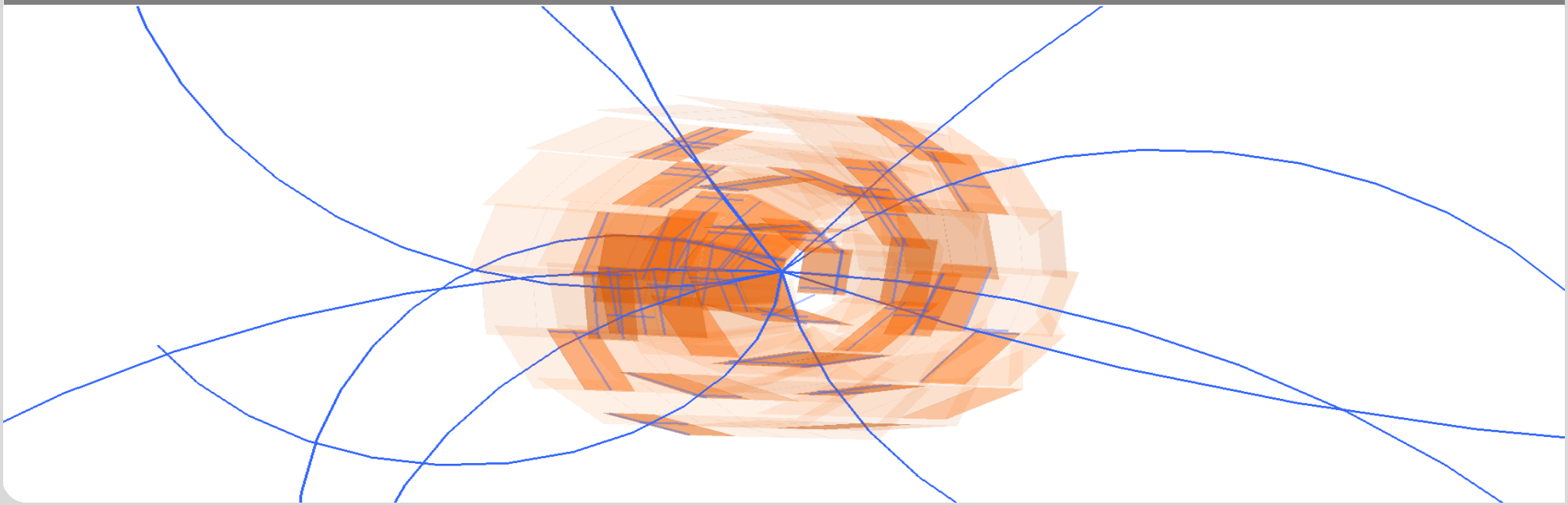
MVA Quality Estimator for Full Track

F2F Tracking Meeting (Part I VXD) @ KEK, Tsukuba, Japan

Sebastian Racs | 3rd February 2018

INSTITUT FÜR EXPERIMENTELLE TEILCHENPHYSIK (ETP)



Introduction

Moving from SVD to a Full Track

- Previously shown MVA QualityEstimation for SVD using VXDTF2's SpacePointTrackCandidates @ Pisa F2F
- Using this output for SVDRecoTrack
- Has to be used in some way for full tracking

$\Rightarrow$ Doing a new MVA QualityEstimation for the full RecoTrack using sub-RecoTrack QualityIndicators and other features as input

List of variables/features

- Categories of variables that get collected for each RecoTrack:
  1. Whole Event in general and comparisons between current and other RecoTracks
  2. Current RecoTrack's information including the fit
  3. Sub-RecoTracks information including their sub-QIs and fit
  4. Combining information form the RecoTracks hits

- color-marked features were (currently) rated important in FastBDT evaluation
- gray-marked features were rated completely unimportant
- Values that don't exist for different reasons get filled with:
  - -1, if the feature should always be positive
  - $-$float::max, if it can also be negative

Information from the whole Event

- # of RecoTracks and sub-RecoTracks (PXD, SVD, CDC)
- Δ total number # of
  - PXD vs. SVD RecoTracks
  - SVD vs. CDC RecoTracks
- current RecoTrack vs. any other RecoTrack in the event: save minimal Δ (and corresponding Index)
  - $\vec{p}_{\text{seed}}$ components *RT.getMomentumSeed()*
    - magnitude *.Mag()*
    - **transversal** *.Pt()*
  - $\vec{r}_{\text{seed}}$ components *RT.getPositionSeed()*
    - azimuthal angle θ *.Theta()*
    - **polar angle** φ *.Phi()*

Possible Problems

- Bias of total events, might cause problem for certain analysis channels

Information from the RecoTrack

- # of total Hits and PXD, SVD, CDC Hits
- get **position** and **momentum** components at perigee (do extrapolation to beamline):
  - longitudinal/z *.Z()*
  - transversal *.Pt()*
  - **magni**tude *.Mag()*
  - **azimuthal angle** θ *.Theta()*
  - polar angle φ *.Phi()*
- **position seed** and **momentum seed** components (see above)
- **charge seed** and **time seed**
- genfit **TrackFitStatus** values
  - Fit successful?
  - χ^2
  - ***p*-Value**
  - **# degrees of freedom (NDF)**
  - charge
  - # failed points

Information from sub-RecoTracks

- QualityIndicator from **SVD**, ~~CDC~~, ~~PXD~~, ~~merger~~
- SVD, CDC fit successful?
- SVDRecoTrack has a SPTackCand? (true, if created by vxdtf2)
- get **position** and **momentum** component differences and χ^2 between CDC and SVD extrapolated to the
  - **CDCwall** (like `VXDCDCTrackMerger`)
  - **perigee**

    components:

    - longitudinal/z *.Z()*
    - transversal *.Pt()*
    - magnitude *.Mag()*
    - azimuthal angle θ *.Theta()*
    - polar angle φ *.Phi()*
    - pseudorapidity η *.Eta()*

Combined information from RecoHits

get KalmanFitterInfo from trackPoints related to RecoTrack

- count # of missing TrackPoints and KalmanFitterInfos
- get **weight** and **smoothedChi2** and save per RecoTrack:
  - maximum
  - minimum
  - **mean**
  - **median**
  - value @ last SVD hit
  - value @ first CDC hit

Missing features

Being developed or finalized now

- CDC QualityIndicator ($\rightarrow$ Michael Eliachevitch)
- Information from the CKF-merger(s) ($\rightarrow$ Nils Braun)

Could still be added

- **PXD** and SVD hit pattern (which layer was hit how often)
- other/better vector components

Any suggestions?

- ...

As seen before: SVDonly QualityIndicator Cuts

- left: same results as in PR-1330, but re-done with newer samples
  - ⇒ already lower fake rate than in Pisa
- right: same weight-file and sample, but on current master
  - ⇒ Fake rate reduced even further by changes in the mean-time
  - ⇒ MVA QI is still better than previous one

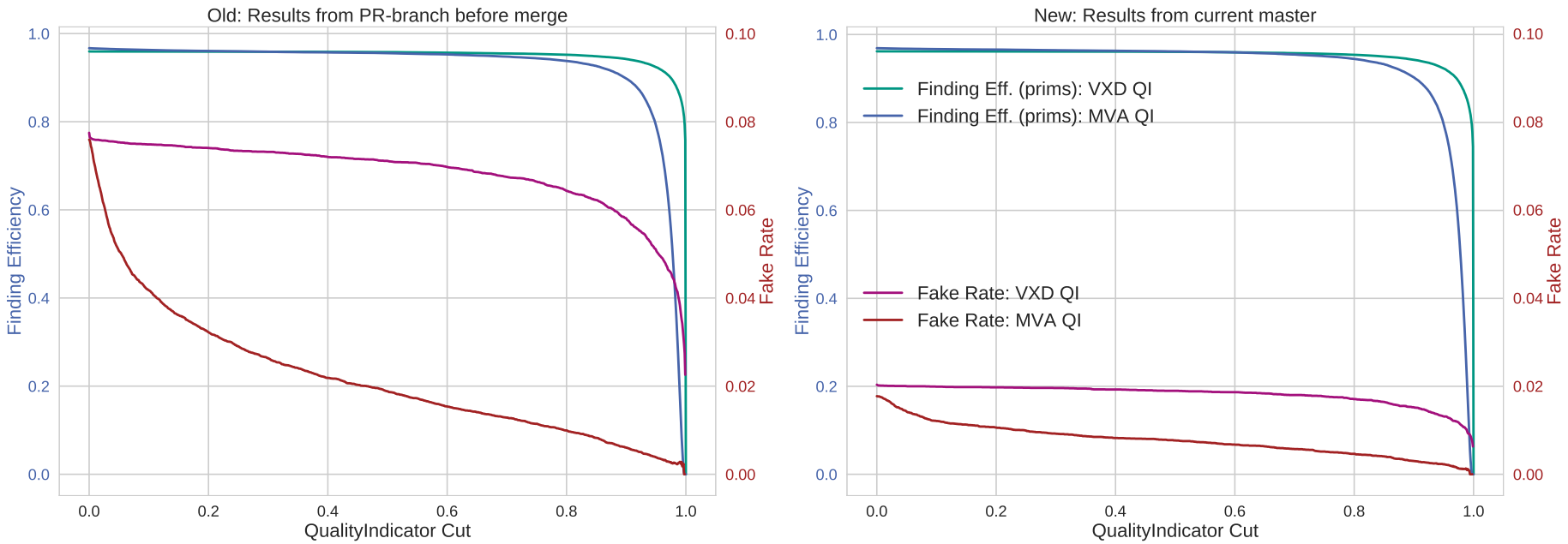


Figure: Finding Efficiency and Fake Rate depending on QI Cuts

SVDonly Fake Rate vs. Finding Efficiency

- Without cuts, the VXD QI from the master is even better than the old MVA results
- In general MVA is still a lot better

$\Rightarrow$ gives possibility to cut fakes without losing many matched SVD tracks

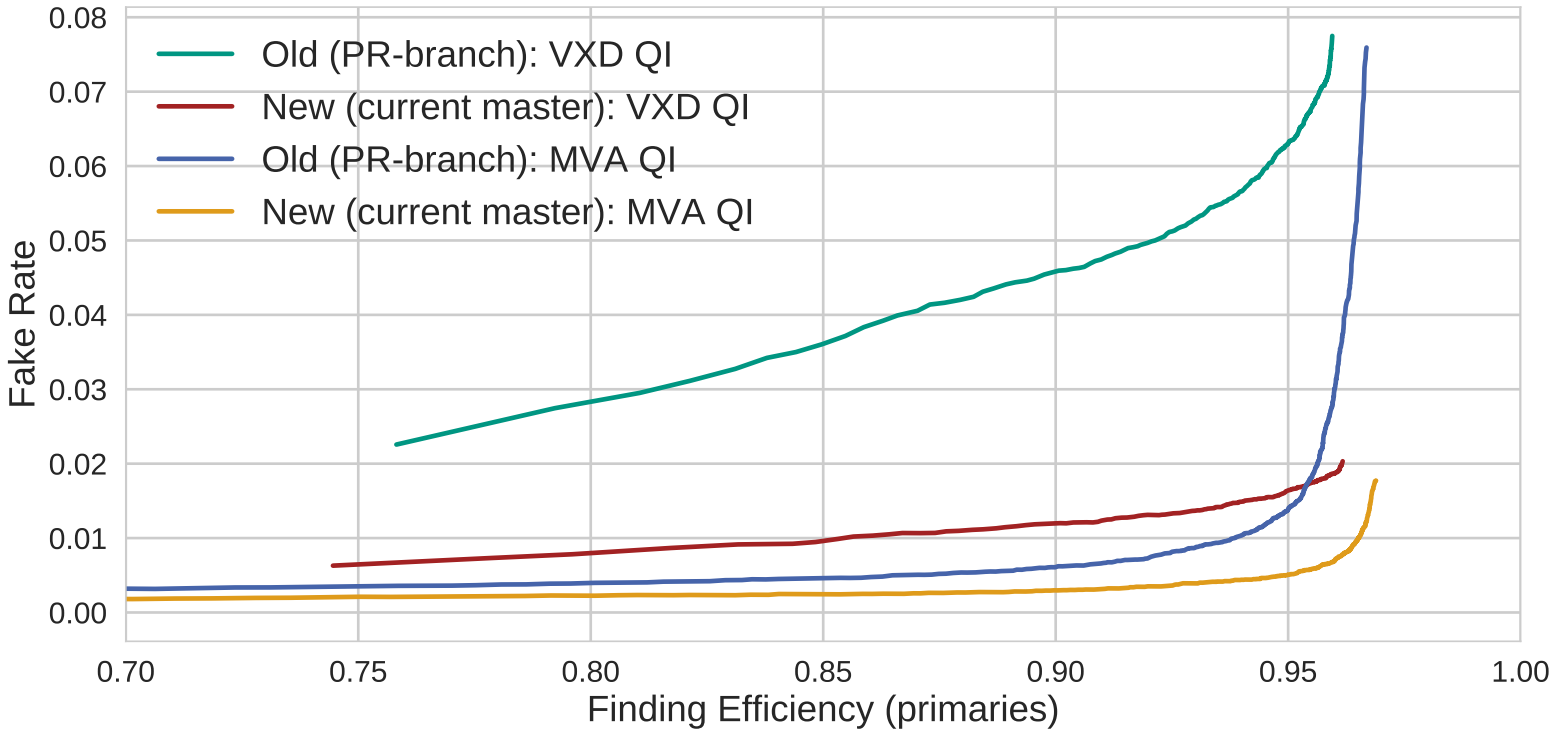


Figure: Fake Rate vs. Finding Efficiency (for same QICuts)

Current Results for Full Track MVA

- Training: 100k $\Upsilon(4S)$ events, Validation: 5k $\Upsilon(4S)$ events,
- Compare just using the SVD QI for any track that contains SVD hits vs. the Full RecoTrack MVA QI with the features from above

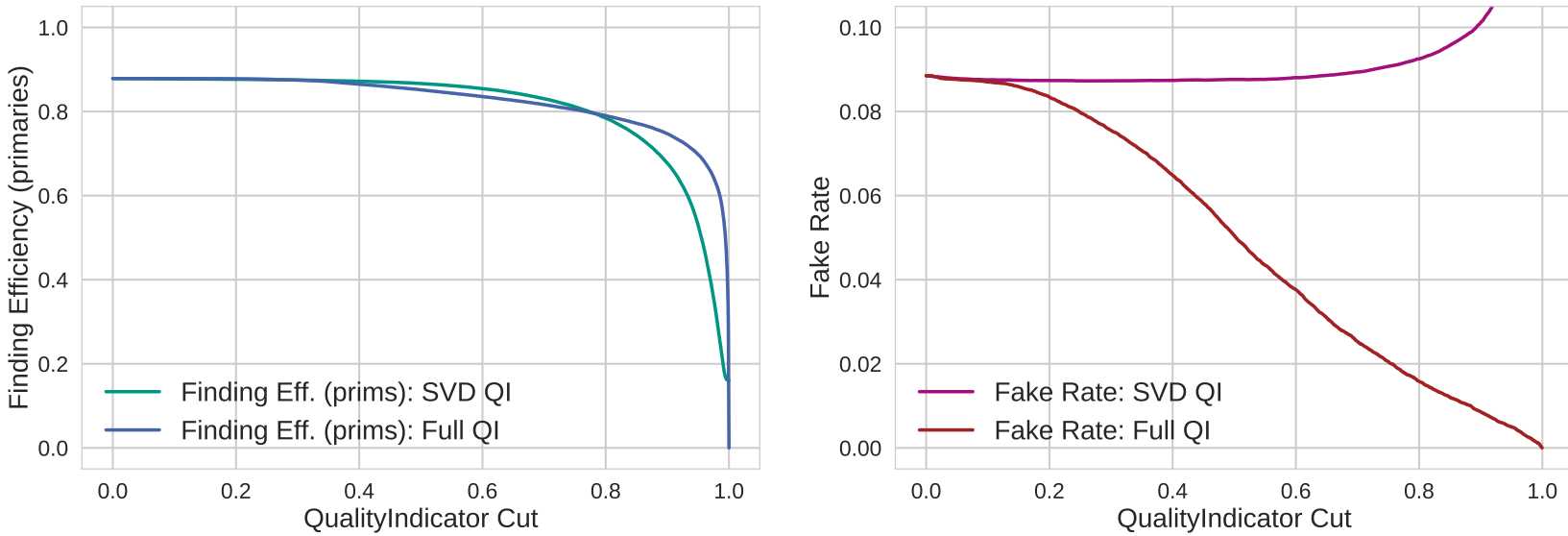


Figure: Using QI from SVDRecoTracks vs. Full RecoTrack MVA QI

- SVD QI's finding Eff. sometimes better, but fake rate goes up
- Full QI's fake rate does not drop as fast as @ SVDonly

Current Results for Full Track MVA

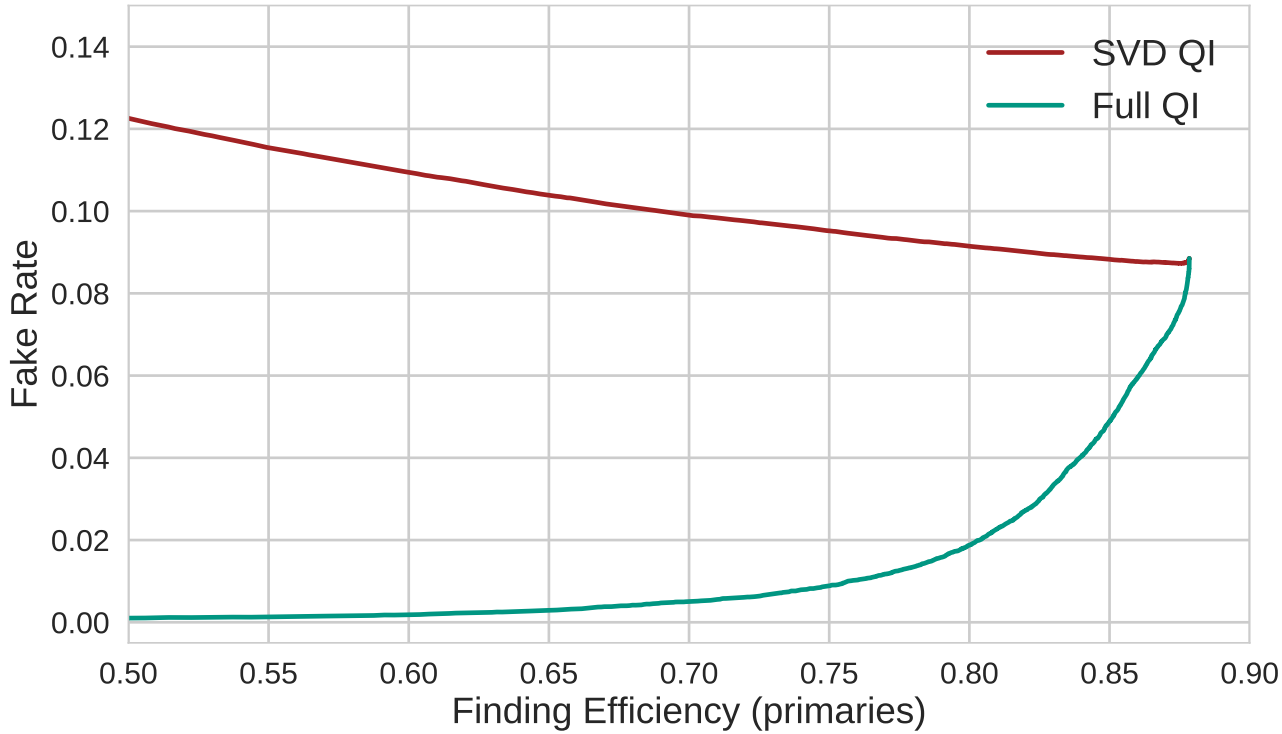


Figure: Fake Rate vs. Finding Efficiency for SVD and Full QI

- Just using the SVD QI (previously discussed for SVDonly tracking) for the full tracking and using it to cut tracks with SVD hits, does not make sense
- The MVA QualityEstimation for the full track can be used for cutting
- Cutting fakes will cut the finding efficiency as well

$\Rightarrow$ Can be possibly improved with the additional info discussed before

Outlook

- See if MVA training needs to be done after other changes
- Put good(&final) weightfiles into the DB
- See if sub-RecoTrack QIs can be used for merging
- Check if cutting some SVD SpacePointTrackCandidates before OverlapCheck makes sense