AGATA Week 2019

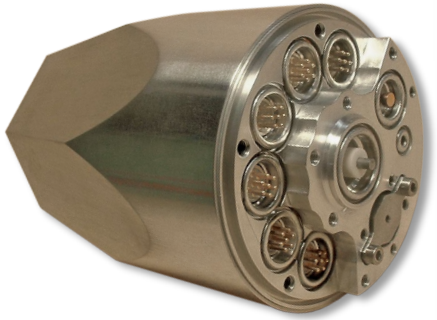
Machine Learning and Topological Data Analysis for Pulse Shape Analysis

Fraser Holloway

UNIVERSITY OF LIVERPOOL
LIV.DAT
Science & Technology Facilities Council
AGATA ADVANCED GAMMA TRACKING ARRAY

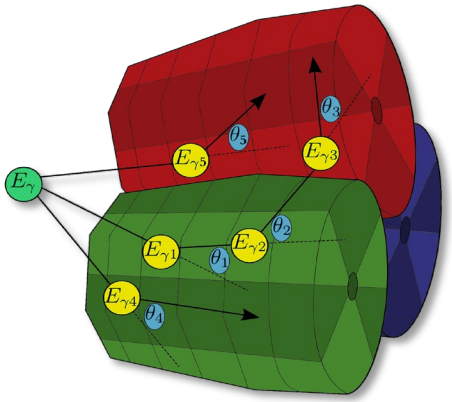
Pulse Shape Analysis (PSA)

- γ-ray tracking requires positions at resolution ~5mm FWHM at ~5kHz/CPU.
- Positions must be inferred from electrical response (PSA).
- Complex detector response makes parametric methods insufficient.
- Instead we simulate the detector response in ADL.
- Interaction locations are then determined by optimisation metrics:



$$Figure\ of\ Merit = \sum_{j}\sum_{t_i}\left|A_m^j\,[t_i] - A_s^j[t_i]\right|^p$$

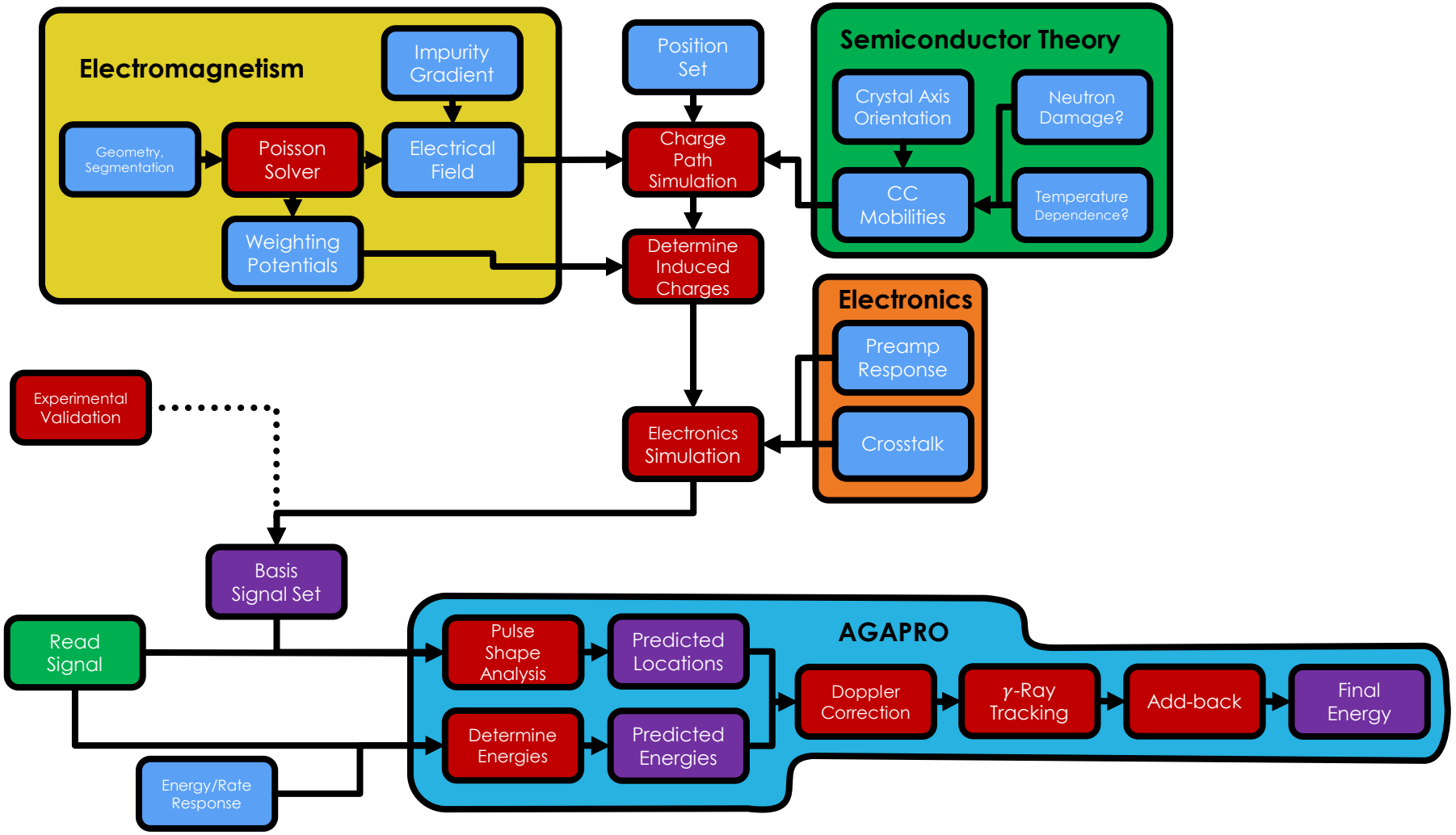
For signals of segment j at time step t_i with p typically =2



- Other metrics can be used to highlight different sensitivities.
  - Different exponents, weighting for segments.

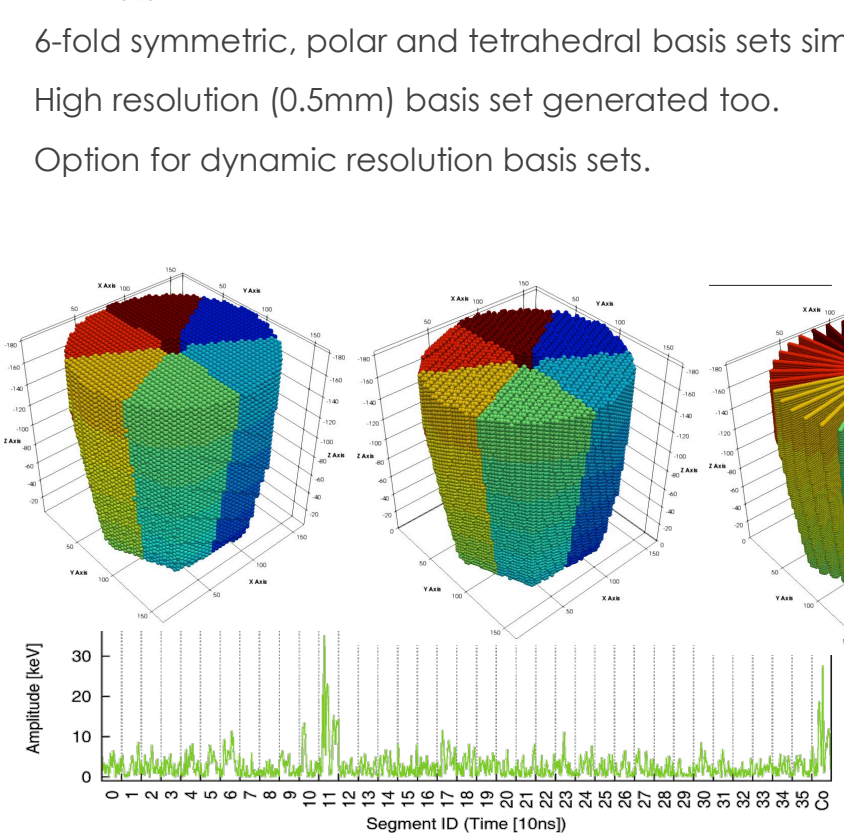
- My work is on developing Novel PSA techniques for AGATA.

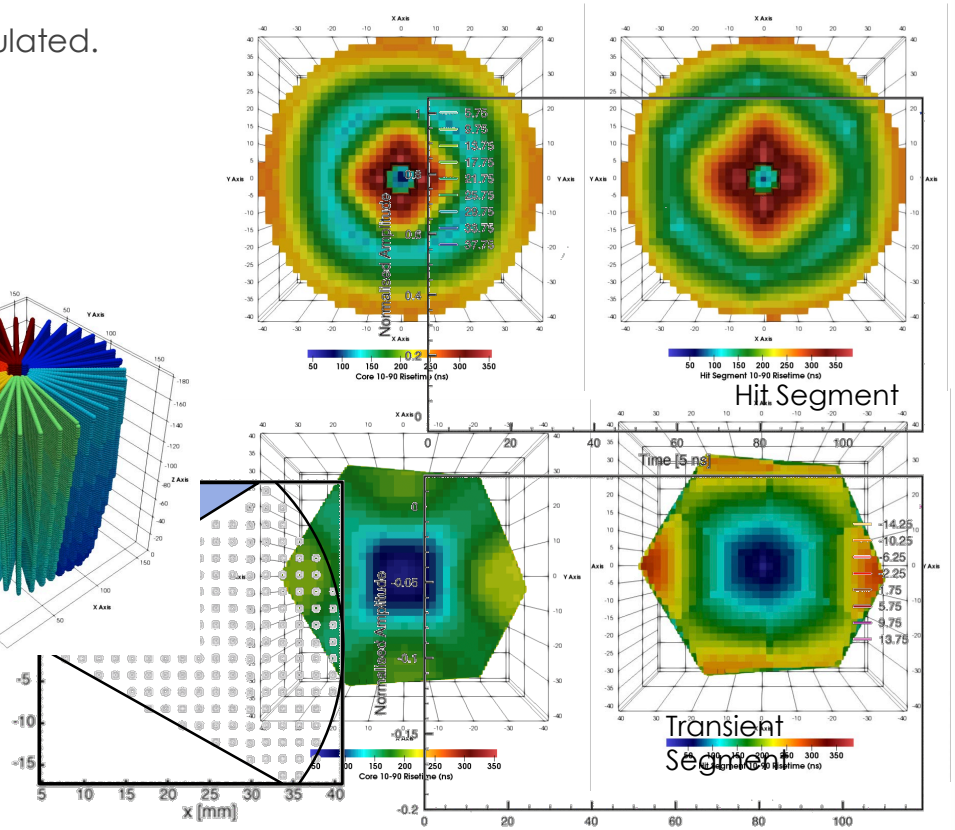
Fraser Holloway - F.Holloway@liverpool.ac.uk


Electromagnetism
Impurity Gradient
Geometry, Segmentation
Poisson Solver
Electrical Field
Weighting Potentials
Position Set
Charge Path Simulation
Determine Induced Charges
Semiconductor Theory
Crystal Axis Orientation
Neutron Damage?
CC Mobilities
Temperature Dependence?
Electronics
Preamp Response
Crosstalk
Electronics Simulation
Experimental Validation
Basis Signal Set
Read Signal
Energy/Rate Response
AGAPRO
Pulse Shape Analysis
Predicted Locations
Determine Energies
Predicted Energies
Doppler Correction
γ-Ray Tracking
Add-back
Final Energy

Detector Simulation

- Simulated data looks reasonable as expected.
- Parametric trends are seen in the data, useful for clustering
  - $T_{10\text{-}90}$, charge asymmetry, knee-point, skewness etc.
- 6-fold symmetric, polar and tetrahedral basis sets simulated.
- High resolution (0.5mm) basis set generated too.
- Option for dynamic resolution basis sets.



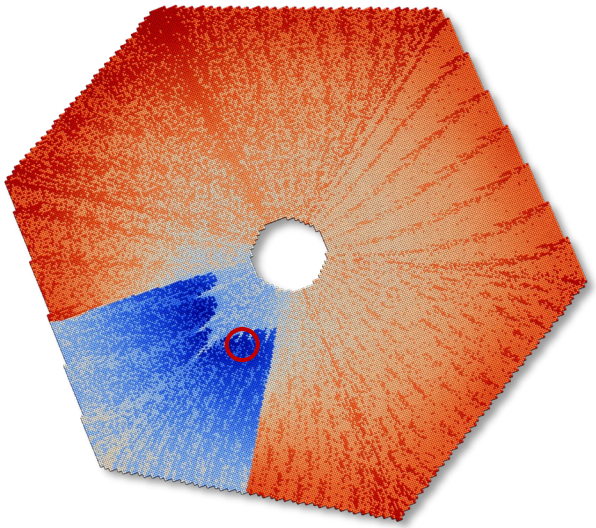


B.Bruyneel – Eur. Phys. J. A (2016)

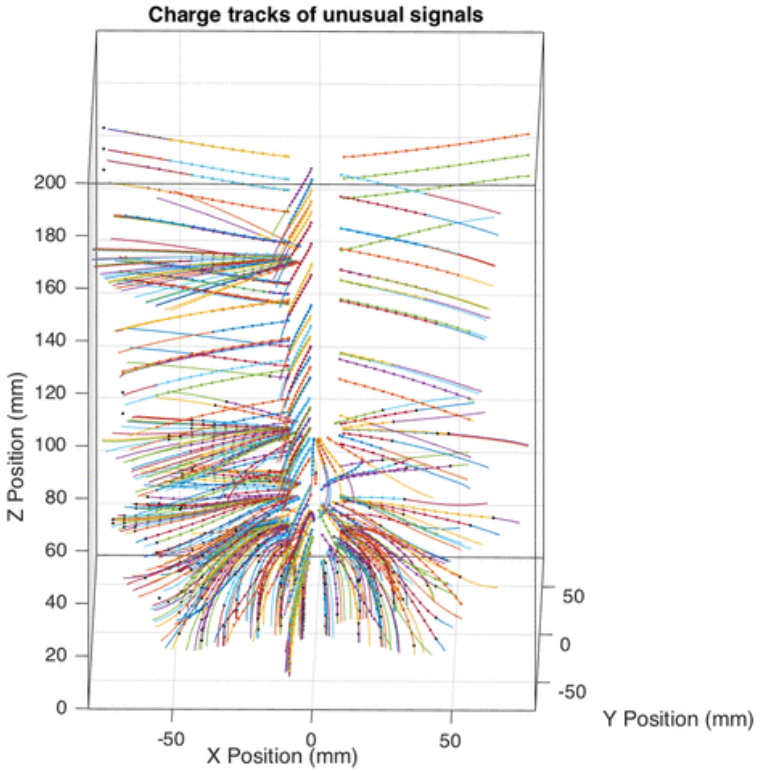
Fraser Holloway - F.Holloway@liverpool.ac.uk

Simulation Limitations

- Field simulation limited to 1mm spacing
- Odd effects seen at segment boundaries & high resolution
  - Unexplained charge sharing between segments
  - Sharp discontinuities at edge changes.



0.5mm FoM Plot showing odd effects

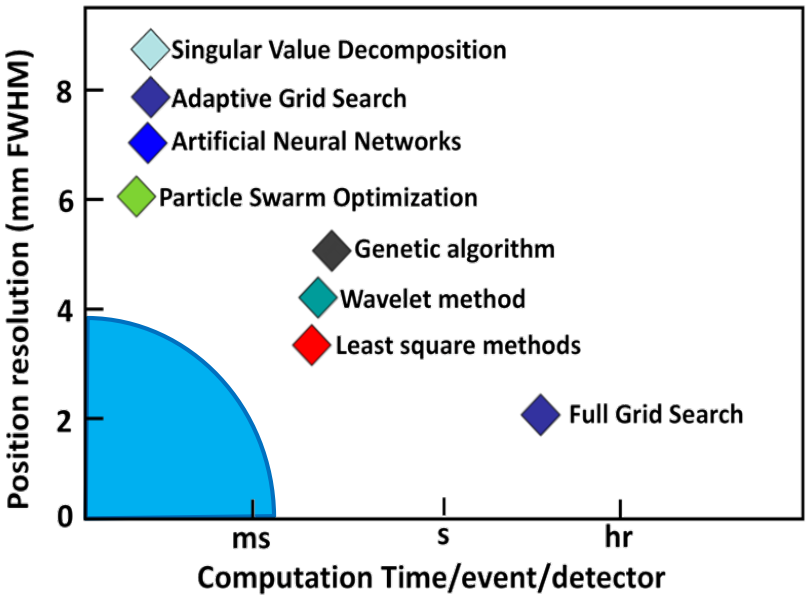


Fraser Holloway – F.Holloway@Liverpool.ac.uk

Algorithm Development

Topological Data Analysis (TDA) techniques try to organize data and form efficient search spaces.

- Generally kd-ball or cover trees used.
- Less prone to local minima.
- Search algorithms aren't naïve.
- Each step made moves search closer to optimum.
- Searching n points can be $\mathcal{O}\log(n)$.



Machine Learning (ML) uses the simulated basis to learn trends via feature extraction.

- No searching is performed whatsoever.
- Simulated basis only needed for training.
- Needs an appropriate model & good data.

Fraser Holloway - F.Holloway@liverpool.ac.uk

Algorithm Development

Tree-based search approaches:

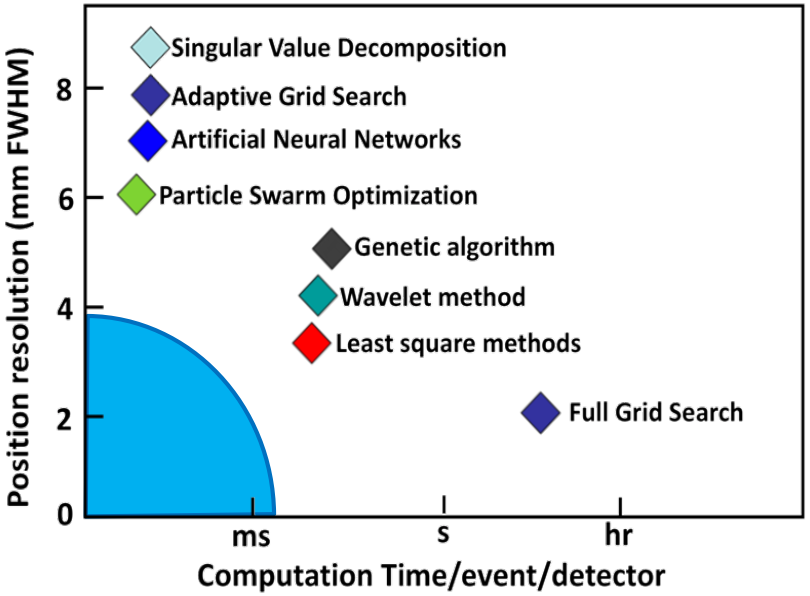
- kNN
- LSH
- Dual-Tree / Single-Tree MKS

Machine Learning options:

- Signal Classification
- Regression (CNN)
- Autoencoding/Fingerprinting (BVAE)

Other options:

- GPU Acceleration



Fraser Holloway - F.Holloway@liverpool.ac.uk

GPU Acceleration

- GPUs have advanced significantly (10x) since the last investigation.
- GPU acceleration can be used on embarrassingly parallel problems:
  - Exhaustive search
  - Adaptive Grid search (two step)
  - Matrix manipulations
    - Figure of merit (although matrix sum $\mathcal{O}\log_2(n)$)
- Shared memory makes things complicated
- Multiple languages can use GPU accelerated code:
  - C, C++ (NVCC)
  - Python (with Numba)
- Programs can be compiled to use NVBLAS:
  - MLPACK (Armadillo)

Nvidia P5000
(277 GFLOPS)

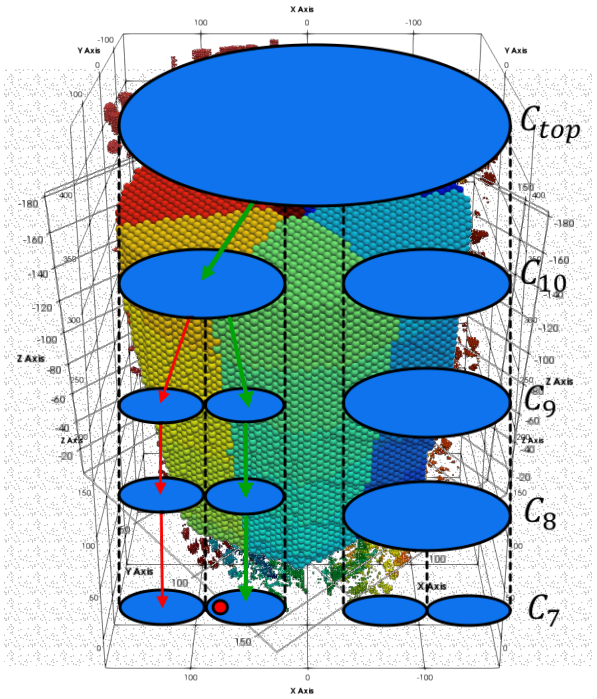
- GPUs are **very** powerful for ML approaches.

Routine	Types	Operation
GEMM	S, D, C, Z	Multiplication of 2 matrices.
SYRK	S, D, C, Z	Symmetric rank-k update
HERk	C, Z	Hermitian rank-k update
SYR2k	S, D, C, Z	Symmetric rank-2k update
HER2k	C, Z	Hermitian rank-2k update
TRSM	S, D, C, Z	Triangular solve (right angled)
TRMM	S, D, C, Z	Triangular matrix-matrix multiply
SYMM	S, D, C, Z	Symmetric matrix-matrix multiply
HEMM	C, Z	Hermitian matrix-matrix multiply

Fraser Holloway - F.Holloway@liverpool.ac.uk

Cluster Optimisation & Tree Building

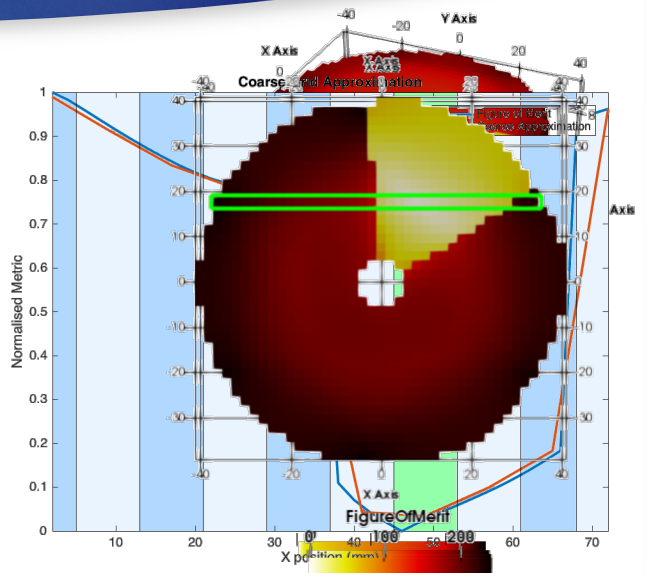
- Basis was converted to Cover Tree using parametric splitting.
  - Segment number → $T_{10\text{-}90}$ → Charge asymmetry → Transient Signal Fingerprint → FoM
- Resolution of metrics inversely related to execution time.
- FoM test only applied at lowest level.
- Most Parametric methods break down at higher multiplicity.
  - Either use fold-invariant metrics or add all combinations to tree.
- Cluster distribution shows evidence of variable detector sensitivity.

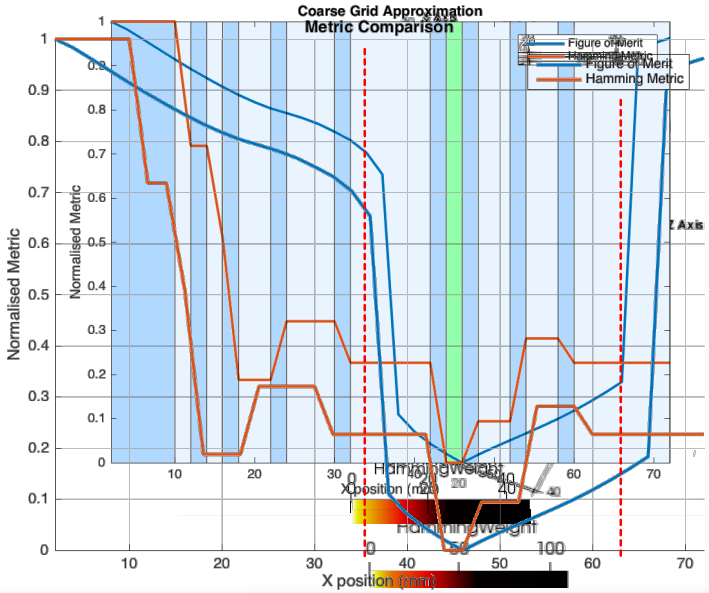


Fraser Holloway - F.Holloway@liverpool.ac.uk

Cluster Optimisation & Tree Building

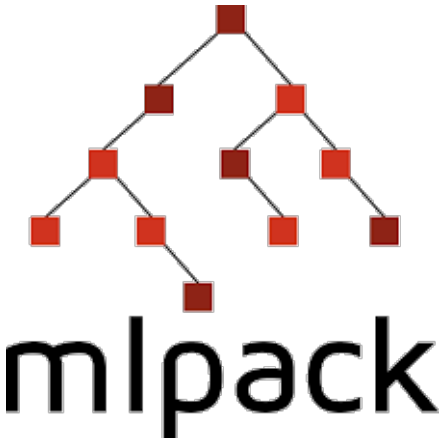
- Basis was converted to Cover Tree using parametric splitting.
  - Segment number → T_{10-90} → Charge asymmetry → Transient Signal Fingerprint → FoM
- Resolution of metrics inversely related to execution time.
- FoM test only applied at lowest level.
- Most Parametric methods break down at higher multiplicity.
  - Either use fold-invariant metrics or add all combinations to tree
- Cluster distribution shows evidence of variable detector sensitivity.





Fraser Holloway - F.Holloway@liverpool.ac.uk

Automated TDA Searching



- Established C++ Library MLPACK used for KNN & MKS operations
- GPU acceleration possible using NVBLAS
- Additional Python API & Command line interfaces available
- Modular design allows for custom Figures of Merit, segment handling
- Prefers smooth & convex search spaces
  - Doesn't like searching multiple segments
    - Metric penalizes segments far from interaction
- *Should* work for multiple interactions within the same segment
  - Combinations need to be precomputed
    - Outrageous memory costs if implemented
- Currently 3 techniques look applicable to Fold-1 searches:
  - kNN
  - LSH
  - **MKS**

Fraser Holloway - F.Holloway@liverpool.ac.uk

FastMKS Searching

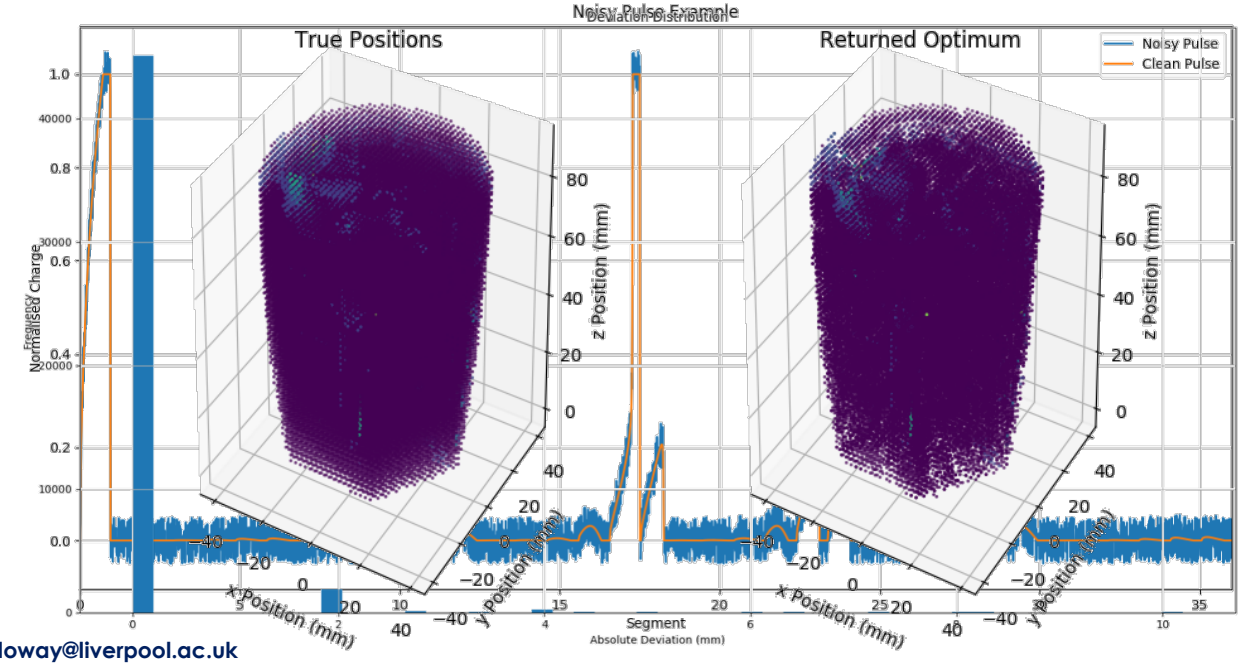
- Fast Maximum Kernel Search uses two trees to search an ordered data structure.
- First tree is used to convert reference set into structured data.
- Second tree is then dynamically built using query set.

- Efficient comparisons mean that the space can be searched quickly.
- Mercer Kernels allow for modifications of phase space, improve separations.
  - More complex kernels have execution penalty.

Fraser Holloway - F.Holloway@liverpool.ac.uk

Fast-MKS Preliminary Results

- 10% Gaussian noise added to simulated database for preliminary validation.
- MKS with Gaussian kernel used to return top 5 solutions of kernel search.
- 95% of fold-1 events identified at input location.
- 99% of fold-1 events within 2mm.
- Discrete distances due to finite grid size.
- Currently deviations are not well understood, needs further analysis.



Fraser Holloway - F.Holloway@liverpool.ac.uk

Signal Discrimination with ML

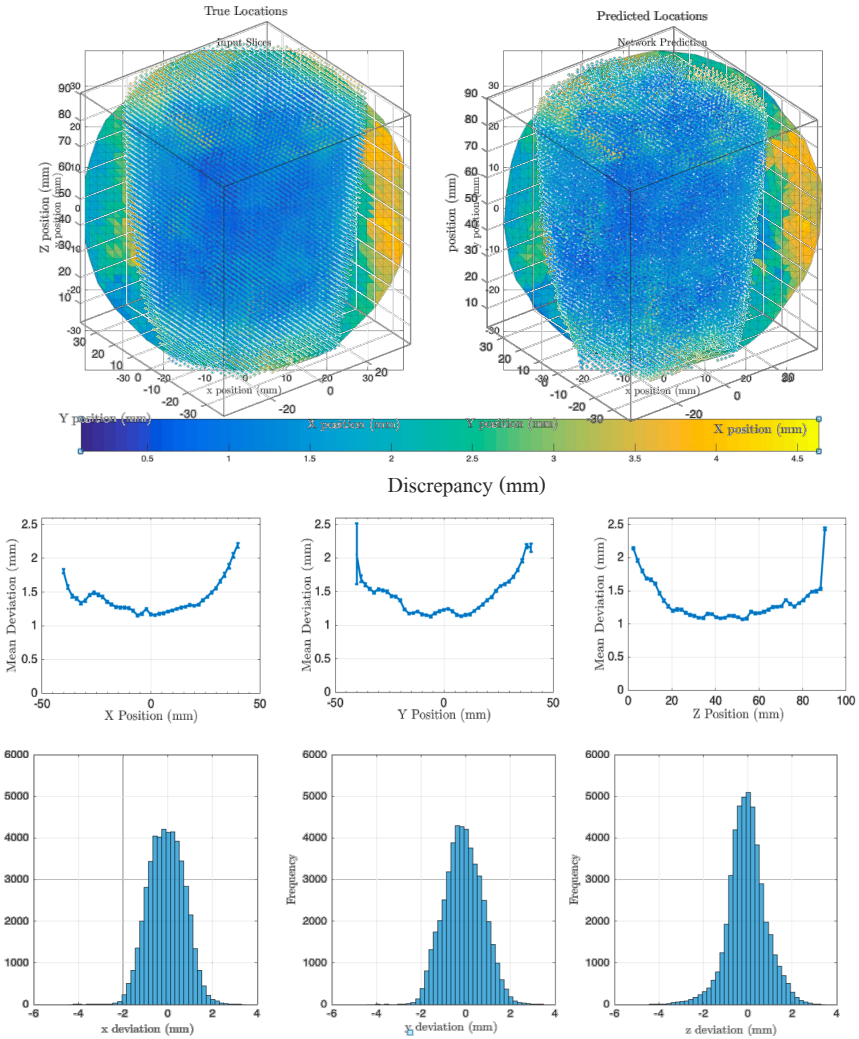
- Main motivation of this method was to distinguish interesting sections of the interaction from noise.
  - Possible groundwork for software-based trigger.
  - Because of this these networks need to be fast (and likely simple).
- Position gated pulses used to generate database of hit, transient & noise samples.
- Various networks trained to predict category.
- Ultimately the cut is arbitrary, open to interpretation.
- Doesn't offer much above traditional methods.
  - However if we really want to look for something it's pretty useful.

Method	Agreement with Midas Label	Execution Time (μs)
Multi-Level Perceptron	~68%	9
Binary Perceptron	~87%	9
Neural Network	~94%	22
Convolution Neural Network	~97.6%	26

- Additional investigation into predicting fold of AGATA events, not mentioned in this talk.

CNNs for Regression

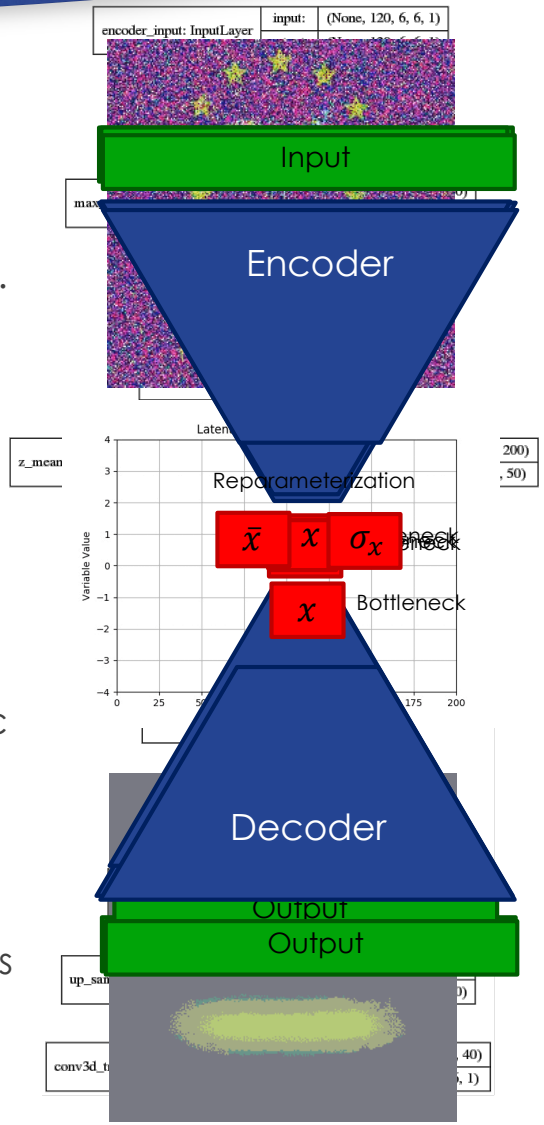
- Instead of searching basis set the neural network is used to directly infer locations.
- Trained on 6x8x120 tensor (core excluded).
  - Column repeats used for CNN windows.
- ResNet architecture used for robustness.
- Gaussian noise layers & Dropouts used to improve reliability.
  - Should use experimental noise instead.
- Works well on detectors with high connectivity.
- Currently only implemented for fold-1 events.
  - Training on multi-fold requires separate networks.
  - This isn't difficult, I'm just waiting for an accurate simulation of multiple fold events.
- Reasonable execution time ~300μs.
- Variable FWHM, performs worse at boundaries.
  - Will likely decrease with realistic data.



Fraser Holloway - F.Holloway@liverpool.ac.uk

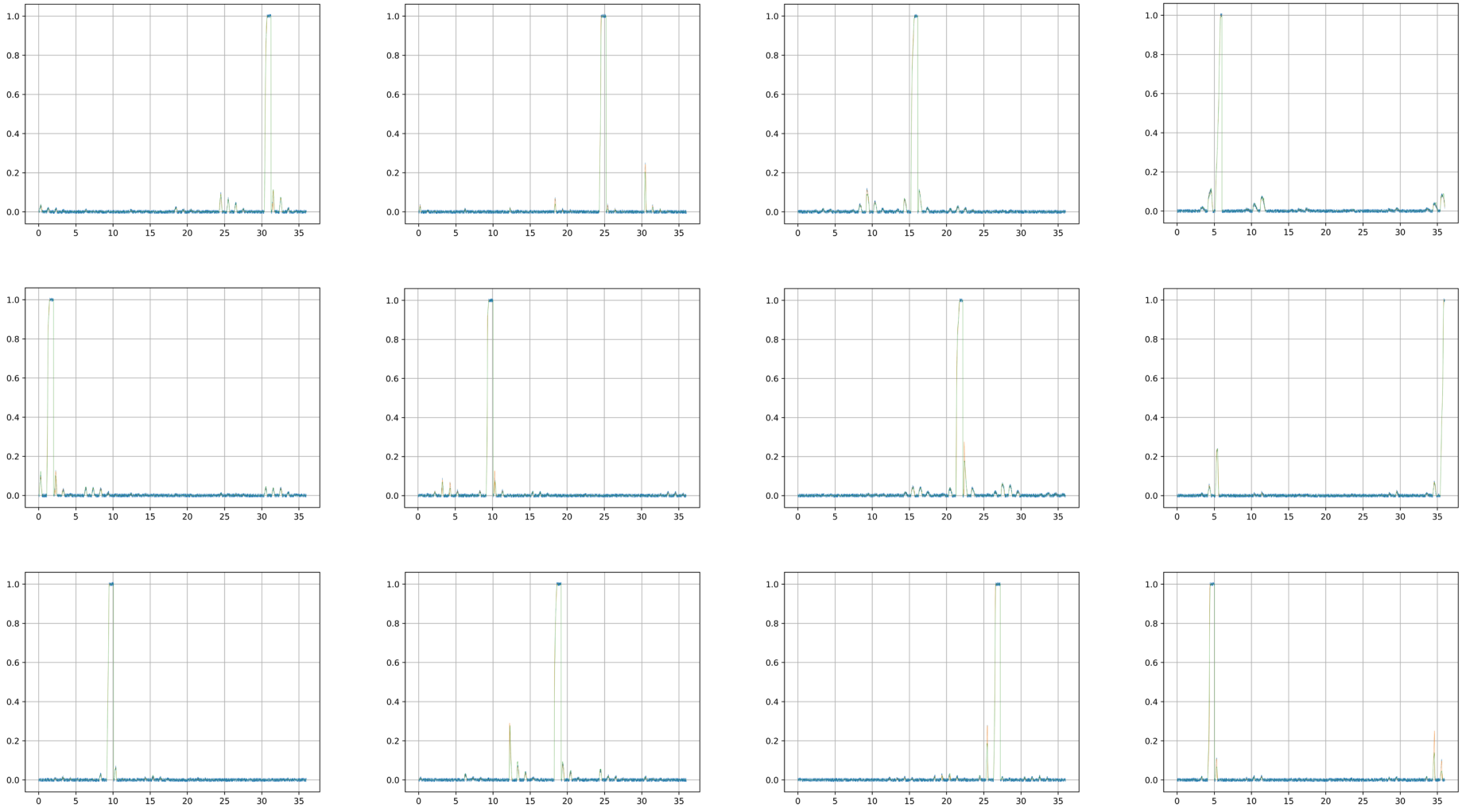
Autoencoders for Tagging & Compression

- Autoencoders combine two separate networks to function:
  - Encoder: converts input to a learned latent space via feature extraction.
  - Decoder: converts latent space into a reconstructed output.
- As a whole the network replicates a denoised input.
  - Signal is intelligently denoised, small transients are unaffected.
- Autoencoders become more useful when split into parts:
  - The Encoder and Decoder compress data far better than traditional methods.
  - The latent representation can be used to express parametric trends.
    - This requires disentangling the latent space (difficult)
    - Can this be used for tagging?
- Compression isn't necessarily bad, oddly the reconstructed pulses could end up being better than the inputs due to denoising.

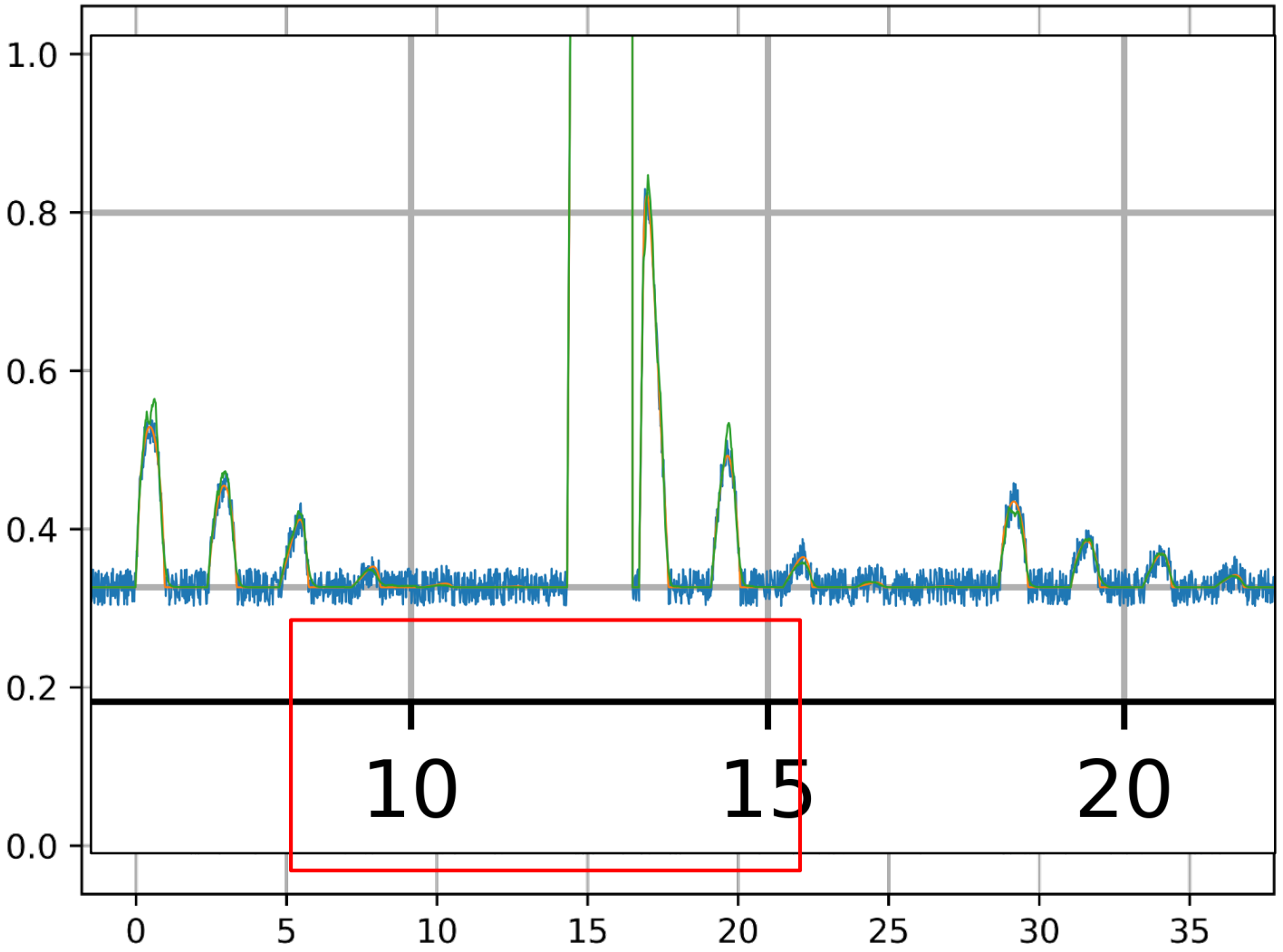


Example Reconstructions, ~44x compression

Randomly selected from reconstructed database



Example Reconstructions, ~44x compression

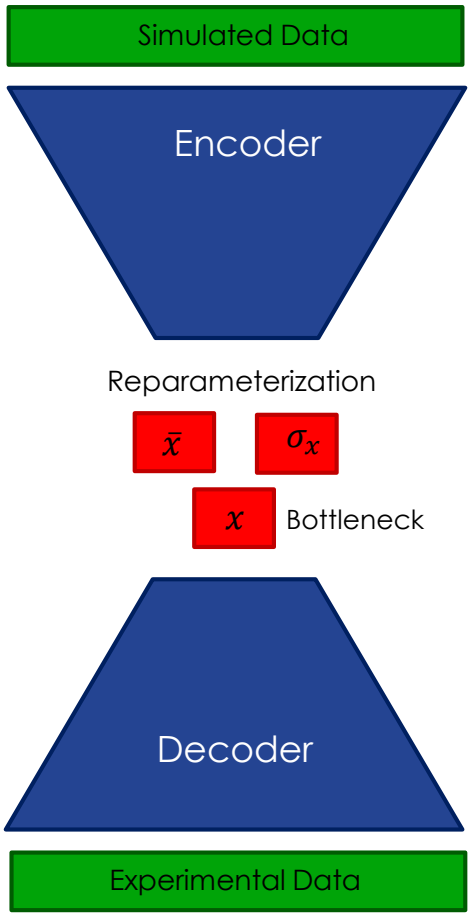


Autoencoders for Basis Correction

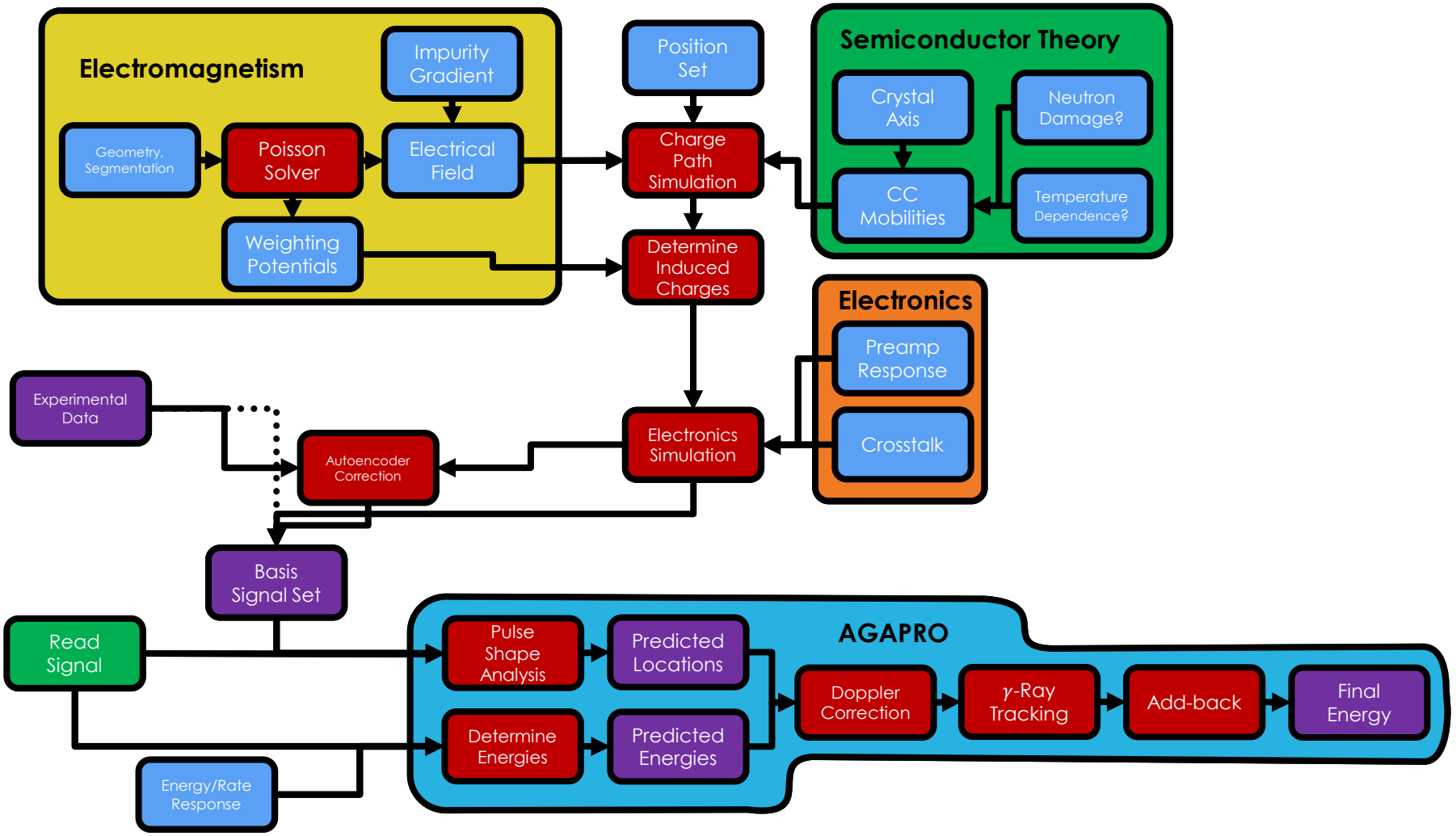
- What happens if we use experimental data as validation?
- PSA and GRT perform differently when given real & simulated data.
- Therefore there's likely some form of discrepancy between the two.
- How about using ML to transform simulated into real data?
- Simulation reduced to latent & then converted to experimental.

- This approach requires very good experimental data:
  - Full x, y, z characterisation of the crystal.
  - No guarantee that trained model can be adapted to different crystals.

- Validation data for A005 will be taken anyways.
  - May as well test the feasibility of this method.
- Transform of preamplifier response also possible.
  - ***Way*** easier



Fraser Holloway - F.Holloway@liverpool.ac.uk

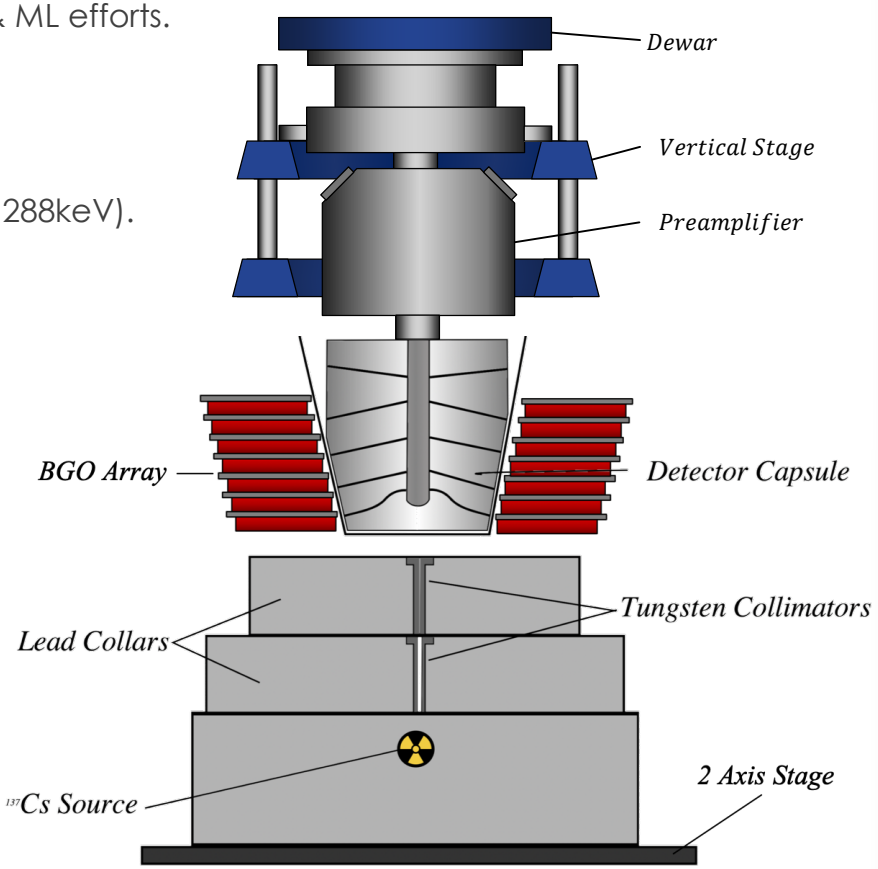

Electromagnetism
Impurity Gradient
Geometry, Segmentation
Poisson Solver
Electrical Field
Weighting Potentials
Position Set
Charge Path Simulation
Determine Induced Charges
Semiconductor Theory
Crystal Axis
Neutron Damage?
CC Mobilities
Temperature Dependence?
Electronics
Preamp Response
Crosstalk
Electronics Simulation
Experimental Data
Autoencoder Correction
Basis Signal Set
Read Signal
Energy/Rate Response
AGAPRO
Pulse Shape Analysis
Predicted Locations
Determine Energies
Predicted Energies
Doppler Correction
γ-Ray Tracking
Add-back
Final Energy

Experimental Validation

- Coincidence scanning will be used to validate simulation & ML efforts.
- 1GBq ^{137}Cs source collimated to 1mm on x, y stage.
- Vertical stage added to apparatus for z movements.
- 90° scatter gating using BGO array & energy gating (374 & 288keV).

- I'm currently writing the MTSort code for acquisition.

- Typical validation measurements will be taken:
  - ^{241}Am surface scan for alignment.
  - Gated cross & circle measurements for CAO.
  - Gated coarse cubic grid using vertical stage.
  - High-resolution pencil beam of front segmentation.
  - (Time permitting) Automated High-resolution scan.



Coincidence scanning apparatus found at Liverpool

Fraser Holloway - F.Holloway@liverpool.ac.uk

Conclusion

- GPUs have advanced significantly over the last decade, likely to continue in the future.
  - Definitely should be revisited considering future projections.
- Tree-based search methods are incredibly efficient but difficult to adapt to high fold.
  - Very applicable for Fold-1 regardless.
- ML approaches offer good learned relationships but need adaptions to high fold.
- We have a good standing for more ambitious ML techniques.
  - Discrimination
  - Regression
  - Auto-tagging / Fingerprinting
  - Compression
  - Basis Correction
- I can't take all these methods to completion

Fraser Holloway - F.Holloway@liverpool.ac.uk

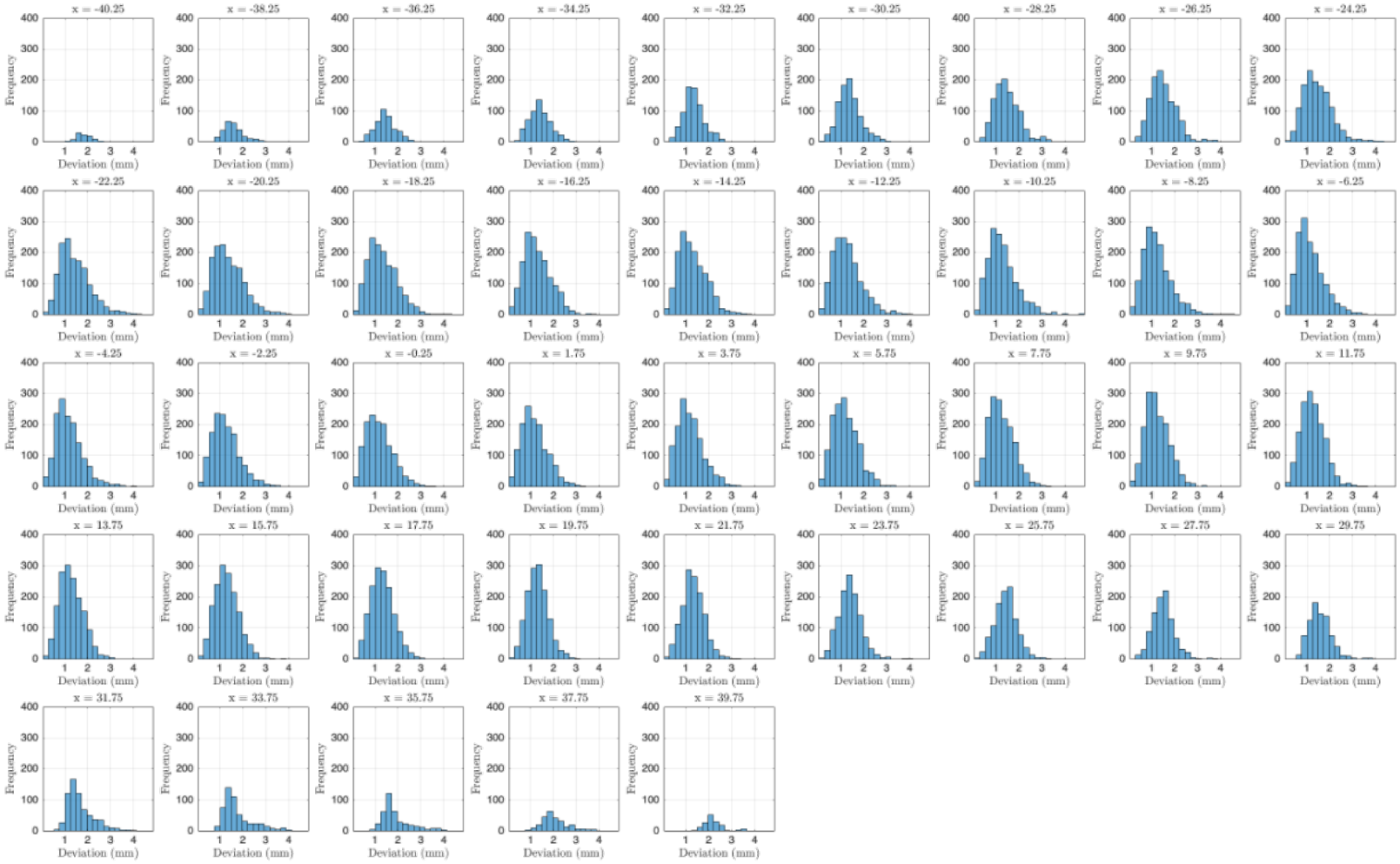
Thanks for Listening

Any Questions?

Fraser Holloway – F.Holloway@Liverpool.ac.uk

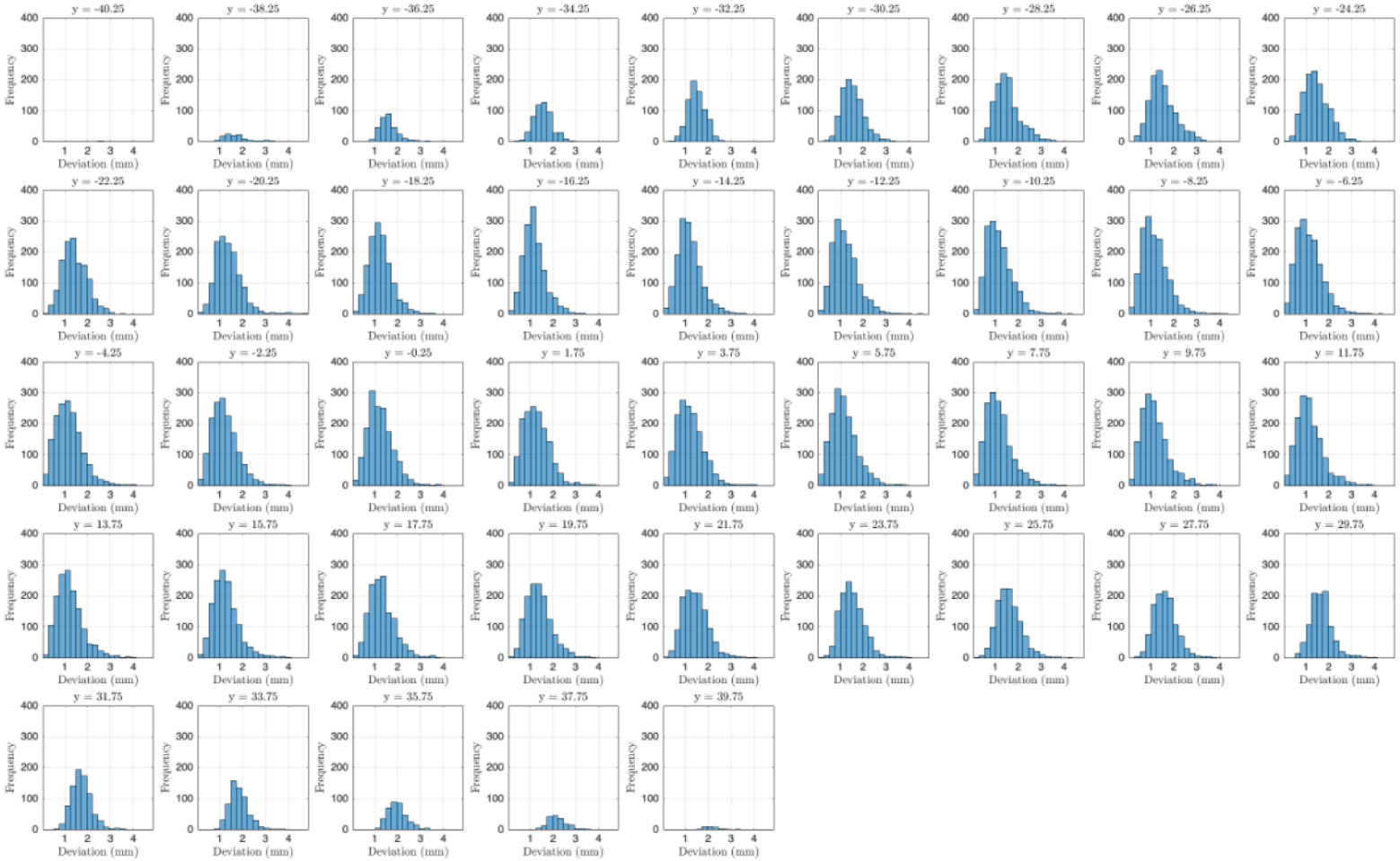
This project has received funding from STFC under grant reference ST/P006752/1

CNN x Deviations



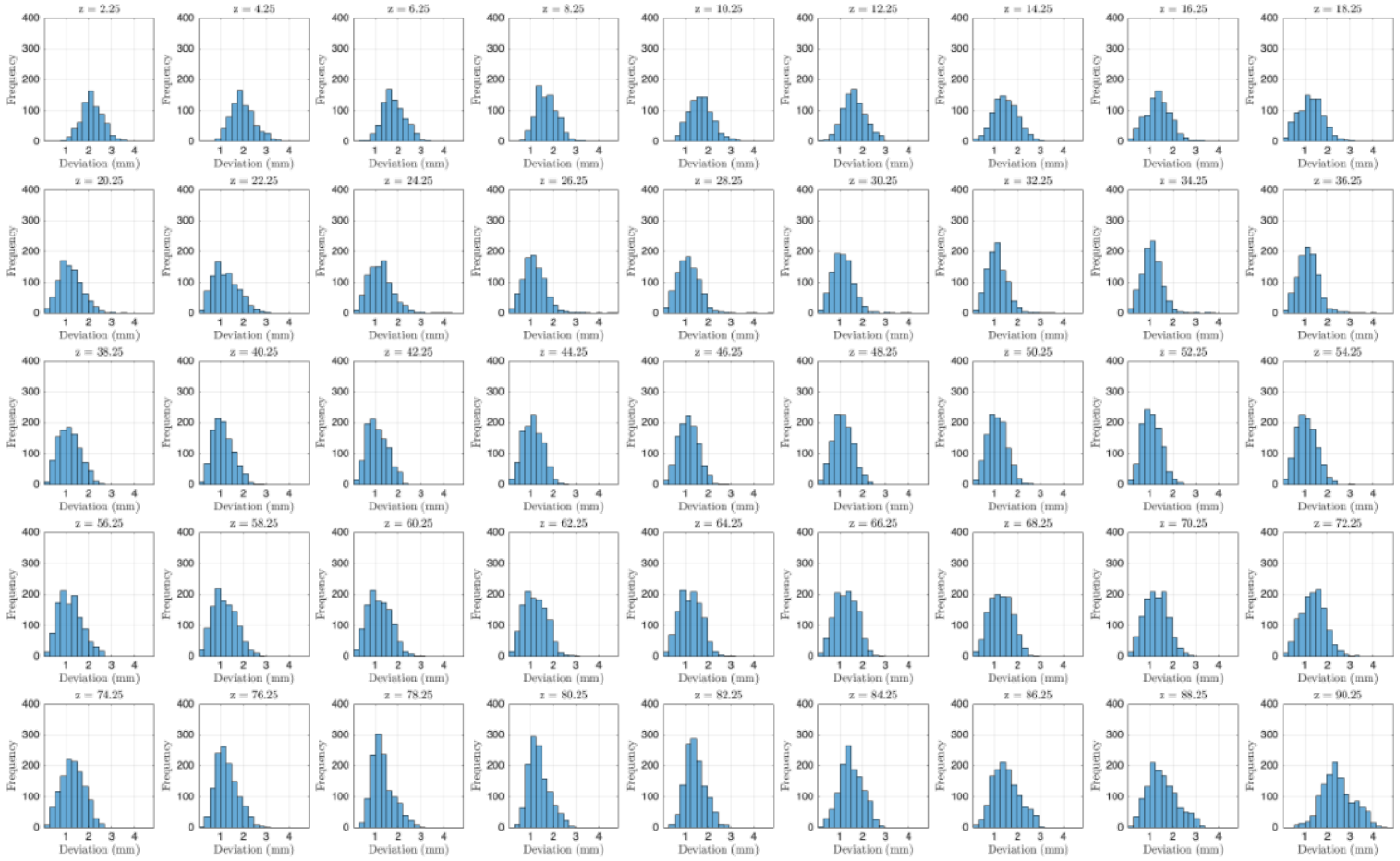
Fraser Holloway - F.Holloway@liverpool.ac.uk

CNN y Deviations



Fraser Holloway - F.Holloway@liverpool.ac.uk

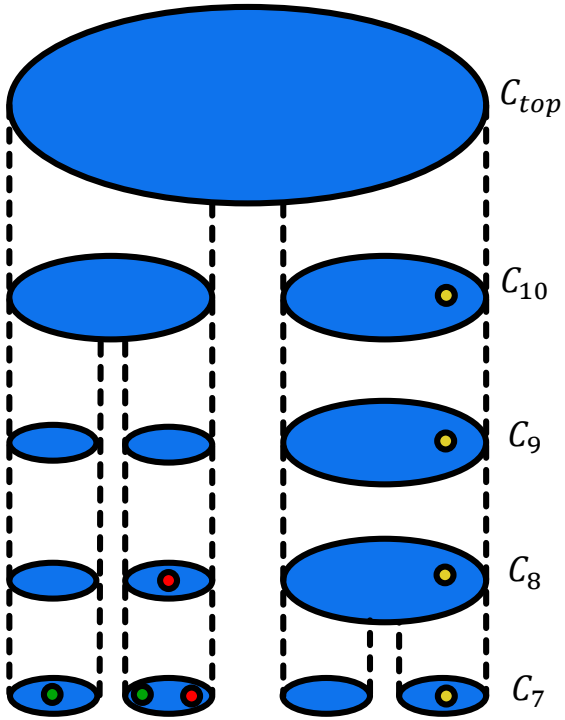
CNN z Deviations



Fraser Holloway - F.Holloway@liverpool.ac.uk

Cover Tree Rules

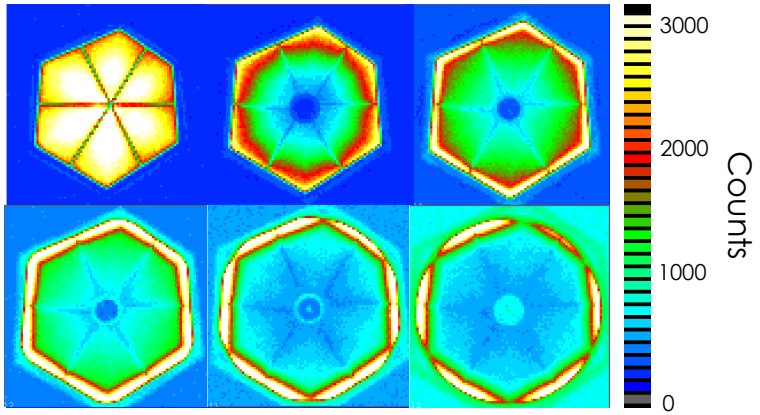
- For a collection of points C_i on level i of T which represent a subset of points in S the following rules must be enforced:
  - $C_i \subset C_{i-1}$ - Nesting: any point $p \in S$ that exists in C_i must have an associated node in all lower levels.
  - $\forall_p \in C_{i-1}$ - Covering: for every $p \in C_{i-1}$ there exists one $q \in C_i$ such that $d(p,q) \leq 2^i$ where the node for q is the sole parent of the node for p.
  - $\forall_p, r \in C_i, d(p,r) > 2^i$ – Separation: For all $p, r \in C_i$ then $d(p,r) > 2^i$

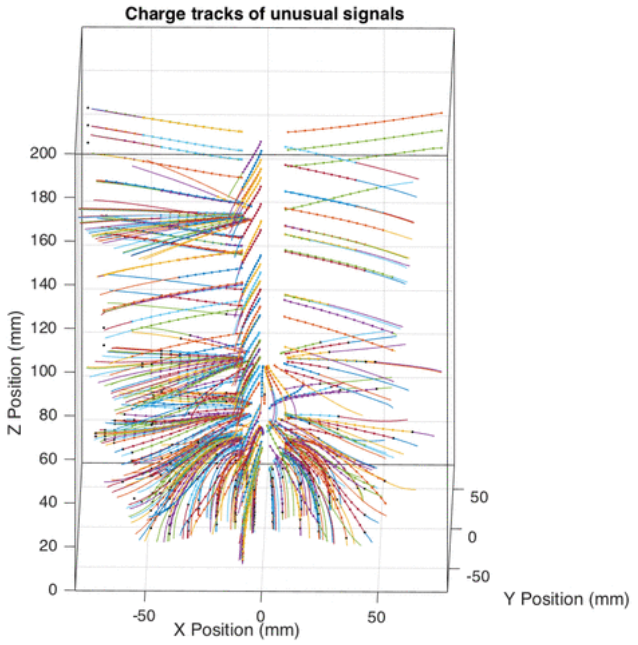


Fraser Holloway - F.Holloway@liverpool.ac.uk

In Summary

- Several algorithms have been developed for fold-1
- Adaptions for multiplicity are hard
- Database needs to be validated experimentally
- Odd effects in basis set need to be investigated

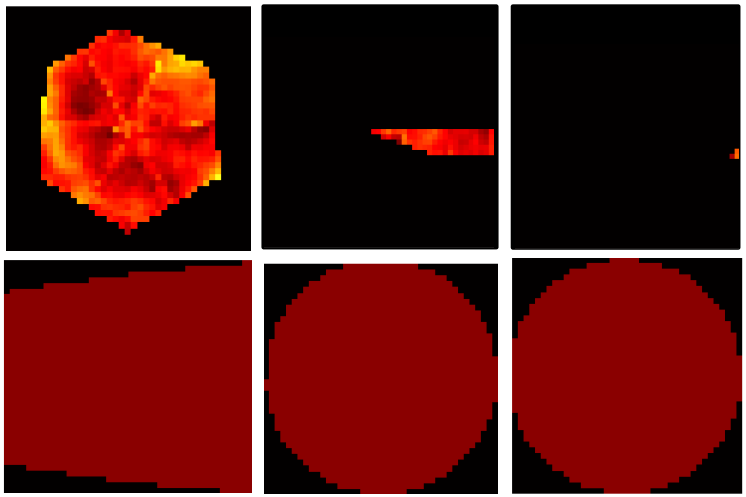


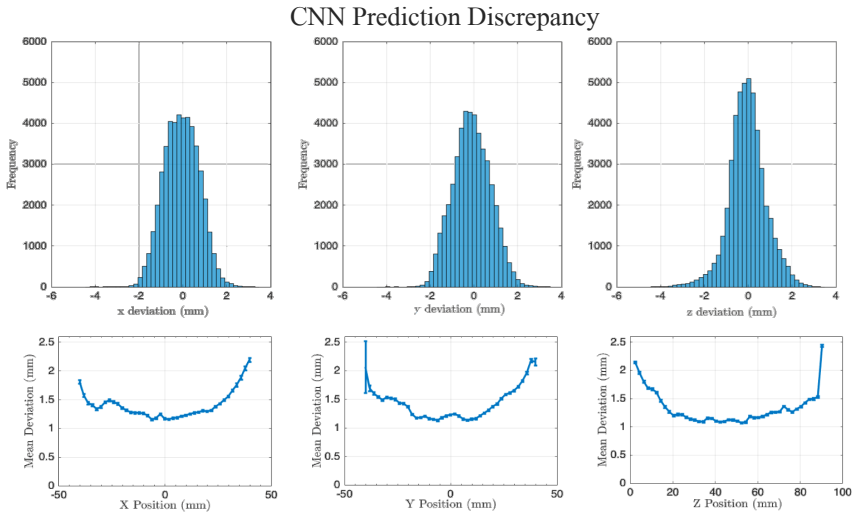


Fraser Holloway - F.Holloway@liverpool.ac.uk

Position Regression with Machine Learning

- Training set taken from ADL simulated pulses, Gaussian noise added
- CNN attempts to predict interaction location from superpulse
- Currently limited to fold-1 events, may be mitigated by using windows





Fraser Holloway - F.Holloway@liverpool.ac.uk