

CDF Production Farms Reconstruction

Summary: Fall, 2000 through early 2003

Jaroslav Antos Yen-Chu Chen Roman Lysak
Miroslav Siket Stephen Wolbers GP Yeh

v2.0, January 2, 2004

1 Introduction

This update includes all first-pass data processing of data collected up to the shutdown of September, 2003. It does not include the reprocessing of all data taken from February, 2002 through September, 2003. That will be included in a future update or a separate note.

The CDF offline production farms have been reconstructing data from CDF since the commissioning run in fall, 2000. Many versions of ProductionExe have been used during that time. In addition, some of the data has been reconstructed multiple times. Information about the streams and datasets can be found by interrogating the farms web pages (fnppcc.fnal.gov), or by looking in the CDF Data File Catalog (DFC). However, summary information is not so simple to obtain. This note will give a summary of the processing that has occurred so far on the farms. This information is useful in understanding what the farms have done, what they are capable of, and in trying to sort out the history of processing so far.

2 Description of Run Periods

The CDF experiment has collected data in Run 2 for quite some time. The first data which was reconstructed on the production farms was collected

during the commissioning run in the fall of 2000. Since that time data was collected and reconstructed in blocks of time separated by shutdowns or by significant changes to the detector, triggers or data organization. During the summer of 2001 the concept of “good runs” was created to identify runs that were meant for reconstruction on the production farms, as opposed to test runs, cosmic runs or runs which had significant detector problems. Prior to that time lists of runs were generated by hand to identify samples that were destined for reconstruction on the production farms.

The run periods are listed in Table 1.

Table 1: **CDF Run Periods 2000-2002**

I	Commissioning Run	Fall, 2000
II	1x8	March, 2001
III	36x36	May, 2001
IV	June	June-July, 2001
V	August	August-September, 2001
VI	December	December,2001-February, 2002
VII	8 stream data	February, 2002-September, 2003

The run period VII is the period that is most interesting. The triggers were well-established by that time and the data was split into the familiar 8 streams (A,B,C,D,E,G,H,J) and written to tape in that format. This is the period that was reprocessed with the newest executable (5.1.x) beginning in September, 2003.

For each run period data was reconstructed 0, 1, 2 or more times. Each processing pass is identified by a ProductionExe version number. A table identifying all of the run periods and each ProductionExe version and the number of events processed can be found as Table 2. One cannot determine from the table how many times each event was processed, or the first and last run number of the event sample. It should be noted that the definition of which data were to be processed by each version was not always well-defined. Only in June 2001 was the “good run” first defined and available for use by the offline production farms to clearly define these runs.

As can be seen by the table, many event samples have been processed more than one time. In some cases the event sample was divided into multiple subsets and each subset was reconstructed with a slightly different version of

Table 2: Executables and Event Counts

Run Period	Exe. Vers.	No. Events	Time/Event
I (Commissioning)	3.11.0g	9,775,297	1.2 s/event
I (Commissioning)	3.12.0	9,866,564	1.5 s/event
II (1x8)	3.14.0	2,539,217	0.5 s/event
III (36x36)	3.15.0c	5,113,927	1.0 s/event
III (36x36)	3.16.0	5,113,927	2.3 s/event
IVa (June, 2001)	3.17.1	12,322,465	3.5 s/event
IVb (June, 2001)	3.18.0	21,993,586	
V (August, 2001)	4.0.0i	9,882,486	
V (August, 2001)	4.1.0	38,857,497	
V (August, 2001)	4.2.0a	8,685,950	
VI (December, 2001)	4.2.0b	6,445,760	
VI (December, 2001)	4.2.0c	26,944,076	
VI (December, 2001)	4.3.1	14,519,033	
VII (8 stream data)	4.3.1b	10,369,500	
VII (8 stream data)	4.3.2	10,109,874	
VII (8 stream data)	4.3.2a	38,690,220	
VII (8 stream data)	4.5.2	125,161,272	
VII (8 stream data)	4.8.0	16,572,810	
VII (8 stream data)	4.8.4a	367,471,080	
VII (8 stream data)	4.8.4g1	152,573,285	
VII (8 stream data)	4.8.5	106,488,677	
TOTAL		999,496,503	

the CDF code.

The total number of events processed (this includes some of the reprocessing done) is almost 1 billion, not including the reprocessing of approximately 588 million events for the winter 2004 conferences.

3 Plots

It is interesting to see the time development of the reconstruction for all versions of the reconstruction. Figure 1 shows the number of events processed as a function of time through March 2003. (Maybe we should update this.) This figure includes each event every time it was processed. The total number of events in the figure is higher than the table because it includes some of the Monte Carlo, test processing, and other versions of code that are not included in table 2.

Figure 2 shows the raw data collected and processed since February, 2002. In this plot the processed data is shown only once – the first time an event is processed. Reprocessing is not included. This plot is up-to-date through early September, 2003.

The processing progress for selected reconstruction code versions can be seen by reconstruction version in figures 3 to 7. In addition, the CPU time/event, shown as a histogram where each entry is the average CPU time/event for a file, can be seen in figures 8 to 12. The CPU time was converted to 500 MHz equivalent seconds.

Figure 1: Total Event Reconstruction on the CDF Farms

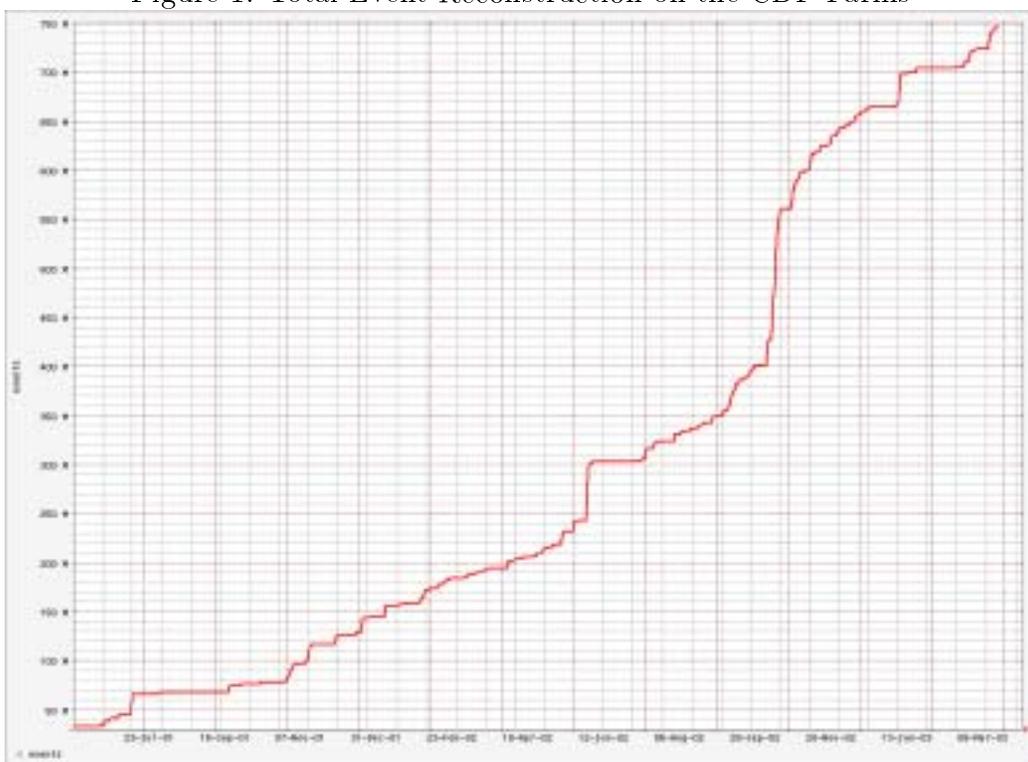


Figure 2: Event Processing since February, 2002 on CDF Farms

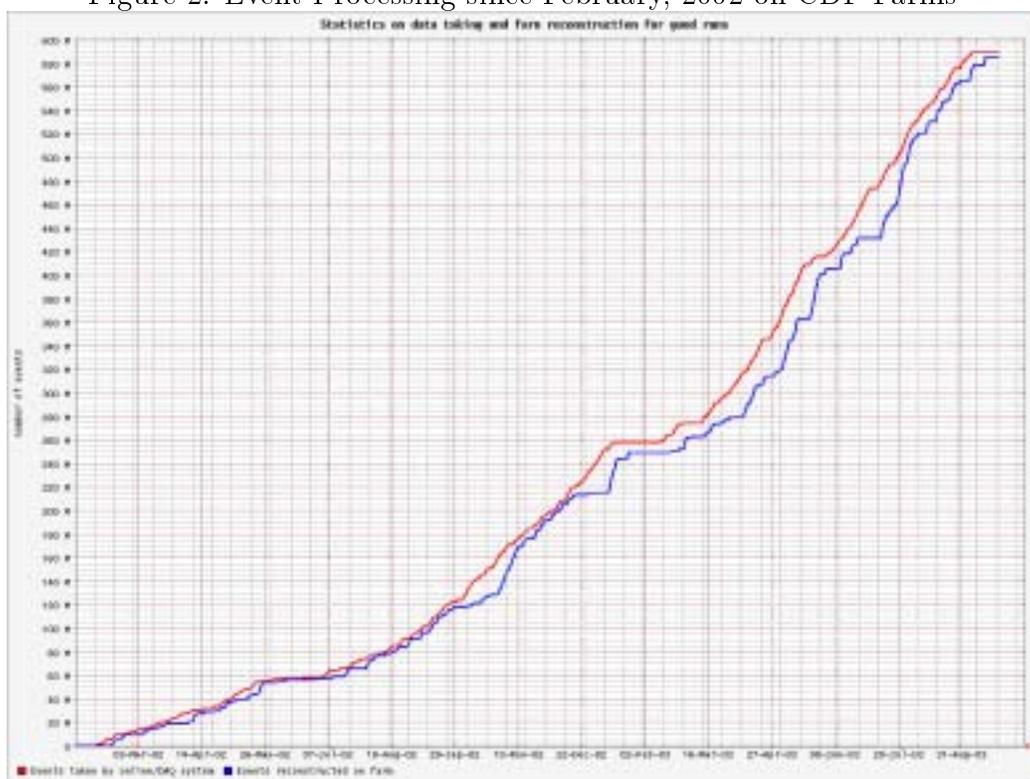


Figure 3: Event Reconstruction Progress



Figure 4: Event Reconstruction Progress

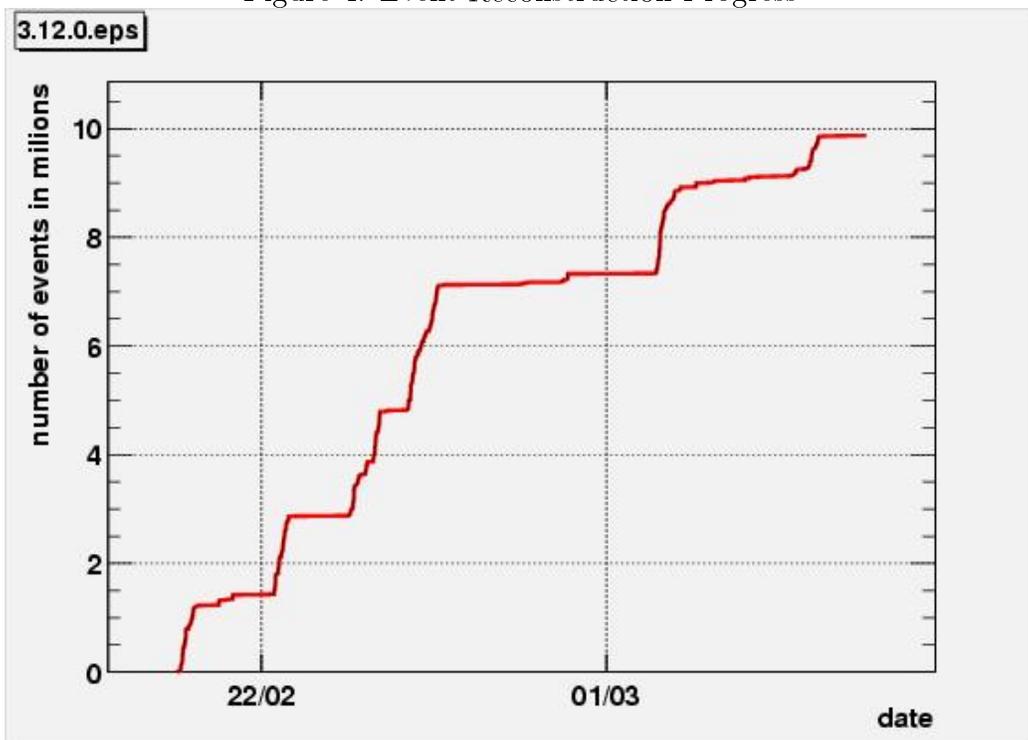


Figure 5: Event Reconstruction Progress

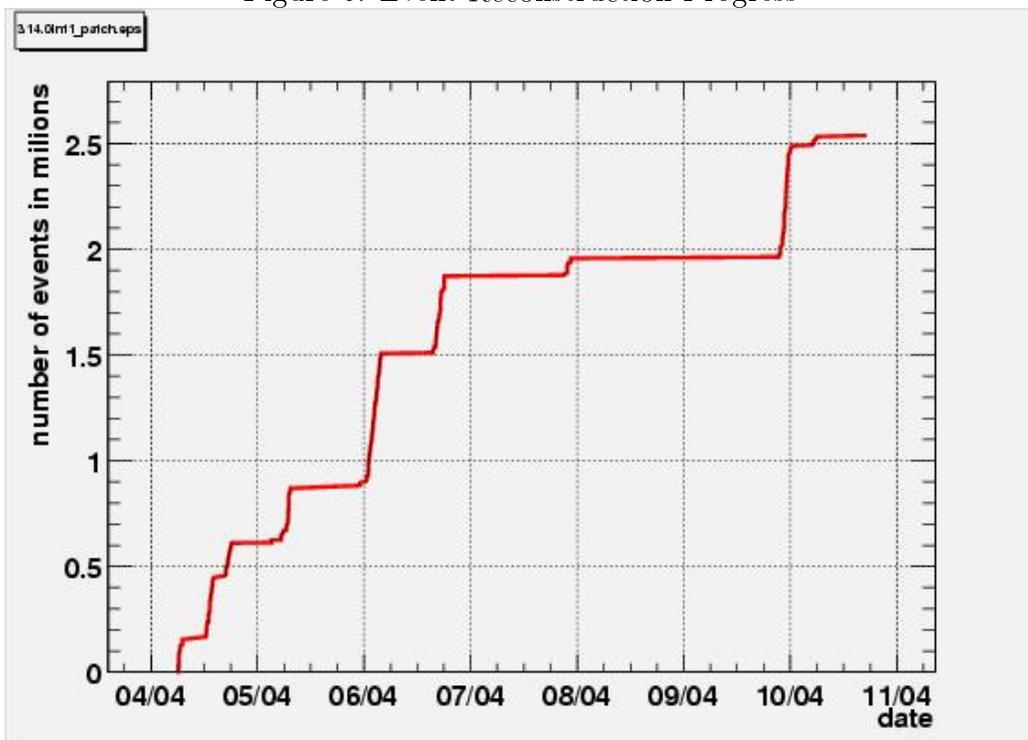


Figure 6: Event Reconstruction Progress



Figure 7: Event Reconstruction Progress

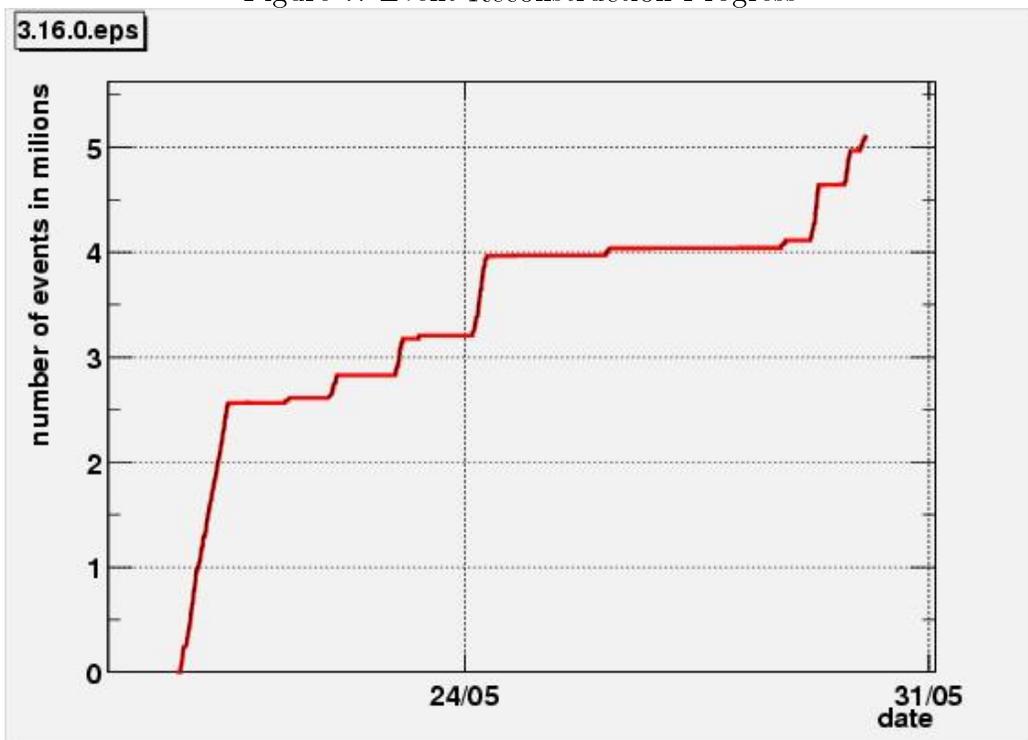


Figure 8: 3.11.0g CPU time/event (500 MHz seconds)

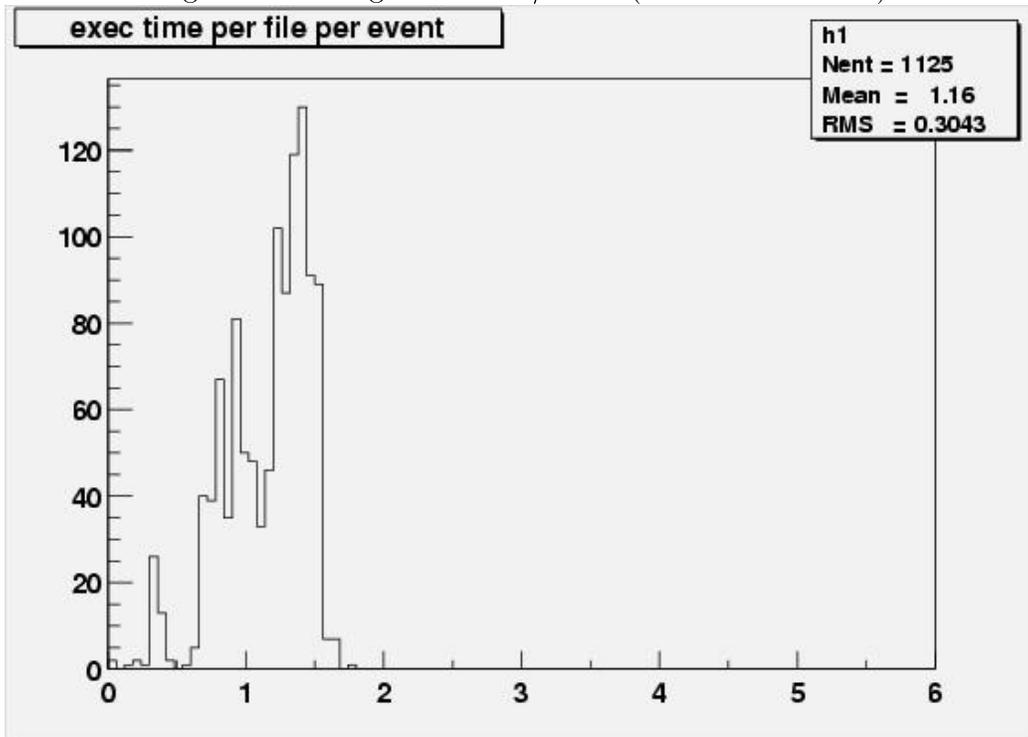


Figure 9: 3.12.0 CPU time/event (500 MHz seconds)

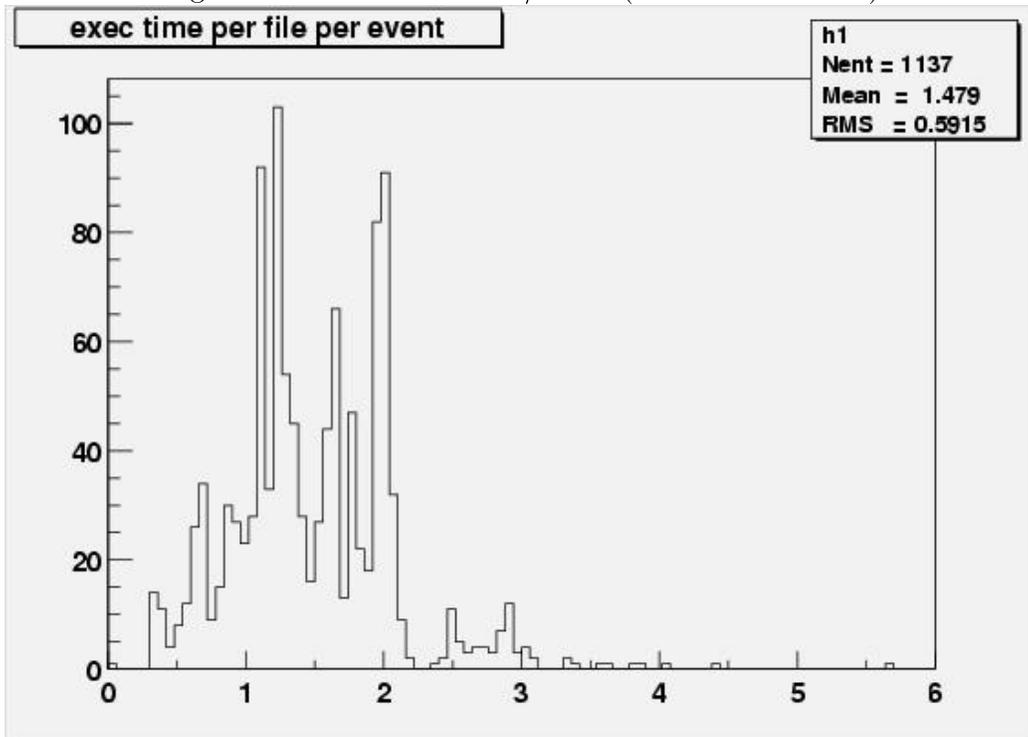


Figure 10: 3.14.0int1_patch CPU time/event (500 MHz seconds)

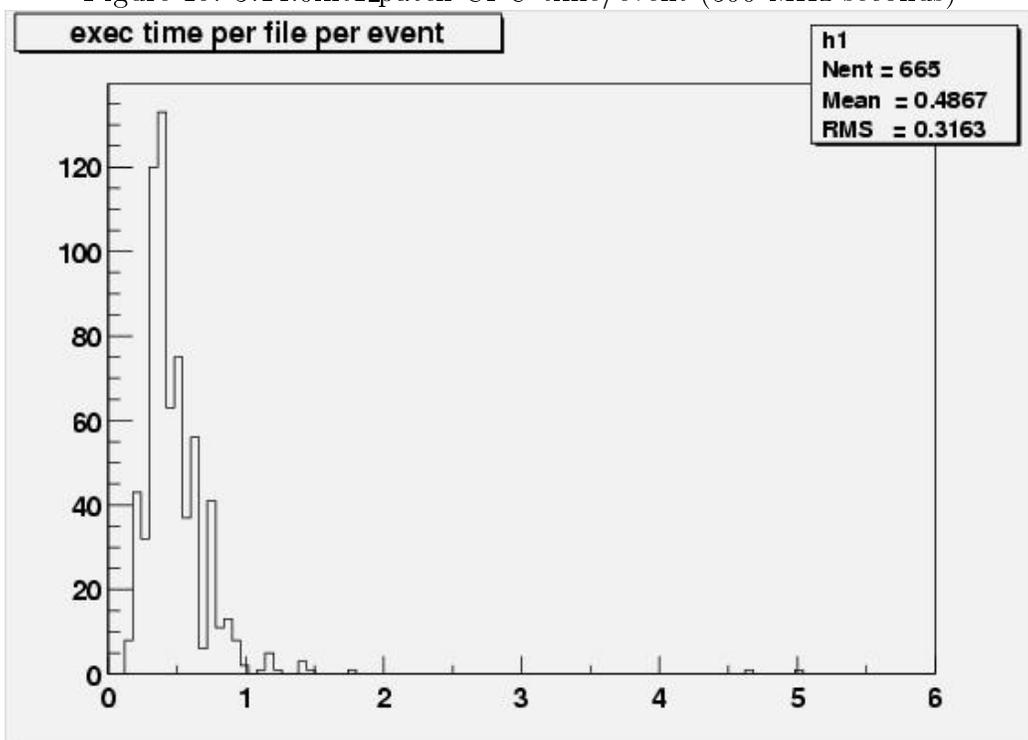


Figure 11: 3.15.0c_calib CPU time/event (500 MHz seconds)

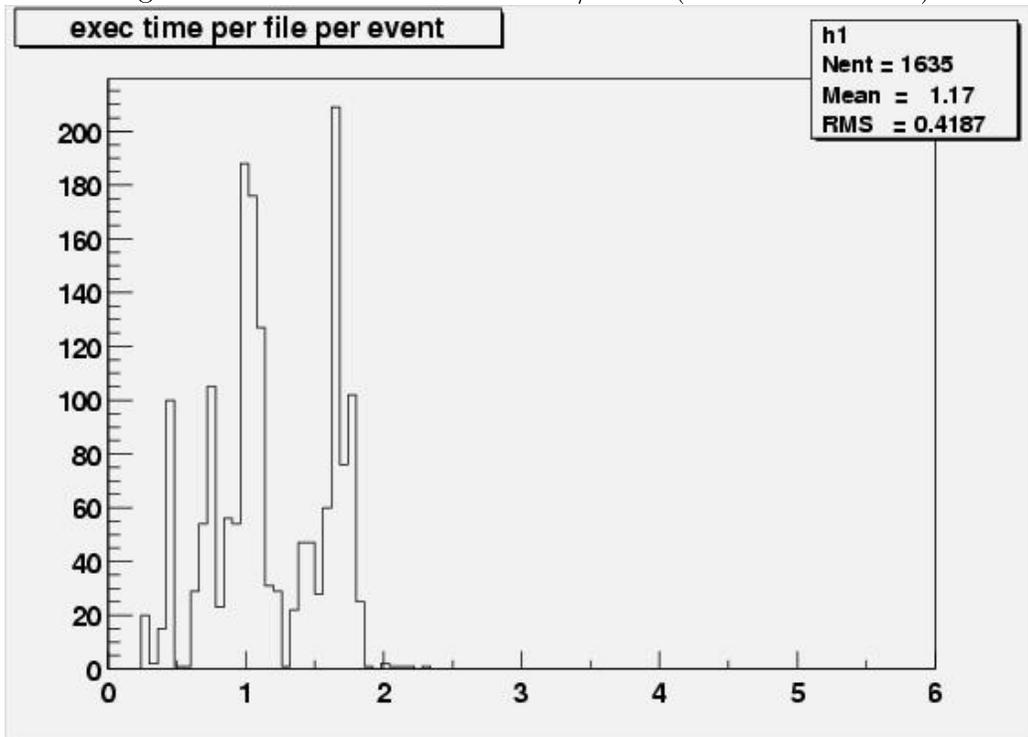


Figure 12: 3.16.0 CPU time/event (500 MHz seconds)

