



Workshop on the Architecture of Smart Cameras,  
27-28 June 2018, Coimbra, Portugal



Advanced Imaging S.L.

# Smart CMOS Image Sensors

Renato Turchetta

[renato.turchetta@imasenic.com](mailto:renato.turchetta@imasenic.com)

Barcelona, Spain

# The silicon retina

imasenic

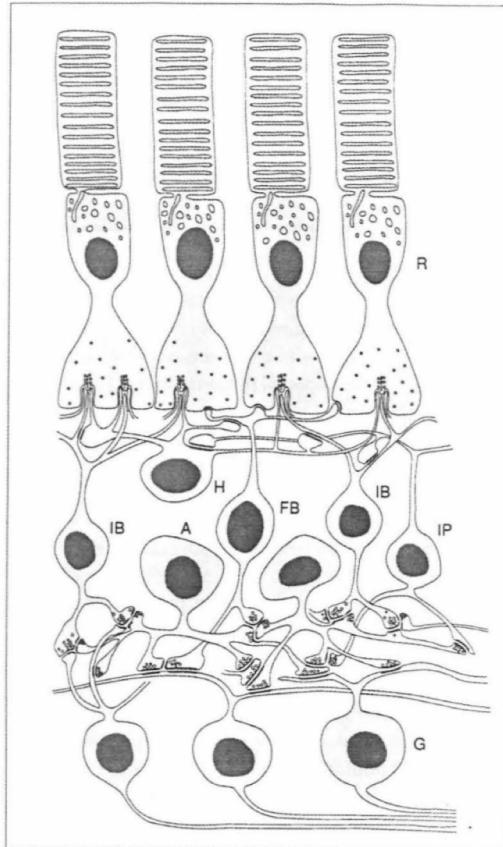


Figure 1.

Cross section through the biological retina. R, photoreceptor; H, horizontal cell; IB, invaginating bipolar cell; FB, flat bipolar cell; A, amacrine cell; IP, interplexiform cell; G, ganglion cell. The outer-plexiform layer is beneath the foot of the photoreceptors. The invagination into the foot of the photoreceptor is the site of the triad synapse. In the center of the invagination is a bipolar cell process, flanked by two horizontal cell processes.

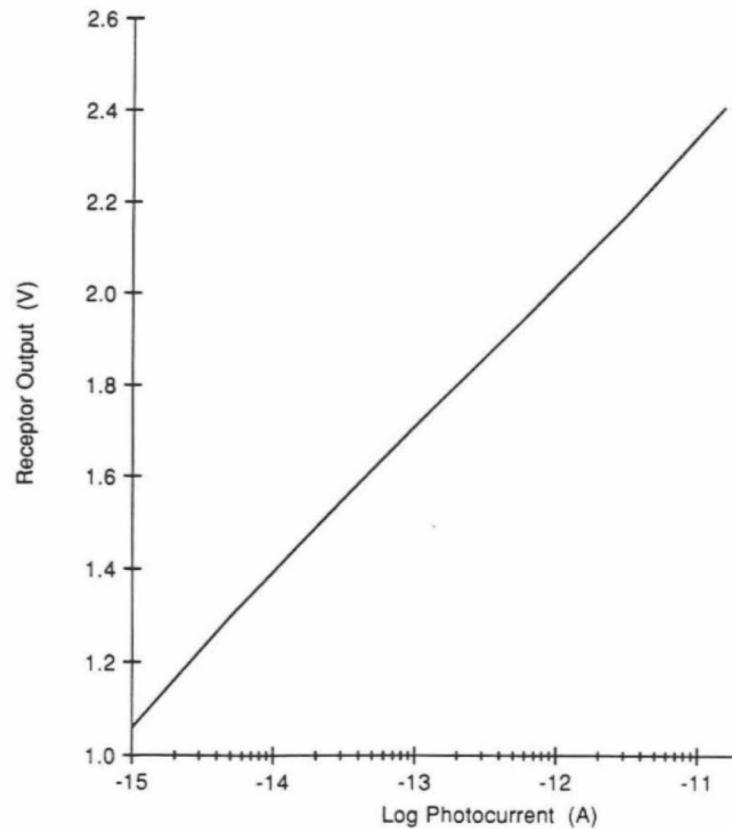


Figure 2.

Measured response of logarithmic photodetector. Photocurrent is proportional to incident light intensity. Response is logarithmic over more than four orders of magnitude in intensity.

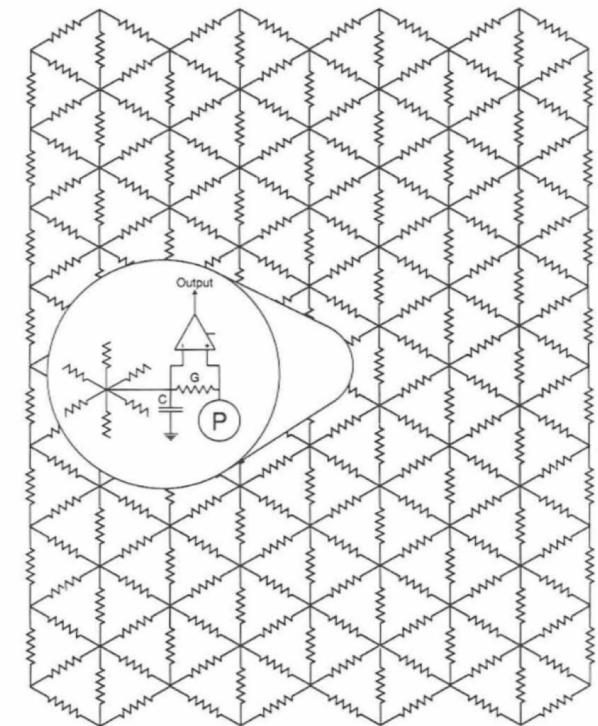
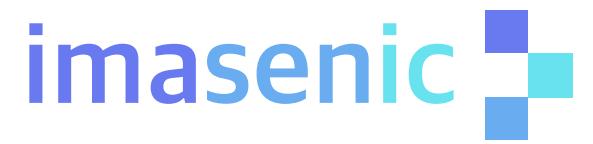


Figure 3.

The silicon retina. Diagram of the resistive network and a single pixel element, shown in the circular window. The silicon model of the triad synapse consists of the conductance ( $G$ ) by which the photoreceptor drives the resistive network, and the amplifier that takes the difference between the photoreceptor ( $P$ ) output and the voltage on the resistive network. In addition to a triad synapse, each pixel contains six resistors and a capacitor  $C$  that represents the parasitic capacitance of the resistive network. These pixels are tiled in a hexagonal array. The resistive network results from a hexagonal tiling of pixels.

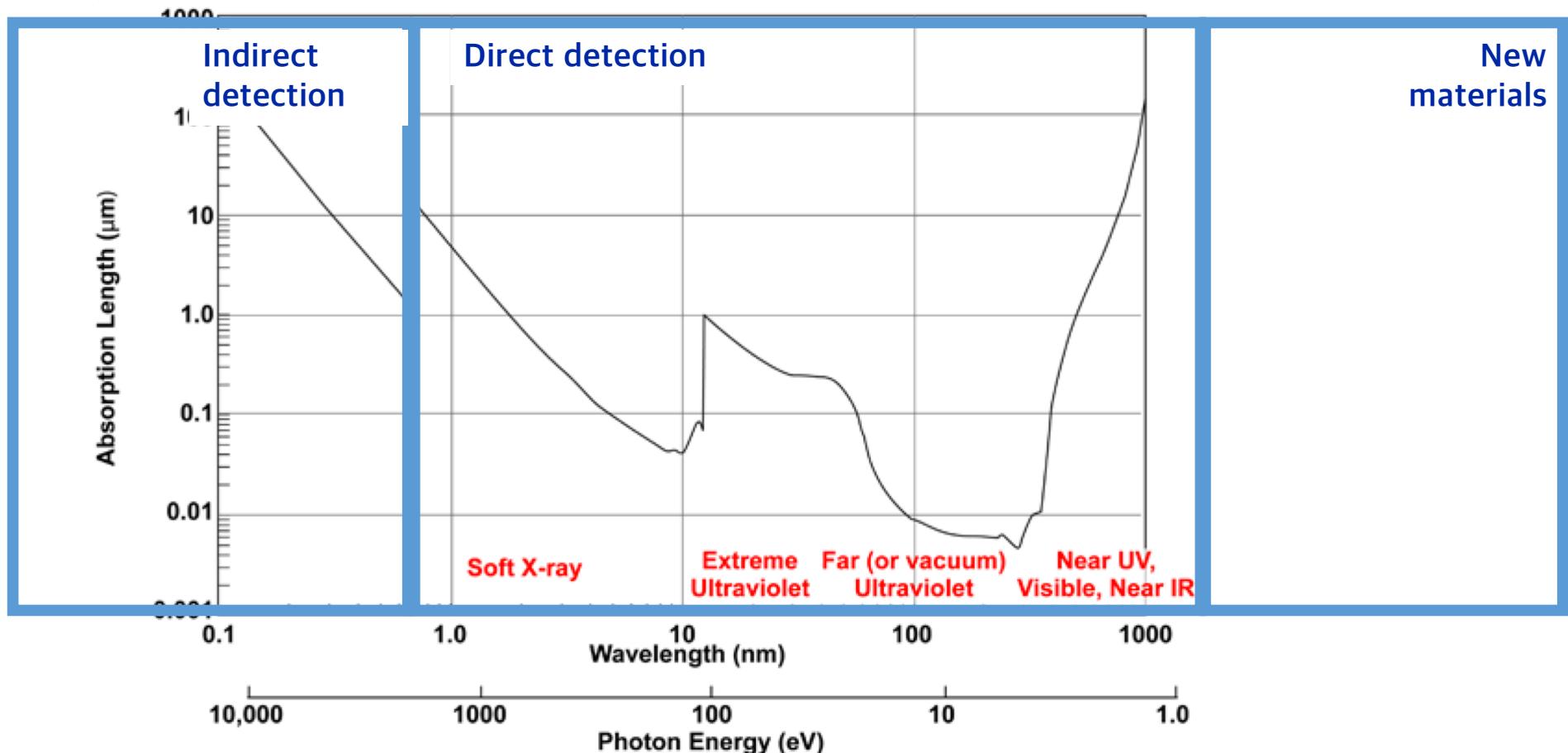


# Introduction



# Photons absorption

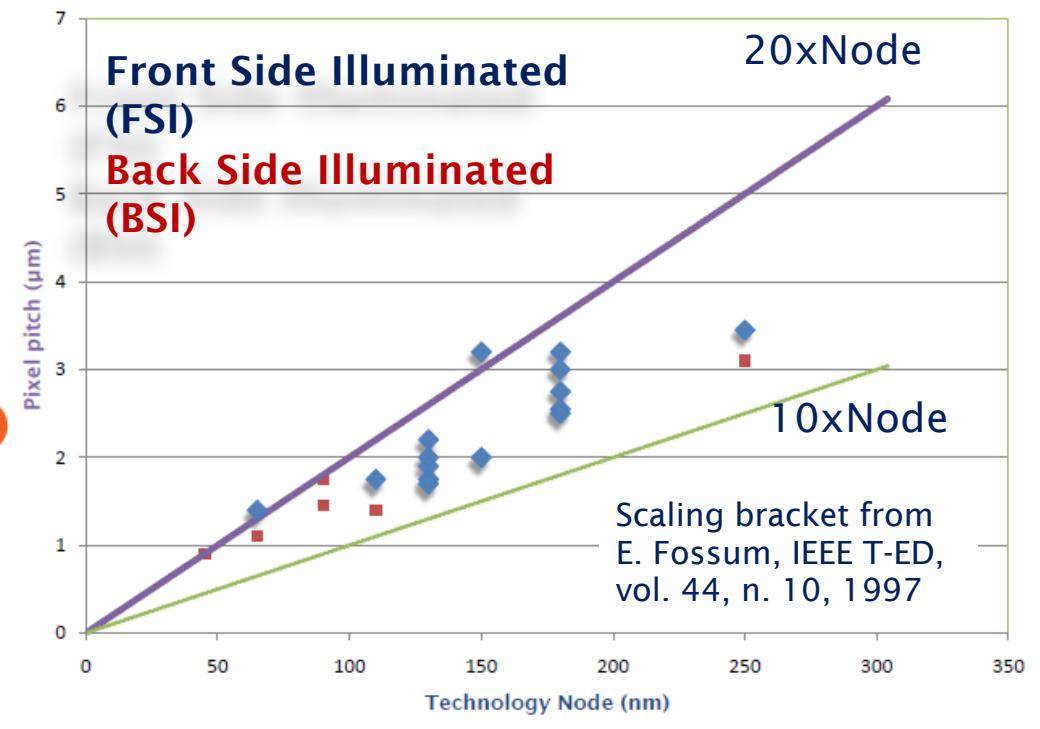
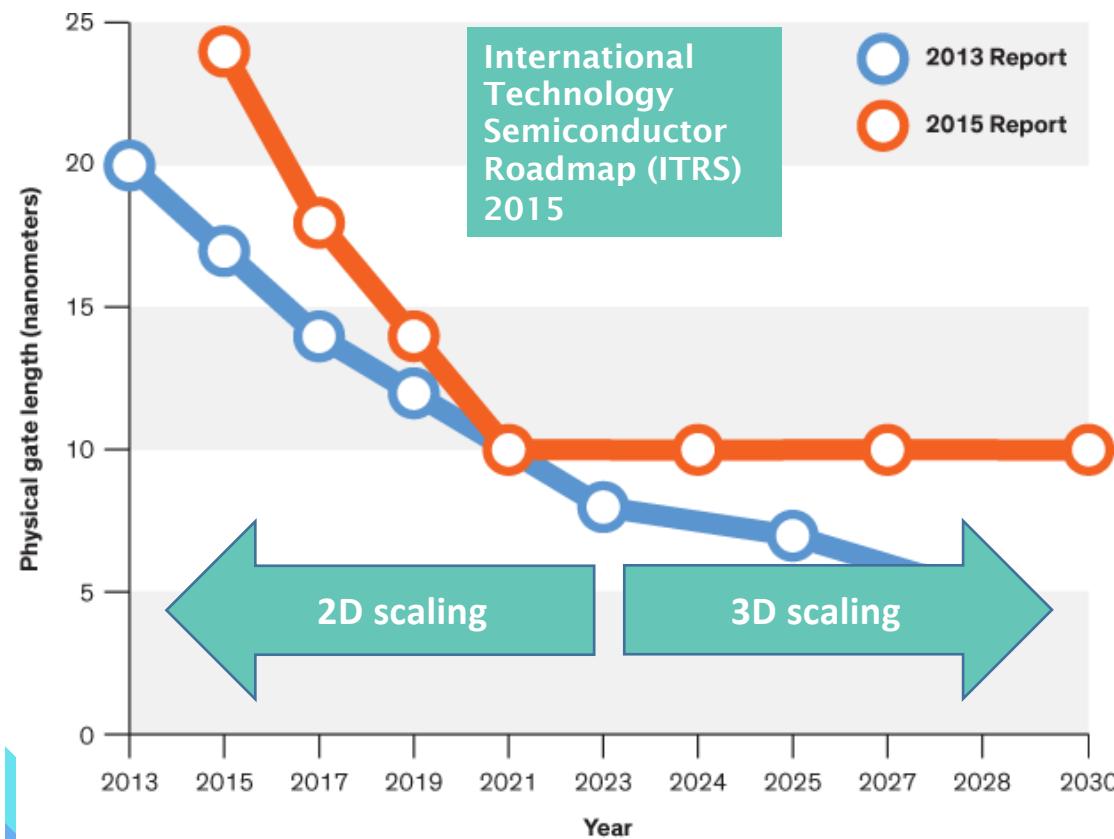
imasenic 



Silicon cut-off at  $1.1 \mu\text{m}$

# Moore's law and pixel scaling

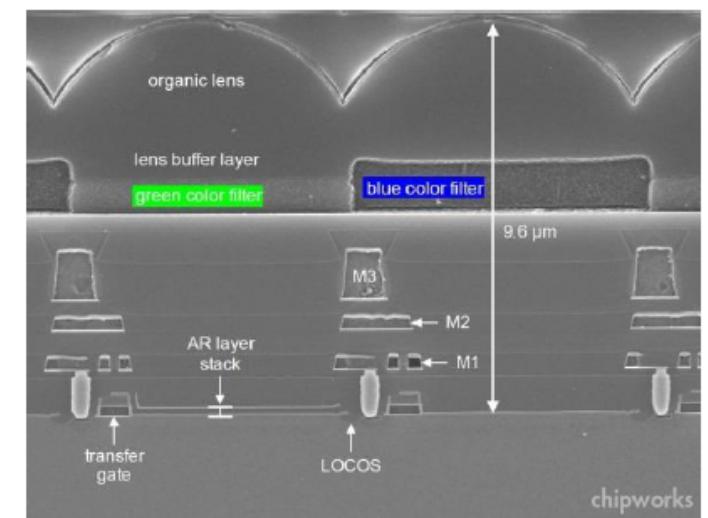
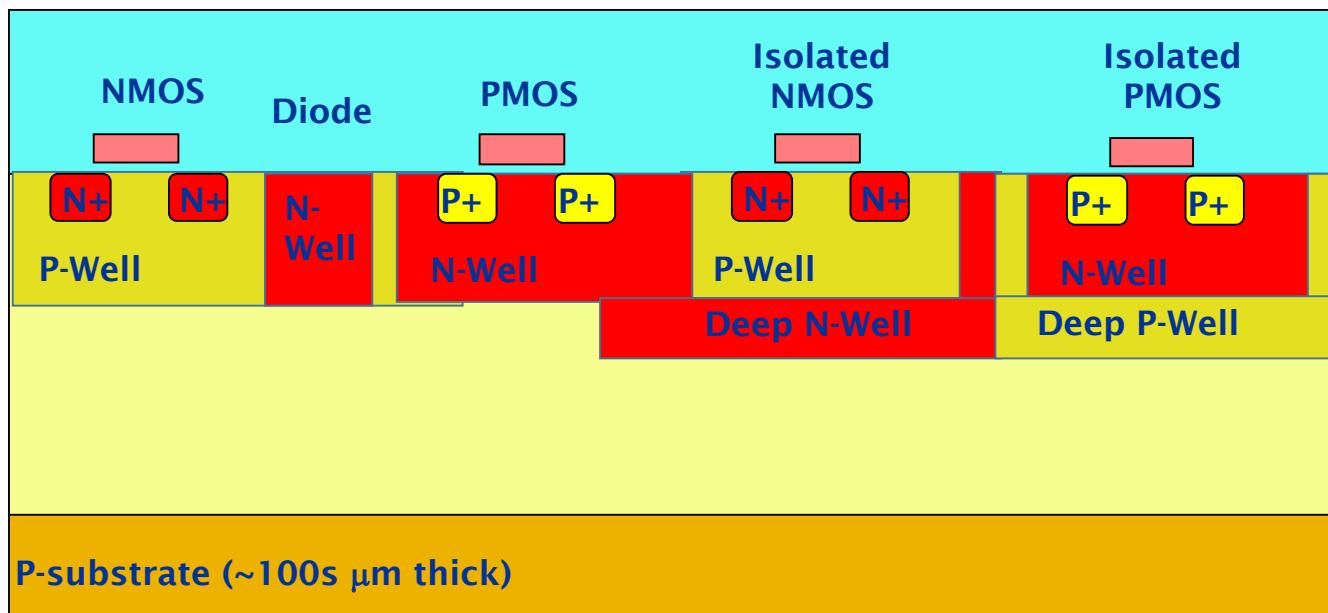
imasenic



Adapted from S. Wuu (TSMC) et al., 2009 IISW, June 2009

# CMOS process cross-section

imasenic 

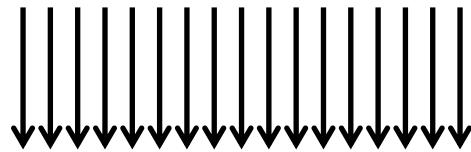


R. Fontaine: Innovative technology  
elements for small and large CIS pixels,  
IISW 2013

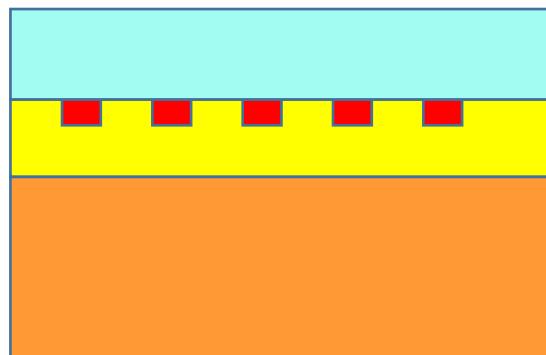
# Front- vs Back-side illumination



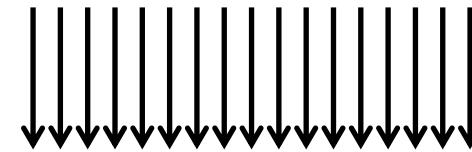
**Front Side Illumination (FSI)**



**SiO<sub>2</sub>+interconnect**  
**Epitaxial layer**  
**Substrate**

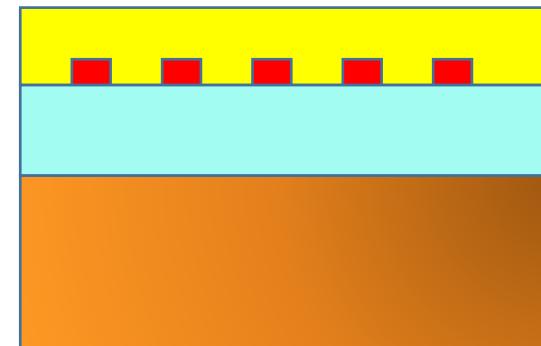


**Back Side Illumination (BSI)**



Used for improving quantum efficiency of shallow penetrating particles / photons, e.g. low energy electrons / X-rays, EUV

**Epitaxial layer**  
**SiO<sub>2</sub>+interconnect**  
**Handle wafer**



# BSI processing

## 1) Mounting on handling wafer:

- glueing
- wafer bonding

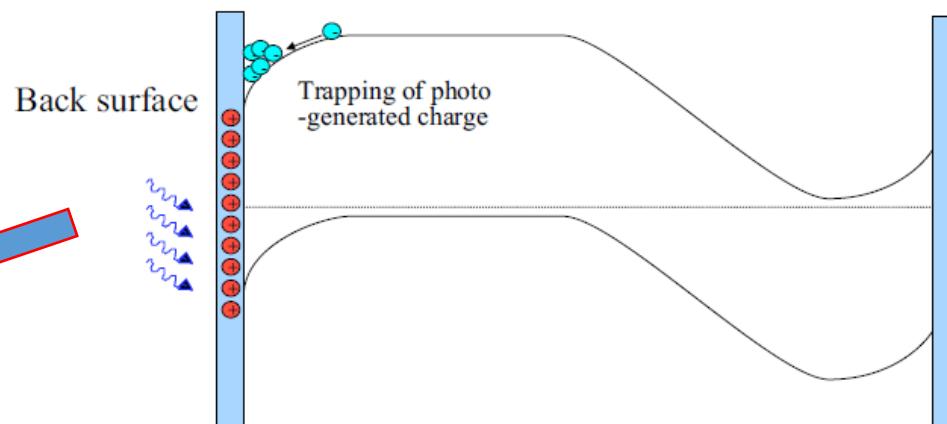
## 2) Surface annealing:

- Laser annealing
- delta doping
- Passivation (also ARC)

### BI device dilemma: Physics of Si / SiO<sub>2</sub> Interface

Si/SiO<sub>2</sub> interface charging produces spontaneous “self-bias”:

- QE hysteresis: Low and unstable low quantum efficiency.
- Back illuminated silicon detectors are useless without surface passivation.



Copyright 2009 California Institute of Technology. Government Sponsorship Acknowledged.

# 3D integration

imasenic 

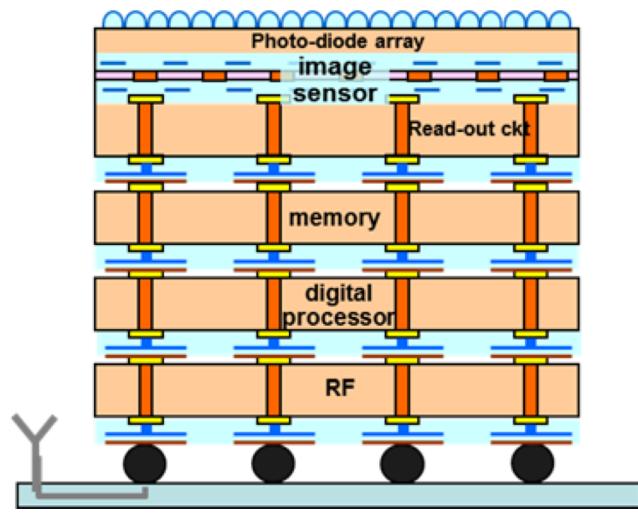
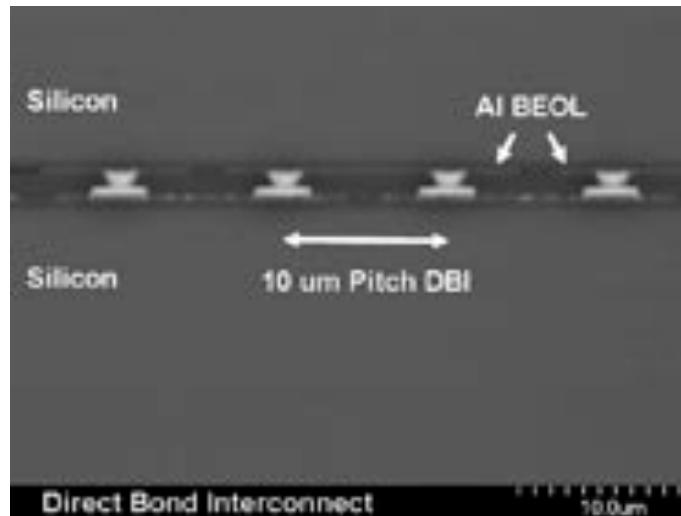


Image sensor stack

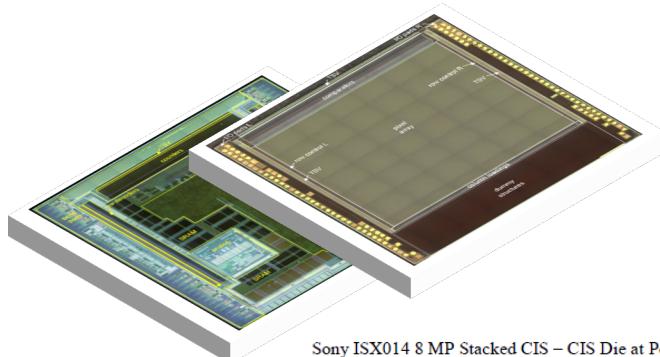


Direct Bond Interconnect (DBI)

# Stacked CIS



First sensor announced by Sony in 2012



Sony ISX014 8 MP Stacked CIS – CIS Die at Poly Level

90 nm CIS

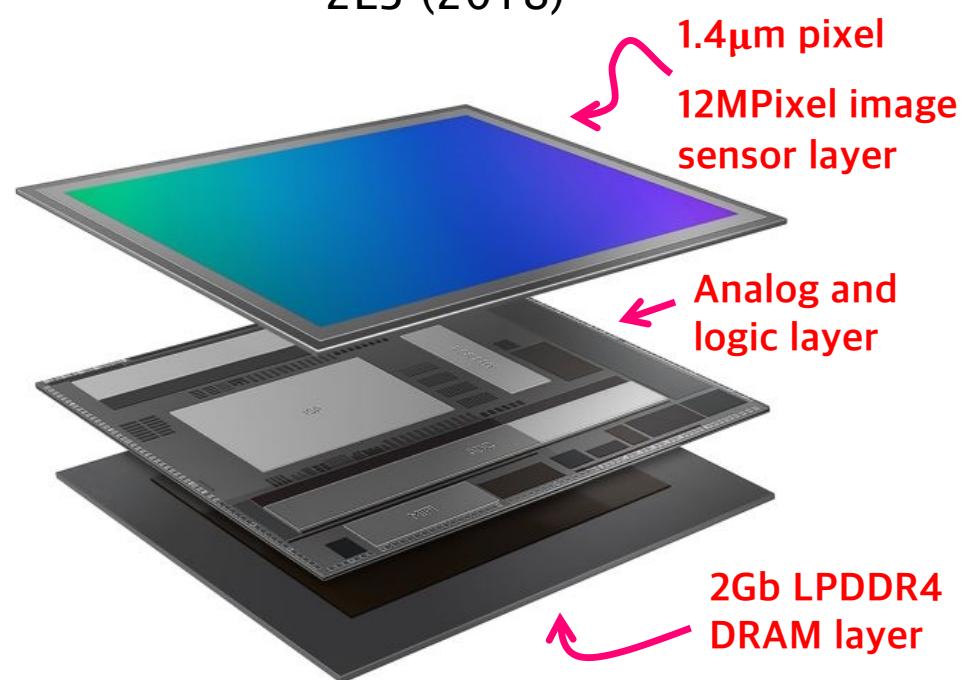
65 nm CIS

2014: Sony sensor for Apple

iPhone 6/6 Plus iSight:

90 nm Sony CIS + 40 nm TSMC

SAMSUNG ISOCELL Fast  
2L3 (2018)



Memory layer to allow recording of super-slow motion video at up to 960 frames per second (fps).



# High Dynamic Range



# Pixel limitation

Standard pixel DR is sub-optimal

Readout noise =  $O(100\mu V)$

Voltage swing =  $O(1V)$

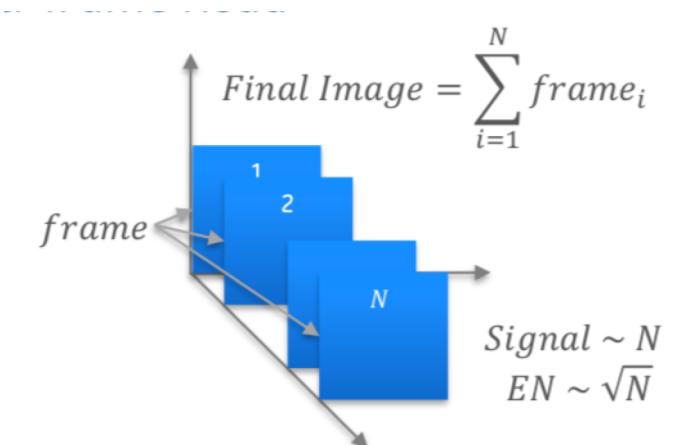
Dynamic range = Voltage swing / Readout noise  
 $= O(10^4) = O(13 \text{ bits})$

HDR without changing sensor architecture

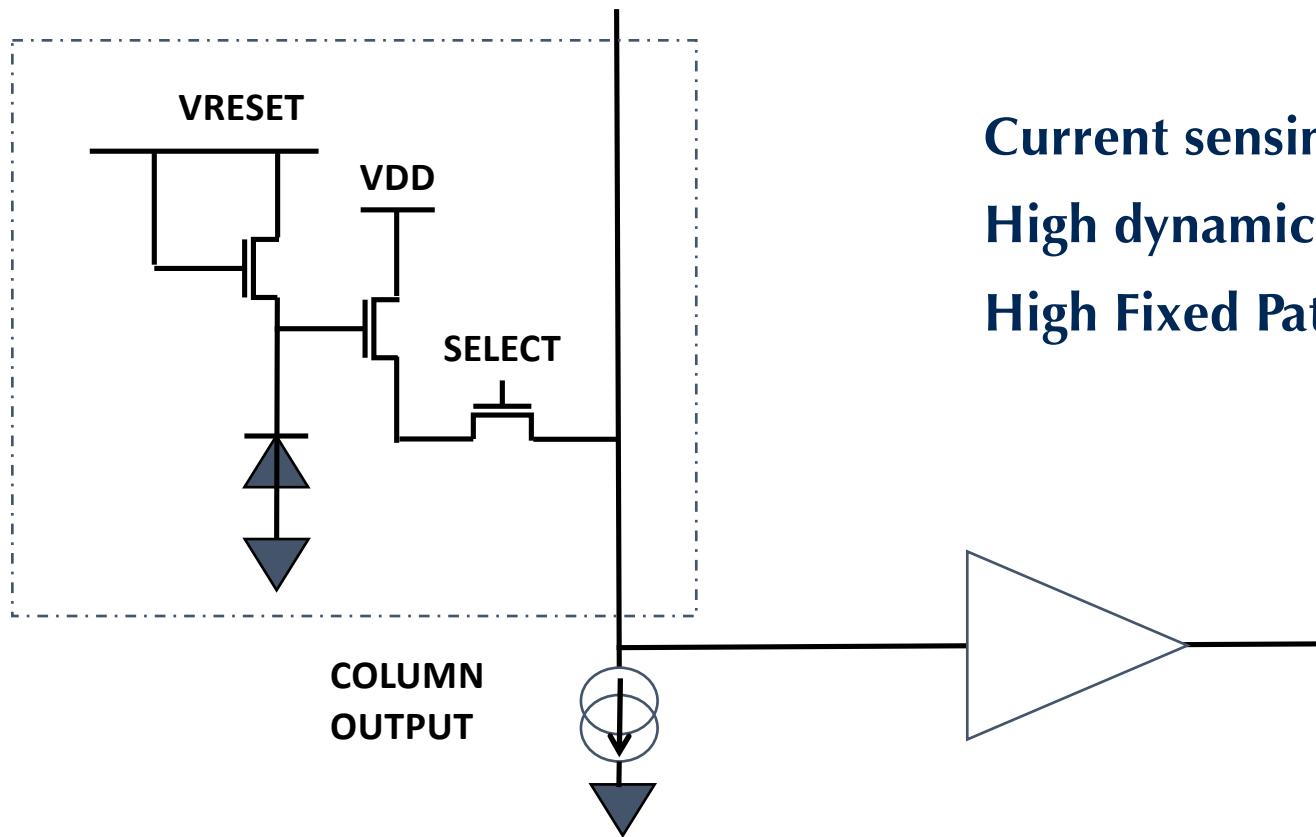
Multiple reads: reads with different integration time

- + Simple, no change to sensor architecture

- Reduce speed. Increased motion blur

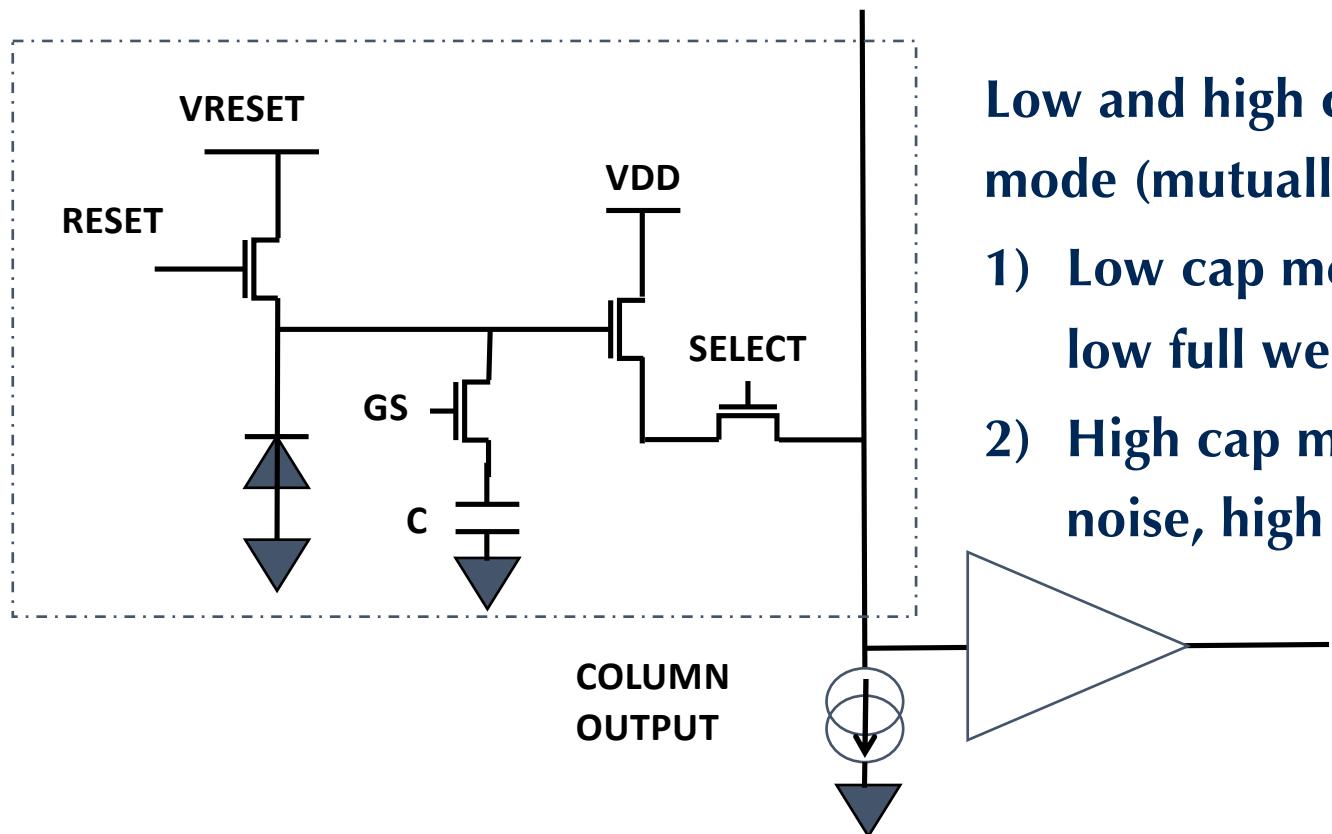


# Logarithmic pixel



**Current sensing**  
**High dynamic range**  
**High Fixed Pattern Noise (FPN)**

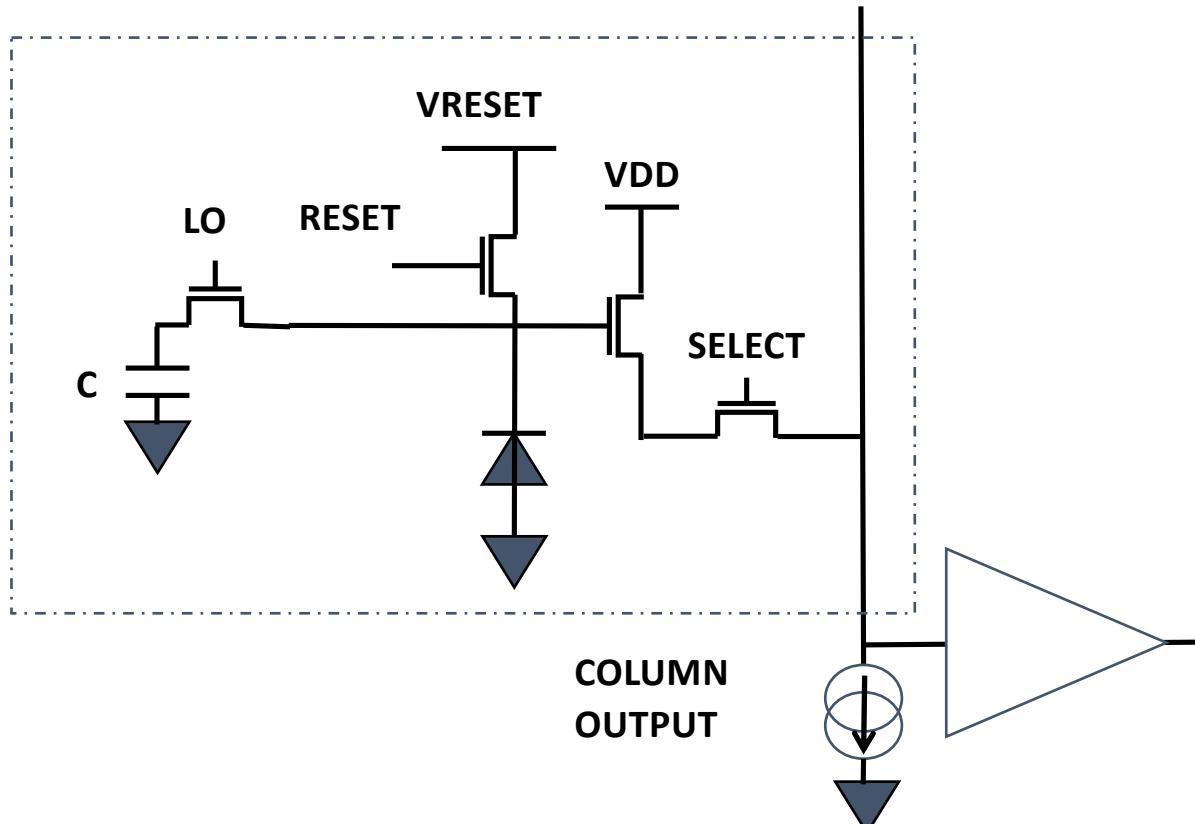
# Dual gain pixel



**Low and high capacitance mode (mutually exclusive)**

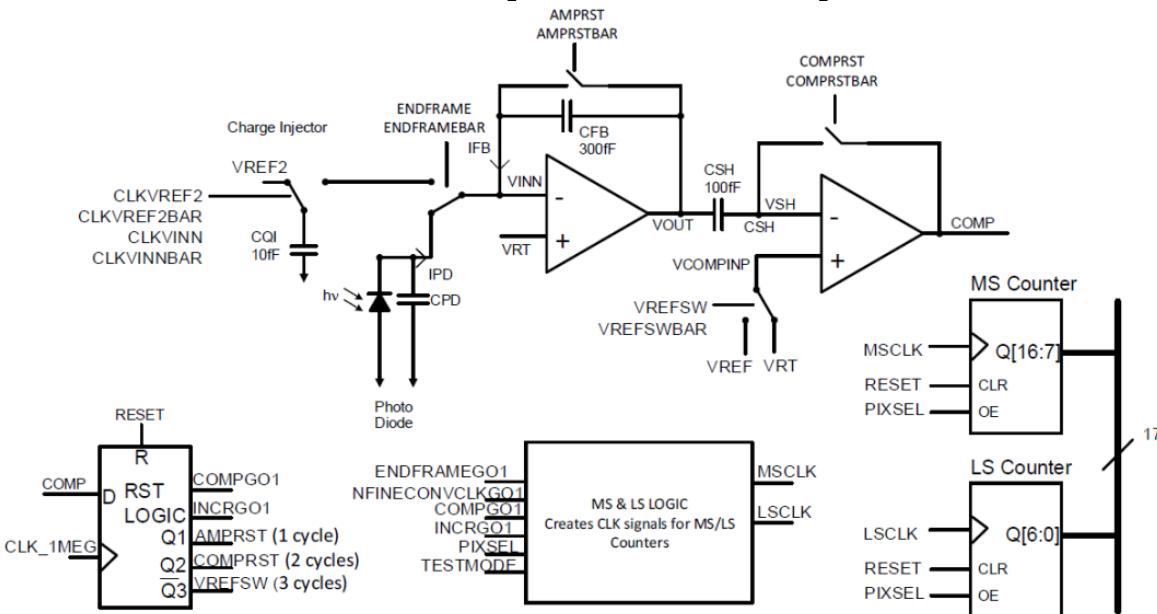
- 1) Low cap mode → low noise, low full well**
- 2) High cap mode → high noise, high full well**

# Lateral overflow



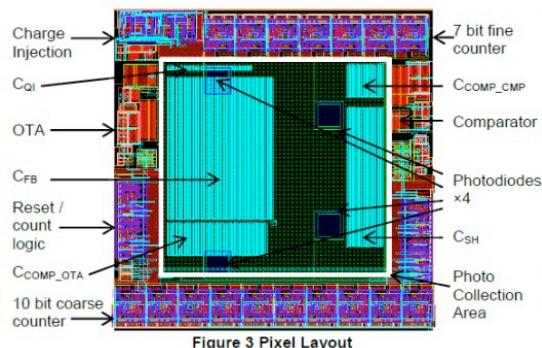
- LO biased at an intermediate level**
- Low and high capacitance mode (in the same frame)**
- 1) **Low cap mode → low noise, low full well**
  - 2) **High cap mode → high noise, high full well**

# Analogue-Digital pixel with Range Extremely Wide (ANDREW)



Type	1	2	3	4	5	6	7
Collection Node	1*	1*	1*	4*	1*	1*	4*
	7x7µm	7x7µm	4x4µm	4x4µm	4x4µm	4x4µm	4x4µm
Implants	NWell	NW+DNW	NW	NW	NW_PIX	NW_PIX	NW_P
C <sub>FB</sub>	300fF						
Fill Factor	50%	50%	50%	50%	50%	50%	50%
QE	32%	32%	26%	49%	26%	25%	51%
I <sub>DARK/Pix</sub>	127fA	128fA	116fA	169fA	114fA	112fA	181fA
Q <sub>inj</sub> Noise	2.3	2.4	2.4	2.2	2.1	2.2	2.3
Q <sub>inj</sub> Range	90	88	90	82	88	87	88
SNR <sub>Qinj</sub>	31.9dB	31.3dB	31.3dB	31.4dB	32.4dB	31.7dB	31.5dB
Ramp Noise	0.27	0.29	0.23	0.28	0.28	0.28	0.44
Ramp Range	97	97	97	97	97	97	97
SNR <sub>Ramp</sub>	51.0dB	50.5dB	52.4dB	52.1dB	50.9dB	50.8dB	46.8dE

Table 1 Photodiode Types and Pixel Performance



ST's 1P4M 65nm (BEOL) / 90nm (FEOL) BSI imaging process. The pixel contains 63 GO2 transistors for the analogue and 812 GO1 for digital. The test device has an array format of 64(H)×68(V) pixels, 50µm pixel

J. Raynor (2015)



# **Hyper-Active Pixel Sensors (HAPS)**



# Hyperactive pixel sensors

- Forget a few transistors per pixel
- Go wild and just add functionalities

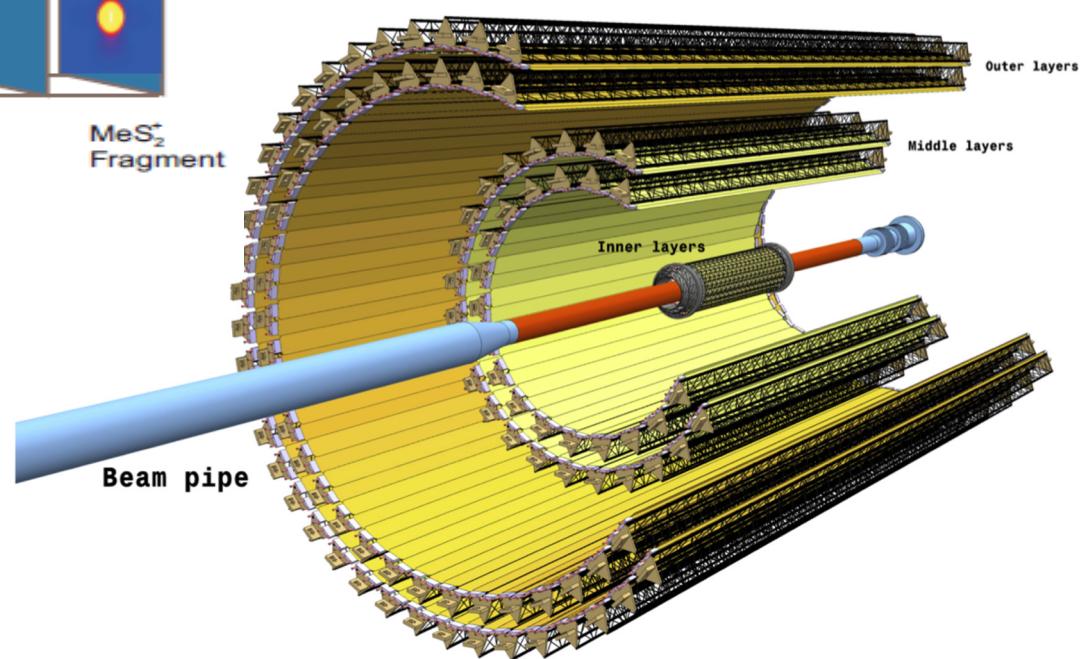
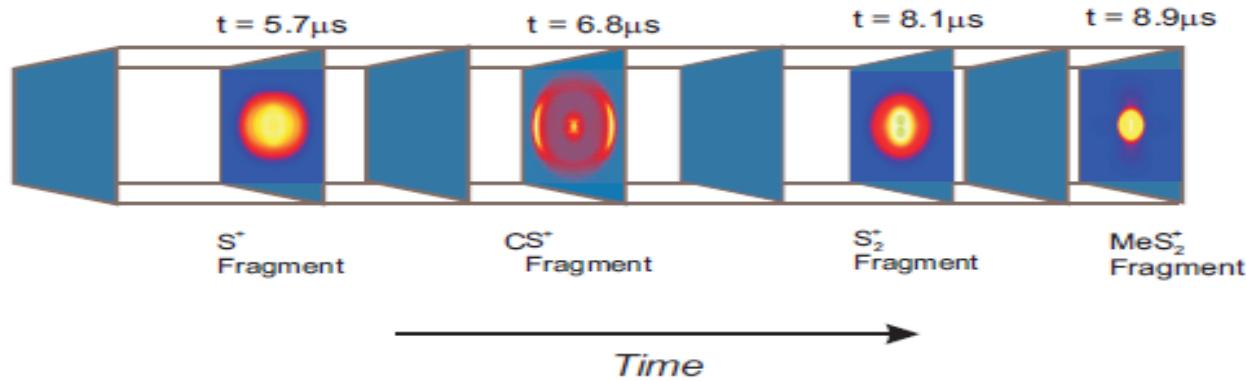
# Why ?

# Hyperactive pixel sensors

- Scientific applications
- ADC in pixels
- High Dynamic Range
- Focal plane processing
- Neuromorphic sensors

# Scientific applications

- Mass spectroscopy



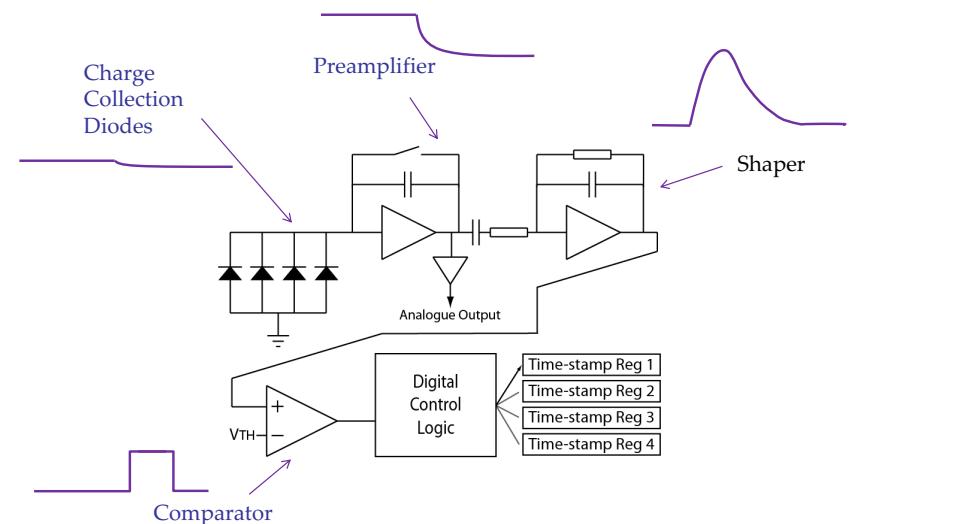
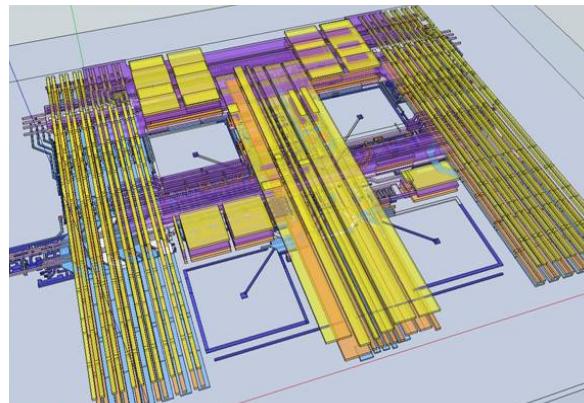
- Particle Physics

# Scientific imagers

imasenic

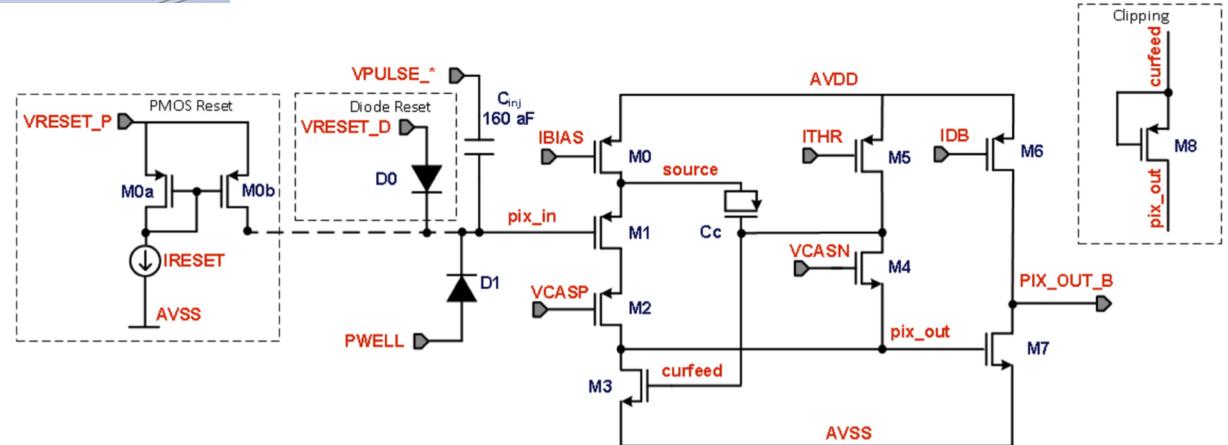
- Mass spectroscopy

(Sedgwick 2012)



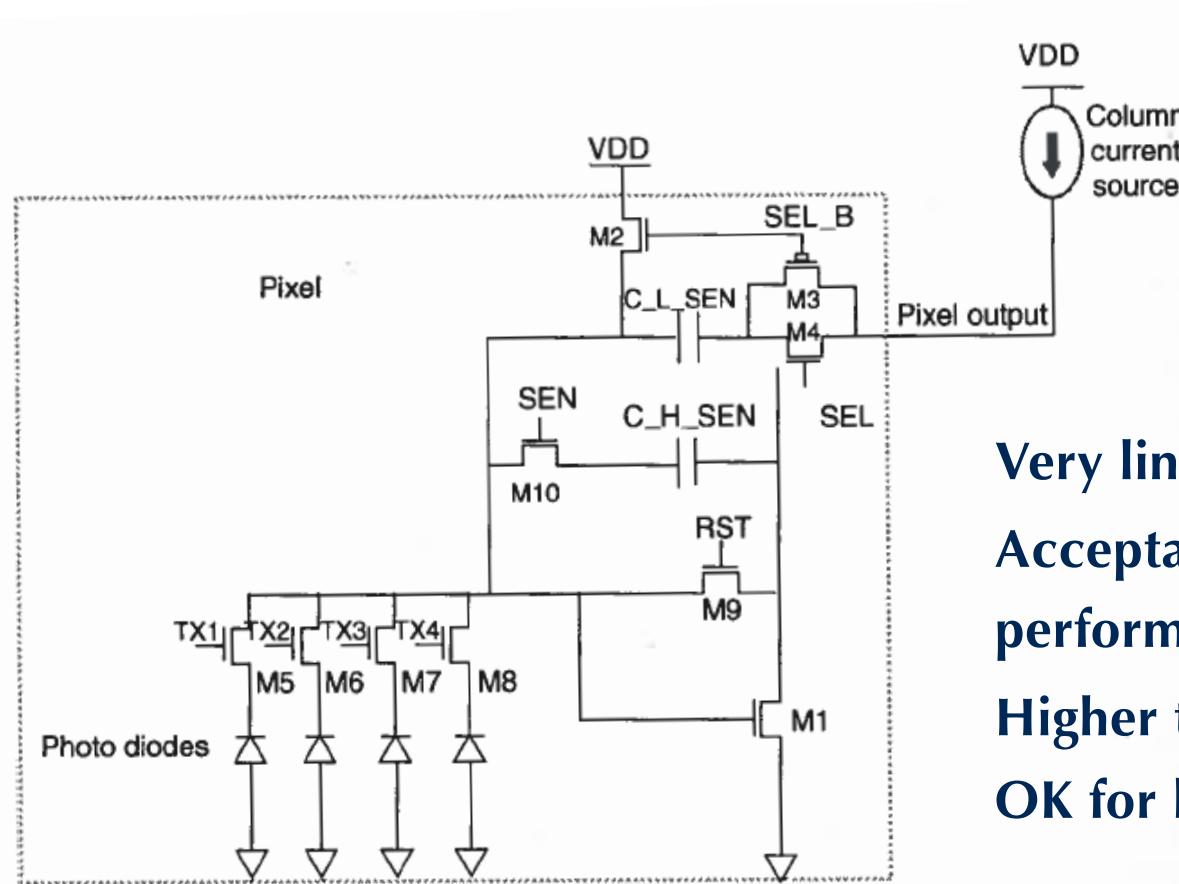
- Particle Physics

(Snoeys 2017)



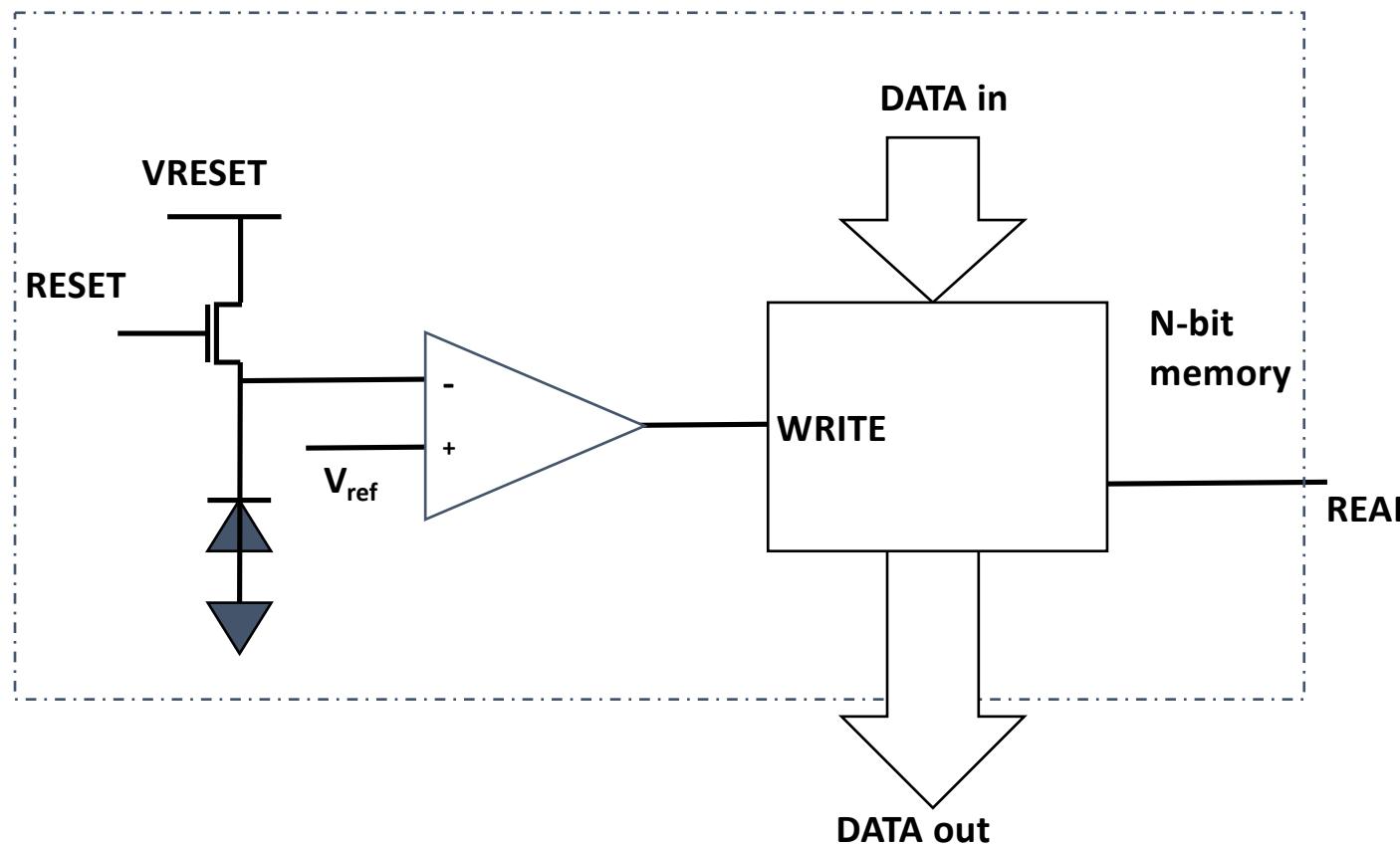
# In pixel Charge amplifier for X-ray sensors

imasenic 



**Very linear**  
**Acceptable noise**  
**performance**  
**Higher transistor count:**  
**OK for large pixels**

# In pixel ADC



- Complex pixel
- Low fill factor
- Low yield
- Ideally high speed



# Focal plane processing



# Edge extraction, centroid tracking, and high-dynamic-range

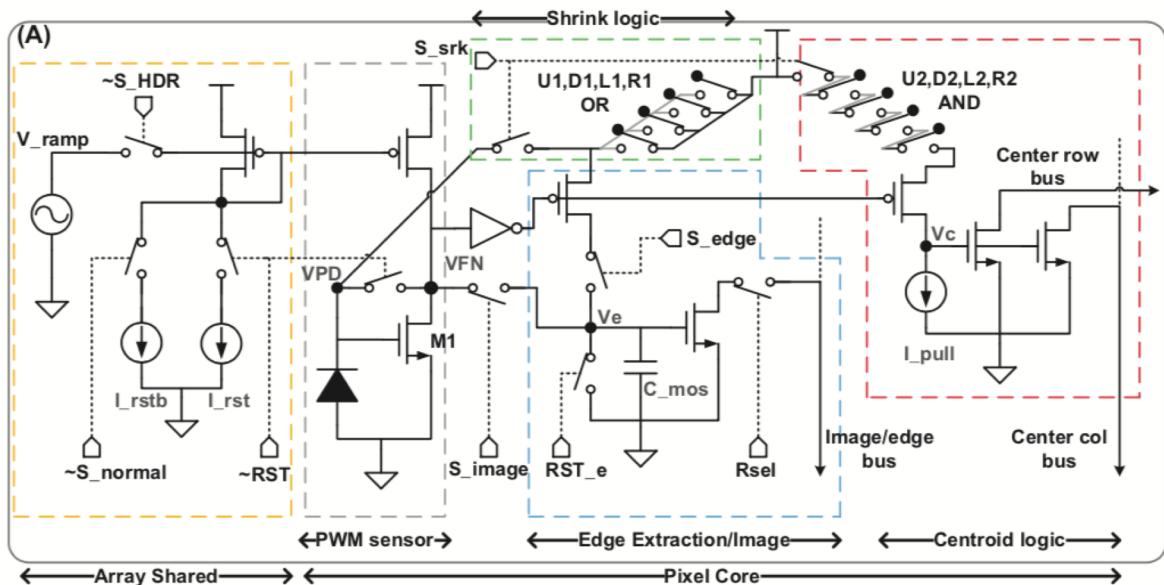


Fig. 2. Pixel core description (a) Schematic (b) Timing diagram

Function	edge filtering, tracking, HDR
Technology ( $\mu\text{m}$ )	0.18
Supply voltage (V)	0.5
Array size	$64 \times 64$
Pixel pitch ( $\mu\text{m}$ )	20
Fill factor (%)	32.4
Dynamic range (dB)	105
Power (nW/px·frame)	1.25

Q. Yin (2013)

# Gaussian pyramid

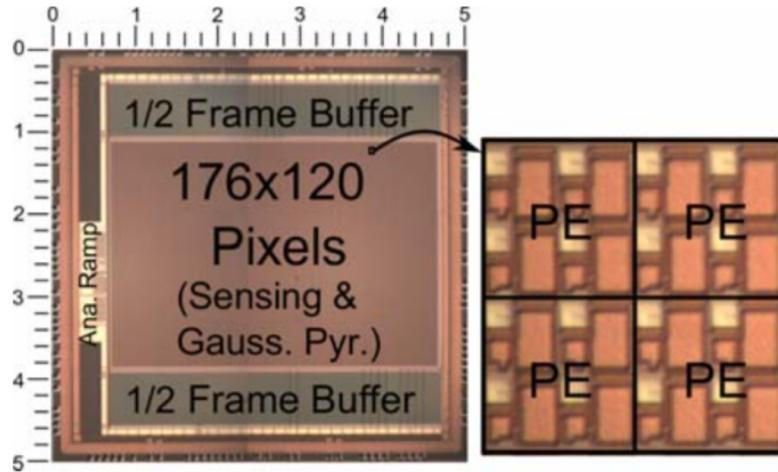


Fig. 1. Chip micrograph with dimensions (in mm) and a close-up of the PEs.

M. Suarez (2014)

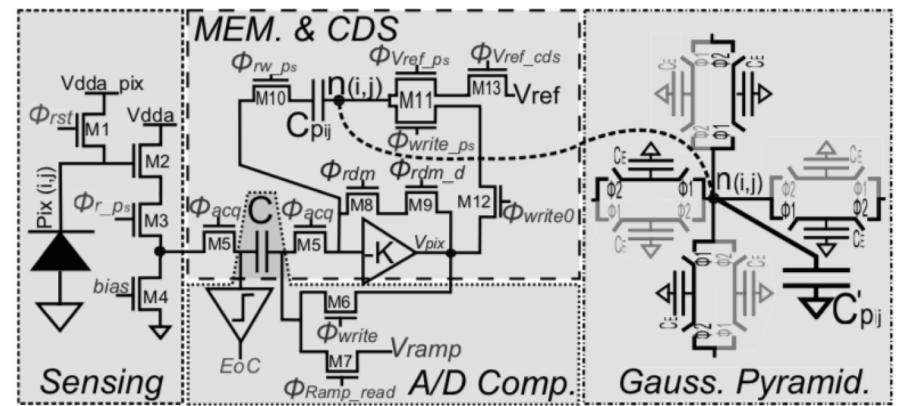
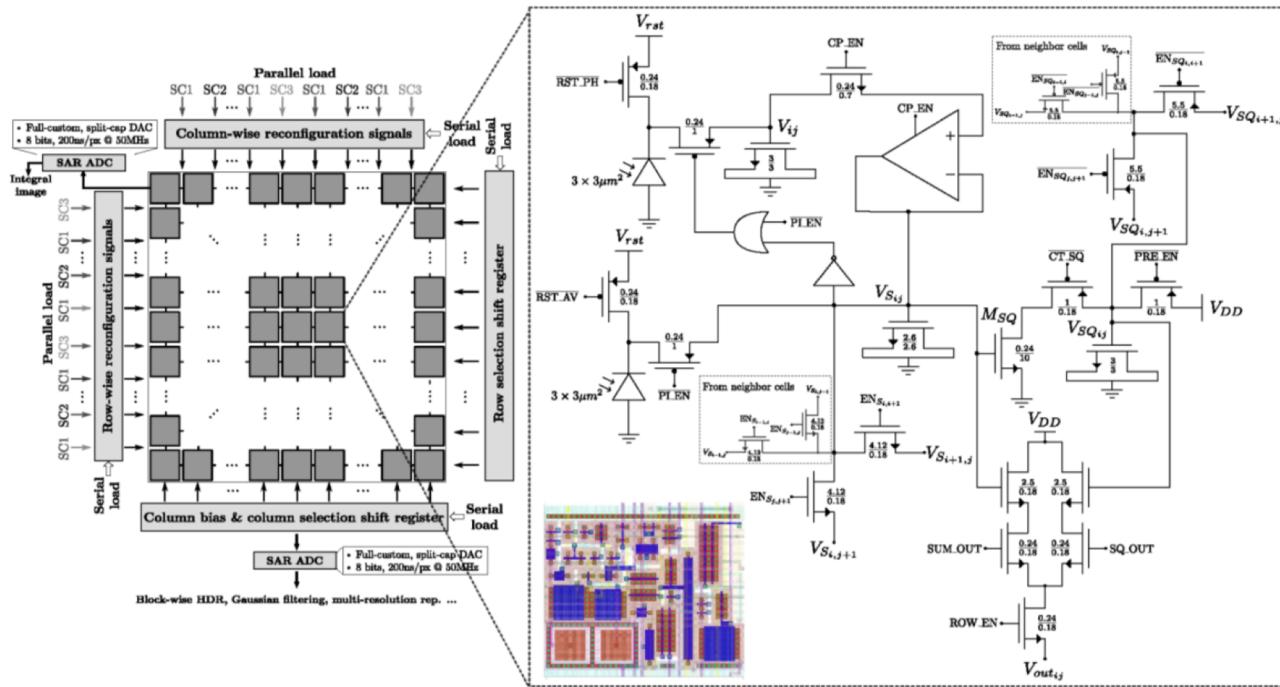


Fig. 2. Processing Element (PE) of the chip.

Function	Gaussian filtering
Technology ( $\mu\text{m}$ )	0.18
Supply voltage (V)	1.8
Array size	$176 \times 120$
Pixel pitch ( $\mu\text{m}$ )	44
Fill factor (%)	10.25
Dynamic range (dB)	—
Power (nW/px·frame)	26.5

# HDR, Gaussian filter, multi-resolution



<b>Function</b>	HDR, Gaussian filter, integral image, multiresolution
<b>Technology (μm)</b>	0.18
<b>Supply voltage (V)</b>	1.8
<b>Array size</b>	320 × 240
<b>Pixel pitch (μm)</b>	19.6
<b>Fill factor (%)</b>	5.4
<b>Dynamic range (dB)</b>	102
<b>Power (nW/px·frame)</b>	23.9

Fig. 1. Functional diagram of the chip architecture and schematic of the processing element.



# **Neuromorphic sensors**



# Neuromorphic imagers

- Inspired by the way our vision system works
- Event-driven imagers

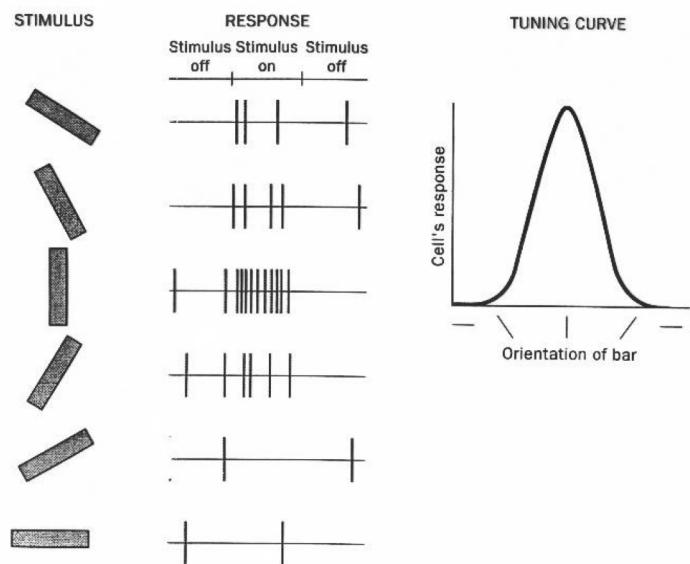
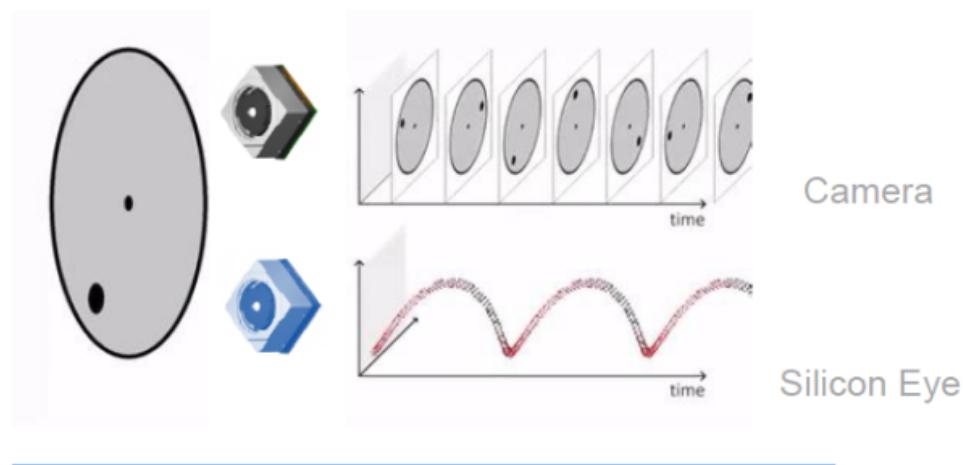


FIGURE 4.8 Response of a single cortical cell to bars presented at various orientations.

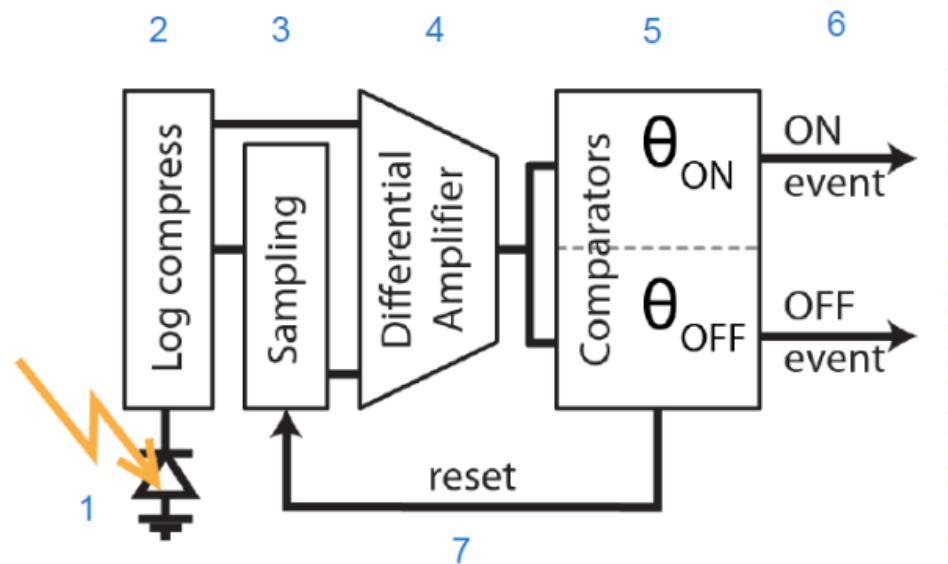


## Event-Based Vision

The Silicon Eye only reports intensity changes which results in more information encoded in less data

(Berner 2018)

# The Silicon Eye Principle



- 1) The light falling onto the pixel is converted into a photocurrent
- 2) The photocurrent is converted into a logarithmic voltage
- 3) The log. voltage is sampled
- 4) Changes in the log. voltage are amplified
- 5) The changes are compared against fix thresholds
- 6) Once the change exceeds the threshold, an event is generated
- 7) The pixel is reset and the latest log. voltage is sampled

The computation performed in the pixel is analog (no large digital memory needed), using sub-threshold currents (super low power consumption)

# Advantages



## Low Data Rate

The sensor does only produce data if new information is captured (~**2MBps**)



## High Temporal Resolution

The vision sensor captures motion with a high temporal resolution (~**100us\***)



## Low Latency

The sensor has a very fast reaction time to capture changes in a scene (**a few 100us\***)



## High Dynamic Range

The pixels may operate independently under extreme light conditions (spanning **>100dB** from dark to bright)



## Low Power

The sensor burns little power (**<70mW**) and the low data rate requires less compute power

- Large pixel (18.5μ pitch)
- Cumbersome asynchronous data interface
- Cumbersome non-standard configuration interface
- Needs FPGA to control and interface
- Jitter dependent on event rate

	Silicon Eye	Conv. Camera*
Resolution	QVGA	VGA
Temp. resolution	10'000fps	60fps
Power consumption	~70mW	320mW
Data rate	~2MB/s	27MB/s
Dynamic range	> 100dB	55dB
Optics	1/3.2 inch	1/3 inch

# History

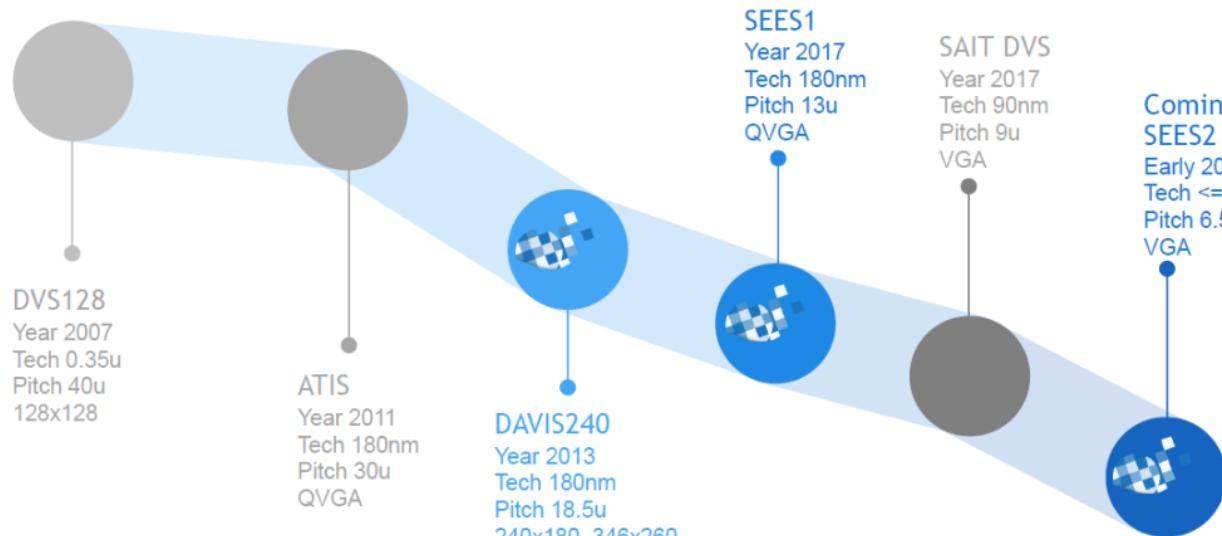
- Two main players today:
- Prophesee (formerly known as Chronocam)
- Insightness

Table 1 Comparison between AER vision sensor devices. Pixel size is given both in lambda (the scaling parameter) and um units. Power consumption is at chip and not board or system level. Best metrics are in bold. (Extended from [5].)

	Prior work						This session		
Year	2001	2003	2005	2006	2008	2009	2010	2010	2010
Source	Zaghoul, Boahen [30]	Ruedi et al. [16]	Mallik et al.[9, 32]	Lichtsteiner et al. [5, 33]	Massari et al. [27]	Ruedi et al.[31]	Posch et al. 2010 [34]	Linares- Barranco et al. [2]	Culurciello et al. [3]
Functionality	Asynchronous spatial and temporal contrast,	Frame-based spatial contrast and gradient direction, ordered output	Temporal frame- difference intensity change detection APS imager	Asynchronous temporal contrast dynamic vision sensor (DVS)	Binary spatial and temporal contrast	Digital log pixel + RISC proc.	Async. Time-based Image Sensor (ATIS) proc.	Async. Weber Contrast (SC), with either rate or ITFS coding	Temporal intensity change or spatial difference can trigger readout
Type (Sec.3)	SC TD AE	SC FE	TD FE	TC AE	SD TD FE	SC, embedded	TC AE	SC AE	TD SD FE
Gray picture output			•			•	•	•	•
Pixel size um (lambda)	34x40 (170x200)	69x69 (276x276)	25x25 <b>100x100</b>	40x40 (200x200)	26x26.5 (130x130)	<b>14x14</b> (311x311)	30x30 (333x333)	80x80 (400x400)	16x21 (?)
Fill factor (%)	14%	9%	17%	8.1%	20%	20%	10%(TC)/20%(gray)	2.5%	<b>42%</b>
Fabrication process	0.35um 4M 2P	0.5um 3M 2P	0.5um 3M 2P	0.35um 4M 2P	0.35um 4M 2P	<b>180nm</b> 1P6M	<b>180nm</b> 4M 2P MIM	0.35um 4M 2P	180nm SiGe BiCMOS 7M
Pixel complexity $T=MOS, C=cap$	38T	>50T, 1C	<b>6T (NMOS)</b> <b>2C</b>	26T(14 anal), 3C	45T	-80T, 1C	77T, 4C, 2PD	131T, 2C	<b>11T</b>
Array size	96x60	128x128	90x90	128x128	128x64	<b>320x240</b>	304x240	32x32	128x128
Die size mm <sup>2</sup>	3.5x3.5	~10x10	3x3	6x6.3	11	5.2x8.4	9.9x8.2	2.5x2.6	??
Power consumption	62.7mW @ 3.3V	300mW @ 3.3V	30mW @ 5V (50 fps)	24mW @ 3.3V	<b>100uW@2V, 50fps</b>	<b>80mW</b> (11mW sensor)	50-175mW	0.66- 6.6mW	< <b>1.4mW</b> @3V
Dynamic range	~50dB	120dB	51dB	120dB 2lux to >100 klux scene	100dB	<b>132dB</b> 6V/lux s 39dB SNR	<b>143dB</b> (static) 125dB@30FPS 56dB SNR	100dB 1lx to 100klx @550nm, 1.14uV/e <sup>-</sup>	2V/s/(uW/cm <sup>2</sup> ) @550nm, 1.14uV/e <sup>-</sup>
PD dark current@25C	?	300fA	?	4fA (~10nA/cm <sup>2</sup> )	NA	44mV/s	1.6nA/cm <sup>2</sup>	NA	
Response latency, frames/sec (fps), events/sec (eps)	~10Meps	< 2ms 60 to 500 fps	< 5ms? 200 fps?	15μs @ 1 klux chip illumination 2Meps	Max 4000fps	30fps	<b>3.2us</b> @1klux 30Meps peak, 6Meps sustained	100us @50klx 66Meps	200-800fps dep. on mode 13Meps
FPN matching	1-2 decades	2% contrast	<b>0.5%</b> of full scale, 2.1% TD change	2.1% contrast	10% contrast	<b>0.8%</b>		0.87% contrast	

(Delbruck 2010)

# Evolution of pixel pitch



- The output data of event-based vision sensors is very different from the data of image sensors
- Standard computer-vision algorithms cannot be used
- New algorithms have to be developed, and our customers won't do that
- To commercialize EBVS, we need to deliver sensors and the according algorithms

Resolution	320 x 262
Pixel pitch	13μm
Optical format	1/3.2 inch
Die size	5.3mm x 5.3mm
Technology	180nm 6M1P
Temporal resolution	Up to 10kHz
Sensitivity	Down to 50% contrast
Dynamic range Events	>100 dB (1)
Min. latency	100μs
Latency 100 lux	<1ms (2)
Readout bandwidth	>50 Meps
Power nominal	80 mW (3)
Low power mode (1kHz temp. res., no gray scale)	<20 mW

# DAVIS

imasenic

2334

IEEE JOURNAL OF SOLID-STATE CIRCUITS, VOL. 49, NO. 10, OCTOBER 2014

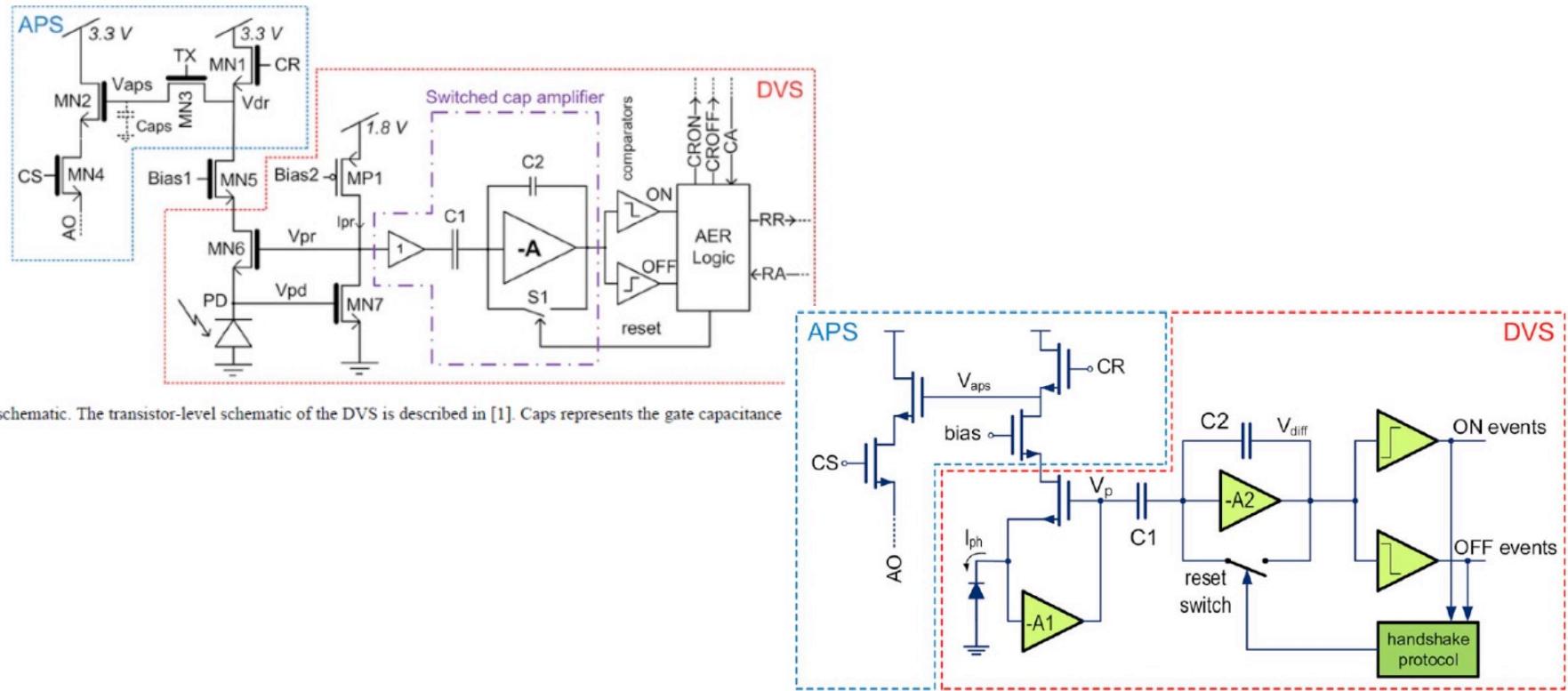


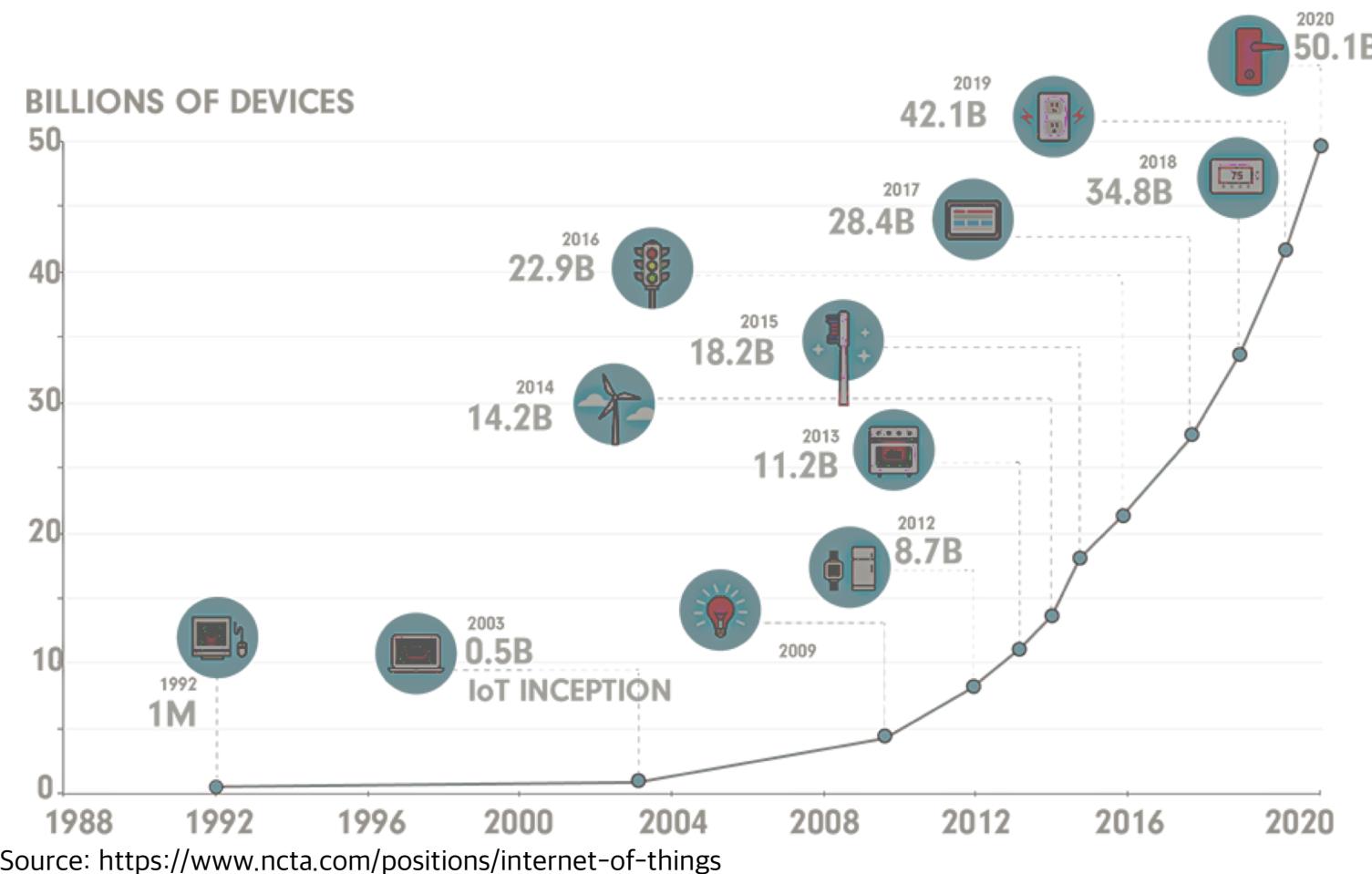
Fig. 1. DAVIS pixel schematic. The transistor-level schematic of the DVS is described in [1]. Caps represents the gate capacitance



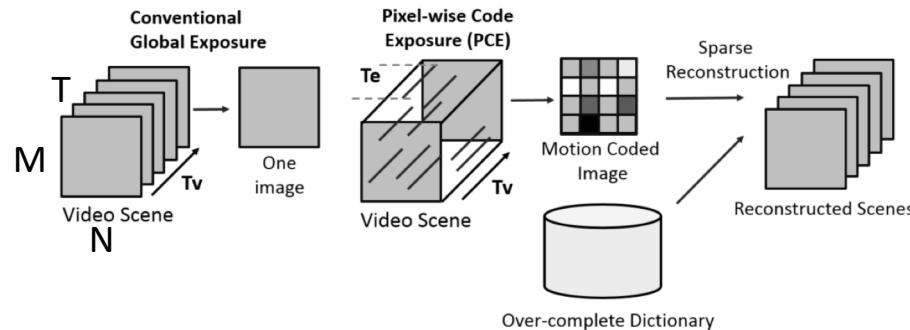
# Compressive sensing



# The Data challenge



# Spatio-temporal compressive sensing

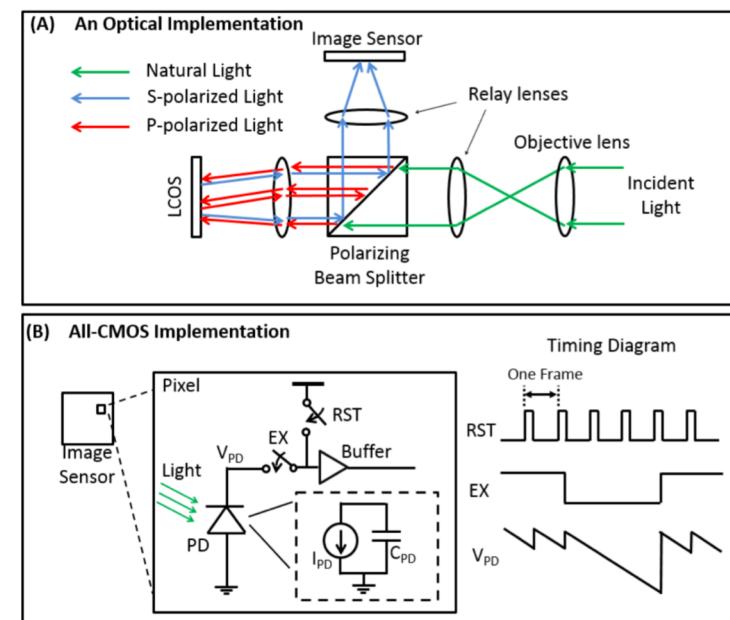


$$\mathbf{X} \in \mathbb{R}^{M \times N \times T}$$

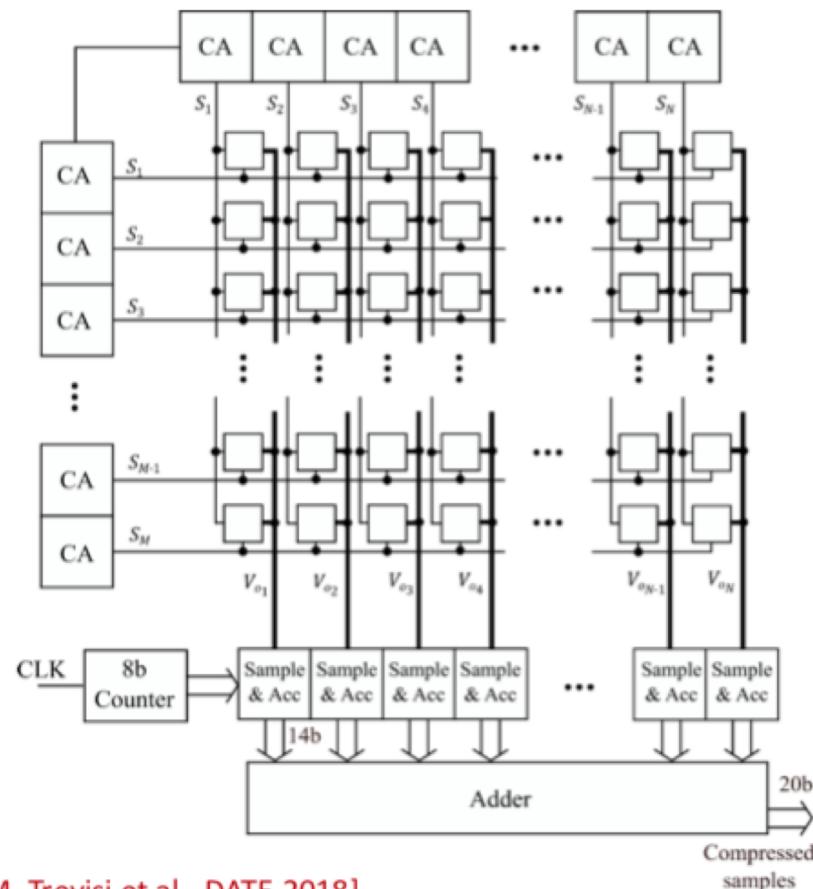
$$\mathbf{S} \in \mathbb{R}^{M \times N \times T}$$

$$\mathbf{Y}(m, n) = \sum_{t=1}^T \mathbf{S}(m, n, t) \cdot \mathbf{X}(m, n, t)$$

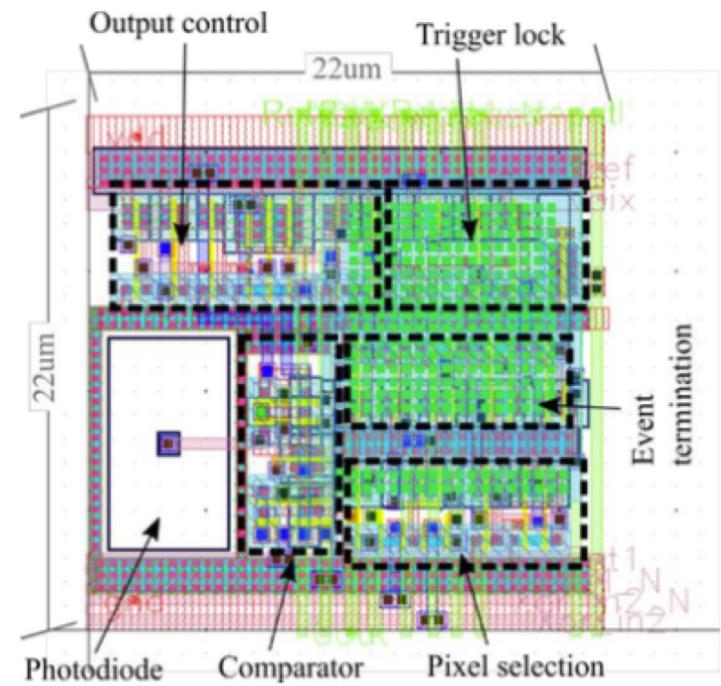
$$\hat{\mathbf{X}}(m, n) = \operatorname{argmin}_a \| a \|_0 \text{ s.t. } \| Y - S \mathbf{D} a \|_2 \leq \varepsilon$$



# Compressive sensing

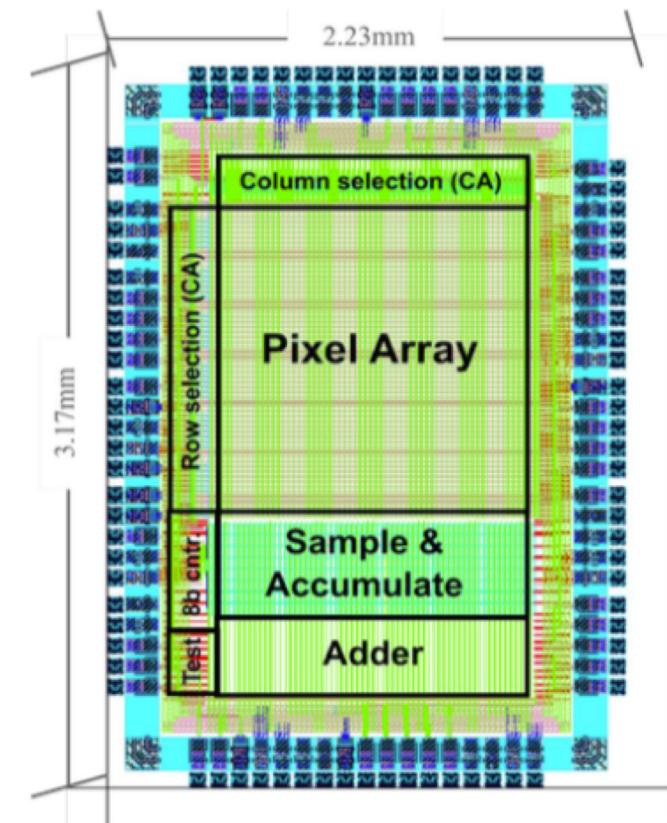


[M. Trevisi et al., DATE 2018]



# Compressive sensing

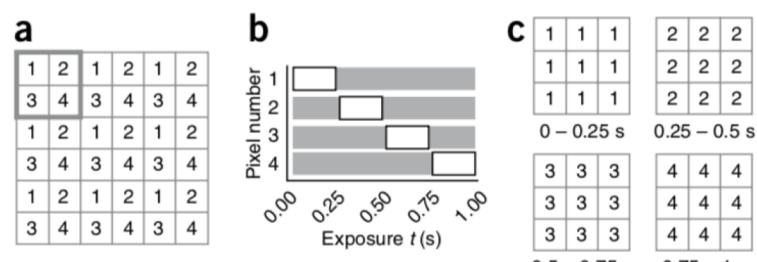
Technology	CMOS 0.18µm 1P6M
Die size (w. pads)	3174µm × 2227µm
Pixel size	22µm × 22µm
Fill factor	9.2%
Image resolution	64 × 64 pixels
Photodiode type	n-well/p-substrate
Power supply	3.3V-1.8V
Predicted power consumption	<100mW
Frame rate	30fps
Max. compressed sampling rate	50kHz
Clock Freq.	24MHz



[M. Trevisi et al., DATE 2018]

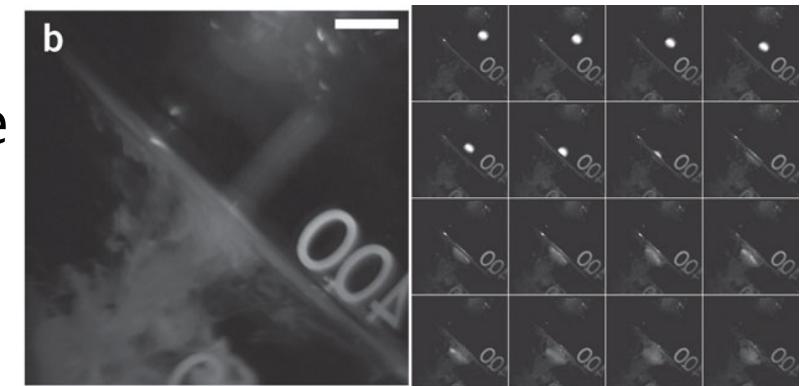
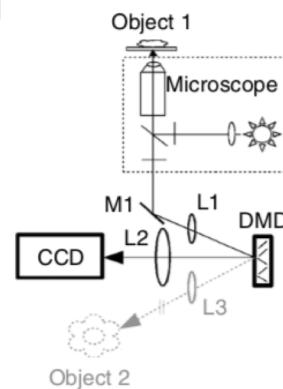
# Time pixel multiplexing

- Time-coded pixels
- Each pixels correspond to a different frame



G. Bub, 2010

Conventional camera used together with  
Digital Micromirror Devices (DMD)



High resolution still picture  
acquired to a lower resolution,  
high-speed video

# Aperture Compressive CIS

- Multiple arrays on single chip
- High-speed compressive imaging

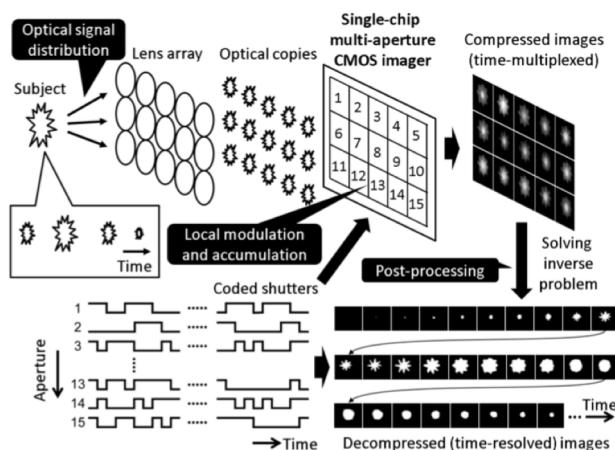


Figure 6.4.1: Overall flow of image acquisition and reproduction with a multi-aperture single-chip CMOS imager.

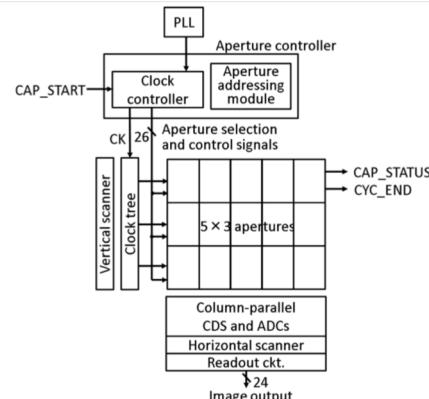


Figure 6.4.2: Sensor architecture.

F. Mochizuki, 2015

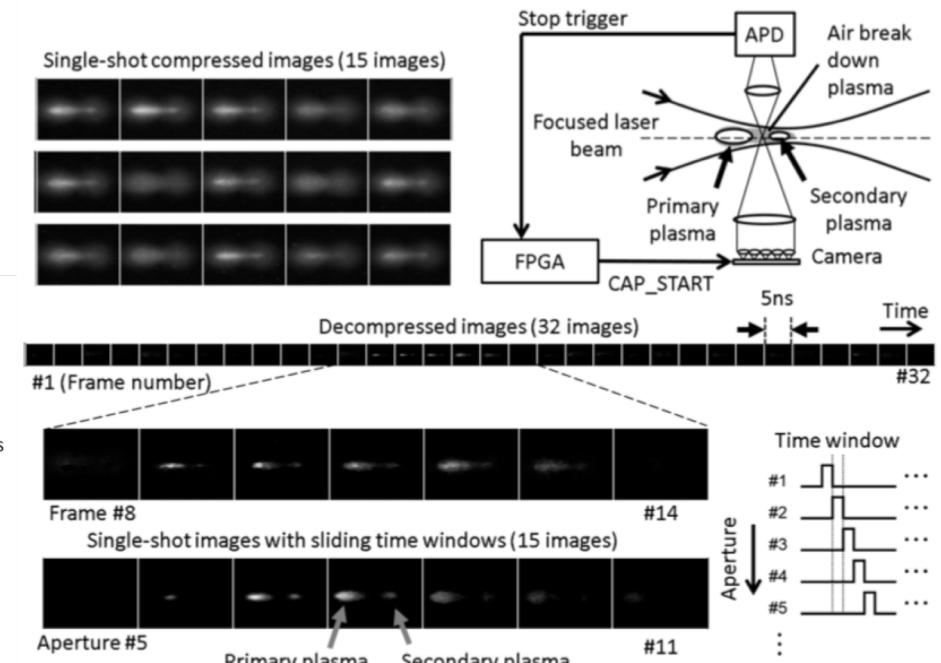
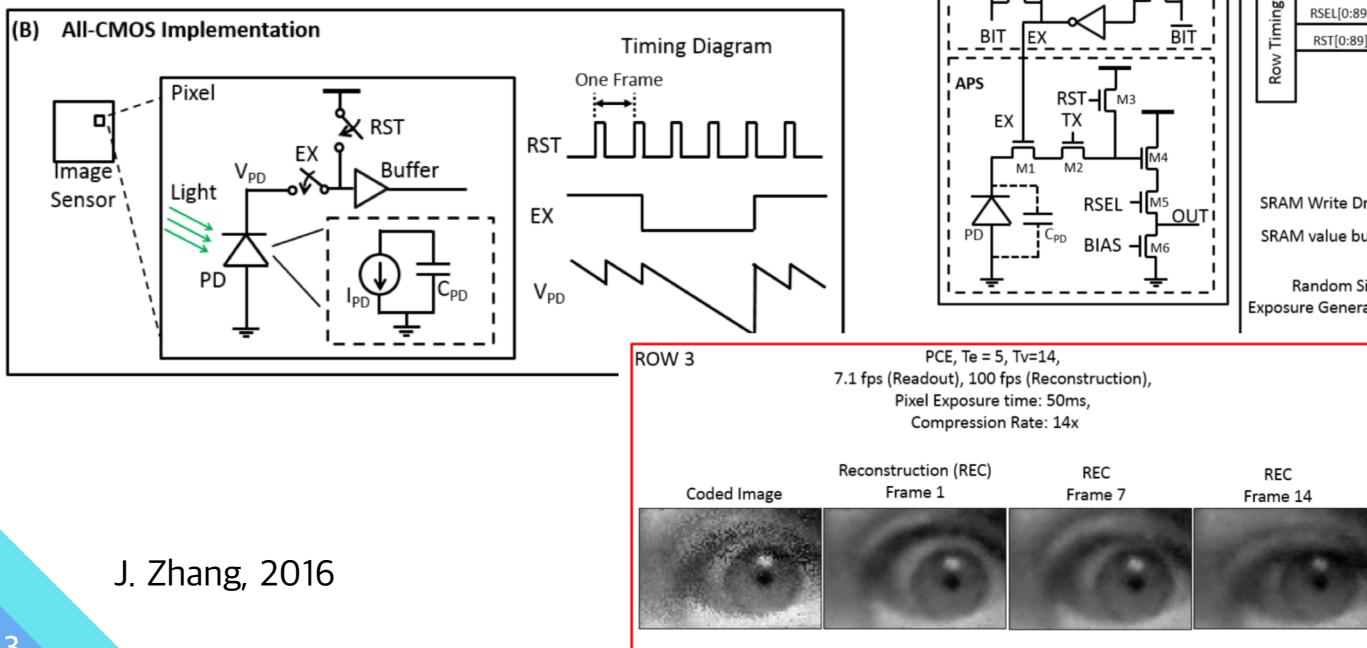


Figure 6.4.5: Experimental results of single-shot image acquisition of air breakdown plasma.

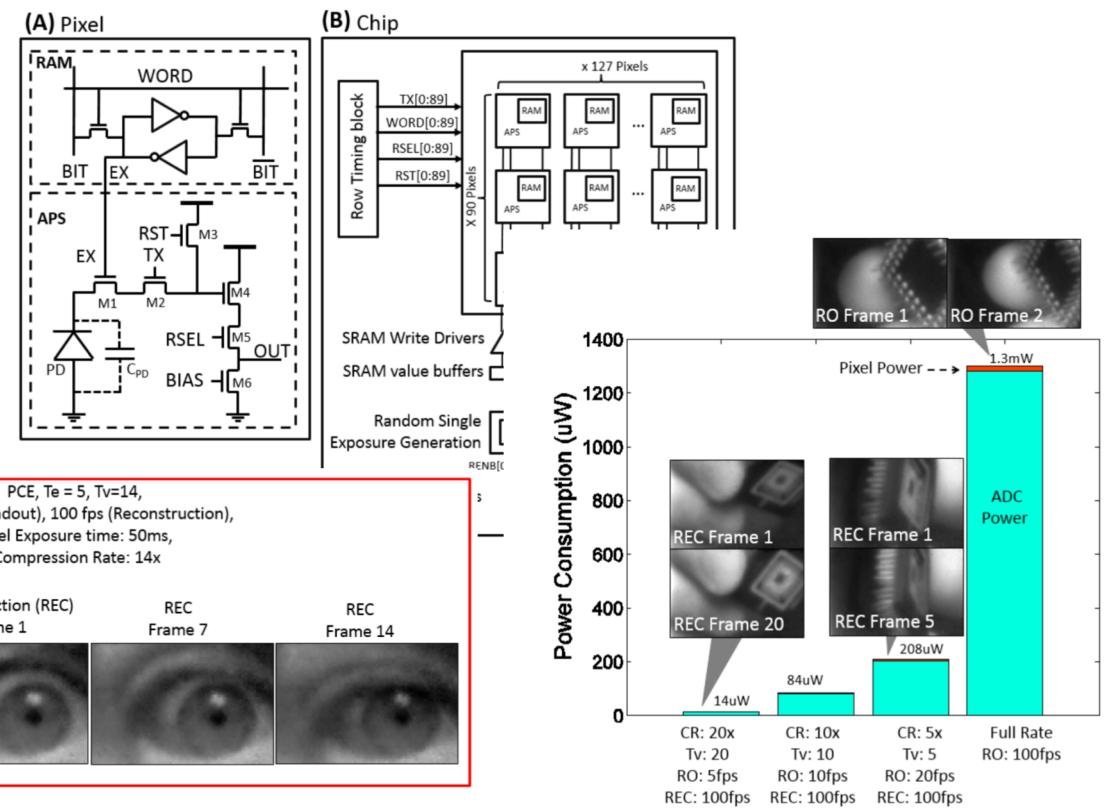
# Pixel-wise coded aperture

- Sensor designed to perform the compressive sampling
- Power efficient



J. Zhang, 2016

43

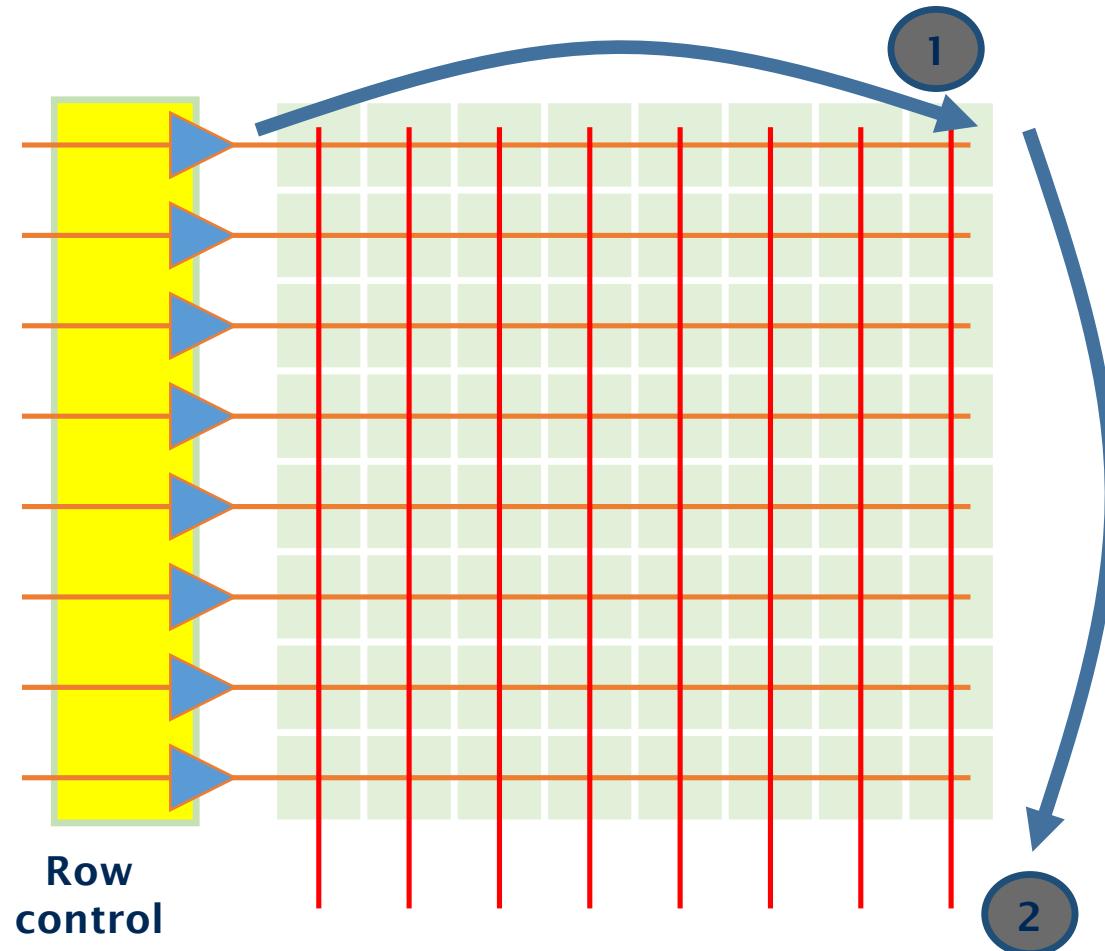




**High speed  
imaging**

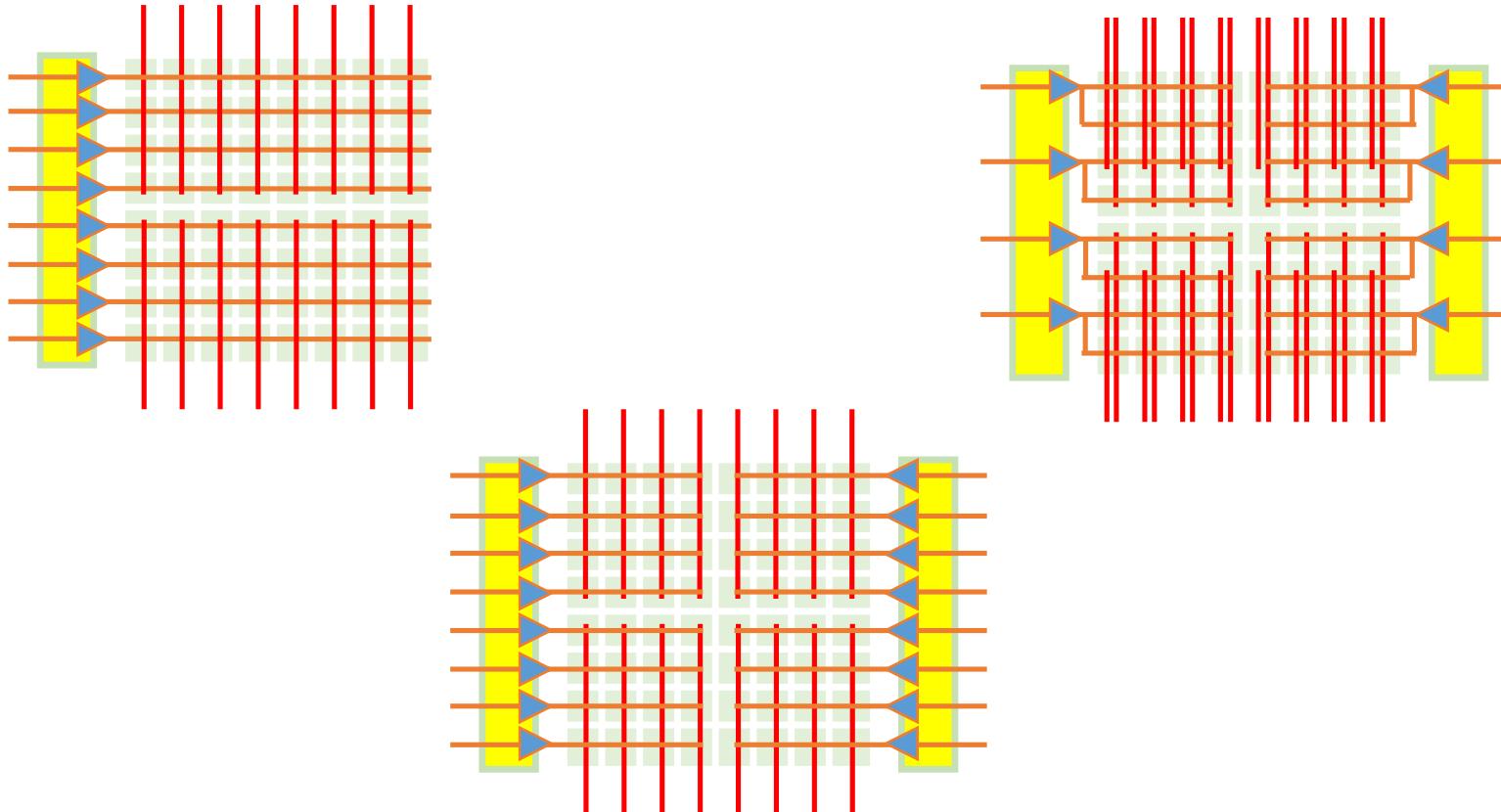


# Readout architecture. Sampling data from pixel.



Data out of the pixel  
to the periphery:  
Two main time  
constants:  
1) time to select the  
row  
2) Time to settle the  
data to the periphery

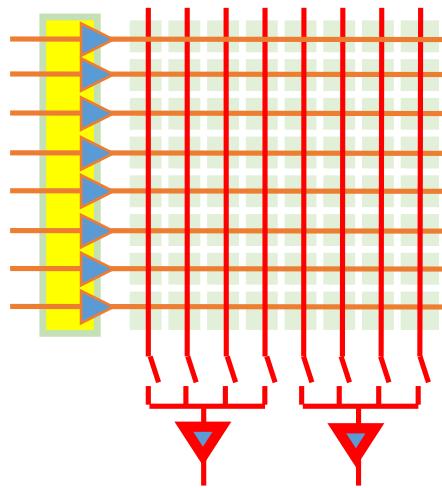
# Readout architecture.



**Save time by reading from two sides, controlling from two sides  
or reading rows in parallel**

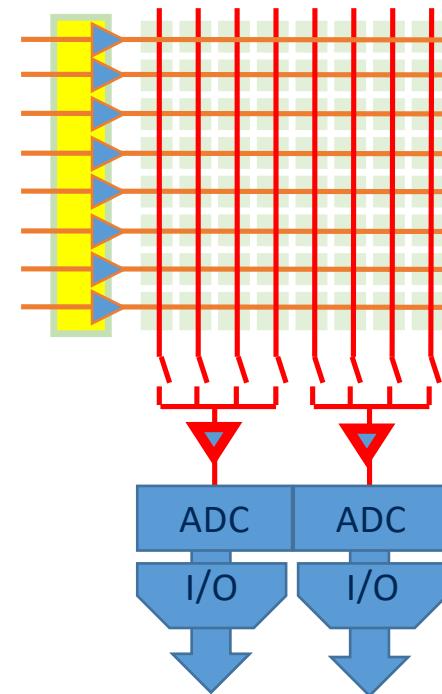
# Output path

Analogue output



Reduced design complexity  
Reduced area  
Reduced power  
but complexity, area and  
power moved to camera  
design!

On-chip ADC



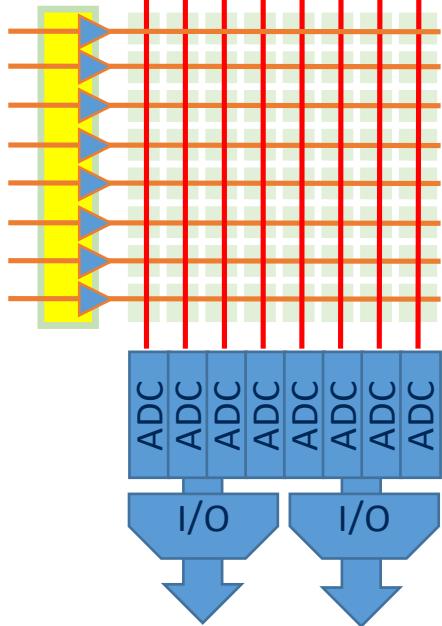
Limited design complexity  
More area for ADC  
Limited power  
Limited speed

(Leñero 2014)

# Output path

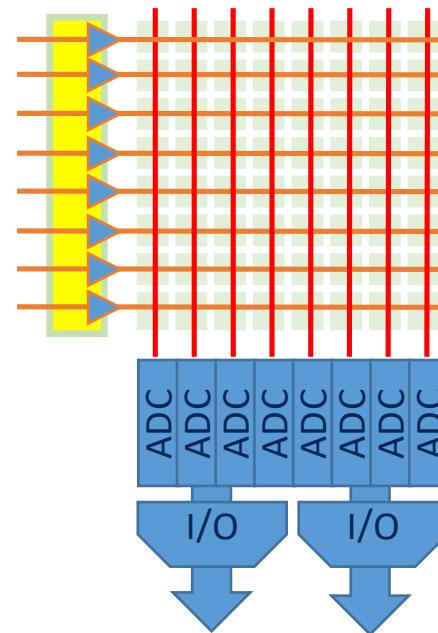


Column-level ADC



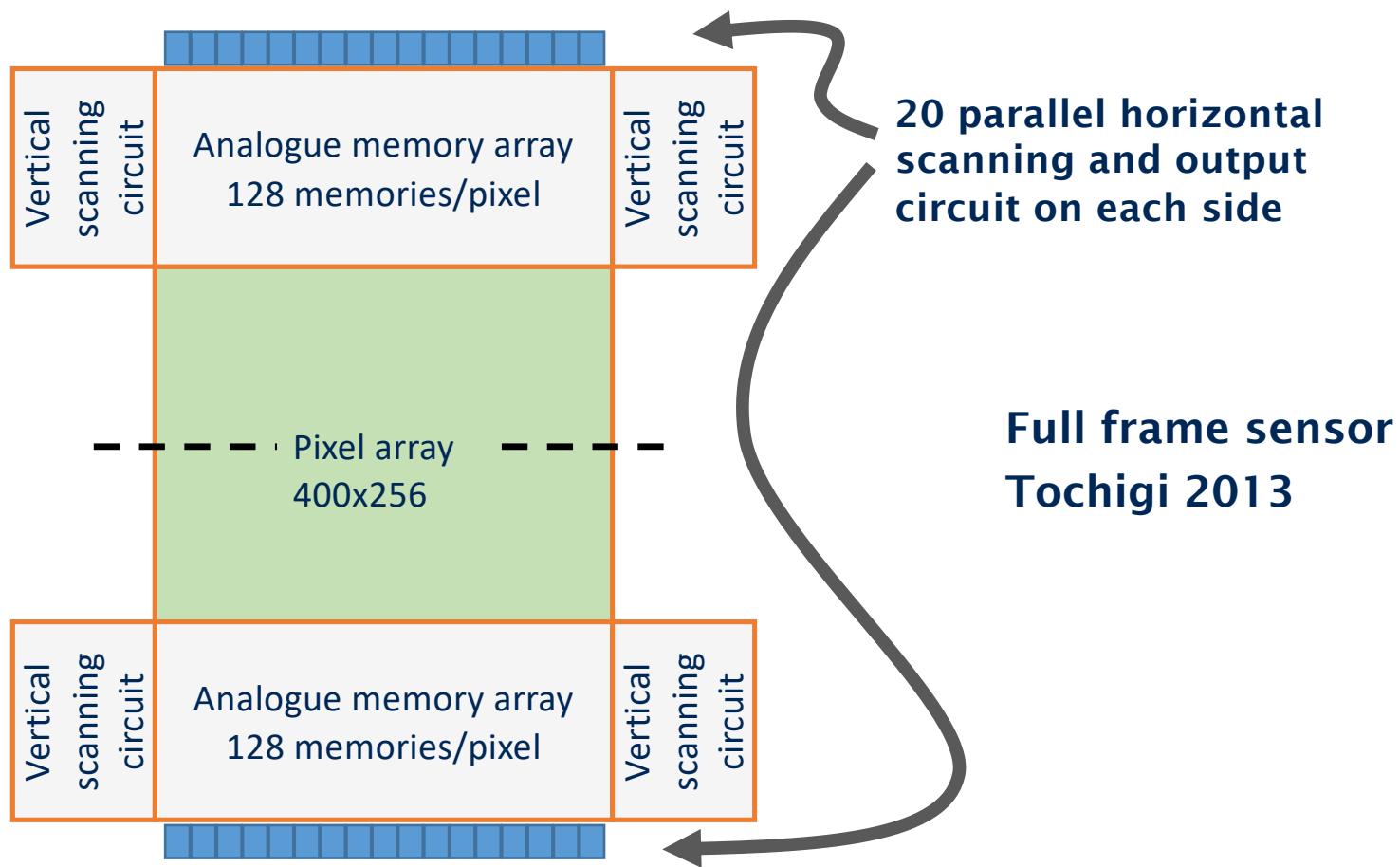
Small area for ADC  
Low power budget per ADC  
Medium speed  
High-bandwidth digital I/O

Pixel-level ADC

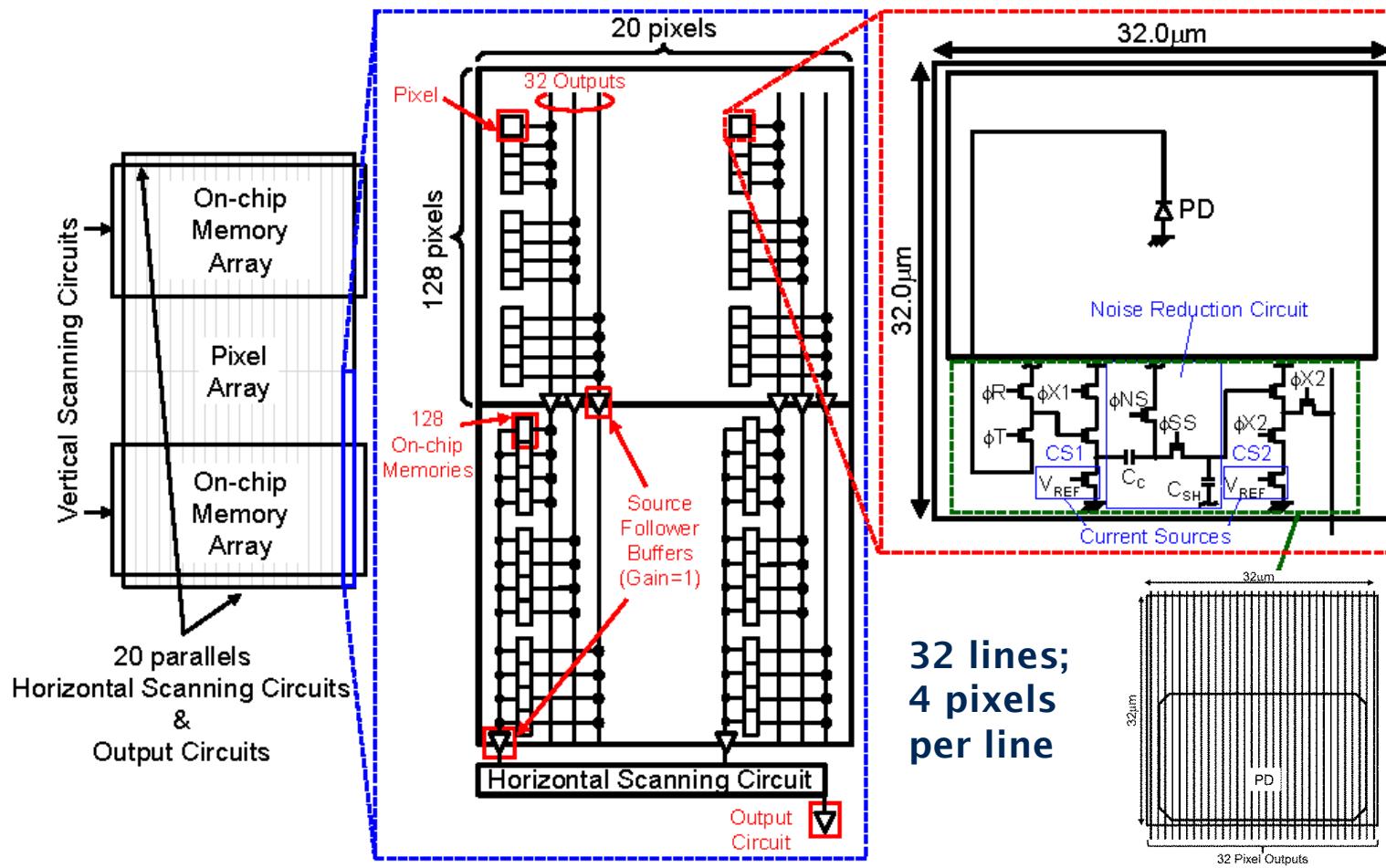


Minimal area for ADC  
Low power  
Lowest speed  
Add complexity to the pixel ( $\rightarrow$  3D?)  
High-bandwidth digital I/O

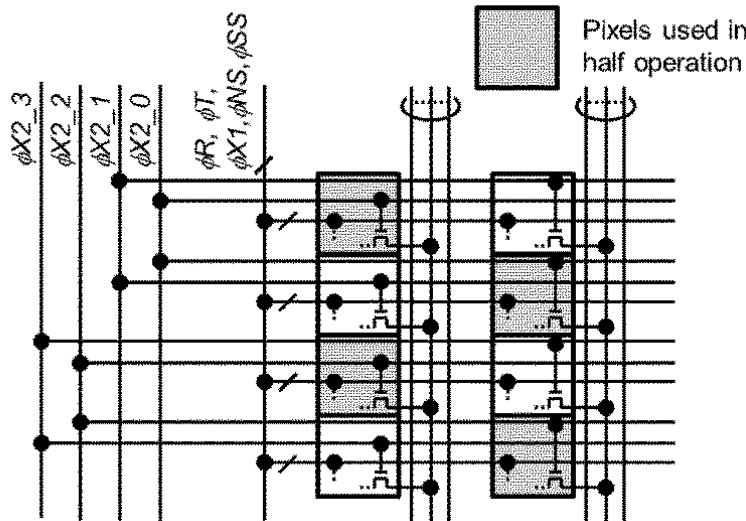
# Full frame voltage storage



# Full frame voltage storage



# Full frame voltage storage



**Full mode operation**

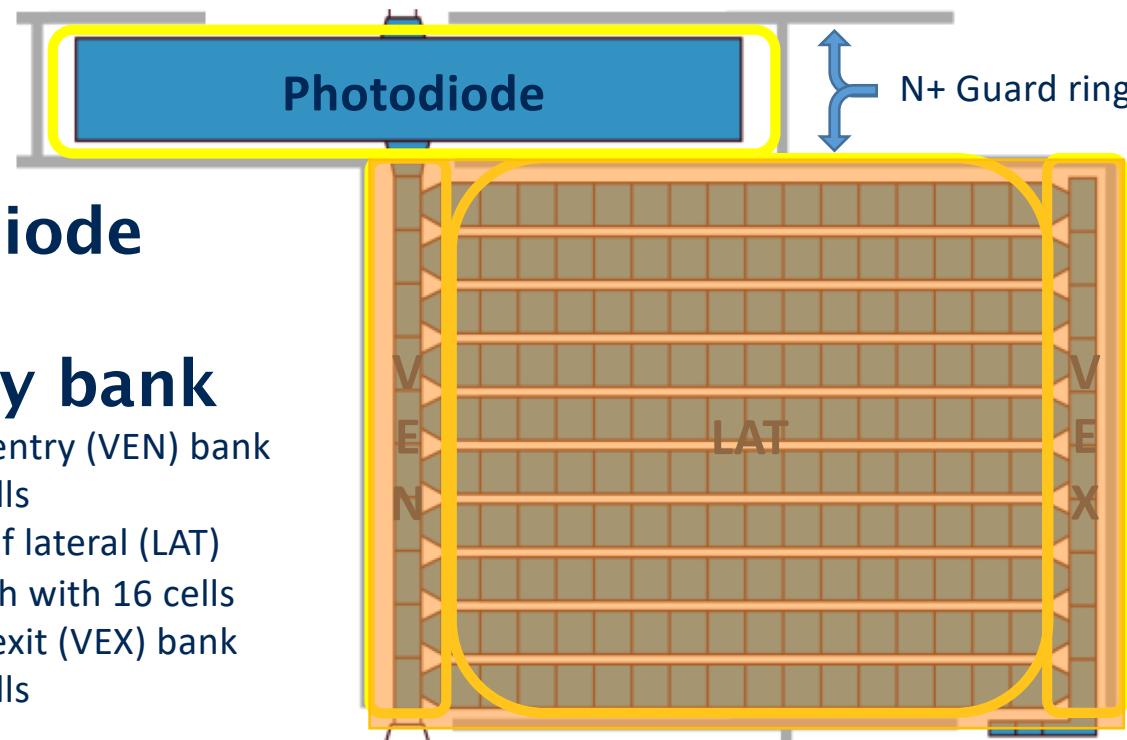
**Half mode operation: half  
the number of pixels,  
double the number of  
memory cells**

# Charge storage

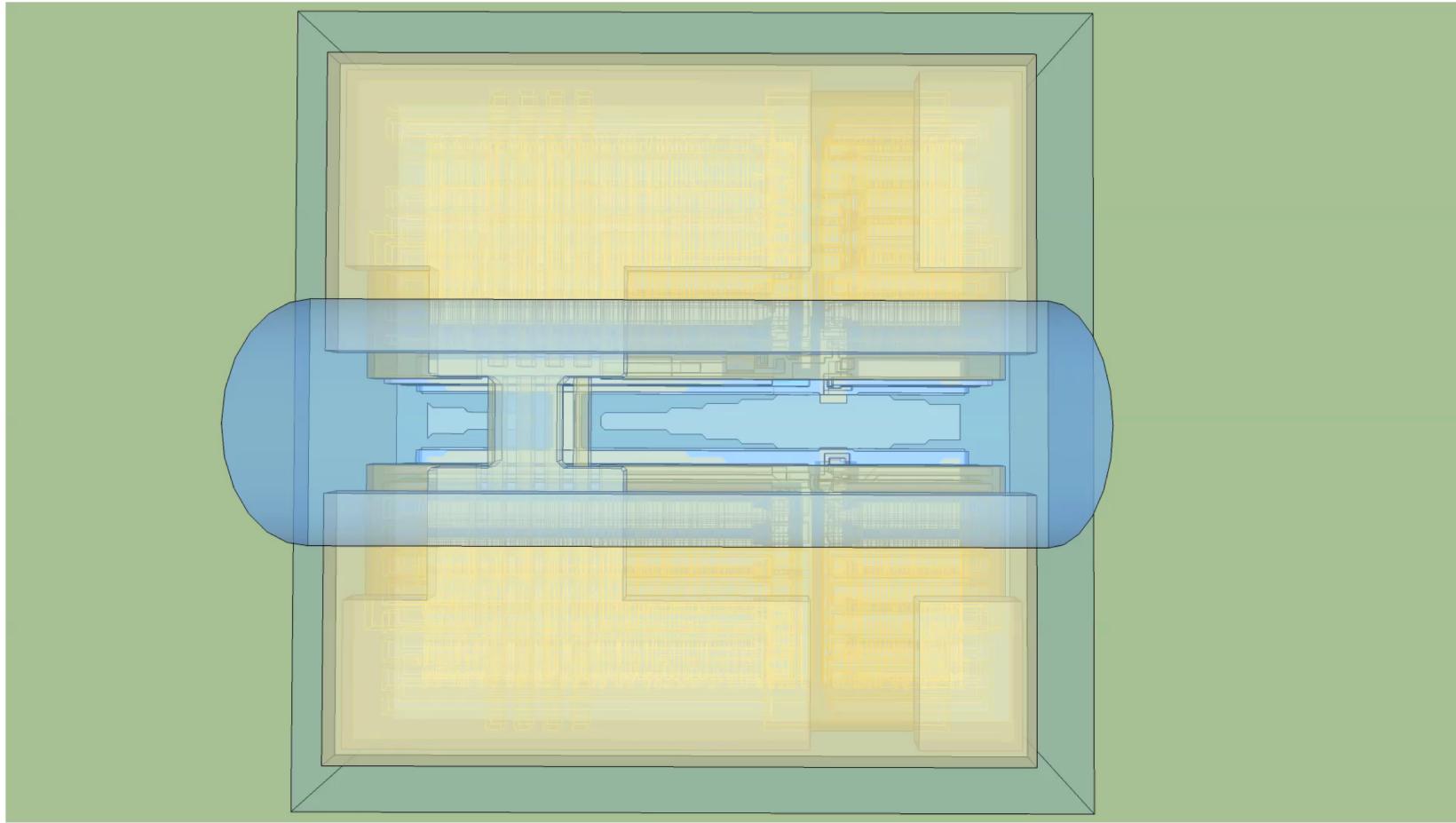
**Photodiode**

**Memory bank**

- A vertical entry (VEN) bank with 10 cells
- Ten rows of lateral (LAT) banks, each with 16 cells
- A vertical exit (VEX) bank with 10 cells



# The Kirana pixel



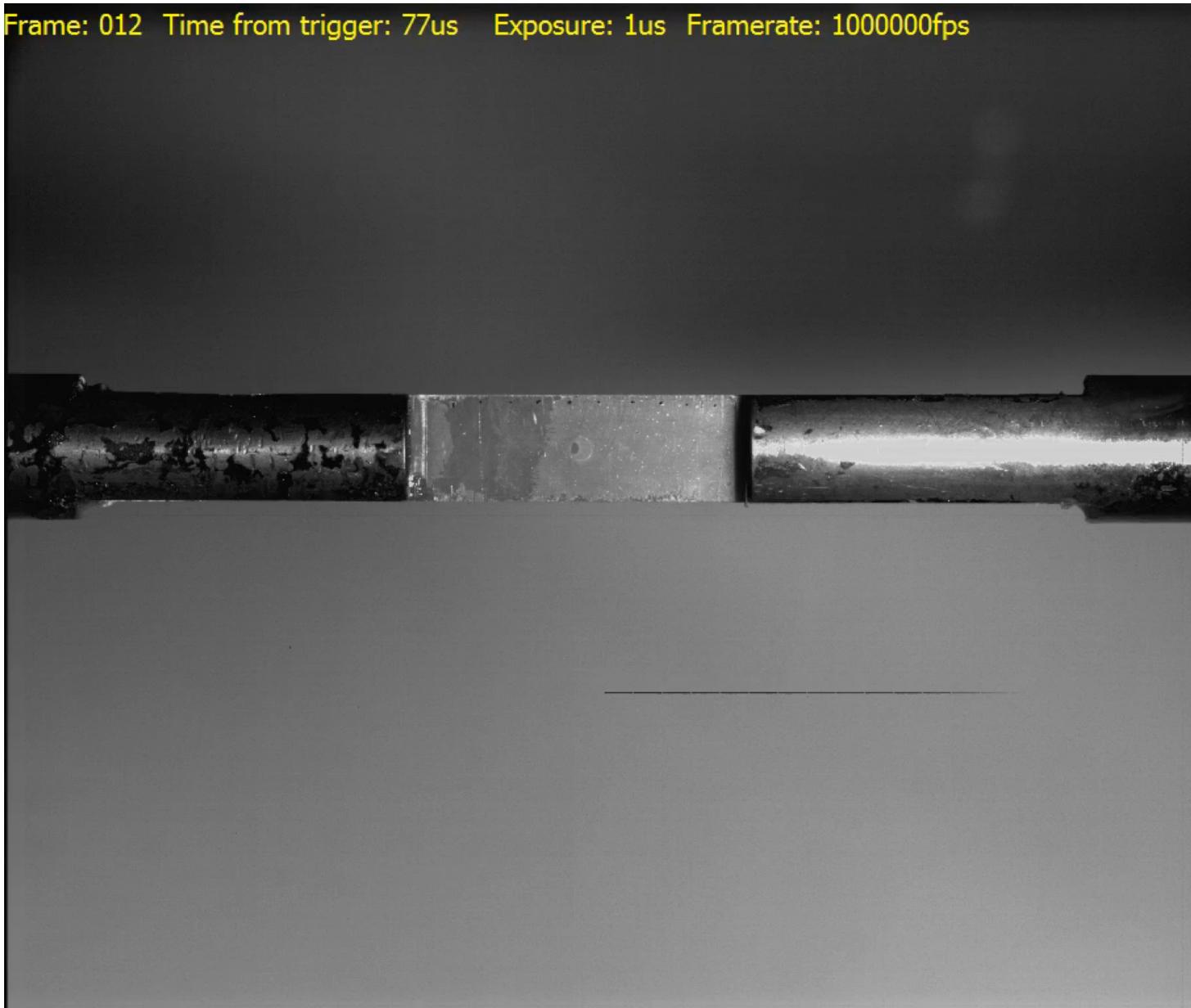
# The Kirana sensor



## Highly scalable architecture:

- Number of memory cells
- Number of pixels
- Low power

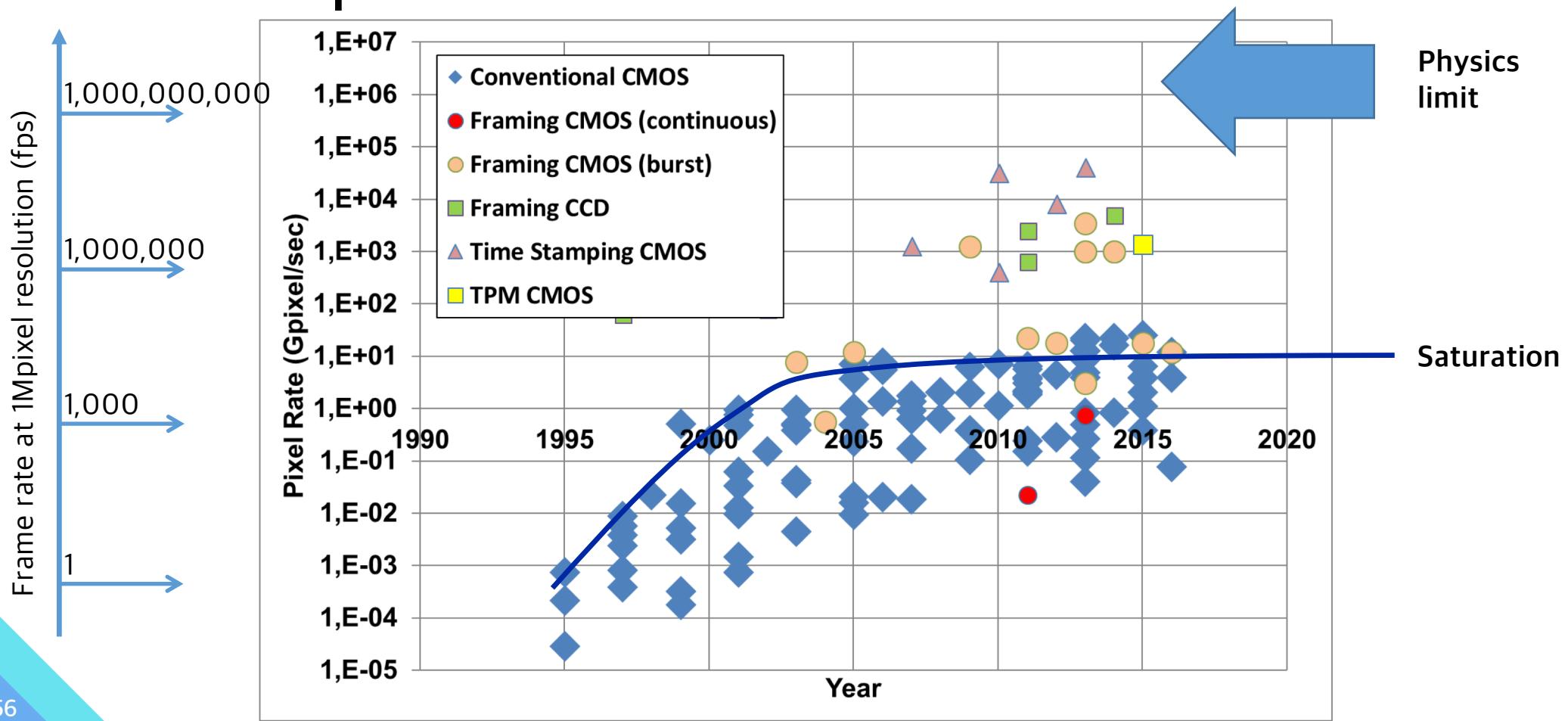
Frame: 012 Time from trigger: 77us Exposure: 1us Framerate: 1000000fps



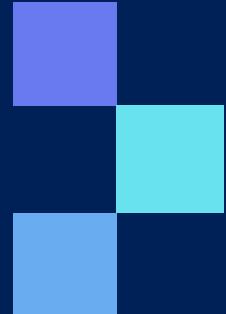
enic

# Evolution of CMOS image sensor speed

imasenic 



*Thank you!*

imasenic 

[www.imasenic.com](http://www.imasenic.com)