Design of a smart camera SoC in a 3D-IC technology

R. Carmona-Galán, J. Fernández-Berni, S. Vargas-Sierra, G. Liñán-Cembrano, Á. Rodríguez-Vázquez, V. Brea-Sánchez[(*)], M. Suárez-Cambre[(*)], D. Cabello-Ferrer[(*)]

Institute of Microelectronics of Seville (IMSE-CNM), CSIC-Universidad de Sevilla (Spain)

[(*)]Information Technology Research Center (CITIUS) Univ. de Santiago de Compostela (Spain)

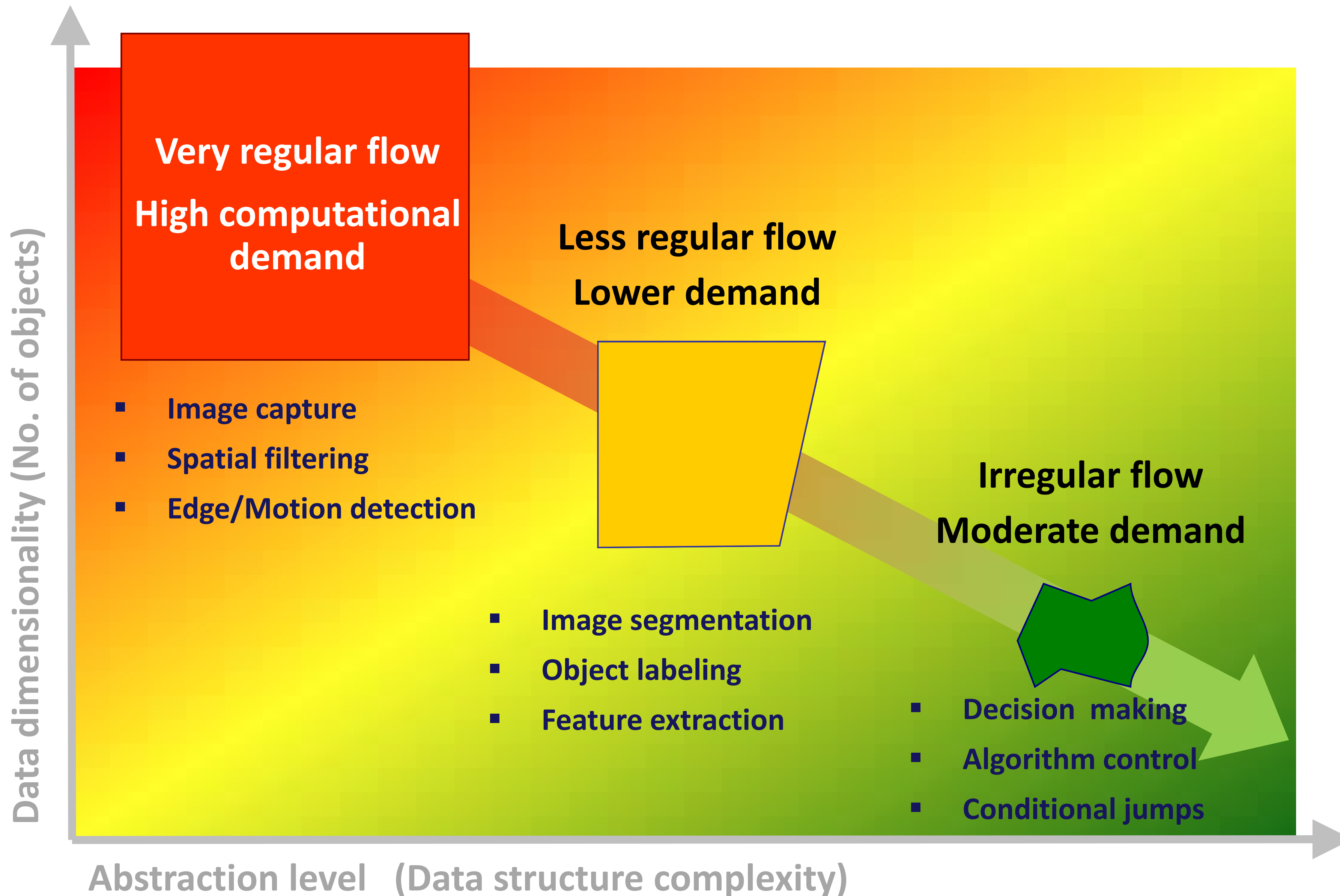
Workshop on Architecture of Smart Camera

Clermont-Ferrand, France
April 5-6, 2012

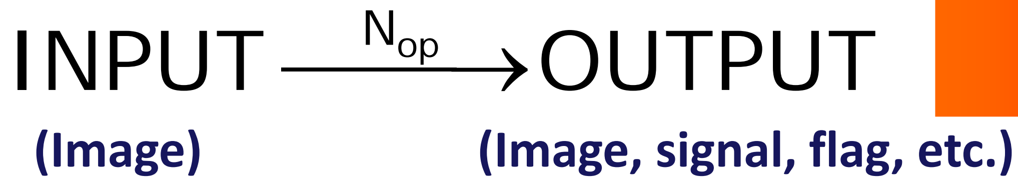
Main lines

- Conventional digital signal processing architectures introduce **data bottlenecks** and are inefficient when dealing with **multidimensional sensory signals**
- Architectures **adapted** to the nature of the stimulus are **more efficient** in terms of power consumption per operation but…
- Concurrent sensing, processing and memory in **planar technologies** introduces serious **limitations to image resolution** and **image size** via the penalties in fill factor and pixel pitch
- **3D integrated circuit technologies** with a dense TSV distribution permits **eliminating data bottlenecks** without degrading image resolution and size.

Computational demand in artificial vision



Power-speed trade-off



Time-critical applications

$$\text{Speed} = \frac{1}{T_{tot}} = \frac{N_{proc}}{N_{op}} \cdot \frac{1}{t_0}$$

Energy:

$$E_{tot} = N_{op} \cdot e_0$$

Power-aware applications

$$\text{Power} = \frac{E_{tot}}{T_{tot}} = N_{proc} \cdot \frac{e_0}{t_0}$$

Time:

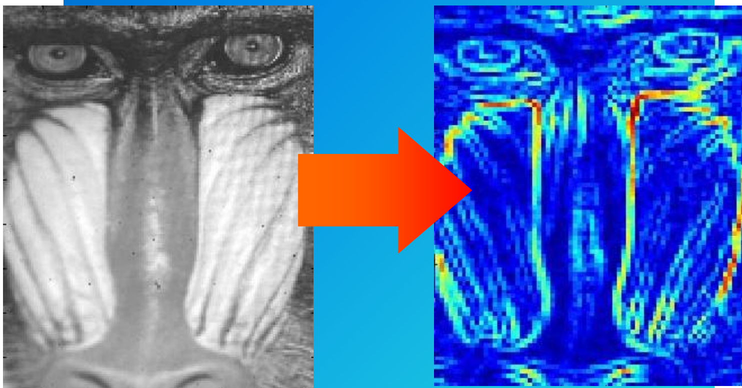
$$T_{tot} = N_{op} \cdot \frac{t_0}{N_{proc}}$$

Power and time-critical applications

$$\text{FOM} = \frac{\text{Speed}}{\text{Power}} = \frac{1}{E_{tot}} = \frac{1}{N_{op} e_0}$$

Strategies for E_{tot} minimization

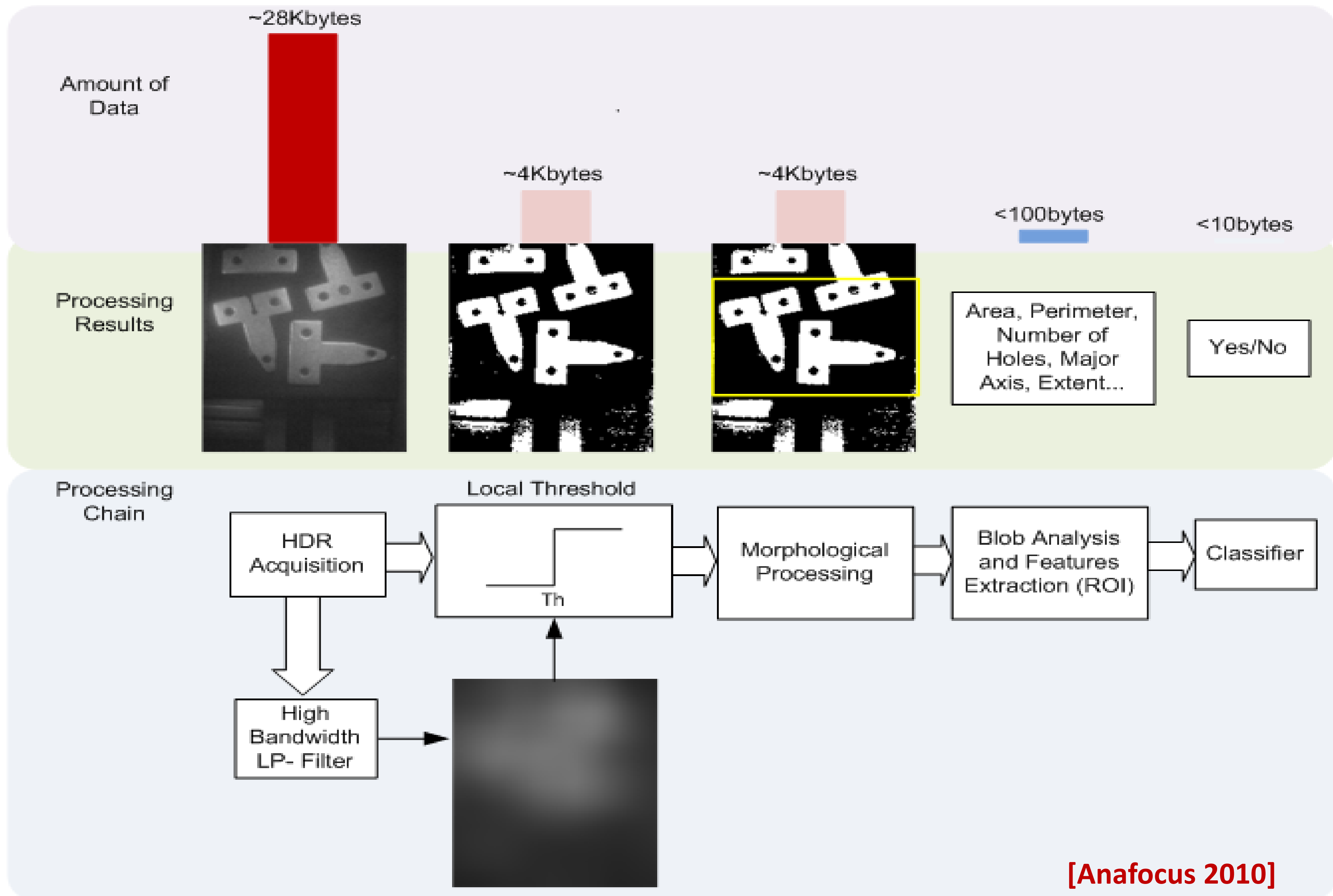
Minimization of N_{op}
(number of operations)



- Simplified image
- Hierarchical processing
- Sparse representation
- Compressed sensing

Minimization of e_0
(energy per operation)

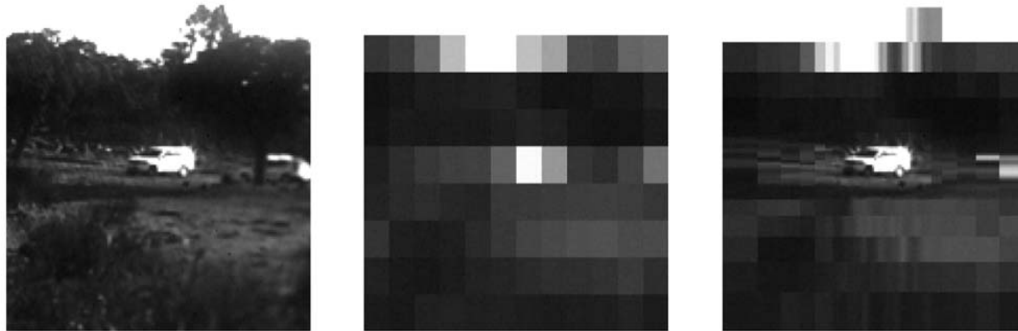
Hierarchical processing and data reduction



Feature based processing

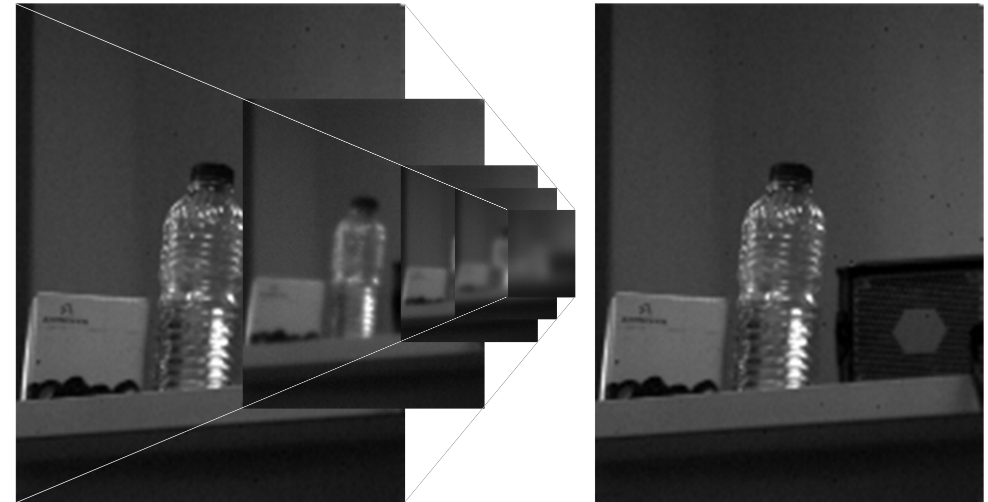
[Fernández-Berni et al. 2011]

Multiresolution and foveation

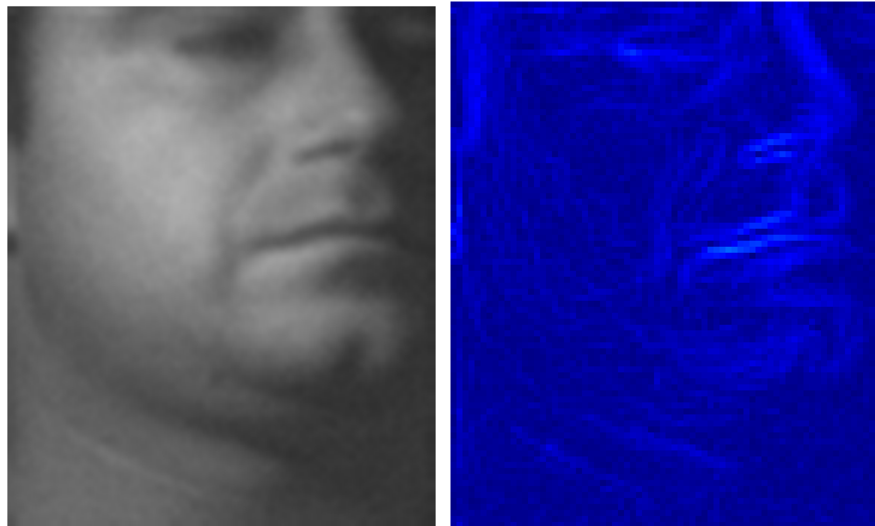


Original image | 12×16 px | Foveation

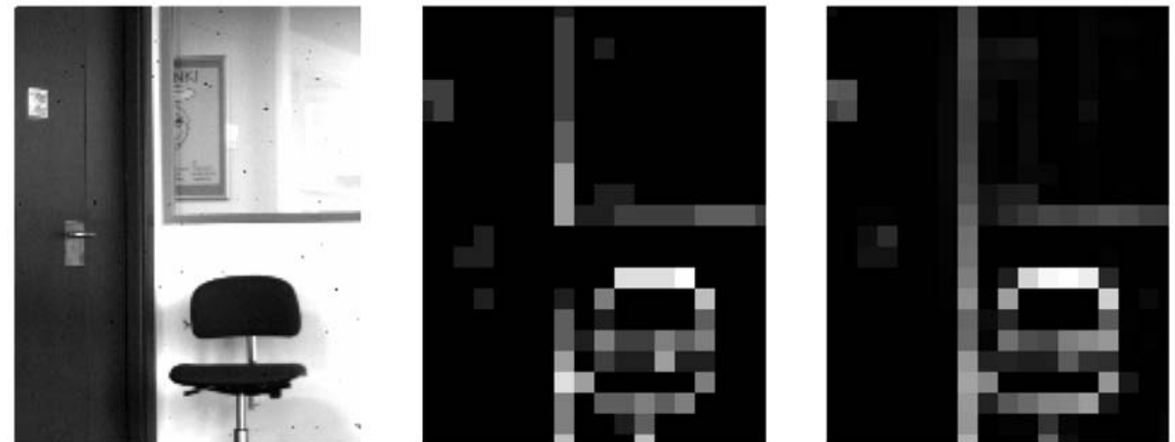
Gaussian pyramid and scale-space



Edge extraction



Energy-based representation and saliency

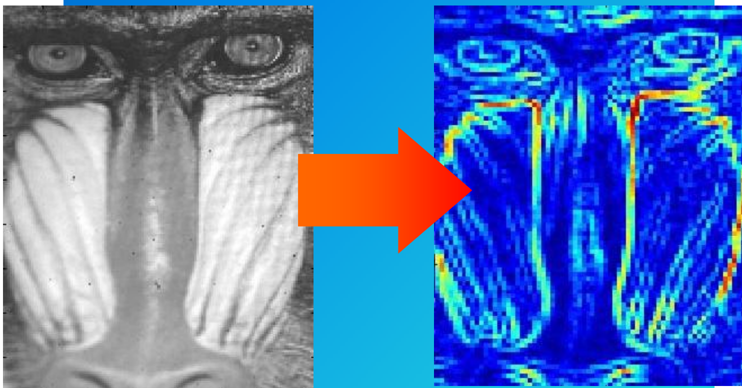


Original image | On-chip processing | Ideal processing

Strategies for E_{tot} minimization

Minimization of N_{op}

(number of operations)

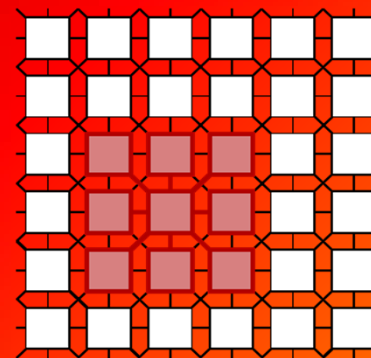


- Simplified image
- Hierarchical processing
- Sparse representation
- Compressed sensing

Minimization of e_0

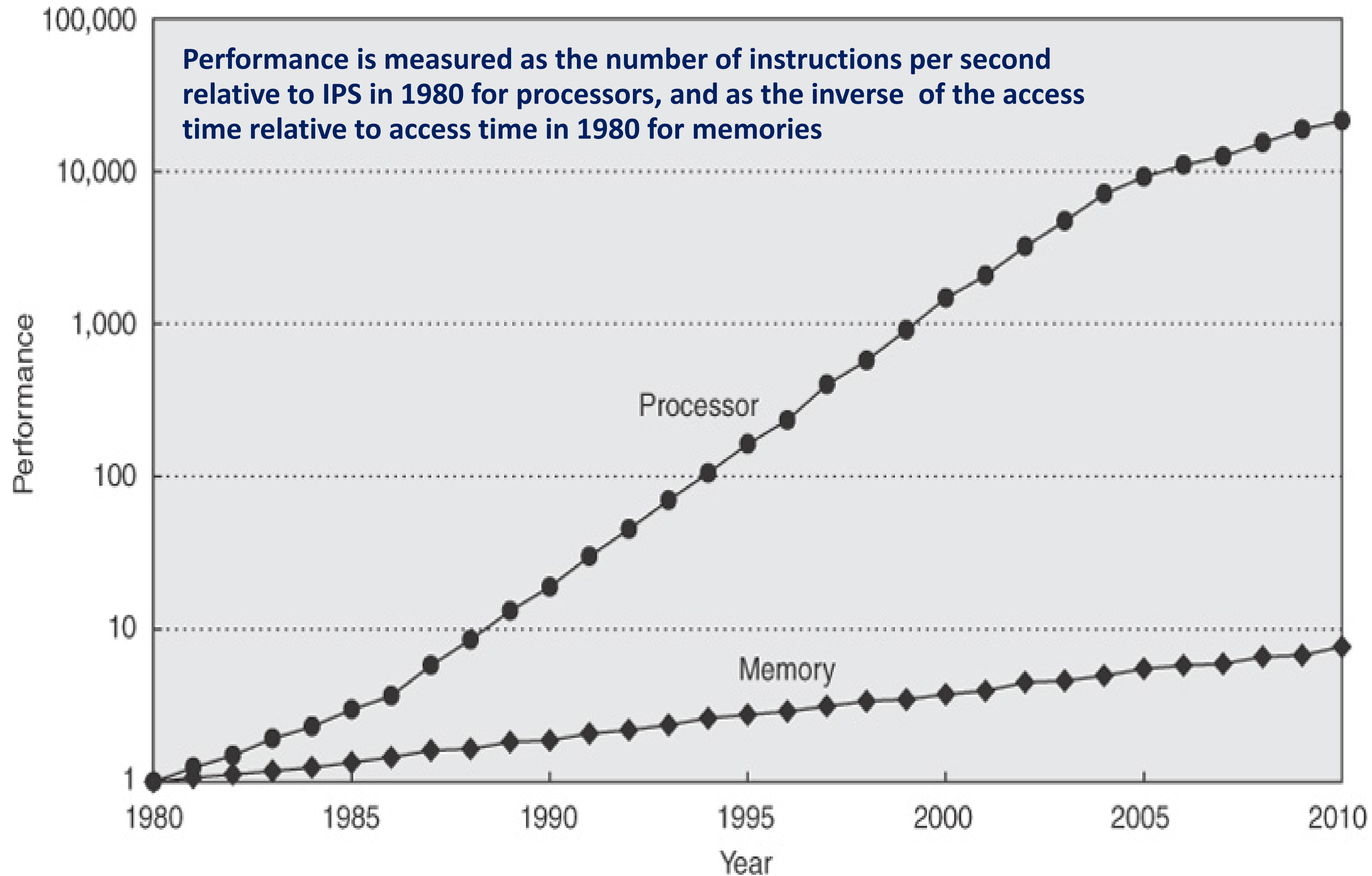
(energy per operation)

@ system level:



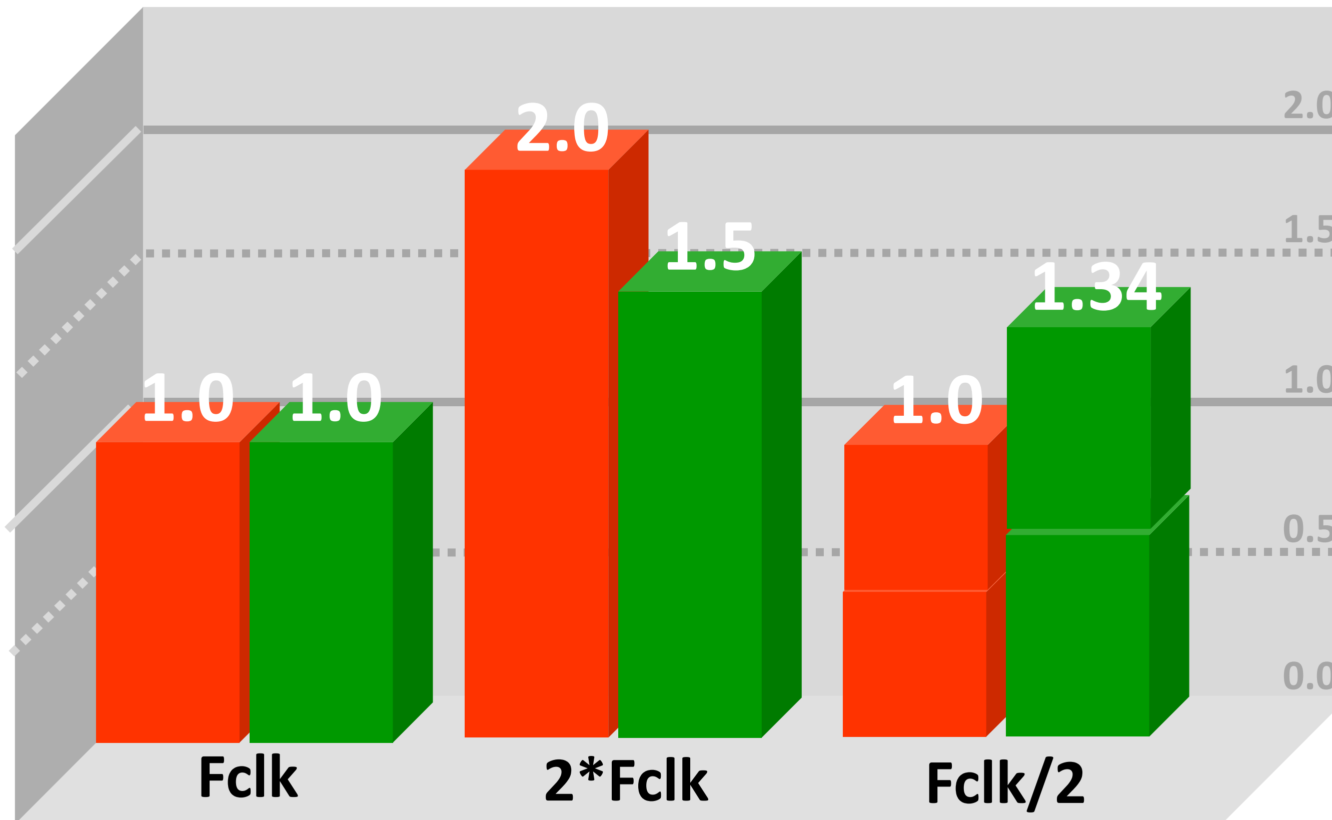
- Distributed processing
- Distributed memory
- Distributed ADC

Processor/memory performance gap



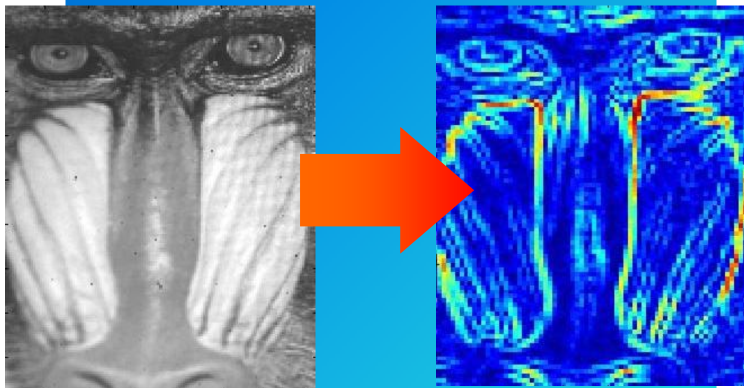
[Hennessy & Patterson 2006]

Multicore architectures



Strategies for E_{tot} minimization

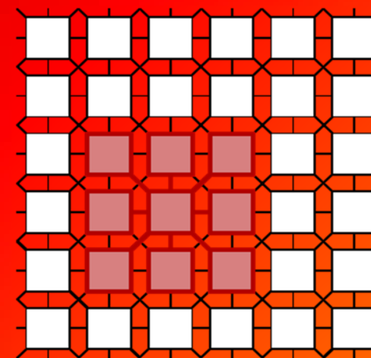
Minimization of N_{op}
(number of operations)



- Simplified image
- Hierarchical processing
- Sparse representation
- Compressed sensing

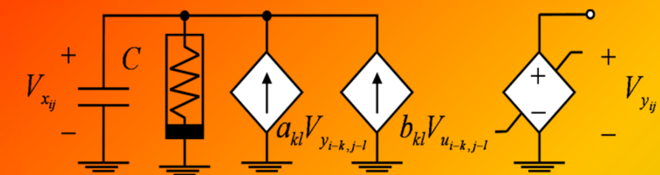
Minimization of e_0
(energy per operation)

@ system level:



- Distributed processing
- Distributed memory
- Distributed ADC

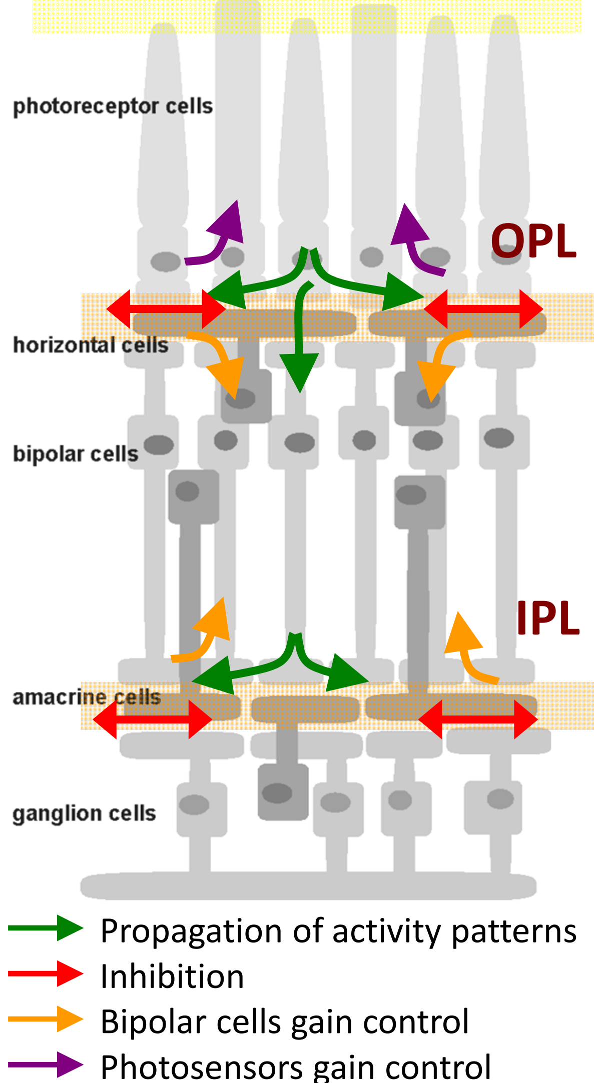
@ PE level:



- Power efficient circuits
- High signal/bias ratio
- Complex dynamics

CNN model for retinal signal processing

4 types of interaction

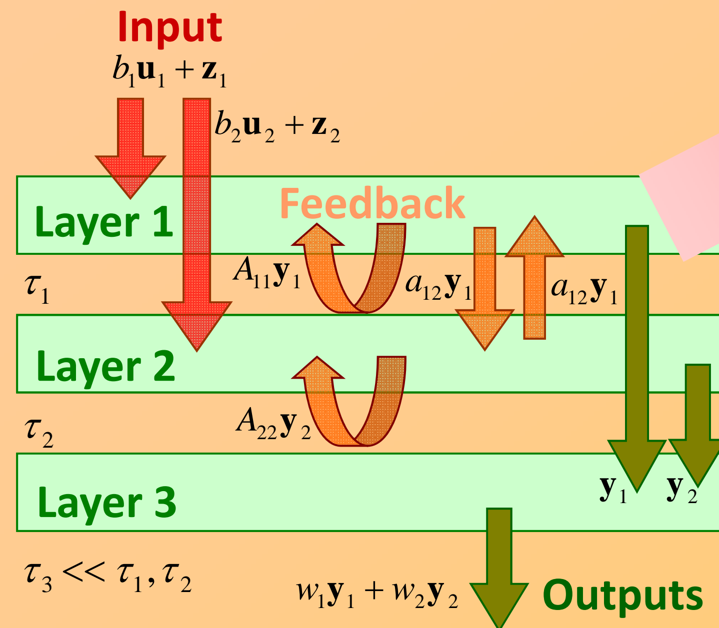


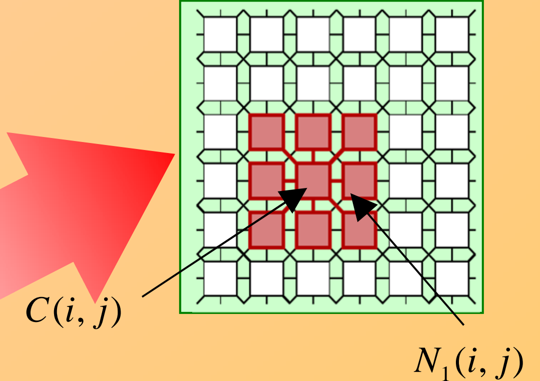
- Propagation of activity patterns
- Inhibition
- Bipolar cells gain control
- Photosensors gain control

[Roska & Werblin 2001]

CNN models for the OPL and IPL

[Rekeczky, Balya et al. 2000]





- Non-linear dynamic processors
- Local interactions by means of continuous signals (in amplitude and time)
- Interconnection pattern (cloning template) = analog program

[Chua & Yang 1988]

$$\tau_k \frac{dx_{ij_k}}{dt} = -g[x_{ij_k}(t)] + \sum_{l=1}^{3} \sum_{m=-1}^{1} \sum_{n=-1}^{1} [a_{mn_{kl}} y_{(i+m)(j+n)_l} + b_{mn_{kl}} u_{(i+m)(j+n)_l}] + z_{ij_k}$$

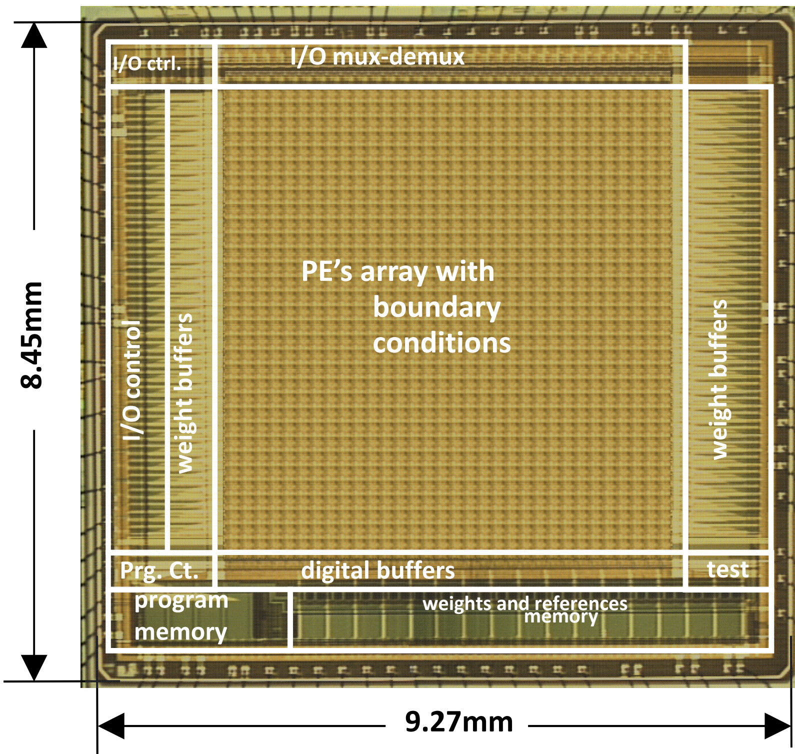
losses — feedback — feedforward — bias

2-CNN-layer (in-plane) chip

Analog Parallel Array Processor with 1024 PE's

- 0. 5mm standard CMOS
- 2 CNN layers of 32 x 32 nodes
- Programmable time constant ratio
- Local logic unit and local memories
- 24 programmable weights

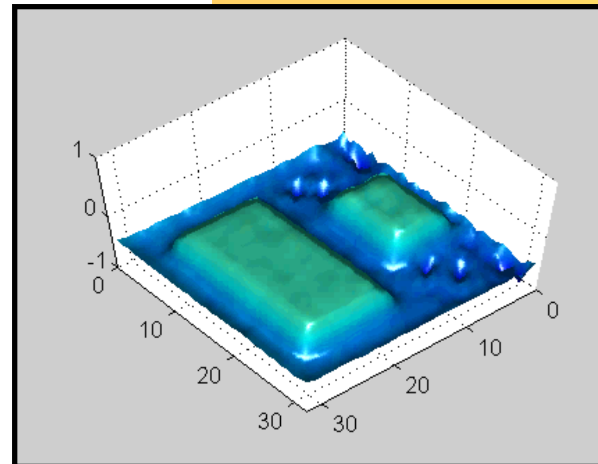
[Carmona et al. 2002]

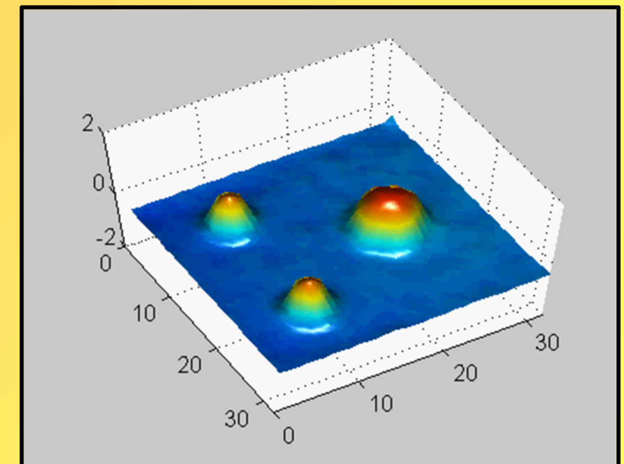


Realizes a set of coupled reaction-diffusion equations

- Wave phenomena in active media
- Pattern generation
- Retinal dynamics emulation

[Petrás et al. 2003]



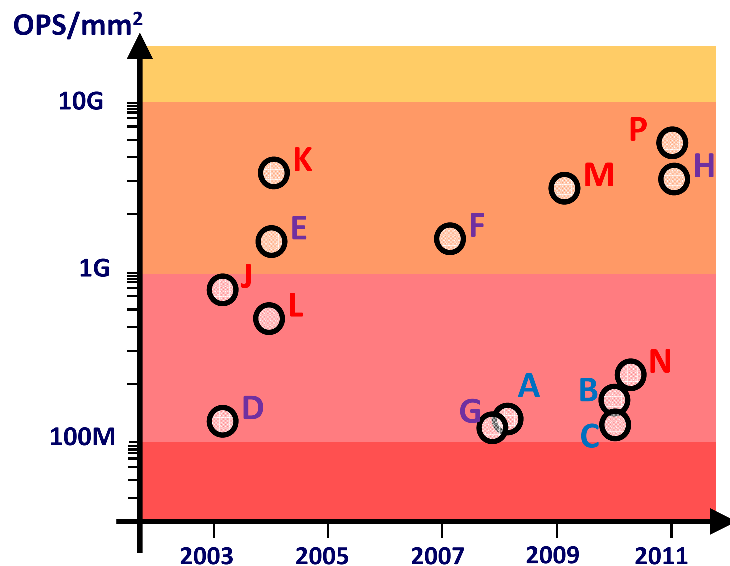


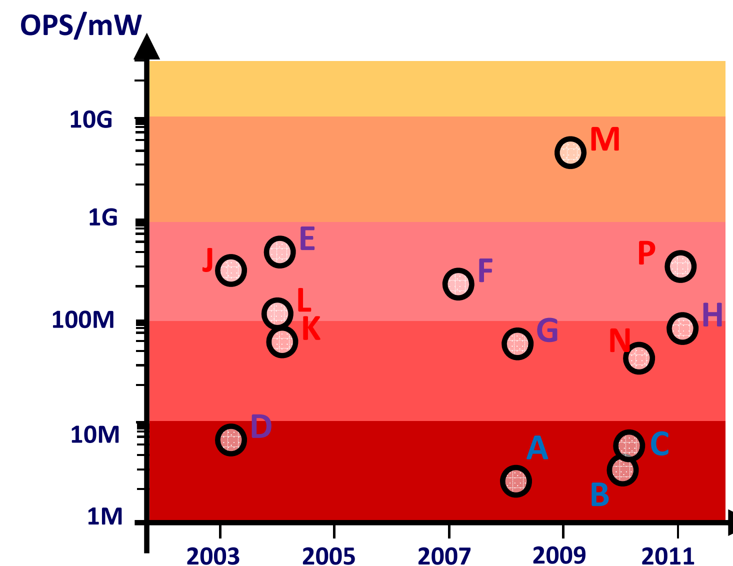
$$\frac{d}{dt}\phi_i(x,y,t)-c_i\nabla^2\phi_i(x,y,t)=\alpha_i\phi_i(x,y,t)+\beta_i\phi_i(x,y,t_0)+\gamma_{ij}\phi_j(x,y,t)$$

diffusion | reaction

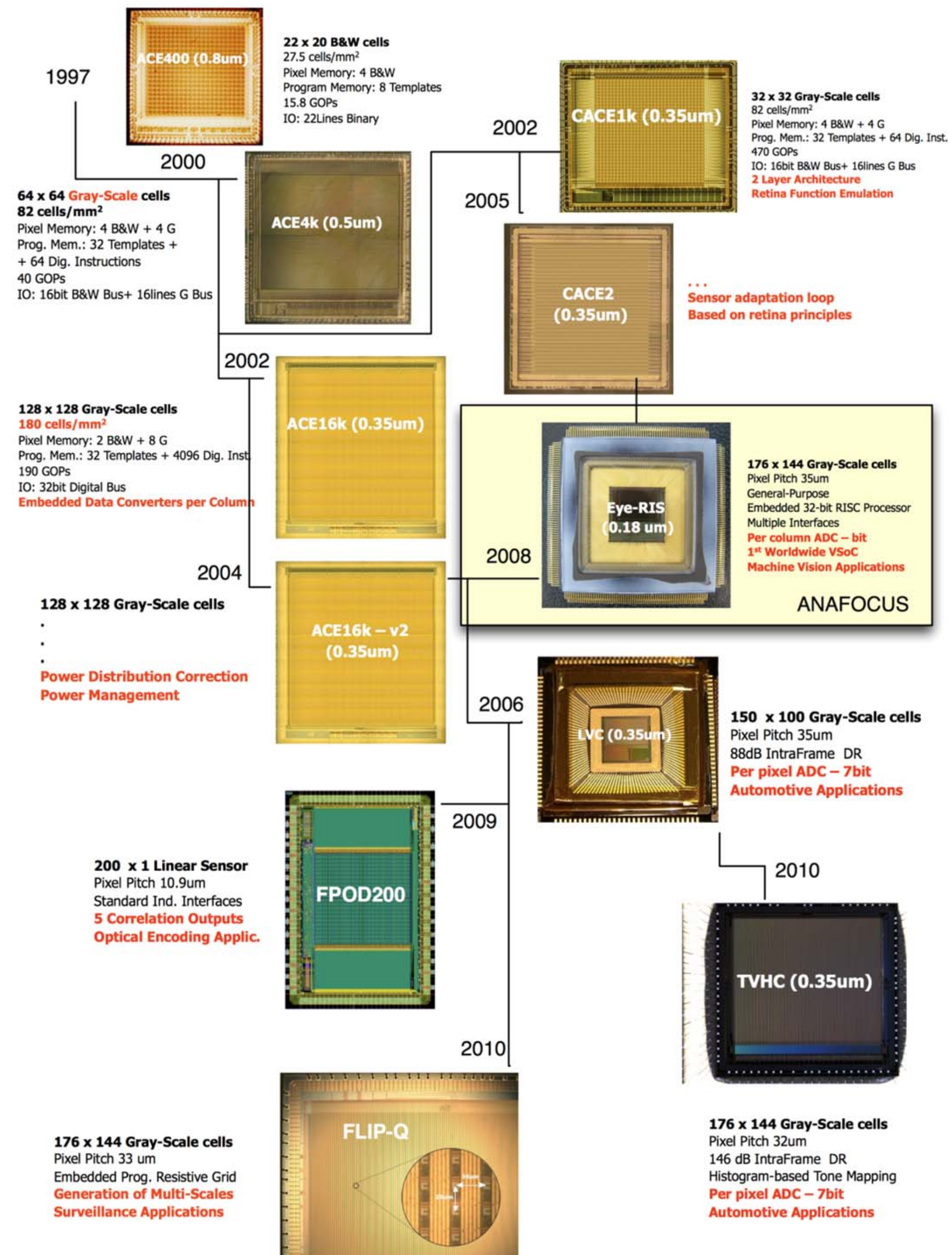
Performance chart

		Chip	Tec.	Description	Res.	Clk (MHz)	PE's/mm²	OPS/mm²	OPS/mW
CPUs + GPUs	A	[Intel 2008]	45n	Atom Single-core	64b	1730	0.038	0.125G	1.32M
	B	[Intel 2010]	45n	Atom dual-core	64b	1300	0.023	0.160G	1.64M
	C	[Nvidia 2010]	40n	Tegra (2ARM9+8CPU)	32b	1000	0.204	0.047G	4.60M
Digital SIMD	D	[Raab 2003]	350n	Parallel array 16 PEs	32b	100	0.080	0.104G	6.60M
	E	[Komuro 2004]	500n	SIMD 64 x 64 PEs	1b	10	140	1.40G	365M
	F	[Abbo 2007]	180n	Xetal-II Het. Multicore 320PEs	16b	84	4.32	1.45G	178.3M
	G	[Miao 2008]	180n	SIMD 16 x 16 PEs	4b	300	833.3	0.094G	24.4M
	H	[Zhang 2011]	180n	Multi-level SIMD 32+32 x 128	8b	100	317.5	3.4G	97.8M
Focal-plane processors	J	[Carmona 2003]	500n	RD CNN 2 x 32 x 32 cells	8b	10	58.4	0.963G	250M
	K	[Liñán 2004]	350n	Parallel array 128 x 128 cells	8b	100	180	3.20G	82.5M
	L	[Dudek 2004]	350n	Current mode SIMD 39 x 48 PEs	6b	2.5	410	0.513G	104M
	M	[Gottardi 2009]	350n	Parallel array 128 x 64cells	8b	80	409.6	2.8G	4G
	N	[Lopich 2010]	350n	Cellular Proc. 19 x 22 cells	8b	75	85.5	0.25G	38M
	P	[Lee 2011]	130n	Digital CNN 80 x 60 + 120PEs	8b	200	1093	5.33G	285.7M





Smart CIS based on FPP @IMSE



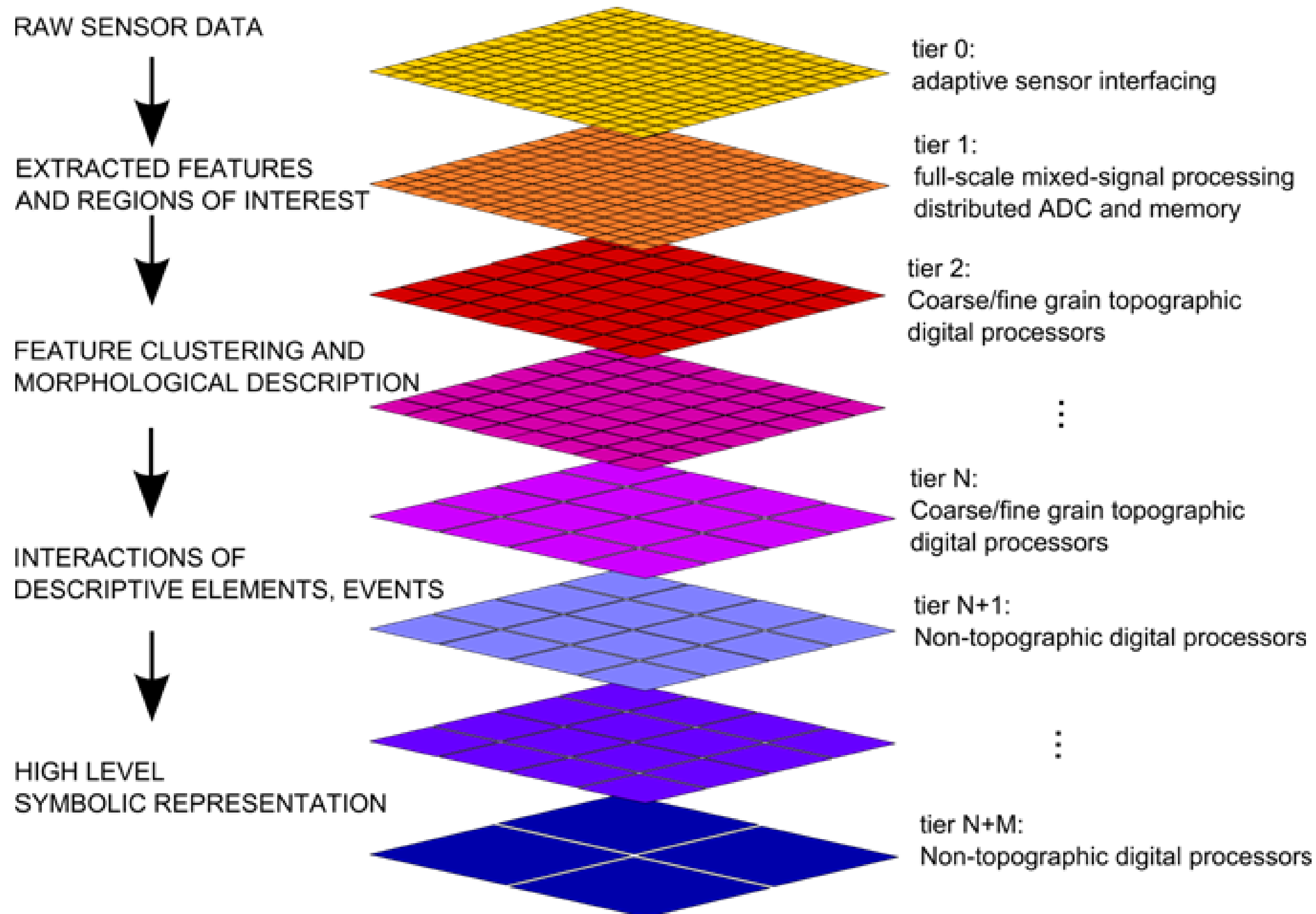
Major achievements

- Fully programmable features
- Large variety of functional targets
- Image-to-Decision at >1,000fps using 60nW per pixel
- Spatio-temporal filtering @22nJ/cycle
- Content-aware HDR acquisition with >145dB intra-frame DR

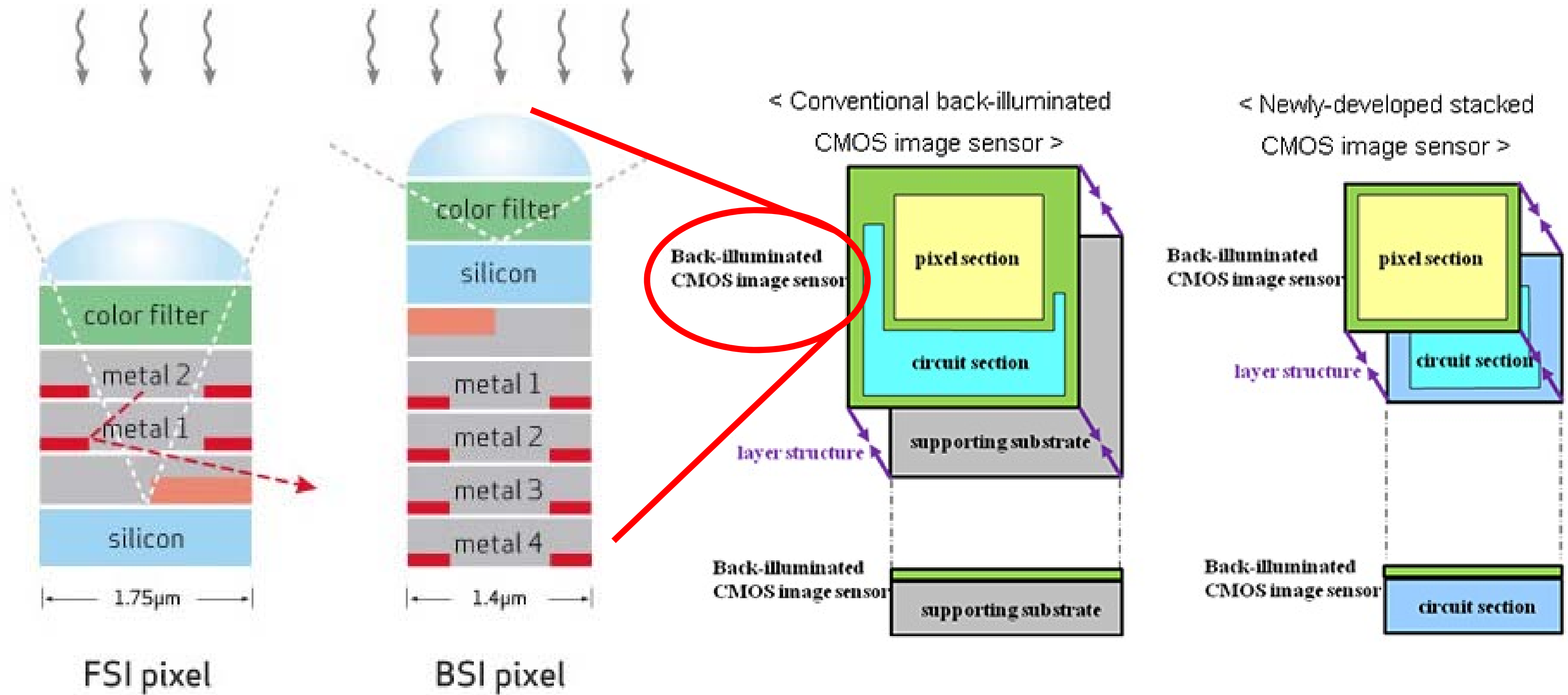
Major drawbacks

- Reduced fill factor
- Large pixel pitch
  - → **Small image size**
  - → **Limited resolution**
  - → **Sensitivity vs. resolution trade-off**

Multilayer hierarchical vision architecture



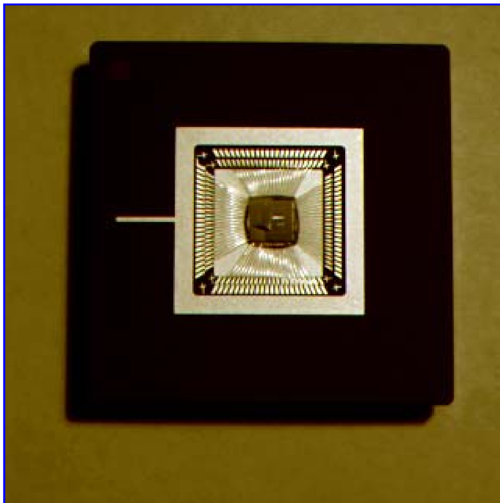
3D integration for CMOS image sensors



[OmniVision 2010]

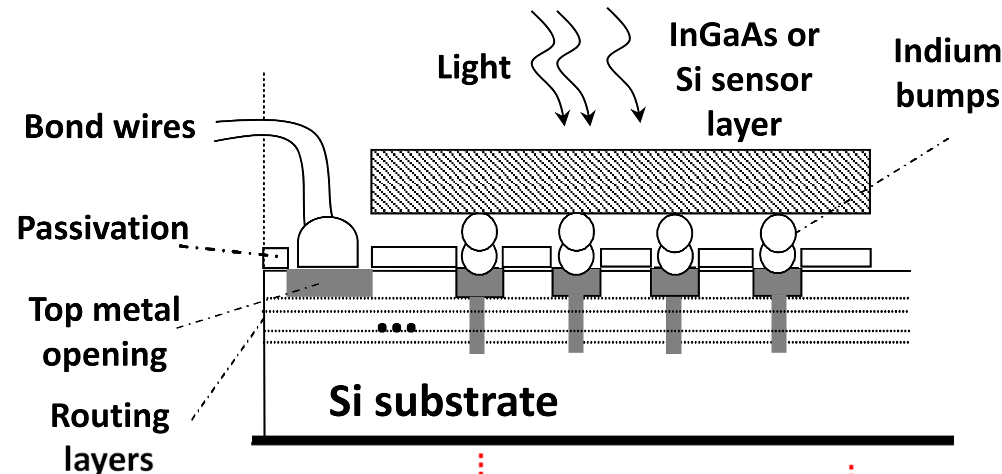
[Sony 2012]

1st attempt: bump-bonded sensor layer

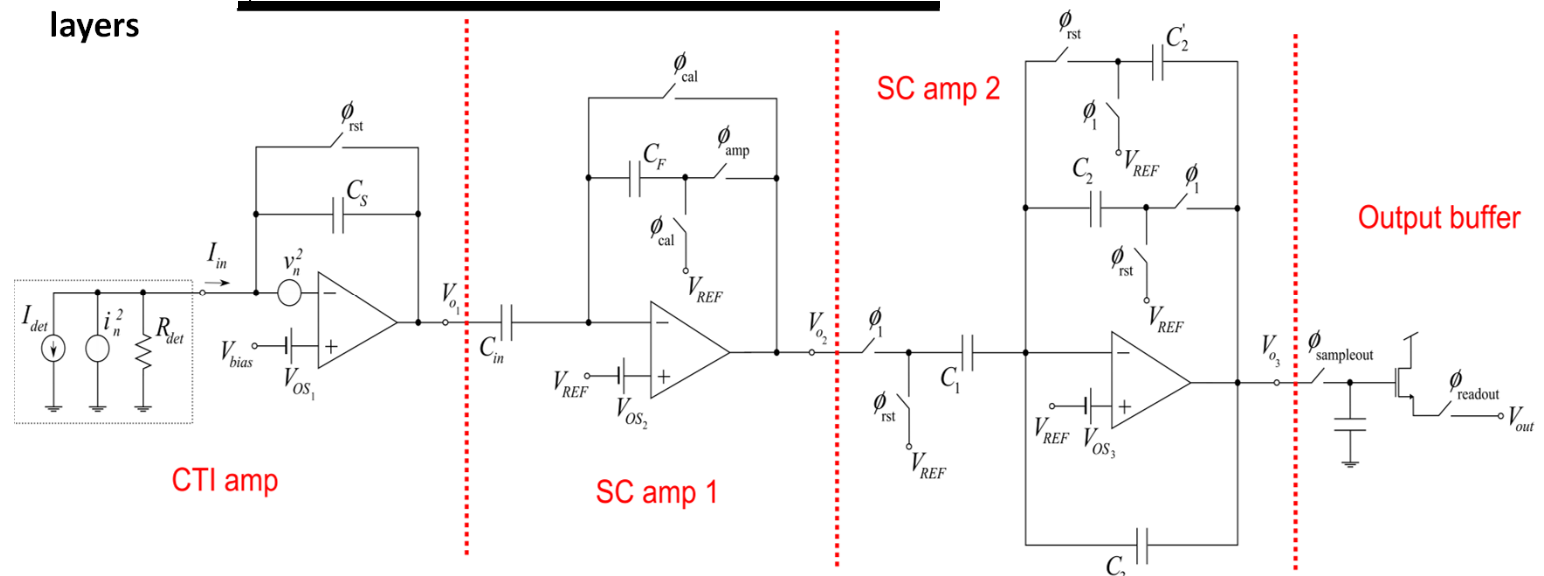


[Rekeczky et al. 2007]

Xenon-NC V1	
Technology	0.18um UMC
Die size	5x5 mm2
# Pixels	8x8
# Pixel pitch	125um
# PEs	8
Int. word length	24b
Clock frequency	80 MHz
Local memory	64 words



- CMOS compatible
- High fill factor
- Custom spectral responsivity



CTIA

- Zero-bias detection
- Sense capacitor matching

Full-custom ROIC

- Reconfigurable gain
- Adaptive sensing

2nd attempt: VISCUBE 3D IC stack

Operation control

Sensor array (320x240) — Thin-film diodes

Data I/O

320x240 bump bonds

3D IC stack

Mixed-signal layer (160x120) — tier 3: sensor interface & feature extraction

I/O pads

160x120 TSVs

Frame buffer (160x120x(6x8b+1b)) — tier 2: distributed image memory

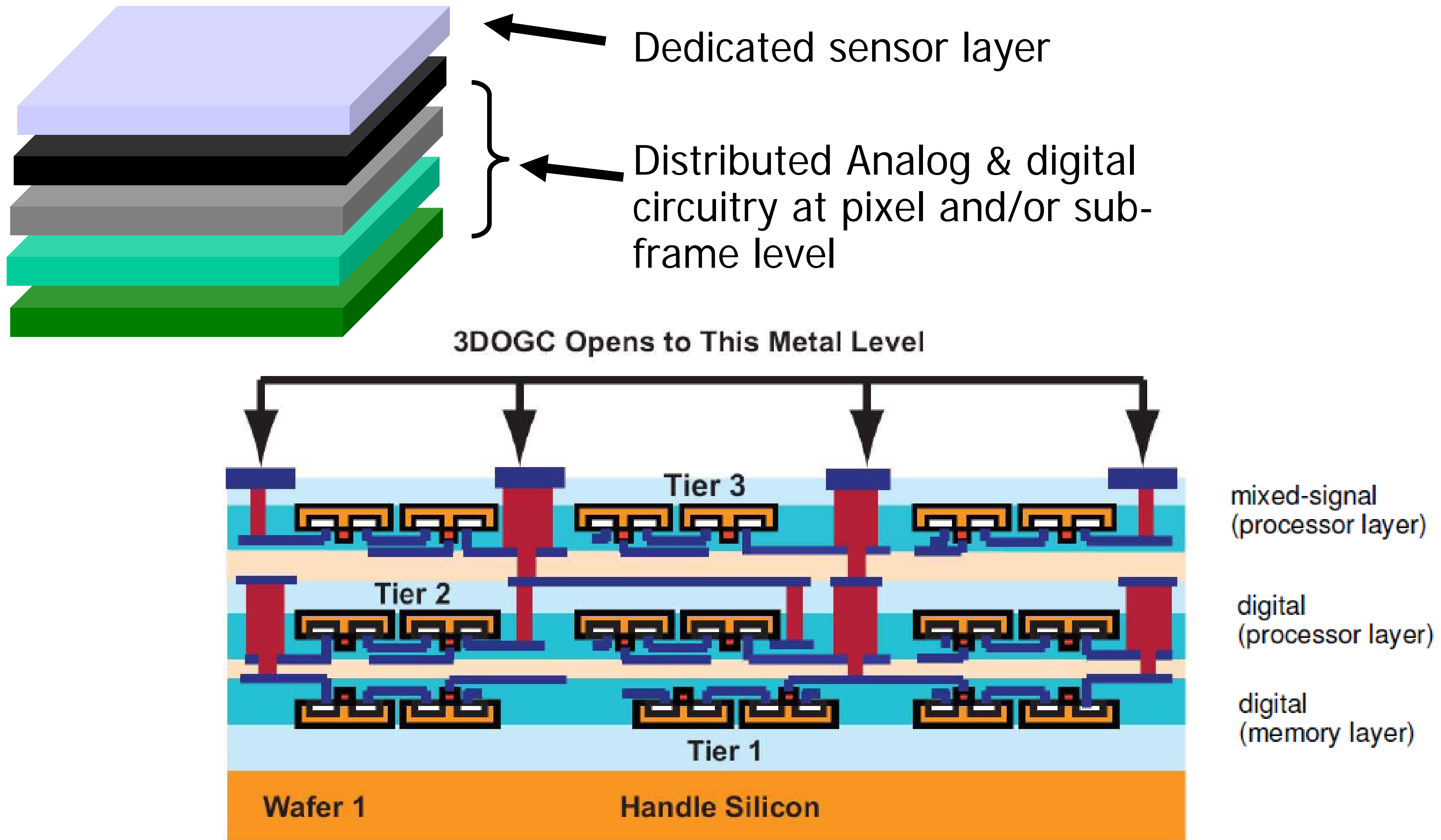
32x32 TSVs

Digital processor (32x32) — tier 1: digital fovea

Project partners:

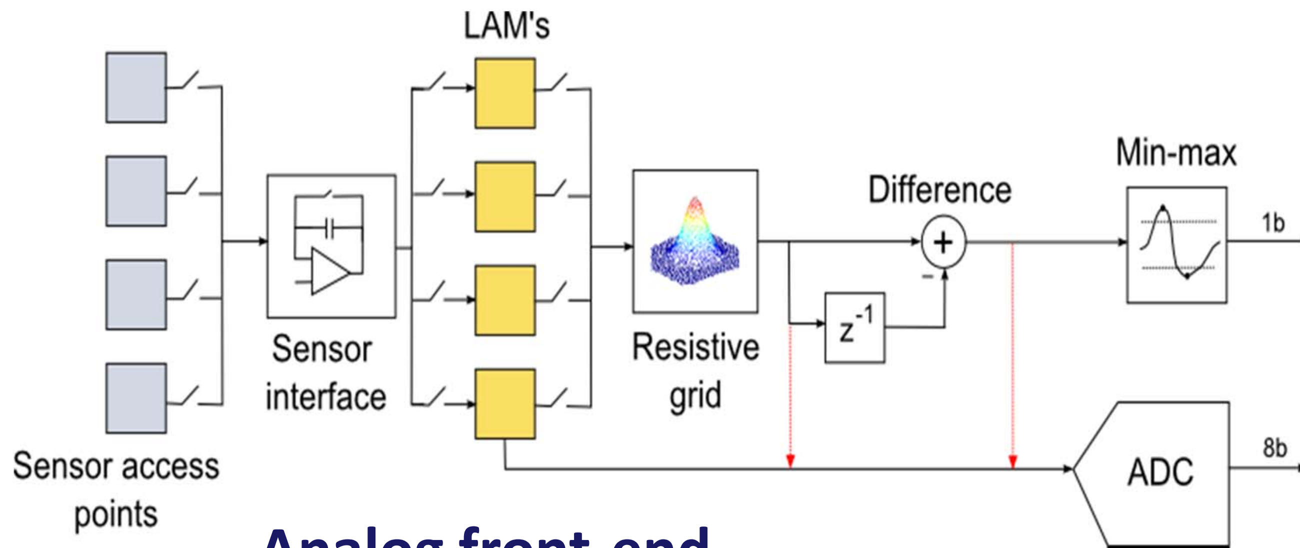
Funding agency:

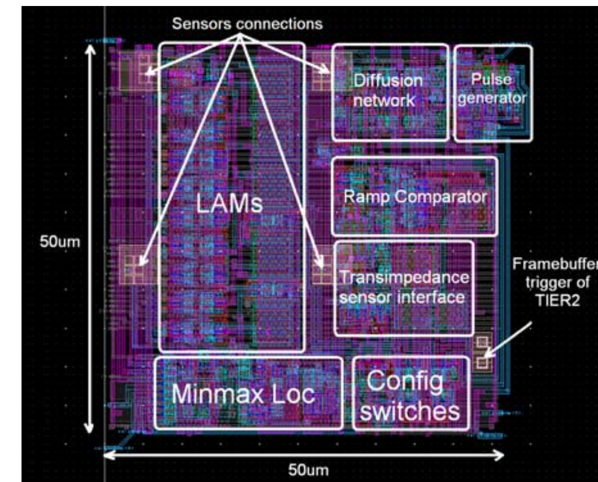
3D-IC fabrication process



MIT Lincoln Labs 0.18um FDSOI CMOS process (funded by DARPA)

Tier 3: sensor interface and feature extraction



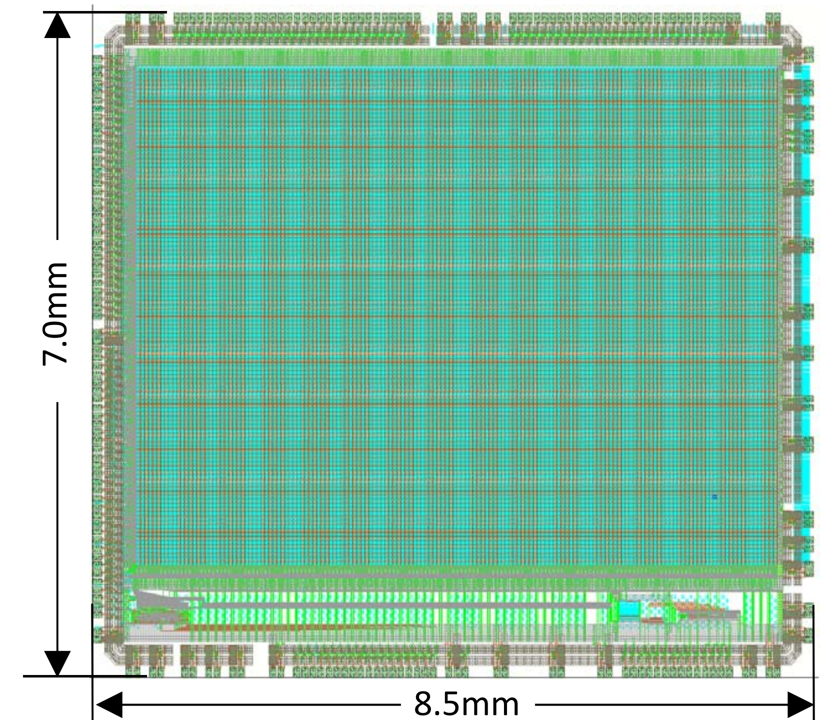


Analog front-end

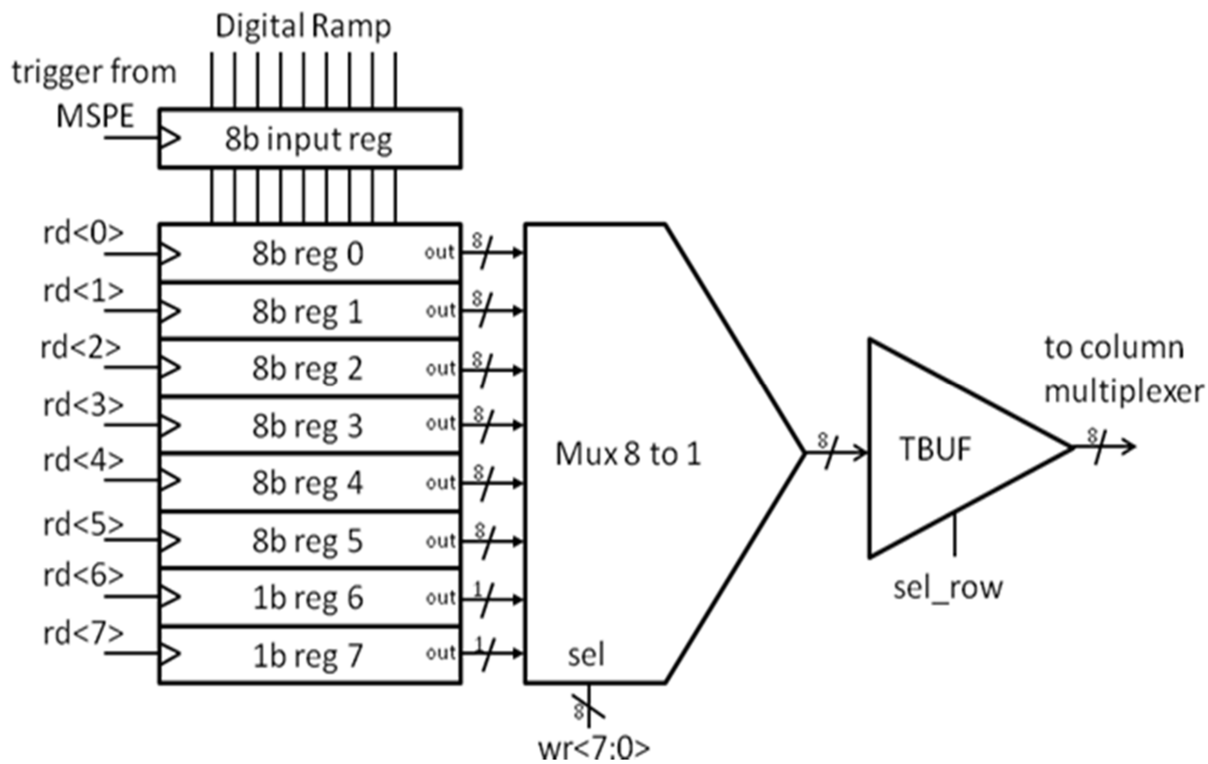
- Capacitive transimpedance amplifier
- Multiplexed sensor interface
- Full frame refresh at 1kfps

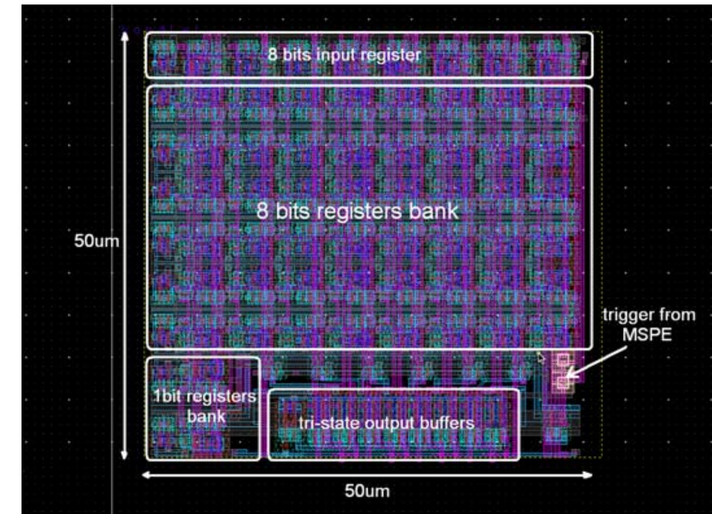
Focal-plane processing

- A/D conversion of 320x240 raw image
- Binning of the 4 sensors
- Filtering at 2 user-selected scales
- Subtraction of the two scales
- Local maxima and minima detection



Tier 2: distributed image memory





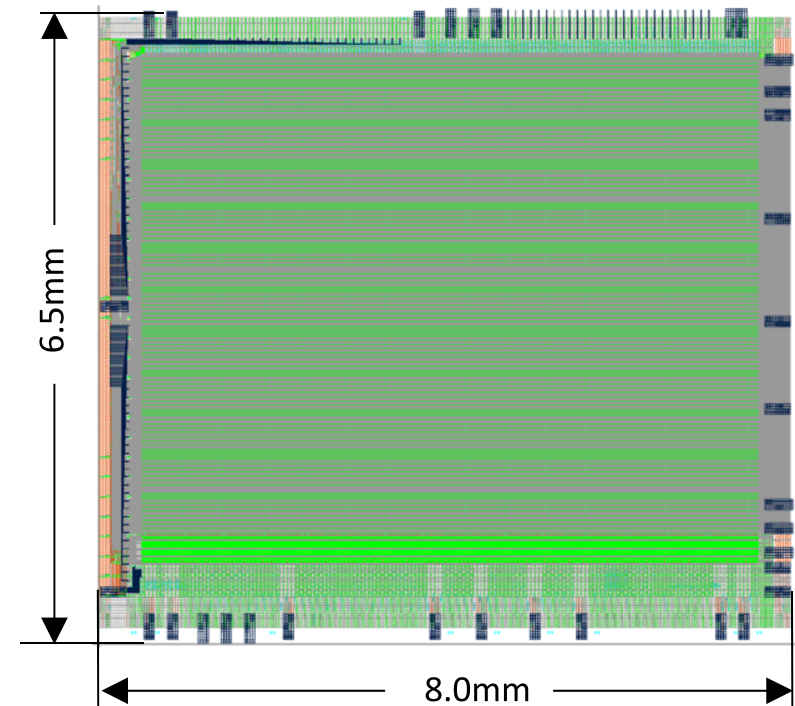
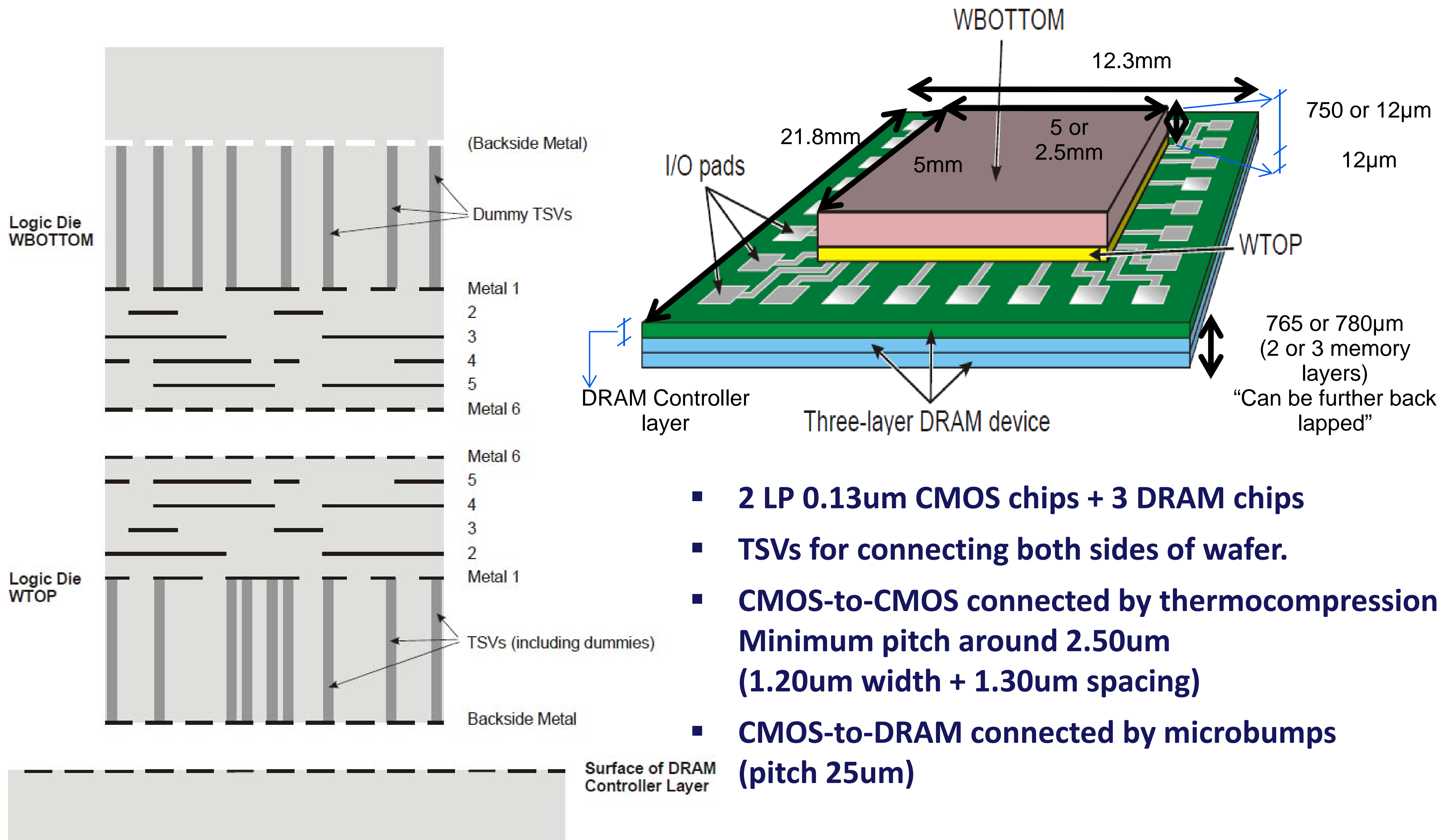


Image frame buffer

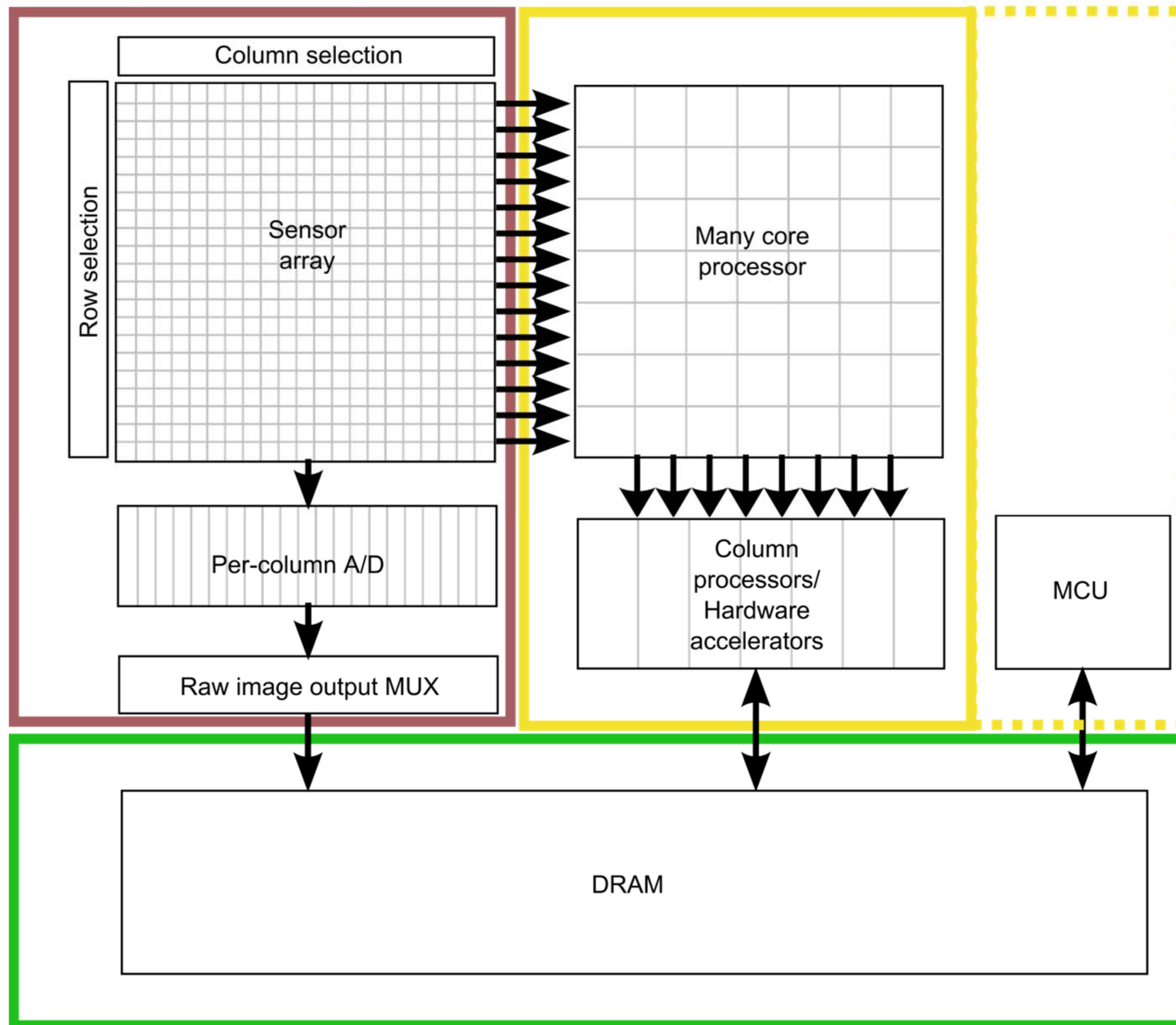
- Pitch-aligned with MS-layer (160*120 TSVs)
- Dual-port 160*120*(6x8b + 2x1b) SRAM
- 8b parallel I/O

3rd attempt: Tezzaron's 3D IC stack



- 2 LP 0.13um CMOS chips + 3 DRAM chips
- TSVs for connecting both sides of wafer.
- CMOS-to-CMOS connected by thermocompression
  Minimum pitch around 2.50um
  (1.20um width + 1.30um spacing)
- CMOS-to-DRAM connected by microbumps
  (pitch 25um)

3rd attempt: Tezzaron's 3D IC stack



Tier2 (WBOTTOM)

- 800 x 640 px.
- BSI sensors
- Global shutter
- In-pixel CDS
- Raw image ADC to DRAM

Tier1 (WTOP)

- Smaller image size
- Bidirectional access to memory (foveation)
- Anisotropic diffusion
- Gaussian and Laplacian pyramids
- Min/Max detection
- Operation control?

Conclusions

- Conventional data processing architectures introduce **data bottlenecks** and are inefficient when dealing with **multidimensional sensory signals**

- Architectures **adapted** to the nature of the stimulus are **more efficient** in terms of power consumption per operation

- Concurrent sensing, processing and memory in **planar technologies** introduces serious **limitations to image resolution** and **image size** via the penalties in fill factor and pixel pitch

- **3D integrated circuit technologies** with a dense TSV distribution permits **eliminating data bottlenecks** without degrading image resolution and size.

Acknowledgments

This work is financially supported by Andalusian Regional Government, through project 2006-TIC-2352, the Spanish Ministry of Economy and Competitiveness, through projects TEC 2009-11812 and IPT-2011-1625-430000, both co-funded by the EU-ERDF and by the Office of Naval Research (USA), through grant N000141110312.

Project Part-Financed by the European Union

European Regional Development Fund