Towards Self-Supervised Optical Reconstruction in Liquid Argon Time Projection Chambers

Carolyn Hellerqvist Smith

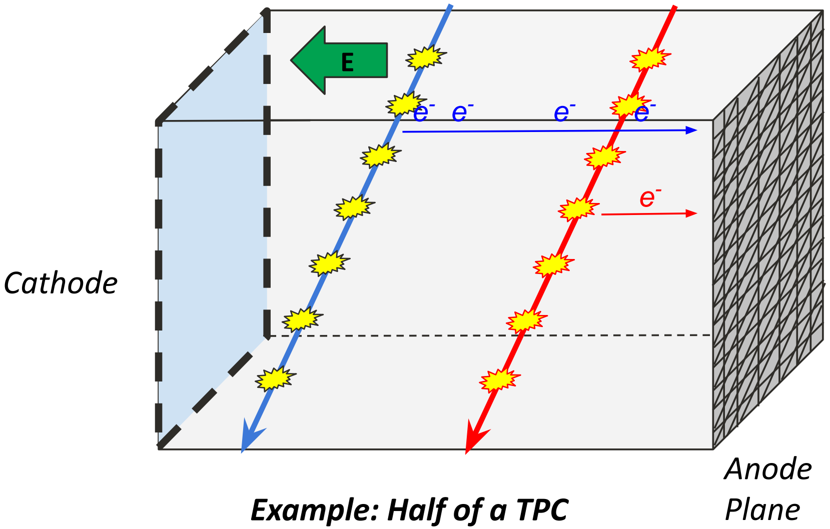
SLAC NATIONAL ACCELERATOR LABORATORY

Stanford University

Outline

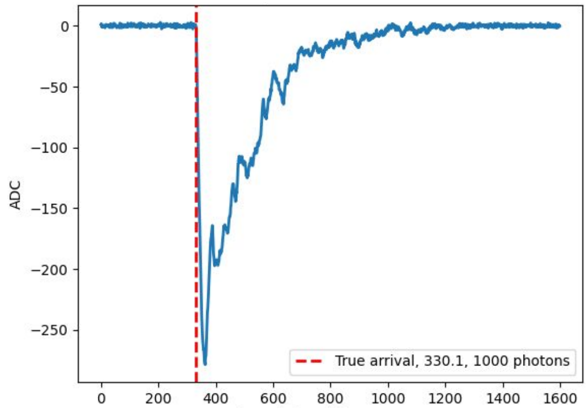
- **Motivation and Optical Reconstruction Goals**
- Data Simulation
- Supervised Models on Single-PMT Waveforms and Architecture Selection
- Self-Supervised Models on Single-PMT Waveforms
- Self-Supervised Models on Multi-PMT Waveforms

LArTPC Optical Waveforms & Pileup Challenge

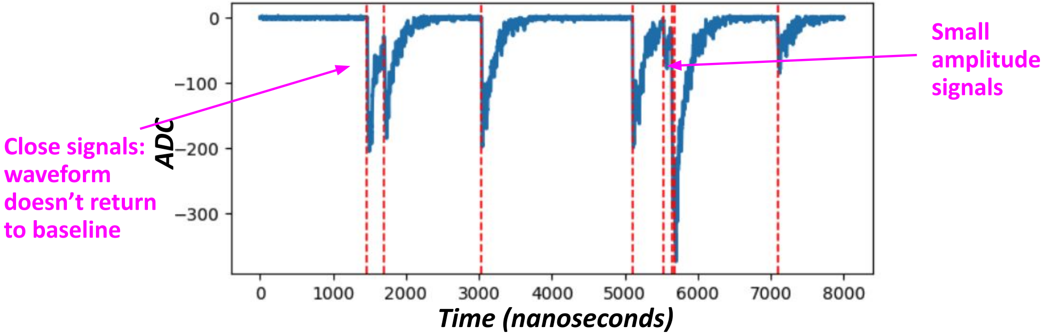


Example: Half of a TPC

Optical Signal From 1 Interaction With 1000 Photons



Example Waveform With Pileup

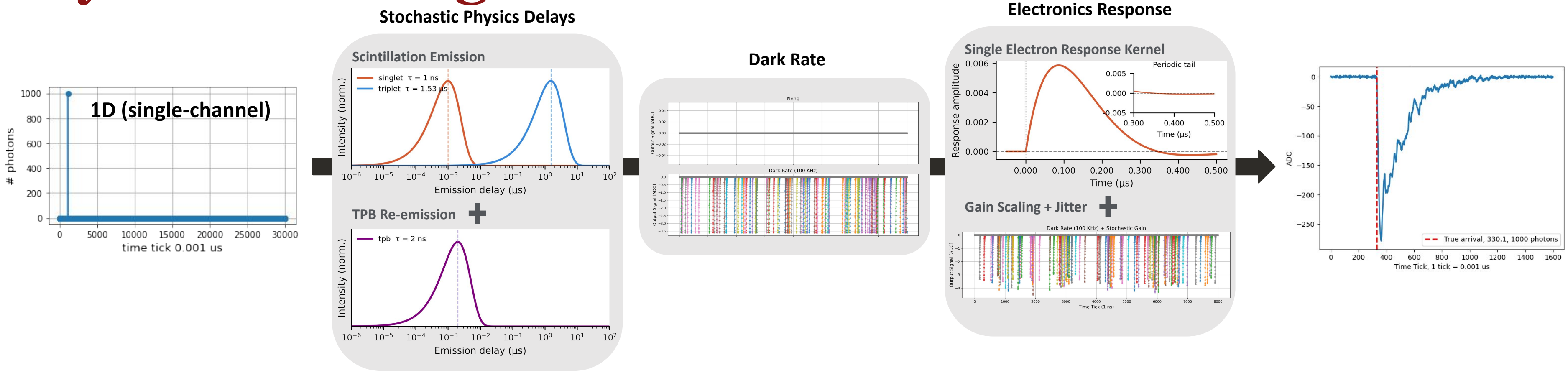


Can we build a reconstruction framework that reliably identifies individual flash arrival times in high-pileup optical waveforms?

Outline

- ➔ Motivation and Optical Reconstruction Goals
- ➔ **Data Simulation**
- ➔ Supervised Models on Single-PMT Waveforms and Architecture Selection
- ➔ Self-Supervised Models on Single-PMT Waveforms
- ➔ Self-Supervised Models on Multi-PMT Waveforms

Synthetic Single-PMT Waveform Data

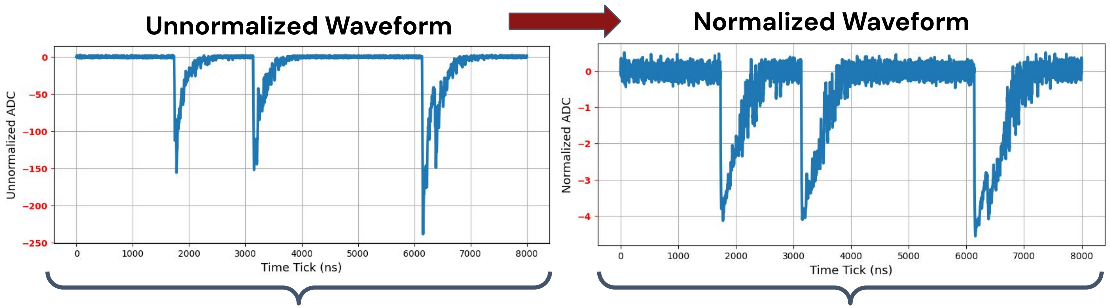


- Specify how many photons are produced by a point source directly in front of the PMT (no geometric optics/pathtracing)
- Sample stochastic delays for each photon: prompt or delayed scintillation, TPB re-emission
- Inject dark rate
- Convolve photon arrival time distribution with functional form of single-photoelectron response kernel (damped sinusoid).

For this project, we are interested in the general shape of the signal and not the amplitude scale (can vary based on gain of experiment-specific circuits).

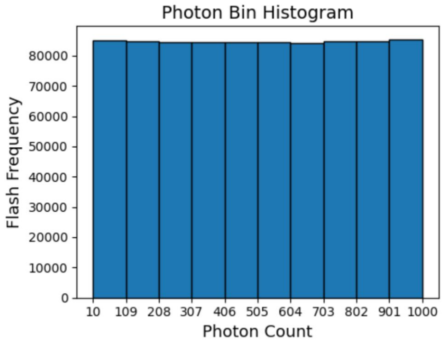
Single-PMT Dataset

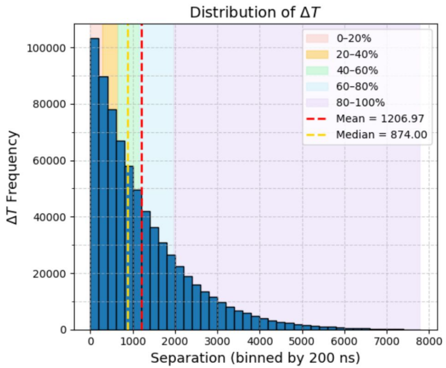
- Training set has **180k waveforms**, and validation set has **20k waveforms**.
- Each waveform (model input) has the following properties:



- 8 µs long, with 1 ns resolution
- 2-8 signals, with 10-1000 photons each
- ADC values normalized nonlinearly:

$$\mathbf{x}' = \operatorname{arcsinh}\left(\frac{\mathbf{x}}{5}\right)$$





- Dataset has a uniform distribution over photon counts per interaction
- High pileup conditions - 50% of interactions occur within 1 microsecond of another.

Outline

- ➔ Motivation and Optical Reconstruction Goals
- ➔ Data Simulation
- ➔ **Supervised Models on Single-PMT Waveforms and Architecture Selection**
- ➔ Self-Supervised Models on Single-PMT Waveforms
- ➔ Self-Supervised Models on Multi-PMT Waveforms

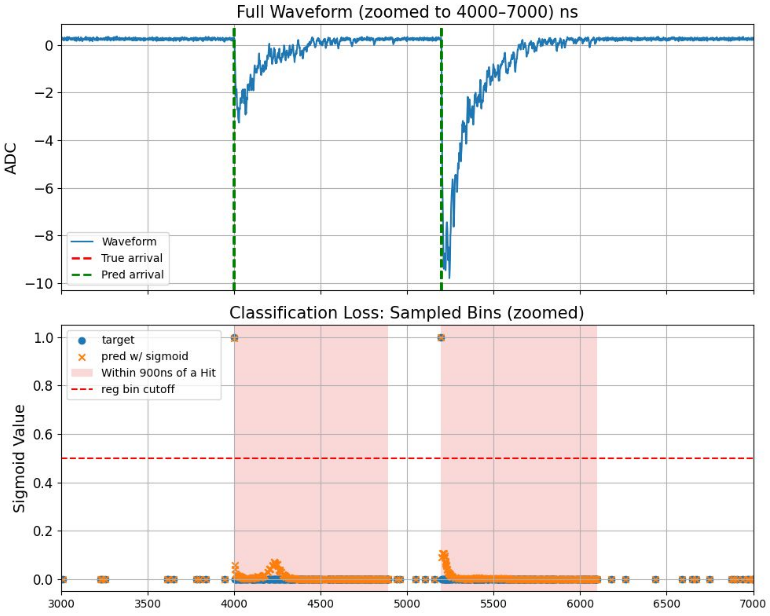
Reconstruction Task: t_0 Detection

- Framed as dense per-bin classification of interaction arrival times
- Post-processing step: merging bins within some skip tolerance to create a "t_0 window"

Final prediction is a 10 ns window that contains the true arrival time.

Red = true flash arrival time

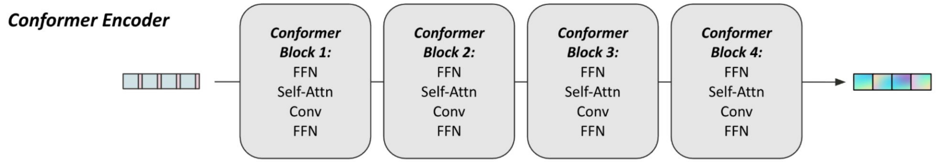
Green = positive predicted bin

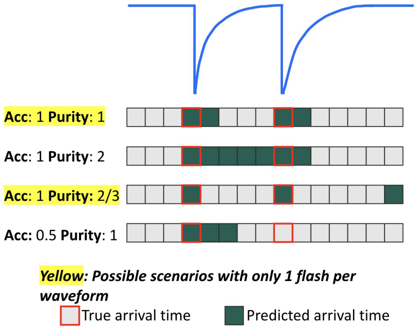


- Weighted BCE loss with hard-negative mining:
  - all positive bins
  - 500 tail negatives (red regions)
  - 100 background negatives
- Positive bins weighted by inverse class frequency.

Benchmarking Architectures

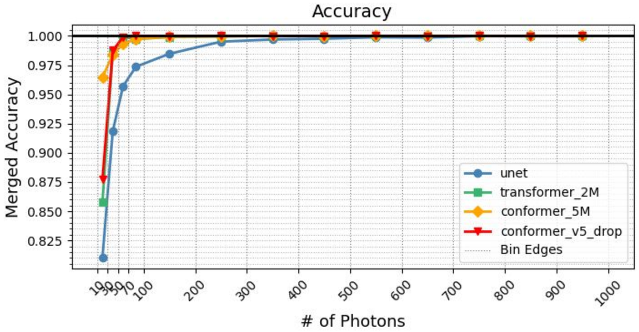
- UNet, Transformer, Conformer, Conformer + Multi-Level Tokenizer

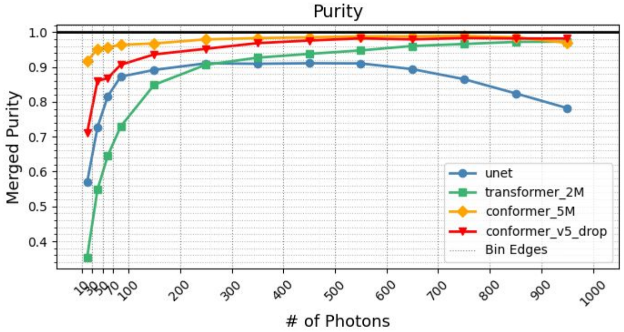


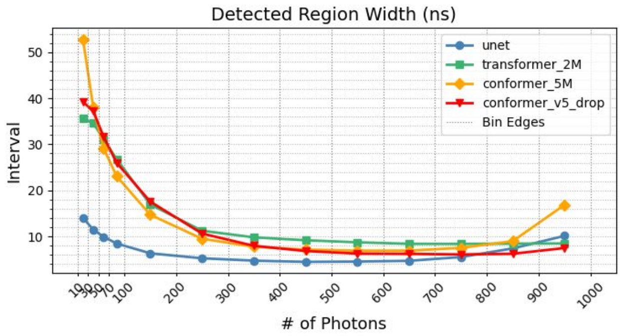


Intensity Benchmark (measure of t_0 detection performance as a function of flash intensity)

- 99,100 waveforms with exactly 1 interaction per waveform.
- Photon count Nγ ∈ [10, 1000] with 100 waveforms generated per photon value



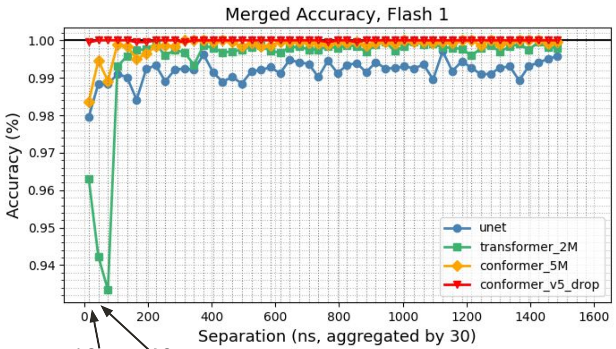




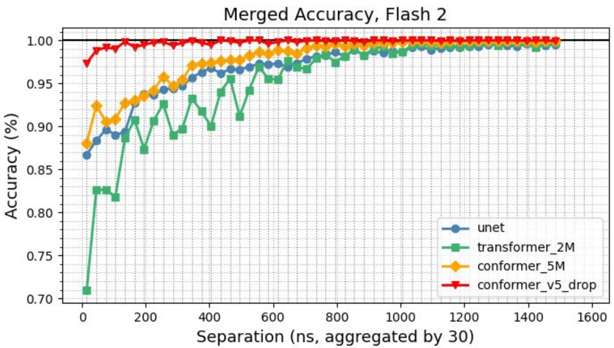
Benchmarking Architectures

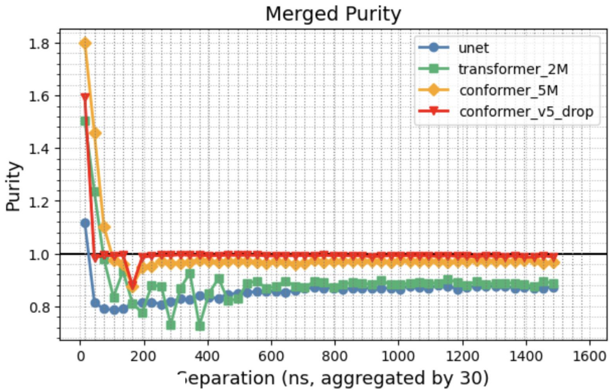
Δt Benchmark (measure of t_0 detection performance as a function of flash separation)

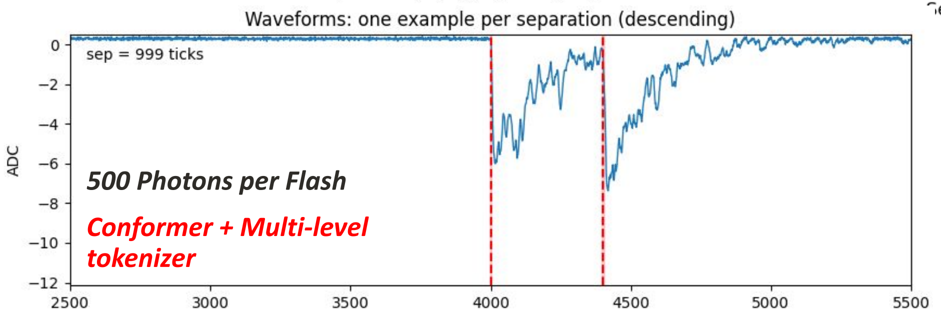
- 100,000 waveforms with exactly 2 interactions per waveform.
- Flash separation $\Delta t \in [0, 1500]$ ns uniformly.



10 ns 40 ns







🔴**True Hit Time** (1 ns)

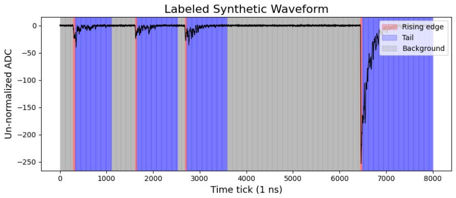
🟪**Predicted Hit Region** (can span multiple ns)

Outline

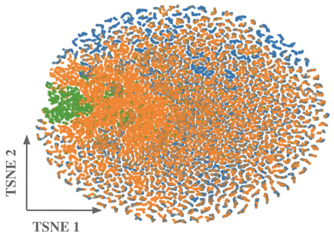
- ➔ Motivation and Optical Reconstruction Goals
- ➔ Data Simulation
- ➔ Supervised Models on Single-PMT Waveforms and Architecture Selection
- ➔ **Self-Supervised Models on Single-PMT Waveforms**
- ➔ Self-Supervised Models on Multi-PMT Waveforms

Pretraining Tasks and Representation Learning

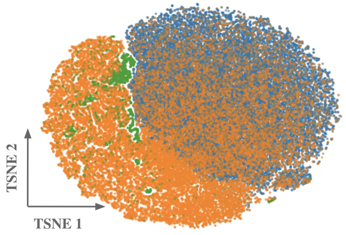
- I pre-trained the Conformer encoder on a 20k-waveform dataset using various objective combinations (masked autoencoding, self-distillation, supervised contrastive)
- Final configuration was chosen by evaluating validation embeddings with PCA and t-SNE to assess cluster quality.



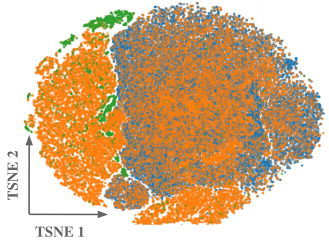
Token labels for contrastive term.



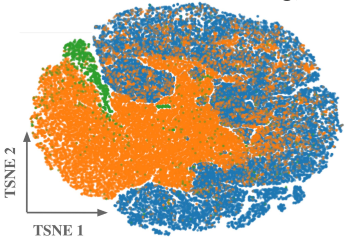
DINO (60% masking)



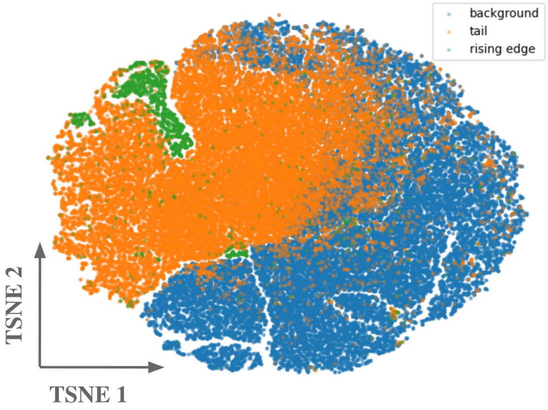
MAE (60% masking)



MAE (75%) + DINO (60%)



MAE (75%) + DINO (60%) + Cont.



For final pretraining, I used 180k waveforms and trained for 12 epochs using **MAE (75% masking) and contrastive objectives**.

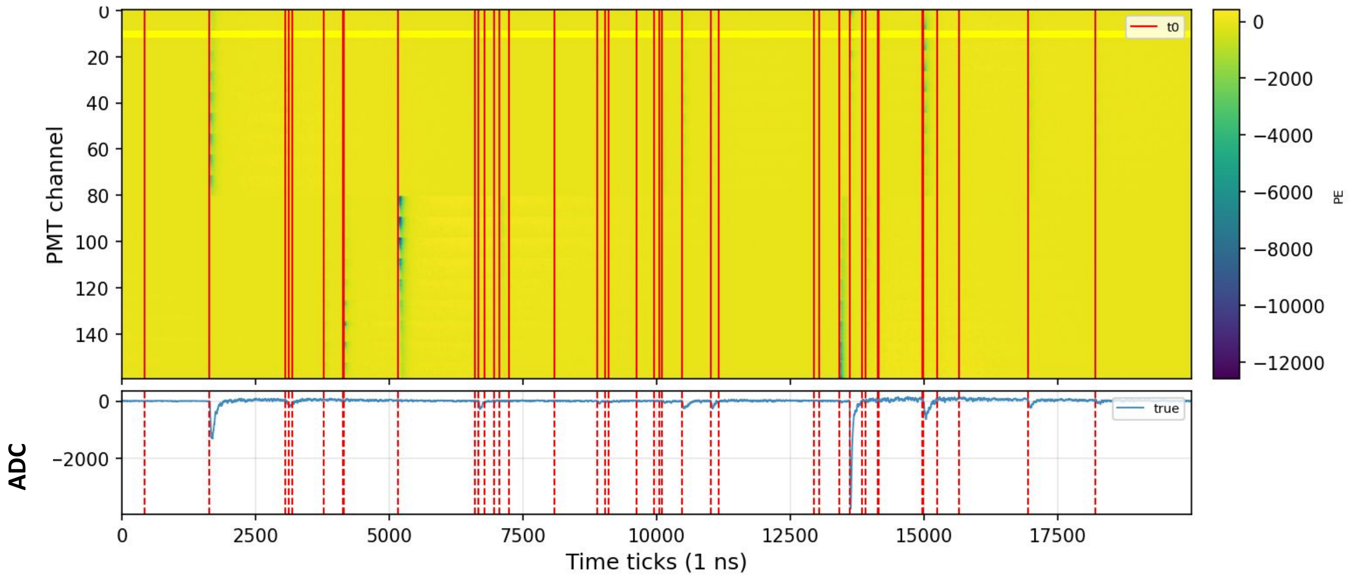
Finetuning Results

- Fine-tuning a pretrained Conformer backbone achieves comparable t_0 detection accuracy to the fully supervised model while requiring only 30% of labelled data during the downstream training stage.

Outline

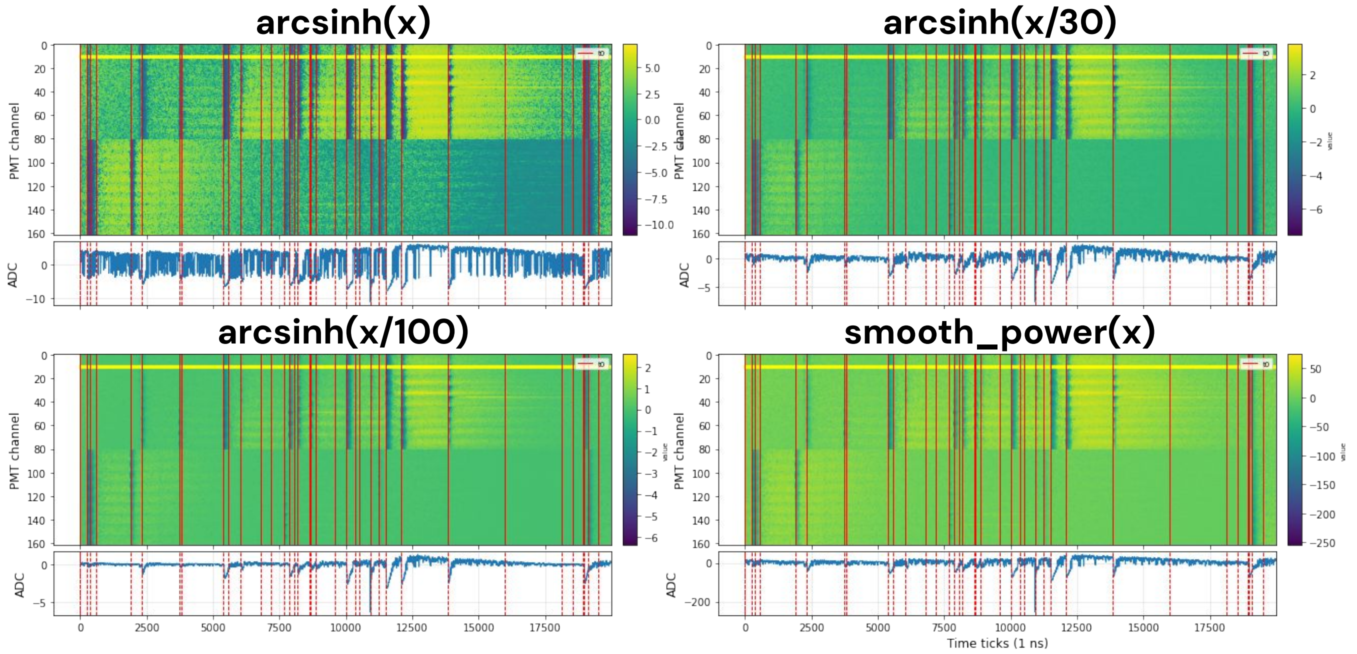
- → Motivation and Optical Reconstruction Goals
- → Data Simulation
- → Supervised Models on Single-PMT Waveforms and Architecture Selection
- → Self-Supervised Models on Single-PMT Waveforms
- → **Self-Supervised Models on Multi-PMT Waveforms**

Multi-PMT Data



- Generated using a full optical simulation (geometric optics, stochastic physics processes, electronics response)
- Range of ADC values spans ~4 orders of magnitude

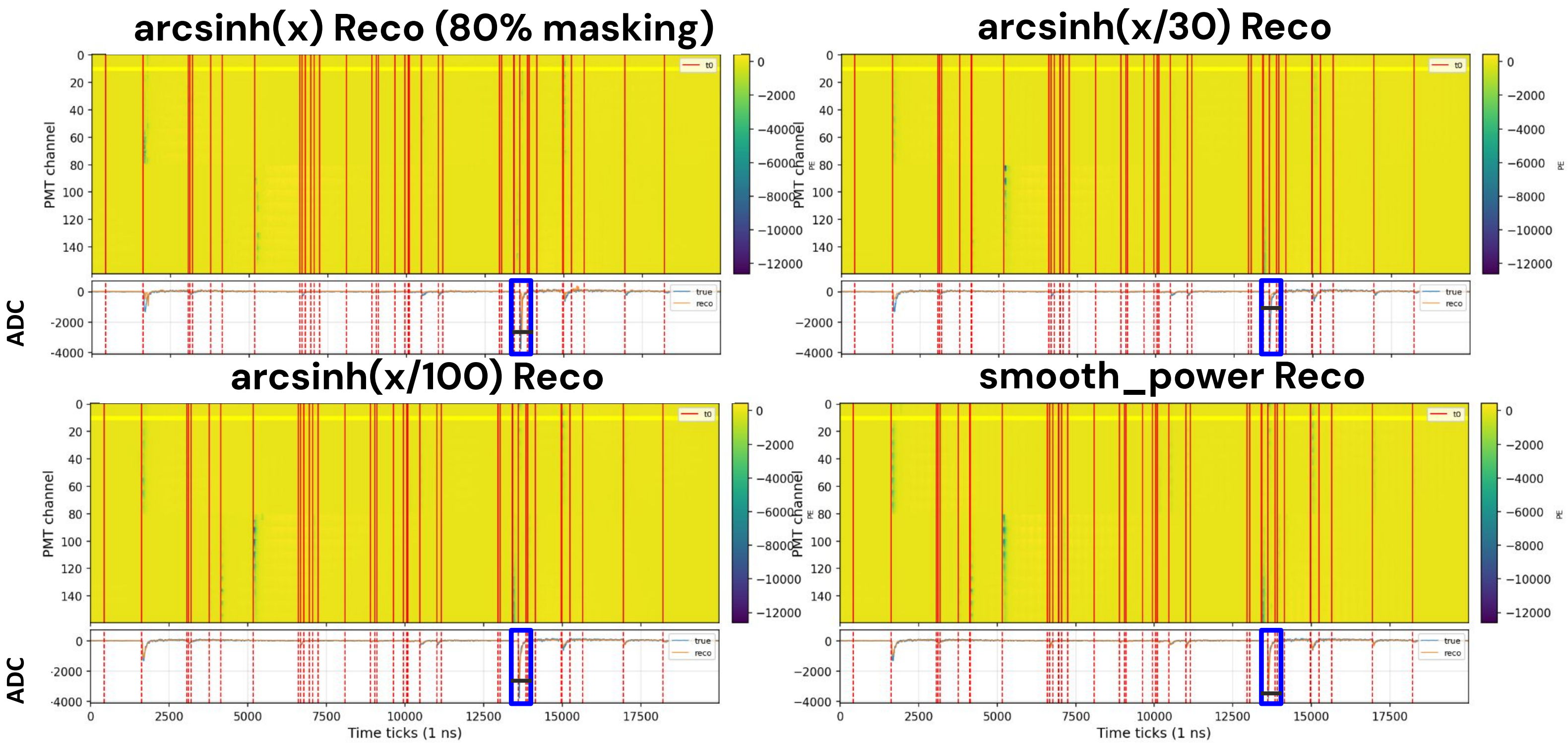
Challenge: Data Normalization



smooth_power: $f(x) = \mathrm{sgn}(x)\,\frac{\beta}{p}\left[\left(1+\frac{|x|}{\beta}\right)^{p} - 1\right]$

- β=10 ADC, p = 0.25
- Linear below noise threshold (β); $\rightarrow$ $x^{0.25}$ for large x

Masked Autoencoding Normalization Study

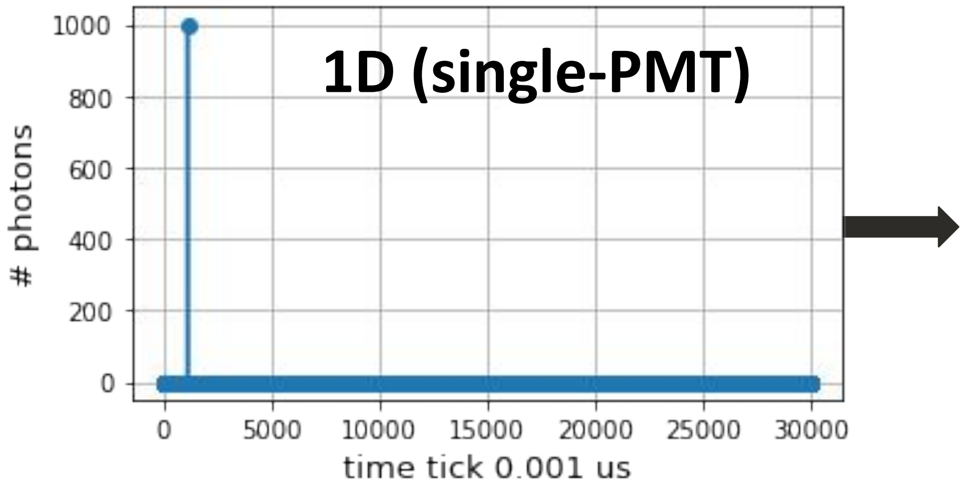


Conclusion

- Conformer-based architectures achieve the strongest t_0 detection performance for optical waveform analysis, outperforming pure convolutional and attention-based models.
- Supervised training enables strong flash localization and timing resolution for sufficiently bright interactions, even in high pileup environments.
- Self-supervised pretraining with masked autoencoding learns meaningful waveform representations and improves downstream label efficiency, achieving comparable performance with substantially less labeled data during fine-tuning.
- Extending representation learning to realistic multi-PMT waveforms introduces new challenges due to the detector's large dynamic range, but can likely be mitigated with careful data transformations.
    - This is a focus for future work!

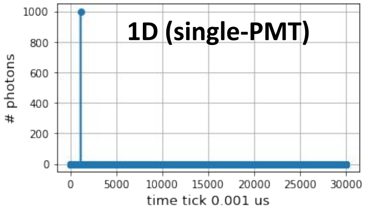
Backup Slides

Synthetic Single-PMT Waveform Data

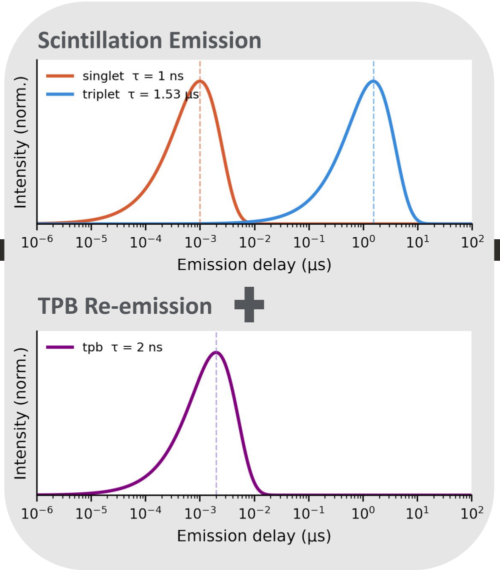


- Start by specifying how many photons are produced by a point source directly in front of the PMT (no geometric optics/pathtracing)

Synthetic Single-PMT Waveform Data

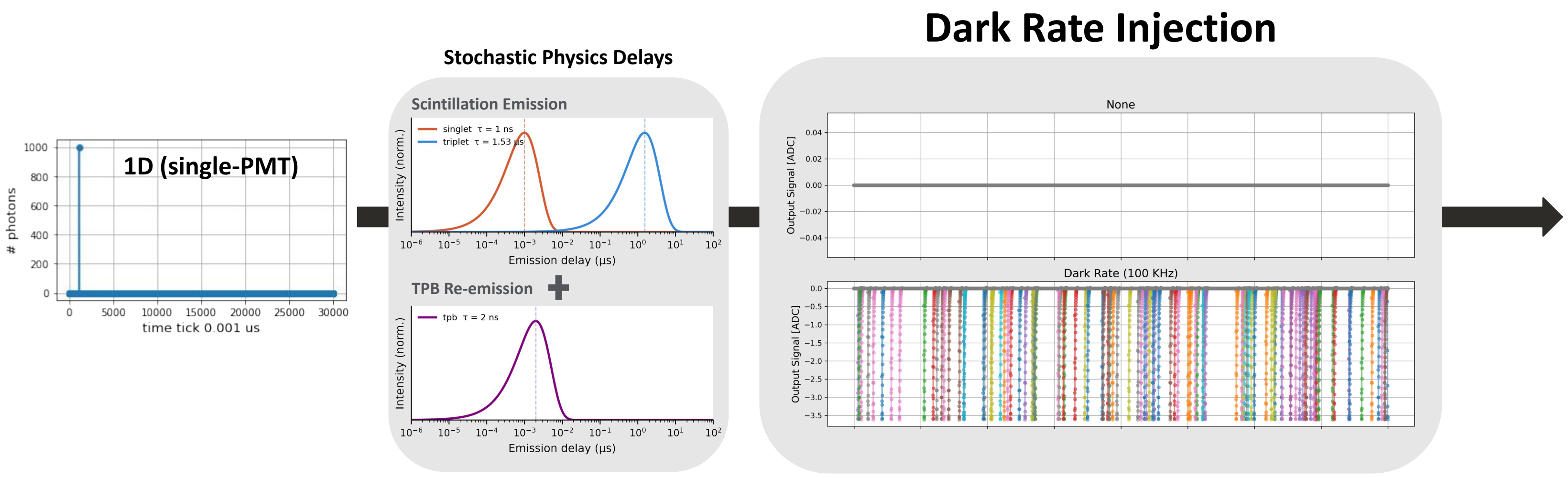


Stochastic Physics Delays



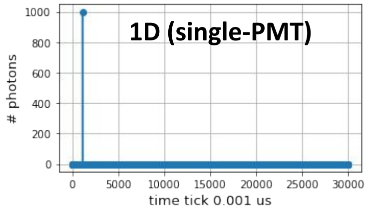
- Sample stochastic delays for each photon: prompt or delayed scintillation, and TPB re-emission

Synthetic Single-PMT Waveform Data

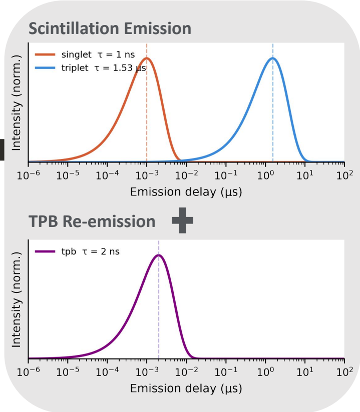


- Inject 100 KHz dark rate (aggressive)

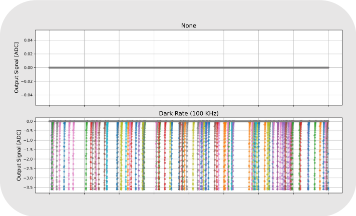
Synthetic Single-PMT Waveform Data



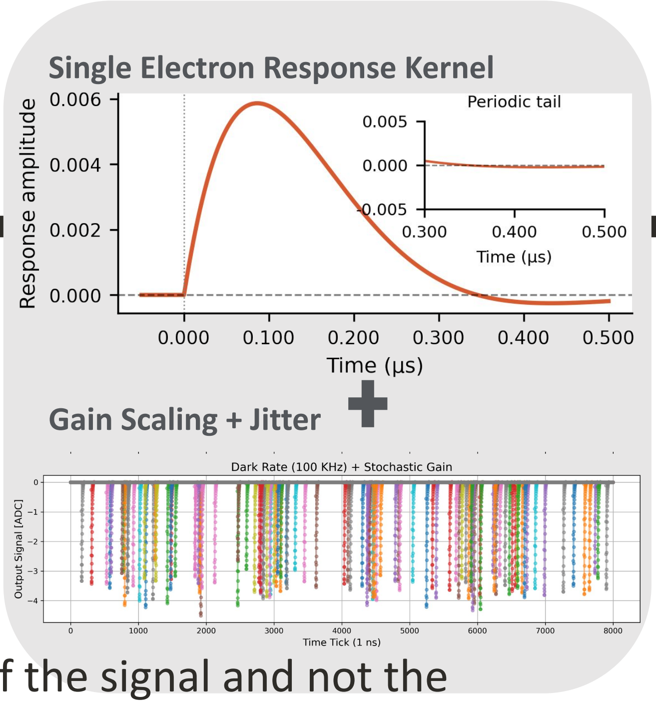
Stochastic Physics Delays



Dark Rate Injection

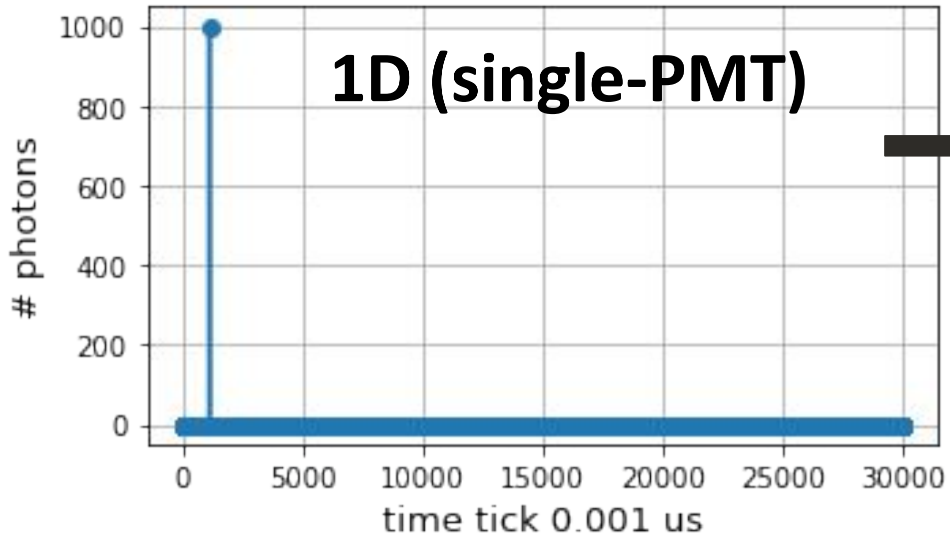


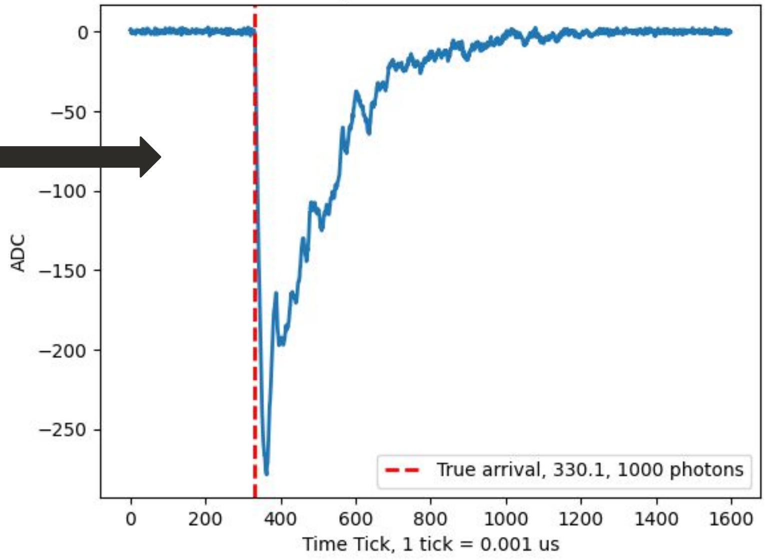
Electronics Response



- Convolve photon arrival time distribution with functional form of single-photoelectron response kernel (damped sinusoid).
- For this project, we are interested in the general shape of the signal and not the amplitude scale (can vary based on gain of experiment-specific circuits).

Synthetic Single-PMT Waveform Data



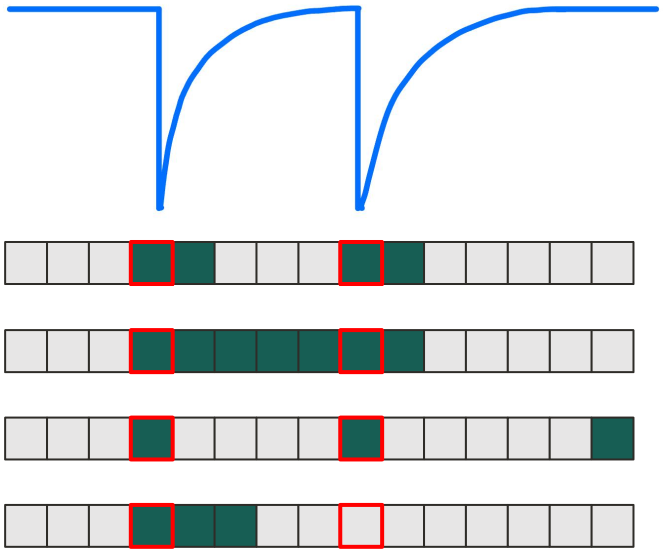


Single Flash Performance Benchmark

How does performance depend on flash amplitude? Can we handle small flashes?

Dataset: 100,000 waveforms; 1 flash per waveform; 100 per photon count (10–1000)

- **Classification Accuracy:** Fraction of true flashes that are correctly identified
- **Classification Purity**: (# correct flashes / # predicted regions)
- **Window Width**: Average size of the predicted flash region after merging
- **Reconstructed Photon Fraction**: Ratio of predicted photons to true photons



Acc: 1 **Purity**: 1

Acc: 1 **Purity**: 2

Acc: 1 **Purity:** 2/3

Acc: 0.5 **Purity**: 1

Yellow: Possible scenarios with only 1 flash per waveform

True arrival time Predicted arrival time

Double Flash Performance Benchmark

How does performance depend on proximity between flashes?
Can we handle close-together flashes?

Dataset: 100,000 waveforms; 2 flashes per waveform; 0-1500 ns apart

- **Classification Purity**: (# correct flashes / # predicted regions)
- **Window Width:** average predicted flash size

- **Accuracy (per flash)**: fraction of true flashes correctly identified
- **Photon Fraction (per flash)**: predicted / true photons