ZH at high energy

Matthias Weber (CERN)

Baseline analysis strategy

HZ →bb qq at $\sqrt{s_{eff}} > 2500$ GeV characterised by two high-energy boosted fat jets, back-to-back in azimuth

Excellent jet mass resolution helps to discriminate between signal and background signatures

- Investigate subjet behavior and jet substructure
- Concentrate on H→bb, use b-tagging information to reject backgrounds and to help selecting H jet

→accurate measurement of underlying $\sqrt{s_{eff}}$ beneficial for EFT fits

Objects

Jets defined using VLC algorithm with β=γ=1.0, run in exclusive mode with R=0.7with n_{jets}=2
→tight timing and p_T selection applied on particle flow objects for jet clustering

Check for isolated leptons and photons with E>10 GeV
Requirement: relative isolation relIso<0.10 within a cone of 10 degrees
→lepton veto in event selection

Ordering jets by masses (j1) >m(j2), treat j1 as H jet, j2 as Z jet

Signal and backgrounds for $\sqrt{s}_{reco}$>2500 GeV

Total Cross-section: ee→qqqqqq sum of 14 different six quark datasets

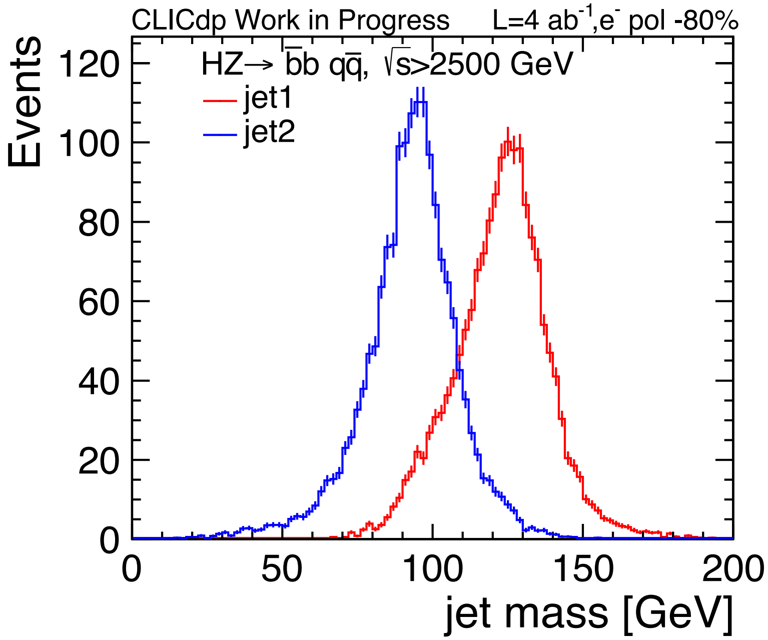
Process	Cross-section in fb Polarization e^- -80 %	Cross-section in fb Polarization e^- +80 %
HZ → H qq	3.78	2.67
ee→qq	1269	786
ee→qqqq	902	120
ee→qqqqqq	64.4	22.3

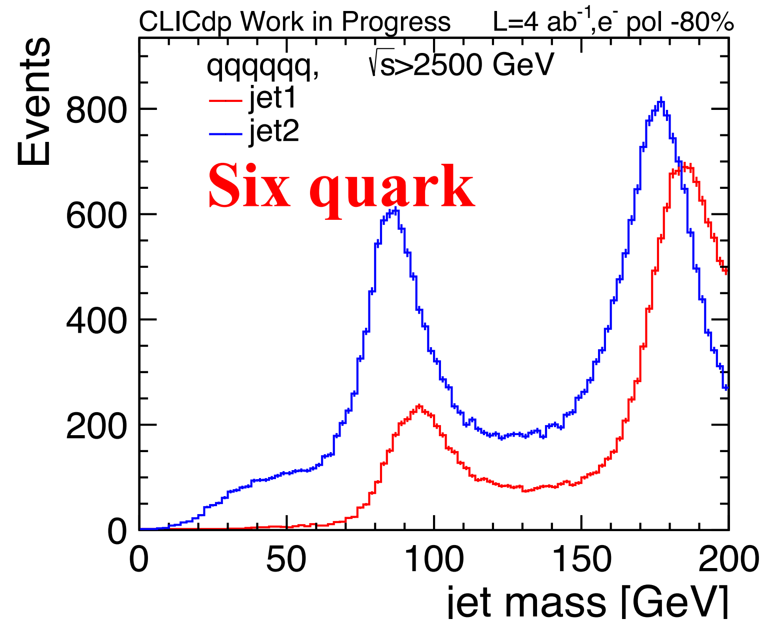
Event numbers for cut on reconstructed $\sqrt{s}$ > 2500 GeV

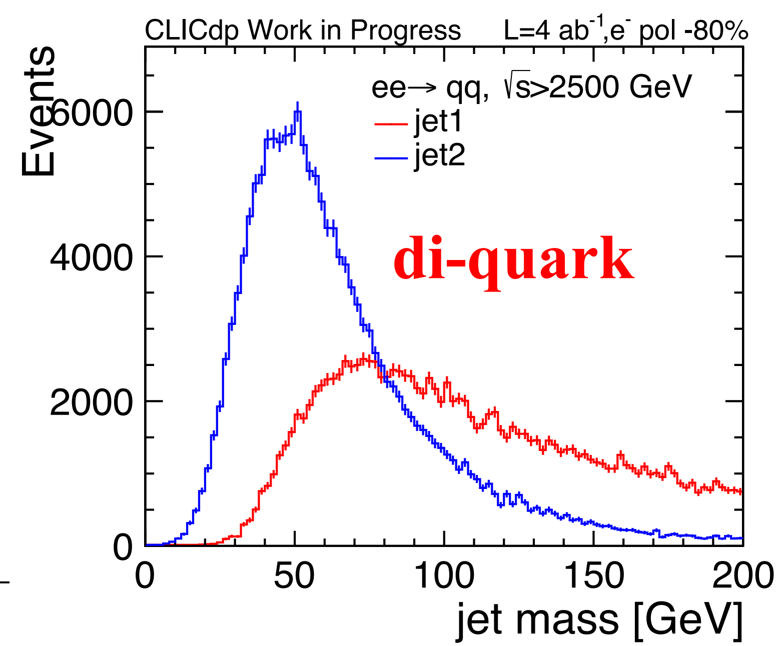
Process	$\sqrt{s}_{reco}$>2500 GeV e^- pol -80 %, L=4 ab^{-1}	$\sqrt{s}_{reco}$>2500 GeV e^- pol +80 %, L=1 ab^{-1}
HZ → bb qq	1660	291
ee→qq	165 000	17 800
ee→qqqq	239 000	8 090
ee→qqqqqq	31 300	3 010

Leading jet masses

HZ signal





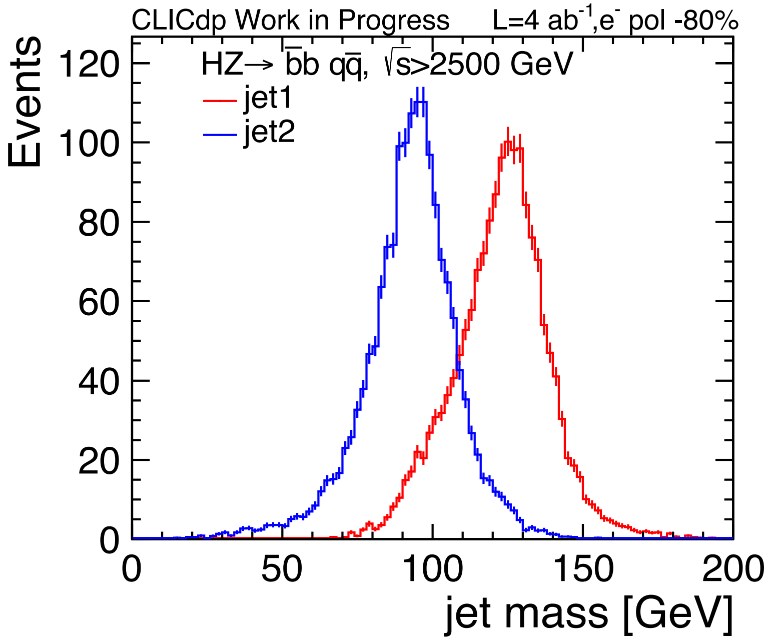


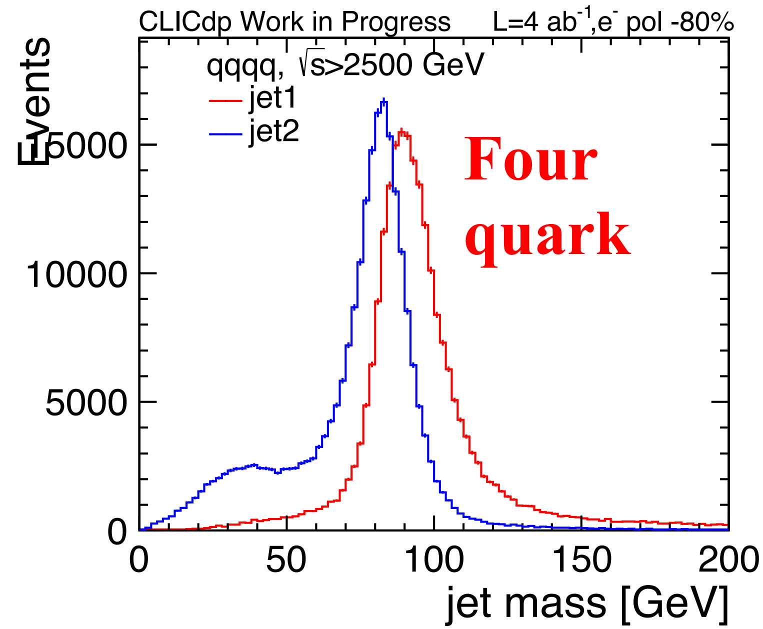
Clear peak at Z and H mass for signal

- for diquark sample peaks at lower values with long tail
- Six quark sample has leading peak around top mass, subleading peak around W mass

Cut on jet masses

HZ signal



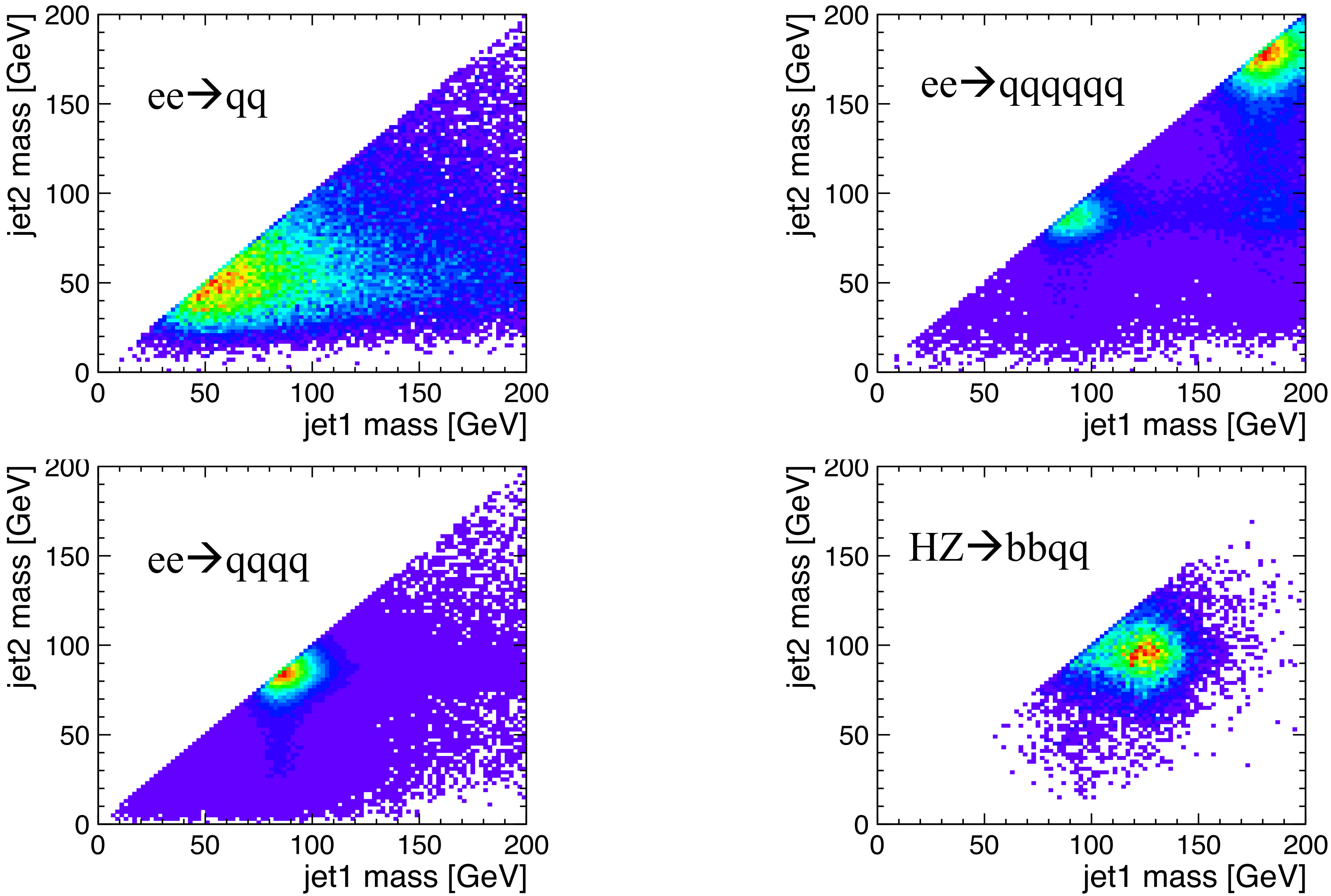


Four quark

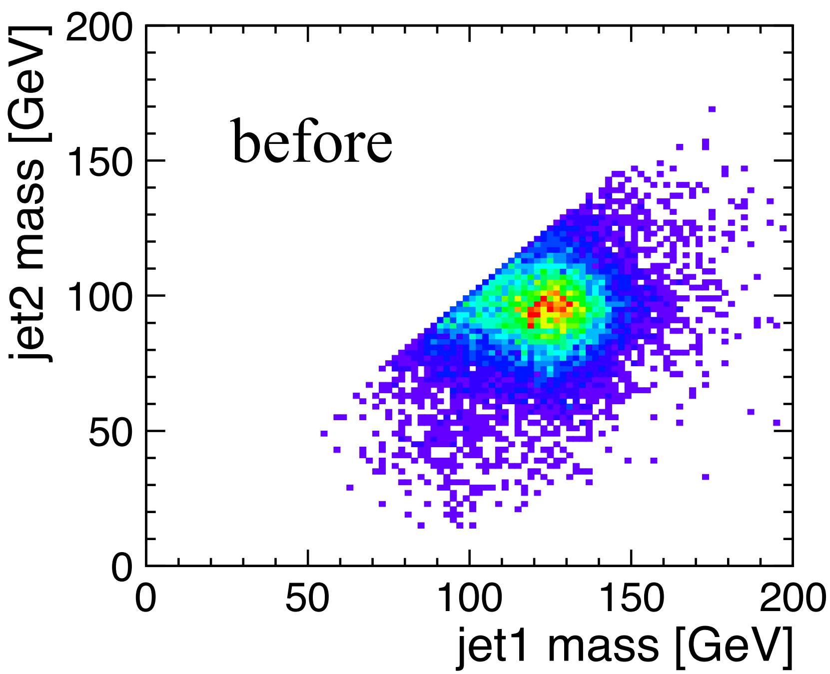
Four quark sample has narrow peaks around W and Z masses → sizeable tails to higher masses

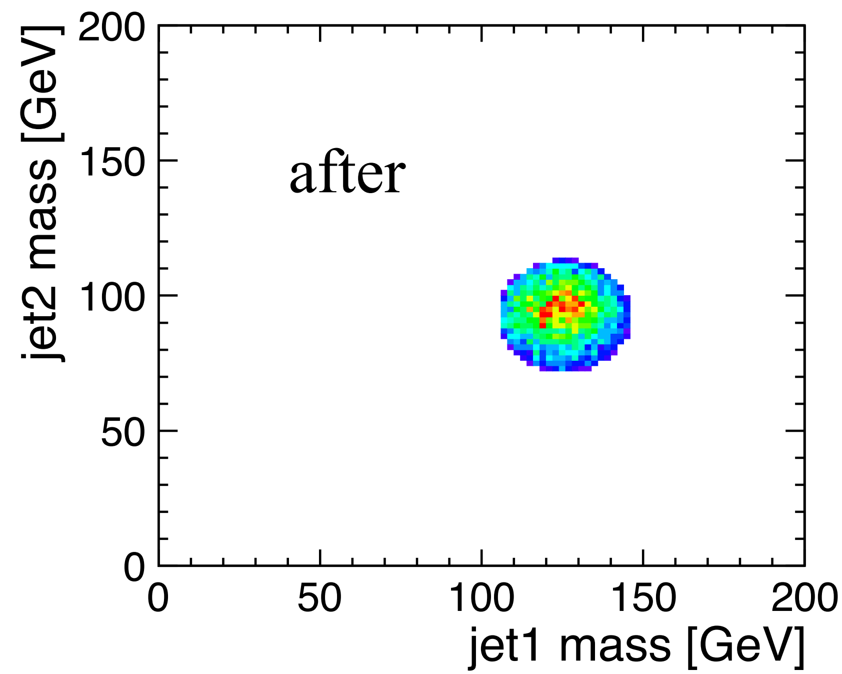
Apply two dimensional ellipse cut on jet1-jet2 mass plane around Z and H mass with windows of $\Delta m(j1) = \Delta m(j2) = 20$ GeV
→ Cuts more on the lower tail of m(j1), needed to reject qqqq background

Background jet 1 mass vs jet2 mass



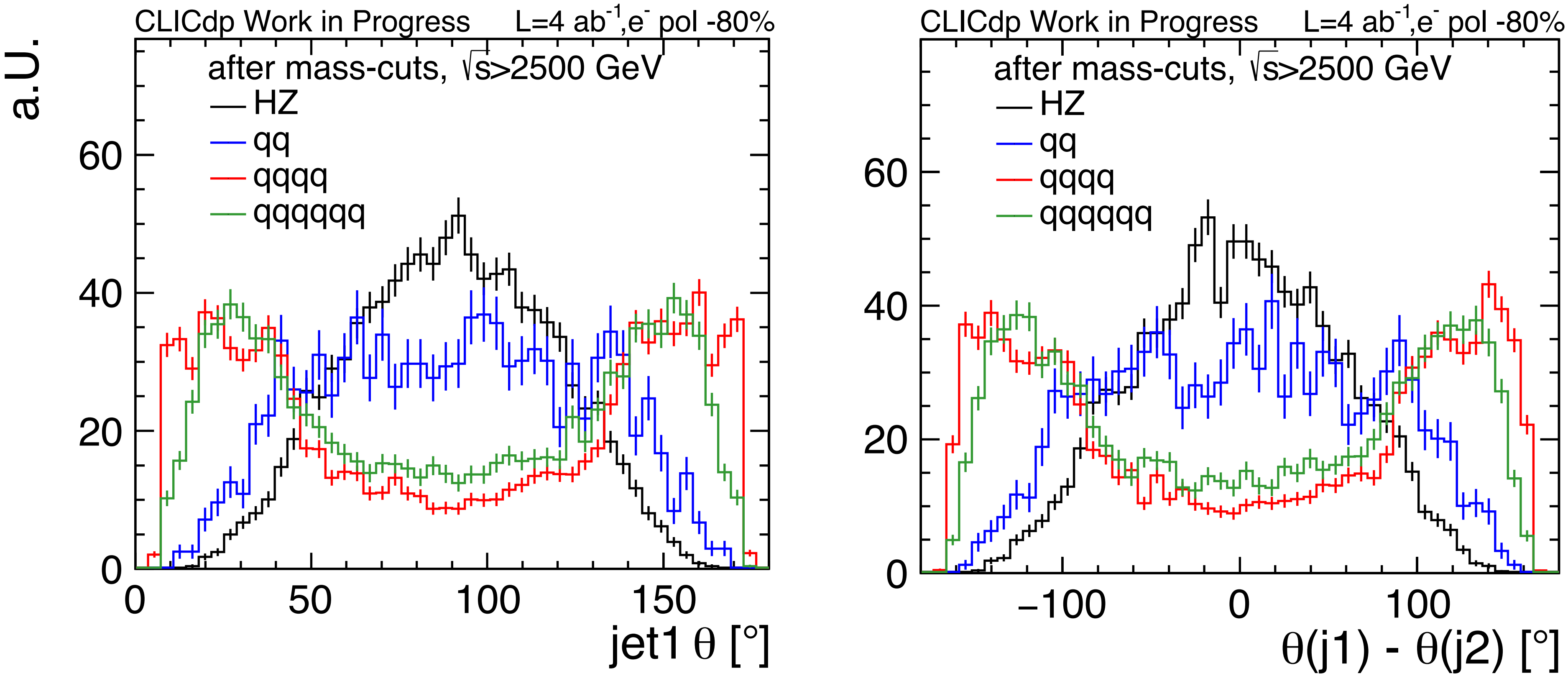
Mass Cuts for $\sqrt{s}_{reco}$>2500 GeV





Process	$\sqrt{s}_{reco}$>2500 GeV e^- pol -80 %, L=4 ab^{-1}	Efficiency In %	$\sqrt{s}_{reco}$>2500 GeV e^- pol +80 %, L=1 ab^{-1}	Efficiency In %
HZ → bb qq	1020	61	179	62
ee→qq	8020	4.9	860	4.8
ee→qqqq	20399	8.5	771	9.5
ee→qqqqqq	1080	3.4	49	1.6

Mass and Theta Cuts for $\sqrt{s}_{reco}$>2500 GeV



For four and six quark samples jets forward peaked, jets typically in opposite hemispheres

Keep events with Δθ (j1,j2)<100° and 20°< θ(j1)<160°

→ Cut could be potentially tightened

Theta on top of mass cuts for $\sqrt{s_{reco}}>2500$ GeV

Process	$\sqrt{s_{reco}}>2500$ GeV e^- pol -80 %, L=4 ab^{-1}	Total (to prev cut) Efficiency In %	$\sqrt{s_{reco}}>2500$ GeV e^- pol +80 %, L=1 ab^{-1}
HZ → bb qq	956	57 (93)	167
ee→qq	6550	4.0 (82)	699
ee→qqqq	8215	2.7 (32)	337
ee→qqqqqq	523	1.7 (49)	24

Cuts on polar angle keeps most of the HZ and di-quark dataset, removes significant parts of remaining four and six quark events

BTagging procedure

Use particles clustered in both large jets used as input for vertex reconstruction and flavour tagging procedure

- Jet Refiner running on jet particles, clustering the input in two jets, VLC7
- Each fat jet has two subjets

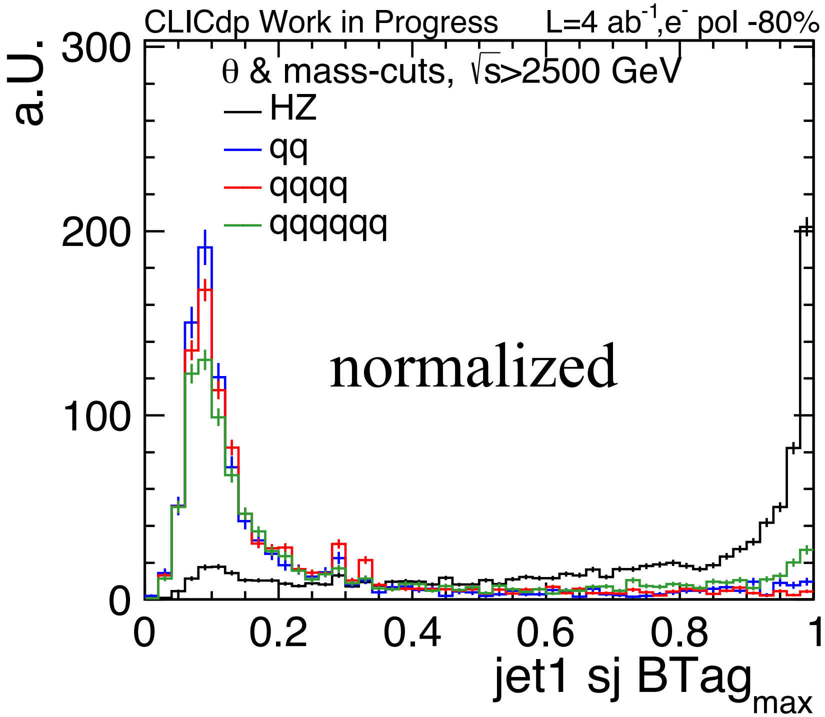
→ Since LCFIPlus uses vertex jets as seed for jet clustering, typically splitting in refined jets considerably different from FastJet subjet declustering, e.g. after LFCIPlus leading subjet harder

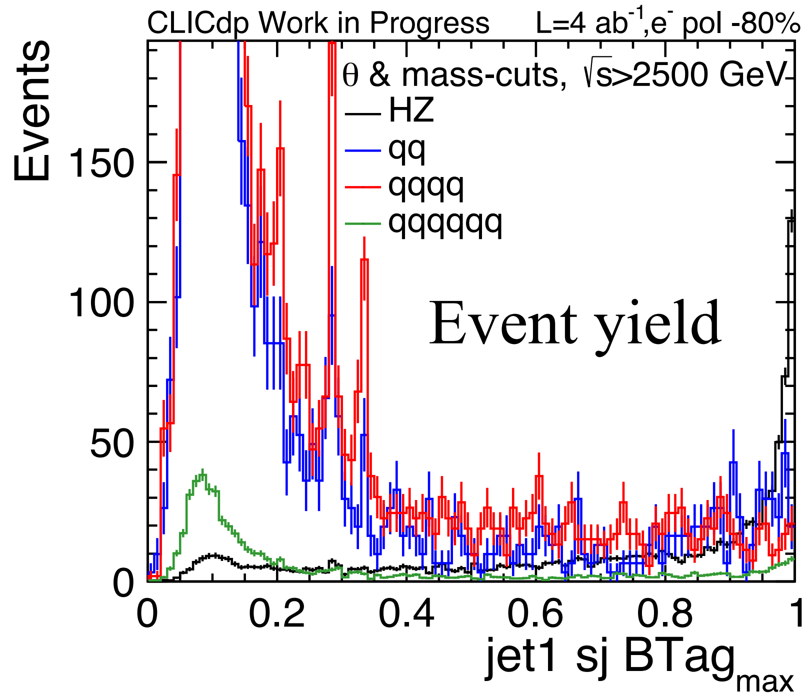
→ For HZ signal, the subjet with the largest BTag belongs to the more massive jet

→ For background subjet with largest BTag belongs to both jets

⇒ Cut on largest BTag value of subjets assigned to the leading fat jet BTag>0.9

After mass, theta and Btag>0.9 cuts





Process	$\sqrt{s}_{reco}$>2500 GeV e^- pol -80 %, L=4 ab^{-1}	Total (to prev cut) Efficiency In %	$\sqrt{s}_{reco}$>2500 GeV e^- pol +80 %, L=1 ab^{-1}
HZ → bb qq	407	24.5 (42.7)	68
ee→qq	262	0.16 (4.0)	22
ee→qqqq	147	0.11 (1.8)	10
ee→qqqqqq	42	0.13 (8.1)	2.5

Transverse Momentum Correlation Functions

Cuts on jet masses, polar angle and BTagging information reduce background significantly, di-quark dominated the background events

→ Use substructure information within fat jets, for signal two subjets within those fat jets

Transverse momentum and energy correlation functions

- Introduced in JHEP 06 (2013) 108 by A.J Larkoski, G.P Salam and J Thaler

$$\mathrm{ECF}(N,\beta) = \sum_{i_1<i_2<\ldots<i_N\in J} \left(\prod_{a=1}^{N} E_{i_a}\right)\left(\prod_{b=1}^{N-1}\prod_{c=b+1}^{N}\theta_{i_b i_c}\right)^{\beta}$$

Maybe more suitable for e^+e^- collisions

$$\mathrm{ECF}(N,\beta) = \sum_{i_1<i_2<\ldots<i_N\in J} \left(\prod_{a=1}^{N} p_{T i_a}\right)\left(\prod_{b=1}^{N-1}\prod_{c=b+1}^{N}R_{i_b i_c}\right)^{\beta}$$

Used by the LHC experiments
So far investigated these variables

$$\mathrm{ECF}(0,\beta) = 1,$$

$$\mathrm{ECF}(1,\beta) = \sum_{i\in J} p_{T i},$$

$$\mathrm{ECF}(2,\beta) = \sum_{i<j\in J} p_{T i}\, p_{T j} (R_{ij})^{\beta},$$

Both definitions available in FastJet

Energy Correlation Double Ratios

Energy correlation double ratios $C_N^{(\beta)}$ defined as

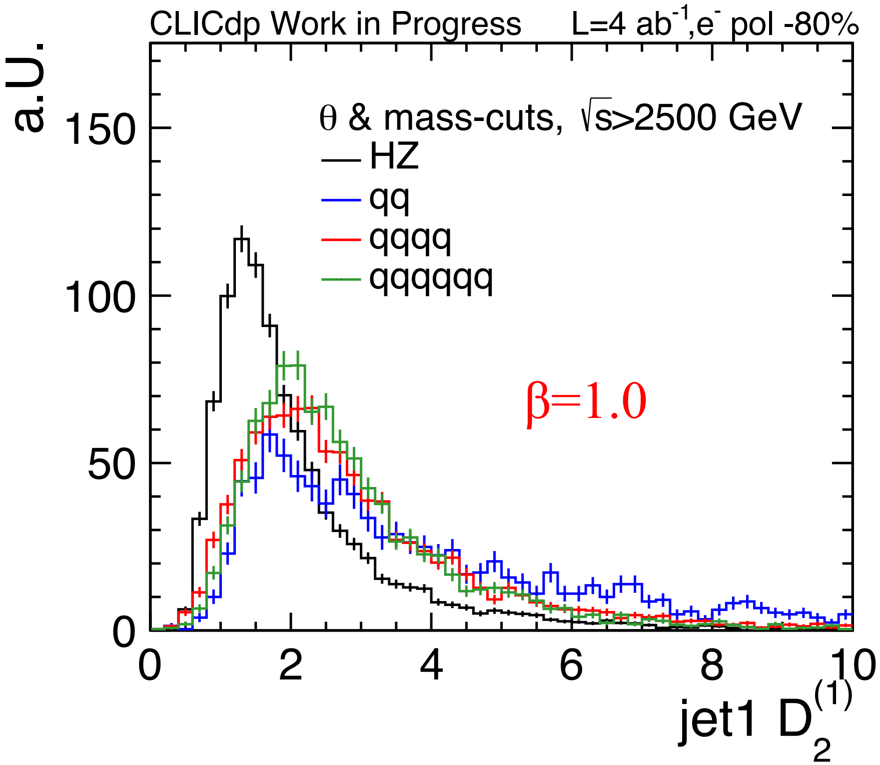
$$C_N^{(\beta)} \equiv \frac{r_N^{(\beta)}}{r_{N-1}^{(\beta)}} = \frac{\mathrm{ECF}(N+1,\beta)\,\mathrm{ECF}(N-1,\beta)}{\mathrm{ECF}(N,\beta)^2}$$

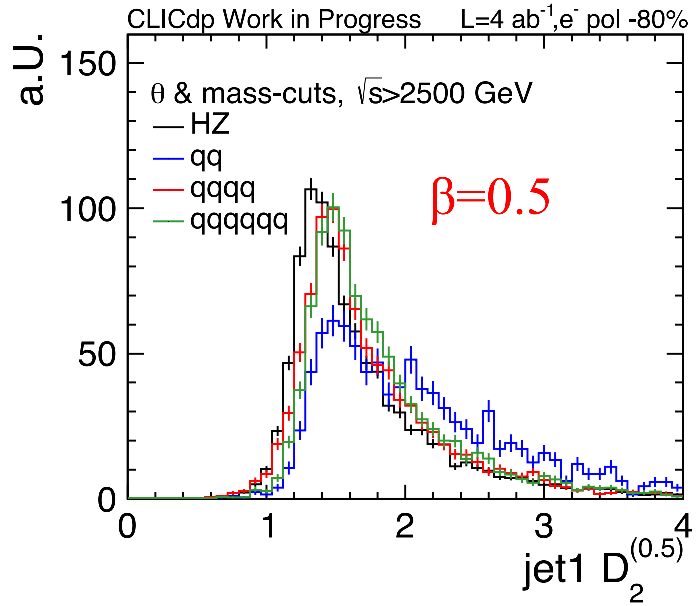
$D_2^{(\beta)}$ introduced in JHEP 12 (2014) 009 (https://arxiv.org/abs/1409.6298) by A.J Larkoski, I. Moult and D. Neill as

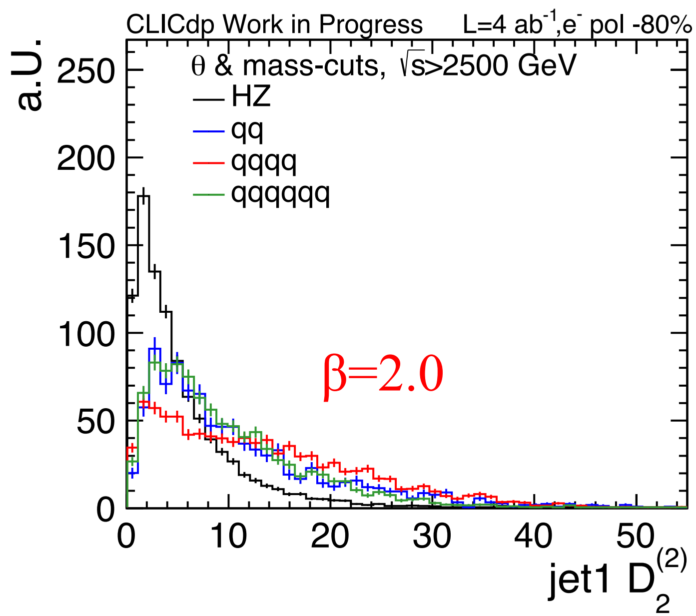
$D_2^{(\beta)} = \mathrm{EFC}(3, \beta)/\mathrm{EFC}(2, \beta)^3$
$C_2^{(\beta)} = \mathrm{EFC}(3, \beta)*\mathrm{EFC}(1, \beta)/\mathrm{EFC}(2, \beta)^2$

Leading jet Energy Correlation Double Ratio $D_2^{(\beta)}$

Studied three different beta values: 0.5, 1 and 2, variables least discriminant using β=0.5



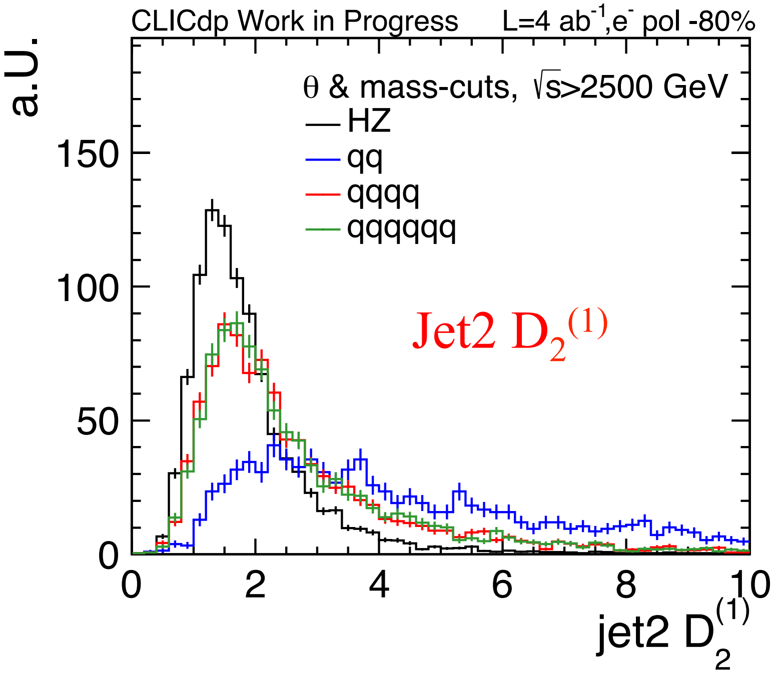


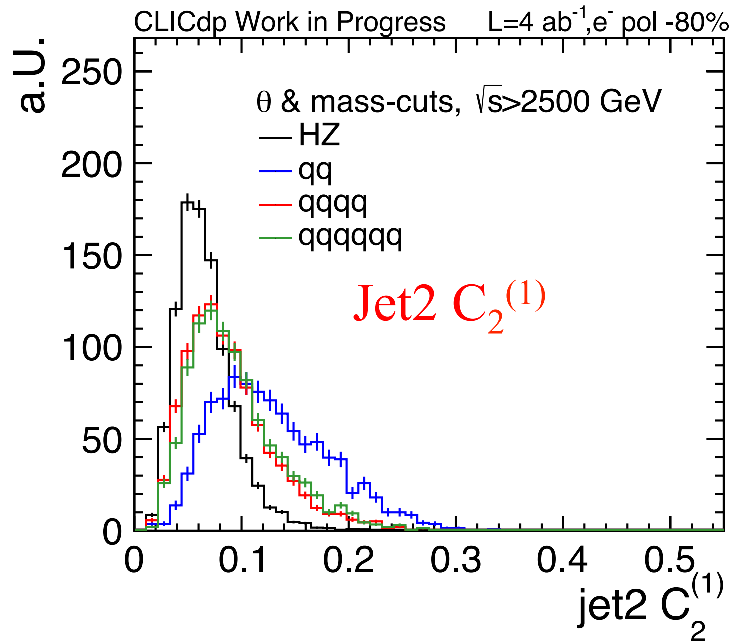


Use β=1.0 for now

→for HZ dataset distribution peaked at lower values

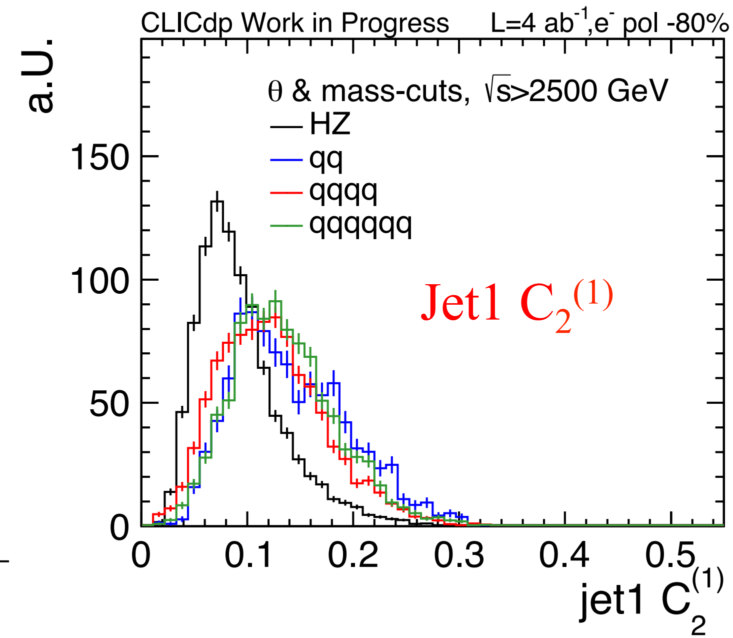
Energy Correlation Double Ratios $C_2^{(1)}$ and $D_2^{(1)}$



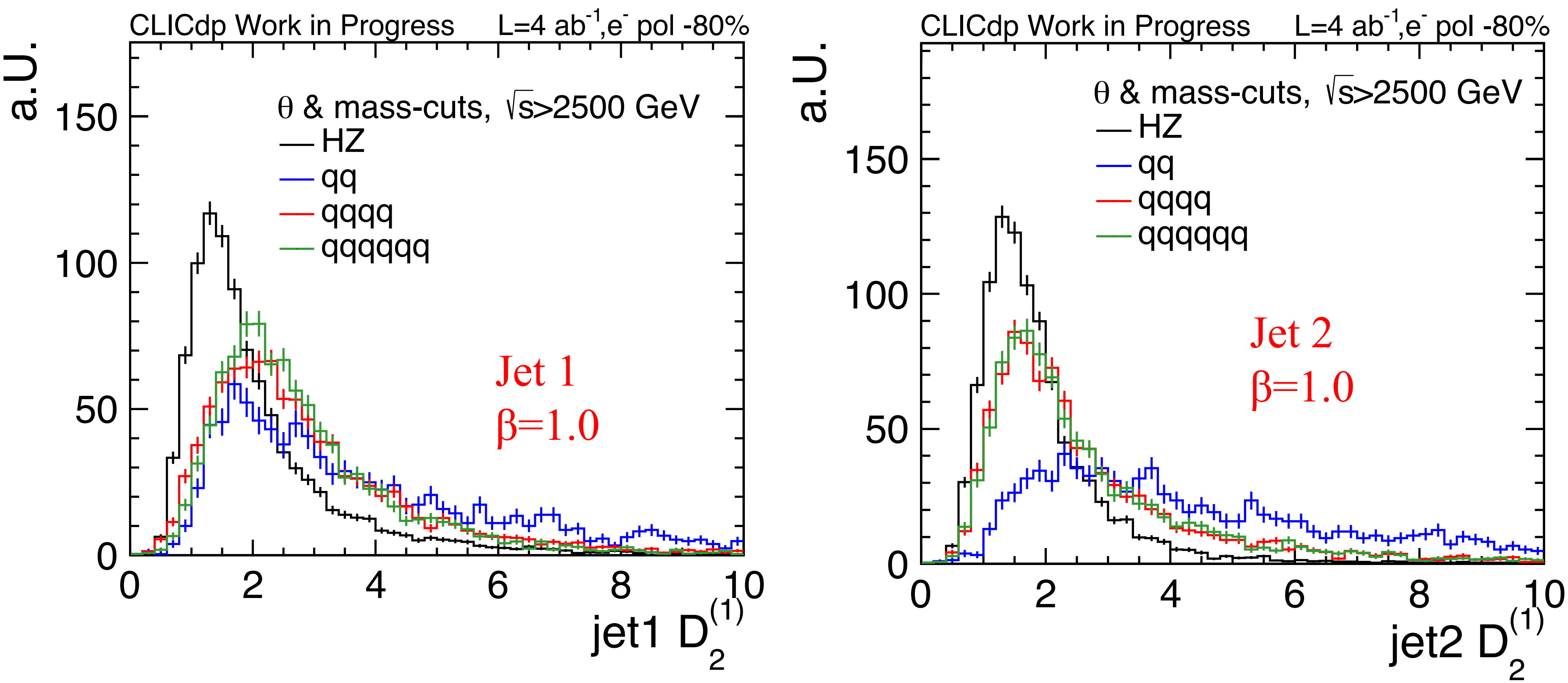


Jet Energy correlation ratios of both jets additional handle to discriminate between signal and backgrounds

→ D_2 and C_2 quite correlated, choose D_2



Energy Correlation Ratio $D_2^{(1)}$

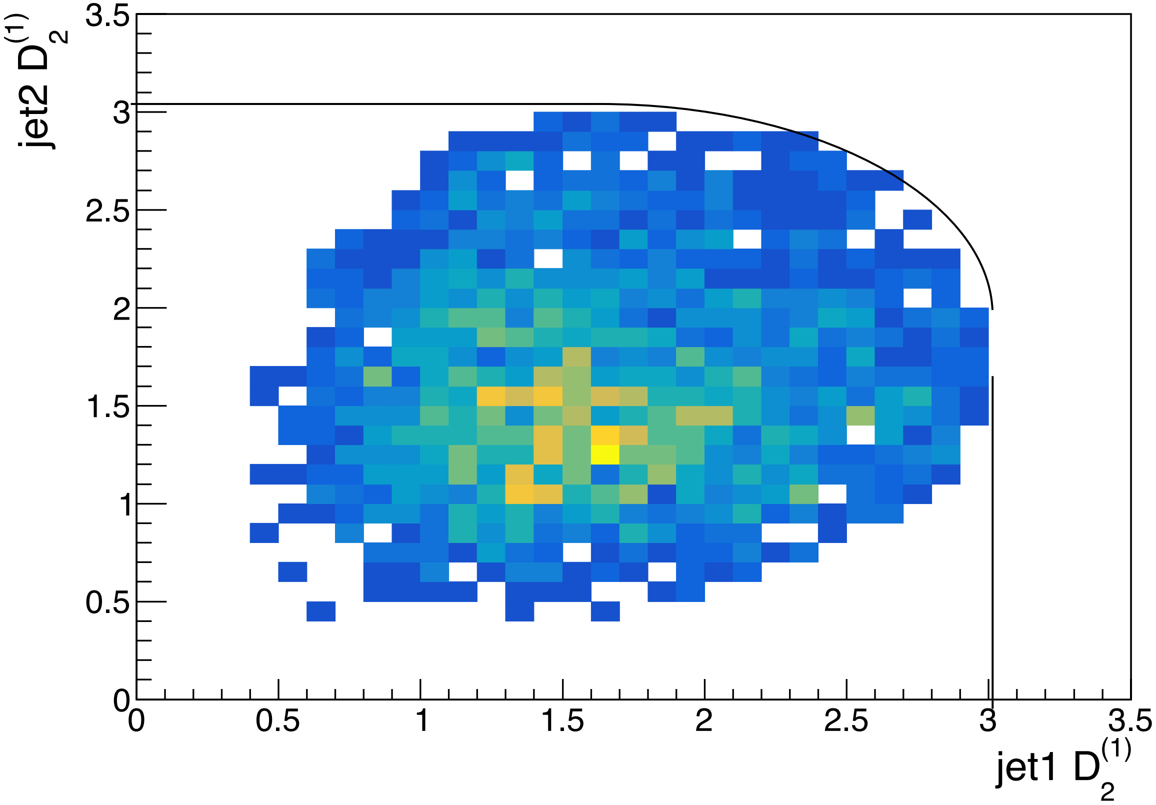


Most powerful against ee→qq, for four and six quark backgrounds leading jet D_2 more discriminant

→ Cut on 2D plane on upper end

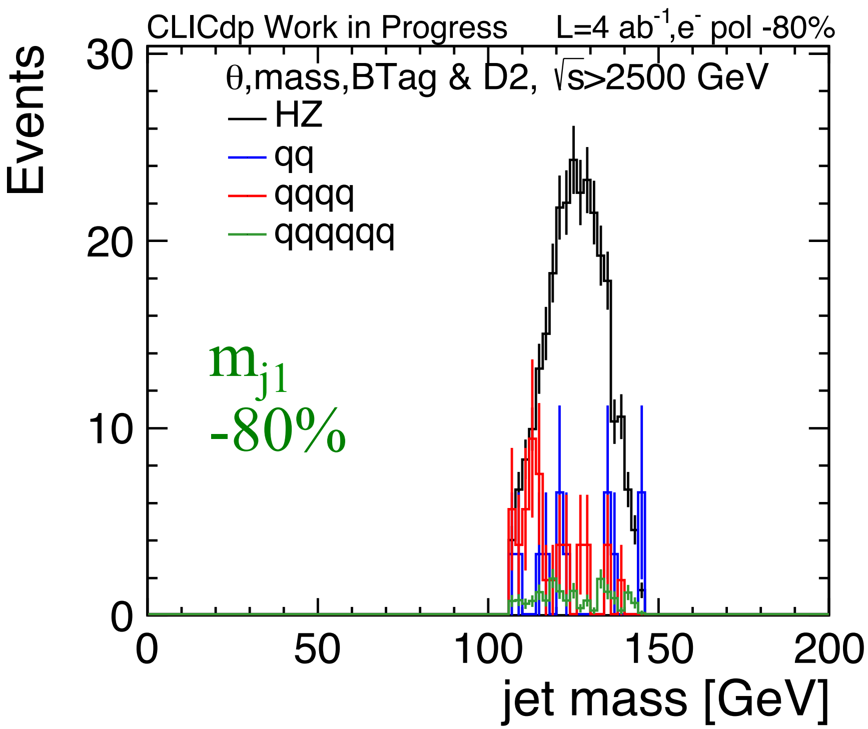
Energy Correlation 2D cut selection on HZ

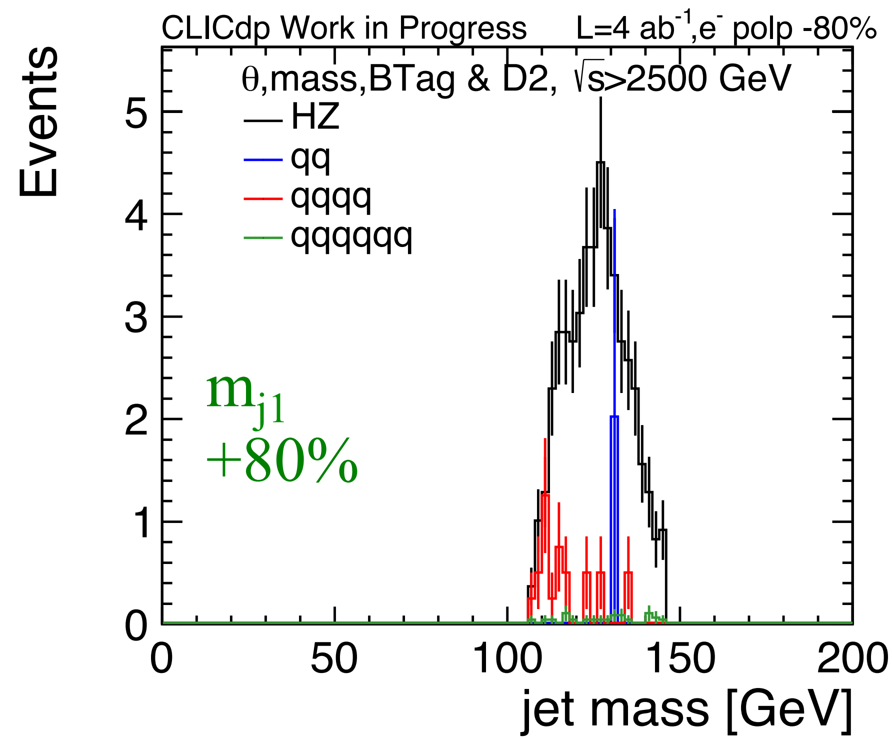
Cut on 2 dimensional distributions of jet substructure variables



Selection keeps most of the signal, but removes the tails of all backgrounds

Final cut flow default example

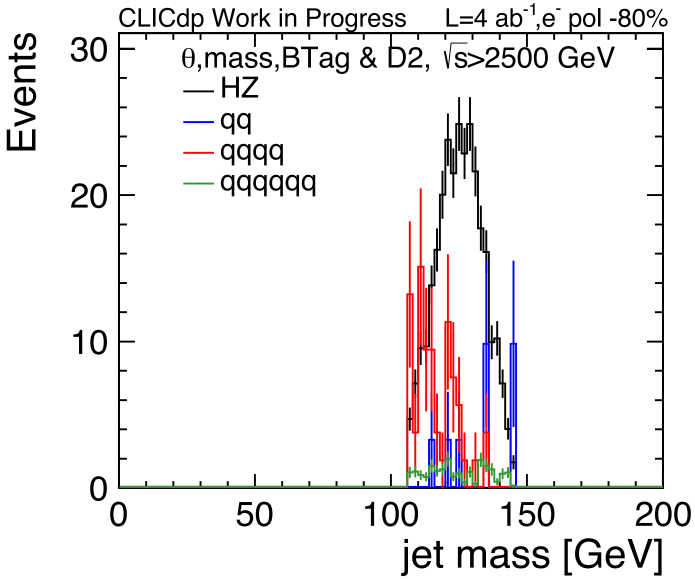


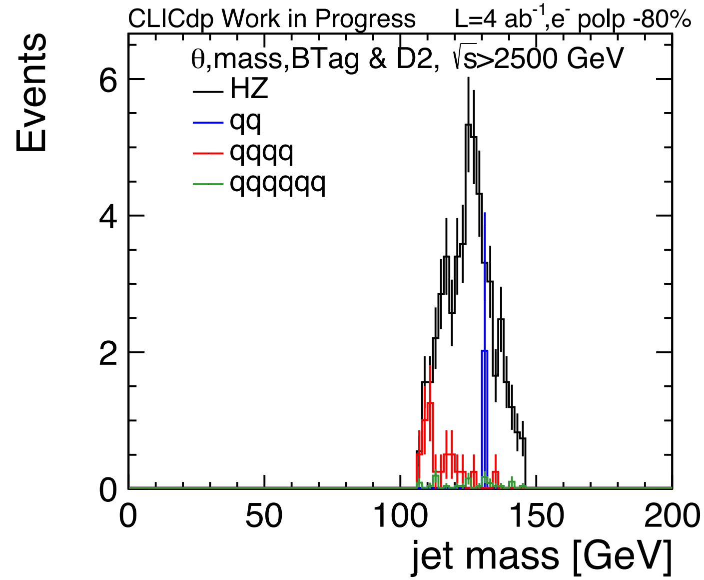


Process	$\sqrt{s_{reco}}$>2500 GeV e^- pol -80 %, L=4 ab^{-1}	Total (to prev cut) Efficiency In %	$\sqrt{s_{reco}}$>2500 GeV e^- pol +80 %, L=1 ab^{-1}
HZ → bb qq	282	17 (69)	48
ee→qq	39	0.02 (15)	2.0
ee→qqqq	57	0.02 (38)	5.0
ee→qqqqqq	18	0.06 (43)	0.9

Alternative Cut Flow

Keep jet mass and polar angle cuts, relax cut on Btag>0.8, tighten cut on D2 → increase in ee→qqqq background, slight decrease for ee→qq





Process	$\sqrt{s}_{reco}$>2500 GeV e^- pol -80 %, L=4 ab^{-1}	Total (to prev cut) Efficiency In %	$\sqrt{s}_{reco}$>2500 GeV e^- pol +80 %, L=1 ab^{-1}
HZ → bb qq	288	17.2 (54)	52
ee→qq	30	0.018 (6.6)	2.0
ee→qqqq	90	0.037 (25)	5.3
ee→qqqqqq	21	0.07 (31)	1.1

Open H selection, default cuts

So far always select H→bb, based on MC truth. In case we remove this cut based on MC information, the number of events increases from 282 to 288 for the dataset with negative polarization.
The amount of H→bb within that sample is 98 %

For the +80 % polarization sample H→bb contributes to 98 % of the final selected events as well

Process	$\sqrt{s}_{reco}$>2500 GeV e^- pol -80 %, L=4 ab^{-1}	Total (to prev cut) Efficiency In %	$\sqrt{s}_{reco}$>2500 GeV e^- pol +80 %, L=1 ab^{-1}
HZ → H qq	288	17 (69) on H→bb	49
ee→qq	39	0.02 (15)	2.0
ee→qqqq	57	0.02 (38)	5.0
ee→qqqqqq	18	0.06 (43)	0.9

Without any signal selection the jet with the larger mass is matched to the H in around 85 % of cases. After the final selection this is the case in more than 98 %

Conclusions

First look at HZ signal with H→bb and backgrounds from ee→qq, ee→qqqq and ee→qqqqqq

Discrimination between signal and backgrounds by cuts on

- Jet masses against all three backgrounds
- Jet polar angle (rejects mainly ee→qqqq and ee→qqqqqq)
- Flavour identification, using BTagging on the leading jets
- Energy correlations addressing the differences in jet substructure → particularly helpful to discriminate against ee→qq

→ Further tuning limited by available statistics in ee→qq and ee→qqqq datasets

→ Signal selection almost exclusively selects H→bb events, more massive jet is matched to H with over 98 %

Next step:

Use most discriminating variables in BDT

MC Production has been started to increase statistics of di-quark and four quark