abstract

I report on some modifications made on the original simulation and reconstruction program written by Molly Swanson\textsuperscript{1} for CHICOS. The modifications include simulation of the background and noise effect in the real event in the Monte Carlo simulation program, a procedure to cut on some reconstruction events to improve the accuracy of the reconstruction, and a procedure to filter out the noise in the cosmic ray data. Also, I present some preliminary analysis of the cosmic ray events from 1\textsuperscript{st} May – 31\textsuperscript{th} July 2002.

\textsuperscript{1}Refer to the Senior Thesis by Molly Swanson
1 Introduction
This summer (18th June – 20th Aug 2002) Chan Pak Yuen and I went to the Kellog Radiation Laboratory in California Institute of Technology (Caltech) to do a research project. I worked on a project called CHICOS (California HIgh school Cosmic ray ObServatory) under Prof. Robert McKeown’s supervision.

CHICOS is an array of detectors designed to observer air showers caused by cosmic rays with energy above $10^{18}$ eV. The detectors are installed in high schools throughout the Los Angeles area and linked back to Caltech via the internet for data analysis. There are currently 22 detectors sites\(^2\) being installed. A search program is already written to search for the potential air showers by looking for the coincidences between detectors at several different sites\(^3\), and a reconstruction program is written to reconstruct the cosmic ray from these air showers (i.e. to find out the direction of the incident cosmic ray, the location of the hit core of the shower on the horizontal plane, the time hit of the core and the energy of the incident cosmic ray, which is totally 6 parameters)\(^4\). Also, a simulation program is used to test the accuracy of the reconstruction program.

And my work was mainly on the modification of the simulation and the reconstruction program so that a more realistic and accuracy result can be obtained.

2 Goodness of the reconstruction program
The reconstruction program is based on the minimization of the chi-square function $\chi^2$ and negative log-likelihood function $-\log(L)$ with respect to the arrival direction $(\theta, \phi)$, arrival time $t$, core location $(x, y)$ and the normalization constant $C$,\(^5\) which are defined by

\[
\chi^2 = \frac{1}{n - 3} \sum_{\text{detector}} \left( \frac{T_i - T_f(R_i, \theta, \phi) - T_d(\rho_i, R_i) - t}{T_s(\rho_i, R_i)/\sqrt{\sigma_i \rho_i}} \right)^2
\]

\[
-\log(L) = -\log \left( \prod_{\text{detector}} \left( \frac{1}{\sigma_i \sqrt{2\pi}} \right) \exp \left[ -\frac{1}{2} \sum_{\text{detector}} \left( \frac{\rho_i - \rho(R_i)}{\sigma_i} \right)^2 \right] \right)
\]

\(^2\)up to 30th Aug 2002
\(^3\)refer the the Senior Thesis by Molly Swan
\(^4\)In the following of this report, $\theta$ and $\phi$ is used for polar angle and azimuthal angle of the direction of the incident cosmic ray respectively, $x$ and $y$ is used for the location of the hit core of the air shower on the horizontal plane where the detector at Caltech is at origin, $t$ is used for the time of the shower hit on the ground and $E$ is used for the energy of the cosmic ray.
\(^5\)The detail of the reconstruction algorithm can be found in Molly’s senior thesis
where $T_i$ is the time detector $i$ was hit, $\rho_i$ is the particle density for detector $i$, $a_i$ is the scintillator area of the detector $i$, $T_f$ is the time the shower plane hits each detector, $\sigma_i = \rho_i/a_i$, and $T_d$ and $T_s$ are calculated as:

$$T_d(\rho, R) = 2.6\left(1 + \frac{R}{30}\right)^{1.5} \rho^{-0.5} \text{nec}$$ \hspace{1cm} (3)

$$T_s(\rho, R) = 2.6\left(1 + \frac{R}{30}\right)^{1.5} \rho^{-0.3} \text{nec}$$ \hspace{1cm} (4)

$$C = 2.636 \times 10^{-14} E\text{[eV]}^{0.985}$$ \hspace{1cm} (5)

From figure 1, 5000 simulated events with random incident direction and core location but constant energy is reconstructed and the histogram of (reconstructed parameter - simulated parameter) is shown. It can notice that the reconstructed parameters is not so good, especially for the angle $\theta$ and $\phi$. There is a long tail on the plot which means that we can not obtain the correct incident direction of the cosmic ray which is interested for high energy cosmic events ($> 10^{20}\text{eV}$).

So, it is necessary to find out some method so that the goodness of the reconstruction program can improve. The method found is to cut off some reconstructed events with large error calculated by the package MINUIT from CERNlib. The package MINUIT is the program used in minimizing the chi square function $\text{chi}^2$ (1) and the negative likelihood function (2) in the reconstruction program.

In MINUIT, it can calculate the parabolic error which is from the error matrix and also the assymetric positive and negative errors. In figure 2, we can notice that a linear relation can be found between the parabolic error and the difference of the reconstructed parameters with the simulated parameters.

In fact, a more clear relation can be obtain if we can filter out some points first. By MINUIT, the assymmetric positive and negative errors is sometimes give the zero which mean that the error is ”at limit”. And if we first filter out the points which has either the zero positive or negative error of $\theta$ and $\phi$, the linear relation is much better (figure 3).

---

6the package can be found in webpage:
http://wwwinfo.cern.ch/asdoc/minuit/minmain.html
MINUIT Reconstruction Results

Shower energy = 1E+19 eV

Example reconstruction test: original result using sq_all.dat

Figure 1: original results of the reconstruction program. the histogram of the (reconstructed parameter - simulated parameter) for each parameters are plotted. $\sigma$ displays in each graph is the standard derivation of the distribution on each plot.
Figure 2: above: the plot of $\log(\|\theta_{stimated} - \theta_{reconstructed}\|)$ against $\log$ (parabolic error of $\theta$) below: the plot of $\log(\|\phi_{stimated} - \phi_{reconstructed}\|)$ against $\log$ (parabolic error of $\phi$)
Figure 3: The point with zero positive or negative error is eliminated in the graph. Above: the plot of \( \log\left(\frac{k_{\text{stimulated}}}{k_{\text{reconstructed}}}\right) \) against \( \log\left(\text{parabolic error of } \theta\right) \). Below: the plot of \( \log\left(\frac{\phi_{\text{stimulated}} - \phi_{\text{reconstructed}}}{\phi_{\text{reconstructed}}}\right) \) against \( \log\left(\text{parabolic error of } \phi\right) \). We see that the points far away from the linear relation is removed.
Figure 4: the histogram of different parameters after a cut on parabolic error of $\theta$ and $\phi$ is shown again. (cut on parabolic error of $\theta$ = cut on parabolic error of $\phi = 0.3$) The efficiency of the cut is $\sim 72\%$ with cosmic ray energy $= 10^{19}$ eV. We see that the standard derivation is reduced in all the plots.
So, from figure 3, we can see that if we set a cut on the parabolic error, such that the reconstructed events which have a large parabolic error of $\theta$ and $\phi$ is neglected. A much better reconstruction result can obtain (figure 4). (Of course, the cost is the decrease in the efficiency of successfully reconstruction and the efficiency of the cut as a function of the incident cosmic ray energy is shown in figure 5 and 6.)
3 modification of the simulation program

The simulation program used in the analysis above is just simulate for the 2-D detector array (ie. all detectors are on the same horizontal plane). And the noise in the photomultiplier tube (PMT) and the background hit is not considered also. Also, the signal height of the PMT simulated in the original simulation program can only give the exact value (ie. no fluctuation in the signal height) so that we can always get the exactly correct particle density from detector which it is obviously impossible. So, to make the simulation becomes more realistic, a modified simulation program is necessary to include all of these things.

A typical signal pulse height distribution of the PMT is shown in figure 7 and we know that the background rate of the PMT used in CHICOS is about 50 Hz. Also, in practical, the number of the particles hited at a detector is calculated by the signal detected so that it is impossible to obtain the exact number of particle hited. In the modified simulation program, monte carlo method is used to simulated the noise effect, and the signal of the particle hited is simulated as the Gaussianian distribution.

The reconstruction result with the modified simulation program used the 3-D modified detector arrays is shown in figure 8. We can see that the result is much worse and it is make sense since more uncertainties is included in the simulation program. And we can
also observe that the cosmic ray is more probable to have a larger reconstructed energy, it is because if the noise is included in an air shower, more particles is counted as generated by the incident cosmic ray and thus the energy of this cosmic ray reconstructed must large than that it should be.

4 procedure to remove the noise in an shower event

From figure 8, we see that the reconstruction is not good enough to calculate the parameters of the incident cosmic ray, and most of the additional error is come from the noise and background of the data. So, a procedure is purposed to eliminate such noise and background.

The first thing used is to consider the chi square function $\chi_i^2$ for both the time and density at each site which is defined by:

$$\chi_i^2 = \left( \frac{T_i - T_f(R_i, \theta, \phi) - T_d(\rho_i, R_i) - t}{T_s(\rho_i, R_i) / \sqrt{a_i \rho_i}} \right)^2 + \left( \frac{\rho_i - \rho(R_i)}{\sigma_i} \right)^2$$  \hspace{1cm} (6)

where the meaning of the variables is same as before and i is refer to the detector which
MINUIT Reconstruction Results

Shower energy = 5e19 eV

Example reconstruction test: recon_mod.c using sg_hs_3D.dat

Figure 8: the histogram of the reconstruction result with the use of the modified simulation program with the cut on the parabolic error of $\theta$ and $\phi$ is used. Efficiency of the cut used here $\sim 40\%$. (the cut used here is the cut on both the parabolic error of $\theta$ and $\phi$ is $< 0.3$ (in radian)) Note that there is often an overestimation of the cosmic rays energy which can be explain that sometimes the noise and the background is also considered as the real particles generated by the cosmic ray.
detect the signal in this event. Obviously, $\chi_i > 0$ for all detectors site the reduced chi square $\chi_{\text{reduced}}$ which is defined by:

$$\chi_{\text{reduced}} = \frac{1}{2n-6} \sum \chi_i$$

where $2n-6$ is refer to the number of degree of freedom. From the well-known statistical theory, it can be prove that $\chi_{\text{reduced}} = 1$. So, we can see that in order to make sure $\chi_{\text{reduced}} \sim 1, \chi_i < 1$ for all $i$. This mean that after the reconstruction, a set of the parameters can be obtained. With the use of these parameters, $\chi_i$ for each detector $i$ can be calculated. And if some of the $\chi_i > 1$, it seems that this data should be the noise and can be removed. Under this consideration, we can try to remove the noise in the shower event.

But in fact, if we just consider this constraint, most of the noise cannot be removed. And a more complicated algorithm is used here to filter out the noise by also consider the distance of the corresponding detectors from the hit core and the signal height of the PMT detected. Since we know that the particle density is small if it is far away from the core location, so if the detector far away from the hit core detect a signal, it is less probable to detect a real coherent particle and thus the signal is probable to be the noise or background. We also know that if the signal from the PMT is too small, it is probable to be the noise of the PMT.

By consider all of this criteria, an algorithm (A1) to filter out the noise is purposed:

1. Run the original reconstruction program for the whole set of data.
2. Find the $\chi^2_i$ for each data point with the use of the reconstructed parameters. (If the reconstruction is not successful, the estimate parameters are used)
3. If the particle density at one site $< 0.5$ (ie. it is probable to be a noise), eliminate this point if $\frac{\chi^2_i}{\rho_i} > C_1$, where $C_1$ is a parameter which is fitted by the data.
4. Run the reconstruction program again for the remaining data point.
5. Calculate $\chi^2_i$ for the remaining data point again with the new set of parameters. (6) If the largert $\chi^2_i > C_2$, eliminate this data point. ($C_2$ is another parameter fitted by the data)
6. Repeat (4)-(6) until all the remaining data point with $\chi^2_i < C_2$ or the number of data points $< 4$.

---

\(^7\) the data point here mean the detected time and the number of particles hit on the detector
\(^8\) the method to estimate the initial parameters can be found in Molly’s thesis
Figure 9: the histogram with the modified simulation program with the cut on the parabolic error of $\theta$ and $\phi$ is used. An a reconstruction program include the suggested algorithm A1 is used also. Efficiency of the cut used here is $\sim 35\%$ (the cut used here and after is the same as the one used in figure 8)
With the use of this algorithm, the histogram of the reconstruction is shown in figure 9. Compare this graph with figure 8, especially from the energy histogram which can reflect the noise contain in a shower event, we notice that the result can get improve which mean that the algorithm suggested can remove the some noise but not the all. So, a modified algorithm (A2) with additional procedures is suggested.  

(1) Use the algorithm (A1) to remove all the possible noise and calculate the reduced chi square function $\chi^2_0$ for the whole set of data. 

(2) Eliminate one date point $i$ in the data set and run the reconstruction program for the remaining data point. Then calculate the reduced chi square function $\chi^2_{\text{reduced}}$ and the distance $r_i$ which is the distance from the eliminated data point to the core location reconstructed by the remaining data point. 

(3) Calculate the function $F_i$ for each detector $i$ which is defined by:

$$F_i = \frac{p_i}{r_i} \{ \chi^2_{\text{reduced}}(1 + \| \frac{\Delta \cos \theta}{\cos \theta} \|)(1 + \| \frac{\Delta C}{C} \|)(1 + \| \Delta r \|)(1 + \| \Delta y \|) \times (1 + \| \Delta \phi \|)(1 + \| \Delta t \|) \}$$

(4) Obtain $F_i$ for all the possible cases. (ie. Try to eliminate one of the data point in the original data set in each time can calculate the corresponding $F_i$) 

(5) Find the minimium $F_i$ among all $F_i$ with non-zero positive and negative error. 

(6) Compare $F_0'$ and $F_i'$ which is defined as follow:

$$F_0' = \{ \chi^2_0(1 + \| \frac{\Delta \cos \theta}{\cos \theta} \|)(1 + \| \frac{\Delta C}{C} \|)(1 + \| \Delta r \|)(1 + \| \Delta y \|) \times (1 + \| \Delta \phi \|)(1 + \| \Delta t \|) \}$$

$$F_i' = \{ \chi^2_i(1 + \| \frac{\Delta \cos \theta}{\cos \theta} \|)(1 + \| \frac{\Delta C}{C} \|)(1 + \| \Delta r \|)(1 + \| \Delta y \|) \times (1 + \| \Delta \phi \|)(1 + \| \Delta t \|) \}$$

\[\text{we assume that the original algorithm A1 is good enough but sometimes one noise is left in the data set after this algorithm.}\]
Figure 10: the histogram with the modified simulation program with the cut on the parabolic error of $\theta$ and $\phi$ is used. An a reconstruction program include the suggested algorithm A2 is used also. Efficiency of the cut in this plot $\sim 62\%$.
If $F_0' < F_i'$, we argue that the original result is good enough. Otherwise, we argue that the i-th data point is a noise and we remove this data point and reconstruct the parameters by the remaining data point.

With the use of this algorithm A2, the histogram of the reconstruction parameters compare with the simulated parameters is shown on figure 10. We note that the result of the energy reconstruction is better compare with the algorithm A1. And in fact, the standard derivation of the distribution of the energy histogram is nearly the same as that without the noise. In fact, the angle reconstruction $(\theta, \phi)$ is not very good, it is mainly because the uncertainties in the particle density obtain from the pulse height conversion by the PMT (figure 11).
MINUIT Reconstruction Results

Shower energy = $5 \times 10^{19}$ eV

reconstruction test: recon_mod.c without b/g but varied pulse height using sg_hs

Figure 11: the histogram of reconstruction with the simulation program without background noise but varies signal pulse height with the cut on the parabolic error of $\theta$ and $\phi$ is used. Note that it is still have an asymmetric and large error in the histogram of the angle $\theta$ and $\phi$ so we can conclude that the error in the reconstruction is come from the uncertainties in the particles density detected. Efficiency of the cut in this plot is 67%
Data analysis of the real shower data

After the algorithm A2 has developed to remove the noise, the real data from CHICOS from 1st May – 31st July 2002 is analyzed. And the energy histogram of all the successfully reconstructed data is shown in figure 12. There are totally 47 shower events with energy larger than $10^{18}$ eV within this 3 months obtained from the reconstruction program.

In figure 13, we see that all the reconstructed cosmic rays are concentrated into 2 small region. And this result is encouraging since that if more detectors are installed, the number of cosmic rays obtained should be increased.

In figure 14, the aperture of the current detector arrays is shown. We see that the aperture is small very ( $\sim 0$ ) for low energy cosmic ray (with energy $\sim 10^{18}$ eV). It is because the detector density is small in the current set up and thus it is difficult for a low energy cosmic ray to have a 4 site hit $^{10}$. And it can explain why there are no event with energy at about $10^{18}$ eV can be reconstructed by the current detector array.

$^{10}$the data from the detectors will consider as a potential air shower event if there are at least 4 sites hit within a short time interval ( $\sim 0.1$ ms), we call this as trigger event.
Figure 13: The plot of the location of the detector site and the core location of the reconstructed cosmic ray. Square is the detector site and cross is the core location of the reconstructed cosmic ray.

Figure 14: Apperture of the current detector sites in CHICOS (Jul 2002)
We can find that the number of events of high energy cosmic rays is much larger than what it should be,\textsuperscript{11} even though both the aperture and the efficiency of the cut are considered. This can be explain that some of the successfully reconstructed events should not be the real cosmic rays data. But another reason may be because the reconstruction program is often reconstruct a larger energy than the cosmic rays should be. (figure 15) But this error is reduced in higher incident energy ($>10^{19}$ eV) or a stronger cut on the parabolic error of $\theta$ and $\phi$ is used. (figure 16) So, we can conclude that our program can work better in a higher energy range ($>10^{19}$ eV).

From the energy histogram of the real data and also the aperture and efficiency of the reconstruction program for the detector array (figure 6), an estimation on the energy spectrum of the cosmic rays of ultra-high energy ($>10^{19}$ eV) can be obtained.

\textsuperscript{11}refer to M. Nagano and A.A. Watson. "Observations and implications of the ultrahigh-energy cosmic rays." Rev. Mod. Phys. 72, 689-732(2000), and the senior thesis from Molly for an estimation of the number of cosmic rays with energy larger than E.
6 Conclusion
In this report, a procedure to cut some reconstructed events is purposed and a modified simulation program which include the background noise is written. And we see that if the modified simulation program is used, the original reconstruction program is not good enough. So two algorithm A1 and A2 is purposed to remove the noise and background event in the potential air shower data and the result is tested. Finally, the preliminary air shower data is analyzed and we can obtain a number of air shower events and the number of these air shower events is comparable with the theoretical value. It shows that the CHICOS array is functioning well and will be able to collect more air shower data in the future.

7 Acknowledgement
I thank Professor McKeown of California Institute of Technology and Department of Physics, CUHK for granting me this opportunity to learn in Kellog Radiation Laboratory in Caltech.
**Reference**

1. Webpage: http://chicos.caltech.edu


8. P.R. Bevington, D.K. Robinson, Data reduction and Error Analysis for the Physical Sciences