

# Software Readiness for 2019 Data

---

Norman Graf (SLAC)  
HPS Software Meeting  
September 16, 2020

---

# Issues

- We have promised our funding agencies that we will conduct a “timely” analysis of the 2019 data.
  - It’s been over a year since the run ended
- We have another data taking run coming up
  - Run is scheduled in less than a year
- What software is needed and when is it needed to accomplish our goals?
  - Simulation
  - Reconstruction
  - Analysis (not addressed today)
    - Assume 2016 analysis frameworks will form basis for 2019.

# Simulation Software I

- We will need large samples of MC in order to fully understand our detector performance and provide input to the physics analyses.
- hps-mc fairly mature, working at both Jlab and SLAC
  - Investigating use on OSG grid resources
- Event generation
  - Can we run MadGraph inline in hps-mc or do we need to generate/distribute stdhep files?
- Beam backgrounds
  - Investigating use of min-bias data overlay
  - If feasible, can these files be made available on the grid via xrootd?

# Simulation Software II

- Can we use current geometry or do we need to wait for the final alignment?
  - ditto for target position
- Will we generate one “2019” sample or will we need run-specific samples?
- Output MC files are bloated with extraneous secondaries
  - Needs work to understand and prune
- Critical need for biasing in detector simulation
  - Preferential WAB photon conversion in silicon layers
  - Preferential charge track scattering in silicon layers

---

# MC Reconstruction Software

- Is current trigger simulation sufficient?
  - Can we run multiple trigger simulations at once?
- Will we process with one “2019” set of conditions or will we need run-specific processing?
- How will we handle dead/hot channels?
- Clear need for communication between software, MC and analysis groups.

# Data Reconstruction Software

- ECal finishing up gains, sampling fractions and timing.
- Hodoscope software OK?
- SVT APV25 waveform fitting
  - Is the current fitting sufficient for our track timing?
  - replacing simplex with migrad improves fitting, gives uncertainties, but takes more time.
    - Need to study this ASAP, as we plan to drop raw data from output.
- SVT trigger phase needs to be fixed for certain runs.
- SVT actively working on alignment/calibration
  - PF has either ported or provided bindings to the C++ version of GBL code used to impose constraints on the alignment.
- Tracking group actively improving CPU performance
  - PF has replaced lcsim matrices and vectors with ejml
  - Robert actively developing Kalman Filter
  - Need characterization and performance evaluation
- Need a 2019 Event Flag Filter to remove obviously bad events
  - skip “monster” SVT events, wrong SVT position, wrong SVT voltage, etc.
- Output lcio files are bloated with extraneous data.
  - Remove extraneous Drivers
  - Need to prune our data tree and remove unnecessary collections from lcio output
- Memory footprint needs to be below 1GB to be efficient at JLab.

# Logistics

- We need good estimates of our CPU needs to process the full 2019 “good” data sample
  - ~50 Billion events
- We need good estimates of the amount of computing power we can rely on.
  - will be competing with CLAS for processing resources
- We need good estimates of our storage needs
  - ~600TB of evio data
  - Will tape access be an issue?
- Will inform the overall HPS data processing plan
  - e.g. do we start MC generation with our existing detector geometry?
  - e.g. do we “pre-process” the SVT data now while we wait for recon improvements?