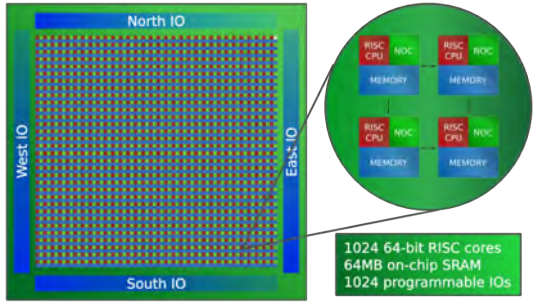


Benchmarking the Epiphany processor as a reference neuromorphic architecture

International Workshop on Edge Artificial Intelligence for Industrial Applications (EAI4IA)
25-26 July 2022
Vienna

Maarten Molendijk[1,2], Kanishkan Vadivel[2], Gert-Jan van Schaik[1], Amirreza Yousefzadeh[1], and Henk Corporaal[2]

[1]imec, Eindhoven, Netherlands
[2]Technical University of Eindhoven, Netherlands

KDT JU
Key Digital Technologies Joint Undertaking
ECSEL
Joint Undertaking

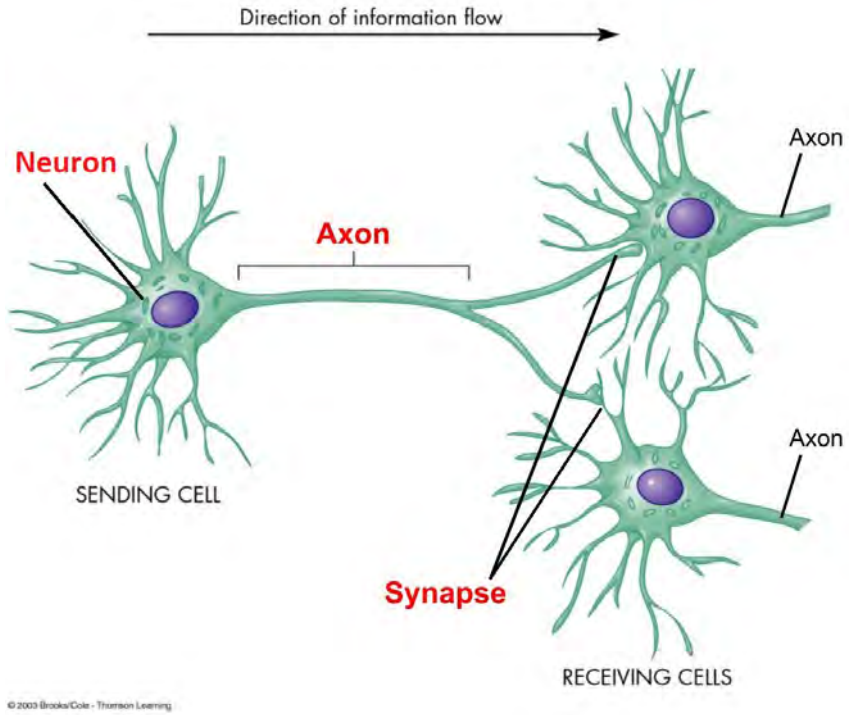
ANDANTE

TEMPO

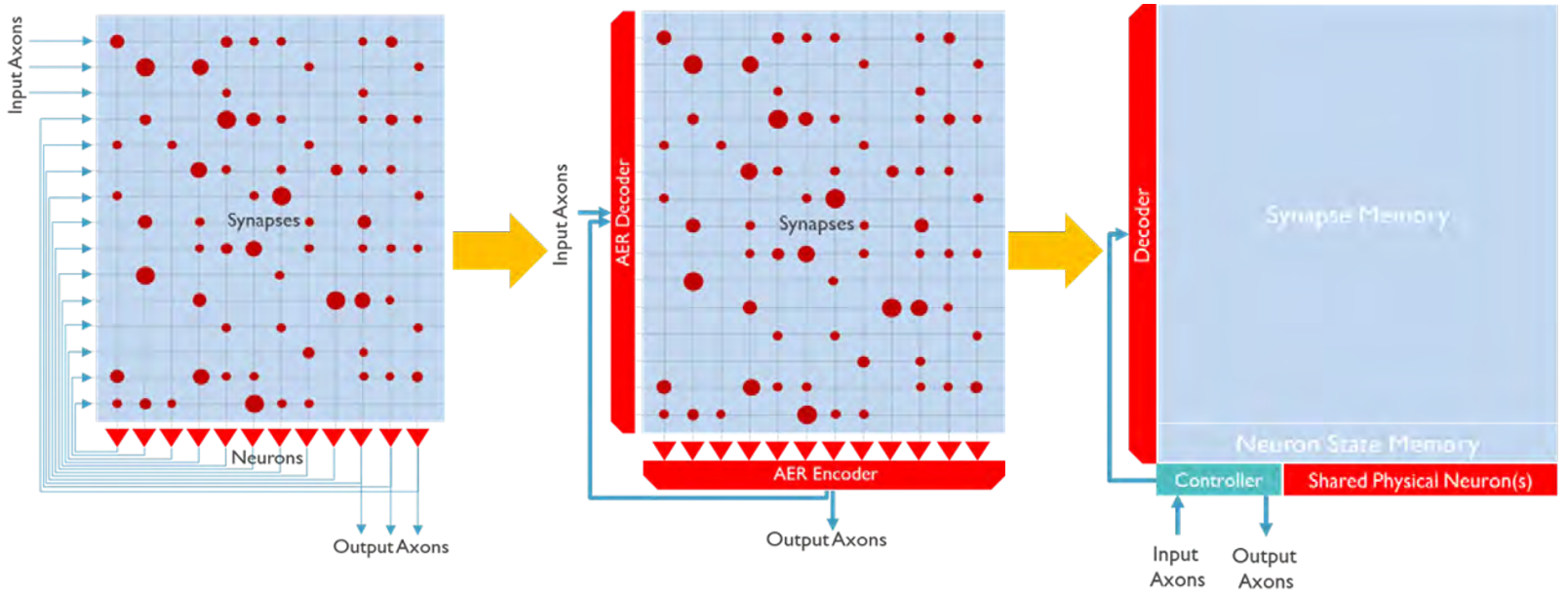
imec

Bio-inspired processing

- Energy efficient natural signal processing
- Interesting features:
  - Sparsity exploitation
  - Data-flow parallel processing
  - Scalable
  - Low-precision parameters
  - Asynchronous and non-deterministic
  - Adaptative (fault-tolerance)



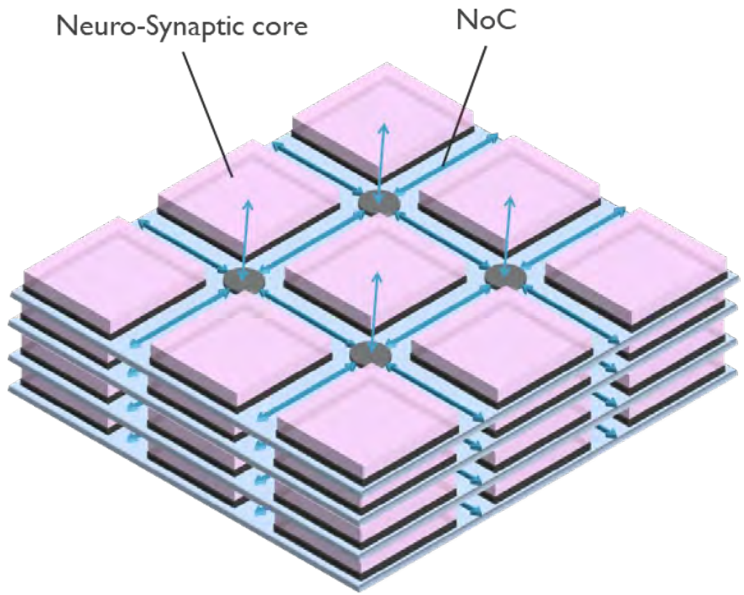
Digital Neuromorphic Processor



A - No time multiplexing

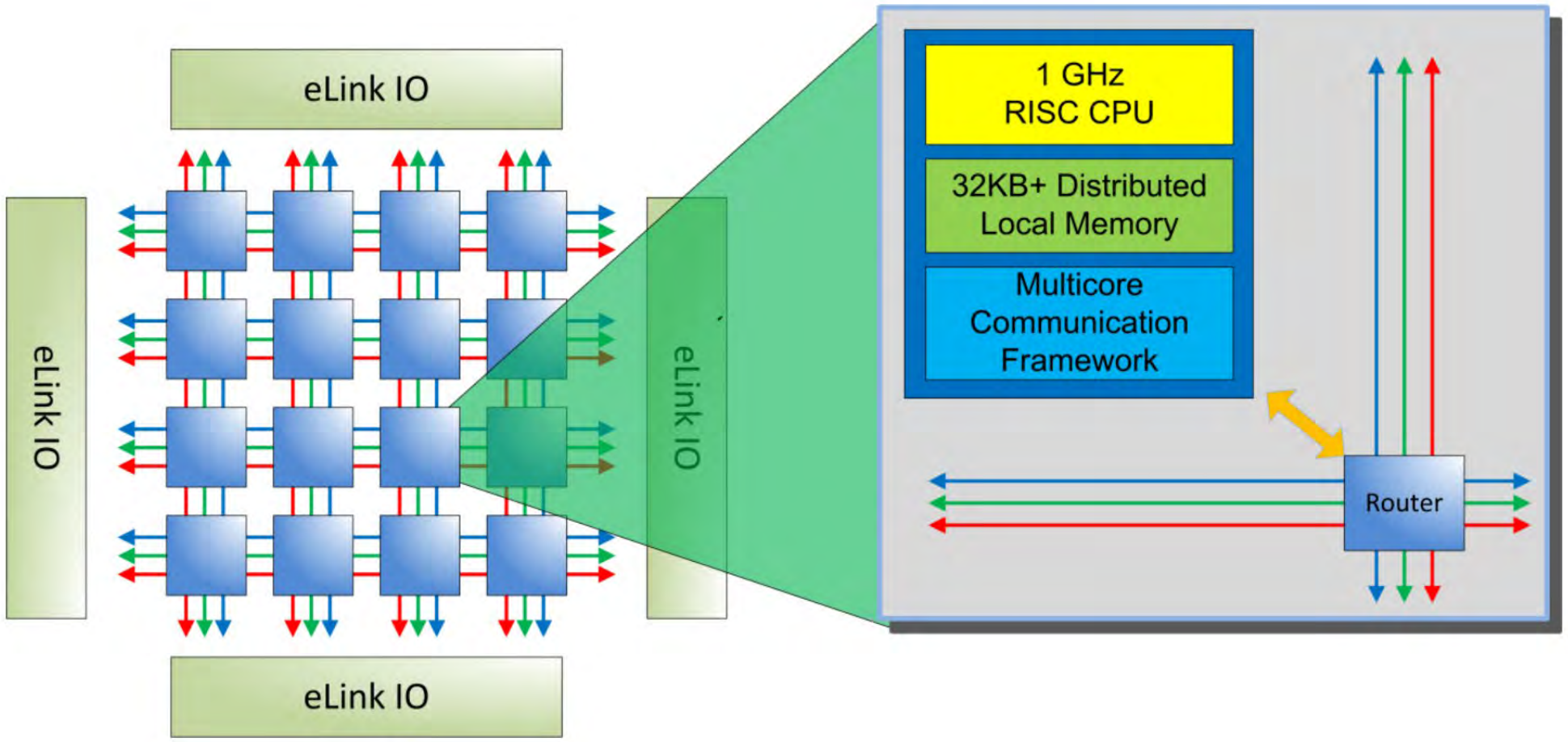
B – Axon time multiplexing

C – Neuron & Axon time multiplexing



D – Interconnecting the cores

Epiphany

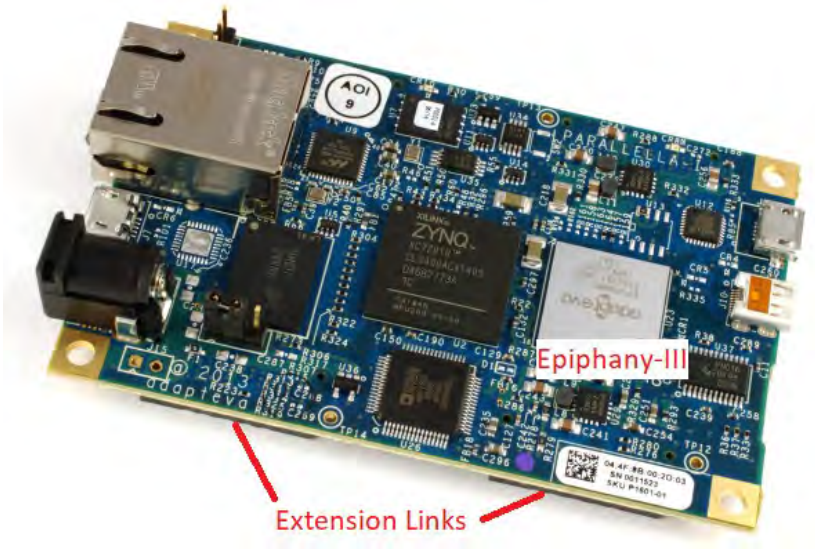


Introduced in 2009

Failed as a general-purpose processor!

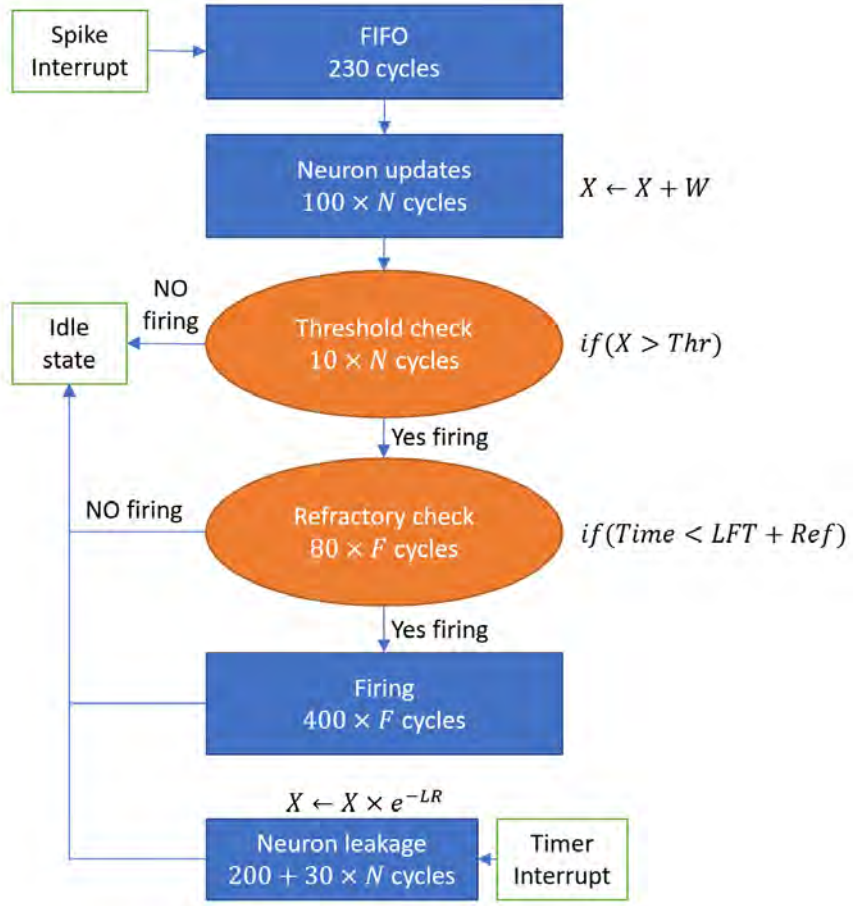
Why Epiphany is a good base-line processor?

- Easy Access ($100 for SBC)
- Flexible memory allocation (soft partitioning)
- Simple Network on Chip
- Flexible processing model make it possible to implement and test:
  - Various neuron models
  - Various learning algorithms



Profiling Epiphany for event processing

- Implementation of simple LIF neurons
  - N : Number of neurons
  - F : Number of Firings
  - X : Neuron state
  - W: Synaptic weight
  - Thr: Firing threshold
  - LFT: Last Firing Time
  - Ref: Refractory time
  - LR: Leak Rate



Profiling Epiphany for event processing

1GHz clock, 1024 neurons in a core

- Receiving the spike: $230\ cycles\ \approx 0.2\mu s$
- Updating neurons: $1024 \times 100\ cycles\ \approx 102\mu s$
- Checking the Thresholds: $1024 \times 10\ cycles\ \approx 10\mu s$

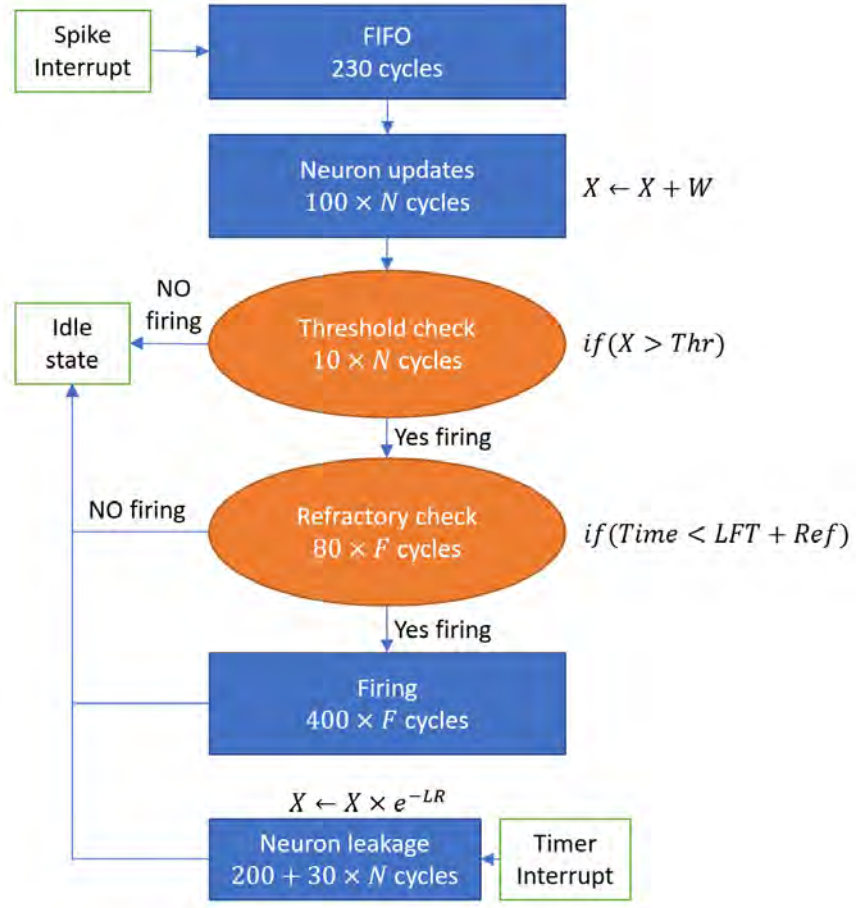
10% over threshold:

- Refractory check: $1024 \times 10\% \times 80\ cycles\ \approx 8\mu s$

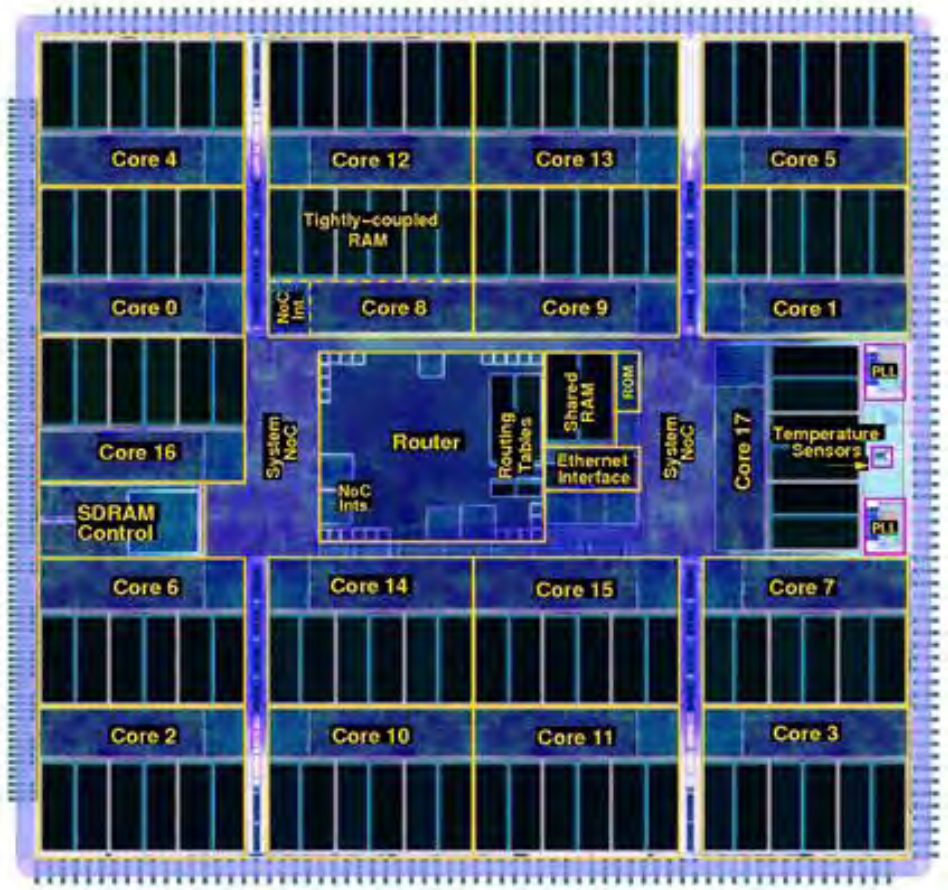
1% firing:

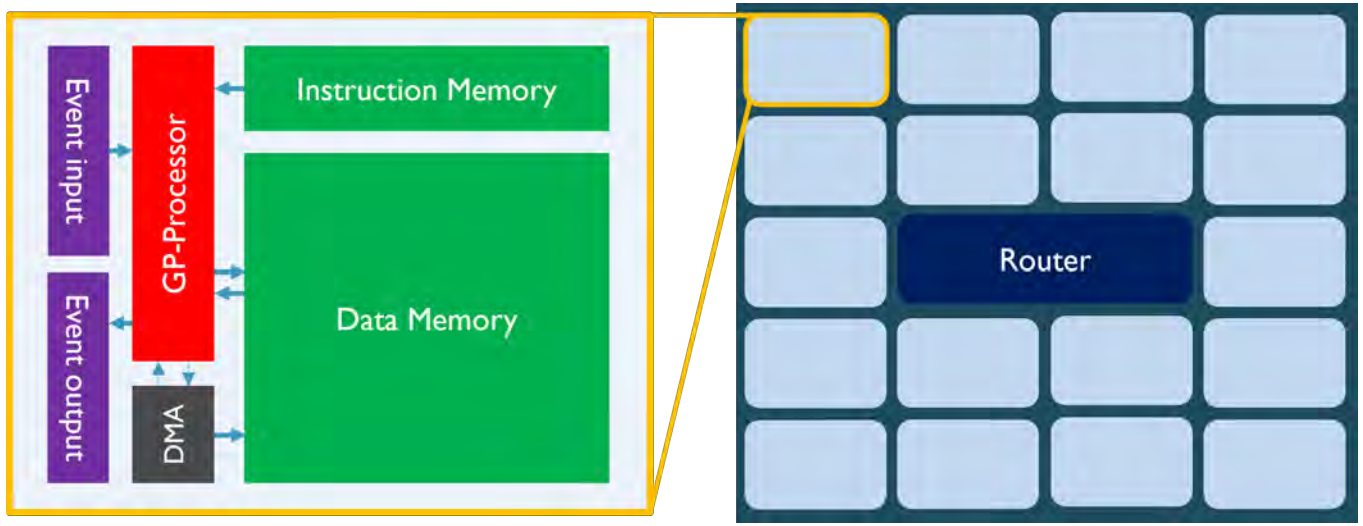
- Firing: $1024 \times 1\% \times 400\ cycles\ \approx 4\mu s$

Periodic leak: $200 + 1024 \times 30\ cycles\ \approx 31\mu s$

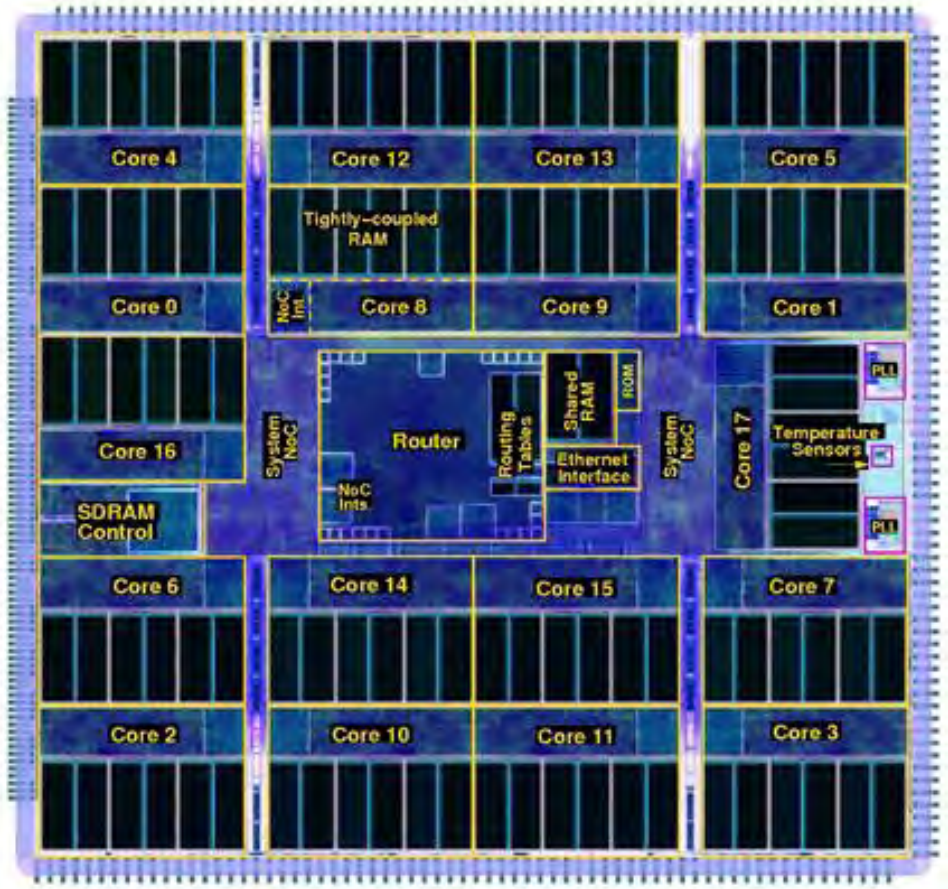


Comparison to SpiNNaker





Comparison to SpiNNaker



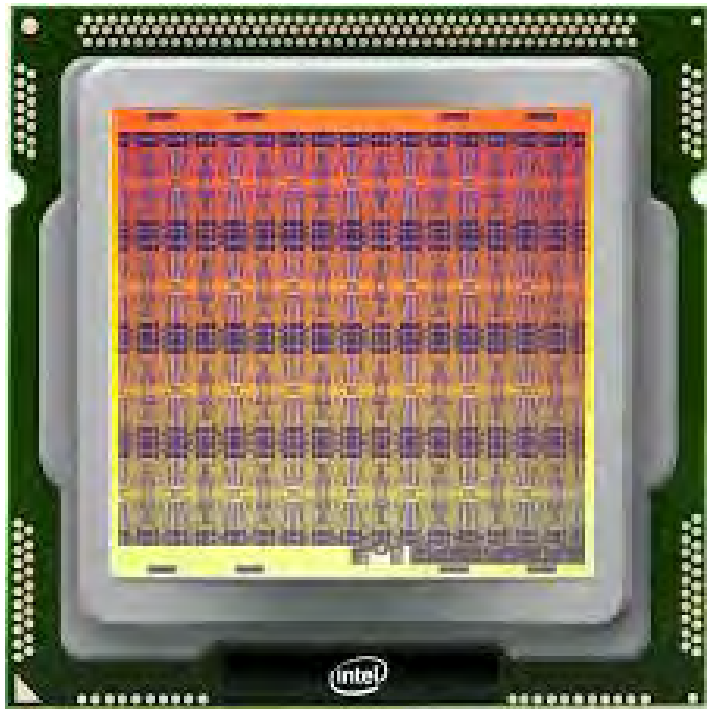
SpiNNaker1 Differences:

- ARM processors [only integer]
- GALS
- Multi-Casting NoC + 6 IO Links
- Off-chip memory access
- Separated IRAM/DRAM

SpiNNaker2 Differences:

- ARM processors + Accelerated MACs
- GALS
- Multi-Casting NoC + 6 IO Links
- Off-chip memory access
- Separated IRAM/DRAM

Comparison to Loihi



LOIHI1 Differences:

- 128 dedicated cores:
  - Fixed neuron model
  - Fixed learning algorithm
- Asynchronous
- Separated Neuron/Synapse/Axon memories

LOIHI2 Differences:

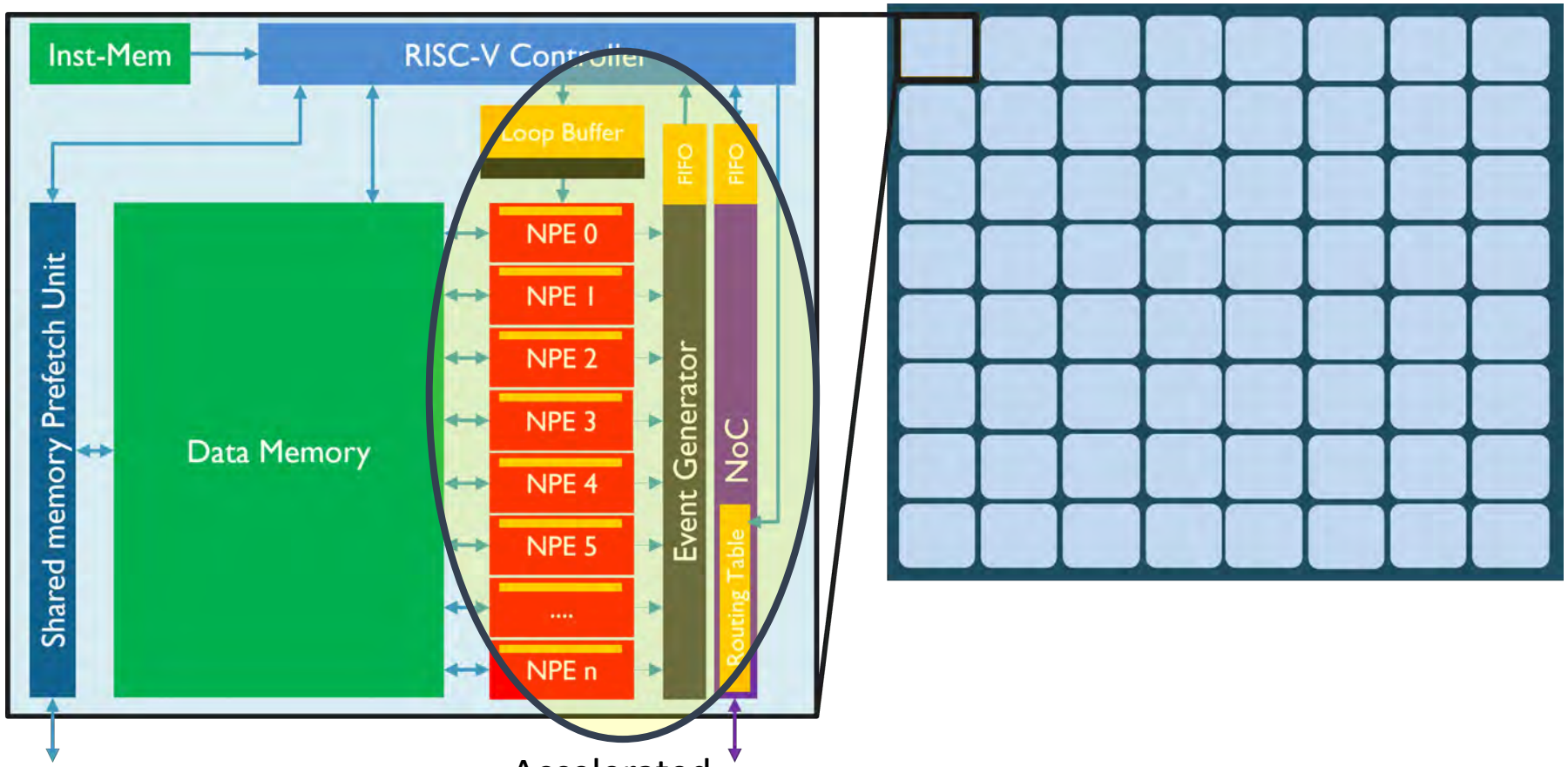
- 128 dedicated cores:
  - Programmable neuron model
  - Programmable learning algorithm
- Asynchronous

Lesson learned

- General Purpose processor provides high amount of flexibility
- However, it is inefficient compared to the dedicated logics
  - Loop over the instruction memory is inefficient

- Solution:
  - Accelerating the most common operations
    - 95% accelerated
    - 5% General purpose

SENeCA

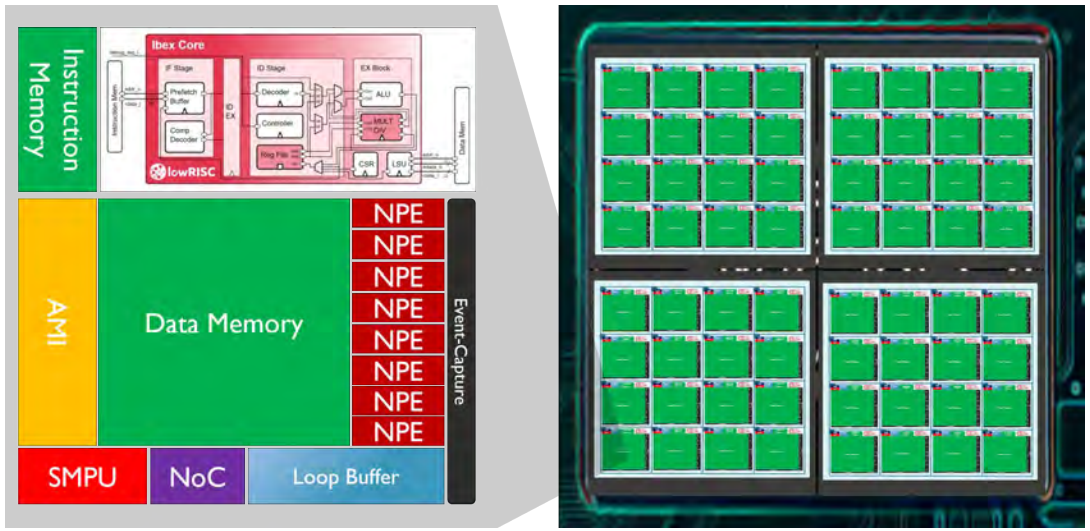


Accelerated neuromorphic processing

Synthesis results and power profiling for a SENeCA core

(CADENCE GENUS-JOULES)

Module	Area (kum2)	Peak Power (mW)
AMI (event-based interface)	12 (2%)	0.1
RISC-V (IBEX)	23 (4%)	0.8
NCP (8xNPE)	38 (7%)	8
Inst Memory (128Kb)	28 (5%)	2
Data Memory (2Mb)	443 (80%)	32



- 400MHz clock, 3.2G Synaptic Operations per second per core
- Area is reported by using the GF-22nm FDSOI
- Power is reported for a three-layer keyword spotting application reported in:

  Blouw, Peter, Xuan Choo, Eric Hunsberger, and Chris Eliasmith. "Benchmarking keyword spotting efficiency on neuromorphic hardware." In *Proceedings of the 7th annual neuro-inspired computational elements workshop*, pp. 1-8. 2019.