Analysis Development

Paul King, Kent Paschke

Analysis development timescale

Analysis tools for the Integrating and Counting data streams must be “fully” developed for commissioning, characterizing and operating the MOLLER apparatus by the time we start operations

This summer: need functioning integrating analysis for beam studies

Between summer ’23-December ’24: Beam studies for beam and monitors

Phase 2 Experimental Readiness Review (likely summer ’24): need a baseline design, credible schedule, demonstrable on-schedule progress

Run 1 (early ’26): prepared for commissioning, calibration, characterization, and operation by late ’25

Counting / Tracking Analysis

- optics calibration
- point back to backgrounds sources
- detector alignment, overlap, response model (position, angle)
- rate profile, confirm yield simulation models
- demonstrate optics model, confirm acceptance model and analyzing power calibration
- background measurement, identification
  - neutral (non-track-like) hits for photon backgrounds
  - broad acceptance tracking (background search)

Proposed baseline software: Build from the PREX and SBS Podd analyzers. SBS has had a lot of GEM analysis development. Basic framework and starting functionality (zero-field tracking) will already exist

For development:

helpful to make G4 simulation ⟶ mock data for tracking and analysis development

JAPAN for integrating analysis

Just Another Parity ANalyzer was developed out of the Qweak analysis framework, and was used for the PREX-II and CREX experiments.

- Decode event data, apply calibrations and normalizations, and calculate composite quantities (e.g., BPM positions)
- Apply event-level cuts. Use recent history of events to apply beam trip cuts and stability cuts
- Use data exchange classes to calculate new composite quantities (e.g., position corrected event yields, event-level correlation matrices)
- Collect the events for a complete pattern; calculate the yield and asymmetries for all quantities. Blinding is applied during the asymmetry calculation.
- Use data exchange classes to calculate new composite quantities (e.g., position corrected asymmetries, asymmetry correlation matrices).
- ROOT histogram and tree outputs can be selected at event-level, pattern-level, or minirun-level
- If there have been sufficient patterns accumulated to complete a minirun, calculate and store the average quantities and rms.

Second-pass through the output data used to perform beam corrections

- Regression (from multiplets), Dithering from “raw” yield outputs

First Steps

Mock-data generator within JAPAN creates realistic data files with time-dependent and randomized generation of beam parameters and detector signals, including correlations

Analysis of these data files allows testing of throughput and processing algorithms

- Processing time for initial mock-data (216 PMT channels and 50 beamline channels) is ~5ms per event (this is faster than we had expected from scaling arguments but only tested on short data segments)
- This summer: Paul has undergrad, will focus on mock data with focus on producing a large mock data set and more complete JAPAN throughput test
- This December: Project requires specifications for analysis workstations, so we need a pretty good idea of required resources and throughput expectation

This summer: injector beam tests

- Can provide some base data for analysis optimization
- Actual 2kHz beam data, perhaps with full muliplet-64s if helicity board is ready

Priorities

JAPAN throughput should be examined

- We should aim to understand if the underlying structure is suitable. Obviously online monitoring would be complicated by the 1/10th real time analysis rate.
- Profiling and optimization should provide throughput gains, as this wasn't carefully explored for large channel counts before PREX-2. The key question is whether significant architectural changes are needed.

Mock data or beam test data may be an important resource to test design decisions. Some examples:

- Convenient 20Gb file-split size, about ~150 seconds. A convenient unit for a "minirun"? Plan needed for efficient "extent cut" execution across boundaries.
- With 64plets, we will have Yield and Asym (or Diff). What about RMS? Max, Min?
- With new ADCS: individual windows will have RMS, Max, Min. How to make use of these (e.g. error flagging, accumulate distributions for miniruns, output at low duty cycle)?
- What sort of correlations or moments to output? (X / Y dipole, open/transition/closed, etc.)

Design decisions are needed alongside development

Aggregation / Reporting

- Past experience shows that aggregation is time and effort expensive
- There are very many important combinations and moments, multiple detector systems.
- As with all steps of analysis, must avoid overburdening storage, processing, or human cognition
- Flexibility for new outputs will be needed
- “aggregation mock data” may itself be valuable here (e.g. simulate minirun results from models, figure out what is needed to catch problems)

Thoughtful choices and design here will be important

Online monitoring, Feedback

- These need rapid (at least real time) processing.
- Can be lightened (some components dropped, or low duty cycle, etc)
- Data structure, data persistence, interface all need design thought (reminder: PREX used “bounded” root file, panguin display)

Summary

- Natural place to start
  - Integrating: evaluate throughput (and start performance improvement) for JAPAN
  - Counting: Adapt SBS / Podd. Develop remoll mock data capability.
- Design decisions are needed: on data output, aggregation, monitoring. Hard to think about these without a prototype approach.
- We should restart the parity analysis group discussions.
  Would need an interested cohort, and some applied person-power.

backups

Integrating Analysis components

Mock data generator

Experimental condition database

Feedback / Online Monitor

- EPICS communication
- feedback logging
- reference histograms
- automated evaluation and/or alerts
- alert and/or histogram logging

Aggregation / Reporting

- daily compilations
- grand averaging
- Lagrange multiplier analysis and diagnostics

Documentation

Prompt / Offline

- data reduction (multiplet, + n seconds, minirun)
- Hardware streams
  - MD sectors
  - SAM / LAM / Auxs
  - Showermax
  - Pion Detectors
  - Combinations / moment building
- correlation analysis
  - grand correlation matrices
- dithering calibration
  - phase advance diagnostic
- beam corrections

Analysis tools

- pedestal calibrations
- pita
- relative bandwidth pulse measurements

Blinding

Rings 2-6,
Showermax

$$\delta_B^M \equiv \delta_{APV}^B(90\,\mathrm{deg}) \qquad \delta_{B,i}^M = \frac{\langle A_{PV}^M \rangle_i}{A_{PV}^M(90°)} \qquad \phi_i^M = \frac{Y_i^M}{\Sigma Y_i^{tot}}$$

$$\delta_i^B = \delta_{B,i}^M \phi_i^M + \delta_{B,i}^{ep} \phi_i^{ep}$$

Choose both δ's from flat boxes at 10σ where σ is from total extraction (i.e. half ppb on Moller) with independent randomization

We will make a documented proposal for this scheme and post in docdb, present in a meeting.

Everyone should think about what they will try to do in the analysis… does this get in the way?

Or… how they could engineer a way to extract δ_B from the data!