SENECA: Building a fully digital neuromorphic processor, design trade-offs and challenges

Guangzhi Tang

guangzhi.tang@imec.nl

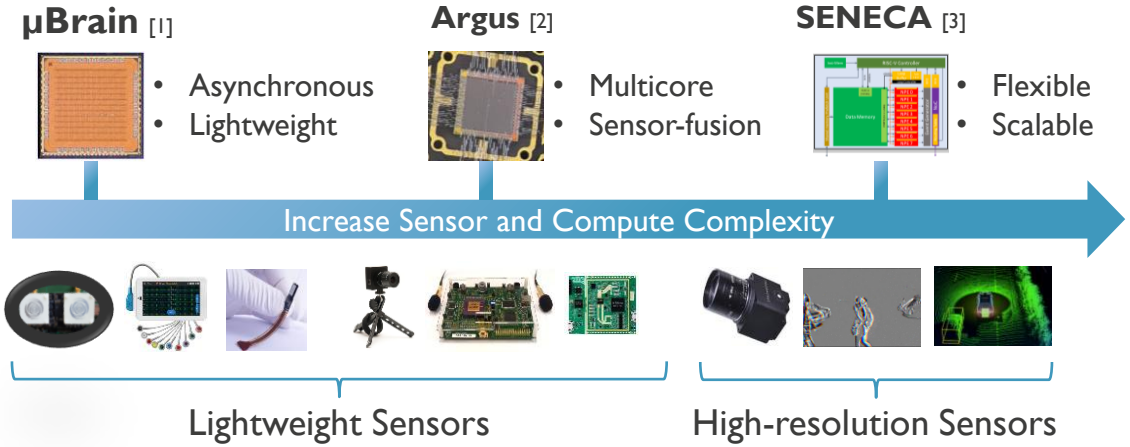
Hardware Efficient AI, imec the Netherlands

Neuromorphic Computing in the Netherlands Workshop (NCN2023) – October 2nd, 2023 – Groningen

Neuromorphic and Hardware Efficient AI Research

@ imec the Netherlands

Holst Centre
Powered by imec & TNO



[1] Stujit, et al., Frontiers in Neuroscience, 2021.
[2] Detterer, et al., in submission.
[3] Tang, et al., Frontiers in Neuroscience, 2023.

Neuromorphic and Hardware Efficient AI Research

@ imec the Netherlands

Holst Centre
Powered by imec & TNO

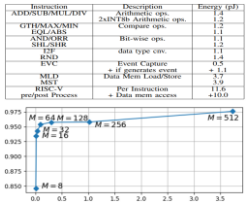
µBrain [1]
- Asynchronous
- Lightweight

Argus [2]
- Multicore
- Sensor-fusion

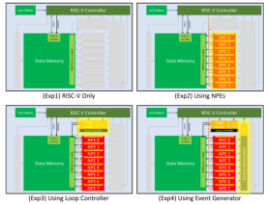
SENECA [3]
- Flexible
- Scalable

Increase Sensor and Compute Complexity

HW-Algorithm Co-optimization

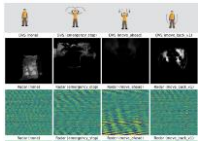


Instruction-level benchmarking [4]

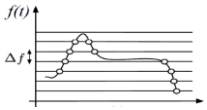


Design space exploration [5]

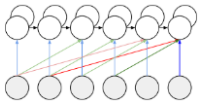
Neuromorphic Algorithms & Datasets



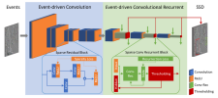
Sensor Fusion Dataset [6]



Delta DNN [8]

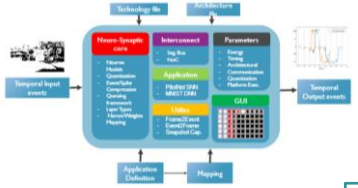


Synaptic Delay [7]



Event RNN [9]

Software Design



Event-driven Simulator [10]

SDK & Mapper

[4] Tang, et al., ISCAS, 2023.
[5] Shidqi, et al., under review.
[6] Muller, et al., RadarConf, 2023.
[7] Patiño-Saucedo, et al., ISCAS 2023.
[8] Yousefzadeh, et al., IJCNN, 2022.
[9] Wang, et al., in submission.
[10] Nembhani, et al., under review.

Neuromorphic and Hardware Efficient AI Research

@ imec the Netherlands

Holst Centre
Powered by imec & TNO

µBrain [1]
- Asynchronous
- Lightweight

Argus [2]
- Multicore
- Sensor-fusion

SENECA [3]
- Flexible
- Scalable

Increase Sensor and Compute Complexity

HW-Algorithm Co-optimization

Instruction-level benchmarking [4]

Design space exploration [5]

Neuromorphic Algorithms & Datasets

Sensor Fusion Dataset [6]

Delta DNN [8]

Synaptic Delay [7]

Event RNN [9]

Software Design

Event-driven Simulator [10]

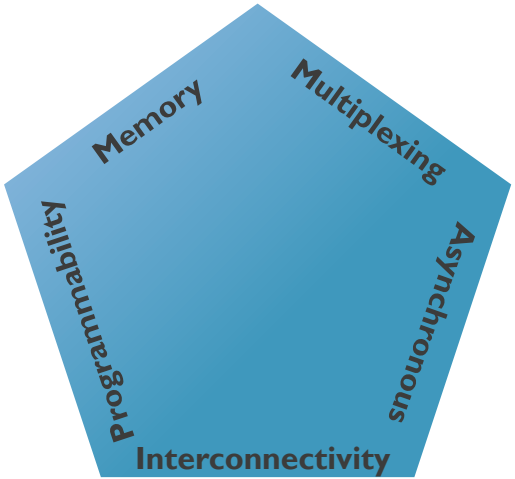
SDK & Mapper

[4] Tang, et al., ISCAS, 2023.
[5] Shidqi, et al., under review.
[6] Muller, et al., RadarConf, 2023.
[7] Patiño-Saucedo, et al., ISCAS 2023.
[8] Yousefzadeh, et al., IJCNN, 2022.
[9] Wang, et al., in submission.
[10] Nembhani, et al., under review.

Trade-offs in Digital Neuromorphic Architecture Design

Memory
- Size and flexibility
- Technology and hierarchy

Multiplexing
- Axon and neuron
- Area efficiency
- Memory and control cost



Programmability
- Events and precision
- Synaptic and neural models
- Architecture
- Learning

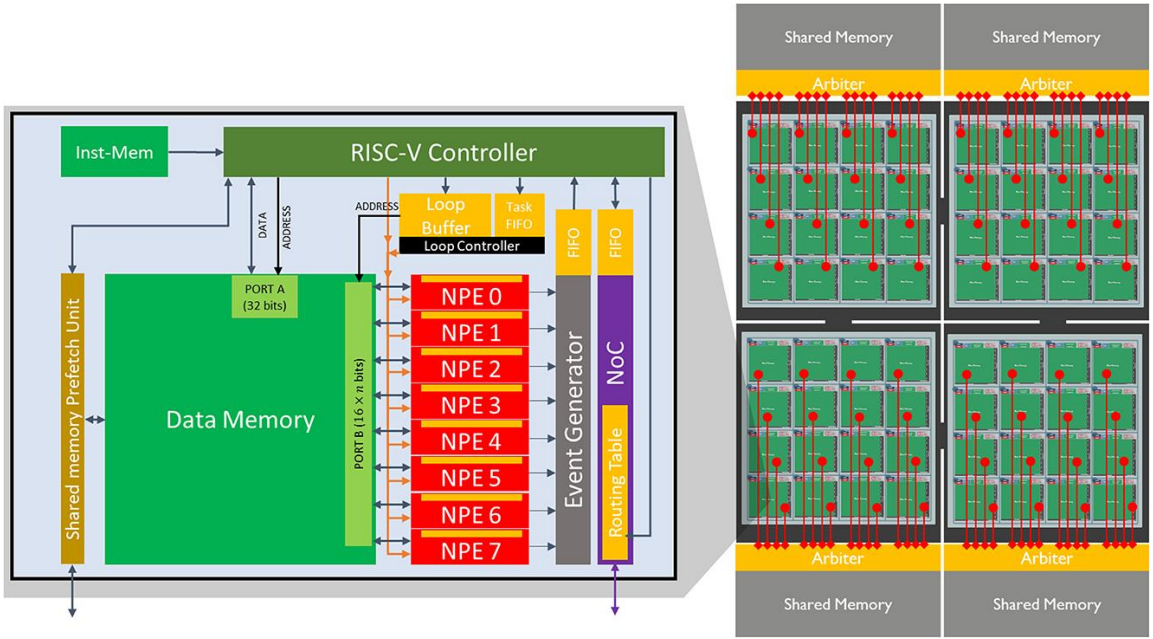
Asynchronous
- Full asynchronous
- Self-clocked cores, global synchronous
- Synchronous cores, global asynchronous

Interconnectivity
- Multicasting
- Compression

Tang, et al., Frontiers in Neuroscience, 2023.

SENECA

Scalable **E**nergy efficient **NE**uromorphic **C**omputer **A**rchitecture



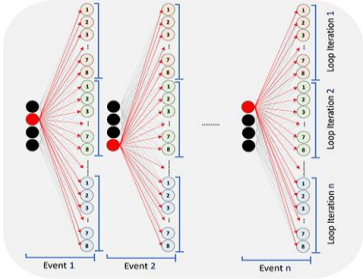
- **Memory** – Unified, 3-level hierarchy with register, SRAM, shared memory
- **Multiplexing** – Axon and neuron
- **Programmability** – Fully programable for synapse, neuron, architecture, learning
- **Asynchronous** – Core-to-core asynchrony
- **Interconnectivity** – Multicasting NoC, software compression

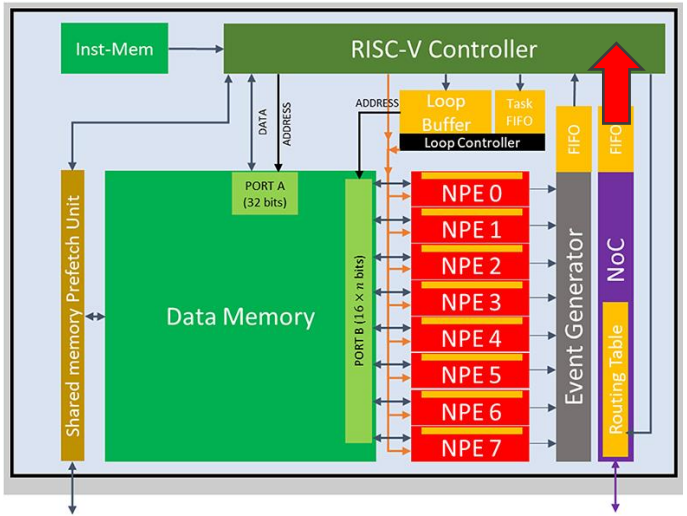
Tang, et al., Frontiers in Neuroscience, 2023.

Optimizing event-driven processing on SENECA

Event-driven neural network processing

Event-driven Network
(Input event integrate to all post-synaptic states)





Event Receiving

Task Decoding

Neural Processing

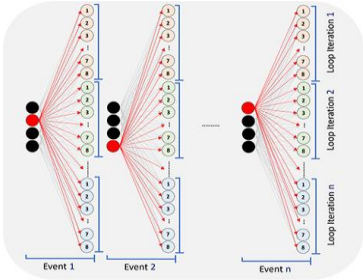
Event Generation

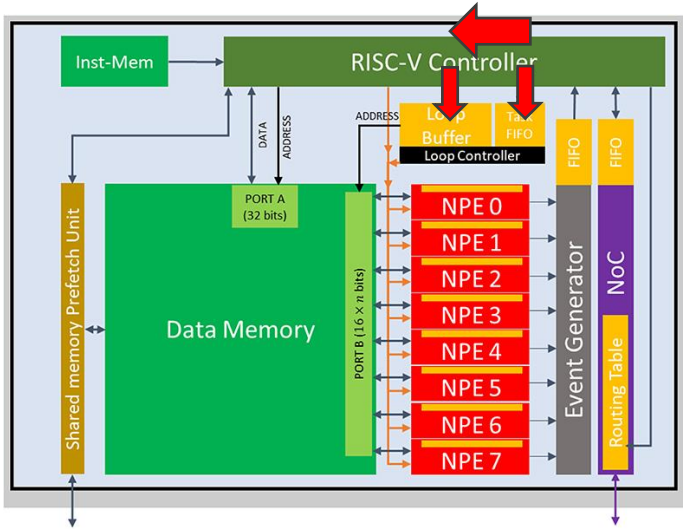
Tang, et al., Frontiers in Neuroscience, 2023.

Optimizing event-driven processing on SENECA

Event-driven neural network processing

Event-driven Network
(Input event integrate to all post-synaptic states)





Event Receiving

Task Decoding

Neural Processing

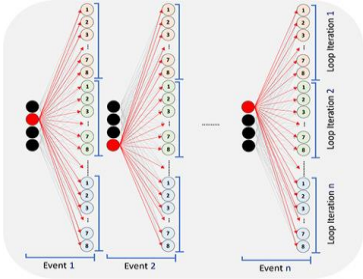
Event Generation

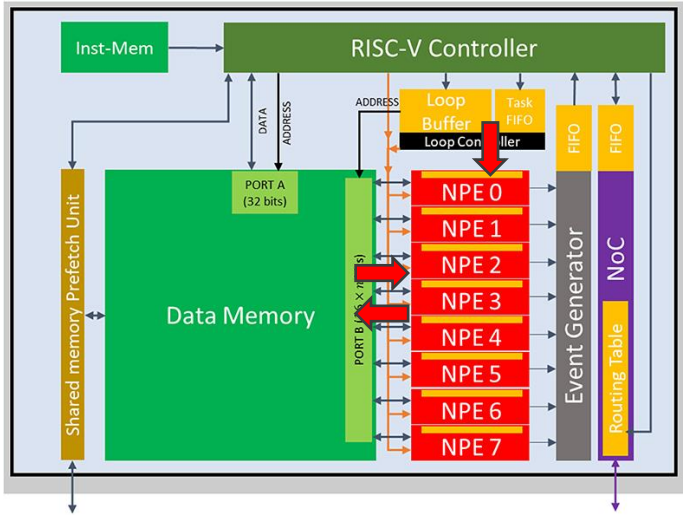
Tang, et al., Frontiers in Neuroscience, 2023.

Optimizing event-driven processing on SENECA

Event-driven neural network processing

Event-driven Network
(Input event integrate to all post-synaptic states)





Event Receiving

Task Decoding

Neural Processing

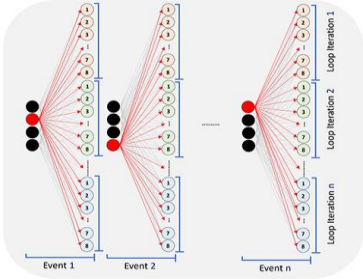
Event Generation

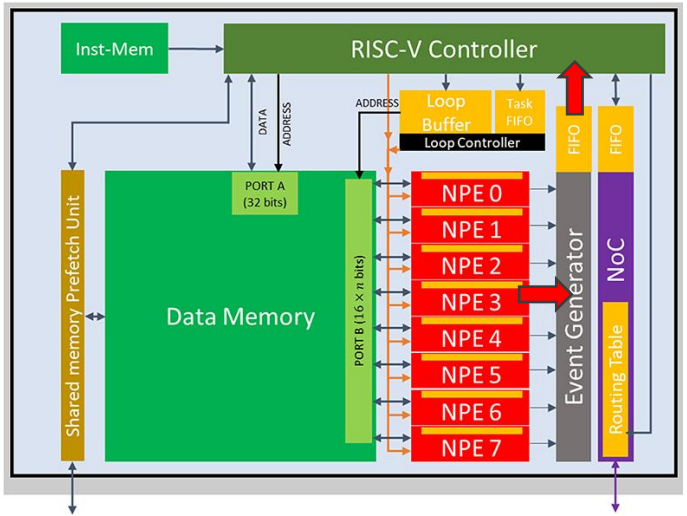
Tang, et al., Frontiers in Neuroscience, 2023.

Optimizing event-driven processing on SENECA

Event-driven neural network processing

Event-driven Network
(Input event integrate to all post-synaptic states)





Event Receiving

Task Decoding

Neural Processing

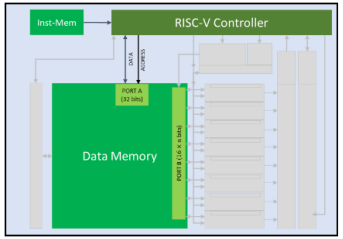
Event Generation

Tang, et al., Frontiers in Neuroscience, 2023.

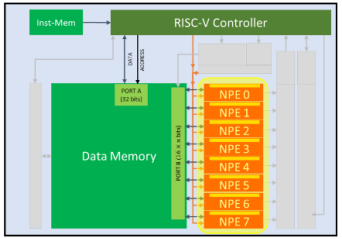
Optimizing event-driven processing on SENECA

Design space exploration

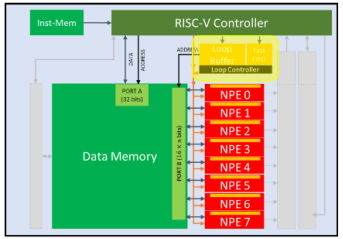
Event-driven Network
(Input event integrate to all post-synaptic states)



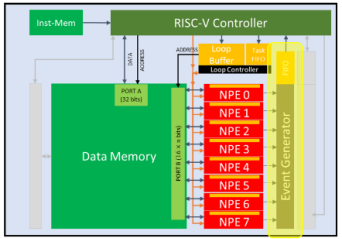
(Exp1) RISC-V Only



(Exp2) Using NPEs



(Exp3) Using Loop Controller



(Exp4) Using Event Generator

Experiment	Latency (µS)	Energy (µJ)
RISC-V Only	6625	34.0
+NPEs	1098	6.71
+Loop Controller	541	2.81
+Event Generator	**400**	**2.10**

>15x HW Improvement

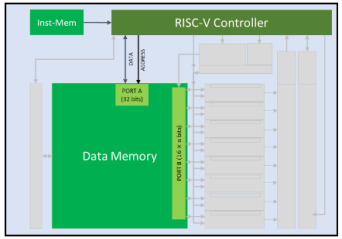
** Results generated by Cadence Xcelium and JOULES ** KWS Network: 390-256-256-29*

Shidqi, et al., under review.

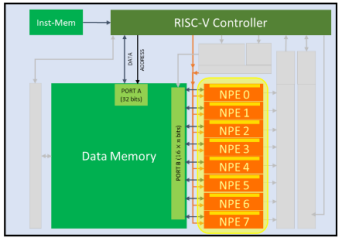
Optimizing event-driven processing on SENECA

Design space exploration

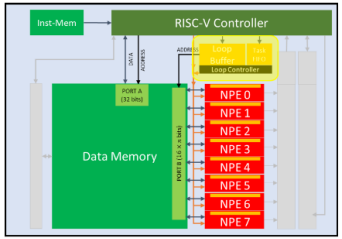
Event-driven Network
(Input event integrate to all post-synaptic states)



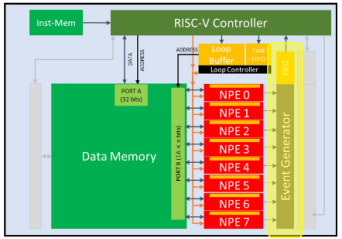
(Exp1) RISC-V Only



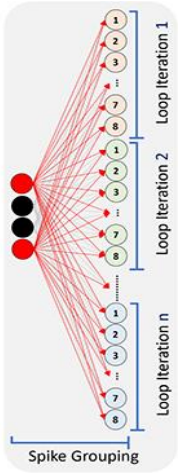
(Exp2) Using NPEs



(Exp3) Using Loop Controller



(Exp4) Using Event Generator



Spike Grouping
Reduce data movements

Experiment	Latency (µS)	Energy (µJ)
RISC-V Only	6625	34.0
+NPEs	1098	6.71
+Loop Controller	541	2.81
+Event Generator	400	2.10
+Spike Grouping	**218**	**1.20**

300x vs Loihi

6x vs SpiNNaker2

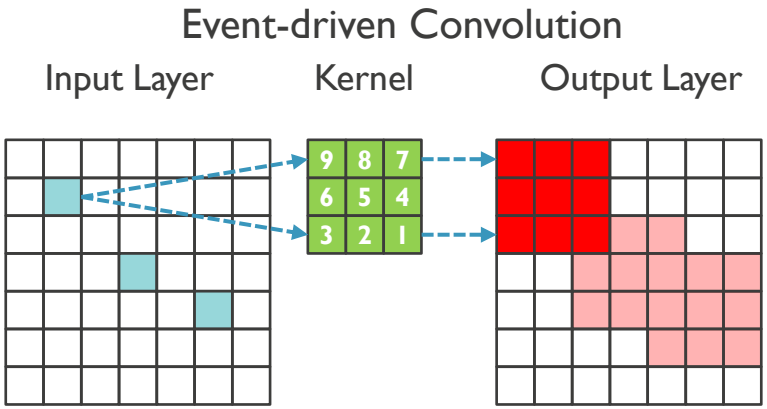
imec

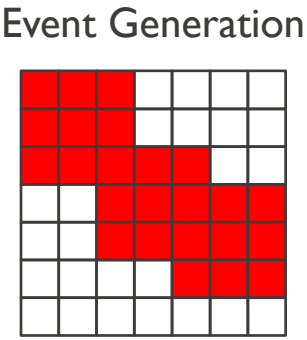
** Results generated by Cadence Xcelium and JOULES ** KWS Network: 390-256-256-29*

Shidqi, et al., under review.

Event-driven Convolutional Neural Network

Existing problems of convolutional neural network on large-scale digital neuromorphic HW





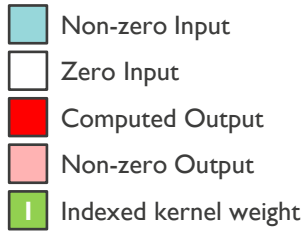
* *Take Loihi as example*

Problem 1: State Memory

- Store all states on-chip
- Unbearable for high-res

Problem 2: Layer Latency

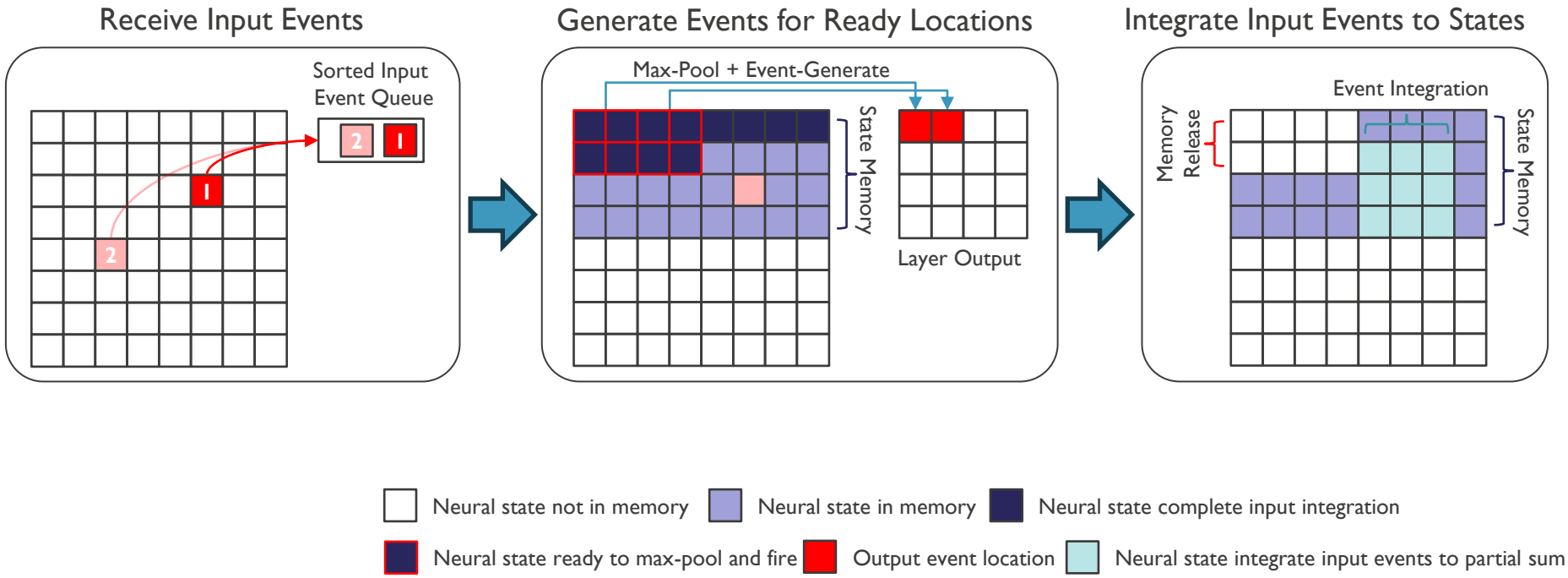
- Event generate after step sync
- Additional latency per layer



Event-driven Convolutional Neural Network

Event-driven Depth-first Convolution on SENECA

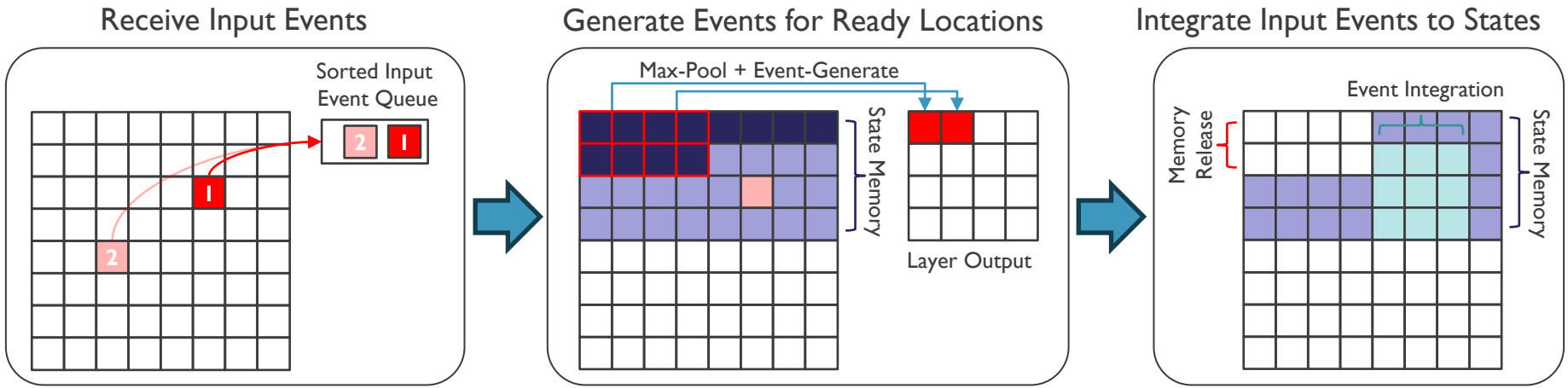
Single SENECA core processing 3x3 Convolution + 2x2 Max-Pooling



Event-driven Convolutional Neural Network

Event-driven Depth-first Convolution on SENECA

Single SENECA core processing 3x3 Convolution + 2x2 Max-Pooling



Reduce state memory cost

X*Y*C to **min(X,Y)*C*4**

Reduce layer latency

Full layer to **First event**

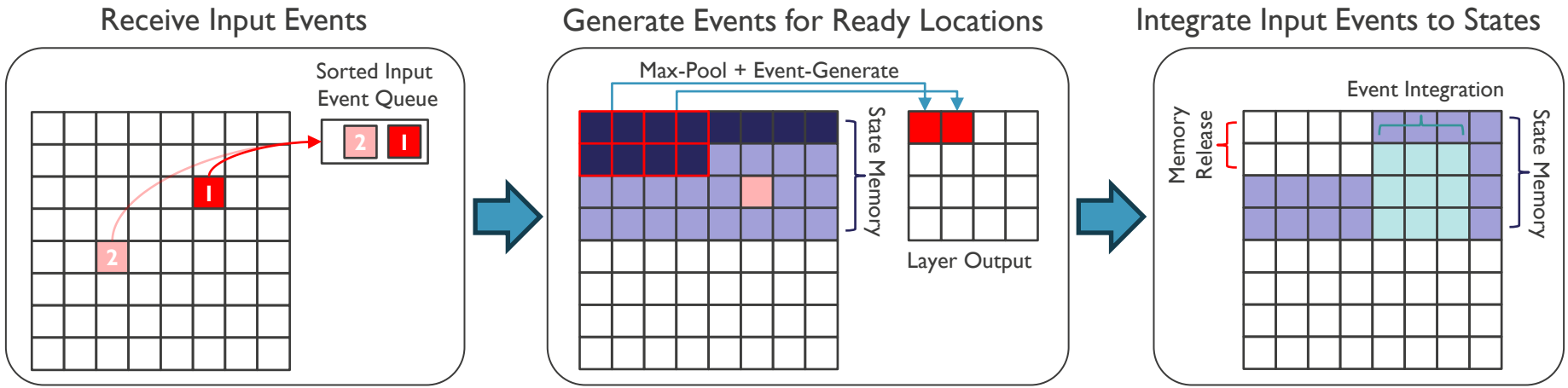
Improve vs	Energy	Latency	Area
Loihi	>50x	>3x	>3x
TrueNorth	>9x	>3x	>100x

** 5-layer CNN: (8c3k2p)-(16c3k2p)-(32c3k2p)-128-class*

Event-driven Convolutional Neural Network

Event-driven Depth-first Convolution on SENECA

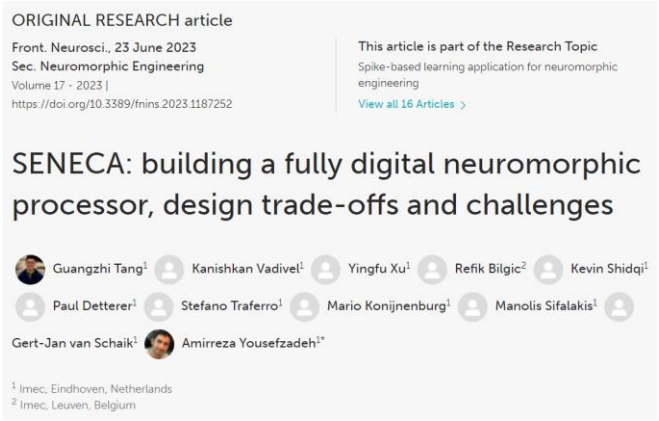
Single SENECA core processing 3x3 Convolution + 2x2 Max-Pooling



Limitations:

1. Overheads on buffering and sorting input events.
2. Difficulties for synchronizing a multi-core single-layer mapping in asynchronous system.
3. Only non-stateful convolutional layer can benefit on state memory reduction.

Hardware Efficient AI Team @ imec the Netherlands



ORIGINAL RESEARCH article

Front. Neurosci., 23 June 2023
Sec. Neuromorphic Engineering
Volume 17 - 2023 |
https://doi.org/10.3389/fnins.2023.1187252

This article is part of the Research Topic
Spike-based learning application for neuromorphic engineering
View all 16 Articles

SENECA: building a fully digital neuromorphic processor, design trade-offs and challenges

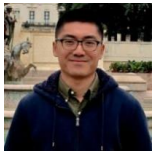
Guangzhi Tang[1] Kanishkan Vadivel[1] Yingfu Xu[1] Refik Bilgic[2] Kevin Shidqi[1]
Paul Detterer[1] Stefano Traferro[1] Mario Konijnenburg[1] Manolis Sifalakis[1]
Gert-Jan van Schaik[1] Amirreza Yousefzadeh[1*]

[1] Imec, Eindhoven, Netherlands
[2] Imec, Leuven, Belgium

This technology is partially funded and initiated by the Netherlands and European Union's Horizon 2020 research and innovation projects **TEMPO** (ECSEL Joint Undertaking under grant agreement No 826655), **ANDANTE** (ECSEL Joint Undertaking under grant agreement No 876925), **DAIS** (Key Digital Technologies Joint Undertaking under grant agreement No 101007273) and **REBECCA** (Key Digital Technologies Joint Undertaking under grant agreement No 101097224).

Team Members


Amirreza


Guangzhi


Kanishkan


Paul


Gert-Jan


Manolis


Yingfu


Kevin


Alexandra

Students


Refik


Prithvish


Shenqi


Cina


Roy

...


imec
embracing a better life