Deep Neural Networks in EXO-200

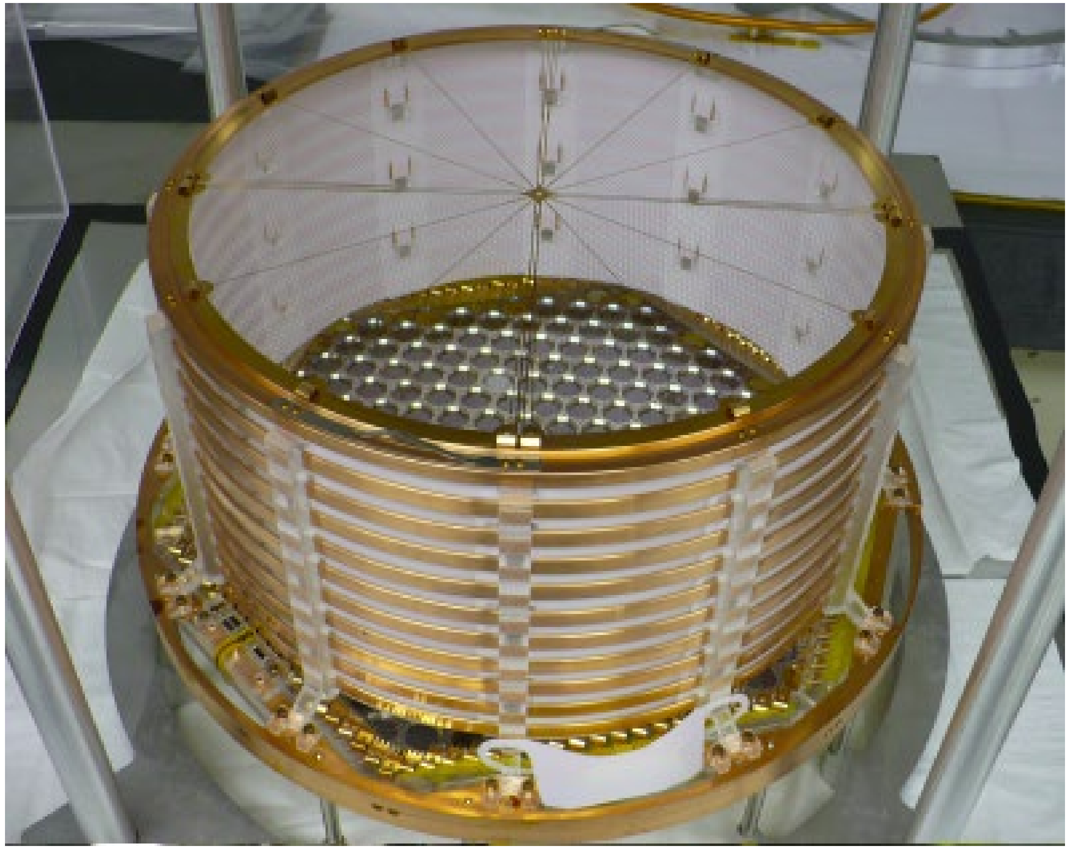
Igor Ostrovskiy – University of Alabama

IHEP EPD Seminar

March 2024

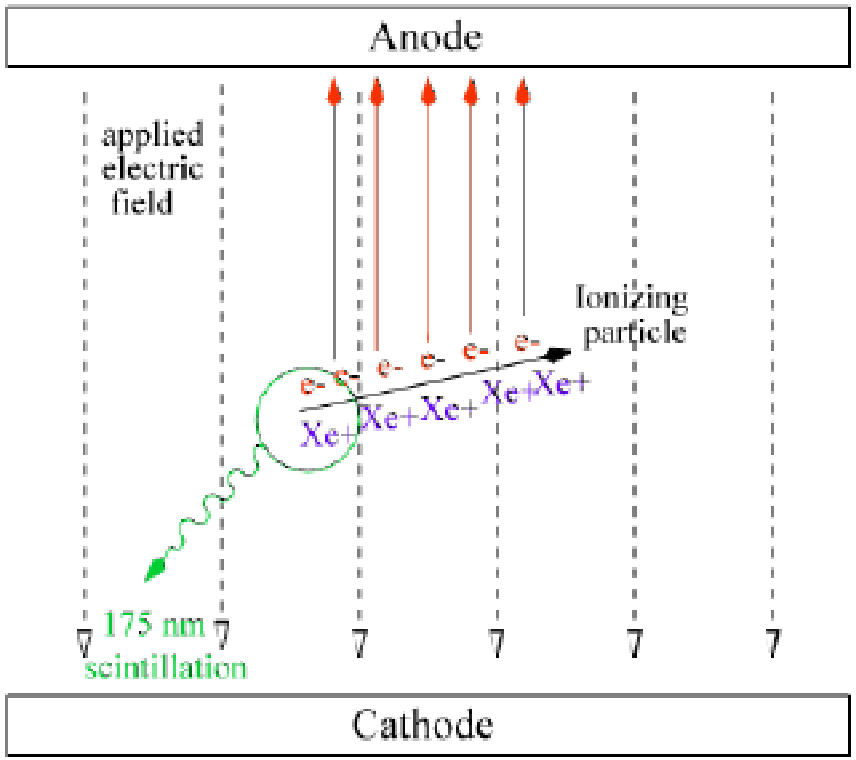
EXO-200

- Double-sided with shared cathode
  - One side shown
  - -8 kV (-12 kV) on cathode in Phase I (II)
- Single-phase LXe TPC
  - Enriched to 80.6% in ^{136}Xe
  - ~175 kg in liquid phase
  - ~90 kg fiducial mass
- Retired in December 2018

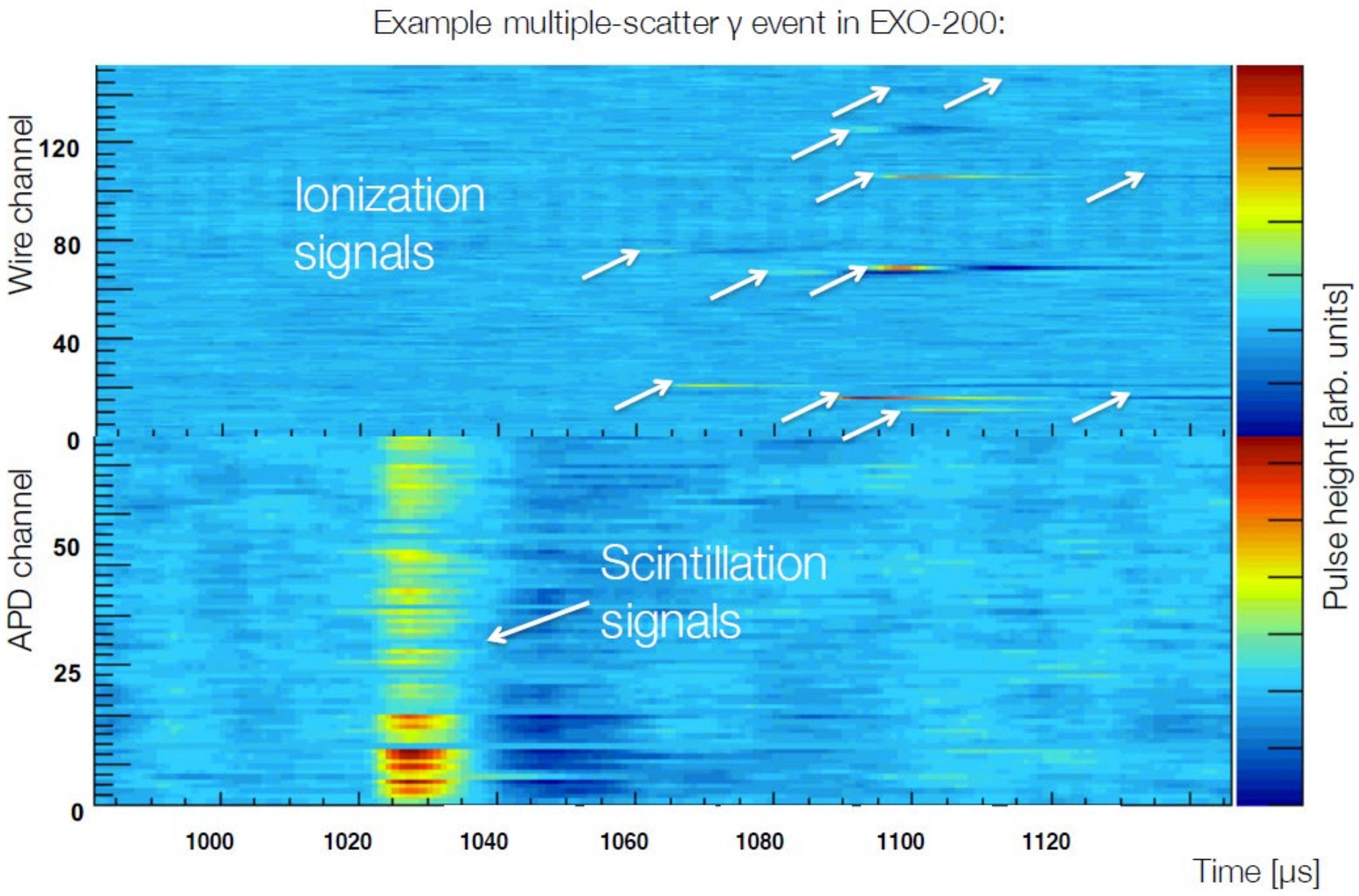


EXO-200

- Each side detects both charge and light
- 38x2 U-wire channels for charge collection
  - 800 e- noise per wire
- 38x2 V-wire channels for charge induction
  - Crossed at 60^o with U-wires
- 74x2 APD channels for light
  - Each channel is a chain of 7 LAAPDs
  - Cathode is mostly transparent (mesh)
  - Cylindrical Teflon reflector

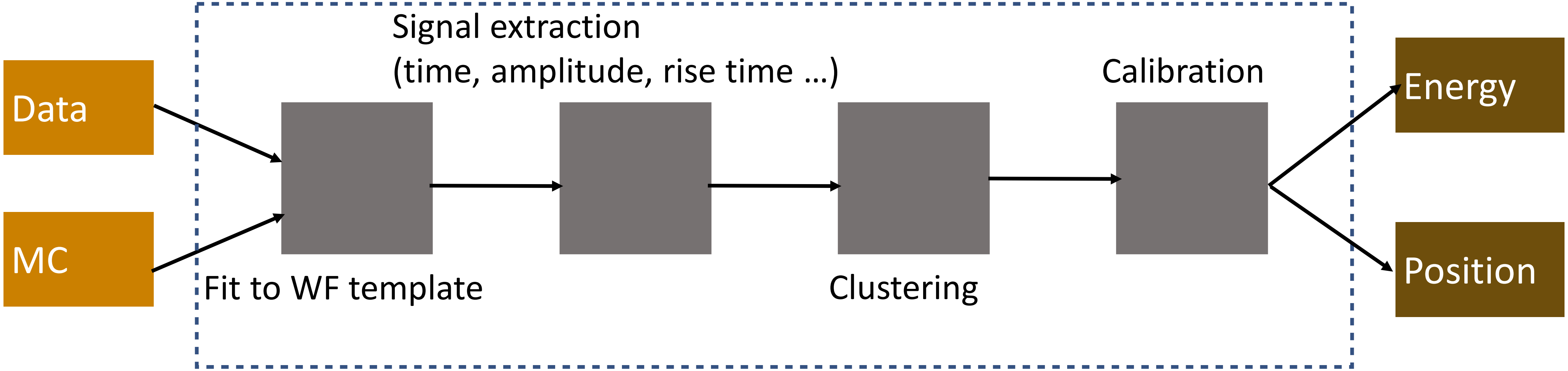


EXO-200 data



Example multiple-scatter γ event in EXO-200:
Wire channel
120
80
40
0
Ionization
signals
APD channel
50
25
0
Scintillation
signals
1000
1020
1040
1060
1080
1100
1120
Time [μs]
Pulse height [arb. units]

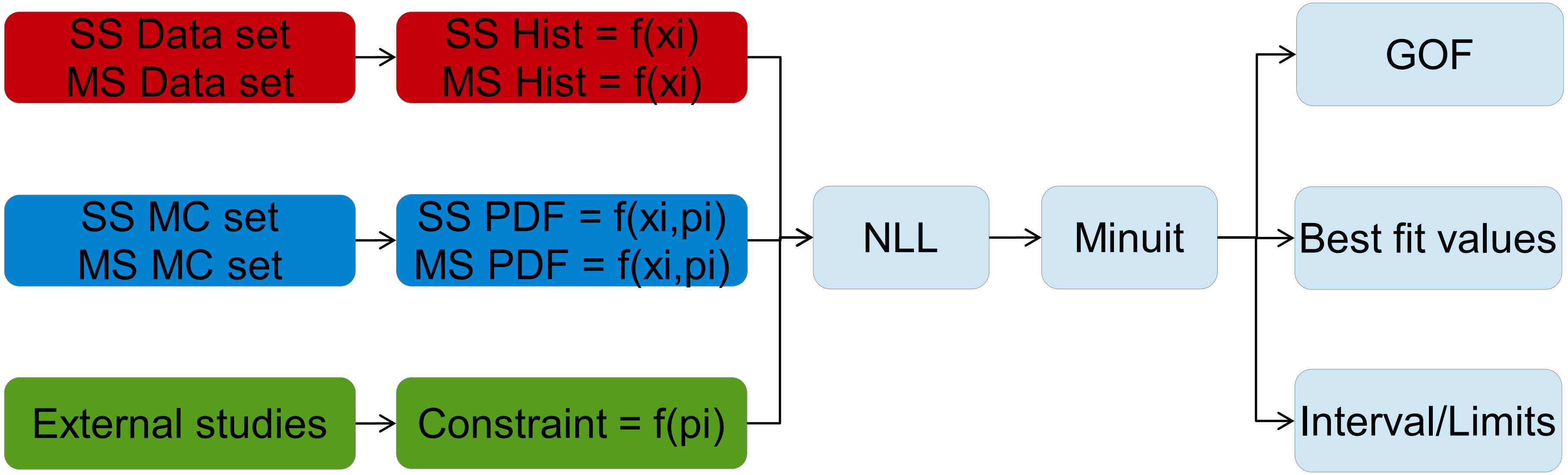
EXO analysis in broad strokes: reconstruction



- Multiple algorithmic steps
- Done by different people over the course of several years
- Imperfections in each step can add systematics

"grey" boxes

EXO analysis in broad strokes: point/interval estimation



- MC based PDFs, binned extended NLL with systematics constraints
- Profile likelihood for interval construction
- Systematics due to recon and MC errors. Measured or estimated using calibration data

Deep Neural Networks (DNN) in broad strokes

- DNN contains many tunable (trainable) parameters
- Training is done by minimizing discrepancy between truth and network’s output
  - E.g., RMS deviation between known and predicted energy
- Minimization is done, essentially, by gradient descent (like MIGRAD), but with some new tricks to efficiently handle the multitude of parameters

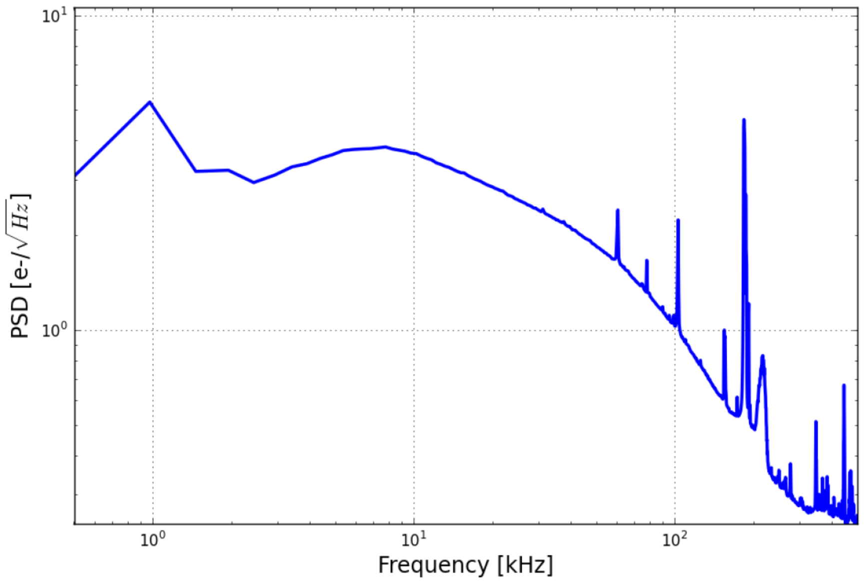
Deep Neural Networks in EXO

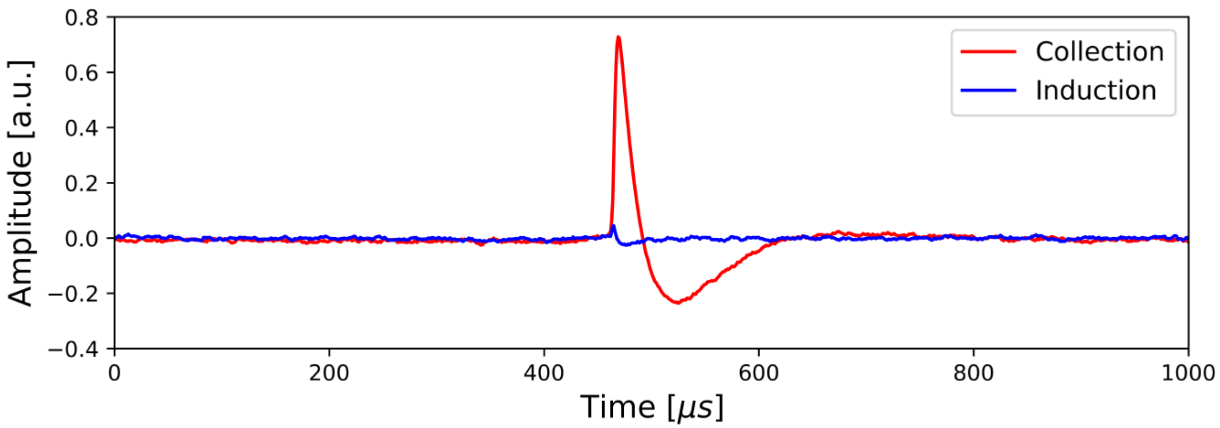
- Can circumvent intermediate steps and extract high level information directly from raw waveforms?
  - **YES**
- Can validate results on real detector data, not just MC?
  - **YES**
- Even then, if using MC truth during training, would be limited by how well MC models data (as some standard analysis steps are). Can reduce reliance on traditional MC?
  - **YES (Sometimes)**

- *JINST* **13** P08023 (2018), *Phys. Rev. Lett*. **123** 161802 (2019), *JINST* **18** P06005 (2023)
- Note: not covered in this talk are EXO-200 works that use non-DNN ML or use DNN with high-level info as input

First application: charge energy reconstruction

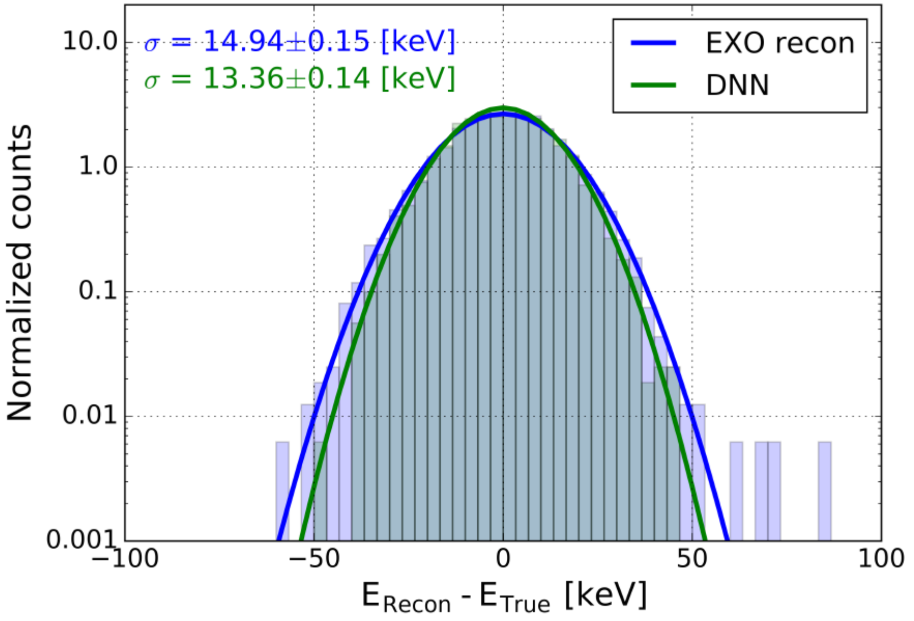
- The main challenges of charge reconstruction are nontrivial noise and disentangling U-wire signal into induction and collection

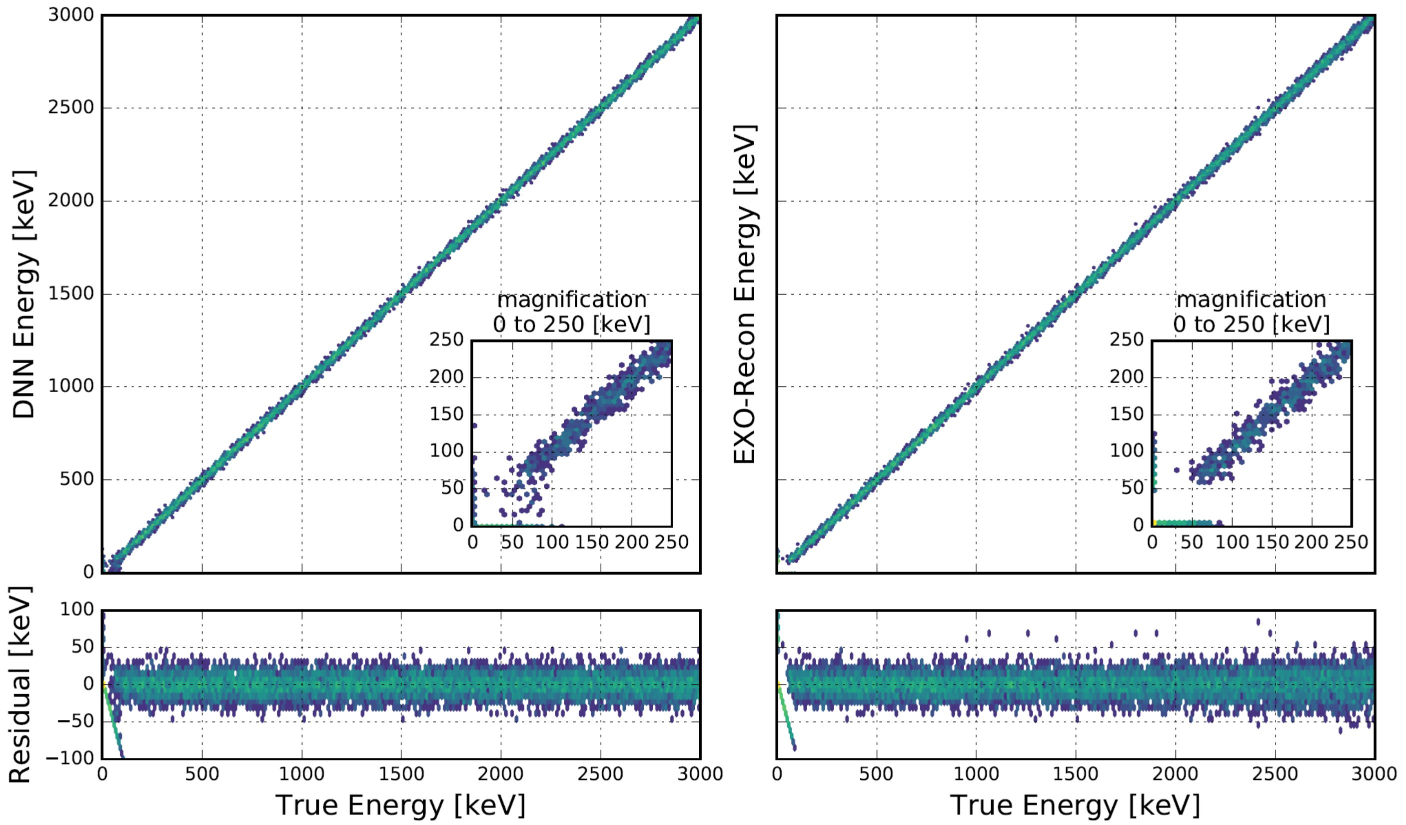




First application: charge energy reconstruction

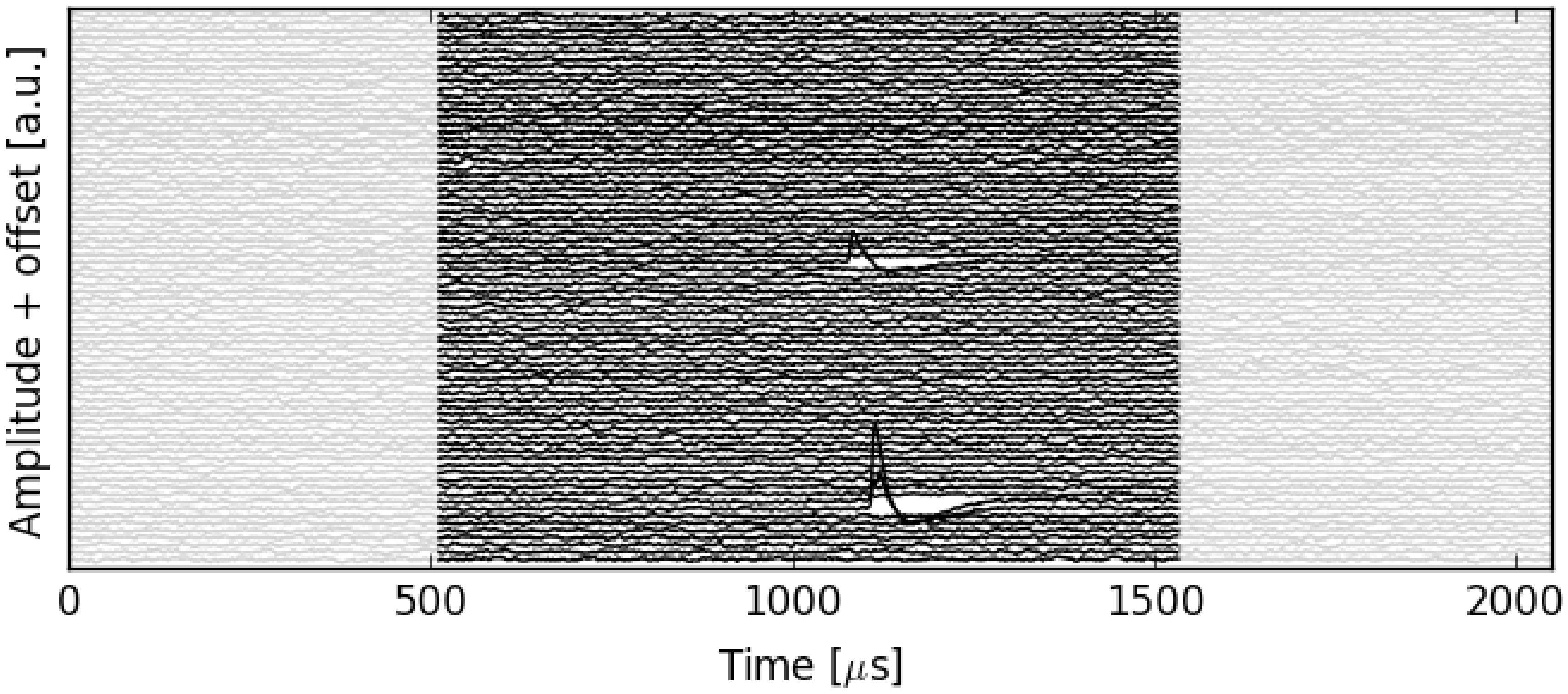
- Starting with single wire





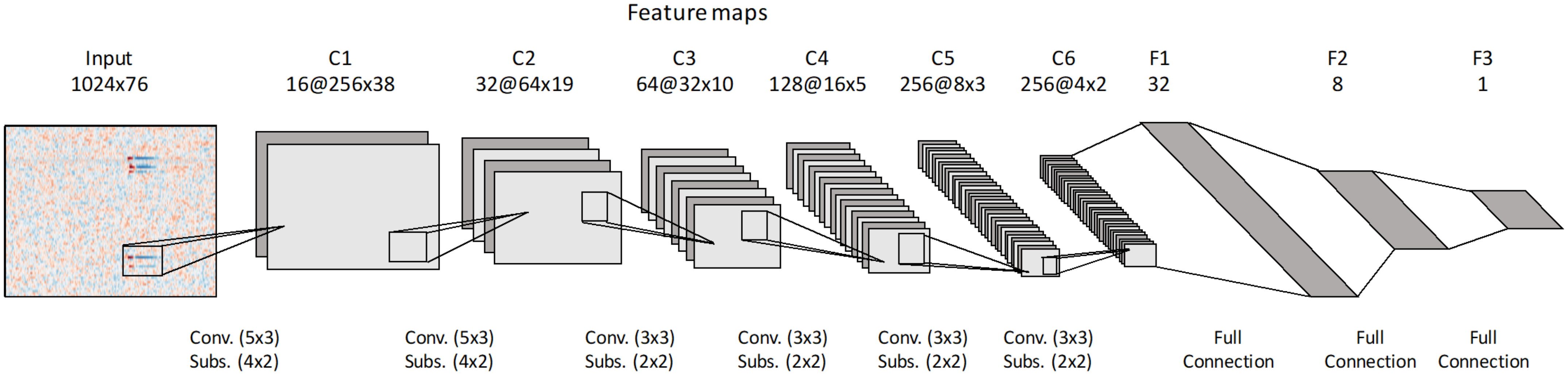
First application: charge energy reconstruction

- Now full events – all 76 U-wire waveforms (1024 time samples)
- Minimal Preprocessing: correct channel gains + crop waveforms



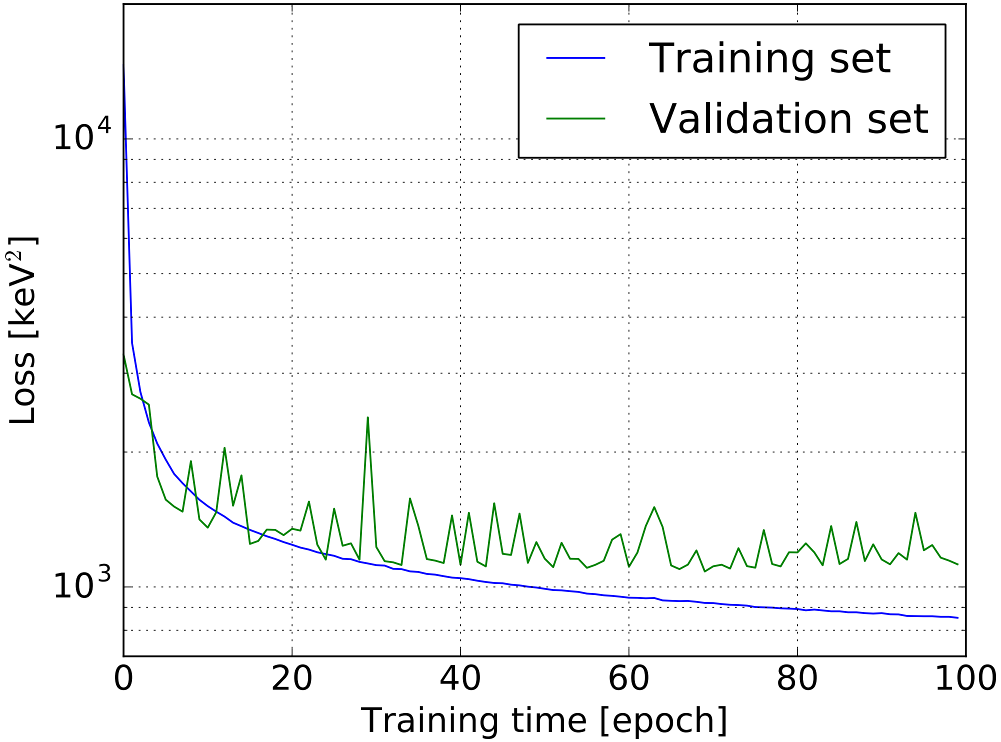
First application: charge energy reconstruction

- Input waveform image
- Convolutional part extracts features from image
- Dense part extracts target variable(s) from features



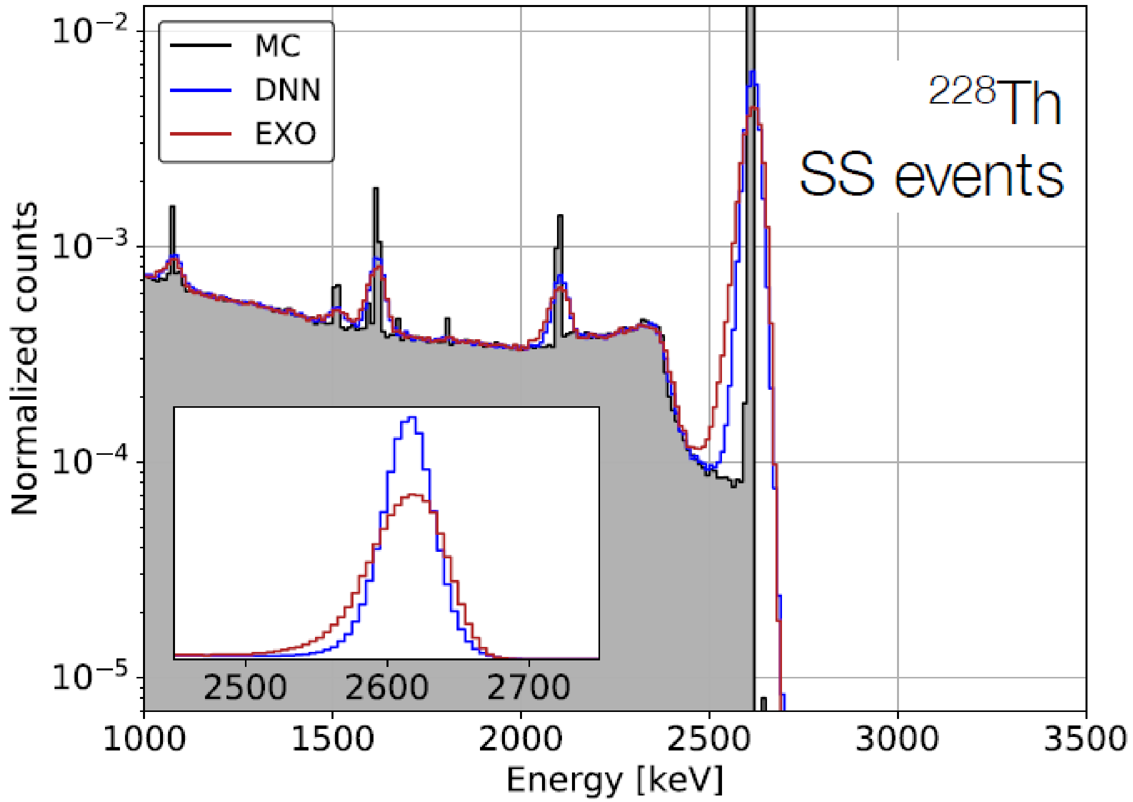
Charge reconstruction training details

- Training data:
  - Simulated events
  - Gamma ray source
  - Detector response uniform in energy
- Training:
  - 720 000 training events
  - 100 epochs
- Technical details:
  - Adam optimizer
  - Minimize mean square error
  - L2 regularization
  - ReLU activation
  - Uniform Glorot initialization



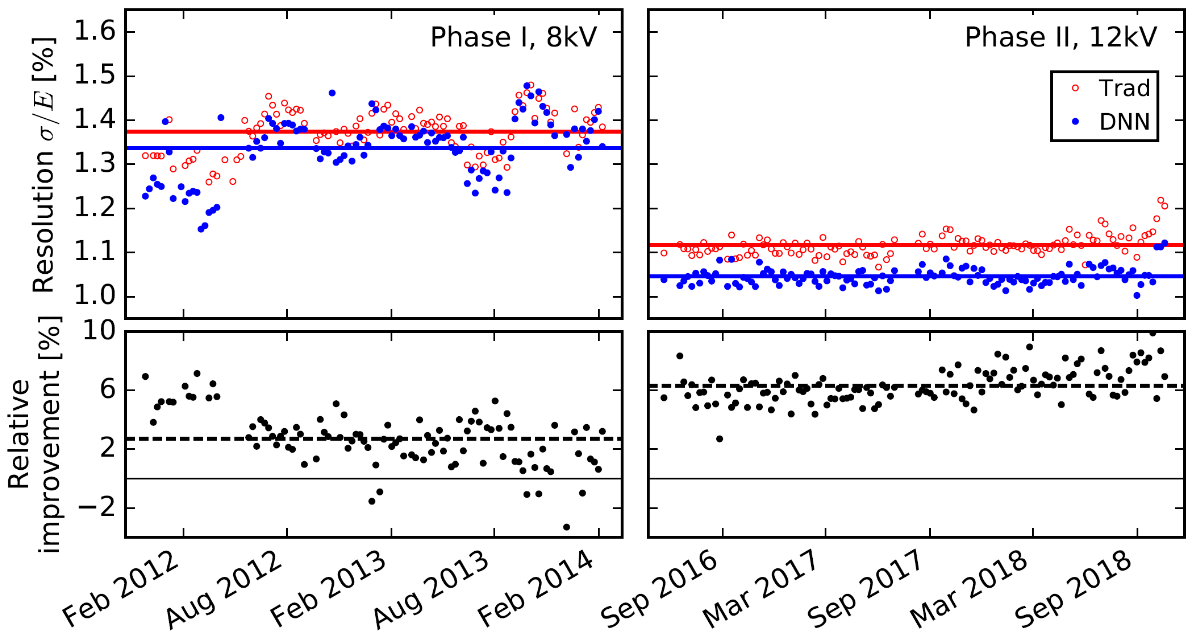
First application: charge energy reconstruction

- Reconstruction works on MC over the energy range under study
- Resolution (σ) at the ^{208}Tl full absorption peak (2615 keV):
  - **DNN: 1.21% (SS: 0.73%)**
  - **EXO Recon: 1.35% (SS: 0.93%)**
- Network outperforms in disentangling mixed induction and collection signals
  - See valley before ^{208}Tl peak, right in $0\nu\beta\beta$ ROI!



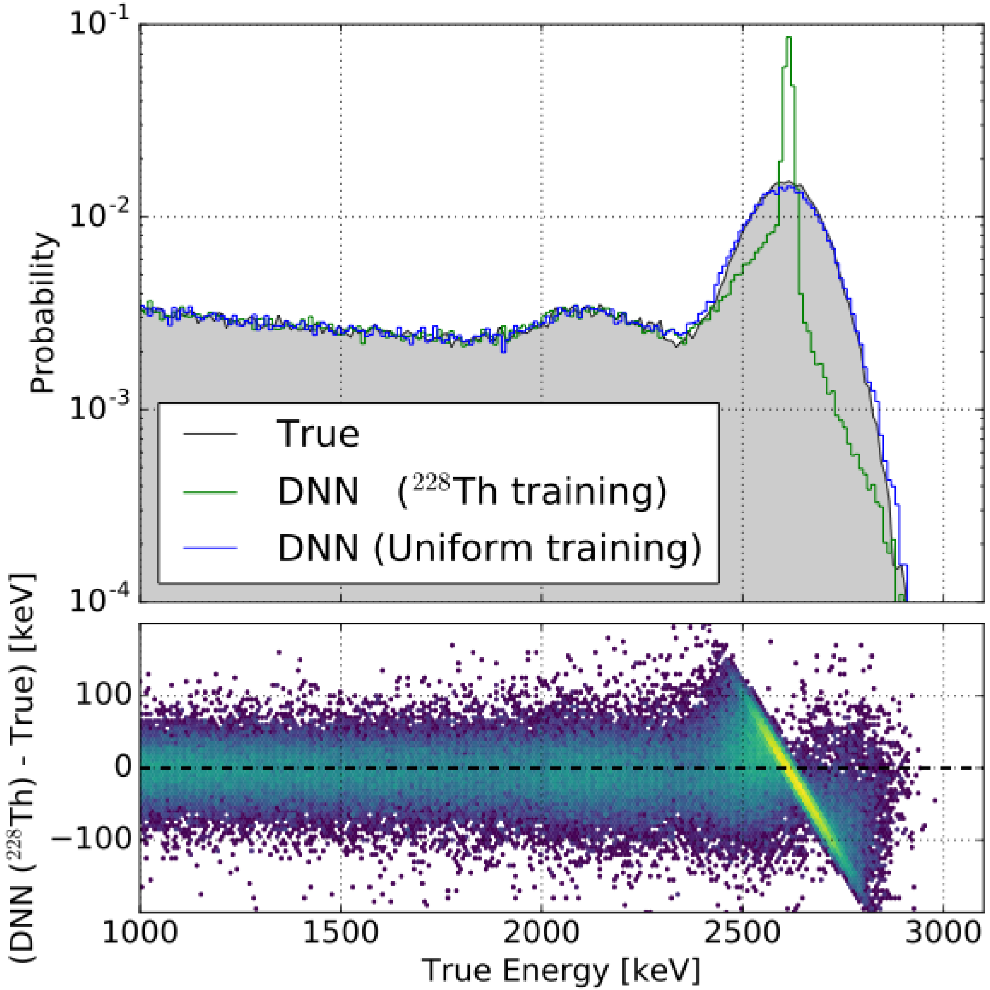
First application: charge energy reconstruction

- Applied to data and anti-corrleated with scintillation, the DNN based „rotated“ resolution outperforms EXO by 2-6%, depending on the week
- The better performance of the DNN alerted that something was lacking in the traditional approach and triggered improvements in EXO-recon
- While the cause is now largely understood (handling of mixed induction and collection signals), the developed traditional solution in EXO-recon is still outperformed by the DNN



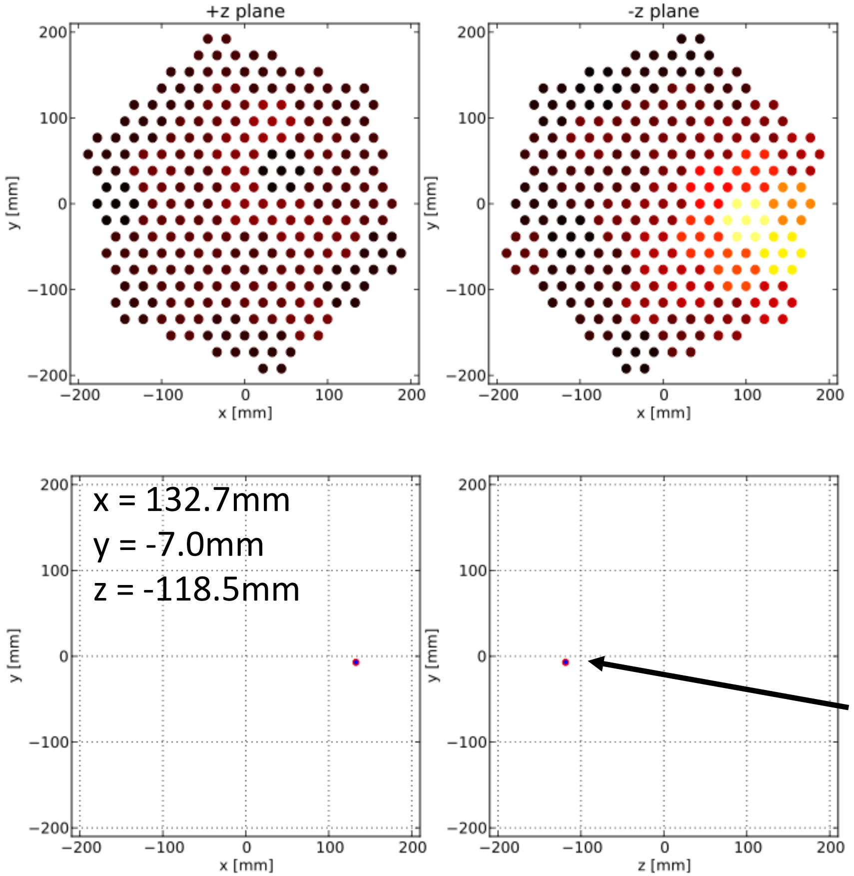
First application: Pitfalls of DNNs

- Potential danger of DNN is that they learn to reproduce the training data well but perform poorly on real data.
  - **Validation on real data is critical**
- We saw this in EXO-200:
  - DNN over-trains on sharp MC training peaks and shuffles independent validation events towards the sharp peaks → resolution too good to be true
  - Mitigated by using training events with uniform energy distribution



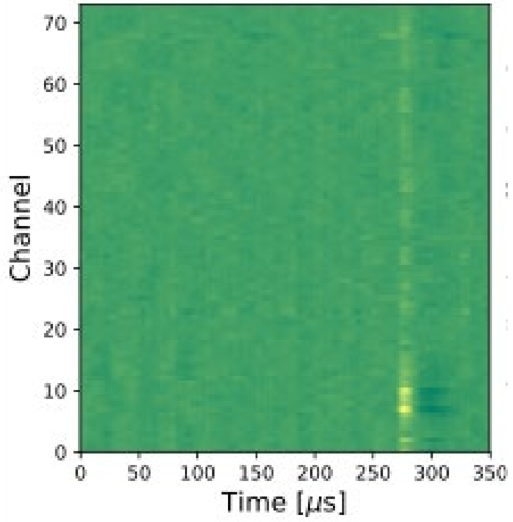
Second application: light position reconstruction

- Event position reconstruction from scintillation light
  - Truth label provided by ionization information of real data
  - Input are all 74 raw APD **real data** waveforms cropped to 350 µs



Event position is encoded in APD pattern

The time dimension adds information on waveform shape and noise



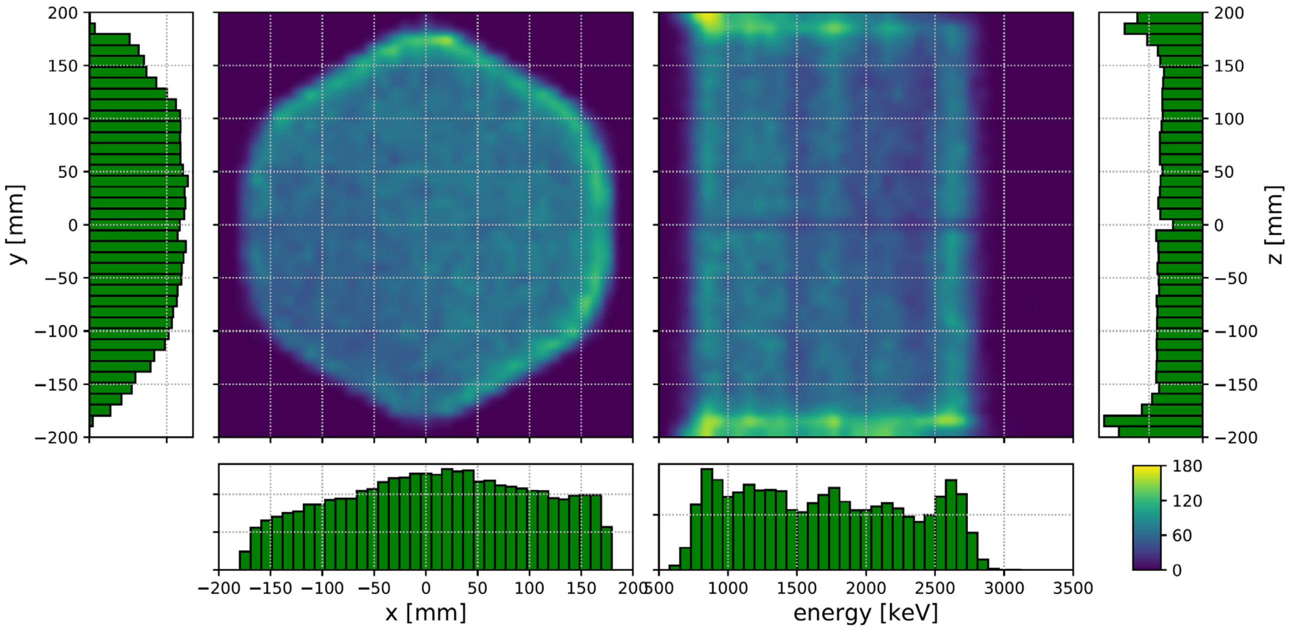
Truth information extracted from ionization signal

Light reconstruction details

- Waveform image is fed to CNN consisting of 4 convolutional and 3 fully connected layers
- Output has three units corresponding to event x-, y-, z-coordinates
- Loss function is Euclidean loss with L2 regularization

$$L = C + \lambda \cdot R \quad \text{where} \quad C = \frac{1}{3m} \sum_{i=1}^{m} \sum_{k=1}^{3} \left(y_i^k - \hat{y}_i^k \right)^2$$

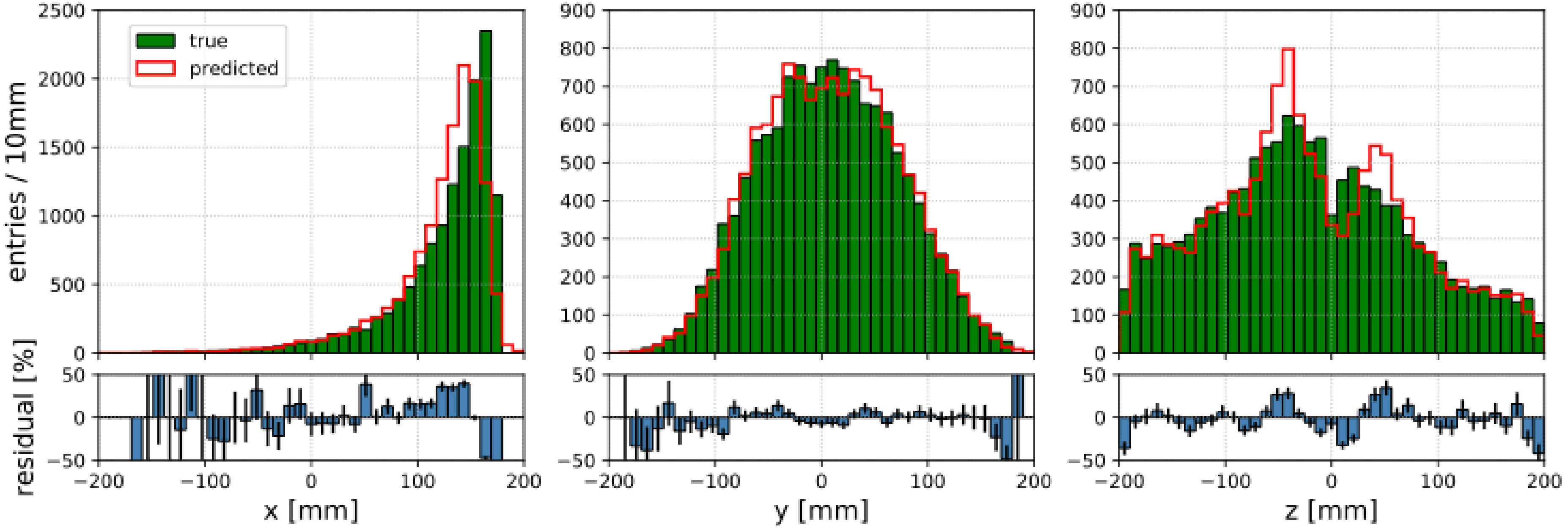
- Training is done on **real** calibration data with uniform distribution in space and energy



Second application: light position reconstruction

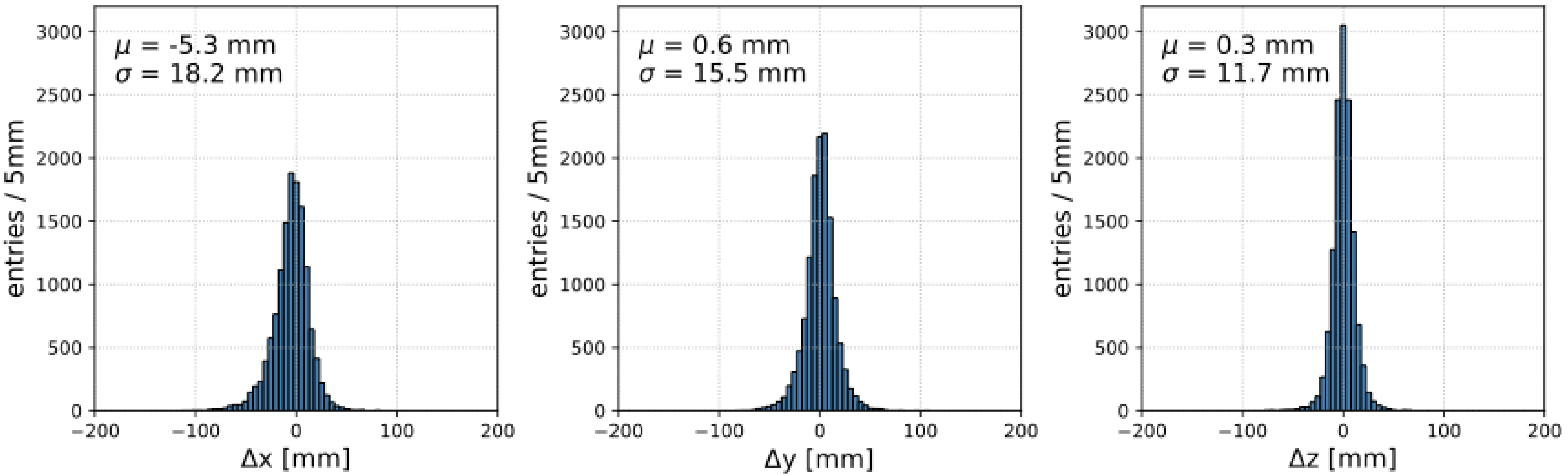
- Loss function reaches 200 mm^2 after training the DNN for 200 epochs
- The corresponding resolution in 3D is 25 mm
- The model is tested on different types of source data at different locations
- No alternative light position reconstruction in standard analysis, so uncontested

Accuracy: 22.5mm (d_x = 13.6mm, d_y = 11.3mm, d_z = 8.1mm) corresponding to R^2 = 0.99



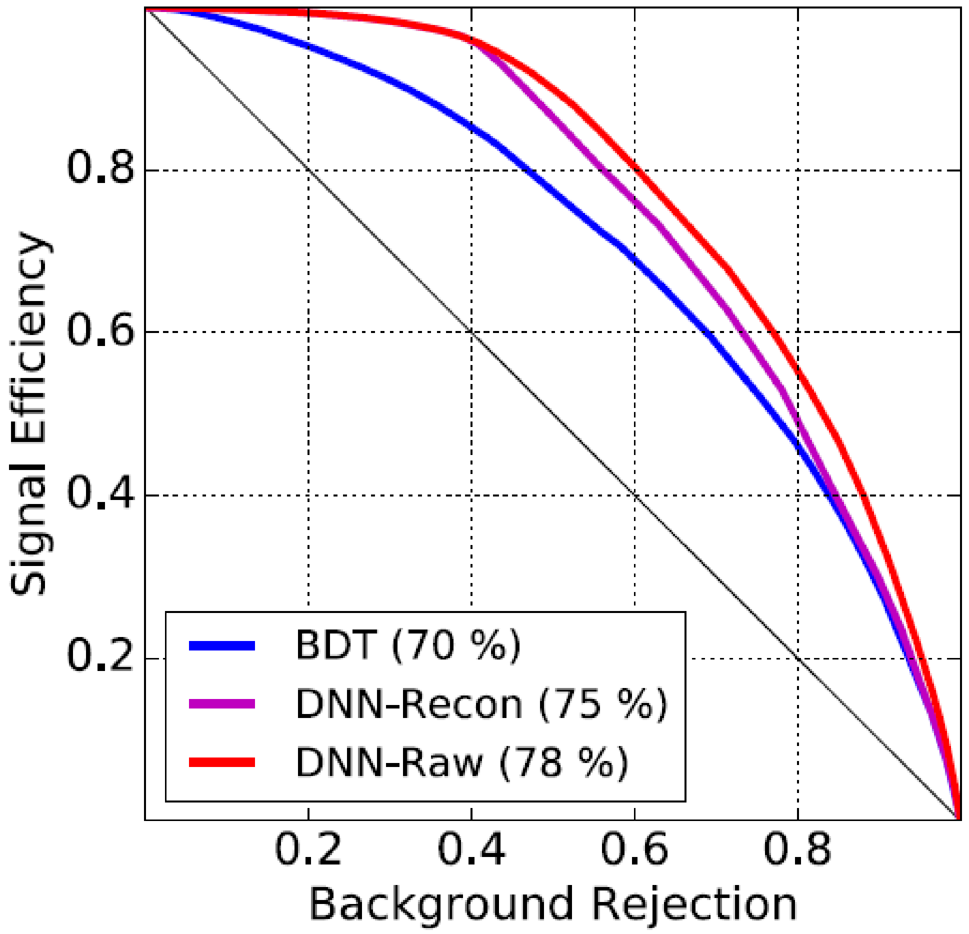
Second application: light position reconstruction

- Loss function reaches 200 mm^2 after training the DNN for 200 epochs
- The corresponding resolution in 3D is 25 mm
- The model is tested on different types of source data at different locations
- No alternative light position reconstruction in standard analysis, so uncontested

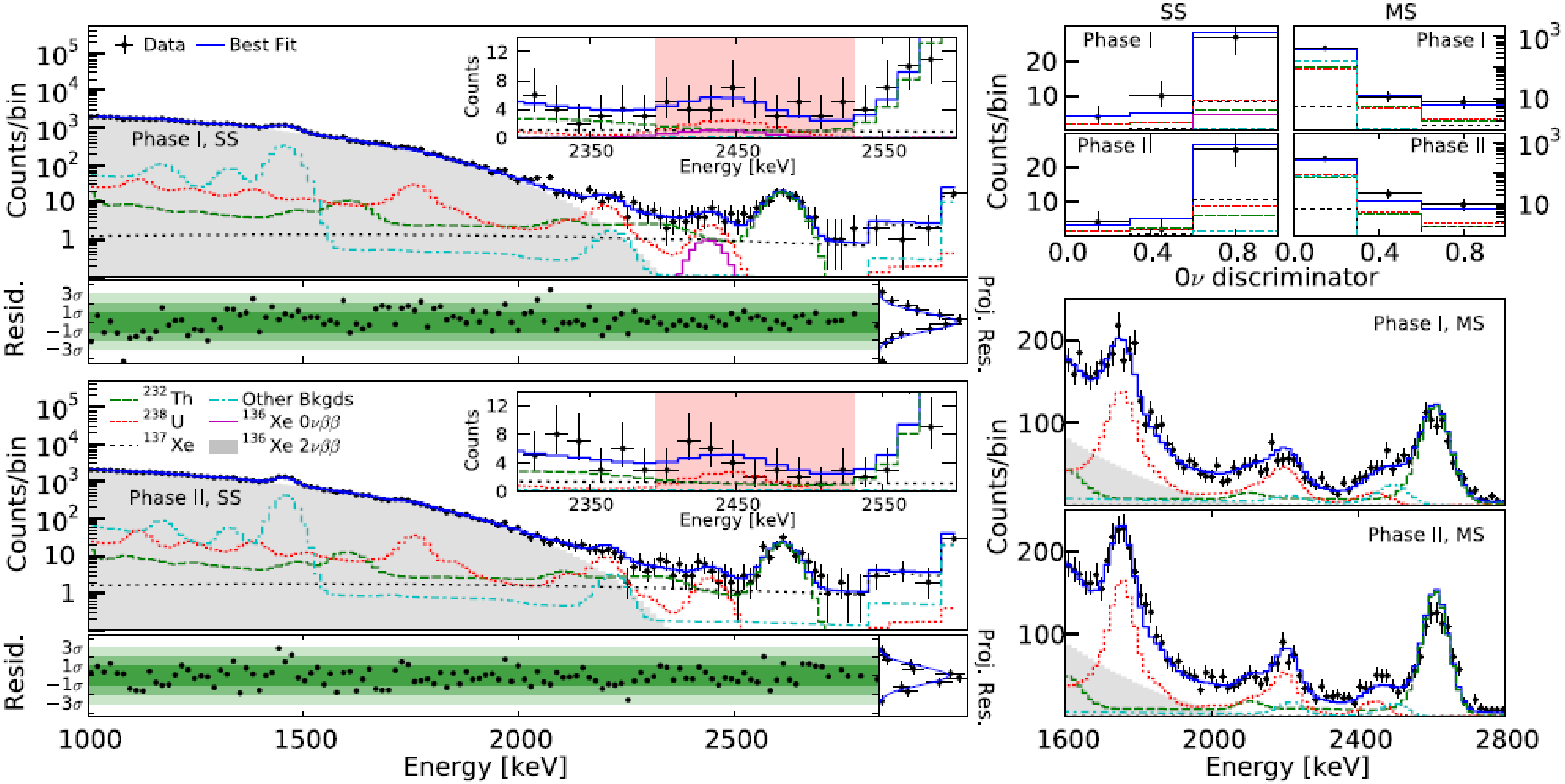


Third application: Signal/Background Discrimination

- Binary ($\beta\beta$ vs γ) DNN based discriminator as an additional variable to the "traditional" ML fit
- DNN trained on waveforms re-generated from EXO recon'd signals (not on raw waveforms)
- DNN outperforms previously used BDT discriminator
- Overall, a 25% sensitivity improvement, compared to non-ML based analysis
  - *Phys. Rev. Lett.* ***123****, 2019, 161802*
  - Kudos to grad. students who made this happen (Tobias Ziegler&Mike Jewel most of all)

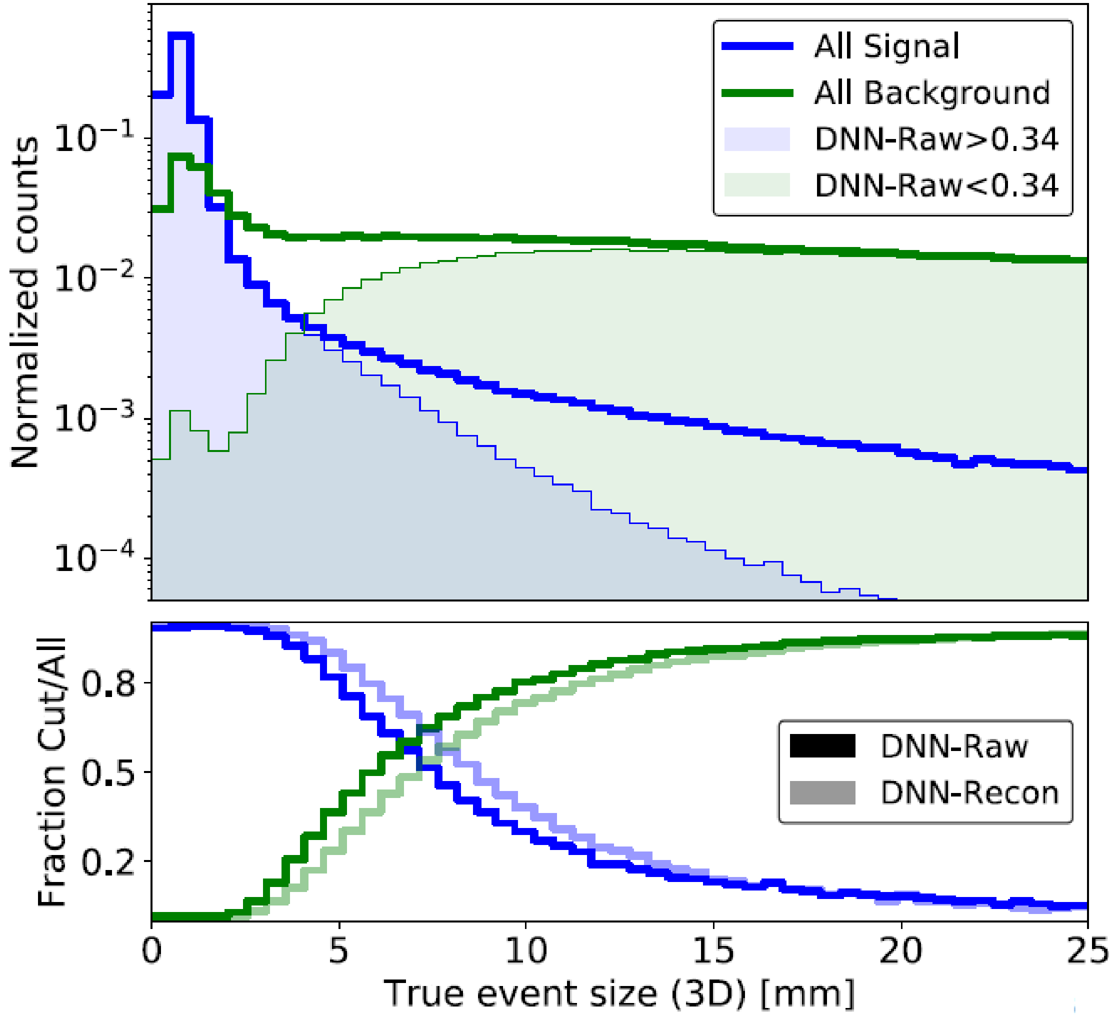


- Energy spectra: SS (left) and MS (bottom right)
- DNN spectra: SS/MS (top right) of projected for ROI events



Third application: Signal/Background Discrimination

- $\beta\beta$ events are more localized than γ
- DNN efficiency demonstrates correlation with the true event size in the MC
- Indicates that the DNN picks up correct features of the waveform when reconstructing events
- Data/MC agreement of the "DNN variable" validated with real calibration data
    - Agreement not perfect, but comparable to other "shape" errors.

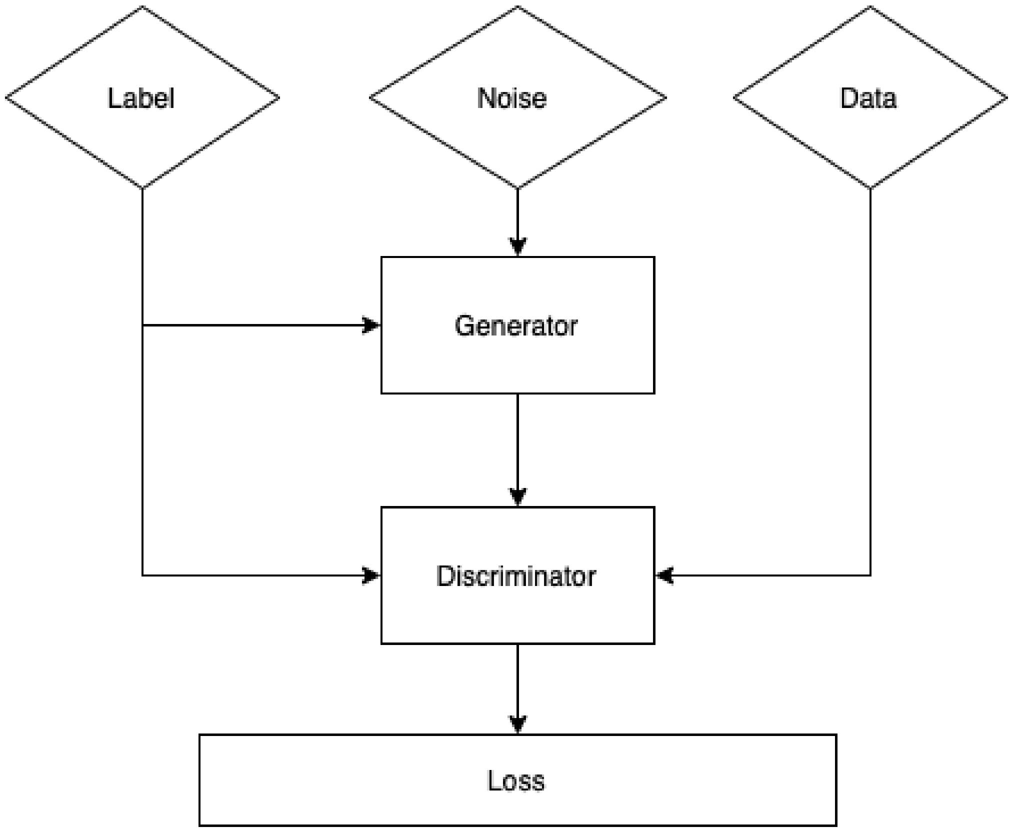


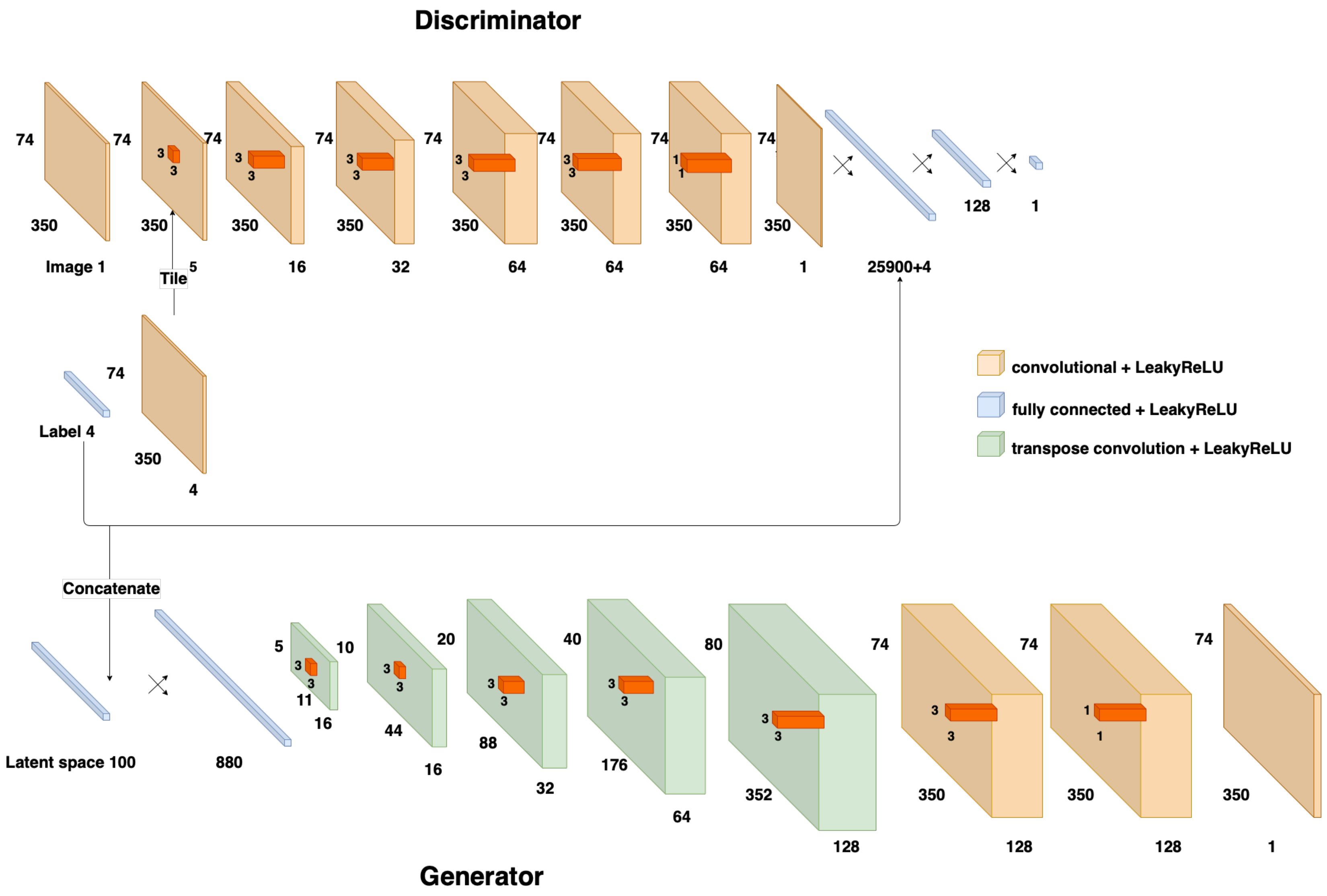
Most recent: MC with GANs

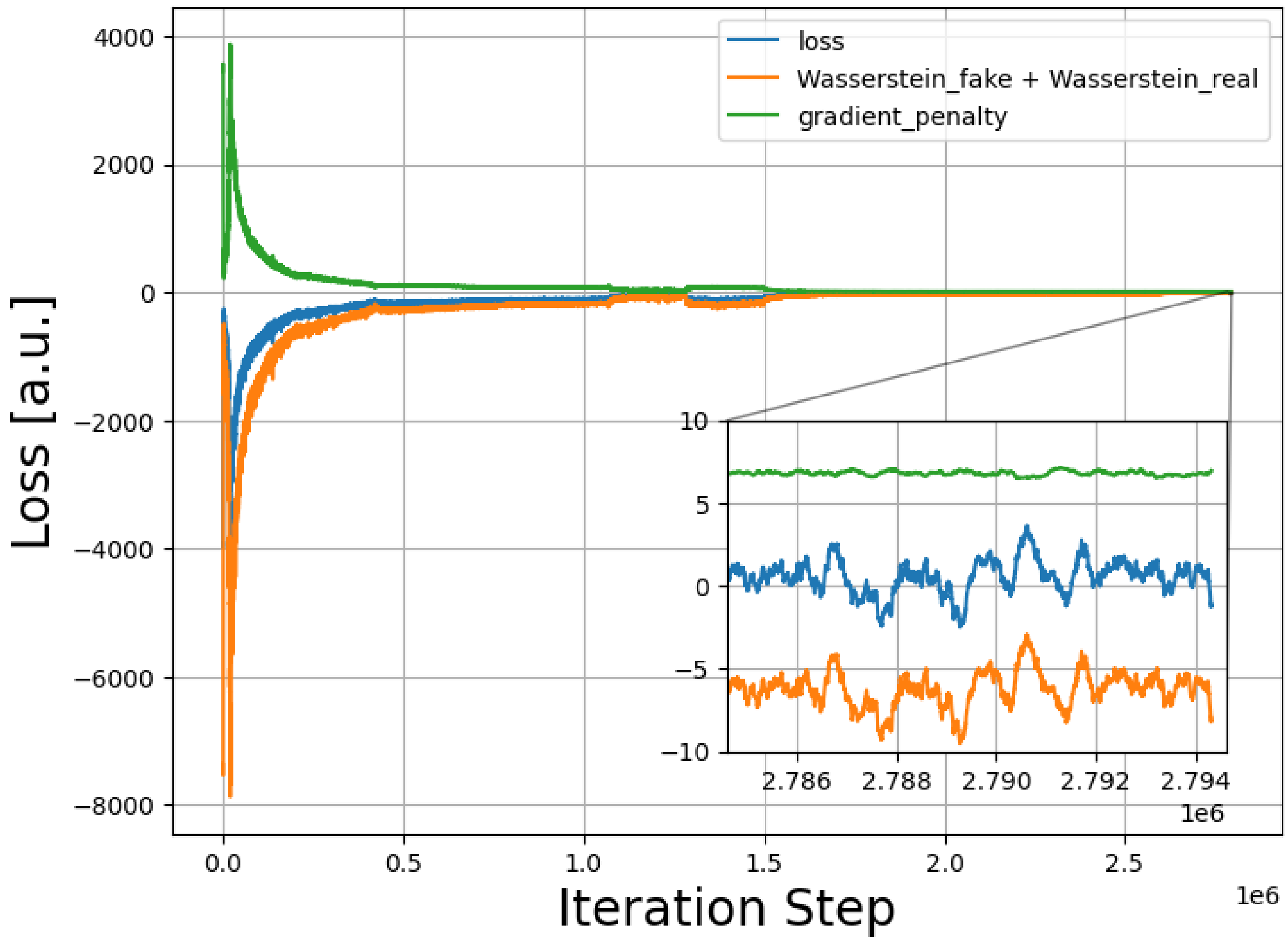
- EXO-200’s earlier attempts to develop a detailed photon-tracking MC did not succeed
  - Poor agreement with data, possibly due to imperfect knowledge of optical properties or shortcuts in geometry implementation
  - It was also very resource-consuming to track photons
  - A simple parametric simulation of the overall light yield per one array of APDs was used instead, only for limited purposes
- We showed that one can train a GAN network directly with waveforms from calibration data, bypassing the needs for detailed knowledge of optical properties and detector geometry
  - *Importantly, we compared the output at all levels – from raw waveforms to signal amplitude and its position dependency, to reconstructed energy spectra*
  - *JINST* **18** P06005 (2023)

Most recent: MC with GANs

- Generator starts from white noise and label with requested position, energy
- Critic (discriminator) compares the generated waveform to data sample
- Wasserstein (Васерштейн) distance, aka Kantorovich distance, as a metric for comparison (more stable than standard GAN)
- Constrainer: supervises training and ensures the generated waveform conforms to the requested label



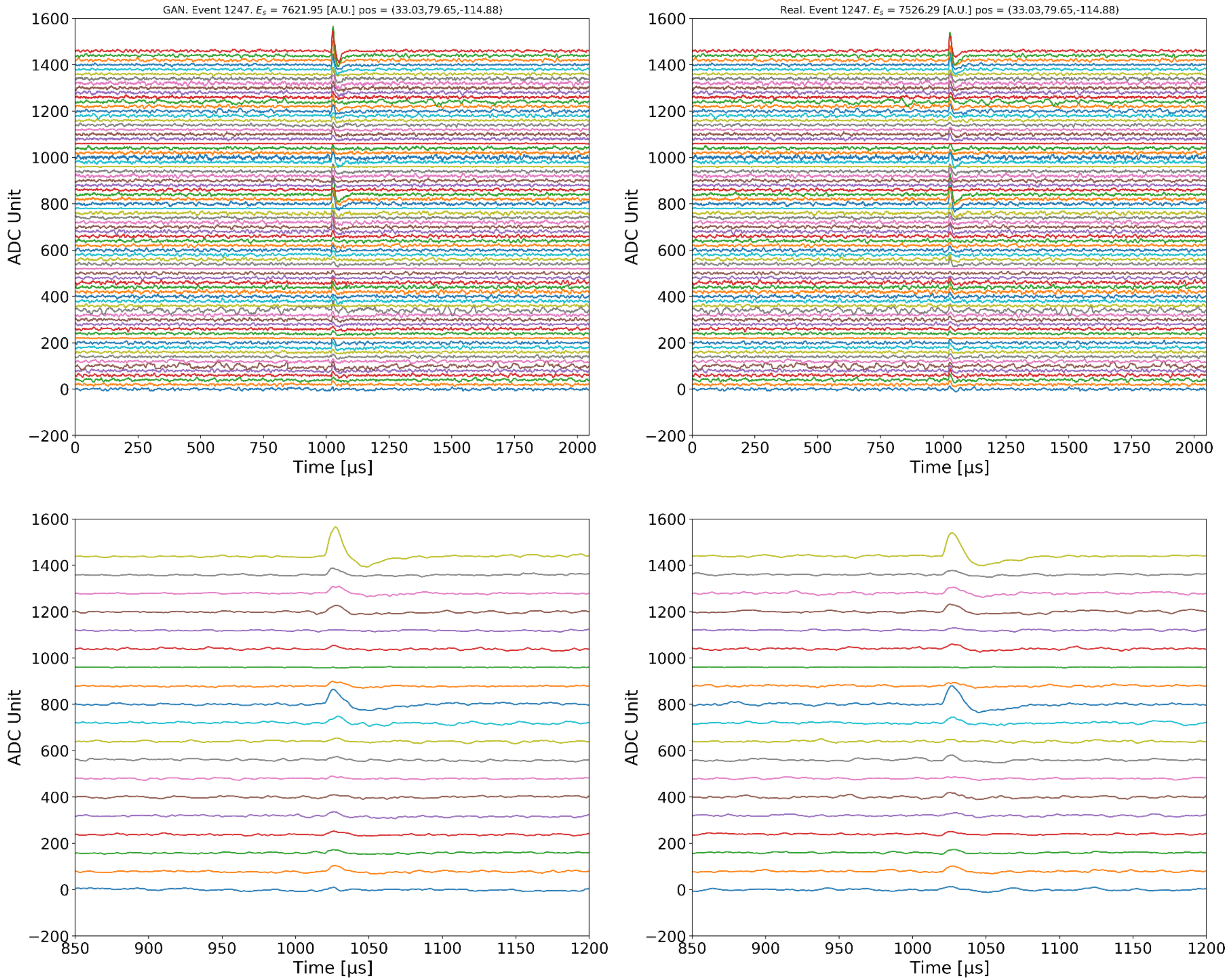

Discriminator
74
350
Image 1
Tile
5
16
32
64
64
64
1
25900+4
128
1
74
Label 4
350
4
convolutional + LeakyReLU
fully connected + LeakyReLU
transpose convolution + LeakyReLU
Concatenate
Latent space 100
880
5
10
11
16
20
44
16
40
88
32
80
176
64
74
352
128
74
350
128
74
350
128
350
1
Generator



loss
Wasserstein_fake + Wasserstein_real
gradient_penalty
Loss [a.u.]
Iteration Step
1e6
4000
2000
0
−2000
−4000
−6000
−8000
0.0
0.5
1.0
1.5
2.0
2.5
10
5
0
−5
−10
2.786
2.788
2.790
2.792
2.794
1e6

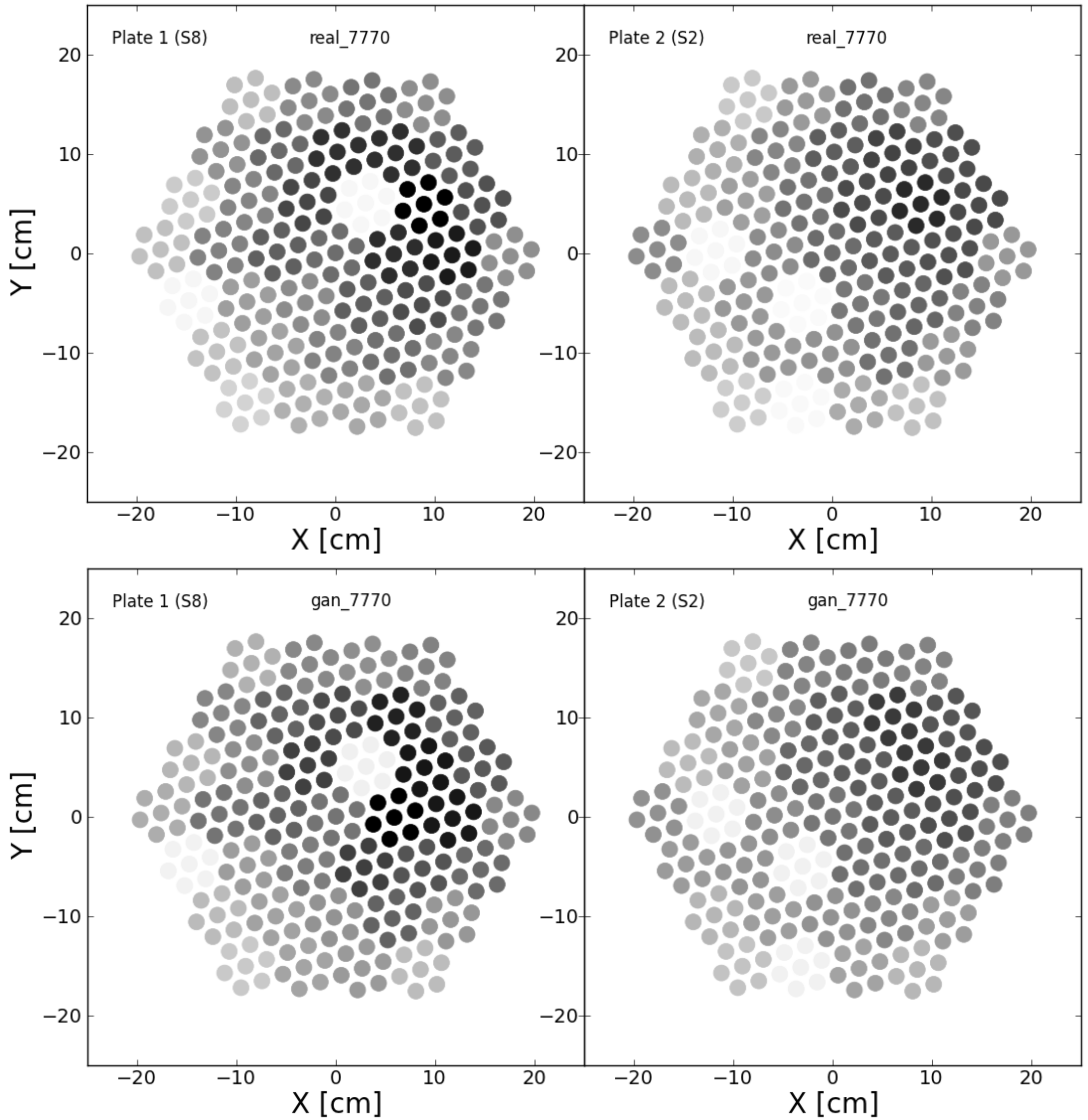
Most recent: MC with GANs

- Raw waveform comparison
- GAN generates waveforms more than an order of magnitude faster than the standard EXO approach
  - that does not even include photon tracking



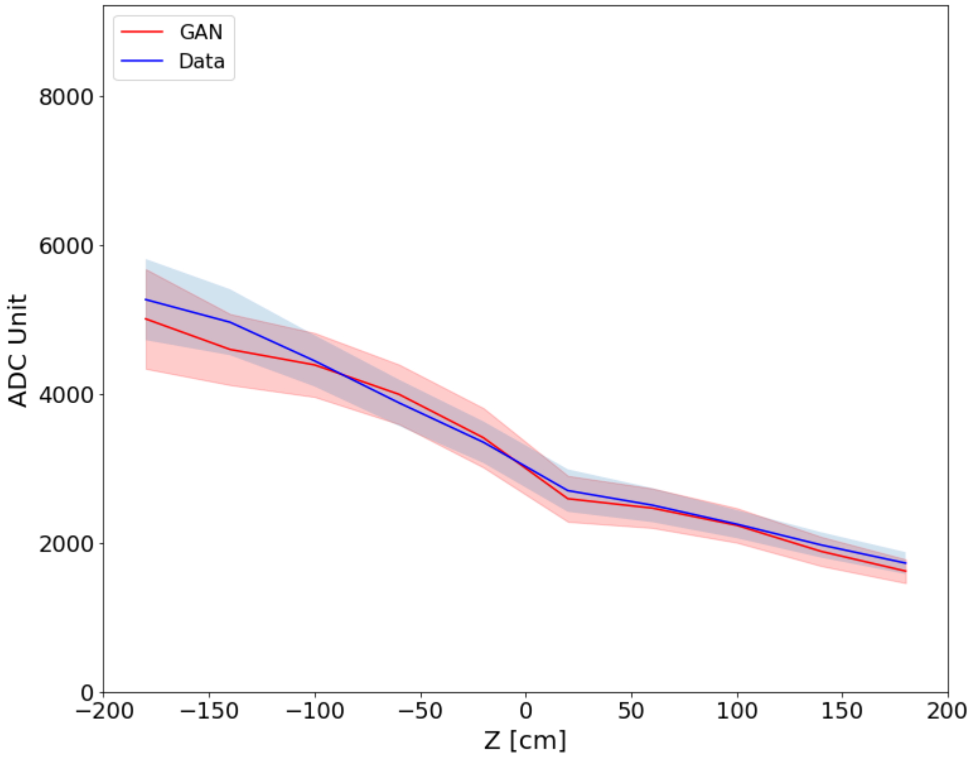
Most recent: MC with GANs

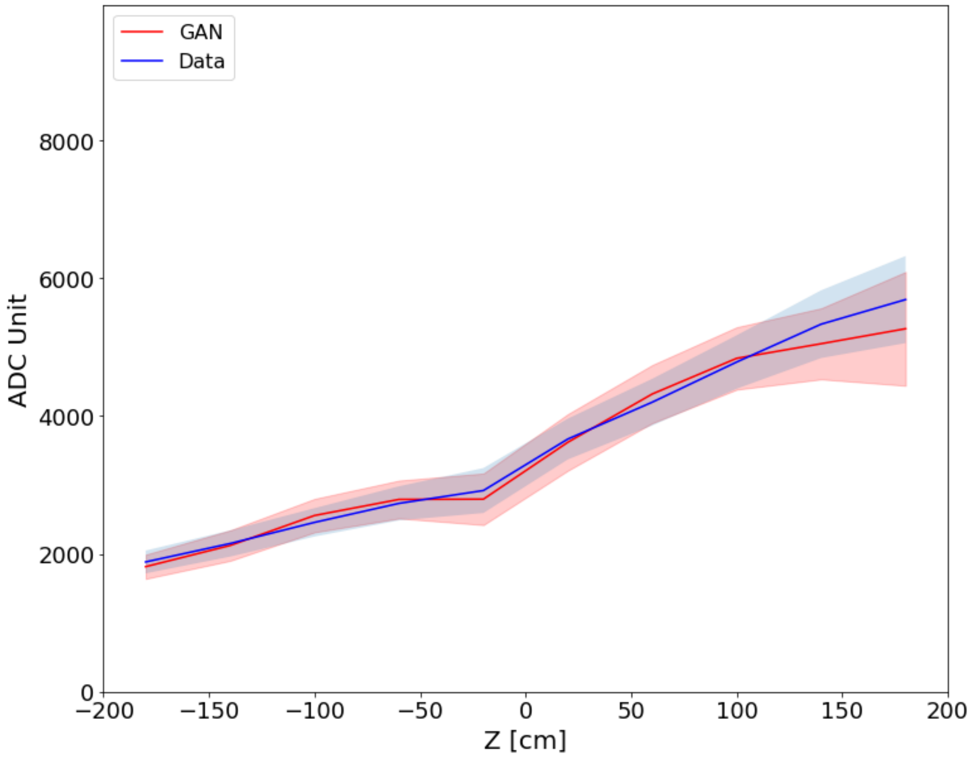
- Summed amplitude per APD gang
  - GAN reproduces the dead channels



Most recent: MC with GANs

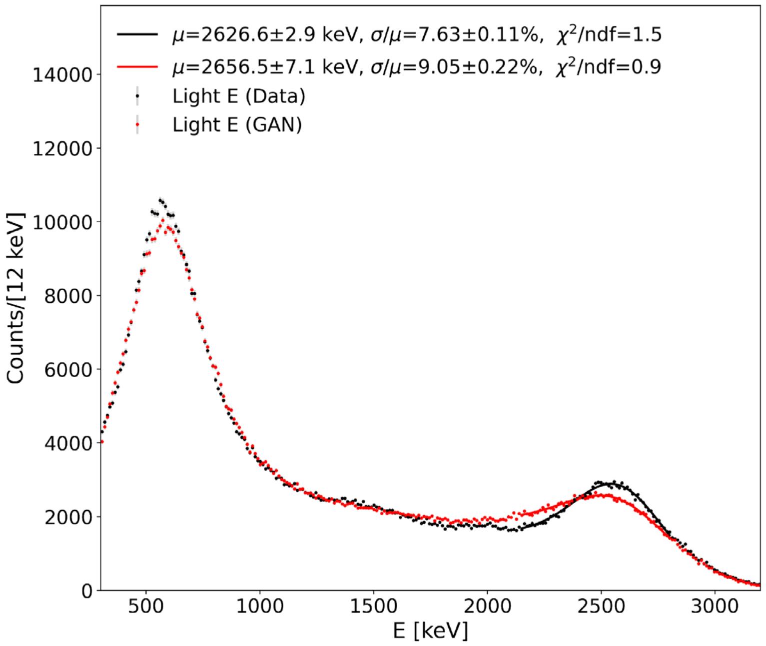
- Position dependence of light response reproduced

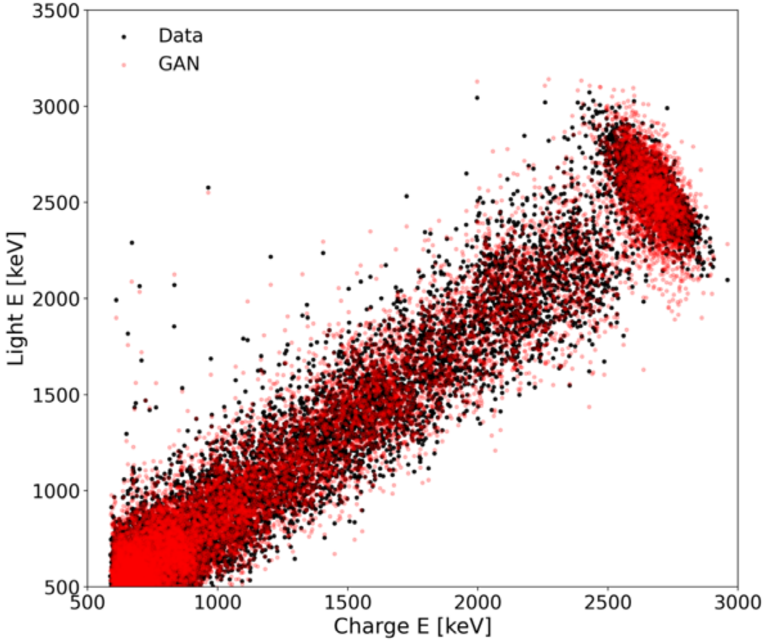




Most recent: MC with GANs

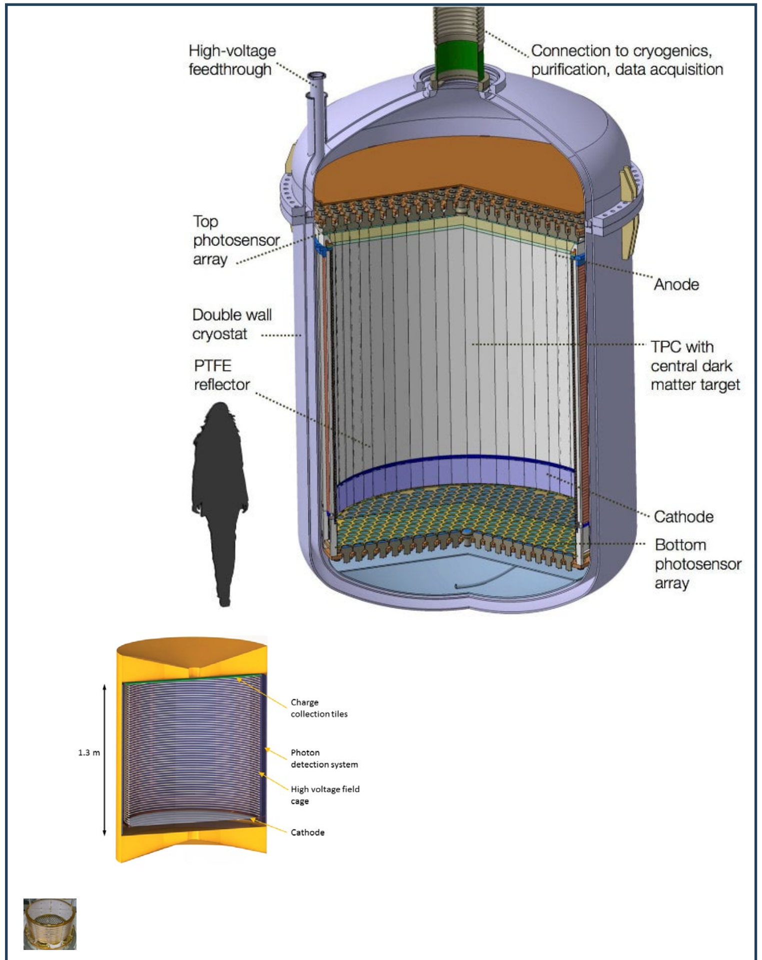
- Anti-correlation between charge and light signals reproduced
  - optimal angle is slightly different
- Light-only energy spectrum looks good but does not reproduce the resolution exactly
  - Consistent with the extra uncertainty added by imperfect truth labels. Experiments that could train on calibration data with more precise labels can do better





Towards next-generation LXe TPC experiments

- DNNs are being used to help guide design
  - See, e.g., D.Bajpai's talk at APS April Meeting
  - Naturally, limited to MC so far, so not particularly interesting in the context of this talk
- A couple of general notes:
  - Should be careful with treating DNNs as Deus Ex Machina to justify nonideal design choices (à la "who needs Frisch grids, can overcome long induction tails with DNN magic")
  - If want waveforms as input, scaling up to DARWIN, nEXO may be difficult
    - In EXO-200, 0.5M training events are 0.25 TB full (ROOT), but this gets down to 25 GB when cropped and pre-selected (hdf5)
    - Long-baseline LAr TPC are bigger, so are dealing with this issue already (sparse networks, reducing resolution of non-critical input, etc.)



EXO-200, nEXO, DARWIN in approx. proportions

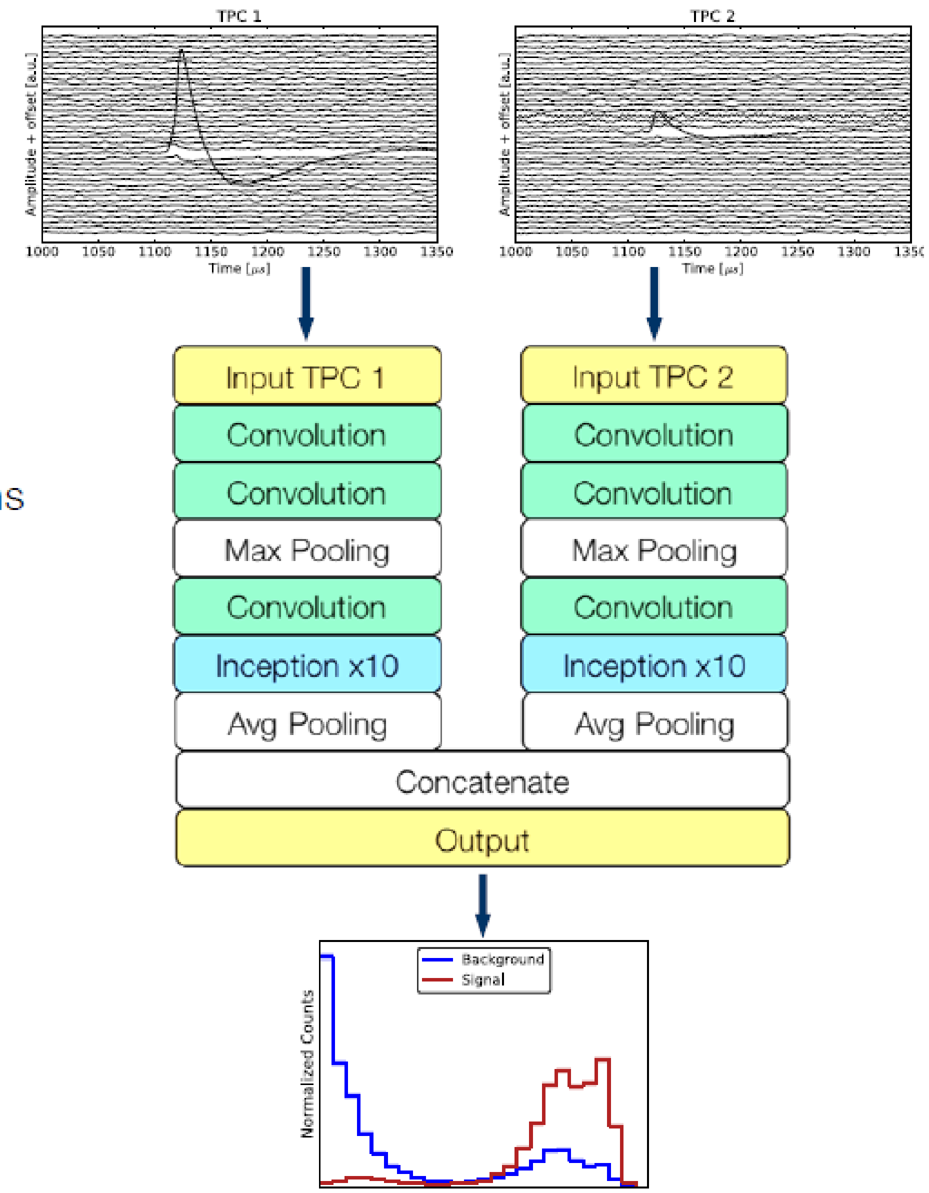
Summary

- EXO-200 has demonstrated the potential of deep neural networks for the data analysis of a 0ν experiment directly from raw data
  - Improved energy resolution compared to standard approach
  - Improved sensitivity to neutrinoless double beta decay
  - Reconstructed position using scintillation light without using MC
  - Data-driven MC of signal waveforms, faster than traditional approach
  - Validated on real detector data

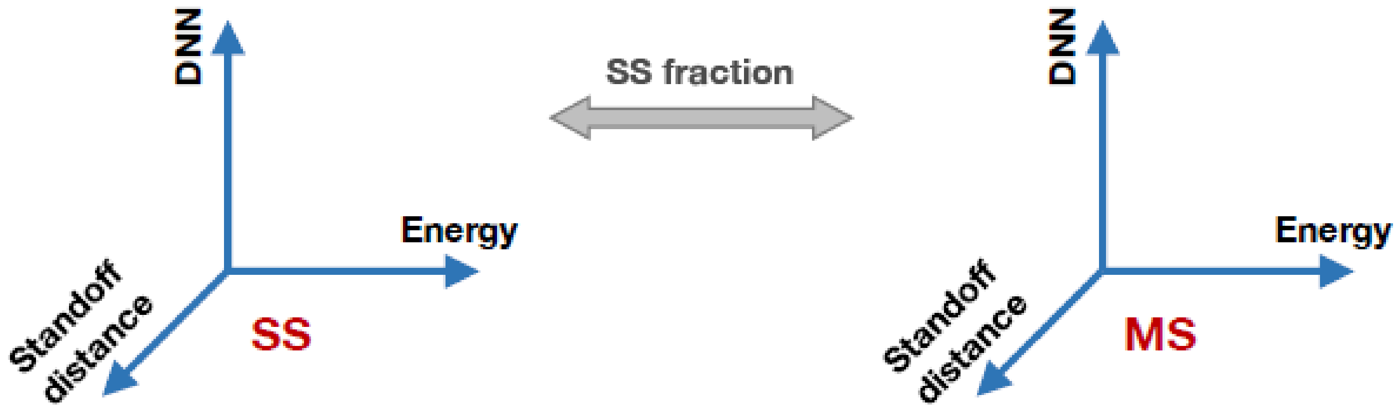
- DNNs are revolutionizing the way we do analysis
  - While the field is somewhat overhyped currently, there is no doubt that once the dust settles the CNNs will stay as a new staple tool in physicist's arsenal. Like the BDT was during the past several decades. The jury is still out for GANs, GNNs and other advanced tools
  - Can go from waveforms directly to the physics result? Still an open question. If 'yes', then could reduce the need for the dedicated experiment-specific (or even field-specific) software frameworks. The advantage is less overhead for doing physics

Backup slides

- Binary discriminator for ββ vs γ events
- Training data is identical to energy DNN
  - 50% ββ signal, 50% γ background
- MC event distributions uniform in detector volume
  - Topological discrimination only
  - No assumption on spatial distributions
- MC event distribution uniform in energy
  - validation on $2\nu\beta\beta$ data possible
- DNN architecture inspired by the Inception architecture
- Shared weights in TPC braches



- Blinded analysis performed
- SS/MS classification
- 3-dimension fit in both SS and MS events:
  Energy + DNN (topology) + Standoff distance (spatial)
  - Make the most use of multi-parameters for background rejection
  - SS, MS relative contributions constrained by SS fraction
  - Fit Phase-1 and Phase-2 separately
- Improvement of ~25% in $0\nu\beta\beta$ half-life sensitivity
  compared to using energy spectra + SS/MS alone

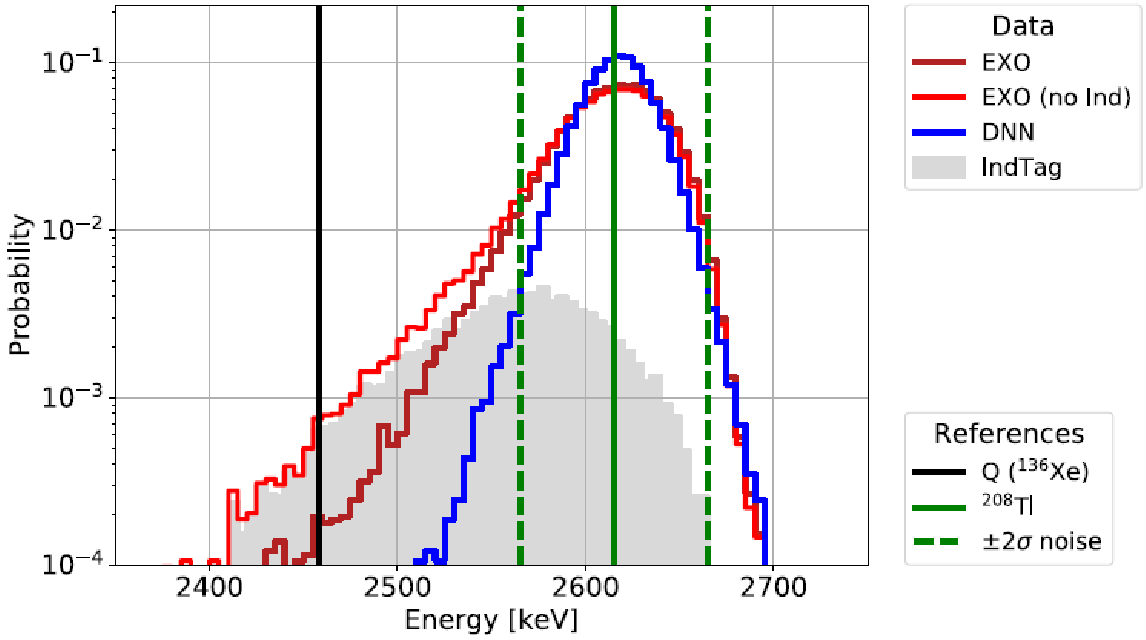


$$L = \underbrace{\mathop{\mathbb{E}}_{\tilde{x}\sim P_g}[D(\tilde{x})] - \mathop{\mathbb{E}}_{x\sim P_r}[D(x)]}_{\text{Wasserstein distance}} + \underbrace{\lambda \mathop{\mathbb{E}}_{\hat{x}\sim P_{\hat{x}}}[(\|\nabla D(\hat{x})\|_2 - 1)^2]}_{\text{gradient penalty}} \tag{3.1}$$

where D is the discriminator, $\hat{x} = \epsilon x + (1-\epsilon)\tilde{x}$, $\epsilon \in U(0,1)$, λ is the gradient penalty's weighting coefficient, and $\|\|_2$ denotes the Euclidean norm. The gradient penalty term, $(\|\nabla D(\hat{x})\|_2 - 1)^2$, encourages the norm of the gradient to go towards 1. The point x used to calculate the gradient norm is any point sampled between the GAN-generated distribution, P_g, and real data distribution, P_r. A gradient penalty is a soft version of the Lipschitz constraint that removes the undesirable behaviour of gradient explosion/vanishing when the weight clipping parameter is not carefully tuned in the earlier Wassertein GAN design.

First application: A note on the "Black box"

- The better performance of the DNN alerted that something was lacking in the "traditional" approach and triggered improvements in EXO recon
- While the cause is now largely understood (handling of mixed induction and collection signals), the developed "traditional" solution is still outperformed by the DNN



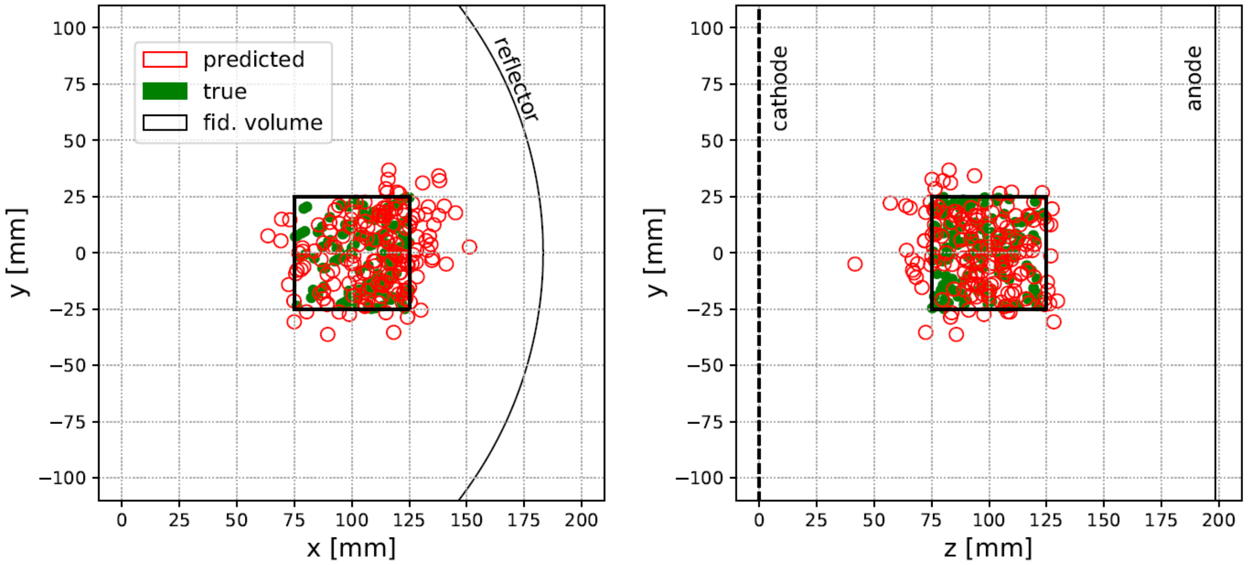


Figure 28. Evaluation of the trained model on an independent set of test data. The events are selected to be in a tight fiducial volume of 50 mm × 50 mm × 50 mm at position $x = 100$ mm, $y = 0$ mm and $z = 100$ mm and have energies above 2400 keV.