Chicago LC Workshop, January 7-9, 2002

Direct measurement of the top-Higgs Yukawa coupling in an e+e- Linear Collider at 500 GeV

A. Juste (Fermilab)

OUTLINE

- Motivation
- Signal and background processes
- Simulation aspects
- Experimental aspects
- Semileptonic channel: a first look
- Summary and conclusions

Motivation (I)

- ***GOAL***: evaluate a realistic uncertainty for the direct measurement of the top-Higgs Yukawa coupling at 500 GeV

  => need to understand the potential of this channel in the first likely energy stage of a Linear Collider.

- By that time there will probably be some "**indirect measurement" of the top-Higgs Yukawa coupling** => made possible by recent (and future!) theoretical progress in the predictions of tt threshold observables (e.g. $\Delta\sigma_{tt}/\sigma_{tt} \sim 2\%$)
  - h-exchange between tops affects the interquark potential near threshold ($\sqrt{s} \sim 350$ GeV) => mainly $\sigma_{tt}(m_h, g_{tth})$
  - experimental analysis (M. Martinez and R. Miquel, Sept 16. 2001):
    - 300 fb^{-1} in a 10 point scan
    - Assume 1% systematic uncertainty in total cross-section.
    - Use 1S mass definition
    - 4-parameter fit (m_t, Γ_t, g_{tth} and α_s) to three threshold scan observables (σ_{tt}, dN/dp_t, A_{FB}^t):

Δm_t = 30 MeV;	$\Delta\alpha_s$=0.001 (external constraint)
$\Delta\Gamma_t$ = 33 MeV;	$\Delta g_{tth}/g_{tth}$= +0.33 –0.57 (correlations up to 85%)

Motivation (II)

Direct measurement of the top-Higgs Yukawa coupling at $\sqrt{s}$=800 GeV

A. Juste and G. Merino (hep-ph/9910301)

- Considered both semileptonic and hadronic channels (~90% of decays)
- "Realistic" detector effects and reconstruction procedures (EF, b-tagging, etc)
- Backgrounds considered: interfering (ttZ) and non-interfering (tt, qq, WW, ZZ, ZH).
- Made use of multivariate techniques for efficient signal/background discrimination. Main background after selection is tt.
- Assuming m_h=120 GeV, L=1 ab^{-1} and after combining both channels:

$$(\Delta g_{tth}/g_{thh})_{tot} = 4.2\%(stat) \oplus 3.6\%\ (syst)$$

5% syst. bckg normalization

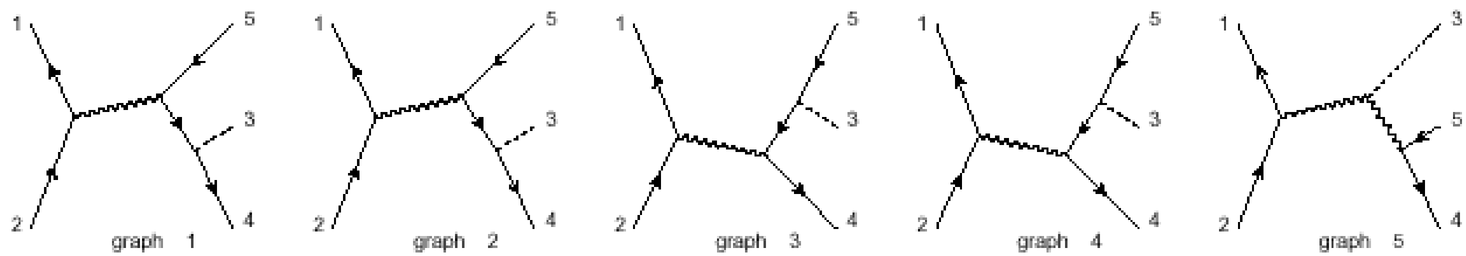
NAÏVE EXTRAPOLATION TO $\sqrt{s}$ = 500 GeV (with L=1 ab^{-1})

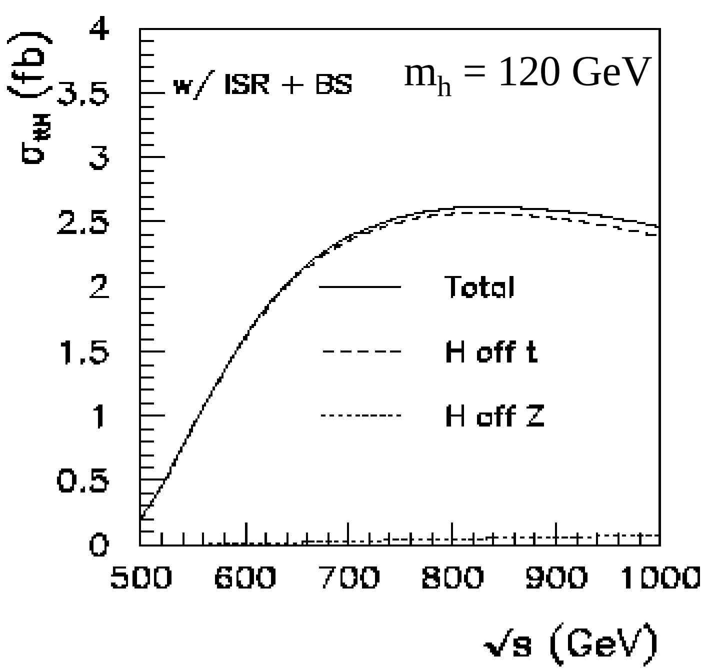
- Since sensitivity is a factor ~3.4 worse, and assuming same ε and ρ can be achieved:

$$(\Delta g_{tth}/g_{thh})_{tot} = 14.3\%(stat) \oplus 3.6\%\ (syst)$$

Signal Processes (I)

At LO 5 diagrams contribute to this process:





Dominant contribution: γ-exchange.

The diagram in which the h is radiated off the Z is just a small correction

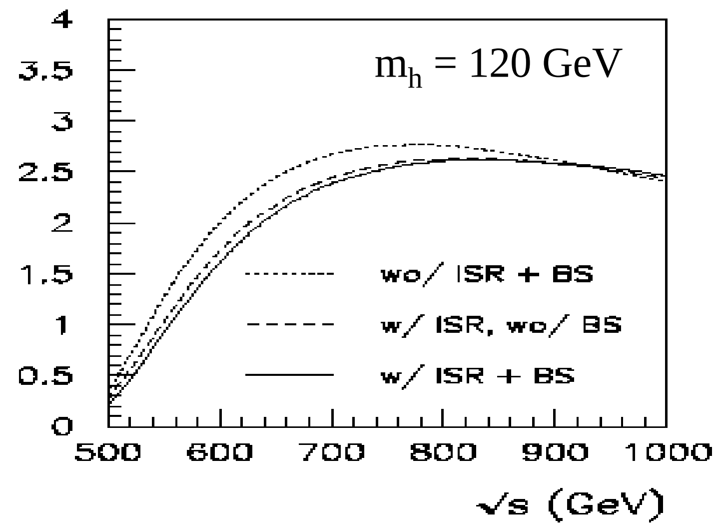
=> to a good approximation:

$$\sigma_{tth} \propto g_{tth}^2$$

Signal Processes (II)

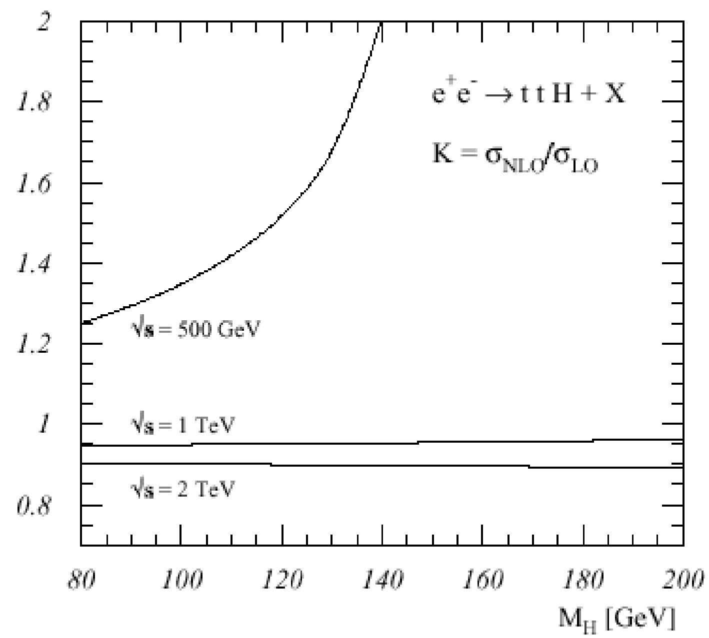
RADIATIVE CORRECTIONS

- QED: ISR (Bremsstrahlung) and Beamstrahlung.
  Significantly distort the tth lineshape.
  Main effect is to shift the maximum towards higher √s. At 500 GeV, $\sigma_{ttH}\downarrow$ by a factor ~2



- WEAK: in general small corrections.
  Can be taken into account by judicious choice of dressed couplings and propagators
  => Improved Born Approximation

- QCD: can be large corrections,
  Dominant effect is from rescattering diagrams generated by Coulombic gluons exchange between the top quarks near the tt threshold.

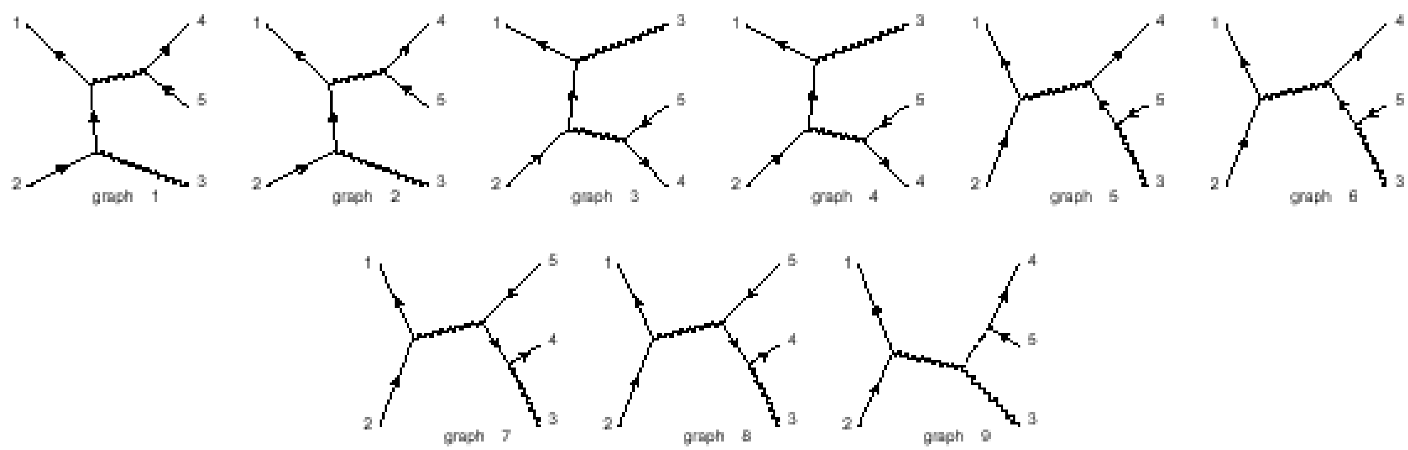


S. Dawson and L. Reina, Phys. Rev. D59 (1998) 054012
S. Dittmaier et al, Phys. Lett. B441 (1998) 383

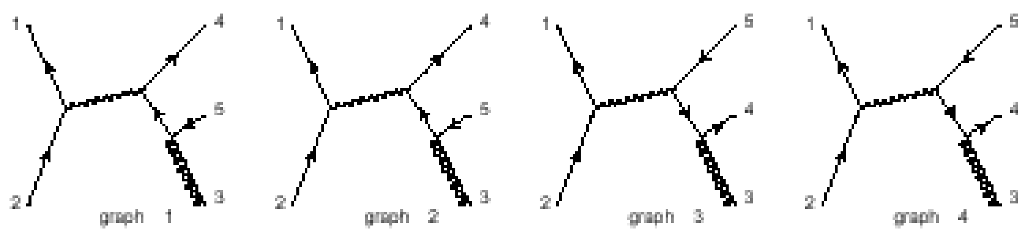
Background Processes (I)

Interfering:

- Electroweak: dominated by far by ttZ. Particularly dangerous if the Higgs mass is close to M_Z.

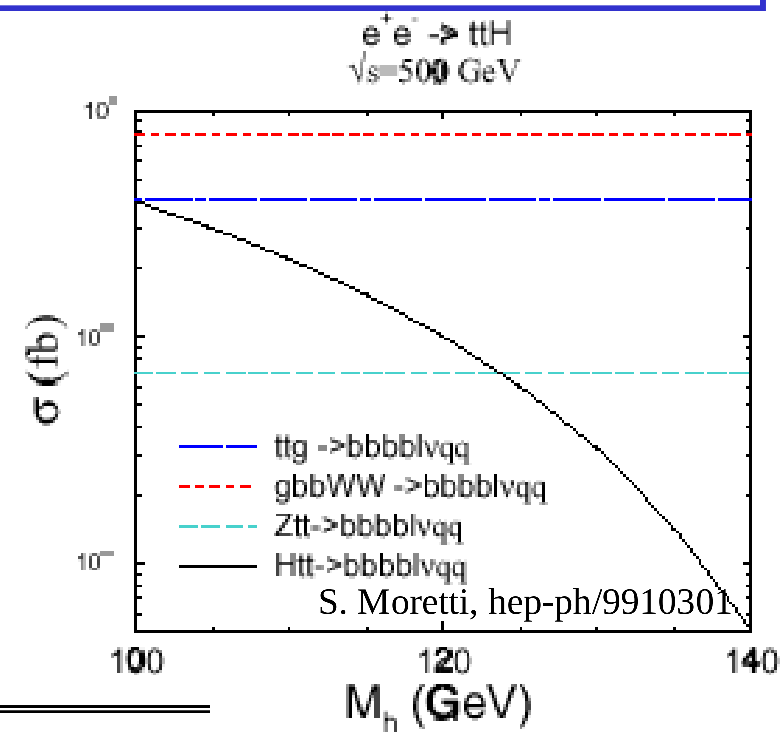


- QCD: formaly dominant interfering background, with g^*->bb. The g can be radiated from the t or b line. Since $m_t \gg m_b$, g off t is radiated at much larger angles (larger "dead cone effect").



Background Processes (II)

- Interfering backgrounds have cross-sections comparable to that of the signal.
- QCD interfering background can be greatly reduced by applying a cut to the bb invariant mass distribution (b-jet energy logarithmically enhanced at low values).



√s = 500 GeV, ISR+BS

2 →4 Process	**σ(fb)**	**σ_{cut}(fb)**
$e^+e^- \to tth \to ttbb$ (5 diags)	0.108	0.108
QCD: $e^+e^- \to ttg^* \to ttbb$ (8 diags, g off t)	0.586	0.021
EW : $e^+e^- \to ttZ \to ttbb$ (9 diags)	0.094	0.093
EW : $e^+e^- \to$ all but ttZ $\to ttbb$ (31 diags)	0.013	0.010
EW : $e^+e^- \to$ all $\to ttbb$ (40 diags)	0.106	0.094

10% interference effects!

CUT: min(E_b)>20 GeV, m_{bb}>50 GeV

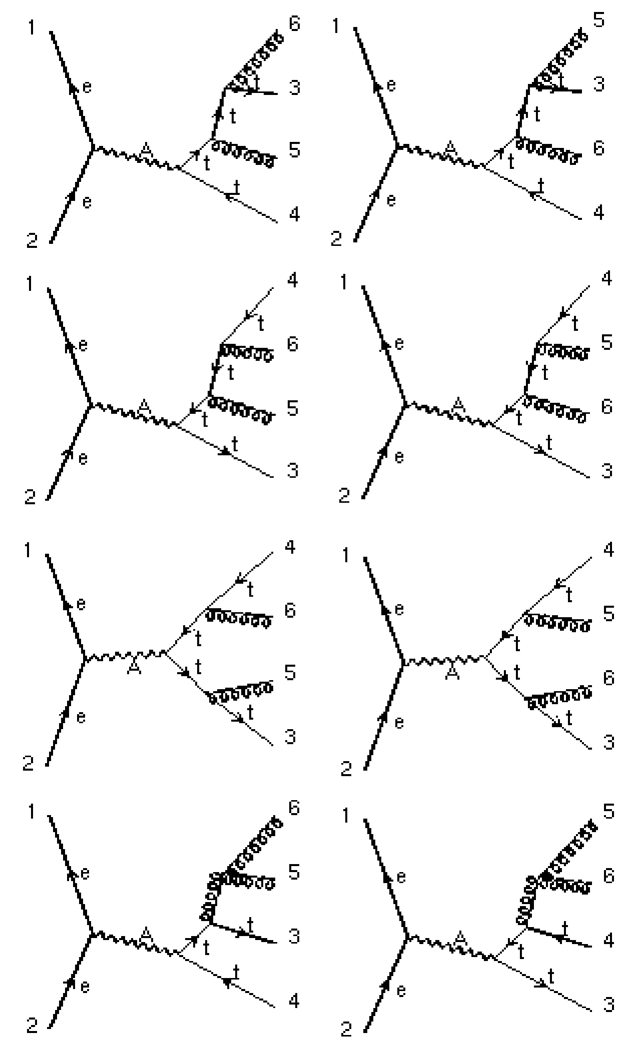
Background Processes (III)

Non-Interfering:

- Traditionally dismissed by most of previous analysis.
- Have formally lower number of partons in the final state, but due to hard gluon radiation, detector limitations and their huge cross-sections, enter the final selection and may eventually dominate.

Process	σ(fb) $\sqrt{s}$ = 500 GeV, ISR+BS
$e^+e^- \to tt$	521.0
$e^+e^- \to qq$ (n_f=5)	3951.8
$e^+e^- \to W^+W^-$	7890.4
$e^+e^- \to ZZ$	577.2
$e^+e^- \to ZH$	67.8

⇒ Need to be computed to higher orders in QCD in order not to jeopardize the measurement. Typically tt is the dominant contribution after selection.



Simulation Aspects: General Considerations

- Difficulty of the measurement requires high degree of sophistication in the analysis

  => aim at maximal use of kinematical information: e.g. angular correlations (see later), multijet invariant masses, etc.

  => Signal and main interfering and non-interfering backgrounds with proper spin transmission from the tops to the final state fermions.

  => make 2->n (n=6,7,8) generators.
- Interferences neglected so far. Will be examined in more detail at a later stage.
- Higher order QCD corrections:
  - Mainly relevant for signal and tt+X backgrounds.
  - Eventually should start making use of the K-factors for a more realistic estimate. Not used yet.
  - Concern about ME ⇔ PS matching/double-counting…

    Strategy so far (to be revised/improved):
    - Let PYTHIA handle gluon radiation from q=u,d,s,c,b => ME matching to PS already implemented.
    - Implement ME calculations for real gluon radiation from top up to NNLO. Let PYTHIA do PS on the final state, trying to avoid double-counting:

      e.g. veto PS from g in ttg events => already accounted for by $ttg^*(g^*\text{->}qq)$ and $ttg^*(g^*\text{->}gg)$ events.

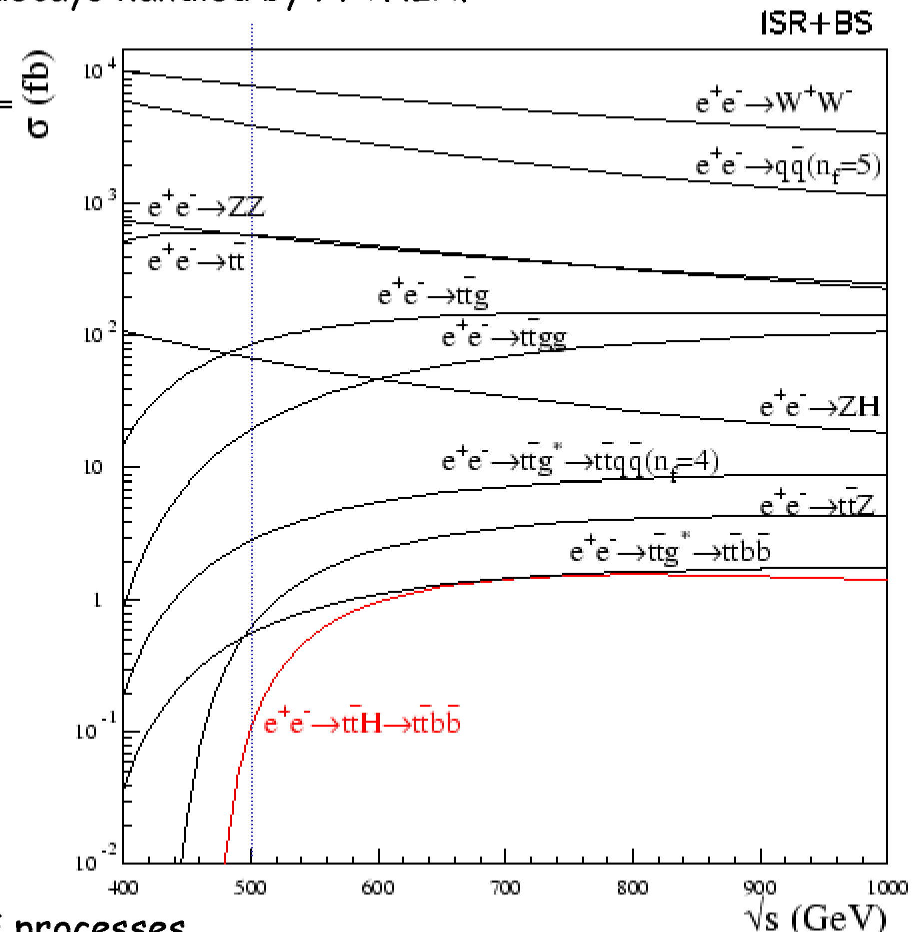
Simulation Aspects: Event Generation (I)

- $\sqrt{s}$=500 GeV; ISR (SF approach) and Beamstrahlung (CIRCE-NLC/JLC 2001) included.
- m_t=175 GeV, m_h=120 GeV
- Fragmentation, hadronization and particles' decays handled by PYTHIA.

	Process	σ(fb)	Nevt	L (ab^{-1})
ME CALCULATION	$e^+e^- \to tth \to 6f+h$	0.16	100k	621.1
	$e^+e^- \to tth \to 6f+bb$	0.11	100k	934.6
	Interfering backgrounds (interferences neglected)			
	$e^+e^- \to ttZ \to 8f$	0.58	100k	173.6
	$e^+e^- \to ttg^* \to 6f+bb$	2.10	100k	47.7
	$e^+e^- \to ttg^* \to 6f+qq$ (n_f=4)	27.49	100k	3.6
	Non-interfering backgrounds			
	$e^+e^- \to ttg \to 6f+g$	107.57	600k	5.6
	$e^+e^- \to ttgg \to 6f+2g$	28.98	150k	5.2
	$e^+e^- \to tt \to 6f$	Not used yet		
PYTHIA	$e^+e^- \to tt$	521.0	10M	19.2
	$e^+e^- \to qq$ (n_f=5)	3951.8	10M	2.5
	$e^+e^- \to W^+W^-$	7890.4	10M	1.3
	$e^+e^- \to ZZ$	577.2	2M	3.5
	$e^+e^- \to ZH$	67.8	2M	29.5

(*) $\sqrt{s}$ > 100 GeV for all processes

(**) "Canonical cuts": E_i^{LAB} > 5 GeV. m_{ij}> 5 GeV for ME processes



Simulation Aspects: Detector Simulation

- Fast simulation of the response of a "Large Detector" for NLC (SIMDET v3.1):

SUBDETECTOR	RESOLUTION	ACCEPTANCE	GRANULARITY
Vertex Detector (CCD in 1 cm beampipe)	$\sigma_{xy(rz)} = 3.8\mu m \oplus 6.3\mu m/(p \sin^{3/2}\theta)$ in central region: $26^0 < \theta < 90^0$	$9.5^0 < \theta < 90^0$	
Tracker System B=2T (main tracker, forward tracker and forward muon tracker)	$\Delta p_T/p_T \leq 0.8 \times 10^{-4}$ For $p \geq 50$ GeV/c	$\theta > 5^0$	
EM Calorimeter HAD Calorimeter	$\sigma_E/E = 10\%/\sqrt{E} \oplus 1\%$ $\sigma_E/E = 40\%/\sqrt{E} \oplus 2\%$		0.9^0 2^0
TOTAL		$\theta > 57$mrad	

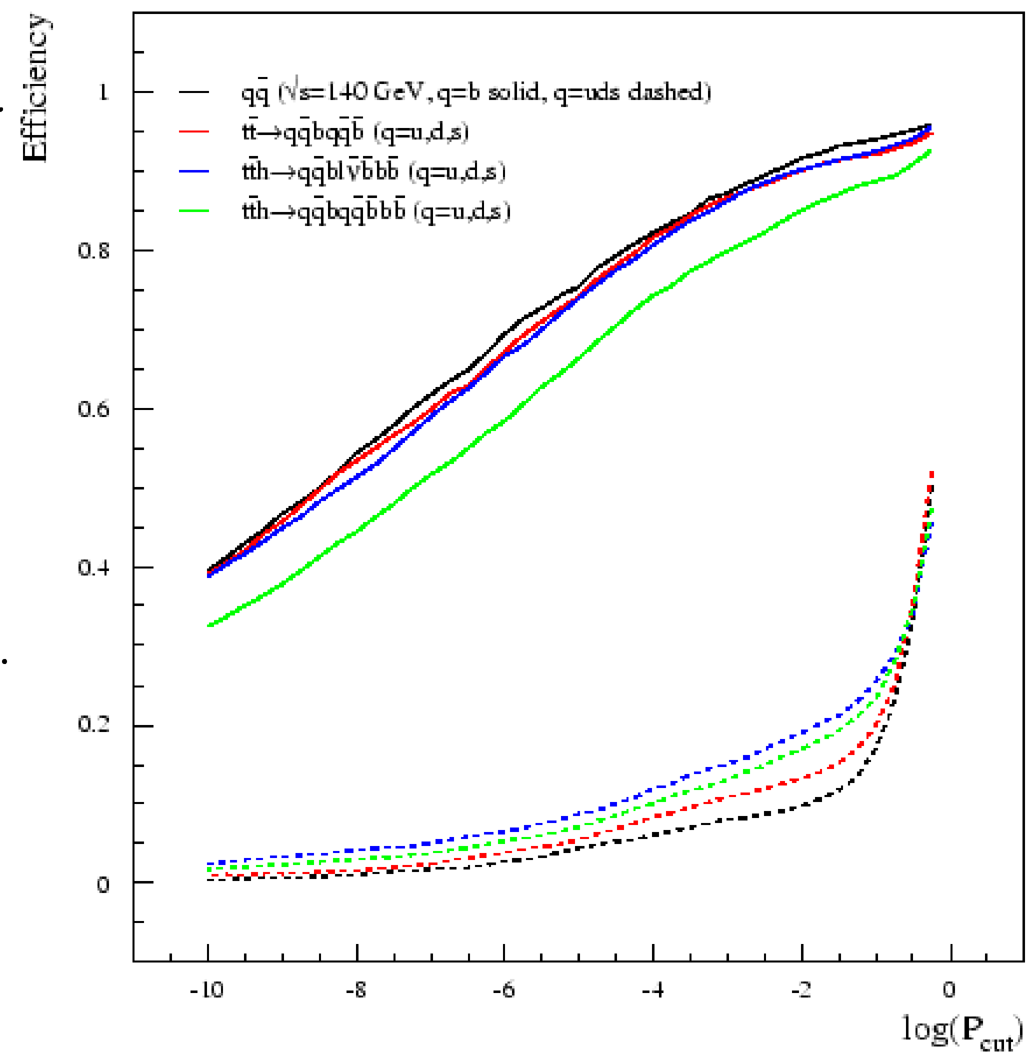
- Parameterization of EM and HAD shower deposits obtained from a full GEANT simulation. Includes calorimeter cluster finding algorithm.
- Pattern recognition emulated.
- Energy Flow algorithm incorporating track-cluster matching.

Experimental Aspects: b-tagging

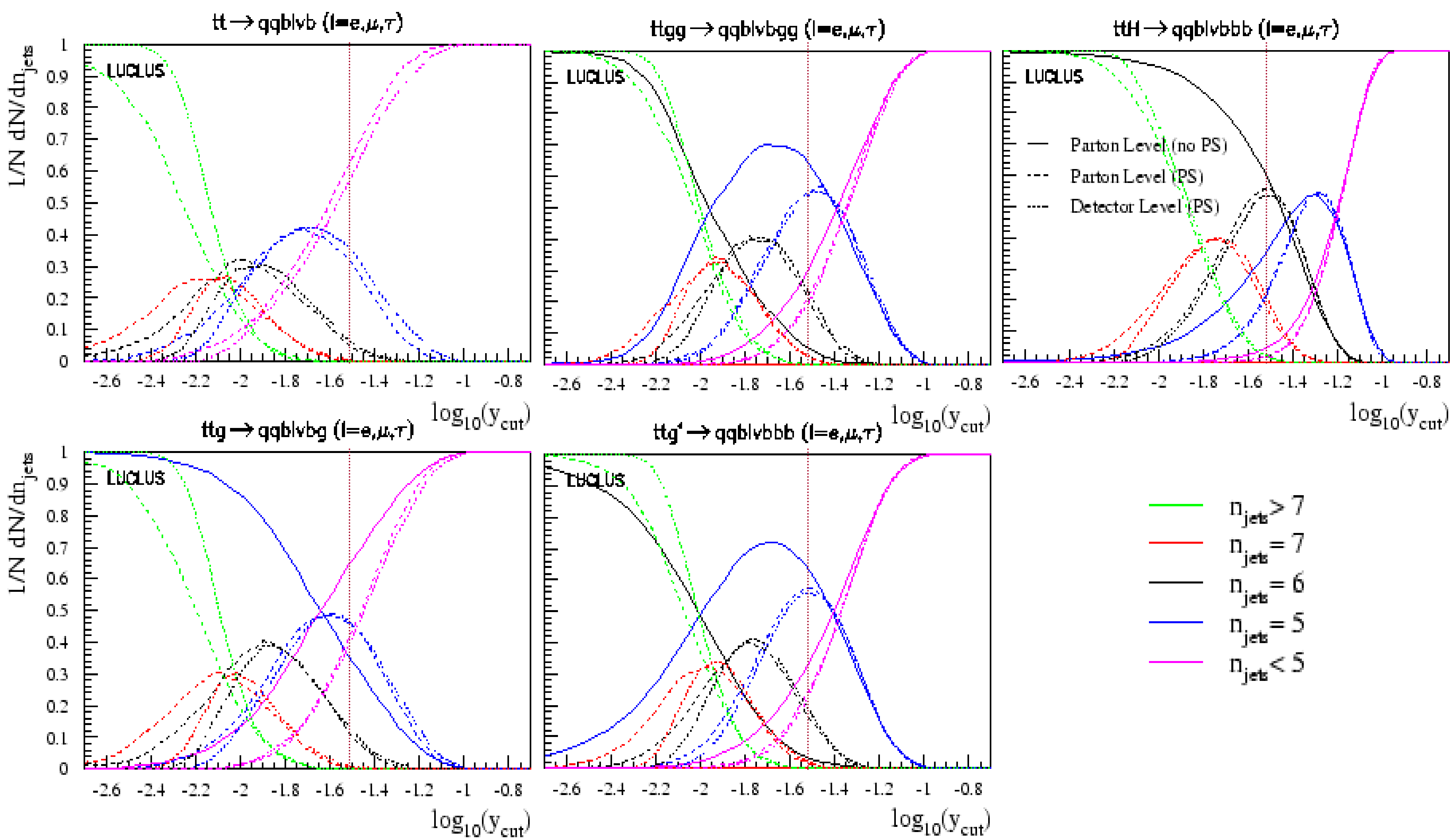
- Probably the most crucial experimental issue.
- Tag B hadrons lifetime using VDET
- Parameterize the pdf of impact parameter significance for no-lifetime tracks in $Z \to qq$ (q=uds).
- Assign a probability of "being consistent with coming from primary vertex" to each track/jet/event.
- B-tagging algorithm is kept simple so that success of the analysis doesn't depend on detector details.

  Rather "standard performance":

  $\varepsilon_b \sim 50\%$ with $\varepsilon_{uds} \sim 1\%$
- More efficient algorithms can be developed (e.g. multivariate b-tagging)



Experimental Aspects: hard gluon radiation



Experimental Aspects: Sensitivity

- Expected statistical and (some) systematic uncertainties on g_{tth} for a given selection with efficiency ε and purity ρ:

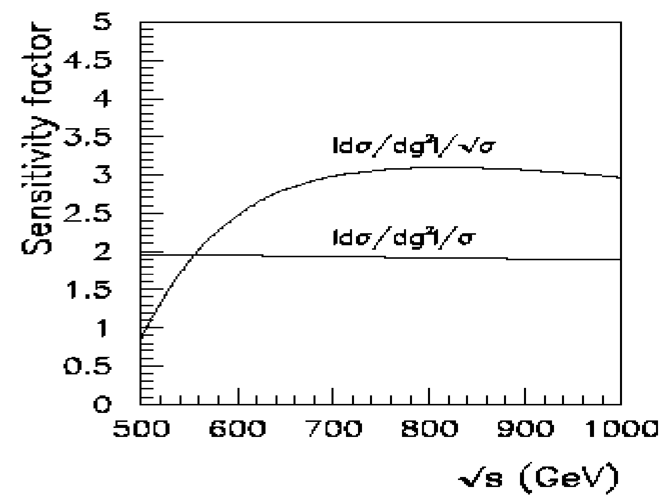
$$\left(\frac{\Delta g_{tth}}{g_{tth}}\right)_{stat} \cong \left(\Delta g_{tth}^2\right)_{stat} = \frac{1}{S_{stat}\left(g_{tth}^2\right)\sqrt{\varepsilon\rho L}}$$

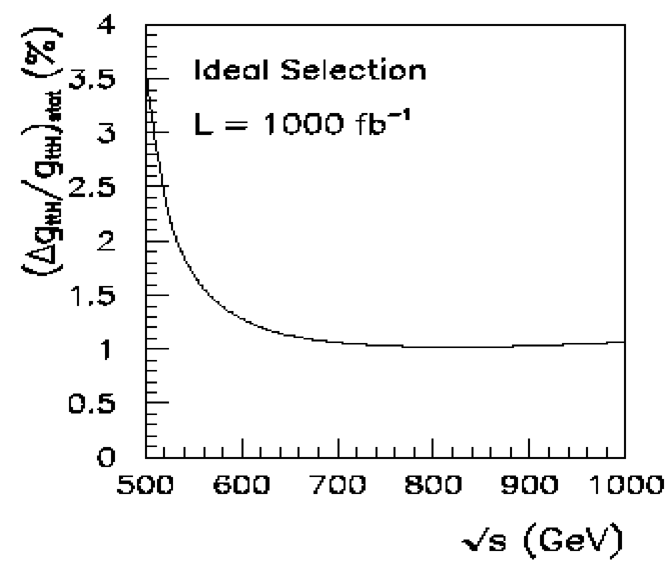
$$\left(\frac{\Delta g_{tth}}{g_{tth}}\right)_{syst} \cong \left(\Delta g_{tth}^2\right)_{syst} = \frac{1}{S_{syst}\left(g_{tth}^2\right)}\left[\frac{1-\rho}{\rho}\frac{\Delta\sigma_{bckg}^{eff}}{\sigma_{bckg}^{eff}} \oplus \frac{1}{\rho}\frac{\Delta L}{L} \oplus \frac{\Delta\varepsilon}{\varepsilon}\right]$$

$$S_{stat}\left(g_{tth}^2\right) = \frac{1}{\sqrt{\sigma_{tth}}}\left|\frac{d\sigma_{tth}}{dg_{tth}^2}\right|$$

$$S_{syst}\left(g_{tth}^2\right) = \frac{1}{\sigma_{tth}}\left|\frac{d\sigma_{tth}}{dg_{tth}^2}\right|$$

- Sensitivity a factor of ~3.4 lower at 500 GeV than at 800 GeV.
- Systematic from uncertainty in background normalization will dominate total error on g_{tth} unless high purity is achieved.





Experimental Aspects: Strategy

- The measurement requires high luminosity: assume L = 1000 fb^{-1}
- So far have considered only semileptonic channel.
- Selection optimized for $h \to bb$ (we treat the Higgs decay inclusively).
- No τ-ID => further loss of efficiency
- Despite the apparently clean signature of the signal, the measurement has many difficulties:
  - Very limited signal statistics (expect ~70.8 tth semileptonic decays with L = 1000 fb^{-1} BEFORE selection!!)
  - Backgrounds orders of magnitude larger than the signal (e.g. tt alone is >3 orders of magnitude larger).
  - Limitations of jet-clustering algorithms in properly reconstructing ≥6 jet in the final state: hard gluon radiation, jet mixing, etc
  - B-tagging far from optimal. Not fully exploiting the potential of precise VDET.
  - …

PHILOSOPHY:

Soft Preselection +

Multivariate Analysis (to optimally increase the statistical sensitivity of the signal) +

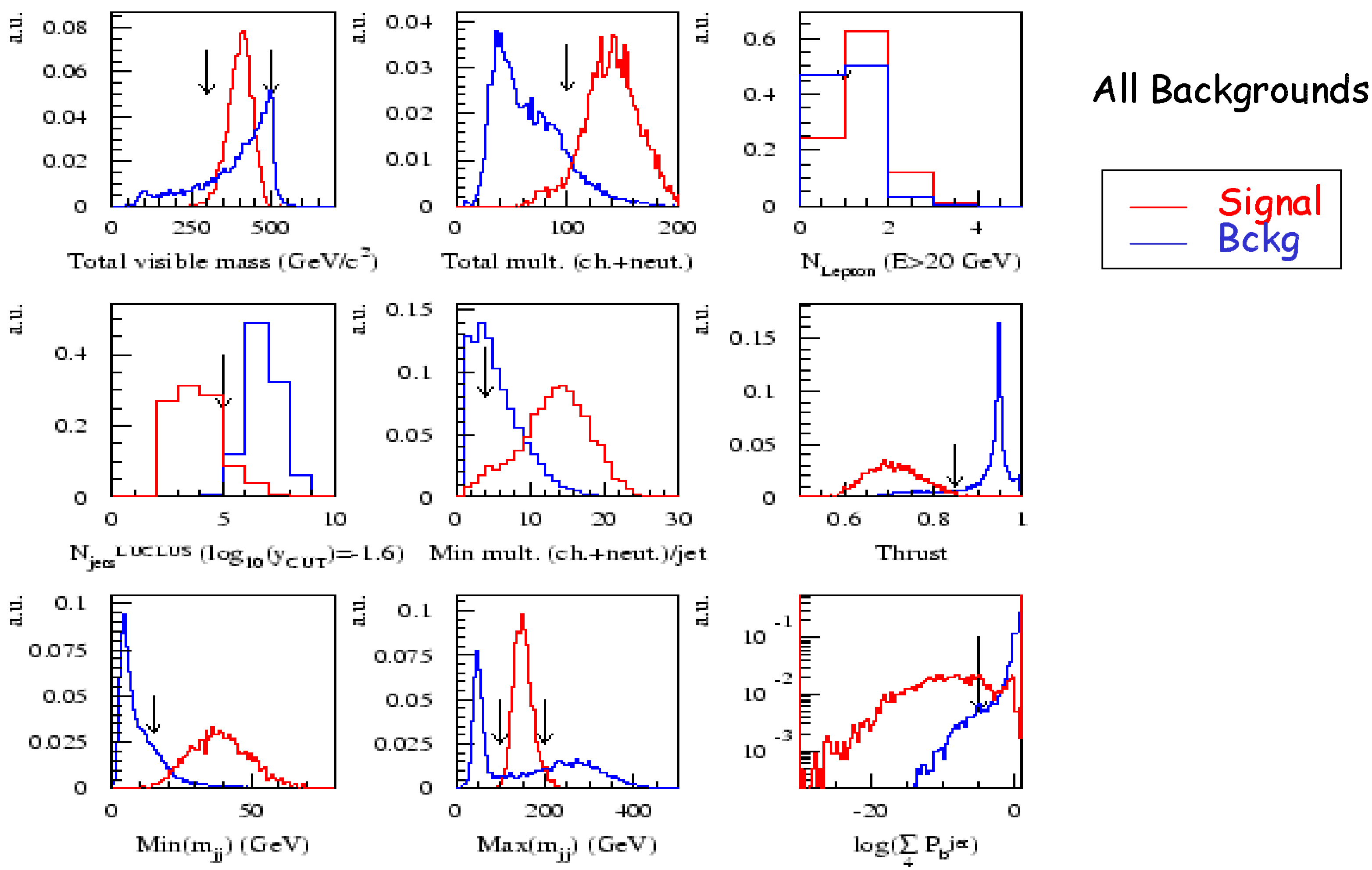
Maximal Use of kinematical information

Done

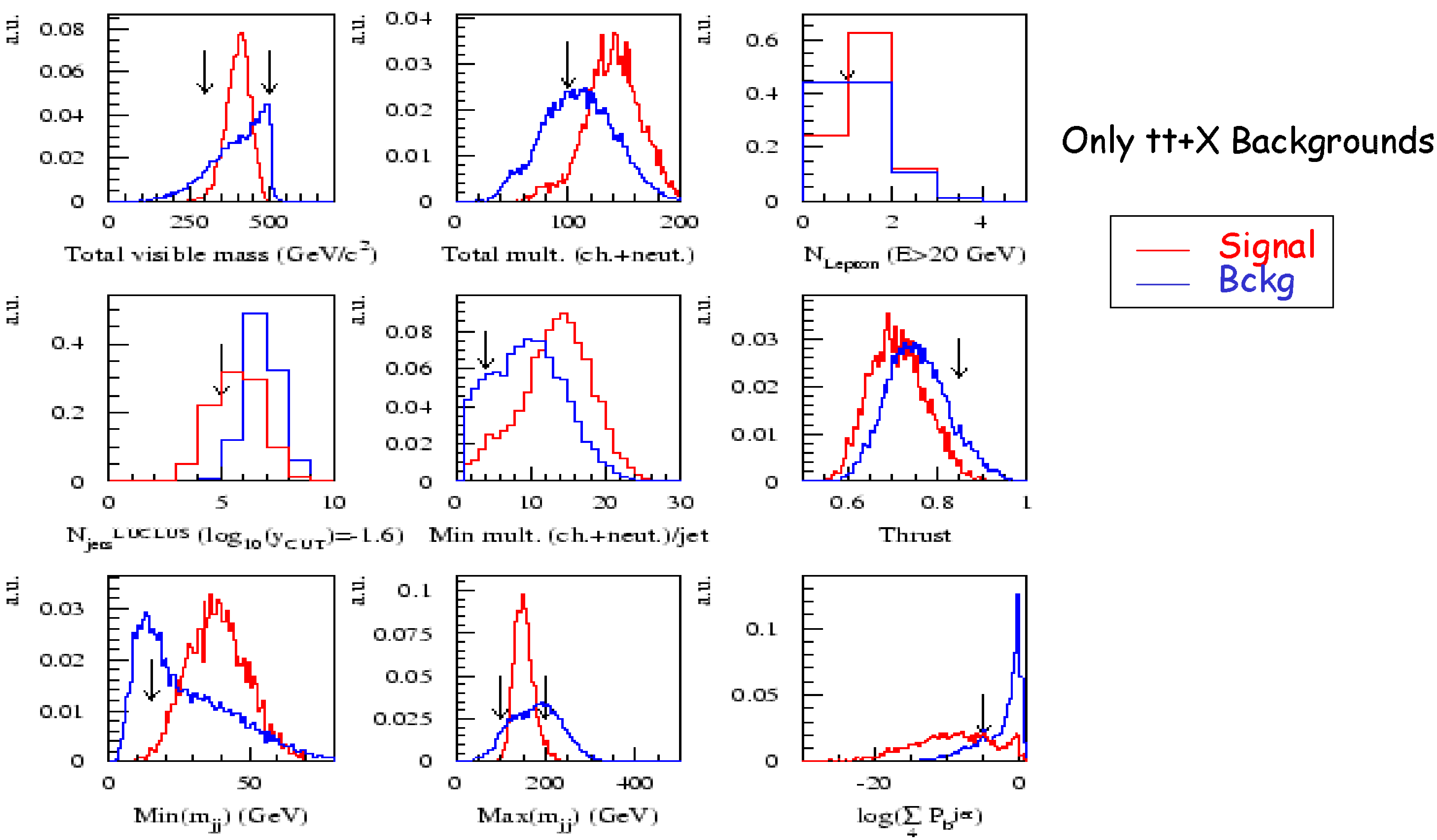
Not yet

Semileptonic Channel: Preselection (I)

- $\geq$1 e/μ E_l>20 GeV. Choose the one with highest energy.
- Force the rest of the event to 6 jets using LUCLUS



Semileptonic Channel: Preselection (II)



Semileptonic Channel: Preselection (II)

Process	σ(fb)	ε_{prsel}	σ_{eff}(fb)	Nevt(L=1 ab^{-1})
$e^+e^- \to tth \to$ semilept	0.0708	44.53%	0.0315	31.5
$e^+e^- \to tth \to$ had+lept	0.0902	4.88%	0.0044	4.4
			Total Signal:	35.9
$e^+e^- \to tt$	521.0	3.62e-3	1.8860	1886.0
$e^+e^- \to ttg \to 6f+g$	107.57	7.56e-3	0.8132	831.2
$e^+e^- \to ttgg \to 6f+2g$	28.98	1.73e-2	0.5014	501.4
$e^+e^- \to ttZ \to 8f$	0.58	7.03e-2	0.0405	40.5
$e^+e^- \to ttg^* \to 6f+bb$	2.10	4.37e-2	0.0917	91.7
$e^+e^- \to ttg^* \to 6f+qq$ (n_f=4)	27.49	1.46e-2	0.4027	402.7
$e^+e^- \to qq$ (n_f=5)	3951.8	2.70e-6	0.0107	10.7
$e^+e^- \to W^+W^-$	7890.4	1.03e-5	0.0813	81.3
$e^+e^- \to ZZ$	577.2	6.50e-6	0.0038	3.8
$e^+e^- \to ZH$	67.8	2.00e-5	0.0014	1.4
			Total Background:	3850.7

- After preselection: ε_{prsel} = ***44.53%*** and ***N/S~100***.
- Inclusive leptonic selection results in low efficiency in τ channel:

$$\varepsilon_{prsel}(tth \to qqbl\nu bbb;\ l=e,\mu) = 68.27\%$$
$$\varepsilon_{prsel}(tth \to qqbl\nu bbb;\ l=\tau) = 12.70\%$$

Semileptonic Channel: Multivariate Analysis

- Need to optimize selection in order to minimize <u>total uncertainty.</u>
- Efficient discrimination between signal and backgrounds can only be accomplished by combining many variables.
- NN techniques:
  - Optimal use of information in a set of discriminant variables: x_i, i=1,..,N
  - 1-D projection without loss of sensitivity
  - Statistical interpretation in terms of a-posteriori Bayesian probability

  ⇒ *A* very convenient feature:

  ⇒ Optimal test of hypothesis

  ⇒ Train a NN to disentangle between signal and only one tt+X background at a time

  e.g. tth->qqblvbbb vs tt-> all

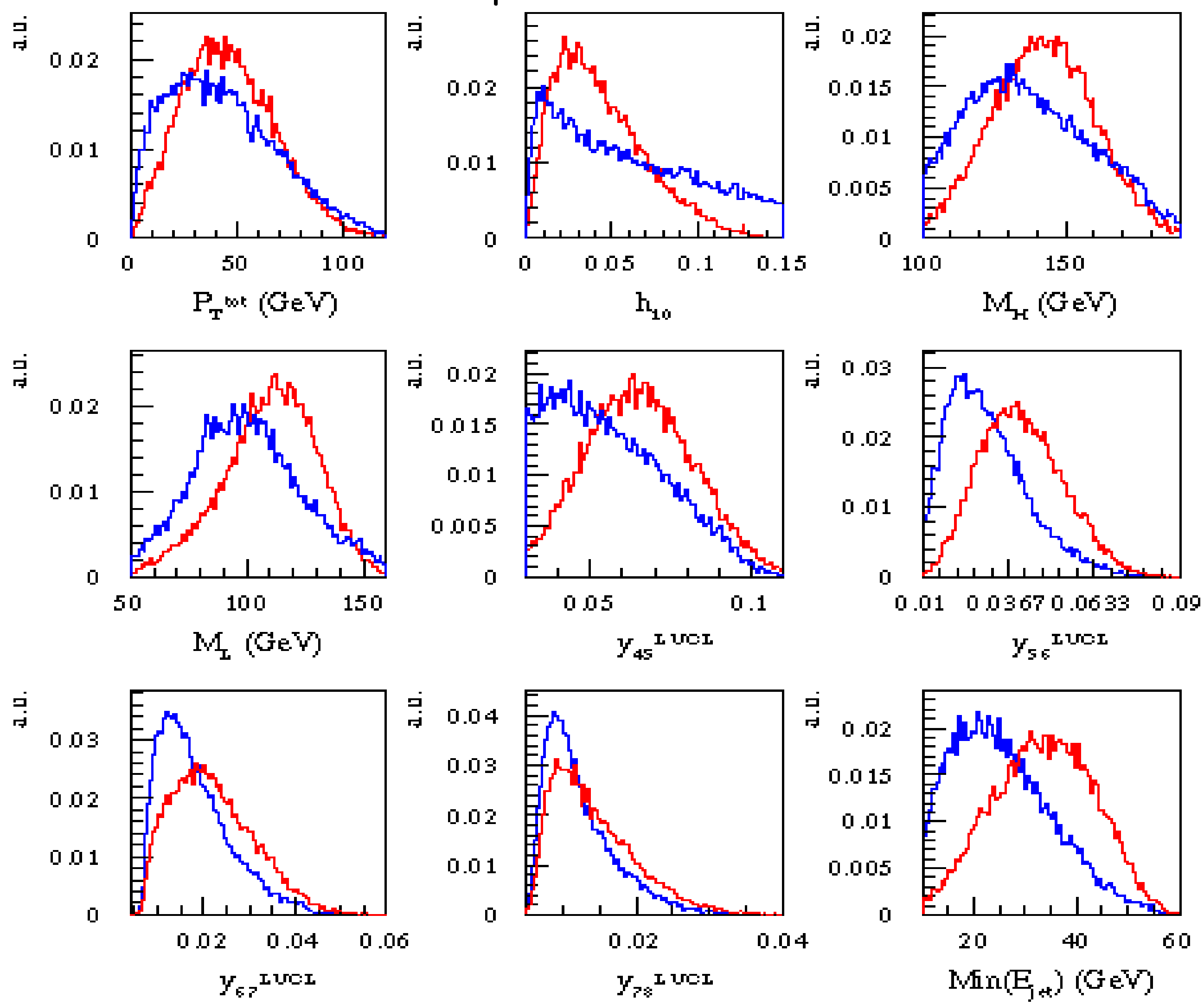
$$O_{tt}(\vec{x}) \cong P_{\text{tth vs tt}}(\text{tth} \rightarrow \text{qqbl}\nu\text{bbb} \mid \vec{x}; \rho_{tth} = \rho_{tt} = 0.5)$$

⇒ Build the correct Bayesian probability taking into account all background classes:

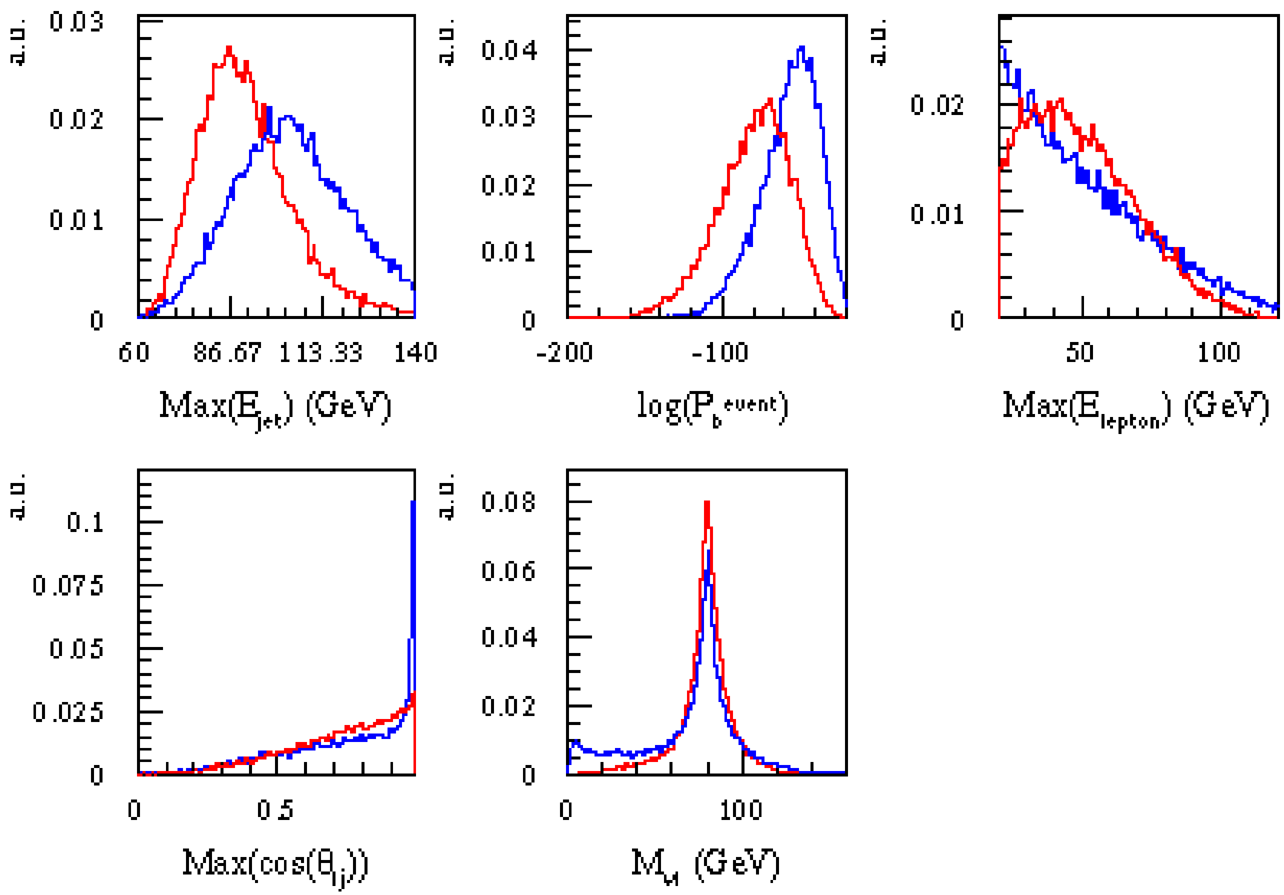
$$P(\text{tth} \rightarrow \text{qqbl}\nu\text{bbb} \mid \vec{x}) = \left[1 + \sum_i \frac{\sigma_i^{eff}}{\sigma_{tth}^{eff}} \left(\frac{1 - O_i(\vec{x})}{O_i(\vec{x})}\right)\right]^{-1}$$

Semileptonic Channel: Multivariate Analysis

- Select a set of 23 variables: 9 preselection variables + 14 more.

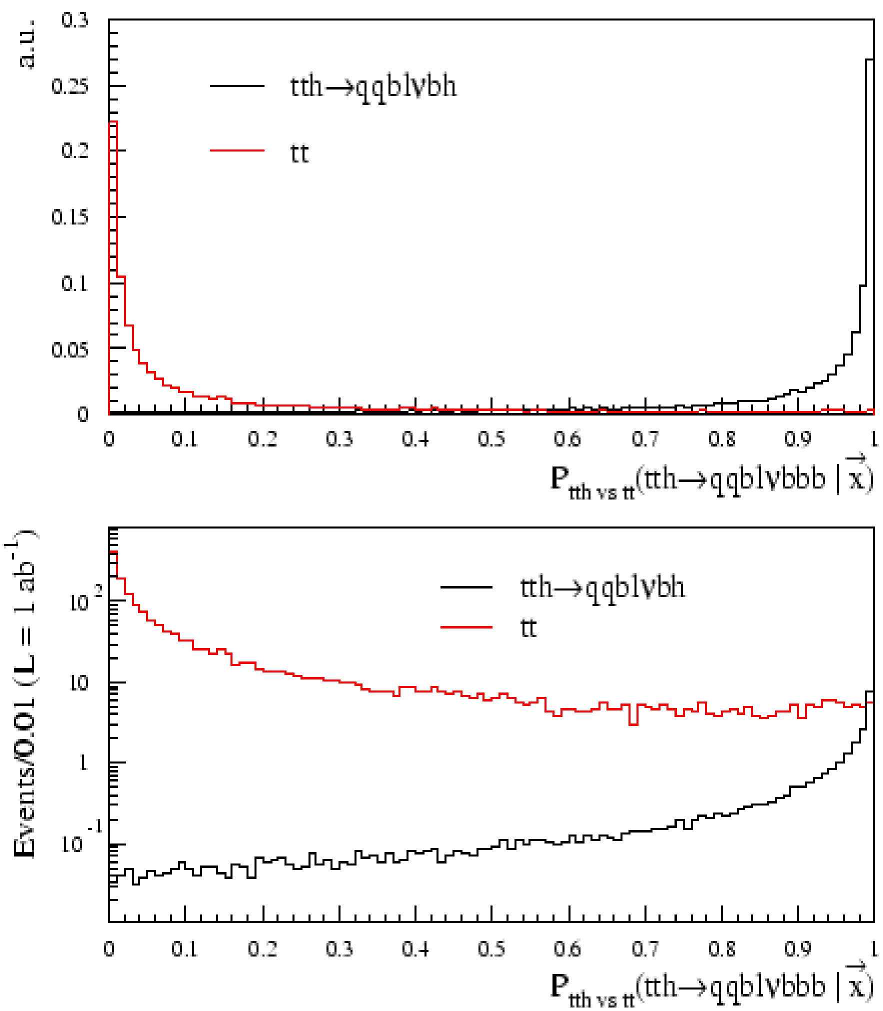


Semileptonic Channel: Multivariate Analysis

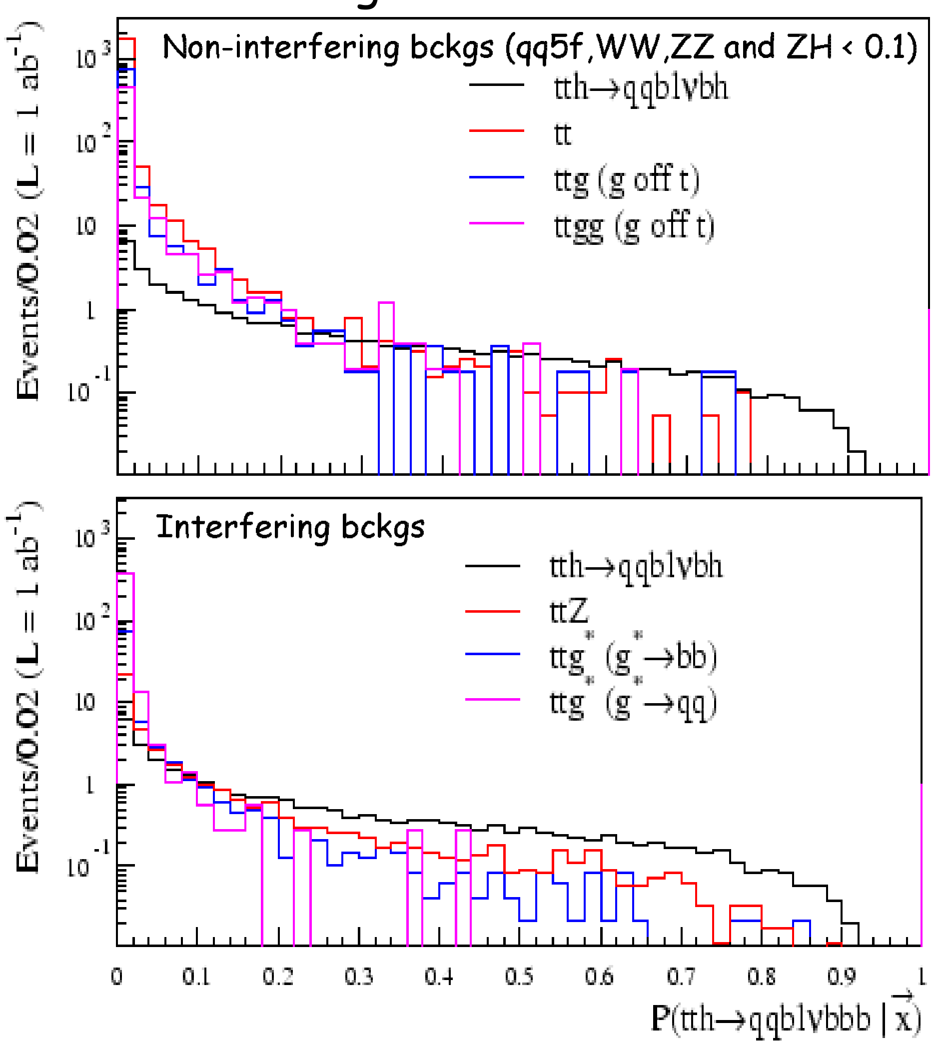


Semileptonic Channel: Multivariate Analysis

tth vs tt only



The global discriminant



Semileptonic Channel: Multivariate Analysis

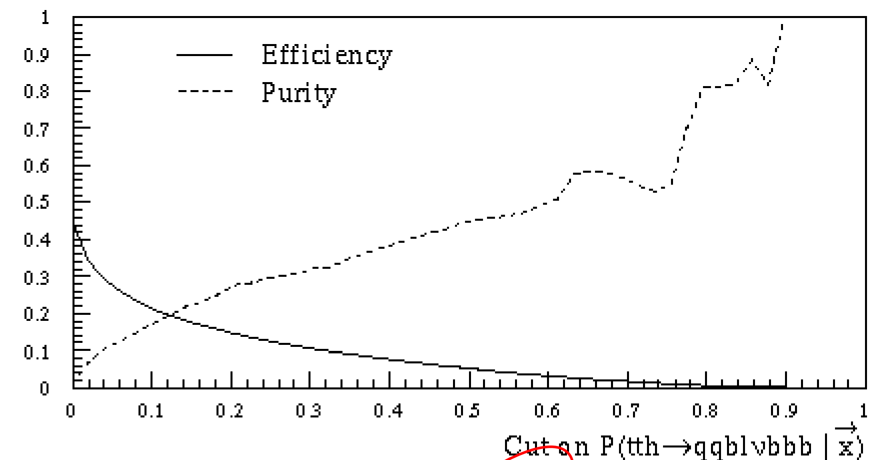
- By cutting on the Bayesian probability, the ε and ρ are optimally varied (given the set of variables used) and can look for the minimum in total uncertainty.
- From the semileptonic channel, and assuming L = 1000 fb^{-1}:

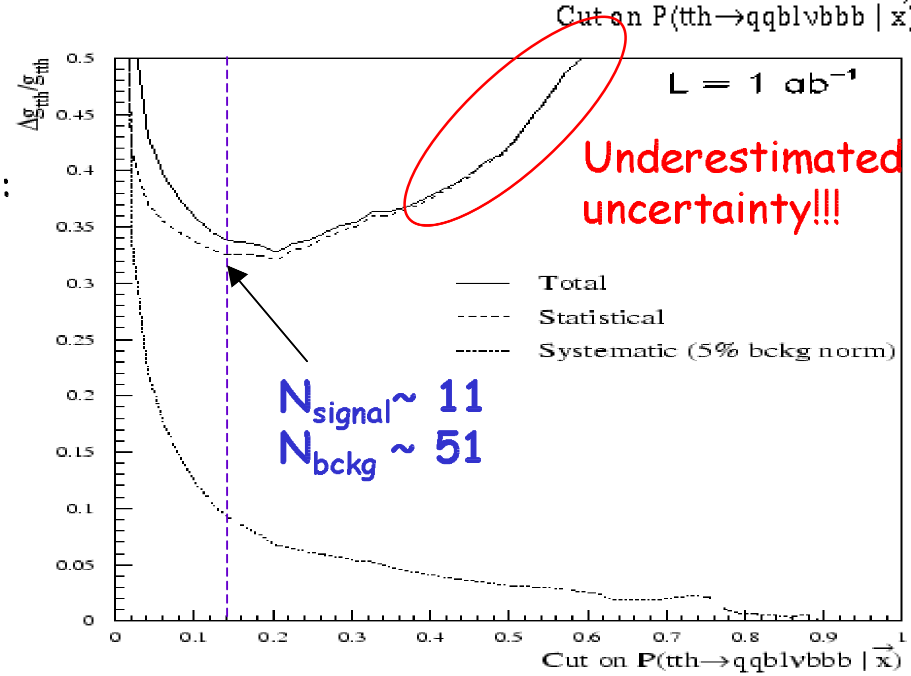
$$\left(\frac{\Delta g_{tth}}{g_{tth}}\right)_{stat} \approx 33\%$$

- Assuming $\sqrt{2}$ improvement from the combination with the fully hadronic channel:

$$\left(\frac{\Delta g_{tth}}{g_{tth}}\right)_{stat} \approx 23\%$$

=> believe this result can be further improved…



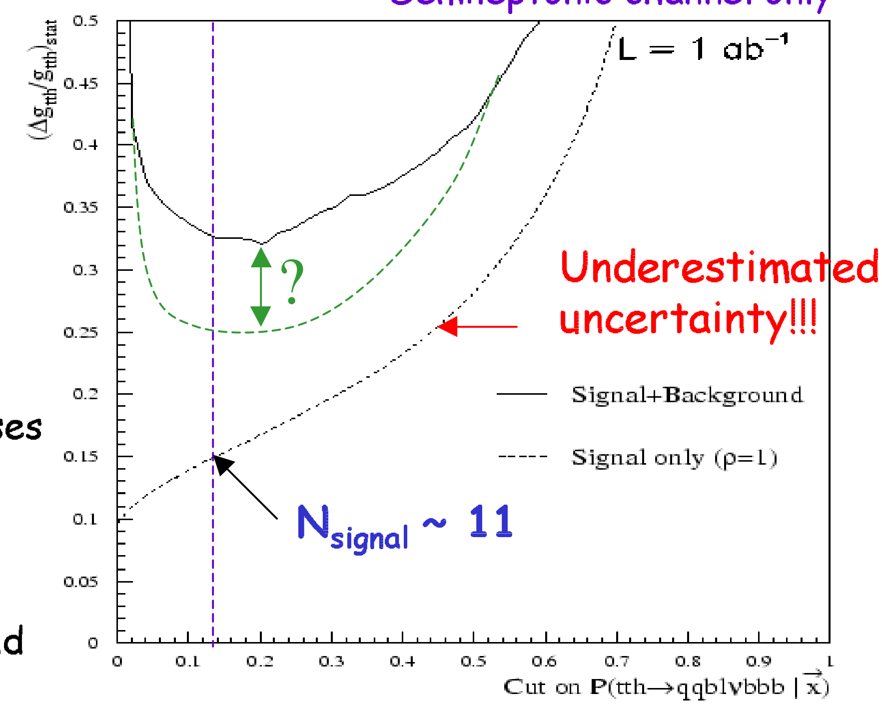


How could we improve?

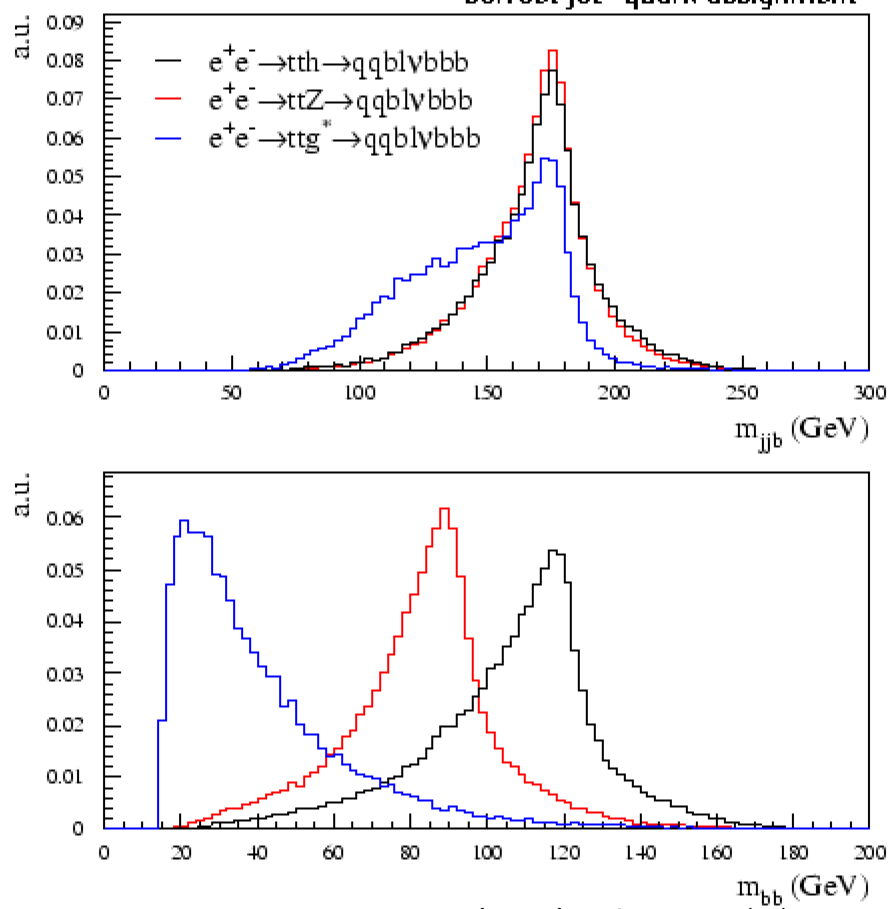
- Didn't take into account NLO QCD K-factor for tth:
  K~1.5 for m_h=120 GeV and $\sqrt{s}$ = 500 GeV
  => 22% decrease in statistical uncertainty.
  => actually K will probably be >1.5 due to the effect of ISR+BS

  S. Dawson and L. Reina, Phys. Rev. D59 (1998) 054012
  S. Dittmaier et al, Phys. Lett. B441 (1998) 383
- Dedicated selection for τ channel.
- Improve b-tagging: rather inclusive and therefore not fully efficient.
  Won't change it to keep the analysis "conservative" to some extent.
- Make explicit use of kinematical information:
  - Apply energy dependent corrections to jet energies (in particular b-jets) in order to correct for ν-related energy losses
    => crucial to improve bb invariant mass resolution!!!
  - Avoid use of "best masses":
    as the number of jet increases background peaks more and more in the signal region.
  - Make use of angular distributions: should allow to disentangle between tth, ttZ and misreconstructed tt events.

Semileptonic channel only

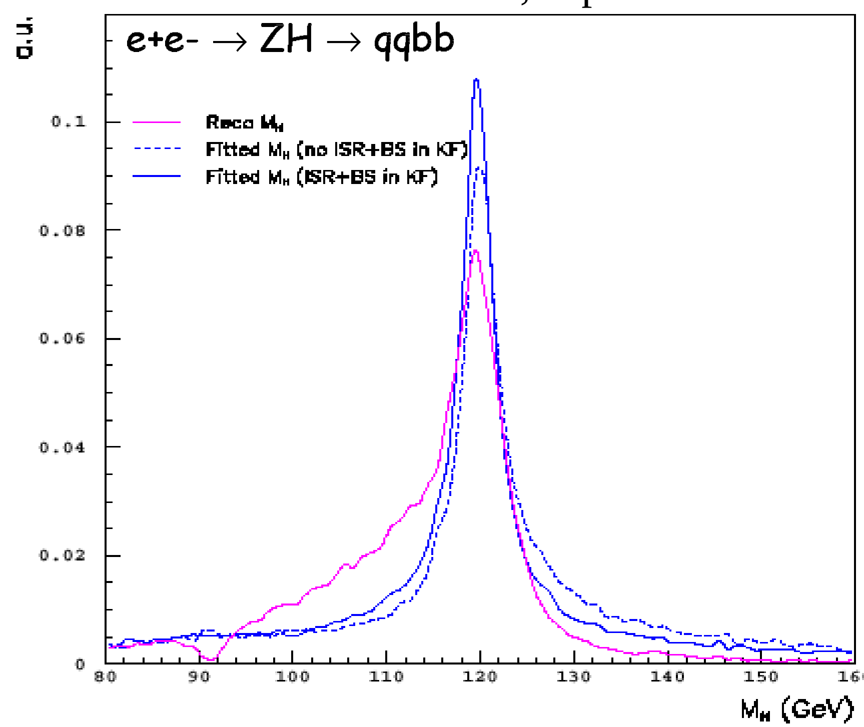


Multijet invariant masses



Potential improvement from b-jet energy corrections:

A. Juste, hep-ex/9912041



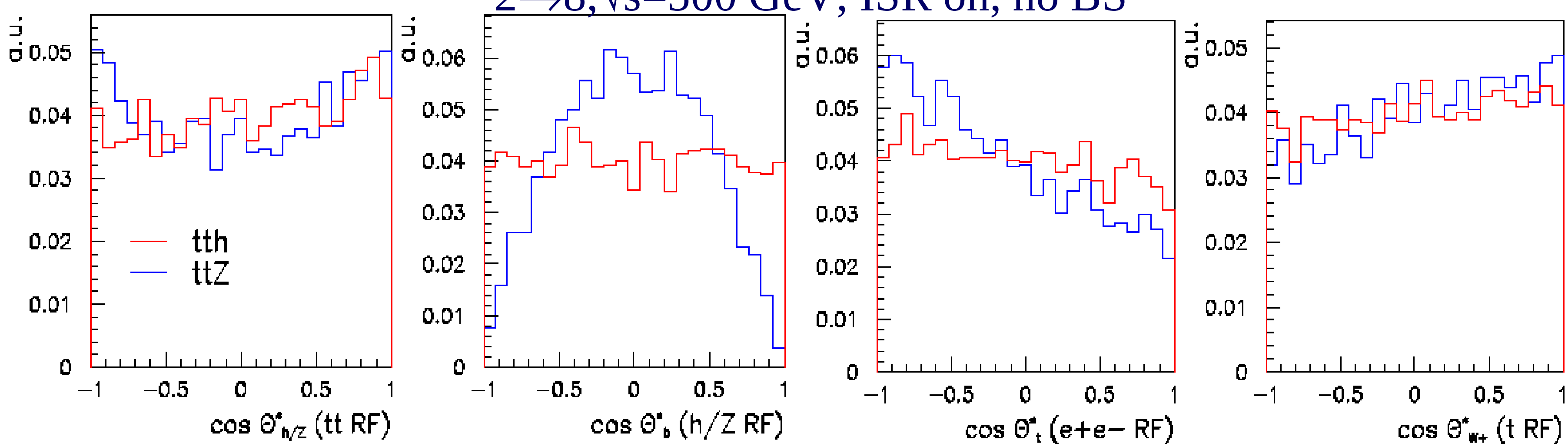
- To maximize signal-to-background discrimination from invariant masses:
  => improve multijet invariant mass resolution
  => reduce combinatorial background: use of b-tagging crucial!!
  12 (perfect b-tagging) vs 180 (no b-tagging) jet permutations in semileptonic channel

Angular Distributions/Correlations in ttX

- Angular distributions/correlations in misreconstructed tt events probably very different from nominal, whereas they are preserved to some extent in properly reconstructed ttX* (X*→bb), with X* = h, Z or g.
- Angular distribution of b in h/Z rest frames should help to discriminate between tth and ttZ.
- Combinatorics even more problematic here due to q<->$\bar{q}$ ambiguity (flavor tagging for W->cs, quark-charge tagging, etc):

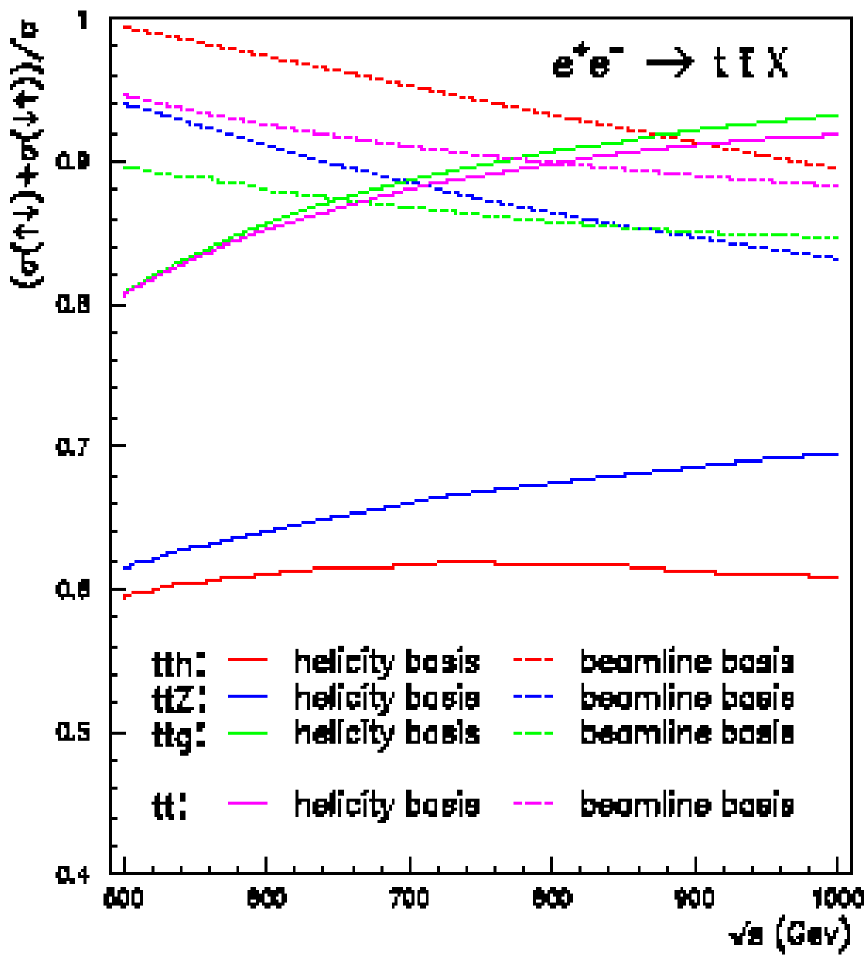
Possible permutations: 6! = 720
Perfect b-tagging, no q-flavor tagging: 48
Perfect b-tagging, perfect q-flavor tagging: 12

2→8,√s=500 GeV, ISR on, no BS

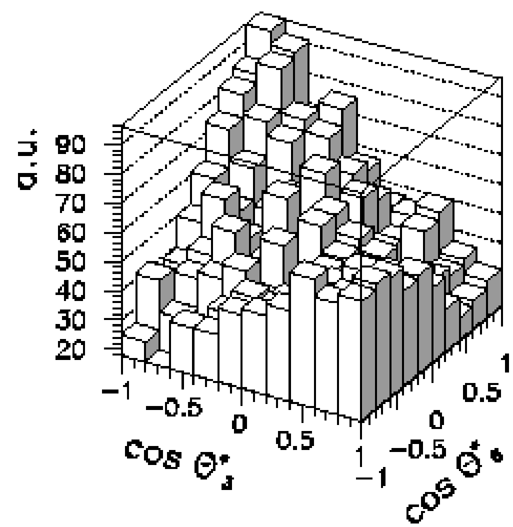


Angular Distributions/Correlations in ttX

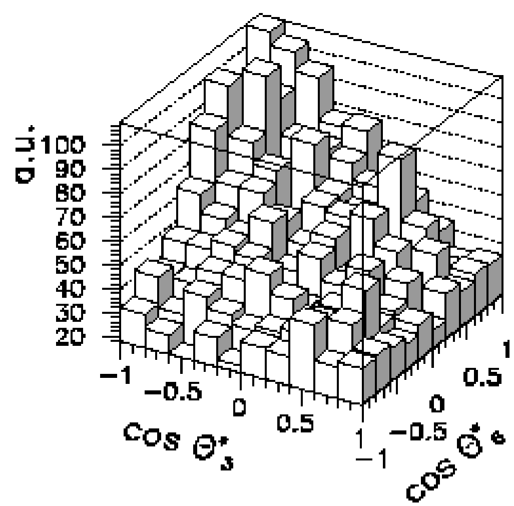
- Comparing results in different spin quantization axes might help to discriminate among the different interfering backgrounds.



Beamline Basis



tth



ttZ

Summary and conclusions

- First attempt at a realistic estimate on the direct determination of the top-Higgs Yukawa coupling at $\sqrt{s}$=500 GeV => first likely energy stage of the NLC.
- Improved (more complete!!) background simulation (interfering + non-interfering processes) but just getting started, several issues pending:
  - proper implementation of higher order QCD corrections (in particular for the dominant tt+jets background):
    - how to properly internormalize tt, ttg and ttgg? (effect of missing virtual+soft corrections?)
    - implement $\alpha(Q^2)$, Γ_t^{NLO},...
  - K_{NLO} for tth and ttZ.
  - effect of interferences between backgrounds (EW and QCD).
  - ...
- Measurement rather challenging: high luminosity+sophisticated analysis required:
  - done so far: multivariate selection

    =>

$$\left(\frac{\Delta g_{tth}}{g_{tth}}\right)_{stat} \approx 33\%$$

semileptonic channel
m_h= 120 GeV
$\sqrt{s}$=500 GeV, L = 1 ab^{-1}

Summary and conclusions (cont'd)

- *next step* is to try to make an exhaustive use of the kinematical information: multijet invariant masses, angular correlations, etc
- believe the estimate is conservative:
  - improve b-tagging
  - dedicated τ selection
  - etc

- A priori it looks it should be possible to ultimately obtain:

$$\left(\frac{\Delta g_{tth}}{g_{tth}}\right)_{stat} \leq 20\%$$

hadronic+semileptonic channels
m_h= 120 GeV
$\sqrt{s}$=500 GeV, L = 1 ab^{-1}