



***On the Design of
CMOS
Vision Systems on Chip***

Angel Rodríguez-Vázquez

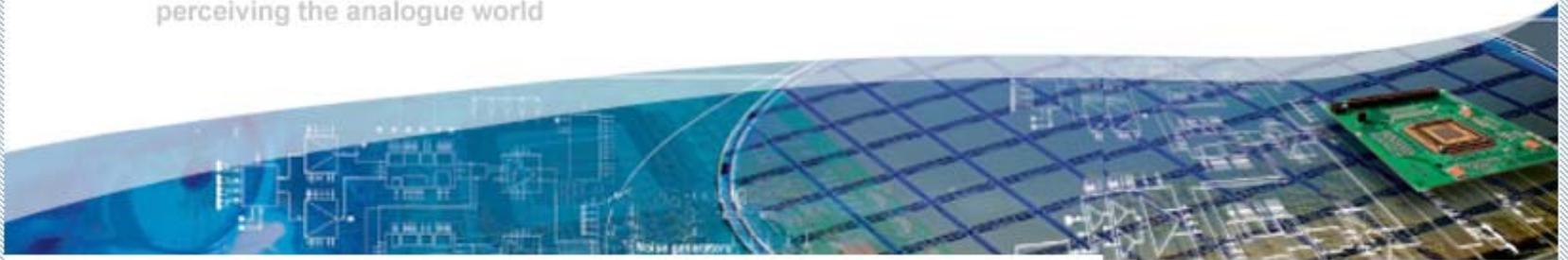
***IEEE CAS Workshop 2008
Rio de Janeiro, August 27 - 2008***

Outline of the Talk

- ***Some Basic Concepts:***
 - ◆ ***Concept of Vision Systems and VSoCs***
 - ◆ ***Conventional Vision System Architecture***
 - ◆ ***Rationale for Using CMOS***
- ***Architectures for Vision Systems***
 - ◆ ***Progressive Distributed Processing***
 - ◆ ***Bio-Inspired Vision System Architectures***
 - ◆ ***Shifting the Analog-to-Digital Border***
- ***Topographical Sensor-Processors***
 - ◆ ***Concept of Concurrent Sensory-Processing***
 - ◆ ***Multi-Functional Pixel***
 - ◆ ***Rationale for Analog Pre-Processing***
- ***The Eye-RIS Vision Systems***
 - ◆ ***Architecture***
 - ◆ ***The System in Operation***
 - ◆ ***Some Prospects for Future Developments***



perceiving the analogue world

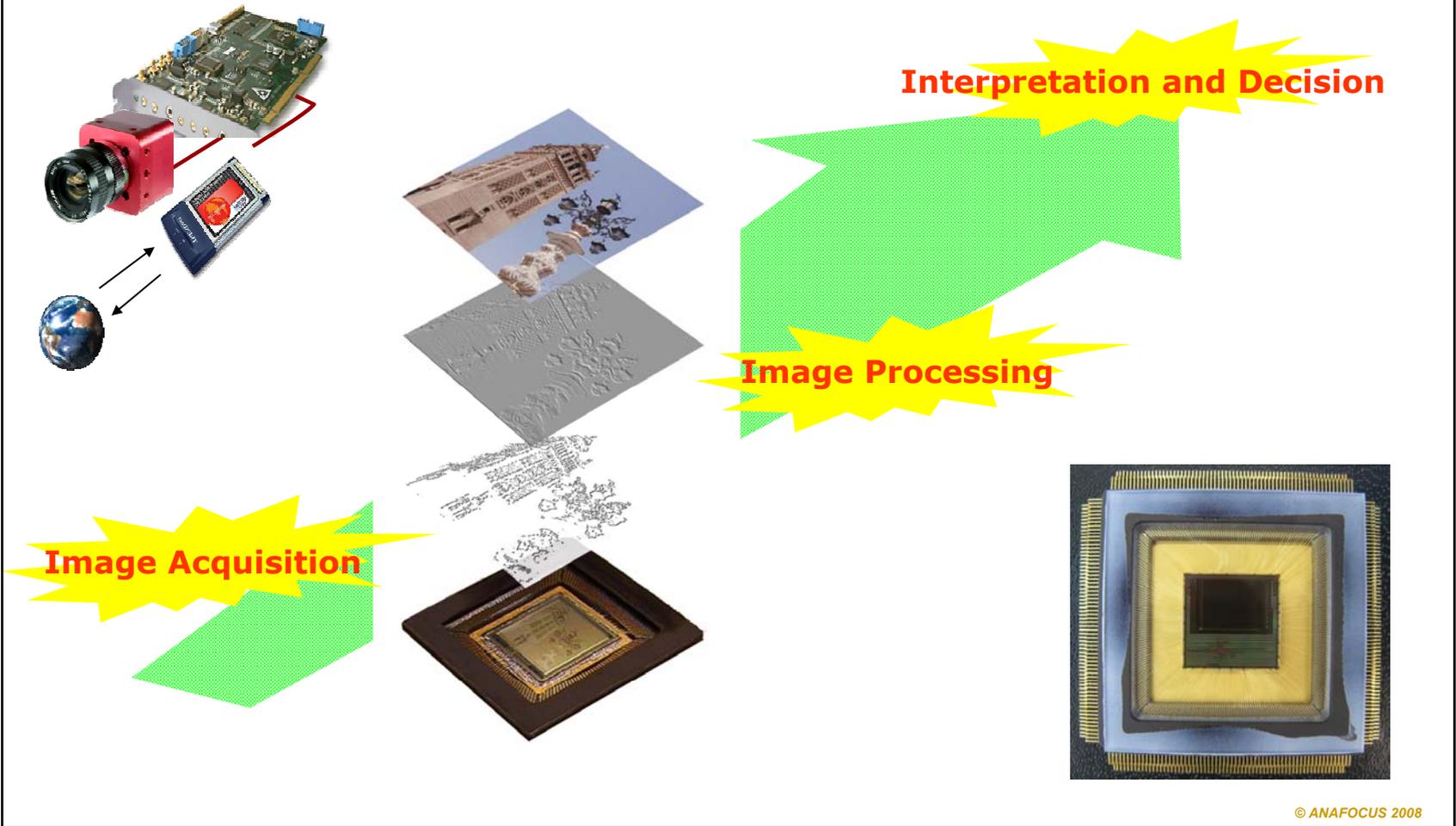


Some Basic Concepts

Concept of Vision Systems

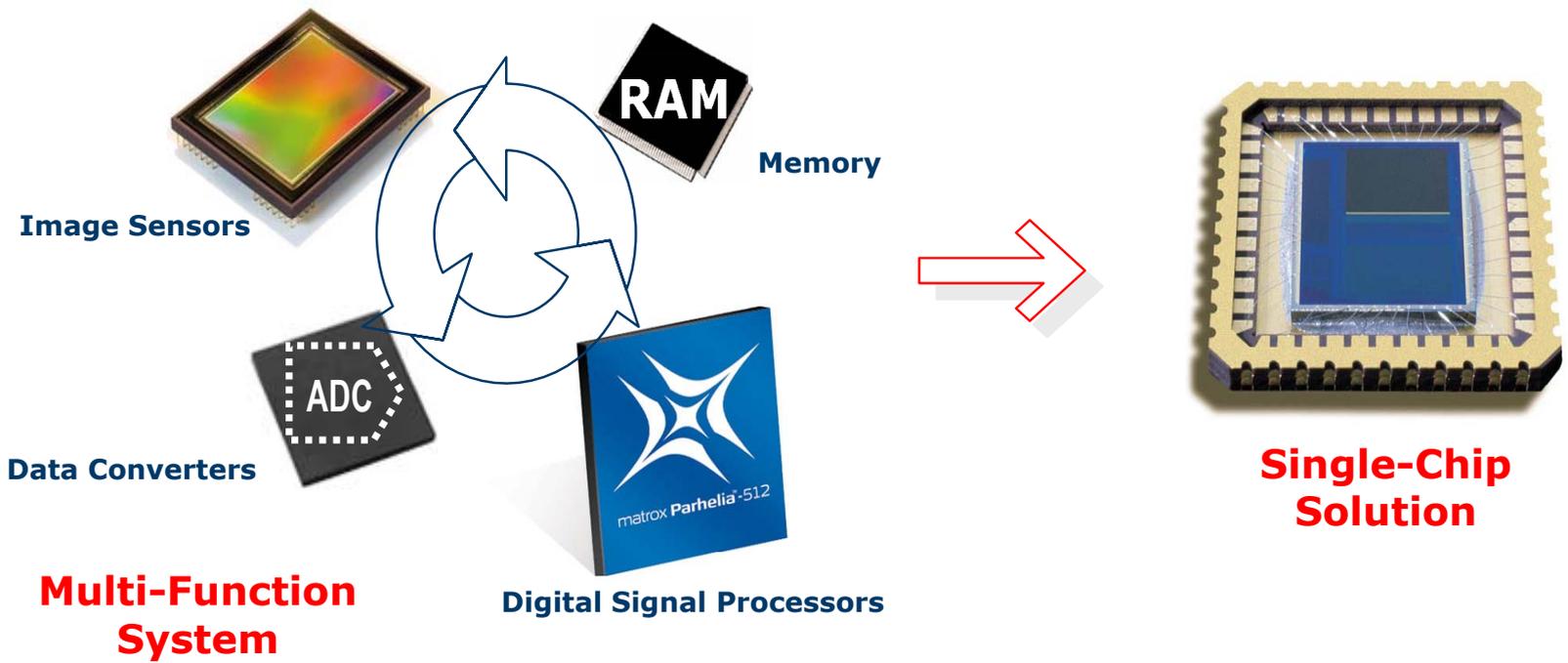
Conventional MV Architectures

Rationale for CMOS Implementation

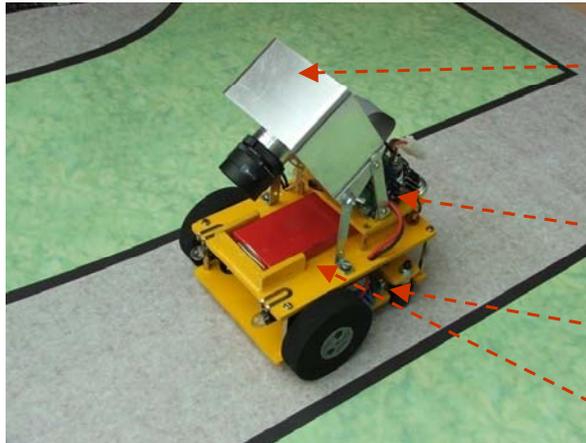




All functions for smart imaging and vision systems on chip



Systems capable to make autonomous decisions for vision-based guidance



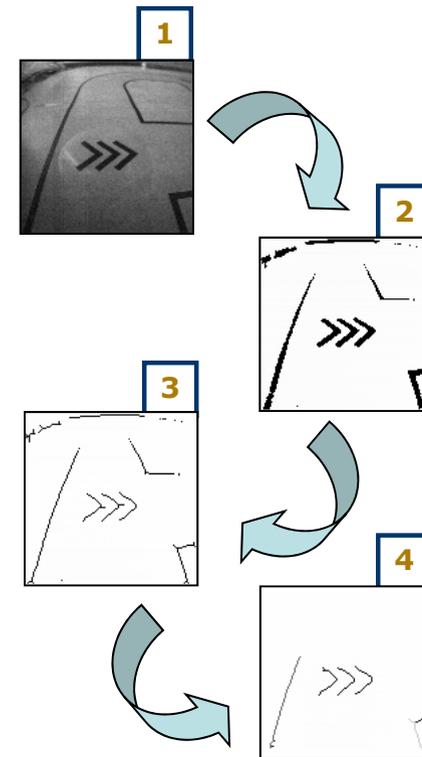
Eye-RIS

Voltage regulator board

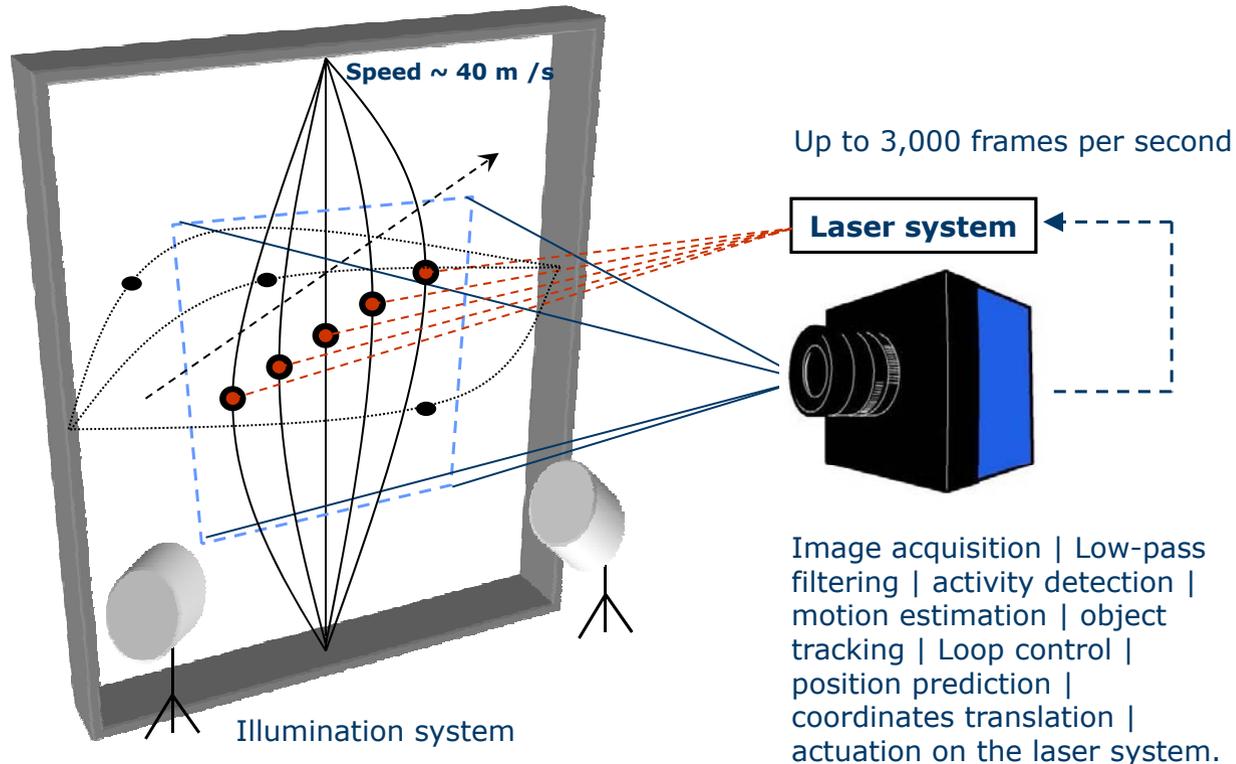
Motor controller board

Battery

Image acquisition | Low-pass filtering | thresholding | skeletonization | morphological filtering | signal recognition | road-lanes recognition | actuation on the robot wheels.



Systems to close the **Perception-to-Action Loop** at **Thousands Images-per-Second**





- ⊙ **Perceive lightness and colour under various illuminations**
- ⊙ **Detect intensity changes and perform 2-D segmentation**
- ⊙ **Infer 3-D structures from stereo or motion images**
- ⊙ **Organize surface and regions into objects of interest**
- ⊙ **Generate description of objects and recognize them among a potentially large class of objects**
- ⊙ **Make non-visual inferences about the scene based on visual processing (abstraction)**

Emulate the capabilities of human visual systems



Task n°	Processing Task	Example	Amount of data
①	Image Acquiring	Light sensing	$b \times N^2$
②	Noise reduction	Lowpass, median filtering...	$b \times N^2$
③	Segmentation	Edge detection	$b \times N^2$
④	Feature extraction	shape detection	N^2
⑤	Classification	identification	N
			1

\downarrow N=image width & height
 \downarrow b=number of bits of resolution



☹️ **Huge amount of data at Early Stages**

RGB 8-bit coded @ 512 x 512 implies 200Mbit/s data rate

☹️ **Most data are useless**

Per-Pixel 3x3 Linear Convolution

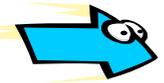
9 Products

+

8 Sums

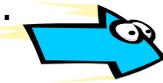
17 Operations

If the image is $N \times M$



$17 \times N \times M$ Oper.

If the algorithm contains N_s convs.



$17 \times N_s \times N \times M$ Oper.

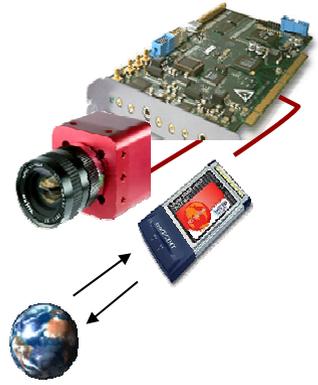
At F Frames per second



$17 \times F \times N_s \times N \times M$ OPS

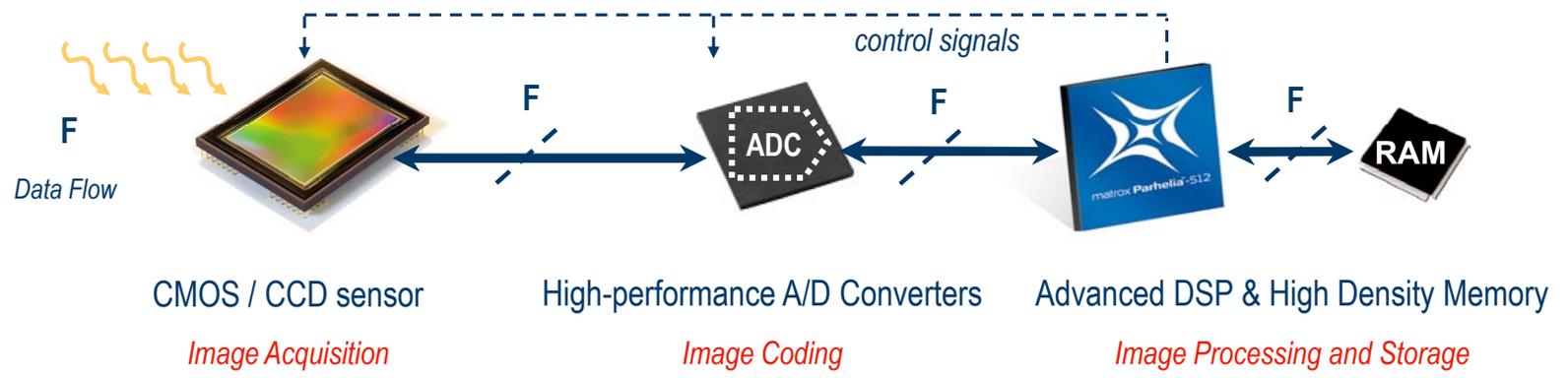
10 convolutions on 8-bit gray-scale VGA images at 100 FPS requires 5 GOPS

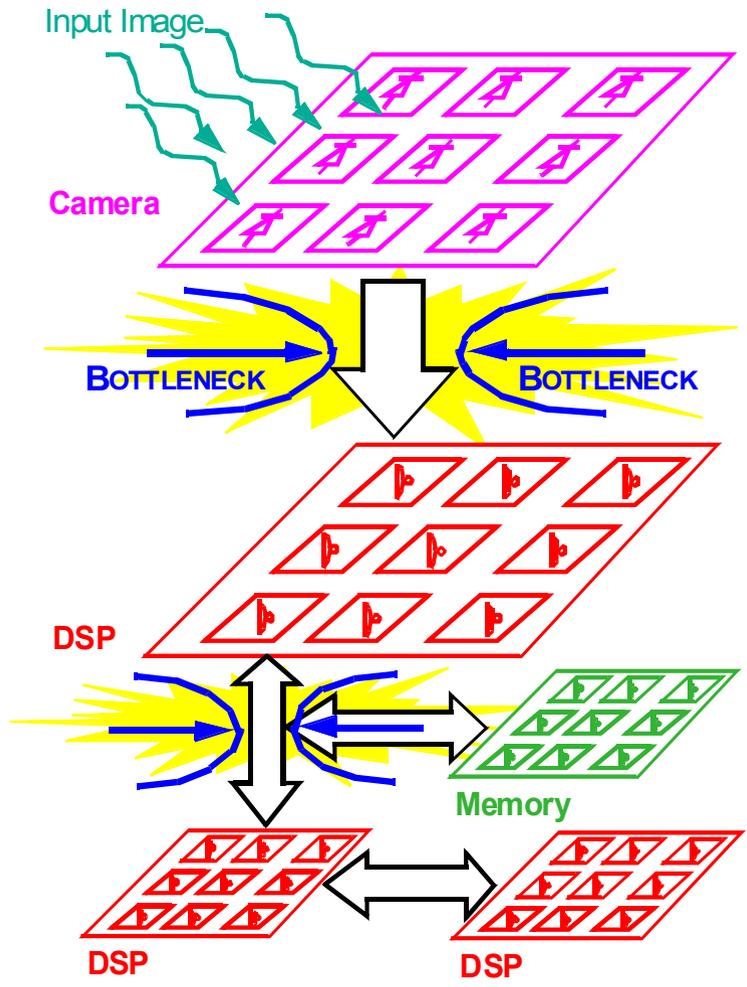
Memory operations must be accounted for as well



- **Conventional systems are heterogeneous**
Sensors, Memory, Processors, Communications

- **Involve multiple technologies**





- **Physical separation:**

Sensors

Processors

- **Heterogeneous technologies**

CCD or CMOS for sensing

CMOS: FPGA, Power-PC, DSP, etc for processing

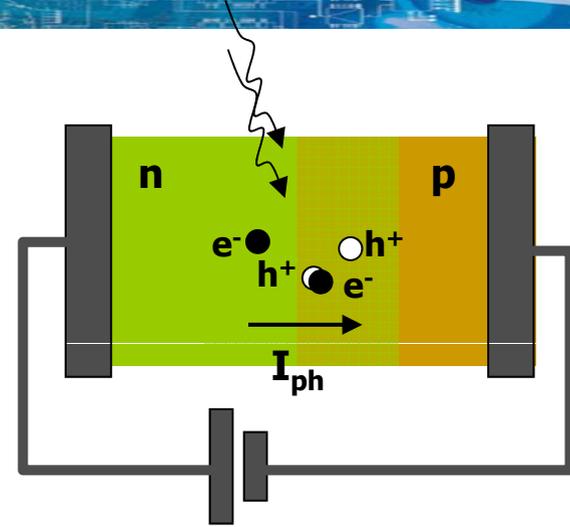
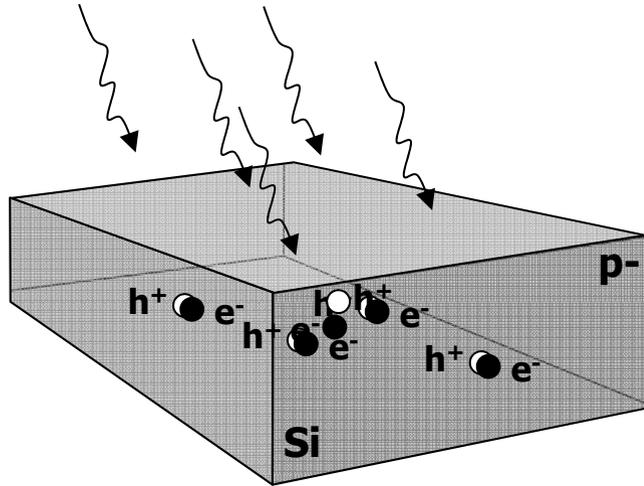
- **Bottlenecks at different stages**

Image coding

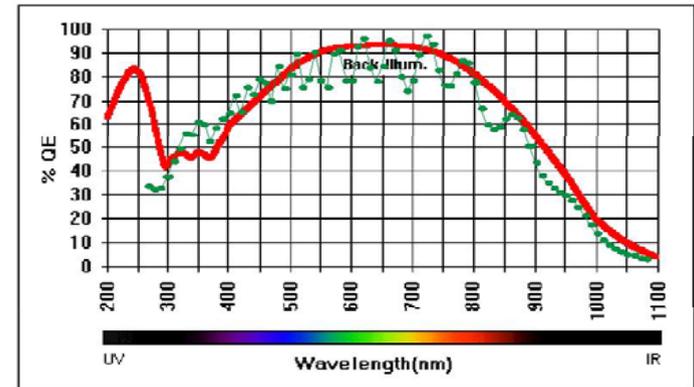
Image transmission

Image processing

Photo-transduction Mechanisms

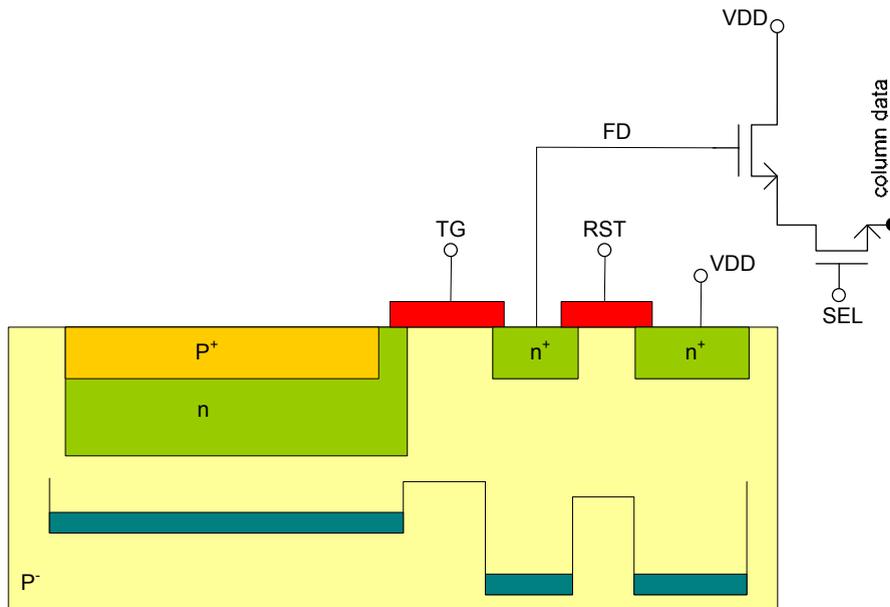


- Incident photons create electron-hole pairs
- The absorption length depends upon the wavelength
- An electrical field separate the hole-electron pairs



CCD: Hamamatsu S7030
 CMOS: Perkin Elmer RL0512T

Pinned Photodiodes

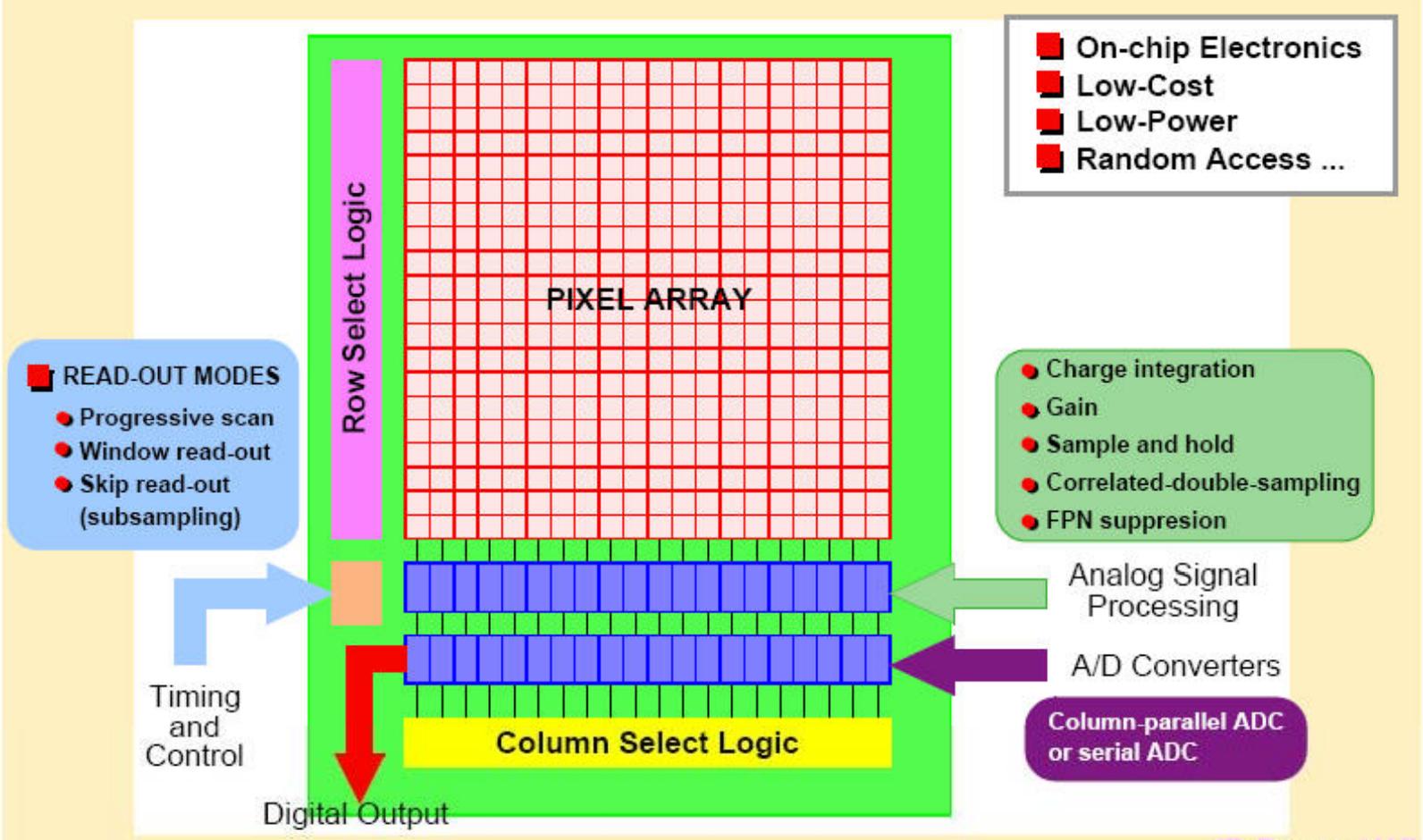


Features:

- $p^+ n p^-$ structure
- Low-dark current
- Operation similar to photogate
- Decrease reset noise (**Correlated Double Sampling**)
- Better response to blue

During integration phase, photo-generated majority carriers are stored in the depletion region. (Device is fully depleted)

Fossum's Camera on Chip



- On-chip Electronics
- Low-Cost
- Low-Power
- Random Access ...

- READ-OUT MODES
- Progressive scan
 - Window read-out
 - Skip read-out (subsampling)

- Charge integration
- Gain
- Sample and hold
- Correlated-double-sampling
- FPN suppression

Analog Signal Processing

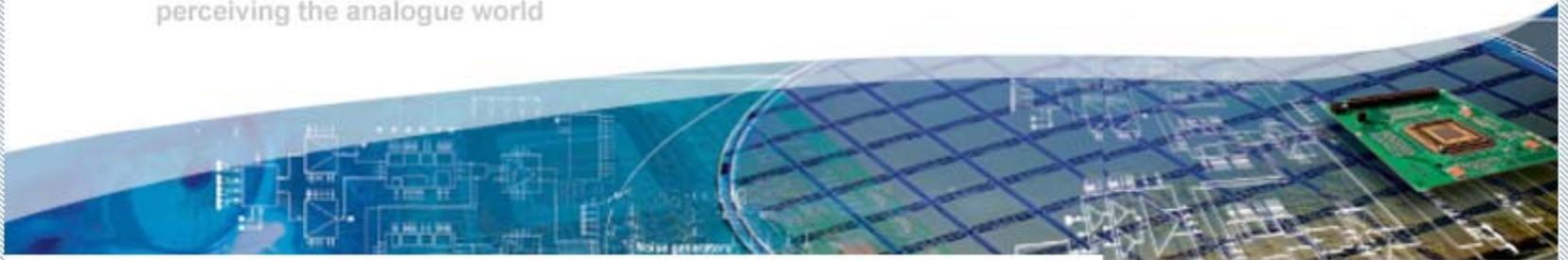
A/D Converters

Column-parallel ADC or serial ADC



ANAFOCUS

perceiving the analogue world



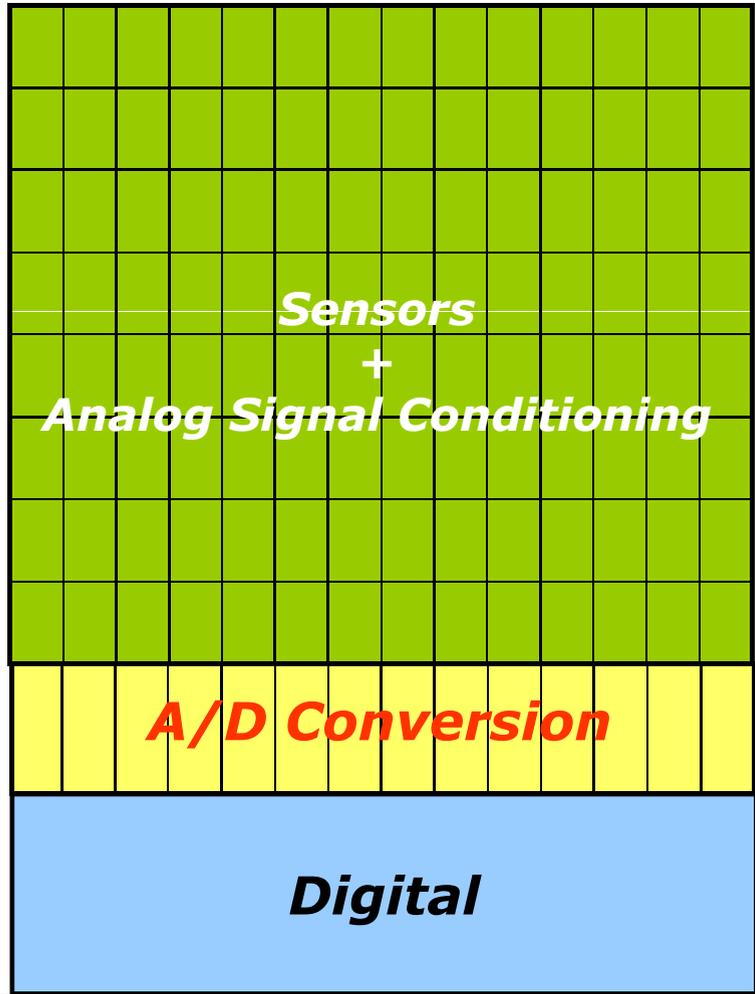
Architectures for Sensory-Processing

Progressive Distributed Processing

Bio-Inspired Vision Architectures

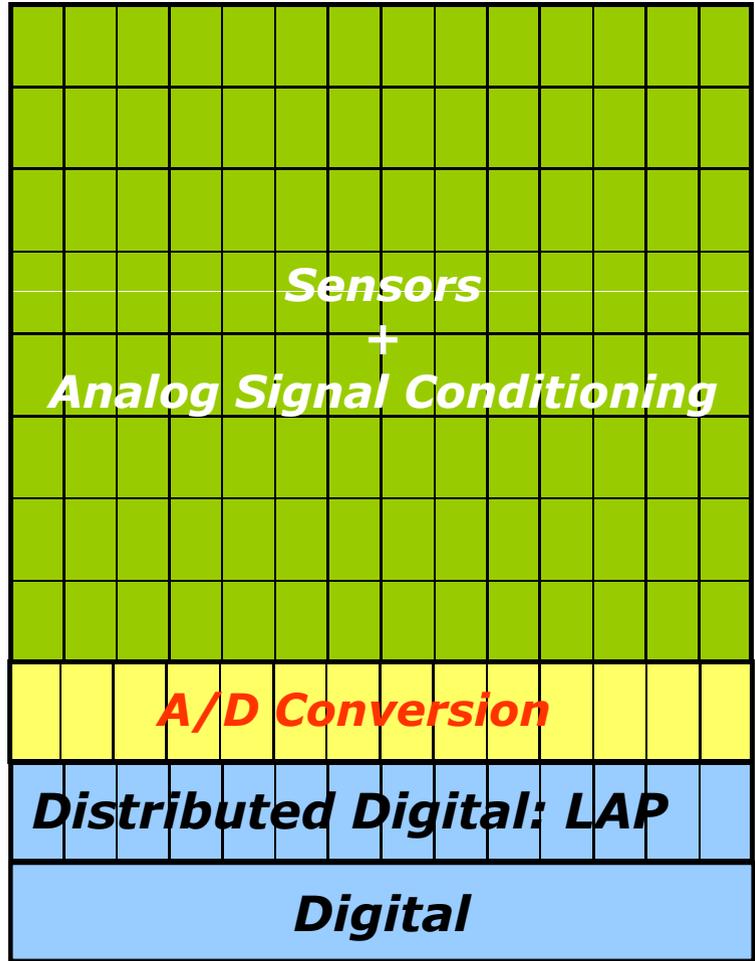
Shifting the Analog-to-Digital Border

Conventional Architecture



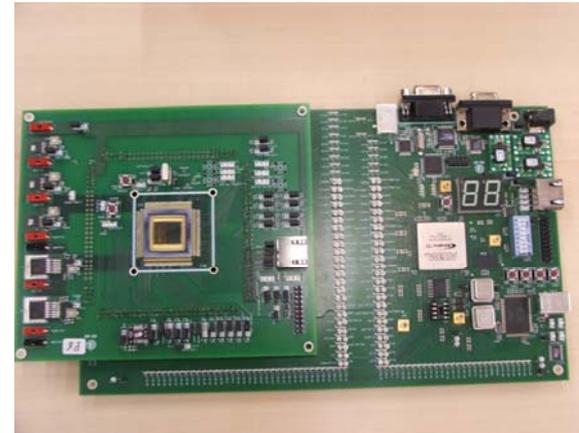
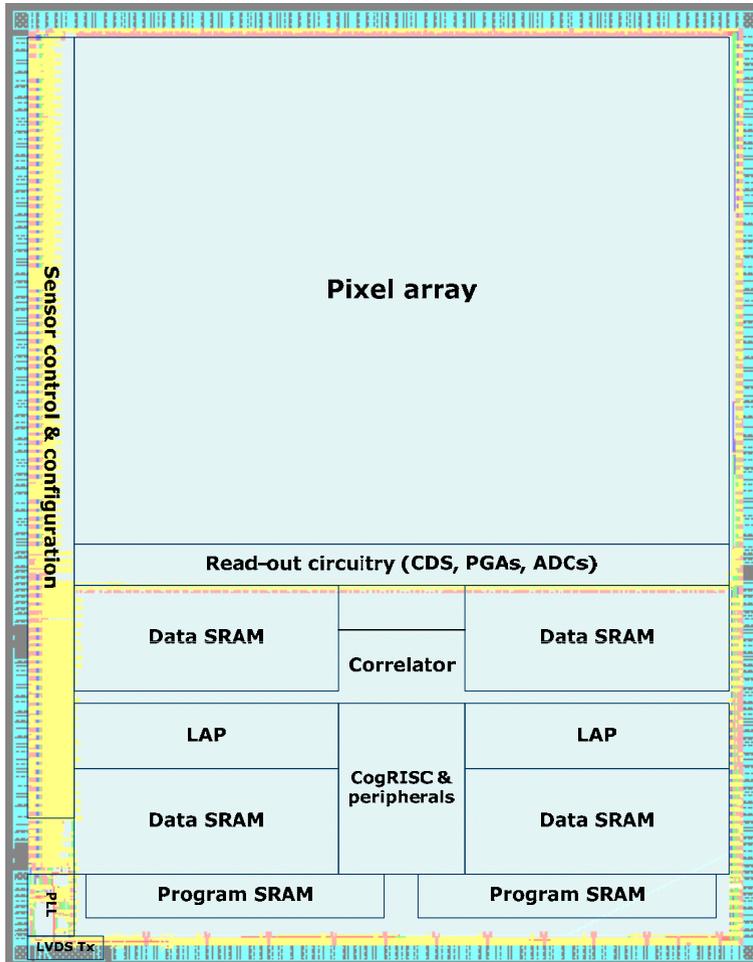
- ☺ **Large spatial resolution**
limited only by pixel SNR
- ☺ **Large circuit operation**
predictability and robustness
analog only at the ADCs
- ☺ **Small spurious signal**
interactions
- ☺ **Large flexibility and pro-**
grammability:
mostly digital circuits and codes
- ☹ **Small computational power**
and efficiency
- ☹ **Large memory requirements**
- ☹ **Data transfer bottlenecks**

Linear Array of Digital Processors



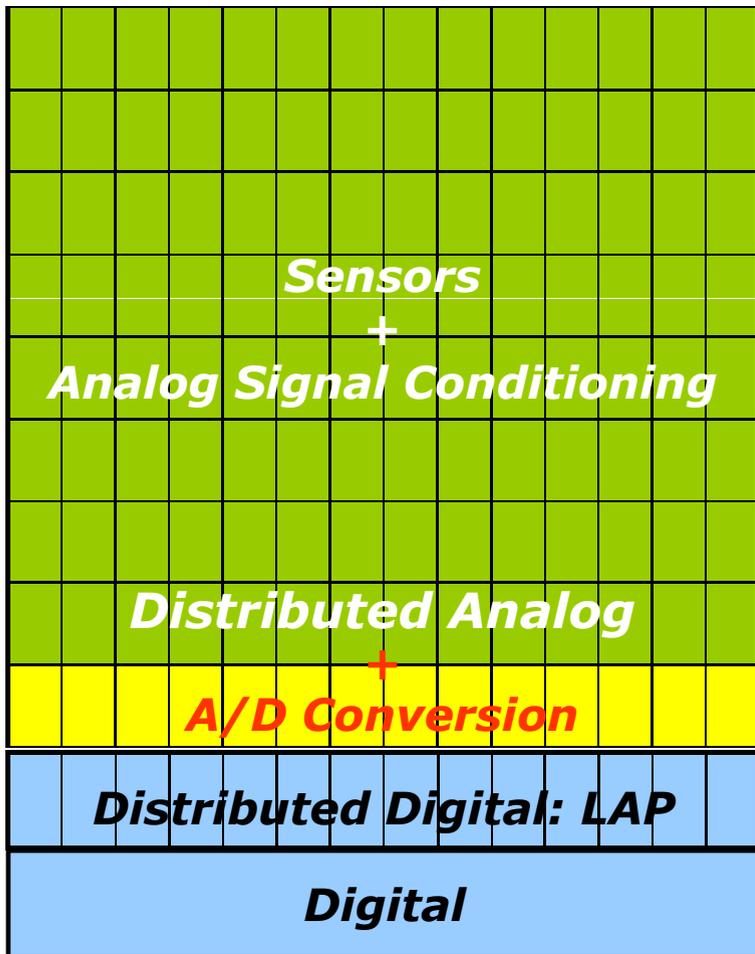
- ⚠ **Smaller spatial resolution:**
trade-off with processing
- 😊 **Large circuit operation**
predictability and robustness:
analog only at the ADCs
- 😊 **Small spurious signal**
interactions
- 😊 **Large flexibility and pro-**
grammability:
mostly digital circuits and codes
- ⬆ **Larger computational power**
and efficiency
- ⬆ **Smaller memory requirements**
- 😞 **Data transfer bottlenecks**

A True LAP VSoC for Machine Vision



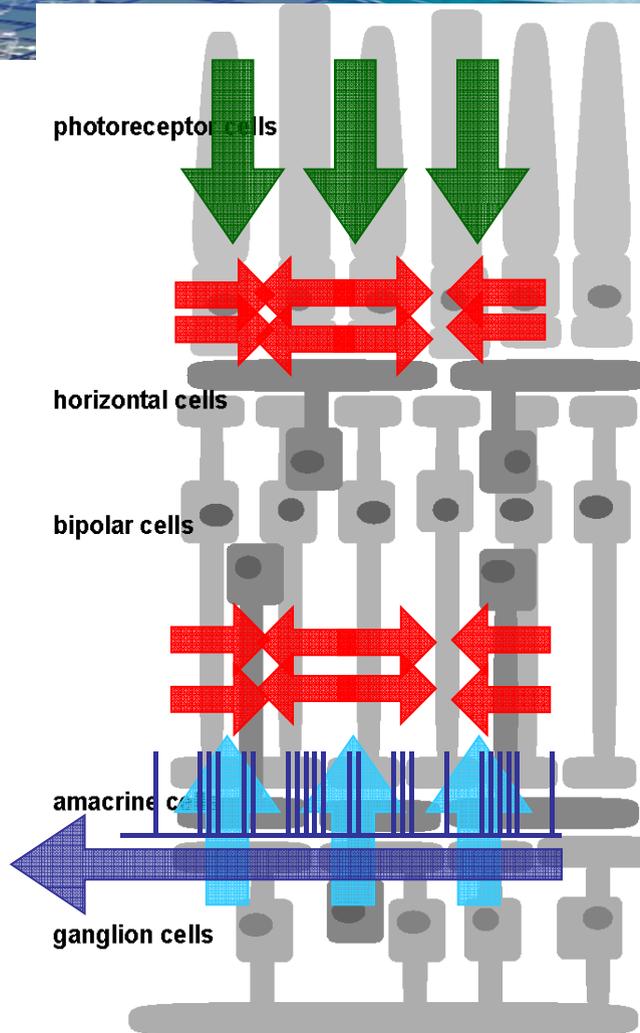
- Pinned photodiode pixel
- High-speed rolling and global electronic shutter with programmable exposure time.
- Programmable controls: gain, offset, frame rate and frame size (RoI)
- 100MHz clock frequency
- Per-column readout path permitting up to 500fps readout speeds
- Multiple I/Os and high-speed communications

Linear Array of Analog-Digital Processors



- ⚠ **Smaller spatial resolution:**
trade-off with processing
- ⚠ **Smaller circuit operation predictability and robustness:**
analog also for processing
- ⚠ **Larger spurious signal interactions**
- ☹ **Smaller flexibility and programmability:**
both analog and digital circuits
- ⬆ **Larger computational power and efficiency**
- ⬆ **Smaller memory requirements**
- ⬆ **Smaller transfer bottlenecks**

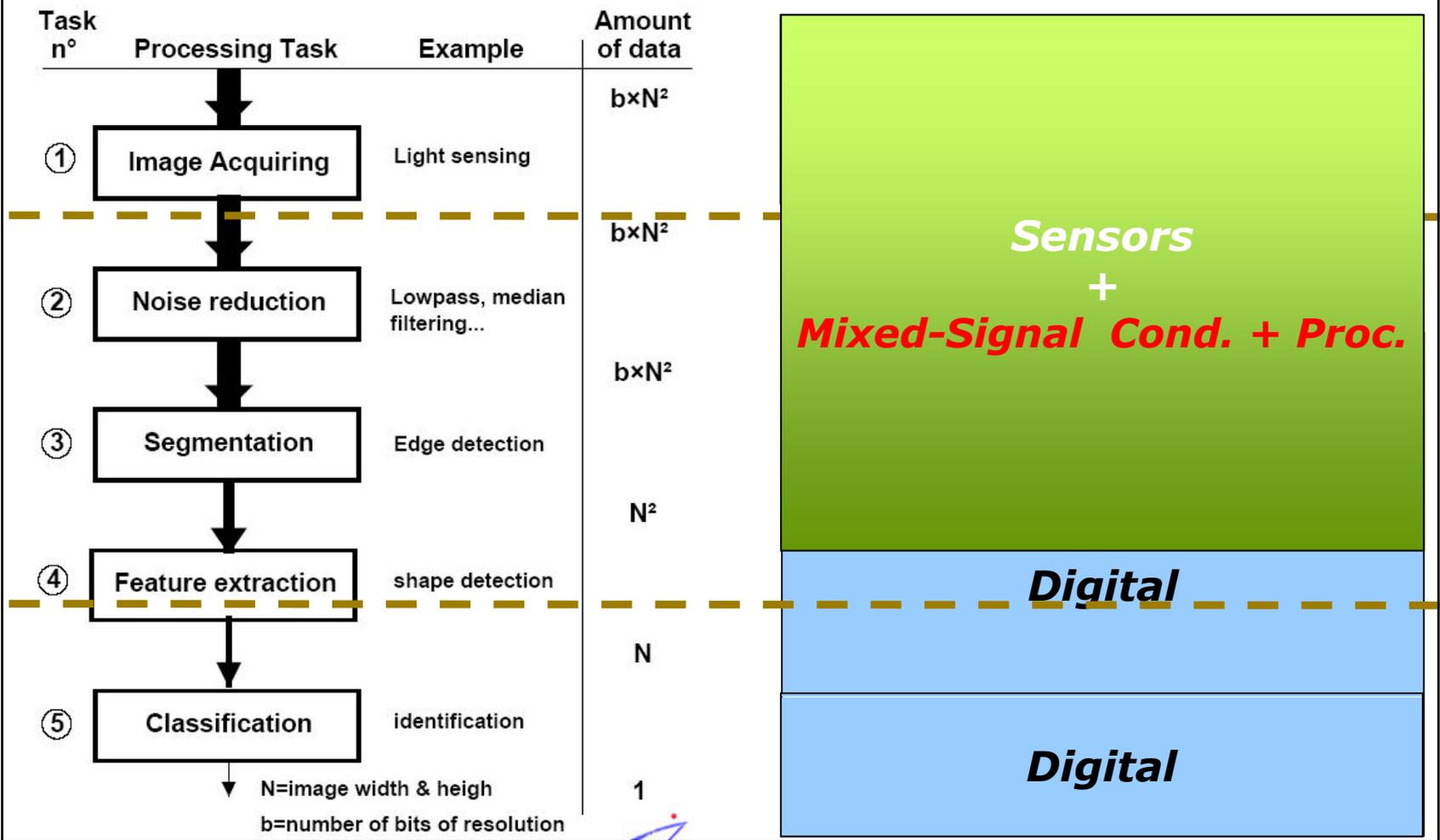
How Does The Retina Work ?²¹

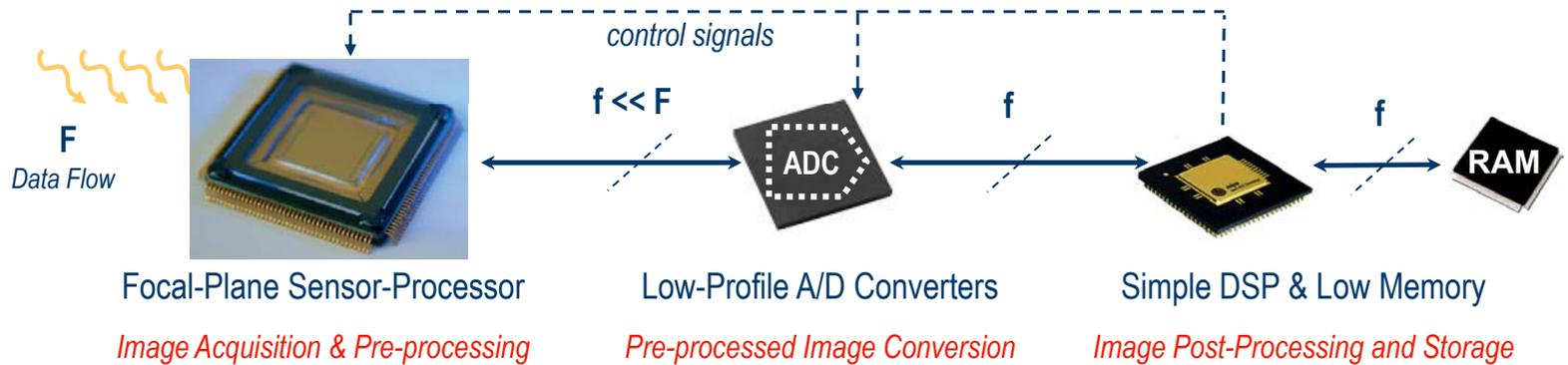
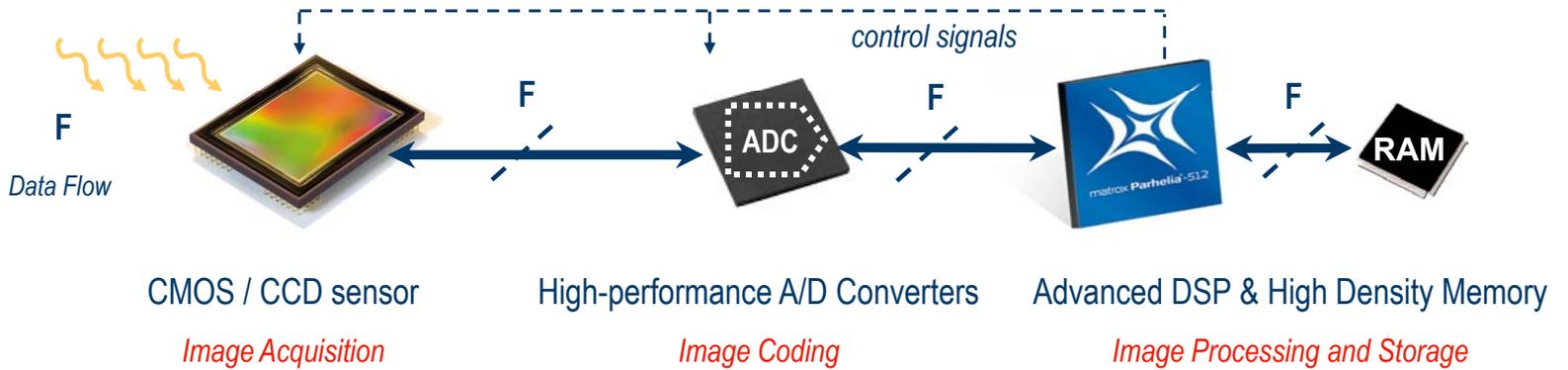


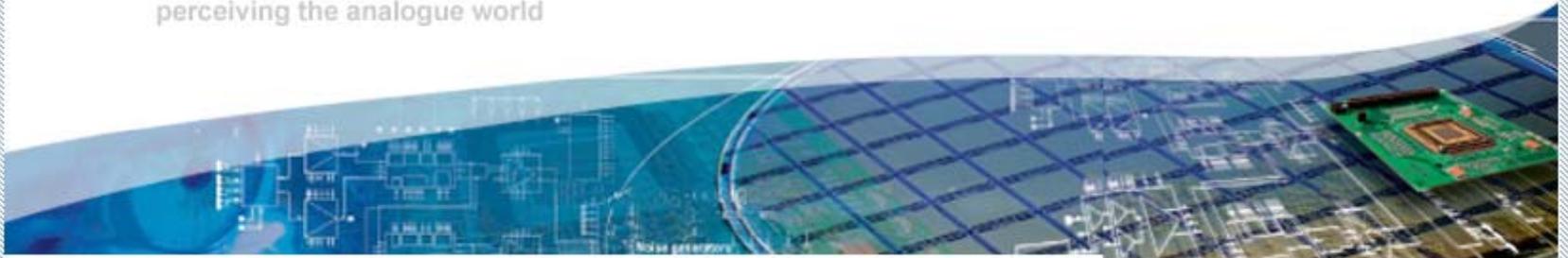
- **Sensors and processors are merged**
- **Processing and sensing are simultaneous**
- **Significant data compression is achieved**



[Roska & Werblin 2001]





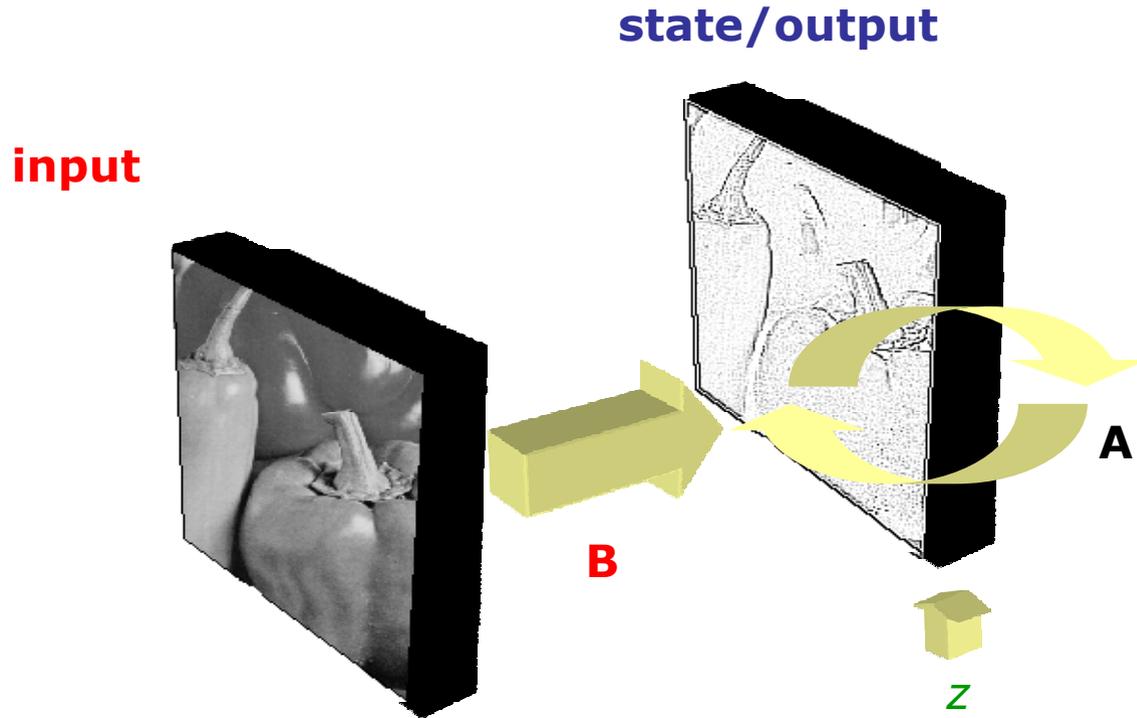


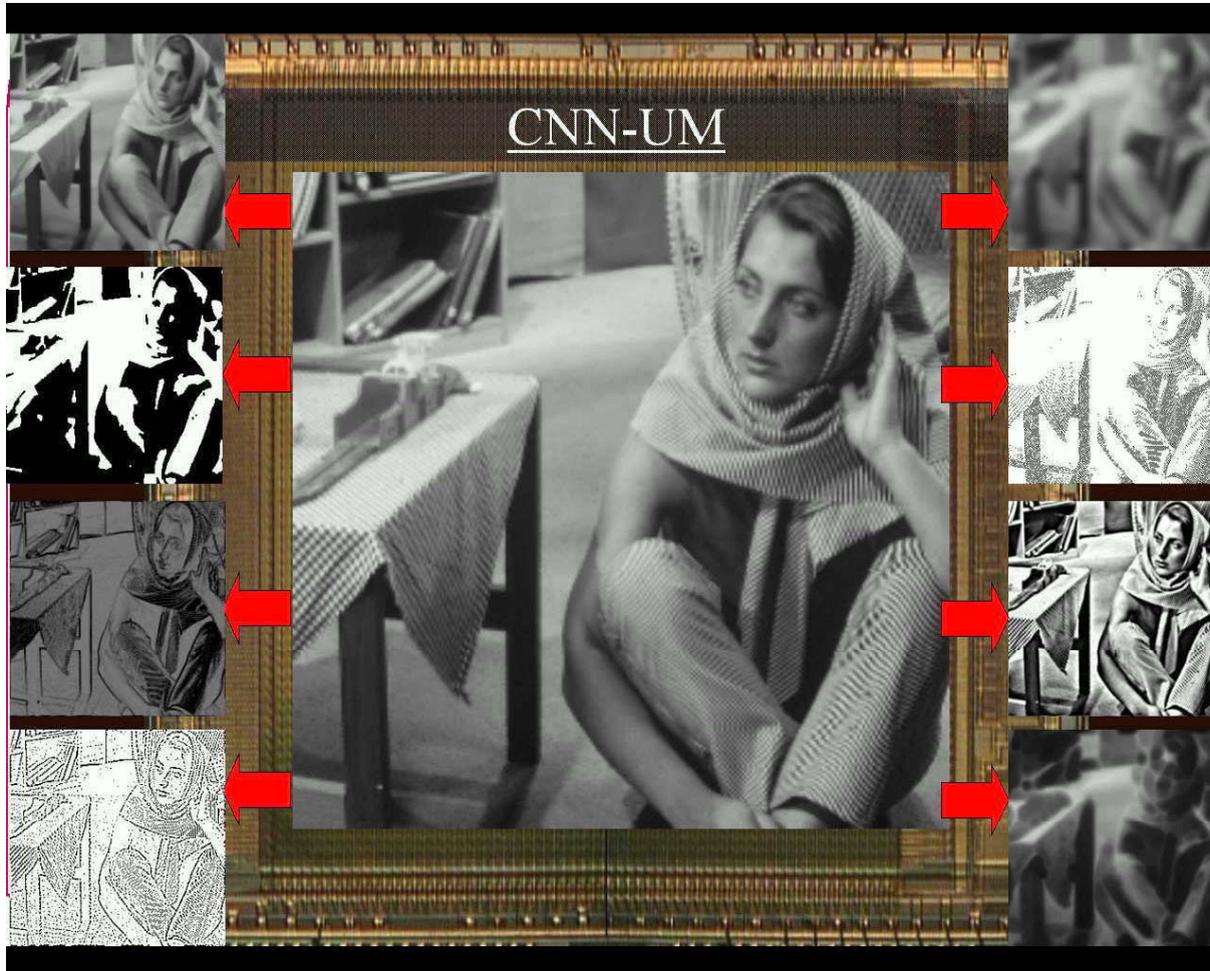
Topographic Sensor-Processors

Concept of Concurrent Sensory-Processing

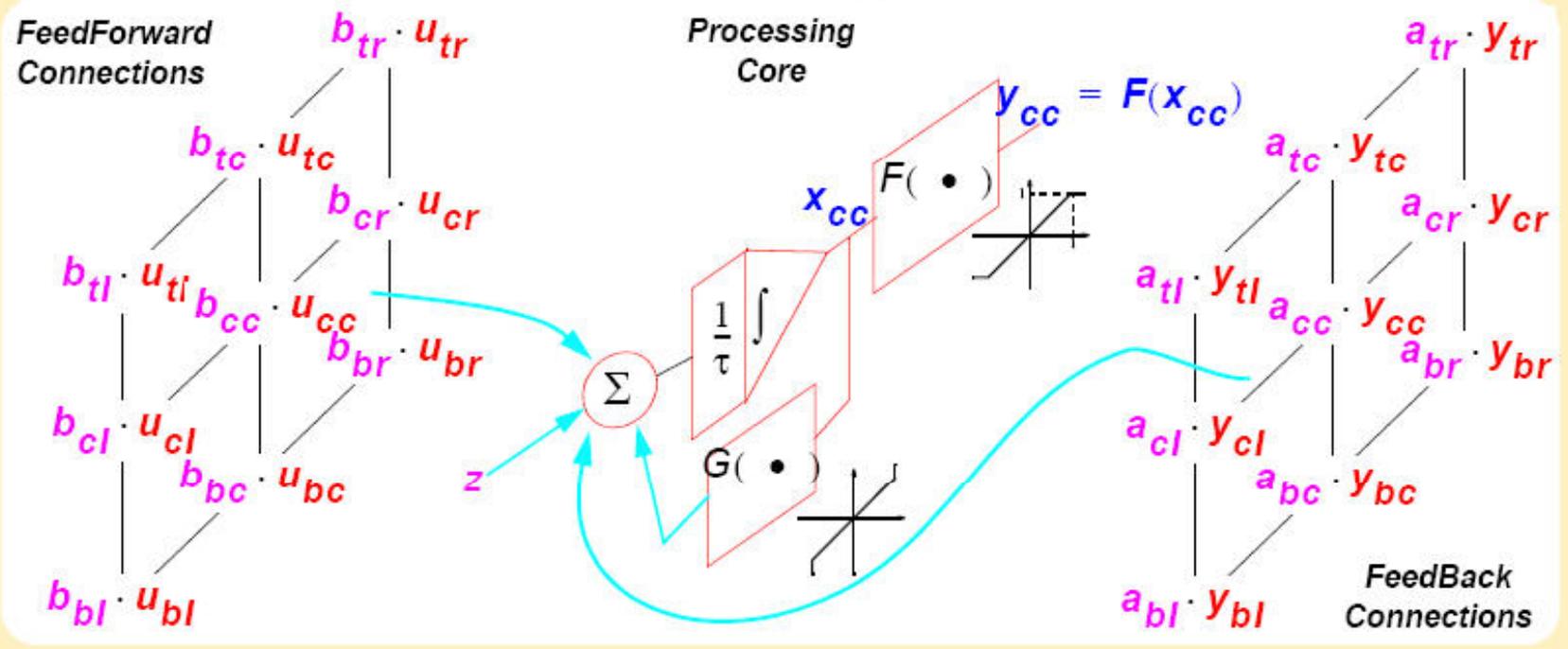
Multi-Functional Pixels

Rationale for Analog Pre-Processing

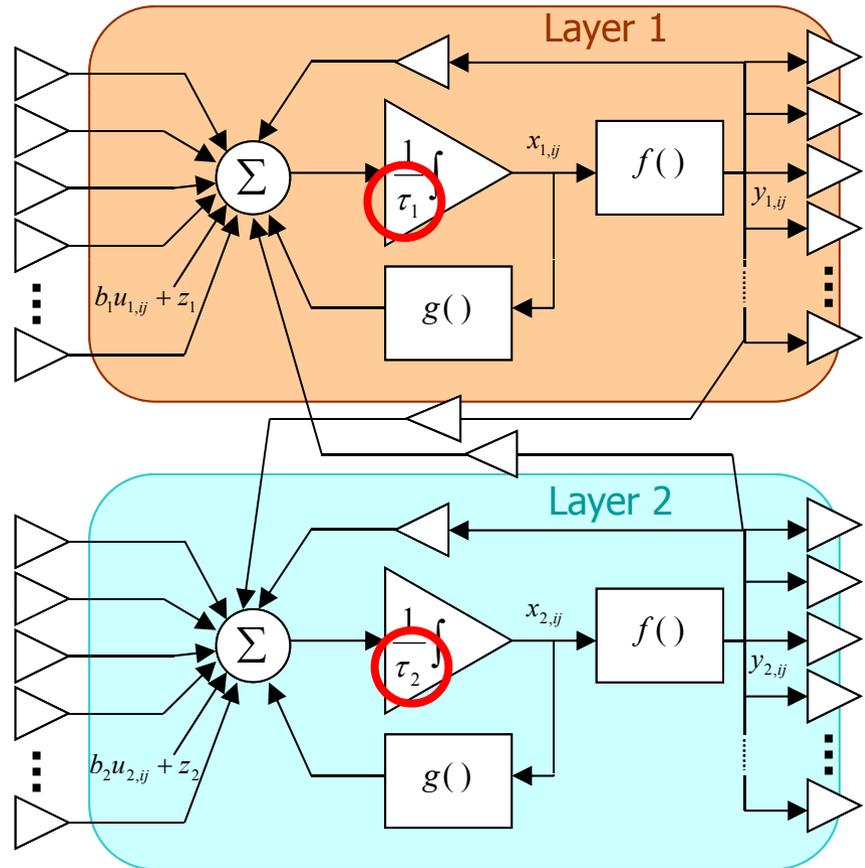
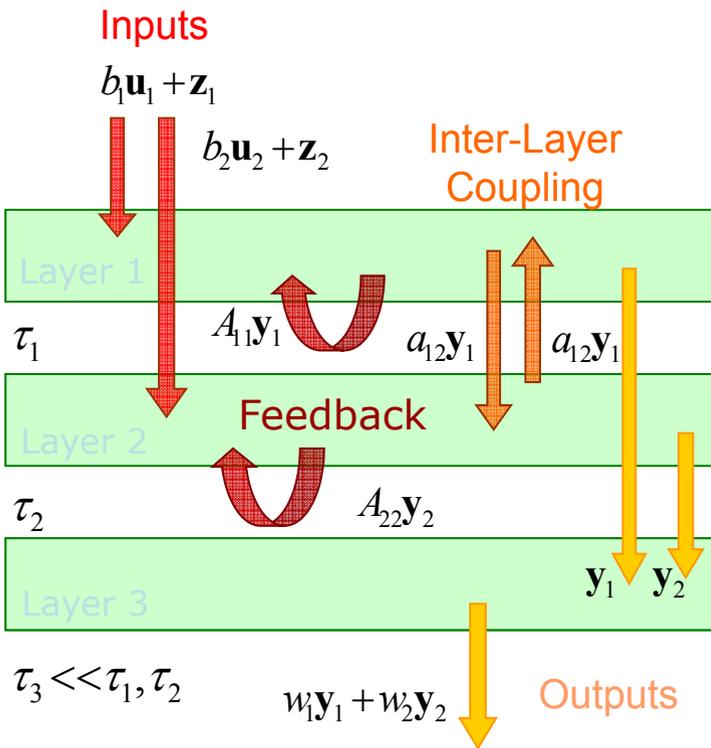




CNN Single Layer Interactions



CNN Multi-Layer Interactions



Multi-Layer Interactions

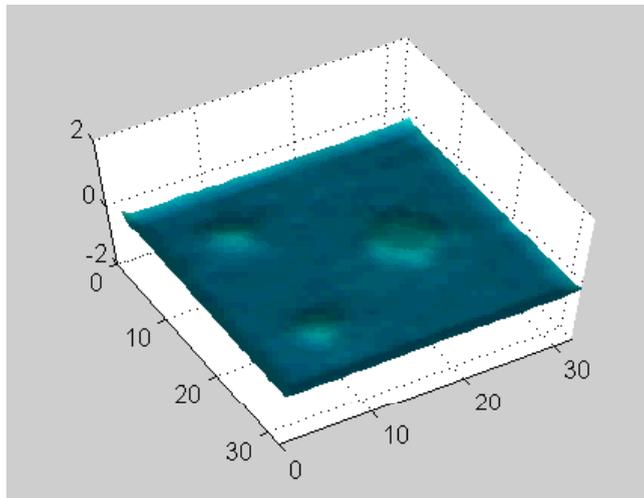
$$\frac{d}{dt} \phi_i(x, y, t) - c_i \nabla^2 \phi_i(x, y, t) = \alpha_i \phi_i(x, y, t) + \beta_i \phi_i(x, y, t_0) + \gamma_{ij} \phi_j(x, y, t)$$

diffusion

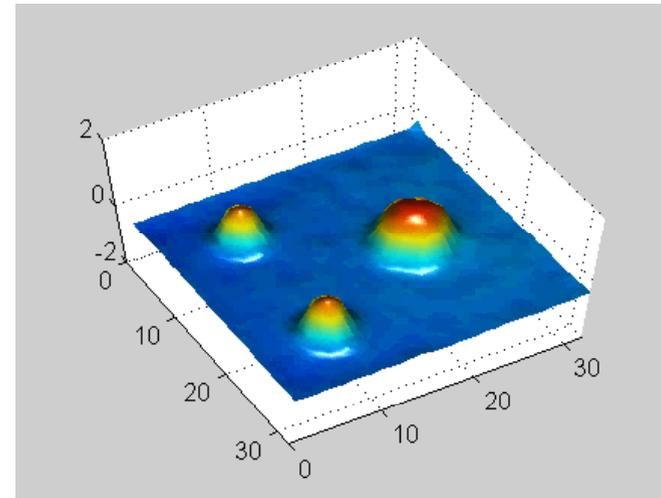
reaction

Wave propagation in active media

[Perona & Malik 1990]



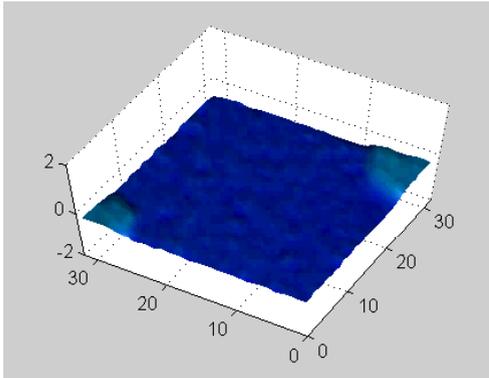
Layer 1 (slow)



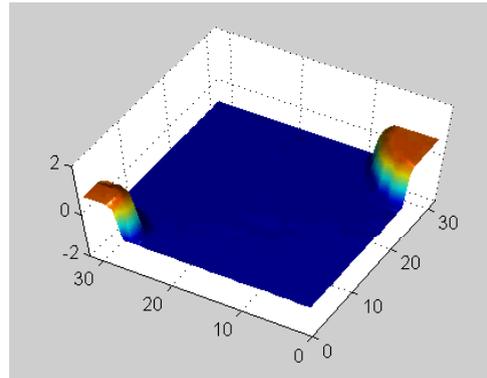
Layer 2 (fast)

Multi-Layer Interactions

Traveling Waves



Layer 1 (slow)



Layer 2 (fast)

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1.5 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad \mathbf{A}_2 = \begin{bmatrix} 0.6 & 0.8 & 0.6 \\ 0.8 & -0.6 & 0.8 \\ 0.6 & 0.8 & 0.6 \end{bmatrix}$$

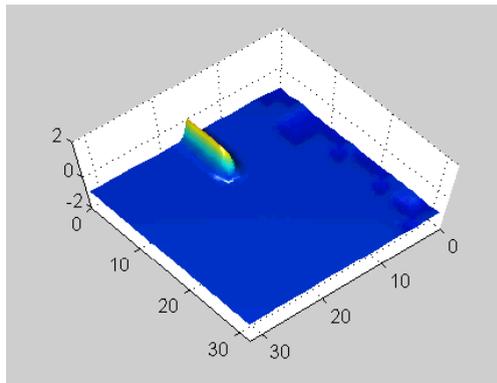
$$a_{21} = 4.5 \quad a_{12} = -4.2$$

$$b_1 = 0 \quad b_2 = 5$$

$$z_1 = 0.6 \quad z_2 = -2.7$$

$$\tau_1 : \tau_2 = 16 : 1$$

Spiral Wave



Only fast layer

$$\mathbf{A}_1 = \begin{bmatrix} 0.6 & 0.8 & 0.6 \\ 0.8 & -0.3 & 0.8 \\ 0.6 & 0.8 & 0.6 \end{bmatrix} \quad \mathbf{A}_2 = \begin{bmatrix} 0.6 & 0.8 & 0.6 \\ 0.8 & -0.9 & 0.8 \\ 0.6 & 0.8 & 0.6 \end{bmatrix}$$

$$a_{21} = 6 \quad a_{12} = -3$$

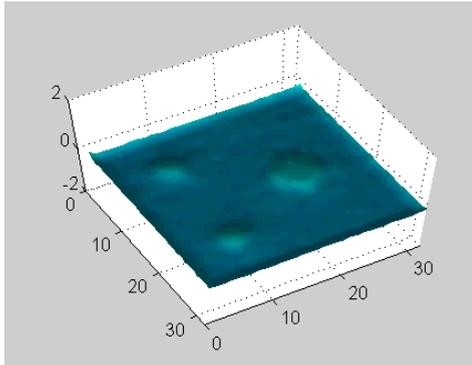
$$b_1 = 0.0 \quad b_2 = 3.9$$

$$z_1 = -4.5 \quad z_2 = -3.6$$

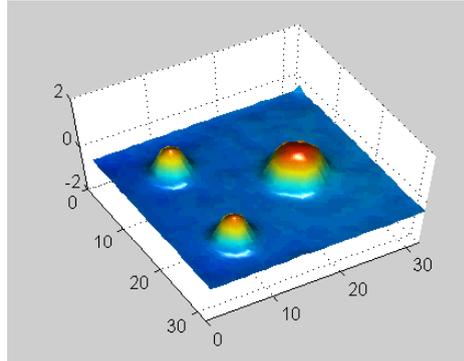
$$\tau_1 : \tau_2 = 8 : 1$$

Multi-Layer Interactions

Long Distance Inhibition in the IPL



Layer 1 (slow)



Layer 2 (fast)

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1.5 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad \mathbf{A}_2 = \begin{bmatrix} 0.6 & 0.8 & 0.6 \\ 0.8 & -0.6 & 0.8 \\ 0.6 & 0.8 & 0.6 \end{bmatrix}$$

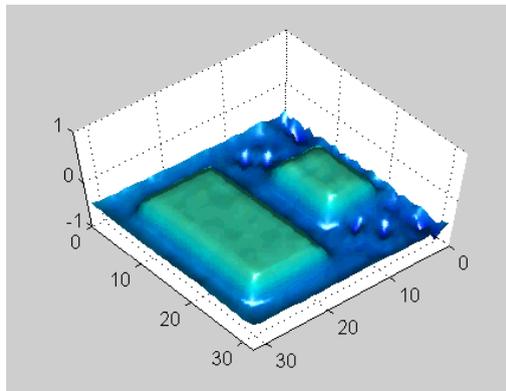
$$a_{21} = 3.3 \quad a_{12} = -5.4$$

$$b_1 = 0 \quad b_2 = 3$$

$$z_1 = 0.6 \quad z_2 = -2.7$$

$$\tau_1 : \tau_2 = 16 : 1$$

Spatial-Temporal Detection of Edges in the OPL



Only fast layer

$$\mathbf{A}_1 = \begin{bmatrix} 0.5 & 0.8 & 0.5 \\ 0.8 & -3.6 & 0.8 \\ 0.5 & 0.8 & 0.5 \end{bmatrix} \quad \mathbf{A}_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

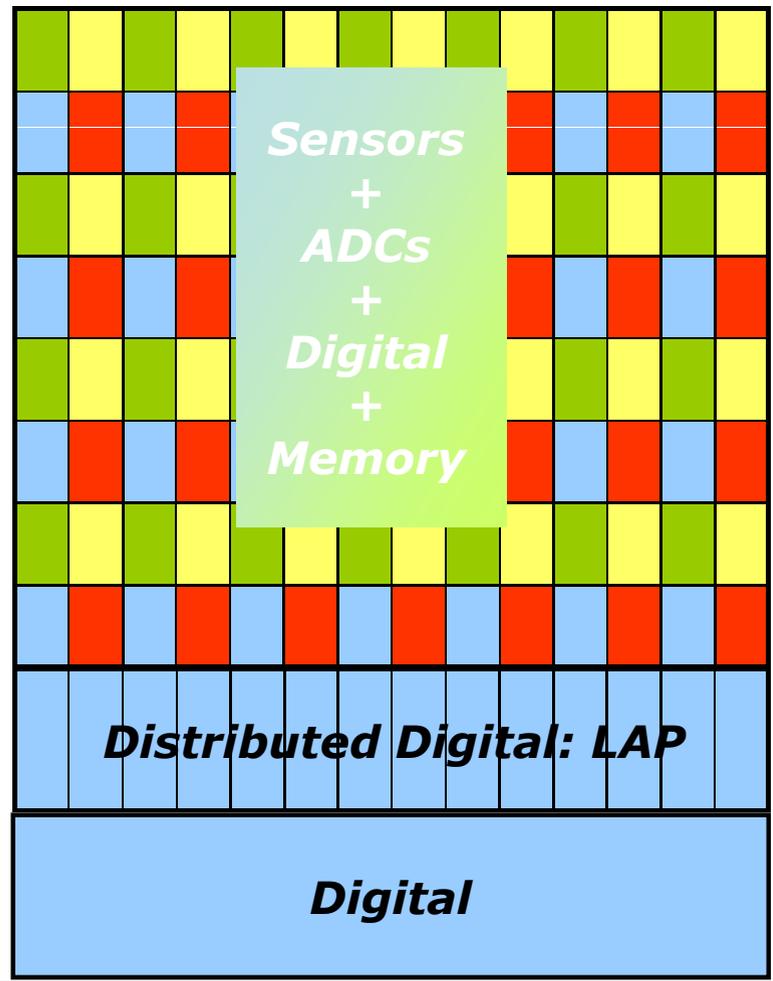
$$a_{21} = 4.5 \quad a_{12} = -5.7$$

$$b_1 = 0.0 \quad b_2 = 3.0$$

$$z_1 = -9 \quad z_2 = 0$$

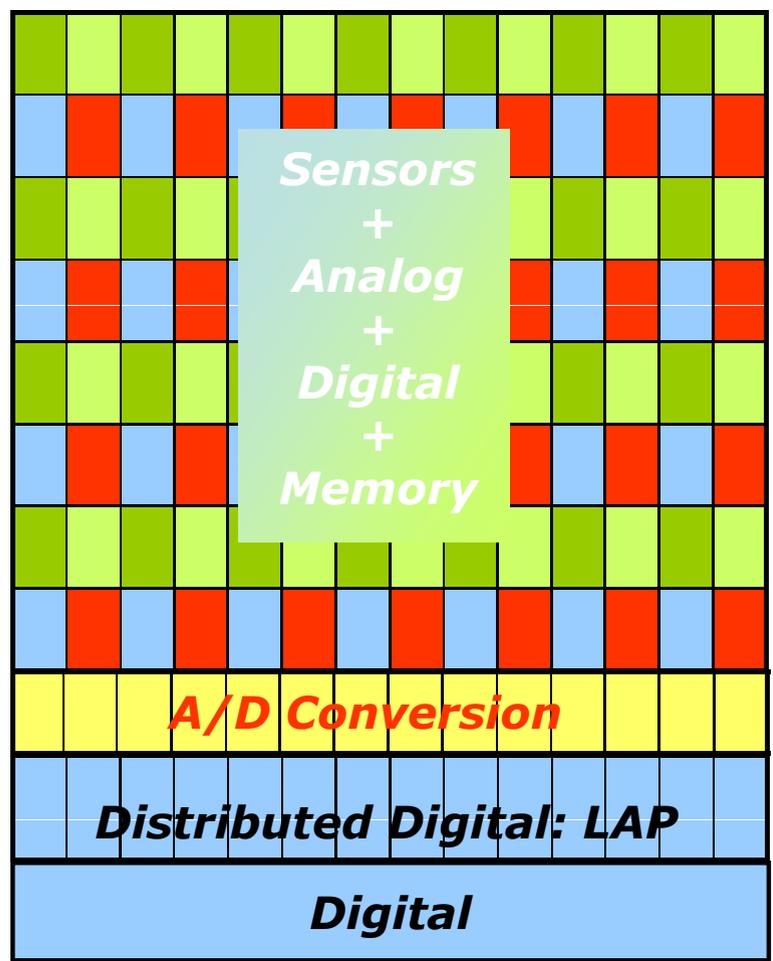
$$\tau_1 : \tau_2 = 6 : 1$$

Topographic ADCs + Digital Processing



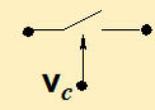
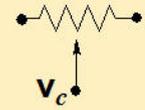
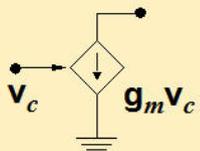
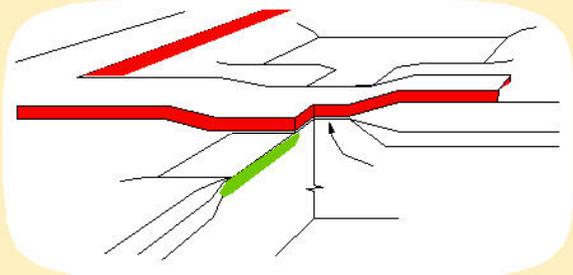
- ⬇️ **Smaller spatial resolution:**
trade-off with processing
- ⬆️ **Larger system operation predictability and robustness:**
analog only for ADCs
- ⬆️ **Smaller spurious signal interactions**
- ⬆️ **Larger flexibility and programmability:**
only digital processing
- ⬇️ **Functionality and efficiency compromised by ADC accuracy**
- ⬇️ **Large in-pixel memory requirements**

Topographic Mixed-Signal Processing



- ☹️ **Low spatial resolution:**
trade-off with processing
- ☹️ **Involved circuit operation**
predictability and robustness:
analog also for processing
- ☹️ **Larger spurious signal**
interactions
- ☹️ **Involved flexibility and**
programmability:
both analog and digital
circuits
- 😊 **Large computational power**
and efficiency
- 😊 **Small memory requirements**
- 😊 **Small transfer bottlenecks**

The Functional Power of MOST



✓ Ohmic SI-Region:

$$I_D \approx k \frac{W}{L} V_{DS} (V_G - V_{T0} - nV_S)$$

V_{DS} must be kept constant

✓ Saturation SI-Region:

$$I_D \approx \frac{k}{2n} \frac{W}{L} (V_G - V_{T0} - nV_S)^2$$

Early Voltage is low $\propto L$

✓ Saturation WI-Region:

$$I_D \approx k \frac{W}{L} e^{\frac{(V_G - V_{T0} - nV_S)}{nU_T}}$$

✓ Ohmic Region:

$$R_{\Omega} = \frac{L}{kW(V_G - V_{T0} - nV_S)}$$

V_{DS} must be kept small

Large Distortion

$$R_{\square} \sim 10K\Omega$$

✓ Ohmic Region:

$$R_{\Omega} \propto \frac{L}{kW}$$

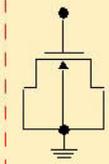
No offset

Cut-off Region:

$$R_{ON} \rightarrow \infty$$

Charge Injection

✓ Strong-Inv. Region:



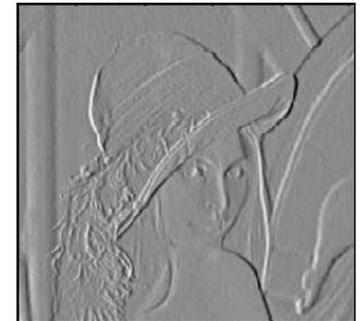
$$C \propto C_{OX} LW$$

$$C_{OX} \sim 3fF/mm^2$$

Nonlinear

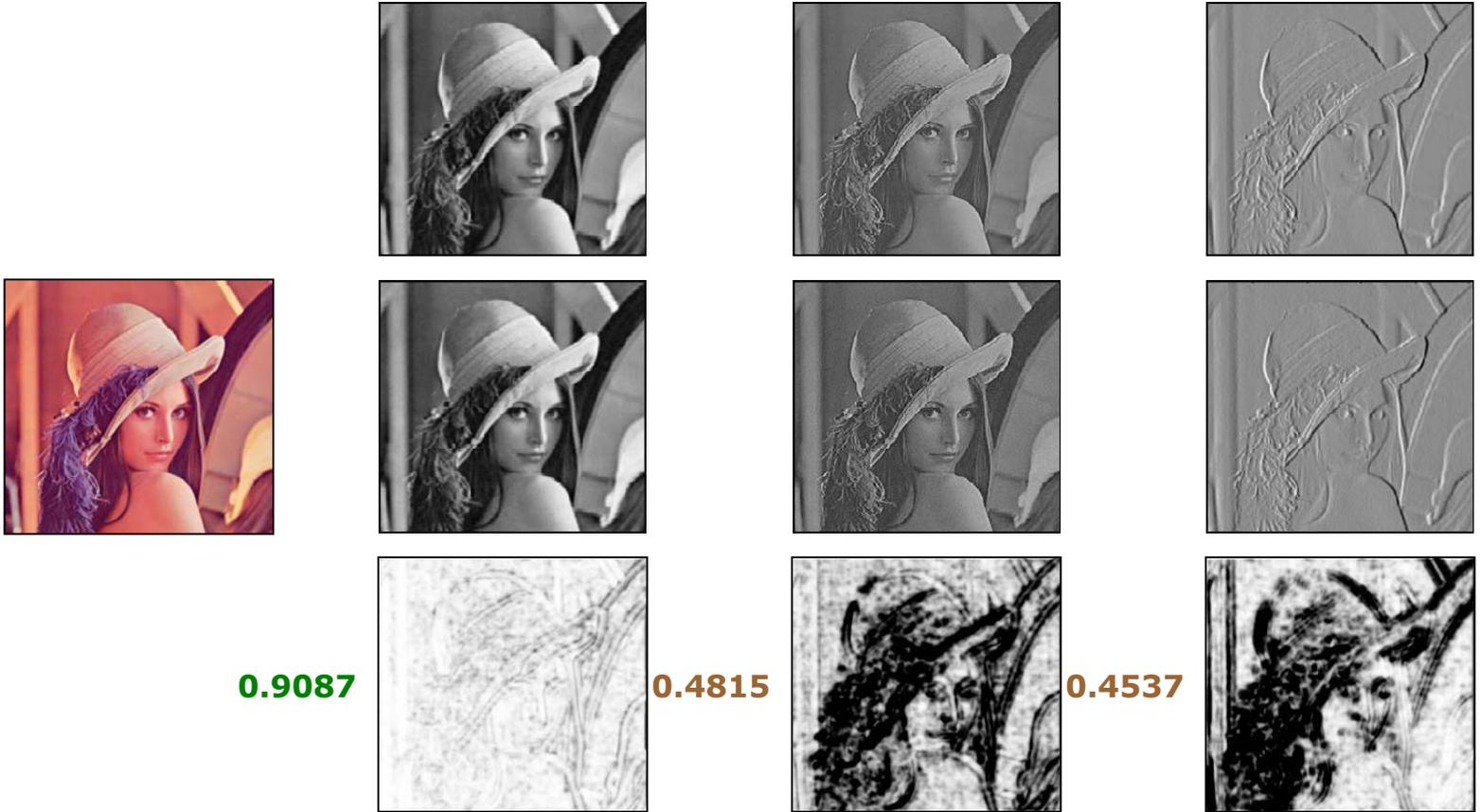
✓ Long-Term Memory with Floating Gate

Visual Inspection of Results

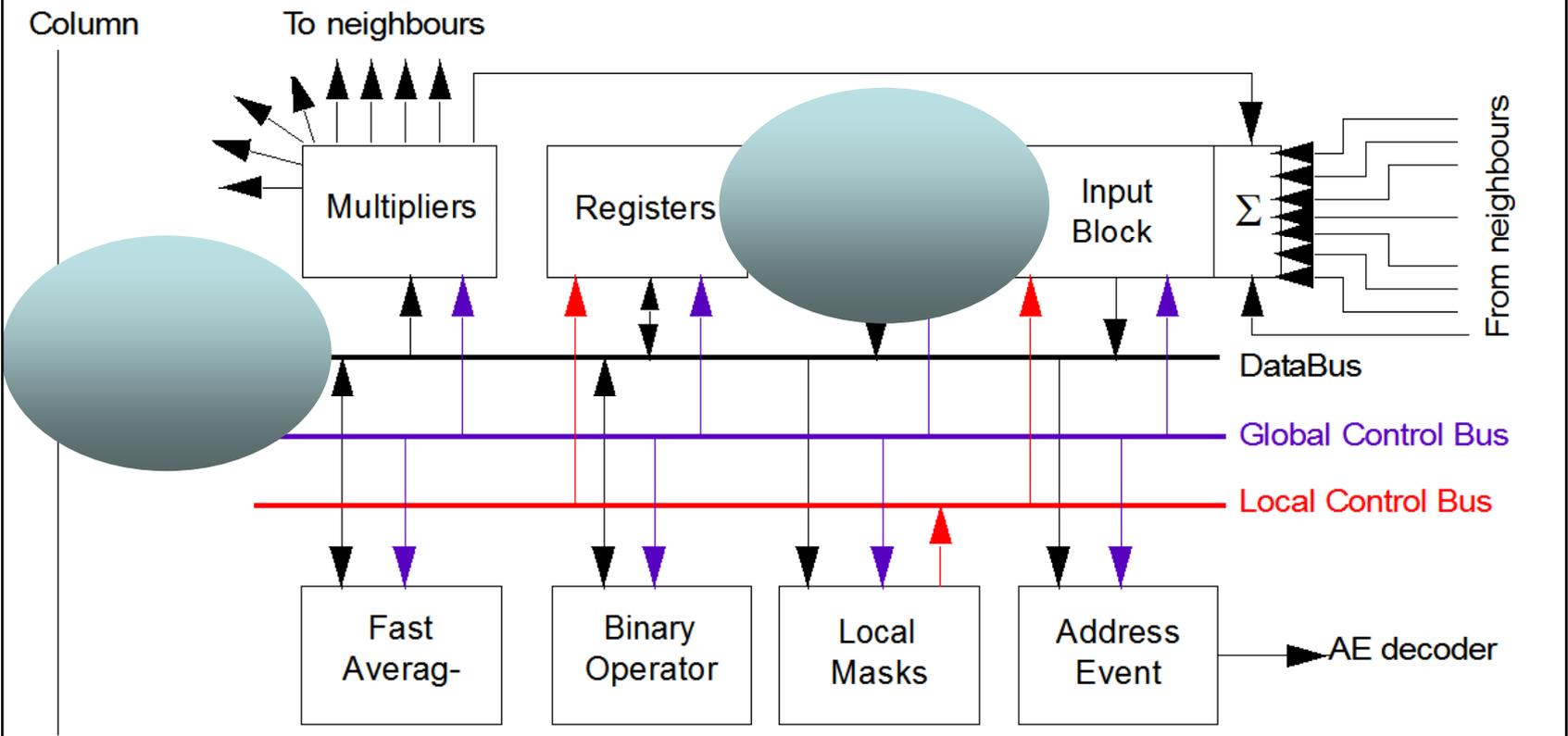


Adding random noise ($\sigma=4\text{LSB}$) to: 1) Input Image; 2) Coefficients of the convolution kernel at every position; 3) Output Image

Using Quality Assessment Methods



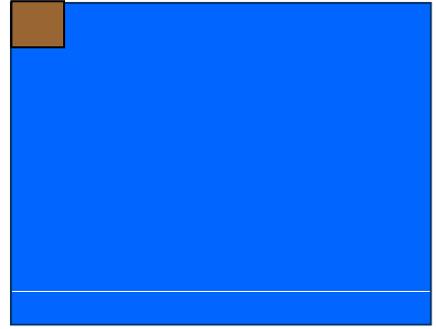
Z. Wang, et al., "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, Apr. 2004.





Size of Sensory Pixel

Size of Sensory-Processing Pixel



How Important is Resolution ?





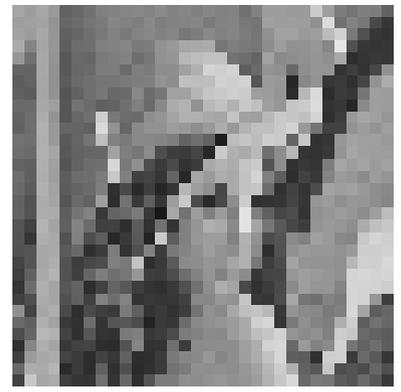
512 x 512



128 x 128



64 x 64



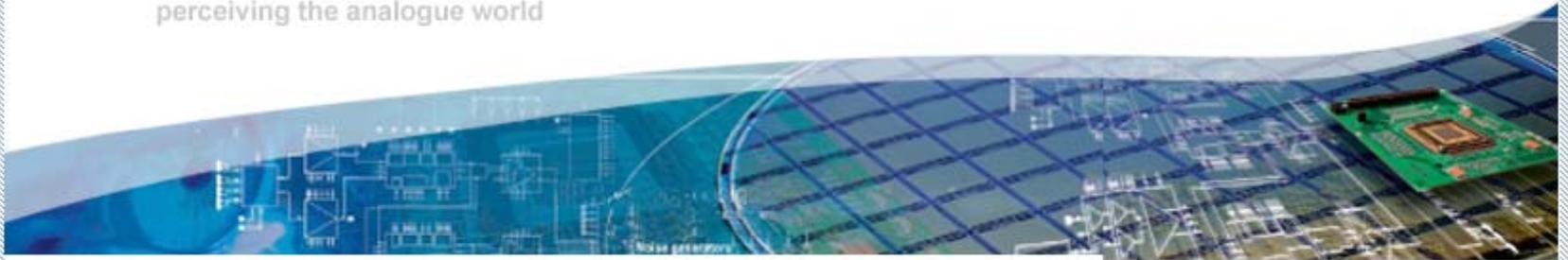
32 x 32

- **Vision is possible with 25 x 25 "pixels" (within limited field of view)**
- **Text can be read at 200 words/min (300 words/min with normal vision)**
- **Students can navigate in complex environments (maze) with confidence**

Resolution can be increased through proper design

- **VGA and up to 1.3Mpixels resolution for Surveillance**
- **> 1Mpixels for Machine Vision and Automotive**

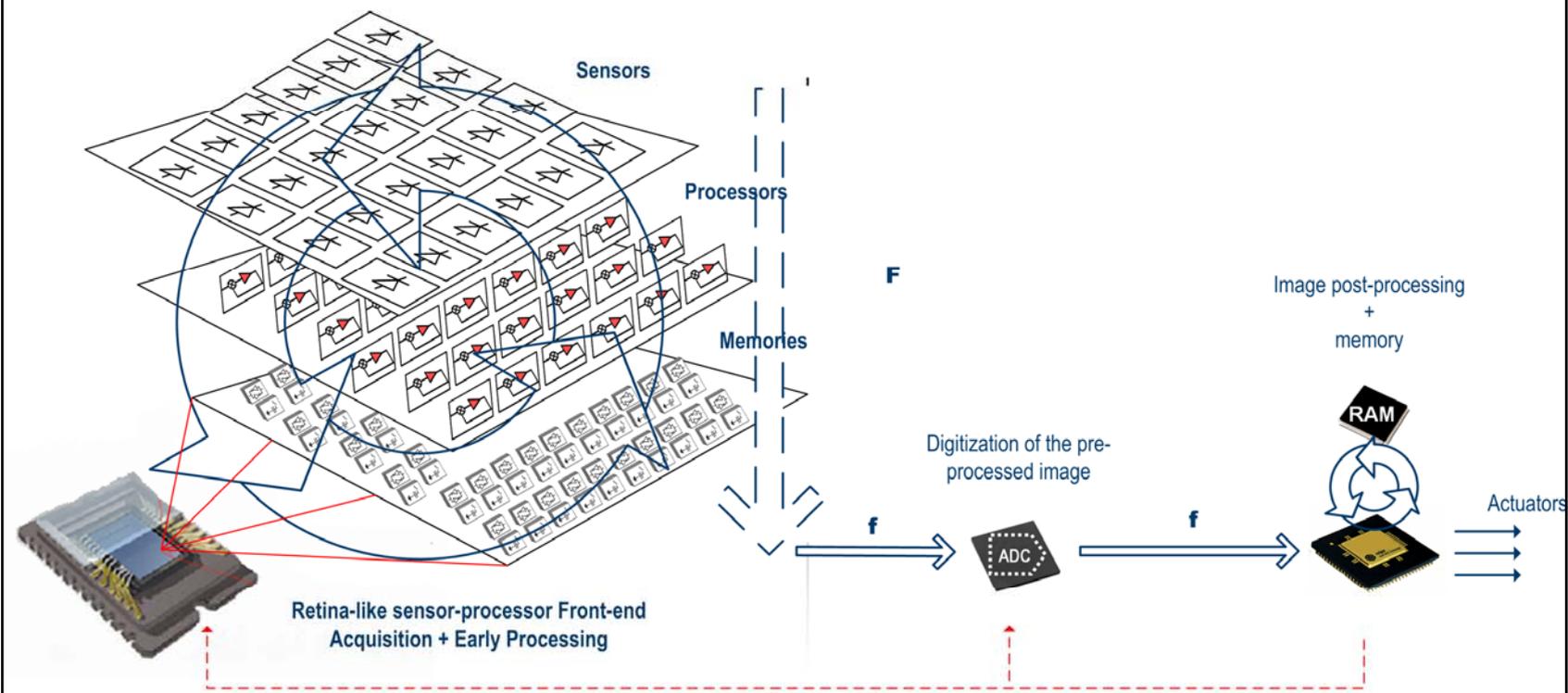




The Eye-RIS system

Architecture

The System in Operation

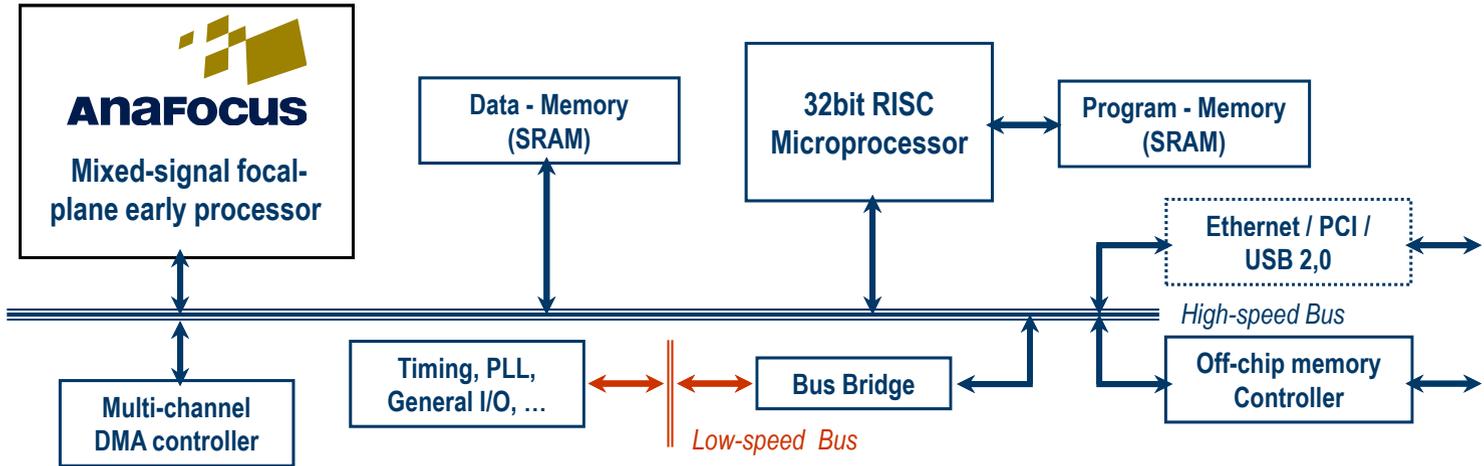


F – Input data flow @ bits per second (bps)

f – Data flow after early processing

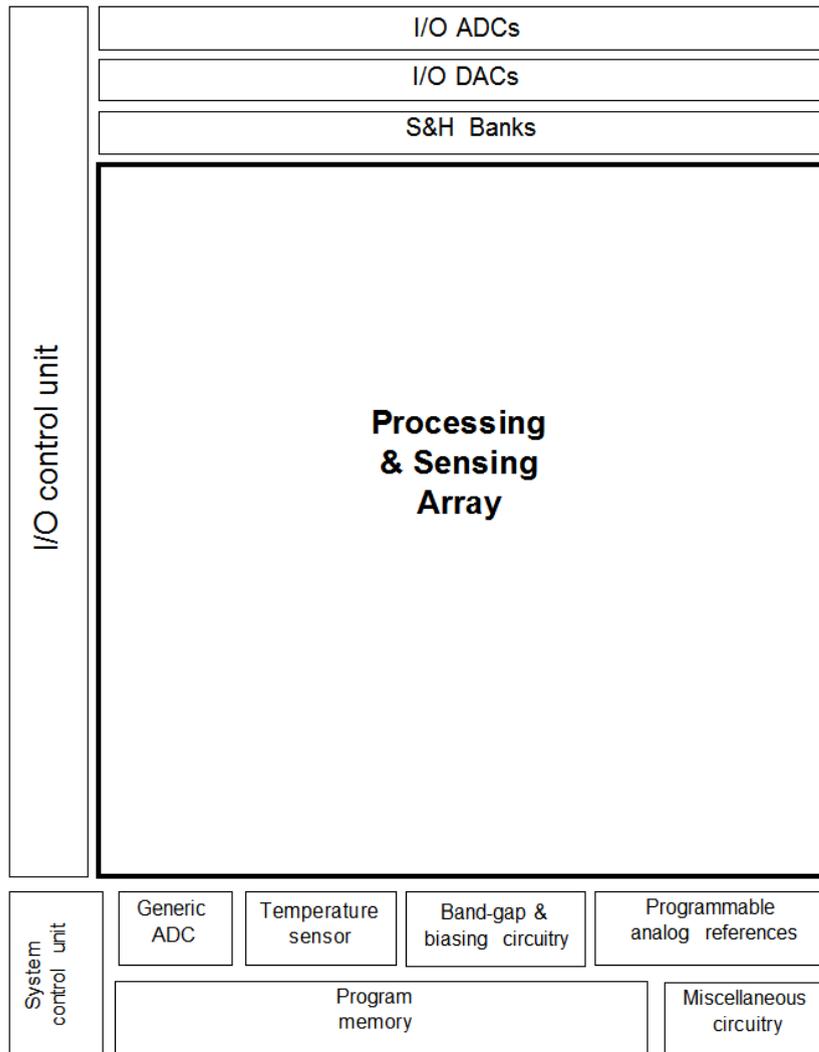
$$f \ll F$$

Configurable system which provides an efficient solution for low-to-medium resolution and high frame rates applications

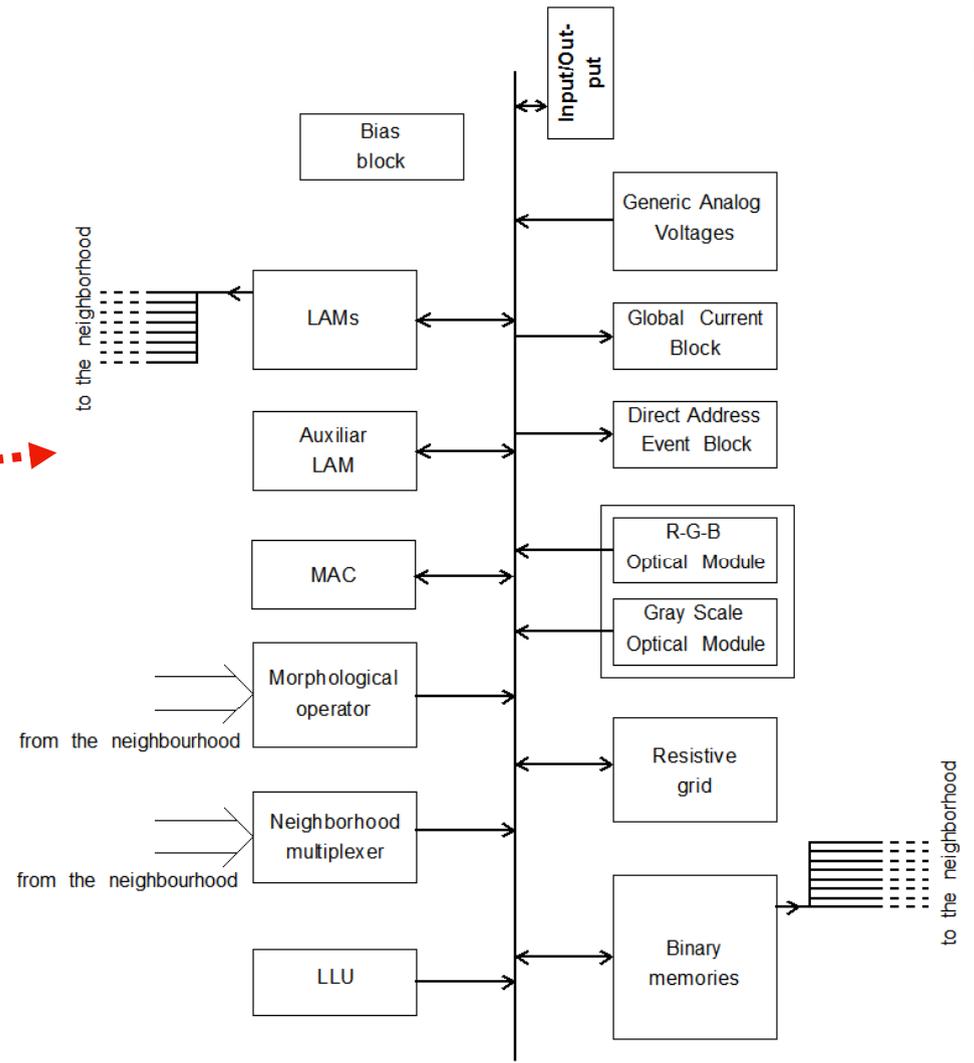
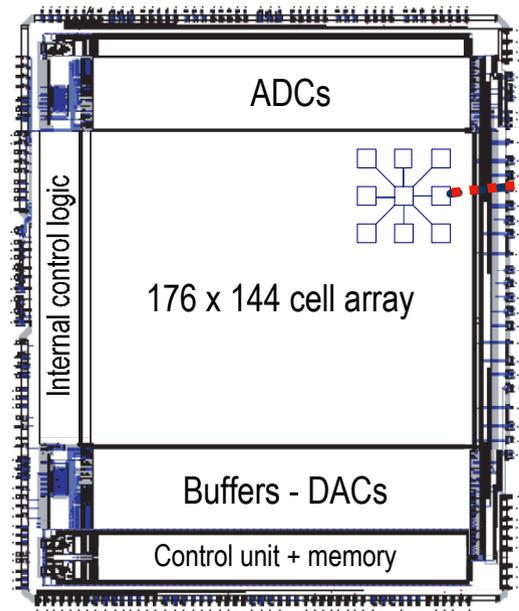


Ultra-high frame rate: up to 10,000fps@QCIF image resolution

Ultra-low power consumption: < 10mW@30fps@QCIF image resolution



- ✓ UMC 0.18 μ m CMOS 1P5M (1.8V/3.3V) - mixed-signal.
- ✓ High-performance Smart Image Sensor
 - 176 x 144 grey-scale pixels with 29.1 μ m pitch
 - High-speed non-rolling electronic shutter. Programmable exposure time (controlling step-down to 20ns)
 - Approximated sensitivity of 3V/lux-sec at 550nm
 - 4 + 1 (two banks) high-retention analog and 4 binary memories
 - Analog multiplexer for image shifting and analog MAC unit
 - Programmable, 3 x 3 neighbourhood pattern matching with 1/0/d.n.c. pattern definition (fast morphological functions)
- ✓ On-chip bank of high-speed ADCs and DACs.
- ✓ Multiple I/Os and high-speed communication ports



Main Features

Q-Eye focal-plane processor

- 176 x 144 spatial resolution
- Monochrome image sensor with 3.2V/(lux·sec) sensitivity
- Maximum frame rate (sensing + processing) of over 10,000fps
- Advanced pixel architecture combining image acquisition, image processing and storage:
 - Multi-mode image sensing, analogue & binary memories, analogue multiplexor for image shifting, analogue MAC unit, programmable LUT, resistive grid for controllable smoothing...

Digital control & post-processing

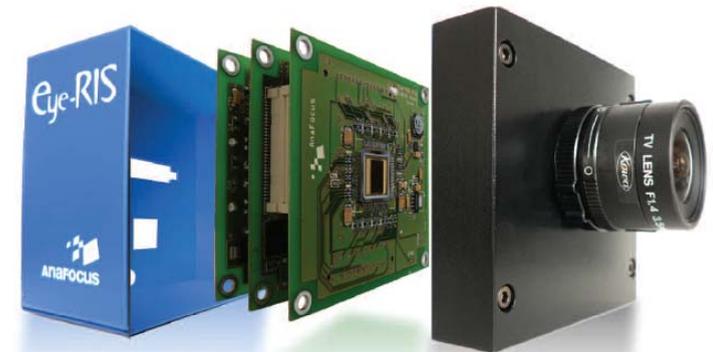
- ALTERA NIOS-II 32-bit RISC microprocessor
- 1.17 DMIPS/MHz performance at 70MHz operation frequency
- 32Mb SDRAM for program and image/data storage

Communications

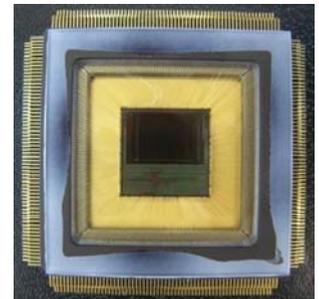
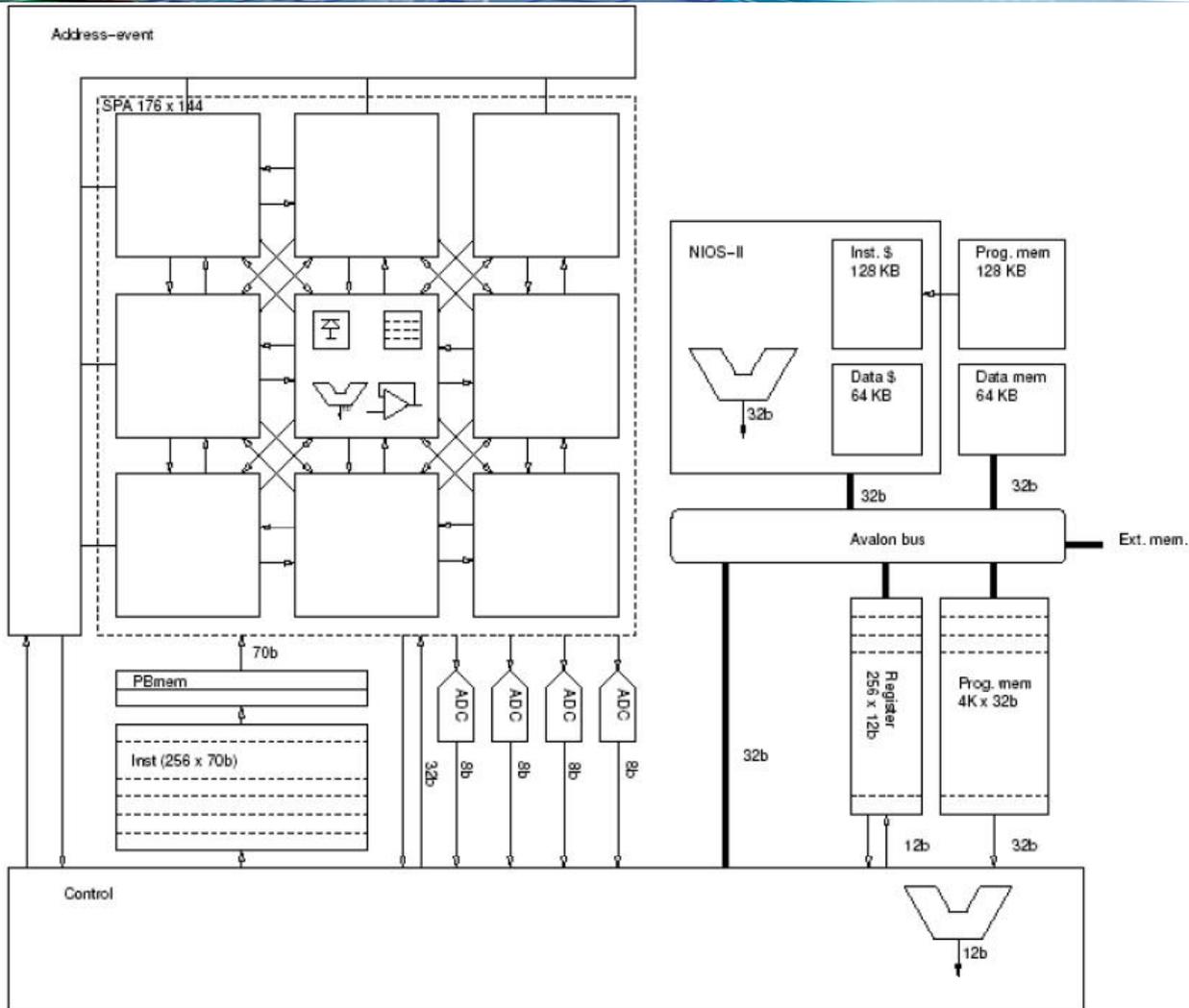
- SPI port, UART, 2xPWM ports and GPIOs, USB 2.0, and GigE
- JTAG Controller
- 1.5W typical power consumption

Application development kit

- *Application Development Kit* including: Project builder, C compiler, assembler, linker, and source-code debugger.
- *Image-processing library* including basic routines such as: point-to-point operations, spatial filtering operations, morphological operations and statistical operations



Eye-RIS™ v1.2 vision system



Main Features

Main Features

Q-Eye focal-plane processor

- 176 x 144 spatial resolution
- Monochrome image sensor with 3.2V/(lux·sec) sensitivity
- Maximum frame rate (sensing + processing) of over 10,000fps
- Advanced pixel architecture combining image acquisition, image processing and storage:
 - Multi-mode image sensing, analogue & binary memories, analogue multiplexor for image shifting, analogue MAC unit, programmable LUT, resistive grid for controllable smoothing...

Digital control & post-processing

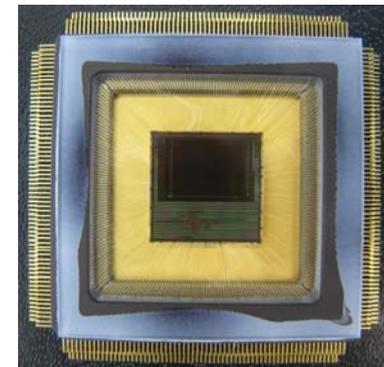
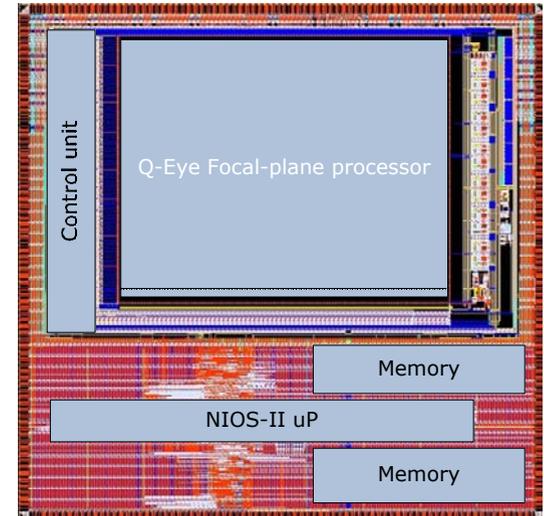
- ALTERA NIOS-II 32-bit RISC microprocessor
- 1.17 DMIPS/MHz performance at 100MHz operation frequency

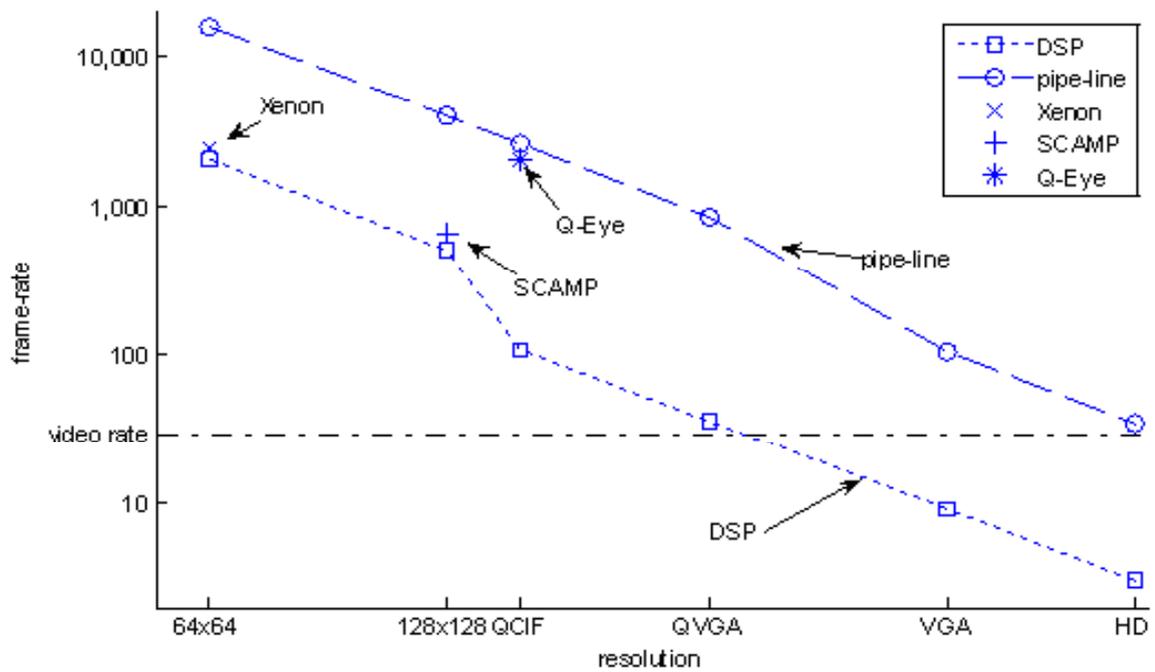
Communications

- SPI port, UART, 2xPWM ports and GPIOs, USB 2.0 interface, and
- JTAG Controller
- <700mW typical power consumption

Application development kit

- *Application Development Kit* including: Project builder, C compiler, assembler, linker, and source-code debugger.
- *Image-processing library* including basic routines such as: point-to-point operations, spatial filtering operations, morphological operations and statistical operations





Benchmark: Each of the processors calculated 20 convolutions, 2 diffusions, 3 means, and 40 morphology and 10 global ORs. Note: Only the DSP and the pipe-line multi-core (FPGA) architectures support trading between resolution and frame-rate



	General								
	Type	Tech. [nm]	Area [mm]	Data [Mb]	Pow [W]	Clock [MHz]	GOPs	mW GOP	
Xetal-1	SIMD LPA	180	25	0,05	1,6	24	5	320	
IMAP	SIMD LPA	180	121	2	2	100	51,2	39,06	
Xetal-2	SIMD LPA	90	74	10	0,64	84	107	6	
Q_Eye	SIMD SPA	180	37,5	Int.	0,1	50	250	4	
EyeRIS	SIMD SPA	180	92	0,5	1	50	250	4	
(NIOS)				0,5					
CVSoC	SIMD LPA	180	165		0,8	90	92,16	8,681	

	DSP (DaVinci [†])	Pipe-line (FPGA ⁺⁺)	Course-grain (Xenon)	Fine-grain (SCAMP/Q-Eye)
<i>Silicon technology (micron)</i>	0.065	0.065	0.18	0.35/0.18
<i>Silicon area mm²</i>			25	50/60
<i>Power consumption</i>	1.25 W	2-3W	0.08 W	0.20 W
<i>Arithmetic proc. clock speed</i>	600 MHz	250 MHz	100 MHz	1,2 - 2.5 MHz
<i>Number of arithmetic proc.</i>	8	120	256	16384
<i>Efficiency of arithmetic calc.</i>	75% *	100%	80% **	50% *
<i>Arit. computational speed</i>	3.6 GMAC	30 GMAC	20 GMAC	~20GOPS***
<i>3x3 convolution time</i>	42.3 μs****	4.9 μs	12.1 μs	22 μs ***
<i>Arithmetic speed-up</i>	1	8.6	3.5	1.9
<i>Morph. proc. clock speed</i>	600 MHz	83 MHz	100 MHz	1,2 - 5 MHz
<i>Number of morphologic proc.</i>	64	864	2048	147456
<i>Morphologic processor kernel type</i>	2 × 32 bit	96 × 9 bit	256 × 8 bit	16384 × 9 bit
<i>Efficiency of morphological calc.</i>	28% *	100%	90% **	100%
<i>Morphologic computational power</i>	10 GOPS	71 GOPS	184 GOPS	737 GOPS
<i>3x3 morphologic operation time</i>	13.6 μs****	2.05 μs	1.1 μs	0.2 μs
<i>Morphologic speed-up</i>	1	6.6	12.4	68.0

† Texas Instrument Davinci video processor (TM320DM043)

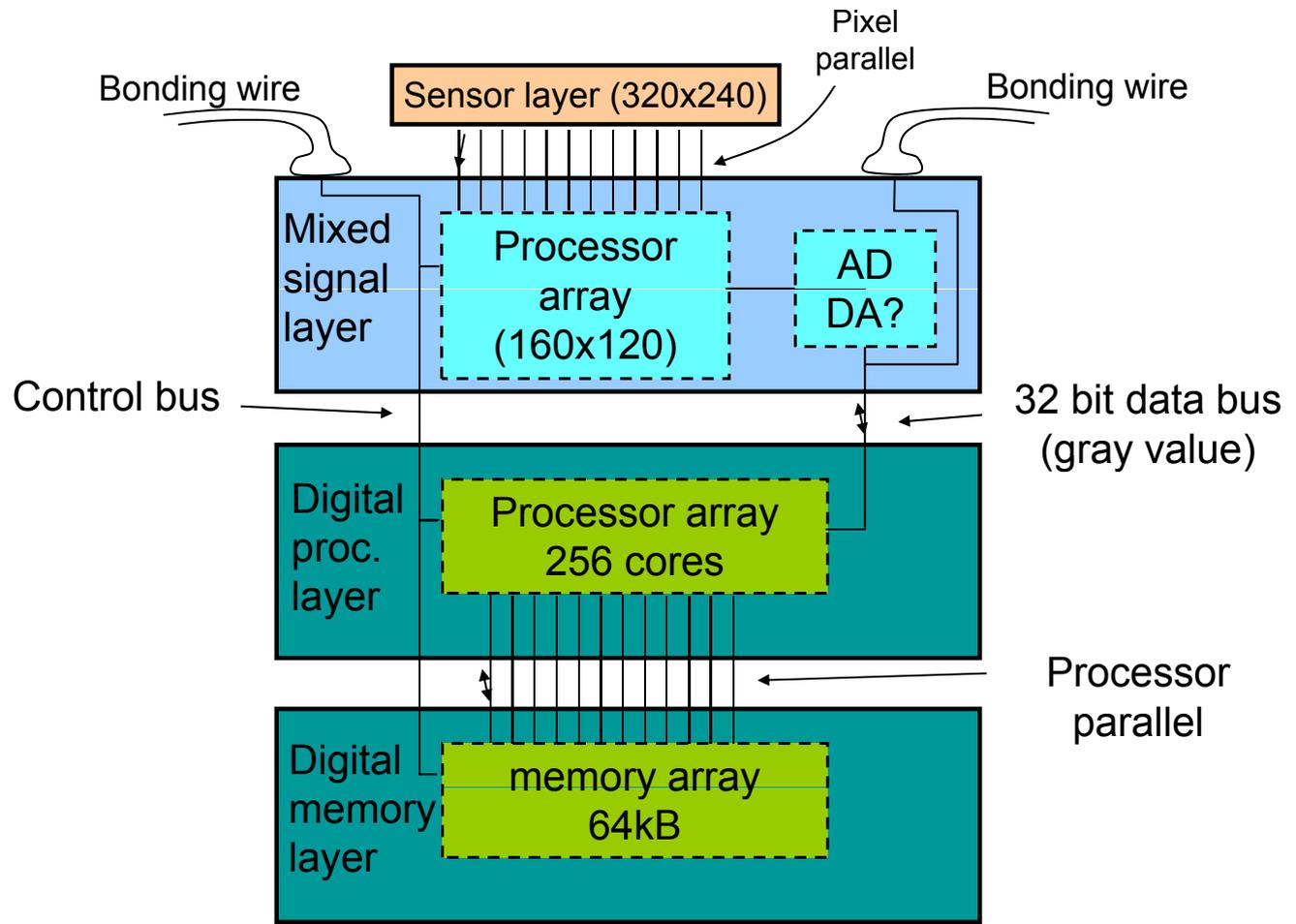
++ Xilinx Spartan 3A DSP FPGA (XC3SD3400A)

* due to data access

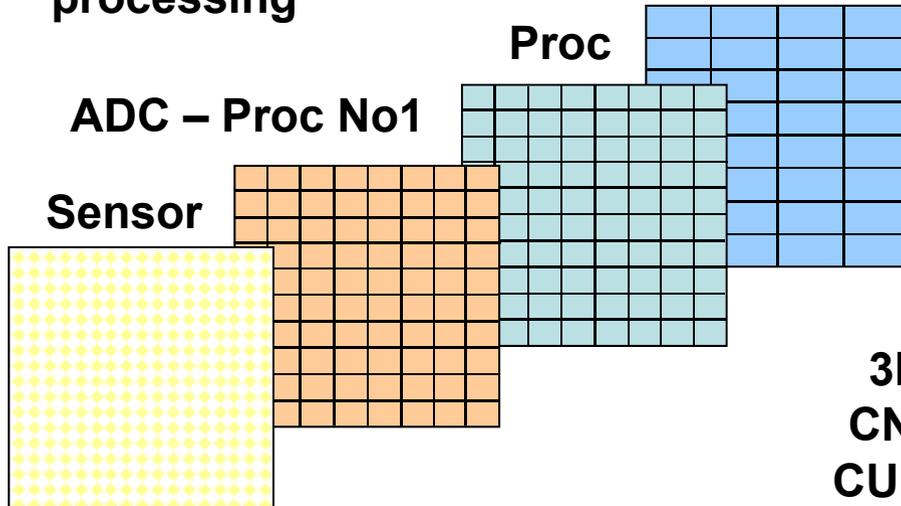
** due to pipe-line operations

*** no multiplication, scaling with a few discrete values

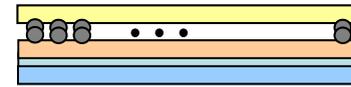
**** these data intensive operators slow down significantly when the image does not fit to the internal memory (typically above 128x128 for a DaVinci, which has 64kByte internal memory)



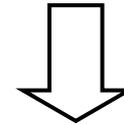
Topographic,
layered
sensing-
processing



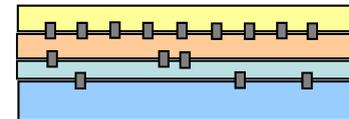
AFRL BB Process



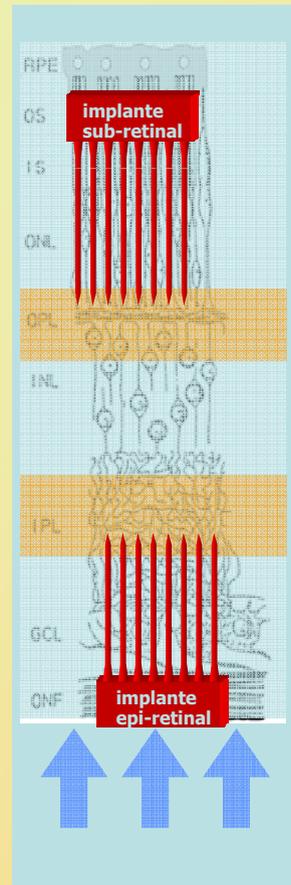
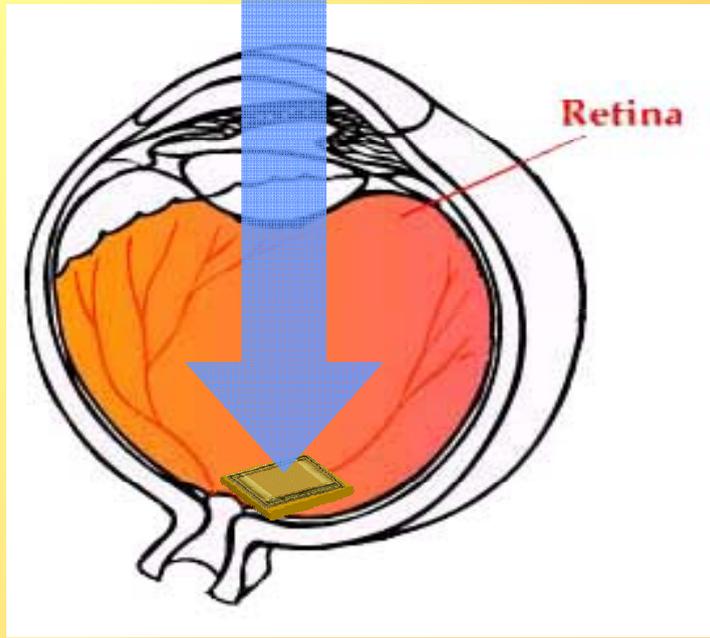
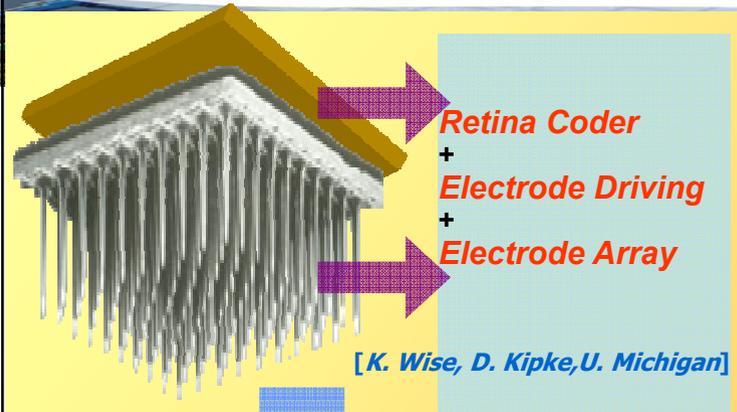
Bump bonding:



3D stacking:

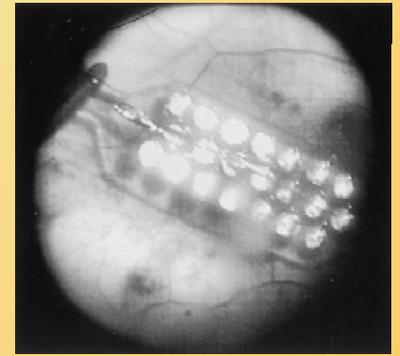


MITLL Low-Power
FDSOI CMOS Process



Sub-Retinal Implant

[Chow et. al 2001]



Epi-Retinal Implant

[Humayun et. al 1999]

Discussion

→ **VSoCs** are still like **science fiction** toys for industry

→ Even **smart CMOS sensors** are **not common yet**

A real challenge for engineering !!

→ We can **take advantage of nature** through understanding

- **Parallel and concurrent sensory-processing**

- **Multi-scale representation**

- **Signal coding**

→ **Close interaction and collaboration between analog and digital circuitry is needed for efficient VSoC design.**