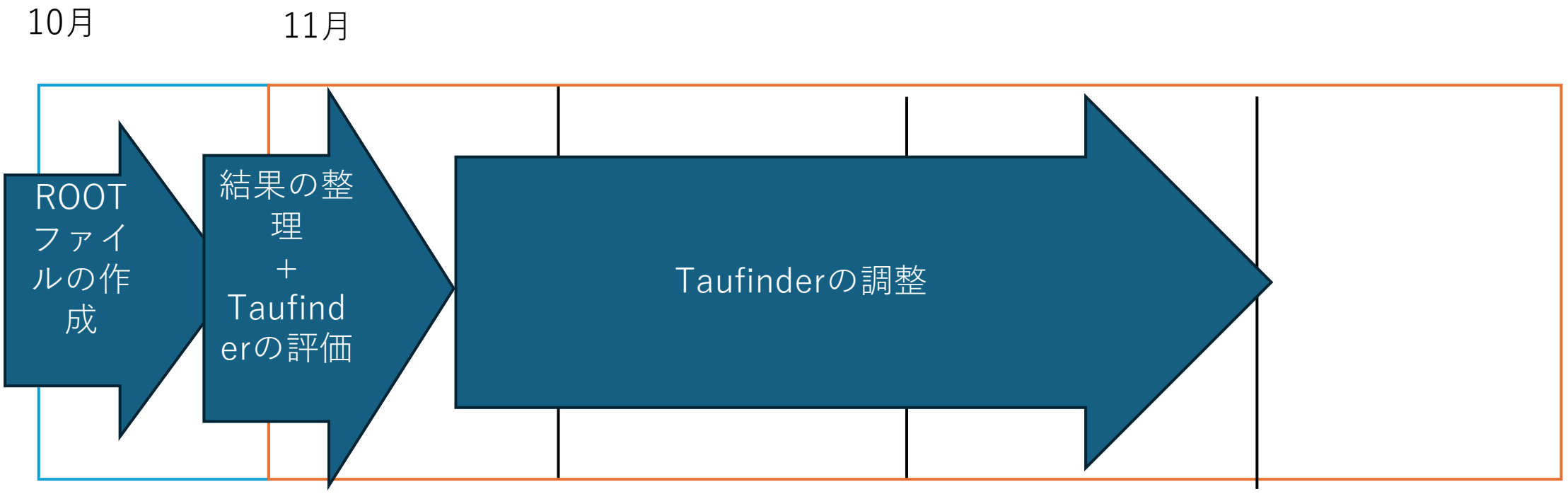
2026_11_4

進捗報告

伊藤悠風

活動内容

- 結果の整理
  - バックグラウンド$WW \to \mu\nu q\bar{q}, \tau\nu q\bar{q}$の$\tau\nu q\bar{q}$について評価
  - 現状のtaufinderの性能についての評価
  - 中間報告準備



重み付けのやり直し

- **課題**
  - 本解析の初期段階において、清野さんの修論に記載された期待事象数と、自身の解析による重み付け事象数に大きな乖離が生じていた。
- **原因の特定（断面積の不整合）**
  - 事象の重み（Weight）は通常 (Cross Section × Luminosity) / (生成イベント数) で算出される。
  - 調査の結果、計算ロジック自体は正しいものの、**ROOTファイル内にメタデータとして保存されていた自動生成の Cross Section の値が、実際の解析に用いるべき実効断面積と異なっていた**ことが原因と判明した。
  - ※ファイル内の値は、ビーム偏極の条件やジェネレータの初期設定が最終的な物理解析の条件を正確に反映していなかった可能性が高い。
- **解決策と結果**
  - ROOTファイル内の値を盲信せず、先行研究（修論）で物理的に正しく評価・採用されていた Cross Section の値を直接用いて再計算を行った。
  - 結果として、先行研究の期待事象数とほぼ完全に一致し、正しいスケールファクター（Weight）で以後の解析を進める土台が整った。

重み付けの結果

生成された生イベント数と重み付けされたイベント数

	プロセス	ファイル名	生イベント数	重み付けされたイベント数
signal	$ZH \to \nu\bar{\nu}s\bar{s}$	nnss	200000	13.8
背景事象	$ZH \to \nu\bar{\nu}b\bar{b}$	nnbb	200000	40287.0
	$ZH \to \nu\bar{\nu}c\bar{c}$	nncc	200000	1999.0
	$ZH \to \nu\bar{\nu}g\bar{g}$	nngg	200000	5650.0
	$ZH \to \nu\bar{\nu}ZZ^*$	hzz	1000000	1816.0
	$ZH \to \nu\bar{\nu}WW^*$	hww	1000000	14707.0
	$ZH \to \nu\bar{\nu}\tau^+\tau^-$	htautau	1000000	4318.0
	$ZZ \to q\bar{q}l^+l^-/q\bar{q}\nu\bar{\nu}$	4fZZsemileptonic	3099600	311961.0
	$\nu\bar{\nu}Z \to \nu\bar{\nu}q\bar{q}$	4fZnusemileptonic	2298400	235599.0
	$WW \to \mu\nu q\bar{q}, \tau\nu q\bar{q}$	4fWWsemileptonic	48112000	4189999.9
	$WW \to e\nu q\bar{q}$	4fWsl	45357600	2287805.0
	$WW/ZZ \to l\nu q\bar{q}$	4fmixleptonic	8400000	850430.0

*今回作ったのはeL_pRのみ

Pre-cutの結果

- 条件付けによって、背景事象を一定数まで減らすことに成功。
- しかし、シグナルを失わずに背景だけをこれ以上削ることは、単純なカット（直線の境界線）では難しい。

⇒複数の変数の相関関係から複雑な境界線を引ける機械学習（BDT）へ移行

適用した条件(清野さんの修論と同じものを使用)

- Missingmass<140GeV,
- 100GeV<dijetmass<200GeV,
- 5GeV< Tmomentum <100GeV,
- P_{max}<80GeV,
- 9<$N_{charged}$,
- isolepton=0.

File Name	Original	After Cut	Efficiency
4fWWsemilep	4.19e+06	1.48417e+06	35.42 %
4fWsl	2.28781e+06	92479.9	4.04 %
4fZZsemilep	311961	129121	41.39 %
4fZnusemilep	235599	95223	40.42 %
4fmixlep	850430	36046.9	4.24 %
htautau	4318	347.215	8.04 %
hww	14707	7236.74	49.21 %
hzz	1816	971.146	53.48 %
nnbb	40287	37660.7	93.48 %
nncc	1999	1952.23	97.66 %
nngg	5650	5587.23	98.89 %
Total Background (B)	7.94457e+06	1.8908e+06	23.80 %
Signal (nnss) (S)	13.8	13.5311	98.05 %

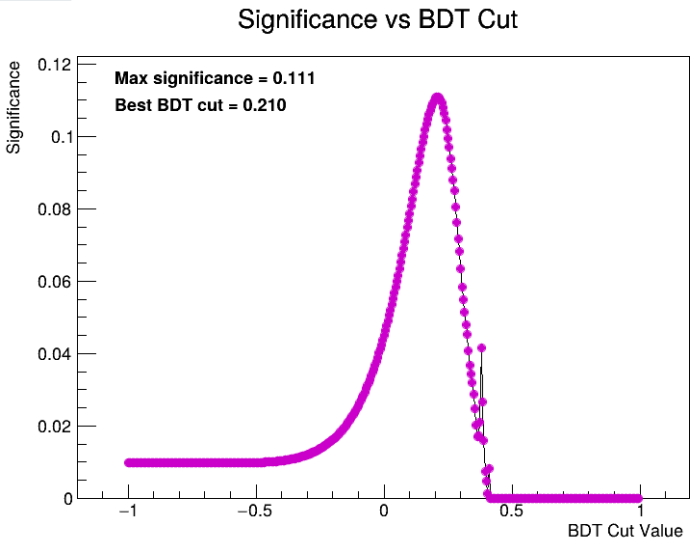
Original Significance : 0.0049
Pre-cut Significance : 0.0098

最適カット値の探索の結果

$$100 \times \frac{After\ BDT\ Cut}{Full\ events}$$

Process	File Name	Full events	After Pre-Cut	After BDT Cut	Efficiency
$WW \to \mu\nu q\bar{q}, \tau\nu q\bar{q}$	4fWWsemilep	4.19e+06	1.48417e+06	145.002	0.0035%
$WW \to e\nu q\bar{q}$	4fWsl	2.288e+06	92479.9	20.3775	0.0009%
$ZZ \to q\bar{q}l^+l^-/q\bar{q}\nu\bar{\nu}$	4fZZsemilep	311961	129121	782.821	0.2509%
$\nu\bar{\nu}Z \to \nu\bar{\nu}q\bar{q}$	4fZnusemilep	235599	95223	490.797	0.2083%
$WW/ZZ \to l\nu q\bar{q}$	4fmixlep	850192	36046.9	10.0229	0.0118%
$ZH \to \nu\bar{\nu}\tau^+\tau^-$	htautau	4317.87	347.215	0.125222	0.0029%
$ZH \to \nu\bar{\nu}WW^*$	hww	14706.6	7236.74	17.5307	0.1192%
$ZH \to \nu\bar{\nu}ZZ^*$	hzz	1809.97	971.146	6.75007	0.3729%
$ZH \to \nu\bar{\nu}b\bar{b}$	nnbb	40287	37660.7	0.604305	0.0015%
$ZH \to \nu\bar{\nu}c\bar{c}$	nncc	1999	1952.23	22.9685	1.1490%
$ZH \to \nu\bar{\nu}g\bar{g}$	nngg	5650	5587.23	160.968	2.8490%
	Total Background (B)	7.94433e+6	1.8908e+06	1657.97	0.0209%
$ZH \to \nu\bar{\nu}s\bar{s}$	Signal (nnss) (S)	13.8	13.5311	4.5253	32.792%

Best BDT cut = 0.21



BDT cut Significance :0.110986

BDTカットの確認

My result
Significance= 0.110986

Seino's result
Significance=0.118

Process	File Name	Full events	After Pre-Cut	After BDT Cut	Efficiency	Full events	After BDT Cut	Efficiency
$WW \to \mu\nu q\bar{q}, \tau\nu q\bar{q}$	4fWWsemilep	4.19e+06	1.48417e+06	145.002	0.0035%	4.19e+6	245.7	0.0059%
$WW \to e\nu q\bar{q}$	4fWsl	2.288e+06	92479.9	20.3775	0.0009%			
$ZZ \to q\bar{q}l^+l^-/q\bar{q}\nu\bar{\nu}$	4fZZsemilep	311961	129121	782.821	0.2509%	311961	423.6	0.136%
$\nu\bar{\nu}Z \to \nu\bar{\nu}q\bar{q}$	4fZnusemilep	235599	95223	490.797	0.2083%	235599	264.2	0.112%
$WW/ZZ \to l\nu q\bar{q}$	4fmixlep	850192	36046.9	10.0229	0.0118%	850430	4.6	0.0005%
$ZH \to \nu\bar{\nu}\tau^+\tau^-$	htautau	4317.87	347.215	0.125222	0.0029%	4318	0.0	0%
$ZH \to \nu\bar{\nu}WW^*$	hww	14706.6	7236.74	17.5307	0.1192%	14707	10.4	0.071%
$ZH \to \nu\bar{\nu}ZZ^*$	hzz	1809.97	971.146	6.75007	0.3729%	1816	3.9	0.215%
$ZH \to \nu\bar{\nu}b\bar{b}$	nnbb	40287	37660.7	0.604305	0.0015%	40287	0.0	0%
$ZH \to \nu\bar{\nu}c\bar{c}$	nncc	1999	1952.23	22.9685	1.1490%	1999	8.1	0.405%
$ZH \to \nu\bar{\nu}g\bar{g}$	nngg	5650	5587.23	160.968	2.8490%	5650	113.9	2.016%
	Total Background (B)	7.94433e+6	1.8908e+06	1657.97	0.0209%	5.65e+6	1074	0.019%
$ZH \to \nu\bar{\nu}s\bar{s}$	Signal (nnss) (S)	13.8	13.5311	4.5253	32.792%	13.8	3.9	28.261%

末原先生のコメント
もうちょっと救えるようになるといいな

残存背景事象の確認

目的・状況:

- BDTの最適カット（BDT > 0.21）を適用した結果、シグナル領域に $WW \to l\nu qq$ が残存していることが判明。
- ニュートリノチャンネル（レプトン0個のトポロジー）において、本来 isolepton=0 のPrecutで弾かれるべきWからのレプトンがなぜ漏れているのか、その原因究明が必要

Process	File Name	Original	After Pre-Cut	After BDT Cut	Efficiency
$WW \to \mu\nu q\bar{q}, \tau\nu q\bar{q}$	4fWWsemilep	4.19e+06	1.48417e+06	145.002	0.0035%
$WW \to e\nu q\bar{q}$	4fWsl	2.288e+06	92479.9	20.3775	0.0009%
$ZH \to \nu\bar{\nu}s\bar{s}$	Signal (nnss)	13.8	13.5311	4.5253	32.792%

バックグラウンド残存要因の切り分け

- **現状の課題：**
  - BDTの最適カット適用後も、$WW \to l\nu qq$ がシグナル領域に残存している。
  - 本来であれば、Wボソン由来のレプトンはpre-cutの isolepton カットで識別・除外（Veto）されるはずである。
- **問い：**
  - tauのハドロニック崩壊が原因で、tauがレプトンとして認識されずアルゴリズムをすり抜けているのか？
  - tauは関係なく、単に isolepton finder の性能限界で電子・ミューオン（e, μ）を見逃しているだけなのか？
- **本研究のアプローチ：**
  - MC Truth（モンテカルロシミュレーションの真値）情報を抽出し、事象内の真のタウの個数と崩壊モードを追跡。残存の根本原因を特定する。

残存事象における真のタウの有無の確認

- **目的:**
  - BDTカット通過後も残存している $WW \to \ell\nu qq$事象の内訳が、真のタウを含む事象（$\tau\nu qq$）と含まない事象（$\mu\nu qq$ 等）のどちらに偏っているかを定量的に評価する。
- **手法:**
  - MC Truth情報（n_true_tau）を用い、プレカットおよびBDTカット（>0.21）を通過した事象内における真のタウの個数分布をプロットした。

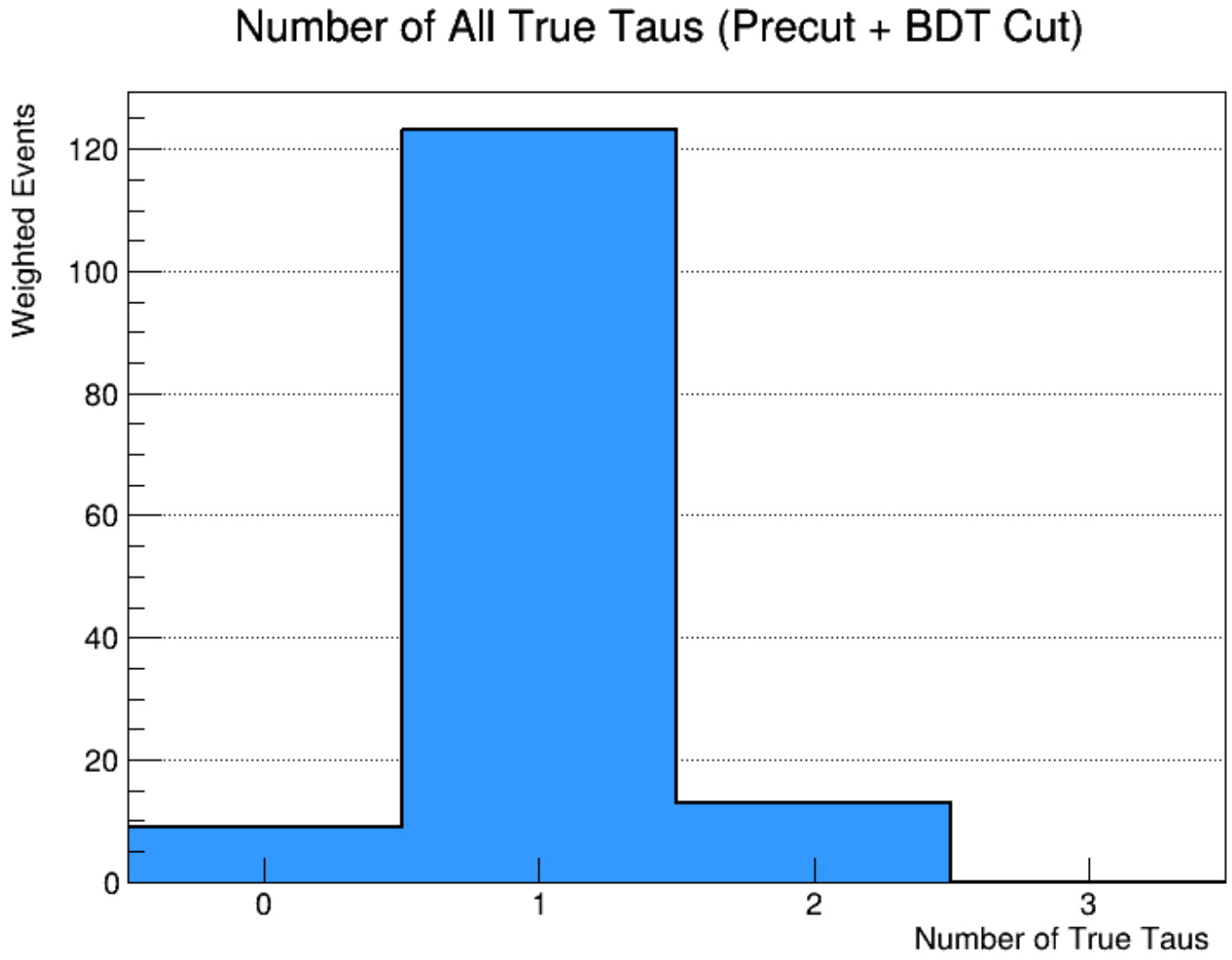
Process	File Name	After BDT Cut
$WW \to \mu\nu q\bar{q}, \tau\nu q\bar{q}$	4fWWsemilep	145.002

残存するものが$\mu\nu q\bar{q}$か$\tau\nu q\bar{q}$なのかで、偏りがあるかどうか確かめる

タウの個数分布と残存事象の偏り

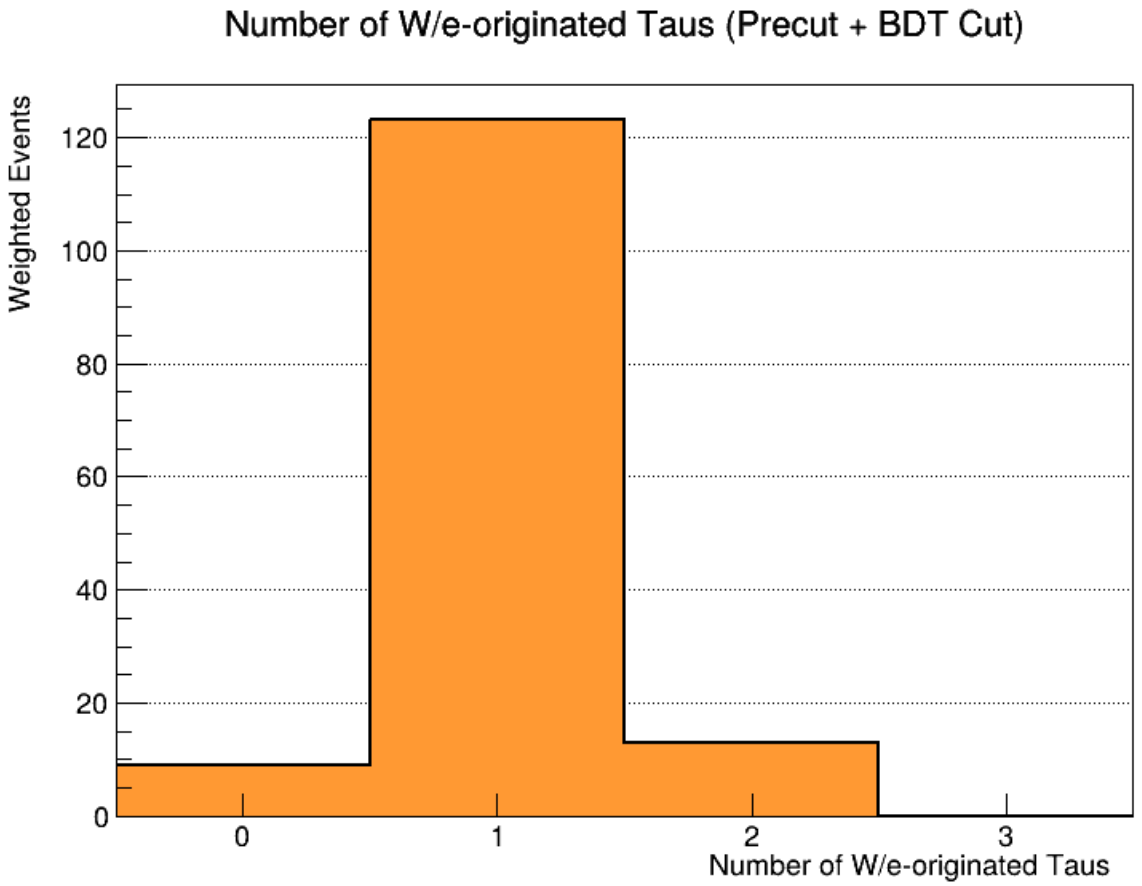
- **結果:**
  - タウを含まない事象（0-tau, 直接のミューオン漏れ）：約9事象（約6%）
  - タウを含む事象: 約136事象（約94%）
- **結論:**
  isolepton カットによる直接のミューオン漏れ（$\mu\nu qq$）は少なく、残存事象の主要因はタウを含む事象（$\tau\nu qq$）であることが示された

All True Taus (Precut + BDT > 0.21)
Total Surviving Events : 145.002
Events with 0 taus : 8.88303
Events with 1 tau : 123.144
Events with 2 taus : 12.9762



タウの起源(W/e由来)の絞り込みとDouble Countの排除

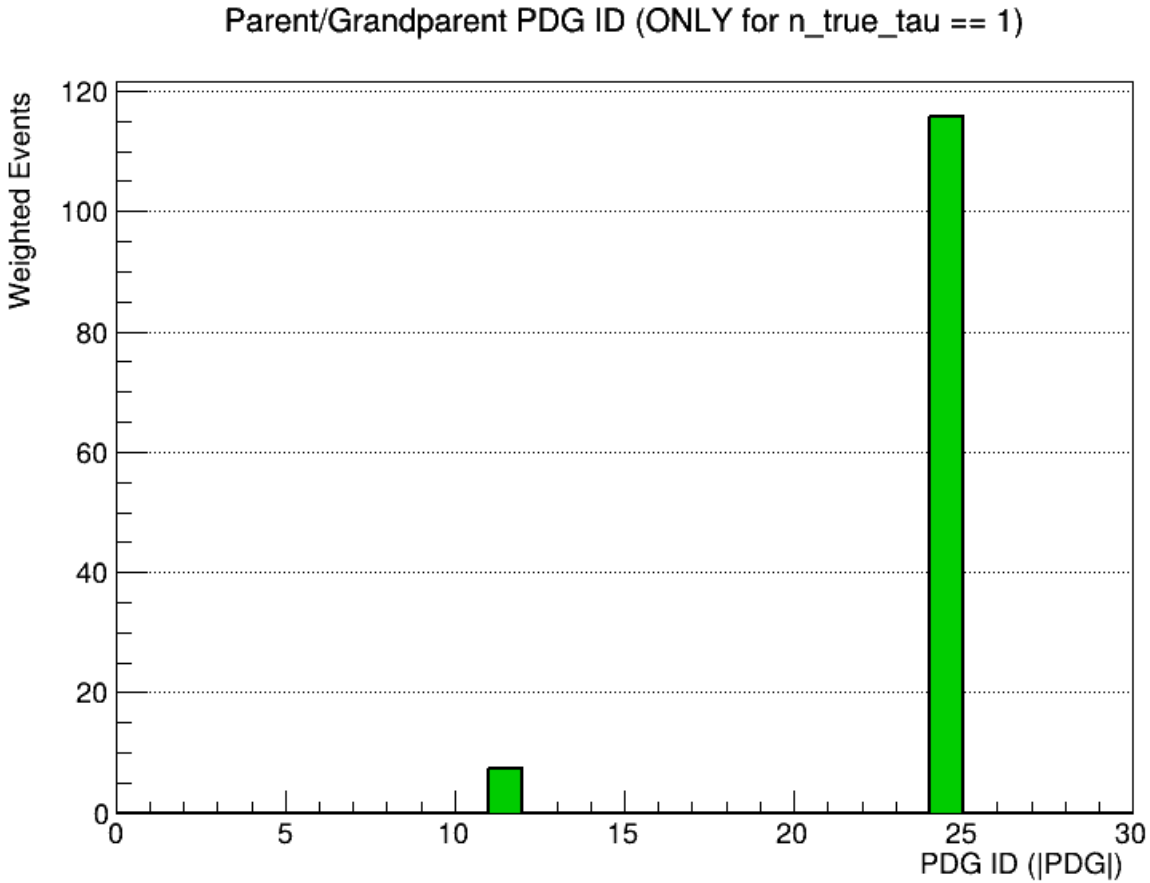
- **目的:**
  - MCシミュレータの都合上、生成過程で中間状態のWボソンが省略され、電子（e）から直接タウが生成されたように記録される事象が存在する。
  - 終状態輻射（FSR）等により1つのタウが複数回（tau2）記録されるDouble Count現象を排除し、純粋なタウの起源を評価するため、親粒子条件による絞り込みを行った。
- **結果:**
  - タウの親粒子をWボソンまたは電子に限定した結果、事象ごとの追跡により tau2 から約0.09事象が減少し（$2 \to 1$への移動）、同時に tau1 からも約0.09事象が減少（$1 \to 0$への移動）していることが判明した。
- **結論:**
  - 見かけ上 tau1 の総数（123.1事象）は変動していないが、内部ではクォーク由来タウの除外（$1 \to 0$）とタウとクォーク由来タウの混在事象の解消（$2 \to 1$）が同数（約0.087事象）だけ発生し、相殺されていた。
  - 残存する tau2（12.8事象）は純粋なダブルカウント（FSR等）であると断定できる



W/e-originated Taus (Precut + BDT > 0.21)
Events with 0 W/e-taus : 8.97012
Events with 1 W/e-tau : 123.144
Events with 2 W/e-taus : 12.8891

tau1事象における親粒子の内訳

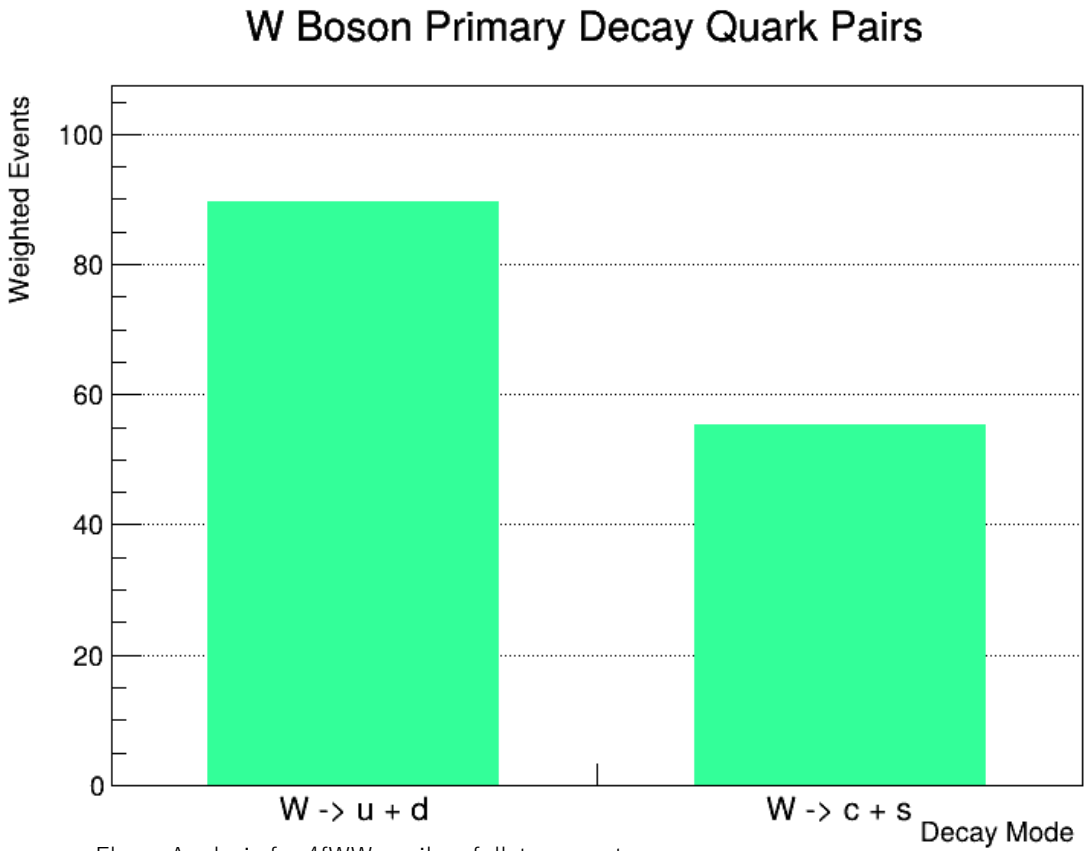
- **手法:**
  - ダブルカウントの影響がない n_true_tau =1 の事象（123.143事象）を対象に、タウの親粒子（Parent/Grandparent PDG ID）を追跡・分類した。
- **結果:**
  - 残存タウの約94%（115.7事象）はWボソン（PDG: 24）由来、約6%（7.3事象）は電子（PDG: 11）由来であった。
  - しかし、約0.087事象が、Wボソンでも電子でもZボソンでもないその他の親（Other）から生成されたtauであることが判明した。これが前スライドにおいて $1 \rightarrow 0$へと移動した事象の正体である。



Breakdown of Tau Origins (ONLY for n_true_tau == 1)
Total 1-Tau Events : 123.143
Events from W (PDG 24) : 115.741
Events from e (PDG 11) : 7.31543
Events from Z (PDG 23) : 0
Events from other : 0.0861842

Wボソンのハドロン崩壊先（クォークペア）の特定

- **目的:**
  - $WW \to \ell\nu qq$事象において、クォーク側（qq）が具体的にどのフレーバーのペアに崩壊しているかを特定する。
- **手法:**
  - ハードプロセス直後のクォークを抽出し、W ボソンからの主要な崩壊モードを分類した。
- **結果と考察:**
  - グラフが示す通り、$W \to ud$ と $W \to cs$の2つの主要な崩壊モードがあることが確認された。

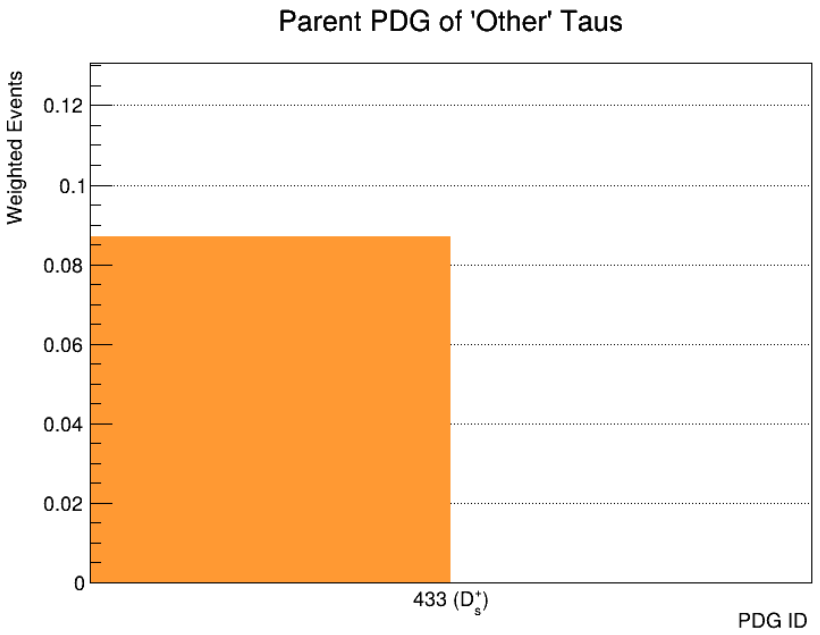



```
=== Flavor Analysis for 4fWWsemilep_full_tmva.root ===
[1] Primary W Boson Decay Quark Pairs
  Total Events              : 145.002
  W -> c + s (Primary)      : 55.4753
  W -> u + d (Primary)      : 89.5269
------------------------------------------------
[Extra] Parton Shower Effects (Subset of above)
  Events with b quark (g->bb) : 0.0870885
  -> Note: This 'b' originates from gluon splitting in the shower, not W decay.-------
```

未知のタウ起源の特定（クォークからの二次生成）

- **目的:**
  - スライド13で確認されたW/e/Z由来ではないタウ（Other: 約0.086事象）の真の起源を特定する。
- **手法:**
  - 当該タウの親・祖父母粒子のPDG IDを精査し、崩壊の起点となった重いハドロン（ID最大値）を抽出した。
- **結果と結論（1 →0事象の正体）:**
  - D_s^+中間子（PDG ID: 433）：親粒子の追跡から、チャームクォーク（c）と反ストレンジクォーク（$\bar{s}$）からなるD_s^+中間子が特定された。
  - **崩壊メカニズム:** 前スライドで確認した$W \to cs$崩壊によって生成されたcクォークがD_s^+中間子となり、そこから二次的にタウが崩壊生成された事象であると判明した。
  - **事象の全体像:** レプトン側がミューオン（μ）等に崩壊して isolepton カットをすり抜け、同時にクォーク側からタウが出たケースであり、これが$1 \to 0$へ移行した事象の正体であることが証明された。

※レプトン普遍性により、レプトン側がτの場合（$2 \to 1$へ移動する事象）でも、クォーク側では全く同じ現象が起きていると結論付けられる



Parent PDGs of 'Other' Taus (Not W, e, Z)
PDG: 433 (Weight: 0.0870885)　（D_s^{*+} 中間子）

タウの親粒子特定手法と $W \to cs$崩壊の物理的根拠

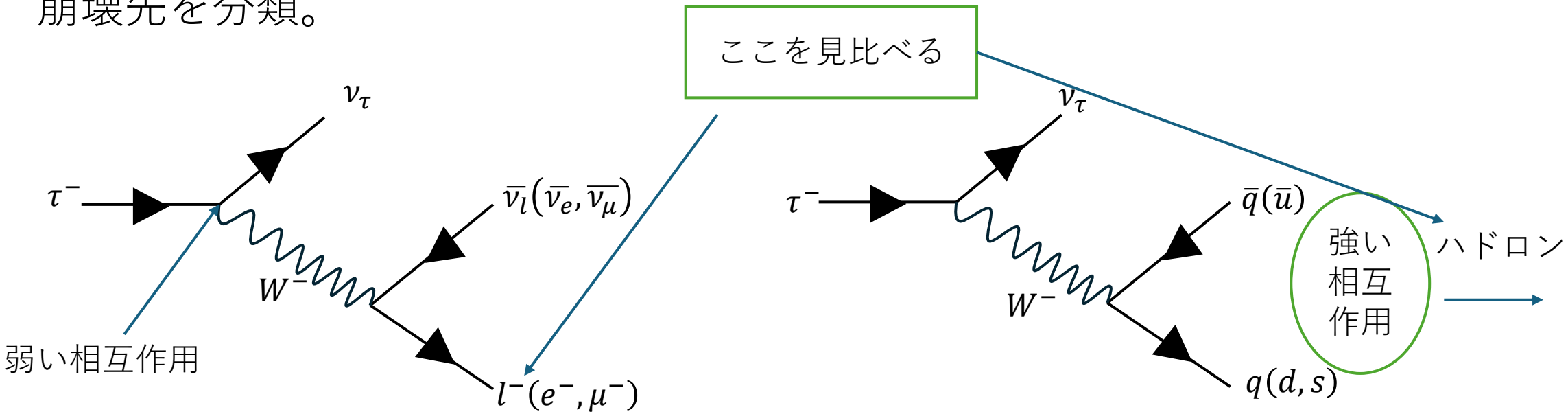
1. なぜPDG IDの最大値からクォーク由来だと分かるのか？
   - PDG IDの命名規則： 中間子のIDは、内部に含まれるクォークが重いほど大きな数字が割り当てられる（例：ストレンジ系は300番台、チャーム系は400番台）。
   - 崩壊連鎖の逆探知： ハドロン崩壊は重い状態から軽い状態へと連鎖する。タウの直前の履歴（親・祖父母）からIDの最大値を抽出することで、寿命の短い中間状態のノイズを排除し、崩壊の起点となった大元の重いハドロンを特定できる。
2. なぜ$D_s^{*+}(433)$の発見が$W \to cs$の証明になるのか？
   - 粒子の構成要素： D_s^{*+}中間子は、チャームクォーク（c）と反ストレンジクォーク（$\bar{s}$）で構成されている。
   - Wボソンの崩壊則（電荷保存則）： Wボソン（電荷 ± 1）がハドロン崩壊する際、CKM行列の制約により主要な崩壊先は udペアまたは csペアに限られる。

結論：
事象内に cクォークを含む D_s^{*+}が生成されたという事実は、大元のWボソンが csペアに崩壊したことの証拠となる。

（補足）なぜ $W \to ud$からはタウが生じないか:
u, d クォークから形成されるハドロン（$\pi^{\pm}$ 等）は質量が軽すぎ（~ 0.14 GeV）、タウ（~ 1.78 GeV）を生成するためのエネルギー保存則を満たさない。

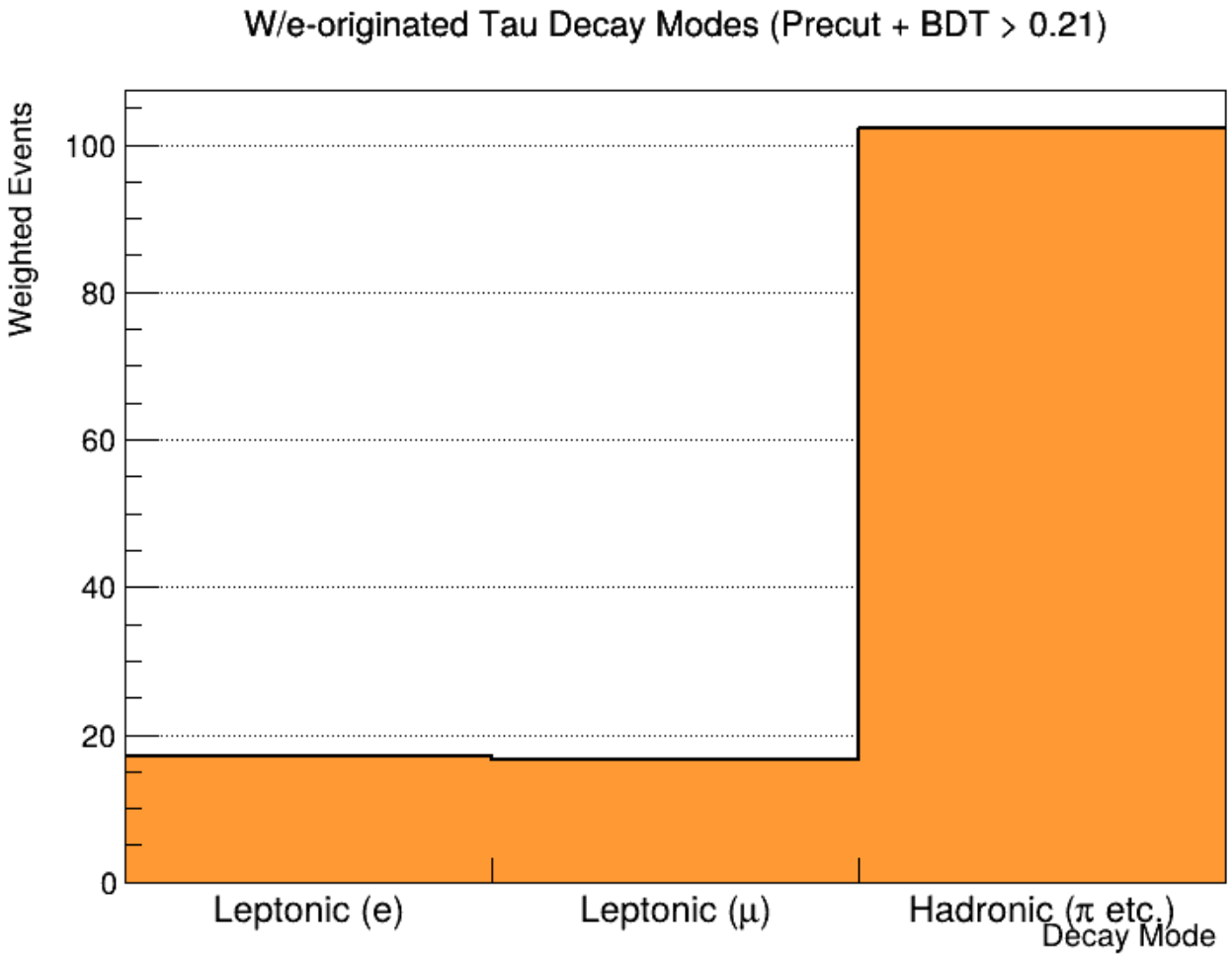
タウの崩壊モード特定

- **目的:** 残存したタウがレプトン（e, μ）とハドロン（π 等）のどちらに崩壊したかを特定する。ハドロニック崩壊であれば、既存のisoleptonで弾けないのはアルゴリズムの仕様上不可避であると証明できる。
- **解析条件と理由:**
  - **条件:** MC Truth情報からタウが正確に1つだけ存在する事象（n_true_tau == 1）のみを抽出して解析。
  - **理由:** $WW \to l\nu qq$プロセスにおいて、Wボソンから直接生成される本物のtauは原理的に1つのみであるため。2つ以上のタウが記録されている事象（二重カウントや、ジェット内での二次生成等のノイズ）を排除し、純粋なWボソン崩壊由来のタウの振る舞いのみを評価する。
- **手法:** 上記条件を満たす事象のタウ娘粒子情報（true_tau_dau_pdg）を追跡し、崩壊先を分類。



タウの崩壊モード特定

- **結果:**
  - レプトニック崩壊（$\rightarrow e, \mu$）：約34事象（約25%）
  - ハドロニック崩壊（$\rightarrow \pi$ 等）：約102事象（約75%）
- **結論:** 全体の約7割を占めるハドロニックタウは、既存のisoleptonでは原理的に除外が難しく、専用のTauFinderによるタウVetoが必須であるという根拠を得た。



=== Breakdown of W/e-originated Taus in Surviving Events (Precut + BDT > 0.21) ===
Leptonic decay (-> e) : 17.1564
Leptonic decay (-> mu) : 16.6339
Hadronic decay (-> pi) : 102.242
Total : 136.032

ハドロニックタウが isolepton をすり抜ける理由

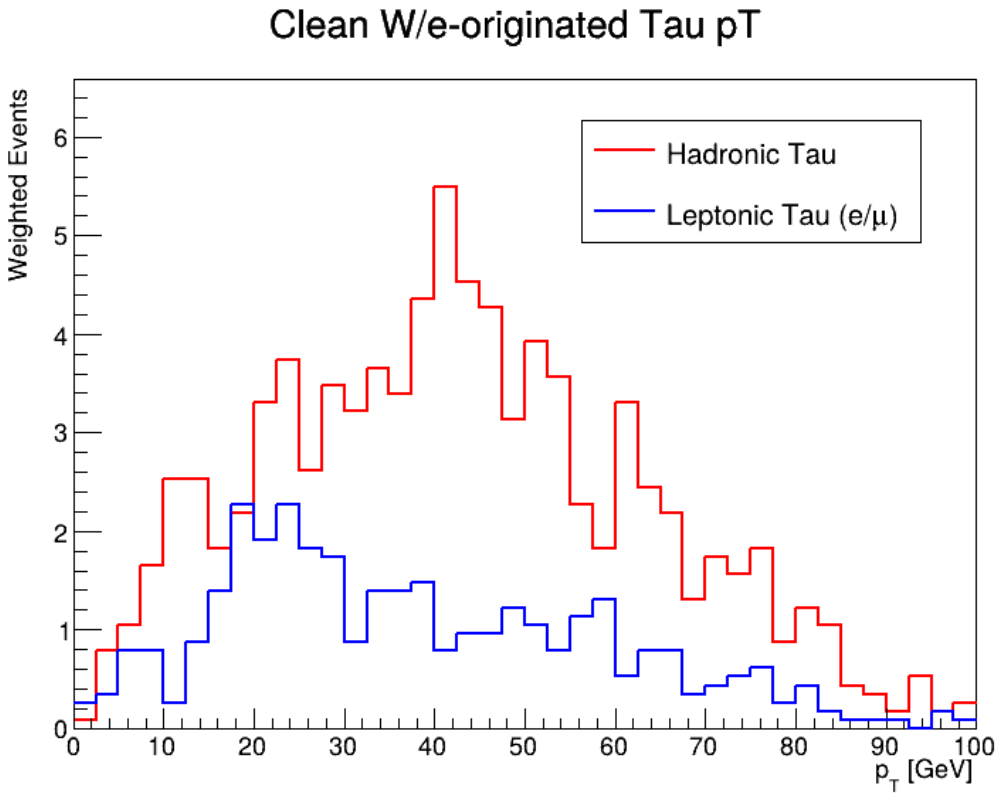
- **isolepton finder の探索対象:**
  - このアルゴリズムは、電磁カロリメータ（Ecal）での特有のシャワー形状を持つ電子（e）や、ハドロンカロリメータ・ミューオン検出器を貫通するミューオン（μ）の信号を探すように設計・学習（MVA）されている。
- **ハドロニック崩壊の物理的性質:**
  - タウの約66%はパイ中間子（$\pi^{\pm}, \pi^{0}$）などのハドロンに崩壊する。
  - これらは強い相互作用を起こすため、検出器内では電子やミューオンではなく細くて粒子数の少ないクォークジェット（ミニジェット）と全く同じ振る舞いとして記録される。
- **結論：**
  - ハドロニックタウは物理現象としてジェットそのものであるため、電子・ミューオン専用のカッター（isolepton）では原理的に識別不可能である。
  - したがって、通常のクォークジェットと、タウ由来の細いジェットを見分けるための専用モジュール（TauFinder）の導入と最適化が望ましい。

運動学的要因の解明

- **目的:** そもそもなぜレプトン（直接の μやタウ由来の e, μ）が isolepton で見逃されたのか。アルゴリズムの運動量閾値（5 GeV）に対して、エネルギーが低すぎたのかを特定する。
- **手法:** レプトニック崩壊のtauとハドロニック崩壊のtauの、横運動量（p_T）の分布をプロットし、閾値との関係を評価する。
- p_Tを用いる理由: ビームパイプ方向へ逃げる粒子の影響を受けず、アルゴリズムのカット閾値への影響を最も正確に評価できるため。

運動学的要因の解明（結果①：tauのp_T分布）

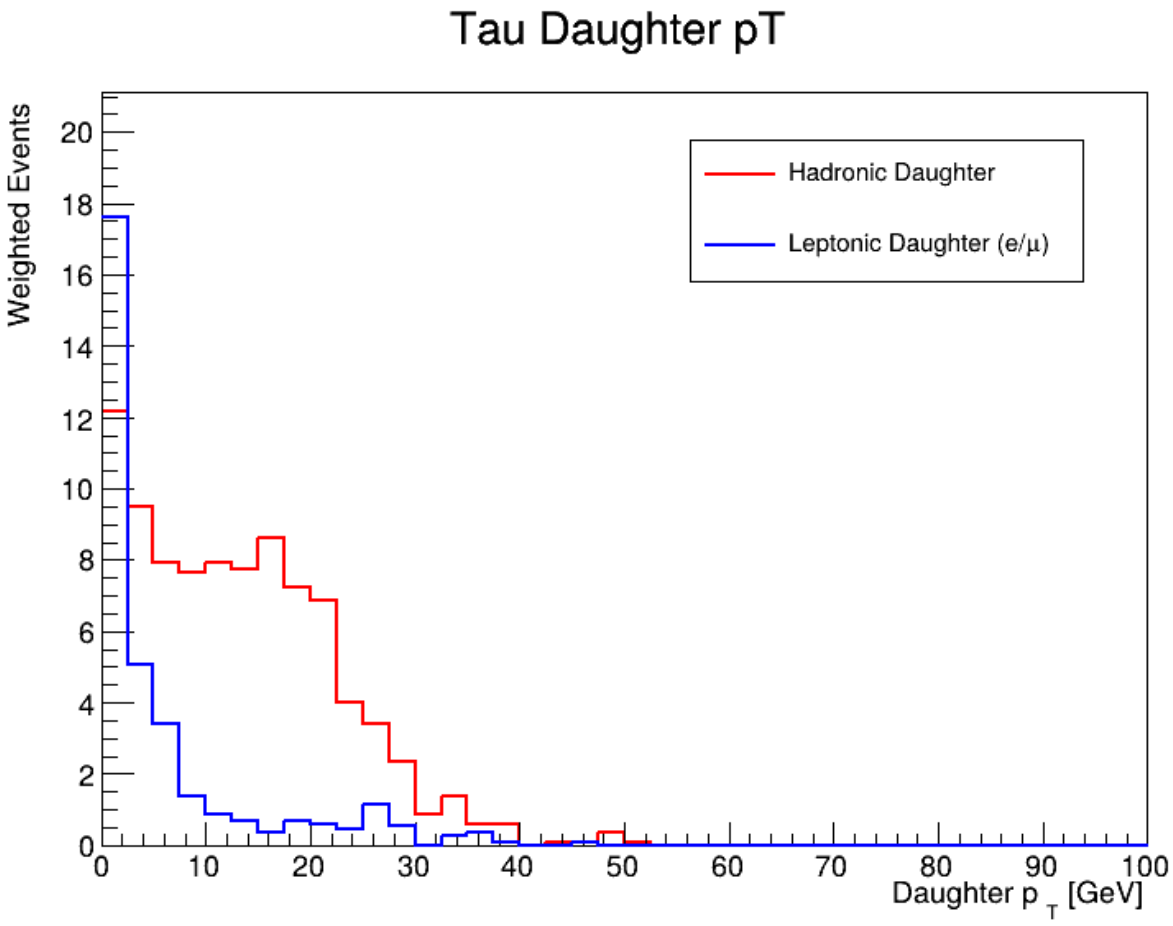
- **結果:** タウハドロニック崩壊およびレプトニックの p_T（横運動量）の大部分が10～60 GeVの範囲に分布している。
- **システム設定との比較:** isolepton finder（seinoproc_tau.xml）におけるレプトン判定の運動量閾値は $P > 5\text{GeV}$ に設定されている。
- **小結論:** 大半の事象はこの5 GeVの足切りラインを超えており、エネルギー不足がすり抜けの主原因ではないことがわかった。



末原先生より、
tauが崩壊した際のニュートリノにより運動量が大きく持ってかれてしまっている可能性があるので、tau自体の横運動量を見るよりかはtauが崩壊して出したレプトンかハドロンの運動量を新たに見た方がいいかも

運動学的要因の解明（結果②：娘粒子のp_T分布）

- **末原先生からのご指摘に基づく再評価**
  - タウ崩壊時に放出されるニュートリノ（ν_τ等）により、観測される粒子（娘粒子）の運動量はタウ自身の運動量から大きく欠損する可能性がある。
  - そこで、タウ自身のp_Tではなく、実際に検出器が捉えるのに近い娘粒子のp_Tの分布を再評価した。
- **結果**
  - 前スライドのタウ自身のp_T分布と比較し、娘粒子のp_T分布は低エネルギー側へシフトしていることが確認された。
  - 特にレプトニック崩壊（青線）において、ニュートリノ2つ分のエネルギー欠損により、事象の大半が $p_T < 10$ GeVの領域に集中している。
- **結論**
  - エネルギー欠損の結果、少くない事象がisolepton の判定閾値である $P > 5$ GeVを下回ってしまい、レプトンとして識別されずすり抜けていることが判明した。



運動学的要因の解明

- **結論：isoleptonカットすり抜けの主因は娘粒子の低運動量化にある**
  - タウ自身の p_T はisoleptonの判定閾値（$P > 5$ GeV）を十分に満たしていた。
  - しかし、タウ崩壊時のニュートリノ放出に伴うエネルギー欠損により、実際に検出器が捉える娘粒子の運動量は著しく低下し、5 GeV 以下の領域に多数の事象が落ち込んでしまうことが判明した。
- **物理的背景：**
  - レプトニック崩壊においてはニュートリノが2つ放出されるため、このエネルギー欠損の影響が特に顕著に表れ、レプトン（e, μ）として識別されるべき粒子が判定閾値を下回ってしまう。
- **次の方針（TauFinderの最適化へ）：**
  - この低運動量化したタウ崩壊生成物を正しく拾い上げるためには、単純にisoleptonの閾値を下げるだけでは他の背景事象（QCDノイズ等）の増大を招く危険がある。
  - したがって、低運動量領域においてもタウ特有のジェット構造（トラックの孤立性や広がり等）を多角的に評価できる **TauFinder アルゴリズムのパラメータ最適化が不可欠**となる。

未調整TauFinderのベースライン評価（手法と定義）

- **目的:**
  - TauFinderのパラメータチューニングを開始する前に、デフォルト設定における検出能力と誤検知の割合を測定する。
- **手法（Particle-Level Strict Matching）:**
  - 事象単位の曖昧な判定ではなく、TauFinderが捉えた飛跡（トラック）1本1本が、真のタウ由来であるかをシミュレータのMC Truth（RecoMCTruthLink）を用いて直接照合し、評価を行う。
  - **プレカットや事象の重み付け（Weight）は一切行わず、生の生成イベント数のみで評価**する。
- **評価指標の定義:**
  - **Efficiency（真のタウを見つける確率）:**
    - MC上で生成された全タウ（クローン重複排除済） に対し、TauFinderが正しく見つけ出してマッチした個数の割合。
  - **Fake Rate（タウと誤認する確率）:**
    - TauFinderがtauだと判定した全オブジェクト のうち、実際にはtau以外の粒子（クォークジェット等）であった個数の割合。

現状のTauFinder設定と評価条件（ベースライン）

- **評価対象のアルゴリズム（TaJetFinderProcessor）**
  - 入力コレクション： PandoraPFOs（全再構成粒子）
- **主なデフォルトパラメータ設定値**
  - TauMass（想定されるタウの質量）: 2.0 GeV
  - MinimumTrackEnergy（トラックの最小エネルギー）: 1.0 GeV
  - MinimumJetEnergy（ジェットの最小エネルギー）: 5.0 GeV
  - ConeMaxCosAngle（探索コーンの最大角度）: 1.0
  - *※この設定をチューニングの起点（ベースライン）とする。*
- **評価条件（Particle-Level Strict Matching）**
  - **プレカットや事象の重み付けを一切行わず、**代表サンプル（17,200事象）に対して粒子レベルのマッチング評価を実行した。
  - 4fWWsemilep($WW \to \mu\nu q\bar{q}, \tau\nu q\bar{q}$)のlcioファイルにて評価を実行した

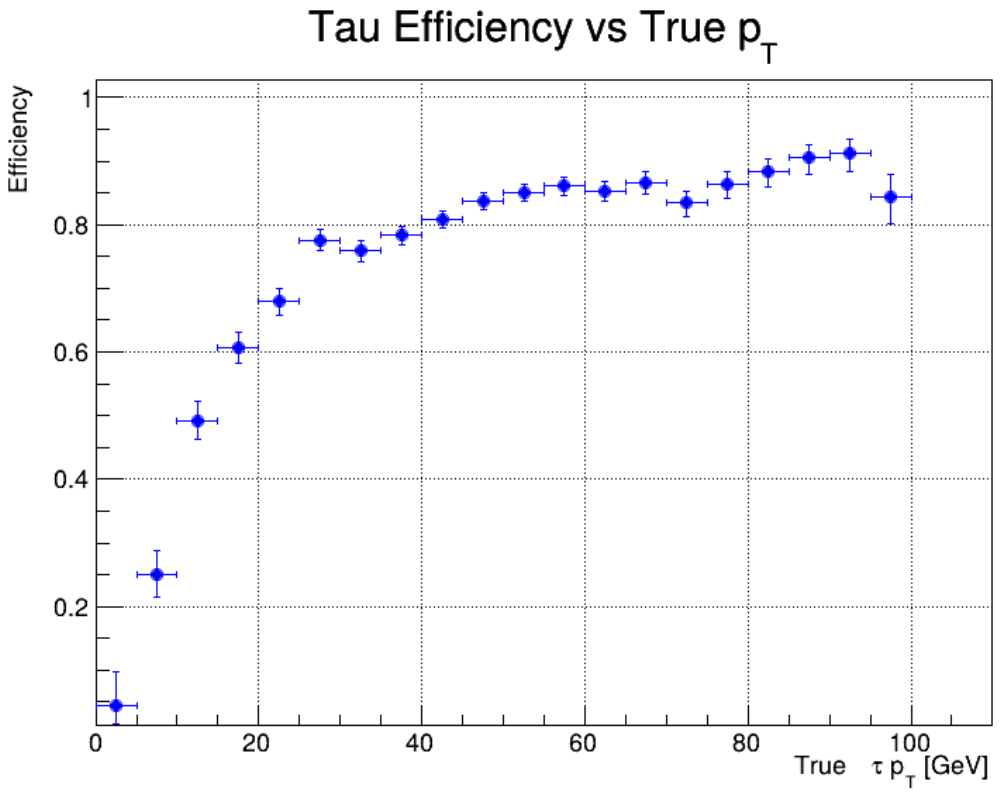
粒子レベルベースライン評価結果（全体像）

```
============================
TauFinder Particle-Level Performance (No Cuts, Raw Counts)
============================
 [Analyzed Events] : 17200 events
--------------------------------------------
 [Efficiency: 本物のタウを正しく捕まえた割合]
  - MC True Taus     : 9547 particles
  - Found & Matched  : 7490 particles
  => Efficiency      : 78.45 %
--------------------------------------------
 [Fake Rate: クォーク等をタウと誤認した割合]
  - Total Reco Taus  : 31938 particles
  - Fake (Not a Tau) : 25077 particles
  => Fake Rate       : 78.52 %
```

- **粒子レベル評価結果（全事象・カット無し）**
  - **Efficiency（正解率）：71.87 %**
    - （MC上の真のタウ 9,547個のうち、6,861個を正しく特定）
  - **Fake Rate（誤認率）：78.52 %**
    - （TauFinderが判定した31,938個のうち、25,077個がクォーク等）
- **全体像からの考察**
  - 見つけたオブジェクトの約8割が偽物（クォークジェット等の誤検知）であり、過検出の状態にある。
- **次ページ以降の目的**
  - Fake Rate の原因と、見逃してしまった 30% のタウの正体を探るため、**運動量（p_T）、質量、飛跡数の3つの観点からTauFinderの性能を解剖する。**

タウの運動量（p_T）と検出効率の関係

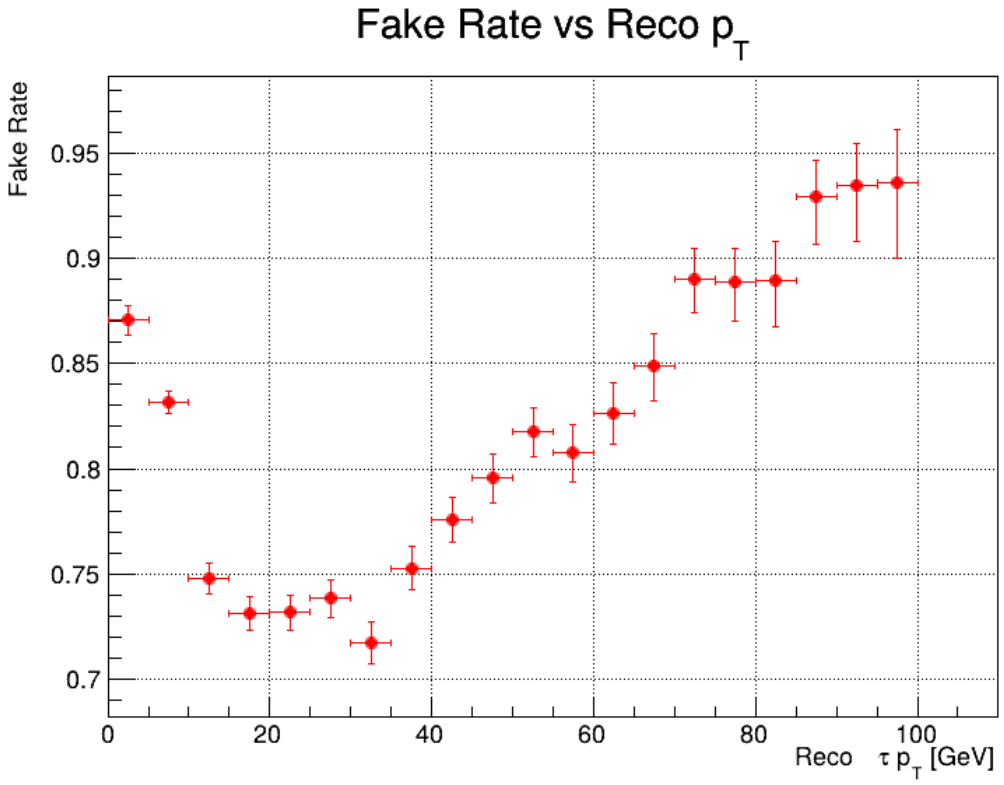
- **結果と考察**
  - **高エネルギー帯での高い性能:** p_Tが 30 GeV を超える領域では、Efficiencyは80%～90% 程度で安定しており、ある程度検出能力を持っている。
  - **低エネルギー帯での性能低下:** 一方で、p_T が 20 GeV を下回る低エネルギー領域では効率が低下し、数 GeV のtauはほとんど見逃されている。
- **チューニングへの示唆**
  - MinimumTrackEnergy などのエネルギー閾値を下げることで、低エネルギーのタウを拾い上げ、全体のEfficiencyを底上げできる可能性がある。
  - 下げ過ぎもよくない気がするけど。



横軸はシミュレータ上の真のタウの運動量（p_T）、縦軸はそのタウを正しく発見できた割合（Efficiency）

再構成運動量（p_T）と誤検知率（Fake Rate）の関係

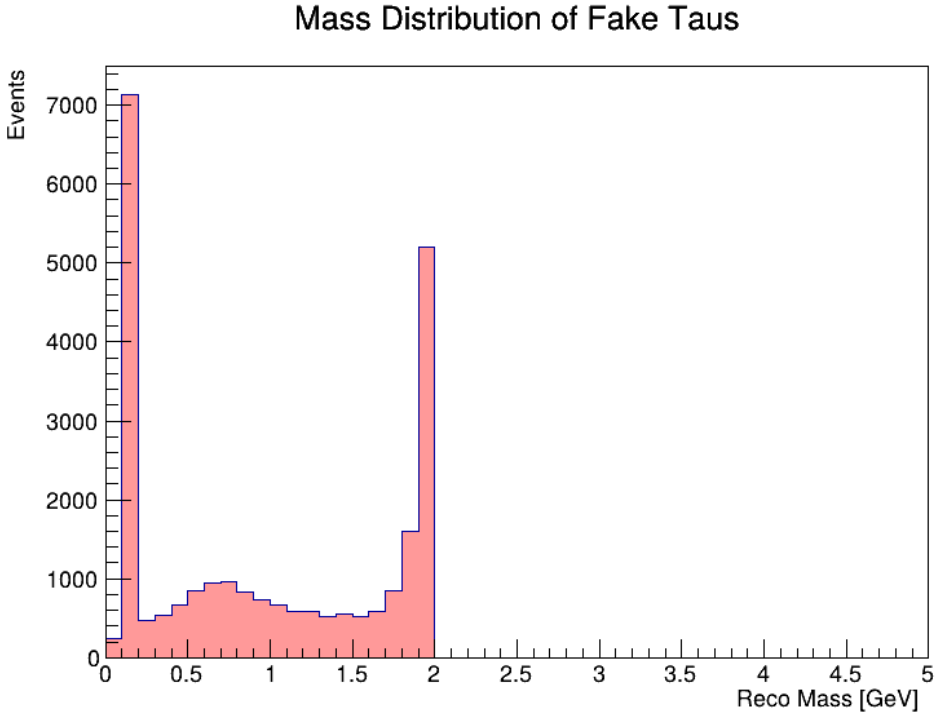
- **結果と考察**
  - **高エネルギー帯での著しい誤検知:** p_Tが高くなるほどFake Rateが悪化し、80 GeV 以上の領域では判定したオブジェクトの 90% 以上が偽物（クォークジェット等）である。
  - **中間エネルギー帯の谷:** $p_T \approx 30$ GeV付近でFake Rateが底を打つ（約72%）が、それでも絶対値としては高い。
- **チューニングへの示唆**
  - 高エネルギーのクォークジェットをtauと誤認するケースが多発している。ジェット特有の広がりを抑えるため、探索コーンサイズ（ConeMaxCosAngle）を狭める等の対策が必要である。



横軸はTauFinderがtauだと判定したオブジェクトの運動量（p_T）、縦軸は実はtauではなかった割合（Fake Rate）

誤検知（Fake）の正体①：質量分布

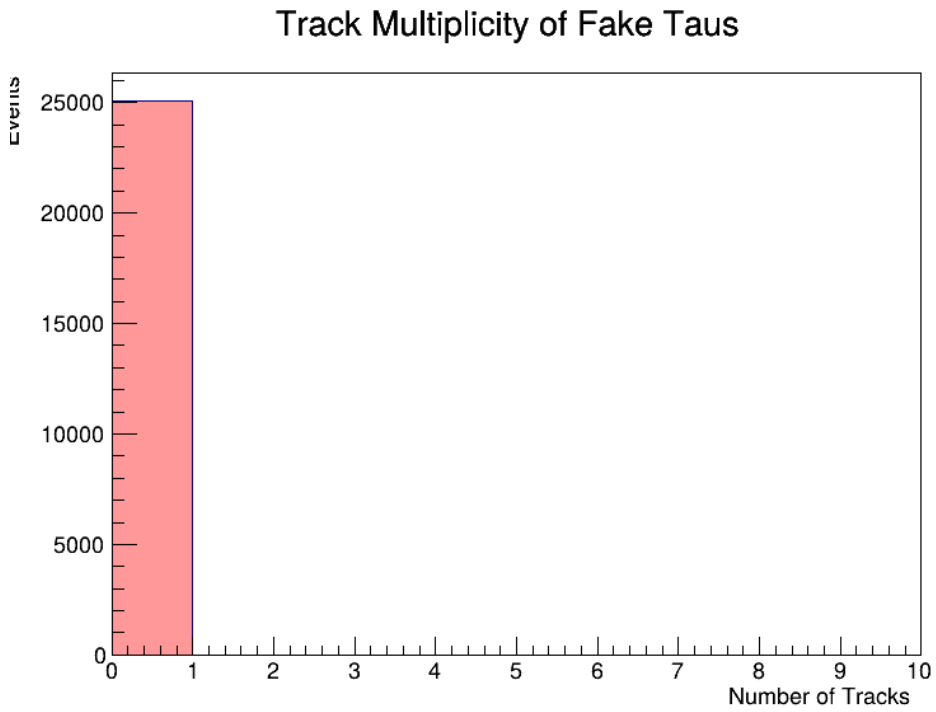
- **結果と考察**
  - 分布は 0 GeV 付近と、2.0 GeV 付近で2極化している
  - **2.0 GeV のピーク:** これはTauFinderのパラメータ TauMass の上限設定値（2.0 GeV）に張り付いていることを意味し、本来もっと重いクォークジェットを無理やりタウとして切り出している証拠である。
  - **0 GeV のピーク:** 質量がほぼゼロの単一の粒子（パイ中間子など）を誤検知している。
- **チューニングへの示唆**
  - 質量カットのウィンドウ（上限・下限）を適切に狭めることで、この両極端に存在する大量のFakeを効果的に削ぎ落とせる可能性がある。



TauFinderがタウと誤認した偽物（Fake）オブジェクトの、再構成質量（Mass）の分布

誤検知（Fake）の正体 ②：飛跡数（Track Multiplicity）

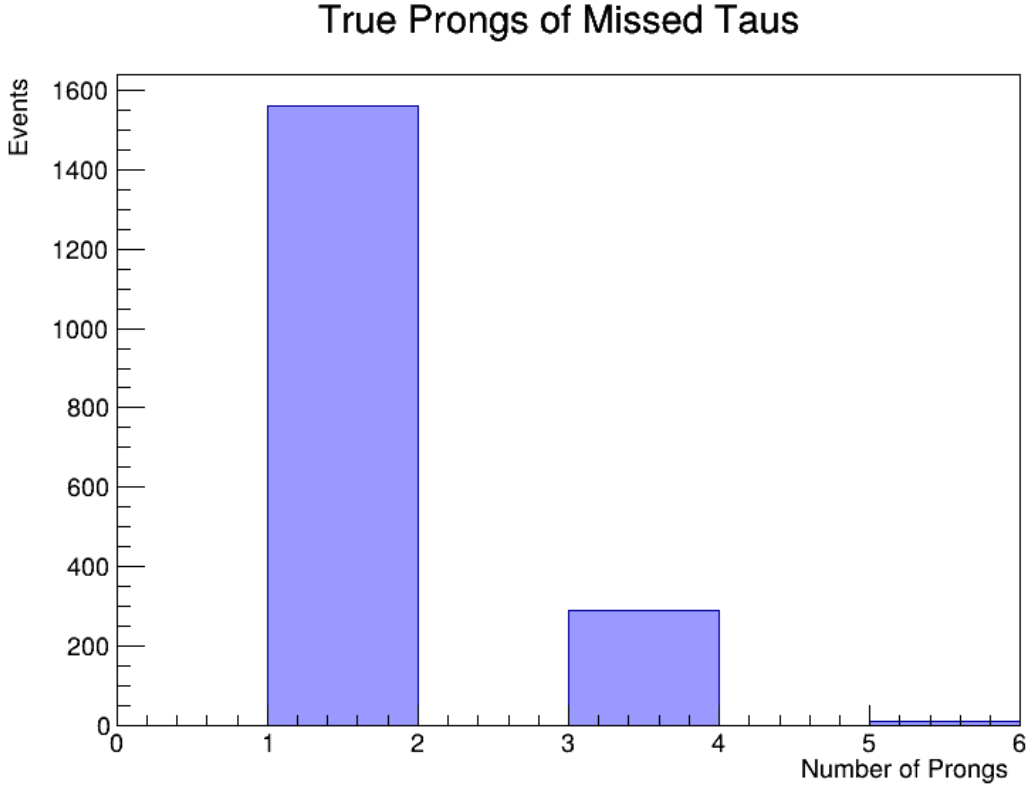
- **結果と考察（大発見）**
  - **Fakeのほぼ 100% が1-prong（飛跡1本）に集中している**
  - クォークジェット特有の多重飛跡（3本、5本…）はTauFinderによって正しく弾かれている。
- **チューニングへの最大の鍵**
  - Fake Rate（78%）を下げるための最大の急所は、**1-prongのFakeをいかに削減するか**にある。
  - IPCutsSigmaOneProng（1-prongに対するインパクトパラメータ要求）を厳しくする、あるいは特定の運動学条件を満たす単一飛跡を弾くことで、性能向上が見込める。



TauFinderがタウと誤認した偽物（Fake）オブジェクトに含まれる、飛跡（Track）の本数の分布

見逃されたタウ（False Negative）の正体解剖 ①：飛跡数

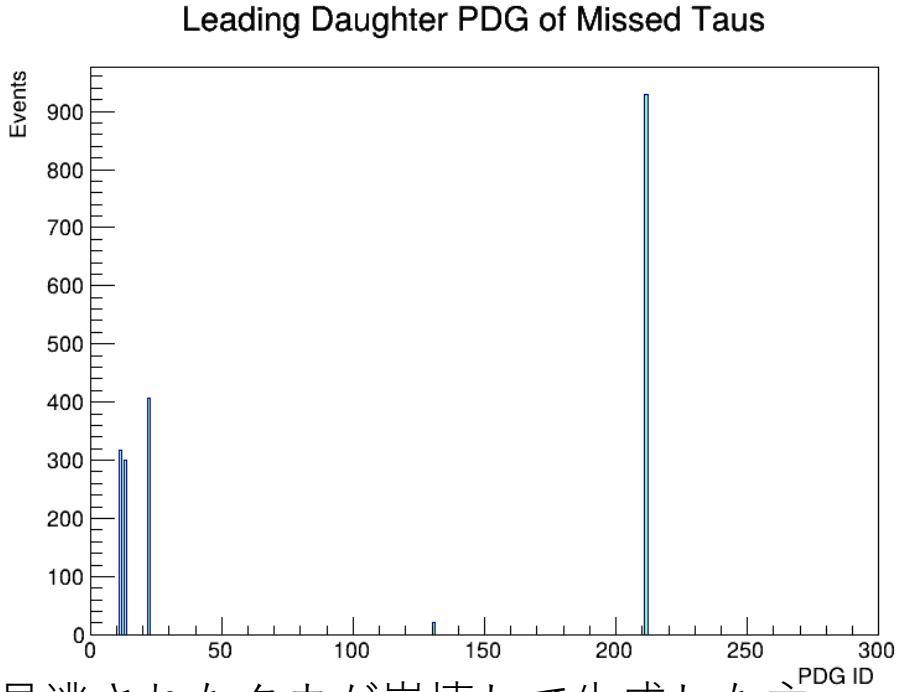
- **結果と考察**
  - **見逃しの大部分は 1-prong に集中している:**
    - 見逃されたタウの圧倒的多数（約1500事象以上）が飛跡1本の 1-prong 崩壊であった。
    - 3-prong の見逃し1-prongよりは少なく（約300事象）、TauFinderは多重飛跡のタウは比較的上手く捕らえられている。
- **小括**
  - 前述のFakeの大部分が1-prongであるという結果と合わせると、**TauFinderは1-prongのオブジェクトの真贋判定に苦戦している**ことがわかった。



TauFinderが見逃してしまった真のタウが、元々何本の飛跡（Prong）を持っていたかを

見逃されたタウ（False Negative）の正体解剖 ②：崩壊モード

- **結果**
- **レプトニック崩壊の見逃し:**
  - PDG 11 (e) : 317
  - PDG 13 (mu) : 301
  - 電子やミューオンに崩壊したタウが合計で約600事象見逃されている。
- **ハドロニック 1-prong 崩壊の取りこぼし:**
  - PDG 211 (pi+/-) : 929
    見逃しの最大要因は、荷電パイ中間子（$\pi^{\pm}$）に崩壊した 1-prong タウであり、900事象以上を取りこぼしている。
  - PDG 22 (gamma) : 406
    これは $\pi^{\pm}\pi^{0}$崩壊などで生じた $\pi^{0} \to \gamma\gamma$の光子がLeadingとして記録されたもの。
- **考察**
  - 低運動量（$p_T < 20$ GeV）の $\pi^{\pm}$が、飛跡のエネルギー閾値（MinimumTrackEnergy等）を満たせずに足切りされている可能性が高い。



見逃されたタウが崩壊して生成した主要な娘粒子（Leading Daughter）のPDG ID

[Missed Taus Leading Daughter: 見逃したタウの主な正体]
- PDG 11 (e (Electron)) : 317 particles
- PDG 13 (mu (Muon)) : 301 particles
- PDG 22 (gamma (Photon)) : 406 particles
- PDG 130 (K_L0 (Long Kaon)) : 21 particles
- PDG 211 (pi+/- (Charged Pion)) : 929 particles
- PDG 321 (K+/- (Charged Kaon)) : 70 particles
- PDG 2112 (中性子（Neutron）) : 4 particles
- PDG 2212 (陽子（Proton）) : 9 particles

性能解剖の総括とパラメータ最適化方針

- **ベースライン性能の総括（現在の課題）**
  - **Fake（過検出）の正体:** 高運動量帯に分布する **1-prong（単一飛跡）のクォークジェット**をtauとして誤認している。
  - **見逃しの正体:** 低運動量帯に分布する **1-prongの真のタウ（主に $\pi^{\pm}$崩壊）** を足切りしてしまっている。また、レプトニック崩壊も捉えられていない。
- **最大のトレードオフ（ジレンマ）の発見**
  - 偽物（Fake）も本物（True）も1-prongに集中しているため、安易に1-prongのカットを厳しくすると、本物のタウまでさらに見逃してしまう。
- **次の方針（TauFinderパラメータ最適化フェーズへ）**
  - 1-prong 領域において、Fakeを落としつつ、本物を救うための最適動作点（Sweet Spot）を以下のパラメータチューニングで探索する。
    - **インパクトパラメータの最適化:** IPCutsSigmaOneProng を調整し、タウ特有の有限の飛程（寿命）を活かしてクォーク由来のFakeを弾く。
    - **エネルギー閾値の緩和:** MinimumTrackEnergy 等を下げ、低 p_Tの $\pi^{\pm}$崩壊タウを回収してEfficiencyを底上げする。
    - **質量の適正化:** TauMass の上限設定（2.0 GeV）や質量ウィンドウを見直す。

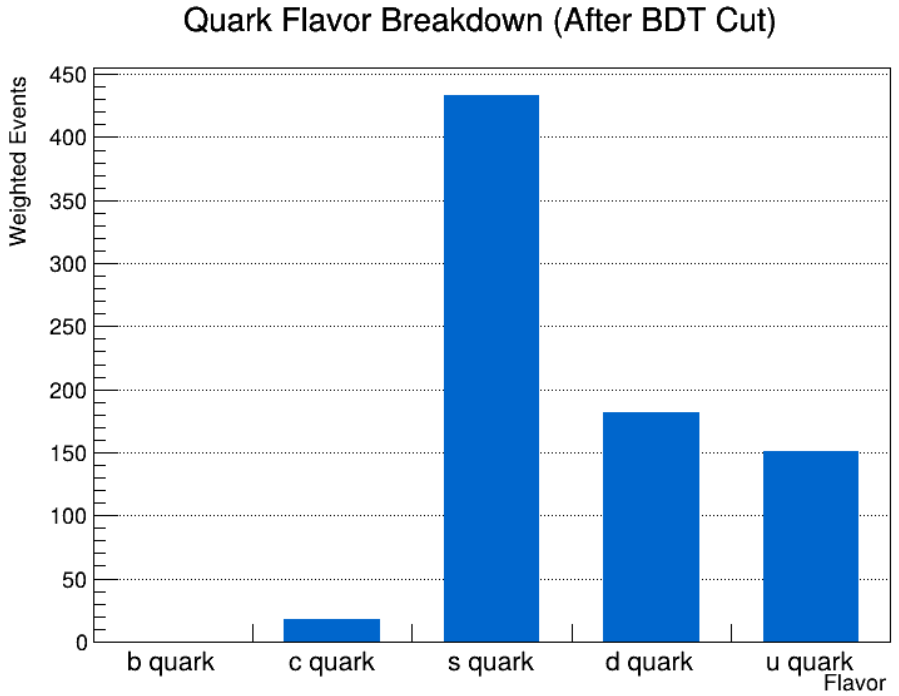
残存ZZ系背景事象のフレーバー解析

- **目的:**
  - Pre-cut+BDTカット（>0.21）を通過してシグナル（$H \to s\bar{s}$）領域に残存してしまう主要な背景事象（ZZ および $\nu\bar{\nu}Z$）が、なぜシグナルとして誤認されたのかを明らかにする。
- **手法:**
  - MC Truth情報（mc_quark_pdg）から、ハードプロセス直後のクォーク（PDG ID: 1～5）を直接抽出した。
  - ダブルカウントを防ぐため、事象内に含まれる最も重いクォークを基準として（$b \to c \to s \to d \to u$ の順）、事象のフレーバーを排他的に分類した。

Process	File Name	Full events	After Pre-Cut	After BDT Cut	Efficiency
$ZZ \to q\bar{q}l^+l^-/q\bar{q}\nu\bar{\nu}$	4fZZsemilep	311961	129121	782.821	0.2509%
$\nu\bar{\nu}Z \to \nu\bar{\nu}q\bar{q}$	4fZnusemilep	235599	95223	490.797	0.2083%
$ZH \to \nu\bar{\nu}s\bar{s}$	Signal (nnss) (S)	13.8	13.5311	4.5253	32.792%

$ZZ \to q\bar{q}\ell^+\ell^-/q\bar{q}\nu\bar{\nu}$事象におけるフレーバー内訳

- **重フレーバー（b, c）の排除は機能している**
  - BDTカット通過後の約782事象のうち、b クォーク由来は0.1事象、c クォーク由来も17事象まで減っている。
  - BDT（およびLCFIPlusのフレーバータグ）は、重いクォークの除去という役割を完璧に果たしている。
- **残存事象の過半数は $Z \to s\bar{s}$（約55%）**
  - 最も多く残存しているのは sクォーク由来（約433事象）である。
  - シグナル（$H \to s\bar{s}$）と全く同じフレーバーを持つため、原理的に分離が困難な背景事象である。
- **軽クォーク（u, d）の残存**
  - 次いで dクォーク（181事象）、u クォーク（150事象）が残存している。
  - 質量が軽く、s ジェットとトポロジーが酷似しているため、現在のフレーバータギングの性能限界により残存していると考えられる。



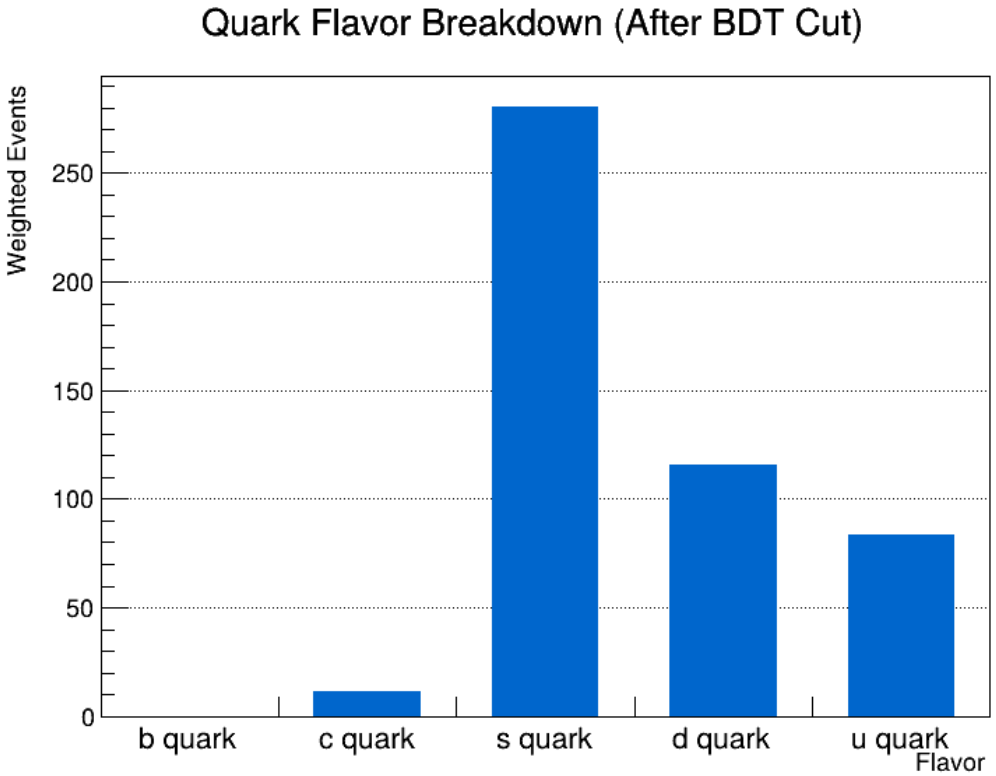
Flavor Breakdown for: 4fZZsemilep_tmva.root
Condition: Precut + BDT > 0.21
==========
Total Surviving Events : 782.821

-> b quarks (Z->bb etc) : 0.100646
-> c quarks (Z->cc etc) : 17.2104
-> s quarks (Z->ss etc) : 433.279
-> d quarks (Z->dd etc) : 181.363
-> u quarks (Z->uu etc) : 150.868

$\nu\bar{\nu}Z \to \nu\bar{\nu}q\bar{q}$事象におけるフレーバー内訳と結論

$\nu\bar{\nu}Z$背景事象（残存約490事象）においても、ZZ事象と同様の解析を行った。

- **重フレーバーの排除と $Z \to s\bar{s}$の残存**
  - bクォーク由来は**0事象**、c クォーク由来も約11事象のみであり、ここでも重フレーバーの除去が機能している。
  - 残存事象の過半数（約280事象、全体の約57%）が sクォーク由来であった。次いで d クォーク（約115事象）、u クォーク（約83事象）と続く。
- **結論（Zボソン由来の背景事象の総括）**
  - シグナル領域にZZ系背景事象が残存する最大の要因は、**Zボソンが $s\bar{s}$（および類似する軽クォーク）に崩壊した事象**であることが2つの独立したサンプルから証明された。



==========
Flavor Breakdown for: 4fZnusemilep_tmva.root
Condition: Precut + BDT > 0.21
==========
Total Surviving Events : 490.797

-> b quarks (Z->bb etc) : 0
-> c quarks (Z->cc etc) : 11.3781
-> s quarks (Z->ss etc) : 280.455
-> d quarks (Z->dd etc) : 115.626
-> u quarks (Z->uu etc) : 83.3371

今後1週間の予定と方針

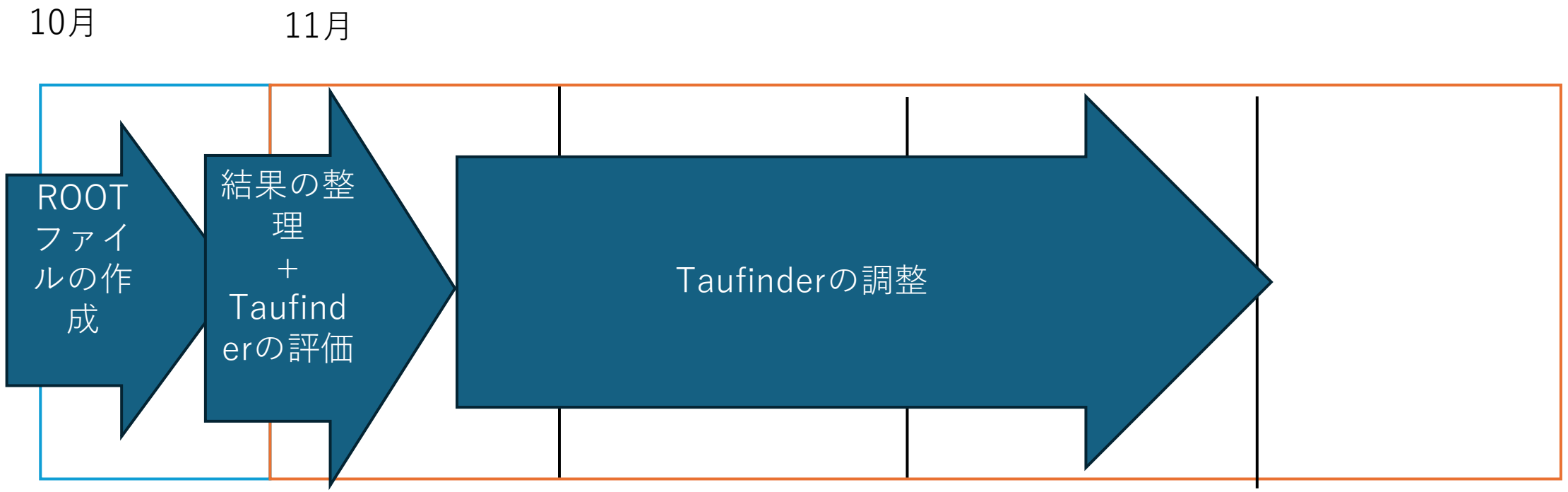
現在のステータス

Taufinderの調整

- **Taufinderパラメーターを操作し変化を確認（現在実行中）**

今後1週間の具体的なTo-Do

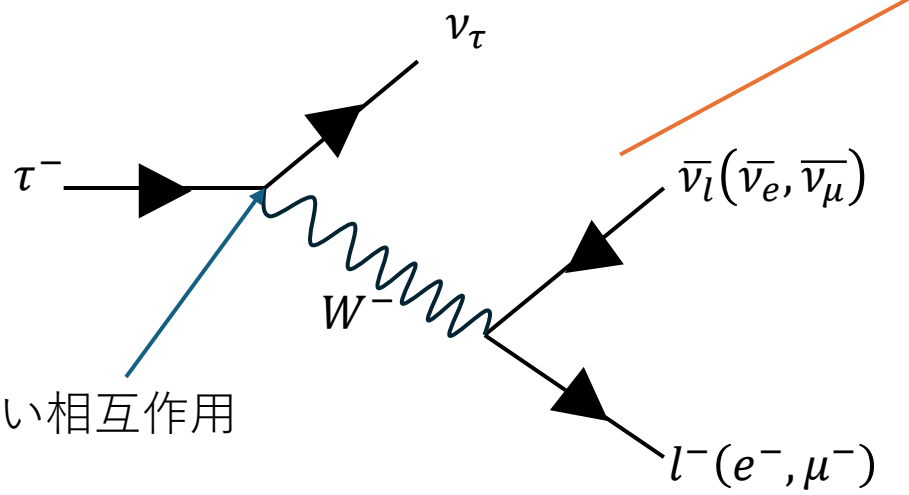
- **Taufinderパラメーターによる変化をまとめる**
- **中間発表の準備**

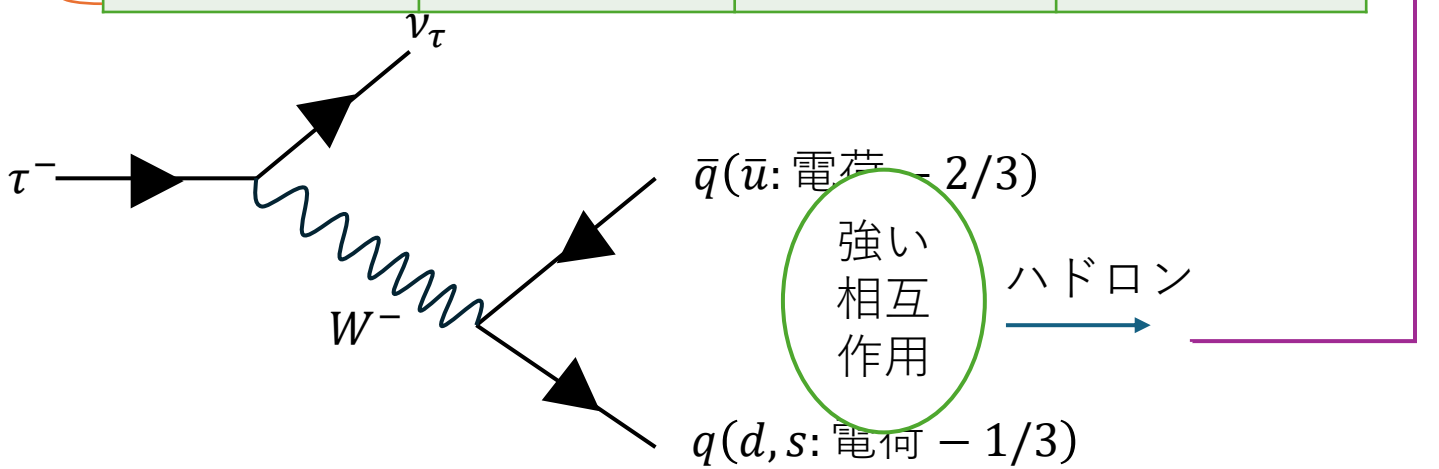


Tau lepton

質量(静止): $1.77\ GeV$
平均寿命: $2.90 \times 10^{-13}\ s$
電気素量(e)と等しい電荷と1/2スピンをもつ
弱い相互作用でハドロンに崩壊しうる唯一のレプトン
レプトンに崩壊する時も弱い相互作用

分類 (合計分岐比)	崩壊生成物	生成物記号の例 (τ^-の場合)	分岐比
ハドロン崩壊 (約64.79%)	荷電パイ中間子, 中性パイ中間子, タウニュートリノ	π^-, π^0, ν_τ	25.49%
	荷電パイ中間子, タウニュートリノ	π^-, ν_τ	10.82%
	荷電パイ中間子, 中性パイ中間子, タウニュートリノ	$\pi^-, 2\pi^0, \nu_\tau$	9.26%
	荷電パイ中間子, タウニュートリノ	$\pi^-, \pi^+, \pi^-, \nu_\tau$	8.99%
	荷電パイ中間子, 中性パイ中間子, タウニュートリノ	$\pi^-, \pi^+, \pi^-, \pi^0, \nu_\tau$	2.74%
	荷電パイ中間子, 中性パイ中間子, タウニュートリノ	$\pi^-, 3\pi^0, \nu_\tau$	1.04%
レプトン崩壊 (約35.21%)	電子, 電子反ニュートリノ, タウニュートリノ	$e^-, \bar{\nu}_e, \nu_\tau$	17.82%
	ミューオン, ミューオン反ニュートリノ, タウニュートリノ	$\mu^-, \bar{\nu}_\mu, \nu_\tau$	17.39%



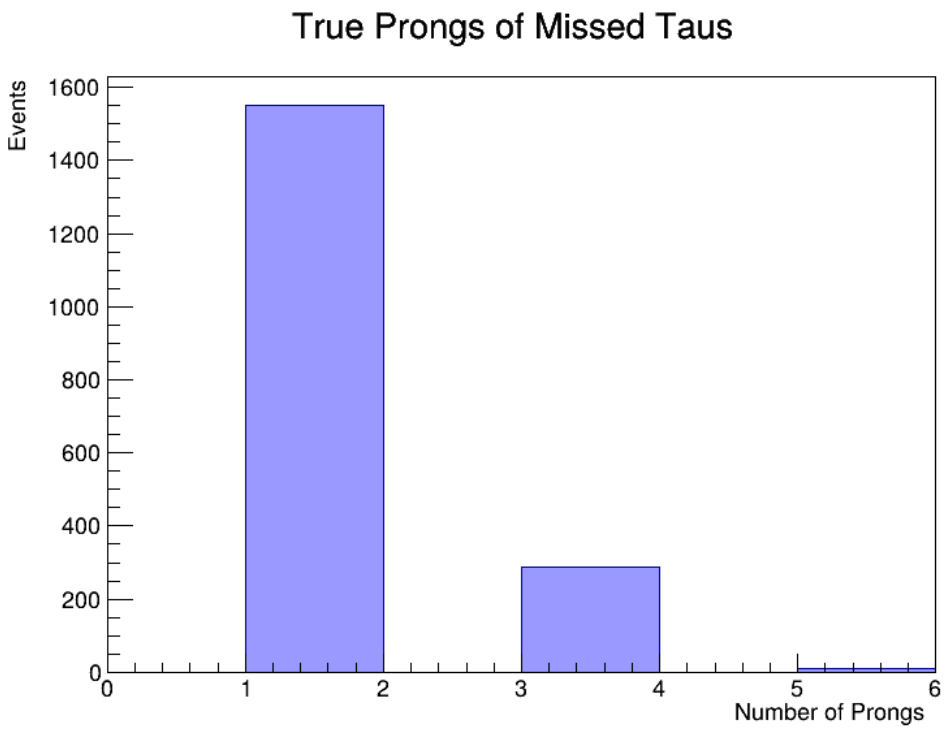


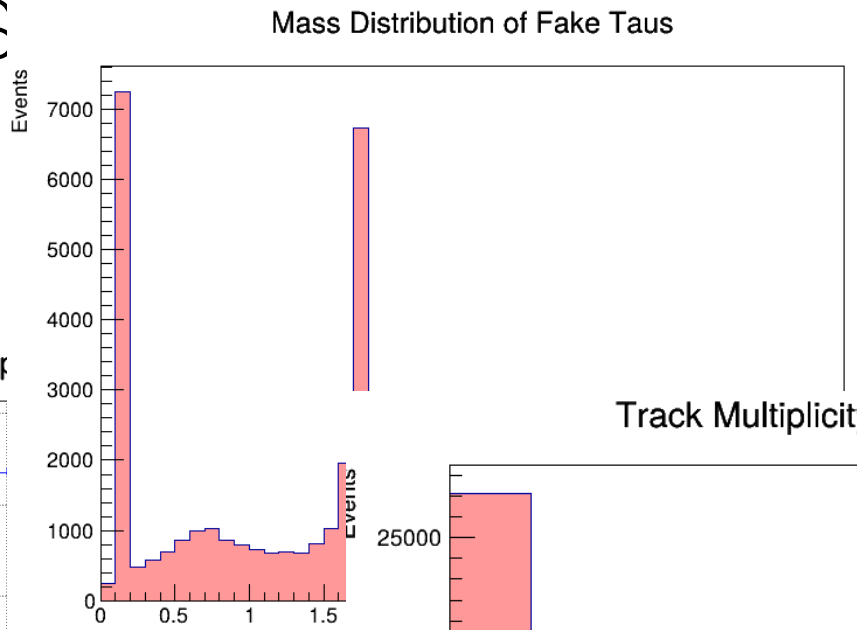
変更箇所：TauMass を 2.0 から 1.8 に変更する

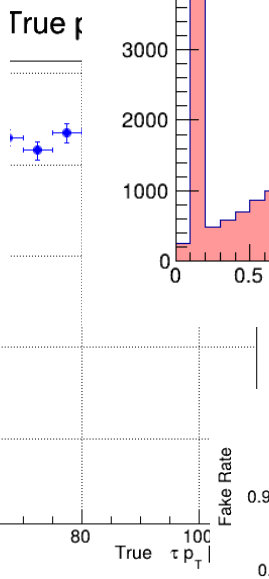
1番
fake落とし

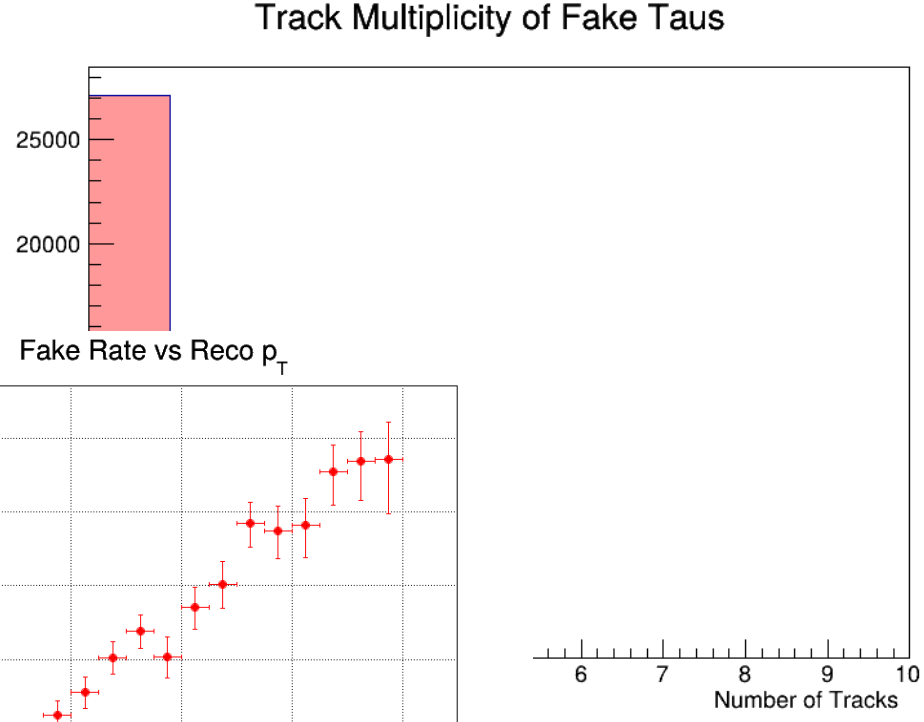
- **XMLの変更イメージ:** <parameter name="TauMass" type="float"> 1.8 </parameter>
- **狙いと見るべきポイント:**
  - 先ほどの質量分布のグラフで、「Fakeの質量が上限の 2.0 GeV に異常に張り付いている」という結果が出た。
  - タウの本来の質量は約 1.777 GeV です。上限を 1.8 GeV に厳しく絞ることで、誤検知（Fake）の正体 ①：質量分布で2.0 GeV にいた「無理やりタウのフリをしていた重いクォークジェットのFake」を削ぎ落とせるはずです。
  - **結果の確認:** fake_mass.png で 2.0 GeV のピークが消えるか？ そして全体の Fake Rate がどれくらい下がるかを確認。

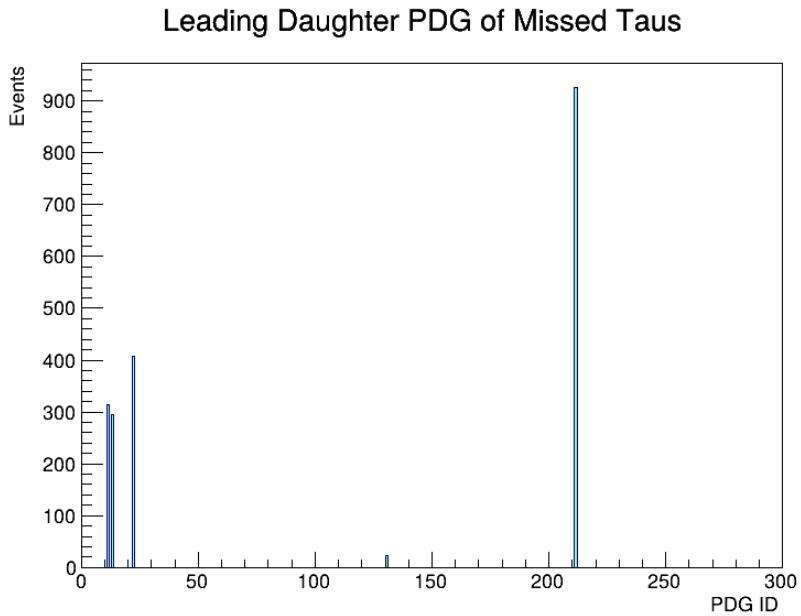
SS

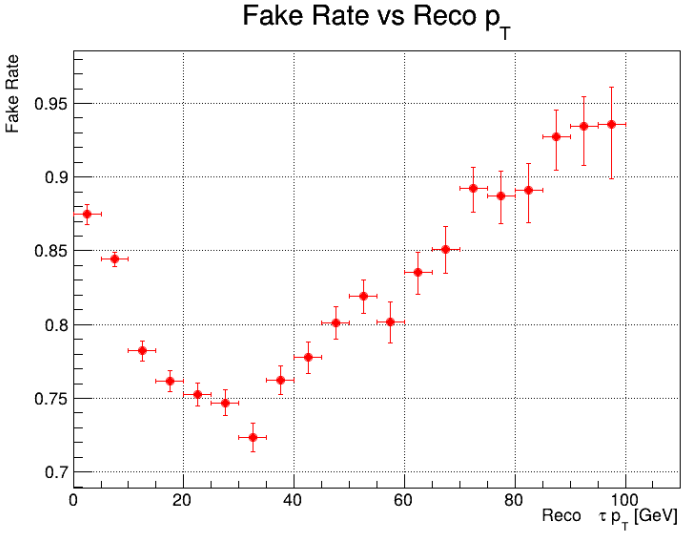

True Prongs of Missed Taus
Events
Number of Prongs


Mass Distribution of Fake Taus
Events




Track Multiplicity of Fake Taus
Number of Tracks


Leading Daughter PDG of Missed Taus
Events
PDG ID


Fake Rate vs Reco p_T
Fake Rate
Reco τ p_T [GeV]

悪化してない？

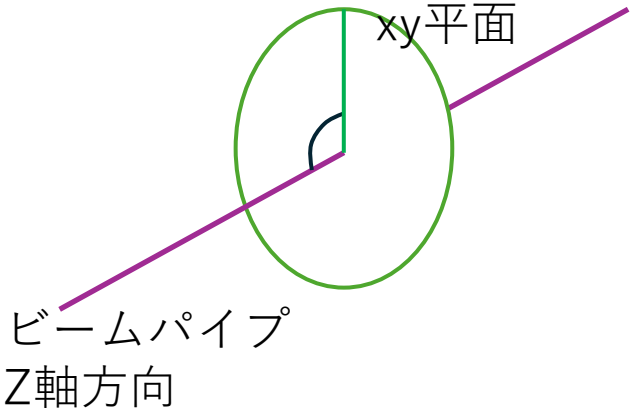
変更箇所：MinimumTrackEnergy を 1.0 から 0.5 に変更する

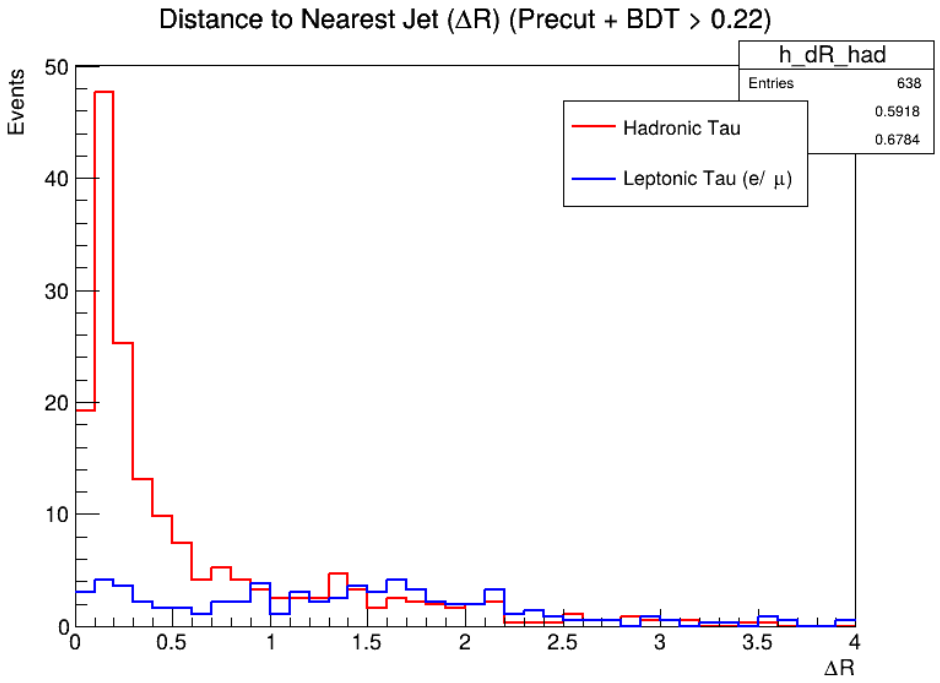
2番
見逃した本物（True）を救う

- **XMLの変更イメージ:** <parameter name="MinimumTrackEnergy" type="float"> 0.5 </parameter>
- **狙いと見るべきポイント:**
  - 先ほどの解析で、「見逃している本物のタウの大半は、低運動量のパイ中間子（$\pi^{\pm}$）」であることが判明しました。
  - 飛跡のエネルギーの足切りラインを 1.0 GeV から 0.5 GeV に下げることで、これまで切り捨てていた低エネルギーのタウを拾えるようになるはずです。
  - **結果の確認:** eff_vs_pT.png の左端（20 GeV以下の低運動量領域）のEfficiencyが持ち上がるか？ そして、全体の Efficiency がどこまで上がるか（逆にFake Rateがどれだけ悪化してしまうか）を確認します。

運動学的要因の解明（結果2：ΔR分布）

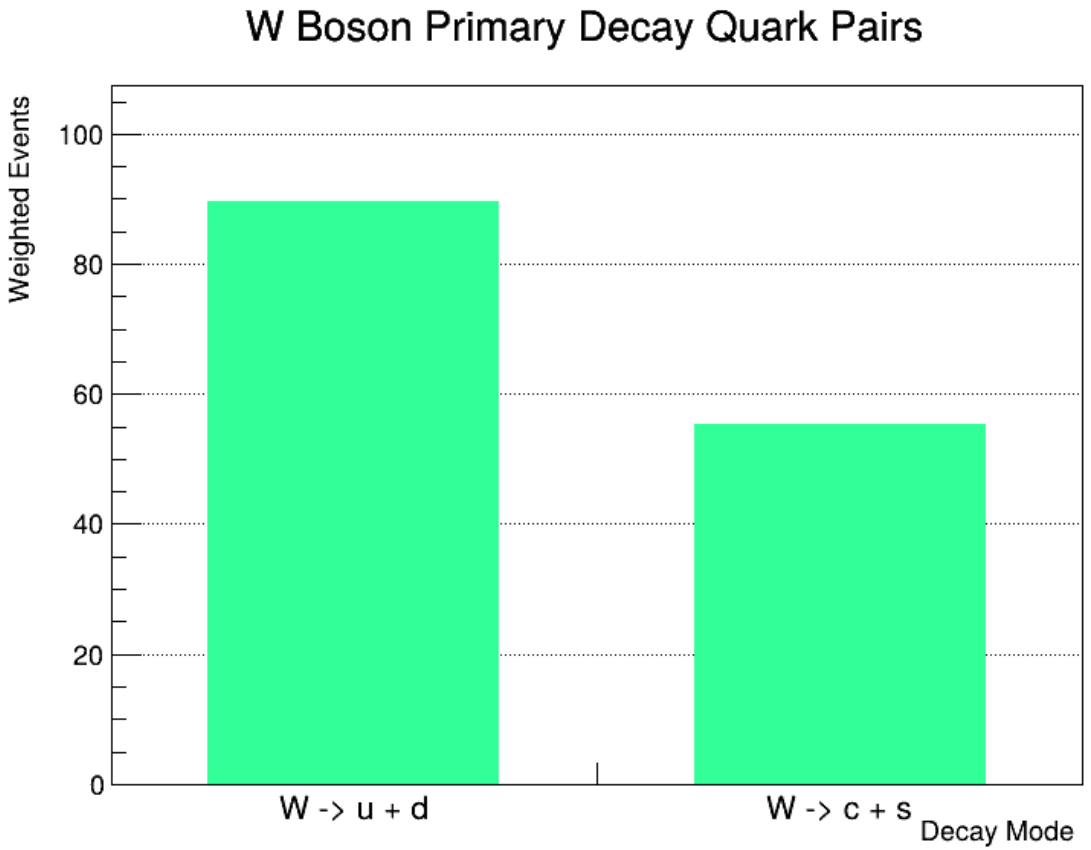
- **評価指標 (ΔR):** 粒子間の空間的な近さ（角度差）を表す指標。$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ で定義される。
  - η（擬ラピディティ）：ビーム軸に対する角度（極角 θ）を変換した座標。前方・後方への広がりを表す。
  - ϕ（方位角）：ビーム軸周りの角度（xy平面の角度）。
- **結果:** ハドロニックタウ・レプトニックタウを問わず、大半の事象が $\Delta R < 0.5$の極めて狭い領域にピークを持っている。
- **小結論:** すり抜けたタウ（および直接のレプトン）は、Wボソン由来のクォークジェットのすぐ近く、実質的にジェット内部を飛んでいたことがわかる。





未知のタウ起源の特定（クォーク崩壊由来のタウ）

- **目的:**
  - 前段で確認されたW/e/Z由来ではないタウ（Other: 約0.086事象）の真の起源を特定する。
- **手法:**
  - 該当事象のタウの親・祖父母粒子のPDG IDを精査し、その中で最も重い粒子（ID最大値）を代表的な親として抽出した。
- **結果と結論:**
  - これは、$WW \to \ell\nu qq$ のクォーク側（qq）で生成された cクォークが D_s^{*+}中間子となり、そこから二次的にタウが崩壊した事象である。
  - 事象の全体像としては、レプトン側がミューオン（μ）等に崩壊して isolepton カットをすり抜け、同時にクォーク側からタウが出たというケース（これが $1 \to 0$へ移動した事象の正体）であることが完全に証明された。
    ※レプトン普遍性により、レプトン側が τの場合（$2 \to 1$ へ移動する事象）でも、クォーク側では全く同じ現象が起きていると結論付けられる。




```
=== Flavor Analysis for 4fWWsemilep_full_tmva.root ===
[1] Primary W Boson Decay Quark Pairs
  Total Events              : 145.002
  W -> c + s (Primary)      : 55.4753
  W -> u + d (Primary)      : 89.5269
------------------------------------------------
[Extra] Parton Shower Effects (Subset of above)
  Events with b quark (g->bb)  : 0.0870885
  -> Note: This 'b' originates from gluon splitting in the shower, not W decay.-------
[2] Parent PDGs of 'Other' Taus (Not W, e, Z)
  PDG: 433 (Weight: 0.0870885)    (D_s^{*+} 中間子)
```

システム要因の裏付け：Marlin設定の確認

- **目的:** 運動学的な結論（ΔR が小さいと弾かれる）が、実際のソフトウェアの挙動と一致しているかを確認する。
- **手法:** ステアリングファイル(seinoproc_tau.xml)のプロセッサ実行順序とパラメータを確認。

システム要因の裏付け：Marlin設定の確認

結果:

- **処理順序:** <execute> ブロックにおいて、IsolatedLeptonTagging →JetClustering の順で処理されていることを確認。
- **MVAカット:** レプトン判定にはMVA出力による孤立条件（Electron: 0.5, Muon: 0.7）が課されており、ΔR が小さく周囲のエネルギー被りがある粒子は見逃される仕様と完全に符合した。
- **TauFinderの現状:** 独立して動作する taufinder プロセッサが存在するが、ConeMinCosAngle=0.9 などのコーン角・エネルギー比のパラメータが未調整である。現状では正確なタウVetoとして機能しておらず、後段のJetClusteringにも結果が反映されていない。

重み付けとは

- **シミュレーションと現実のギャップを埋める操作**
  - シミュレーション（MC）では、解析に必要な統計精度を確保するため、実際の発生確率（断面積）に関わらず、意図的に大量の事象（例：数十万〜数千万件）を生成している。
- **本来の物理的な発生スケールへの変換**
  - 生成したそのままの事象数で解析を行うと、現実と全く異なるS/B比（シグナルと背景の比率）になってしまう。
  - そこで、実際のILC実験の条件（積分ルミノシティ $900\ \mathrm{fb}^{-1}$ など）を想定し、各事象が現実の検出器で何回観測されるかという本来の期待値にスケールを合わせる必要がある。
- **重み付けのメリット・目的**
  - 現実の物理環境（正しい事象数分布）を正確に再現できる。
  - これにより、シグナルの有意性（Significance：どれだけ背景に埋もれず発見できるか）の正しい評価や、最適なカット条件の探索が可能になる。

重みと重み付けされたイベント数

重み

$$Weight = \frac{\mathcal{L} \times \sigma}{N_{gen}} \times W_{pol} \times BR$$

$\mathcal{L}$(Luminosity):積分ルミノシティ　$900\ \mathrm{fb}^{-1}$

σ(Cross Section): 事象の断面積

N_{gen}(Total Generated Events):ある事象についての全生成事象数

W_{pol}:偏極の重み 0.585 *今回作ったのはeL_pRのみ

BR(Branching Ratio):崩壊分岐比

重み付けされたイベント数

シミュレーションで生成されたすべての事象の重みを足し合わせた合計値

$$N_{weighted} = \sum_{i}^{N_{gen}} Weight_i$$

$$= N_{gen} \times \left(\frac{\mathcal{L} \times \sigma}{N_{gen}} \times W_{pol} \times BR \right)$$

$$= \mathcal{L} \times \sigma \times W_{pol} \times BR$$

想定ルミノシティと（偏極や分岐比を考慮した）物理的な断面積を掛け合わせた値になる
つまり、実験を行った際に、検出器で実際に観測されると思われる、本来の事象数

＊BRを考慮する必要があるのは今回はZH随伴生成のブランチであるものに対してだけで、そのほかの事象に対してはBR=1

重み付けの結果

生成された生イベント数と重み付けされたイベント数

	プロセス	ファイル名	生イベント数	重み付けされたイベント数
signal	$ZH \to \nu\bar{\nu}s\bar{s}$	nnss	200000	16.5
背景事象	$ZH \to \nu\bar{\nu}b\bar{b}$	nnbb	200000	38721.6
	$ZH \to \nu\bar{\nu}c\bar{c}$	nncc	200000	1952.9
	$ZH \to \nu\bar{\nu}g\bar{g}$	nngg	200000	5751.2
	$ZH \to \nu\bar{\nu}ZZ^*$	hzz	1000000	1771.7
	$ZH \to \nu\bar{\nu}WW^*$	hww	1000000	14428.3
	$ZH \to \nu\bar{\nu}\tau^+\tau^-$	htautau	1000000	4241.3
	$ZZ \to q\bar{q}l^+l^-/q\bar{q}\nu\bar{\nu}$	4fZZsemileptonic	3099600	321101.2
	$\nu\bar{\nu}Z \to \nu\bar{\nu}q\bar{q}$	4fZnusemileptonic	2298400	238962.4
	$WW \to \mu\nu q\bar{q}, \tau\nu q\bar{q}$	4fWWsemileptonic	36086800	9887219.3
	$WW \to e\nu q\bar{q}$	4fWsemileptonic	45357600	5404003.8
	$WW/ZZ \to l\nu q\bar{q}$	4fmixleptonic	8400000	861910.4

*今回作ったのはeL_pRのみ

Pre-cutの結果

- 条件付けによって、背景事象を一定数まで減らすことに成功。
- しかし、シグナルを失わずに背景だけをこれ以上削ることは、単純なカット（直線の境界線）では難しい。

⇒複数の変数の相関関係から複雑な境界線を引ける機械学習（BDT）へ移行

適用した条件(清野さんの修論と同じものを使用)

- Missingmass<140GeV,
- 100GeV<dijetmass<200GeV,
- 5GeV< Tmomentum <100GeV,
- P_{max}<80GeV,
- 9<$N_{charged}$,
- isolepton=0.

File Name	Original	After Cut	Efficiency
4fWWsemilep	9.88722e+06	3.50232e+06	35.42 %
4fWsl	5.404e+06	218446	4.04 %
4fZZsemilep	321101	132905	41.39 %
4fZnusemilep	238962	96582.4	40.42 %
4fmixlep	861910	36533.5	4.24 %
htautau	4241.26	341.019	8.04 %
hww	14428.3	7099.68	49.21 %
hzz	1771.67	947.553	53.48 %
nnbb	38721.6	36201	93.49 %
nncc	1952.86	1907.32	97.67 %
nngg	5751.2	5687.56	98.89 %
Total Background (B)	1.67801e+07	4.03897e+06	24.07 %
Signal (nnss) (S)	16.5087	16.1884	98.06 %

Original Significance : 0.004
Pre-cut Significance : 0.0081

モデルの学習

目的: 多変量解析ツール（TMVA）による識別器の構築

- 44個の運動学的変数（質量、運動量、フレーバータグなど）を入力とし、シグナルと背景を最適に分離する決定木（BDT）を学習させる。

i=1,2

変数	定義
m_i	ジェットiの質量
$\boldsymbol{p}_i$	ジェットiの運動量
p_T	横方向運動量
p_L	縦方向運動量
M_{jj}	2ジェットの不変質量
m_{miss}	消失質量
P_{max}	イベント内の粒子が持つ運動量ノルムの最大値
$N_{charged}$	イベント内の荷電粒子数
$\cos\theta_i$	$(\boldsymbol{p}_i)_z/\lvert\boldsymbol{p}_i\rvert$
$\cos\theta_{opening\ angle}$	ジェット1,2がなす角の余弦
$N_{iso\gamma}$	Isolated photonの数
E_γ	Isolated photonのエネルギーの和
Y_{34}	Durhamにおいて，4本のジェット候補を3本にまとめる際の最小のy値
Y_{23}	3本のジェット候補を2本にまとめる際の最小のy値
Y_{12}	2本のジェット候補を(仮想的に)1本にまとめる際の最小のy値
Flavor tag value	各種フレーバーに対するスコア(11種)

全データへの適用

目的: 学習済みのモデルを用いて、本来の物理スケールでの事象選別を行う

- 学習させ作った重みファイルを、実際のルミノシティ重みを持つ全データに適用し、全事象にシグナルらしさ（BDTスコア）を付与する。
- **課題:** 膨大なデータ処理と、解析に必要な変数の保持
- データ量が約1億件と極めて膨大であり、単純な直列処理では時間がかかりすぎる。
- 学習時にBDTのカンニング防止のために隠していた変数（τ の数など）は、最終的なtauを含む事象がどれだけあるか確認するのに必要になるため、消さずに引き継ぐ必要がある。
- **工夫と解決策:** ROOTの機能活用とバッチジョブによる並列化
- ROOTの機能（CloneTree）を活用し、元の変数を失うことなく、新たにBDTスコアの列だけを横に追記して保存する処理を実装した。
- 全12種類の物理プロセスをバッチジョブ（bsub）に並列で投げることで、全事象のファイルのスコア計算をメモリ溢れなく完遂した。

最適カット値の探索

目的:最も物理的な発見確率（Significance）が高まる境界線の決定

- BDTが各事象に付与したスコア（-1.0 ～ 1.0）に対して、どこでカットを行えば、シグナルを最大限残しつつ背景を極限まで削れるかを探索する。

課題:単純なシグナルの数だけでは最適解が決まらない

- カットを緩くすればシグナルは増えるが、同時に膨大な背景事象も混入してしまう。逆に厳しくしすぎると、シグナル自体が消滅してしまう。
- Significance（$S/\sqrt{S+B}$）
  という統計的な優位性を最大化する必要がある。

工夫と解決策: 全スコア領域での自動スキャンと可視化

- BDTスコアに対するカット値を0.005 刻みで少しずつ変化させ、すべての地点での Significance を自動計算する
- ⇒算出された Significance の推移をグラフ化し、最大のピークを迎えるベストなカット値を特定する。

Taufinderのbaseの状態

