Bio-inspired Vision
and what electronics and computers can learn from nature

Christoph Posch
Austrian Institute of Technology AIT

TWEPP 2011
Inspired by Biology?

- Biology: **Highly efficient** machines that greatly **outperform** any man-made technology …

- Even **small/simple animals** like the bee displays flight motor skills and cognitive behaviors, **out of reach for any artificial sensor/processor/actuator system**.
  - body weight of < 1 gram
  - brain weighing few micrograms
  - dissipating power of ~ 10µW

- Nature achieves **efficient and reliable** computation based on **fuzzy input** data in an **uncontrolled** environment
  - How is nature doing this?
  - Can we **learn** from nature?
Brain vs. Computer

Biological brains and digital computers are both complex information processing systems. But here the similarities end

<table>
<thead>
<tr>
<th>Brains:</th>
<th>Computers:</th>
</tr>
</thead>
<tbody>
<tr>
<td>imprecise</td>
<td>precise</td>
</tr>
<tr>
<td>error-prone</td>
<td>deterministic</td>
</tr>
<tr>
<td>slow</td>
<td>fast</td>
</tr>
<tr>
<td>flexible</td>
<td>inflexible</td>
</tr>
<tr>
<td>concurrent</td>
<td>serial</td>
</tr>
<tr>
<td>adaptive - tolerant of component failure</td>
<td>susceptible to single-point failure</td>
</tr>
<tr>
<td>autonomous learning</td>
<td>program code</td>
</tr>
</tbody>
</table>

- Can understanding of brain function point the way to more efficient, fault-tolerant computation?
- Why is it important?
Energy Efficiency

Progress of electronic information processing over past 60 years:
- dramatic improvements:
- from 5 Joules / instruction (vacuum tube computer, 1940s)
- to 0.0000000001 Joules / instruction (ARM968)
- 50,000,000,000 times better
- Raw performance increase about 1 million

Energy efficiency
- Chip: $10^{-11}$ J/operation
- Computer system level: $10^{-9}$ J/operation
- Brain: $10^{-15}$ J/operation
- Brain is 1 million times more energy efficient!!!

Where is the Energy?

1 Million

- Cost of elementary operation – turning on transistor or activating a synapse – is about the same. \((10^{-15}\text{J})\)

- Lose a factor 100 because:
  - capacitance of gate is a small fraction of capacitance of the node
  - spend most energy charging up wires

- Use many transistors to do one operation (typically switch \(10000\)).
  - information encoding: “0”, “1”
  - elementary logic operations (AND, OR, NOT)

C. Mead: “We pay a factor 10000 in energy for taking out the beautiful physics from the transistor, mash it up into “0” and “1” and then painfully building it back up with gates and operations to reinvent [e.g.] the multiplication …”

---

Can we get away with this?

so far we can … but for how much longer?

- From observation to planning tool
- Industry invests to make it happen

... things in silicon VLSI are getting tough ...

Credit: S. Furber, Manchester Univ.
Limitations I: Devices

- Control of electron cloud depends critically on statistics of the components: fewer components $\rightarrow$ less robust statistics
- Transistors less predictable, less reliable

- Graphene transistors?
- Nano wires?
- Fin-FETS?
- Molecular transistors?

Limitations II: Computing Architecture

Clock race: Single processors running faster and faster …

- no more!!

- Going parallel!
- + probably need to:
  - avoid synchronicity
  - abandon determinism
Computing Power: Human Brain vs. Computer

- Massive parallelism (10^{11} neurons)
- Massive connectivity (10^{15} synapses)
- Low-speed components (~1 – 100 Hz)
- >10^{16} complex operations / second (10 Petaflops!!)
- 10-15 watts!!
- 1.5 kg

„K computer“ (RIKEN, Japan)

8.162 petaflops
9.89 MW


http://www.nsc.riken.jp
Neurons are similar across a wide range of biological brains, and like logic gates that are “universal” in the sense that any digital circuit can be built using the same basic gates.

Both are multiple-input single-output devices, but:

<table>
<thead>
<tr>
<th>Neurons:</th>
<th>Logic Gates:</th>
</tr>
</thead>
<tbody>
<tr>
<td>fan-out: 1000-10000</td>
<td>fan-out: 2-4</td>
</tr>
<tr>
<td>dynamic, several time constants</td>
<td>static internal process</td>
</tr>
<tr>
<td>output is time-series of “spikes”</td>
<td>output is well-defined stable function of inputs</td>
</tr>
<tr>
<td>fires at 10s to 100s Hz</td>
<td>defined by boolean logic</td>
</tr>
<tr>
<td>information is encoded in timing of spikes</td>
<td></td>
</tr>
</tbody>
</table>

http://webspace.ship.edu/cgboer/theneuron.html
Communication and Processing

- Neurons talk to each other via of dendrites and axons
- Transmitting electrical impulses - ’spikes’ - from one neuron to another
- Most of the processing happens in the junctions between neurons – the synapses
- Storage (synapse stores state) and processing (evaluating incoming signal, previous state and connection strength) happen at the same time and in the same place.
- This locality is one key to energy efficiency
„Neuromorphic“ Engineering


- Silicon VLSI technology can be used to **build circuits** that mimic **neural functions**

- Silicon primitive: **transistor** – much physics similar to **neurons**

- Building blocks: **neurons, axons, ganglions, photoreceptors, …**

- Biological **computational primitives**: logarithmic functions, excitation/inhibition, thresholding, winner-take-all selection …

Building Brains

- Mostly **limited-scale**: multi-neuron chips, synapse arrays, convolution chips etc.
- Initially for pure scientific purposes – now more and more for solving real-world **engineering and computing problems**
- Emerging technologies like **memristors** are investigated
- Some remarkable “big-scale” projects attempt the **scale** of a mammalian brain → final frontier: the **human brain**

- **“Neurogrid”** – K. Boahen, Stanford: 1 million neurons with 6 billion synapses in mixed analog/digital VLSI.
- **“SpiNNaker”** – S. Furber, Manchester Univ., (20 processors/chip each simulating ~1000 neurons, 65000 chips in 2D toroidal mesh)
- **“Brainscales/FACET”** – K.H. Meier. Univ. Heidelberg: CMOS wafer-scale integration of analog multi-neuron chips (400 neurons/10000 synapses per chip, up to $10^8$ neurons on the wafer system)
- **“Blue Brain” Project** – Henry Markram, EPFL Lausanne → final goal: human brain running on IBM Blue Gene/L (360TFLOPS)
Biology-Inspired „Neuromorphic“ Vision

- Very successful branch of neuromorphic engineering: sensory transduction → vision
- “Silicon Retina” (Mahowald, Mead; 1989)
- Neuromorphic vision sensors sense and process visual information in a **pixel-level, event-based, frameless** manner
- Vision **processing** is practically **simultaneous** to vision **sensing**
- Only **meaningful information** is sensed, communicated, and processed
Fig. 4. Exterior view of the electronic model of the retina.

History of integrated silicon retina vision sensors

- 1990: Mahowald/Mead: Silicon Retina (SciAm 91)
- 1995: CSEM: VISe contrast vision sensor (JSSC 03)
- 2000: JHU/UPenn: “Octopus” imager (ISSSC 01)
- 2005: Stanford/UPenn: 5-Layer silicon retina (SciAm 05)
- 2005: JHU: Temporal Change Detection Imager (JSSC 07)
- 2008: IMSE: Spatial Contrast Silicon Retina (TCAS 08)
- 2011: ATIS – Asynchronous Time-based Image Sensor (JSSC 2011)
Limitations of Conventional Image Sensing

Conventional image sensors acquire the visual information as “snapshots”
time-quantized @ frame rate

- World works in **continuous time**: things happen **between** frames
- **Each frame** carries the information from all pixels – whether or not this information has **changed** since the last frame

**Biological approach**: Not “blindly” sense/acquire redundant data but **respond** to visual information:

- Generate only **meaningful** data near **real-time**
- **Reduce data** rate → decrease demands on bandwidth / memory / computing power for data transmission / storage / post-processing
The Human Retina

- **135 million** photoreceptors – detection threshold (rod): 1 photon
- **1 million** ganglion cells in the retina **process** visual **signals** received from groups of (few to several hundred) **photoreceptors**.
- Analog **gain control**, **spatial** and **temporal filtering**: ~ 36 Gb/s HDR raw image data is compressed into ~ 20 Mb/s spiking output to the brain
- Retina encodes **useful** spatial-temporal-spectral **features** from a redundant, wide dynamic range world into a small internal signal range.
- Power consumption: ~ 3.5 mW
Biology-Inspired Dynamic Vision

Neuromorphic Dynamic Vision Sensor (DVS)

- Array of light-controlled “integrate-and-fire” neurons driving local bipolar cells
- Sensor models the Magno-cellular transient pathway, constitutes a simplified three layer model of the human retina
- Individual pixels respond to relative change (temporal contrast) by generating asynchronous pulse events
- Pixels operate autonomously – no external timing signals → No frames!
- Sensor is event-driven instead of clock-driven → responds to “natural” events happening in the scene
- Temporal resolution is not quantized to a frame-rate

Lichtsteiner, P.; Posch, C.; Delbruck, T., "A 128 × 128 120dB 30mW asynchronous vision sensor that responds to relative intensity change," ISSCC 2006 (JSSC 2008)
DVS Operation

- Pixel **asynchronously** and in **continuous time** responds to **relative change** in illuminance
- Generates ‘**spike**’ events
- For each spike the **x,y-address** is put on an asynchronous bus → **Address-Event-Representation**

**Key characteristics:**
- Wide dynamic range (> 120dB)
- High temporal resolution (< 1µs)
- Low latency (< 10µs)
- High contrast sensitivity (~ 8%)

Lichtsteiner, P.; Posch, C.; Delbruck, T., "A 128 × 128 120dB 30mW asynchronous vision sensor that responds to relative intensity change," **ISSCC 2006 (JSSC 2008)**
Temporal Contrast Events
High Speed - Wide Dynamic Range

- **Temporal resolution**: μs-range (equivalent 100,000 to 1,000,000 frames/s)
- **Dynamic range**: >120 dB (standard CMOS/CCD: 60 – 70dB)
Next-generation bioinspired imaging „where“ and „what“!

Magno- and Parvo- ganglion cells – have very different spatio-temporal characteristics

- **Transient** Magno-cellular pathway – alerting $\rightarrow$ “where” system
- **Sustained** Parvo-cellular pathway – detailed vision $\rightarrow$ “what” system
- Next step: Add biological “What” function to transient temporal contrast sensing.

- **Sustained pathway** principle encodes absolute intensity (gray-levels) in asynchronous pulse events.

$\rightarrow$ Array of **pixels** that:

- **individually** and **autonomously** react to scene changes **and**
- acquire illumination information **conditionally** and **event-driven**

**Asynchronous Time-based Image Sensor – ATIS**
ATIS Pixel – Basics

- Two blocks: DVS change detector and exposure measurement
- Change detector triggers exposure measurement only after a detected change in the pixel’s FoV
- Continuous-time asynchronous operation
- Communicates detected changes AND new exposure information independently and asynchronously – NO FRAMES
- Intensity-encoding is time-based
ATIS Pixel – The Complete Picture

pixel detail: change detector

change detector - waveforms

typical pixel signals

ATIS pixel

photo measurement

change detector

logic

comparison

handshake protocol
ATIS Pixel Layout – CMOS 0.18µm

- 0.18µm CMOS
- 30 × 30µm²
- 77T, 4C, 2 PDs
- Fill factor: 30%
  10% CD
  20% EM
- Analog part:
  - change detector
  - integration
  - comparator
- Digital part
  - pixel-level state logic
  - communication
  - handshaking
ATIS Concept – Implications

Pixel-autonomous change-detector controlled operation

- Pixel does not rely on any external timing signals
- Pixel that is not stimulated visually does not produce output
- Complete suppression of temporal data redundancy = lossless pixel-level video compression

Asynchronous time-based encoding of exposure information

- **Avoids** the time quantization of frame-based acquisition and scanning readout.
- Allows each pixel to choose its own optimal integration time instead of imposing a fixed integration time for the entire array.
- Yields exceptionally high dynamic range (DR) and improved signal-to-noise-ratio (SNR).
- DR is not limited by power supply rails
# ATIS – Specifications

<table>
<thead>
<tr>
<th>Specification</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fabrication process</td>
<td>UMC L180 MM/RF 1P6M CMOS</td>
</tr>
<tr>
<td>Supply voltage</td>
<td>3.3V (analog), 1.8V (digital)</td>
</tr>
<tr>
<td>Chip size</td>
<td>9.9 × 8.2mm² (6.8 Mio. transistors)</td>
</tr>
<tr>
<td>Optical format</td>
<td>2/3“</td>
</tr>
<tr>
<td>Array size</td>
<td>QVGA (304 × 240)</td>
</tr>
<tr>
<td>Pixel size</td>
<td>30µm × 30µm</td>
</tr>
<tr>
<td>Pixel complexity</td>
<td>77T, 3C, 2PD</td>
</tr>
<tr>
<td>Fill factor</td>
<td>30% (20% EM, 10% CD)</td>
</tr>
<tr>
<td>Integration swing ΔV&lt;sub&gt;th&lt;/sub&gt;</td>
<td>100mV to 2.3 V (adjustable)</td>
</tr>
<tr>
<td>SNR typ.</td>
<td>&gt;56dB (9.3bit) @ ΔV&lt;sub&gt;th&lt;/sub&gt; = 2V, &gt;10Lx</td>
</tr>
<tr>
<td>SNR low</td>
<td>42.3dB (7bit) @ ΔV&lt;sub&gt;th min&lt;/sub&gt; (100mV), 10Lx</td>
</tr>
<tr>
<td>t&lt;sub&gt;int&lt;/sub&gt; @ ΔV&lt;sub&gt;th min&lt;/sub&gt; (100mV)</td>
<td>2ms @ 10Lx (500 fps equ. temp. res.)</td>
</tr>
<tr>
<td>DR (static)</td>
<td>143dB</td>
</tr>
<tr>
<td>DR (30fps equivalent)</td>
<td>125dB</td>
</tr>
<tr>
<td>PRNU / FPN</td>
<td>&lt;0.25% @ 10Lx (with TCDS)</td>
</tr>
<tr>
<td>Power consumption</td>
<td>50mW (static), 175mW (high activity)</td>
</tr>
<tr>
<td>Readout format</td>
<td>Asynchronous AER, 2 × 18bit-parallel</td>
</tr>
</tbody>
</table>
Ultrahigh Dynamic Range

- 143dB for static scene
- 125dB for 30fps (video speed) equivalent temporal resolution
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
Behind the Scenes
Conclusions

DVS real-time contrast data

- Wide dynamic range (> 120dB)
- High temporal resolution (< 1µs)
- Low latency (< 10µs)
- Contrast sensitivity ≥ 10%

High-speed dynamic machine vision

- Industrial robotics
- Micromanipulation
- Autonomous robots and AUVs
- Automotive

... where need for speed meets uncontrolled lighting conditions!

ATIS frame-free, compressed video

- Wide dynamic range (125–143 dB)
- Fast (500 fps equ.temp.res. @ 10lx)
- High image quality (56dB SNR)
- Compression up to 1000 (static)

Low-data rate video

- Wireless sensors
- Sensor networks
- Web video

Wide DR, high-quality imaging/video

- Cell monitoring
- X-ray crystallography
- Astronomy
Thank you for your attention!
Outline

- Bio-inspired?
- Computers and brains: similarities and differences
- Limitations to digital, synchronous information processing
- Computational primitives – neurons vs. logic gates
- “Neuromorphic Engineering” – building brains
- Vision – biological / bio-inspired
- Modeling the retina – ”Silicon Retina”
- CMOS implementations of bio-inspired vision chips
- Applications
- Outlook
Neural Circuits – Cortical Architecture

- **Regular high-level** structure
  - e.g. 6-level cortical micro architecture
    - low-level vision
    - language, ...

- **Random low-level** structure
  - adapts over time
  - synaptic connections change
  - weights change
  \( \rightarrow \text{learning!!} \)
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. (10^{-15} \text{ J})

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

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

Biological Vision – Retina Ganglion Cells

Two different types of retinal ganglion cells and corresponding retina-brain pathways:

**Magno- and Parvo-Cells** – very different spatio-temporal characteristics

**Magno-cellular pathway** – *transient channel*
- receptors *evenly distributed* over retina, big (low spatial resolution)
- *short latencies*, rapidly conducting axons (high temporal resolution)
- respond to *changes*, movements, onsets, offsets (transient response)
- biological role in *alerting*, detecting dangers in our peripheral vision
- “Where” system

**Parvo-cellular pathway** – *sustained channel*
- receptors concentrated in the fovea, *small* (high spatial resolution)
- have *longer latencies* and slower conducting axons (low temporal res.)
- respond as long as *visual stimulus is present* (sustained response)
- transportation of *detailed visual information* (spatial details, color)
- “What” system

(A. v.d.Heijden, „Selective attention in vision“)
Bioinspired Vision – Events vs. Frames

- Conventional imagers:
  - **Neglect dynamic** visual information
  - Acquire **frames** at discrete points in time

- A lot of interesting **information** in the **dynamic** contents of a **scene**
- Things happen between frames …
- New paradigm of visual sensing and processing
  → “Event-based vision”
Going past the retina and simple vision
Event-based Visual Cortex – Convolution

Bernabé Linares-Barranco, Instituto de Microelectrónica de Sevilla
Projective AER convolution hardware module

AER = Address-Event Representation

Linares-Barranco, IMSE, Sevilla
Building a Self-Learning Visual Cortex with Memristors

Figure 21: (A) Projection Field Topology of V1 layer in Visual Cortex. (B) Hybrid CMOS-memristor arrangement with CMOL style tilted lines.
Selected publications


- Lichtsteiner, P.; Posch, C.; Delbruck, T., "A 128 × 128 120db 30mW asynchronous vision sensor that responds to relative intensity change," *Solid-State Circuits, 2007 IEEE International Conference, Dig. of Technical Papers, ISSCC 2006*. “*ISSCC 2006 Jan Van Vessem Award for Outstanding European Paper*”

Applications of DVS

DVS events are well suited to drive computer vision systems

- Fast visual feedback loops
- Motor control
- Industrial robotics
- Autonomous robots and AUVs
- Automotive

... where **need for speed** meets **uncontrolled lighting conditions**!

applies DVS in
- Traffic data acquisition
- People counting, people flow monitoring
- Ambient assisted living (fall detection)
- Gesture recognition
Event-based vs. Frame-based Processing

Frame-based (conventional)
- camera captures a sequence of frames
- each frame is transmitted to a computing system
- processed by sophisticated image processing algorithms for achieving some kind of recognition
- computing system needs to have all pixel values of a frame before starting any computation
- reality is binned into compartments of duration $T_{\text{frame}}$
- computing system has to process the full frame, handling large amounts of data

Event-based (bio-inspired)
- pixel sends an event (usually its own x,y coordinate) when it senses something – asynchronous, real-time
- events are transferred to computing system as they are produced
- computing system updates its state after each event
- events are processed as they flow – sensing and processing is done concurrently – no need to wait for frames
- For performing recognition not all events are necessary
Pixel circuit – “Integrate-and-Fire” neuron

- **Photoreceptor:** logarithmic intensity → gain control mechanism that is sensitive to **temporal contrast = relative change**
- **Differencing amplifier** removes DC, amplifies pos/neg transients
- Two **threshold comparators** monitor ganglion output, spike generation
- **Output:** asynchronous spike events (circuit is reset after each event)
Encoding Gray-Level Data in Time: PWM Imaging

Measured SNR vs. integration swing $\Delta V_{th}$ and light intensity:

SNR = 56dB for $\Delta V_{th} = 2V/10Lx$

(42.3dB for $\Delta V_{th} = 100mV/10Lx/2ms$)

$\rightarrow$ 9.3 Bit grayscale resolution

FPN < 0.25%
CMOS Neuromorphic Vision Sensors Gallery

**Tmpdiff128 – 128 × 128 pixel array sensor**
high-dynamic-range, low-power temporal contrast dynamic vision sensor with in-pixel analog signal processing

**DLS – 2 × 256 pixel line sensor**
dual-line optical transient sensor with on-chip precision time stamp generation and digital arbitration for high-speed vision
ISSCC 2007

**DVS-IR – 64 × 64 pixel IR array sensor**
 transient vision sensor for the thermal infrared (IR) range with micro-machined bolometer IR detector technology

**ATIS – QVGA event-driven dynamic vision and image sensor**
QVGA ultra-wide dynamic range CMOS imager and dynamic vision sensor with focal-plane lossless video compression
Wiring – address-event representation

- How to get the data off the array quickly?
- Mobility of electrons in silicon is about $10^7$ times that of ions in solution,
- Signal transmission speed: lightspeed vs. few meters/second
- Solution: **Softwiring**
  - Each neuron is assigned an “address” → Address Event (AE)
  - When neuron fires it pushes its address onto a shared asynchronous bus
  - Asynchronous digital circuits map and route the address events to other nodes or different chips or (external) processing units
Processing and Storage

- N1 spikes—pulse travels down the axon to the synapse of target N2.
- The synapse of N2—having stored its own state locally—evaluates the importance of the information coming from N1 by integrating it with own previous state and strength of connection to N1.
- Two pieces of information—signal from N1 and state of N2's synapse—flow toward body of N2.
- When information reaches N2, there is only a single value—all processing has already taken place during the information transfer.
- Storage and processing happen at the same time and in the same place.
- This LOCALITY is one of main reasons for energy efficiency of biological brains.