Stress Testing of Spiking Neural Network-based TDC-less dToF

Jack I. MacLean[1], Brian D. Stewart[2], Istvan Gyongy[1]

[1]The University of Edinburgh, Institute for Integrated Micro and Nano Systems, EH9 3BF, Edinburgh, U.K.
[2]STMicroelectronics Imaging Division, EH3 5DA, Edinburgh, U.K.
istvan.gyongy@ed.ac.uk Tel: +44 131 651 9031

***Abstract*— Spiking Neural Networks (SNNs) are naturally suited for processing discrete events, such as those produced by single-photon avalanche diode (SPAD) sensors, with the promise of lower energy consumption, hardware resources and latency compared with traditional artificial Neural Networks (ANNs). We explore a recently proposed direct Time-of-Flight (dToF) scheme that uses an SNN with a Legendre Memory Unit (LMU) architecture for processing SPAD data. We "stress-test" this SNN in simulations, evaluating its performance in operating conditions beyond those considered during training, such as multi-surface signals, broader SPAD instrument response functions (IRFs), and elevated ambient levels. The SNN shows good robustness in these new operating regimes.**

I. INTRODUCTION

Combining arrays of single-photon avalanche diodes (SPADs) with integrated processing enables solid-state direct time-of-flight (dToF) sensors to be built, with the potential of compact, robust, and low-cost 3D sensing solutions for applications ranging from smartphones to robotics and automotive. To manage the high data rates produced by SPADs, architectures with on-chip histogramming have been proposed [1]. However, the associated memory and power requirements limit the scalability of such architectures. Partial histogramming can overcome scalability issues, but at the cost of reduced laser power efficiency and increased susceptibility to motion artifacts [2]. Furthermore, as in full histogramming, high-frequency reference clocks are typically used for photon timing, and hence the chip power consumption can still be significant. In recent years, several alternatives to and variants of histogramming have been proposed, including spline-based sketches [3], count-free histograms [4], histogram-free processing of photon time stamps [5], and neural network-based processing [6].

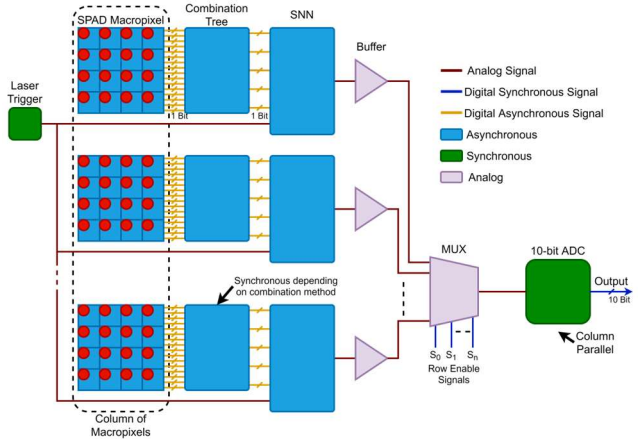

Figure 1. Proposed architecture of TDC-less dToF sensor (reproduced from [7]) showing a column of pixels. Each pixel has a 4×4 array of SPADs followed by a combination tree (e.g., an asynchronous adder), whose output is presented to an SNN. The outputs of the SNNs are in turn fed to a column parallel ADC.

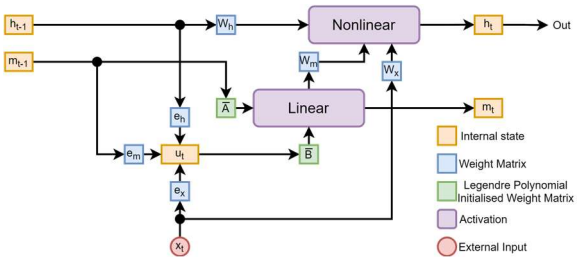

Figure 2. Block diagram of LMU SNN [8], reproduced from [7]

This paper follows-on from [7], where a new single-photon dToF processing scheme was developed using spiking neural networks (SNN) with Legendre Memory Unit (LMU) architecture [8]. The LMU operates by implementing a memory cell which approximates the unit delay ($y(t) = u(t - \theta)$) such that any input seen within θ seconds is represented within d memory units and can be approximately reproduced. This behaviour allows for the network to store all the SPAD events over the whole exposure to make its final prediction.

The scheme (Figures 1 and 2) benefits from the low-power capabilities of SNNs [9], together with the ability of LMUs to analyse long sequences of time series data. In tests using simulated SPAD data, the scheme is found to have lower precision than conventional (centre-of-mass, CMM) dToF processing on histogram data, but higher overall accuracy, successfully learning to compensate for SPAD saturation effects (Figure 3). When run on measured data (Figure 4), the SNN is seen to successfully reproduce the main features of a mannequin's head, despite having been trained on synthetic data only.

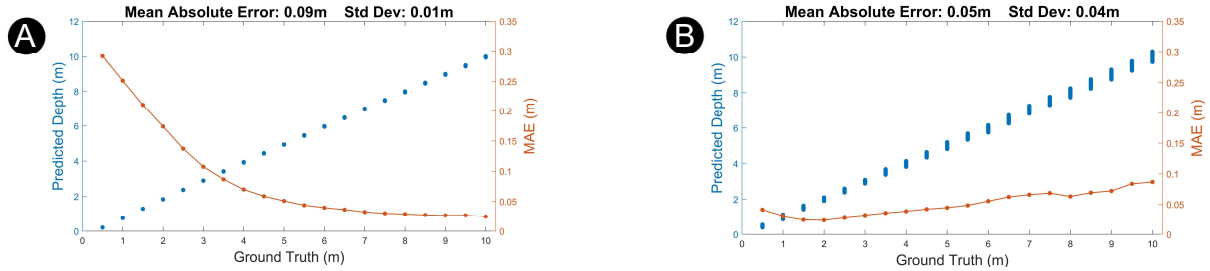

Figure 3. Depth estimation of CMM (panel a) and SNN processing (panel b), based on synthetic photon data. The graphs include scatter plots of estimated depth (blue points) as well as the mean absolute depth error (orange) calculated using repeated simulations at a range of assumed distances. The CMM results assume histogram generation (with bin size 500 ps) prior to depth processing. The following system parameters were assumed: 45 cycles of 940 nm laser source with 4 ns pulse duration and 640 nJ pulse energy. SPADs are taken to be 10 µm in size; an f1.2 objective with a 10 nm bandpass filter is presumed to be in front of the array. The ambient level is 1 klux and object reflectivity is 25%.

II. ARCHITECTURE

The memory cell m is implemented using Legendre polynomials in a state space representation, as shown below:

$$\dot{m}(t) = A\, m(t) + B\, u(t)$$

Where $m(t)$ represents the current state of the memory cell at time t, $u(t)$ the current input, while A and B are matrices of the state space representation which are initialised using Legendre polynomials:

$$A = [a]_{ij} \in \Re^{d\times d}, \quad a_{ij} = (2i+1), \quad \begin{cases} -1 & i < j \\ (-1)^{i-j+1} & i \geq j \end{cases}$$

$$B = [b]_i \in \Re^{d\times 1}, \quad b_i = (2i+1)(-1)^i \quad i,j \in [0, d-1]$$

Where d is the order of the memory cell m. The input signal can be reproduced using the state space equation:

$$y(t) = C\, m(t) + D\, u(t)$$

The C matrix needed to reproduce the input signal from θ seconds ago is simply all ones, However the C matrix can be modified to produce outputs for different delays within $0 \leq \theta' \leq \theta$. The D matrix is always zero. This operation can be visualised using Figure 5, where the response for an ANN and an SNN memory cell of the order 56 stores and then reproduces a SPAD signal, resulting from synchronous summation technique (SST) [10], for a θ of 3.87 µs.

The LMU network uses this memory cell to store previously seen information in a compressed format.

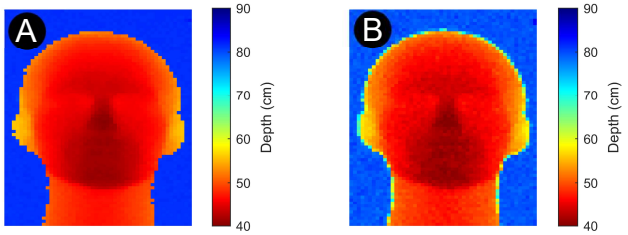

Figure 4. Depth maps generated by applying CMM (panel a) and SNN (panel b) processing to real single photon LIDAR data from [3], converted into SPAD events. Good correspondence can be seen even though the IRF is non-gaussian, compared with the gaussian IRF used in SNN training.

Another ensemble of neurons, the hidden units, can then use the compressed representation to make a prediction based on all the information seen within the last θ seconds. In theory, as the SNN is making predictions based on a sliding window of a set number of laser cycles, then once the network reaches a steady state it will be capable of producing a surface depth prediction after every laser cycle.

The network is trained using the Python package Nengo-DL [11], which trains the SNN using surrogate gradient descent based on approximated firing rates of the networks neurons. Nengo-DL uses Tensorflow as the backend enabling the network to be trained using a GPU, which in this case was an RTX 3090.

III. RESULTS

Figure 6 plots the mean error in the depth estimate in the case when the SPAD macropixel observes two distinct surfaces. The plots indicate that the SNN reports a weighted average of the two surfaces, with the weighting favouring the surface that is closer to the sensor (as expected, due to the higher return signal from the latter). End effects are noted when the 2nd surface is close to the sensor. In Figure 7 heat maps are used to capture, for the single-surface case, the variations in the mean error (panel a) and standard deviation (std) of the error (panel b) as the SPAD IRF and ambient level are increased beyond the 100 ps, and maximum of 30 klux that the SNN was trained for. As the IRF is made wider, an increased bias is observed in the depth estimate, although it is somewhat negated by an apparent bias in the opposite direction at high ambient levels. On the other hand, the widened IRF is seen to have a relatively minor effect on the std of the error (quantifying the precision of the depth estimates). Although the std does increase with the ambient level, the increase is seen to be linear. Overall, no sudden, drastic deteriorations in performance can be observed from the graphs.

In case STT (rather than an asynchronous adder) is used to combine SPAD events, it becomes useful to assess the impact of reducing the sampling frequency, with no retraining. Fig. 8 shows a deterioration in accuracy (panel a) but only a minor decline in precision (panel b).

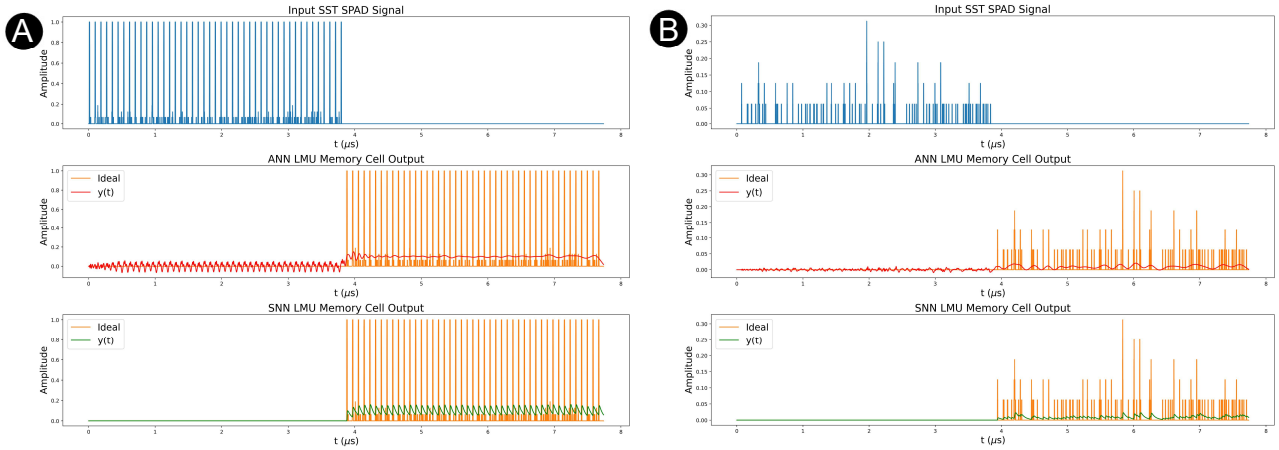

Figure 5. Operation of the 56 ordered memory cell if *u(t)* is a series SPAD events combined using SST with a window size θ of 3.87 µs (corresponding to 45 laser cycles). The cases of high signal level (panel a) and lower signal level (panel b) are shown. The ideal signal recreation of the input signal (blue) using the current state of the memory cell is shown in orange. The approximated recreation using an ANN version of the network is shown in red while the recreation using an SNN is shown in green. Neither ANN nor SNN version can make a perfect recreation of the input signal, as they can only store a finite amount of information. The recreation of the input can be improved by increasing the order d of the memory cell, at the cost of more neurons and therefore an increased demand on power and silicon area.

Figure 9 depicts the evolution in the SNN output during exposure for two pixels in the mannequin data. We note that a reasonable depth map is produced even early in the exposure, raising the possibility of "early sampling" when a high signal level is sensed.

The power consumption of the SNN is estimated to be in the nW range, assuming typical firing rates seen in simulations, and the generation of depth estimates at video rates [7]. As regards to the area of a potential circuit implementation, we note that compact (0.04 um^2) leaky integrate-and-fire (LIF) neurons have been proposed in advanced technology nodes (7nm FinFET) that can fire at the 1 GHz rates required here [12]. Noting that the SNN presented here has 610 neurons and 3506 synapses, and assuming a synapse size similar to that of neurons, a total circuit area in the region of 164 um^2 is estimated, which would fit under a typical 40×40 um SPAD macropixel [1].

IV. CONCLUSION

These preliminary results suggest that the SNN is reasonably robust to unfamiliar operating conditions, which bodes well for applications such as a low power object detection for robots or drones, avoiding the need for re-training, which can be time-consuming. Modifications are being considered to the SNN architecture to enable longer ranging distances, and the detection of multiple surfaces. In addition, adjustments will be explored to the window size and operation to enable faster and more accurate depth prediction after every laser cycle.

ACKNOWLEDGMENTS

Jack MacLean's PhD research was supported by STMicroelectronics and the Scottish Research Partnership in Engineering under Grant SRPe-IDP/017.

REFERENCES

[1] Gyongy, I., Dutton, N. A., & Henderson, R. K. (2021). Direct time-of-flight single-photon imaging. *IEEE Transactions on Electron Devices*, 69(6).
[2] Taneski, F., Al Abbas, T., & Henderson, R. K. (2022). Laser power efficiency of partial histogram direct time-of-flight LiDAR sensors. *Journal of Lightwave Technology*, 40(17).
[3] Sheehan, M. P., Tachella, J., & Davies, M. E. (2024). Spline sketches: An efficient approach for photon counting lidar. *IEEE Transactions on Computational Imaging*, 10.
[4] Ingle, A., & Maier, D. (2023). Count-free single-photon 3D imaging with race logic. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
[5] Tontini, A., Mazzucchi, S., Passerone, R., Broseghini, N., & Gasparini, L. (2023). Histogram-less LiDAR through SPAD response linearization. *IEEE Sensors Journal*, 24(4).
[6] Milanese, T., Zhao, J., Hearn, B., & Charbon, E. (2023). Histogram-less direct time-of-flight imaging based on a machine learning processor on FPGA. *In Proc. IISW 2023*
[7] MacLean, J. I., Stewart, B., & Gyongy, I. (2024). TDC-less direct time-of-flight imaging using spiking neural networks. *IEEE Sensors Journal*, 24(20).
[8] Voelker, A., Kajić, I., & Eliasmith, C. (2019). Legendre memory units: Continuous-time representation in recurrent neural networks. *Advances in neural information processing systems*, 32.
[9] Yamazaki, K., Vo-Ho, V. K., Bulsara, D., & Le, N. (2022). Spiking neural networks and their applications: A review. *Brain sciences*, 12(7), 863.
[10] Patanwala, S. M., Gyongy, I., Mai, H., Aßmann, A., Dutton, N. A., Rae, B. R., & Henderson, R. K. (2021). A high-throughput photon processing technique for range extension of SPAD-based LiDAR receivers. *IEEE OJSSCS*, 2.
[11] Rasmussen, D. (2019). NengoDL: Combining deep learning and neuromorphic modelling methods. *Neuroinformatics*, 17(4).
[12] Mushtaq, U., Akram, M. W., Prasad, D., & Islam, A. (2024). An energy and area-efficient spike frequency adaptable LIF neuron for spiking neural networks. *Computers and Electrical Engineering*, 119.

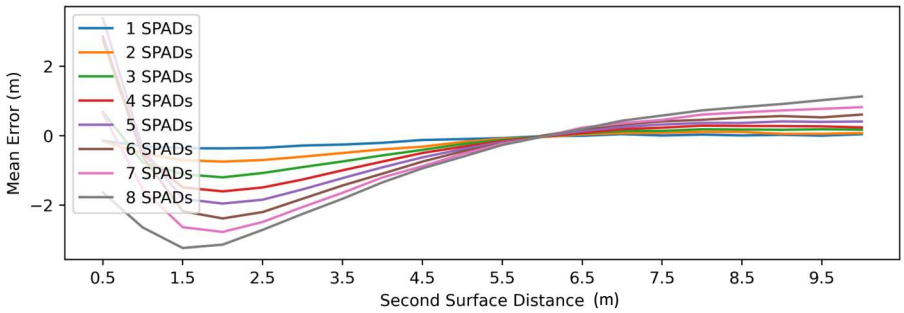

Figure 6. Mean error in SNN depth estimate of primary surface (situated at 6 m distance) in the case of a second surface (at different distances). Each plot corresponds to the secondary surface covering a different number of SPADs (within the 16 SPADs constituting a pixel). Surface reflectivities of 40% and an ambient level of 25 klux are assumed.

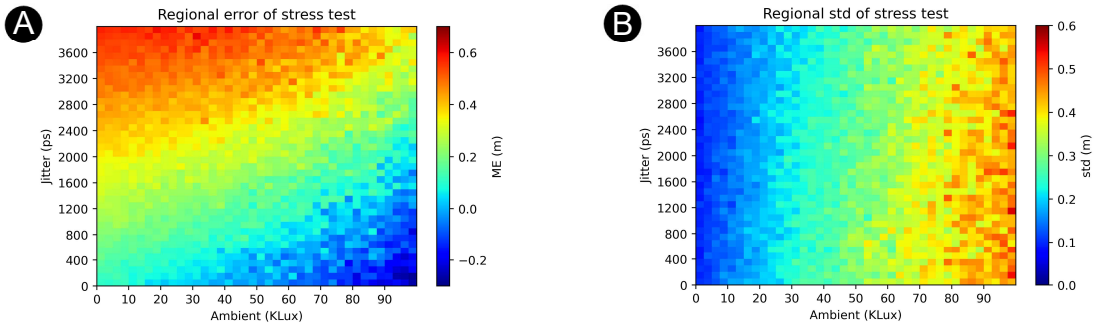

Figure 7. Heat maps of the mean error (panel a) and standard deviation of the error (panel b) for varying levels of SPAD IRF width and ambient level. A target distance of 10 m with surface reflectivity of 40% is assumed.

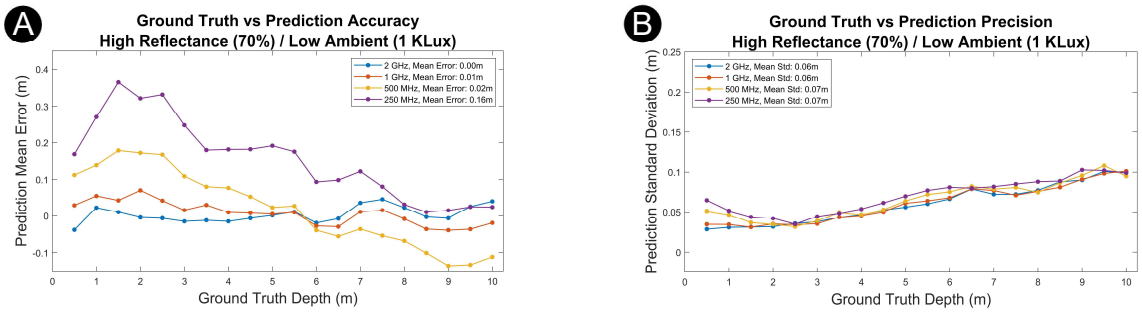

Figure 8. Ranging performance of SNN, based on synthetic data, in the case STT is used for SPAD combination, and the sampling frequency is decreased from the nominal 2 GHz used in training. The performance is quantified in terms of accuracy (panel a) and precision (panel b); the same system parameters are assumed as in Figure 4. In terms of accuracy, we observe only modest a change at 1 GHz, but a marked decrease at lower frequencies (this may be countered by re-training the network). On the other hand, the sampling frequency does not have a significant impact on the depth precision, apart from at close range (possibility to the SPAD IRF becoming compressed due to SPAD saturation).

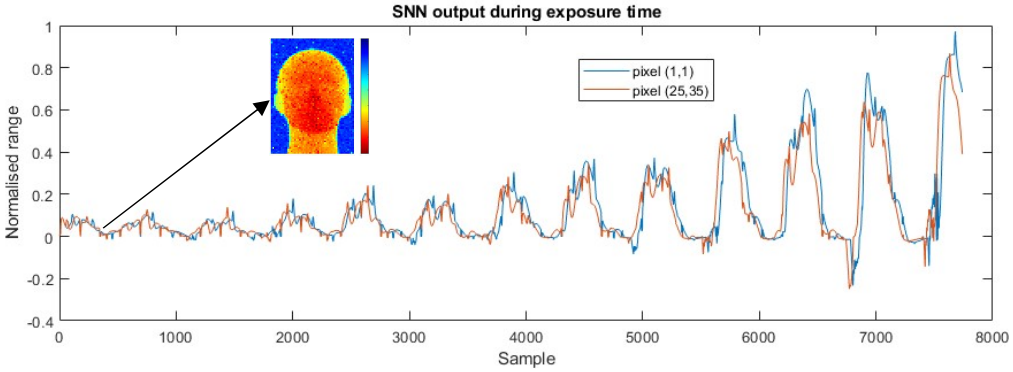

Figure 9. Output of SNN for the same data as in Figure 3. The outputs corresponding to two pixels (one in the background, and one in the middle of the head) are shown. Whilst the output is seen to oscillate over the exposure time, reasonable depth maps may be obtained even early on during the exposure (such as at sample 364 as depicted) suggesting that early sampling of the SNN output may be useful in some circumstances.