Improved jet clustering algorithm with vertex information for multi-b final states

Taikan Suehara,

Tomohiko Tanabe, Satoru Yamashita

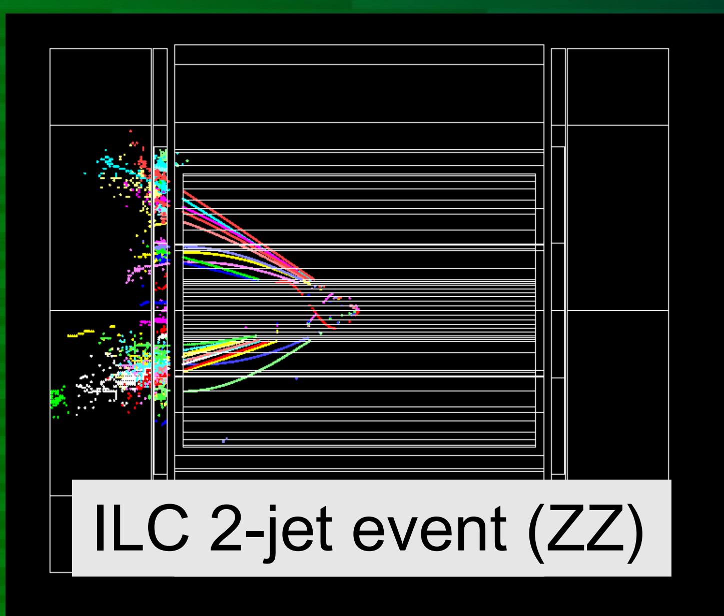
(The Univ. of Tokyo)

TeV physics in ILC/LHC/…

- Higgs → bb (2-jet), WW (up to 4-jet)
- SUSY → ex. missing + W (2-jet): 4-jet in pair production
- exotics

Final states with many jets (4/6/8/…)

Jet clustering is a major performance driver



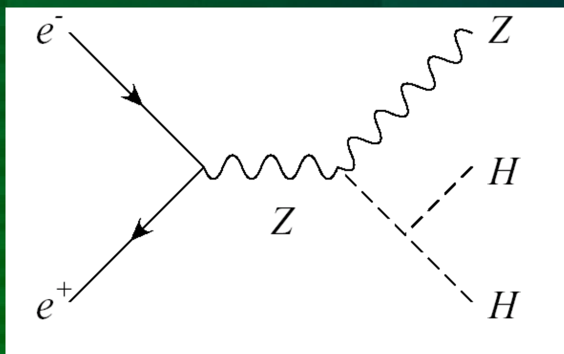
ILC 2-jet event (ZZ)

ILC 6-jet event (ZHH)

Example: ZHH in ILC

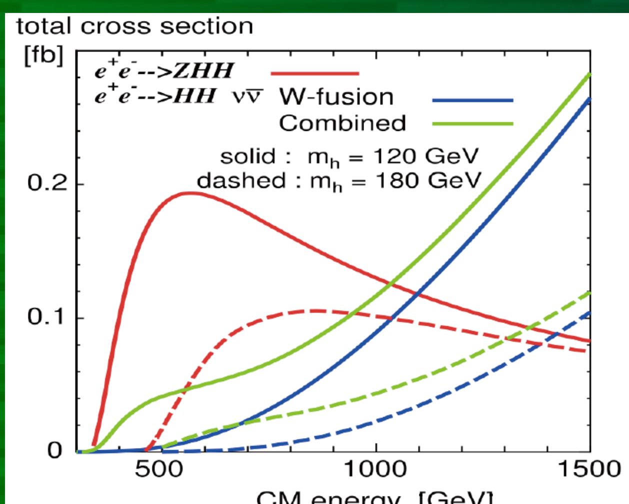
HHH coupling: a key to prove Higgs mechanism

m_H=120 GeV



Double Higgs-strahlung:
largest xsec around 500 GeV

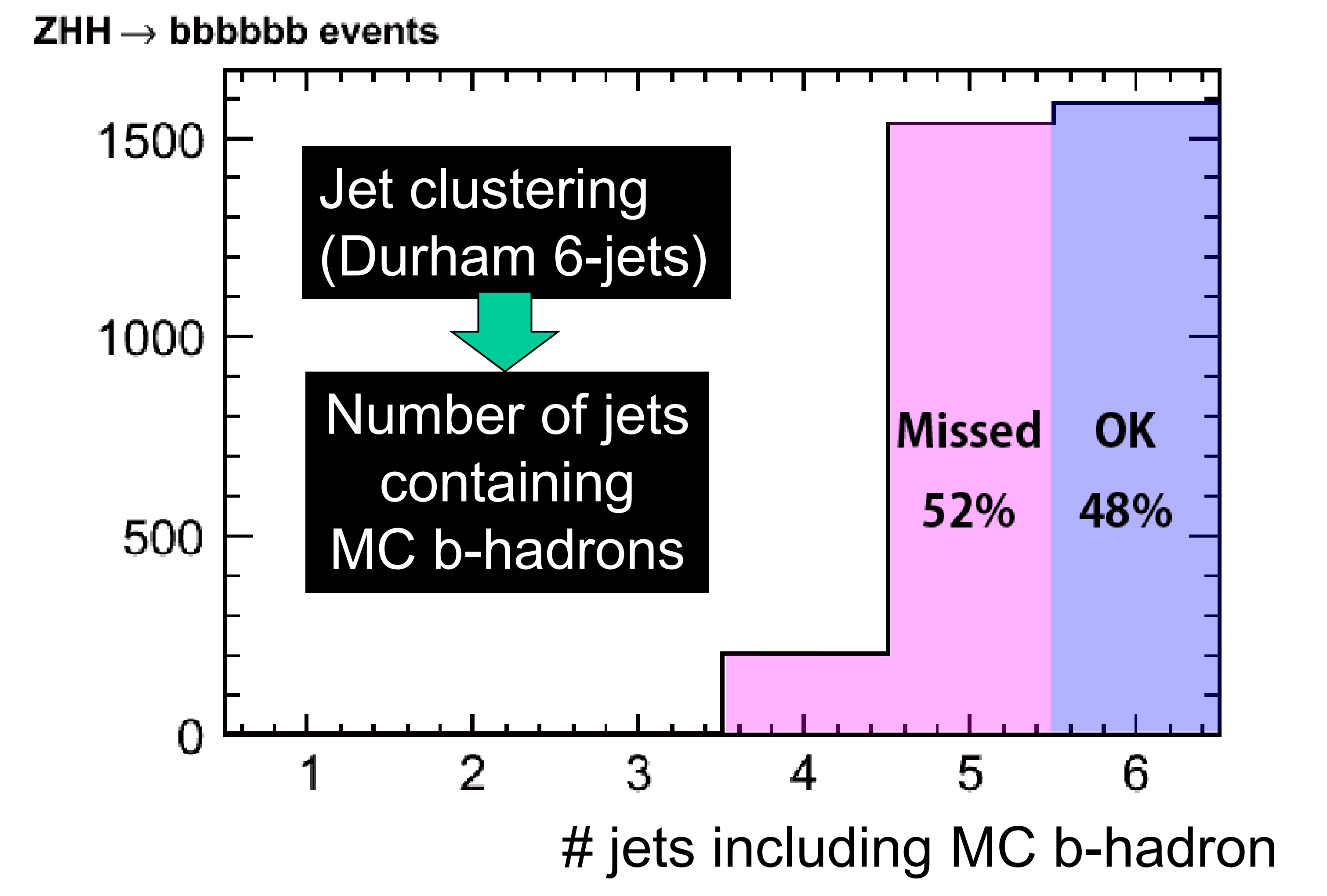
Decay mode	BR.	# events in 1 ab^{-1}
qqbbbb	32%	73
ννbbbb	9%	21
qqbbWW*->qqbbqqqq	6%	14
llbbbb	4%	10
qqbbWW*->qqbbqqlν	3%	7
qqbbWW*->qqbblνqq	3%	7
others	43%	97
tt -> bbqqqq		~400,000



Extremely small cross section of 0.2fb
Background (esp. top-pair) must be very strongly suppressed

Excellent b-tag in 6-jet environment

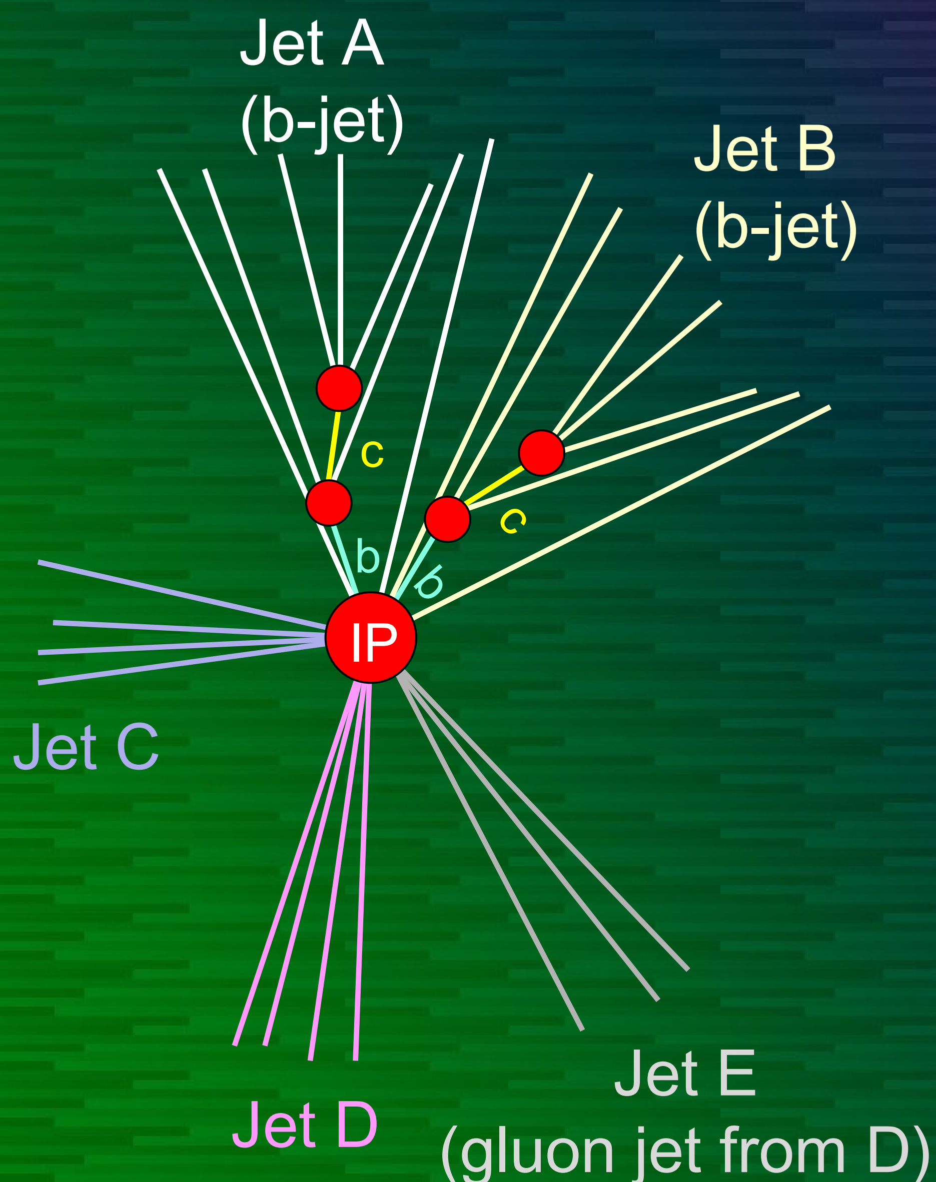
‘# of b jets’ in ZHH



# of b-jets is reduced due to mis-jet-clustering

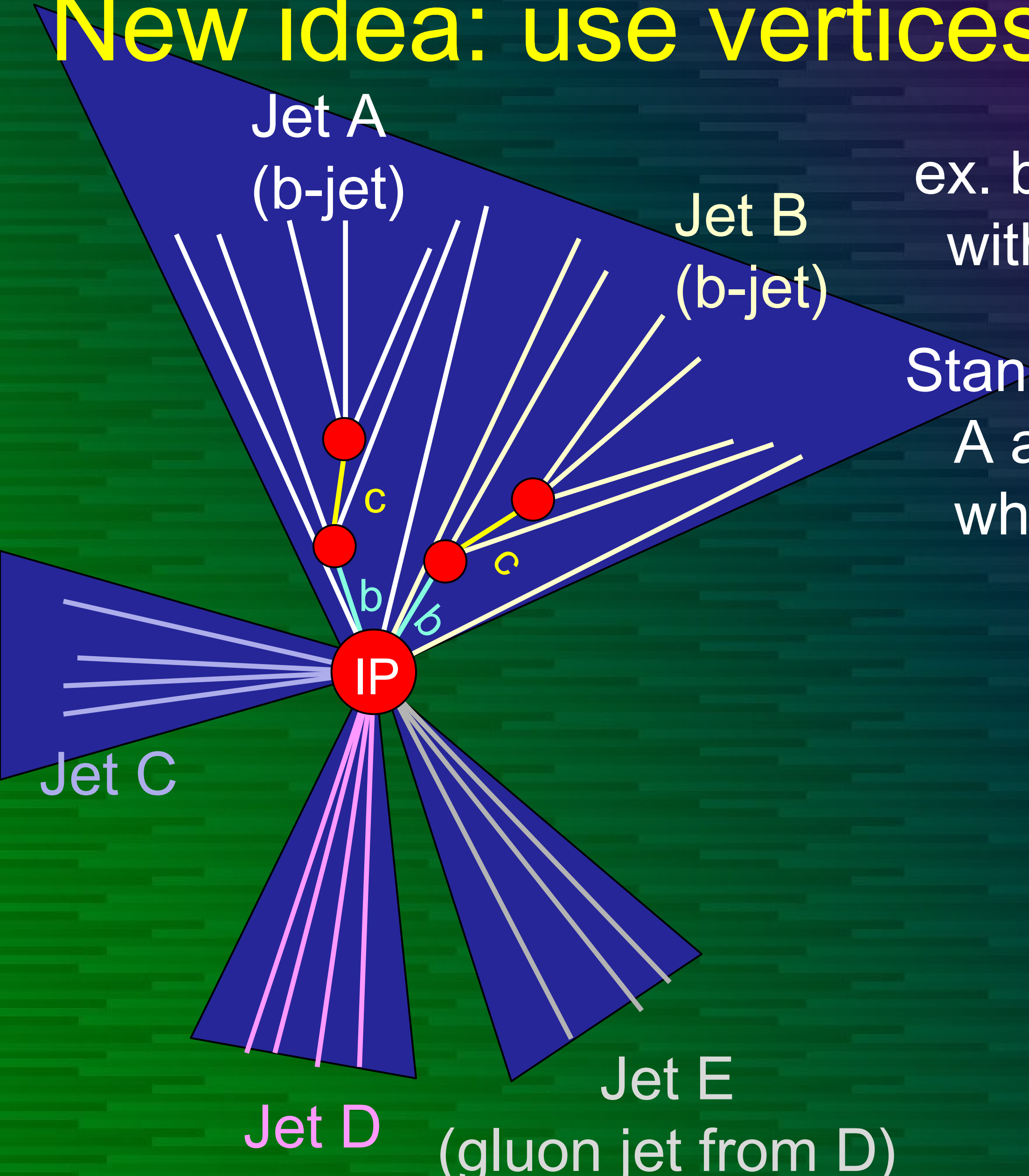
→ Major problem in counting b-quarks

New idea: use vertices in jet clustering



ex. bbqq with a hard gluon
with 4-jet configuration

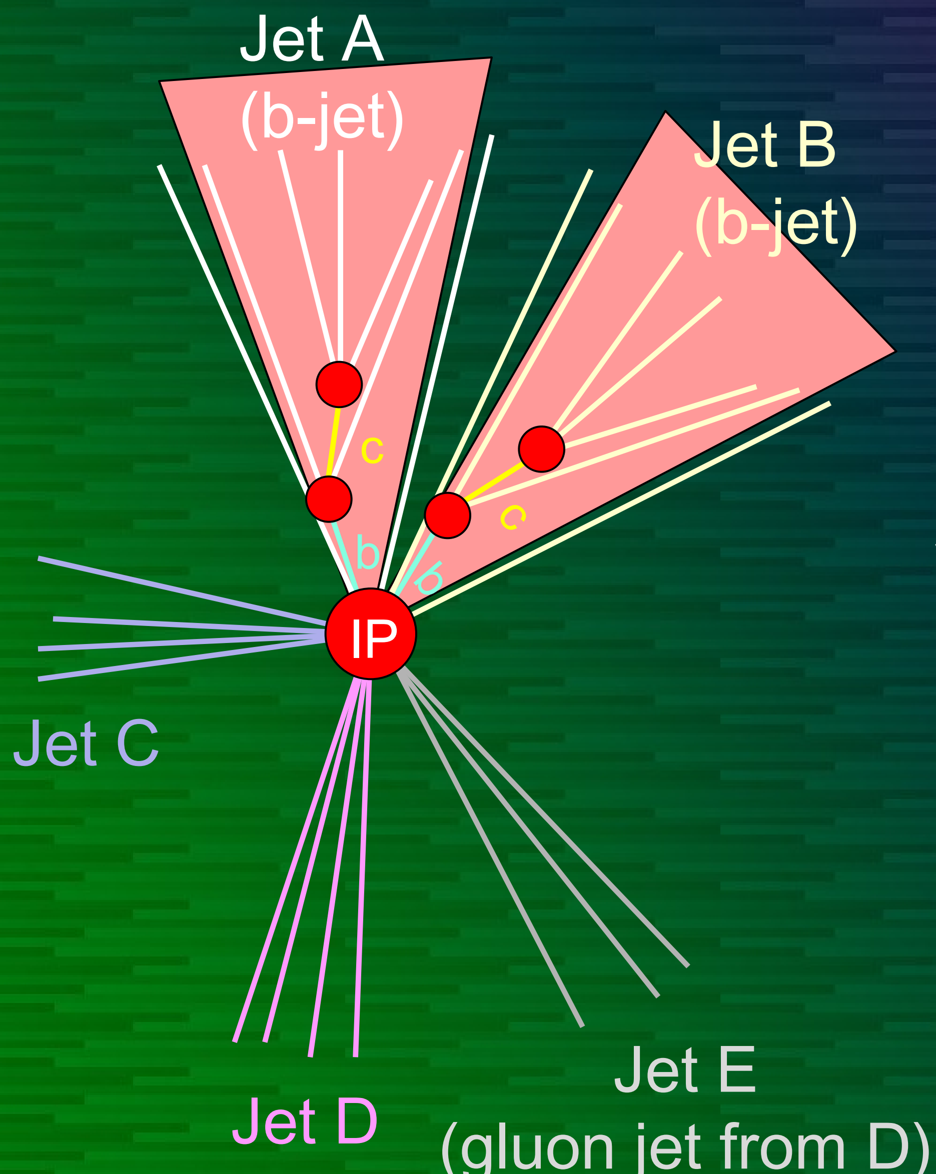
New idea: use vertices in jet clustering



ex. bbqq with a hard gluon
with 4-jet configuration

Standard jet clustering:
A and B might be combined
while E separated

New idea: use vertices in jet clustering

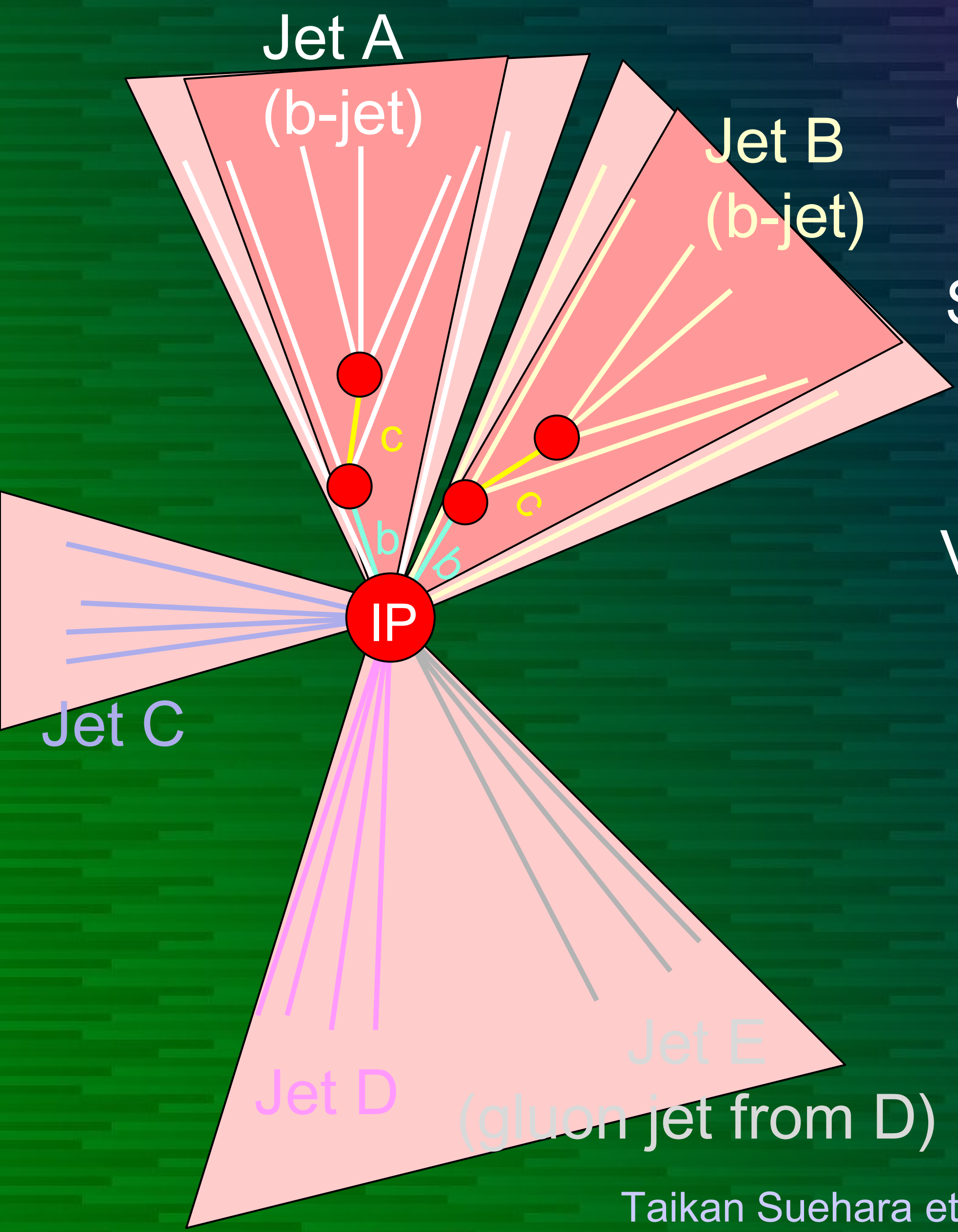


ex. bbqq with a hard gluon
with 4-jet configuration

Standard jet clustering:
A and B might be combined
while E separated

Vertex clustering:
Two b-jets can be separated
with vertex information

New idea: use vertices in jet clustering

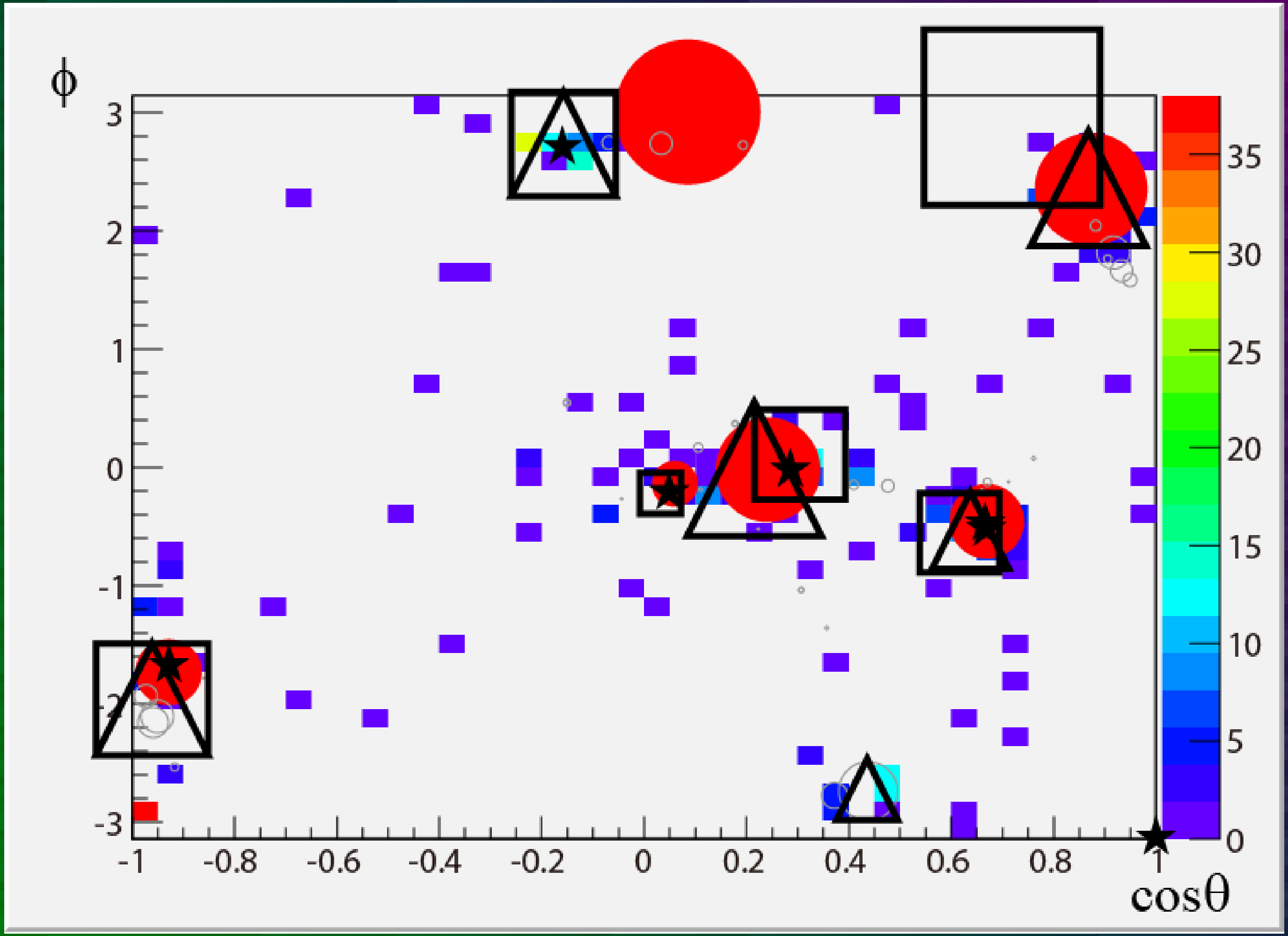


ex. bbqq with a hard gluon
with 4-jet configuration

Standard jet clustering:
A and B might be combined
while E separated

Vertex clustering:
Two b-jets can be separated
with vertex information

Sample in MC (2D extraction)



Red circle: MC b (before gluon emission), Star: vertex found
Triangle: Durham jets, Square: Our jet clustering

Details (a bit) of our method

1. Vertex finder
2. Secondary Muon ID
3. Vertex combination
4. Jet clustering using vertices
5. Jet flavor tagging

1. A new build-up vertex finder

High-purity vertex finder is critical in this method

Fake vertices significantly degrade performance!

Problem:

Usual vertex finders assume jet clustering is correct
… and use the jet direction to improve purity

We cannot use jet direction since we search for vertices first

Original vertex finder

- Build-up method (pairing tracks -> association)
- Not include new idea; main effort on optimization
  - mass based cuts, track combination order, etc.

For implementation details, see backup slides

2. Secondary muon ID

- Secondary muons can also be used to identify heavy-flavor jets
  - Secondary electrons are currently not used (because of non-trivial separation from pions)
- Secondary muon criteria:
  - Require hit in muon detector
  - Impact parameter > 5 σ, < 5 mm
  - ECAL, HCAL energy deposit
- Secondary muons are treated similarly to vertices (with muon direction as vtx. direction)

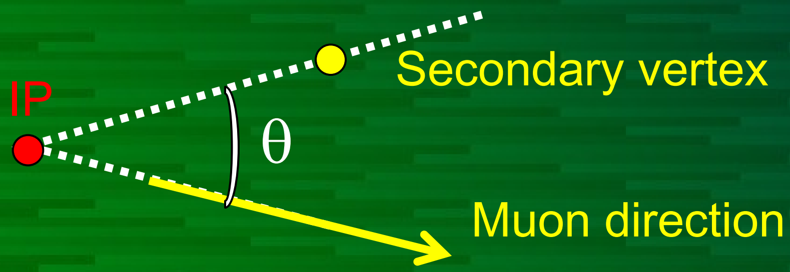
3. Vertex combination

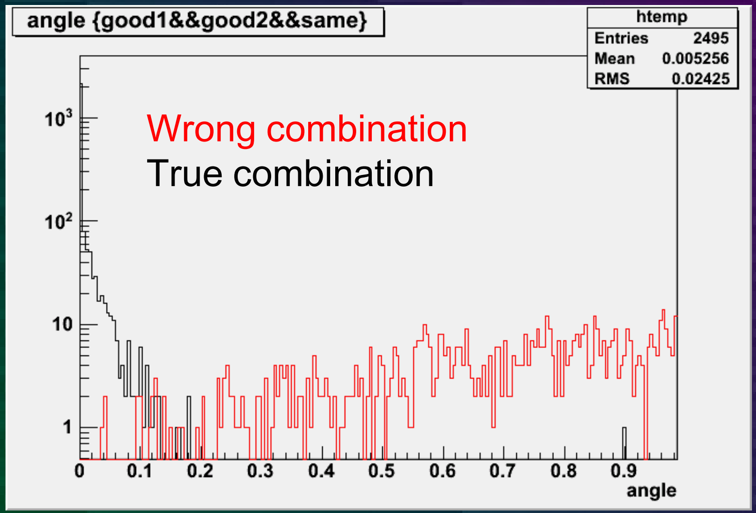
Our jet clustering strategy:

- Identify heavy flavor jets using vertices
- Separate heavy flavor jets using vertices

→ b & c vertices must be combined
others must remain separated

- Simple combination criteria
  - Opening angle to IP < 0.2 rad.
  - In muon case; < 0.3 rad.

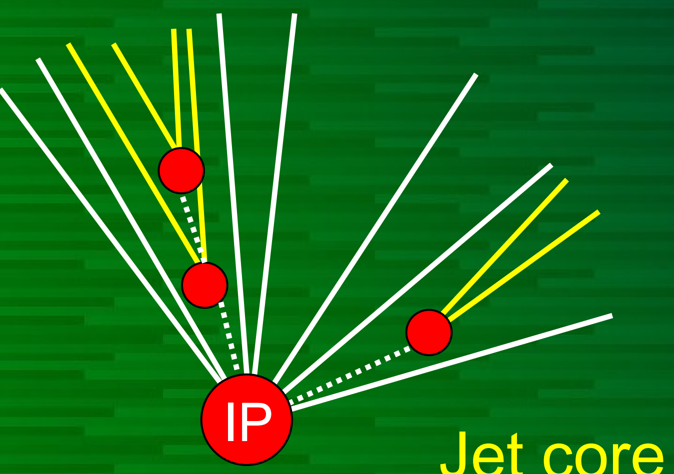




4. Jet clustering using vertices

1. Combined vertices are listed as ‘jet core’
2. All particles within 0.2 rad. to the jet core are associated to the core
3. All associated jet cores and residual particles are associated with Durham y criteria

$$y = \frac{2\min(E_1, E_2)^2(1-\cos\theta_{ij})}{Q^2}$$

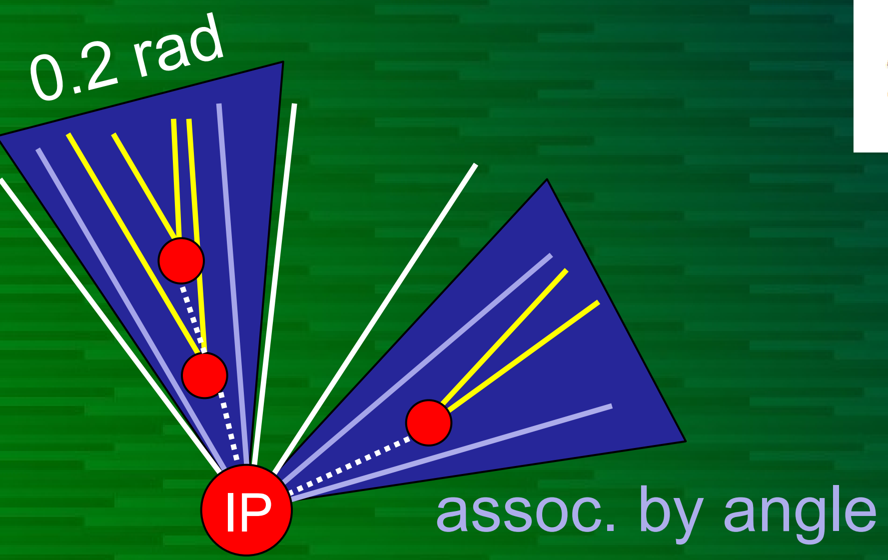


* Jet cores with vertices are never combined to each other (y value is set to +inf.)

4. Jet clustering using vertices

1. Combined vertices are listed as 'jet core'
2. All particles within 0.2 rad. to the jet core are associated to the core
3. All associated jet cores and residual particles are associated with Durham y criteria

$$y = \frac{2\min(E_1, E_2)^2(1-\cos\theta_{ij})}{Q^2}$$

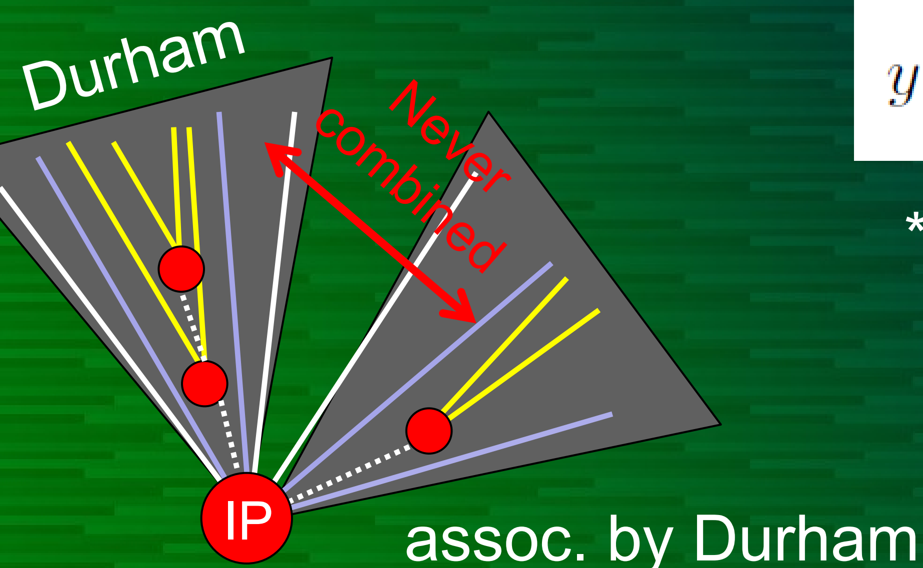


* Jet cores with vertices are never combined to each other (y value is set to +inf.)

4. Jet clustering using vertices

1. Combined vertices are listed as ‘jet core’
2. All particles within 0.2 rad. to the jet core are associated to the core
3. All associated jet cores and residual particles are associated with Durham y criteria

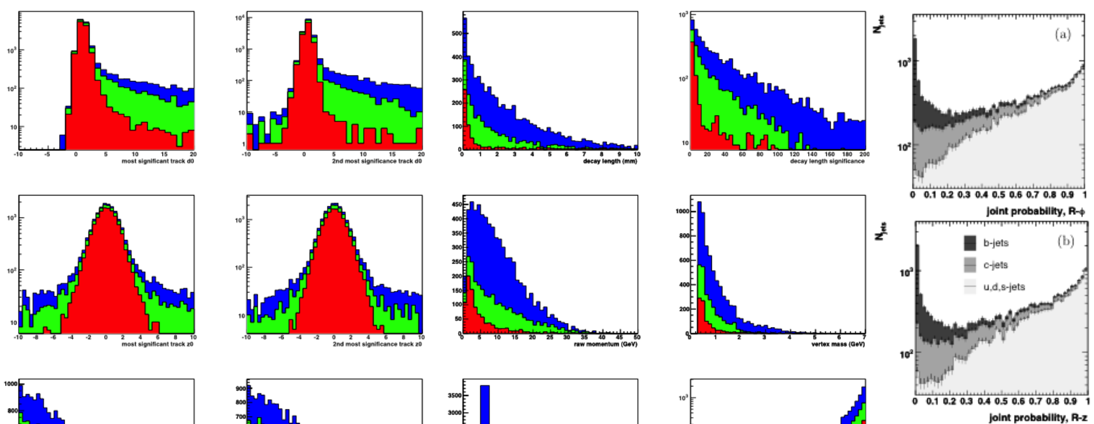
$$y = \frac{2\min(E_1, E_2)^2(1-\cos\theta_{ij})}{Q^2}$$



* Jet cores with vertices are never combined to each other (y value is set to +inf.)

5. Jet flavor tagging (LCFIVertex)

Neural-net based flavor tagging package of standard ILC analysis
(now improving also by us, but new version not available yet)



joint prob.

input variables (tracks)

input variables (vertex)

neural net

b-likeness
c-likeness
parton charge

LCFI Collaboration: NIM A 610 (573) [arxiv:0908.3019]

Performance

1. MC number of b-jets
2. MC number of tracks from b-hadrons
3. ZHH vs. top-pair b-tagging performance
4. Effect on b-tag cut in ZHH analysis

1. MC number of b-jets

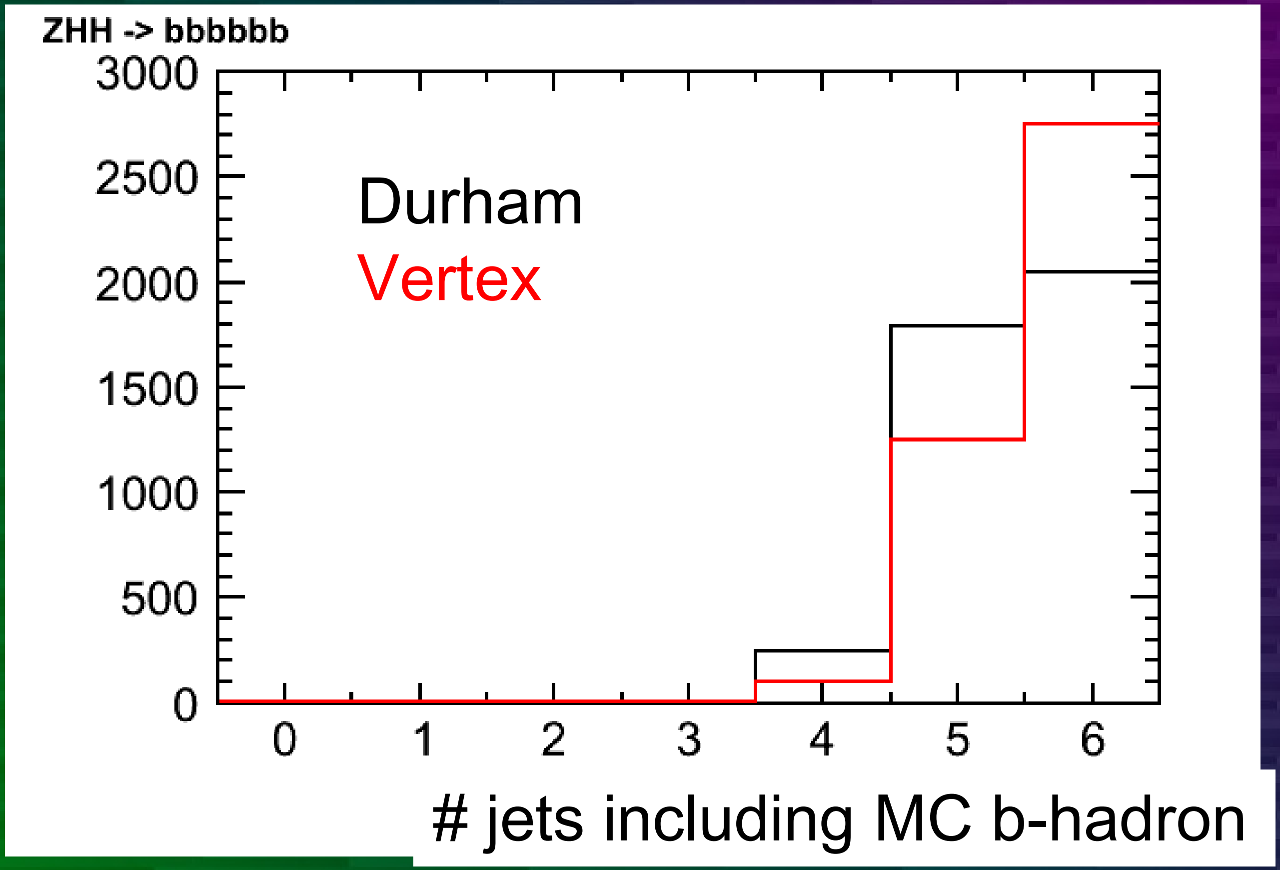
Quick view of the vertex effect

Using ZHH -> bbbbbb events (# b-jets should be 6)

Procedure:

1. Jet clustering (Durham / Vertex)
2. Listing b-hadrons (MC information)
   Listing tracks from b-hadrons
3. Associate b-hadrons to reconstructed jets
   Jet including largest # of tracks from a b-hadron is associated to the b-hadron
4. Counting jets associated to b-hadrons

1. MC number of b-jets



All jets including b – 52% -> 66%

Significant improvement!

2. MC number of tracks from b-hadrons

Effect on b-tagging at MC level

Using
ZHH -> qqHH (H->bb: 68%, Z decays to every flavor)
tt -> bbcssc (each W decays to c and s quarks)

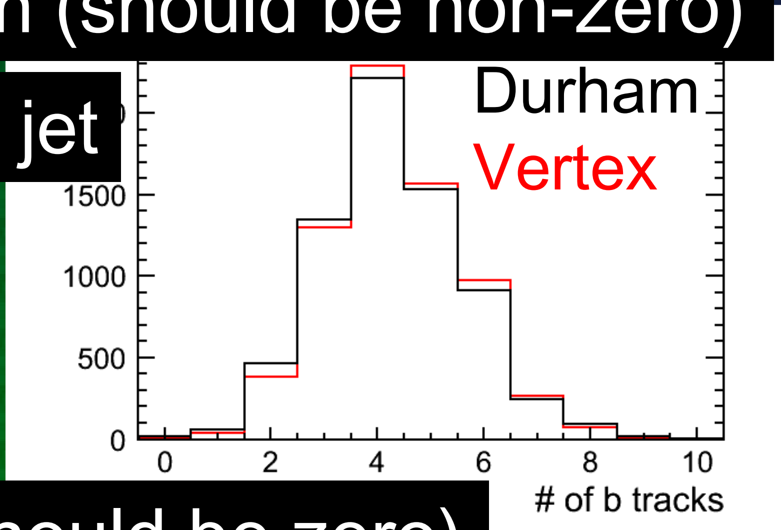
Procedure:

1. Listing b-hadron tracks with MC information
2. Counting b-hadron tracks in each jet
3. Sort jets by number of b-hadron tracks
4. See 3rd and 4th jets:
   - qqHH: #b=4; # b-hadron tracks should be > 0
   - tt: #b=2; # b-hadron tracks should be 0

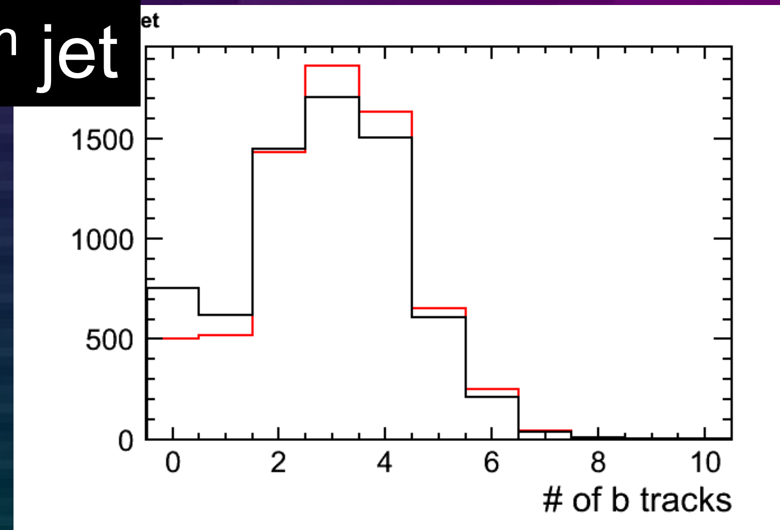
2. MC number of tracks from b-hadrons

qqhh (should be non-zero)

3rd jet

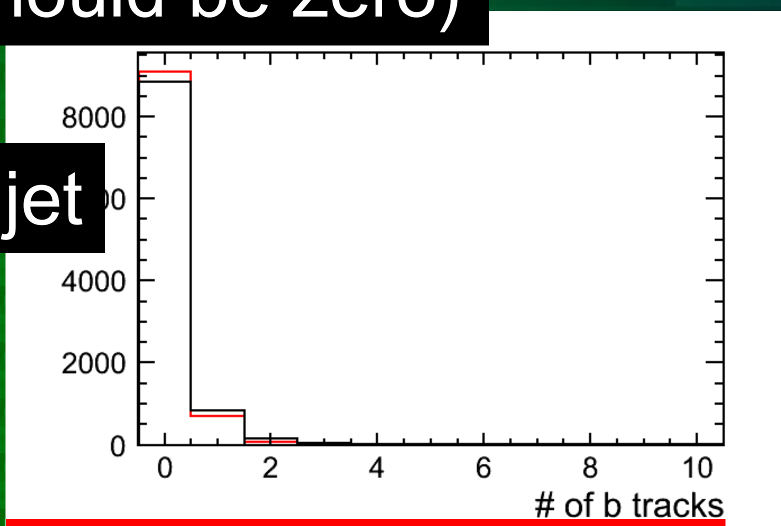


4th jet

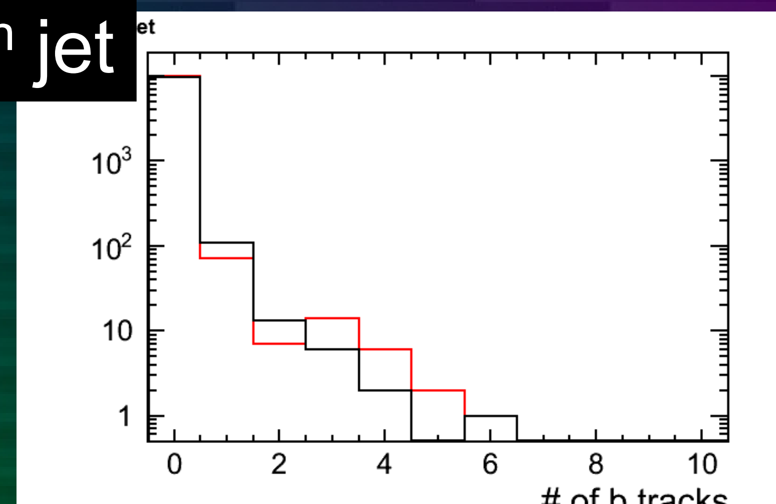


tt (should be zero)

3rd jet



4th jet



tt rejection improved

qqHH acceptance improved

3. ZHH vs. top-pair b-tagging performance

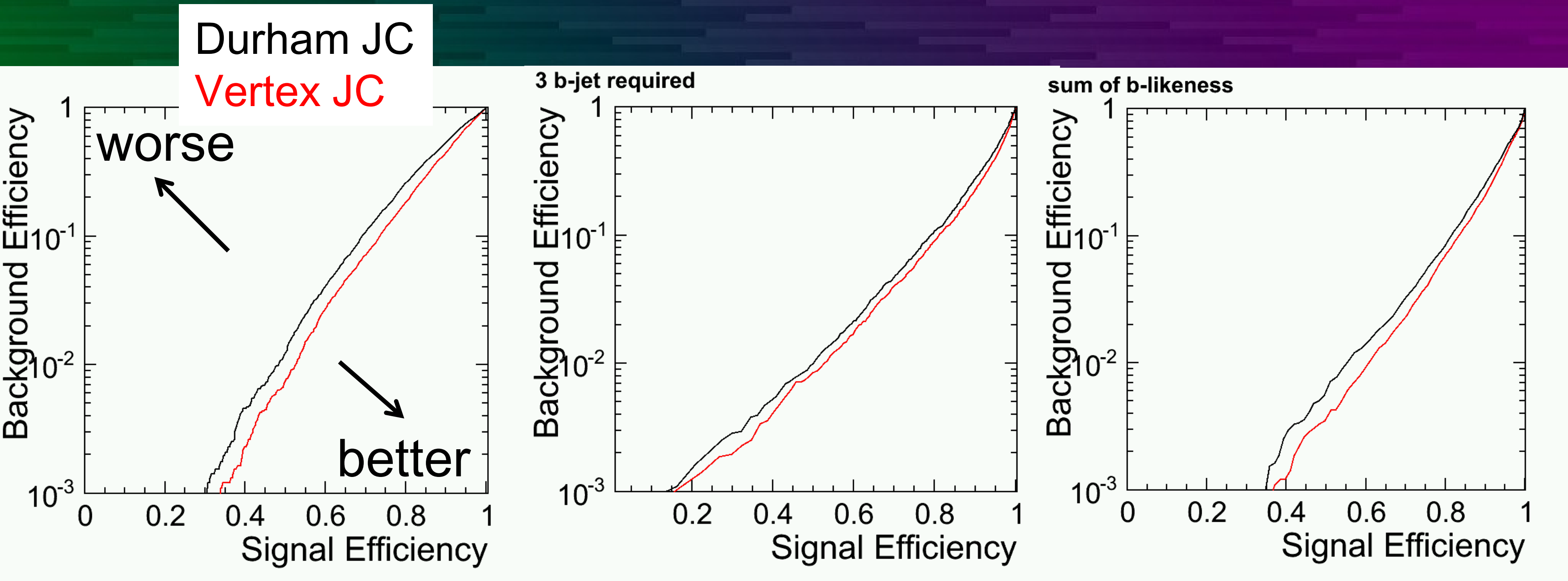
Study with realistic b-tagging

Using ZHH -> qqHH & tt -> bbcssc (same as 2.)

Procedure:

1. Jet clustering (Durham / Vertex)
2. Flavor tagging by LCFIVertex
   - obtain b-likeness value (0 to 1) for each jet
3. Check qqHH vs bbcssc acceptance with varying b-likeness threshold
   - Both 3rd and 4th jets and sum b-likeness over all jets are examined

3. ZHH vs. top-pair b-tagging performance



Cut on 4th jet | Cut on 3rd jet | Cut on sum b-likeness

- Improvements seen in all criteria
- Improvements are particularly significant for high-purity region (signal eff. < 60%)

4. Effect on b-tag cut in ZHH analysis

Practical impact on the physics study

Using ZHH -> qqHH & tt -> bbcssc (same as 2. & 3.)

1. Jet clustering & flavor tagging
2. Determine cut value of b-likeness at signal efficiency = 50%
   - High purity is needed to suppress enormous tt background
3. Count the remaining number of events and scale to 1 ab^{-1} luminosity

4. Effect on b-tag cut in ZHH analysis

Numbers in parentheses are at 1 ab^{-1}

Vertex jet clustering

	No cut	4th jet cut	3rd jet cut	3rd & 4th jet cut
qqHH (H -> bb)	8352 (73)	4233 (37)	4367 (38)	3163 (28)
bbcssc	9930 (100000)	95 (960)	113 (1140)	20 (201)

Durham jet clustering

	No cut	4th jet cut	3rd jet cut	3rd jet cut
qqHH (H -> bb)	8352 (73)	4277 (37)	4382 (38)	3116 (27)
bbcssc	9930 (100000)	145 (1460)	137 (1380)	29 (292)

30% improvement!

Summary

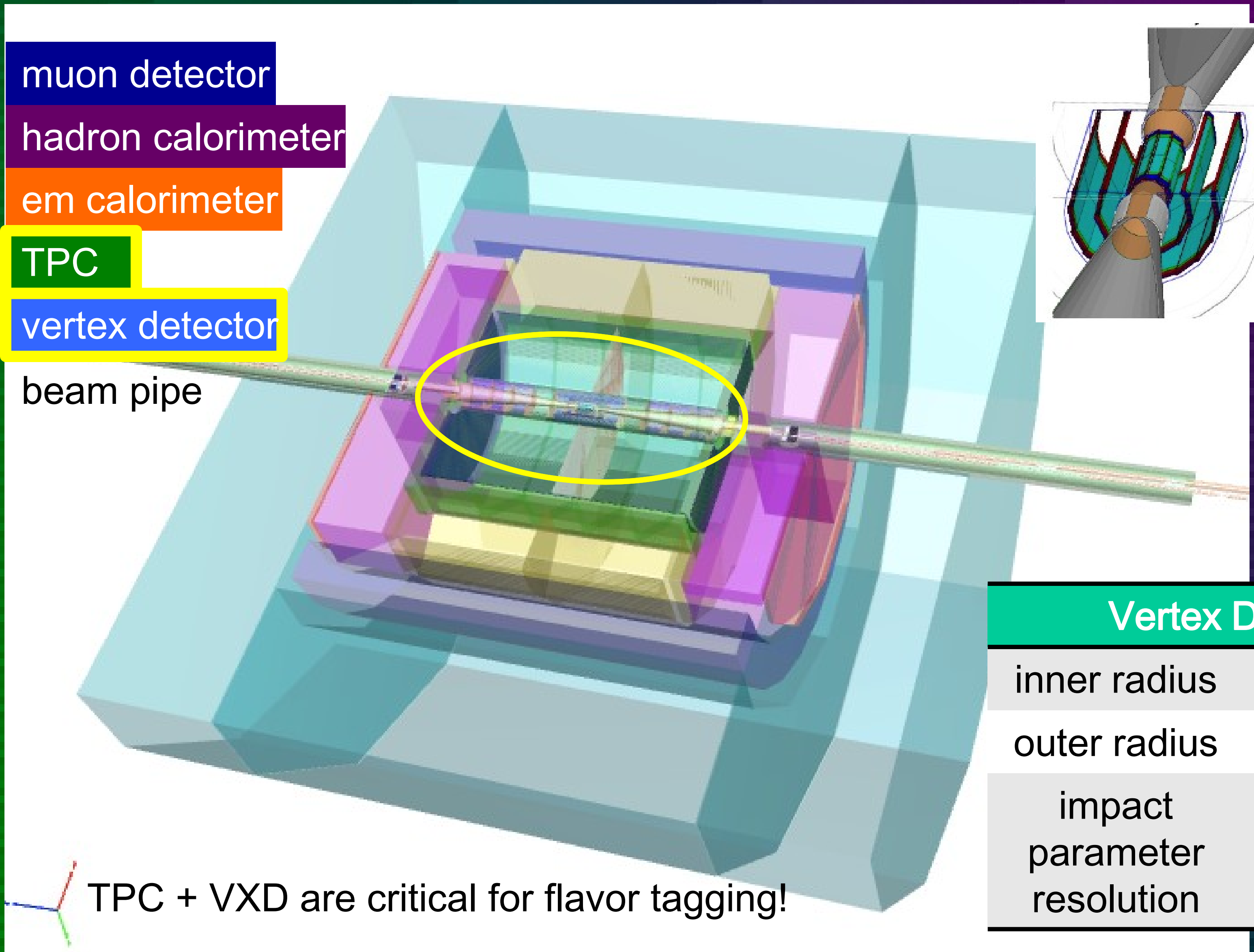
Vertex clustering can improve jet clustering performance for counting b-jets

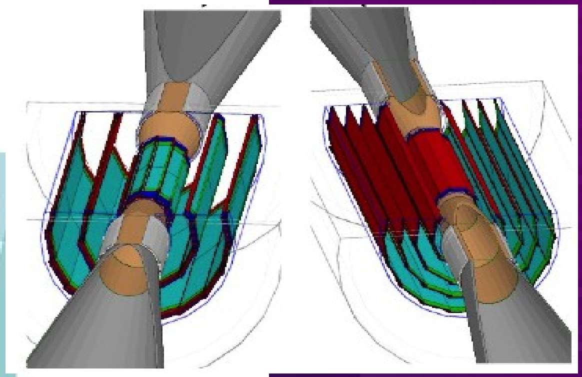
- Simple analysis with ZHH shows 30% improvement in reducing tt background

Backup

- ILD detector
- ZHH analysis by J. Tian
- Vertex finder details
- LCFIVertex input variables

ILD Detector



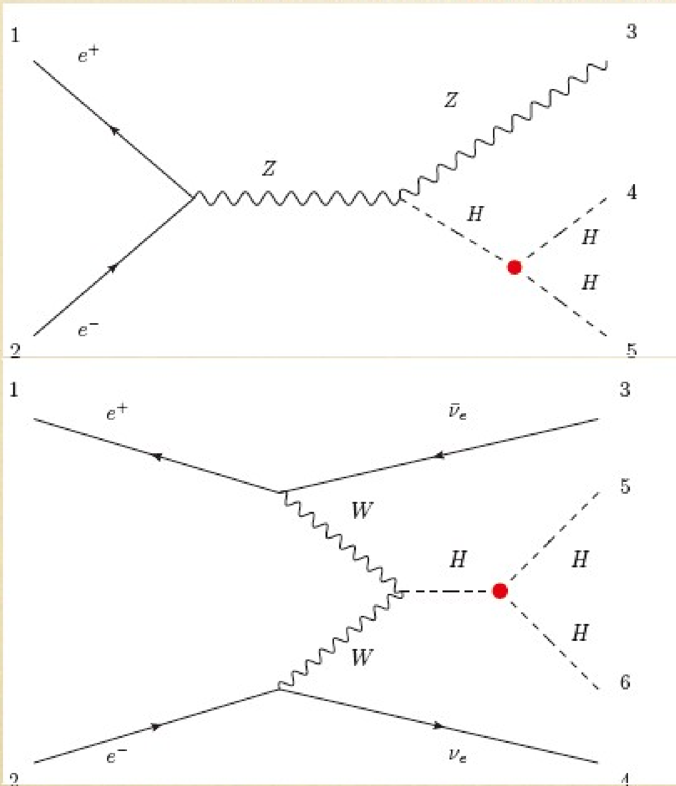


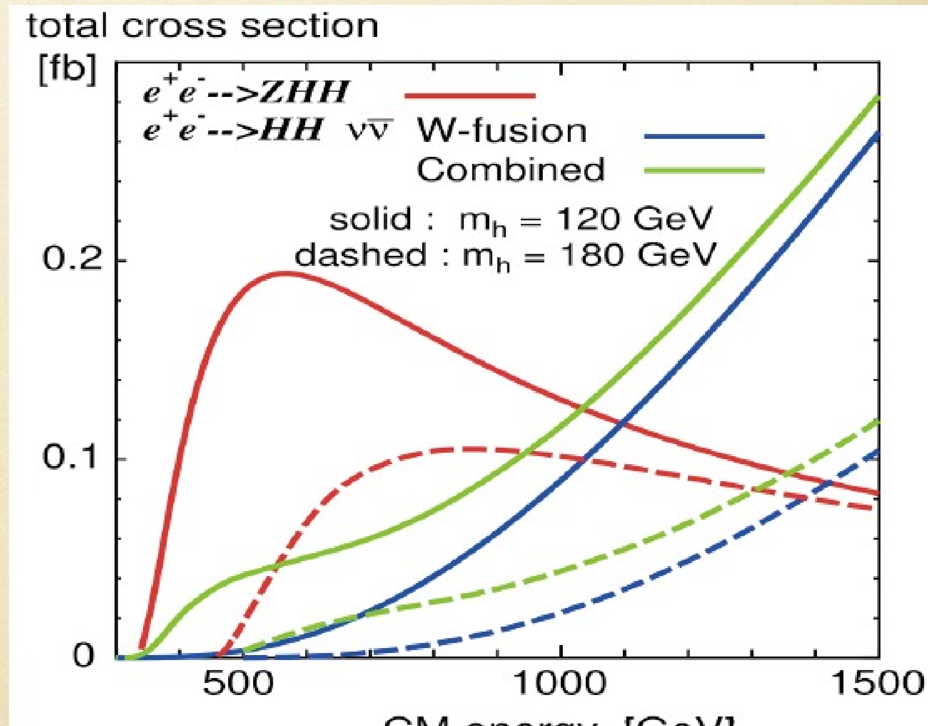
Vertex Detector	
inner radius	15 mm
outer radius	60 mm
impact parameter resolution	< 5 mm (high momentum)

TPC + VXD are critical for flavor tagging!

Measurement of the trilinear Higgs self-coupling @ ILC

- double Higgs-strahlung (dominate at lower energy)
- WW fusion (dominate at higher energy)





σ = 0.2 fb:
Only 100 events in 500 fb^{-1}

qqhh (with Z-like pair)

reduction table (probZ1+probZ2 > 0.9)

P(e-,e+)=(-0.8,+0.3) $E_{cm} = 500\mathrm{GeV}, M_H = 120\mathrm{GeV}$ $\int Ldt = 2\mathrm{ab}^{-1}$

normalized	expected	MC	pre-selection	probZ1+probZ2>0.9	Evis>400 MissPt<60 ! (Plmax>20&&Econe>10)	MLP_bbbb>0.2	MLP_bbqqqq>-0.3	MLP_qqbbbb>-0.6	Bmax3>0.76 Bmax4>0.33
qqhh(qqbbbb)	313(138)	117173	82.0(65.1)	15.5(13.8)	13.9(13.0)	13.1(12.3)	12.7(11.9)	12.1(11.4)	8.50(8.15)
qqbbbb	192	59994	50.9	3.17	2.97	2.01	1.75	1.28	0.55
qqqqH(ZZH)	381	49702	45.8	6.58	5.72	5.11	4.80	4.14	2.70
bbcsdu	394548	710285	3016	29.7	29.1	22.3	14.9	13.5	1.38
bbuddu	199165	109200	374	10.5	7.92	5.37	5.37	5.37	0.28
bbcssc	197790	359084	4904	58.4	53.8	47.9	39.2	36.5	2.01
ttqq	2169	9999	170	10.0	5.08	4.83	4.70	4.49	1.85
bbbb	40824	198431	4722	598	494	2.83	2.20	1.80	1.27
lvbbqq	821199	797027	12216	230	33.2	6.18	6.18	4.39	0.07
BG			25509	951	636	100	82.2	73.7	11.7

qqhh (without Z-like pair) reduction table

(probZ1+probZ2 < 0.9)

P(e-,e+)=(-0.8,+0.3) $E_{\mathrm{cm}} = 500\mathrm{GeV}, M_H = 120\mathrm{GeV}$ $\int Ldt = 2\mathrm{ab}^{-1}$

normalized	expected	pre-selection	probZ1+probZ2<0.9	Plmax<20 || Econe>10	Evis>400 MissPt<60	MLP_bbbb>-0.2	MLP_bbqqqq>0.16	Bmax3>0.8 Bmax4>0.52
qqhh(qqbbbb)	313(138)	82.0(65.1)	66.4(51.3)	63.0(50.9)	57.6(48.7)	54.9(47.1)	33.1(29.1)	16.6(15.1)
qqbbbb	192	50.9	47.7	47.4	44.9	36.2	11.7	6.00
qqqqH(ZZH)	381	45.8	39.2	38.2	35.0	32.3	15.5	7.65
bbcsdu	394548	3016	2986	2973	2869	2581	469	42.2
bbuddu	199165	374	364	364	356	[illegible]	[illegible]	5.37
bbcssc	197790	4904	4845	4825	4616	[illegible]	[illegible]	39.6
ttqq	2169	170	159	107	79.4	78.4	42.8	13.7
bbbb	40824	4722	4124	4106	3368	70.1	18.2	9.12
lvbbqq	821199	12216	11986	8041	1641	297	49.4	4.34
BG		25509	24557	20509	13015	7555	1298	129

ttbar with mis-b-tag

put all together
(preliminary)

Polarization: (e-,e+)=(-0.8,0.3) $e^+ + e^- \to ZHH$ $M(H) = 120\text{GeV}$ $\int Ldt = 2\text{ab}^{-1}$

Energy (GeV)	Modes	signal	background	significance	
				excess (I)	measurement (II)
500	$ZHH \to (l\bar{l})(b\bar{b})(b\bar{b})$	6.4	6.7	2.1σ	1.7σ
500	$ZHH \to (\nu\bar{\nu})(b\bar{b})(b\bar{b})$	5.2	7.0	1.7σ	1.4σ
500	$ZHH \to (q\bar{q})(b\bar{b})(b\bar{b})$	8.5	11.7	2.2σ	1.9σ
		16.6	129	1.4σ	1.3σ

we are interested in:

A. the combined significance of ZHH excess.

B. the combined precision of measured ZHH cross section.

Overall signal excess: 3.9σ in 2 ab^{-1} -> not enough??

1. Vertex finding (1)

- Original jet finder based on “build-up” method
  - ZVTOP cannot be used without tuning
    - It’s designed to be used after jet clustering
    - Too many fakes – without “jet-direction” parameter

- IP tracks are firstly removed (tear-down)
- Vertices are calculated for every track-pair
  - Calculate nearest points of two helices
    - Geometric calculation for the start point
    - Minuit minimization using track error-matrices

1. Vertex finding (2)

- Pre-selection
  - Mass < 10 GeV (B: ~5 GeV)
  - Momentum & vertex pos: not opposite to IP
  - Vertex mass < energy of either track
    - This selection is very effective for dropping fakes
  - Vertex distance to IP > 0.3 mm
  - Track chi2 to the vertex < 25
- Associate more tracks to passed vertices
  - Using same criteria as above
- Sort & select obtained vertices by probability
  - Associated (# tracks >=3) vertices are prioritized
  - Example in next slide…

1. Vertex finding (3)

- Example: 5 vertices are found

Vertex #	Tracks included	Probability
1	1,2,3	0.9
2	2,4,5	0.4
3	3,4	0.8
4	5,6	0.6
5	6,7	0.5

1. Vertex finding (3)

- Example

Vertex #	Tracks included	Probability	
1	1,2,3	0.9	Adopted!
2	[illegible],4,5	0.4 -> 0.6	
3	[illegible],4	0.8	Removed!
4	5,6	0.7	
5	4,7	0.5	

1. Vertex finding (3)

- Example

Vertex #	Tracks included	Probability
1	1,2,3	0.9
4	5,6	0.7
2	4,5	0.6
5	4,7	0.5

Adopted!

1. Vertex finding (3)

- Example

Vertex #	Tracks included	Probability	
1	1,2,3	0.9	Adopted!
4	5,6	0.7	Adopted!
2	4,5	0.6	Removed!
5	4,7	0.5	

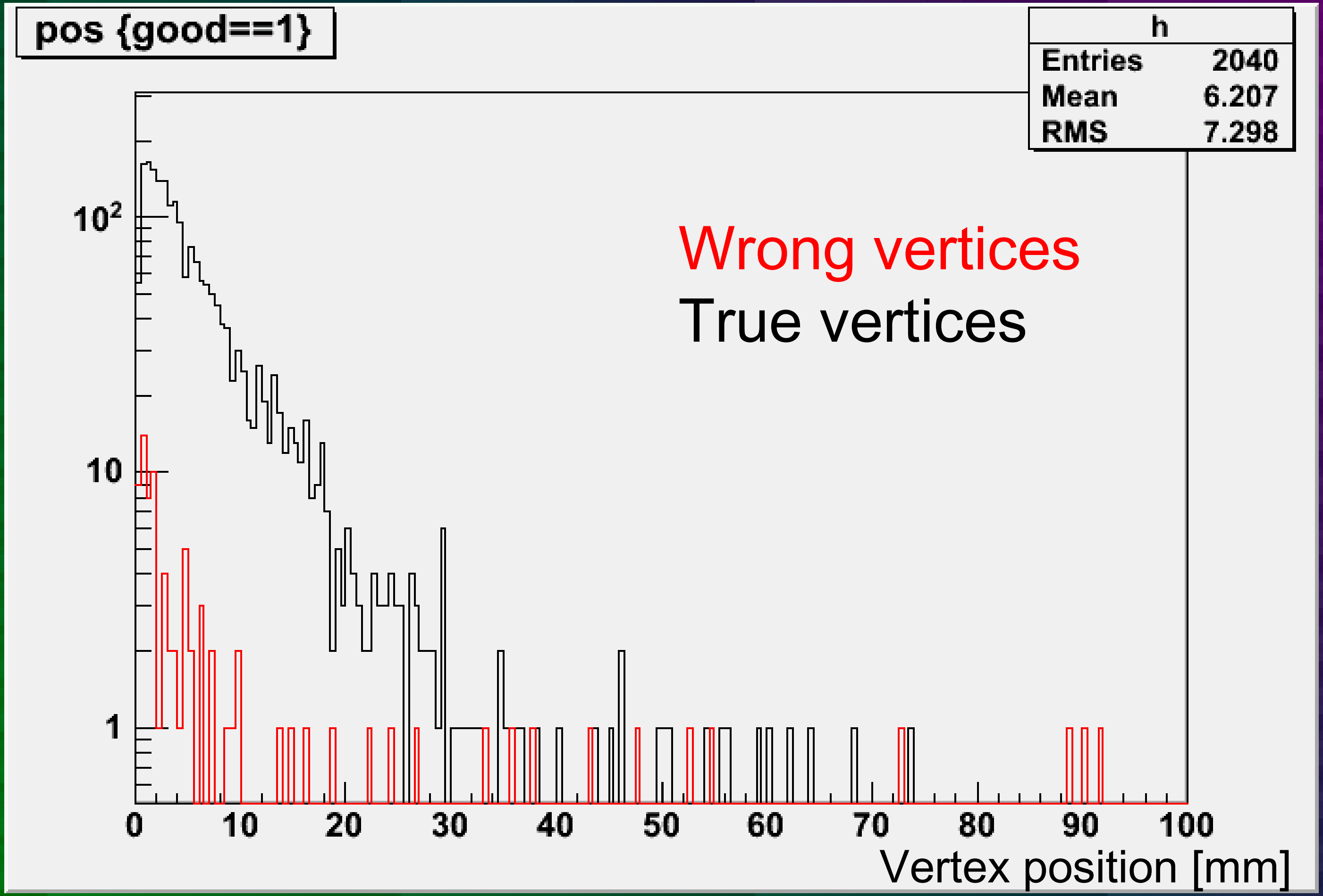
1. Vertex finding (3)

- Example

Vertex #	Tracks included	Probability	
1	1,2,3	0.9	Adopted!
4	5,6	0.7	Adopted!
5	4,7	0.5	Adopted!

Finally, three vertices are adopted.

Vertices in ZHH -> bbbbbb



In 347 bbbbbb events:
Good = 2040, bad (not from B-semistable) = 90

2. Vertex selection & muons

- Vertex selection
  - K0 vertices are removed (mass +- 10 GeV)
  - Vertex position > 30 mm are removed
    - Mostly s-vertices
- Secondary muons
  - Following tracks are treated as same as vertices
    - With muon hit
      - Currently > 50 MeV energy deposit
    - Impact parameter (> 5 sigma & not too much)
    - Ecal, Hcal energy deposit

Selection performance

- Vertex selection (347 bbbbbb events)

	Good vtx	Bad vtx	Purity
No cut	2040	90	96%
K0 and pos cut	1960	61	97%

Optimized for efficiency (bad contains partially bad)

- Lepton selection (347 bbbbbb events)

	Secondary μ	Other μ	Others	Purity
No cut	430	585	23168	1.8%
Muon hit > 50 MeV	267	23	49	79%
All cuts	178	4	5	95%

Optimized for purity (not so good efficiency)

LCFI input variables

- LCFI input variables:
  - three categories, trained independently:
    - # vertex = 0
    - # vertex = 1
    - # vertex >= 2
  - for # vertex = 0 (8 variables):
    - d_0 impact parameter (1)
    - d_0 impact parameter (2)
    - z_0 impact parameter (1)
    - z_0 impact parameter (2)
    - track momentum (1)
    - track momentum (2)
    - d_0 joint probability
    - z_0 joint probability
  - for # vertex = 1, >=2 (8 variables):
    - d_0 joint probability
    - z_0 joint probability
    - vertex decay length
    - vertex decay length significance
    - vertex momentum
    - pt-corrected vertex mass
    - vertex multiplicity
    - vertex probability from the fitter

(1) and (2) indicate the most and second most significant track.

“joint probability” – probability that a track comes from the IP, computed *a priori* using the distribution of impact parameter significance (separately for d_0 and z_0), multiplied for all tracks in the jet