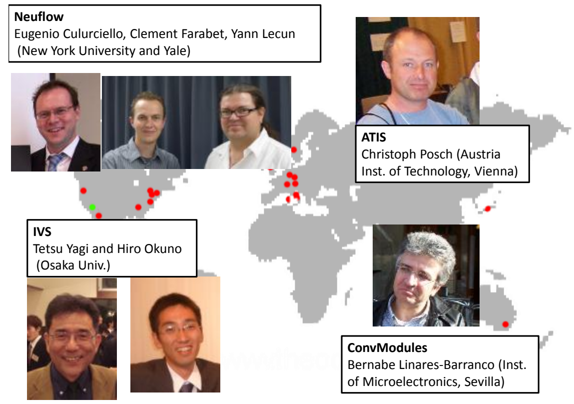
Bio-inspired (electronic) vision

Tobi Delbruck
Inst. of Neuroinformatics
University of Zurich and ETH Zurich
www.ini.uzh.ch/~tobi
2011 VLSI Circuits Workshop, 14 June:
Bio Inspired Computation - What Electronics can Learn from Bio.

With big thanks to ….

The organizers Michael Flynn (Univ. of Michigan) and Makoto Ikeda (Univ. of Tokyo)

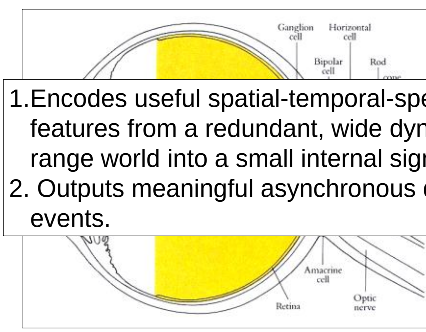
And to ….



Outline

- **Biological vision as an inspiration for better machine vision:** computation and costs in the retina and cortex.
- **Silicon retina vision sensors:** the **Scanned "Intelligent Vision Sensor",** and then **"event-based" frame-free asynchronous** sensors – the **Dynamic Vision Sensor** and the **Asynchronous Time-Based Image Sensor**.
- **Applications of these sensors** in surveillance and robotics.
- **Going past simple object tracking: Convolutional networks** - both **event-based** hardware and **"LeNet" in Neuflow project**.
- Retrospective view **comparing natural and artificial computation:** What do we need?

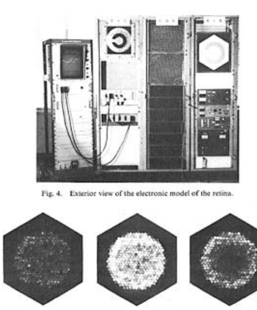
Function of the retina

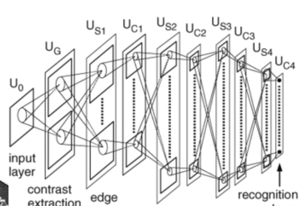


1. Encodes useful spatial-temporal-spectral features from a redundant, wide dynamic range world into a small internal signal range.
2. Outputs meaningful asynchronous digital events.

Rodieck, 1998

Fukushima's / NHK Research Labs 1970 Electronic Retina, 1980 Neocognitron





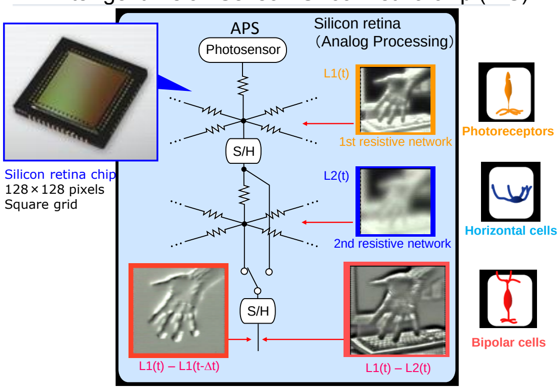
http://www4.ocn.ne.jp/~fuku_k/index-e.html

Historical development of integrated silicon retina vision sensors

1990 1995 2000 2005

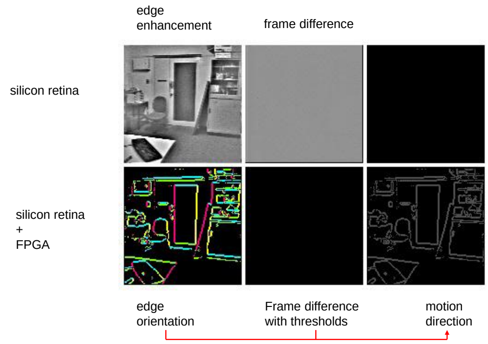
- Mahowald & Mead outer retina (SciAm 91)
- UPenn Magno-Parvo silicon retina (TBME 02)
- CSEM VISe steerable-filter contrast vision sensor (JSSC 03)
- Yale Univ. "Octopus" imager (JSSC 04)
- JHU Temporal Change Threshold Detection Imager (JSSC 07)
- **Osaka Univ. Intelligent Vision Sensor (JSSC 08)**
- **ETH/UZH Dynamic Vision Sensor (JSSC 08)**
- Sevilla Spatial Contrast Silicon Retina (TCAS 08)
- AIT Asynchronous Time-based Image Sensor (ISSCC 10)

"Intelligent Vision Sensor" Silicon retina chip (IVS)



Real-time feature extraction with the IVS

Taking advantages of edge enhancement and frame subtraction images, our system can extract multiple features in parallel in a single frame.

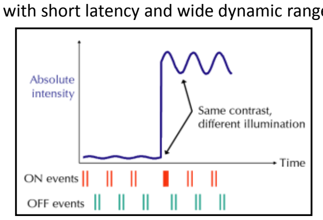


Principle of ETH/UZH Dynamic Vision Sensor (DVS)

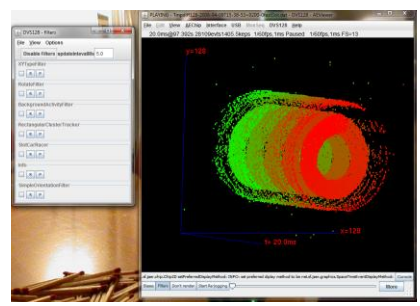
- Quick, sparse, & informative output for dynamic vision

- This vision sensor asynchronously emits digital *address-events* that encode the *addresses* of changing pixels.
- Each event means that the log intensity has changed by a quantized amount.

This operation efficiently encodes local changes in scene reflectance with short latency and wide dynamic range

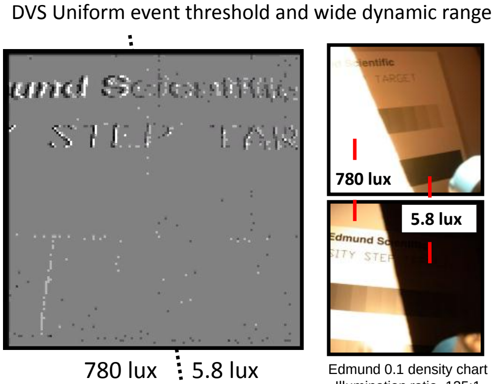


Demonstration of the DVS



Siliconretina.ini.uzh.ch

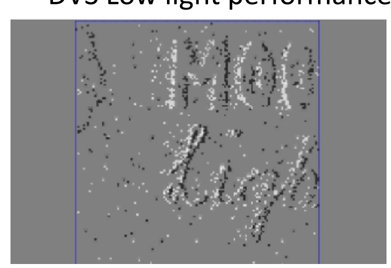
DVS Uniform event threshold and wide dynamic range



Siliconretina.ini.uzh.ch

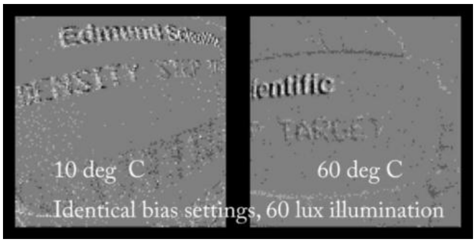
Edmund 0.1 density chart
Illumination ratio=135:1

DVS Low light performance



Shot under moonlight (<0.1 lux) with high contrast text
Photocurrent is <20% of dark current!
Keys to this ability
1) Low threshold mismatch
2) Pixels remember all change since last event

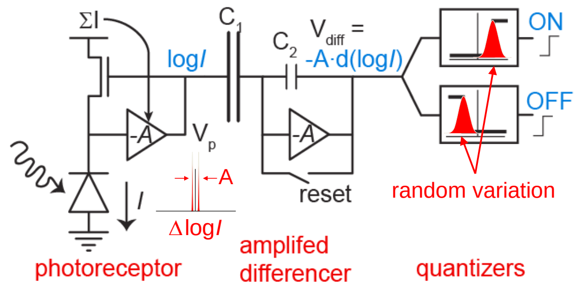
DVS Integrated biases enable unadjusted operation over a wide temperature range



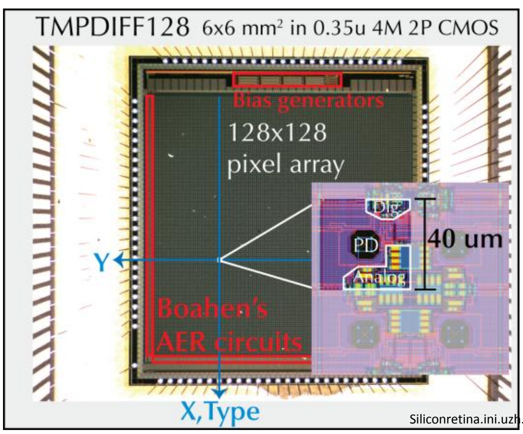
Siliconretina.ini.uzh.ch

How do the DVS pixels work?

Objectives:
Good event-threshold uniformity
Fast response under wide illumination range



Siliconretina.ini.uzh.ch



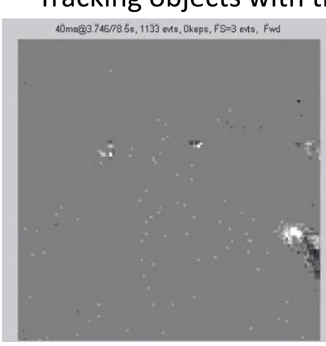
Siliconretina.ini.uzh.ch

DVS128 silicon retina cameras

Stream time-stamped address-events over high speed USB interface

Siliconretina.ini.uzh.ch

Tracking objects with the DVS events



Each event either

- ❑ moves an existing containing cluster, *or*
- ❑ Spawns a new cluster

- Starved clusters are pruned
- Overlapping clusters can be merged

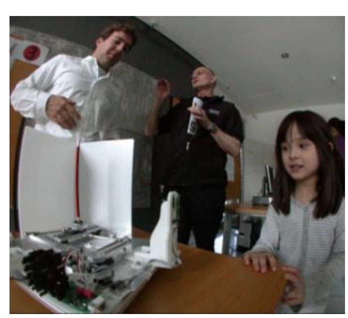
Advantages: 1. No frame memory (100 bytes/object).
2. No frame correspondence problem

20

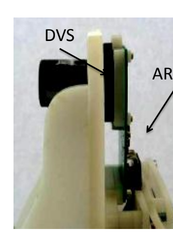
Siliconretina.ini.uzh.ch

Embedded pencil balancer with DVS Sensors

Jorg Conradt, Matthew Cook



eDVS
(Embedded DVS)
DVS+ARM7 microcontroller



Conradt et al., 2009

DVS application areas

- Low level feature extraction (Delbruck, Zurich)
- Car and person counting (AIT, Vienna)
- Fast robotic vision (Delbruck, Zurich)
- Neuromorphic spike-based hardware systems: CAVIAR
- Assembly line part identification (AIT, Vienna)
- Tracking grasping for spinal cord recovery (Rogister, Zurich)
- Eye tracking (Ersboell, DTU Lyngby, EU NoE COGAIN)
- Sleep – humans, mice, worms (Tobler/Winsky, UZH Zurich)
- Hydrodynamics (Hafliger and Jensen, Oslo)
- Tracking fruit fly wing beats (Fry, UZH-ETH Zurich)
- Tracking walking flies (Dickenson lab, Caltech)
- Human movement analysis (Perona lab, Caltech)
- Locust antennal movements (Huston, Caltech)
- Microscopic organisms and Brownian motion (Wu, Caltech)
- Tracking satellites (Assad, JPL)
- Fluorescence / Phosphorescence imaging (Arian, JPL)
- Calcium imaging of neural activity (Kanold, Maryland)
- Driving with spikes (Besselmann & Delbruck, Zurich)
- Reinforcement learning for slot car racing (Riedmiller, Germany)

Historical development of event-based integrated silicon retina vision sensors

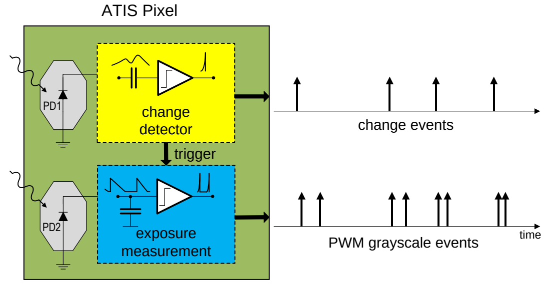
1990 1995 2000 2005

Mahowald & Mead outer retina (SciAm 91)
UPenn Magno-Parvo silicon retina (TBME 02)
CSEM VISe steerable-filter contrast vision sensor (JSSC 03)
Yale Univ. "Octopus" imager (JSSC 04)
JHU Temporal Change Threshold Detection Imager (JSSC 07)
ETH/UZH Dynamic Vision Sensor (JSSC 08)
Sevilla Spatial Contrast Silicon Retina (TCAS 08)
AIT Asynchronous Time-based Image Sensor (ISSCC 10)

25

Asynchronous Time-based Image Sensor (ATIS)

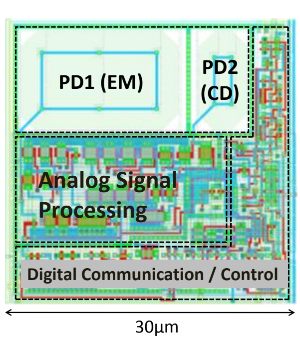
C. Posch, Austria Inst. of Technology



Posch et al. JSSC 2010

ATIS Pixel Layout – CMOS 0.18µ 6M MiM

- 77T, 4C, 2 PDs
- Fill factor: 30%
  10% CD
  20% EM



Posch et al. JSSC 2010

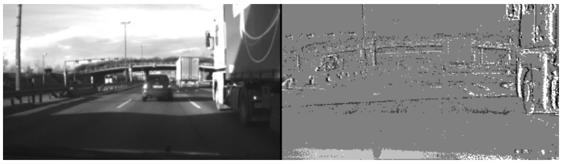
High Speed - Wide Dynamic Range

ATIS
vs.
Conventional Camera

- Temporal resolution: µs-range (10kfps – 1Mfps)
- DR >120 dB (standard CMOS/CCD: ~ 60-70dB)

Posch et al. JSSC 2010

ATIS Ego-Motion
DVS output in automotive applications



Posch et al. JSSC 2010

ATIS Ultrahigh Dynamic Range

125dB for **30fps** (video speed)

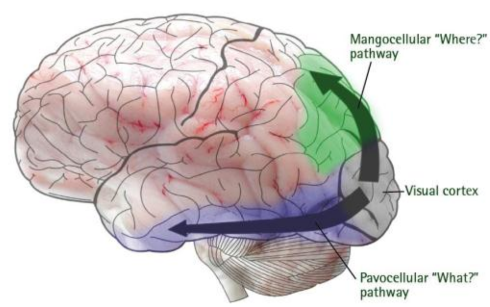
Posch et al. JSSC 2010

ATIS Pixel-level Video Compression

- ~QVGA continuous-time video stream
- 2.5k – 50k events/sec
- with 18bit/event **45k – 900k bit/sec**
- 30fps×8bit×QVGA = 18Mbit/sec (raw)
- Variable compression factor: **20 – 400** at the raw sensor output (no other embedded compression)

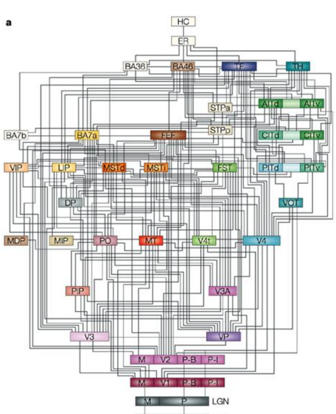
Posch et al. JSSC 2010

Going past the retina and simple vision



Culurciello

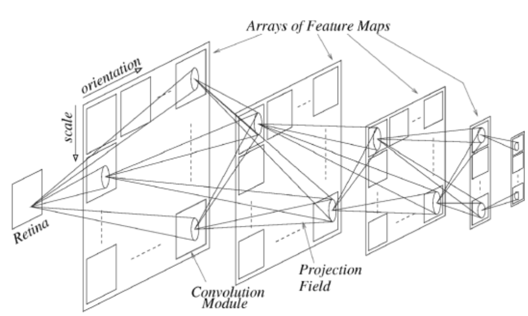
The visual cortex: A hierarchy of about 30 visual areas



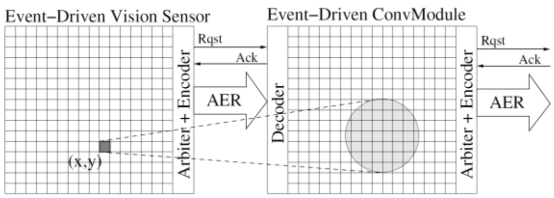
Felleman & van Essen 1991

Culurciello

Event-driven convolution processing.
Bernabe Linares-Barranco, IMSE, Sevilla



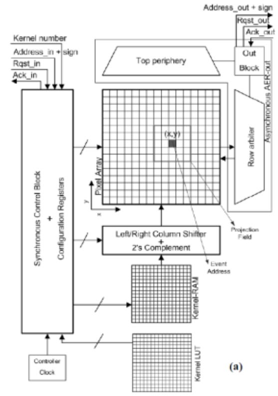
Projective AER convolution hardware module

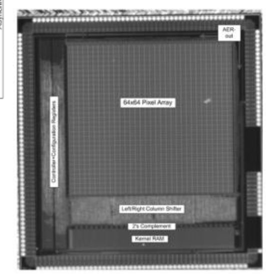


AER=Address-Event Representation

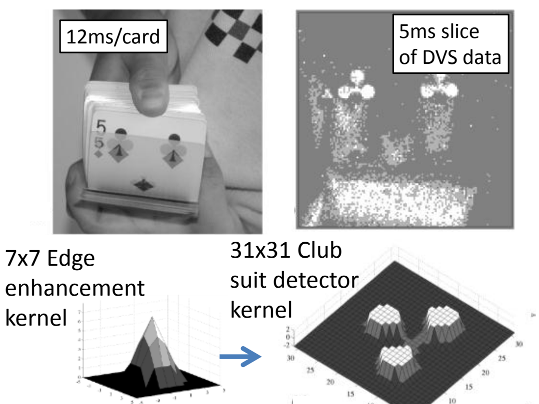
Linares-Barranco, IMSE, Sevilla

64x64 ConvModule

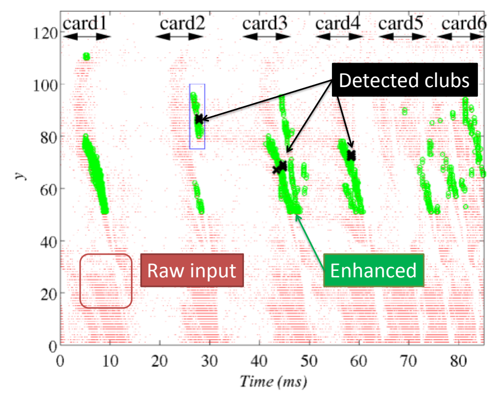




L. Camunas-Mesa et al., 2010



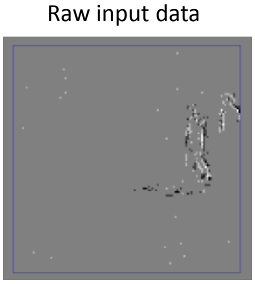
L. Camunas-Mesa et al., submitted 2010



L. Camunas-Mesa et al., submitted 2010

Gabor filtering with spike-based convolutions on FPGA

Raw input data

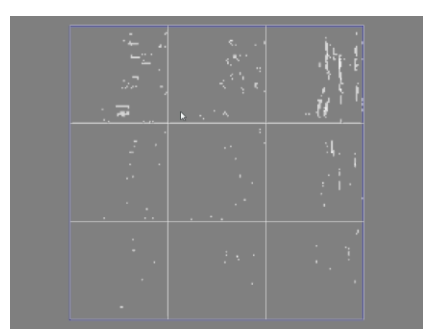


- Spike-based convolutional network is implemented in a Virtex-6 FPGA
- 3 orientations X 3 scales

Carlos Zamarreño Ramos, B. Linares-Barranco (unpublished)

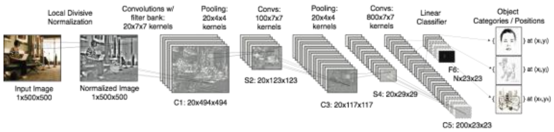
Gabor filtering with spike-based convolutions

9 kernels running in parallel on Virtex 6 FPGA



Carlos Zamarreño Ramos, B. Linares-Barranco (unpublished)

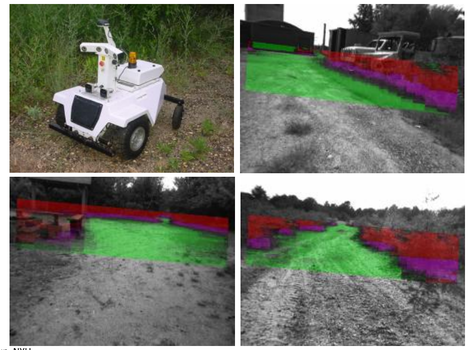
multi-layer convolutional neural networks LeCunn's "LeNet" and Neuflow.org



- convolutional layers, non-linearities, subsampling, convergence
- deep feed-forward multi-layer network: hierarchical system = invariance
- all parameters are learned from data
- implements models of the mammalian visual system

Neuflow.org, Le Cunn, E. Culurciello, 2010

- Obstacle avoidance, real-time guidance



Yann LeCun, NYU

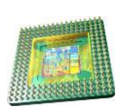
- and face detection / pose estimation

Yann LeCun, NYU

- scene parsing with deep neural networks

Farabet, Culurciello, LeCun 2011

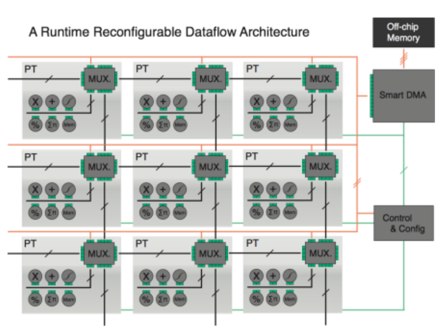
NeuFlow approach: fully digital!



- NeuFlow Processor! A DARPA-funded Flow-based Processor with a streaming instruction set
- An instruction set that allows macroscopic vision operations, e.g. 2D convolutions, local contrastive normalization, etc.
- Can implement complex image processing chains using a simple API in C/C++
- can be implemented in programmable hardware (FPGA) and also on custom VLSI micro-chips
- is essentially a SPECIALIZED GPU for models of vision!!!!

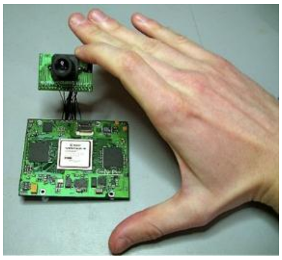
Farabet, Culurciello, LeCun 2011

inside the NeuFlow Processor



Farabet, Culurciello, LeCun 2011

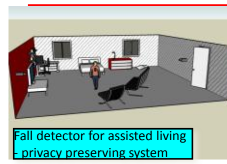
NeuFlow fits anywhere!

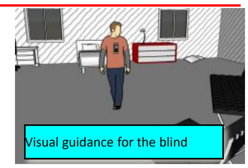


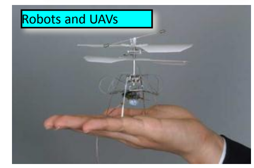
a home-made PCB that includes a Virtex4 and some quite large bandwidth to/from QDR memories

Farabet, Culurciello, LeCun 2011

Example applications of Neuflow







Farabet, Culurciello, LeCun 2011

Neuflow technical notes

- Fills a $2k Virtex 6 XC6VLX240T model... so quite large. Fit = uses all available routing fabric, logic utilization is actually ~25%. This FPGA has 240k logic cells, 3000kb of distributed RAM, 768 multipliers, and 14,000kb of distributed SRAM. Power: ~5-7W.
- IP-free, and entirely described in Verilog. Overall, it represents about 100,000 lines of Verilog code, and most of it is 'generative' meaning that the code replicates itself with different parameters.
- Uses a custom CPU (64-bit, with a pseudo vector-oriented instruction set), which can reconfigure most of the connections within the grid (between ALUs, from each ALU to the DMA ports), and the role of each ALU, at runtime.
- Compiler takes xFlow and automatically generates binary code for this CPU. You can see this binary as being the runtime sequence of grid reconfigurations and DMA (memory) accesses required to compute the algorithm described in xFlow.
- Targeting a standard 45nm IBM process, it requires 5x3mm of silicon (with less than 100kB SRAM overall). MPW prototype will cost only $8k! (special deal obviously)

Neuflow.org

Retrospective: Computer vs. Brain

At the system level, brains are at least 1 million times more power efficient than computers. Why?

Cost of elementary operation (turning on transistor or activating synapse) is about the same. It's not some magic about physics.

Computer	Brain
Fast global clock	Self-timed, data driven
Bit-perfect deterministic logical state	Synapses are stochastic! Computation dances digital→analog→digital
Memory distant to computation	Synaptic memory at computation
Fast, high resolution, constant sample rate analog-to-digital converters	Low resolution adaptive data-driven quantizers (spiking neurons)

Mobility of electrons in silicon is about 10^7 times that of ions in solution.

53/30