GSIM Acceptance of Real CLAS Data

Craig Bookwalter
on behalf of the g12 rungroup

September 9, 2010

Abstract:

Detector simulations are a heavily-used tool in analyses of CLAS data. Our collaboration trusts the `GSIM` software suite to provide an accurate reproduction of the real-world structure and response of CLAS. This note details a study undertaken to verify the quality of this reproduction by using real CLAS data as input into the `GSIM` toolchain. These data are taken from a single raw data file from the g12 run, which had photon beam energy from 1 to 5.5 GeV and a target Z-position of -90 cm.

1 Introduction

Many CLAS measurements rely on acceptance corrections provided by the `GSIM` software suite, and thus it is important to periodically verify that `GSIM` is indeed giving an accurate representation of CLAS. Many techniques exist to achieve this end. We in the g12 rungroup elected to study the response of the `GSIM` toolchain when it is asked to process events that have already been detected by its real-life counterpart. Since these events must exist in kinematic regions accepted by CLAS, we expect that passing these events through the `GSIM` toolchain will result in 100% reconstruction efficiency, modulo tracks lost to multiple scattering or secondary interactions within the `GEANT` physics models.

The `GSIM` toolchain, for the purposes of this note, is defined as:

- `gsim_bat` for track swimming and hit-generation,
- `gpp` for adjusting photon timing,
- `a1c` running in Monte Carlo mode for event reconstruction.

Builds of `gsim_bat` and `gpp` were created using revision 1199 of the CLAS Subversion repository, then stored under `/group/clas/builds/32bit/STABLE/build/bin`. The build of `a1c` was from the `clasg12-pass1-trunk` version of the CLAS Subversion repository, in order to stay in step with production g12 Monte Carlo.

The event sample used for this study originates from one g12 raw data file, found on the Jefferson Lab Mass Storage System at this path:

```
/mss/clas/g12/data/clas_056855.A10
```

This file was chosen because 56855 was the run best-calibrated in terms of the time-of-flight, and the time-of-flight calibrator performed this study. The file was re-cooked with the version of `a1c` described above, but with all banks saved. From this output, exclusive $\gamma p \to p\pi^+\pi^-$ events with their vertex (according to the `MVRT` bank) within the g12 target volume were selected and corrected for energy lost in the target. The size of this sample was 6246 events.

2 Procedure

Consider an event sample identified and written to disk in BOS format as `feedback_orig.bos`. Refined over many variations, in its final form the feedback procedure consists of the following:

- Create a new BOS file, `feedback.part`, with two banks per event:

- TAGR sector 0, containing the energy of the trigger photon from the corresponding event in feedback_orig.bos,
- PART sector 0, containing all of the information from the PART sector 1 bank of the corresponding event in feedback_orig.bos.

In practice, the standard CLAS tool gamp2part was used along with several scripts to accomplish this goal, but one well-written program could do the job.

- Run gsim_bat with the following arguments:

```
]$ gsim_bat -ffread feedback.ffread -kine 1 -mcin feedback.part \
   -bosout feedback.gsim -trig 100000
```

The FFREAD file used here can be found in Appendix A.

- Run gpp with the following arguments:

```
]$ gpp -R56855 -P0x20 -s -ofeedback.gpp feedback.gsim
```

These arguments cause gpp to process only the tagger so that ttag and tpho in the TAGR bank can be properly adjusted so that the vertex time is 0.

- Run a1c with the following arguments:

```
]$ a1c-g12-pass1 -T4 -sa -ct1930 -cm0 -cp0 -X0 -d1 -F \
   -P0x1bff -z0,0,-90 -Aprlink_tg-90pm30.bos -ofeedback.a1c feedback.gpp
```

3 Results

As one might expect, the first pass through this procedure did not produce the desired result. Using the default g12 production Monte Carlo settings, we recovered little more than half of the events introduced. After several passes the procedure became more refined, and it became clear that issues costing us reconstructed events fell into a few categories. Some events were lost due to improper execution of this study, others to GEANT's post-vertex physics interactions, and others were lost due to actual bugs in the code. Additionally, examination of lost events was further complicated by the nature of GEANT's random number generation. Use of the RNDM card in the FFREAD file was important for reproducibility, but the sequence of events in the input file also affects the outcome as well. Extracting an event into a separate file for more detailed investigation effectively places it at a different point in GEANT's random number sequence, leading to different results. This can be allayed by examining the contents of the MCEV bank, which contains the positions in the random number sequence for each event as it is written to disk by GSIM . A single event can then be extracted, the RNDM card set to the values in the MCEV bank, and then fed into gsim_int for interactive analysis.

The best example of a mistaken conception we held in performing this study is selection of a vertex position for the input tracks. One could choose to start the tracks from their common event vertex, or from their own individual track vertices as determined by their DOCA to the beamline. After much ado, we found that the optimal solution is to allow each track its own vertex, because some events have tracks on the edge of CLAS acceptance, and when forced to originate from a common, displaced vertex their path no longer passes through the active volumes of CLAS. Also, the omission of events with common vertices outside the experimental target resulted in a substantial improvement in the reconstruction efficiency.

The vast library of post-vertex physics interactions present in GEANT, some of which are summarized in Table 1, also contributed to a loss of reconstruction efficiency. Picking the right interactions to deactivate proved challenging; for example, turning off all secondaries in GEANT also turns off hits in the drift chambers for a large portion of tracks. The best combination of cards for the purposes of this study are catalogued in the FFREAD file in Listing 1.

Finally, only once did we encounter something that could be considered a software bug. It was found in the creation of the SCRC bank of the TOF reconstruction code. The SCRC bank is made up of merged hits from the SCR bank. If two or more hits in the SCR bank are in adjacent paddles, the SCRC creation routine looks at the errors in their time and energy, and if the adjacent hits have values within those errors, it creates a new row in the SCRC bank with the times and energies of the individual paddles merged into a single hit. In real data, this merging of adjacent hits occurs in less than 3% of events. However, the standard Monte Carlo configuration of the calibration constants for the TOF is to set all uncertainties to 0. During reconstruction this leads to most SCR hits having `NaN` as the value for their errors in energy, time, and hit position. Coupled with GEANT-produced secondaries in the TOF, the SCRC bank in events reconstructed from `GSIM` is often populated with clusters with times averaged between the actual hit and the secondaries in adjacent paddles. These averaged times are always greater than the time of the hit associated with the track, so tracks that should be labelled pions are called kaons, and kaons can be shifted into the proton band. 14.5% of events passed through GSIM and reconstructed with `a1c` experienced this SCRC clustering. We corrected this problem by putting realistic values into the calibration database for the Monte Carlo run range (runs 1-10) for these TOF calibration constants:

- `SC_CALIBRATIONS_V2/atten_u/left` and `right`
- `SC_CALIBRATIONS_V2/NMIP_ADCu/left` and `right`
- `SC_CALIBRATIONS_V2/veffu/left` and `right`

So-called "realistic values" were obtained by examining the actual parameters in the calibration database for run 56855 and selecting the median value for each field. After adding these values to the database, SCRC merging occurred at a rate of 6% in the reconstruction of Monte Carlo data. This problem had not been discovered prior to this study since many rungroups use `user_ana` for event reconstruction rather than `a1c`, and `user_ana` does not use the SCRC bank.

After all of these changes were put in place, we successfully recovered 5790 of 6246 events (93% efficient). Single track efficiency for π^+ was 96.4%; for π^-, 97.1%; for protons, 98.1%. At this point we examined the distributions of the lost tracks, shown in Figure 1. Seeing no clear dependency on lab angle or physics angles, we elected to include this 7% loss as a systematic error in all g12 analyses rather than continue deeper into a regime of diminishing returns.

Table 1: A table of the GSIM FFREAD cards controlling physics interactions (courtesy of Maurik Holtrop)

Card	**# of arguments**	**Type**	**Name in code**	**Description**	**Initial value**
HADR	1	I	IHADR	hadronic processes	1
LABS	1	I	ILABS	Cerenkov light absorbtion	0
LOSS	1	I	ILOSS	energy loss	2
MULS	1	I	IMULS	multiple scattering	1
MUNU	1	I	IMUNU	muon nuclear interaction	1
PAIR	1	I	IPAIR	pair production	1
PFIS	1	I	IPFIS	photofission	0
PHOT	1	I	IPHOT	photo electric effect	1
RAYL	1	I	IRAYL	Rayleigh scattering	0
STRA	1	I	ISTRA	energy fluctuation model	0
SYNC	1	I	ISYNC	synchrotron radiation generation	0

Listing 1: `feedback.ffread`, the FFREAD file used in this study

```
CUTS   5.e-3 5.e-3 5.e-3 5.e-3 5.e-3
DCCUTS 1.e-4 1.e-4 1.e-4 1.e-4 1.e-4
ECCUTS 1.e-4 1.e-4 1.e-4 1.e-4 1.e-4
SCCUTS 1.e-4 1.e-4 1.e-4 1.e-4 1.e-4

MAGTYPE 2
MAGSCALE 0.500 0.000
PTGIFIELD 0

STTYPE 1
STZOFF -90.0
TGPOS 0. 0. 0.
TARGET 'g11a'
TGMATE 'PROT'
POSBEAM 0.0 0.0

GEOM 'ALL' 'ST'
NOGEOM 'MINI' 'PTG '
BEAM 0 0 5.744

KINE 1
MULS 0
HADR 0
DCAY 0
AUTO 1

RUNG 56855
RNDM 128 128
TIME 1000000 1000000 1000000
TRIG 1000000

STOP
```

Figure 1: Distributions of tracks lost after passing through GSIM. **Top left:** Charge times the magnitude of the momentum for lost tracks. **Top right:** Charge times θ_{lab} for lost tracks. Large-angle negative tracks could be lost at an enhanced rate since they at the rearward edge of CLAS for g12 kinematics. **Middle left:** Charge times ϕ_{lab} for lost tracks. **Middle right:** Vertex-Z distribution for events with lost tracks. **Bottom left:** $cos\theta$ in the Gottfried-Jackson frame for events with lost tracks. **Bottom right:** ϕ in the Gottfried-Jackson frame for events with lost tracks.