Tracking Algorithm Recommendation Committee Report

H. T. Diehl, A. Kharchilava, R. van Kooten, V. Kouznetsov, J. Qian (Chair)
October 2, 2002

I: Charges:

In July 2002, the DØ physics and computing/software management commissioned a Tracking Algorithm Recommendation Committee (TARC) with the following charges:

1) Collect information and get data on the performance of the different tracking algorithms. These measures of performance should include efficiencies, fake rates and mis-reco rates, using the standard procedures developed by the global tracking group. In addition, measurements of efficiencies and fake rates for particle id obtained from standard datasets (both data and Monte Carlo) accumulated by physics/id groups should be made. Finally, average CPU time per event, memory usage, and luminosity dependence of the CPU time should be determined. Input from physics/id/algorithms groups should be solicited.

2) Make recommendation(s) on how we should run tracking in p13 on the farm, taking into account the maximum CPU time budget. This budget is set by the number of available farm nodes, the average farm efficiency and the average DAQ rate. Based on what will be available on the farms in October, and an official average DAQ rate goal of 25 Hz, the maximum allowed CPU time per event on a 500 MHz machine is 58 seconds. Assigning 50% of this budget to finding tracks results in a maximum allowed average time per event for tracking algorithms of 29 seconds (normalized to 500 MHz units).

The committee was asked to submit its recommendation(s) by September 15, 2002, which was subsequently postponed to Sept. 20 due to scheduling problems. This aggressive schedule is largely set by the DØ p13 software release schedule. This report briefly summarizes the work done by the committee, the global tracking group, physics/ID groups and others over the last two months.

II: Details:

DØ global tracking group are currently developing four different tracking algorithms:

1) Global TRack finder (GTR): based on TRF++, uses specific paths (roads) during track finding.
   Author: GTR group

2) Histogramming Track Finder (HTF): divides the detector into slices in (φ, ρ) and uses Hough transform to reduce initial number of combinations. Existing algorithm can use either CFT, SMT hits or combinations of them to construct tracks.
   Author: S. Khanov

3) The ELAstic algorithm (ELA) finds additional tracks by using the clusters left unassociated with tracks reconstructed by other algorithms. Seeds are formed by filtering a 5 dimensional histogram

focused on primary vertex regions of interest. The seeds are then fit to clusters and the cluster associations are updated simultaneously.
Author: A. Hass

4) Alternative Algorithm (AA): starts from any combination of 3 hits in different SMT super-layers. Track candidates are extended towards CFT. It uses 1D SMT clusters while all others use 2D clusters.
Author: G. Borissov

More details about these algorithms can be found from the introduction talk given at an All DØ Meeting (ADM) on August 9, 2002 by Kuznetsov. As stated by the above charges, the committee was asked to evaluate the performances of these algorithms and their combinations. The committee met a number of times in the two months period. Some history of the committee's work can be found from the TARC web page. Upon the recommendations of the tracking group, the committee decided to evaluate the following algorithms and their combinations:

1) GTR alone with a pT cut of 0.4 GeV
2) HTF alone with a pT cut of 0.5 GeV
3) GTR+ELA: ELA runs on unused hits
4) HTF+ELA: ELA runs on unused hits
5) GTR+HTF: pT>0.4 GeV for GTR and pT>0.5 GeV for HTF
6) AA alone with a pT cut of 0.2 GeV

Other combinations* are not considered due to technical (for example, 1D clusters used by the AA and 2D clusters used by others) and/or schedule difficulties. In GTR combinations with others, the GTR algorithm was turned off in the overlap region, the region with partial CFT coverage. The committee then requested physics/ID groups to provide their preferred data and Monte Carlo samples on which the algorithm performance could be evaluated. The following samples were selected for processing with different algorithms:

Data:

Sample	Events	SAM output locators
J/ψ	30k	%dimuon_third_merged%tk-p11.11-%.root
μ+jets	10k	%merge_mujet%tk-p11.11-%.root
Z→μμ	3k	%pick_dimuon%tk-p11.11-%.root
Z→ee	7k	%pick_diem%tk-p11.11-%.root
Run 155554†	10k	%reco_all_0000155554%tk-p11.11-%.root
Run 157708	100k	%reco_all_0000157708%tk-p11.11-%.root

Monte Carlo:

Sample	Events	SAM input/output locators
Ttbar	10k	%ttbar-wjj+wlnu%tk-p11.11-%.root
Higgs	10k	%bbh-bbbb%tk-p11.11-%.root
Z→ee	5k	%z-ee%tk-p11.11-%.root
τ	10k	%tau_tauhcw%tk-p11.11-%.root
b stuff	25k	%bbbarQQ%tk-p11.11-%.root
Light quark	~10k	%ddh-dddd%tk-p11.11-%.root

* A combination of all tracking algorithms was also studied for cross checks.
† Run 155554 is for test purpose only.

Here "%" is the SAM wild card. The last "%" stands for one of the algorithms (GTR, HTF, GTRELA, HTFELA, AA, TRKALL). For Monte Carlo events, the version of simulation is generally different from different samples.

Special p11.11 *dØreco* executables were made with all algorithms. Algorithms were selected by RCP parameters. The data were processed at the Fermilab central farm while the Monte Carlo events were processed at the UTA farm. The committee expresses its appreciation to the offline resource board, the central farmers (Heidi and Mike) as well as the UTA crew (Mark Sosebee, Jae Yu and others) for their assistance in timely processing of these events. About 3% of all jobs crashed in *dØreco* and ~15% in *dØanalyze*. The causes of these crashes, though unclear, are likely to be outside of the tracking algorithms under the study.

Physics/ID groups were asked to compare performances of different algorithms using *gtr_analyze* and *gtr_examine* packages provided by the tracking group and to compare efficiency and fake rate for final ID objects as well as try to determine overall impact on physics analyses. The study was culminated in a mini-workshop held on Sept. 18 with the following agenda:

```
                 MiniWorkshop on Tracking Algorithm Studies
                   1:30-5:30pm, Sept. 18, 9th Circle

1) 1:30-2:00  Sasha Khanov (tracking)
   Tracking status and performance on MC and data

2) 2:00-2:25  Bram Wijngaarden (b-id)
   lifetime B tagging

3) 2:25-2:50  Silke Duensing (tau-id)
   Tracking for taus

4) 2:50-3:15  Vivek Jain (B physics)
   Comparison of tracking algorithms

5) 3:15-3:40  Robert Zitoun (EM id)
   Electron track matching

   3:40-4:00   Break

6) 4:00-4:25  Ryan Hooper (mu id/np)
   Checking different tracking algorithms with a dimuon sample

7) 4:25-4:55  Lorenzo Feligioni (Higgs)
   Higgs group report

8) 4:55-5:15  Elizaveta Chabalina (top)
   Tracking in top samples

9) 5:15-5:35  Ariel Schwartzman (top)
   Secondary vertex b-tagging in top MC samples
```

It should be noted that many others had shown a significant amount of work at individual tracking, physics and ID working group meetings before the mini-workshop. Interested readers are encouraged to get details there (from working group web pages for example).

III: Observations & Recommendations:

The workshop was followed by two presentations at the ADM meeting two days later. Representing the committee, Diehl summarized the results of the workshop and Qian presented the committee's recommendations. The followings are some of the results shown at the workshop:

- The CPU speed of different algorithms varies from ~ 4 GHz-sec/event for the AA to ~ 12 GHz-sec/event for the GTR+HTF, the corresponding *dØreco* speed is between 9-20 GHz-sec/event. The memory consumption is between 580-750MB with the HTF using the least amount and the AA using the most amount. All these are measured using Run 157708.
- The efficiencies for isolated high pT tracks in data are between 45-65% with GTR+ELA and HTF+ELA combinations at the high end and single algorithms at the low end. In a sample of 160 $Z \rightarrow ee$ events, the overlap of reconstructed candidates by AA, GTR, and HTF is only 20%.
- The tracking efficiency in jets for MC events is between 60-80% in the central region with a 5-10% fake rate. The winner is GTR+HTF and HTF+ELA in efficiency and is AA in fake rate.
- The efficiency for b-tagging with secondary vertices increases from about 40% at pT=10 GeV to about 60% at pT=50 GeV for GTR+HTF and HTF. The corresponding fake rate varies from 2% to about 5%. Other algorithms have slightly lower efficiencies with comparable fake rates. No b-tagging studies were done for the AA algorithm for technical reasons.
- For low mass resonance reconstructions, the AA algorithm appears to be superior compared with all others, possibly due to the low pT cut applied. For MC taus, the HTF+ELA and GTR+HTF are among the best.
- In general, the AA algorithm has low efficiencies for both data and Monte Carlo, but it also has the smallest fake rates for most cases.

In brief, the committee

- is impressed by the results of all the alternative algorithms presented at the workshop. However, current performance on data of all the algorithms still show deficiencies, particularly in tracking efficiency for isolated high-pT tracks.
- believes that all algorithms have potential for significant improvements and that it is obviously in our best interest to further develop them and continue the study as a part of an ongoing process.
- notes that there is no single best algorithm (or combination of algorithms) that meets all our physics needs at this time,
- despite this, it is clear that in almost all cases, the other alternative algorithms perform better than the current algorithm, GTR alone.

Therefore, the committee recommends to

- ***implement GTR+HTF combination as the default algorithm for p13.***
- ***encourage all developers to continue the code development for p14.***
- ***carry out a similar study for p14, but focus more on data.***

In addition, the committee supports the tau group request to include axial-only tracks in *dØreco* outputs and has learnt that it is being implemented for p13. We also note that the tracking group is studying to lower the HTF pT threshold used in GTR+HTF to see whether it can efficiently reconstruct low pT tracks without blowing our CPU budget for p13.

The recommendation for GTR+HTF is based on the fact that it

- is one of the better performers for most physics,
- offers improvement while providing continuity,
- runs at ~ 12 GHz-sec/event (~ 20 GHz-sec/event for *dØreco),* well within our CPU budget, at least for now.