Duplicate studies T5

TRES REID, PETER WITTICH [CORNELL]
BEI WANG [PRINCETON]
PHILIP CHANG, BALAJI VENKAT SATHIA NARAYANAN, SLAVA KRUTELYOV, AVI YAGIL [UCSD]
T5

Made $d(1/pt,\eta,\phi)$ plots for all T5-T5 pairs.
  - Initial (old) selections: 0.1, 0.075, 0.075 -> should be loosened

Duplicates are T5s that are matched to the same sim track

Fakes are T5s that are not sim matched
  - A fake will never be included in the duplicate collection by this definition, but maybe have a high number of the same hits

Hit Matching suggests that we can remove all duplicates if fake tracks can be correctly identified
  - Implemented a track score
    - Line defined from first 2 hits
    - Line defined from first and last hit
    - Line defined by least square of all hits (works the best)
    - $Score = layer \times \sum(abs(residuals))$ [since there are no variable errors on the hits]
D(1/pt) (needs to be increased)

NON-DUPLICATE PAIRS

DUPLICATE PAIRS

6/22/2021

SEGMENT LINKING
dEta
dPhi (no zeros)
hit matching between all T5

NON-DUPLICATE HIT MATCHES (NO 0) CAN INCLUDE FAKES

NON-DUPLICATE HIT MATCHES (NO 0) WHERE BOTH T5S ARE REAL
hit matching between all T5

DUPLICATE HIT MATCHES (BOTH T5S MUST BE REAL)

NON-DUPLICATES WITH ONLY 1 REAL AND 1 FAKE
T5 matching grouped by original track id excluding fakes of second track collection

NON- DUPLICATE (ALL)

NON-DUPLICATE (NO FAKEs)
T5 matching grouped by original track id excluding fakes of second track collection

**DUPLICATEs**

We should be able to get rid of nearly all duplicates requiring \( \geq 6 \) hits matched if we can properly discriminate against fake T5s.
Track Score Difference

REAL – FAKE T5 SCORE
NMATCHED >=6

KEEP LOWER SCORE

4238(27.6%): positive difference
11126(72.4%): negative difference
- Similar fractions for >=5,6 hits
- None with exact same score

Linear fit using least squares formula for all 10 hits. To get slope and intercept

Score = modulesInGPU.layers[lowerModule1] * sum(abs(slope*(z)+b) – r)
- Score for tracks starting in second layer are doubled->prefer innermost T5
No Duplicate removal

Efficiency of Track Candidate of all types
Sample: PU200  Version tag:56b4fa9D/RTY_GPU_unified_newgrid
|η| < 2.4, |VTX| < 30 cm, |Vtx_x| < 2.5 cm

Duplicate Rate of Track Candidate of all types
Sample: PU200  Version tag:56b4fa9D/RTY_GPU_unified_newgrid
|η| < 2.4
No Duplicate removal

Efficiency of Quintuplet of all types
Sample: PU200  Version tag:58b4fa9D/RTY_GPU_unified_newgrid  
$|\eta| < 2.4$, $|v_{x,y}| < 30$ cm, $|v_{x,y}| < 2.5$ cm

Duplicate Rate of Quintuplet of all types
Sample: PU200  Version tag:58b4fa9D/RTY_GPU_unified_newgrid  
$|\eta| < 2.4$
With duplicate removal (nMatched>=7)

Score using least square method*layer, keeping lower score
nMatched >=7; d(pt,eta,phi) requirement as above
T5 objects: 9339-> 3230

Efficiency of Track Candidate of all types
Sample: PU200 Version tag:93bca6cDIRTY_GPU_unified_newgrid
|h| < 2.4, |Vb_x| < 30 cm, |Vb_y| < 2.5 cm

Duplicate Rate of Track Candidate of all types
Sample: PU200 Version tag:93bca6cDIRTY_GPU_unified_newgrid
|h| < 2.4
With duplicate removal (nMatched>=7)

Efficiency of Quintuplet of all types
Sample: PU200  Version tag:93bca6cDIRTY_GPU_unified_newgrid
|η| < 2.4, ||< 30 cm, |< 2.5 cm

Duplicate Rate of Quintuplet of all types
Sample: PU200  Version tag:93bca6cDIRTY_GPU_unified_newgrid
|η| < 2.4
With duplicate removal (nMatched>=6)

Score using least square method*layer, keeping lower score
nMatched >=6; d(pt,eta,phi) requirement as above
T5 objects: 9339->2902

Efficiency of Track Candidate of all types
Sample: PU200  Version:93bc96cDIRTY_GPU_unified_newgrid
|h| < 2.4, |Vxy| < 30 cm, |Vxb| < 2.5 cm

Duplicate Rate of Track Candidate of all types
Sample: PU200  Version:93bc96cDIRTY_GPU_unified_newgrid
|h| < 2.4
With duplicate removal (nMatched>=6)
With duplicate removal (nMatched>=7)

No pt, eta, phi cuts
T5 objects: 9339->2521

Efficiency of Track Candidate of all types
Sample: PU200  Version tag:93bca6cDIRTY_GPU_unified_newgrid
$|\eta| < 2.4$, $|V_{xy}| < 30$ cm, $|V_{xz}| < 2.5$ cm

Duplicate Rate of Track Candidate of all types
Sample: PU200  Version tag:93bca6cDIRTY_GPU_unified_newgrid
$|\eta| < 2.4$
With duplicate removal (nMatched>=7)
With duplicate removal (nMatched&ge;=6)

No pt, eta, phi cuts
T5 objects: 9339-> 2099

Efficiency of Track Candidate of all types
Sample: PU200 Version tag: 93bca6cDIRTY_GPU_unified_newgrid
|η| < 2.4, |Vxy| < 30 cm, |Vxy₇| < 2.5 cm

Duplicate Rate of Track Candidate of all types
Sample: PU200 Version tag: 93bca6cDIRTY_GPU_unified_newgrid
|η| < 2.4
With duplicate removal ($n_{\text{Matched}} \geq 6$)
Summary

Hit matching should be sufficient for duplicate removal if we can perfectly discriminate against fake tracks

- $\geq 7$ tends to be a bit better than $\geq 6$ likely because of hit to the efficiency in the 6 case
- $D(pt, \eta, \phi)$ cuts were estimated to be $0.1, 0.075,$ and $0.075$ but will possibly need to be revisited since there are better results without any cut applied
- Track score = layer * $\sum(\text{abs(residuals)})$ where the linear fit is made using a least squares method from all hits. Keep track with lower score
  - Correctly keeps the real track $\sim 75\%$ of the time
dR

dR of T5 pairs with nMatched <=6
At least one T5 is real
Large number of non-duplicate tracks with dR=0
This dR calculation is just asking if the hit in the second layer is close (or the same). Might be better to take a value from the linear fit instead? Compare slope and intercept (maybe with additional phi requirement)?