



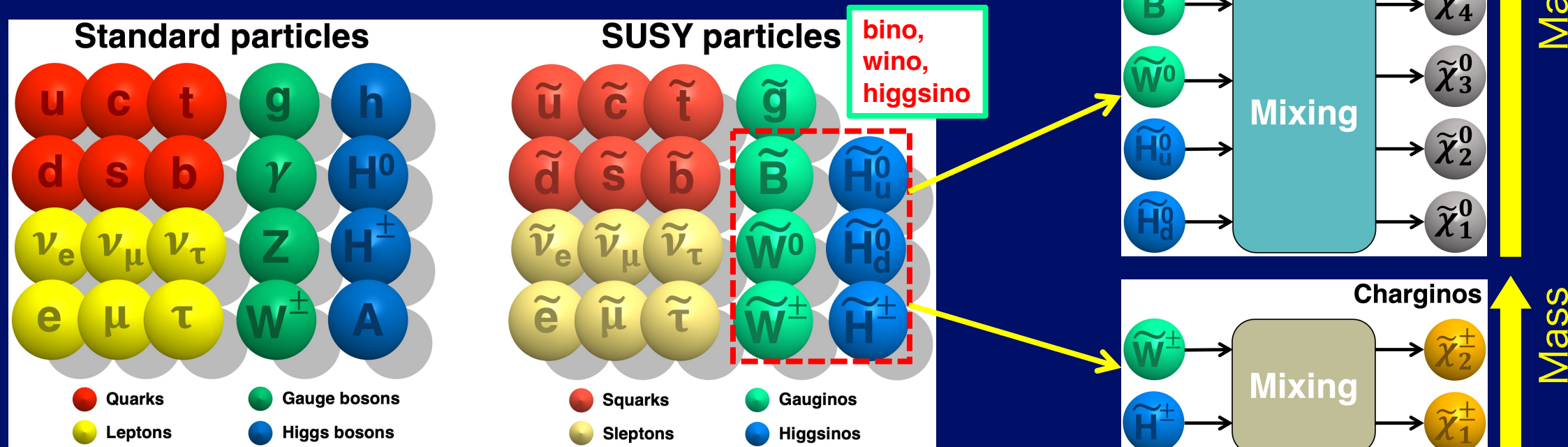
LHC-ATLAS 実験における 深層学習を用いた 超低運動量電子の識別

ML@HEP Workshop

9th July 2022

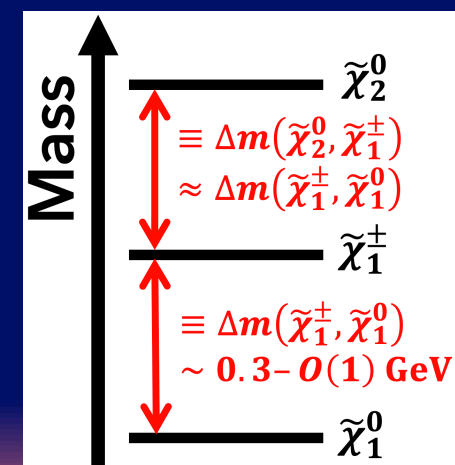
Kaito Sugizaki (The University of Tokyo ICEPP D2)

超対称性 (Supersymmetry, SUSY)



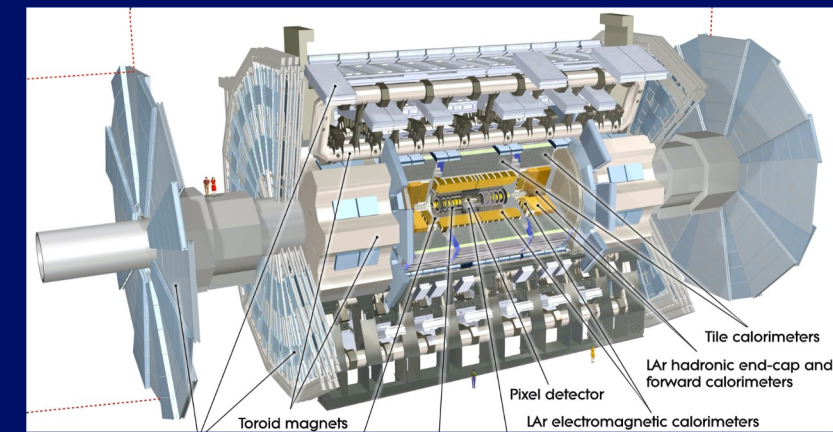
ヒグシーノだけが軽い物理シナリオ

- SUGRA の制約のもとで暗黒物質を説明する有力なシナリオ (ヒグシーノ暗黒物質 $\tilde{\chi}_1^0$)
 - $\tilde{\chi}_1^0, \tilde{\chi}_1^\pm, \tilde{\chi}_2^0$ がほぼ純粋なヒグシーノ：典型的な質量差は $0.3-O(1)$ GeV
 - 加速器実験・暗黒物質の残存密度の制約によって許される質量は $100-1000$ GeV 程度
- ヒグシーノ質量が $100-200$ GeV の場合、SUSY の naturalness との相性が良い

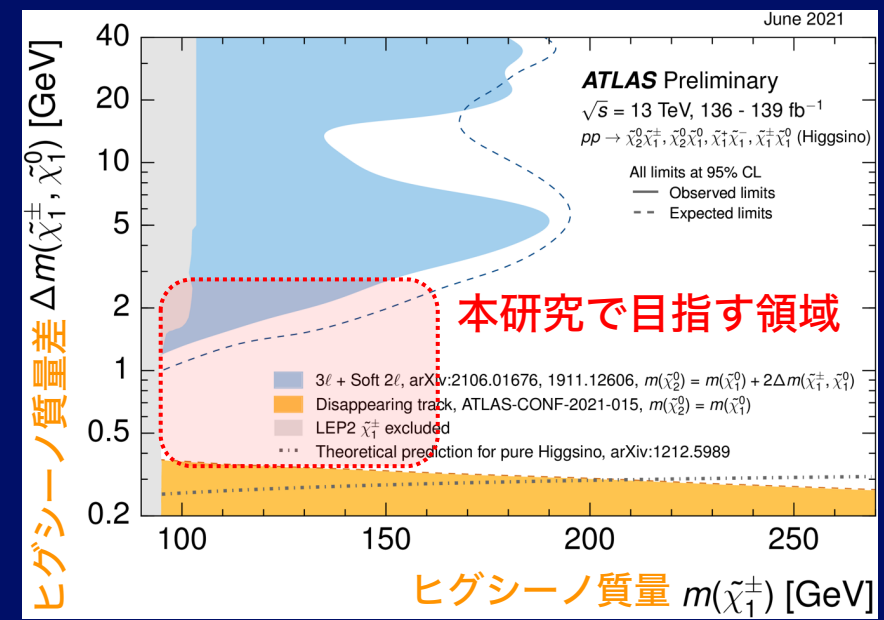
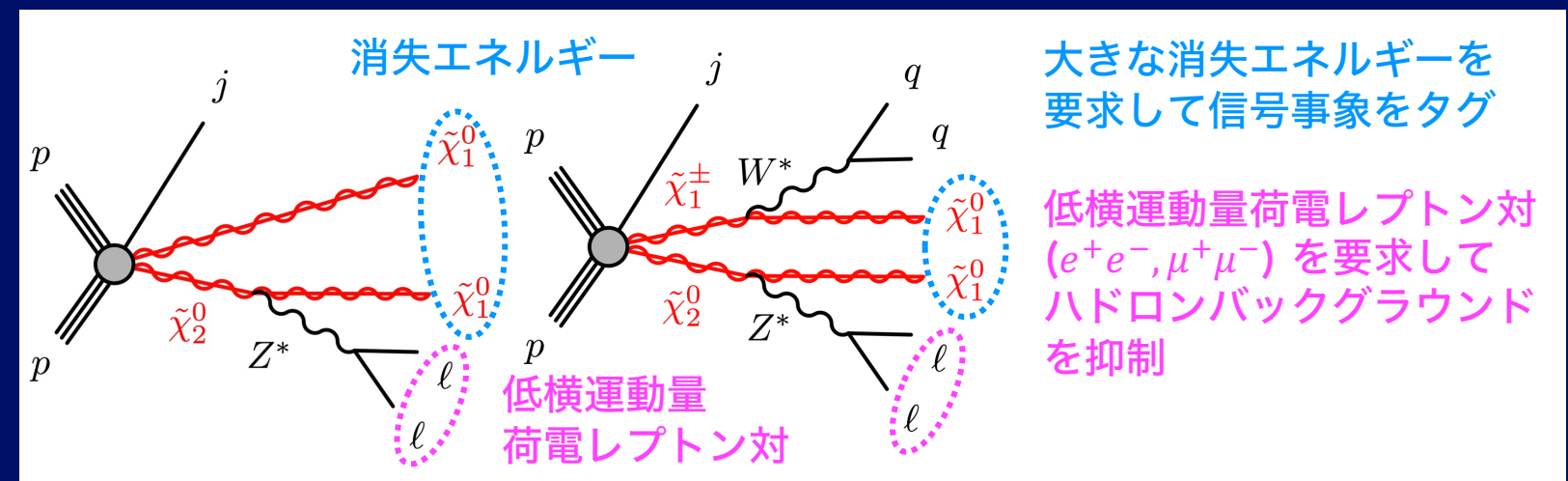


LHC-ATLAS 実験

- 世界最高エネルギーの陽子陽子衝突から標準模型を超える物理を探索
- ATLAS は内部飛跡検出器・カロリメータ・ミュオン検出器から構成
 - ・ 内部飛跡検出器：荷電粒子の横運動量 (p_T) を測定
 - ・ 電磁/ハドロンカロリメータ：電子と光子/ハドロンエネルギーを測定



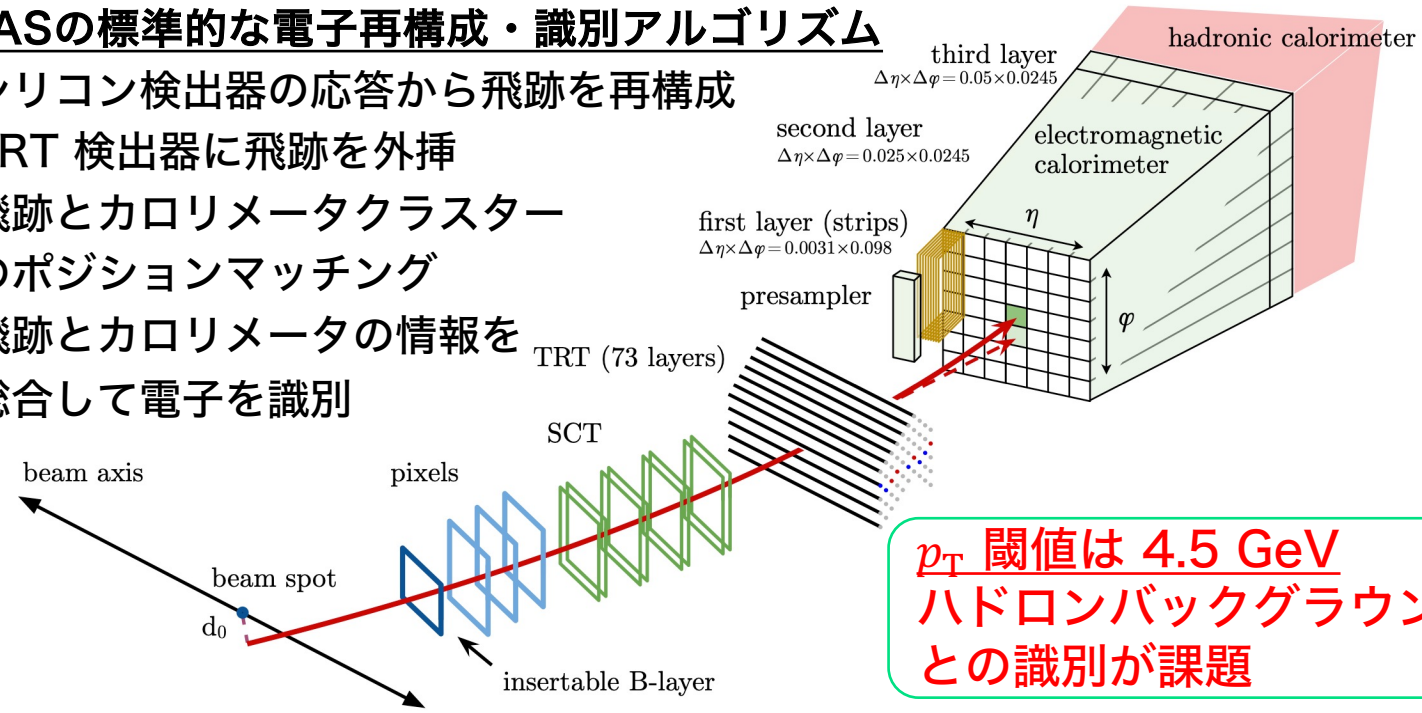
LHC-ATLAS実験におけるヒグシーノ探索の現状



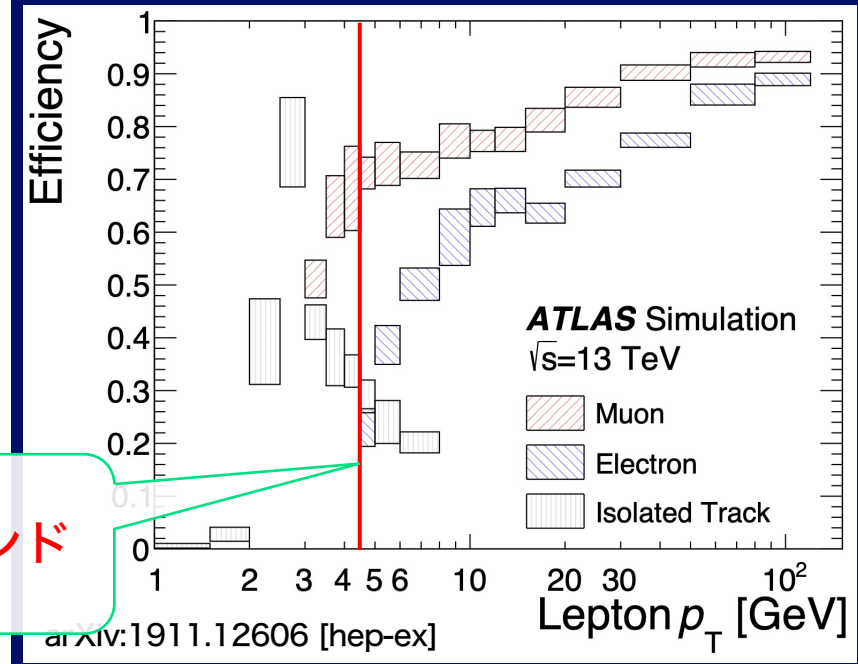
質量差が小さい領域では低運動量レプトンの識別効率が探索感度を大きく制限

ATLASの標準的な電子再構成・識別アルゴリズム

1. シリコン検出器の応答から飛跡を再構成
2. TRT 検出器に飛跡を外挿
3. 飛跡とカロリメータクラスターのポジションマッチング
4. 飛跡とカロリメータの情報を総合して電子を識別



p_T 閾値は 4.5 GeV
ハドロンバックグラウンドとの識別が課題

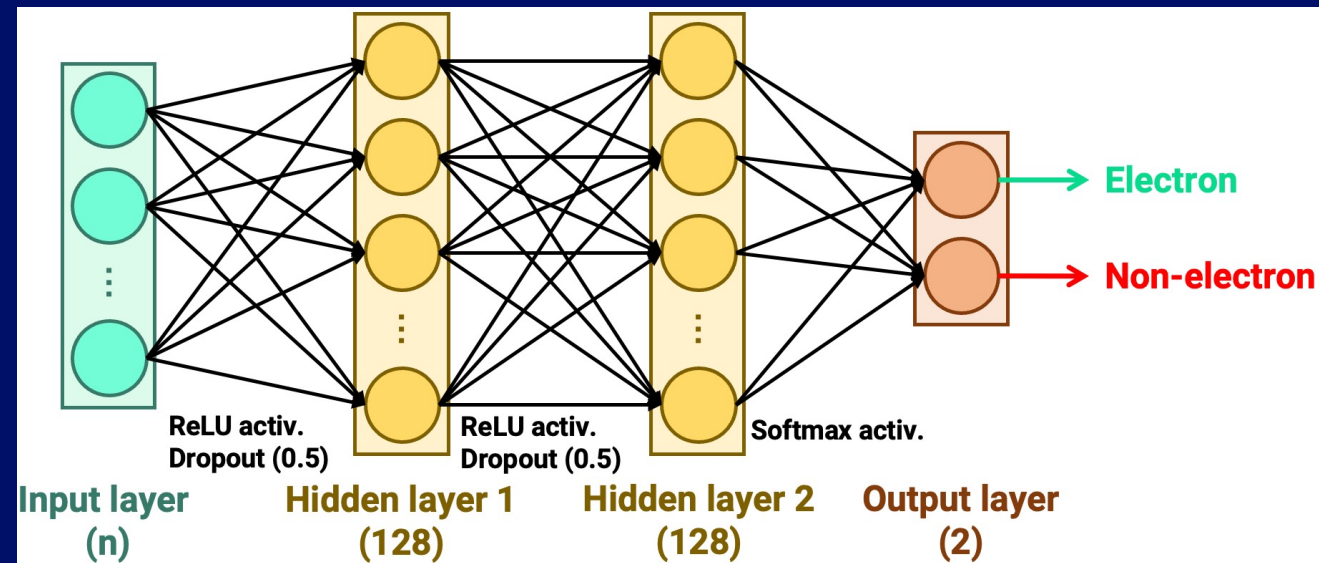


ATLAS検出器における電子の経路の模式図 (arXiv:1902.04655 [physics.ins-det])

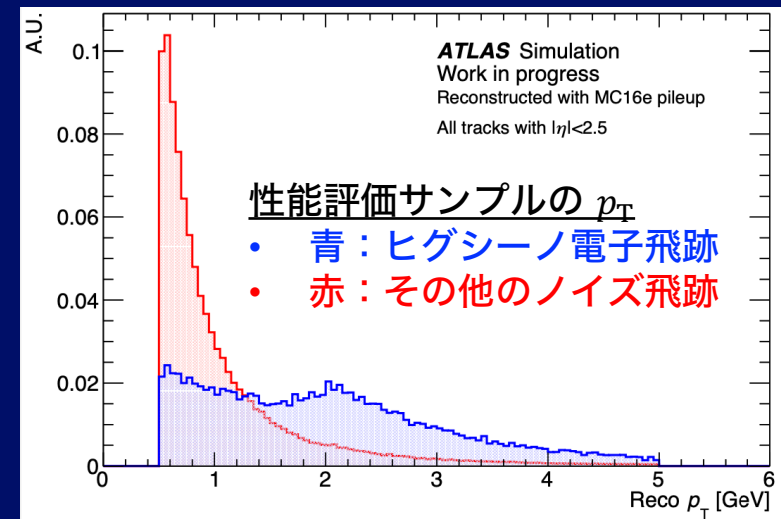
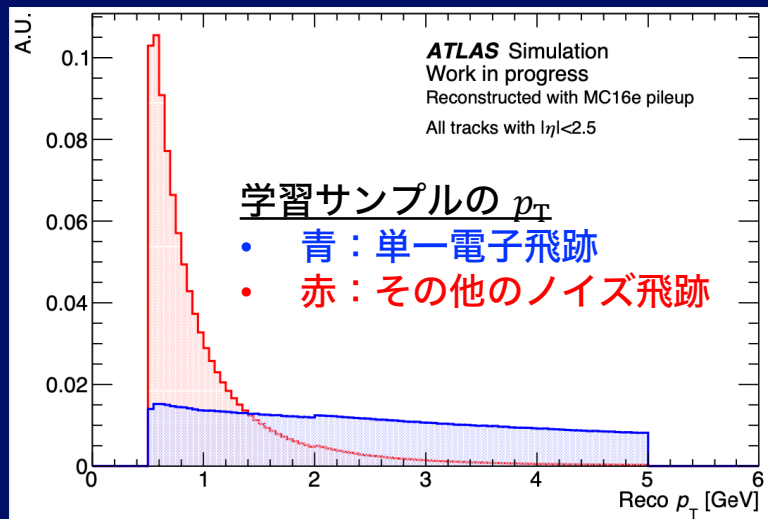
$p_T = 0.5-5$ GeV の超低運動量電子を捉えることに最適化されたアルゴリズムを開発し、ヒグシーノ探索を始めとする重要な物理探索のアクセプタンスを切り拓きたい

$p_T = 0.5\text{--}5\text{ GeV}$ の超低運動量電子を捉える深層学習識別器を新たに開発

- 荷電飛跡が電子であるか否かの二項分類
 - 入力は飛跡の測定量や検出器応答などの変数
- 一般的な sequential モデルを採用
 - Fully dense
 - Two hidden layers with 128 nodes each
 - Dropout of 0.5 to prevent overtraining
- Training parameters
 - Optimizer: Adam (learning rate = 0.001)
 - Loss: categorical_crossentropy
 - Batch size = 512



- ❖ Run 2 を模した pileup データを重ねて再構成した ATLAS 検出器シミュレーションサンプルを使用
 - 学習サンプル： $p_T = 0.5-5$ GeV の単一電子事象サンプル
 - 性能評価サンプル： $m(\tilde{\chi}_2^0, \tilde{\chi}_1^0) = (153, 150)$ GeV を仮定した $pp \rightarrow \tilde{\chi}_2^0 \tilde{\chi}_1^0$ のヒグシーノ生成事象サンプル

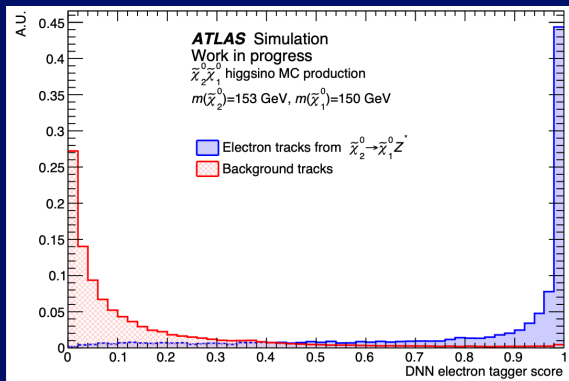


- ❖ 入力変数
 - シリコン検出器による飛跡情報 (+ TRT の応答情報) (+ カロリメータの各層の損失エネルギーなどの情報)
 - TRT 情報の有無、カロリメータの情報の有無によって飛跡を4つのカテゴリーに分類

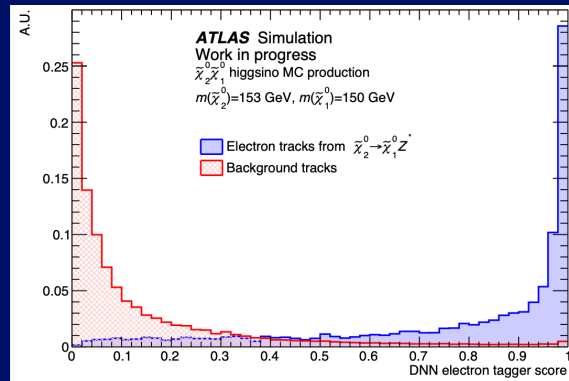
| | w/ TRT & w/ calo | w/o TRT & w/ calo | w/ TRT & w/o calo | w/o TRT & w/o calo |
|--------------------|------------------|-------------------|-------------------|--------------------|
| ヒグシーノ由来の電子飛跡における割合 | 52.2% | 22.9% | 16.1% | 7.6% |
| DNN への入力変数の数 | 79 | 68 | 60 | 49 |

超低運動量電子識別器の性能

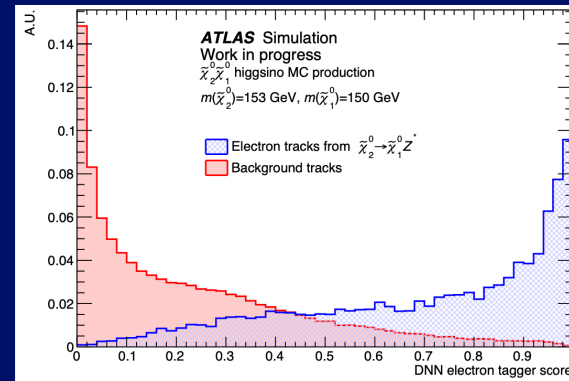
w/ TRT & w/ calo



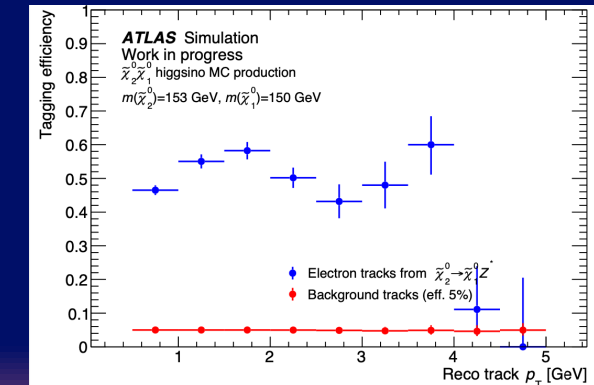
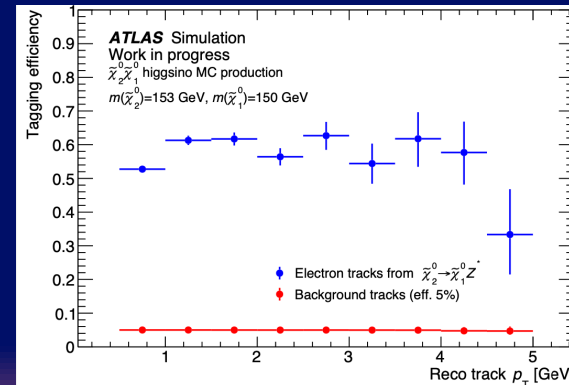
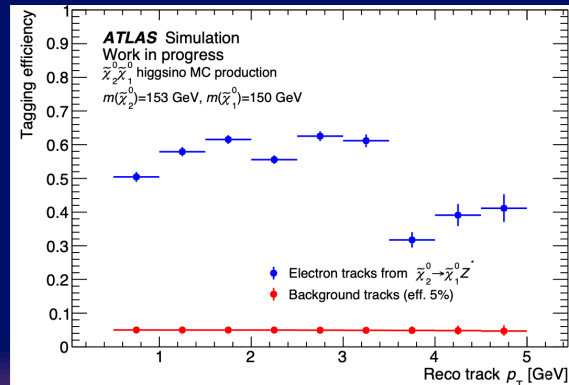
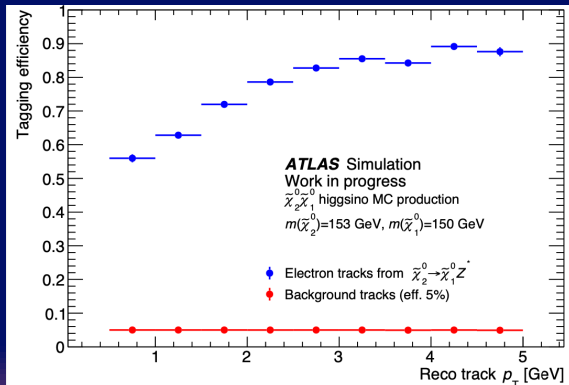
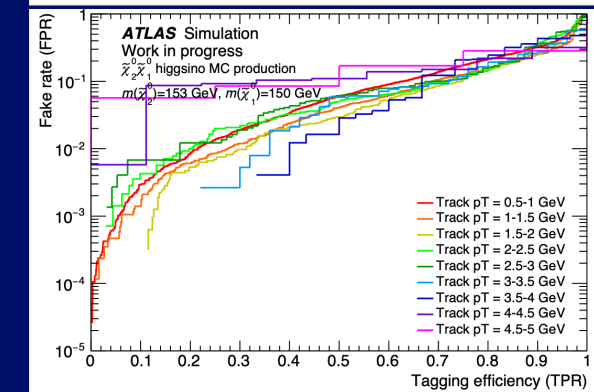
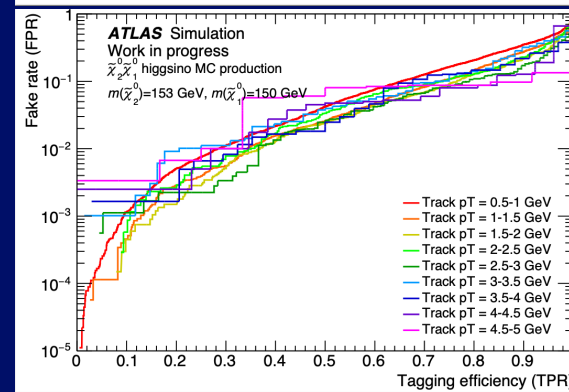
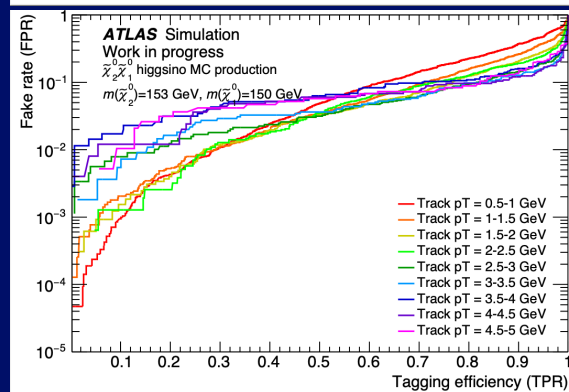
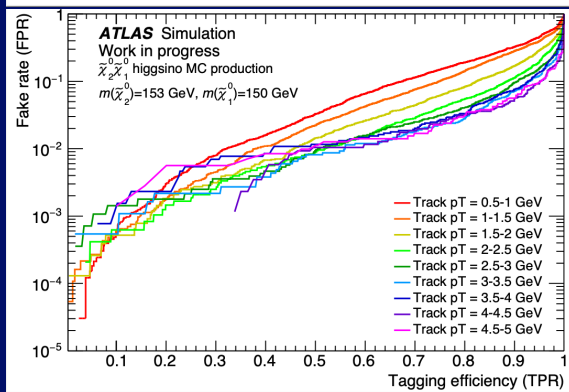
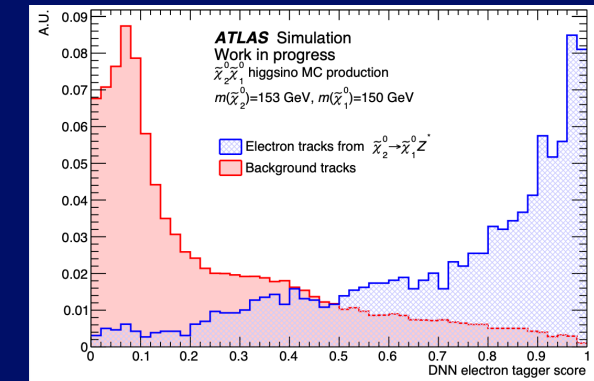
w/o TRT & w/ calo



w/ TRT & w/o calo



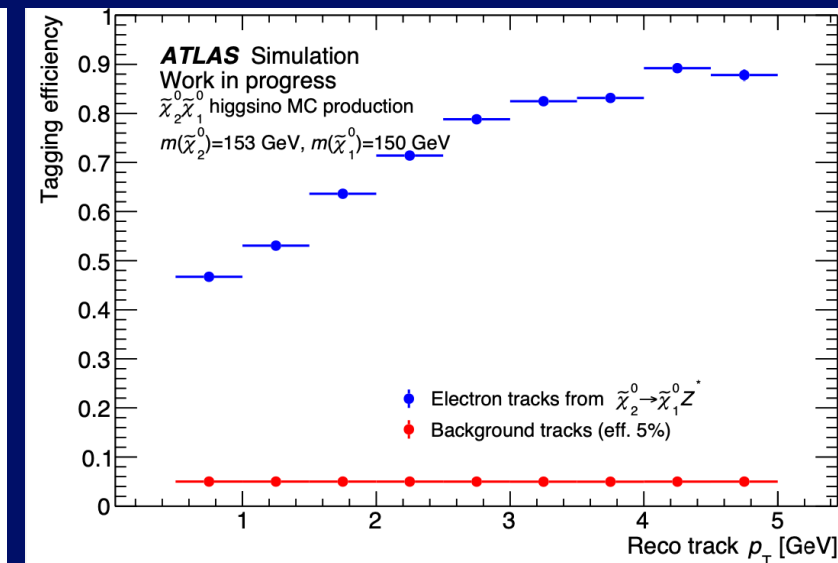
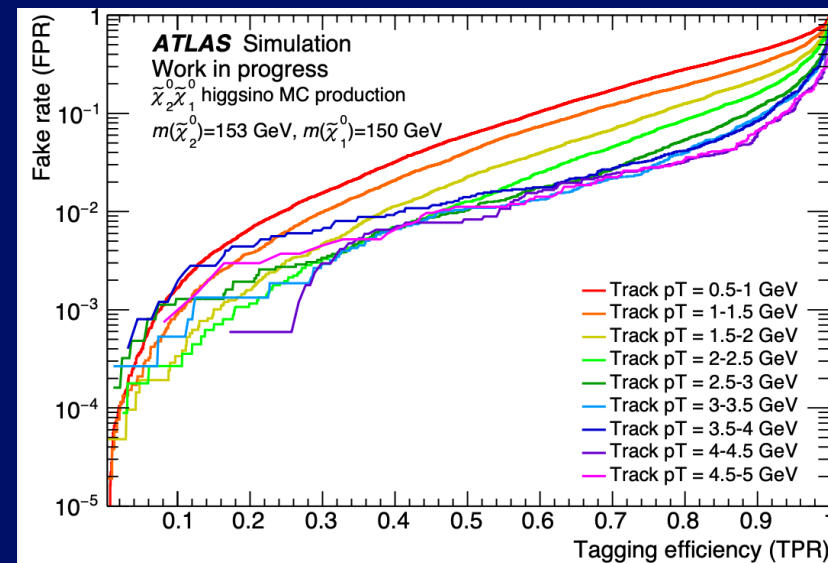
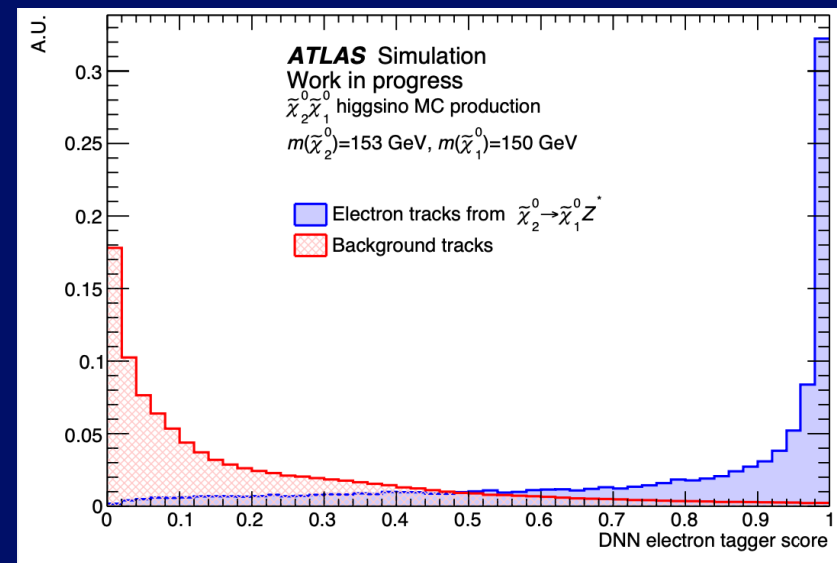
w/o TRT & w/o calo



❖ 4つのカテゴリーにおけるスコア出力を統合

➤ これまで全く見られていなかった超低運動量領域で電子飛跡の識別を高効率で実現

- Fake rate 5% の場合、 $p_T = 1$ GeV で 50%、2 GeV で 60-70%、3 GeV 以上で 80% 以上の識別効率



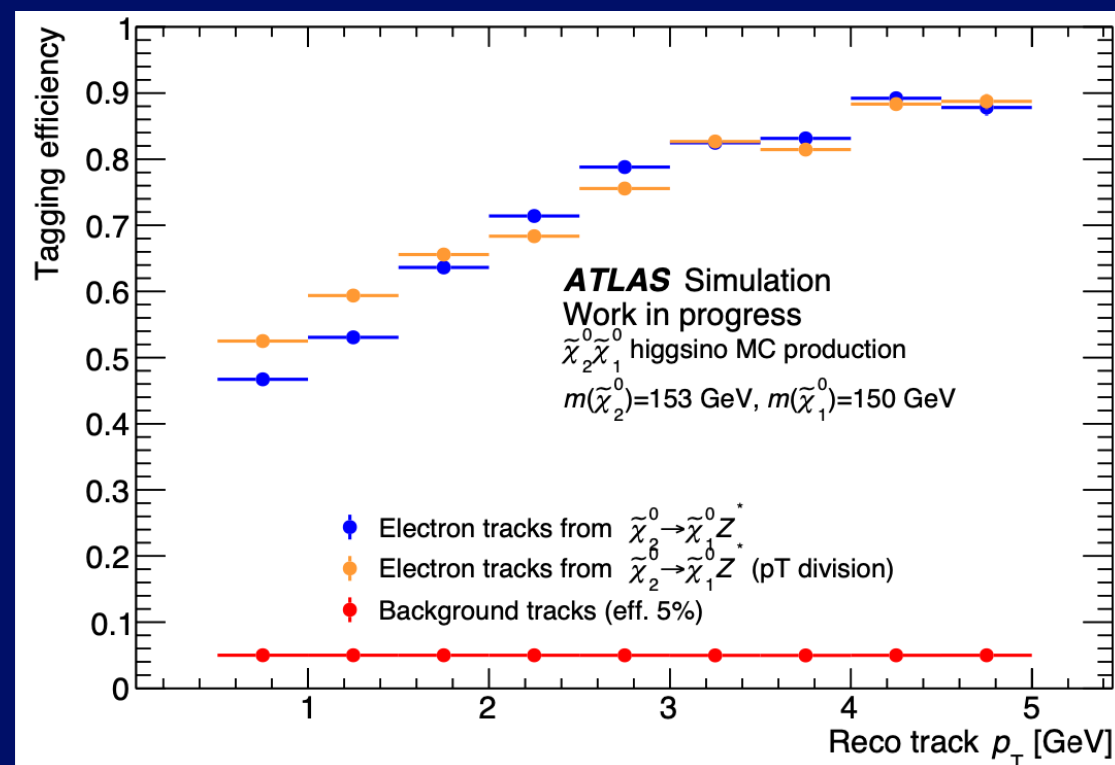
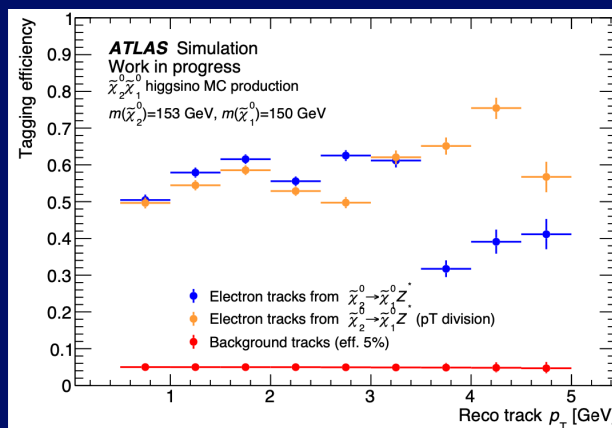
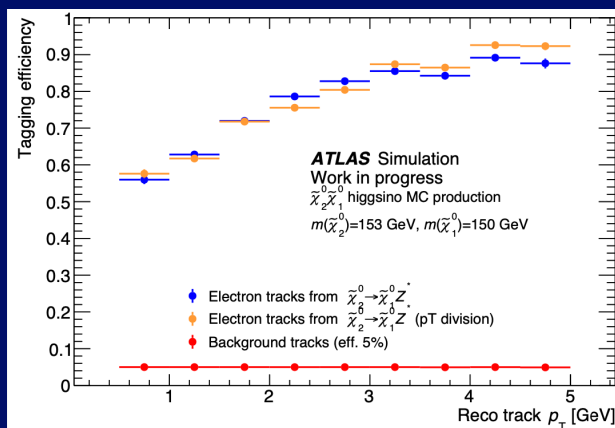
pT の領域分けによる性能の変化

❖ p_T 領域で分けてそれぞれで学習・テストを行うと $p_T < 2$ GeV における識別効率がやや向上する

- p_T 領域によって学習の仕方が異なることが原因と推測
- 一括で学習・テストを行なって同様の性能を実現するには？

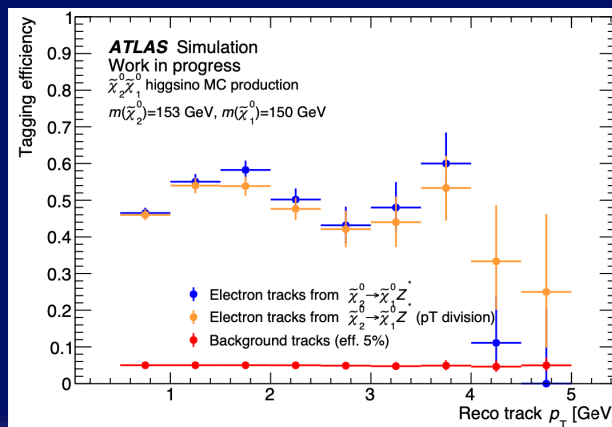
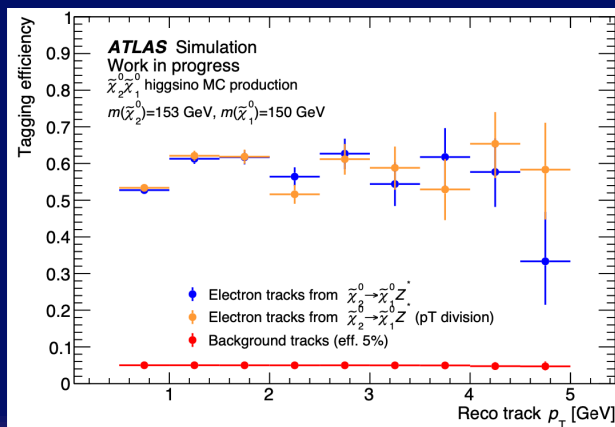
w/ TRT & w/ calo

w/o TRT & w/ calo



w/ TRT & w/o calo

w/o TRT & w/o calo



- ❖ 4つのカテゴリーを1つに統一してなるべくシンプルな構造にしたい
 - カテゴリーごとに学習の仕方が異なる場合（特徴量の重要度が異なる場合）、1つの NN で最大の性能を引き出せるのか
- ❖ シンプルさ vs 性能
 - 経験的に入力変数が多い方が性能が良い場合が多い
 - 入力変数が多いほどそれらの重要度や相関の理解、calibration が大変
- ❖ 実データを用いたバリデーションや誤差の評価

質量が縮退したヒグシーノの探索に向けて、深層学習を用いた超低運動量電子識別器を新たに開発し、シミュレーションサンプルを用いた性能評価を通してその実用可能性を明らかにした

- ❖ $p_T = 0.5\text{--}5$ GeV の超低運動量電子飛跡を識別することに成功し、ヒグシーノ探索の物理アクセプタンスを切り拓いた (5% の fake rate で 40–90% 程度の識別効率)

今後の展望

- ❖ 超低運動量電子識別器の実用化および探索感度の最大化研究
 - データを用いた性能評価
 - 背景事象の理解・精査
 - イベントトポロジーを最大限活用した多変数解析の導入
- ❖ 先を見越した、識別器のさらなる性能向上に向けて
 - p_T などの重要変数に関する扱いの吟味
 - よりハドロン飛跡を落とせるような入力変数やアルゴリズムの追加
 - ・ カロリメータのセル情報を用いた CNN など

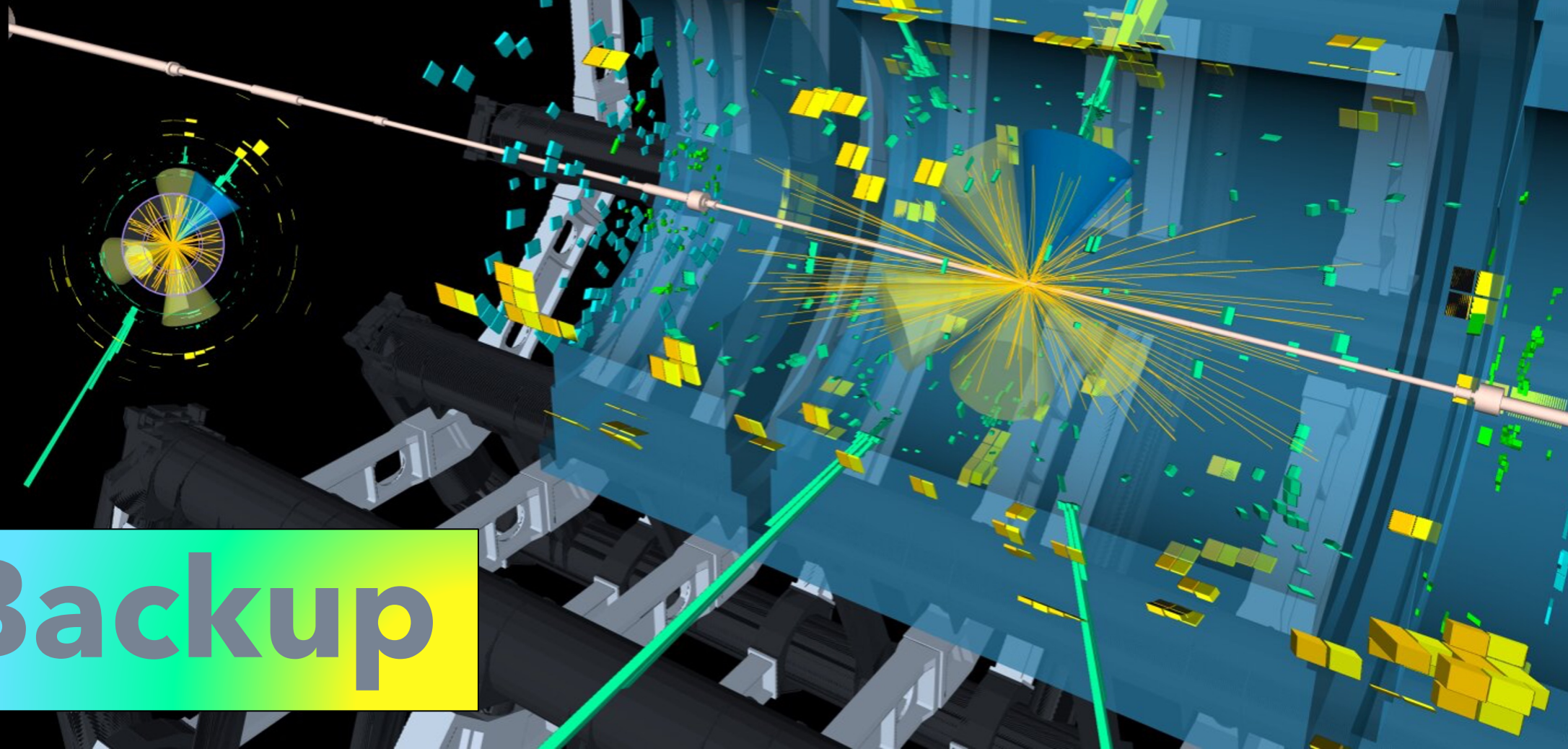


ATLAS EXPERIMENT

Run: 303079

Event: 197351611

2016-07-01 05:01:26 CEST



Backup

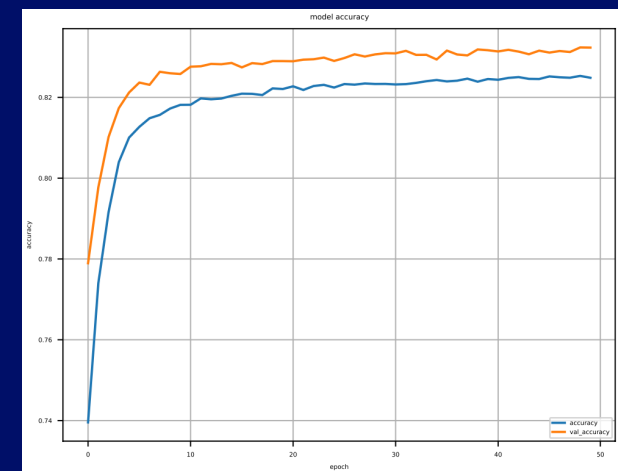
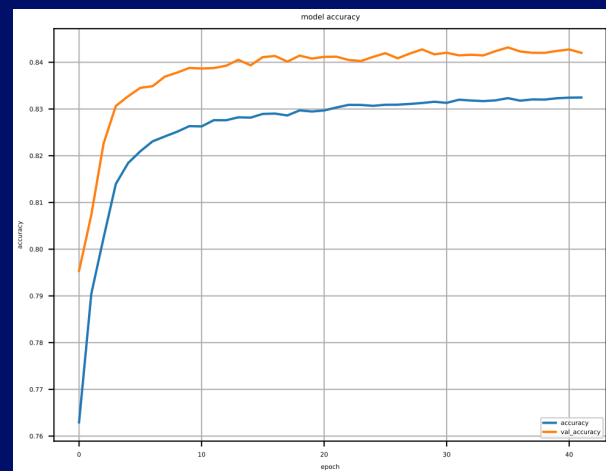
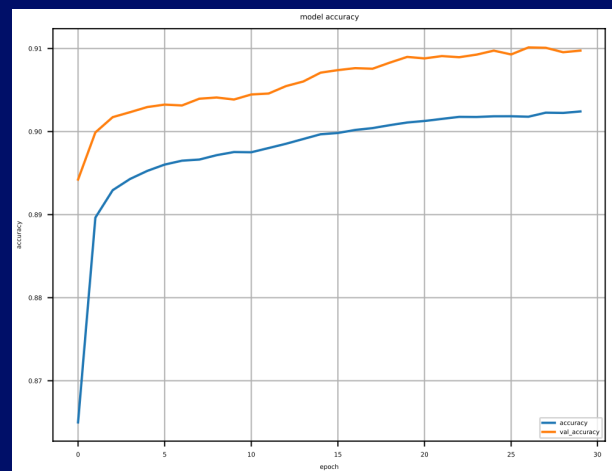
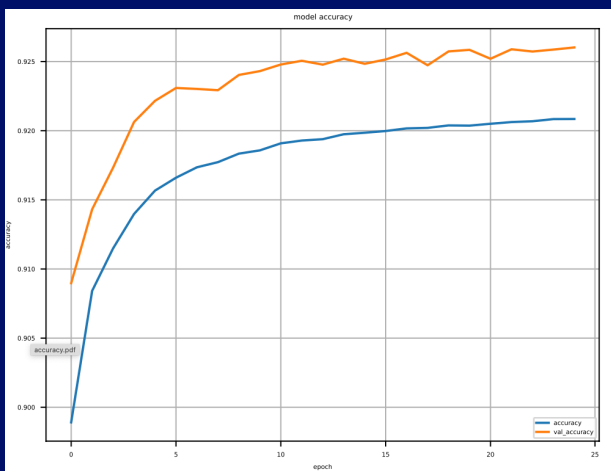
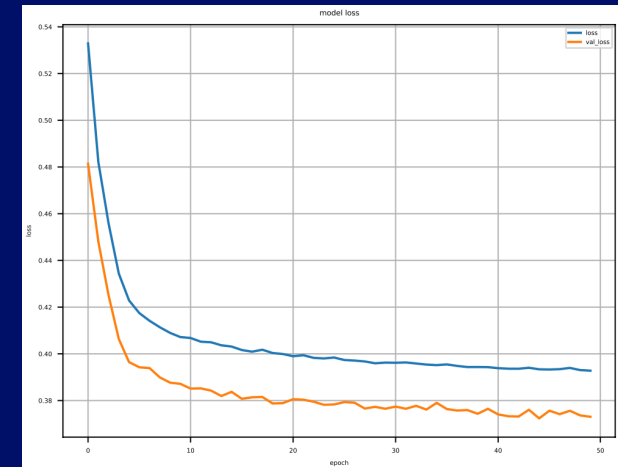
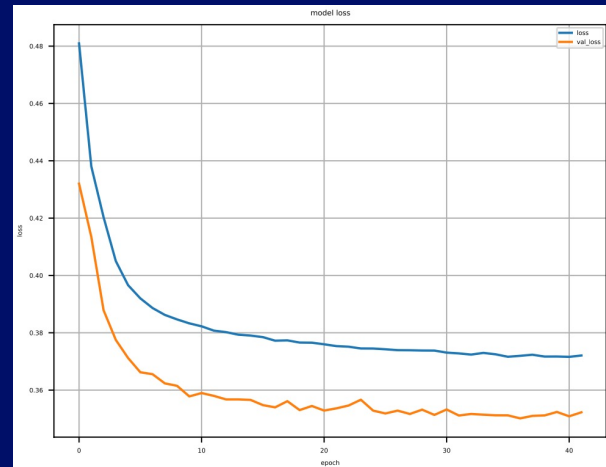
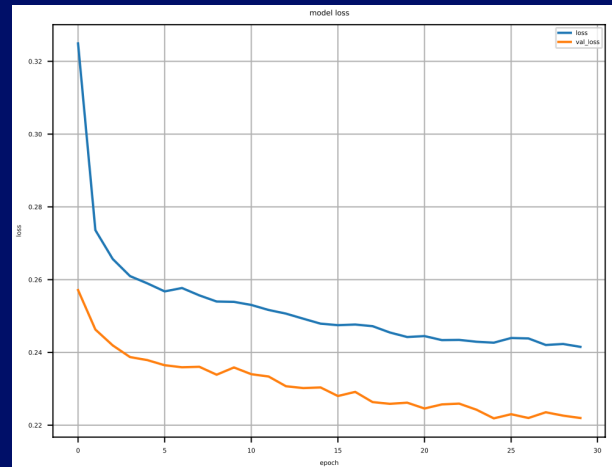
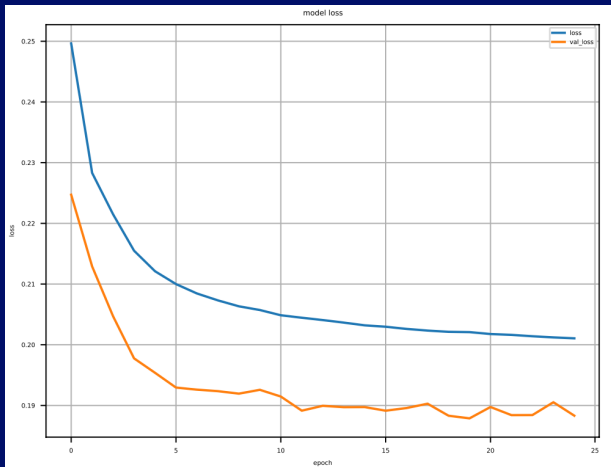
Loss and accuracy for each category

w/ TRT & w/ calo

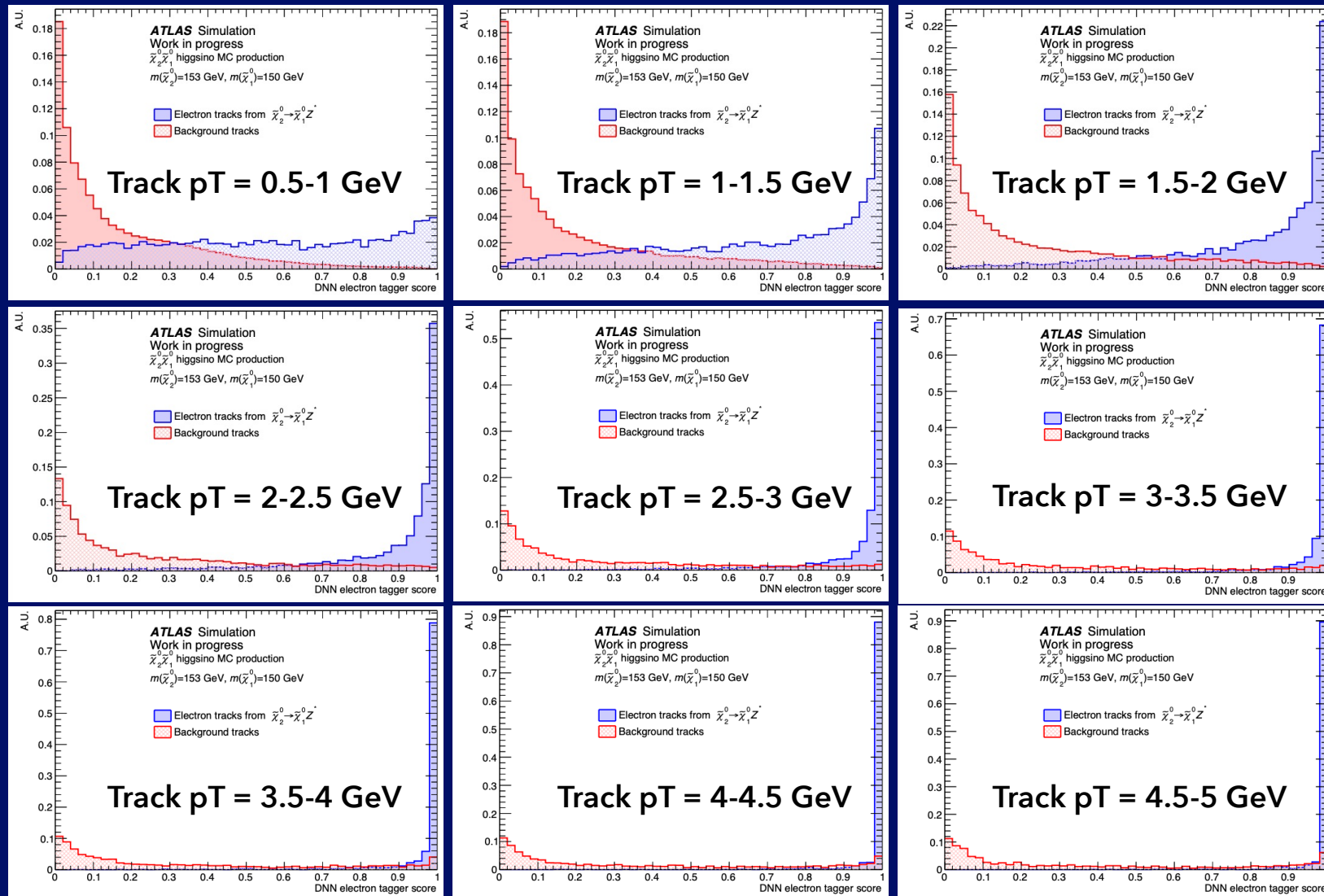
w/o TRT & w/ calo

w/ TRT & w/o calo

w/o TRT & w/o calo

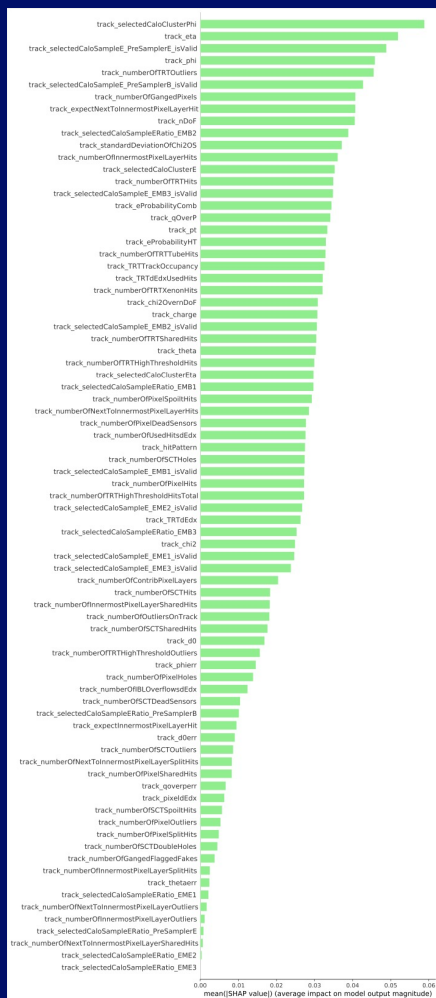


Tagger score for each pT regime

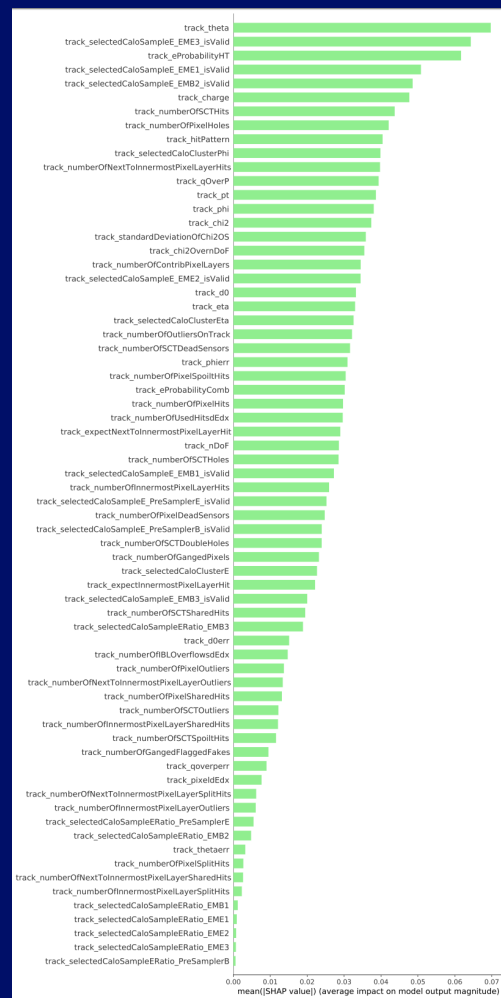


Feature importance using SHAP

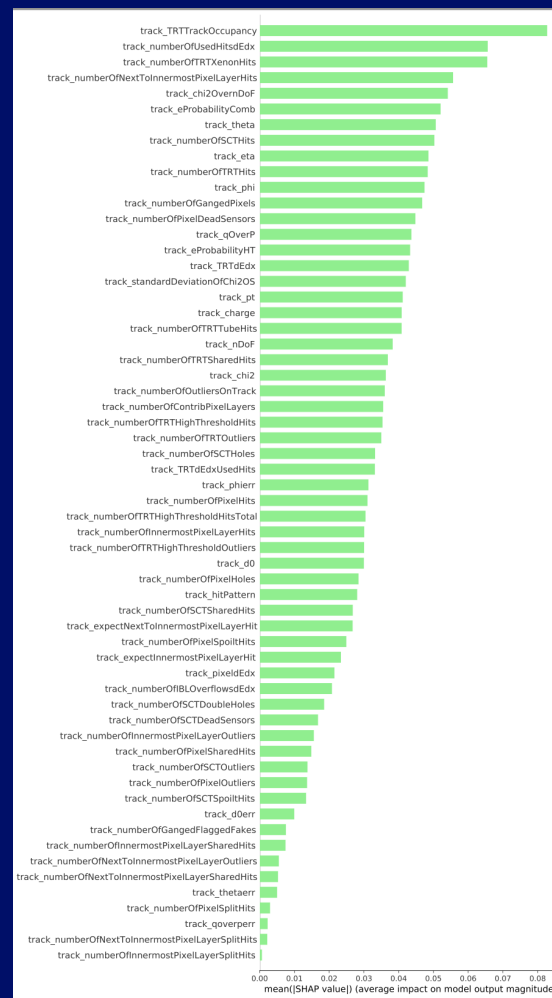
w/ TRT & w/ calo



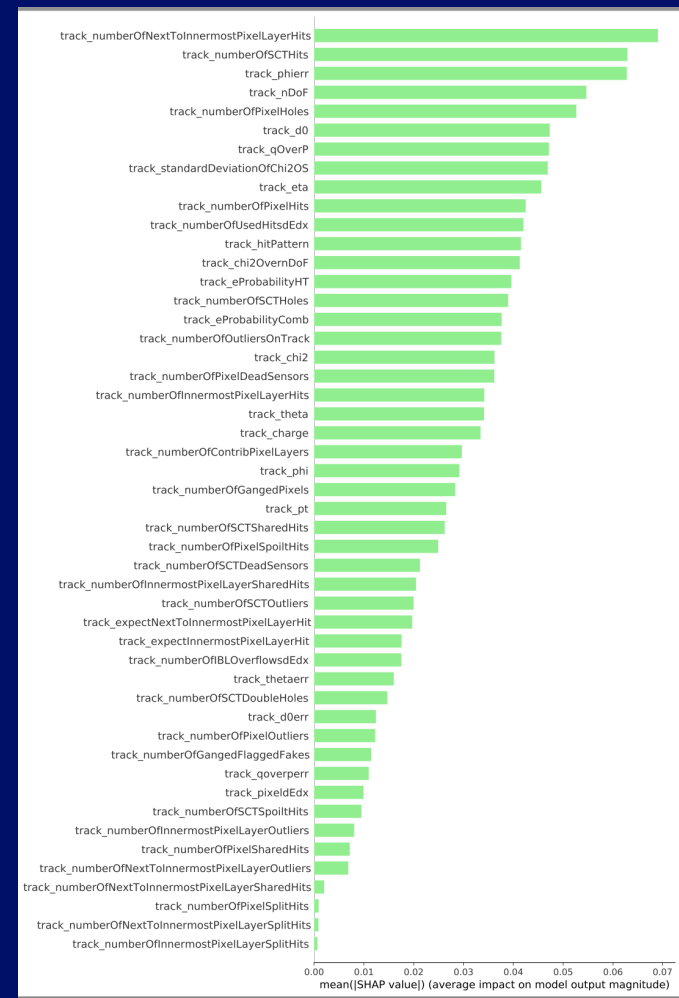
w/o TRT & w/ calo

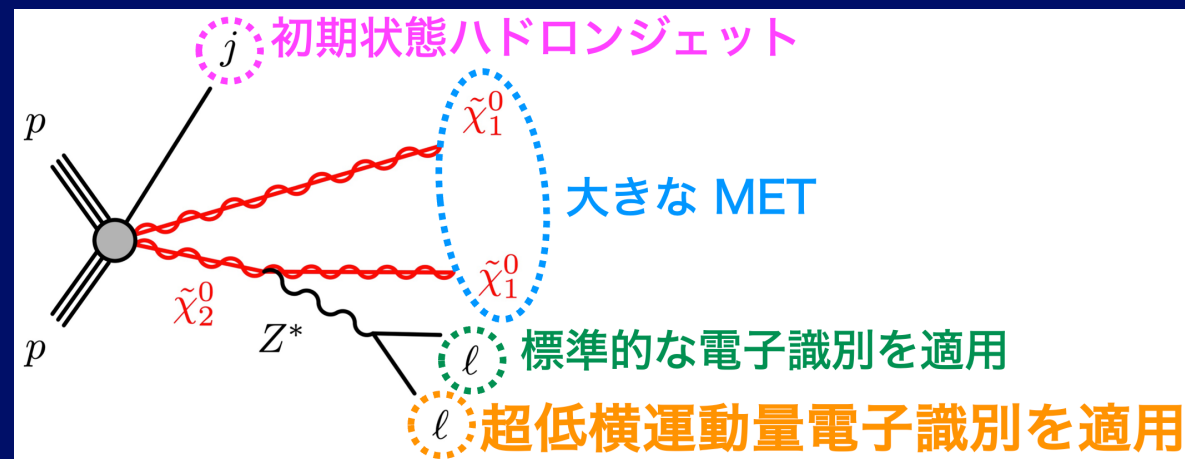


w/ TRT & w/o calo



w/o TRT & w/o calo



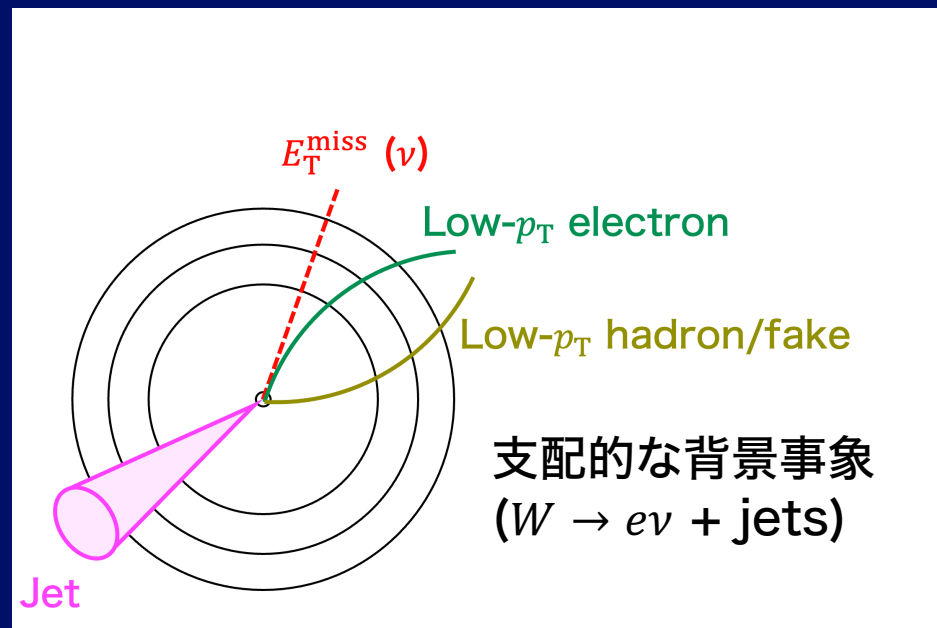
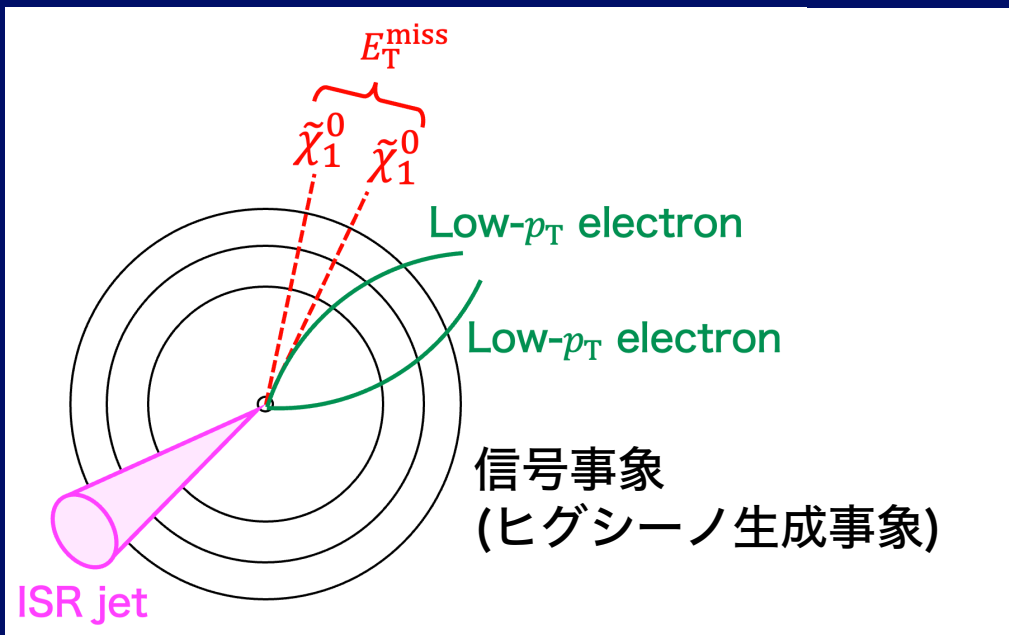
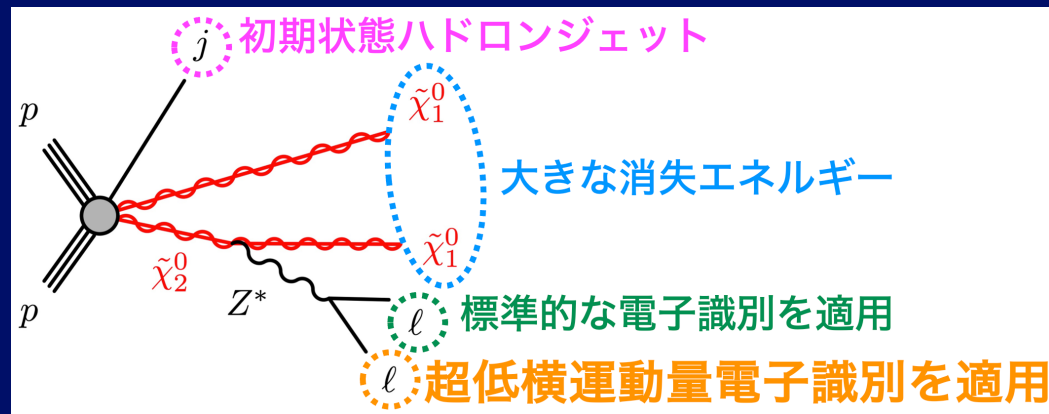


イベント選別の条件 (模擬的な signal region)

- MET > 200 GeV
- Leading jet $p_T > 100$ GeV
- 標準的な電子識別を通過する電子がただ1つ
 - LHVeryLoose ID
 - $p_T < 10$ GeV
- 上記の電子と対をなす Z^* candidate track が1個以上
 - Opposite charge
 - $0.5 < \text{track } p_T < 5$ GeV
 - $\text{PtCone30}/\text{Pt} < 0.06$
 - $0.1 < dR(\text{track}, e) < 1.2$
 - $m(\text{track}, e) < \Delta m(\tilde{\chi}_2^0, \tilde{\chi}_1^0)$
(J/psi veto で 3.0–3.2 GeV は除外)
 - $\Delta d_0(\text{track}, e) < 0.3, \Delta z_0(\text{track}, e) < 0.5$
 - $\Delta\phi(\text{track}, \text{MET}) < 1.5$
 - $(p_{T,\text{track}} + p_{T,e})/\text{MET} < 0.05$
- Z^* candidate track に超低横運動量電子識別を適用し、通過した track があるイベントを信号事象として定義

感度向上度の評価方法

- 先行研究に倣い、模擬的な signal region を定義
 - 標準的な電子と対をなす候補飛跡を選別し、これに超低運動量電子識別器を適用して信号イベントを選別
- 背景事象は $W \rightarrow ev + jets$ を使用
- $\tilde{\chi}_2^0 \tilde{\chi}_1^0$ チャンネルを使用し、 $\tilde{\chi}_2^0 \tilde{\chi}_1^\pm$ も考慮するために生成断面積をスケール
- 全イベント数は Run 2 フルデータの 139 fb^{-1} に規格化



ヒグシーノ探索感度に与えるインパクトの見積り

◆ 超低運動量電子識別器の適用によりヒグシーノ探索感度の大きな向上が見込める

➢ $W \rightarrow e\nu + \text{jets}$ 背景事象の大幅な削減が可能

$$m(\tilde{\chi}_2^0, \tilde{\chi}_1^0) = (103, 100) \text{ GeV}$$

$$m(\tilde{\chi}_2^0, \tilde{\chi}_1^0) = (128, 125) \text{ GeV}$$

$$m(\tilde{\chi}_2^0, \tilde{\chi}_1^0) = (153, 150) \text{ GeV}$$

$$m(\tilde{\chi}_2^0, \tilde{\chi}_1^0) = (155, 150) \text{ GeV}$$

