# Filter Proposal for EHE analysis with IC40

Aya Ishihara for EHE PWG*

January 28, 2008

## 1 Physics Motivation

Purpose of this EHE filtering is to efficiently transfer high energy events i.e. EHE neutrino induced event candidate and high energy muon bundle events. The main physics analysis from the EHE filtered sample are the EHE neutrino search, high energy muon (bundle) structure study (collaboration with CR WG [1]) and monopole/exotic event search (Exotic WG [2]).

## 2 Basis of the EHE Filtering 2008

EHE WG has been using number of photo-electrons (NPE) as the basis of the analysis, because the NPE parameter shows a good correlation with the energy of the particles near the detector. The muon energy and NPE relation is shown on the left plot in Fig. 1 and the muon energy and Nch relation is shown on the right.

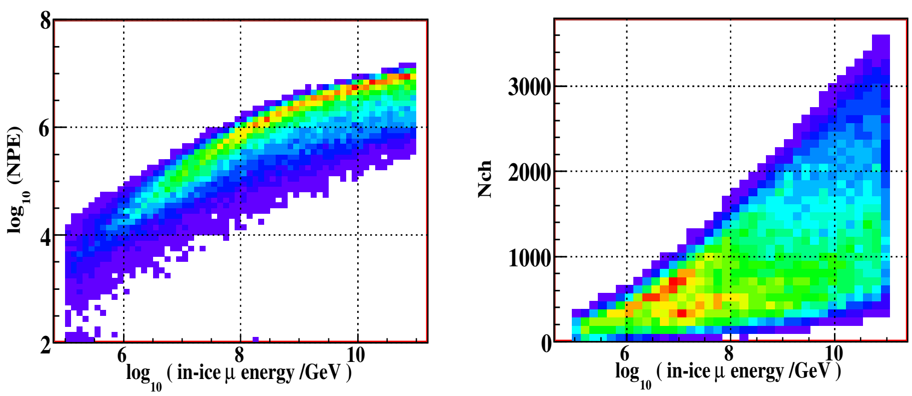


Figure 1: Correlation between MC truth energy and NPE on the left and between MC truth energy and Nch shown on the right. log NPE is more sensitive parameter to the particle energy compared to Nch.

In 2006 and 2007, Nch > 80 filtering (Nch filter) is used for EHE sample filtering because Nch filter does select large NPE events. However, it is observed that Nch filter still retain some fraction of small NPE events which have not been used for analysis. As the more number of strings, the more effecient background filtering become important, we propose to use NPE as the basis of the 2008 EHE filtering in stead of Nch filter. Also Nch cut biases to NPE distribution in mid-NPE region (up to approximately log NPE $\sim$ 2.6) as seen in the left plot of Fig.2. These events in the NPE region are currently disregarded in the offline EHE analysis but for the further muon bundle background study and the improvement of the EHE neutrino sensitivity a portable access to the unbiased distribution is strongly preferred in 2008.

*Technical Contact:aya@hepburn.s.chiba-u.ac.jp

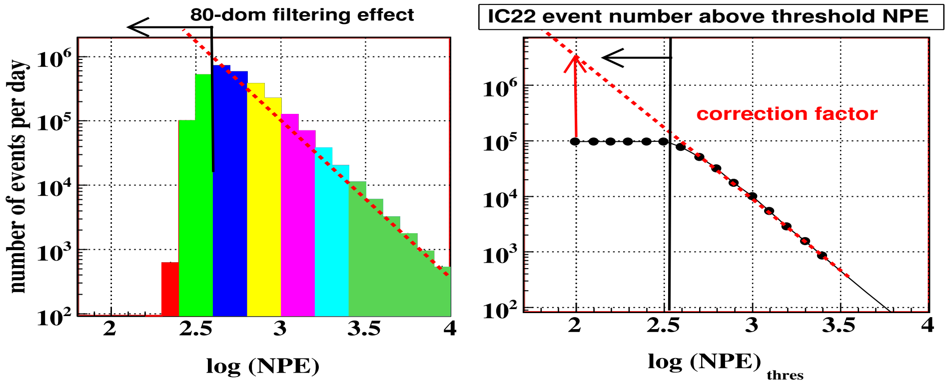


Figure 2: Left: IC22 event rate distribution normalized per day as function of log NPE. NPE distribution indicate simple power-low behavior down to log NPE $\sim$ 2.6 and Nch filter effect is visible in the lower NPE region. Right: the integrated IC22 event rate per day as function of threshold log NPE. To estimate event rate without Nch filter bias, a correction factor multiplication will be applied on the Nch filtered integrated event rate distribution assuming the high NPE slope is unchanged in the lower region.

# 3 Event Rate and Data Volume Estimation

In this study, I used data from 4 runs, Run00108986, Run00108987, Run00108988 and Run00108989 in Aug. 2nd, 2007. The runs correspond to $\sim$ 25 hours of livetime containing 115635 events after Nch filter. The total event size after Nch filter of these 4 runs are 2.7 GB. The average EHE event size can be approximately converted to 23kB/events. As it can be expected that the size of physics data (no geometry, calibration, detector information) in 2008 is similar to that with IC22 as long as the same NPE region is considered, in the following this value is used for the filtered data volume estimation in 2008. The NPE distributions of this sample is shown on the left in Fig.2 and integrated number of events above threshold log NPE is shown on the right-hand side. It can be seen in the plots that a bias from Nch filter condition become visible below log NPE $\sim$ 2.6. To estimate event rate in 2008 without Nch filter bias, correction factors will be applied on the Nch filtered sample below log $\mathrm{NPE}_{thres}$ = 2.6. Assuming the high NPE event rate behavior continues down to the log NPE $\sim$2.0, the correction factors at log $\mathrm{NPE}_{thres}$ = 2.0, 2.2, 2.4, 2.6 are estimated to be x30, x10, x3 and x1.1 respectively and the same factor is used in the case for the other number of strings.

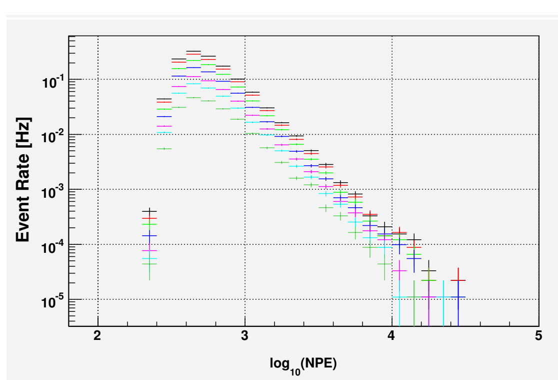


Figure 3: Event rate as function of log NPE. Colors indicate different numbers of strings used to calculate the total NPE. Colors for 22, 20, 18, 15, 13, 11 and 9 strings are black, red, light green, blue, pink, light blue and green respectively.

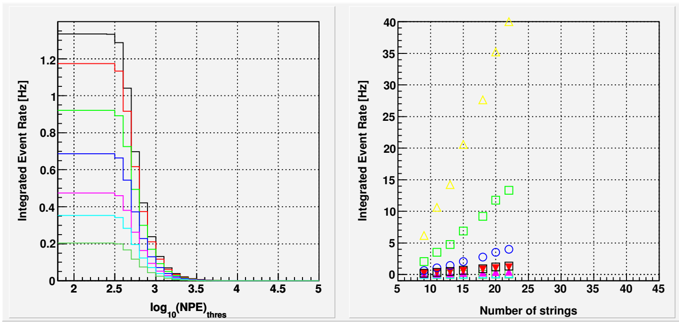


Figure 4: Integrated event rate as function of threshold log NPE and number of strings. Nch filter bias correction is applied on the right plot. Colors on the left indicate different numbers of strings used to calculate the total NPE as in Fig. 3 and on the right indicate different threshold log NPE. Colors on the right is log NPE$_{thres}$ of 2.0, 2.2, 2.4, 2.6, 2.8, 3.0, 3.2 are yellow, light green, blue, red, pink, light blue and green. Black are events which passed Nch filter.

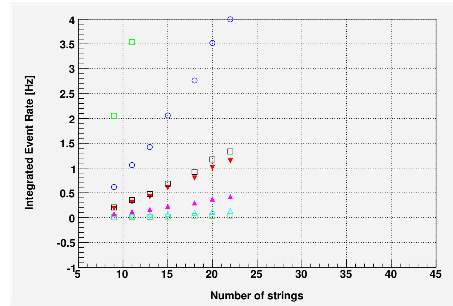


Figure 5: The integrated event rate as function of number of strings as the right plot in Fig. 4 but now focusing on the higher log NPE$_{thres}$. Colors on the right is log NPE$_{thres}$ of 2.4, 2.6, 2.8, 3.0, 3.2 are blue, red, pink, light blue and green. Black are events which passed Nch filter. Extrapolation of the data points to 40 strings gives an approximate event rate with IC40.

In Fig. 3, the event rates per second are shown for different number of strings indicated by different colors from 9 strings to 22 strings. The rates for N string $< 22$ are obtained by applying the string number reduction to the IC22 data and Nch filter is re-applied. No Nch bias correction is applied in the plot. Given on the left in Fig. 4 is the integrated event rate as functions of threshold log NPE for each string numbers. Shown on the right is also the integrated event rate but now as functions of number of strings for different threshold log NPE indicated by different colors and the Nch filter bias correction factor is applied. The 40 strings event rates can be inferred from the plot on the right in Fig. 4 and Fig. 5 (the same plots with different axis range) by extrapolating total event rate values from $N_{string} \leq 22$ to 40 strings, as it can be seen in the plot that the expected event rates show approximately linearly increase with the number of strings. This gives an estimation that threshold log NPE value of 2.6 gives ∼3.0 Hz with 40 strings while Nch $>$ 80 filter will yield ∼3.8 Hz of event rate in 2008. EHE PWG propose to use threshold log NPE value of 2.6 for the 2008 EHE filtering conditon, i.e. All events with $\log_{10}$ NPE above 2.6 will be transfered to North. This ∼2.2 Hz event rate can be converted to expected data size of 4.37 GB per day as summarized in Table 1.

| Filtering Condition | IC40 Event Rate [Hz] | IC40 Data Size [GB/day] |
|---|---|---|
| Threshold $log_{10}$ NPE $> 2.6$ (proposed) | 3.0 | 5.9 |
| Threshold Nch $> 80$ (old) | 3.8 | 7.55 |

Table 1: Expected data volume per day (in *.i3.gz format).

# 4 Example Filter Module and Python script and CPU

Projects used in this filter are
1) Portia
http://code.icecube.wisc.edu/svn/projects/portia/trunk r41017 or after
and 2) EHEFilter

http://code.icecube.wisc.edu/svn/sandbox/ishi-work/EHEFilter

An example script can be found at
http://code.icecube.wisc.edu/svn/sandbox/ishi-work/EHEFilter/resources/scripts/ExampleScript_EHEFilter08.py
Time used to run this script is compared with the time used for simple Nch filter, e.g.,
http://code.icecube.wisc.edu/svn/sandbox/ishi-work/EHEFilter/resources/scripts/ExampleScript_NchFilter08.py

Time consumed by Nch filter is xxx sec, time used by portia plus NPE filter is yyy sec.

## Appendix

# A The NPE calculation and The Baseline Estimation

The filtering variable proposed to use in this request is NPE calculated using the PORTIA project (http://code.icecube.wisc.edu/svn/projects/portia) in which NPE is defined as total integrated charge divided by a SPE charge. The only ambiguity may come from the baseline determination as there is no perfect baseline estimation method yet exists for every type of signals.
In the PORTIA project, there is a set of option to choose how to decide baseline. The most appropriate method for EHE analysis has been studied using the standard candle sample [3], channel-by-channel basis and summarized as *eheoptimized* option in the PORTIA.
Here the observation and the option are summarized.

For ATWD, it has been observed that waveforms with small pulses (contains a few PEs), iteration methods work the best as known from hardware people. However this would not work for waveforms with a large number of PEs. In this case the average of first 2 bins are the best, generally. However for this option we can not avoid the possibility that the effect from a random photon or external effect (e.g. previous events).

Considering these situation, baseline estimation for ATWD with *eheoptimized* option would do:
1. Take the first 2 bin average.
2. if the value is smaller than -1mV, larger than 2mV, consider the baseline as 0mV.
3. Then calculate NPE.
4. If the obtained the first 2 bin average is between -1mV and 2mV and NPE value is less than 20, calculate the best baseline with the iteration method.
Also note that the systematic error due to the baseline mis-identification by 1mV from ATWD in total number of pe by $\mathrm{NPE}_{error} = \pm$ 1(mV)*3.3(ns)*128(bin)/43(ohm)/1.6(pC) = $\pm 6.1 pe$ per DOM.

Similarly for FADC, baseline estimation with *eheoptimized* option would do:
1. Use the average of the first 3 bins.
2. If the value is not within $\pm$0.5 mV, then assume that baseline is at 0mV to avoid strangely triggered pulses.
For FADC it need to be reminded that because of large undershoots, we count value only <u>above</u> baseline to get integrated charge, while for ATWD, we use both positive and negative entry with an assumption that if we have correct baseline and we don't have any pulse in it, we get zero integrated charge.
The systematic error due to the baseline mis-identification by 1mV from FADC in total number of PE are estimated as, $\mathrm{NPE}_{error} = \pm$ 1(mV)*25(ns)*256(bin)/43(ohm)/1.6(pC) = $\pm 93 pe$ per DOM.

# References


[1] the Cosmic-Ray working group, Cosmic Ray Filter Request for 2008

[2] the Exotic Physics working group,

[3] http://www.ppl.phys.chiba-u.jp/∼aya/SC/