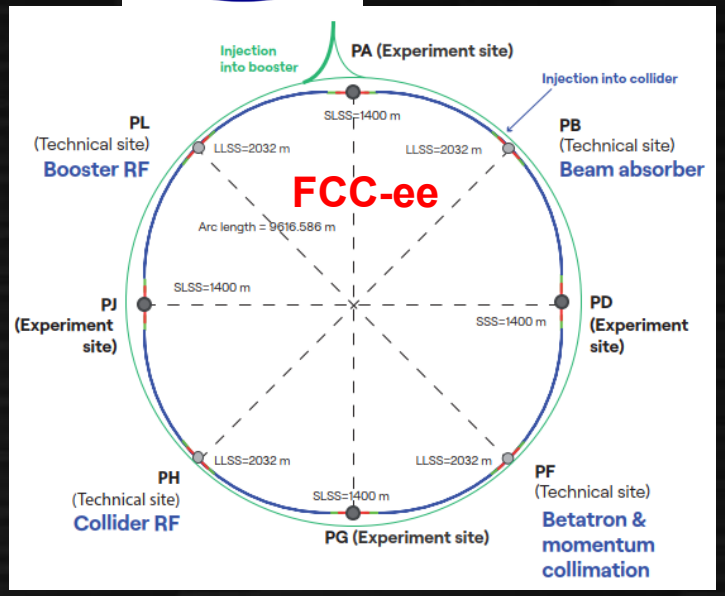
ヒッグスファクトリーの物理解析と深層学習

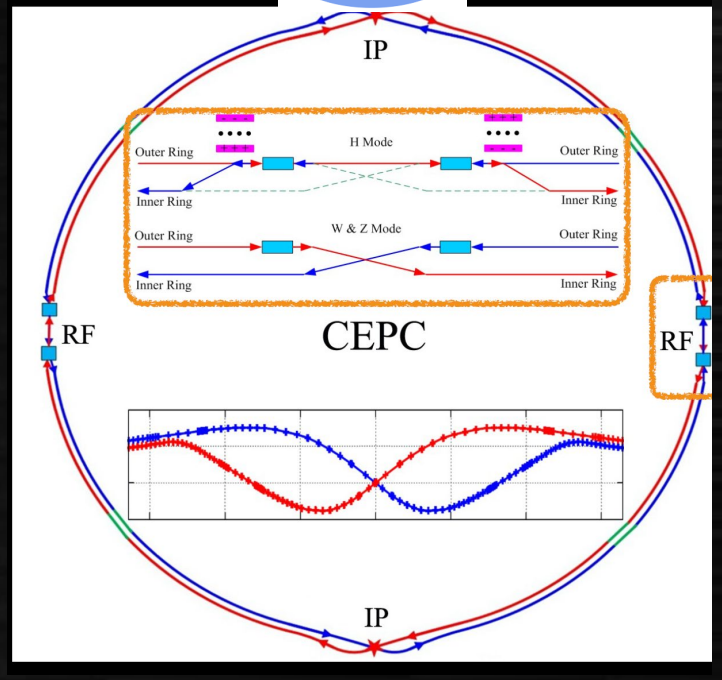
Taikan Suehara
(ICEPP, The University of Tokyo)

(少しだけ)ヒッグスファクトリーの物理

世界のヒッグスファクトリー計画

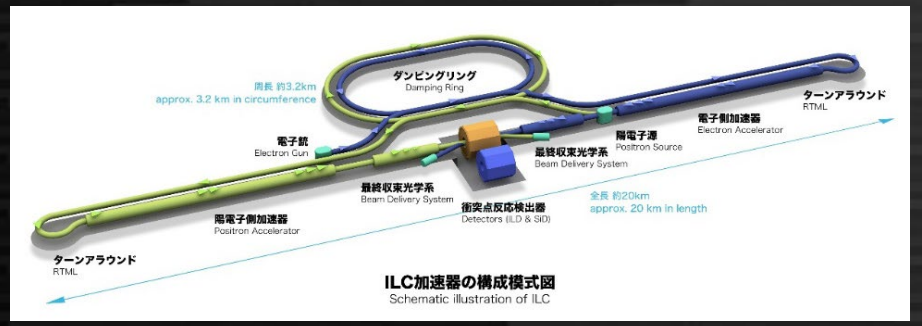


円形のe+e−コライダー計画
90 kmのトンネルで最大
365 GeVの衝突エネルギー
を実現 (~2048)
将来は100 TeVハドロン
コライダーを目指す (~2070)



中国で検討中の100km
円形コライダー計画
FCC-eeと同様の計画だが
2030年代後半の実現を
目指す

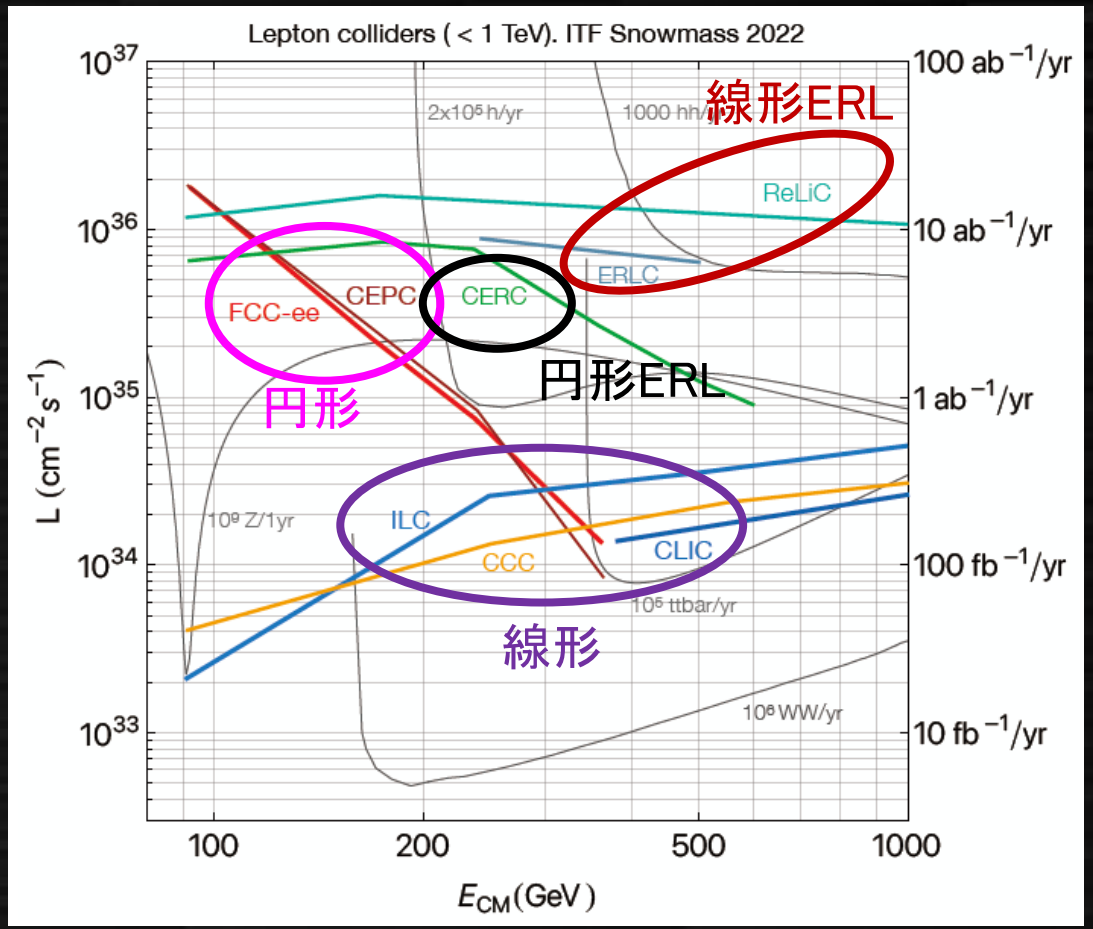
/ LCF



線形ヒッグスファクトリー計画
ILC: 20 kmトンネル, 1IP
LCF@CERN: 30 kmトンネル, 2IP

250 GeVの初期計画から500–550
GeVのアップグレードへ
さらに将来は1 ～ multi TeVへの
アップグレードを目指す

線形コライダーと円形コライダー



- 91–160GeV
  円形コライダーが有利
  （線形ERLでは円形に近いluminosityに到達可）
- 250 GeV
  円形コライダーがやや高いluminosity
  線形コライダーの偏極により物理reachは同等
- 350 GeV
  円形コライダーでも可能だが線形が有利
- ＞ 500 GeV
  線形コライダーでのみ実現可能
  （円形ERLで550 GeVはぎりぎり到達可能）

Note: FCCeeの建設費はILC500と1000の間くらい ppにしないとコスパ的にunreasonable
円形 ＞ 250 GeVは電力消費が非常に大きい（現実的にはERLが必要か）
ERL技術は原理的実証はされているが実装には多くの課題がある

電子・陽電子 Target Energies

91~250 GeV — Oblique parameter決定, W/Z mass, b/τの精密測定

250 GeV — Higgs 結合定数測定 (~1%), Higgs稀崩壊 (軽い新物理)
(TeV新物理間接探索)

350 GeV — Top質量の精密測定 → 真空の安定性

Linear only

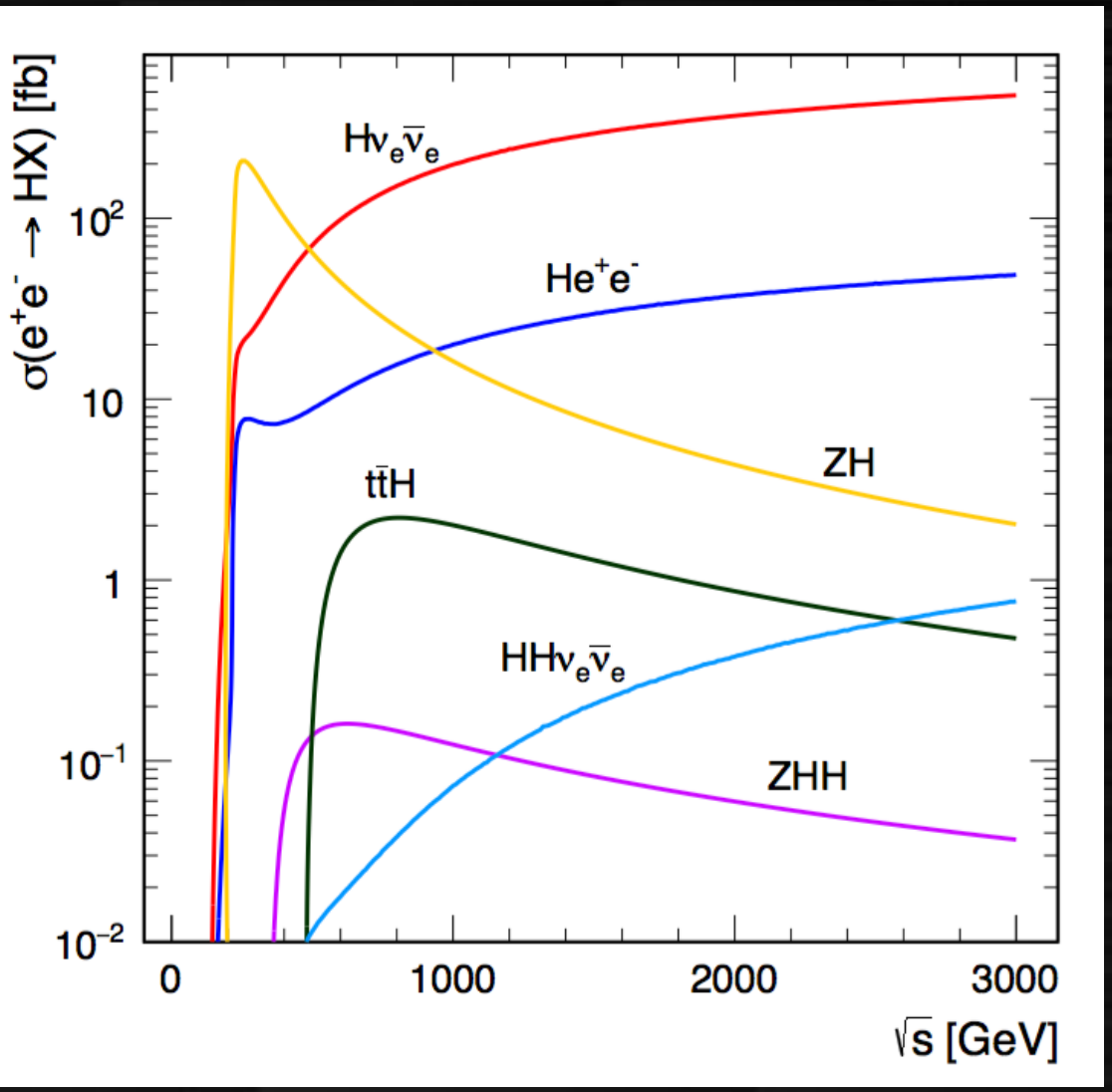
500–550 GeV — Higgs自己結合 (20–30%), ttH結合

1 TeV — Higgs自己結合 (10%) → 物質生成

250 GeV – a few TeV — TeV新物理直接探索

Natural SUSY (250 GeV – 1 TeV)　1 TeV Higgsino　3 TeV Wino

Higgs production and CM energies



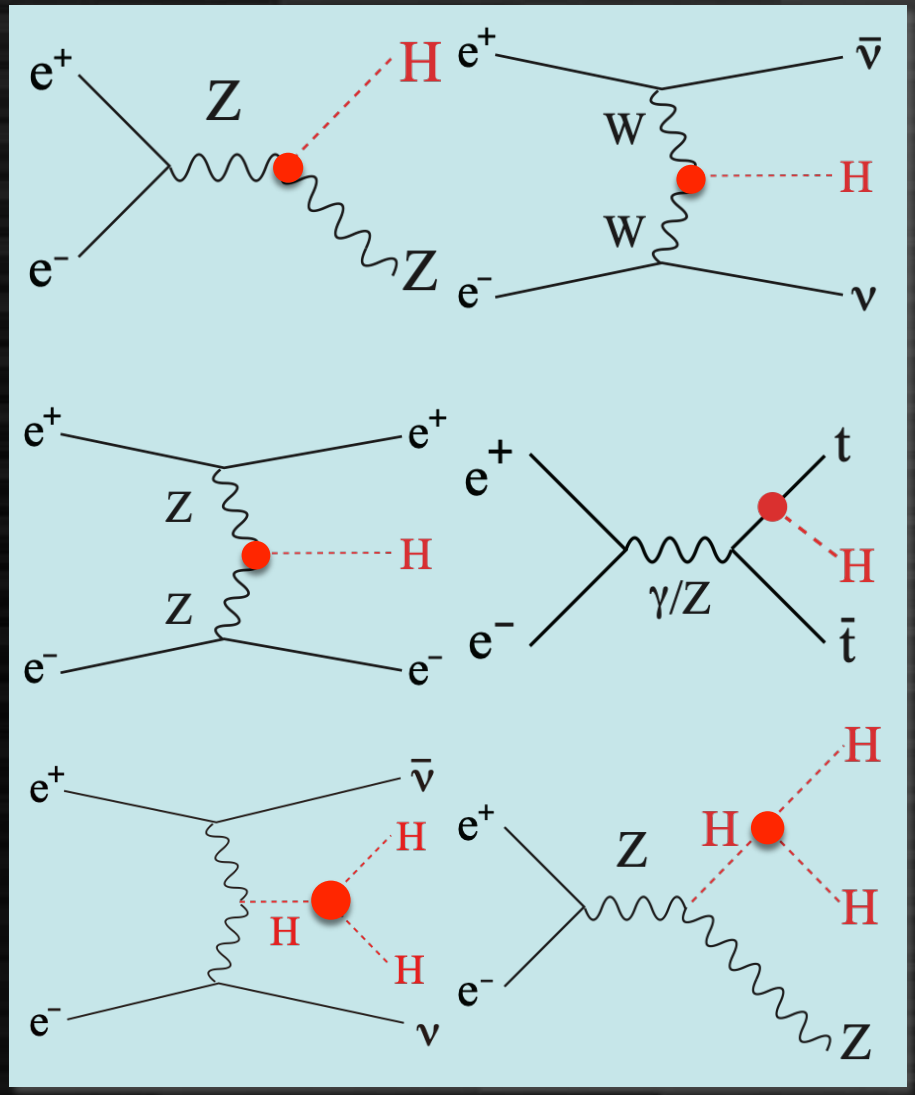
240-250 GeV:
ZH @ ~100 fb

350 GeV
ννH @ ~50 fb

500-600 GeV:
ννH @ ~100 fb
ttH @ ~1 fb
ZHH @ ~0.1 fb

1000 GeV:
ννHH @ ~0.1 fb

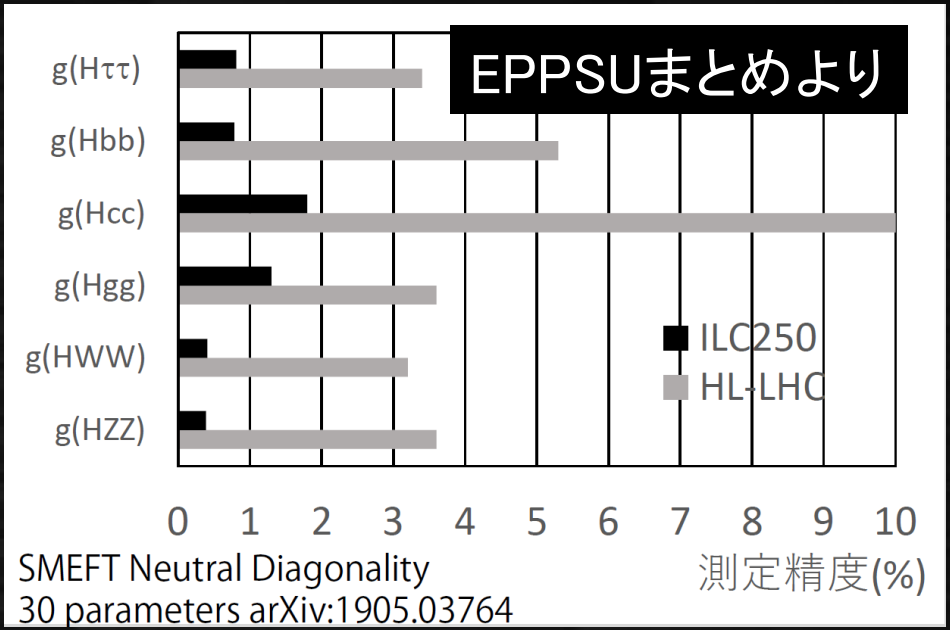
2000-3000 GeV:
ννHH @ ~0.5 fb



Higgs結合の精密測定とBSMの識別

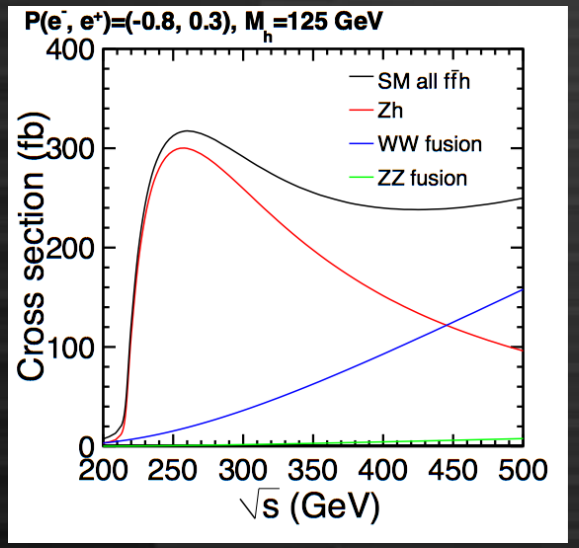
Higgs崩壊モードの分類

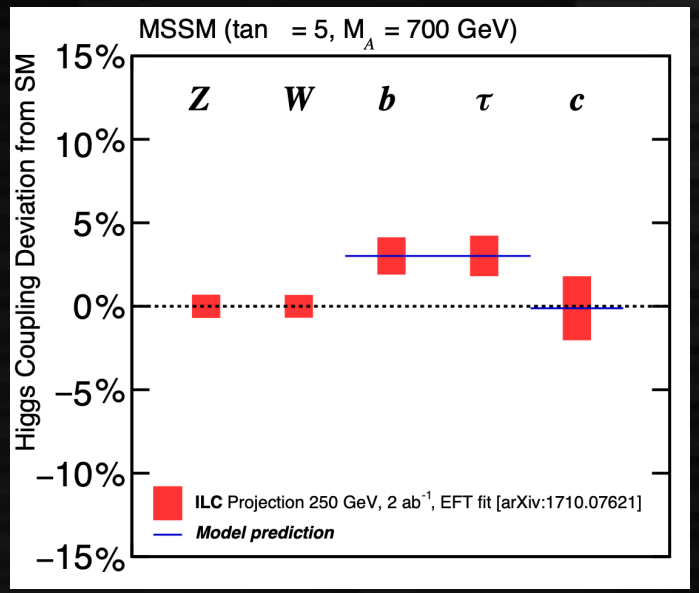
結合の相対的な強さがメインの測定
となるハドロンコライダーに比べて
結合の絶対値の精度が高い

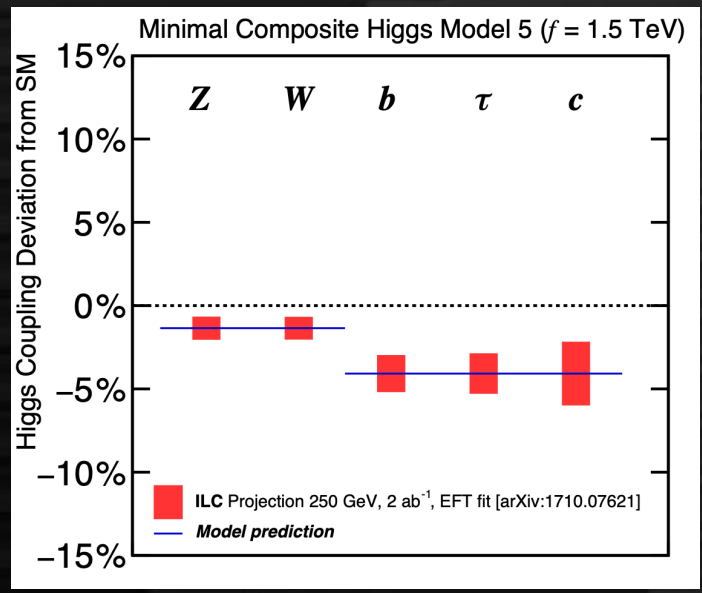


ざっくり一桁の精度向上が期待される

~250 GeVで
ZH生成断面積が最大



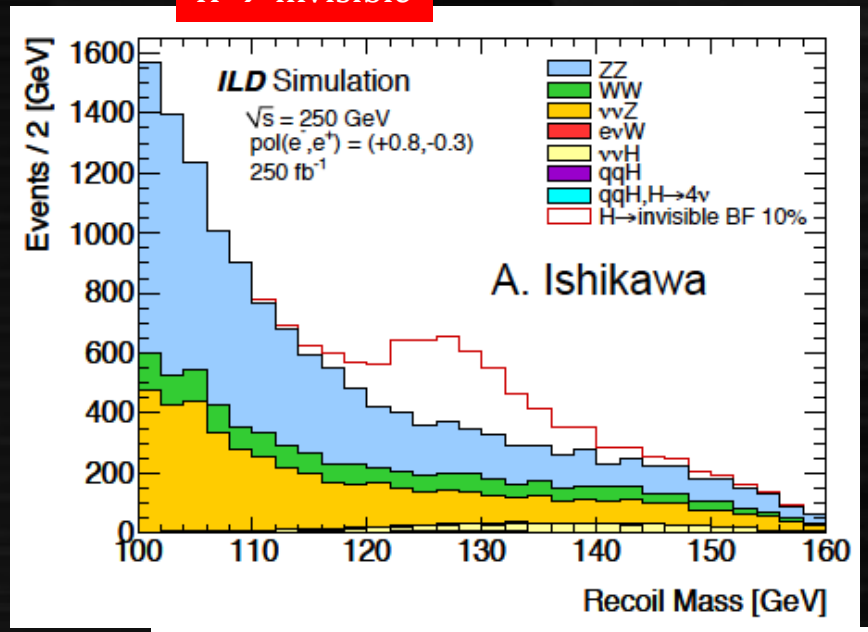




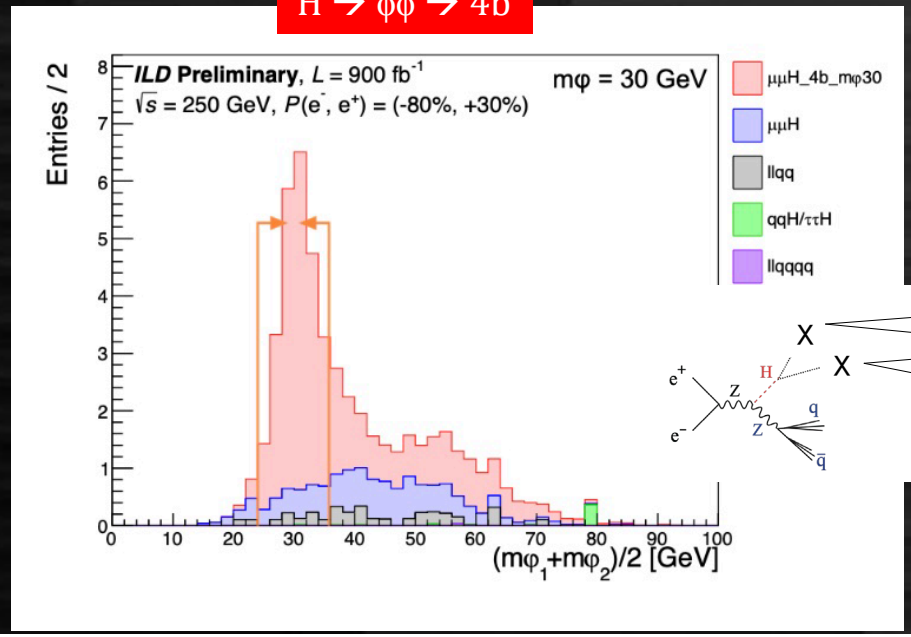
O(%)の結合定数測定は1 TeVの新物理の
間接探索になる。パターンによりモデル識別

Higgs 稀崩壊 (軽い暗黒物質探索)

H → invisible



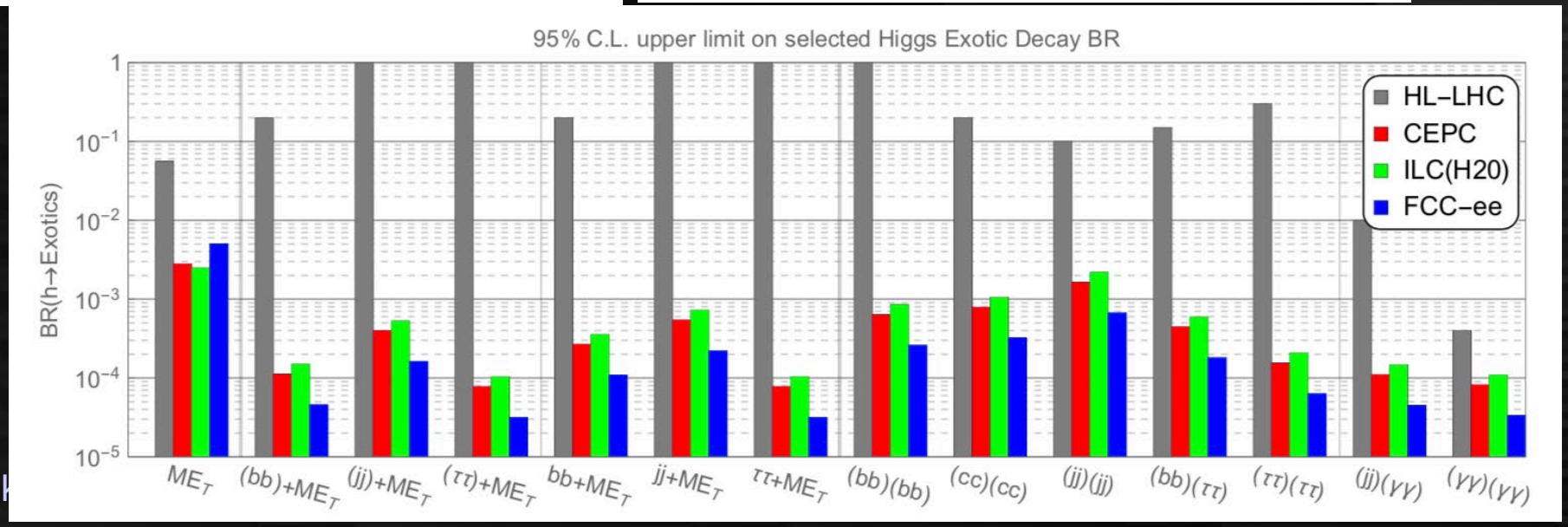
H → ϕϕ → 4b



H → SMでない終状態

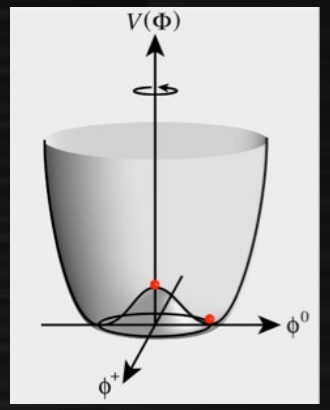
H → ff → なにか

- bbbb
- bb + DMDM
- その他いろいろ

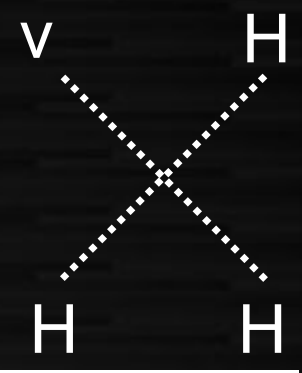


Taik

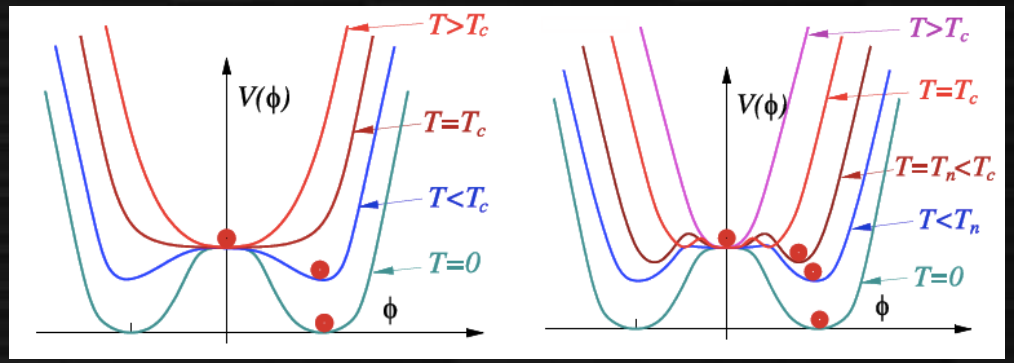
Higgs自己結合



ヒッグス
ポテンシャル
の4次の項
真空の構造
を決める

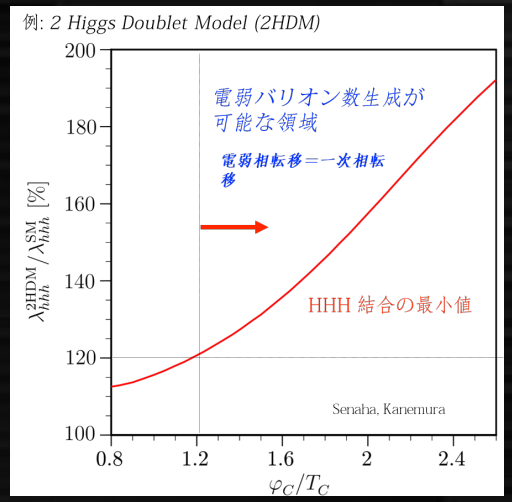


$$V(\eta_H) = \frac{1}{2}m_H^2\eta_H^2 + \lambda v\eta_H^3 + \frac{1}{4}\lambda\eta_H^4$$

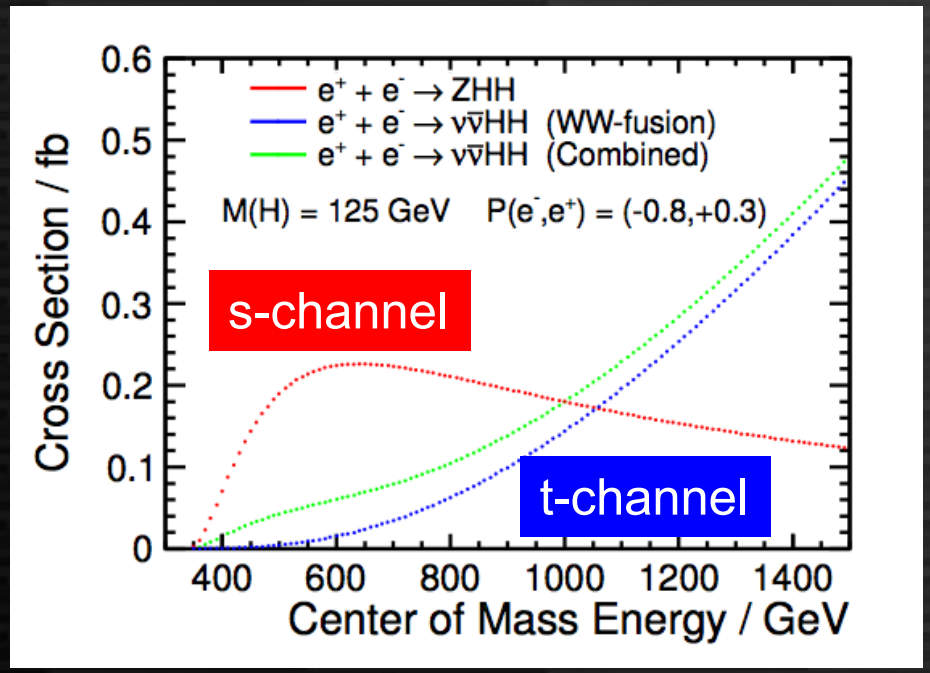


真空の2次相転移　真空の1次相転移

電弱バリオジェネシス(有力な物質生成仮説の一つ)
には真空の一時相転移が必要



一次相転移なら
λの値は
O(10%)以上
増加する



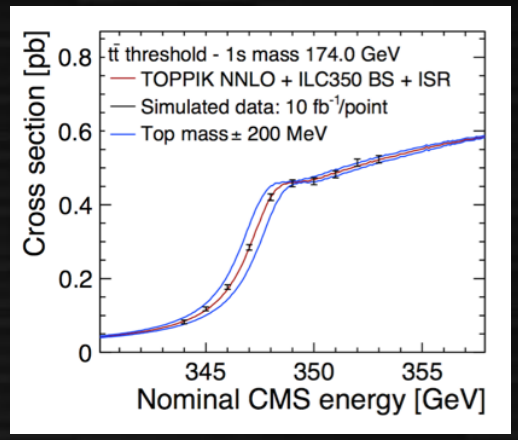
断面積が小さく困難。
1 ab^{-1}でO(100)事象

channel	√s[GeV]	L [ab^{-1}]	λ精度
s (正の干渉)	550	8	~13%
t (負の干渉)	1000	4	<10%

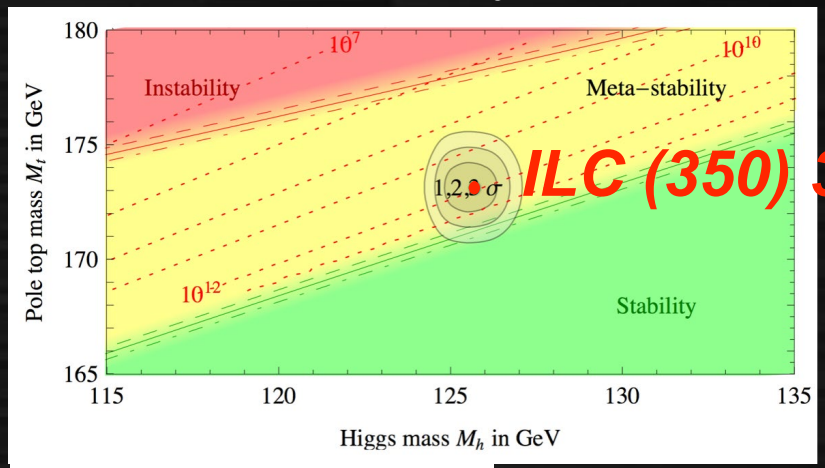
Linear Collider (550 GeV以上)なら
10%程度の感度でλの測定が可能

ttbar threshold and ttH

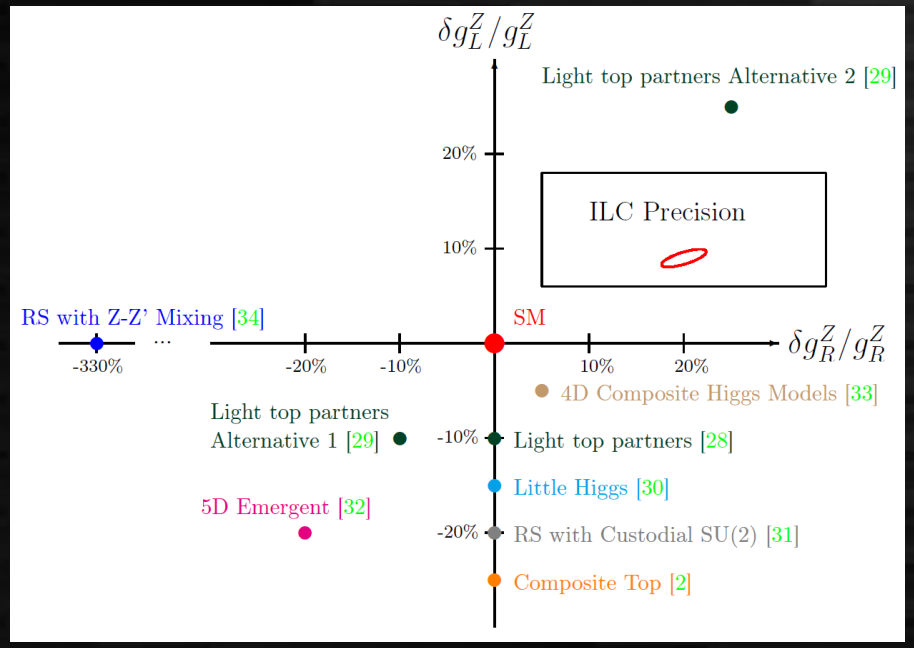
Threshold scan @ 350 GeV



"Potential-subtracted mass" which is theoretically compatible to Msbar mass can be directly observed



arXiv:1205.6497, Degrassi et al.



偏極を用いて右巻き、右巻きの結合定数を求め、Topが関係する新物理の探索・モデル識別が可能

$$\Delta m_t(\overline{MS}) \lesssim 50\,\mathrm{MeV}$$
$$\Delta m_h \simeq 14\ \mathrm{MeV}$$

Final answer on stable/metastable vacuum

Direct top-Yukawa measurement needs > 550 GeV CM energy

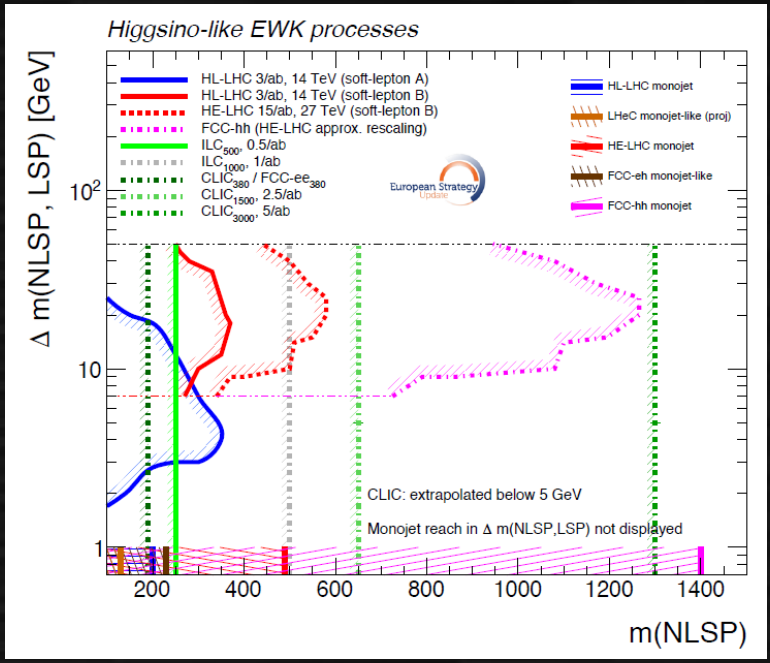
~2.8% possible

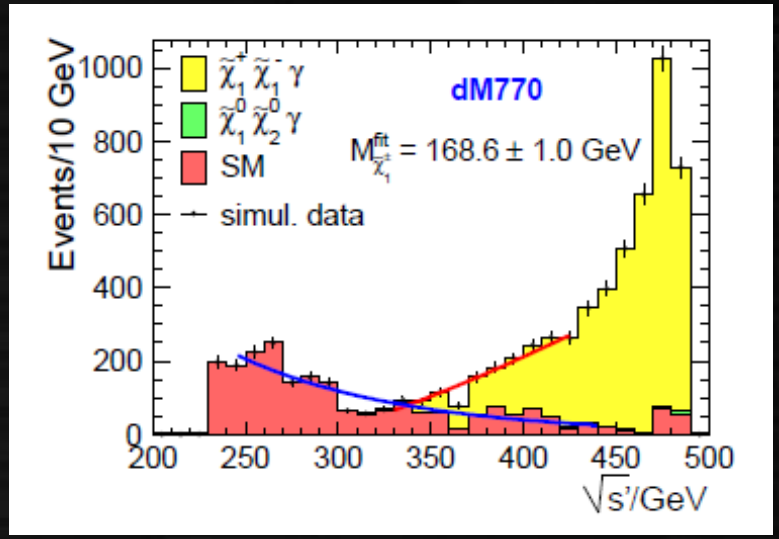
550 GeV can prove HH self coupling as well

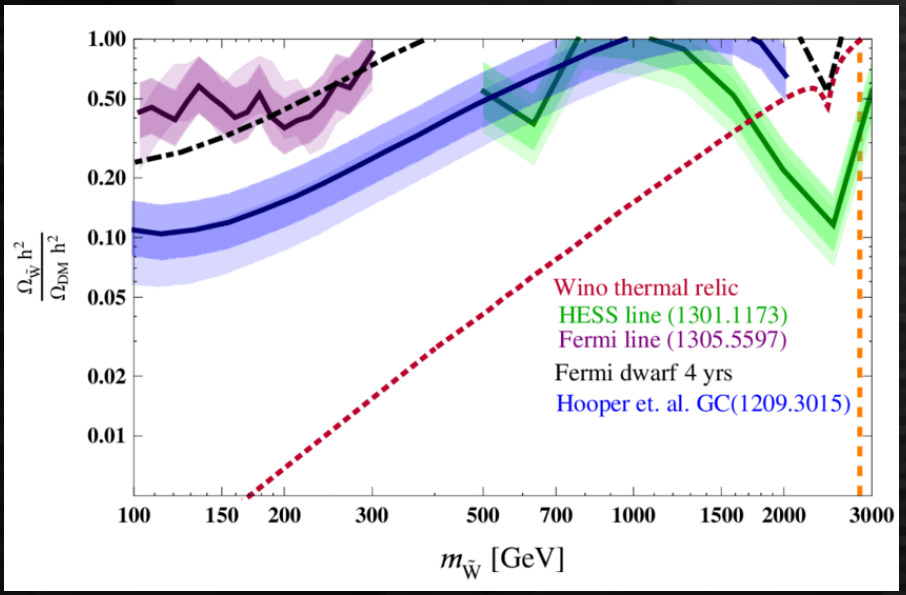
Other EW precision variables like 2f cross section (sensitive to Z'/WIMP) triple gauge coupling etc.

Ultimate target: direct search of TeV WIMP

- Big motivation of SUSY consistent with thermal DM
  - < 3 TeV Wino → 6 TeV collider needed (probably needs novel acceleration)
  - 1 TeV Higgsino → 2 TeV collider (~final target of SC or NC RF accelerator)
- Degenerated SUSY: easy to fill the gap by e+e- collider







物理のまとめ

- 240-250 GeV: e+e-ヒッグスファクトリーの共通第一ターゲット
  - ヒッグス結合定数の精密測定 → TeV新物理の包括的探索
  - ヒッグス稀崩壊 → 軽い新物理に感度
- リニアコライダーの物理
  - 電子・陽電子偏極を活かして様々な物理の感度向上
  - 550 GeV, 1000 GeV: ヒッグス自己結合に感度 (~10%)
    - ヒッグスポテンシャルの測定, 物質生成機構に感度
  - A few TeVまでの新物理直接探索 (特に 1 TeV Higgsinoに注目)
- 円形コライダーの物理
  - 240 GeVでリニアコライダーより数倍高いluminosity
  - 91 GeV (Tera-Z): 電弱精密測定, フレーバー物理 (ただし超高ルミノシティが必要で技術的困難が大きい)

解析技術と深層学習

物理解析の流れ

イベント生成
(シグナル/バックグラウンド)

モデル→粒子

測定器シミュレーション

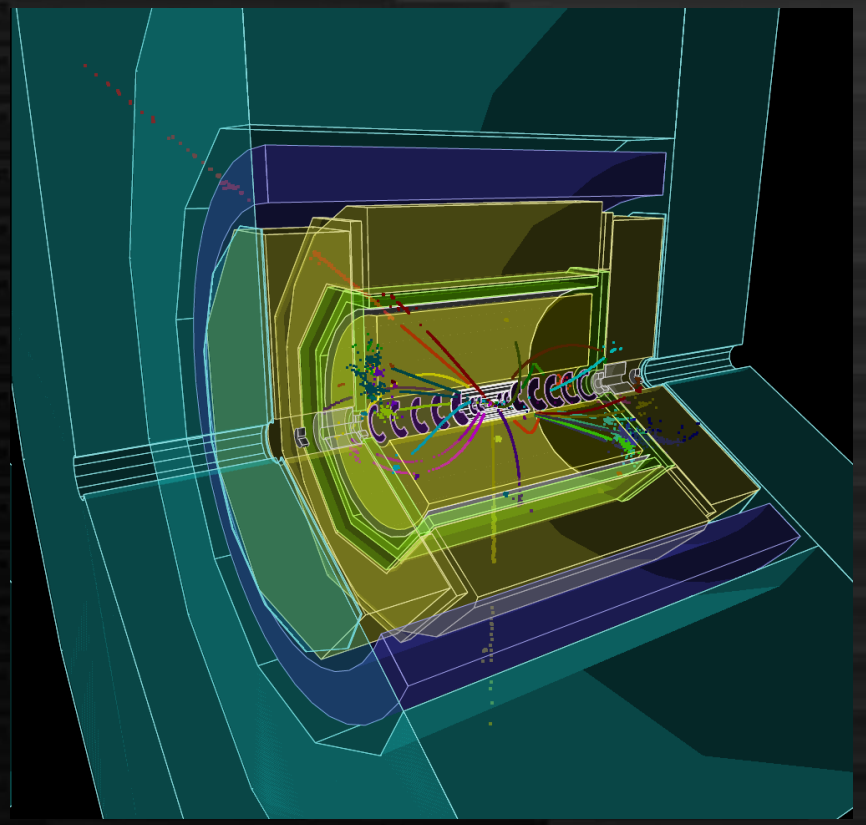
粒子→ヒット

再構成

ヒット→粒子

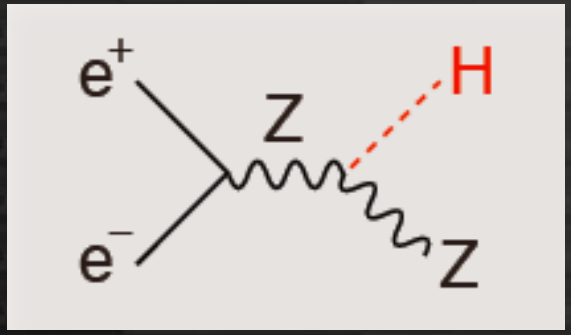
解析 (カット、フィット、
物理パラメータ導出など)

粒子→モデル



ZH → μμbb at ILD 250 GeV

イベント生成



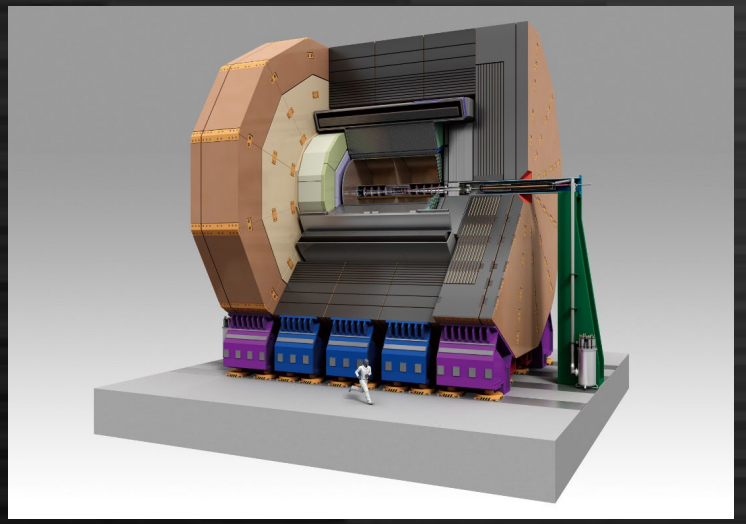
- モデルからMatrix elementを計算
  - 標準理論、いろんなBSM理論
  - シグナル / バックグラウンドそれぞれ
- Phase spaceを積分して微分断面積を計算
  - 4元運動量、偏極、各粒子
- 断面積の分布に従ってランダムにイベント生成
  - ビームスペクトル、ISR/FSRを考慮
- カラー粒子のハドロン化
  - 中寿命粒子(b,cハドロン)の崩壊
  - タウの崩壊は特別なコードを使う

ILDでのイベント生成
- Diagram計算: whizard
- ハドロン化: pythia
- タウ崩壊: tauola

ILDでのイベント生成
- Diagram計算: whizard
- ハドロン化: pythia
- タウ崩壊: tauola

測定器シミュレーション

- 生成した粒子を「測定器」内で発生させる
  - 衝突角, ビームスポット
  - 測定器中の物質との相互作用
  - 長寿命粒子の崩壊(K_0など)
- Geant4 (汎用測定器シミュレーションコード)
  - 空間に測定器の物質を配置し、粒子をトラッキング
    - 各種相互作用(physics list)をシミュレート
  - 多量の計算機パワーが必要 (O(1分)/event)
- 軽量シミュレーション (fastsim)もある
  - 発生した粒子の4元運動量などを測定器の期待される分解能でランダムにsmearingする
    - 測定器シミュレーション(or実データ)で校正する必要がある
  - 測定器、実験ごとに定義された「Delphes Card」を用いる



GEANT4 A SIMULATION TOOLKIT

https://geant4.web.cern.ch/

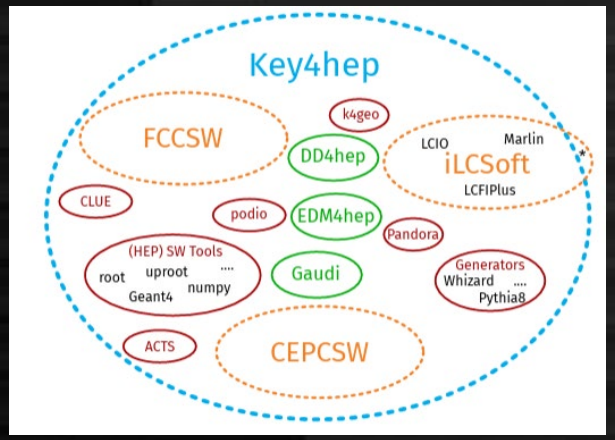
DELPHES fast simulation

https://delphes.github.io/

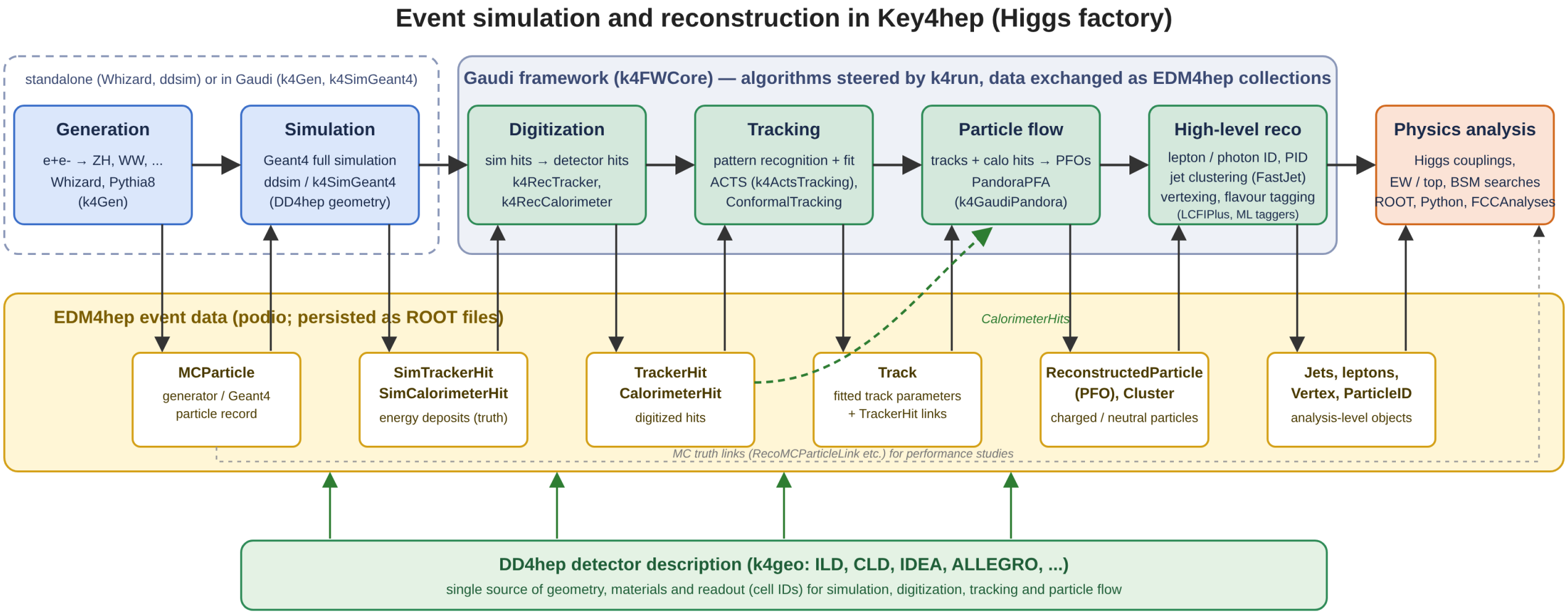
Key4hep framework

https://key4hep.github.io/

- 将来コライダー等のためのシミュレーション・再構成フレームワーク
  - Event Data / Persistency model (edm4hep / podio)
    - 真の粒子(MCParticle)、再構成粒子(Reconstructed Particle)、飛跡 (track)、クラスタ (cluster)、測定器ヒット (TrackerHit, CalorimeterHit)などのシミュレーション・再構成に必要な要素のオブジェクトフォーマットとファイル構造を定義
    - Key4hepではROOTファイルで保存 (xxx.edm4hep.root)
  - 測定器ジオメトリーの定義 (k4geo)
    - 各測定器要素の位置や構造が定義されている
    - たとえばILDのgeometryを使ったり、パラメータ調整もできる
  - 測定器シミュレーション (ddsim) Geant4のラッパー
    - 上記の測定器ジオメトリーをgeant4に自動で読み込んでイベントファイルをプロセスする
    - シミュレーション結果をedm4hepファイルに保存
  - 事象再構成 (後述)

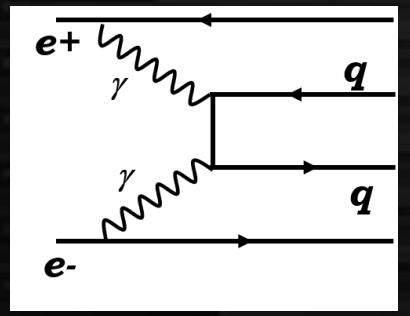


イベント再構成


Event simulation and reconstruction in Key4hep (Higgs factory)
standalone (Whizard, ddsim) or in Gaudi (k4Gen, k4SimGeant4)
Gaudi framework (k4FWCore) — algorithms steered by k4run, data exchanged as EDM4hep collections
Generation
e+e- → ZH, WW, ...
Whizard, Pythia8
(k4Gen)
Simulation
Geant4 full simulation
ddsim / k4SimGeant4
(DD4hep geometry)
Digitization
sim hits → detector hits
k4RecTracker,
k4RecCalorimeter
Tracking
pattern recognition + fit
ACTS (k4ActsTracking),
ConformalTracking
Particle flow
tracks + calo hits → PFOs
PandoraPFA
(k4GaudiPandora)
High-level reco
lepton / photon ID, PID
jet clustering (FastJet)
vertexing, flavour tagging
(LCFIPlus, ML taggers)
Physics analysis
Higgs couplings,
EW / top, BSM searches
ROOT, Python, FCCAnalyses
EDM4hep event data (podio; persisted as ROOT files)
CalorimeterHits
MCParticle
generator / Geant4
particle record
SimTrackerHit
SimCalorimeterHit
energy deposits (truth)
TrackerHit
CalorimeterHit
digitized hits
Track
fitted track parameters
+ TrackerHit links
ReconstructedParticle
(PFO), Cluster
charged / neutral particles
Jets, leptons,
Vertex, ParticleID
analysis-level objects
MC truth links (RecoMCParticleLink etc.) for performance studies
DD4hep detector description (k4geo: ILD, CLD, IDEA, ALLEGRO, ...)
single source of geometry, materials and readout (cell IDs) for simulation, digitization, tracking and particle flow
Generation / simulation
Reconstruction algorithms (Gaudi)
EDM4hep collections
Analysis
Tool names are examples; Marlin processors can also be run in the same chain through k4MarlinWrapper (EDM4hep ↔ LCIO conversion) when needed.

再構成1: digitization

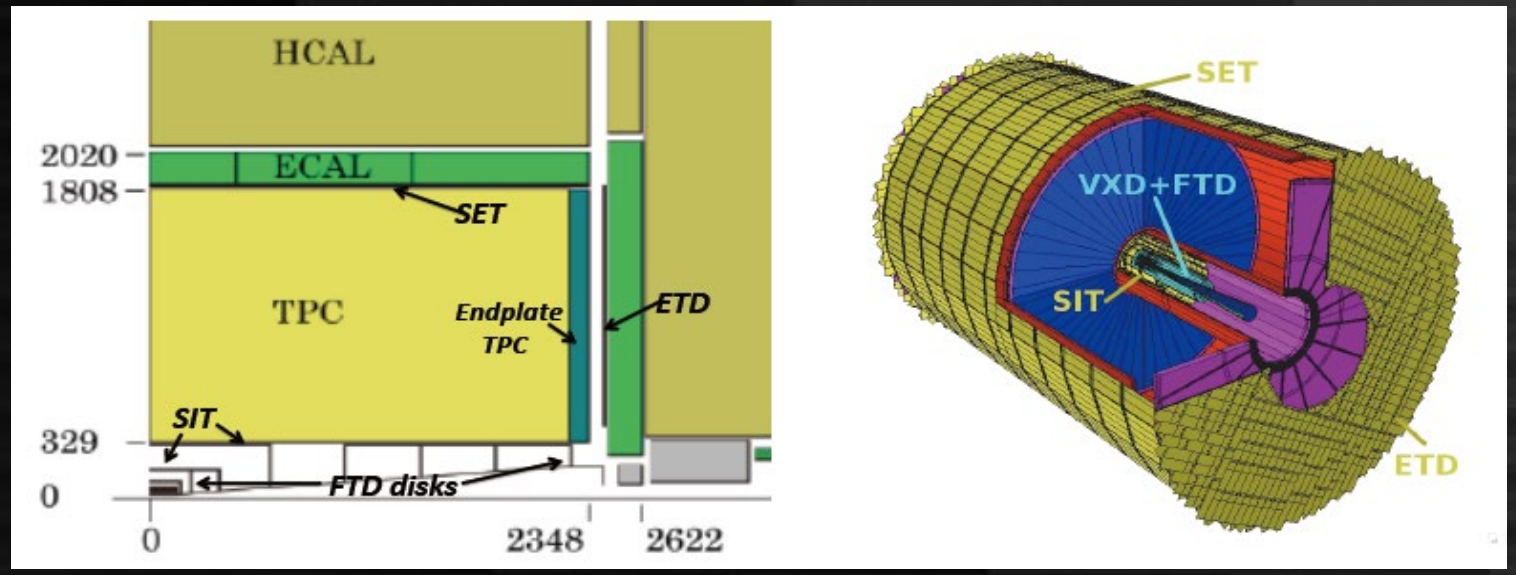
- シミュレーションの結果はenergy deposit
- 測定器の性能により得られる反応はenergy depositより劣化する
  - 有限のピクセルサイズ
  - ドリフト中の電荷の広がり(TPC)など
- これらの効果を入れて「測定器のヒット」を生成
- Pileup事象をここでoverlayすることも
  - $\gamma\gamma$ → hadron pileup @ILC
    1.7 / bunch @ 500 GeV,
    4.1 / bunch @ 1 TeV



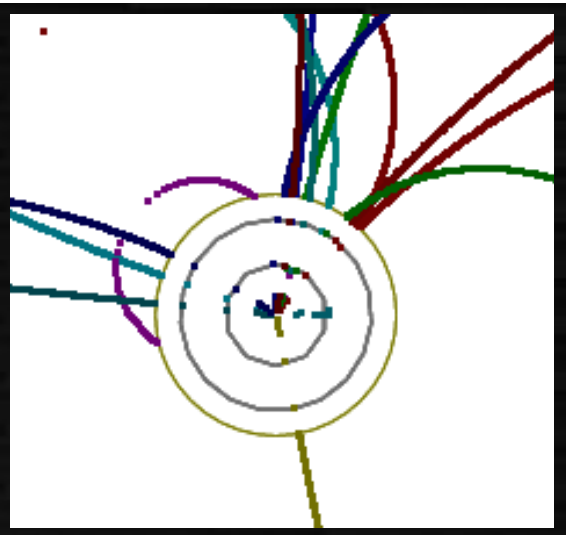
- ここで実データと比較できるヒットが完成
  ここから先は実データの時と同じことをする

再構成2: tracking

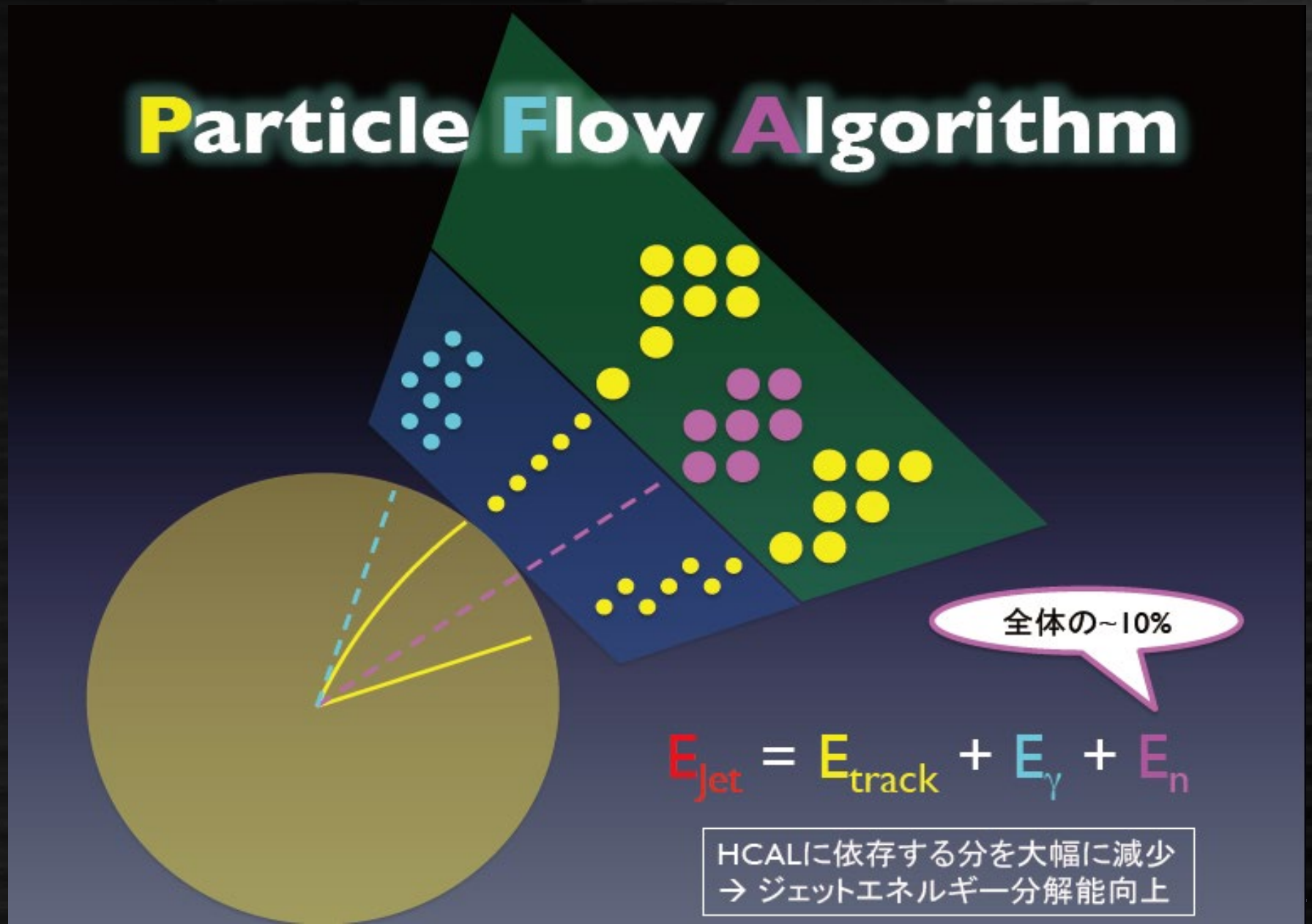
Vertex detector
　trackがどこから来たか
Silicon inner tracker
　タイムスタンプ
Time projection chamber
　運動量測定のメイン



- Track seedを作る
  Vertex+SIT, TPCそれぞれ
- つなげていく
  Helix fit, Kalman filter, など
- Silicon trackとTPC trackをつなげる
  Refitする
- dE/dxなどを使ってParticle ID



再構成3: Jet energy測定

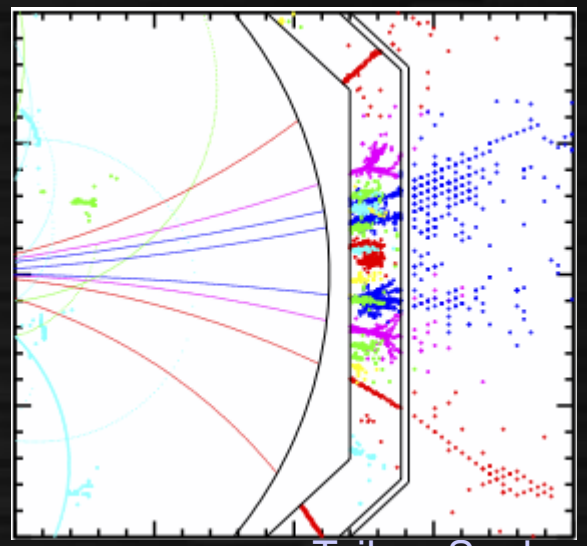


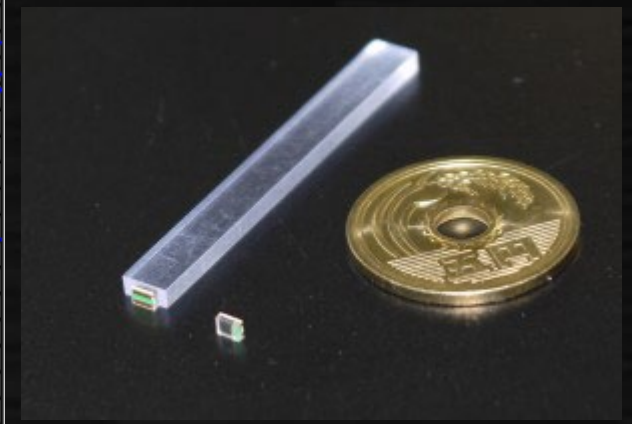
“Real” Particle Flow Calorimetry

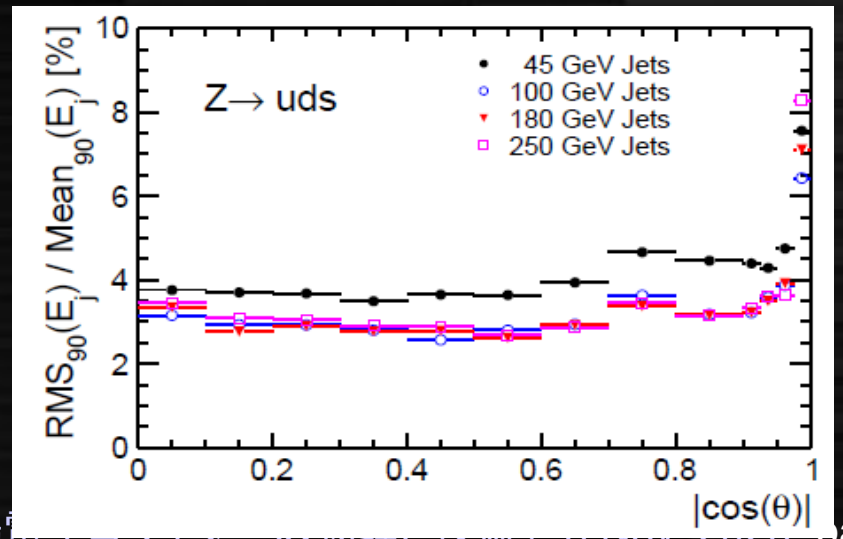
Separate “particle-by-particle” in jets
- need high granular calorimeter – typ. 5x5 mm^2 in ECAL

	Detector	Fraction	Resolution	Identical case (100 GeV)
Charged Particle	Tracker	~ 60%	0.1% / $p_T \sin\theta$ (GeV) for each	negligible contribution
Photon	ECAL	~ 30%	15% / $\sqrt{E}$ (GeV)	15 / $\sqrt{30}$ x 30% = 0.82
Neutral Hadron	HCAL	~ 10%	60% / $\sqrt{E}$ (GeV)	60 / $\sqrt{10}$ x 10% = 1.90
ILC Calorimetry	All		30% / $\sqrt{E}$ (GeV)	0.82+1.90 = 2.72 GeV

ILD





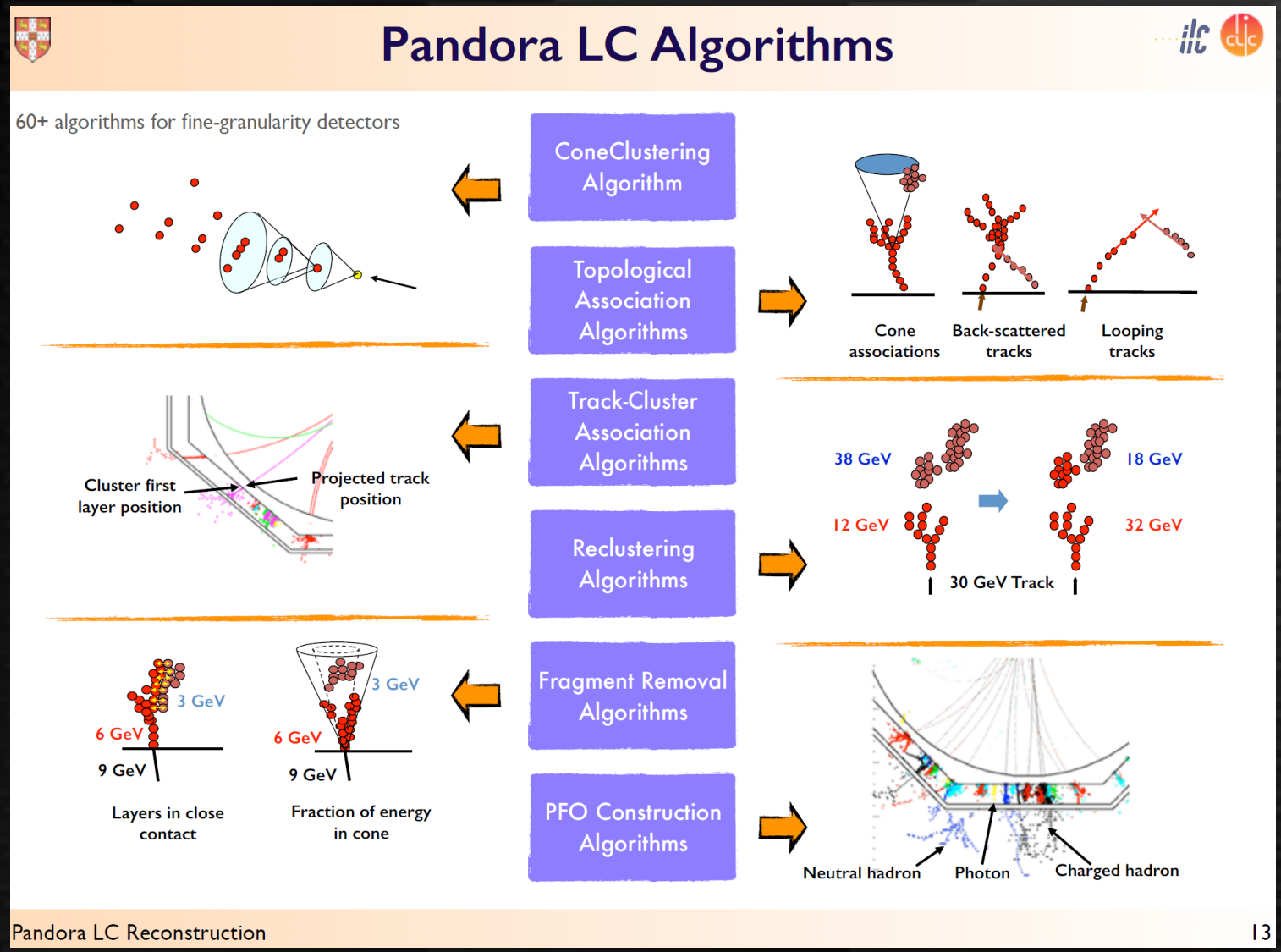


Pandora PFA

2008年頃に開発された
trackとhitから粒子を作る
アルゴリズム
(開発者: M. Thomson)
現在も使われている

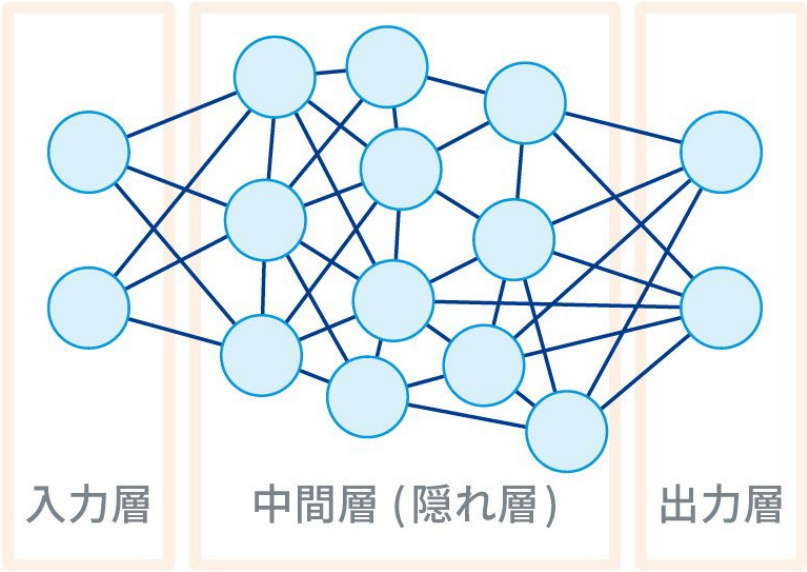
- Calorimeter Hitをクラスタ
- クラスタをマージ
- Trackと結びつける
- Reclustering
- Trackのmomentumを
  使って微調整

→ output

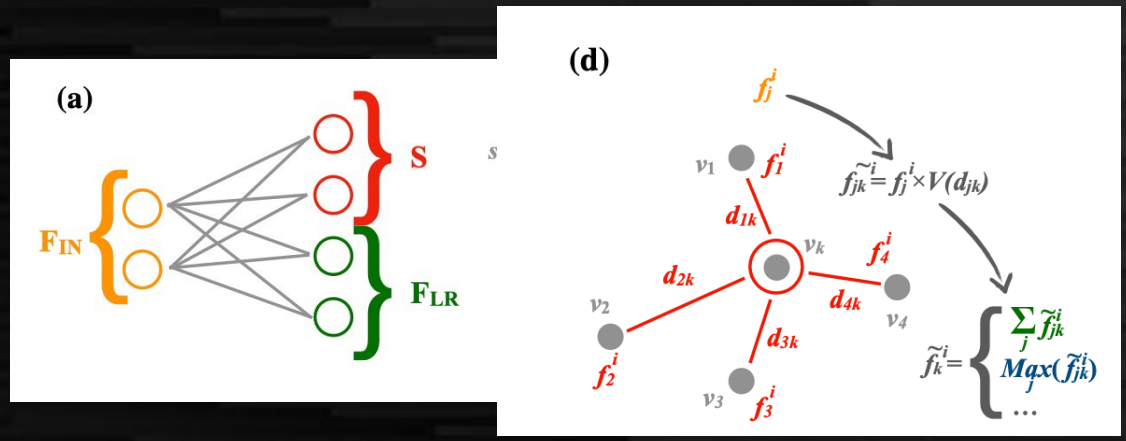


ML-based particle flow

- 最近Pandoraを上回るalgorithmがDNNベースで作られている
- 主な発想: graph neural network and/or transformer



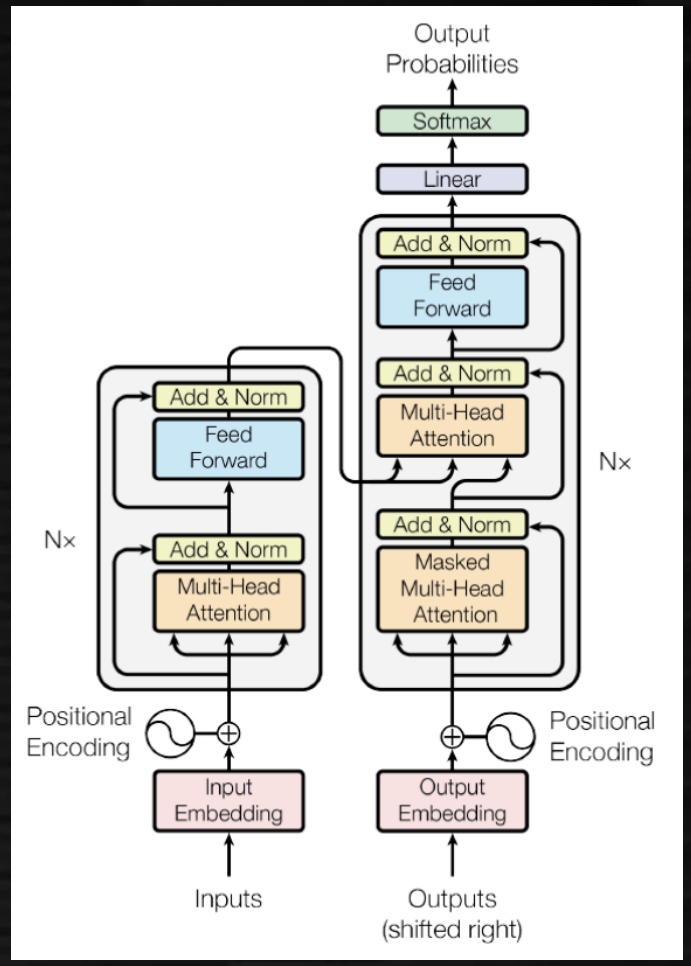
普通の全結合ネットワーク



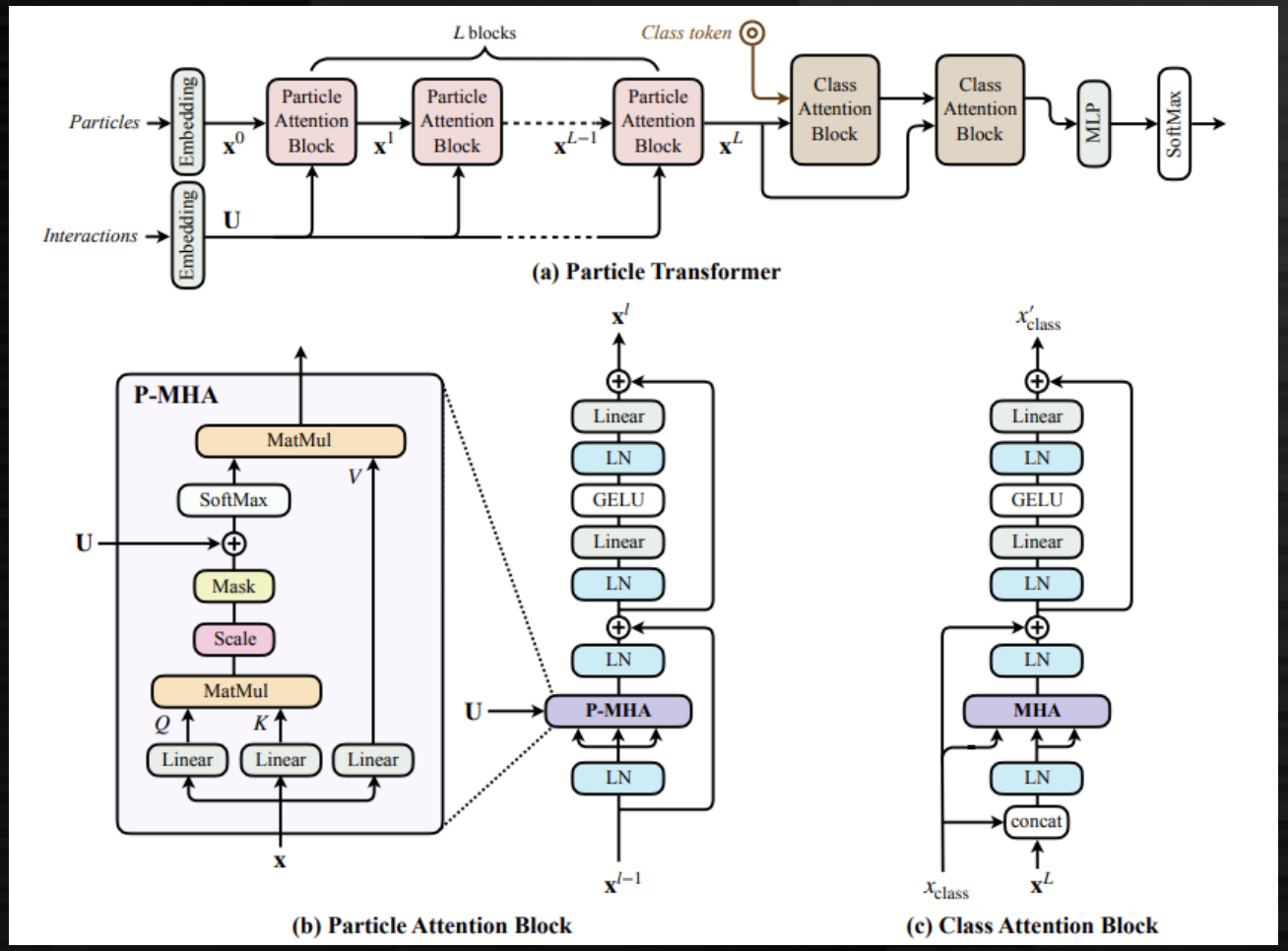
グラフニューラルネットワーク

各「object(node)」間でデータを送って
Nodeの情報を更新する
Node間の関係性を学習させる

Transformerとその応用



Regular Transformer

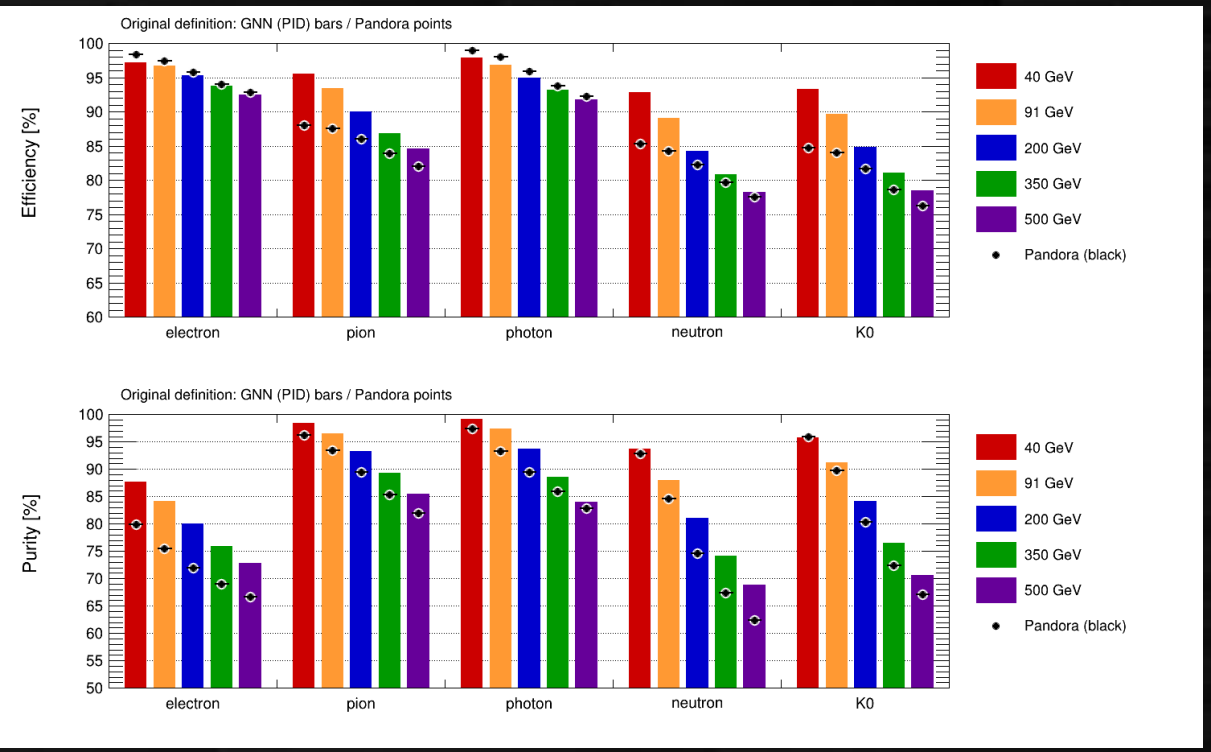


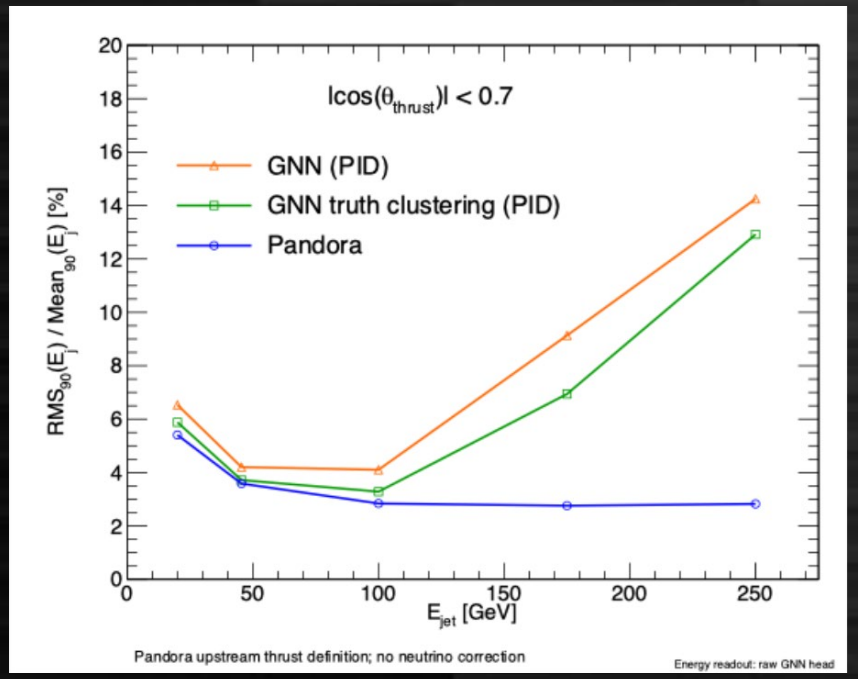
Particle Transformer

MHA – MultiHeadAttention

Note: P-MHA – Augmented version of MHA by Particle Transformer that involves Interactions Embeddings instead of Positional Embeddings

ML-based PF: some results





東大の開発ではジェット分解能はあと一歩

クラスターの分離効率(efficiency, purity)はPandoraPFAより大幅に向上

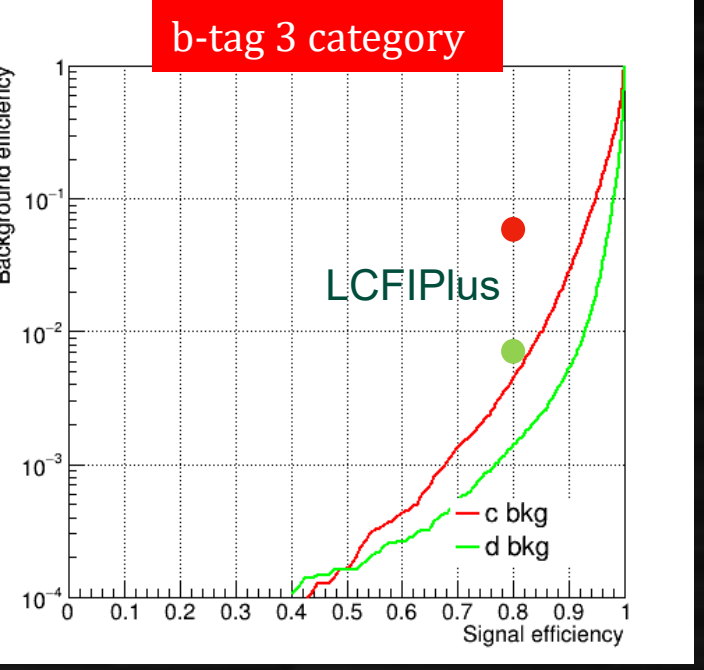
国際競争が激しくなっている
(条件によるがPandoraを上回る性能を示しているモデルも出始めている)

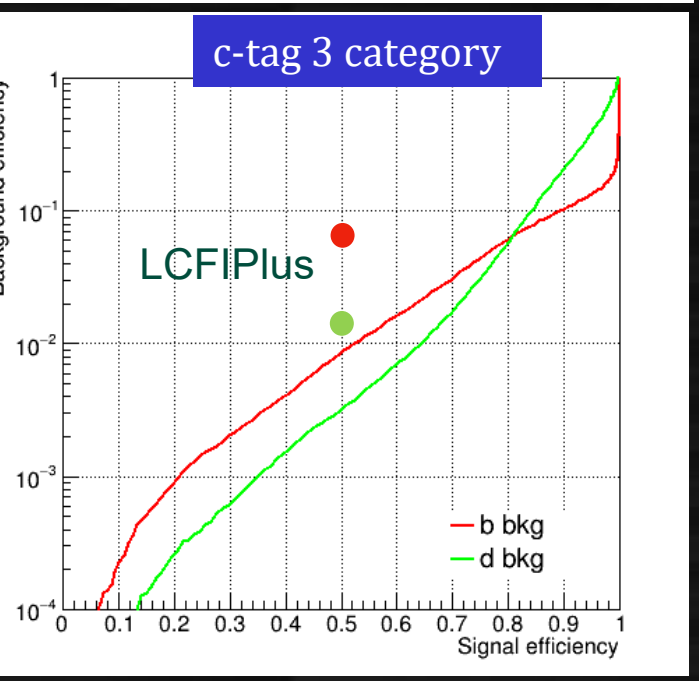
High level reconstruction

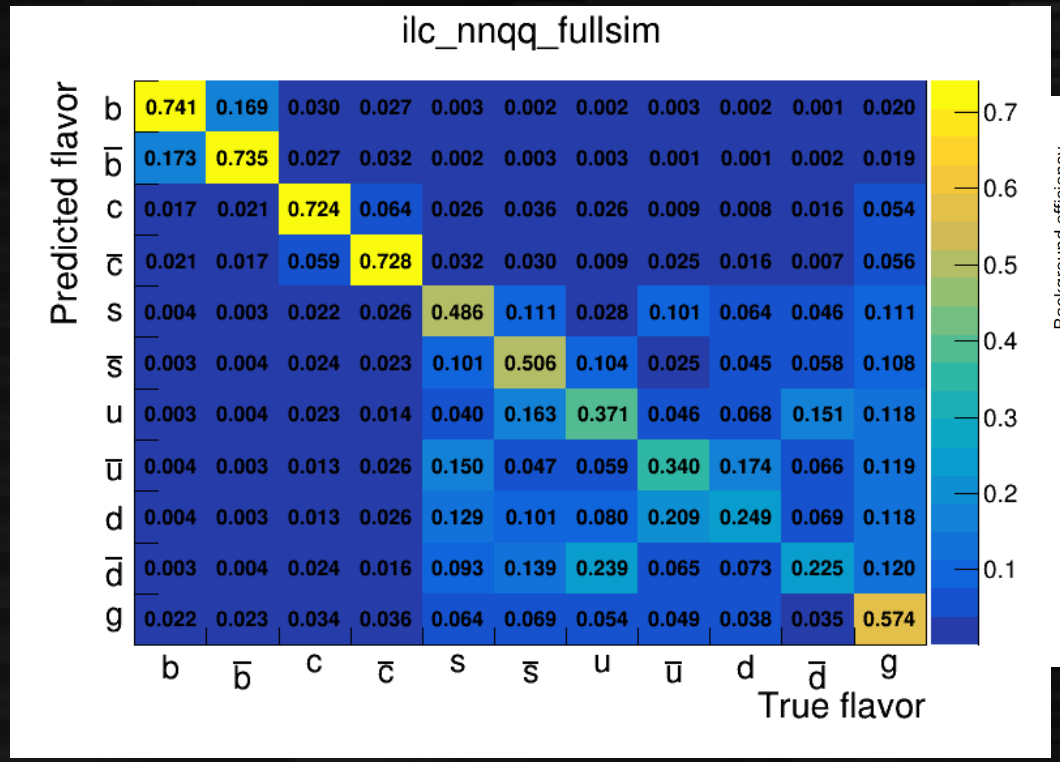
- 物理解析と再構成オブジェクトをつなぐ様々なアルゴリズム
  - Vertex reconstruction
  - Jet clustering / jet flavor tagging
  - Isolated Lepton / Tau / photon finder
  - …
  - 解析に応じて必要なアルゴリズムは変わる
    - 基本的に解析specificに走らせる (jet数などは解析により異なる)
    - Vertex reconstructionだけはILDでは標準再構成に含まれる

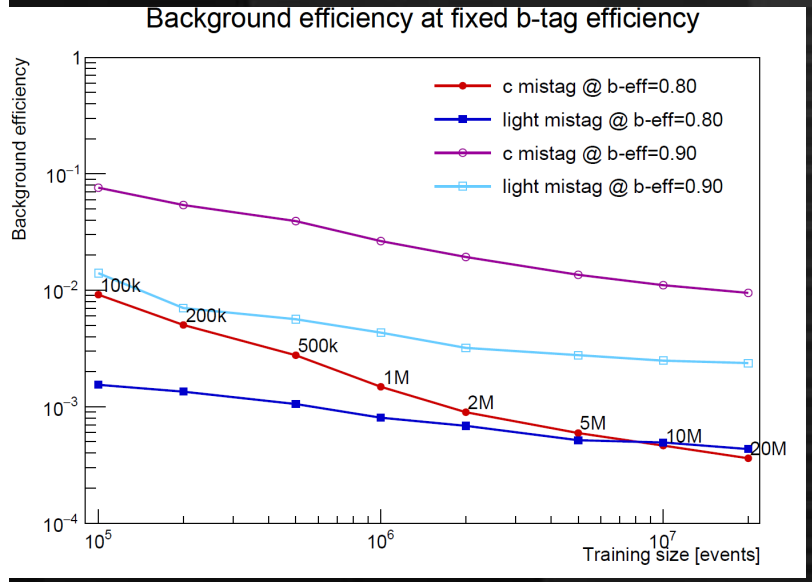
ML-based flavor tagging

圧倒的な性能: 旧来より~10倍よい背景分離能力
学習サンプルを増やすと更によくなっていく
従来できなかったストレンジ識別、ジェット電荷識別も
学習の効率やパラメータ依存性等を調査中



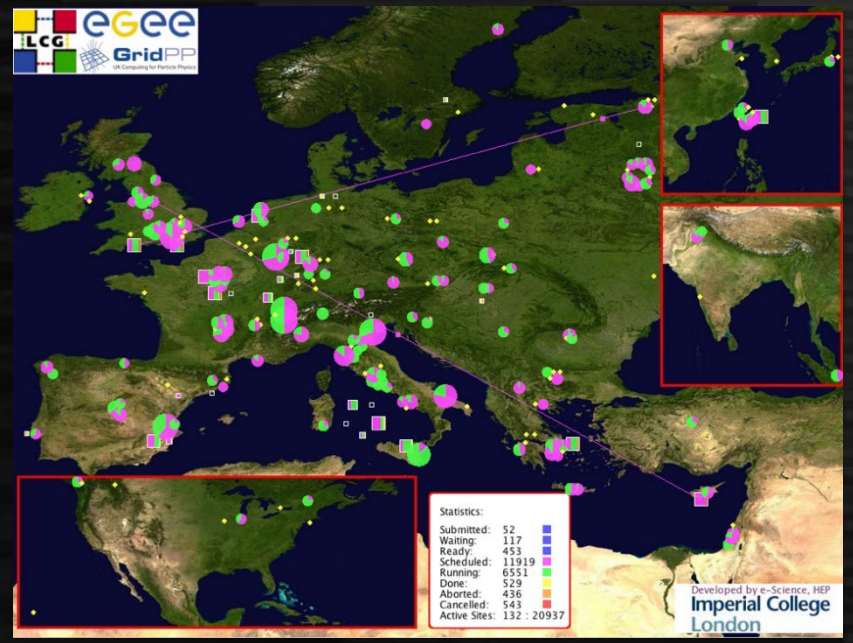






計算機資源

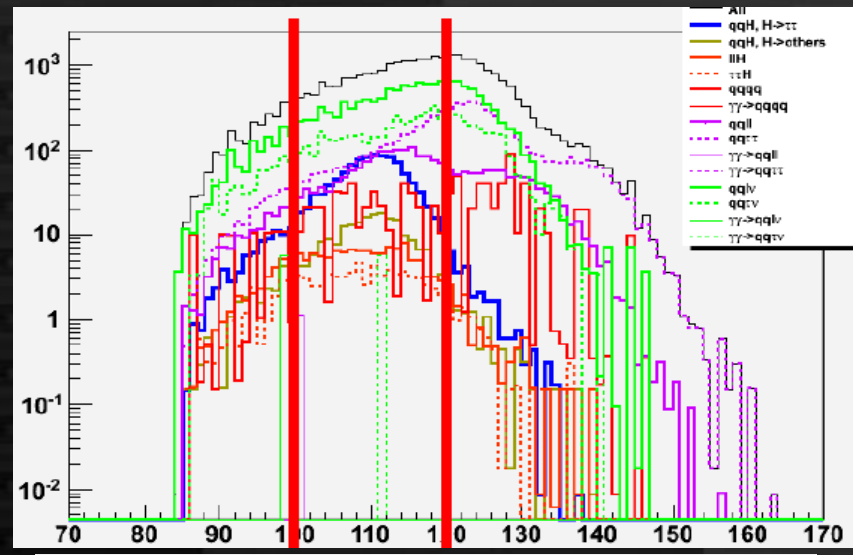
- 再構成や物理解析には大量の計算機資源が必要
  - シミュレーション等: CPU並列計算 (MCはif文の集合でGPUは不適)
    - Transformer等のtrainingには多数のサンプルが重要
  - 深層学習: GPU (大きな行列計算が得意)
- Grid
  - 世界中で共有する計算機資源
  - 主にMC jobを投入/データ共有
- local
  - KEKやCERN等のローカル資源 (CPUクラスタ)
  - 手元や大学の共用計算機 (GPU等)



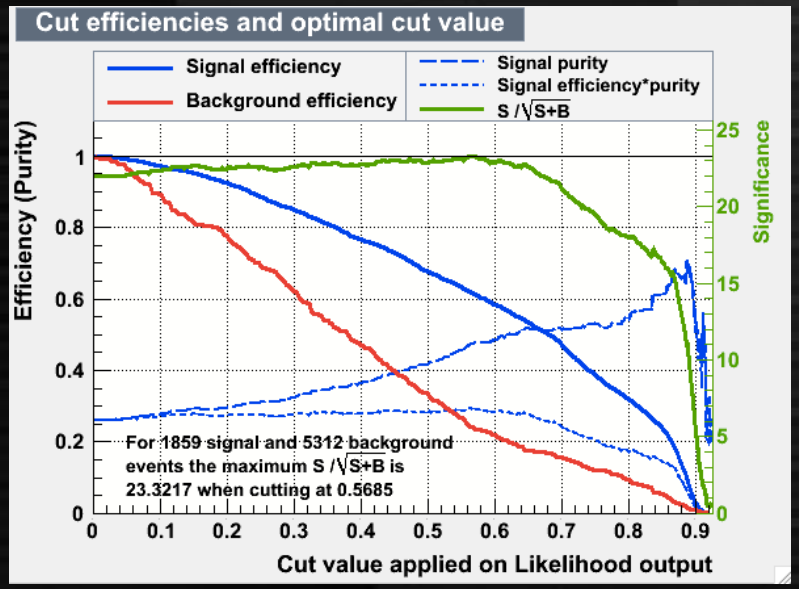
物理解析 (少しだけ)

モデル ← 信号事象
モデル ← 背景事象

- 一致・不一致の検証
- パラメータの決定



1. 事象識別 (信号をできるだけ選び、背景を捨てる)
- ジェットやレプトンの数など
  - “n-jetらしさ”, レプトン選別のパラメータなど
- ジェットフレーバー
  - bがいくつあるか、cがいくつあるか等
- 運動学的情報
  - 2ジェットで組むZやHの質量, missing E/m等
- これらを組み合わせてカットや多変量解析



Counting experiment

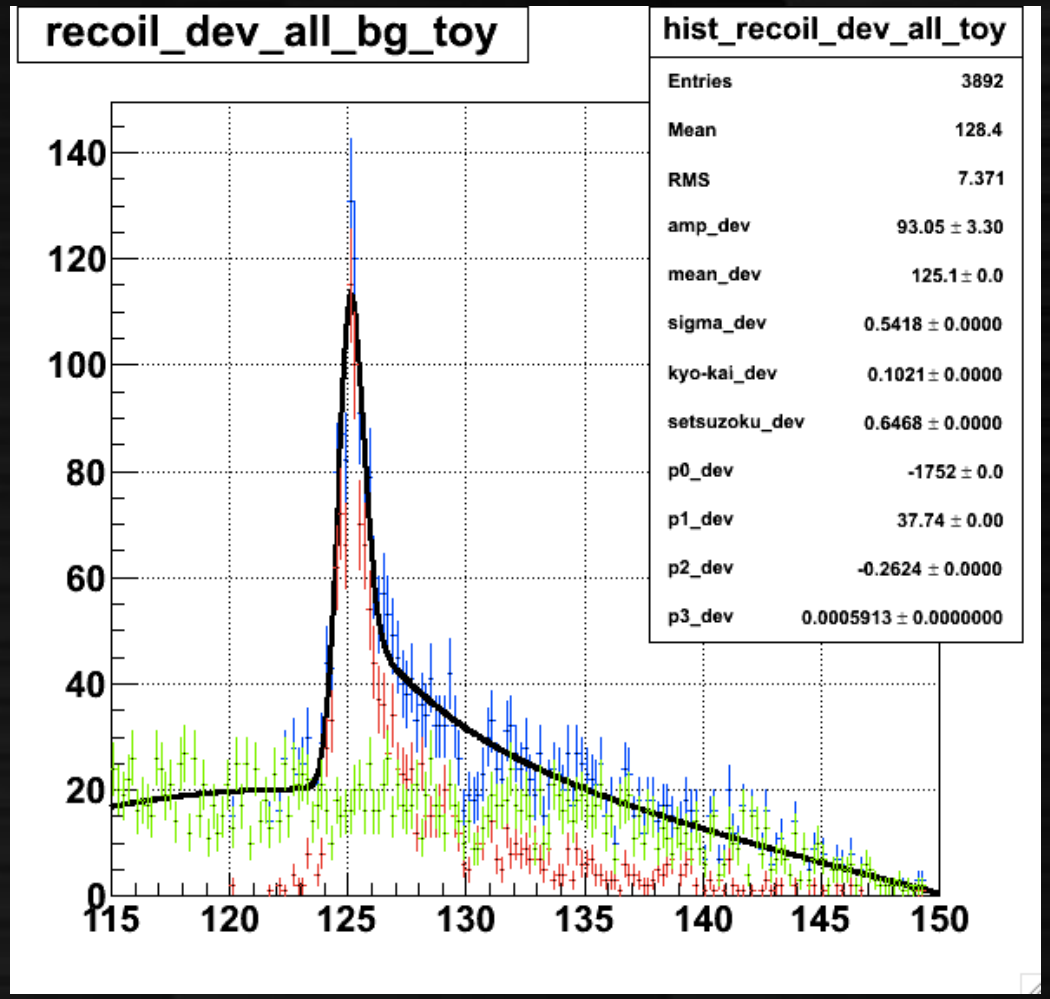
P(e-,e+)=(-0.8,0.3) $e^+ + e^- \to ZHH$ $M(H) = 120\text{GeV}$ $\int Ldt = 2\text{ab}^{-1}$

Energy (GeV)	Modes	signal	background	significance	
				excess (I)	measurement (II)
500	$ZHH \to (l\bar{l})(b\bar{b})(b\bar{b})$	3.7	4.3	1.5σ	1.1σ
		4.5	6.0	1.5σ	1.2σ
500	$ZHH \to (\nu\bar{\nu})(b\bar{b})(b\bar{b})$	8.5	7.9	2.5σ	2.1σ
500	$ZHH \to (q\bar{q})(b\bar{b})(b\bar{b})$	13.6	30.7	2.2σ	2.0σ
		18.8	90.6	1.9σ	1.8σ

S / √N S / √(N+S)

Selection後の信号と背景の数を数えて、信号の有意度や断面積を調べる
複数の測定チャンネルを組み合わせたり、モデルパラメータをスキャンして棄却領域を定めたりなど

フィッティングによる物理量抽出



フィットパラメータを使って
見たい変数を抽出

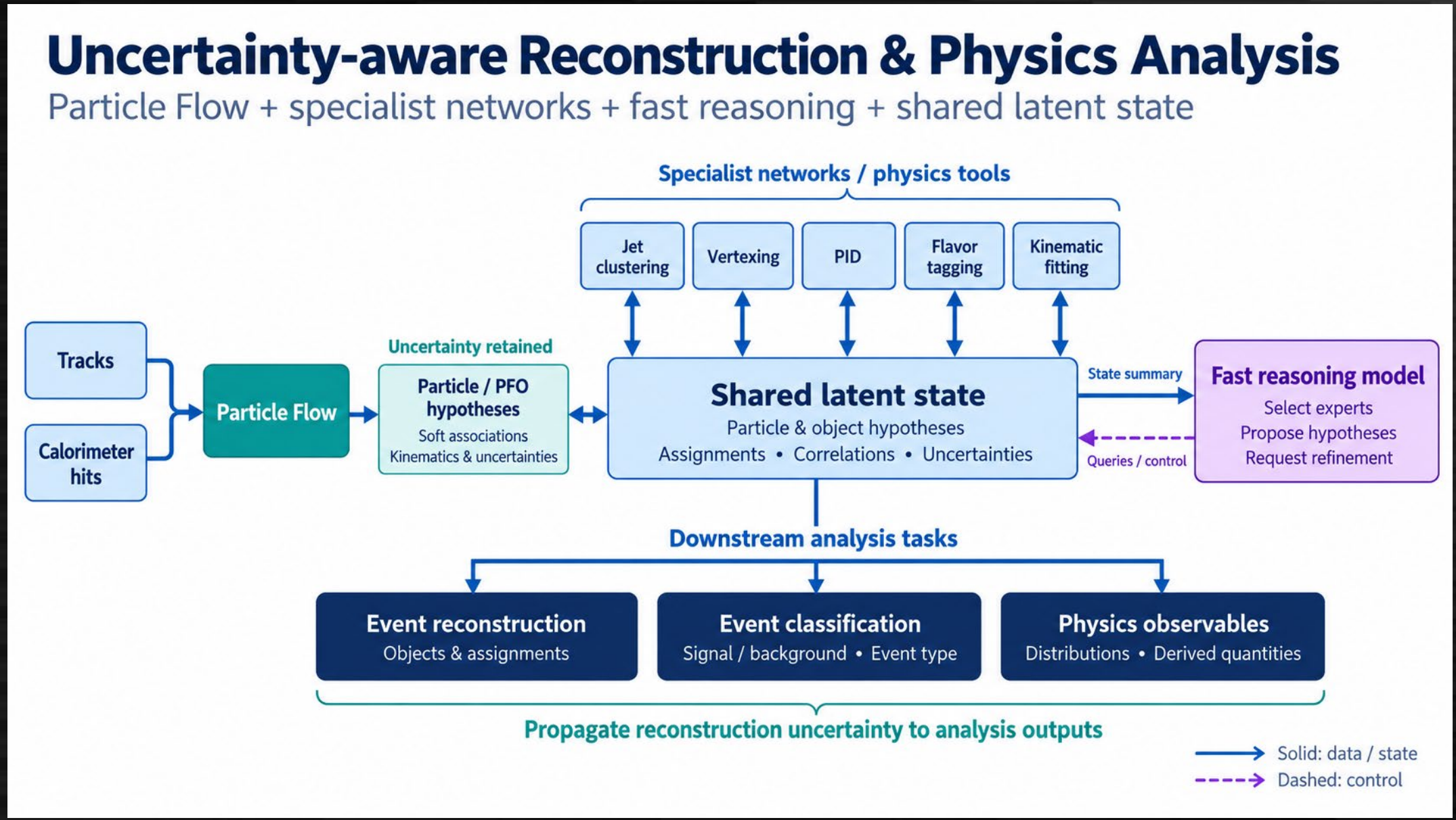
フィットエラーが見たい
変数の測定精度となる

χ^2 / NDFをチェック
→ 1に近ければfitはよい
1より大きい場合fitが
うまく合っていない
(モデル関数の不一致等)
1より小さい場合たいてい
エラーの付け方が
間違っている

物理解析と再構成

- 事象再構成の精度は解析後の物理性能に直結
  - ジェット分解能が向上 → 質量による信号/背景分離向上
  - フレーバー識別能が向上 → HiggsとZ/W/tの分離性能向上
    - 特に自己結合などHが複数あると質量だけでは組み合わせ不定性で分離困難
  - 飛跡の運動量分解能が向上 → H/Z等の質量精度向上
    - 分離にも使える
  - Hadron PID (π/K/p分離)向上 → bの崩壊モード分離やs識別に重要
- 深層学習(transformer等)で再構成性能が近年大幅向上
  - 多くの開発要素
  - 物理解析への適用も進行中
- 測定器のパラメータと性能の関係も重要

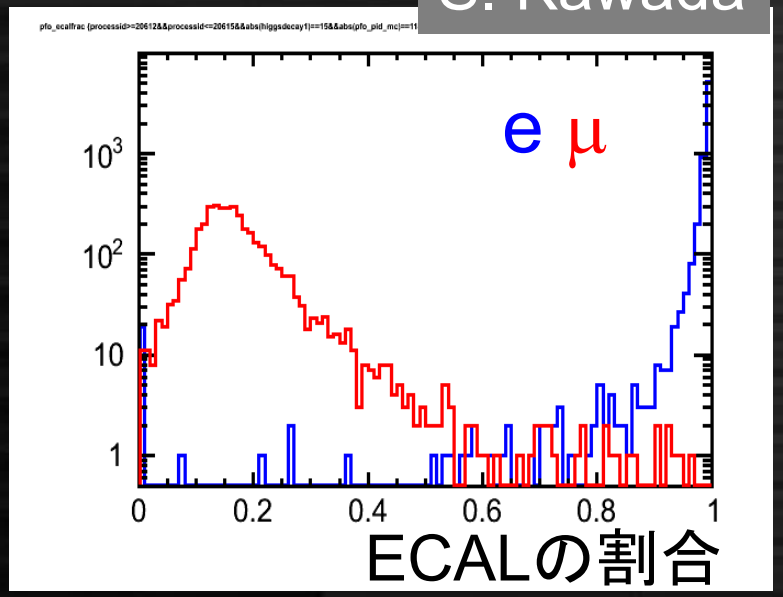
将来像?

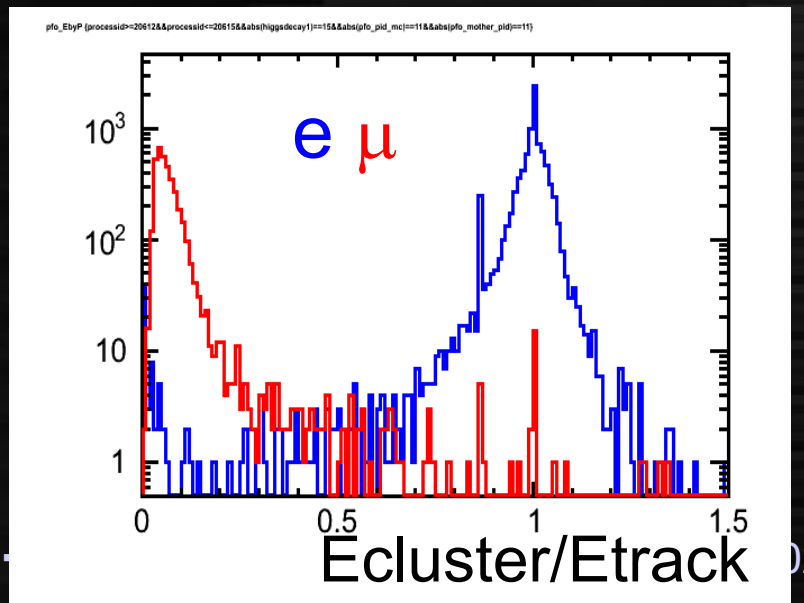

Uncertainty-aware Reconstruction & Physics Analysis
Particle Flow + specialist networks + fast reasoning + shared latent state
Specialist networks / physics tools
Jet clustering
Vertexing
PID
Flavor tagging
Kinematic fitting
Tracks
Calorimeter hits
Particle Flow
Uncertainty retained
Particle / PFO hypotheses
Soft associations
Kinematics & uncertainties
Shared latent state
Particle & object hypotheses
Assignments • Correlations • Uncertainties
State summary
Queries / control
Fast reasoning model
Select experts
Propose hypotheses
Request refinement
Downstream analysis tasks
Event reconstruction
Objects & assignments
Event classification
Signal / background • Event type
Physics observables
Distributions • Derived quantities
Propagate reconstruction uncertainty to analysis outputs
Solid: data / state
Dashed: control

再構成4: lepton ID

- isolate lepton (jet化前に取り除く)
  - ECALに落ちたエネルギーの割合
  - cluster energyとtrack energyの比
  - muon detector
  - isolation (見たいプロセスで異なる)
    - cone energyを使うことが多い
- jet lepton (flavor taggingに使う)
  - 同様のものをフレーバー識別コード内に実装
- 本当はcalorimeter内の形状も使える

S. Kawada





イベントのリストアップと重み付け

イベントのリストアップ

- シグナル　→　自分で用意することが多い
  SMヒッグスなどは用意されている場合も
- SMバックグラウンド　→　再構成済の「標準サンプル」を主に使う
  効きそうなプロセスだけ入れる or 全部入れる

イベントの重み付け

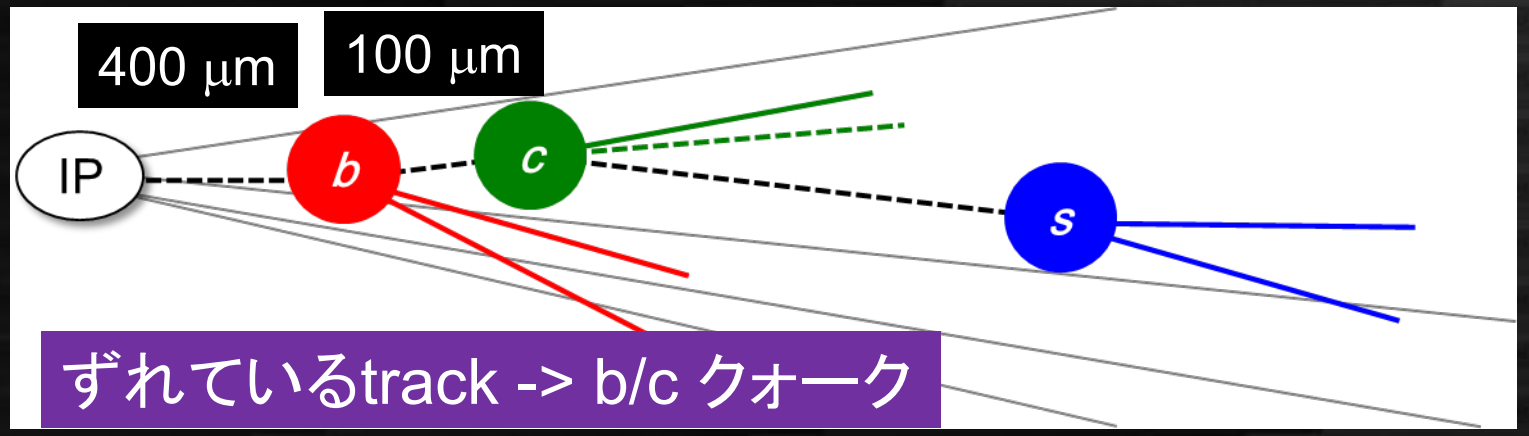
$$w = \sigma \times L \times p(e^-) \times p(e^+) / S$$

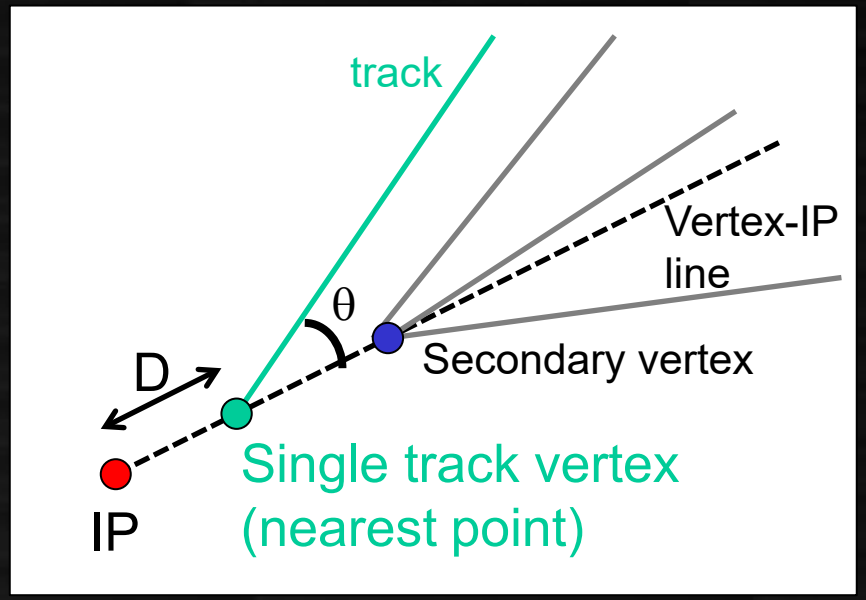
σ: cross section, L: luminosity, S: 持っているイベント数
$p(e^-)$: e^-のpolarizationが一致する割合 (90% or 10%)
$p(e^+)$: e^+のpolarizationが一致する割合 (65% or 35%)

重み付けの適用

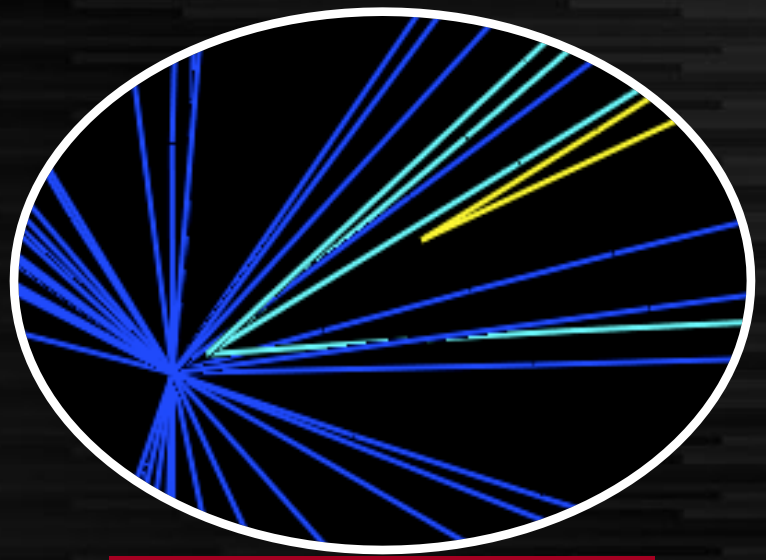
- 重み付けの割合だけ選ぶ (誤差が正しい)
- 各イベントに重みをつける (高統計)

再構成5: バーテックス検出





1本しかなくても!



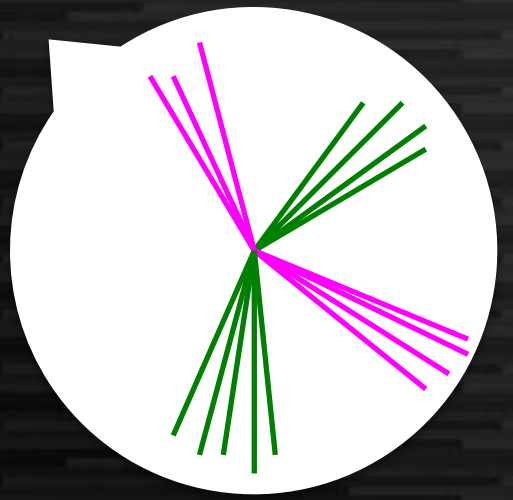
実際のMCイベント

再構成6: jet clustering

- yが一番小さいpairをつけていく

$$y = \frac{2\min(E_1, E_2)^2(1-\cos\theta_{ij})}{Q^2}$$

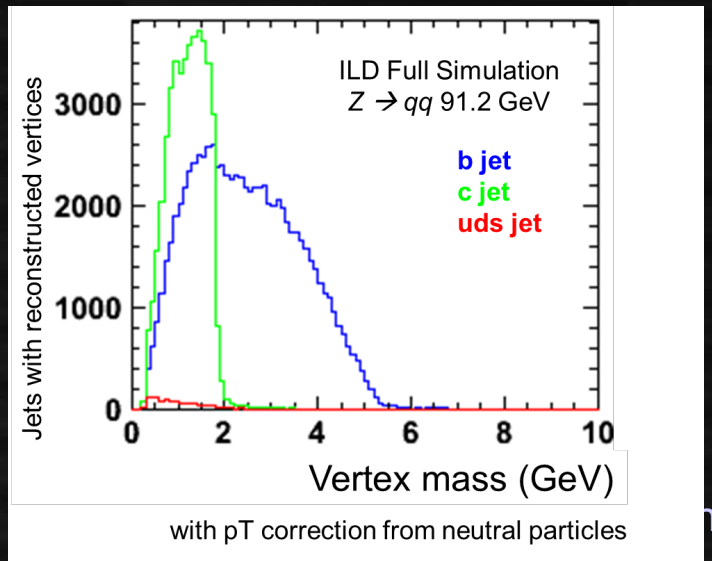
(Durham, 他のもある)



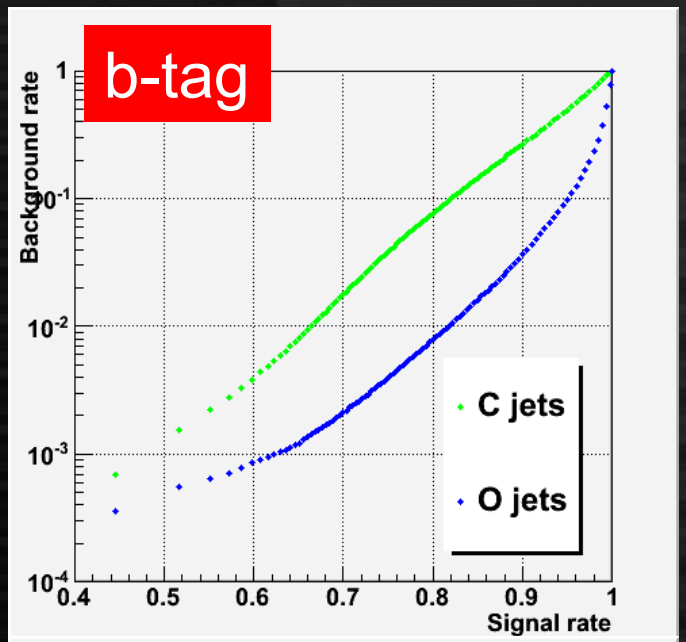
- multi-jetではどうしても間違える→多大な影響
  - そもそもjetの存在自体が近似
- vertexを使ってbを優先的にとる方法
- 前方のbeam起因粒子を落とすのにも使う
  - anti-kT clustering etc.

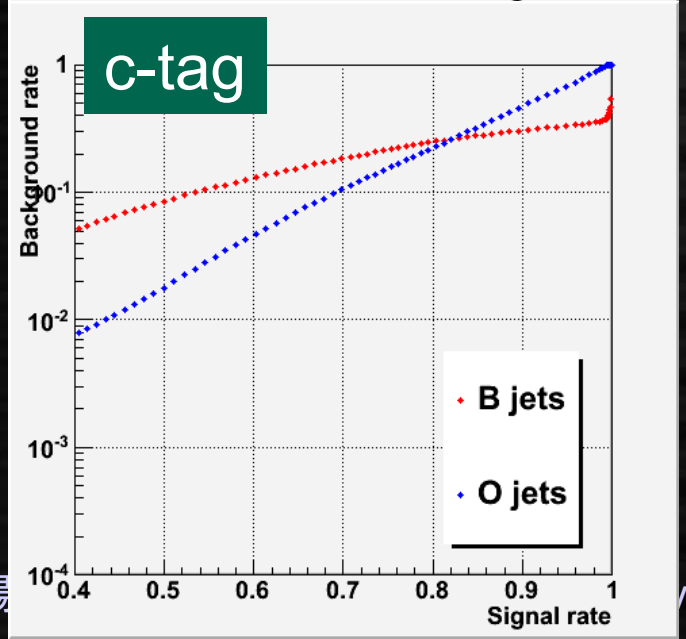
再構成7: フレーバー識別

- 多変量解析で分離
  - vertexの数(通常, single)
    - 実際には場合分け
  - vertex mass
  - trackのimpact parameter
  - 関連した変数(~20)



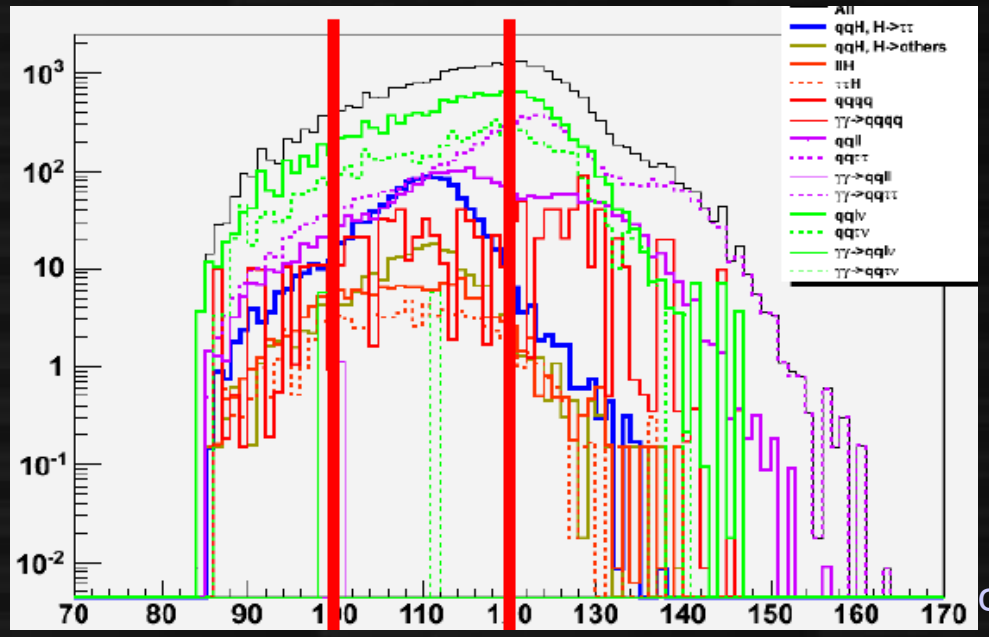
with pT correction from neutral particles

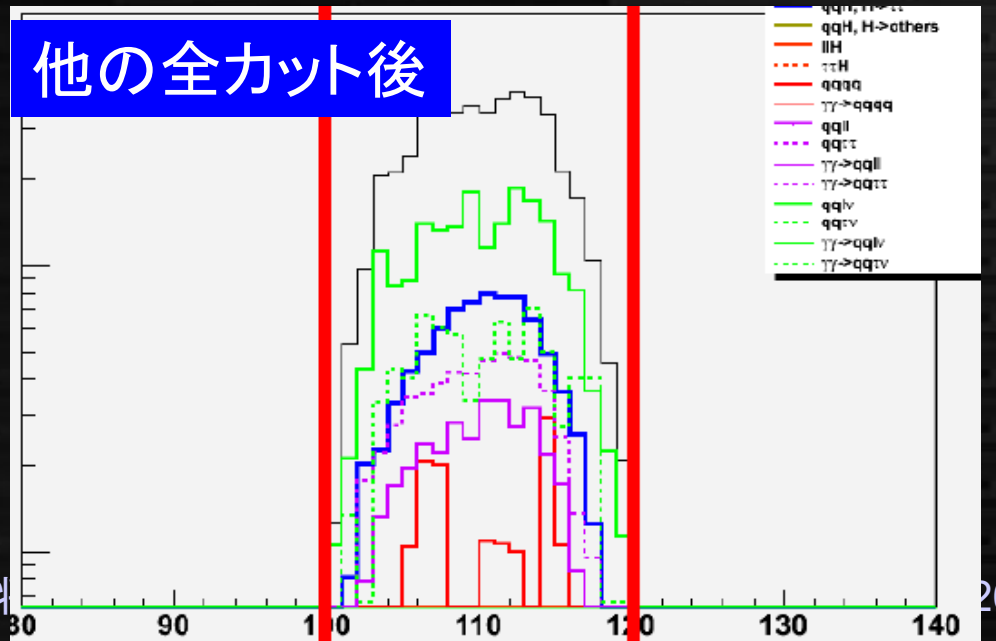




イベント識別1: カット

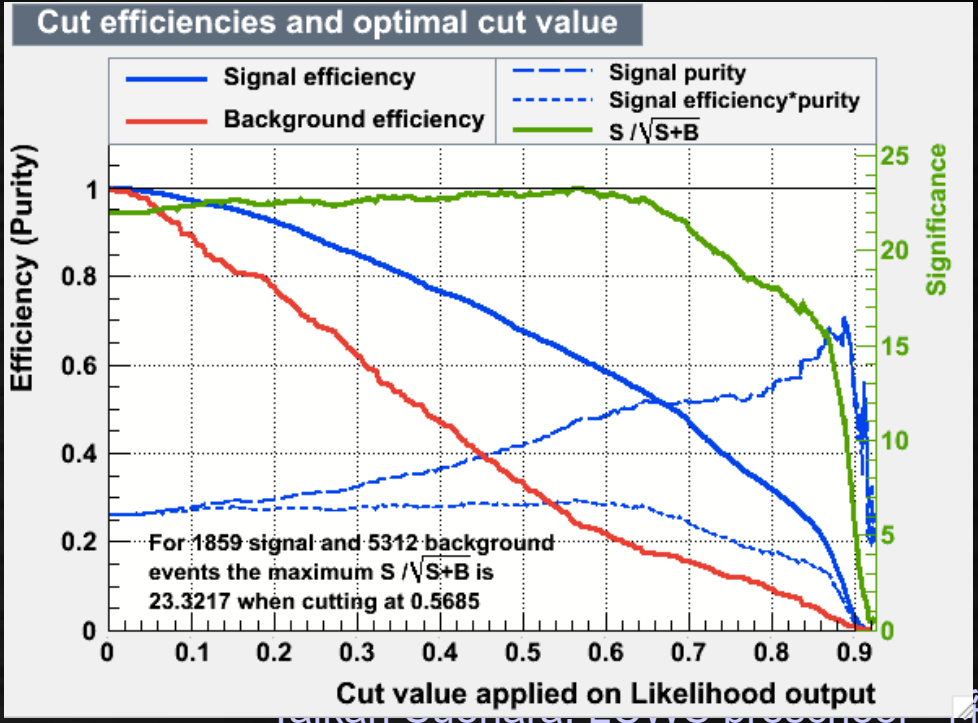
- シグナルが少ない領域を捨てS/Nを向上
- シグナルを多く切りすぎないよう注意
- 他のすべてのカットをかけた後
  再度分布をチェックする
- 系統誤差の影響を評価しやすい





イベント識別2: 多変量解析(TMVA)

- いくつかの変数を同時に使って“シグナルらしさ”を抽出
- いくつかの方法
  - ニューラルネット
  - BDT
  - likelihood etc.



使う変数を構成

↓

トレーニング

↓

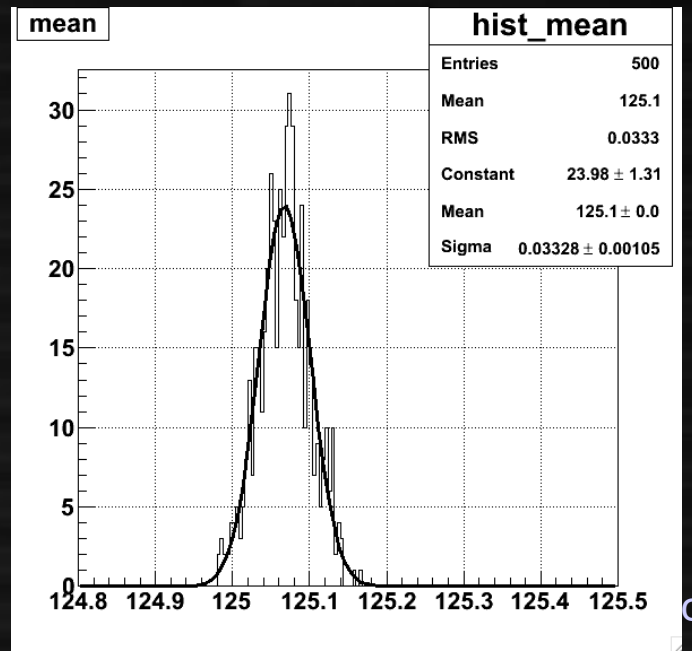
カット値の決定

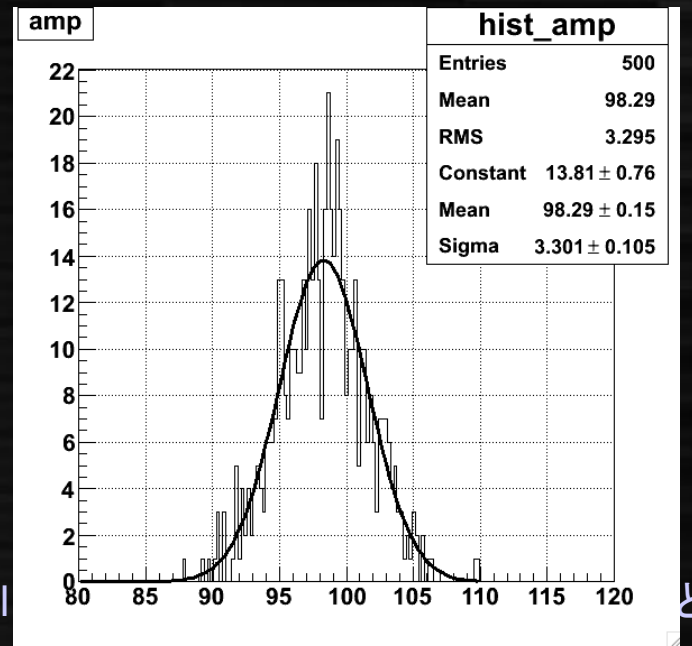
↓

TMVA outputでカット

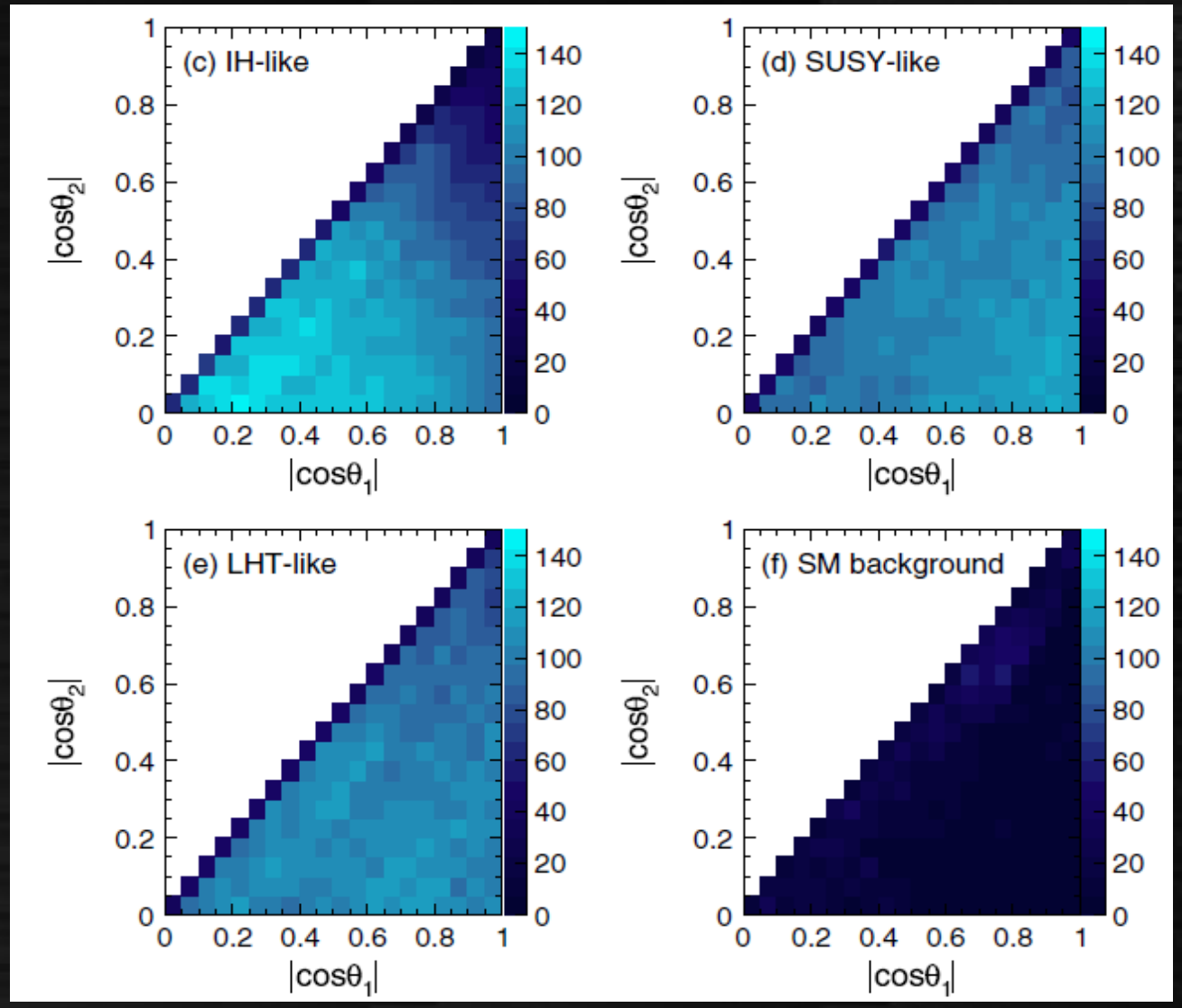
精度の見積4: Toy-MC

- 各binの値をpoisson分布で振る
  (振る前は高統計サンプルが望ましい、またはフィットした分布を使う)
- フィット等をやり直して測定値を求める
- 多数回繰り返して測定値分布を作る





精度の見積3: template fit



それぞれのtemplate
と"data"のχ^2を
調べる
(この場合は
Toy-MCを併用)

モデル識別性能
を評価

パラメータ決定
にも使える
(隣接パラメータの
内挿を使う等)

解析手法まとめ

- イベント生成
- 測定器シミュレーション
- 再構成
- シグナル選別
- 値の抽出

→ ILCでの測定感度

ビームバックグラウンド

incoherent scattering of beamstrahlung photons to

- e+e- pairs
  vertex detectorのoccupancy とforward detectorに効く
  物理インパクトは未確認

- qq pairs (minijets)
  物理インパクト大
  物理イベントに重ねる処理
  1.7 / bunch @ 500 GeV,
  4.1 / bunch @ 1 TeV

