

修士論文

LHC-ATLAS 実験における 2 本の飛跡で
再構成されたタウ粒子の同定手法改善と
物理解析への応用

2022/03/25

早稲田大学 先進理工学研究科 物理学及応用物理学専攻
寄田浩平研究室

5320A071-6 山際美由希

概 要

宇宙で起こる全ての現象を説明することを目的とする素粒子物理学において、素粒子標準模型は大きな成功を収めている。しかしながら、素粒子標準模型では説明しきれない現象や実験事実が複数確認されている。この問題を解決するために新たな物理モデルが様々な考案され、実験による探索が進められているが、未だ発見には至っていない。

世界最高エネルギーでの陽子・陽子衝突型実験を行っている施設の 하나가 LHC-ATLAS 実験である。LHC(Large Hadron Collider) は重心系エネルギー 13TeV で運用を行っている。ATLAS 実験は LHC で行われている実験の 1 つであり、汎用検出器を用いてさまざまな物理事象探索を可能としている。熱心に探索が進められている物理事象として、MSSM(Minimal Supersymmetric Standard Model) におけるヒッグス粒子や、第三世代に崩壊するレプトクォークがある。特に MSSM の中性ヒッグスが 2 つのタウ粒子対に崩壊するモードは、 $\tan\beta$ の大きなパラメータ領域を広く探索することのできる解析モードである。またレプトクォークは B アノマリーから動機づけられている新粒子候補であり、第三世代に崩壊するモードが最も有力だと考えられている。これらの解析においてタウ粒子は非常に重要なオブジェクトである。また現在の探索状況から、これらの新粒子候補は TeV 領域以上の高質量であると考えられており、したがって高運動量のタウ粒子が重要となっている。しかしながら、タウ粒子を再構成しオブジェクトを同定することは電子やミュオンと比べて技術的に難しく、特に高運動量領域においては再構成効率が低くなっているのが現状である。

タウ粒子は寿命約 2.90×10^{-13} sec で崩壊し、ハドロン崩壊するタウ粒子は奇数本の飛跡を残す。この特徴を用いて粒子同定をしているが、飛跡の検出効率や分類が完全ではないため 2 本の飛跡で再構成されてしまうことがある。2 本の飛跡で再構成されるタウ粒子は、同定効率が低いことと電荷を完全に測定できないことに問題があるため、終状態にタウ粒子を含む多くの解析では 1 本または 3 本の飛跡で再構成されるタウ粒子しか用いられていない。特に高運動量領域では 2 本の飛跡で再構成されるタウ粒子の割合が増加するため、高運動量のタウ粒子を用いる物理事象探索感度に影響が出ている。

本研究ではまず新物理探索 $H/A \rightarrow \tau\tau$ 解析において 2 本の飛跡で再構成されたタウ粒子を含む事象数をモンテカルロシミュレーションサンプルを用いて調べた。その結果それらが一定の割合で存在していることがわかった。よって 2 本の飛跡で再構成されたタウ粒子を物理解析で使えるようにするための開発を行った。電荷が ± 2 で再構成されたタウ粒子の電荷の識別効率を調べ、1 本または 3 本の飛跡で再構成されるタウ粒子の識別効率と比較した。また電荷が 0 で再構成されたタウ粒子の電荷を識別する手法を、真の飛跡数が 1 本と 3 本の場合のそれぞれで開発した。またそれらを適用するために、真の飛跡数が 1 本か 3 本かを識別する手法についても開発し、それらの総合的な性能を評価した。低運動量領域における電荷識別効率は、真の飛跡数が 1 本のときに 85%以上、3 本のときに 70%を達成した。1TeV 程度の高運動量領域では 1 本のときに約 70%、3 本のときは約 55%となった。真の飛跡数識別は、全ての運動量領域で約 85%の識別効率を達成した。さらに総合的な性能は、低運動量領域では約 80%、1TeV 程度で約 70%程度の識別効率となった。また、本研究で開発した電荷識別手法を $H/A \rightarrow \tau\tau$ 解析の信号領域で適用したところ、従来解析と同様に 2 つのタウ粒子が反対電荷となる事象を含めると、信号領域の事象数が 1TeV のヒッグス粒子の場合約 18%、2TeV の場合 25%以上増加することが見込まれる結果が得られた。

目次

1	序論	4
1.1	素粒子標準模型	4
1.2	素粒子物理学における課題	5
1.3	$H/A \rightarrow \tau\tau$ の物理	6
2	LHC-ATLAS 実験	8
2.1	LHC	8
2.2	ATLAS 実験	9
2.3	ATLAS 検出器	9
2.3.1	ATLAS 座標系	9
2.3.2	内部飛跡検出器	10
2.3.3	カロリメータ	11
2.3.4	ミューオン検出器	13
2.4	トリガーシステム	14
2.5	オブジェクトの再構成	14
3	タウ粒子	16
3.1	タウ粒子の性質	16
3.2	タウ粒子の再構成	17
3.3	タウ粒子の同定	19
3.3.1	電子の排除	20
3.4	タウ粒子トリガーシステム	20
3.4.1	L1 tau Trigger	20
3.4.2	HLT tau Trigger	20
4	ATLAS 実験における $H/A \rightarrow \tau\tau$ 過程の探索	22
4.1	解析の概要	22
4.2	$H/A \rightarrow \tau\tau$ の探索結果	23
4.3	$H/A \rightarrow \tau\tau$ の解析感度向上のための試み	24
4.4	2-prong のタウ粒子を使用した解析の検討	25
4.4.1	検討手法	25
4.4.2	信号領域の事象選択	26
4.4.3	信号領域の事象数	26
5	2-prong のタウ粒子の真の電荷の識別	28
5.1	再構成される飛跡の本数とタウ粒子の電荷	28
5.2	真の飛跡数の識別	32
5.2.1	真の電荷を識別する流れ	32
5.2.2	BDT を用いた真の飛跡の本数の識別器の開発	32
5.3	truth 1-prong $\rightarrow$ reco 2-prong のタウ粒子の電荷の識別	37
5.4	truth 3-prong $\rightarrow$ reco 2-prong のタウ粒子の電荷の識別	42
5.5	電荷識別性能の評価	43
5.6	$H/A \rightarrow \tau\tau$ 解析への電荷識別手法の適用	43
6	まとめ	46

7	考察と今後の展望	47
8	謝辞	49
A	多変量解析 Boosted Decision Trees	53
A.1	ブースティング手法	53
A.1.1	Adaptive Boost (AdaBoost)	54
A.1.2	Gradient Boost	54
A.2	入力変数	55
A.3	超パラメータ	55
B	タウ粒子同定 BDT の入力変数の説明	56
C	タウ粒子の識別効率	57
D	真の飛跡数識別の BDT で用いる入力変数の分布	57
E	truth 1-prong $\rightarrow$ reco 2-prong のタウ粒子の電荷識別の BDT で用いる入力変数の分布	63

表目次

1.1	2HDM のタイプと結合の大きさ	6
3.1	タウ粒子の崩壊分岐比	16
3.2	タウ粒子同定 BDT の入力変数	19
3.3	BDT によるタウ粒子同定の基準点の取得効率	19
4.1	$H/A \rightarrow \tau\tau$ 現行解析の各チャンネルで用いたトリガー	22
4.2	$H/A \rightarrow \tau\tau$ 現行解析の信号領域の事象選択	23
4.3	$H/A \rightarrow \tau\tau$ 現行解析の主要な系統誤差	23
4.4	信号領域の事象選択	26
4.5	3 種類の信号領域の事象数	27
5.1	崩壊モードの定義	28
5.2	4 つのタウ粒子 p_T 領域の定義	33
5.3	分離能力の上位 10 変数 [pt low, medium]	34
5.4	分離能力の上位 10 変数 [pt high, veryhigh]	34
5.5	BDT 超パラメータの候補値と決定値 (真の飛跡数の識別)	36
5.6	分離能力の上位 10 変数 [pt low, medium]	38
5.7	分離能力の上位 10 変数 [pt high, veryhigh]	38
5.8	BDT 超パラメータの候補値と決定値 (truth 1-prong $\rightarrow$ 2-prong のタウ粒子の電荷識別)	40
5.9	3 種類の信号領域の事象数 (電荷識別適用後)	45
A.1	BDT における超パラメータ	56

図目次

1.1	標準模型の素粒子	4
1.2	$R(D)$ と $R(D^*)$ の実験値と理論値 [4]	5
1.3	ミューオンの異常磁気能率 $g - 2$ の実験値と理論値 [8]	5
1.4	MSSM ヒッグスの探索状況 [36]	7
1.5	中性ヒッグス粒子の生成過程	7
2.1	LHC 立体図 (写真提供 CERN アトラス実験グループ) [3]	8
2.2	ATLAS 実験 Run2 の積分ルミノシティ [14]	9
2.3	ATLAS 実験 Run2 の平均パイルアップ数分布 [14]	9
2.4	ATLAS 検出器全体図 [13]	10
2.5	動径方向から見た内部飛跡検出器 [16]	11
2.6	カロリメータ全体図 [13]	12
2.7	ミューオン検出器全体図 [13]	13
2.8	ATLAS トリガーシステム [17]	15
3.1	トラックの分類の流れ	18
3.2	Run2 における HLT tau trigger の性能 [17]	21
4.1	$H/A \rightarrow \tau\tau$ 解析の制限結果 [28]	24
4.2	2-prong のタウ粒子の同定性能	25
4.3	信号事象の飛跡数と電荷情報による分類	26
4.4	信号領域の事象数の増加割合	27
5.1	再構成される飛跡の数の割合	28
5.2	崩壊モードごとに分けた、再構成される飛跡の数の割合	29
5.3	2-prong のタウ粒子のイメージ図のタウ粒子	29
5.4	1-prong のタウ粒子と 3-prong のタウ粒子の電荷の正解率	30
5.5	2-prong のタウ粒子の再構成される電荷の組み合わせ	31
5.6	truth 3-prong $\rightarrow$ reco 2-prong のタウ粒子の理由 2	31
5.7	2 本のタウトラックが同電荷の 2-prong のタウ粒子の電荷の正解率	31
5.8	目指す 2-prong のタウ粒子の識別の流れ	32
5.9	各 p_T 領域における BDT score の分布	36
5.10	ROC curve (真の飛跡数識別)	36
5.11	1 変数でタウ粒子の電荷識別をしたときの正解率 (truth 1-prong $\rightarrow$ reco 2-prong)	37
5.12	決定した BDT の入力変数 (low p_T 領域)	39
5.13	各 p_T 領域における BDT score の分布	40
5.14	ROC curve (truth 1-prong $\rightarrow$ reco 2-prong)	41
5.15	タウ粒子の p_T と電荷の正解率の関係 truth 1-prong $\rightarrow$ reco 2-prong)	41
5.16	タウ粒子の p_T と電荷の正解率の関係 (truth 3-prong $\rightarrow$ reco 2-prong)	43
5.17	カテゴリーごとの正解率 (truth 3-prong $\rightarrow$ reco 2-prong)	43
5.18	電荷の正解パターン	43
5.19	合わせた電荷識別効率	44
5.20	電荷 0 の 2-prong のタウ粒子の決定した電荷識別効率	44
5.21	2-prong のタウ粒子の決定した電荷識別効率	44
5.22	信号領域の事象数の増加割合 (電荷識別適用後)	45
A.1	1 つの決定木のイメージ	53
C.1	再構成されたタウ粒子に対する ID の効率	58

D.1 真の飛跡数識別の BDT で用いる入力変数の分布 (low p_T 領域)	59
D.2 真の飛跡数識別の BDT で用いる入力変数の分布 (medium p_T 領域)	60
D.3 真の飛跡数識別の BDT で用いる入力変数の分布 (high p_T 領域)	61
D.4 真の飛跡数識別の BDT で用いる入力変数の分布 (veryhigh p_T 領域)	62
E.1 決定した BDT の入力変数 (medium p_T 領域)	63
E.2 決定した BDT の入力変数 (high p_T 領域)	63
E.3 決定した BDT の入力変数 (veryhigh p_T 領域)	63

1 序論

1.1 素粒子標準模型

素粒子物理学は、宇宙で起こる現象の全てを記述することを目標として、物質を構成する最小単位である素粒子とそれらの相互作用を研究する学問である。現在の素粒子物理学において素粒子標準模型と呼ばれる理論は多くの実験結果を非常によく説明する。本章では素粒子標準模型と終状態にタウ粒子を含む物理について説明する。

標準模型の素粒子は、物質を構成するフェルミオンと相互作用を担うボソンに大きく分類されており、図 1.1 のようにまとめられる。フェルミオンはさらにクォークとレプトンに分かれそれぞれ第 1 世代から第 3 世代までの階層構造を持っていて、各世代に電荷の異なる 2 つの素粒子がある。クォークの上段のアップクォーク、チャームクォーク、トップクォークは電荷が $+2/3e$ 、下段のダウルクォーク、ストレンジクォーク、ボトムクォークは電荷が $-1/3e$ である。これらは単位電荷 e よりも小さい単位であるが、それぞれ単体で存在することはできない。クォーク同士は強い力の相互作用により常に 3 つまたは 2 つのクォークの結合状態として存在しており、それぞれをバリオン、メソン (中間子) と呼ぶ。レプトンの上段は荷電レプトンであり、電荷は -1 である。これと対になる下段はニュートリノであり、電荷は 0 である。

ボソンは力を媒介する粒子である。自然界には 4 つの力、電磁気力、弱い力、強い力、重力が存在し、このうち重力以外の 3 つの力が標準模型で記述されている。電磁気力を媒介するのは γ である。質量は 0 で、光速で運動する。弱い力を媒介するのは $W^\pm, Z$ である。 $W^\pm$ 粒子の質量は $80.385 \pm 0.015 \text{ GeV}$ 、 Z 粒子の質量は $91.1876 \pm 0.0021 \text{ GeV}$ と重い質量となっている。強い力を媒介するのは g (グルーオン) である。グルーオンは質量が 0 で、カラー荷を持つ。最後に質量の起源となる H (ヒッグス粒子) がある。2012 年に LHC で発見された最後の標準模型の粒子で、その質量は $125.10 \pm 0.14 \text{ GeV}$ である。

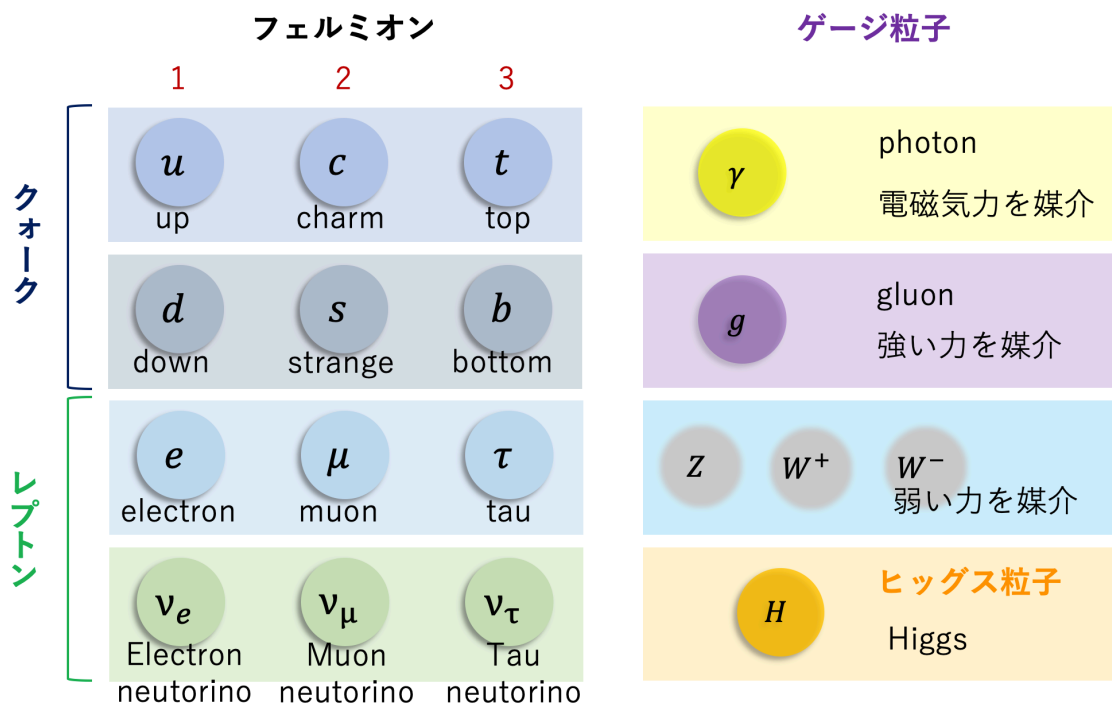


図 1.1: 標準模型の素粒子

1.2 素粒子物理学における課題

標準模型はよく成功した理論であるが、全ての現象を完全に記述することはできていない。

現在の素粒子物理学の実験領域において重要な課題は主に2つに大別することができる。1つは、標準模型の精査である。ヒッグス粒子やトップクォークの質量や、ヒッグス粒子とその他の粒子の結合定数、またヒッグス粒子の自己結合定数の測定を精密に行い、より詳細に標準模型の各種値を測定していくことが必要である。精密測定を実現することは、新物理の理論モデルのパラメータ領域を制限することに繋がる点でも重要である。また標準模型の理論計算とのズレが見つかった場合、それは新物理の重要な兆候となる。現在兆候が確認されている理論値とのズレとして代表的なものに、B アノマリーとミュオン $g-2$ がある。一般的に標準模型との差が 5σ で観測されたときに「発見」となる。現在の実験結果では、B アノマリーは実験値の平均が標準模型の理論値から約 3.4σ ([4])、ミュオン $g-2$ では約 4.2σ のズレ ([8]) が観測されている (図 1.2, 図 1.3)。取得データ量の蓄積に加えて、より精密な実験と解析を行っていき、このズレが真に存在するものかどうかに関着をつけることが必要である。

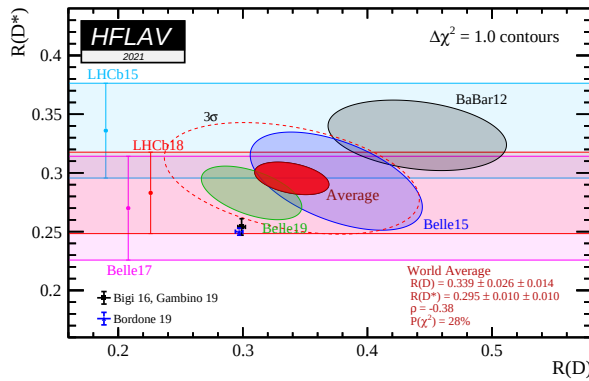


図 1.2: $R(D)$ と $R(D^*)$ の実験値と理論値 [4]

横軸の $R(D)$ は、 $B \rightarrow D l \nu_l (l = e, \mu)$ の崩壊分岐比に対する $B \rightarrow D \tau \nu_\tau$ の崩壊分岐比の値、縦軸の $R(D^*)$ は、 $B \rightarrow D^* l \nu_l (l = e, \mu)$ の崩壊分岐比に対する $B \rightarrow D^* \tau \nu_\tau$ の崩壊分岐比の値である。灰色の領域は BaBar 実験、青色の領域は Belle 実験の 2015 年の結果、緑色の領域は Belle 実験の 2019 年の結果、縦軸のみの結果として水色と赤色の領域は LHCb の 2015 年、2018 年の結果、ピンク色の領域は Belle 実験の 2017 年の結果を表している。これらの実験値の平均が赤色の領域である。一方で十字の記号が示す値は標準模型の理論値となっており、実験値の平均と約 3.4σ のズレが観測されている。

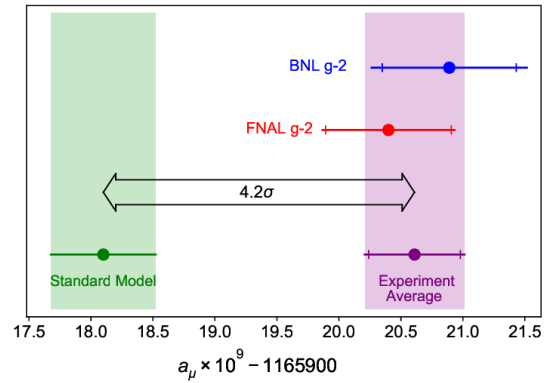


図 1.3: ミュオンの異常磁気能率 $g-2$ の実験値と理論値 [8]

青の値は BNL E821 の結果、赤の値は FNAL の結果であり、その平均値が紫の領域である。緑の値は標準模型の理論値を示しており、実験値の平均と約 4.2σ のズレが観測されている。各実験値の線についている小さな目盛りは、全ての不確実性に対する統計誤差の割合を示している。共に統計誤差が支配的な実験である。

2つ目は標準模型を超える新物理モデルの粒子を発見することである。標準模型の精査により見つかったズレが真に存在するとき、それらを解決する新物理が存在するはずである。B アノマリーが真に存在していた場合、ゲージボソンとレプトンの結合定数はフレーバーに依らないというレプトンフレーバー普遍性が第三世代で破れていることになる。特に第三世代のレプトンと強く結合する、レプトンとクォークに結合するレプトクォークはこの問題の解決策の有力な候補である。ミュオン $g-2$ のズレが真に存在していた場合、ミュオンが未知の粒子と結合することを示唆している。その候補は多数に上るが、フェルミオンとボソンの対称性を考える超対称性理論で、今までの実験結果と $g-2$ の結果を解決する理論が可能であると考えられている。その他標準模型で記述されていない課題の1つに暗黒物質の存在がある。渦巻き銀河の回

転速度分布の観測結果から、通常の光を用いて観測することができないが重力を感じる物質が存在することが立証されており、これを暗黒物質と呼称している。しかし暗黒物質の発見には至っていない。これまでの観測結果から、暗黒物質は

- ・電荷を持たない。
- ・質量がある。
- ・安定して存在する。

性質を持つことがわかっている。この有力候補の1つがニュートラリーノである。ニュートラリーノは超対称性理論において存在が予言される、ズィーノ (Z 粒子のパートナー)、フォティーノ (光子のパートナー)、中性ヒグシーノ (中性ヒッグス粒子のパートナー) による混合状態であるフェルミオンである。ニュートラリーノが超対称性粒子の中で最も軽い粒子であるとき、これが暗黒物質になり得ると考えられている。新物理探索では、様々な実験結果や現象から予測される新物理モデルにおける新粒子の存在とその性質をある程度仮定し、その粒子の探索を行っている。以上のように、標準模型の精査と新物理探索の両輪を回し、この宇宙の物理法則のより正確な記述を目指して研究が進められている。

1.3 $H/A \rightarrow \tau\tau$ の物理

新物理探索において、タウ粒子を終状態とする事象の探索は数多く進められている。超対称性理論を含むいくつかの新物理モデルでは重い中性ヒッグス粒子が存在し、それがタウ粒子対に崩壊する過程があると考えられている。ここではこの物理事象のモチベーションについて述べる。

ATLAS 実験と CMS 実験は、2012 年に発見した質量 125GeV の新粒子が標準模型で予言されていたヒッグス粒子と一致することを様々な研究で示してきた。しかし依然としてこの発見されたヒッグス粒子はより拡張されたモデルのうちの1つの粒子に過ぎない可能性が提唱されている。標準模型を超える多くの物理理論モデルにおいては、拡張ヒッグスセクターを含んでいる。最もシンプルな拡張はヒッグス二重項モデル 2Higgs Doublet Model(2HDM) である。このモデルでは2種類のヒッグス粒子 H_u^0 と H_d^0 が導入される。それぞれの真空期待値を ν_u, ν_d としたとき、 $\sqrt{\nu_u^2 + \nu_d^2} = \nu$ (ただし ν は標準模型における真空期待値) を満たす。このモデルでは、標準模型のスカラーヒッグス (h) に加え、重いスカラーヒッグス (H)、重い擬スカラーヒッグス (A)、荷電ヒッグス ($H^\pm$) の5つのヒッグス粒子が存在する。このモデルにおける重要なパラメータが2つの真空期待値の比である $\tan\beta$ であり次の式 1.1 で定義される。

$$\tan\beta = \nu_u/\nu_d \quad (1.1)$$

2HDM には、各ヒッグスセクターとフェルミオンとの結合の組み合わせによって幾つかの種類が存在する。主要なタイプを表 1.1 にまとめる

表 1.1: 2HDM のタイプと結合の大きさ

フェルミオンの種類	Type-I	Type-II	Type-X	Type-Y
up type quark (u,c,t)	小	小	小	小
down type quark (d,s,b)	小	大	小	大
charged lepton (e, μ, τ)	小	大	大	小

その中で Type II と呼ばれる 2HDM は、超対称性を実現する標準模型で最小の拡張モデルである最小超対称標準模型: Minimal Supersymmetric Standard Model (MSSM) におけるヒッグスセクターとなっている。ツリーレベルでは、MSSM のヒッグスセクターの性質は、重い擬スカラーヒッグス A の質量 m_A と真空期待値の比 $\tan\beta$ の2つのパラメータのみに依存する。この二次元平面で解釈した MSSM ヒッグス粒子

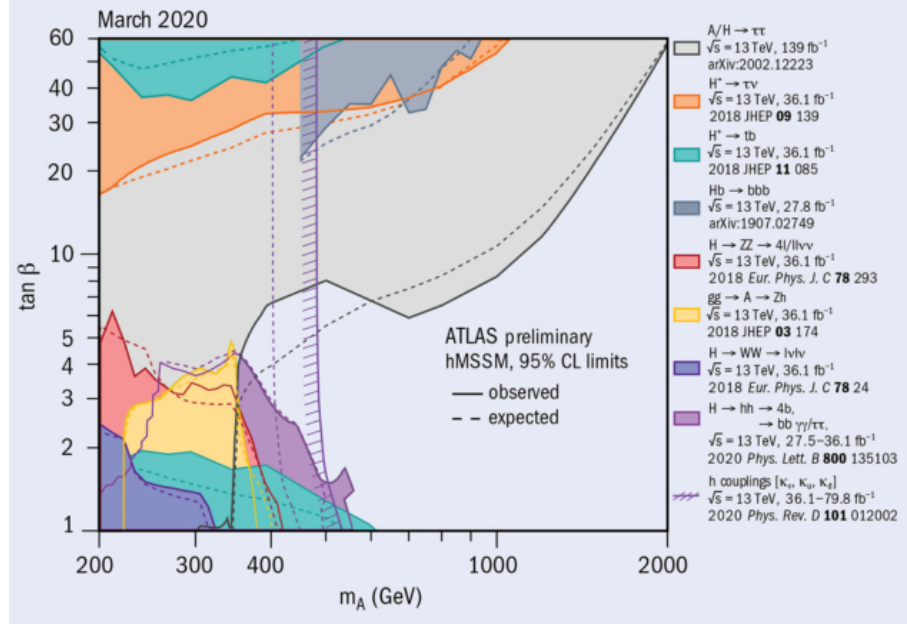


図 1.4: MSSM ヒッグスの探索状況 [36]

の ATLAS 実験における探索状況を図 1.4 に示す。グレーの領域で示されている $H/A \rightarrow \tau\tau$ 解析は、 $\tan\beta$ の大きいパラメータ領域を広く探索することのできる重要な解析モードであることがわかる。

中性ヒッグス粒子の生成過程について説明する。生成過程はグルーオンフュージョン (ggF) か b クォークを伴う過程 (bbH) が考えられている (図 1.5)。MSSM ヒッグス粒子は、 $\tan\beta$ の大きい領域ではダウンタイプのフェルミオンとの結合が標準模型と比べて大きくなるため、 b クォークを伴う生成過程の生成断面積は $\tan\beta$ に伴って大きくなる。また崩壊においても $\tan\beta$ が大きいほどタウ粒子や b クォークへの崩壊分岐比が大きくなる。このため、中性ヒッグス粒子が $\tau\tau$ や bb へ崩壊するモードの探索がこれまでも盛んに行われてきた (LEP[30], Tevatron[31],[32], LHC[33], [34], [35])。

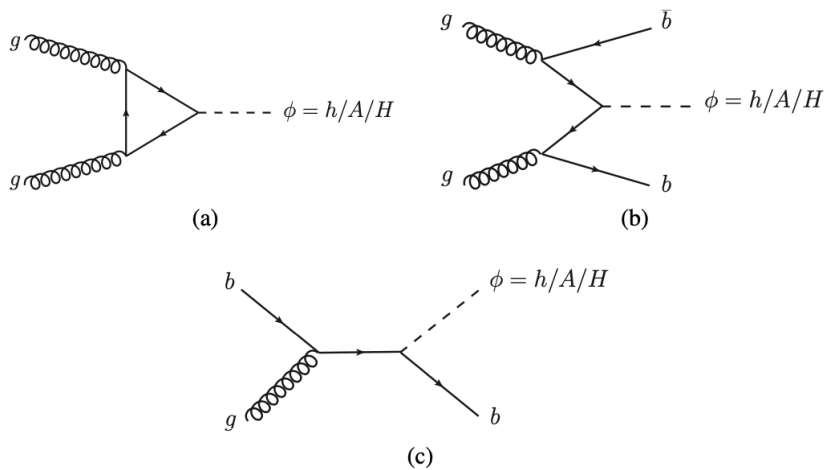


図 1.5: 中性ヒッグス粒子の生成過程

2 LHC-ATLAS 実験

LHC(Large Hadron Collider) は欧州原子核研究機構、通称 CERN で稼働している世界最大エネルギーの大型陽子-陽子衝突型加速器である。LHC を用いた実験はいくつかあるが、その中の ATLAS 実験に参加し研究を行った。この章では LHC と ATLAS 実験について述べる。

2.1 LHC

LHC は 2022 年現在において世界最大エネルギーを実現している、全長 27km の巨大円形加速器である。スイスのジュネーブ郊外に位置し、地下 100m のトンネル内に設置している。図 2.1 のように、陽子ビームの衝突点は 4 箇所ありそれぞれ ATLAS 実験、CMS 実験、LHC-b 実験、ALICE 実験が、独立した実験を行なっている。

加速器の内部で超高真空のビームパイプ内で光速近くまで加速された、反対方向に進む 2 つの陽子を衝突させる。陽子は種々の超伝導電磁石による強力な磁場でリング内を曲げられ走行する。粒子の衝突の可能性を高めるため、衝突の直前には四極超伝導磁石を用いてビームを絞っている。LHC は重心系エネルギー $\sqrt{s}=14\text{TeV}$ 、瞬間ルミノシティ $1.0 \times 10^{-34}\text{cm}^{-1}\text{s}^{-1}$ で設計されており、2015 年から 2018 年に行われた Run2 では重心系エネルギー $\sqrt{s}=13\text{TeV}$ で稼働した。2021 年 3 月現在は 2022 年 5 月より行われる Run3 に向けてロングシャットダウン中である。世界で最も高エネルギーでの陽子-陽子衝突実験を行なっているのが CMS 実験と ATLAS 実験である。

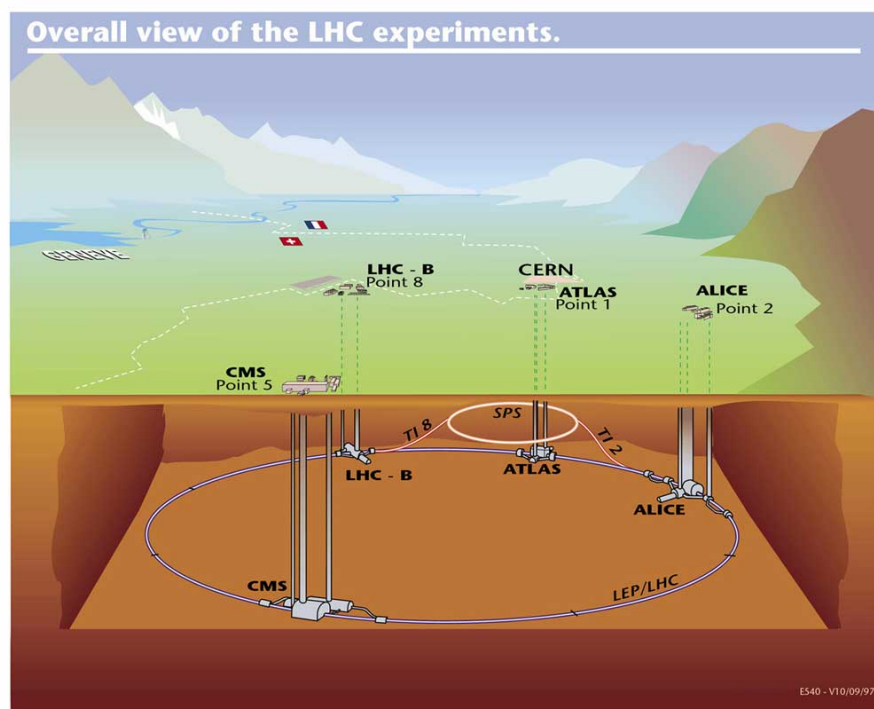


図 2.1: LHC 立体図 (写真提供 CERN アトラス実験グループ) [3]

2.2 ATLAS 実験

CMS 実験と ATLAS 実験は、両実験ともに幅広い物理事象を調べるための汎用検出器を開発し使用している。2つの実験はそれぞれ独立して行われている。このような体制を取っているのは、新しい発見をした際にその結果を相互に確認することが必要不可欠なためである。Run2 の結果、ATLAS 実験では約 139fb^{-1} の積分ルミノシティのデータを物理解析に用いることができるようになった。ATLAS 実験の Run2 における積分ルミノシティの推移を図 2.2 に示す。

ATLAS 実験において重要となるのがパイルアップと呼ばれる、一回のバンチ衝突に際して衝突する事象の数である。これが重要である理由は、興味のある物理事象、例えば標準模型ヒッグスが生成される事象に対して、陽子陽子衝突であるが故の QCD 事象の数の方が圧倒的に多く、この脅威的な数の比が物理探索を非常に困難とさせているからである。Run2 における平均パイルアップ数の変化と分布を図 2.3 に示す。このように一回のバンチ衝突で数十の衝突が起こっており、Run3 や High Luminosity LHC (HL-LHC) にはさらに大きくなることが予想される。

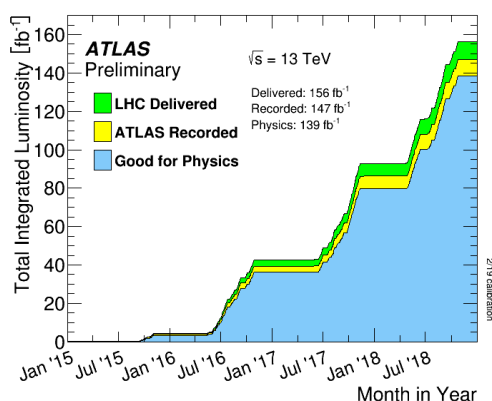


図 2.2: ATLAS 実験 Run2 の積分ルミノシティ [14]

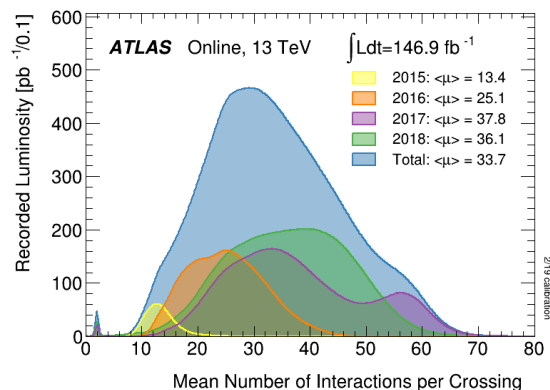


図 2.3: ATLAS 実験 Run2 の平均パイルアップ数分布 [14]

2.3 ATLAS 検出器

ATLAS 検出器の概要と座標系について述べる。ATLAS 検出器の模式図を図 2.4 に示す [13]。ATLAS 検出器は直径 25m、全長 44m の大型検出器であり重さは 7000 トンに上る世界最大容量の粒子検出器である。内側から順に内部飛跡検出器、電磁・ハドロンカロリメータ、ミューオン検出器で構成される。検出器は円柱を横に倒したような形状をしており、これを樽に見立てて側面部分をバレル領域、底面部分をエンドキャップ領域と呼ぶ。

2.3.1 ATLAS 座標系

まず ATLAS 実験において使用される座標系を定義する。原点をビームラインの衝突点の中心にとり、x 軸の正の方向は円形加速器の中心に向かう方向、y 軸の正の方向は地面に垂直な上方にとる。z 軸は右手系をなすようにビームライン状にとる。衝突実験の解析においては (x, y, z) の座標系ではなく (r, η, ϕ) の座標系が一般的であり、 (p_x, p_y, p_z) の代わりに (p_T, η, ϕ) を用いる。ただし p_T は横運動量、 η は擬ラピディティ、 ϕ は方位角である。 p_T と η は次式で定義される。ただし θ は天頂角である。

$$p_T = \sqrt{p_x^2 + p_y^2} \quad (2.1)$$

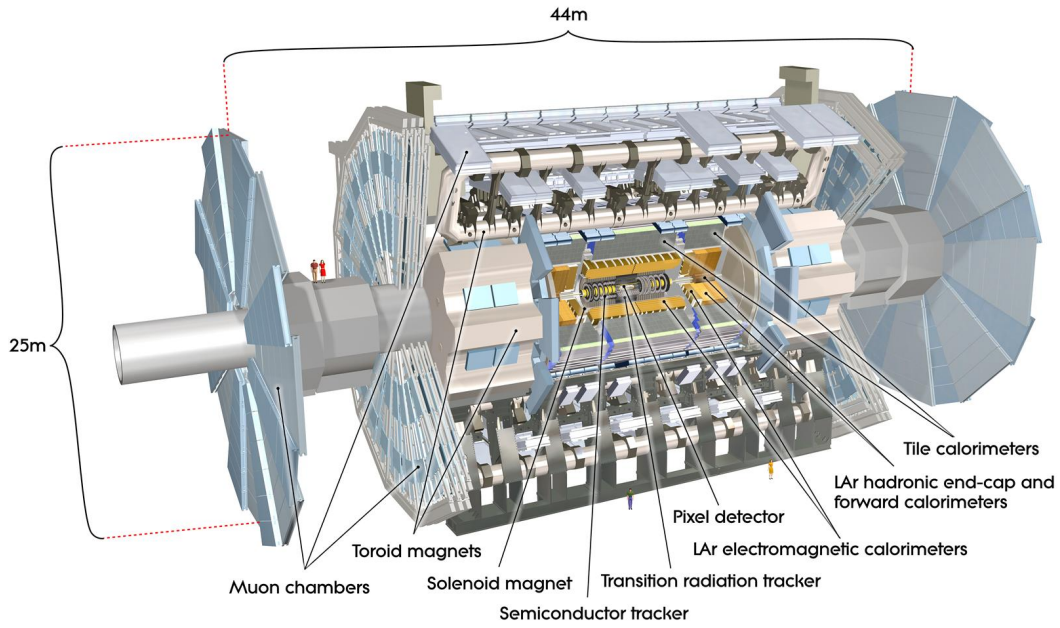


図 2.4: ATLAS 検出器全体図 [13]

$$\eta = -\ln \tan(\theta/2) \quad (2.2)$$

ATLAS 実験の粒子同定や物理解析において重要となるパラメータに、 ΔR という $\eta - \phi$ 角度平面上での距離を表すパラメータがある。ジェット内の飛跡同士の距離やオブジェクト間の距離を表すために用いられ、式 2.3 で定義される。

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \quad (2.3)$$

2.3.2 内部飛跡検出器

内部飛跡検出器は検出器の最内層に位置する検出器の部分である。 $|\eta| < 2.5$ の領域で荷電粒子の飛跡を検出し、各飛跡の運動量や一次衝突点・二次衝突点の位置を精密に測定する。内部飛跡検出器は 3 種類の検出器から構成されており、内側から順に IBL(Insertable B-layer)+Pixel 検出器、SCT(Semi Conductor Tracker)、TRT(Transition Radiation Tracker) である (図 2.5)。それらの全てが超電導ソレノイド磁石で覆われ、2T の磁場がかかっている。

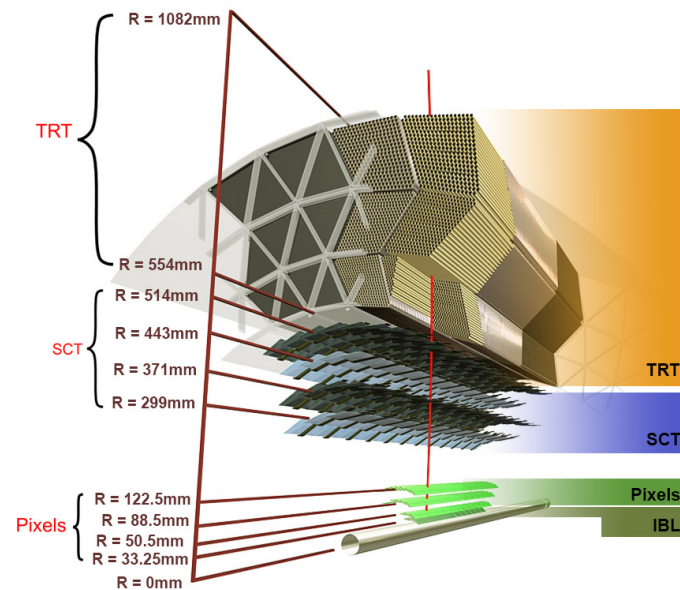


図 2.5: 動径方向から見た内部飛跡検出器 [16]

● IBL, Pixel 検出器

ATLAS 検出器で最もビームパイプの近くに位置する検出器であり、最内層の IBL1 層と Pixel3 層から成る。これらは非常に高い粒度のピクセル状の半導体検出器で構成され、各層は荷電粒子の通過位置を 2 次元で読み出す。IBL は Run2 から追加された層であり、ビーム軸から 33.25mm に位置する。これにより精密で堅牢な飛跡の再構成が可能となっている。

● SCT

ピクセル検出器の外側に位置する検出器であり、バレル領域では 4 層、エンドキャップ領域では 9 層からなる。それぞれはストリップ型の半導体検出器で構成されている。ストリップ検出器は 1 次元の読み出しであるが、それらを並行に並べて構成した 2 枚のセンサーを ϕ 方向に 40mrad ずらして重ねて配置することで、2 次元読み出しを可能としている。よって理想的にはバレル領域でヒットを 8 層に残すこととなる。

● TRT

内部飛跡検出器の最も外側に位置する検出器である。直径 4mm のストロー型気体検出器からなる。内部は X_e , CO_2 , O_2 の混合気体で満たされており、荷電粒子が通過することにより発生した電荷を電気信号として検出する。バレル領域ではビーム中心から 554mm-1082mm の広い範囲を覆っており 1 本の飛跡は平均して 36 のヒットを残し、このおかげで後段で行われる飛跡のパターン認識がより強化される。

2.3.3 カロリメータ

内部飛跡検出器の外側に位置するのがカロリメータである。図 2.6 に示すように電磁カロリメータとハドロンカロリメータの 2 つで構成され、電磁カロリメータのすぐ外側にハドロンカロリメータが位置する。飛跡検出器では検出することのできない中性粒子の検出と、通過した粒子のエネルギーの測定が行われる。カロリメータでは、オブジェクトの正確なエネルギー測定とその外側に位置するミュオン検出器でのミュ

オン測定のために、電磁シャワーやハドロンシャワーを起こす物理オブジェクトのエネルギーを全て落としきることが求められる。エネルギーを落としやすくするため、吸収層は物質量の大きな物質で構成される。

● 電磁カロリメータ

電子、光子の電磁シャワーを利用しエネルギーを測定する。 $|\eta| < 3.2$ の領域を覆っている。吸収層が鉛、検出層が液体アルゴンで構成されるサンプリング型カロリメータである。

内部飛跡検出器を覆っている領域と一致する領域 ($|\eta| < 2.5$) では、両者の情報を合わせて電子や光子の精密な測定を可能とするため、細かい粒度で構成されている。電磁カロリメータはバレル領域では 22 放射長以上、エンドキャップ領域では 24 放射長以上の厚みとなっている。

● ハドロンカロリメータ

ハドロンと原子核の強い相互作用で生じるハドロンシャワーを利用してエネルギーを測定する。ハドロンカロリメータはバレル領域、エンドキャップ領域に加えフォワード領域と呼ばれるよりビーム軸に近い前後方領域も覆っており、 $|\eta| < 4.9$ の広い領域をカバーしている。バレル領域は吸収層に鉄、検出層にタイル状のプラスチックシンチレータを用いている。エンドキャップ領域は吸収層に銅、検出層に液体アルゴンを用いている。フォワード領域は内部飛跡検出器等の容量の関係から限られた領域に設置しなければならず、より密度の大きい物質で構成する必要がある。よって吸収層は最内層が電磁シャワーの測定に対応した銅、その他はタングステンで、液体アルゴンを用いて検出している。

ハドロンカロリメータは電磁カロリメータよりも粗い粒度で構成されているが、ジェットの再構成のための十分な精度を持っている。またハドロンカロリメータまでの全層を合わせて、全領域で約 10 相互作用長の厚みとなっている。

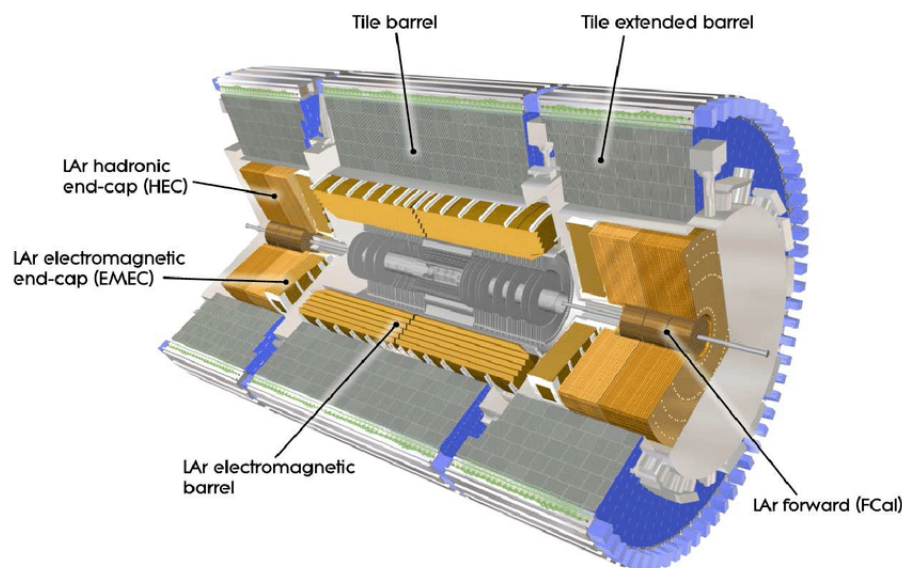


図 2.6: カロリメータ全体図 [13]

2.3.4 ミューオン検出器

ミューオンを検出するための検出器である。電子と比べて物質との相互作用が小さいミューオンを検出するため、ATLAS 検出器の最外層に位置し、 $|\eta| < 2.7$ の領域を覆っている。図 2.7 のようにミューオン検出器は 4 つの部分から構成されており、MDT(Monitored Drift Tubes)、CSC(Cathode Strip Chambers)、RPC(Resistive Plate Chambers)、TGC(Tin Gap Chambers) からなる。MDT と CSC は精密な飛跡測定、RPC と TGC はトリガーのための情報取得を行う。それぞれの覆っている領域は、MDT が $|\eta| < 2.7$, CSC が $2.0 < |\eta| < 2.7$, RPC が $|\eta| < 1.05$, TGC が $1.05 < |\eta| < 2.7$ である。

ミューオン検出器内はバレル領域の大型トロイド磁石とエンドキャップ磁石によって磁場が形成されている。これらの組み合わせにより、ミューオン検出器の全領域において飛跡とほとんど直交する磁場を作ることができる。バレル領域ではビーム軸を中心とする 3 層の円筒形のチェンバーで飛跡が測定される。エンドキャップ領域と、それらの間であるトランジション領域では、ビームに垂直な向きで設置された同じく 3 層のチェンバーで測定される。

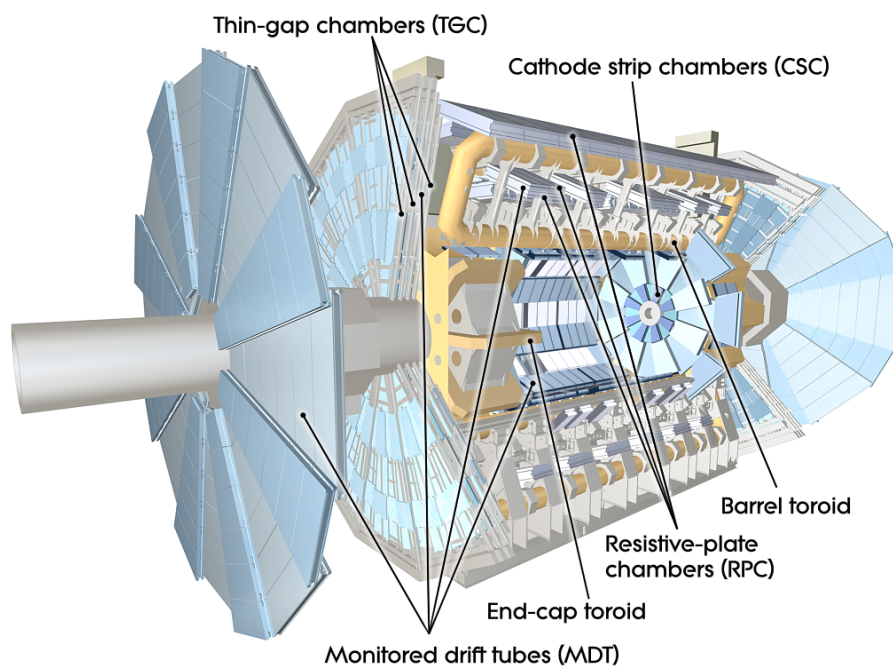


図 2.7: ミューオン検出器全体図 [13]

2.4 トリガーシステム

陽子陽子の衝突は約 40MHz で起こるため、技術的な問題とストレージの容量の問題で全ての事象を記録することは現在の技術では不可能である。よってトリガーシステムを用いて興味のあるイベントを効率的に残すシステムを採用している。対象の物理事象に応じて様々な種類のトリガーを用意している。

ATLAS 実験では 2 段階のトリガーシステムを用いている。1 段階目はハードウェアベースの Level 1 Trigger(L1 Trigger)、2 段階目はソフトウェアベースの HLT(High Level Trigger) である。ATLAS 検出器のトリガーシステムについて図 2.8 に示す。

• L1 Trigger

L1 Trigger では、40MHz のイベントレートを 100kHz まで落とす。1 イベントにつき $2.5\mu\text{s}$ での処理が要求されている。L1 Trigger ではミューオン検出器とカロリメータの情報のみを用いる。これらを用いて高い横運動量を持つミューオン、電子、光子、ジェット、ハドロン崩壊するタウ粒子、大きな MET を探す。ミューオンは RPC と TGC の情報を用いて同定される。カロリメータは、全てのカロリメータの粒度を小さくした情報を用いている。これらを用いて選ばれた結果は中央トリガー処理器にて処理される。L1 Trigger を通過したイベントは各検出器のエレクトロニクスと DAQ システムに渡される。

また各イベントにおいて、L1 Trigger は 1 つ以上の Regions-of-Interest(RoI) という $\eta - \phi$ 座標で記述されるコーン状の領域を定義する。RoI データはその領域が持つ特徴の種類と閾値等の基準に関する情報を含む。この情報は後段の HLT に送られ使用される。

• HLT

HLT では約 100kHz のイベントレートのデータをオンラインで処理する。HLT では L1 Trigger にて定義された RoI に基づいて選別が行われる。これは全イベントデータの約 2% に該当するデータ量である。HLT では初めに専用の高速トリガーアルゴリズムを用いてイベントレートを減らし、その後オフラインに近い再構成を行い最終的な選別を行なっている。HLT では基本的に全ての検出器情報を用いて RoI 内の飛跡の再構成を行い、事象の選別が行われる。HLT で選別されたイベントは結果として平均 1.2kHz であり、それらのデータがストレージに保存される。

2.5 オブジェクトの再構成

ATLAS 検出器を用いて、物理オブジェクトを再構成する手法について述べる。

電子 電子候補は電磁カロリメータに落としたエネルギーと内部飛跡検出器で測定された飛跡を用いて再構成される。検出器の覆う範囲から、 $|\eta| < 2.47$ を満たす必要がある。バレル領域とエンドキャップ領域の間の $1.37 < |\eta| < 1.52$ の領域は除去される。電子候補は、likelihood を用いて同定される。電子の取得効率により *loose, medium, tight* の基準点が設けられている。

ミューオン ミューオン候補は $|\eta| < 2.5$ の領域でミューオン検出器で測定された飛跡と内部飛跡検出器で測定された飛跡をマッチングして再構成される。複数変数を用いて粒子同定され、取得効率により *loose, medium, tight* の基準点が設けられている。

ジェット ジェットはカロリメータにエネルギーを落としたトポロジカルクラスターを用いて再構成される。 $\Delta R = 0.4$ のコーン状で再構成される。パイルアップ由来のジェットを除去するために、飛跡情報と一次衝突点の情報を利用する。

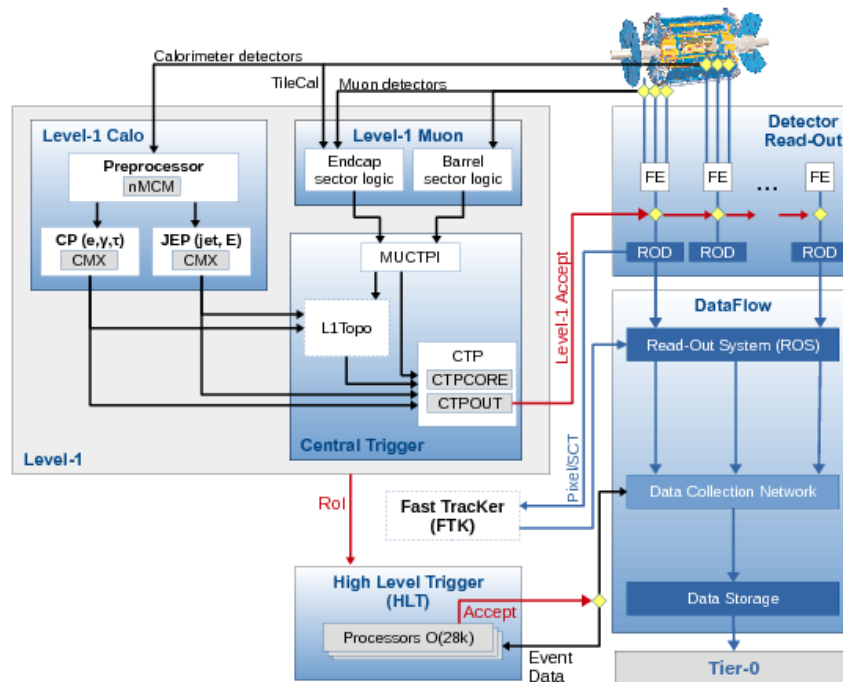


図 2.8: ATLAS トリガーシステム [17]

b -ジェット (b -jet) ジェットの中でも b -hadron 由来のジェットは長寿命であり二次崩壊点を持つという特徴がある。多変量解析を用いた MV2 というアルゴリズムを用いて、ジェットの中から b -jet を同定する。 b -jet の取得効率とそれに伴う背景事象除去効率からいくつかの基準点が設けられている。

タウ粒子 短寿命のタウ粒子は検出器に入る以前にほとんどが崩壊し、その終状態に荷電レプトンを含むか含まないかでレプトン崩壊とハドロン崩壊に大別される。レプトン崩壊するタウ粒子は、電子またはミューオンとして再構成される。ハドロン崩壊するタウ粒子は、ジェットをシードとして再構成される。 $1.37 < |\eta| < 1.52$ を除く $|\eta| < 2.5$ の領域に存在すること、そして荷電飛跡の本数が 1 または 3 であること、電荷が ± 1 であることが要求される。多変量解析を用いて粒子同定され、*loose*, *medium*, *tight* の基準点が設けられている。タウ粒子の再構成、同定手法については後に詳しく述べる。

消失横運動量 (MET) 陽子-陽子衝突型実験では、 z 軸方向の運動量は保存しないが、横方向のエネルギーは正確に保存する。この特性を生かして、検出器と相互作用しないニュートリノ等の横運動量の和を消失横運動量 (Missing E_T) として再構成することができる。あるイベント内において再構成された全てのオブジェクトのベクトル和の、逆ベクトルで計算される。

Overlap Removal いくつかの物理オブジェクトが再構成され同定の基準をクリアしたとき、複数のオブジェクトが空間的に被った位置に存在してしまっていることがある。このとき ΔR で決めた距離内にオブジェクトが重なり合っていた際、どのオブジェクトを優先しそれを排除するかを決め実行する。この基準は解析により多少異なるが、多くの場合ミューオン→電子→タウ粒子→ジェットの順で優先され、優先順位が下となったオブジェクトは除去される。

3 タウ粒子

本章では、本研究対象であるタウ粒子の性質と、その ATLAS 実験における再構成の仕組みについて述べる。

3.1 タウ粒子の性質

タウ粒子は第三世代の荷電レプトンであり、負の電荷を持つ。反粒子は電荷が正でその他は全く同じ性質を持つ。質量は 1776.99 ± 0.12 MeV、平均寿命 τ は 2.90×10^{-13} sec、崩壊長 $c\tau = 87.03 \pm 0.15 \mu\text{m}$ である。タウ粒子には反粒子が存在し、電荷が正でその他の性質は全く同じである。

タウ粒子の崩壊には、大きく分けてレプトン崩壊とハドロン崩壊の2つがある。タウ粒子の下世代である電子またはミューオンに2つのニュートリノを伴って崩壊するモードをレプトン崩壊と呼ぶ。ニュートリノとハドロンに崩壊するモードをハドロン崩壊と呼ぶ。

タウ粒子の電荷は+1 または-1 であり、また電荷は崩壊前後で保存することから、ハドロン崩壊するタウ粒子は荷電粒子が奇数個で崩壊する。その中でも1本の荷電粒子を伴って崩壊することが最も多く、3本、5本... と割合は減っていく。ハドロン崩壊するタウ粒子は、その荷電粒子の個数により 1-prong, 3-prong, ... と呼ばれる。レプトン/ハドロン崩壊や終状態で分類した時のタウ粒子の崩壊分岐比を、表 3.1 に示す。ただし、荷電粒子には $K^\pm$ も含まれるが表中では $\pi^\pm$ で統一して表記している。

表 3.1: タウ粒子の崩壊分岐比

崩壊の種類	終状態	分岐比 [%]
レプトン崩壊	$e\nu\nu$	17.8
	$\mu\nu\nu$	17.4
ハドロン崩壊	$\pi^\pm\nu$	10.8
	$\pi^\pm\pi^0\nu$	25.5
	$\pi^\pm 2\pi^0\nu$	9.3
	$3\pi^\pm\nu$	9.0
	$3\pi^\pm\pi^0\nu$	2.7
	Others	7.5

3.2 タウ粒子の再構成

ATLAS 検出器においてタウ粒子を再構成し粒子の同定を行うまでの一連の流れを説明する。タウ粒子のほとんどは Pixel 検出器の一層目に到達する前に崩壊する。よってレプトン崩壊するタウ粒子は通常の電子あるいはミューオンと区別がつかず、タウ粒子として同定することは不可能である。しかしハドロン崩壊するタウ粒子は QCD からくるジェットに比べて細く、また飛跡数が 1 本または 3 本という特徴があるため、タウ粒子として同定することが可能である。ハドロン崩壊するタウ粒子の終状態には必ずニュートリノを含むがその成分は検出することができない。今後本論文でタウ粒子という表現を用いる場合、断りのない限り、ニュートリノ以外の検出可能な部分で再構成したタウ粒子のことを指す。

タウ粒子のシードジェットの再構成

タウ粒子の候補は、再構成されたジェット中から選ばれる。まずジェットの再構成法について述べる。ジェットは Anti-kt アルゴリズム [38] を用いて $|\Delta R| < 0.4$ で再構成される。カロリメータのセルの 3 次元のクラスターを再構成するトポクラスタリング [39] を行い、ローカルハドロンキャリブレーション (LC) [40] と呼ばれる手法でそのエネルギーを較正する。再構成されたジェットのうち、タウ粒子の候補となるジェットとして $p_T > 10\text{GeV}$ と $|\eta| < 2.5$ を要求する。カロリメータのバレル部分とフォワード部分の隙間である $1.37 < |\eta| < 1.52$ の領域のジェットは除去する。これらを通過したタウ粒子候補のジェットを、シードジェットと呼ぶ。

タウ粒子の再構成

タウジェット独自の生成点の再構成アルゴリズムである Tau Jet Vertex Association (TJVA) が実行される。ここでタウ粒子の再構成において用いられる領域の定義として、コア領域とアイソレーション領域を定義する。シードジェットの向きに対して $\Delta R < 0.2$ の領域をコア領域、 $0.2 < \Delta R < 0.4$ の領域をアイソレーション領域と呼ぶ。コア領域に存在するトラックが 1 本以上のタウ粒子候補に対して TJVA が実行される。タウ粒子の生成点は、そのような条件を満たすジェット内のトラックの運動量の合計が最も大きくなるように選択される。

タウ粒子の向き (η, ϕ 方向) は、上記の手法で再構成されたタウ粒子の生成点を原点とし、シードジェットのコア領域のトポクラスターのベクトル和を用いて計算される。タウ粒子のエネルギーは Tau Energy Scale (TES) と呼ばれる専用の較正アルゴリズムによって計算される。

トラックの分類

タウ粒子の向きから $\Delta R < 0.4$ 内の全てのトラックに対して、多変量解析の 1 つである Boosted Decision Tree (BDT) を用いたトラックの分類手法が実行される。この BDT にはトラックの運動量や各検出器に残したヒット数等の情報が用いられる。この分類の結果、タウ粒子の飛跡数が再構成される。ここではトラックの分類手法について詳細に説明する。

全てのトラックは図 3.1 のように 2 段階の BDT により 4 種類のトラックに分類される。それぞれのトラックの種類は次のようになっている。

タウトラック タウ粒子の直接の崩壊粒子による荷電トラック。

対生成トラック 直接の崩壊粒子のうち π^0 の、 $\pi^0 \rightarrow \gamma(\rightarrow ee)\gamma(\rightarrow ee)$ の反応による電子のトラック。

孤立トラック アイソレーション領域に存在するその他のトラック。

フェイクトラック パイルアップ由来のトラックや、ヒットをランダムに結んだ偽のトラック。

タウジェット内のトラックの分類の結果、タウトラックになったトラックの合計本数が、再構成された飛跡の本数となる。後段のタウ粒子同定においては飛跡数が 1 本と 3 本のものをターゲットに手法が構築されているため、再構成された飛跡数は非常に重要である。

今後表記の簡略化のために、本来の直接の崩壊粒子数が n 個であり、 m 本のタウトラックで再構成されたタウ粒子のことを、「truth n -prong $\rightarrow$ reco m -prong のタウ粒子」と表記する。

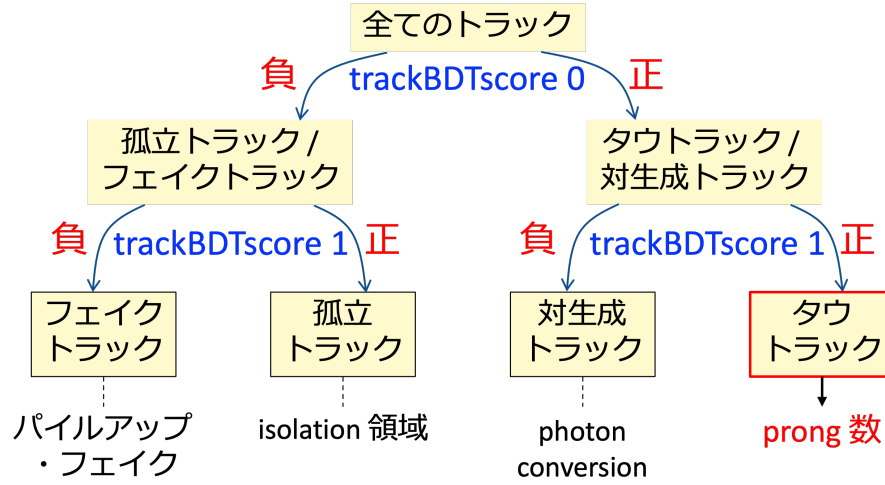


図 3.1: トラックの分類の流れ

1 段階目の BDT では限定的な種類への分類を行っており、タウトラックまたは対生成トラックの種類と、孤立トラックまたはフェイクトラックの種類に分類する。2 段階目で、1 段階目の結果に従ってタウトラックか対生成トラックの分類、孤立トラックかフェイクトラックの分類を行う。本論文では、1 段階目の BDT score を trackBDTscore_0、2 段階目の BDT score を trackBDTscore_1 と呼ぶ。この分類手法では図 3.1 のようにツリー状に 3 つの BDT が存在し、分類の際には BDT score が閾値以上か閾値未満かで分類が決定する。この閾値はどれも 0 にとても近い値となっており、タウ粒子の p_T 等によらず常に一定である。

これらの BDT は、 $\gamma^* \rightarrow \tau\tau$ のモンテカルロシミュレーションサンプルを用いて開発された。BDT の入力変数は次の 13 個である。3 つの BDT は全て同じ入力変数を用いており、学習時に用いる信号事象と背景事象のサンプルの種類分けのみが異なっている。

jetSeed p_T ... シードジェットの横運動量

track η ... トラックの η

$z_0 \sin \theta_{\text{wrt T JVA}}$... T JVA で計算された生成点に対して計算された $z_0 \sin \theta$

dRJetSeedAxis ... シードジェットの向きとトラックの間の ΔR

d_0 ... d_0

qOverP ... 電荷を運動量で割った値

numberOfInnermostPixelLayerHits ... Pixel 最内層 (つまり IBL) のヒット数

numberOfPixelSharedHits ... Pixel 検出器において他のトラックと共有されているヒット数

numberOfSCTSharedHits ... SCT 検出器において他のトラックと共有されているヒット数

eProbabilityHT ... トラックの電子らしさ

nPixHits ... ピクセル検出器におけるヒット数とデッドセンサーの数の和

nSiHits ... ピクセル検出器と SCT 検出器におけるヒット数とそれらのデッドセンサーの数の和

3.3 タウ粒子の同定

ジェットを候補として選択されたタウ粒子候補の中にはタウ粒子だけではなく、QCD によるクォークやグルーオン由来のジェットが多数含まれている。それらの背景事象を効率よく排除しタウ粒子を取得するために、タウ粒子の同定手法が実行される。同定には BDT を用いている。BDT はトラック分類の結果の飛跡数によって、1-prong と multi-prong の 2 つに分けて実行される。再構成飛跡数が 0 本または 1 本のとき 1-prong 用 BDT が、2 本以上のとき multi-prong 用 BDT が実行される。それぞれの手法の入力変数は 1-prong 用 BDT で 9 個、multi-prong 用 BDT で 10 個あり、表 3.2 の通りである。それぞれの変数の説明は付録 B に記載する。

表 3.2: タウ粒子同定 BDT の入力変数

変数	変数名	1-prong 用 BDT	multi-prong 用 BDT
f_{cent}	Central energy fraction	○	○
$f_{\text{leadtrack}}^{-1}$	Leading track momentum fraction	○	○
$R_{\text{track}}^{0.2}$	Track radius	○	○
$ S_{\text{leadtrack}} $	Leading track IP significance	○	-
$f_{\text{iso}}^{\text{track}}$	Fraction of tracks p_T in the isolation region	○	-
ΔR_{Max}	Maximum ΔR	-	○
S_T^{flight}	Transverse flight path significance	-	○
m_{track}	Track mass	-	○
$f_{\text{EM}}^{\text{track-HAD}}$	Fraction of EM energy from charged pions	○	○
$f_{\text{track}}^{\text{EM}}$	Ratio of EM energy to track momentum	○	○
$m_{\text{EM+track}}$	Track-plus-EM-system mass	○	○
$p_T^{\text{EM+track}}/p_T$	Ratio of track-plus-EM-system to p_T	○	○

信号事象は $Z \rightarrow \tau\tau$ のシミュレーションサンプル、背景事象は 2015 年の取得データから QCD ジェットの豊富なデータを用いて BDT は学習されている。1-prong 用 BDT は、truth 1-prong $\rightarrow$ reco 1-prong のタウ粒子と、1-prong に再構成されたジェット事象を用い、multi-prong 用 BDT は、truth 3-prong $\rightarrow$ reco 3-prong のタウ粒子と、3-prong に再構成されたジェット事象を用いて学習されている。BDT による粒子同定の基準点として、*loose*, *medium*, *tight* の 3 種類を用意している。*tight* から順にタウ粒子の取得効率が低く、背景事象の除去効率が低い基準点となる。それぞれの基準点において、タウ粒子取得効率がタウ粒子の p_T に依存しないように設計されている。また BDT の入力変数はパイルアップ数に応じて較正されており、その結果取得効率はパイルアップ数にも依存しない結果となっている。ただしこれらの基準点は、1-prong 用 BDT は truth 1-prong $\rightarrow$ reco 1-prong のタウ粒子に対して、multi-prong 用 BDT は truth 3-prong $\rightarrow$ reco 3-prong のタウ粒子に対して設計されたものである。取得効率は 1-prong と multi-prong で異なり、表 3.3 のようになっている。

表 3.3: BDT によるタウ粒子同定の基準点の取得効率

基準点	1-prong 用 BDT	multi-prong 用 BDT
<i>loose</i>	0.85	0.75
<i>medium</i>	0.75	0.60
<i>tight</i>	0.60	0.45

3.3.1 電子の排除

電子の背景事象を減らすために、再構成された電子から $\Delta R < 0.4$ の距離に存在する 1-prong のタウ粒子候補に対して電子排除が実行される。排除される条件は、その電子の同定手法である likelihood score が、基準点 *veryloose* を満たすことである。この likelihood score によるカットは、タウ粒子の p_T と η に対して 95% の効率でフラットに取得できるように調整されている。

3.4 タウ粒子トリガーシステム

タウ粒子を含むイベントは標準模型の精査や新物理の探索において非常に重要なイベントの一種であるため、探索する物理事象に応じてタウ粒子に関連するさまざまなトリガーを用意している。

3.4.1 L1 tau Trigger

L1 Trigger ではカロリメータの情報をもとにタウ粒子候補を同定する。EM カロリメータとハドロンカロリメータの両方で、コア領域とアイソレーション領域の 2 つの領域が定義される。L1 のカロリメータでは粒度は $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ で $|\eta| < 2.5$ の領域が対象である。タウ粒子の横エネルギーの和である E_T は、EM カロリメータのコア領域で最もエネルギーを落とした 2 つの領域とハドロンカロリメータの 2×2 の領域のエネルギーの和で計算される。この値は電磁スケールで較正される。また、アイソレーション領域で EM カロリメータに落としたエネルギー E_T^{EMisol} は EM カロリメータにおいて $\eta \times \phi = 0.2 \times 0.2 - 0.4 \times 0.4$ の間の領域に落としたエネルギーから計算される。背景事象を抑えつつ高い効率でタウ粒子を同定するため、60GeV 以下のタウ粒子候補に対して $E_T^{\text{EMisol}} \leq (E_T/10 + 2)$ のカットを適用する。60GeV 以上の候補に対してはアイソレーション領域のエネルギーに関する要求はしていない。

3.4.2 HLT tau Trigger

HLT のタウトリガーは 3 つのステップに分かれている。1 つ目がカロリメータオンリープリセクション、2 つ目がトラックプリセクション、3 つ目がオフラインライクセクションだ。各ステップにおいて異なるタウ粒子候補を選別していき、後段にいくほど精密な再構成が行われる。

カロリメータオンリープリセクション

カロリメータの情報のみでタウ粒子を再構成する。L1 で定義された RoI 内のカロリメータのセルの情報が読み込まれ、トポクラスタリングアルゴリズムが実行される。カロリメータ情報は最大粒度の情報を用いておりバンチ衝突ごとにパイルアップ補正もかけているため、再構成されるエネルギーの値の精度はオフラインに近くなっている。これらのクラスターのエネルギーはローカルハドロンキャリブレーション (LC) によって較正され、それらのベクトル和がタウ粒子の候補のシードジェットとして用いられる。タウ粒子のエネルギーは、シードジェットのコア領域内の LC クラスターのエネルギー和で計算される。タウ粒子に特化したエネルギー較正もこの段階で行われるが、用いることのできる情報が限られているためこの精度はオフラインと比べて劣っている。以上の再構成が行われたのちに p_T の閾値を通過したタウ粒子候補が次のステップへ進む。

トラックプリセクション

飛跡の情報もタウ粒子の再構成に用いられる。飛跡の再構成は、2 段階に分割された独自の高速飛跡再構成アルゴリズムを用いて行われる。高速飛跡再構成がシードジェットを中心とした領域で実行される。1 段階目では、 $p_T > 1$ GeV を満たす最も高い運動量を持つ飛跡 (leading track) をタウ粒子候補の細いコーン内で探す。それが見つからなかった場合、タウ粒子候補は除去される。2 段階目では RoI の中心から $\Delta R < 0.4$ の範囲内で leading track (最も $p_T > 1$ の高いトラック) と、 $|z_0|$ の近い領域でタウ粒子候補に付随する他の

飛跡を探す。2段階の高速飛跡再構成の後、 $p_T > 1$ GeV のトラック数を数え上げ、タウ粒子候補のコア領域の飛跡の本数 $N_{\text{core}}^{\text{trk}}$ とアイソレーション領域の飛跡の本数 $N_{\text{isol}}^{\text{trk}}$ をそれぞれ算出する。 $1 \leq N_{\text{core}}^{\text{trk}} \leq 3$ かつ $N_{\text{isol}}^{\text{trk}} \leq 1$ を満たすタウ粒子候補のみが HLT tau Trigger の最終ステップに進む。

オフラインライクセレクション

最終段階ではタウ粒子候補に付随するトラックをより正確に測定し、3.3節で説明したオフラインと同様のBDTを用いた同定アルゴリズムを適用する。前段階の高速飛跡再構成で特定されたトラックをシードとして、オフラインと同等の精度の飛跡再構成アルゴリズムが実行される。それにより測定された飛跡とカロリメータの情報を用いてタウ粒子同定のBDTのための入力変数が計算される。オンラインとオフラインで大きく異なる変数が $|S_{\text{leadtrack}}|$ である。オフラインではタウ粒子候補に関連する衝突点が求められこの値が計算されるが、オンラインではビームスポットの位置に対して計算される。また各入力変数がパイルアップ数によって較正された値を用いる点は同様だが、オンラインのパイルアップ数によって計算されている点で異なる。計算されたBDT scoreによりタウ粒子が同定され、各トリガーメニューの選別で用いられる。

HLT のタウトリガーの性能は、 $Z \rightarrow \tau\tau$ とダイジェットのシミュレーションサンプルを用いて、真のタウ粒子とオフラインでタウ粒子の同定基準を満たしたものをを用いて評価される。オフラインで1-prong または3-prong のタウ粒子に再構成され、HLT において p_T の基準と飛跡数の基準を満たすタウ粒子候補に対するBDTアルゴリズムの性能を図3.2に示す。タウ粒子同定の基準点として、タウ粒子の取得効率の高い方から順に *loose*, *medium*, *tight* の3点を示している。

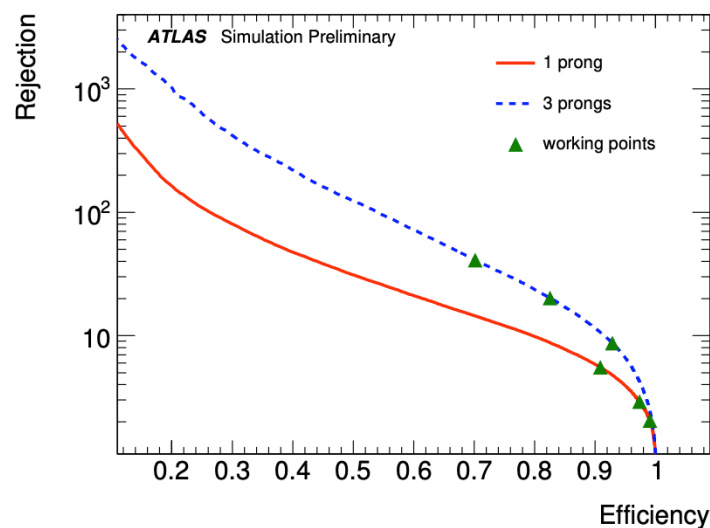


図 3.2: Run2 における HLT tau trigger の性能 [17]

タウ粒子を用いたトリガーは解析の対象となる物理事象に応じて様々に用意しており、1つの高運動量タウ粒子を要求するもの、2つのタウ粒子を要求するもの、1つのタウ粒子に加えて他の粒子やMETを要求するもの等がある。

4 ATLAS 実験における $H/A \rightarrow \tau\tau$ 過程の探索

ここでは、ATLAS の $H/A \rightarrow \tau\tau$ の現行解析について簡単にまとめる。現行解析は Run2 の全データ 139fb^{-1} を用いて行われおり、詳しい情報は文献 [27], [28], [29] にまとめてある。この解析で探索している質量領域では、H ボソンと A ボソンの質量差は実験分解能よりも極めて小さく、質量は縮退しているものとして扱っている。

4.1 解析の概要

解析チャンネルは、タウ粒子の崩壊のモードにより 2 つのチャンネルに分かれており、 $\tau_{\text{lep}}\tau_{\text{had}}$ チャンネルと $\tau_{\text{had}}\tau_{\text{had}}$ チャンネルを考慮している。タウ粒子の崩壊分岐比を考慮すると、2 つのタウ粒子の崩壊モードのうち $\tau_{\text{lep}}\tau_{\text{had}}$ チャンネルは全体の約 46% を、 $\tau_{\text{had}}\tau_{\text{had}}$ チャンネルは約 42% をカバーする。また各チャンネルで、b-veto カテゴリーと b-tag カテゴリーを定義する。b-veto カテゴリーではグルーオンフュージョンで生成されたヒッグス粒子を、b-tag カテゴリーでは b クォークを伴って生成されたヒッグス粒子を目標に探索をする。

各チャンネルで用いたトリガーを表 4.1、各チャンネルにおける信号領域の事象選択を表 4.2 にまとめる。 $\tau_{\text{lep}}\tau_{\text{had}}$ チャンネルではシングルエレクトロントリガーあるいはシングルミューオントリガー、 $\tau_{\text{had}}\tau_{\text{had}}$ チャンネルではシングルタウトリガーが用いられている。ただし、事象選択中に用いられている変数である横質量 $m_{\text{T}}(l, E_{\text{T}}^{\text{miss}})$ の定義は次の式 4.3 である。

$$m_{\text{T}}(l, E_{\text{T}}^{\text{miss}}) = \sqrt{2p_{\text{T}}^l E_{\text{T}}^{\text{miss}} [1 - \cos\Delta\phi(p_{\text{T}}^l, E_{\text{T}}^{\text{miss}})]} \quad (4.1)$$

表 4.1: $H/A \rightarrow \tau\tau$ 現行解析の各チャンネルで用いたトリガー

ピリオド	$e\tau_{\text{had}}$ チャンネル	$\mu\tau_{\text{had}}$ チャンネル	$\tau_{\text{had}}\tau_{\text{had}}$ チャンネル
2015 年	e24_lhmedium_L1EM20VH, e60_lhmedium, e120_lhloose	mu20_iloose_L1MU15, mu50	HLT_tau80_medium1_tracktwo_L1TAU60, HLT_tau125_medium1_tracktwo, HLT_tau160_medium1_tracktwo
2016 年	e26_lhtight_nod0_ivarloose, e60_lhmedium_nod0, e140_lhloose_nod0	mu26_ivarmedium, mu50	
2017 年			HLT_tau160_medium1_tracktwo, HLT_tau160_medium1_tracktwo_L1TAU100
2018 年			HLT_tau160_medium1_tracktwoEF_L1TAU100

各チャンネルにおいて、b-tag された $p_{\text{T}} > 20\text{GeV}$, $|\eta| < 2.5$ を満たす b-jet が 1 個以上あるイベントが b-tag カテゴリー、0 個のイベントが b-veto カテゴリーに分類される。タウ粒子の崩壊モードと b-tag による 4 つのカテゴリーはいずれも直交する領域となる。

$\tau_{\text{lep}}\tau_{\text{had}}$ チャンネルにおける主要な背景事象は、ジェットが τ_{had} にフェイクすることにより生じる。これらの背景事象は大きく 2 つに分けられ、レプトンが終状態に含まれる $W+\text{jet}$ 事象や $t\bar{t}$ 事象や、マルチジェット事象がレプトンに誤同定されてしまう事象がある。このような理由により生じる背景事象の数はモンテカルロシミュレーションで正確に見積もることが困難であるため、データを用いて見積もられる。

$\tau_{\text{had}}\tau_{\text{had}}$ チャンネルにおける主要な背景事象はマルチジェット事象であり、この事象数もデータを用いて見積もられる。マルチジェット事象以外で無視できない背景事象として、b-veto カテゴリーでは $Z/\gamma^* \rightarrow \tau\tau$ 事象、b-tag カテゴリーでは $t\bar{t}$ 事象がある。これらを含むその他の背景事象はモンテカルロシミュレーションにより事象数を見積もる。

系統誤差は、積分ルミノシティ、検出器シミュレーション、生成断面積の理論値、背景事象のモデリングを考慮して見積もられる。主要な系統誤差を表 4.3 に示す。質量 400GeV の点では系統誤差は約 50% で

表 4.2: $H/A \rightarrow \tau\tau$ 現行解析の信号領域の事象選択

対象	$e\tau_{\text{had}}$ チャンネル	$\mu\tau_{\text{had}}$ チャンネル	$\tau_{\text{had}}\tau_{\text{had}}$ チャンネル
τ_{had}	1 ≤ 個 $p_T \geq 25 \text{ GeV}, \eta < 2.3,$ ID medium	1 ≤ 個 $p_T \geq 25 \text{ GeV}, \eta < 2.3,$ ID medium	2 ≤ 個 $\tau_{\text{had}1} : p_T \geq 85 \text{ GeV}, \eta < 2.3$ ID medium, トリガーマッチ $\tau_{\text{had}2} : p_T \geq 65 \text{ GeV}, \eta < 2.3$ ID loose
e	=1 個 $p_T > 30 \text{ GeV}$, トリガーマッチ	0 個	0 個
μ	0 個	=1 個 $p_T > 30 \text{ GeV}$, トリガーマッチ	0 個
dilepton veto	$p_T > 15 \text{ GeV}$ ID loose が 2 個以上	$p_T > 7 \text{ GeV}$ ID loose が 2 個以上	-
電荷	$q_e \times q_{\tau_{\text{had}}} < 0$	$q_\mu \times q_{\tau_{\text{had}}} < 0$	$q_{\tau_{\text{had}1}} \times q_{\tau_{\text{had}2}} < 0$
$\Delta\phi(\tau_1, \tau_2)$	> 2.4	> 2.4	> 2.7
$m_{vis}(l, \tau_{\text{had}})$	$80 <, < 110$	-	-
$m_T(l, E_T^{\text{miss}})$	< 40	< 40	-

あるが、高い質量点ほど小さい誤差となっている。系統誤差と統計誤差の大きさを比較したところ、質量 600 GeV 以上の領域においては統計誤差が支配的となっている。

表 4.3: $H/A \rightarrow \tau\tau$ 現行解析の主要な系統誤差

原因	ggF(400GeV)	ggF(1TeV)	bbH(400GeV)	bbH(1TeV)
タウ粒子同定効率	0.14	0.16	0.12	0.08
Tau Energy Scale	0.33	0.09	0.22	0.03
$Z + jet$ 背景事象のモデリング	0.27	0.19	0.08	0.04
フェイクタウ粒子による背景事象	0.22	0.01	0.14	0.02
その他	0.09	0.04	0.11	0.02
合計	0.54	0.28	0.45	0.13

最終的な感度は、式 4.2 で定義される全横質量 m_T^{tot} 分布から binned likelihood 関数によって求められる。ただし式中の τ_1 と τ_2 は、 $\tau_{\text{lep}}\tau_{\text{had}}$ チャンネルでは電子もしくはミューオンと τ_{had} 、 $\tau_{\text{had}}\tau_{\text{had}}$ チャンネルでは 2 つの τ_{had} を意味する。

$$m_T^{\text{tot}} = \sqrt{m_T^2(E_T^{\text{miss}}, \tau_1) + m_T^2(E_T^{\text{miss}}, \tau_2) + m_T^2(\tau_1, \tau_2)} \quad (4.2)$$

ここで $m_T(a, b)$ は、式 4.3 で定義される。

$$m_T(a, b) = \sqrt{2p_T(a)p_T(b)(1 - \cos\Delta\phi(a, b))} \quad (4.3)$$

4.2 $H/A \rightarrow \tau\tau$ の探索結果

Run2 の全データ 139fb^{-1} を用いて $H/A \rightarrow \tau\tau$ 事象を探索した結果、期待される標準模型の背景事象数を上回る有意な信号の超過は見られなかった。そのため中性スカラーヒッグス粒子 (ϕ) の生成断面積 $\times \tau\tau$ 対への崩壊分岐比の 95%信頼区間の上限値を、ヒッグス粒子の質量の関数として設定した。ヒッグス粒子の質量は、0.2-2.5 TeV までを考慮している。

ggF, bbH の生成断面積 $\times \tau\tau$ 対への崩壊分岐比の制限結果は図 4.1 の (a), (b) である。観測されたデータによる上限値は、 $m_\phi = 1 \text{ TeV}$ で ggF で 1.8 fb, bbH で 1.1 fb である。さらにこの結果を MSSM の制限値と

して解釈し、 m_A と $\tan\beta$ の二次元平面での制限結果を示したものが図 4.1 の (c) である。観測されたデータにより、 $m_A = 1.0$ TeV で $\tan\beta > 8$ を、 $m_A = 1.5$ TeV で $\tan\beta > 21$ を 95%信頼区間で棄却している。

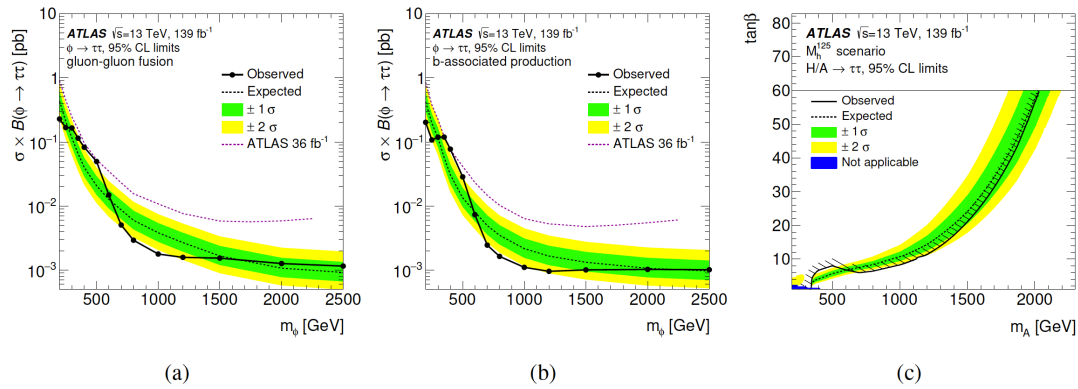


図 4.1: $H/A \rightarrow \tau\tau$ 解析の制限結果 [28]

4.3 $H/A \rightarrow \tau\tau$ の解析感度向上のための試み

$H/A \rightarrow \tau\tau$ 探索の感度をより上げるために可能な試みについて述べる。

- 統計量を増やす。

解析に使用することのできるデータ量が大きくなれば、その分感度を伸ばすことができる。簡単に考えると統計量を N 倍に増やすことができると感度は $\sqrt{N}$ 倍良くなる。Run3 では約 160fb^{-1} のデータ量の取得を目指しており、また HL-LHC までに全データ量は 3000fb^{-1} に上る見込みである。よって単純計算では、データ量は約 21.6 倍となり約 4.65 倍の感度向上が見込まれる。

- タウ粒子同定効率の系統誤差を小さくする。

今後の探索でより重要となってくるのは高質量の領域である。高質量領域における系統誤差の多くはタウ粒子同定効率が占めている。統計量の伸びに応じて $H/A \rightarrow \tau\tau$ の探索感度をあげていくためには、この系統誤差を小さくする必要がある。そのためには、高運動量領域のタウ粒子の実データを用いて、精度良く測定を行うことが重要である。

- 信号のアクセプタンスを上げる。

多くの新物理探索は信号事象の生成断面積が標準模型背景事象と比較して小さいことが、信号の発見を困難としている。そのため信号事象のアクセプタンスを上げ、なるべく多くの信号を解析できるようにすることが重要である。 $H/A \rightarrow \tau\tau$ の信号事象のアクセプタンスを上げるための一つの方法として、タウ粒子の prong 数の条件を緩めることが考えられる。現行解析では 1-prong と 3-prong のタウ粒子しか用いていないが、2-prong のタウ粒子を用いることでアクセプタンスを向上させることができる。

- 背景事象を効率よく除去する。

信号事象のアクセプタンスを増やすと、それに伴って背景事象も増える。感度を伸ばすためには信号事象のアクセプタンスを増やした分以上に背景事象を除去する必要がある。現行解析では信号事象と背景事象の分離をカットベースで行っている。これを機械学習や深層学習等を用いて分離を試みることで、より効率よく背景事象を識別し感度を向上させることができる可能性がある。

4.4 2-prong のタウ粒子を使用した解析の検討

4.4.1 検討手法

信号事象のアクセプタンスを上げるために 2-prong のタウ粒子を使用することについて検討する。Run2 の 2018 年の取得データ 58.45fb^{-1} 用のモンテカルロシミュレーションサンプルを用いて、従来解析の信号領域の取得事象数と 2-prong のタウ粒子を用いたときの取得事象数を比較する。今回はグルーオンフュージョンで生成する中性ヒッグス粒子が (図 1.5 (a))、2 つのハドロン崩壊するタウ粒子に崩壊する過程についてシミュレーションを行なった。b-jet が伴う生成過程では b-jet のタウ粒子へのフェイクが考えられるためその影響を排除して考えることができ、また 2-prong のタウ粒子を用いることによる取得効率の向上率が高いと考えられるため、この過程を選択した。使用した信号事象のモンテカルロシミュレーションサンプルの質量は、600, 1000, 2000, 2500 GeV の 4 つである。

2-prong のタウ粒子を解析で使用する際に留意する点が 2 つある。まず、2-prong のタウ粒子に特化した粒子同定アルゴリズムが解析ソフトウェア内で未実装である点である。よって本研究では 2-prong のタウ粒子の粒子同定効率を仮定し、event weight にその効率をかけることで事象数の見積もりを行った。現在 ATLAS 実験の他研究グループにおいて 2-prong のタウ粒子に特化した粒子同定アルゴリズムの開発が行われており、その性能を図 4.2 に示す [41]。

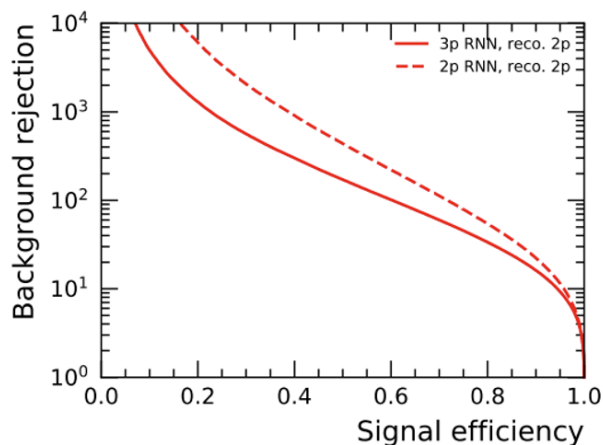


図 4.2: 2-prong のタウ粒子の同定性能

横軸を信号取得効率、縦軸を背景事象除去率とした図の場合、曲線が右上であるほど性能が良い識別器である。1-prong のタウ粒子、3-prong のタウ粒子の性能曲線を 2-prong のタウ粒子の性能と比較したところ、性能は $1\text{-prong} < 2\text{-prong} < 3\text{-prong}$ の順になっており、また背景事象の数は $1\text{-prong} > 2\text{-prong} > 3\text{-prong}$ の順で多くなっている。つまり、2-prong のタウ粒子の粒子同定効率は 1-prong と 3-prong の中間に設定することが妥当であると考えられる。表 3.3 の同定効率の基準点の取得効率を踏まえると、2-prong のタウ粒子の基準点を *loose* で 0.80, *medium* で 0.70 と仮定する。本来はタウ粒子の横運動量やパイルアップ数に依らず一定の同定効率となるような補正が行われるが、今回はそのような影響を無視して見積もっている。

2 つ目は電荷の識別ができない場合があることである。2-prong のタウ粒子の中には、2 本の飛跡の電荷の和が 0 となりタウ粒子の電荷がわからないことがある。本研究では、電荷の和が 0 の場合の電荷識別手法の開発を行なったが、複数の解析手法の検討のため、電荷が ± 2 で再構成された 2-prong のタウ粒子を含む場合と電荷が 0 で再構成された 2-prong のタウ粒子を含む場合を分け、それぞれを従来解析に追加した際に見込まれる事象数を見積もった。

2つのタウ粒子の再構成飛跡数と電荷情報により、信号事象を図 4.3 に示すように 3 種類に分類する。1-prong と 3-prong のタウ粒子のみを用いる「従来解析」、電荷が ± 2 の 2-prong のタウ粒子を含む「追加可能」、電荷が 0 の 2-prong のタウ粒子を含む「検討可能」の 3 種類に分類した。OS は Opposite Sign : 反対電荷 の要求を行うことを示しており、「従来解析」と「追加可能」にはこれを行い、「検討可能」には OS の要求をしていない。

$\tau_{\text{had } 1} \backslash \tau_{\text{had } 2}$		1-prong / 3-prong	2-prong	
			電荷 = ± 2	電荷 = 0
1-prong / 3-prong		OS 従来解析	OS 追加可能	- 検討可能
2-prong	電荷 = ± 2	OS 追加可能	OS 追加可能	- 検討可能
	電荷 = 0	- 検討可能	- 検討可能	- 検討可能

OS : 反対電荷

図 4.3: 信号事象の飛跡数と電荷情報による分類

4.4.2 信号領域の事象選択

4.1 章にて述べた先行研究の事象選択と基本的には同一としているが一部に変更があるため改めて表 4.4 に記す。なお 1-prong, 3-prong のタウ粒子には粒子同定を要求しているが、2-prong に対しては event weight に仮定した同定効率をかけることで事象数を見積もる。1 番目のタウ粒子が 2-prong であった場合は medium ID なので 0.70 をかけ、1 番目は 1-prong または 3-prong で 2 番目が 2-prong であった場合は loose ID なので 0.80 をかける。

表 4.4: 信号領域の事象選択

対象	$\tau_{\text{had}}\tau_{\text{had}}$ チャンネル
トリガー	シングルタウトリガー
$\tau_{\text{had}1}$	$p_T > 85\text{GeV}$, ID medium
$\tau_{\text{had}2}$	$p_T > 55\text{GeV}$, ID loose
e	$p_T > 15\text{GeV}$, ID loose
μ	$ \eta < 2.5$, ID loose
電荷	$q_{\text{had}1} \times q_{\text{had}2} \leq 0$
$\Delta\phi(\tau_{\text{had}1}, \tau_{\text{had}2})$	> 2.7

4.4.3 信号領域の事象数

事象選択を通過した事象を 3 種類に分類したときのそれぞれの事象数を表 4.5 にまとめる。「追加可能」のみを加えた場合と「追加可能」と「検討可能」の 2 つを加えた場合に、従来解析と比較して取得事象数の増加した割合を図 4.4 に示す。

「追加可能」により約 10% の事象数増加が見込まれる。「検討可能」を含めるとさらに多数の事象数増加が見込まれるが、これらは電荷の識別ができていないため信号感度を高められない可能性がある。これら

の事象をより効果的に解析に用いるために、電荷 0 で再構成された 2-prong のタウ粒子の電荷識別手法を開発する。

表 4.5: 3 種類の信号領域の事象数

種類	600 GeV	1000 GeV	2000 GeV	2500 GeV
従来解析	6639.7	480.10	3.3609	0.5279
追加可能	398.1	39.94	0.3930	0.0676
検討可能	689.7	69.72	0.8067	0.1412

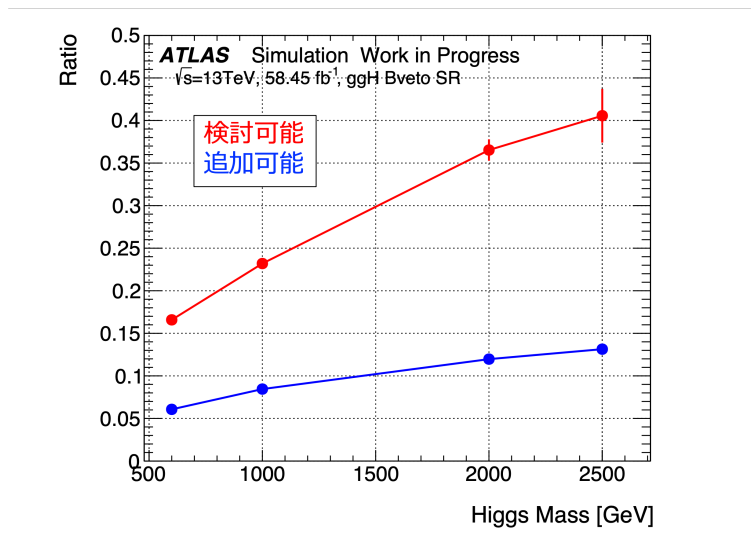


図 4.4: 信号領域の事象数の増加割合

5 2-prong のタウ粒子の真の電荷の識別

5.1 再構成される飛跡の本数とタウ粒子の電荷

まずは、現状のタウ粒子の再構成の性能について確認する。性能として今回重要視する点は、再構成される飛跡の本数と再構成されるタウ粒子の電荷の2点とする。

再構成される飛跡の数は、トラックの分類の結果タウトラックに分類されたトラックの合計本数である。まず、truth 1-prong と truth 3-prong の場合に分け、それぞれにおいて何本の飛跡に再構成されているのかを調べる。図 5.1 に再構成されるタウ粒子の横運動量の関数として、タウ粒子の飛跡再構成数の割合を示す。

truth 1-prong の場合は、1-prong で再構成される割合が80%以上を保っている。次に多いのが0-prong、その次が2-prong となっている。1-prong で再構成される割合は高運動量ほど下がっており、代わりに0-prong や2-prong が増えている。

一方で truth 3-prong の場合は、低運動量領域では3-prong での再構成が最も多いが、高運動量領域でその割合が急減し、700GeV 以上では2-prong での再構成が最も多くなっている。2-prong と同様に高運動量領域ほど1-prong, 0-prong での再構成の割合も増加している。

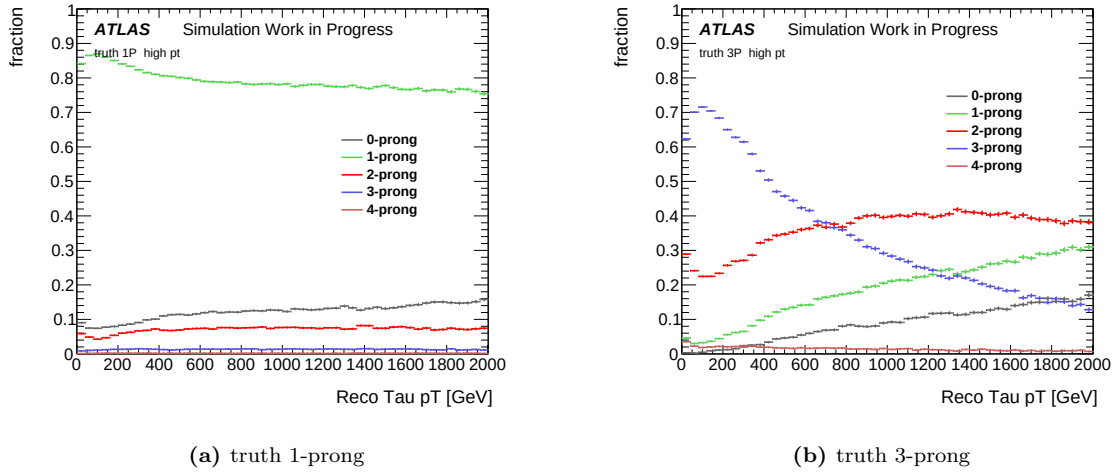


図 5.1: 再構成される飛跡の数の割合

次に、これらの特徴をより詳しく調べるために、崩壊時に伴う π^0 の個数を元に定義した崩壊モードによって再構成される prong 数が異なっているかを調べた。崩壊モードの定義は表 5.1 の通りである。なお、真の飛跡数が5本以上のタウ粒子は多く存在しないため、本研究では考えない。以上の定義に基づいた5つの崩壊モードごとに再構成される飛跡の数をまとめたものが図 5.2 である。左上から右上にかけて 1p0n, 1p1n, 1pXn, 下段左から 3p0n, 3pXn となっている。

表 5.1: 崩壊モードの定義

崩壊モード名	真の飛跡数	伴う π^0 の数
1p0n	1	0
1p1n	1	1
1pXn	1	$2 \leq$
3p0n	3	0
3pXn	3	$1 \leq$

truth 1-prong の場合を見てみると、 π^0 が 0 個の場合に 1-prong で再構成される割合が最も高い。2-prong 以上で再構成されることは数 % 以下で、1-prong 以外の場合ほとんどが 0-prong での再構成となっている。またタウ粒子の p_T が高いほどその割合は高くなっている。次に π^0 を伴う場合を見てみると、1-prong での再構成率が減り、2-prong での再構成の割合が増えている。1p0n と同様にタウ粒子の p_T が高いほど 0 または 2-prong の割合が増えているが、その傾向は 1p0n ほど顕著ではない。1p0n、1p1n、1pXn を比較すると 1-prong での正しい再構成の割合は π^0 の数が増えるほど減っており、2-prong での再構成割合が増えている。これらの状況を踏まえると、 π^0 の典型的な崩壊である $\pi^0 \rightarrow \gamma(\rightarrow e^+e^-)\gamma(\rightarrow e^+e^-)$ の終状態の電子のうちの 1 本を、タウトラックとして誤分類してしまった結果、2-prong のタウ粒子で再構成されたと考察できる。

truth 3prong の場合 2-prong で再構成される原因は主に 2 つある。

1. 低い p_T を持つ 1 本のタウトラックが、タウトラックとして認識されないこと
2. 高運動量領域で、2 本のタウトラック同士の距離が近いいため内部飛跡検出器のヒットを共有し、1 本として再構成されてしまうこと

であると考えられる。

以上をふまえ、2-prong のタウ粒子のイメージを図 5.3 に示す。

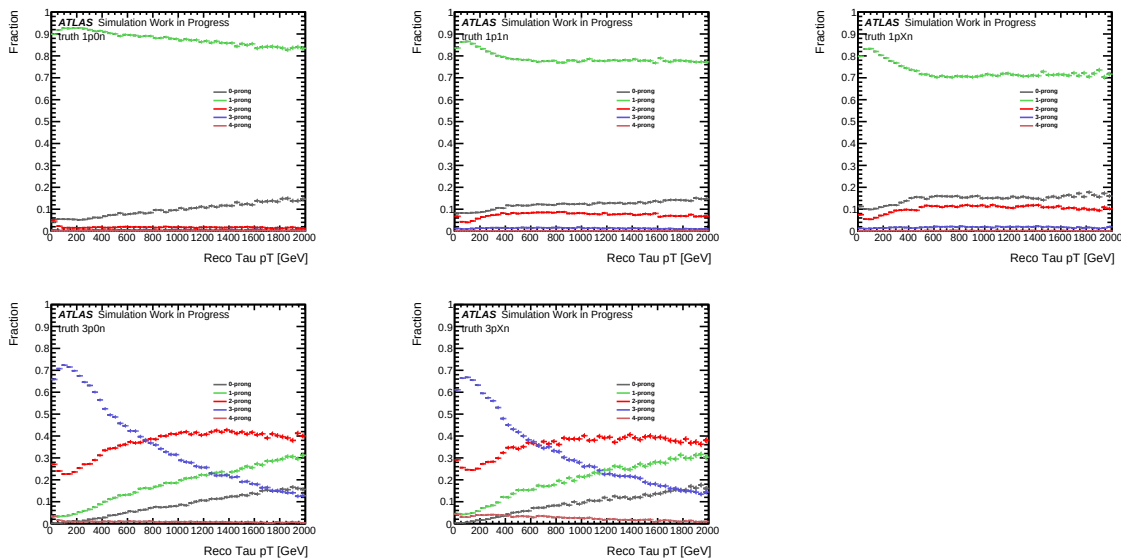


図 5.2: 崩壊モードごとに分けた、再構成される飛跡の数の割合

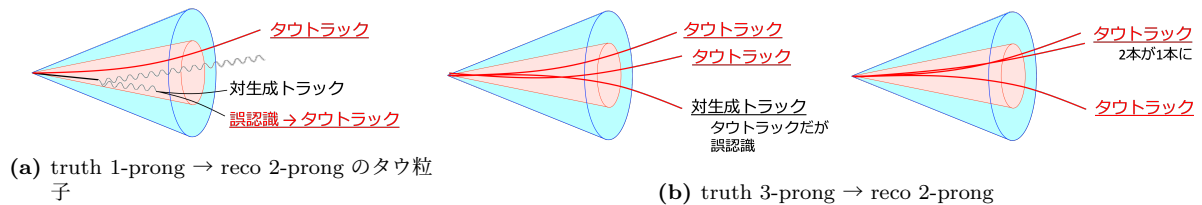


図 5.3: 2-prong のタウ粒子のイメージ図のタウ粒子

次に再構成されるタウ粒子の電荷について述べる。タウ粒子の電荷は、タウトラックの電荷の総和で再構成される。まず、通常の物理解析で標準的に用いられている 1-prong と 3-prong のタウ粒子の再構成される

電荷をタウ粒子の p_T の関数として調べた正解率を図 5.4a と図 5.4b に示す。ここで、3-prong の場合再構成される電荷は+3, +1, 0, -1, -3 の 4 パターンあるが、正解としているのは、+1 と-1 に再構成されている場合のみとしている。これは、通常の物理解析で電荷が ± 1 のタウ粒子のみを用いているからである。共に 100GeV 以下程度の領域では約 98%程度という高い識別効率を達成している。高運動量領域では識別効率は下がり、その下がり方は truth 3-prong $\rightarrow$ reco 3-prong のタウ粒子の方が大きい。これは 3 本の飛跡が存在しているためであると考えられる。

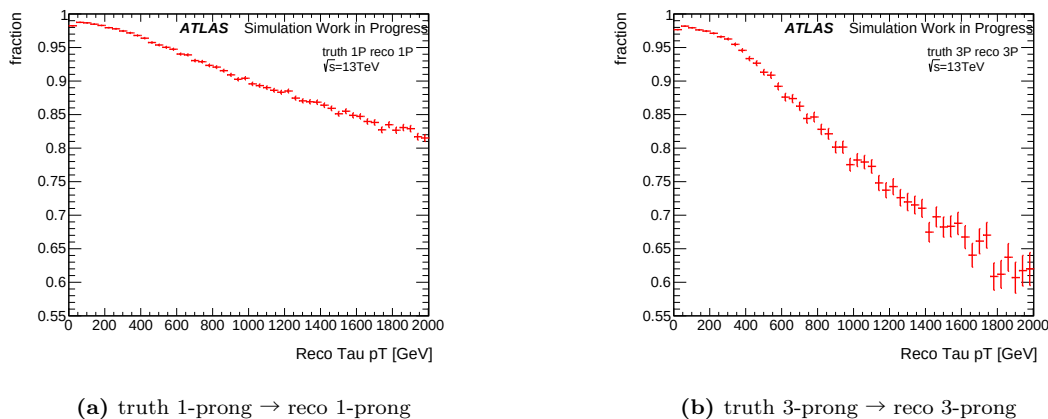


図 5.4: 1-prong のタウ粒子と 3-prong のタウ粒子の電荷の正解率

2-prong のタウ粒子の場合、再構成される電荷は +2, -2, 0 の 3 パターンである。これらの割合がどの程度存在しているかを、truth 1-prong と truth 3-prong で分けてタウ粒子の p_T の関数として調べたものが図 5.5 である。全てのパターンでタウ粒子の p_T 依存は、ほとんどないことが分かる。truth 1-prong の場合、2 本目のタウトラックは誤ってタウトラックに識別された対生成トラック由来であることが多く、その電荷に真のタウ粒子の電荷との相関はないため、2 本の電荷の組み合わせは確率的にランダムになっている。

truth 3-prong の場合の結果を 2-prong として再構成される理由から考える。理由 1 の「低い p_T を持つ 1 本のタウトラックが、タウトラックとして認識されないこと」が理由であると考え、3 本のタウトラックのうち失った 1 本のトラックの電荷がタウ粒子の電荷と同じ場合に、電荷の和が 0 となる。つまりその確率は 2/3 である。

次に理由 2 「主に高い p_T で、2 本のタウトラック同士の距離が近く 1 本として再構成されてしまうこと」であると考え。このとき図 5.6 のように、3 本のタウトラックのうち 2 本が 1 本として再構成されるパターンを洗い出すと便利である。この図では真のタウ粒子の電荷を +1 としている。ここでいくつかの仮定をする。1 本がそのまま 1 本で再構成されたタウトラックは正しい電荷で再構成されているとする。また 1 本として再構成されてしまう 2 本のタウトラックの組み合わせはランダムであり、同電荷同士の結合の場合はその電荷で再構成され、反対電荷同士の場合は電荷はランダムに決まるとする。以上の仮定をすると、真のタウ粒子の電荷が +1 の場合、次の A,B,C の 3 パターンがそれぞれ等確率で起こる。

- A. 同電荷の 2 本が 1 本に結合し、その電荷で再構成される。2 本の電荷の和は 0 となる。
- B. 反対電荷の 2 本が 1 本に結合し、マイナス電荷として再構成される。2 本の電荷の和は 0 となる。
- C. 反対電荷の 2 本が 1 本に結合し、プラス電荷として再構成される。2 本の電荷の和は +2 となる。

この中で電荷の識別が可能なパターンは C ののみとなり、その確率はおよそ 1/3 である。

このようにどちらの理由で 2-prong となったとしても 2/3 の確率で電荷が 0 になると考えられ、図 5.5 の結果がほとんどこれと一致していることが確認できた。

このように、真の飛跡数が 1 本の場合も 3 本の場合も 50% を超える割合で電荷が 0 となっていることが分かった。物理解析で 2-prong のタウ粒子を 1-prong や 3-prong のタウ粒子と同様に扱えるようにするため

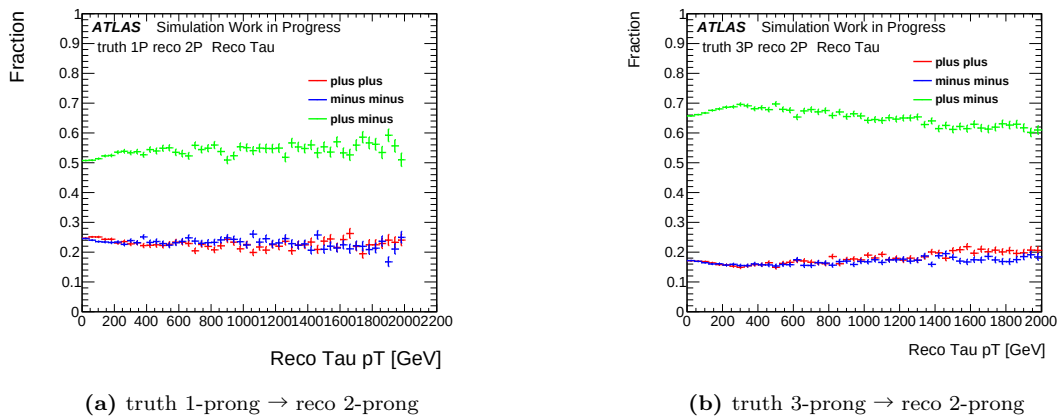
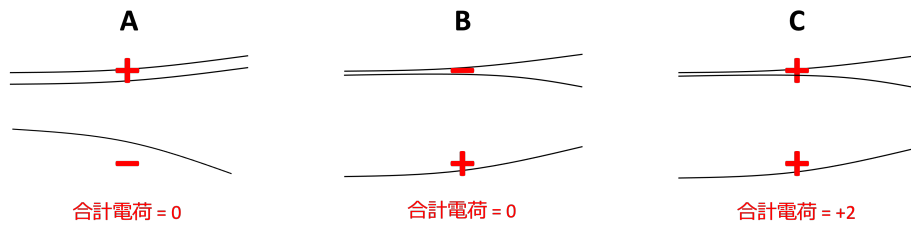


図 5.5: 2-prong のタウ粒子の再構成される電荷の組み合わせ

図 5.6: truth 3-prong $\rightarrow$ reco 2-prong のタウ粒子の理由 2

には、電荷が 0 で再構成された 2-prong のタウ粒子の電荷を正しく識別することが課題である。

最後に、2-prong のタウ粒子が +2 か -2 の電荷に再構成されたとき、その電荷の符号が真のタウ粒子の電荷の符号と一致していた正解率を図 5.7 に示す。1-prong のタウ粒子と 3-prong のタウ粒子の電荷の正解率 (図 5.4a, 図 5.4b) と比較したとき、この 2-prong のタウ粒子の電荷の正解率はほとんど同等である。

truth 3-prong $\rightarrow$ reco 3-prong と truth 3-prong $\rightarrow$ reco 2-prong の電荷の正解率を比較すると、後者の方が高い結果となっている。このように、電荷が ± 2 で再構成された 2-prong のタウ粒子に関しては、それらを +1, -1 の電荷のタウ粒子として扱うことに問題がないと言える。

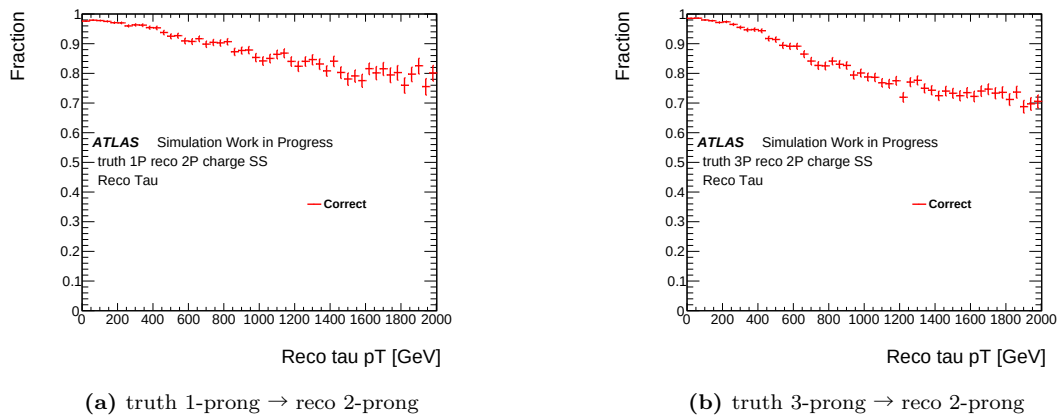


図 5.7: 2 本のタウトラックが同電荷の 2-prong のタウ粒子の電荷の正解率

5.2 真の飛跡数の識別

5.2.1 真の電荷を識別する流れ

前節にて、2-prong のタウ粒子として再構成される理由が truth 1-prong と truth 3-prong で異なることについて述べた。その結果電荷が 0 となる理由やその割合が異なっていることを理解した。このことから電荷が 0 となった 2-prong のタウ粒子の電荷の識別手法を、それぞれの真の飛跡数ごとに行うことでより高い正解率を達成する。本研究では電荷が 0 の 2-prong のタウ粒子をまず truth 1-prong 由来のタウ粒子か truth 3-prong 由来のタウ粒子かを識別する手法を開発する。また truth 1-prong $\rightarrow$ reco 2-prong に対する電荷 0 のタウ粒子の正しい電荷の識別手法と、truth 3-prong $\rightarrow$ reco 2-prong に対する手法をそれぞれ開発する。これら 3 つの識別手法を図 5.8 のような流れで適用し、2-prong のタウ粒子の電荷を決定する。

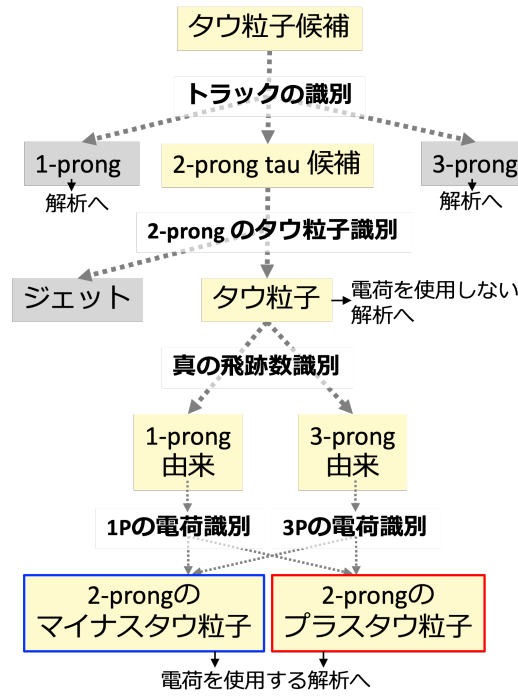


図 5.8: 目指す 2-prong のタウ粒子の識別の流れ

5.2.2 BDT を用いた真の飛跡の本数の識別器の開発

truth 1-prong $\rightarrow$ reco 2-prong のタウ粒子と truth 3-prong $\rightarrow$ reco 2-prong のタウ粒子を識別する手法を開発する。複数変数を用い教師ありでクラス分類を行う機械学習アルゴリズムのうち、今回は線形でない相関を持つ変数を持つ場合に有効に機能する BDT を用いて開発する。真の飛跡数が 1 本の事象を信号事象、3 本の事象を背景事象として開発する。まずは識別に有効な変数を探す。候補とする変数は、タウジェットの変数と、2 本のタウトラックの持つ変数とする。それらの違いを式 5.1 で定義する分離能力で順位づける。その際、タウ粒子の p_T により変数の分布が異なるので、再構成されたタウ粒子の p_T で、表 5.2 のように 4 つの領域に分け、それぞれの領域で順位づけた変数と分離能力の値を表 5.3 と表 5.4 にまとめる。

$$\langle \hat{S}^2 \rangle = \frac{1}{2} \int \frac{(\hat{y}_S(y) - \hat{y}_B(y))^2}{\hat{y}_S(y) + \hat{y}_B(y)} dy \quad (5.1)$$

分離能力は、1 変数の分布が信号事象と背景事象でどれだけ分離しているかを表す指標の一つである。ここで y は信号事象と背景事象それぞれの確率密度関数である。分離能力は 0 から 1 の値をとり、信号事象と背景事象の分布が完全に一致するときに 1 を、信号事象と背景事象の分布に全く重なりがない場合に 0 を取る。

表 5.2: 4 つのタウ粒子 p_T 領域の定義

領域名	low	medium	high	veryhigh
タウ粒子の p_T [GeV]	< 50	$50 \leq, < 100$	$100 \leq, < 500$	$500 \leq$

表 5.3: 分離能力の上位 10 変数 [pt low, medium]

順位	low 変数名	low 分離能力	medium 変数名	medium 分離能力
1	trackBDTScores_0[lowscoreTrack]	2.53E-01	trackBDTScores_0[lowscoreTrack]	2.16E-01
2	flagSet[lowscoreTrack]	2.20E-01	trackBDTScores_1[lowscoreTrack]	1.99E-01
3	trackBDTScores_1[lowscoreTrack]	1.88E-01	etHotShotDR1	1.93E-01
4	ClustersMeanEMProbability	1.62E-01	EMFracFixed	1.88E-01
5	PFOEngRelDiff	1.54E-01	ClustersMeanCenterLambda	1.75E-01
6	ptPi0	1.51E-01	ClustersMeanEMProbability	1.68E-01
7	etHotShotDR1	1.47E-01	ptPi0	1.67E-01
8	qOverP[lowscoreTrack]	1.46E-01	PFOEngRelDiff	1.67E-01
9	dRmax	1.46E-01	etHotShotDR1OverPtLeadTrk	1.67E-01
10	etHotShotDR1OverPtLeadTrk	1.28E-01	PSSFraction	1.58E-01

表 5.4: 分離能力の上位 10 変数 [pt high, veryhigh]

順位	high 変数名	high 分離能力	veryhigh 変数名	veryhigh 分離能力
1	EleMatchLikelihoodScore	2.91E-01	EMFracFixed	3.50E-01
2	etHotShotDR1	2.88E-01	etHotShotDR1	3.40E-01
3	EMFracFixed	2.62E-01	ClustersMeanFirstEngDens	3.35E-01
4	BDTEleScoreSigTrans	2.44E-01	ClustersMeanCenterLambda	3.22E-01
5	centFrac	2.36E-01	PSSFraction	3.04E-01
6	etHotShotDR1OverPtLeadTrk	2.32E-01	EMFRACTIONATEMSCALE_MOVEE3	3.01E-01
7	ClustersMeanCenterLambda	2.31E-01	EleMatchLikelihoodScore	2.87E-01
8	BDTEleScore	2.28E-01	etHotShotWin	2.82E-01
9	etHotShotWin	2.24E-01	centFrac	2.74E-01
10	PSSFraction	2.03E-01	BDTEleScoreSigTrans	2.40E-01

決定した入力変数は以下の 10 変数である。

- 2 本目のタウトラックの `bdtScores_0`
2 本のタウトラックのうち `bdtScores_1` の低い方のトラックの持つ、`bdtScores_0`
- 2 本目のタウトラックの `bdtScores_1`
2 本のタウトラックのうち `bdtScores_1` の低い方のトラックの持つ、`bdtScores_1`
- 1 本目の tau トラックの `bdtScores_0`
2 本のタウトラックのうち `bdtScores_1` の高い方のトラックの持つ、`bdtScores_0`
- `etHotShotDR1`
 $\Delta R < 0.1$ の領域で最もエネルギーの高い部分の E_T
- `EMFracFixed`
カロリメータに落としたエネルギーのうち電磁カロリメータに落としたエネルギーの割合
- `EleMatchLikelihoodScore`
電子らしさを表す likelihood score
- `ptPi0`
 π^0 として再構成されたオブジェクトの p_T
- `PSSFraction`
全てのエネルギーのうち、pre-sampler と電磁カロリメータの 1 層目に落としたエネルギーの割合
- `centFrac`
タウ粒子候補がカロリメータのコア領域に落とした横エネルギーのうち $\Delta R < 0.1$ の領域に落とした横エネルギーの割合。タウ粒子同定変数の 1 つ。
- `dRmax`
タウ粒子候補の方向と、タウ粒子に関連するトラックとの ΔR の最大値。タウ粒子同定変数の 1 つ。

次に BDT の超パラメータを決定する。BDT 超パラメータは次の 2 点を満たすように選択する。

- 過学習がなく安定していること。
- より高い性能となること。

BDT 超パラメータの候補とした値と決定した値を次の表 5.5 にまとめる。超パラメータの候補値は、1 つ以外の超パラメータを固定した状態で 1 つの超パラメータの値を動かして性能を確認し、上記 2 点の条件を満たすと思われる候補値を 1 つあるいは複数選択した。Boosting 手法は、どの超パラメータ値であっても Gradient Boost の方が Adaptive Boost よりも性能が高かったため、Gradient Boost に決定した。また Gradient Boost の場合 BaggedSampleFraction の値を変化させても性能が変化しなかったため、デフォルト値である 0.5 に決定した。以上のように決定した候補値で総当たりで性能を確認し、上記 2 点の条件を満たす超パラメータの値のセットを決定した。

以上のように開発した BDT による真の prong 数識別器の性能を図 5.9, 5.10 に示す。図 5.9 はそれぞれの p_T 領域の BDT score 分布であり、大きく過学習をしていないことが確認できる。図 5.10 は 4 つの p_T 領域の ROC curve を重ね描いた結果である。全ての p_T において、truth 1-prong $\rightarrow$ reco 2-prong のタウ粒子を約 85% の効率で取得するとき、truth 3-prong $\rightarrow$ reco 2-prong のタウ粒子を約 85% で取得できる識別器を開発することができた。

表 5.5: BDT 超パラメータの候補値と決定値 (真の飛跡数の識別)

超パラメータ	候補値	決定値
Ntrees	200, 300, 400, 500, 700, 1000	500
Shrinkage	0.1, 0.2, 0.3	0.1
MaxDepth	2, 3, 4, 5, 7, 10	4
MinNodeSize [%]	0.5, 1.0, 2.0	1.0
BaggedSampleFraction	0.5	0.5
nCuts	20, 30, 50	20
Boosting	Adaptive, Gradient	Gradient

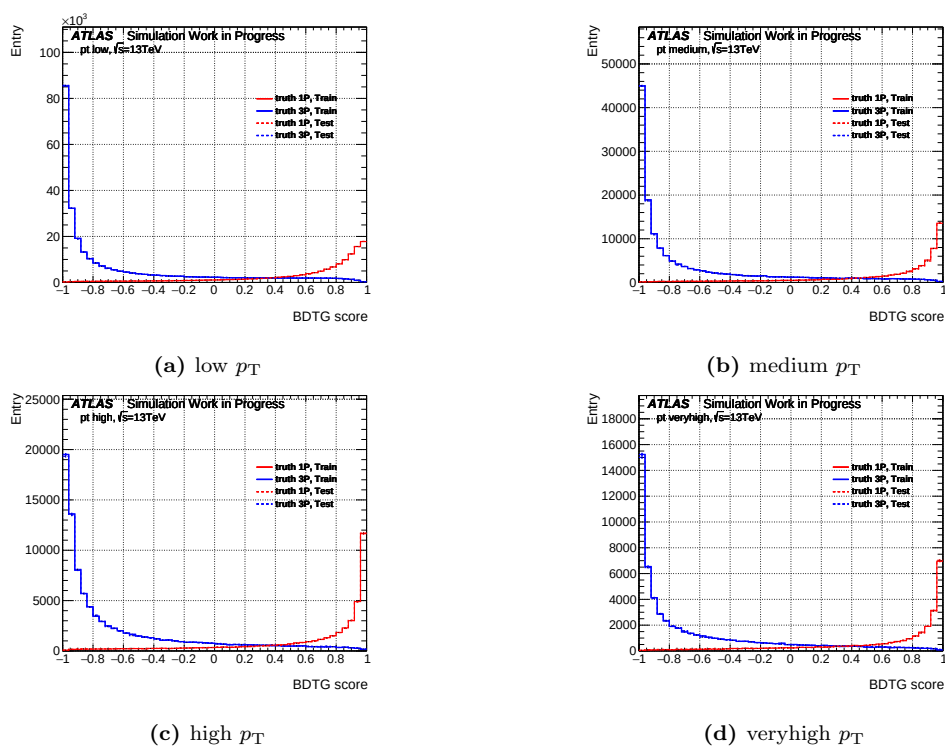
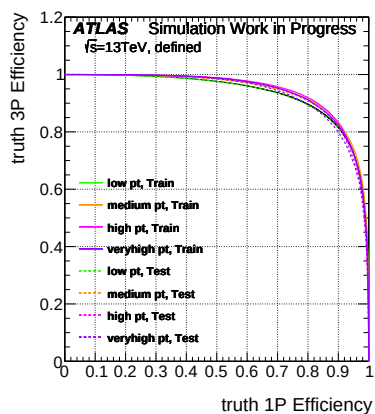
図 5.9: 各 p_T 領域における BDT score の分布

図 5.10: ROC curve (真の飛跡数識別)

5.3 truth 1-prong $\rightarrow$ reco 2-prong のタウ粒子の電荷の識別

truth 1-prong $\rightarrow$ reco 2-prong のタウ粒子のうち、電荷が 0 で再構成されたタウ粒子の真の電荷を識別する手法の開発について述べる。2 本のタウトラックのうち、1 本は偽のタウトラックである。そのため、2 本のタウトラックのタウトラックらしさを示す変数である trackBDTscore_1 に差があると考えられる。この特性を用いて電荷の識別ができるかどうかを確認するため、まずは trackBDTscore_1 のみを用いて電荷の識別ができるかどうかを確認する。2 本のタウトラックのうち trackBDTscore_1 の値がより大きい方のトラックを真のタウトラックと考え、これが真のタウ粒子の電荷と一致した割合を示した結果が図 5.11 である。100GeV 程度までの低い p_T では約 80% で電荷を正しく識別することができている。 p_T が高くなるにつれ正解率は低くなっており、1TeV 程度の高い p_T 領域では約 60% 程度の正解率となっている。

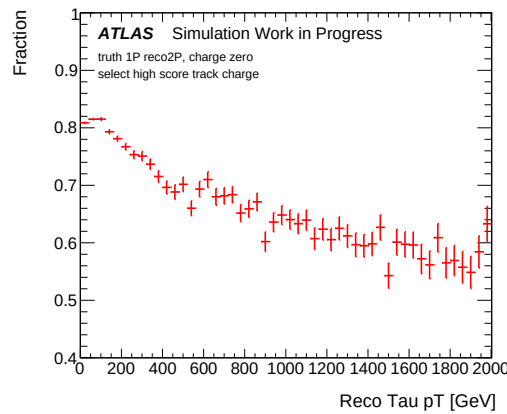


図 5.11: 1 変数でタウ粒子の電荷識別をしたときの正解率 (truth 1-prong $\rightarrow$ reco 2-prong)

trackBDTscore_1 とともに複数変数を用いることでより高い精度で正しい電荷を識別する手法を BDT を用いて開発する。

真にプラス電荷のタウ粒子を信号事象、真にマイナス電荷のタウ粒子を背景事象として、より効率よく 2 つの事象を分離することを目指す。BDT は、2 本のタウトラックの電荷の和が 0 のタウ粒子を用いて学習させテストを行う。2 本のタウトラックのうちプラス電荷のタウトラックの持つトラックの変数と、マイナス電荷のタウトラックの持つトラックの変数の中から BDT の入力変数を決める。真にプラス電荷のタウ粒子の場合、プラス電荷のタウトラックが真のタウトラックであり、マイナス電荷のトラックは偽のタウトラックである。マイナス電荷のタウ粒子の場合はその逆となる。このとき、真の電荷がプラスとマイナスの事象を比較すると、プラストラック同士あるいはマイナストラック同士のタウトラックらしさを表す trackBDTscore_1 の変数の分布に差が生じるため、電荷を識別することが可能となると考えられる。真の飛跡数識別のときと同様に、タウ粒子の p_T によって変数の分布が異なるため、今回も表 5.2 で定義した 4 つの p_T 領域に分けて BDT の学習、テストを行う。

以上の考えをもとに、BDT の入力変数を決定する。まずは入力変数を順位づけする。トラックの変数の分離能力をそれぞれのタウ粒子 p_T 領域ごとに調べた結果の、上位 10 変数の名前と分離能力の値を表 5.6 と表 5.7 に示す。

なおこれらの分離能力を調べる際に、プラストラックの変数とマイナストラックの変数を区別することに意味はないため、真のタウトラックであると思われる真のタウ粒子と同電荷のトラックの持つ変数と、反対電荷のトラックの持つ変数を信号事象と背景事象と考え算出した分離能力を記している。

表 5.6: 分離能力の上位 10 変数 [pt low, medium]

順位	low 変数名	low 分離能力	medium 変数名	medium 分離能力
1	trackBDTScores_0	3.70E-01	trackBDTscore_1	2.78E-01
2	qOverP	2.64E-01	trackBDTScores_0	2.34E-01
3	trackBDTscore_1	2.61E-01	qOverP	1.77E-01
4	rConv	8.69E-02	rConvII	7.55E-02
5	dRJetSeedAxis	8.44E-02	rConv	7.49E-02
6	rConvII	8.17E-02	dRJetSeedAxis	5.90E-02
7	dPhiJetSeedAxis	4.92E-02	dPhiJetSeedAxis	3.94E-02
8	dEtaJetSeedAxis	4.68E-02	numberOfInnermostPixelLayerHits	3.76E-02
9	z0TJVA	2.85E-02	dEtaJetSeedAxis	3.54E-02
10	d0	2.68E-02	nPixHits	2.71E-02

表 5.7: 分離能力の上位 10 変数 [pt high, veryhigh]

順位	high 変数名	high 分離能力	veryhigh 変数名	veryhigh 分離能力
1	trackBDTscore_1	2.12E-01	trackBDTscore_1	4.86E-02
2	trackBDTScores_0	9.22E-02	qOverP	3.40E-02
3	qOverP	7.53E-02	rConvII	3.09E-02
4	numberOfContribPixelLayers	5.28E-02	nSiHits	2.94E-02
5	nPixHits	5.15E-02	trackBDTScores_0	2.93E-02
6	numberOfInnermostPixelLayerHits	5.03E-02	d0	2.83E-02
7	numberOfPixelHits	4.57E-02	numberOfContribPixelLayers	2.55E-02
8	nSiHits	3.49E-02	nPixHits	2.53E-02
9	rConvII	3.08E-02	numberOfPixelHits	2.52E-02
10	dRJetSeedAxis	2.85E-02	expectInnermostPixelLayerHit	1.81E-02

4つの p_T 領域で共通に上位に出てくる変数のうち、BDT でより効果的となる入力変数を選ぶために、

- 分離能力が共通で高いこと
- 変数の信頼性が高いこと
- 他の入力変数との相関係数が小さいこと

の3つの基準から、次の3つの変数を暫定的に選択した。

- trackBDTScores_0
- trackBDTScores_1
- rConvII

これら3つの変数の分布を図 5.12 に示す。なお真のタウトラックと偽のタウトラックの分布はそれぞれ1に規格化して描いている。いずれも真のタウトラックの分布と偽のタウトラックの分布が異なる形をもち、BDT において有効に機能する見込みがあることがわかる。この3つの変数を選んだとき、プラスとマイナスのトラックがあるため、実質的には6変数をBDTの入力変数として用いている。

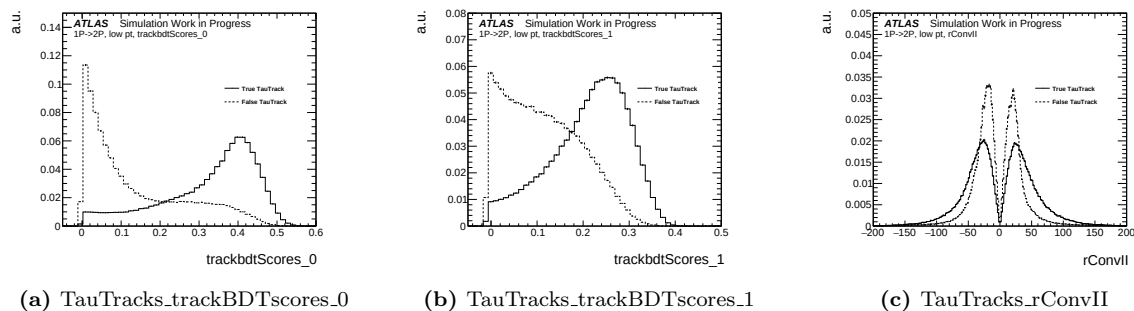


図 5.12: 決定した BDT の入力変数 (low p_T 領域)

以上の3変数に加えて別の変数をBDTに入力したときに正解率に有意に変化があるかどうかを確認した。正解率は、プラス電荷の取得効率とマイナス電荷の取得効率で描いたROC curve (Receiver Operatorating Characteristic curve) の積分値で比較した。他の候補変数を1つずつ追加して積分値を比較したところ、大きな変化が見られる変数は存在しなかった。よって入力変数を前述の3変数に決定する。

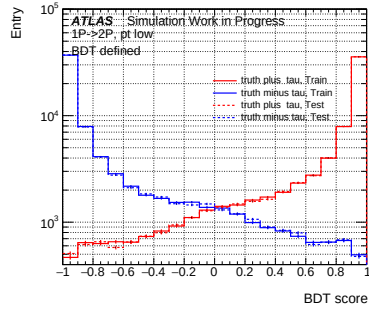
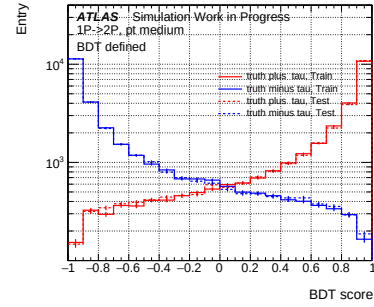
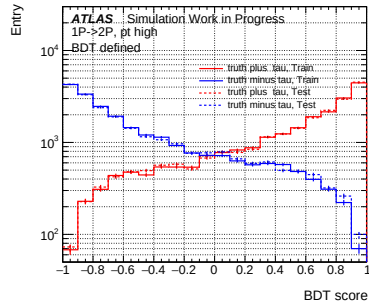
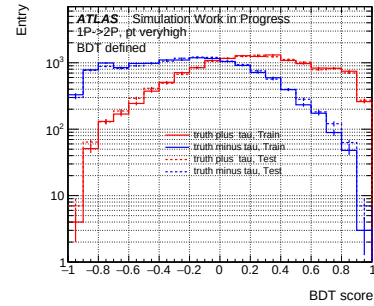
次にBDT超パラメータの調整を行う。BDT超パラメータの候補とした値と、過学習がなくより高い性能となることの2点の条件を満たすことを要求し決定した値を表 5.8 にまとめる。truth 1-prong $\rightarrow$ reco 2-prong のタウ粒子の電荷識別のためのBDTセットアップでは、超パラメータを変化させても性能に大きな違いが見られなかったため、過学習が起こらないという安定性をより重視して決定した。

以上のようにして決定したBDTによる電荷識別器の性能を図 5.13 から図 5.15 に示す。図 5.13 はそれぞれの p_T 領域のBDT score 分布であり、過学習をしていないことが確認できる。またほとんど左右対称の分布であるという特徴がある。low p_T 領域におけるROC curve が図 5.14 である。図 5.13 と同様に $y = x$ に対してほとんど対称となっている。これはプラス電荷とマイナス電荷は同確率で存在し対照関係にあることから期待される結果であると言える。BDT score が0以上のときにそのタウ粒子をプラス電荷とみなし、0より小さいときにマイナス電荷とみなして電荷の正解率を確認する。

図 5.15 では、本研究でBDTを用いて開発した電荷識別器の性能を、TauTracks.trackBDTScore_1の1変数のみで電荷を識別したものと比較している。BDTの結果も同様にhigh p_T 領域で識別効率が下がる傾向がある。1変数での識別と比べて全ての p_T 領域で識別効率が向上していることを確認できた。その伸び幅は特に100GeV以下のlow p_T 領域で大きく、BDTの識別効率は約85%以上を実現した。また1TeV程度のhigh p_T 領域においても約70%程度の識別効率を維持している。

表 5.8: BDT 超パラメータの候補値と決定値 (truth 1-prong $\rightarrow$ 2-prong のタウ粒子の電荷識別)

超パラメータ	候補値	決定値
Ntrees	50,100,200,300,500,1000	200
Shrinkage	0.05, 0.1, 0.2, 0.3, 0.4, 0.5	0.1
MaxDepth	2,3,4,5,6	3
MinNodeSize [%]	0.5, 1, 2, 3, 5	0.5
BaggedSampleFraction	0.3, 0.5, 0.7	0.5
nCuts	10, 20, 30	10
Boosting	Adaptive, Gradient	Gradient

(a) BDT score 分布 (low p_T)(b) BDT score 分布 (medium p_T)(c) BDT score 分布 (high p_T)(d) BDT score 分布 (veryhigh p_T)図 5.13: 各 p_T 領域における BDT score の分布

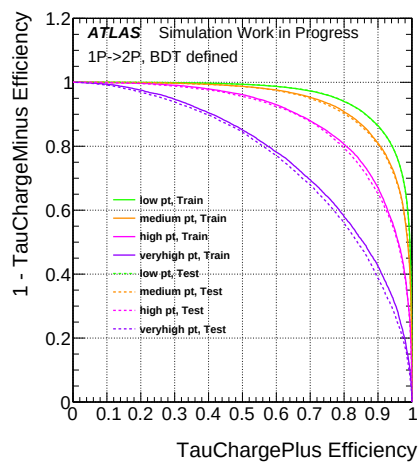


図 5.14: ROC curve (truth 1-prong
→ reco 2-prong)

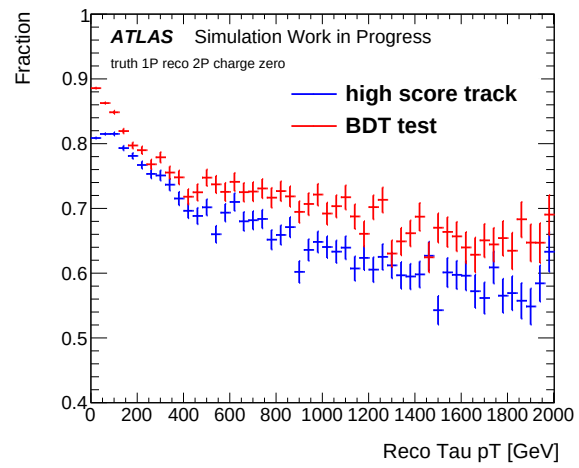


図 5.15: タウ粒子の p_T と電荷の正解率の関係
(truth 1-prong → reco 2-prong)

5.4 truth 3-prong $\rightarrow$ reco 2-prong のタウ粒子の電荷の識別

truth 3-prong $\rightarrow$ reco 2-prong のタウ粒子のうち、電荷が 0 で再構成されたタウ粒子の真の電荷を識別する手法の開発について述べる。3-prong 由来が 2-prong のタウ粒子となる理由は次の 2 つであった。

1. 低い p_T を持つ 1 本のタウトラックが、タウトラックとして認識されないこと
2. 主に高い p_T で、2 本のタウトラック同士の距離が近く 1 本として再構成されてしまうこと

それぞれの理由から電荷を識別する手法を検討する。

理由 1 の場合は、タウトラックと分類されなかったトラック (これを 3 本目のタウトラックと呼ぶこととする) を正しく見つけ出すことで電荷の識別が可能となると考えられる。3 本の真のタウトラックの全てが正しい電荷で再構成されていると仮定すると、2 本のタウトラックが反対電荷の組み合わせであるとき 3 本目のタウトラックは真のタウ粒子の電荷と一致する。よって、3 本目のタウトラックを正しく見つけ出すことでタウ粒子の電荷を識別することが可能となる。

次に理由 2 の場合を検討する。図 5.6 とそのときにおいた仮定を考えると、パターン A とパターン B は等確率で発現する。真のタウ粒子の電荷がプラスのときとマイナスのときのそれぞれで A, B のパターンが存在しており、ある電荷 0 の 2-prong のタウ粒子があったときにどのパターンに該当するかを特定することができれば電荷の識別が可能となる。しかしながら、その識別は困難であると考えられる。2 本が 1 本で再構成されたトラックがどちらかを、ピクセル検出器のエネルギー損失 dE/dx の値を用いて特定できる可能性はあるが、それを特定できたとしても、その 1 本で再構成されたトラックが同電荷同士の結合か反対電荷同士の結合かを識別することが必要である。現状これらを特定することは技術的に難しいと考えられるため、理由 2 の場合を想定した手法の開発は断念し、理由 1 を想定した手法を開発する。

3 本目のタウトラックを、タウトラックと分類されたトラック以外から見つけ出す手法について述べる。タウトラックらしさは `TauTracks.trackBDTScores_1` と `TauTracks.trackBDTScores_0` の値で決まる。タウトラックに分類されなかった場合、3 本目のトラックは対生成トラックに分類されると考えられる。よって、対生成トラックに分類されたトラックが 1 本以上存在する場合、その中で最も高い `trackBDTScores_1` を持つトラックを 3 本目のタウトラックとする。対生成トラックに分類されたトラックが 0 本だった場合は、孤立トラック/フェイクトラックに分類されたトラックのうち、`trackBDTScores_0` の最も大きいトラックを 3 本目のタウトラックとする。ここで、対生成トラックから 3 本目を見つけれられたタウ粒子を「カテゴリー 1」、対生成トラックがなく孤立/フェイクトラックから見つけたタウ粒子を「カテゴリー 2」と名づけ、区別する。

このように 3 本目のタウトラックを発見し、タウ粒子の電荷を正しく識別できた割合を図 5.16 に示す。横軸は再構成されたタウ粒子の p_T である。縦軸の正解率は、選択した 3 本目のタウトラックの電荷がタウ粒子の電荷と一致する割合である。

p_T が 400 GeV 程度までの領域では約 70% の正解率での識別を実現している。一方で 400 GeV 以上の p_T 領域では正解率は下がり続け、800 GeV 以上の p_T 領域では 50% をわずかに上回る程度にとどまっている。

次に、カテゴリー別に電荷の正解率を示したものが図 5.17 である。カテゴリー 1 の正解率がカテゴリー 2 と比べて高くなっている。カテゴリー 2 は p_T が 100 GeV 以上の領域では約 50% の正解率となっており、電荷識別器として全く機能していない結果となった。

今回開発した手法は理由 1 による truth 3-prong $\rightarrow$ reco 2-prong のタウ粒子を想定したものであったため、カテゴリー 1 の場合は電荷を識別できた。しかしカテゴリー 2 の場合は理由 2 による 2-prong のタウ粒子であり、そもそも 3 本目のタウトラックが存在しないため電荷を識別できなかった。

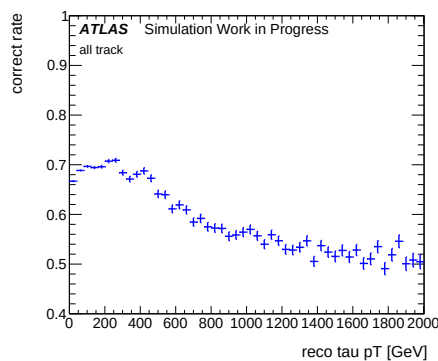


図 5.16: タウ粒子の p_T と電荷の正解率の関係 (truth 3-prong $\rightarrow$ reco 2-prong)

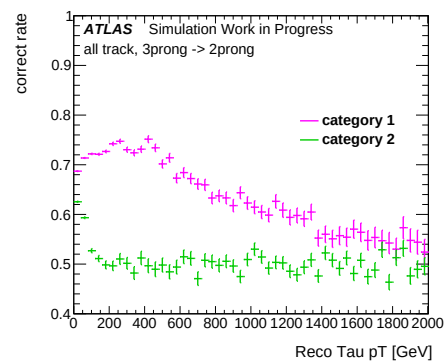


図 5.17: カテゴリーごとの正解率 (truth 3-prong $\rightarrow$ reco 2-prong)

5.5 電荷識別性能の評価

真の飛跡数を識別する BDT score の閾値の決定を行う。閾値の決定は、真の飛跡数識別と後段の電荷識別性能を行い、正解率が最大となる閾値を選択する。電荷が結果として正解するパターンは 2 種類存在する。パターンを図 5.18 にまとめる。飛跡数識別と電荷識別を共に正解する場合だけでなく、飛跡数識別が不正解の場合も 50% の割合で電荷は正解する。よって 2 つの正解パターンを合わせた正解率を確認する。BDT score の閾値は -0.3 から 0.4 まで 0.1 刻みで 8 箇所を試行する。

②電荷識別 \ ①飛跡数識別	①飛跡数識別	
	正解○	不正解×
正解○	正解○	ランダムで正解
不正解×	不正解×	ランダムで不正解

図 5.18: 電荷の正解パターン

truth 1-prong $\rightarrow$ reco 2-prong の正解率を図 5.19a に、truth 3-prong $\rightarrow$ reco 2-prong の正解率を図 5.19b に、2 つを合わせた reco 2-prong の正解率を図 5.19c に示す。truth 1-prong $\rightarrow$ reco 2-prong は閾値が低いほど正解率が高く、truth 3-prong $\rightarrow$ reco 2-prong では反対の傾向となっている。これは、truth 1-prong $\rightarrow$ reco 2-prong 用の電荷識別性能がより高いため、これをより多く取得できる場合に正解率が高くなるためである。一方で reco 2-prong の図 5.19c を見ると正解率は閾値によらず 1% 以下程度のズレしかなく共通となっていることがわかる。よって簡単のため、BDT score の閾値は 0.0 に決定する。

次に決定した電荷識別性能を 5.20 に示す。また、電荷 0 で再構成された 2-prong のタウ粒子だけではなく電荷 ± 2 で再構成された 2-prong んもタウ粒子も合わせた電荷識別効率を 5.21 に示す。電荷 ± 2 で再構成されたタウ粒子のみの正解率は図 5.7 であり、電荷 0 の場合よりも高い効率のため、総合すると低運動領域では 80% 以上の識別効率を達成した。1TeV 程度の高運動量領域でも約 70% の識別効率となった。

5.6 $H/A \rightarrow \tau\tau$ 解析への電荷識別手法の適用

最後に、電荷識別手法を $H/A \rightarrow \tau\tau$ の解析のシミュレーションにおける「検討可能」の種類の電荷 0 の 2-prong のタウ粒子に対して適用し、2 つのタウ粒子が反対電荷の組み合わせとなった事象数を見積もる。まず真の飛跡数識別を行い、1 本由来と識別されたタウ粒子に対して 1 本由来用電荷識別の BDT を適用し

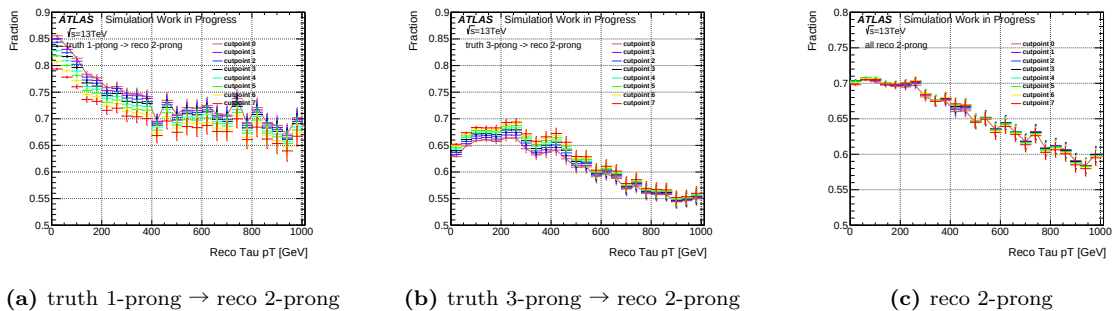


図 5.19: 合わせた電荷識別効率

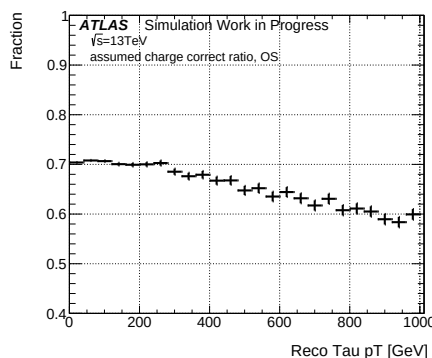


図 5.20: 電荷 0 の 2-prong のタウ粒子の決定した電荷識別効率

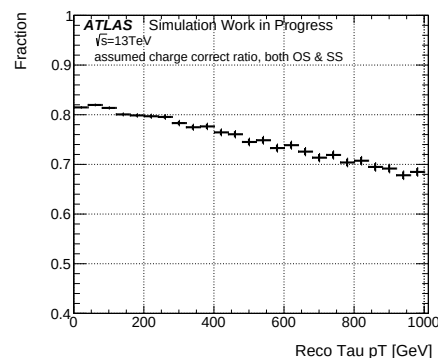


図 5.21: 2-prong のタウ粒子の決定した電荷識別効率

電荷を識別する。3 本由来と識別されたタウ粒子では 3 本目のタウトラックを探す、トラック数が 2 本のみで 3 本目を探すことができない場合がある。その場合、真の飛跡数識別を間違えており 1 本由来であると考え直し、1 本由来の BDT を適用し電荷を識別することも可能であるが、今回は簡単のためその処置は行わず、識別不能として処理する。

表 4.5 の事象数に加えて「検討可能」の事象のうち 2 つのタウ粒子が反対電荷となった事象「電荷 OS」の数をまとめたものが表 5.9 である。また図 4.4 に加えて「電荷 OS」の事象数の増加割合を描いたものが図 5.22 である。

「検討可能」のうち 2 つのタウ粒子が反対電荷となる事象を含めると、従来解析と比較して 1000GeV では約 18%、2000GeV では 25% 以上の事象数を増加させる見込みがある結果となった。また、「検討可能」のうち反対電荷となった事象の割合について考察を行う。「検討可能」のうち、600GeV では約 70%、2000GeV では約 62% の事象が 2 つのタウ粒子が反対電荷である結果となった。今回使用したモンテカルロシミュレーションサンプルは信号事象のサンプルであり、また信号領域の事象選択を行なっているため、「検討可能」の事象は 2 つのタウ粒子が反対電荷の事象が非常に多く含まれると考えられる。600GeV のヒッグス粒子が典型的に 300GeV のタウ粒子に崩壊し、そのうち 1 つが電荷 0 の 2-prong のタウ粒子であると考えられる。図 5.20 より、300GeV では約 70% の電荷識別効率であることから、反対電荷となる割合が約 70% であることの理解ができる。2000GeV のヒッグス粒子で同様の考え方をすると、1000GeV では約 60% の電荷識別効率であり、同様に理解できる結果となっている。このようにモンテカルロシミュレーションサンプルにおいては、開発した結果と整合性のある結果が得られることを確認できた。今後はデータを用いてこの性能を評価していく必要がある。また、本研究では信号事象のみしか調査していないため、今後は背景事象数の見積もりの精査を行い、どのような信号領域を設定するかを検討していく必要がある。

表 5.9: 3 種類の信号領域の事象数 (電荷識別適用後)

種類	600 GeV	1000 GeV	2000 GeV	2500 GeV
従来解析	6639.7	480.10	3.3609	0.5279
追加可能	398.1	39.94	0.3930	0.0676
検討可能	689.7	69.72	0.8067	0.1412
電荷 OS	486.0	47.18	0.4960	0.0833

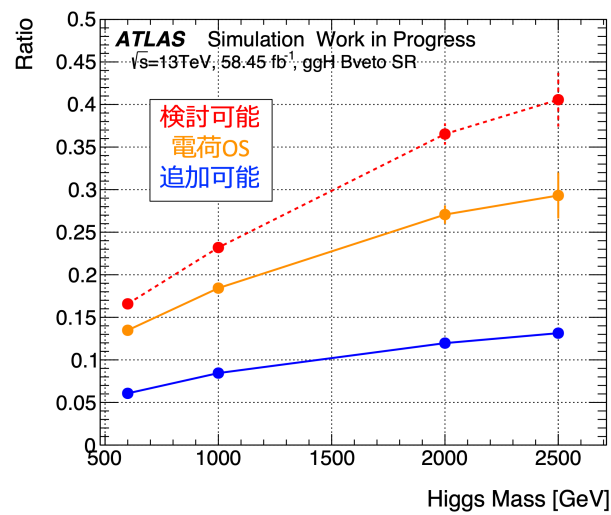


図 5.22: 信号領域の事象数の増加割合 (電荷識別適用後)

6 まとめ

本研究では、物理解析において 2-prong のタウ粒子を用いることで物理感度を高められる可能性があるという見込みに基づき、2-prong のタウ粒子を物理解析に使用できるようにするための研究を行なった。そのために必要な開発は主に同定効率の改善と電荷の識別の 2 つである。同定効率の改善は ATLAS 実験グループとして共通で開発中のため、本研究では電荷識別手法について開発しその性能を評価した。まずタウ粒子の真の飛跡数と再構成される飛跡数に着目し、現状の飛跡数再構成性能や電荷再構成性能を確認した。2-prong に再構成されたタウ粒子の電荷の識別は、2 本の飛跡が同電荷の場合と反対電荷の場合で異なる。電荷が ± 2 で再構成される場合にその符号をタウ粒子の電荷と考えると、1-prong のタウ粒子や 3-prong のタウ粒子と比較して遜色ない電荷の正解率となっていることがわかった。一方で電荷が 0 で再構成されたタウ粒子は真の電荷がプラスかマイナスかがわからないため、この場合に真の電荷を識別する手法の開発を行った。2 本の飛跡で再構成されてしまう理由は、タウ粒子の真の飛跡数が 1 本のときと 3 本のときで異なっている。よって真の飛跡数ごとの識別手法を開発し、それらを適用させるために 2 本が反対電荷の組み合わせの 2-prong のタウ粒子の真の飛跡数を識別する手法も開発した。

まず真の飛跡数の識別は、BDT を用いて開発した。タウ粒子の p_T により 4 つの領域に分割し、真の飛跡数が 1 本と 3 本の場合に違いが生じる変数を精査しセットアップを決定した。結果として全ての運動量領域で、真の飛跡数が 1 本のタウ粒子を約 85% の効率で取得するとき、真の飛跡数が 3 本のタウ粒子を約 85% の効率で取得できる BDT を開発することができた。次に truth 1-prong $\rightarrow$ reco 2-prong のタウ粒子に特化した電荷の識別手法は、2 本の飛跡のうち 1 本は偽のタウトラックであることを利用して、真の電荷がプラスかマイナスかを識別する BDT を開発した。その結果 100GeV 以下の低運動量領域では約 85% 以上、約 1TeV 程度の高運動量領域では約 70% での電荷の識別を行うことができた。truth 3-prong $\rightarrow$ reco 2-prong のタウ粒子に特化した電荷の識別手法は、本来 3 本存在するはずの 3 本のタウトラックのうち、タウトラックに分類されなかった飛跡を探し出すことによって、真の電荷を識別する手法を開発した。その結果 400GeV 程度までの運動量領域では約 70% 程度の効率で識別することができた。しかしそれ以上の超高運動量領域では有効に機能しておらず、改善の余地があるという結果となった。また、2-prong のタウ粒子の総合的な電荷識別性能を評価した。総合的な性能は、低運動量領域では約 80%、1TeV 程度で約 70% 程度の識別効率となった。

また、本研究で開発した電荷識別手法を $H/A \rightarrow \tau\tau$ 解析の信号領域で適用し、事象数とその増加割合を見積もった、従来解析と同様に 2 つのタウ粒子が反対電荷となる事象を含めると、信号領域の事象数が 1TeV のヒッグス粒子の場合約 18%、2TeV の場合 25% 以上増加することが見込まれる結果が得られた。

7 考察と今後の展望

まず、電荷識別効率のさらなる向上を目指す。BDT を用いて開発を行なった真の飛跡数の識別と truth 1-prong $\rightarrow$ reco 2-prong のタウ粒子の電荷識別に関しては、RNN 等の深層学習を用いることを検討する。本研究では有効に機能する入力変数を精査することができた。さらに BDT から RNN へと、手法自体をアップデートすることでその性能をさらに向上させることができると見込んでいる。また本研究では Run2 データを用いた解析用に作成されたモンテカルロシミュレーションサンプルを使用して開発を行なった。今後は Run3 に向けた検出器やソフトウェアのアップデートに従い開発をしていく必要がある。次に truth 3-prong $\rightarrow$ reco 2-prong のタウ粒子の電荷識別に関しては、2つのアプローチを行う。より重要な課題は、2-prong のタウ粒子となる理由 2 である「主に高い p_T で 2 本のタウトラック同士の距離が近く 1 本として再構成されてしまうこと」に着目した電荷識別手法を検討することである。truth 3-prong $\rightarrow$ reco 2-prong のタウ粒子の電荷識別の性能が truth 1-prong $\rightarrow$ reco 2-prong のタウ粒子と比較して劣っているのは、理由 2 の場合に機能する識別手法を開発できていないことが主な理由である。理由 2 の場合の電荷識別を、現状技術的に見ることのできなかったエネルギー損失 dE/dx 等を見ることが実現できる可能性がある。2 本が 1 本で再構成されてしまったトラックは、通常のトラックと比較して dE/dx が大きいと考えられる。この性質を利用し、2 本が 1 本で再構成されてしまったトラックを同定することができると考えている。図 5.6 のラベルを利用すると、本論中では A, B, C の割合がほとんど等しく $1/3$ ずつとなっていると考えられることを議論した。しかし、同電荷の 2 本のトラックが 1 本に再構成される確率と反対電荷の 2 本のトラックが 1 本に再構成される確率が異なっている可能性がある。これらについて精査し、2 本のトラックが 1 本に再構成されたトラックが 2 本の飛跡のうちどちらかを同定することで、タウ粒子の電荷の識別ができる可能性について探っていく。もう 1 つのアプローチは、低運動量領域における 3 本目のタウトラックの探索手法を、複数変数を用いて行うことである。図 5.16 からわかるように、 $p_T < 100\text{GeV}$ の領域では、それ以上の領域と比べて 5% ほど電荷の正解率が低くなっている。図 5.17 を見るとこの領域ではカテゴリー 2 から発見されたタウトラックでも正解する割合が高いことから、低運動量領域では真の 3 本目のタウトラックが、最もタウトラックの分類に近いトラックではない場合もある可能性がある。真の 3 本目のタウトラックの特性を他のトラックとのキネマティカルな特性を含めて精査し、より精度良く 3 本目のタウトラックを選択する手法を検討していく。

今回の電荷識別手法はモンテカルロシミュレーションサンプルを用いて開発した。物理解析に 2-prong のタウ粒子を使用するためには実際のデータにおいてその性能がどれだけ実現できているかを確認することが必要である。そのために 2 つの手法で調査を行う。まず、2-prong のタウ粒子の電荷識別性能をデータを用いて直接的に確認する。これらを確認するための信号領域を、真の 2 つのタウ粒子が反対電荷である事象であり、かつそれが $\tau_{\text{lep}}\tau_{\text{had}}$ の組み合わせとなる領域に設定する。このときレプトン崩壊したタウ粒子の再構成された電荷が正しいと仮定したとき、ハドロン崩壊したタウ粒子の電荷はレプトンの電荷と反対電荷であると考えられる。この信号領域を用いて、2 本の飛跡が反対電荷で再構成された 2-prong のタウ粒子に今回開発した電荷識別手法を適用し、その正解率を確認する。信号領域の選択について、 $Z \rightarrow \tau\tau$ 事象は事象選択と質量の再構成により低運動量領域のタウ粒子を含む事象をある程度確保できると考えられる。また $H \rightarrow \tau\tau$ 事象も同様に用いることが可能であると考えられる。しかし課題となるのは、重粒子の探索で特に重要となる高運動量のタウ粒子の統計量を確保することとなる。そこで $t\bar{t}$ 事象の 2 つの W ボソンが $W \rightarrow \tau\nu$ 崩壊をする事象も用いることも検討する。 $t\bar{t}$ 事象を用いる際に課題となるのが、100% の割合で存在する 2 つの b-jet のタウ粒子へのフェイク確率と、 $t\bar{t}$ 事象以外の事象の混在である。この事象を利用できるかどうかを含めて今後検討していく。次に別のアプローチ手法として、本研究で用いた各種変数がデータとモンテカルロシミュレーションサンプルで一致しているかを確認する。本研究で重要な役割を果たす trackBDTScores や、高運動量領域におけるタウ粒子同定 BDT の入力変数の比較はこれまで見られてこなかった。特に高運動量領域のタウ粒子の統計を確保する解析はこれまで多く行われてこなかった。そこでタウ粒子を多く取得できる物理事象である $W \rightarrow \tau\nu$ 事象を用いて、データとモンテカルロシミュレーション

サンプルの比較を調べる。

本研究で開発した 2-prong のタウ粒子の電荷識別手法は物理解析において使用することを目標としている。 $H/A \rightarrow \tau\tau$ 解析の、従来解析を元にした信号領域において 2-prong のタウ粒子を使用することで見込まれる事象数の増加について見積もった結果、解析感度に影響を与え得る可能性がある結果となった。今後は背景事象数についても見積もり、信号領域の設定や解析手法を含めてさらに検討していく。

8 謝辞

まずはじめに、ATLAS 実験への参加という研究機会を与えてくださり、指導教官として3年間ご指導くださいました寄田浩平教授に感謝申し上げます。物理を知らず研究に対する姿勢が全くなっていなかった私に対して時に厳しく指導してくださり、素粒子物理とその研究の面白さを教えてくださいました。議論はいつも緊張感があり深い思考を促される刺激的な時間でした。また飲み会で様々な話をして、人生について熟考することになったことにも感謝しています。

次に、田中雅士准教授に感謝いたします。発表の場にていただく指摘により自分の理解や思考の足りない点に気づきました。また修士課程以降はミーティングにて多くの議論を交わすのを目の当たりにして、疑問の持ち方や議論の仕方等大変多くを学ばせていただきました。

また駒宮幸男上級研究員にも多大なる感謝を申し上げます。学部4年のゼミ期間中から長らくお世話になっており、節目のタイミングで議論を交わす機会はとても貴重でした。

坂本敦子秘書には研究室の引っ越しや事務手続きで大変お世話になりました。研究室の環境をよくしたいという同じ思いがあり、協力できたことを嬉しく思います。

ATLAS 実験の先輩である加持俊瑛さん・伊藤紘貴さんには、物理の話から ATLAS ソフトウェアの技術的な話までたくさん相談させていただきました。特に伊藤さんはタウ粒子に関連する研究を始めたこともあり、身近な先輩として、そして共に研究を推進する仲間として常々お世話になりました。特別に感謝したいのは同期である秋山大也さん、鈴木隆仁さん、孔さんの三人です。秋山とは物理の話をたくさんしました。知識不足で教えてもらえばかりではありましたが、議論を交わす時間はとても楽しかったです。SUSY の発見を期待しています。鈴木には公私ともに仲良くしていただき、どんな話もすることができた大切な存在でした。孔さんは2年間一緒にタウ粒子に取り組んできた仲間でした。今後の道はそれぞれ異なりますがまた一緒に集まって話ができたらと思っています。学部4年に研究を始めたばかりのときから修士論文の書き上げまでずっと指導して下さったのが三谷貴志さんです。三谷さんは一緒に考えるような研究指導をしてくださり、また技術的にもたくさん助けていただきました。また既に卒業し新天地でご活躍されている方を含めた先輩方・後輩等にも大変お世話になりました。隣の席で朝から晩まで研究に取り組んでいた木村真人さんの研究姿勢から多くを学びました。ANKOK 実験、GRAMS 実験の皆さんは関わることは多くはなかったですが、魅力的な研究内容について教えてくださいたり日頃の研究姿勢から常に刺激を受けたりと、私の研究生活に大きく影響を及ぼしてくださいました。鶴田亮さん・小津龍吉さんは、在学中にお世話になったことはもちろん、卒業後も気遣い応援してくださり、心の支えになっておりました。

忘れてはいけないのが ATLAS 日本の Working Group1-tau のミーティングにて約2年間毎週議論を交わして下さった、東大素粒子センターの奥村恭幸さん・江成祐二さん・増渕達也さん、KEK 素核研の津野総司さん、筑波大数理の廣瀬茂輝さんです。ATLAS 実験の測定器、Combined Performance、物理解析に詳しい上記の皆様からいただくご指摘なくして今回の修士論文を完成させることはできませんでした。皆様に本当に感謝しています。

最後となりますが、幼少期からこれまでの間ずっと私自身が進みたい道を歩めるよう全面的に支えてくださったのは他ではない両親、家族です。高校生の時に漠然と憧れた素粒子物理学の世界を目指して進学しその分野で修士号を取れるのは、各岐路で私の希望・選択を後押ししてくれたからです。本当にありがとうございました。

参考文献

- [1] Particle Data Group, <http://pdg.lbl.gov/>
- [2] CERN HP, <https://home.cern/>
- [3] ATLAS Japan HP, <https://atlas.kek.jp/>
- [4] Heavy Flavor Averaging Group, <https://hflav-eos.web.cern.ch/hflav-eos/semi/spring21/main.shtml>
- [5] Svjetlana Fajfer (Ljubljana U. & Stefan Inst., Ljubljana), Theory Status - Puzzles in B Meson Decays and LFU ,SciPost Phys. (2019), <https://scipost.org/SciPostPhysProc.1.010/pdf>
- [6] R. Aaij et al., Test of Lepton Universality using $B^+ \rightarrow K^+ l^+ l^-$ Decays, <https://journals.aps.org/prl/pdf/10.1103/PhysRevLett.113.151601>
- [7] R. Aaij et al., Test of lepton universality with $B_0 \rightarrow K^{*0} l^+ l^-$ decays, <https://arxiv.org/pdf/1705.05802.pdf>
- [8] Muon g - 2 Collaboration, Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm, <https://journals.aps.org/prl/pdf/10.1103/PhysRevLett.126.141801> (2021)
- [9] LHC long term, <https://lhc-commissioning.web.cern.ch/schedule/LHC-long-term.htm>
- [10] HL-LHC plan, <https://hilumilhc.web.cern.ch/sites/hilumilhc.web.cern.ch/files/inline-images/HL-LHC-plan-2020-Plan-2.pdf>
- [11] list of τ , <https://pdg.lbl.gov/2020/listings/rpp2020-list-tau.pdf>
- [12] τ Branching Fractions, <https://pdg.lbl.gov/2019/reviews/rpp2019-rev-tau-branching-fractions.pdf>
- [13] ATLAS detector. <http://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08003/meta>
- [14] ATLAS EXPERIMENT PUBLIC RESULTS LuminosityPublicResultsRun2, <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>
- [15] ATLAS EXPERIMENT PUBLIC RESULTS Event Displays from Run 2 physics analyses not included in publications, <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/EventDisplayRun2Physics>
- [16] The ATLAS Collaboration, Alignment of the ATLAS Inner Detector in Run-2, <https://arxiv.org/pdf/2007.07624.pdf>
- [17] The ATLAS Collaboration, Operation of the ATLAS trigger system in Run 2, <https://arxiv.org/pdf/2007.12539.pdf>
- [18] ATLAS Japan HP, <https://atlas.kek.jp/>
- [19] The ATLAS Collaboration, Muon reconstruction performance of the ATLAS detector in proton-proton collision data at $\sqrt{s} = 13$ TeV, <https://arxiv.org/pdf/1603.05598.pdf>
- [20] The ATLAS Collaboration, Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC, <https://cds.cern.ch/record/2064383/files/ATL-PHYS-PUB-2015-045.pdf>

-
- [21] The ATLAS Tau Trigger in Run 2, <https://cds.cern.ch/record/2274201/files/ATLAS-CONF-2017-061.pdf>
 - [22] Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at $\sqrt{s}=13\text{TeV}$, <https://cds.cern.ch/record/2261772/files/ATLAS-CONF-2017-029.pdf>
 - [23] Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment, <https://cds.cern.ch/record/2688062/files/ATL-PHYS-PUB-2019-033.pdf>
 - [24] 竹田康亮, LHC-ATLAS 実験における高エネルギータウ粒子の再構成効率の改善, <https://kds.kek.jp/event/30559/contributions/112422/attachments/88211/104751/14aK406-11.pdf>
 - [25] Kosuke Takeda, High-pT 2-prong tau performance, https://indico.cern.ch/event/799509/contributions/3324594/attachments/1798110/2932163/TauCP_ktakeda_0219.pdf
 - [26] Dirk Duschinger, Track Classification - R21 Finalization, https://indico.cern.ch/event/615208/contributions/2481630/attachments/1415563/2167141/TauCP_TrackClassificationTaskforce_20170220.pdf
 - [27] Search for beyond the standard model Higgs bosons decaying to ditau with the full ATLAS Run 2 dataset, <https://cds.cern.ch/record/2666953/files/ATL-COM-PHYS-2019-208.pdf>
 - [28] Search for heavy Higgs bosons decaying into two tau leptons with the ATLAS detector using pp collisions at $\sqrt{s} = 13\text{TeV}$, <https://arxiv.org/pdf/2002.12223.pdf>
 - [29] Search for the neutral heavy Higgs bosons decaying to hadronic ditau in proton-proton collisions at $\sqrt{s} = 13\text{ TeV}$, <https://cds.cern.ch/record/2713501/files/CERN-THESIS-2019-321.pdf>
 - [30] DELPHI, OPAL, ALEPH, LEP Working Group for Higgs Boson Searches, L3, Search for neutral MSSM Higgs bosons at LEP, <https://arxiv.org/abs/hep-ex/0602042>
 - [31] CDF Collaboration, Search for Higgs bosons predicted in two-Higgs-doublet, models via decays to τ lepton pairs in 1.96 TeV proton-antiproton collisions, <https://arxiv.org/abs/0906.1014>
 - [32] D0 Collaboration, Search for Higgs bosons decaying to τ pairs in pp collisions with the D0 detector, <https://arxiv.org/abs/0805.2491>
 - [33] ATLAS Collaboration, Search for additional heavy neutral Higgs and gauge bosons in the ditau final state produced in 36 fb^{-1} of pp collisions at $\sqrt{s} = 13\text{ TeV}$ with the ATLAS detector, <https://arxiv.org/abs/1709.07242>
 - [34] ATLAS Collaboration, Search for additional neutral MSSM Higgs bosons in the $\tau\tau$ final state in proton-proton collisions at $\sqrt{s} = 13\text{ TeV}$, <http://arxiv.org/abs/1803.06553>
 - [35] LHCb Collaboration, Limits on neutral Higgs boson production in the forward region in pp collisions at $\sqrt{s} = 7\text{ TeV}$, <https://arxiv.org/abs/1304.2591>
 - [36] The ATLAS Collaboration, hMSSM summary plots from direct and indirect searches, <https://cds.cern.ch/record/2713580/files/ATL-PHYS-PUB-2020-006.pdf>
 - [37] I-Huan Chiu, Search for the neutral heavy Higgs bosons decaying to hadronic ditau in proton-proton collisions at $\sqrt{s} = 13\text{ TeV}$, <https://cds.cern.ch/record/2713501/files/CERN-THESIS-2019-321.pdf>

-
- [38] The anti- k_t jet clustering algorithm, <https://arxiv.org/abs/0802.1189>
- [39] ATLAS Collaboration, Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1, <https://arxiv.org/abs/1603.02934>
- [40] Local Hadronic Calibration, <https://cds.cern.ch/record/1112035>
- [41] TUNING THE TAU ID ALGORITHM FOR RUN 3,
https://indico.cern.ch/event/885011/contributions/4029288/attachments/2113406/3555229/ATLAS_Tau_and_HLeps_workshop.pdf
- [42] K. Albertsson, S. Gleyzer, A. Hoecker, L. Moneta, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, H. Voss, S. Wunsch, TMVA Users Guide, <https://root.cern.ch/download/doc/tmva/TMVAUsersGuide.pdf>
- [43] Junpei Maeda, 多変数解析ことはじめ, <https://www.icepp.s.u-tokyo.ac.jp/atlas/tutorials/2016/TMVA.2016-12-28.pdf>

A 多変量解析 Boosted Decision Trees

複数の変数を用いて2値分類問題を解く教師あり機械学習の手法の1つである。学習サンプルとテストサンプルに分割し、学習サンプルに対して最適化された性能をテストサンプルで評価する。1つの決定木は図 A.1 のように2つ以上の変数を用いてカットを繰り返す。各分岐点におけるデータの集合体をノードという。各ノードにおいて信号事象と背景事象を最もよく分離させる変数が用いられる。したがって1つの決定木の中で複数のノードで同じ変数を使用されることもあれば一度も使用されない変数があることもある。学習サンプルを用いて分類したとき、各最終ノードで信号事象と背景事象のどちらが多かったかにより、その最終ノードの分類が決定する。決定木は階層を深くするほど複雑な相関を含めて識別することができるが、一方で学習サンプルの持つバイアスを深く学習してしまう危険がある(過学習)。

この問題を改善するために、複数の決定木を用いて学習を行う手法が開発された。このように弱い識別器を多数集めることで強い識別器を作成する手法をブースティング(Boosting)といい、ブーストされた決定木を用いる機械学習という意味で Boosted Decision Trees (BDT) という [42]。1つの決定木を作成しその分類の結果から、事象を再び重み付けして次の決定木を作成する。これを繰り返し、最終的に各決定木の加重平均により、BDT の結果を出力する。ブースティングを行うことで学習サンプルの変動やバイアスに対する結果の出力を安定させることができるとともに1つの決定木と比較して性能を大幅に向上させることができる。弱点として、1つの決定木の場合結果を容易に解釈できるが、BDT の場合複雑化しており決定木をみてその結果や特性を解釈することは困難となることがある。しかしいくつかの決定木を見れば一般的な構造を理解することは可能である。

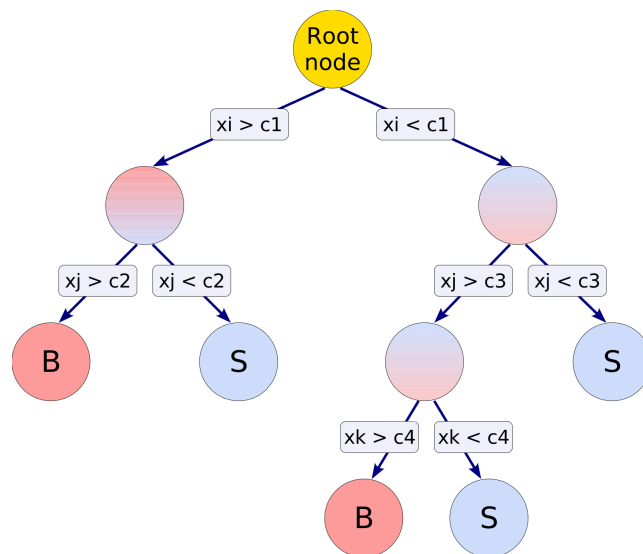


図 A.1: 1つの決定木のイメージ

x_i, x_j, x_k は入力変数を表し、 c_1 - c_4 は各分岐点における変数カットの閾値を表す。最終ノードの S と B はそれぞれ信号事象と背景事象のラベリングを表す。

A.1 ブースティング手法

ブースティングの手法にもその計算方法によりいくつか種類がある。ここでは AdaptiveBoost (AdaBoost) と Gradient Boost(GradBoost) の2つのブースティング手法を説明する。

A.1.1 Adaptive Boost (AdaBoost)

最もよく使用されるブースティング手法である。学習サンプルのある決定木において誤分類された事象には、次の決定木でより高い重み付けがされる。各事象の重みを event weight と呼び、前の決定木に対する次の決定木の event weight を boost weight と呼ぶと、boost weight は式 A.1 で計算される。ただし式中の err は、前の決定木で誤分類された事象の割合を表す。event weight の合計は boost weight をかけた前後で常に保存される。

$$\alpha = \frac{1 - \text{err}}{\text{err}} \quad (\text{A.1})$$

各分類器の結果を $h(\mathbf{x})$ で表す。ただし $\mathbf{x}$ は入力変数のセットを表す。信号事象と背景事象のラベルはそれぞれ $h(\mathbf{x}) = +1$ と -1 である。ブーストされた事象の分類の結果を $y_{\text{boost}}(\mathbf{x})$ とすると、式 A.2 で与えられる。

$$y_{\text{boost}}(\mathbf{x}) = \frac{1}{N_{\text{collection}}} \cdot \sum_i^{N_{\text{collection}}} \ln(\alpha_i) \cdot h_i(\mathbf{x}) \quad (\text{A.2})$$

本論文では $y_{\text{boost}}(\mathbf{x})$ を BDT score と呼ぶ。BDT score は-1 から 1 までの数値であり、1 に近いほど信号事象らしく-1 に近いほど背景事象らしいことを意味する。BDT score に対して閾値を決めることで、信号事象と背景事象を分離することができる。

AdaBoost は各決定木の深さが 2-3 程度の、単独では分離能力が極めて小さい、弱い決定木で通常よく機能する。弱い決定木を与えることで、単純な 1 つの決定木と比べて過学習が起きづらくなり、また性能も格段に向上する。また学習率 β を小さくし多数回ブーストさせることで性能はより向上する。 β は、式 A.1 における α を α^β に置き換えたときの β である。

ブースティングによる出力の関数の推定は、単純な加法展開により理解することができる。出力の関数 $F(\mathbf{x})$ は、弱い学習器である $f(\mathbf{x}; a_m)$ の重み付き和であるとする、式 A.3 で表される。いま、 $f(\mathbf{x}; a_m)$ は各決定木に対応している。

$$F(\mathbf{x}) = \sum_{m=0}^M \beta_m f(\mathbf{x}; a_m); P \in \beta_m; a_m \quad (\text{A.3})$$

ここで関数 $F(\mathbf{x})$ と学習サンプルから得られる真の値 y との差が最小となるように、ブースティング手法を用いてパラメータ P を調整する。この差は損失関数 $L(F, y)$ で測定される。Adaptive Boost においては、損失関数は指数関数の式 A.4 で計算される。

$$L(F, y) = e^{-F(\mathbf{x})y} \quad (\text{A.4})$$

指数関数形の弱点は、外れ値や誤分類されたデータの対応に弱いことである。このため Adaptive Boost は外れ値等の多いデータセットでは性能が低下する性質を持つ。

A.1.2 Gradient Boost

Gradient Boost は、Adaptive Boost の持つ弱点を改善するような損失関数の計算を行い、式 A.5 で表される。

$$L(F, y) = \ln(1 + e^{-2F(\mathbf{x})y}) \quad (\text{A.5})$$

この損失関数に対応するブースティング手法は、最急降下法を用いる。最急降下法では、損失関数の現在の勾配 (gradient) を計算し、決定木中の各領域の勾配の平均値に一致するようにそれぞれのノードの値を調整した決定木を作成することによって行われる。この手順を繰り返すことで、損失関数を最小化する決定木の集合体を作成する。なお最急降下法を用いた Gradient Boost、は勾配の計算が可能であればどのような損失関数でも適用することができる。

Adaptive Boost と同様に、Gradient Boost は深さが 2 - 4 程度の小さな決定木に対して最もよく働く。小さな決定木を用いることで、単純な決定木に比べて過学習してしまう可能性を大幅に下げることが可能

となる。その堅牢性は、個々の決定木の重みを制御する Shrinkage (収縮率) のパラメータを調節し、手法の学習率を下げることでより強くすることができる。一般的には、Shrinkage が 0.1 - 0.3 程度の小さい値で設定すると、このときより多くの決定木の作成が必要とはなるが難しい分類問題において性能を大きく向上させることが可能となっている。

また決定木を作成するために学習サンプルのランダムなサブサンプルを使用する、Bagging のような手法を導入することでさらに性能を向上させることができる。これは Stochastic Gradient Boosting (確率的勾配ブースティング) と呼ばれる。各決定木を作成する際に使用するサンプルを一部のみにすることで、各決定木間の相関が小さくなり、全体としてバイアスや過学習の可能性を低下させることが可能となる。通常 0.5 - 0.8 程度の割合で良い結果が得られている。

Gradient Boost は分類問題において優れており、回帰問題でも使用することは可能ではあるがあまり有効でないことが確認されている。本研究では Adaptive Boost と Gradient Boost を試行し、Gradient Boost を採用した。

A.2 入力変数

BDT に入力する変数として重要な点が 2 つある。1 つ目は、信号事象と背景事象で分離した分布を持つことだ。これは BDT は決定木でカットを繰り返すためである。本論中でも述べた、次式で定義される分離能力が大きい変数は識別において有効である。

$$\langle \hat{S}^2 \rangle = \frac{1}{2} \int \frac{(\hat{y}_S(y) - \hat{y}_B(y))^2}{\hat{y}_S(y) + \hat{y}_B(y)} dy \quad (\text{A.6})$$

2 つ目は入力変数同士の相関係数が小さいことである。BDT はカットを繰り返して決定木を作成するため、相関の強い変数を複数用いたとしても有効に機能しないためである。2 変数 X と Y の相関係数 ρ は式 A.7 で定義される。ここで $\text{cov}(X, Y)$ は X と Y の共分散であり σ_X, σ_Y はそれぞれの変数の分散を表す。

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} \quad (\text{A.7})$$

相関係数は -1 から 1 の値となり、正の場合正の相関、負の場合は負の相関を意味する。2 変数が独立変数のとき $\rho = 0$ となるがその逆は成り立たないことに注意する。 $\rho = \pm 1$ となるのは 2 変数が線形の関係にあるときでありかつそのときに限る。

A.3 超パラメータ

BDT を実行する際のオプションはブースティング手法の他にも存在する。これらを超パラメータといい、代表的な超パラメータを表 A.1 に一覧で示す。

ブースティング手法や Bagging の選択、また解くべき問題の設定により、適切な超パラメータは異なる。BDT の性能を最大化させ、かつ過学習を抑えるためには適切な超パラメータセットを探し決定することが必要である。入力変数を決定後に、超パラメータを 1 つずつパラメータスキャンして試すべき値を決定し、さらにそれぞれの変数の総当たりによって超パラメータセットを確定する。リソース等に余裕があれば、超パラメータセットの決定後に再度入力変数の精査と超パラメータセットの決定を繰り返し、同じ結果に収束することを確認する。

表 A.1: BDT における超パラメータ

超パラメータ	意味
NTrees	決定木の個数
MaxDepth	各決定木で許可された最大の深さ
MinNodeSize	学習サンプルにおける各ノード内に存在してよい事象数の最小の割合
nCuts	各ノードにおけるカットの閾値候補の数
BoostType	ブースティング手法
BaggedSampleFraction	データサンプルの本来の大きさに対するバグ事象サンプルの相対的な大きさ

B タウ粒子同定 BDT の入力変数の説明

Central energy fraction (f_{cent}) タウ粒子候補がカロリメータのコア領域に落とした横エネルギーのうち、 $\Delta R < 0.1$ の領域に落とした横エネルギーの割合。EM エネルギースケールで較正した値を用いる。

Leading track momentum fraction ($f_{\text{leadtrack}}^{-1}$) タウ粒子候補がコア領域に落とした横エネルギーの和のうち、コア領域で最も高い p_T を持つトラックの横運動量の割合の逆数。EM エネルギースケールで較正した値を用いる。

Track radius ($R_{\text{track}}^{0.2}$) タウ粒子候補のコア領域に存在するトラックのみを用いて、 p_T で重みづけした ΔR の距離。

Leading track IP significance ($|S_{\text{leadtrack}}|$) コア領域のトラックの中で最も高い p_T を持つトラックの横方向のインパクトパラメータの絶対値をその誤差で割った値。TJVA で計算された Tau Vertex に対して計算された値が用いられる。

Fraction of tracks p_T in the isolation region ($f_{\text{iso}}^{\text{track}}$) タウ粒子の $\Delta R < 0.4$ 以内のトラックの p_T のスカラー和のうちタウ粒子のアイソレーション領域のトラックの p_T のスカラー和の割合。

Maximum ΔR (ΔR_{Max}) タウ粒子候補の方向と、タウ粒子に関連するトラックとの ΔR の最大値。コア領域のトラックのみが考慮される。

Transverse flight path significance (S_T^{flight}) 横方向平面における、Tau Vertex と Secondary Vertex の距離をその誤差で割った値。multi-track のタウ粒子候補のみに対して計算される値である。

Track mass (m_{track}) コア領域とアイソレーション領域に存在する全てのトラックの四元運動量を用いて計算した不変質量。全てのトラックにパイオンの質量を仮定して計算する。

Fraction of EM energy from charged pions ($f_{\text{EM}}^{\text{track-HAD}}$) タウ粒子のコア領域のトラックの電弱エネルギーの割合。分子はコア領域のトラックの運動量の和とハドロンカロリメータに落としたトポクラスターのエネルギーの和の差で計算される。ただしハドロンカロリメータの計算には電磁カロリメータの 3 層目に落としたエネルギーも含む。分母は電磁カロリメータのトポクラスターのエネルギーの和で計算される。全てのクラスターのエネルギーは LC エネルギースケールで較正される。

Ratio of EM energy to track momentum ($f_{\text{track}}^{\text{EM}}$) コア領域のトラックのエネルギーの和に対する、電磁カロリメータのトポクラスターのエネルギーの和の比。全てのクラスターのエネルギーは LC エネルギースケールで較正される。

Track-plus-EM-system mass ($m_{\text{EM}+\text{track}}$) コア領域のトラックと、コア領域の中で最も高いエネルギーを落とした 2 つの電弱クラスターで構成される系の不変質量。ただし電弱クラスターの四元運動量は、その質量を 0 と仮定し、またトポクラスターのシードの方向を用いて計算される。

Ratio of track-plus-EM-system to p_T ($p_T^{\text{EM}+\text{track}}/p_T$) カロリメータのみから計算されたタウ粒子候補の p_T に対する、コア領域のトラックとコア領域の中で最も高いエネルギーを落とした 2 つの電弱クラスターの p_T の比。

C タウ粒子の識別効率

次に、現状のタウ粒子の識別効率を再構成された飛跡数ごとに確認する。ここではタウ粒子として再構成されたオブジェクトのうち、Loose, Medium, Tight の要求を通った割合を、再構成されたタウ粒子の p_T を横軸に図 C.1 に示す。

図 C.1a は truth 1-prong $\rightarrow$ reco 1-prong のタウ粒子に対する識別効率である。1-prong で再構成されるタウ粒子は、真の飛跡数が 1 のものが支配的であるため、ほとんど設計通りの識別効率となっていることがわかる。次に図 C.1d は truth 3-prong $\rightarrow$ reco 3-prong のタウ粒子に対する識別効率である。これもおよそ設計通りの識別効率になっているが、タウ粒子の p_T が 1 TeV 以上の領域では少し識別効率が悪くなっている。次に図 C.1c は truth 3-prong $\rightarrow$ reco 2-prong のタウ粒子に対する識別効率である。図 C.1d と比べて識別効率が低く、またタウ粒子の p_T に対してフラットとなっていないことがわかる。図 C.1b は truth 1-prong $\rightarrow$ reco 2-prong のタウ粒子に対する識別効率である。これもタウ粒子の p_T に対してフラットとなっていないことがわかるのに加えて、その効率が図 C.1a と比べて低いことがわかる。2-prong のタウ粒子の識別効率が低くなっているのは、multi-prong 用 BDT は truth 3-prong $\rightarrow$ reco 3-prong のタウ粒子を用いて学習されており、2-prong で再構成されたタウ粒子に対して最適化や設計がされていないためである。特に truth 1-prong $\rightarrow$ reco 2-prong タウ粒子は、本来は 1-prong のタウ粒子のため、truth 3-prong $\rightarrow$ reco 2-prong のタウ粒子に比べて同定効率が特に低い結果となっている。

このように、truth 1-prong $\rightarrow$ reco 1-prong のタウ粒子と truth 3-prong $\rightarrow$ reco 3-prong のタウ粒子については設計通りで良い性能となっているが、2-prong で再構成されるタウ粒子については識別効率が低く、またタウ粒子の p_T に対して不安定な性能となっていることがわかった。特に truth 1-prong $\rightarrow$ reco 2-prong に対しては識別効率の低さが際立っている。よって 2-prong に特化したタウ粒子の識別器を開発することの重要性が確認できる。

D 真の飛跡数識別の BDT で用いる入力変数の分布

BDT に用いた入力変数の分布を図 D.1a から図 D.4j に示す。全部で 10 変数を入力している。これらは真の飛跡数が 1 本の分布と 3 本の分布のそれぞれで 1 に規格化して描いている。

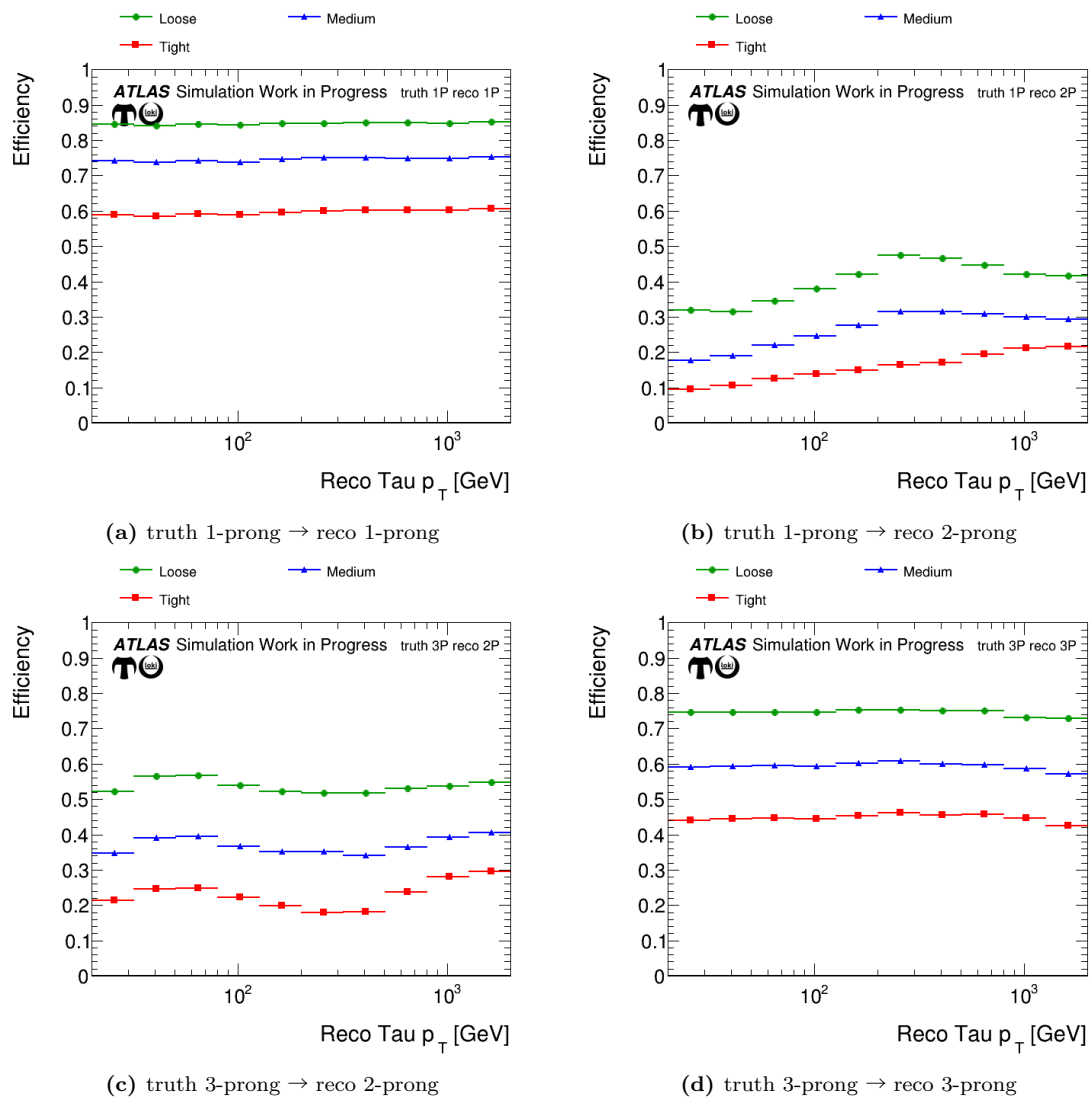
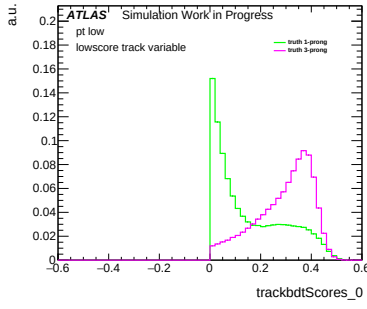
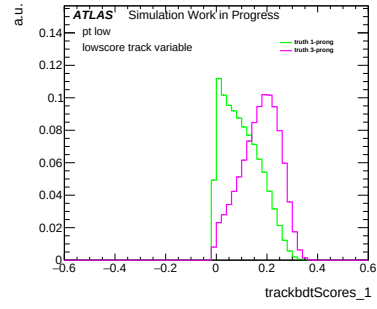


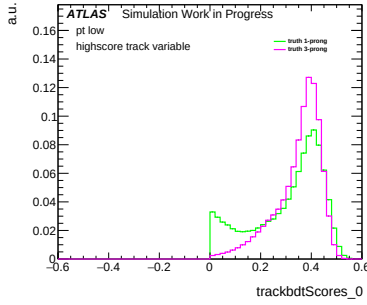
図 C.1: 再構成されたタウ粒子に対する ID の効率



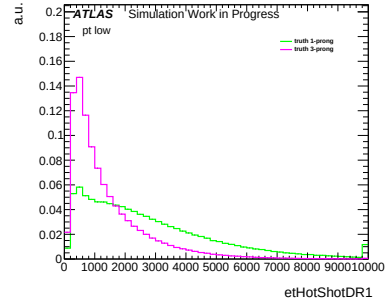
(a) 2 本目のタウトラックの trackBDTScores_0



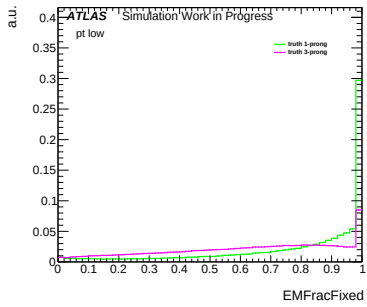
(b) 2 本目のタウトラックの trackBDTScores_1



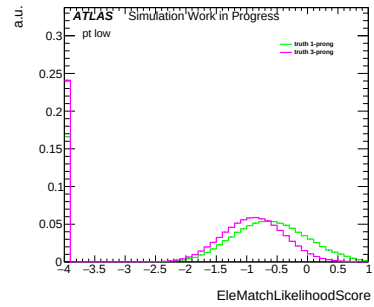
(c) 1 本目のタウトラックの trackBDTScores_0



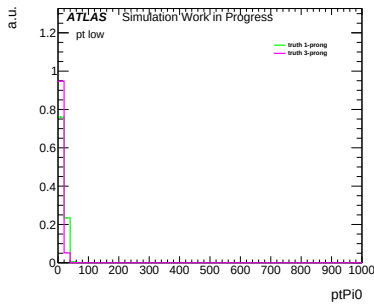
(d) etHotShotDR1



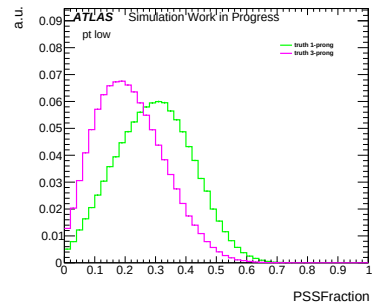
(e) EMFracFixed



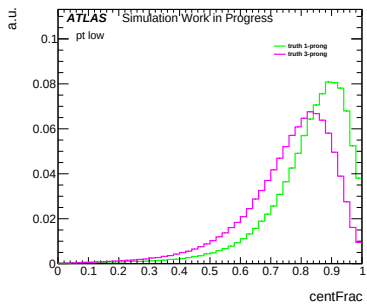
(f) EleMatchLikelihoodScore



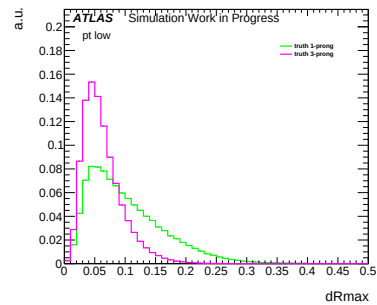
(g) ptPi0



(h) PSSFraction

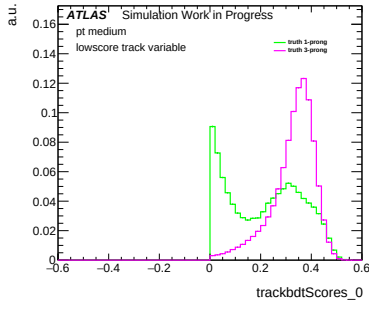


(i) centFrac

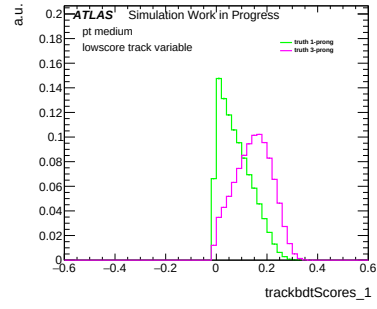


(j) dRmax

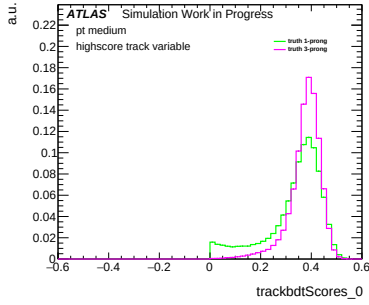
図 D.1: 真の飛跡数識別の BDT で用いる入力変数の分布 (low p_T 領域)



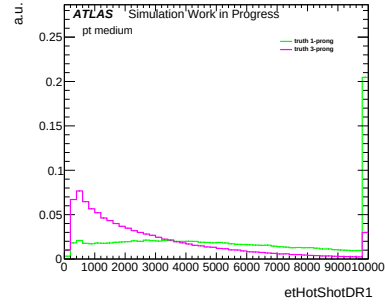
(a) 2 本目のタウトラックの trackBDTScores_0



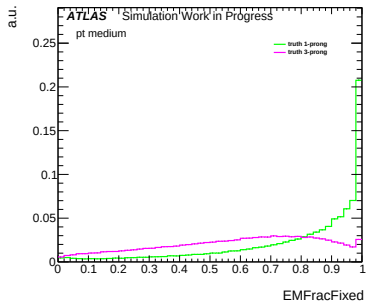
(b) 2 本目のタウトラックの trackBDTScores_1



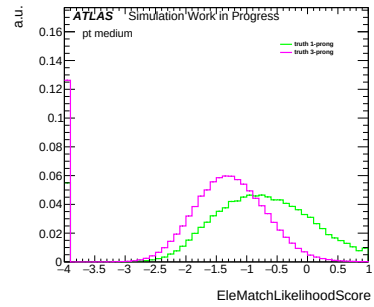
(c) 1 本目のタウトラックの trackBDTScores_0



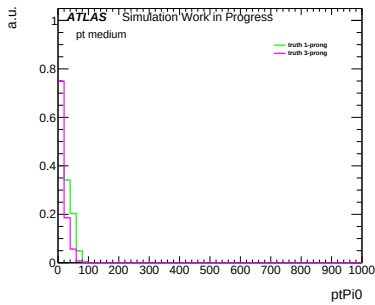
(d) etHotShotDR1



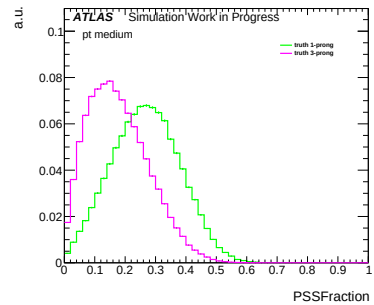
(e) EMFracFixed



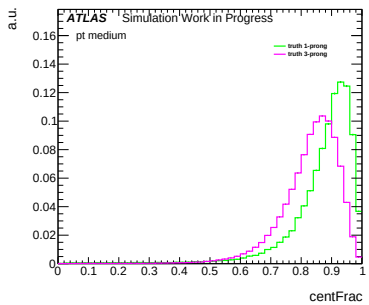
(f) EleMatchLikelihoodScore



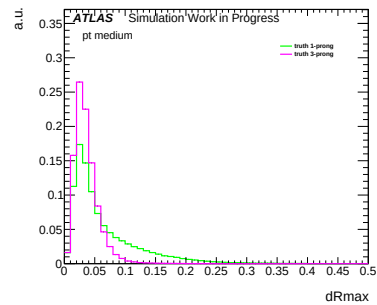
(g) ptPi0



(h) PSSFraction

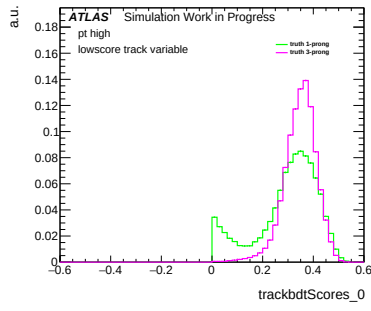


(i) centFrac

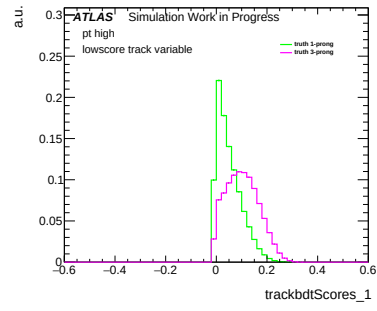


(j) dRmax

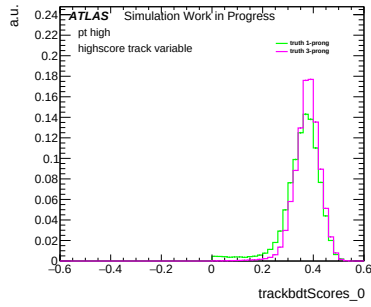
図 D.2: 真の飛跡数識別の BDT で用いる入力変数の分布 (medium p_T 領域)



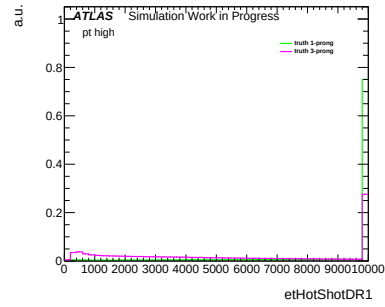
(a) 2 本目のタウトラックの trackBDTScores_0



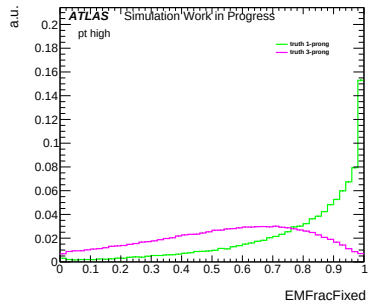
(b) 2 本目のタウトラックの trackBDTScores_1



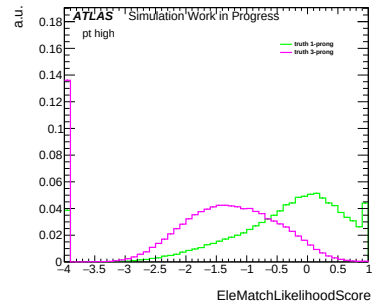
(c) 1 本目のタウトラックの trackBDTScores_0



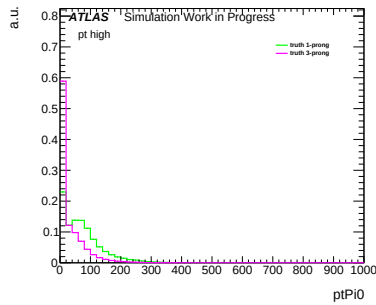
(d) etHotShotDR1



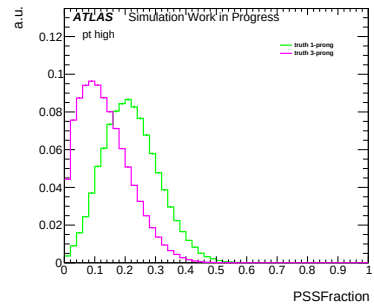
(e) EMFracFixed



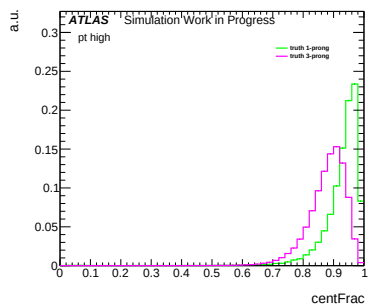
(f) EleMatchLikelihoodScore



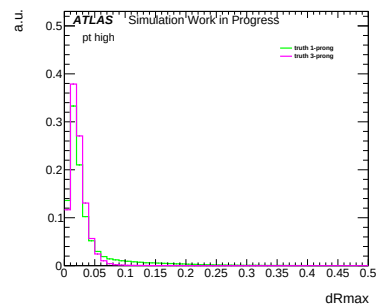
(g) ptPi0



(h) PSSFraction

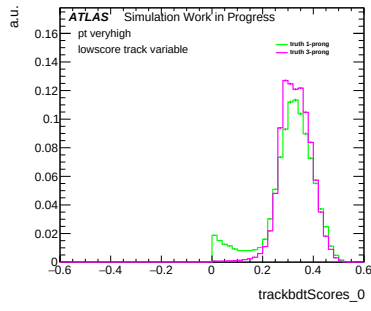


(i) centFrac

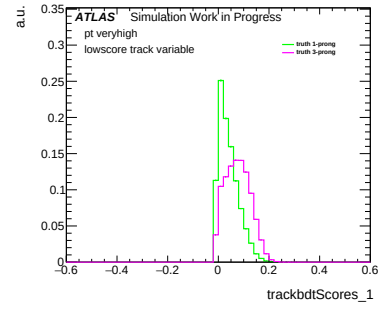


(j) dRmax

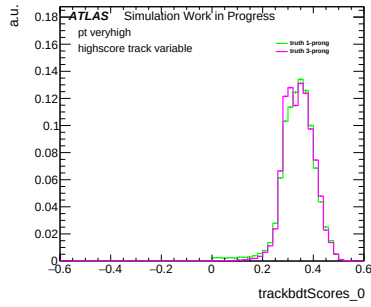
図 D.3: 真の飛跡数識別の BDT で用いる入力変数の分布 (high p_T 領域)



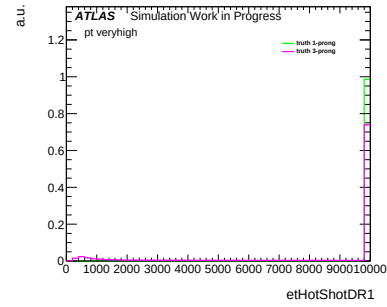
(a) 2 本目のタウトラックの trackBDTScores_0



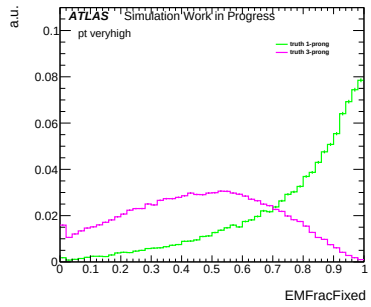
(b) 2 本目のタウトラックの trackBDTScores_1



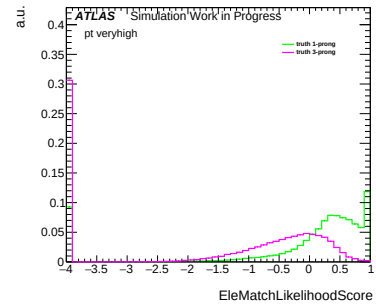
(c) 1 本目のタウトラックの trackBDTScores_0



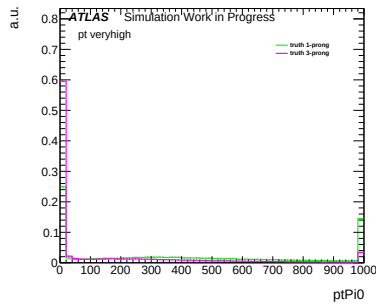
(d)



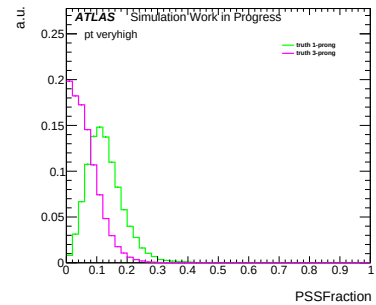
(e) EMFracFixed



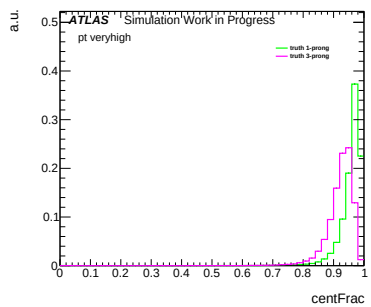
(f) EleMatchLikelihoodScore



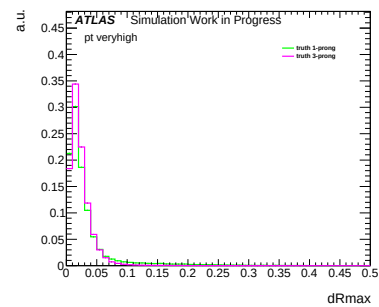
(g) ptPi0



(h) PSSFraction



(i) centFrac



(j) dRmax

図 D.4: 真の飛跡数識別の BDT で用いる入力変数の分布 (veryhigh p_T 領域)

E truth 1-prong $\rightarrow$ reco 2-prong のタウ粒子の電荷識別の BDT で用いる入力変数の分布

medium, high, veryhigh p_T 領域における入力変数の分布を図 E.1 から E.3 に示す。

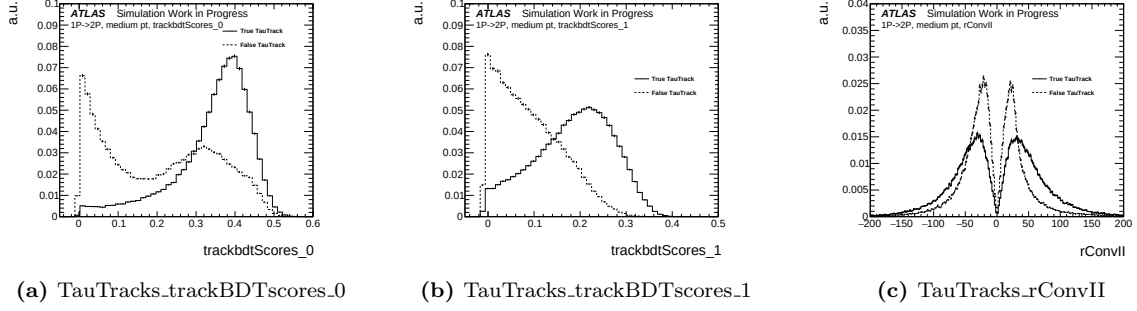


図 E.1: 決定した BDT の入力変数 (medium p_T 領域)

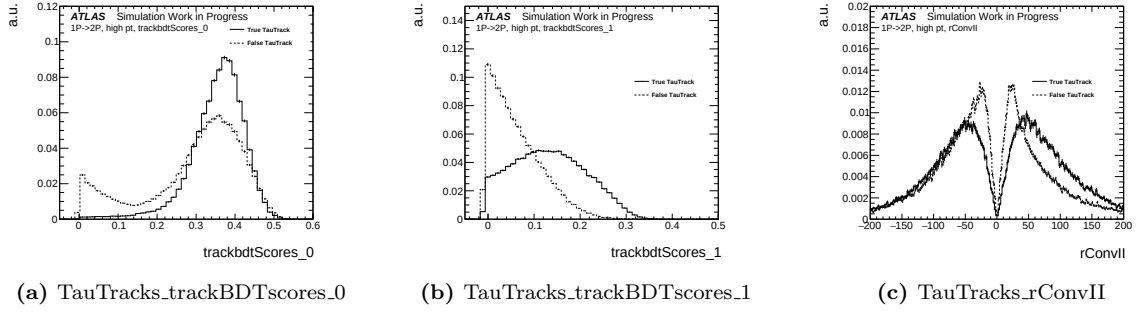


図 E.2: 決定した BDT の入力変数 (high p_T 領域)

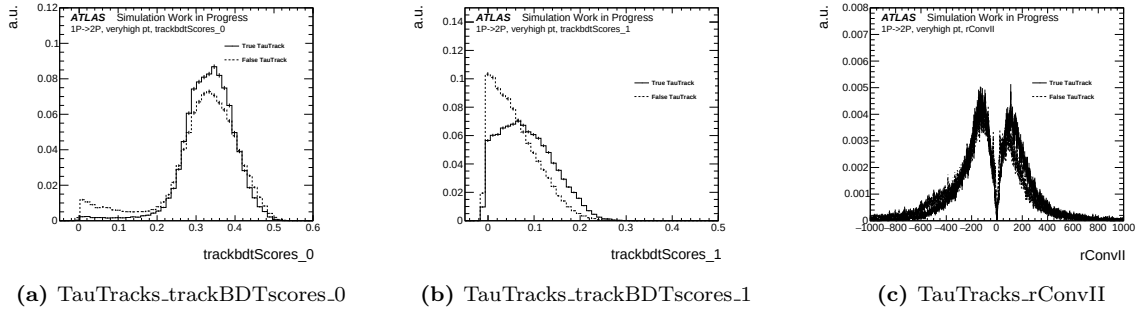


図 E.3: 決定した BDT の入力変数 (veryhigh p_T 領域)

学籍番号: 5320 A091-6

研究活動の不正行為 定義確認書

2022 年 1 月 20 日

早稲田大学大学院先進理工学研究科長 殿

早稲田大学大学院先進理工学研究科

物理学及应用物理学 専攻 2 年

本人氏名 山際 美由希

(自署)

私は、裏面に記載のある研究活動の不正行為に関する定義を理解し、修士論文提出において、不正行為または不正行為と疑われる行為を一切行っていないことを誓約します。
なお、当該行為を行なうと、厳重な処分（無期停学・当該学期成績無効・修士論文不合格等）を受けること、学位取得後であっても学位取消となることを十分に認識した上で、論文を執筆しました。

研究倫理教育受講確認書

2022 年 1 月 20 日

本人氏名 山際 美由希 

私は、修士論文の執筆にあたり、以下の研究倫理教育を受講しました。

✓	GEC 設置科目「研究倫理概論」の受講
	専攻設置科目の受講 科目名 ()
	その他 () ※受講前に指導教員を通じて専攻主任の承認を受けること。 ※受講証明を添付すること。

研究活動の不正行為（捏造、改ざん、盗用、剽窃）について

修士論文を作成するに当たっては、以下の点に十分留意してください。

- ◆既存の文書・資料や自ら取得したデータ等に関し、「捏造・改ざん」は絶対に行ってははいけません。これらの行為は、社会人はもちろん、学生であっても当然守らなくてはならないルールです。

※捏造：事実でないことを事実のように仕立て上げること。でっちあげること。

※改ざん：文書やデータ等の全部あるいは一部を、故意に本来のものでない内容・形式・時期等に変更すること。悪意の有無は問わない。

- ◆学問の体系に新たな知見を加えるとき、その信頼性は命です。学術研究でも、そのための対応が求められます。そこでは上記に加え、資料・データ等の適切な利用と管理、情報取得に当たってのインフォームドコンセントの適用、取得した個人情報の保護等に配慮しなければなりません。

他人の著作物（書籍や論文などに加え、講演での発言やインターネットに掲載された文章・図表・データ等も含む）を活用する場合には、「盗用・剽窃」に十分配慮してください。

盗用・剽窃は、著作権法で禁止された行為です。社会人はもちろん、学生であっても当然守らなくてはならないルールです。

※盗用・剽窃：引用元を適切に記載せずに、他人の文章、結果、意見、アイデア、理論、学説などを自分のものとして発表すること。

- ◆学問の発展は独創性・独自性が基盤です。初めにそれを公表した人のオリジナリティを尊重し、敬意を払うことは学問の府に身を置く者の当然の責務です。学術論文においても、自分の考えと他人の意見を明確に区別し、表現しなければなりません。

このためには、適切な「引用」が重要です。学術論文では、他人の意見・アイデア・理論などを参照・参考にした箇所ごとに番号を付け、巻末や脚注で、その出所を明らかにすることが一般的です。学会等によって、その記載順序が若干異なりますが、以下を標準にします。

- ・論文・総説：著者、題名、掲載誌名、巻号、ページ、発表年（月）等
 - ・国際会議・シンポジウム・口頭発表：著者、題目、シンポジウム等の名称、場所、年月、巻号、ページ、出版元、出版年等
 - ・著書：題目、著者、ページ、出版元、出版年等
- インターネットのからの引用では、以下の通りとします。
- ・著者、タイトル、URL、検索日

- ◆プレゼンテーション用のパワーポイントでは、上記の引用方法を簡略化して用いることを認めます。簡略化の原則は確認・参照の容易性です。例えば、論文からの引用では、著者、掲載誌名、発表年月のみの記載を認めます。著書や論文、インターネットなどの文章をそのまま記載する場合は、「」を付け、その文章を誰が公表しているかを明示するのが一般的です。