Subject: Data organization question
Posted by Brian Larsen on Tue, 25 Nov 2008 14:56:44 GMT
View Forum Message <> Reply to Message

All,

I am having a challenge here figuring out how to efficiently store a data set in IDL and am looking for some suggestions on what to do, linkedlist, pointers in some other form, or something else I haven't thought of.

This is output from a fortran code (EGS5) that given an incident electron of a certain energy computes all the products on the other side of my shielding material and geometry. This is a one to many problem, meaning that one incident electron can cause a shower of zero or more secondaries in three species (electrons, photons, or positrons). I am shooting 100,000 particles in each of 21+ bins in a Monte Carlo way then tracking all the secondaries.

The data file I am writing out from EGS5 looks like this:

```
 Start history  44
                 0.805        0         2.0
                18.727       -1          6.8
 Start history  45
                 0.231        0         1.8
                 0.104        0         3.5
                19.197       -1          1.4
 Start history  46
                 2.452        0         6.1
                 0.217        0         5.2
                 1.629        0         3.5
                 0.391        0         4.1
                14.846       -1          5.6
```

(continue for a gig or so)

I need to keep track of the incident energy (given in another part of the file), the history number, the secondary energy (1st column), the charge (2nd col), and angle (3rd col).

I am currently using a giant array for this:
[[history], [incident energy], [secondary E], [charge], [angle]]
the annoying thing with this is that the array gets huge and I am afraid of the dreaded "array to large" message and there is a lot of repeated information in the array.

Things I need to do with the data (as this might affect the storage decision):
- Monte Carlo error bars: on the above array is done by resampling the

history numbers and computing the output spectra from there, repeat.
- Compute secondary spectra from all the input histories
- Keep track of the 2-d histogram of primary energy and secondary energy

Does this make sense? Need more info? Thanks for any suggestions.

Brian

---------- ----------
Brian Larsen
Boston University
Center for Space Physics
http://people.bu.edu/balarsen/Home/IDL