Hera-B 99-063
Software 99-007

Recommendations for Evaluating the Performance of HERA-B Reconstruction Modules

Rainer Mankel

Humboldt University Berlin

12-Jan-1999

---

1. Introduction

The complete HERA-B reconstruction will be achieved with a delicate interplay of many different program modules performing reconstruction tasks in different parts of the detector. It is essential to assess the performance of each module in a transparent and consistent way. This is also important in cases where several reconstruction approaches are investigated and the most suitable one for a specific reconstruction task has to be selected.

In the following, some guidelines for a consistent performance evaluation are collected.

2. The reference set

For experiments as HERA-B, the particles relevant for physical analysis are accompanied by a host of additional particles, as background from superimposed interactions, or secondary particles produced through interactions within the detector material. It is neither feasible nor advisable to

consider the reconstruction of all these particles. Hence, a set of *objects of interest* has to be selected for which efficiencies are to be calculated. Typical *interest sets* are the tracks from the golden B decay, or charged particles with momenta above 1 GeV from a primary vertex.

A second ingredient is the geometrical range of the detector. It does not make sense to judge the performance of a reconstruction program on tracks which which are outside the acceptance of the detector, or just straddling its border. Objects of interest within the geometrical acceptance region define the *reference set*.

In many cases, the geometrical criterion can be conveniently defined by the number of detector elements, eg. the number of layers in a tracking system, which have been passed by the particle, which can be derived from the *Monte Carlo Impact Points (MIMPs)* provided by the detector simulation. Example: for the vertex detector, tracks within the acceptance region pass at least three superlayers. Assuming that each superlayer has four logical GEDE planes on two detector modules per quadrant, giving rise to two MIMPs per superlayer, a requirement of at least 6 MIMPs for a reference particle may be appropriate. It is permissible to discard MIMPs outside the sensitive region from this count but it should be stated.

3. Geometrical acceptance

The definition of the reference set can be regarded as a definition of geometrical acceptance for the particle type X:

$$\varepsilon^{X}_{geo} = N^{X}_{ref} / N^{X}_{tot}$$

which is independent of the hit efficiencies of the individual detectors. Since the definition of the reference set has potential influence on the reconstruction efficiency, the corresponding geometrical acceptance should always be considered as well.

4. Efficiency

In order to define the reconstruction efficiency, a reconstructed object must be *assigned* to a Monte Carlo particle. There are two basic approaches for such an assignment:

- parameter matching
- hit matching

Parameter matching is sometimes the only way because it requires not direct Monte Carlo relation information between generated particle and the detector hits used in the reconstruction. In high particle densities however, it bears the danger to accept random coincidences between Monte Carlo particles and ghost reconstructions and can lead to the paradox impression that the reconstruction efficiency *improves* with increasing hit density.

The recommended technique for HERA-B is therefore based on hit matching: the reconstructed track is *assigned* to the Monte Carlo particle which contributes the largest fraction of its hits if this fraction exceeds a certain limit. The contribution fraction of a Monte Carlo particle *m* is defined as

$$f_m = N^m_{hits} / N_{hits}$$

where N_{hits} is the total number of detector hits on the reconstructed track and N^m_{hits} the number of detector hits to which the MC particle *m* has contributed. Note that since several MC particles may contribute to the same strip cluster etc, the sum over all f_m for a reconstructed track can in general exceed unity.

The track is assigned to the particle of largest contributed fraction if it exceeds a lower limit f_{min}. A typical choice for tracking applications is f_{min}=70%.

A particle is called reconstructed if at least one reconstructed object is assigned to it. The reconstruction efficiency is then

$$\varepsilon = N_{ref}^{reco} / N_{ref}$$

ie. the fraction of reference tracks which are reconstructed.

5. Ghosts

Reconstructed objects which are not assigned to a Monte Carlo particle are called *ghosts*.

A ghost rate can be defined as

$$\varepsilon_{gh} = N_{ghost} / N_{ref}$$

It is informative to also provide the mean number of ghosts per event.

6. Clones

The above definitions for efficiency and ghost rates are intentionally insensitive to multiple proper reconstructions of particles. Such redundant reconstructions are called *clones*. For a given particle *m* with N_m^{reco} reconstructed objects assigned to it, the number of clones is

$$N_m^{clone} = \begin{cases} N_m^{reco} - 1 & \text{, if } N_m^{reco} > 0 \\ 0 & \text{, otherwise} \end{cases}$$

and one can define the *clone rate* as

$$\varepsilon_{clone} = N^{clone} / N_{ref}$$

7. Parameter Estimates

The quality of reconstructed particle parameters in case of reconstruction in a detector subsystem is essential for matching and propagation into other subsystems. For the whole detector, it determines directly the physics performance.

The quality of the parameter estimate is reflected in the parameter residual

$$R(X_i) = X_i^{REC} - X_i^{MC}$$

where X_i is a reconstructed particle parameter (eg. x, y, t_x, t_y, Q/p, ...). The corresponding Monte Carlo value can usually be inferred from the Arte tables `MTRA`,`MIMP`,`MCAL` etc. From the parameter residual distribution, one can then obtain the parameter estimate bias, $\langle R(X_i) \rangle$, and the parameter resolution given by the width, which can be expressed in terms of the result of a Gaussian fit, the RMS or FWHM, whichever appears appropriate with regard to the shape of the distribution.

If the reconstruction tool provides also an estimate of the parameter covariance matrix (C_{ij}), it is very informative to investigate the *normalized parameter residual* or *pull*, which is defined as

$$P(X_i) = (X_i^{REC} - X_i^{MC}) / (C_{ii})^{1/2}$$

Ideally, the pull should follow a Gaussian with mean value zero and standard deviation one.