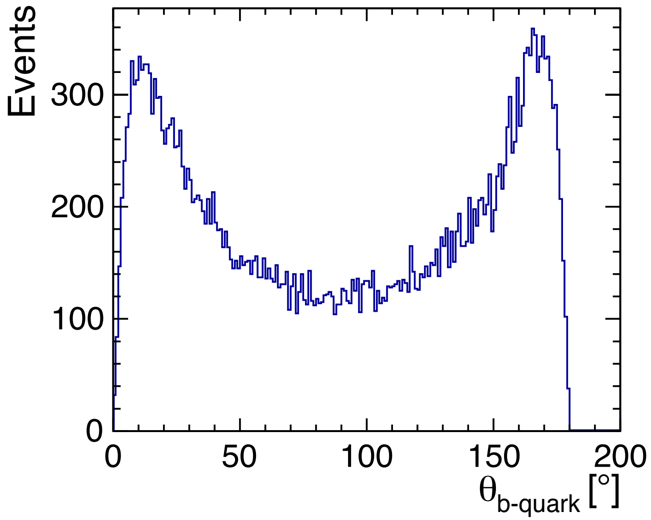


- Effect of bkg on jet energy and mass clustering:
  ΔE of same events w/wo bkg
- *"Toy bkg"*: 200 particels of 1 GeV overlaid on signal
- VLC more robust

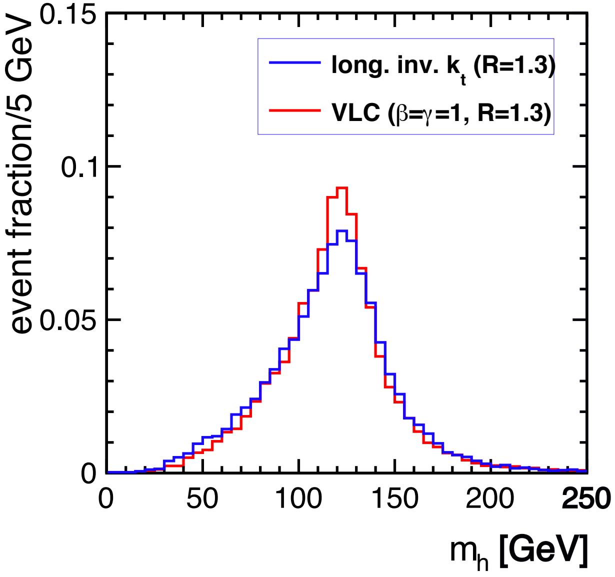
Realistic bkg case in the next slides

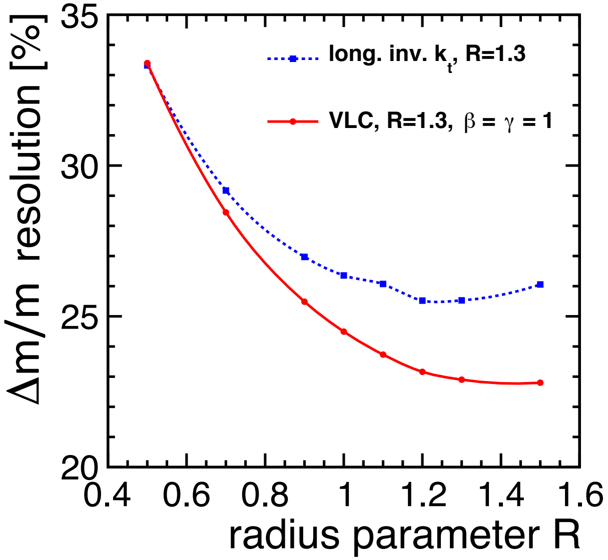
Performance in HH events

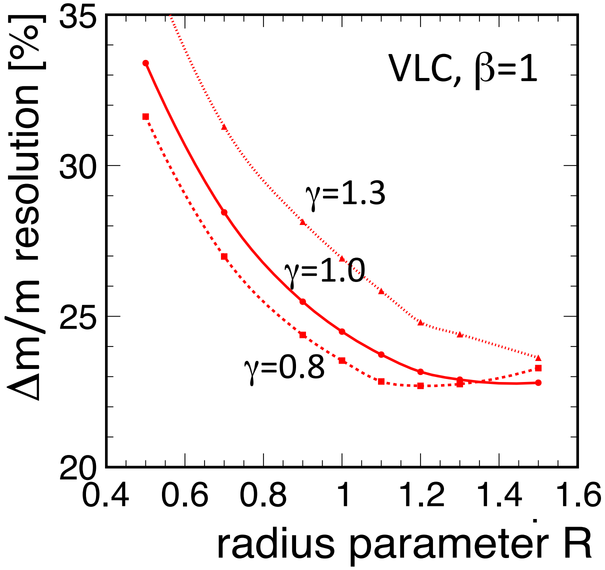




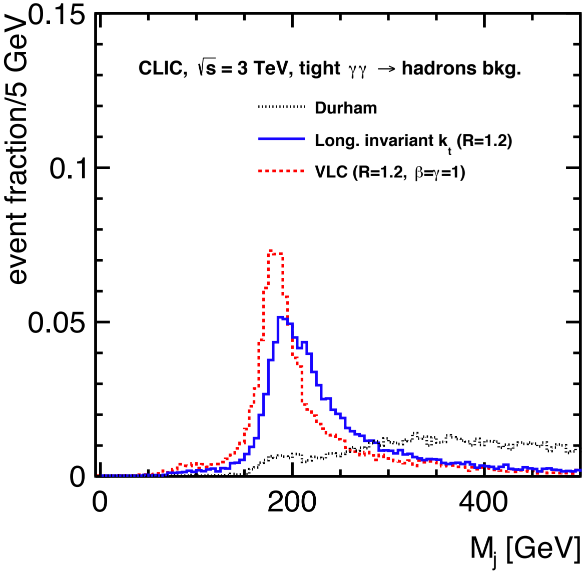
- *HHνν → bbbbνν* at 3 TeV, full simulation
- Very forward topology: challenging with bkg
- Exclusive 4 jets clustering
- Mass resolution crucial to separate signal from other SM processes

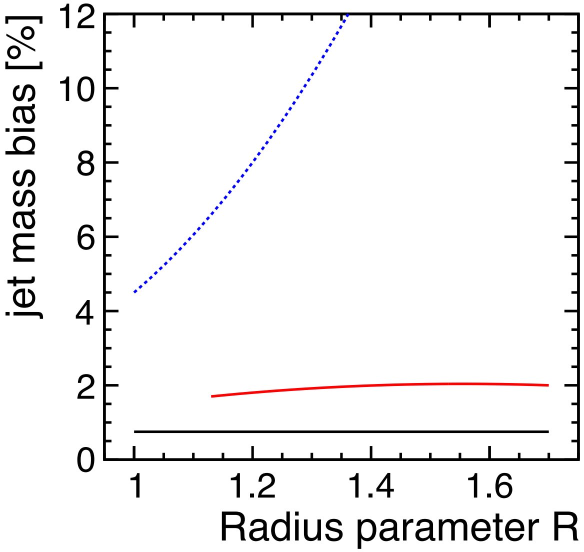


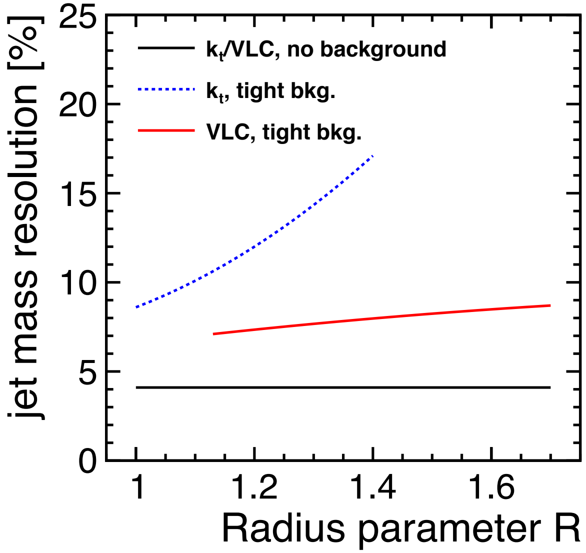




Performance in hadronic tt events

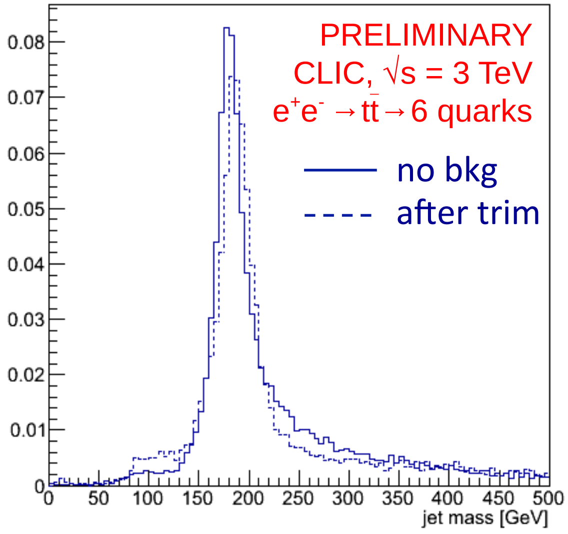






- *ee→tt→bbqqqq* at 3 TeV, full simulation
- Boosted top: decay products in 2 jets of R=1.2
- γγ→hadrons bkg have large effect on mass resolution
- *Trimming restore jet mass resolution to ~4%:*
  - 3+3 Valencia trimming with R=0.2 subjets
- Just started systematic studies to optimise:
  - Jet algorithm, R parameter and threshold
  - Look at *Johns Hopkins top tagger*

Valencia trimming



Conclusions

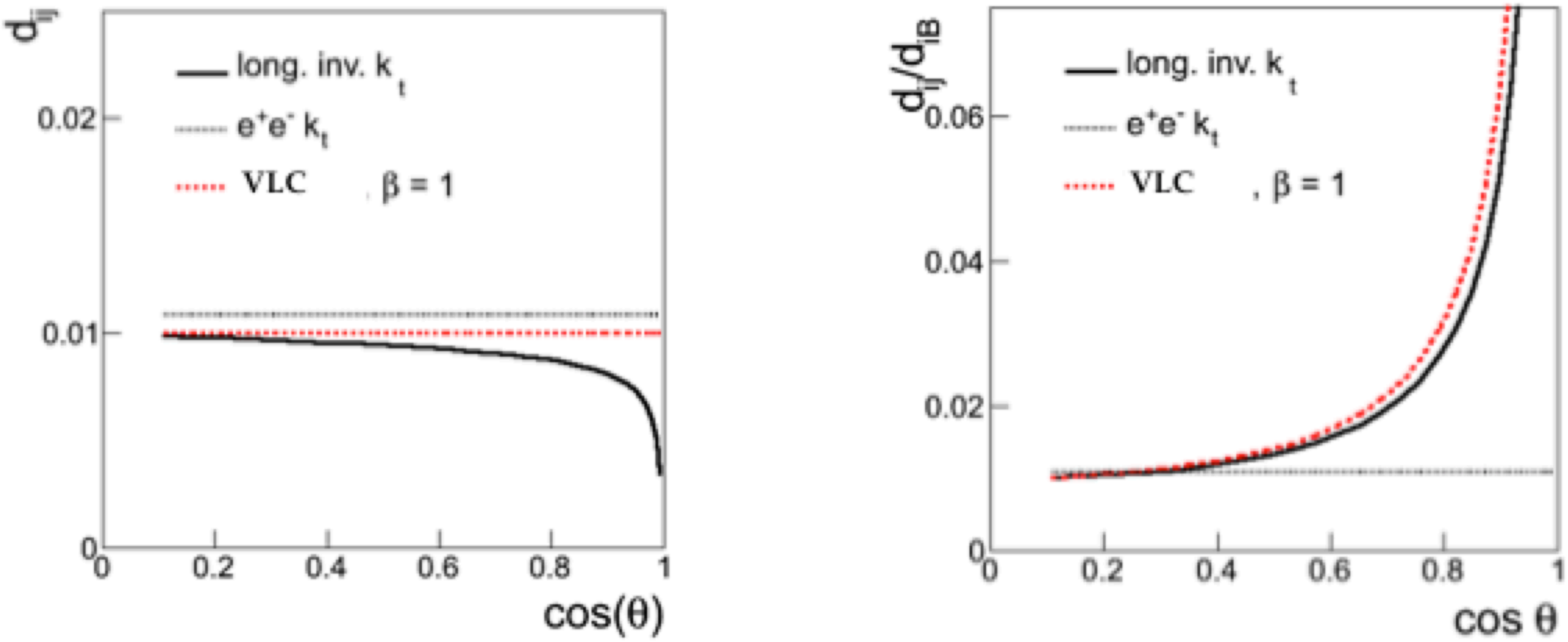
- A lepton collider is one of the options for post-LHC
- CLIC studies are well under way → European strategy in 2019-2020
  http://clicdp.web.cern.ch/
- Uniqueness of CLIC is the possibility to go to high centre-of-mass energies → √s = 3 TeV
  - Relative high level of beam induced background
- Very granular calorimeters with imagining capabilities designed for particle flow
- Rethink the jet reconstruction algorithm for lepton colliders
  - In same analyses, VLC shows some improvements w.r.t. inv long kt
    http://arxiv.org/abs/1607.05039 http://arxiv.org/pdf/1404.4294.pdf
  - Is it the way to go? Other ideas?
- High granular calorimeters are promising for jet substructure techniques → potential at CLIC to be fully explored

BACK-UP

Options for future lepton colliders

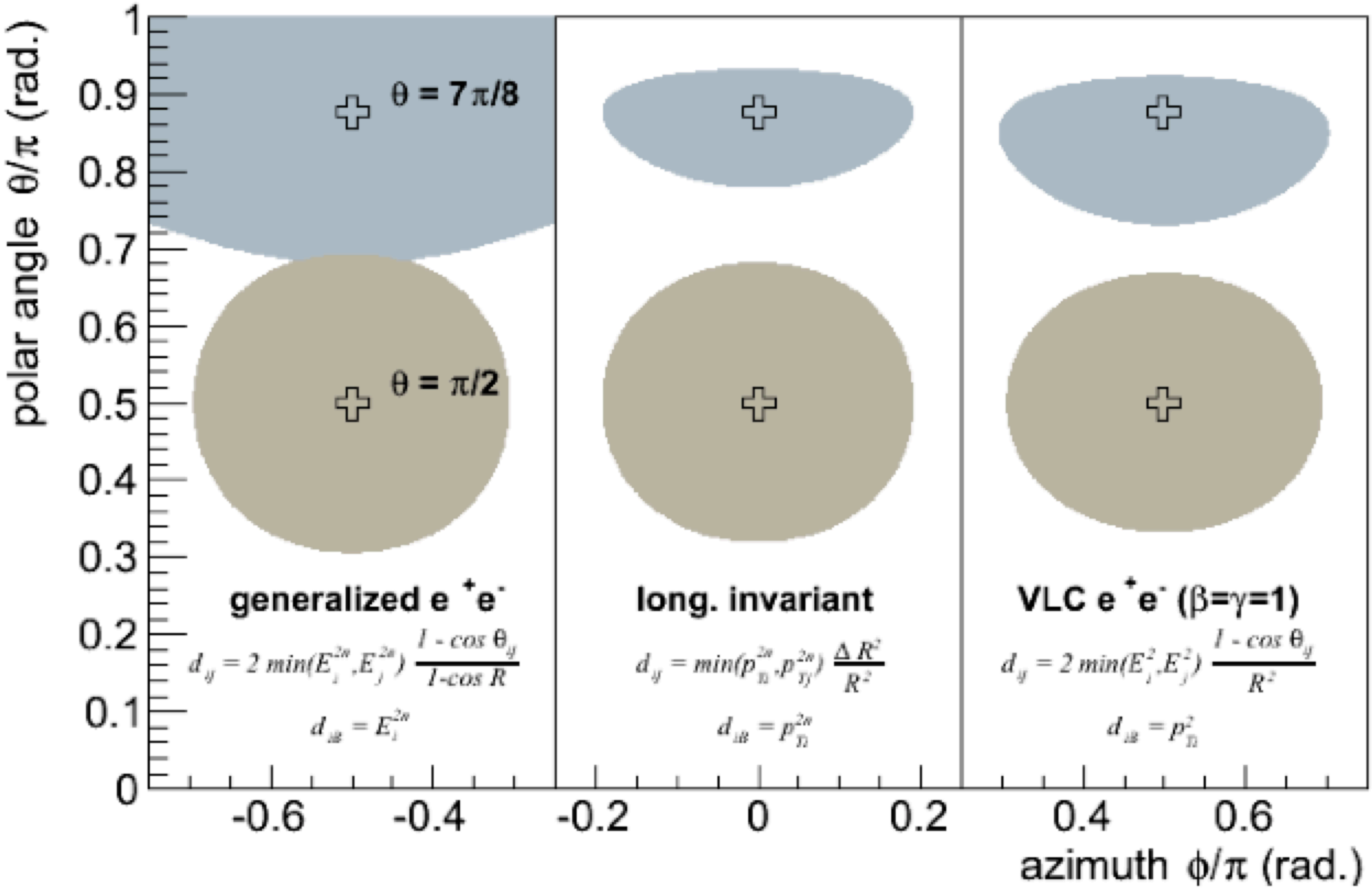
- Compact Linear Collider (CLIC):
  - ❑ CERN
  - ❑ http://clicdp.web.cern.ch/
  - ❑ √s: 380 GeV (1.4 km), 1.5 TeV (29.0 km), 3 TeV (50.1 km)
- International Linear Collider (ILC)
  - ❑ Japan (Kitakami)
  - ❑ https://www.linearcollider.org/ILC
  - ❑ √s: 500 GeV (31 km), possible upgrade to 1 TeV (50 km)
- Future Circular Collider ee (FCC-ee, old TLEP):
  - ❑ CERN
  - ❑ http://tlep.web.cern.ch/
  - ❑ √s: 90-400 GeV (circumference 80 – 100 km)
- Circular Electron Positron Collider (CEPC)
  - ❑ China (Qinghuada)
  - ❑ http://cepc.ihep.ac.cn/index.html
  - ❑ √s: 240 – 250 GeV (circumference 50 – 70 km)
- μ collider
  - ❑ http://map.fnal.gov/
  - ❑ √s: 2 TeV (?) (diameter 2 km ?, circumference 6-7 km ?)

Comparison of the distance criteria



- Two test particles with constant energy (E = 1 GeV) and fixed polar angle separation (100 mrad)
- The ratio of the inter-particle distance and the beam distance: d_{ij}/d_{iB} drives the robustness to bkg: the decision to assign the particle to final state or beam jets depends on this ratio (and R)
- Long. inv. k_t robustness is indeed due to its increasing d_{ij}/d_{iB} ratio
- VLC with β=1 is similar to long. inv. k_t

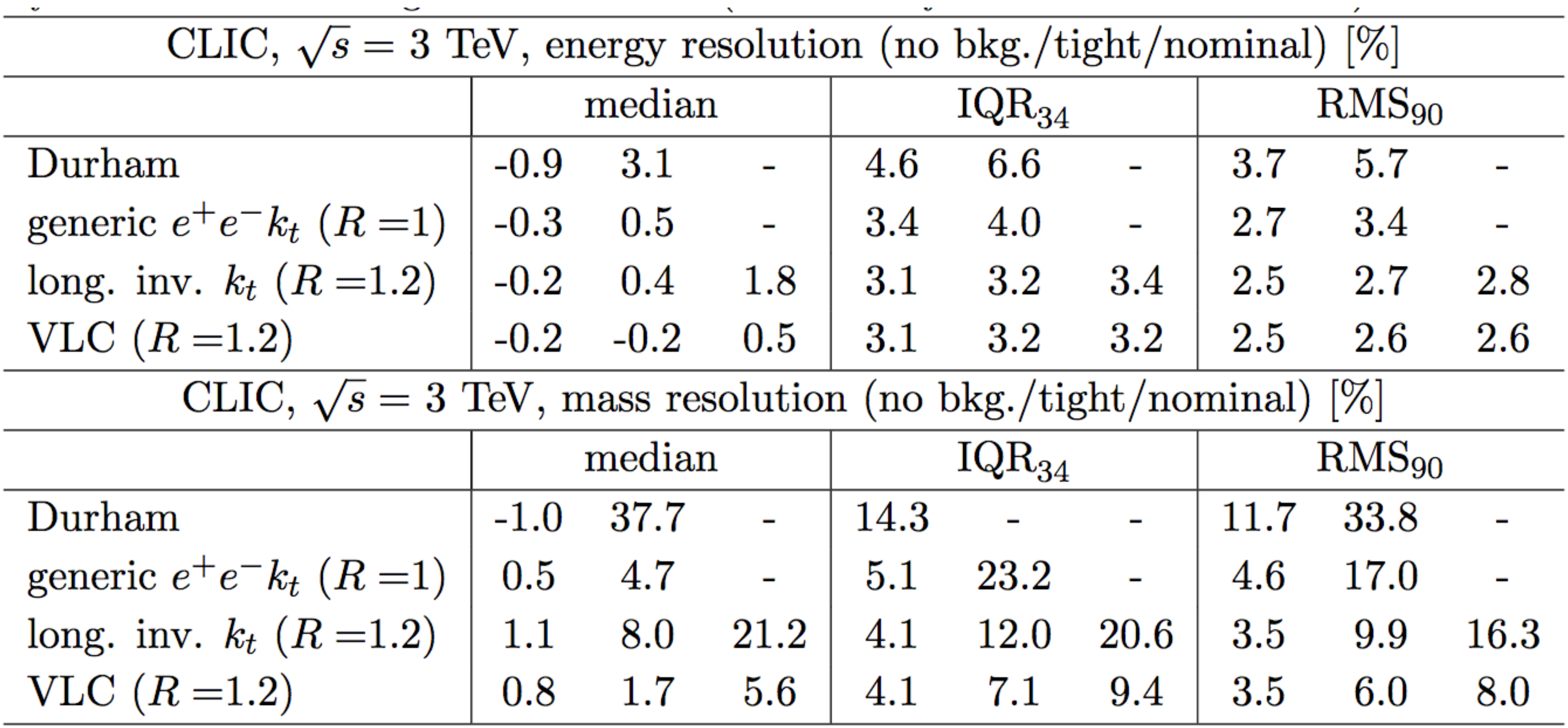
Jet footprint



ttbar table with full comparison

http://arxiv.org/pdf/1607.05039v1.pdf

- The bias and resolution of jet energy and mass for reco top jets



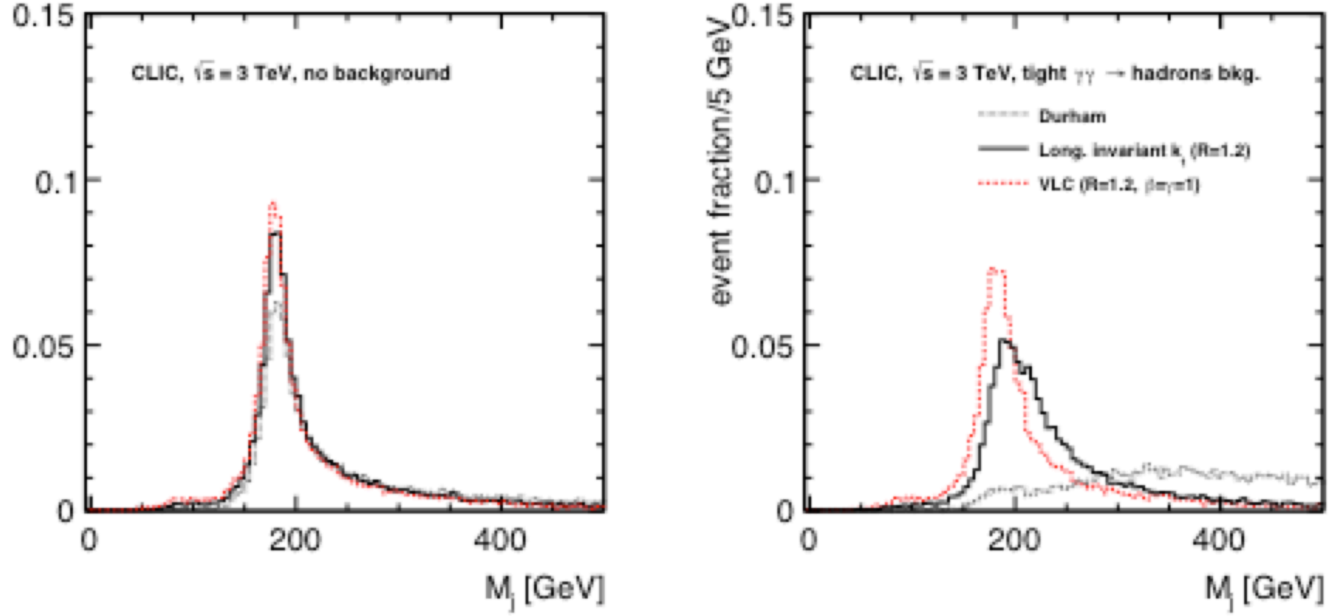
CLIC, $\sqrt{s} = 3$ TeV, energy resolution (no bkg./tight/nominal) [%]

	median			IQR$_{34}$			RMS$_{90}$		
Durham	-0.9	3.1	-	4.6	6.6	-	3.7	5.7	-
generic $e^+e^- k_t$ (R =1)	-0.3	0.5	-	3.4	4.0	-	2.7	3.4	-
long. inv. k_t (R =1.2)	-0.2	0.4	1.8	3.1	3.2	3.4	2.5	2.7	2.8
VLC (R =1.2)	-0.2	-0.2	0.5	3.1	3.2	3.2	2.5	2.6	2.6

CLIC, $\sqrt{s} = 3$ TeV, mass resolution (no bkg./tight/nominal) [%]

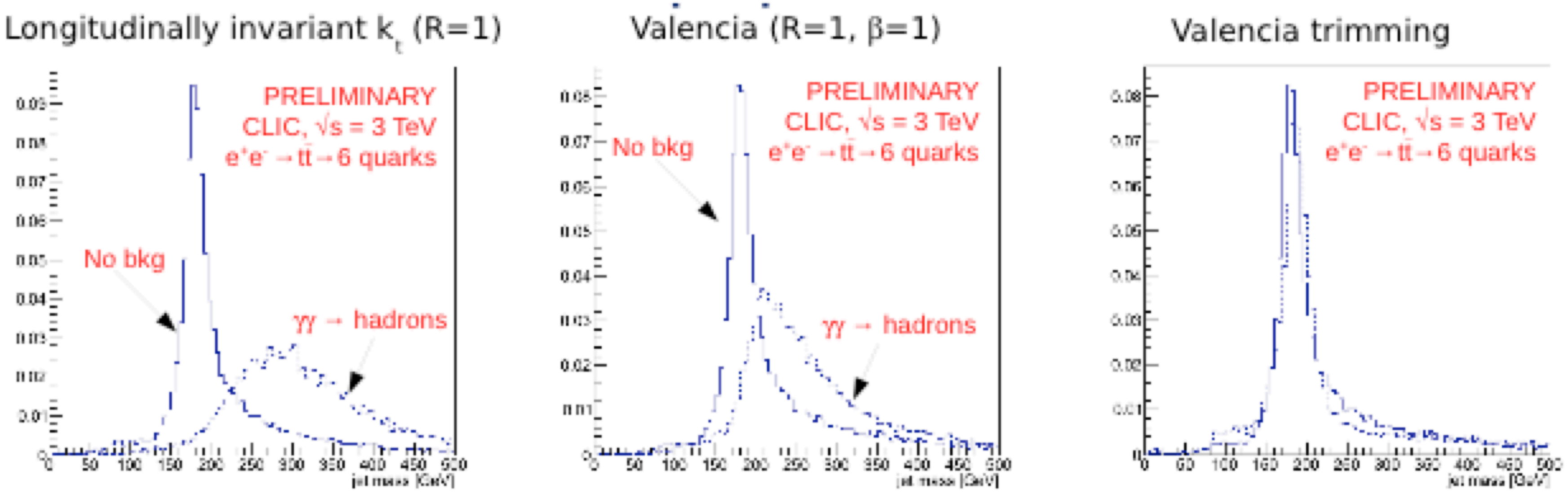
	median			IQR$_{34}$			RMS$_{90}$		
Durham	-1.0	37.7	-	14.3	-	-	11.7	33.8	-
generic $e^+e^- k_t$ (R =1)	0.5	4.7	-	5.1	23.2	-	4.6	17.0	-
long. inv. k_t (R =1.2)	1.1	8.0	21.2	4.1	12.0	20.6	3.5	9.9	16.3
VLC (R =1.2)	0.8	1.7	5.6	4.1	7.1	9.4	3.5	6.0	8.0

Hadronic tt at 3 TeV



with bkg overlayed

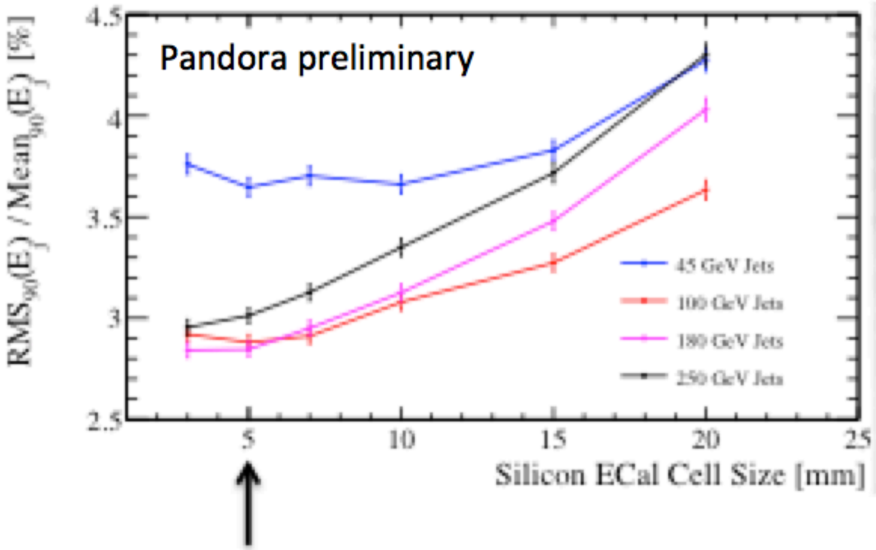
Algorithm	RMS90
Durham	33.8
Gen e+e- kt	17.0
Long inv kt	9.9
VLC	6.0


Longitudinally invariant k_t (R=1)

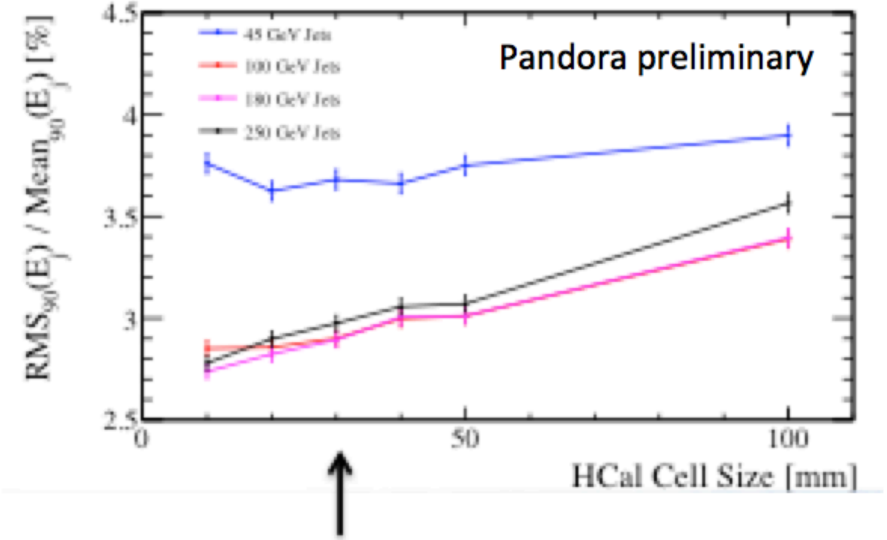
Valencia (R=1, β=1)

Valencia trimming

Cell size optimisation study



- Zuds events
- ECAL cell size: 5×5mm²
- Similar results with Scintillator as active material

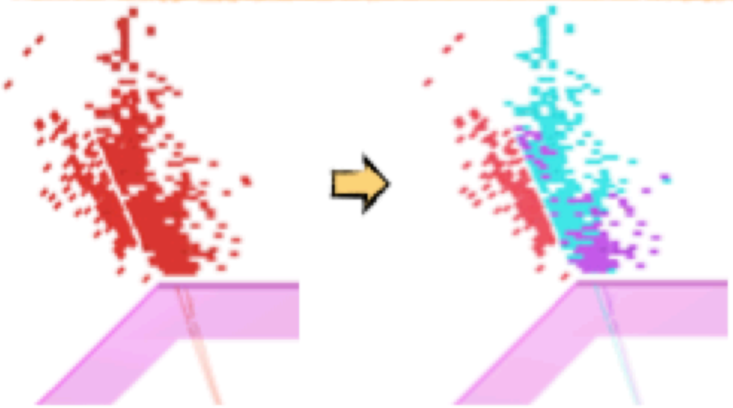


- Zuds events
- HCAL cell size: 30×30mm²

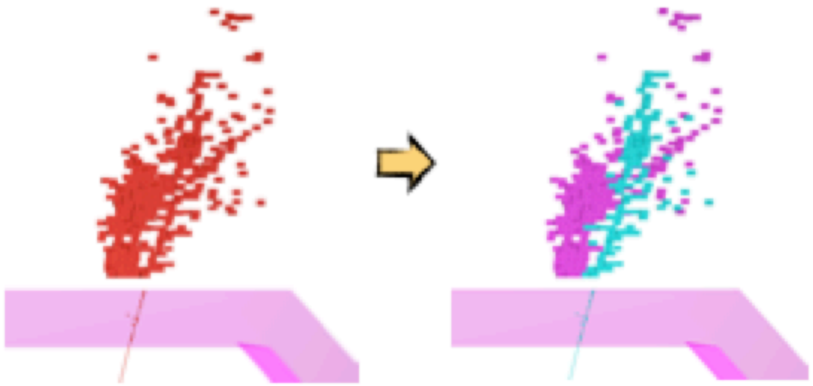
PandoraPFA

"The Pandora Software Development Kit for Pattern Recognition" → *EPJC.75.439, http://arxiv.org/abs/1506.05348*

- Pandora: approach to automated computer pattern recognition
- Exploit calorimeter granularity to gradually build-up picture of events
- Uses large numbers (70+) of algorithms, each carefully developed to address specific event topologies, with very few mistakes, and very careful to avoid accidentally merging energy deposits from separate particles
    - Cone Clustering, Topological Association, Track-Cluster Association, Reclustering Association, Fragment Removal, PFO Construction Algorithms



1. Multiple tracks associated to single cluster – split cluster.



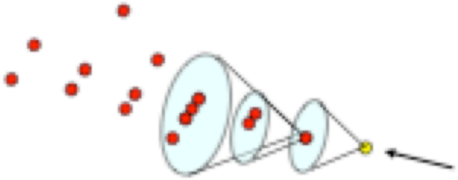
2. Cluster energy much greater than track momentum – split cluster.

from J. Marshall's talk at LCWS15

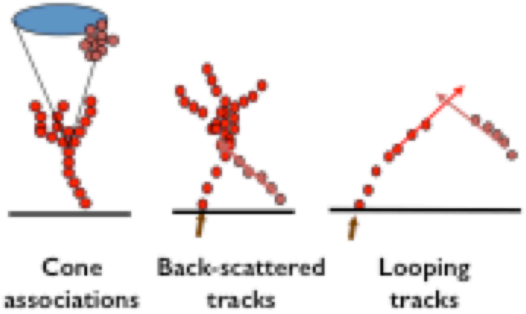
Pandora LC algorithms

from J. Marshall's talk at LCWS15

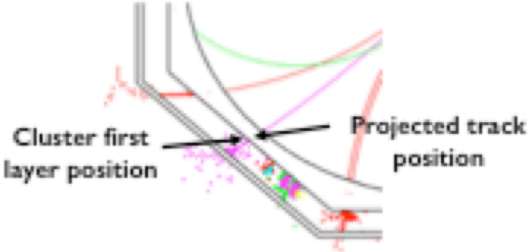
ConeClustering Algorithm



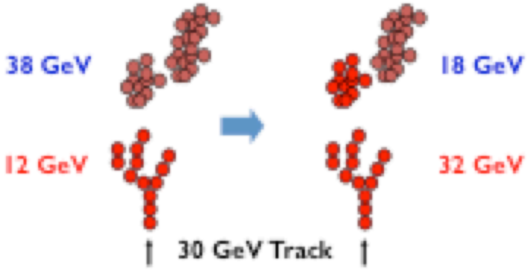
Topological Association Algorithms



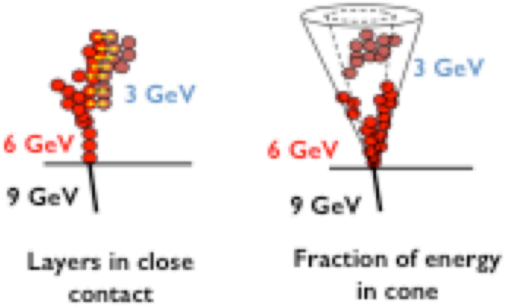
Track-Cluster Association Algorithms



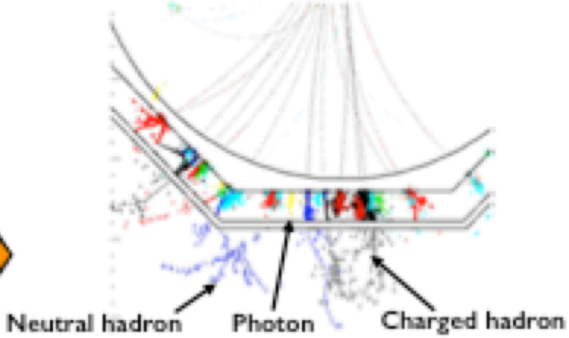
Reclustering Algorithms



Fragment Removal Algorithms

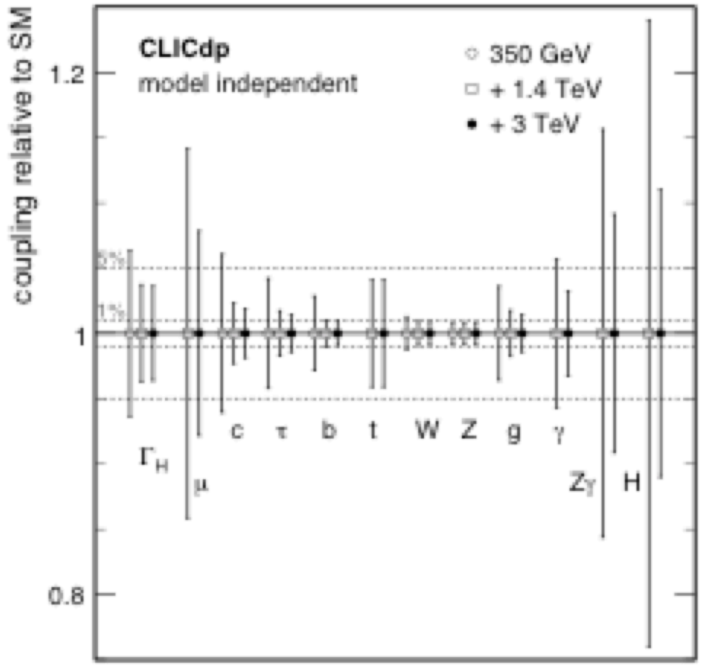


PFO Construction Algorithms



Higgs couplings

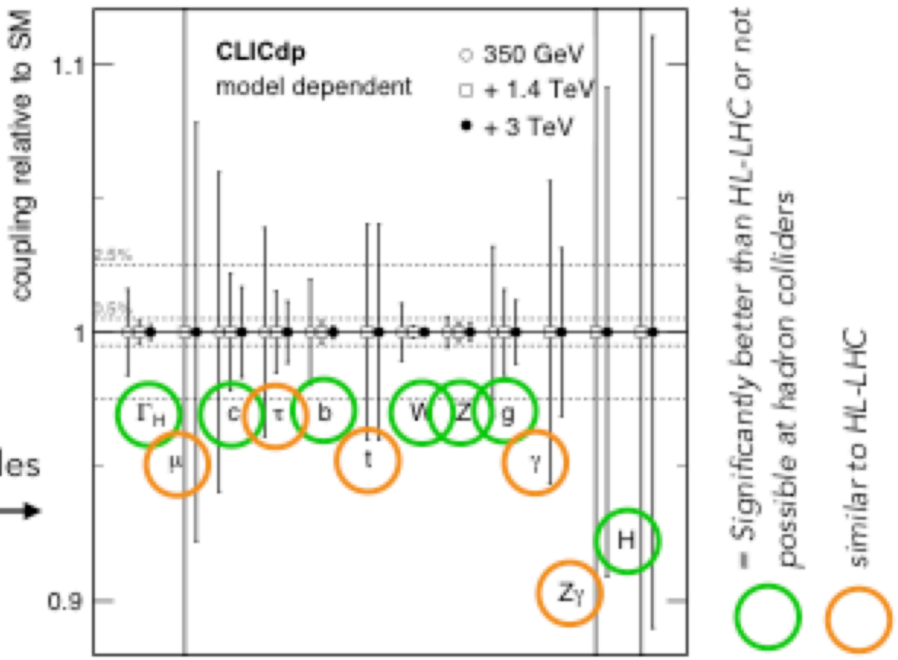
Model-independent global fits



Higgs width is a free parameter allows for additional non-SM decays.

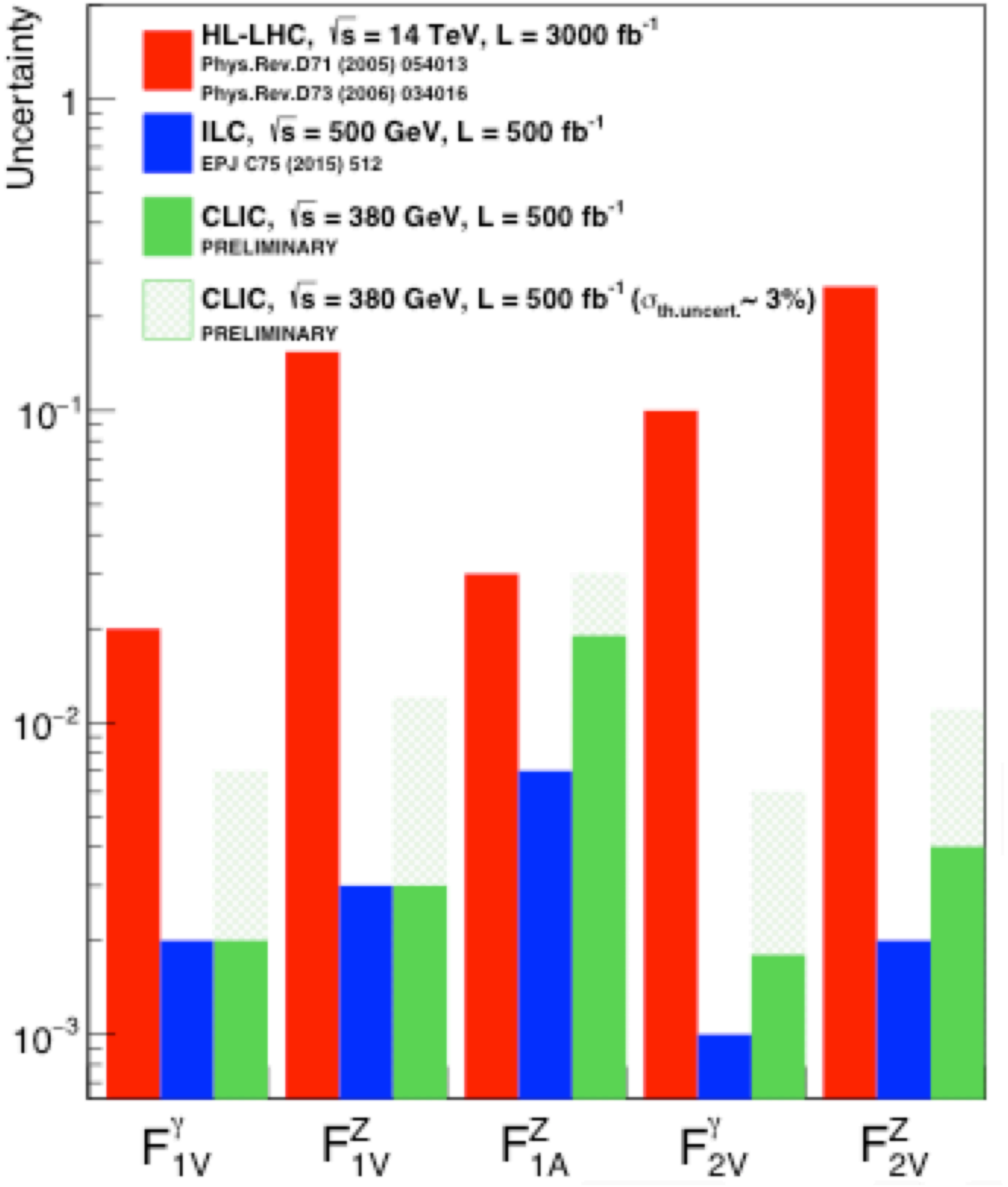
Note the different scales

Model-dependent global fits



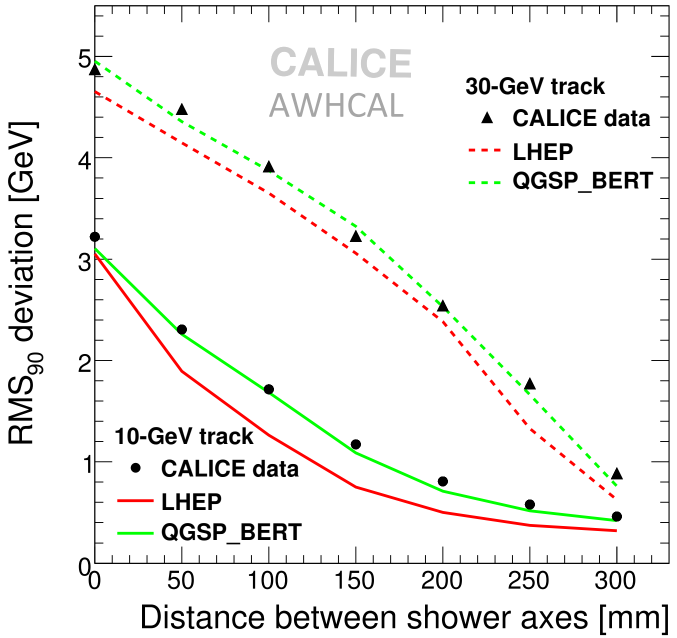
Constraining "LHC-style" fits assuming no invisible Higgs decays (model-dep.)

Uncertainties on top form factors



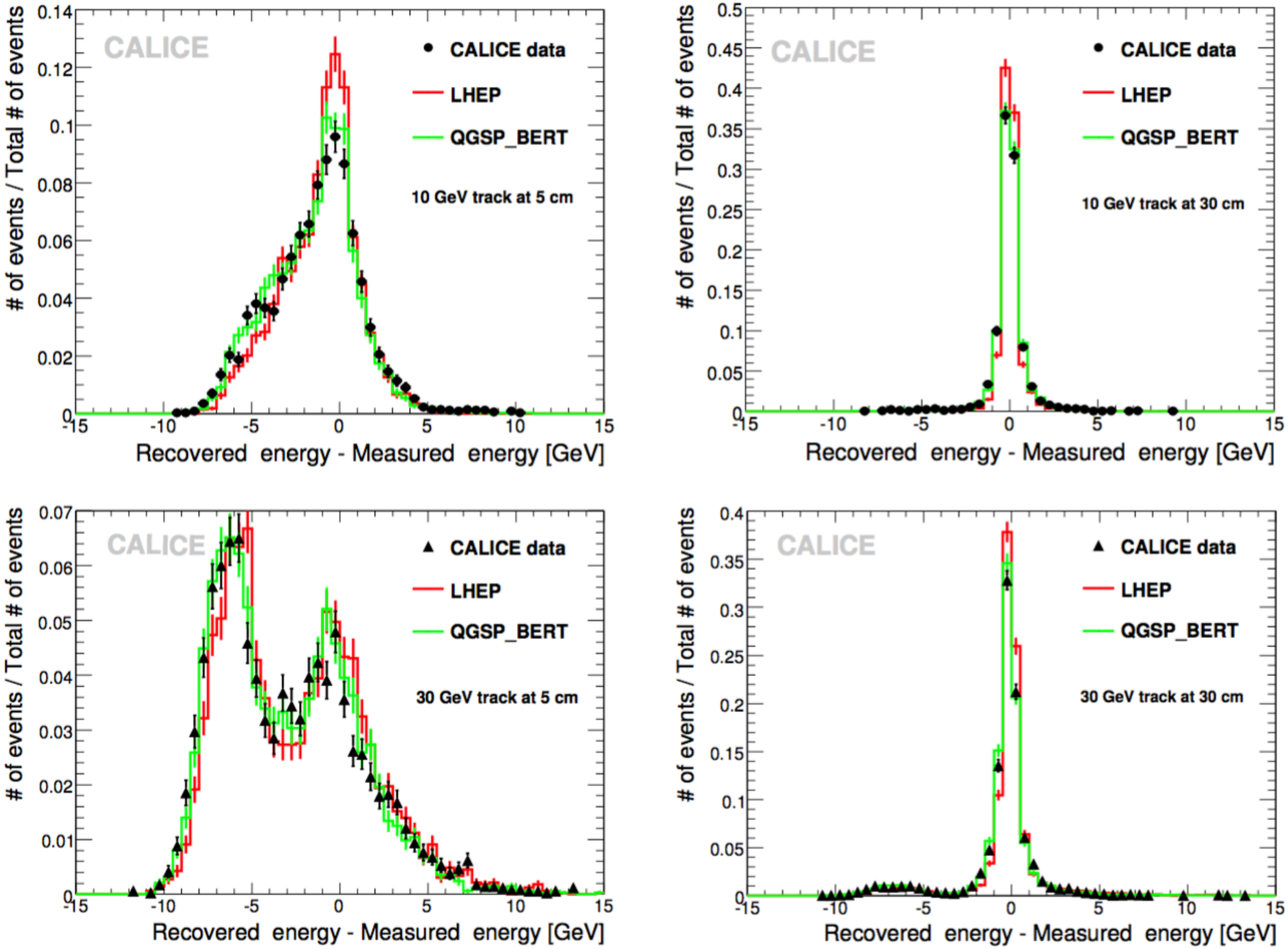
Confusion effects

- Overlay two test-beam events to study the reconstruction performance as a function of the transverse separation, *arXiv:1507.05893*
- Confusion effect quantified the deterioration of the neutral particle measurement induced by the presence of a nearby charged particle shower
- Moderate effect on jet energy resolution:
  - Small probability of 30 GeV particle in 100 GeV jets
  - Probability for very small distance small too → here no B, also less separation

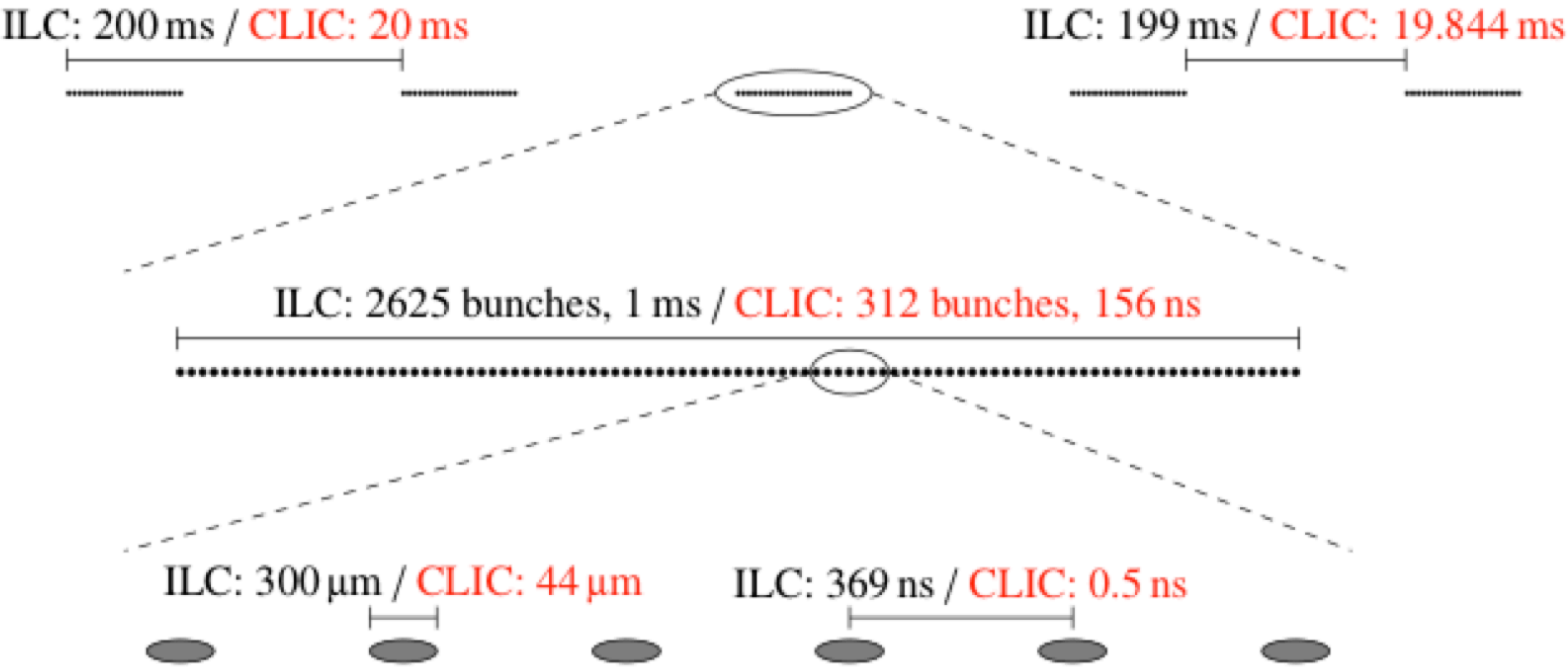


Confusion effects

arXiv:1507.05893



CLIC and ILC – beam structure



ILC: 200 ms / CLIC: 20 ms

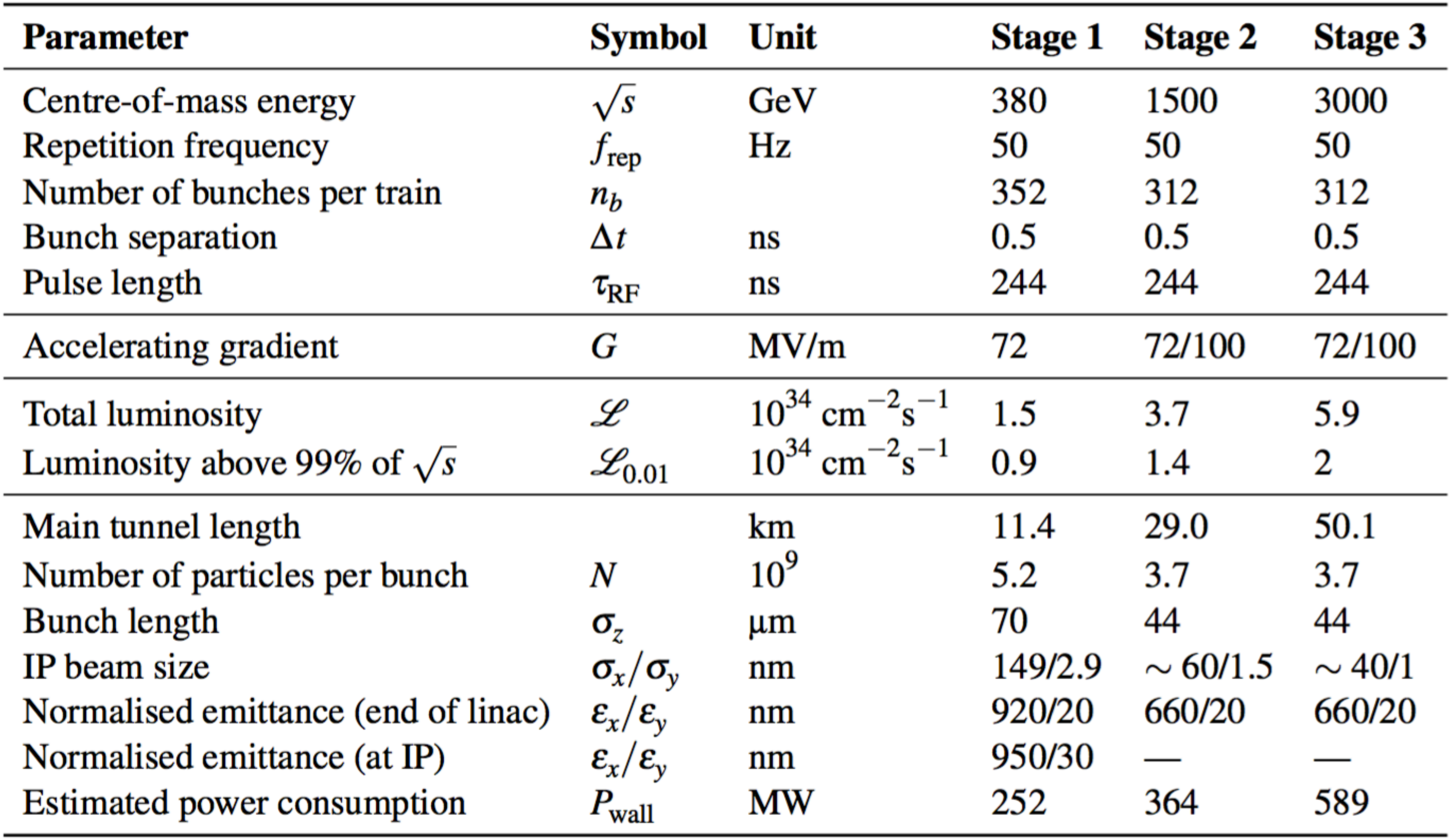
ILC: 199 ms / CLIC: 19.844 ms

ILC: 2625 bunches, 1 ms / CLIC: 312 bunches, 156 ns

ILC: 300 µm / CLIC: 44 µm

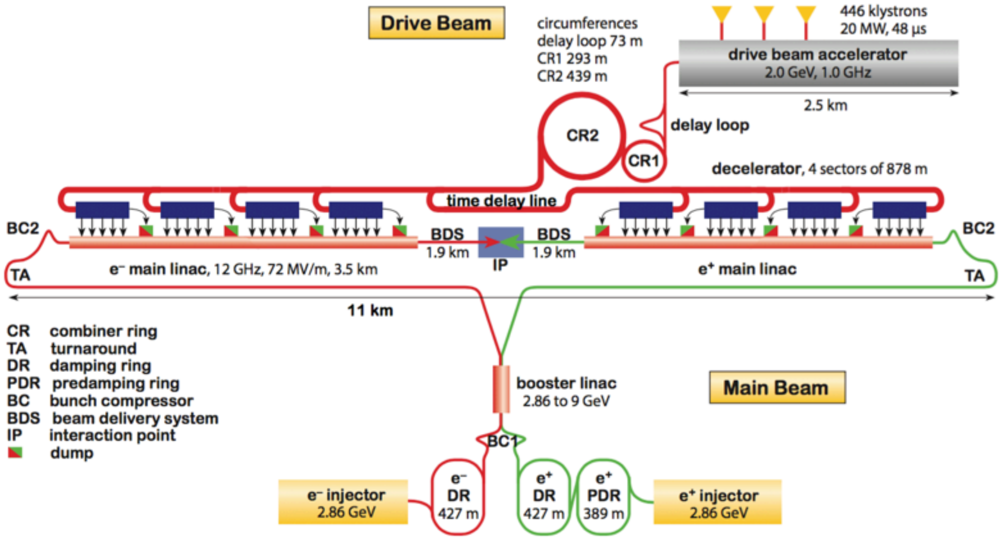
ILC: 369 ns / CLIC: 0.5 ns

CLIC machine parameters



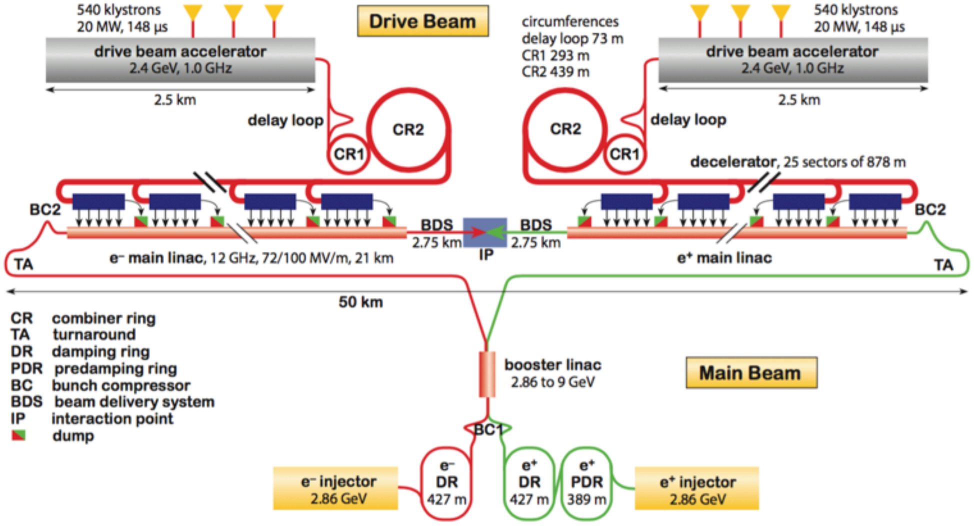
Parameter	**Symbol**	**Unit**	**Stage 1**	**Stage 2**	**Stage 3**
Centre-of-mass energy	$\sqrt{s}$	GeV	380	1500	3000
Repetition frequency	f_{rep}	Hz	50	50	50
Number of bunches per train	n_b		352	312	312
Bunch separation	Δt	ns	0.5	0.5	0.5
Pulse length	τ_{RF}	ns	244	244	244
Accelerating gradient	G	MV/m	72	72/100	72/100
Total luminosity	$\mathscr{L}$	10^{34} cm^{-2}s^{-1}	1.5	3.7	5.9
Luminosity above 99% of $\sqrt{s}$	$\mathscr{L}_{0.01}$	10^{34} cm^{-2}s^{-1}	0.9	1.4	2
Main tunnel length		km	11.4	29.0	50.1
Number of particles per bunch	N	10^9	5.2	3.7	3.7
Bunch length	σ_z	µm	70	44	44
IP beam size	σ_x/σ_y	nm	149/2.9	$\sim$ 60/1.5	$\sim$ 40/1
Normalised emittance (end of linac)	$\varepsilon_x/\varepsilon_y$	nm	920/20	660/20	660/20
Normalised emittance (at IP)	$\varepsilon_x/\varepsilon_y$	nm	950/30	—	—
Estimated power consumption	P_{wall}	MW	252	364	589

CLIC dual beam scheme

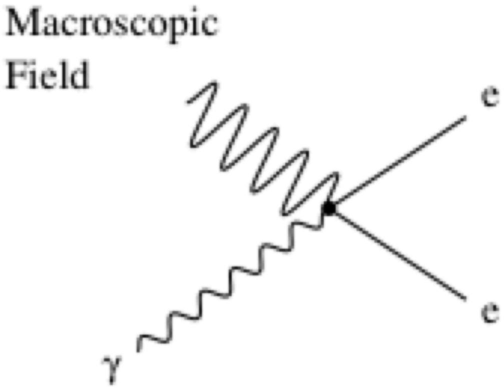


1 drive beams accelerator for 380 GeV and 1.5 TeV

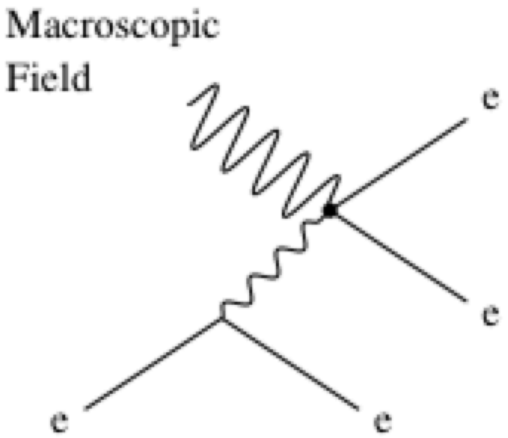
2 drive beams accelerators for 3 TeV



Pair production

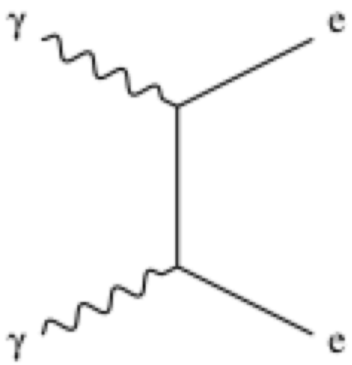


(a) Coherent Pair Production

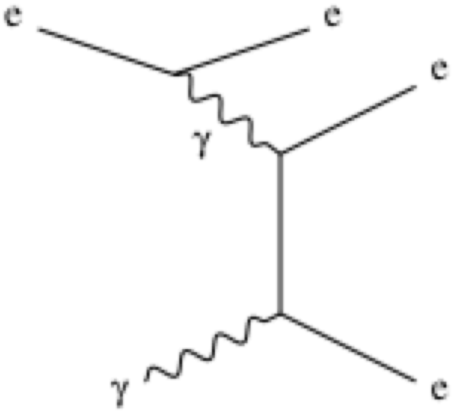


(b) Trident Pair Production

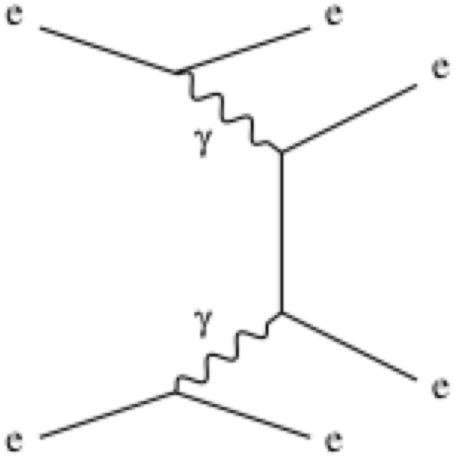
Incoherent Pair



(a) Breit-Wheeler



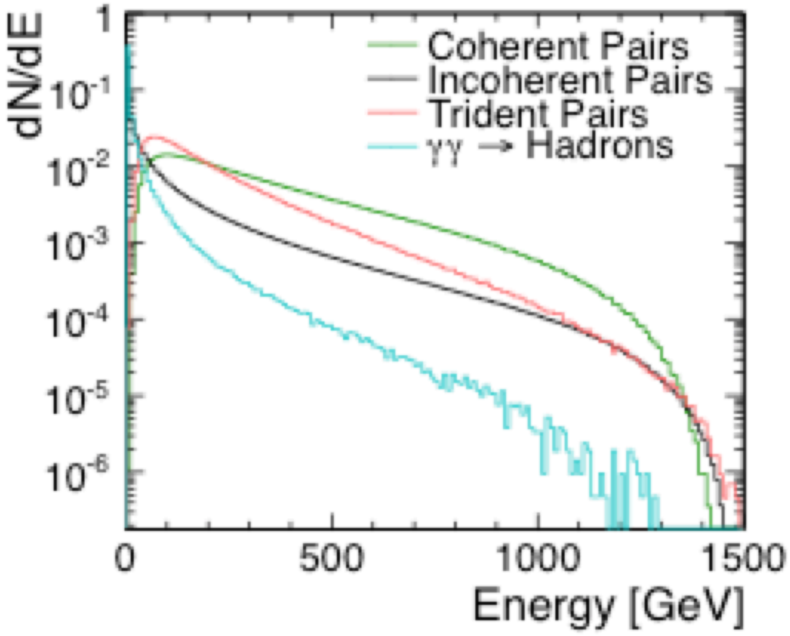
(b) Bethe-Heitler

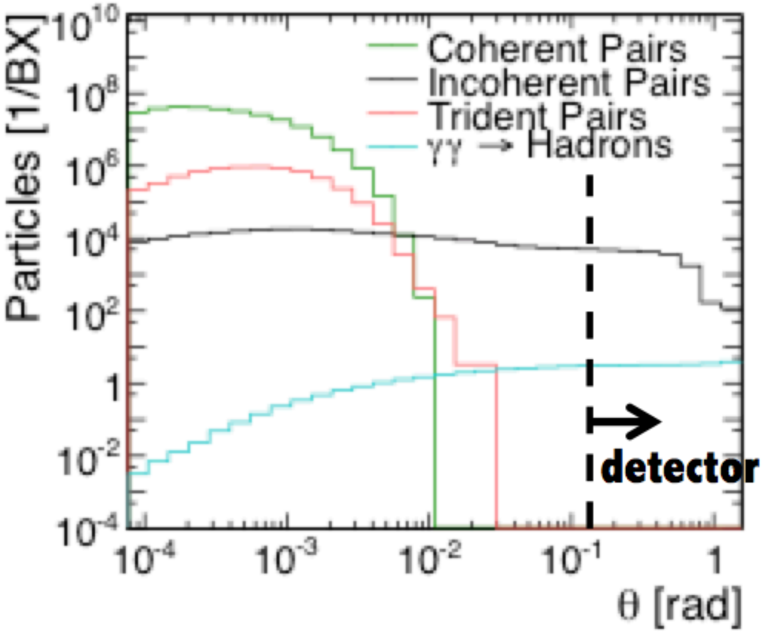


(c) Landau-Lifshitz

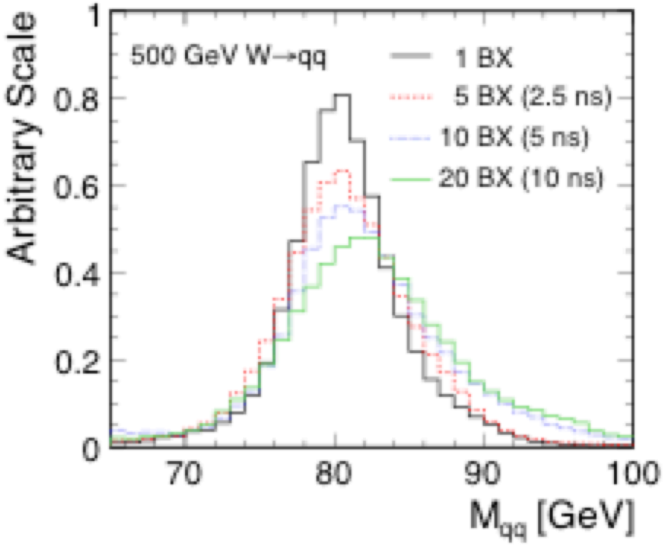
Background characterisation

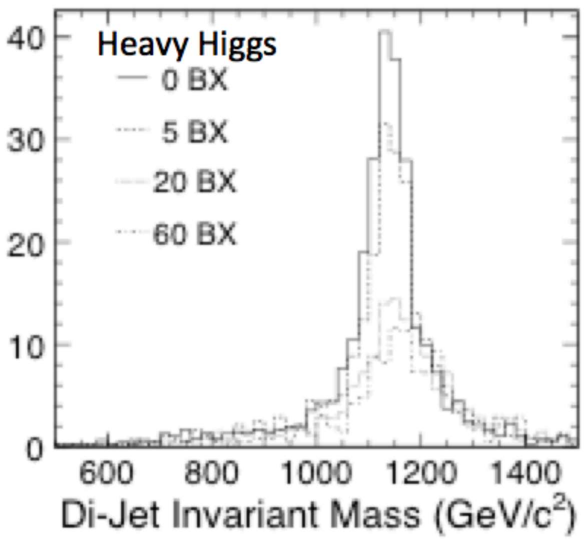
- High background levels at low p_T, θ
  - **Incoherent pair production**
    - 3×10^5 / BX
    - Mostly low angle
  - **γγ->hadrons**
    - 3.2 / BX for E(γγ) > 2 GeV
    - Main bkg in central tracking and calorimeter systems
- Read out the full train → > 19 TeV visible energy per 156 ns train

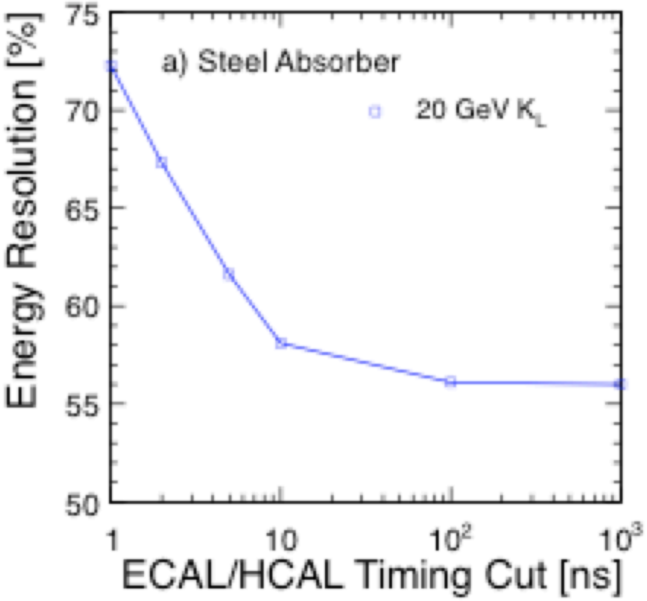


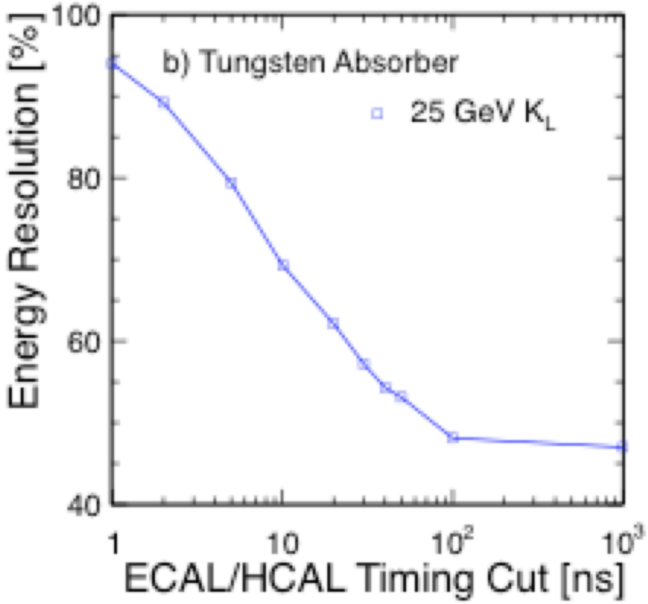


Timing cuts


500 GeV W→qq
1 BX
5 BX (2.5 ns)
10 BX (5 ns)
20 BX (10 ns)
Arbitrary Scale
0
0.2
0.4
0.6
0.8
1
70
80
90
100
M_{qq} [GeV]


Heavy Higgs
0 BX
5 BX
20 BX
60 BX
0
10
20
30
40
600
800
1000
1200
1400
Di-Jet Invariant Mass (GeV/c^2)


a) Steel Absorber
20 GeV K_L
Energy Resolution [%]
50
55
60
65
70
75
1
10
10^2
10^3
ECAL/HCAL Timing Cut [ns]


b) Tungsten Absorber
25 GeV K_L
Energy Resolution [%]
40
60
80
100
1
10
10^2
10^3
ECAL/HCAL Timing Cut [ns]

PFO selection – Loose

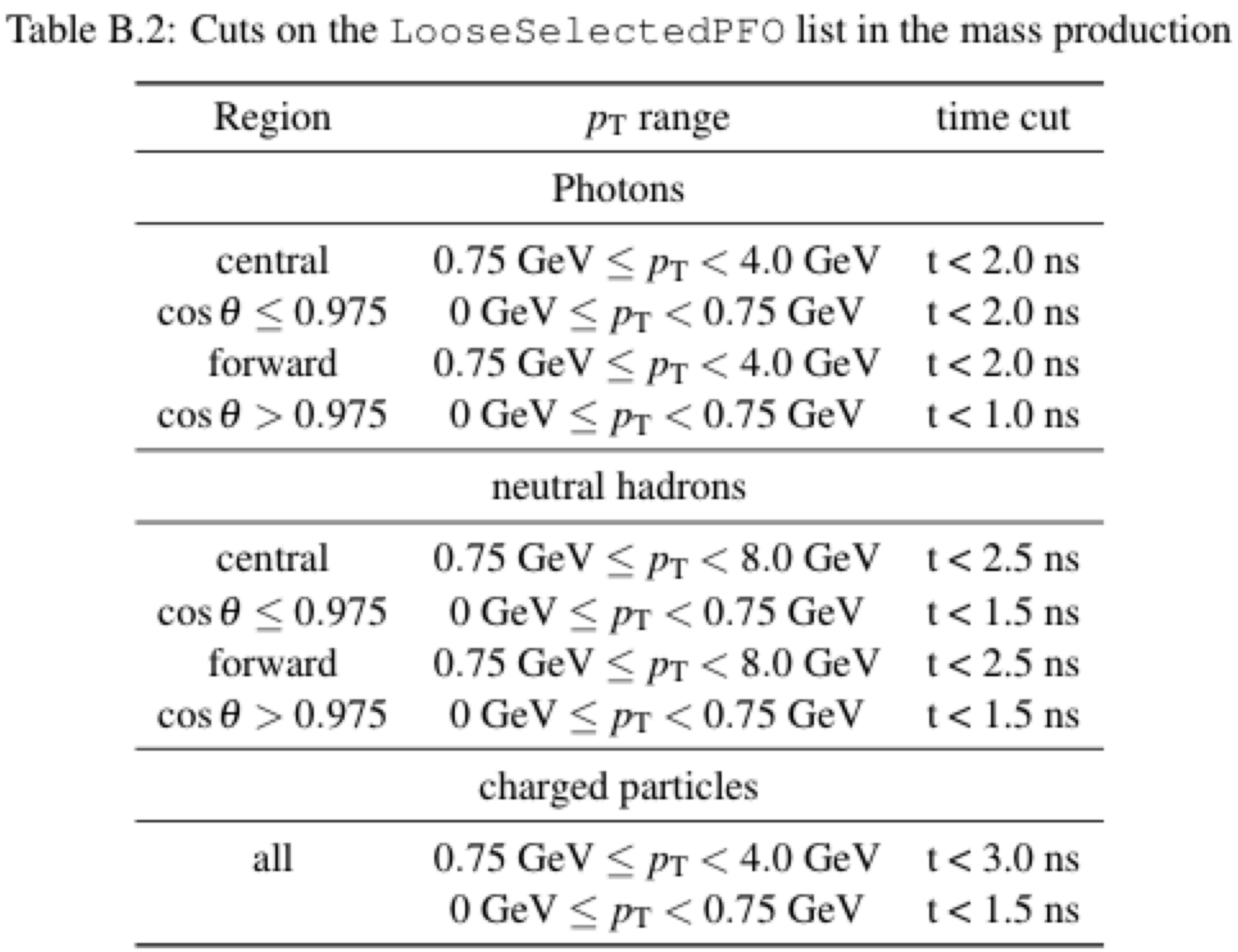


Table B.2: Cuts on the `LooseSelectedPFO` list in the mass production

Region	p_T range	time cut
Photons		
central $\cos\theta \leq 0.975$	$0.75\ \text{GeV} \leq p_T < 4.0\ \text{GeV}$	$t < 2.0$ ns
	$0\ \text{GeV} \leq p_T < 0.75\ \text{GeV}$	$t < 2.0$ ns
forward $\cos\theta > 0.975$	$0.75\ \text{GeV} \leq p_T < 4.0\ \text{GeV}$	$t < 2.0$ ns
	$0\ \text{GeV} \leq p_T < 0.75\ \text{GeV}$	$t < 1.0$ ns
neutral hadrons		
central $\cos\theta \leq 0.975$	$0.75\ \text{GeV} \leq p_T < 8.0\ \text{GeV}$	$t < 2.5$ ns
	$0\ \text{GeV} \leq p_T < 0.75\ \text{GeV}$	$t < 1.5$ ns
forward $\cos\theta > 0.975$	$0.75\ \text{GeV} \leq p_T < 8.0\ \text{GeV}$	$t < 2.5$ ns
	$0\ \text{GeV} \leq p_T < 0.75\ \text{GeV}$	$t < 1.5$ ns
charged particles		
all	$0.75\ \text{GeV} \leq p_T < 4.0\ \text{GeV}$	$t < 3.0$ ns
	$0\ \text{GeV} \leq p_T < 0.75\ \text{GeV}$	$t < 1.5$ ns

PFO selection – Default

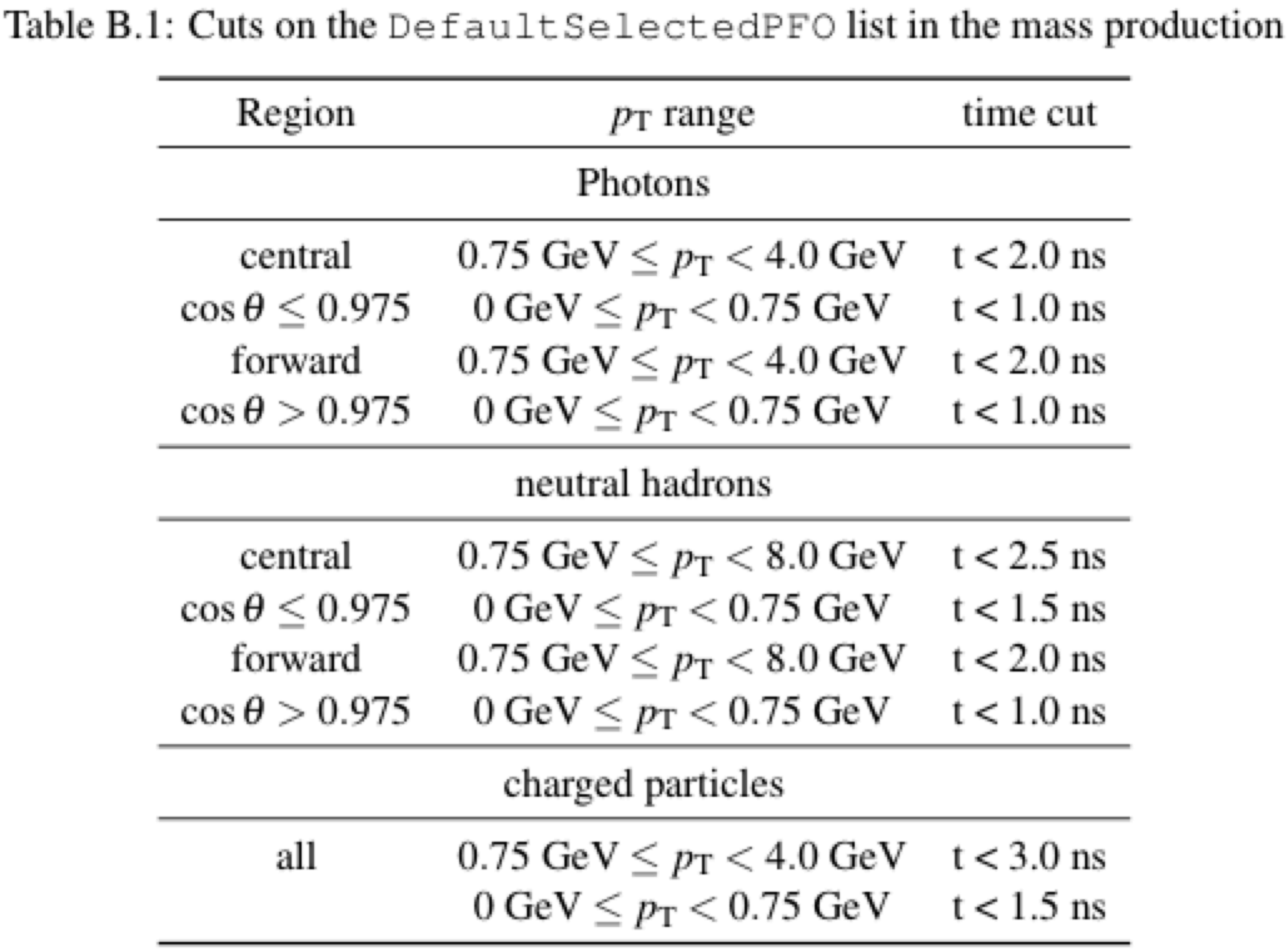


Table B.1: Cuts on the `DefaultSelectedPFO` list in the mass production

Region	p_T range	time cut
	Photons	
central	$0.75\ \text{GeV} \leq p_T < 4.0\ \text{GeV}$	$t < 2.0$ ns
$\cos\theta \leq 0.975$	$0\ \text{GeV} \leq p_T < 0.75\ \text{GeV}$	$t < 1.0$ ns
forward	$0.75\ \text{GeV} \leq p_T < 4.0\ \text{GeV}$	$t < 2.0$ ns
$\cos\theta > 0.975$	$0\ \text{GeV} \leq p_T < 0.75\ \text{GeV}$	$t < 1.0$ ns
	neutral hadrons	
central	$0.75\ \text{GeV} \leq p_T < 8.0\ \text{GeV}$	$t < 2.5$ ns
$\cos\theta \leq 0.975$	$0\ \text{GeV} \leq p_T < 0.75\ \text{GeV}$	$t < 1.5$ ns
forward	$0.75\ \text{GeV} \leq p_T < 8.0\ \text{GeV}$	$t < 2.0$ ns
$\cos\theta > 0.975$	$0\ \text{GeV} \leq p_T < 0.75\ \text{GeV}$	$t < 1.0$ ns
	charged particles	
all	$0.75\ \text{GeV} \leq p_T < 4.0\ \text{GeV}$	$t < 3.0$ ns
	$0\ \text{GeV} \leq p_T < 0.75\ \text{GeV}$	$t < 1.5$ ns

PFO selection – Tight

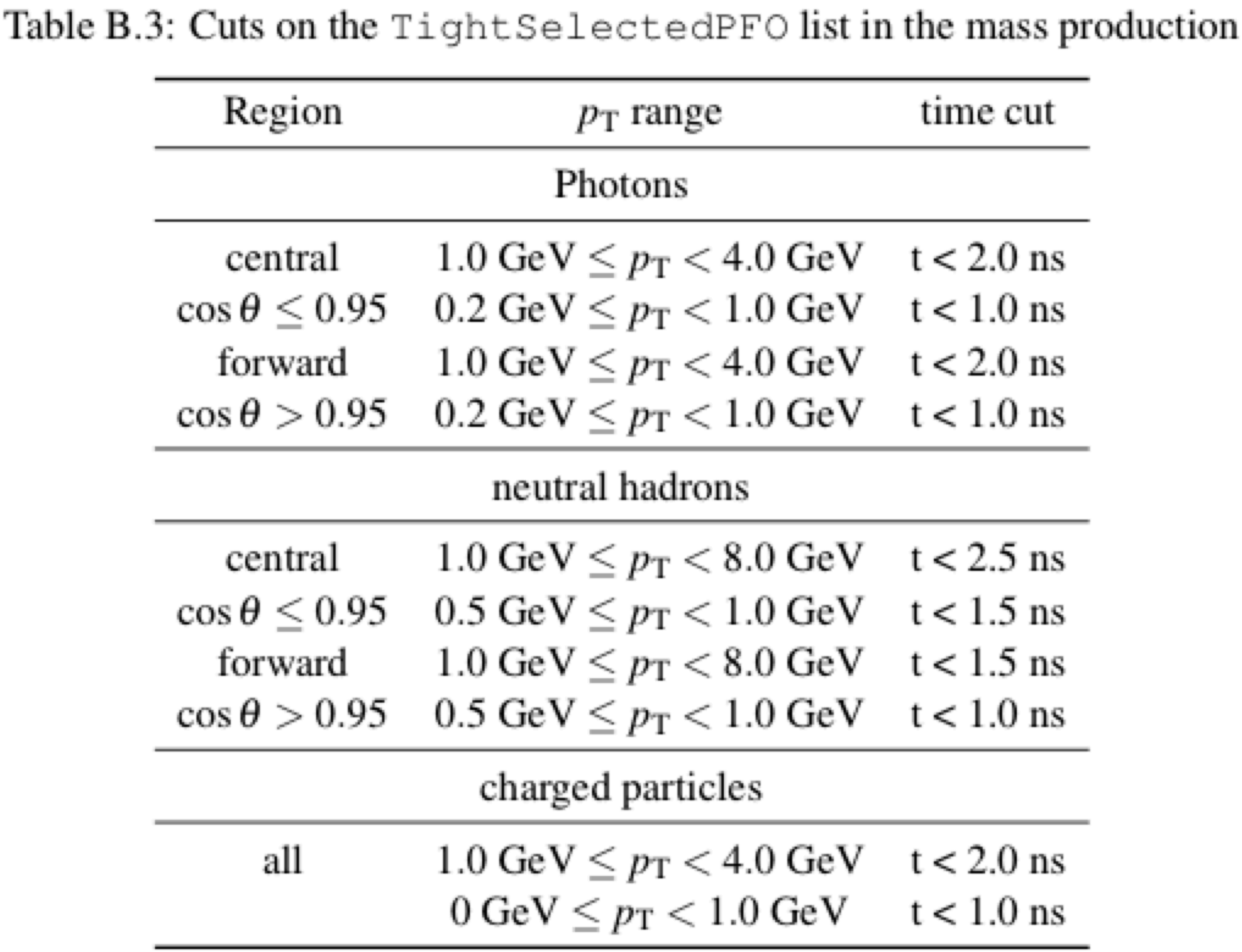
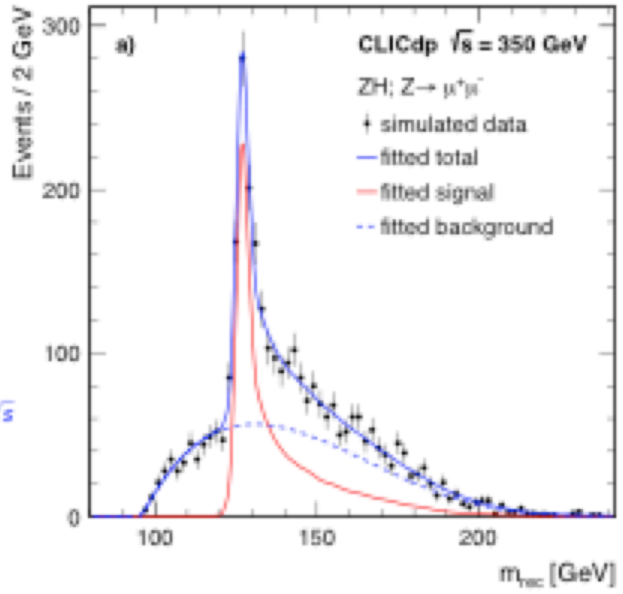
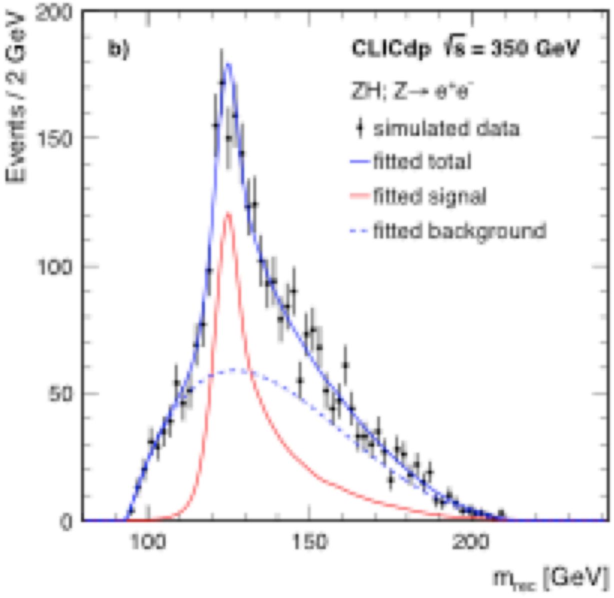


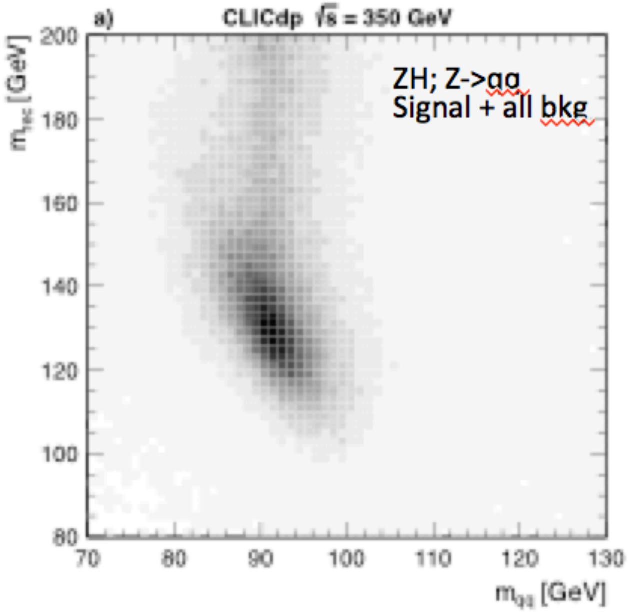
Table B.3: Cuts on the `TightSelectedPFO` list in the mass production

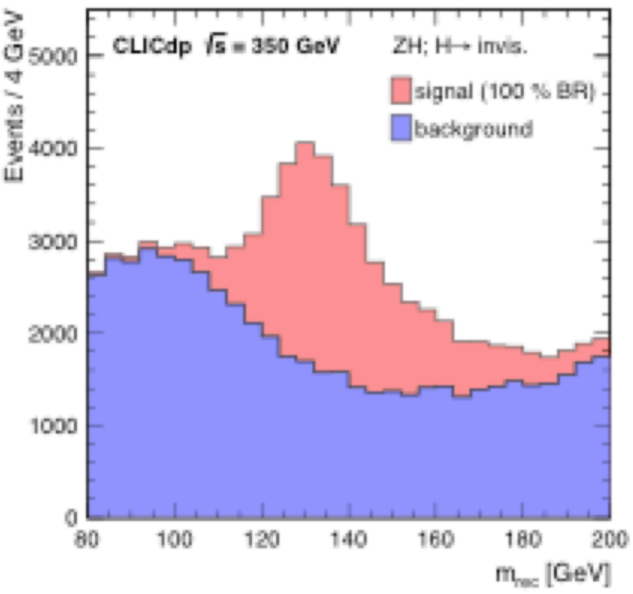
Region	p_T range	time cut
Photons		
central	$1.0\ \text{GeV} \leq p_T < 4.0\ \text{GeV}$	t < 2.0 ns
$\cos\theta \leq 0.95$	$0.2\ \text{GeV} \leq p_T < 1.0\ \text{GeV}$	t < 1.0 ns
forward	$1.0\ \text{GeV} \leq p_T < 4.0\ \text{GeV}$	t < 2.0 ns
$\cos\theta > 0.95$	$0.2\ \text{GeV} \leq p_T < 1.0\ \text{GeV}$	t < 1.0 ns
neutral hadrons		
central	$1.0\ \text{GeV} \leq p_T < 8.0\ \text{GeV}$	t < 2.5 ns
$\cos\theta \leq 0.95$	$0.5\ \text{GeV} \leq p_T < 1.0\ \text{GeV}$	t < 1.5 ns
forward	$1.0\ \text{GeV} \leq p_T < 8.0\ \text{GeV}$	t < 1.5 ns
$\cos\theta > 0.95$	$0.5\ \text{GeV} \leq p_T < 1.0\ \text{GeV}$	t < 1.0 ns
charged particles		
all	$1.0\ \text{GeV} \leq p_T < 4.0\ \text{GeV}$	t < 2.0 ns
	$0\ \text{GeV} \leq p_T < 1.0\ \text{GeV}$	t < 1.0 ns

Higgsstrahlung – recoil mass





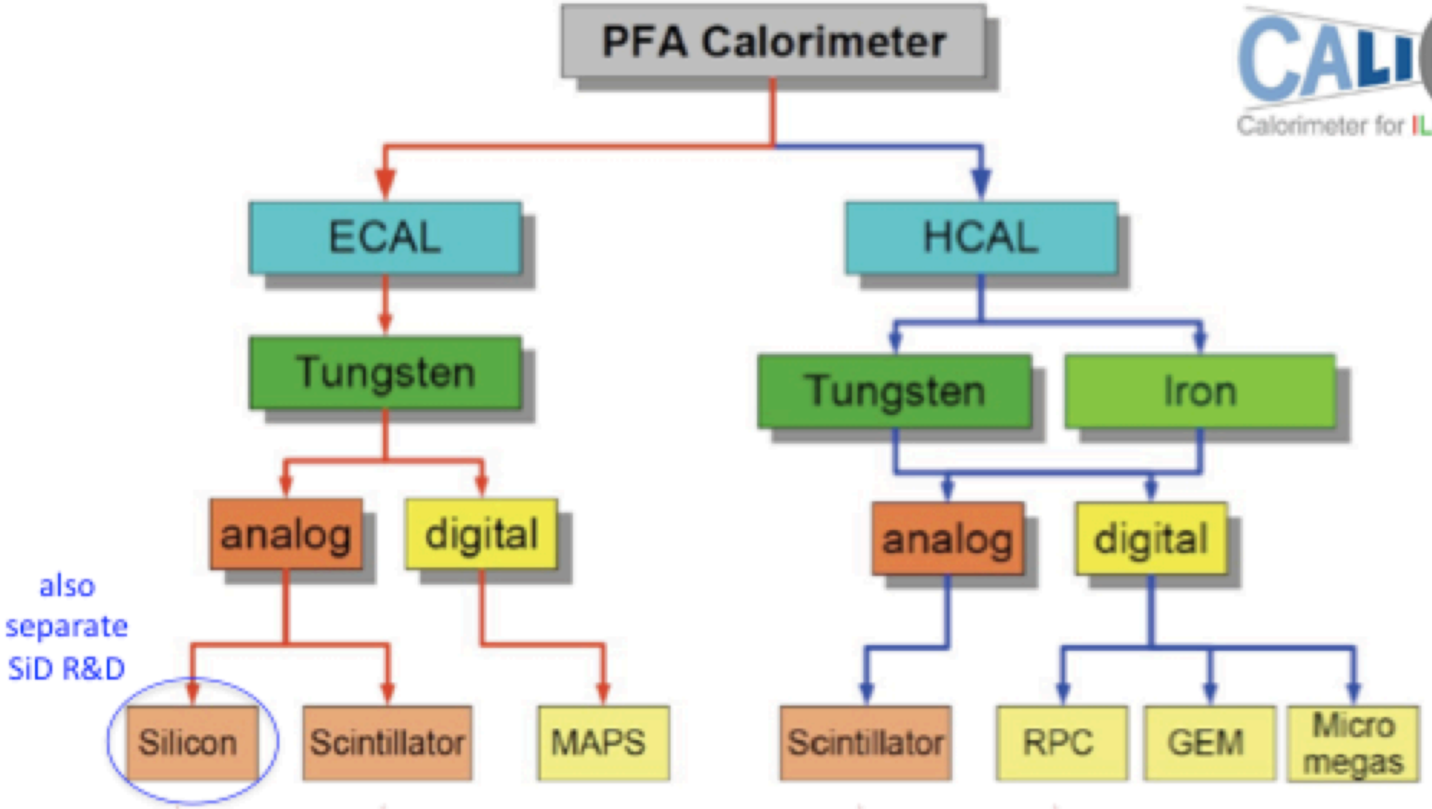




CALICE program

https://twiki.cern.ch/twiki/bin/view/CALICE/CaliceCollaboration

- CALICE test beam of novel technologies for ECAL and HCAL prototypes
  - ❑ Demonstration of detector calibration capabilities
  - ❑ Characterisation of prototypes: linearity, resolution
  - ❑ Measurements of particle shower evolution and comparison with Geant4



CALICE Calorimeter for ILC