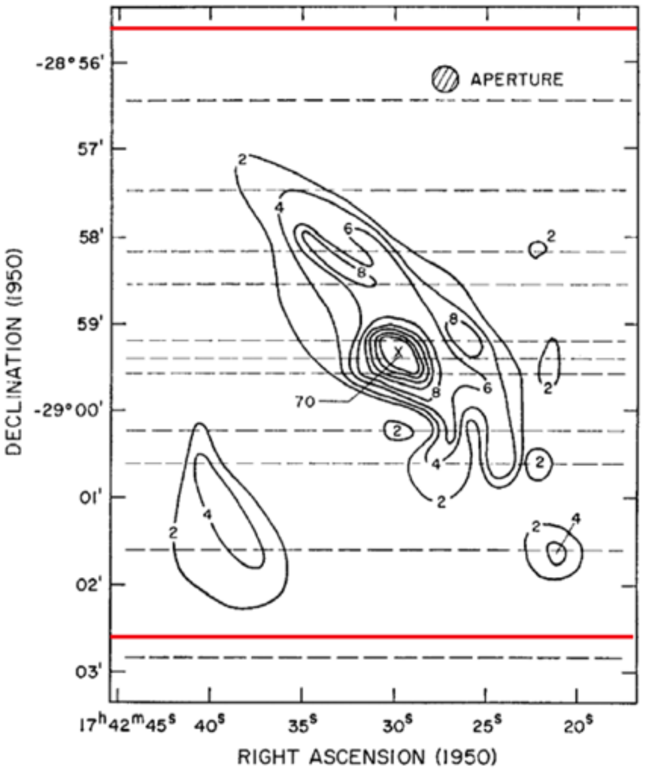
Introduction to the JWST Detector Arrays

George Rieke
The University of Arizona

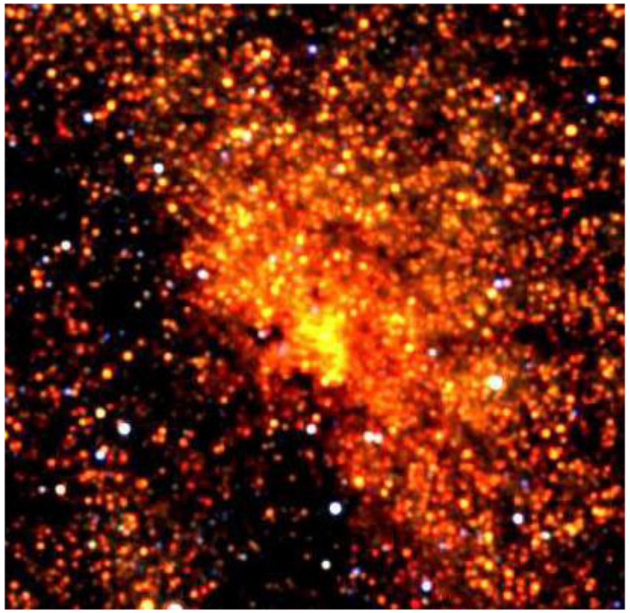
This lecture was funded in part by the Marie Curie Initial Training Network
of the European Commission under contract PITN-GA-2008-214227

The Array Revolution in Infrared Astronomy:
The Most Important Aspect of JWST!

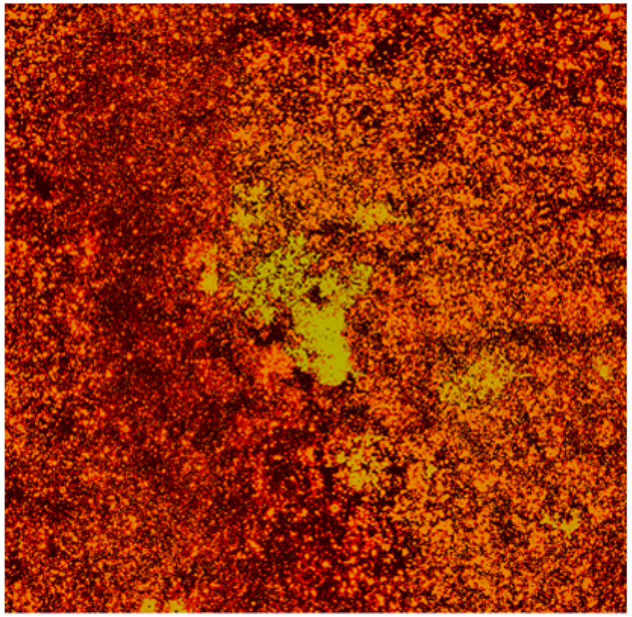
1968, 5-m telescope, 3 nights, single detector



~ 2000, 1.3-m telescope, 8 seconds, 256 X 256



~ 2006, 6.5-m telescope, 1 hour, 1024 X 1024 (4 minutes with the 2x2 2048 X 2048 NIRCam mosaic)

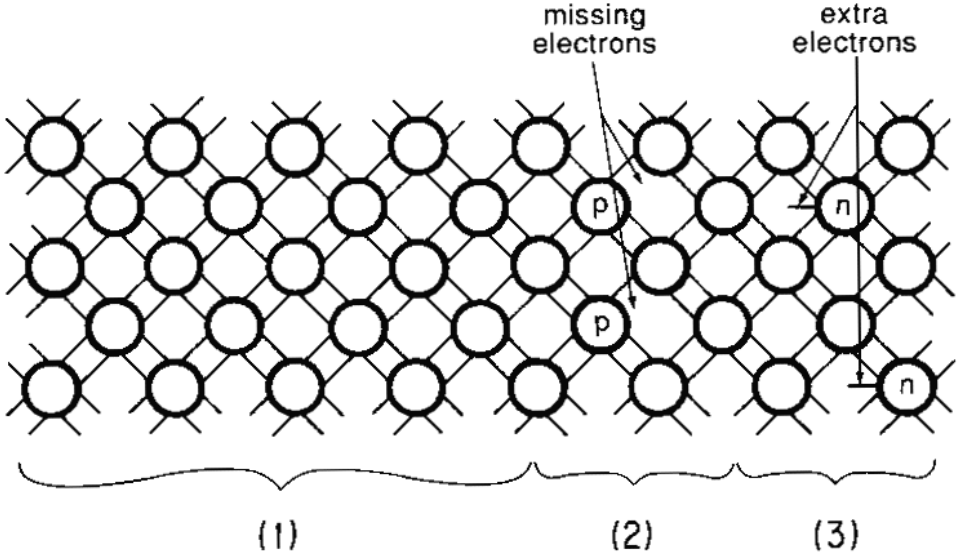


Three Ways to Detect Light

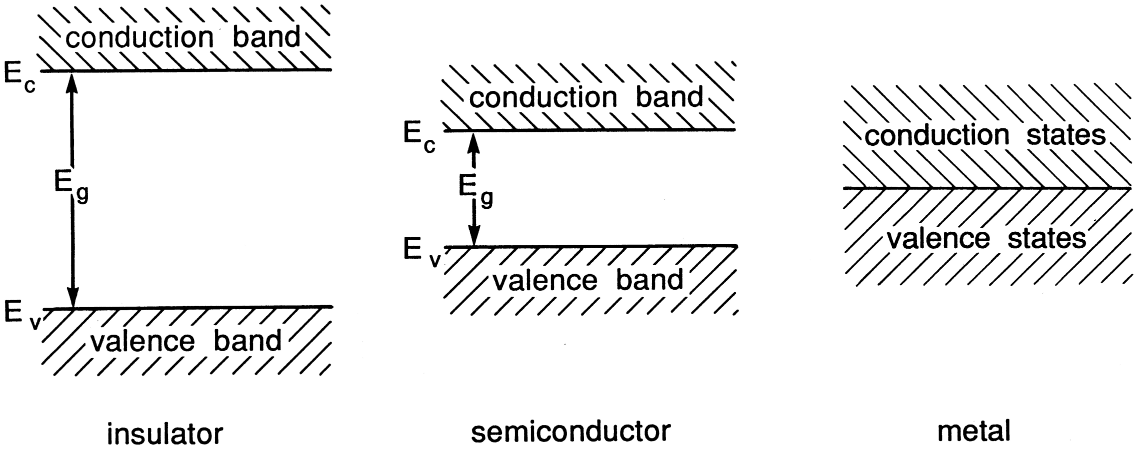
- In **photo-detectors**, the light interacts with the detector material to produce free charge carriers photon-by-photon. The resulting miniscule electrical currents are amplified to yield a usable electronic signal. The detector material is typically some form of semiconductor, in which energies around an electron volt suffice to free charge carriers; this energy threshold can be adjusted from ~ 0.01 eV to 10eV by appropriate choice of the detector material.
- In **thermal detectors**, the light is absorbed in the detector material to produce a minute increase in its temperature. Exquisitely sensitive electronic thermometers react to this change to produce the electronic signal. Thermal detectors are in principle sensitive to photons of any energy, so long as they can be absorbed in a way that the resulting heat can be sensed by their thermometers. Thermal detectors are important for the X-ray and the far-infrared through mm-wave regimes.
- In **coherent detectors**, the electrical field of the photon interacts with a locally generated signal that downconverts its frequency to a range that is compatible with further electronic processing and amplification. Downconversion refers to a multi-step process in which the incoming photon electrical field is mixed with a local electrical field of slightly different frequency. The amplitude of the mixed signal increases and decreases at the difference frequency, termed the intermediate frequency (IF). This signal encodes the frequency of the input photon and its phase.

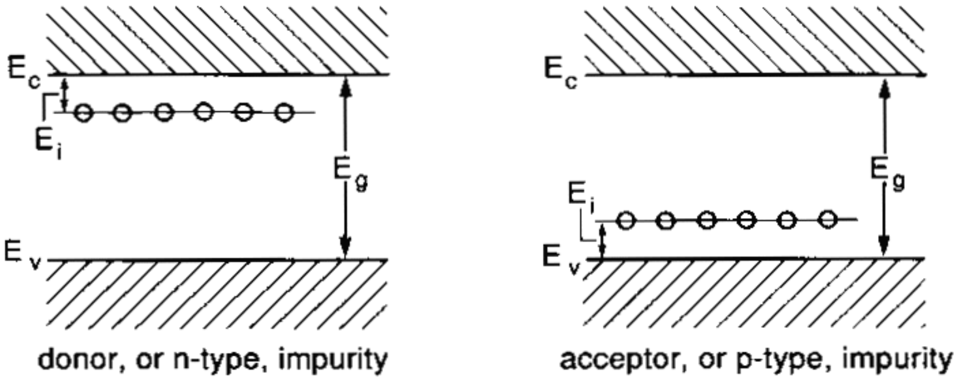
JWST uses photo-detectors (as is generally the case for UV, visible, NIR,& MIR)

Photon absorption and creation of free charge carriers in semiconductors is the basic process behind photo-detectors.



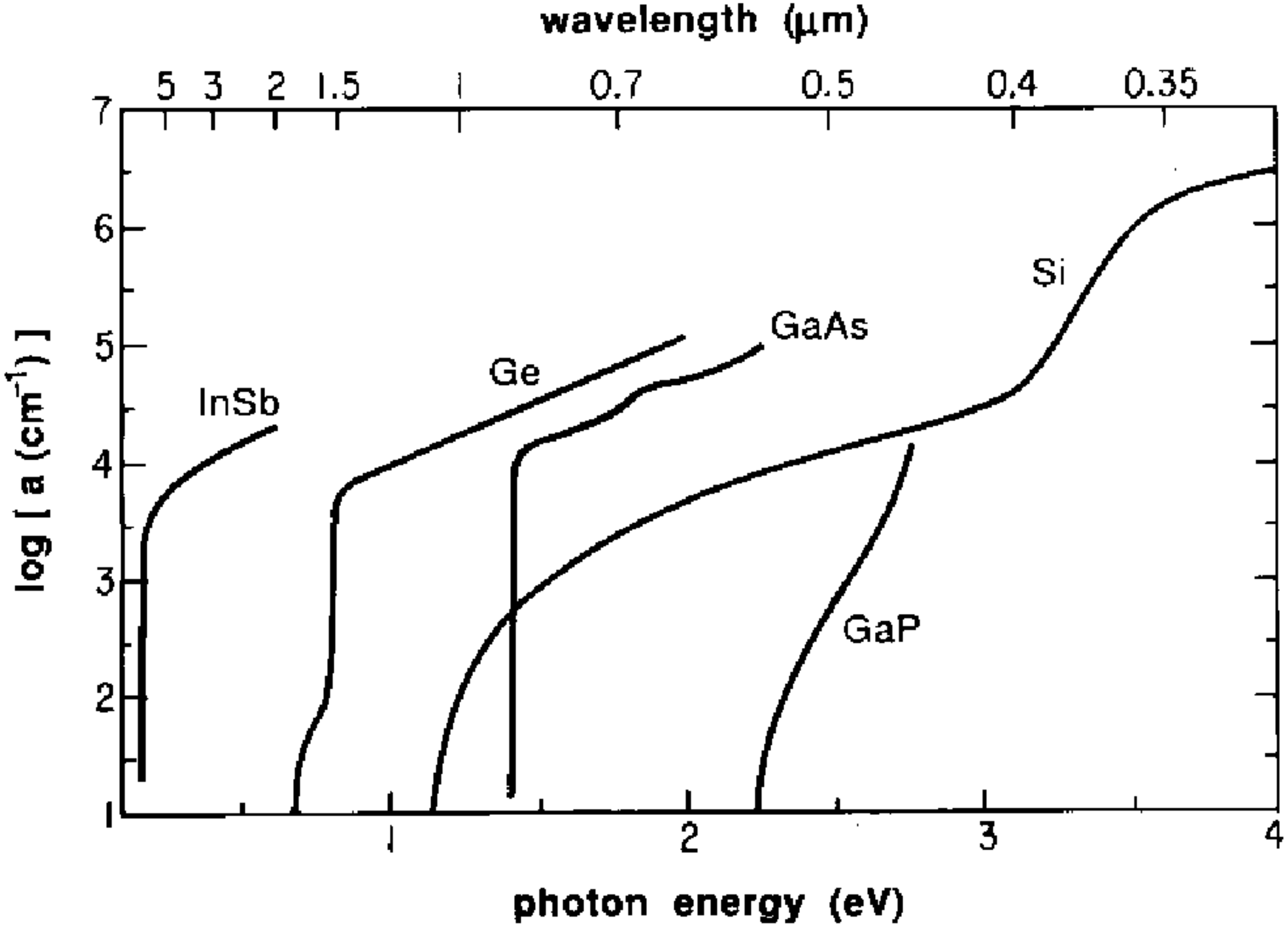
Compare the diagram of crystal structure (above) with the band gap diagrams (below). To free an electron in intrinsic material (1) requires a certain energy indicated by the band gap. It takes less energy to free charge carriers from impurities (2) and (3). A freed charge carrier can move through the detector to produce a photocurrent, which is what we measure.



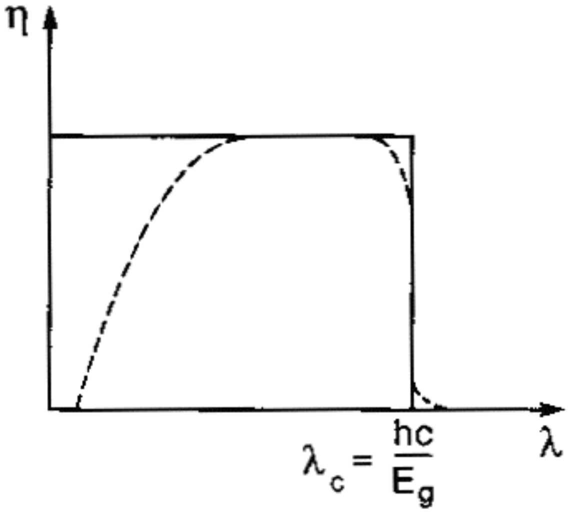


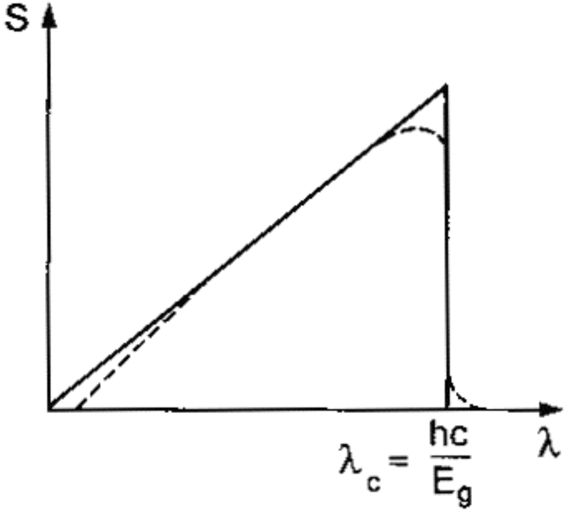
The net absorption is characterized by the absorption coefficient, *a*. Note the difference between direct and indirect absorption (e.g., silicon vs. GaAs) .

The **quantum efficiency** is: $$\eta_{ab} = \frac{S_0 - S_0\, e^{-a(\lambda)\, d_1}}{S_0} = 1 - e^{-a(\lambda)\, d_1} \, ,$$



The **spectral response** drops abruptly to zero at the band gap or excitation energy. The **responsivity** is the amps out per watt of signal in. It rises linearly to the cutoff wavelength for an ideal detector (assuming one charge carrier per absorbed photon).

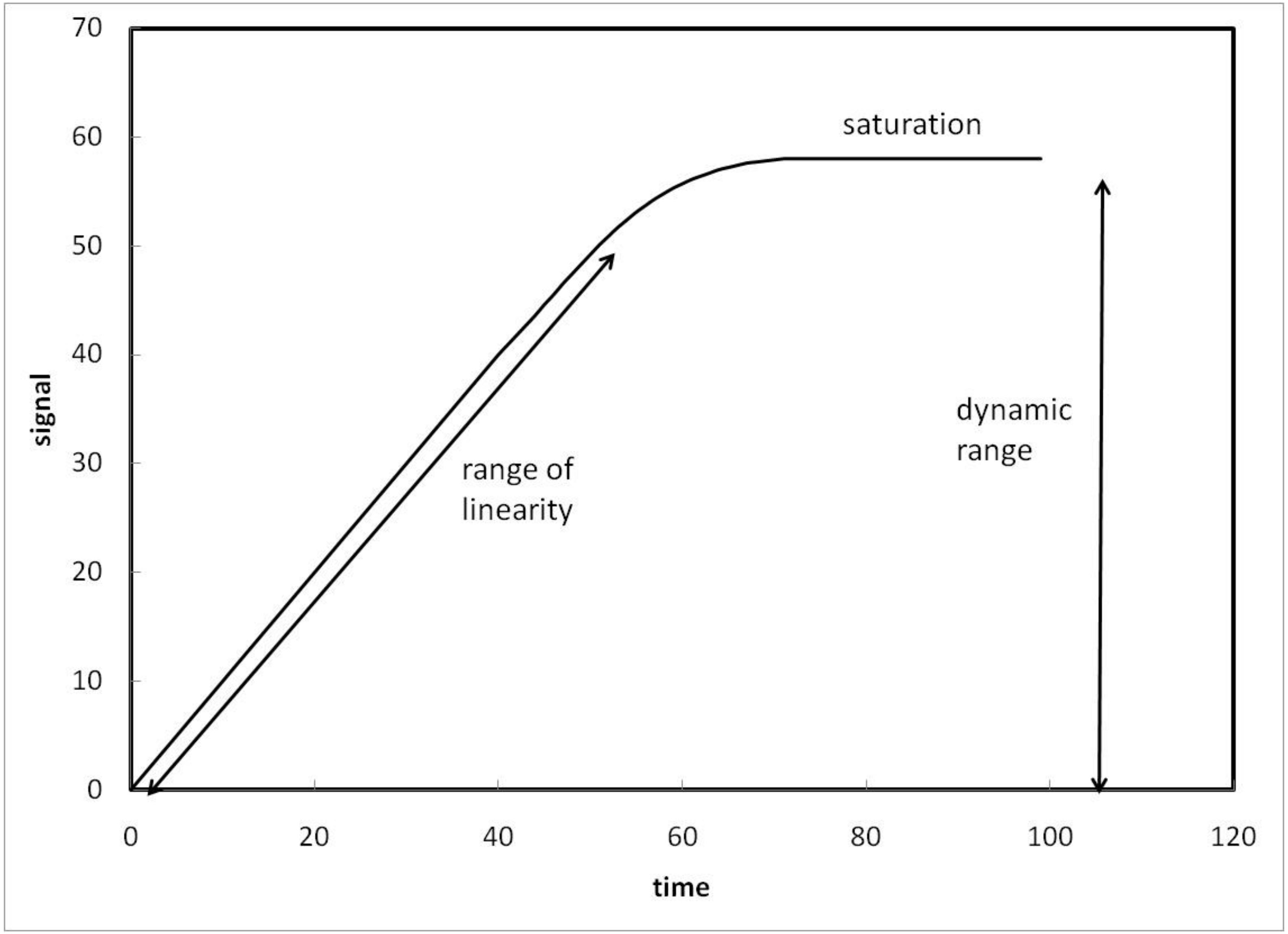




Although the **time response** can have complex behavior, we will deal mostly with simple resistance-capacitance (RC) behavior:

$$v_{out} = \frac{v_0}{\tau_{RC}} e^{-t/\tau_{RC}} , \qquad t \ge 0$$

A decent detector will have close to **linear** response over some range of signal, and will completely saturate at some high level. In between, it is possible to recover information – the range of signal where useful information can be obtained is the **dynamic range**.



General issues in making infrared detectors:

- Minimizing thermal noise (due to Brownian motion of charge carriers) requires large resistance:

$$\langle I_J^2 \rangle = \frac{4kTdf}{R}$$

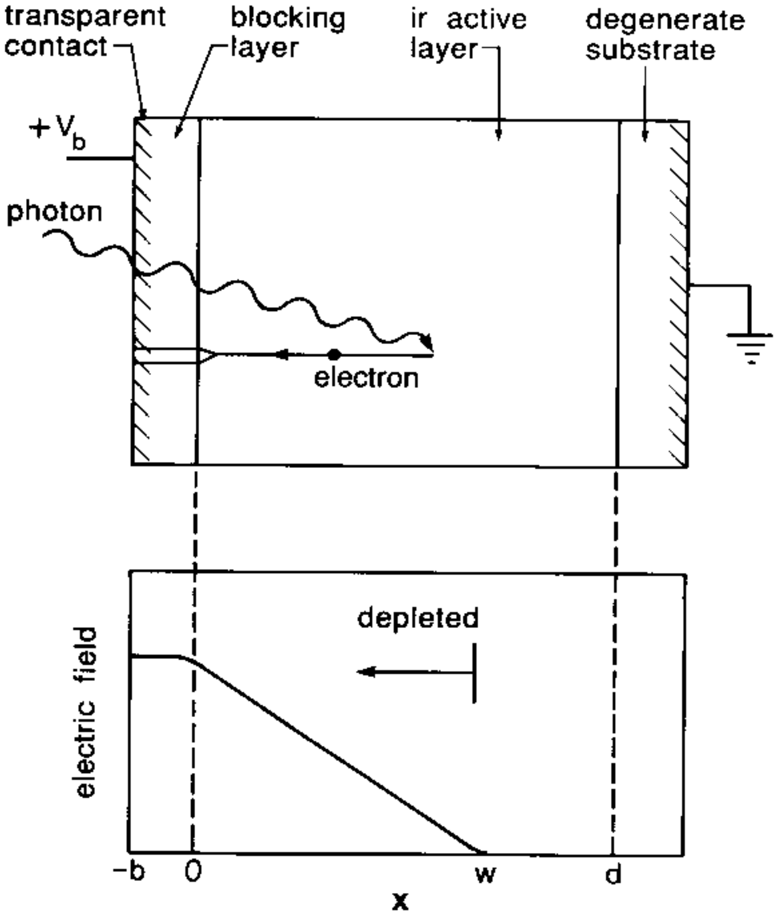
- Very large resistances result in long time constant response due to τ = RC; with the fancy name of "dielectric relaxation" this behavior makes it very difficult to calibrate detectors where there is no other solution to the large-resistance imperative (examples: IRAS, ISO, Spitzer far infrared photoconductors)

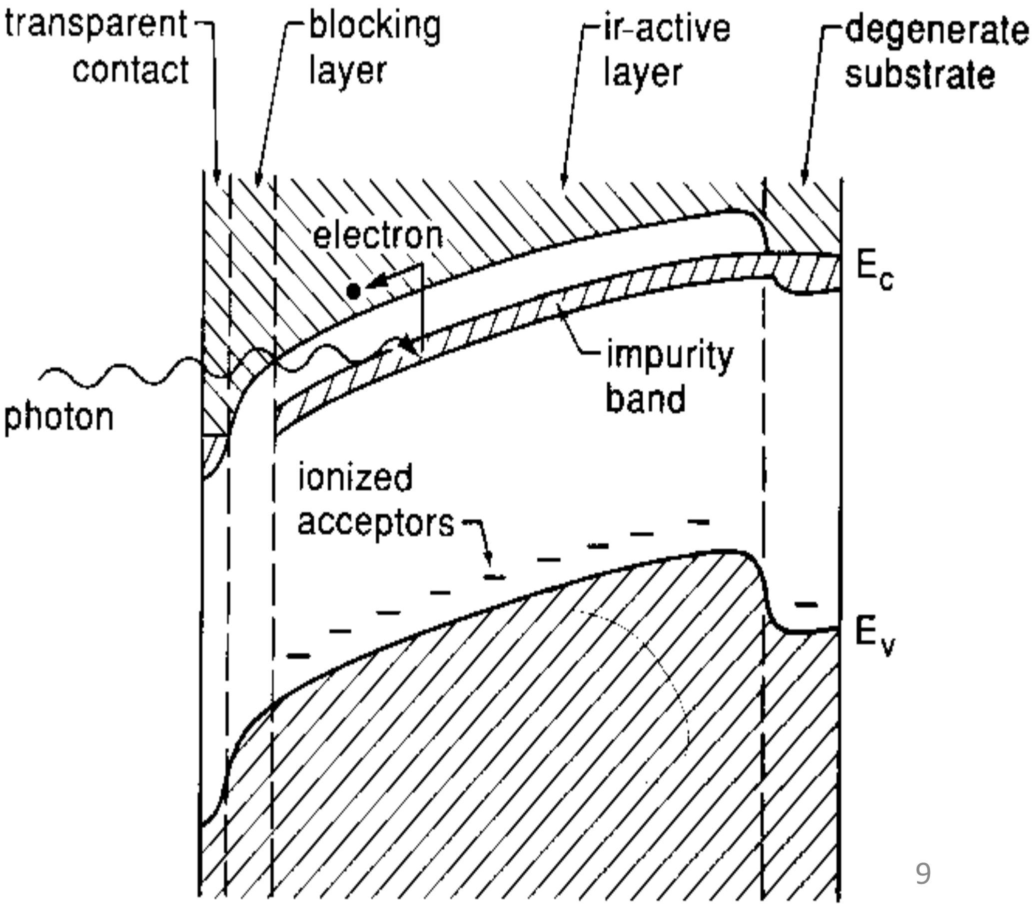
- High-performance detectors need to find a way to maintain the large resistance to thermally-generated charge carriers, but somehow let the photo-generated ones through

- Even with good detector designs, charge carriers can be lost – we want them ***all*!**
  - Traps – charge carriers may get captured temporarily in "traps," which occur at impurities or crystal flaws. In particular, there are many uncompleted bonds at the crystal boundary, which act as traps.
  - Inadequate charge collection – detectors need to be built in ways that allow all the charge carriers to be collected or sensed before they recombine with crystal atoms.

Detector type #1, Si:X IBC

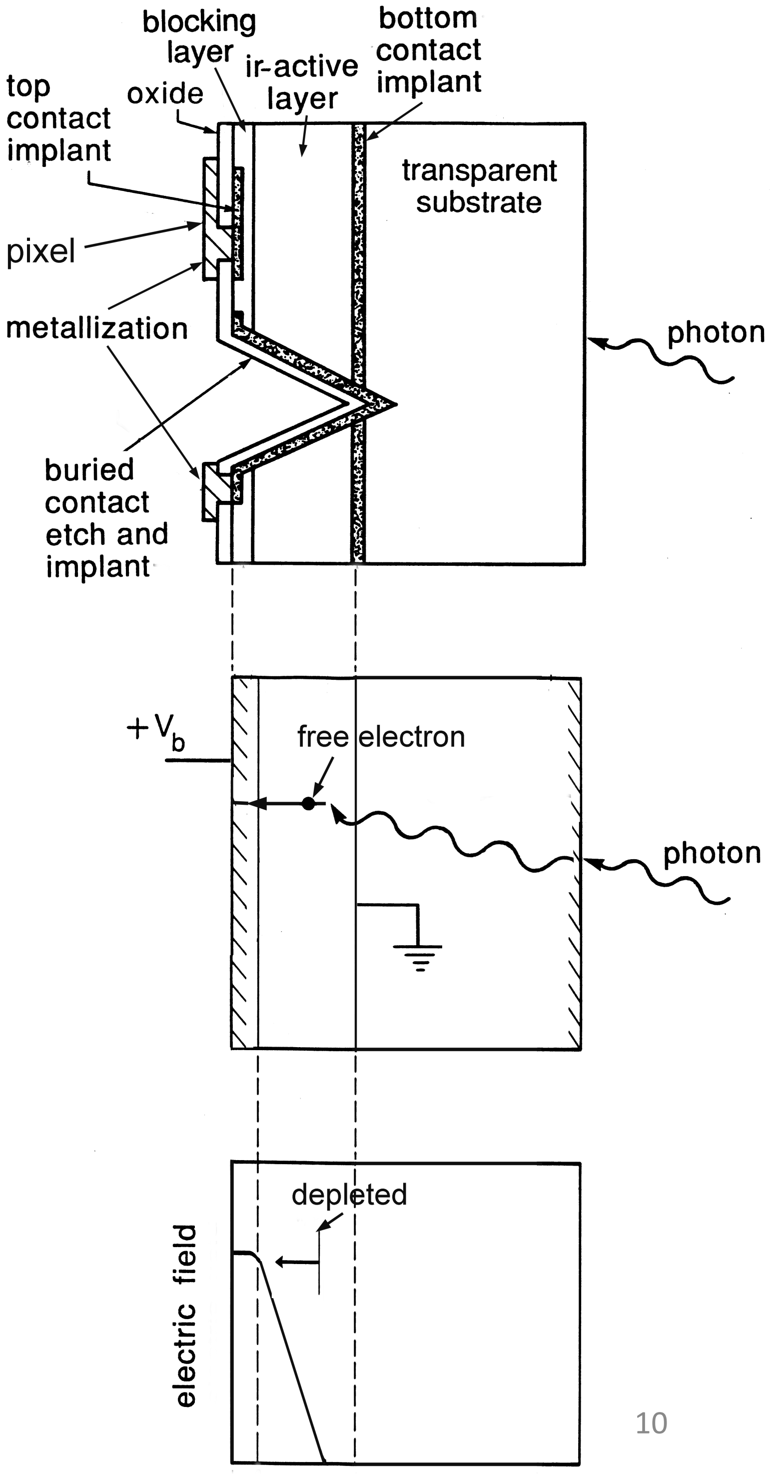
- Physical structure to left, band diagram to right; structure is a thin intrinsic layer, then to right of it a heavily doped absorbing layer, then to right, a contact
- An absorbed photon elevates an electron to the conduction band, from which it can migrate to the contact unimpeded. Thermal charges in the impurity band are stopped at the blocking layer, so dark current is low.
- Detector type of choice for 5 – 35μm
- Notice the separation of zones for electrical properties and photo-response





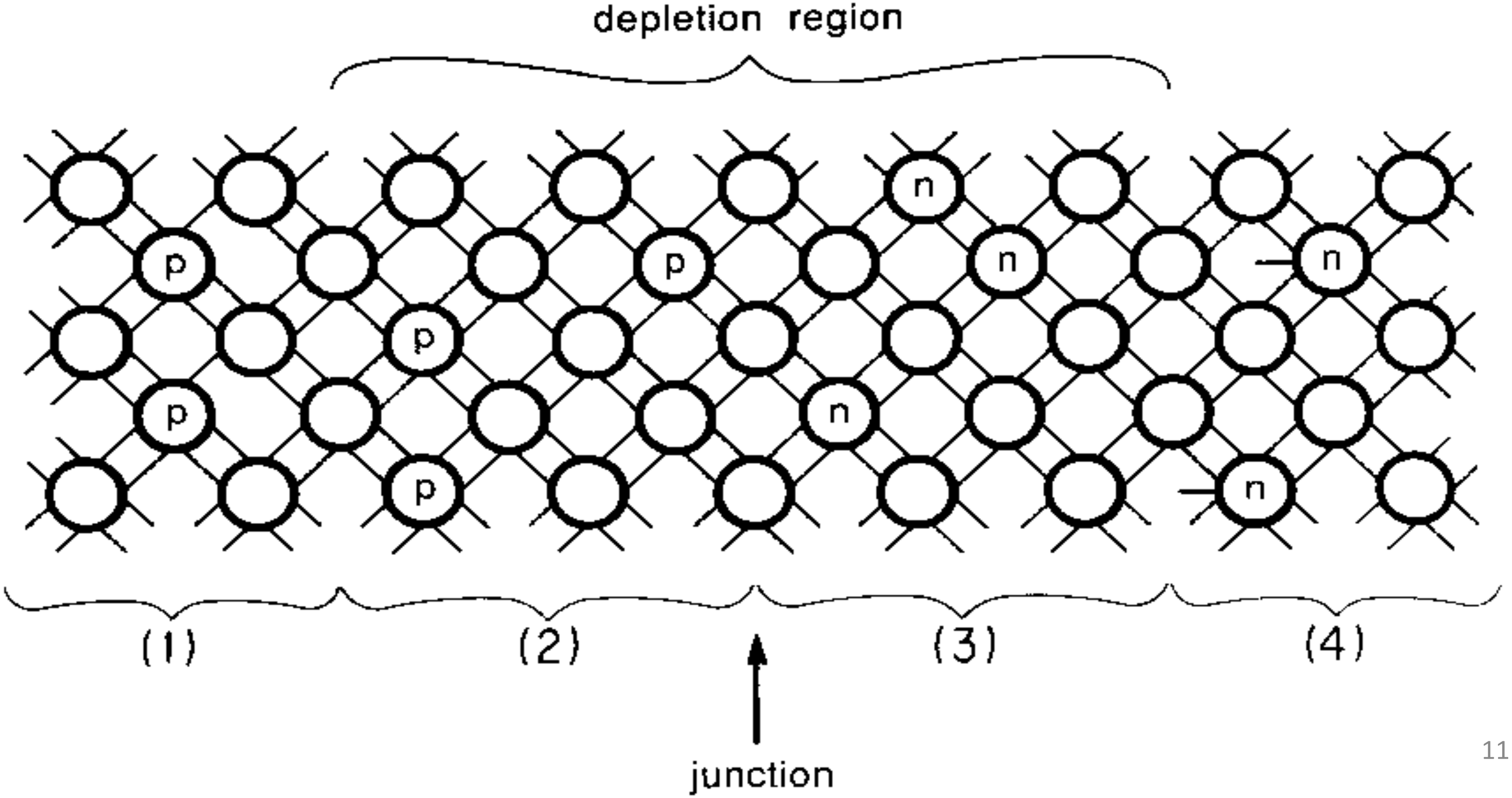
Use of these detectors in an array requires some architecture changes, to allow attaching the readout (to the left in these drawings).

Also, very high purity must be achieved in the silicon to allow for complete depletion of the IR-active layer, or not all of the charge carriers will be collected and the quantum efficiency will suffer. If one tries to increase the depleted zone by increasing the bias, avalanching can occur where the field is largest, increasing the noise.



Detector Type #2, Photodiodes

- A depletion region is created by doping to form a junction between n-type and p-type material. Because all the bonds are complete, this region has high impedance. The charge imbalance creates a field across the depletion region and drives out any remaining free charge.



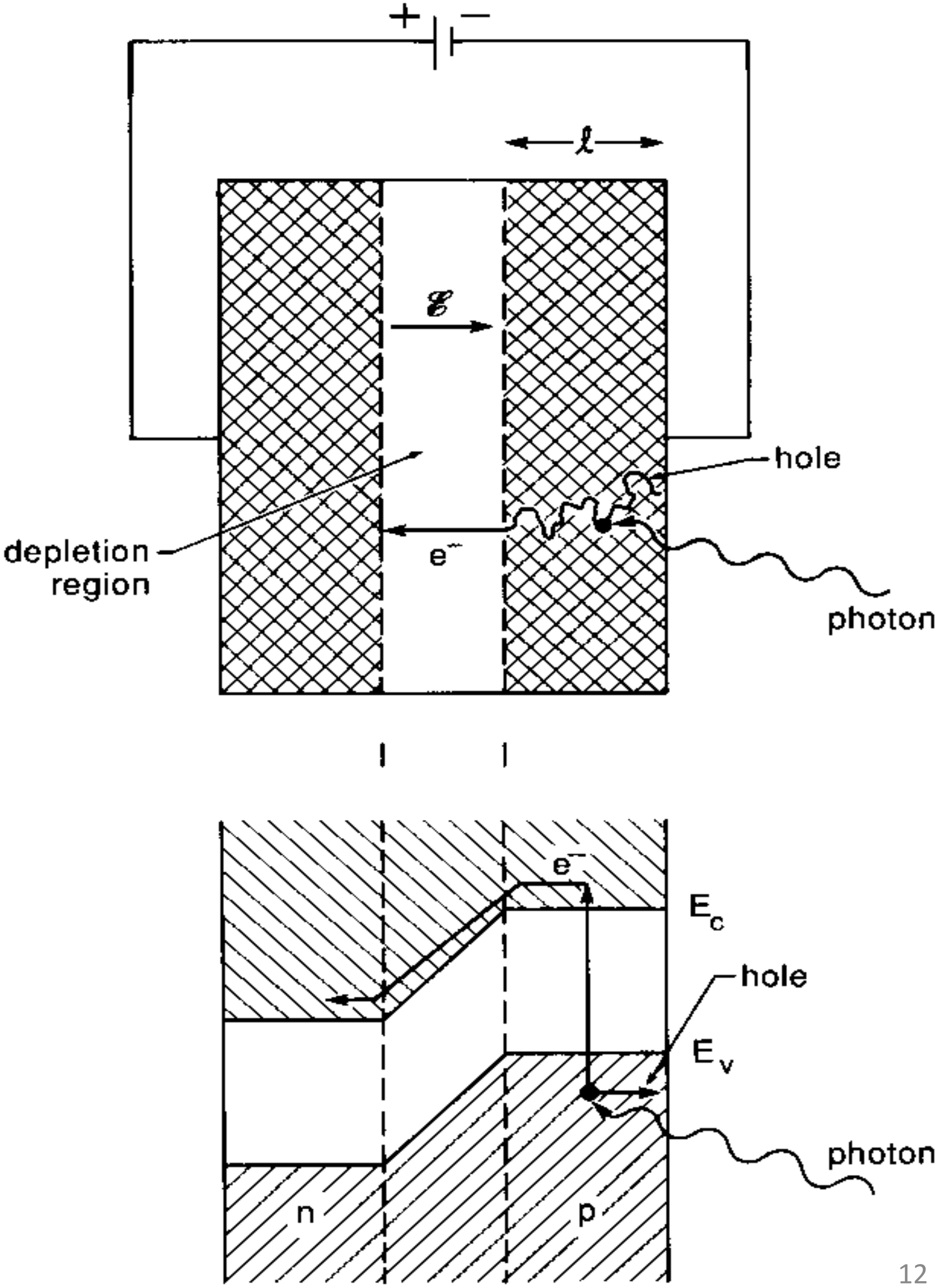
When a photon is absorbed, the freed charge carriers diffuse through the material overlying the junction until one of them encounters the junction; it is then driven across by the internal field to create the photocurrent. The diffusion coefficient and length are:

$$D = \frac{\mu k T}{q},$$

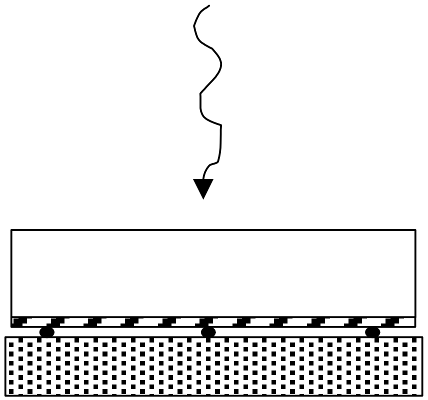
$$L = \sqrt{D\tau}\ .$$

where m is the mobility (characterizes ability of charge carrier to migrate) and τ is the recombination time (goes as $T^{1/2}$).

***L* goes as $T^{3/4}$**



- For operation with low dark current, low temperatures are needed
- The diffusion length is then very small
- Absorbing layer may need to be thinned to 10 - 20μm to collect charge carriers at these temperatures
- Two approaches are shown below
- The yields in doing this can be low, partly explaining the high prices of these arrays.



Diodes on “back” of

- Here are some photodiode materials and their cutoff wavelengths.

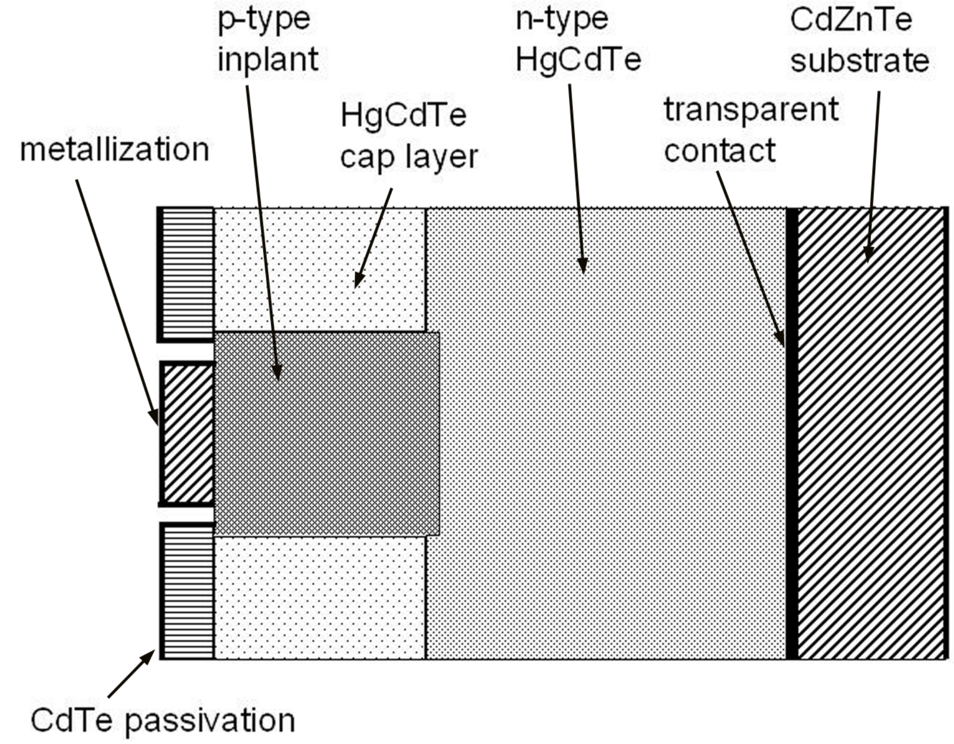
HgCdTe has a variable bandgap set by the relative amounts of Hg and Cd in the crystal. AlGaAsSb behaves similarly.

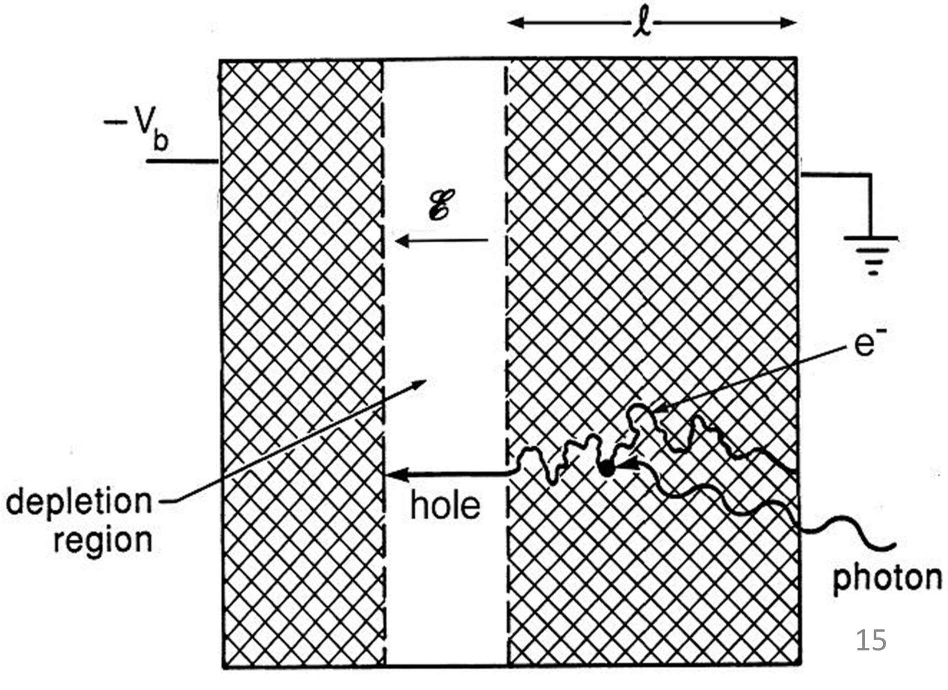
- Indirect absorbers will have poor QE just short of the cutoff

Material	Cutoff wavelength (μm)
Si	1.1 (indirect)
Ge	1.8 (indirect)
InAs	3.4 (direct)
InSb	6.8 (direct)
HgCdTe	~1.2 - ~ 15 (direct)
GaInAs	1.65 (direct)
AlGaAsSb	0.75 – 1.7 (direct)

An example: Teledyne HgCdTe for all the JWST instruments except MIRI. Here is the architecture of a pixel. Photons come in from the right and the contact to the readout is to the left. The cap layer has the Hg/Te ratio adjusted to increase the bandgap, so free charge carriers are repelled without actually having a discontinuity in the crystal, reducing trapping.

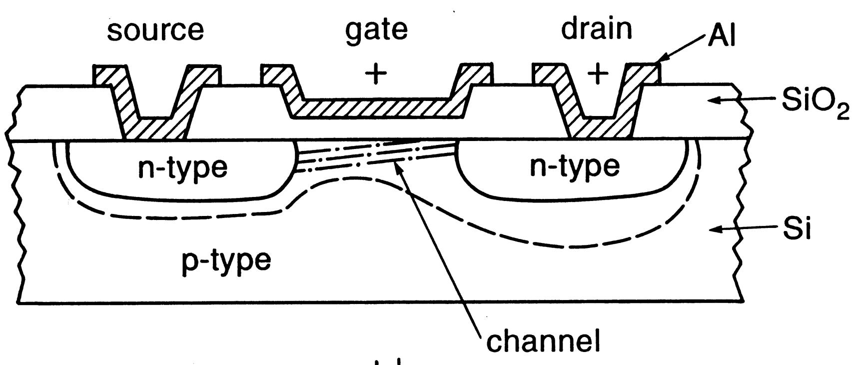
The JWST arrays come with either 2.5 or 5μm cutoff. A competing technology is diodes in InSb, with a cutoff just beyond 5μm.

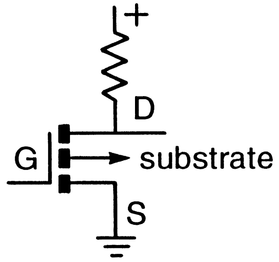


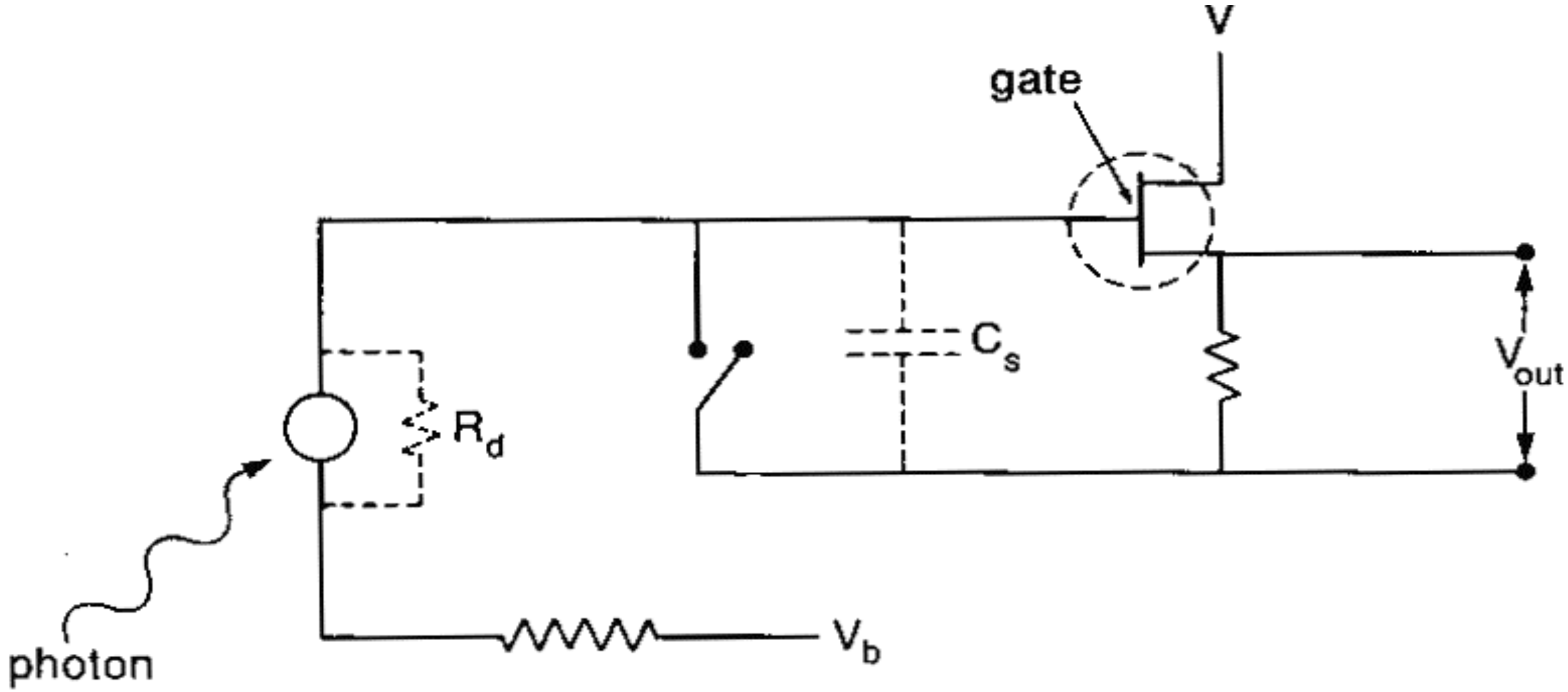


The readout: a source follower simple integrating amplifier

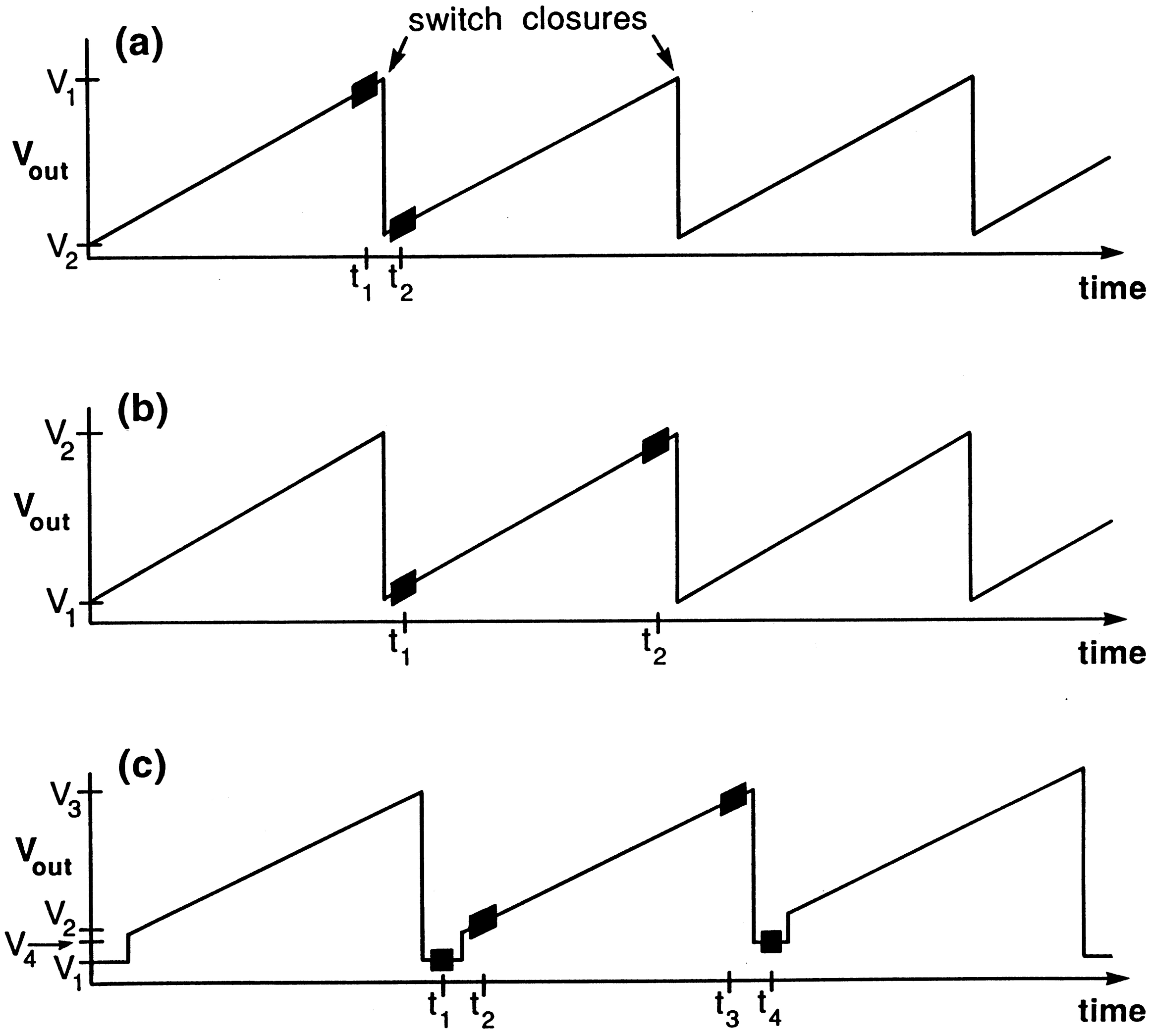
- The photocurrent is conveyed to the gate of a MOSFET
- As charge accumulates on the gate capacitance, it increases the depletion under the gate and modulates the current in the channel
- To keep from saturating, the charge is reset from time to time







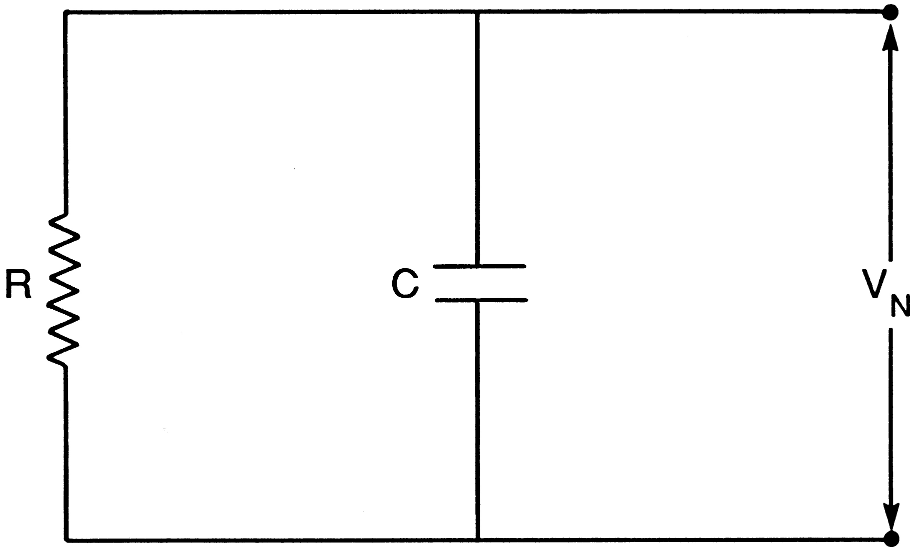
How should we sample the output?



kTC or Reset Noise

- The circuit below has both potential energy (charge on the capacitor) and kinetic energy (Brownian motion of electrons in the resistor)
- From thermodynamics, we have kT/2 of energy with each degree of freedom, leading to:

$$\frac{1}{2} C \left\langle V_N^2 \right\rangle = \frac{1}{2} kT \qquad \left\langle I_J^2 \right\rangle = \frac{4kTdf}{R}$$



From the left expression we get:

$$\left\langle Q_N^2 \right\rangle = \frac{kT\,C_S}{q^2}. \qquad (24)$$

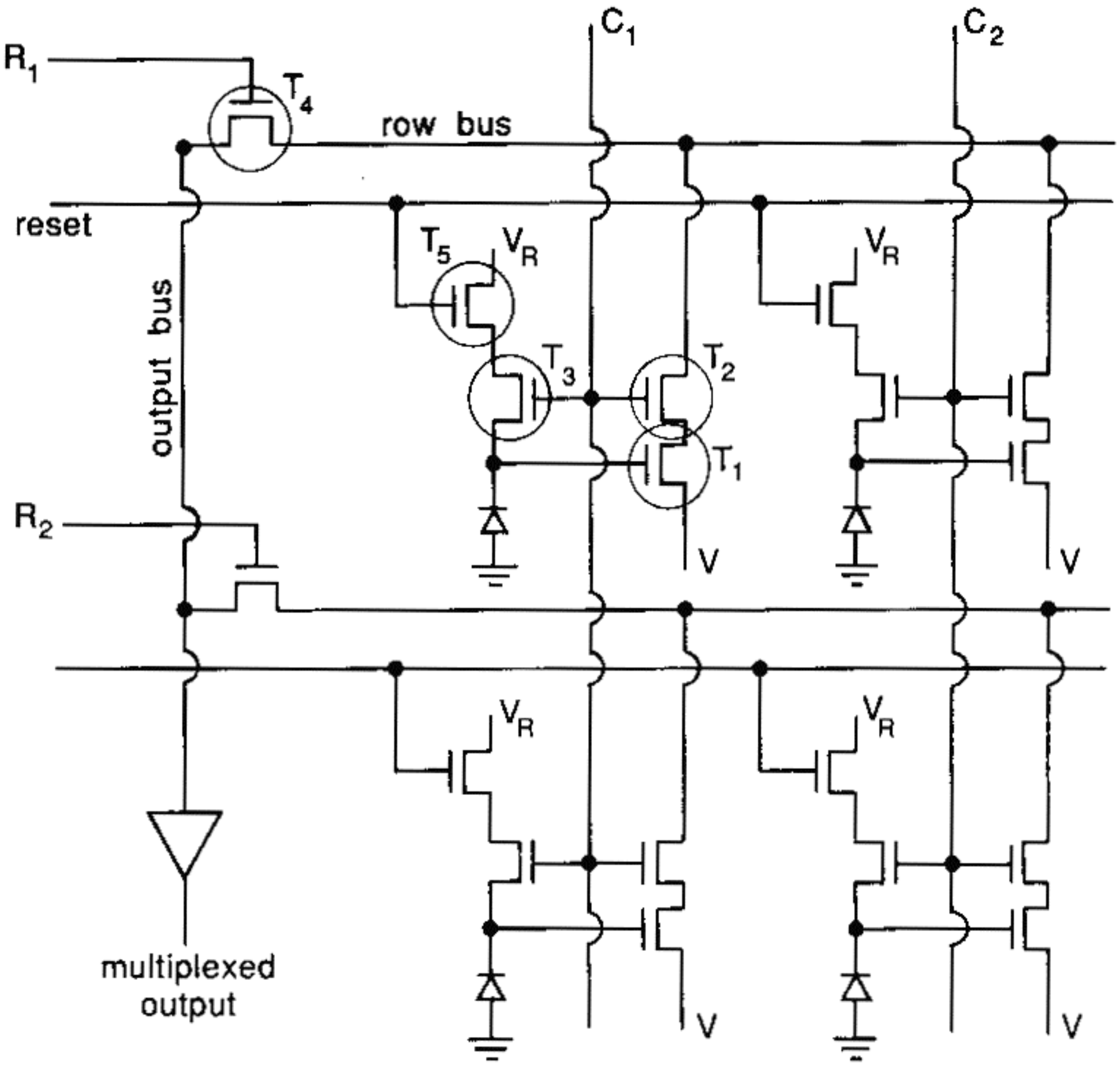
For C = 10^{-13} F and T = 40K, the noise is about 45 electrons rms. A NIRCam array has a read noise of 6 - 7 electrons. How is this done?

- To avoid reset or kTC noise, we have to use readout strategy (b) or (c)
- Then, if R = $10^{19}\ \Omega$ (say), τ = RC = 10^6 sec
- (c) lets us take out some forms of slow drift by sampling with the reset switch closed. However it adds root2 more amplifier noise
- Modern arrays can support strategy (b) even for integrations of 1000 - 2000 seconds
- The kTC noise is then (45e)*(1 – $e^{-t/RC}$) = 0.1e for 2000 seconds
- In general, the reset noise is

$$read\ noise = \sqrt{\frac{kTC}{q^2}}\,(1 - e^{-\frac{t}{RC}})$$

with additional components from amplifier noise and other sources.

Now consider the operation of an array of readout amplifiers:

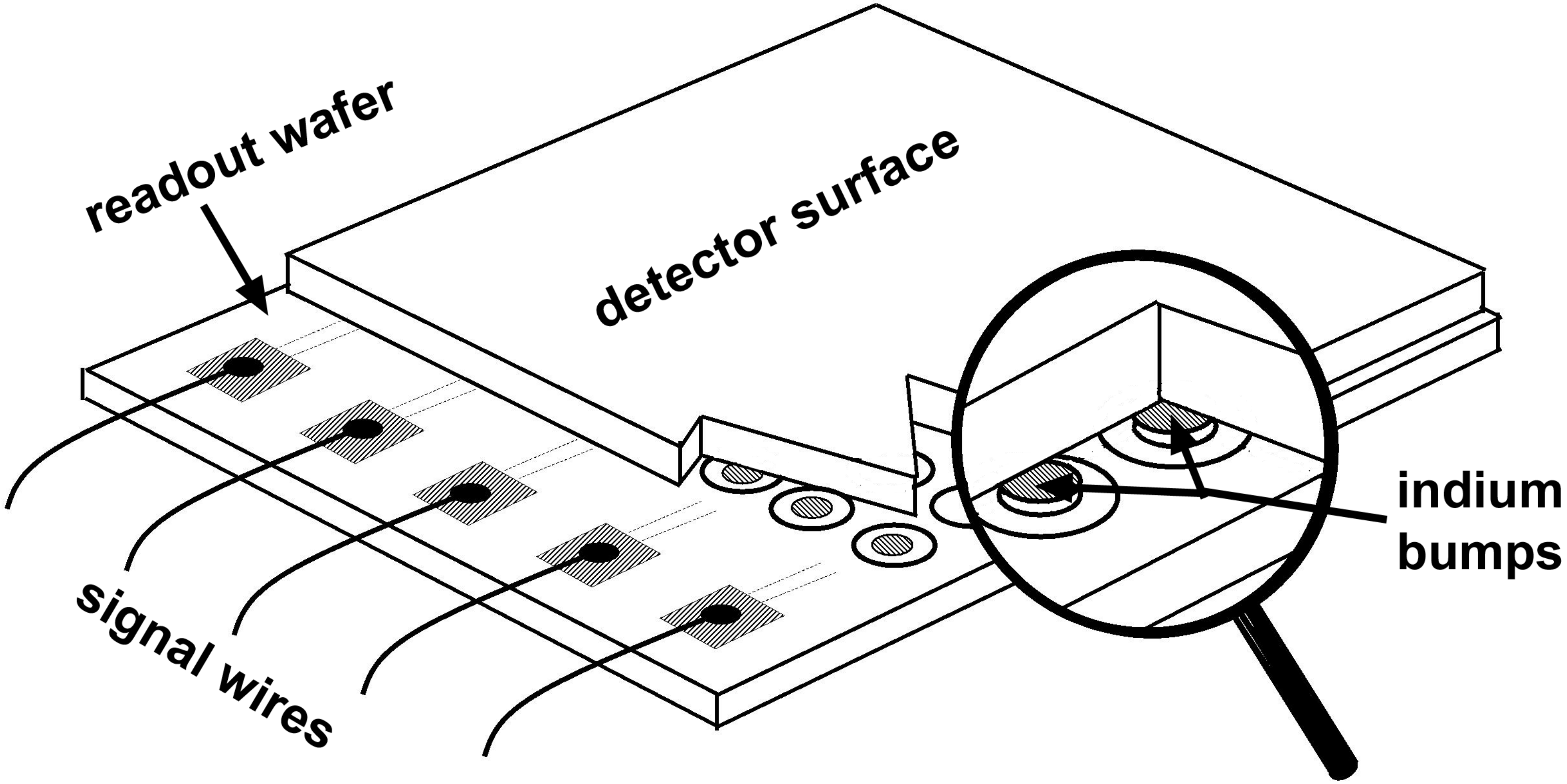


The source follower amplifier is T_1. T_5 is the reset switch. T_2 and T_3 can turn off power to T_1 and isolate it, allowing signal to accumulate on its gate with no power. When T_2,T_3, and T_4 are turned on the signal comes out the row bus to the output. If desired, T_1 is then reset by closing T_5.

This approach has random access and reads nondestructively. That is, we can read out any pixel we want, and we can read it and then leave it exactly as it was with the same signal. This allows interesting read out patterns, like Fowler sampling (multiple samples to drive down amplifier noise) or sampling up the integration ramp.

Infrared Detector Arrays

- Best performance with silicon integrated circuit readout
  - Cannot manufacture high quality electronics in other semiconductors
  - CCD-type readout has charge transfer problems at cold temperatures
  - Therefore provide an amplifier for each detector (as in preceding slide)
- Direct hybrid construction
  - Fields of indium bumps evaporated on detector array, readout amplifiers
  - Aligned and squeezed together - very carefully

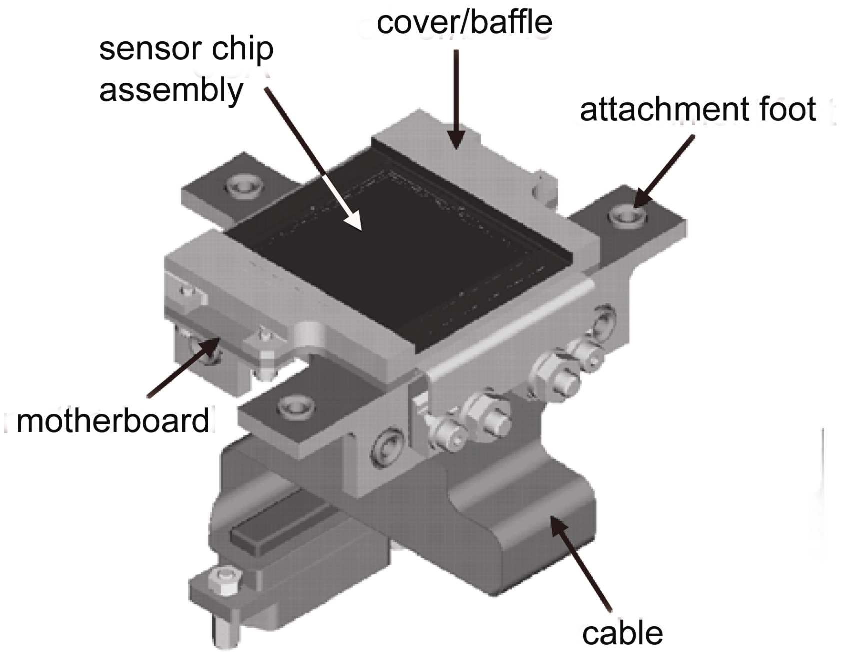


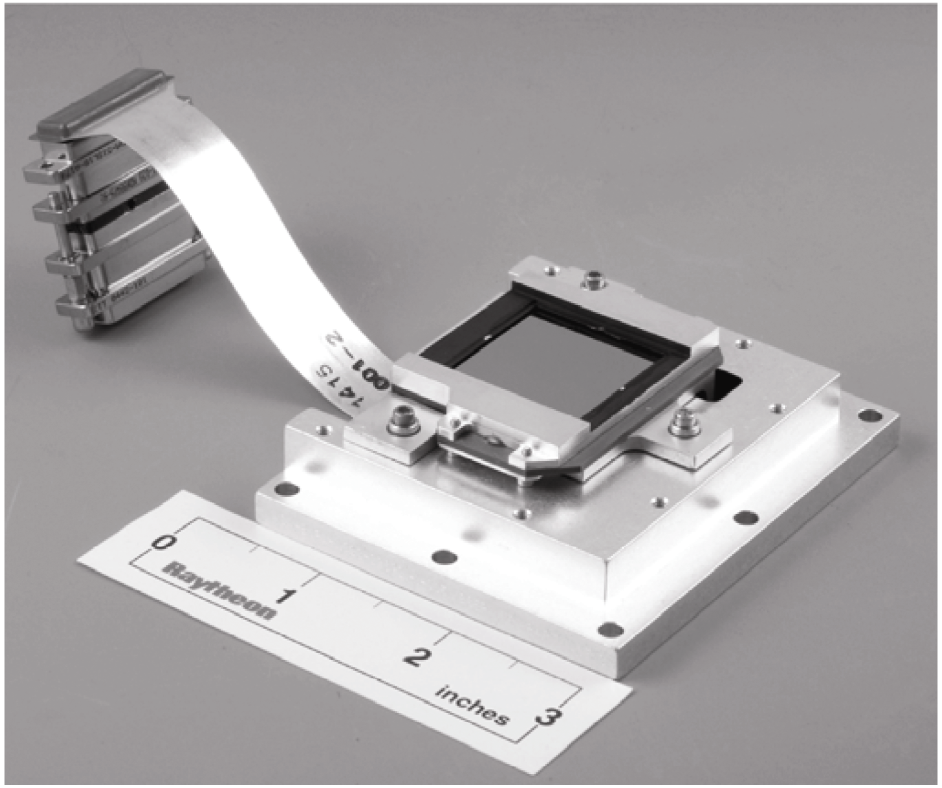
Here is a state-of-the-art Si:As IBC array (made by Raytheon for MIRI on JWST).

It is 1024 X 1024 pixels and at ~ 6.7K delivers dark current < 0.1 e/s, read noise of ~ 15 e rms, and quantum efficiency > 60% from 8 to 26μm (~ 10% of maximum still at 28.3μm and > 40% at 5μm). The pixels are 25μm on a side.

A special process is used for the readout circuit so it works well at such low temperatures.

Similar devices but somewhat lower performance have been made by DRS Technologies for WISE.

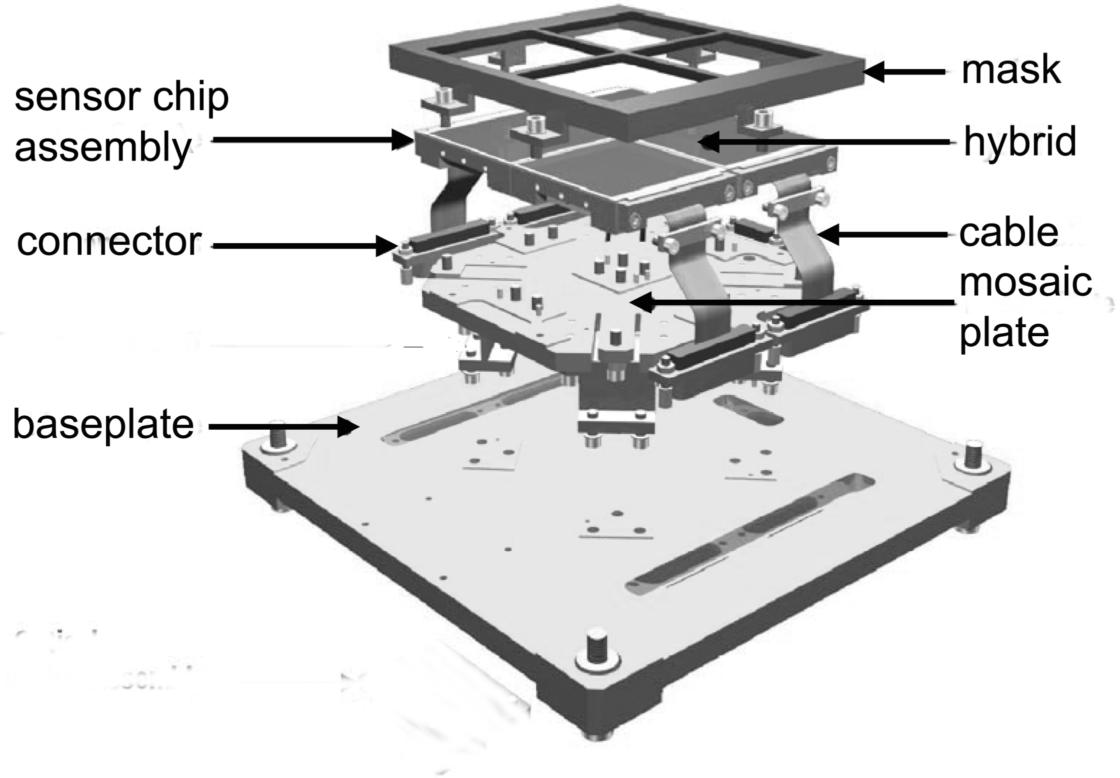


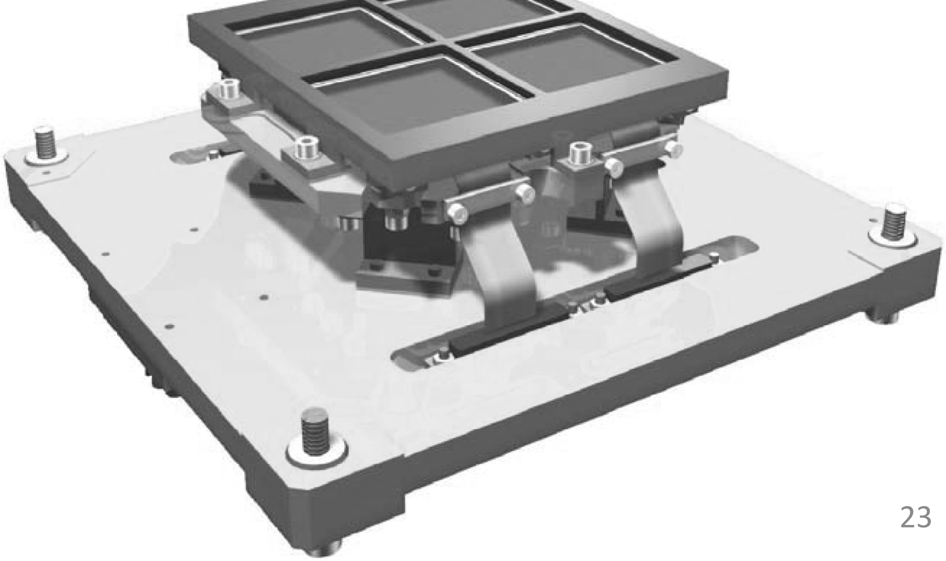


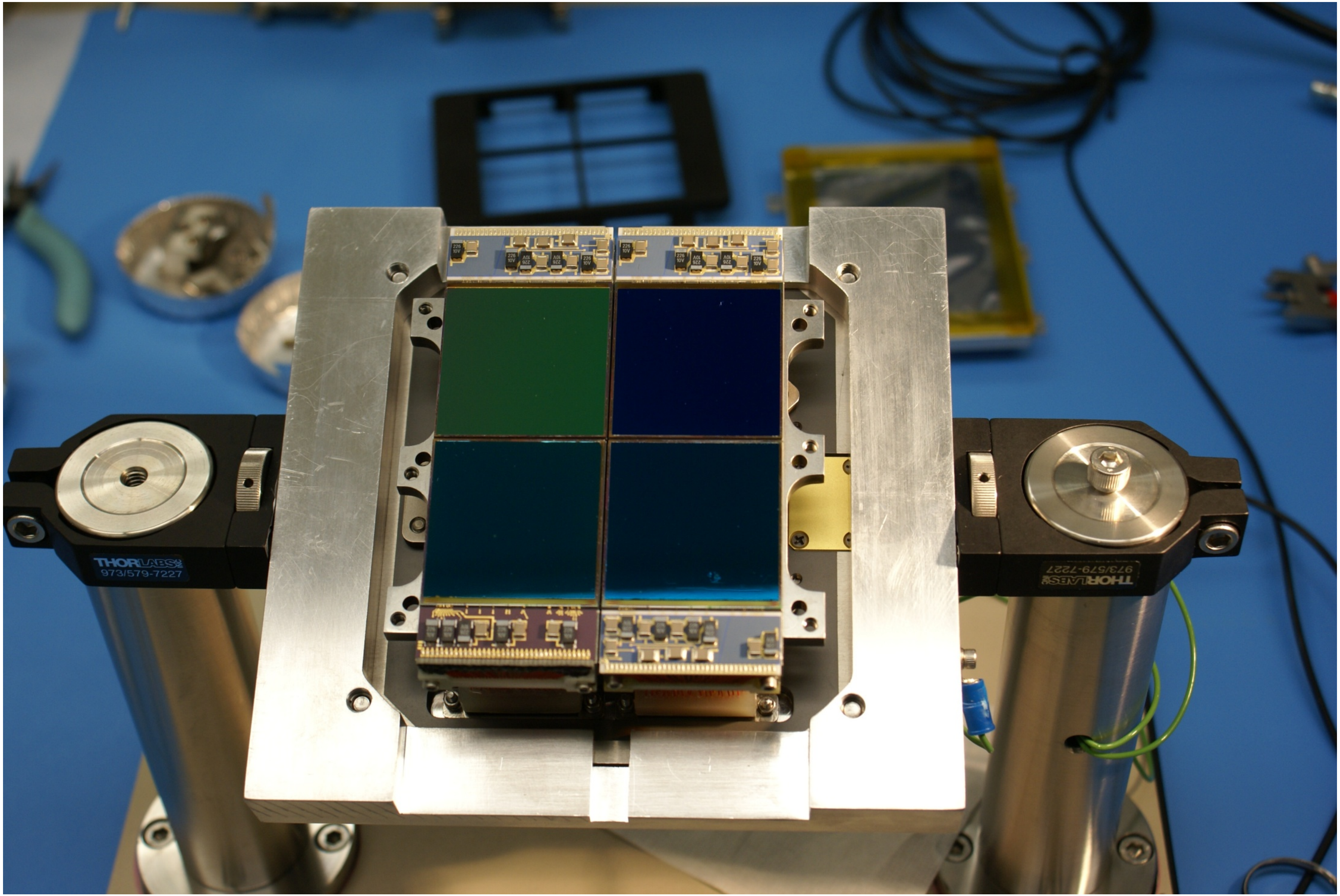
An example: the HgCdTe arrays for NIRCam and other JWST instruments.

They are 2048 X 2048 pixels in size, with 18μm pixels. Dark currents at ~ 37K are 0.004 e/s for the short cutoff and 0.01 e/s for the long. The QE is > 90% from 1μm to close to the cutoff and > 70% between 0.5 and 1μm. The read noise is ~ 6-7e rms.

Similar devices but somewhat lower performance have been made by Raytheon for VISTA.






THORLABS
973/579-7227
THORLABS
973/579-7227

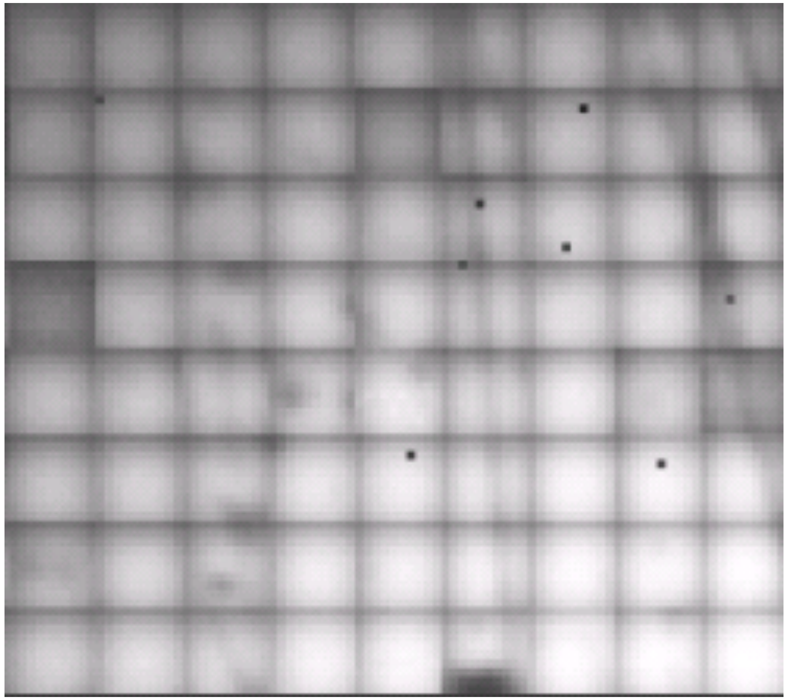
OK, it’s a lovely detector, but how do we get the most out of it?

Here is a menagerie of problems:

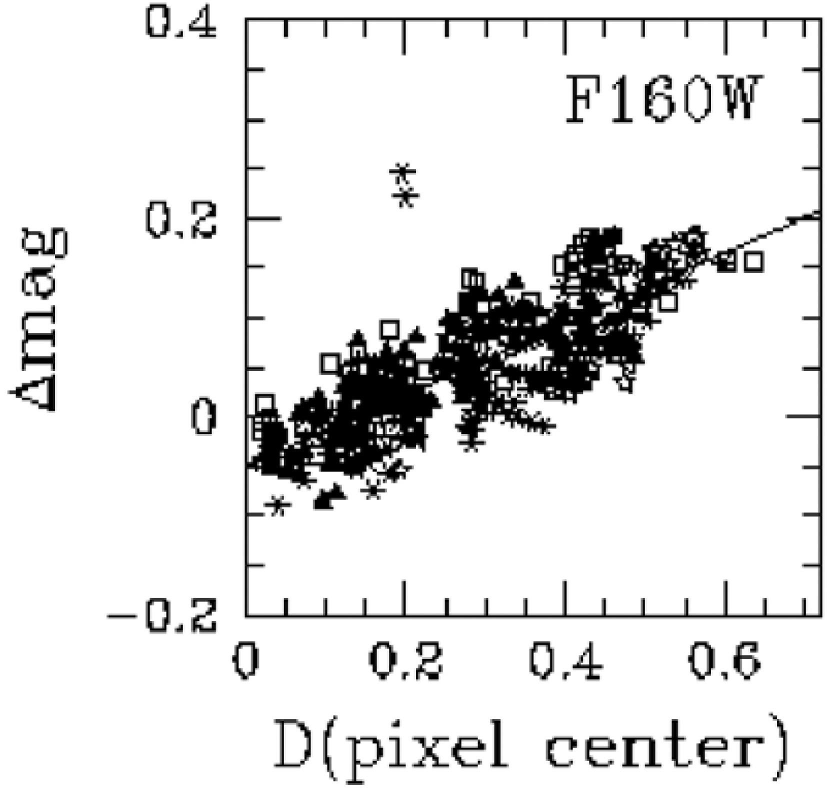
ghost images

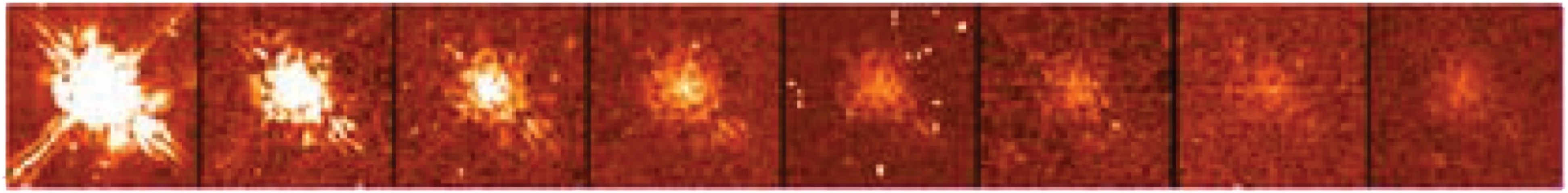
$\cos^N\theta$

response gaps between pixels



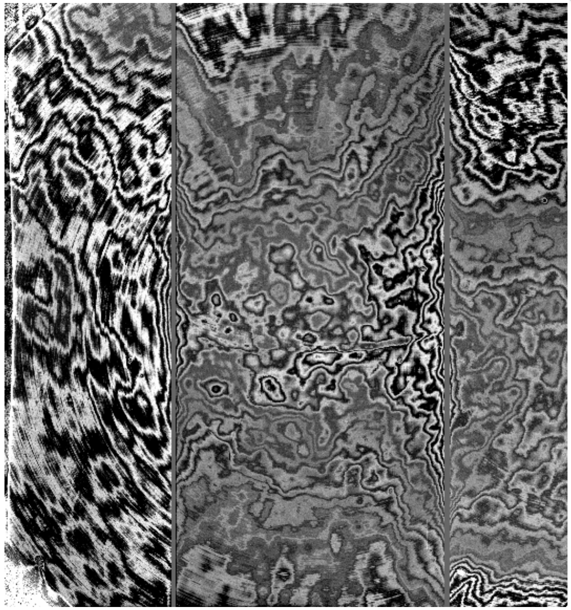
response variations across a pixel



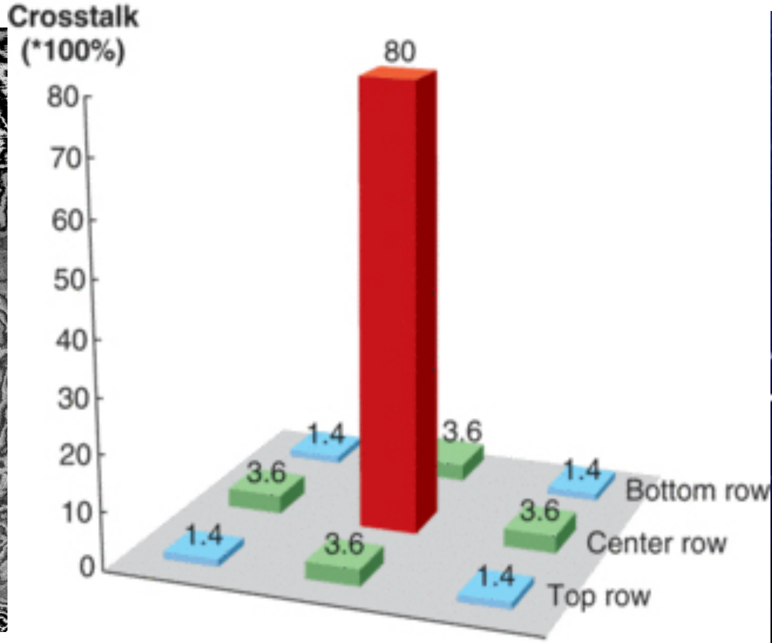


latent images

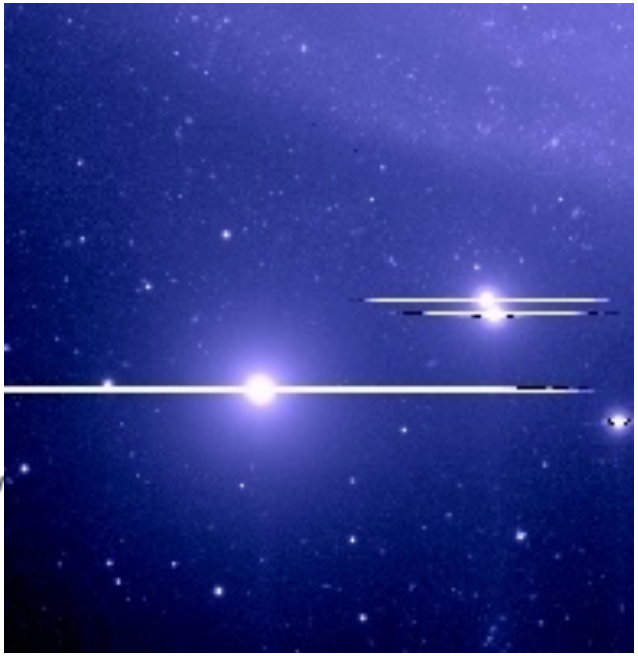
fringing

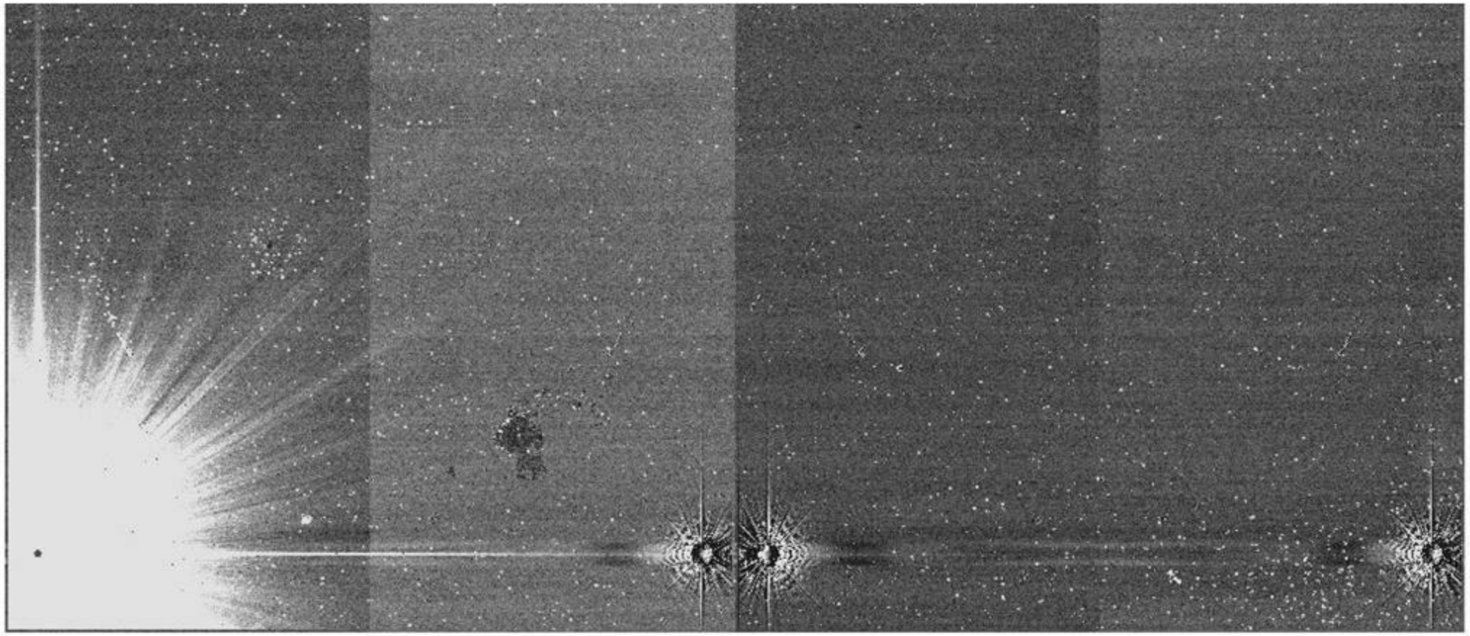


cross talk from one pixel to another in general



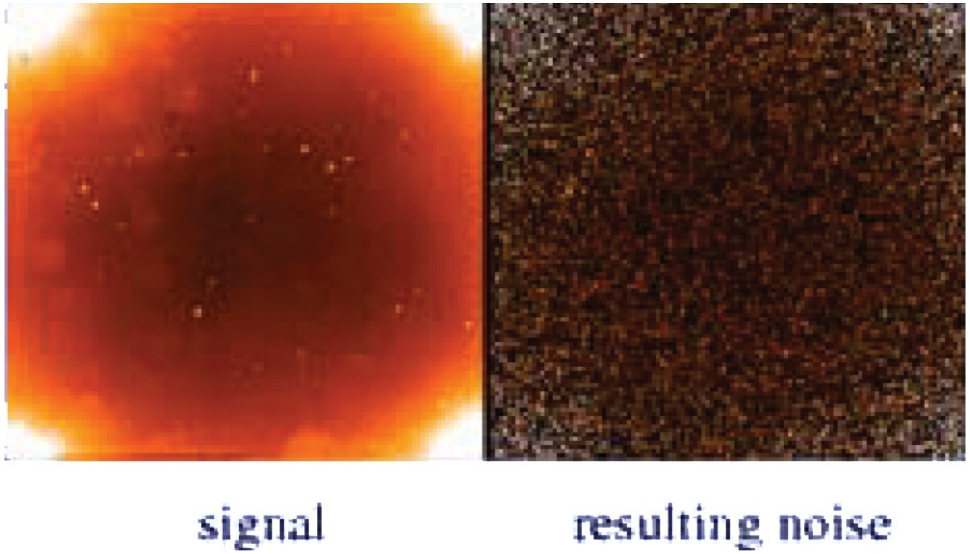
bleeding





electronic ghosts

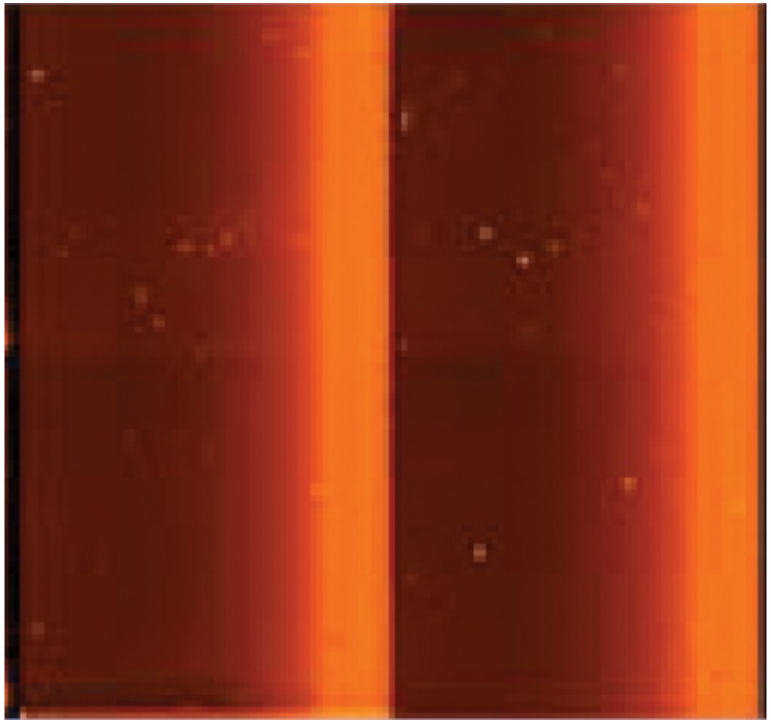
amplifier glow



signal

resulting noise

pedestal effect



Hot and dead pixels

Fixed patterns due to readout and array growth processes - see Orion arrays to right - 2K X 2K InSb

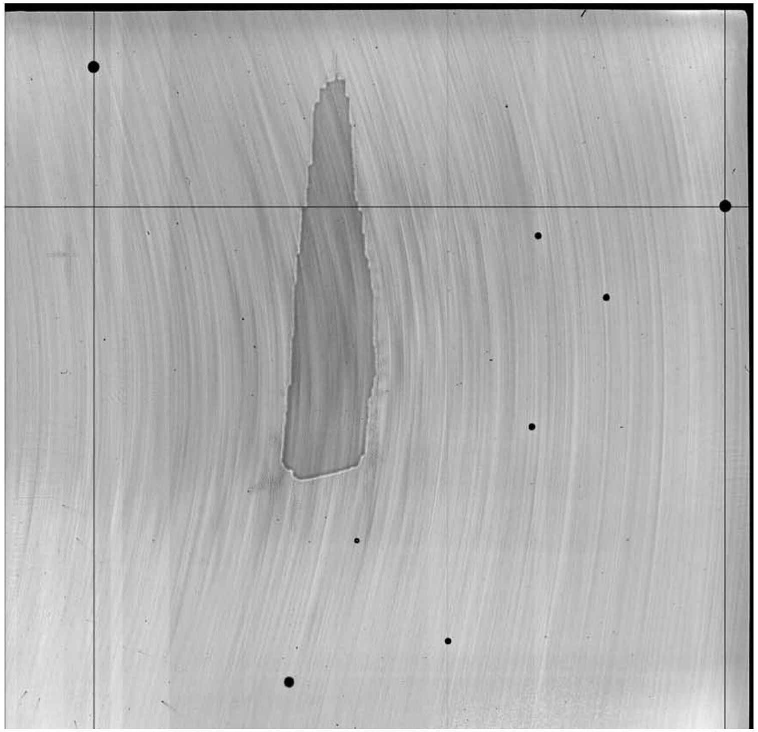
Cosmic ray hits

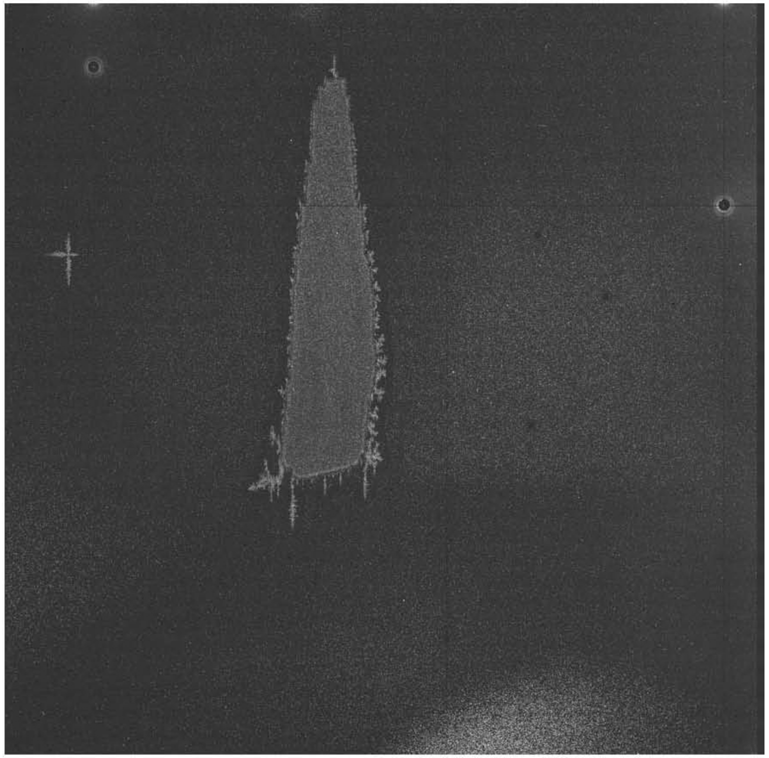
Amplifier transients

Nonlinearity and soft saturation

Photon emitting defects (PEDs)
See Orion arrays to right

Thermal drifts
Freeze-out of charge carriers





Don’t Despair!!

- **Repetition is important**
  - Lets you remove transient signals like cosmic ray hits, amplifier transients
- **Not changing things is important**
  - Many array artifacts like pedestal effects and MUX glow can be removed almost perfectly if the observing conditions are not changed
  - Thermal drifts are minimized by keeping a constant cadence on the array
- **Dithering is also important - putting the signal on a variety of pixels**
  - Lets you replace bad pixels with good data
  - Allows generating calibration frames from the sky - can be the best kind of calibration
  - Can allow you to identify and remove latent images
- **Reference pixels (IR arrays) or overscan (CCDs) may help solve some problems**
  - Let you track the behavior of the readout electronics and fix it in data reduction
- **Pixel scale is important**
  - Coarse pixel scales (relative to Nyquist) lose information irretrievably
  - Coarse pixels make it difficult to remove intra-pixel sensitivity variations, make your measurements susceptible to inter-pixel gaps and cross talk, make cross talk a more significant problem, and so forth.
  - Fine pixels limit the field of view and may increase the effective noise

The first rule of array imaging is repetition

- Multiple images allow systematic identification of outlier signals due to cosmic rays and other transients
- They also allow replacing areas compromised by cosmic rays, latent images, hot or dead pixels, ghosting, etc. with *real data.*
- By dithering the pointing on the sky between exposures (moving the telescope slightly so the images fall on different parts of the array), the sky signal itself can be used to flatten the image (as discussed below). If the sky dominates the signal, then fringing effects are removed to first order, along with many other potential contributors to non-flatness.
- Properly sampled images are another form of repetition - more than one pixel contributes to the signal. Accurate photometry benefits from spreading the light over multiple pixels (which can also be done with dithering a lot).

- Modern arrays often include non-active pixels that are electrically identical to those that detect photons. For CCDs, some benefit is obtained by overscanning the array, while for infrared arrays they are physical outputs called reference pixels. They can be used to correct the images for slow drifts in the electronics.

The second rule of imaging photometry is don't change anything

- Artifacts like MUX glow, pedestals, and many others will disappear from your reduced data virtually completely if you are careful to take all your data - science and calibration frames - in identical ways (for example, the identical exposure times and readout cadences)
- Detector arrays also perform better when they reach equilibrium, i.e., constant exposure times and readout cadences, plus constant temperatures, backgrounds, and etc.

A good strategy for imaging is:

- Take repeated exposures of the field, moving the source on the array between exposures
- Generate the response frame by a median average of these frames – sources will disappear because they do not appear at the same place on any two frames
- Obtain dark frames with the same exposure time as used for the data and response frames
- Subtract dark from data and response (also takes out offset); divide corrected data by corrected response
- Shift frames to correct for frame-to-frame image motions
- Median average again to eliminate bad pixels and cosmic rays, while gaining signal to noise on the source image

In general, the image reduction software will include standard or recommended procedures to generate the necessary calibration frames from your data, and to shift and add all your science frames into one high-quality image.

Is the job done?

Recall the list of possible array problems we discussed earlier. Some of them should be taken care of at this stage, although they might have required some extra processing: 1.) fringing; 2.) hot and dead pixels; 3.) cosmic ray hits; 4.) latent images; and 5.) MUX glow. The next to last item might require generating a special flat field frame designed to just capture the latents. You might also have to identify electrical and optical ghost images and other such effects and fix them by hand or with custom routines.