Abstract

We propose a WAveform Vector Exploitation (WAVE) deep neural network for full-waveform Time-Of-Flight (TOF) physics detectors, and evaluate its performance against traditional reconstruction techniques via Monte Carlo study of a small plastic-scintillator scatter camera. Ultralytics LLC (https://www.ultralytics.com) provides WAVE freely under the open source GPL-3.0 license at https://github.com/ultralytics/wave.

Keywords: Machine Learning, Deep Neural Networks, Time-Of-Flight, Scintillation, SiPM
Traditional TOF relies on a two step process: first timing waveform arrivals at separate locations, and subsequently reconstructing the source of the waveforms given the measured times. This technique is heavily reliant on accurate timing of the waveform arrivals. In contrast, WAVE directly solves for the waveform’s source in a single pass, without the need for any explicitly calculated features. This means that WAVE does not need to solve separately for the arrival time of either waveform, a novel concept which appears to provide performance benefits in our study.

2. Time-Of-Flight Methods

Assuming known propagation speeds \( v \) and similar waveform shapes, one may solve for the waveform origin time \( t \) and position \( x \) via Equations 1 and 2, respectively.

\[
t = \frac{t_1 + t_2 - L/v}{2}
\]

\[
x = v \left( \frac{t_1 - t_2}{2} \right)
\]

where \( t_1 \) and \( t_2 \) are the measured waveform arrival times at sides 1 and 2, and \( L \) is the distance between measurement locations. \( x_0 \) is at the center of the fiber, and \( -x \) trend towards side 1.

In our example application, Equations 1 and 2 are applied to the particle scatter camera shown Figure 1 to reconstruct the scatters within the detector volume. The detector consists of one hundred 5 × 5 × 200 mm² plastic scintillating fibers arranged in a 10 × 10 grid. Gamma or neutron scatters in these fibers create isotropic scintillation light. For our proof of concept simulations, we assume a response consistent with the ‘fast’ output of commercially available SiPMs, and assume that these voltage waveforms are digitized by DRS4 integrated circuits[8], which are differentially-driven switched capacitor array ASICs developed by the Paul Scherrer Institut (PSI) in Switzerland. The DRS4 can sample a waveform at up to 5 GigaSamples Per Second (GSPS) with a buffer depth of 1024 samples. Similar switched capacitor array chips include LABRADOR[9] and PSEC4[10].

We use a validated GEANT[11] + MATLAB[12] MC engine to model our waveforms. This MC package was created by Ultralytics for the mini-TimeCube[13] and Large Area Picoscend Photodetector (LAPPD)[14] collaborations, and allows quick prototyping of new detector designs. Waveforms modeled by this package, along with known truth information are used to analyze the TOF performance of 3 methods:

1. Constant Amplitude (CA) TOF
2. Constant Fraction (CF) TOF
3. Machine Learning (WAVE)

The MC engine places ‘perfect’ point sources randomly within the scintillating fibers, randomized in time and energy, to create a dataset of 100,000 events, which form the basis for all the results published in this paper. The timing of these point sources was randomized to start anywhere within a 3 ns window, which is a similar trigger jitter to what the DRS4 evaluation board experiences on internally triggered events. The energy of the points was randomized per an exponential distribution with a 250 keVee (keV electron equivalent) exponential constant, to mimic the low energy events a plastic ~MeV neutron scatter camera would likely target.

Digital readout noise was modeled as white and Gaussian with a standard deviation of 0.5 mV. Electronics saturation was modeled at ± 1.1 V. Single Photo-Electron (PE) waveform amplitudes in our model averaged 0.4 mV ± 10% at +5 V SiPM overvoltage. All SiPM characteristics including Photo-Detection Efficiency (PDE), spectral response, dark count rates (80 kHz/mm² at 20° C) and microcell crosstalk (though not afterpulsing) were modeled per the SensL J-60035 datasheet[15]. A Gaussian Transit Time Spread (TTS) of 120 ps 1σ was modeled for SensL J-60035 PEs as well.

Optical properties of the scintillating fibers, including scintillation spectrum, index of refraction and yield of 8000 photons/MeV were modeled using a Saint Gobain BCF-10 datasheet[5]. Optical surfaces are assumed to be perfectly flat.

Although we have modeled the full detector, the relevant quantities of comparison for this study are the capability to lo-
calize the interaction position along the longitudinal axis of the fiber, as well as the absolute time of the interaction. To that end, the comparisons between the following techniques are done on their performance for a single bar at a time, not including effects of optical crosstalk that might be present in the full detector.

2.1. Constant Amplitude (CA) thresholding

CA thresholding supplies Equation 1 with crossing times $t_1$ and $t_2$ extracted from the side 1 and 2 waveforms at the same amplitude. This amplitude is typically set low, in order to extract the very beginning of a waveform risetime, yet high enough to avoid false positives from the noise (which in our model is $0.5 \text{ mV} \sigma$). Figure 3 shows a study we performed on this threshold amplitude to determine the optimum value in our application. This search yielded a best amplitude of 6 mV, which we use in reporting our CA results.

2.2. Constant Fraction (CF) thresholding

CF thresholding supplies Equation 1 with crossing times $t_1$ and $t_2$ extracted from the waveforms at the same fraction of max amplitude. This fraction is typically set at 1/2. Figure 4 shows a study we performed to determine an optimum value for this parameter. The best results are observed around 0.3 - 0.5 fraction. 0.5 is a commonly used, so we will also use 0.5 for reporting CF results.

2.3. Machine Learning

WAVE is a 5-layer function-fitting feedforward network, shown in Figure 5. We constructed WAVE in Python using the PyTorch 0.4.0[16] package, and made it freely available under the open source GPL-3.0 license at https://github.com/ultralytics/wave. A 512-length input layer, composed of two joined waveforms, connects to 3 hidden layers of size 76, 23, and 7, with an output layer of size 2: 1 position $x$ and 1 time $t$. Side 1 and Side 2 waveforms shown in Figure 2 of length 256 each are joined sequentially to create each 512-length input vector.

All hidden layer neurons are linear (weights + bias), with tanh() activations, except layer 4, which lacks activation. This structure was settled on using trial and error, though the exact geometry is not critical, as we observed that our Section 3 results were robust to moderate changes in network geometry. Our 100,000 event dataset was split into 3 groups:

1. **Train (70%).** Used during WAVE training to update network parameters (weights and biases).
2. **Validate (15%).** Used to stop WAVE training to prevent over-training. Not used to update network parameters.
3. **Test (15%).** Used to evaluate performance of CA, CF and WAVE. Not used during WAVE training.

We used an Adam optimizer with near-default hyperparameters ($lr = 0.001$, $\epsilon = 1 \times 10^{-8}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$) and a Mean Squared Error (MSE) loss criterion. Supervised learning (on the Train and Validate datasets) was performed on a Google Cloud Platform (GCP) Compute Engine Virtual Machine (VM) instance with one Nvidia Tesla P100 GPU, taking ~5 minutes of wall-clock time for 50,000 epochs, and yielding a minimum validation loss of 0.023 after 14,278 epochs. The epoch 14,278 model was frozen for inference on the Test dataset to produce the Section 3 results.
3. Results

Results are summarized in Table 1 and detailed in Figure 6 as a function of scatter energy. WAVE performance exceeds CA and CF in our study by a wide margin in both position and time reconstruction. The largest performance increases are observed at the lowest energy events, the most challenging region of the parameter space to reconstruct.

| method | candidates fraction | position $x$ (mm) $1\sigma$ | time $t$ (ps) $1\sigma$ |
|--------|---------------------|-------------------------------|------------------------|
| CA - 6 mV | 0.95 | 38 | 62/ |
| CF - 0.5 | 1.00 | 26 | 150 |
| WAVE  | 1.00 | 11 | 90 |

Table 1: MC study results (lower 1$\sigma$ is better), showing WAVE exceeding CA and CF performance in both spatial and temporal reconstruction. This table summarizes Figure 6, which plots the above metrics vs scatter energy.

4. Conclusion

Table 1 and Figure 6 show clear performance increases for WAVE compared to traditional techniques across all regions of the parameter space evaluated in our study.

The next steps are the application of this technique to real-world data. Challenges to real-world implementation include the need to accumulate training data of sufficient quantity and variation to generalize well in real-world scenarios. For applications where such training data is attainable, it appears WAVE may offer significant performance benefits over traditional solutions.

5. Acknowledgements

This effort was initiated and funded by Ultraalytics LLC (https://www.ultraalytics.com) in collaboration with the University of Hawaii and Lawrence Livermore National Laboratory. Lawrence Livermore National Laboratory is operated by Lawrence Livermore National Security, LLC, for the U.S. Department of Energy, National Nuclear Security Administration under Contract DE-AC52-07NA27344. Release LLNL-JRNL-761634.

References

[1] W. Adams, A. Elagin, H.J. Frisch, R. Obaid, E. Oberla, A. Vostrikov, R.G. Wagner, J. Wang, M. Wetstein. Timing characteristics of Large Area PicoSecond Photodetectors. Nuclear Instruments and Methods in Physics Research Section A, Volume 795, 21 September 2015, Pages 1-11, ISSN 0168-9002, http://dx.doi.org/10.1016/j.nima.2015.05.027
[2] A. Abashian, et al., The Belle detector, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, Volume 479, Issue 1, 21 February 2002, Pages 117-232, ISSN 0168-9002. http://dx.doi.org/10.1016/S0168-9002(01)02013-7
[3] Wei-Dong Li, Ya-Jun Mao, and Yi-Fang Wang, Int. J. Mod. Phys. A 24, 9 (2009). http://dx.doi.org/10.1142/S0217751X09046424
[4] Grabas, H. and Obaid, R. and Oberla, E. and Frisch, H. and Genat, J.-F. and Northrop, R. and Tang, F. and McGinnis, D. and Adams, B. and Wetstein, M., “RF strip-line anodes for psec large-area mcp-based photodetectors” Nuclear Instruments and Methods, A 711 (0) (2013) 124-131 http://dx.doi.org/10.1016/j.nima.2013.01.055
[5] Saint Gobain Scintillating Fibers website
Figure A.7: Results vs scatter position (in fiber). Edge effects appear in all 3 fitting techniques, though interestingly WAVE is the only technique to show improved performance near the fiber edges. WAVE outperform CA and CF across all regions of this parameter space (scatter position space), as it does in Figures 6 and A.8.

Figure A.8: Results vs scatter time. No significant edge effects are observed. WAVE outperform CA and CF across all regions of this parameter space (scatter time space), as it does in Figures 6 and A.7.

Figure A.9: WAVE training loss vs number of epochs trained. Validation and Test losses reach a minimum around 20,000 epochs, after which they begin to increase. Test losses continue to decrease. Training is terminated after 5,000 validation epochs fail to register a new absolute minimum.
Figure A.10: Hyperparameter study: effect of learning rate on best validation loss and epochs required to reach best validation loss. The default learning rate of 0.001 appears to be well placed, as it allows for a near optimal final loss in a relatively short training time.

Figure A.11: Activation function study. 5 activation function candidates are evaluated for use in WAVE. ELU and Tanh both appear well suited to the problem. Softsign produces the worst results.

Figure A.12: Hidden layer structure study. Different hidden layer structures were analyzed, starting at 1 hidden layer with 23 neurons and moving to 5 hidden layers with 181 neurons in the first layer, 64 in the second layer, etc. The largest performance increase was observed when moving from 1 to 2 layers, with corresponding training speed improvements as well, yet with diminishing returns beyond 3-4 layers.

Figure A.13: WAVE pytorch training speed on various hardware, measured in epochs per second. About 20,000 epochs are required for training, corresponding to ~100 seconds on a Nvidia Tesla P100 GPU.