Back-thinned CMOS Sensor Optimisation

Paul Jerram[a], David Burt[a], Neil Guyatt[a], Vincent Hibon[b], Joel Vaillant[b], Yann Henrion[b]
[a]e2v technologies, Waterhouse Lane, Chelmsford, CM1 2QU, UK
[b]e2v semiconductors SAS, Avenue de Rochepleine, Saint-Egrève, France

ABSTRACT

Back-thinning of a CCD image sensor is a very well established process for achieving high quantum efficiency and the majority of high-specification space and science applications have used such back-thinned devices for many years. CMOS sensors offer advantages over CCDs for a number of these applications and, in principle, it should be possible to back-thin CMOS devices and obtain the same performance as the CCD. This has now been demonstrated by e2v and results from two recent programmes to back-thin CMOS sensors show excellent quantum efficiency values.

1 INTRODUCTION

The technology of thinning and back-illumination is now well established for CCD image sensors. By avoiding losses in the electrode structure, a high quantum efficiency can be obtained right across the useful spectral range of silicon from X-ray to near infra-red wavelengths. Such devices are now the primary choice for high-specification space and science applications.

CMOS image sensors offer certain performance advantages over CCD, such as the possibilities for integrating much of the imaging system on the chip and the provision of digital outputs, and it seems possible that the same back-illumination technology can be applied to these devices.

This paper firstly describes the basic principles and then reports results from two recent programmes that have successfully produced back-illuminated CMOS sensors.

2 BASIC PRINCIPLES

The following analysis gives the basic requirements for achieving high quantum efficiency.

If monochromatic light is incident on silicon and the incident flux is F photons/cm²/s, then ignoring possible reflection the rate of carrier generation at depth x is given by the following.

$$G(x) = (F/L)\exp(-x/L) \text{ electrons/cm}^3\text{/s}$$

where L is the absorption length. Alternatively, $1/L = \alpha$, where α is the absorption coefficient.

The equation shows that for any value of L the rate of generation has a maximum value of (F/L) near the surface, and that the rate then falls with distance x according to the exponential factor reducing to about 37% at x = L, to about 14 % at x =2L, and so on.

L is wavelength dependent [1] and has a minimum value of about 10 nm at UV wavelengths increasing to 1-10 µm at visible wavelengths and hundreds of microns in the near infra-red. Thus at UV wavelengths the generation is very close to the surface, at visible wavelengths the generation is spread over a depth of about 20 µm and at near infra-red wavelengths the generation can be spread over many hundreds of microns. This large spread in the range of charge generation can be used to explain the spectral response typically seen with CCD and CMOS sensors operated with front-face illumination.

2.1 Front illumination quantum efficiency

Typical CCD cross-sections and the corresponding spectral response curves are shown in figures 1 and 2, respectively.

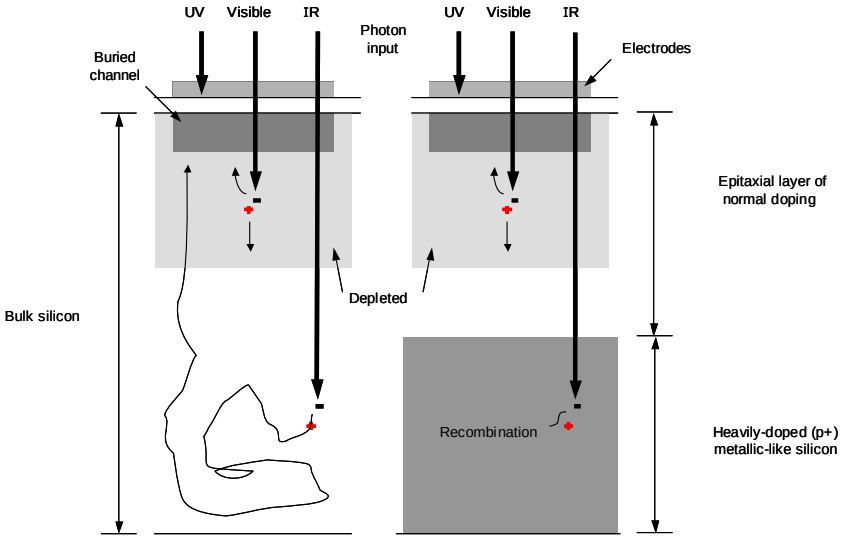

Fig 1 – CCD cross-sections: a) bulk silicon, b) epitaxial silicon

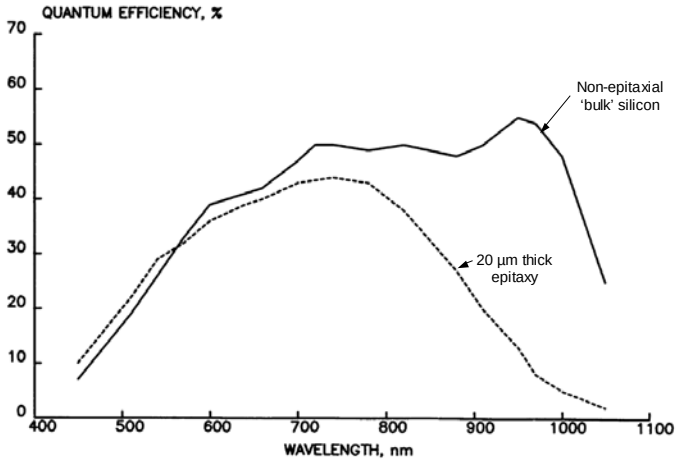

Fig 2 – Measured front-face CCD spectral response

The electrodes are typically 0.5 µm thick and being of silicon (in polycrystalline form) adsorb all the photons of wavelength shorter than about 400 nm leaving none to enter the active silicon and generate signal charge. The quantum

efficiency QE, the fraction of the incident photons giving useful signal, is therefore zero. The adsorption length increases at longer wavelengths such that the photons enter the active silicon and generate signal The electrodes do however reflect some light and there are additional losses through multiple reflections and interference with the result that the QE reaches a maximum of only about 40-50% at a wavelength of about 700 nm. In the case of devices fabricated on uniformly doped substrate material of thickness in the region of 600 µm, the response is maintained out to wavelengths of about 1 µm, but then falls through the value of L becoming very large in relation to the device thickness. It has generally been the case that readily available silicon has a minimum doping concentration of about 10^{14} atoms/cm^3 and that the resulting depth of depletion is no higher than about 10-20 µm. The deeply-generated charge can diffuse laterally in the underlying field-free silicon giving rise to poor spatial resolution. The solution adopted is the use of epitaxial silicon. The silicon wafers are fabricated with the underlying regions heavily doped to reduce the minority carrier diffusion length, but leaving an active 'epitaxial layer' of normal silicon at the surface, as can be seen in figure1b. The layer thickness is chosen to give the best trade-off between response and resolution. An added benefit of the structure found in the early days of its use (1980s) was a considerable improvement in the dark current performance, especially a reduced incidence of white-spot defects. This was found to be due to an inherent 'intrinsic gettering' mechanism [2] attracting contamination away from the active region and into the underlying inactive regions below. It may be noted that the success of this feature had lead other types of semiconductor manufacture to also use epitaxial silicon, e.g. CMOS and MOS memory devices.

For CMOS sensors the QE curve has a different form. A device cross-section is shown in figure 3. It will be noted that a fraction of the surface of the device is now covered in metal tracks and this area is insensitive to all illumination. The transistors will collect any charge generated directly within their p-well region, but not necessarily that generated in the silicon below it. The remaining area with the photo-diode has an oxide and possibly nitride coating and so can have measurable QE below 400nm (unlike CCD with the losses in the electrodes). The critical parameter in determining the QE of a CMOS detector at shorter wavelengths is therefore the open area ratio (or 'fill-factor') and this will vary significantly depending on the number of tracks and transistors used and the size of the pixel. A spectral response curve for a device with an open area ratio of 55% is shown later in figure 11. As a large pixel tends to have a large open area ratio the benefits of back-illumination are not great although it does allow the addition of an AR coating, whereas for small pixels with complex structures back-illumination becomes increasingly essential. This is the opposite of the

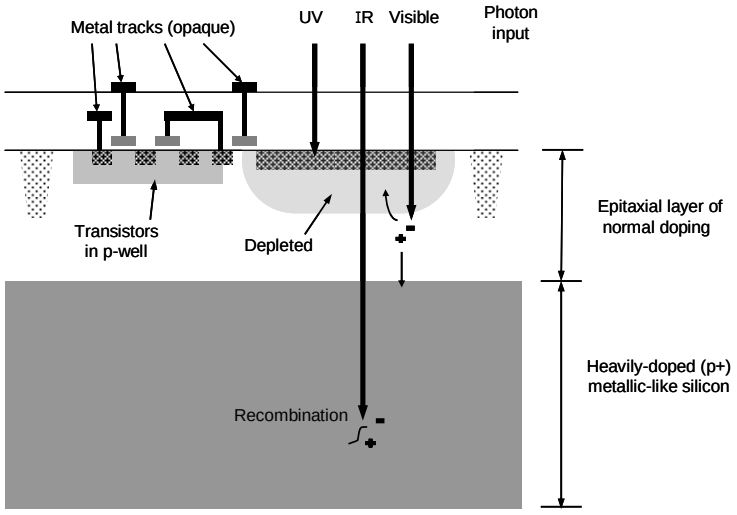

Fig 3 – CMOS sensor Cross Section

situation with CCDs where the QE does not vary significantly with pixel size. Note that these CMOS sensors are usually fabricated on epitaxial silicon (as shown in figure 3) with a layer that is thinner than typically used for CCDs, with the consequence that the QE falls off faster towards the infra-red.

2.2 Improving red response

More recent results are that bulk material of much lower doping concentration is now more readily available and that CCD fabrication processes are of a far higher quality such that intrinsic gettering is no longer essential for high-yield manufacture. The problem of loss of spatial resolution is being addressed by applying a back-substrate bias to deplete the full silicon thickness up to 300µm and ensure that charge drifts to the point of collection with minimal sideways spreading. This new approach is termed ‘Hi-rho’ by e2v technologies.

Further experimentation will be necessary to determine whether such technology can be applied to CMOS sensors.

2.3 Improving UV response

Clearly, whatever the substrate properties, there can be no response to UV light with a basic front-illuminated CCD. A first solution is to use an ‘open-electrode’ structure [3] with areas of the pixel not covered with polysilicon. There is a trade-off with open area and full-well capacity and a QE of about 25% is readily achieved. Similar results can be achieved with a CMOS sensor with a high open area ratio. The problem in both cases is that most of the pixel will have no response, so the approach is not suitable for resolving fine image detail. The final universally accepted solution for achieving both UV response and a high QE over a wide spectral range is to illuminate the device on the back face to avoid the losses in the front-face structures.

Various manufacturers have developed fabrication technologies by for producing back-face illuminated devices, the process developed by e2v technologies for CCD image sensors and now being applied to CMOS sensors is next described.

The process has been developed in the days of the silicon having a doping concentration that provides a depth of depletion in the region of 10-20 µm. Thus, simply illuminating the back of the 600 µm thick device is not sufficient as charge spreading from the shorter wavelength generation near the back surface gives an unacceptable loss of resolution. The solution is to thin the silicon by etching such that its thickness becomes comparable with the depth of depletion. However, silicon of this thickness is quite flexible and a mechanical support is necessary. A back-etched surface also tends to have a high concentration of generation-recombination centres or traps that can cause any photo-generated charge to immediately recombine. A ‘passivation’ layer is therefore required. A further problem is that the silicon surface has a high reflectance and an anti-reflection coating AR is essential for achieving high QE (but not, however, for devices intended for soft X-ray applications). The means used for achieving these objectives are as follows.

The fabrication sequence is shown in figure 4.

a) The silicon wafer with completed devices is glued or molecularly-bonded face-down on a blank silicon wafer to be used for support purposes.
b) The device-wafer silicon is etched down to the desired thickness. This is facilitated with the use of epitaxial silicon as the etch concentration can be such as to slow at the change of substrate doping, thereby tending to produce more uniform thickness. A variation of typically ± 0.5 µm is typically achieved.

c) The passivation and AR coating are processed; more details are given later
d) The silicon above the metal bond pads is etched away.
e) The wafer is sawn to release individual chips and these are packed in the same way as front-face chips (but with the wire bonds now effectively made to the opposite side of the metal pads).

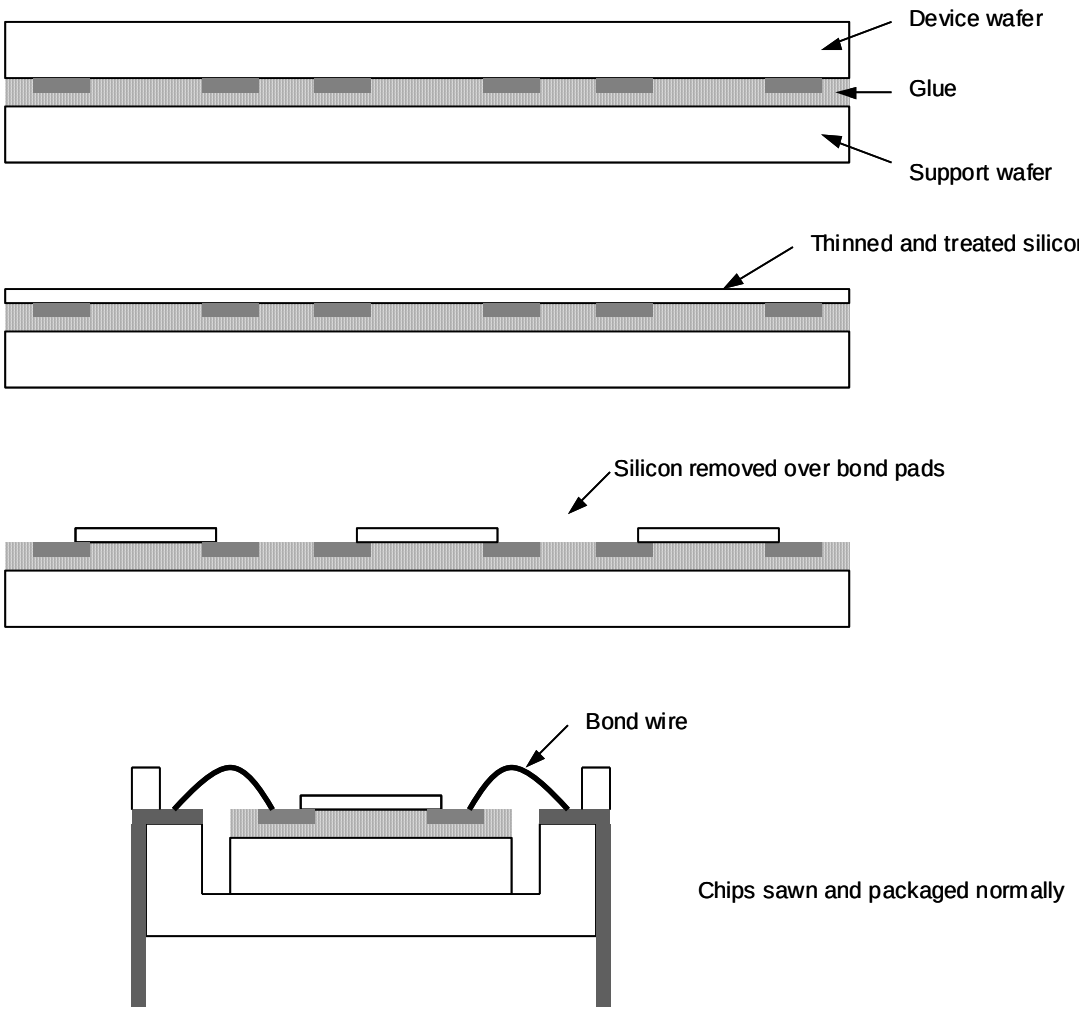

Fig 4 – Back-face fabrication sequence

Details of the back-surface treatments are as follows.

2.4 Passivation

There are basically two methods for minimising the effects of the back-surface traps. A first is to cause the native oxide to hold a negative charge such that electrons are repelled from the surface regions. A second, as preferred by e2v technologies on account of potentially higher stability, is to introduce a thin layer of higher doping concentration at the back surface, p+, as shown in figure 5 for a CCD, CMOS is similar. Since in any semiconductor the local product of the

hole and electron concentrations is always attempting to reach the equilibrium value of n_i^2, the minority carrier electron concentrations in the p+ layer and the underlying silicon are in inverse ratio to the p-type doping concentrations. The probability of an electron within the silicon diffusing back into the p+ layer is therefore simply this ratio. It may be noted that the inhibiting nature of the p+ layer can also be described in terms of an equivalent voltage step (as shown) or field, but the net effect is exactly the same.

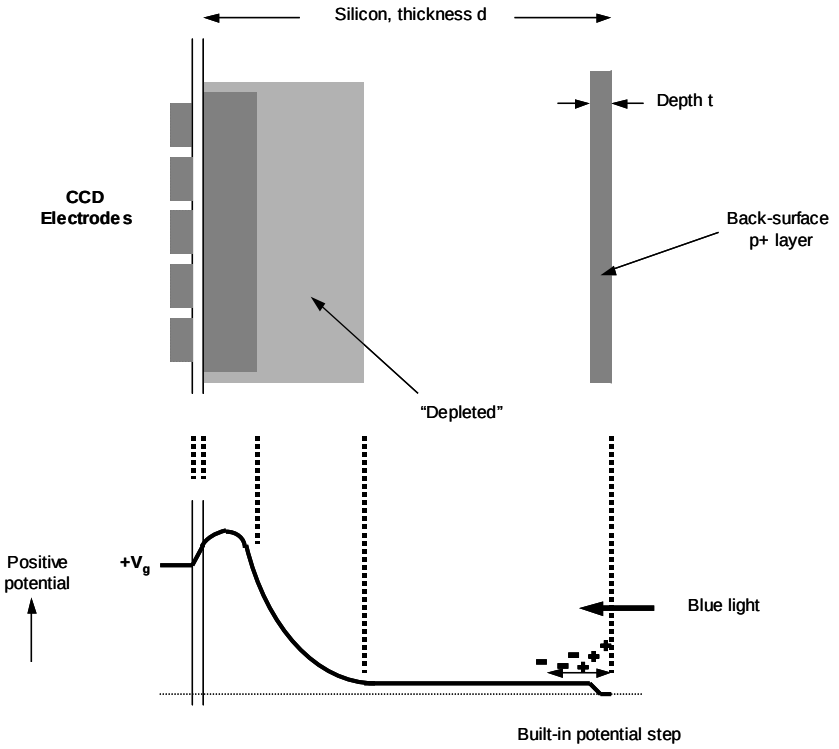

Fig 5 – Incorporation of back-surface p+ layer

Analysis of the charge transport assuming a uniformly doped p+ layer [4] shows that the probability of a photo-generated electron being collected as signal is P_o at the surface, then increasing to 100% at the edge of the p+ layer and within the active silicon. The actual value for P_o depends on a variety of factors including the trap density and the oxide charge, neither of which can be independently measured. However, for a high trap concentration P_o becomes zero, and this can be taken as a worse case.

Combination of this probability function with the rate of generation given by equation 1 leads to a differential equation that can be solved to give the QE, assuming zero reflectance.

$$QE \sim P_o + (1 - P_o)\,(L/t)\,(1 - e^{-t/L}) - e^{-d/L}$$

In this d is the total silicon thickness.

Under conditions where the ratio t/L is larger than about 3 (generally blue and UV wavelengths) and where $P_o \sim 0$, the equation may be simplified to the following.

$$QE \sim L/t$$

Clearly, therefore, t should be a small as possible to obtain maximum response at the UV end of the spectral range.

At longer wavelengths where t/L becomes small, the QE approximates to $(1- e^{-d/L})$ and the mid-band response is therefore independent of the p+ layer thickness. The response at longer wavelengths is dependent on the total silicon thickness d.

The method used by e2v technologies to form the p+ layer is a low-energy boron implantation. Normally such implants require a high temperature anneal to activate the boron (i.e. cause the atoms to substitute in the silicon lattice), but this cannot be a furnace treatment with the glued option because the glue cannot withstand temperatures higher than about 150°C. The solution has been to use a UV laser to heat to the required temperature just a shallow depth of silicon. This is normally done with a small area of illumination stepped-and-repeated across the device surface that has typically left a low-contrast pattern visible in the UV response. Recent optimisation has however resulted in such patterns becoming virtually invisible. An alternative process for devices made using molecular bonding is to carry out a thermal anneal. However, the maximum temperature that can be used is still no higher than 400°C and this causes only a fraction of the born atoms to be substituted into the lattice. This process does however give a QE that is acceptable in the visible region, but drops rapidly in the UV and is then more dependent on operating temperature.

Typically the p+ layer is doped to at least 10^{18} atoms/cm³ and of a thickness t in the region of 0.05 µm. For the more demanding applications a thinner layer is obtained by etching off some of the silicon.

It is reported that molecular beam epitaxy can be used to form an ultra-thin p+ layer [5], which possibly gives the ultimate level of performance for this approach, but such equipment is not generally available throughout the industry.

2.5 Anti-reflection coatings

Across the useful response spectrum the refractive index of silicon is about 3.5 in the infra-red, gradually increasing to about 7.0 at 400 nm wavelength, then decreasing rapidly in value at UV wavelengths with a growing imaginary component. The corresponding reflectivity increases from about 30% to about 60%. The classic method to minimise such reflection is to use a coating of refractive index equal to the geometric mean of the silicon and air. Zero reflection is then obtained at the wavelength where the layer is a quarter-wavelength thick.

Figure 6 shows that near-ideal performance can be obtained at wavelengths longer than about 400 nm using a layer of hafnium oxide with a refractive index of about 1.9. Good results can also be achieved at shorter wavelengths, but performance is less-ideal because of the complex behaviour of silicon. Other layer materials can also be used, e.g. Si_3N_4.

Figure 7 shows that a generally broader response can be achieved with a two-layer coating, where the upper layer has a refractive index the geometric mean of air and the next layer.

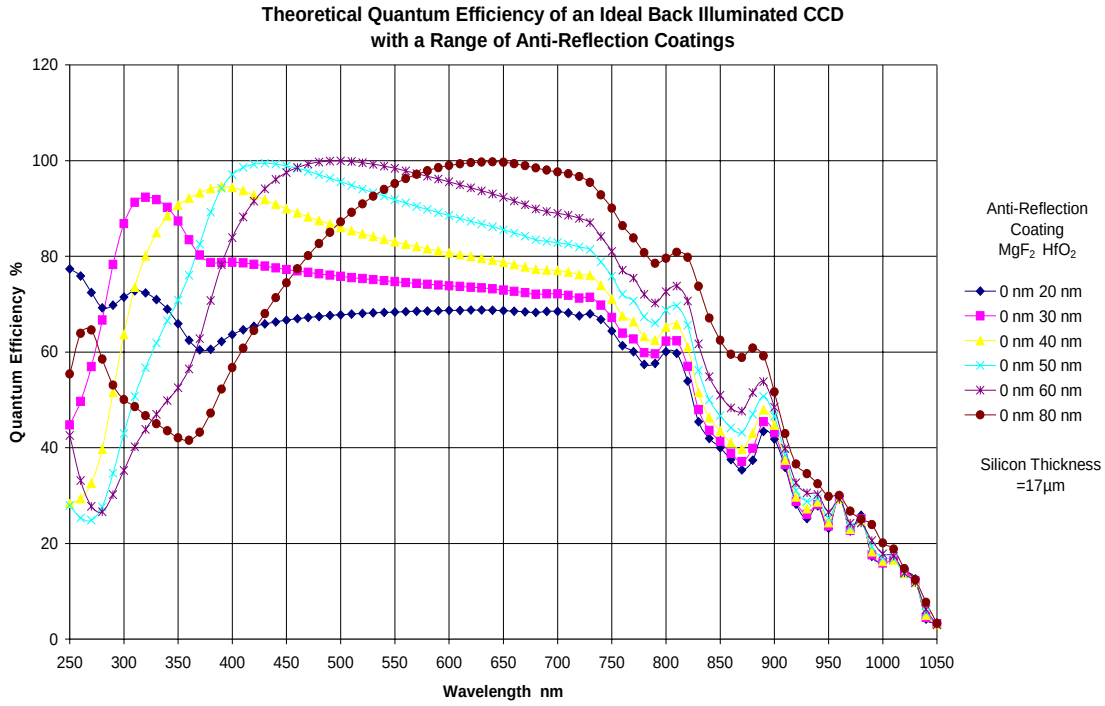

Fig 6 – Response using a single-layer AR coating

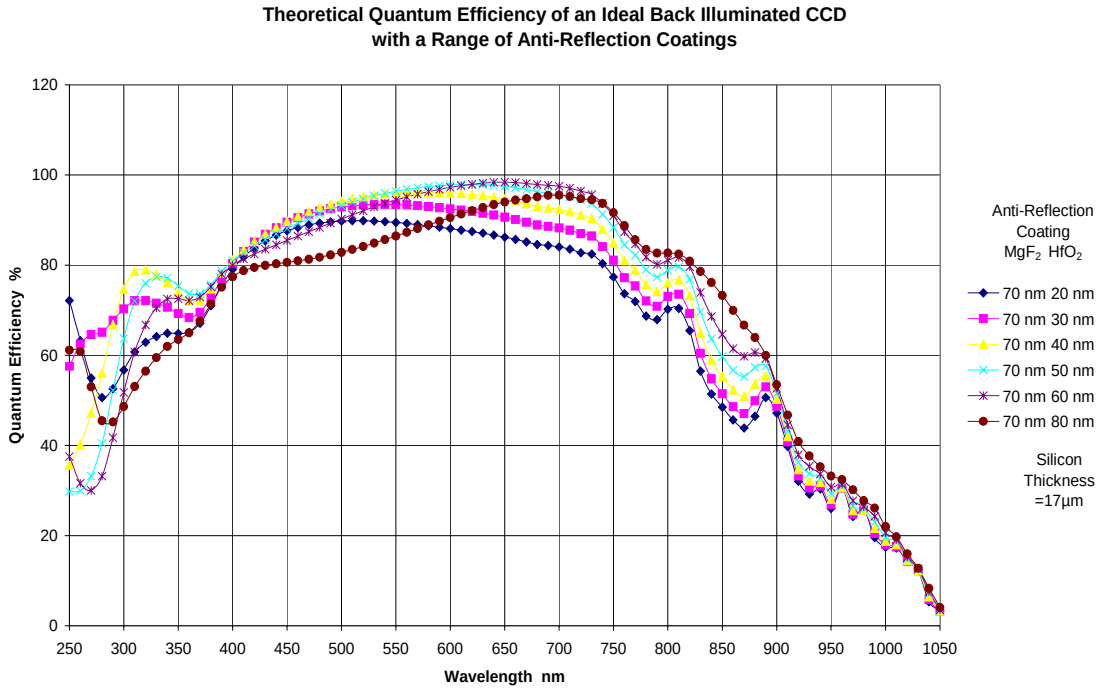

Fig 7 – Response using a two-layer coating

3 PRACTICAL RESULTS WITH CCD SENSORS

As an example of what can be achieved with CCD sensors, figure 8 shows results obtained from the Gaia flight devices [6]. The QE values are in excellent agreement with the theoretical prediction based on the p+ layer thickness and AR coating that are used, and are very consistent from device to device.

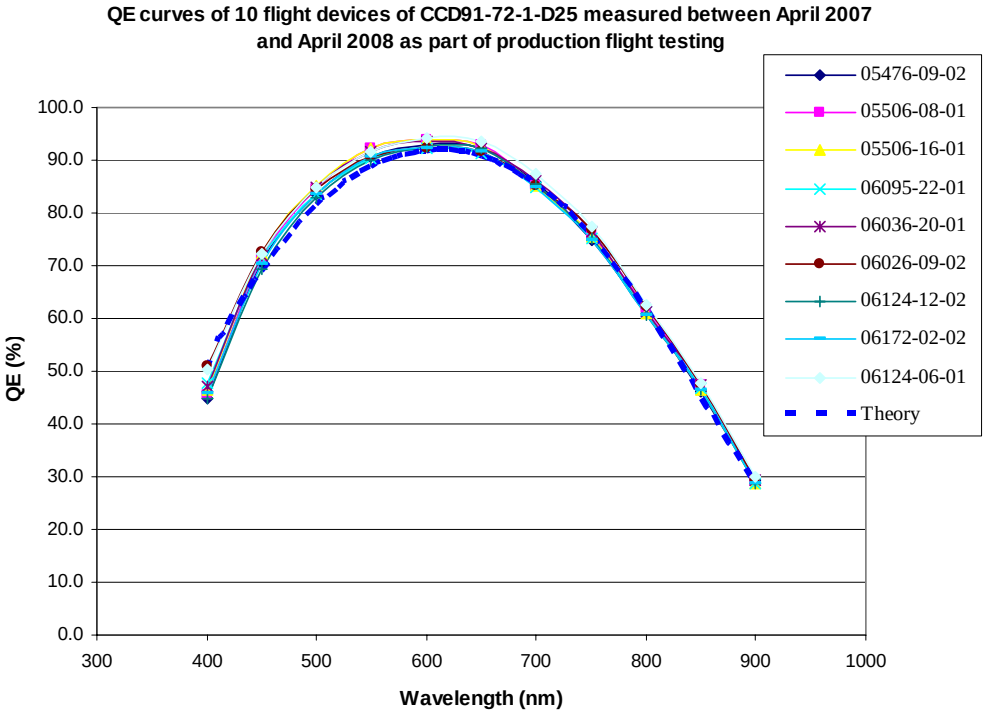

Fig 8 – Spectral response curves for Gaia CCD sensors

4 PRACTICAL RESULTS WITH CMOS SENSORS

Since CMOS sensors are generally fabricated on silicon of epitaxial format, it is possible to use the same etching techniques as are used for CCD. However, as the layer thickness can be as small as 5 µm and the etching normally goes a few microns into the layer, using a standard device can result in the silicon becoming unacceptably thin. It is therefore necessary for devices to be fabricated on silicon with an increased layer thickness.

Practical results are as follows.

4.1 0.5Mpixel sensor

This sensor has 864 x 640 pixels each 5.8 µm square and is intended for general-purpose industrial applications. The pixel is a 5T type with a pinned-photodiode and a shutter function. The frame rate is 60 Hz with an 8-bit digital output.

Increased red and near infra-red response the devices are fabricated on epitaxial silicon of increased layer thickness. After thinning the silicon is in the region of 8 µm thick. The AR coating is 85 nm of Si_3N_4. Results are shown in figure 9.

The fact that the actual and modelled performances are close suggests that the internal losses are very small. This is an important result meaning that the p-well in which the in-pixel transistors are fabricated provides an effective shield to prevent pick-up of charge that is generated in the underlying silicon (i.e. nearly all the photo-generated charge flows to the photodiode). The basic CMOS technology is therefore ideal for scientific applications requiring the highest QE. The AR coating is not of course fixed but can be tailored to suit particular requirements.

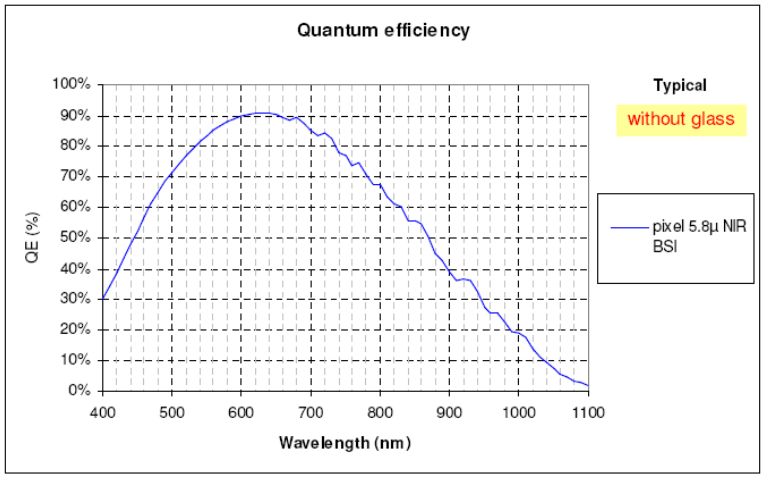

Fig 9 – 0.5Mpixel CMOS sensor back-face response

The global shutter efficiency has also been measured under back-illumination and values are shown as a function of wavelength in figure 10. At shorter wavelengths when all of the electrons are generated in the field free region near the back surface of the silicon the efficiency is approximately 500:1, i.e. only 0.2% of the charge reaches the detection node. This ratio implies that the fields are favourable in that they draw the electrons to the photodiode. At longer wavelengths charge is generated deeper within the silicon and therefore directly at the detection node and a high efficiency is much more difficult to obtain, hence the values are now approximately 200:1. It may be noted that this device was not designed for back-illumination and with improvements to the geometry it will certainly be possible to improve on these ratios. For comparison, the front-illuminated global shutter efficiency is 1700:1 at 480nm and 600:1 at 660nm.

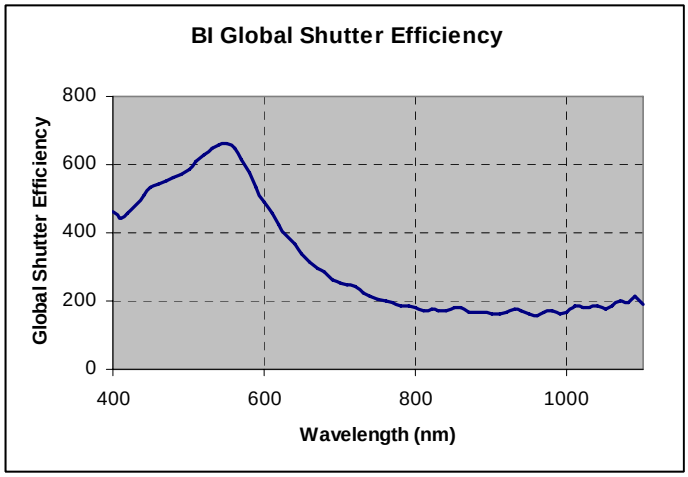

Fig 10 – 0.5Mpixel CMOS sensor shutter efficiency

4.2 2Mpixel sensor

This sensor has 1415(H) x 1430(V) pixels each 14.81 µm (H) x 11.53 µm (V) with 4 analogue outputs operating at 10 MHz. It was designed and developed by Astrium in cooperation with ISAE/CIMI for the GOCI space instrument, with space qualification and assembly by e2v technologies.

Although the sensor was designed primarily for front-illuminated operation, two wafers were thinned for back-illumination. The first wafer was processed using a low temperature thermal anneal and hence has relatively poor QE in the UV region, the AR coating thickness being approximately 65nm. The QE curves from the back- and front-illuminated versions of this sensor are shown in figure 11, the front-face open area ratio being 55 %. The fact that the QE values are measured with a narrow bandwidth monochromator shows up the fluctuations in the front-face response that are due to optical interference in the multiple dielectric stack over the photodiode (other broader bandwidth methods would show an artificially smoothed curve), These fluctuations are absent in the response of the back-illuminated sensor at visible wavelengths, but there are now fluctuations at longer wavelengths from interference between the front and back surfaces of the silicon.

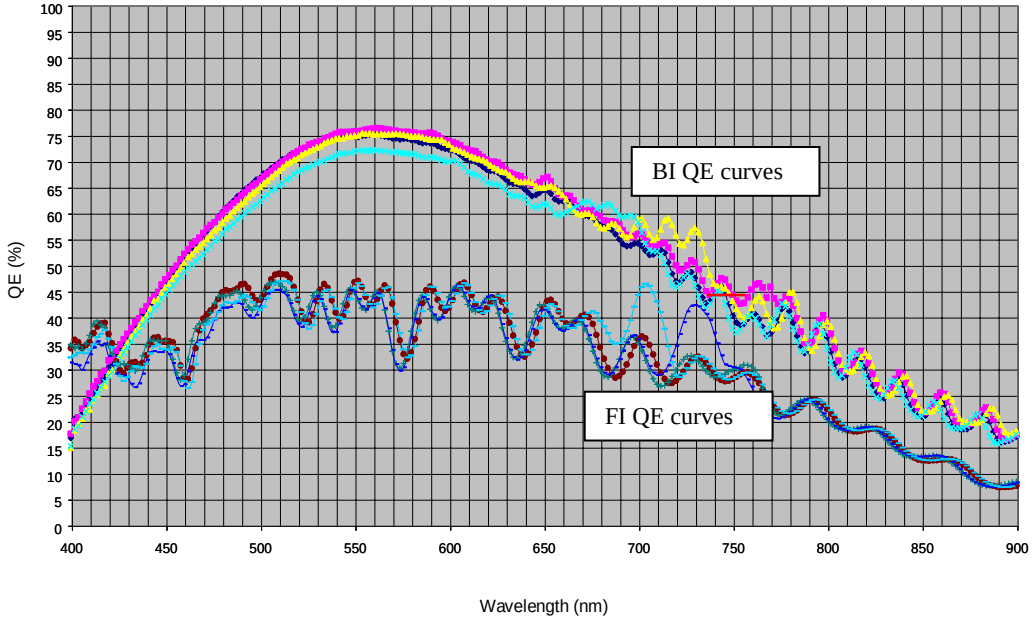

Fig 11 – 2Mpixel CMOS sensor, comparison of front-face and back-face responses

The second wafer was processed with a shallow boron implant and a variety of laser-anneal processes intended to minimise the melt depth of the silicon and hence to give a higher QE in the UV region. A total of 12 different laser anneal processes were used and QE values for all were measured at 320nm wavelength. Because of the limited number of devices available, each has three different laser-anneal stripes across the active area. A flat field image from one device is shown in figure 12 to illustrate the different responses. The QE at 320 nm obtained from each process is shown in the table 1. As can be seen, most variants gave QE values higher than 25% with the best at around 31%, implying that the precise process conditions are not too critical. Note that these results compare very favourably with what we would have expected from a UV-optimised process with the p+ layer thinned down after the laser–anneal. This wafer had an AR coating that was not optimised for UV performance, with an optimised coating the QE would have been above 60%.

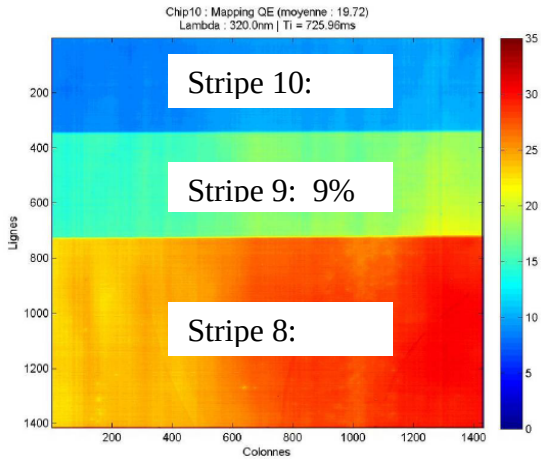

Fig 12 – Flat Field image from a device with three different laser processes

Variant	Measured QE at 320nm
1	25 %
2	28 %
3	28 %
4	32 %
5	32 %
6	31 %
7	26 %
8	16 %
9	9 %
10	7 %
11	15 %
12	21 %
Low temp anneal	4 %
FF sensor	13 %

Table 1 – QE at 320nm from different laser anneal process variants

5 CONCLUSIONS

Practical results have shown that the thinning and back-illumination techniques now standard for CCD sensors can also be successfully applied to CMOS from four different foundries with a range of pad architectures. The response towards the infra-red has not been particularly high because the silicon is relatively thin. Future devices will attempt to use thicker material, which will need to be lightly doped to achieve full depletion with the low voltages typical of CMOS technology, or possibly Hi-rho techniques may be used. Without such optimisation the spatial resolution is likely to be poor.

ACKNOWLEDGEMENTS

Our thanks to Michel Breart De Boisanger and Olivier Saint-Pe from Astrium for their inputs to the work on the 2MPixel CMOS device and particularly for the test results.

REFERENCES

[1] Dash W C and Newman R, “Intrinsic optical absorption in single crystal germanium and silicon at 77K and 300K”, Physics Review, Vol 99, No 4, August (1955)

[2] Kishino S, Nagasawa K & Iizuka T, “Heavy metal gettering by an intrinsic gettering technique using microdefects in Czochralski-grown silicon wafers” Japanese Journal of Applied Physics, Vol 19, No 8, (1980)

[3] Short A, Keay A & Turner M, “Performance of the XMM EPIC MOS CCD detectors” Proc. SPIE, Vol. 3445, No 13, (1998)

[4] Stern R,Shing L & Blouke M, “Quantum efficiency measurements and modeling of ion-implanted, laser-annealed charge-coupled devices: x-ray, extreme-ultraviolet, ultraviolet, and optical data” Applied Optics, vol 33, No. 13, pp 2521-2533, (1994)

[5] Hoenk M et al, “Growth of a delta-doped silicon layer by molecular beam epitaxy on a charge coupled device for reflection-limited ultraviolet quantum efficiency” Applied Physics Letters, Vol 61, pp 1084-1086, (1992)

[6] Walker A et al, “The Gaia challenge: Testing high performance CCDs in large quantities” Proc. SPIE, Vol 7106-54, (2008)