Assessment of SBS Software Progress and Timeline for Readiness (Preliminary)

Andrew Puckett[1] and Eric Fuchey[1]
[1] *University of Connecticut, Storrs, CT, USA, 06269*
(Dated: December 15, 2019)

I. OVERALL SOFTWARE PROGRESS AND TIMELINE FOR READINESS

The SBS software effort suffered a significant setback with the departure of Seamus Riordan from the collaboration and from the field of nuclear physics altogether. However, under the leadership of the UConn group, the software effort is recovering well and nearly back on track. The UConn group has been the leading contributor to the development and maintenance of the GEANT4 Monte Carlo simulation package *g4sbs* since 2014. *g4sbs* is a mature, stable, and relatively full-featured simulation package for all the major experiments in the SBS program. The geometry descriptions for both GMN and GEN-RP in *g4sbs* are essentially complete, and *g4sbs* has already been used to inform experiment design and planning for several years.

A. Progress since August and current status

Since the last SBS collaboration meeting in August, major, steady progress has occurred in the software development effort. In this section we summarize the recent progress and current status of the major project milestones.

1. Successful test of GEM reconstruction/tracking code with cosmics (≈ 100% complete)

The UConn group, led by professor Puckett, on the timescale of approximately 2 weeks of Professor Puckett's FTE effort, developed and tested a new GEM reconstruction, tracking, and alignment code and demonstrated the ability to do a complete analysis and performance characterization using cosmic rays of the INFN GEMs. For details, see section II of this document and the recent slideshow. The new code has been distributed to the various GEM groups and will be used in the near future to analyze the four-layer and/or five-layer cosmic data from the UVA GEM test stand at JLab. It will also be used to perform a complete analysis and performance characterization of all of the INFN GEM cosmic ray data collected since November 2018. This code is written so that it can also handle the planned "UV" GEM layers being constructed at UVA without modification of the source code. In parallel to the ongoing analysis of incoming data using the "standalone" version of these codes, the developed and tested algorithms are being written into SBS-offline, the Podd-based, "production" analysis library for the SBS program.

2. Progress in "simulation digitization" library development (≈ 90% complete)

Dr. Eric Fuchey has led the development of the "simulation digitization" library that takes the output of the Monte Carlo simulation, and produces realistic pseudo-raw data with background superimposed on top of signal events at a level appropriate to the expected luminosity for the experiment. This simulated raw data, together with the Monte Carlo "truth" information from *g4sbs*, will be used for the testing and benchmarking of the "production" event reconstruction algorithms in SBS-offline.

3. "SBS-offline" development (≈ 40% complete)

"SBS-offline" is the "production" analysis library that will be used for online and offline analysis and event reconstruction for the SBS program. The raw data decoders for both real and simulated data and the "skeleton" analysis classes for all major SBS subsystems and all GMN subsystems are essentially complete. The major remaining work consists of writing the "database" for each detector, writing the low-level "detector" reconstruction algorithms as well as the high-level "spectrometer" reconstruction algorithms, and putting together the high-level analysis "scripts" which tell the Hall A analyzer how to process the data depending on the experiment. Additionally, separate calibration scripts will likely be required for each subsystem for things like gain-matching, HV plateau, optimization of optics models, etc. Many of the required reconstruction and/or calibration algorithms already exist in specialized, "standalone" ROOT macros and other relevant codes from previous experiments, and simply need to be properly migrated into the SBS-offline framework. We note that the HCAL group is already using SBS-offline for the analysis of cosmic ray data collected in the test lab. The ingredients of the reconstruction "database" include:

1. The detector map, which consists of the mapping between physical locations of detector channels and the unique DAQ crate, slot, and channel information in the case of real data or the unique "channel index" from *g4sbs* in the case of simulated data.
2. The geometry database, which consists of all relevant geometrical parameters of the detector re-

quired for event reconstruction, including individual detector physical dimensions, positions and orientations in the local coordinate system of the detector or spectrometer of which it is a component, relevant fiducialization points and local $\leftrightarrow$ global coordinate transformations

3. The "calibration constants" database, which consists of all relevant constants required to convert raw detector information such as ADC and TDC values into physical quantities like energy deposition, event start times, particle four-momenta and trajectories, time-of-flight, vertices, etc. Examples include gain coefficients in the case of detectors which measure energy depositions such as calorimeters or scintillators, time offsets and walk correction parameters, effective propagation speeds for scintillation detectors with PMT readout at both ends, pedestals, any relevant software thresholds for zero-suppression and/or clustering algorithms, spectrometer optics model parameters, etc., etc.

B. Timeline for remaining work and readiness for beam

The major outstanding work to demonstrate readiness for beam consists of:

- **"end-to-end" simulated experiment analysis using the production SBS-offline code. Estimated completion time: April 2020.** Given the nearly complete and "production ready" status of the simulation digitization library, we will start new simulations with all detectors, magnets and magnetic fields, and beamline elements of the elastic signal, inelastic physics background, and beam-induced "soft" background for the highest Q^2 point of GMN and the GEN-RP kinematics in early January 2020. In parallel to this, we will put together the "simulation" databases for SBS-offline for both GMN and GEN-RP and begin to migrate existing reconstruction algorithms into SBS-offline. Some modest specialized code development is likely required for the unique situation of the GEN-RP analysis.

- **Demonstrate decoding of real detector data, analysis and histogramming of raw data from individual detector channels, and production of "online" diagnostic plots and histograms. Estimated completion time: July 2020**[1] We note that SBS-offline is already being used to decode and analyze data from the HCAL cosmic test stand, and will be used in the near future to analyze cosmic ray data from the UVA and/or INFN GEM test stands.

- **Collecting input from subsystem contacts on calibration and commissioning procedures with beam and run plans. Estimated completion time: March 2020**[2]

- **Demonstration and validation of calibration and commissioning procedures using simulation: Estimated completion time: August 2020**

- **End-user documentation: Estimated completion: September 2020**

- **Demonstrated software readiness for beam: Estimated completion: September 2020**

Note that in all these estimates, I am only assuming the current manpower of the UConn group dedicated to SBS software development (Puckett at ≈0.25 FTE during academic year, ≈0.8 FTE summer 2020, Fuchey at approximately 0.75 FTE dedicated to software between now and August 2020, two new Ph.D. students Datta and Seeds expected to transition to full-time research by summer 2020). Beyond summer 2020, the manpower of my group will depend on my funding situation. I fully expect other groups will eventually ramp up their contributions to the software effort, after the "base" libraries and end-user documentation are stable enough to facilitate efficient contributions from the broader collaboration.

II. GEM TRACKING STATUS

A. TreeSearch code

Until recently, the GEM analysis for Super BigBite was relying on the TreeSearch library, which performs, in this order:

- GEM decoding and hit reconstruction separately for each strip direction;
- Separate tracking for 2D projections of the track along each strip direction ("2D tracking") using the tree search algorithm;
- Combination of "2D tracks" into "3D tracks", by combining the hits belonging to the 2D tracks on basis of amplitude matching.

[1] The timeline for this milestone is dependent on the actual progress of the installation of the GMN and GEN-RP detectors in Hall A, the instrumentation of the full DAQ system, and the production of cosmic and/or other data for testing/commissioning.

[2] This milestone is dependent on cooperation from the SBS collaboration at large

This library allows to restrict the region in which to do the tracking, based on the information of the other detectors. However, this search region is applied *after* the hit reconstruction is performed. Moreover, this was not even being done for the analysis on which the "3 Hz" rate presented at several previous meetings was based. For those analyses, the TreeSearch pattern recognition was processing the entire GEM active area *before* introducing any restrictions from other detectors. Essentially, by considering the entire GEM area for clustering and hit reconstruction *and* pattern recognition/2D tracking, the analysis rate was artificially reduced by a large factor compared to how fast it can be in principle. In the case of G_M^n, with the background conditions expected (100-120 kHz/cm^2 with beamline shielding), the decoding and hit reconstruction was responsible for more than 80% of the total analysis time.

B. New tracking code

In the production analysis code, the reconstruction of BigBite will start with the preshower and shower calorimeters. Because it is a total absorption calorimeter with a relatively high threshold and good energy resolution, the reconstruction of the cluster corresponding to the primary electron track will be relatively clean, as experience has shown with similar detectors in even higher-luminosity environments. For example, the GEp-III BigCal was used successfully for electron detection in Hall C at luminosities up to twice as high as those expected in GMN, and that calorimeter had no magnet in front. The logic of the event reconstruction in BigBite is as follows:

- The BigBite shower cluster defines a region of approximately 7.5×7.5 cm^2 (at the 3σ level) within which to search for hits in the fifth GEM layer located between the GRINCH detector and the preshower.

- Based on the expected occupancy of the fifth GEM layer, we expect, on average, a 65% probability of one or more background hits overlapping the region defined by the shower cluster. Most of the time, we will have 0, 1, or 2 background hits in this region of the fifth GEM layer in addition to the primary hit from the good electron track.

- The combination of one precise coordinate measurement from a low-occupancy GEM layer and the rough energy measurement with approximately 8% resolution from the BigBite shower calorimeter provides a powerful constraint on the properties of the good electron track. Given the optics of the BigBite dipole magnet, the thickness of the GMN target and the narrow horizontal acceptance of the BigBite magnet, the combination of the shower cluster and the fifth GEM layer allows us to restrict the search region in each plane of the front INFN GEMs to approximately 5×5 cm^2, or roughly 0.4% of the total area of each plane. This restriction requires no assumptions about the location of the interaction vertex or any requirement that the electron track satisfy elastic or quasi-elastic kinematics.

- On average, within this restricted search region at each plane, we expect 0.9 background hits per plane per event in addition to the primary hits from the good electron track. This means that most of the time, we will have 0, 1, or 2 background hits per plane in our allowed search region. The combinatorics associated with such multiplicities are very forgiving, and tracking with such powerful pre-existing constraints can even be done using "brute force" combinatorics without the need to resort to specialized algorithms like TreeSearch.

- Naively, based on the ratio of search areas alone, the analysis will be sped up by a factor of approximately 240 compared to the "3 Hz" rate that has been presented before based on doing things in the wrong order. In reality, the gain in analysis rate will likely be even greater than this, since the combinatorics and computation time scale very nonlinearly with the effective occupancy/background hit multiplicity within the effective search region. As such, the "3 Hz" rate reported in the past should not be taken seriously as a baseline for the reconstruction rate.

- Realistically, one can conservatively expect 1 kHz analysis rate or more in BigBite. In summary, this means that the time spent reconstructing hits may be slashed by at least a factor of 100, simply by reordering the procedure.

Since modifying TreeSearch to make it fit the SBS GEM analysis would alter it too much, we will write a new library for the Hall A analyzer based on an already-existing 2D clustering and tracking code used successfully for the analysis of real cosmic ray data from the INFN GEMs, with the following broad procedure:

- Decoding of raw data

- Restriction of the search region for hit reconstruction based on constraints from external detectors

- Direct bi-projection ("2D") hit reconstruction, properly enforcing amplitude and time correlation;

- Direct use of "2D hits" for 3D tracking (using a "Kalman filter-like" pattern recognition and track finding algorithm).

The G_M^n BigBite analysis will include restriction of the search region from BigBite calorimeter, the GRINCH, and the clean signal on the UVA GEM plane (about 35Khz/cm^2) sandwiched between the GRINCH and the calorimeter. A slideshow with more details of the new 2D clustering and tracking algorithm and the latest results achieved in the INFN cosmic ray analysis can be found here.