

深層学習の科学研究へのインパクト (素粒子物理学を例として)

野尻美保子

「学習物理学」

MLPhYs Foundation of "Machine Learning Physics"
Grant-in-Aid for Transformative Research Areas (A)

CONTACT

Members only



En

Jp

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



1. 高エネルギー物理学と深層学習

2. ジェット分類を例にした深層学習応用の例
(自分でやってること)

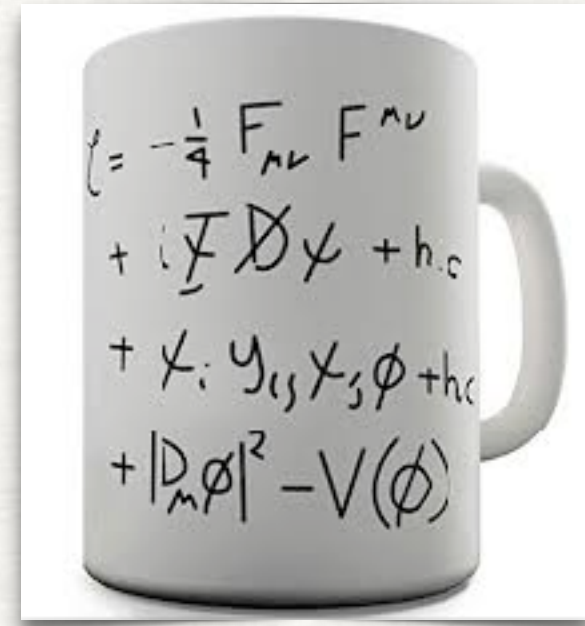
3. これからの理論の課題

1. 高エネルギー物理の目指すもの と深層学習

深層学習と素粒子物理

- 実用的な深層学習利用
 - 画像生成(Diffusion model)・翻訳(Transformer)
 - それっぽいものをAIが作り、質保証は人間がやる。
- 素粒子物理学(精密科学)における深層学習→質補償が絶対的に重要（最初のころは、深層学習になんでもやらせればいいのかいっていた）
- バイアス、誤差推定、推定プロセス、再現性

素粒子理論と実験



素粒子の標準モデル

+

標準模型以外の粒子
や相互作用

"New Physics"

複数の精密データ

理論予言
(任意のスケールでの
素粒子の相互作用)

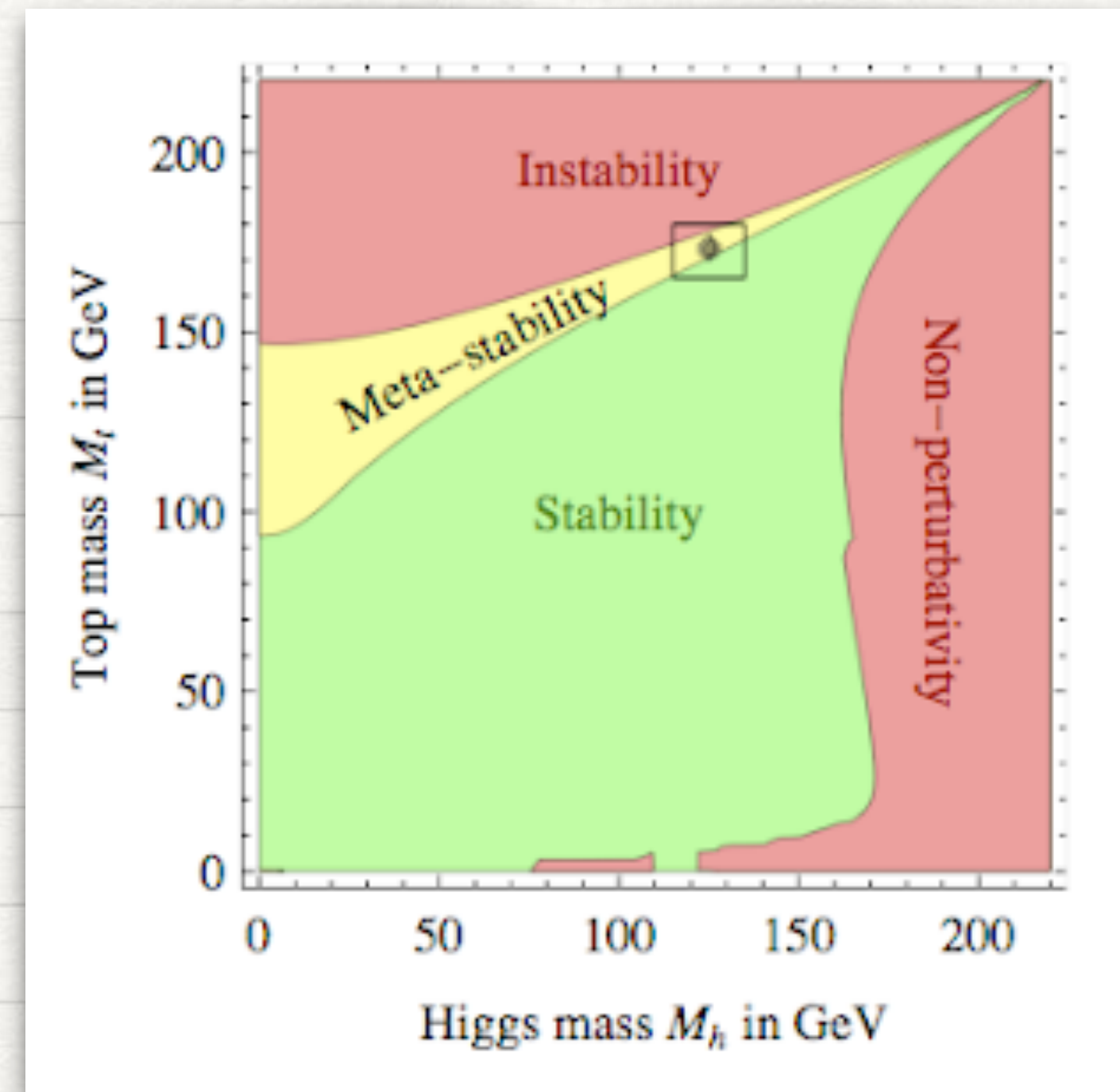
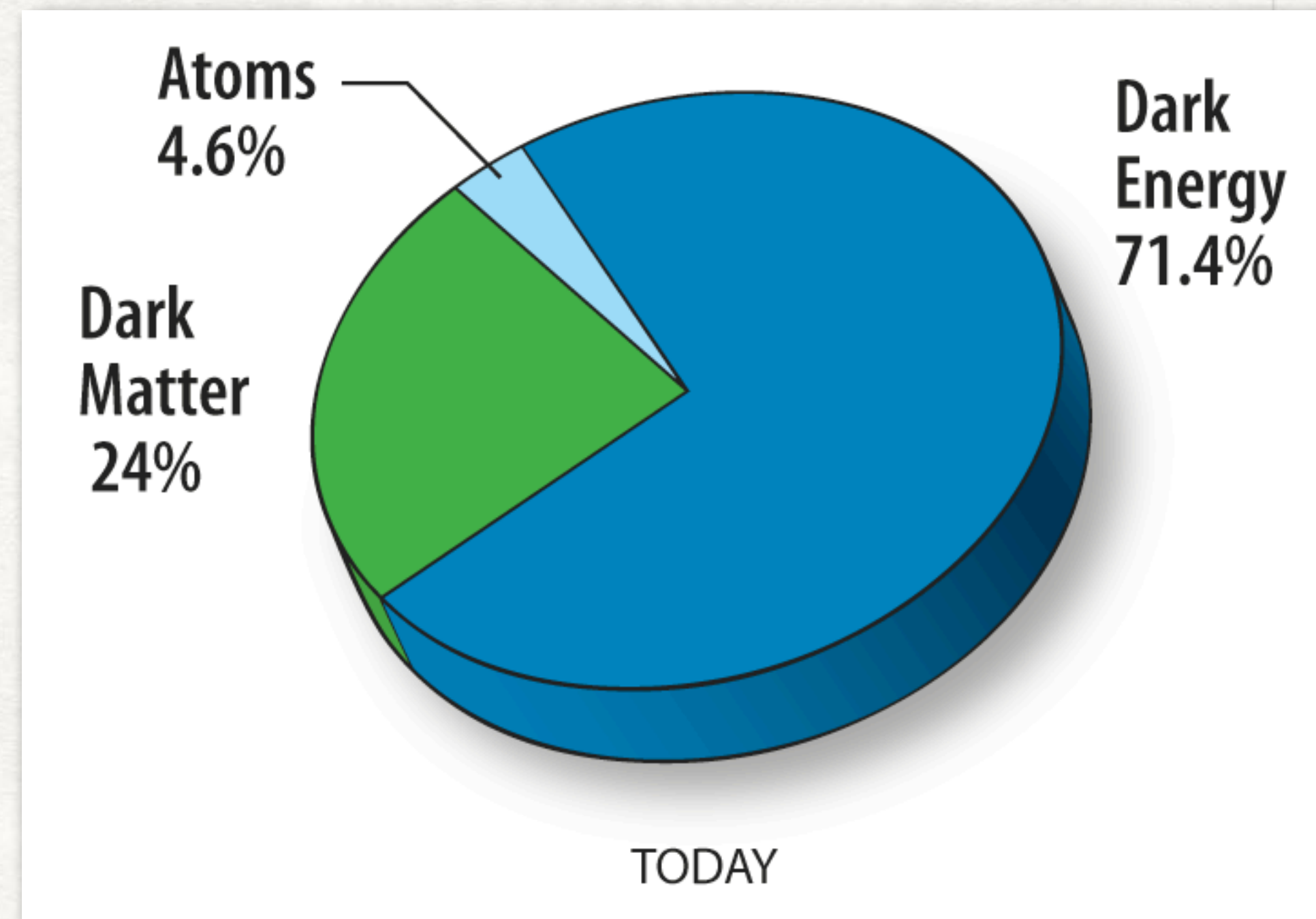


場の理論

Other data

現象に対する予言

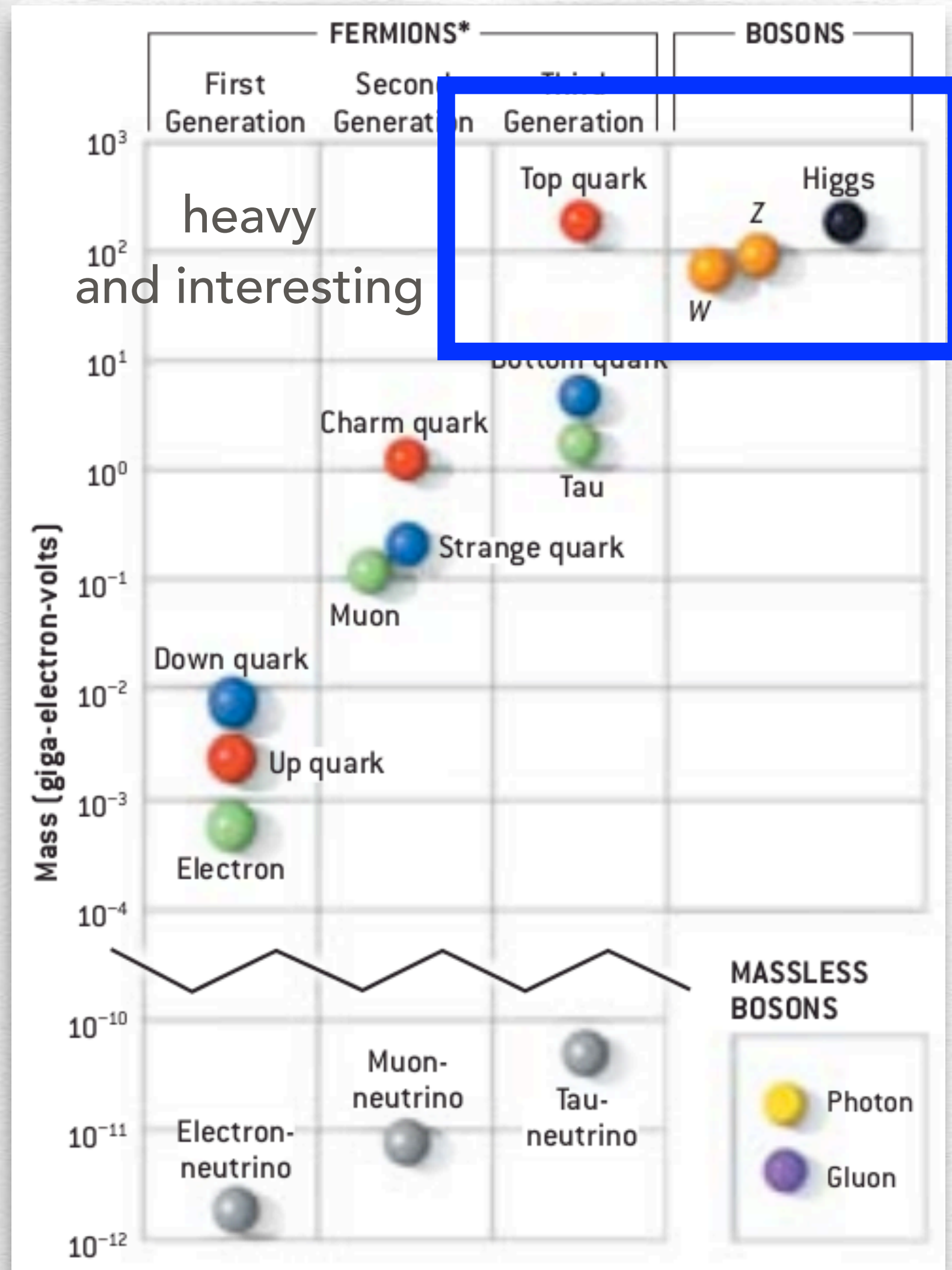
- 宇宙の歴史
- Higgs 場の真空の不安定性
- 暗黒物質の存在 → dark sector
- バリオン数、レプトン数...



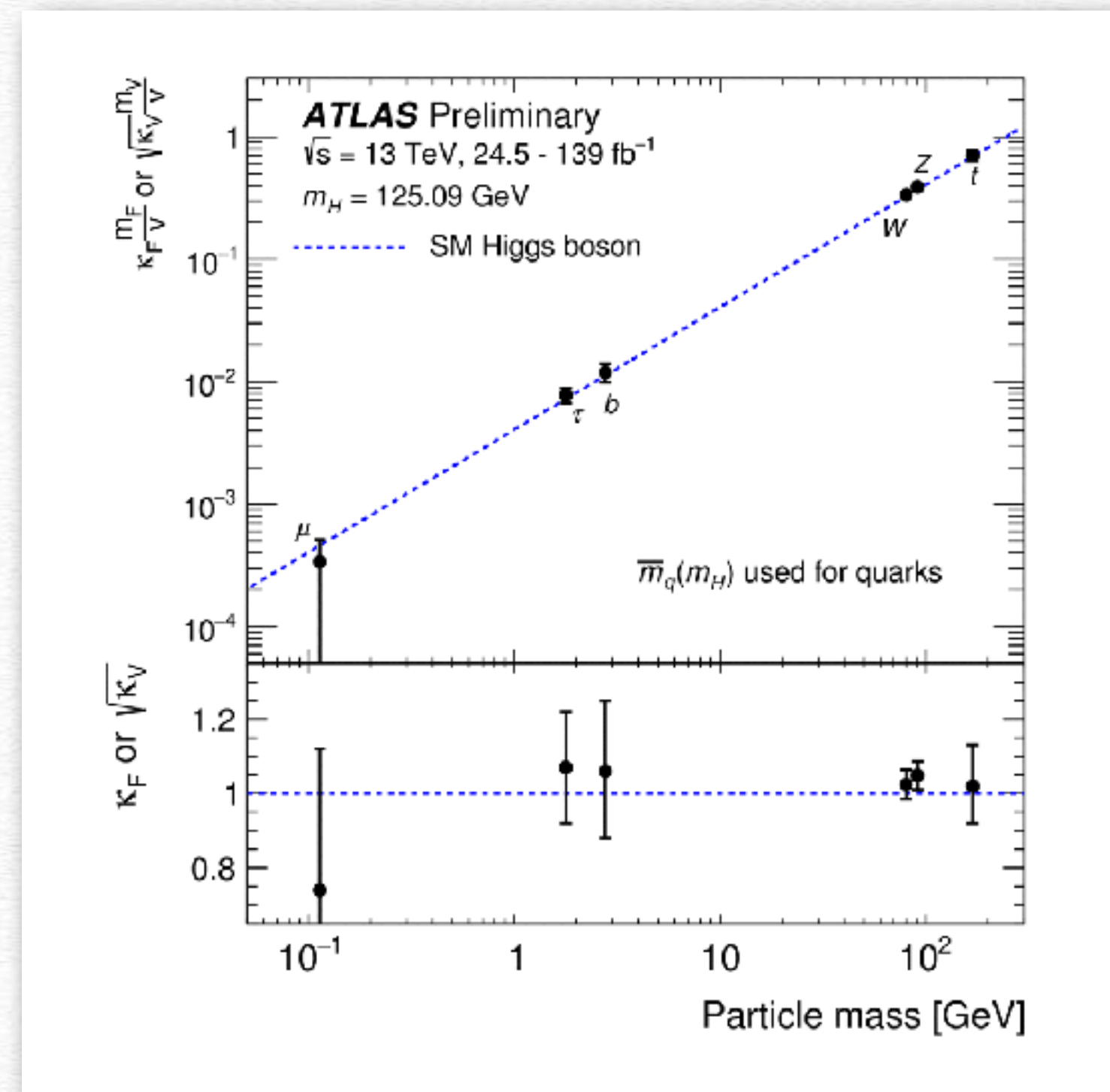
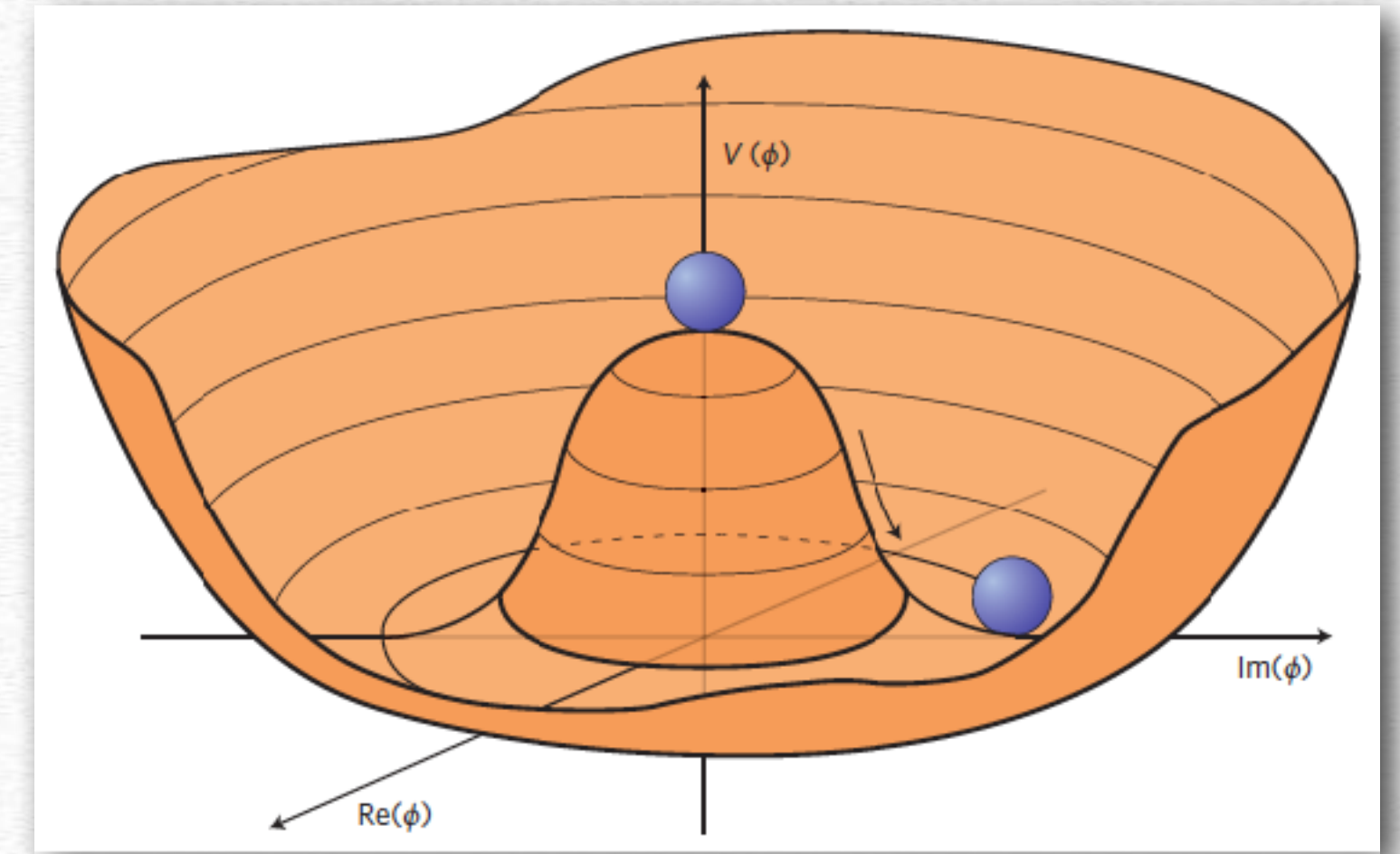
Axion, Supersymmetry... ,



ヒッグス粒子の発見(2012)



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

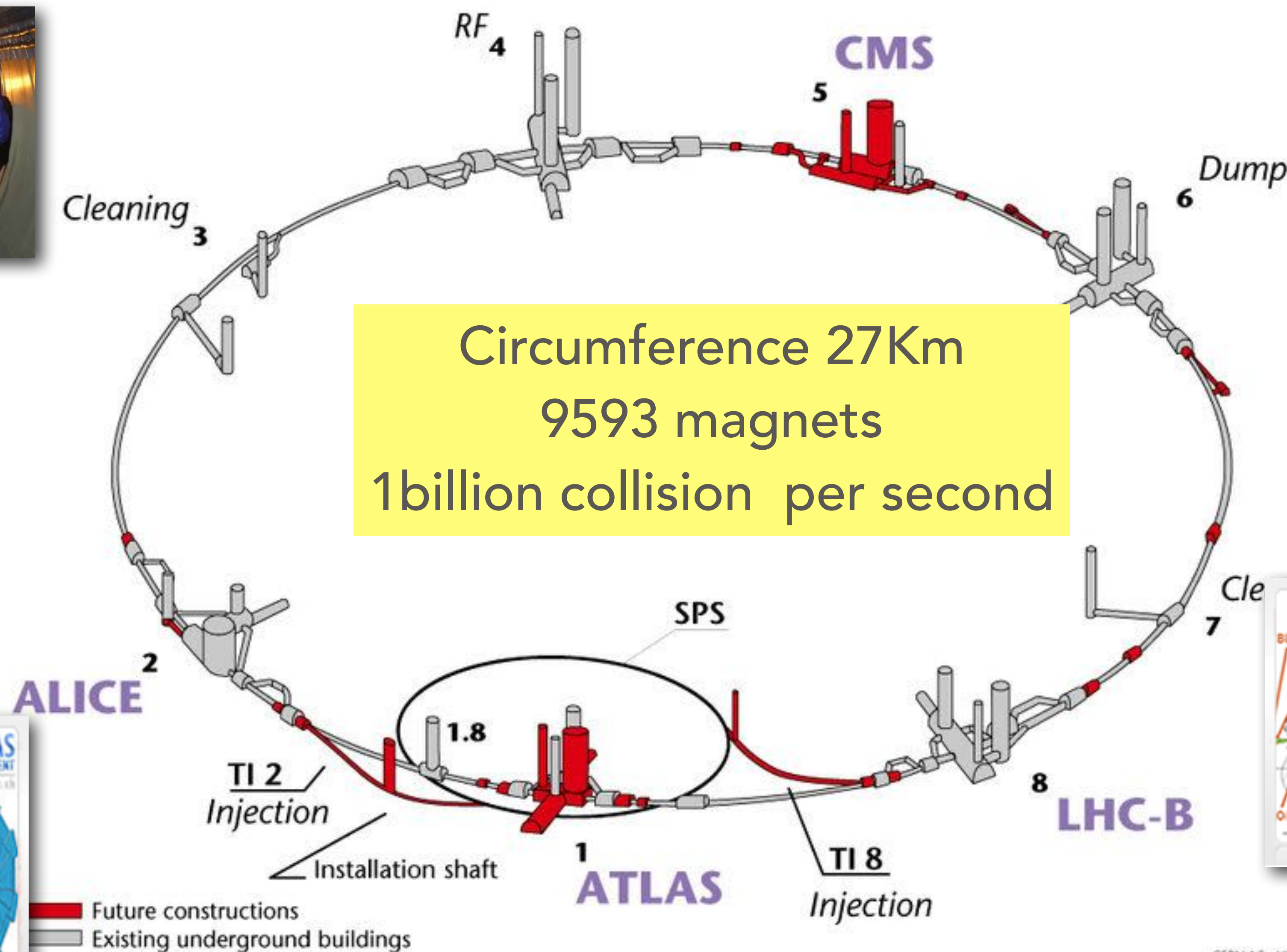
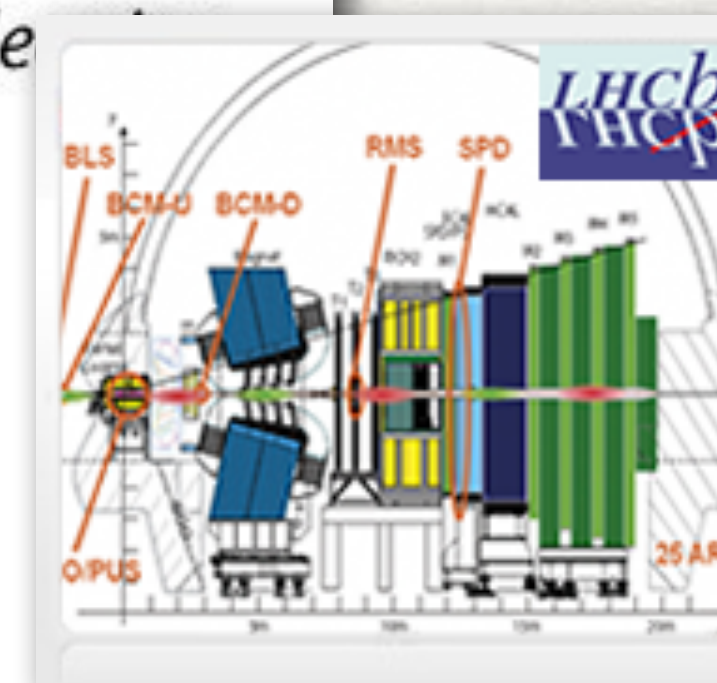
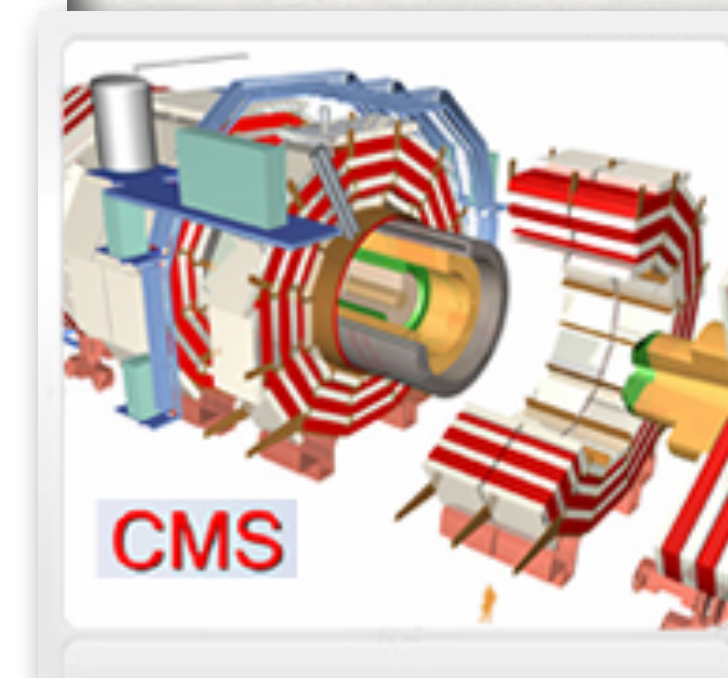
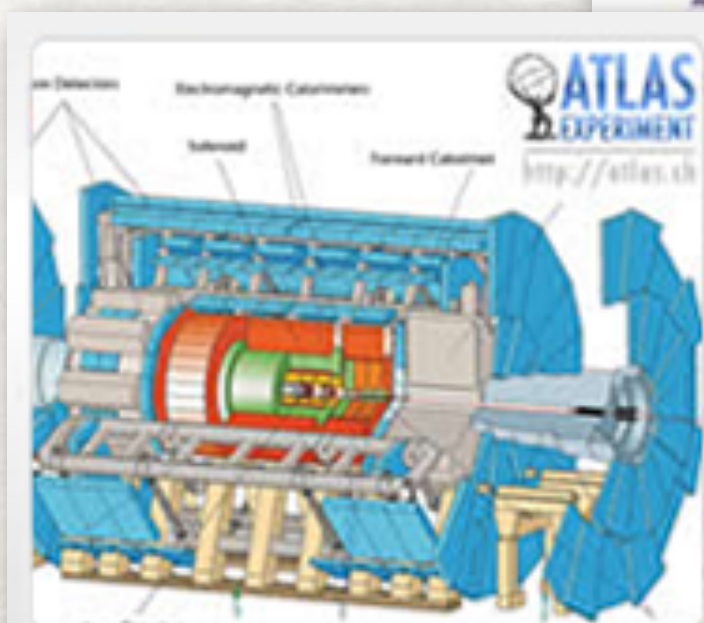
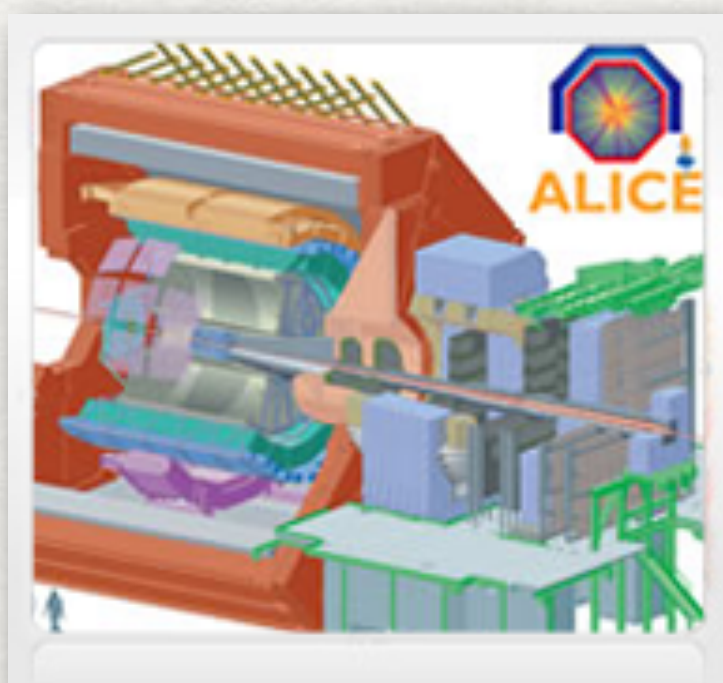


Higgs sector の何らかの拡張を
 実験がどの程度制限できるか

SMEFT (SM の重要な対称性を
 維持した高次元項)

新粒子探索

Layout of the LEP tunnel including future LHC infrastructures.



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

ヨーロッパ戦略で議論

FCC -ee → FCC-hh or Linear Collider)

暗黒物質

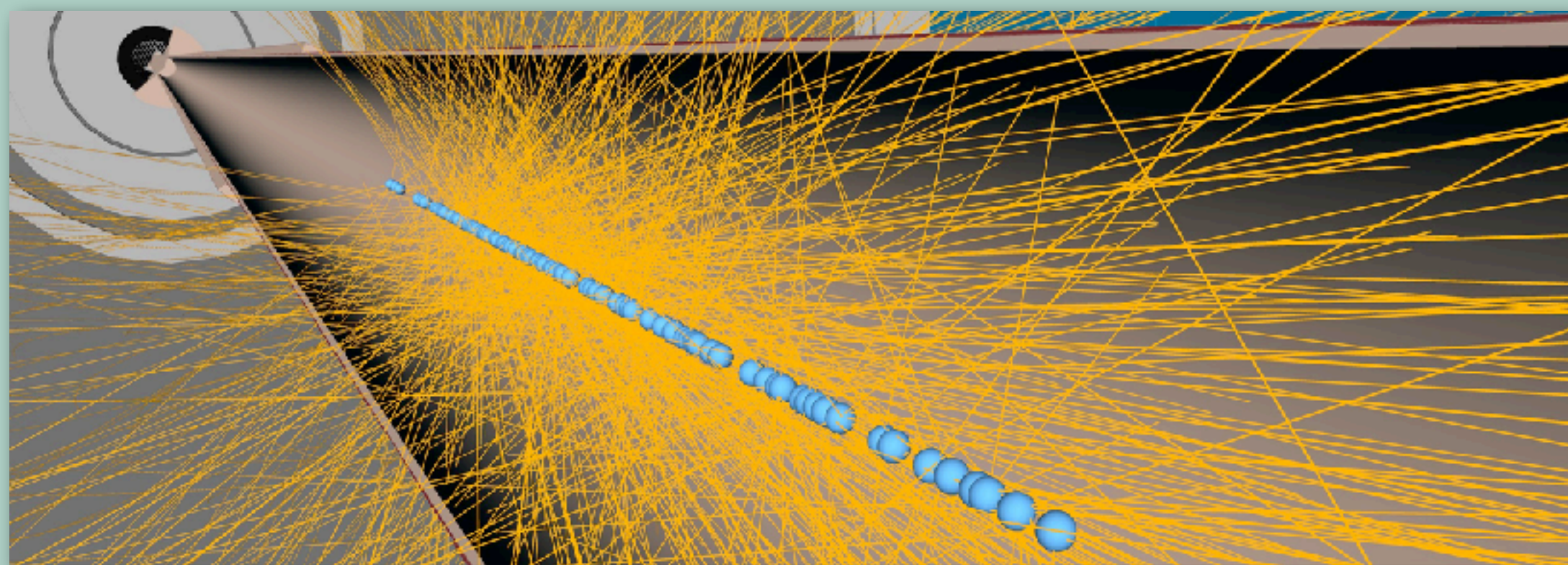
直接探索, 間接探索

フレーバー関連

Belle-II 実験 LHC b , muon g-2

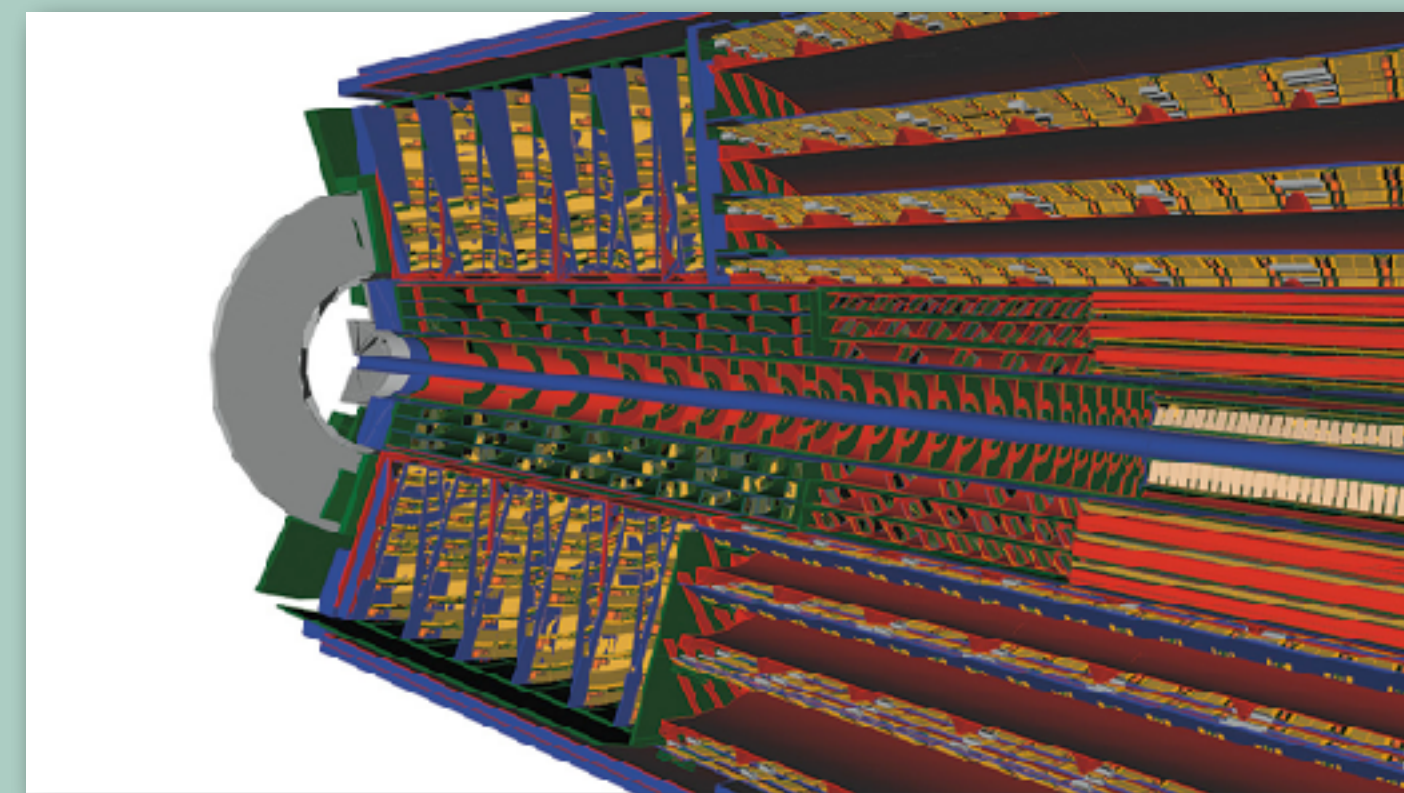
ニュートリノ振動実験

一回のバンチ衝突で 200個の衝突

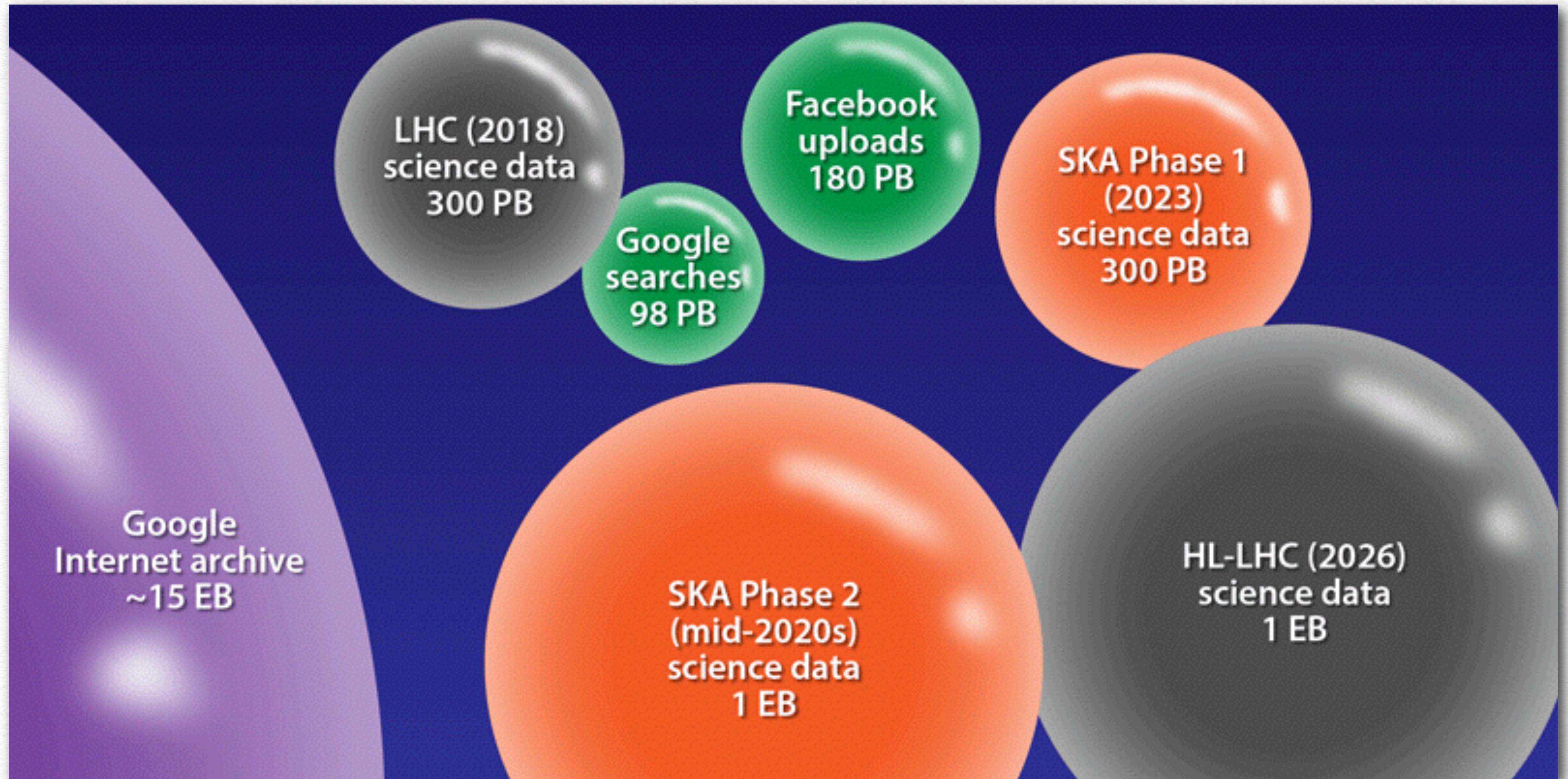


VS

up grade した測定器で解析
(日本も相当量を制作)

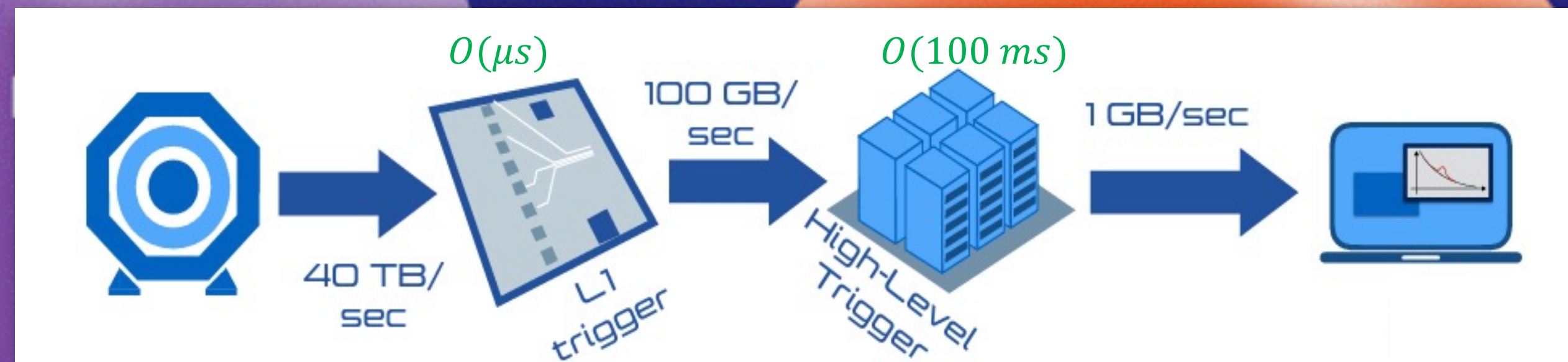


HL-LHC はデータサイエンス上のチャレンジ



HL-LHC はデータサイエンス上のチャレンジ

全宇宙でもっともデータレートが高い
AI (深層学習) でこれをなんとかできるか



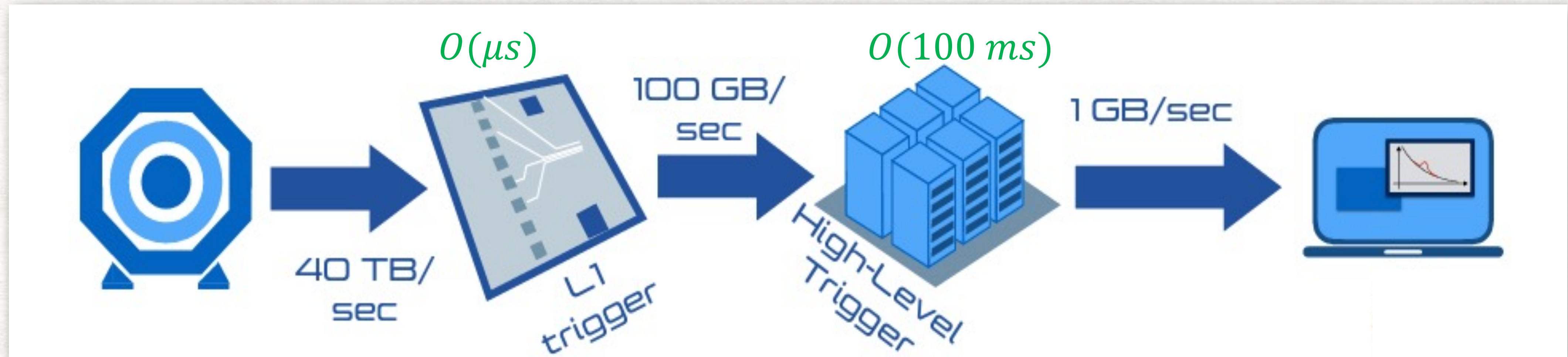
LHC (2018)
science data

Facebook
uploads
180 PB

SKA Phase 1
(2022)
data
B

HL-LHC (2026)
science data
1 EB

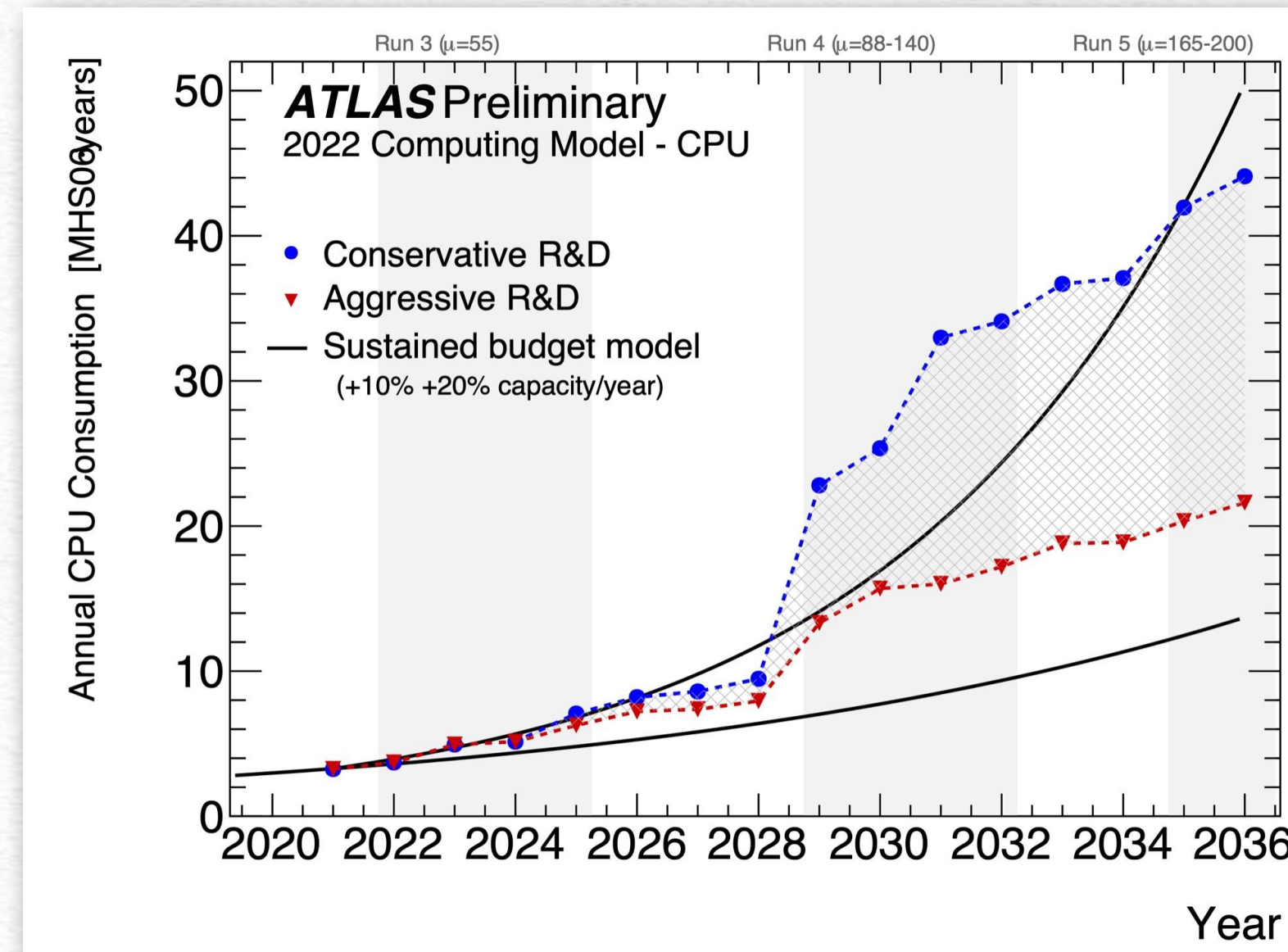
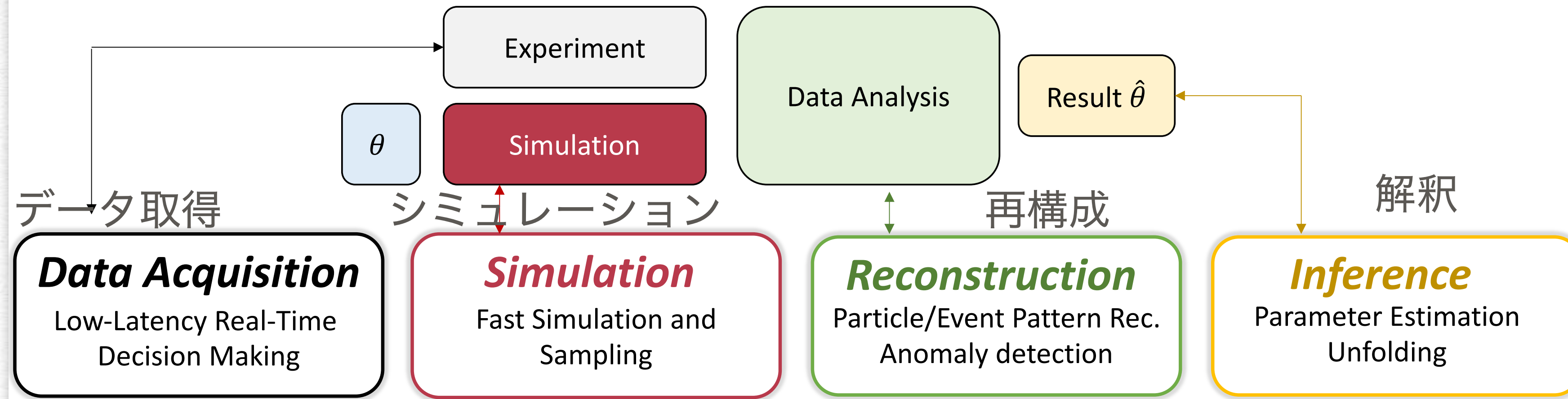
HL-LHC実験の計算コスト



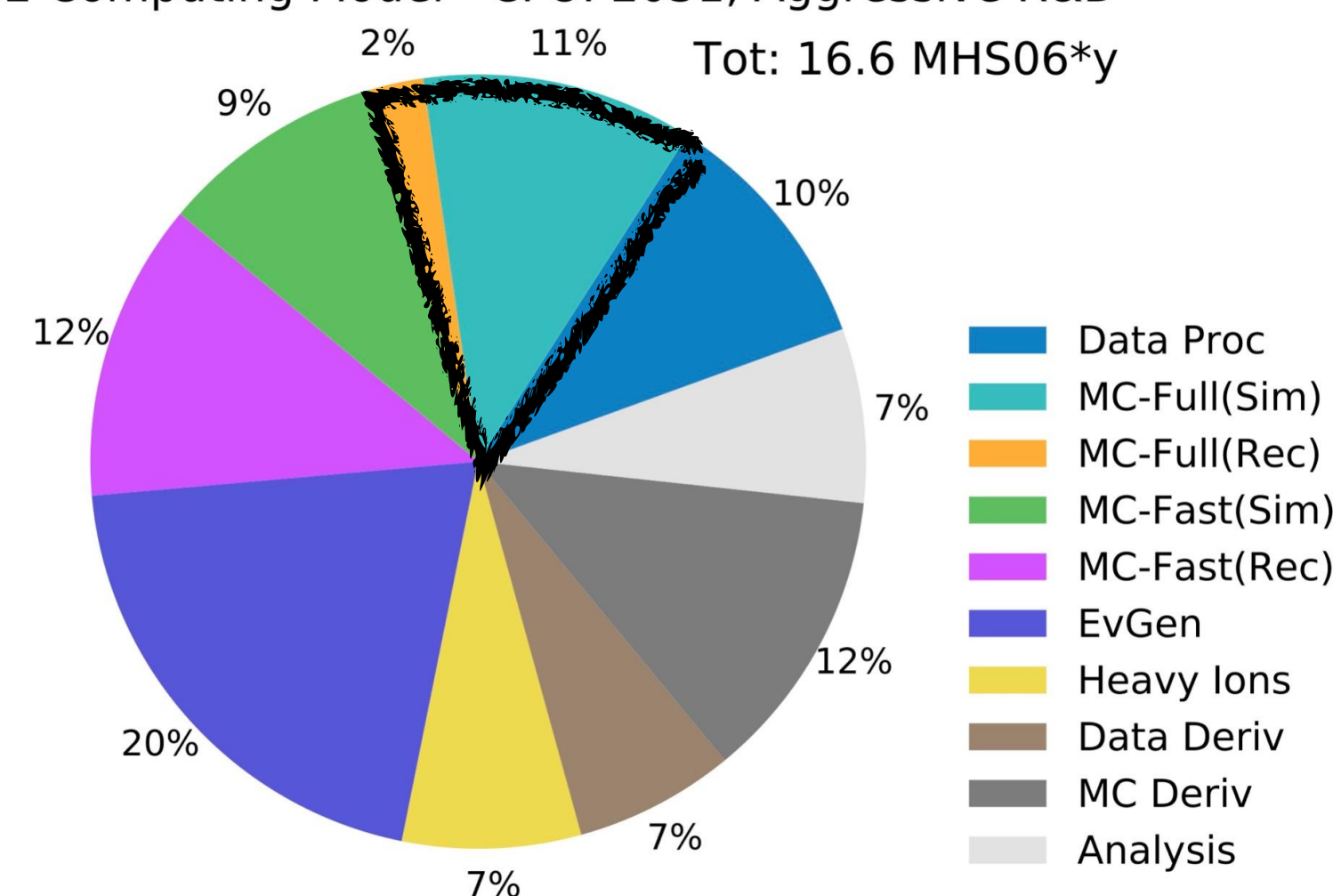
- 40TB/sec (factor 100 increase)
- 10倍くらい計算を高速化することを想定している。（測定器のシグナルを元の粒子と同一化する、重要な事象を認識して保存するなど。）
- 計算量を確保するための予算が多額にならないようにしたいが、単純に考えると足りていない。今の工程表は、研究開発によって、計算予算が削減できることを過程している。

Machine Learning Across Data Analysis

4



4



Kagan "ML at HEP"

データ取得にMLを直接載せることができる集積回路(FPGA)

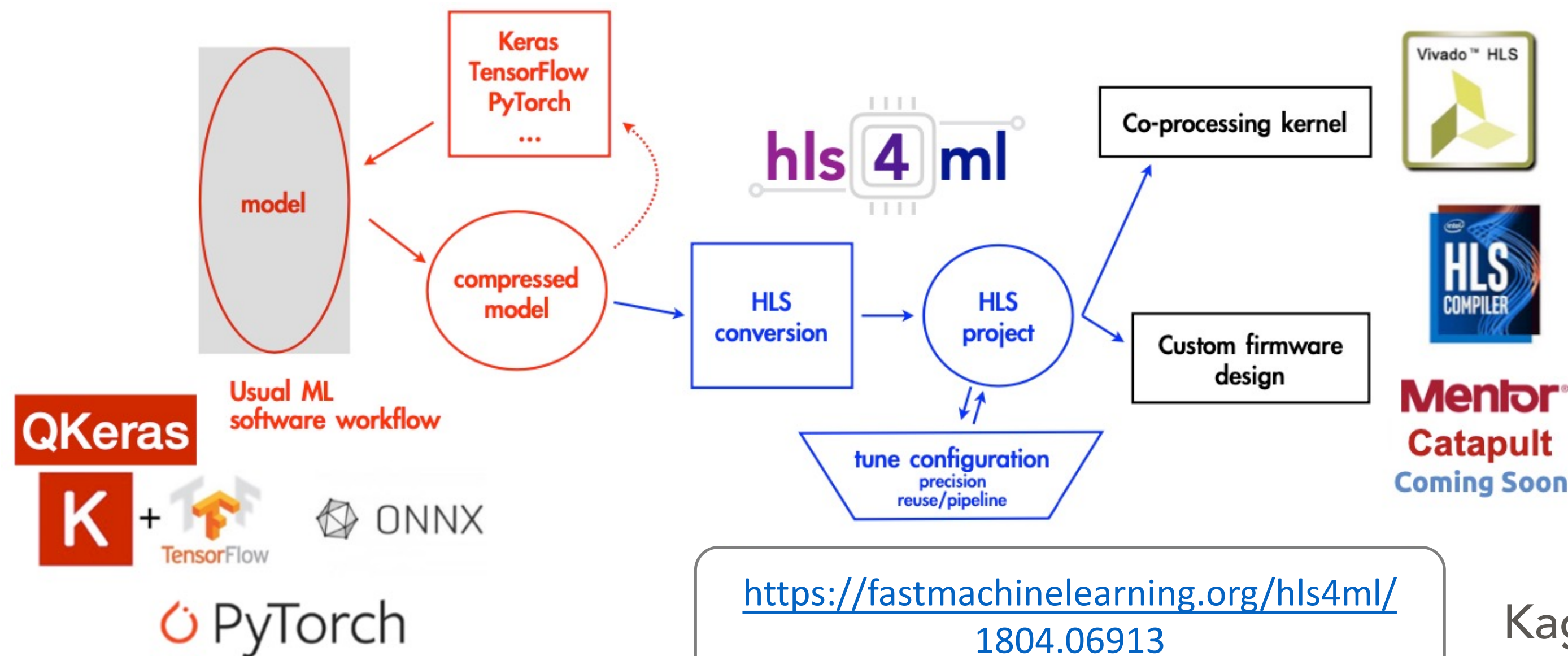
Real-Time Low Latency ML

Goal: Move improvements from ML for reconstruction to on

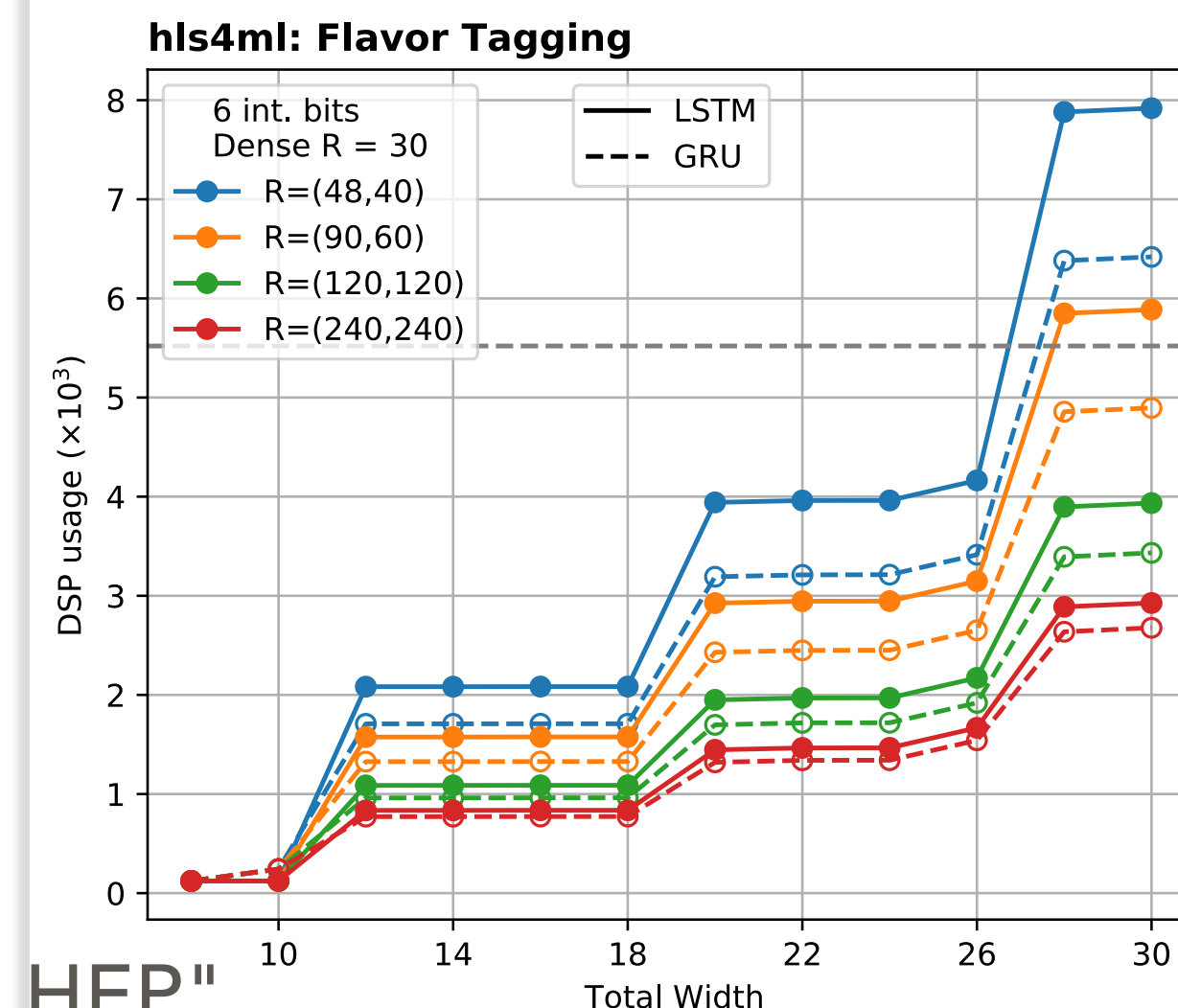
FPGA= Field Programmable Gate Array
HLS =its compiler

hls4ml:

- HEP Community developed tool do deploy NN as electronic circuit on hardware
- FPGA-as-accelerator library → Whole model on chip
- Model Compression essential



Example: Resource Usage in Jet Classification with RNNs



Kagan "ML at HEP"

Khoda, ..., MK, et. al, [2207.00559](#)

深層学習を使ったイベント生成

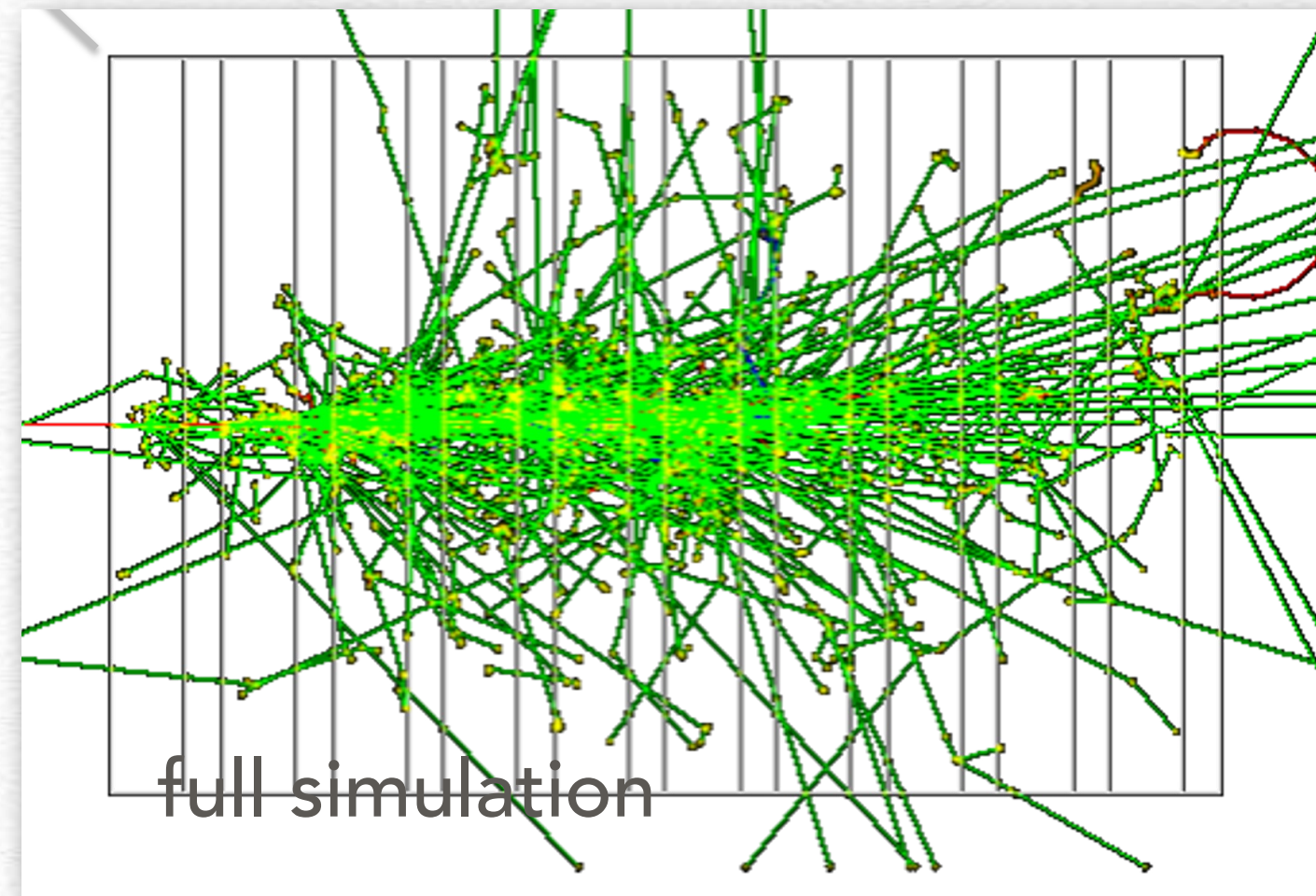
カロリメータ

イベント一つ作るのに 3分

Bang!



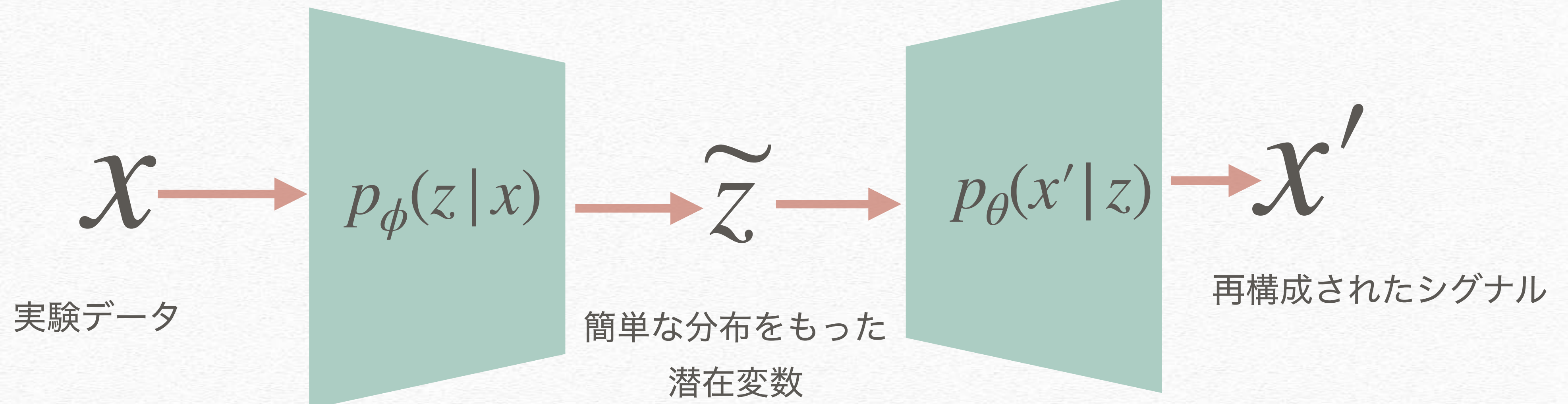
入射粒子 $\{x\}$



それぞれの測定器
での信号

$$P(\{E\} | \{x\})$$

Variational AutoEncoder(VAE): bottle
neck structure $x \rightarrow z \rightarrow x'$



深層学習を使ったイベント生成

カロリメータ

イベント一つ作るのに 3分

Bang!



入射粒子 $\{x\}$

ML model

それぞれの測定器
での信号

$$P(\{E\} | \{x\})$$

Variational AutoEncoder(VAE): bottle
neck structure $x \rightarrow z \rightarrow x'$

x

実験データ

$$p_{\phi}(z | x)$$

$\tilde{z}$

簡単な分布をもった
潜在変数

$$p_{\theta}(x' | z)$$

x'

再構成されたシグナル

深層学習を使ったイベント生成

Calorimeter

Bang!



入射粒子 $\{x\}$

ML models



それぞれの測定器
での信号

$P(\{E\} | \{x\})$

ML toolbox

GAN:

competition between
generator and discriminator

Normalizing Flows:

map simple distribution to
nontrivial function

Diffusion : Add noise step by step

もしこのマップが完全なら、だれでも測定器シミュレーションができる。
(たとえば理論屋さんでも)

深層学習モデルは微分可能なので、パラメータサーチも可能ML

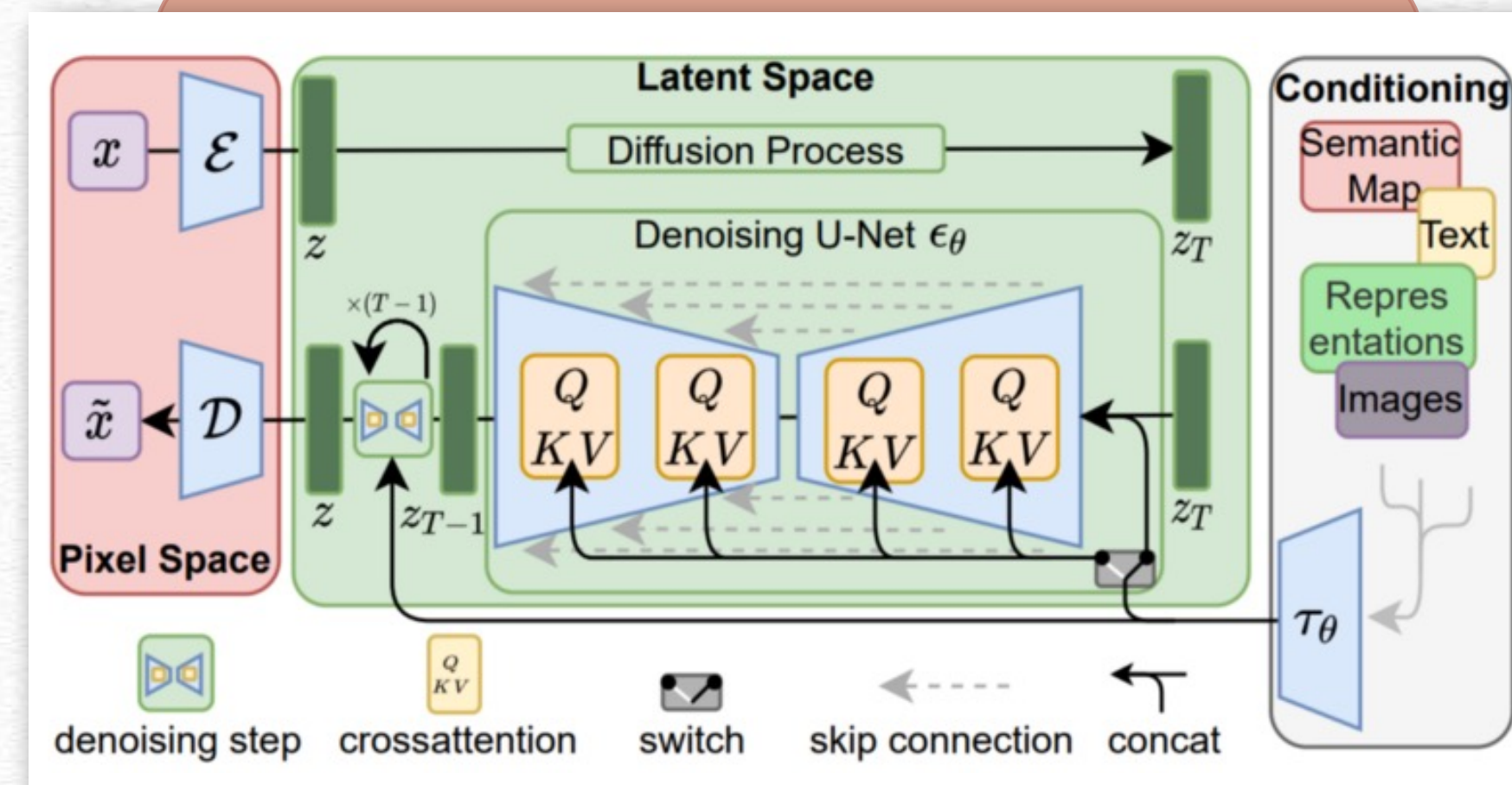
深層学習を使ったイベント生成

Calorimeter

Bang!



入射粒子 $\{x\}$



それぞれの測定器
での信号

$$P(\{E\} | \{x\})$$

ML toolbox

GAN:

competition between
generator and discriminator

Normalizing Flows:

map simple distribution to
nontrivial function

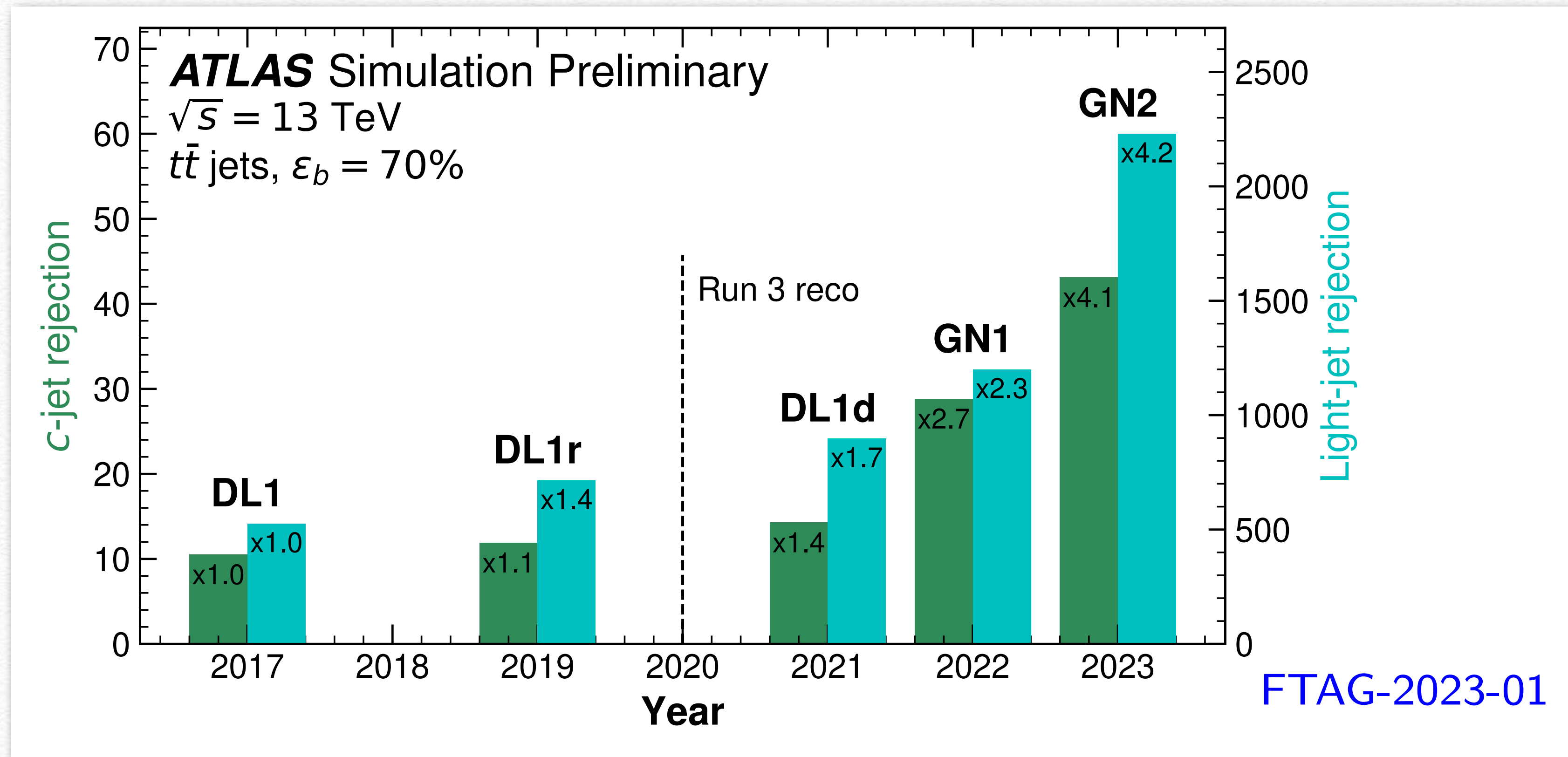
Diffusion : Add noise step by step

もしこのマップが完全なら、だれでも測定器シミュレーションができる。
(たとえば理論屋さんでも)

深層学習モデルは微分可能なので、パラメータサーチも可能ML

深層学習による粒子再構成

ボトムクォークがどのくらいクリーンに捉えることができるようになるか



bottom quark からくるジェットを70% 以上捉えた時に、

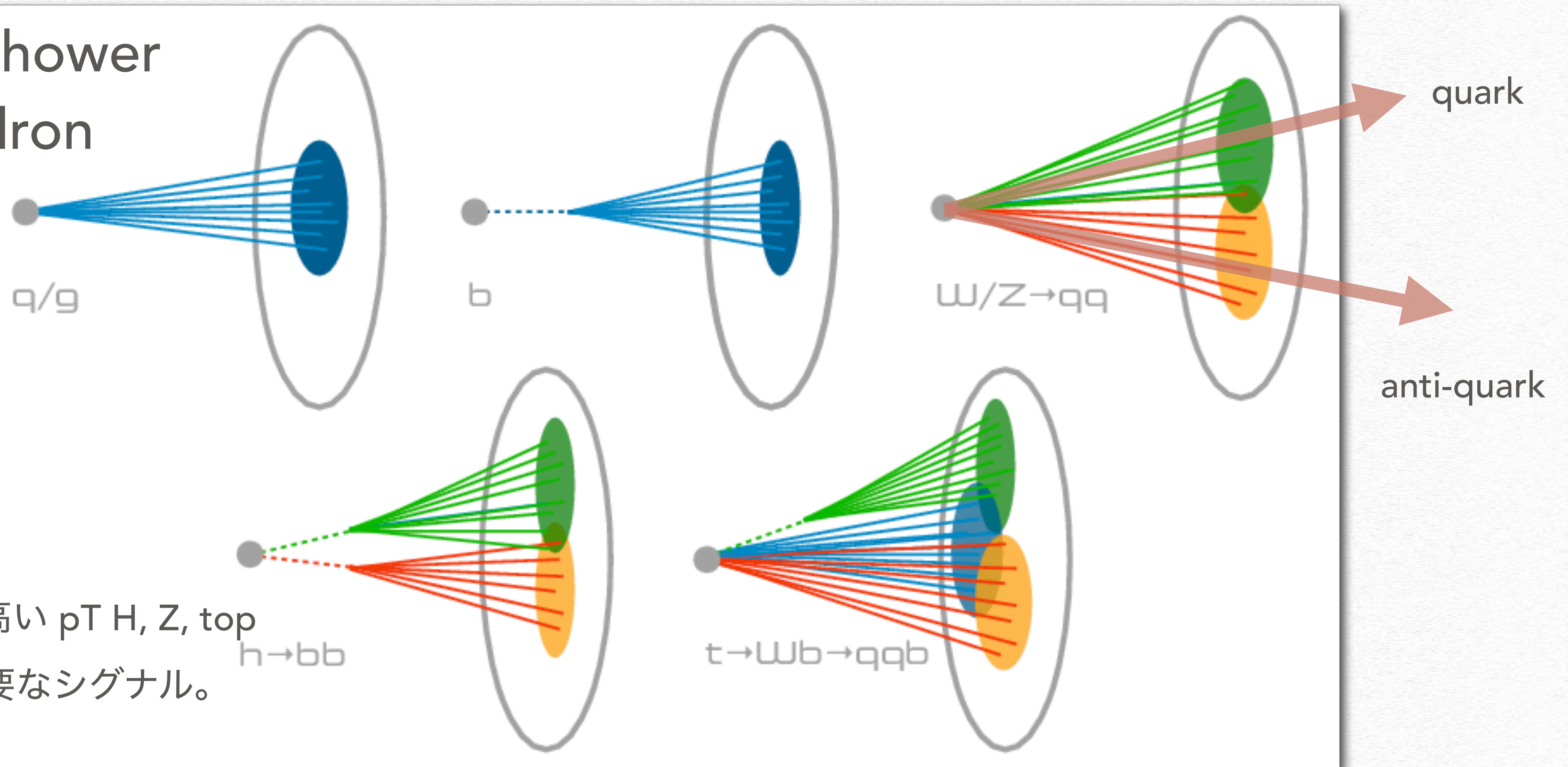
light quark c quark からからくるジェットをどの程度排除できるか

$t \rightarrow b W$ 崩壊や $H \rightarrow b\bar{b}$ 崩壊を捉えるのとても重要

2. ジェット分類を例にして 深層学習モデルと物理

ジェット分類と深層学習

Parton shower
→ hadron



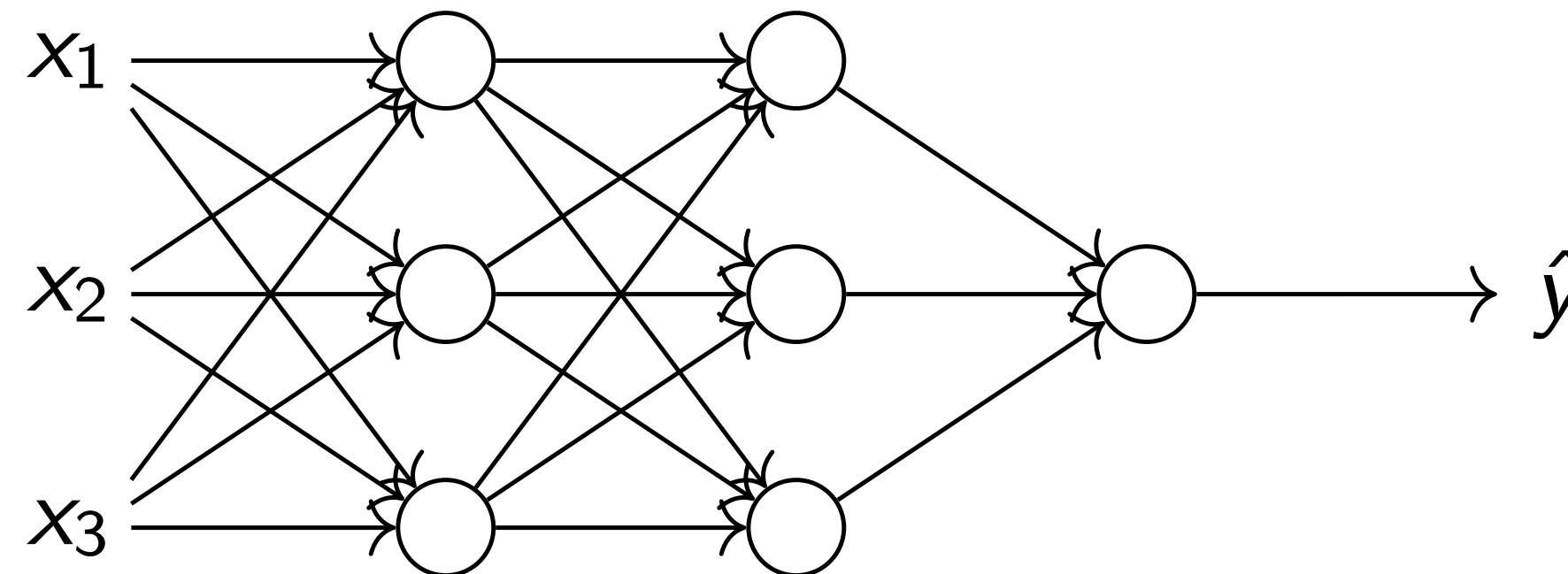
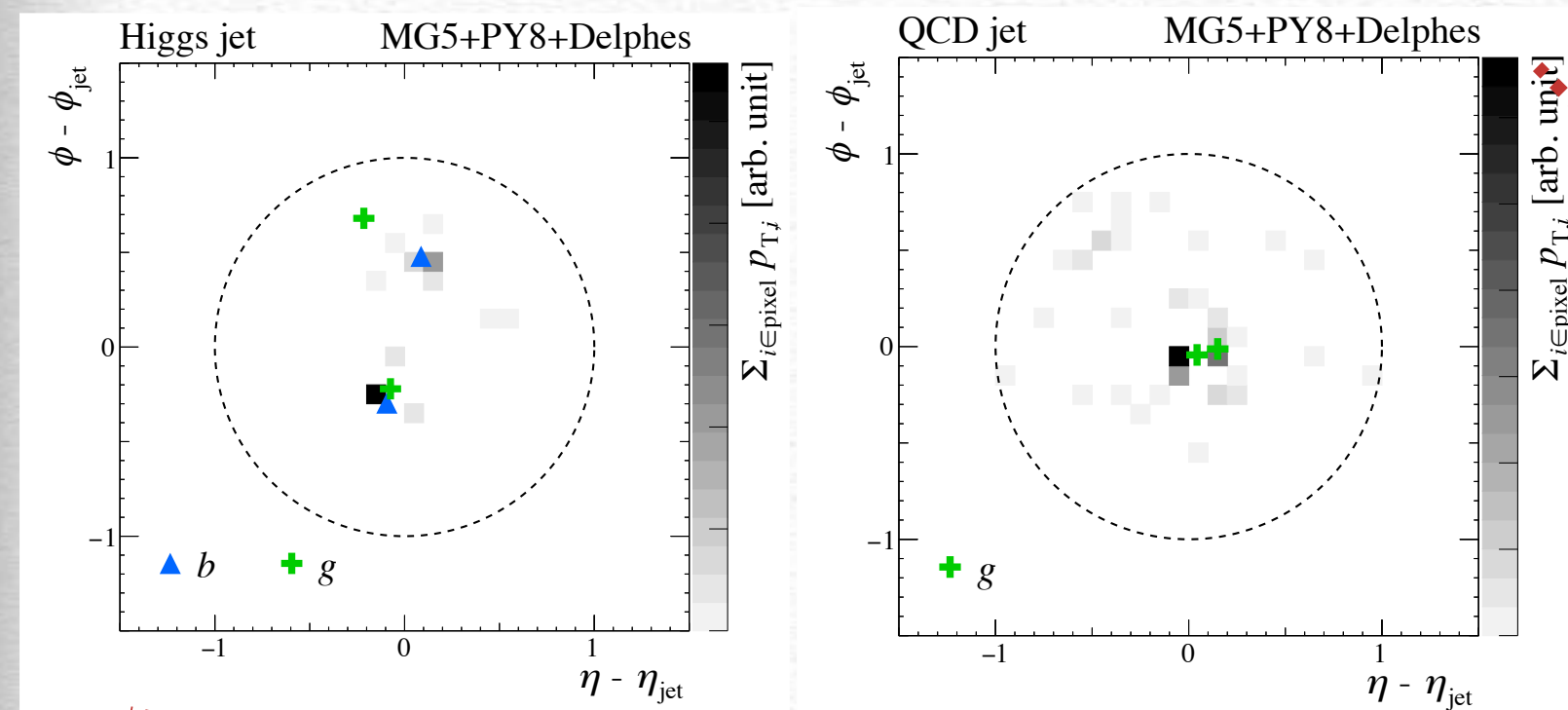
エネルギーの高い p_T H , Z , top
はBSM の重要なシグナル。

MLP (multiple layer perceptron) の基本構造

Input: Jet images

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

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

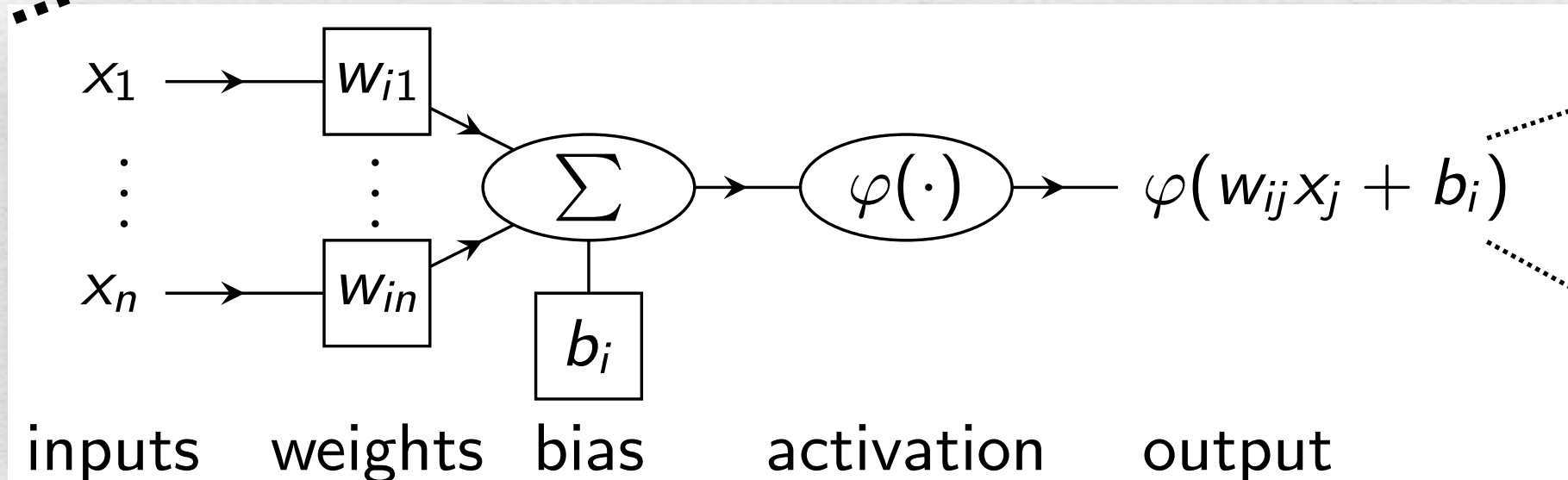
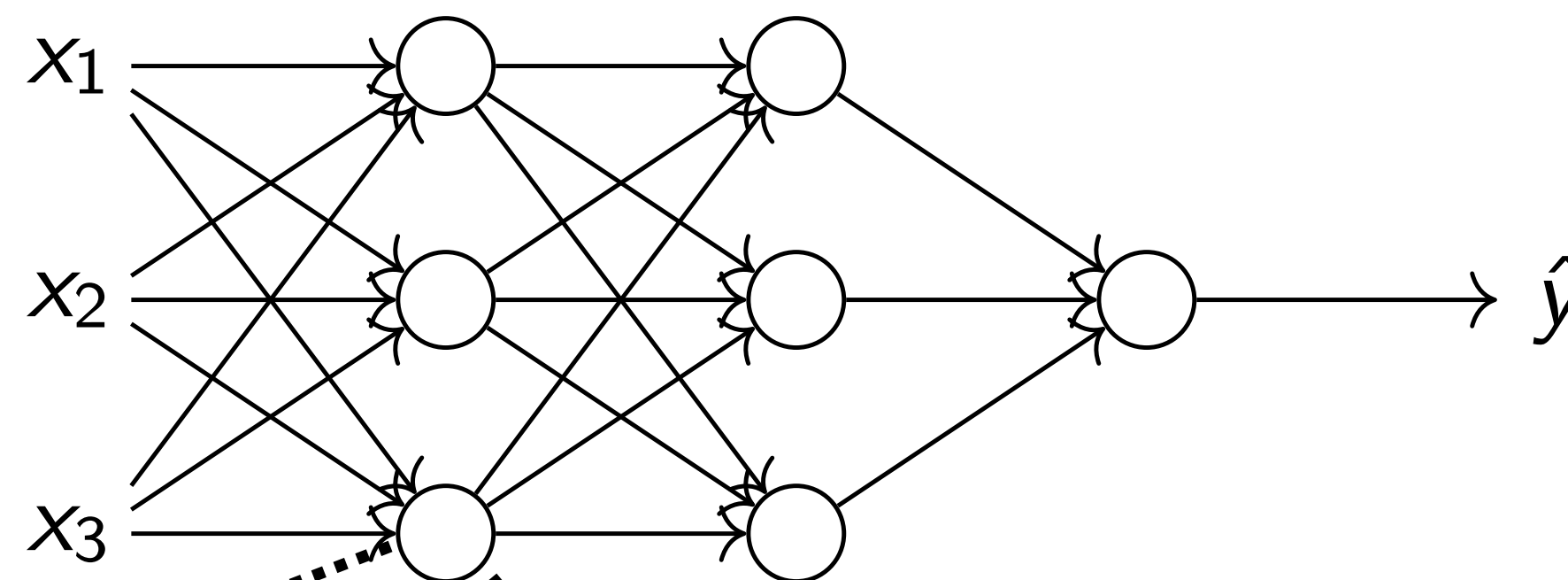
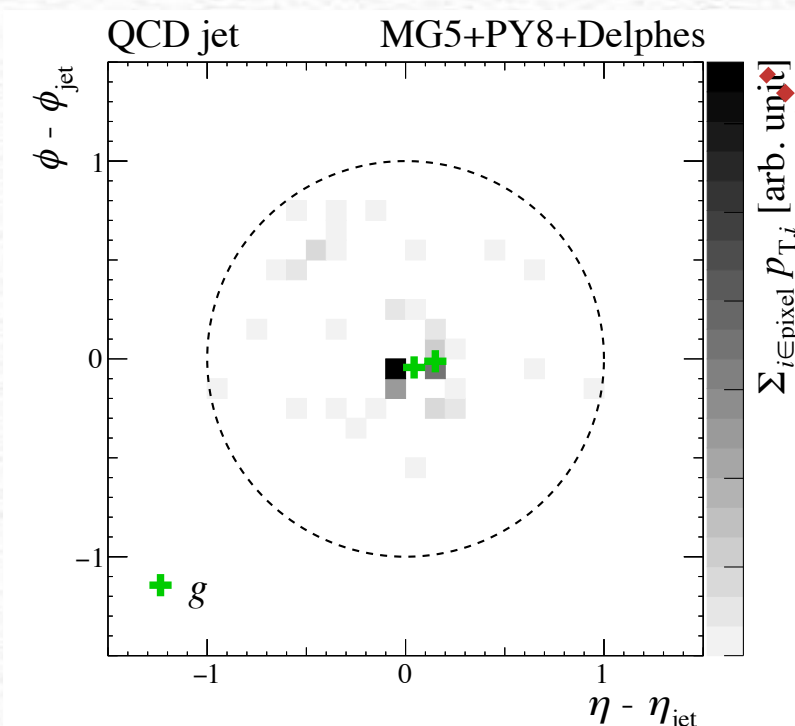
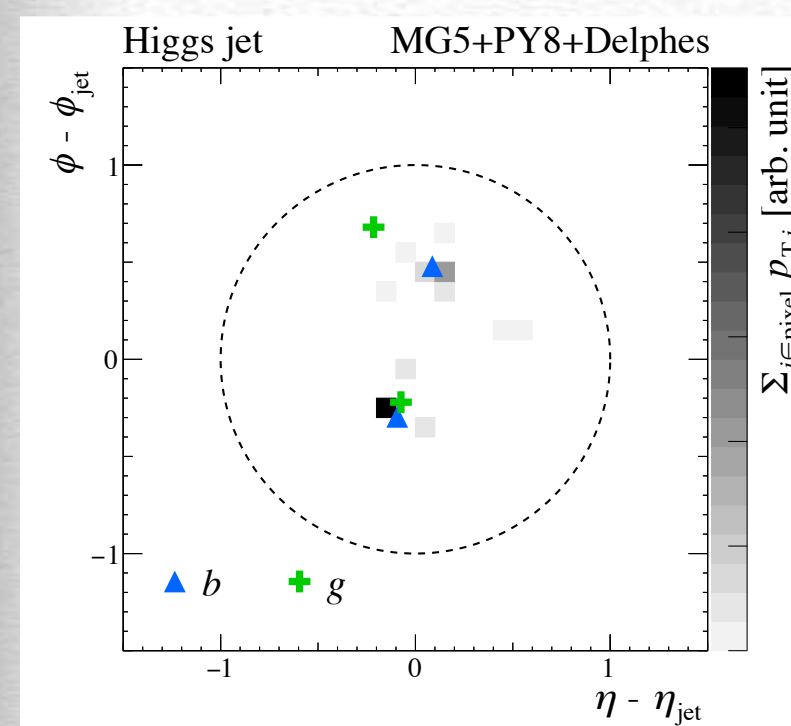


MLP (multiple layer perceptron) の基本構造

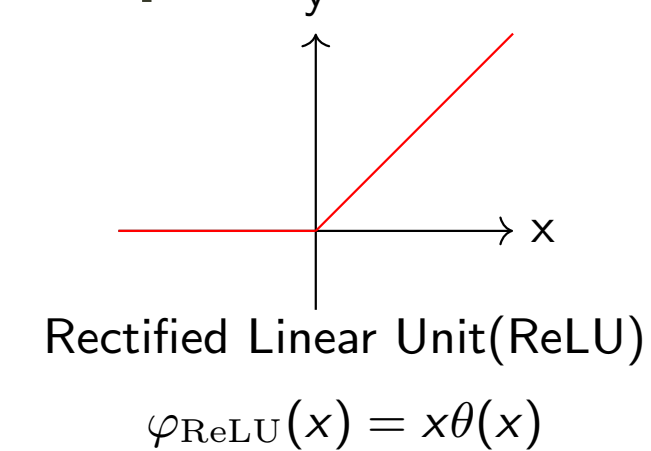
Input: Jet images

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

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



ϕ : 活性化関数 (非線形性)



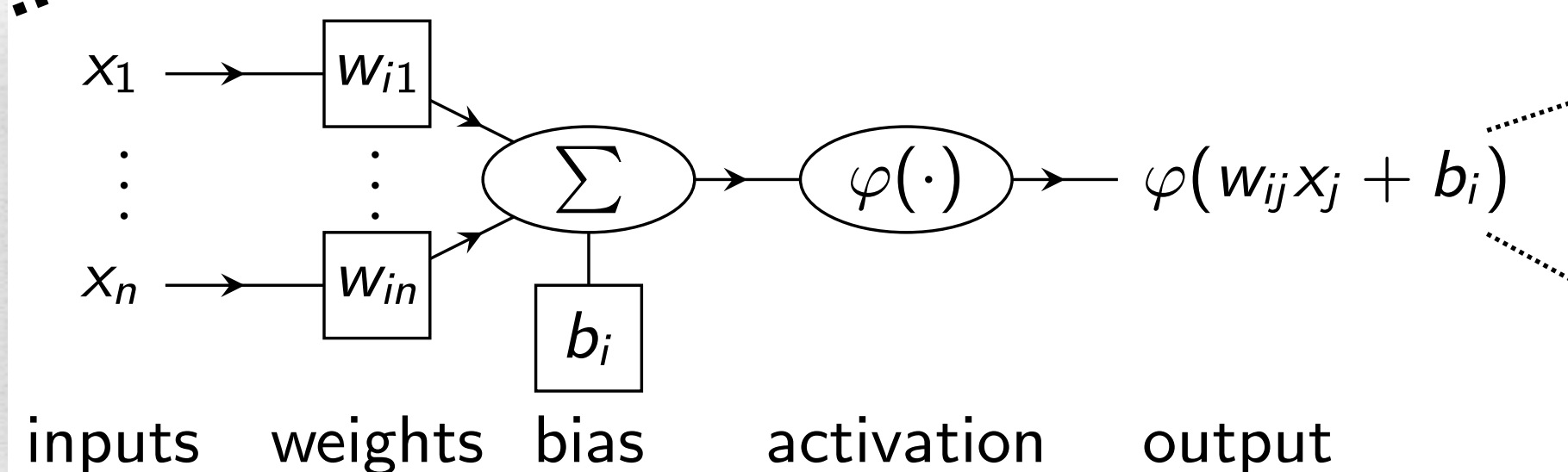
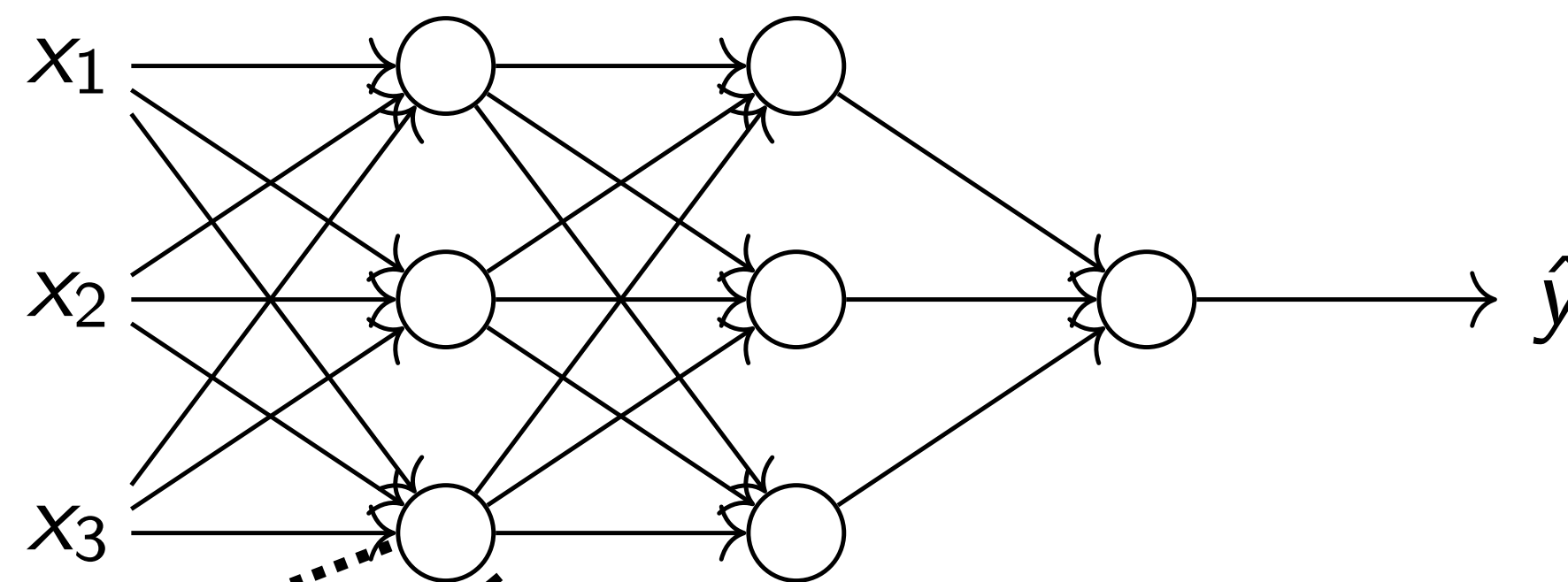
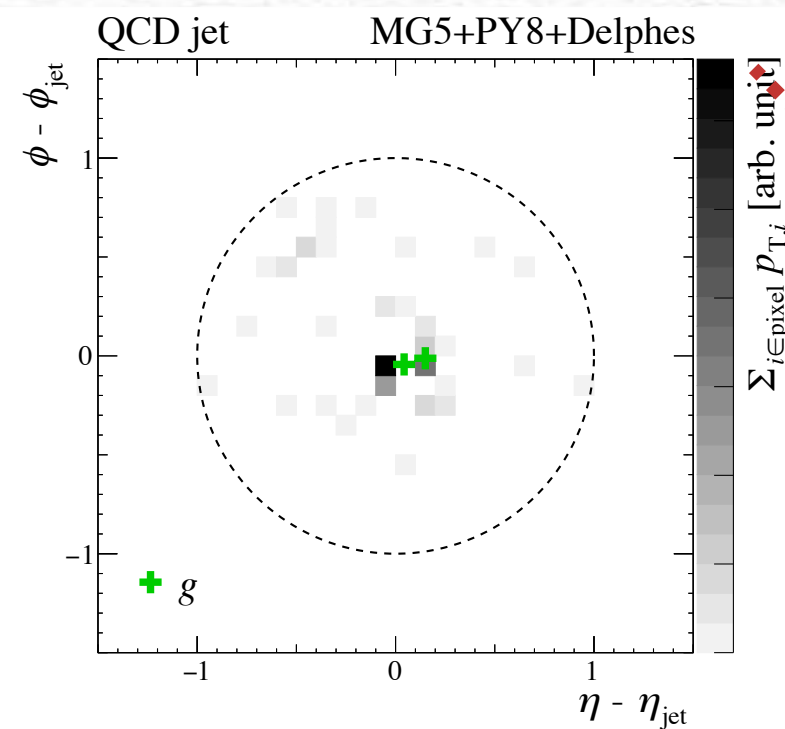
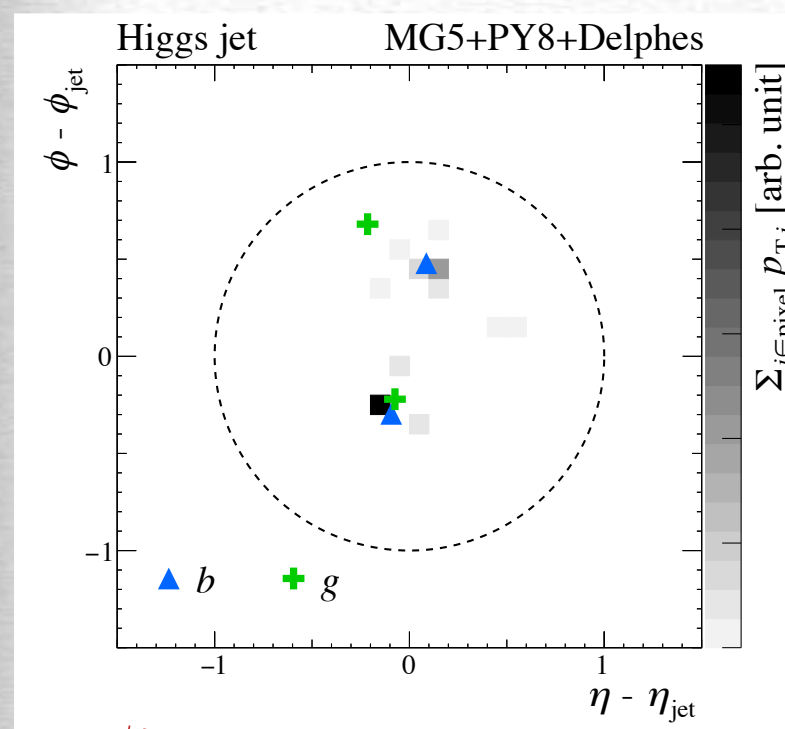
MLP (multiple layer perceptron) の基本構造

Input: Jet images

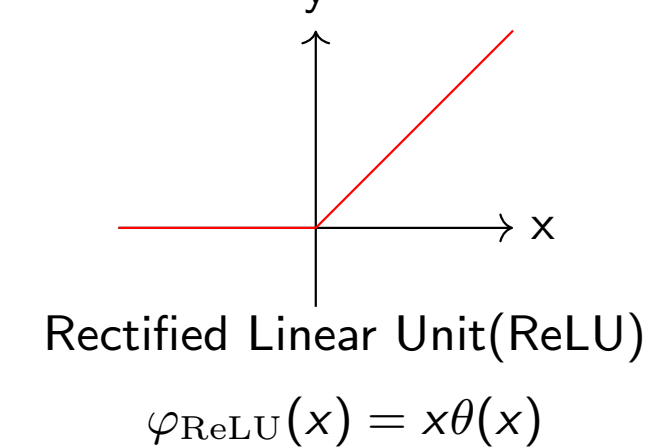
Higgs $h+Z(\rightarrow \nu\nu)$

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

訓練による潜在変数の決定: $\mathbf{w}_{ij}, \mathbf{b}_i$



ϕ : 活性化関数 (非線形性)

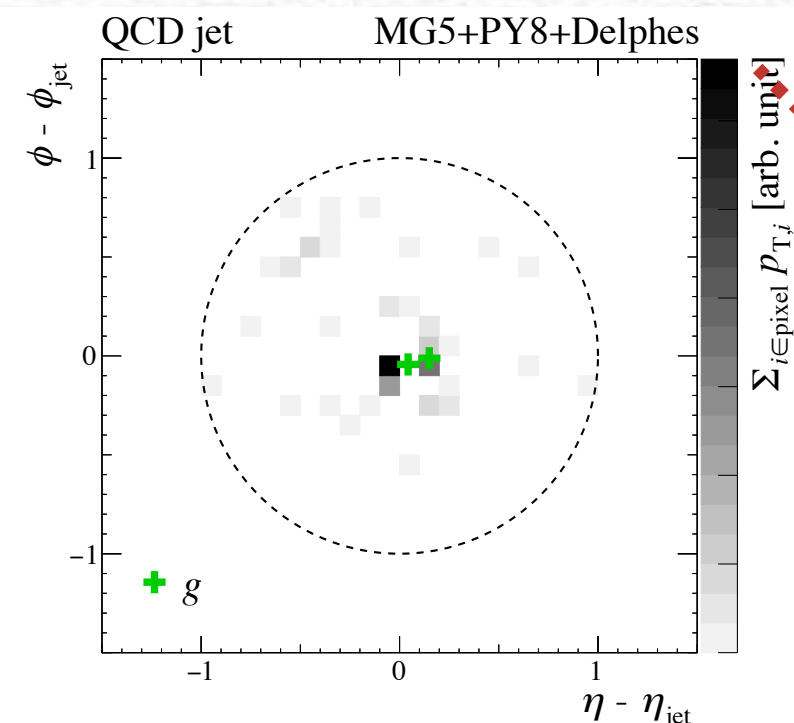
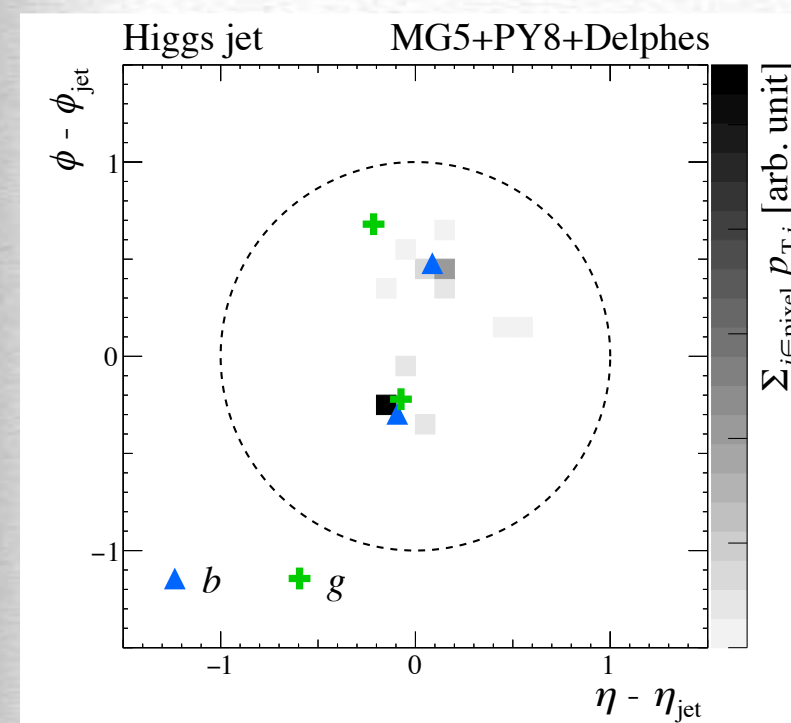


MLP (multiple layer perceptron) の基本構造

Input: Jet images

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

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



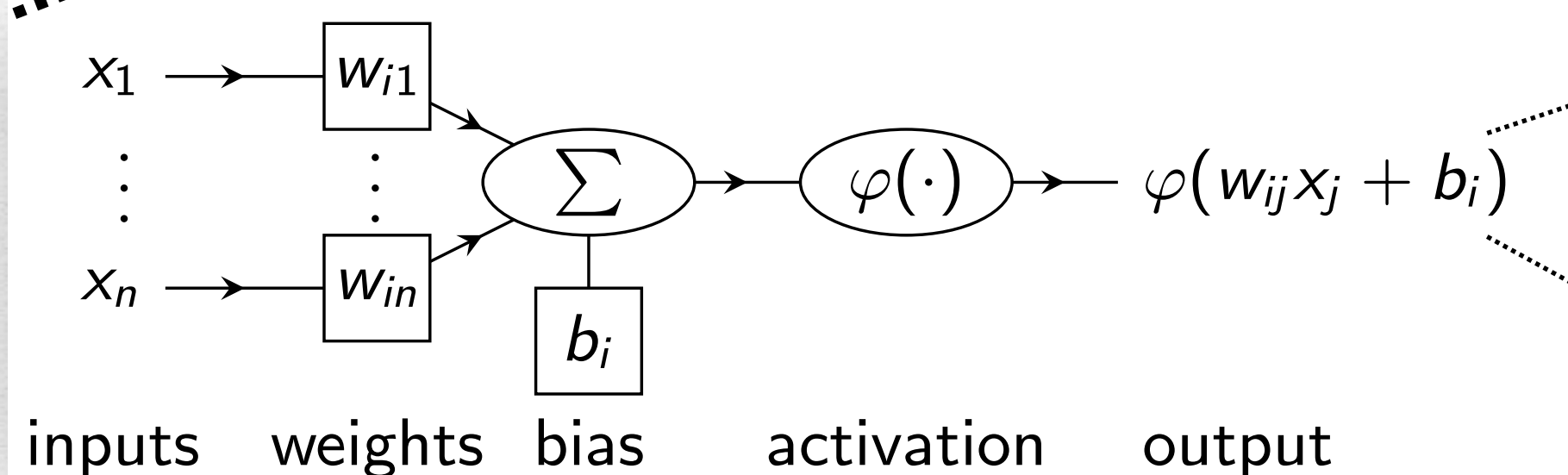
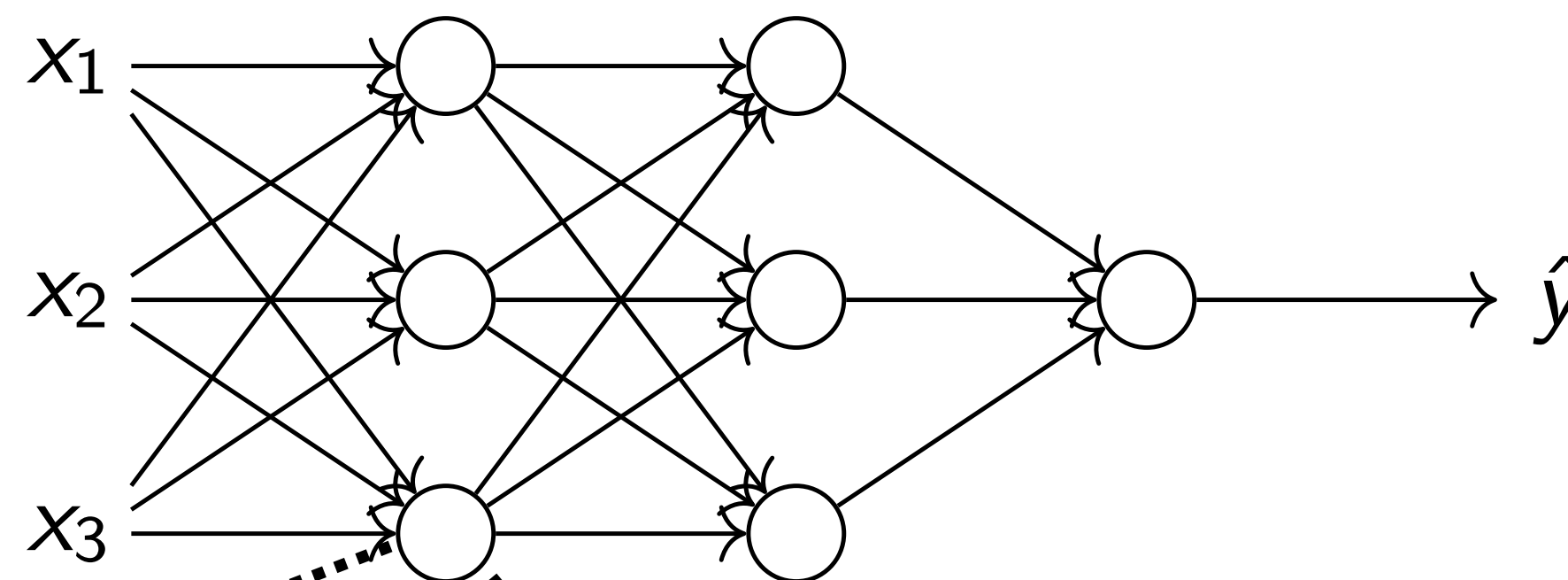
訓練による潜在変数の決定: $\mathbf{w}_{ij}, \mathbf{b}_i$

損失関数 (loss function の最適化)

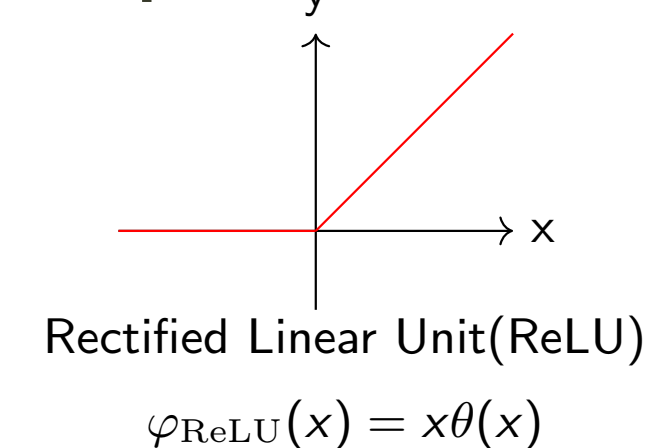
$$L(y, \hat{y})$$

Higgs
 $y = (1, 0)$

QCD
 $y = (0, 1)$



ϕ : 活性化関数 (非線形性)

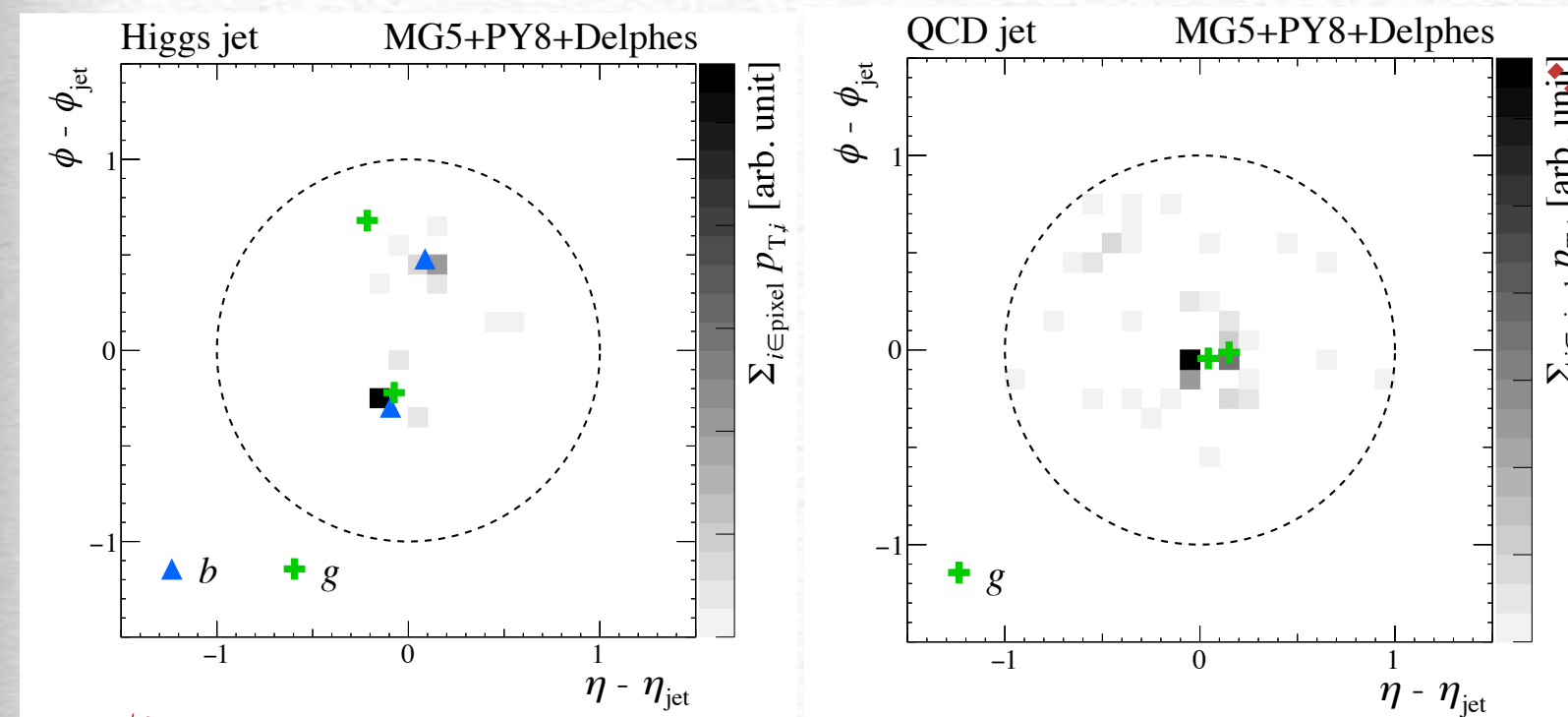


MLP (multiple layer perceptron) の基本構造

Input: Jet images

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

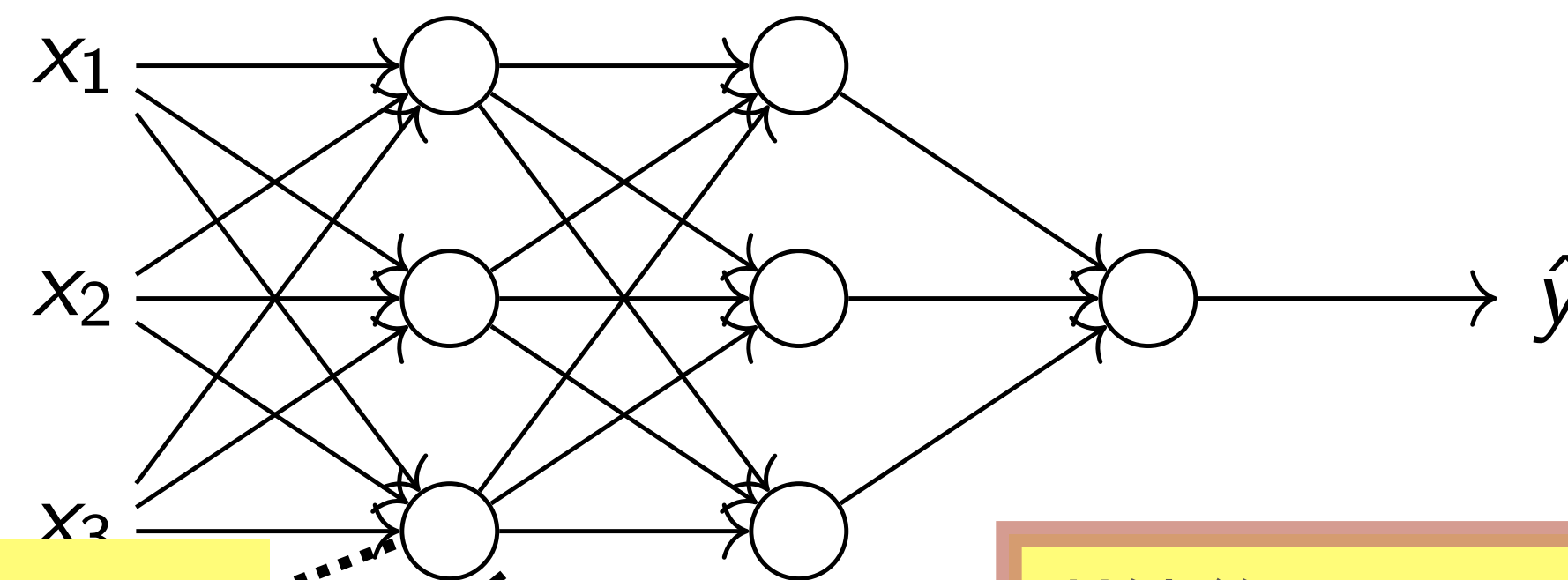
QCD $(j+Z(\rightarrow \nu\nu))$



訓練による潜在変数の決定: $\mathbf{w}_{ij}, \mathbf{b}_i$

損失関数 (loss function の最適化)

$$L(y, \hat{y})$$

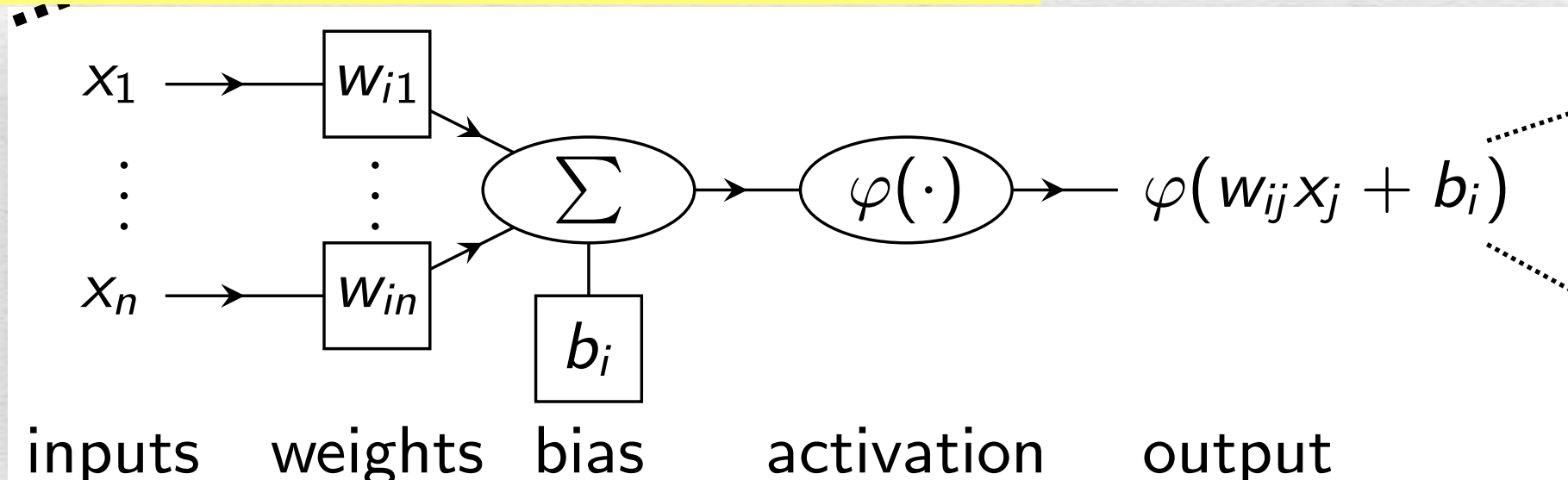


Higgs
 $y = (1, 0)$

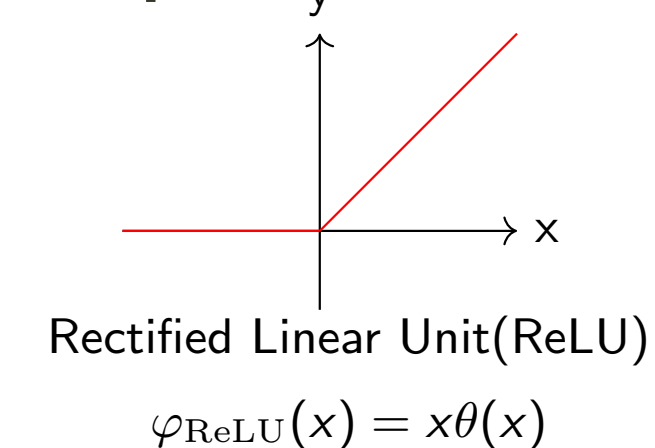
QCD
 $y = (0, 1)$

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



ϕ : 活性化関数 (非線形性)



訓練とは:

Loss function とbackward propagation

素粒子物理への応用の場合に気になること

- ❖ データ $d=\{X, y\}$ 入力データ $X=\{x_i\}$ とその意味 (y)
 - ❖ 動物の画像と動物の種類(こぶた、たぬき、きつね、ねこ...)
 - ❖ 文章とその翻訳
 - ❖ [素粒子の場合] 測定器の反応と実際に起こったこと
- ❖ 意味を提供せず深層学習アルゴリズムに判断させることもできる
(unsupervised learning)
 - ❖ 異常検知、要約 →応用上の問題点 偏見

深層学習における正確さ:認知バイアス

❖ 人間の場合

- ❖ 科学者→男性 政治家→男性
- ❖ 看護師→女性 小学校教員→女性
- ❖ 与えられたデータの中にある相関を勝手に読み込む。翻訳において性別補完をしたり、特定の人種、性別の画像だけ作ってきたりする。[広報とかで使おうとしているが今のところ炎上の元]
- ❖ 深層学習は統計的な判断しかしていない。
 - 使う側がデータの癖を理解していないとほぼ役に立たない。)
 - 解釈可能性(interpretation) やバイアス訂正機能が重要
- ❖ 物理応用としては、今まで物理解析に使わないからと放置していた相関に関して、より真剣に取り組むようになった。
- ❖ 画像生成も仔細にみるとやばくなっていることがある。→異常検知ってほんとにできるの？

ゾンビか？割と気づかないが
怖い部分

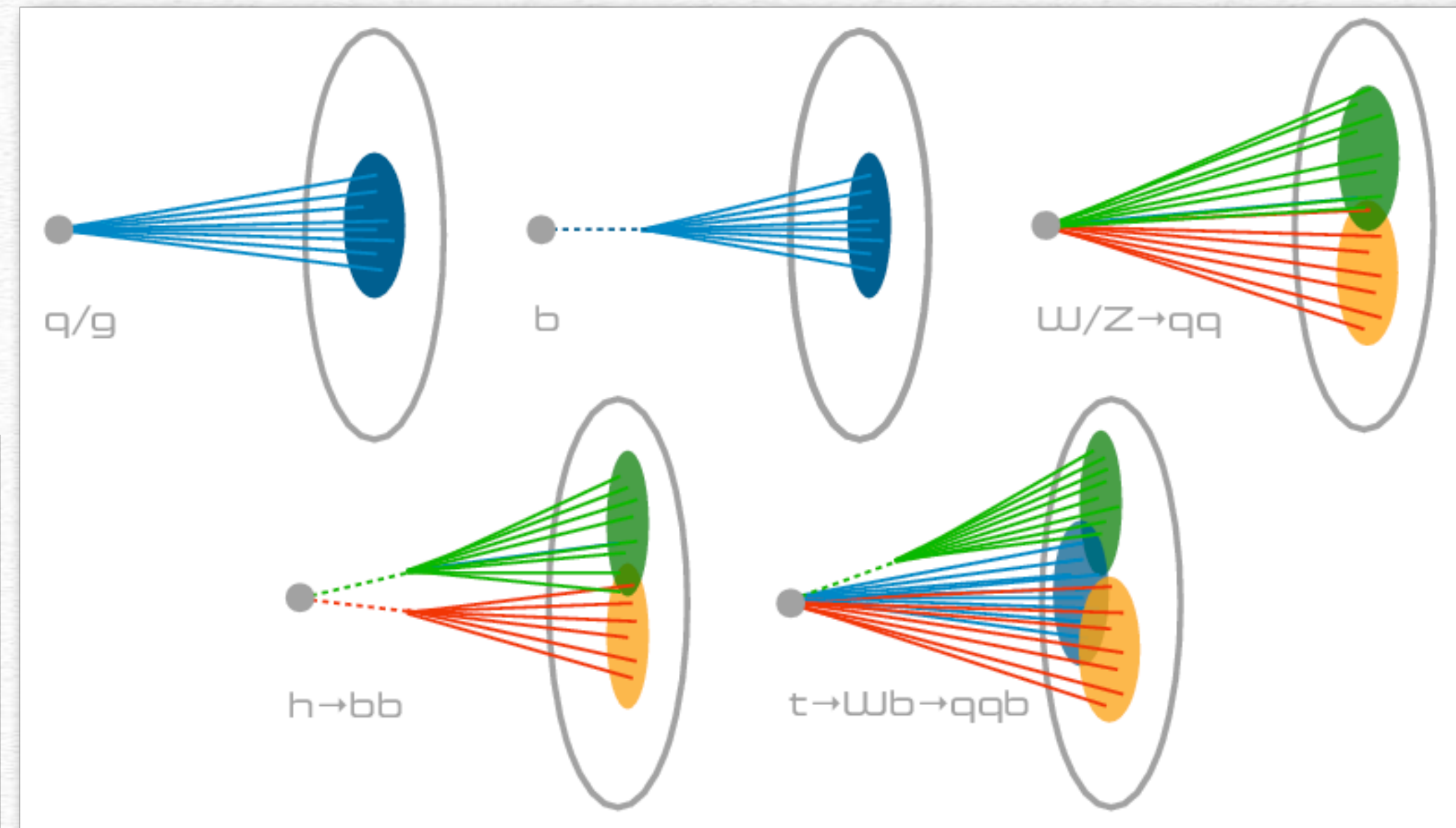
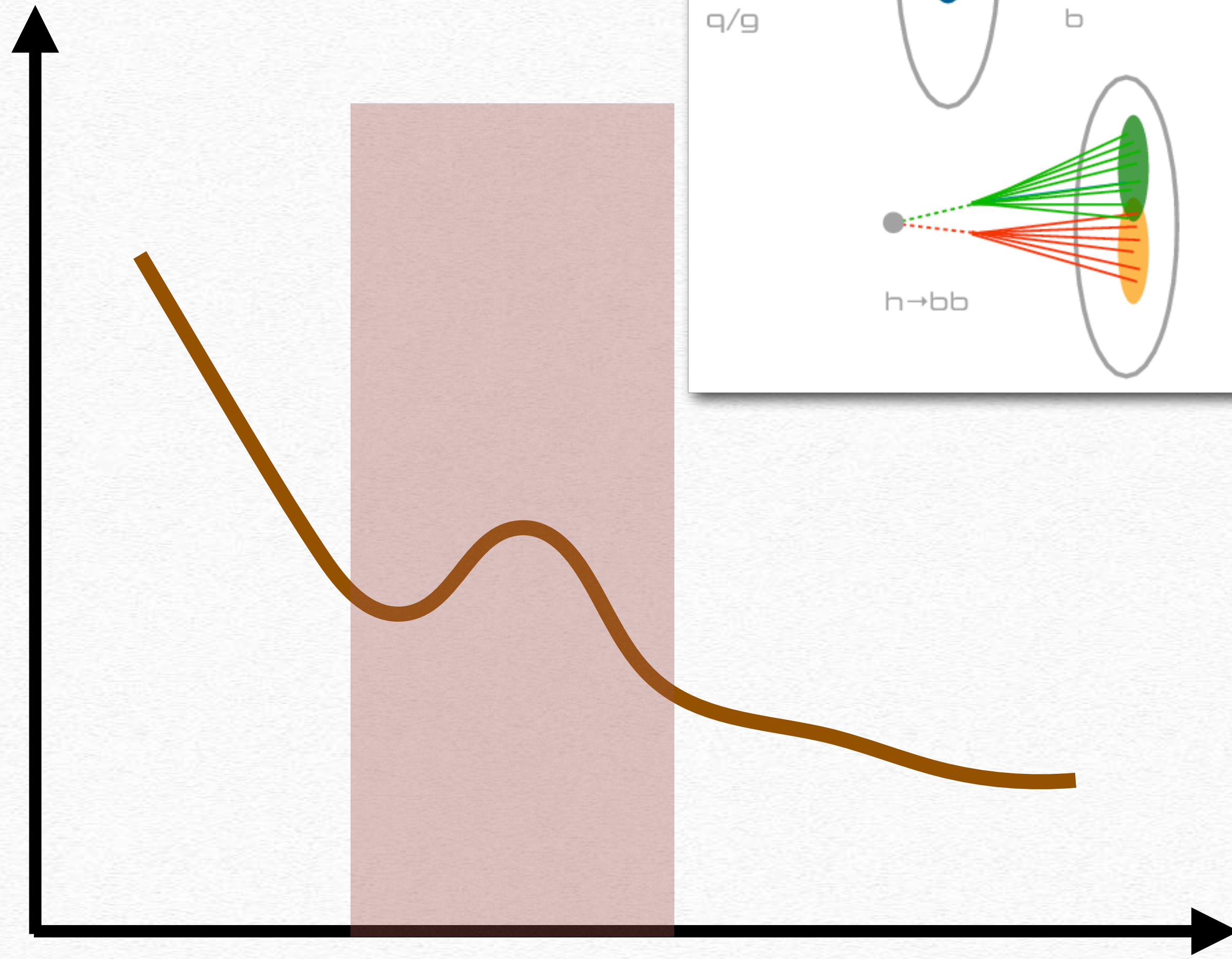
日本人が生成されない



指が6本

腕が途切れる

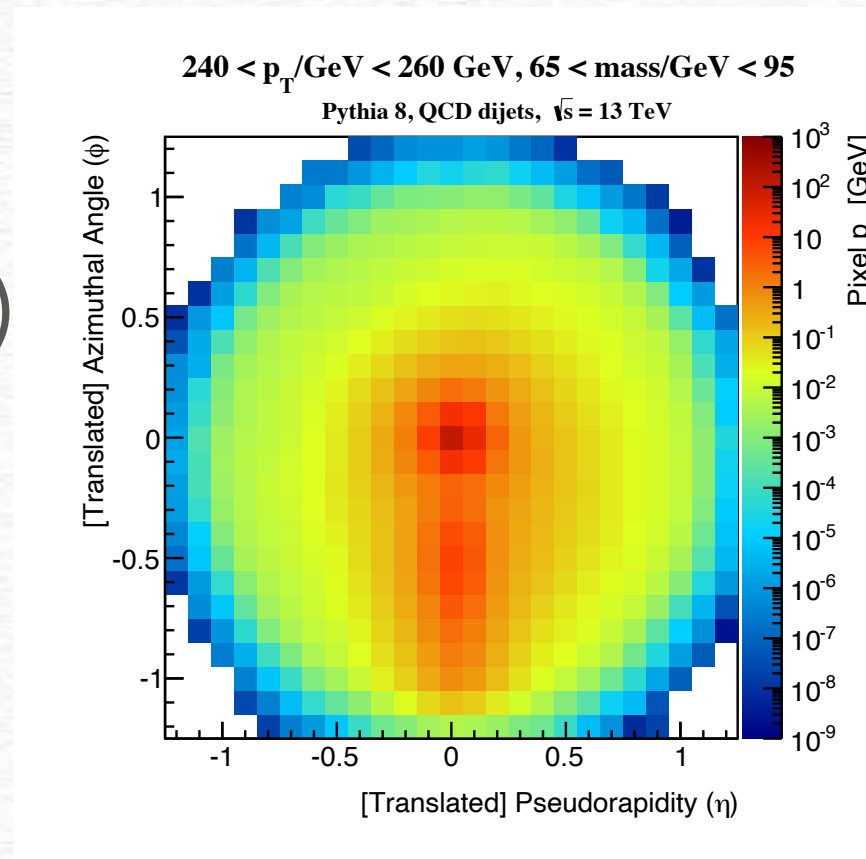
異常検知



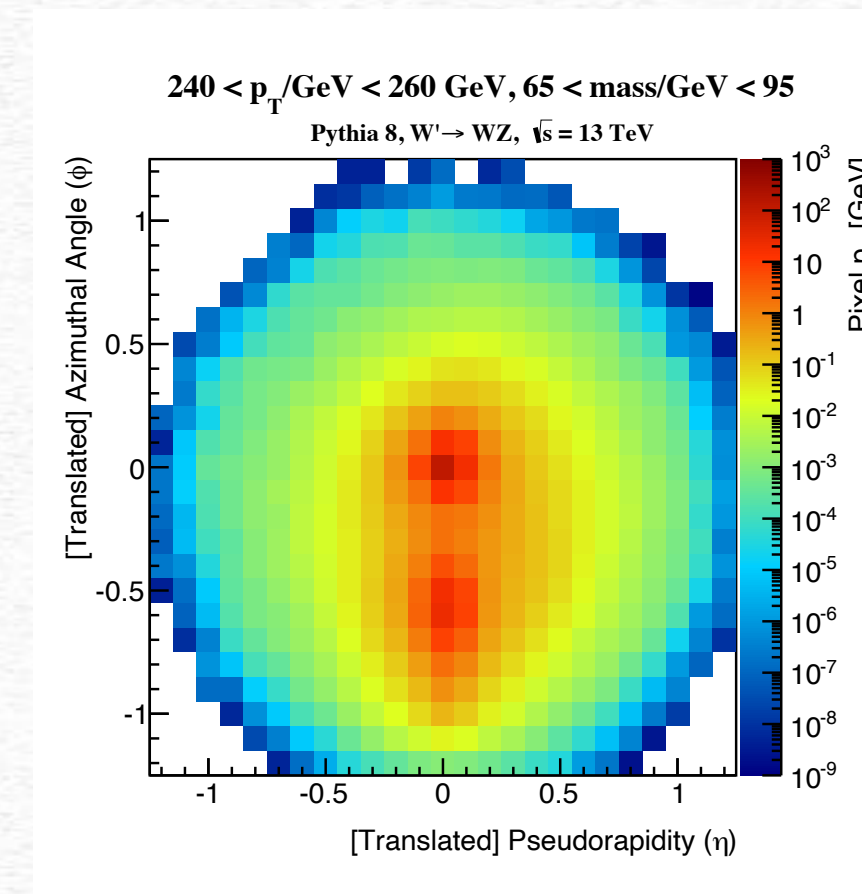
m_{jj}

Jet classification using ML

QCD jet
(in W mass region)

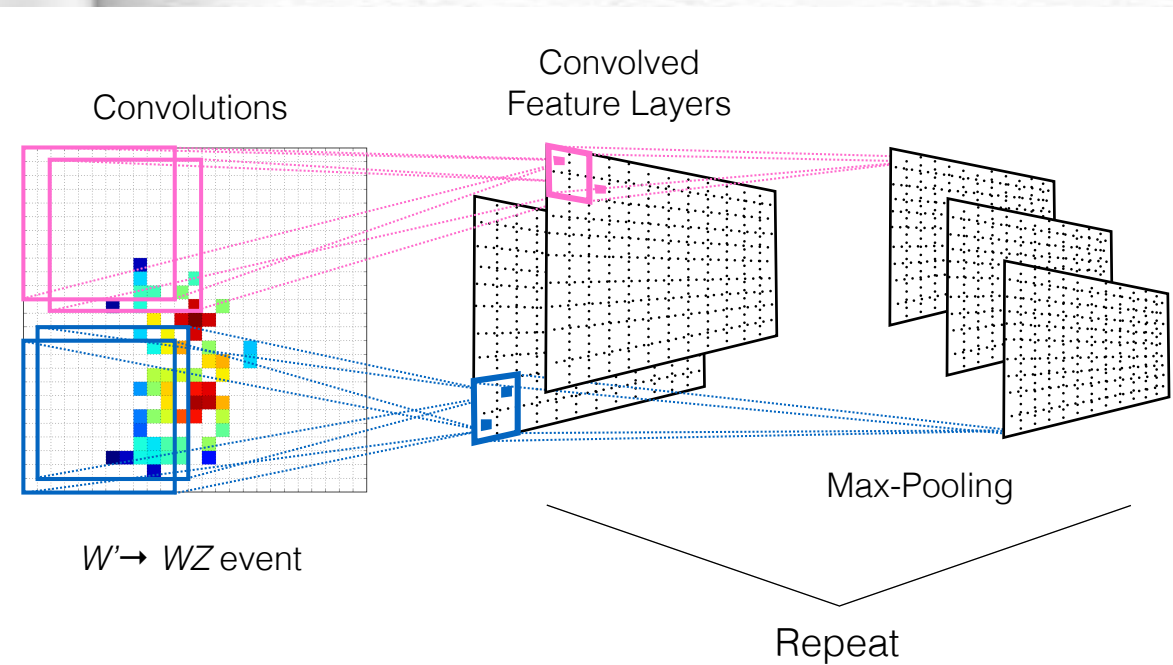


W jet



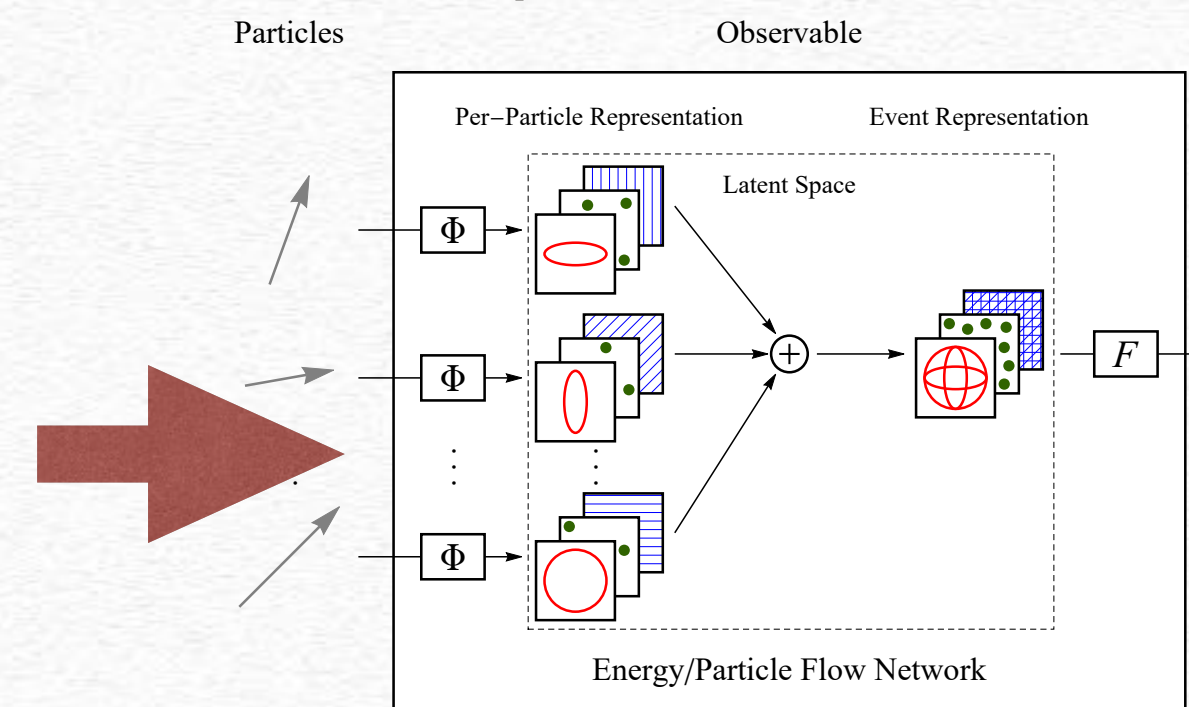
from Schwartzman et al
<https://iopscience.iop.org/article/10.1088/1742-6596/762/1/012035>

Jet Image



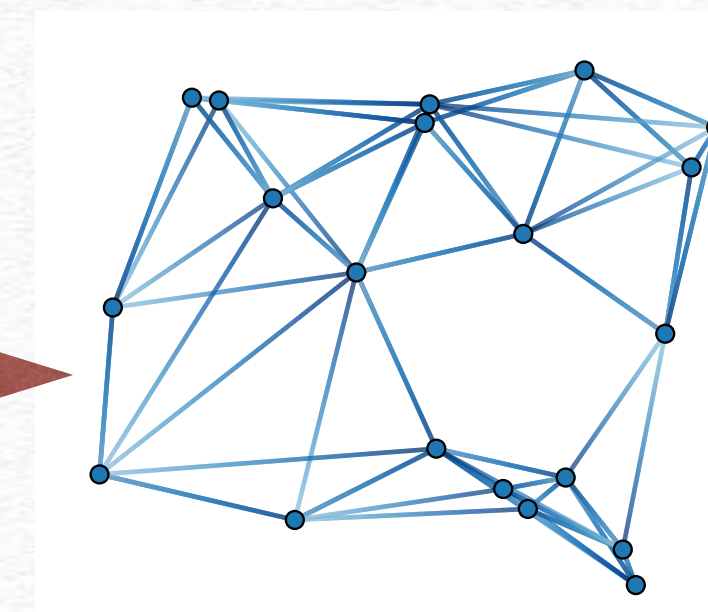
CNN Oliverira et al
(1511.05190)

集合として



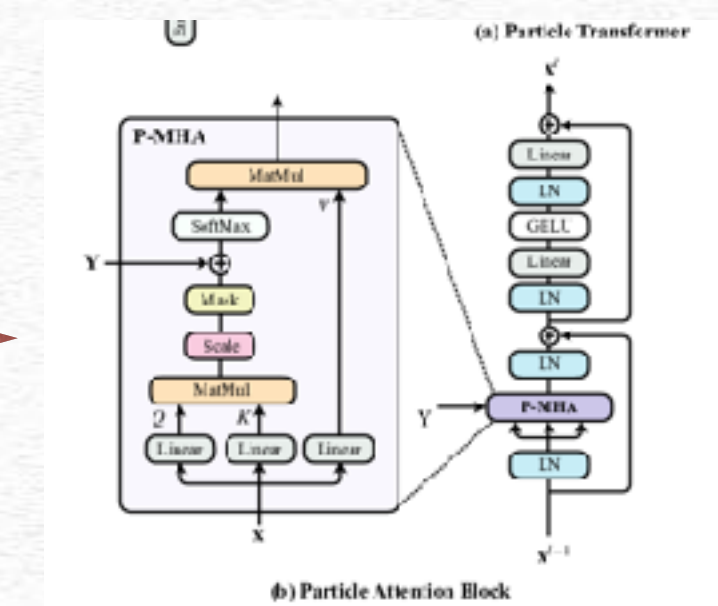
permutation invariance
(Energy Flow Network and
Particle Flow Network 1810.05165)

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

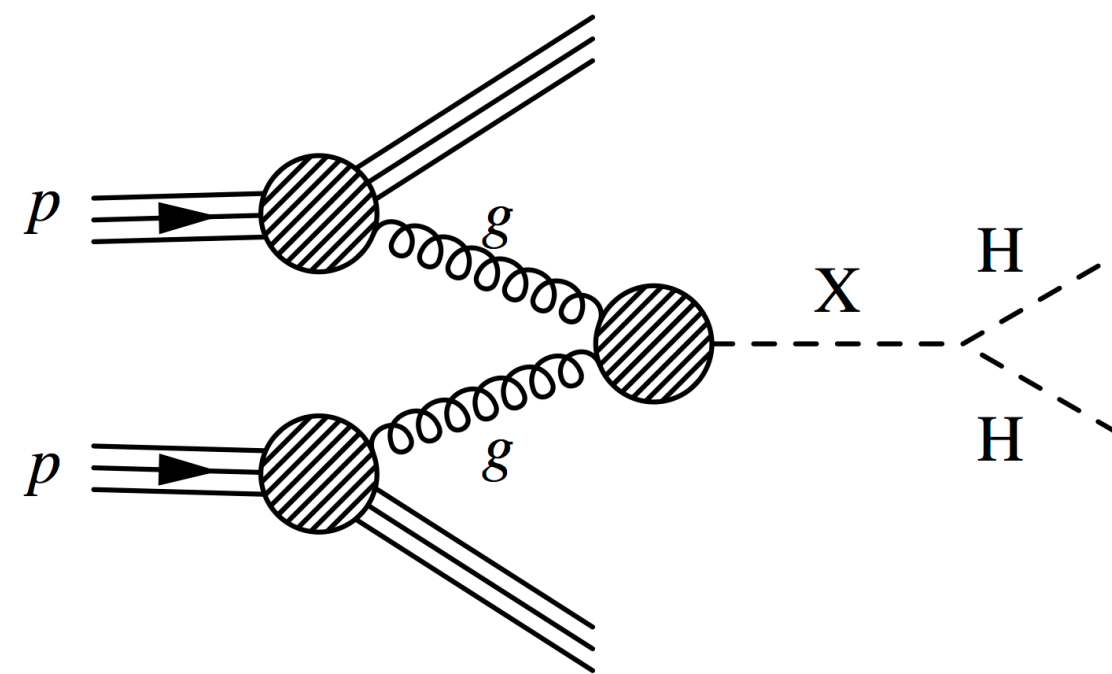


building key
and query
2202.03772

Bogatskiy et al PELICAN (2211.00454)

レゾナンスの再構成

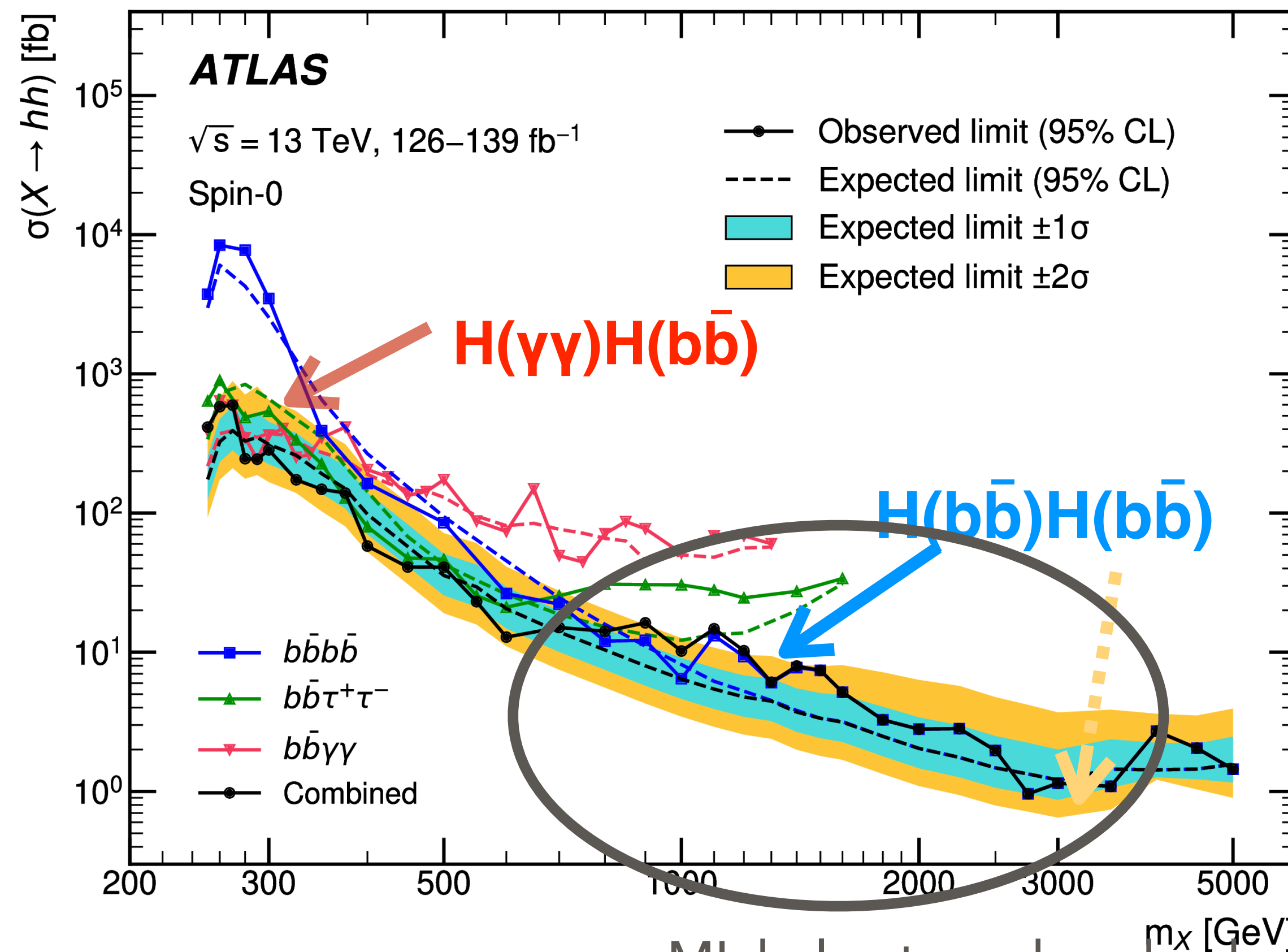
$X \rightarrow HH$



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

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

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

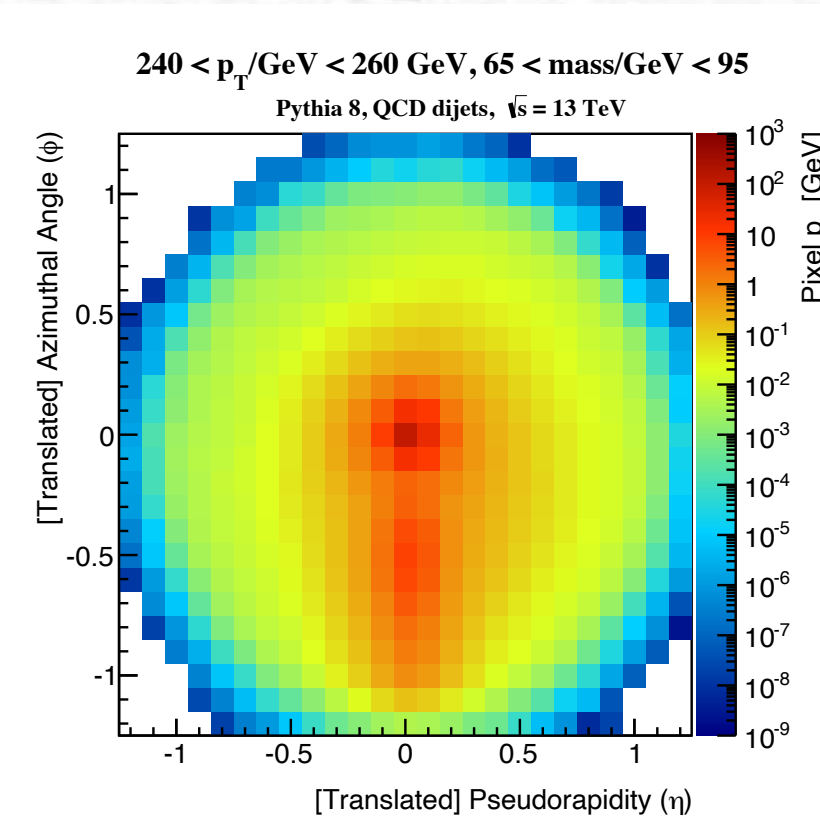


ML helps to reduce background

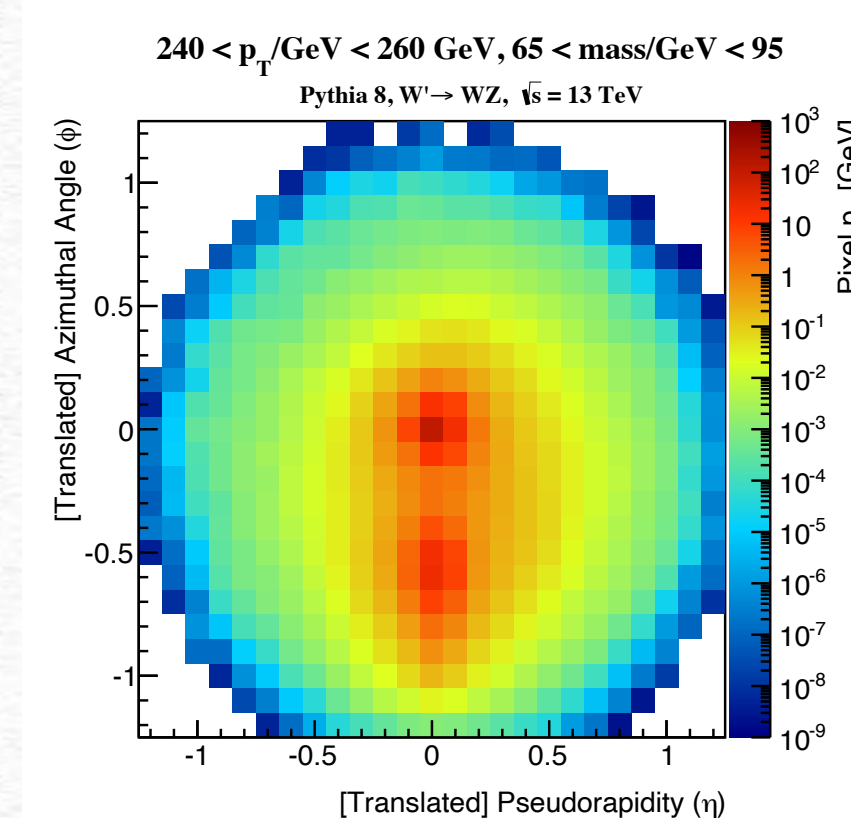
CMS 実験では
次のjet IDは単一の
深層学習モデルに

Jet classification using ML

QCD jet
(in W mass region)

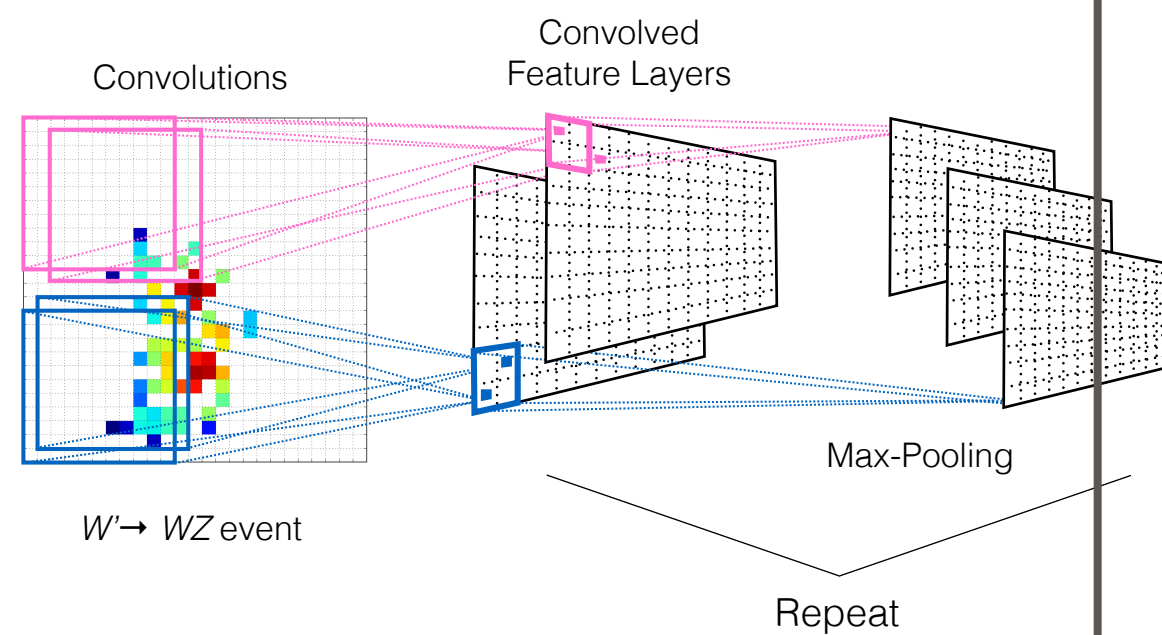


W jet



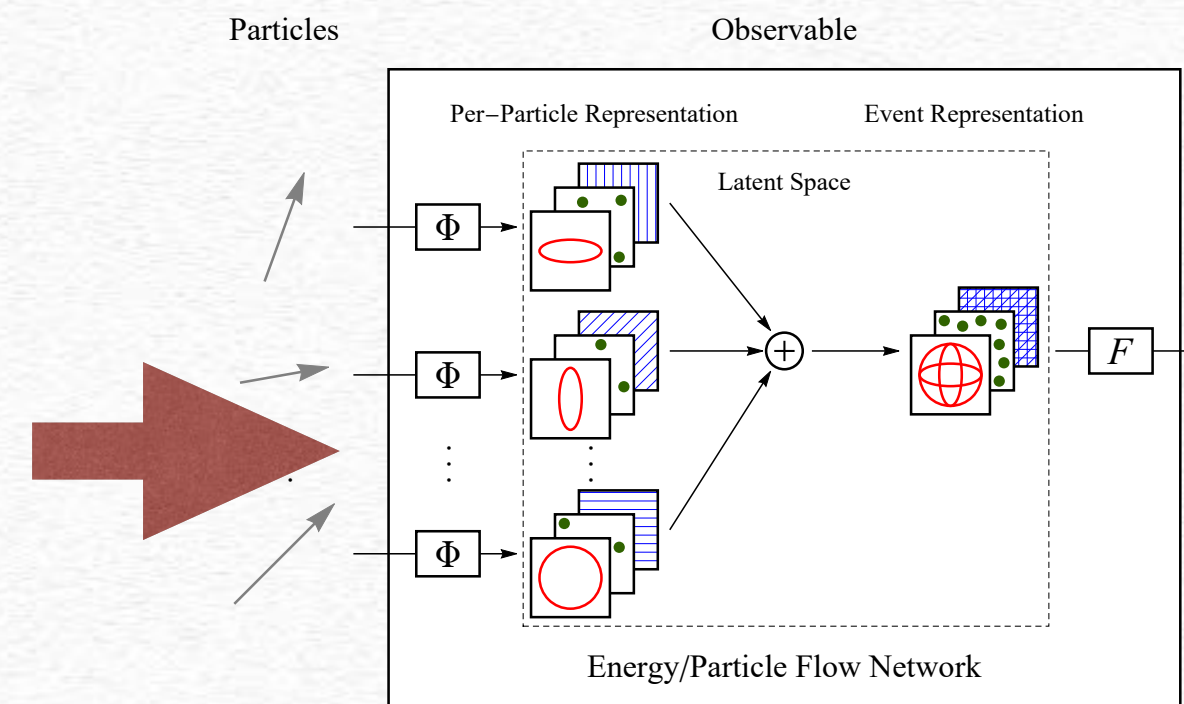
from Schwartzman et al
<https://iopscience.iop.org/article/10.1088/1742-6596/762/1/012035>

Jet Image



CNN Oliverira et al
(1511.05190)

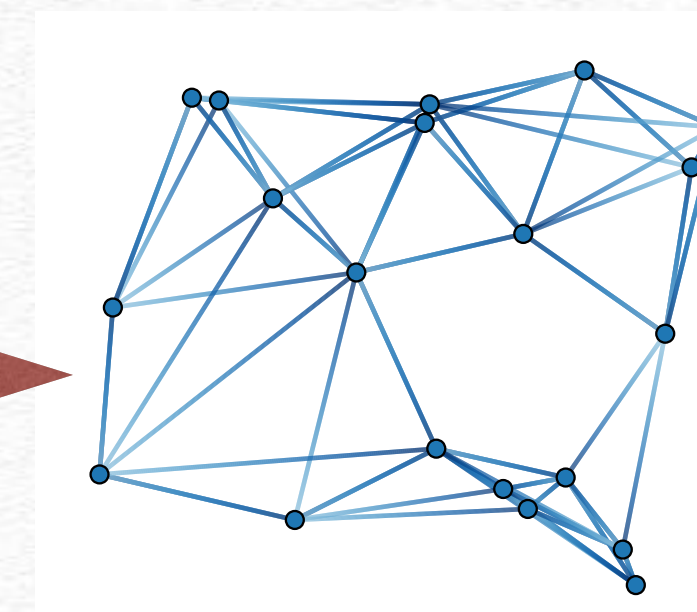
集合として



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

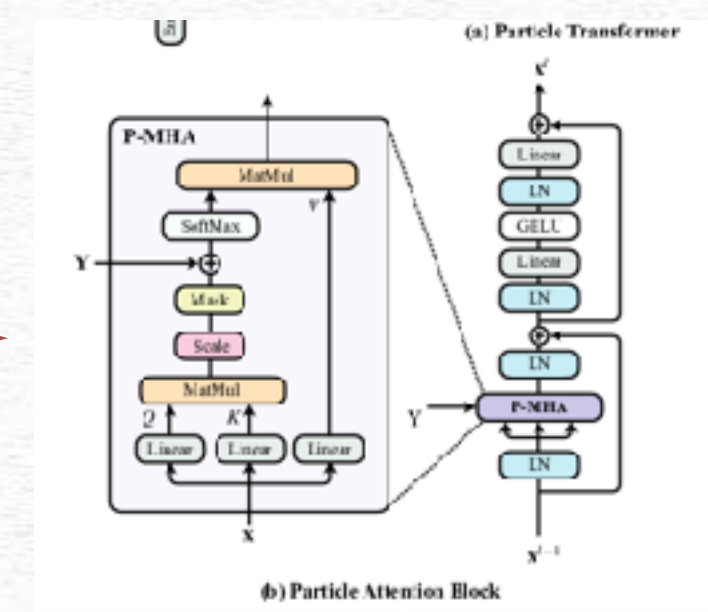
Bogatskiy et al PELICAN (2211.00454)

グラフとして



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

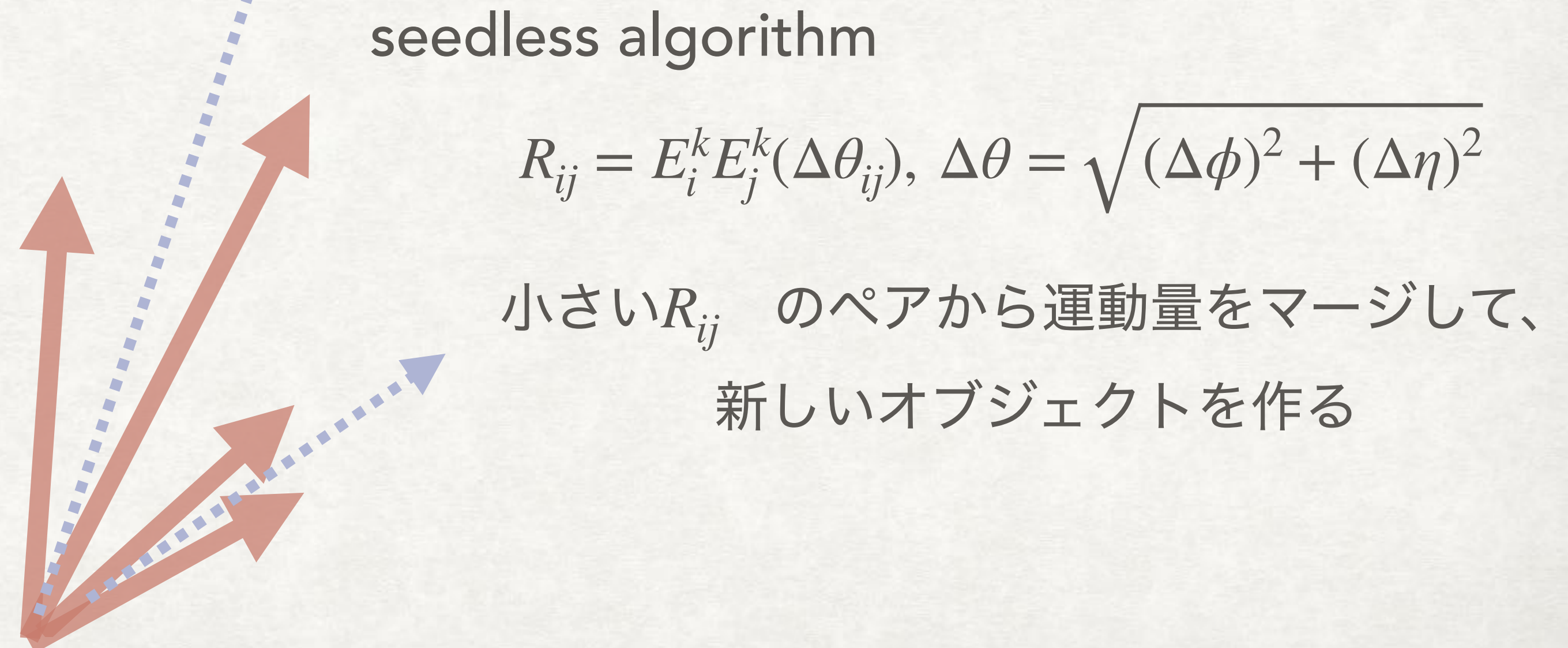
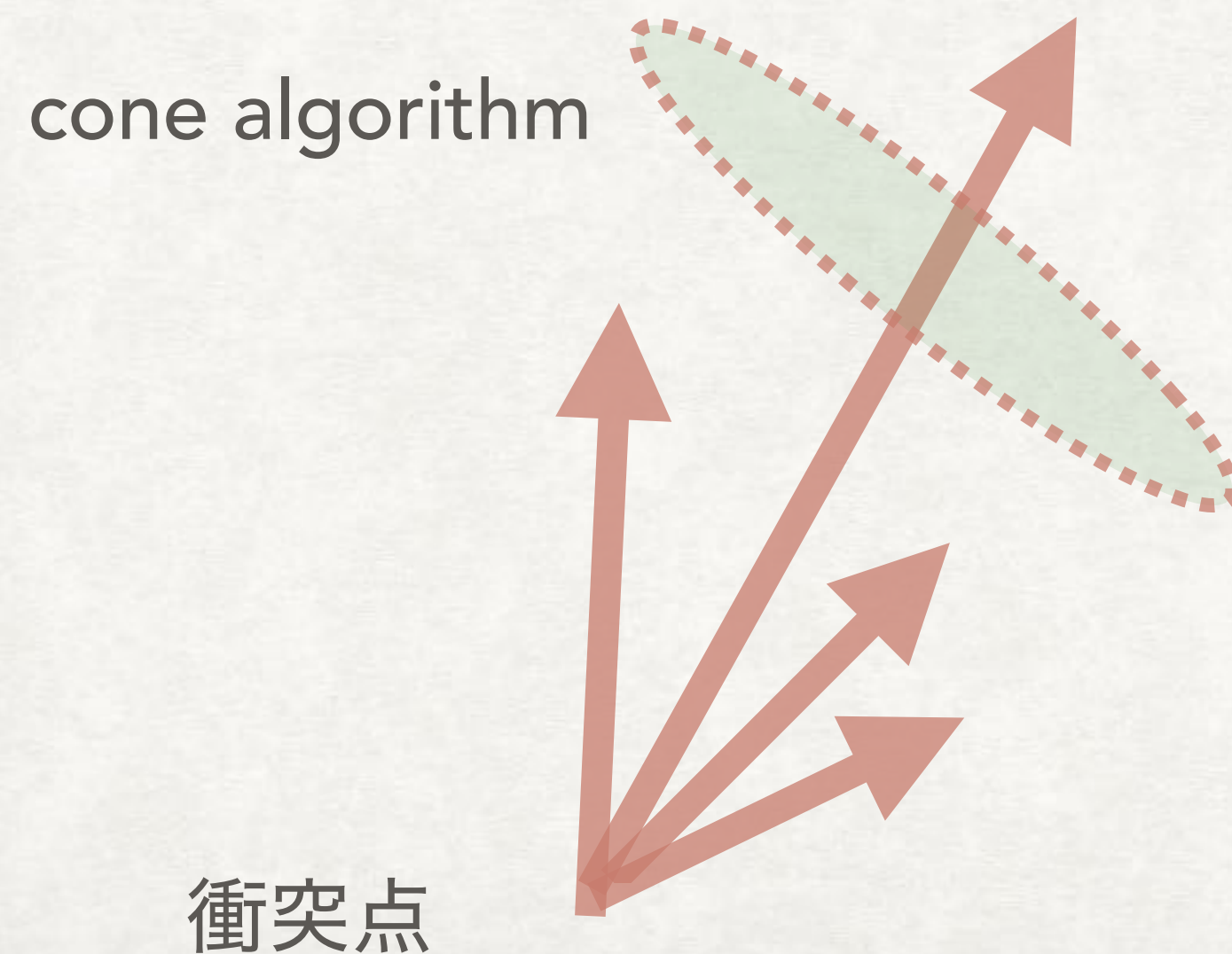
transformer



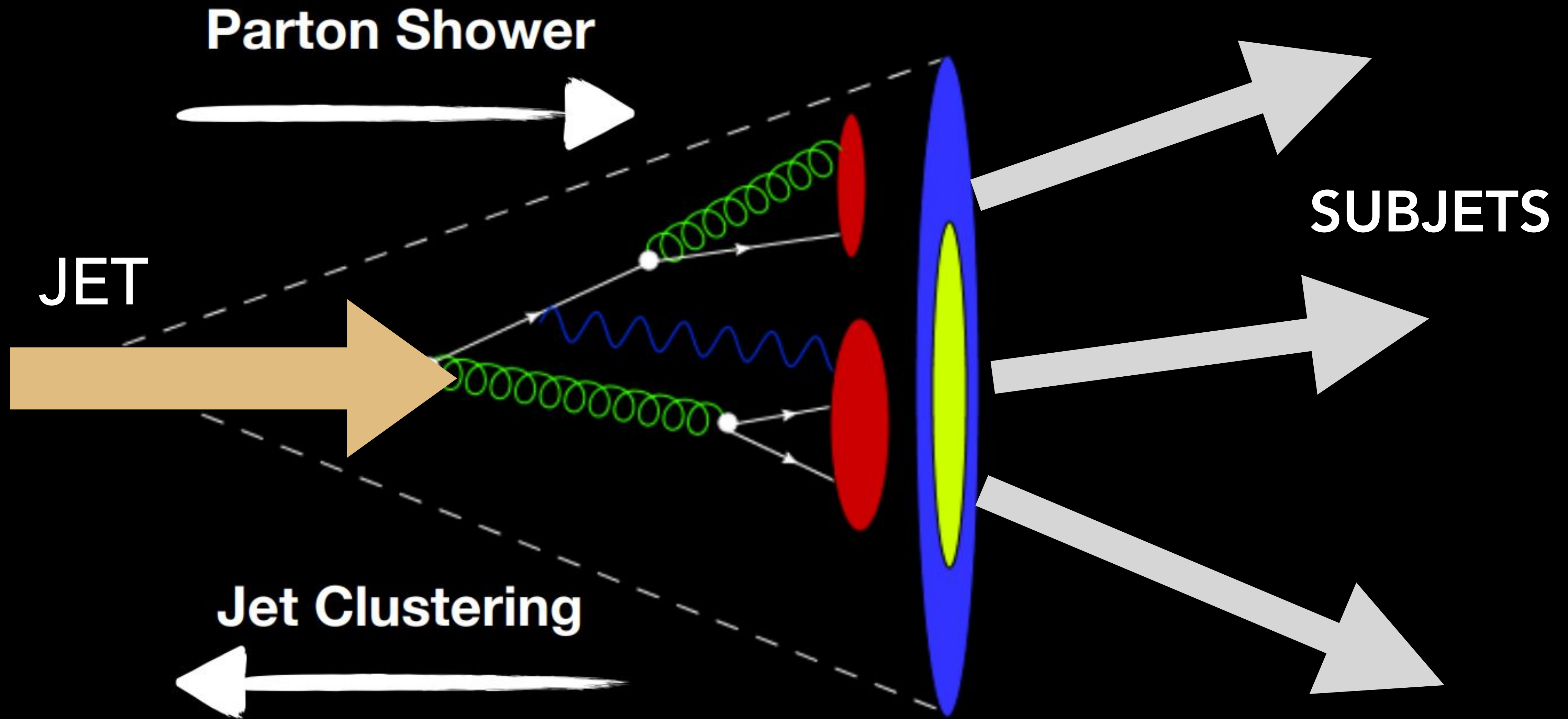
building key
and query
2202.03772

粒子入れ替えに対する対称性(DEEP SET)

- ジェット内部の粒子に「順番」はないので、粒子の集合に対して関数を作るべき
- DEEP SET(集合) ネットワークを作る場合は、粒子の入れ替え対象性を保証することが重要
- 画像認識問題や、言語の問題では順番に対して意味があることが多いので、ここで必ず素粒子用の工夫が入る。
- **set に対して作用するという概念は素粒子分野では新しいものではなくて Jet の定義は 1990年代後半から、粒子集合にたいして働くように工夫されている。**

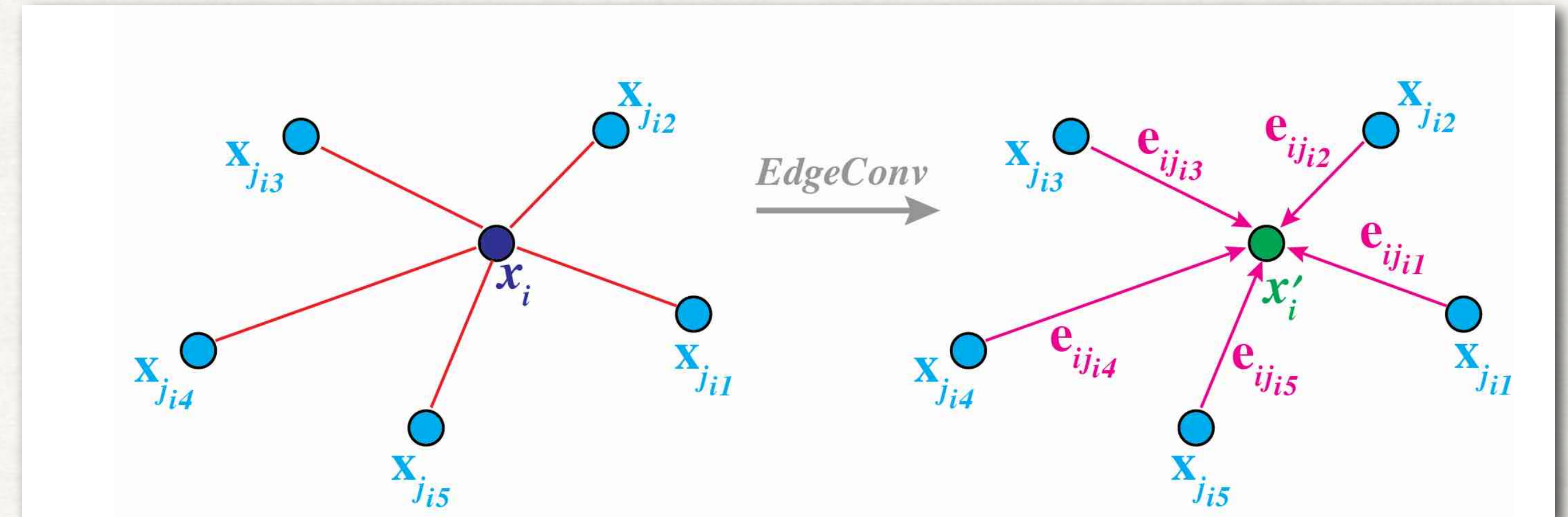
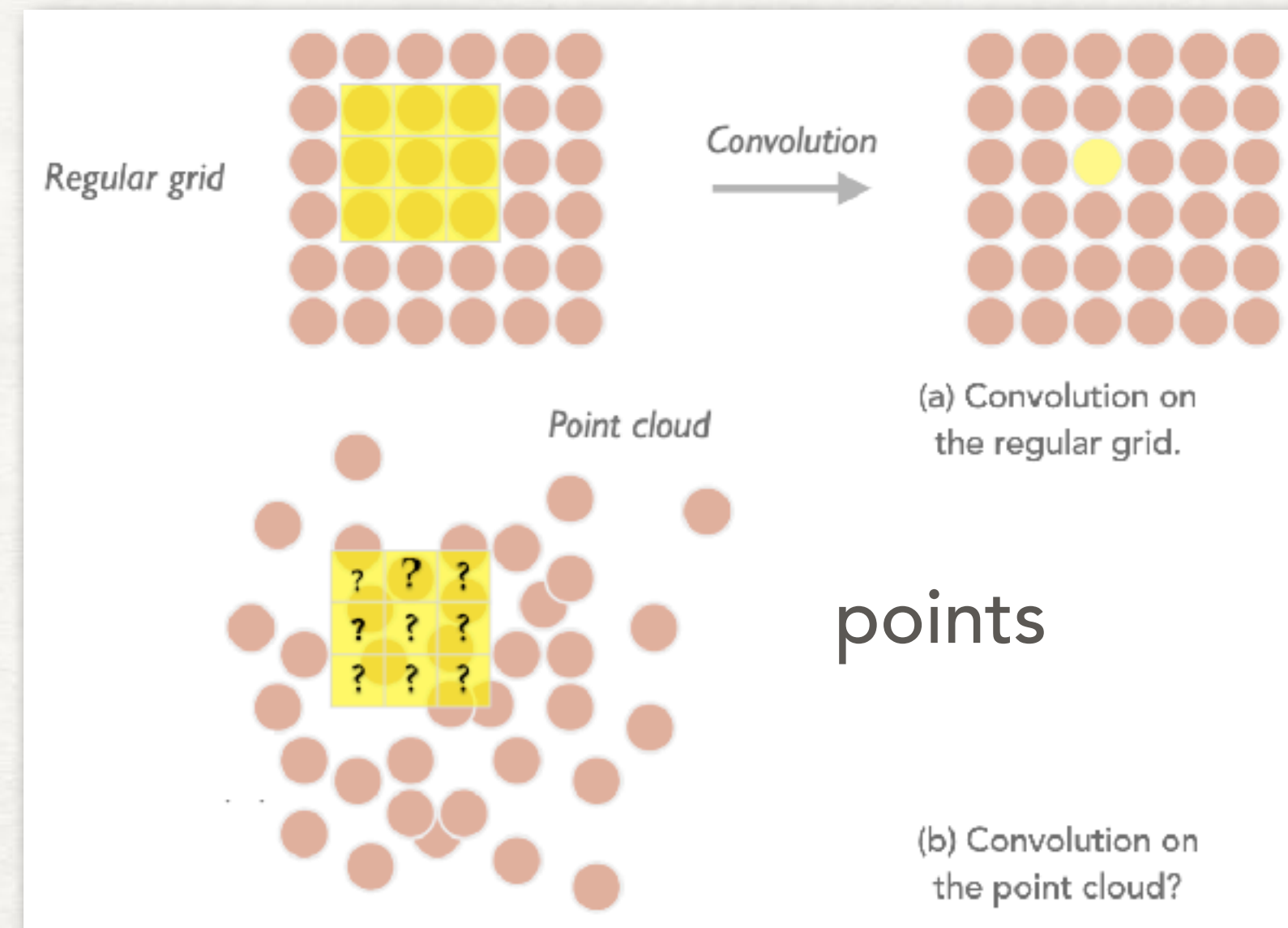


SUBJET の分布をカットベースで調べていくことで、ジェット分類が始まった



QCD community have developed sophisticated theoretical treatment about this.

PARTICLE NET(1902.08570)



$$x'_i = \square_{j=1}^k h_{\Theta}(x_i, x_{ij}) \quad x_i \in R^F$$

エッジ関数 $h_{\Theta}(x_i, x_{ij}) : R^F \times R^F \rightarrow R^{F'}$

$\square$: channel wise symmetric operation (最大、平均、和など)

k ; 近傍の粒子数 7~16

ジェットは疎なデータ(sparse data)
 なので画像と捉えるのはあまり
 賢くない。

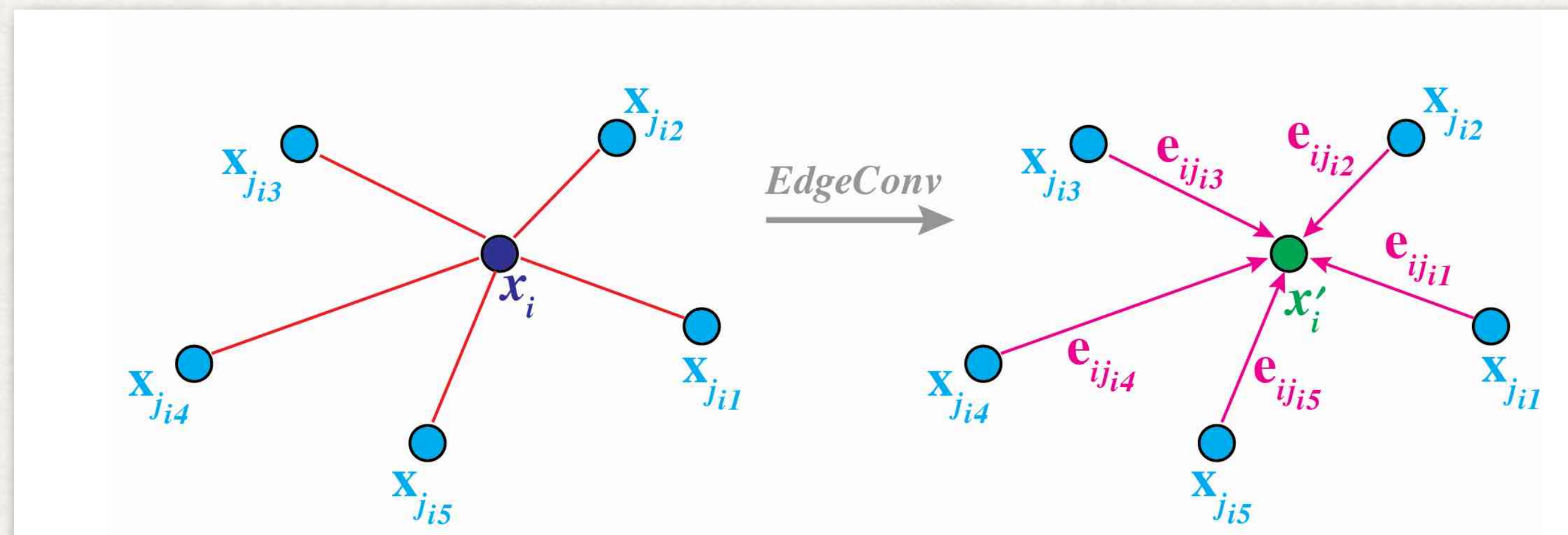
物理をもとにした深層学習の取り組み LUND NET(2012.08526)

Particle NET : 粒子一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

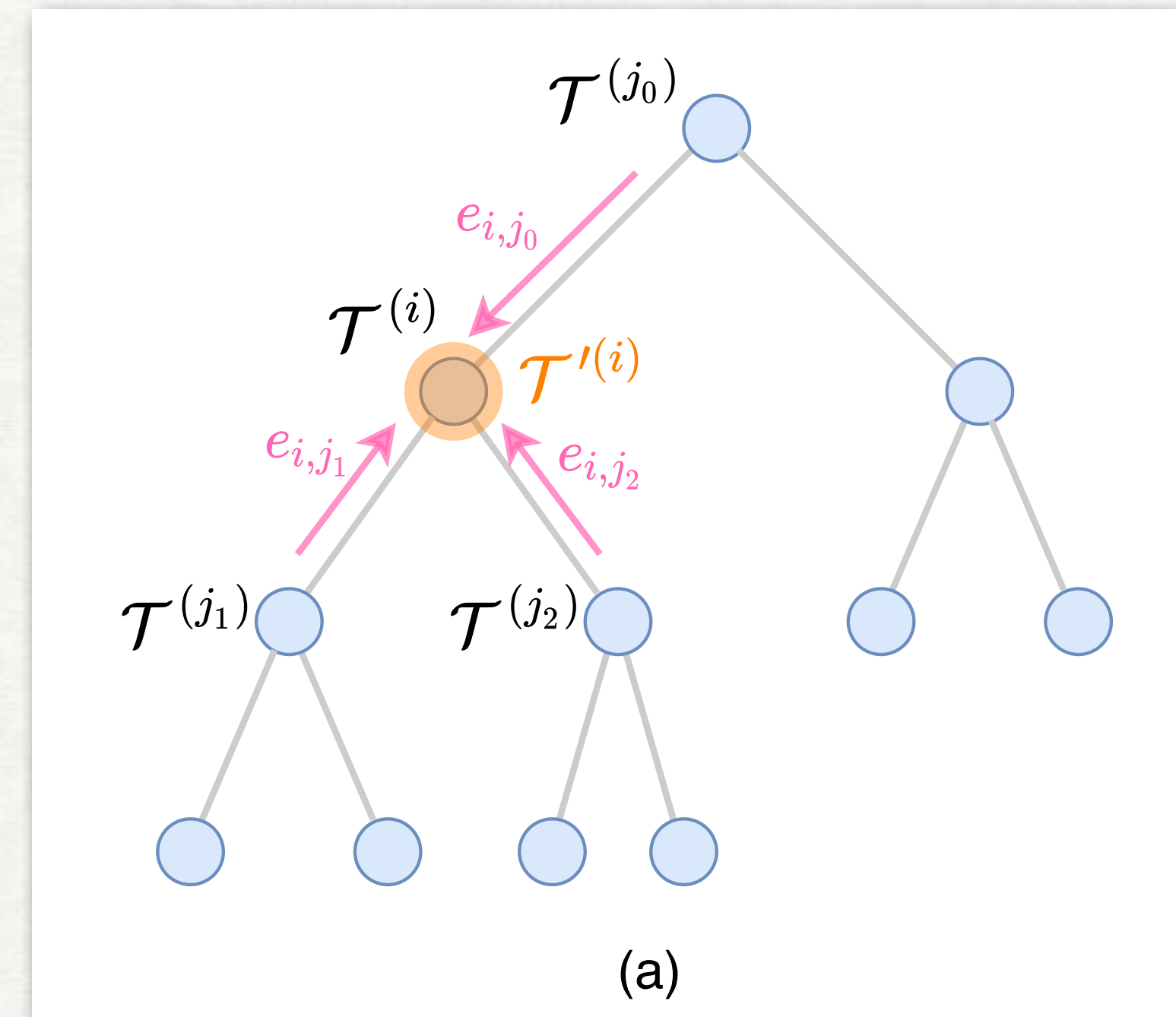
LundNet: 粒子の情報:サブジェット情報

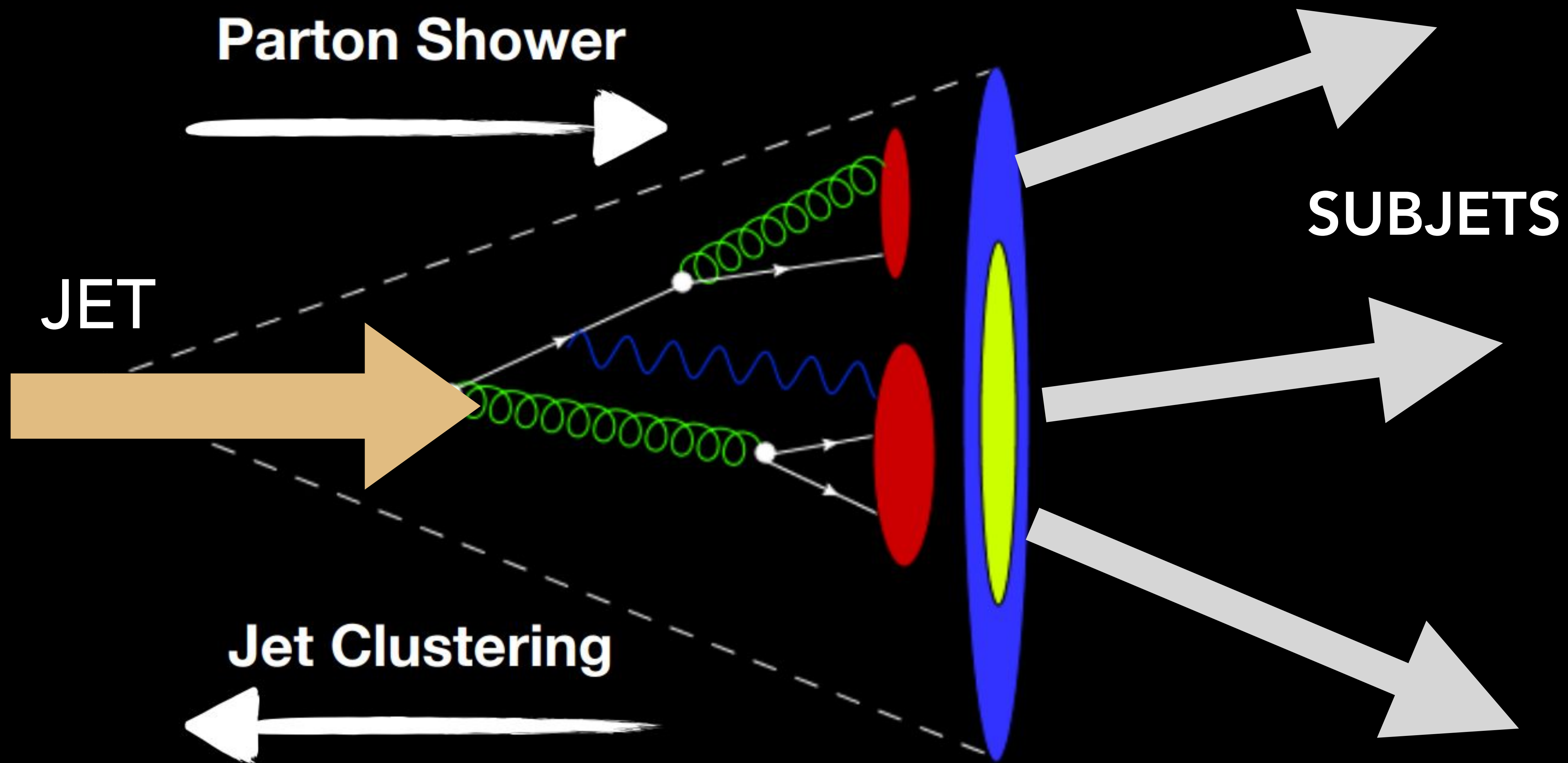
グラフ:ジェットクラスタリング過程

「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

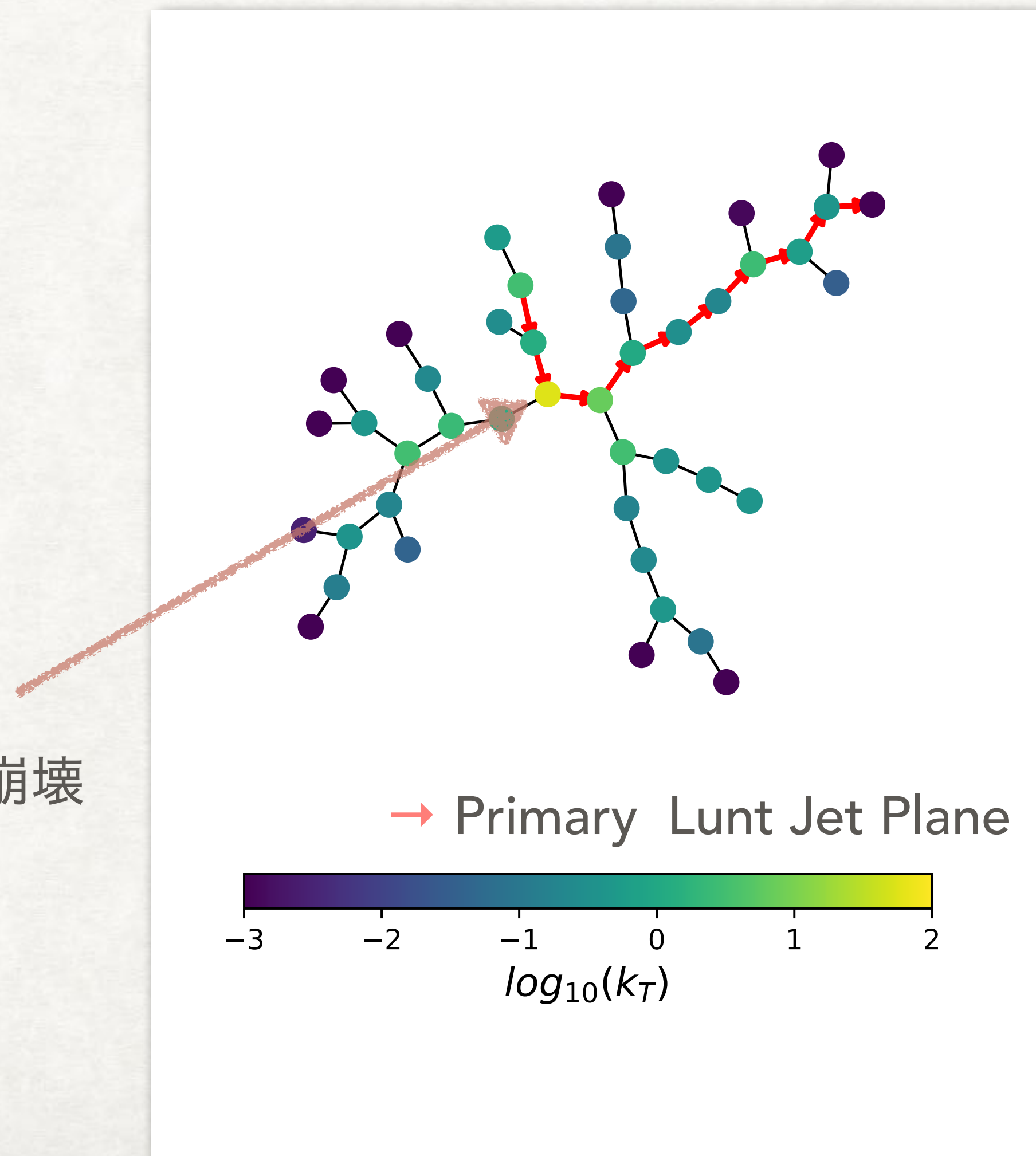




QCD community have developed sophisticated theoretical treatment about this.

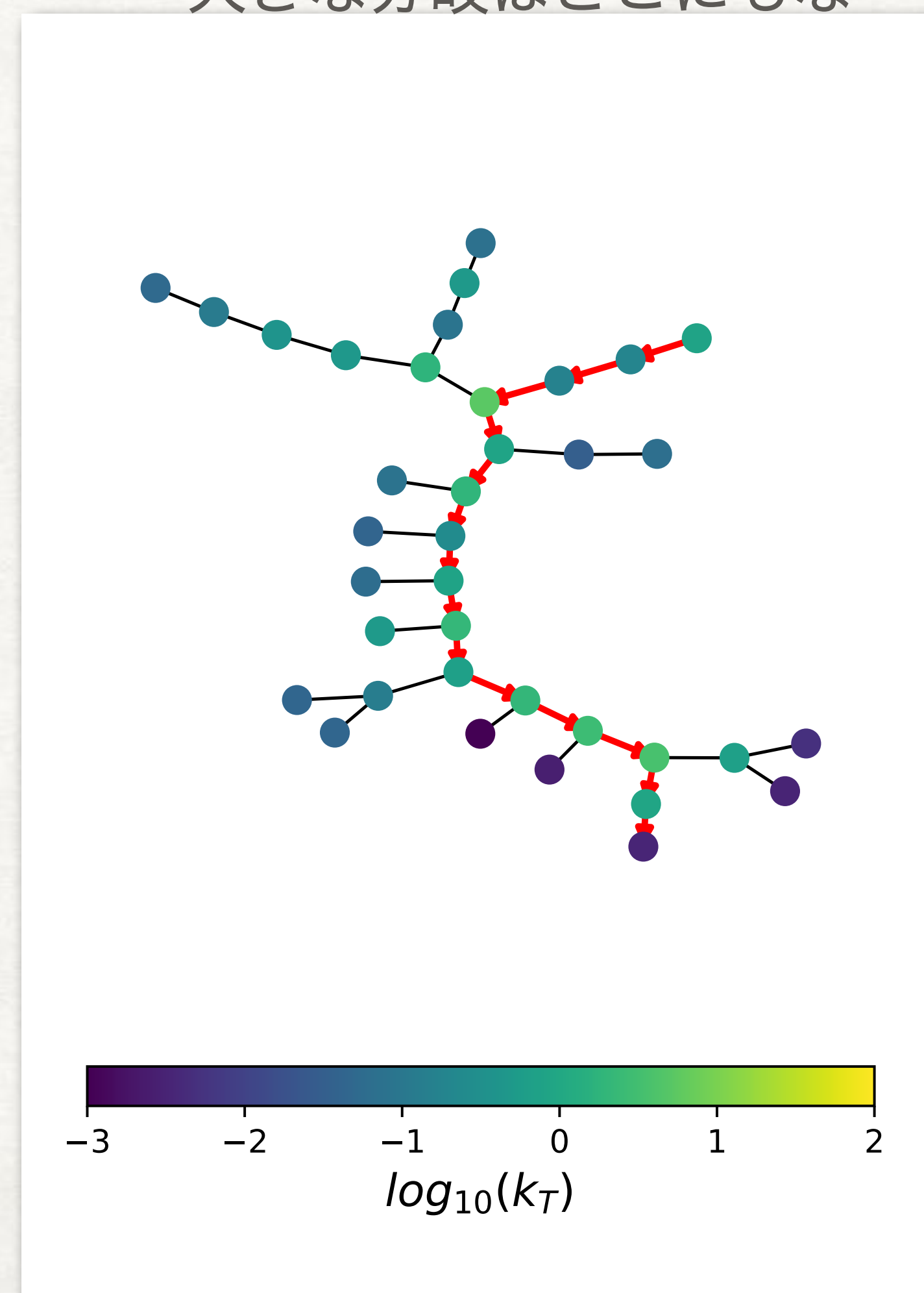
$\log(k_T)$ LUND TREE

W jet 大きな分岐が存在する



QCD JET

大きな分岐はどこにもな



パフォーマンスと理論不定性のやりとり

小さい横方向運動量の分岐を取り込んでいくほど、性能の良くなっていく。

これはネットワークを工夫することにより低エネルギー粒子の相関をとりいれていくということに対応。

低エネルギーの粒子の分布を正しく計算する理論が存在しないので、ここで理論と実験データの関係がとれなくなって詰む

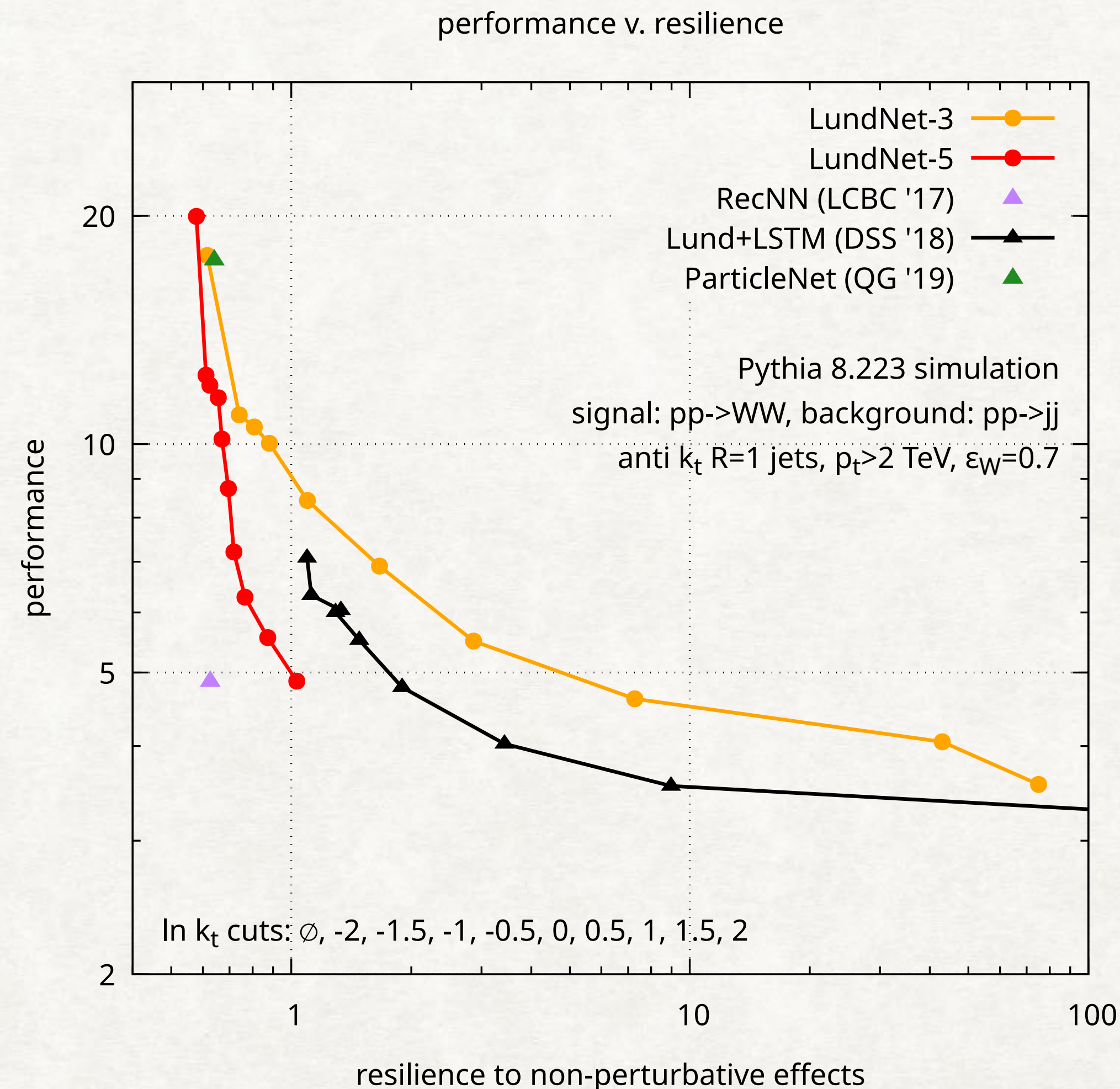
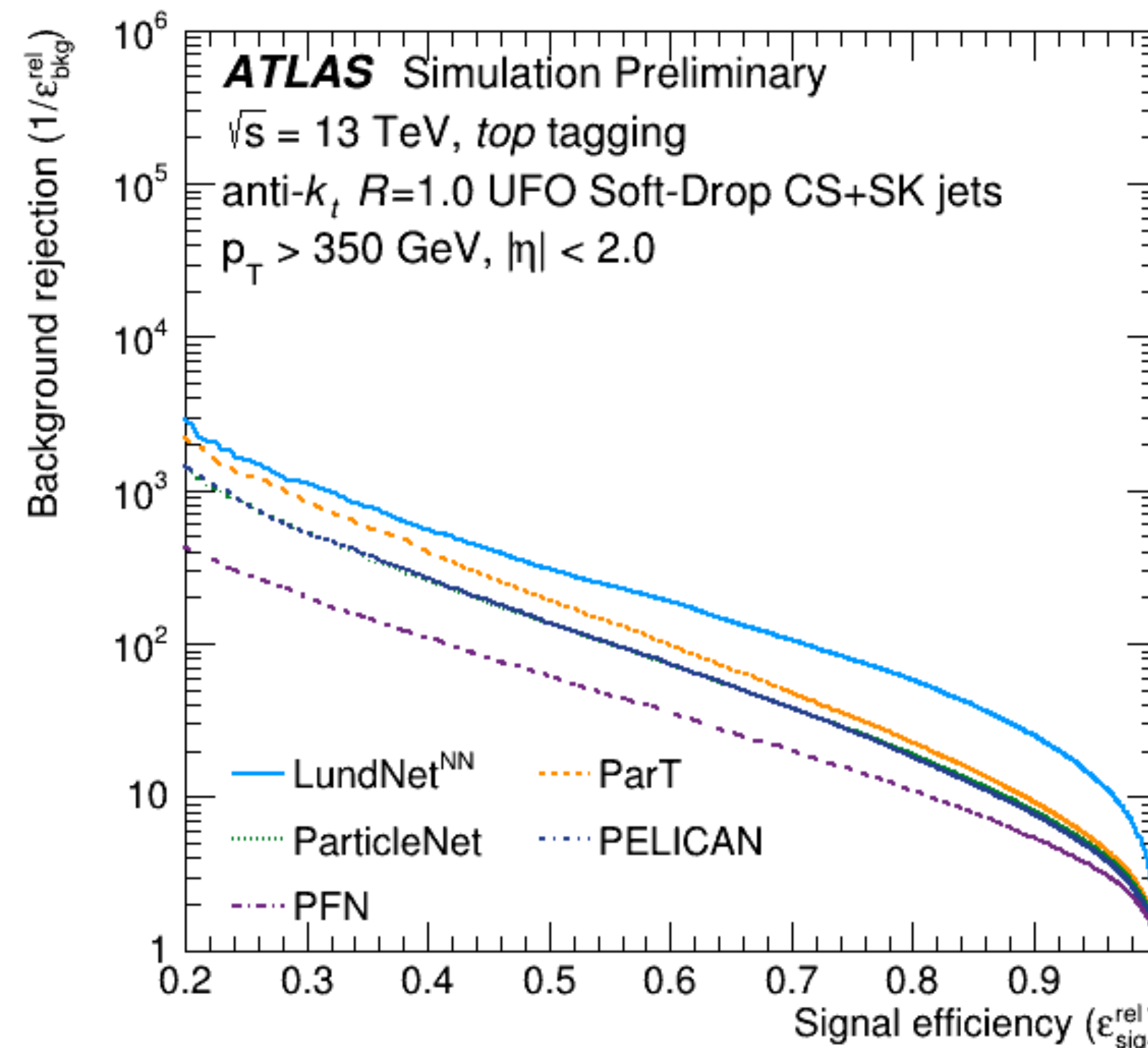
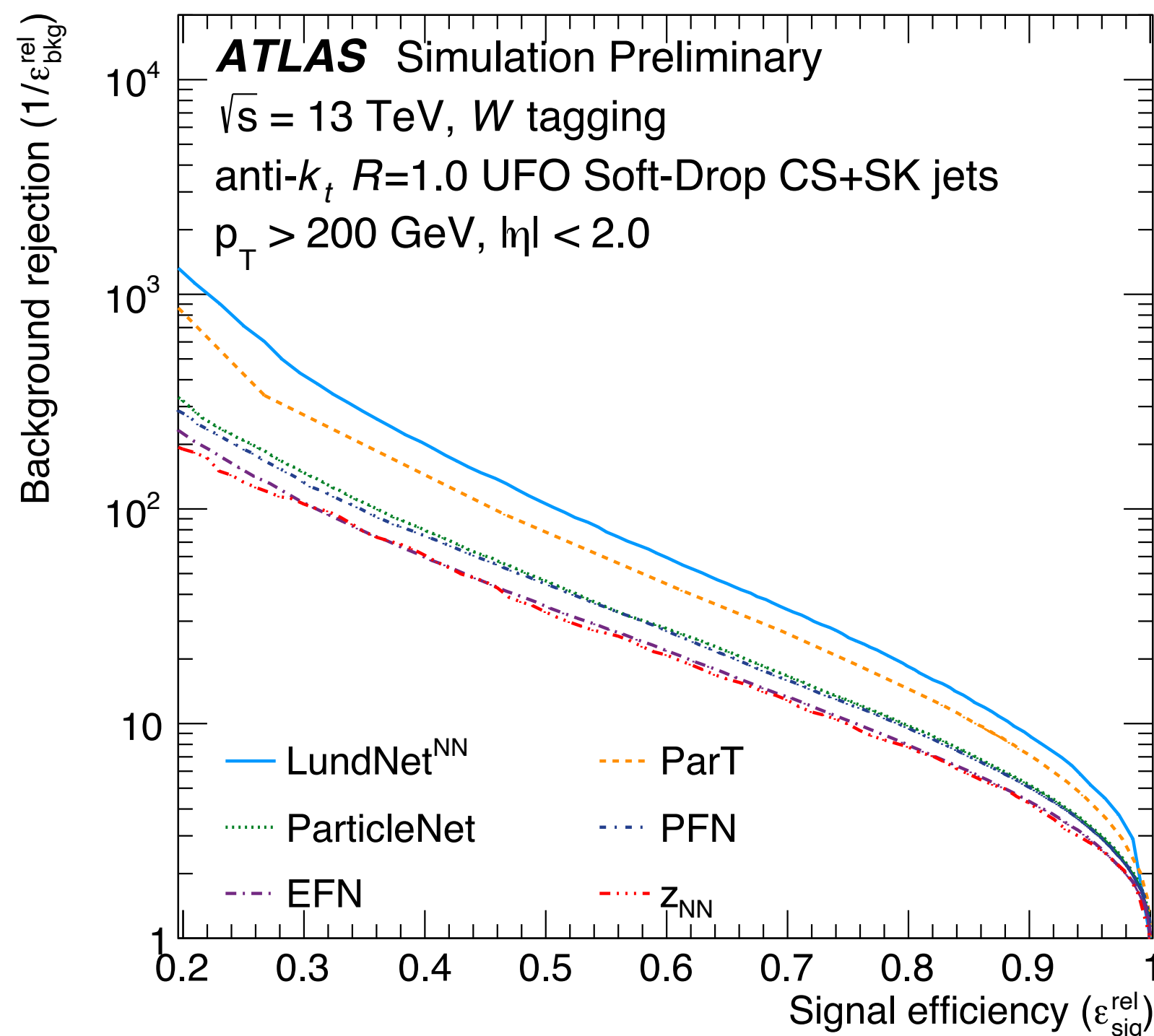


Figure 8. Performance $\frac{\epsilon_W}{\sqrt{\epsilon_{QCD}}}$ versus resilience to non-perturbative effects.

Of all explored architectures in ATLAS, the physics driven Lund-net tagger performs the best so far

- Better than standard transformers or symmetry driven PELICAN
- Is stronger hyper-parameter optimization needed?
- How does the modeling dependence fit into this picture?
- Are detector details important?



Particle transformer や
 PELICAN といった性能の高い
 ネットワークより安全なところで
 運用したLundNet の方が、
 系統誤差が少なく良い結果を
 出したりする時もある。

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はず。

- 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)

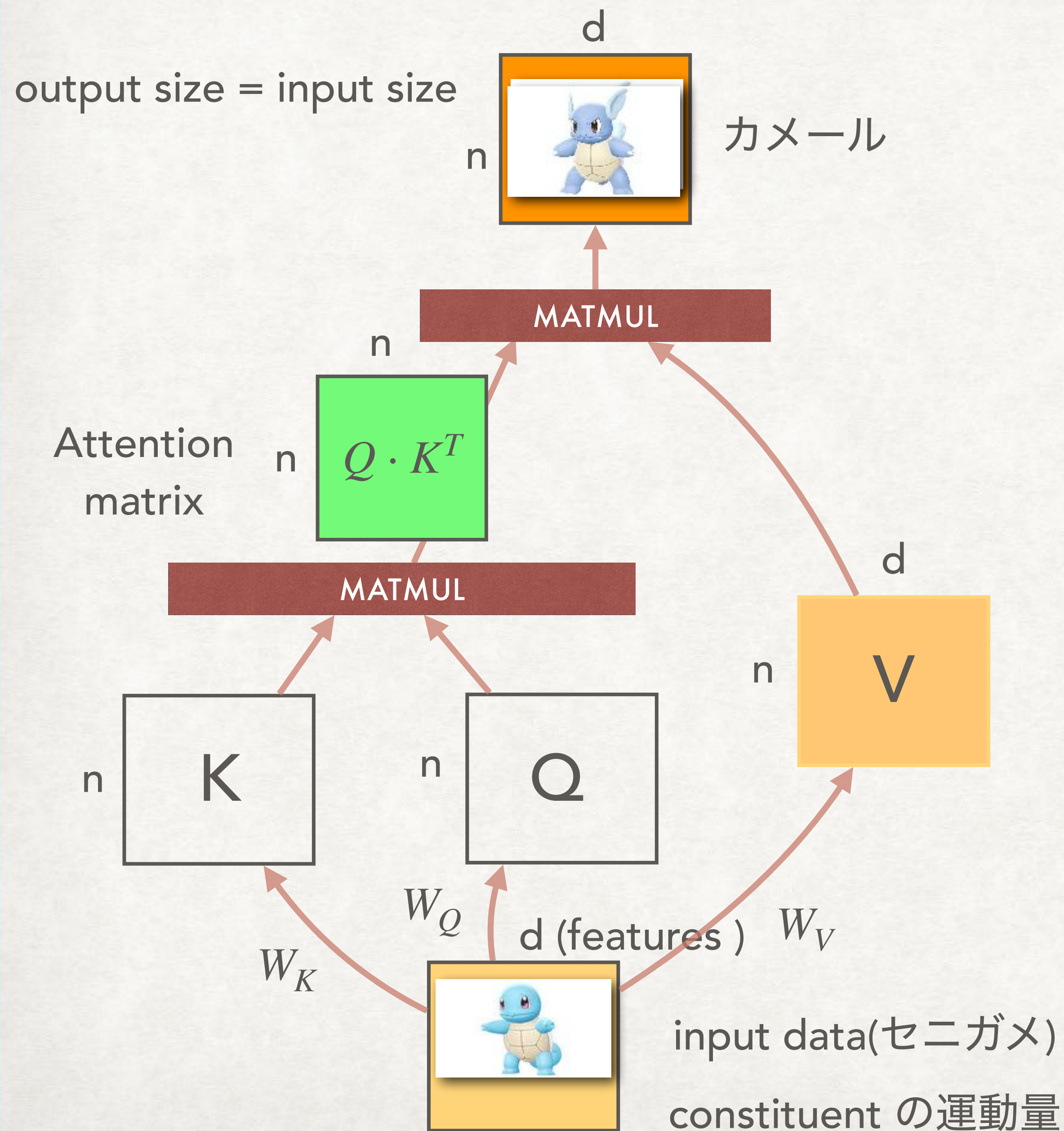
Inputs $p_i \rightarrow p_i \cdot p_j$ permutation invariance $F(\pi \circ T_{ijk\dots}) = \pi \circ F(T_{ijk\dots})$, Use equivariant layer $T'_{ij} = \sum_{k,l=1}^N B_{ijkl}^a T_{kl}$

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)

“TRANSFORMER” :SELF ATTENTION LAYERS



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

- 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_{exp} is restricted to use only the 5-class experimentally realistic particle identification information, while QG_{full} uses the full set of particle identification information in the dataset and further distinguish between different types of charged hadrons and neutral hadrons.

Category	Variable	Definition	JETCLASS	TOP	QG _{exp}	QG _{full}
運動量	$\Delta\eta$	difference in pseudorapidity η between the particle and the jet axis	✓	✓	✓	✓
	$\Delta\phi$	difference in azimuthal angle ϕ between the particle and the jet axis	✓	✓	✓	✓
	$\log p_{\text{T}}$	logarithm of the particle's transverse momentum p_{T}	✓	✓	✓	✓
	$\log E$	logarithm of the particle's energy	✓	✓	✓	✓
	$\log \frac{p_{\text{T}}}{p_{\text{T}}(\text{jet})}$	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	✓	✓	✓	✓
	ΔR	angular separation between the particle and the jet axis ($\sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$)	✓	✓	✓	✓
電荷、particle ID など	charge	electric charge of the particle	✓	—	✓	✓
	Electron	if the particle is an electron (<code> pid ==11</code>)	✓	—	✓	✓
	Muon	if the particle is an muon (<code> pid ==13</code>)	✓	—	✓	✓
	Photon	if the particle is an photon (<code>pid==22</code>)	✓	—	✓	✓
	CH	if the particle is an charged hadron (<code> pid ==211 or 321 or 2212</code>)	✓	—	✓	✓ ^a
	NH	if the particle is an neutral hadron (<code> pid ==130 or 2112 or 0</code>)	✓	—	✓	✓ ^b
衝突点から飛んだ距離	$\tanh d_0$	hyperbolic tangent of the transverse impact parameter value	✓	—	—	—
	$\tanh d_z$	hyperbolic tangent of the longitudinal impact parameter value	✓	—	—	—
	σ_{d_0}	error of the measured transverse impact parameter	✓	—	—	—
	σ_{d_z}	error of the measured longitudinal impact parameter	✓	—	—	—

^a (`|pid|==211`) + (`|pid|==321`)*0.5 + (`|pid|==2212`)*0.2

^b (`|pid|==130`) + (`|pid|==2112`)*0.2

自分のやってることもはなしたい

高いスケールの理論

「学習物理」での研究

高エネルギー衝突のスケール構造(QCD)と

相性の良いネットワーク

素粒子標準型作り
+新物理

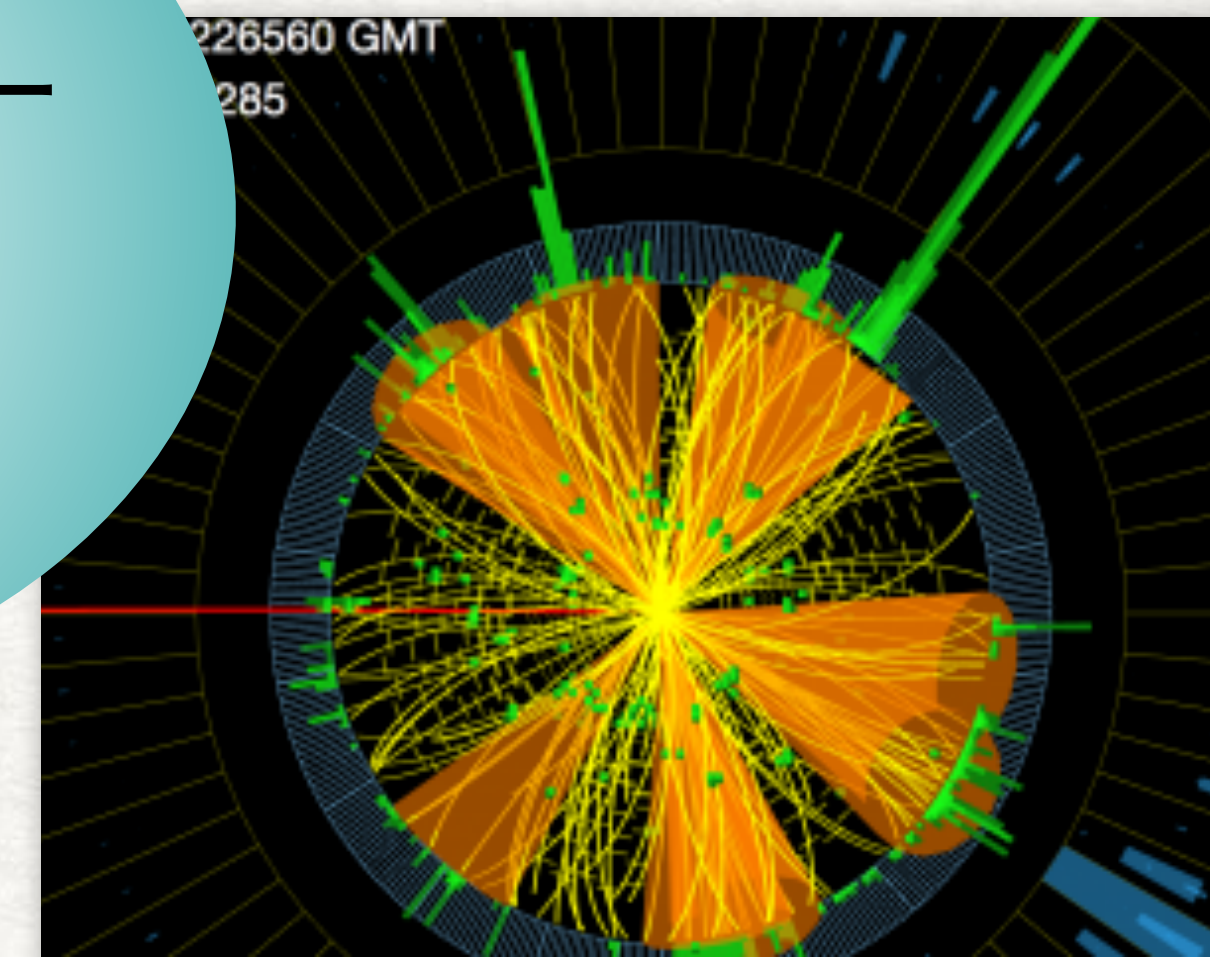
強い相互作用の
高次補正

強い相互作用
の理論

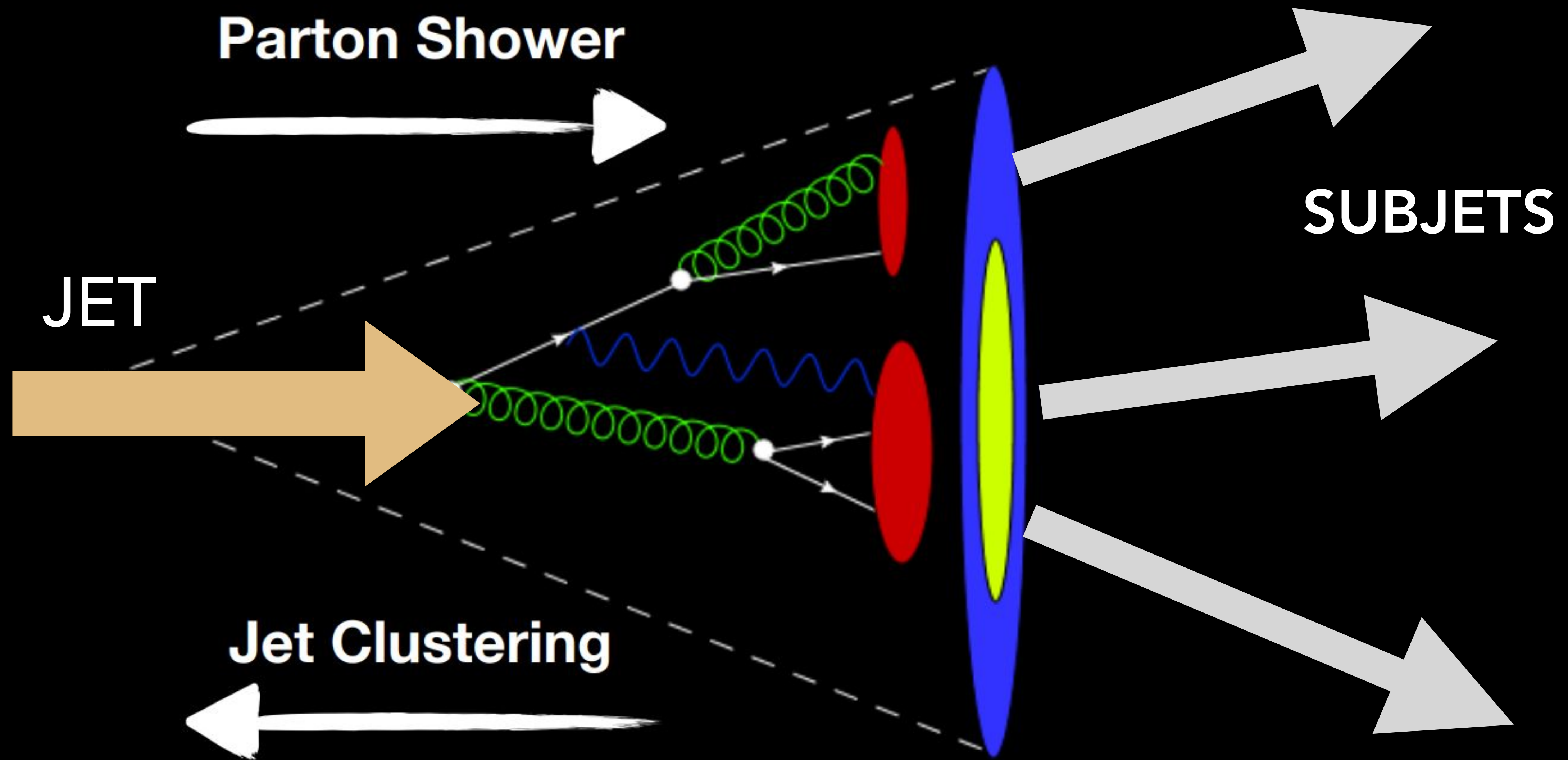
パートンシャワー
ジェット

高次元相関

→事象のモンテカルロ計算



But what is the connection to this picture??



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

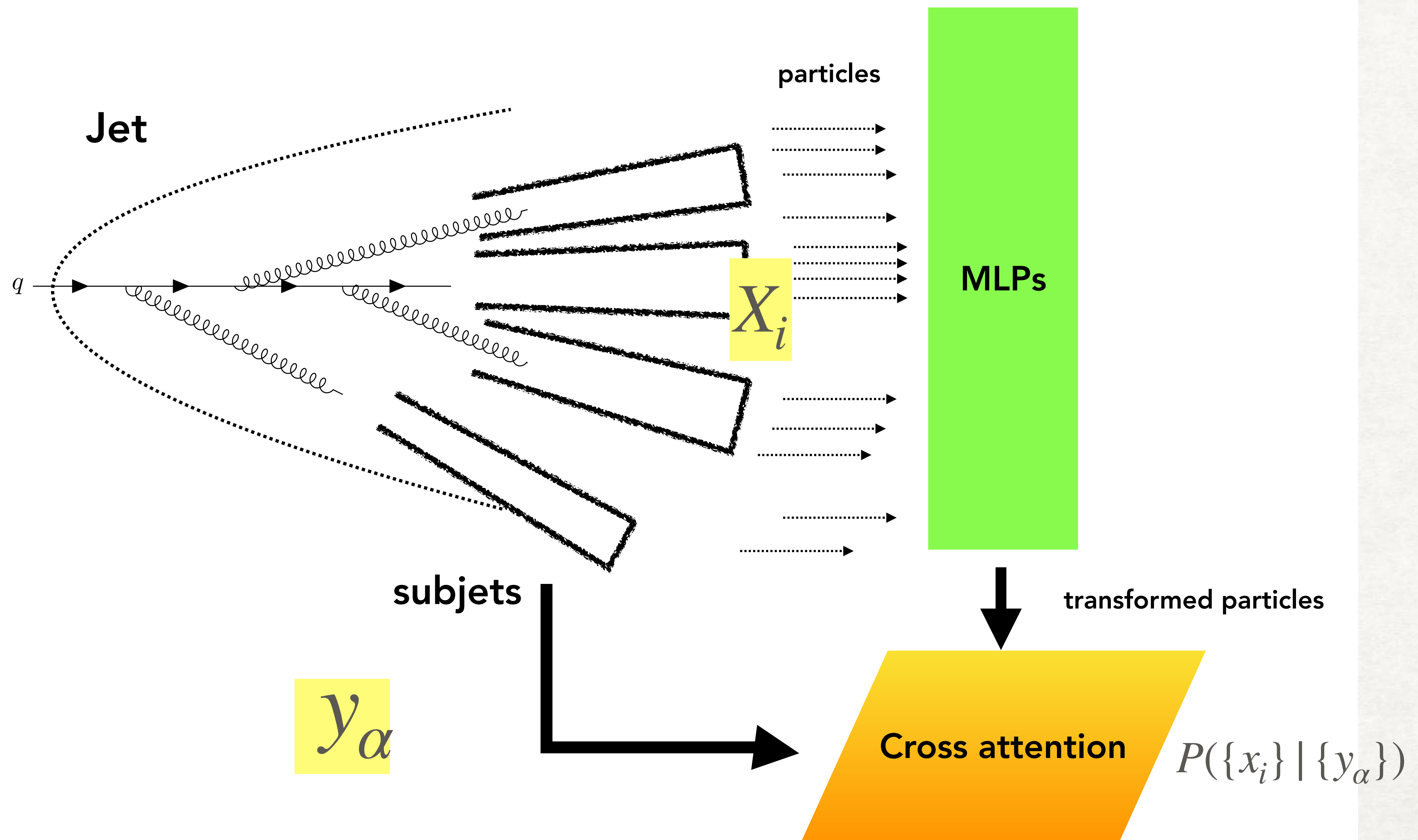
$$P(\text{hadrons in jets} | \text{parton} \sim \text{jet}) = P(\{x_i\} | \{y\})$$

- 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。

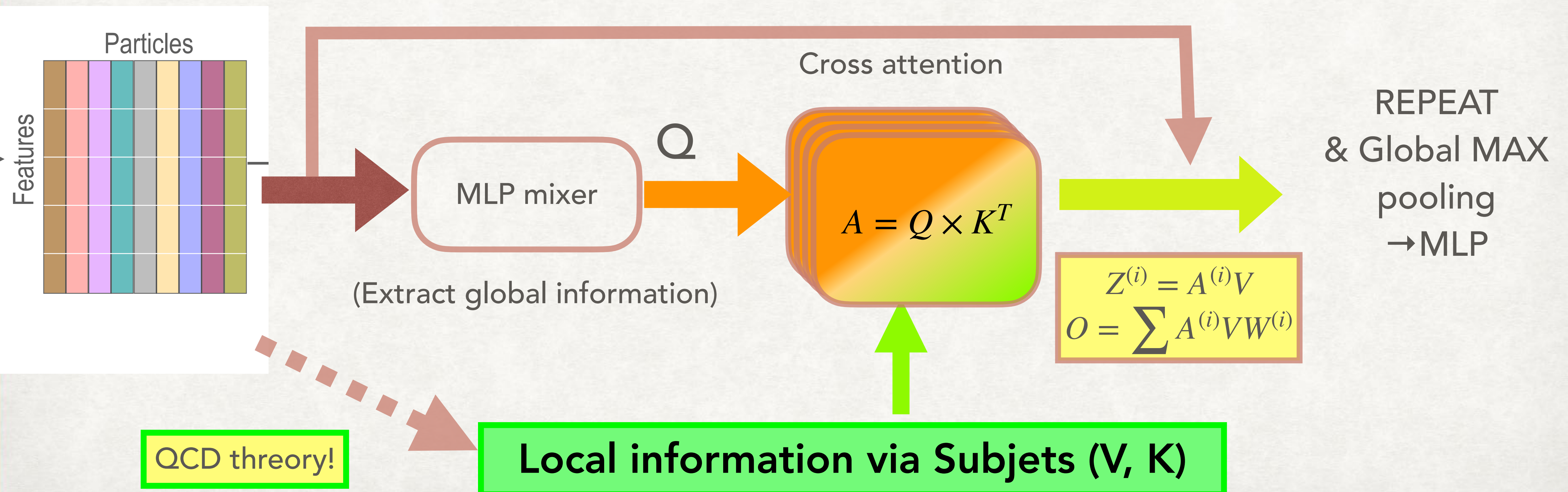


QCDのファクトリゼーション構造をとらえたネットワーク

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

input X

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



CROSS ATTENTION LAYERS

output



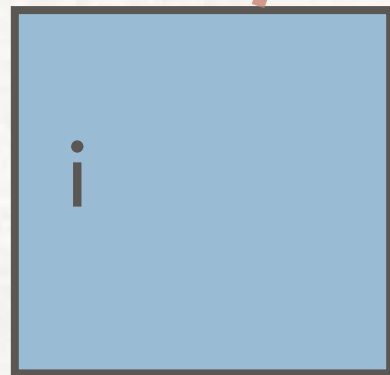
f

i



ヤドキング

n



cross
attention
matrix

f

i

Q

n

K

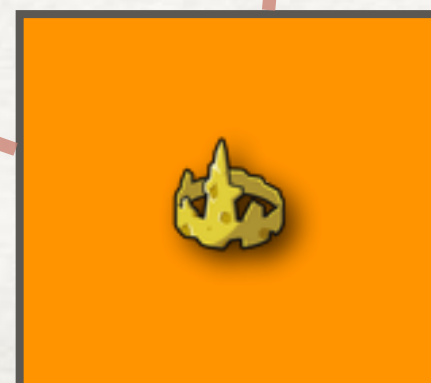
f

jet
constituents

やどん

W_K

n



subj
kinematics

おうじゃのしるし

W_V

j

- restrict network to cross attention (subjct) x (jet constituent) $A=QK^T$
- jet constituent Q
- subjct : parton shower, hadronization K V
- サブジェットの運動量(パートンシャワーや重たい粒子のスケールの物理を反映) が、対応するジェット内部に粒子(1GeVのスケール) に意味づけを与える構造を持っている.

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

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

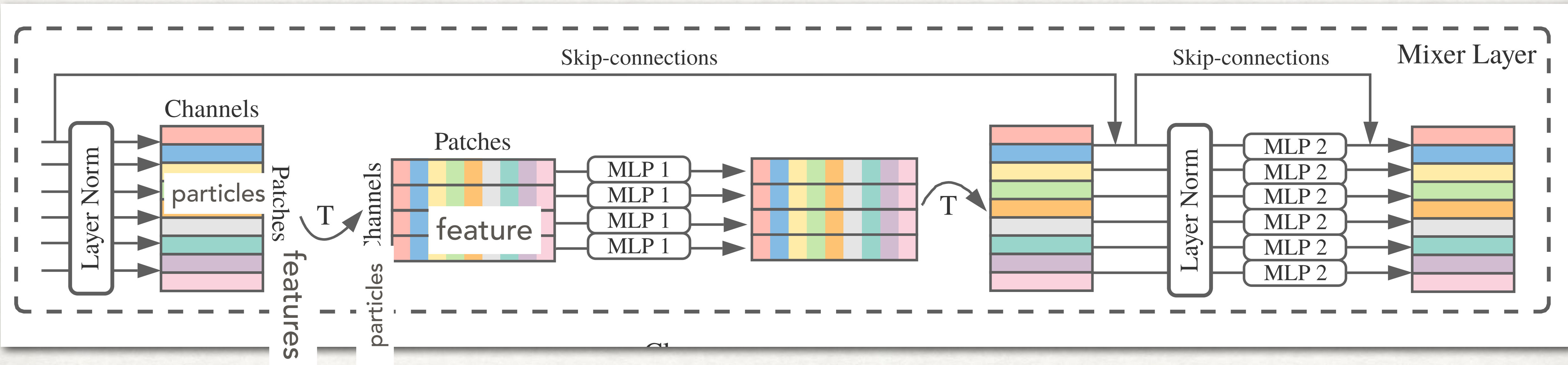
全てのデータをまとめてself attention するとこの構造は実現しない

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

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、粒子をまぜる。

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

HIGH PERFORMANCE WITHOUT
USING LORENTS INVARIANCE

*Subjet cone size $R=0.3$

*HDBSCAN is algorithm without distance measure

APPLICATION 2

TOWARD GLOBAL EVENT ANALYSIS

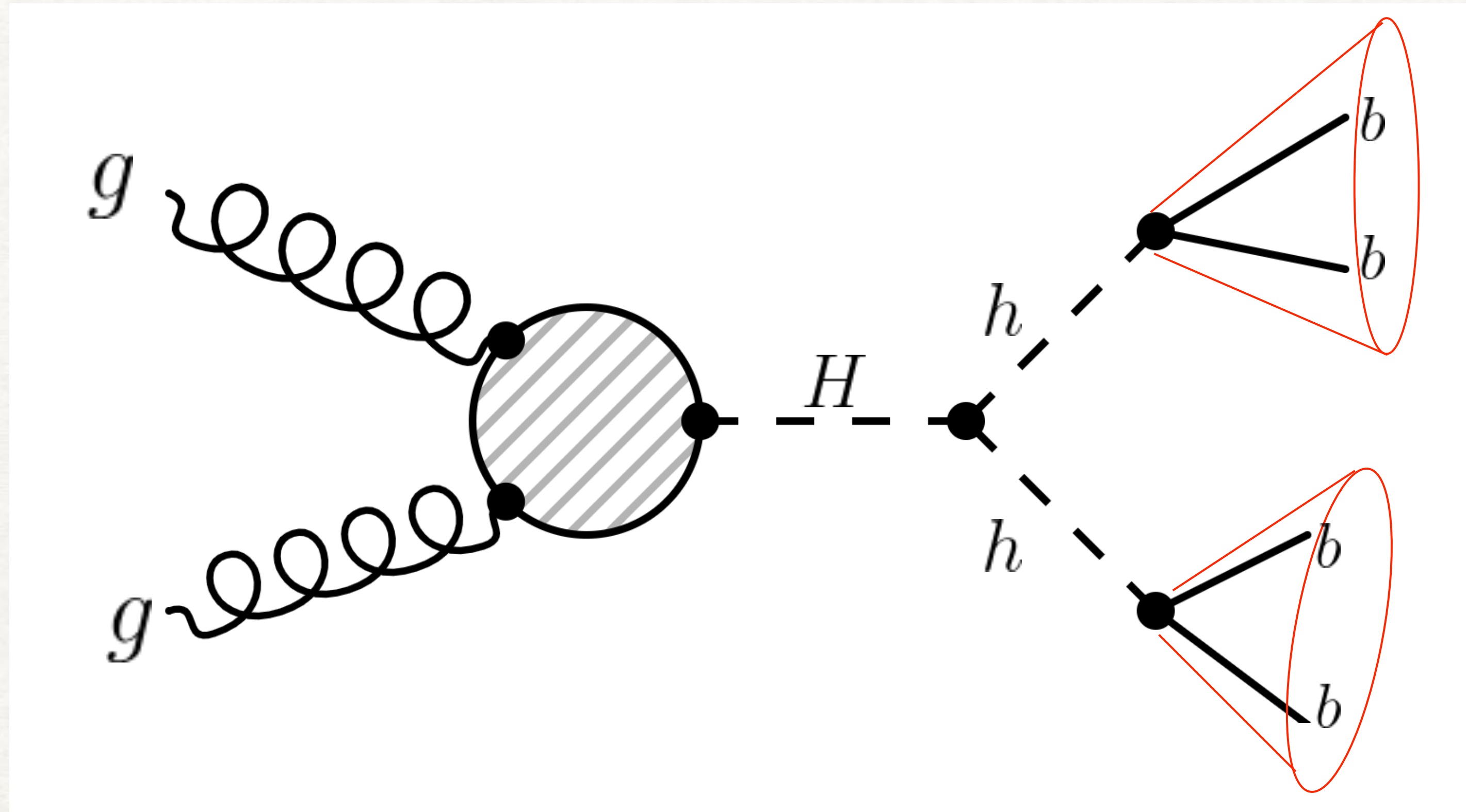
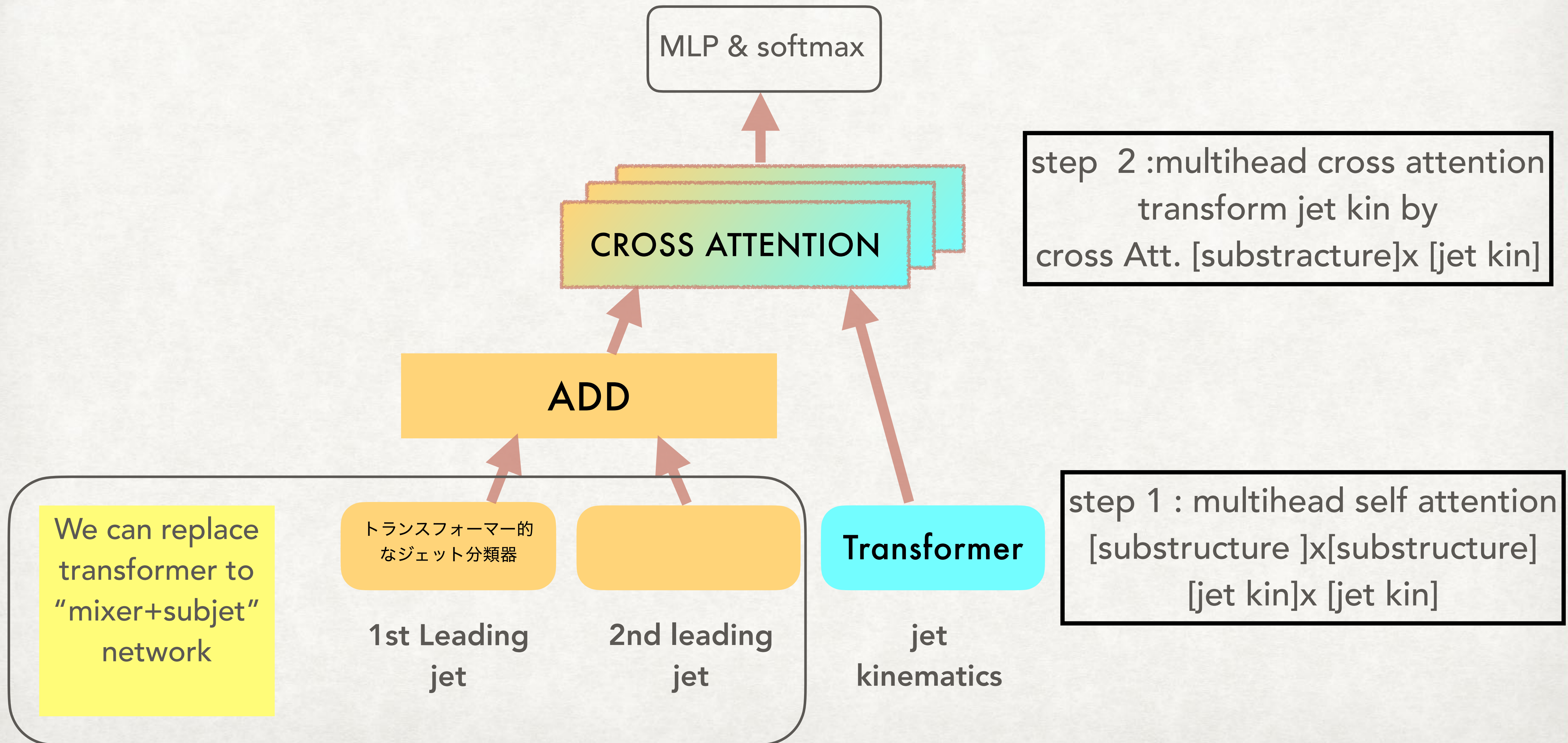


Figure 2: Feynman diagram for the signal process.

cross attention motion for 2 fatjet events



INPUT TO NETWORK : EVENT KINEMATICS

測定器の円筒座標

Kinematical inputs (3, 6)

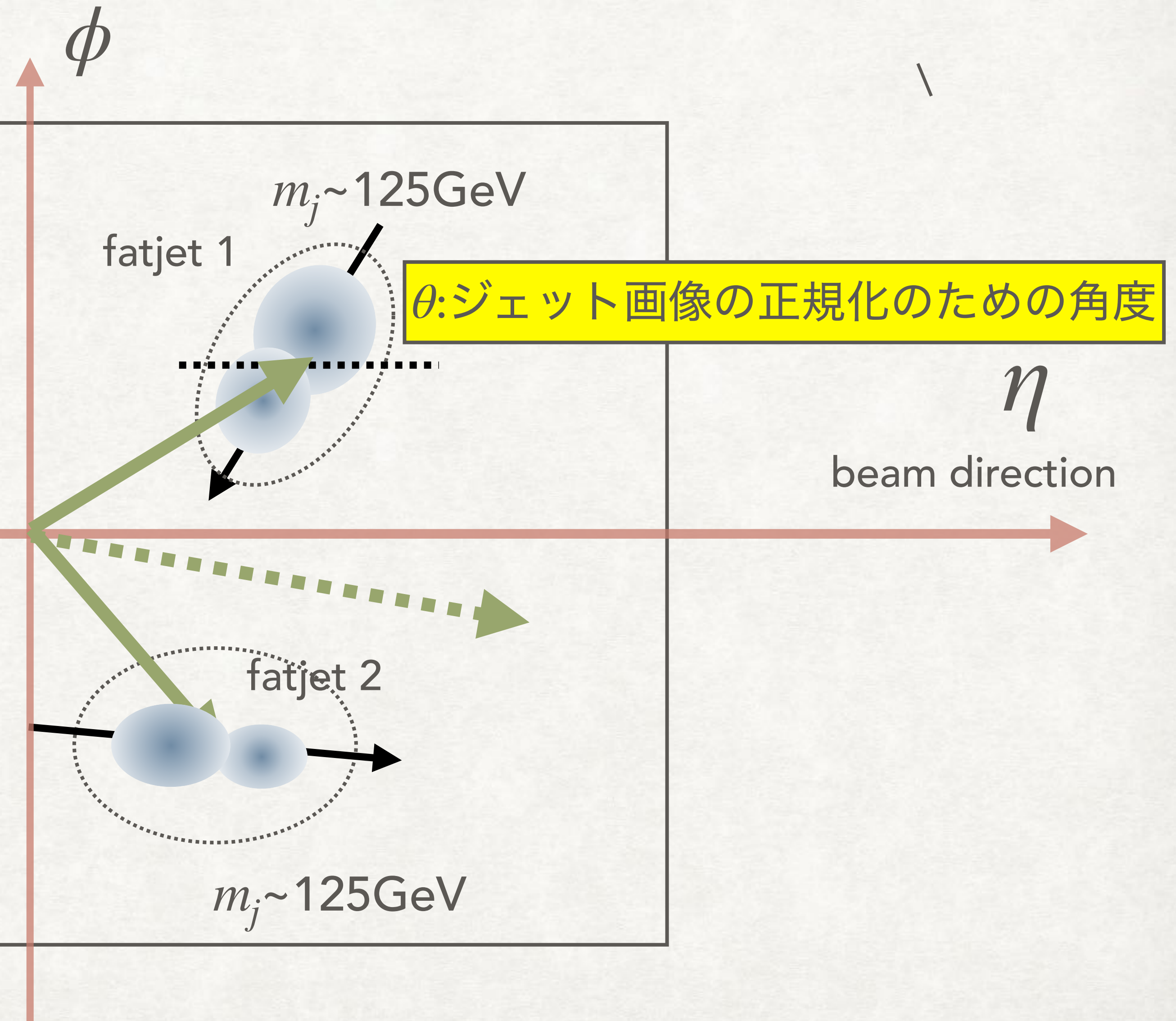
fatjet 1 = $(m_1, \eta_1, \phi_1, p_{T1}, E_1), \theta_1$

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 :

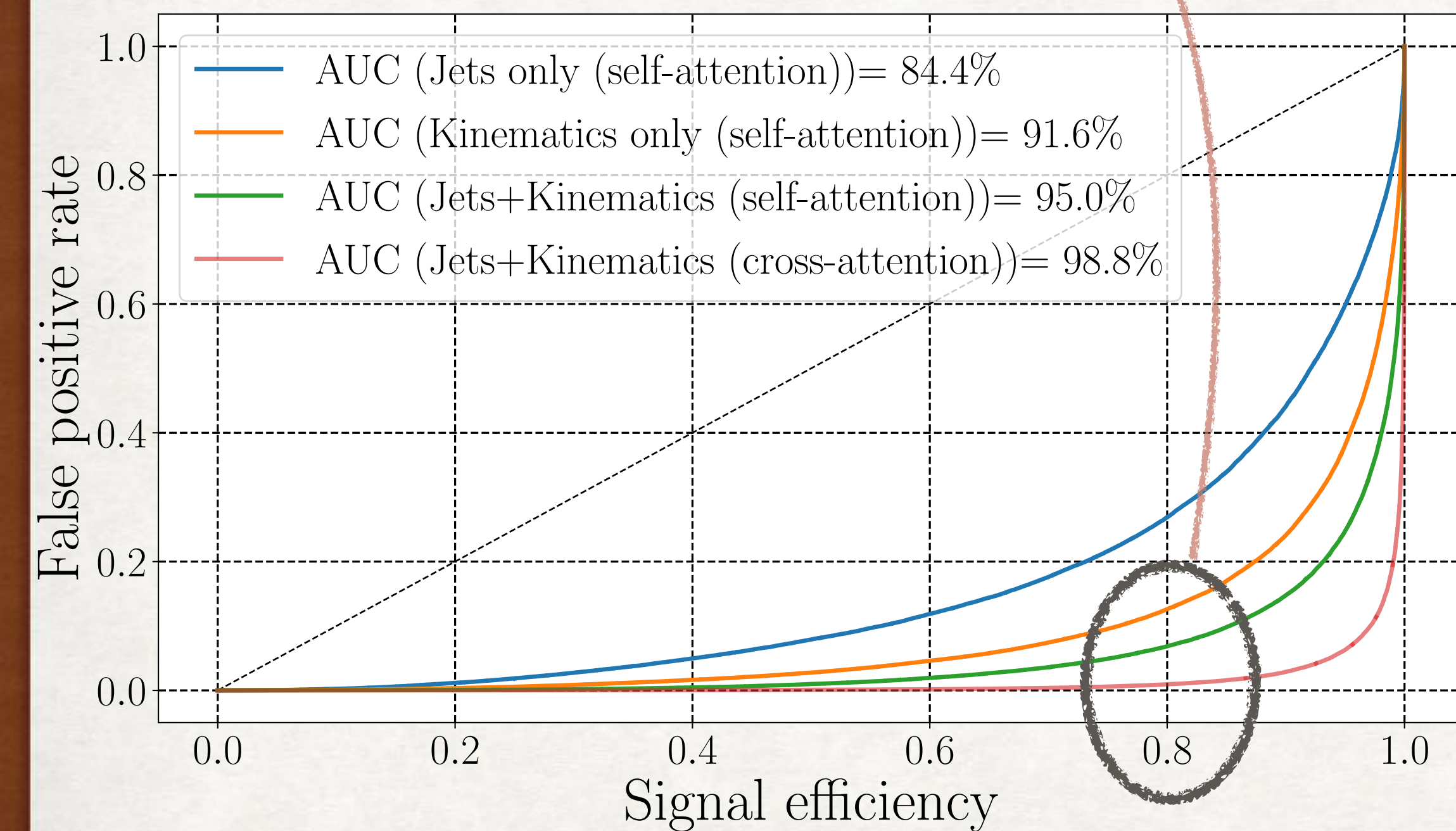
1. "5 inputs for 4 momentum" ,
2. H candidate momentum as sum of the fat jet momentum.
3. add " θ " :the correlation beyond a subjet



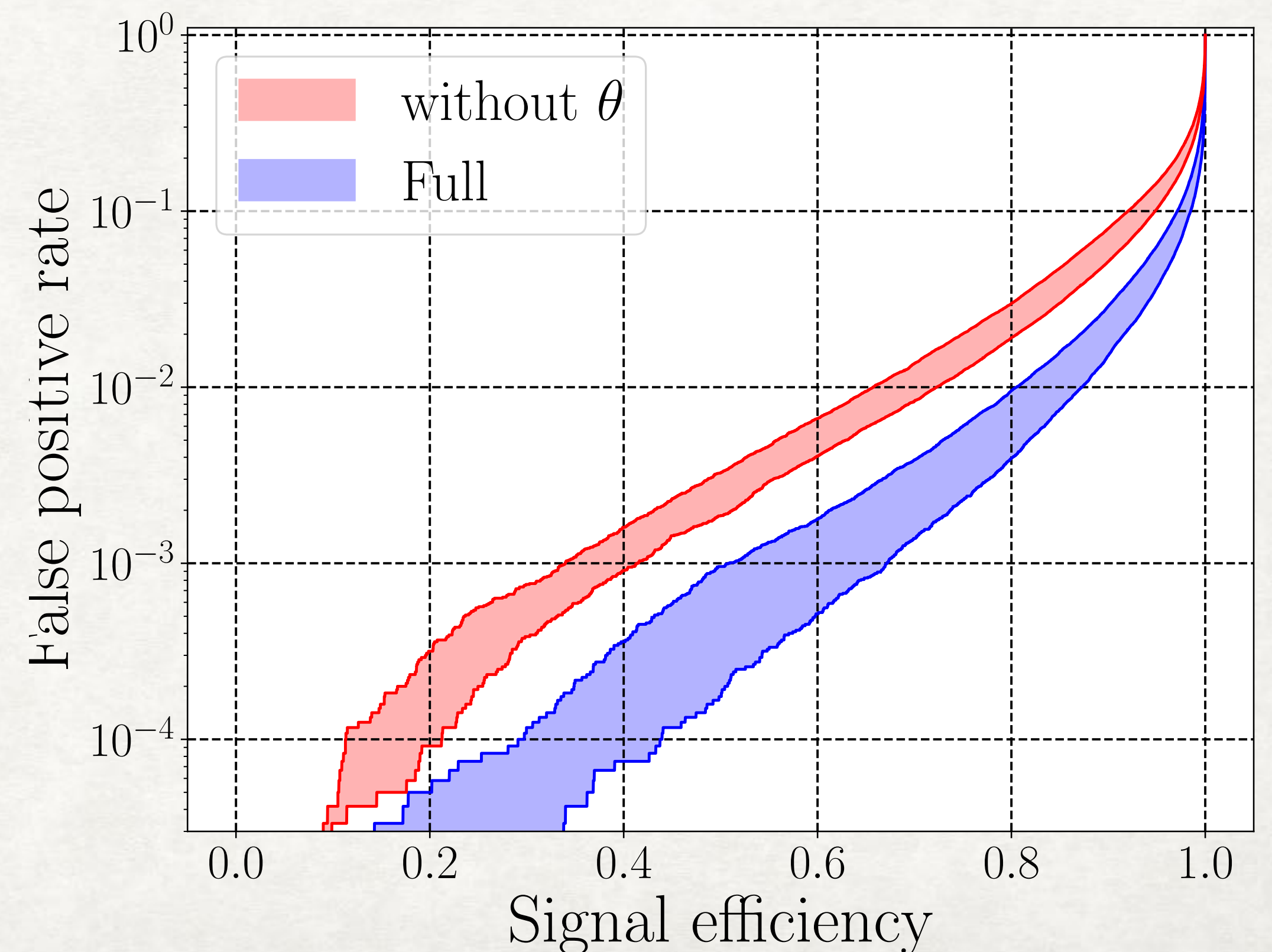
IMPROVEMENT USING CROSS ATTENTION

factor 5 improvement at the same acceptance.

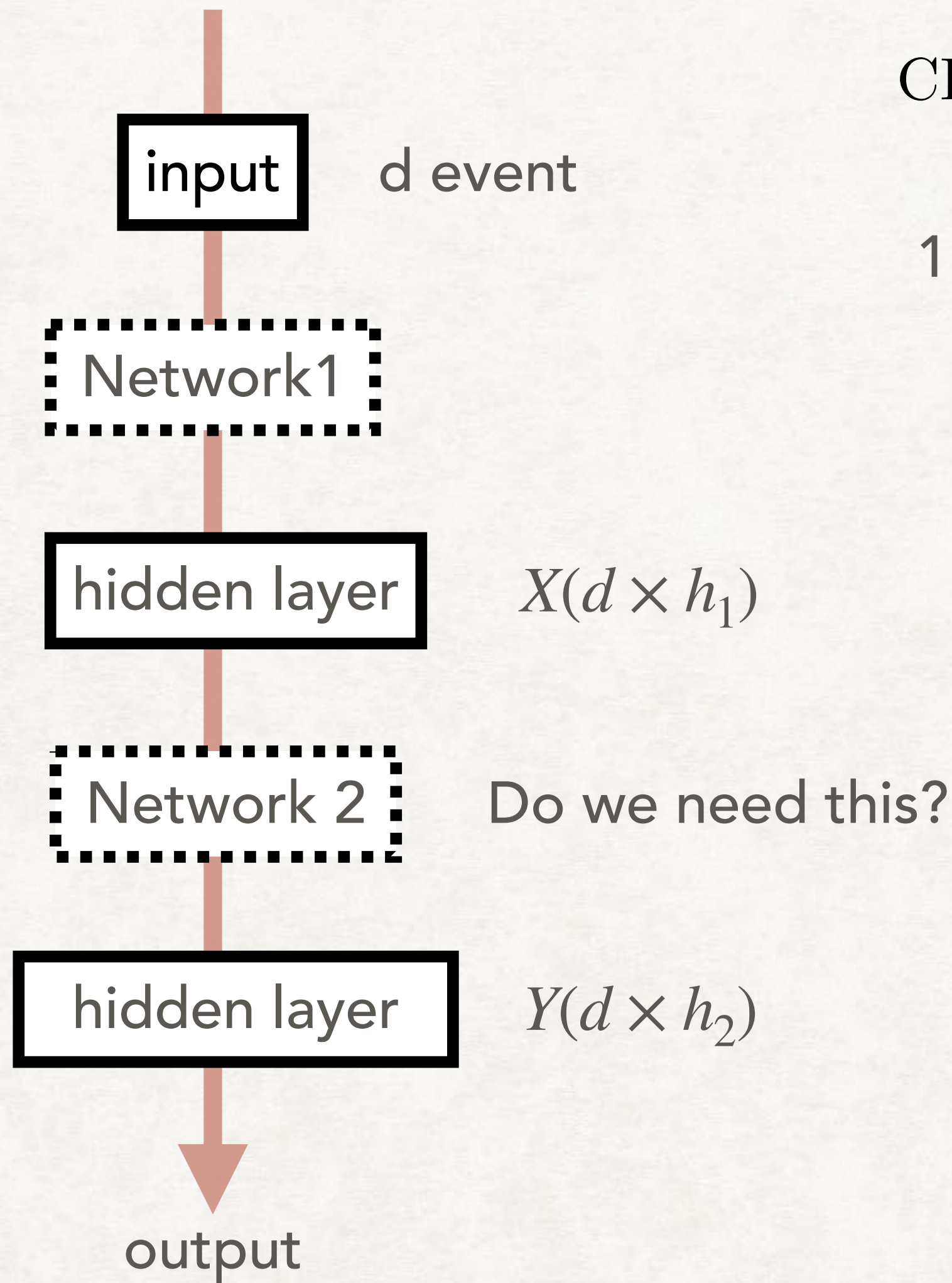
Decay correlation is important
(because QCD background is correlated)



Cross attention improves the rejection efficiency significantly



色々な解釈 CKA 類似性を例に



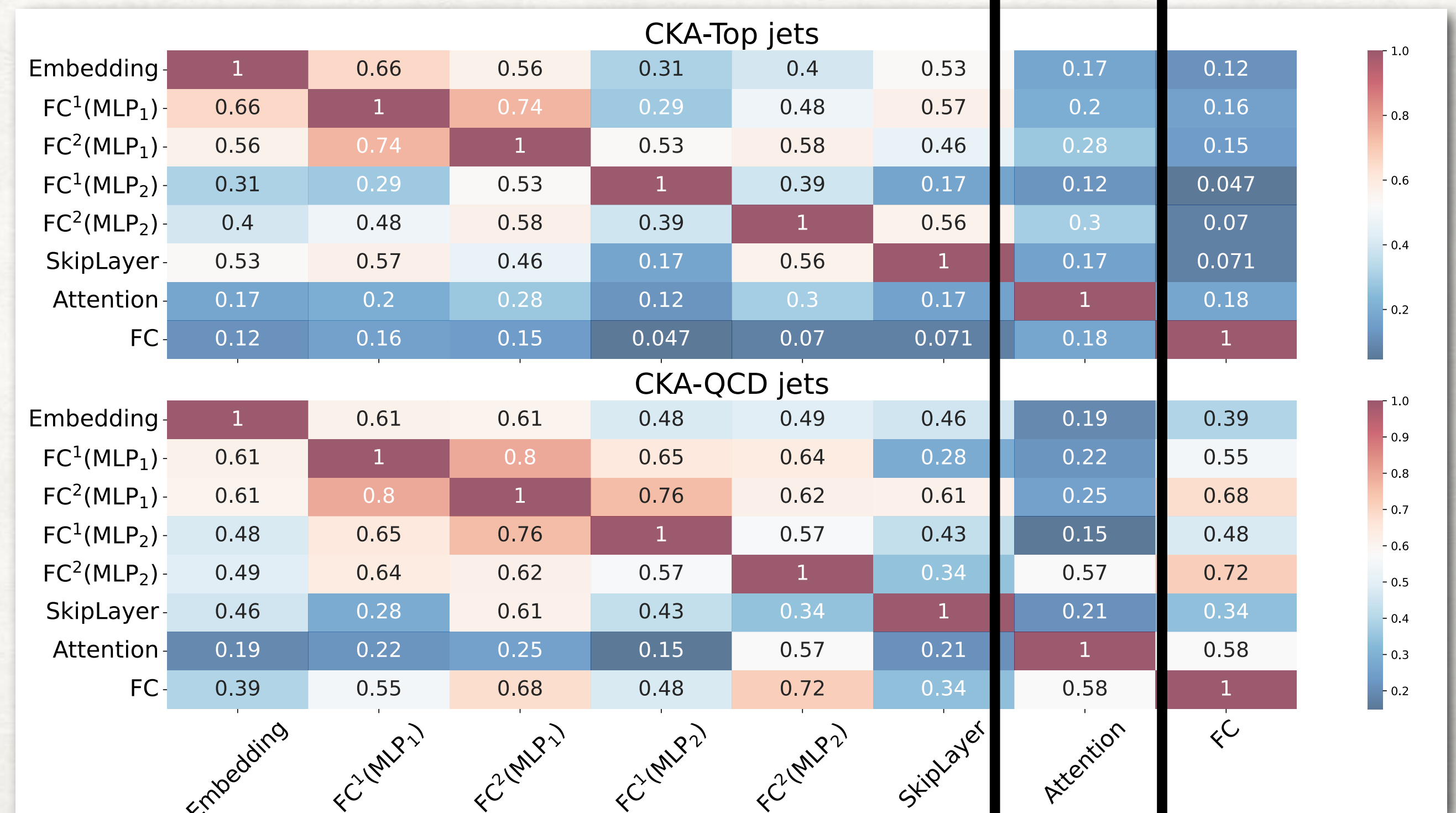
$$\text{CKA}(M,N) = \frac{\text{HSIC}(M,N)}{\sqrt{\text{HSIC}(M,M)\text{HSIC}(N,N)}},$$

1 if they are same (no improvement)

$$\text{HSIC}(M,N) = \frac{1}{(d-1)^2} \text{Tr}(MHNH)$$

$$H = \delta_{ij} - \frac{1}{d}$$

Efficient!



$M = XX^T, N = YY^T$ は
入力情報数x 入力情報数の固定サイズ

3. これからの理論の課題

実験は頑張る。だから理論も頑張れ。

データ解析には、理論分布の予言が重要

高いスケールの理論

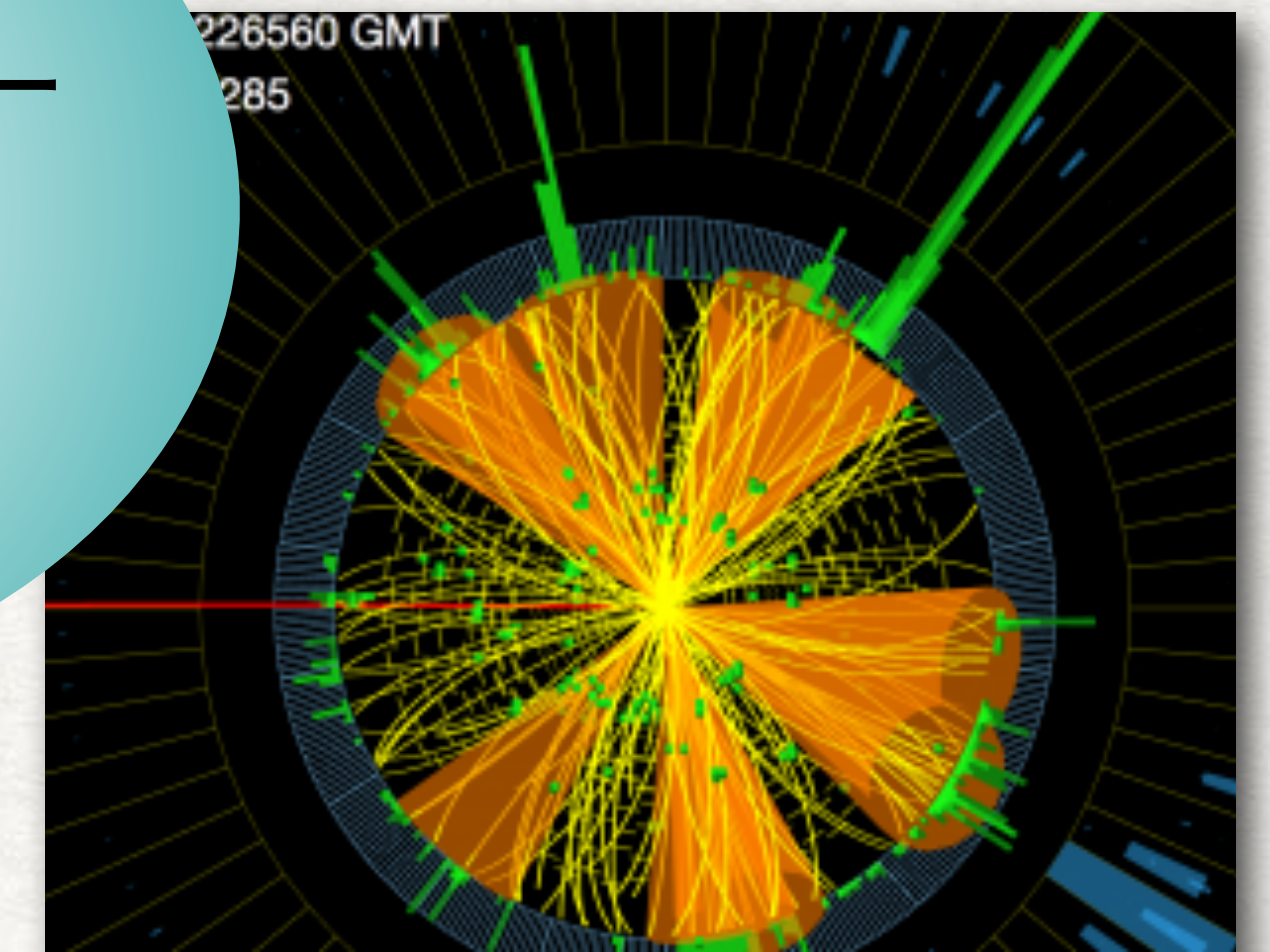
素粒子標準型作り
+新物理

強い相互作用の
高次補正

強い相互作用
の理論

パートンシャワー
ジェット

QCD計算の精度が上がるほど、
実験と理論予言の精度がよくなり
その結果として、
探索感度が上がってきた歴史がある



高エネルギー物理の理論における進展

- ❖ QCDのループ補正を高次までやる。QCD のNNN次数 (α_s^3) が目標。

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

- ❖ 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。

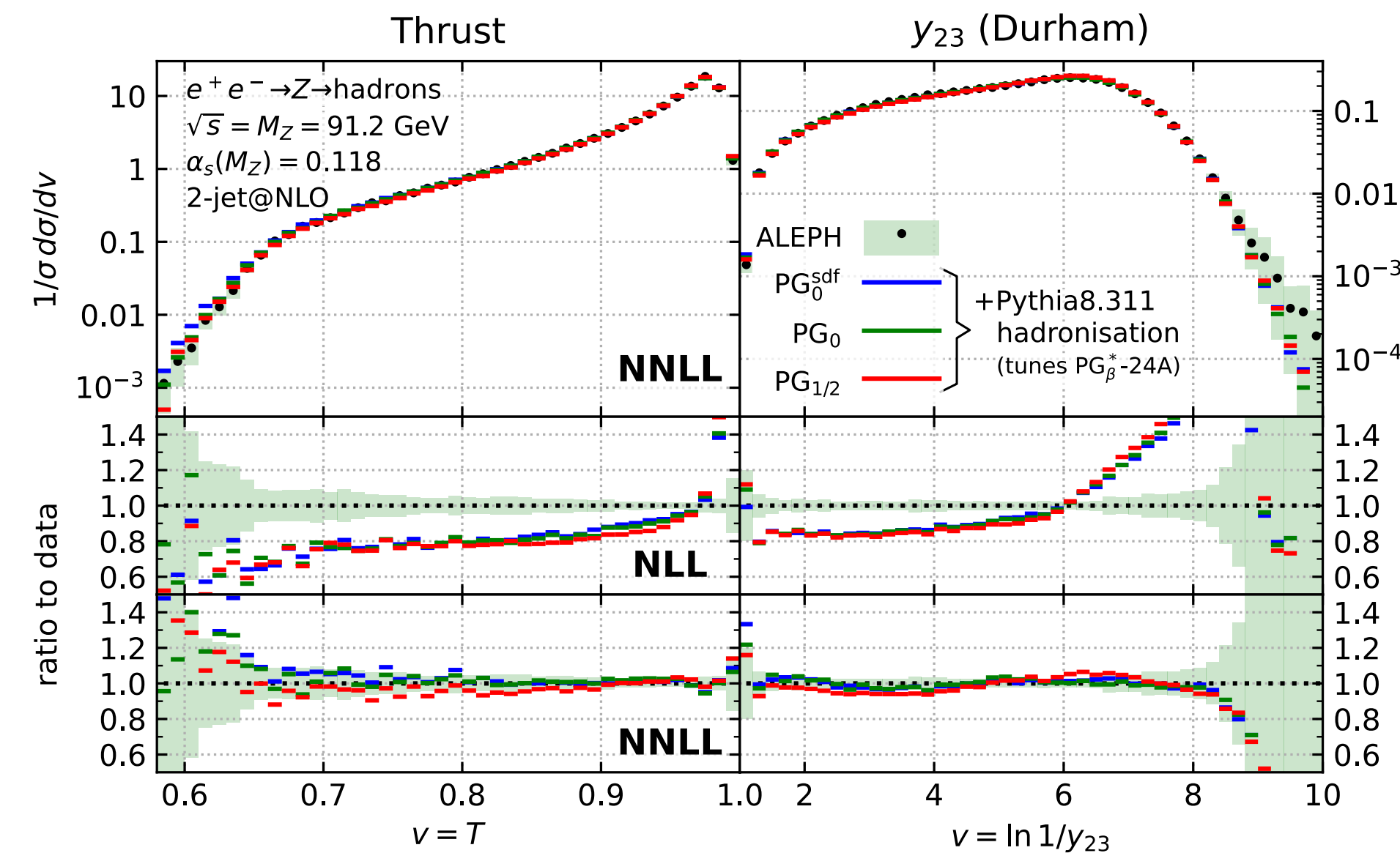
高性能計算(HPC)の進展

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)

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 α_s^{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 α_s と
 摂動計算の α_s が初めて一致

LUND PLANEの測定 理論との比較 (ATLAS 2004.03540)

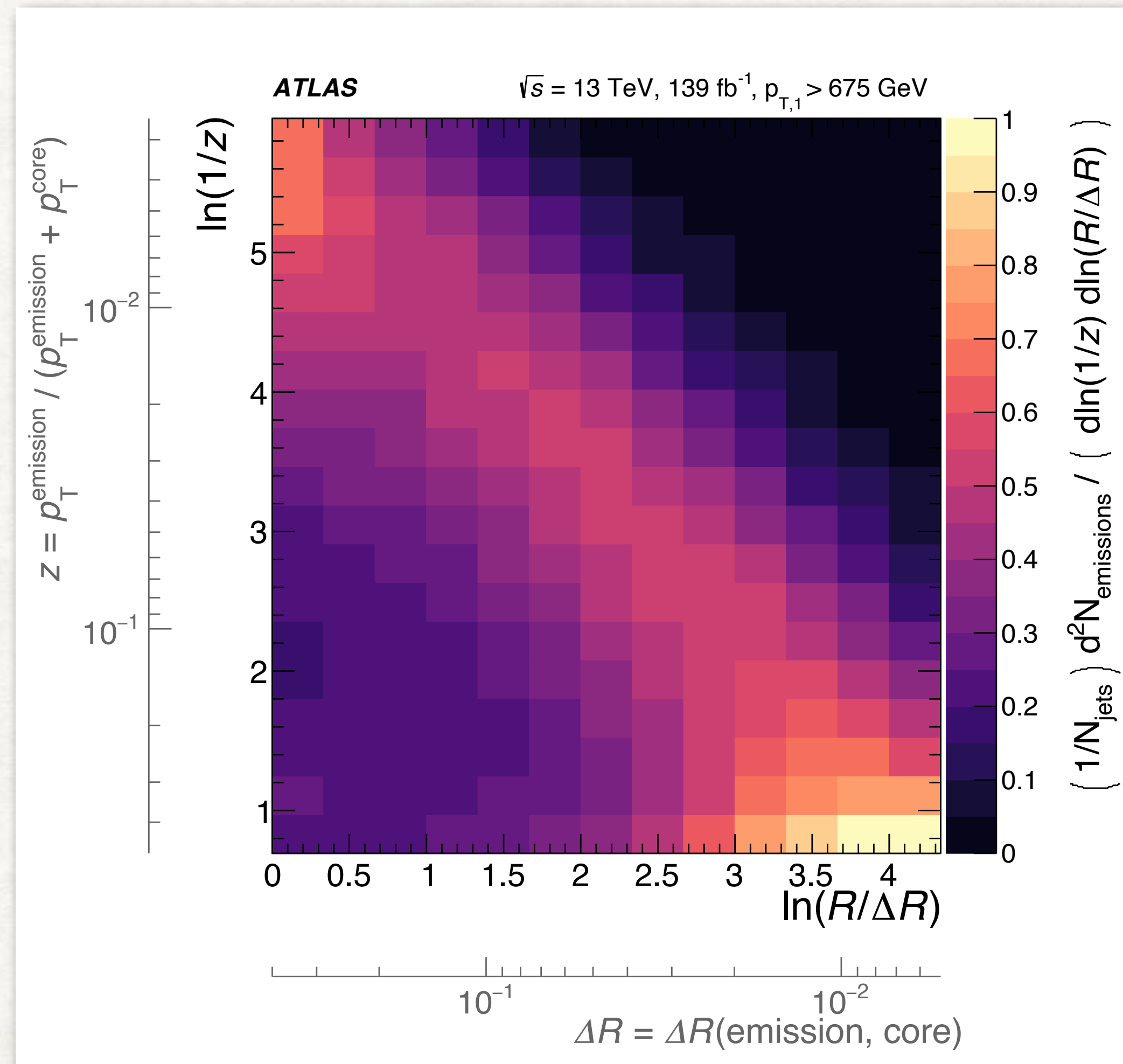
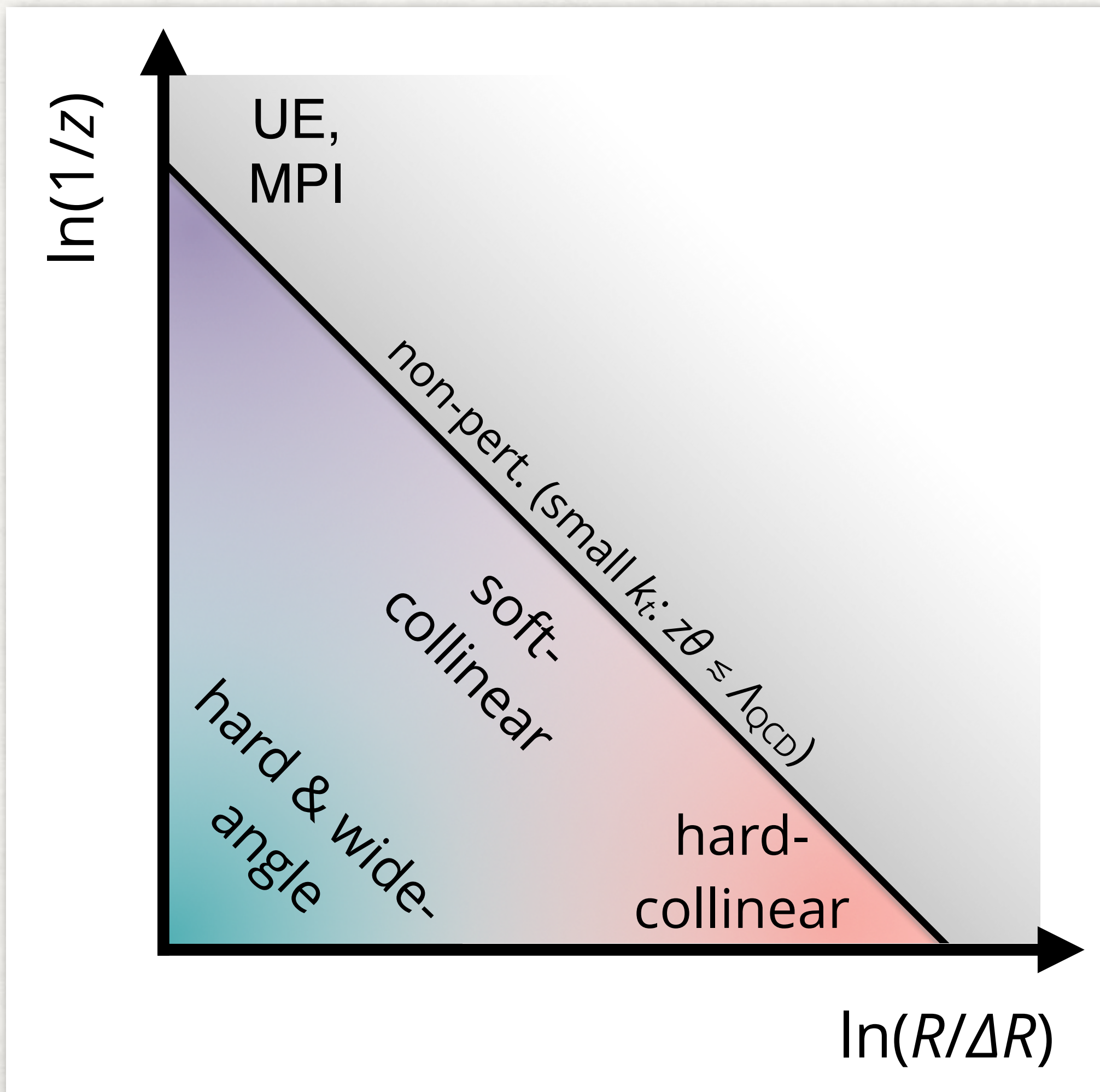


Figure 2: The LJP measured using jets in 13 TeV pp collision data, corrected to particle level. The inner set of axes indicates the coordinates of the LJP itself, while the outer set indicates corresponding values of z and ΔR .

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。