

***Optical
Radiation Detectors
for Astronomy***

Astronomical Optical Radiation Detectors

Detectors in astronomy have a very long and fraught history. Through the late part of the 19th and first two-thirds of the 20th, the effective QE of photographic plates improved by a factor of order 100, (culminating in Kodak IIIa-J, at ~0.5%). Plates were never satisfactory for photometry except in the very most careful hands--which existed, but were not very numerous. They are nonlinear, nonuniform, and cannot be calibrated after the fact.

After the Second World War, the photomultipliers developed during the war, the 931A, the 1P21, and later the red-sensitive RCA7102, came to be used for astronomy. They were single-channel devices, suitable for aperture photometry and, again in capable hands, spectrophotometry. They had quantum efficiencies approaching 20 percent in the blue and a few percent in the red.

In the 1960s, the first photoelectric image tubes became available, both as a result of military research and research sponsored by astronomy, particularly through the Carnegie institution of Washington. This led to magnetically-focussed devices with good photocathodes and ~20 micron resolution. They were light amplifiers, and the recording medium was again photographic plates.

Electronic detectors in the form of low-light-level vidicon camera tubes, all employing the charge multiplication of energetic (typically a few keV) electrons bombarding some kind of 'sticky' target, were developed in the 1960s and early 1970s. (SEC, SIT) Analog-to-digital converters were used from the very beginning with these devices, and the resulting pixellated images stored on magnetic tape and/or disk. Various schemes were also implemented for 1 and 2-d photon counting using electron multipliers of very high gain; some of these are still in use today in the UV. (GALEX)

The first CCDs used in astronomy came in the middle 1970s; devices from RCA, Fairchild, and TI were used, but by far the best devices were the TI ones developed with NASA support for Voyager and later the mission which became Hubble. These CCDs were developed for low light-level scientific applications and were the first high-sensitivity, low-noise, very high performance CCD devices available. The arrays were 500 x 500 Pixels and later 800 x 800 pixels, 15 microns square. They were thinned (see later) and illuminated from the back. QEs were in the neighborhood of 60%

Solid-state Photodetector Imagers

CCDs, CMOS imagers, IR imagers using InSb and HgCdTe photodiode arrays, Reticons—essentially all of the current zoo of solid-state detectors, use sophisticated semiconductor technology, and could not have been developed much before they were even if someone had thought of them; the technical problems would have been much too severe.

Before we talk about how they work, let us list their advantages over essentially all of their predecessors.

- 1. Very high quantum efficiency, often limited entirely by the reflection losses at their surfaces (they all are made of materials which have very high refractive indices, for good physical reasons) CCDs in the visible and red can have QE as high as 90%, 200 times better than the best plates, and five times better than the best photocathodes of the 1970s.***

2. *Very low noise. CCDs have on-chip amplifiers with very low capacitance, achieving sensitivities of the order of one or even a few microvolts per electron, and the readout noise using proper circuitry is as small as a few electrons. (2, currently)*

*NB!! The noise in a signal of N electrons is \sqrt{N} , so if the read noise is n electrons, detected photon statistics will dominate for all signals greater than n^2 electrons. If n is 3, which is achievable, then read noise is negligible for all observations for which the $S/N > 3$, $n > 10$ or so for measurements **INCLUDING** the background.*

3. *Very large dynamic range. For a device with a read noise of 4 electrons and a capacity of 100,000 electrons, which is not unusual, the dynamic range is 25,000. The useful dynamic range of a photographic plate is less than 10, and for photoelectric imagers at most a few hundred, except for true zero-background photon counters, for which the concept is not well defined, though there are always limitations at the high-count rate end set by circuit speed.*

4. *Very good linearity, often better than a few percent over the full very large dynamic range of the device. The nonlinearity which exists is stable and can be calibrated relatively easily.*
5. *Good devices from the past two decades or so are quite uniform in response over their surfaces, and again the nonuniformities are stable. The big problem with calibrating the devices over their surface for imaging is not the device itself but scattered light in the telescope.*
6. *Well-defined and stable intrapixel response, facilitating PSF photometry with barely adequately sampled images and accurate astrometry.*
7. *Large sizes. The largest commercial scientific devices currently are ~4000 15-micron pixels square, a single integrated circuit 60mm on a side. A 9000-pixel square device with pixels too small to be of direct interest in astronomy for large telescopes (8 microns) is commercially available.*

How do they work?

Essentially all of the current crop of detectors, including all of the laundry list a few slides ago, are MOS devices with the active circuitry (though not, in some cases, the photosensitive parts) made of silicon.

The property of silicon which allows the devices in principle to be made is simply that it is a semiconductor; in practice the availability of these devices is a direct consequence of the enormous commercial development of silicon semiconductor technology for innumerable other uses. The choice of silicon over other semiconductors is motivated by several factors, not the least of which is that silicon can be oxidized to form a robust dielectric layer, allowing devices to be built of several vertically differentiated and electrically isolated layers. MOS devices make use of M(etal) gates isolated by O(xide) from the S(emiconductor) substrate. More anon.

What is a semiconductor, and what semiconductor properties do imagers use?

Semiconductors are loosely defined as solids

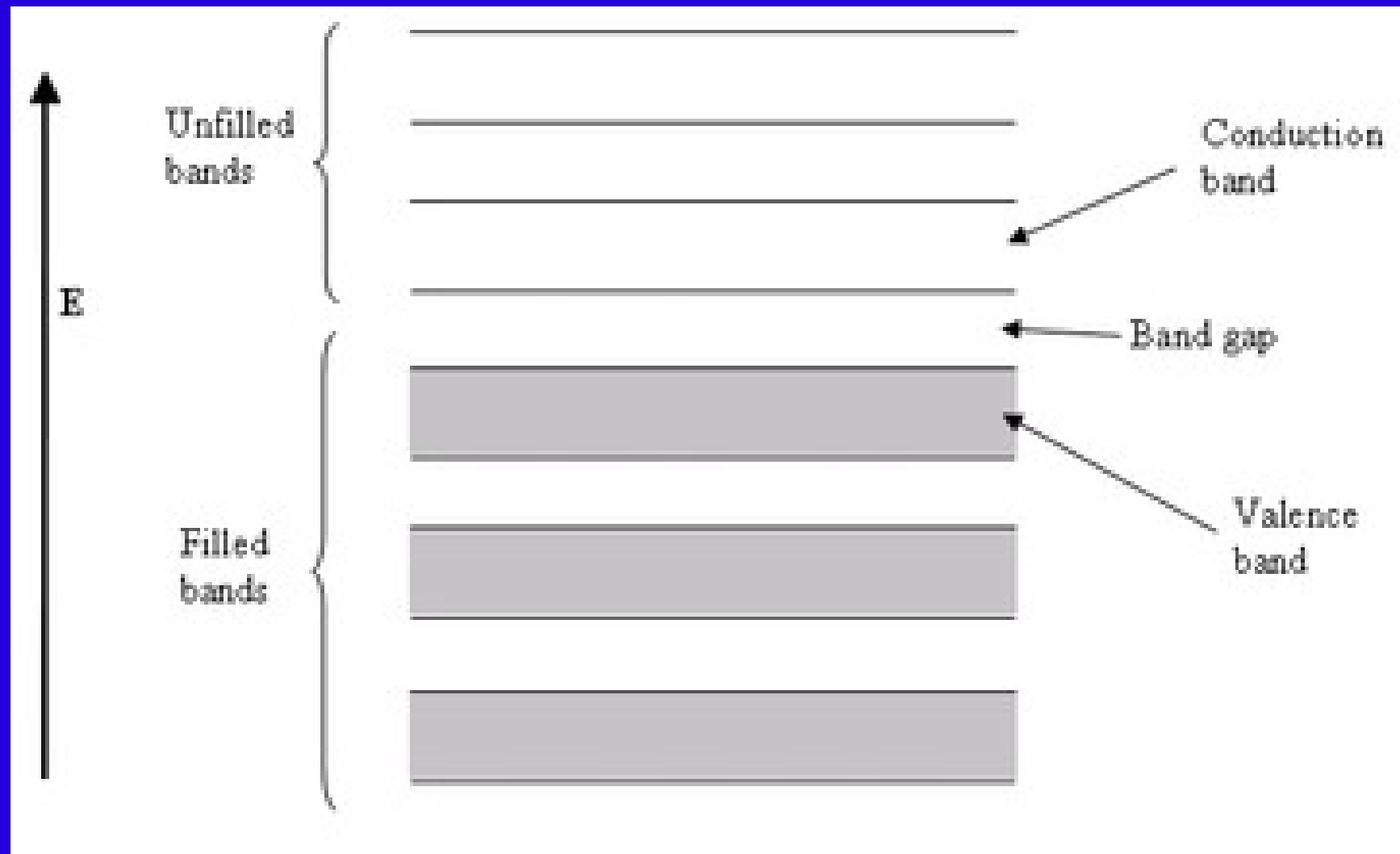
a) which have well-defined crystalline structures

b) in which the energy bands containing the valence electrons of the constituent atoms are filled. (Metals have partially filled bands, allowing electrons to carry momentum and current with very little energy input).

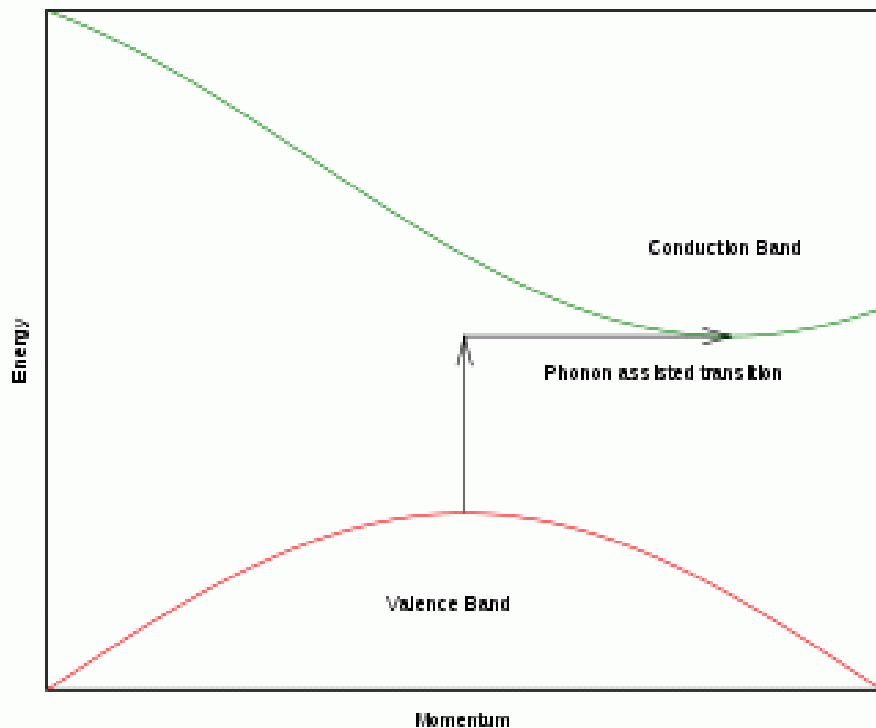
c). in which there exist continuum bands at most about three volts above the top of the filled valence band. (This energy difference is called the BAND GAP) The distinction between dielectrics (insulators) and semiconductors is fuzzy, but generally insulators have larger bandgaps.

The bandgap in silicon is about 1.12eV at room temperature, corresponding to a photon at about 11000 Å; this sets a firm upper limit on the wavelength of light which can be detected with silicon photodetectors.

The electron energy band structure in a semiconductor looks like this:

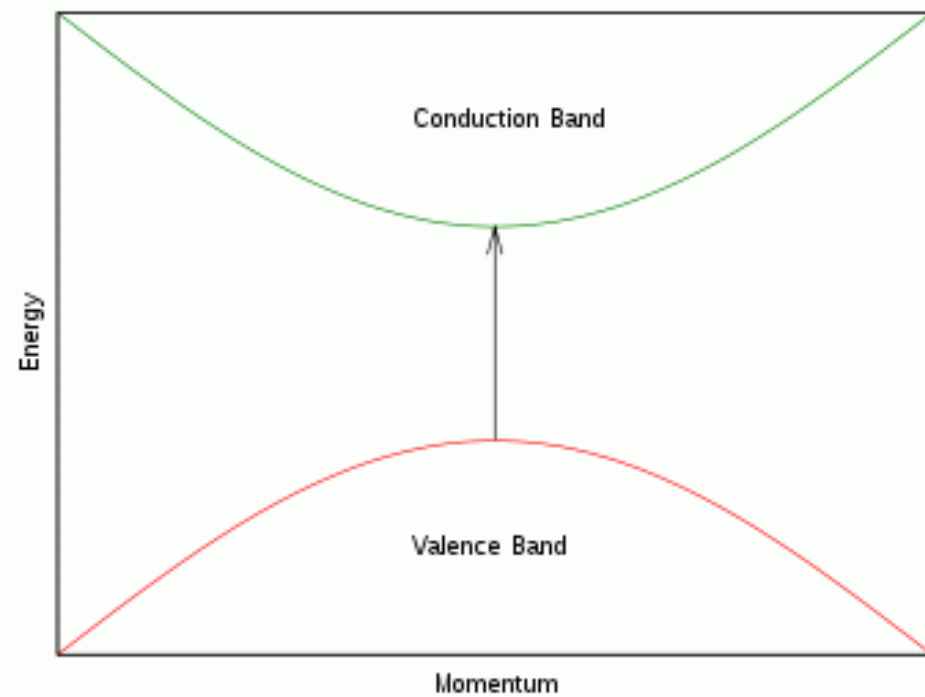


Indirect Band Gap



***Indirect band-gap semiconductor (eg Si, Ge)
phonon-assisted emission or absorption
small opacity or emissivity just shortward
of bandgap energy. Si CCDs need to be
thick to have high QE in the 1-micron region***

Direct Band Gap



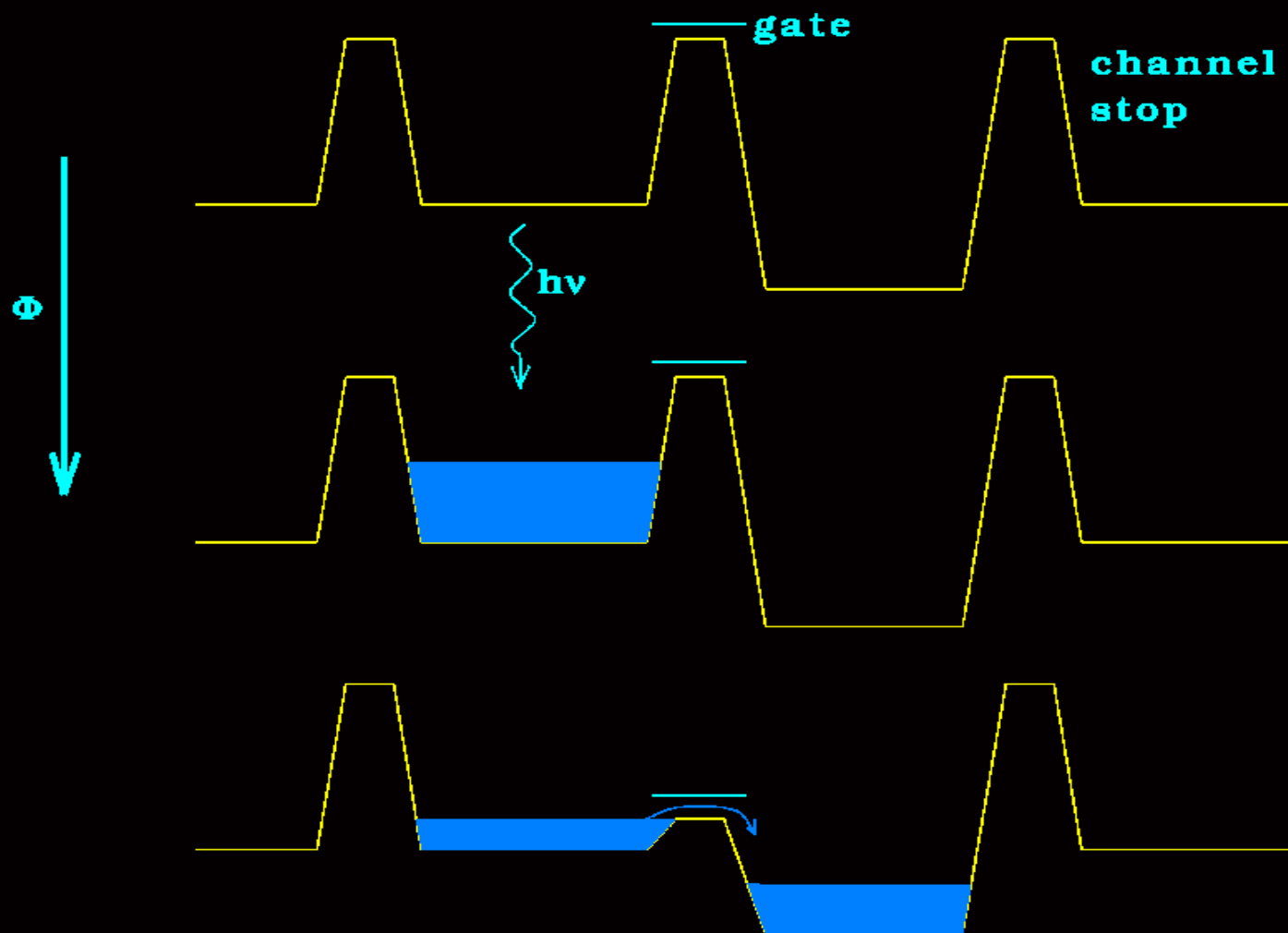
***Direct band-gap semiconductor (eg GaAs,
Most IR photomaterials) No phonon assist
necessary, very high opacity immediately
shortward of bandgap energy.***

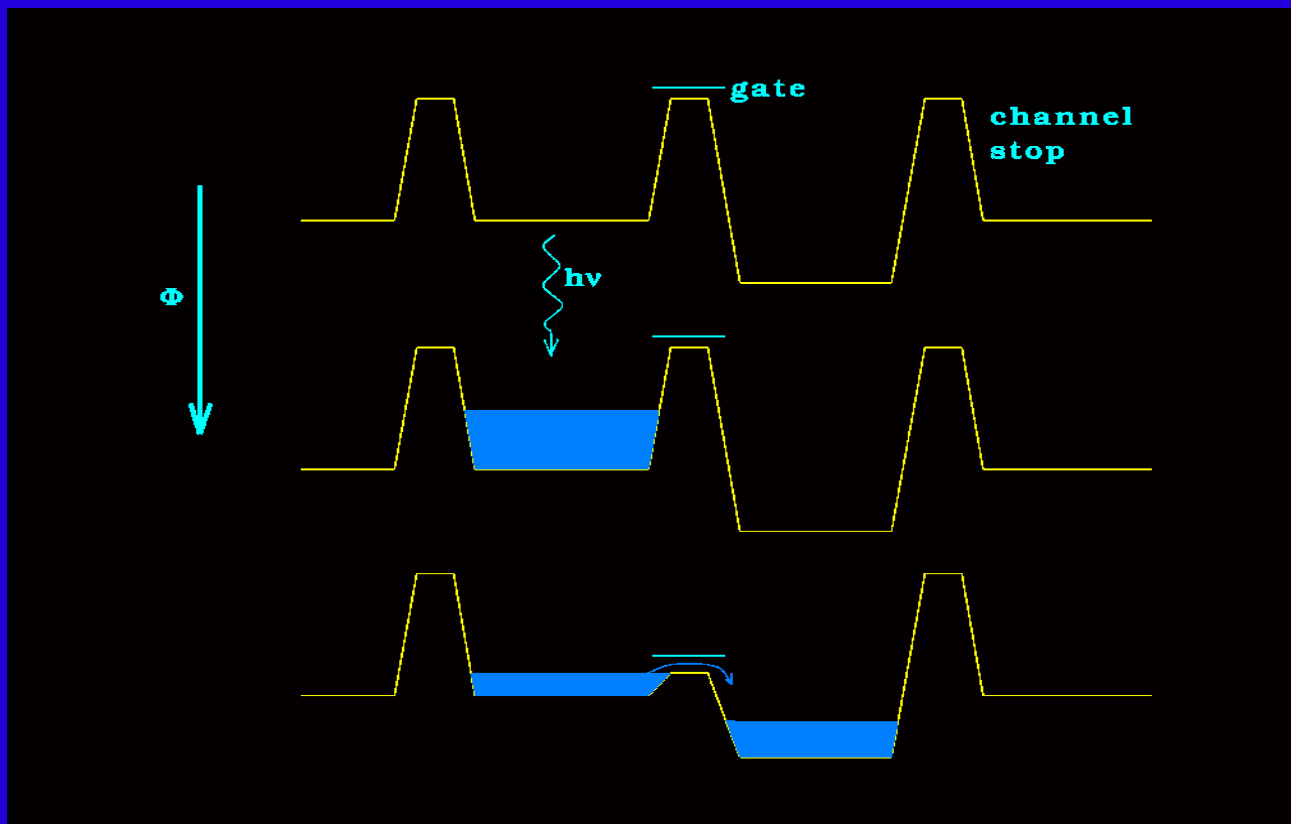
Semiconductors in general and silicon in particular can be DOPED. Silicon has 4 valence electrons, but one can insert very small quantities (10^{-8} to 10^{-3} by number) of impurity atoms which are happy to reside in the place of a silicon atom in the crystal lattice, impurity atoms which have more (eg phosphorous, with 5) or fewer (eg boron with 3) valence electrons than silicon. These are called, respectively, DONORS and RECEPTORS, and can supply electrons (or holes, which are quantum-mechanically essentially identical to electrons except for the sign of the charge. Silicon so treated is called n-type (phosphorus) or p-type (boron) material. Most CCDs are built on a thin (5-30 microns thick) n-type 'epitaxial' layer grown on a p-type substrate. By selectively doping the thin n-type layer with ion beam deposition the potential of the bands in the silicon can be BENT substantially, by several volts. In addition, metal electrodes can be deposited, either in ohmic contact with the silicon or insulated from it by a thin oxide film (hence MOS). The latter structure is called a GATE, and the application of a voltage to a gate also bends the bands underneath the gate basically to follow the potential of the gate.

The potential gradient perpendicular to the surface can be manipulated as well as the layer is grown, and can incorporate an electric field which attracts electrons toward the surface. Thus a photon above the bandgap energy which interacts with the silicon can knock a valence electron into the conduction band, and physically the electron will be pulled toward the surface of the epitaxial layer, where it can be moved about by manipulating the voltages on gates built on the surface.

You can get structures like this:

Note that I will be plotting potentials upside down, so electrons go DOWNHILL; it is visually more pleasing.



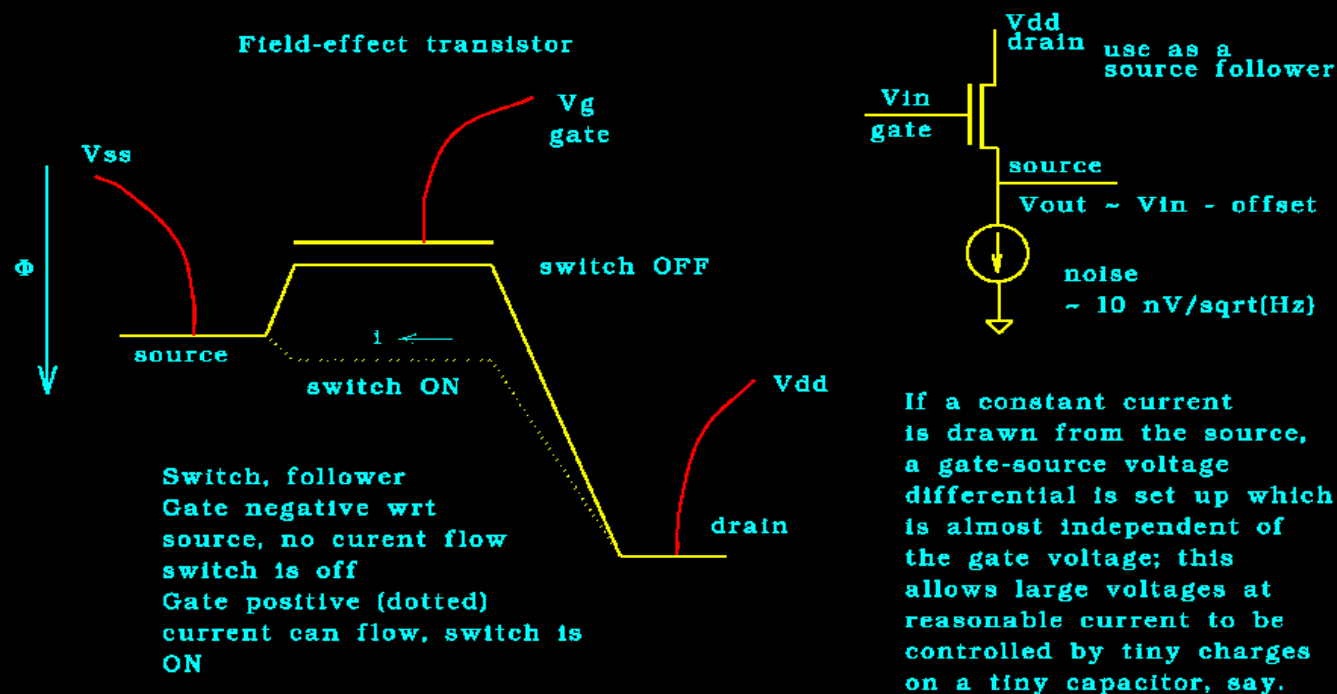


In this picture I have shown a section of a device with three implanted negative barriers (channel stops) and a positive depression which forms a well. There is a metal gate over one of the channel stops.

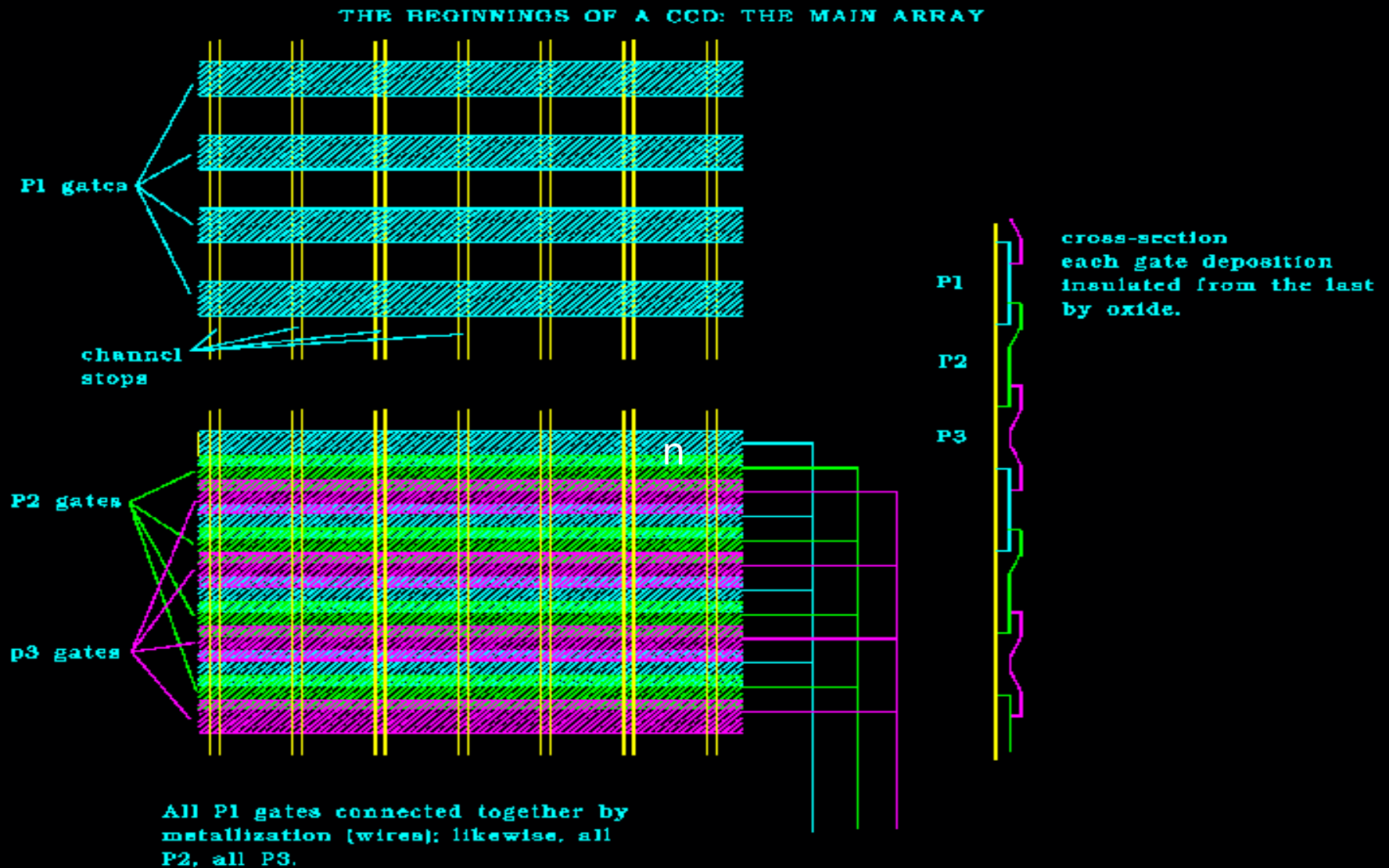
A source of electrons (a part of an illuminated

image, perhaps), supplies some electrons which are trapped between the left and center channel stop. As this well gathers more electrons, the potential associated with the electrons becomes more negative (just Q/C), where C is the capacitance associated with the site. At some point, the voltage on the gate is made more positive than this, and electrons flow into the deeper well until their potential is equal to that of the barrier. This is not a sensible device, but illustrates how one can use implants and gate voltages to control charges collected in a MOS device.

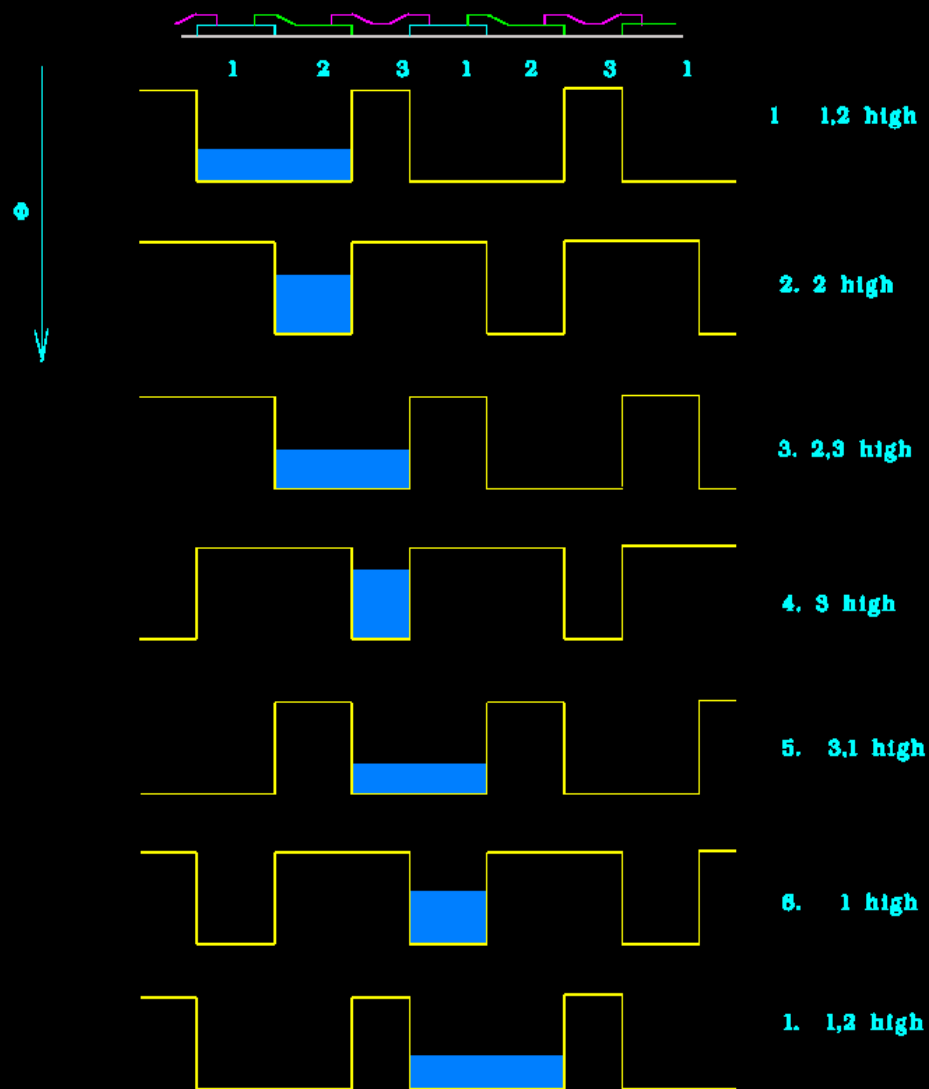
A MOSFET transistor looks like this: If the source electrode is grounded, voltage on the gate can be manipulated to make it act like a switch; if the drain is connected to a positive supply and the source to a resistor or current source, it makes a follower, in which voltages which can drive macroscopic currents can be controlled by voltages generated by only a few electrons on the gate. These are used as output devices in CCDs.



Let us start building a CCD. We implant some strongly negative channel stops, to form the CCD columns. These keep photo-electrons from moving sideways. Then, for a 3-phase CCD, deposit some conducting gates, like this:



These gates allow us to MOVE charge along the CCD columns by manipulating the voltages on the three gate structures, like this:



The PIXELS of the CCD are defined horizontally by the channel stops, but only temporarily vertically by the gate voltages. The configuration at the top is the integration mode; we collect photoelectrons under gates 1 and 2, and the pixel is this collection area and half of the gate 3 extent on either side.

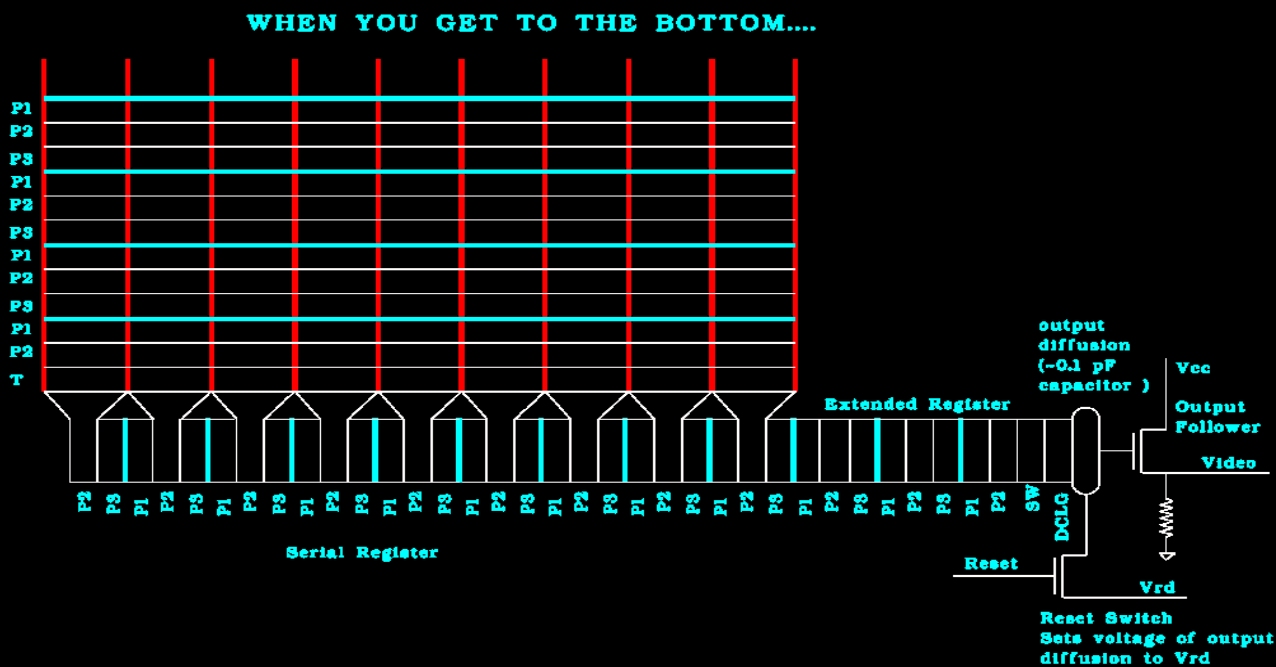
**BEWARE OF
IRREGULARITIES WHICH
CAN TRAP AND LEAVE
CHARGE BEHIND !!!!**

Charge-Transfer Efficiency

Well-designed CCDs made of good material achieve levels of transfer of charge in a single pixel (three-phase) transfer of CTE = .99998 or better; at this level it is clearly better to use the charge transfer INEFFICIENCY $CTI = 1 - CTE \leq 2 \times 10^{-5}$. This number means that a fraction of the charge CTI is left behind in a transfer. After N transfers, a charge packet of size Q has a tail which is roughly $Q \cdot \exp(-N \cdot CTI \cdot n)$, where n is the index of the pixel following the packet (n=0 is the packet). The CTI is unfortunately nearly always a function of the size of the packet; defects tend to trap amounts of charge somewhere between a fixed QUANTITY of charge and a fixed FRACTION of the charge—for many devices a reasonable approximation is that the trapped charge is roughly the square root of the size of the charge packet, so small packets (faint objects?) suffer more, fractionally, than big ones. In broadband photometry, the sky usually keeps things OK, but in high-dispersion spectrographs, for example, where there is essentially no background, this can be a serious problem.

The complete logical structure of a small CCD is shown here, with the CCD register at the bottom of the array (the serial register) which is used to carry the charge to the amplifier node. The output circuitry is shown schematically. The output diffusion is a well isolated from the substrate. Charge from a pixel is dumped onto it each complete pixel transfer of the serial register. The gate of the output FET follower sees the voltage on this capacitor, and its source is the output of the CCD. After the voltage is recorded, the voltage of the output node is

RESET with the reset switch transistor. Thereby hangs a tale.



Readout noise in CCD-like detectors

Resistors at finite temperature develop JOHNSON NOISE, the expression for which is

$\langle V^2 \rangle = 4kTRB$; B the bandwidth in frequency units of the measurement.

If you imagine charging a capacitor (like the output node of a CCD) to an infinitely stable voltage source through some resistor for a few RC time constants and then opening the switch, the Johnson noise of the resistor will surely appear on the capacitor in some way. The bandwidth of the connection is of order $1/(RC)$, so the noise is

$\langle V^2 \rangle \sim kT/C$, INDEPENDENT of the resistor value. All switches have some resistance, so this is the intrinsic thermal noise associated with resetting a capacitor. (the coefficient is, in fact, unity).

How bad is it? The output node capacitance in modern detectors is of order 0.1 pf (1.6mV/e). kT is $\sim 4e-21$ J at room temperature, so $(kT/C)^{1/2}$ is about 200 microvolts—about 160 electrons.

The designers work very hard to build very low-noise FETs for the output, and achieve noise figures like 10 nV/sqrt(Hz) and whiteish noise above ~ 10 kHz.

If we read the CCD at 5 microseconds per pixel, much of that is taken up with clocking the charge, so we have something like a microsecond to measure the charge on the output well, and the frequency is about one MHz. The noise from the output transistor is then about 10 microvolts, about 6 electrons. But we are completely killed by the thermal noise.

Enter DOUBLE CORRELATED SAMPLING:

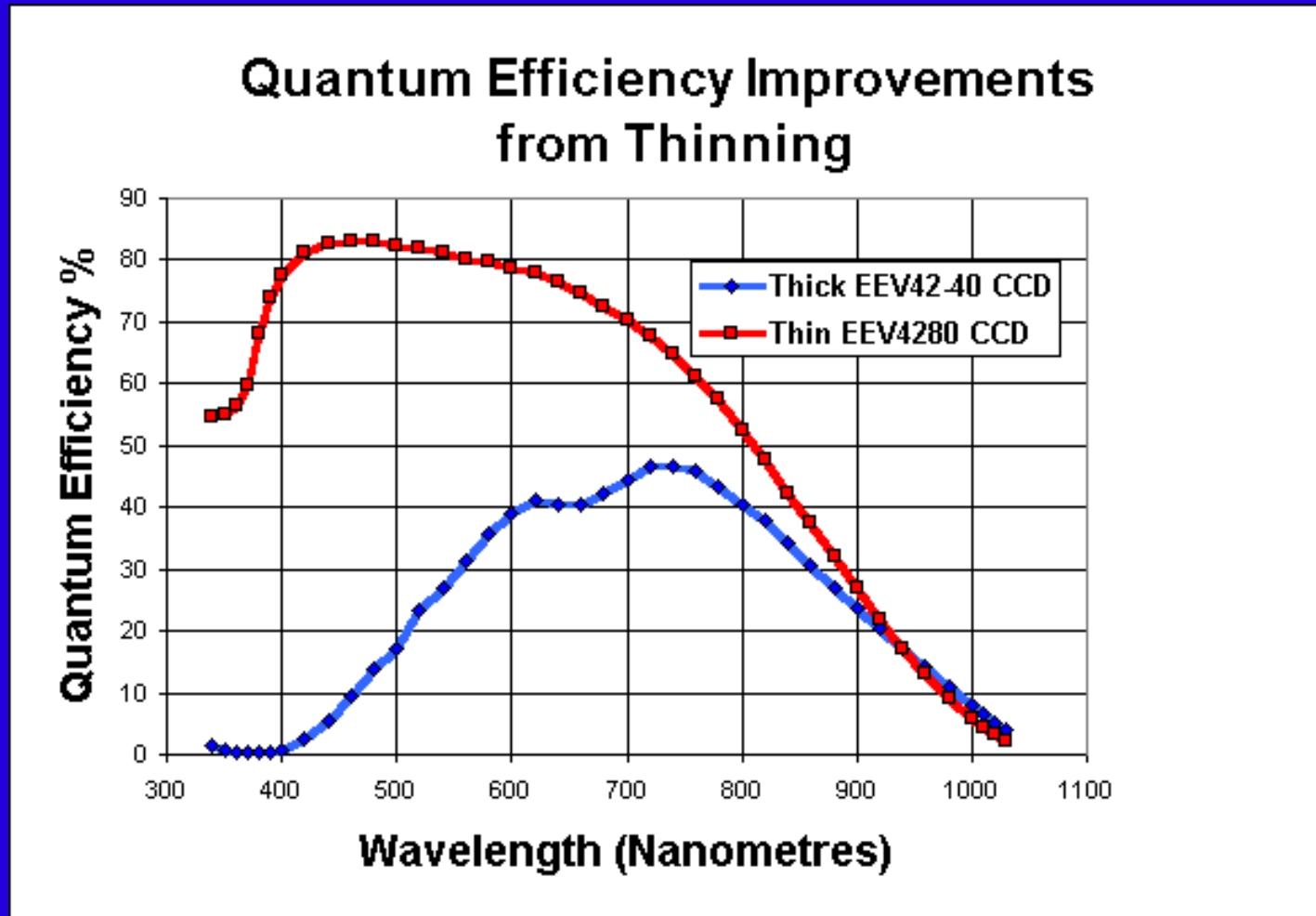
Simple. Reset the output node, THEN measure the output voltage for a microsecond, THEN clock the next pixel's charge onto the output node, and subtract. 1.4 from the differencing, but much better than $(kT/C)^{1/2}$.

WARTS:

CCDs are almost perfect, but not quite:

- 1. In order to take full advantage of the potential sensitivity, they must be THINNED to the thickness of the n-type layer, leaving a device only a few tens of microns thick, which is illuminated from the BACK; the gates even if made of polysilicon instead of metal are pretty opaque, especially in the blue. The thinning is generally a low-yield process, and mounting the thinned device in a manner which keeps it flat and reasonably robust is a nightmare.*
- 2. They must be run under vacuum cryogenic conditions, at $T \sim -40 \rightarrow -100\text{C}$ to keep thermally-generated carriers from swamping the signal. The thermal dark noise goes like $T^{3/2}\exp(-DE/kT)$, where DE is the bandgap; this is a doubling about every 7 degrees at -80C , at which temperature a good detector will have $\sim 0.01\text{e/s}$ dark current or less.*

Front vs. Backside CCD QE



3. They require complex circuitry to run, and many accurately regulated voltages, including some which are fairly complicated dynamic loads (the gates).

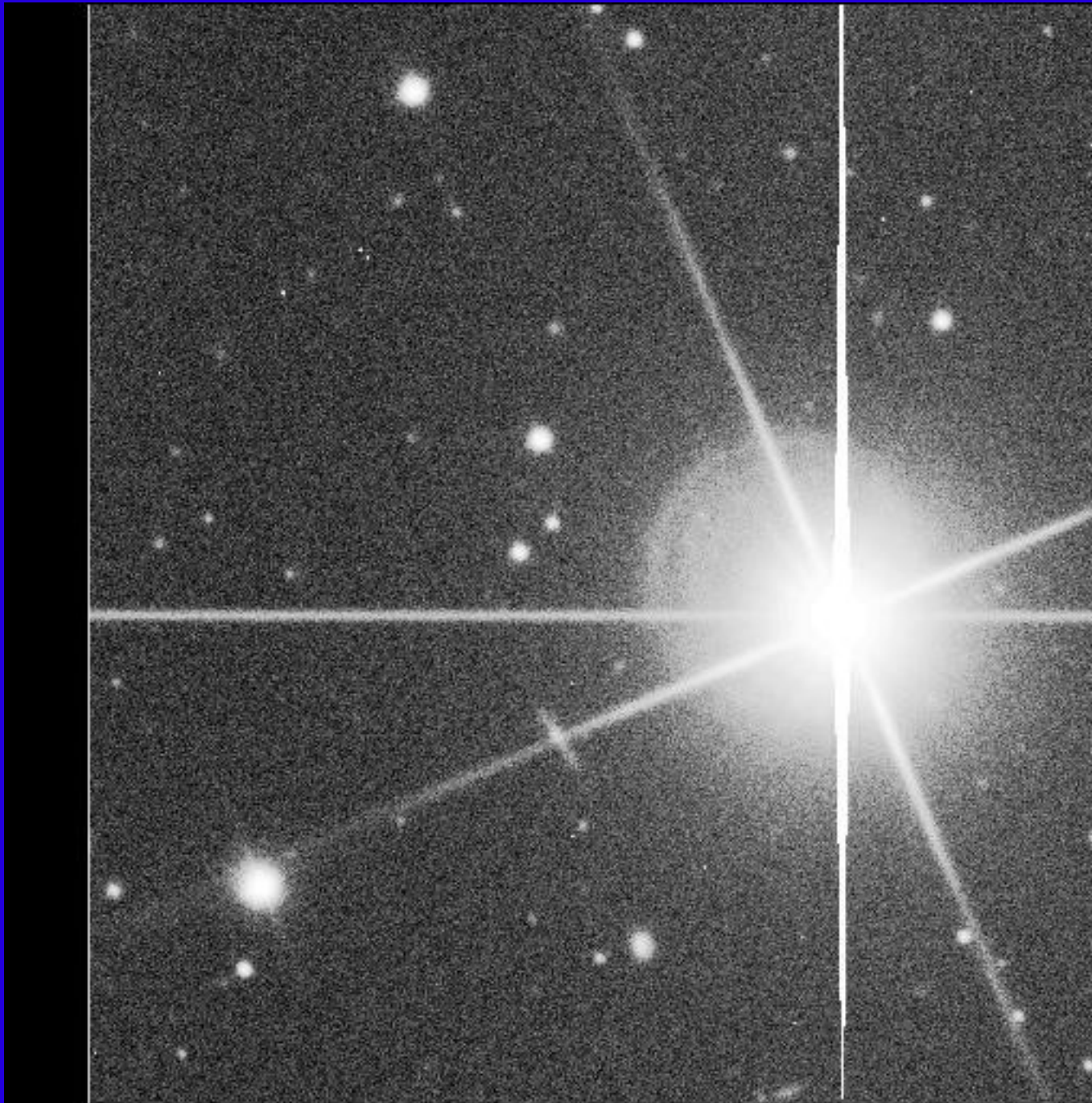
We live with these. but they also have some failings as detectors, which are more serious.

4. Silicon will detect only out to its bandgap energy, which corresponds to about 11000 Å; As the push to higher redshifts in galaxy research goes on, this is more and more serious.

5. Connected with the same phenomenon, the detectors become more and more transparent as one approaches the bandgap energy from above. Because the index of refraction is so high (associated with the small bandgap) internal reflections between the front and back of the device become serious, causing Fabry-Perot-like FRINGING patterns in the response. The surfaces are far from optical quality, and the fringes are very irregular and difficult to calibrate out.

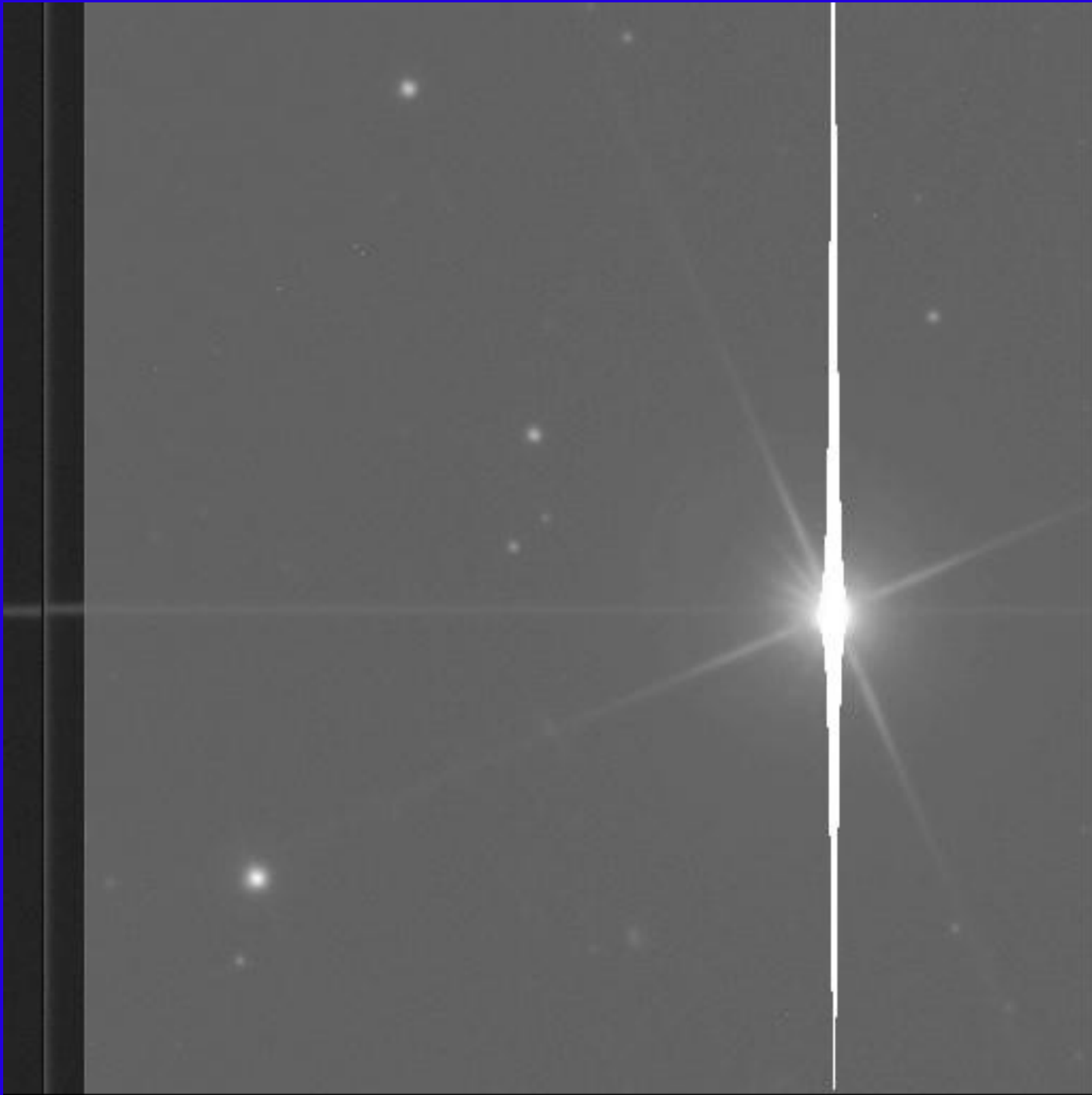
- 6. CCDs do not saturate very gracefully; bright stars cause charge to run up and down the columns, and VERY bright objects even across the channel stops into adjacent columns.*
- 7. There is no way to electronically shutter a CCD; you must have a mechanical shutter.*
- 8. Large silicon wafers are almost never perfect, and defects in the silicon cause isolated hot pixels (dark current defects) and traps (CTE defects)*
- 9. Long-term damage (CTE degradation, dark current degradation) accompany prolonged exposure to energetic charged-particle fluxes, making the use of CCDs in space always a touchy proposition. But they work, clearly.*
- 10. For all these reasons, scientific-grade CCDs are and will remain very expensive.*

Some Warts:



This is a high-contrast rendition of an SDSS image containing a bright star. In addition to the diffraction spikes from the spider of the telescope, there is a saturation bleed trail (the vertical trail above and below the star) and a horizontal trail which is caused by the fact that the serial register is not covered. The SDSS scans continuously, and when bright objects cross the serial, defects like this are produced.

Same image, much lower contrast to show the saturated vertical track and the very faint horizontal serial trail.



A BETTER CCD?

Work has been going on at LBL to develop a more Red-sensitive CCD for SNAP/JDEM. They are in use for BOSS and will be used for DES. Hamamatsu are building them as well, and they are In use at Subaru/SuprimeCam

They are built in a p-type epitaxial layer on an n-type substrate, so all the voltages are inverted with respect to 'normal' CCDs.

They are VERY thick, 200-300 microns, and employ a transparent electrode on the back side (they are not thinned in the ordinary way, but ARE illuminated from the back) which establishes a large electric field through the bulk of this material. Holes are collected over the whole thickness of the device, and the large field keeps the lateral diffusion much less than one 15-micron pixel.

The CCD has high QE up almost to the bandgap energy, and remains opaque by virtue of its thickness; there is very little fringing.

The p-type device is much more radiation-resistant than normal CCDs (by about an order of magnitude) making the new device more suitable for space missions.

BUT---

The cosmic ray rate for COSMETICS is proportional to the surface area of the rectangular parallelepiped bounding the pixel, and is much larger than for conventional thinner CCDs. With a rate of about 1 proton above any reasonable shieldable energy (~10 MeV) gives probability that a pixel will be hit in a 1000-sec exposure of about 12 percent. Can YOU process data with an eighth of the pixels bad?

Photon absorption length

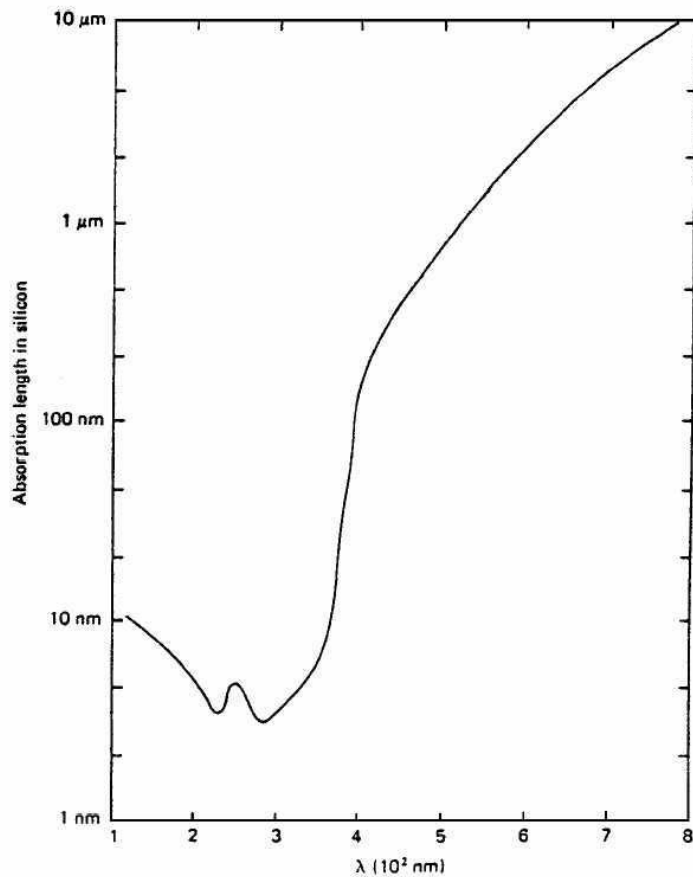


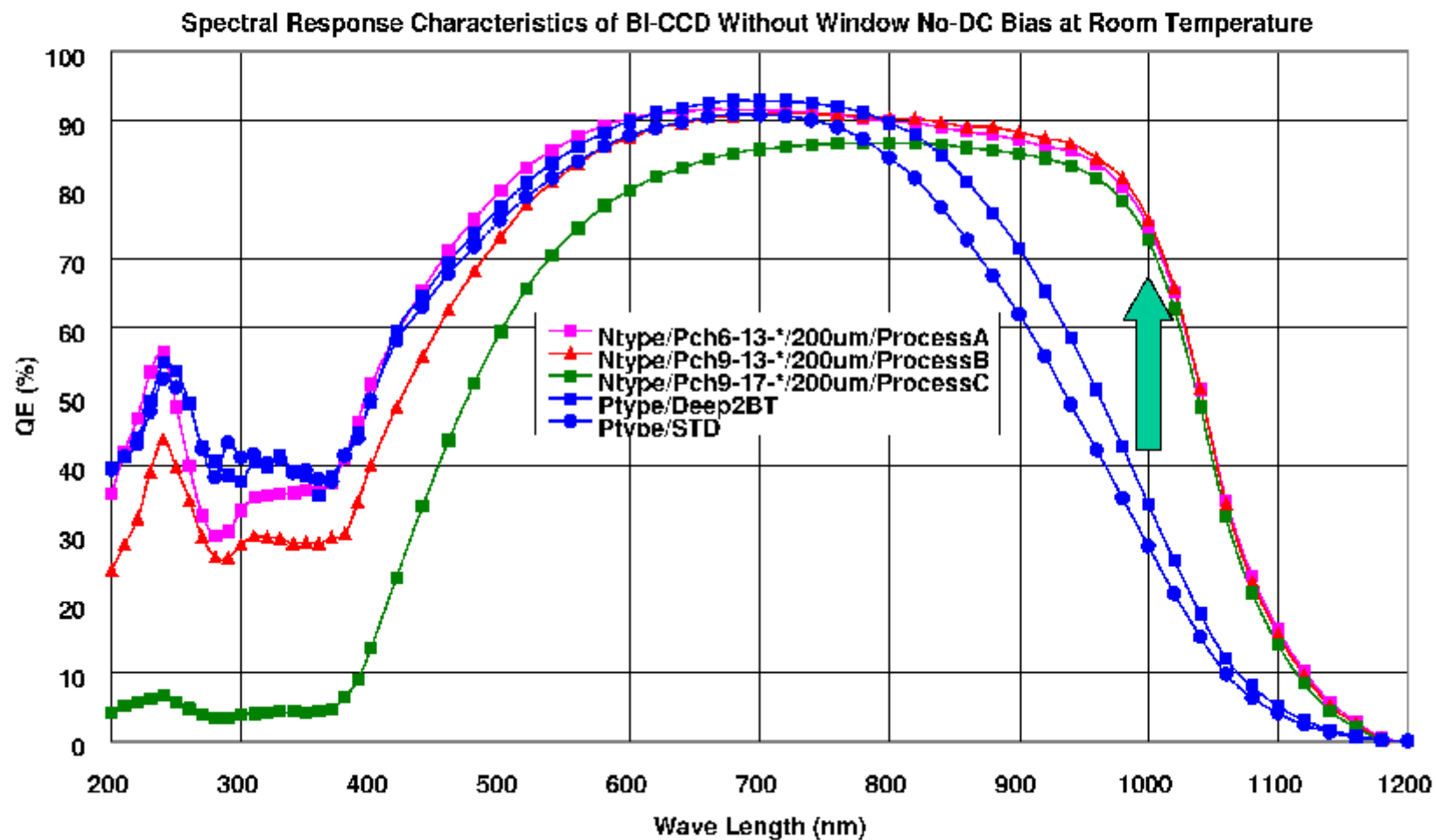
Fig. 3.1. The photon absorption length in silicon is shown as a function of wavelength in nanometers. From Reicke (1994).

Semiconductor	T (K)	$\Delta (E_{\text{Cond}} - E_{\text{Val}})$ (eV)	λ_c (nm)
CdS	295	2.4	500
CdSe	295	1.8	700
GaAs	295	1.35	920
Si	295	1.12	1110
Ge	295	0.67	1850
PbS	295	0.42	2950
InSb	295	0.18	6900

λ_c : *beyond this wavelength
photodiodes become insensitive.*

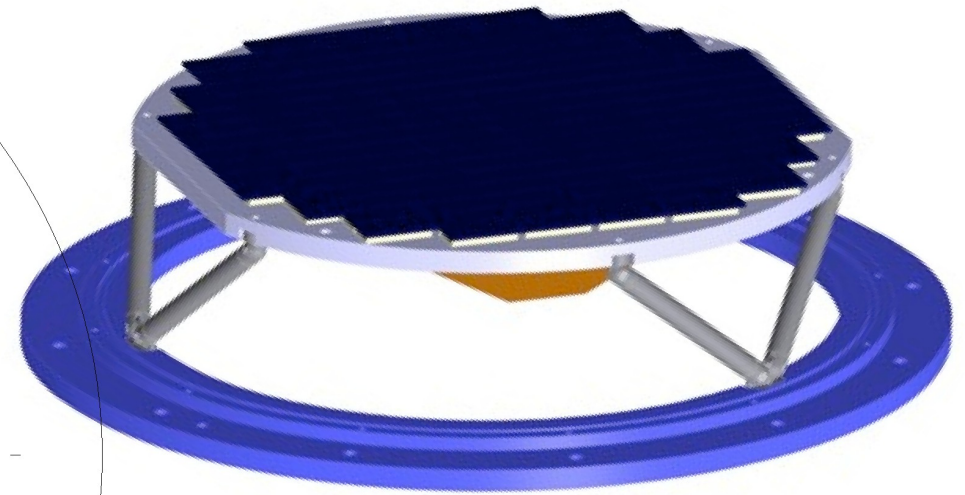
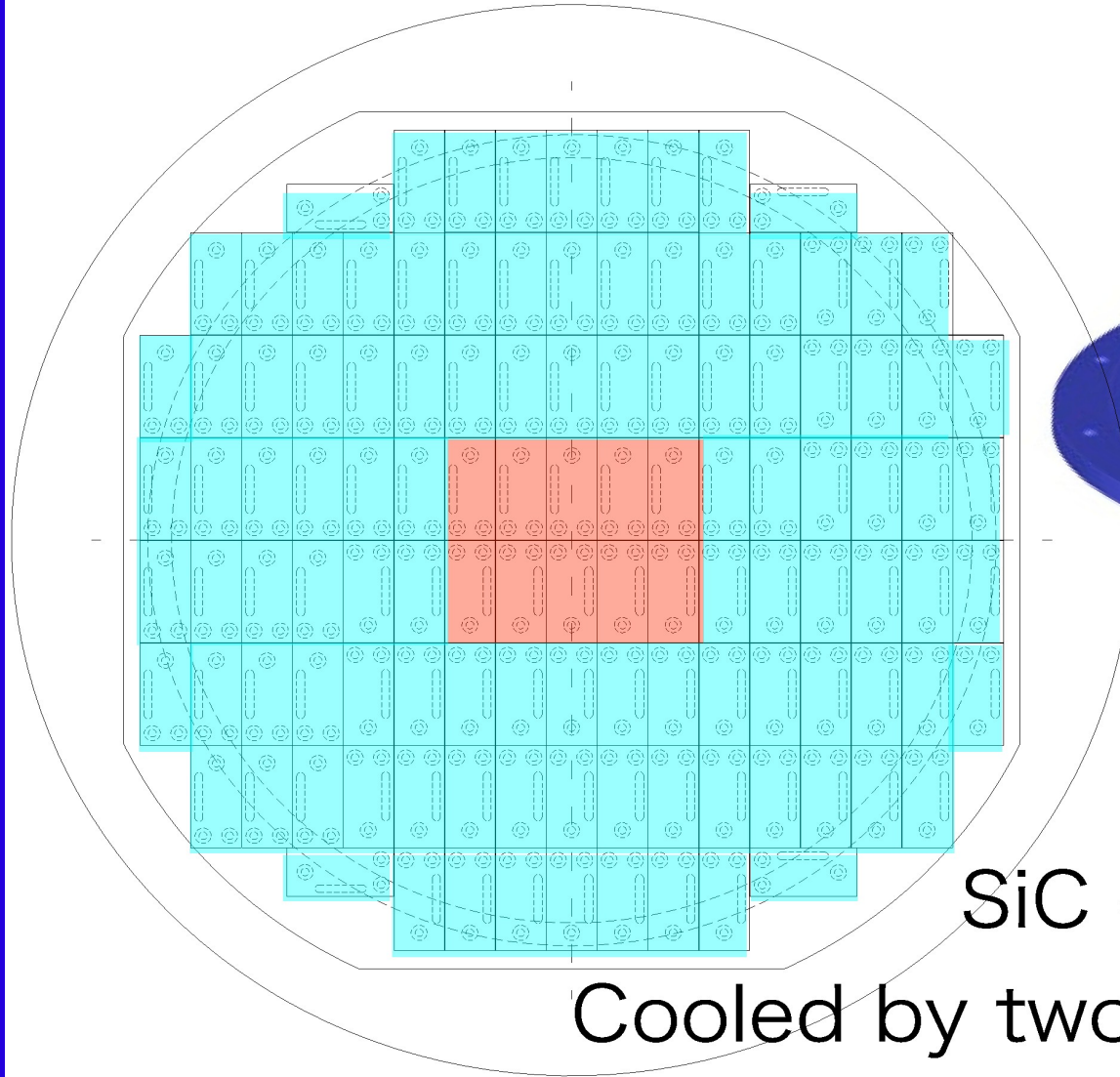
CCDs

- Quantum Efficiency





HSC Focal Plane



112 + 4 Guides

SiC cold plate

Cooled by two pulse tube coolers

45 W@-100 C each

CMOS sensors for the optical

Silicon, very well-understood technology (like your Pentium or CoreX or Opteron or G4). Same general idea as the IR sensors do be discussed next, with one amplifier per pixel and switches onto a bus network. Have not achieved CCD-level noise performance yet, and since the switches are not UNDER the pixel but beside it, the pixels do not fill the device. Promising, but not yet. LSST was pushing them (at least in part because they are electronically shutterable, and nobody wanted to think about building a shutter for the LSST camera.) But they will use CCDs.

The promise is that when they get there, they will be really cheap.

Tomorrow.....

The ideal detector would be noiseless, have 100 percent QE, be available with any desired pixel size, and above all, measure x , y , λ , and t for each photon incident on it, through the visible and into the mid-IR. We are not there yet.

But kT at 4K is $4e-4$ eV, and even a 5 micron photon is 0.2eV, so the physics says that without dilution refrigerators or such, resolution of 500 should be possible in the mid-ir, and 5000 in the optical, if a system for which kT was the energy resolution could be found. Not yet, but there are many clever ideas which have not worked. One will, eventually.

