Neutral meson and photon ntuples explanation, and examples

Gabor David

Stony Brook University, Brookhaven National Laboratory

π^0-s and photons (γ-s)

Based on real data this write-up introduces you how to find π^0-s and photons in an electromagnetic calorimeter (EMcal). They contain information on *clusters* – small, contiguous regions of energy deposit in the EMCal – which are usually considered as *energy deposited by one single particle*. While this is not always true, this is our first working hypothesis, but we should always be aware that a cluster might contain energy from more than one particle. Also, not all clusters correspond to photons; many are from hadrons, especially at low and very high p_T (transverse momentum)[1]. In fact, it's going to be your job to find selection criteria ("cuts") which eliminate hadrons (the *contamination*) but still preserves most of the photons (*signal*) in the sample.

As for π^0, recall that most of them decay via $\pi^0 \to \gamma\gamma$, so they can be reconstructed from photon candidate *pairs* using the invariant mass

$$m_{\gamma\gamma} = \sqrt{2E_1E_2(1 - cos(\theta))}$$

where E_1, E_2 are the energies of the two photon candidates and θ is the *opening angle* between them. Note that the p_T of the π^0 is simply the vector sum of the p_T-s of the two photons paired.

Please note that – as per PHENIX standards – time is always given in nanoseconds, distance in centimeters and energy in GeV.

The *gnt* and *ggntuple* Ntuples

The two files *MBntup.root* and *ERTntup.root* are produced using a small fraction of the $\sqrt{s_{NN}}$ =200 GeV p+Au data collected by PHENIX in 2015. They are published for educational

[1] This may sound surprising in an electromagnetic calorimeter, but it is true: many high p_T single clusters are actually from two overlapping photons from a π^0 decay, i.e. they are hadrons.

purposes only. You will not be able to derive any publishable physics results from them, however, you can learn some basic techniques of particle identification (PID), practice how to make different cuts and estimate their effect, how the signal to background ratio can be improved at the expense of some loss in the signal and how to find an optimum between the two trends. You may also learn that PID itself is also not an unambiguous procedure: you can apply much looser cuts on the "photonness" of the clusters when reconstructing π^0, because you identify the π^0 by the *correlation* of the two "photons", namely, whether their invariant mass is in the expected range (around 0.135 GeV) or not. Only photon pairs from the decay of the same π^0 are truly correlated, random pairs of photons, photon-hadron or hadron-hadron pairs only rarely give invariant masses in the "π^0 window". Therefore, you can allow yourself to have more contamination is the "photon" sample – and gain significant statistics in reconstructed π^0 (why?). – On the other hand for reconstructing (inclusive) photons you don't have such a help: if you decide that a cluster is a photon, that's it. In order to keep the contamination from misclassified hadrons you might want to make your photon PID cuts stricter, in order to increase *purity* (in other terminology decrease *contamination*) at the expense of *efficiency*.

The first Ntuple is called *ggntuple* (gamma-gamma ntuple) and contains information on EMCal cluster *pairs* and the two clusters in the pair (see Table I). In each event all clusters in a sector are paired with all other clusters, and the pair variables (*pt, costheta, phi, mass, asym*) calculated, even if they are obviously not photons from a π^0 decay (for instance because their invariant mass is, say, 0.4 GeV). These "obviously wrong" random pairs are called the *combinatorial background* and they help you to estimate the background in the π^0 peak when you'll try to extract the number of π^0 in the sample. The momenta attributed to a cluster (particle candidate) is calculated from the vector connecting the collision point (*vtxZ*) with the impact point on the EMCal, and under the assumption that the particle is a photon (i.e. $p = E$, the measured energy). From these, the pair momentum *pt* is calculated as the vector sum of the transverse components; *costheta* is the cosine of the opening angle between the two momentum vectors, *phi* is the azimuth of the pair momentum (i.e. of the parent π^0), *mass* is the invariant mass calculated with the formula above, and *asym* is the energy asymmetry α of the two clusters

$$\alpha = \frac{|E_1 - E_2|}{E_1 + E_2}$$

The rest of the variables describe the two clusters (*1,2*), their meaning is spelled out below.

Variable name	Description
cent	Event centrality
vtxZ	z-vertex of the event
pt	Transverse momentum of cluster *pair* (π^0 candidate)
costheta	*cos* of the opening angle between the two clusters
phi	Azimuthal angle of the direction of the pair momentum (assumed π^0)
mass	Invariant mass calculated from the two clusters (energy and position)
asym	Energy asymmetry ($\lvert E_1 - E_2\rvert/(E_1 + E_2)$) of the two clusters
sec1	EMCal sector where the first cluster is
Ecore1	"Core" energy of the first cluster
tof1	Time-of-flight of the first cluster
twrhit1	Number of towers in the first cluster
prob1	Probability that the first cluster is a photon (based on χ^2)
chisq1	χ^2 from expected photon shape for the first cluster
stoch1	Combined variable to describe "photonness" of first cluster
sec2	... Same quantities as above for the second cluster of the pair

TABLE I. Fields in the γ-pair Ntuple (*ggntuple*).

The second Ntuple is called *gnt* and contains information on individual clusters (again, many of which are *not* photons!). Here *pt* is the transverse momentum of the (single) particle, calculated under the assumption that it is a photon (straight path from the collision vertex and $p = E$), *costheta* is the cosine of the polar angle, *phi* is the azimuth, and *sec* is the sector number (0-5 are lead scintillator sectors, PbSc, 6-7 are lead glass, PbGl). For both detectors the smallest, individually read out units are called *towers*.

The field *ecore* is an estimate of the particle energy (using various corrections to account for detector effects), *ecent* is the energy in the central (highest energy) tower of the cluster, *tof* is the time-of-flight measured in the central tower and corrected with the flight-path s/c such that for photons *tof* is a distribution around zero. Based on testbeam data we have a model (parametrization) of electromagnetic showers depending on energy, impact point and angle of the photon; the *chisq* is calculated as the deviation of the actual shower (energy deposit pattern in the towers) from the "model" shower of the same energy, impact point and angle. The *prob* is calculated from this χ^2, usually showers with *prob* > 0.05 are very likely indeed photons; *disp* is the 2D dispersion, *twrhit* is the number of towers hit (i.e. included in the cluster), finally *stoch* is a combination of various probability measures to characterize "photonness", higher values mean higher probability that the cluster is indeed a photon. The calculated impact point of the particle is given as x, y, z in the PHENIX absolute coordinate system ($x > 0$ is the West Arm, negative z is South).

The *MBntup.root* file is produced from minimum bias data (although with a lower limit $p_{\mathrm{T}} > 2.0\,\mathrm{GeV}/c$ in *gnt* and pair p_{T} in *ggntuple*), whereas in *ERTntup.root* the threshold for single cluster p_{T} in *gnt* is 8 GeV, and the threshold for pair p_{T} in *ggntuple* is also 8 GeV. Note that since here we restrict only the pair p_{T}, the energy of the individual clusters can be (and often is) significantly lower. Important: both the MB and the ERT *ggntuple*-s are written out with the condition that the asymmetry is less than 0.8.

Variable name	Description
cent	Event centrality
vtxZ	z-vertex of the event
pt	Transverse momentum of the cluster (γ-candidate)
costheta	Polar angle of the cluster (γ-candidate)
phi	Azimuthal angle of the cluster (γ-candidate)
sec	EMCal sector of the cluster (γ-candidate)
ecore	"Core" energy of the cluster (γ-candidate)
ecent	Energy in the central tower of the cluster (γ-candidate)
tof	Time-of-flight in the central tower of the cluster (γ-candidate)
prob	Probability that the cluster is a photon (based on χ^2)
disp	Dispersion of the cluster (γ-candidate)
chisq	χ^2 from expected photon shape of the cluster (γ-candidate)
twrhit	Number of towers in the cluster (γ-candidate)
stoch	Combined variable to describe "photonness" of the cluster (γ-candidate)
x	x-position of impact point on the EMCal surface
y	y-position of impact point on the EMCal surface
z	z-position of impact point on the EMCal surface

TABLE II. Fields in the γ Ntuple (*gnt*).

Examples

Here are a few simple example plots illustrating what you can do, based on the MB and ERT (high p_T) data. (The amount of data, i.e. the statistics might be different in the released datafiles, but you should be able to reproduce the overall features of the histograms.)

```
f1 = new TFile("MBntup.root");
f2 = new TFile("ERTntup.root");
f1->cd();
ggntuple->Draw("mass","mass<1.0");
f2->cd();
ggntuple->Draw("mass","mass<1.0");
f1->cd();
ggntuple->Draw("mass>>htemp1","mass<1.0");
ggntuple->Draw("mass>>htemp2","mass<0.4&&chisq1<3.0&&chisq2<3.0");
htemp2->SetLineColor(2);
htemp1->Draw();
htemp2->Draw("same");
```

For characteristic results see Fig. 1. The left panel shows the $m_{\gamma\gamma}$ distribution in the MB data (no cut on p_T or any other observable). While the true π^0 peak dominates (0.135 GeV), there is a pronounced peak at 0.03 GeV (primarily due to "minimum ionization" hadrons), also, the combinatorial background under the true π^0 peak is substantial. The middle panel is the same $m_{\gamma\gamma}$ distribution from the ERT data. The fake peak at 0.03 GeV/c is gone, the combinatorial background is much smaller, and now you clearly see the peak at the η meson mass as well. The right panel is once again MB data; the blue histogram is the (uncut) $m_{\gamma\gamma}$ (same as on the left panel), the red is the same $m_{\gamma\gamma}$ if we require that both clusters of the pair satisfy the $\chi^2 < 3$ condition (i.e. are consistent with the photon cluster shape). Note the log scale. As you can see, the cut on "photonness" leaves the π^0 peak area nearly unchanged, but reduces the combinatorial background drastically.

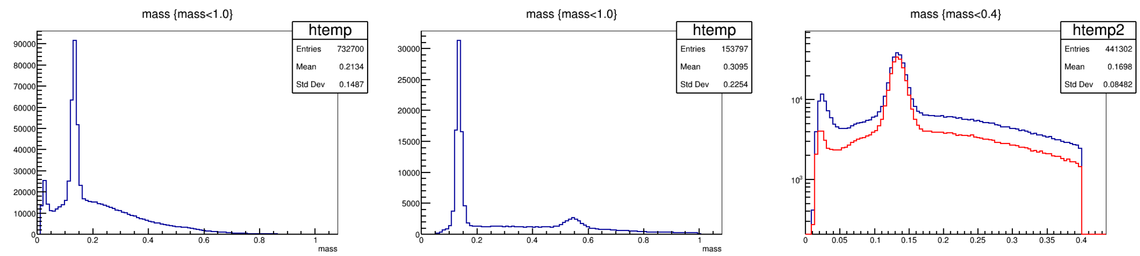

FIG. 1. Left panel: Invariant mass from the MB data in the 0-1 GeV region. You can see a strong π^0 peak and maybe a hint of η. Middle panel: π^0 peak for pairs from the ERT data (i.e. the pair p_T is greater than 8 GeV/c. The combinatorial background is much smaller and the η peak is now clearly seen. Right panel: the effect of a strong χ^2 cut on the photon pairs in the MB data. The blue histogram is the same as in the left panel, the red is with the χ^2 cut on both clusters. While you barely lose any signal (due to the small inefficiency of the cut), the signal to background ratio improves dramatically.

The (energy) asymmetry of the two π^0 decay photons is a flat distribution by kinematics (at least in 4π solid angle). This offers another opportunity to improve the signal/background ratio when extracting the π^0-s. In the left panel of Fig. 2 the asymmetry distribution of all $m_{\gamma\gamma}$ pairs is shown (MB data). The enhancement at high asymmetries comes from the abundant low p_T photons and hadrons depositing small energy - i.e. non-π^0, combinatorial background pairs. By applying a modest, say $\alpha < 0.6$ asymmetry cut the S/B ratio can be significantly improved at a small cost of lost true π^0-s, as demonstrated in the right panel of Fig. 2. The ROOT commands are here:

```
ggntuple->Draw("asym","mass<1.0");
ggntuple->Draw("mass>>htemp1","mass<1.0");
ggntuple->Draw("mass>>htemp2","mass<1.0&&asym<0.6");
htemp1->Draw();
htemp2->Draw("same");
```

Let's make a crude estimate of the *raw* π^0 and η spectrum, i.e. the pair p_T distribution of those pairs which give an invariant mass in the π^0 or η "mass window". The estimate will

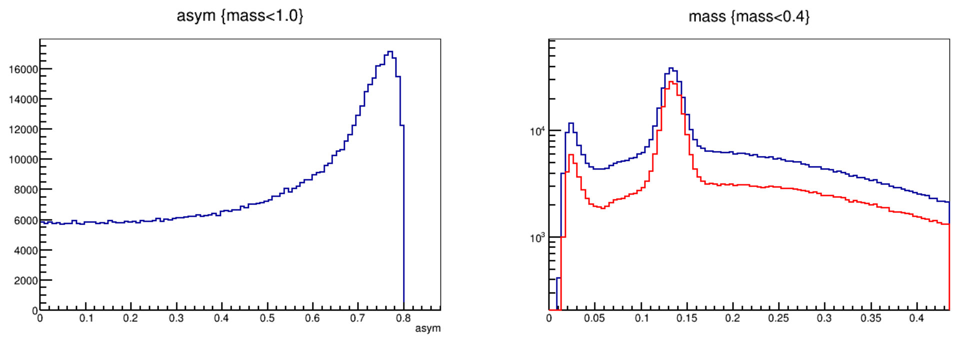

FIG. 2. MB data, plots from the pair ntuple. Left panel: energy asymmetry distribution for all pairs. Right panel: invariant mass distribution without asymmetry cut (i.e. with the default $\alpha < 0.8$) and with a tighter $\alpha < 0.6$ cut (red histogram. The signal to combinatorial background ratio is significantly improved.

be crude because we don't attempt to account for the combinatorial background. However, we can try to clean up the spectra by making a cut on the "photonness" of both clusters. This is illustrated in Fig. 3 for π^0 in both MB and ERT data, and η (ERT only). The blue histograms are the p_{T} distribution of pairs in the mass window, the red histoes the same but with an additional cut on the *chisq* variables. The principal commands:

```
ggntuple->Draw("pt","abs(mass-0.135)<0.015");
ggntuple->Draw("pt","abs(mass-0.135)<0.015&&chisq1<3.0&&chisq2<3.0");
ggntuple->Draw("pt","abs(mass-0.548)<0.1");
ggntuple->Draw("pt","abs(mass-0.548)<0.1&&chisq1<3.0&&chisq2<3.0");
```

Note that the mass window for π^0 is much narrower than for η. In the MB π^0 spectra at low p_{T} the *chisq* cut eliminates some 20-30% of π^0 counts (red vs blue histogram), in line with what the left plot in Fig. 1 suggests (the estimated magnitude of the combinatorial background). In ERT there is very little change, again in line with the small combinatorial on the middle panel in Fig. 1. Finally, for η there is a significant drop in the yields for the entire (high p_{T}) spectrum, in line with the large combinatorial (as compared to the small peak) in Fig 1. Also, the spurios "flattening" at very high p_{T} in the blue histogram is completely gone.

As stated earlier, the time-of-flight (*tof*) for clusters is re-adjusted such that it will be zero for true photons. In the left panel of Fig. 4 for cluster pairs which have an $m_{\gamma\gamma}$ in the

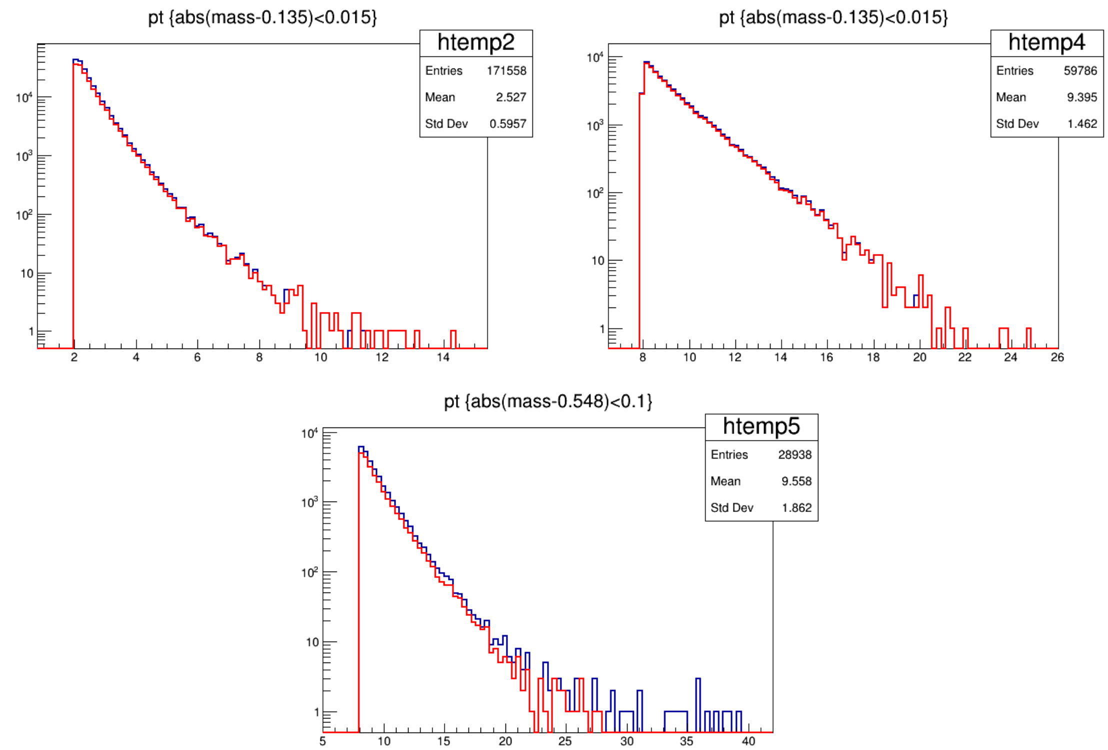

FIG. 3. Transverse momentum (p_{T}) distributions of photon candidate pairs in the π^0 window (top panels) and η window (bottom panel). Blue histograms: p_{T} for all pairs (no identification of the cluster as photon). Red histograms: p_{T} for pairs where both clusters satisfy the *chisq* "photonness" criterion. Top left: MB data, top right and bottom: ERT data.

π^0 window we plot *tof* values vs each other. True π^0-s are in the big peak at (0,0). If we now look at the horizontal band to the right of the peak, the first cluster is still likely to be a photon (small *tof*), but the second is more likely to be a hadron. Is this reflected in other observables characterizing the clusters, too? The answer is yes: on the right panel in Fig. 4 we plot the "photonness" variables *chisq1* in blue, *chisq2* in red. The second cluster, presumed hadronic, has on average higher *chisq2* value.

There is one more important lesson to be learned here. Real-life analyses are complicated. There is no "silver bullet", single cut, that, by miracle, selects *all* photons and nothing else. Cuts are characterized by their *efficiency* (how much of the desired signal they preserve) and *contamination* (how much of the undesired signal passes). Tighter cuts may reduce contamination, but always at the cost of efficiency. The art of analysis is to optimize the signal/background, without sacrificing too much signal. This is usually achieved by combining different cuts with AND or even OR operations. In the last decades probabilistic approaches of combining cuts are also gaining ground.

```
ggntuple->Draw("tof1:tof2","abs(mass-0.135)<0.015&&abs(tof1)<20.0&&abs(tof2)<20.0","colz
TH1D *hsmall = new TH1D("hsmall","hsmall",60,0.0,30.0);
TH1D *hlarge = new TH1D("hlarge","hlarge",60,0.0,30.0);
ggntuple->Draw("chisq1>>hsmall","abs(mass-0.135)<0.015&&abs(tof1)<3.0&&abs(tof2)>5.0");
ggntuple->Draw("chisq2>>hlarge","abs(mass-0.135)<0.015&&abs(tof1)<3.0&&abs(tof2)>5.0");
hlarge->SetLineColor(2);
hlarge->SetLineWidth(2);
hsmall->SetLineWidth(2);
hsmall->Draw();
hlarge->Draw("same");
```

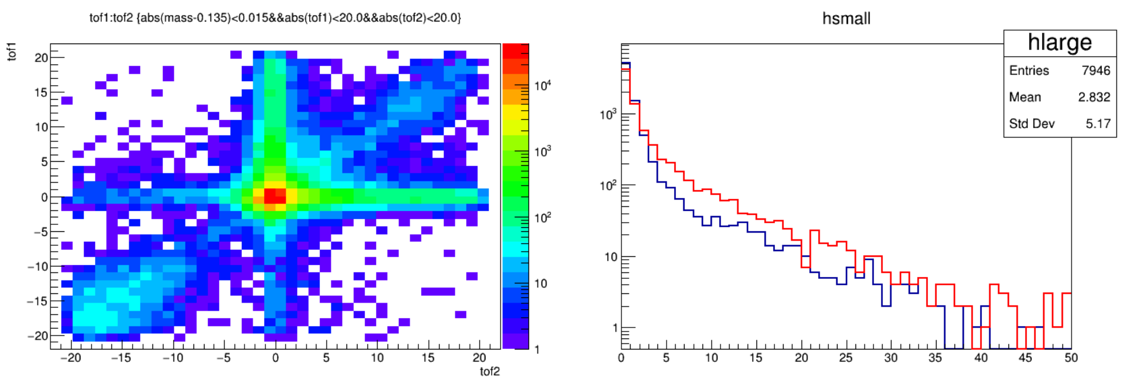

FIG. 4. Left panel: time-of-flight values of two clusters which produce an invariant mass in the π^0 window. TOF for true photons is per construction around zero so the true (non-combinatorial) pairs are in the sharp peak at (0,0). Right panel: *chisq* (“photonness”) of the clusters in the horizontal band above the main peak at (0,0). Blue histogram is *chisq1* of the presumed true photon, the red histogram is the *chisq2* of the presumed hadron. See text for more details.

Let’s turn to single clusters (photon candidates, *gnt* Ntuple), and look first at the position of the clusters (assumed impact point) in a particular sector. In Fig. 5 the y, z positions of the clusters are shown from MB and ERT data. The depletion at the edges is due to the algorithm (clusters too close to the edge are not reconstructed because energy may leak out and the resulting *ecore* will be incorrect). The white areas represent “dead” channels that never fire, and the blue areas have only few hits, although our expectation is that the clusters should be approximately uniformly distributed over the sector. Such dead or low-sensitivity areas should be taken account in a real analysis when the acceptance and efficiency of the detector are calculated.

```
f1->cd();
gnt->Draw("y:z","sec==1","colz");
f2->cd();
gnt->Draw("y:z","sec==1","colz");
```

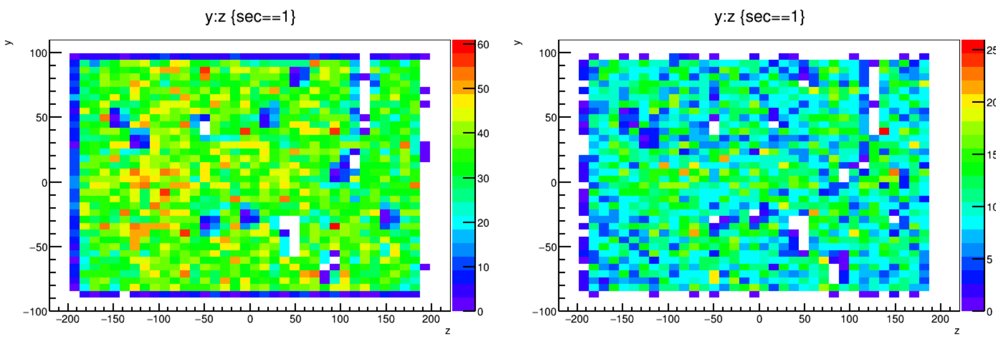

FIG. 5. Position of cluster centers (calculated impact point) in sector 1 (PbSc West). Left panel: MB data. Right panel: ERT data.

Unfortunately not all clusters in *gnt* are photons. The *chisq* variable is an estimate of the "photonness" based on a parametrized shower profile, but its detailed description is beyond the scope of this write-up. However, it is interesting to look at the compactness of the shower. Electromagnetic showers are narrow, most of the energy is deposited in a single tower, while hadronic showers are fairly wide. Besides the total energy (*ecore*) the Ntuple also stores the energy in the "central" (most energetic) tower (*ecent*). The ratio *ecent/ecore* is expected to be high for photons, lower for hadrons. In Fig. 6 this ratio is plotted versus the total energy for MB data, first without any other cuts, then cutting on the *chisq* variable, first deliberately selecting "photon-like" showers, then "un-photon-like" showers. Without cuts the *ecent/ecore* has two characteristic peaks, which then get almost perfectly resolved by the "photon - not photon" cut. Of course that also means that *ecent/ecore* in its own right can be a powerful cut to select photons.

```
f1->cd();
gnt->Draw("ecent/ecore:ecore","ecore<4.0","colz");
gnt->Draw("ecent/ecore:ecore","ecore<4.0&&chisq<3.0","colz");
gnt->Draw("ecent/ecore:ecore","ecore<4.0&&chisq>5.0","colz");
```

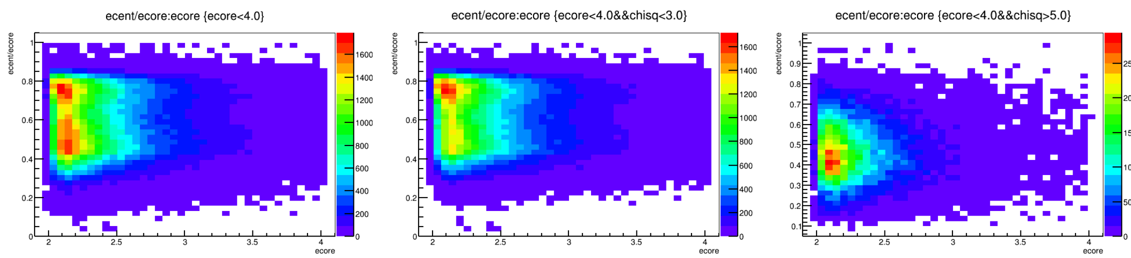

FIG. 6. Shower compactness (*ecent/ecore*) vs shower energy in MB data. Left panel: no cuts. There is a characteristic double-peak structure. Using the *chisq* ("photonness") variable on the middle panel we select showers that are likely photons (small *chisq*), on the right panel clusters that are most likely not photons (large *chisq*). The compactness variable is quite powerful distinguishing between them (i.e. for photon identification).

As mentioned before, the *chisq* variable is a powerful tool to enhance photons and reject hadrons in the sample. In Fig. 7 we plot the p_{T} distribution of the clusters in the ERT data sample, first with no cut (blue histogram), then with a cut on *chisq* (red histogram). Without the cut, the spectrum spectrum "flattens out" at high p_{T}, which is non-physical. The effect of the cut becomes visible early on, and will be quite dramatic above $15\,\mathrm{GeV}/c$. Up to about $25\,\mathrm{GeV}/c$ the red histogram shows the expected power-law shape, above that it also flattens out. You may try to cut also on other *gnt* variables to get rid of the nonphysical flat part of the red histogram.

```
TH1D *h1 = new TH1D("h1","h1",100,0.0,50.0);
TH1D *h2 = new TH1D("h2","h2",100,0.0,50.0);
h1->SetLineWidth(2.0);
h2->SetLineWidth(2.0);
h2->SetLineColor(2);
gnt->Draw("pt>>h1","");
gnt->Draw("pt>>h2","chisq<3.0");
```

```
h1->Draw();

h2->Draw("same");
```

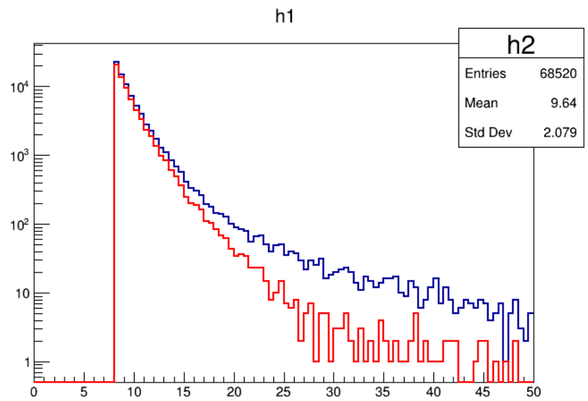

FIG. 7. Blue histogram: p_T spectrum from the ERT data, *gnt* ntuple. Red spectrum: same, but with a cut on "photonness" (*chisq*). Much - albeit not all - of the nonphysical flattening at high p_T is gone.

Please continue to play with the data! The variables in the ntuples are there for a reason! Try plotting them against each other, making them part of cuts when you plot a spectrum. Make invariant mass plots in bins of pair p_T, fit the combinatorial background with some simple function (first, second order polynomial) and try to get a "clean" raw π^0 yield as a function of p_T. Try to imagine in advance, what effect a cut would have, then actually make it and compare the result with your expectations. If you are open-minded and play creatively with them, *the data will start speaking to you and telling you, how to analyze them.*

Have fun!