#### Part III: Noise in Image Sensors

#### Introduction

- We have seen how pixels are designed to maximise the sensitivity to illumination
- However, this is only part of the story
- The overall performance of the sensor is ultimately limited by the noise that is added by the system to the signal
- In this sense, the noise figure of the detector system is a measure of its "perfection"
- Noise comes from numerous sources and its minimisation requires optimisation of many individual parts of the system
- Our discussion will not consider external noise sources, such as electrical pick-up
  - » the only "external" noise we will include is noise in the optical signal itself
- The treatment of noise is a complex subject, and it is even harder to measure the individual components accurately
  - » the theoretical treatment is important, however, as a design tool for optimising the performance of specific stages in the system

### **Types of Noise**

- "Noise" in image sensors is typically separated into two categories
  - » random noise
  - » pattern noise
- Random noise is what you might call "real" noise
  - » it is temporally random and is not constant from frame to frame in the image
  - » hence, it can be reduced by averaging successive frames
  - » and is described by statistical distributions
- Pattern noise is effectively a spatial noise as seen by the observer of the image
  - » it does not change significantly from frame to frame
  - » and so cannot be reduced by frame averaging

#### • Pattern noise is divided into two components

- » fixed pattern noise (FPN)
- » photo-response non-uniformity (PRNU)

#### **Pattern Noise**

- FPN is the component of pattern noise measured in the absence of illumination
- It is mainly due to variations in
  - » detector dimensions
  - » doping concentrations
  - » contamination during fabrication
  - » characteristics of MOSFETs (V<sub>T</sub>, gain, W, L, etc.)
- PRNU is the component of pattern noise that depends on the illumination

#### PRNU depends on

- » detector dimensions
- » doping concentrations
- » thicknesses of overlayers
- » wavelength of illumination (spectral response)
- Historically, pattern noise (FPN in particular) has been the factor limiting the acceptability of CMOS imagers
  - » PRNU is not often mentioned ...
  - » shortly, we will see how FPN can be reduced

### **Describing Noise**

- Pattern noise is usually specified in terms of the variation in the signals from individual pixels under uniform illumination
  - » usually as a percentage of the saturation output
- Random noise is expressed in terms of parameters which describe the statistical distribution of voltage or current
- If there are n samples of the signal

» x<sub>1</sub>, x<sub>2</sub>, x<sub>3</sub>, . . . x<sub>n</sub>

- then the mean is  $x = (x_1 + x_2 + x_3 + ... + x_n) / n$
- However, the mean for many noise sources is zero
  - » leaving the DC level of the signal unaffected
- So a more useful description of the noise is either the variance (<x<sup>2</sup>>) or the standard deviation ( <x<sup>2</sup>>, in rms units)
  - » which measures the scatter of the data points about the mean

$$\langle \mathbf{x}^2 \rangle = \frac{1}{n} \frac{n}{j=1} \left( \mathbf{x}_j - \mathbf{x} \right)^2$$

To sum noise sources, we have to add the variances

$$\left\langle x^{2}\right\rangle =\left\langle x_{1}^{2}\right\rangle +\left\langle x_{2}^{2}\right\rangle +\left\langle x_{3}^{2}\right\rangle +...\left\langle x_{n}^{2}\right\rangle$$

• or the standard deviation is given by

$$\left\langle x\right\rangle = \sqrt{\left\langle x_{1}^{2}\right\rangle + \left\langle x_{2}^{2}\right\rangle + \left\langle x_{3}^{2}\right\rangle + ...\left\langle x_{n}^{2}\right\rangle}$$

#### **Importance of Noise**

• We can illustrate the importance of the noise on the overall sensor performance as follows



- Dynamic range = (saturation signal / rms noise level)
  - » saturation 200,000 e<sup>-</sup>, noise 40 e<sup>-</sup> rms
  - » typical value is 5,000 for a PD (~75dB)
  - » assuming dark current is not the limiting factor

- Responsivity = (# electrons / light intensity)
  - » in linear portion of the curve (electrons.cm<sup>2</sup>/ $\mu$ W)
- Provided that the dark current is small, the minimum resolvable signal is determined by the noise in the system
- Hence, a good responsivity is not enough to ensure a good signal at low light levels
  - » a low "noise floor" is also required
- In a convenient model, the rms system noise is

$$\left< n_{sys} \right> = \sqrt{\left< n_{shot}^2 \right> + \left< n_{floor}^2 \right> + \left< n_{pattern}^2 \right>}$$

- » where the floor is determined by the amplifier noise, the reset noise, and the analog-to-digital converter noise
- » the noise floor is often referred to as the <u>read noise</u>
- The other noise included above is called the shot noise, which arises because of the statistical arrival of electrons
  - » due to the photo-generation of the electrons
  - » and the thermal generation of electrons
- We will now examine some of the noise sources present in image sensors

### **Thermal Noise**

- Thermal noise is a white noise
  - » the noise power is constant over all frequencies
- For a resistor, the thermal noise root mean square voltage is given by

$$\left< v_{th} \right> = \sqrt{4kTBR}$$

- » where R is the resistance, and B is the noise equivalent bandwidth
- Since the thermal noise covers the entire frequency range, the bandwidth determines the actual amount measured
- So the <u>open circuit</u> equivalent circuit is



• Alternatively

$$\left< \mathbf{i_{th}} \right> = \sqrt{\frac{4kTB}{R}}$$

 However, an important factor is the noise equivalent bandwidth for use in the calculation

# **Noise Equivalent Bandwidth**

- This is defined as the voltage-gain-squared bandwidth of the circuit
- The ideal case is that the (gain)<sup>2</sup> is constant at a value of A<sub>0</sub><sup>2</sup> up to the bandwidth (A<sub>0</sub> = voltage gain)



• Mathematically, this is given by

$$\mathbf{B} = \frac{1}{\left|\mathbf{A}_{0}\right|^{2}} \left|\mathbf{A}(\mathbf{f})^{2}\right| d\mathbf{f}$$

- So in the ideal case bandwidth =  $\frac{1}{A_0^2} |A|^2 df = \frac{A_0^2 B}{A_0^2} = B$
- If we take the example of an RC low pass filter



• Calculating the transfer function

A() = 
$$\frac{v_{out}}{v_{in}} = \frac{\frac{1}{j C}}{\frac{1}{j C} + R}$$
  
=  $\frac{1}{1+2 fRC}$  since = 2 f  
=  $\frac{f_0}{jf + f_0}$  where  $f_0 = \frac{1}{2 RC}$ 

- At f = 0,  $A(f) = A_0 = 1$  for this circuit
- Now we can calculate the noise equivalent bandwidth, using A<sub>0</sub> = 1

$$B = \frac{f_0}{\sqrt{f^2 + f_0^2}} df$$
$$= f_0^2 \left( f_0^2 + f^2 \right)^{-1} df$$
$$= -\frac{1}{2} f_0$$

 The reason for choosing this example is that it is directly applicable to the resetting of photodiodes and the output nodes of CCD and photogate pixels

#### **Reset Noise**

 If we consider a diffusion (either a floating diffusion or a photodiode) being reset through a MOSFET



• Effectively, this is a capacitance being charged through the resistance of the MOSFET channel



• So the ac-equivalent circuit is



• From before, the bandwidth is

$$\mathbf{B} = \frac{1}{2}\mathbf{f}_0 = \frac{1}{4\mathbf{R}\mathbf{C}}$$

• So we find the rms noise voltage

$$\left< \mathbf{v_{out}} \right> = \sqrt{\frac{\mathbf{kT}}{\mathbf{C}}}$$

- Usually, the noise voltages is expressed in terms of electrons, in order to compare directly with the electrons in the well
- In which case the reset noise on the capacitor is calculated from Q = nq = Cv<sub>out</sub>, and the rms noise <u>electrons</u> is given by

$$\left< n_e \right> = \frac{C}{q} \sqrt{\frac{kT}{C}} = \frac{\sqrt{kTC}}{q}$$

- This noise is generally called "kTC noise" or, in this case, reset noise
- Calculating this out at room temperature gives  $\left< n_{kTC,RT} \right> = 400 \sqrt{C(pF)}$
- For a floating diffusion C ~ 20fF, so n<sub>kTC</sub> = 55 e<sup>-</sup>
- For a (10µm)<sup>2</sup> photodiode, C ~ 60pF, so n<sub>kTC</sub> = 100 e<sup>-</sup>
  - » currently, reset noise limits the read noise in PDs

### Shot Noise

- Shot noise is another white noise that arises from the discrete nature of the electrons themselves
  - » i.e. the random arrival of particles of charge
- This is the result of the random generation of carriers
  - » either by thermal generation within a depletion region (i.e. shot noise of the dark current)
  - » or by the random generation of photo-electrons, caused in turn by the random arrival of photons
- The rms signal is given by

 $\left< i \right> = \sqrt{2 q I_{dc} B}$ 

- If the noise statistical distribution is described by a Poisson distribution
  - » the variance is equal to the mean

 So, if electrons are generated with a current density, J<sub>dark</sub>, in a sensor of area, A, over an integration time, t<sub>int</sub>, the shot noise variance is

$$\left\langle n_{dark}^2 \right\rangle = n_{dark} = \frac{J_{dark}At_{int}}{q}$$

• Similarly, the photo-electron shot noise variance is given by

$$\left\langle n_{pe}^{2}\right\rangle = n_{pe} = I_{0}At_{int}$$

- » where  $I_0$  is the photon flux (photons/cm<sup>2</sup>s) and is the quantum efficiency
- So the total rms shot noise contribution from the sensor is

$$\begin{split} \left\langle n_{shot} \right\rangle &= \sqrt{\left\langle n_{dark}^2 \right\rangle + \left\langle n_{pe}^2 \right\rangle} = \sqrt{n_{dark} + n_{pe}} \\ &= \sqrt{\frac{J_{dark}At_{int}}{q}} + I_0At_{int} \end{split}$$

- For example, with
  - »  $J_{dark} = 200 n A/cm^2$
  - »  $A = (10 \mu m)^2$
  - »  $t_{int} = 30ms$
  - »  $I_0 = 10^{13} \text{ photons/cm}^2 \text{s}$
  - » and = 0.5
- we find  $\langle n_{shot} \rangle = (37,500_{dark} + 150,000_{pe}) = 430 e^{-1}$

# Flicker (1/f) Noise

- At any junction, including metal-to-metal, metalto-semiconductor, and semiconductor-tosemiconductor, conductivity fluctuations occur
  - » the causes of these are still not completely understood
- The rms 1/f noise current is given by



- 1/f noise arises mainly in amplifier circuits where there are numerous such contacts
- At low frequencies, 1/f noise can be the dominant component
  - » but, at higher frequencies, the 1/f noise drops below the thermal noise
  - » the frequency at which this happens depends on the situation





## **"Referred" Noise Figures**

- Conventionally, the noise figures are referred either to the final output or to the output of the optical detector
  - » i.e. to be compared directly with the number of electrons generated by the detector
  - » called input referred noise
- For input referred noise, the noise of later stages must be divided by the gains of the intermediate stages
- Or vice versa for output referred noise
- Usually, authors in CMOS circles use the input referred figure
  - » but this is tough to obtain for intermediate stages in the circuit owing to uncertainties in the gains of each stage
  - only the overall figure in electrons is practically feasible because the appropriate inverse-conversion efficiency (e<sup>-</sup> per μV) is only known for the entire output circuit

# **Typical Noise Figures**

• From Mendis, the calculated and measured input referred noises for a 128x128 element photogate array are

Noise source	Calculated	Measured
	rms	rms
<b>kTC from reset of FD</b>	negligible	negligible
In-pixel amp. 1/f	111µV	
kTC from column	93µV	
	A ( <b>X</b> 7	
Column source	46µ V	
follower 1/f		
Total column noise	86µV	120µV
Total noise	152µV	151µV
Total noise electrons	41 e <sup>-</sup>	41 e <sup>-</sup>

- Mendis also reported a photodiode read noise of ~80 e<sup>-</sup> rms
- Typical read noises for CMOS sensors

Technology	<b>RMS read noise</b>
Photodiode APS	50 - 80 e <sup>-</sup>
Photogate APS	20 - 40 e <sup>-</sup>
Logarithmic APS	700 e <sup>-</sup>
Passive pixels	200 - 300 e <sup>-</sup>

 Remember that this does not include shot noise or pattern noise

#### **Fixed Pattern Noise**

- Fixed Pattern Noise is due to pixel-to-pixel variations in the absence of illumination
- The main cause of FPN in CMOS imagers is variations in  $V_{\rm T}$ 
  - » between reset and buffer MOSFETs in the pixel
  - » and between MOSFETs in the column circuits
- FPN can also arise from repeating irregularities in the array clocking
  - » allowing small variations in integration time etc.
- In very large arrays, resistive drops in reset buses may lead to a "droop" in the voltage to which the pixels are reset,
  - » but this is not usually significant in CMOS imagers
- FPN is just as valid as a "noise" as the temporal variety
  - » both affect the actual output voltage that the pixel produces
  - » in a way that is not directly related to the illumination to be measured

# PRNU

- The issue of photo-response non-uniformity has not historically received much attention in the CMOS imager community
  - » although there is now some occasional mention of "gain nonuniformity"
- Like FPN, PRNU is essentially time-independent, but it is signal-dependent
- Both types of pattern noise can be specified in terms of either an rms or a peak-to-peak value, referenced to an average value
  - » e.g. the full-well capacity
- A histogram of output signals is built up in the dark or light, as appropriate
  - » PN<sub>rms</sub> = rms of distribution / average value
  - » PN<sub>p-p</sub> = peak-to-peak variation / average value
- Since PRNU is signal dependent, it is often expressed as a multiplier of the number of photons
  - $\sim < n_{PRNU} > = Un_{pe}$

### **Minimum Noise**

 In principle, the noise floor and dark current can be reduced so that the system is photon shot noise limited

$$\left< n_{sys} \right> = \left< n_{pe} \right> = \sqrt{n_{pe}}$$

- » this approximation is sometimes used to calculate the pixel sensitivity ( $\mu V/e^{\text{-}})$
- But there will never be zero PRNU, so a more achievable value would be

$$\left\langle n_{sys} \right\rangle = \sqrt{n_{pe} + \left( U n_{pe} \right)^2}$$

- » The worst case when  $n_{pe} = n_{full-well}$
- If we plot out this limiting noise as a function of PRNU, it looks like



# Noise Reduction Techniques

- Having seen some of the common sources of noise in CMOS imaging systems, how might we go about reducing them?
- Essentially, there are three classes of noise
  - » those we can do nothing about, such as photon shot noise
  - » those we can reduce by careful design of circuit components, such as thermal noise
  - » those we can reduce by circuit design, such as FPN
- These techniques are inter-dependent
  - » we shall see that adding extra circuitry to reduce FPN also introduces extra 1/f and kTC noise
  - » so the optimisation of noise is a system issue, not just a question of optimising each element individually
- We will look at the general techniques for reducing noise in electronic devices, as well as circuit techniques for pattern noise etc
- The study of noise is a specialised topic, and we will only look at the essentials

### Shot Noise

- As we remarked earlier, photon shot noise is dependent on the illumination level, and there is not much we can do about it
  - » except reduce the QE of the detector, which we don't want to do!
- Shot noise also arises from the pixel dark current
  - » which we can alter
- By changing doping levels, we can reduce the dark current
  - » but, in a regular photodiode, at the expense of QE
- And by removing the collection area away from the surface
  - » this is another advantage of the pinned photodiode

# • The magnitude of the dark current is, of course, dependent on the pixel area

- » so the shot noise will be smaller for smaller pixels
- » although the perimeter component of the dark current means that S/N still gets worse as pixel dimensions are reduced
- But shot noise is not usually the limiting factor

#### **Thermal Noise**

- Thermal noise is important mainly in the input stages of amplifiers
  - » because of the (4kTBR<sub>channel</sub>) from the MOSFETs
- In general, the power spectrum of the thermal noise will be proportional to (W/L)<sup>-1</sup>
- But it is also dependent on the current through the devices
- A common way of expressing thermal noise is the <u>noise electron density (NED)</u>

$$NED(f) = \frac{e_n(f)C_t}{q}^2$$

- » where e<sub>n</sub>(f) is the total equivalent noise voltage at the output stage (e.g. a floating diffusion)
- and C<sub>t</sub> is the total capacitance present at the input, including diffusion capacitance, gate capacitance, and everything else (to convert to electrons)
- e<sub>n</sub>(f) represents the device noise, referred to the input
  - » and so includes all factors such as the transistor geometry, i<sub>DS</sub>, device area etc. that affect the gain



- Increasing W, for a fixed L, increases C<sub>t</sub> because of the device area
- But increased W decreases e<sub>n</sub>(f)
  - » because the current i<sub>DS</sub>, and the gain of the circuit, are increased
  - » therefore reducing the input referred noise
- So the optimum condition is to keep W 15µm in these transistors
  - » too small a W reduces the gain
  - » and too large a W increases the C
- Also the NED decreases with L, which makes smaller devices advantageous

### 1/f Noise

- 1/f noise arises mainly from trapping detrapping of electrons at the Si-SiO<sub>2</sub> interface
- So we can do two things to minimise 1/f noise
  - » reduce the device area, W x L
  - » use a buried channel device to separate the channel from the interface

# • While in standard CMOS we cannot do much about the second option, we could reduce W x L

- » but the gain of the amplifier is dependent on W/L
- » so reducing L is the best choice!
- In the pixel, this is a good thing since we want to minimise the areas of the transistors anyway, to achieve a high fill factor
  - although care is needed to ensure that the pixel source follower can adequately drive the column capacitance (i.e. enough W/L)
- Note that we need the noise added at early stages of the process to be minimised
  - » since this is amplified at all subsequent stages

# **Correlated Double Sampling**

- Reset noise is difficult to design out of the system
  - » since the properties of the transistor cancel out
  - » although reducing the capacitance of the node is useful for both kTC and conversion efficiency
- So the most common solution is to measure the value of the reset noise and then to subtract it from the signal
- A generic circuit for achieving this in a CCD or floating gate APS would be



- During the sample and hold period, the appropriate switches are pulsed on & off
  - » to leave the voltages stored on the capacitors

 The sequence of events for a CCD or photogate would be



#### • Here, the reset signal is given by

»  $V_{reset} = [V_{DD} - (V_T \pm V_T)] \pm (V_{kTC}) \pm (V_{part})$ 

• Here the  $V_T$  is the component of FPN arising from mismatches between the reset transistors

» and is approximately the same for each frame

#### • V<sub>kTC</sub> is the reset noise

- » and is different from frame to frame
- Note that we are considering <u>voltages</u> (not electrons) at this stage
  - » so the "kTC" noise is given by (kT/C), and is therefore reduced for larger C

- V<sub>part</sub> refers to what is called the partition noise
  - » when the reset FET turns off, the channel charge moves either to the source (= FD) or to the drain
  - » but we do not know exactly how much goes to each
- This type of sample-and-hold technique is known as <u>correlated double sampling (CDS)</u>
- The "correlated" part comes about because the noise component of the two signals is correlated, and can therefore be subtracted out
- In a CCD, a single CDS circuit is needed because there is only one floating diffusion output node
- In CMOS APS, there is an output node per pixel
  - » but practically, we need only one CDS circuit per column of the array
  - » and the S&H is carried out for all columns in parallel



## **CDS for Photodiode APS**

- This form of CDS works very well for pixels with a floating diffusion output node
  - » photogate, and pinned photodiode with transfer gate
- Indeed, the main advantage of using the photogate structure is to facilitate the removal of reset noise
  - » since the improvement in conversion efficiency is offset by the lower QE
- In photodiode designs, double sampling can only remove the FPN that results from mismatches
  - » this is because the double sampling is <u>not</u> correlated
- In FD designs, the signal was added to the existing (and stored) reset value
  - » so the subtraction was of <u>exactly the same</u> noise signals
- In the photodiode, there is no separate output node, so the signal must be read out first
  - » and this signal includes the original reset voltage on the photodiode
  - » which in turn includes FPN and kTC noise

- We can now reset the pixel again and subtract this value
  - » the FPN will be much the same as that which was included when we sampled the signal
  - » but the kTC noise will be different, i.e. not correlated
  - » remember kTC is the rms value of a distribution

#### • So now the sequence of events is



- This would be better termed pixel double sampling
- Or, alternatively, a graphical representation is as follows ...



- This is the reason why reset noise is now the limiting noise source in photodiode circuits
- Note that even this noise reduction is unavailable to the logarithmic pixels, hence their poor FPN characteristics
- The conventional CDS circuit used in CMOS sensors is shown below



- Of course, the additional circuitry required for the CDS implementation adds further noise to the signal
  - » kTC from the sample-and-hold capacitors
  - » 1/f and thermal noise from the transistors
- But usually in CMOS sensors, the FPN is the more critical issue

### **Column FPN**

- The other issue with using column-wise CDS is that FPN is then added by the CDS circuits themselves
  - » appearing as vertical streaks in the image
- This can be removed by storing and subtracting column reference signals off chip
- Alternatively a second stage of double sampling is performed
  - » where, after the readout of the differential signal, the S&H capacitors are shorted together
  - » this results in a differential output that is a measure of the mismatches between the two sets of output stages
  - » Mendis calls this a "crowbar" circuit and the process delta difference sampling



## **Typical Figures**

- Typical figures for FPN are hard to define because it depends so much on the precise process used
- For photogates with a 2µm CMOS process, Mendis reported a p-p FPN of 1% – 2.5% saturation with the CDS circuit
  - » falling to ~0.1% sat. with the DDS as well
- A photodiode fabricated similarly showed a p-p FPN of ~0.5% sat. after CDS, and ~0.1% after CDS + DDS
  - » typical raw data are about 2 3% p-p sat.
- For a 0.35µm process, the raw FPN for a PG array was 6% sat.
  - » reducing to 0.4% after off-chip correction
- Mansoorian et al. give a final FPN of 0.6% sat. pp for both PG and PD using a 0.55µm process
  - » using a similar DDS technique
- For logarithmic pixels, IMEC report a raw FPN of ~100% of the useable signal range!

#### **Patent Issues**

- Other methods of reducing noise are possible, although probably not so good
- Hitachi have several patents which cover the idea of active pixel sensors and the use of CDS in these devices
  - » in their CMOS digital still camera, VLSI Vision use a mechanical shutter in order to measure a true dark image for subsequent subtraction

#### One possibility is to smooth out large signal variations between neighbouring pixels

- » a smooth curve is fitted through points either side of the test point, and the test point moved to fit that curve
- » the smoothing is improved if the number of neighbours is increased, but the "sharpness" of the image is lowered



### Feedthrough & Crosstalk

- We have considered here "natural" sources of noise such as 1/f, thermal, and shot noise
- And technological noise, such as FPN and PRNU
- In addition to these, there can be unwanted signals in one part of the circuit due to the operation of another part
  - » these can be addressed in the design of the array and circuits
  - » although some sources are not easy to eliminated
- Feedthrough of digital signals from control lines into the analog parts of the circuit can be a problem
  - » analog and digital sections of the chip can be separated to some extent
  - » but the array itself, and much of the analog signal processing is intrinsically both analog and digital
- The minimisation of these effects requires careful layout
  - » and mixed signal design is currently a hot topic in many areas, such as A-D & D-A conversion, DSP etc.

#### **References – Part III**

- » H.W. Ott (1988), "Noise reduction techniques in electronic systems", Wiley
- » G.C. Holst (1996), "CCD Arrays, cameras and displays", SPIE Press
- » T.E. Jenkins (1987), "Optical sensing techniques and signal processing", Prentice Hall
- S.K. Mendis (1995), "CMOS active pixel image sensors with on-chip analog-to-digital conversion", PhD Thesis, Columbia University, USA
- » B. Mansoorian et al. (1997), "Megapixel CMOS APS with analog and digital outputs", IEEE CCD and AIS Workshop, Bruges, Belgium, June 5 - 7, 1997
- S.K. Mendis et al. (1997), "Active pixel image sensors in 0.35µm CMOS technology", IEEE CCD and AIS Workshop, Bruges, Belgium, June 5 - 7, 1997
- » 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
- » B. Dierickx et al. (1997), "Offset-free offset correction for active pixel sensors", IEEE CCD and AIS Workshop, Bruges, Belgium, June 5 - 7, 1997