Compute Resource Estimates

- **Status**: Current estimates on computing resources not reliable:
    - For accurate estimates, a prototype for reconstruction of physics events from 1ms time slices is needed.
        - **Details on next slide.**
        - **Question:** Do we use time frames or slices as our nomenclature?
    - Need for reliable estimates on the fraction of background events and their impact on reconstruction.
    - Understanding speed at which these events can be discarded is crucial.
    - Importance of defining alignment and calibration methods for each subsystem.
    - Detailed discussions on fast alignment and calibration techniques are necessary.
- **Goal**: Reliable compute resource estimates before TDR.

INTERLUDE: Prototype for Frame-Based Reconstruction

For the TDR, we limit the scope of the first study to the track reconstruction only:

Correlate hits in a realistic time frame to the various events in the time window of the MAPS.

We need to accomplish the following tasks (15 FTE months):

- **Framework Tests and Development**:
  - **01/24** Integrate Jana2's built-in workflow for supporting frames in and events out in EICrecon.
- **Simulation**:
  - **12/23** Prepare simulation productions, using detailed information on FEEs for tracking detectors, utilizing the full, wide MAPS integration window for tracking purposes.
  - **01/24** Implement and utilize the frame-building infrastructure post-Geant4 and post-digitization.
- **Reconstruction Process**:
  - **02/2024** Adapt the reconstruction process to work with frames, making it frames-aware.
  - **03/2024** Demonstrate tracking from realistic frames
  - **04/2024** Deliver first estimate of reconstruction time from frames.

Scale of Computing

Based on the number of electronic channels in the detector, and the occupancy you expect, what is the expected frame size?

- The frame size (1ms) based on our current detector readout design is 10MB when running at peak luminosity and in standard operating conditions.

How many events do we expect to record and simulate per year, respectively?

- Assuming a 50% up-time for ½ year, we will record 15.5 billion frames in a year.
    - **Note**: Should we use a *snowmass* year instead (running 1/3 of the year, 10^7s)?
- The event rate at peak luminosity is 500kHz, which gives roughly 4 x 10^{12} events (60% background, 40% bunch crossing related):
    - This will of course be much lower at start of operations, where the luminosity will be lower (but relatively speaking the background rate is expected to be higher).
    - The actual physics events related to key EIC observables is only a very small fraction of the total physics bunch crossings. The expected number of DIS events / physics event of interest for one year of running at peak luminosity is ~ 10^{10}. This is the number that drives our simulation needs, and we expect to simulate 10x events for each event of interest, yielding O(10^{11}) simulated events. While considerable (~ 60k core years on today's hardware), this should be a realistic target in a decade.

How many core-seconds on a typical modern machine does our reconstruction and simulation take today, respectively?

- Our current simulations of background embedded events take ~17s for simulation and ~ 2s for reconstruction, per event.

How much do you think this will change once we add a realistic frame to event reconstruction?

- We do not currently have an estimate for how event reconstruction time will change once signal events are embedded in realistic frames. For the TDR we expect to have first estimates based on the initial implementation of frames in the reconstruction.

Discussion

Input from the Electronics and DAQ and Streaming Computing Model WGs:

- What is our best knowledge of the 1ms time frames?
  - Data Model Definition?
  - Data Format?
  - What is the time structure of the various event types?
- How close are we to put a 1ms time frame together?
  - Who could provide that?
- Could we document the above in a brief technical note?
  - Who would lead the effort on that?