

PR #238:

Merge T5s After Building

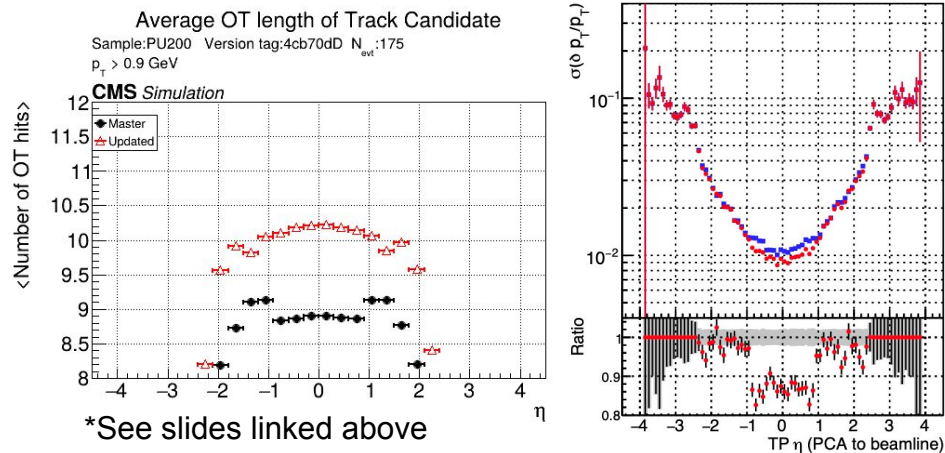
Gavin Niendorf

Overview

- Moved the **T5-T5 merging kernel** from the end of LST to immediately after the T5s are built.
 - Allows downstream duplicate cleaning and object building kernels to make use of longer **track length**.
- Revisited the T5 duplicate cleaning kernels after moving the merging kernel to improve **displaced efficiency**.

Motivation

- Current **T5-T5 merging kernel** (see previous slides [here](#)) increases TC length and improves momentum resolution significantly, but merging happens **after TCs are fully built**.
 - Duplicate cleaning kernels and object building kernels (e.g. pT5 building) can't make use of **longer track length** to improve performance.
- Moving the merging kernel from the end of LST to immediately after T5s are built allows for potential improvements in duplicate cleaning and object building.



PR Changes

- Moved T5-T5 merging kernel from the end of LST to after T5s are built, and **optimized the kernel** significantly so that the timing difference is negligible.
 - Previous merging kernel only merged built TCs, so the number of T5-T5 pairs it considered was significantly smaller.
- Made improvements to the **three T5 duplicate cleaning kernels** after moving the merging kernel - *RemoveDupQuintupletsAfterBuild*, *RemoveDupQuintupletsBeforeTC*, and *CrossCleanT5*.
 - **Longer tracklets** are now kept over shorter ones, and longer tracklet pairs have **tighter cuts** to be considered duplicates of each other.

Duplicate Cleaning Changes - After Build

RemoveDupQuintupletsAfterBuild (Kernels.h):

- Shared-hit threshold changed from flat ≥ 7 hits to **proportional threshold** based on tracklet length ranging from 6-8 hits (see backup).
- **The longer tracklet** is kept when a pair gets flagged as duplicates. For pairs of the same length, the tracklet with the **better T5 DNN score** is kept for low pT tracks ($pT < 5$ GeV) and the tracklet with the lower chi-squared score is kept for high pT tracks ($pT > 5$ GeV).
 - The pT-dependent tiebreaker was added to maintain efficiency at high pT. The T5 DNN needs to be retrained in the future with high pT tracks upweighted so that this can be simplified.

Duplicate Cleaning Changes - Before TC / Cross Clean

RemoveDupQuintupletsBeforeTC (Kernels.h):

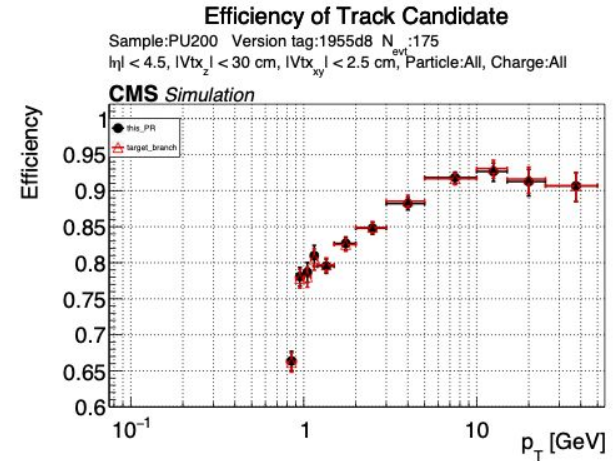
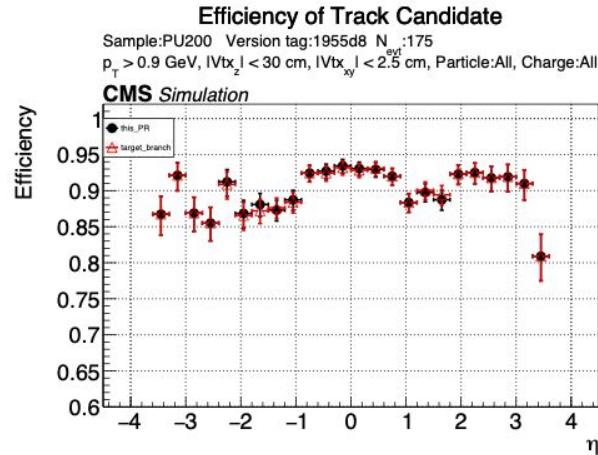
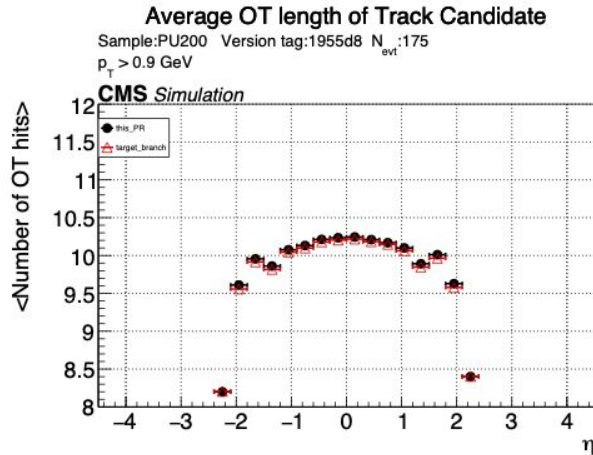
- Same pT-dependent tiebreaker added as in the after-build kernel.
- Embedding-distance threshold tightened from 1.0 to 0.25, chosen at the 99th percentile of true-duplicate pairs for 100 PU200 events to avoid displaced efficiency loss.

CrossCleanT5 (TrackCandidate.h):

- T5s that got merged at least once get tighter embedding cuts (thresholds (0.1,1.0) \rightarrow (0.03,0.3)) to avoid displaced efficiency loss from cross cleaning.
- A fallback based on **shared-hit count** was added to catch true duplicates missed from the above tightened selection.

Physics Performance - Standalone

- Minimal improvements in TC track length and efficiency at low p_T from the duplicate cleaning kernel changes (taken from the CI, this PR is black).

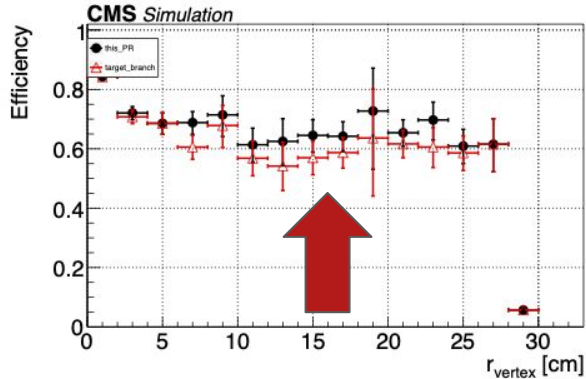


Physics Performance - Standalone Cont.

- **Significant increase in displaced efficiency** for PU200, slight decrease in fake rate, minimal differences in duplicate rate in the barrel.

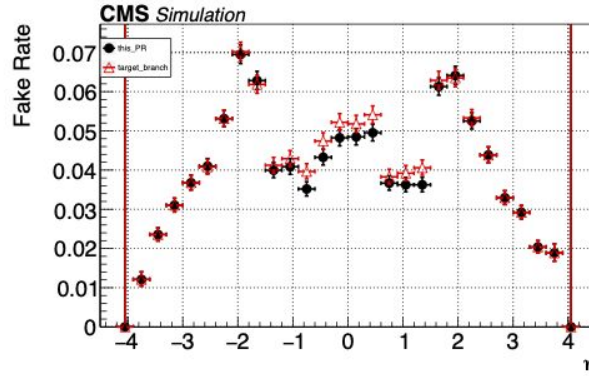
Efficiency of Track Candidate

Sample:PU200 Version tag:1955d8 N_{evt}:175
 $|\eta| < 4.5$, $p_T > 0.9$ GeV, $|Vtx_z| < 30$ cm, Particle:All, Charge:All



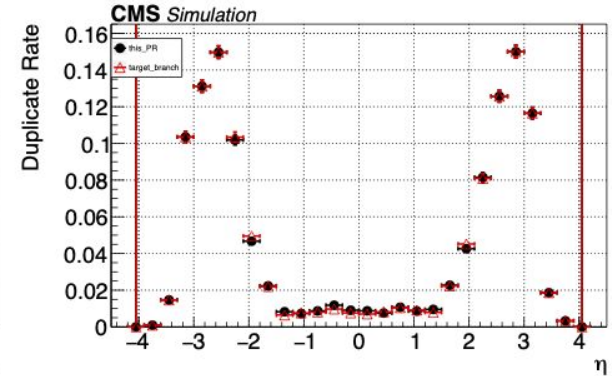
Fake Rate of Track Candidate

Sample:PU200 Version tag:1955d8 N_{evt}:175
 $p_T > 0.9$ GeV



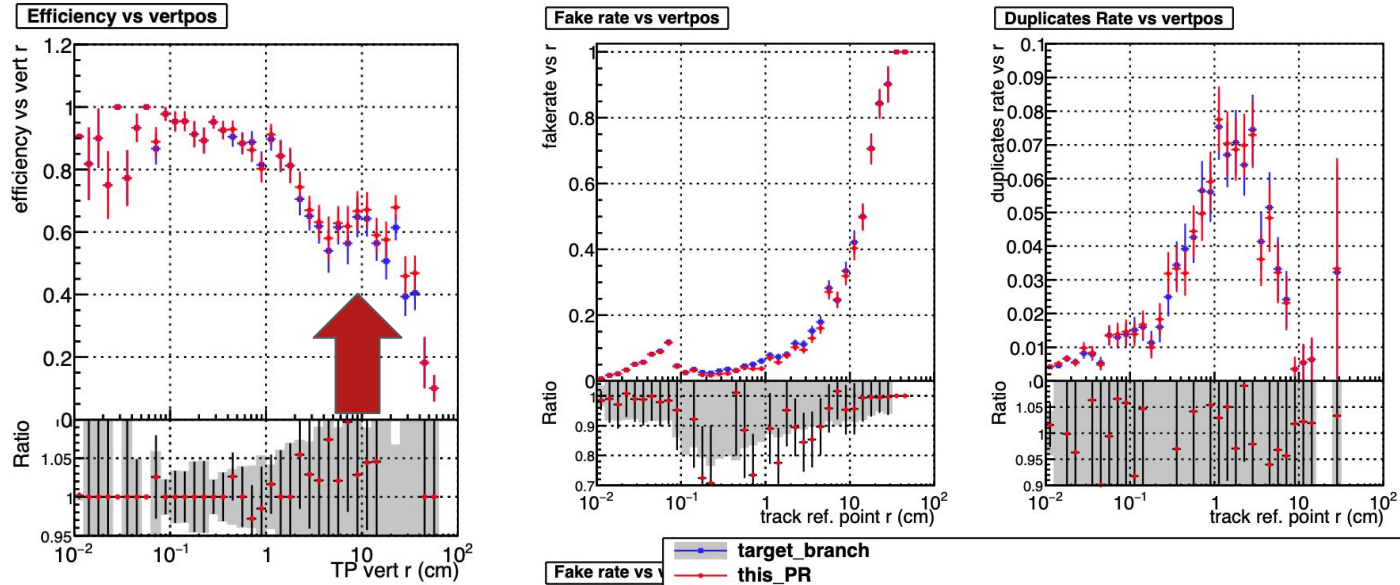
Duplicate Rate of Track Candidate

Sample:PU200 Version tag:1955d8 N_{evt}:175
 $p_T > 0.9$ GeV



Physics Performance - CMSSW (HLT)

- **Significant increase in displaced efficiency** for PU200, slight decrease in fake rate, minimal differences in duplicate rate (red is this PR, blue master).



Timing Comparison - CPU (Negligible Changes)

Master

```
Total Timing Summary
Average time for map loading = 369.094 ms
Average time for input loading = 6352.85 ms
Average time for lst::Event creation = 0.00156675 ms
```

Evt	Hits	MD	LS	T3	T5	pLS	T4	pT5	pT3	TC	Reset	Event	Short	Rate	
avg	14.7	90.0	73.8	64.2	24.1	331.5	5.7	23.1	44.5	118.8	0.1	790.6	444.3+/- 106.5	790.6	explicit[s=1]
avg	14.7	90.0	74.1	64.6	24.4	332.0	5.7	23.1	44.5	119.0	0.1	792.4	445.6+/- 106.9	202.1	explicit[s=4]
avg	15.1	90.2	74.6	65.2	24.7	331.5	5.8	23.2	44.6	119.1	0.2	794.3	447.7+/- 107.5	57.5	explicit[s=16]
avg	16.0	90.3	75.0	65.9	25.1	332.3	5.9	23.4	44.8	119.4	0.3	798.3	450.0+/- 107.7	33.2	explicit[s=32]
avg	19.7	97.6	81.8	72.5	28.1	359.4	6.4	25.1	47.6	126.6	0.5	865.4	486.2+/- 117.1	20.3	explicit[s=64]

This PR

```
Total Timing Summary
Average time for map loading = 371.179 ms
Average time for input loading = 6358.75 ms
Average time for lst::Event creation = 0.00170299 ms
```

Evt	Hits	MD	LS	T3	T5	pLS	T4	pT5	pT3	TC	Reset	Event	Short	Rate	
avg	14.3	90.0	73.5	63.2	31.6	316.3	5.5	26.5	43.9	108.9	0.0	773.8	443.1+/- 106.0	773.9	explicit[s=1]
avg	14.4	90.1	74.6	64.0	32.2	317.2	5.6	26.3	43.8	108.5	0.1	776.8	445.2+/- 106.8	197.9	explicit[s=4]
avg	14.8	90.2	74.2	64.7	32.4	316.8	5.7	26.4	43.9	108.6	0.3	777.9	446.4+/- 107.0	56.2	explicit[s=16]
avg	15.4	90.3	74.5	64.9	33.0	316.9	5.7	26.6	44.0	108.6	0.3	780.3	448.0+/- 107.1	32.5	explicit[s=32]
avg	19.3	97.3	81.8	71.7	36.8	340.5	6.3	28.4	46.8	114.8	0.5	844.2	484.3+/- 117.0	19.8	explicit[s=64]

Timing Comparison - GPU (Negligible Changes)

Master

```
Total Timing Summary
Average time for map loading = 465.488 ms
Average time for input loading = 7109.9 ms
Average time for lst::Event creation = 0.000613076 ms
```

Evt	Hits	MD	LS	T3	T5	pLS	T4	pT5	pT3	TC	Reset	Event	Short	Rate	
avg	0.6	0.2	0.2	0.5	0.7	0.2	0.4	0.3	0.3	1.0	0.0	4.4	3.6+/- 2.5	4.5	explicit[s=1]
avg	0.7	0.4	0.3	0.6	0.8	0.3	0.6	0.5	0.4	1.3	0.0	5.7	4.8+/- 2.7	3.0	explicit[s=2]
avg	1.0	0.5	0.5	0.9	1.0	0.3	0.8	0.6	0.5	1.9	0.0	8.1	6.8+/- 3.0	2.1	explicit[s=4]
avg	1.3	0.8	0.8	1.2	1.3	0.4	1.2	0.8	0.7	2.5	0.0	11.0	9.3+/- 3.8	2.0	explicit[s=6]
avg	1.6	1.2	1.0	1.6	1.7	0.5	1.5	1.0	0.9	3.2	0.0	14.3	12.2+/- 3.9	1.9	explicit[s=8]

This PR

```
Total Timing Summary
Average time for map loading = 483.814 ms
Average time for input loading = 7130.15 ms
Average time for lst::Event creation = 0.000703903 ms
```

Evt	Hits	MD	LS	T3	T5	pLS	T4	pT5	pT3	TC	Reset	Event	Short	Rate	
avg	0.6	0.2	0.2	0.5	0.8	0.2	0.4	0.4	0.2	0.9	0.0	4.5	3.7+/- 2.6	4.6	explicit[s=1]
avg	0.7	0.3	0.3	0.6	0.9	0.3	0.6	0.6	0.4	1.1	0.0	5.8	4.9+/- 2.8	3.0	explicit[s=2]
avg	1.0	0.6	0.5	0.8	1.3	0.3	0.8	0.8	0.6	1.6	0.0	8.3	7.0+/- 3.1	2.2	explicit[s=4]
avg	1.3	0.8	0.7	1.2	1.6	0.4	1.1	0.9	0.7	2.3	0.0	11.1	9.5+/- 3.7	2.0	explicit[s=6]
avg	1.6	1.2	1.1	1.5	2.0	0.5	1.5	1.3	0.9	3.0	0.0	14.7	12.7+/- 4.3	1.9	explicit[s=8]

Possible Future Improvements

- Retrain T5 DNN with high-pT tracks upweighted so that same-length tiebreaker can be simplified to keeping the tracklet with better DNN score.
- Use merged hits in pT5 building kernel to potentially improve physics performance (currently only the base T5 hits are considered).
- Revisit the T5 DNN architecture (started in PR [#200](#)) to add an explicit displaced output score like the T3 DNN has to recover some more displaced efficiency.

Summary

- Moved the **T5-T5 merging kernel** from the end of LST to immediately after T5s are built and revisited the T5 duplicate cleaning kernels to make use of increased track length.
- Improvements to duplicate cleaning kernel after moving the merging kernel result in a significant improvement in **displaced efficiency** for PU200 with negligible differences in timing.

Backup Slides

Duplicate Merging Threshold

- Shared-hit threshold is computed as $\text{int}(0.6 \times \min(2 \cdot n\text{Layers}_i, 2 \cdot n\text{Layers}_j))$, i.e. 60 % of the shorter tracklet's hit count. This gives the 6/7/8-hit ladder below.

pair (nHits × nHits)	shared-hit threshold (≥)
10 × 10	6
10 × 12	6
10 × 14	6
12 × 12	7
12 × 14	7
14 × 14	8