

# 複数の物理スケールのあるプロセスに 配慮した深層学習モデル

野尻美保子 KEK

A. Hammad Mihoko M. Nojiri Streamlined jet tagging network assisted by jet prong structure 2024.14677 JHEP  
+ S. Moretti Multi-scale cross-attention transformer encoder for event Classification 2401.00452 JHEP

# 目次

- Higgs の物理においてなぜ深層学習が重要か
- スケールのある物理

具体的には LHCにおけるハードプロセス(1TeV~100GeV)→パートンシャワ(10GeV)

→ハドロニ化 (1GeV)を学習したいネットワークの優位性

- 他の物理を重視したネットワーク
- 最近の理論の発展と深層学習

# 「学習物理学」

Overview

Organization

Events

Achievements

Outreach

## Overview

### message

Head Investigator

## Koji Hashimoto

Professor

Particle Physics Theory Group

Department of physics, Kyoto University



The research area "Machine Learning Physics" will begin with the aim of discovering new laws and pioneering new materials

B01 Akinori Tanaka (Riken AIP) Math and application of DL

B02 Yoshiyuki Kabashima (Tokyo) Statistical data and ML

B03 Kenji Fukushima (Tokyo) Topology and Geometry of ML

A01 Akio Tomiya (IPUT Osaka) Lattice

A02 Mihoko Nojiri HEP

Junichi Tanaka (ICEPP Tokyo, ATLAS)

Masako Iwasaki (Osaka Metropolitan Belle II)

Noriko Takemura and Hajime Nagahara (Data Science)

A03 Tomi Ohtsuki (Sophia U) Condensed Matter

A04 Koji Hashimoto Quantum and Gravity

Ahmed Hammad

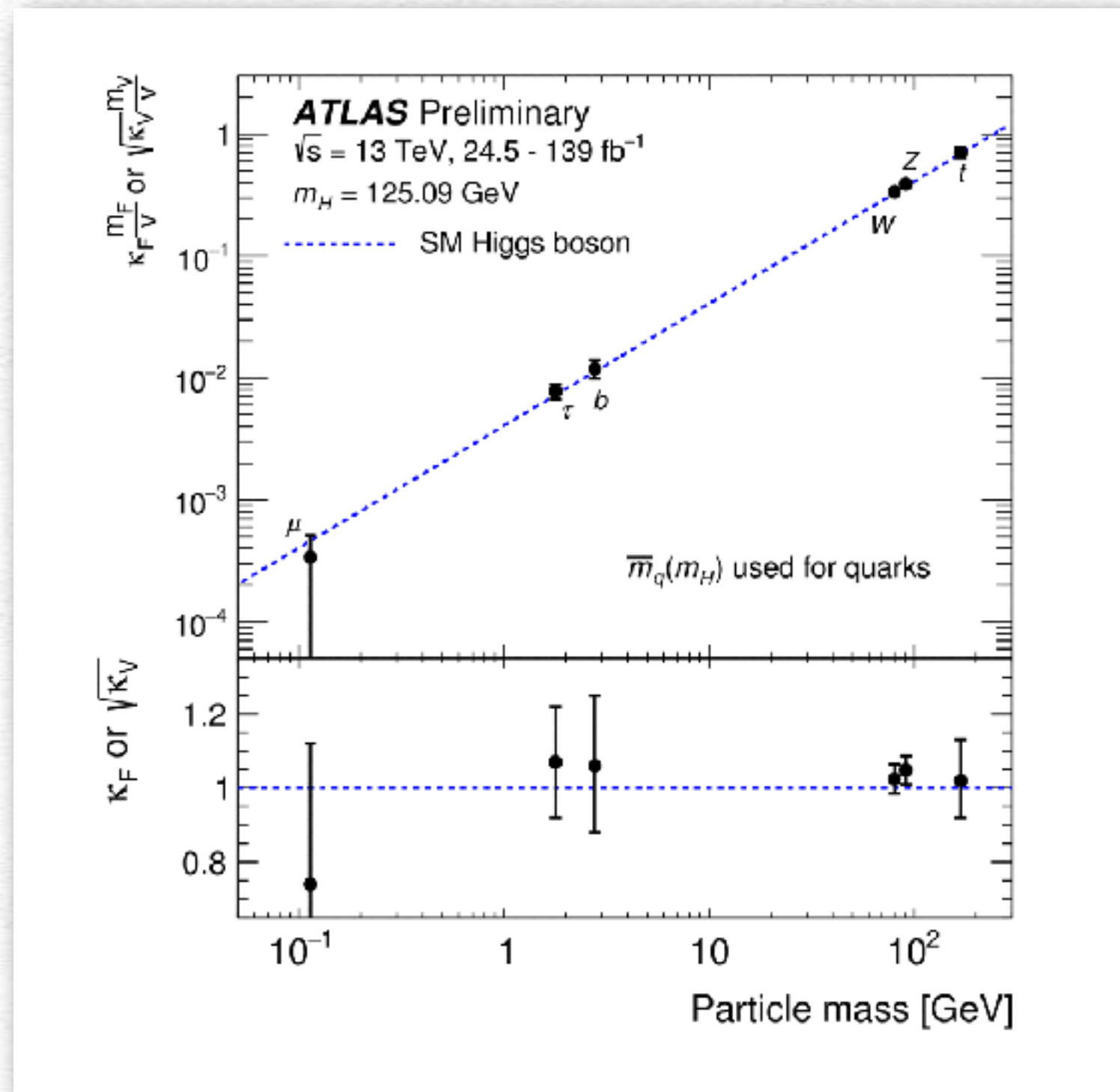
2017-2020: Ph.D Basel University,  
Basel Switzerland

2020-2023: SeoulTech, Korea

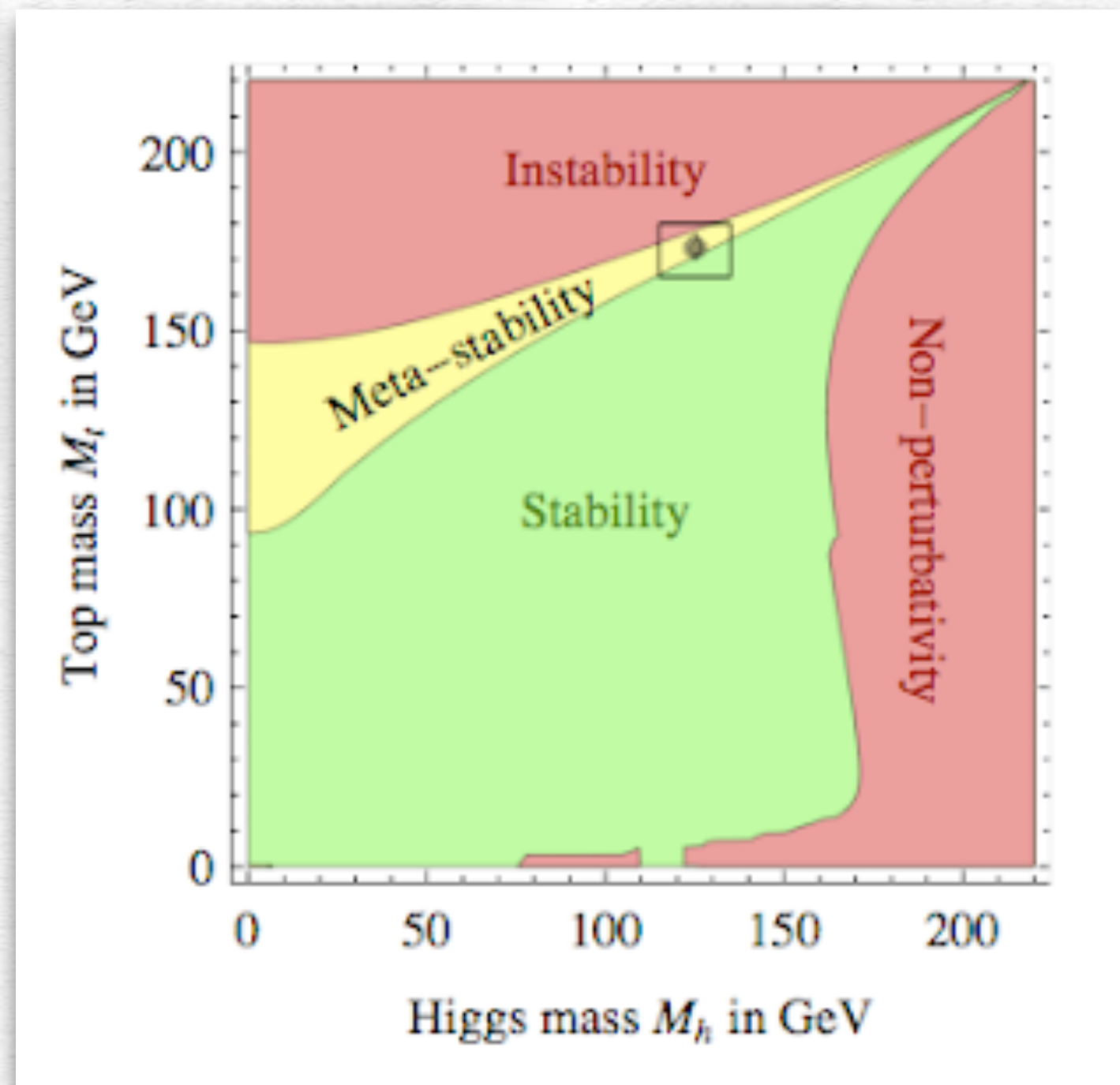
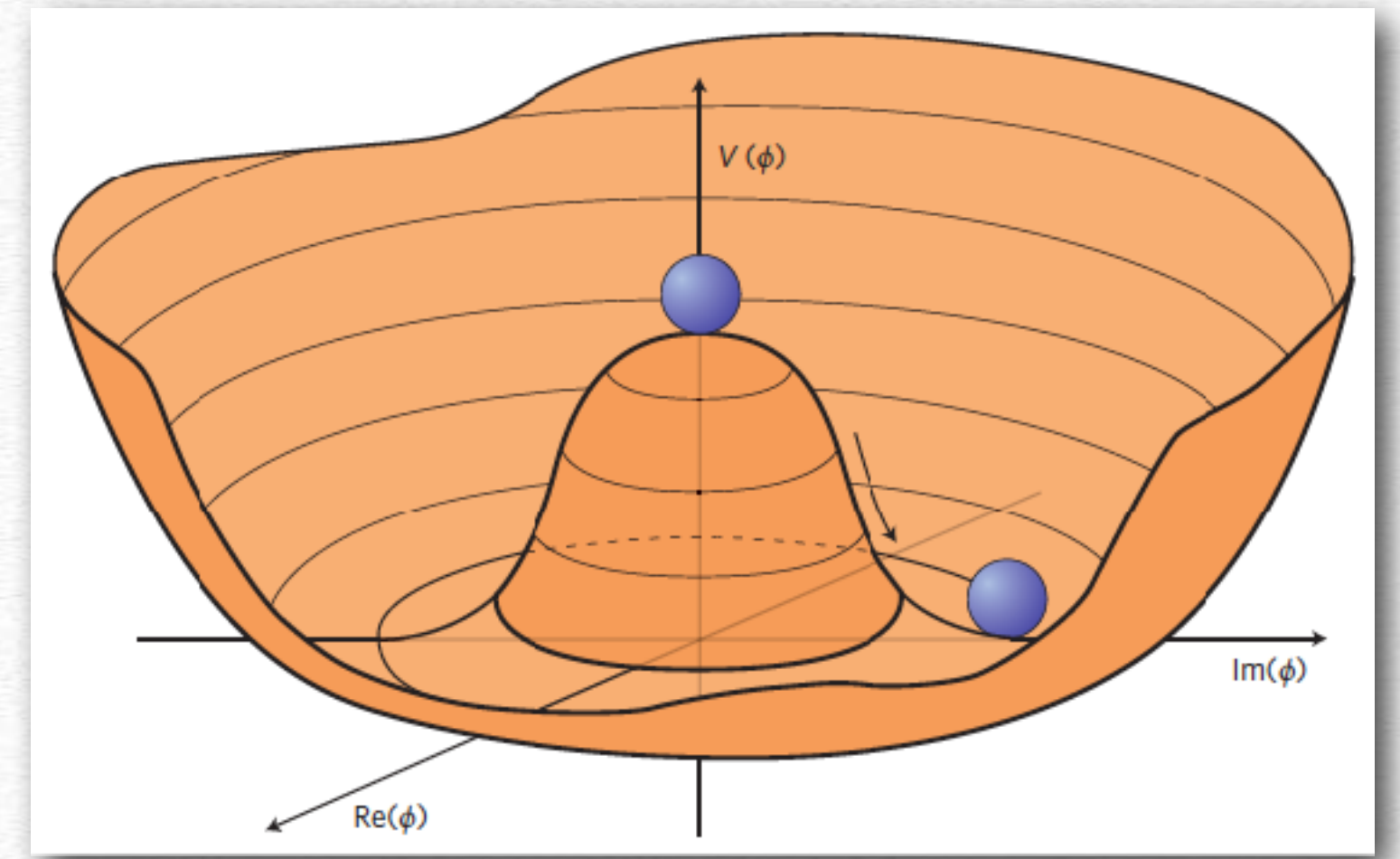
2023- KEK



# ヒッグス粒子の精密測定



Higgs order param.  $v$   
 "yukawa coupling"  $y$   
 "particle mass"  $\sim yv$



Higgs sector の何らかの拡張を  
 実験がどの程度制限できるか

SMEFT (SM の重要な対称性を  
 維持した高次元項)

新粒子探索

# LHC → HL-LHC

Higgs boson, dark sector 探索

LHC(2025 年 run 3 終了 300fb-1 )

HL-LHC(2029年から 高輝度化 3000fb-1 )

(その後の計画は来年の

ヨーロッパストラテジーで議論

FCC -ee → FCC-hh or Linear Collider )

## 暗黒物質

直接探索, 間接探索

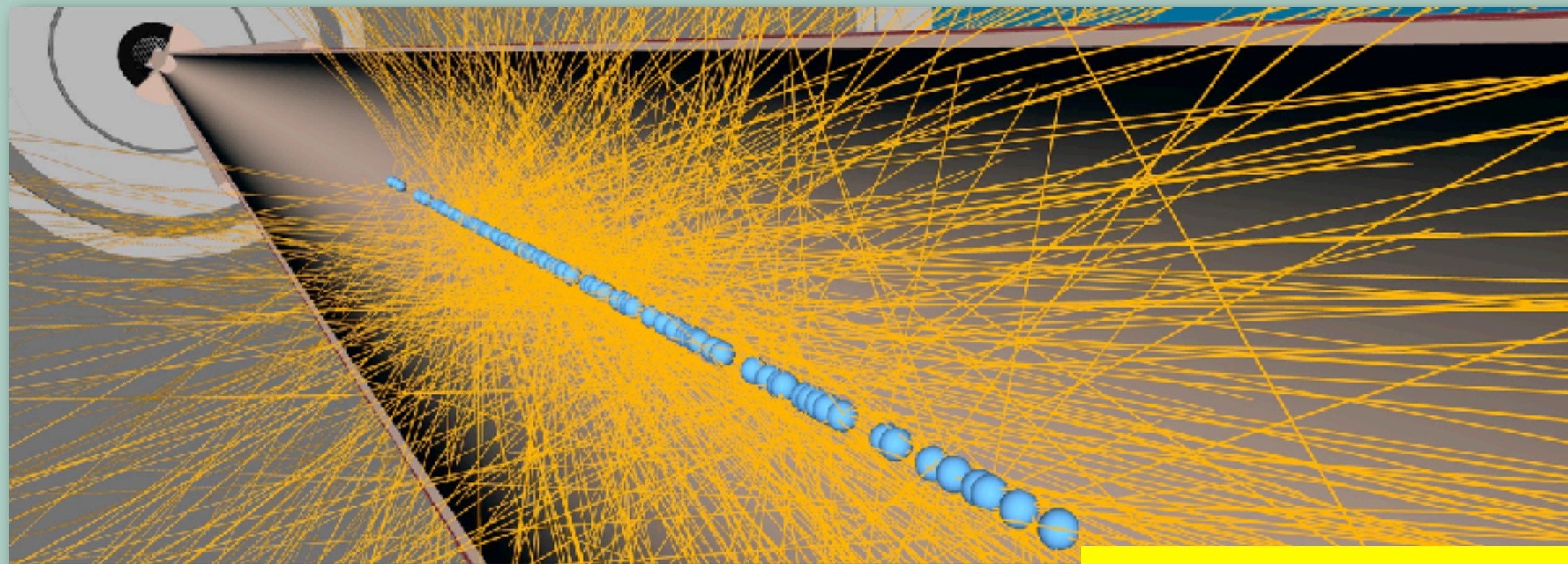
フレーバー関連

Belle-II 実験 LHC b , muon g-2

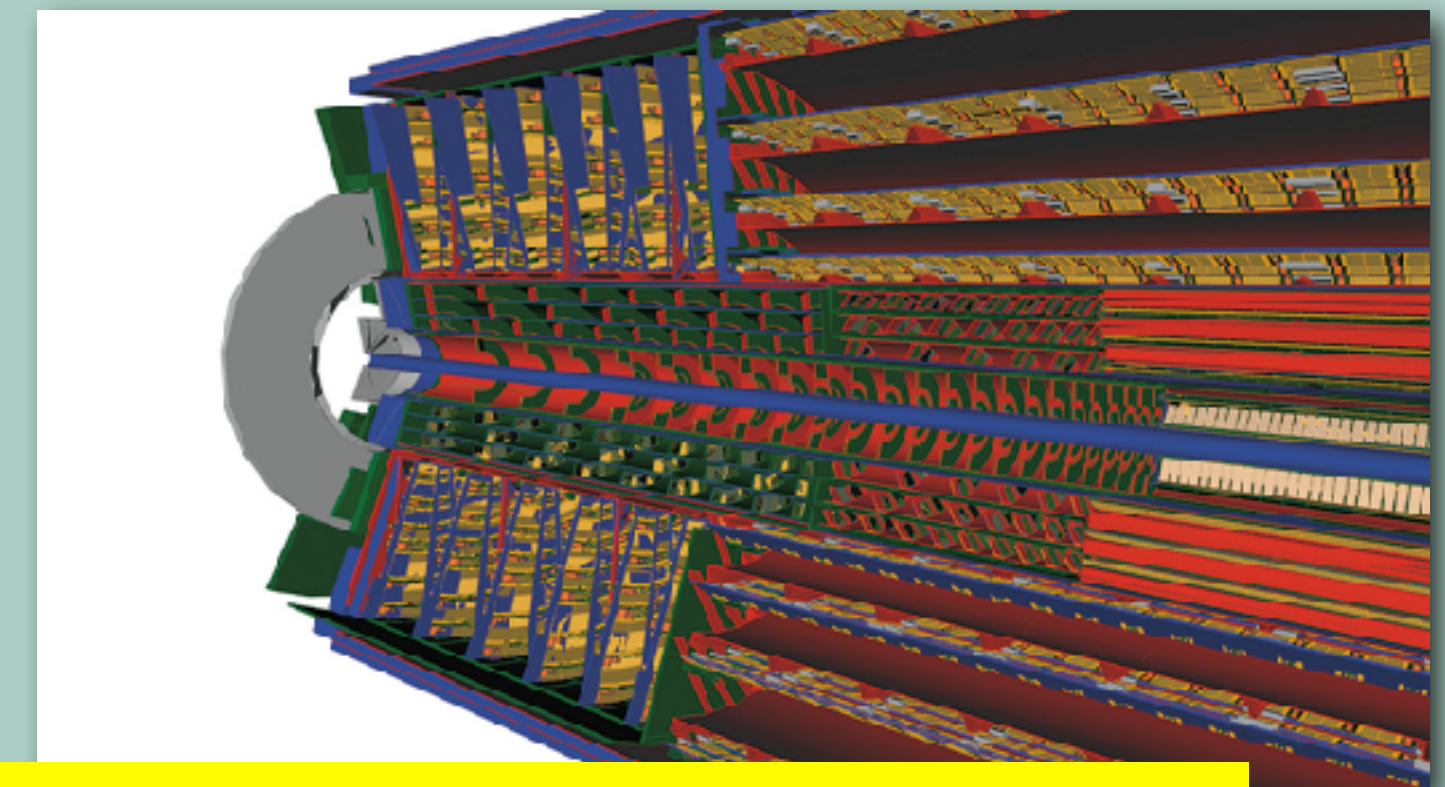
ニュートリノ振動実験

一回のバンチ衝突で 200個の衝突

up grade した測定器で解析 (日本も相当量を制作)



VS

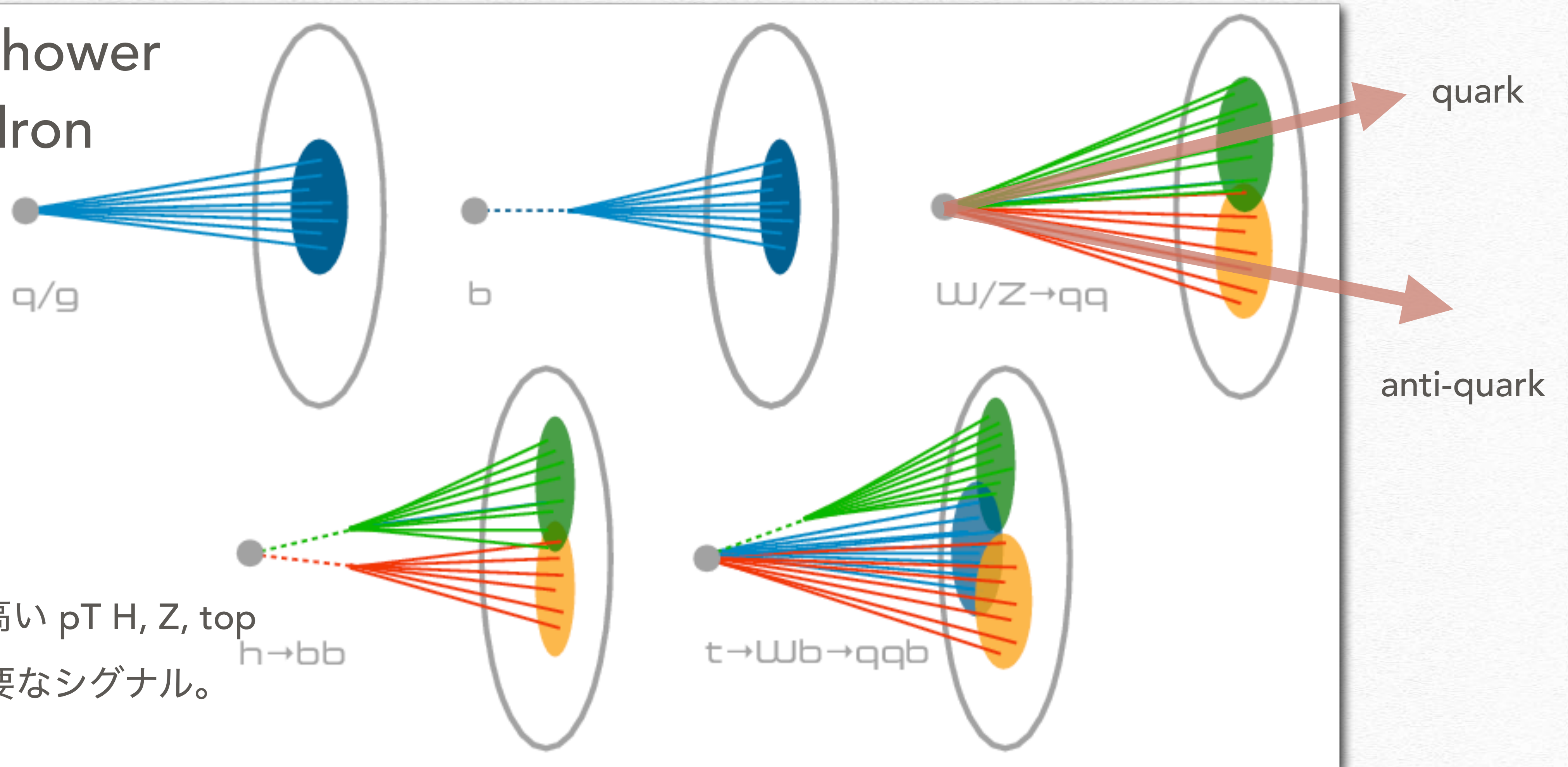


この環境でHiggs 粒子をどう研究するか。

# LHC実験におけるジェット分類と深層

## 学習

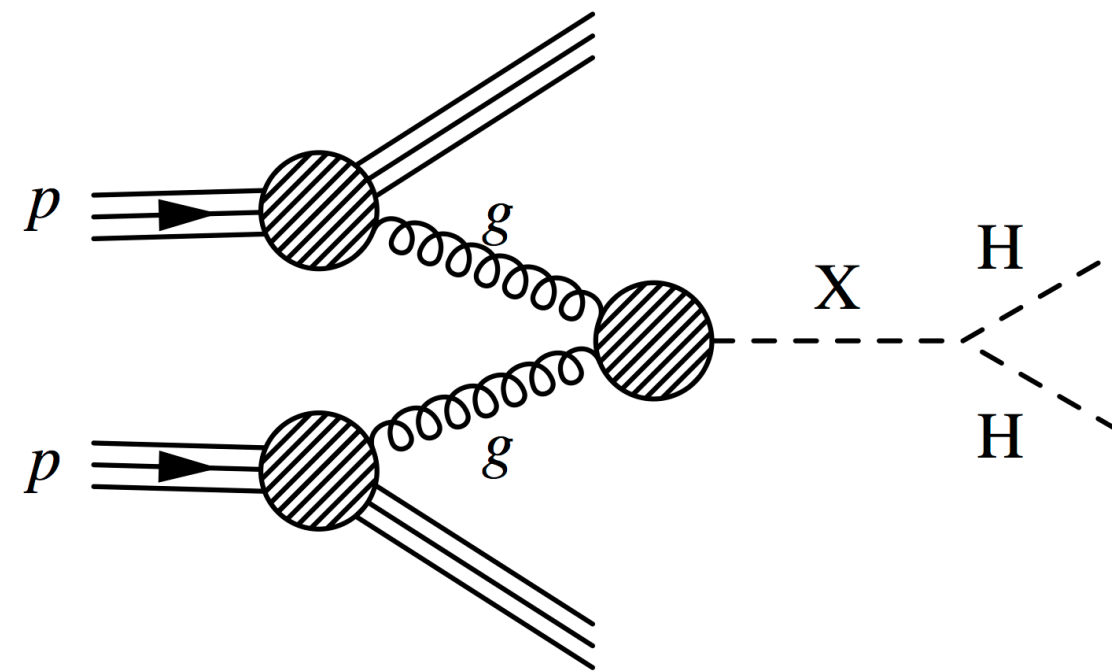
Parton shower  
→ hadron



エネルギーの高い  $p_T$   $H, Z, top$   
はBSM の重要なシグナル。

# 例: レゾナンスの再構成

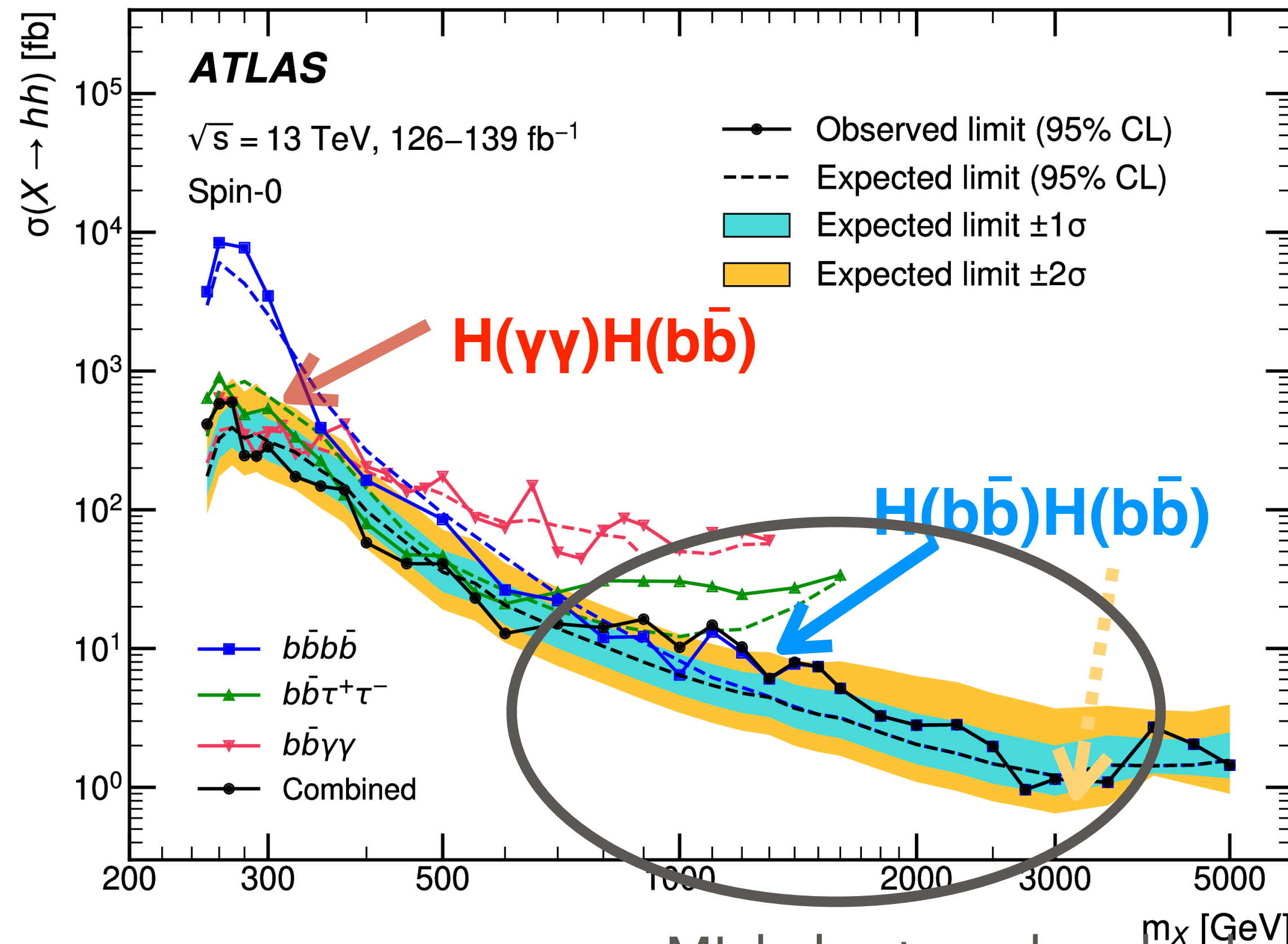
$X \rightarrow HH$



[Phys. Rev. Lett. 132 \(2024\) 231801](#)

$H(b\bar{b})H(b\bar{b})$  most sensitive channel for  $m_X > 400/500$  GeV

$H(\gamma\gamma)H(b\bar{b})$  complement in the low mass



ML helps to reduce background

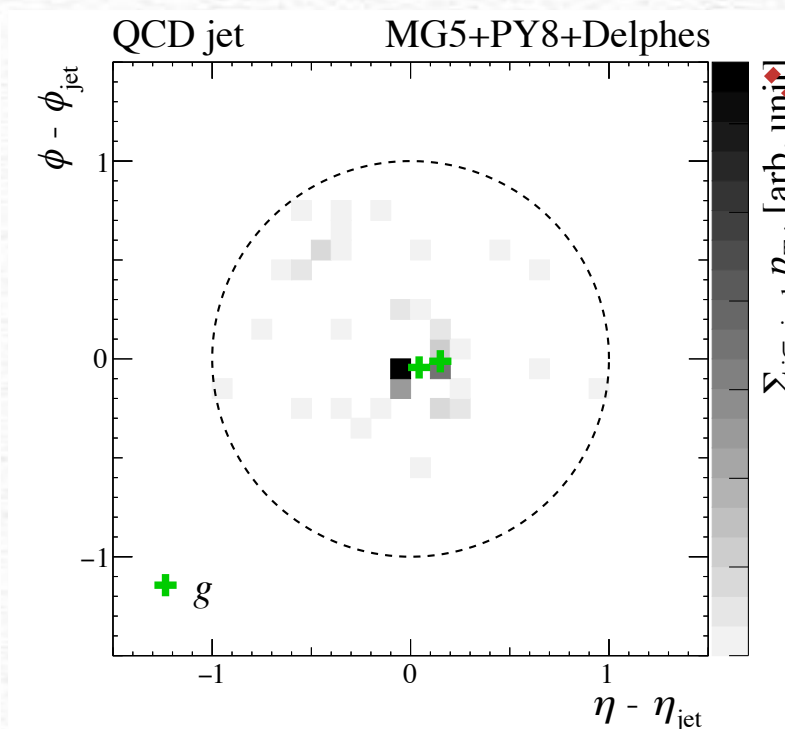
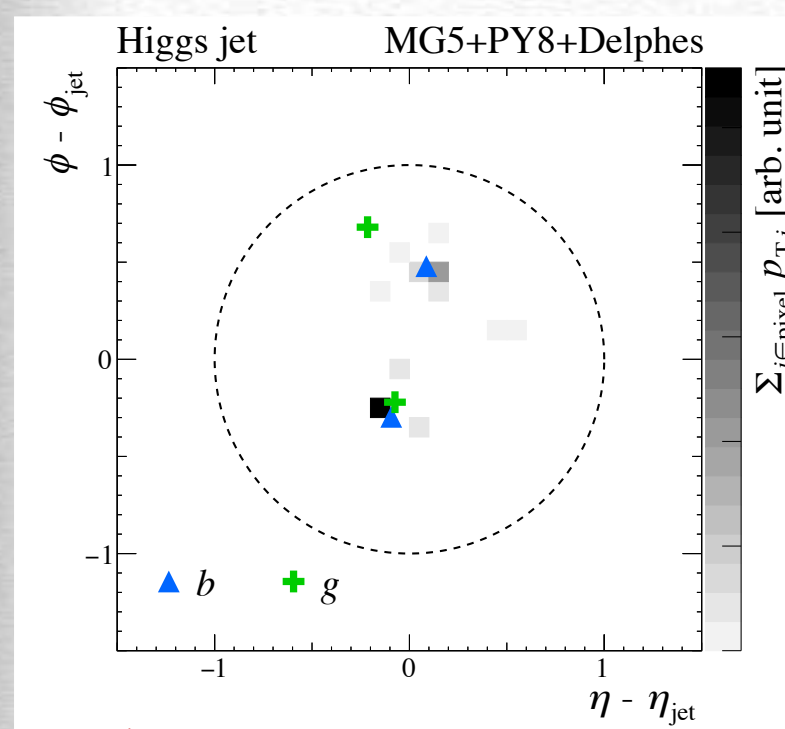
CMS 実験では  
 次のjet IDは単一の  
 深層学習モデルに

# MLP (multiple layer perceptron) の基本構造

Input: Jet images

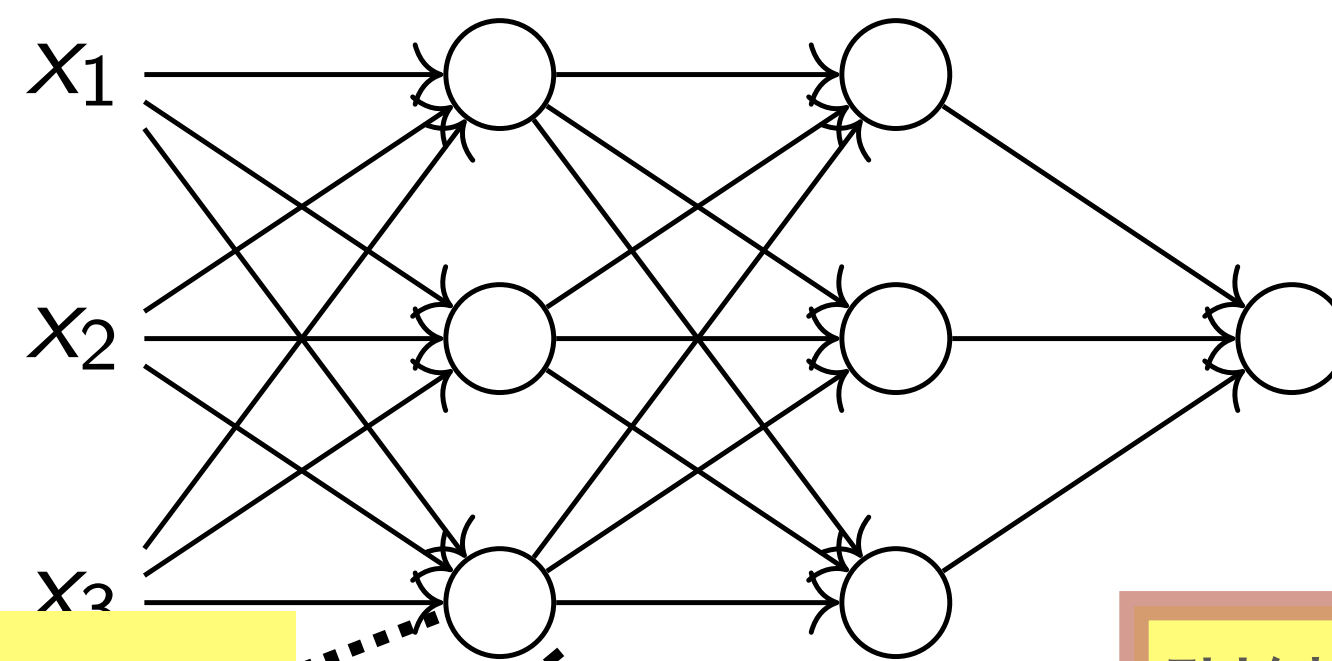
Higgs  $h+Z(\rightarrow \nu\nu)$

QCD ( $j+Z(\rightarrow \nu\nu)$ )



訓練による潜在変数の決定:  $w_{ij}, b_i$

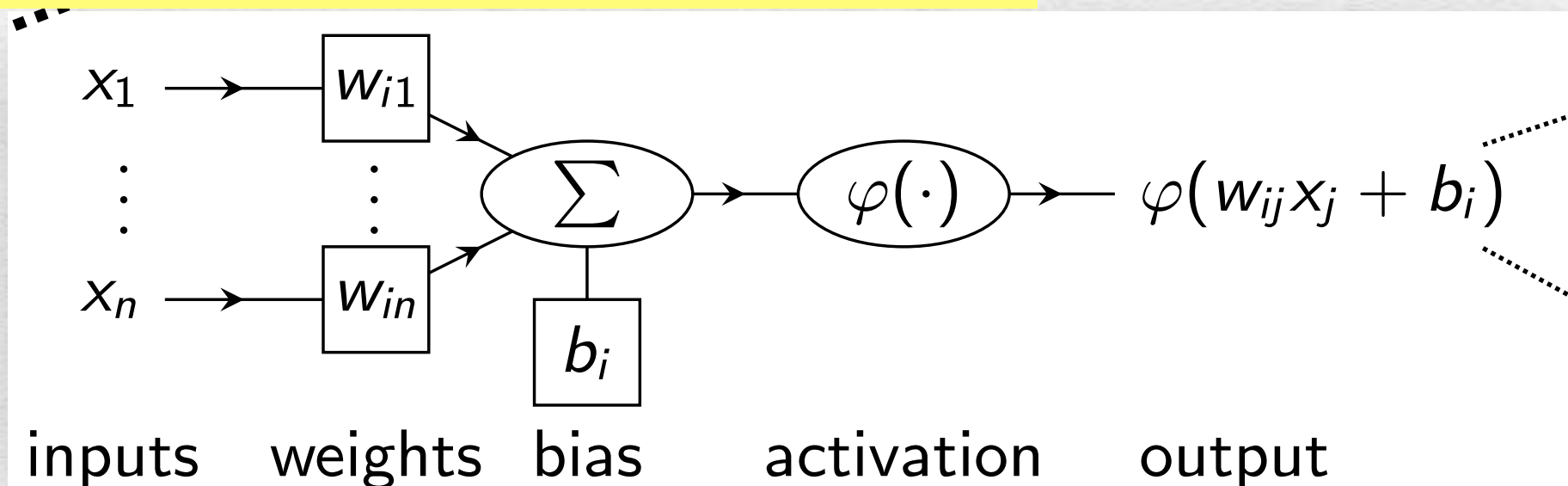
損失関数 (loss function の最適化)



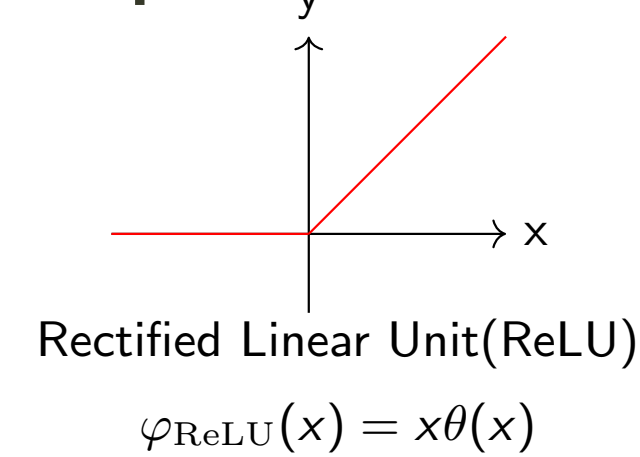
Higgs  
 $y = (1, 0)$   
 QCD  
 $y = (0, 1)$   
 $L(y, \hat{y})$

訓練後  $\hat{y}$  を使って、粒子らしさを推定

- ✓ 表現力 expressive power
- ✓ データを学習 learn from data
- ✓ 微分可能 Simple linear algebra + activation



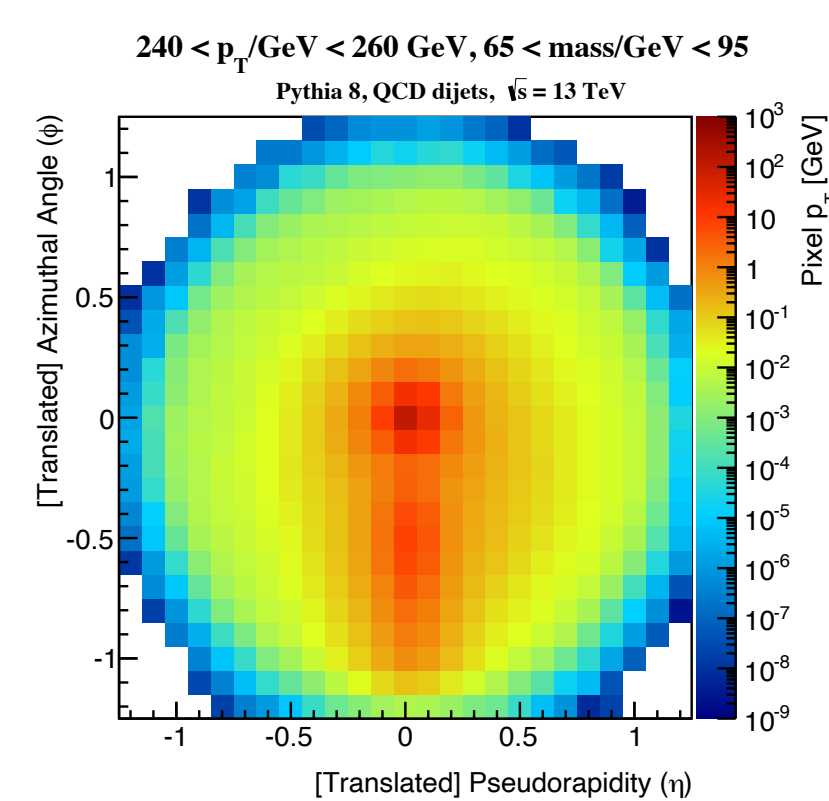
$\phi$ : 活性化関数 (非線形性)



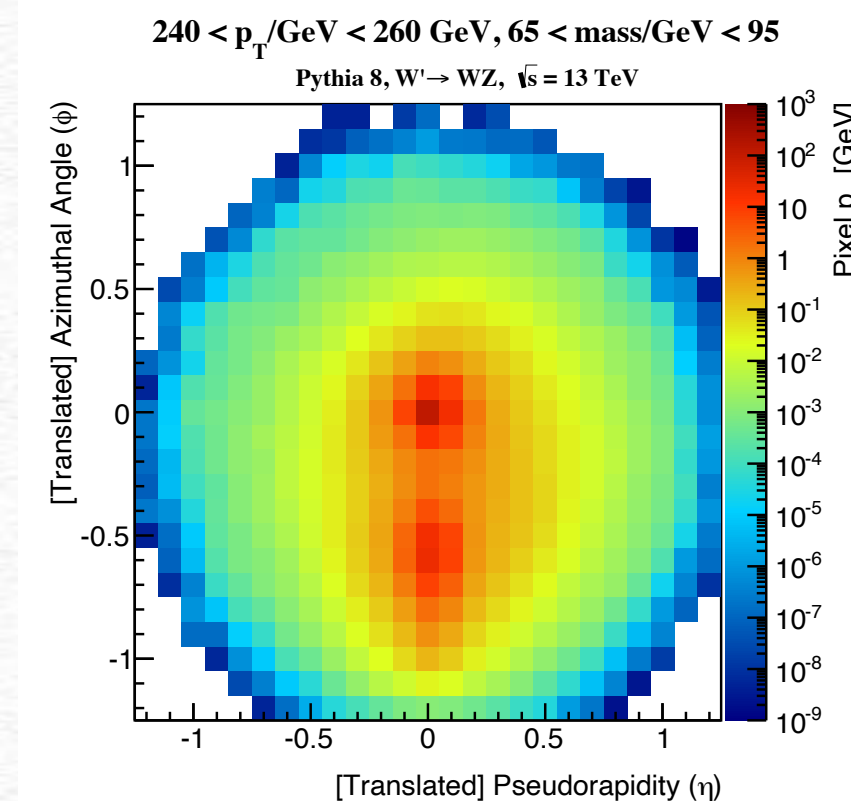


# Jet classification using ML

QCD jet  
(in W mass region)

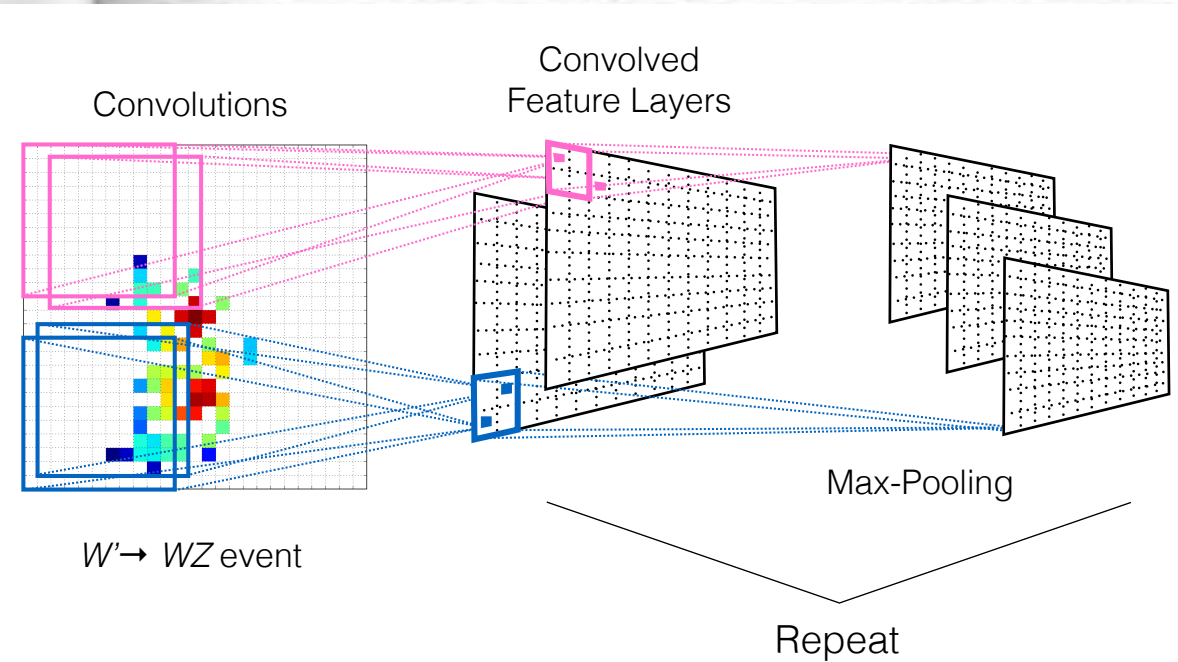


W jet



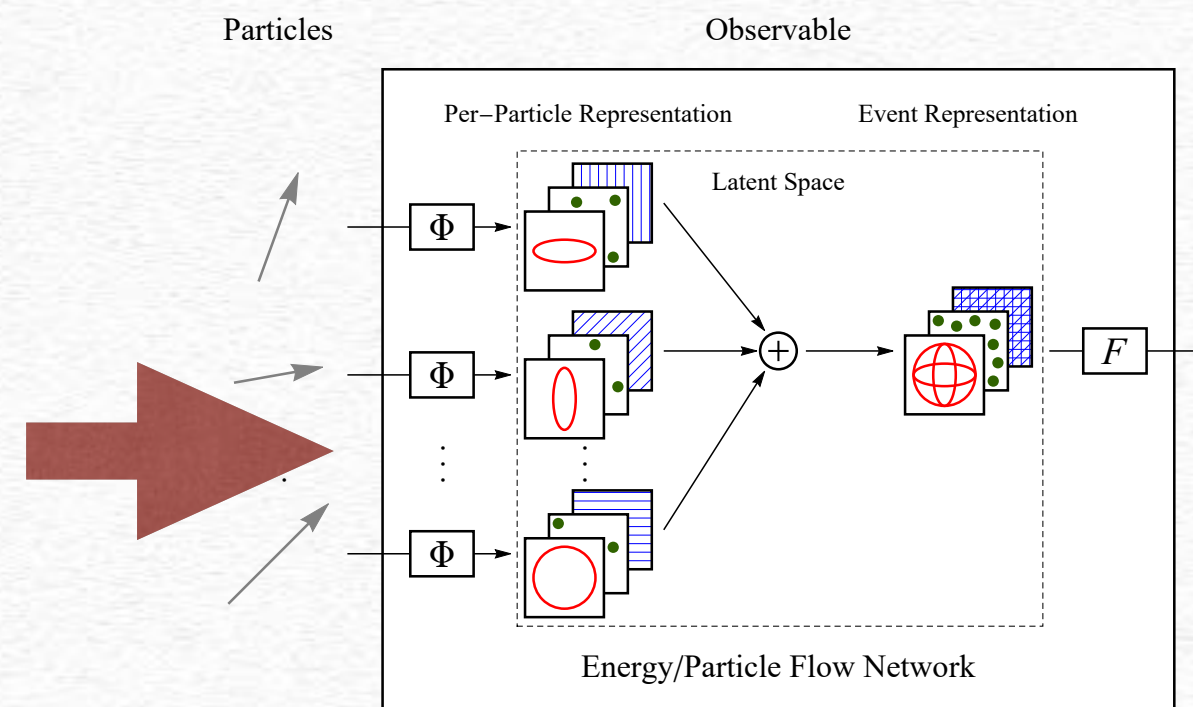
from Schwartzman et al  
<https://iopscience.iop.org/article/10.1088/1742-6596/762/1/012035>

Jet Image



CNN Oliverira et al  
(1511.05190)

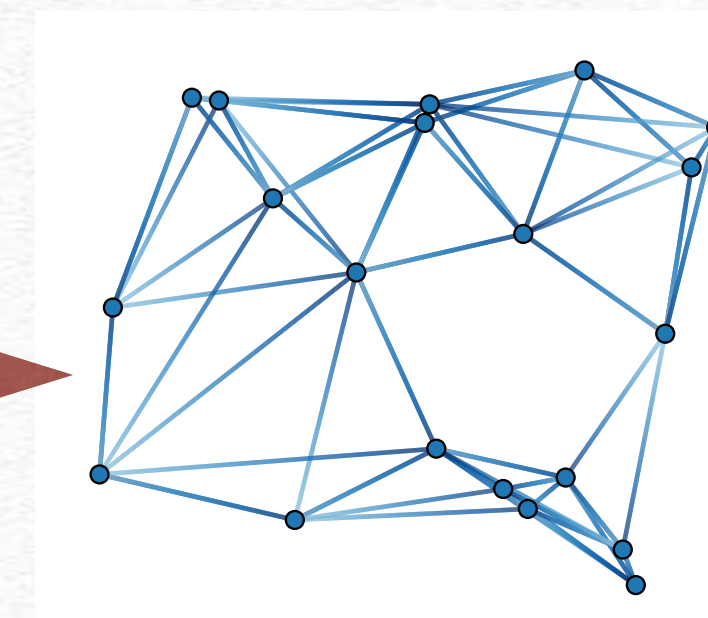
集合として



permutation invariance  
(Energy Flow Network and  
Particle Flow Network 1810.05165)

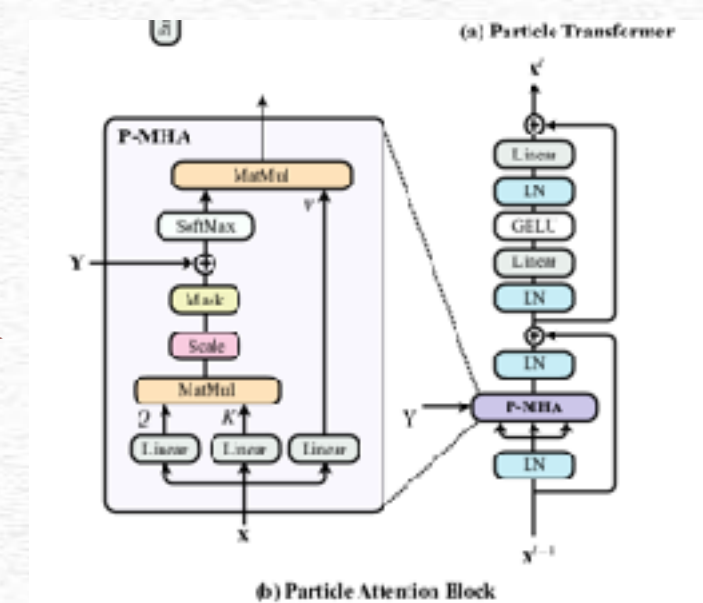
Bogatskiy et al PELICAN (2211.00454)

グラフとして



sparse data  
1902.08570 Particle Net  
Dreyer et al LundNet (1807.04758)  
Gong et al LorentzNet(2201.08187)

transformer



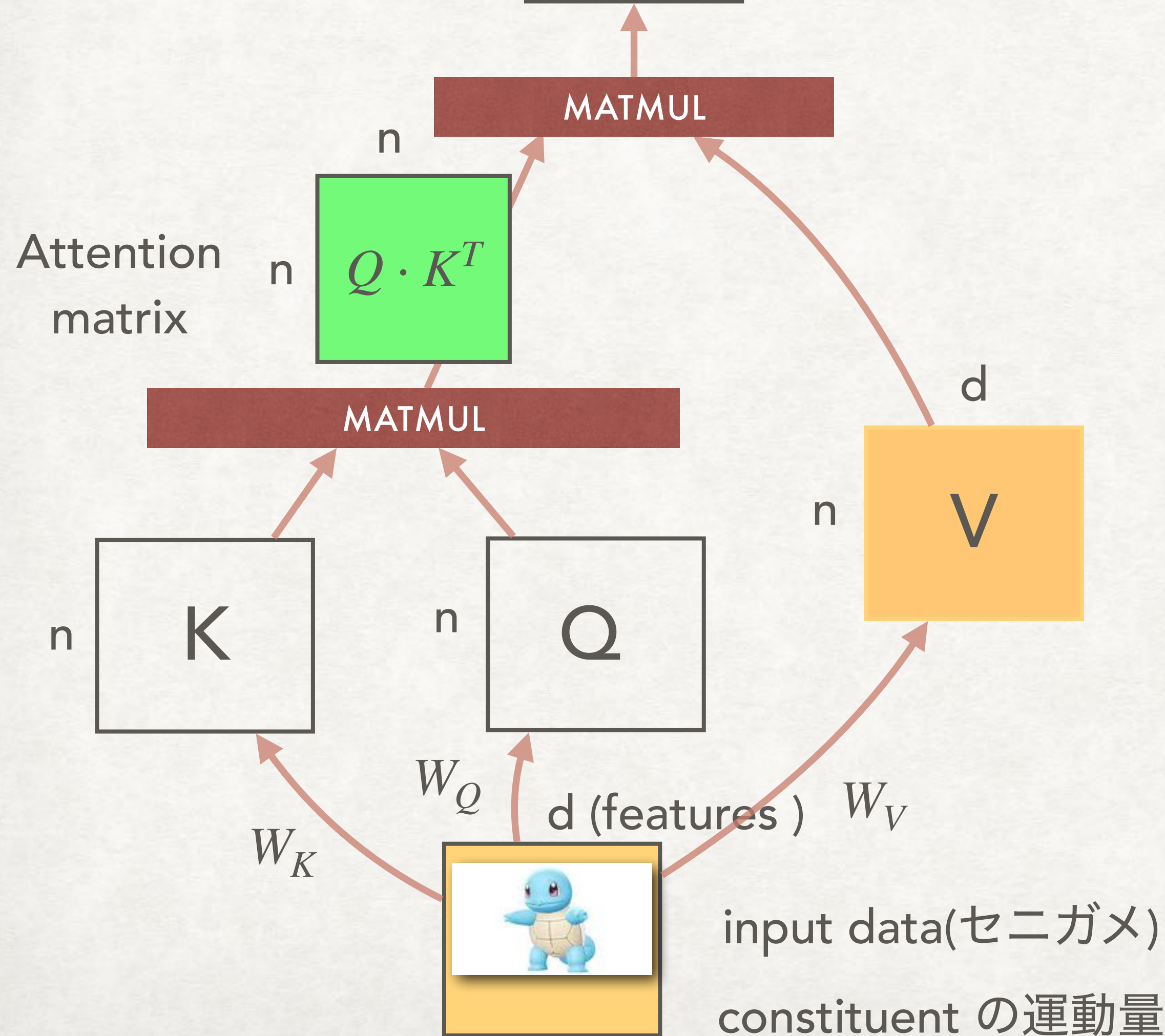
building key  
and query  
2202.03772

# “TRANSFORMER” : SELF ATTENTION LAYERS

output size = input size



$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$



- 言語モデルとして重要
- トランスフォーマーのコア部分は 「Attention Matrix」
- 全ての粒子の間の訓練 (W)可能な相関を特徴すべてを使って構成する。
- アテンションブロックの入力と出力の次元が同じ。
- $X \rightarrow X' \rightarrow X''$  と入力がどんどん進化していく。

CMS will have jet trigger using transformer soon

# なんでもデータとしていられるのが楽しい

Table 2. Particle input features used for jet tagging on the JETCLASS, the top quark tagging (TOP) and the quark gluon tagging (QG) datasets. For QG, we consider two scenarios: QG<sub>exp</sub> is restricted to use only the 5-class experimentally realistic particle identification information, while QG<sub>full</sub> uses the full set of particle identification information in the dataset and further distinguish between different types of charged hadrons and neutral hadrons.

運動量

電荷、particle ID など

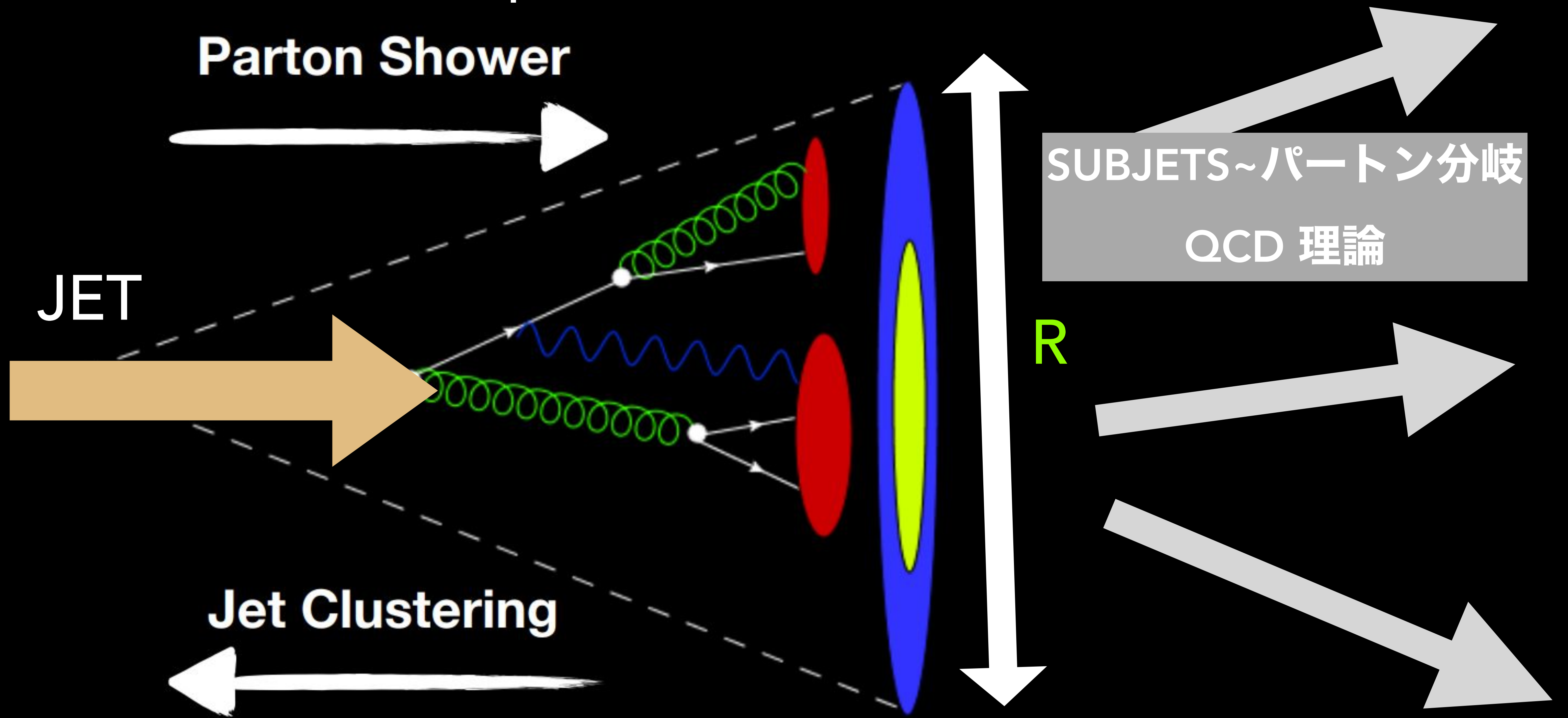
衝突点から飛んだ距離

Category	Variable	Definition	JETCLASS	TOP	QG <sub>exp</sub>	QG <sub>full</sub>
Kinematics	$\Delta\eta$	difference in pseudorapidity $\eta$ between the particle and the jet axis	✓	✓	✓	✓
	$\Delta\phi$	difference in azimuthal angle $\phi$ between the particle and the jet axis	✓	✓	✓	✓
	$\log p_T$	logarithm of the particle's transverse momentum $p_T$	✓	✓	✓	✓
	$\log E$	logarithm of the particle's energy	✓	✓	✓	✓
	$\log \frac{p_T}{p_{T(\text{jet})}}$	logarithm of the particle's $p_T$ relative to the jet $p_T$	✓	✓	✓	✓
	$\log \frac{E}{E(\text{jet})}$	logarithm of the particle's energy relative to the jet energy	✓	✓	✓	✓
	$\Delta R$	angular separation between the particle and the jet axis ( $\sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ )	✓	✓	✓	✓
Particle identification	charge	electric charge of the particle	✓	—	✓	✓
	Electron	if the particle is an electron ( $ \text{pid} ==11$ )	✓	—	✓	✓
	Muon	if the particle is an muon ( $ \text{pid} ==13$ )	✓	—	✓	✓
	Photon	if the particle is an photon ( $\text{pid}==22$ )	✓	—	✓	✓
	CH	if the particle is an charged hadron ( $ \text{pid} ==211$ or $321$ or $2212$ )	✓	—	✓	✓ <sup>a</sup>
	NH	if the particle is an neutral hadron ( $ \text{pid} ==130$ or $2112$ or $0$ )	✓	—	✓	✓ <sup>b</sup>
Trajectory displacement	$\tanh d_0$	hyperbolic tangent of the transverse impact parameter value	✓	—	—	—
	$\tanh d_z$	hyperbolic tangent of the longitudinal impact parameter value	✓	—	—	—
	$\sigma_{d_0}$	error of the measured transverse impact parameter	✓	—	—	—
	$\sigma_{d_z}$	error of the measured longitudinal impact parameter	✓	—	—	—

<sup>a</sup> ( $|\text{pid}|==211$ ) + ( $|\text{pid}|==321$ )\*0.5 + ( $|\text{pid}|==2212$ )\*0.2

<sup>b</sup> ( $|\text{pid}|==130$ ) + ( $|\text{pid}|==2112$ )\*0.2

より物理的なpicture がほしい



QCD community have developed sophisticated theoretical treatment about this.

# ネットワークを作る前に物理を考える

- 最初の高エネルギー衝突 Partons(quarks and gluons)  $\{y\}$
- 構造のあるジェット  $P(\{x_i\} | \{y_\alpha\})$
- ジェットの中のハドロン

$$P(\text{hadrons in jets} | \text{parton} \sim \text{jet}) = P(\{x_i\} | \{y\})$$

- 複数の構造のあるイベントをもつ事象

$$P(\{x_i\}, \{x'_j\}, \{y_\alpha\}, \{y'_\beta\}) \sim P(\{x_i\} | \{y_\alpha\}) P(\{x'_j\} | \{y'_\beta\}) P(\{y_\alpha, y'_\beta\})$$

$y$ : ハードな衝突なら 1TeV 前後のスケール    ジェット内部構造なら 100GeV スケール    ハドロンは 1GeV

QCD のもつ構造 (factrization, parton shower) がイベントを作ってる。

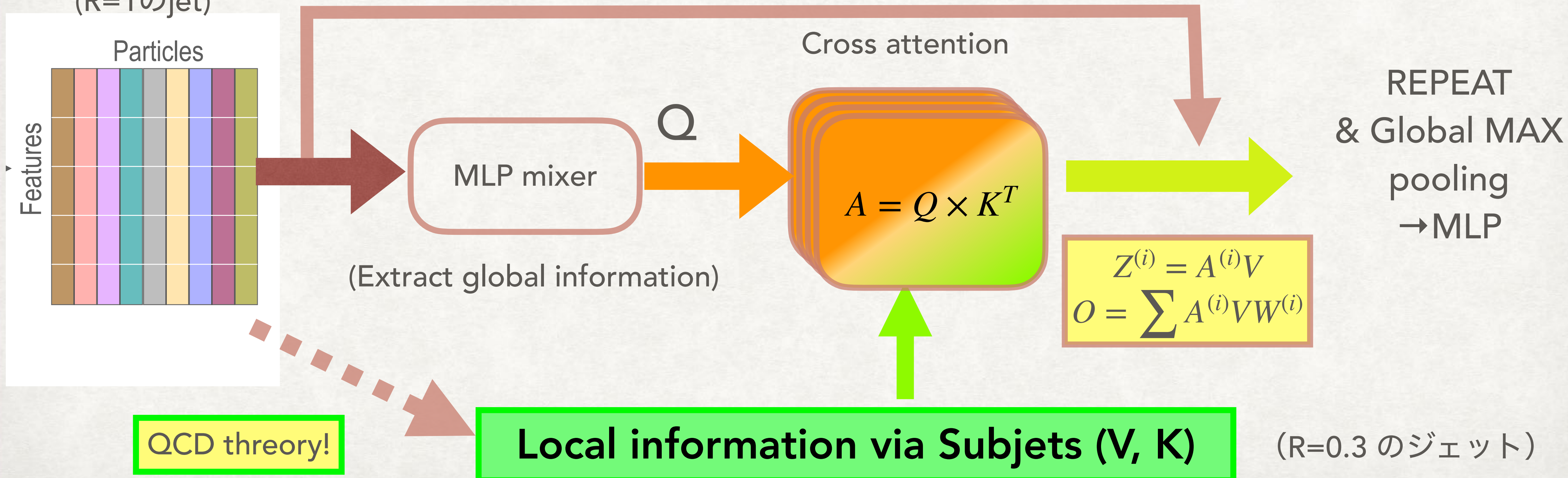
# QCDのファクトリゼーション構造をとらえたネットワーク

ATTENTION → CROSS Attention による  $P(h| \text{subjets})$  評価

input  $X$

( $R=1$ のjet)

skip connection  $\tilde{X} = X + O$



構造があるものに関してはその構造を尊重したネットワークが良い

LHC process

Hard scattering

Jet function Parton shower

$$\sigma(pp \rightarrow a, b \rightarrow N \text{jets}) \sim H_N \left[ B_a B_b \prod_{k=1}^N J_k \right] \otimes S_N,$$

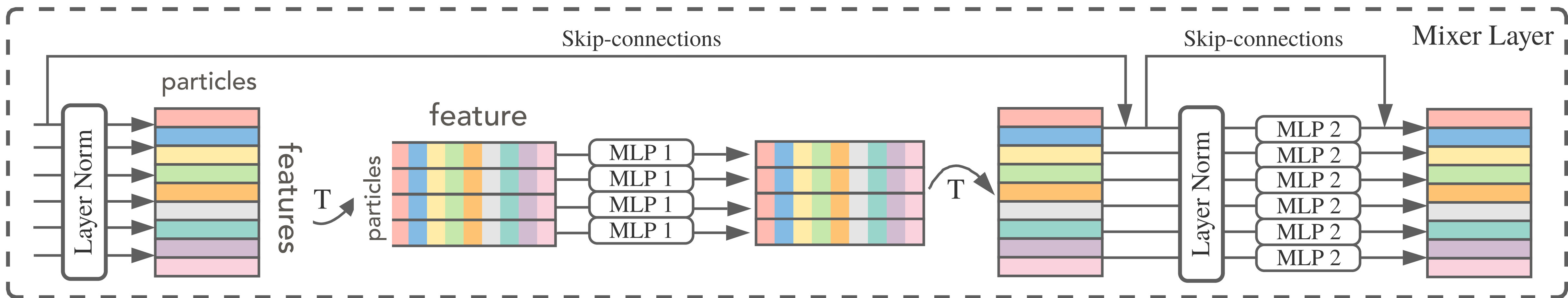
全てのデータをまとめてself attention するとこの構造は実現しない

$$A V = \begin{pmatrix} Q(\text{constituent}) \times K(\text{constituent}) & Q(\text{constituent}) \times K(\text{subj}) \\ Q(\text{subj}) \times K(\text{constituent}) & Q(\text{subj}) \times K(\text{subj}) \end{pmatrix} V = Q(\text{subj}) K(\text{subj}) V(\text{subj}) + \dots$$

一方 cross attention にすると  $Q(\Phi_{1\theta_1}(I_1)) \cdot K(\Phi_{2\theta_2}(I_2))$

と、必ず二つの量の積として確率が計算されるので、スケールヒエラルキーのある物理を上からしたまで記述するのに有効

大枠は捉えた。ソフトな情報はMLP MIXERで



MLP 1 : feature だけに作用する。

MLP 2: feature に作用せず、粒子を混ぜる。

MLP-Mixer は画像に認識の  
アルゴリズムで有名



# Performace comparable to Particle Transformer but much faster and lighter

Models	AUC	R50%	#Parameter	Time (GPU%)
ParT	0.9858	413+-16	2.14M	612
Mixer+subjet (CA)	0.9856	392+-5	86.03K	33
(AK)	0.9854	375+-5	86.03K	33
(HDBSCAN)	0.9859	416+-5	86.03K	33
LorentzNet	0.9868	498+-18	224K	
PELICAN (Lorents Invariance)	0.9869	-	45K	-

\*Subjet cone size  $R=0.3$

\*HDBSCAN is algorithm without distance measure

Performace comparable to Particle Transformer but much faster and lighter

Models	AUC	R50%	#Parameter	Time (GPU%)
ParT	0.9858	413+-16	2.14M	612
Mixer+subjet (CA)	0.9856	392+-5	86.03K	33
(AK)	0.9854	375+-5	86.03K	33
(HDBSCAN)	0.9859	416+-5	86.03K	33
LorentzNet	0.9868	498+-18	224K	-
PELICAN (Lorents Invariance)	0.9869	-	45K	-

SMALL SIZE

612

FAST

\*Subjet cone size  $R=0.3$

\*HDBSCAN is algorithm

HIGH PERFORMANCE WITHOUT USING LORENTS INVARIANCE

without distance measure

# APPLICATION 2

## TOWARD GLOBAL EVENT ANALYSIS

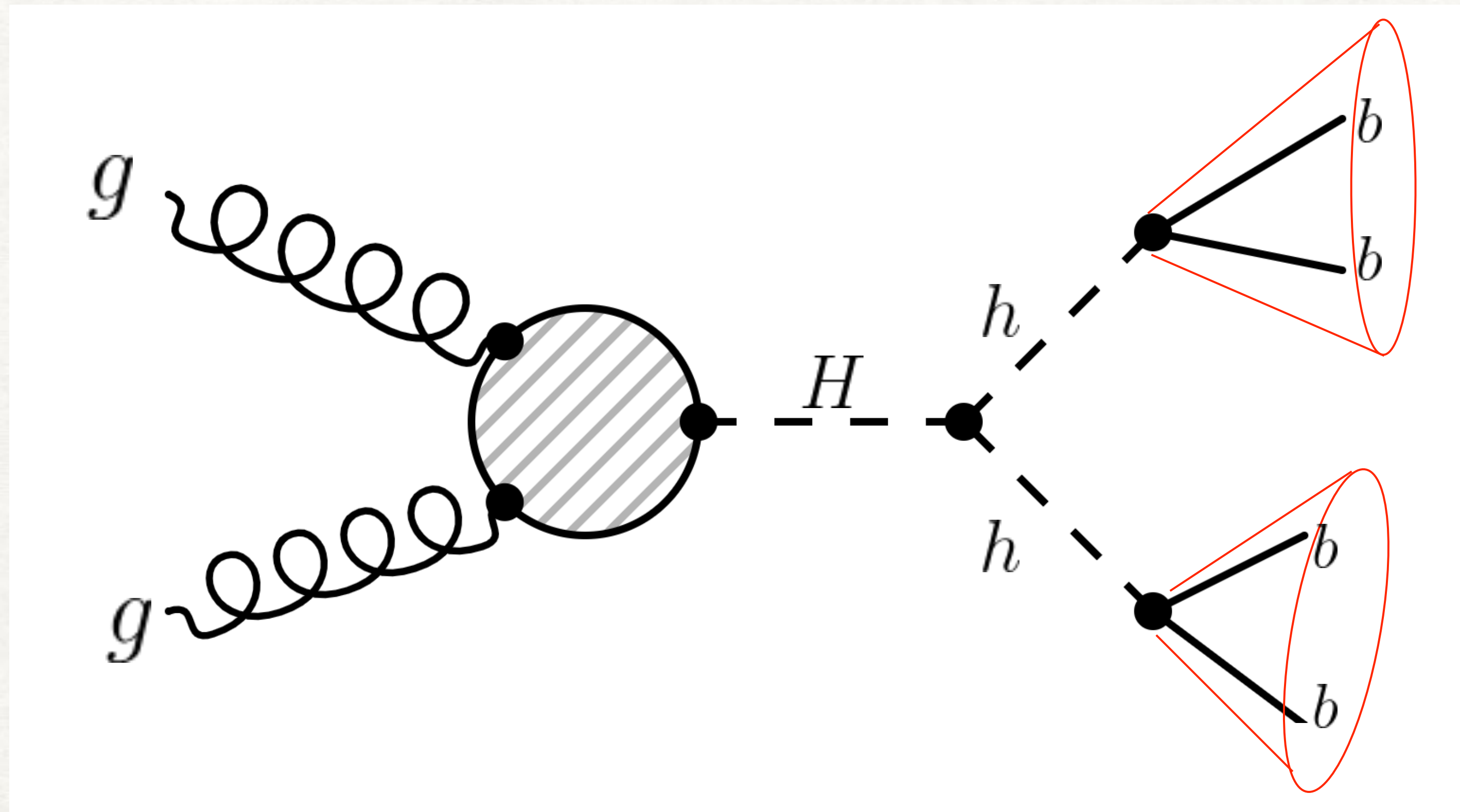
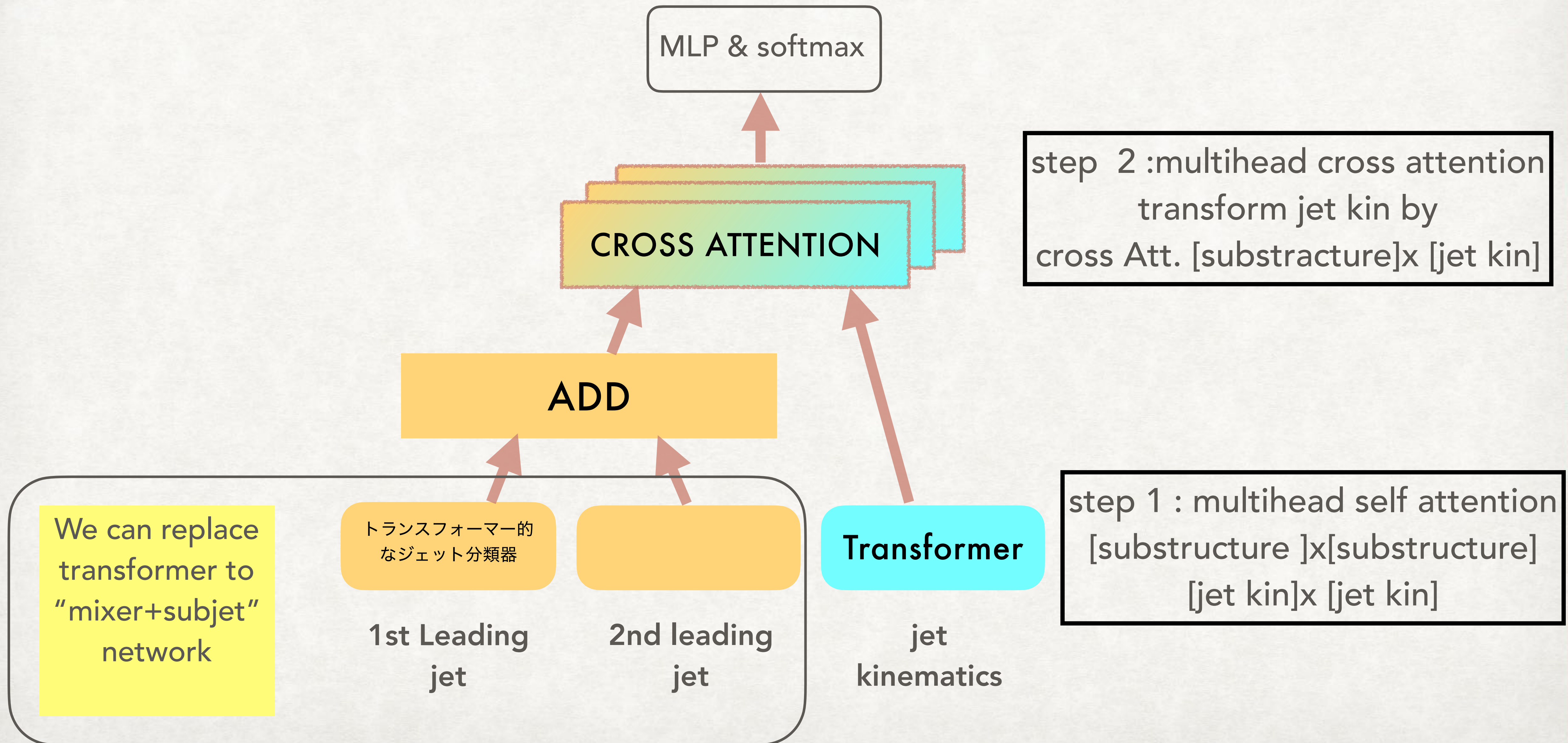


Figure 2: Feynman diagram for the signal process.

# cross attention motivation for 2 fatjet events



# INPUT TO NETWORK : EVENT KINEMATICS

## 測定器の円筒座標

Kinematical inputs (3, 6)

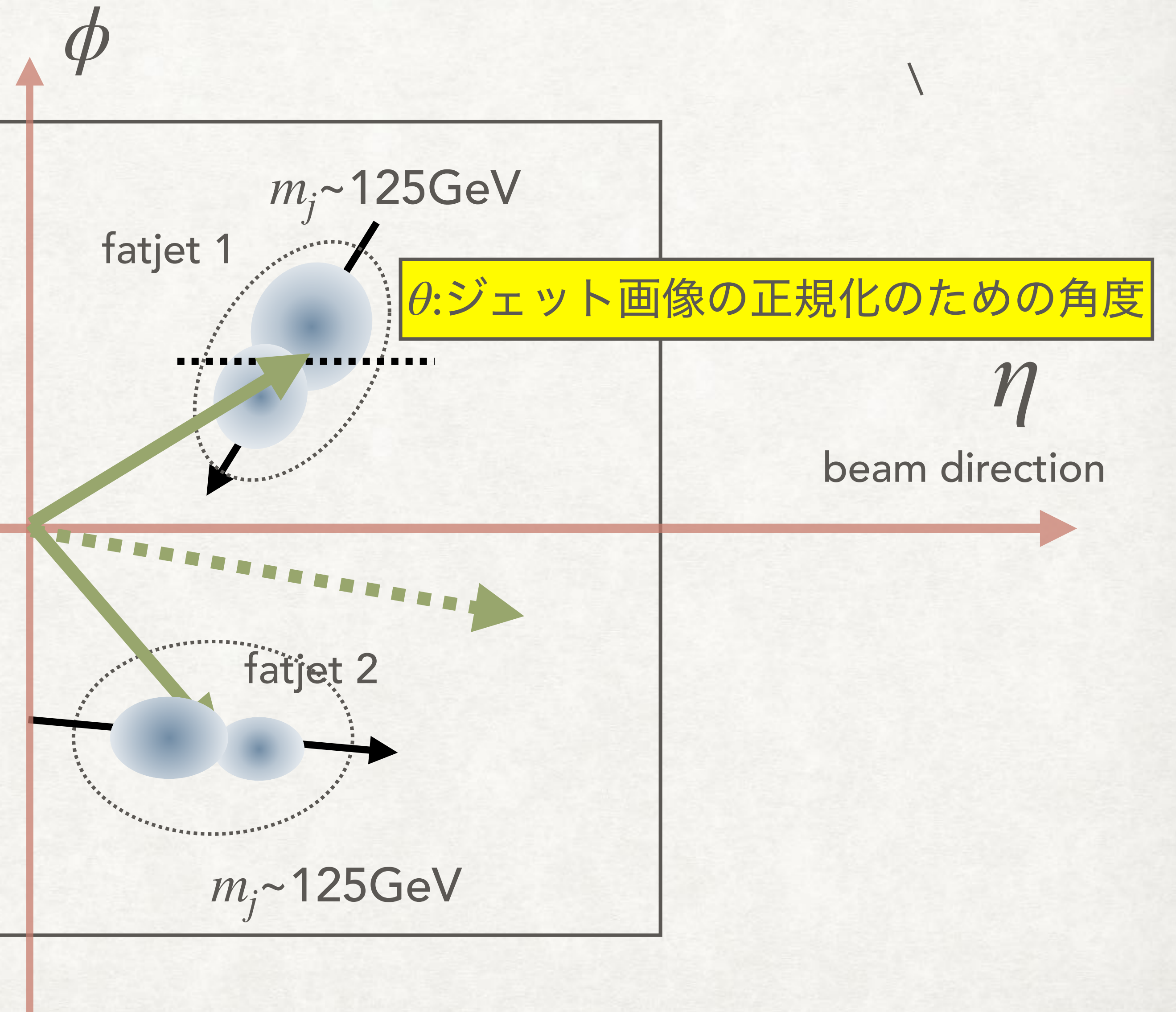
fatjet 1 =  $(m_1, \eta_1, \phi_1, p_{T1}, E_1), \theta_1$

fatjet 2 =  $(m_2, \eta_2, \phi_2, p_{T2}, E_2), \theta_2$

H candidate =  $(m_{12}, \eta_{12}, \phi_{12}, p_{T12}, E_{12}), \theta_{12} = 0$

NOTE :

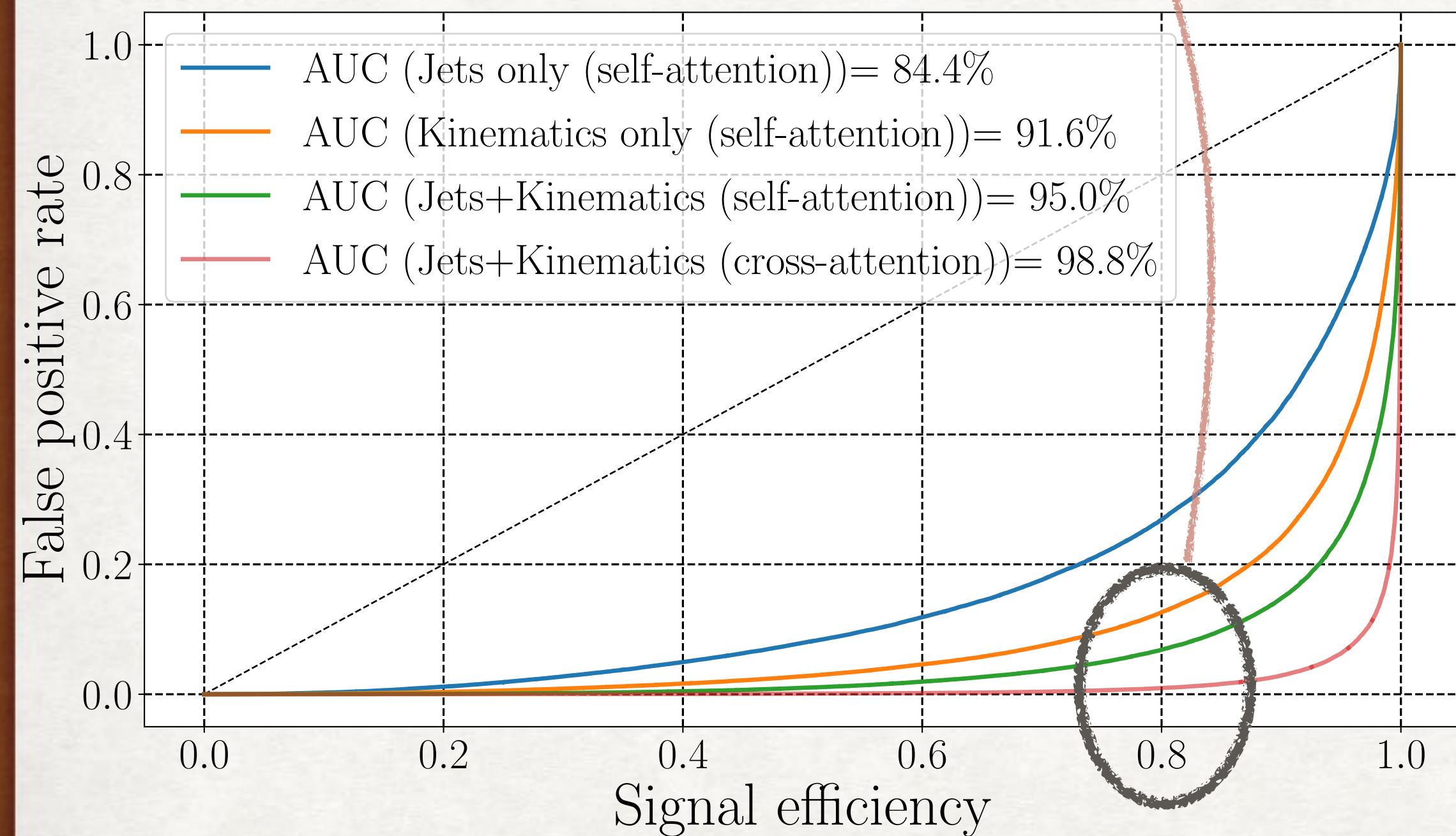
1. "5 inputs for 4 momentum" ,
2. H candidate momentum as sum of the fat jet momentum.
3. add " $\theta$ " :the correlation beyond a subjet



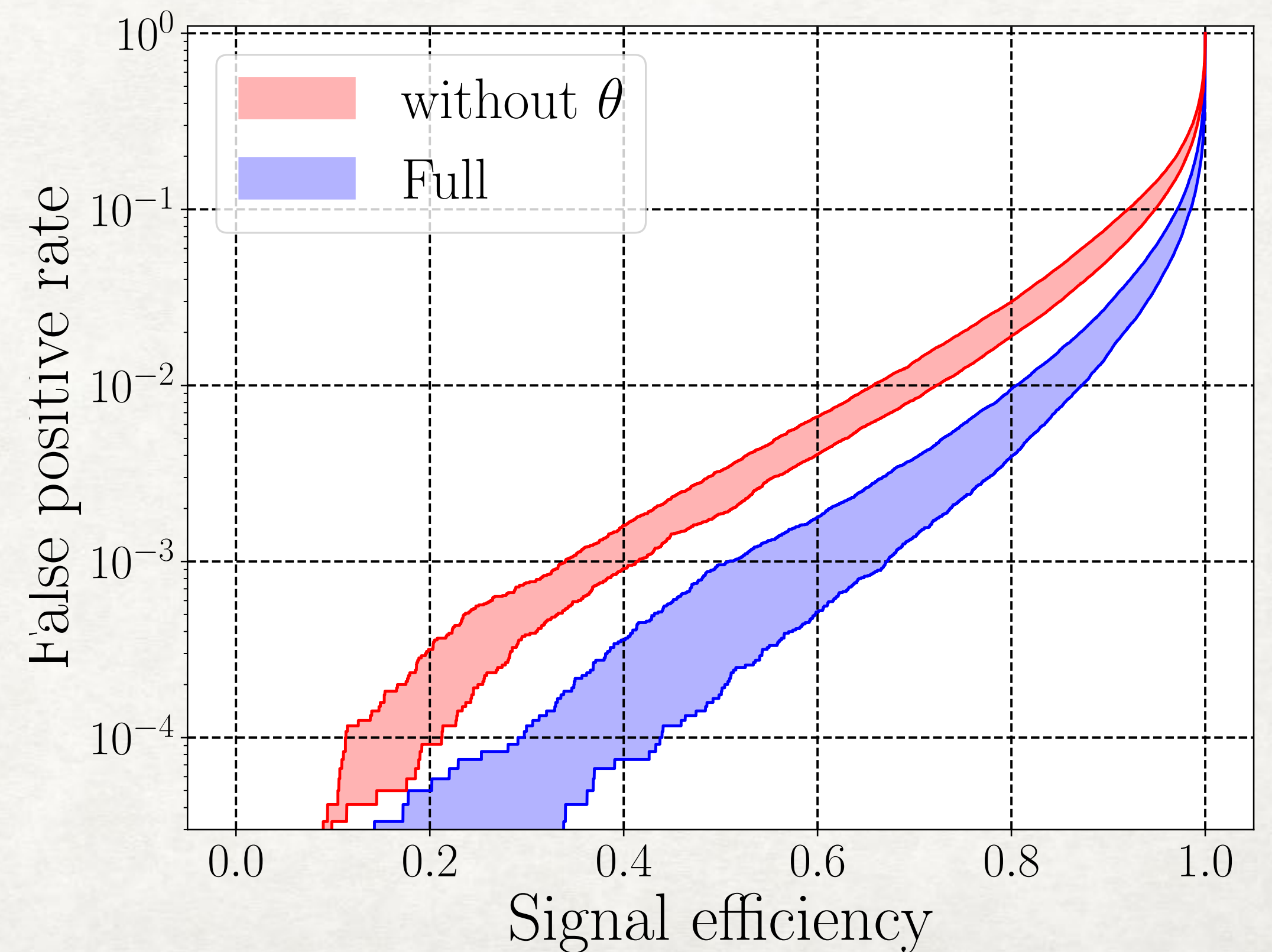
# IMPROVEMENT USING CROSS ATTENTION

factor 5 improvement at the same acceptance.

Decay correlation is important  
(because QCD background is correlated)



Cross attention improves the rejection efficiency significantly



## 今後：素粒子における深層学習の応用

懐疑的に考えられていた、深層学習の利用はこの十年で市民権を獲得した。

素粒子実験のデータ生成やデータ解析は相当部分深層学習に置き換わると考えられる。

深層学習が物理をやってくれるわけではないが、物理と深層学習を正しく理解していれば、やくには立つ。素粒子理論→他粒子の相関の計算 は深層学習が使うような情報を正しく  
予言できるようにupdate されるべき。

これが新しい発見や深層学習の一般的な応用へのインパクトにつながるかはよくわからない。

# 物理と深層学習

- ファクトリゼーションの構造(このトーク)
- Parton branch
- ローレンツ不変性
- ジェット記述



# ローレンツ対称性への配慮

★ジェットの分類結果は、ローレンツフレームによらないことを深層学習に組み込みたい

- データのフレームを固定する(楽。ジェットの軸をそろえたりする。)
- データを回転する等して水増しする (データサイエンスの世界ではありがち。無駄。)
- ネットワークそのものをローレンツ不変に組む(最強だが面倒。測定エラーに弱いかもという話も)

Inputs  $p_i \rightarrow p_i \cdot p_j$  permutation invariance  $F(\pi \circ T_{ijk\dots}) = \pi \circ F(T_{ijk\dots})$ , Use equivariant layer  $T'_{ij} = \sum_{k,l=1}^N B_{ijkl}^a T_{kl}$

といった対応がとられる。

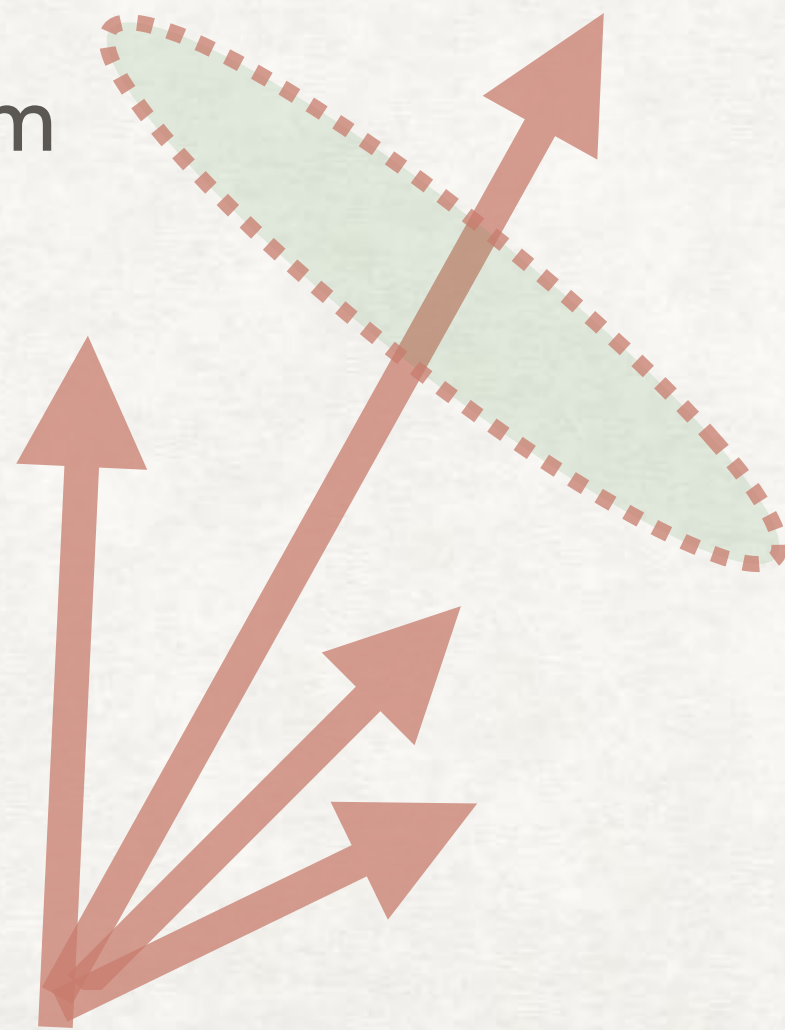
ローレンツ不変な量だけを使ったネットワークは、比較的少数のイベントで、良い分類結果を出す。データを同等なクラスに事前に分類していることに対応しているため。参考 LorentzNet 2201.08187 PELICAN

(arXiv 2211.00454) L-GATr 2405.14806 (Lorentz Geometric Algebra transformer)

# ジェットクラスタリングアルゴリズムを深層学習に生かす

- ジェットの定義—距離の近い粒子をマージして新しい粒子を定義することを繰り返す  
seedless アルゴリズム( 粒子集合に働くアルゴリズム)
- $J \rightarrow J_1^{(1)}, J_2^{(1)}, J_1^{(1)} \rightarrow J_1^{(2)} J_2^{(2)}, J_1^{(2)} \rightarrow \dots$

cone algorithm

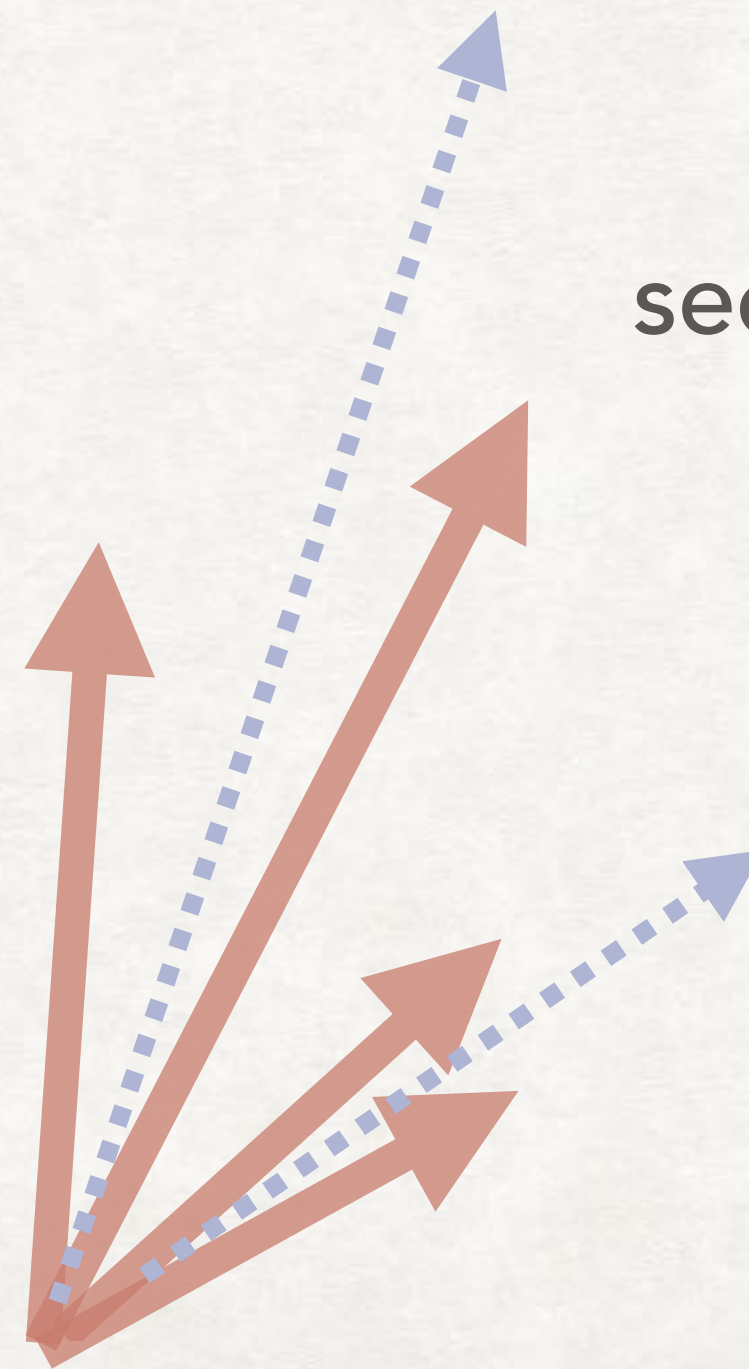


衝突点

seedless algorithm

$$R_{ij} = E_i^k E_j^k (\Delta\theta_{ij}), \quad \Delta\theta = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$$

小さい  $R_{ij}$  のペアから運動量をマージして、  
新しいオブジェクトを作る



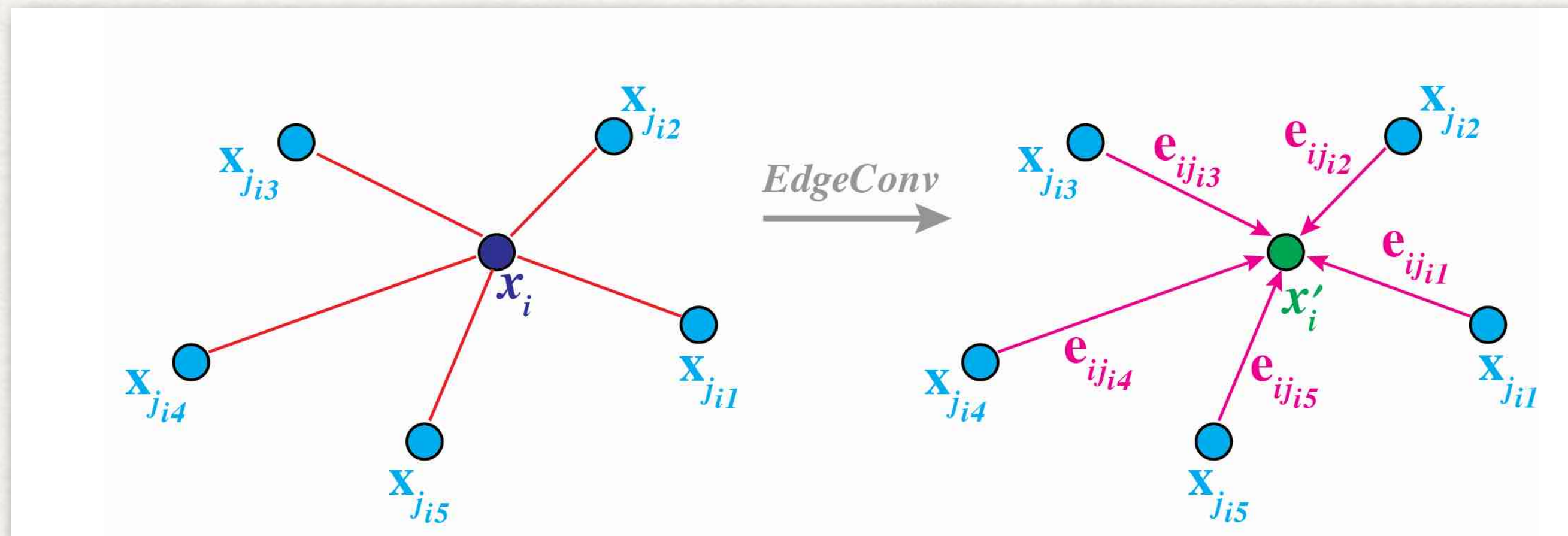
# 物理をもとにした深層学習の取り組み LUND NET(2012.08526)

Particle NET : 粒子-粒子間の繋がりをふやすと計算量が膨大に→ GPUメモリが爆増

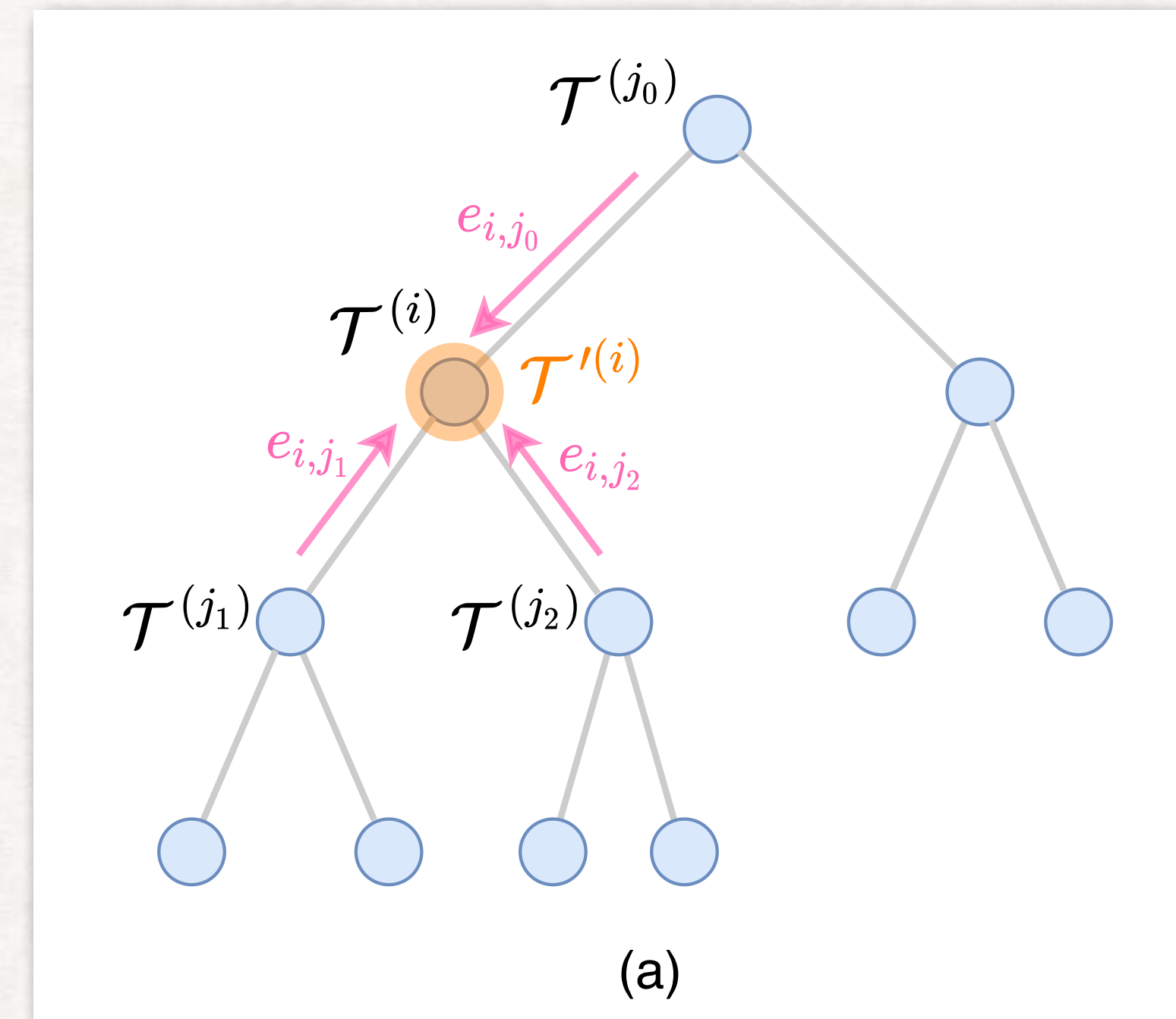
LundNet: 粒子の情報:サブジェット情報

グラフ:ジェットクラスタリング過程

「問題によらないモデルの構成」



「物理的な知見」を踏まえた模型



結果がソフトな末端の分岐にとってもセンシティブ

# 深層学習が使うような粒子相関を正しく理論で記述したい。

❖ QCDのループ補正を高次までやる。QCDの $N^2LL$ 次数( $\alpha_s^3$ )が目標。

❖ **パートンシャワーの高度化**  $\sum_N [\alpha_s \log Q^2/\mu^2]^n$  (LL)  
+  $\sum_N \alpha_s [\alpha_s \log Q^2/\mu^2]^n$  (NLL) +  $\sum_N \alpha_s^2 [\alpha_s \log Q^2/\mu^2]^n$

❖ 多数の終状態に対応する。(計算の並列化, Parton shower ML のマッチング)

❖ ハドロン生成 (パートンシャワーを対応後になんとかする。)

2000年あたりから、着実に進展し、実験データと数値計算が合うように。10%以下の精度を出すにはNNLL程度を目指す必要がある。

高性能計算(HPC)の進展

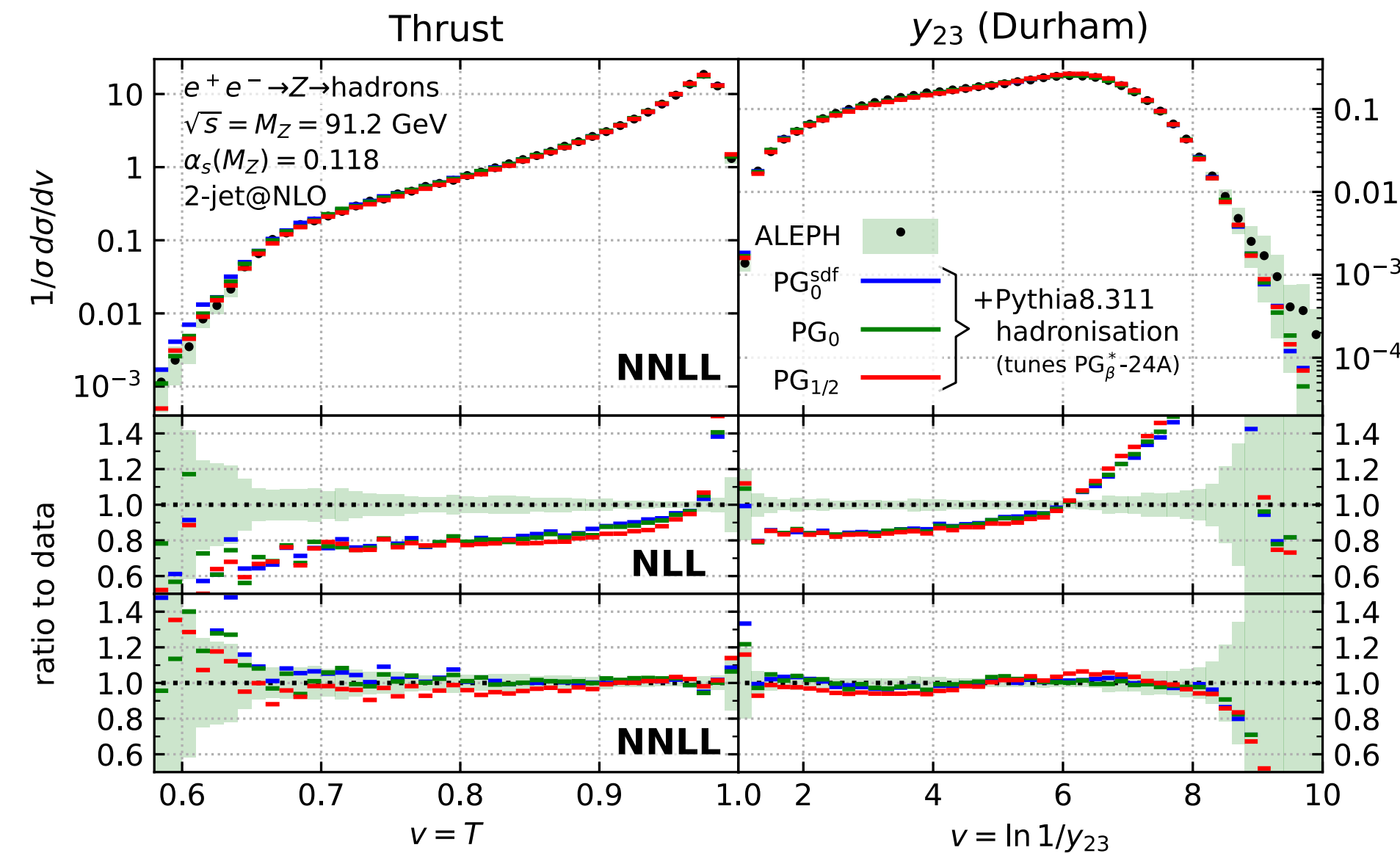
CPU → 多重コア

→ GPU[深層学習]

(簡単な計算をたくさんするのが得意になった)

## これからのジェットの物理

# Comparison to LEP data



$$\alpha_s(m_Z) = 0.118$$

Colour is handled using the NODS scheme which gives full colour accuracy at NLL for global observables (includes those shown)

\*This should be taken as an average  $\alpha_s^{\text{eff}}$  not an  $\alpha_s^{\overline{MS}}$

- Inclusion of NNLL potentially resolves the issue of needing an anomalously large value of  $\alpha_s(m_Z)$  to achieve good agreement with LEP data. ( $\alpha_s(m_Z) = 0.137$  in Pythia's Monash 13 tune \*)

[arxiv:1404.5630](https://arxiv.org/abs/1404.5630), Skands, Carrazza, Rojo )

- Some caution needed as no 3-jet NLO matching, which is known to be relevant away from the 2-jet region.
- A comprehensive study of shower uncertainties is still to be done.

パartonシャワーにNNLL補正を  
 入れることで  
 パartonシャワーの $\alpha_s$ と  
 摂動計算の $\alpha_s$ が初めて一致

## 今後：素粒子における深層学習の応用

懐疑的に考えられていた、深層学習の利用はこの十年で市民権を獲得した。

素粒子実験のデータ生成やデータ解析は相当部分深層学習に置き換わると考えられる。

深層学習が物理をやってくれるわけではないが、物理と深層学習を正しく理解していれば、やくには立つ。素粒子理論→他粒子の相関の計算 は深層学習が使うような情報を正しく  
予言できるようにupdate されるべき。

これが新しい発見や深層学習の一般的な応用へのインパクトにつながるかはよくわからない。