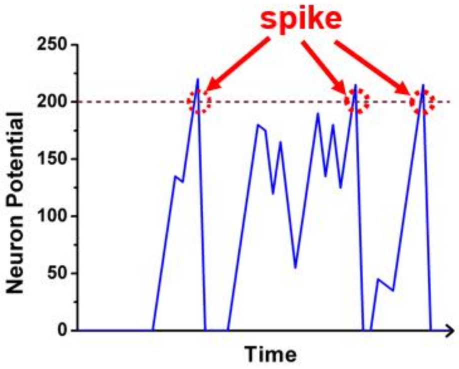
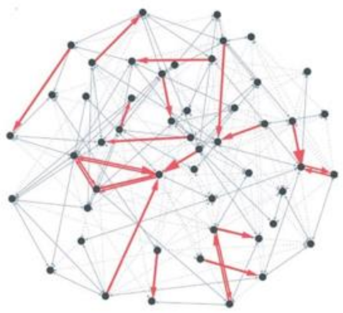
# Efficient Neuromorphic Computing with Dynamic Vision Sensor, Spiking Neural Network Accelerator and Hardware-Aware Algorithms

Jae-sun Seo
Arizona State University

# Event-Driven Neuromorphic Computing

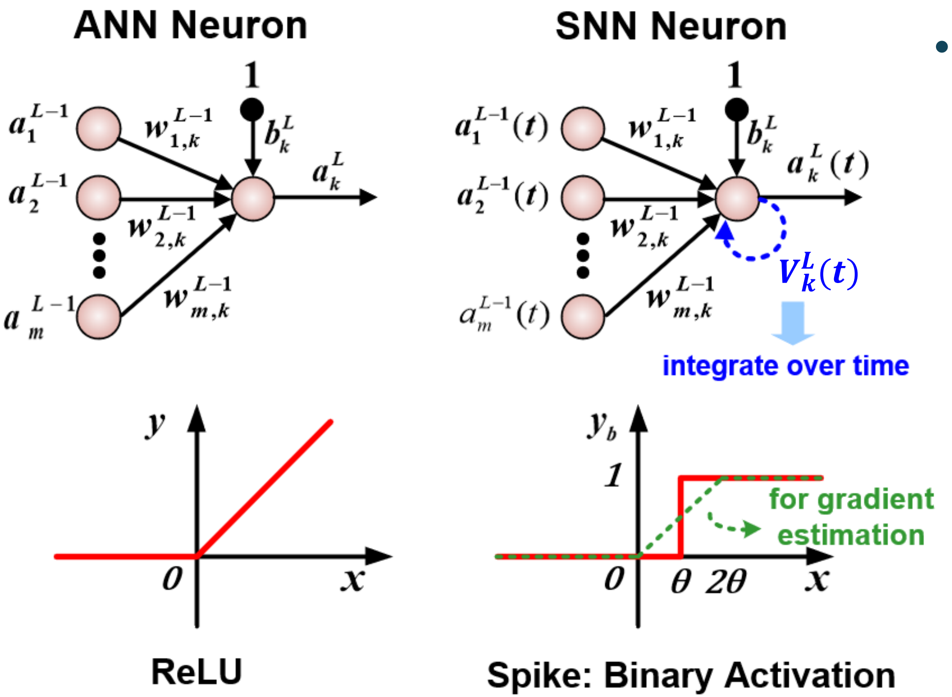
- Neurons in biological nervous systems communicate with 'spikes' → binary pulses
- Spike represents events in spiking neural networks (SNNs)
- Spiking activities are sparse
- Major dynamic energy only consumed when spikes occur
- Spike-based, event-driven computing: Low energy consumption



[1] S. Song, PLoS Biology, 2005

# Artificial vs. Spiking Neuron Model




[2] S. Yin et al., BioCAS 2017

- Neuron behavior: integrate & fire
  - Integrate membrane potential ($V_k$) over time with weighted sum of spikes & weights

$$V_k^L(t) = V_k^L(t-1) + \sum_{i=1}^{m} a_k^{L-1}(t) \times w_{i,k}^{L-1} + b_k^L$$

  - If membrane potential crosses threshold,

$$V_k^L(t) > \theta$$

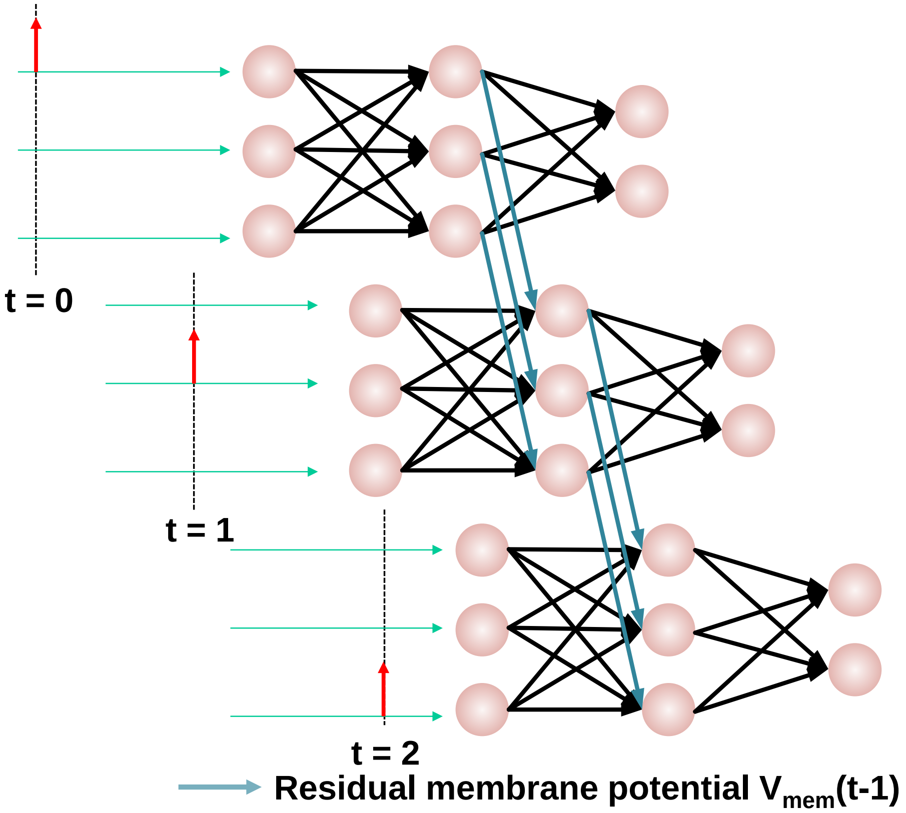
    - Neuron spikes

$$a_k^L(t) = 1$$

    - Membrane potential resets

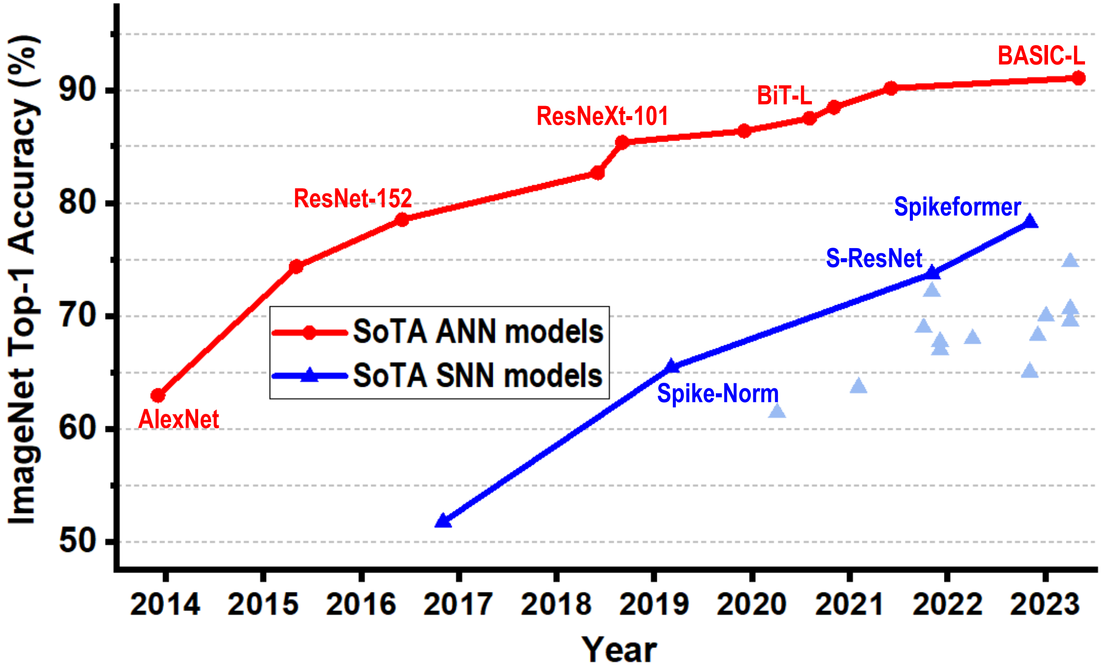
$$V_k(t) = V_{reset}$$

# SNN Training with Temporal Information



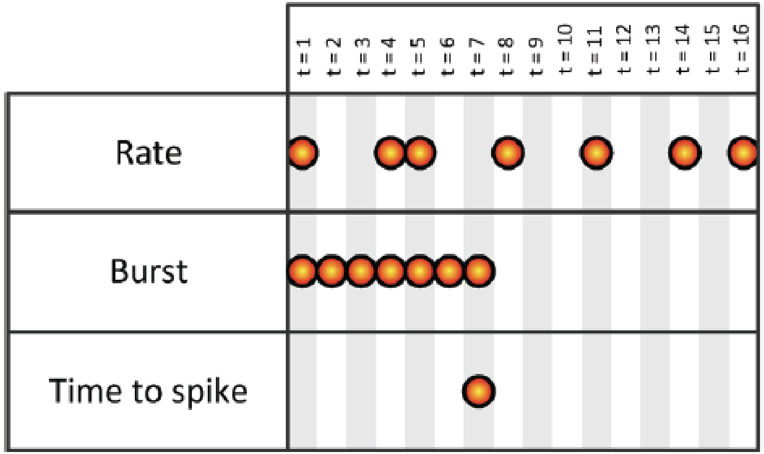

- Conventional training:
  - 2-D tensor for input
  - batch + neurons
- SNN training w/ time:
  - 3-D tensor for input
  - batch + neurons + time
- Difference with RNN:
  - Include all time steps with same importance (membrane potential integration)

# ANN vs. SNN Accuracy for ImageNet Dataset

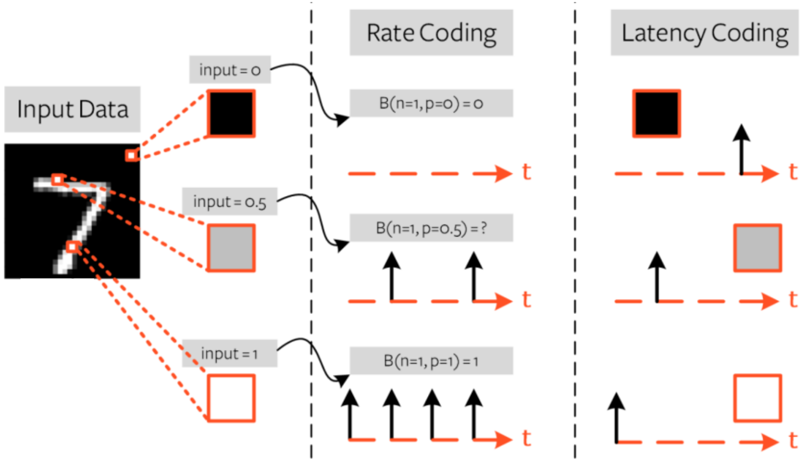



- Similar to ANN's success on accuracy improvement for ImageNet dataset, recently there have been large improvements on SNN algorithms for ImageNet

# Spike Encoding for Input Images




[3] A. Andreopoulos, et al.,
IBM Journal of Res. & Dev., 2015



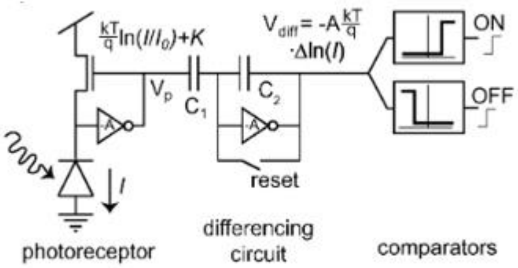

Source: https://snntorch.readthedocs.io/

- Natural images (for computer vision tasks) are not in spiking format
  → each input image needs to be converted into spikes via rate, burst, or latency encoding
- Such spike encoding implementation adds overhead on latency, energy, and area.

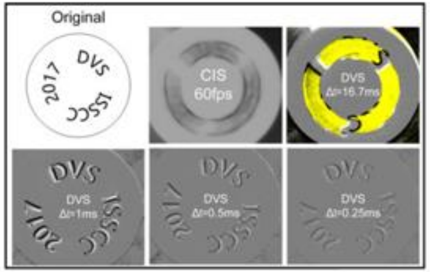
# Event-based Vision Sensor for SNNs

- If front-end sensor has outputs that are spikes, it can directly connect to SNNs
- Dynamic vision sensor, or event-based camera, sends spikes only when events occur e.g. changes in pixels (instead of sending full image at 30 fps irrespective of events)
  → *fast event detection possible with low power*

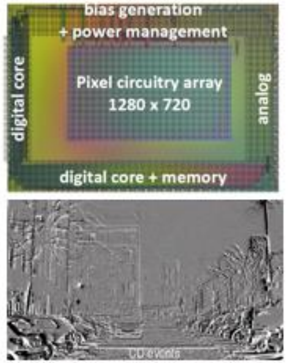
**Dynamic vision sensor (DVS) or Event-based vision sensor**




128x128
[4] ETH Zurich, JSSC 2008



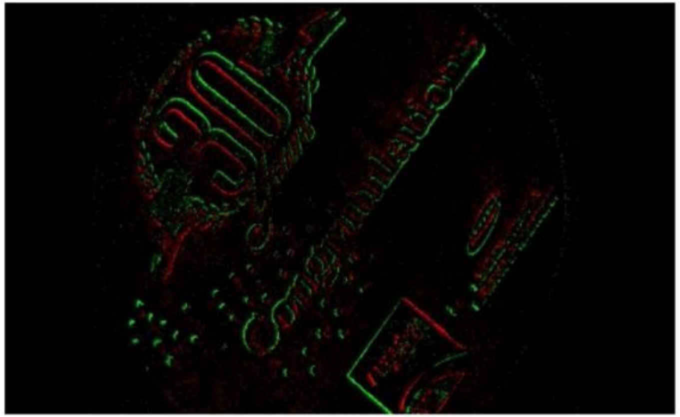

640x480
[5] Samsung, ISSCC 2017




1280x720
[6] Prophesee, ISSCC 2020

# Advantages of Event-based Camera

CIS image for rotating object (~220 rpm)
Represented with 8.3 msec time resolution
(120 fps)

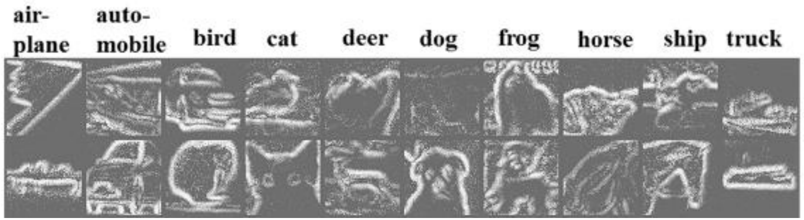


DVS image for rotating object (~220 rpm)
Represented with 0.5 msec time resolution
(effectively 2,000 fps)

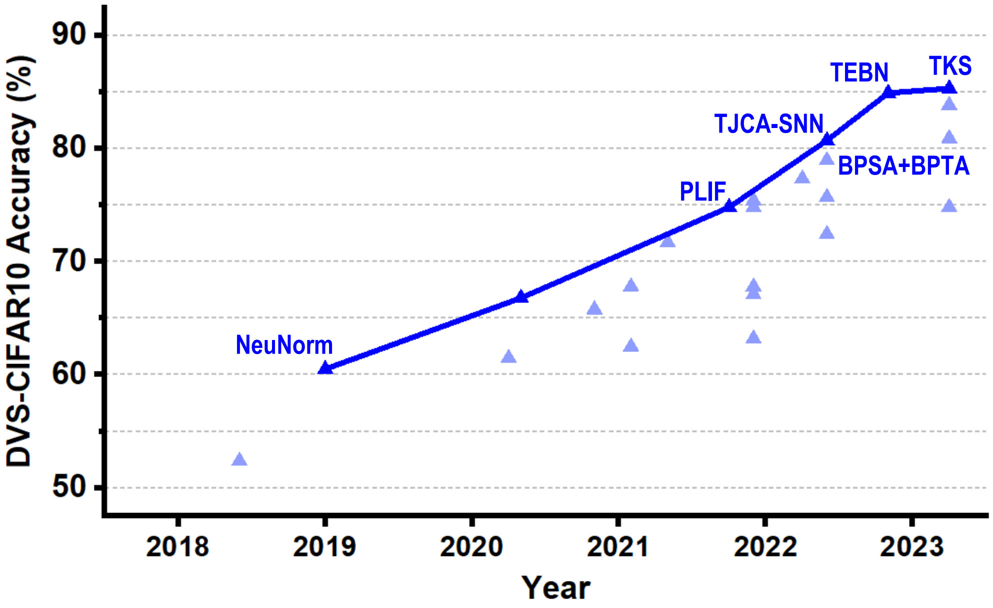
[5] B. Son, ISSCC, 2017

- Suitable applications: fast moving object detection, control based on it, etc.
- Conventional CIS will always have blurry images for fast moving objects

# SNN Accuracy for DVS-CIFAR10 Dataset




[7] H. Li, Front. Neuroscience, 2017

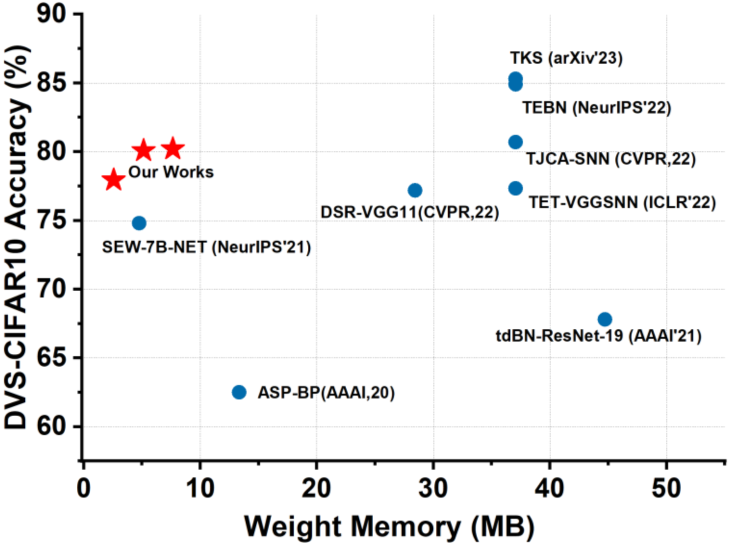



- DVS-CIFAR10: dataset of event-stream recordings (done with DVS camera & image movement) for classification of 10 different objects

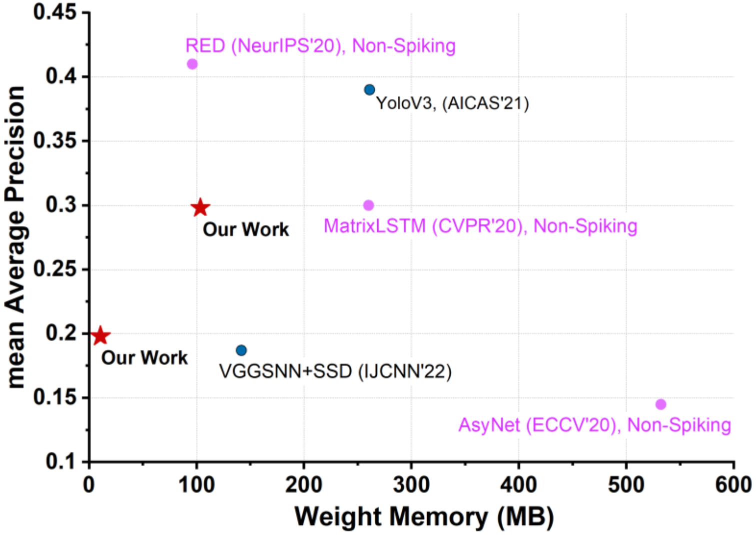
- Various training techniques have been largely improving the SOTA accuracy for event-based DVS-CIFAR10 dataset

# SNN Accuracy vs. Model Size

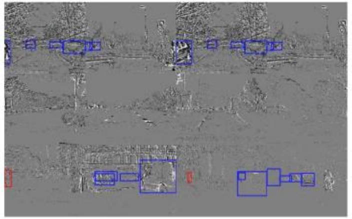
DVS-CIFAR10 Dataset




Prophesee Gen1 Dataset



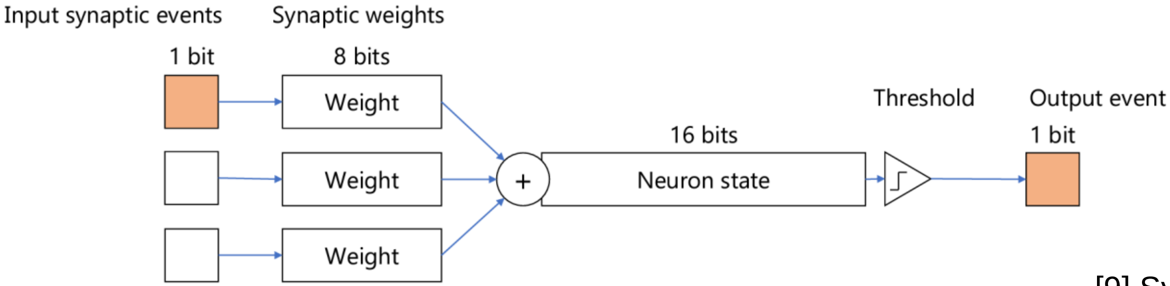

Prophesee Gen1 dataset:
- 228,123 cars
- 27,658 pedestrians
- 39 hours, recorded w/ Prophesee Gen1 sensor (1M pixel)



[8] E. Perot, NeurIPS, 2020

- For on-device SNN inference, model size and total memory footprint are important
- Many SNN algorithms have been using FP32 precision to achieve high accuracy
- Our proposed SNN works: learnable threshold with low-precision quantization

# Total Memory Footprint for SNNs



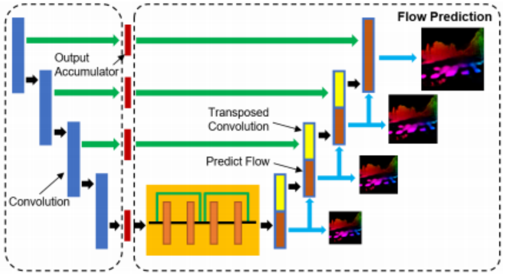

[9] SynSense, TinyML Talks, 2021

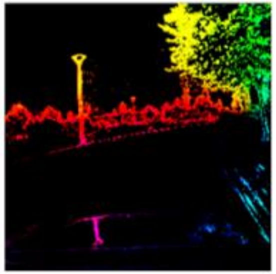
- For ANNs/DNNs, you don't use the previous frame's high-precision weighted sum value in the current/future frames, so don't need to store the weighted sums for each neuron
- For SNNs, events occur over time, and you need to accumulate membrane potential (neuron state) over time for every neuron in the entire SNN
  - Every neuron's membrane potential is different → need to store each of them in entire SNN

- **Total SNN memory** = (weight_precision × # of weights) + (1-bit × # of neurons) + (mem_pot_precision × # of neurons) + etc.
- Membrane potential memory: more significant for activation-heavy SNNs (e.g. high input resolution)
  - For Prophesee Gen1 dataset, memory of VGG-11+SSD [10] - weight: 35MB, mem. pot.: 40MB

# Event-based Computer Vision Algorithms

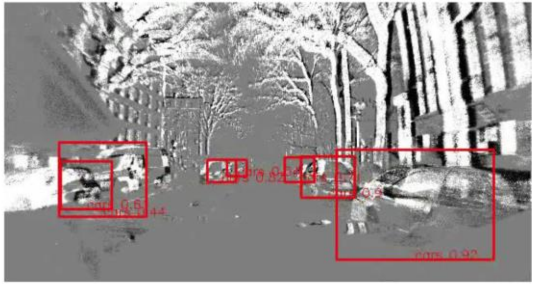
- Spike-FlowNet: spike-based optical flow detection



[11] C. Lee, ECCV, 2020

- Prophesee: real-time detection of fast-moving objects



MobileNet-V2
7.1M weights

[12] Prophesee, CVPR Workshop, 2019

Prophesee, Edge AI & Vision Alliance, 2019

# Digital Neuromorphic Chips in the Literature

| Company/Lab | Chip type | #Neurons/ synapses | On-chip learning | Power | Software | Applications |
|---|---|---|---|---|---|---|
| TrueNorth/IBM (9) | Digital | 1 M/256 M (in 4 K cores) | N | ~0.3 W | Custom | DNN acceleration |
| SpiNNaker/University of Manchester (13) | Digital | 1B/10 kilobytes (in 64 K x 18 ARM cores) | Y | ~kW | PyNN, NEST | Real-time simulation of SNN; HPC |
| Loihi/Intel Labs (12) | Digital | ~128,000/128 M per chip (scalable) | Y | ~1 W | Lava | Research chip |
| Dynap-CNN/ SynSense | Digital | ~327,000/278,000 | N | ~5 mW | Rockpool, PyTorch | Smart sensing |
| BrainChip/Akida | Digital | Configurable, 8-Mb SRAM | Y | ~30 mW | TensorFlow, CNN → SNN | Smart sensing, one-shot learning |
| Tianjic/Tsinghua University (34) | Digital | 40,000/10 M (on 156 cores) | N | ~1 W | Custom | ANN/SNN acceleration |

[13] Y. Sandamirskaya, Science Robotics, 2022

- A number of industrial/academic neuromorphic chips have been presented to date
- Quickly evolving SNN algorithms need to be accommodated

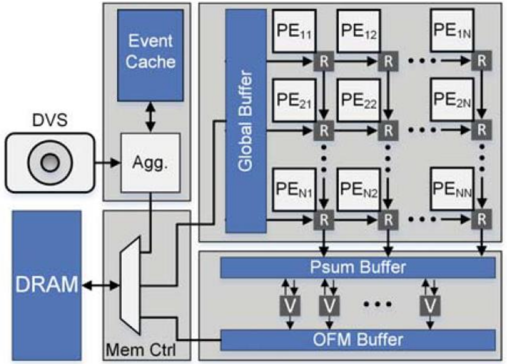
# Custom Hardware for Event-based SNN

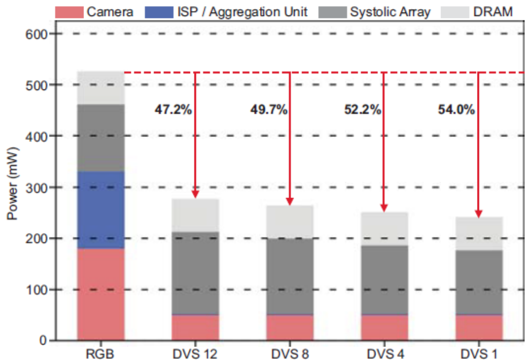
- Algorithm:
  - SNN trained with Prophesee's Gen1 dataset
  - Object detection: YOLOv3 variant (backbone: ResNet18)
  - After training, 8-bit quantization
  - Frame stacking improves mAP
- Hardware:
  - Custom SoC simulated: 16nm, 600MHz, w/ DRAM model

DETECTION RESULTS ON GEN1 AUTOMOTIVE DETECTION

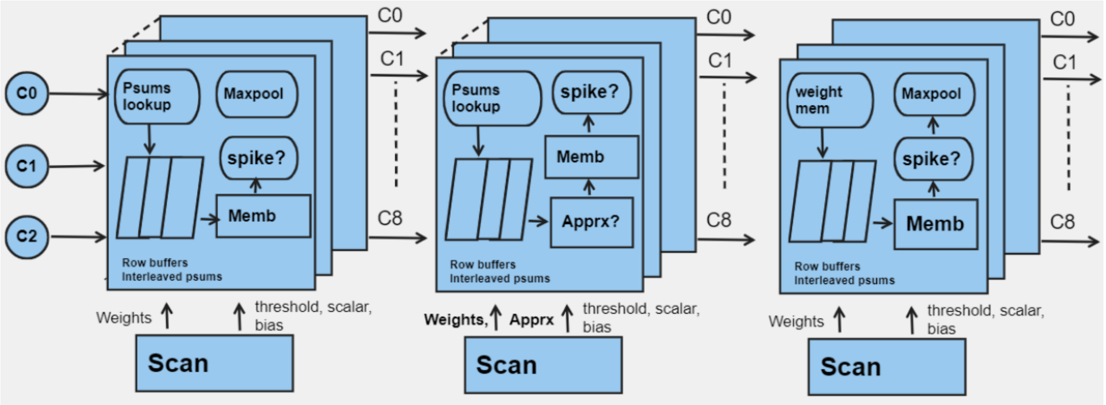
| Camera | # Frames | mAP (%) | Energy (mJ) | Latency (ms) |
|---|---|---|---|---|
| Optical | 1 | - | 4.37 | 33.7 |
| DVS | 1 | 32.0 | 4.15 | 32.0 |
| DVS | 4 | 37.3 | 4.48 | 34.5 |
| DVS | 8 | 38.6 | 4.91 | 37.9 |
| DVS | 12 | 39.6 | 5.34 | 41.2 |

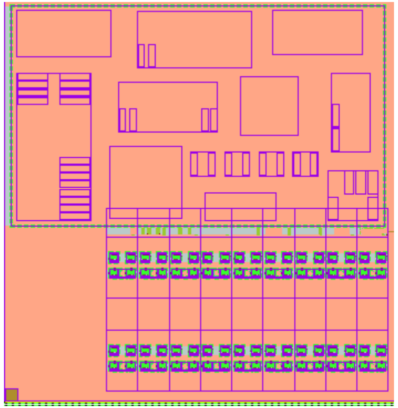
[14] B. Crafton, AICAS, 2021



# ASIC Chip Design for Mobile-SNN

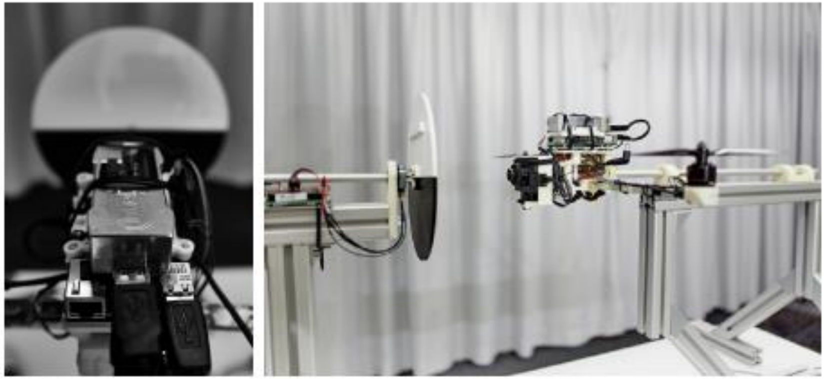


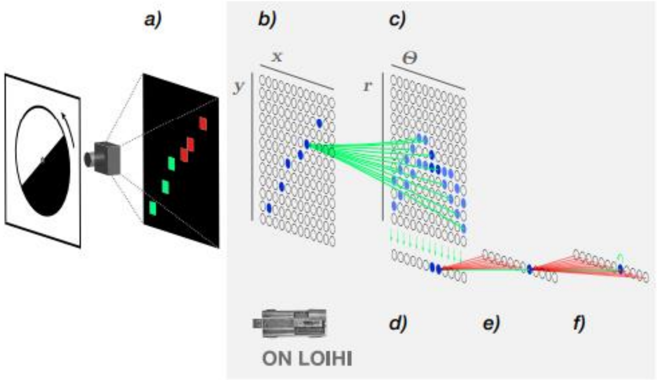



Anupreetham et al., ASU

- Efficient SNN for compact MobileNet architecture with configurable approximate computing
- Fine-grain pipelined architecture for low-precision SNN algorithms
- Prototype chip implemented with Intel 16 CMOS technology, 2mm x 2mm chip area

# Event-based Vision & Control for UAVs w/ Loihi






[15] A. Vitale, ICRA, 2021

- Event-based cameras produce a sparse stream of events that can be processed more efficiently and with a lower latency than images, enabling ultra-fast vision-driven UAV control.
- Event-based vision algorithm implemented as SNN on Loihi chip → used in drone controller.
- Seamless integration of event-based perception on Loihi chip leads to faster control rates and lower latency

# Summary

- Ideal end-to-end neuromorphic computing system will require & integrate:
  - Front-end DVS: high-resolution event sensor w/ sparse spike outputs
    - Mature event-based cameras are being commercially available
  - Back-end SNN accelerator: low-power custom hardware fully exploiting sparsity
    - Commercial/academia chips exist, but support for SOTA SNN algorithms isn't clear
  - Hardware-aware SNN algorithms: high-accuracy and compact algorithms
    - SOTA SNN algorithms have shown noticeable accuracy improvement, while many still use high-precision floating-point precision
    - Large event-based dataset available and could be forthcoming
- Fitting neuromorphic applications:
  - High-speed motion/object tracking, latency-sensitive tasks
  - Smart drones, robotics with fast perception and control

# Acknowledgements


- Faculty: Priya Panda (Yale)
- Students: Jian Meng, Ahmed Hasssan, Anupreetham (ASU)

- Sponsors:

# References


[1] S. Song et al., "Highly Nonrandom Features of Synaptic Connectivity in Local Cortical Circuits," *PLOS Biology*, 2015.

[2] S. Yin et al., "Algorithm and Hardware Design of Discrete-time Spiking Neural Networks based on Back Propagation with Binary Activations," *IEEE BioCAS*, 2017.

[3] A. Andreopoulos et al., "Visual Saliency on Networks of Neurosynaptic Cores," *IBM Journal of Research and Development*, 2015.

[4] P. Lichtsteiner et al., "A 128×128 120 dB 15 μs Latency Asynchronous Temporal Contrast Vision Sensor," *IEEE JSSC*, 2008.

[5] B. Son et al., "A 640×480 Dynamic Vision Sensor with a 9μm Pixel and 300Meps Address-event Representation," *IEEE ISSCC*, 2017.

[6] T. Finateu et al., "A 1280×720 Back-Illuminated Stacked Temporal Contrast Event-Based Vision Sensor with 4.86μm Pixels, 1.066GEPS Readout, Programmable Event-Rate Controller and Compressive Data-Formatting Pipeline," *IEEE ISSCC*, 2020.

[7] H. Li et al., "CIFAR10-DVS: An Event-Stream Dataset for Object Classification," *Frontier of Neuroscience*, 2017.

# References


[8] E. Perot et al., "Learning to Detect Objects with a 1 Megapixel Event Camera," *NeurIPS*, 2020.

[9] M. Sorbaro et al., “Always-on visual classification below 1 mWwith spiking convolutional networks on Dynap™-CNN,” *TinyML Talks*, 2021.

[10] L. Cordone et al., “Object Detection with Spiking Neural Networks on Automotive Event Data”, *IJCNN*, 2022.

[11] C. Lee et al., “Spike-FlowNet: Event-based Optical Flow Estimation with Energy-efficient Hybrid Neural Networks,” *ECCV*, 2020.

[12] A. Sironi et al., “Learning from Events: on the Future of Machine Learning for Event-based Cameras,” *CVPR Workshop*, 2019.

[13] Y. Sandamirskaya et al., “Neuromorphic Computing Hardware and Neural Architectures for Robotics,” *Science Robotics*, 2022.

[14] B. Crafton et al., “Hardware-Algorithm Co-Design Enabling Efficient Event-based Object Detection,” *IEEE AICAS*, 2021.

[15] A. Vitale et al., “Event-driven Vision and Control for UAVs on a Neuromorphic Chip,” *IEEE ICRA*, 2021.