
CERN

Hashing and similarity learning for tracking

with the HL-LHC ATLAS detector

Moritz Kiehn[1], S. Amrouche[2], N. Calace[1], T. Golling[2], A. Salzburger[1]

[1] CERN
[2] Université de Genève

Why new tracking techniques?

Tracking is the core of almost every particle physics experiment

Tracking at the High-Luminosity LHC means pile-up 200

Combinatorial reconstruction approaches explode

$\Rightarrow$ Better algorithms, better physics!

Hashing?

A typical hash function

$$h(\text{complex object}) = \text{unique number}$$

used e.g. in hash maps (Python `dict`, `std::unordered_map`)

Hashing?

A perfect tracking hash function

$$h(\text{hit}) = \text{track number}$$

does not exists.

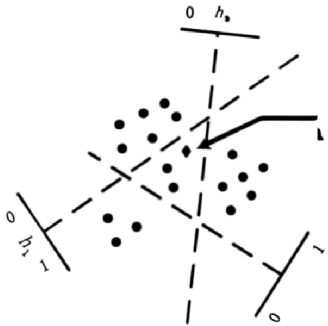
Hashing?

But, approximate grouping of hits

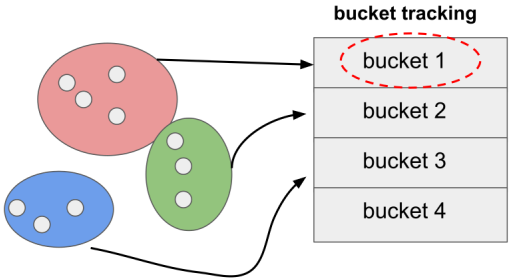
$$h(\text{track 1 hit 0}) = \text{group x}$$
$$h(\text{track 1 hit 1}) = \text{group x}$$
$$h(\text{track 0 hit 1}) = \text{group x}$$
$$\dots$$

can be constructed.

Hashing as bucketing strategy

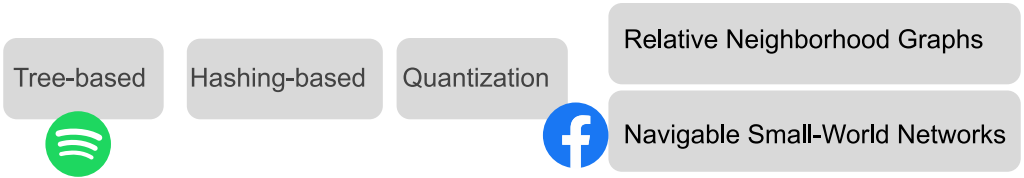


Random projections to create item hash

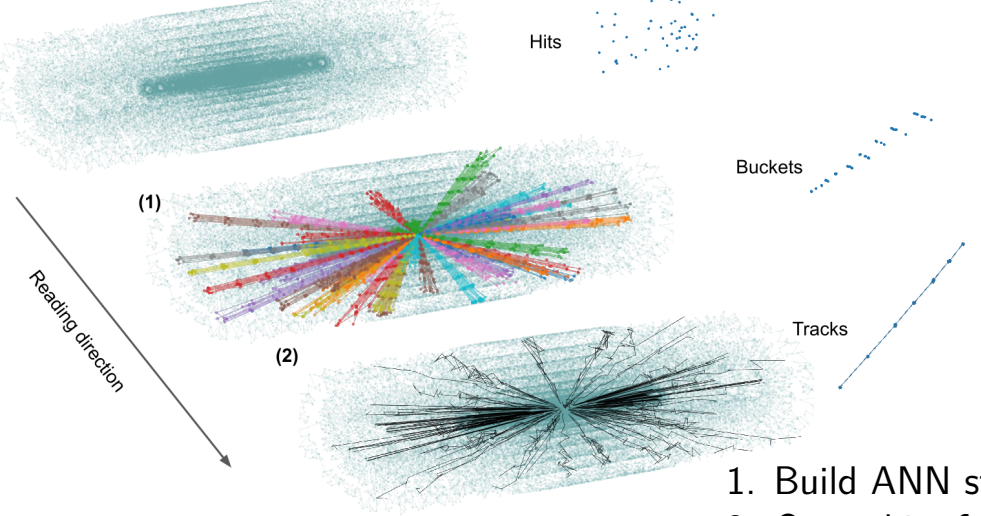


Groups similar items/ neighbors for a given metric

General: approximate nearest neighbors

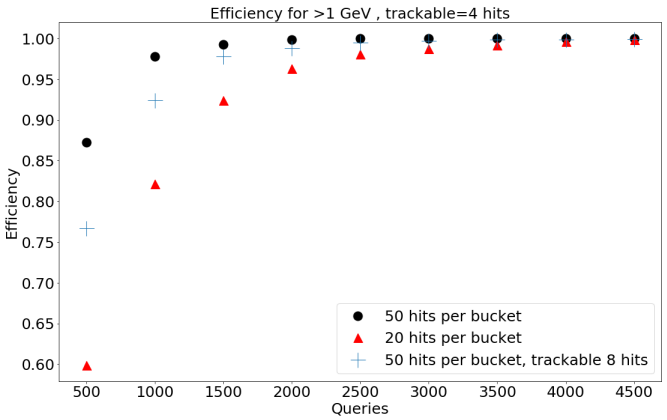


Our choice: `https://github.com/spotify/annoy`



1. Build ANN structure
2. Query hits for neighbors
3. Local reconstruction

Reconstruction quality



TrackML dataset: $t\bar{t}$, $\mu = 200$

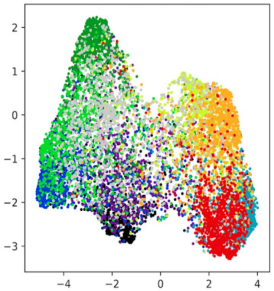
Select buckets/neighborhoods of hits using randomly drawn hits as query points

$$\text{Efficiency} = \frac{\text{trackable particles in buckets}}{\text{total trackable particles}}$$

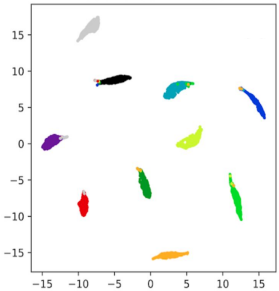
Find (almost) all tracks with sufficient queries

Possible improvement: metric learning

Ex.: initial space



Ex.: learned space



Neighbor definition depends on a metric
(so far: angular distance)

Learn metric/transformation based on known grouping for max. separability

$\Rightarrow$ Better metric, better buckets

Metric learning for tracks

Fig 1 Transverse view (x/y)

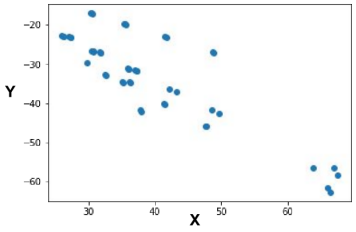
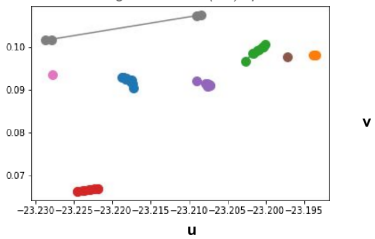


Fig 2 Learned (u/v) space



Uses Local Fisher Discriminant Analysis

Spatially close tracks (left) separated in transformed space (right)

Caveat: uses TrackML dataset with additional hit features
Not easily translatable to other cases (yet)

Can this run on accelerators?

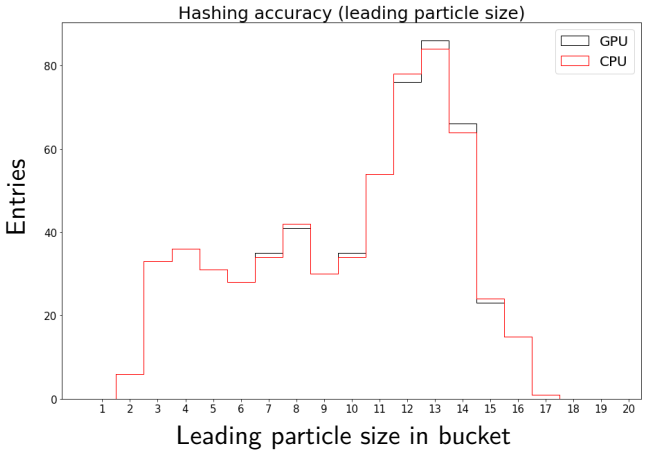
(Approximate) nearest neighbor search is a general problem

Think: similar images, movies, people, ...

E.g. "Billion-scale similarity search with GPUs"

arXiv:1702.08734 from Facebook, based on a different technique

CPU/GPU comparison



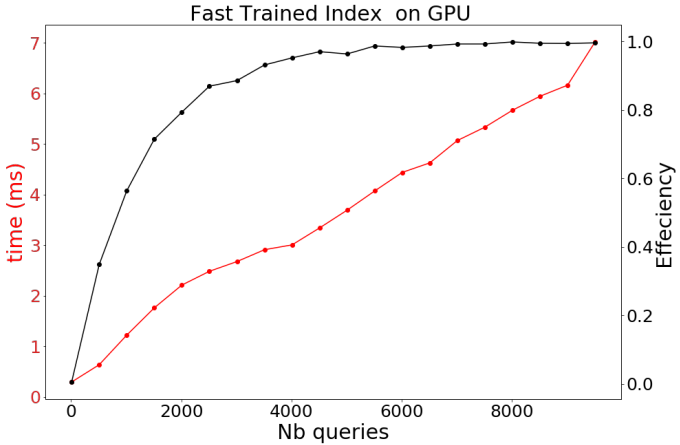
TrackML dataset
$t\bar{t}$, $\mu = 200$

GPU NVidia Tesla K40m
12Gb RAM, 2880 CUDA Cores

Separate CPU/GPU implementations, same input data, same algorithm

$\rightarrow$ Consistent bucket definition

GPU performance



TrackML dataset
$t\bar{t}$, $\mu = 200$

GPU NVidia Tesla K40m
12Gb RAM, 2880 CUDA Cores

Select buckets of hits using randomly drawn hits as query points

Queries are fully paralellized

Running on ATLAS ITk

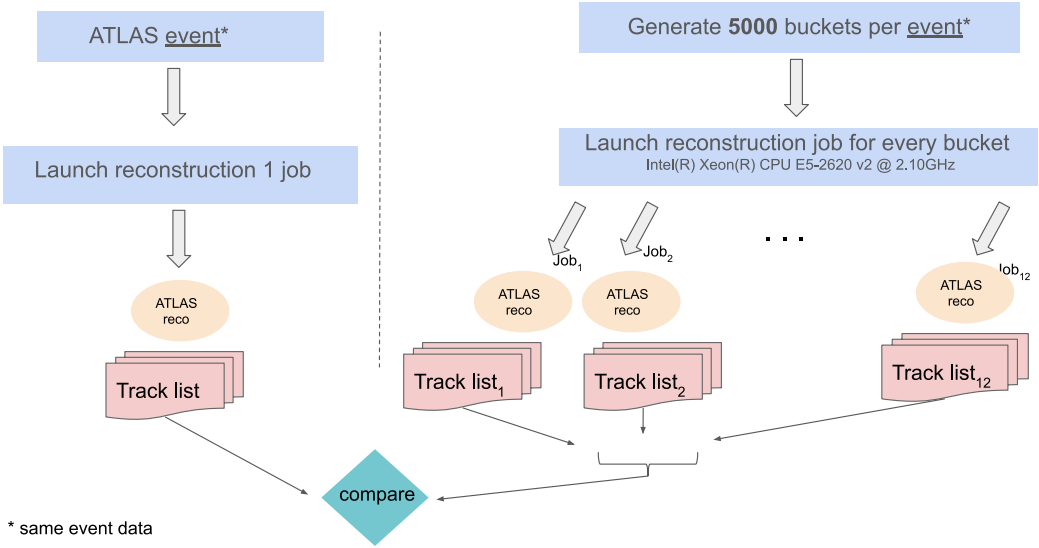
$t\bar{t}$ with $\mu = 200$ and ATLAS ITk geometry

Use initial method to select buckets of hits

Apply ATLAS reconstruction in buckets

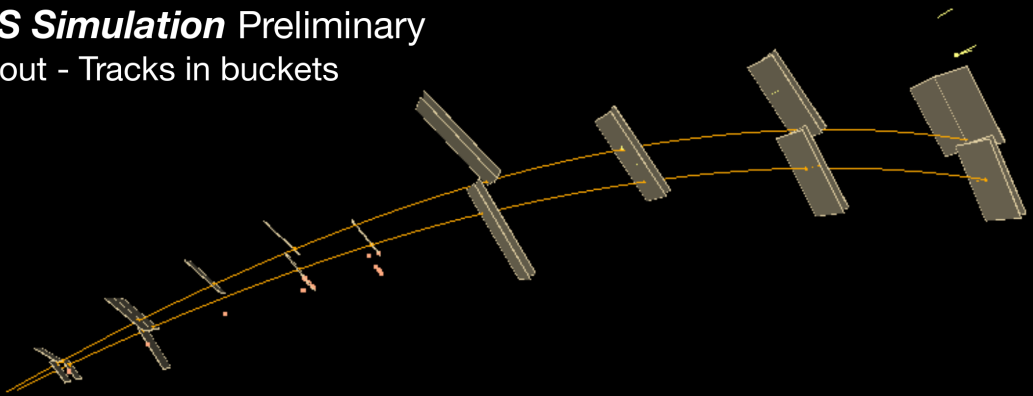
As a proxy for (future, optimized) local reconstruction

ATLAS-IDTR-2019-008

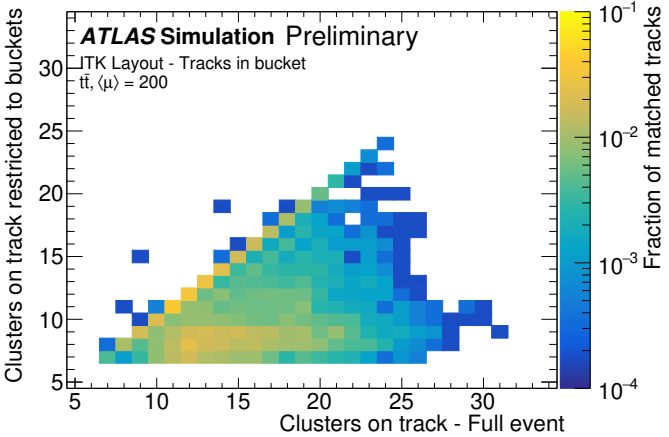

ATLAS event*
Launch reconstruction 1 job
ATLAS reco
Track list
Generate 5000 buckets per event*
Launch reconstruction job for every bucket
Intel(R) Xeon(R) CPU E5-2620 v2 @ 2.10GHz
Job1
Job2
Job12
ATLAS reco
ATLAS reco
ATLAS reco
Track list1
Track list2
Track list12
compare
* same event data

ATLAS Simulation Preliminary

ITk layout - Tracks in buckets



Reconstruction in bucket vs full event



clusters on tracks with full event reconstruction versus restricting the reconstruction to a bucket of 50 hits.

Diagonal: fully matched tracks

Below: bucket track is missing hit(s)

Above: bucket track contains additional hit(s)

ATLAS-IDTR-2019-008

Summary

Tracking via approximate-nearest-neighbor methods: hashing and similarity learning

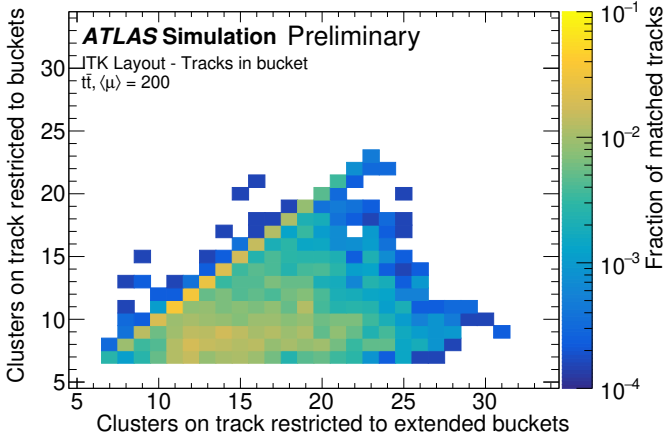
General, detector-independent techniques

Promising results on TrackML and ATLAS ITK upgrade simulation dataset

Ongoing: work towards full reconstruction chain


CERN

Reconstruction in extended buckets



As before, but buckets extended to contain all truth hits for contained tracks to emulate optimized track road.

Diagonal: fully matched tracks

Below: bucket track is missing hit(s)

Above: bucket track contains additional hit(s)

ATLAS-IDTR-2019-008