

高輝度LHC-ATLAS実験に向けた GPUを用いた飛跡再構成の性能評価

神戸大学 M1 浅見優輝

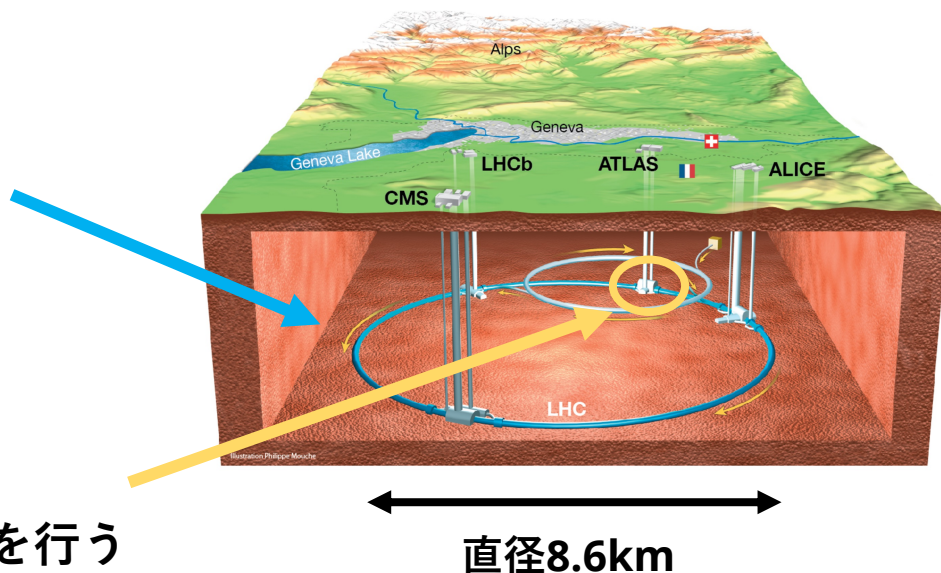
高輝度LHC-ATLAS実験

LHC (Large Hadron Collider)

- 世界最大の陽子陽子衝突型加速器
- 陽子衝突頻度を増加するための高輝度化アップグレードを予定

LHC-ATLAS実験

- LHC加速器での陽子衝突実験
- 標準模型の精密測定や新物理探索を行う

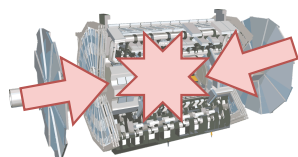


ATLASトリガーシステム

- 興味のある事象を選択・保存する

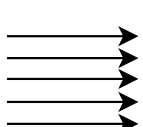
全てのデータを保存したら
~100TB/s!

陽子陽子衝突



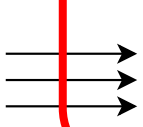
ATLAS検出器

40MHz



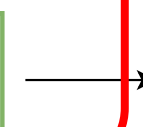
Level-0
Trigger

1MHz



Event Filter

~ 10kHz



Data
Storage

ハードウェアベースの
高速で粗い選別

ソフトウェアベースの
より詳細な選別

本発表のメイン

荷電粒子の飛跡
情報を用いる

トリガーシステム

第31回ICEPPシンポジウム

高輝度LHCに向けた課題

- 高輝度LHCでは陽子の衝突頻度が増加予定
 - ATLAS実験では内部飛跡検出器をアップグレード
= 検出器の細分化（読み出しチャンネル数の増加）
- 検出器内のシグナル数が増加
→ 飛跡再構成の計算量も大幅に増加
→ 必要な計算機資源が大幅に増加



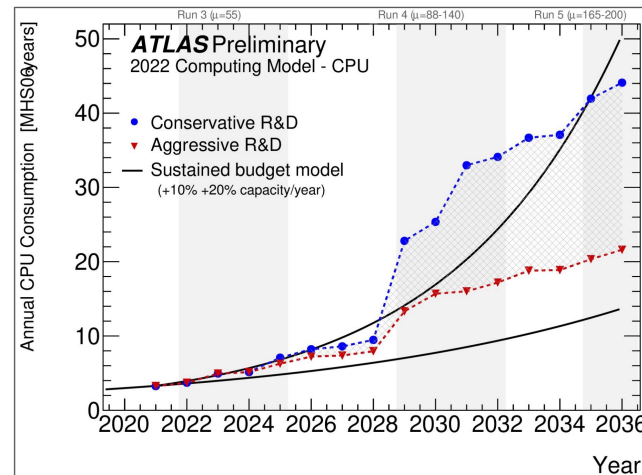
Event Filterにおける新たな飛跡再構成手法をいくつか検討

そのうちの 하나가GPUによる飛跡再構成 (Traccc ← 次のスライド)

GPUによる飛跡再構成のモチベーション

- GPUは並列処理に優れる
- 飛跡再構成の計算内には並列処理可能な部分が多数存在

➡ GPUを用いることで飛跡再構成の性能向上が期待！



GPUを用いた飛跡再構成

Traccc (ACTS* Parallelization Project)

- ACTSをGPU上で走らせるプロジェクト
- ATLAS Event Filter開発での採用を目指す
(2025年秋Technology Choice)
- 複数のGPU言語で開発
(CUDA, CYCL, Alpaka, etc.)
→ 今回は**CUDA**を使用



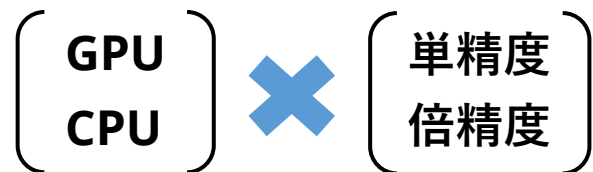
*A Common Tracking Software

- 高エネルギー実験のための汎用的な飛跡再構成ツール

本発表の概要

- 飛跡再構成の計算時間
- 飛跡再構成の計算精度
- 新たに取り組んでいること (行列乗算におけるレジスタ使用の最適化)

本発表では、GPUとCPUによる飛跡再構成の性能を以下の4通りで比較する

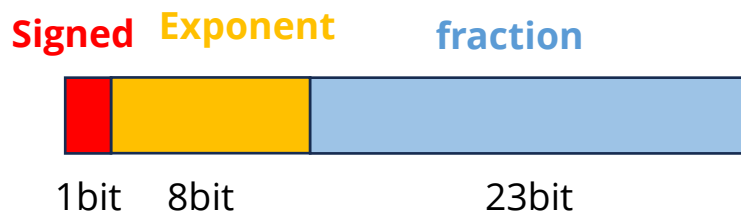


単精度・倍精度

コンピュータ内で数値を何桁の精度で扱うかの指標

- 計算を繰り返すとどんどん誤差が大きくなる
- 科学技術計算では一般的に倍精度(64bit)が用いられる

単精度 32bit (single, float)



$$\text{Signed} \quad \text{fraction} \quad \text{Exponent} \\ + 2.99792458 \times 10^8$$

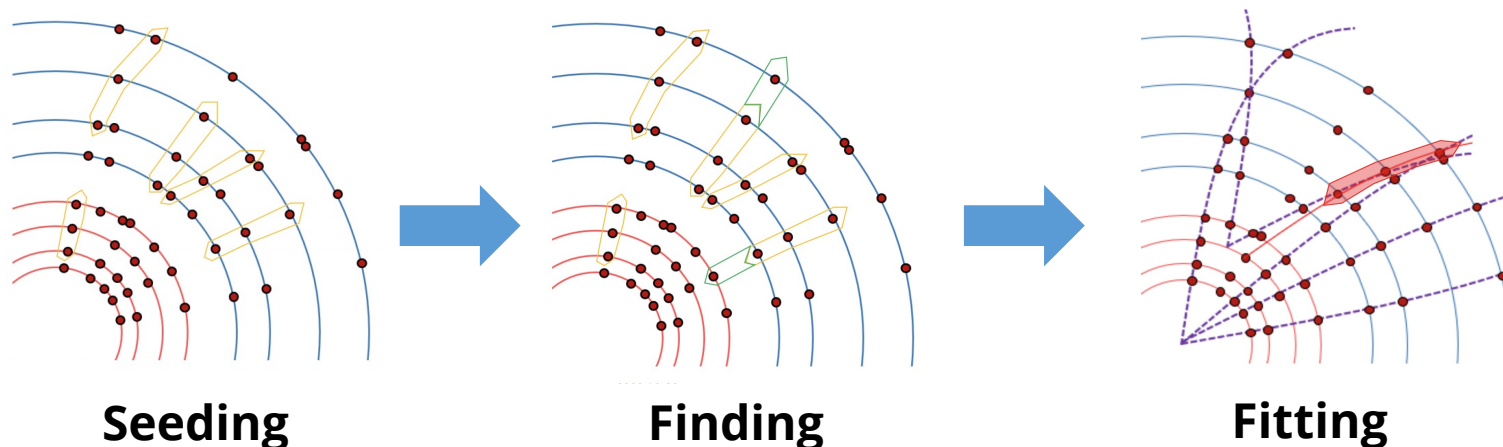
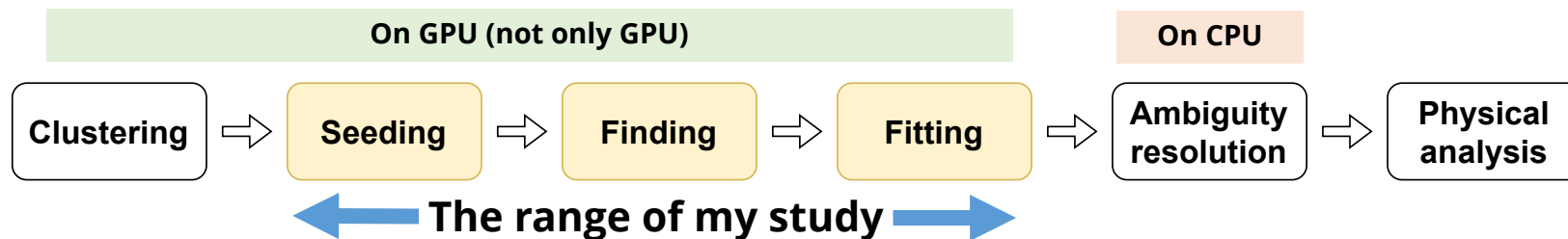
※実際は二進法

倍精度 64bit (double)



Tracccでの飛跡再構成アルゴリズム

Tracking chain



<https://atlassoftwaredocs.web.cern.ch/internal-links/tracking-tutorial/>

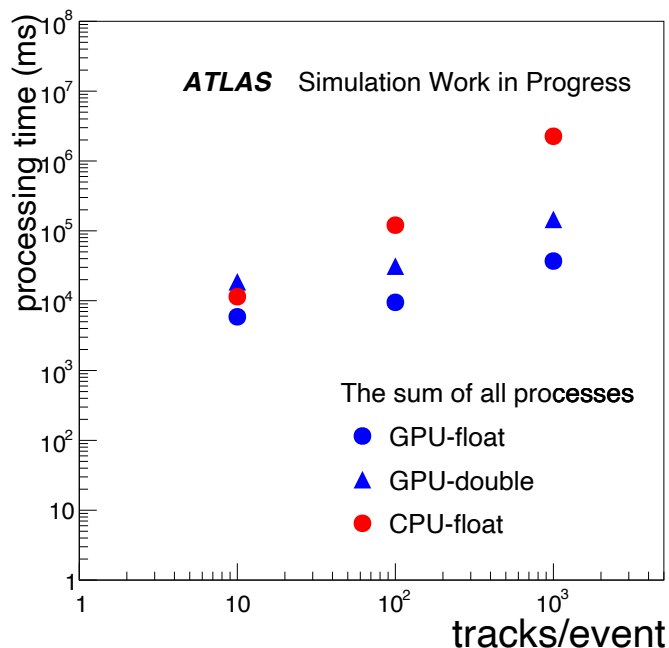
Seeding: 飛跡として矛盾しないヒット 3 個を一組にまとめる (= a seed)

Finding: seedと矛盾しないようなヒットを繋げていく

Fitting: track parametersを得る

飛跡再構成の計算時間の比較

計算時間の比較



検出器のジオメトリ

ITk (HL-LHCで使用される新しいATLAS内部飛跡検出器)

生成トラック

100events,

10GeV<p<100GeV, -4.0< η <4.0

パイルアップはなし

非物理的な粒子を使用(massless, q = -1)

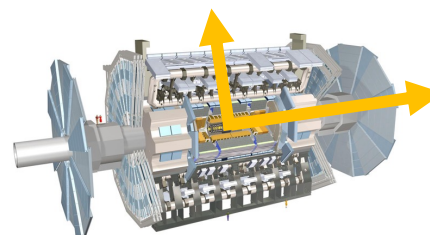
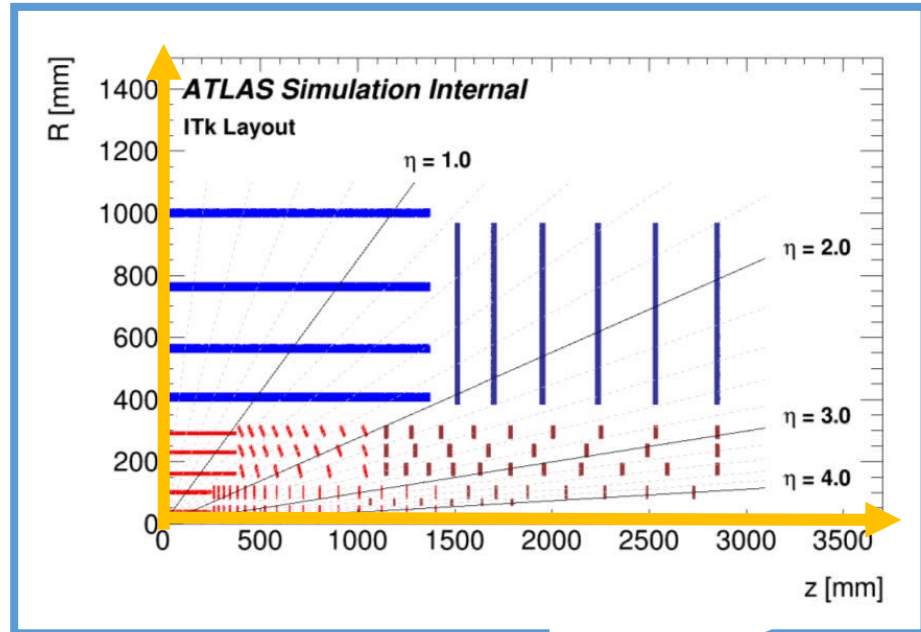
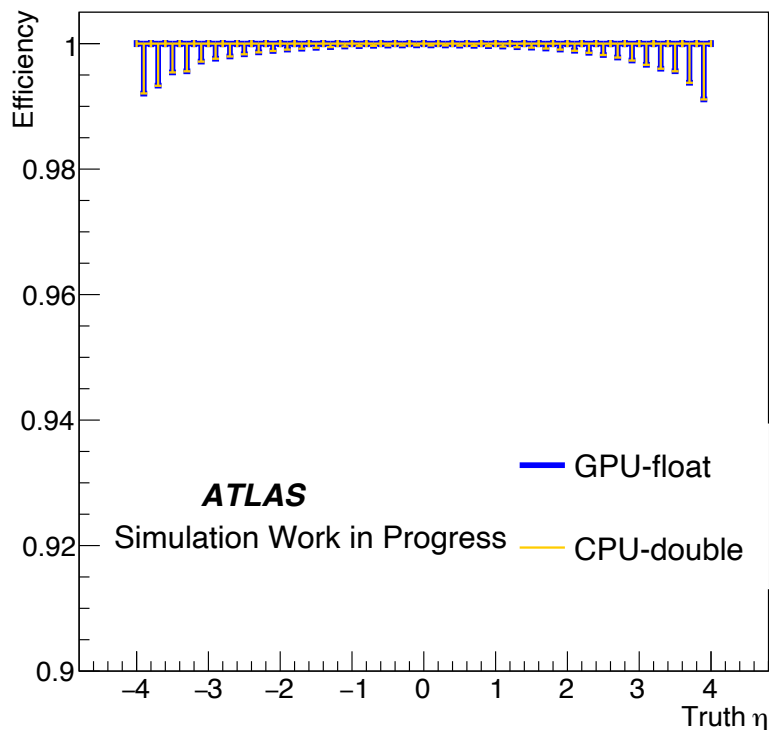
- GPUの方が高速（1000 tracksでは100倍くらい速い！）
- Track数が多いほど、GPUが優位になる
- GPUにおいて単精度の方が倍精度より高速

Question : 計算精度は単精度で十分か？

単精度(32bit)と倍精度(64bit)で計算精度があまり変わらないなら、より高速な単精度を使いたい！

計算精度 (Seeding)

Efficiency (truth $p_T > 1\text{GeV}$ のみプロット)



ATLAS検出器の" η "座標

Efficiencyの観点からは、**単精度GPU**は**倍精度CPU**と一致している

→ **seeding**は**単精度**でも**十分な精度**！

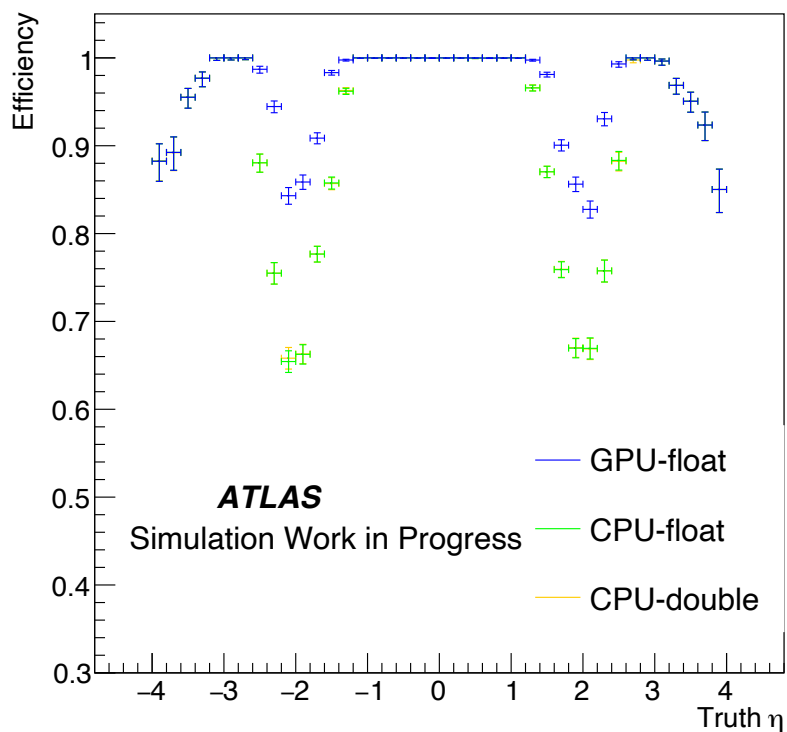
Efficiencyの定義

あるtruth particle由来のヒットのみからなるreco trackがある場合、そのtruth particleは"完全に再構成された"とみなされ、efficiencyの分子に加えられる。

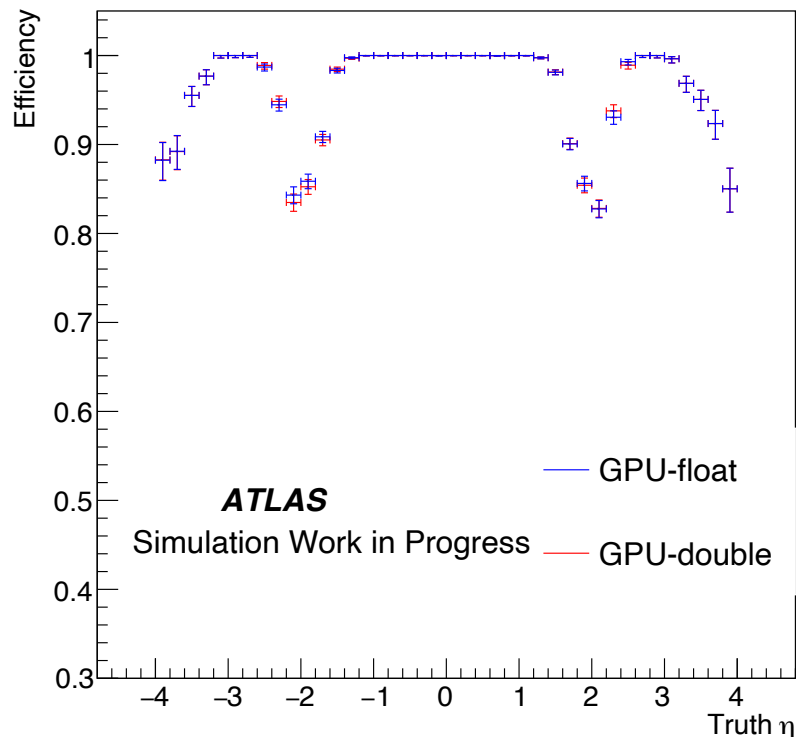
計算精度 (Finding)

Efficiency

(plot only truth $p_T > 0.1\text{GeV}$)



GPU-float vs CPUs



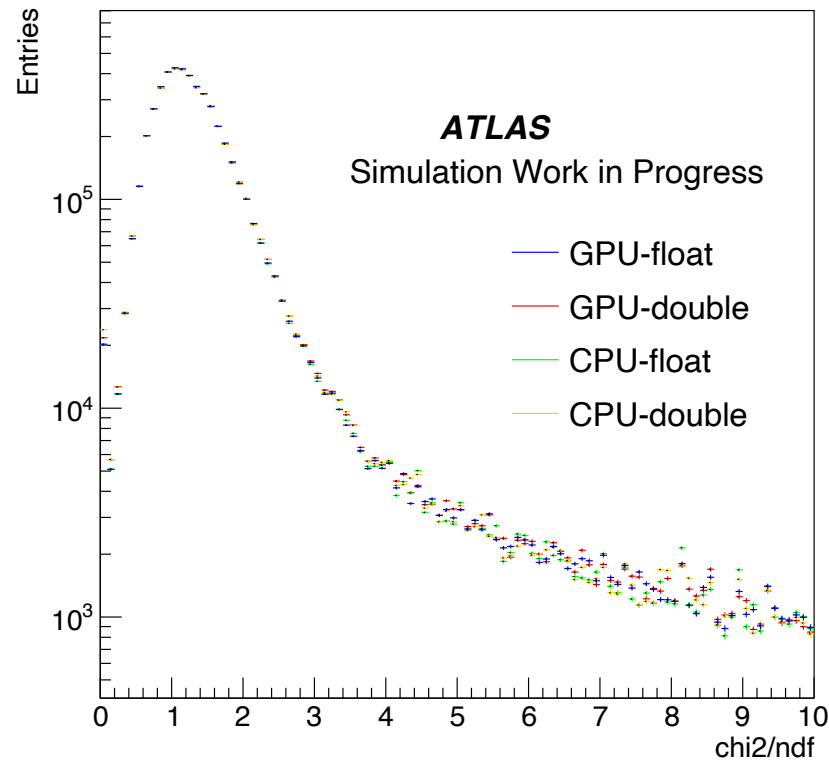
GPU-float vs GPU-double

- GPUの方がCPUよりも効率が良い (左図)
 - 単精度GPUは、倍精度GPUとほとんど同じperformanceを示す (右図)
- findingにおいても単精度GPUで十分な精度

計算精度 (Fitting) 1/3

ここからはfitting parameterについて見ていく

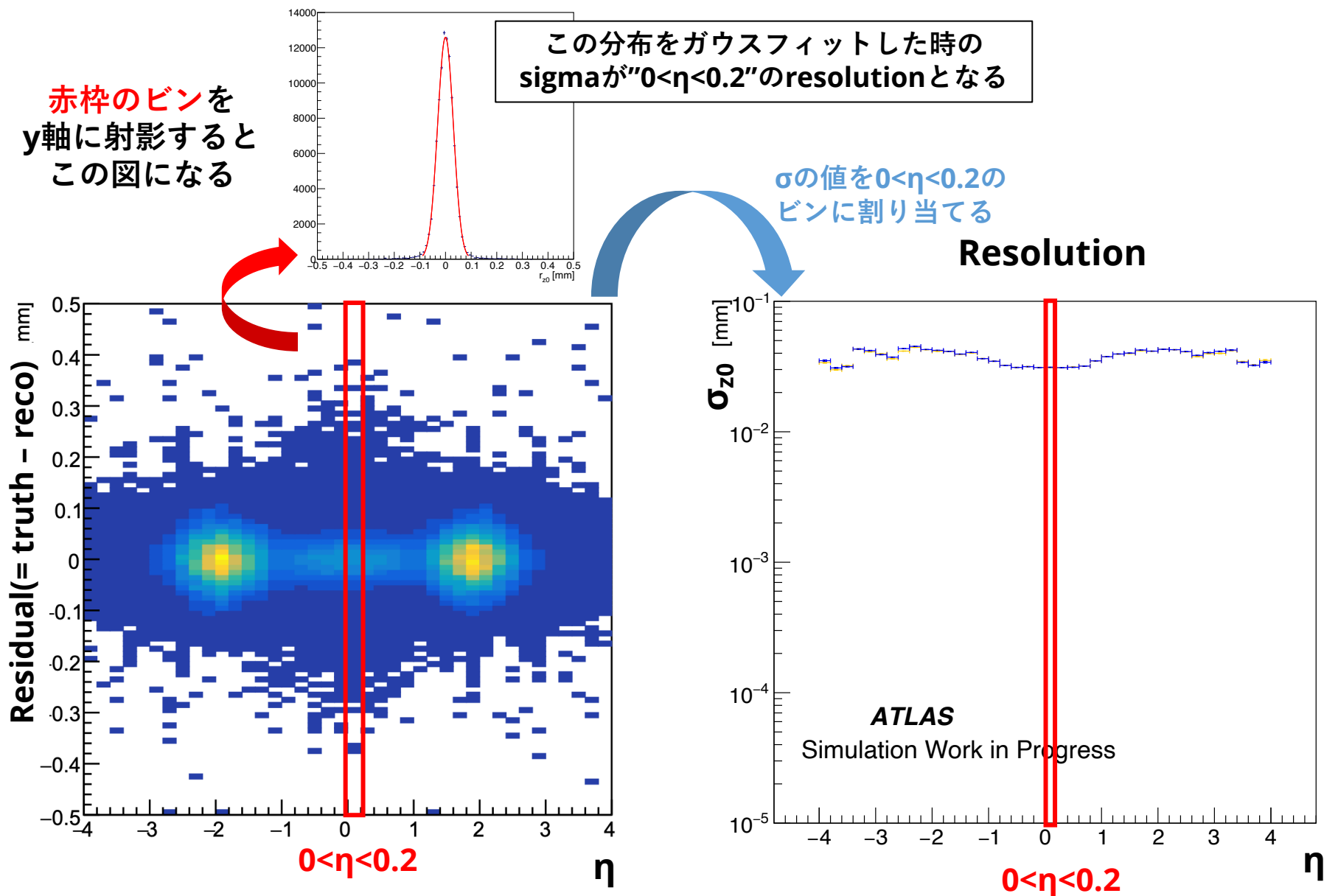
Chi2/NDF



Chi2/NDFの分布は4種類の間でおおむね一致している

計算精度 (Fitting) 2/3

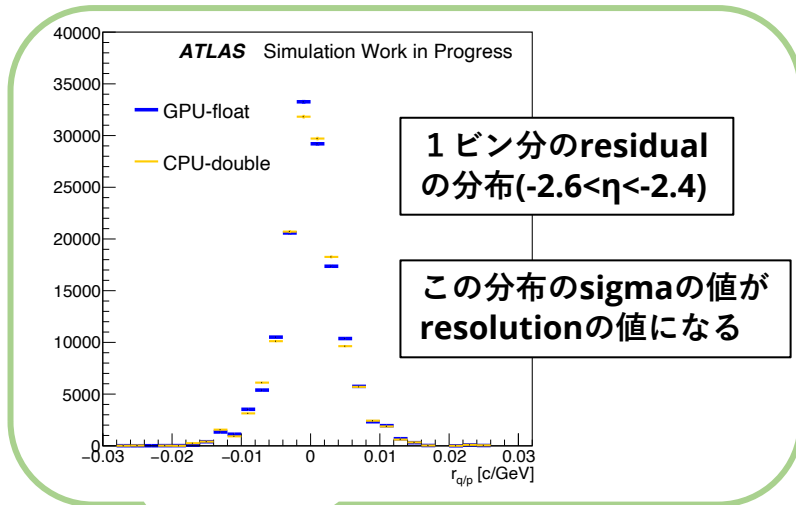
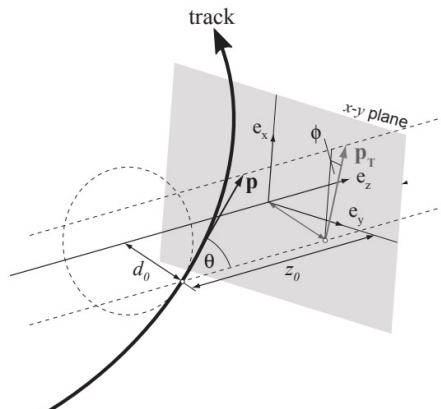
※residualはtracccのコードから自動的に作成されるが、resolutionは自身で作成した



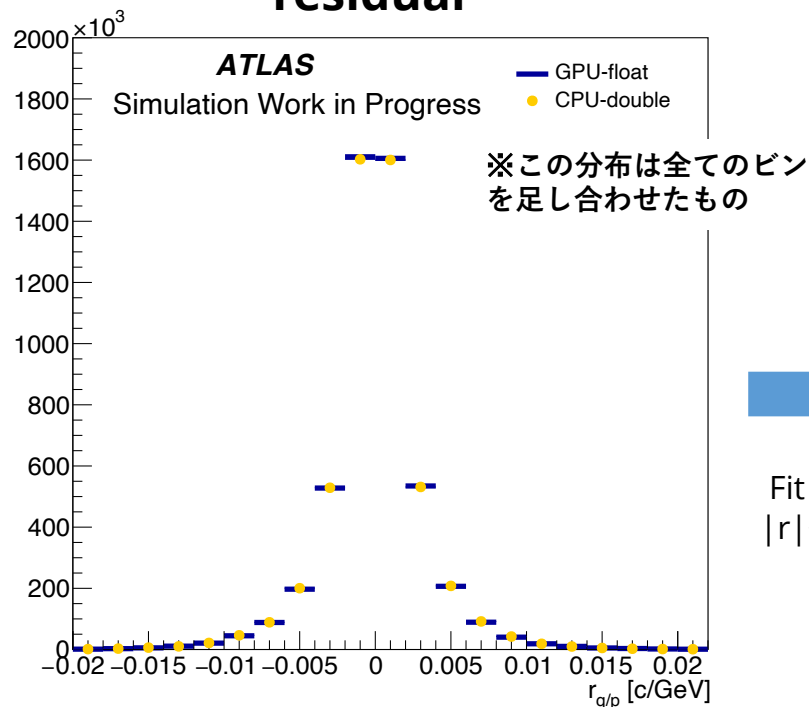
計算精度 (Fitting) 3/3

q/p

トラックの曲率
(曲がり具合) を表す

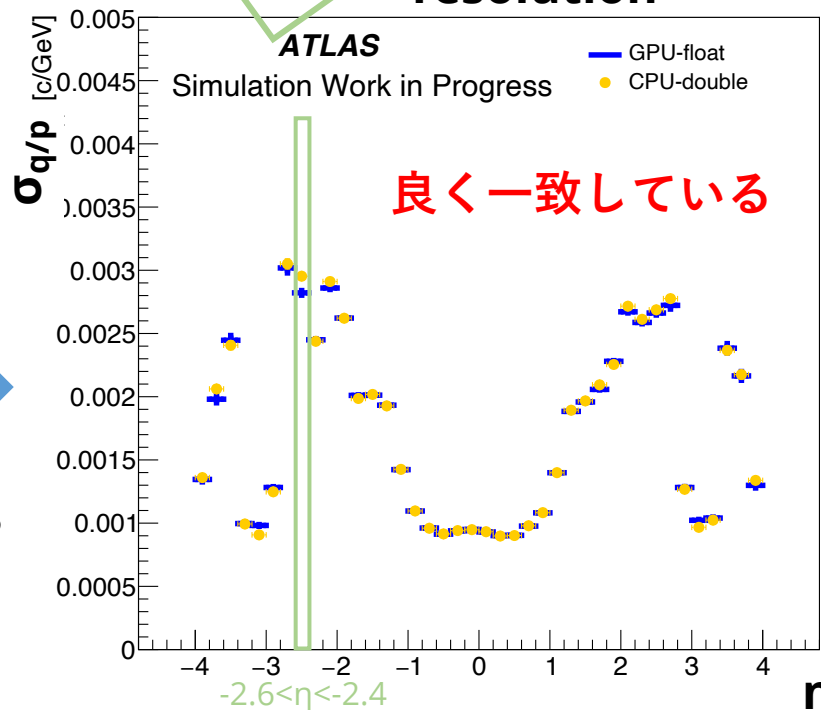


residual



Fit range
 $|r| < 0.005$

resolution



- この場合も単精度GPUと倍精度CPUのresolutionはほとんど一致している
- 他のtrack parameterでも一致していることを確認 (back up)

新たに取り組んでいること

「GPUを用いた行列演算におけるレジスタ使用の最適化」

レジスタとは

- CPUやGPUの中にある、演算等を行う際に必要なデータを一時的に格納するメモリ
- メインメモリ(RAM)よりも読み書きが高速だが、容量は小さい

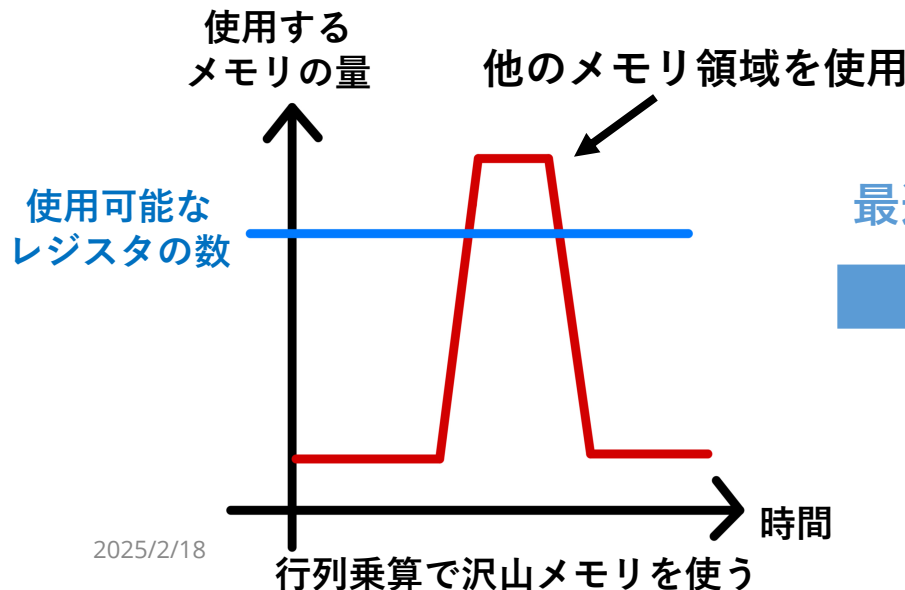
モチベーション

- FittingやPropagatorで行列乗算が多く行われる
 - これらが計算時間の大部分を占めている
- 最適化によって処理時間が短縮の可能性

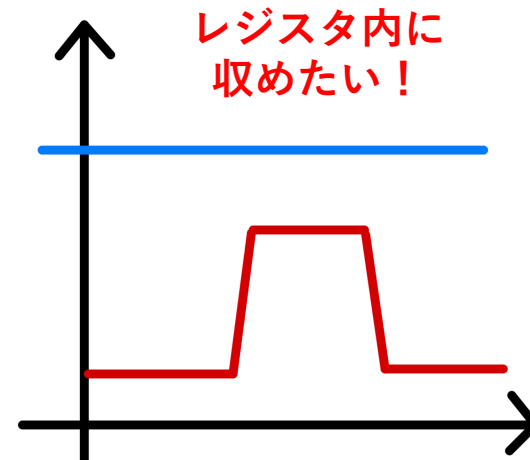
NVIDIA Nsight compute

Function Name	Demangled Name	Duration [ms]
fit	..concentric_cylin...	17.86
propagate_to_next_surface	..pointwise_mater...	5.82
propagate_to_next_surface	..pointwise_mater...	5.54
		5.19
		5.19
		5.16
		3.59
find_tracks	..find_tracks..findi...	0.14

Fittingだけで
17.86ms / 49.55ms



最適化



どのようにしてレジスタの範囲内に収めるか

キーポイント「行列の成分が0になっている部分の計算を省く」

$$\begin{pmatrix} 3 & 0 & 1 & 1 & 0 \\ 0 & 6 & 0 & 1 & 0 \\ 1 & 0 & 2 & 0 & 0 \\ 0 & 1 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} 4 & 1 & 2 & 1 & 0 \\ 1 & 7 & 0 & 0 & 0 \\ 2 & 0 & 2 & 1 & 0 \\ 1 & 0 & 1 & 8 & 0 \\ 0 & 0 & 0 & 0 & 5 \end{pmatrix} = ?$$

例えば、(5, 5)成分の計算は各行列の(5, 5)の積に簡略化できる！

- 飛跡再構成の計算内にも行列の成分が常に0になっている部分が存在
- この部分で変数宣言をしない（=レジスタの使用予約をしない）ことによってメモリを削減 → レジスタだけで賄える（高速化が期待）

まとめ

- 必要な計算資源の大幅な増加に対応するため、LHCの高輝度化に向けてトリガーシステムのアップグレードが必要
 - **GPUを用いた飛跡再構成を検討 (Traccc)**

結果

- GPUによって計算時間は大幅に短縮可能
 - 飛跡数が増加するほどGPUが優位
- 計算精度が単精度 (32bit)でも十分であることを確認

今後の展望

- 行列乗算におけるレジスタ使用の最適化 → 飛跡再構成の速度の向上

Back Up

NVIDIA RTX 6000

SPECIFICATIONS

GPU memory	48GB GDDR6
Memory interface	384-bit
Memory bandwidth	768 GB/s
Error-correcting code (ECC)	Yes
NVIDIA Ampere architecture-based CUDA Cores	10,752
NVIDIA third-generation Tensor Cores	336
NVIDIA second-generation RT Cores	84
Single-precision performance	38.7 TFLOPS⁷
RT Core performance	75.6 TFLOPS⁷
Tensor performance	309.7 TFLOPS⁸
NVIDIA NVLink	Connects two NVIDIA RTX A6000 GPUs¹²
NVIDIA NVLink bandwidth	112.5 GB/s (bidirectional)
System interface	PCIe 4.0 x16
Power consumption	Total board power: 300 W
Thermal solution	Active
Form factor	4.4" H x 10.5" L, dual slot, full height
Display connectors	4x DisplayPort 1.4a⁹
Max simultaneous displays	4x 4096 x 2160 @ 120 Hz, 4x 5120 x 2880 @ 60 Hz, 2x 7680 x 4320 @ 60 Hz
Power connector	1x 8-pin CPU
Encode/decode engines	1x encode, 2x decode (+AV1 decode)
VR ready	Yes
vGPU software support	NVIDIA vPC/vApps, NVIDIA RTX Virtual Workstation
vGPU profiles supported	See the Virtual GPU Licensing Guide
Graphics APIs	DirectX 12 Ultimate, Shader Model 6.6, OpenGL 4.6¹⁰, Vulkan 1.3¹⁰
Compute APIs	CUDA 11.6, DirectCompute, OpenCL 3.0

NEXT

Intel(R) Xeon(R) Gold 5318Y CPU@2.10GHz

Essentials

[Download Specific](#)

Product Collection	3rd Gen Intel® Xeon® Scalable Processors
Code Name	Products formerly Ice Lake
Vertical Segment	Server
Processor Number ?	5318Y
Lithography ?	10 nm
Recommended Customer Price ?	\$1483.00

CPU Specifications

Total Cores ?	24
Total Threads ?	48
Max Turbo Frequency ?	3.40 GHz
Intel SpeedStep® Max Frequency ?	3.40 GHz
Processor Base Frequency ?	2.10 GHz
Cache ?	36 MB
Intel® UPI Speed	11.2 GT/s
Max # of UPI Links ?	3
TDP ?	165 W

各プロセスにおける処理されたseed/trackの数

	seeding	finding	fitting
GPU-float	2,158,529	5,212,913	5,212,913
GPU-double	2,158,550	5,209,885	5,209,885
CPU-float	2,158,529	5,206,287	5,206,287

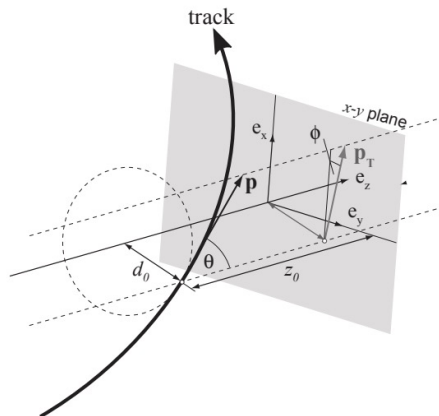
これら3つの中での差は最大でも0.13%

→ これらを比較しても問題ない

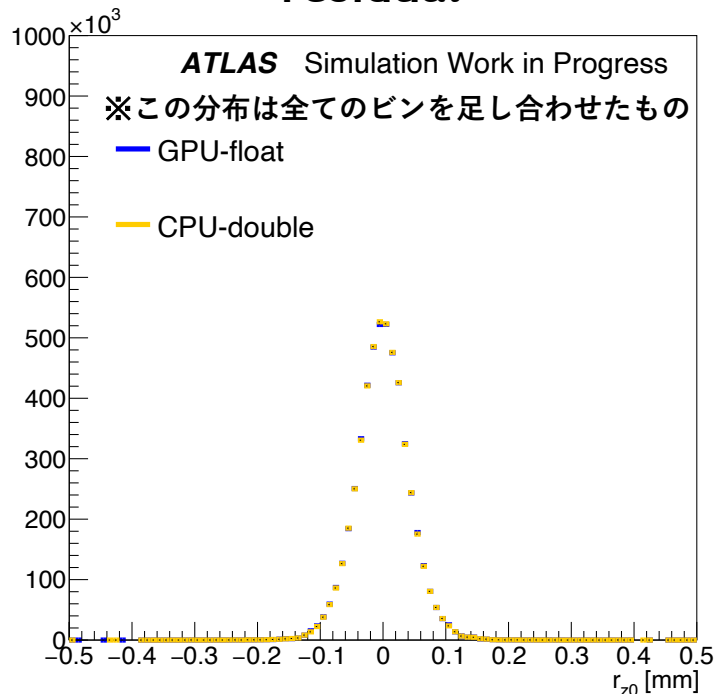
計算精度 (fitting)

z_0

ビーム軸方向における
トラックの位置を表す

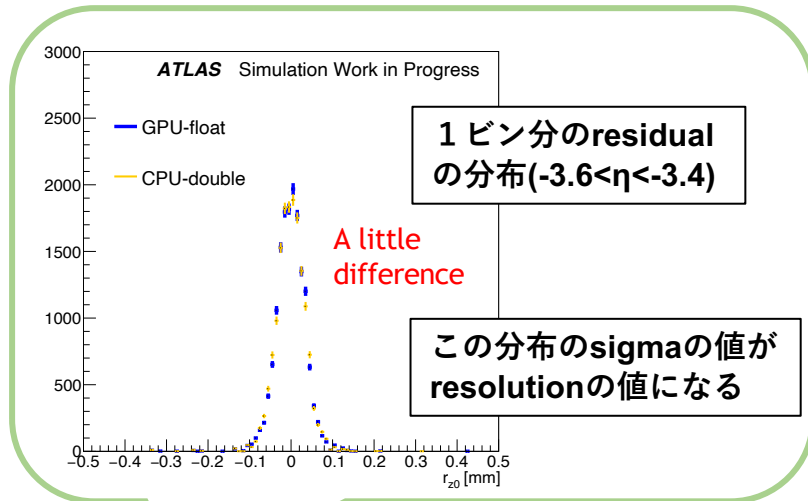
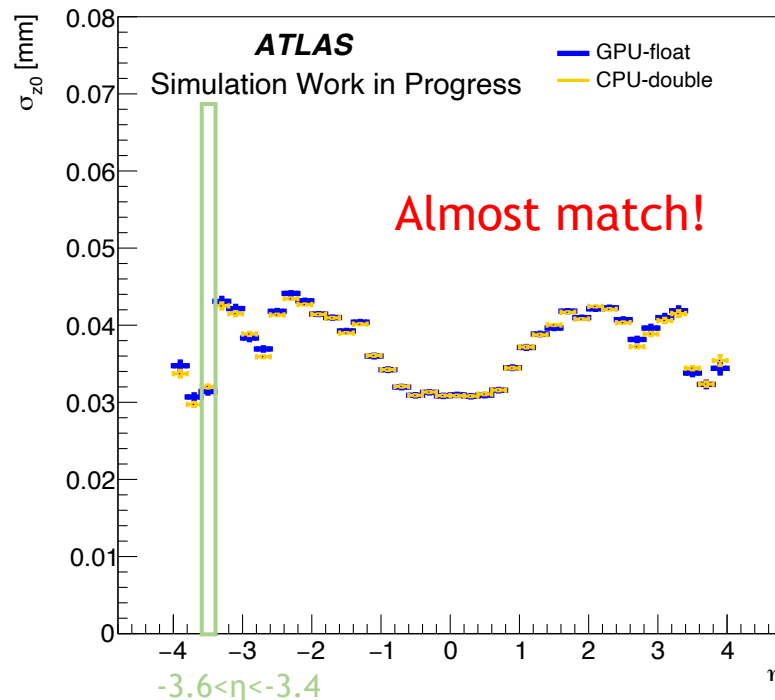


residual



Fit range
 $|r| < 0.1$

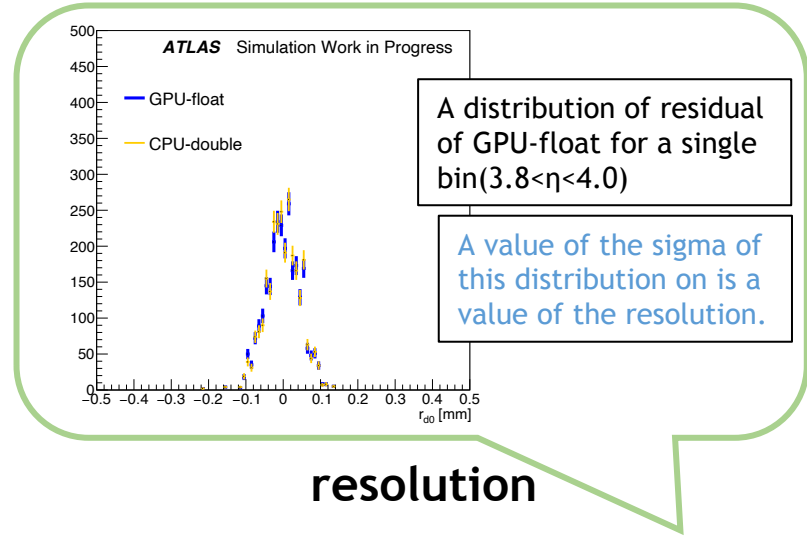
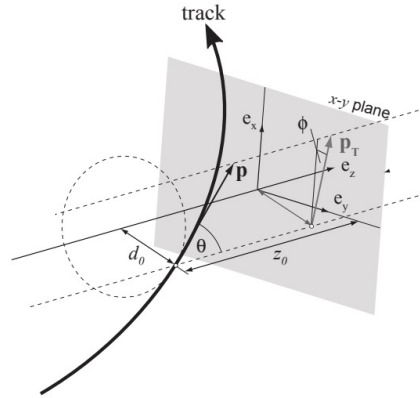
resolution



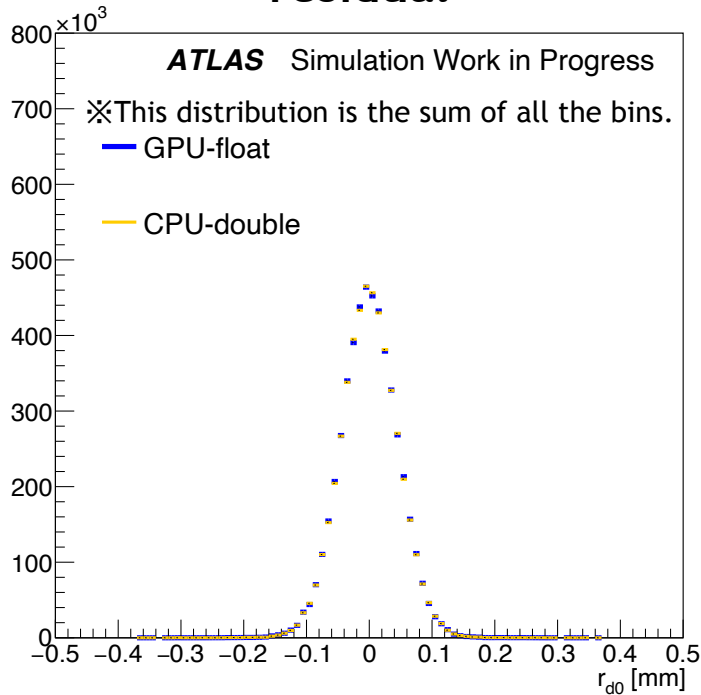
計算精度 (fitting)

d_0

d_0 , one of the track parameters represents a distance between a beam axis and a track.

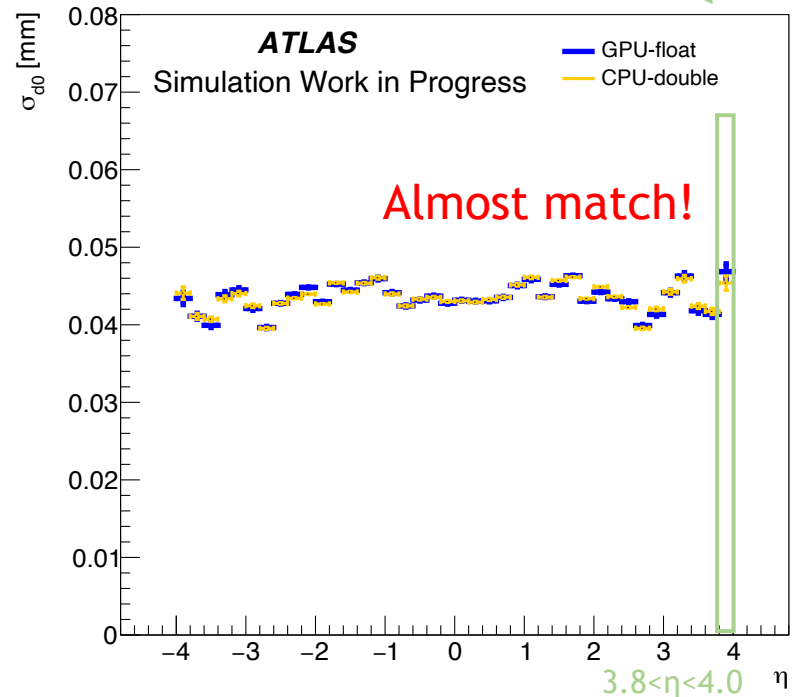


residual



Fit range $|r| < 0.1$

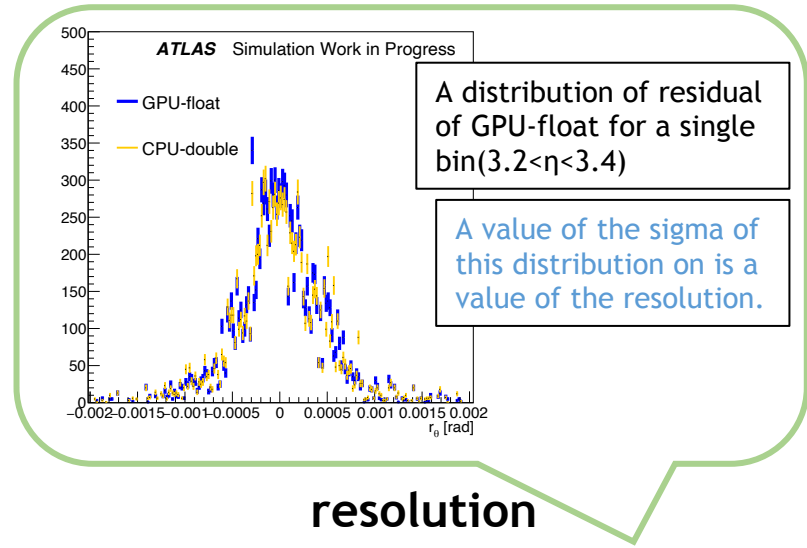
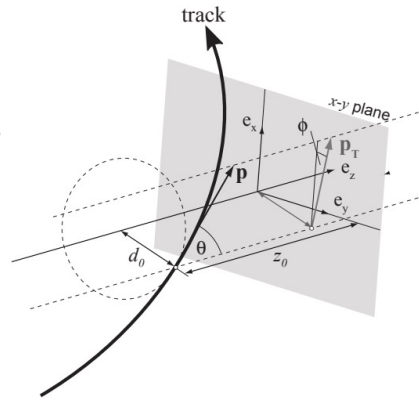
resolution



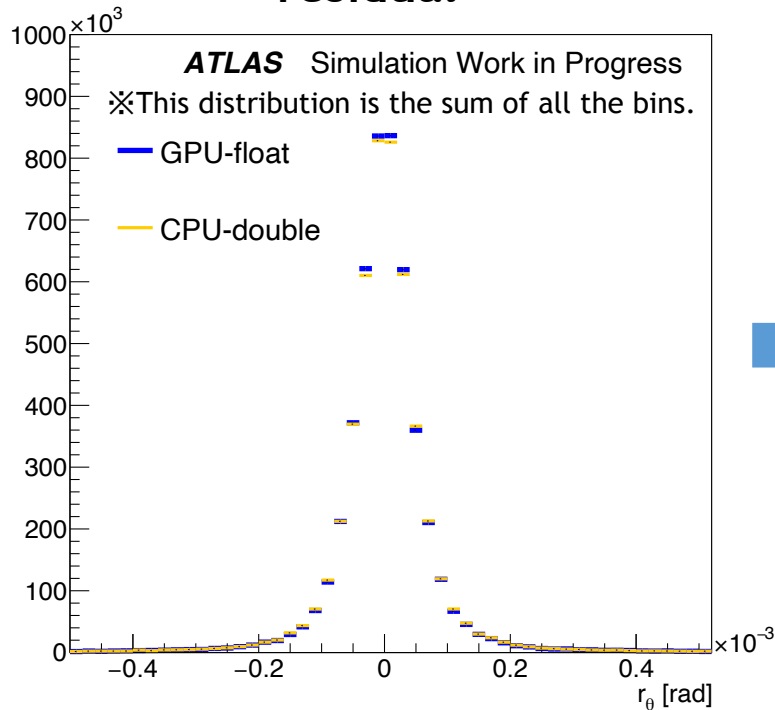
計算精度 (fitting)

theta

theta, one of the track parameters represents an angle from the reference plane.

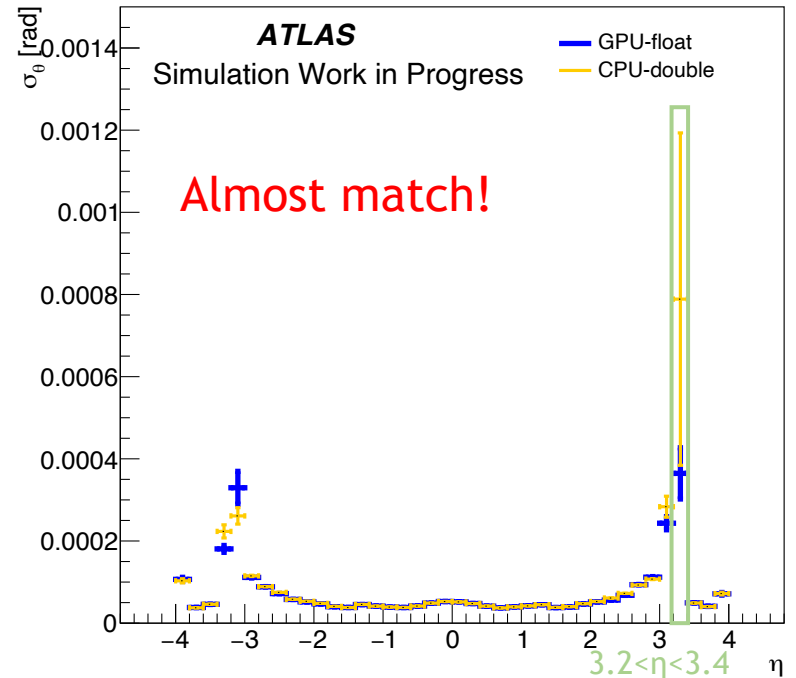


residual



Fit range
 $|r| < 0.1$

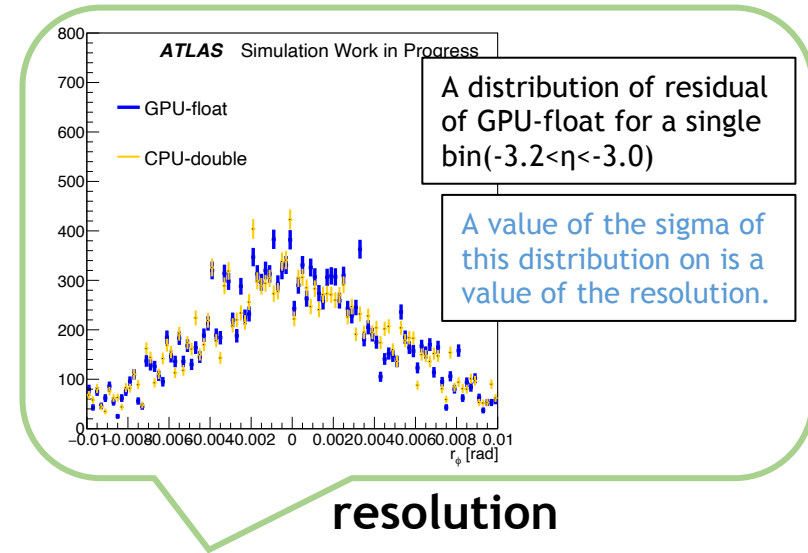
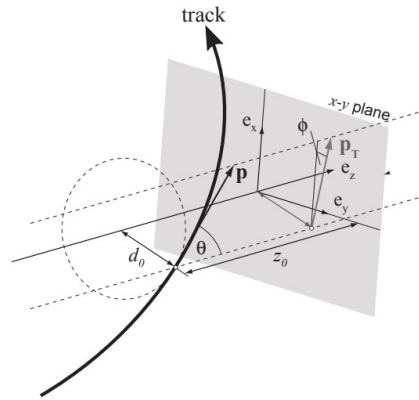
resolution



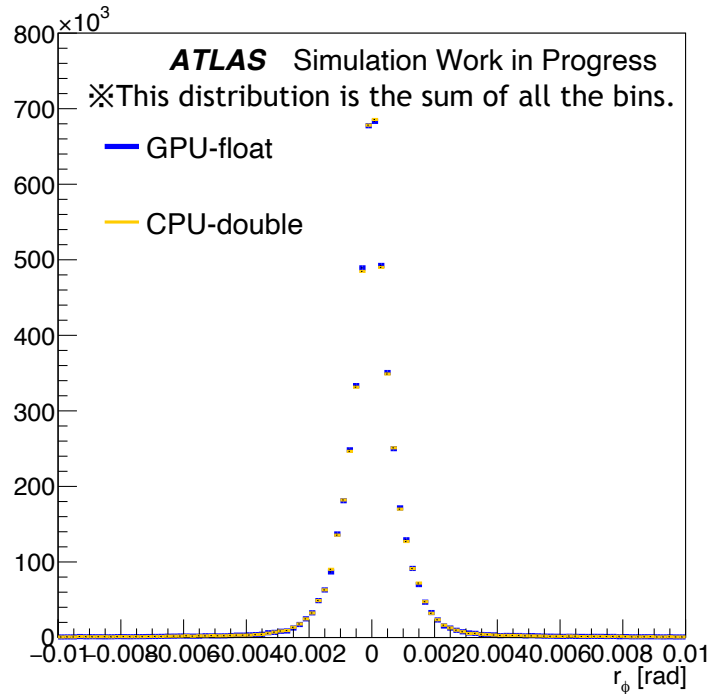
計算精度 (fitting)

phi

phi, one of the track parameters represents azimuthal angle.



residual



Fit range
 $|r| < 0.1$

