The Calculation of Detector Bin Cluster Efficiencies

Cedric M. Guss
18 October, 1995

Abstract

The method for calculating cluster efficiencies in the E811 detectors is described here. In runs 745+746 of E811, we find severe variations in efficiency from 32 % to 100 % that, left uncorrected, can create extremely distorted measurements of background and elastic dN/dt data.

I. Method for Obtaining Cluster Efficiencies

Each detector was divided into 30 vdas cell by 30 vdas cell bins (about 1.35 mm by 1.35 mm). Without care for the type of events in the data sample, we selected all events that had a toe or heel counter in time with a cluster found in the corresponding sci-fi bundle (which cluster efficiency was to be found), along with the requirement that the heel adc be above threshold. A summed grey count per cluster distribution was obtained for each bin for all clusters which peak vdas cell grey count was found in that bin in the detector for all the events meeting the above conditions in runs 745+746 (low t, no main ring). Because most bins had few statistics (10s to 100s of events), we could not fit each bin's distribution to a straightforward function such as a Gaussian. Only the highest statistics bins (with a few thousands of events) could be fit, and a Gaussian function was used. For those bins which could not be fit, the only relevant information that could be extracted was the raw mean of each distribution.

For fitted distributions, one could easily find the percent of the distribution above vdas threshold (for the detector and bin in question) by integrating the Gaussian fit parameters, using an Error function for the entire range of the function above vdas threshold. This percent would thus be the cluster efficiency for that bin. However, the low statistics bins have only distorted distribution mean information available, and therefore something else needed to be done.

For low statistics bins, simple counting of distribution datum points in order to extrapolate a trend below vdas threshold was found to be very unreliable. However, if we could make some assumptions about what each bin's distribution would look like, we could still use an Error function to calculate expected distribution percentages above vdas threshold. The first assumption we made was to state the Ansatz that each bin's distribution behavior is the same and could be described by Gaussian behavior. Since we know high statistics bin distributions do behave this way, the physical causes of this distribution behavior remain unchanged from bin to bin to bin; only the dark current voltage pedestal with respect to hardware vdas threshold can change from bin to bin to bin. This changing dark current as a function of CCD chip position is blamed for the shifting mean pulse heights per distribution bin in the first place. Secondly, and more controversial, we made the assumption that the distribution widths DO NOT CHANGE from bin to bin to bin. Again, the physical reasons giving rise not only to Gaussian statistics but also to distribution sizes of each bin's pulse height distribution are still the same. Hence, we assume that the fit width of one high statistics bin distribution can be used for ALL bins within the same detector. Figures 1 show a sample high statistics bin distribution for each of the 4 detectors and the overlaying Gaussian fit. It can be seen that in each case, the fit describes the available data (above vdas threshold) very well. The 0 point can be interpreted as the vdas threshold of each distribution.

Now, for the Error function technique to work, we only need to concern ourselves with distribution shape and mean with respect to some threshold; so all that is left to input into each bin's Error function is the bin distribution mean. We might be tempted to use the average of the available data within the bin, but this is distorted because the true mean is the average of all the data, above and below vdas threshold. We could calculate the distortion in a straightforward way if we knew apriori HOW MUCH of the distribution was missing, but we don't. (If we did, we would already know the cluster efficiency.) Therefore, we have to reach back to our knowledge of the statistics of Gaussian uncertainties, and realize that, on average, the mean of the MEASURED Gaussian distribution is too high by $1/2\ \sigma$ from the real mean of the seen portion of distributions where there is a random threshold cutoff and where further information is unavailable due to poor statistics. Finally, the distortion can be calculated:

$$\mu_{true} = \{\mu_{meas} \sum_{x=thres}^{\infty} [G_x(\mu_{meas}, \sigma)] + \sum_{x=-\infty}^{thres} [x G_x(\mu_{meas}, \sigma)]\}/\{\sum_{x=-\infty}^{\infty} [G_x(\mu_{meas}, \sigma)]\},$$

where G_x is the value of the Gaussian function at summed grey count point x, μ_{true} is the true mean of the distribution, μ_{meas} is the distorted MEASURED mean calculated from the seen data above vdas threshold (where 1/2 σ is subtracted off), and thres represents the vdas threshold x point.

A point needs to be made here about the vdas threshold x points that were used. It was different for each detector, and the summed grey count point 0 was not used. The reason for this has to do with the cluster definition itself; in most cases, 3 out of 9 vdas cells in a block had to be above vdas threshold. But this creates a situation where the software definition has its own cluster recognition efficiency, and it changes drastically as a function of summed grey count in the region above 0; the pattern is also different for each detector. Hence, we determined a "safe" summed count threshold for each detector where we were confident that the software recognition efficiency was 100 % at and above this level. For detectors 1-4, it was determined to be 30, 40, 100, and 20 summed counts, respectively. These were the numbers used for the x "thres" point in the above equation for each detector's bin efficiency map; a cluster with summed count below this number is thrown out in the analysis, but accounted for in the above calculation for its detector bin cluster efficiency.

Now, by inputting the starting and end summation points of the above equation into the mathematical Error function, using the distribution width from the fit of data from a high statistics bin from the same detector and the distorted distribution mean from the bin in question (with 1/2 σ subtracted) as its parameters, one can calculate the above equation in order to find the true mean of the bin's distribution. With knowledge of each bin's true distribution mean and the assumed distribution width, σ, one can input these parameters into the Error function in order to integrate the Gaussian distribution over the desired starting and end points, and thus find each bin's summed grey count distribution percentage above vdas threshold, defined as that bin's cluster efficiency:

$$\varepsilon_{clus} = \{\int_{x=thres}^{x=\infty} G_x(\mu_{true},\sigma)dx\}/\{\int_{x=-\infty}^{x=\infty} G_x(\mu_{true},\sigma)dx\},$$

where x, G_x, thres, σ, and μ_{true} are defined as before.

II. Results

Below are listed four tables showing the results from our study of no main ring runs 745-746. The first lists the number of events available that met the above criteria for analysis of each distribution bin of each detector. The 7 lines per detector represent rows of y, the first row being closest the beam, etc. The 10 columns represent sections along x where the centermost columns are closest the beam. The second table lists the corrected means of the summed cluster grey count distributions for each bin. The third and fourth tables are the resultant cluster efficiencies (using the formula above) and cluster efficiency errors, respectively, for each bin.

Detector 1

61	198	326	1732	436	18	1	1	1	1
214	807	2395	7027	3469	722	280	12	3	1
141	365	866	1862	1153	458	205	14	2	1
96	229	441	806	571	315	165	6	2	1
13	96	251	380	308	136	31	7	1	1
3	41	124	224	150	88	24	1	1	1
1	1	5	27	25	12	1	1	1	1
1	1	1	3	1	1	1	1	1	1

Detector 2

1	1	1	1	1	1	1	1	1	1
1	6	472	4691	1406	300	58	6	1	1
84	441	2606	11049	6973	1578	724	179	12	1
78	251	729	2390	2206	820	453	117	4	1
65	168	418	951	1128	534	263	78	5	1
11	60	249	475	563	306	71	6	1	1
1	26	107	228	235	133	27	1	1	1

Detector 3

1	37	678	2957	1336	569	179	4	1	1
97	404	1828	5825	2513	938	304	14	4	1
89	337	863	1830	1305	574	191	7	1	1
48	212	518	841	662	291	59	6	3	1
12	77	290	458	382	122	27	6	1	1
2	32	101	147	92	59	11	1	1	1
1	3	15	15	16	10	1	1	1	1

Detector 4

113	458	1699	4253	1580	411	75	3	1	1
157	586	1827	3719	1970	585	111	9	1	1
114	365	908	1463	995	410	78	4	5	1
52	168	430	705	538	197	41	5	2	1
12	114	249	318	287	94	10	5	1	1
6	19	49	55	54	15	1	6	3	1

Table 1. Events per bin available for obtaining cluster efficiencies in all bins of all detectors in runs 745+746.

Detector 1

522.3	385.3	162.7	210.4	146.4	-44.8	361.9	361.9	361.9	361.9
762.3	704.2	529.0	627.2	670.2	623.4	618.0	690.2	361.9	361.9
705.6	768.1	635.3	672.8	722.0	726.8	708.6	531.4	361.9	361.9
719.7	761.6	733.2	676.1	690.2	712.9	706.4	931.6	361.9	361.9
188.6	611.8	720.5	683.9	688.1	701.4	687.4	120.0	361.9	361.9
-68.7	351.7	697.6	608.9	632.1	670.2	650.2	361.9	361.9	361.9
361.9	-112.2	81.1	540.4	585.1	246.1	897.1	361.9	361.9	361.9

Detector 2

499.8	499.8	499.8	499.8	499.8	499.8	499.8	499.8	499.8	499.8
499.8	58.5	158.5	148.5	254.6	284.4	178.7	367.1	499.8	499.8
345.7	386.8	395.3	385.6	417.8	428.4	383.4	377.8	405.5	499.8
336.3	376.3	405.6	430.7	419.1	407.6	360.1	348.4	114.6	499.8
287.7	336.8	385.7	416.0	406.8	390.5	337.1	285.3	647.5	499.8
224.5	317.7	366.6	385.8	388.5	358.2	331.6	320.1	499.8	499.8
499.8	221.1	330.0	350.2	333.0	282.2	274.8	499.8	499.8	499.8

Detector 3

-18.3	-4.4	-52.7	-28.2	14.0	65.8	123.4	194.0	361.5	361.5
83.7	78.0	64.1	113.5	196.0	291.6	291.6	511.4	166.8	361.5

315.7	280.6	324.0	345.1	410.5	419.2	449.9	100.8	443.3	361.5
174.3	353.8	412.4	480.1	501.0	444.5	171.2	2.8	114.6	361.5
-72.8	69.4	291.3	398.2	475.1	295.5	40.8	-67.6	-166.4	361.5
-137.6	-114.6	19.5	79.8	86.2	45.6	-105.8	-151.3	361.5	361.5
361.5	-144.5	-135.9	-137.4	-141.2	-135.2	-157.2	361.5	361.5	361.5

Detector 4

479.4	479.4	479.4	-69.0	479.4	479.4	479.4	479.4	479.4	479.4
131.6	92.0	41.8	14.2	21.4	15.8	45.6	-85.6	479.4	479.4
339.2	225.1	160.3	117.4	101.1	89.9	112.8	5.5	479.4	479.4
370.2	358.7	308.4	235.7	169.2	178.7	169.3	246.9	479.4	479.4
463.8	393.5	338.2	276.8	267.8	251.2	188.1	-77.4	479.4	479.4
261.2	372.3	292.8	250.0	248.6	290.1	204.6	479.4	479.4	479.4
-30.8	221.4	319.6	263.2	337.6	209.8	-82.7	479.4	479.4	479.4

Table 2. Corrected bin distribution means for all detectors in runs 745+746.

Detector 1

0.914	0.838	0.644	0.692	0.627	0.418	—	—	—	—
0.979	0.969	0.917	0.951	0.962	0.950	0.949	0.967	—	—
0.970	0.980	0.954	0.963	0.973	0.973	0.970	0.918	—	—
0.972	0.979	0.975	0.964	0.967	0.971	0.970	0.994	—	—
0.670	0.947	0.972	0.965	0.966	0.969	0.966	0.599	—	—
0.392	0.814	0.968	0.946	0.953	0.962	0.957	—	—	—
—	0.347	0.556	0.922	0.938	0.726	0.992	—	—	—

Detector 2

—	—	—	—	—	—	—	—	—	—
—	0.561	0.836	0.815	0.962	0.978	0.874	0.997	—	—
0.994	0.998	0.998	0.998	0.999	0.999	0.998	0.997	0.999	—
0.993	0.997	0.999	0.999	0.999	0.999	0.996	0.995	0.731	—
0.980	0.993	0.998	0.999	0.999	0.998	0.993	0.979	—	—
0.936	0.989	0.997	0.998	0.998	0.996	0.992	0.990	—	—

—	0.933	0.992	0.995	0.992	0.977	0.974	—	—	—

Detector 3

0.360	0.376	0.322	0.349	0.398	0.459	0.528	0.612	—	—
0.480	0.474	0.457	0.516	0.614	0.718	0.718	0.893	0.580	—
0.742	0.707	0.751	0.770	0.826	0.832	0.855	0.501	0.850	—
0.589	0.778	0.827	0.874	0.887	0.851	0.585	0.385	0.518	—
0.301	0.463	0.718	0.816	0.871	0.722	0.429	0.306	0.211	—
0.237	0.259	0.404	0.476	0.483	0.435	0.267	0.224	—	—
—	0.230	0.238	0.237	0.233	0.239	0.219	—	—	—

Detector 4

—	—	—	0.336	—	—	—	—	—	—
0.702	0.634	0.541	0.489	0.503	0.492	0.549	0.308	—	—
0.936	0.836	0.748	0.679	0.650	0.630	0.671	0.472	—	—
0.952	0.947	0.915	0.848	0.761	0.775	0.761	0.860	—	—
0.983	0.962	0.935	0.889	0.881	0.864	0.788	0.321	—	—
0.874	0.953	0.903	0.863	0.862	0.901	0.810	—	—	—
0.405	0.831	0.923	0.877	0.935	0.817	0.313	—	—	—

Table 3. Cluster efficiencies for all bins in all detectors in runs 745+746.

Detector 1

0.164	0.162	0.162	0.160	0.162	0.198	0.415	0.415	0.415	0.415
0.160	0.160	0.160	0.160	0.160	0.160	0.161	0.168	0.273	0.415
0.161	0.160	0.160	0.160	0.160	0.160	0.160	0.176	0.314	0.415
0.161	0.160	0.160	0.160	0.160	0.160	0.161	0.163	0.314	0.415
0.206	0.162	0.160	0.160	0.160	0.161	0.163	0.245	0.415	0.415
0.324	0.171	0.161	0.161	0.161	0.161	0.165	0.415	0.415	0.415
0.415	0.502	0.274	0.168	0.167	0.205	0.183	0.415	0.415	0.415

Detector 2

0.160	0.160	0.160	0.160	0.160	0.160	0.160	0.160	0.160	0.160

0.160	0.258	0.161	0.160	0.160	0.160	0.166	0.162	0.160	0.160
0.160	0.160	0.160	0.160	0.160	0.160	0.160	0.160	0.160	0.160
0.160	0.160	0.160	0.160	0.160	0.160	0.160	0.160	0.273	0.160
0.161	0.160	0.160	0.160	0.160	0.160	0.160	0.161	0.160	0.160
0.176	0.161	0.160	0.160	0.160	0.160	0.160	0.165	0.160	0.160
0.160	0.167	0.160	0.160	0.160	0.161	0.163	0.160	0.160	0.160

Detector 3

0.506	0.179	0.161	0.160	0.161	0.161	0.164	0.292	0.441	0.441
0.168	0.162	0.160	0.160	0.160	0.161	0.162	0.180	0.294	0.441
0.167	0.162	0.161	0.160	0.160	0.161	0.162	0.248	0.391	0.441
0.175	0.163	0.161	0.160	0.160	0.161	0.172	0.255	0.330	0.441
0.208	0.170	0.162	0.161	0.161	0.165	0.186	0.247	0.438	0.441
0.340	0.178	0.167	0.165	0.168	0.173	0.208	0.447	0.441	0.441
0.441	0.291	0.194	0.194	0.192	0.209	0.443	0.441	0.441	0.441

Detector 4

0.199	0.199	0.199	0.316	0.199	0.199	0.199	0.199	0.199	0.199
0.166	0.162	0.160	0.160	0.160	0.162	0.170	0.311	0.199	0.199
0.161	0.161	0.160	0.160	0.160	0.161	0.166	0.231	0.199	0.199
0.161	0.160	0.160	0.160	0.161	0.161	0.167	0.236	0.169	0.199
0.161	0.161	0.160	0.160	0.161	0.162	0.172	0.263	0.181	0.199
0.186	0.161	0.161	0.161	0.161	0.163	0.202	0.169	0.199	0.199
0.256	0.182	0.164	0.166	0.163	0.189	0.490	0.167	0.174	0.199

Table 4. Runs 745+746 cluster bin efficiency errors. (A factor of 16 added in quadrature due to the $1/2\ \sigma$ correction in the mean.)

From Table 1, one can see the poor statistics that was generally available for each bin. Table 2 gives an indication of the extreme changes in pulse height from bin to bin (relative to 0, representing the vdas threshold point), and which allows for the extreme changes in efficiency found in Table 3, where efficiencies vary anywhere from 32.2 % to 99.9 %. And Table 4 shows the error in each efficiency, most of which is due to the $1/2\ \sigma$ correction in the mean, and not so much due to statistics; the reason for this lies in the fact that even though the statistics were poor, we assume we know the shape of the distribution, and that it can be described perfectly with a Gaussian of known distribution width. The blanks represent cells with no available data for doing an efficiency calculation, and the negative means in Table 3 are either due to very weak pulse heights in those cell regions, or due to very poor statistics in determining the measured means that are then set below vdas threshold (0) when $1/2\ \sigma$ is subtracted off.

The final point to be made is that it is crucial for the experiment to find these cluster efficiencies with great accuracy (i.e. more statistics) and with greater bin resolution (i.e. again more statistics) in order to make proper corrections to the missing signal due to detector inefficiency, and to do proper background extrapolation from the normalization region to the signal region; otherwise ρ can never be obtained. We used about 200,000 events for this study; we will need at least another order of magnitude more statistics for obtaining cluster efficiencies in a proper analysis of the experiment. The good news is that any halo or elastic trigger can be used for this efficiency study, but preferrably not beam gas since it tends to produce big and ugly clusters that distort the summed cluster grey count distributions of true halo and elastic clusters.