



Heavy flavor tagging and collider searches for stop and sbottom

D. Bortoletto
Purdue University

**Help from: Rott, Gonzalez- Lopez,
Silvestris, Cucciarelli, Palla, Rozanov,
Wright, Incliffe, Duperrin, Greenlee,
Guimaraes da Costa, Watts and more**

daniela.bortoletto



TEV4LHC



Outline

- **Searches for stop and bottom**
 - **Current status at CDF and D0**
 - **Plans at CMS and Atlas**
- **The role of heavy flavor tagging**
 - **Algorithms**
 - **CDF and D0 experience**
 - **Differences between TeV and LHC**
- **What CMS and Atlas want to know from the Tevatron?**



Third generation squarks

- Third generation squarks could be light, due to large mixing

$$\tilde{q}_1 = \tilde{q}_L \cos \Theta_{\tilde{q}} + \tilde{q}_R \sin \Theta_{\tilde{q}}$$

Light **Sbottom**: large $\tan\beta$

Light **Stop**: large m_t

Mass matrix mixing term

$$m_q (\mathbf{A}_q - \mu\kappa)$$

$\kappa = \tan\beta$
down type
quarks

$\kappa = 1/\tan\beta$
up type
quarks

- To suppress FCNC and CP violation might require that first and second generation of squarks is heavy ($M \sim 20$ TeV)
- Stop and sbottom should still be much lighter ($M < 1$ TeV) to maintain naturalness

Cohen, Kaplan,
Nelson

Stop and sbottom searches RUN 1

- mSUGRA ($m_0, m_{1/2}, A_0, \tan\beta, \mu$)
- Most analysis assumed $R_p = (-1)^{3B+L+2}$ conservation \Rightarrow stable LSP \Rightarrow MET
- Few used b-tagging

Top to Stop PRD 63, 091101(R) (2001)

$p\bar{p} \rightarrow t\bar{t} \rightarrow (\tilde{t}\tilde{\chi}_1^0)(\tilde{t}\tilde{\chi}_1^0) \rightarrow (b\tilde{\chi}_1^\pm)\tilde{\chi}_1^0(\bar{b}\tilde{\chi}_1^\mp)\tilde{\chi}_1^0$
Signature: ≥ 1 lepton, ≥ 2 jets, MET

Stop/Sbottom in c/b+LSP PRL 84, 5704 (2000)
PRL 93, 011801 (2004)

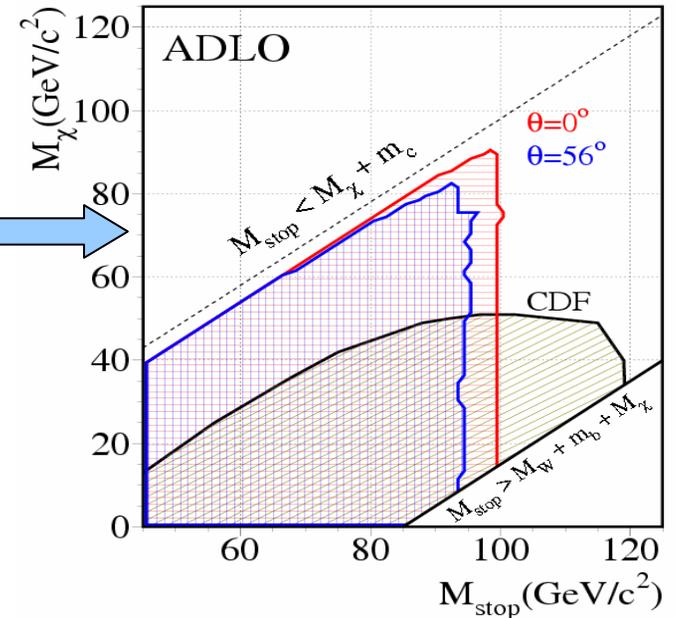
$p\bar{p} \rightarrow \tilde{t}\tilde{t} \rightarrow (c\tilde{\chi}_1^0)(\bar{c}\tilde{\chi}_1^0)$
Signature: ≥ 2 HF jets, MET

Stop in lepton + b-jets PRL 84, 5273 (2000)

$p\bar{p} \rightarrow \tilde{t}\tilde{t} \rightarrow (b\tilde{\chi}_1^+)(\bar{b}\tilde{\chi}_1^-) \rightarrow (b\ell^+\tilde{\chi}_1^0)(\bar{b}\ell^-\tilde{\chi}_1^0)$
Signature: ≥ 1 lepton, b-jet (tagged), MET

Stop in MET+ leptons PRL 84, 5273 (2000)
PRL 88, 171802-1 (2002)

$p\bar{p} \rightarrow \tilde{t}\tilde{t} \rightarrow (b\ell\tilde{\nu})(\bar{b}\ell\tilde{\nu})$
Signature: 2 opposite sign leptons, MET

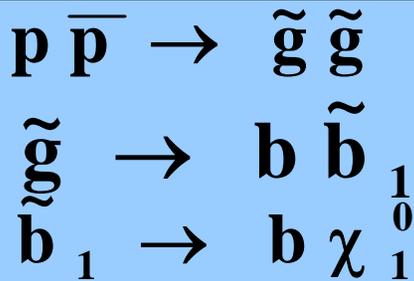


PRL 93, 011801 (2004)

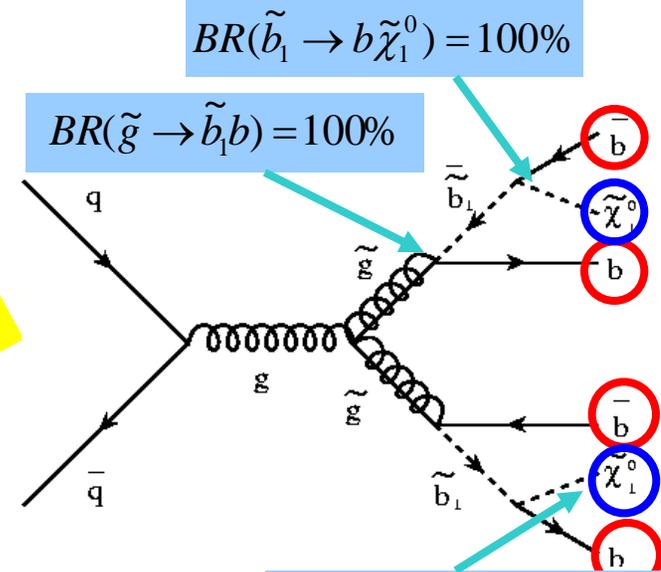
$\tilde{t} \rightarrow b\tilde{\chi}_1^\pm \rightarrow b\ell\nu\tilde{\chi}_1^0$
No b-tagging

Sbottom searches Run II

- CDF Run II search in light sbottom scenario [Z. Phys. C64, 499 (1994)]



Spectacular signature: 4 b-jets + MET



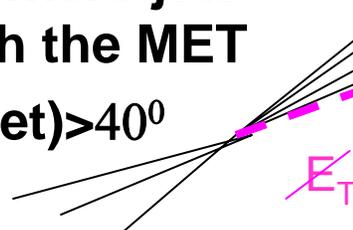
$$BR(\tilde{b}_1 \rightarrow b \tilde{\chi}_1^0) = 100\%$$

$$BR(\tilde{g} \rightarrow \tilde{b}_1 b) = 100\%$$

$$m(\tilde{\chi}_1^0) = 60 \text{ GeV}/c^2$$

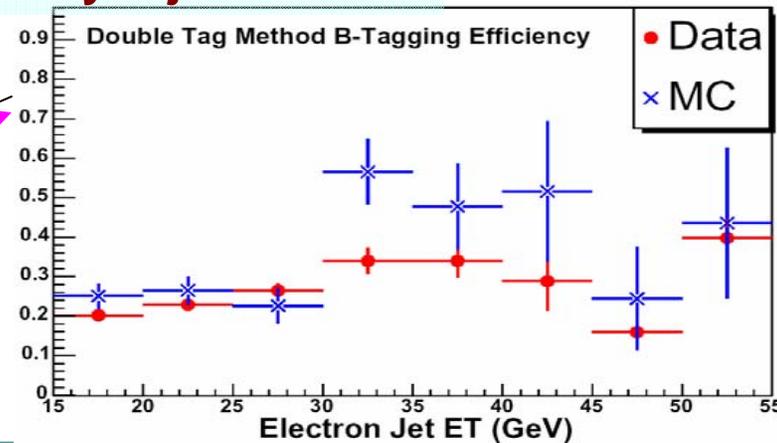
- MET >35GeV+ jet trigger
- 3 or more jets
- Reject events where jets are aligned with the MET

$$\bullet \Delta\phi(\text{Met}, 1\text{-}3.\text{jet}) > 40^\circ$$

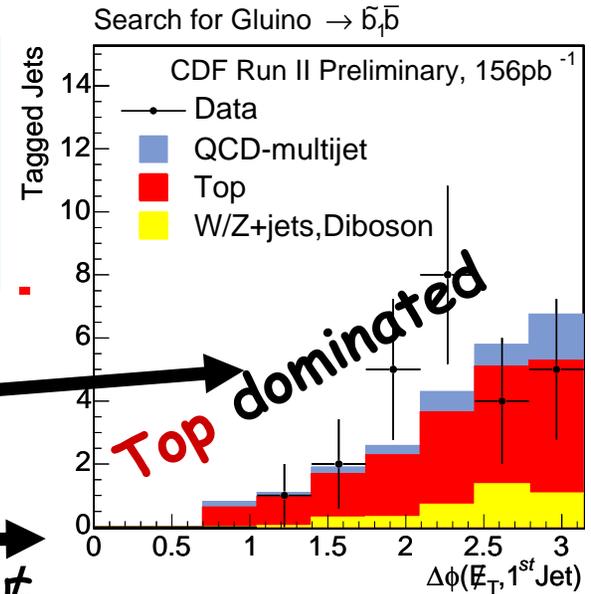
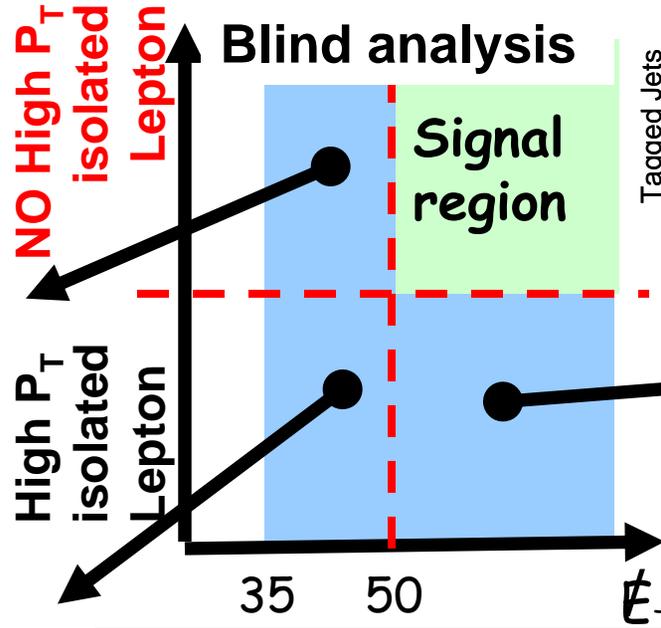
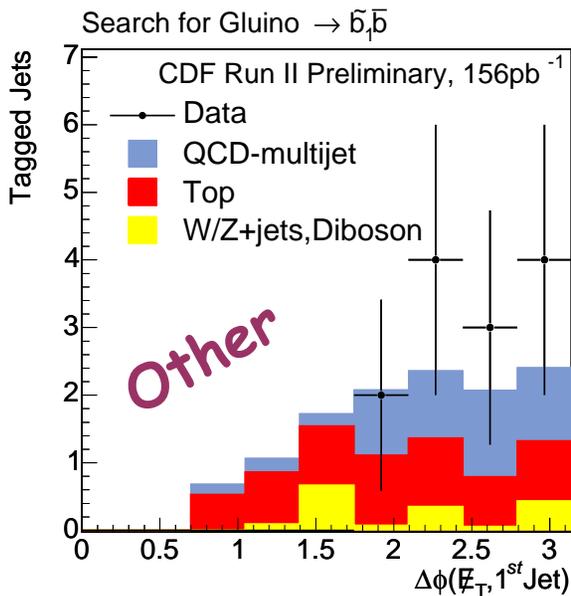
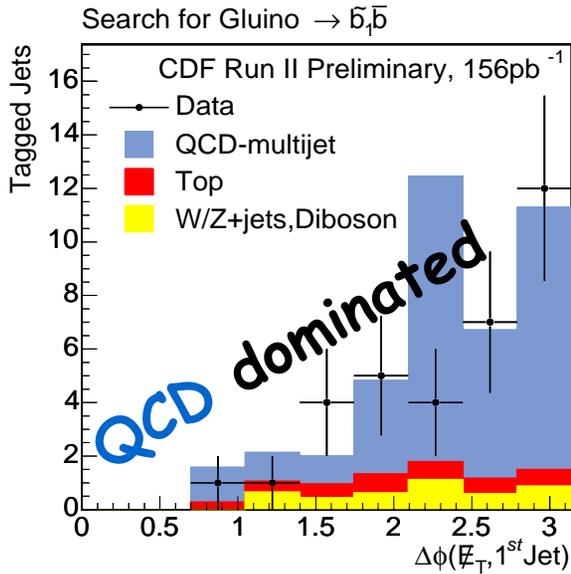


- Separate analysis for single and double b-tag

Fake rate/jet $\sim 0.3\%$
Efficiency/b-jet 25-40 %



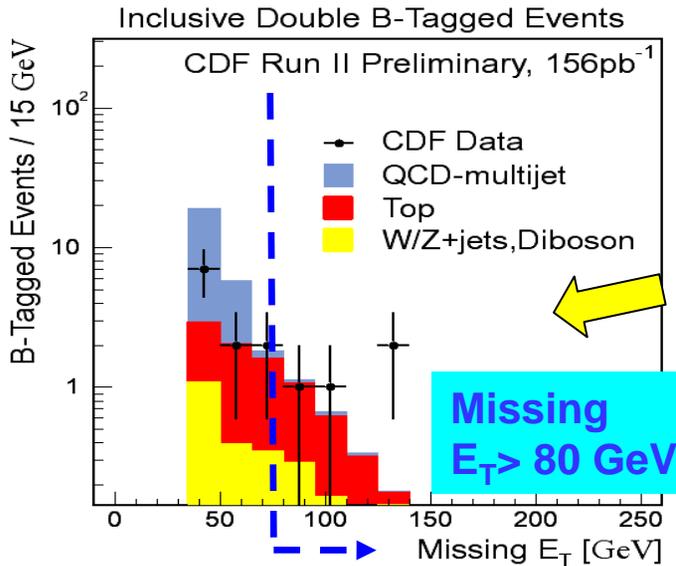
Control regions



	Other	QCD dominated	Top dominated
EW	3.9±0.8	11.0±1.2	9.6±1.2
Top	11.7±0.2	8.2±0.1	35.2±0.3
QCD	19.2±4.1	129.6±17.3	10.9±4.5
BCK	34.8±4.2	148.8±17.3	55.7±4.7
OBS	36	121	63

TEV4LHC

Signal region

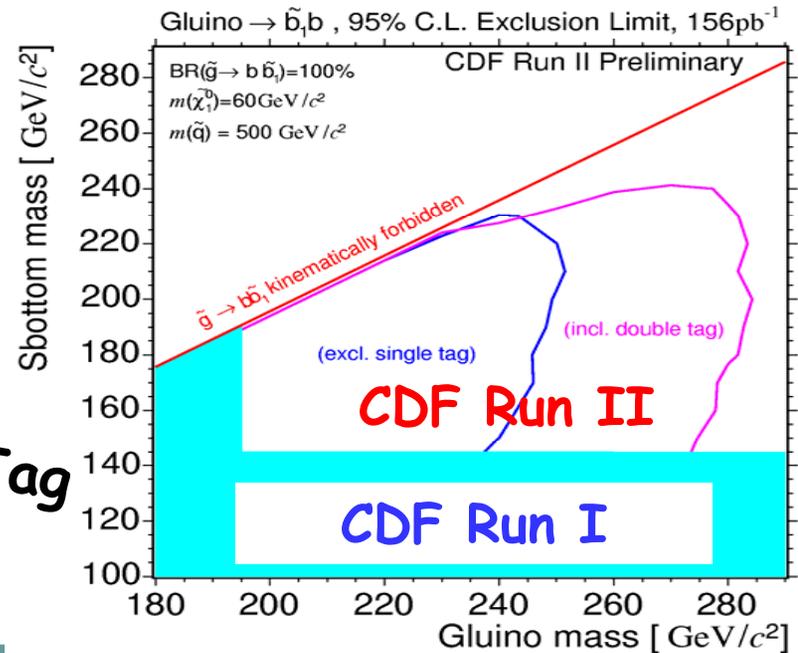
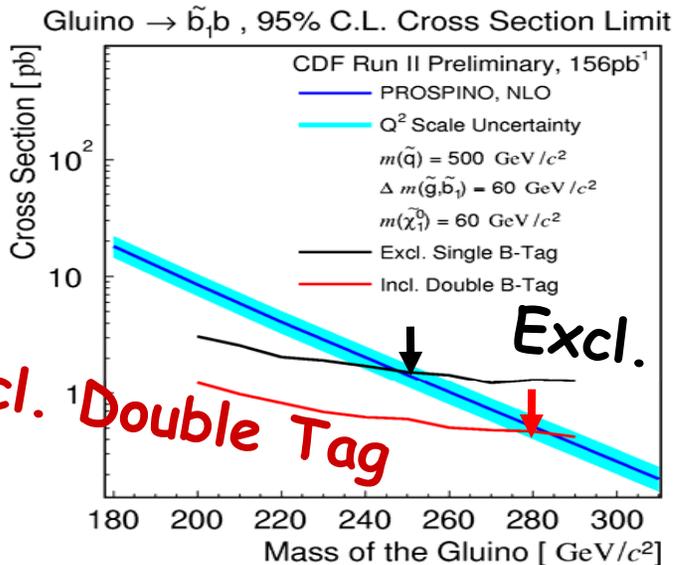


156pb^{-1} of data taken 2002-2003

Expected Excl. single tag 16.4 ± 3.7
Observed 21

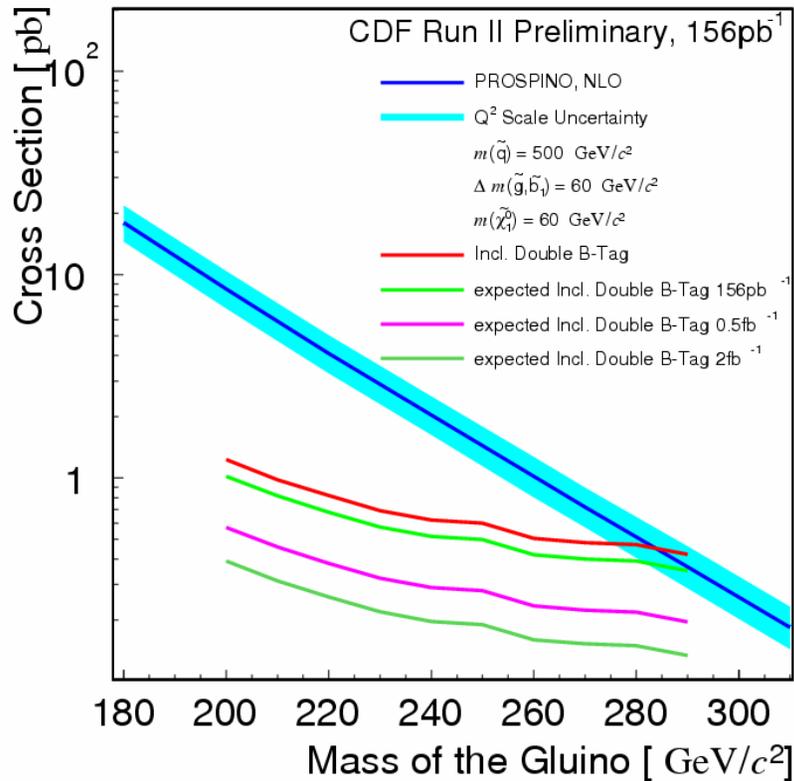
Expected incl. double tag 2.6 ± 0.7
Observed 4

Acceptance: Exclusive Single Tag = 7.7%
Inclusive Double Tag = 9.0%

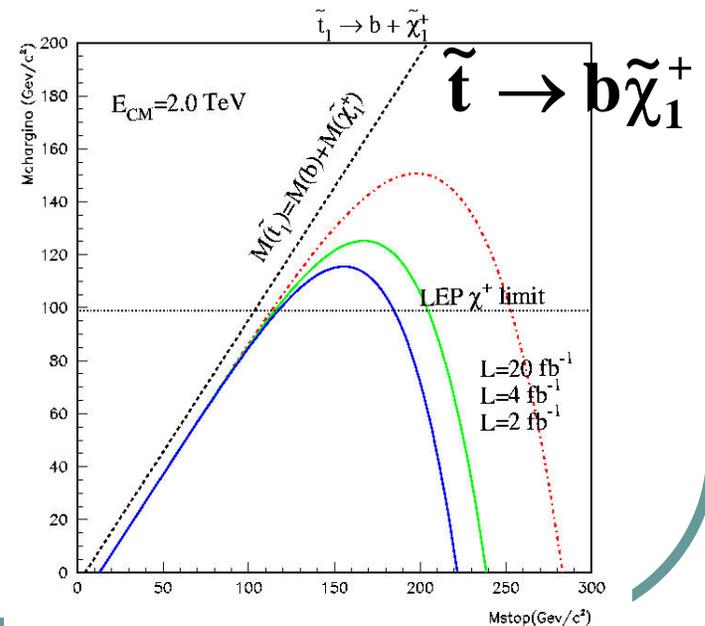
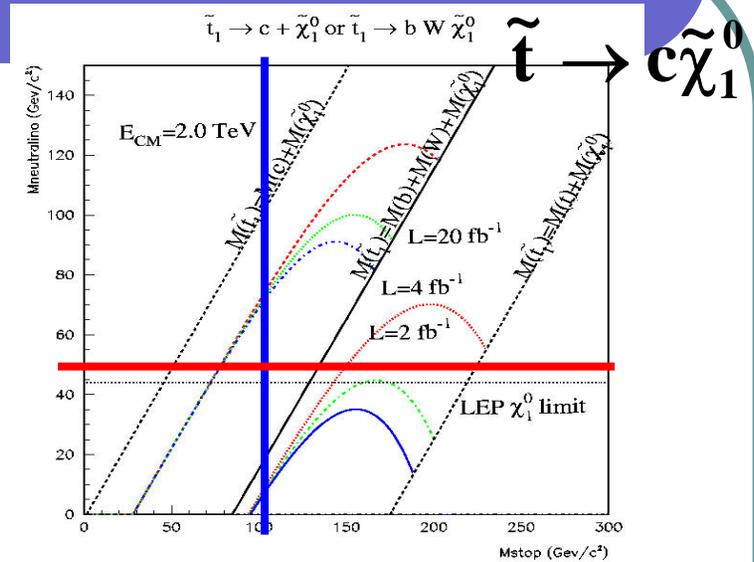


Prospect for run II

Glucino $\rightarrow \tilde{b}_1 b$, 95% C.L. Cross Section Limit



TeV will explore stop and sbottom well above top mass



Beyond Minimal SUSY

- Even the MSSM (100+ parameters) may not be general enough.
- **For example, the LEP-II bound on the Higgs mass implies that the stop mass is greater than ~ 700 GeV.**
 - This would make looking for stops pretty hopeless at run II.
 - It's disturbing because it starts to recreate the hierarchy problem.
- Specific extensions (**NMSSM**, "**Gauge Extended MSSM**", "**Fat Higgs**", etc.) are already appearing to relax this requirement.
 - They often have more states (i.e. **Z's**, **exotic Higgses**).
 - There are also often special properties for the new states related to the 3rd family, like enhanced decays to **τ** , **b** , and **t** .
- What about R-parity violation? It still allows the hierarchy solution, though it does give up SUSY dark matter. It's interesting!
- **MORAL 1: Keep looking for those stops!**
- **MORAL 2: Realistic SUSY theories are probably somewhat richer than we naively expect. The first hints may not look like SUSY.**

Tim Tait's talk at CDF
Collaboration meeting

Searches at the LHC

- Scenario has been investigated by CMS. For point B, gluino heaviest sparticle, neutralino is LSP

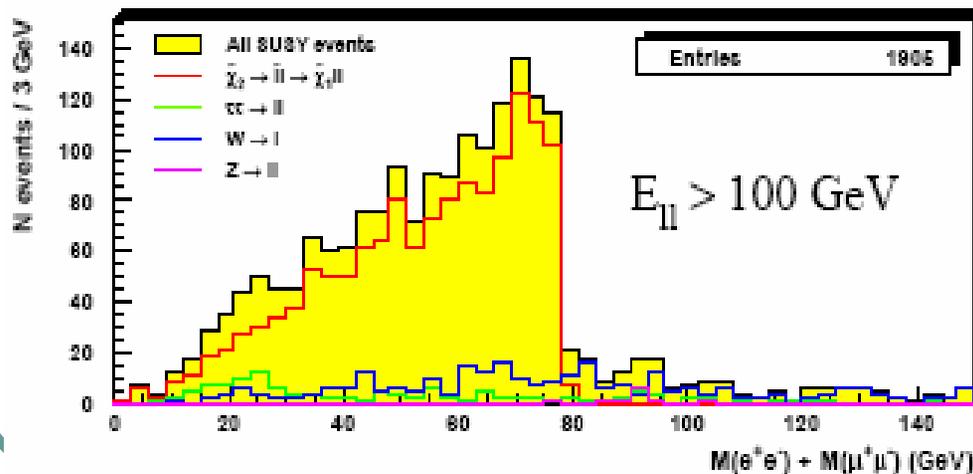
$$m(\tilde{g}) = 595 \text{ GeV} \quad m(\chi_1^0) = 95 \text{ GeV}$$

$$\text{BR}(\tilde{g} \rightarrow \tilde{b}_1 b) + \text{BR}(\tilde{g} \rightarrow \tilde{b}_2 b) \approx 27.5\%$$

- Fast detector simulation
- Selection starts with dilepton end-point invariant mass distribution

\tilde{b}_2	
Decay channel	BR(%)
$\tilde{\chi}_1^\pm t$	34.06
$\tilde{\chi}_1^0 b$	17.32
$\tilde{\chi}_2^0 b$	25.04
$\tilde{\chi}_3^0 b$	0.17
$\tilde{\chi}_4^0 b$	1.53
$\tilde{t}_1 W$	2.19

\tilde{b}_1	
Decay channel	BR(%)
$\tilde{\chi}_1^\pm t$	46.37
$\tilde{\chi}_2^0 b$	37.05
$\tilde{\chi}_1^0 b$	5.35
$\tilde{\chi}_3^0 b$	0.19
$\tilde{\chi}_4^0 b$	1.80
$\tilde{t}_1 W$	9.23



$$M_{l^+l^-}^{\max} = \frac{\sqrt{(M_{\tilde{\chi}_2^0}^2 - M_{\tilde{l}}^2)(M_{\tilde{l}}^2 - M_{\tilde{\chi}_1^0}^2)}}{M_{\tilde{l}}}$$

Searches at the LHC

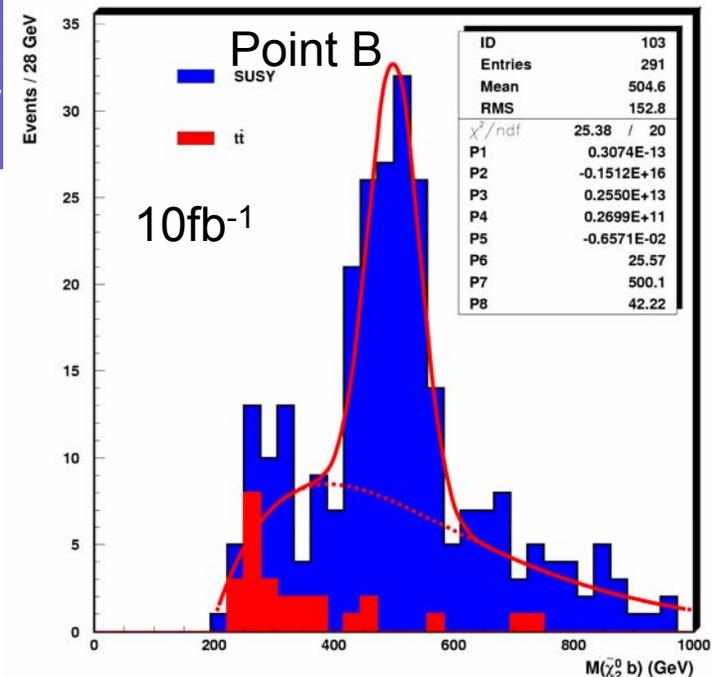
- B-tagging used to reconstruct sbottom and gluino decay
- Also used to anti select b-jets for squark reconstruction
- Select b-jet with $E_{b\text{-jet}} > 250$ GeV:

$$M(\tilde{\chi}_2^0 b) = (500 \pm 7) \text{ GeV}$$

- **Reconstruction of strongly interacting sparticle possible**

Point B:

- first few weeks of LHC running period:
 - reconstruction of squark (resolution $\sim 12\%$)
- first year:
 - reconstruction of sbottom and gluino (resolutions $\sim 6\div 8\%$)
 - reconstruction of gluino in the squark chain (independent channel)



M. Chiorboli, Physics at the LHC- Prague 2003

Point I ($\tan \beta=35$):

- no reconstruction possible in the leptonic final state even with high accumulated statistics
- tau final state is under investigation

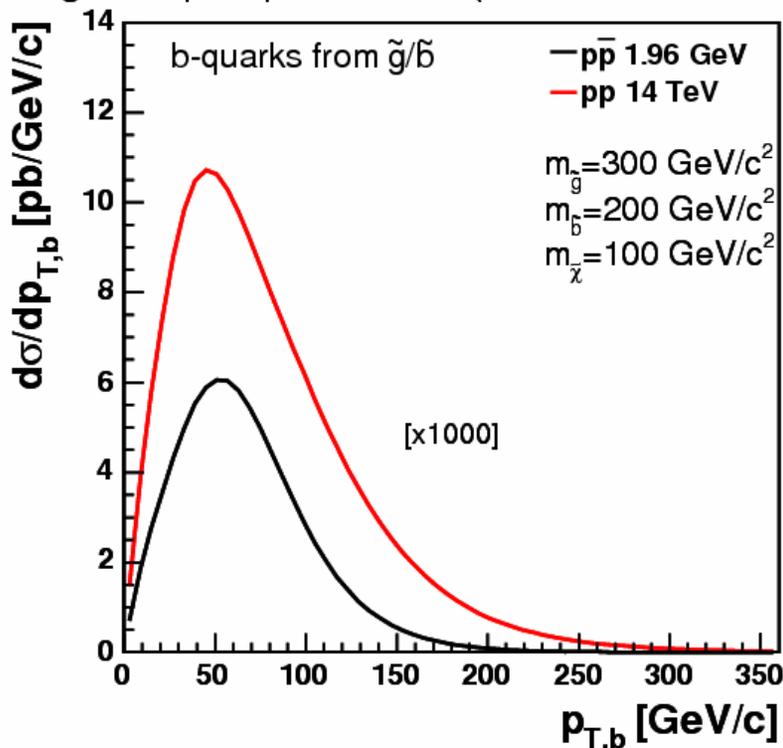
Sbottom from gluino decays at LHC

- The search for sbottom from gluino decay in MET + 2b-tag jets is also interesting for the LHC
- Analysis is just starting for CMS. Huge cross section!!!!

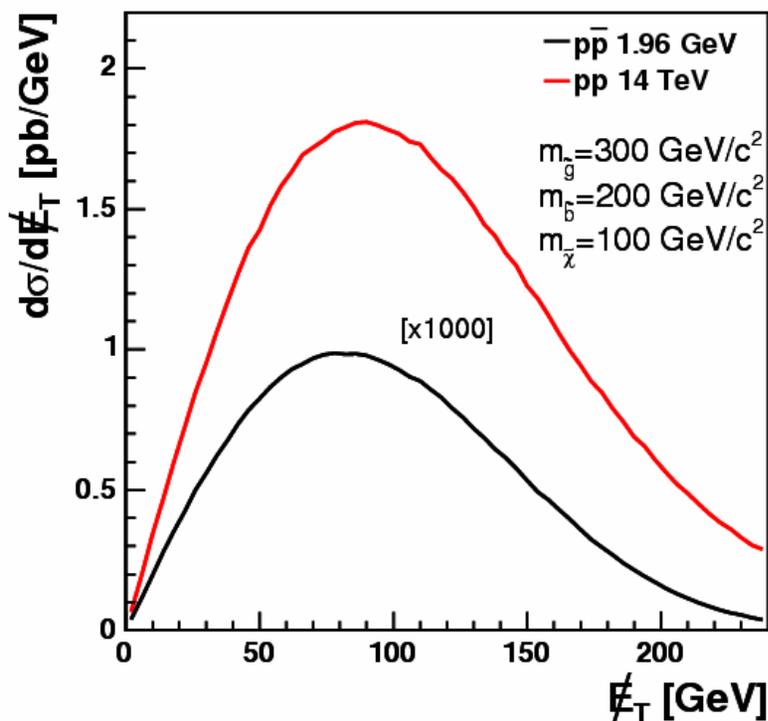
$$\text{BR}(\tilde{b}_1 \rightarrow b\tilde{\chi}_1^0) \approx 5\%$$

$$\text{BR}(\tilde{b}_2 \rightarrow b\tilde{\chi}_1^0) \approx 17\%$$

gluino pair production (ISAJET, CTEQ6L1)



gluino pair production (ISAJET, CTEQ6L1)

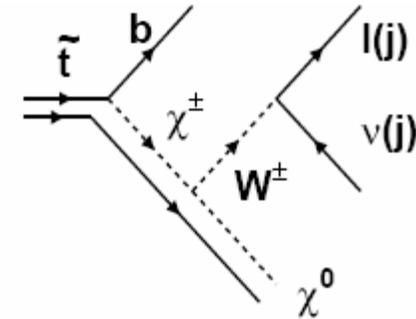
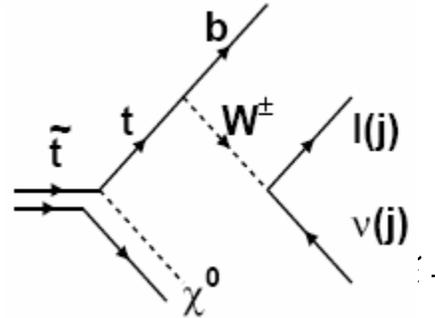


Searches for the stop at the LHC

Decay chain for

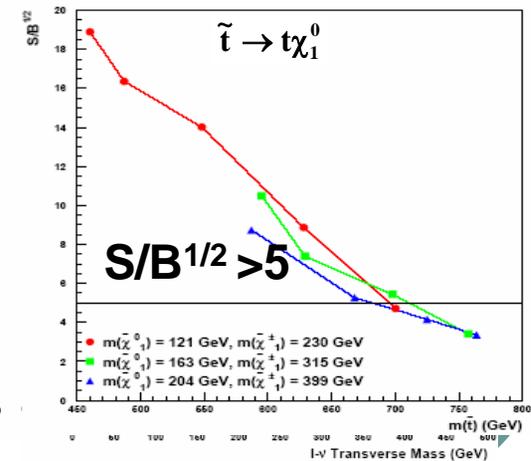
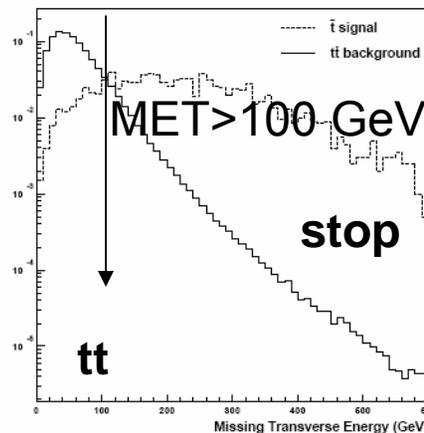
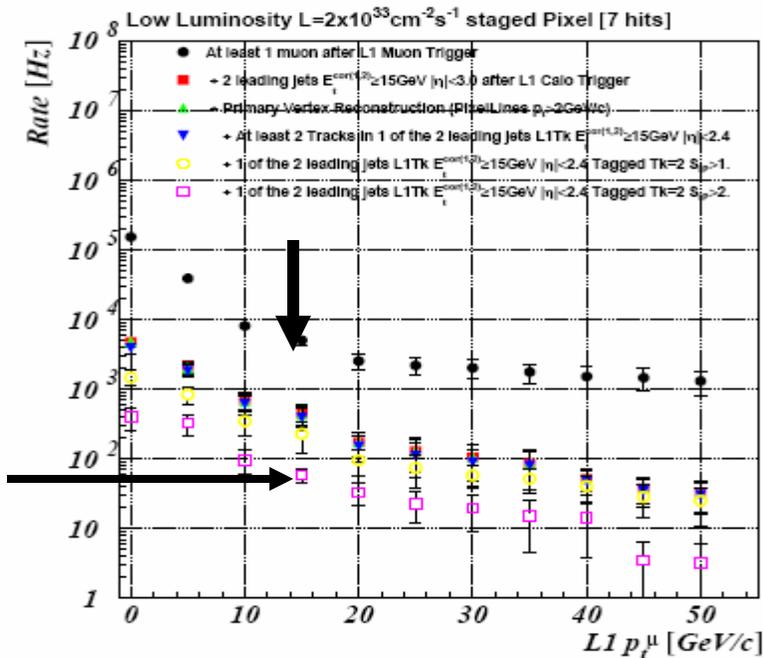
$$\tilde{t} \rightarrow t\chi_1^0 \quad \text{and} \quad \tilde{t}_1 \rightarrow b\chi_1^\pm$$

Trigger rate can be reduced to 30Hz by requiring at least one of the b-jets to be tagged



Large tt background can be reduced by MET > 100 GeV

Stop can be discovered up to about 600- 700 GeV



Gabriele Segneri- CMS

Impact Parameter Methods

■ Track Counting Methods

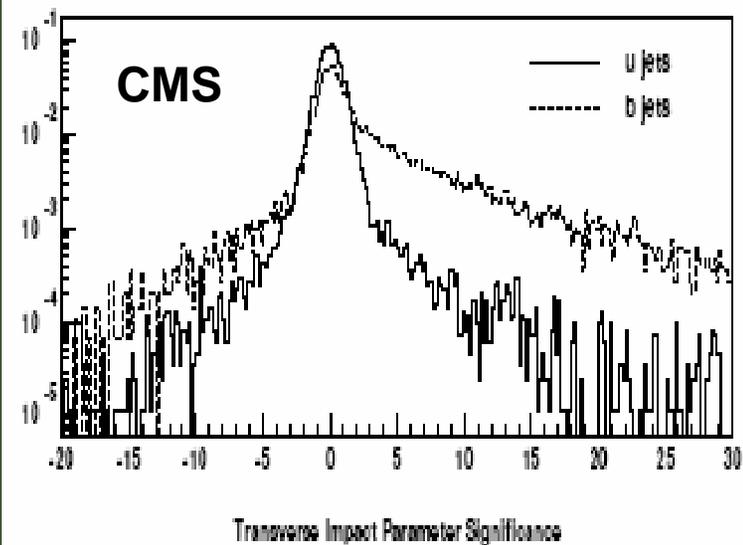
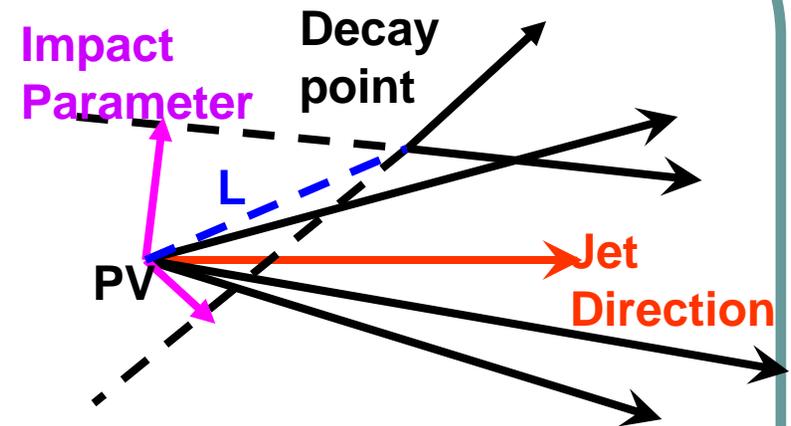
- Minimum N(good tracks) with IP Significance S_{IP}

$$S_{IP} = \frac{IP}{\sigma_{IP}} > 2 \text{ or } 3$$

- Quick optimization and feedback

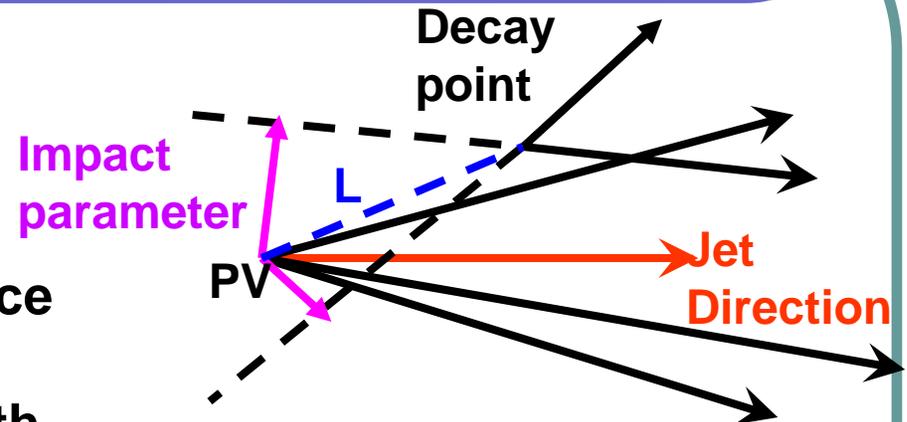
■ Probability method

- Estimate probability that tracks come from primary (2D or 3D)
- Computed using $S_{IP} < 0$. The IP is negative if tracks appear to originate from behind PV
- Could be optimized for b, c, u
- Continuous Output {0-1} but tagging efficiency, fakes can be determined for fixed values (1%, 5%)

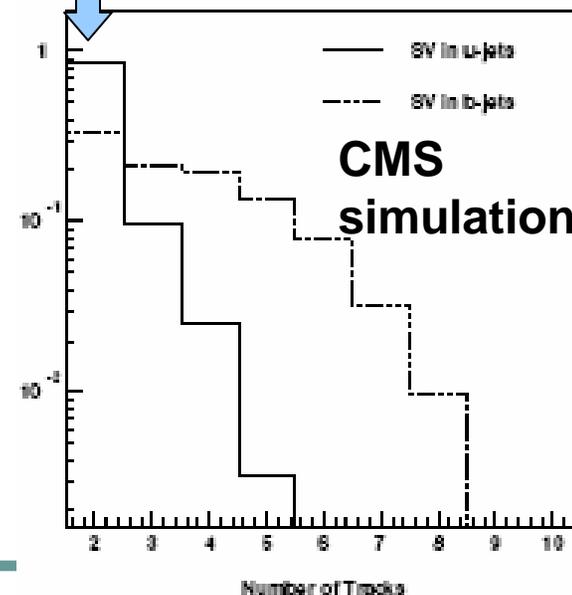
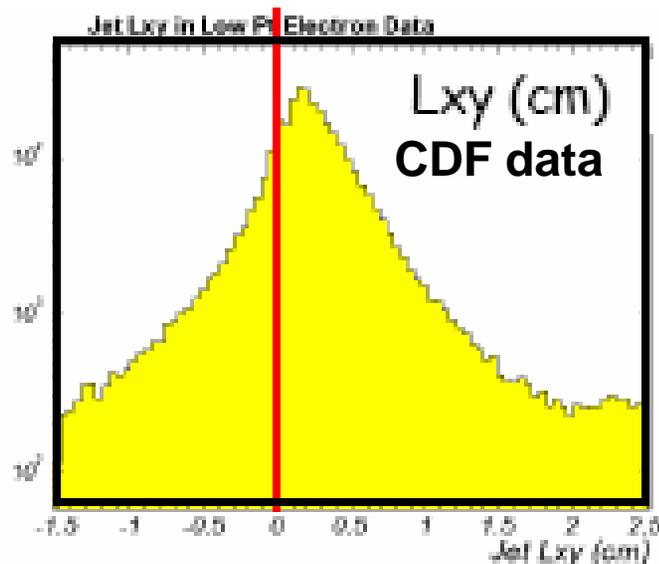


Secondary Vertex Methods

- Use tracks with large S_{IP} to find secondary vertex
- Find >2 or 3 tracks secondary vertices
- Calculate 2D or 3D decay distance L and classify SV based on L
- Fake vertices due interaction with material and long lived particles



2 track vertices dominate by light quark jets



B-tagging at CDF and D0

CDF

- **Jetprob** (Probability algorithm)
- **SecVTX** (secondary vertex algorithm)
 - Loose and tight tagger
- **Track counting**
- **Jet Probability method**
- **Secondary Vertex reconstruction Method**

CMS

- **Counting Signed Impact Parameter (CSIP)**

$$S_{IP}^{RAW} = \frac{IP}{\sigma_{IP}^{RAW}}$$

- Tag if >2 tracks has $S > 3$
- Tag if >3 tracks have $S > 2$
- **Jet Lifetime Probability (JLIP)**
 - Probability algorithm
- **Secondary Vertex reconstruction (SVT)**
- Loose, Medium and Tight for mistag rate of 1%, 0.5%, 0.25%
- **Likelihood Method Approach**

D0

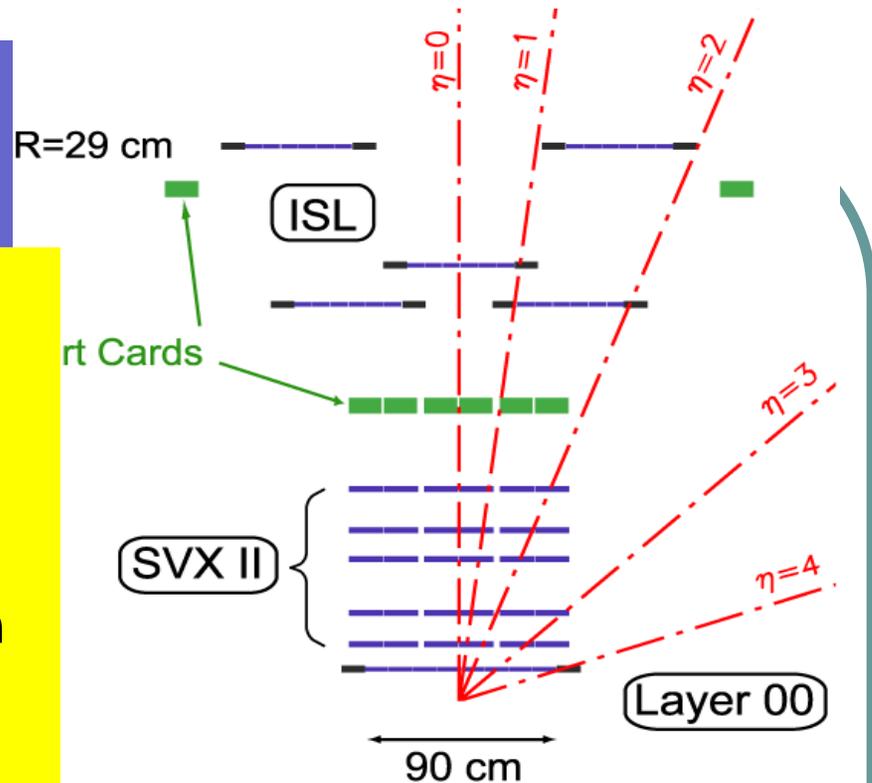
Atlas

$$r_i = \frac{f_b(S_{IP})}{f_u(S_{IP})}$$

B-tagging at CDF

- Tagging algorithms depends on:
 - Hardware performance
 - Tracking
 - Vertexing
 - Alignment
- Retuning must be done after initial detector and reconstruction performance is understood.
 - Hardware changes
 - Improved tracking, primary and secondary vertexing, alignment
 - Hardware performance needed to tune realistic MC

- SeCVTX algorithm developed in run I. Ported to run II in 2000
 - Pass 1: 3 track vertices
 - Pass 2: 2 track vertices



Run II started in APRIL 2001

- ISL cooling
 - April 2003 ISL included in tracking
 - May 2003 ISL and z-side included in SECVTX
- Detector noise on L00
 - L00 to be included in next release (later in 2004)

B-tagging timeline at CDF

■ Dec. 2002 SecVtx optimization for 2003 Winter conference:

- SVX alignment and tracking code not optimal
 - Tight track selection
 - Loose vertex quality cuts
- Run averaged beamline

■ May 2003:

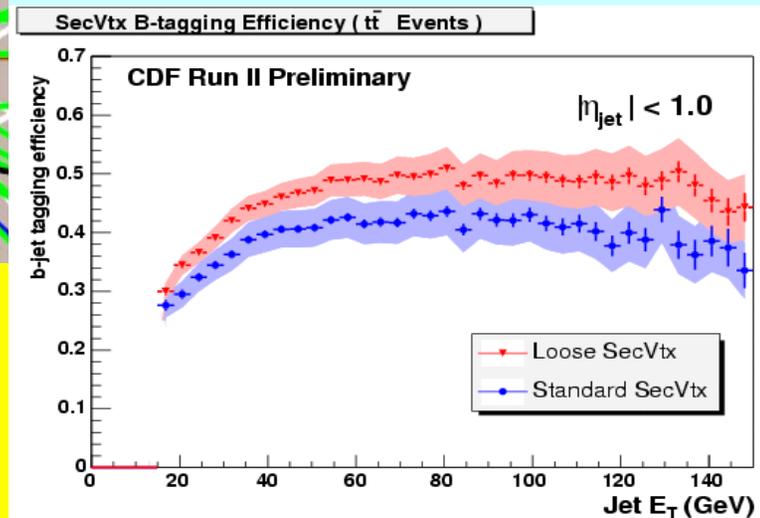
- Event by event primary vertex with beam line constraint
- Beam width function of z-position
- Account beam variation in a run



■ Summer 2003: first top and exotic analysis blessed using SecVtx

■ Improved Tracking and alignment in Fall 2003

- Loosen track selection
- Development Loose SECVTX for double tag analysis (allow larger fake rate)
- First blessing of double tag analysis Summer 2004

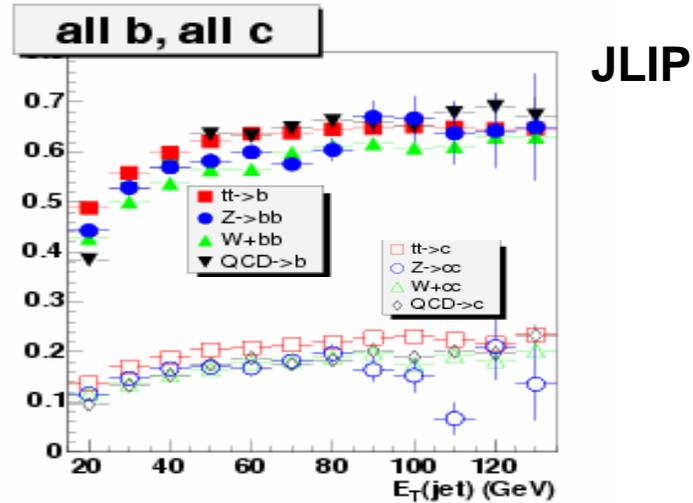
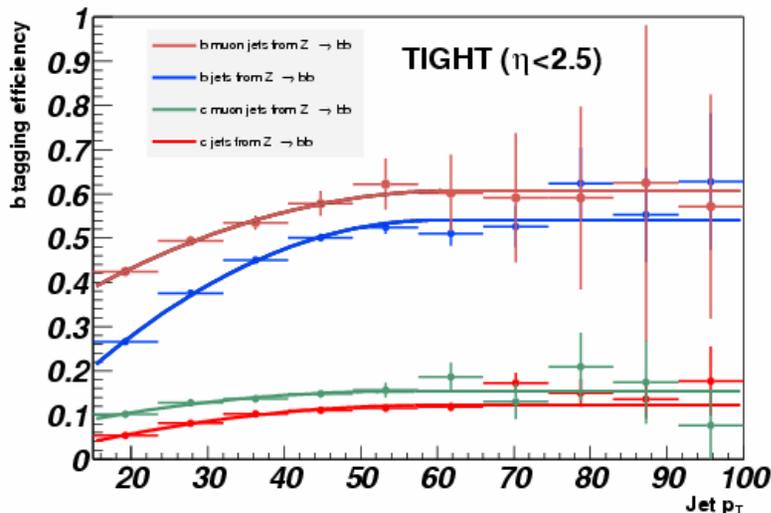
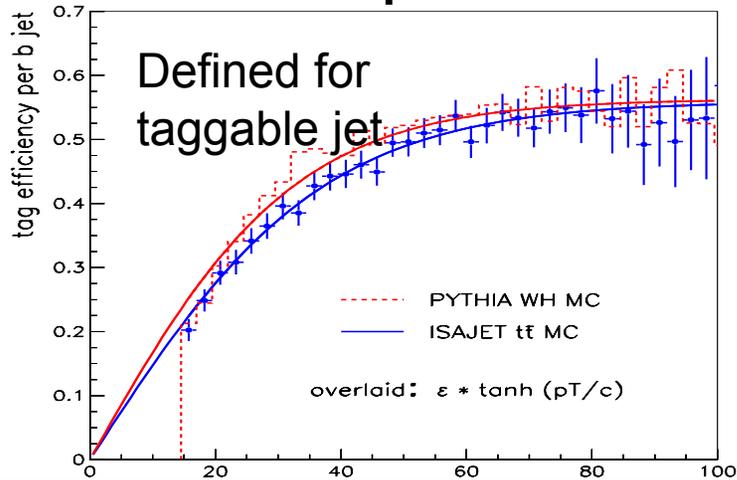


Organizational experience at CDF

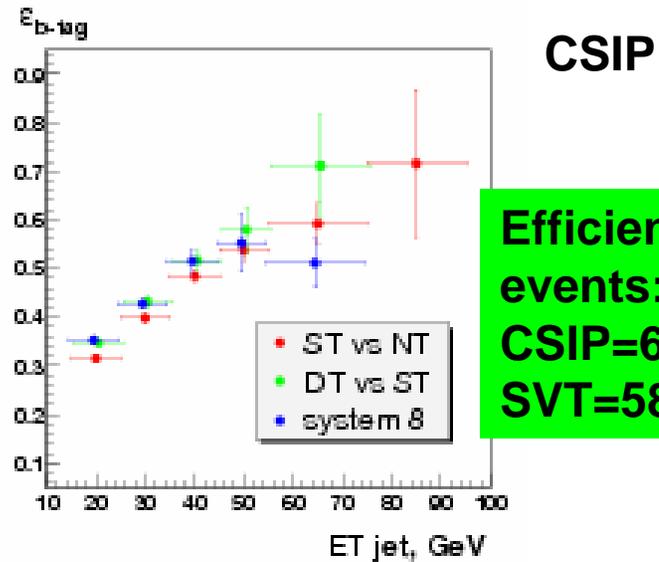
- **Current b-tagging improvements:**
 - tracking up to $\eta \sim 1.2$ (IO tracks)
 - More realistic MC simulation with new charge deposition model
 - Better alignment in Z.
- **Future: Would like to include forward tracking**
- **B-tagging is a high level tool built on lower level objects subjected to change \Rightarrow Maintenance and upgrades are necessary**
- **Once a b-tagging algorithm is working, it is difficult to motivate people to keep improving it**
- **Algorithm development at CDF was mostly by top group for top analysis**
 - Top mass analysis in all hadronic channel needs very pure b-tagging
- **Optimization for searches**
 - Improved c-tagging ($t \rightarrow c$ LSP)
 - Or charm rejection
 - Sbottom search largest background is top (which yields b-jets)
 - might need algorithm with high b-tagging efficiency and “sufficient” rejection of light quark jet

Predictions and performance

D0 Predictions HEP 2001
Budapest - Chiarelli



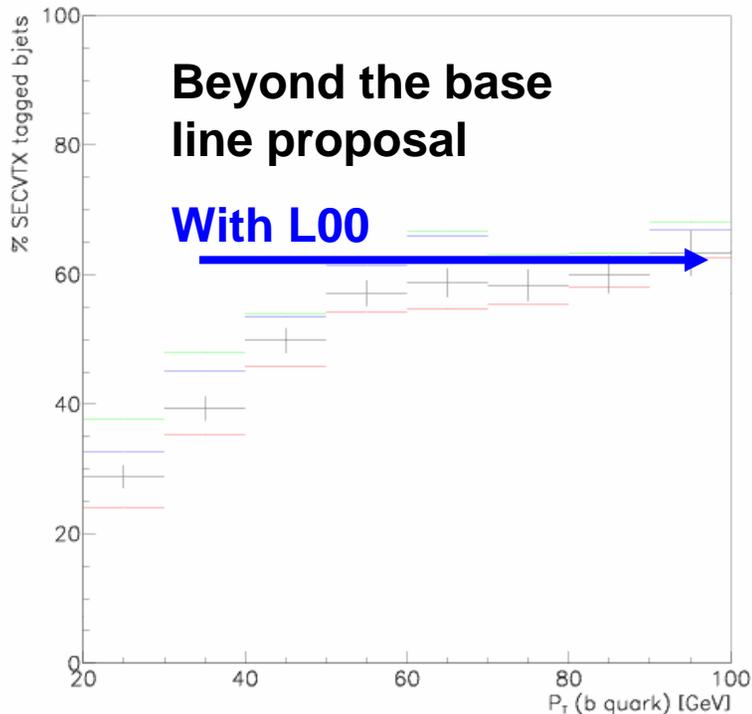
SVT



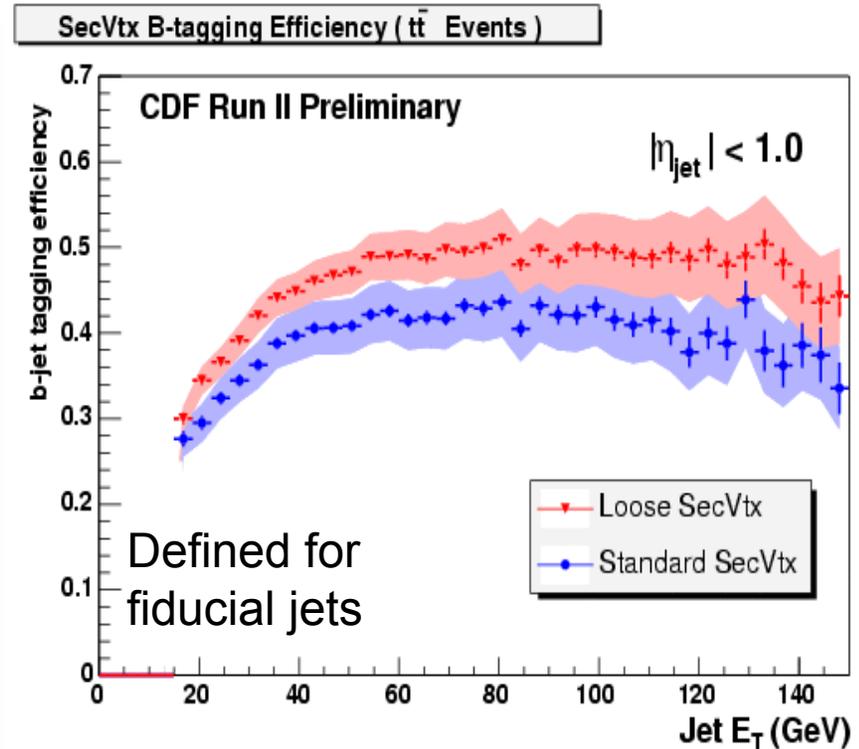
Efficiency in top events:
CSIP=61%
SVT=58%

Predictions and performance

■ CDF



Expected efficiency in top events 65%



Efficiency in top events:
Winter 03 : 44.3%;
Summer 03: 52.3%;
Summer 04: 58.6% (loose tagger) and still improving

LHC challenges

■ Luminosity

- low-luminosity: $10^{33}\text{cm}^{-2}\text{s}^{-1}$ (first 3 years)
- high-luminosity: $10^{34}\text{cm}^{-2}\text{s}^{-1}$
 - ~20 minimum bias events per bunch crossing
 - ~1000 charged tracks per event

- Radius: 2cm 10cm 25cm
- N_{Tracks} 10.0 1.0 0.10 / $\text{cm}^2\text{25ns}$
- bigger probabilities of jet overlaps.

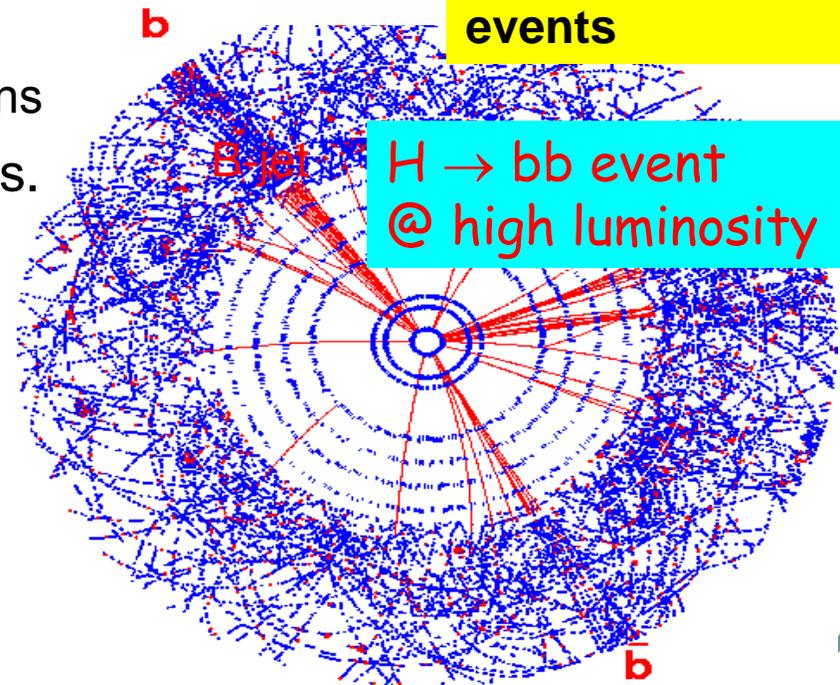
■ LHC detectors have been built for this environment

- Fine granularity
- Fast response time
- Excellent IP resolution
- 90% reconstruction efficiency for high pt tracks inside jets

How a "real" 2- jet event looks like:

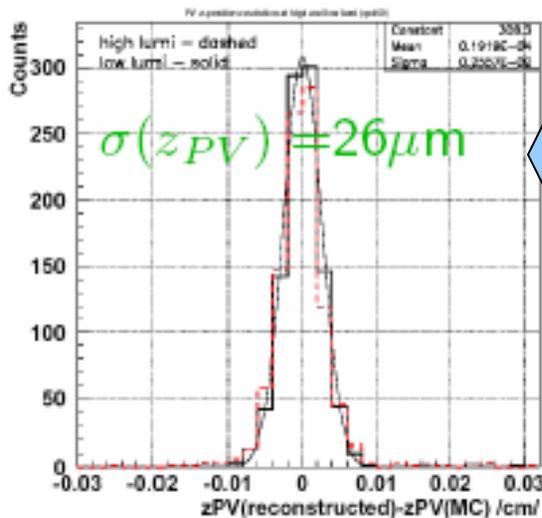
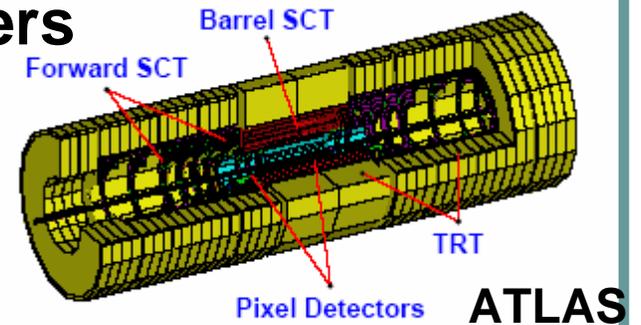
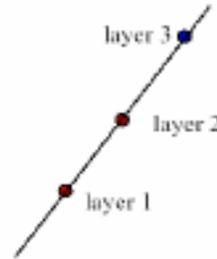
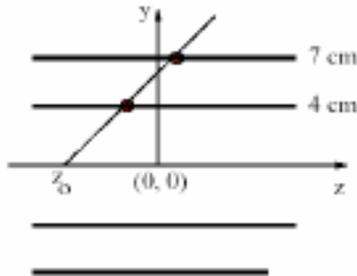
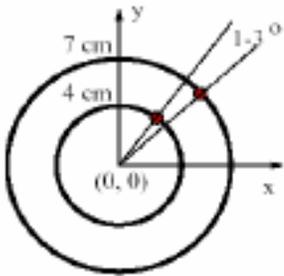


17 superimposed events



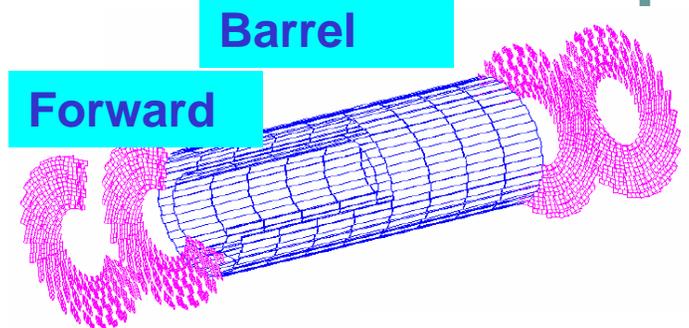
LHC challenges

- Fine pixel sensors close to beam provide excellent and fast primary vertex reconstruction
- Track seeding starts from pixel layers

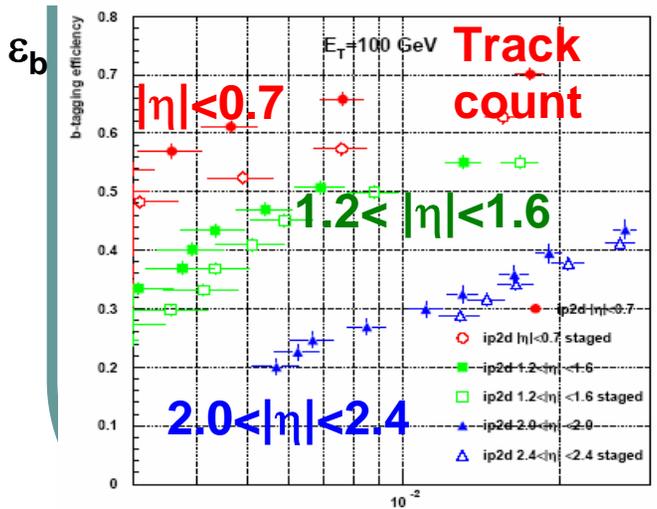
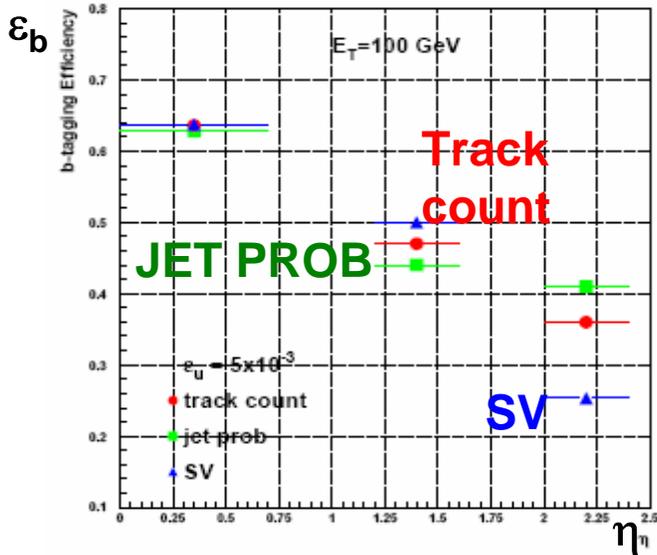


- Select PV with at least 3 tracks that cross the beam axis

- Full tracker resolution 15 μm



Expected LHC performance



CMS **ATLAS**

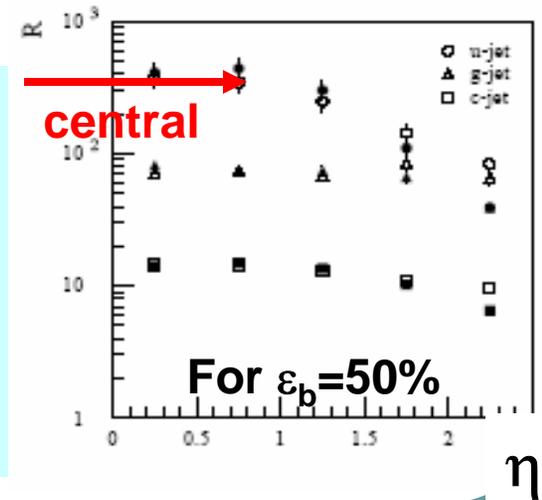
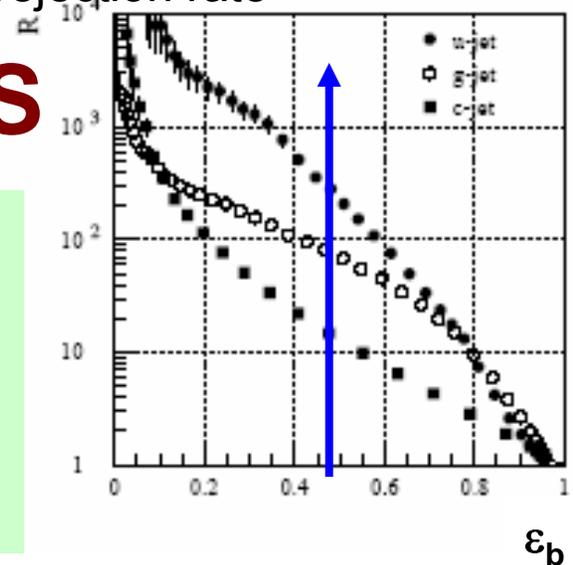
Both Atlas and CMS achieve:

- Mistagging rate for u-jet $\sim 5 \times 10^{-3}$
- Efficiency $\sim 50\%$

Staging the pixels: 2 barrels layers + 1 disk

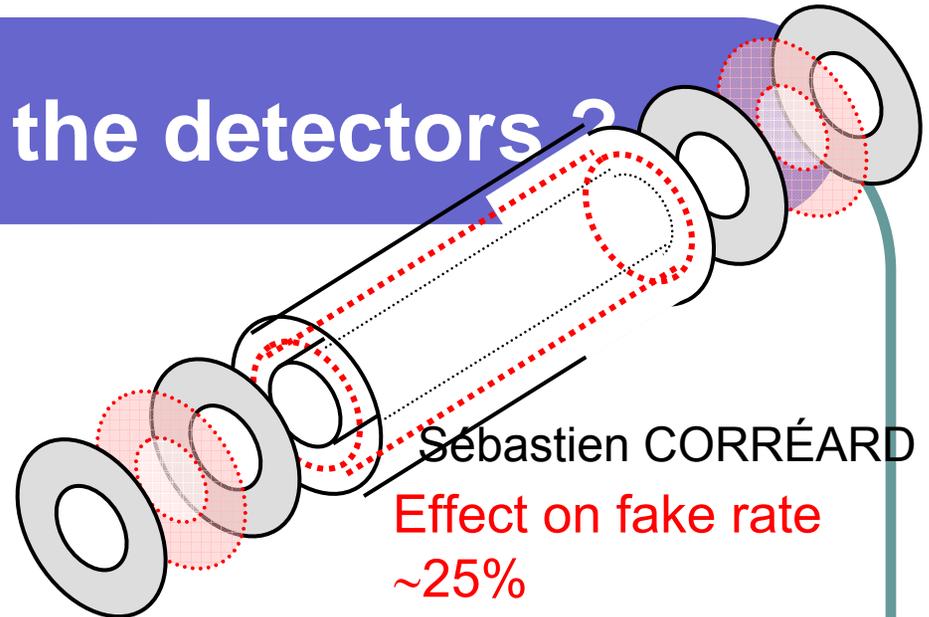
Loss of 10% average b-tagging efficiency

R=rejection rate



How well will we know the detectors ?

- Effect of staging in ATLAS
 - no pixel barrel #1
 - no pixel disk #2
 - no CTRT wheels
 - a *b*-layer pitch = 400 μm

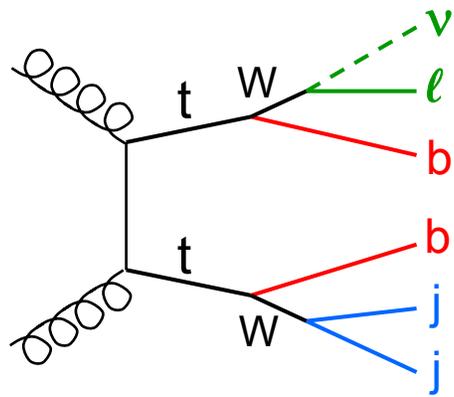


time	Inefficiency on pixels & SCT		Inefficiency on <i>b</i> -layer pixels		Precision on alignment	
	module	chip	module	chip	pixels	SCT
Period 1 (months 1-2)	4%	3%	2%	1.5%	100 μm	300 μm
Period 2 (months 3-4)	2%	2.5%	1%	1.25%	20 μm	60 μm
Period 3 (months 5-6)	1%	2%	0.5%	1%	10 μm	30 μm
Period 4 (months 7-8)	1%	2%	0.5%	1%	5 μm	15 μm
Period 5 (after...?)	1%	2%	0.5%	1%	-	-

Calibration at the LHC

- Large data samples for calibrations:

- Top



Sébastien CORRÉARD

- $Z \rightarrow \mu\mu \sim 1$ event/s

- Calibrate absolute momentum scale (B field)

- Calibrate b-tagging efficiency using top ℓ +jet decay?

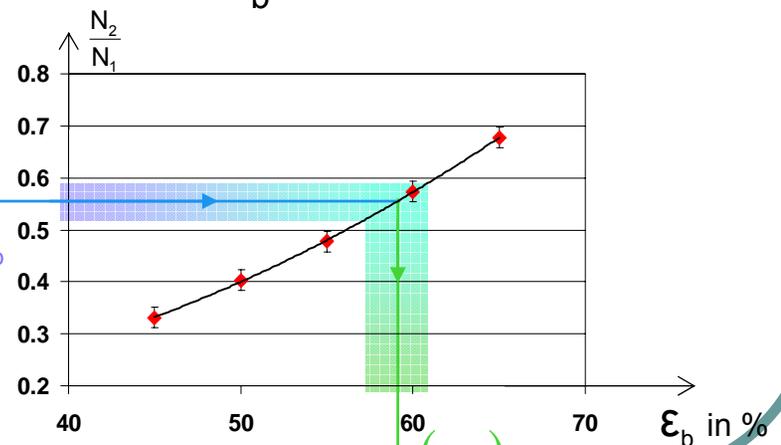
- 2 b -jets
- high cross-section: 2.5 M events 1 year at low luminosity

- Predict the value of N_2/N_1 , for different values of ϵ_b
- Measure N_2/N_1 in data
- Derive ϵ_b

$$\epsilon_b \propto \frac{N_2}{N_1}$$



$$\left(\frac{N_2}{N_1}\right)_{\text{exp}}$$

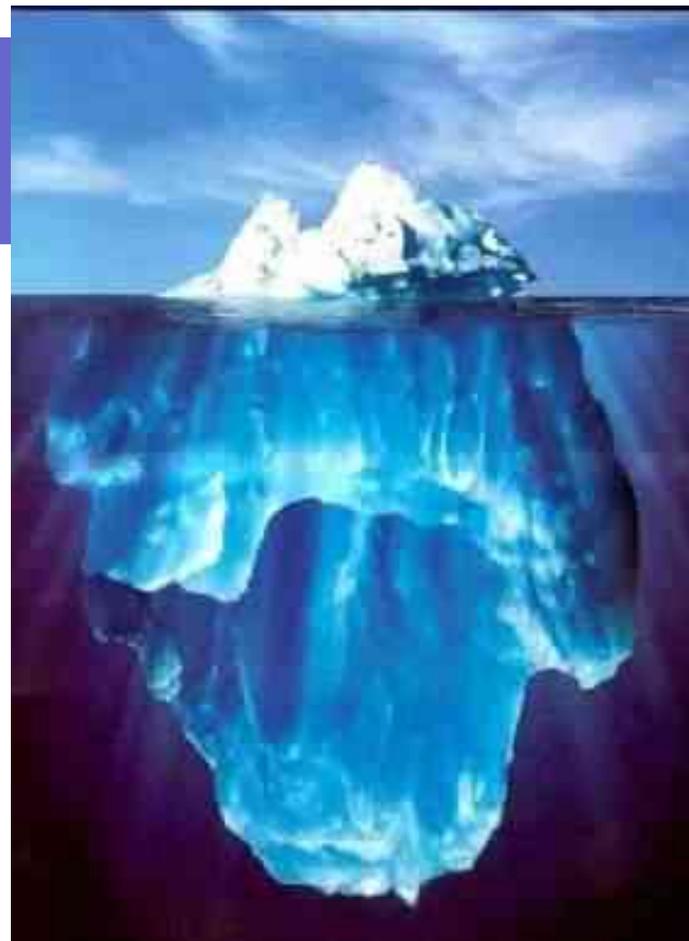


LHC challenges

- **What is the effect of the detector response, physics object definitions, calibration, and alignment procedures on the expected performance?**
- **CMS plans a TDR with feasibility studies of Higgs, SUSY, etc**
 - **Small number of full analyses in the most realistic possible scenarios**
 - **Analyses will include backgrounds, misalignment, and miscalibration**
- **Alignment**
 - **State of the art techniques ala SLD with 96 pixel elements:**
 - **Determine 578 corrections x from 2108 coefficients from residual fits C**
 - **$Ax=C$ where A is the design matrix**
 - **Find A^+ such that $x=A^+c$ minimizes $|Ax-C|$**
 - **CMS has 20,000 independent silicon sensors.....**

Outlook

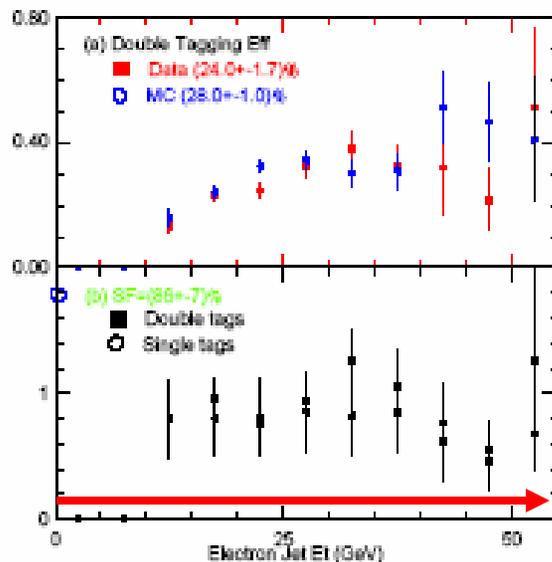
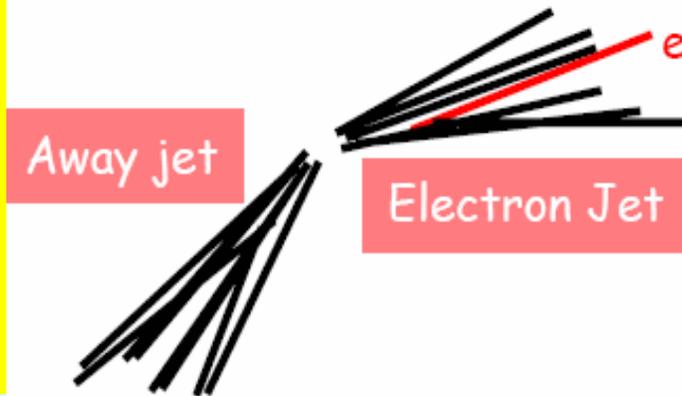
- We might find SUSY at the Tevatron \Rightarrow would give insight for LHC.
- Implementation of TeV analysis in LHC framework, validation of LHC techniques on TeV data might facilitate preparation for discovery
- Questions/ requests:
 - Was the simulation different from the real life at the TeV startup ?
 - Do the b-tagging efficiency and fake rate change with luminosity and # of primary vertices?
 - Micro-DST of the good TeV events with and without b-jets to tests LHC software
 - How much the mis-alignment detector affect the results?



- How do you calibrate the results.. using MC generators, which ones?
- As usual: Hard work is necessary

B-tagging efficiency

- **B-tag efficiency depends on**
 - b- momentum and lifetime
 - Decay Charge multiplicity
 - Tracking performance (efficiency, resolution, material)



Not a great match to top and exotic physics

■ Measurement

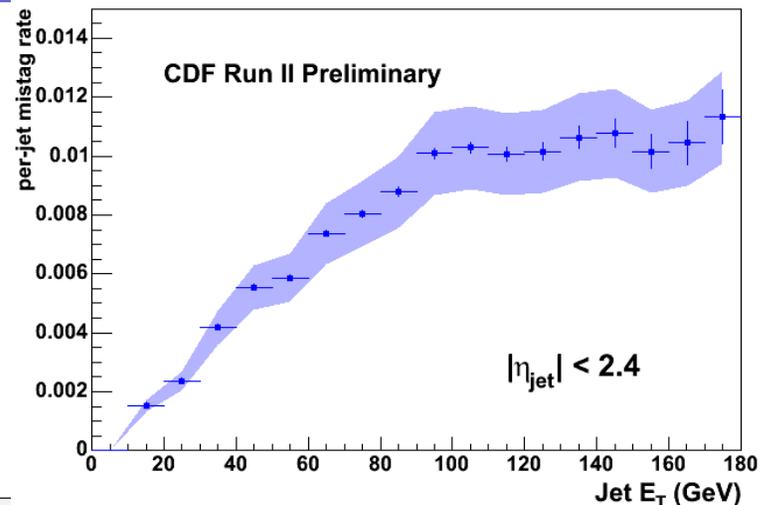
- Use high statistics, high purity b-jets sample such as electron data
- Both Jets with $E_T > 15$ GeV
- Tag away jet to increase purity
- Measure efficiency in data and MC and correct with a scale factor
- Aim: scale factor $S=1$

■ **Scale factor**
 $S = 0.82 \pm 0.06$

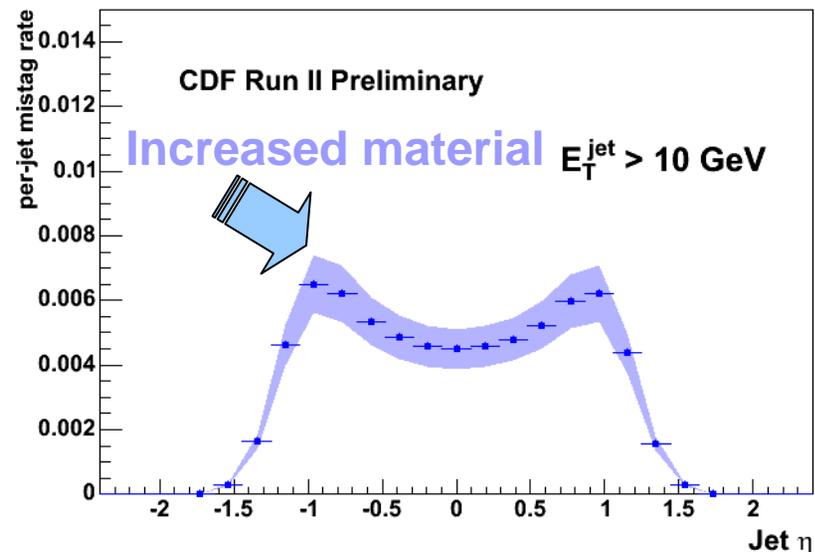
Mistagging

- Mis-tagging probability depends on :
 - Jet E_T
 - Jet η and ϕ
 - # of tracks in jet
 - Energy in the event
- Determined from the negative rate ($-L_{xy}$) in jet data. Corrected for material effects and long lived particles
- Build fake matrix
 - Use jet sample collected with different thresholds (20, 50, 70, 100 GeV)
 - Test on 4 jets sample and $\Sigma E_T > 100$ GeV
- Correlation with MET not explored. But is important for searches

SecVtx Mistag Rate

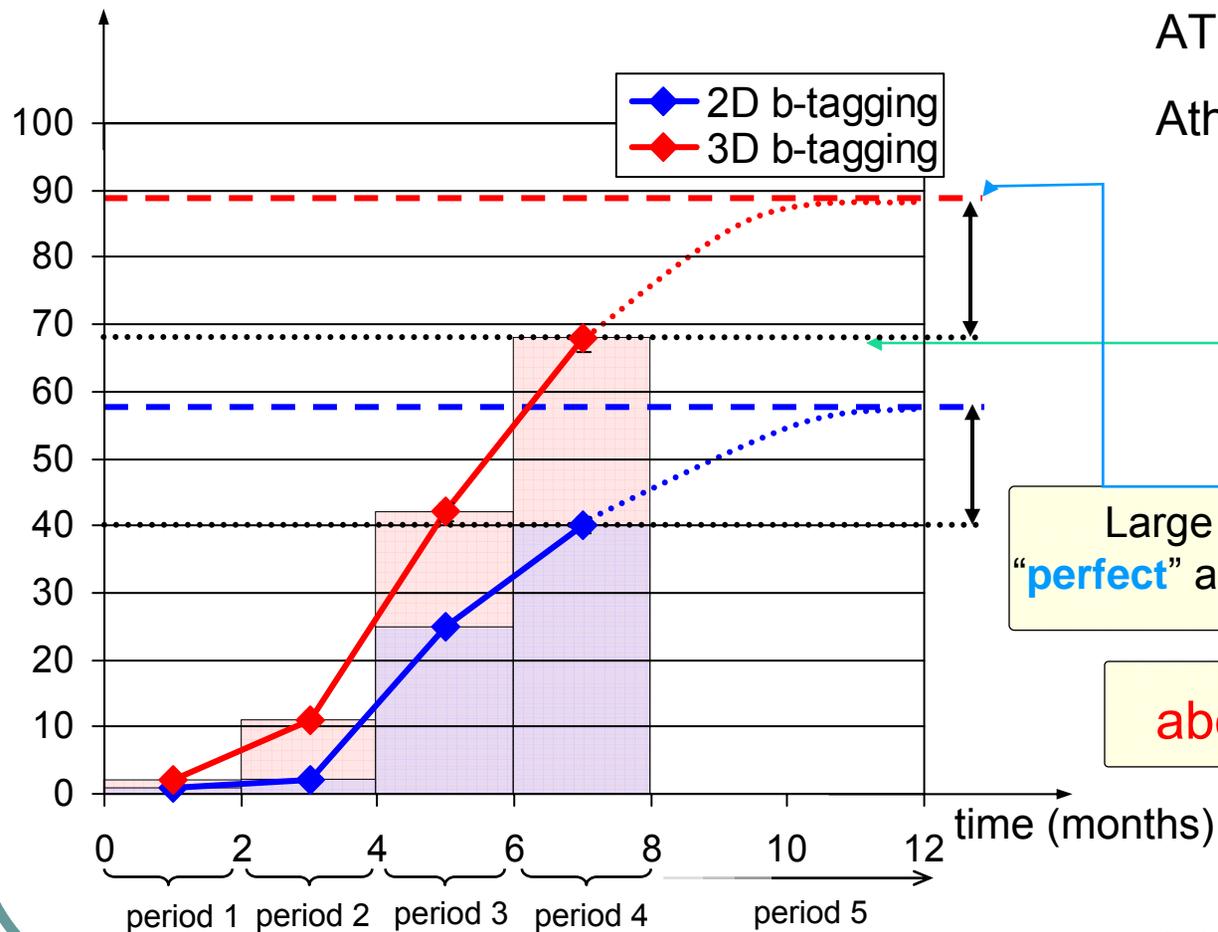


SecVtx Mistag Rate



Effect on performance

R_u : light jet rejection for $\varepsilon_b=60\%$



ATLAS

Athen's meeting 2003

Large difference between
"perfect" and "nominal" alignment:

about 25% !

Sébastien CORRÉARD

Alignment at Atlas

- Alignment must not degrade any track parameters by 20%

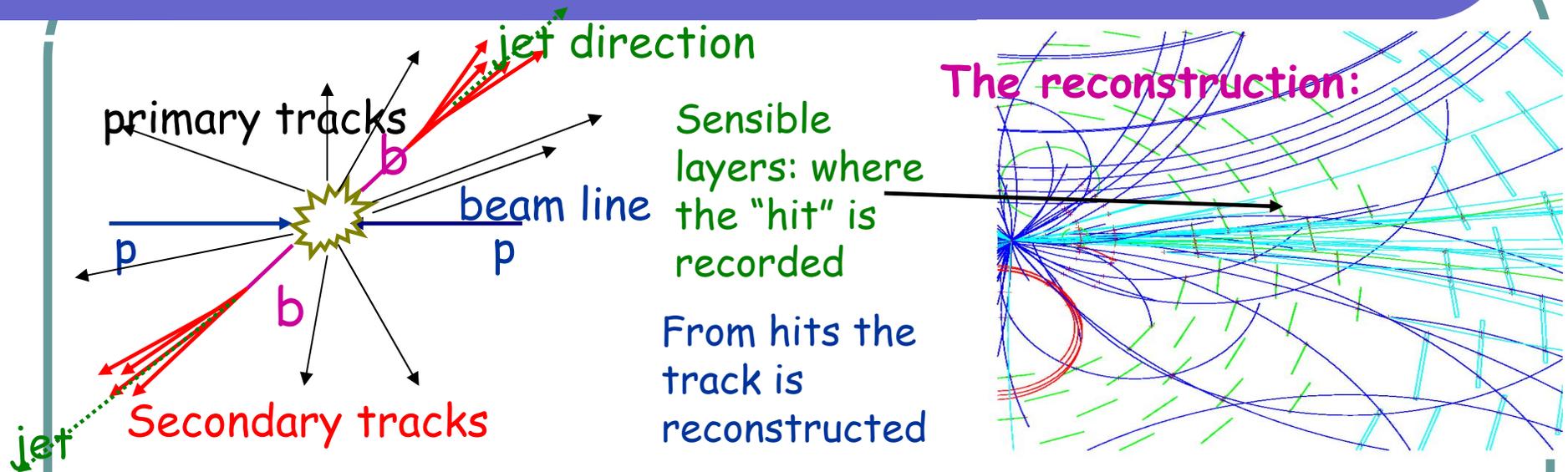
- Precision placement during building: pixels (10/100 μm), strips (50/250 μm)
- X-ray survey determine actual position 10/50 μm
- Continuous monitoring of deformation (scanning interferometer)

- Track alignment gives better local precision
- $M(w) < 15 \text{ MeV}$,
 - 1 μm in $r\text{-}\Phi$
 - B field to 0.02%
 - Detector Material known to 1%
 - P_T resolution to 1%
- Calibration with large sample of $Z \rightarrow \mu\mu$ possible

- Alignment algorithms under development:

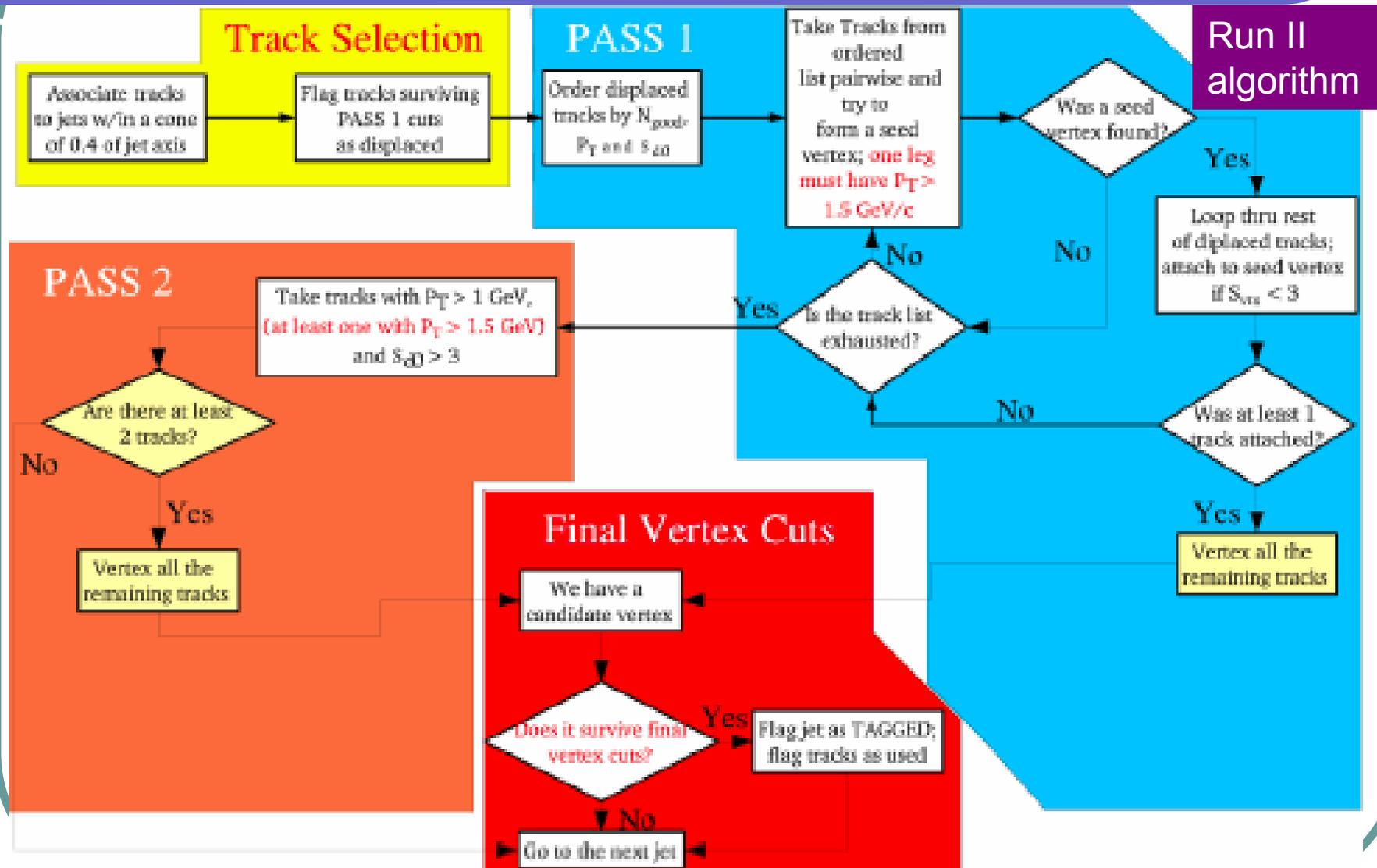
- χ^2 minimization (ALEPH and SLD)
 - Huge $30,000 \times 30,000$ symmetric matrix
- Iterative track fitting
 - Fit track
 - Plot hit residuals/module
 - Move module

CMS Trigger a b Jet with HLT



- From **pixel hits** and **calorimeters**:
 - The seed for tracks reconstruction is created around the **LVL1 jet** direction
 - Primary vertex is calculated
- Tracks are reconstructed **in a cone of $\Delta R > 0.15$** around the jet direction
- Tracks are **conditionally** reconstructed
- The **Jet direction is refined** using the reconstructed tracks

SECVTX ALGORITHM



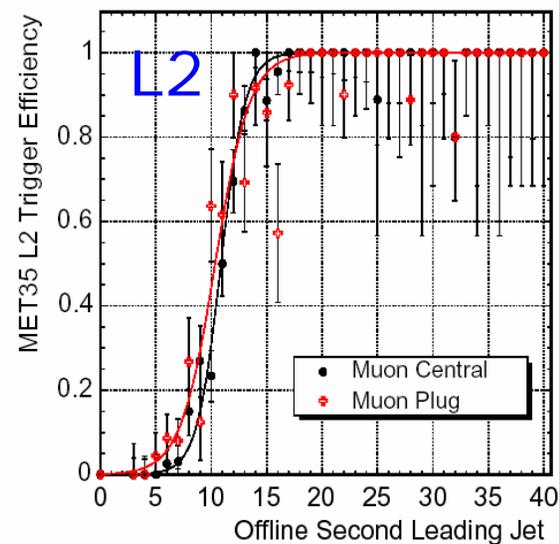
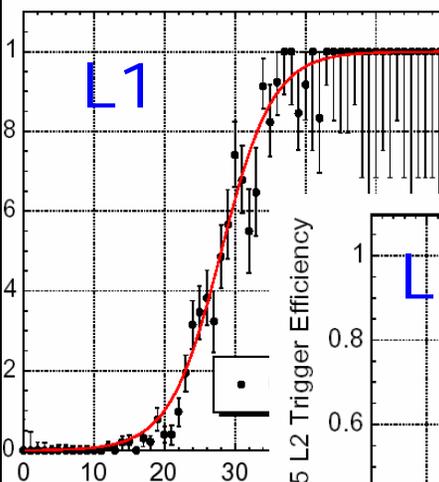
Trigger and systematics

• L1: $\cancel{E}_{T\ trg} \geq 25$ GeV

L2: Two clusters $E_T \geq 10$ GeV, $\eta = (0, 3.6)$

L3: $\cancel{E}_{T\ off} \geq 35$ GeV

Signal acceptance Systematic Uncertainties	Exclusive Single b- tag	Inclusive Double b- tag
	10.0%	10.0%
	7.5%	5.0%
	3.0%	3.0%
	2.0%	2.0%
	6.0%	6.0%
	2.0%	2.0%
	7.0%	14.0%
	2.5%	2.5%
	0.5%	0.5%
	0.5%	0.5%
Total	16.5%	19.5%



Backgrounds

Process	Exclusive Single B-Tag	Inclusive Double B-Tag
EWK	$5.66 \pm 0.76(stat) \pm 1.72(sys)$	$0.61 \pm 0.21(stat) \pm 0.19(sys)$
TOP	$6.18 \pm 0.12(stat) \pm 1.42(sys)$	$1.84 \pm 0.06(stat) \pm 0.46(sys)$
QCD	$4.57 \pm 1.64(stat) \pm 0.57(sys)$	$0.18 \pm 0.08(stat) \pm 0.05(sys)$
Total Predicted	$16.41 \pm 1.81(stat) \pm 3.15(sys)$	$2.63 \pm 0.23(stat) \pm 0.66(sys)$

$Wjj \rightarrow e\nu jj$

$Wjj \rightarrow \mu\nu jj$

$Wjj \rightarrow \tau\nu jj$

$Zjj \rightarrow eejj$

$Zjj \rightarrow \mu\mu jj$

$Zjj \rightarrow \tau\tau jj$

$Zjj \rightarrow \nu\nu jj$

$Zjj \rightarrow b\bar{b}jj$

ZZ

WZ

WW

EWK backgrounds estimated using
ALPGEN + HERWIG MC samples

- QCD-multijet

- HF QCD MC Fake b-tags are estimated using parameterization of the neg. tag rates obtained from data
- fake b-tags

- QCD uncertainty dominated by jet energy scale uncertainty