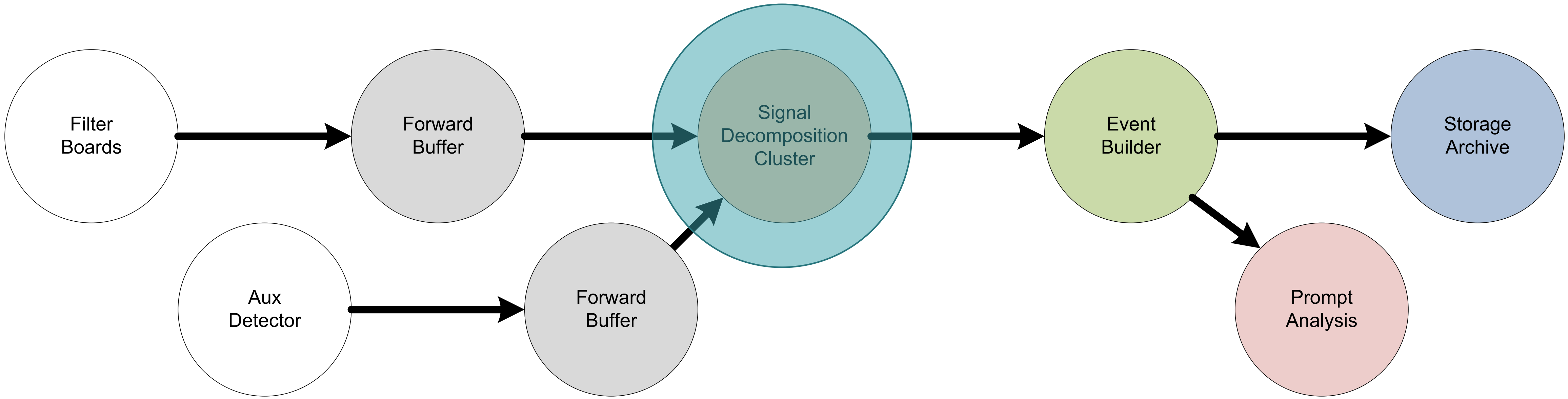
GRETA Signal Decomposition Status

Gustav Jansen, David Radford

Third AGATA-GRETA MEETING, ANL

October 2019

Decomposition



Goal:

Transform time-dependent electrical signals from all segments to number, position, and energy information of interactions

GRETA Requirements

- Requirements
  - 480k events per second (15k per second per crystal)
  - Sets the scale of the computing cluster
- Resource estimate
  - 10 ms per event per core (2016)
  - 5000 cpu cores
- Notes
  - Estimate made on old hardware
  - Estimate uses a single core per event

Strategy

- Determine best fit by minimizing the χ^2 difference between actual and simulated signal for each event
- Different minimization methods
  - Grid search: Brute force search on a coarse grid
  - Non-linear least square: Modified gradient descent algorithm
- Different types of signals
  - Single segment: Test for one, two, and three interactions
  - Multiple segments: Test for one or more interactions in each segment
  - Significant increase of computational complexity for multiple interactions

Update of GRETINA Code

- Refactoring of old GRETINA code for GRETA
- Prototyping of new GRETA system
  - Decouple signal decomposition from I/O, slow control, and monitoring subsystems
    - Experimenting with the algorithm is easier if all components are self-contained
  - Optimize for new processors with high core count
  - Investigate the use of GPUs in the signal decomposition cluster
  - Determine computing resources in production system
  - Investigate the use of libraries in minimization algorithm

Prototype Status

- Refactoring
  - Overall design completed
  - Interface to event builder and forward buffer is in progress
  - Almost ready to have signal decomposition functionality in development environment
- Library usage
  - Current matrix inverse implementation is faster than tested libraries for relevant matrix sizes
  - Will investigate minimization libraries
    - Depends on refactoring the algorithm code itself

CPU vs. GPU

1. CPU only: Decomposition algorithm is implemented on a standard CPU (GRETINA implementation)
2. Matrix offload: Matrix-inversion operations in the algorithm are offloaded to a GPU
3. GPU-based grid search: The grid-search part of the algorithm is implemented on the GPU
4. GPU only: Both the grid-search and non-linear least-squares parts are implemented on the GPU

CPU vs. GPU

1. CPU only
2. Matrix offload
   - Minimal development efforts
   - Matrices are too small for GPU matrix inversion
   - Multiple simultaneous inversions?
     - More complex code - possible better performance
   - Only 14 % of runtime is spent in matrix inversion

CPU vs. GPU

1. CPU only
2. Matrix offload
3. GPU-based grid search
   - Minimal development effort
   - Significant speedup compared to CPU only
     - Will need to run full node tests to evaluate cluster performance
   - Can use a finer grid and add optimization parameter (t0)
     - Leads to a less costly non-linear least square
   - Currently only 9% of total runtime is spent in grid search
     - More if finer grid and additional optimization parameters

CPU vs. GPU

1. CPU only
2. Matrix offload
3. GPU-based grid search
4. GPU only
   - Gradient descent uses 88% of runtime
   - Limited parallelism in algorithm
   - Will require massive rewrite of code
     - Is it possible to run multiple minimizations simultaneously?
   - Difficult to maintain as low-level code is necessary for optimal performance

CPU vs. GPU

Conclusion:

GPU usage is possible for an expanded grid search algorithm. It has the potential for significant speedup and can enable a better signal decomposition algorithm.

Next Steps

- Integrate signal decomposition with the GRETA development system
- Implement and validate new grid-search algorithm with additional parameters (CPU and GPU versions)
- Investigate the use of minimization libraries
- Optimize detector model code
  - Currently 75% of runtime is spent on evaluating the detector model given a set of interactions

Field Calculation and Cross-Talk Fitting

Additional improvements are planned or underway for auxiliary codes

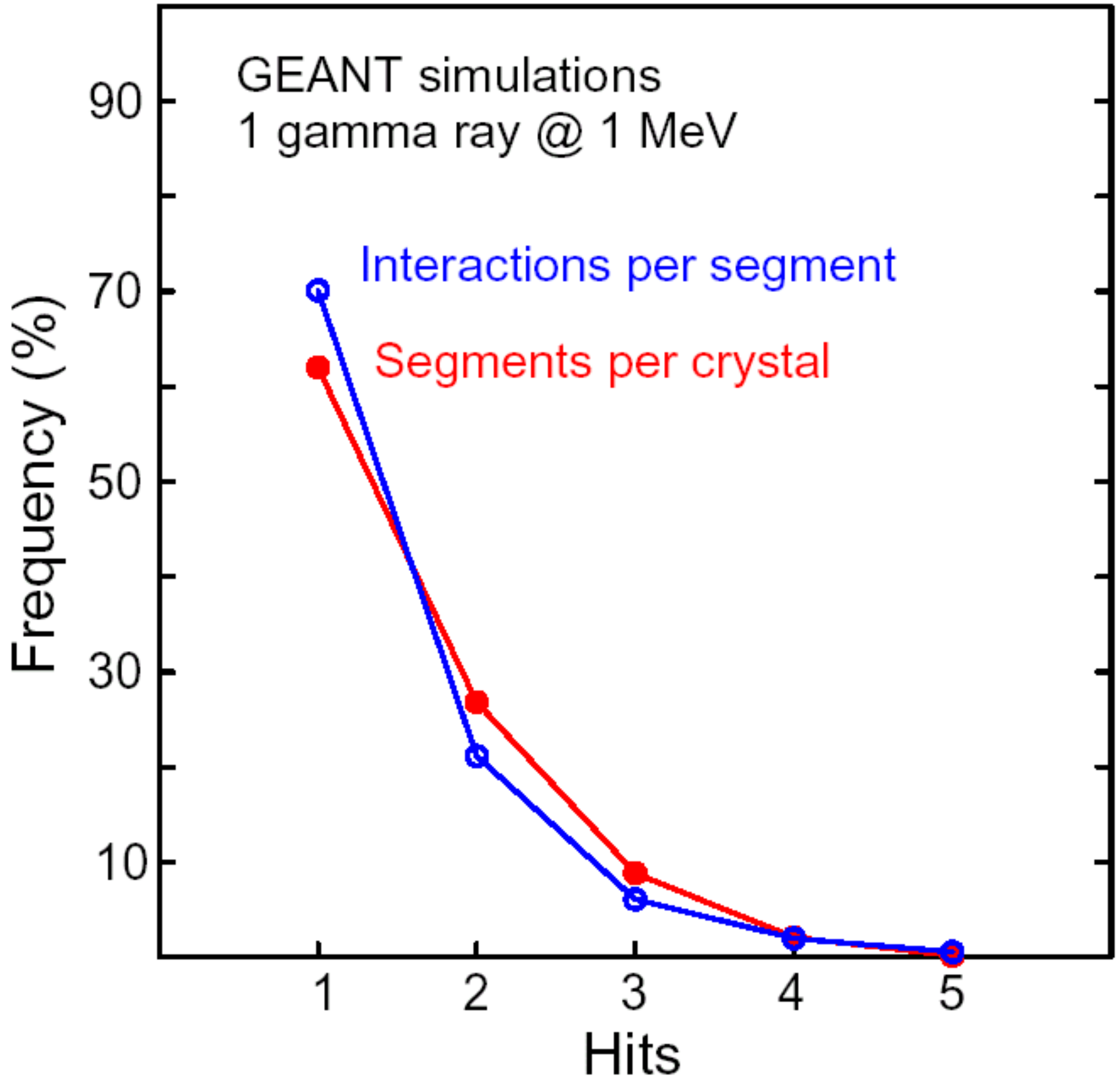
- `fieldgen`
  - Adapt improvements from MAJORANA
  - Significant improvement in speed
  - Allows for smaller grid sizes
- `xtalk`:
  - Investigate the use of libraries in minimization algorithm

Extras

Expected Distribution of Hits

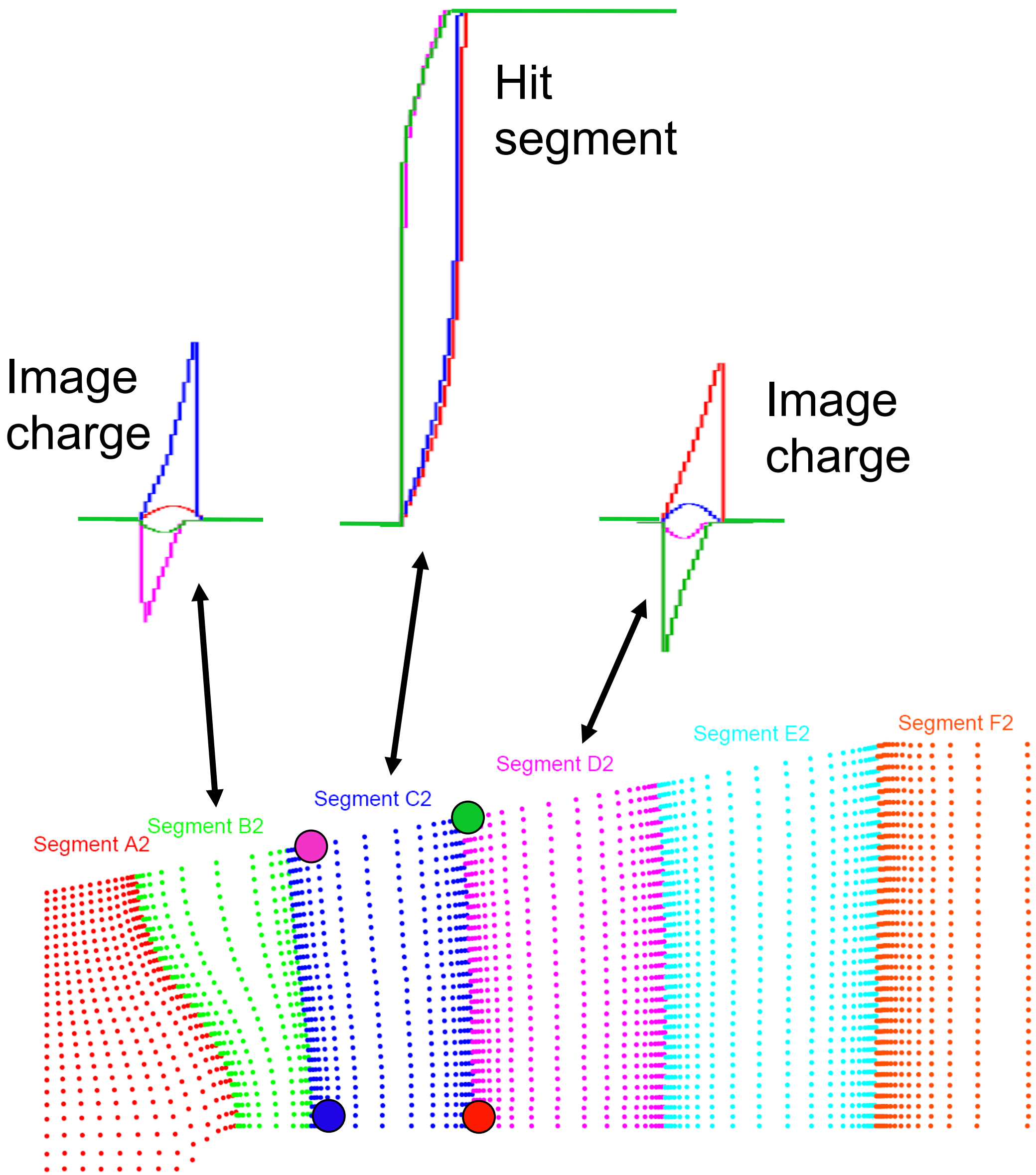
GEANT simulations
1-MeV gamma into GRETA

- Most hit crystals have one or two hit segments
- Most hit segments have one or two interactions



Position Sensitivity of Signals

- Calculated signals (without preamp response)
- Color-coded for position of interaction (energy deposition)



Performance: Example Fits

- Concatenated signals from 36 segments, 500ns time range
- Typical multi-segment events (red)
- Linear combination of basis signals fitted by decomposition algorithm (blue)