Validation of GENIE – Using a Simplified Geometry and Flux in the NOνA-ART Framework

Robert Hatcher[1]

[1]rhatcher@fnal.gov – *Fermi National Accelerator Laboratory*

May 31, 2012

1 Introduction

The GENIE simulation when used for event generation brings together geometry, flux and cross-sections in a conceptually straightforward manner, but in actual practice with complex geometries and flux profiles (both in energy and direction) it is difficult to discern whether the numbers are correct. For instance, the use of of GENIE for event generation should give consistent numbers of events per POT (protons-on-target) independent on how the maximum path length through the detector is determined. And the overall normalization of events can only be worked out analytically in specialized cases. In this note we attempt to perform some of these tests in a manner that explores some potential pitfalls.

2 The flux

The flux consists of a `GSimpleNtpFlux` format file populated with 50,000,000 divergentless rays of 1 GeV ν_μ neutrinos uniformly illuminating a $3m \times 3m$ area. This is arbitrarily assigned the value of 500,000 protons-on-target.

3 The cross-section

The ν_μ-C^{12} and ν_μ-Fe^{56} cross-section was extracted from the spline file used by NOνA: `/grid/fermiapp/nova/aux/gxspl-NUMI-R2.6.0.xml`. The cross-sections in the XML file were converted to ROOT `TGraph` objects using `gspl2root` which were then queried for their interpolated value at 1 GeV. In units of $10^{-38} cm^2$ the values for were: 15.019, 70.3381 and 0.847475 for C^{12}, Fe^{56} and H^1 respectively.

4 The geometry

For these tests a very simple GDML geometry was constructed that has a minimal number of components. The top level is a `vWorld` box (12.5m x 12.5m x 12.5m) within which there is a `vDetEnclosure` box (5m x 5m x 5m), both constructed of "vacuum" ($10^{-25} g/cc$ of H). Inside that are a `vCylinder` and a `vBlock`. The `vDetEnclosure` volume is positioned so that the front face is $z = 0$ of the world.

The cylinder has a radius of 50 cm and length of 1 m, positioned so the downstream surface is at $z_{mother} = 0$ and centered in x and y. This gives it an area of 7854 cm^2 and volume of 785398.2 cm^3. It is constructed of "graphite" ($1g/cc$ of C^{12}) for a total mass of 785398.2 g.

The block has a width of 75 cm, height of 125 cm and a length of 50 cm, positioned so the upstream surface is at $z_{mother} = 0$ and again centered in x and y. This gives it an area of 9375 cm^2 and a volume of 468750 cm^3. It is constructed of "steel" ($7.87g/cc$ of Fe^{56}) for a total mass of 3689062.5 g.

In world coordinates the z position of the plane dividing the cylinder from the block is at 250 cm. There are two additional elements positioned in the `vDetEnclosure` both made of "vacuum" in order to satisfy the needs of the `Geometry` package in the NOνA ART framework. Variants of the geometry were also made where one or the other of the cylinder or block was made out of "vacuum" but otherwise unchanged.

5 Pathlength lists

As GENIE is running neutrino rays through the geometry it determines the amount of material traversed along that ray; for each ray this is broken down by isotope (A,Z) as the interaction probability depends on the amount of material × the cross-section for that material. GENIE reads in a pre-computed list of cross-section splines that are separate for each isotope. GENIE also uses a list of maximum path lengths in each potential isotope to set the overall normalization. It is known that GENIE will abort if it ever encounters a pathlength of a isotope longer than the value in its maximum list, so it is crucial to not underestimate any length. On the other hand, for efficiency reasons one does not want to severely overestimate any lengths either. GENIE acquires the maximum pathlength list in one of three ways:

- by reading in an XML file
- by scanning the geometry using rays starting at random points on a bounding box, with random orientations
- by scanning the geometry using rays derived from the flux driver

The use of a file is definitely a time saver. But care must be used in order not to invalidate the constraint of underestimating a pathlength. For the tests described in this paper the maxpathlength list was one of the following five cases:

- **box:** recalculated on the fly using random rays on a bounding box
- **flux:** recalculated on the fly using rays from the flux generator
- **readin:** read from a file created by the first method
- **scaled:** read from a file where all the lengths were scaled by 2
- **scaledfe:** read from a file where the Fe length was scaled by 2

6 Event rate predictions

In all cases with all other things being the same the event rate should **not** depend at all on how maxpathlength list. The absolute event rate should be:

$$\frac{events}{POT} = \frac{mass}{A_{isotope}} N_A \times \sigma(E_\nu, nucleus) \times \Phi_\nu$$

where Φ_{ν_μ} is the number of ν_μ through unit area (cm^2) per POT. Thus for this specialized flux file Φ_{ν_μ} is

$$\frac{5 \times 10^7}{300cm \times 300cm} \times \frac{1}{5 \times 10^5} = 1.1111 \times 10^{-3} cm^{-2}$$

For clarity we collect the common factors of N_A, 10^{-38} from the cross-section and Φ together as:

$$6.022141 \times 10^{23} \times 10^{-38} cm^2 \times 1.1111 \times 10^{-3} cm^{-2} = 6.69126 \times 10^{-18}$$

For the cylinder we have:

$$\frac{785398.16g}{12g} \times 15.019 \times 6.69126 \times 10^{-18} = 6.57745 \times 10^{-12} \frac{events}{POT}$$

For the block we have:

$$\frac{3689062.50g}{56g} \times 70.3381 \times 6.69126 \times 10^{-18} = 3.10045 \times 10^{-11} \frac{events}{POT}$$

The total rate is the linear sum: 3.758195×10^{-11}.

Event generation proceeded for a fixed number of events and reported the POT used so the inverses are: 1.520344×10^{11} and 3.22532×10^{10} $\frac{POT}{event}$ respectively, with the total being 2.66085×10^{10} The relative ratio of events in the block to those in the cylinder should be $31.0045/6.57745 = 4.7137$.

7 Event generation production

Jobs were run in the NOνA-ART framework. Individual generator jobs did nothing more than create empty records via the `EmptyEvent` module and then run the `GENIEGen` module to create individual interactions, and then the `RootOutput` module to write the event. A second job was run on each created file to extract event vertices for determining which volume the event was generated in.

Files were generated for configurations involving different combinations of: geometry (cylinder-only, block-only, both); maxpathlength method and number of events in a job. Each job ran until it reached a fixed number of events and then reported the POT used. More jobs were run for configurations that targeted a fewer number of events. In particular the combinations of (jobs,events) were: (1000,1000), (100,10000) and (10,100000). The output files were processed to determine the number of events in the cylinder and the block. In total 13320 jobs were run; not all completed successfully. In this past such failures have generally been grid related, the actual cause in the 502 cases of failure have not been definitely determined. Files that could not report the used POT did not contribute to any event counts either.

8 Results

A simple inspection of Figure 1 and Tables 1 and 2 makes clear that our assumption about the independence of reported event rate vs. method of determining the maxpathlength is violated. The average POT/event for all methods other than *flux* appear to be quite consistent and stable which is shown in Figure 2. Since for the *flux* case the average POT/job decreases towards the more stable number as the events per file increases we can infer that rays that were generated to explore the geometry were contributing to the overall POT accounting for the file. This shouldn't be true and is a failure of the code.

The issue of timing is complicated. One would naively expect that jobs producing few events in a file would take longer per event due to initialization issues. One would also expect that jobs that have to explore the geometry to determine the maxpathlength would take longer. The trends in the reported timing data don't seem to bear this out, which implies that there are other more important effects contributing to the processing time, perhaps the I/O.

9 Mitigation

The over-estimation of the POT used has effected some the NUMI experiments simulations. For MINERVA, using the `gNuMIExptEvGen` code supplied with GENIE, the default was switched from using the bounding box method to the flux method for determining the maxpathlength list on 2010-08-05. For the experiments using `GENIEHelper` in the ART framework (NOνA and `LArSoft`) the default was to use the box method and there should be no skew. Both frameworks could be configured to use either alternative so one must understand the conditions under which simulations were actually run to definitively determine whether there is an effect. The magnitude of the effect depends on in part on geometry and flux as well as the relative ratio of rays used in geometry exploration and those used for event generation. Preliminary numbers from MINERVA indicate that this could be a source of a 2% effect for their ν_μ simulations, and 5% for their $\bar{\nu}_\mu$ simulations.

The code in the disparate frameworks has a bit that is essentially the same:

```
GMCJDriver * mcj_driver = new GMCJDriver;
mcj_driver->UseFluxDriver(flux_driver);
mcj_driver->UseGeomAnalyzer(geom_driver);
mcj_driver->UseMaxPathLengths(gOptExtMaxPlXml);
mcj_driver->Configure();
mcj_driver->UseSplines();
mcj_driver->ForceSingleProbScale();
```

It is during the `Configure()` call that the maxpathlength list is determined and flux rays might be used. Ideally this section of code would then be followed by a call:

```
flux_driver->ResetTotalExposure();
```

but such an interface does not uniformly exist for flux drivers in GENIE. New code needs to be written for the GENIE flux drivers to make this possible.

In the mean time, one can use the proper POT accounting calculation to determine the POT usage at this point and subtract it off from the appropriate places. Alternatively, all frameworks continue to support the use of the bounding box method and the use of XML files which don't distort the POT accounting. NOνA is working towards using the XML files as that also reduces some of the start-up time.

Figure 1: POT/event: prediction from Section 6 is the gray dotted line. Left-most clusters are the POT/event average for jobs of 1000 events (error bars show RMS); middle and left clusters for 10000 and 100000 event jobs. Colors distinguish maxpathlength method.

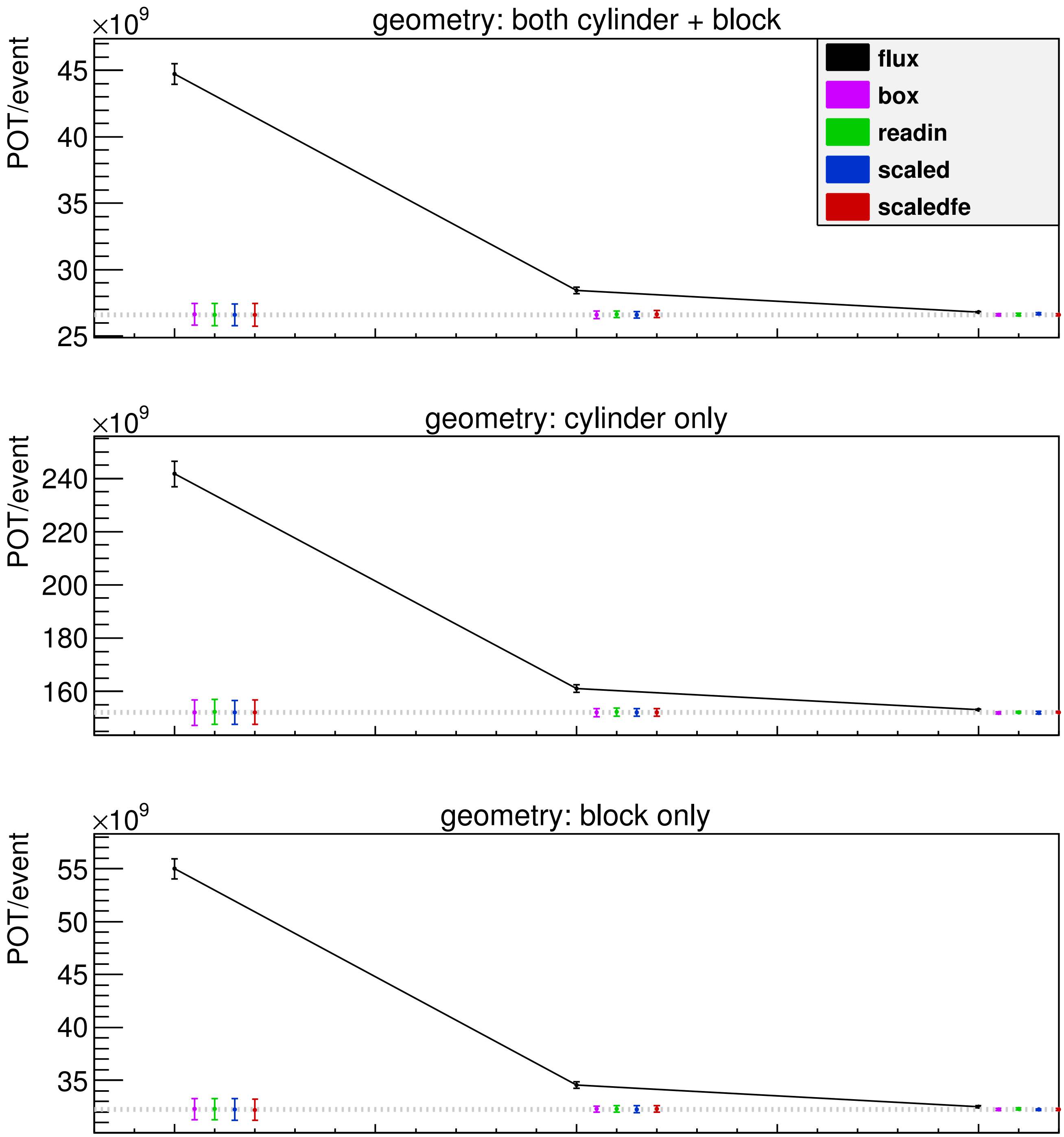

Figure 2: POT/event: as Figure 1 but without the case with maxpathlength being recalculated.

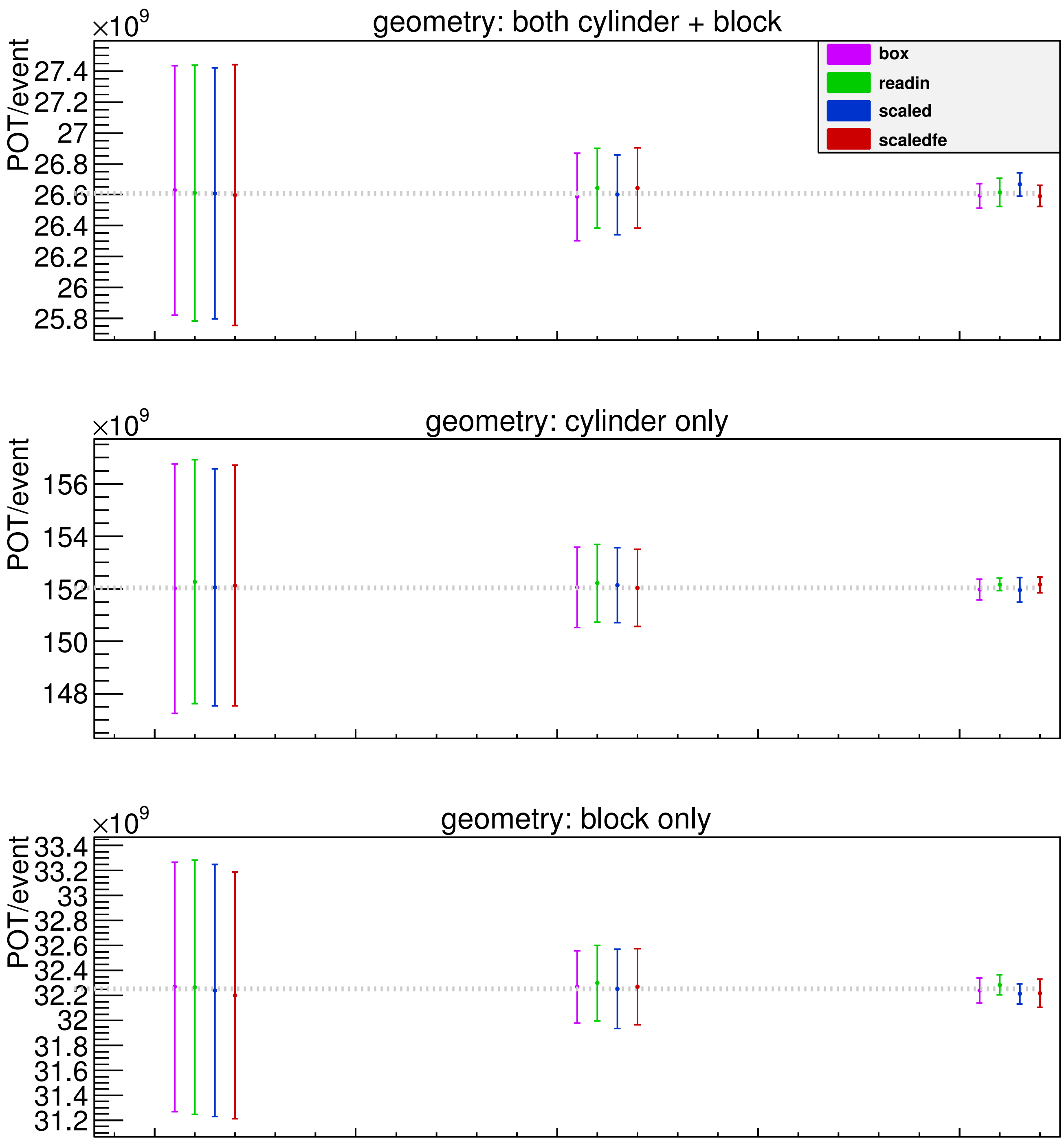

Table 1: POT statistics for successful jobs

geometry	maxpathl	evt/job	avg POT/job	err on mean	RMS	low	high	successful	total POT
both	box	1000	$2.66287\ 10^{13}$	$2.55468\ 10^{10}$	$8.07862\ 10^{11}$	2.42268	2.94578	1000	$2.66287\ 10^{16}$
both	box	10000	$2.65862\ 10^{14}$	$2.83549\ 10^{11}$	$2.83549\ 10^{12}$	2.58395	2.73038	100	$2.65862\ 10^{16}$
both	box	100000	$2.65926\ 10^{15}$	$2.46456\ 10^{12}$	$7.79361\ 10^{12}$	2.64279	2.67265	10	$2.65926\ 10^{16}$
both	flux	1000	$4.47409\ 10^{13}$	$2.48797\ 10^{10}$	$7.84795\ 10^{11}$	4.24508	4.72330	995	$4.45172\ 10^{16}$
both	flux	10000	$2.84277\ 10^{14}$	$2.38744\ 10^{11}$	$2.38744\ 10^{12}$	2.76936	2.90403	100	$2.84277\ 10^{16}$
both	flux	100000	$2.68091\ 10^{15}$	$1.45523\ 10^{12}$	$4.60185\ 10^{12}$	2.67461	2.69093	10	$2.68091\ 10^{16}$
both	readin	1000	$2.66105\ 10^{13}$	$2.61643\ 10^{10}$	$8.27387\ 10^{11}$	2.41253	2.89377	1000	$2.66105\ 10^{16}$
both	readin	10000	$2.66428\ 10^{14}$	$2.58402\ 10^{11}$	$2.58402\ 10^{12}$	2.59980	2.72329	100	$2.66428\ 10^{16}$
both	readin	100000	$2.66160\ 10^{15}$	$2.87999\ 10^{12}$	$9.10731\ 10^{12}$	2.64194	2.67445	10	$2.66160\ 10^{16}$
both	scaled	1000	$2.66092\ 10^{13}$	$2.56923\ 10^{10}$	$8.12462\ 10^{11}$	2.34565	2.94254	1000	$2.66092\ 10^{16}$
both	scaled	10000	$2.66012\ 10^{14}$	$2.58378\ 10^{11}$	$2.58378\ 10^{12}$	2.59983	2.73340	100	$2.66012\ 10^{16}$
both	scaled	100000	$2.66666\ 10^{15}$	$2.41752\ 10^{12}$	$7.64487\ 10^{12}$	2.65446	2.67958	10	$2.66666\ 10^{16}$
both	scaledfe	1000	$2.65985\ 10^{13}$	$2.66915\ 10^{10}$	$8.44061\ 10^{11}$	2.40951	2.96554	1000	$2.65985\ 10^{16}$
both	scaledfe	10000	$2.66446\ 10^{14}$	$2.60886\ 10^{11}$	$2.60886\ 10^{12}$	2.60317	2.72182	100	$2.66446\ 10^{16}$
both	scaledfe	100000	$2.65915\ 10^{15}$	$2.17514\ 10^{12}$	$6.87840\ 10^{12}$	2.64595	2.66695	10	$2.65915\ 10^{16}$
cylinder	box	1000	$1.52008\ 10^{14}$	$1.50496\ 10^{11}$	$4.75910\ 10^{12}$	1.37657	1.69937	1000	$1.52008\ 10^{17}$
cylinder	box	10000	$1.52058\ 10^{15}$	$1.53878\ 10^{12}$	$1.53878\ 10^{13}$	1.48613	1.55404	100	$1.52058\ 10^{17}$
cylinder	box	100000	$1.51964\ 10^{16}$	$1.24239\ 10^{13}$	$3.92879\ 10^{13}$	1.51072	1.52492	10	$1.51964\ 10^{17}$
cylinder	flux	1000	$2.41745\ 10^{14}$	$1.51244\ 10^{11}$	$4.77798\ 10^{12}$	2.28000	2.56855	998	$2.41261\ 10^{17}$
cylinder	flux	10000	$1.61064\ 10^{15}$	$1.43012\ 10^{12}$	$1.43012\ 10^{13}$	1.58261	1.64179	100	$1.61064\ 10^{17}$
cylinder	flux	100000	$1.53138\ 10^{16}$	$7.80618\ 10^{12}$	$2.46853\ 10^{13}$	1.52576	1.53510	10	$1.53138\ 10^{17}$
cylinder	readin	1000	$1.52270\ 10^{14}$	$1.46932\ 10^{11}$	$4.64639\ 10^{12}$	1.38144	1.64534	1000	$1.52270\ 10^{17}$
cylinder	readin	10000	$1.52215\ 10^{15}$	$1.47442\ 10^{12}$	$1.47442\ 10^{13}$	1.46624	1.54992	100	$1.52215\ 10^{17}$
cylinder	readin	100000	$1.52166\ 10^{16}$	$7.63403\ 10^{12}$	$2.41409\ 10^{13}$	1.51886	1.52575	10	$1.52166\ 10^{17}$
cylinder	scaled	1000	$1.52058\ 10^{14}$	$1.42991\ 10^{11}$	$4.51951\ 10^{12}$	1.36759	1.64835	999	$1.51906\ 10^{17}$
cylinder	scaled	10000	$1.52142\ 10^{15}$	$1.42852\ 10^{12}$	$1.42852\ 10^{13}$	1.48350	1.55421	100	$1.52142\ 10^{17}$
cylinder	scaled	100000	$1.51955\ 10^{16}$	$1.46754\ 10^{13}$	$4.64078\ 10^{13}$	1.51261	1.52807	10	$1.51955\ 10^{17}$
cylinder	scaledfe	1000	$1.52125\ 10^{14}$	$1.45090\ 10^{11}$	$4.58815\ 10^{12}$	1.37772	1.67083	1000	$1.52125\ 10^{17}$
cylinder	scaledfe	10000	$1.52039\ 10^{15}$	$1.47532\ 10^{12}$	$1.47532\ 10^{13}$	1.49178	1.56338	100	$1.52039\ 10^{17}$
cylinder	scaledfe	100000	$1.52155\ 10^{16}$	$9.37424\ 10^{12}$	$2.96440\ 10^{13}$	1.51773	1.52672	10	$1.52155\ 10^{17}$
block	box	1000	$3.22681\ 10^{13}$	$3.16019\ 10^{10}$	$9.99338\ 10^{11}$	2.95359	3.55486	1000	$3.22681\ 10^{16}$
block	box	10000	$3.22686\ 10^{14}$	$2.90219\ 10^{11}$	$2.90219\ 10^{12}$	3.16761	3.30119	100	$3.22686\ 10^{16}$
block	box	100000	$3.22410\ 10^{15}$	$3.15456\ 10^{12}$	$9.97561\ 10^{12}$	3.20842	3.24295	10	$3.22410\ 10^{16}$
block	flux	1000	$5.49919\ 10^{13}$	$3.03099\ 10^{10}$	$9.58485\ 10^{11}$	5.21129	5.80541	1000	$5.49919\ 10^{16}$
block	flux	10000	$3.45456\ 10^{14}$	$2.87040\ 10^{11}$	$2.87040\ 10^{12}$	3.37310	3.52329	100	$3.45456\ 10^{16}$
block	flux	100000	$3.24816\ 10^{15}$	$3.19879\ 10^{12}$	$1.01155\ 10^{13}$	3.23441	3.26556	10	$3.24816\ 10^{16}$
block	readin	1000	$3.22661\ 10^{13}$	$3.42667\ 10^{10}$	$1.01825\ 10^{12}$	2.93045	3.58369	883	$2.84909\ 10^{16}$
block	readin	10000	$3.22985\ 10^{14}$	$3.01934\ 10^{11}$	$3.01934\ 10^{12}$	3.14950	3.31359	100	$3.22985\ 10^{16}$
block	readin	100000	$3.22849\ 10^{15}$	$2.53839\ 10^{12}$	$8.02708\ 10^{12}$	3.21630	3.24290	10	$3.22849\ 10^{16}$
block	scaled	1000	$3.22399\ 10^{13}$	$3.68384\ 10^{10}$	$1.00953\ 10^{12}$	2.95037	3.54396	751	$2.42122\ 10^{16}$
block	scaled	10000	$3.22532\ 10^{14}$	$3.17674\ 10^{11}$	$3.17674\ 10^{12}$	3.14504	3.30320	100	$3.22532\ 10^{16}$
block	scaled	100000	$3.22118\ 10^{15}$	$2.51501\ 10^{12}$	$7.95315\ 10^{12}$	3.20886	3.23735	10	$3.22118\ 10^{16}$
block	scaledfe	1000	$3.21993\ 10^{13}$	$3.34327\ 10^{10}$	$9.87257\ 10^{11}$	2.89191	3.50863	872	$2.80778\ 10^{16}$
block	scaledfe	10000	$3.22686\ 10^{14}$	$3.04166\ 10^{11}$	$3.04166\ 10^{12}$	3.15847	3.32617	100	$3.22686\ 10^{16}$
block	scaledfe	100000	$3.22192\ 10^{15}$	$3.56259\ 10^{12}$	$1.12659\ 10^{13}$	3.19980	3.23630	10	$3.22192\ 10^{16}$

Table 2: Event location

geometry	maxpathlength	total evt	total POT	in cylinder	in block
both	box	3000000	$7.98075\ 10^{16}$	525316	2474684
both	flux	2995000	$9.97540\ 10^{16}$	524791	2470209
both	readin	3000000	$7.98693\ 10^{16}$	524910	2475090
both	scaled	3000000	$7.98770\ 10^{16}$	525402	2474598
both	scaledfe	3000000	$7.98346\ 10^{16}$	526087	2473913
cylinder	box	3000000	$4.56030\ 10^{17}$	3000000	0
cylinder	flux	2998000	$5.55463\ 10^{17}$	2998000	0
cylinder	readin	3000000	$4.56651\ 10^{17}$	3000000	0
cylinder	scaled	2999000	$4.56003\ 10^{17}$	2999000	0
cylinder	scaledfe	3000000	$4.56319\ 10^{17}$	3000000	0
block	box	3000000	$9.67777\ 10^{16}$	0	3000000
block	flux	2935000	$1.22191\ 10^{17}$	0	2935000
block	readin	2830000	$9.30743\ 10^{16}$	0	2830000
block	scaled	2588000	$8.86772\ 10^{16}$	0	2588000
block	scaledfe	2788000	$9.25656\ 10^{16}$	0	2788000

Table 3: Event rate

geometry	maxpathlength	total POT/evt	in cylinder	in block
both	box	$2.66025\ 10^{10}$	$1.51923\ 10^{11}$	$3.22496\ 10^{10}$
both	flux	$3.33068\ 10^{10}$	$1.90083\ 10^{11}$	$4.03828\ 10^{10}$
both	readin	$2.66231\ 10^{10}$	$1.52158\ 10^{11}$	$3.22693\ 10^{10}$
both	scaled	$2.66257\ 10^{10}$	$1.52030\ 10^{11}$	$3.22788\ 10^{10}$
both	scaledfe	$2.66115\ 10^{10}$	$1.51752\ 10^{11}$	$3.22706\ 10^{10}$
cylinder	box	$1.52010\ 10^{11}$	$1.52010\ 10^{11}$	
cylinder	flux	$1.85278\ 10^{11}$	$1.85278\ 10^{11}$	
cylinder	readin	$1.52217\ 10^{11}$	$1.52217\ 10^{11}$	
cylinder	scaled	$1.52052\ 10^{11}$	$1.52052\ 10^{11}$	
cylinder	scaledfe	$1.52106\ 10^{11}$	$1.52106\ 10^{11}$	
block	box	$3.22592\ 10^{10}$		$3.22592\ 10^{10}$
block	flux	$4.16324\ 10^{10}$		$4.16324\ 10^{10}$
block	readin	$3.28884\ 10^{10}$		$3.28884\ 10^{10}$
block	scaled	$3.42648\ 10^{10}$		$3.42648\ 10^{10}$
block	scaledfe	$3.32014\ 10^{10}$		$3.32014\ 10^{10}$