LHCb-INT-2009-023
LPHE-2009-07
version **1**
October 8, 2009

A b jet “seed” finder

Aurelio Bay[1], **Cedric Potterat**[2]

Laboratory for High Energy Physics, Ecole Polytechnique Fédérale de Lausanne

Abstract

We dicuss a method to find jets initiated by b quarks. An implementation in DaVinci is presented.

---

[1] E-mail:`Aurelio.Bay@epfl.ch`
[2] E-mail:`Cedric.Potterat@epfl.ch`

1 Introduction. Jet algorithms and beauty tagging

Jets of particles are produced by the hadronization of hard quarks and gluons. Several algorithms are available for the identification of jets and for the determination of the associated four-momenta. We consider here only the "cone algorithm", which consists in the integration of the energy deposited inside a cone of given radius R_{max} defined in the $\eta - \phi$ space. The axis of the cone (the "seed") must be provided by some *a priori* knowledge of the jet direction. In the case of b jets this direction can be obtained from the position of the secondary vertices found in the decay chain of the b-hadron.

Jet studies in LHCb have already been presented in [1, 2, 3, 4]. See also the note [5] which describes a neural-net based optimization of the energy measured by a cone algorithm.

The b jet seed finder presented here is inspired by the one described in [2]. The study is Monte Carlo (MC) based. We have used events of 120 GeV/c^2 Higgs produced in association with a Z or W, generated by Pythia and fully simulated, assuming a single interaction per event. The event selection requires a prompt and isolated lepton from W or Z.

In addition we have used events produced by a "jet-gun": a quark of chosen flavor is placed in the particle list of the Pythia generator, set up for independent fragmentation. Several thousands of such events have been produced, with all quark flavors (except t), energies $0.2 < \mathrm{E} < 3$ TeV, the gun spraying a little beyond the whole detector acceptance. In order not to exceed too much the expected range of b quarks from Higgs, we have limited our analysis to jets with the generated quark momenta $\mathrm{P} < 2000$ GeV/c and $\mathrm{P_t} < 150$ GeV/c.

The seed algorithm must tag the b jet with optimal efficiency and purity and provide the direction for the cone algorithm, which will result in the best jet momentum resolution. The second criterion is somehow ambiguous, as already discussed in [5], in which we were aiming at the jet energy reconstruction: there we had the choice to optimize the energy measurement taking as reference the visible energy (corrections associated to detector effects only) or, at the other extreme, to try to recover the original b quark energy (compensation for neutrino emission, hard gluon radiation,...). In the present case, at the level of single events, a comparison of the seed found with the MC truth can follow three possibilities: we can compare with the direction of 1) the b quark just after Higgs decay, 2) the b quark after parton shower (in the string), or 3) the corresponding b hadron. Nevertheless the difference is normally not very large, of the order of 1 degree, as shown in Figure 1 for the b jets from Higgs decay. As expected, the hadron direction is very close to the b quark leading the string. We expect deviations to be large in case of hard gluon emission, i. e. for events with more than two b jets, in which case a poor energy estimate is unavoidable.

2 The "VV" seed finder algorithm

The goal of VV is to identify jets which can be associated with the hadronization of b quarks. It consists in identifying all displaced vertices in the event: all possible 2-tracks vertices are formed, and grouped in sets of similar direction with respect to the primary vertex (V^0). Each set is then used to infer an average vector which is taken to be the seed. As previously said, we only consider the situation with an unique interaction per bunch crossing, so there is no ambiguity in the definition of V^0. The procedure is subdivided into seven phases.

Phase 1, track selection

We select tracks with $\mathrm{P_t} > \mathrm{P_{t-min}}$ and we filter ghosts. The plots for the transverse momenta is given in Figure 2.

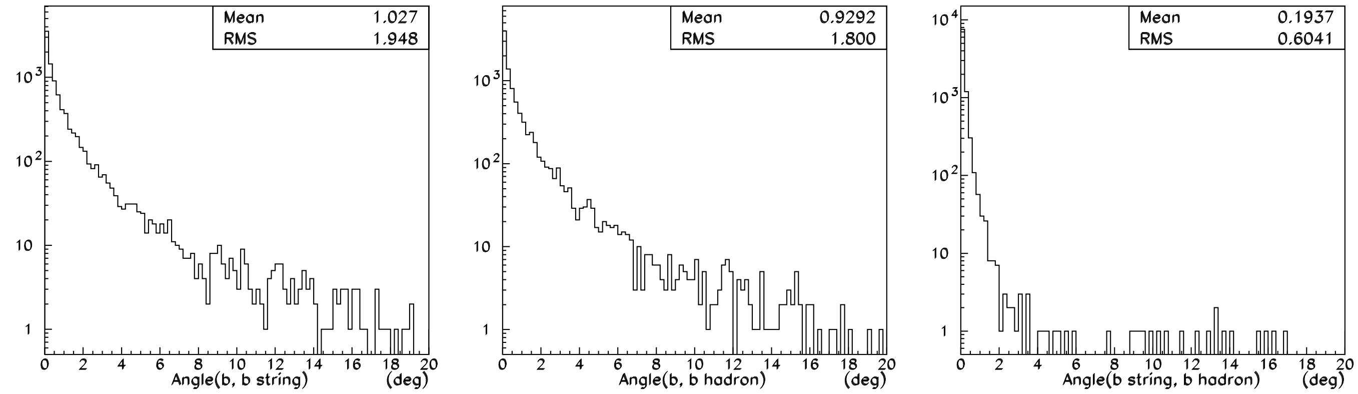

Figure 1: Left: angle from the b quark from Higgs to the b quark before hadronization. Middle: angle from the b quark from Higgs to the b hadron. Right: angle from the b quark before hadronization to the b hadron.

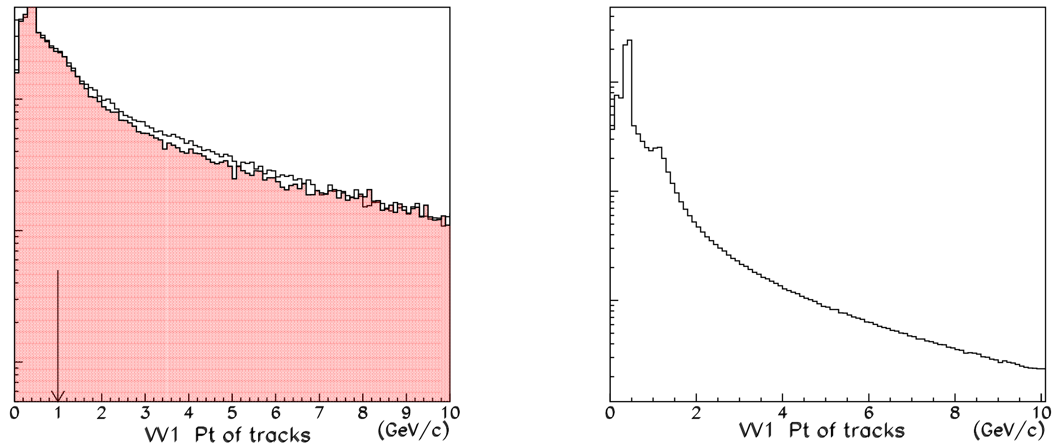

Figure 2: P_t of tracks. Left: events from the jet-gun for b (black) and u quarks (in red), right: from Higgs decay

The arrow indicates the typical cut $P_{t-\mathrm{min}} = 1$ GeV/c. At the left we have the results for b jet-gun data, and from u jets (in red, filled). The two plots are normalized to the same area. On the right the plot from Higgs analysis.

After P_t selection the Impact Parameter (IP) to V^0 is calculated for each track (for jet-gun data, V^0 is the generated origin of the quark, smeared to reproduce the expected V^0 resolution). IP and its significance $\equiv \mathrm{IP}/\sigma(\mathrm{IP})$ are shown in Figure 3. We select tracks with $\mathrm{IP} > \mathrm{IP}_{\mathrm{min}} = 0.1$mm, and $IP/\sigma(IP) > IP_{sig-min} = 3$.

Phase 2, K^0 rejection

We take the selected tracks two by two and discard couples of tracks compatible with K^0: invariant mass within 10 MeV/c^2 and Distance of Closest Approach (DCA) of the two tracks is less than 0.5 mm.

The statistics of the number of tracks selected at this stage is shown in Figure 4. Given the fact

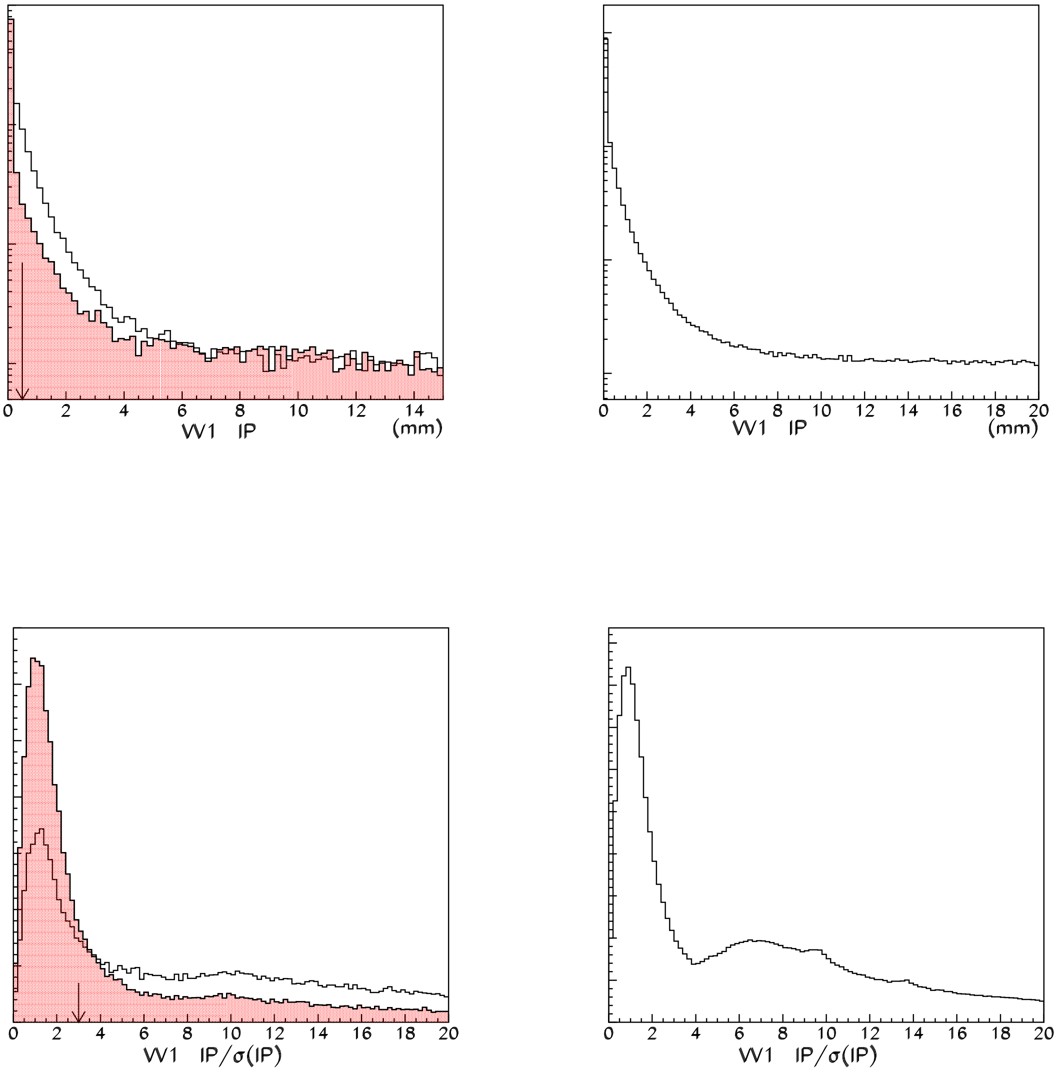

Figure 3: Top: IP of tracks. Bottom: IP significance, IP/σ(IP)

that the Higgs data contains at least two b jets plus the underlying event, it is normal to have more tracks than with single b jets.

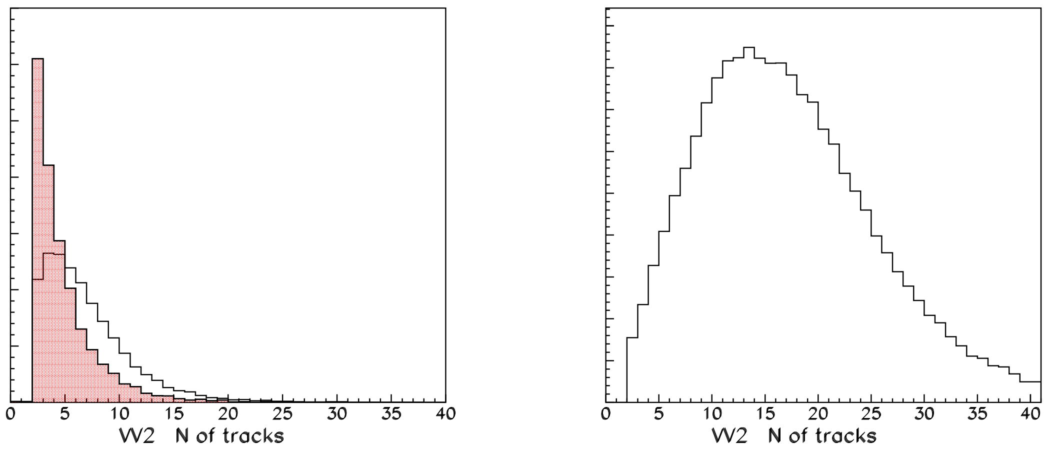

Figure 4: Number of tracks selected after Phase 2

Phase 3, search of displaced vertices

Figure 5 shows the DCA distributions for the couples of tracks. Tracks with DCA $< \mathrm{DCA}_{\mathrm{max}} = 0.5\,\mathrm{mm}$ are selected and used to form a candidate secondary vertex V^{i}. V^{i} is discarded if found upstream: we require $\mathrm{V}^{\mathrm{i}}_{\mathrm{z}} < \mathrm{V}^{0}_{\mathrm{z}}$. Moreover we request V^{i} to be at some minimal distance $\mathrm{d} = |\mathrm{V}^{\mathrm{i}} - \mathrm{V}^{0}|$ from the primary vertex $\mathrm{d} > \mathrm{d}_{\mathrm{min}} = 1.5$ mm, and we limit to the interesting decay region: $\mathrm{d} < \mathrm{d}_{\mathrm{max}} = 200$ mm. Idem in the transverse plane where we compute the radial distance $\mathrm{r} = |\mathrm{V}^{\mathrm{i}} - \mathrm{V}^{0}|_{\mathrm{t}}$, and we require $\mathrm{r}_{\mathrm{min}} < \mathrm{r} < \mathrm{r}_{\mathrm{max}}$, $\mathrm{r}_{\mathrm{min}} = 0.1$ mm, $\mathrm{r}_{\mathrm{max}} = 50$ mm.

The corresponding plots are in Figure 6.

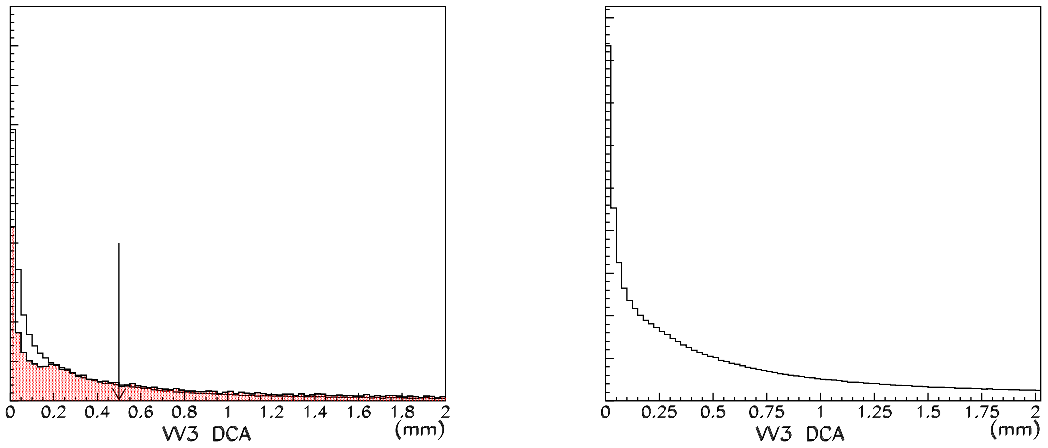

Figure 5: DCA of couples

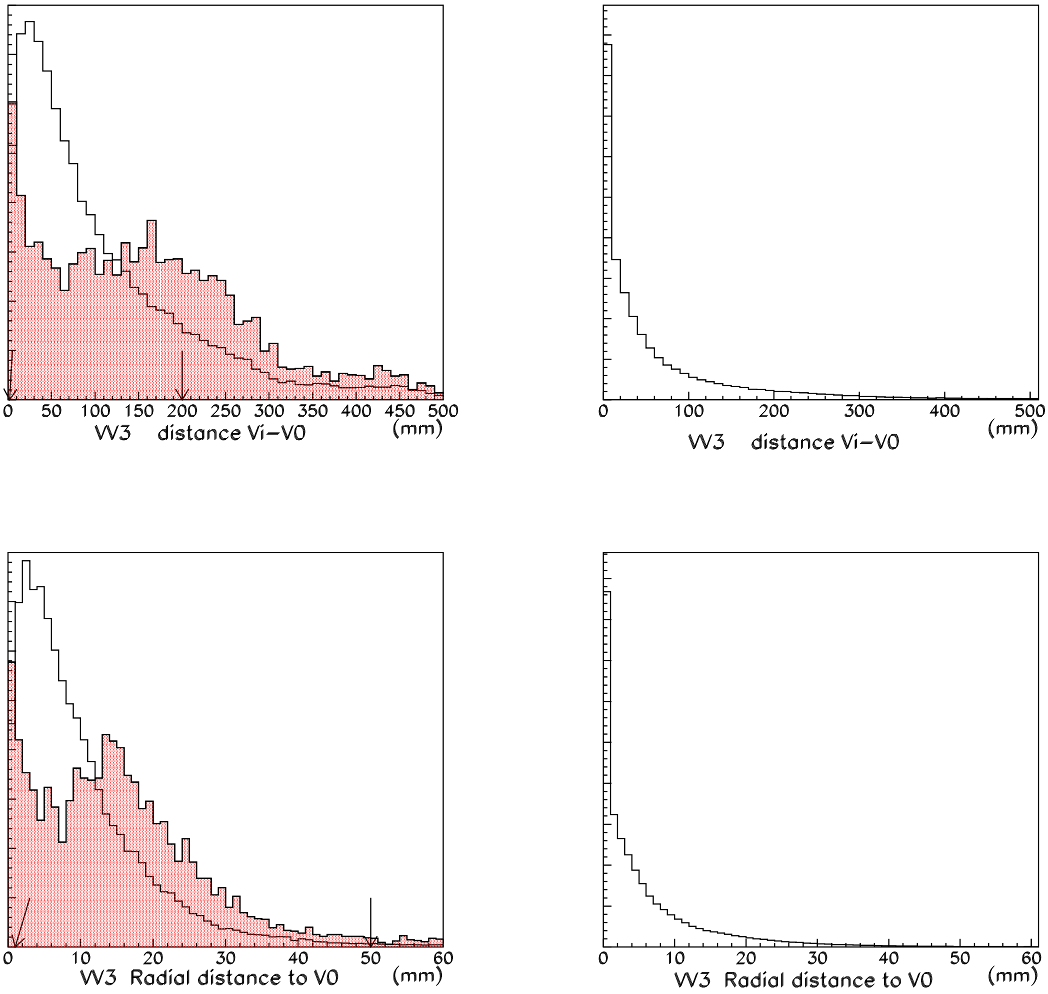

Figure 6: Top: distance of flight $d = |\mathrm{V}^{\mathrm{i}} - \mathrm{V}^{0}|$ in 3D, bottom: projected on the transverse plane $r = |(\mathrm{V}^{\mathrm{i}} - \mathrm{V}^{0})_{\mathrm{t}}|$

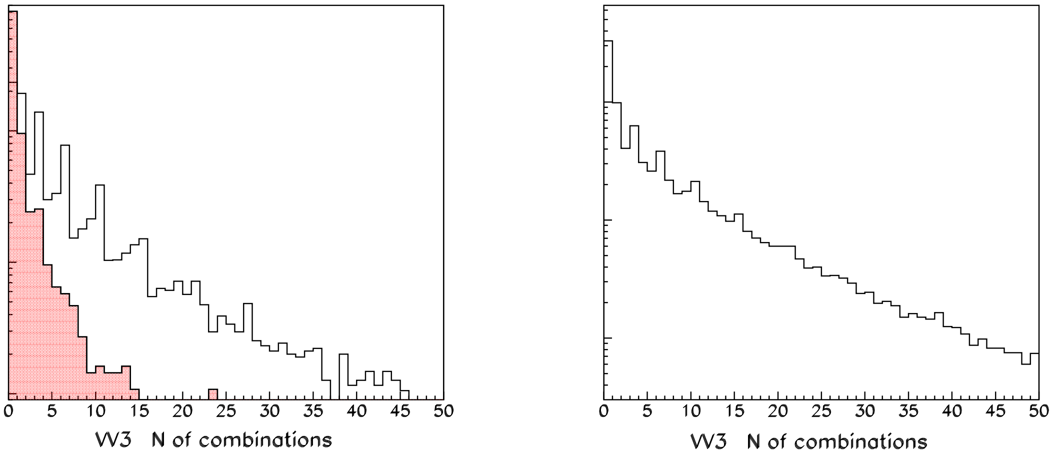

Figure 7: Number of selected two-track combinations after Phase 3

The statistics of the number of $\mathrm{V^i}$ is shown in Figure 7.

Phase 4, construction of "protojets"

By the cone algorithm for each $\mathrm{V^i}$ we construct a "protojet" around the direction from the primary vertex to $\mathrm{V^i}$, with $\eta-\phi$ aperture R_{max} =0.15. The statistics of the number of protojets is shown in Figure 8.

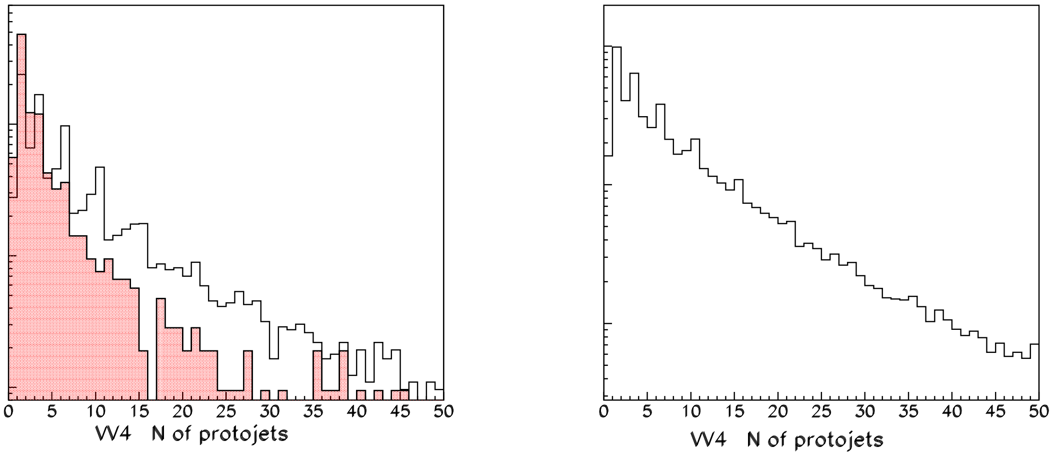

Figure 8: Number of protojets after Phase 4

Phase 5, protojet reduction

We scan all the protojets by pairs and we merge when the $\eta-\phi$ distance is inferior to $\Delta R^{max}_{proto} = 0.45$. This value was chosen because the cone radius for jet calculation is usually larger than 0.5. The procedure is done iteratively: we merge the two closest protojets, and we continue until all the protojets have a distance larger than ΔR_{proto}^{max}. The statistics of the number of protojets after grouping is shown in Figure 9.

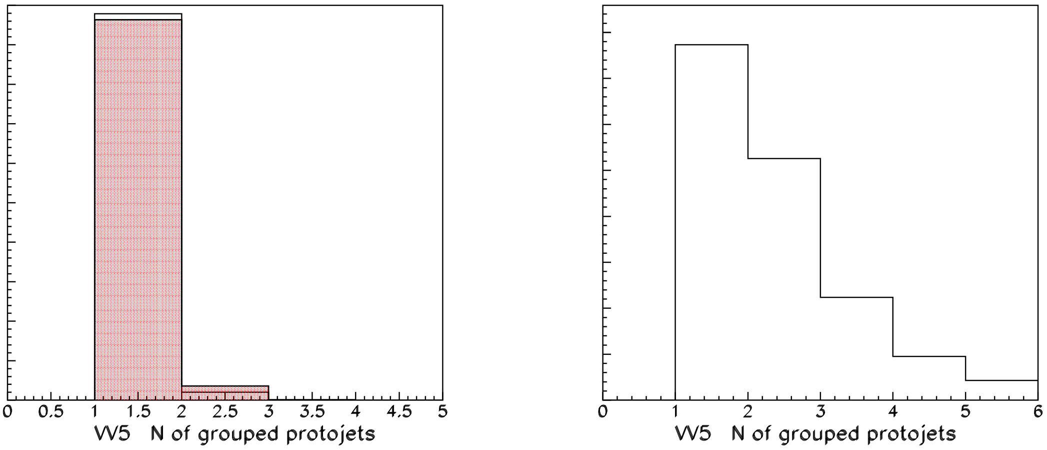

Figure 9: Final number of protojets after grouping in Phase 5

Phase 6, selection of the seeds

We recompute the direction of the protojet with same R_{max} as in Phase 4. We keep the protojet when $\mathrm{P_t} > \mathrm{Pt}_{\mathrm{seed}}^{\mathrm{min}} = 8$ GeV/c, see Figure 10. The protojets surviving are the “seed” candidates $\mathbf{p}^{\mathrm{i}}$. The number of seeds per event that will be transferred to Phase 7 is shown in Figure 11. It must be understood that already about 30% of b jets have been lost at this stage by the previous selection criteria. Note that in this histogram the first bin corresponding to zero seeds only represents the fraction discarded by the $\mathrm{P_t} > \mathrm{Pt}_{\mathrm{seed}}^{\mathrm{min}}$ cut.

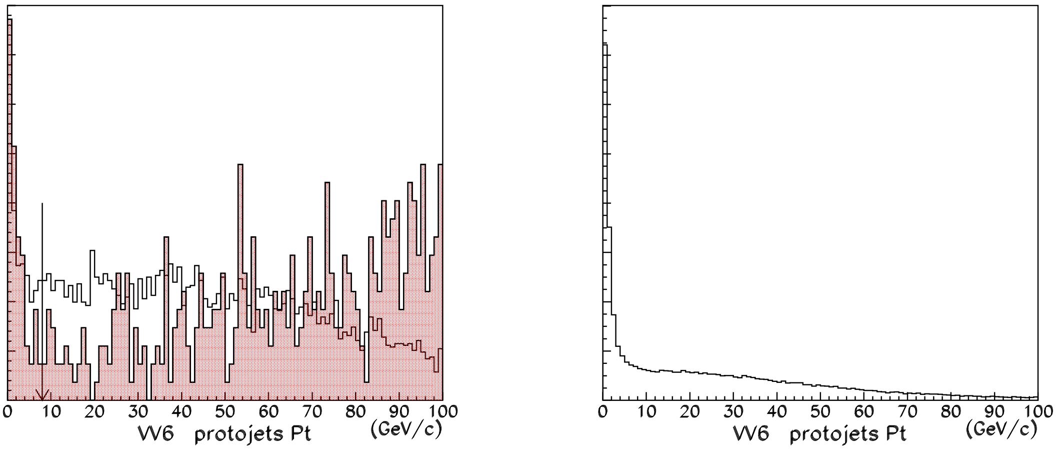

Figure 10: Protojets Pt distribution

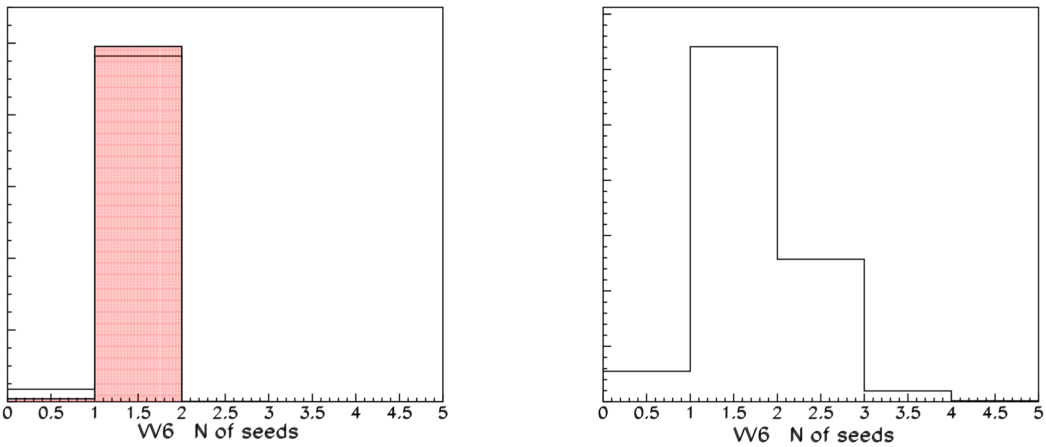

Figure 11: Number of seeds per event after Phase 6

Phase 7, ordering of the seeds

Finally the set of seeds is ordered as a function of the number of V^i in the seeds. For seeds with the same number of V^i they are ordered by decreasing P_t. We expect that the best seeds have a large number of V^i in the associated jet. Figure 12 shows that there is some difference between the number of V^i produced by b jets and what is found in the surviving u jets.

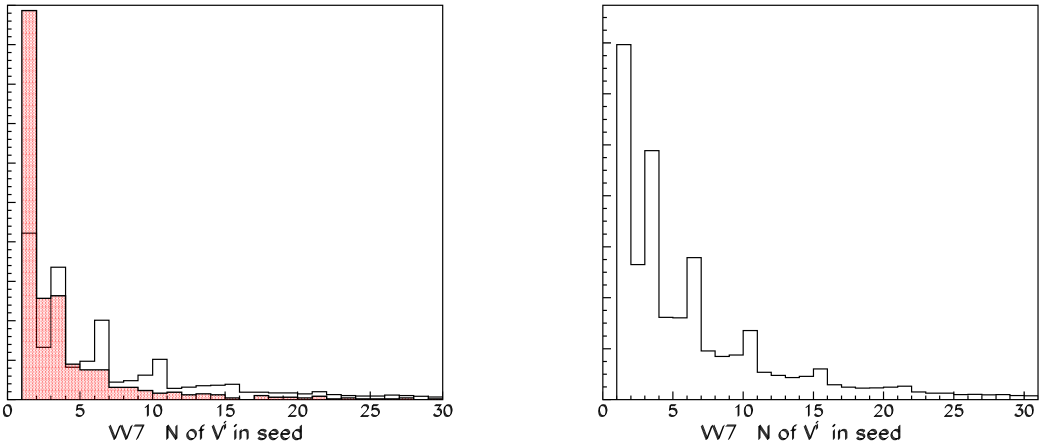

Figure 12: Number of V^i in each seed

3 Tests

From the b jet-gun events and the set of cuts and parameter as indicated, the efficiency for finding at least one seed is of about 68%. The events were generated with flat quark momentum, and azimuthal angle distributions. We have also computed weights which are function of the quark P and P_t to match the two-dimensional spectrum of b quarks produced from Higgs decay. If we use these weights in the calculation of the number of seeds, the efficiency to find the jet in b jet-gun data grows to 74%.

We have tested our setup on the Higgs events, selected to have two b quarks in the acceptance pseudorapidity window of $[1.9, 4.7]$. The distribution of number of seeds found in Higgs events, is shown in Figure 13, together with the di-jet reconstructed mass. We see that 48% of the events have at least two seeds found, while 3% of events have more than two seeds (the extra seed number is controlled by R_{max}, as we will present later). We have fitted a binomial distribution, assuming two b quarks in the acceptance, plus a Poisson function to describe the source of extra seeds. The fitted binomial probability is p=0.66, while the Poissonian component has μ =0.08. So the efficiency to find one seed is 66%, which has to be compared close to the 74% obtained with single jets. This might come from the presence of an underlying event, or from the fragmentation model.

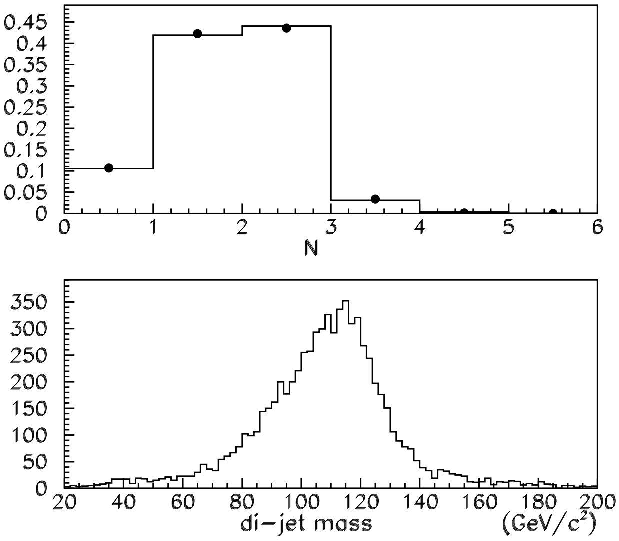

Figure 13: Top: distribution of the number of seeds found in Higgs events, the histogram normalized to 1 and the dots are the result of the fit described in the text. Bottom: the di-jet mass reconstructed from the seeds and using the procedure described in [5]

The angular distance of the seeds from the b quark is shown in Figure 14. We see that we are very close to the four-vector results of the first two histograms of Figure 1. On the right figure we give the distance in $\eta - \phi$ space, with an average of 0.14.

Figure 15 displays the pseudorapidity distribution of all the b quarks, and for the events in which VV did not find seeds. For this analysis we have released the pseudorapidity cut. The difficult region is seen to be for $\eta < 2$. We also observe a drop of efficiency at high η. In order to better understand our efficiency we have restricted to an inner acceptance region requiring the b quarks to fall in the pseudorapidity window $[2.3, 3.5]$. With this restriction the distribution of number of seeds found is shown in Figure 16, left. The fraction of events with at least two seeds is now about 56%, and 4% of the events have extra seeds. The fitted binomial average is p=0.74.

As an alternative we have the possibility during Phase 2 to search for triplets, instead of couples of tracks compatible with a common vertex. Without any change in the parameters and always with the restriction to the inner part of the detector, we find p=0.58, see Figure 16, right.

A further test was done by selecting “visible” Higgs events: as before we require the two b quarks in the inner rapidity region and in addition we ask that the di-jet invariant mass from MC truth be inside the 70–160 GeV/c^2 window (the seeds for the cone jet calculation with the MC four-vectors

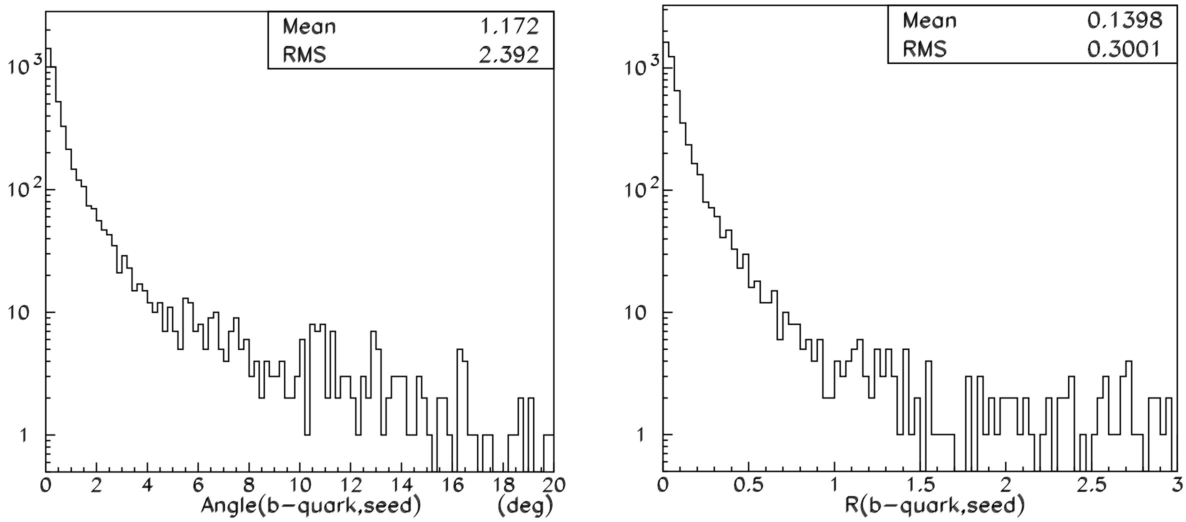

Figure 14: Left: angle between seed and b quark, right: distance in $\eta-\phi$ space

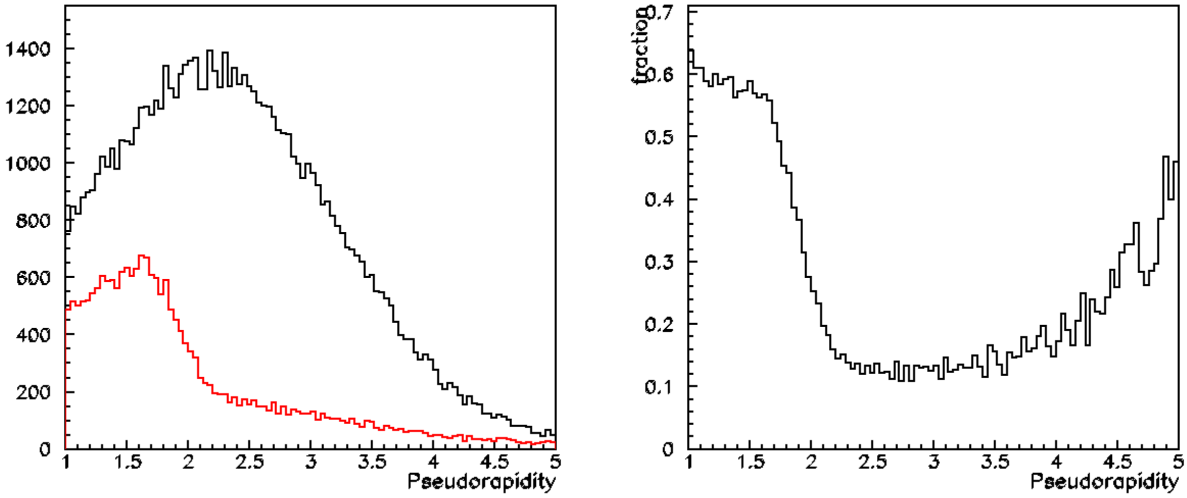

Figure 15: Pseudorapidity of the generated b quarks. In red the pseudorapidity of the two b quarks for events with zero seed found. The fraction of events with zero seeds is shown on the left plot as function of pseudorapidity

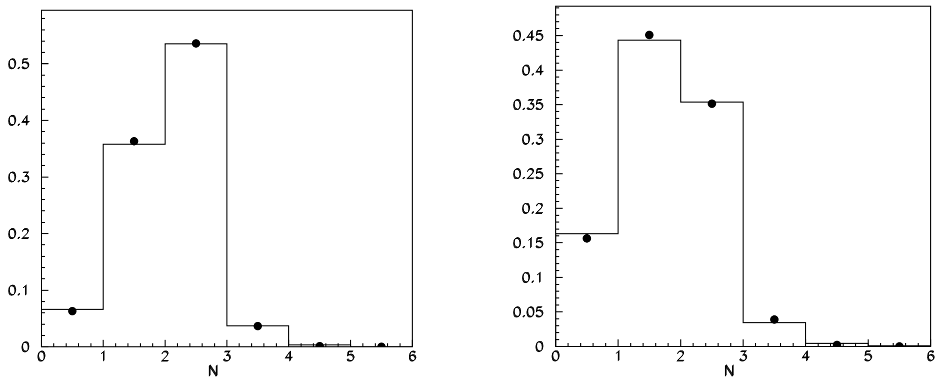

Figure 16: Number of seeds distribution in Higgs events, for b quarks in the inner part of the detector, $\eta = [2.3, 3.5]$. The histogram is normalized to 1. The dots are the result of a fit, see text. On the left the result using doublets of tracks, on the right using triplets

are given by the b quark direction, the particles outside acceptance and neutrinos are discarded). After this tight selection we did not observe any significant change in the number of seeds found. Figure 17 shows the seed number distribution and the reconstructed di-jet invariant mass obtained, the fitted value of p is 0.75.

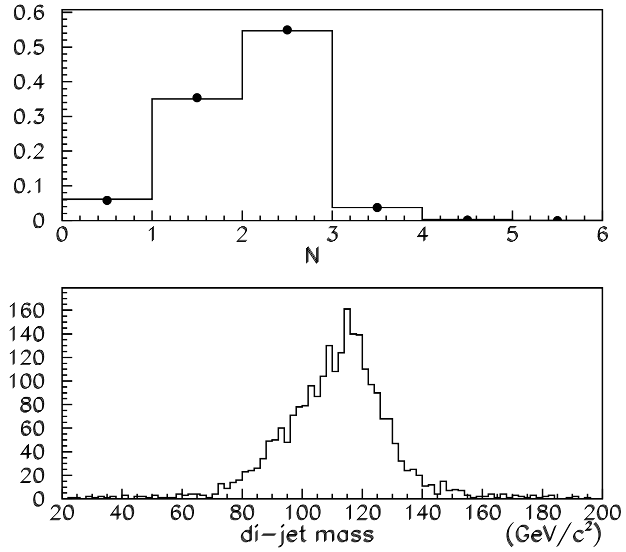

Figure 17: Top: N of seed distribution after selection of the MC truth di-jet mass value in 70–160 GeV/c^2. Bottom: reconstructed di-jet mass

Light flavor discrimination was assessed with the help of jet-gun data. For an efficiency to b quarks of 68%, we have a charm rejection of 72%, and lighter quarks rejection at the level of 90%. These results have been obtained with quarks generated with flat momentum and azimuthal distributions. When re-weighting all the distributions as a function of P and P_t, taking the same weights for all the flavors, the b acceptance is 0.74, as seen before, with no change for the other flavors. These results remain stable if we restrict to the inner detector region.

The choice of the VV parameters was inspired by previous studies, mainly discussed in [2]. The optimization should be done by a scan over some range, the physical variables of interest plotted as a function of the parameter value. Examples are given in Figure 18, where the variables shown are the p and n values fitted to the distributions of the number of seeds found. The analysis was done with the same conditions as before, with an isolated lepton, and two b in the inner part of the detector. We can see that the choice of R_{max} lower than about 0.2 reduces the number of spurious seeds to a very small fraction, while the efficiency stays close to 75%. All the other plots have been obtained with R_{max} =0.15.

Another criterion is to improve the b jet detection efficiency versus other flavor rejection, see Figure 19. Here we have used the jet-gun data, with quarks in the inner part of the detector, and the results weighted to match P and P_t of the b quarks from Higgs. We cans see that a choice of the min IP significance of 40 gives 30% efficiency and a 97% rejection for c and u .

4 Conclusion. Algorithm implementation

We have described an algorithm (dubbed VV) for the determination of the seeds to be used in the reconstruction of b jets. The seeds can be used either in a cone algorithm or sent to the tool presented in [5].

The typical efficiency of VV to discover b jets is of 70%. The rejection of jets initiated by c quarks is of about 70%, and 90% for lighter quarks. For a reduced b efficiency of 30% the other quarks are rejected in 97% of the cases. Enhanced flavour discrimination criteria will be addressed in further studies.

The angular precision to find the original b quark direction is about 1 degree, compatible with four-vector (generator level) predictions.

An experimental version of VV, the “VVSeedsFinder” procedure, is described in the appendix, with an example.

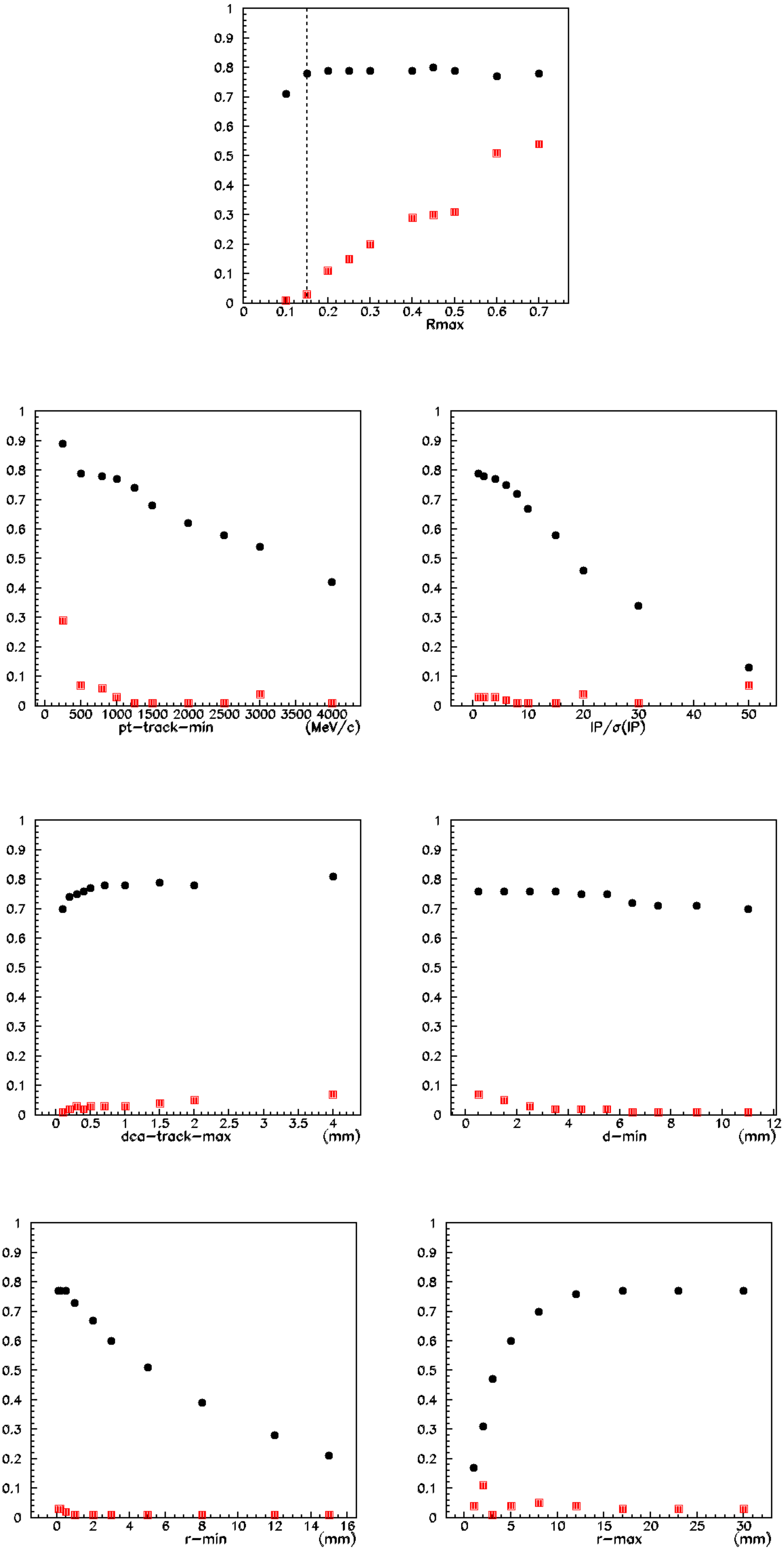

Figure 18: Fitted values of p (black dots) and n (red squares) as a function of various VV parameters. The top plot is for R_{max} with the selected value shown by a dashed line. The other figures, from top left to bottom right, are: tracks P_{t-min}, $IP_{sig-min}$, DCA_{max}, d_{min}, r_{min}, and r_{max}

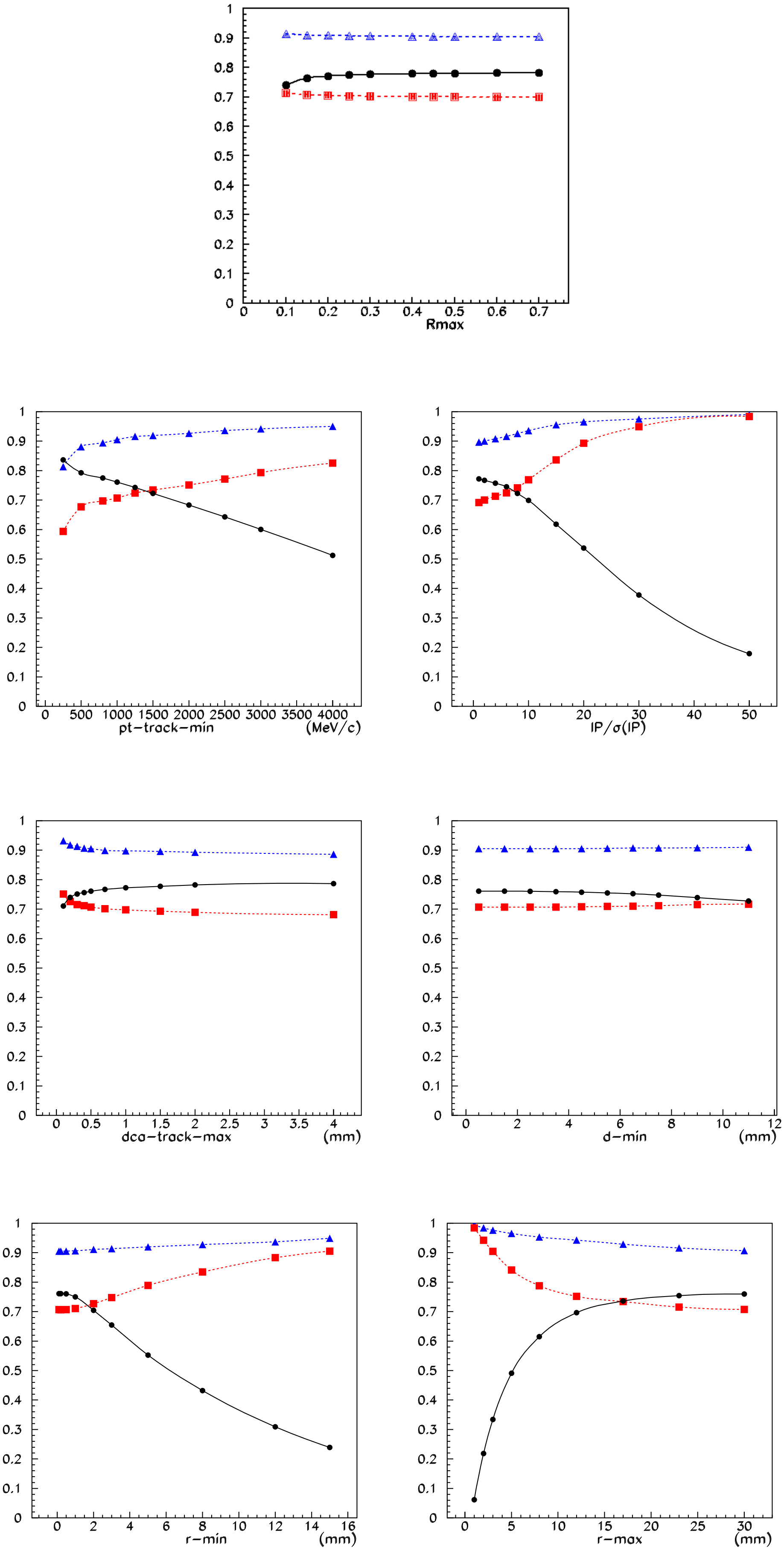

Figure 19: Selection yield for b jets (dots), rejected fraction for c (squares), and u (triangles), as a function of various VV parameters already presented in Figure 18

APPENDIX: the VVSeedsFinder tool

The algorithm finds the seeds to be used in a cone jet algorithm and place them on the TES with a specific identifier (PID). The current version of the code can be used to feed the Neural Network tool described in [5].

C++ code

Add the following line in your requirement file:

```
use                 VVSeedsFinder     v* Phys
```

Then include the tool interface o *"Kernel/IVVSeedsFinder"* and declare a pointer $m\_getVVSeeds$ to the class *IVVSeedsFinder* as a global variable:

```
#include "Kernel/IVVSeedsFinder"
IVVSeedsFinder* m_getVVSeeds = 0;
```

You are now allowed to use the tool. The pointer needs to be initialized once:

```
if ( 0 == m_getVVSeeds )
m_getVVSeeds = tool<IVVSeedsFinder> ( "VVSeedsFinder" , this );
```

The tool requests all the particles and the primary vertex as input. It is recommended to use the *std::vector<LHCb::Particle>*, which provides more information about the construction of the seeds. The tool provides a vector of seeds (*std::vector<LHCb::Particle>* or *std::vector<Gaudi::LorentzVector>*).

```
std::vector<LHCb::Particle> VVSeeds;
LHCb::RecVertex::Container::const_iterator ipv = PVs->begin();
LHCb::RecVertex V0 = *(*ipv);
LHCb::Particle::ConstVector Parts = desktop()->particles();

StatusCode sc =  m_getVVSeeds->GetVVSeeds( VVSeeds, Parts, V0 );
```

VVSeedsFinder options

It is recommended to use StdTightMuons, StdTightElectrons, StdNoPIDPions, and StdLooseAllPhotons, in this specific order, as input particles:

```
MyDVAlg.InputLocations = [ "StdTightMuons","StdTightElectrons",
                           "StdNoPIDsPions", "StdLooseAllPhotons" ]
MyDVAlg.addTool( VVSeedsFinder() )
```

Here is a condensed description of the VVSeedsFinder analysis parameters and options, with their default values, as described in the text (except $d_{min} = 2.$ mm instead of 1.5).

Keyword	default value	name	meaning
General			
FilterPart	true		discard particles loaded twice
JetConeID	10098		PID of the seeds
VVNJetsMin	1		min n of seeds needed (to speed up calculation)
Phase 1			
VVPtTrackMin	1000 [MeV/c]	P_{t-min}	$\mathrm{P_t}$ min for track selection
VVIPmin	0.1 [mm]	IP_{min}	IP min for track selection
VVSignif	3.0	$IP_{sig-min}$	IP significance min for track selection
Phase 2			
VVDMK0	10 [GeV/c^2]		window for K^0 rejection
VVDtrakMaxK0	0.5 [mm]		dca max for pions to form a K^0
Phase 3			
VVtriplets	0		use triplets of tracks when true, else doublets
VVDtrakMax	0.5 [mm]	DCA_{max}	dca max for track doublets (triplets) to form a vertex V^{i}
VVChi2min	0		min chi2 in vertex calculation
VVChi2max	1000		max chi2
VVTseedVtxMin	2.0 [mm]	d_{min}	min distance d to accept V^{i}
VVTseedVtxMax	200 [mm]	d_{max}	max distance d
VVDRmin	0.1 [mm]	r_{min}	min radial distance r to accept V^{i}
VVDRmax	50. [mm]	r_{max}	max radial distance r
Phase 4			
VVRParameter	0.15	R_{max}	radius in $\eta - \Phi$ for "protojet" calculation
Phase 5			
VVDeltaRSeeds	0.45	ΔR^{max}_{proto}	min distance between couples of protojets, else merge
Phase 6			
VVPtSeedsMin	8000 [MeV/c]	Pt^{min}_{seed}	min $\mathrm{P_t}$ of the seed

Example of usage

This program gets the seeds, prints the seed direction, and sends them to the NN algorithm to reconstruct the jets.

```
 main{

 std::vector<LHCb::Particle> VVSeeds;
 LHCb::RecVertex::Container::const_iterator ipv = PVs->begin();
 LHCb::RecVertex V0 = *(*ipv);
 LHCb::Particle::ConstVector Parts = desktop()->particles();

// Get the seeds using VVSeedsFinder:

 StatusCode sc = m_getVVSeeds->GetVVSeeds( VVSeeds, Parts, V0 );

// scan the seed vector:

 for(std::vector<LHCb::Particle>::iterator ip=VVSeeds.begin(): ip != VVSeeds.end(); ++ip)

  cout << "direction of the seed : " << *ip.momentum() << endl;

// feed the NN cone algorithm to get the optimized jet

  sc = m_jetcalib->GetCalibratedJets( VVSeeds,Parts,V0);

// print some information,
// then select jets with invariant mass > Mmin and copy in a Particle vector named Jets

 std::vector<LHCb::Particle> Jets;
 for(std::vector<LHCb::Particle>::iterator ip=VVSeeds.begin(): ip !=  VVSeeds.end(); ++ip)

   cout << "- mass of the jet                       : " << *ip.momentum().M()   << endl;
   cout << "- nb of protojets merged in this jet    : " << *ip.weigth()         << endl;
   cout << "- vertex position of the most                                      "<< endl;
   cout << "  energetic protojet merged in this jet : " << *ip.referencePoint() << endl;

   if (*ip.momentum().M()>Mmin) {
         ... copy VVseds in Jets ...
   }
```

References

[1] C. Currat, A. Bay, M. Koratzinos, “Jet studies in LHCb”, LHCb-99-016 PHYS (1999)

[2] C. Currat, “Direct search for Higgs boson in LHCb”, Thesis of the University of Lausanne (2001) CERN thesis-2001-024

[3] L. Locatelli, “Direct search for Higgs boson in LHCb and contribution to the development of the Vertex Detector”, Thesis of the Ecole Polytechnique Fédérale de Lausanne, `http://lphe.epfl.ch/engl/publications/theses.html` (2007)

[4] V. Coco, “Reconstruction et identification de jets beaux dans l’expérience LHCb en vue d’étudier sa sensibilité à un boson de Higgs standard se desintégrant en paires bbbar”, Thesis of the University of Savoie (2007)

[5] A. Bay, C. Potterat, “Optimization of the b jet energy measurement in LHCb”, LHCb-INT-2009-024 (2009)