Part IV: Imaging System On-a-Chip

Introduction

- **By now, we have a good idea of how to detect light with silicon, how to design optical sensors, and how to treat noise in our detector system**
- **In this section, we will consider what else we need in order to construct a camera on a chip**
- **These system aspects can be divided into two main sections**
  - » the system needed for basic video capture; mostly common to all applications
  - » more specialised systems, such as adaptive circuits, artificial retinas etc
- **The first part will cover the basic elements of the camera – scanning, A-D conversion, video encoding, electronic shutter, and others**
  - » most of the commercially available devices include these standard features to form complete video cameras
- **While some standardisation in approach is becoming apparent, there is still a lot of innovation and change in reported designs**
- **In the second part, more advanced focal plane signal processing will be discussed**

Operating the Array

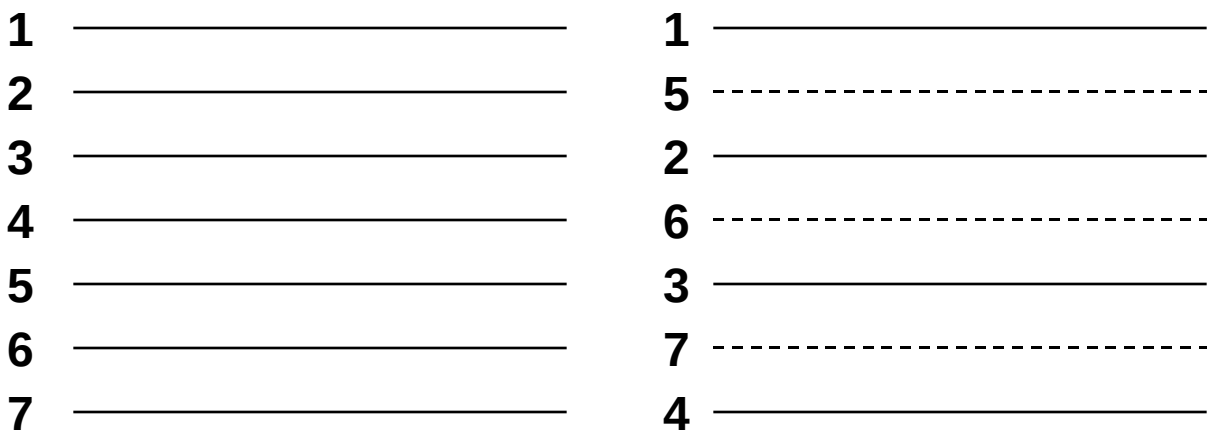
- **The basic scanning of the array requires several sets of inputs**
  - » bias voltages
  - » clocking signals
  - » reset signals
- **Bias voltages are required by the column source followers, and the S&H source followers**
  - » and possibly by subsequent amplifier stages
- **While bias voltages can be simply supplied by a resistive potential divider, this is not usually performed on-chip because**
  - » accurate resistors are hard to fabricate in digital CMOS
  - » potential dividers dissipate power continuously, especially because large resistors are not feasible
  - » and we may well wish to adjust the biases, perhaps to change the operating range of the sensor
- **Usually, some externally adjustable voltages are required to program an internal active reference voltage generator unit**
  - » based on a bandgap reference source for stability

- **An alternative approach is being pursued by JPL for low-power, digitally programmable still camera chip**
- **Here, four 5-bit digital-to-analog converters are used to program the DC reference voltages**
  - » the digital input signal to each DAC can be latched with little power consumption
  - » and the DACs are powered down when their output is not required to reduce power consumption
- **This highlights an important area of development, for battery-powered still cameras in particular**
  - » power conservation by intelligent on-chip power control
  - » e.g. standby modes (40μW for the JPL camera, ~10mW during operation) by turning off analog circuitry
  - » e.g. column-parallel ADCs only operated when column is read out (plus maybe one warming up)

Timing

- **How the timing of the array readout is controlled depends on the application**
- **In many cases, adherence to a standard video output format is not required**
  - » and there is more flexibility in the control methodology
- **If, however, the camera has to output signals in standard TV formats, for example, the frame rate and signal format are precisely defined**
- **The longest integration time allowable for the sensor is 1/(frame rate)**
  - » so, 1/30 s for standard NTSC (National Television System Committee) TV rates
- **Video standards determine every aspect of the signal output**
  - » levels and timing of line, frame, blanking etc. pulses
  - » black levels and white levels
  - » compression of information – gamma correction
  - » frame rates
- **There are two major standards, CCIR & EIA**
  - » International Radio Consultative Committee (Europe)
  - » Electronic Industries Association (North America)

- **And in each standard, there are two major scanning modes**
  - » progressive scan and interlaced scan



- **The progressive scan achieves a high resolution by displaying every line of the image**
  - » e.g. computer monitor
- **The interlaced scan achieves the same apparent resolution, but with a reduced bandwidth, by displaying every other line, and then going back to fill in the gaps**
  - » e.g. all TV sets
  - » originally, the bandwidth of the radio transmission was a major factor in adopting interlaced TV broadcasting
- **Each scan is called a field, and each full picture is a frame**
  - » so 1 frame = 1 field for progressive
  - » and 1 frame = 2 fields for interlaced

- **The frame rates are specified by the standard field frequency**
  - » CCIR: 50 Hz
  - » EIA: 60 Hz
- **So an interlaced EIA system displays 30 frames (i.e. full pictures) per second**
- **In a CMOS imager, the easiest way to achieve interlacing is simply to read out alternate rows of the imager**
  - » so there is effectively an independent array for each field
  - » [this does not work so well for CCDs, since the two fields cannot be reset separately; hence the fields have overlapping integration times, causing odd effects for moving objects]
- **After all the image processing, cameras for standard video formats may include some deliberately non-linear weighting**
  - » known properly as “inverse-gamma correction”
  - » but often called simply “gamma” or “gamma correction”

Gamma Correction

- **The need for gamma correction arose because the inherent behaviour of CRT monitors is non-linear**
  - » the brightness of the monitor is related to the accelerating voltage of the electrons by

$$L_{display} = K(V_{acc})^{\gamma}$$

- **So, if the camera were to have the inverse relationship, the properties of the scene would be maintained**

$$L_{display} = K(V_{acc})^{\gamma} = K\left(V_{scene}^{\gamma}\right)^{\frac{1}{\gamma}} = KV_{scene}$$

- **Of course, this only works provided the correct match of camera and monitor is used**
  - » although image processing systems, including computers, usually offer a choice of gamma correction parameters
- **Because of the non-linearity, image processing should be performed on the linearized image**
  - » even if the signal is gamma corrected again afterwards

- **The precise value of gamma depends on the CRT manufacturer, but standard values are recommended**
  - » NTSC has $\gamma = 2.2$
  - » or an inverse γ of 0.45
- **For low levels of light intensity, the effect of gamma correction is very large, so there is usually some cut-off point below which gamma is not applied**
- **In the camera, the gamma transformation is often performed using a look-up table**
  - » to convert the digital signals appropriately
  - » and in some cases, manufacturers may use their own compression system to enhance contrast or signal-to-noise ratios

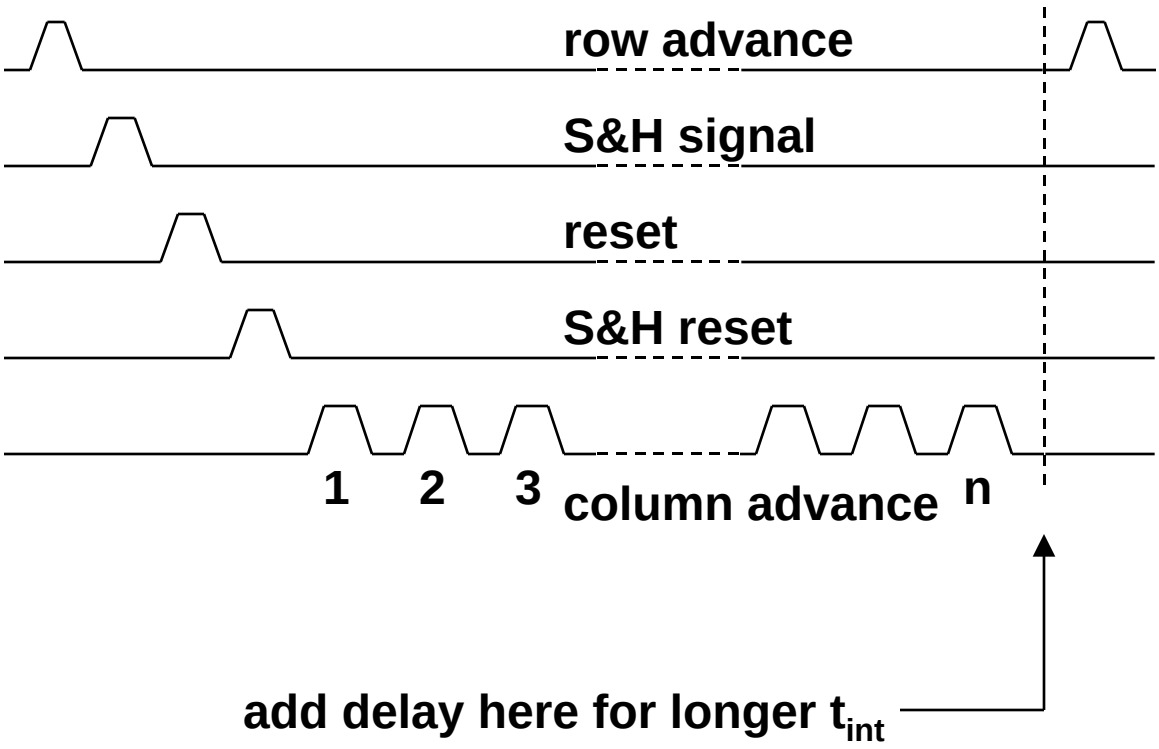
Exposure and Gain Control

- **Exposure control refers to the amount of light that is being sensed by the camera**
  - » primarily controlled by the integration time
- **The simplest way of defining the integration time is to make it equal to the frame time**
  - » i.e. the time between reading the pixel once, and reading it again on the next frame
- **Alternatively, for non-TV applications, an electronic shutter can be used to define the exposure time**
- **An integrated camera must be able to adjust its exposure to the prevailing conditions by adapting its integration time**
  - » but it can also adjust the gain of subsequent amplification stages to achieve the optimum combination of detector sensitivity and output voltage swing

- **In the VLSI Vision 5400 series monochrome camera chips, the gain and integration time are controlled together**
  - » to maintain a relatively constant average output signal
- **The number of pixels in a frame whose signal is above some predefined threshold value is counted**
- **If, for example, there are too few bright pixels**
  - » first the integration time is increased
  - » if the integration time is near the maximum value, the gain is increased and the integration time reduced
  - » to allow room for further increase in the integration time
  - » the total range of exposure control (from min. t_{int} & min. gain, to max. tint & max. gain) for the VVL 5400 chips is about 100,000:1
- **The same chips also allow for longer exposures in part of the image in a "backlit mode"**
  - » for dark objects on a bright background (snow, sand, sky etc)
  - » by applying a higher threshold in the central part of the array
  - » this is a simple version of the methods used to increase the dynamic range (see later)

Clocking and Speed

- **The way in which the array is scanned very much depends on the application**
  - » modifications must be made to accommodate interlacing, variations in integration time, electronic shuttering etc.
- **In the basic system of photodiode readout, we must supply clocks to operate row and column scanners, the two sample and holds, and the pixel reset**

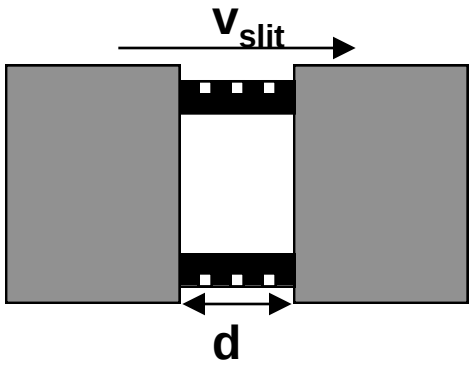


- **Most camera chips generated these clocks on-chip from an externally regulated master clock**
  - » usually a crystal oscillator like a computer to achieve frequency stability
  - » although on-chip oscillators are also used for maximum integration
- **As with all circuits, the readout has a maximum frequency of operation, determined by**
  - » the fabrication technology
  - » the detailed design, both low-level and overall circuit functionality
- **In particular, it is hard to drive large arrays fast, because of the capacitance of the address lines**
  - » $i = C\, dV/dt$
  - » so larger C requires that the drivers of the address lines must be able to source more current
  - » this requires larger transistors, and multiple stages (e.g. cascaded inverters), which also add delays
- **It also makes the use of smaller pixels advantageous**
  - » to reduce the array physical dimensions (hence reduce C) while maintaining resolution
  - » also reduces cost

- **CMOS addressing is typically less capacitative than for CCDs, but the on-chip drive circuitry cannot be too large**
  - » because of the power dissipation
- **For large arrays, such as the VVL 800K-pixel sensor, readout rates are 5 - 10 MPPS**
  - » mega-pixels per second
  - » which is about typical
  - » or a few fps for megapixel arrays, and 30 - 100 fps for VGA (640 x 480) or less
- **For specialised applications, such as industrial inspection, high frame rates may be required**
  - » the Integrate Vision Products (IVP) MAPP2200 can capture 4000 fps with a 256 x 256 array
  - » an impressive 260 MPPS
- **But, in general, limiting the integration time to the frame time results in t_{int} being too long**
  - » so how can we make a widely variable t_{int}?
- **This is also linked to the issue of reducing image blur for rapidly moving objects**
  - » e.g. scissor blades passing at 400 per minute
- **The answer is an electronic shutter, analogous to that found in a mechanical camera**

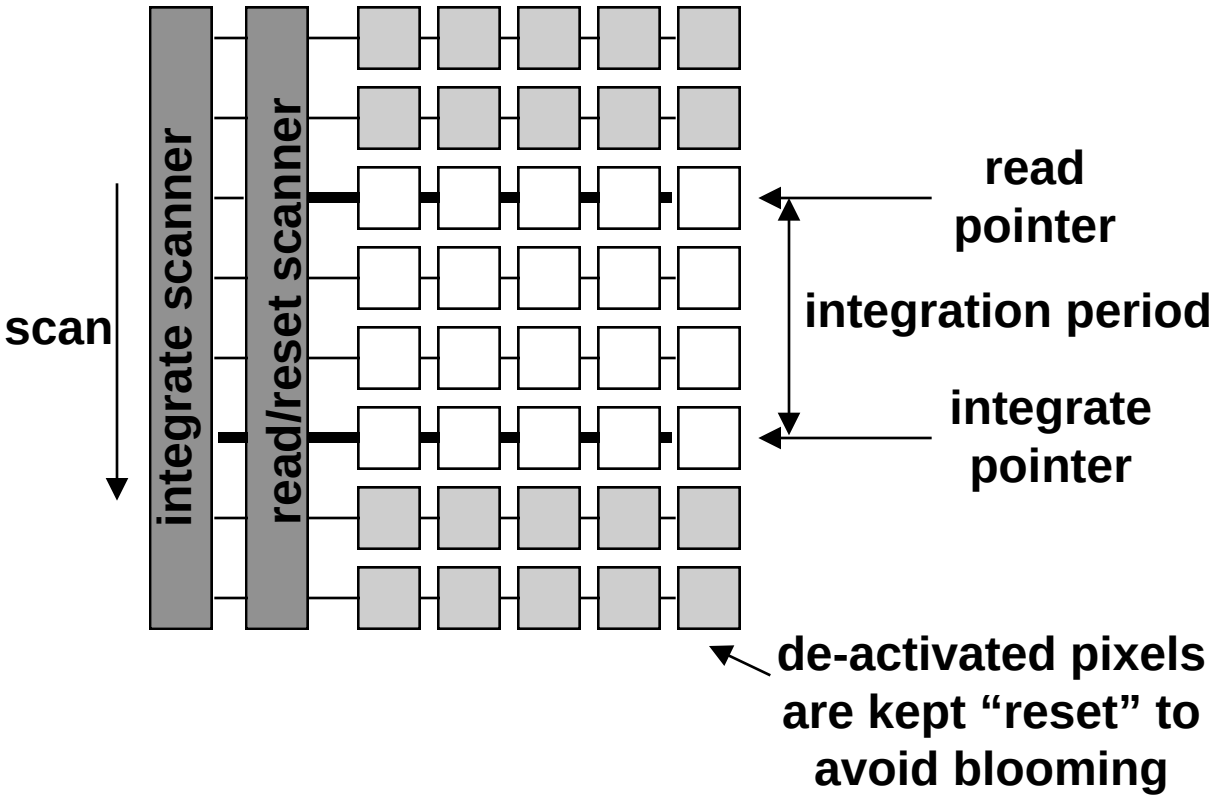
Electronic Shutter

- **In a regular film camera, the shutter consists of a slit made of two parts**
  - » the first edge moves across rapidly, exposing the film to the light
  - » the second edge follows the first, at a distance determined by the desired exposure time, cutting off the illumination
  - » the exposure time is d/v_{slit}, where d is the slit width and v_{slit} is the velocity



  - » this method, called a rolling shutter, is used to obtain improved high shutter speeds
- **The same method is also used to expose CMOS camera chips**
  - » except that opening the “shutter” refers to selective activation of rows of pixels
  - » rather than to a mechanical shutter

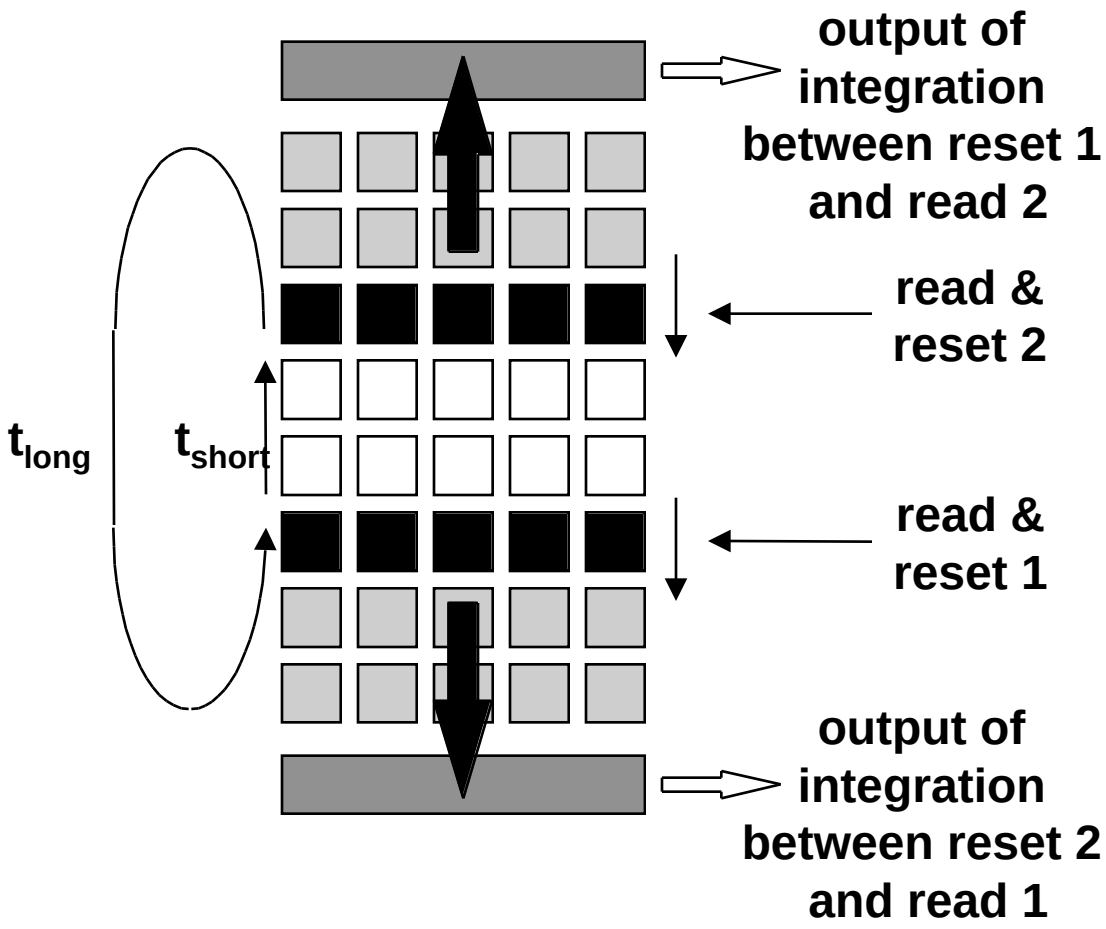
- **Thus the readout is implemented as follows**



- **The integration time is thus variable from (1/m) to (1) times the frame time**
  - » where there are m rows
  - » 70µs - 33ms for a VGA at 30fps
- **A derivative of this technique has also been used to increase the dynamic range of the camera**
  - » by allowing two integration times

WIDyR Sensor

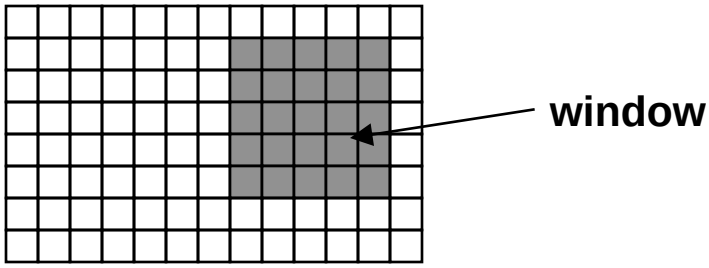
- **The Wide Intrascene DYnamic Range (WIDyR) sensor employs the same principle as the rolling shutter**
  - » but has an additional column readout system



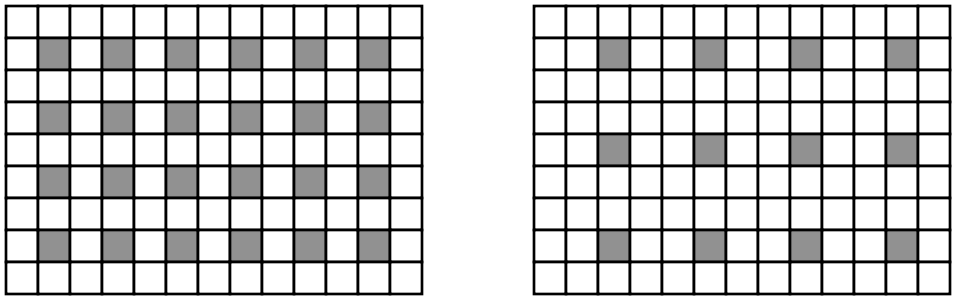
- **The integration times can be controlled**
  - » but not independently
  - » and the longest time depends on dark current

Selective Spatial Scanning

- **Many of the non-video-standard cameras offer two modes of selective spatial scanning**
  - » windowing
  - » sub-sampling
- **Windowing refers to the selective scanning of a restricted area of the array**



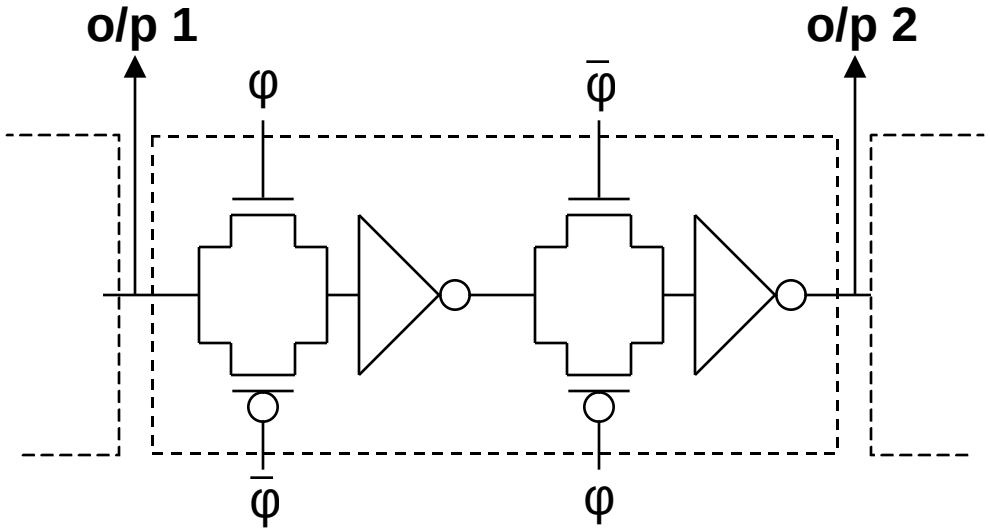
- **Sub-sampling is the readout of every n^{th} pixel**



- **The idea of both schemes is to get information about the scene without collecting all the data**
  - » usually to enable faster scanning

Shift Registers

- **From the previous points, the scanning circuits have to perform several functions**
  - » sequentially scan out the columns and rows
  - » offer enough drive capability for the address lines
  - » be programmable for start, stop, and step
- **There are two major candidates for the scanning circuits**
  - » shift register and counter/decoder
- **Shift registers are simpler and more expandable**

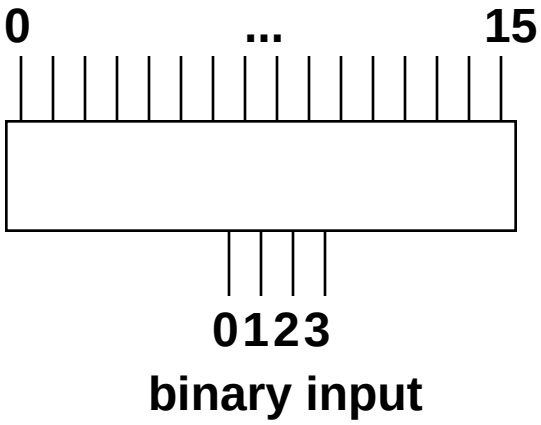


  - » every two clock periods, the signal on o/p1 is transferred to o/p2
  - » so whatever sequence we input at one end is passed along to the array

- **So, to read out all the columns, we put a “1” followed by (n - 1) “0”s into the shift register, and this sequentially activates the column readouts**
- **The real advantage of the shift register is that it can be made infinitely long without any change in architecture**
  - » just by adding on extra cells
  - » the input circuitry etc all remains the same
- **It can also handle two pulses for the shutter and the WIDyR sensor, although it may well be better to use two separate registers**
- **However, the shift register does not easily allow anything more complex than sequential readout**
  - » random addressing is impossible; a “1” has to propagate all the way from the input before it reads out a column at the other end
  - » windowing & sub-sampling; how to determine start and stop? How to skip columns?
- **So shift registers are usually used in cases where unconventional readout is not required**

Decoder Scanning

- **Decoder scanning has more in common with the addressing of computer memory**
  - » the decoder takes a binary input and selects the appropriate sequential output

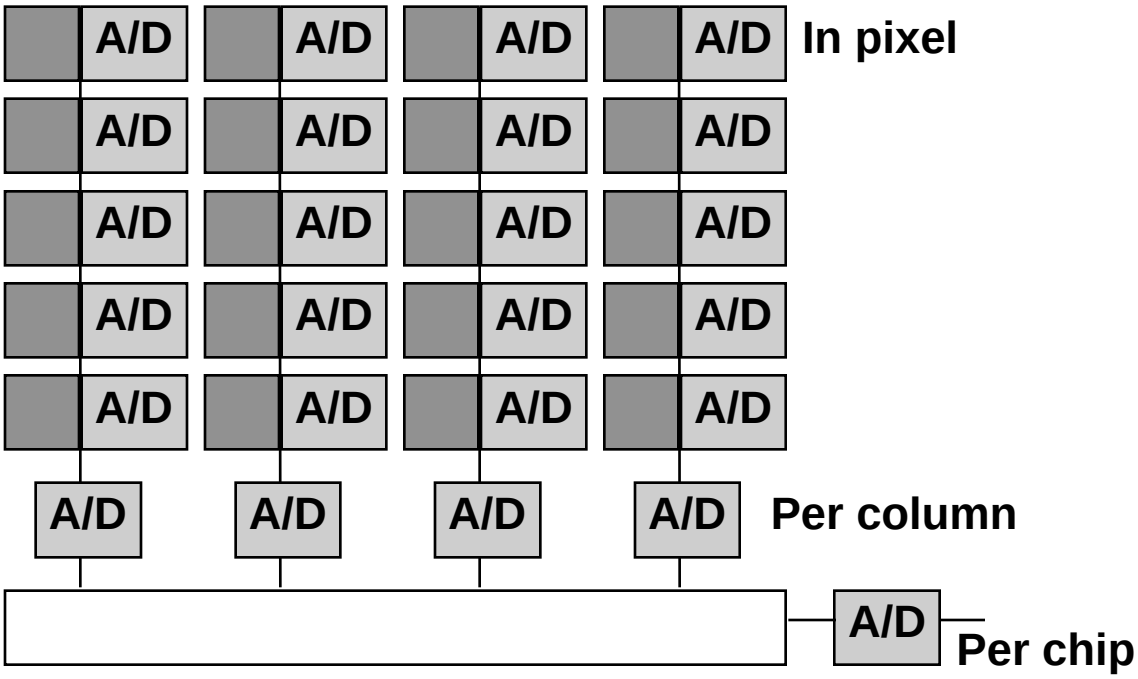


- **This allows any output to be selected at random**
- **Or, to scan the array, a binary counter can provide the input**
  - » this counter can have programmable counting limits for windowing
  - » and programmable counting steps for sub-sampling
- **The main disadvantage is that the decoding circuitry is more cumbersome and less straightforward to expand**
  - » although modern layout tools minimise this problem

Analog-to-Digital Conversion

- **The preceding sections have covered the main building blocks required to operate an integrated camera**
- **However, there is an increasing demand for direct digital signal output, especially for computer interfacing such as**
  - » digital still cameras
  - » machine vision & industrial inspection
  - » interne-based video telephony
- **But there are many possible ways of including ADCs onto the chip**
  - » one ADC per chip
  - » one ADC per column
  - » one ADC per pixel
- **The choice depends on a number of factors, such as**
  - » speed of conversion (1 - 50MPPS for video rates)
  - » space (area of layout, column pitch)
  - » number of bits (at least 8 for most applications)
  - » power consumption (<100mW to avoid local heating)

General Considerations



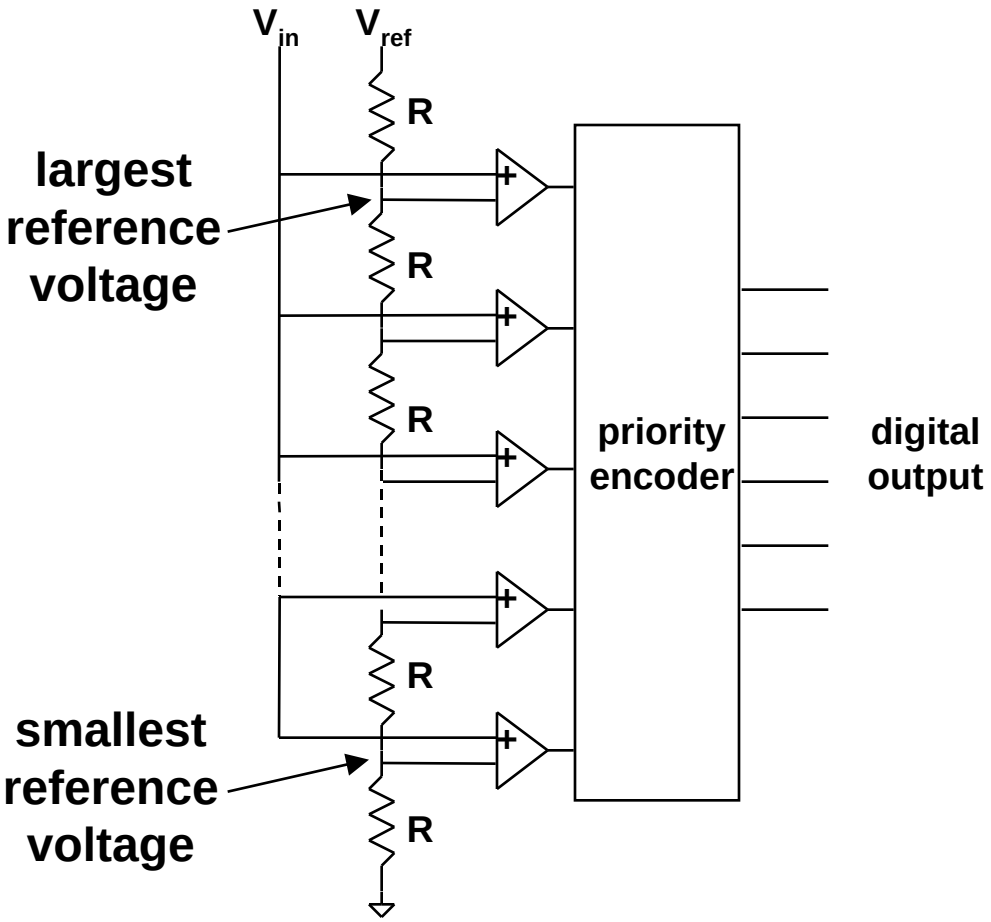
Factor	In pixel	Per column	Per chip
Speed	●	⊙	○
S/N	●	⊙	○
FPN	○	⊙	●
Accuracy	○	⊙	●
Fill factor / resolution	○	●	●
Power consumption†	○	⊙	●

● good ⊙ medium ○ poor

† Power consumption depends on how the system is operated.

Flash ADCs

- **Flash ADCs, sometimes also called parallel ADCs are the fastest converters, owing to their parallel operation**

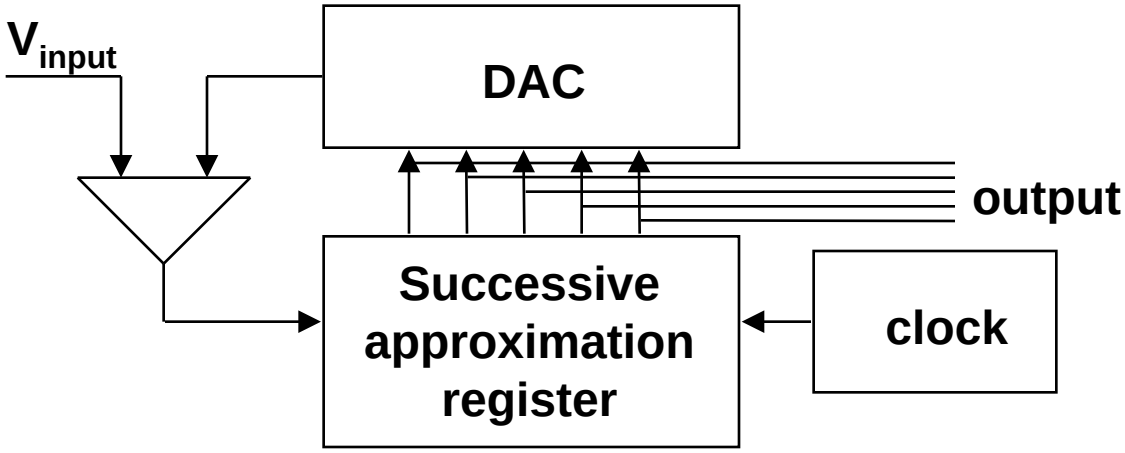


- **All comparators whose reference is below the input voltage gives an output "1"**
  - » and the priority encoder outputs the binary number corresponding to the highest active comparator

- **The flash ADC is fast**
  - » limited only by the delays in the comparators and the encoder
- **But it requires 2^N - 1 comparators (N = # bits)**
  - » so the number ~ doubles for each extra bit
  - » and this requires a large area on the chip
- **Moreover, the resistor chain imposes additional limitations**
- **Firstly, there is a continuous current flow through the resistors**
  - » Rs can be made large to reduce this, at the expense of real estate
- **Secondly, the accuracy of the ADC depends on accuracy to which the resistors can be fabricated**
  - » limited to 8 - 10 bits without special “trimming”

Successive Approximation ADCs

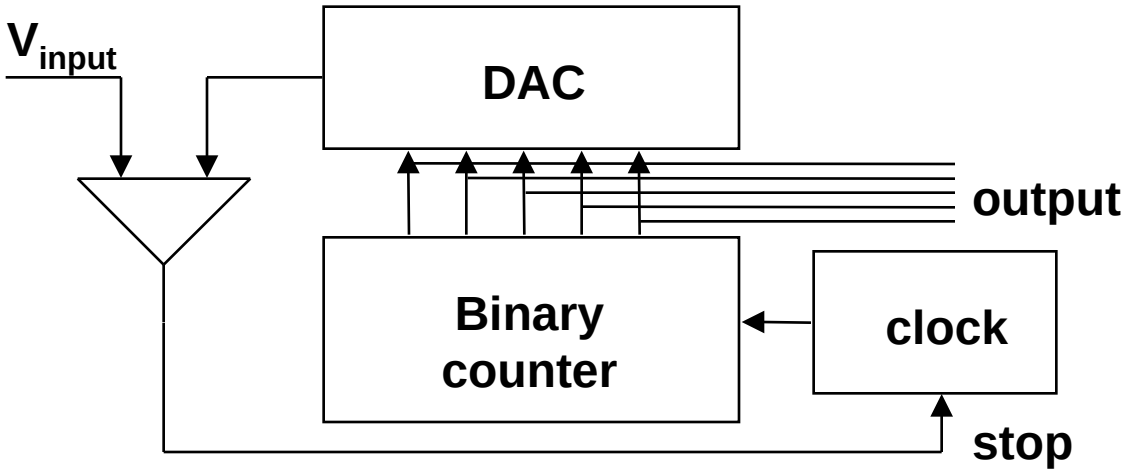
- **These ADCs converge on the correct digital word by comparing the input voltage with the output of a DAC corresponding to that word**
  - » and adjusting the digital word to bring the DAC output closer to the analog input



- **The converter first tests the most significant bit of the digital word**
  - » and turns it on or off according to whether the resulting DAC-ed voltage is greater or less than the input
  - » then each lesser significant bit is tested until the best digital equivalent is reached
- **Thus the successive approximation ADC requires N clock cycles to perform the conversion**

Single-slope ADC

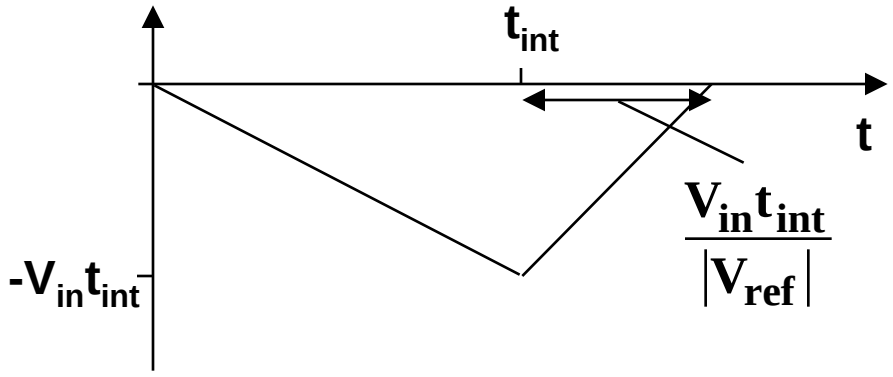
- **Also known as the Counting ADC, this is similar to, but simpler than, the successive approximation type**



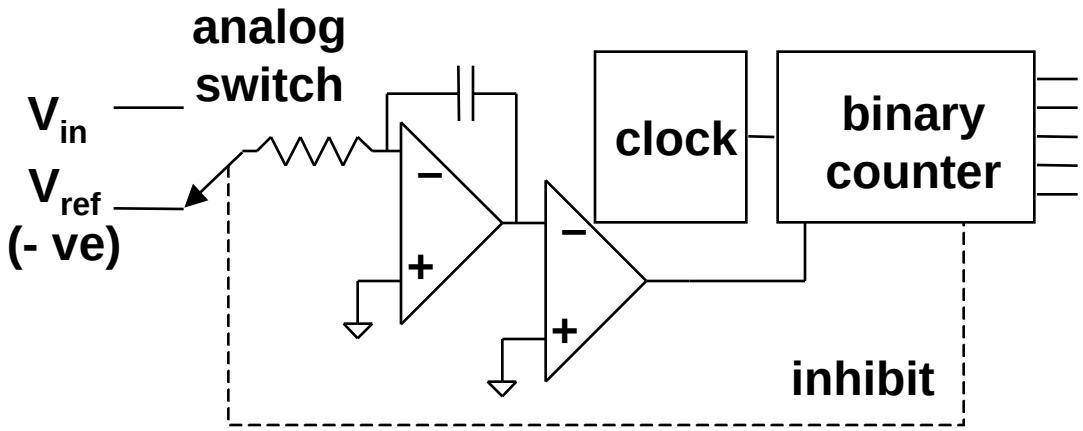
- **Now, the binary counter is incremented (or decremented) until the output from the DAC just exceeds (or falls below) the analog input**
  - » when the comparator inhibits the counter
  - » and the output of the binary counter is the best digital approximation to the analog voltage
- **The speed and accuracy of the conversion are not good**
  - » the speed depends on the voltage to be converted
  - » but the area and the power consumption are relatively low

Dual-Slope ADC

- The dual-slope, or ratiometric, ADC is another counting converter, but is more sophisticated and robust than the single slope method
- Here, the analog input it integrated for a fixed period of time
- And then the time required for a known reference voltage to achieve the same final value is measured by a binary counter



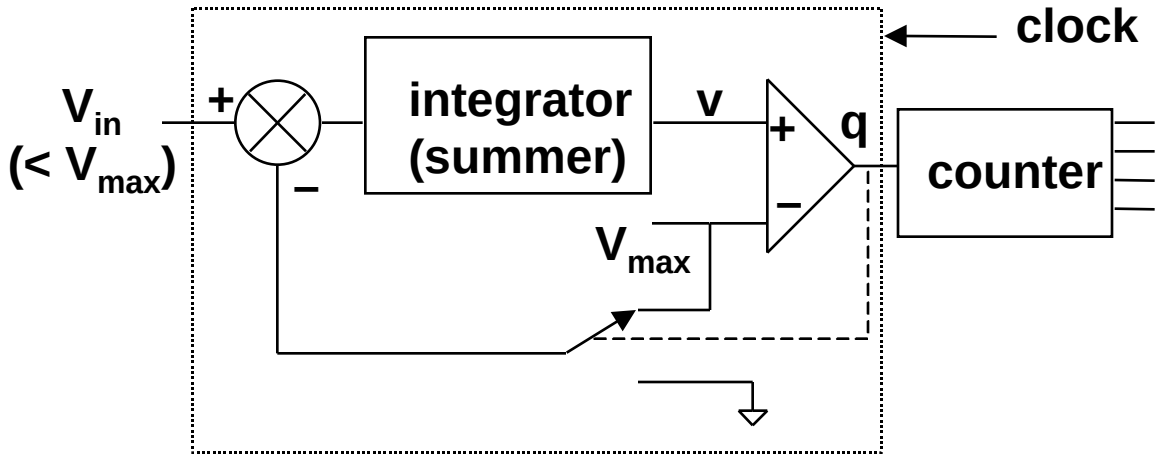
- This is achieved by using an integrator with an analog switch to determine the input



- The accuracy of this technique can be very high, but at the expense of a long integration time
- Moreover, the effects of digital noise can be removed by ensuring that the integration period is an integral multiple of the interfering period
- But the power consumption and area are still low

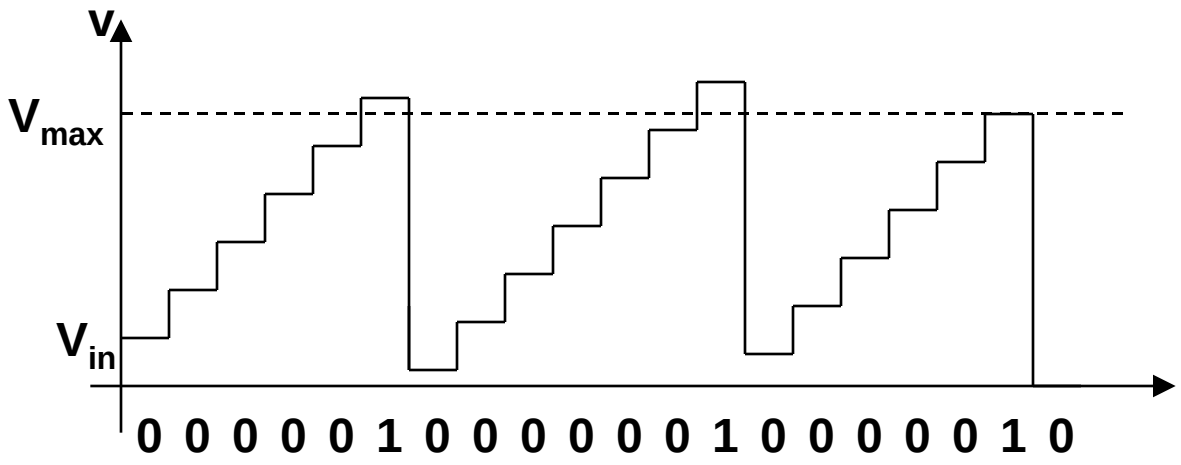
Sigma-Delta ADC

- **Sigma-Delta (∑-Δ) converters offer an excellent immunity to process variations**
  - » whereas previous types rely on high accuracy components
- **This is because it functions on an averaging basis**
  - » in its simplest form the ∑-Δ converter block diagram is as follows



- **During each clock cycle, the integrator accumulates multiples of the input voltage**
  - » until the total voltage exceeds the reference voltage, V_{max}, and the output, q, goes to “1”
  - » at which point, the feedback loop subtracts V_{max} again

- **The process repeats and, after another few periods, q → 1**



- **So, by counting the number of "1"s over 2^N clock periods (N = # bits), a binary counter will give the analog output**
- **More complex versions of the ∑-Δ converter (second order etc) are possible**
  - » they reduce the number of clock cycles for a given number of bits
  - » but at the expense of extra feedback loops, with higher real estate costs

Summary

- **Following Mendis, the merits of the various ADCs can be summarized as follows**

Type	Resol-ution	Speed	Power	Area	Robust-ness
Sigma-Delta	High	Slow	Low	Low	High
Succ. Approx.	Med.	Fast	High	Med.	Med.
Single/ dual slope	High	Slow	Low	Low	Low
Flash	Low	Fast	Very high	High	Med.

- **Here, robustness refers to the immunity of the technique to the process variations inevitable in the use of standard CMOS technology**
- **It is generally considered that flash ADCs are too large for other than one per chip**
  - » but that the others would be suitable for one per column
  - » and one per pixel is not really viable

Examples

- **The commonly accepted design is to use column-parallel ADCs**
- **And the successive approximation approach seems to give the best compromise between resolution, speed and power consumption**
- **JPL is currently the leader in incorporating ADCs into CMOS image sensors**
- **JPL's single chip camera uses 256 10-bit successive-approximation ADCs**
  - » at about a 24µm pitch
  - » taking 1/3 of the chip area
  - » and ~20% of the total chip power
  - » and can achieve > 85KPPS
- **JPL also use a single-slope ADC to achieve a smaller column pitch**
  - » 10µm for a 10-bit converter
- **Commercially, the IMEC FUGA 15 offers a single on-chip 8-bit flash ADC**
- **And the IVP MAPP2200 uses 256 parallel 8-bit ADCs**

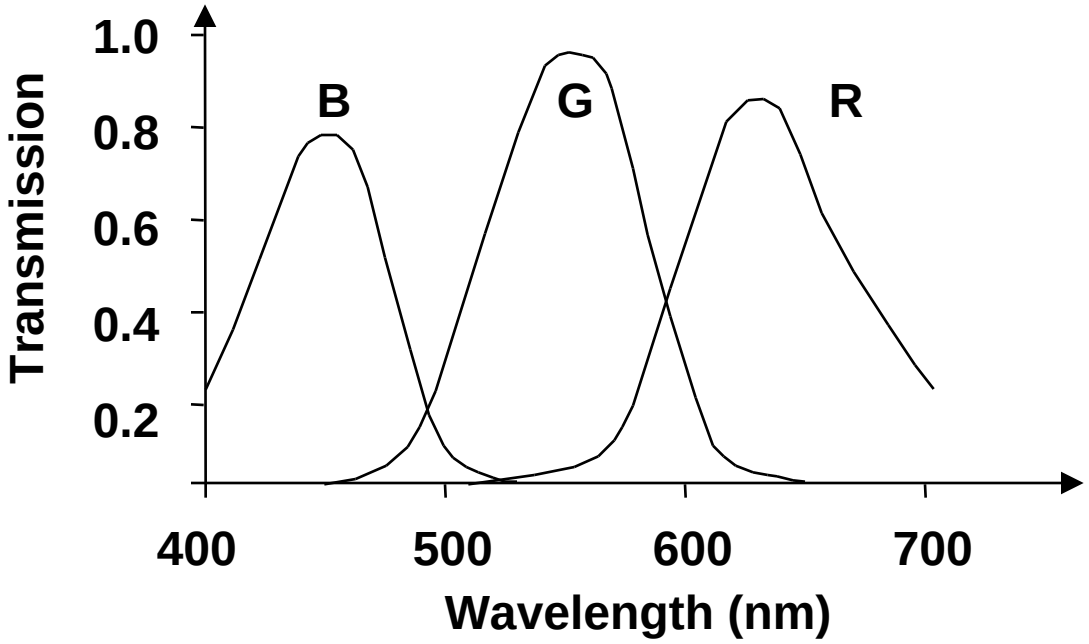
- **Stanford University has been developing in-pixel ADCs**
  - » first a ∑-Δ ADC per pixel
  - » with 22 transistors per pixel
- **And then a ∑-Δ ADC for a block of 4 pixels**
  - » with 4.25 transistors per pixel
  - » 1 FET for each passive pixel
  - » 13 for the ADC
  - » and the signal is reconstructed by counting pulses off-chip (e.g. how many “1”s in 8 clock cycles)
  - » fill factor 30% for 20x20µm pixel
  - » published ADC read time is 1ms per pixel

Colour Processing

- **For most consumer applications, colour images have become mandatory**
- **While the technology to sense colour has been available for a long time, the on-chip colour processing and video encoding is only now becoming feasible**
- **One method for obtaining very high quality colour images is to use three separate sensors**
  - » and to split the light into primary colours (red, green, blue) by a prism
  - » or to use a single sensor and a colour filter wheel
- **The principle of operation is that additive mixing can reproduce the entire spectrum of colours**
- **An important factor when designing colour imaging systems is that the human observer is part of the system**
  - » and so human perception must be considered when determining the performance of the sensor
  - » the human colour perception has been standardized since 1931!

Colour Filter Arrays

- **For single-chip colour sensing, individual pixels in the array are covered with coloured filters**
  - » to form the colour filtered array (CFA)
- **The spectral responses of conventional R, G, B filters are sketched below**



- **An alternative to the RGB system is to use the complementary colours, which have a higher percentage transmission**
  - » yellow (Ye) = R + G = W - B
  - » magenta (Mg) = R + B = W - G
  - » cyan (Cy) = G + B = W - R
  - » where W = white (transparent)

CFA Arrangement

- **Coloured filters are typically arranged either in stripes or in mosaics (checkerboards)**
  - » the mosaic is often called a Bayer pattern after its inventor (and 1976 patent holder)

Stripes

R G B	R G B G	Ye G Cy G
R G B	R G B G	Ye G Cy G
R G B	R G B G	Ye G Cy G
R G B	R G B G	Ye G Cy G

Mosaics

R G R G	R G R G	G Mg G Mg
G B G B	R G R G	Cy Ye Cy Ye
R G R G	G B G B	Mg G Mg G
G B G B	G B G B	Cy Ye Cy Ye

(Bayer pattern) (modified Bayer)

- **While the stripes have some advantages in terms of simplicity of fabrication, the horizontal resolution suffers seriously**
  - » and all CMOS imagers use the Bayer patterns
  - » recall that the VVL 800K-pixel colour sensor uses RGB and “teal” (blue/green) to achieve a proper colour balance with their process

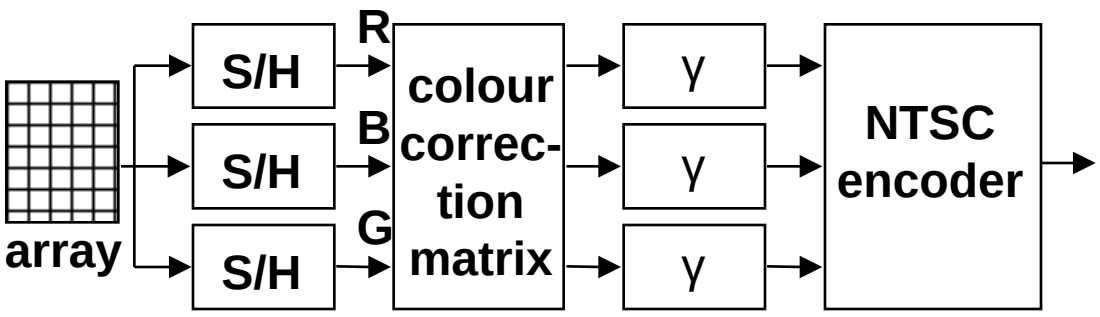
Matching Human Response

- **The human eye derives most of its spatial information from the green part of the spectrum**
- **So the standard reponse calls for 59% of the luminance (brightness) signal to come from green, 30% red, and 11% blue**
- **But, if only 11% of pixels detect blue light, the spatial resolution is too low, so the compromise solution is the Bayer pattern with 50-25-25**
  - » the modified Bayer pattern is sometimes used to facilitate the interlaced scanning by repeating the pattern on adjacent rows
- **The signals are then processed to achieve the correct colour balance**
  - » in NTSC this is known as matrixing

$$\begin{pmatrix} R_2 \\ G_2 \\ B_2 \end{pmatrix} = \begin{pmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \end{pmatrix} \begin{pmatrix} R_1 \\ G_1 \\ B_1 \end{pmatrix}$$

- **A selection of matrices may be available for use under various illumination conditions (scene colour temperatures)**
  - » e.g. tungsten lighting, twilight, full sun

- **The image processing required to NTSC colour processing is quite sophisticated, which is why a single chip version is only now available**
  - » from VVL (ISSCC98)
- **Another potential difficulty that arises from the use of CFAs is the low resolution of each coloured "array"**
  - » leading to the possibility of aliasing
  - » this can be corrected with off-chip filtering, at the expense of further image processing
- **Once the NTSC colour balance has taken place, the signals are then gamma corrected**
- **So the whole system may look as follows**



Advanced focal-plane processing

- **One of the most fascinating aspects of image sensor design is the constant battle to imitate nature**
- **It is a pretty good rule of thumb that if you want to find any engineering solution, look first at how nature does the job and try to imitate it**
  - » mechanics, structures, sensors
- **Of course, there are things that artificial constructions can do that nature cannot, and vice-versa**
  - » and nature often has the edge by combining physics, chemistry and engineering at microscopic level
- **Because of the importance of sight to humans, visual perception has always held a particular interest**
- **So it is not surprising that many attempts have been made to implement biologically-inspired vision systems in VLSI**
  - » these run the range from attempts to verify models of perception, through to engineers trying to satisfy an application

Human Eye vs. Silicon

(B. Dierickx, IMEC)

criterion	eye	CCD	CMOS APS	film
spectral response	400-700 nm peaked at 555	400-1000 nm smooth	400-1000 nm	300-700 nm
peak quantum efficiency	<20%	>50%	>50%	
dynamic range	1e6 logarithmic 1e2 linear	1e4 linear	1e4 non-linear	both
dark limit	0.001 lux 1E-6 W/m^2	typ: 0.1 lux <0.0001 possible	typ: 1 lux 0.001 possible	virtually zero
noise photons (*)	10	10	100	100
integration time (room temperature)	0.3 s	40 ms typical 5 minutes possible	40 ms typical +10 seconds possible	virtually unlimited
max. frame rate	ca. 15 Hz	10 kHz	>> 10 kHz	1 shot only

(*) noise photons is defined as: noise photons = noise electrons / (quantum efficiency * fill factor)

resolution

criterion	eye	CCD	CMOS APS	film
#pixels	120M	typical: 800K record: 60M	typical: 800K record: 4M	typical: 6e6 grains record: paper size
pixel pitch	2-3 µm	5-10 µm	5-10 µm	typ 10 µm grain pattern
focal plane size	3 cm	1 mm ...11 cm	1 mm ... 2 cm	only limited by film size

operating conditions

	eye	**CCD**	**CMOS APS**	**film**
radiation hardness	1 mrad	10 krad	1 Mrad	
operating Temp	36 C	-200 C...+200 C	0 K ... +200 C	0K ... 100 C
power dissipation	< 1mW	500 mW typ	50 mW typ	nihil

on-board image processing

criterion	**eye**	**CCD**	**CMOS APS**	**film**
cosmetic quality	perfect	very good	worse	as good as perfect
color	ideal	poor (only RGB)	poor	poor or print quality
absolute photometry	impossible	easy	easy	possible
focal plane processing	extensive	none	typically none	none
access method	data driven (focus of attention)	serial only	serial, random access, ...	optical only
datapath	5M nerves	8-10 bits	8 bits	none

logistics

criterion	eye	CCD	CMOS APS	film
price	invaluable	typ 10000 BEF	typ 1000 BEF	typ 10 BEF
2nd source	none	few	many	few
technology development cycle	500 Myears	5 years	2 years	20 years
fabs	3E9	15	1000	10

Artificial Eye?

- **Imaging chips made with standard techniques must be planar**
  - » so detector and processor must be side-by-side
  - » to maintain resolution and sensitivity, we usually divide the chip into sensor area and processing areas
  - » which limits the speed and parallelism of the process
- **The maximum area for our system is also limited by the fabrication technology**
  - » which will restrict the total amount of processing power available to us
  - » if we could include a state-of-the-art microprocessor and memory on-chip, the camera would appear much more intelligent! The low-level technology is there, but not the system level.
- **So there are two extreme approaches, and as yet only a few combined approaches**
  - » complex in- and inter-pixel processing at low resolution, with much internal control
  - » higher resolution, but less autonomous, peripheral processor-type devices that we have seen before

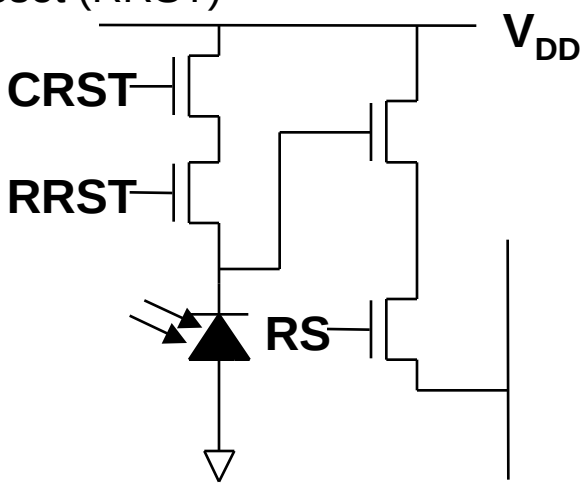
- **So, in this section, we will look at some of the more advanced “engineering” image sensors**
  - » these typically follow the peripheral processor architecture, but with adaptive or programmable capability
  - » but there are not too many of these
- **We will look at some of the things that it would be useful to do and how they are implemented currently**
- **Then we will look into the world of biologically-inspired image sensors**
  - » many designs have been suggested to accomplish both spatial and spatio-temporal image processing following “retinal” lines
  - » more than 50 of these have been collected and reviewed by A. Moini from University of Adelaide, Australia (see references for web page)

Increased Dynamic Range

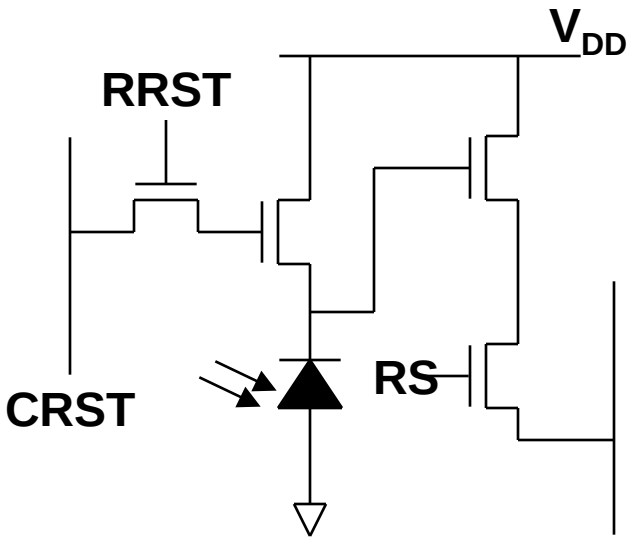
- **We saw in the previous chapter that a modification of the electronic shutter idea could give a sensor with two integration times**
  - » the WiDyR sensor from JPL
  - » the images could be recombined to offer a sensor with a wider dynamic range
  - » a similar concept has been used with CCDs (called the Hyper-D CCD) to give two exposures of different lengths, which are later combined
- **And one of the main reasons for tolerating the disadvantages of logarithmic pixels was the 10^6 dynamic range that could be achieved**
- **One of the potential problems with approaches such as the WiDyR is that of blooming**
  - » in the long t_{int} section, charge from bright pixels will spill over into neighbouring pixels
  - » making their signals indecipherable
- **So apparently the only other viable approach to increasing the dynamic range – by a circuit method – is**
  - » to control the integration time of each pixel individually

Individual Pixel Reset

- **Usually, photodiode pixels are reset by rows after the data has been read out**
  - » hence the natural idea of the electronic shutter
- **But the addition of another reset transistor can make the reset signal X-Y addressable too**
  - » this is achieved with a column reset (CRST) as well as a row reset (RRST)



- **The circuit above is the most obvious method (e.g. Ginosar & Gnusin)**
  - » but this introduces reset anomalies
  - » when CRST is pulsed, node N charges to V_{DD}
  - » this charge is transferred to the diode when RRST is pulsed, even though CRST & RRST were never on simultaneously



- **To remedy this reset problem, the circuit was modified as above**
  - » for use as a star tracker by JPL (Yadid-Pecht)
- **So now the idea is that we can control the integration time of each pixel**
  - » to allow poorly illuminated pixels more time
  - » and to read out and reset brightly lit pixels, thereby avoiding the blooming problem
  - » in this work, the integration time was controlled externally
- **This approach has several important implications for the sensor design**
- **Dark current becomes more of a limitation on the dynamic range than saturation, determining the maximum t_{int}**

- **The more profound departure from normal practice is that the time of the readout is now important, as well as the signal level**
  - » because two pixels may achieve the same signal level, but after different integration times
  - » this makes reconstruction of the image more tricky
- **While JPL's approach was to control the t_{int} externally, several of the "biological" designs allow the pixel to control its own t_{int}**
  - » and maybe those of adjacent pixels as well, to achieve a spatial image processing function rather than a wider dynamic range ***per se***
- **Ginosar & Gnusin proposed a complicated version of the dual-t_{int} approach**
  - » where a latch in each pixel informed control circuits which t_{int} applied to that pixel
  - » and this was adjusted for the next frame according to the value of the present frame relative to a threshold
  - » but the fill factor of a 60 x 70μm pixel was only 13%
- **Other attempts have led to similarly large pixel sizes**
  - » e.g. > 100 x 100μm by Miyagawa & Kanade

Image Compression

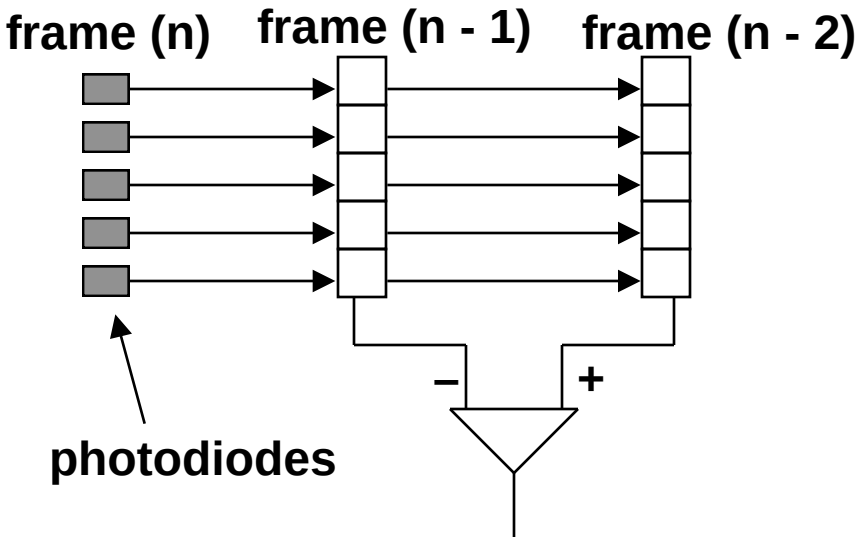
- **The quantity of data contained in images is so large that most transmission systems require the image to be compressed before transmission**
  - » 1000 x 1000 pixels gives 1MByte if 256 grey levels
  - » times however many frames per second
  - » a 38.8k bits/s modem can transmit one uncompressed (100 pixel)^2 8-bit image every ~ two seconds
- **There are many algorithms for compression**
- **Some are standard formats**
  - » JPEG (Joint Photographic Experts Group) for still pictures
  - » MPEG (Motion Picture Experts Group) for moving pictures
  - » H.263 low bit-rate standard, e.g. for video transmission over a telephone line
- **Some are not-yet-standard mathematical tools**
  - » wavelet compression, fractal-based compression
  - » e.g. Microsoft Encarta cd-rom encyclopaedia
- **And others are “natural” methods, such as**
  - » motion detection, foveation, data reduction

Mathematical Compression

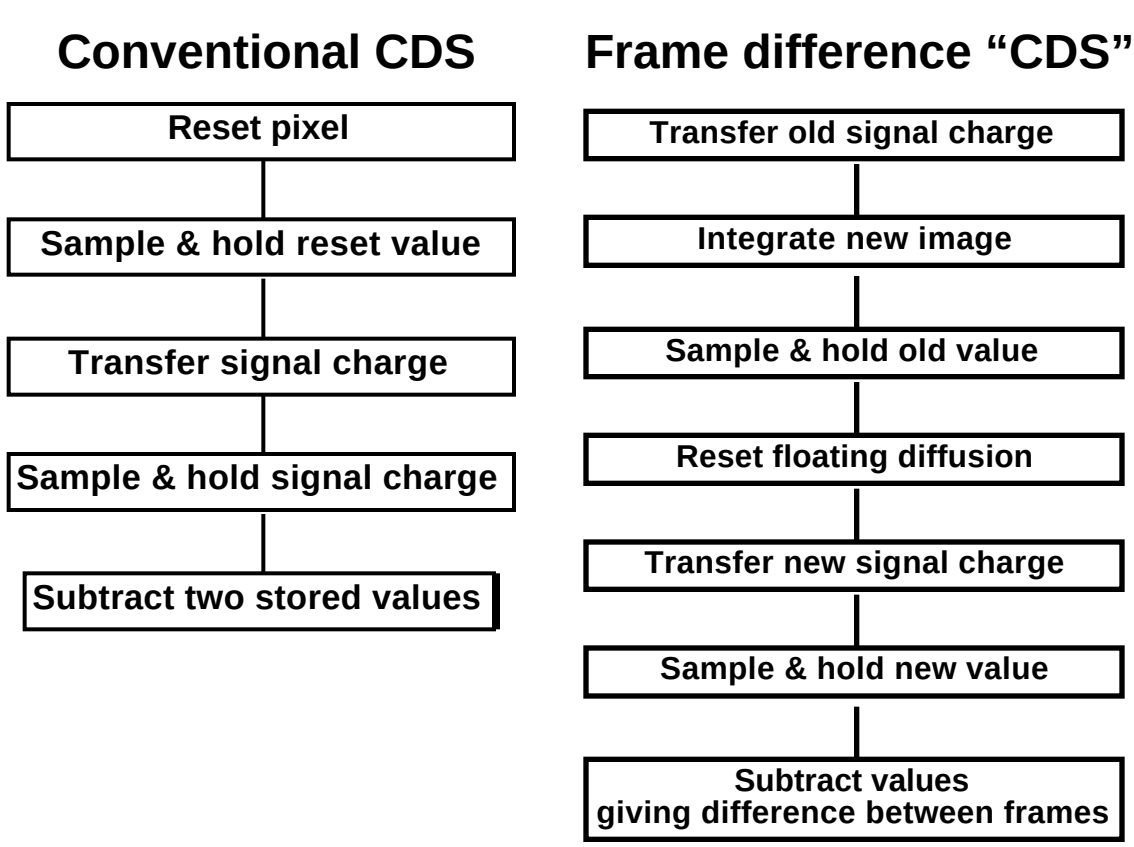
- **To date, there have been few reports of complex compression techniques being performed on-chip**
  - » JPEG encoding is possible in hardware, but has not yet been integrated with an image sensor
  - » MPEG encoding is hard to do on-chip, and the resulting chip would be large and have a high power consumption
- **One of the most recent reports by Kawahito et al. in Dec. 1997 has been of an imager with two-dimensional discrete cosine transform compression, in the analog domain**
  - » a 128 x 128 passive pixel array was read out in blocks of 8 x 8 pixels
  - » the DCT operation is performed by two successive weighted summations, each carried out in parallel, 8-bits at a time
  - » with intermediate results stored in an 8 x 8 memory
- **While the analog processing gave the advantages of speed and small silicon area**
  - » about 1mm^2 for the DCT processor in 0.35µm CMOS
  - » the precision of the method was degraded by mismatches between storage capacitors etc

Motion Detection

- **One of the simplest compression techniques is "motion detection"**
  - » maybe "difference detection" would be a better description (see later)
- **Here, only pixels whose signals have changed since the previous frame are transmitted**
  - » and the receiver can update just those pixels
- **The most straightforward technique for difference detection is frame differencing**
  - » where the newly captured image is compared with the previous image, which is stored somewhere
- **For the technique to be useful, the frame storage must be on-chip**
  - » either in each pixel
  - » or in a separate storage array, similar to a frame transfer CCD
- **In the case of CCDs, the two images are transferred to two readout registers which are subtracted at the output**
  - » ...



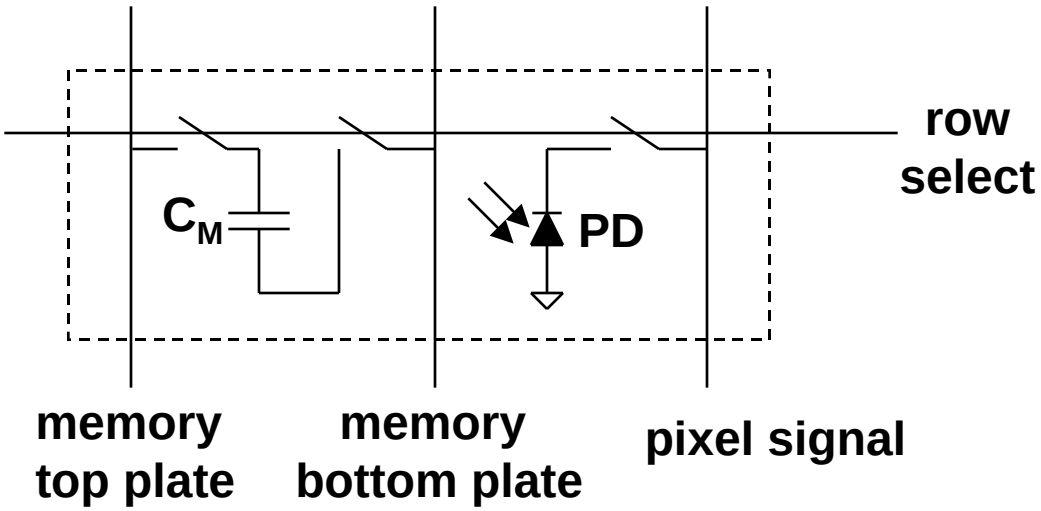
- **For CMOS image sensors, the in-pixel storage scheme is attractive**
  - » for the usual reason of high speed due to parallel processing
- **Usually, the drawback of in-pixel storage is the space required, usually a capacitor**
- **However, a simple and elegant method was implemented by Dickinson et al. using an unaltered photogate pixel**
  - » but changing the timing sequence slightly



- **So the old pixel value was stored on the output node while the new image is collected**
  - » and the CDS operation subtracts the old image from the new
- **Note that the "CDS" is no longer correlated, although FPN is removed**
- **And the accuracy of the method depends on how well the output node is isolated during the second integration**

In-pixel Analog Memory

- **For photodiode pixels, a separate storage node is needed within the pixel**
- **This has been implemented with passive pixels, with a pixel cell as follows (Simoni)**

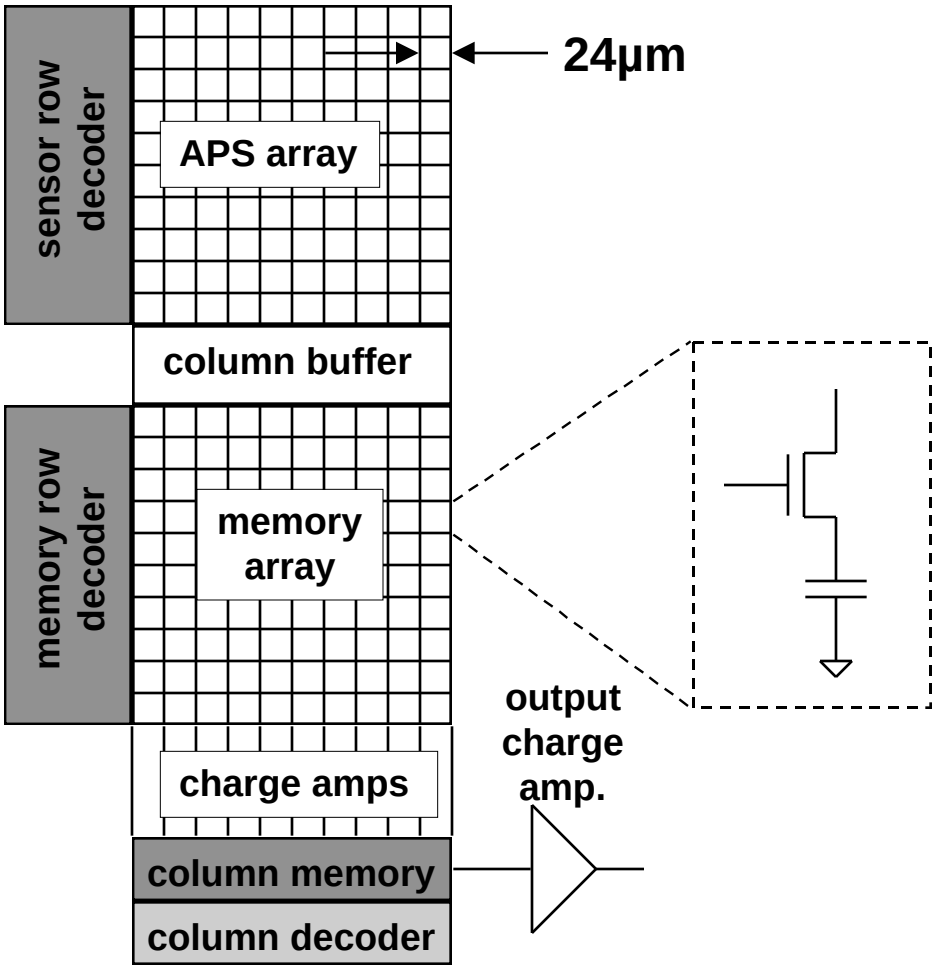


  - » the memory capacitor was designed such that $C_M \approx C_{PD} \approx 0.2pF$
- **Here, the main reason for including the capacitor in the pixel is**
  - » to avoid duplication of scanning electronics
  - » and to ensure that the parasitic capacitances are as equal as possible
  - » which is especially important for passive pixel sensors

- **In this design, a complex set of analog switches was used to transfer signals to the charge amplifier and subtractor from the pixel and memory**
  - » and back to the memory for storage
- **This scheme was demonstrated in 1.2μm CMOS for a 64 x 64 array of (48μm)² pixels**
  - » with a resultant fill factor of 43.5%
  - » and the relatively high power consumption of 10.8mW
  - » pictures of a moving hand were used to show the effect of frame differencing
  - » with a frame rate of 60Hz
- **One of the potential problems with in-pixel storage – in addition to the low fill factor – is a kind of smear effect**
  - » scattered light or electrons from the pixel can affect the integrity of the stored charge

Frame Transfer CMOS

- Recently, a frame transfer CMOS imager was presented for the first time by Zhou et al.
- In this design, the imager array was matched 1-for-1 with an analog memory array



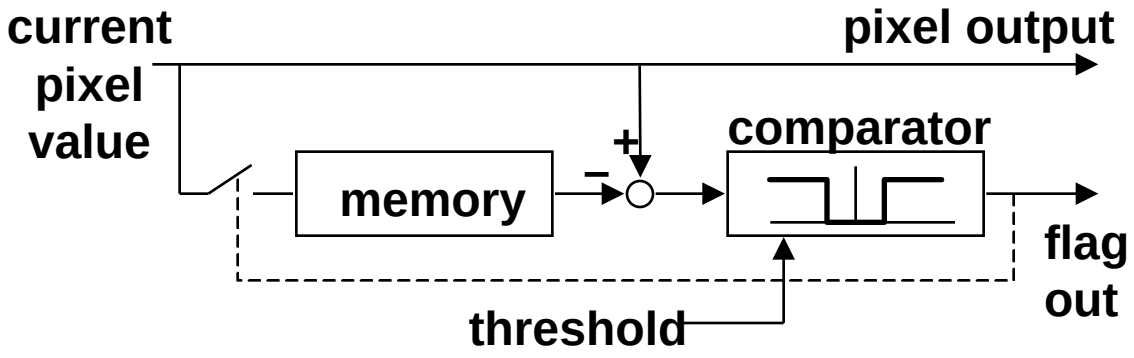
- **Each row of the sensor is transferred in parallel to the corresponding memory row**
  - » taking ~ 2μs per row
  - » unlike a FT-CCD where the transfer register is still illuminated during the transfer, there is no smear here
  - » and rows for transfer can be selected at random if desired
- **The main application for this FT-APS was for <u>pixel binning</u>**
  - » binning has been used in CCDs for many years to increase signal to noise ratio at low signal levels
  - » the signal from a number of vertically adjacent pixels are added together
  - » and SNR increases by a factor ~√N, where N is the number of pixels binned
  - » at the expense of resolution
- **This device was demonstrated for a 32 x 32 array in 1.2μm CMOS process**
  - » with $(24\mu m)^2$ pixels and a fill factor for the photogate APS of 29%
  - » and SNR improved from 15 to 26 when 15 pixels were binned (which is less than √N)

Disadvantages of FT

- **The primary disadvantage of FT as a means for recording frame differences is simply the space required for the memory**
  - » is is hard enough to fit in a large array, without effectively doubling the area by including a memory
  - » this makes FT-APS impractical for consumer applications with reasonable resolution
- **At the University of Waterloo, we are looking at several ways of detecting motion without these disadvantages**
  - » e.g. for using motion to control selective scanning of the imager
  - » so “interesting” regions of a large field of view are captured, keeping data transmission to a minimum by discarding “uninteresting” parts of the scene

Computational Compression

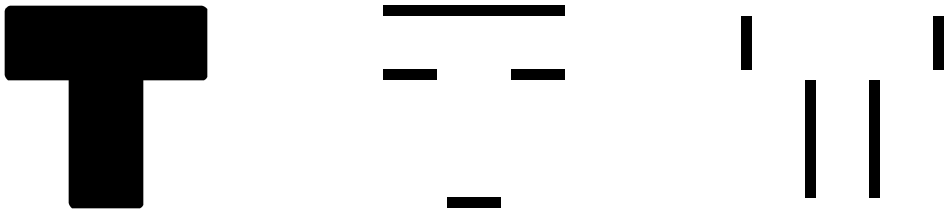
- **Aizawa and co-workers have developed a compression system which is "engineering" in philosophy, but at the "biological" level of focal-plane processing**
- **They argue that the stored-frame differencing approach is insensitive to slow motion (small differences), and prone to circuit mismatches**
  - » their alternative is based on conditional replenishment



- **The idea is that the memory element stores the last recorded value of the pixel**
  - » using a global read signal, the new pixel value is sampled and compared to this stored value
  - » if the magnitude of the difference is greater than a preset threshold, the new pixel value is output and stored in the memory
  - » and a flag is set to indicate that the pixel has a new value

- **The flag indicates to a smart shift register whether the pixel is ready to be read out**
  - » bypass switches are used to skip sections of the shift register
  - » and the X-Y coordinates of active pixels must also be recorded
- **Two modes of operation are possible**
  - » constant threshold, variable number of flagged pixels
  - » constant number of flagged pixels, variable threshold
  - » the optimum conditions depend on the motion rate and the desired frame rate
- **Operation has been demonstrated in both pixel-parallel and column parallel configurations**
  - » with a 32 x 32 array
  - » as usual, the trade-off is between speed, and fill factor & power consumption
  - » the column parallel architecture was also found to be less noisy
- **In the pixel parallel scheme, the fill factor for a $(160\mu m)^2$ pixel was 1.9%, with 33 transistors per pixel**

- **The motion detection was demonstrated for a "T" shape**



"T" flashing **"T" moving up-down** **"T" moving left-right**

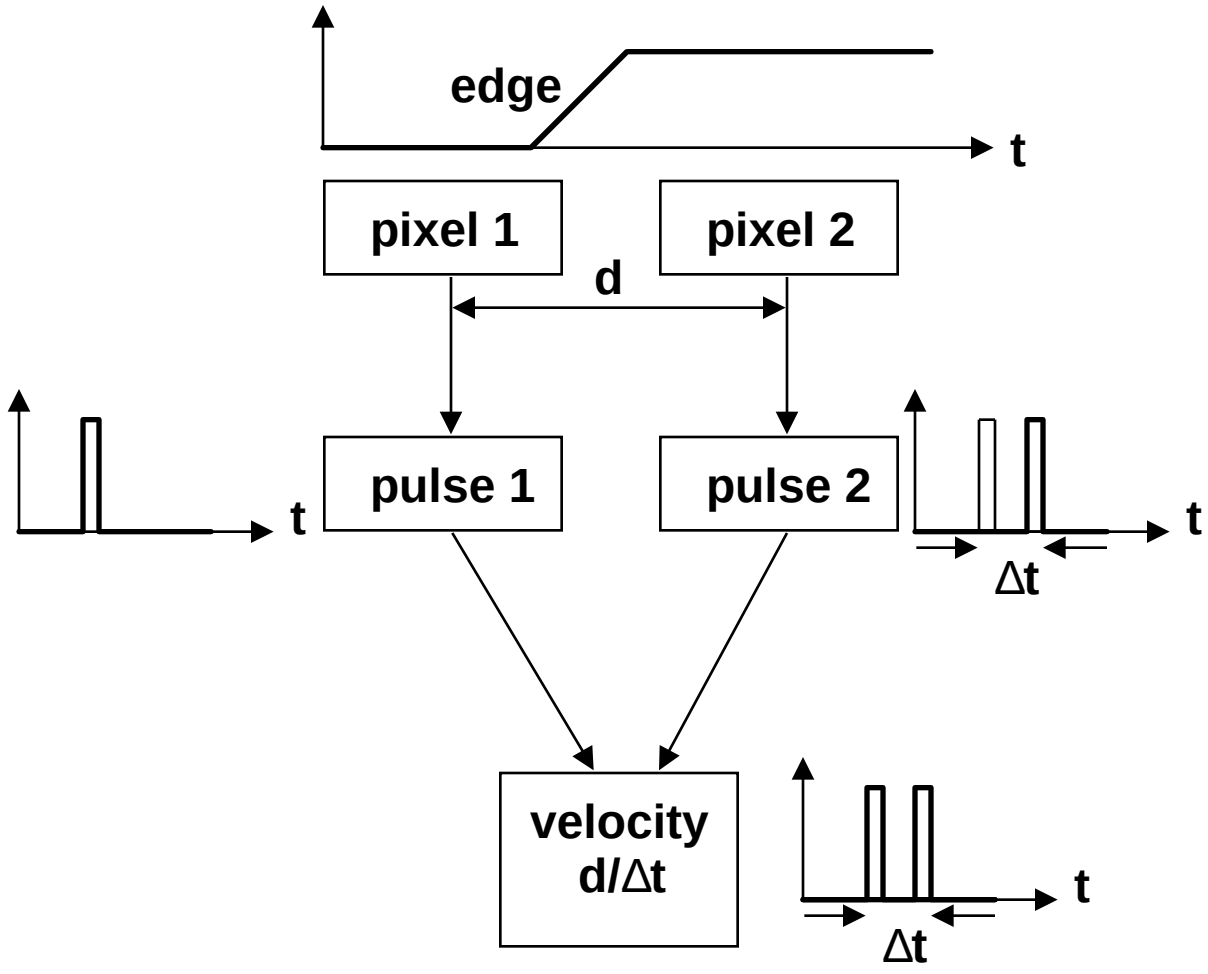
- **This approach has also been extended to control of the pixel integration time**
  - » the pixel integrates until either saturation or motion occurs
  - » at which point, it activates its flag and is reset
- **Here, the idea is both to extend the dynamic range and to reduce the effects of motion blur**
  - » but again, the pixel-parallel approach led to a large number of transistors per pixel (27)

Principle of Motion Detection

- **Strictly, motion detection should determine the optical flow in both space and time**
  - » employing an algorithm to interpret the motion as a whole, rather than just differences between pixels
  - » many “true” motion detectors are not scanned, in contrast to conventional imagers
- **Frame-subtraction schemes only measure snapshots of the scene at discrete intervals**
  - » and so measure simply differences in illumination rather than an overall evolution of the scene
- **Detecting motion is considered to be one of the most difficult of the VLSI imaging problems**
  - » the temporal intensity variations in the image are generally small, and tough to determine reliably
  - » chips tend to be complex because of the storage and/or delay elements required
  - » and the significant amount of analog signal processing involved
- **Three basic types of motion detection are**
  - » elementary motion detector unit
  - » delay and correlation techniques
  - » real-time differentiation

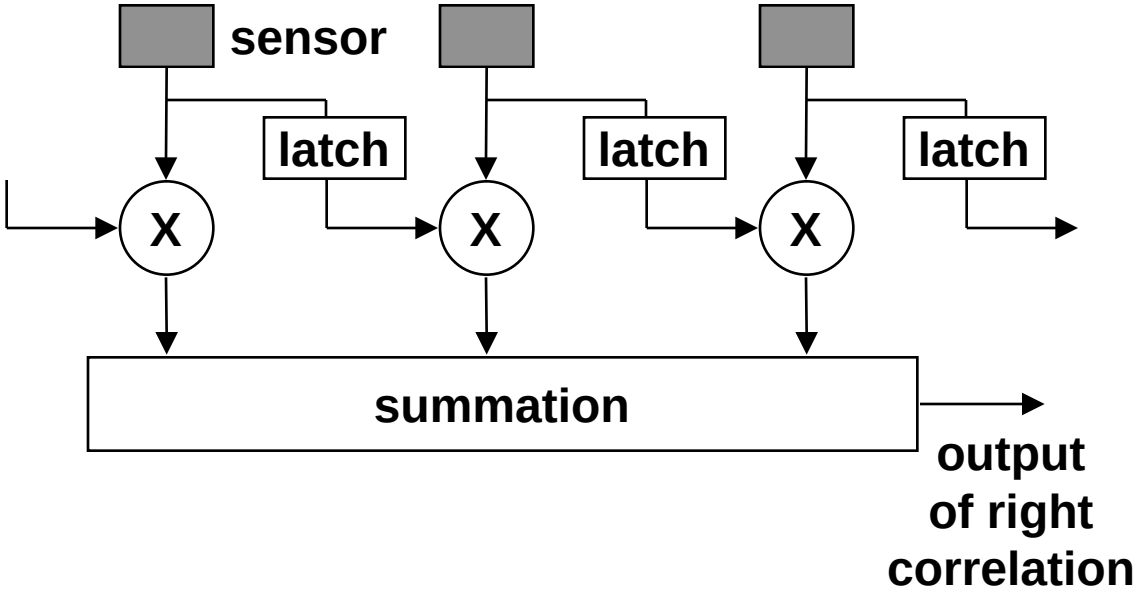
Elementary Motion Detector

- **Many biologically inspired motion detection systems use the elementary motion detector (EMD) unit**
  - » consisting of a pair of sensors which together detect the motion of an edge



Correlated Motion Detection

- **A simple version of a delay & correlate motion sensor was demonstrated by Tanner & Mead in 1984**

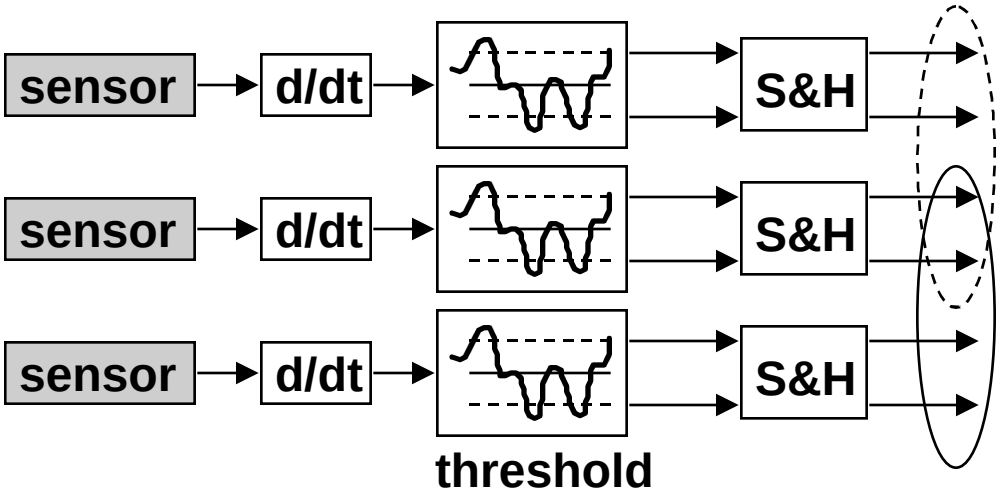


- **At time 1, the outputs from the detectors are thresholded and the 1s or 0s are stored in the latches**
- **At time 2, the latched signals are correlated (multiplied) with the new outputs from the sensors**
  - » and the resulting values summed up for the entire row

- **So, if an edge of bright illumination is moving to the right, the correlated output is relatively large**
  - » because the latched signal is more likely to be a 1, an the current sensor value is therefore more likely to be included in the output
- **Separate rows are needed to provide left-correlation (latches multiplied with left neighbour) and zero correlation (latches multiplied with same pixel)**
- **This idea is essentially an adaptation of the elementary motion detector unit**
  - » in that signals from adjacent pixels were compared
- **But it allows for all elements to contribute to the overall motion detection, rather than treating each pair individually**
- **This idea was adapted for detecting a particular velocity – velocity tuned – by Delbrück**
  - » the timing signals were replaced by a delay element
  - » so that signals which moved “in phase” with the delay were enhanced
- **Many variations on the same theme have been implemented**
  - » see Moini for a full review

Temporal Differentiation

- **Continuous differentiation of the sensor output is another way of achieving motion detection**
  - » the detection of true motion still requiring some coordinated response to these signals
- **In a non-scanned arrangement, for example, the pixel can simply send out a pulse if a significant change in illumination is recorded**
  - » this has been demonstrated in current-mode by Chong et al. but with no motion-detecting algorithm
- **Similar work is reported by Laval University, where groups of four pixels are used to calculate the velocity of motion**
- **A more sophisticated version of the continuous scheme, inspired by insect vision, was implemented by Moini et al.**
  - » and relies on a template tracking approach to reconstruct the true motion information
  - » ...



» the sensor output is differentiated, and fed to a threshold unit

» three outputs are possible – increase, decrease, same – which can be encoded in 2 bits

» an adjacent pair of 2-bit outputs has 9 possibilities

» and each pair is measured at two times to give a template with 81 possible configurations

- **Six template patterns of interest are tracked to follow the evolution of the low-level motion**
- **A critical issue in continuously differentiating systems is the sensitivity of the differentiator**

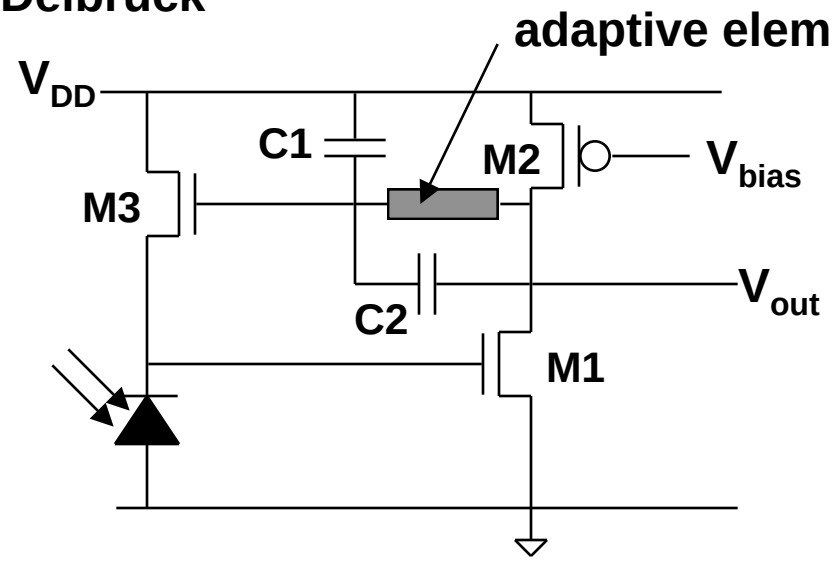
» a simple RC differentiator needs a high rate of change of light intensity to give a significant output

» this necessitates the inclusion of an amplifier, with the resulting increase of pixel complexity

- **Moini's system consisted of a 64-element linear array, thereby allowing unlimited space for the circuitry associated with each sensor**

Adaptive Photocircuits

- **Particularly in motion-detecting chips, a circuit with two modes of operation is desirable**
  - » over relatively long time periods, a logarithmic compression helps cover a large dynamic range
  - » while a high gain is useful for higher frequency variations to permit sensitive motion detection
- **One well known adaptive circuit was presented by Delbrück**



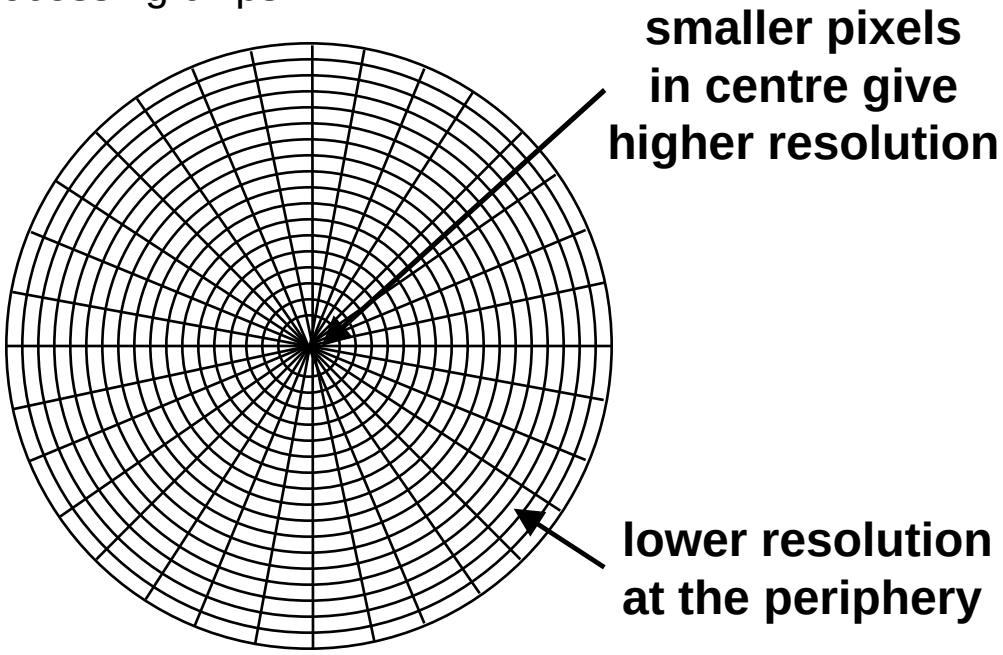
- **For slowly varying signals, the response of the circuit is determined by the adaptive element**
  - » if it is a large resistor, the circuit gives logarithmic compression, with feedback
  - » essentially the same circuit has been used recently by Huppertz et al.

- **At higher frequencies, however, the feedback is dominated by the capacitor divider of C1 and C2**
  - » so the gain can be higher for rapidly varying signals
  - » thus achieving compression for a wide dynamic range and sensitivity for motion detection
- **Delbrück determined that the optimum device for the adaptive element was a PMOS transistor in sub-threshold**
- **It should be noted that adaptability includes many of the features we have already discussed**
  - » variable integration time (individual pixel reset, electronic shutter)
  - » automatic gain
- **Algorithms for the adaptation include the average brightness scheme used by VLSI Vision**
  - » in which pixel outputs above a certain threshold were counted
  - » as well as other simple methods such as maximum signal
  - » these approaches fall under a “system adaptation” category

Spatial Processing

- **In our discussion, we have moved from motion/difference detection as a way of solving engineering problems, to architectures which attempt to imitate biological systems**
- **These spatio-temporal approaches essentially measure temporal effects, as a function of spatial coordinates**
  - » and do not generally include spatial processing of the image
- **Spatial processing is usually taken to mean an array of pixels in which there is an interaction between neighbouring pixels to process the image in some way**
- **This imitates the complex processing in mammalian eyes**
  - » such as inhibition of adjacent sensors, in order to increase local contrast
  - » and edge detection etc.
- **As yet, no high-resolution solution has been proposed to implement such inter-pixel processing**
  - » but a few engineering approximations are available

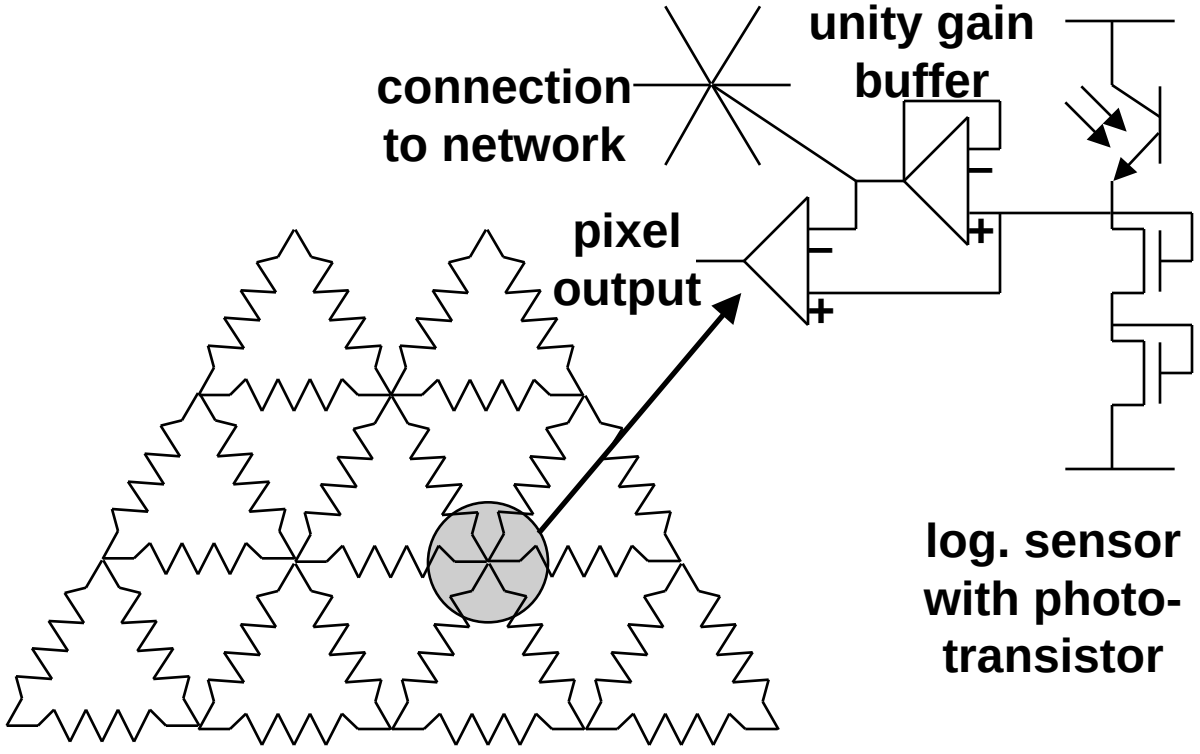
- **Architectures such as IMEC's foveated CMOS chip (and a CCD version) provide physical foveation**
  - » and are included under the loose heading of spatial processing chips



  - » IMEC have produced prototype colour versions of the foveated sensor, but colour processing from the large, low-resolution peripheral pixels is difficult
- **The MAPP2200 from IVP, and its Linköping University predecessors, is the other clearly practical design**
  - » but, unlike the true "silicon retinas", the processing is only in column parallel
  - » and hence is, at best, only 1-dimensional

Silicon Retina

- **The most popular approach to 2-D spatial processing is to approximate a Gaussian weighting function (e.g. Mead)**
  - » so that a particular sensor is affected most by its nearest neighbours and progressively less so by those further away
- **Commonly, this approximation takes the form of a resistive network**
  - » each sensor contributes to the signals at a node on the network, and its individual output is modified by the signal at that node



- **Here, the unity gain buffer sends the sensor output onto the network**
- **The differential amplifier (or a comparator) senses the node signal (including the sensor's own contribution) and modifies the final output of the pixel accordingly**
  - » thus, strong signals from adjacent pixels will inhibit the output
  - » but less so if this pixel also has a strong output itself
- **The resistors used in such a network are not likely to be made from passive resistors but rather from an active MOSFET circuit**
- **As summarized in Moini, other implementations include adaptive sensors, or active circuits to emulate the Gaussian function better**
- **Typically, each of these circuits has a few tens of pixels and is designed to imitate a particular aspect of animal vision**
  - » animals often display more complex functionality, such as two separate smoothing systems, some of which have also been implemented electronically
- **As yet, there has been no practical, compact 2-D spatial processing architecture**

Outlook and Conclusion

Technology

- **Is <0.5µm really necessary?**
  - » scaling issues suggest staying at about the 0.5µm level
  - » plenty of ≥0.5µm Fabs available for the foreseeable future
  - » many opportunities for innovation at the circuit/system level
- **Enhanced CMOS processes?**
  - » limited access & second sources – cost
  - » is performance that much better? If image quality were a priority, would CMOS be the technology of choice anyway?
- **Hybrid technologies?**
  - » this will depend on whether they really offer significant advantages
  - » this has yet to be proven
  - » the sales pitch for e.g. CCD/CMOS seems to fit in the “same as before, only better” category

Consumer

- **Single-chip colour video cameras are now a reality – what next?**
- **Higher resolution**
  - » at reasonable frame rates for video
  - » and for digital still cameras
- **Reduced die size**
  - » for lower costs
- **Digital output**
  - » both still and video
- **Standard compression formats on chip**
  - » digital signal processing
- **Testing and packaging**
  - » ultimate cost:performance is heavily influenced by packaging costs, especially for large area imagers
  - » testing and design for test (DFT) will become particularly important factors in mass market applications

Industrial/Scientific

- **The high levels of integration offered by CMOS cameras has not yet had a significant impact on the industrial market**
- **Technologically, what is needed?**
- **Increased adaptability & autonomy**
  - » increased on-chip feedback
- **Programmable functionality**
  - » digitally controllable signal processing
- **Intelligent power control**
  - » more important as chip complexity increases
- **2-D spatial processing**
- **For scientific applications, CMOS sensors still suffer from too many non-uniformities**
  - » but may be suitable for some low-cost applications

The Final Analysis

- **CMOS imager design has not yet reached maturity**
  - » consistent patterns in design and technology have yet to emerge
  - » but this allows for plenty of innovation
  - » as things settle down, imager design will be increasingly well supported by tools and fabs
- **We have not yet seen the final niche for CMOS imagers**
  - » toy video cameras are just the start
  - » a major application will emerge: internet video conferencing/video e-mail; automotive; security?
- **But it seems that CMOS imagers are here to stay – this time**
  - » the technology has reached a sufficient level to make CMOS imaging competitive in many, if not all, areas
- **So will we still be here in 5 years?**
  - » yes, but the emphasis will be different
  - » less diversity, increasingly subtle design details, more defined application-specific architectures
- **See you in 2003**

References – Part IV

- B. Olsen et al. (1997), IEEE Workshop on CCDs and AIS, Bruges, Belgium, June 5 -7, 1997
- www.vvl.co.uk
- www.ivp.se
- www.imec.be
- O. Yadid-Pecht et al. (1997), “Wide intrascene dynamic range CMOS APS using dual sampling”, IEEE CCD and AIS Workshop, Bruges, Belgium, June 5 - 7, 1997
- K. Parulski (1985), “Color filters and processing alternatives for one-chip cameras”, IEEE Trans. Electron Devices **ED-32**, 1381
- B. Fowler et al.(1994), “A CMOS area image sensor with pixel-level A/D conversion”, ISSCC94, p. 226
- D. Yang et al. (1996), “A 128x128 pixel CMOS area image sensor with multiplexed pixel level A/D conversion”, CICC96
- A. Moini (1997), “Vision chips”, internal report, University of Adelaide, Australia. http://www.eleceng.adelaide.edu.au/Groups/GAAS/Bugeye/visionchips/index.html
- R. Ginosar & A. Gnusin (1997), “A wide dynamic range CMOS image sensor”, IEEE CCD and AIS Workshop, Bruges, Belgium, June 5 - 7, 1997

- R. Miyagawa & T. Kanade (1995), “Integration-time based computational imager sensors”, IEEE CCD and AIS Workshop, Dana Point, CA, 1995
- O. Yadid-Pecht et al. (1997), “CMOS active pixel sensor star tracker with regional electronic shutter”, IEEE J. Solid State Circuits **32**, 285
- S. Kawahito et al. (1997), “A CMOS image sensor with analog two-dimensional DCT-based compression circuits for one-chip cameras”, IEEE J. Solid State Circuits **32**, 2030
- A. Dickinson et al. (1995), “A 256x256 CMOS active pixel image sensor with motion detection”, Proc. 1995 IEEE Solid State Circuits Conference, 226
- A. Simoni et al. (1995), “A single-chip optical sensor with analog memory for motion detection”, IEEE J. Solid State Circuits **30**, 800
- Z. Zhou et al. (1997), “Frame-transfer CMOS active pixel sensor with pixel binning”, IEEE Trans. Electron Devices **44**, 1764
- K. Aizawa et al. (1997), “Computational image sensor for on-sensor compression”, IEEE Trans. Electron Devices **44**, 1724
- K. Aizawa et al. (1997), “Pixel parallel and column parallel architectures and their implementations of on sensor image compression”, IEEE CCD and AIS Workshop, Bruges, Belgium, June 5 - 7, 1997

- T. Hamamoto et al. (1997), “Motion adaptive image sensor”, IEEE CCD and AIS Workshop, Bruges, Belgium, June 5 - 7, 1997
- Tanner & C. Mead (1984), “A correlating optical motion detector”, MIT Advanced Research in VLSI, p.57
- T. Delbrück (1993), “Silicon retina with correlation-based velocity-tuned pixels”, IEEE Trans. Neural Networks **4**, 529
- Chong et al. (1992), “Image motion detection using analog VLSI”, IEEE J. Solid State Circuits **27**, 93
- M. Arias-Estrada et al. (1996), “A focal plane architecture for motion computation”, J. Real Time Imaging ??
- A. Moini et al. (1997), “An insect vision-based motion detection chip”, IEEE J. Solid State Circuits **32**, 279
- T. Delbrück (1993), “Investigations of visual transduction and motion processing”,PhD thesis Caltech 1993, Chapter 2
- J. Huppertz et al. (1997),”Fast CMOS imaging with high dynamic range”, IEEE CCD and AIS Workshop, Bruges, Belgium, June 5 - 7, 1997
- e.g. Pardo et al. (1997), “CMOS foveated image sensor: signal scaling and small geometry effects”, IEEE Trans. Electron Devices **44**, 1731
- C. Mead (1989), “Analog VLSI and neural systems”, Addison Wesley publishers