

ミンコフスキ一汎関数を  
用いた深層学習  
(深層学習から何を学ぶか)

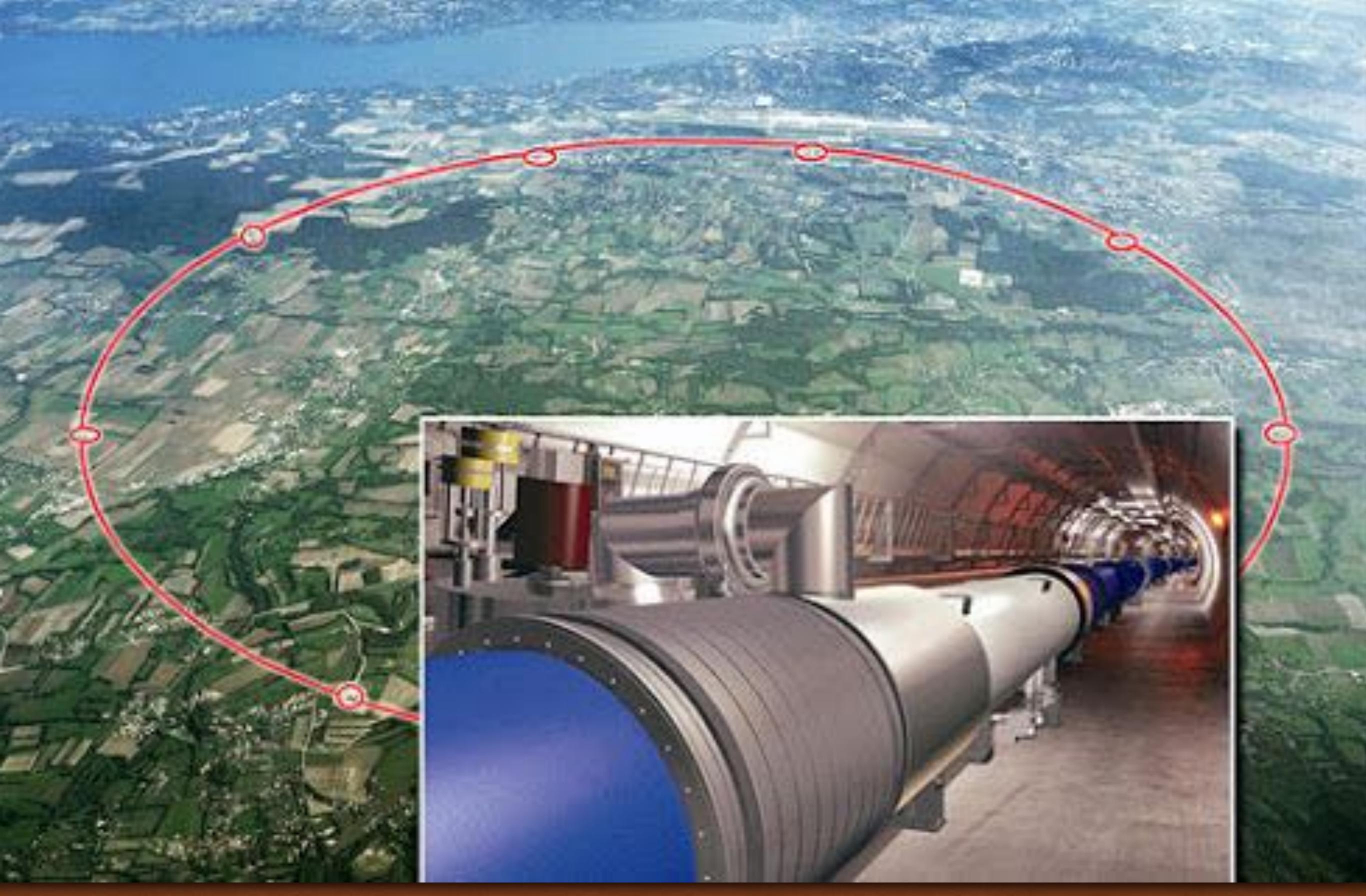
KEK 野尻美保子

(Deep Learning 2020, 10/29, 2020)

# 自己紹介

- 専門 コライダー物理・ダークマター（理論） LHC, ILC, 宇宙線によるダークマターの探索
- コライダー物理 高いエネルギーの粒子を衝突させることによって、より高いエネルギーの物質像を探る。
- LHC ヒッグス粒子を発見(2012) さらにヒッグス粒子の精密測定、暗黒物質の探索などを通じて素粒子の性質を究明
- e-Print 2010.13469, JHEP 07 (2020) 111, JHEP 07 (2019) 135 JEP 10(2018) 181 等に基づいてジェットの深層学習の解釈可能性を調べている。
- 最近やってる別のこと Jeans 方程式をGANで解いて、暗黒物質密度を決める。（また呼んでください。）
- なぜ深層学習を始めたか（もごもご）

# LHC実験とは



# コライダー物理と機械学習・深層学習

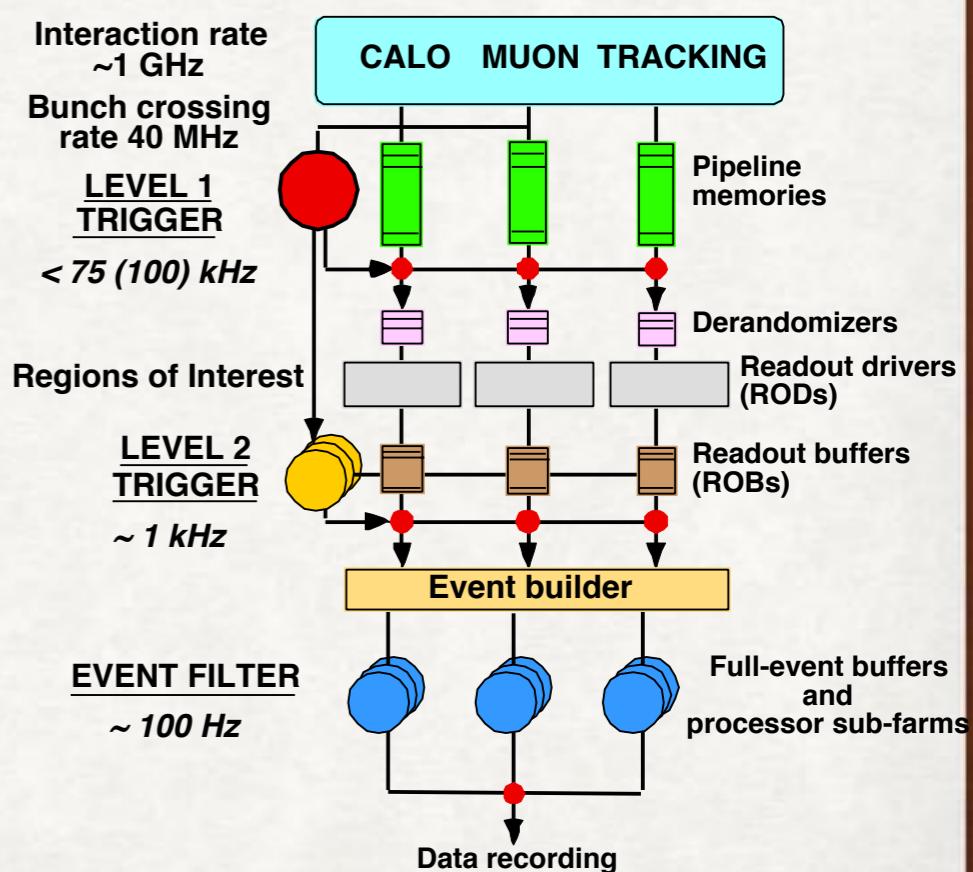
理論の予言

大量のデータ  
実験

新物理（ダークセクターなど）

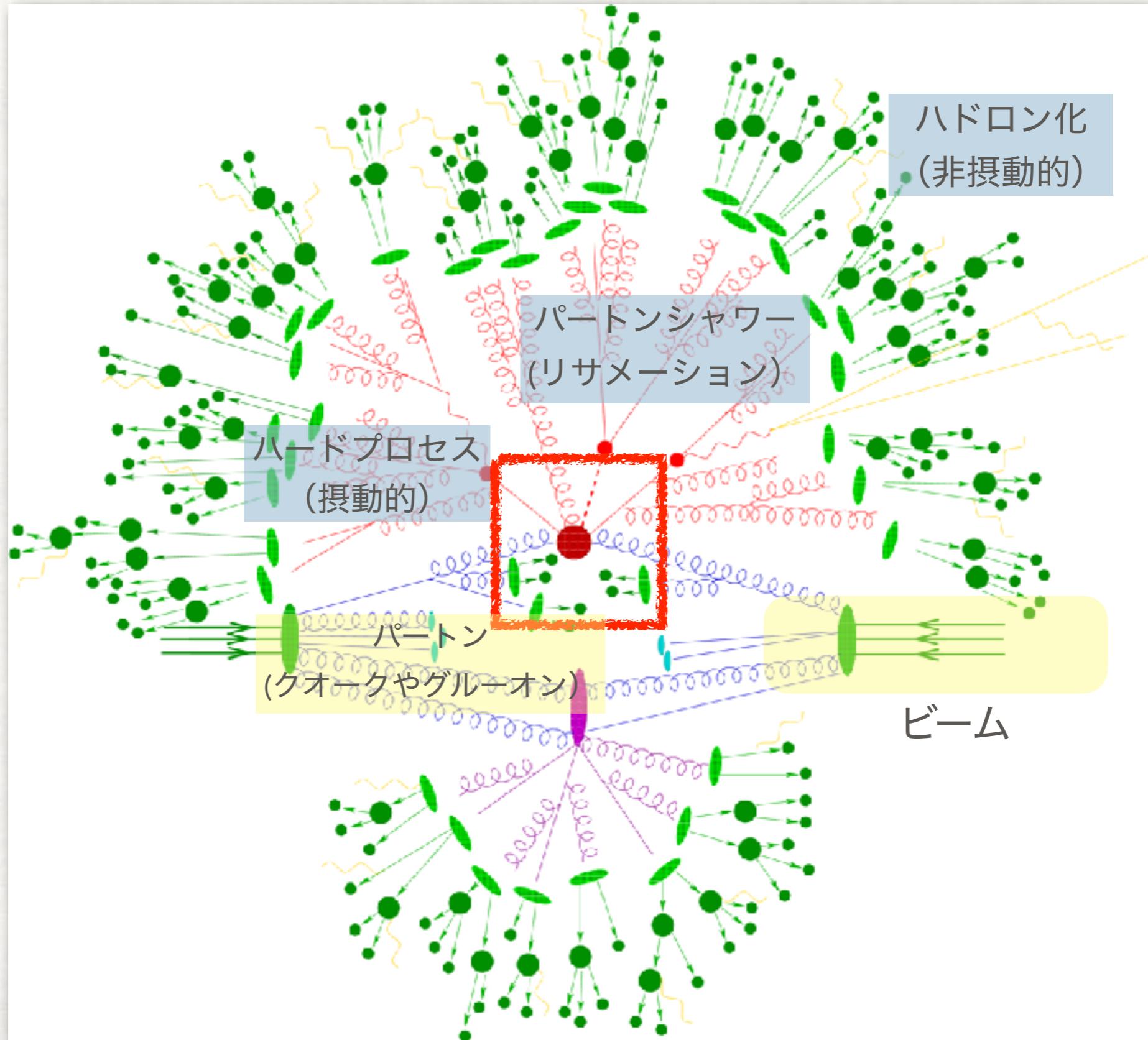
- 「一つしかない自然の理解」 系統誤差のタイミングのよい理解が重要 目標は 0.1%
- ハドロンコライダー特有の難しさ
- 測定側 自動化・高速化(1GHz でイベントが来る、衝突点が100 あるとか) HL-LHC 30倍のデータがくる
- 理論側: QCD の理解
- 機械学習: BDT (決定木) 2002 ~ 深層学習(CNN 等) 5年ほど特に粒子飛跡の解明で大きな成果

LHC実験はデータ取得も大変

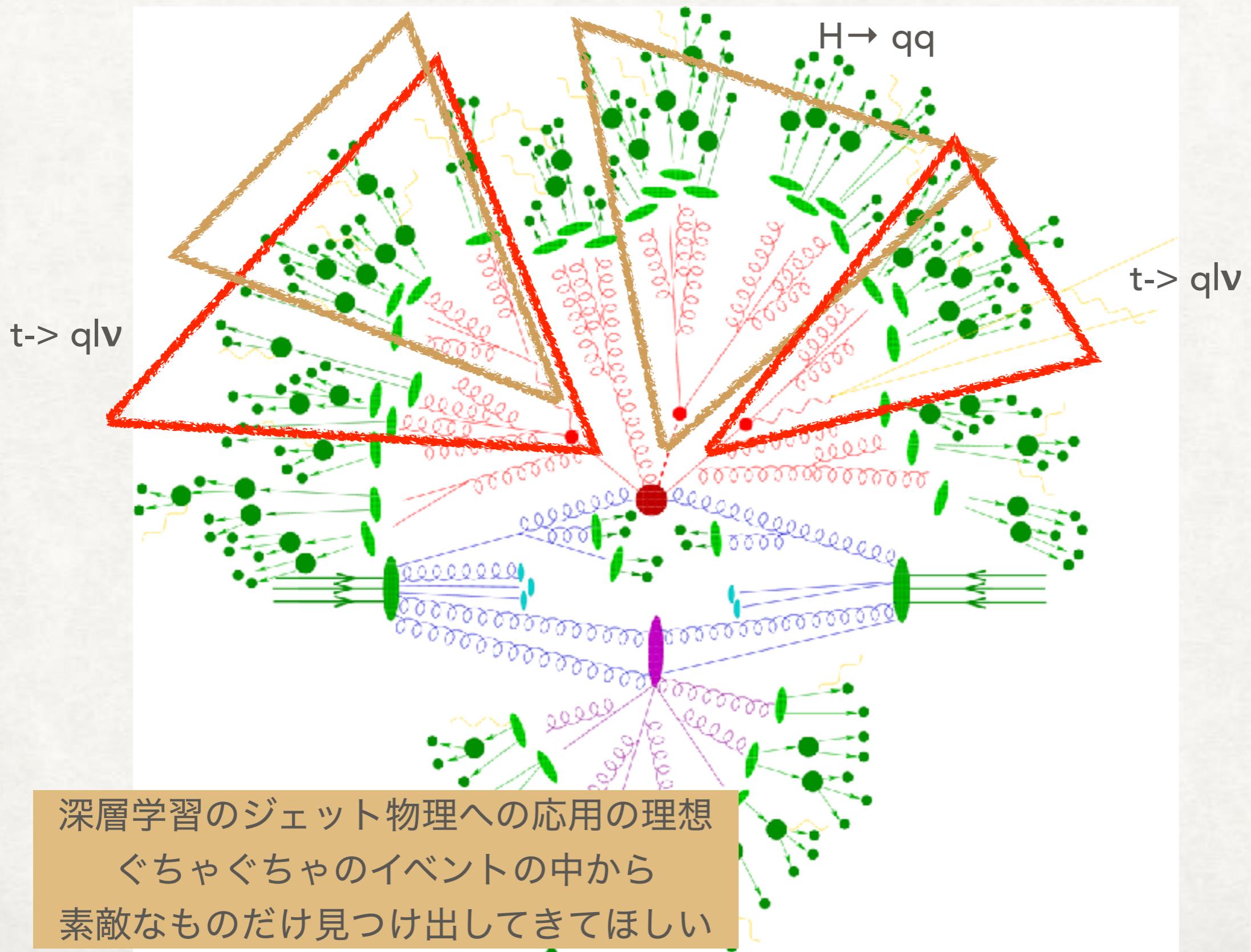


1999年

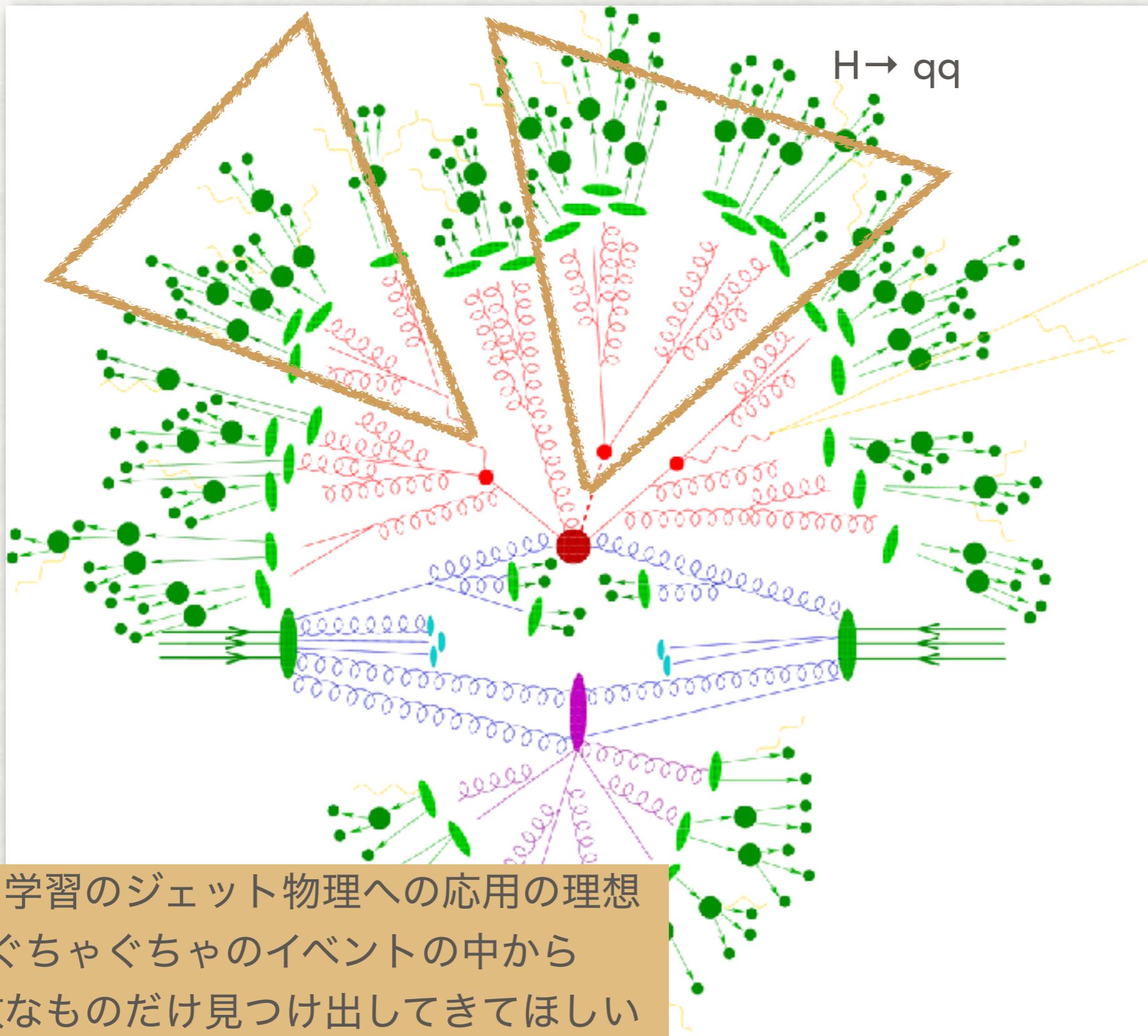
# LHC の物理 (ミニマム)



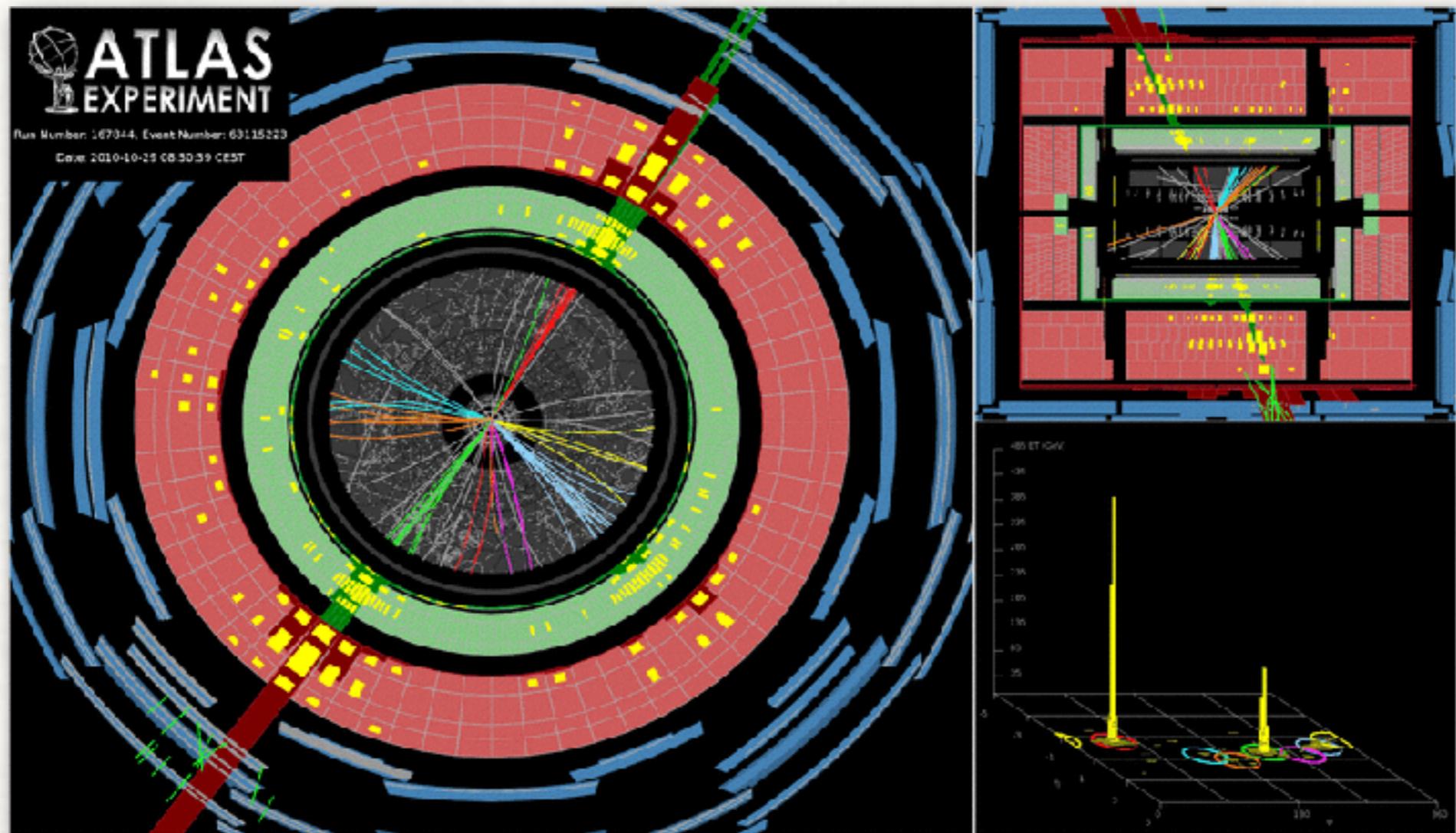
# イベントに隠れる特殊な粒子!



# イベントに隠れるカラーを持たない粒子



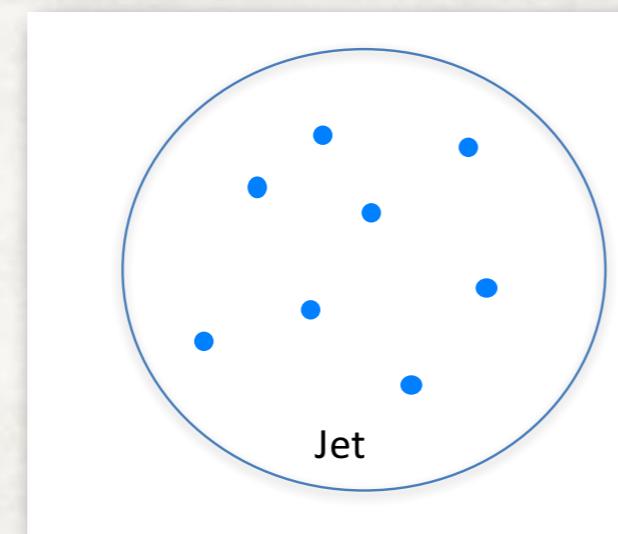
# LHCイベント ジェットとジェットイメージ



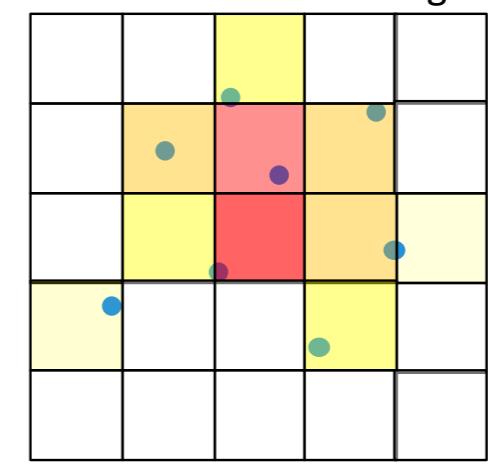
Jet image

$(\eta, \varphi, p_T)$

$$R_{ij} = (\eta_i - \eta_j)^2 + (\varphi_i - \varphi_j)^2$$



Jet Image



$\eta$

$\varphi$

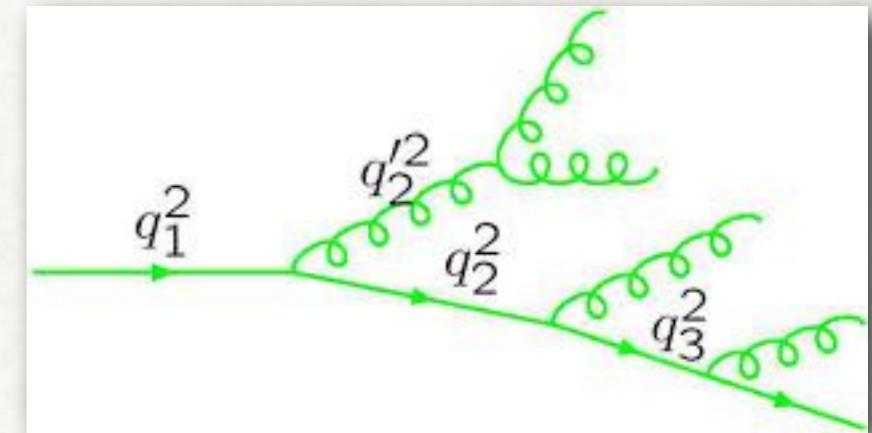
# パートン (q g) → ジェット

- ・ パートンシャワー

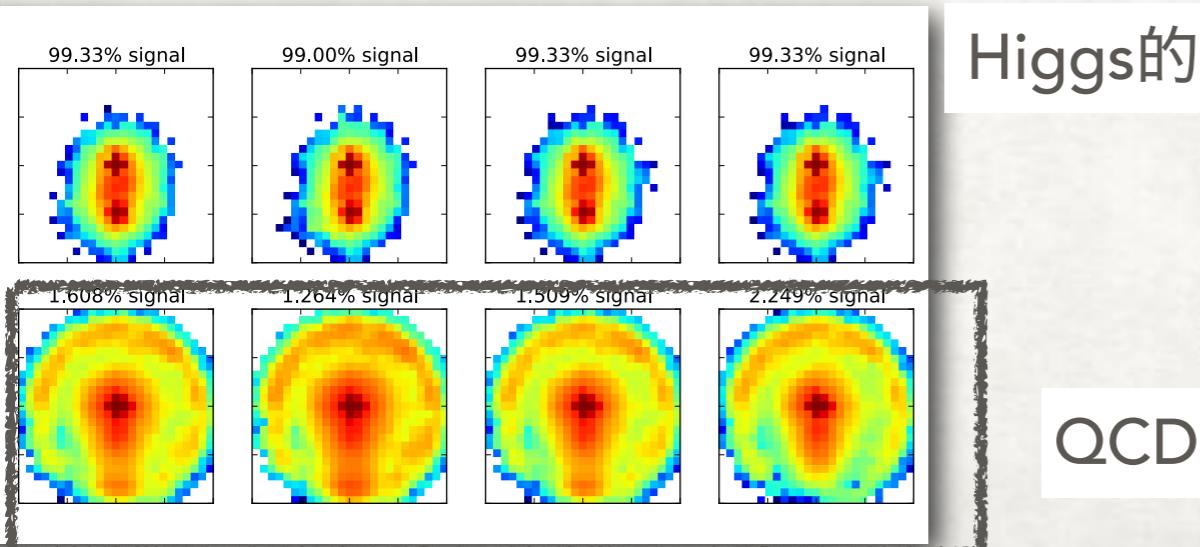
- ・ クオーケやグルーオンがハードなプロセスから生まれるとそこからたくさんのクオーケ、グルーオンが生成される仕組み

- ・ 運動量の関係

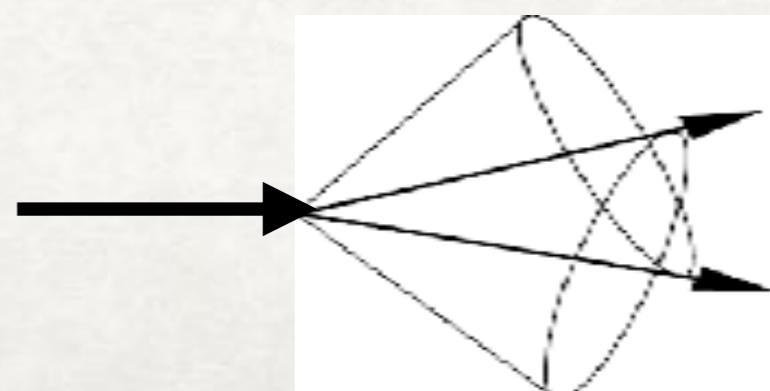
- ・  $q_1^2 \rightarrow q_2^2 \rightarrow q_3^2 \dots$  (mass ordering)
- ・  $p \rightarrow p_1 + p_2, (p^2)^{-1} = (2 p_1 p_2)^{-1}$  散乱断面積は  $p_1 p_2$  が同方向 (collinear) あるいは  $p^2 \sim 0$  で発散 (Soft Collinear Singularity)
- ・  $\theta_1 > \theta_2 > \theta_3 \dots$  (angular ordering) Higgs 粒子 W 粒子 (weak boson)  $\rightarrow q\bar{q}$  parton shower は、 $q\bar{q}$  の間にだけである



$$C_0 \alpha \text{Log}^2(q^1/q^2) + C_1 \alpha \text{Log}(q^1/q^2)$$

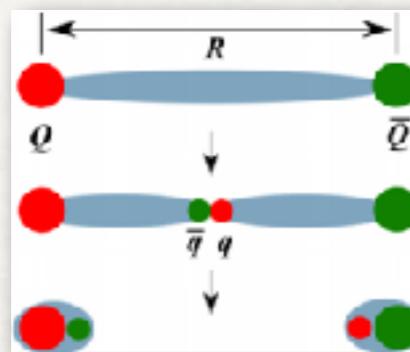


カラーシングレット粒子の崩壊は  
ロングレンジでみるとQCDで粒子を放出しない。

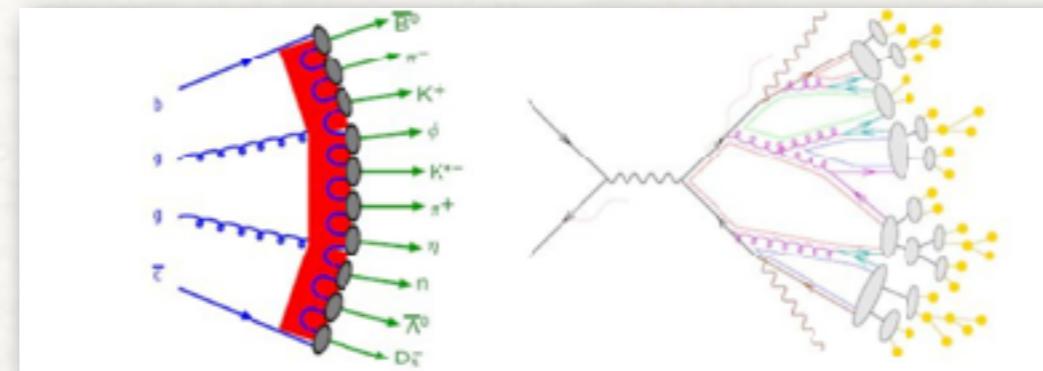


- ・パートンからハドロン(モデルあれど理論なし)
  - ・Quark gluon 同士をつなぐカラーが伸びると  $q\bar{q}$  が生まれてハドロンになる

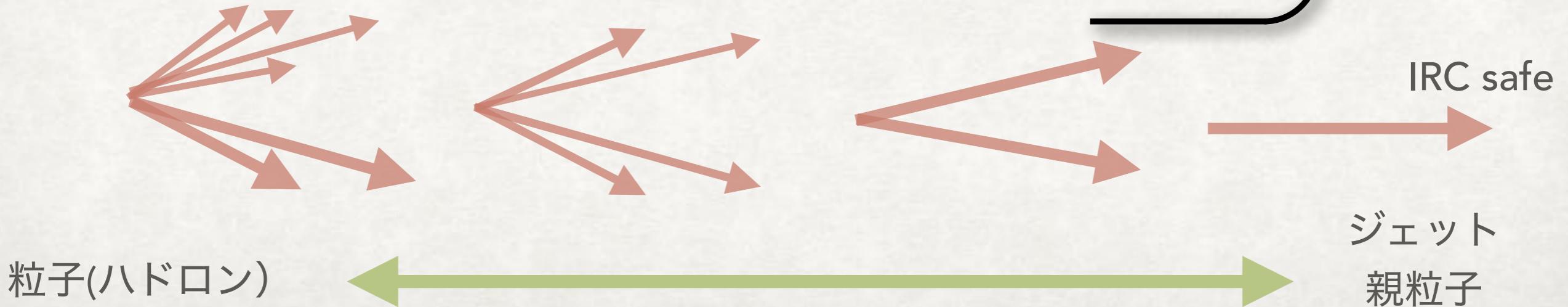
ストリングモデル



ストリングモデルとクラスター モデル

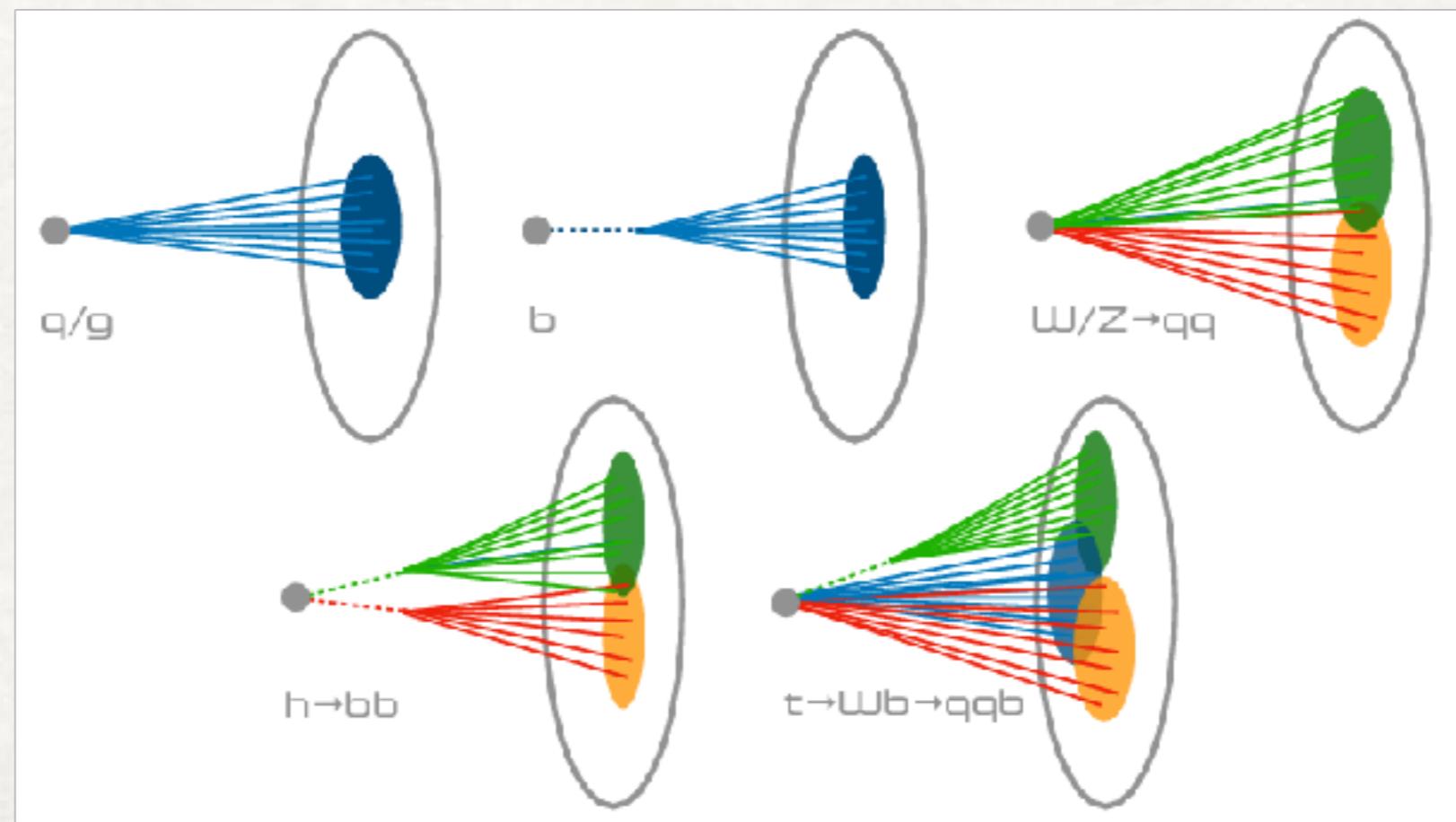


- ・Jet clustering (粒子を相互の関係によってまとめて、ひとかたまりのオブジェクトにまとめる)
  - ・Parton が 2 つにわかれても、ハドロンになってもアウトプットが同じであるようなアルゴリズムを探す(IRC safe)



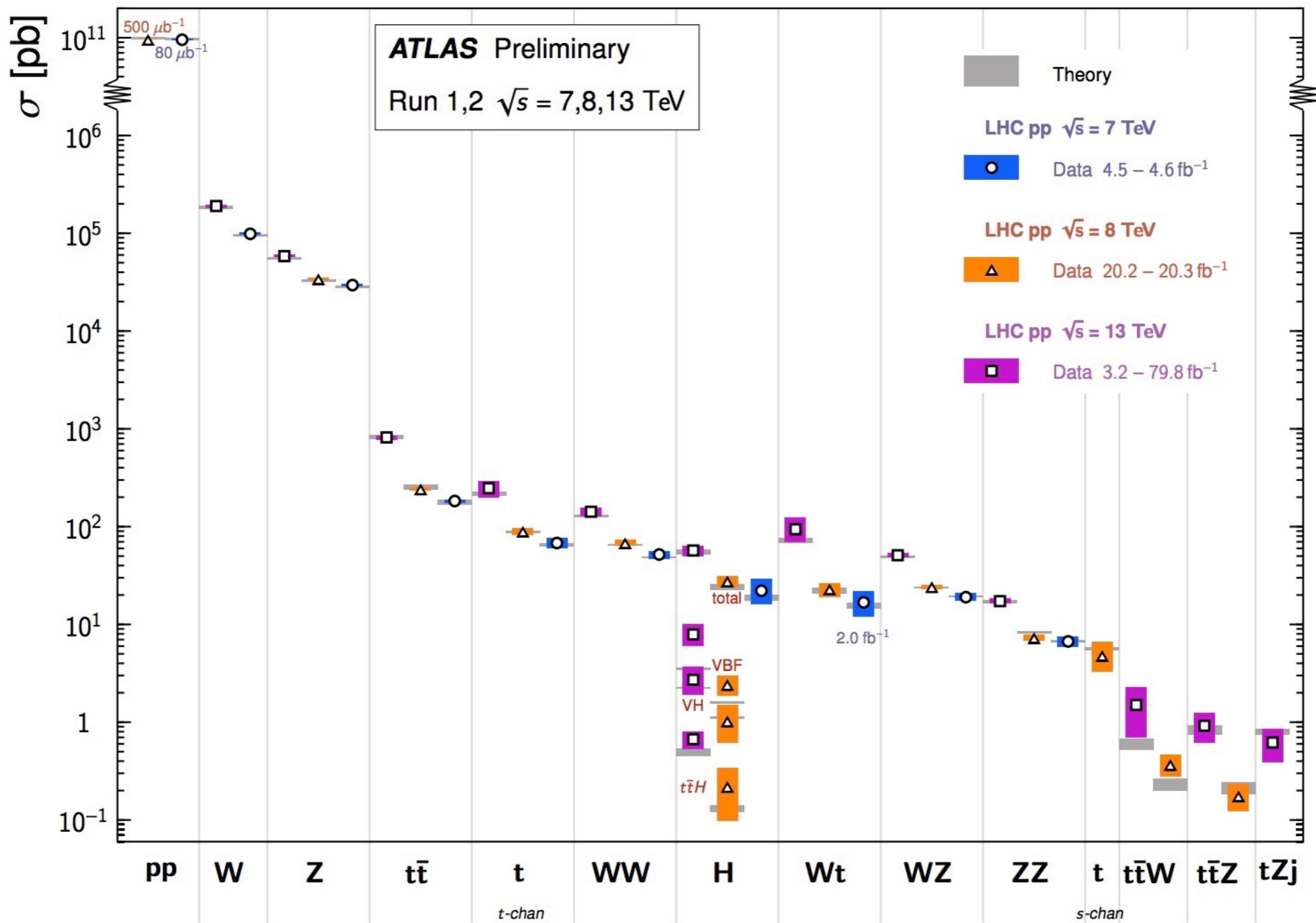
# 今後のLHC 高輝度化

- より高いエネルギーをもつ top quark や、Higgs boson などに興味が移ってくる
- QCD ジェットと boost された top Higgs などの重たい粒子がよく似ている。ここに機械学習や深層学習を使うことが注目されている。



# Standard Model Total Production Cross Section Measurements

Status: July 2018



# 機械学習とPARTICLE TAGGING

## ジェットの特徴量をつかった BDT

ジェットの運動量, 質量, ...

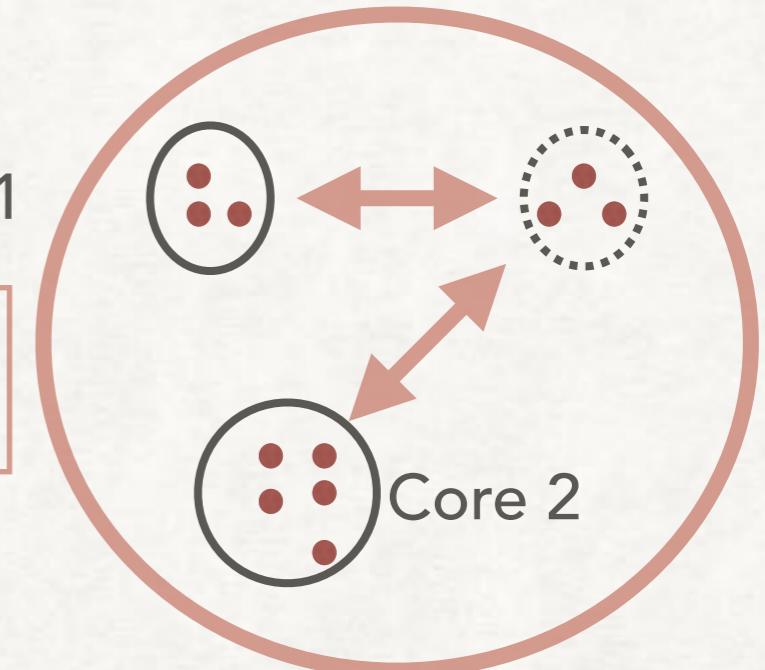
N-subjettiness,

$$\tau_N^{(\beta)} = \frac{1}{p_{TJ}} \sum_{i \in \text{Jet}} p_{Ti} \min \left\{ R_{1i}^{\beta}, R_{2i}^{\beta}, \dots, R_{Ni}^{\beta} \right\}$$

N 個の中心がある  
として最小化

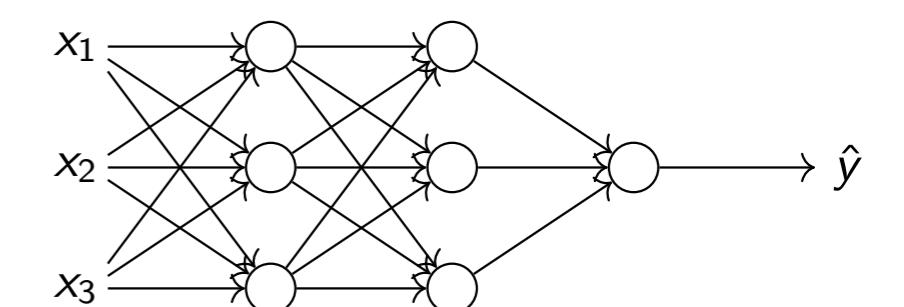
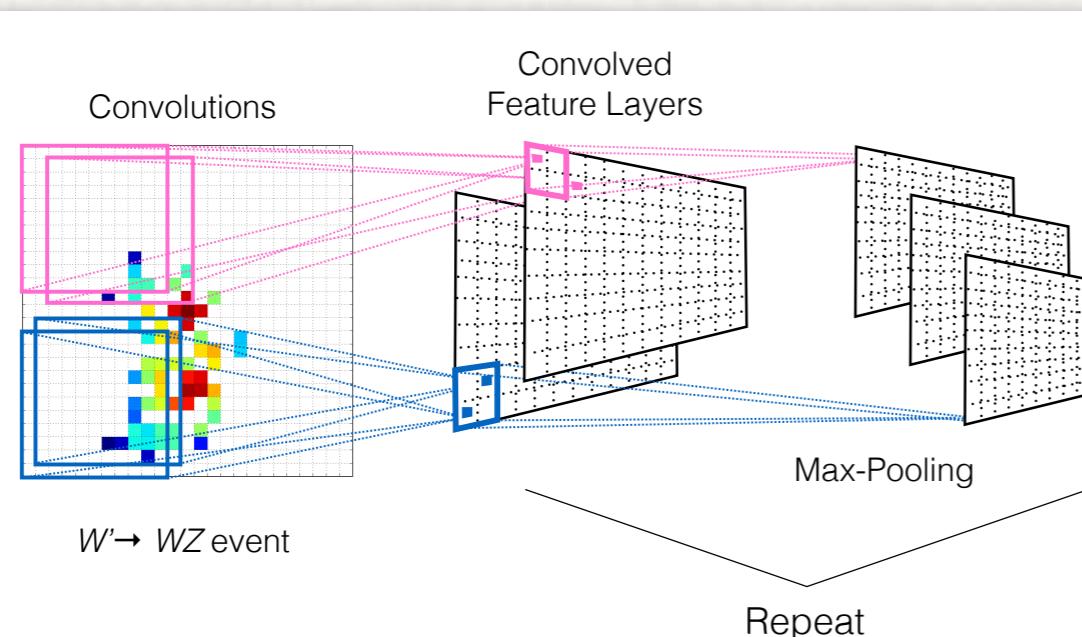
エネルギーで weight するのは、  
IRC safe にするため

Jet 内の全ての粒子で sum



$\tau_2$  LARGE >  $\tau_3$  SMALL

CNN 特徴量とか気にせず, がんがんジェットイメージを入れてやらせてみよう(2015~)

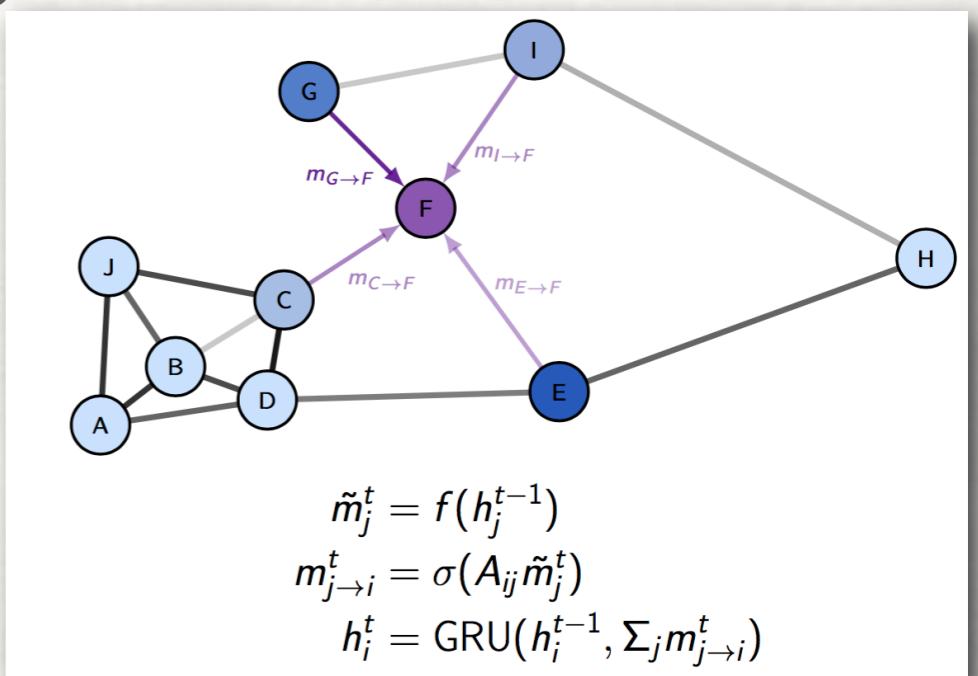
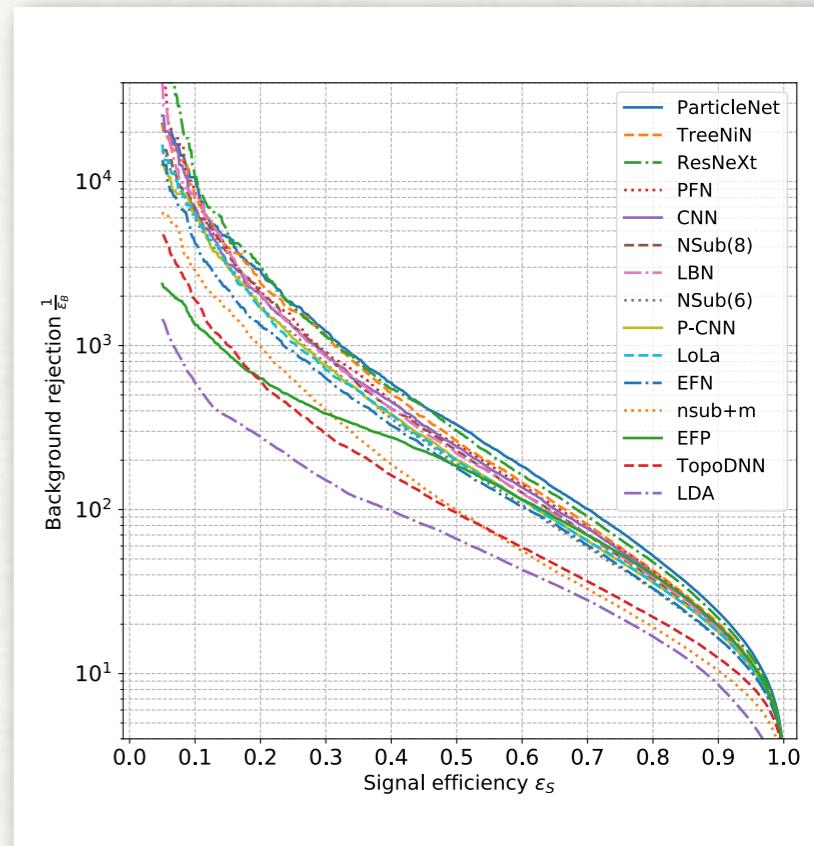


$(p_{Tij}, i, j)$

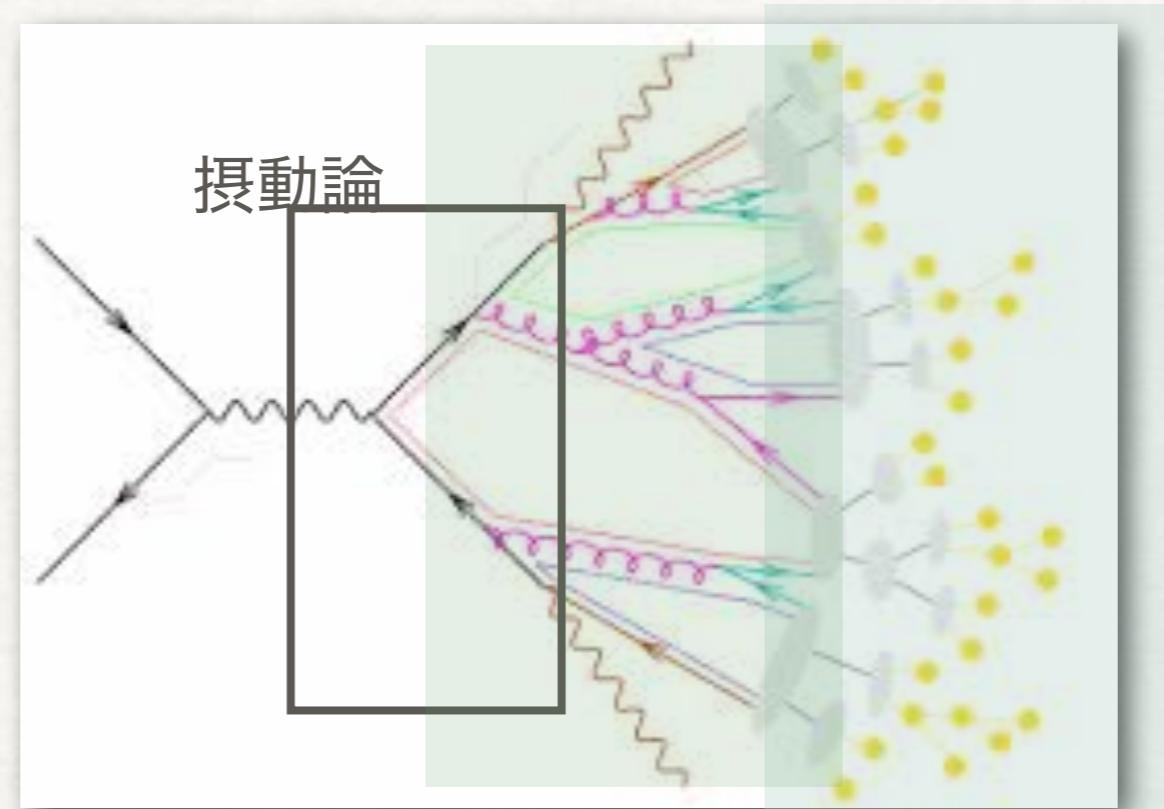
# 深層学習とJETの分類

- 多くの実装: CNN, ResNeXT, Particle Net...
 

>> BDT>> cut base.
- なんの効果でよくなるかが不明(というか気にならないスタンス)
- ジェット画像の特性上, 中心には多くのヒットがあるが、外側のヒットは少ない。分散も大きい。CNN が非効率 →graph network
- Particle net “周辺”の object から output →そこからアウトプット→ ほんとにそこまでしないといけないのか。



# 物理から考える：安心できる量と不安な量



- IRC safe な量: soft or collinear emission に対して安定な量:subjet
- QCD補正に対して不安定な量,例えば運動量の低い粒子数はMC modeling に大きな差 (Pythia vs Hewig vs 実験データなど) 系統誤差に問題あり  
→Jet Image の利用: 両者の間の区別が不明確 系統誤差の評価に問題はないか。実データで補正できるのか。
- 理論：ジェットの構造の中で理論の記述できる量を精査して、それだけ使えばよいのでは？

# IRC safe な量だけではベストな結果はでない

## N-subjettiness MLP classification

$$\tau_N^{(\beta)} = \frac{1}{p_{TJ}} \sum_{i \in \text{Jet}} p_{Ti} \min \left\{ R_{1i}^{\beta}, R_{2i}^{\beta}, \dots, R_{Ni}^{\beta} \right\}$$

$$\left\{ \tau_1^{(0.5)}, \tau_1^{(1)}, \tau_1^{(2)}, \tau_2^{(0.5)}, \tau_2^{(1)}, \tau_2^{(2)}, \tau_3^{(0.5)}, \tau_3^{(1)}, \tau_3^{(2)}, \tau_4^{(0.5)}, \tau_4^{(1)}, \tau_4^{(2)}, \tau_5^{(1)}, \tau_5^{(2)} \right\}$$

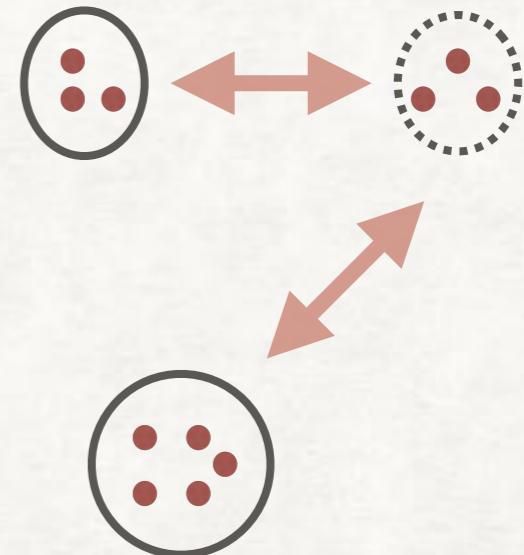
arXiv 1704.08249 Datta Larkoski

## CNN vs N-subjettniess MLPを比較する

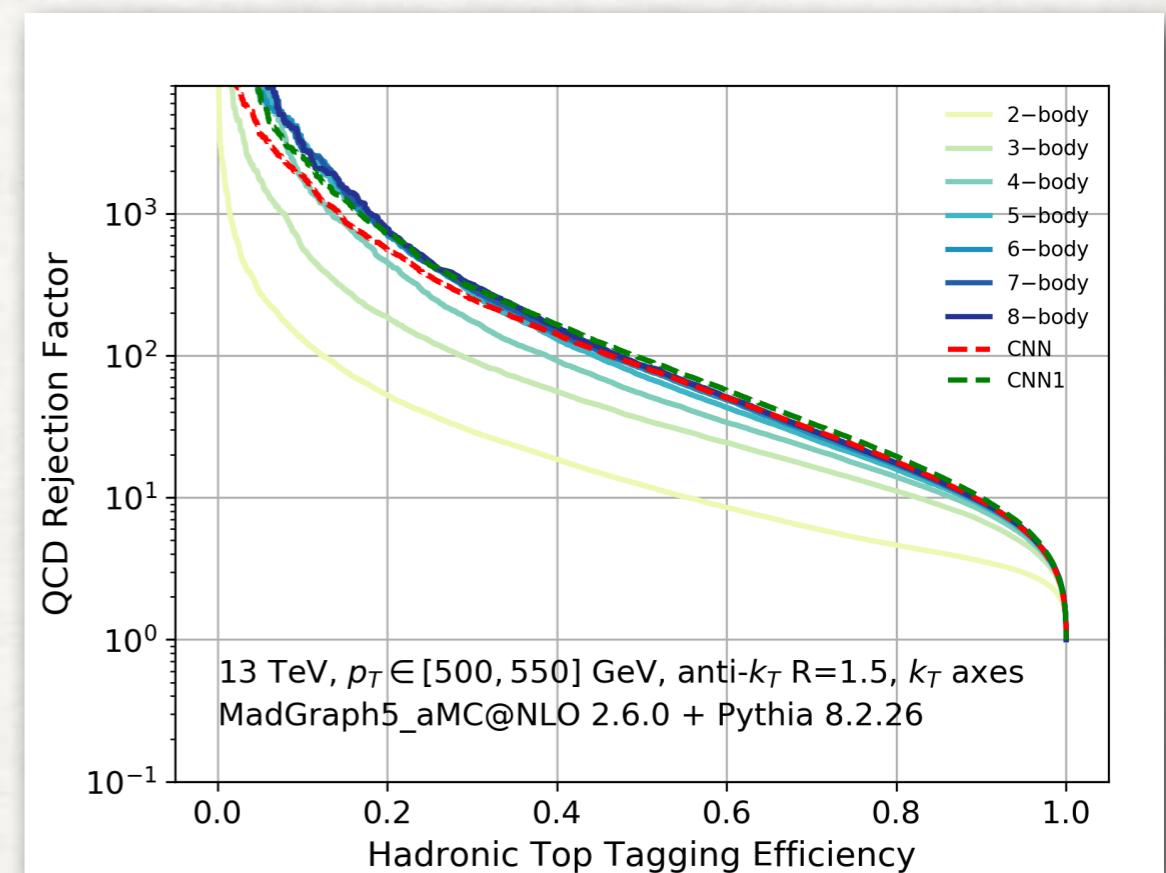
Liam Moore et al 1807.04769

とても高い次数の  $\tau(n$ 点相關)をいれると  
なんとかCNNと同じになる。

なんでそんなものが必要なの？ IRC  
unsafe な量はどういうふうに効いてい  
るの？お任せしちゃって大丈夫なの？



Core2

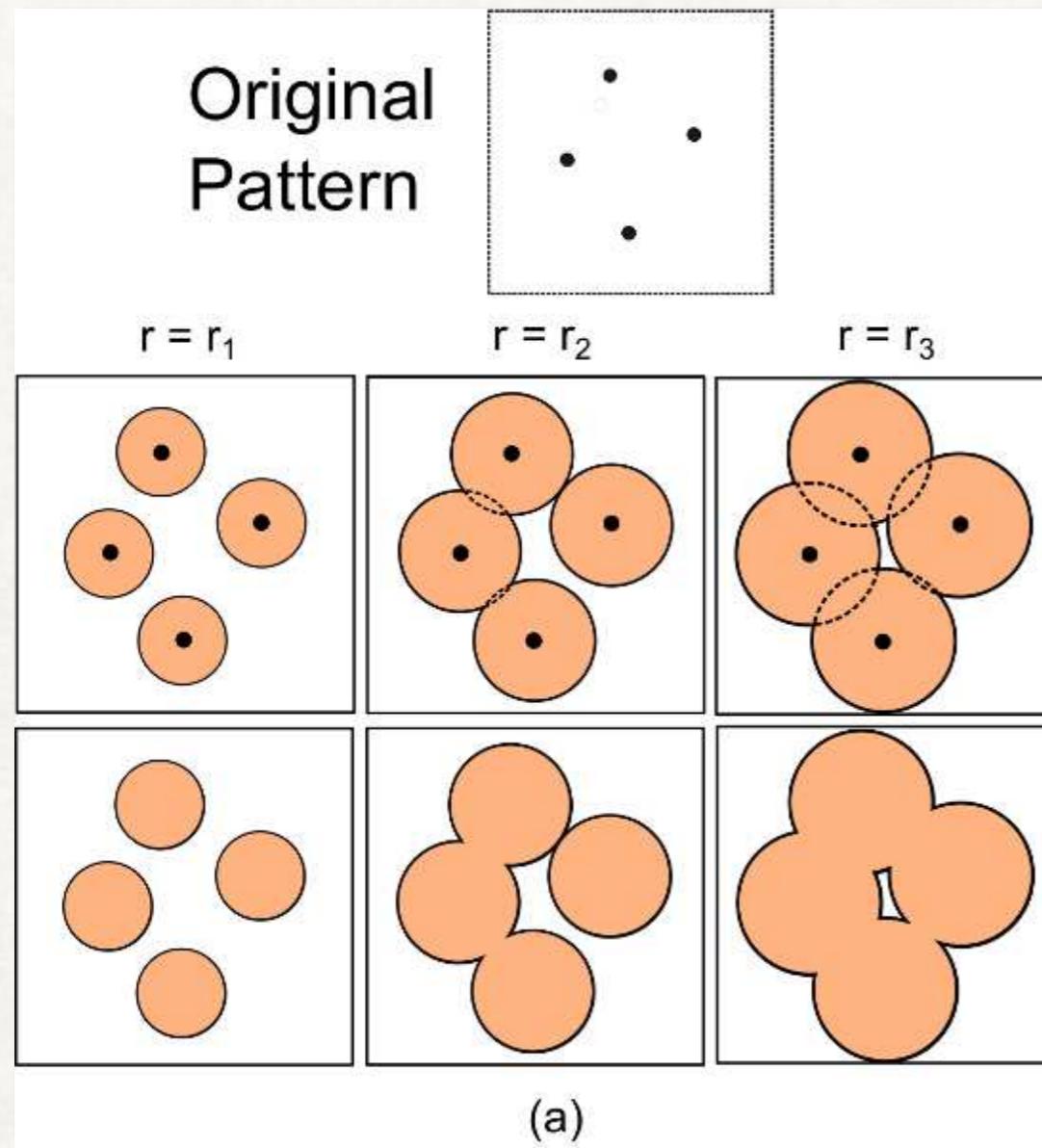


## ここまでまとめ

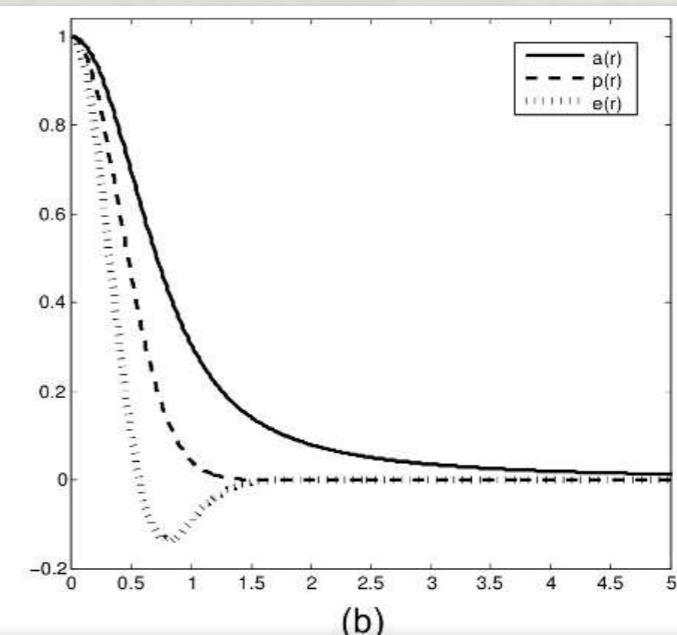
- ジェットの中には、理論的に予言ができる量 (IRC safe な量) と難しい量がある。特に難しいのは、ジェットの中にある、ランダムでエネルギーの低い粒子分布
- CNNはそういう量も使っているのではないか。
- そのようなデータはジェットの外側に分布していて、変動も大きい。

ランダムな点分布を定量化する、よい量はないか。

# ミンコフスキ一汎関数で ジェットのソフトな粒子分布の定量化



Area	$a(r) = \frac{A(r)}{\pi N r^2}$
Perimeter	$p(r) = \frac{P(r)}{2\pi N r}$
Euler characteristic	$e(r) = \frac{\chi(r)}{N}$



$N$  dim 空間  $\rightarrow$   $N+1$  個の独立汎関数が

$M(B_1 \cup B_2) = M(B_1) + M(B_2) - M(B_1 \cap B_2)$  をみたし、並進、回転不变

# MF の他の物理応用

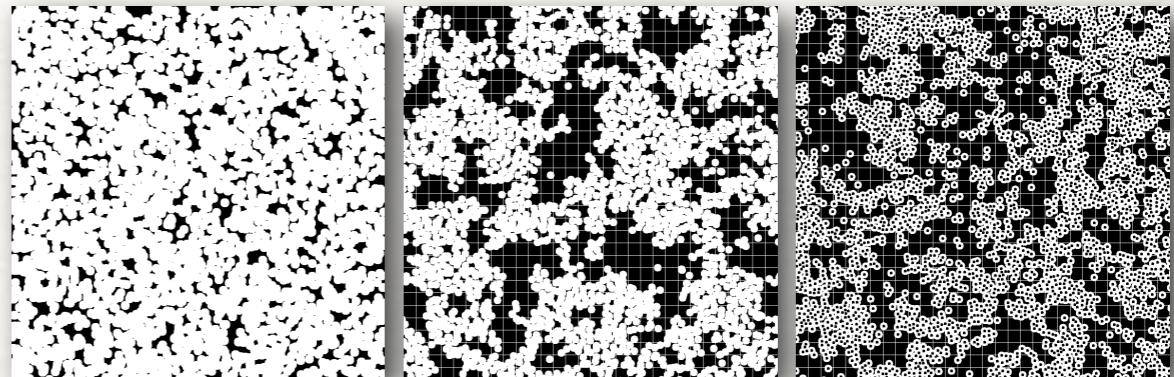
統計物理

左 多孔質体  
真ん中: 微乳濁液  
左 コロイド

体積の占有状況 $V$ , 表面の大きさ( $S$ ) 等に依存して物性が変わる 図は Mecke and Stoyan (2000)

- 天文: 星の分布の定量化、銀河分布、シミュレーション結果の定量化、non-Gaussianity of CMB, weak lensing..

点の集まりの意味を定量的に表現する時に使う



Kratochvil 1109.6334 Proving Cosmology with Weak Lensing Minkowski Functionals

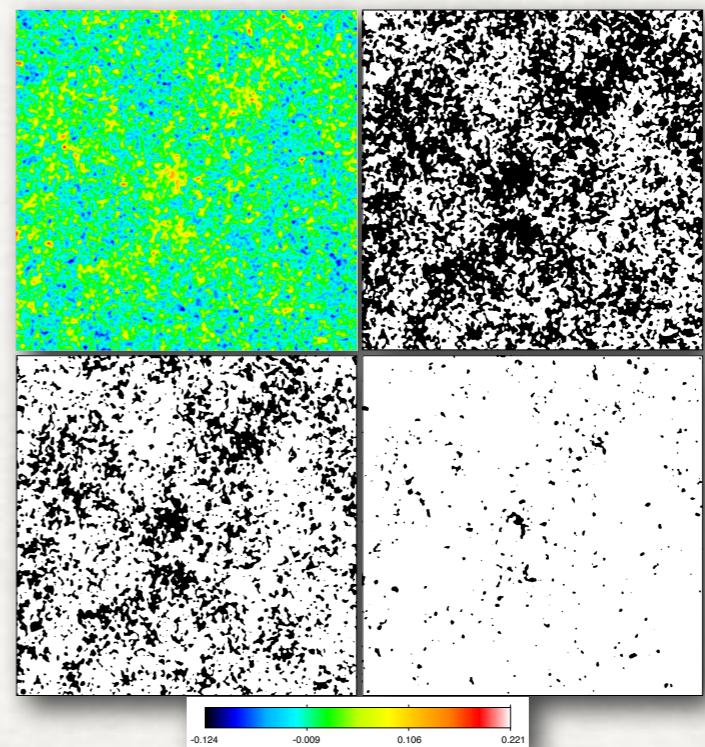
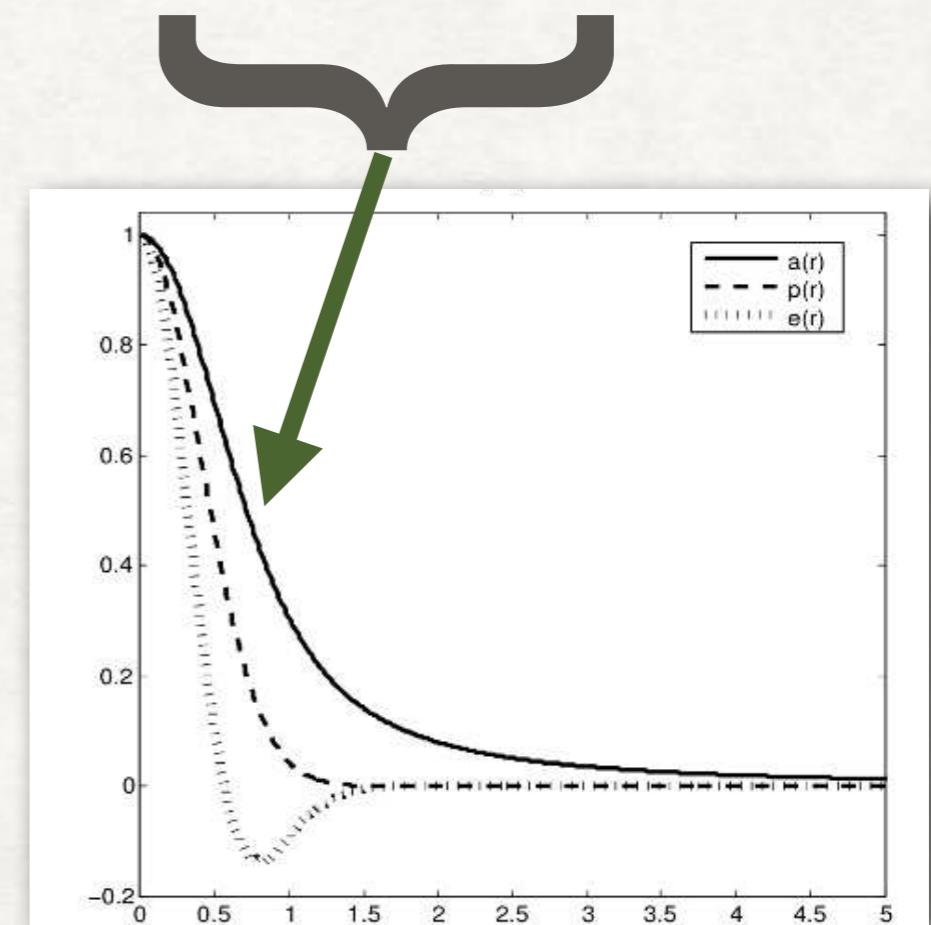
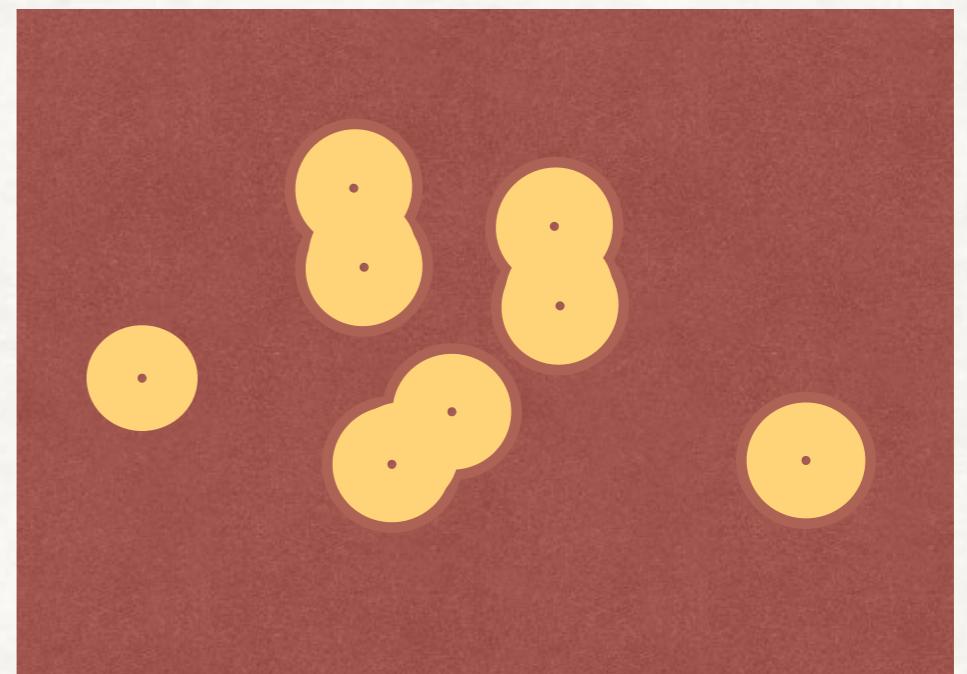


FIG. 1: Top left panel: example of a simulated 12-square-degree convergence map in the fiducial cosmology, with intrinsic ellipticity noise from source galaxies and  $\theta_G = 1 \text{ arcmin}$  Gaussian smoothing. A source galaxy density of  $n_{gal} = 15/\text{arcmin}^2$  at redshift  $z_s = 2$  was assumed. Other three panels: the excursion sets above three different convergence thresholds  $\kappa$ , i.e. all pixels with values above (below) the threshold are black (white). The threshold values are  $\kappa = 0.0$  (top right),  $\kappa = 0.02$  (bottom left), and  $\kappa = 0.07$  (bottom right). The Minkowski Functionals  $V_0$ ,  $V_1$ , and  $V_2$  measure the area, boundary length, and Euler characteristic (or genus), respectively, of the black regions as a function of threshold.

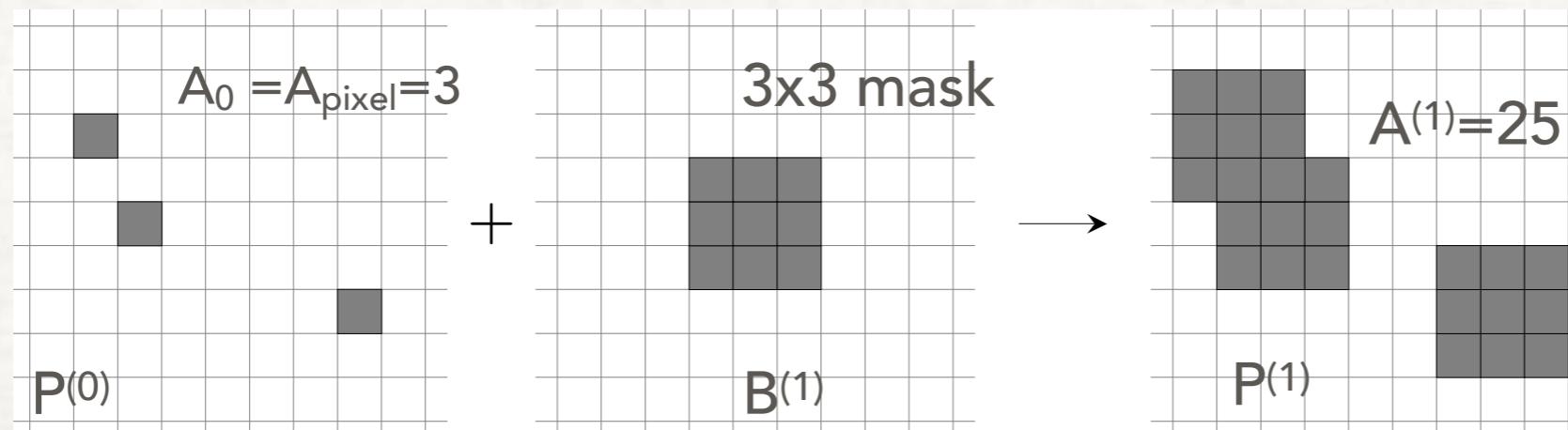
# ジェット物理への応用

- 基本的に良いことばかり
  - 0を並べるような無駄な情報や、位置情報のロスがない。
  - ジェットイメージのピクセルのふらつき  $O(1)$  が  $1/\sqrt{N_{\text{pixel}}}$  に下がっている。Loss 関数の最小化に貢献。
  - すべての点が同等に扱われていることは、現代的なジェットのアルゴリズムに合致する。
  - 同方向のパートンの分岐に関して安定



# ジェットの物理への組み込みかた

- $3 \times 3, 5 \times 5..$  のマスクをヒットに用意  $A_0 \rightarrow A_1 \rightarrow A_2 \dots$

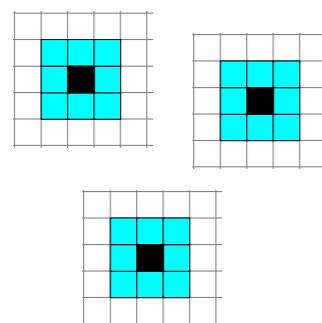


- $A_1/A_0$ の場合

$$P^{(k)} = \{a + b \mid a \in P^{(0)}, b \in B^{(k)}\},$$
$$B^{(k)} = \{(i, j) \mid i, j \in \{-k, -k+1, \dots, k-1, k\}\}$$

Isolated points

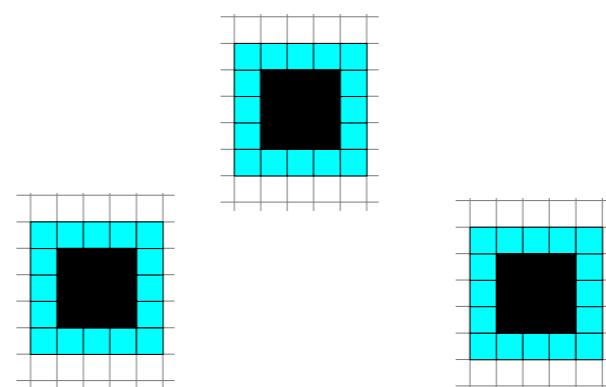
$$A_1/A_0 = 9$$



VS

all pixel appear in  $3 \times 3$

$$A_1/A_0 = 16/9 = 1.78$$



# ミンコフスキー汎関数とCNNの関係

- ミンコフスキー汎関数(MF)は CNN のフィルターの形で書くことができる  
2x2 のフィルター(ルックアップテーブル)を使った実装

$$(A^{(k)}, L^{(k)}, \chi^{(k)}) = \sum_{i,j} \sum_{n,m \in \{0,1\}} \mathbf{v} \left( P_{(i+n)(j+m)}^{(k)} f_{nm} \right) \quad f_{nm} = ((1,2), (4,8))$$

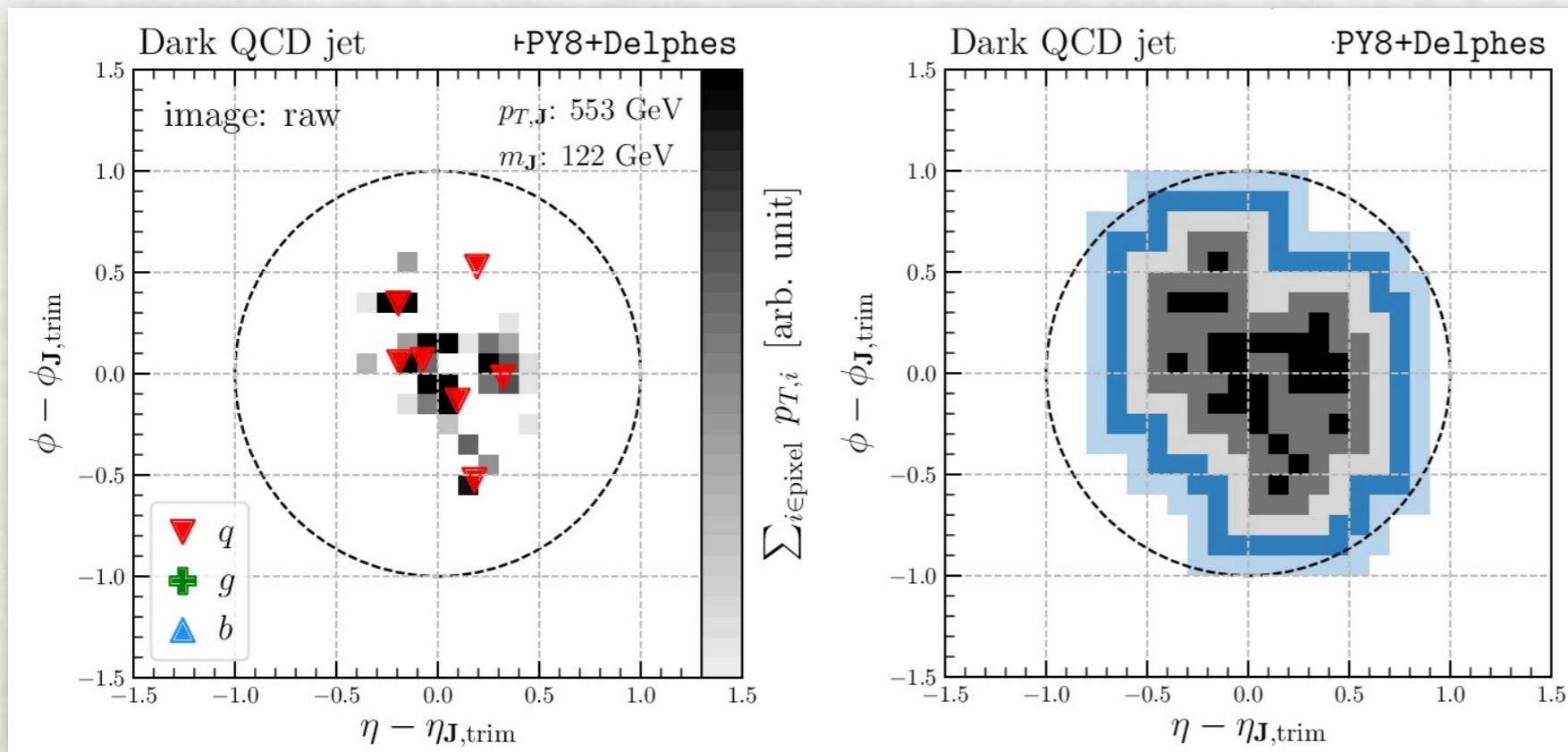
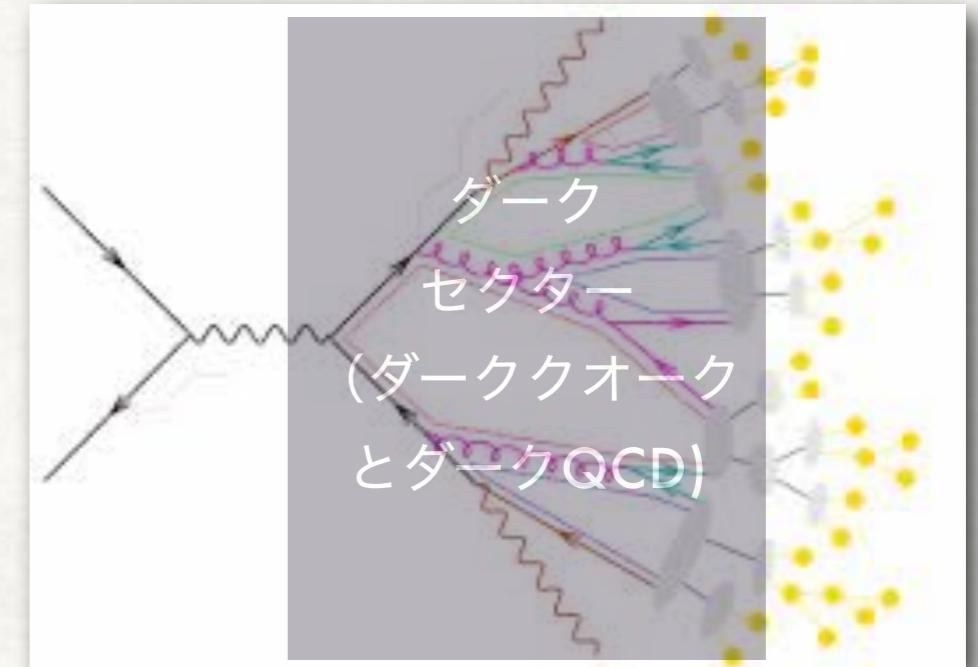
**Table 1.** Look-up table for Minkowski functionals.

Conf.		A	P	$\chi$	Conf.		A	P	$\chi$
1		0	0	0	9		1/4	1	1/4
2		1/4	1	1/4	10		1/2	2	-1/2
3		1/4	1	1/4	11		1/2	1	0
4		1/2	1	0	12		3/4	1	-1/4
5		1/4	1	1/4	13		1/2	1	0
6		1/2	1	0	14		3/4	1	-1/4
7		1/2	2	-1/2	15		3/4	1	-1/4
8		3/4	1	-1/4	16		1	0	0

# DARK JET の場合

Lim, Nojiri in preparation

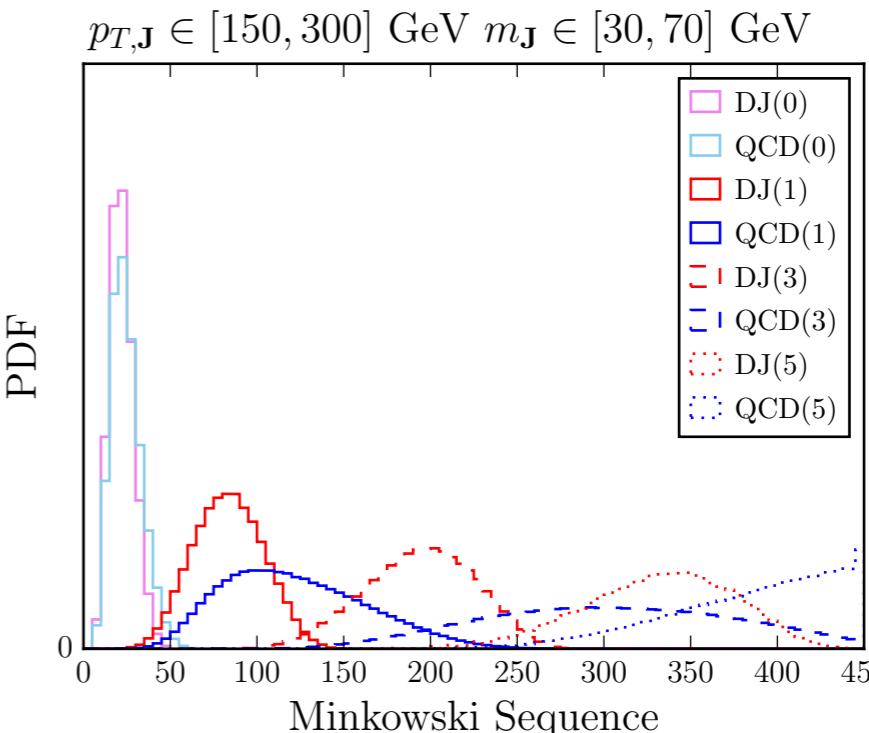
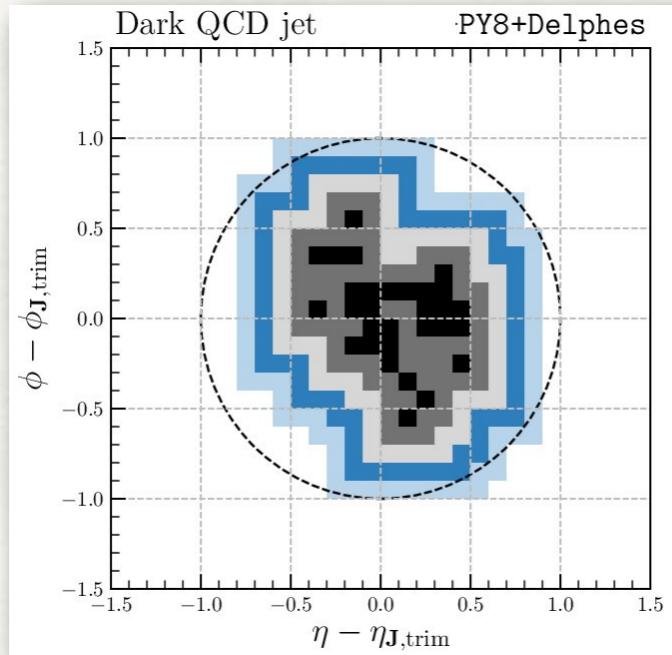
- Dark Jet  $pp \rightarrow Z' \rightarrow qD qD \rightarrow$  dark Parton shower  $\rightarrow \rho_{\text{diag}} \rightarrow qq$
- カラーシングレットなシャワー: 粒子がたくさんあるが、いくつかのカラーシングレットなクラスターになっている状態



# CNN は何を訓練しているか

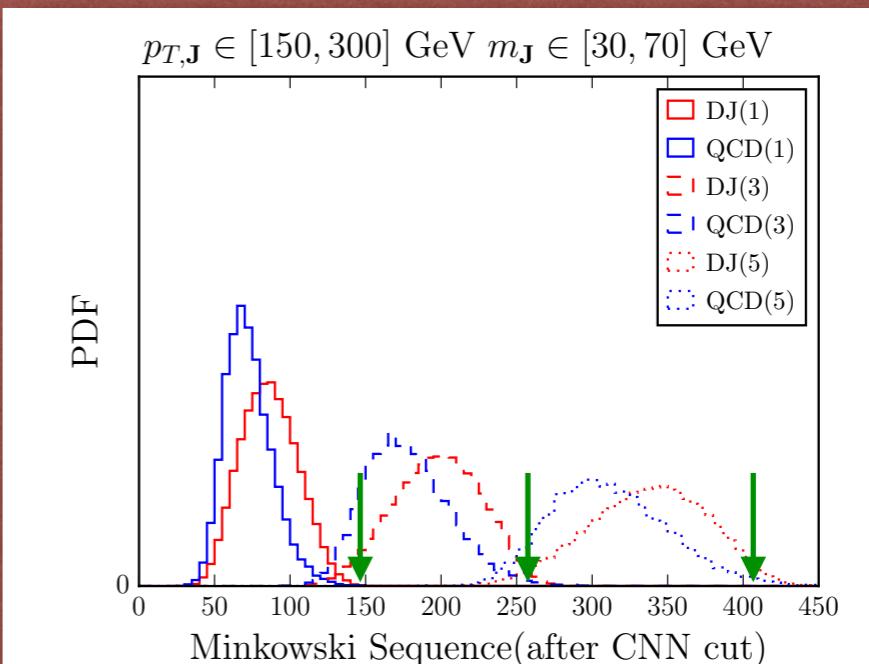
Dark jet と QCD ジェットを判別させる

Original distribution



Cut using CNN

Smaller MF endpoint suggests  
Compact soft activities



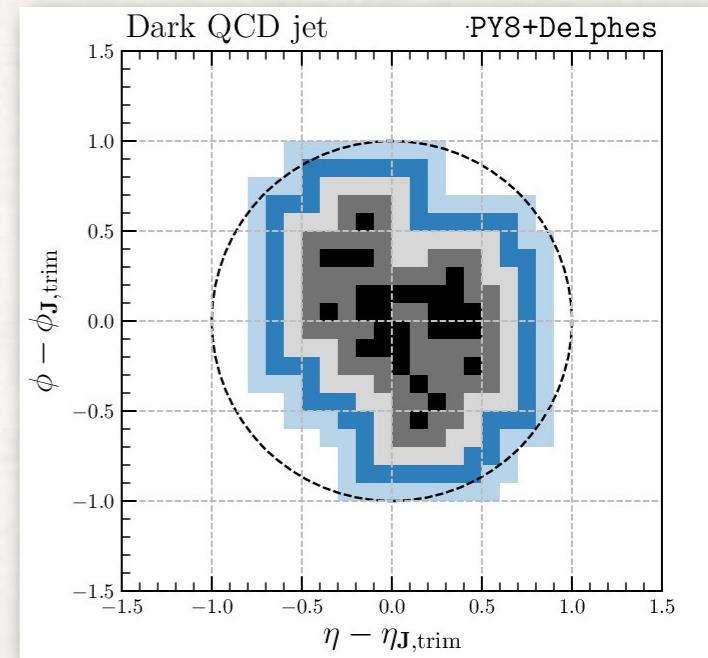
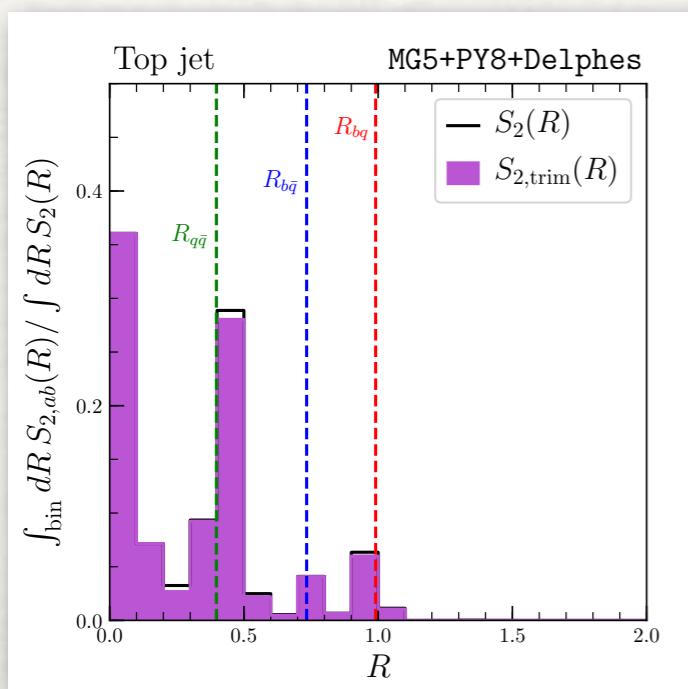
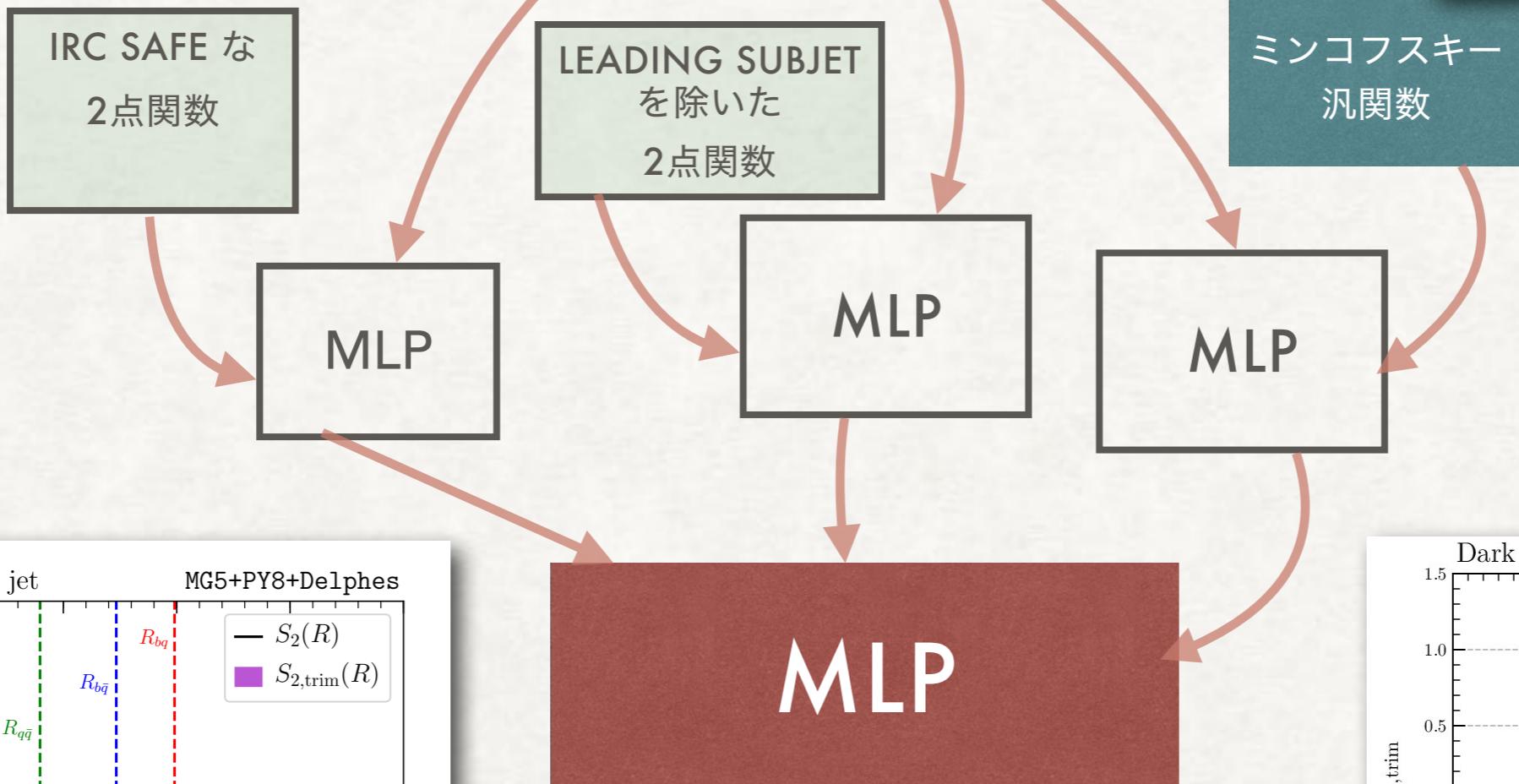
CNN allowing 10% rejection of signal

# RELATION NETWORK+ MF

$$S_{2,ab}(R) = \sum_{i \in a, j \in b} p_{T,i} p_{T,j} \delta(R - R_{ij})$$

ジェットの運動量などの  
大域的な量  $U$

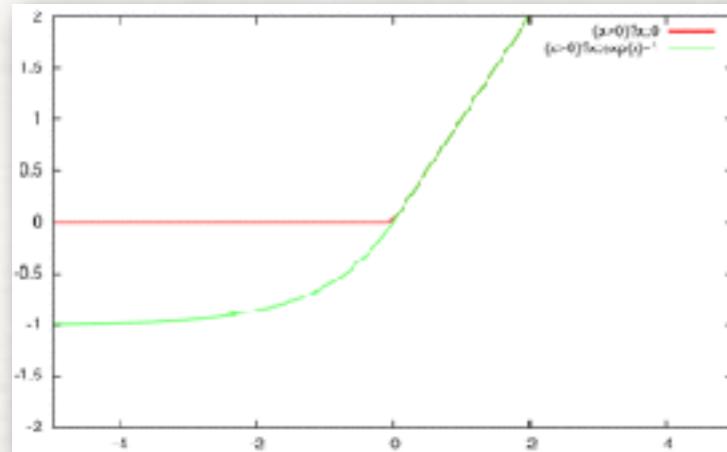
IRC safe でない情報は、  
ここだけからくる



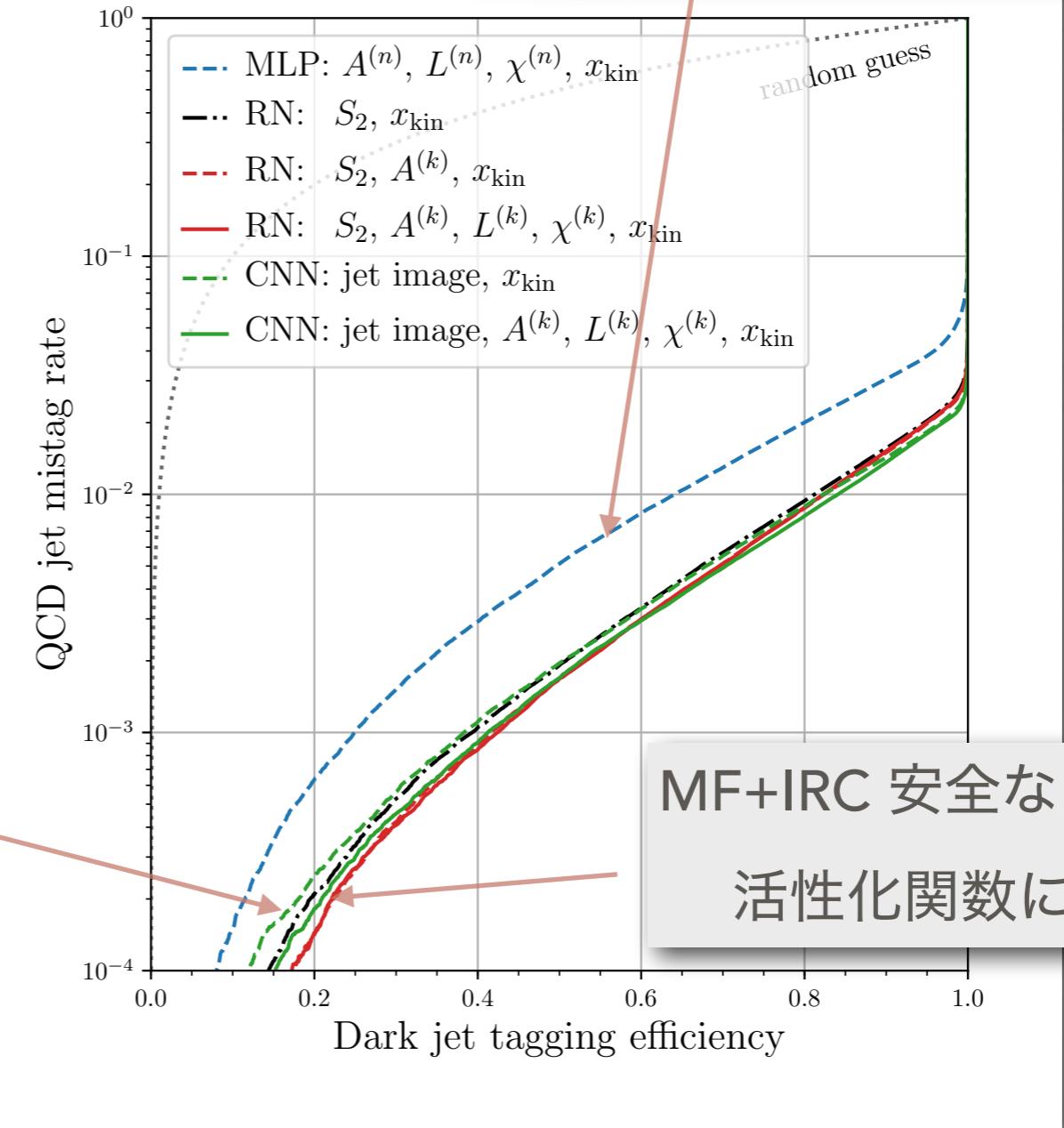
# DARK JET VS QCD

Jet PT 150~300GeV Jet mass 30~70GeV

MF only reject lots of QCD jet

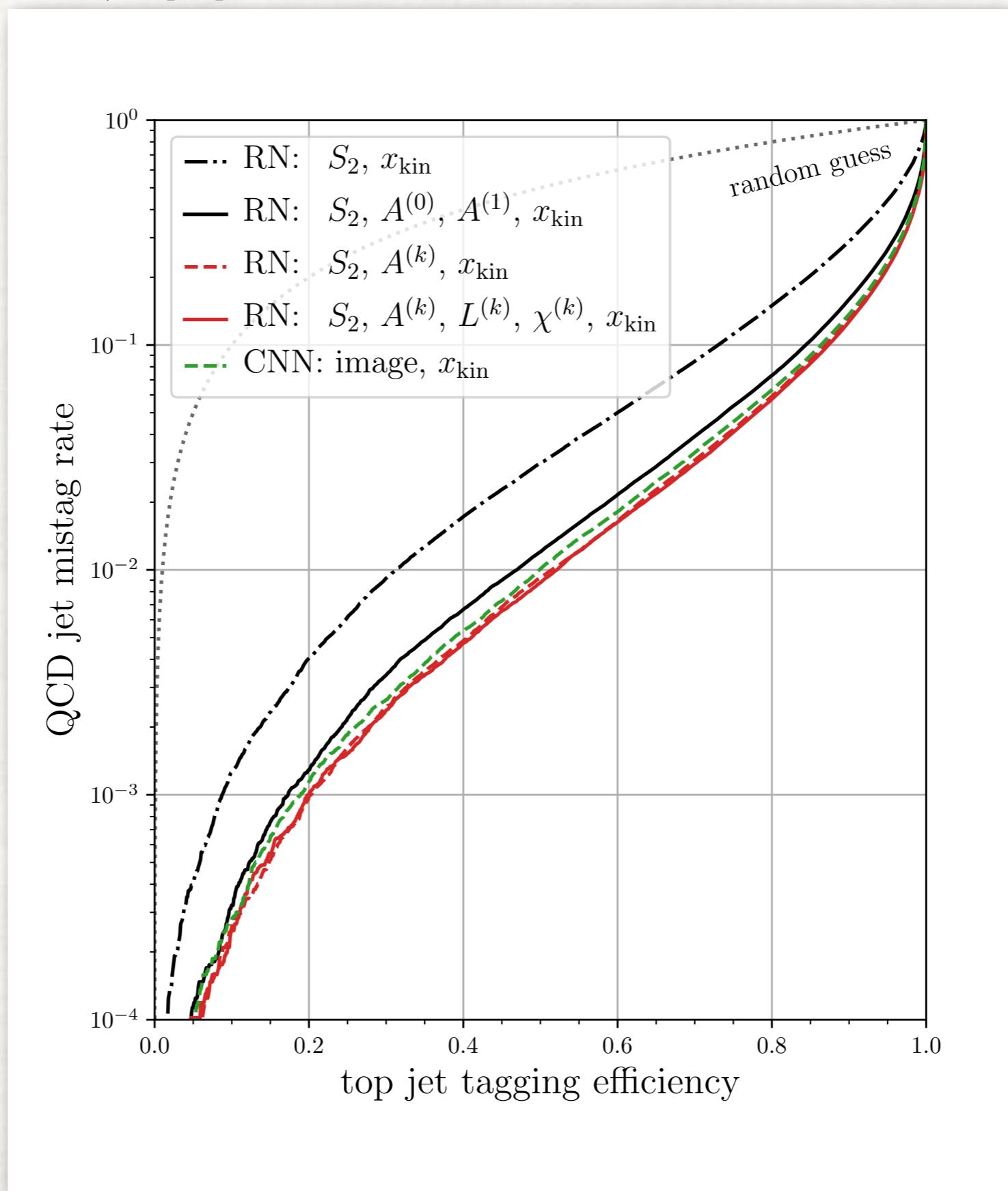


CNN(ReLU)  
ELU にするとすごく悪い  
急な変化をする活性化関数  
を必要としている。  
Bottleneck 必要



RN + MF also improve Top vs QCD (but improvement is smaller )

# Top jet の場合 やっぱり RN+ MF がよい



# ミンコフスキーフン関数とCNNの関係

- ミンコフスキーフン関数(MF)は CNN のフィルターの形で書くことができる  
2x2 のフィルター(ルックアップテーブル)を使った実装

$$(A^{(k)}, L^{(k)}, \chi^{(k)}) = \sum_i \sum_m \mathbf{v} \left( P_{(i+n)(i+m)}^{(k)} f_{nm} \right) \quad f_{nm} = ((1, 2), (4, 8))$$

本来はジェットイメージはRN+MFがもっている情報を使えるはずだ。

Tab

==  
C

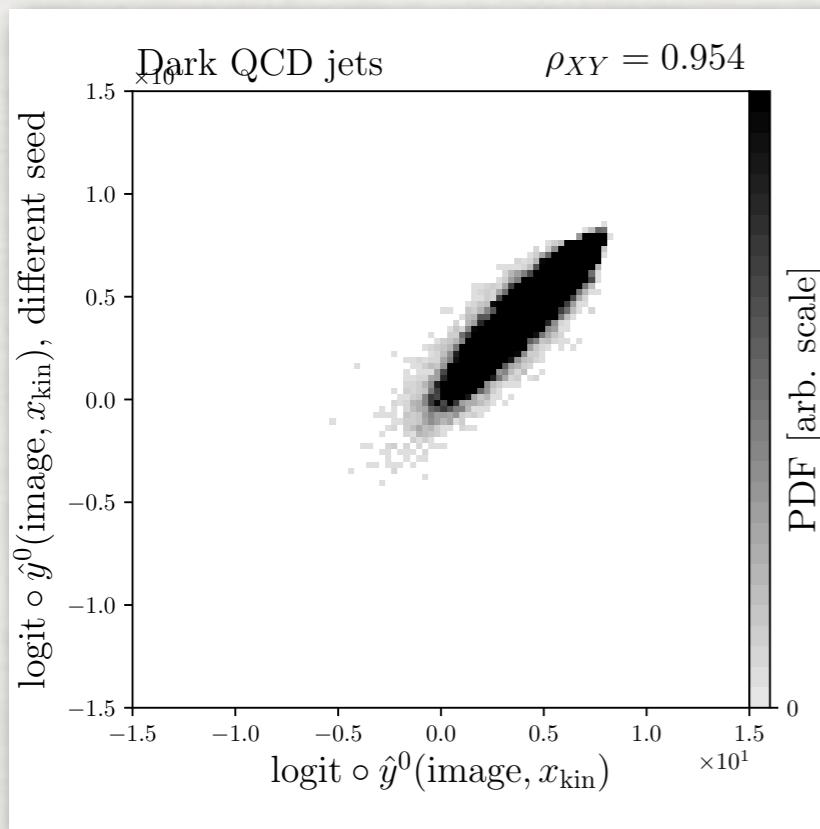
- ジェットイメージがエネルギーでウェイトされていること
- activation, Loss funcは連続関数であることが前提
- でもMFはエネルギーに依存しない関数

5		1/4	1	1/4	13		1/2	1	0
6		1/2	1	0	14		3/4	1	-1/4
7		1/2	2	-1/2	15		3/4	1	-1/4
8		3/4	1	-1/4	16		1	0	0

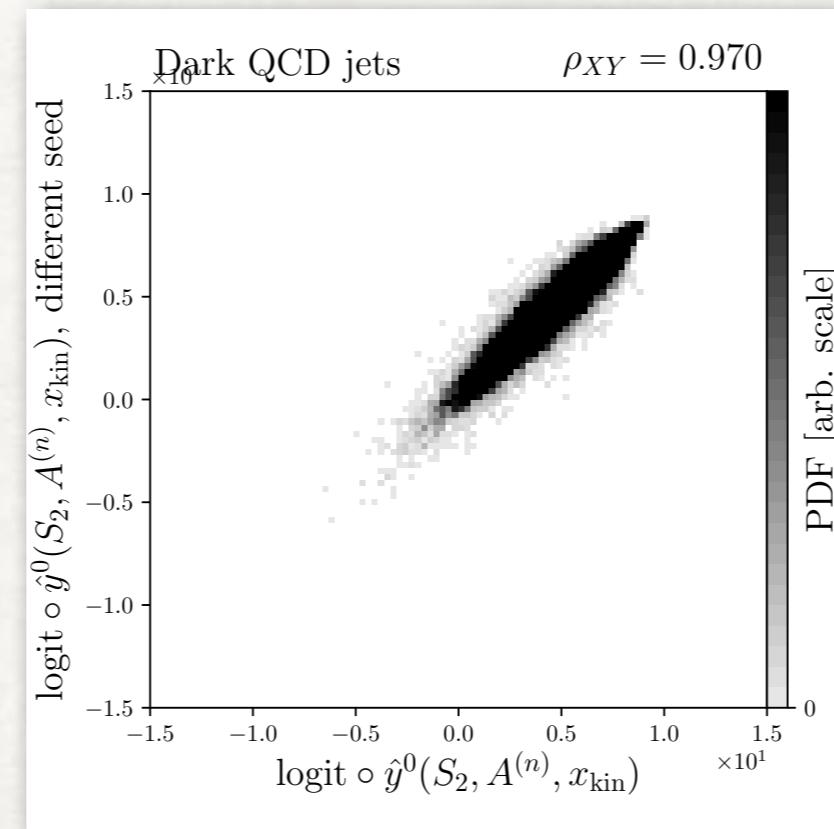
# 何が改善しているか

- 入力が少ない→ loss function の最小化が簡単 true minimum が見つけやすい。
- 時間 1/10 メモリも1/10
- 等価なのか、相関なのか。CNN と RN はどうやら違う判断をしている。

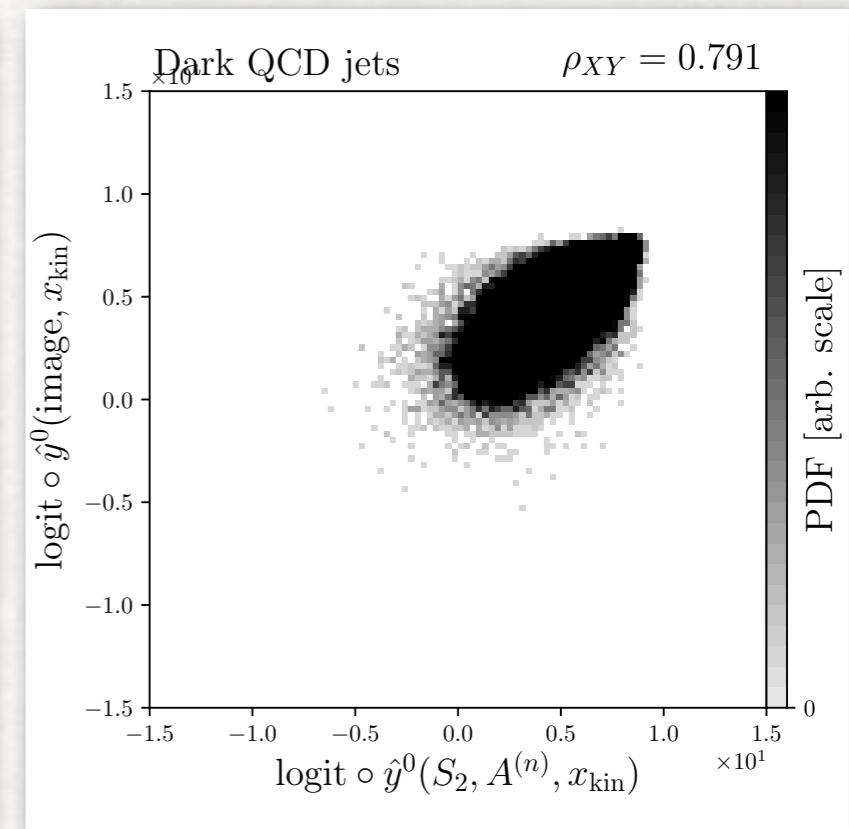
CNN vs CNN



RN vs RN

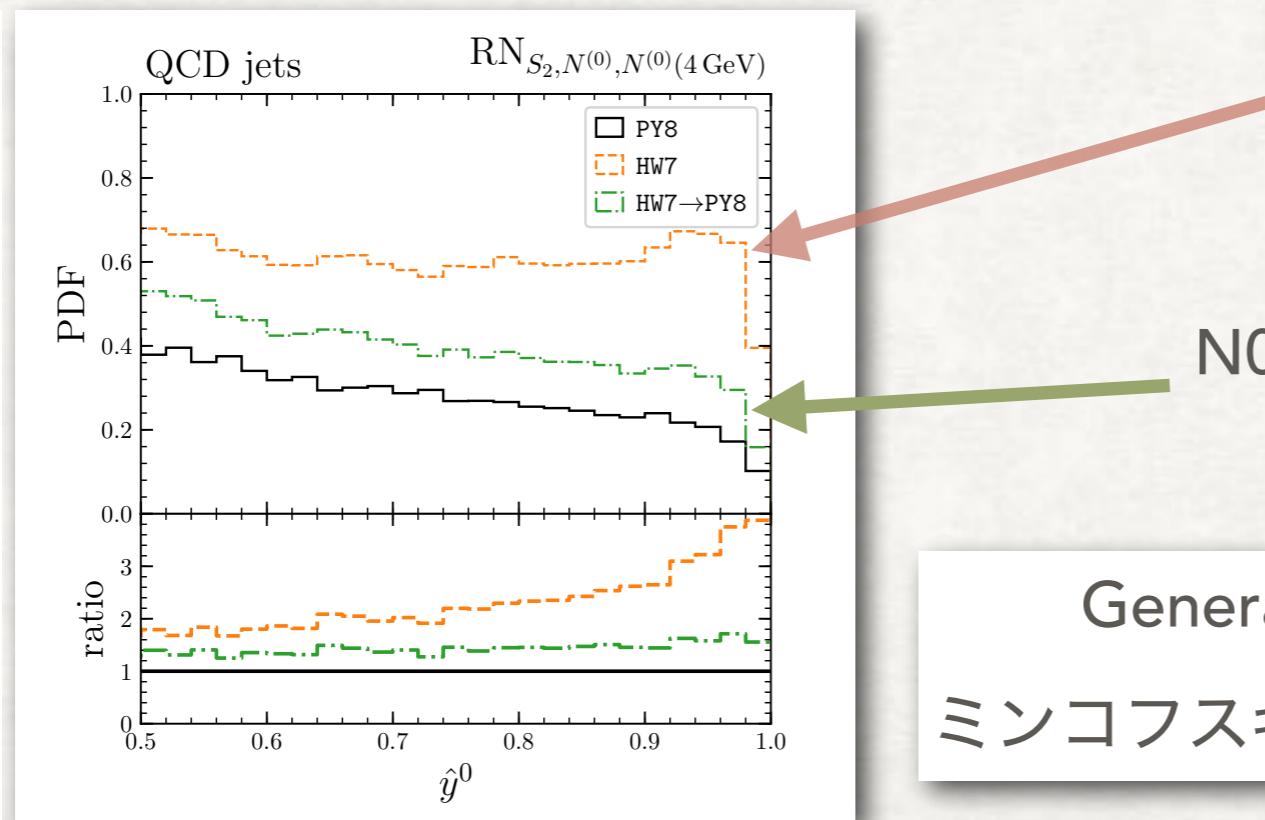
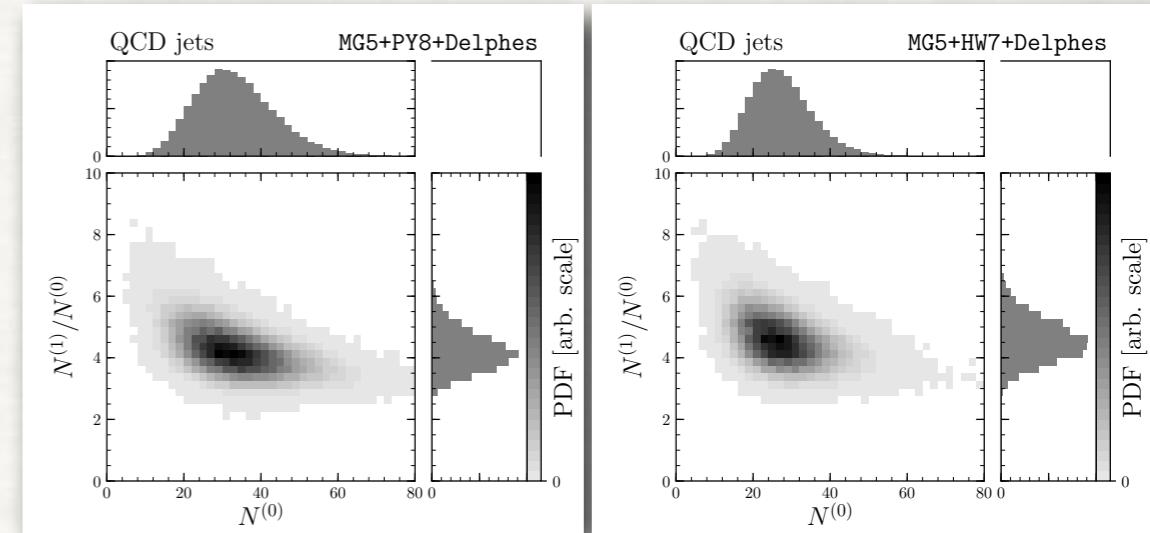


CNN vs RN



# 実データのキャリブレーション

- 既存のMCはQCD ジェットのIRCsafe でない分布をきちんと表していない。(特にグルーオンジェットの粒子数と広がり) Pythia $\neq$  Herwig $\neq$  実データ
- [MC を実データで補正]するときに、MFが実験データと同じになるように weight をつけて補正すると classification の一致は良くなる。



Generator を GAN で補正するときに、  
ミンコフスキーフン関数の分布だけ補正すれば良い?

# 教訓とやれうこと

- 自分の目でイベントをみよう
- CNNなどのジェットイメージを使った訓練は IRC safeでない量も追加して分類を強化している。
- CNNは理解した (1/10まで input 圧縮. パフォーマンスは落ちない) 解釈可能性、速度、安定性、いろいろなプロセスに使える。
- ミンコフスキー汎関数のような「まとめ指標」の方が実データの検定にもよい。パートン→ハドロン化の空間分布をよりよく検定できるかも。