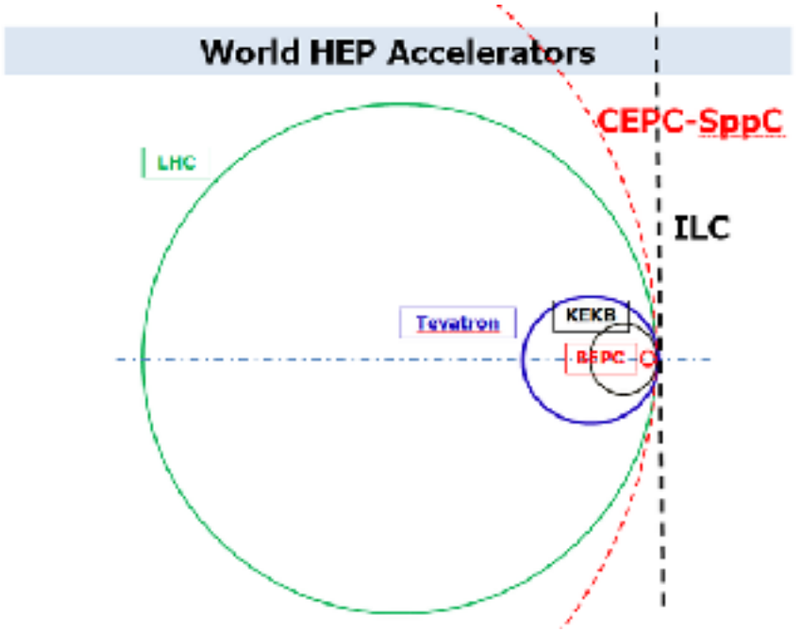
CEPC-SPPC Physics-Software Meeting

H->bb/cc/gg Branch Ratio Measurement in CEPC

Yu Bai (from Southeast University, Nanjing)

On Behalf of CEPC Physics-Software Study Group

November 28, 2016



Introduction: Benchmark Physics Process Measurement Overview

Physics Motivation

- Crucial to understand EW mechanism
  - Reason of fermions mass
  - Is there only one higgs boson?
- CEPC is ideal for such measurement
  - Clean background
  - Utility of recoil mass, free of higgs decay channel
  - High luminosity

One of the benchmark measurement in CEPC

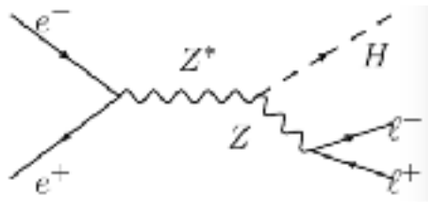
Physics Process	Measured Quantity	Critical Detector	Required Performance
$ZH \to \ell\ \ell\ X$	Higgs mass, cross section	Tracker	$\Delta(1/p_{\perp}) \sim 2 \times 10^{-5}$
$H \to \mu^+\mu^-$	BR($H \to \mu^+\mu^-$)		$\oplus 1 \times 10^{-3}/(p_{\mathrm{T}} \sin\theta)$
$H \to b\bar{b},\ c\bar{c},\ gg$	BR($H \to b\bar{b},\ c\bar{c},\ gg$)	Vertex	$\sigma_{r\phi} \sim 5 \oplus 10/(p \sin^{3/2}\theta)\ \mu$m
$H \to qq,\ VV$	BR($H \to qq,\ VV$)	ECAL, HCAL	$\sigma_E^{\mathrm{jet}}/E \sim 3-4\%$
$H \to \gamma\gamma$	BR($H \to \gamma\gamma$)	ECAL	$\sigma_E \sim 16\%/\sqrt{E} \oplus 1\%$ (GeV)

Excellent flavor tagging capability required

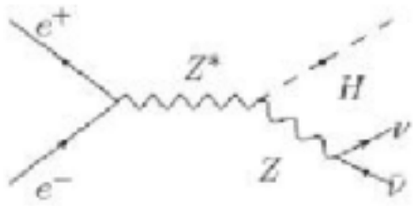
Introduction: Signal Process

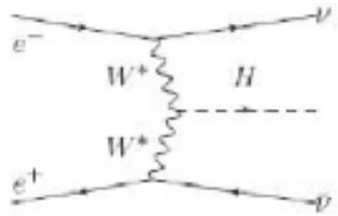
- **ZH->ll+jj (Zhenxing Chen)**
  - ZH production with Z decay to muon pair, H decay to bb/cc quark or gluon pair
  - Very clean signal in muon pair invariant mass and recoil mass
- **ZH->vv+jj(Yulei Zhang, Dikai Li, Hao Liang, Yu Bai)**
  - Via ZH(~86%) or WW fusion(~14%)
  - Clean background
- **ZH->Multi-jet (Boyang Li, Yu Bai)**
  - Both Z and Higgs decay hadronically
  - Much larger cross section than semi-leptonic channel

ZH->ll+jj

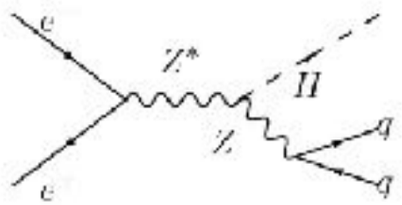


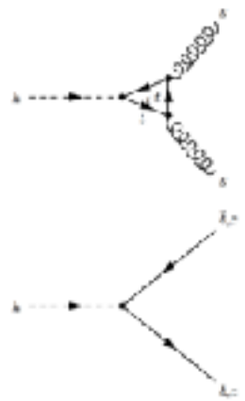
ZH->vv+jj



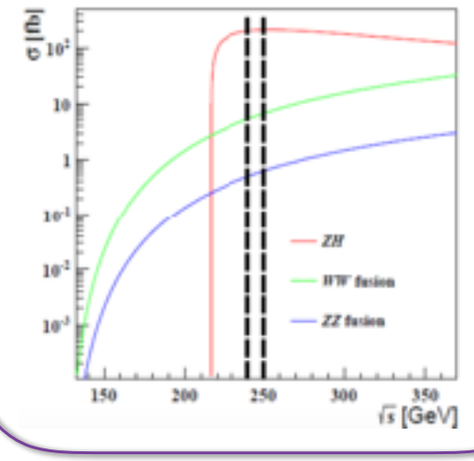


ZH->multi-jets





√s = 250 GeV



Integral luminosity of 5000 fb^{-1} is assumed in these study, corresponding to that of a few years of CEPC running

Upgrade Since Last Software-Physics Meeting

- **Data set and event selection:**
  - nnh: full simulation sample
  - llh: higher performance with more variables
  - qqh: add complete set of qq sample, bug fixed
- **Template fit:**
  - Non higgs background fixed in fit
- **Combination:**
  - Combine the template fit in nnh, llh and qqh channel

Analysis Outline

Event Preselection

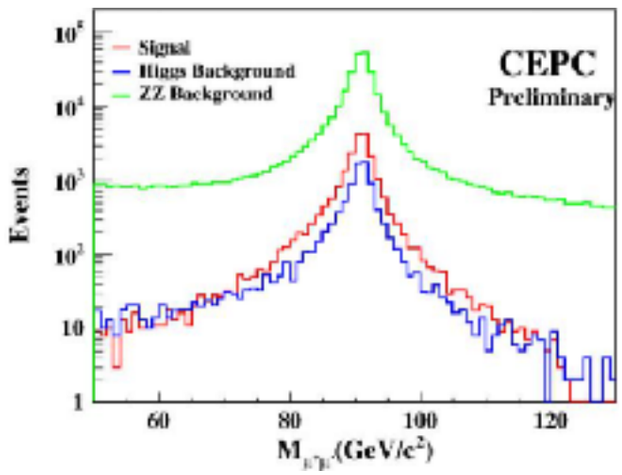
	vvH Channel	llH Channel	qqH Channel
Dominant Background	WW/ZZ->semi-lep, quark pair production, Other Higgs Process	ZZ semi-lep, Other Higgs Processes	WW/ZZ hadronic, quark-pair production
Discriminator	Missing Energy, Jets pair invariant/recoil masses, jet multiplicity(yth-value).	Lepton pair invariant Mass, Lepton pair recoil Mass, Jets invariant Mass, Higgs polar angle	Jet multiplicity, jet paring(invariant mass and angular distribution)

Flavor Tagging: LCFIPlus: MVA flavor weight, Template fit or other methods

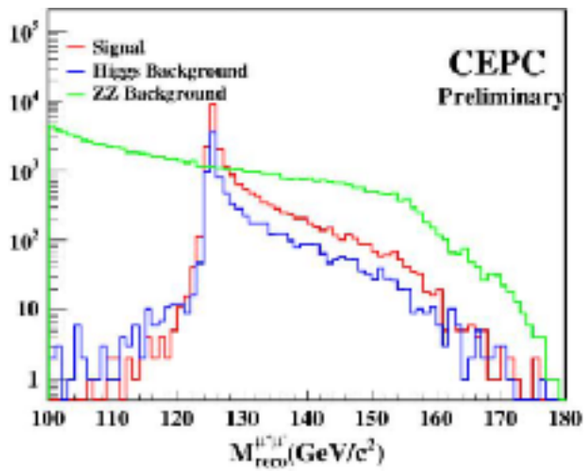
The outcome currently interpreted in terms of σ×*br* with statistic uncertainty

ZH->ll+jj Channel

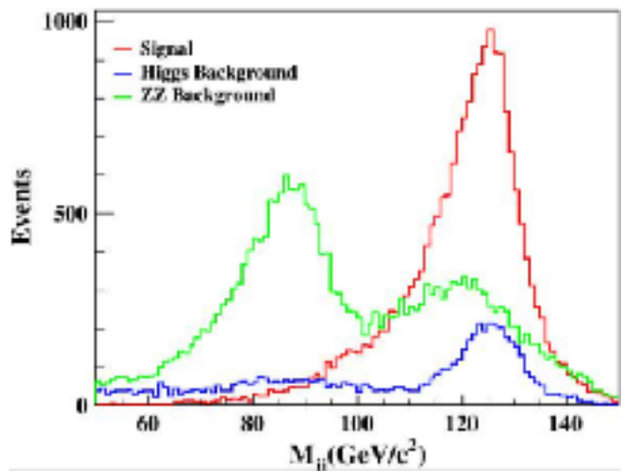
Lepton pair invariant mass between 70 - 105 GeV



Lepton pair recoil mass between 120 - 150 GeV(120 - 160 GeV for ee channel)



jet pair invariant mass between 105 - 135 GeV



Also add cos θ_z cut and y23-y34 cut

|cos θ_z| < 0.78 for ee channel and 0.85 for mumu channel

y23(y34) < 0.362(0.012) for ee channel and 0.365(0.012) for mumu Channel

Results of Event Selection

	signal	higgs bkg	non-higgs bkg
ee channel	9.15(9.42)k	1.10(1.87)k	6.15(11.1)k
$\mu\mu$ channel	12.8(16.5)k	1.48(3.34)k	5.29(7.81)k

$\nu\nu$+jj Channel, Datasets and Event Pre-selection

Datasets:

Full simulated signal and leading backgrounds(background dominantly simulated by arbor v1 or v2_1)

Generator: Whizard, Simulation arbor
Major SM background are with qqln, qqnn and qq events

Event Pre-selection:

- Number of particles(PFO) >=20
- Visible Energy between 110 and 150 Ge
- Isolated electron and isolated muon veto
- y_{12} between 0.15 and 1.0, y_{23}<0.06, y_{34}<0.008
- cos θ between -0.98 and -0.4,
- BDT Cut(redone at 0.07)

Much better than ILC Results:

Event yields after Cut flow

Full Sim Results		
Sig	h_bkg	SM_bkg
85.8k	1.96k	22.88k
Fast Sim Results		
Sig	h_bkg	SM_bkg
76.6k	1.59k	5.31k

Signal and background normalized ~ 5000/ fb

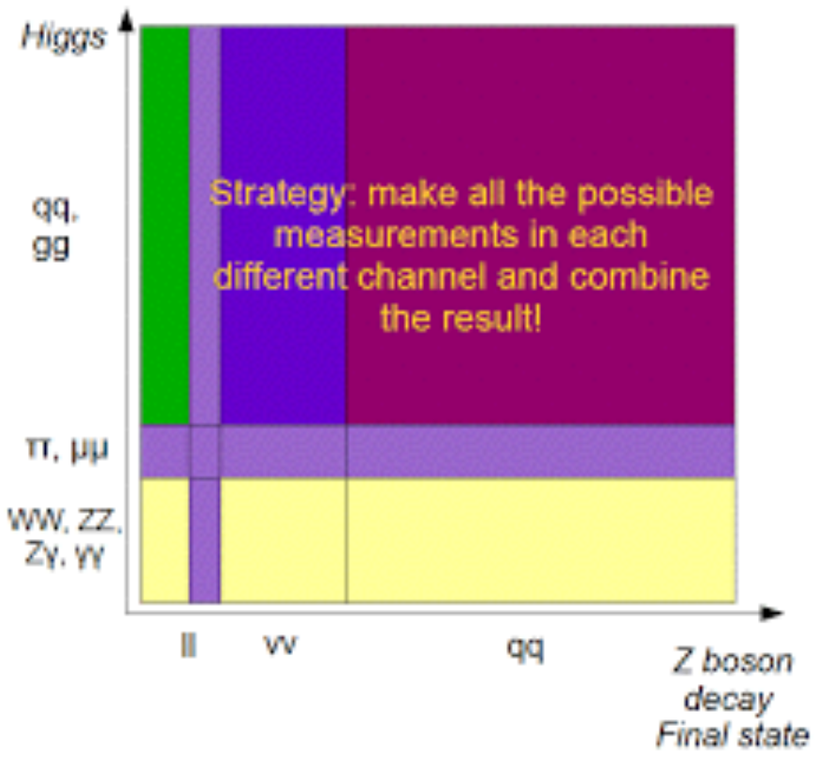
Signal yields: 85.8k, background: 24.84k

Realistic: template fit can be applied to leading bkg

		Signal	Background
Likelihood ratio	$LR > 0.165$	6293	10940

Multi-jets Channel

Majority decay FS in HZ production
Large Statistics



Datasets:

- **Main background: Irreducible background from WW/ZZ and quark pair process**
- **Full simulation sample used**
- **Quark pair sample are filtered due to its huge cross section and low rate to passing the event selection**

With 4 jets in final states, which two jets should be used to do template fit (from higgs) is ***not*** straightforward.

Event Selection and Cut flow

筛选条件	四喷注组合 并去除轻子	可见能量 >206GeV	y_{34} > 0.007	喷注粒子 数>9	$\Delta\theta > 0.92$	X>0.21	BDT>-0.19
所有信号	493947	459972	393979	371240	318163	236652	211281
信号－$b\bar{b}$	413299	381470	325137	305982	261408	197510	177447
信号－$c\bar{c}$	19362	18690	15976	14903	12610	9562	8324
信号－gg	61286	59812	52866	50363	42745	29580	25510
所有背景	75M	50.6M	26.4M	21.4M	13.55M	3.15M	1.52M
希格斯背景	299534	109529	100813	82281	71987	38579	32653
四费米子强子衰变	36.83M	28.19M	21.34M	18.78M	12.16M	2.2M	1.08M
两夸克	23.86M （250M）	20.37M	5.207M	2.601M	1.315M	907188	405567
四费米子半轻子衰变	14.72M	1.167M	218394	27487	4745	3012	580

No Major update

See Boyang’s talk with detail in last physics-software meeting

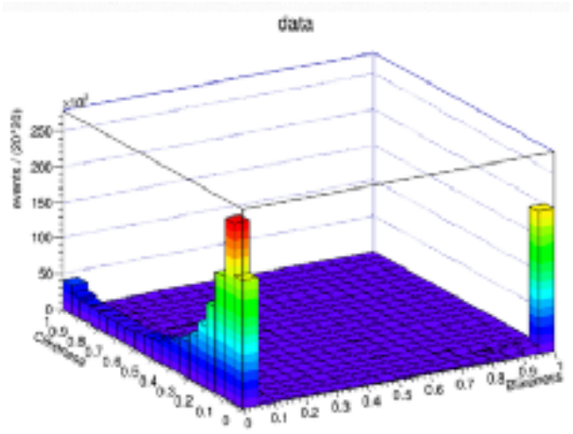
After event selection, the dominant background are WW/ZZ hadronic decay

Better than ILC Results: Likelihood ratio $LR > 0.375$ 13726 166807

Template Fit and ToyMC Results(Old)

$$L_{qq} = \frac{qq\ pair}{qq\ pair\ +\ neither\ is\ q} = \frac{x_q^1\ x_q^2}{x_q^1 x_q^2 + (1-x_q^1)(1-x_q^1)}\ (qq = bb, cc)$$

r_{qq} : $\sigma \times Br_{qq}$



CEPC Results

	e^+e^-H	$\mu^+\mu^-H$	$\nu\nu H$ ($2500\ fb^{-1}$)	qqH	combined
r_{bb}	1.000±0.013	1.000±0.009	0.997±0.006	1.000±0.003	1.000±0.002
r_{cc}	1.002±0.118	0.993±0.095	1.008±0.038	1.000±0.120	1.00±0.02
r_{gg}	0.998±0.106	0.991±0.070	1.020±0.019	1.001±0.189	1.00±0.01

Large uncertainty due to SM bkg

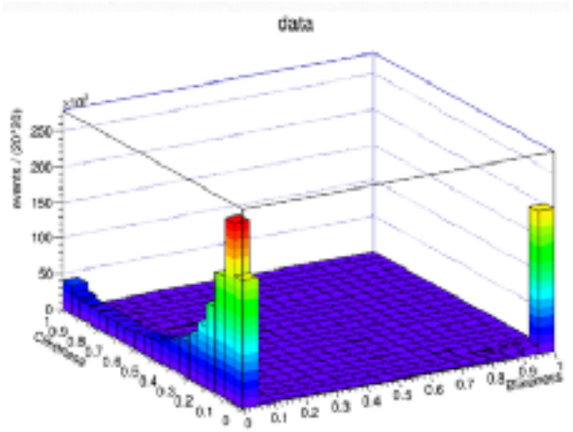
Comparable results in llH channel
Much better result in vvH channel
Much worse results(for r_c and r_g) in qqH channel

ILC Results

	$\nu\bar{\nu}H$	$q\bar{q}H$	e^+e^-H	$\mu^+\mu^-H$	Comb.
$r_{b\bar{b}}$	1.00 ± 0.02	1.00 ± 0.01	1.00 ± 0.04	1.00 ± 0.03	1.00 ± 0.01
$r_{c\bar{c}}$	1.02 ± 0.11	1.01 ± 0.10	1.02 ± 0.27	1.01 ± 0.23	1.02 ± 0.07
r_{gg}	1.02 ± 0.14	1.02 ± 0.13	1.05 ± 0.33	1.02 ± 0.24	1.02 ± 0.09

Template Fit and ToyMC Results

$$L_{qq} = \frac{qq\ pair}{qq\ pair\ +\ neither\ is\ q} = \frac{x_q^1\ x_q^2}{x_q^1 x_q^2 + (1-x_q^1)(1-x_q^1)}\ (qq = bb, cc)$$



CEPC Results r_{qq} : $\sigma \times Br_{qq}$

	e^+e^-H	$\mu^+\mu^-H$	$\nu\nu H$	qqH	combined
r_{bb}	1.000±0.012	0.999±0.011	1.000±0.004	1.000±0.004	1.000±0.003
r_{cc}	0.994±0.144	0.997±0.128	0.999±0.038	1.001±0.080	1.00±0.032
r_{gg}	1.004±0.078	1.003±0.069	1.001±0.016	1.001±0.052	1.00±0.016

SM Background fixed, as well as nn and ll channel

ILC Results

	$\nu\bar{\nu}H$	$q\bar{q}H$	e^+e^-H	$\mu^+\mu^-H$	Comb.
$r_{b\bar{b}}$	1.00 ± 0.02	1.00 ± 0.01	1.00 ± 0.04	1.00 ± 0.03	1.00 ± 0.01
$r_{c\bar{c}}$	1.02 ± 0.11	1.01 ± 0.10	1.02 ± 0.27	1.01 ± 0.23	1.02 ± 0.07
r_{gg}	1.02 ± 0.14	1.02 ± 0.13	1.05 ± 0.33	1.02 ± 0.24	1.02 ± 0.09

Combination

parameter list: 4*NChannel + NBackground*NChannel+2

normalization factor	bb	cc	1-bb-cc(gg)
NChannel	2		free

2+NChannel free parameters, repeat for each channel

background yields

efficiencies for bb/cc/gg in each channel

NBackground*NChannel

3NChannel fixed

template list:

r_{qq} : $\sigma \times Br_{qq}$ = bb * normalization factor

bb/cc/gg and backgrounds template for each channel

Also tried to add higgs background template in the fit. But the result is not good. Might due to very low hbkg efficiency and statistics.

Summary

- H->bb/cc/gg from μμH, llH and qqH is about to finish
  - ZH->ll+jj, analysis with all SM background and apply more cuts to enhance
    signal/background ratio and reduce statistic uncertainty
    - In vv+jj channel, full simulation SM background have been used
    - In qqH->4jets channel, complete background are used(no significant change)
- ToDo
  - Understand why llH accuracy dropped

Backup

nnH Cutflow - I

	Signal	nnH_bkg	swsl_qq(very low stat.)	ww_sl_tauq	ww_sl_mu_up	sznu_nu_up	sznu_up_down	zz_sl_un_up	zz_sl_nu_down
FSClasser	167500	71934	13200	7.11162M	7.12766M	282392	459391	411970	678581
NPFO	167340	41980	12390	7.02154M	6.97721M	238738	415782	361059	620419
E Total	149683	33269	2098	3.02259M	762195	95492	154312	173698	277824
pT	139539	31022	1962	2.57739M	715799	87556	141266	159322	254937
IsoLep Veto	138620	27472	692	2.23187M	218486	87004	140018	158243	252739
M_invariant	132206	25703	341	1.31696M	39686	9409	13771	17122	24247
M_recoil	123733	24379	241	1.0477M	17897	8423	12230	15675	22241
y12	123016	23738	232	965445	16640	8281	11960	15376	21788
y23	124133	13240	128	639871	11117	6641	9284	12428	16729
y34	105451	5608	112	531096	9584	6097	8472	11447	15300
cos θ_{jj}	102186	5078	87	302146	5598	5032	6821	9258	11927
BDT>0.07	82371	1927	1	6637	285	630	931	1097	1609

nnH Cutflow - II

	ww_cuxx	bbrad	ccard	ssrad	uurad	ddrad	bbnonrad	ccnonrad	ssnonrad
FSClasser	1.70919M	234349	267377	291697	296770	291253	218209	236130	267998
NPFO	1.70918M	230511	255381	277154	274663	278266	217751	235012	264488
E Total	15749	154918	185070	211359	208697	214986	4894	3205	2865
pT	2663	820	357	155	50	76	944	296	184
IsoLep Veto	2646	811	355	155	50	76	937	294	184
M_invariant	1181	87	74	59	20	31	644	186	94
M_recoil	1104	66	57	34	9	20	538	152	61
y12	1089	66	56	34	9	19	527	151	54
y23	388	53	48	34	9	18	464	140	47
y34	141	51	48	34	9	18	458	139	46
cos θ_{ij}	92	34	27	7	4	5	344	117	26
BDT>0.07	1	1	0	0	0	0	68	10	0

	uunonrad	ddnonrad	bbfilter	ccfilter	ssfilter	uufilter	ddfilter
FSClasser	258736	267416	728308	1.09601M	788760	1.11429M	794210
NPFO	255376	264585	728291	1.09597M	788562	1.11405M	794032
E Total	2300	2535	10739	7366	4126	4275	3892
pT	75	126	2055	810	278	163	220
IsoLep Veto	75	126	2040	807	278	163	220
M_invariant	40	70	1336	540	178	102	138
M_recoil	21	36	1184	459	113	57	82
y12	18	34	1136	441	103	52	74
y23	13	22	760	310	77	36	53
y34	12	20	646	275	61	28	34
cos θ_{ij}	7	14	526	239	40	14	23
BDT>0.07	0	0	91	23	4	0	2

nnH Cutflow -III

	eeh_bb	eeh_cc	eeh_gg	eeh_bkg	mumuh_bb	mumuh_cc	mumuh_gg	mumuh_bkg	tautauh_bb	tautauh_cc	tautauh_gg	tautauh_bkg
FSClasser	55058	2514	8815	29354	54974	2579	7994	28422	57909	2742	8537	31089
NPFO	55050	2513	8813	20360	54895	2567	7990	19984	57906	2742	8536	22272
E Total	6	0	0	237	86	3	4	291	5792	186	471	4417
pT	4	0	0	214	25	1	0	247	4421	147	396	3415
IsoLep	0	0	0	10	2	0	0	1	3911	134	381	1777
M_invariant	0	0	0	6	1	0	0	1	2390	79	207	1231
M_recoil	0	0	0	5	1	0	0	1	2225	75	197	1034
y12	0	0	0	5	1	0	0	1	2218	75	194	1018
y23	0	0	0	4	1	0	0	1	1818	59	144	510
y34	0	0	0	4	1	0	0	0	1408	46	107	177
cos θ_{ij}	0	0	0	4	1	0	0	0	1333	44	102	150
BDT>0.07	0	0	0	0	0	0	0	0	380	8	38	4

	qqh_bb	qqh_cc	qqh_gg	qqh_bkg	nnh_bb	nnh_cc	nnh_gg	nnh_bkg
FSClasser	417960	19411	61462	223861	140472	6547	20481	71934
NPFO	417960	19411	61462	223602	140333	6531	20476	49180
E Total	85	1	2	16429	123909	6105	19669	33269
pT	31	0	0	14389	115498	5695	18346	31022
IsoLep Veto	31	0	0	10416	114594	5691	18335	27472
M_invariant	8	0	0	7716	108663	5530	18013	25703
M_recoil	7	0	0	7072	101089	5278	17365	24378
y12	7	0	0	6973	100535	5261	17220	23738
y23	5	0	0	4967	91827	4832	14234	13240
y34	3	0	0	3092	88368	4648	12435	5608
cos θ_{ij}	3	0	0	2670	85670	4524	11992	5078
BDT>0.07	2	0	0	357	69820	3029	9522	1927

eeH Cutflow -l

	signal	sze_sl_dd	sze_sl_uu	zzsl_tau_up	zzsl_tau_down
FSClasser	48263	57803	74325	490	861
cos θ_H	38859	27065	31876	287	530
$M_{ee\_invariant}$	38226	4613	4791	67	113
M_{ee_recoil}	36475	3516	3459	24	38
$M_{jj\_invariant}$	26747	1190	1223	2	5
y23&y34	25389	1033	1085	2	4

eeH Cutflow -II

	eeh_bb	eeh_cc	eeh_gg	eeh_bkg	mumuh_bb	mumuh_cc	mumuh_gg	mumuh_bkg	tautauh_bb	tautauh_cc	tautauh_gg	tautauh_bkg
FSClasser	40544	1819	5900	255110	1	0	0	150	191	12	28	550
cos θ_H	32625	1455	4779	19984	1	0	0	126	155	11	23	436
$M_{ee\_invariant}$	32081	1432	4713	18264	0	0	0	63	9	0	0	123
M_{ee_recoil}	30563	1380	4532	16794	0	0	0	38	5	0	0	65
$M_{jj\_invariant}$	22387	1025	3335	5371	0	0	0	0	5	0	0	7
y23_y34	21441	983	2965	3058	0	0	0	0	5			4

	qqh_bb	qqh_cc	qqh_gg	qqh_bkg	nnh_bb	nnh_cc	nnh_gg	nnh_bkg
FSClasser	0	0	0	1588	0	0	0	282
cos θ_H	0	0	0	1286	0	0	0	211
$M_{ee\_invariant}$	0	0	0	578	0	0	0	108
M_{ee_recoil}	0	0	0	324	0	0	0	76
$M_{jj\_invariant}$	0	0	0	87	0	0	0	0
y23_y34	0	0	0	56	0	0	0	0

mumuH Cutflow -I

	signal	zz_sl_mu_down	zzsl_mu_up	zzsl_tau_up	zzsl_tau_down
FSClasser	60414	185637	112410	874	1595
$\cos\theta_H$	52315	122204	71663	586	1084
$M_{\mu\mu\_invariant}$	50548	12946	8226	105	180
$M_{\mu\mu\_recoil}$	49433	10711	6849	53	86
$M_{jj\_invariant}$	40523	3303	2491	5	6
y23&y34	38595	2946	2229	5	5

mumuH Cutflow -II

	eeh_bb	eeh_cc	eeh_gg	eeh_bkg	mumuh_bb	mumuh_cc	mumuh_gg	mumuh_bkg	tautauh_bb	tautauh_cc	tautauh_gg	tautauh_bkg
FSClasser	0			241	50753	2355	7306	30052	235	7	31	692
cos θ_H				191	40567	1895	5853	23928	181	4	27	546
$M_{\mu\mu\_invariant}$				60	39170	1837	5673	21369	6	0	0	112
$M_{\mu\mu\_recoil}$				41	38317	1791	5538	20428	4	0	0	69
$M_{jj\_invariant}$				1	30942	1580	4911	7638	4	0	0	7
y23_y34				1	29771	1514	4380	4123	4	0	0	5

	qqh_bb	qqh_cc	qqh_gg	qqh_bkg	nnh_bb	nnh_cc	nnh_gg	nnh_bkg
FSClasser	0			2047	1	0		245
cos θ_H				1640	0			192
$M_{\mu\mu\_invariant}$				551	0			83
$M_{\mu\mu\_recoil}$				324	0			73
$M_{jj\_invariant}$				102	0			0
y23_y34				69	0			0

qqH Final Cut Results-I

	qqh_bb	qqh_cc	qqh_gg	qqh_bkg	nnh_bb	nnh_cc	nnh_gg	nnh_bkg
Event Number	176478	8256	25244	31810	0			
Weight	1.0018456							
Event Yields	176830.7	8271.2	25290.6	31868.7	0			

	eeh_bb	eeh_cc	eeh_gg	eeh_bkg	mumuh_bb	mumuh_cc	mumuh_gg	mumuh_bkg	tautauh_bb	tautauh_cc	tautauh_gg	tautauh_bkg
Event Number	17	0	4	16	1	0	0	0	343	18	111	206
Weight	0.3818478				0.3586621				0.3581369			
Event Yields	6.49	0	1.53	6.11	0.3586621				122.84	6.45	39.75	73.78

qqH Final Cut Results-II

	ww_ccbs	zz/ ww_cscs	zz_dtdt	zz_cc_nots	zz_utut	ww_uusd	ww_ccds	ww_cuxx	zz_uu_notd	zz/ ww_udud	ww_uubd
Event Number	1911	145577	117663	27546	24109	18327	18853	36959	26004	135807	2421
weight	0.2916197	1.8205249	1.0059087	1.0045608	1.0009637	0.9869772	0.9881102	10.032289	1.0042245	1.8232797	0.0025250
Event Yields	557.29	265026.6	118358.2	27671.6	24132.2	18088.3	18628.8	370783.3	26113.9	247614.1	6.11
Sum 4q	1115980.4										

qqH Final Cut Results-III

	bb	cc	ss	uu	dd
Event Number	12844	21828	10911	18712	11326
weight	5.7120374	5.2311774	5.3775161	5.2450895	5.2513599
Event Yields	73365.4	114186.1	58674.1	98146.1	59476.9
Sum 2q	403848.6				

	zz_sl_mu_up	zz_sl_mu_down	zz_sl_tau_up	zz_sl_tau_down	sze_sl_uu	sze_sl_dd	ww_sl_muq	ww_sl_tauq	sw_sl_qq
Event Number	55	66	132	140	18	17	-	46	-
weight	1.0053624	1.0153959	1.0014406	1.0046875	5.1248076	1.5762269	-	1.0045894	-
Event Yields	55.29	67.02	132.19	140.66	92.25	26.80	-	46.21	-
Sum sl	560.42								

nnH fast cutflow

Cut Definition	*Sig.*	*qq*	*qqnn*	*qqln*	*xxh*
FSClasser output	148955	25M	183687	3698817	63194
$N_{\mathrm{PFO}(E>0.4\mathrm{GeV})} > 20$	148808	23M	163088	3439927	58882
$110 < E_{\mathrm{total}} < 150$	132561	10M	125878	705357	34215
$P_T > 19$	126006	34198	116314	627602	32300
Isolation lepton veto	123586	33775	115867	327206	23773
$100 < M_{\mathrm{inv}} < 135$	117845	9506	10420	162511	21277
$70 < M_{\mathrm{rec}} < 125$	111886	7521	10045	110426	20458
$0.15 < y_{12} < 1$	111353	7405	9702	101797	19983
$y_{23} < 0.06$	105078	6644	8456	69313	14495
$y_{34} < 0.008$	100117	6504	7878	58532	6899
$-0.98 < \cos(\theta^{(2\mathrm{jets})}_{\mathrm{included}}) < -0.4$	97277	5178	5365	33293	6273
$BDT > -0.01$	76666	344	118	69	1594
Significance			**265.20**		
Efficiency			**51.5%**		

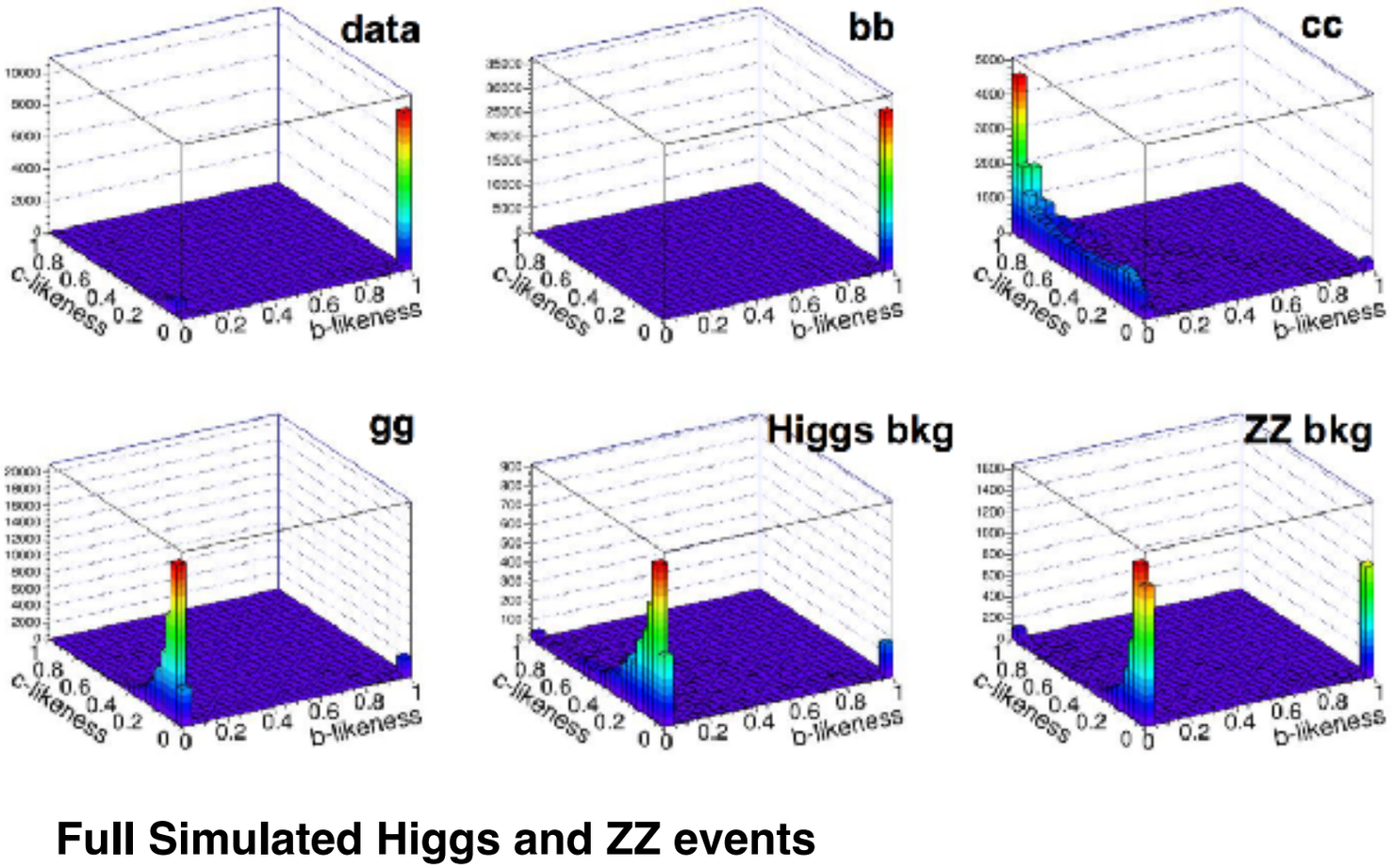
Template Fit in ll+jj Channel

Templates and data

Category	*Signal*	*ZH*	*ZZ*
Pre-selection	21300	9486	381678
$120GeV < M^{\mu^+\mu^-}_{recoil} < 150GeV$	20582	8678	35305
$70GeV < M_{\mu^+\mu^-} < 105GeV$	20121	8340	26525
$105GeV < M_{jj} < 135GeV$	16479	3336	7732

Fit target

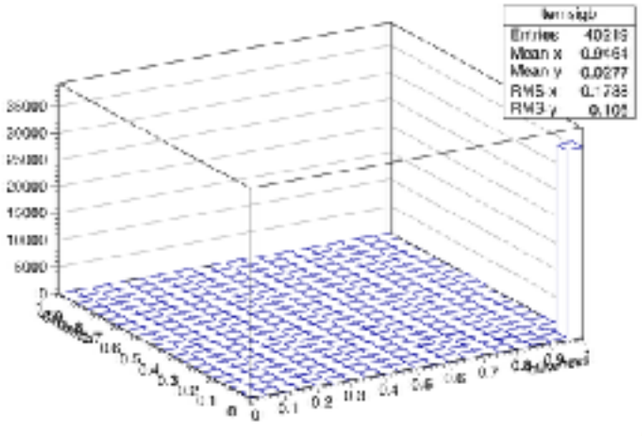
$$L_{qq} = \frac{qq\ pair}{qq\ pair\ +\ neither\ is\ q} = \frac{x_q^1\, x_q^2}{x_q^1 x_q^2 + (1-x_q^1)(1-x_q^1)}\ (qq = bb, cc)$$



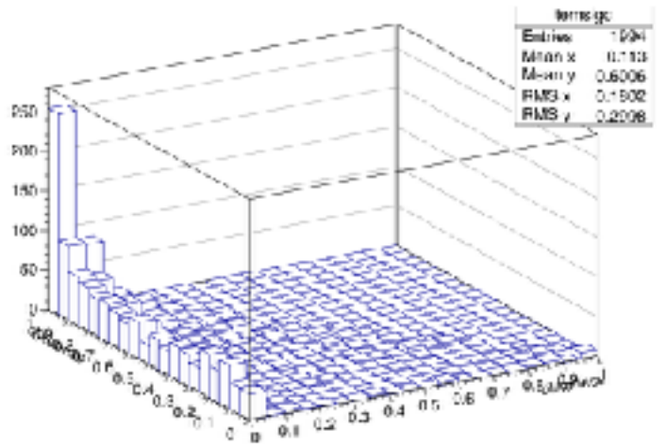
Full Simulated Higgs and ZZ events

vvH Templates

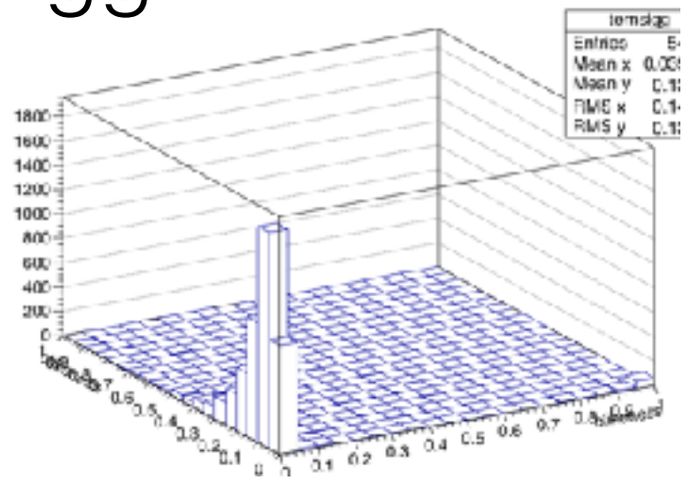
bb



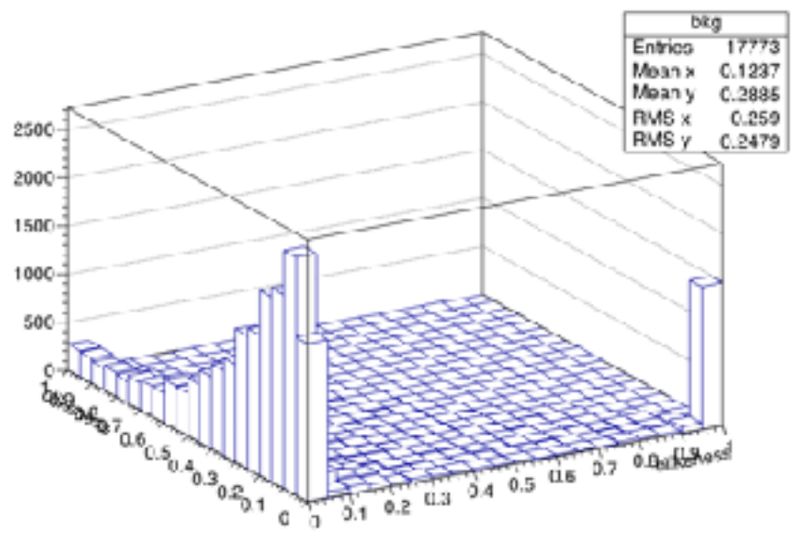
cc



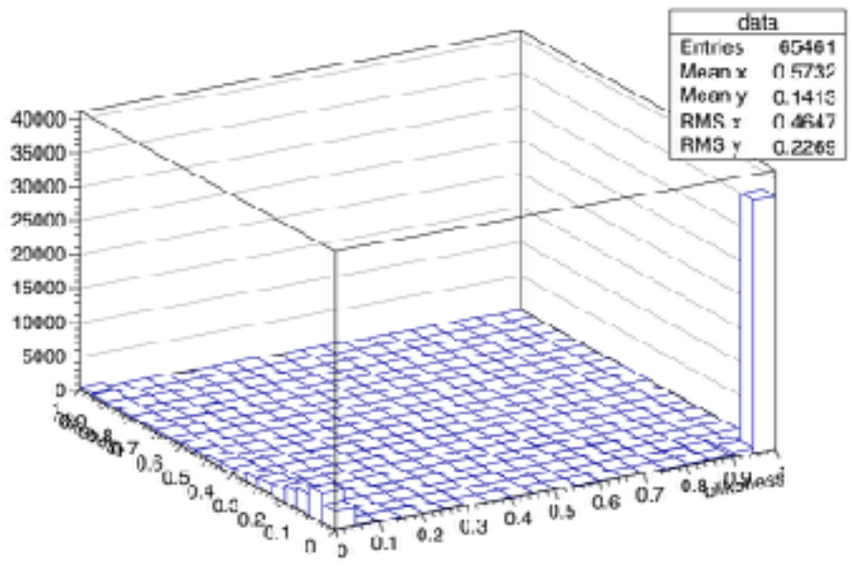
gg



background



data



MVA in vv+jj channel