

# コライダー物理の現象論への機械学習の利用とその効率化の理解

令和4年度 理学研究科 理学専攻 E研, D1, 462201119, 古市亜門, 6月29日 共同研究者: 野尻美保子 (KEK、総研大)、Sung Hak Lim (Rutgers Univ.)



## 研究内容

### 目的

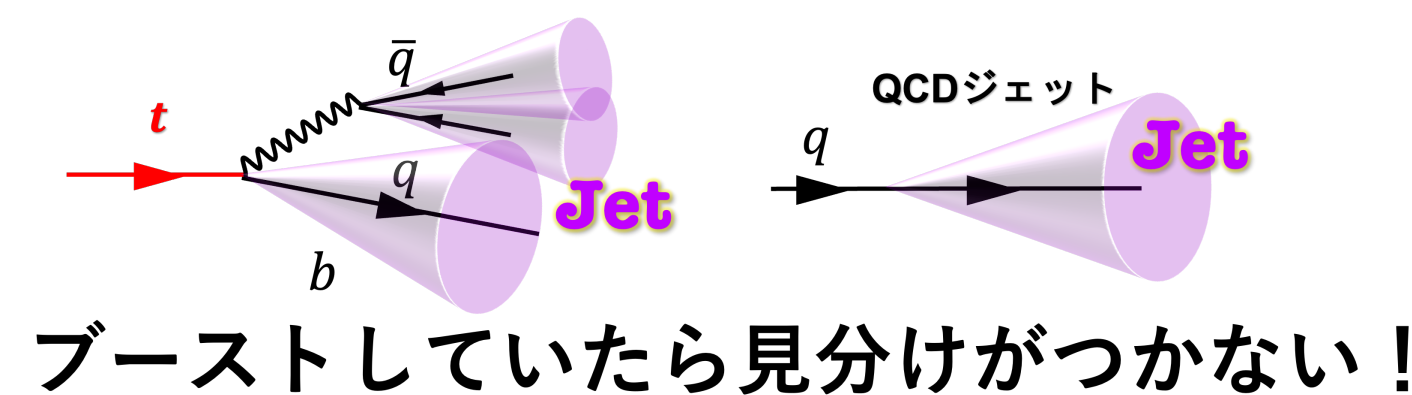
☆ 新物理やSMの精度の高い検証

### 手段・戦略

☆ 高精度なジェットの同定方法の考案・利用  
⇒ ブーストした重い粒子のレアな過程の検証

#### ● ジェット

同じ方向に飛ぶたくさんのハドロン。



ブーストしていたら見分けがつかない!

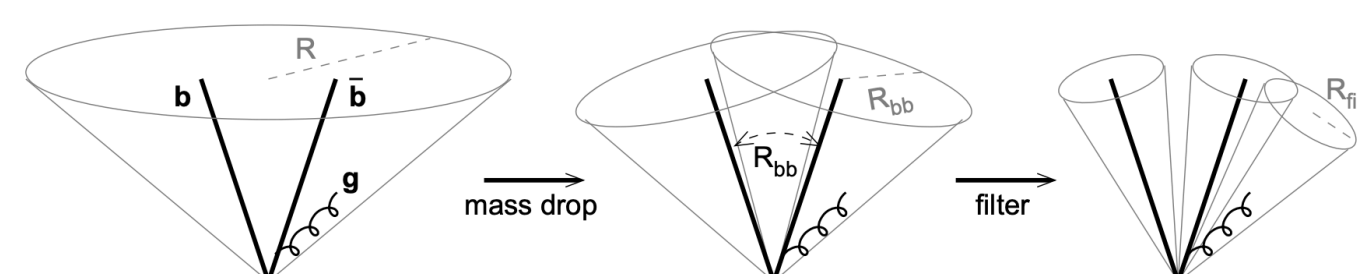
#### ● ジェットの使い所

有効理論や余剰次元模型といった新物理の検証など。

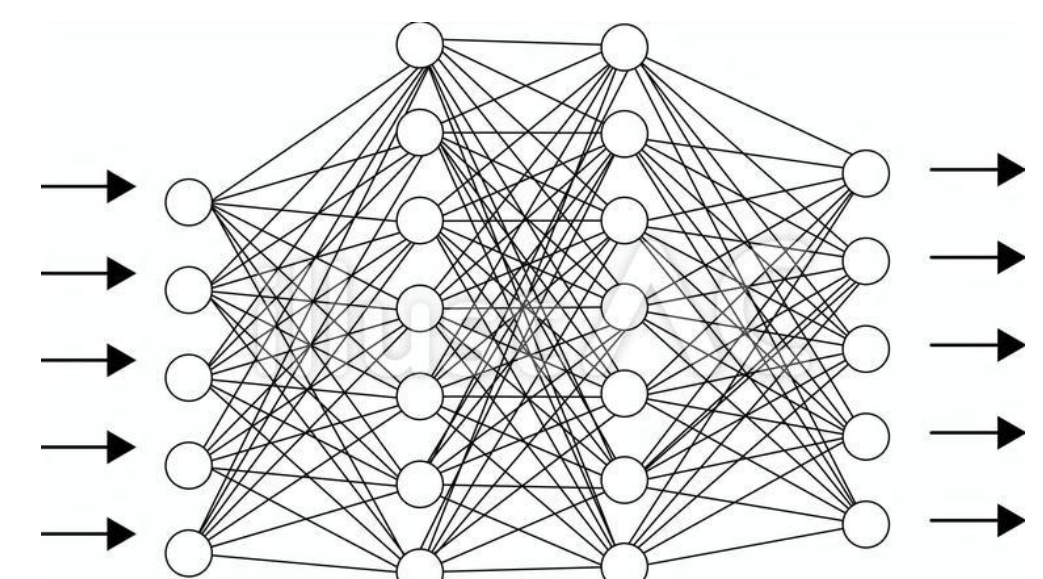
「BSMの重い粒子⇒ブーストしたSMの重い粒子 (t, H, W/Z) ⇒ **ジェット**」

#### ● ジェット分類～堅実なメソッド～

内部構造を定量化。摂動計算可能な量で多変量解析。機械学習を利用。



#### ● ジェット分類～パワフルなメソッド～



low levelな情報で**深層学習**。

分類に重要な情報を自動抽出。

摂動計算できない情報 (粒子数など) も利用。

### 手法・戦術

☆ ジェットのシミュレーションの現象論的補正  
⇒ 深層学習の適切に利用に繋がる

#### ● 「堅実」と「パワフル」の努力

##### 堅実なメソッド

##### 深層学習

● : 分類に重要な情報

● : 分類機に使われる情報

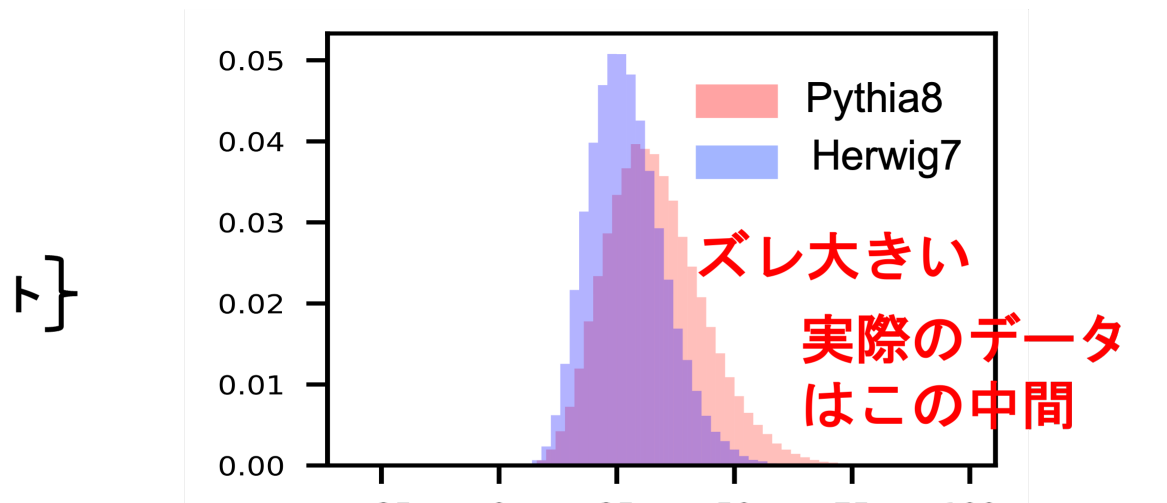
● : 摂動計算可能な情報  
≈ 実際のデータとのズレが小さい情報

➡ : 機械が学習できる情報量の拡張

#### ● 分類機の性能テスト



シミュレーションの答えと照らし合わせて正答率を採点  
⇒ ズレが大きいと適切なテストができない



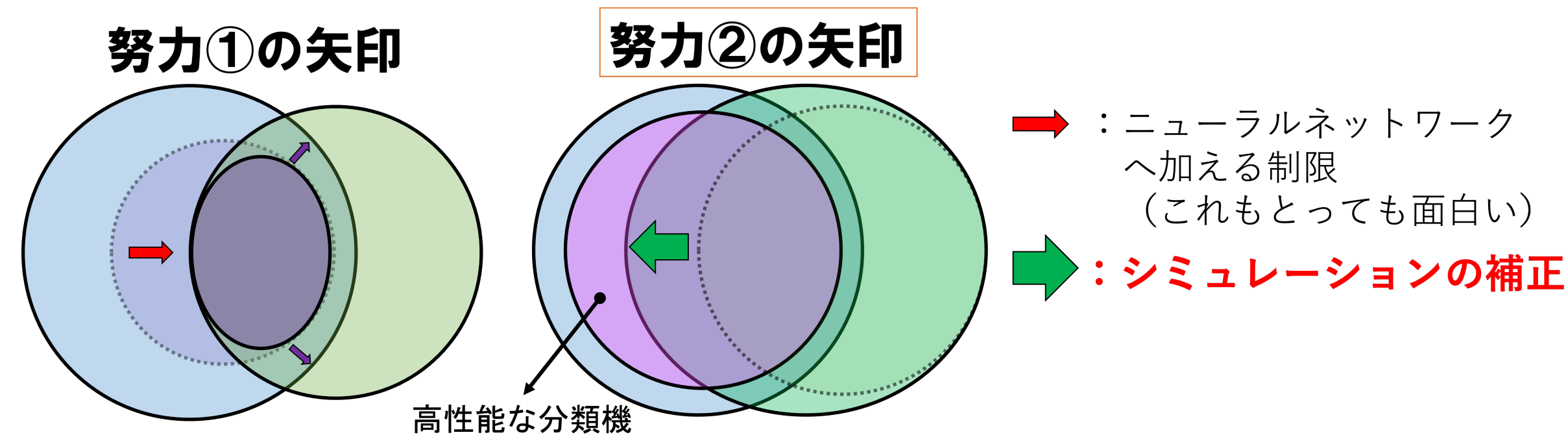
一つのジェットに含まれる粒子数の分布

実際、複雑で高性能な分類機ほどシミュレーションのズレの影響大

## 研究内容 (続き)

### 手法・戦術 (続き)

- 適切なテストのための努力「1」と「2」



### 具体的な方法

☆ ジェットのイベントに**重みw**をつけて補正

$$p_{\text{real}}(x_{\text{jet}}) = w(x_{\text{jet}}) p_{\text{simu.}}(x_{\text{jet}}), \quad x_{\text{jet}}: \text{一つのjetから計算される量}$$

$$p(x): x \text{の生成確率密度}$$

## 実践

### 手順と諸々の設定

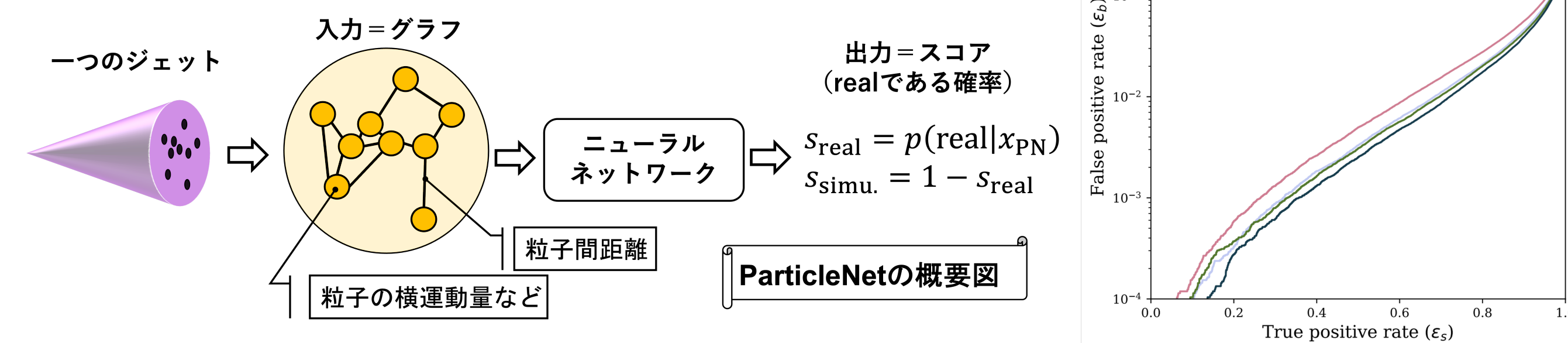
- やりたい事

「分類に重要 ( $p_1$ )」 & 「ズレが大きい ( $p_2$ )」情報 ( $x_{p_1 \wedge p_2}$ ) に関する「イベントごとの重み ( $w(x_{p_1 \wedge p_2})$ )」を推定し「シミュレーションを補正」

$$w(x_{p_1 \wedge p_2}) = \frac{p_{\text{real}}(x_{p_1 \wedge p_2})}{p_{\text{simu.}}(x_{p_1 \wedge p_2})}, \quad p_{\text{real}}(x_{p_1 \wedge p_2}, y) = w(x_{p_1 \wedge p_2}) p_{\text{simu.}}(x_{p_1 \wedge p_2}, y)$$

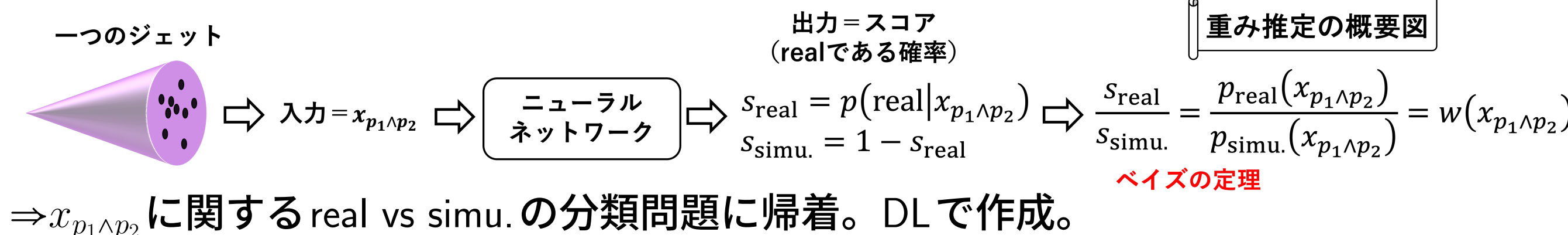
- Step1:  $p_1 \wedge p_2$ の情報の抽出と定量化

real vs simu. の分類機を高性能なDL: **ParticleNet**で作成



「ParticleNetがreal (simu.) だと判断したreal (simu.) のジェット」を調べて、特徴を定量化 ( $x_{PN} \rightarrow x_{p_1 \wedge p_2}, x_{p_1 \wedge p_2} \subseteq x_{PN}$ )。

- Step2: 重み (密度比) 推定



⇒  $x_{p_1 \wedge p_2}$ に関するreal vs simu. の分類問題に帰着。DLで作成。

- Step3: 補正の確認 (簡易ver.)

ParticleNetのスコアを求めた重みで補正

$$\frac{s_{\text{simu.}}^{(\text{new})}}{s_{\text{real}}^{(\text{new})}} \stackrel{\text{def}}{=} w(x_{p_1 \wedge p_2}) \frac{s_{\text{simu.}}(x_{PN})}{s_{\text{real}}(x_{PN})} = \frac{p_{\text{real}}(x_{p_1 \wedge p_2}) p_{\text{simu.}}(x_{PN})}{p_{\text{simu.}}(x_{p_1 \wedge p_2}) p_{\text{real}}(x_{PN})} \stackrel{?}{\rightarrow} 1 \Leftrightarrow s_{\text{simu.}}^{(\text{new})} \stackrel{?}{\rightarrow} 0.5 (\text{理想})$$

⇒ ランダム推定にどれくらい近づくか?

## 実践 (続き)

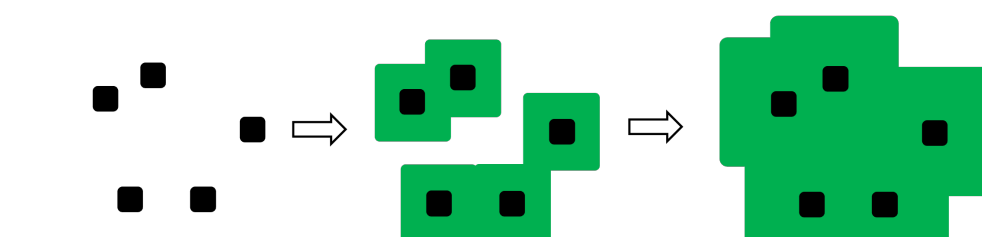
### 結果

☆ 結果 (シミュレータ間の違いの解析)

\* 実際のジェットデータは手元にないので異なるシミュレータで対応。  
real⇒Pythia (PY)、simu.⇒Herwig (HW)

#### 一、点の数、密集度合い

- ⇒ **ミンコフスキー汎関数の利用 (MF)**  
銀河分布や分子分布などで利用される。点の集合を段階的に粗視化し「島」の数と密集度合いなどを定量化。



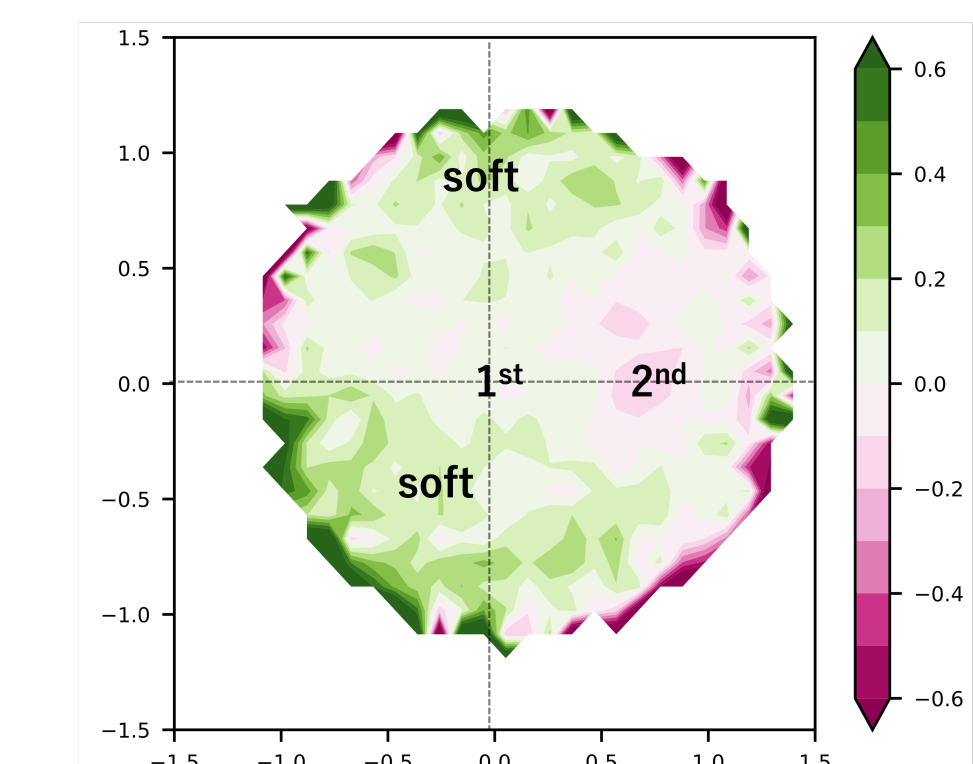
MFの段階的な粗視化 (イメージ)

#### 二、第一と第二サブジェット

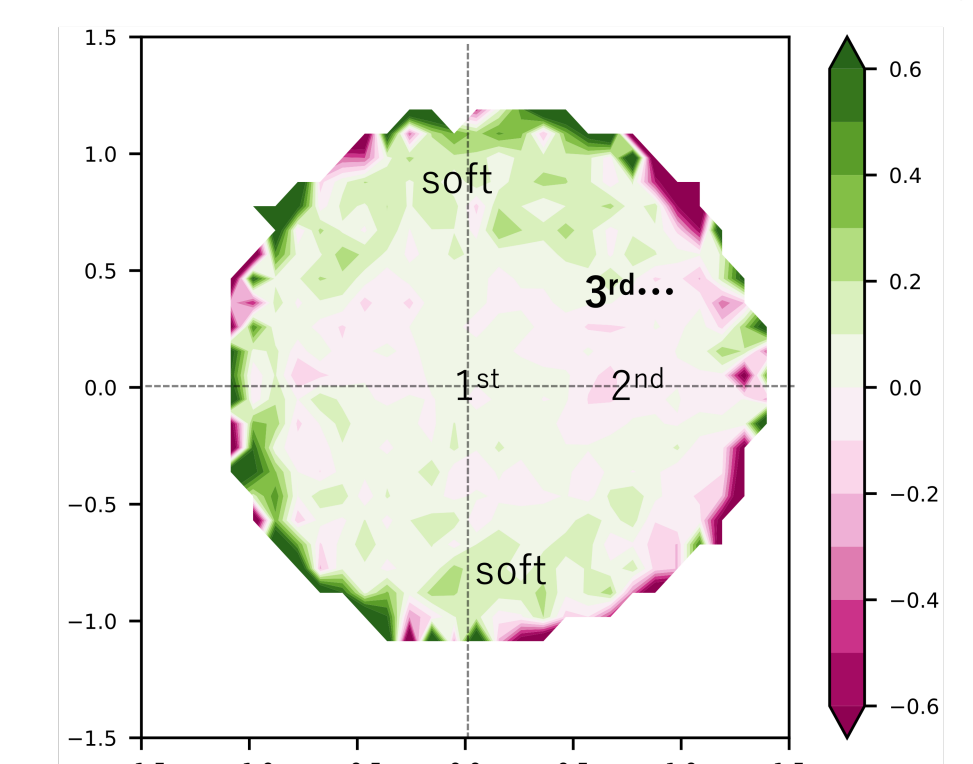
- ジェット内の最も強い二本の軸。  
⇒ **二点エネルギー相関の利用 ( $S_2$ )**

#### 三、第三以降のサブジェット

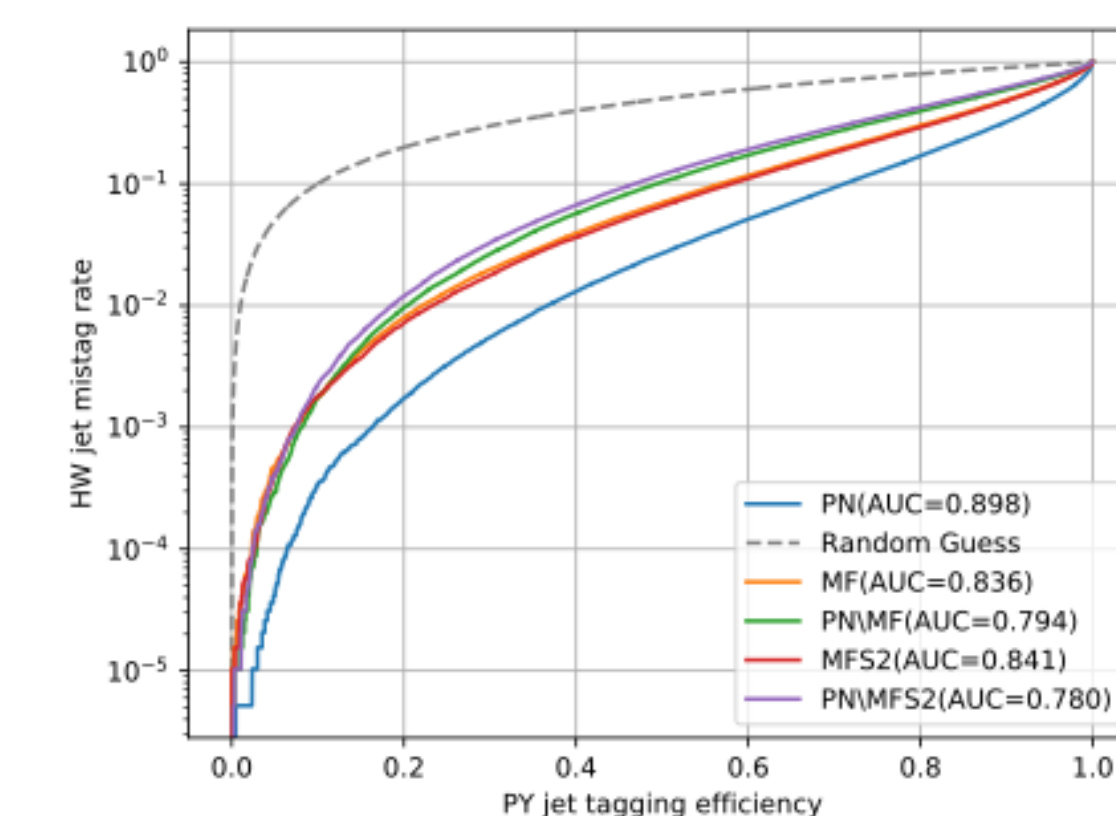
- サブジェット同士のペアの区別が重要か。  
⇒ **三点相関以上の情報の利用 ( $S_3$ )**、  
**各方角のエネルギー量 (zone-count)**、  
**short-MF: 前研究よりも短距離のMF**



正規化したジェットの平均的な差。(PY - HW)/(PY + HW)。  
ParticleNet内で使われた情報の違いを抽出。



ParticleNetをMFと $S_2$ に関して補正した後のジェットの平均的な差。第二、第三ジェットの部分などに違いが残っている。



分類機の補正の結果 ( $x_{p_1 \wedge p_2} = MF + S_2$ )  
ランダム推定になれば完全な補正。  
まだまだ補正の余地はあり。  
⇒  $x_{p_1 \wedge p_2} = MF + S_2 + [\text{short-MF}] + [\text{zone-count}] + S_3$

## これから

### 現研究のその先

- IRC-safe (soft-collinerな分岐にインセンシティブ ≈ 摂動計算可能) な情報の徹底した補正
- SoTAの分類機ParticleTransformer (人工知能分野でもSoTAのTransformerを応用したもの) での試行、
- その他のシミュレータ (Sherpa) での試行

### その他興味があること

- DLの生成モデルでジェットのシミュレーション (VAE、GANなど)
- IRC-safeを強力に保証した分類機の作成
- 対称性を考慮したニューラルネットワーク (Lorentz対称性)
- 上記を組み合わせた深層学習モデルの作成