

Tau/b tag performance

津野 総司 (高工ネ研)

Flavor taggingで期待される物理や効果

- 1) 湯川結合、第3世代Flavor physics (SUSY/b) などの物理量の測定に決定的な役割を果たす。

Higgs粒子が軽い($m \sim 120\text{GeV}$) 時、 $H \rightarrow bb / \tau\tau$ が主要な崩壊モード。
 $gg \rightarrow H$ や、 $qq \rightarrow H$ などのProduction mechanics を組み合わせて、湯川結合の測定。

第3世代への崩壊様式を調べることは、SUSYなどの多様なモデルの判定に対して極めて有効。

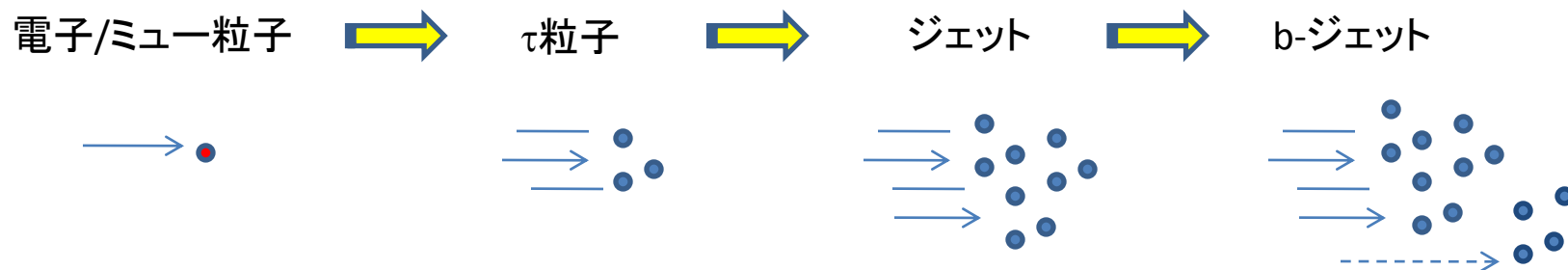
- 2) さらなるバックグラウンド(特にQCD) の除去にも有用。

tagging efficiency $\sim 60\%$ に対して、約 $\sim 1\%$ levelでfakeをsuppressできる。
つまり、S/N の向上。

b-jet vetoのように、bを持つプロセス(topなど)を取り除くことができる。

τ , b-jet 粒子同定 (基礎編)

基本的な描像 (どのように検出されるのか?)



- 一粒子系。とても細い。
- 非常にきれいな信号
- τ 崩壊に依る1、3粒子群。
- 電子より太いが、ジェットより細い。
- 多粒子の束として、観測される。太い。
- ジェットより多粒子。
- 2次崩壊点を持つ。

τ 粒子は、電子とジェットの“中間”のような性質を持つ。(細すぎず、太すぎず。)

b-ジェットは、荷電トラックをたくさん作る“傾向”がある。(ゴージャス)

➡ 一義的なカットで同定するのは困難。(pT, η 依存が大きい。)
多変数likelihoodなどによる同定が最も効果的。

データの理解(DATA \leftrightarrow MC の相互理解) とそれに伴う時間(期間)がキーファクター。

Overview

- 1) アルゴリズムの説明。
 - 2) 期待されるパフォーマンス。
 - 3) データでどう評価し、実験初期になにができるか？
 - 4) 物理解析に於ける適用法と留意点。
 - 5) 弱点の整理となにが求められているのか？
- } 簡単に説明。詳しくはCSC Book参照の事。

τ , b-tagging のポイントは、どこまで実験データを理解できるかということ。

(データの理解):

例えば、LHC実験初期 ... “安全”な変数のみを使う、もしくは、cut-based identification.

例えば、“現在のTevatron実験”... ほとんど、Neural Network Analysis.

Hadronic τ -identification

τ -IDアルゴリズム

τ 粒子は、電子とジェットの“中間”のような性質を持つ。
(細すぎず、太すぎず。)

トラッキング:

- Collimated 1 or 3 (2) tracks in core region.
- Isolation from other tracks.
- Make use of the impact parameter ($c\tau \sim 87\mu\text{m}$).

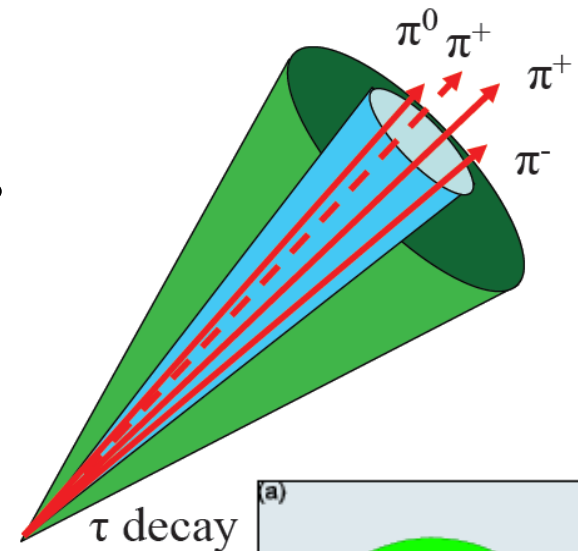
カロリメータ:

- Collimated deposition in EM.
(50% of energy by π^0 in 1-prong τ)
- Use shower shape.
- Both EM and HAD component.
- Isolation cone.
- Reconstruction of π^0 sub-cluster.

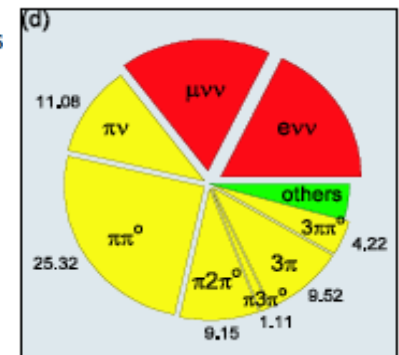
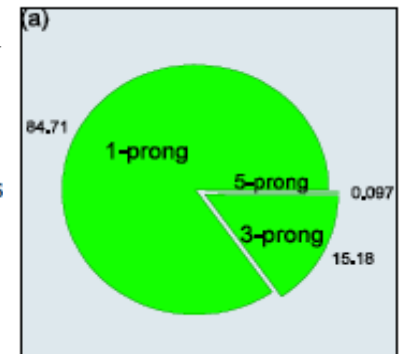
それぞれの変数が、すべて、電子とジェットの中間の性質を持つ。



Cut-basedで決めるのは難しい。(large p_T, η dependence)
Likelihoodなどの多変数同時フィットで同定する。



- Leptonic Decay Modes
 $e/\mu\nu e\mu\nu\tau$ 35%
- 1Prong Hadronic Decay Modes
 $\pi^-\nu\tau$ 11%
 $\pi^-\pi^0\nu\tau$ 25%
 $\pi^-\pi^0\pi^0\nu\tau$ 9%
 $\pi^-\pi^0\pi^0\pi^0\nu\tau$ 1%
 $K^- + \text{Neutrals}$ 1.5%
- 3Prong Hadronic Decay Modes
 $\pi^-\pi^+\pi^-$ 9%
 $\pi^-\pi^+\pi^-\pi^0$ 4.5%
 $K^-\pi^+\pi^-$ 0.4%
- Other Modes ($\sim 3\%$)

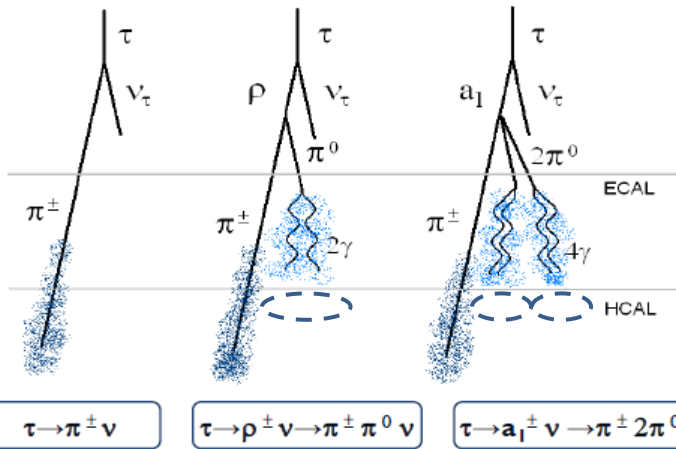


π^0 reconstruction

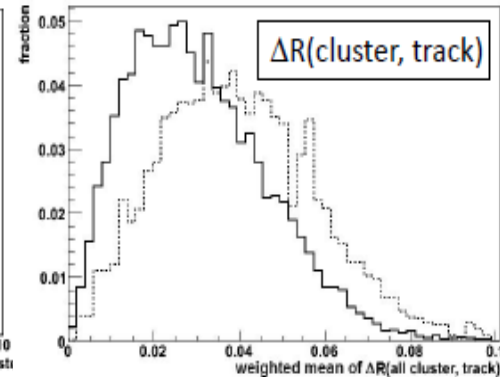
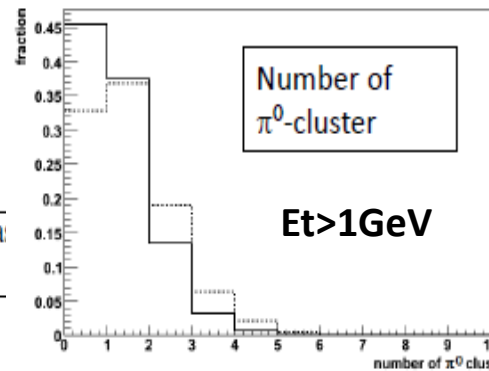
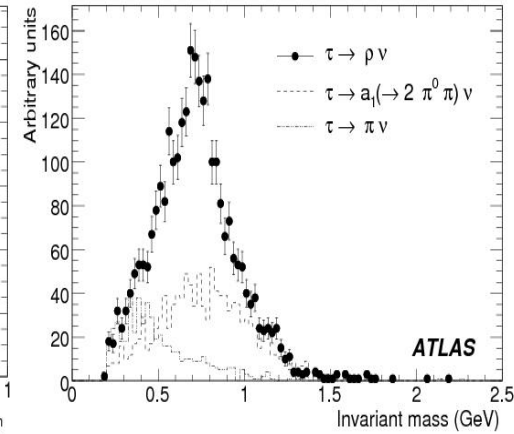
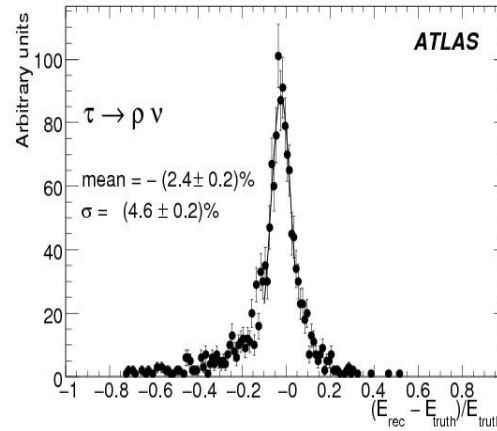
60% of decay products have π^0 .

Try to reconstruct $\pi^0 \rightarrow \text{gamgam}$ clusters.

Cluster depth を見る。



Seeds for track-based τ_{had} -candidates	Reconstructed as single-prong	Reconstructed as three-prong	Reconstructed as two-prong
Electron contamination (from conversion)	1.5%	5.7%	2.9%
$\tau \rightarrow \pi^\pm n \pi^0 \nu$	96.1%	3.8%	23.8%
$\tau \rightarrow 3\pi^\pm n \pi^0 \nu$	3.9%	96.2%	76.2%
Charge misid. (no had. interact.)	1.7%	3.6%	
	0.4%	2.1%	

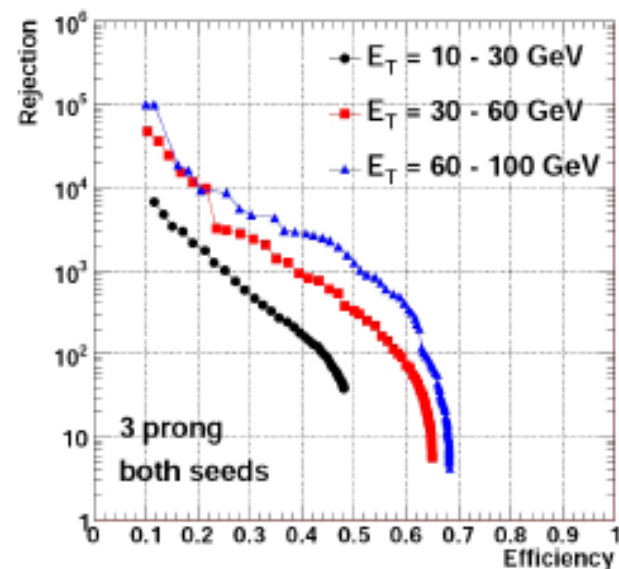
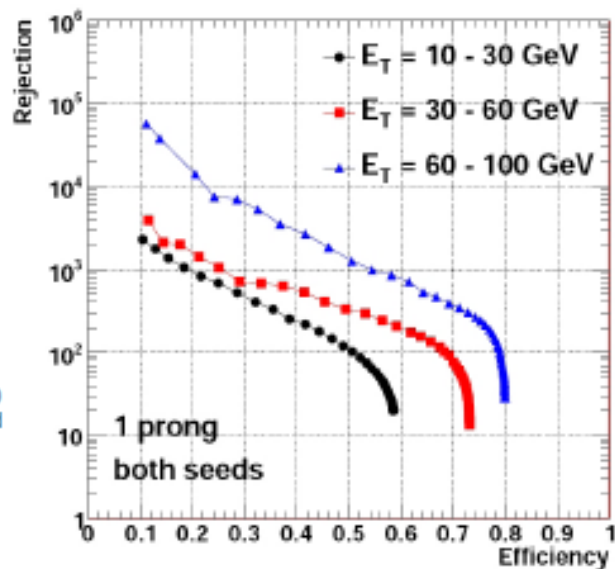
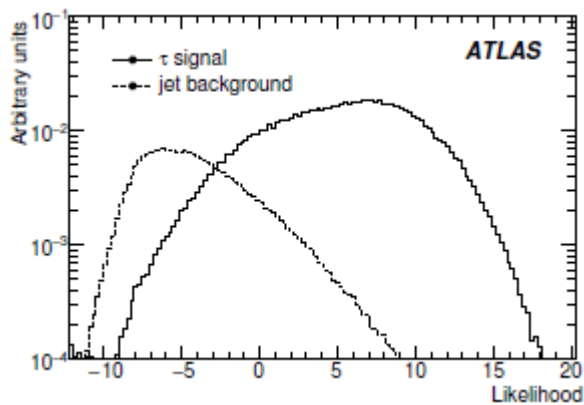


期待されるperformance

Log Likelihood Ratio identification ($= -\ln \sum_i (s_i/b_i)$ w/ 8param.) に対する Performance.

Performance は、 p_T , η と track multiplicity 依存。

$p_T \sim 30-60$ GeV (1-prong) で、eff. $\sim 60\%$ に対して、rejection ~ 200 .
(3-prong) $\sim 60\%$ ~ 80 .



現在、 π^0 の情報は十分に活用されていない。これらの情報も含めると 9% efficiency が向上する。また、2-prong tauのすみ分けも重要。

データの理解と系統誤差

Likelihood Ratio (s/b)は、シグナルは、MCから、ジェットは、実験データ(とMC)で。

シグナル(分子)の系統誤差評価:

実験データ(minimum bias)にMC tau eventを混ぜて評価する。

tracking efficiency in dense environment

shower shape の理解

ジェット(分母)の系統誤差評価:

ジェットのcomposition (gluon-jet? or quark-jet?) による不確かさが最も大きい。

ジェットデータ中、1st, 2nd pT jet (gluon/quark mixed sample) と

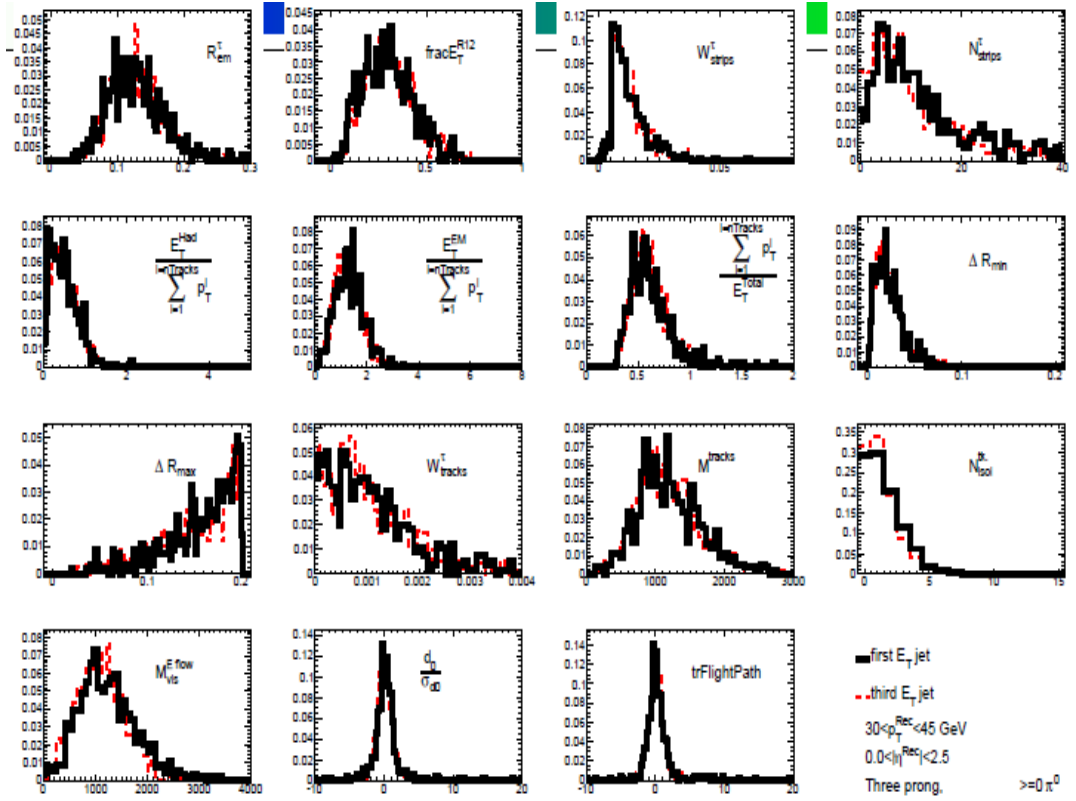
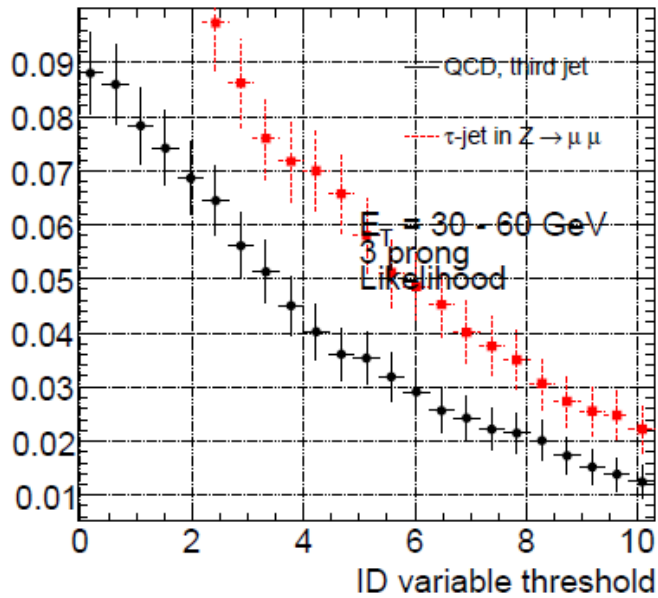
3rd pT jet (gluon-jet dominant).

実験初期には、“安全”な変数のみを用いる。

1st pT jet v.s. 3rd pT jet

違うjetを使って、likelihoodを作る。

3rd pT jet は、gluon dominant.





実験データを使った評価法

$Z \rightarrow \tau\tau$ ($W \rightarrow \tau\nu$) が、唯一の physics calibration process

$Z \rightarrow \tau\tau$ の断面積から efficiency/energy scale を評価する方法 :

- track multiplicity fitting method ... (jet shape の理解が重要)
- same sign charge correlation method ... (理論(MC)による不定性が大きい)
- Z mass side-band subtraction method ... (mass shape の理解)
- a.k.a. System 8 method ... (統計的に厳しい)

弱点:

- 1) Z の生成断面積が相対的に小さい。  利用できる実験データの統計量が少ない。
- 2) tau の p_T range が、比較的限られている。  統計的に High p_T tau の評価が困難。

さらに野心的な方法 :

ジェットによる偽の tau から、MC transfer function を使って引き出す。

 High p_T tau に有効?

実験初期のperformance

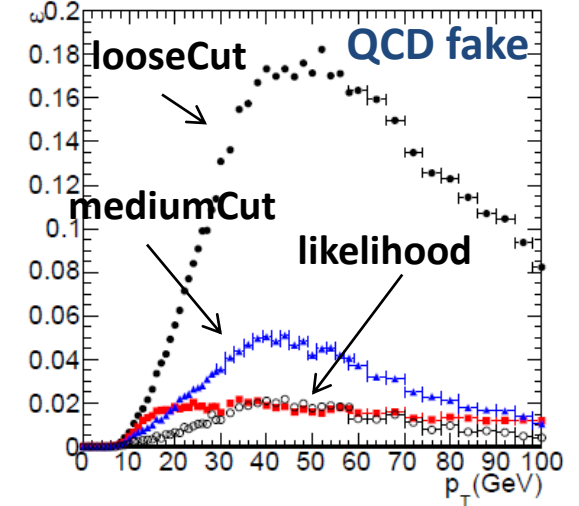
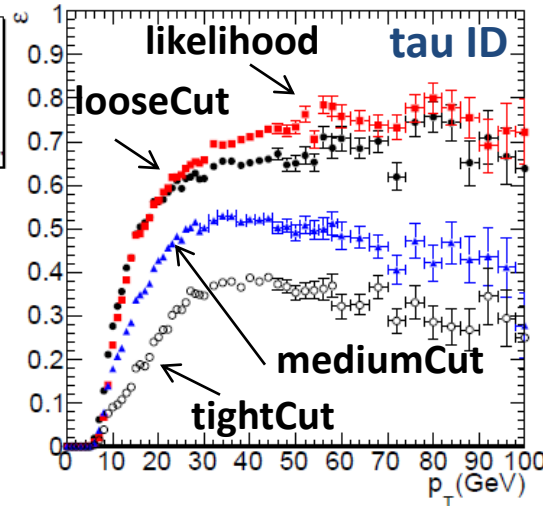
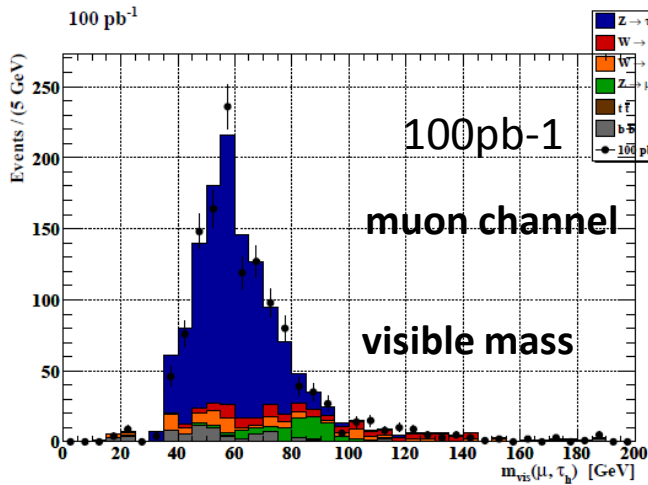
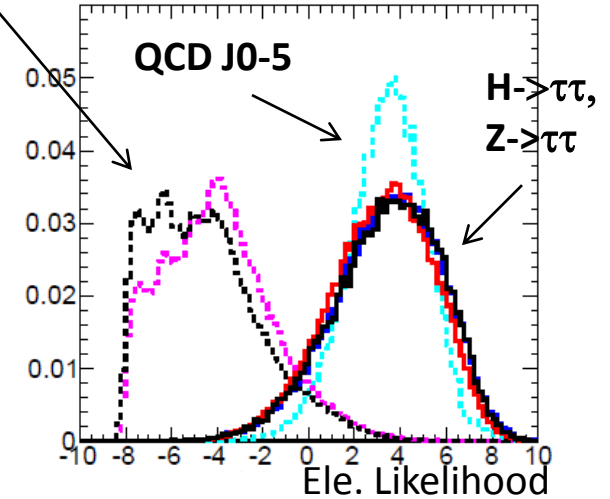
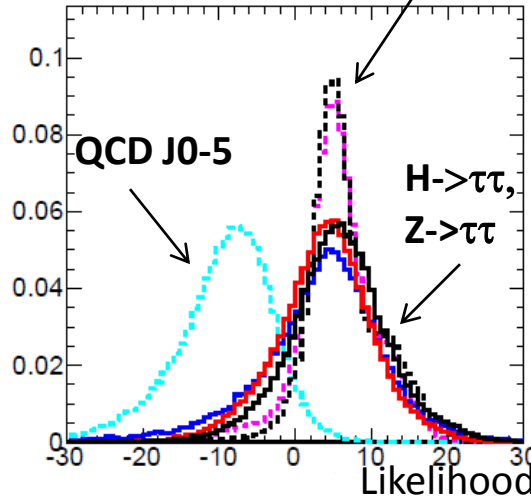
3つのcut-based IDが用意されている。

looseCut eff~70%, rej~16%,
mediumCut eff~50%, rej~5%,
tightCut eff~40%, rej~2%.

エネルギー補正は、Zを使う。

Z->ee, W->enu

Rel.14, 10TeV



b-tagging performance

b-taggingアルゴリズム

荷電トラックをたくさん作る“傾向”がある。

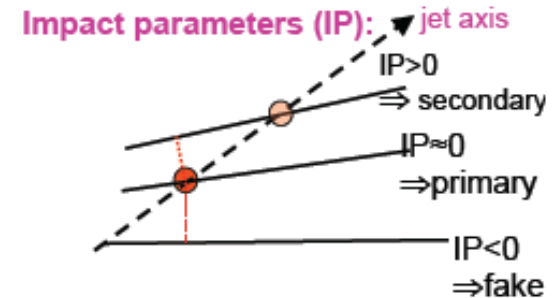
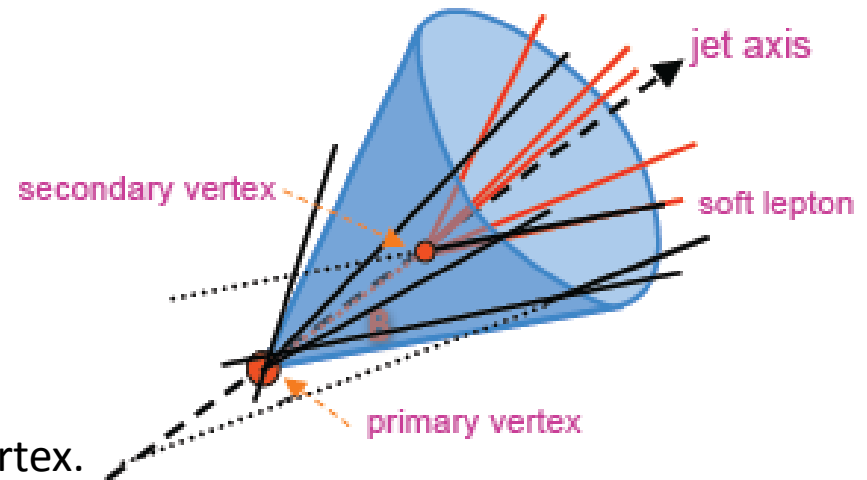
- 符号つきimpact parameterの利用
B-hadron flies ~ a few mm.
Tracks are far away from primary vertex.
- Secondary vertex再構成
direct reconstruction of the displaced vertex.
- Soft-lepton tagging
~40% of B-hadron decays into leptons.
b->l or b->c->l

利用できるparameterは全部使う。

Many taggers: IP2D, IP3D, SV1, SV2, CMB, LF2D, SVBU, LHSIG, SOFTM, SOFTE, JETFIT, JETFITCOMB(NN,COMBBB). 14種類!

Most sophisticated tagger : IP3D+SV1 (best performance)

実験初期 : Jet Probability tagger



Lifetime tagger

B-hadron $c\tau \sim 450\mu\text{m}$.

Track IP resolution : $\sim 40\mu\text{m}$ at $p_T=5\text{GeV}$.

JetProb. Tagger :

実験データ中、Negative IPのみ使って、
Light jet probability を構成する。

⇒ 実験初期からでも比較的利用できる。

Likelihood Ratio :

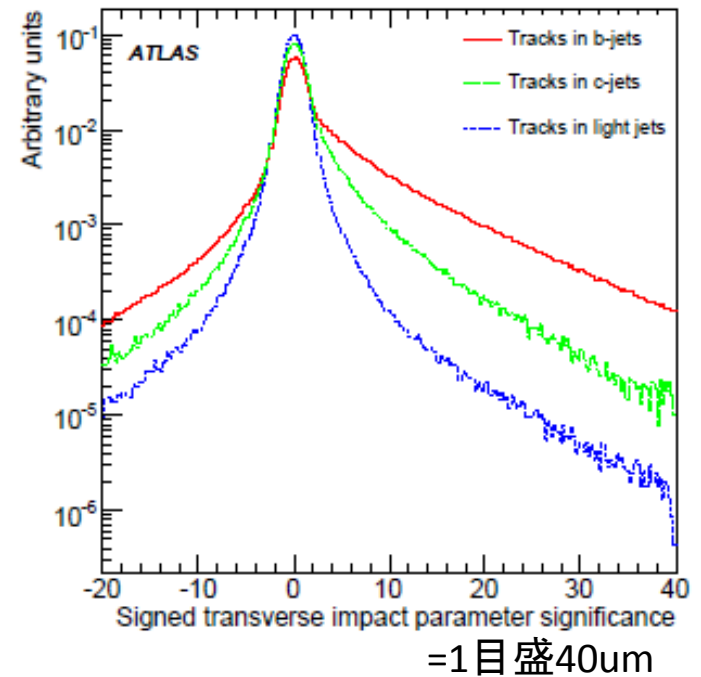
b-, c-, light-jet のprofileからlikelihood を構成。

⇒ どうやって、profileを作る？
large process dependent.

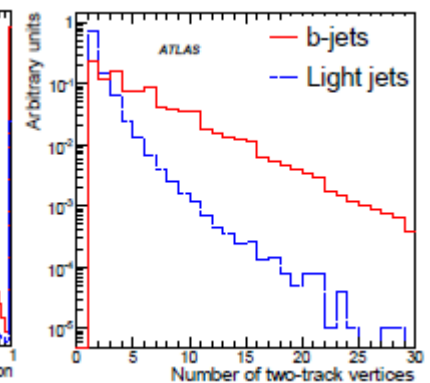
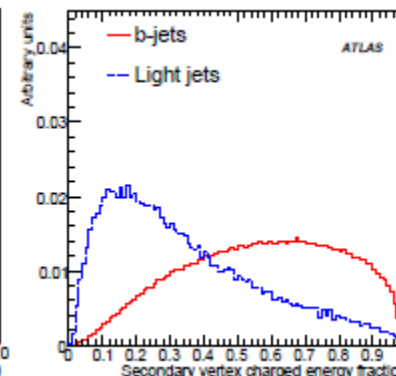
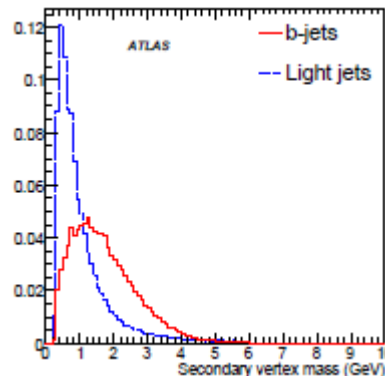
Secondary vertex reco. :

Primaryから離れたvertex
(≥ 2 trks.)を探す。

⇒ efficiencyの理解。



Secondary vertex decay in flight $\sim 3\text{mm}$ at 50GeV



期待されるperformance

IP3D+SV1 : efficiency ~50-60%
with rejection ~100-1000.

JetProb : eff ~50-60%,
with rejection ~ 20-100.

Soft lepton :
eff ~10% for muon,
7% for electron
with rejection 400/100
(incl. Br(b->l)~40%).

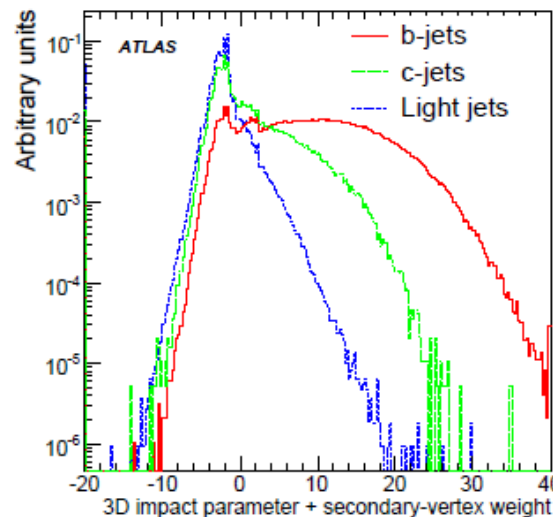
Under pileup L=10³³,
10-15% degradation.

実験初期に有効

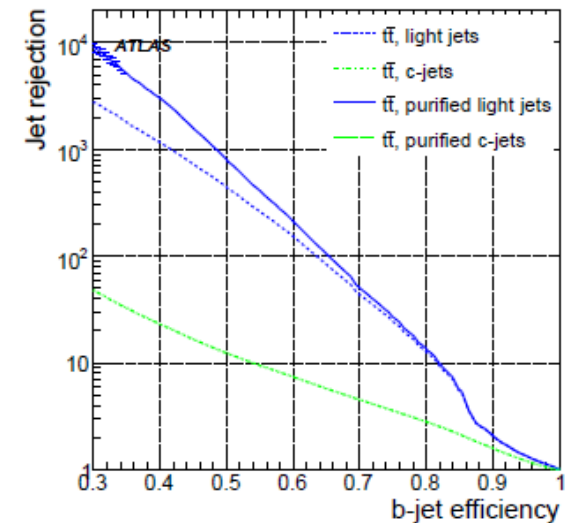
best performance

	JetProb	IP2D	IP3D	IP3D+SV1	IP3D+JetFitter
<i>WH (m_H = 120 GeV) events</i>					
$\epsilon_b = 50\%$	83±1	116±2	190±3	458±13	555±17
$\epsilon_b = 60\%$	30±0	42±0	59±1	117±2	134±2
<i>WH (m_H = 400 GeV) events</i>					
$\epsilon_b = 50\%$	73±1	163±3	179±3	298±7	396±11
$\epsilon_b = 60\%$	27±0	56±1	58±1	96±1	123±2
<i>tt and ttjj events</i>					
Raw, $\epsilon_b = 50\%$	91±0	146±1	232±2	456±4	635±7
Purified, $\epsilon_b = 50\%$	97±0	186±1	310±3	789±10	924±13
Raw, $\epsilon_b = 60\%$	28±0	46±0	67±0	154±1	189±1
Purified, $\epsilon_b = 60\%$	28±0	51±0	76±0	206±1	224±2

IP3D+SV1 : LLHR



ttbar : dense sample



データの理解と系統誤差

1) track finding efficiency

ジェットのpTに強く依存。(ジェット中の粒子密度)

High pT になるとIP resolutionが悪くなる。

track quality の分類: shared hitをどう扱うかなど。

2) 物質とmisaligned/distortion, dead channels

色々なモデルでテスト。~15% degradation

local distortionはそれほど問題なし。

global misalignmentは、secondary vertex

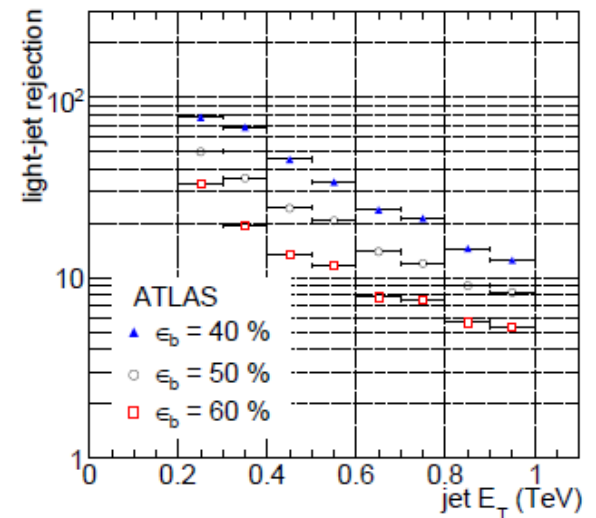
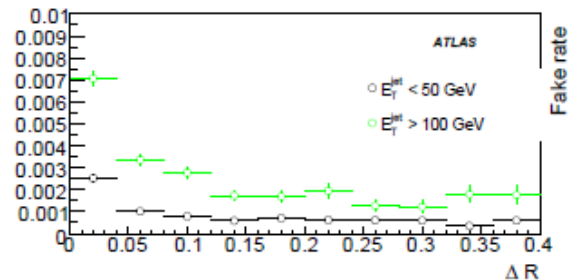
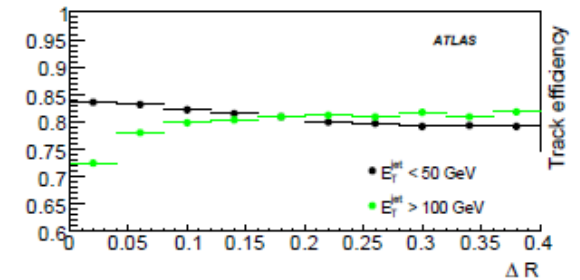
reconstructionに影響。

3) Likelihood のprofileをどう作るか?

light-jet, c-jet, b-jetの定義をどうするか?

例えば、gluon splitting bb/cc は、b(c)-jet or gluon jet?

physics process dependent. (generator dependent)



実験データを使った評価法

top pair production process と QCD bbbar process

ttbar eventを使う方法 :

- tag counting method
b-jetとlight jetの数によってイベントを分類し、
個々のfractionをlikelihoodでフィットする。
S/B~100 at 2b+2j。

➡ 100pb⁻¹ で、2.7% のeff. 統計誤差が期待できる。

- topological / likelihood / kinematic methods
積極的にkinematicsを利用して、ttbar eventを取り出す。
efficiencyのpT, η依存性を評価するのに有効。

➡ 統計量がボトルネック。200pb⁻¹ で、6.4%のeff.統計誤差。

di-jet eventを使う方法 :

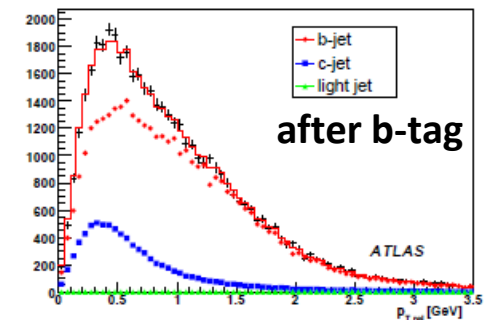
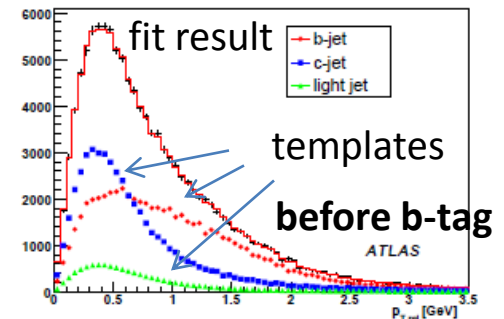
- pT rel method
ジェット軸に対して、semi-leptonic muonの相対的なpT分布を
各b-,c-,light-jetのtemplateとして用いて、fractionをフィットに
よって求める。
- System 8 method
uncorrelated taggerを用いて、8個の変数、8連立方程式から、
直接、efficiencyを得る。(SLT v.s. vertex tagger)

tag counting method

Number of light jets	1	2	3	4	any
2 b-jets, 1 c-jet	15.9 %	10.6 %	3.00 %	0.85 %	30.5 %
2 b-jets, 0 c-jet	-	24.4 %	13.8 %	3.74 %	42.8 %
1 b-jet, 1 c-jet	-	6.46 %	2.38 %	0.74 %	9.75 %
1 b-jet, 0 c-jet	-	-	7.60 %	3.21 %	11.6 %
0 b-jet, 1 c-jet	-	-	0.38 %	0.11 %	0.49 %
0 b-jet, 0 c-jet	-	-	-	0.46 %	0.65 %

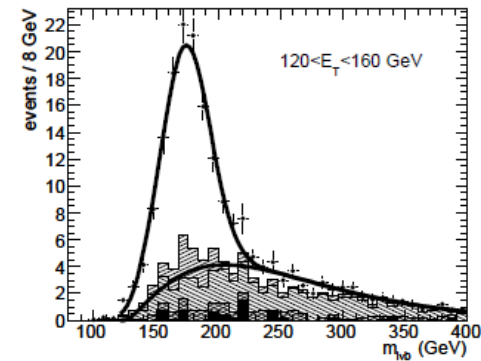
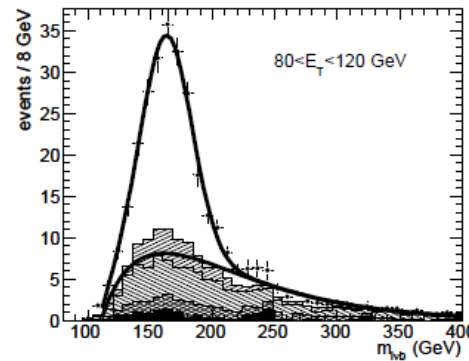
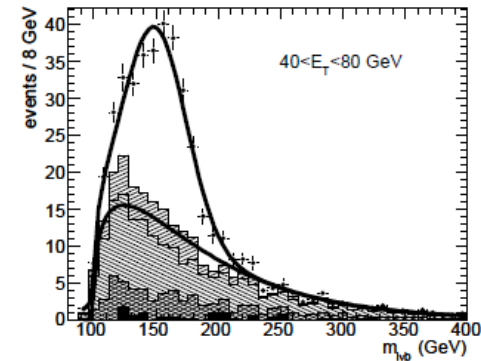
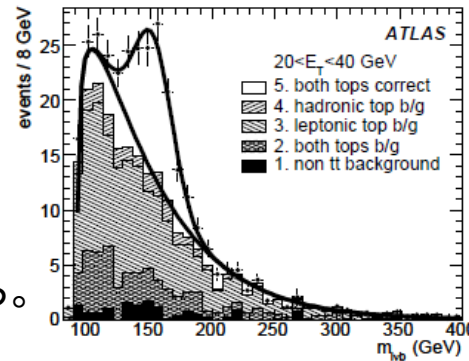
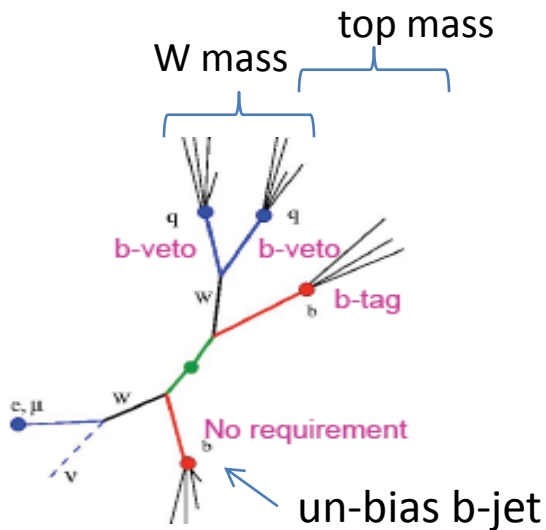
ttbar dominant



pTrel method



ttbar kinematics を使う方法

- 1) hadronic sideに対して、W/top mass kinematical conditionを課して、ttbar control sampleを作る。
- 2) そのsampleに対して、leptonic sideのb-jetにb-taggingをapply。
- 3) backgroundは、side-band subtractionから。





 signal region side-band

**b-jetの各pT, η binでefficiencyを出す。
sampleの統計量がボトルネック。**

注意：私の偏見に基づいています。

物理解析に於ける τ / b-tagging

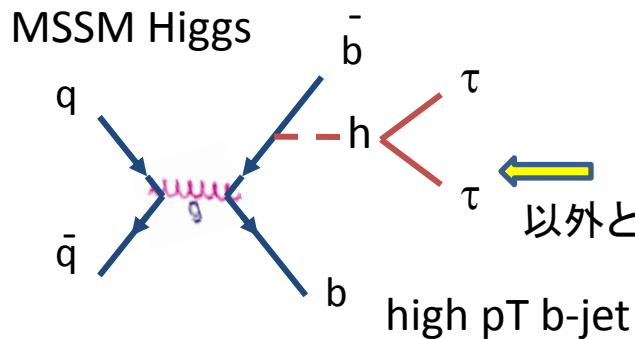
bと τ は相性が悪い。

JetProb : fake 10-20% from τ

tauのlifetimeはc-jetと同程度。
tau IDも状況は同じ。

high p_T τ / b は、どうやってIDするか？
~>500 GeV のsignal (τ or b) に対して、
efficiencyとrejectionは同程度。

	JetProb	IP2D	IP3D	IP3D+SV1	IP3D+JetFitter
<i>c</i> -jet rejection for WH ($m_H = 400$ GeV) events					
$\epsilon_b = 50\%$	7.9 ± 0.1	9.7 ± 0.1	10.7 ± 0.2	12.4 ± 0.2	12.7 ± 0.2
$\epsilon_b = 60\%$	4.7 ± 0.0	5.7 ± 0.1	6.1 ± 0.1	6.8 ± 0.1	7.3 ± 0.1
<i>c</i> -jet rejection for $t\bar{t}$ and $t\bar{t}jj$ events					
$\epsilon_b = 50\%$	8.4 ± 0.0	9.5 ± 0.0	10.6 ± 0.0	12.4 ± 0.1	12.3 ± 0.1
$\epsilon_b = 60\%$	5.1 ± 0.0	5.8 ± 0.0	6.5 ± 0.0	7.4 ± 0.0	7.4 ± 0.0
τ -jet rejection for $t\bar{t}$ and $t\bar{t}jj$ events					
$\epsilon_b = 50\%$	10.2 ± 0.1	13.9 ± 0.1	20.3 ± 0.2	45.2 ± 0.8	36.9 ± 0.6
$\epsilon_b = 60\%$	5.1 ± 0.0	6.4 ± 0.0	8.0 ± 0.1	24.6 ± 0.3	19.3 ± 0.2

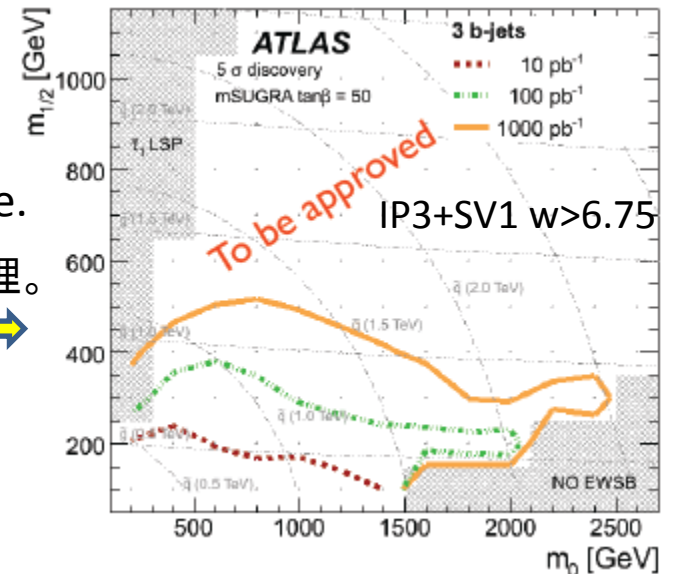


SUSY param. scan
incl. 3-b final state.

JetProbなら、無理。

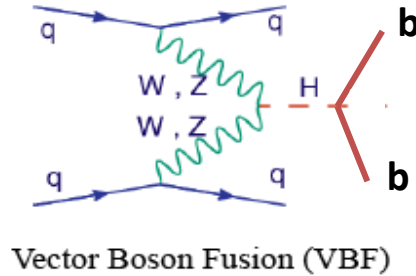
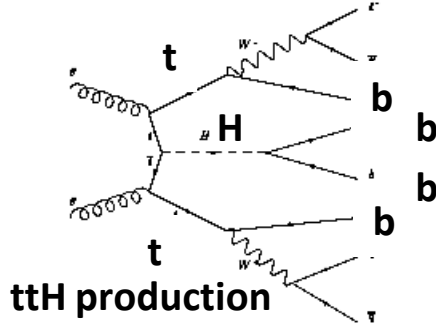


以外と難しいかも？

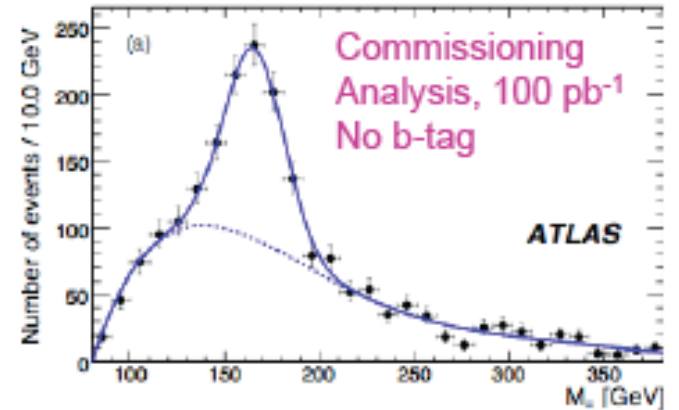


H->bb mass resolution

H->bb の質量分解能を向上させるのが至上命題。



Energy 補正は、top sampleを使う。



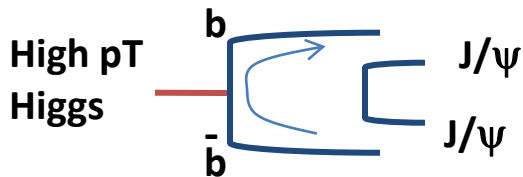
通常のjet energy correctionでは、 $\sigma \sim 20\text{GeV}$

“シロウト的”には

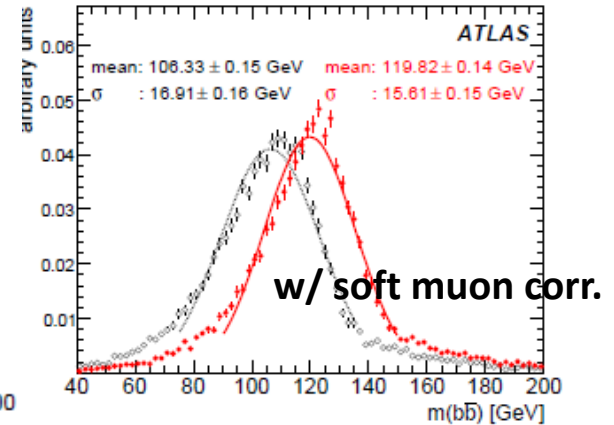
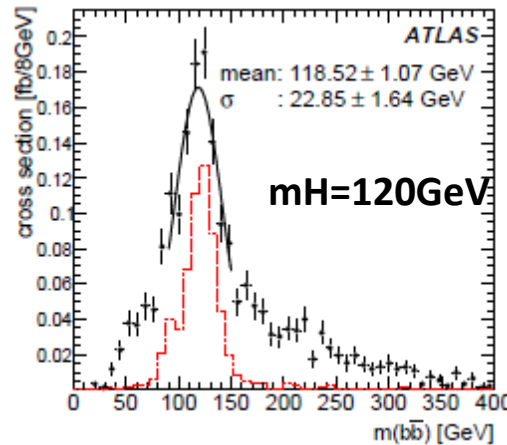
H->bbのcolor singletの性質を生かせないか？

→ high pT boosted Higgs

例えば、Quarkonium を形成するか？



とか。。。 で、spectrumとHiggs massの相関を見るとか。。。



まとめ

τ / b-jetの観測される特徴を最大限に生かすためには、likelihoodなどを用いた同定方法が最も適している。そのため、実験データとMCの相互理解が非常に重要である。

実験初期は、“安全”な変数のみで構成される。

τ ID : (実験初期) eff=50% , rejection \sim 20 , (best performance) eff=70% , rej \sim 50
b-tag : (実験初期) eff=50% , rejection \sim 80 , (best performance) eff=60% , rej \sim 100

Physics processによる実験データを用いた補正:

τ ID : Z $\rightarrow\tau\tau$... pT rangeが限られている。実験初期には、統計量がボトルネック。
b-tag : ttbar , dijet ... 実験初期には統計量がボトルネック。

high pT領域が弱点:

ジェットが細くなる事と、impact parameter resolutionが悪くなることにより、efficiencyとrejectionが共に悪化する。

さらなる改善点:

τ ID : pi0 reconstructionと2-prong tauの分類。さらにenergy flowが進むべき道。

b-tag : 異なるtracking algorithmを並行して使うなどしたtrackの分類付け。

(Non-Gaussian tailの理解)

(pileup toughness等も)

QCDジェットの定義問題: どう定義し、どう解釈するか? process dependenceは克服できるか?

Another topics:

Tau/b IDに於いて、大きなrejection factorは必要か？

- > 大抵の解析に於いて、必要なし。
感覚的には、 $e/\mu\text{on}$ ($\text{rej} \sim 10^4$)が要求されていれば、
fake rate $\sim 1, 2\%$ で対応できる。

足りないもの、 $Z \rightarrow b\bar{b}$ をデータで検証。

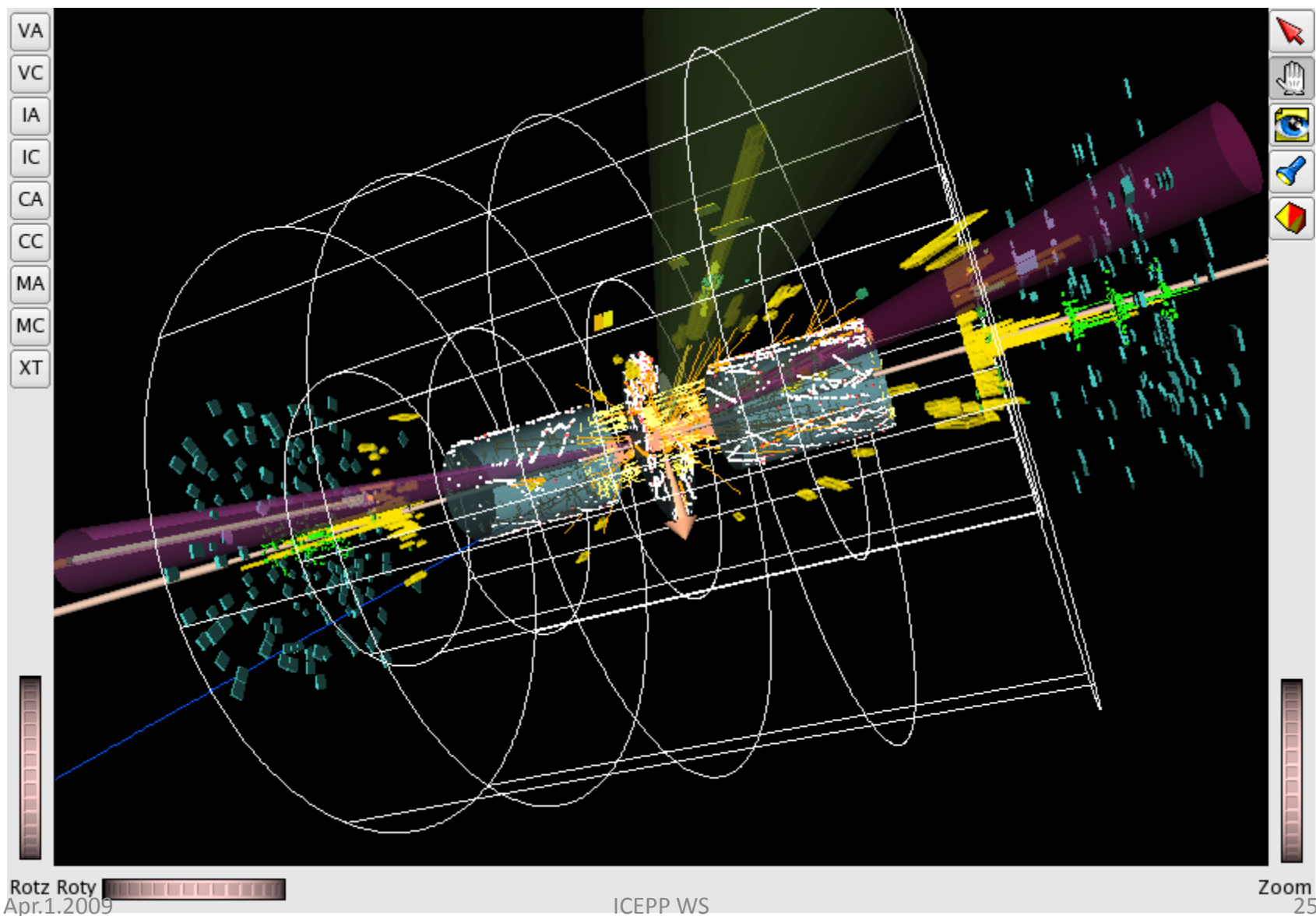
- > $Z + \gamma$ processを使えないか？ (γ でトリガー)

$Z + b\bar{b}$ を真面目にやるべし。

- > double countingの良い検証process
 $H \rightarrow 4\text{lepton}$ に切り込めます。

Backup

Event display (VBF H->tautau)



Z -> $\tau\tau$ measurement (I)

Fit track multiplicity distribution.

(previous method tells nothing about fake-tau.)

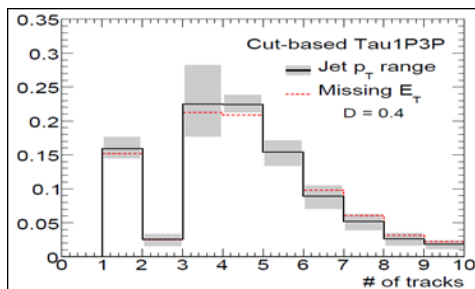
Overall QCD fake estimation (at Control region)

Track multiplicity fit:

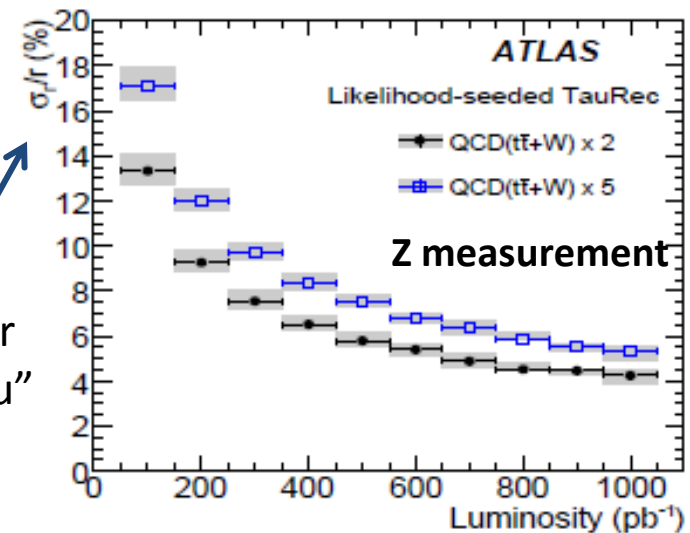
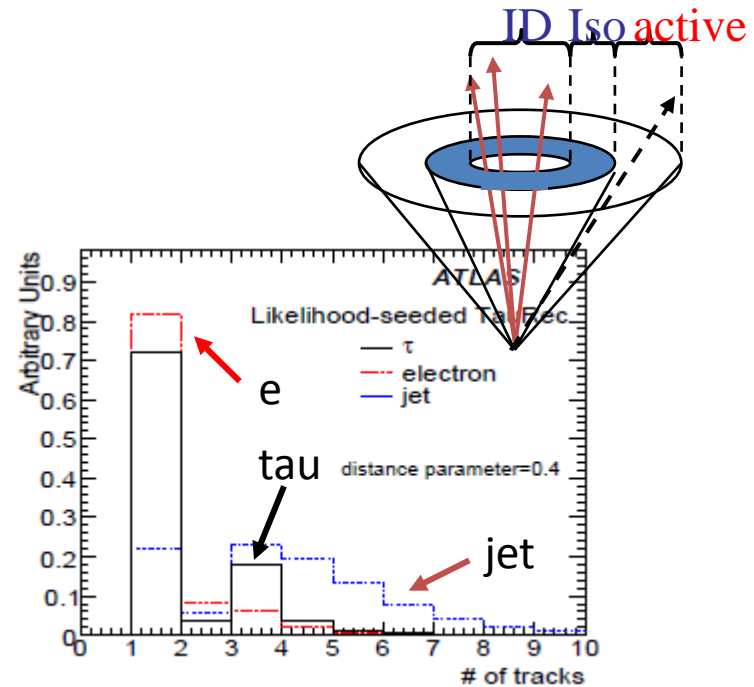
- count the track outside tau core region.
- fit the n-tracks by likelihood.

$$L = \prod_i^N \text{Pois}(n_{csp}^{tot} \times (r_{\tau} f_{\tau}^i + r_{lep} f_{lep}^i + (1 - r_{\tau} - r_{lep}) f_{jet}^i), N_{obs}^i) \\ \times \text{Gaus}(n_{csp}^{tot}, N_{obs}^{tot}) \\ (\times \text{Gaus}(n_{csp}^{tot} r_{lep}, N_{lep}^{measured})) \quad (\text{fit the ratio})$$

- shape (pdf) are modeled by real-data. (ex. QCD jet shape by JET triggered data)



expected error of "ratio of tau" in the events



We don't know the size of QCD jets, so test hypothesis by increasing x2 and x5 QCD events.

Z -> $\tau\tau$ measurement (II)

comment on “same sign subtraction method”.

“SS subtraction method” is used to estimate fake-tau contribution by assuming that the fake-tau has equal probability to be positive or negative charge in tau.

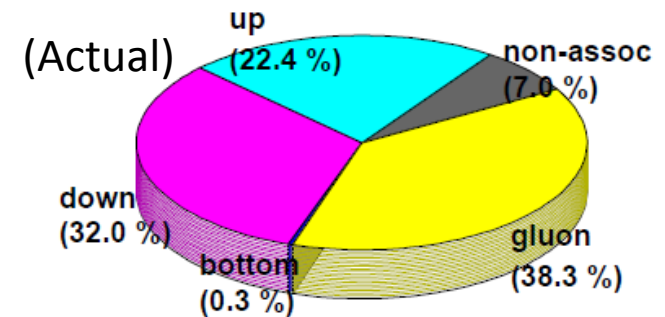
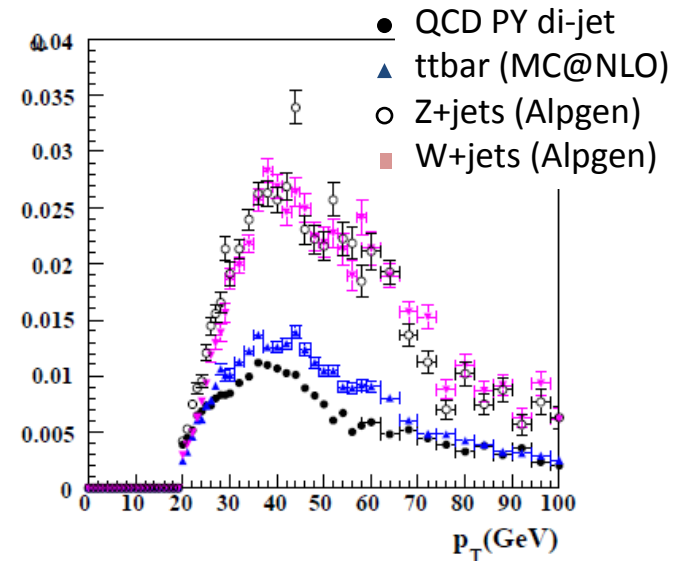
However, QCD fake-tau, $N(OS) \neq N(SS)$ (asymmetric)

The up-type quark jet tends to be positive charge than that the down-type quark be negative charge.

LHC is not Tevatron : (LHC) In W+jets, “gluon”-collision is dominant, so that “quark”-jet is dominant in the final state.

(Tevatron)		(LHC)
$\sigma(qg \rightarrow Wq) \sim 107\text{pb}$	\longleftrightarrow	$\sigma(qg \rightarrow Wq) \sim 3185\text{pb}$
$\sigma(qq \rightarrow Wg) \sim 126\text{pb}$		$\sigma(qq \rightarrow Wg) \sim 939\text{pb}$

On the other hand, **gluon is dominant in QCD di-jet.**

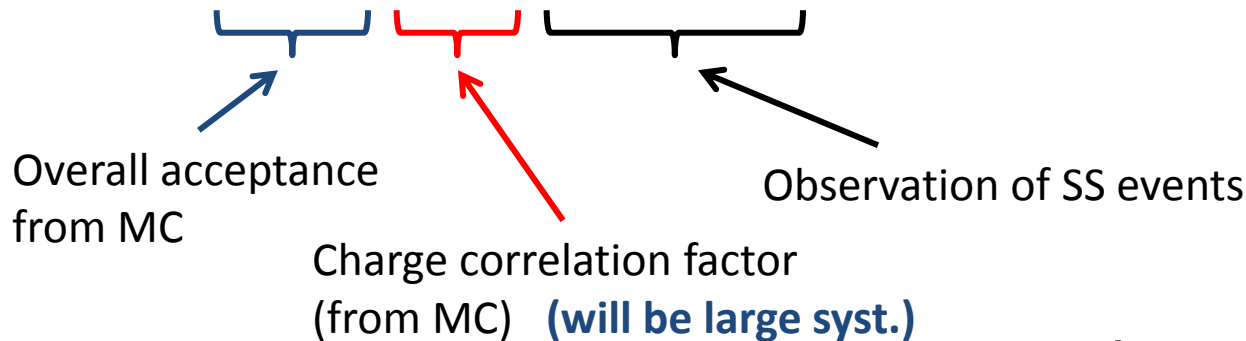


More detailed study and possible proposal will be presented in another next tau meeting.

Possible proposal

If we use SS subtraction method, fake-tau contribution will be :

$$\sigma_{\text{fake-tau}} = f_{W+\text{jets}} \cdot C_{SS}^{W+\text{jets}} \cdot N_{\text{obs}}^{W+\text{jets}}(SS) + f_{t\bar{t}} \cdot C_{SS}^{t\bar{t}} \cdot N_{\text{obs}}^{t\bar{t}}(SS) + f_{\text{QCD}} \cdot C_{SS}^{\text{QCD}} \cdot N_{\text{obs}}^{\text{QCD}}(SS)$$



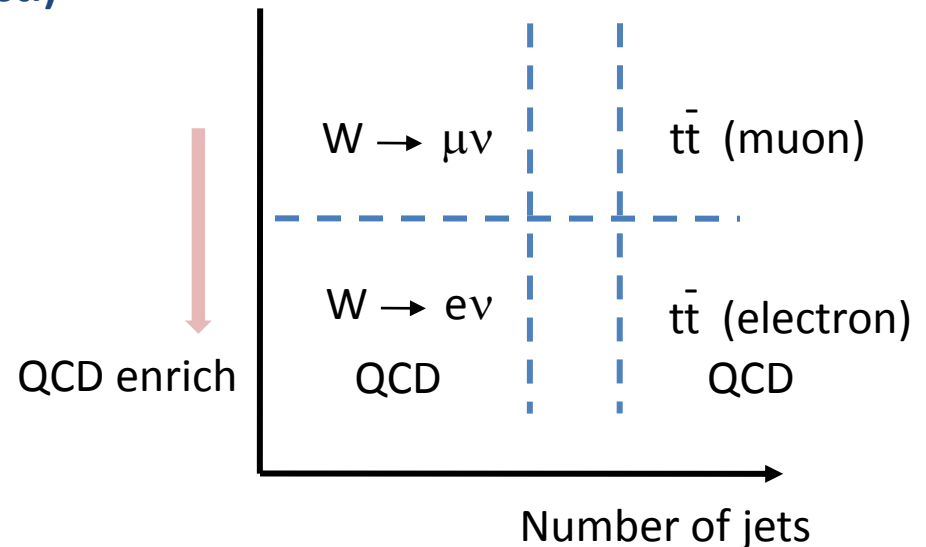
Procedure :

1) Suppose muon channel ($W \rightarrow \mu\nu$ & $t\bar{t}$) is QCD di-jet bkg. free.

A) From low jet-multiplicity region, $N(SS)$ for $W+\text{jets}$

B) From high jet-multiplicity region, $N(SS)$ for $t\bar{t}$

2) In electron channel, estimate QCD di-jet $N(SS) = N_{\text{obs}}(SS) - N_{W+\text{jet}}(SS, \text{muon})$

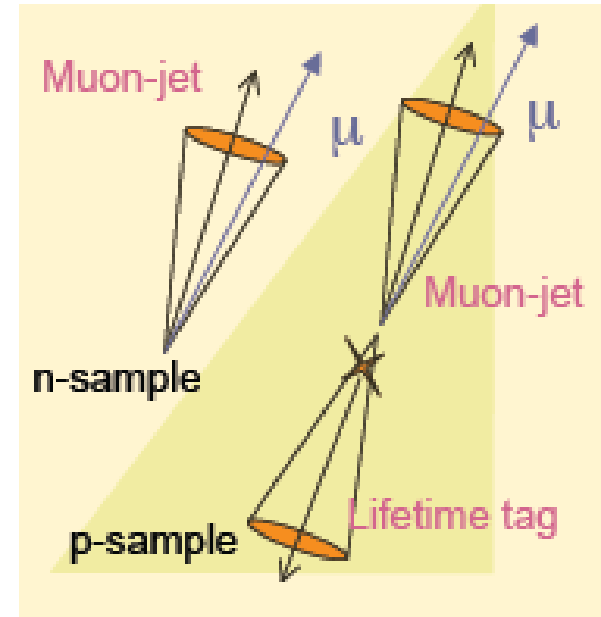


System 8

lifetime taggerとsoft muon taggerは、
uncorrelatedと仮定すると、

$$\begin{aligned}
 n &= n_b + n_{cl} \\
 p &= p_b + p_{cl} \\
 n^{LT} &= \epsilon_b^{LT} n_b + \epsilon_{cl}^{LT} n_{cl} \\
 p^{LT} &= \alpha_6 \epsilon_b^{LT} p_b + \alpha_4 \epsilon_{cl}^{LT} p_{cl} \\
 n^{SMT} &= \epsilon_b^{SMT} n_b + \epsilon_{cl}^{SMT} n_{cl} \\
 p^{SMT} &= \alpha_5 \epsilon_b^{SMT} p_b + \alpha_3 \epsilon_{cl}^{SMT} p_{cl} \\
 n^{both} &= \alpha_1 \epsilon_b^{LT} \epsilon_b^{SMT} n_b + \alpha_2 \epsilon_{cl}^{LT} \epsilon_{cl}^{SMT} n_{cl} \\
 p^{both} &= \alpha_1 \alpha_5 \alpha_6 \epsilon_b^{LT} \epsilon_b^{SMT} p_b + \alpha_2 \alpha_3 \alpha_4 \epsilon_{cl}^{LT} \epsilon_{cl}^{SMT} p_{cl}
 \end{aligned}$$

8個の連立方程式から、8変数の解を得る。



Lxy disribution of b and tau

For fake tau candidates

