Chapter 5

Analysis

In this chapter, we detail the analysis, beginning in event reconstruction and ending in the application of optimal analysis cuts, carried out on raw data – both the ~400 million *real* events written to 6250 bpi tapes during the course of the experiment and the *fake* events obtained by digitizing the output of the MC generator.

5.1 Event reconstruction

Reconstruction proceeds in three stages: PASS0, PASS1, and PASS2. In PASS0, the first 2000 events from each tape are used to collect miscellaneous statistics useful in troubleshooting, beam particle identification, and calibration. These include hot SMD channels, pion and proton peaks in the TRD distributions, and ADC pedestals. During PASS1 and PASS2, various other bookkeeping activities coming under the PASS0 umbrella are carried out on an event-by-event basis.

In PASS1, charged particle trajectories in the four regions of the spectrometer are found by fitting SMD and/or DC hits into straight-line segments called "tracks". Tracks found in the SMD system, with its high resolution and efficiency and low noise, (with corroboration provided by D1 and downstream PWC hits) are linked to tracks downstream of M1 (provided by D2-D4). Assuming a single bend point, linking begins in the y-z plane (on track projections which are straight lines in this appromixation) and continues in the x-z view. The resulting track candidates are then refit using the full magnetic field mapping, and momenta are determined.

In PASS2, the tracks found in PASS1 are projected through the electromagnetic and hadronic calorimeters. This allows for measured showers (energy depositions above threshold in contiguous channels) in the calorimeter to be associated with charged tracks or, if this is not possible, identified with neutral particles. In the former case, the shape and penetration of the shower is used to distinguish electrons and muons from hadrons.

If the charged track is more consistent with being a hadron than a lepton, Čerenkov information is used to distinguish between pions, kaons, and (anti)protons. The track in question is projected through C1 and C2; in each counter, the number of photoelectrons produced in the relevant phototube channel is counted and used to calculate the probabilities that the track in question can be identified with π, K, or p. The final set of particle-ID probabilities is the normalized product of the C1 and C2 probabilities and the *a priori* probabilities[1] for charged secondaries in E769 events.

The final phase of PASS2 is the compilation of a vertex list. Fits are performed on groupings of SMD tracks, to which tracks are iteratively added one by one until the vertices with the largest number of tracks that are consistent with a quality cut of χ^2/dof less than 2 are obtained. The process continues among the leftover tracks. Except for the vertex with the highest number of tracks (which is identified with the primary vertex), none of the vertices in this list are used in subsequent analysis; secondary vertex fits are attempted for *all* track combinations consistent with a particular decay mode hypothesis.

Upon completion of PASS2, nonessential information is discarded and the remaining portion of each event is written in a compressed Data Summary Tape (DST) format onto an 8 mm 2.3 Gbyte exabyte tape. Several hundred tapes comprise the reconstructed data set.

An error was discovered in E769's reconstruction code after the data had been processed. In the software specification of the detector elements, the z positions of two adjacent SMD planes (x and y-views) had been transposed. This mistake obviously impacted the quality of the track fits and increased background by introducing an

[1]These *a priori* probabilities are as follows: 2% e, 1% μ, 81% π, 12% K, and 4% p.

additional source of spurious (by their misplacement) hits. Rather than re-reconstruct the entire data set, the collaboration decided to correct the reconstruction program and run it only on a subset of the data that had been filtered to enhance charm content. This filter, known as the pair strip, selects events containing two-track vertices well separated in z from the primary vertex; it is described in Section 5.3. This procedure (reconstruction $\rightarrow$ pair strip $\rightarrow$ corrected reconstruction) is carried out on both the data and the MC events used in this analysis.

5.2 Cut variables

For purposes of this analysis, each event after reconstruction consists of a set of charged particles whose trajectories, momenta, and charge sign are measured. Assigned probabilities for the possible identities of each charged particle differ from *a priori* values when additional information is available. Each event also contains a vertex list representing plausible groupings of particles based on their proximities at birth. In addition to this "final-state" portion of the event, information relevant to normalization (e.g., spill number and trigger type) and beam particle identification (e.g., DISC and TRD output) is also needed for cross-section measurements. In this chapter, however, we are concerned only with characterization of events in terms of their charm content.

Although three of the four D mesons which we attempt to reconstruct are charged, they are much too short-lived[2] for any attempt to be made to track their paths from production to decay directly. Rather we obtain the 4-momentum of the D by summing the 4-momenta of its charged decay products.

Even after the imposition of a transverse energy trigger, we expect only a minority of the remaining events to contain charm particles; the challenge of isolating these events remains. In addition, E769 events typically contain more than a dozen

[2]See Table 7.2 for a list of average lifetimes τ of the pseudoscalar D mesons. (Vector D's decay to pseudoscalar D's "instantaneously".) Take, for example, the relatively long-lived D^+ ($c\tau = 317\mu$). Boosting from the rest frame of D^+'s produced at some fixed value of x_F, we obtain an exponentially-falling distribution of lab-frame birth-to-death z-spans (Δz) already greatly diminished at a distance of one centimeter. In other words, the vast majority of produced D's decay while still in the target region.

charged tracks. Therefore, even for a *pure* sample of charm events (realizable in MC), we expect invariant mass plots generated from all track triplets[3] to be plagued by an unacceptably high level of combinatoric background. In order to do reduce this background, selection criteria or "cuts" are applied to each triplet; these cuts are designed to favor the passage of real charm decays over randomly-associated tracks (which we'll call "fake decays"). Both the topology of particle production and decay in general and the kinematics of charm decays in particular are exploited in defining quantities whose distributions for real and fake decays differ significantly. If for a particular cut variable this difference is great enough, a cutoff value (above or below which triplets are thrown away) can be found which leads to a high rejection of fake decays while retaining a reasonable fraction of the actual charm signal. In this section, we define the cut variables used in this analysis.

E769's vertex resolution in z is on the order of hundreds of microns. Given that D lab-frame z-spans of millimeters are typical, it is clear that we should be able to make significant measurements of D lifetimes. Moreover, since most tracks in a minimum bias event emanate from the primary interaction point, we expect the separation in z of the primary and secondary vertices to be useful as a cut variable. Rather than cutting on Δz directly, however, the effect of variable z resolution is diminished by measuring Δz in units of $\sigma_{\Delta z}$, where

$$\sigma_{\Delta z} = \sqrt{\sigma^2_{z\,primary} + \sigma^2_{z\,secondary}}. \tag{5.1}$$

The σ's for each vertex ("primary" indicates production, "secondary" indicates decay) are the expected standard deviations in z (based on the number of tracks making up a vertex, their quality, and the effect of multiple scattering) returned by the least-squares vertex fitting algorithm. We call the resulting ratio SDZ (which stands for "significance of Δz"); it is given explicitly by

$$\text{SDZ} = \frac{\Delta z}{\sigma_{\Delta z}} = \frac{z_{secondary} - z_{primary}}{\sigma_{\Delta z}}. \tag{5.2}$$

[3]In this analysis, D^0 decay products form only a pair. Nevertheless, the term "triplet" is used to designate any set of secondary charged tracks nominated as a candidate for a particular charm decay.

From now on we will abbreviate "*primary*" and "*secondary*" to "*pri*" and "*sec*", respectively.

At this point, we should define exactly what is meant by z_{pri}. The vertex in the standard vertex list made up of the most tracks is designated the primary vertex.[4] If a given triplet, however, contains one or more tracks from this vertex, they are removed and the primary vertex is *refit* using only the remaining tracks. Variables such as z_{pri} and $\sigma_{z\,pri}$ are the results of these adjusted fits.

Another property which distinguishes charm decays is the large amount of kinetic energy available to the decay products. The pseudoscalar decays reconstructed in this analysis have Q values on the order of a GeV, leading to decay pions and kaons which are relativistic in the charm parent's rest frame. By summing the squares of p'_T for each decay product, where p'_T is the momentum transverse to the lab-frame composite momentum ($\vec{P} \equiv \sum_i \vec{p}_i$), we obtain a measure (called PT2DK) of the energy of the secondaries:

$$\text{PT2DK} = \sum_i (p'_T)^2_i, \tag{5.3}$$

where i runs over the decay products. Momenta are evaluated in the lab frame.

The hypothesis that a given secondary vertex is a real decay is strengthened if the composite momentum points back to the most probable point of production, the primary interaction vertex. The cut variable DIP is defined as the transverse impact parameter (b_T) of the candidate D with respect to the primary vertex:

$$\text{DIP} = (b_T^{pri})_D. \tag{5.4}$$

The x and y coordinates of the D at z_{pri} are obtained by using the measured slopes dP_x/dP_z and dP_y/dP_z to work backward from the measured position of the secondary vertex.

For a given triplet, we'd also like to have confidence that on average the tracks are more consistent with originating from the secondary vertex than from the primary

[4] Alternative methods for picking the primary vertex, such as using the most upstream vertex in the target region, were compared in MC studies with the procedure actually used; the latter was found to be the best predictor of the true primary interaction point.

vertex. For each decay product, the ratio of the transverse impact parameters with respect to the secondary and primary vertices are calculated. The product of these ratios is called RAT:

$$\text{RAT} = \prod_i \left(\frac{b_T^{sec}}{b_T^{pri}}\right)_i. \tag{5.5}$$

In grouping charged tracks into triplets, we don't want to confuse the correct decay vertex with one which is formed using only an incomplete subset of the tracks actually emanating from the vertex (e.g., selecting a $K\pi\pi$ triplet from a $K\pi\pi\pi$ decay). Therefore, an upper requirement is placed on the smallest $(b_T^{sec})_j$, where j runs over all tracks which are *not* members of the triplet. This minimum transverse impact parameter is called ISO:

$$\text{ISO} = \min((b_T^{sec})_{j,j\neq i}). \tag{5.6}$$

All of these cut variables have been defined without making any assumptions about the identities of either the charged particles associated into a triplet or their putative parent. In addition to these ID-neutral cut variables, we have at our disposal the standard measures of track and vertex quality, χ^2 per degree of freedom (χ^2/dof). These dimensionless quantities are obtained from the least-squares fitting algorithms used to construct tracks from SMD and DC hits or to construct vertices out of groups of said tracks. For each track determined using information from the SMD (the only class of tracks used in this analysis), separate χ^2/dof measures are available corresponding to whether DC and SMD hits or SMD hits alone are used; these are labelled χ^2/dof_{global} and $\chi^2/dof_{silicon}$, respectively. As described below, the cuts that are placed on these quantities in this analysis are too loose to have much impact.

5.3 Pair strip

Following reconstruction, a high-rejection event filter is implemented whose purpose is to produce a data set of manageable size with a fractional charm content greatly enhanced with respect to the generic high-E_T events written to tape. This filter is known as the "pair strip".

Every two-track combination among the set of silicon tracks which pass through the first magnet (and thus have measured momenta) is used to generate a secondary vertex. This vertex is required to have $\chi^2/dof < 5$ and $\sigma_{z\,sec} < 1.8$ mm. With the primary and secondary vertices established, two-track analogues of SDZ, PT2DK, and RAT are calculated. These variables are required to be greater than 6, greater than 0.1 GeV^2, and less than 0.06, respectively. Any event containing at least one pair passing all of these cuts passes the pair strip.

The pair strip, with a rejection factor of about 15, reduces the data set to 43 (partially-filled) tapes, corresponding to 29.9 million events.

5.4 Substrip

The pair strip data set is subjected to a further "substrip" filter. The substrip is comprised of three sets of cuts, each corresponding to a particular pseudoscalar D decay mode. Any event containing a triple which passes through one of these three gauntlets thereby passes the substrip.

In all that follows, triplets are formed from a set of tracks determined by the requirement that they have both silicon and drift chamber hits and pass through at least the first magnet. z_{pri} is constrained to be within the region containing the target foils and interaction scintillator ($-5.5 < z < 0.$ cm). In the target region, the z thickness of the foils ($\sim$100 μ for W, $\sim$250 μ for Be, Al, and Cu) is exploited to obtain a more precise (on average) determination of z_{pri}; the center of the target foil nearest the primary vertex as previously defined is taken to be z_{pri}. The error on z_{pri} is taken to be the RMS deviation of a flat distribution over the foil width, i.e., $\sigma_{RMS} = (z\,\text{width})/\sqrt{12}$. For cut variables that are calculated using z_{pri} (SDZ, DIP, and RAT), a "2" is added to the name of the variable if the modified z_{pri} (or its error) is used (e.g. SDZ2). Note that in the interaction scintillator region, no change is made to z_{pri}.

Decay mode dependent substrip cuts are listed in Table 5.1. In addition to these cuts, any kaons within a given triplet are required to have a Čerenkov probability greater than 13%, higher than the *a priori* value of 12%; this essentially eliminates identified pions as candidate kaons. The relative charge signs of the kaons and pions within the triplet must be consistent with the decay being sought. Very loose quality cuts are placed on the vertices: $\chi^2/dof_{pri} < 3$ and $\chi^2/dof_{sec} < 5$. For each track in the triplet, $\chi^2/dof_{silicon}$ and χ^2/dof_{global} are required to be less than 10.

Cut variable	$D^+ \to K\pi\pi$	$D^0 \to K\pi$	$D_s \to KK\pi$
SDZ2	7	5	7
PT2DK (GeV^2)	—	0.3	—
DIP2 (μ)	100	150	100
RAT2	0.1	0.2	0.1

Table 5.1: Decay mode dependent substrip cuts. See text for a description of generic cuts.

About 3% of data events pass the substrip, coming in under the three decay modes in roughly equal amounts. The purpose of the substrip is simply to facilitate further analysis by obtaining a working data set which can fit on two or three tapes. The cuts are loose,[5] and the set of decay mode cuts under which a particular event passes is not "remembered" in subsequent analysis.

5.5 Cut optimization

For each decay mode studied, a set of optimal cuts is sought which maximizes signal significance.[6] When, for a particular analysis variable, the cut point is not narrowly indicated by significance, both cut efficiency and straight signal-over-background are weighed in deciding where to set the cut. Two measures of significance are examined:

[5]Despite the supposed looseness of the cuts, subsequent cut optimization (see Section 5.5) indicated that some cuts be kept at their substrip values.

[6]Significance is just the ratio of the number of signal events over the uncertainty in this number, given by the square root of the *total* number of events in the signal range, both signal and background.

the first uses MC signal and data background (where the former has been normalized to the latter), the second uses data signal and background. While the first measure does not suffer from statistical fluctuations in the data signal, it loses reliability when the MC and data distributions for a particular variable differ significantly in that variable's natural cut range. In such cases, the second measure is given precedence as long as it is smoothly-varying (i.e., statistical fluctuations are avoided).

As an example, in Fig. 5.1, the following variations versus the DIP2 cut value are shown for the $D^+ \to K\pi\pi$ mode:

(1) Significance #1: Ratio of normalized MC signal to square root of normalized MC signal plus data background

(2) Significance #2: Ratio of data signal to error on data signal (as returned by the fitter)

(3) "Cross-section": Ratio of data signal to MC signal, normalized to average around 1.0

(4) Signal-over-background: Ratio of normalized MC signal to data background

(5) Data signal

It should be noted that the errors indicated at each cut point are correlated.

For differential cross-sections, absolute normalization is not a concern, therefore making the choice of cuts less crucial than it is for the absolute cross-sections. In the latter analysis, however, it is important that the data/MC signal ratio remain stable (within statistics) over the appropriate range (i.e., from no cut to the optimal cut); this is an indication that the MC is modelling the distribution of a given variable in real data reliably. If this is not the case, the acceptance (see Section 7.1) returned by the MC cannot be trusted. As described in the following section, the DIP2 cut is set at different values in the forward and differential cross-section analyses for the $D^+ \to K\pi\pi$ and $D^0 \to K\pi$ modes. In these two cases, the ratio of data to MC signal (proportional to the cross-section) is rather sensitive to the value of the DIP2 cut, requiring that this cut be loosened to a value where this ratio more or less plateaus. For all other cut variables, the data/MC signal ratio is reasonably flat.

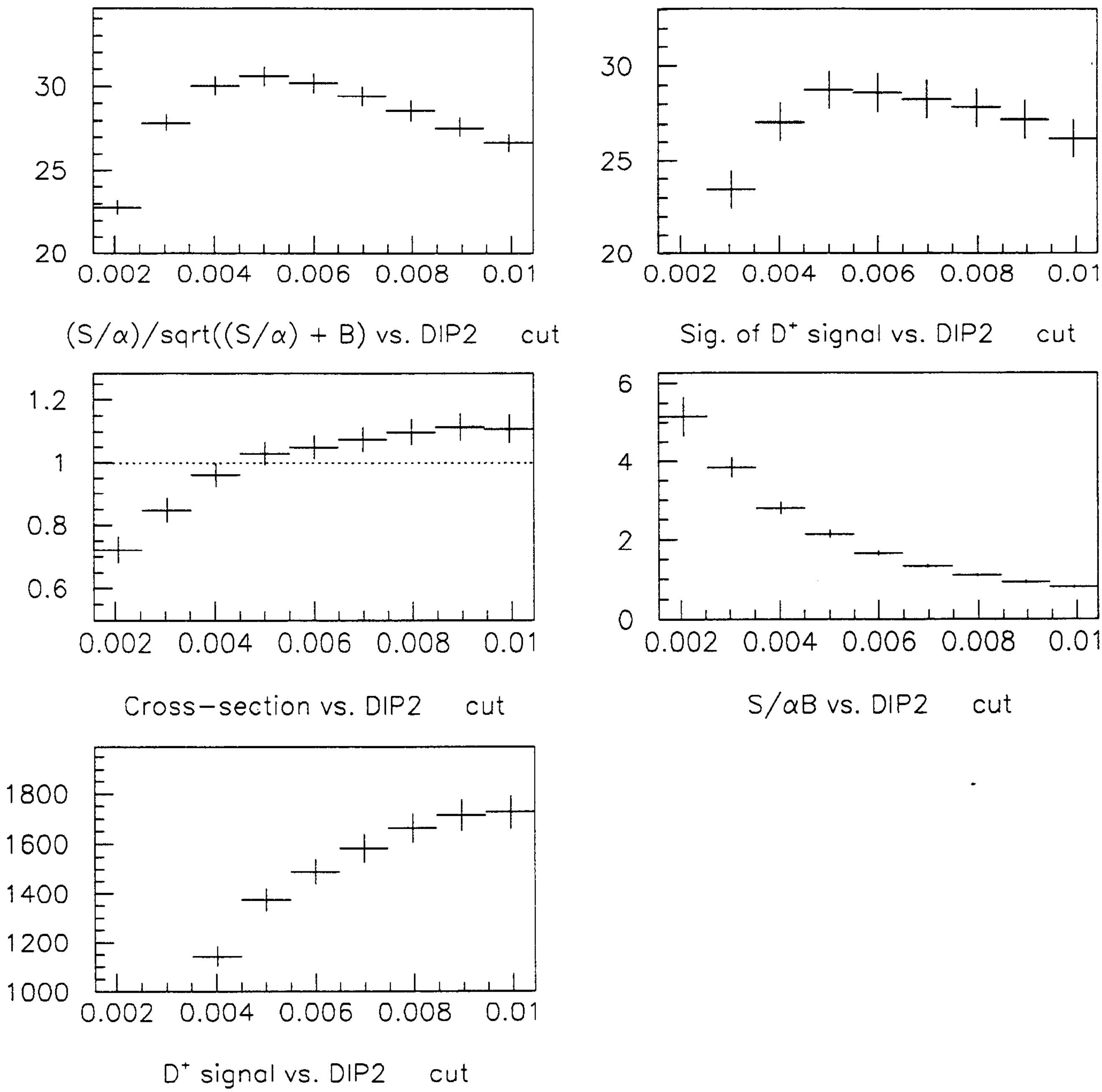

Figure 5.1: $D^+ \rightarrow K\pi\pi$ cut optimization: DIP2 (defined in Section 5.2. For all plots, DIP2 is measured in centimeters.

5.6 Final cuts

In selecting triplets as candidate decays, the same requirements specified in the sub-strip are placed on track and vertex quality, relative charge sign, Čerenkov probabilities for the kaon(s), and z location of the primary vertex. In Table 5.2, optimal cuts for the standard cut variables are listed for each decay mode. Recall that the $D^0 \to K\pi$ mode is used in the D^* analysis; these cuts are defined for the daughter D^0 in this case.

Cut variable	$D^+ \to K\pi\pi$	$D^0 \to K\pi$	$D^* \to D^0\pi$	$D_s \to \phi\pi$	$D_s \to K^*K$
SDZ2	13	10	9	10	13
PT2DK (GeV^2)	0.3	0.4	0.3	0.3	0.3
DIP2 (μ)	60[a]	60[a]	60	70	60
RAT2	.003	0.1	0.06	0.01	0.01
ISO (μ)	40	—	—	30	30

[a]Loosened to 100 μ in the forward cross-section analysis.

Table 5.2: Decay mode dependent final analysis cuts. See text for a description of cuts which are either generic or used specifically for a particular decay mode.

In addition to these cuts, certain cuts are applied to specific decay modes. For $D^+ \to K\pi\pi$, the product of SDZ2 and PT2DK is required to be greater than 15 GeV^2. For $D_s \to K^*K$, the absolute value of the cosine of the angle between the D_s and decay pion momenta (measured in the K^* center-of-mass frame) is required to be greater than 0.2.[7] For D_s decays to $\phi\,\pi$ (K^*K), the invariant mass[8] of the KK

[7]In the decay chain $D_s \to K^*K$, $K^* \to K\pi$, all particles are pseudoscalars (P), except for the K^*, which is a vector (V). In a P $\to$ VP decay, the V is longitudinally polarized. Therefore, if we boost from the rest frame of the parent P to that of the V, the products of a subsequent V $\to$ PP decay will have a $\cos^2\theta$ distribution, where θ is measured with respect to the boost direction.

[8]Invariant masses are calculated by assigning triplet members kaon or pion masses. Given the fact that only loose identification requirements are imposed, one triplet can be consistent with more than one scheme for assigning particle IDs; in these cases, such a triplet can account for more than one entry in an invariant mass plot. (All but the correct entry presumably contribute only to the background continuum.)

(relevant $K\pi$) pair is required to be within 10 (50) MeV of the ϕ (K^*) mass. In the D^* analysis, the invariant mass *difference* between the $K\pi$ pair and the triplet formed by adding the soft D^* decay pion is required to be within 10 MeV of the D^*-D^0 mass difference.

Once the set of all triplets passing the cuts for a given decay mode are found, the invariant masses of these triplets are histogrammed. Fits to these invariant mass distributions allow for the estimation of the number of decays observed in the E769 data set. The results of these fits are given in Chapter 6.