Chapter 2
History of Image Sensor & others research

We would introduce the history of image sensors and others relative circuit's history or theorem in this chapter. From the beginning, we will have a brief section about quantum physics in photoelectric effect. Several photo-detectors, such as photo-diode, photo-transistors, would be described on second section. On third, Couple Charged Device (CCD) image sensor would be compared and described with CMOS LSI technology's image sensor. We can know the differences between these two kinds of sensor, and the reason why we choose CMOS image sensor's architecture as what we want. The interrelated characteristic's research would begin from section 4, wide dynamic range. Most general wide dynamic range research would be introduced. Section 5 shows the basic theorem about differential difference amplifier, the ones which be used in the chip design. For the reason to overcome inaccuracy in amplifier caused from variational temperature, the compensated circuit would be designed and shows on section 6. Finally, we need a stable voltage reference which cannot be disturbed by temperature called bandgap reference circuit. We will show the theorem for this important component and the relevant paper which we referenced from.

2.1 Photoelectric effect

Before we start to realize what image sensor is, material "Light" is the most important topic we have to know. Many of the observable properties of light can be understood if it is regarded as a wave which travels with a finite speed c (3 × 10^8 m/s). James Clerk Maxwell could show that equations which he had derived to describe the behavior of time-varying electric and magnetic fields predicted the existence of propagating waves and his theory yields precisely the speed of light (known from measurements) as their velocity. The picture Fig. 2.1-1 shows a "snapshot" of such an electromagnetic wave: The red and blue arrows symbolize an electric and a magnetic field, respectively, which are perpendicular to one another. The strength of these fields varies along the propagation path; this is indicated by the lengths of the arrows. Both these fields also vary in time and this leads to the propagation of the wave.

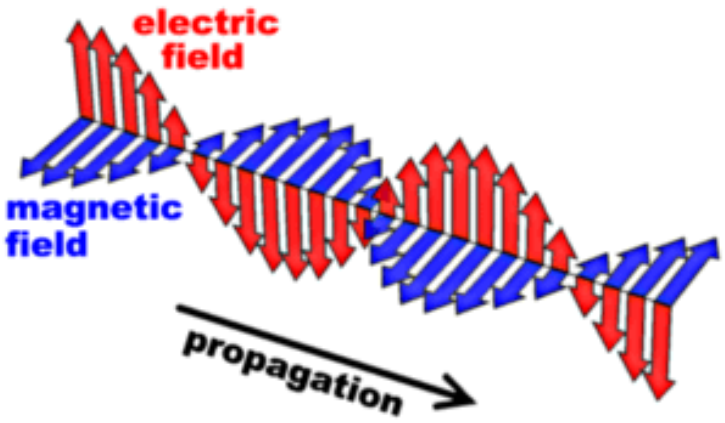


Fig. 2.1-1 Snapshot of Electromagnetic Wave
extracted from MoNOS Lab in Leiden University, Netherlands < 3 >

Two important parameters to characterize an electromagnetic wave are its wavelength λ and its frequency ν. The wavelength λ is the shortest distance between two points along the propagation path at which the electric field strength is the same at any given moment. This is illustrated in the Fig. 2.1-2.

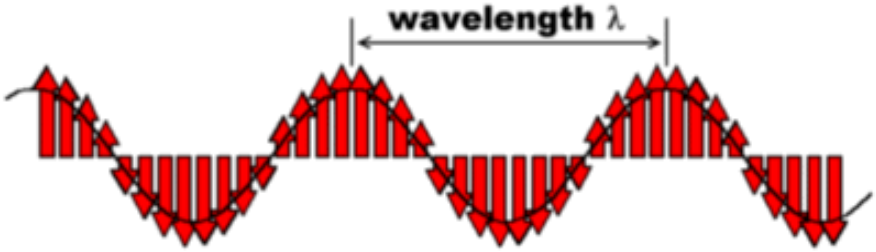


Fig. 2.1-2 Definition of wavelength
extracted from MoNOS Lab in Leiden University, Netherlands < 3 >

Concerning the power issue of electromagnetic wave, though we can describe it from both wave formula and quantum physics view. In this thesis, we would only define the energy in photons angle which in quantum physics. In Max. Planck's research on the radiation from a hot (black) body, he made a simple proposal. He suggested that light consists of photons, and could be display as the Eq.（3）. The energy, E, of each individual photon of a monochromatic light wave, is proportional to the frequency, ν, of the light. Where (h = 6.626*10^{-34} J s) is now known as the Planck constant.

$$E = h \cdot \nu, \tag{3}$$

From now on, the different kinds of frequency caused the different electromagnetic waves. They contain different energy and display as various types in the world, we would discuss them on next paragraph.

2.1.1 Light versus Electromagnetic Waves

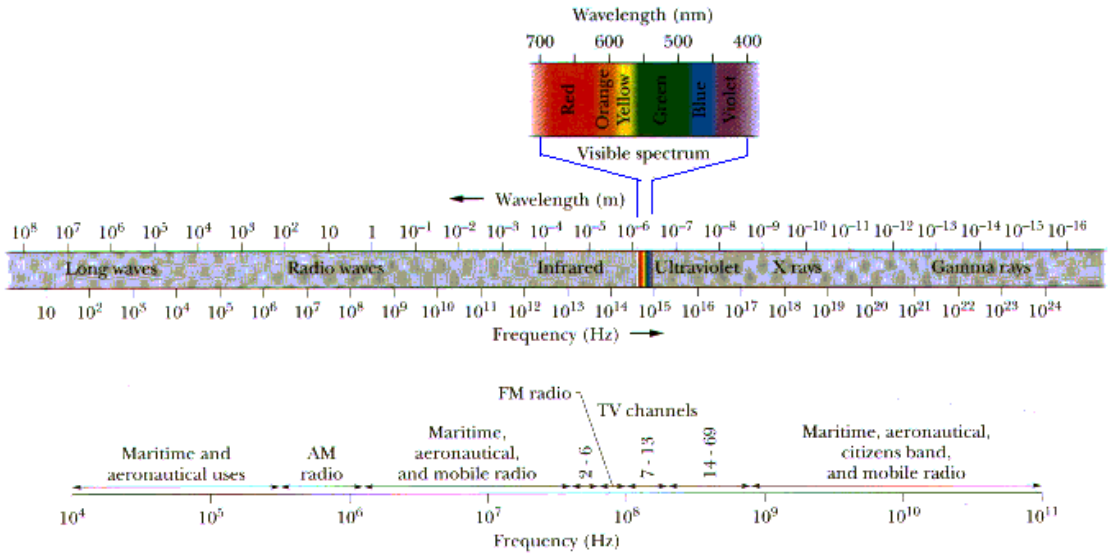


Fig. 2.1-3 Electromagnetic Wave Spectrum
extracted from Physics Demo Lab in NTNU < 2 >

Fig. 2.1-3 shows the electromagnetic wave spectrum, and people gave them different name with varied characteristic. From lowest frequency to highest would be Radio wave, Micro wave, Infrared wave, visible light, Ultraviolet wave, X rays, and Gamma rays.

The wave length which is longer than 0.3m, we call Radio wave. They are used for radio broadcasts, amateur radio, television, and mobile phone. In our radio broadcast system, they have two kinds of modulation, Amplitude Modulation （AM）and Frequency Modulation （FM）. AM frequency band is around 535 kHz ~ 1605 kHz, and this signal easily to be affected by lighting, car ignition system, or even computer. For the reason that noise usually interfere the signal's amplitude, frequency modulation use different frequency as carrier waves that reduce the noise disturbance. FM frequency band is around 88MHz ~ 106MHz. Besides, the energy of radio wave usually far smaller than band gap of valence electron, it won't be absorbed and caused the transition effect. For these reason, the transmission of radio wave almost can pass glass, stone, anything instead of metal obstacle, and that's why radio wave becoming the main technology for wireless communication.

The frequency range belongs in 1GHz to 100GHz we called Microwave. Unlike the radio wave can't pass through the ionosphere, microwave has been extensively used in communication between surface and space. The famous application is microwave

oven. Since water is a polar molecule, which would be affected to the same direction by electric field. While the oven generate the 2.45GHz microwave, water molecule start to oscillate and it's friction generated the heat. Microwave also could use in mobile phones' communication, speed meters, etc.

On 1800, William Herschel measure the heat which generated by refracted sunlight through the prism, and the region on the edge of red light had increase more degree than others on temperature. For the reason, he called these unviable "light" as Infrared wave. Infrared ray's frequency is between 300GHz and 385THz, and it will exist when the material was heated, including human body or animal's body temperature (wave length from 300nm to 10000nm). Infrared ray can pass through the lens or cloud, the characteristic make it been used for detection system such as spy satellites, heat tracing system, like Fig. 2.1-4 , night vision, etc.

Fig. 2.1-4 Infrared Ray Application
extracted from Dipol Website < 4 >

Visible light usually define as the wave length between 390nm ~ 780nm, but human's eyes can detect broadly to 312nm ~ 1050nm. Besides, there are only differences on "visibility" for human beings. Fig. 2.1-5 shows the white light pass the prism and extends into different colors' light.

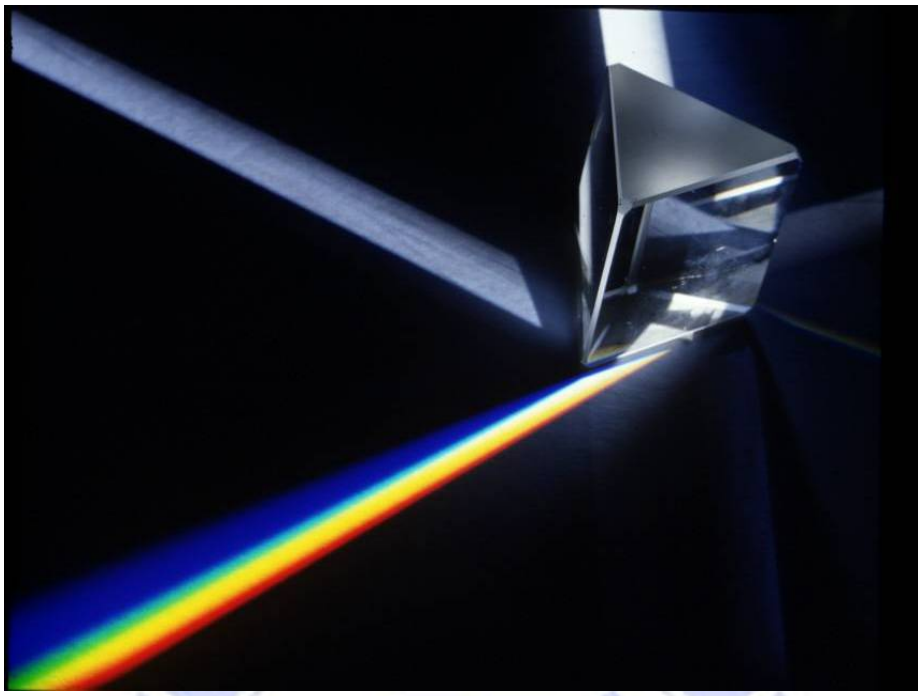


Fig. 2.1-5 White light extend by prism
extracted from DHD Multimedia Gallery < 5 >

Human eye's sensitivity of "white light" should be called as the feeling of sunlight. If the light has similar ratio as sunlight but in different frequency, it may cause the feeling of "white light". For example, mix the red light in 656nm and cyan light in 492nm would have "White Light" feeling. Besides, human's eye cannot distinguish the same color light which mixes in different frequency (ears can distinguish the different timbre). The mix light with 507THz and 509THz look like the same yellow light as 508THz. Unless we use the optical instrument to analysis, human's eye can't discriminate single band's light. Table 2.1.1-1 shows the visible light's frequency and its wavelength.

Table 2.1.1-1 Visible Light's Frequency and Wavelength
extracted from Physics Demo Lab in NTNU < 2 >

Color	Range of Frequency(THz)	Wave-length in Vacuum(nm)
Red	384-482	622-780
Orange	482-503	597-622
Yellow	503-520	577-597
Green	520-610	492-577
Blue	610-659	455-492
Purple	659-769	390-455

Color is not the characteristic itself, and it should be the feeling that vision nerve and brain response. More exactly, we can define “yellow light” as “the light which looks like yellow”. The green and red light which length is 540nm, 690nm in vacuum come into eyes at the same time, and may cause the feeling of yellow light without yellow’s frequency. It is so call “complementary color” effect.

Ultraviolet ray’s frequency belongs in 8×10^{14} Hz to 2.4×10^{16} Hz, were discovered by J. Ritter in 1801. On those days people has known chloride silver would become black after exposure under the sunlight, and he found that the phenomenon went more seriously out of violet band. Thus it is so called “ultraviolet”. Ultraviolet has higher energy than visible light, and easily spoil the link of molecule. It only takes 4eV energy and enough to destroy carbon molecule on 300nm. The wave length shorter than 300nm would cause damage for protein, but how come nature living can survive in the earth? Almost the ultraviolet on space would be absorbed by ozonosphere, and its resonance with N_2, O_2, CO_2, H_2O cause the reason why the sky is blue. For human being, cornea assimilate large part of ultraviolet lead eyes cannot see it. So the cataract patients who get keratectomy have response on ultraviolet. Some animals like honey bees can see ultra-violet light. Some plants have white flowers, at least you think that they are all white, but they may appear to be different colors to a honey bee because of the amounts of ultra-violet light which they reflect.

Fig. 2.1-6 Infrared X-Ray of hand
extracted from ColorXrays.com < 6 >

While the inner-level electron in atom has bandgap transition, the large energy would be demonstrated in X-Rays. It have so much energy and such a short wavelength that they can go right through you with main frequency in 2.4×10^{16} Hz ~ 5×10^{19} Hz. Thus the corresponding energy nearly 100keV-200keV, single X-ray photon can damage substance easily. Medical X-Rays usually have energy in 20keV-100keV. They cannot get through bone as easily as they can get through muscle, because bones contain so much Calcium which cannot absorb it, like Fig. 2.1-6. X-Ray can also be used to find other problems in your body. If the doctors want to look for ulcers in your guts, they can give you a Barium meal. Like Calcium, the Barium absorbs X-Rays so the doctors can look at parts of your guts and find your ulcers.

Gamma Rays are nasty ones. They have highest energy in electromagnet waves, and even go through metals. So they can be used for finding tiny cracks in metals. You cannot see the hairline cracks in an aero-plane wing with the naked eye. How would you like the wings to fall off the plane on your summer holiday flight to Neasden or wherever it is that you go. If the plane has been thoroughly checked you should be safe. Some radioactive materials produce gamma rays. Gamma rays and X-Rays can cause cancer, but gamma rays can also be used to destroy cancer cells: this is radio-therapy.

2.1.2 Theorem of Photoelectric Effect

In 1902, Lenard studied how the energy of the emitted photoelectrons varied with the intensity of the light. He used a carbon arc light, and could increase the intensity a thousand-fold. The ejected electrons hit another metal plate, the collector, which was connected to the cathode by a wire with a sensitive ammeter, to measure the current produced by the illumination. To measure the energy of the ejected electrons, Lenard charged the collector plate negatively, to repel the electrons coming towards it. Thus, only electrons ejected with enough kinetic energy to get up this potential hill would contribute to the current. Lenard discovered that there was a well defined minimum voltage that stopped any electrons getting through, we'll call it V_{stop}. To his surprise, he found that V_{stop} did not depend at all on the intensity of the light! Doubling the light intensity doubled the number of electrons emitted, but did not affect the energies of the emitted electrons. The more powerful oscillating field ejected more electrons, but the maximum individual energy of the ejected electrons was the same as for the weaker

field. But Lenard did something else. With his very powerful arc lamp, there was sufficient intensity to separate out the colors and check the photoelectric effect using light of different colors. He found that the maximum energy of the ejected electrons did depend on the color — the shorter wavelength, higher frequency light caused electrons to be ejected with more energy. This was, however, a fairly qualitative conclusion — the energy measurements were not very reproducible, because they were extremely sensitive to the condition of the surface, in particular its state of partial oxidation. In the best vacuum available at that time, significant oxidation of a fresh surface took place in tens of minutes.

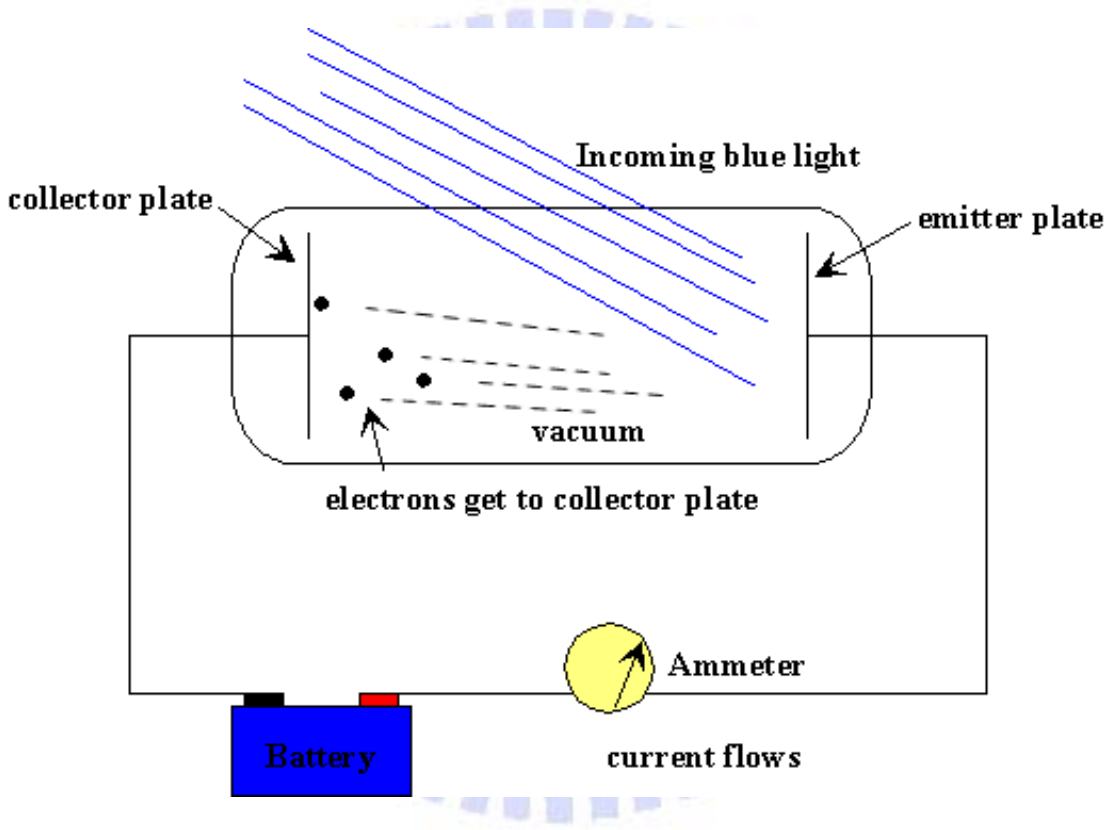


Fig. 2.1-7 Lenard Photon Experiment (get plate)
extracted from Michael Fowler, "The Photoelectric Effect"< 7 >

In the Fig. 2.1-7, the battery represents the potential Lenard used to charge the collector plate negatively, which would actually be a variable voltage source. Since the electrons ejected by the blue light are getting to the collector plate, evidently the potential supplied by the battery is less than V_{stop} for blue light. Show with an arrow on the wire the direction of the electric current in the wire.

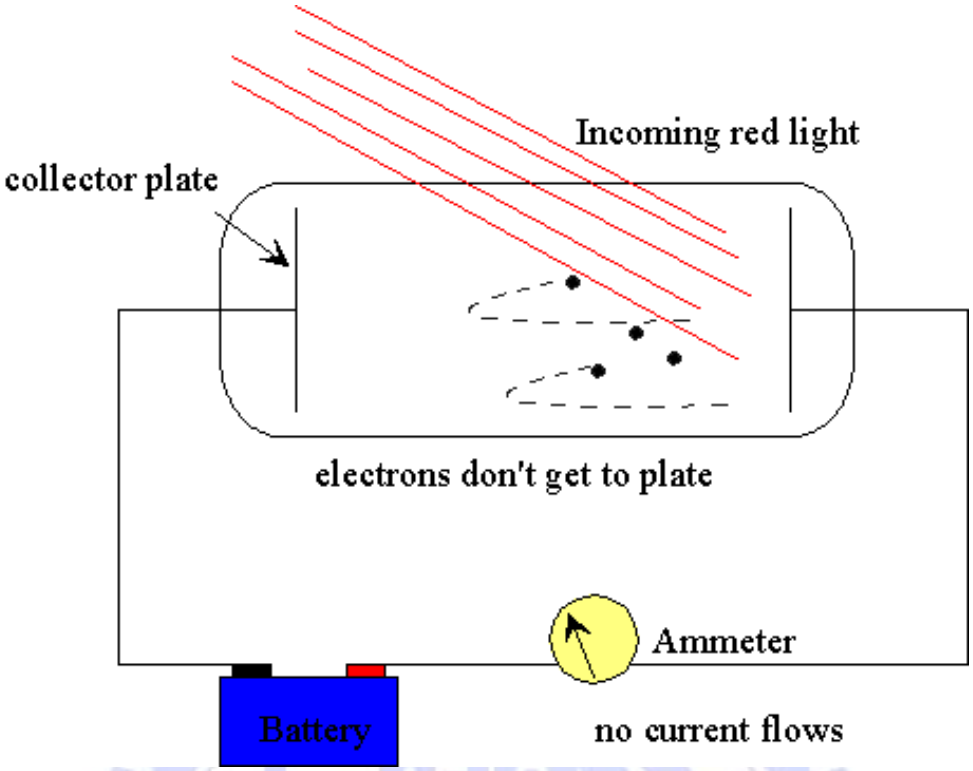


Fig. 2.1-8 Lenard Photon Experiment (don't get plate)
extracted from Michael Fowler, "The Photoelectric Effect"< 7 >

In 1905 Einstein gave a very simple interpretation of Lenard's results with Eq.（4）. He just assumed that the incoming radiation should be thought of as quanta of frequency $h\nu$, with ν the frequency. In photoemission, one such quantum is absorbed by one electron. If the electron is some distance into the material of the cathode, some energy will be lost as it moves towards the surface. There will always be some electrostatic cost as the electron leaves the surface; this is usually called the work function, W. The most energetic electrons emitted will be those very close to the surface, and they will leave the cathode with kinetic energy

$$E = h\nu - W \tag{4}$$

On cranking up the negative voltage on the collector plate until the current just stops, that is, V_{stop}, the highest kinetic energy electrons must have had energy eV_{stop} on leaving the cathode. Thus Eq.（5）,

$$eV_{stop} = h\nu - W \tag{5}$$

Thus Einstein's theory makes a very definite quantitative prediction: if the frequency of the incident light is varied, and V_{stop} plotted as a function of frequency, the slope of the line should be h/e. It is also clear that there is a minimum light frequency for a given metal, that for which the quantum of energy is equal to the work function. Light below that frequency, no matter how bright, will not cause photoemission.

2.2 Photo Detector

The front end of a visual microprocessor must encompass the task of transforming the input visual scene on to a format suitable for processing. This is achieved by using a two-dimensional, spatial arrangement of light –sensitive electronic devices which represent such an image as a matrix $U = \{u_{ij}, 1 \le i \le M, 1 \le j \le N\}$ of electrical voltages or currents. Each of these devices must therefore be capable of transforming the wavelength （color）and intensity of the light which is incident on its area into an associated voltage or current. A priori, this can be achieved by using different types of devices, built with different types of materials in different technologies. Still, for pure imaging tasks, the design of these devices is currently being dominated by CCD technology. In this section, we would only discuss the photo detectors in CMOS VLSI technology for the reason that CCD technology has less integrated ability and limited its application with combinational with other circuit. We would discuss them on next section.

In a standard CMOS process, several structures can be used to convert light into an electrical signal. These structures, which are mainly junction devices, are depicted in Fig. 2.2-1 for an n-well process （the complementary structures are obtained in a p-well process）. The first three structures are p/n junction diodes, whereas the last one is a vertical bipolar junction transistor（BJT）. Diodes are formed by well/substrate （n^- / p^-）（Fig. 2.2-1（a））, source- drain diffusion /substrate （n^+ / p^-）junctions （Fig. 2.2-1（b））and source- drain diffusion/well p^+ / n^- junctions. Vertical p/n/p bipolar junction transistors are obtained by the structure formed a p^+ source-drain diffusion, an n^- well and the p^- substrate （Fig. 2.2-1（d））.

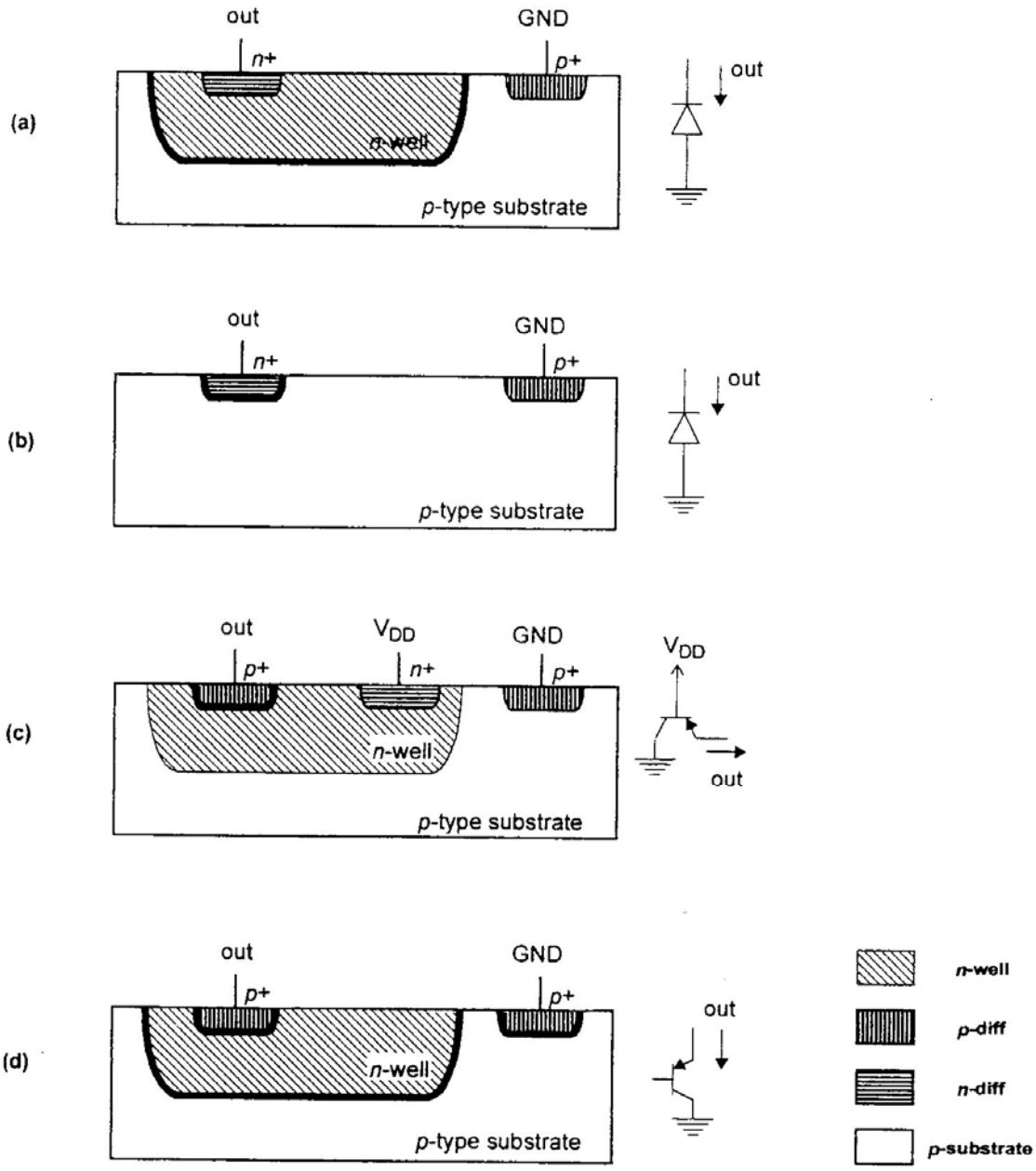


Fig. 2.2-1 Junction devices available in a standard n-well CMOS process for light detection. The complementary structures are obtained in a p-well process. The active junction is shaded for each device.（a）Well/substrate（n^-/p^-）diode;（b）diffusion/substrate（n^+/p^-）diode;（c）diffusion/well（p^+/n^-）diode;（d）（$p^+/n^-/p^-$）vertical bipolar transistor;

extracted from Tamás Roska, Ángel Rodríguez – Vázquez, "Towards the Visual Microprocessor – VLSI Design and the Use of Cellular Neural Network Universal Machines" Chapter 5 Light-sensitive Devices in CMOS Page 185【4】

2.1.1 Photodiodes

When light is incident on a diode, photons absorbed in the depletion region of the p/n junction create electron-hole pairs which are separated by high internal electric field formed across it（Fig.2.2-2）. This charge can be detected as an increase in the reverse current of the device（photoconductive mode of operation）or as a change in the voltage across it when it is left open circuit （photovoltaic mode of operation）. Carriers generated in the silicon bulk at a distance within a minority carrier diffusion length from the depletion region can also contribute to the detected signal, increasing the sensitive volume of the detector.

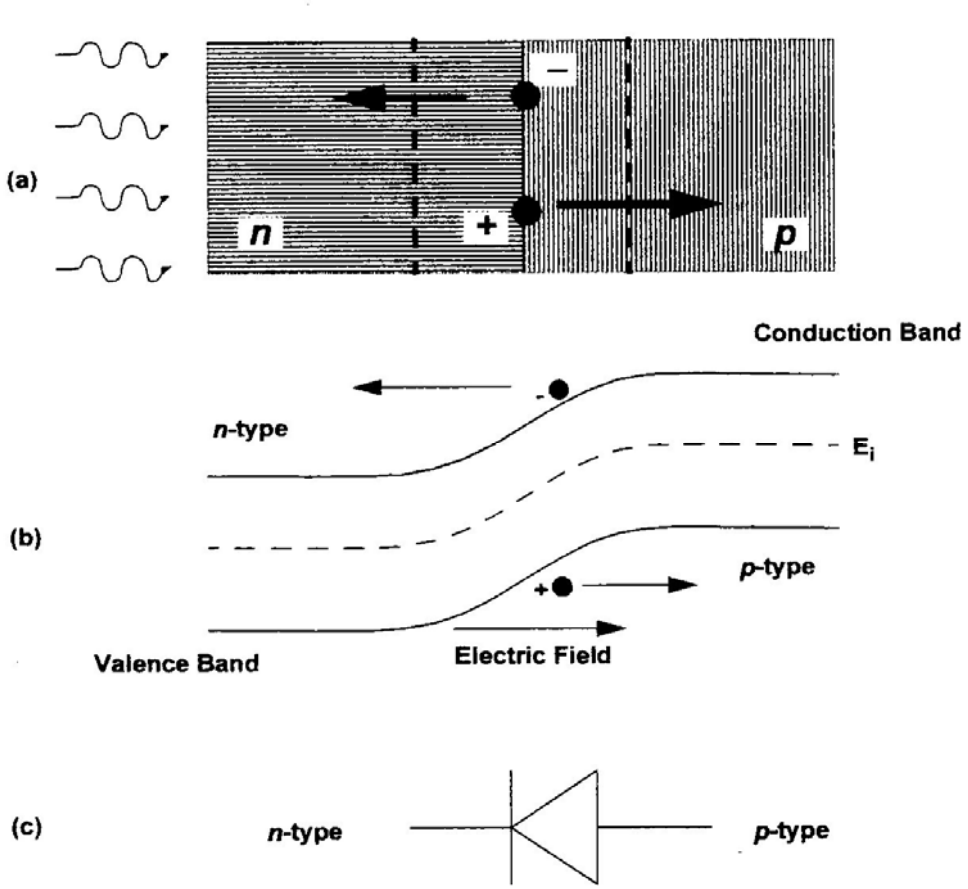


Fig. 2.2-2 Photo-generation of carriers in a photodiode. Electron-hole pairs which are created in the depletion region of the reverse biased junction can be detected as an increase in the reverse current （photoconductive mode of operation）or as a change in the voltage across it（photovoltaic mode of operation）;（a）cross-section of the device structure;（b）energy band diagram;（c）electrical representation of the photodiode

extracted from Tamás Roska, Ángel Rodríguez – Vázquez, "Towards the Visual Microprocessor – VLSI Design and the Use of Cellular Neural Network Universal Machines" Chapter 5 Light-sensitive Devices in CMOS Page 186【4】

The detected signal in diode can be represented by a photo-generated current as Eq.(6).

$$I_{ph} = \frac{q \cdot A \cdot P \cdot \lambda \cdot \eta}{h \cdot c}, \tag{6}$$

where q is the electronic charge, A the detector area, P the incident light power per unit area, λ the wavelength of the incident light, h the Planck constant, c the speed of light in vacuum and η the absolute quantum efficiency of the process at wavelength λ. Here η is defined as the number of electrical charges detected per incident photon.

$$\eta = \frac{no.of\_detected\_charges}{no.of\_incident\_charges}, \tag{7}$$

The quantum efficiency is a quantity specific to each photosensitive device and is a function of the wavelength if the incident radiation. A graph of the quantum efficiency vs. wavelength for a typical photodiode is shown in Fig. 2.2-3.

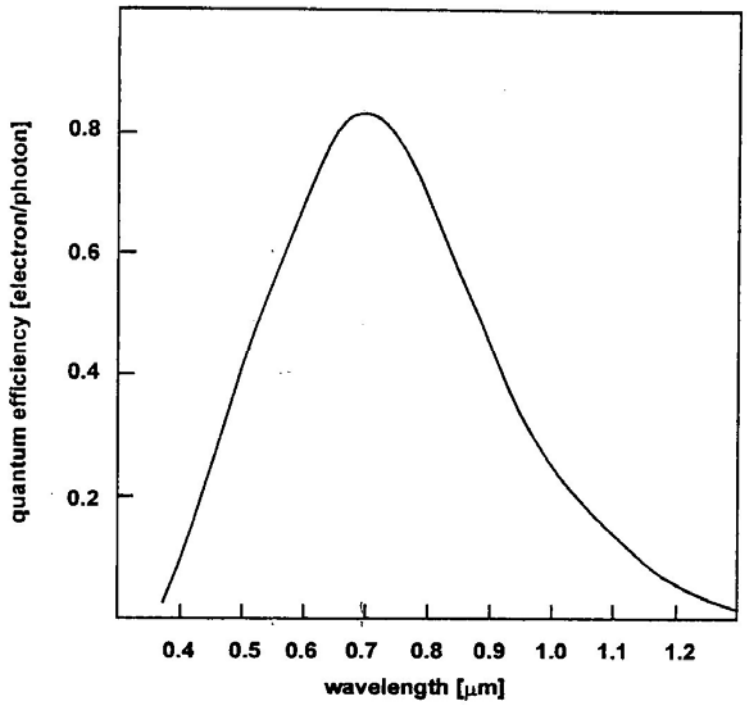


Fig. 2.2-3 Quantum efficiency for a silicion p^+/n photodiode as a function of the wavelength of the incident light

extracted from Tamás Roska, Ángel Rodríguez – Vázquez, "Towards the Visual Microprocessor – VLSI Design and the Use of Cellular Neural Network Universal Machines" Chapter 5 Light-sensitive Devices in CMOS Page 187【4】

One of the factors that determine quantum efficiency is the absorption coefficient α, which is a strong function of the wavelength of the incident light. For silicon, the absorption length, （L（λ）=1/α）, is shorter for shorter photon wavelengths. Being L=0.3μm for blue light （475 nm） while for red light （650 nm）it is L= 3μm. For a given semiconductor, the long cut-off wavelength, λ_c, is determined by the energy gap of the material, E_g. For silicon E_g =1.12eV, which corresponds to λ_c =1.1μm according to the relationship Eq.（8）.

$$E = \frac{hc}{\lambda} = \frac{1.24}{\lambda}, \tag{8}$$

where E is the photon energy. At short wavelengths, the values of α are large and photons are absorbed very near the surface where the recombination time is short. Therefore, photo-generated carriers are recombined before they are collected in the junction, determining the short cut-off wavelength of these devices. For silicon, typical values for the quantum efficiency of about 0.8um are obtained at 0.8-0.9um, as can be seen in Fig.2.2-3. These values can be increased by incorporating anti-reflection coatings on detector surface, improving the efficiency in the absorption process. Quantum efficiency can also be improved by increasing the width of the depletion region, increasing the volume of silicon where electron-hole pairs can be separated. For an abrupt junction diode, the width of the depletion region is given by Eq.（9）.

$$W = \sqrt{\frac{2\varepsilon_s}{q}\left(\frac{1}{N_a}+\frac{1}{N_d}\right)\left(\Phi_{bi}+V_R\right)}, \tag{9}$$

Where ε_s is the permittivity of silicon, Φ_{bi} the built-in potential of junction, V_R the reverse applied voltage, N_d the donor concentration in the n-type silicon and N_a the acceptor concentration in the p-type silicon. As can be seen from Eq.（9）, the depletion region by increasing the reverse bias there will be an improvement in the response of the detector. The depletion region width can also be increased by reducing the doping levels, although this is not possible in a standard CMOS process. However, it is expected that those diodes formed by the least doped materials will present a higher response.

The spectral response of the three previous diodes available in a standard CMOS process is mainly determined by the absorption of light in silicon and wave-length dependent. This dependency cause causes photo-detectors formed with junctions of different junction depths to have different spectral responses. The （n^-/p^-）and （n^+/p^-）diodes behave similarly, with a response very close to Fig.2.2-3. The （p^+/n^-）diodes are more sensitive at shorter wavelengths since the collection of carriers is limited by the junction volume, in contrast to the other diodes where the collection of carriers is limited by the diffusion length of minority carriers in the silicon bulk. This behavior has been observed experimentally by Delbrück [13] He measured those diodes in standard CMOS process with few more junction structures, and Fig. 2.2-4 demonstrated n^-/p^- and p^+/n^- ones he did. Though he didn't measured n^+/p^- but he expected the behavior would be similar as （n^-/p^-）diodes. Soncini[17] confirmed this view with experiment.

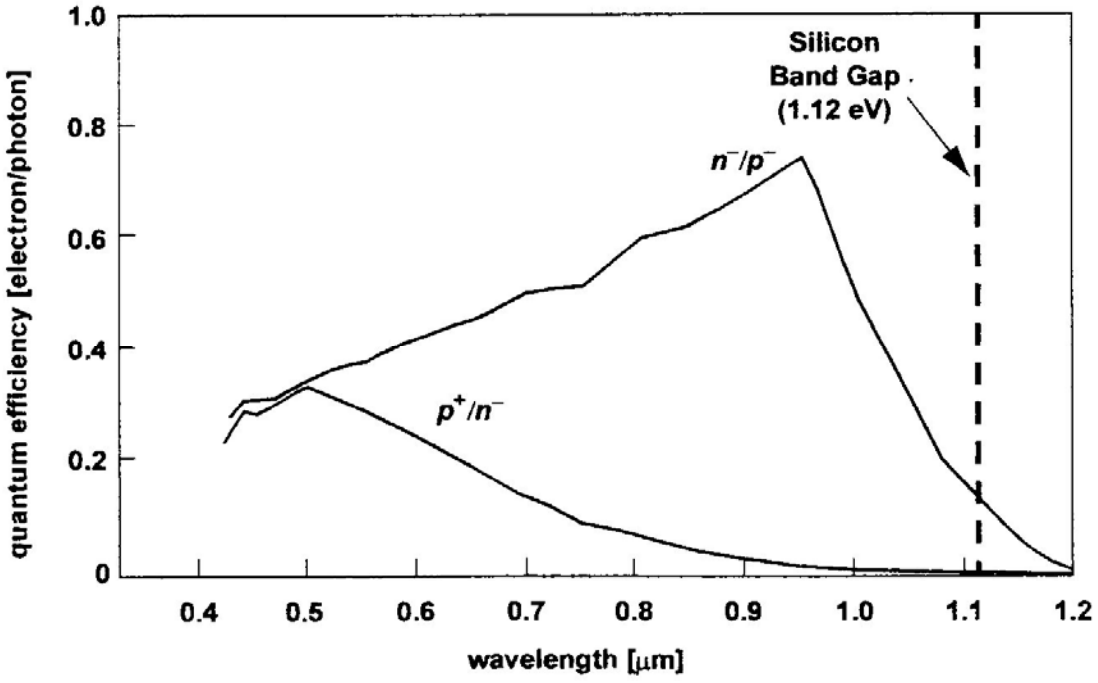


Fig. 2.2-4 Spectral response of n^-/p^- and p^+/n^- diodes as measured by Delbrück in a BiCMOS process

extracted from Tamás Roska, Ángel Rodríguez – Vázquez, "Towards the Visual Microprocessor – VLSI Design and the Use of Cellular Neural Network Universal Machines" Chapter 5 Light-sensitive Devices in CMOS Page 188【4】

2.1.2 Phototransistors

In a bipolar-junction transistor polarized in the active forward region（emitter/base junction forward biased and base/collector junction reverse biased）, whose base terminal（the well）is left floating, the incident light will generate electron-hole pairs in the reverse biased base/collector junction that will induce a current which will be equivalent to a base current （Fig.2.2-5）.

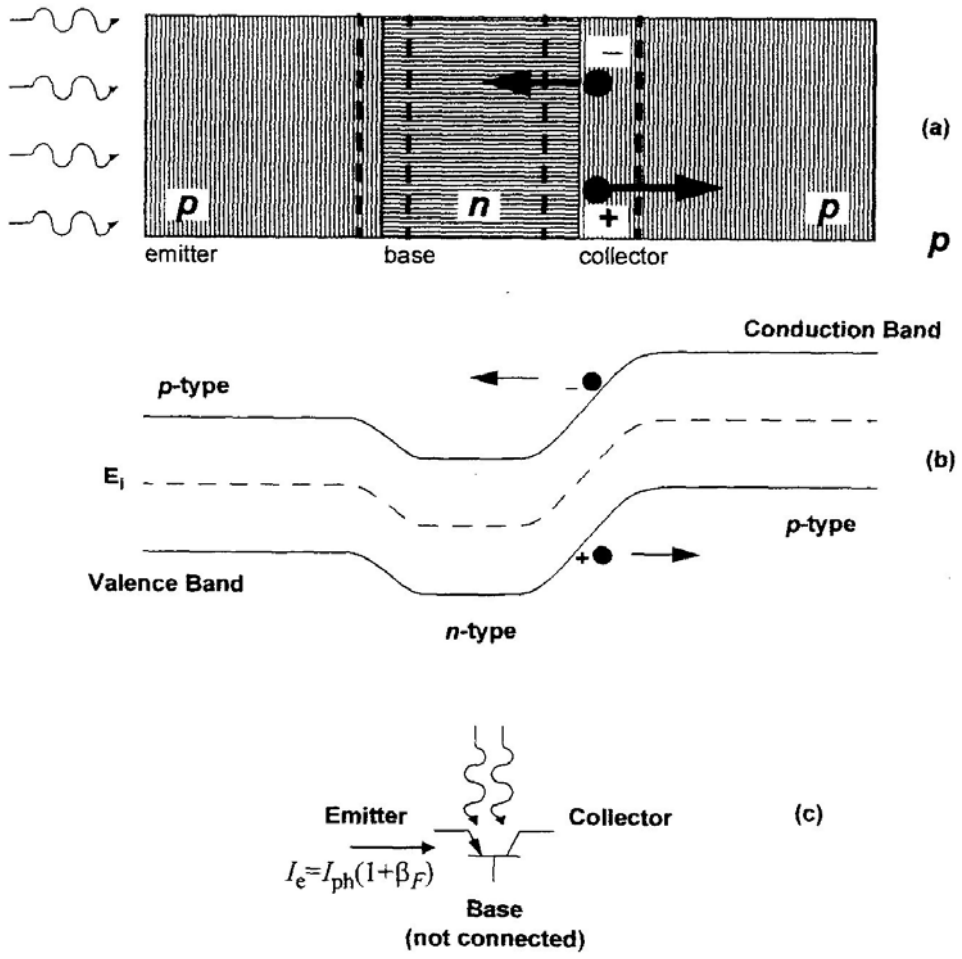


Fig. 2.2-5 Photo-generation of carriers in a phototransistor. Electron-hole pairs created in the depletion region of the reverse biased base/collector junctions will induce a base current, I_{ph} ,which will be amplified by the gain of the transistor, resulting in an emitter current which is （1+β_F）times the detected current. （a）Cross-section of the device structure; （b）energy band diagram;（c）electrical representation of the phototransistor

extracted from Tamás Roska, Ángel Rodríguez – Vázquez, "Towards the Visual Microprocessor – VLSI Design and the Use of Cellular Neural Network Universal Machines" Chapter 5 Light-sensitive Devices in CMOS Page 192【4】

This base current will be amplified by the gain of the transistor, resulting in an emitter current which is （$1+\beta_F$）times the detected current, where β_F is the common-emitter current gain. The emitter current is expected to be equivalent to the detected current in a well/substrate doped of the same well area. But amplified by （$1+\beta_F$）. The value of β_F depends on technological parameters such as base and emitter doping levels and widths【11】. The photo-generated current will be the same as in a photodiode but amplified by（$1+\beta_F$）, being A being the area of the transistor base:

$$I_{ph} = \frac{q \cdot A \cdot P \cdot \lambda \cdot \eta}{h \cdot c}(1+\beta_F) \qquad (10)$$

It is a clear from Eq.（10）that phototransistor offer gain, increasing the level of the detected signal with respect to that of photodiode. The detected signal in a photodiode tends to be very low due to the small area used in large arrays, will be multiplied by （$1+\beta_F$）in a phototransistor, increasing the current level about two orders of magnitude. These detectors will then be suitable for applications where large current levels are needed. However, phototransistors require proportionally more power in order to operate and the fabrication process is, in general, more complex than that of photodiodes. Besides, at low light levels, and hence at low base currents, β_F can drop to quite low values.

The spectral response of phototransistors is very similar in shape to that of an （n^-/p^-）diode, but is increased by the amplifying factor （$1+\beta_F$）, since the active junction in both detectors is the well/substrate junction. This behavior has also observed by Delbrück [13] and can be seen on Fig. 2.2-6 where the spectral response of both well/substrate junction diodes and vertical BJTs are plotted in a logarithmic scale for a better comparison. The response of the （n^-/p^-）diodes and that of the （p^+/n^-）diodes, which is also included in the graph, is the same as in Fig.2.2-4.

After the discussion of these four photo detector, the （n^-/p^-）and （n^+/p^-）diodes are better suit for imaging applications since their higher quantum efficiency in the visible spectrum, simple layout. But since the limitation of TSMC 0.35um 2P4M Mix-Mode Technology which CIC provided, there is no choice on photodiode. For the

reason, and the higher current as the other devices, we chose phototransistors as our photo detector, Table 2.2.2-1 is the collections of these devices' comparison.

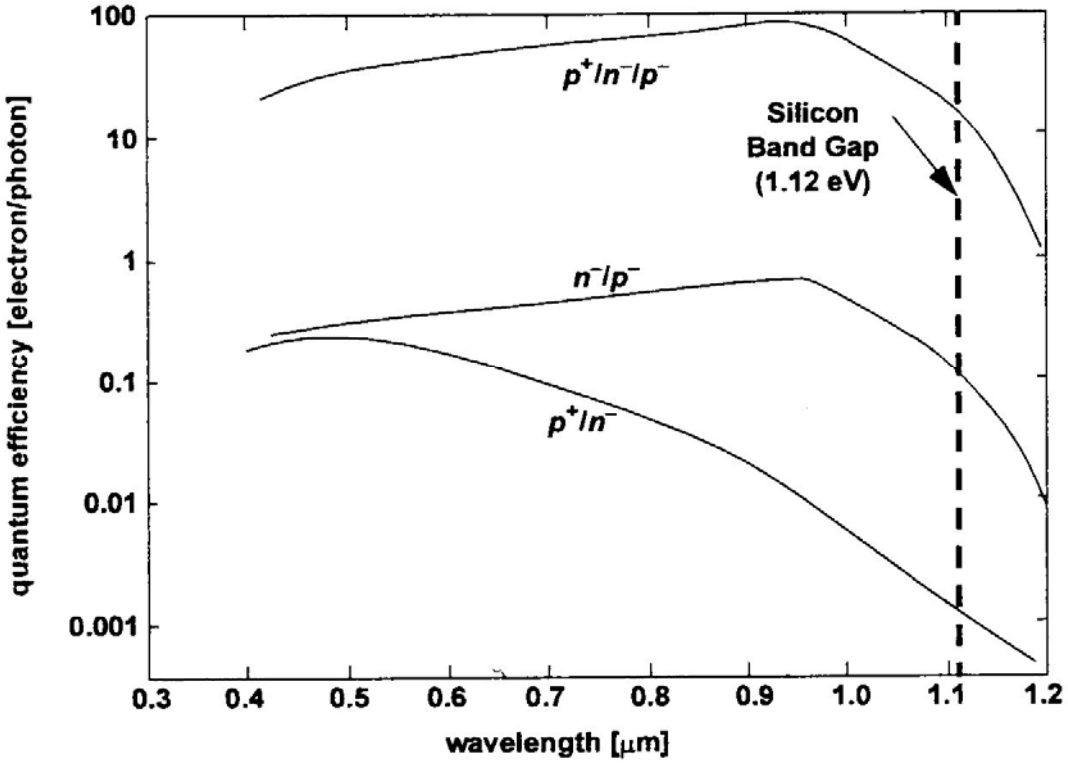


Fig. 2.2-6 Spectral response of n^-/p^- and p^+/n^- diodes and $p^+/n^-/p^-$ vertical transistor as measured by Delbrück in a BiCMOS process
extracted from Tamás Roska, Ángel Rodríguez – Vázquez, "Towards the Visual Microprocessor – VLSI Design and the Use of Cellular Neural Network Universal Machines" Chapter 5 Light-sensitive Devices in CMOS Page 194【4】

Table 2.1.2-1 Main properties of the different photo-detectors available in a standard CMOS Process
extracted from Tamás Roska, Ángel Rodríguez – Vázquez, "Towards the Visual Microprocessor – VLSI Design and the Use of Cellular Neural Network Universal Machines" Chapter 5 Light-sensitive Devices in CMOS Page 189【4】

Detector	n^- / p^-	n^+ / p^-	p^+ / n^-	$p^+ / n^- / p^-$
Time response	Fast	Fast	Fast	Slow
Current Level	Low	Low	Low	High
Dynamic range	5-7 decades	5-7 decades	5-7 decades	3-5 decades
Layout	Simple	Simple	Simple	More Complex
Spectral response	Increasing	Increasing	Decreasing	Increasing

2.3 CCD & CMOS

CCD (Charge Couple Device) and CMOS image sensors are both implemented in silicon. Silicon properties dictate image sensor wavelength response and the sensor's ultimate sensitivity. Fundamentally, CCD and CMOS image sensors can have identical responses to light, since both depend on the absorption of light by silicon and the collection of photo-generated electrons (or holes). They have the most difference in the readout architectures, as Fig. 2.3-1.

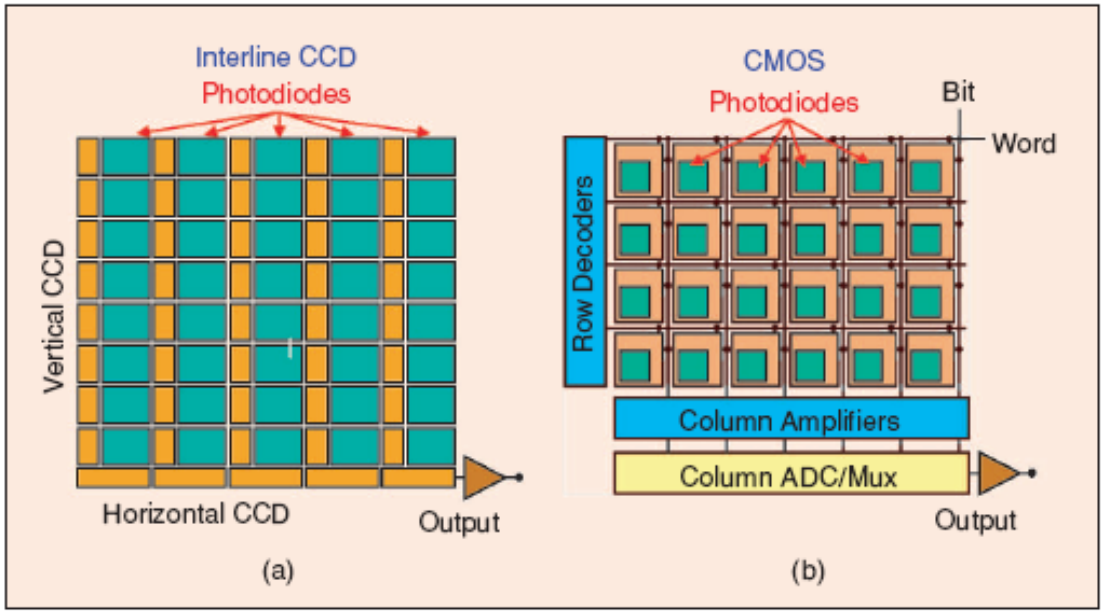


Fig. 2.3-1 （a）Readout architectures of interline transfer CCD and （b）CMOS Image Sensors

extracted from Abbas El Gamal, Helmy Eltoukhy, "CMOS Image Sensors", in IEEE Circuits and Devices Magazine, pp. 6-20 May/ June 2005 [18]

CCD was invented by W. S. Boyle and G. E. Smith, Bell Syst. Tech. 1969[19]. The first generation of CCD solid-state image sensors, have evolved to a high level of performance over the 28 years since their invention. Most of components were built in special technology to increase the sensitivity. Those photo devices sense the light and transfer as the electrical potential signal, and save the electrical value in the coupling capacitance in the photo devices. On the channel design, CCD devices has different phases signal which provided by other chips （Fig. 2.3-2） to change the channel potential and push the signal charge into amplifier, like Fig. 2.3-3. For this architecture, CCD can get larger fill factor and reduce the transfer noise. Though CCD has better quality of frame, but also spend large of power and must compare with other circuit for its' unique process.

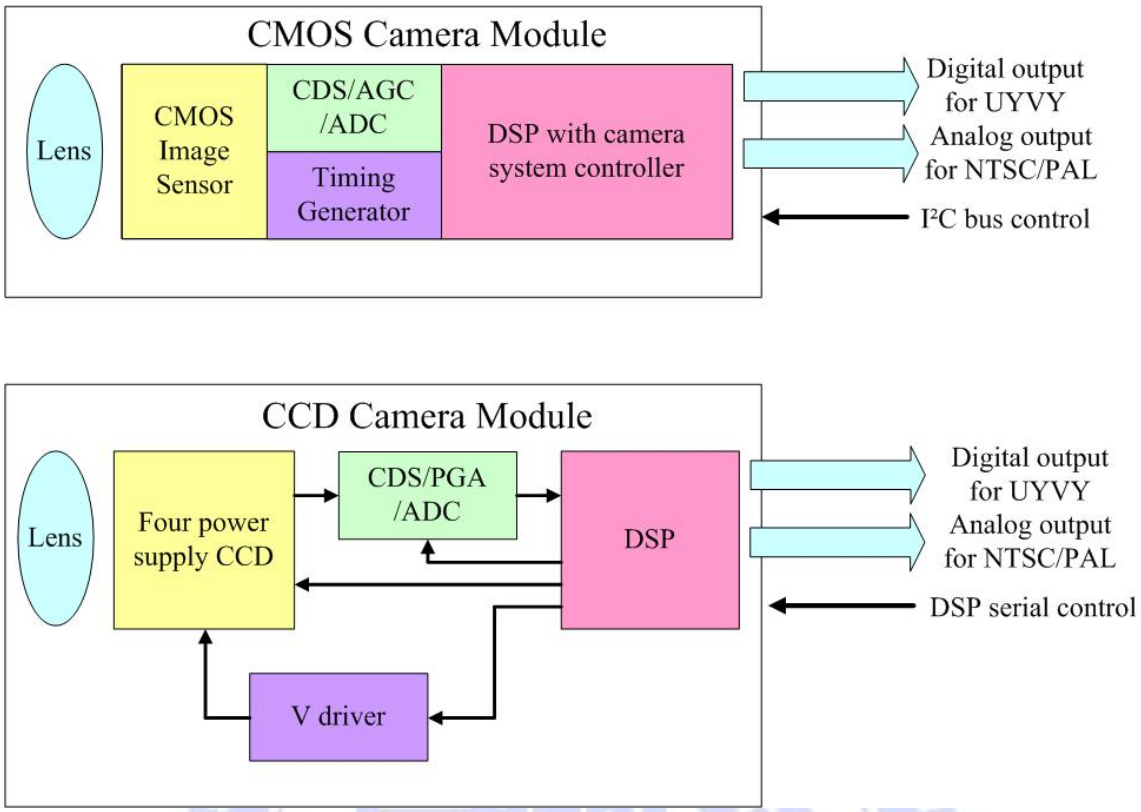


Fig. 2.3-2 The deference of CCD & CMOS on Camera system module

extracted from this thesis'（Shao-Hang Hung. 's ）coordination

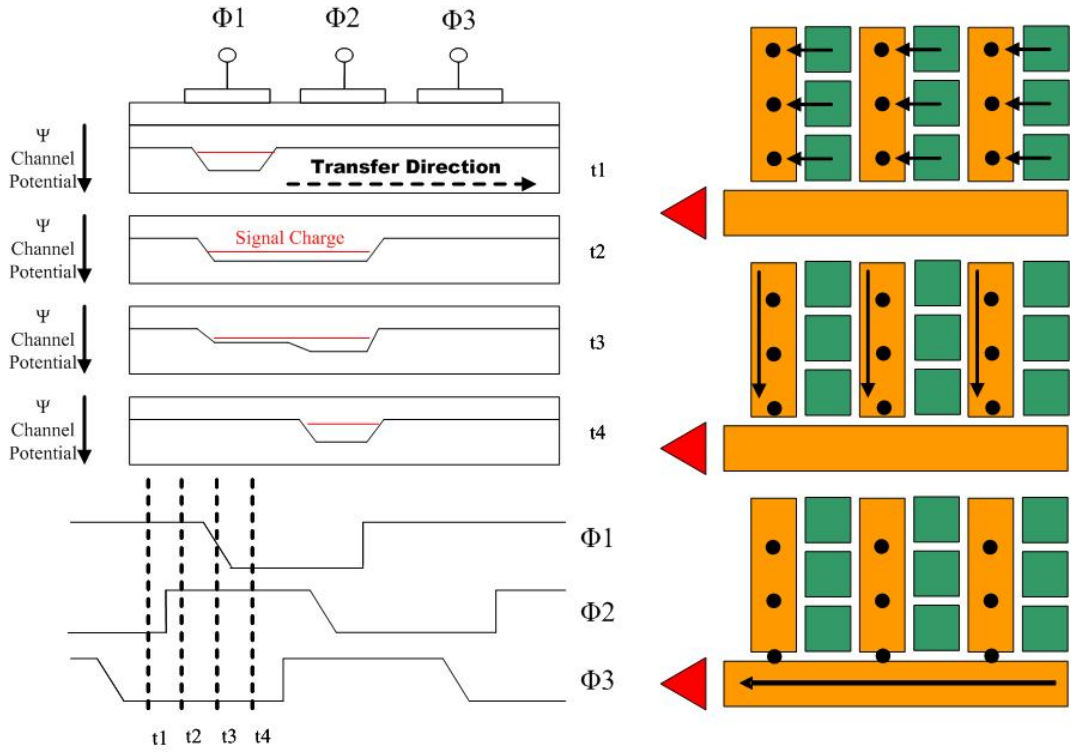


Fig. 2.3-3 CCDs' signal transfer theorem

extracted from this thesis'（Shao-Hang Hung. 's ）coordination

CMOS Image Sensor（CIS） looks like the same as CCD on figure, and also has the ability to record variation on semiconductor technology. But unlike the CCD in particular technology, CIS produce in standard technology like general chips. However, CIS has a little amplifier in the side of each pixel, and cause the different noise result from process variation. It would decrease the frame quality, and getting worse on fast transition. Though CIS has inferiority on image quality, but low cost, low power, integration with other circuits （Fig. 2.3-2）are its superiority. The CIS could be integrated with Analog Digital Converter（ADC）, Programmable Gain Amplifier（PGA）, Timing Generator（TG） etc.（Fig. 2.3-4）, which result in the on-chip-digital-camera. Other smart processing circuit could also be integrated with CIS, like video cooperation, wide dynamic range computation, motion detection, or even retina like neural network circuit, lots of functional circuit can improve the image quality of CIS. Even it is still been extensively used in commercial application such as computer video camera, can't pass though the gap of high class application like high end digital camera or digital video camera. The CIS has invaded the market of image sensor recent years, and threatened original CCD's market.

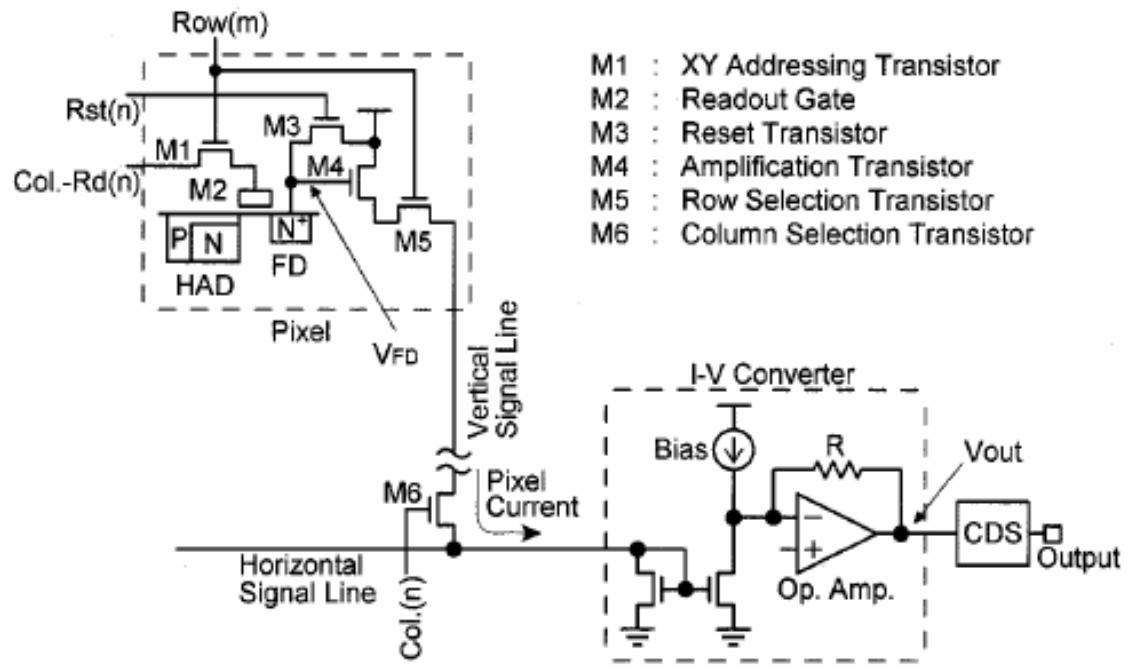
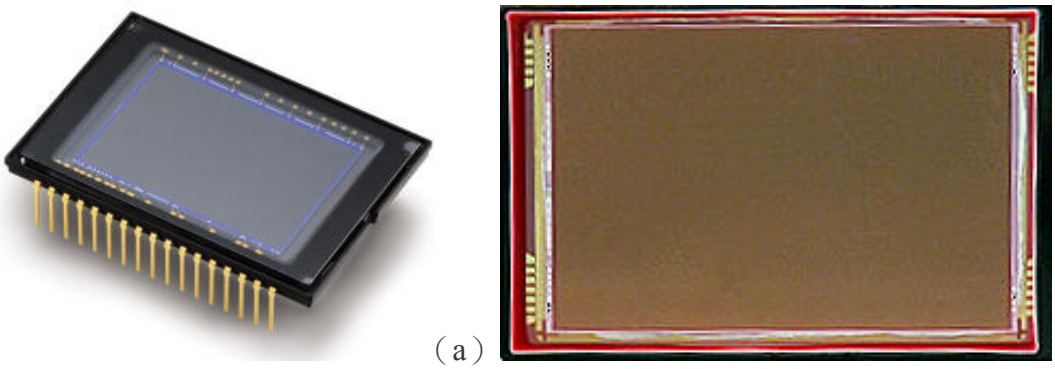


Fig. 2.3-4 Circuit Chain from pixel to CDS

extracted from Kazuya Yonemoto, Hirofumi Sumi, "A Numerical Analysis of a CMOS Image Sensor With a Simple Fixed-Pattern-Noise-Reduction Technology", in IEEE Trans. Electron Devices, Volume 49, Issue 5, pp.746~753, May 2002[20]



（a） （b）

Fig. 2.3-5 （a）Nikon D100 CCD （b）Canon EOS 350D CMOS Image Sensor
extracted from Mr. OH Digital Course Chap. 2 "Sensor Devices: CCD & CMOS"< 8 >

Table 2.1.2-1 Comparison Table on Feature and Performance on CCD & CMOS
extracted from 米本和也 原著；陳榕庭、彭美桂 翻譯，"CCD/CMOS 影像感測器之基礎與應用"，全華科技圖書股份有限公司，2005 年初版。【1】

Comparison	CCD	CMOS
Signal out of pixel	Electron packet	Voltage
Signal out of chip	Voltage (analog)	Bits (digital)
Signal out of camera	Bits (digital)	Bits (digital)
Fill factor	High	Moderate
Amplifier mismatch	N/A	Moderate
System Noise	Low	Moderate
System Complexity	High	Low
Sensor Complexity	Low	High
Camera components	Sensor + multiple support chips + lens	Sensor + lens possible, but additional support chips common
Relative R&D cost	Lower	Higher
Relative system cost	Depends on Application	Depends on Application
Responsively	Moderate	Slightly better
Dynamic Range	High	Moderate
Speed	Moderate to High	Higher
Windowing	Limited	Extensive
Biasing and Clocking	Multiple, higher voltage	Single, low-voltage

2.4 Wide Dynamic Range research

It has been an important issue on photography whether in traditional camera or digital camera, which, the ability to transform illumination of maximum and minimum light. On film's angle, they define this ability in ISO grade. On digital camera's angle, it can be define as dynamic range, as Eq.（11）.

$$Dynamic\_Range(DR) = \frac{Isense_{MAX}}{Isense_{MIN}}, \tag{11}$$

Sensor DR is generally not wide enough to image scenes encountered even in everyday consumer photography. This is especially the case for CMOS image sensors, since their read noise and DSNU are typically larger than CCDs. For reference, standard CMOS image sensors have a DR of 40–60 dB, CCDs around 60–70 dB, while the human eye exceeds 90 dB by some measures. In contrast, natural scenes often exhibit greater than 100 dB of DR. To solve this problem, several DR extension techniques such as well-capacity adjusting [21], multiple capture [22], time-to-saturation [23], and self-reset [24] have been proposed. These techniques extend DR at the high illumination by increasing i_{MAX}. In multiple capture and time-to-saturation, this is achieved by adapting each pixel's integration time to its photocurrent value, while in self-reset the effective well capacity is increased by "recycling" the well. To perform these functions, most of these schemes require per-pixel processing. We would introduce some of them on the following sections.

Fig. 2.4-1 Comparison of CMOS DPS imager versus CCD imager using a HDR scene

extracted from Abbas El Gamal, Helmy Eltoukhy, "CMOS Image Sensors", in IEEE Circuits and Devices Magazine, pp. 6-20 May/ June 2005 [18]

2.4.1 Dual Sampling [25]

The architecture of the new wide intrascene dynamic range (WIDyR) approach is shown in Fig. 2.4-2. In the new architecture, a second column signal processing chain circuit has been added to the upper part of the sensor. As before, row is selected for readout and copied into the lower capacitor bank. Row is reset in the process. However, immediately following, row is selected and copied into the upper capacitor bank. Row is also reset as a consequence of being copied. Both capacitor banks are then scanned for readout. With the difference integration time, which $T1_{\text{int}}$ is the longer one and $T2_{\text{int}}$ is the shorter one, the dynamic range capability is extended by the factor $T1_{\text{int}}/T2_{\text{int}}$. This architecture helps not only extend DR, and also help to raise the Signal to Noise Ratio （SNR）. However, the limitation of this architecture should save large of image data into memories; it would take much longer time and large amount of memories to processing. It is hard to be used on movement application like video stream.

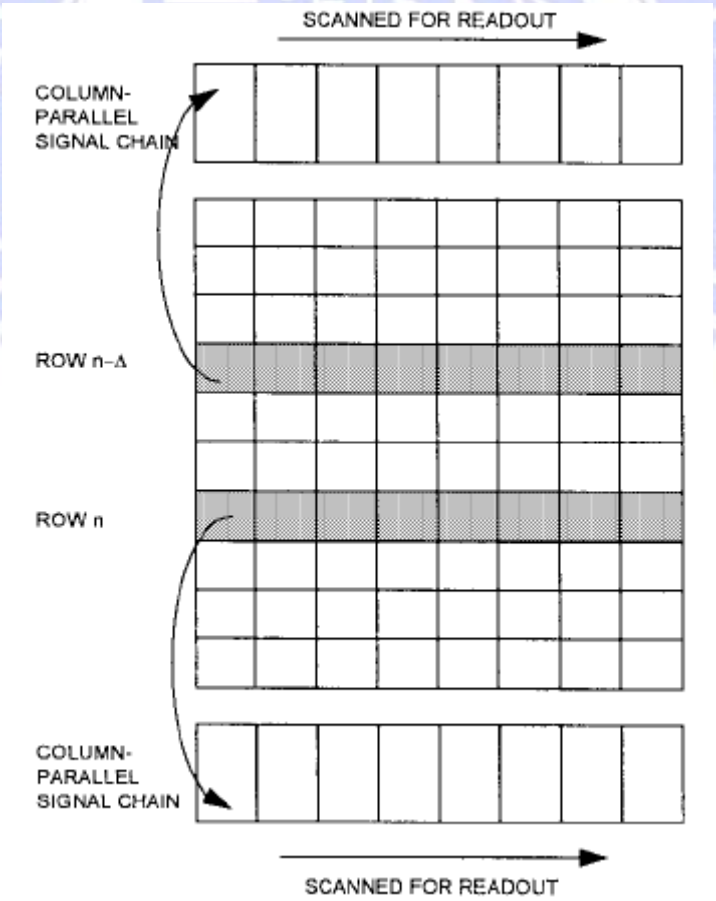


Fig. 2.4-2 Schematic illustration of dual-sample, dual output imager architecture
extracted from O. Y. Petch., E. R. Fossum, "Wide Interscene Dynamic Range SMOS APS Using Dual Sampling," in IEEE Trans. Electron Devices, Vol.44, No.10, pp. 1721~1723,1997.[25]

2.4.2 Logarithm Conversion [26]

Logarithm conversion used the sub-threshold operation region in CMOS transistors. They use the logarithmic characteristic on drain current which related with V_{gs}. Fig. 2.4-3 shows the architecture. The positive end of photodiode connects with NMOS （M1） gate and drain node, thus may change gate voltage with various image current. If the image current I_{ph} flow into M1 which operate in sub-threshold region, the logarithm conversion signal can transfer to V_{gs}. M2 gets V_{gs} too, and helps to save image signal into capacitance C to prevent the timing variation.

While in the small image current, or in the rapid variation of pixel, the logarithm transfer maybe too slow and easily have shadow on the frame. By the way, how to revise the error and fix patter noise on various sub-threshold voltages would be the most important issue to enhance. Even though the noise of the architecture would be very large, but it is still a effect architecture to realize wide dynamic range.

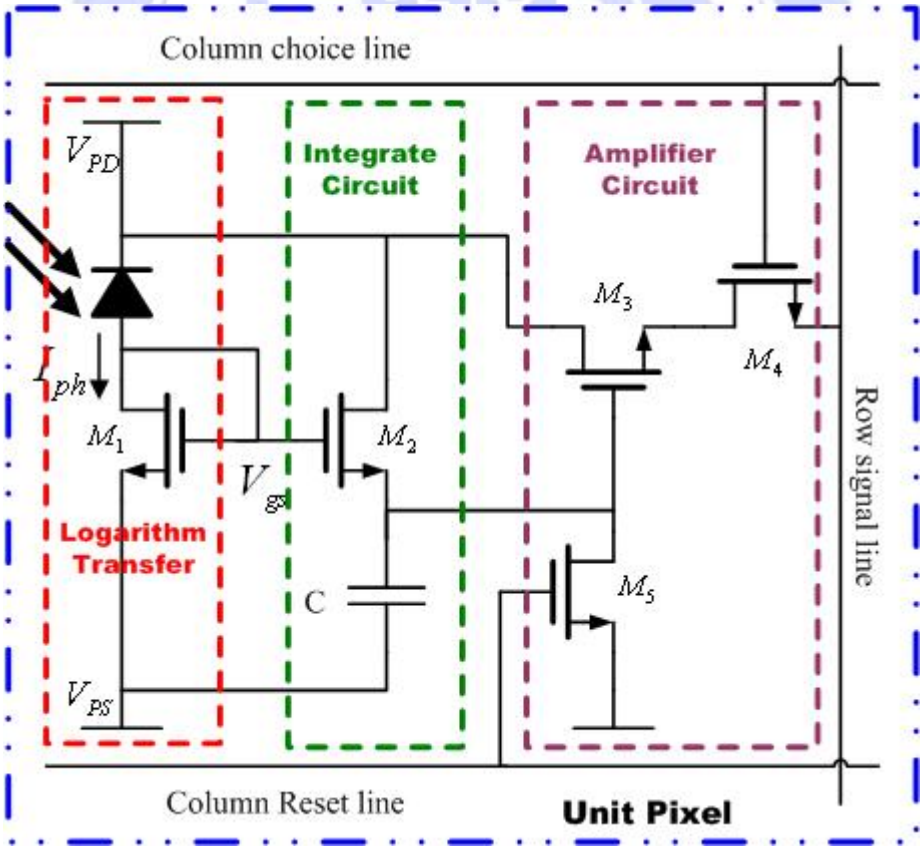
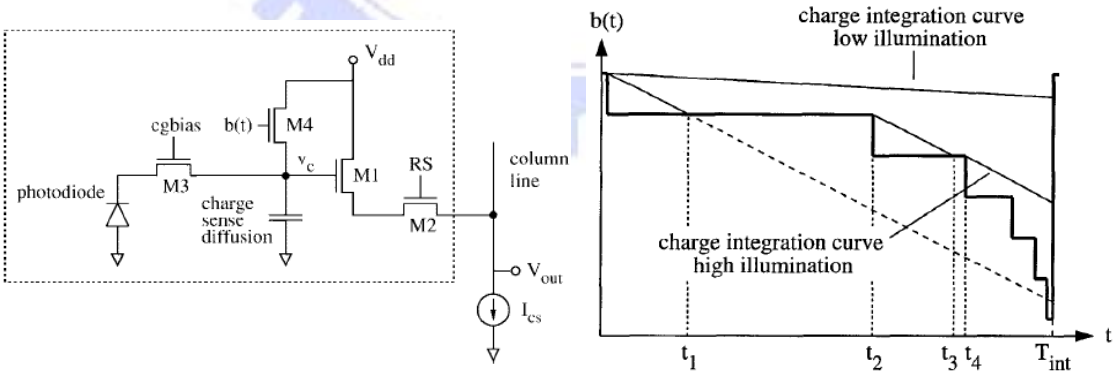


Fig. 2.4-3 Logarithm Conversion Pixel Circuit

extracted from 萩原義雄,ほか,"対数変換形 CMOS えエリア固体撮像素子",映像情報メデイア學会誌, Vol.54, No.2, pp.224~228, 2000.[26]

2.4.3 Capacitance Modulation Conversion [21][27]

The principle of this architecture is to vary maximum value to save electron on photodiode. Except the weak illumination, it use lateral overflow to achieve non-linear wide dynamic range effect. Schematics for the pixel are shown in Fig. 2.4-4（a）. The charge spill gate M3 increases sensitivity of the pixel by acting as a common gate amplifier that photocurrent flows into the low-impedance source node and is discharged into the high-impedance drain. The source follower Ml buffers the pixel from the large column line capacitance. The row-select device M2 connects the source follower output to the column line when the row is read out. The lateral overflow gate M4 increases pixel dynamic range. M4 gate voltage b（t） establishes a potential barrier to electron flow. As photo charge accumulates on the charge sense node, its charge level rises. If it exceeds the barrier level, the excess charge flows to the drain. Dynamic range is increased by decreasing b（t） over the integration period, as shown in Fig. 2.4-4（b）. For low illumination, the integrated charge is unaffected by the barrier, so the pixel retains all of the photo charge. For high illumination, photocurrent spills into the drain between t_1 and t_2, and between t_3 and t_4. This reduces the final integrated charge, as shown by the difference between the dashed and solid lines. As illumination increases, a greater proportion of the photocurrent is diverted to the drain. Depending on the size, number, and timing of the steps in b(t), any arbitrary compression characteristic（final integrated charge vs. illumination） is approximated.



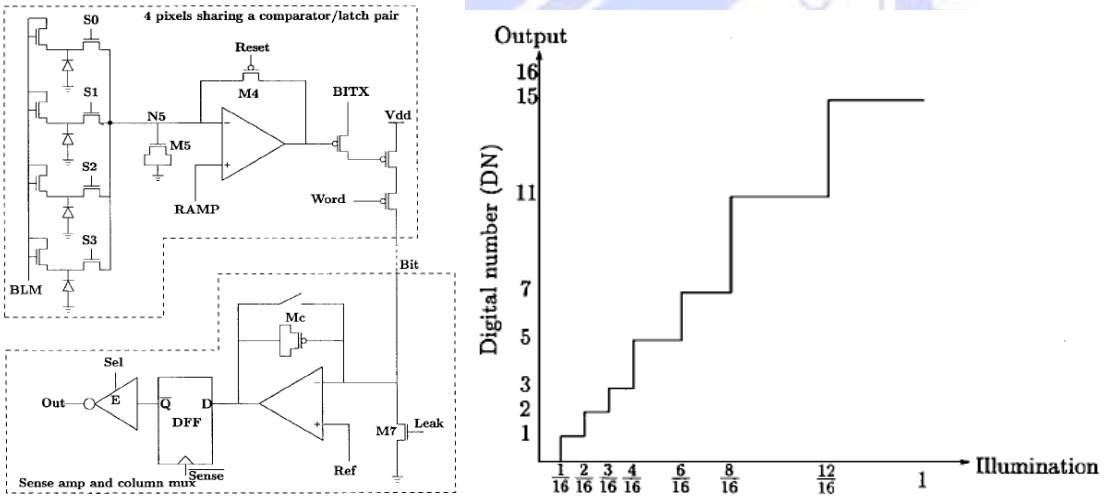
（a） （b）

Fig. 2.4-4 （a）Capacitance Modulation Pixel Circuit
（b）Signal Value versus Capacitance modulation

extracted from S. J. Decker, R. D. McGrath, K. Brehmer, and C. G. Sodini, "A 256 × 256 CMOS imaging array with wide dynamic range pixels and column parallel digital output," in ISSCC, Dig. Tech. Papers, pp.176~177, 1998.[27]

2.4.4 Pixel Level ADC with Multi-Sampling [22][28]

Ones method to embedded pixel level ADC and sampling at exponentially increasing exposure times, such as $T, 2T, ..., 2^k T$. Fig. 2.4-5（a） demonstrates the pixel circuit and column amplifier. To achieve acceptably small pixel size, each ADC, which is bit serial, is multiplexed among four neighboring pixels and generated by performing a set of comparisons between the pixel values and a monotonically increasing staircase RAMP signal. On the certain time, which N5 signal is equal RAMP one, the inversion signal would force BITX signal stay in the word transferring transistor to decide 1 bit data, and the Gray Code would be set up as the transferring reference. With different slope of RAMP, and m-bit serial A-D conversion k+1 times, they can get the higher resolution on digital data. Fig. 2.4-5（b）shows the m=2, k=2 with read serious signal in 4 bits（m + k）. They transfer m-bits in period of T, and conversion another bit in 2T and 4T（$2^2 T$）. This kind of architecture is different as other instead of increasing input signal's value. It uses the high resolution A-D conversion ability to analysis the weak illumination. Besides, if the illumination changed in saving period, the A-D conversion would have weakness on error.



（a） （b）

Fig. 2.4-5 （a）Embedded A/D Pixel block and column sense amplifier circuit （b）Illumination to output digital number transfer curve

extracted from D. Yang, A. El Gamal, B. Fowler, and H. Tian, "A 640X512 CMOS image sensor with ultrawide dynamic range floating-point pixel-level ADC," in IEEE JSSC, vol.34, pp. 1821~1834, Dec. 1999 .[22]

2.4.5 Pixel Level Analog Processing [8][29][30][31][32][33]

While pixel embedded with analog memory, it could be process with various functions in before research. In 2001, Y. Muramatsu et al. [8] use pixel level memory, with two exposure times for one-frame readout and correlated double sampling (CDS) circuit to provide on-chip analog operation, which pixel architecture like Fig. 2.4-6. It has three kinds of operation mode by controlling pixel memory, double sampling skill, and column parallel subtract circuit. They are motion detection, edge detection, and finally wide dynamic range impression.

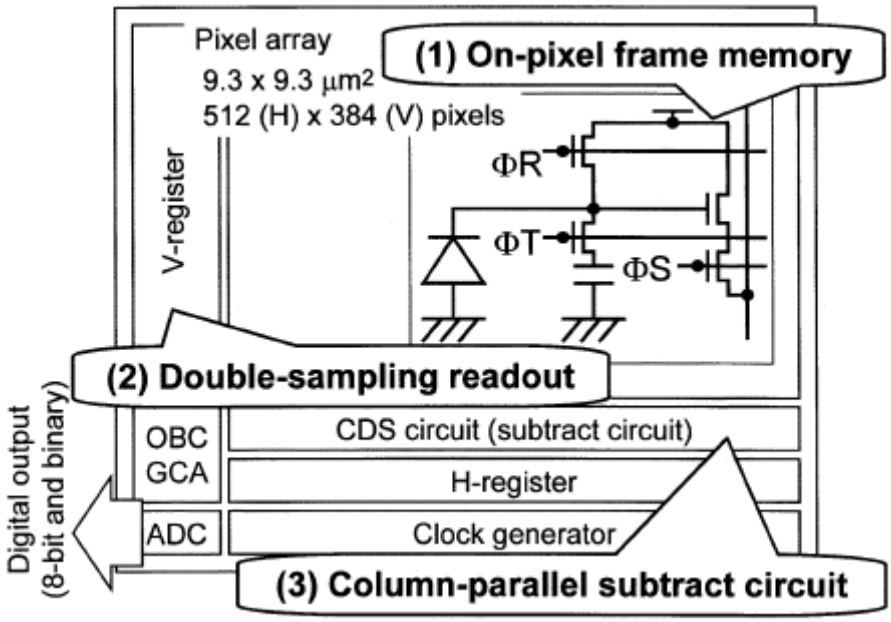


Fig. 2.4-6 Image sensor block diagram
extracted from Y. Muramatsu, S. Kurosawa, M. Furumiya, H. Ohkubo, Y. Nakashiba. "A Signal-Processing CMOS Image Sensor Using a Simple Analog Operation", IEEE JSSC, Vol. 38, No.1, January 2003[8]

Fig. 2.4-7（a） shows the timing chart and potential diagrams for a wide dynamic-range readout. The first exposure begins after the reset line (ΦR) and transfer line (ΦT) are pulsed high to low. The difference at high incident light and at low incident light is whether the excess charge overflows the photodiode into V_{DD} at the first exposure. In the high incident light （Fig. 2.4-7（b））, the first exposure take longer time to get overflow signal. Then, analog memory would turn on and divide first exposure signal into acceptable level. Finally, second exposure takes shorter time to get the rest of signal, and use CDS circuit to read out. The column parallel CDS circuit subtracts the dark signal from the wide dynamic- range photo signal, thus reducing the

fixed pattern noise. The low incident light （Fig. 2.4-7（c））operate in the same procedure, but it won't get the overflow response on second step. From results, we can recognize the architecture could transfer signal with larger illumination. The two kinds of exposure timing would be read by two channel CDS just like dual sampling mode.

As the general common, no mater capacitor which coupling in photo detector or real embedded analog memory, the division of signal electron would cause the worse sensitivity and redistribute noise. Besides, mix these two signal into one pixel may lead the error which is mismatch on saturation signal of photo detectors. However, without any external processing circuit, this architecture could also achieve the multi-functions as wide dynamic range.

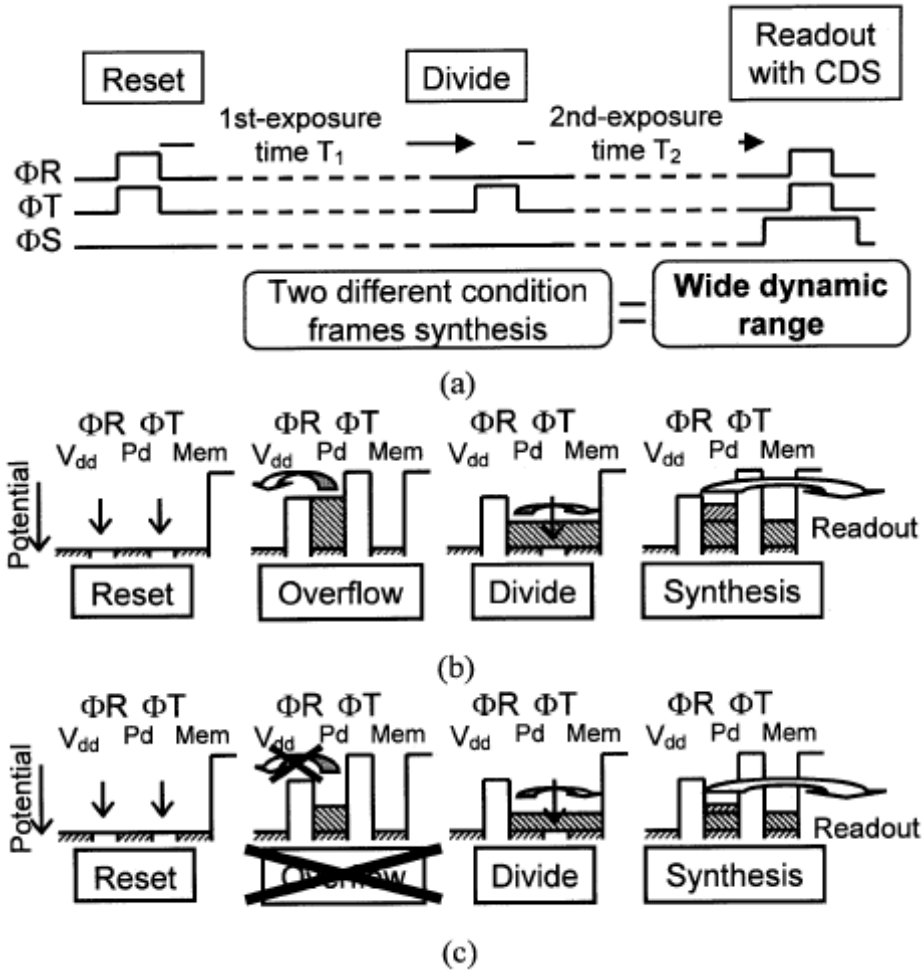


Fig. 2.4-7 Wide dynamic-rage readout (a) Timing chart (b) Potential diagram at high incident light (c) Potential diagram at low incident light

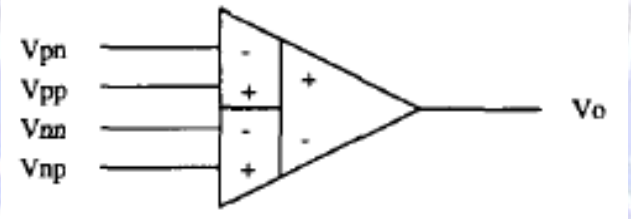
extracted from Y. Muramatsu, S. Kurosawa, M. Furumiya, H. Ohkubo, Y. Nakashiba., "A Signal-Processing CMOS Image Sensor using a Simple Analog Operation", in IEEE JSSC, Vol. 38, No. 1, pp.101~106, January 2003.[8]

2.5 Differential Difference Amplifier（DDA）[9][34]

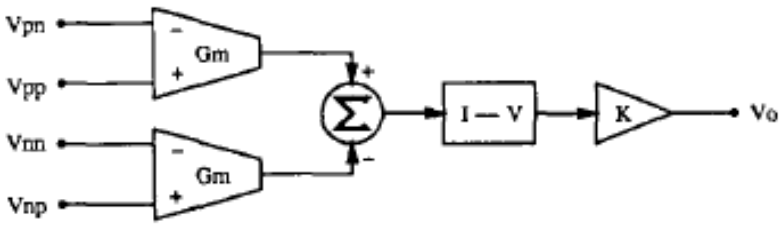
On this thesis, we would use function on double subtract for the signals we get. Therefore, we decide to use the Differential Difference Amplifier （DDA） architecture which bring up by S. C. Huang, et al., 1993 [34], and describe its basic concepts in this chapter. Fig.2.5-1（a） is its symbol, and Fig.2.5-1（b）shows its block diagram. We can realize DDA consist same architecture as general amplifier, with Gm level input stage, I-V transfer, and high gain stage. The main difference is that instead of two single-ended inputs as is the case in op-amps, it has two differential input ports. Therefore, the output of the DDA can be written as Eq.（12）.

$$V_o = A_o\left[\left(V_{pp} - V_{pn}\right) - \left(V_{np} - V_{nn}\right)\right], \tag{12}$$

where A_o is the open-loop gain of the DDA. Like the op-amp, the DDA consists mainly of two stages which show in Fig. 2.5-2 : a differential-input single-ended output transconductance stage (differential pair with active loads) and a second gain stage (common-source amplifier with an active load).



（a）



（b）

Fig. 2.5-1 (a) Symbol of DDA (b) Block Diagram of DDA
extracted from S. C. Huang, M. Ismail, S. R. Zarabadi, "A Wide Range Differential Difference Amplifier: A Basic Block for Analog Signal Processing in MOS Technology", in IEEE Trans. C.A.S.-II：Analog and Digital Signal Processing, Vol. 40, No.5, pp. 289~301, May 1993.[34]

The input differential range of the DDA is determined by the valid area of operation of the differential pairs at the input ports. It is will known that the profile of the differential output current (I_d) of a differential pair versus the differential input voltage (V_d) can be expressed as

$$I_d = \begin{cases} -I_o & ,if V_d \le -\sqrt{\dfrac{2I_o}{K}} \\ \dfrac{1}{2}KVd\sqrt{\dfrac{4Io}{K}} - V_d^2 & ,if\ |V_d| \le \sqrt{\dfrac{2I_o}{K}} \\ -I_o & ,if V_d \ge \sqrt{\dfrac{2I_o}{K}} \end{cases}, \tag{13}$$

where I_o is the tail current and K is the transconductance parameter of MOS transistors. Both transistors carry a current or $|V_d| \le \sqrt{\dfrac{2I_o}{K}}$. Outside this region, either current of the differential pair is zero. The other important parameter of the DDA is its input referred noise. It can be shown that the thermal and the flicker input referred noise of the DDA of Fig. 2.5-2 are given by

$$V_{thermal}^2(f) = \frac{32}{3}kT\left(\frac{1}{g_{mn}}\right) + \frac{16}{3}kT\left(\frac{g_{mp}}{g_{mn}}\right)^2\left(\frac{1}{g_{mp}}\right) \tag{14}$$

$$V_{flicker}^2(f) = \frac{2}{C_{ox}f}\left(\frac{2K_n}{W_nL_n}\right) + \left(\frac{\mu_p}{\mu_n}\right)\left(\frac{K_pL_n}{W_nL_p^2}\right), \tag{15}$$

where k is the Boltzmann's constant ($1.38\times10^{-23}JK^{-1}$), T is the temperature in Kelvin, $g_{mn} = g_{mi}$ (i = 1to 4), $g_{mp} = g_{m5} = g_{m6}$ are the small-signal transconductance of MOS transistors, K_n and K_p are the flicker noise constant for NMOS and PMOS transistors. W and L are the width and the length of the transistors, μ_n and μ_p are the carrier mobility, C_{OX} is the gate oxide capacitance per unit area, and f is the frequency. It is clear that the transconductance (g_{mn}) of the differential pairs should be made as large as possible to minimize the thermal noise. Also, increasing the width of the differential pair transistors clearly minimizes the $1/f$ noise.

The third important parameter is the linearity of the circuit. Since the DDA is used

in a closed-loop configuration, the larger the open-loop gain of the DDA the better is its linearity. The small-signal no-load gain of the DDA can be approximately expressed as

$$A_O = g_{mn}\left(r_{ds1} \| r_{ds3} \| r_{ds5}\right)\left(g_{m9} + g_{m10}\right)\left(r_{ds9} \| r_{ds10}\right), \tag{16}$$

where g_m and r_{ds} are the small-signal transconductance and the output resistance of MOS transistors. Clearly, the gain is directly proportional to g_{mn}. In our thesis, we would start from theorem and characteristic of DDA, and combine with other circuit to achieve certain performance which is needed on automotive application.

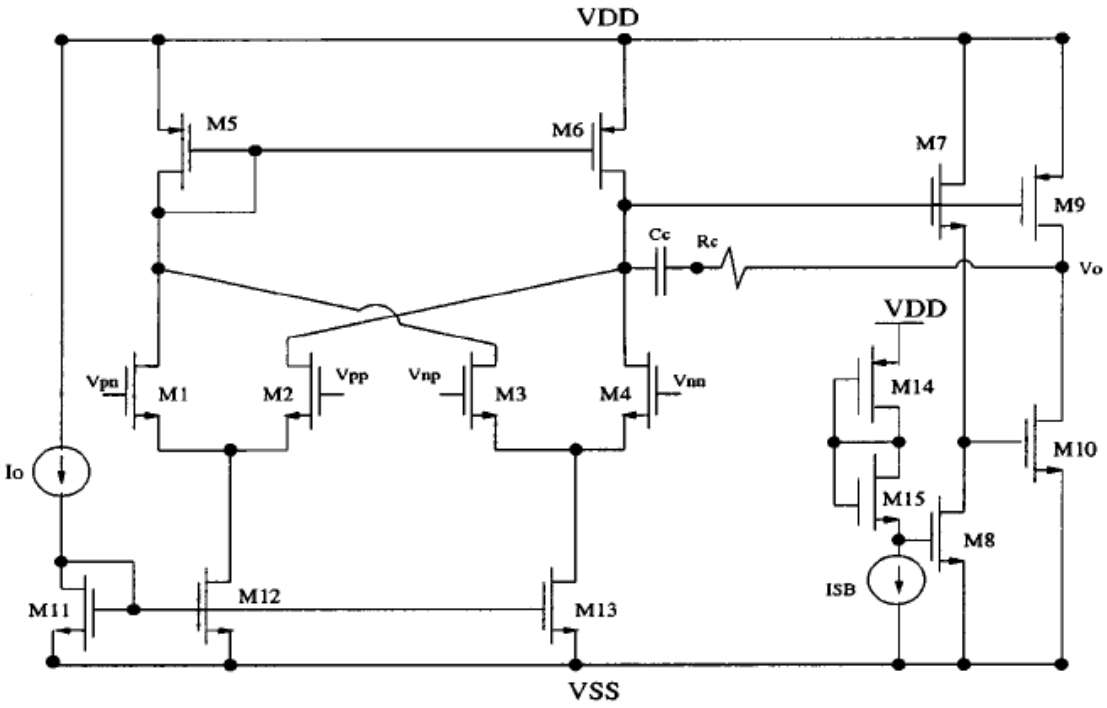


Fig. 2.5-2 Single ended DDA with Class AB realization

extracted from H. Alzaher, M. Ismail, "A CMOS Fully Balanced Differential Difference Amplifier and Its Applications", in IEEE Trans. C.A.S.-II：Analog and Digital Signal Processing, Vol. 48,No. 6, pp.614~620 , June 2001.[9]

2.6 Temperature Compensated Circuit [10]

From section before, it has been descried that gm in differential pair controlled the most of gain on amplifier design. By the way, when the temperature going higher, each transistor's V_{th} would getting larger than the one in general temperature, as Eq. （17） shows, which Φ_{MS} is the difference between the work functions of the polysilicon gate and the silicon substrate, $\Phi_F = \left(kT/q\right)\ln\left(N_{sub}/n_i\right)$, q is electron charge, N_{sub} is

the doping concentration of substrate, Q_{dep} is the charge in the depletion region, and C_{ox} is the gate capacitance per unit area .

$$V_{th} = \Phi_{MS} + 2 \times \Phi_F + \frac{Q_{dep}}{C_{ox}}, \tag{17}$$

$$I_{D,\max} = \frac{1}{2} \mu_n C_{OX} \frac{W}{L} \left(V_{GS} - V_{th} \right)^2, \tag{18}$$

The growing V_{th} would decay of differential pair's current as Eq.（18） shows, which μ_n is the mobility of charge carriers in NMOS. Finally, the transconductance gm in each transistor would also decay like current which proved relationship by Eq.（19）

$$g_m = \left. \frac{\partial I_D}{\partial V_{GS}} \right|_{VDS,const} = \mu_n C_{OX} \frac{W}{L} \left(V_{GS} - V_{th} \right), \tag{19}$$

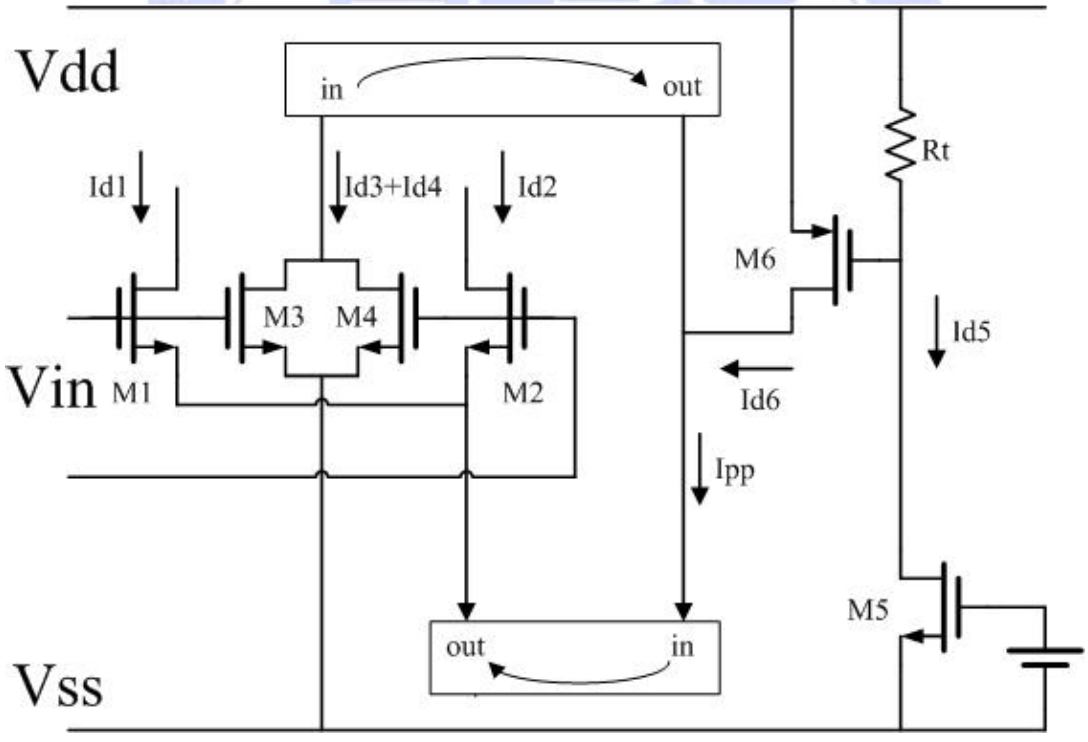


Fig. 2.6-1 Temperature Stable Differential Pair Architecture
extracted from J. A. S. Dias, W. B. de Moraes, "CMOS Temperature- Stable Liberalized Differential Pair", in ELECTRONICS LETTERS, Vol.28 No.25, pp.2350~2351 3rd December 1992. [10]

The paper posted by J.A.S. Dias, et al. on 1992 [10] present a new circuit, which

combine together on Fig. 2.6-1, has try to solve the problem. They use two transistors (M3, M4) with their drains connected to generate a current equal to

$$I_{d3} + I_{d4} = \frac{1}{4} K \left(V_{in} \right)^2 + K \left(V_{SS} + V_{th} \right)^2 , \tag{20}$$

$$K = \mu_n C_{OX} \frac{W}{L}$$

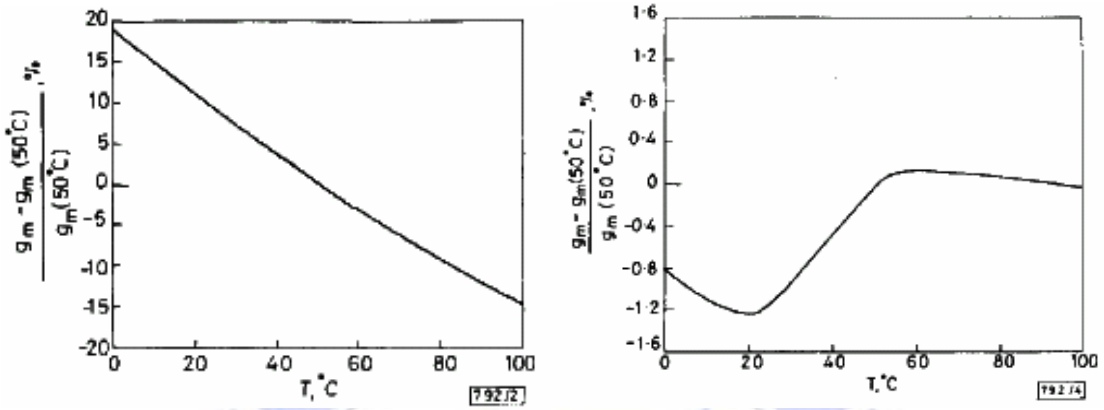
This current is passed through a current mirror, added to I_{pp}, and then is mirrored again to bias the differential pair (Ml-M2). It is easy to show that this type of biasing, with a tail current which is proportional to V_i^2 , results in an almost perfect linear transconductance, even for large values of the input voltage V_i. This imposes limitations in the application of such a technique to implement a circuit where the transconductance amplifier is used in an open-loop configuration.

The transconductance of the differential pair M1-M2, in the circuit presented in Fig. 2.6-1 can be written as

$$g_m = \sqrt{\left\{ K \left[K \left(V_{SS} + V_{th} \right)^2 \right] + K I_{pp} \right\}}, \tag{21}$$

where K is owing to the variation of the mobility, which is proportional to $T^{-3/2}$ and V_{th} ,may decrease linearly with temperature， if the current I_{pp}, is kept constant, the transconductance g_m will present a negative temperature coefficient（TC）, as shown on Fig. 2.6-2（a）. It can be seen from Eq.（19） that, by adjusting the TC of the current I_{pp}, it is possible to make $\partial g_m / \partial T = 0$. It is well known that the temperature coefficient of the drain current in an MOS transistor can be made positive or negative, by properly choosing the voltage V_{gs} and the W/L of the device. Transistor M5 is biased with a constant voltage source V_{gs} so that its drain current has a positive TC. However, we have a product of negative TC, and the current generated in transistor M5 alone is not sufficient to compensate for the high value and nonlinear negative TC of the transconductance g_m . To increase the TC of the current I_{pp}, the solution employed is to use a second transistor. This transistor, M6, is driven by a voltage $V_{gs} = I_{d5} \times R_t$. The value of R_t is chosen in such a way that the resulting V_{gs} forces M6 to operate with a positive TC drain current over the whole temperature range. Furthermore, because the

value of V_{gs} also increases with temperature due to the positive TC of I_{d5}, a high value nonlinear positive TC is obtained for the drain current of M6. This current（I_{d6}）used as I_{pp} in the original circuit. Fig. 2.6-2（b）shows g_m which has been compensated and almost 35 times reduction of the decay.



（a） （b）

Fig. 2.6-2（a）gm without Temperature Stable circuit, 3350ppm/˚C（0-100˚C）
（b）gm with Temperature Stable circuit, 95ppm/˚C（0-100˚C）
extracted from J. A. S. Dias, W. B. de Moraes, "CMOS Temperature- Stable Liberalized Differential Pair", in ELECTRONICS LETTERS, Vol.28 No.25, pp.2350~2351 3rd December 1992. [10]

2.7 Bandgap Circuit 【3】【13】

The principle of the bandgap voltage reference is to balance the negative temperature coefficient, such as the base-emitter voltage if bipolar transistors or forward voltage of pn-junction, with the positive temperature coefficient（TC） of the thermal voltage such as Proportional To Absolute Temperature（PTAT）circuit, we cab develop a reference having a nominally zero temperature coefficient.

To get negative TC, we should analysis the base emitter voltage of bipolar transistor. For a bipolar device we can write $I_C = I_S \exp V_{BE}/V_{th}$, where $V_{th} = kT/q$. The saturation current I_S is proportional to μkTn_i^2, where μ denotes the mobility of minority carriers and n_i is the intrinsic minority carrier concentration of silicon. The temperature dependence of these quantities is represented as $\mu \propto \mu_0 T^m$, where $m \approx -3/2$, and $n_i^2 \propto T^3 \exp\left[-E_g/(kT)\right]$, where $E_g \approx 1.12eV$ is the bandgap energy of silicon.

Thus,

$$I_S = bT^{4+m} \exp\frac{-E_g}{kT}, \tag{22}$$

where b is a proportionality factor. Transfer function as $V_{BE} = V_{th} \ln\left(I_C / I_S\right)$. Taking derivative of V_{BE} to T, and assume I_C is constant. Thus,

$$\frac{\partial V_{BE}}{\partial T} = \frac{\partial V_{th}}{\partial T} \ln\frac{I_C}{I_S} - \frac{V_{th}}{I_S}\frac{\partial I_S}{\partial T}, \tag{23}$$

Combination with Eq.（22）

$$\frac{\partial I_S}{\partial T} = b\left(4+m\right)T^{3+m} \exp\frac{-E_g}{kT} + bT^{4+m}\left(\exp\frac{-E_g}{kT}\right)\left(\frac{E_g}{kT^2}\right), \tag{24}$$

Therefore,

$$\frac{V_{th}}{IS}\frac{\partial I_S}{\partial T} = \left(4+m\right)\frac{V_{th}}{T} + \frac{E_g}{kT^2}V_{th}, \tag{25}$$

with the aid of Eq.（23）and Eq.（25）, we can get

$$\frac{\partial V_{BE}}{\partial T} = \frac{V_{th}}{T}\ln\frac{I_C}{I_S} - \left(4+m\right)\frac{V_{th}}{T} - \frac{E_g}{kT^2}V_{th}, \tag{26}$$

$$= \frac{V_{BE} - \left(4+m\right)V_{th} - E_g/q}{T}, \tag{27}$$

Eq.（27）gives the temperature coefficient of the base-emitter voltage at a given temperature T, revealing dependence on the magnitude of V_{BE} itself. With $V_{BE} \approx 750mV$, and T=300°K, $\partial V_{BE}/\partial T \approx -1.5mV/°K$.

It was recognized in 1964[35] that if two bipolar transistors operate at unequal current densities, then the difference between their base-emitter voltages is directly proportional to the absolute temperature. As Fig. 2.7-1 shows, if two identical transistors（$I_{S1} = I_{S2}$）are biased at collector currents on nI_0 and I_0, where the difference on bipolar's size, their base currents are negligible, then

$$\Delta V_{BE} = V_{BE1} - V_{BE2}, \tag{28}$$

$$= V_{th} \ln\frac{nI_0}{I_{s1}} - V_{th} \ln\frac{I_0}{I_{s2}}, \tag{29}$$

$$= V_{th} \ln n, \tag{30}$$

Thus, the V_{BE} difference exhibits a positive temperature coefficient:

$$\frac{\partial \Delta V_{BE}}{\partial T} = \frac{k}{q} \ln n, \tag{31}$$

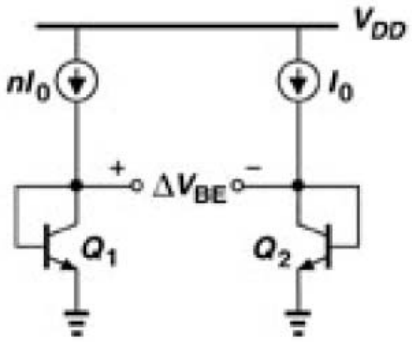


Fig. 2.7-1 Generation of PTAT voltage

extracted from Behzad Razavi, "Design of Analog CMOS Integrated Circuits", McGraw-Hill Companies, Inc. 2004.【3】

Combine the negative and positive TC as Fig. 2.7-2 shows, we can get the reference voltage $V_{REF} = \alpha_1 V_{BE} + \alpha_2 V_{th} (\ln n)$, where $V_{th} \ln n$ is the difference between the base-emitter voltages of the two bipolar transistors operating at different current density.

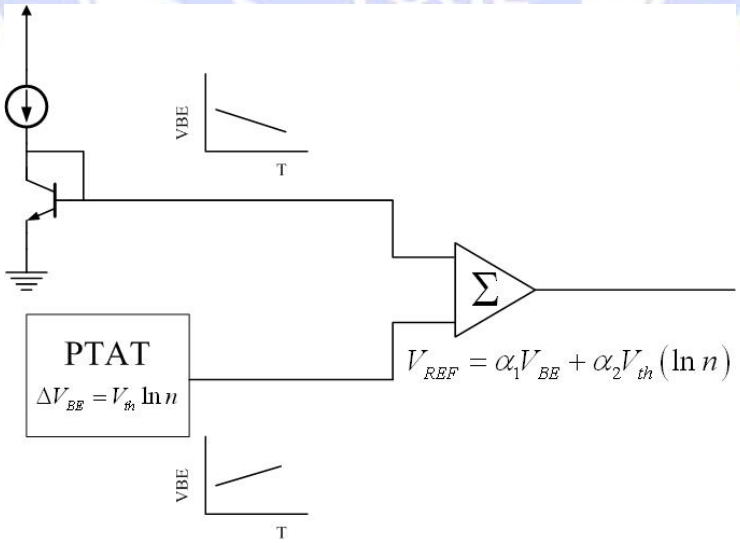


Fig. 2.7-2 Bandgap Reference Theorem

extracted from Prof. G. A. Rincón-Mora, "Bandgap Voltage Reference", Analog Integrated Circuit Handout【13】

Since at the room temperature, it would let $\partial V_{BE} / \partial T \approx -1.5 mV/°K$ and where as

the $\partial V_{th}/\partial T \approx +0.087mV/°K$, we may set α_1 =1 and choose $\alpha_2 \ln n$ such that help$(\alpha_2 \ln n)(0.087mV/°K) = 1.5mV/°K$, that is $\alpha_2 \ln n \approx 17.2$, indicating that for zero TC. For these parameters, we can get $V_{REF} \approx 1.25V$. Fig. 2.7-3 shows the implementation accomplishing help $V_X = V_Y$. Amplifier A1 sense V_X and V_Y , driving the top terminals of R_1 and R_2 （ $R_1 = R_2$ ） such that X and Y settle to approximately equal voltages. The reference voltage is obtained at the output of the amplifier which,

$$V_{out} = V_{BE2} + \frac{V_{th} \ln n}{R_3}(R_3 + R_2), \tag{32}$$

$$= V_{BE2} + (V_{th} \ln n)\left(1 + \frac{R_2}{R_3}\right), \tag{33}$$

For a zero TC, we must have $(1 + R_2/R_3)\ln n \approx 17.2$. For example, we may choose n = 31 and $R_2/R_3 = 4$. From now on, we've described the basic of bandgap reference, and we would analysis the one which used in this thesis on next chapter.

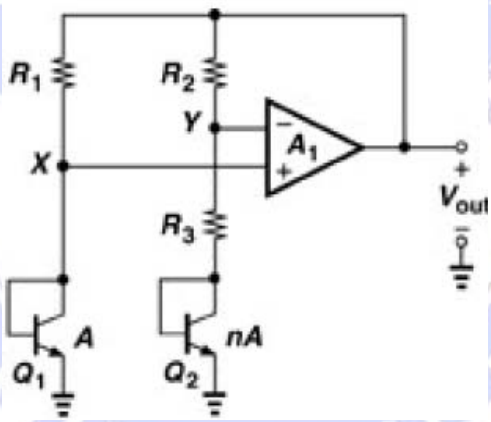


Fig. 2.7-3 Actual implement of the concept of temperature independent voltage
extracted from Behzad Razavi, "Design of Analog CMOS Integrated Circuits", McGraw-Hill Companies, Inc. 2004.【3】