

# Lecture 13: Image Sensor Basics

---

**Yuhao Zhu**

<http://yuhaozhu.com>  
[yzhu@rochester.edu](mailto:yzhu@rochester.edu)

CSC 259/459, Fall 2024  
Computer Imaging & Graphics

# The Roadmap

Theoretical Preliminaries

Human Visual Systems

**Digital Camera Imaging**

Modeling and Rendering



Photographic Optics

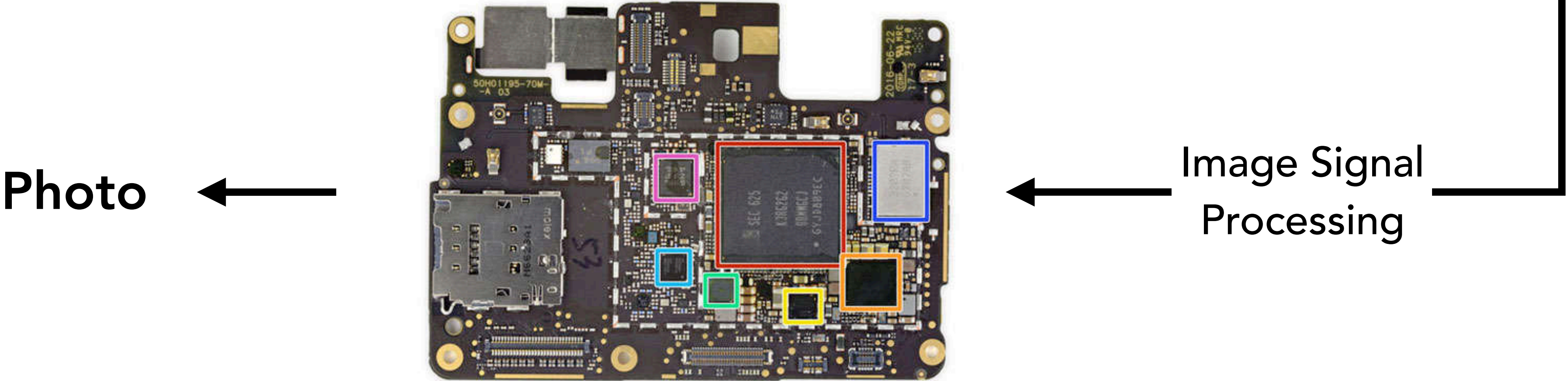
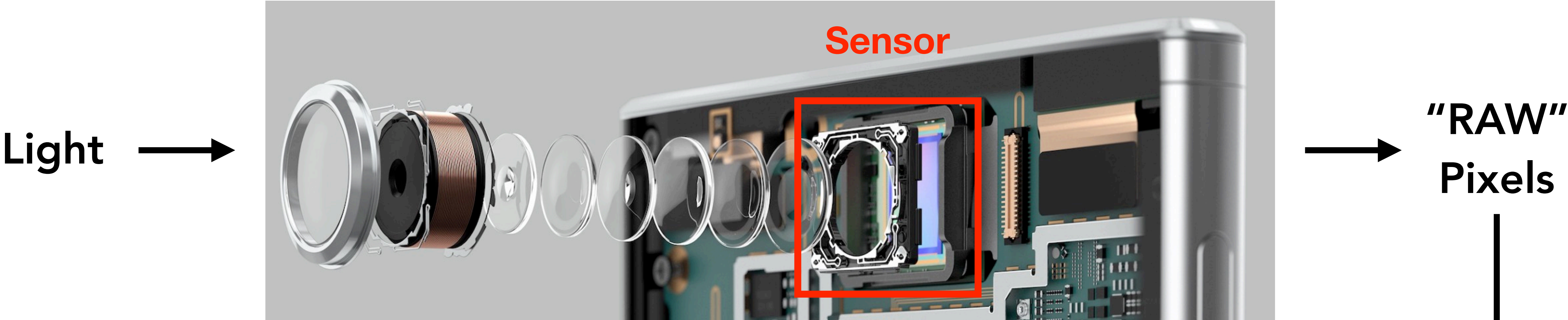
**Image Sensor**

Image Signal Processing

Image/Video Compression

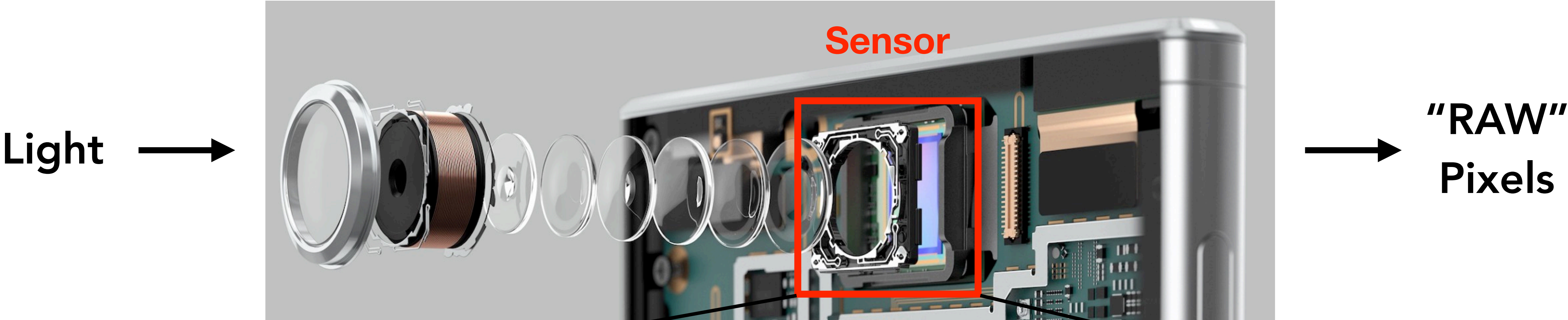
Immersive Content

# Digital Cameras



<https://ph.priceprice.com/Google-Pixel-XL-18713/news/Google-Pixel-XL-Repair-1939/>

# Image Sensor in Digital Cameras



Optical Elements

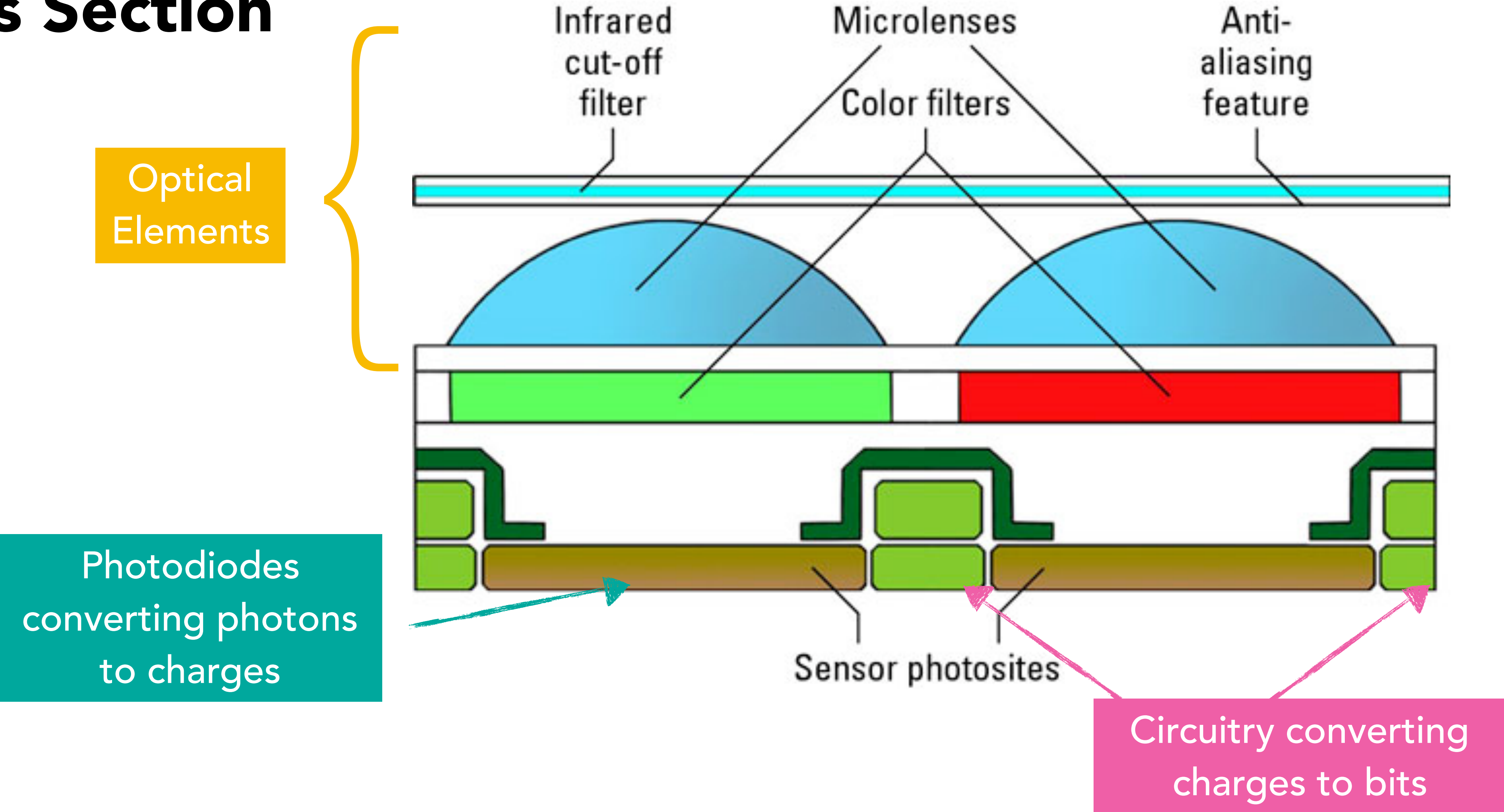


Converting photons to charges

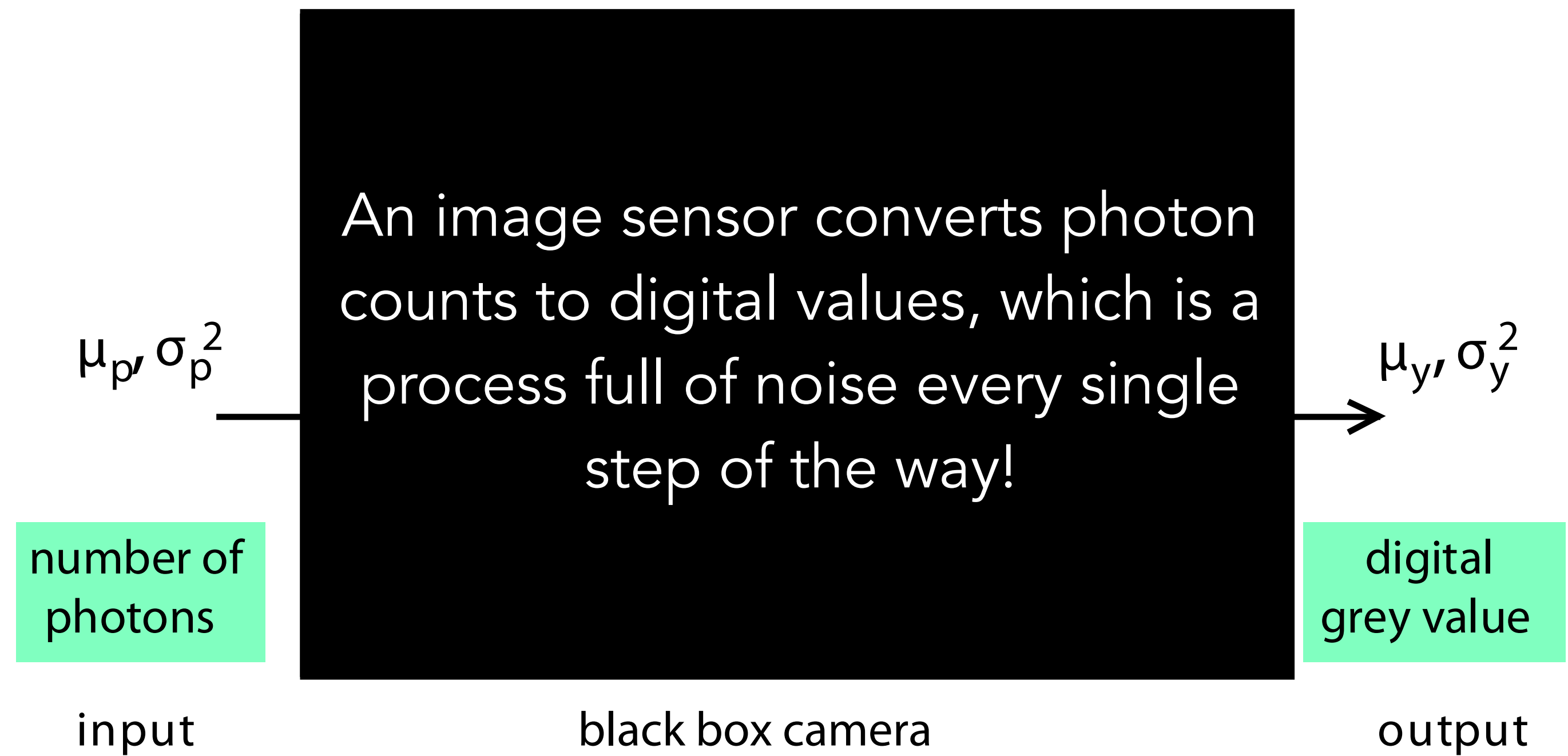
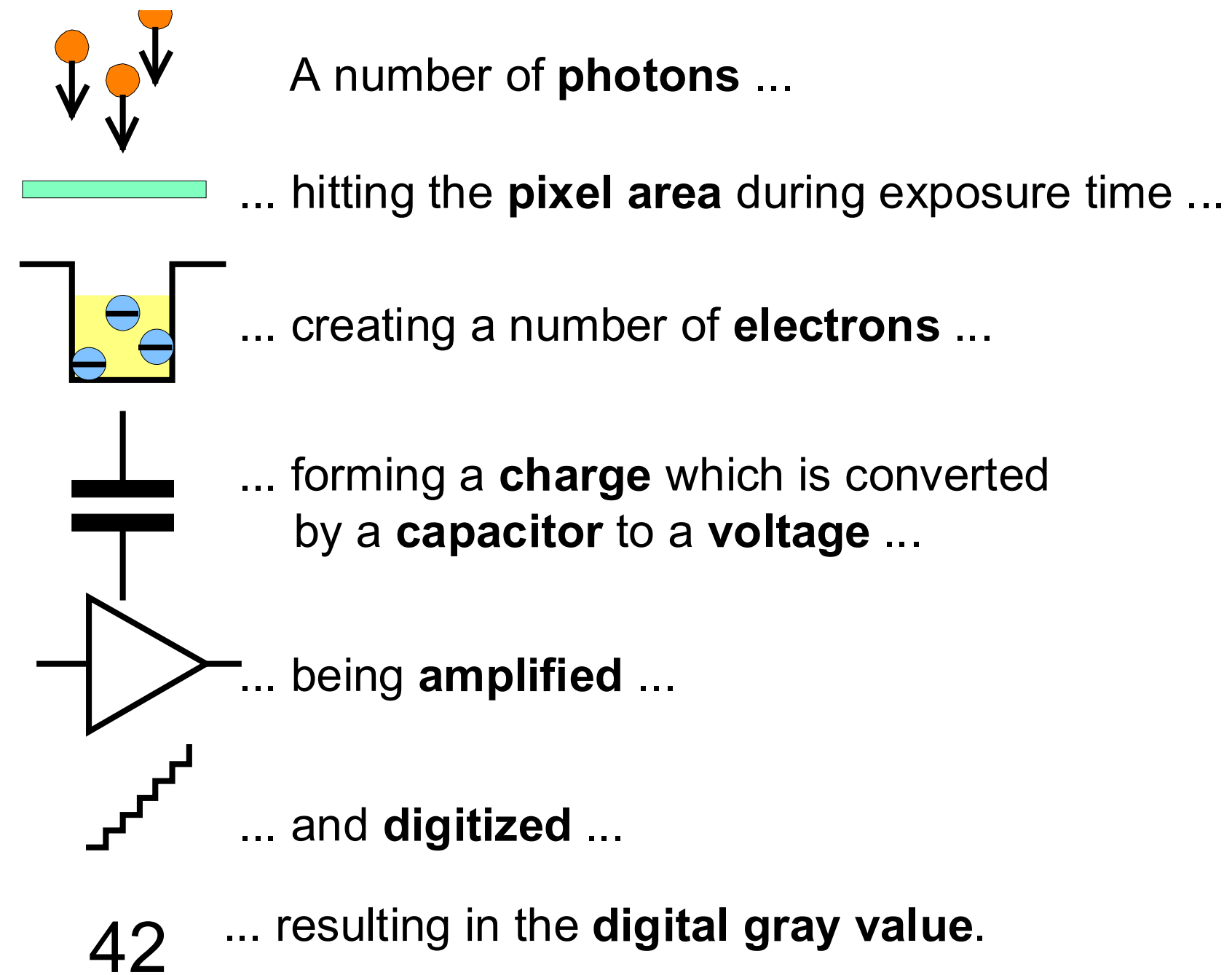


Converting charges to bits

# Cross Section

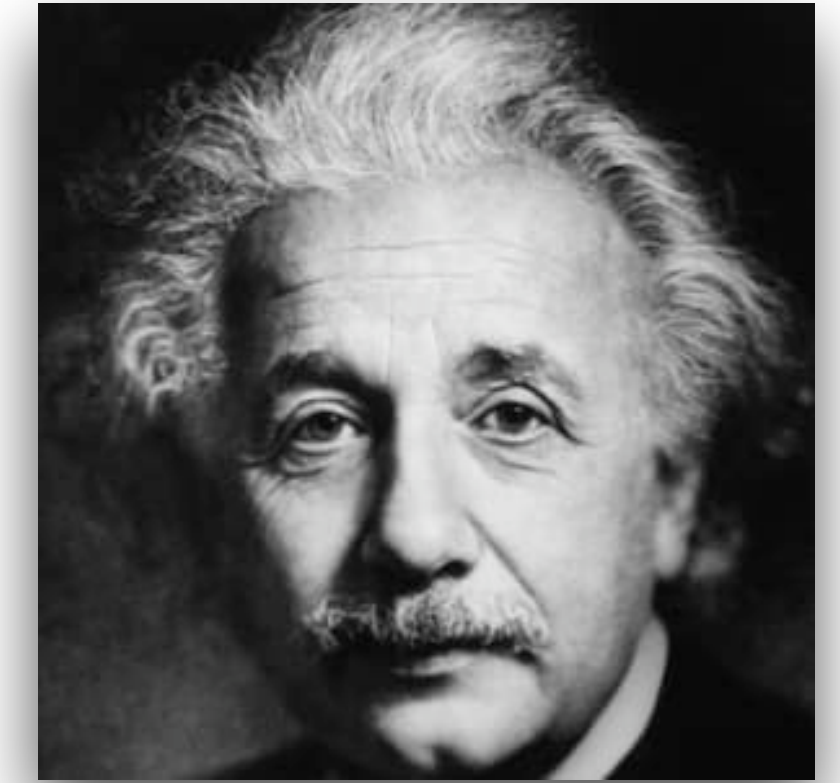


# The Big Picture



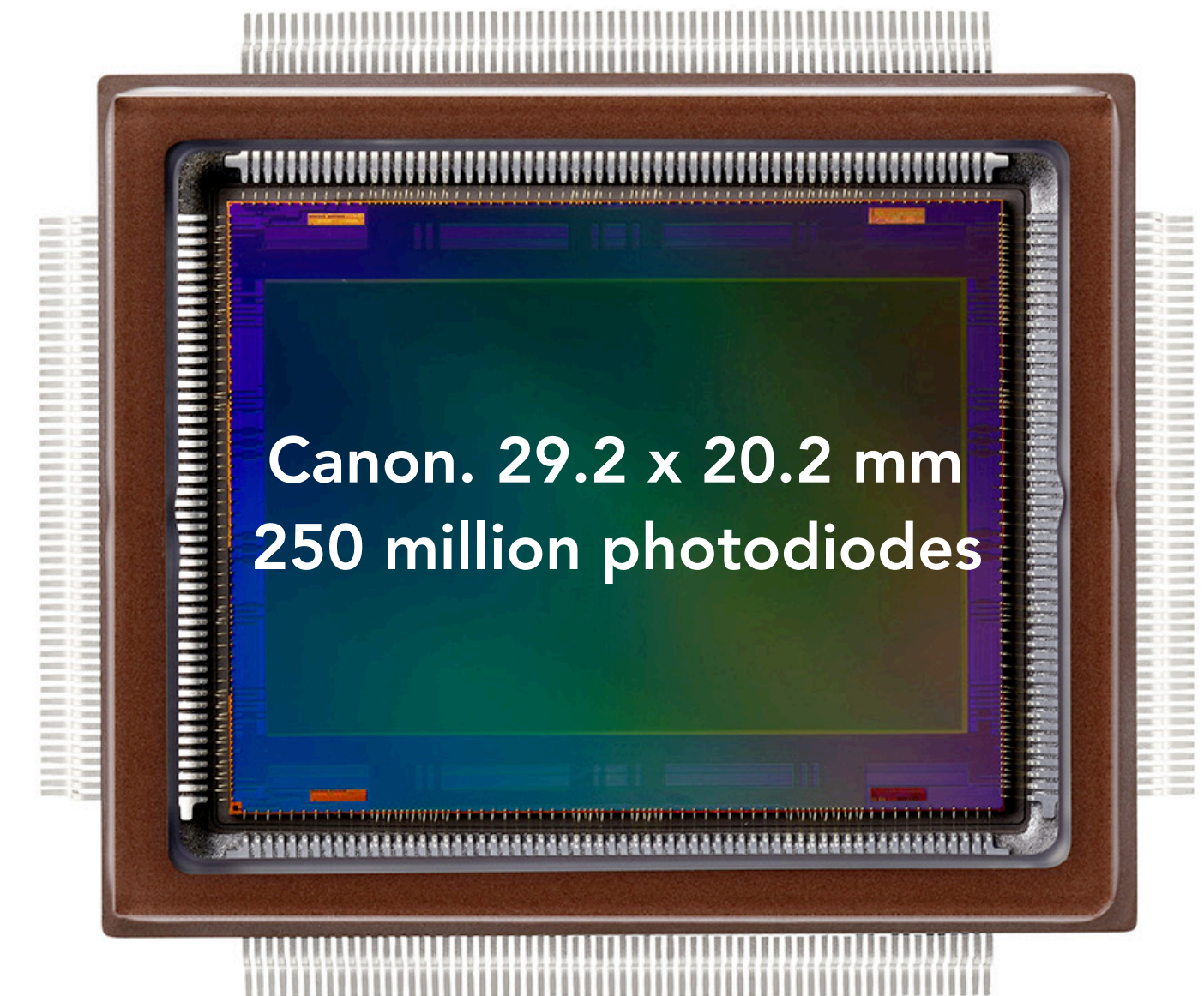
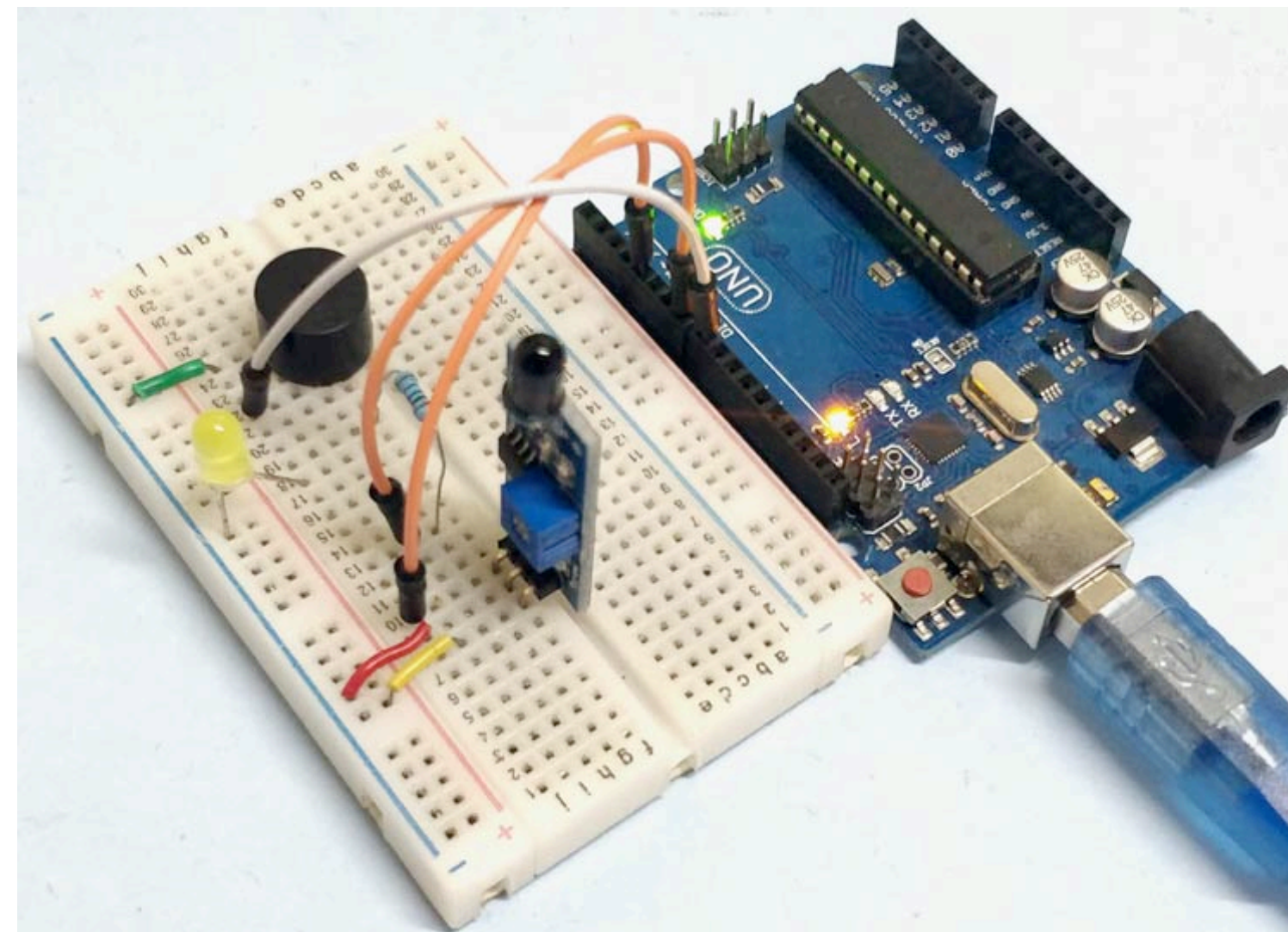
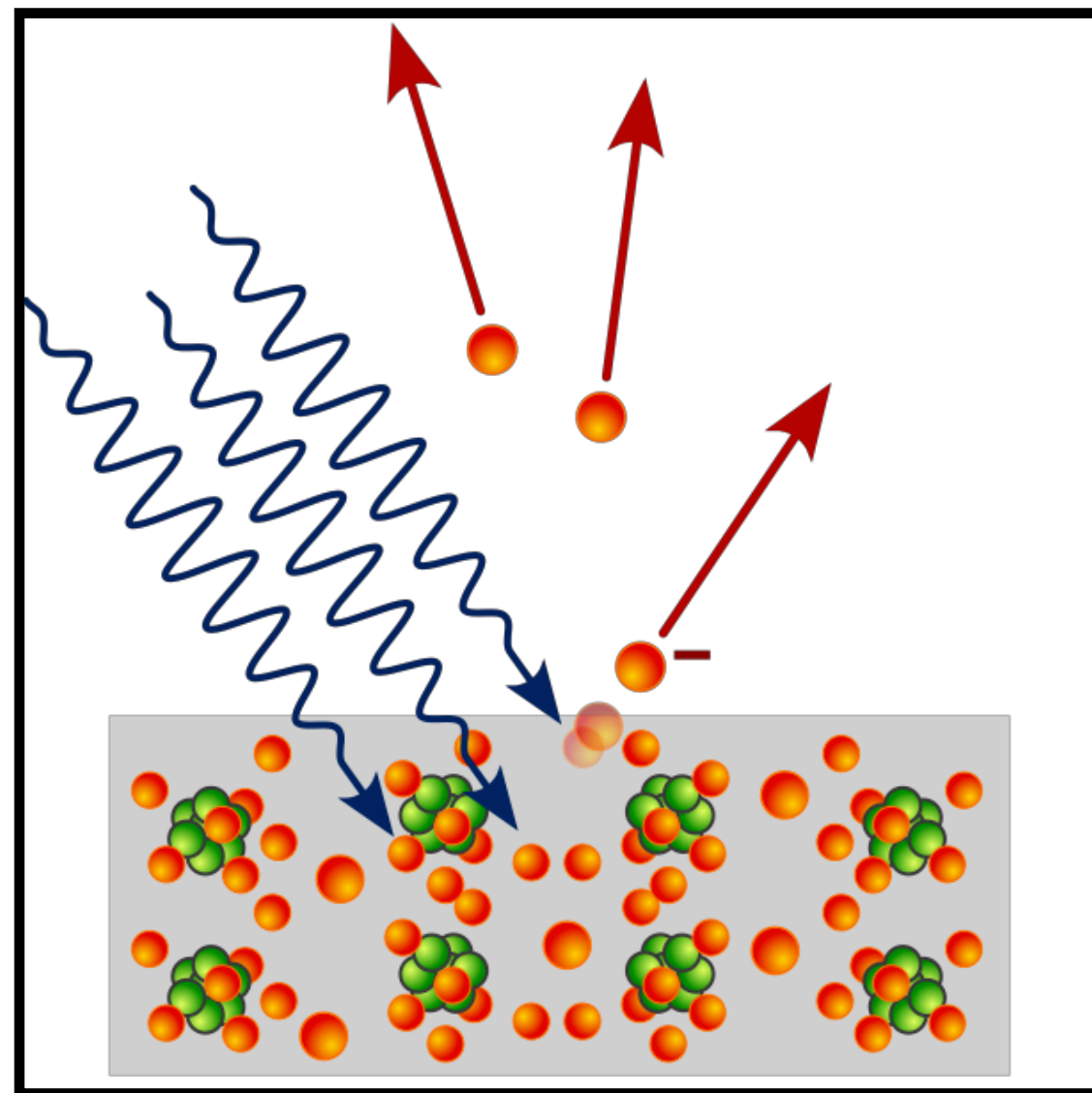
# Pixel: From Photons to Charges

# Photoelectric Effect



When a photon strikes material, an electron *may* be emitted.

- Cameras use silicon
- Quantum nature of light matter interaction.





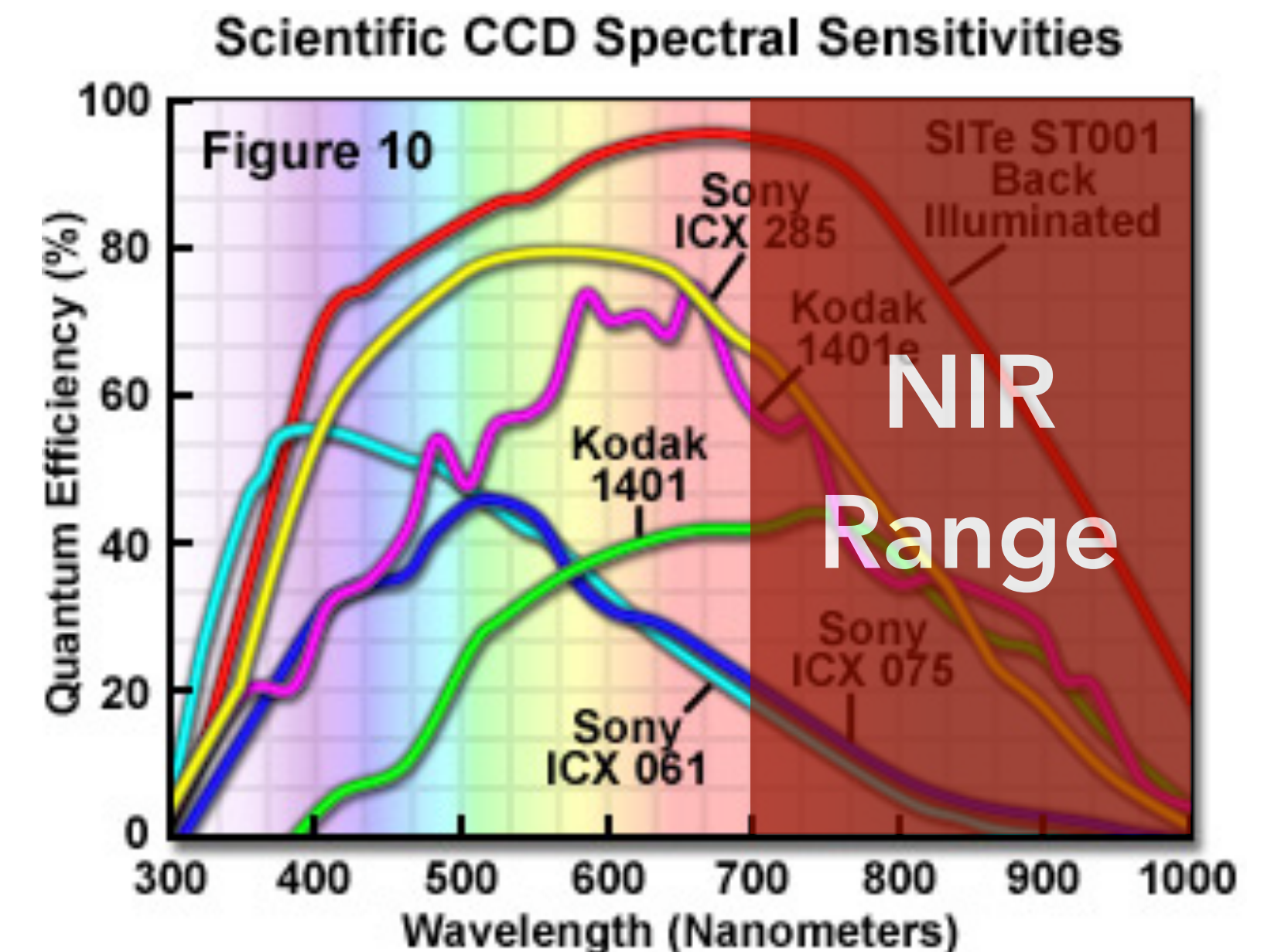
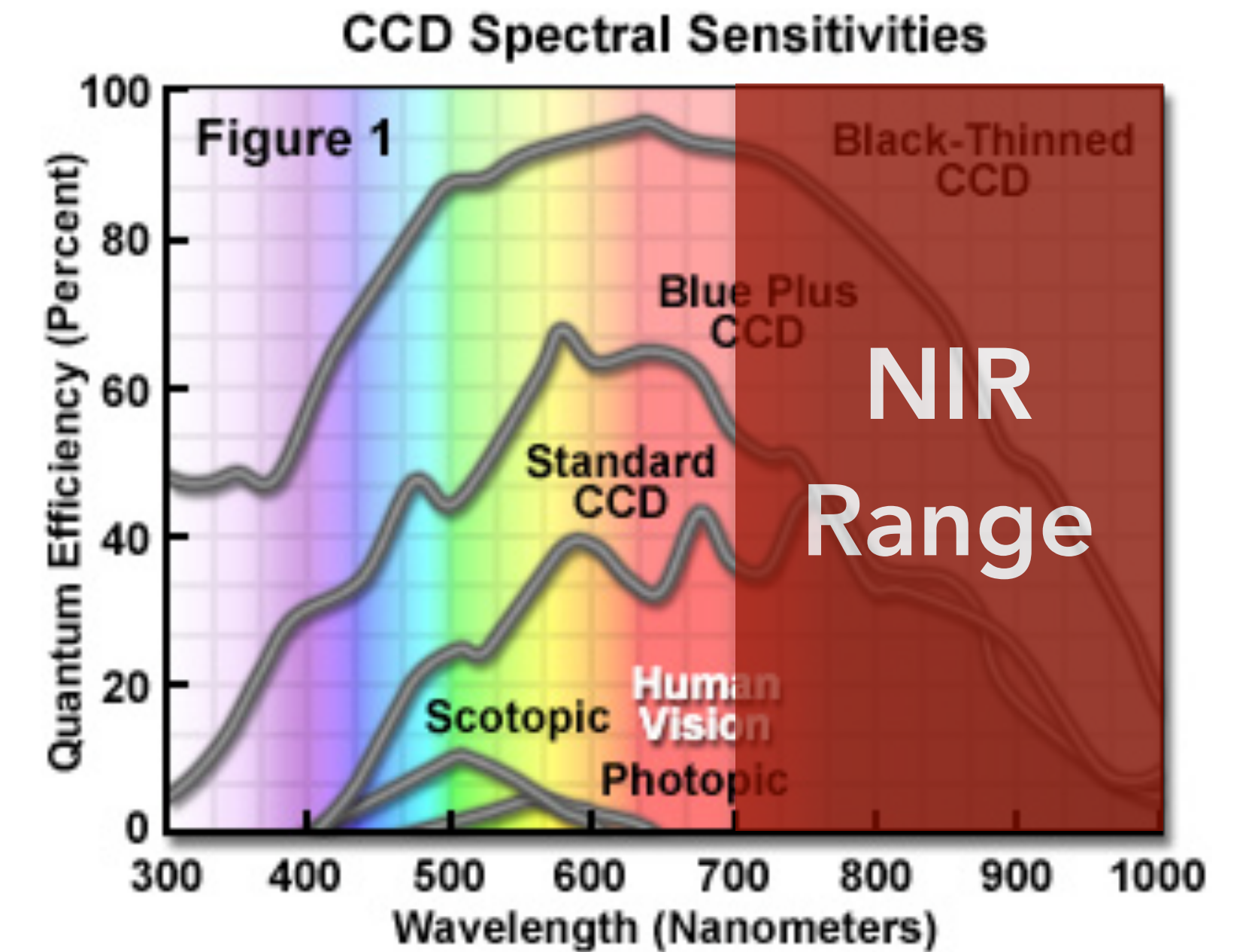
# Quantum Efficiency

Electron emission from photon is probabilistic, which depends on the **quantum efficiency** (QE) of the material. QE is wavelength-dependent (just like about everything in optics).

## Human QE vs. Camera QE

- Human QE is much lower than camera sensor QE.
- Human eyes are not sensitive to IR and UV light, but camera sensors are.

$$QE = \frac{\# \text{ of electrons}}{\# \text{ of photons}}$$



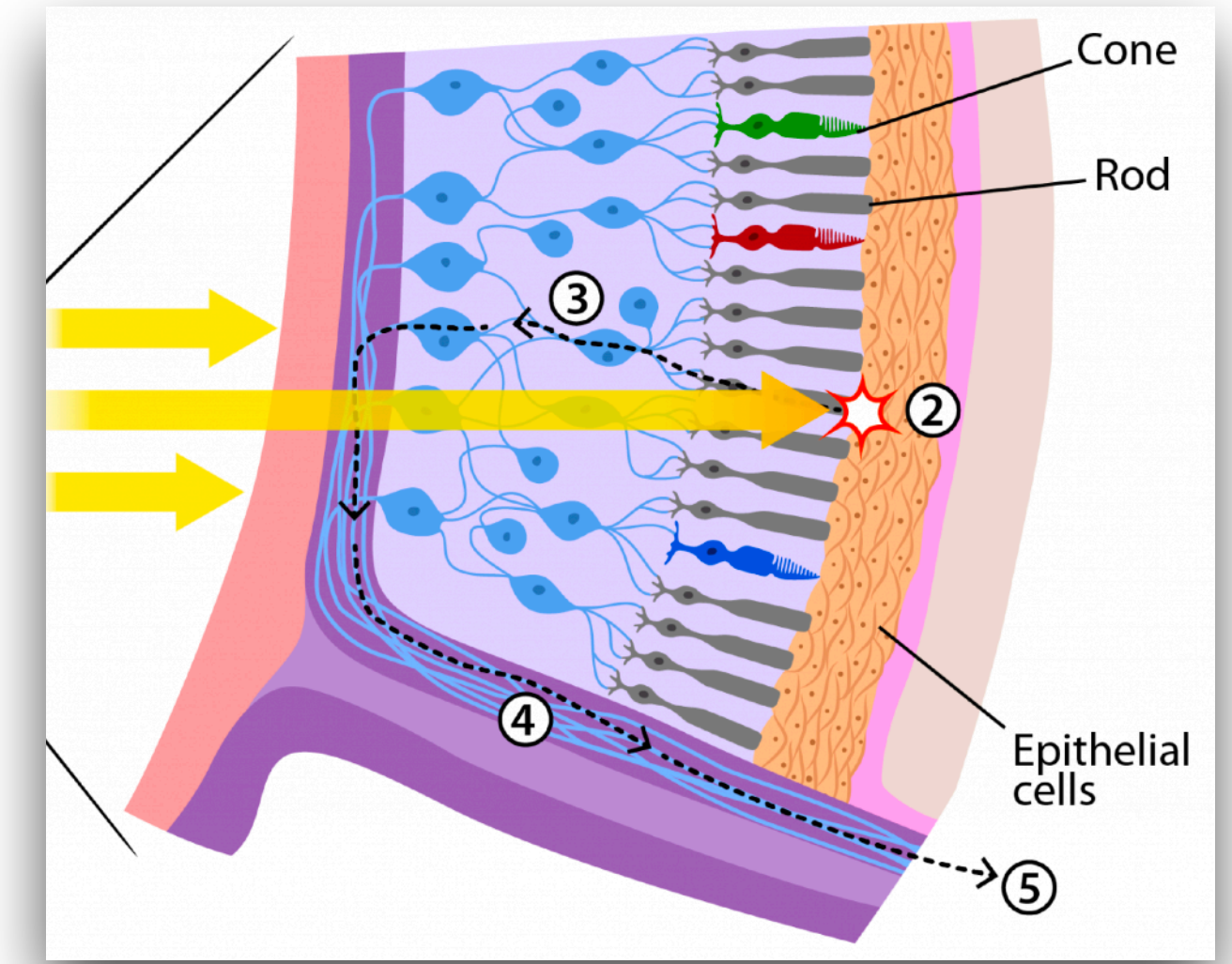
# Other Uses/Forms of Light-Matter Interactions



Solar panels



Photosynthesis



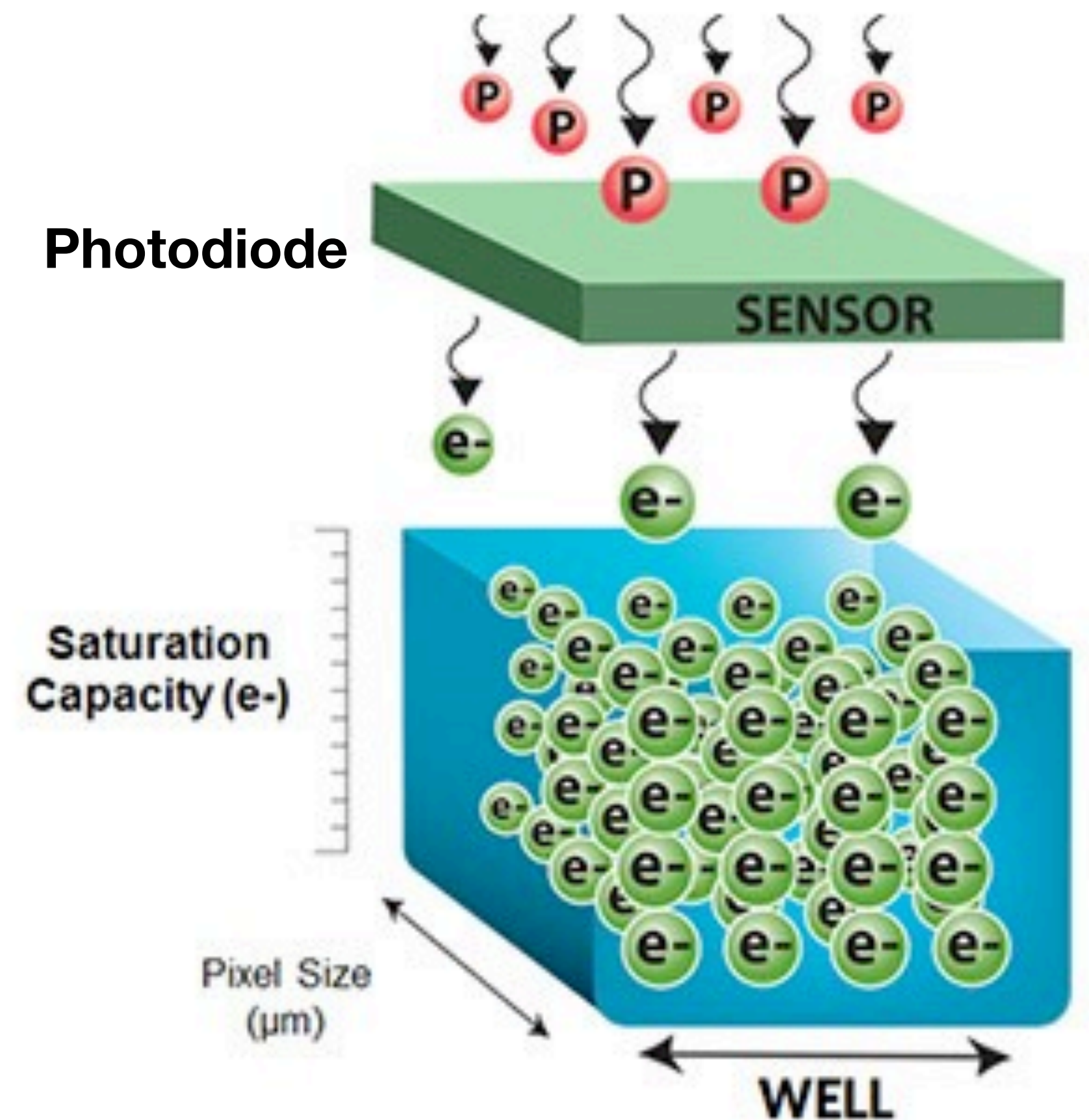
Photoreceptor cells

# Pixel Well

Think of a pixel in a sensor as a well that holds the emitted electrons.

What determines the number of electrons going into the well:

- light power (dictates the number of photons per unit time) and
- exposure time (integration time).



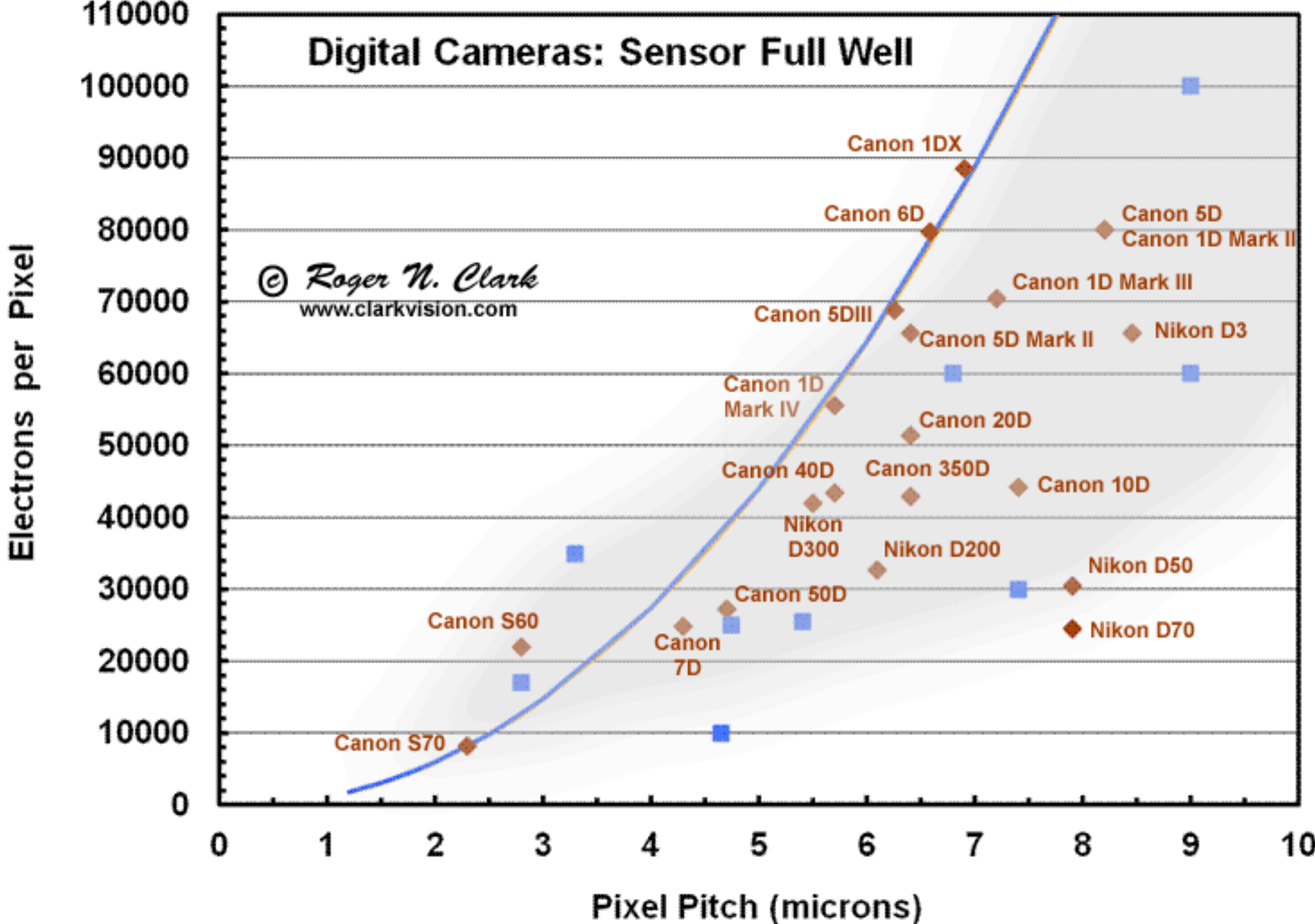
# Full Well Capacity

The **full-well capacity**, the max amount of electrons that can be held, depends on pixel size.

More electrons than the capacity **saturates** the well.

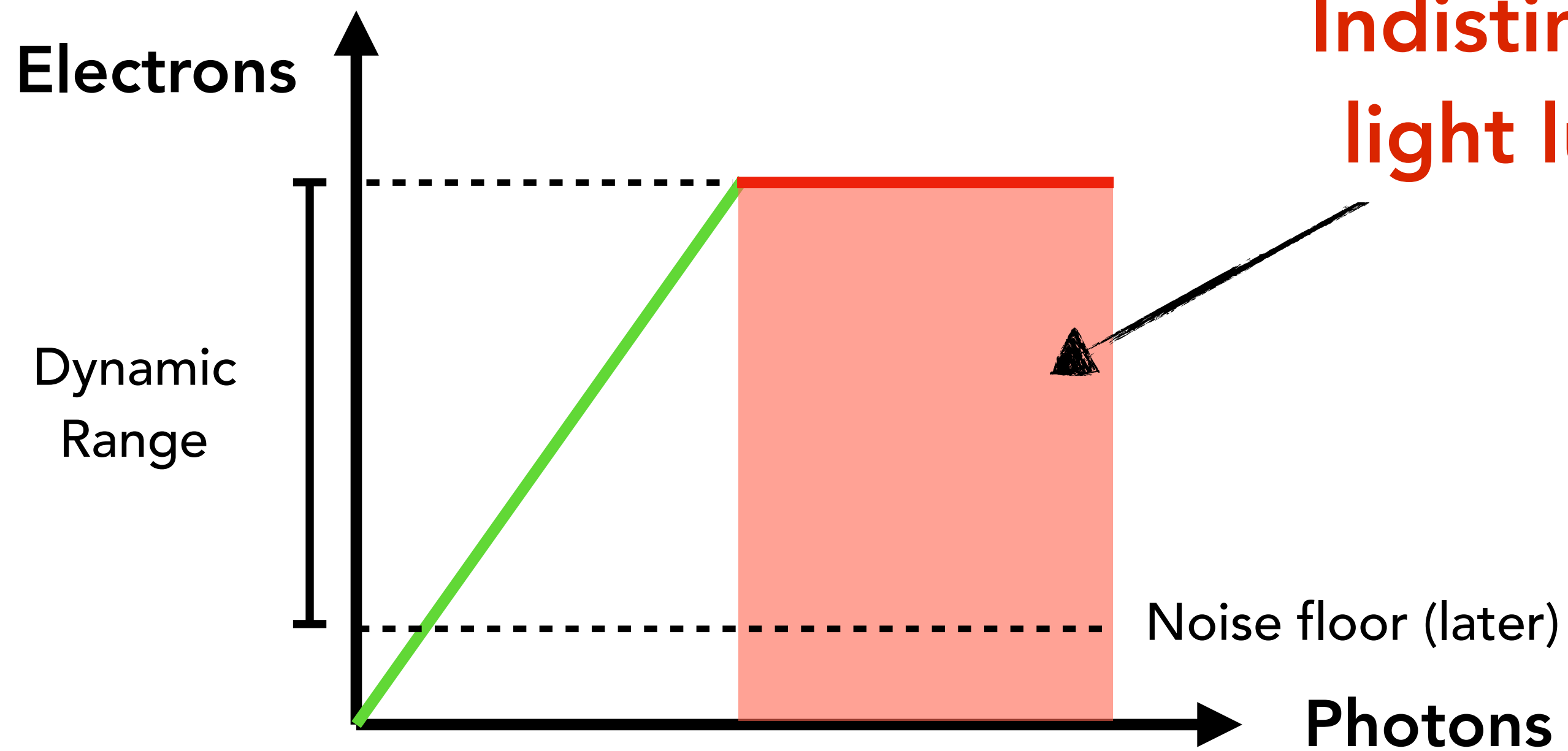
A saturated well doesn't store more electrons.

## Quadratic. Why?

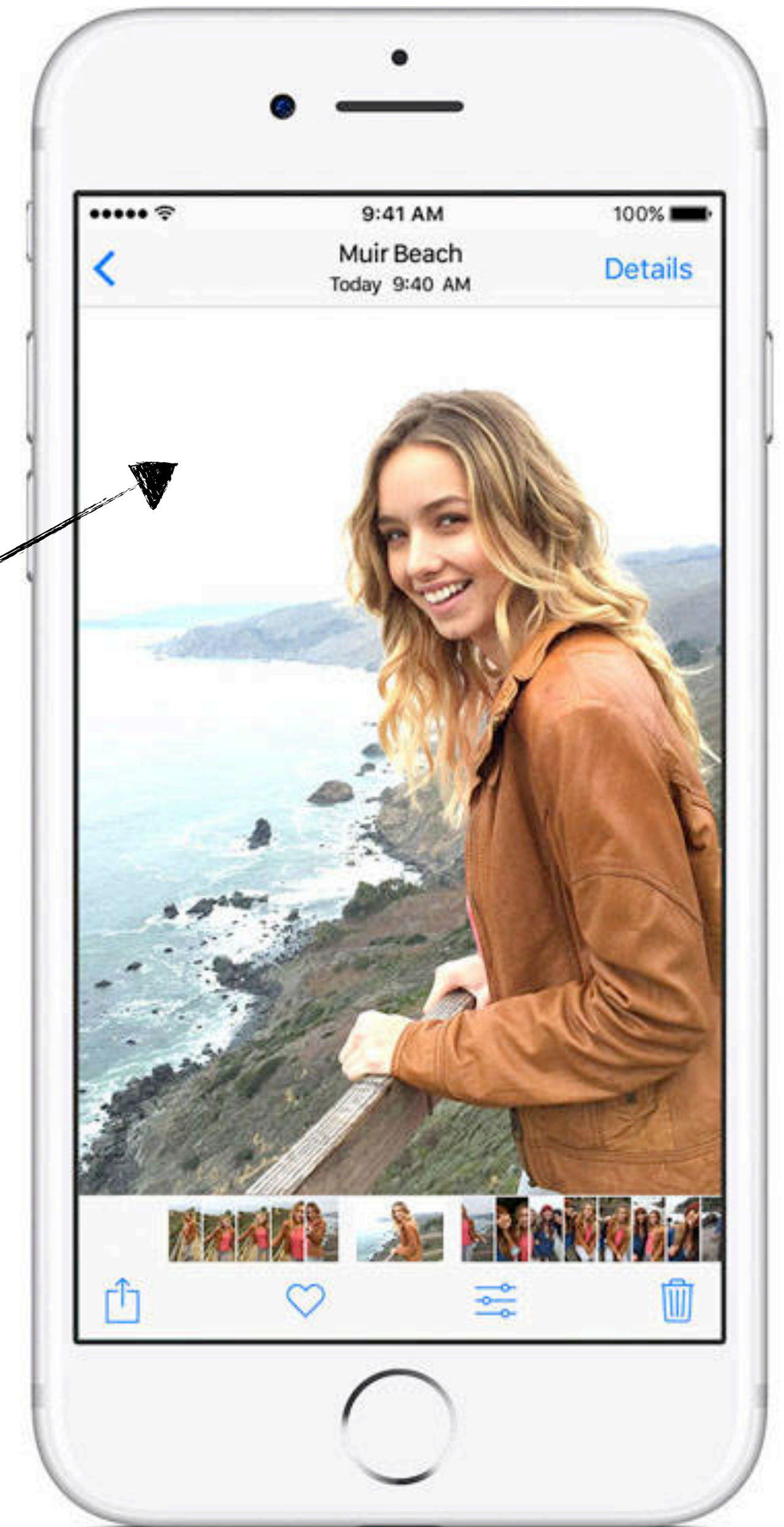


# Dynamic Range (Informal Treatment)

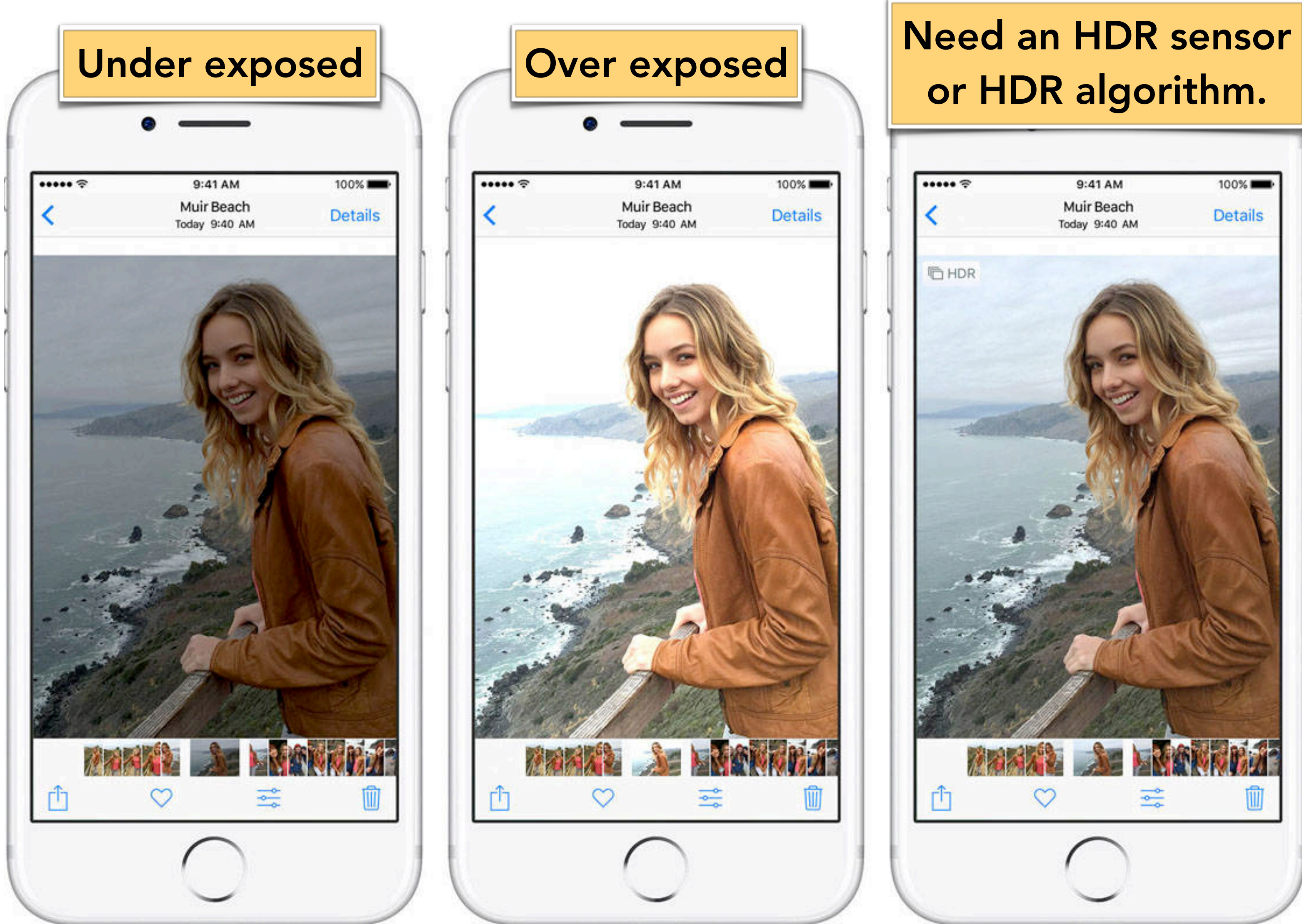
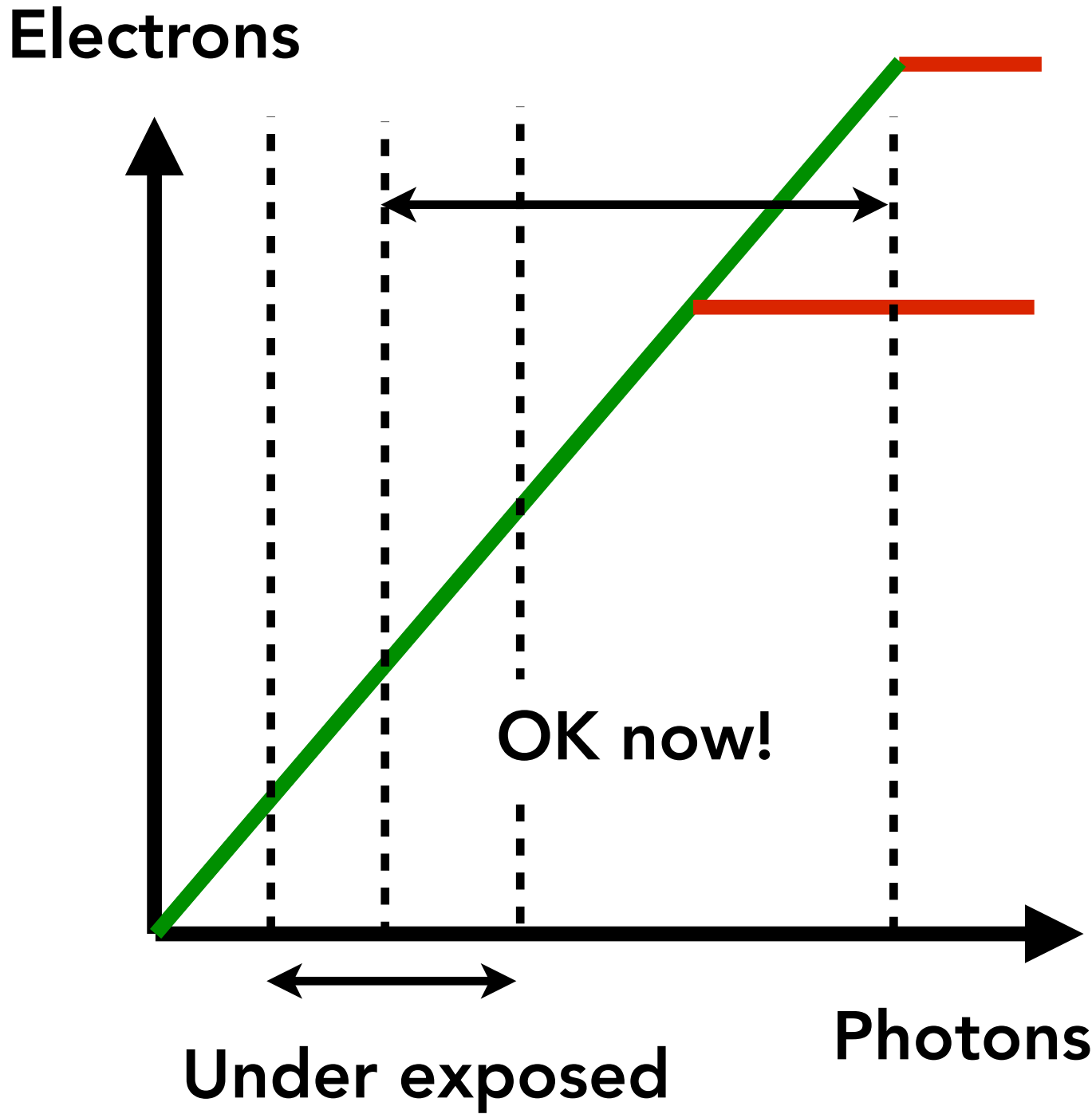
A larger well capacity leads to a higher sensor dynamic range, which can capture/accommodate larger differences in scene luminance.



Indistinguishable light luminance



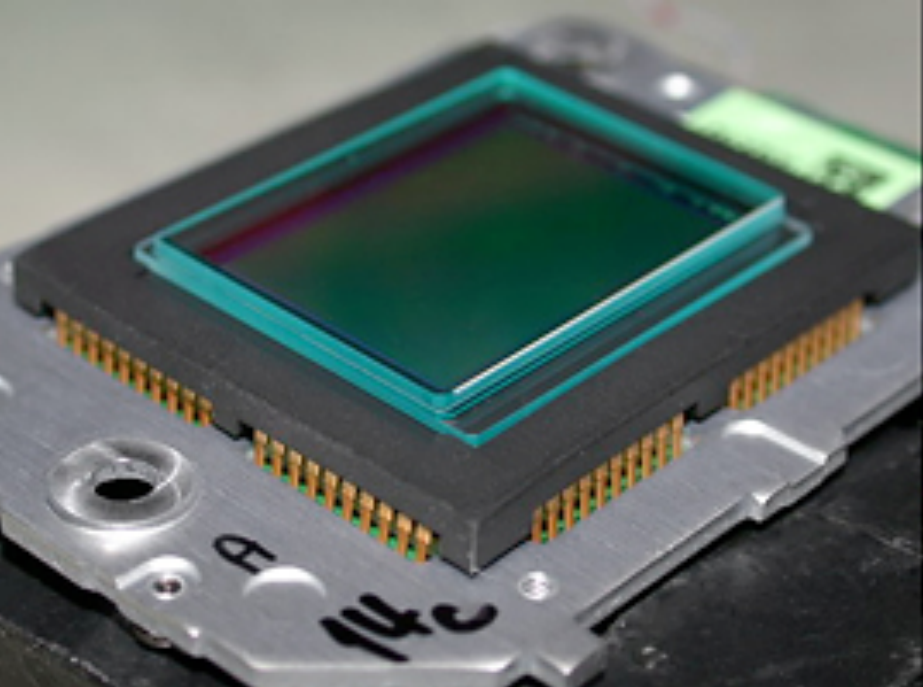
# Dynamic Range (Informal Treatment)



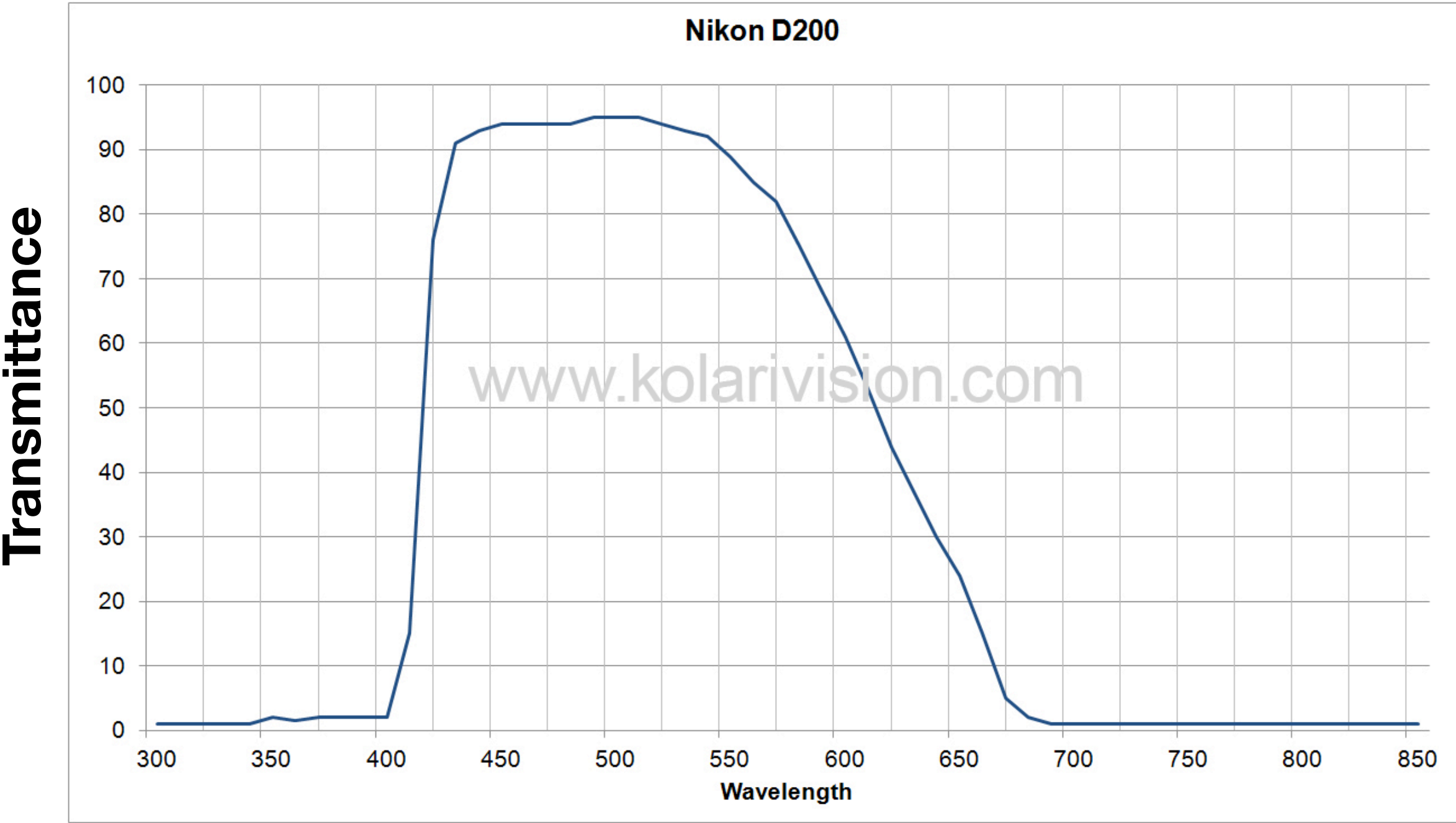
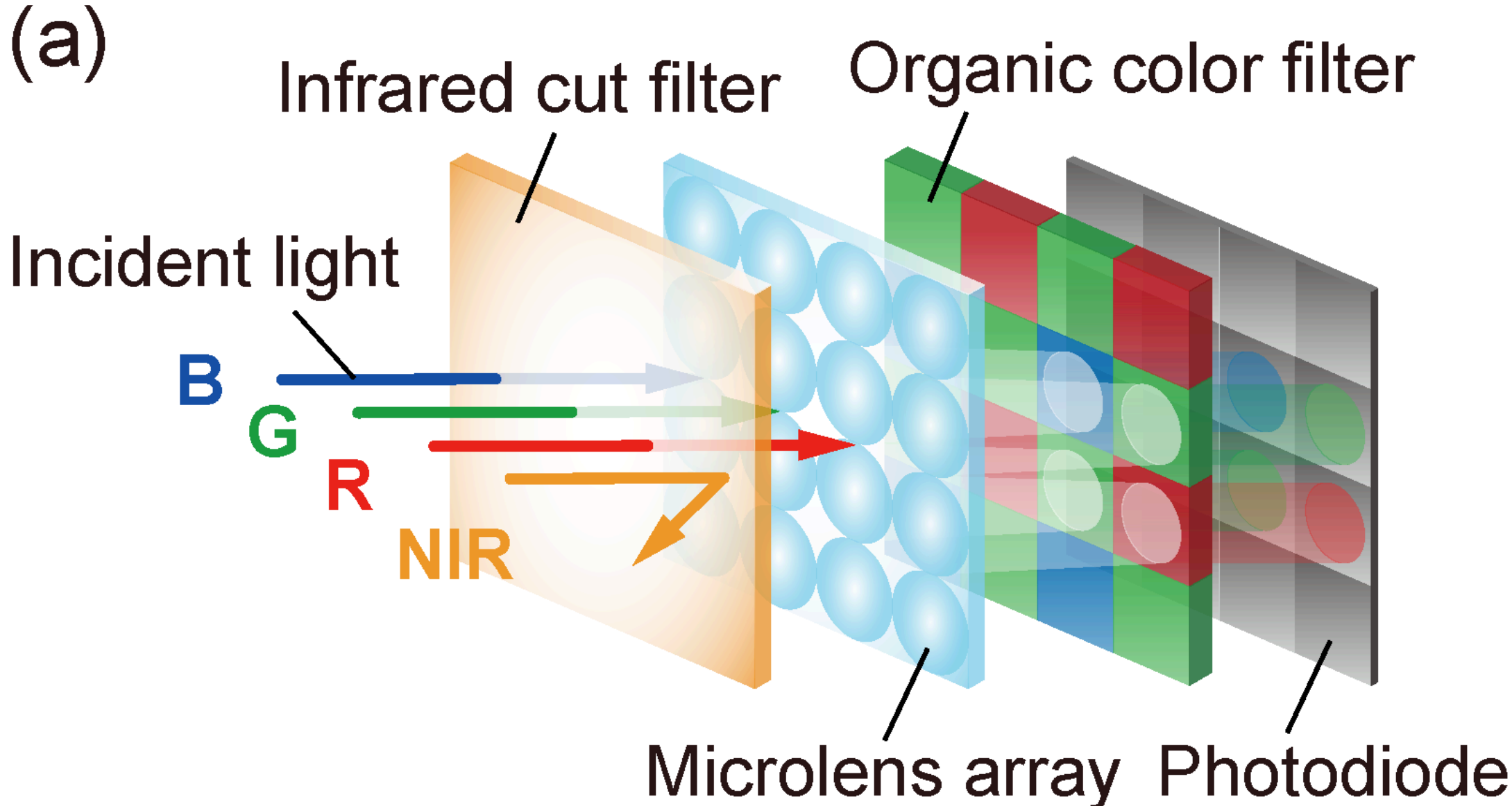
# Optics in Sensor

# Near IR/UV Filters

Remove IR/UV light. Low transmittance in IR/UV ranges.



Nikon D200



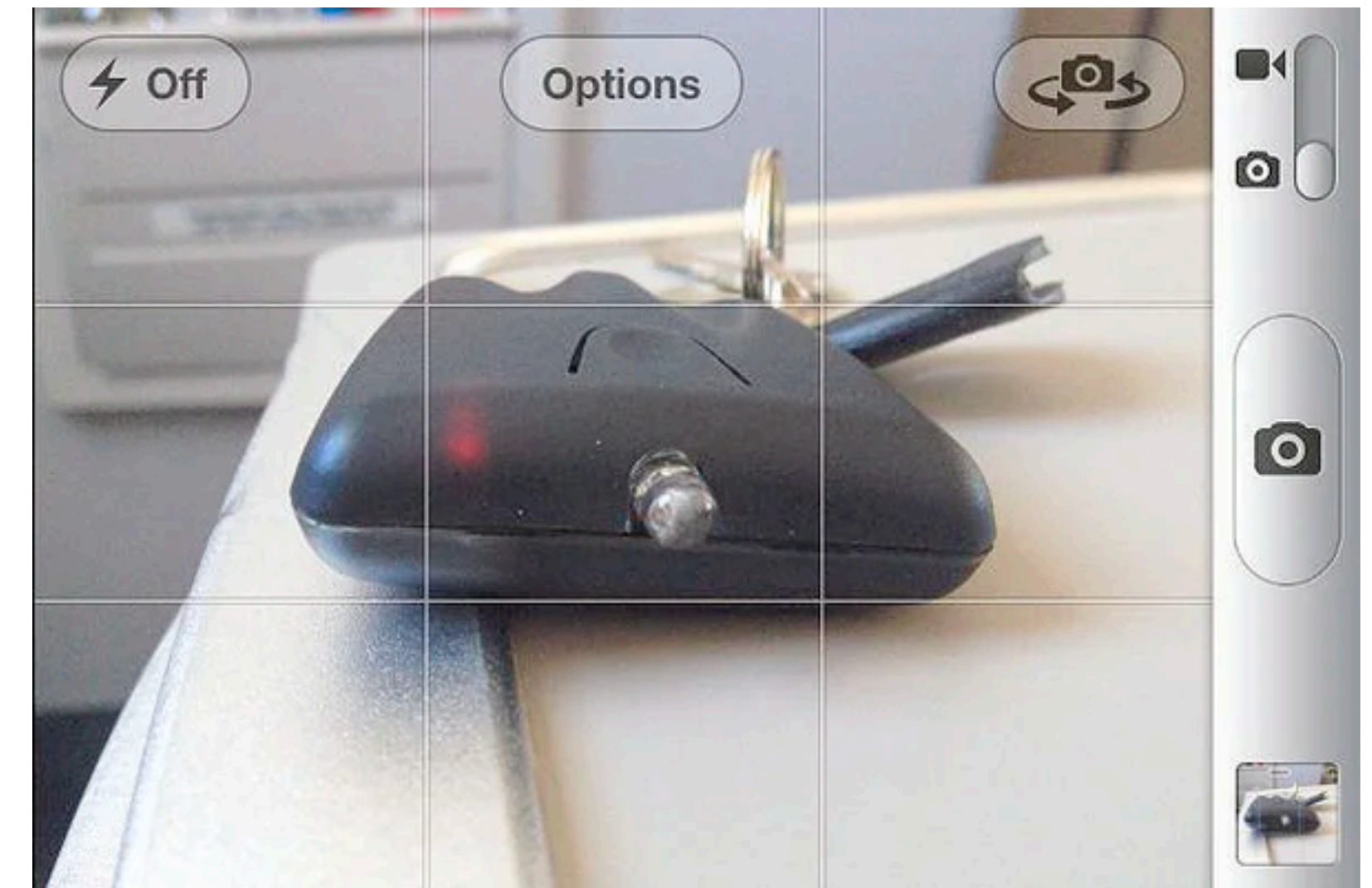
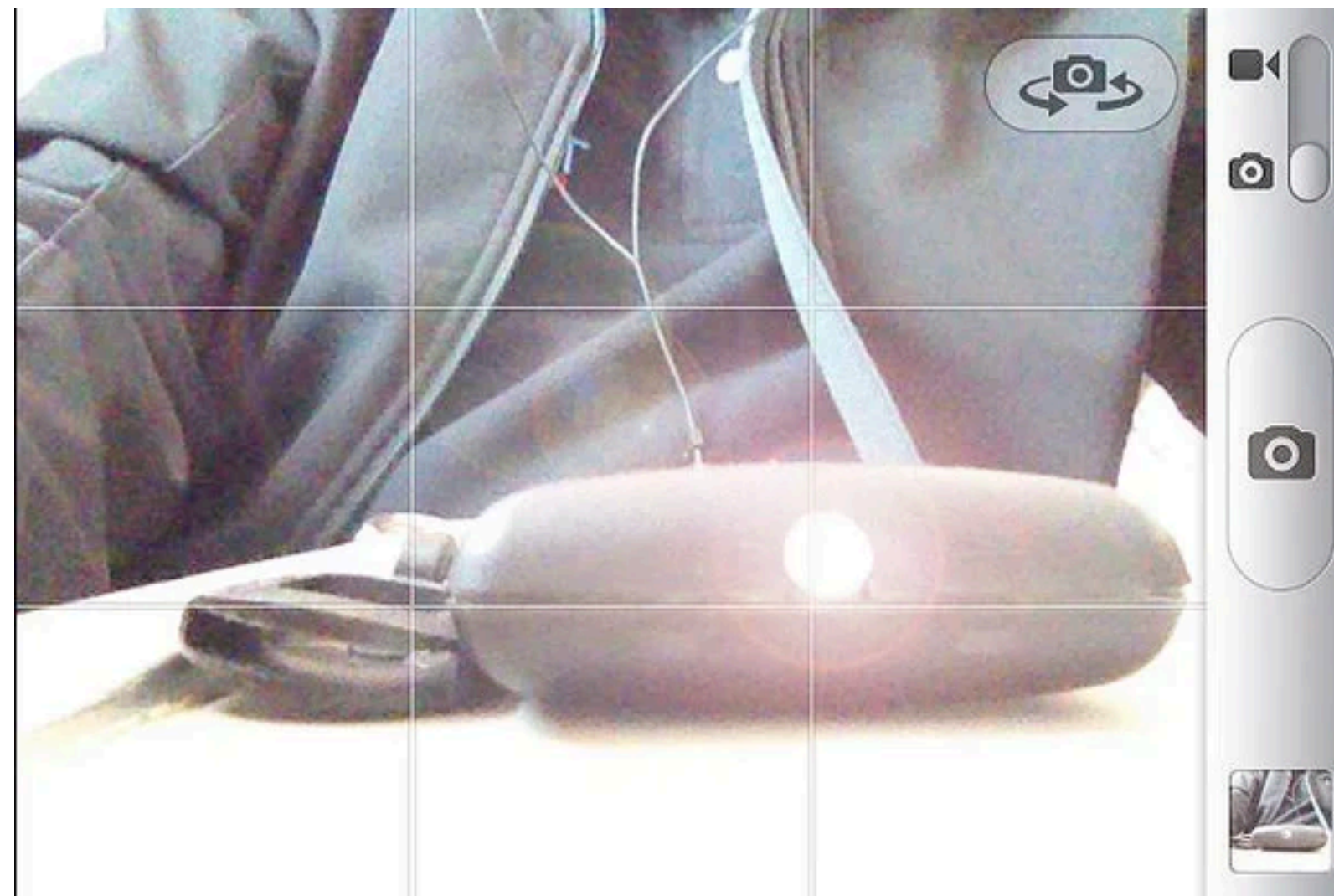


# Near IR/UV Filters

Not all cameras do. What happens if you remove the NIR filter?

- See light that human eyes can't see. NIR photography (false color!).
- Thermographic cameras intentionally sense IR radiations to measure temperature.

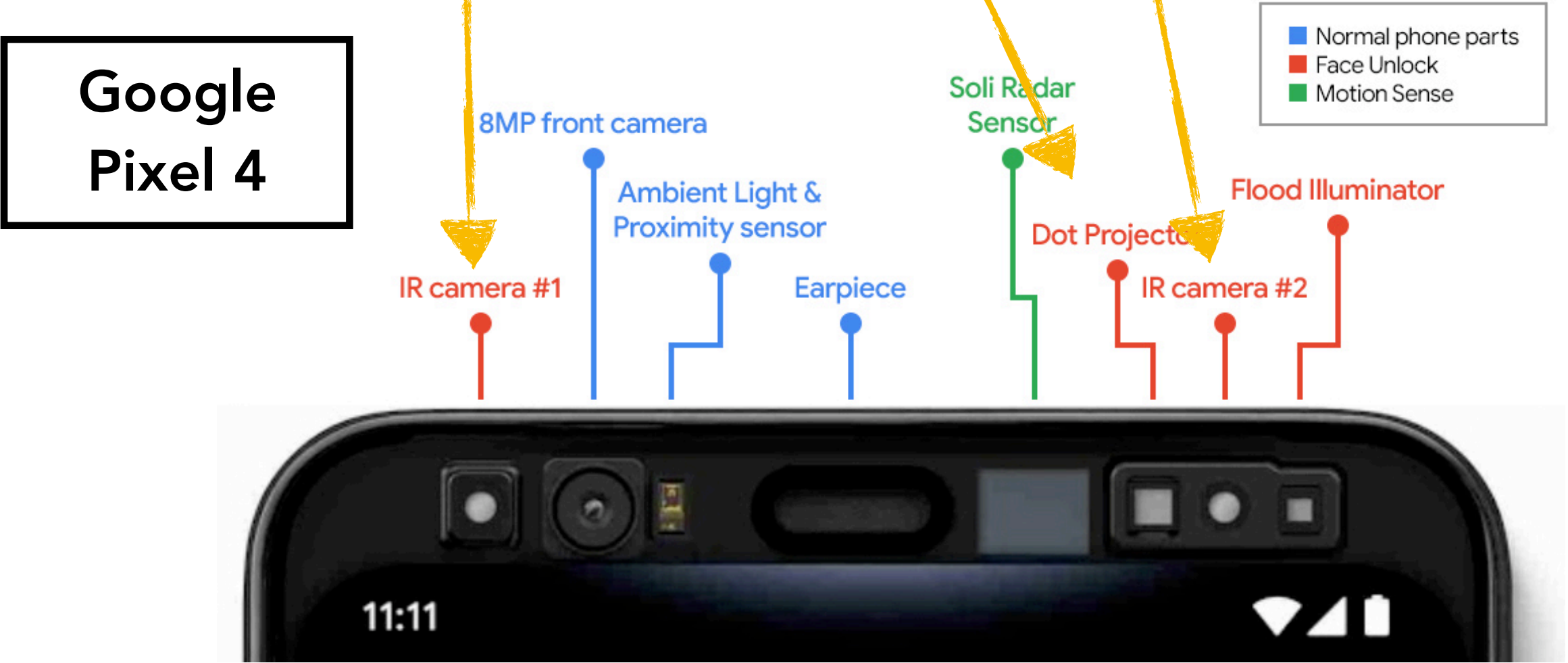
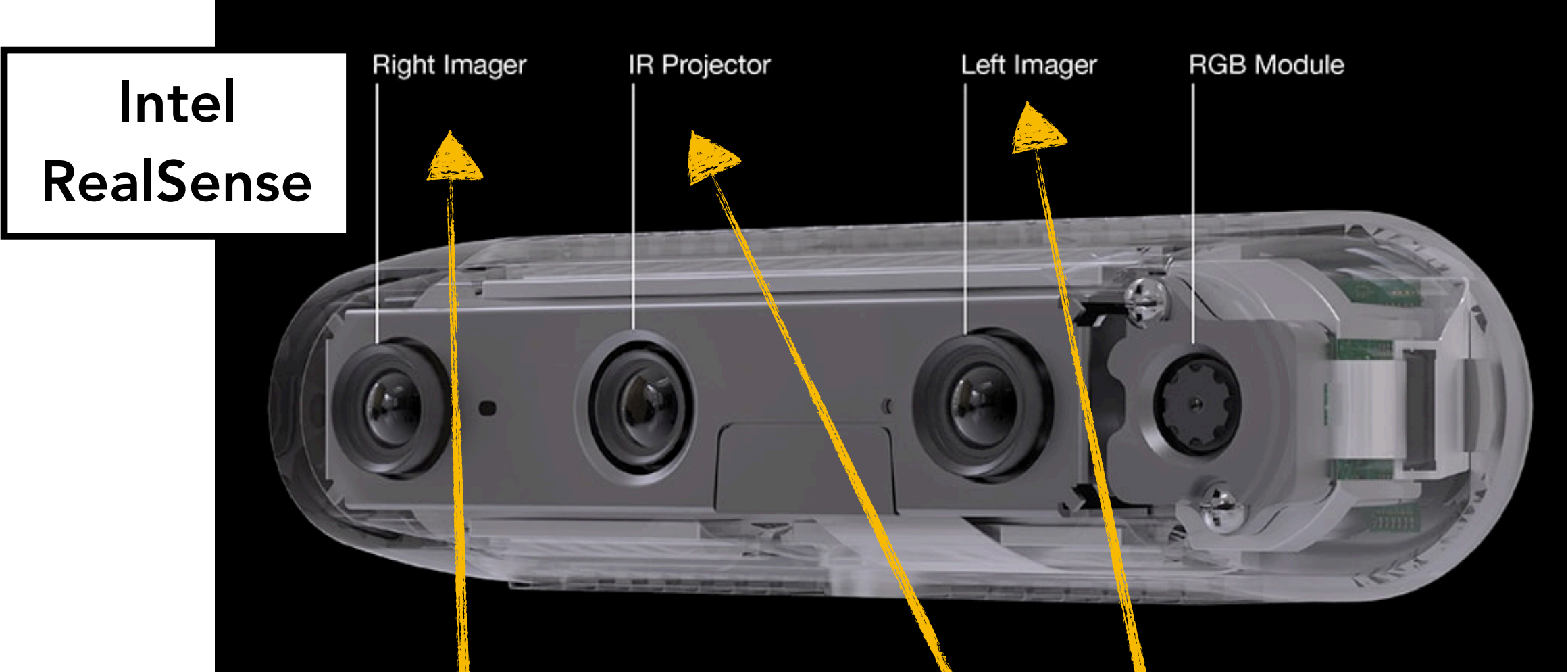
Pointing a TV controller to the front and back cameras of iPhone 4.



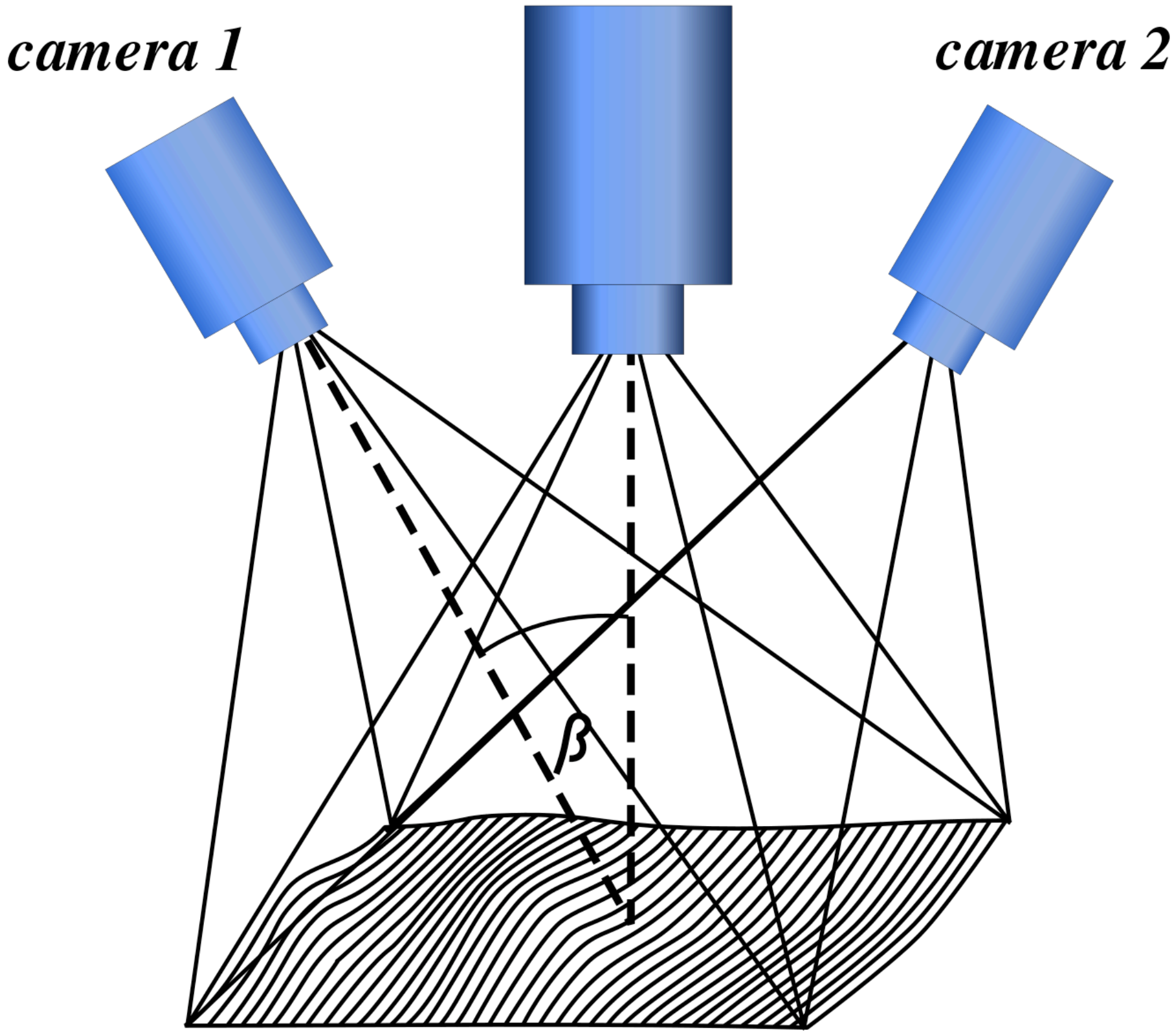




# IR Sensor: Proximity and Depth (Later)



Depth from stereo (one depth cue of human eyes)  
*projector*



<https://www.intelrealsense.com/depth-camera-d435/>

<https://ai.googleblog.com/2020/04/udepth-real-time-3d-depth-sensing-on.html>

[https://en.wikipedia.org/wiki/Structured-light\\_3D\\_scanner](https://en.wikipedia.org/wiki/Structured-light_3D_scanner) 20

# Fill Factor (FF)

Fill factor is the percentage of the photosensitive area (i.e., photodiode) inside a pixel.

Much of the pixel area is occupied by circuitries other than the photodiodes, e.g., analog to digital conversion, read-out, wires, etc.

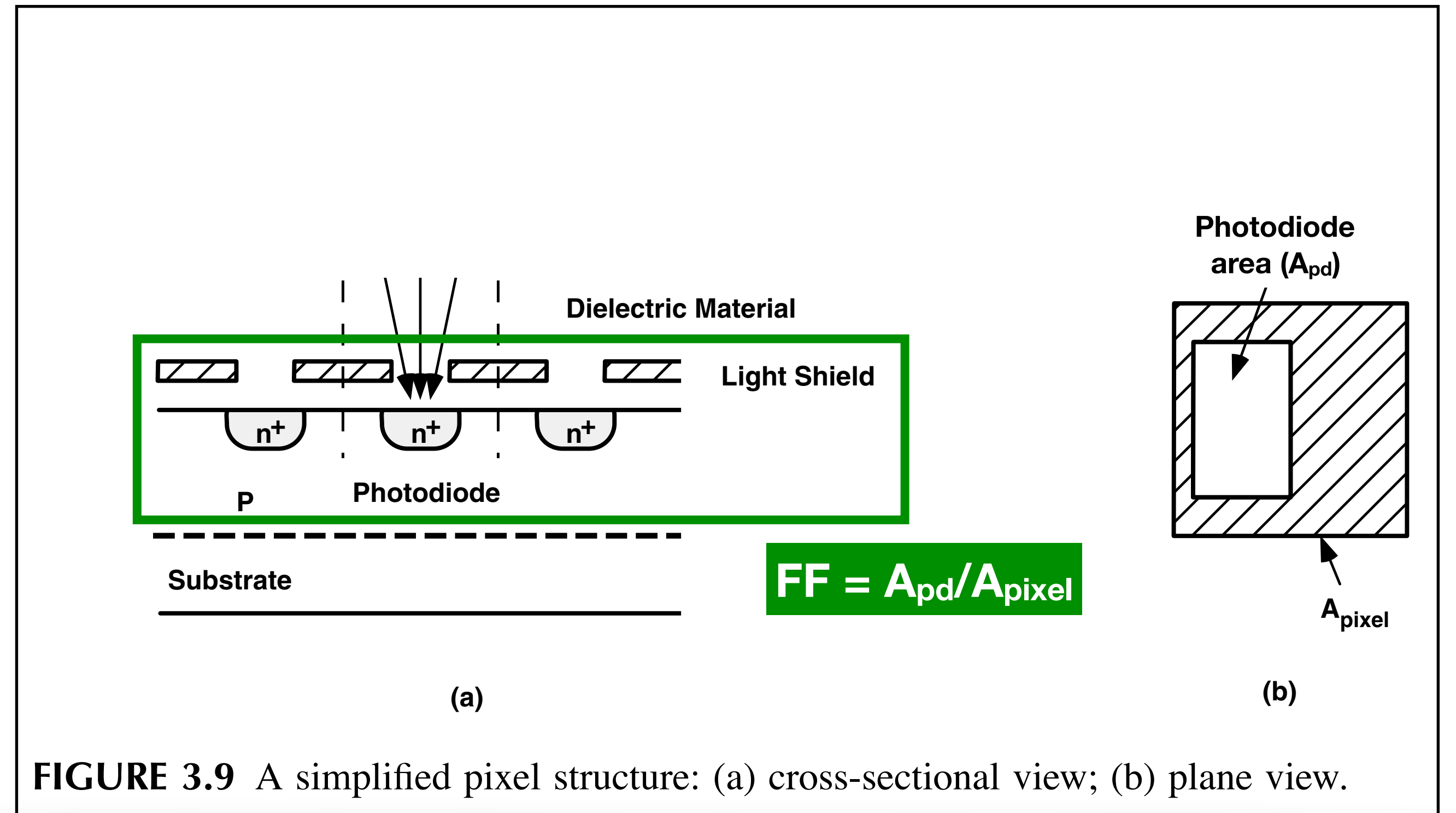


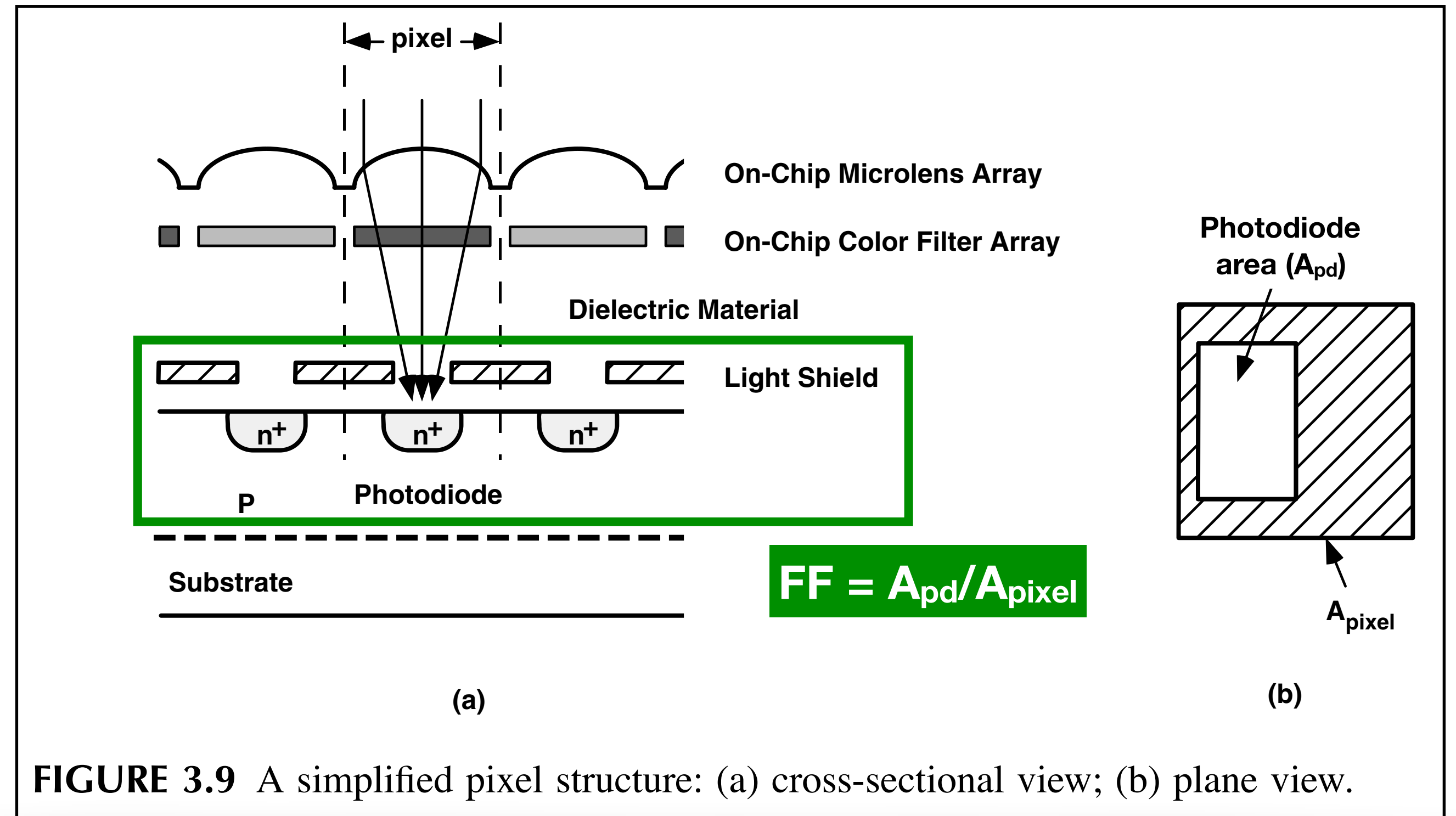
FIGURE 3.9 A simplified pixel structure: (a) cross-sectional view; (b) plane view.

# Fill Factor (FF) and Microlenses

Fill factor affects the sensor dynamic range (higher FF leads to "larger" pixels).

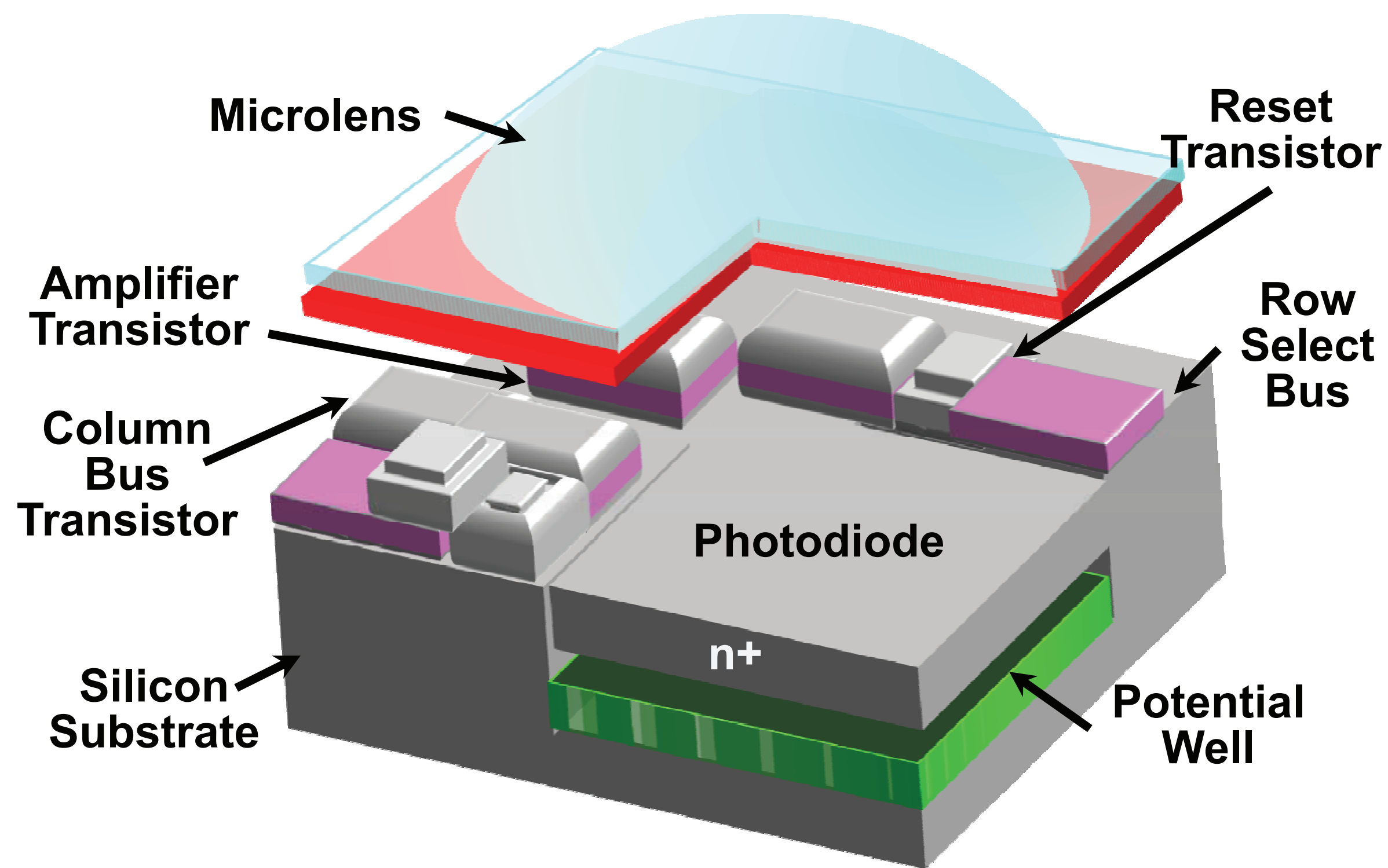
Per-pixel microlenses help significantly increase the FF.

Organic PDs can reach almost 100% FF.



# Anti-Aliasing by Micro-lenses and Pixels

Anatomy of the Active Pixel Sensor Photodiode



Recall: pixels perform spatial sampling, introducing aliasing.

The micro-lenses (and the pixels under) inherently perform a spatial 2D box convolution (which is a low-pass filtering, acting as an anti-aliasing filter)

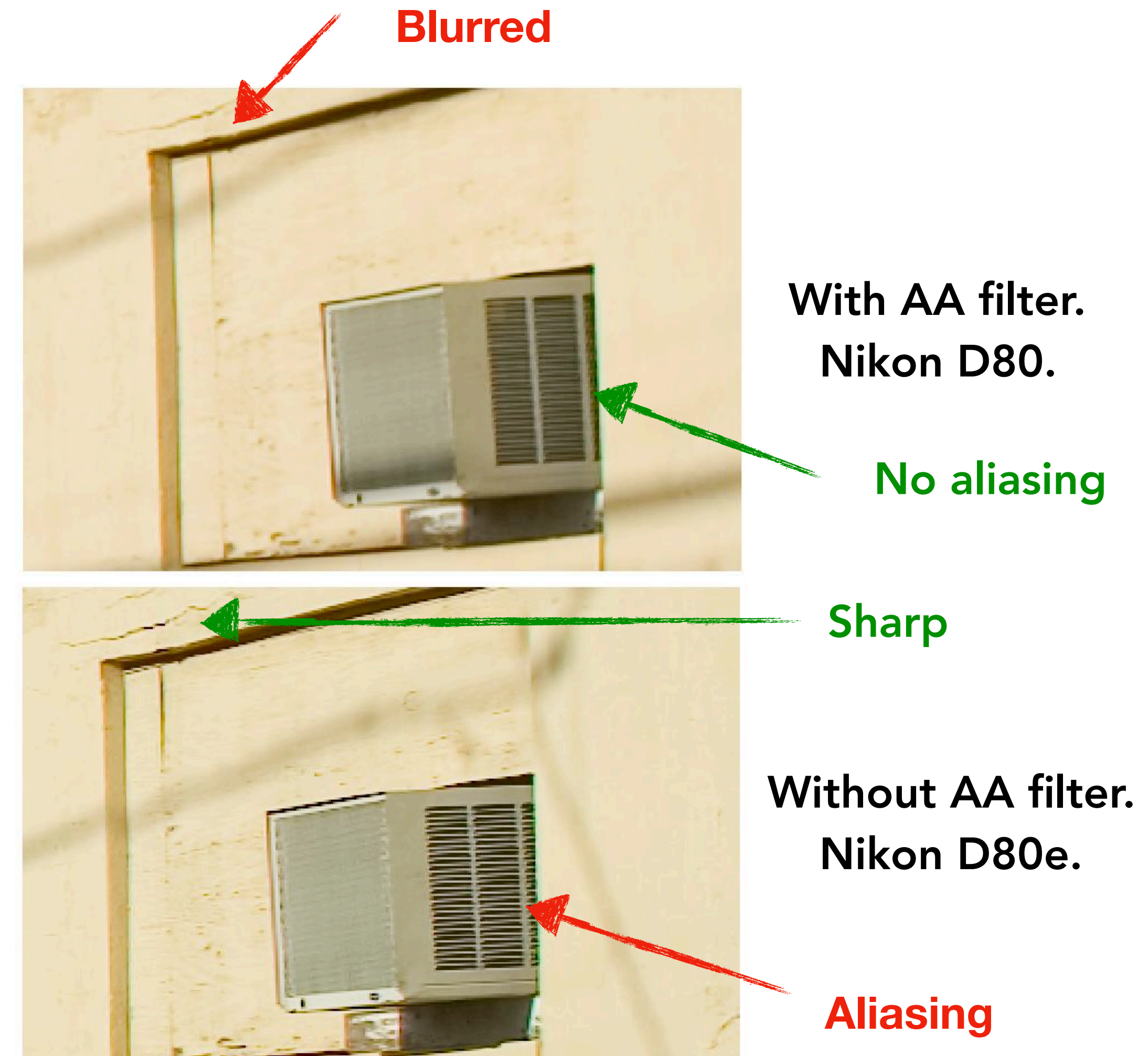
- followed by a 2D sampling, where a sample is taken at each pixel center.

# Anti-Aliasing Filter

Anti-aliasing by micro-lenses and pixels is not enough to combat aliasing.

Use **anti-aliasing filters** (low-pass filter) to remove high spatial-frequency component in light, effectively **blurring** the image.

Removing AA filter increases sharpness at the cost of aliasing.

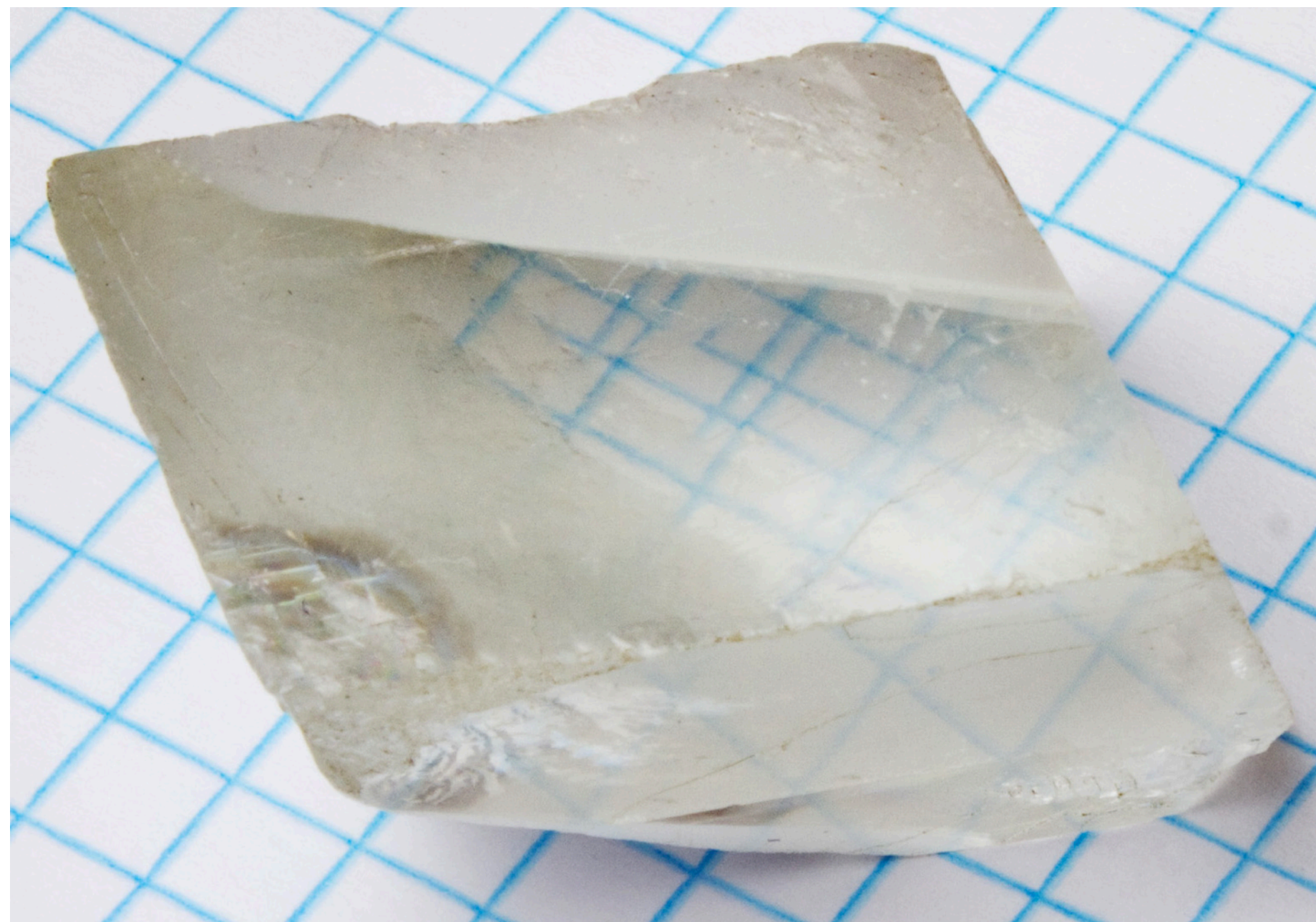




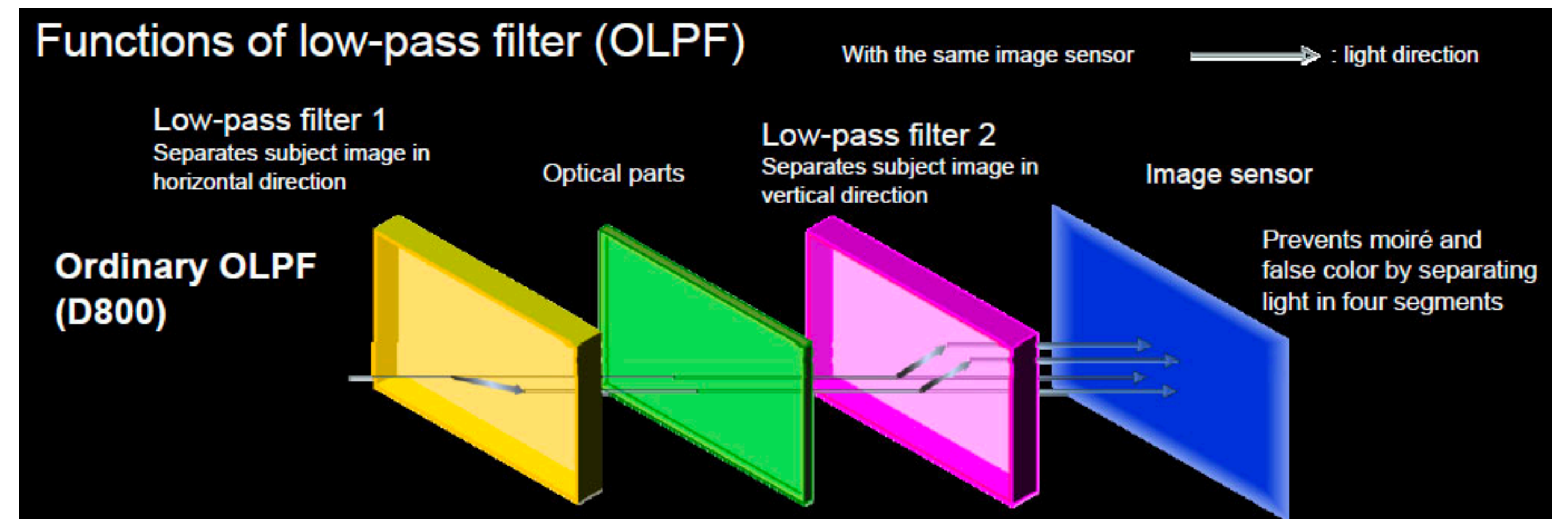
# Anti-Aliasing Filter

Two layers of birefringent material separating one point into four pixels.

One birefringent layer



Combining two birefringent layers



# From Charges to Digital Values

# From Electrons to Pixel Values

Ideally, a pixel's value should reflect the incident light intensity (luminance). So we need to count the number of photons striking each pixel.

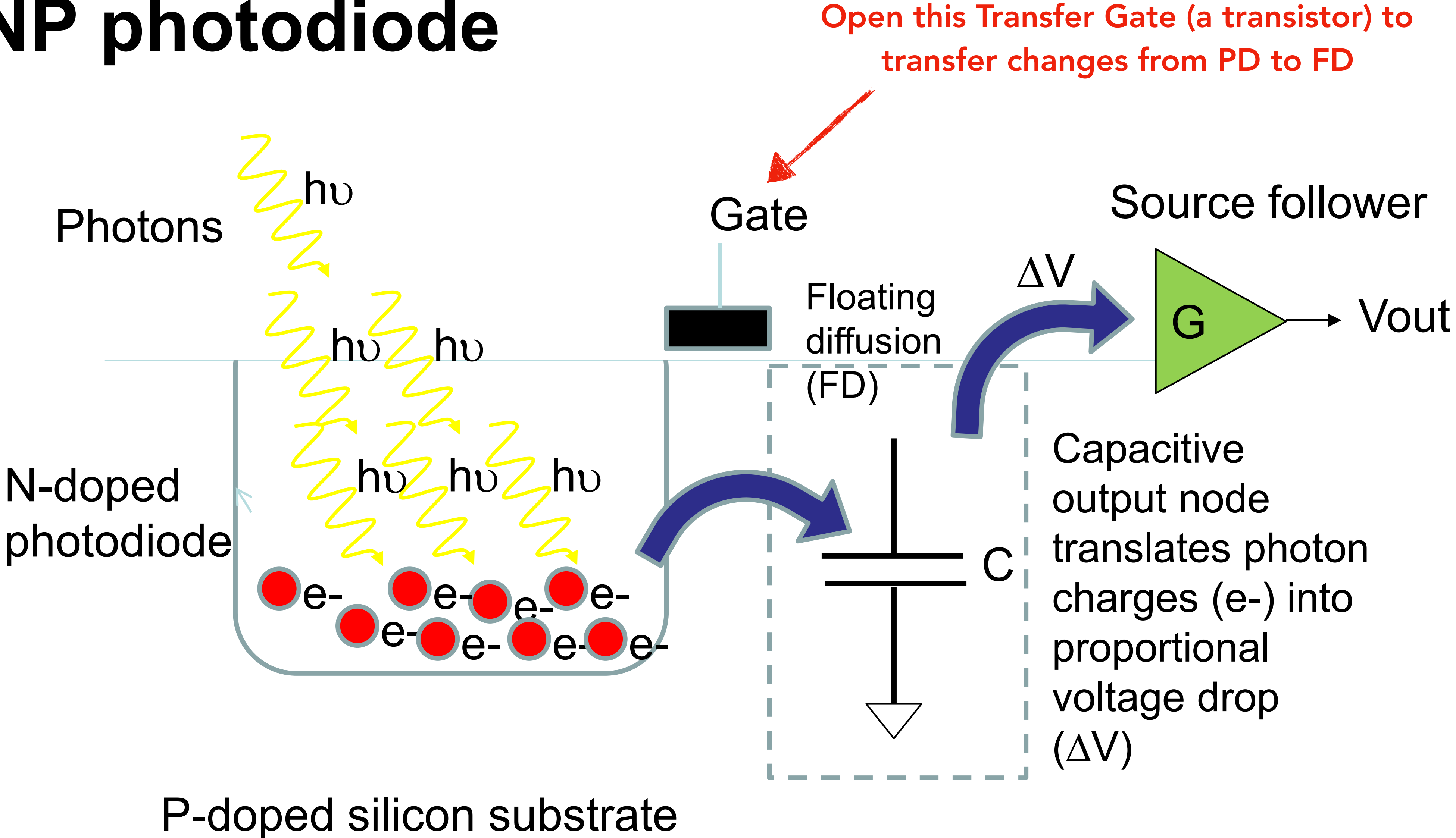
Pixel wells accumulate electrons. The number of electrons is proportional to the number of incident photons.

So counting photons is equivalent to counting electrons.

Count electrons in two steps (implementations differ):

- Convert electron charges to voltage potentials
- Convert voltage potentials to digital values

# NP photodiode



# Charge to Voltage Conversion

Charge in the signal

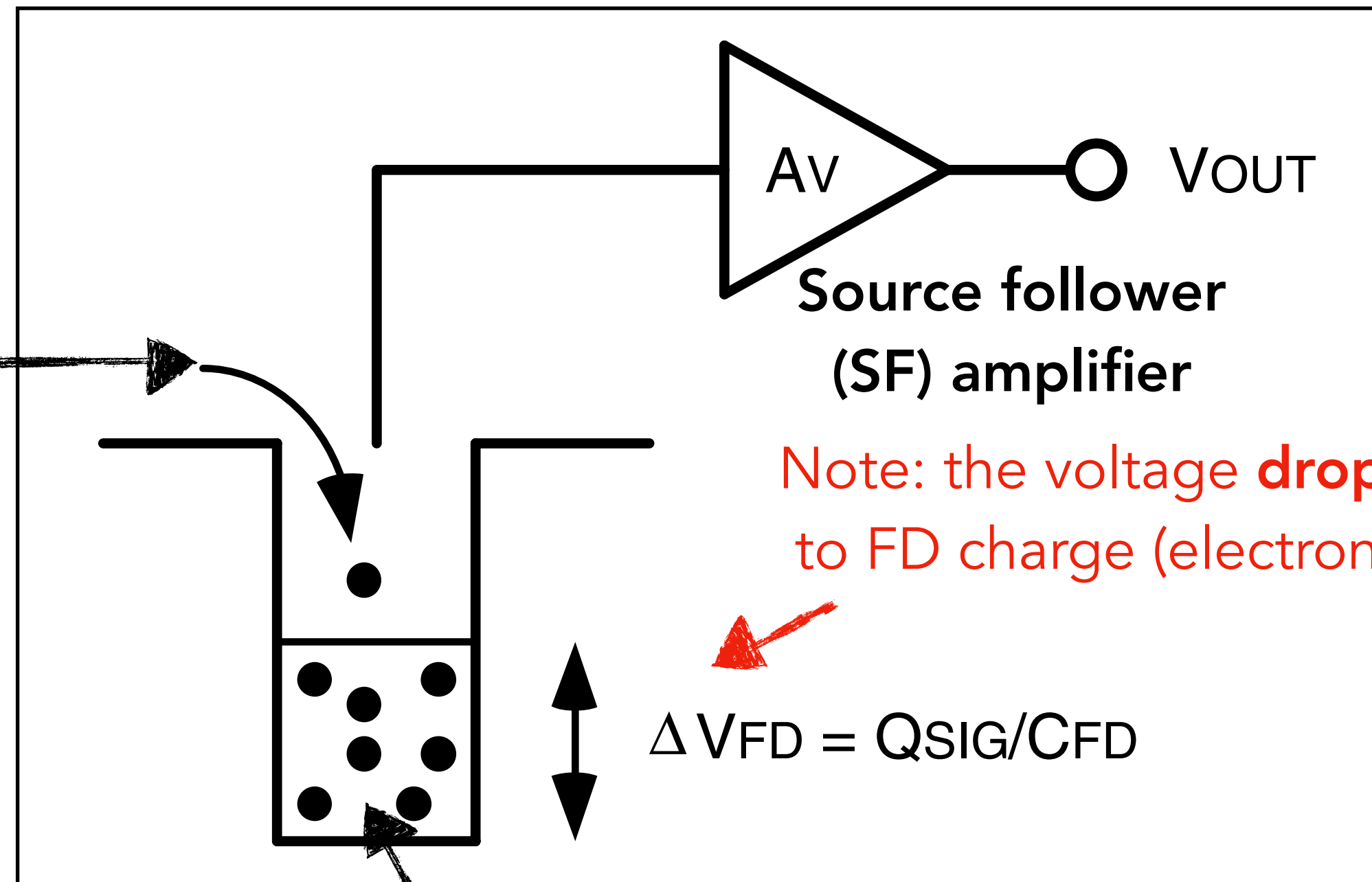
$$\Delta V = \frac{Q_{SIG}}{C_{FD}} \times g$$

Capacitance

Voltage gain from SF

\* The source follower acts as a voltage buffer that allows the voltage to be read. The SF has an inherent amplification (~1).

Charges from Photodiodes (PD)



The capacitor is implemented as a Floating Diffusion (FD)

# Electronic Shutter

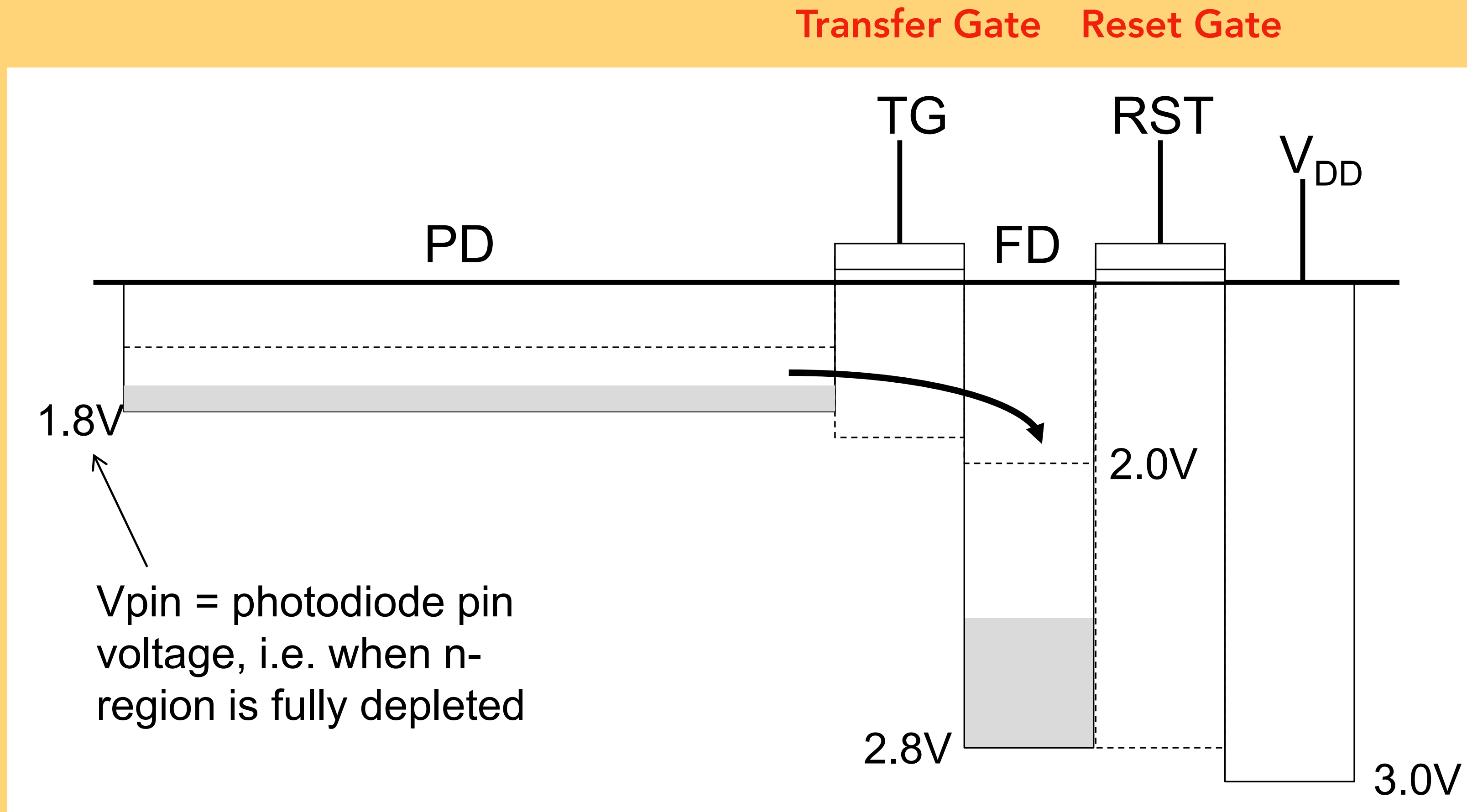
Ideally when we are not capturing the photodiodes should not be exposed to lights. Mechanical shutters do so by physically blocking lights.

With electronic shutters, we expose photodiodes to lights all the time, and when are ready to capture we “reset” the photodiodes.

Reset drains (depletes) the charges accumulated by the photodiode thus far and so prepares it for the actual exposure.

- Change the photodiodes reset time to control actual exposure time.

# Electronic Shutter: Two Resets



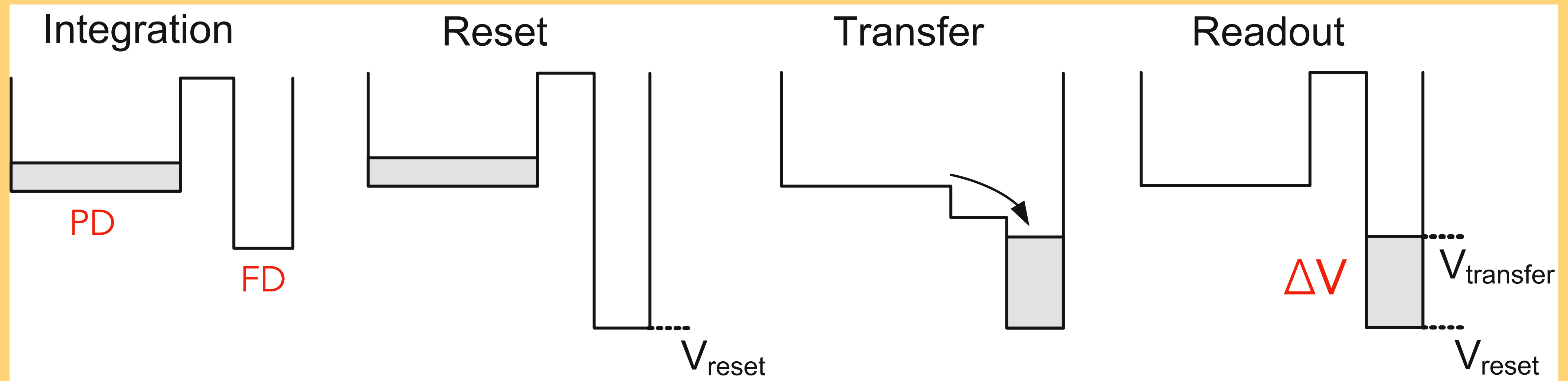
Before exposure, open TG and RST together to **reset the PD to  $V_{DD}$**

- FD will naturally be reset too but it's of no functional use.

After exposure and before read-out, open RST to **reset FD**.

- Then PD to FD transfer.

# FD Reset

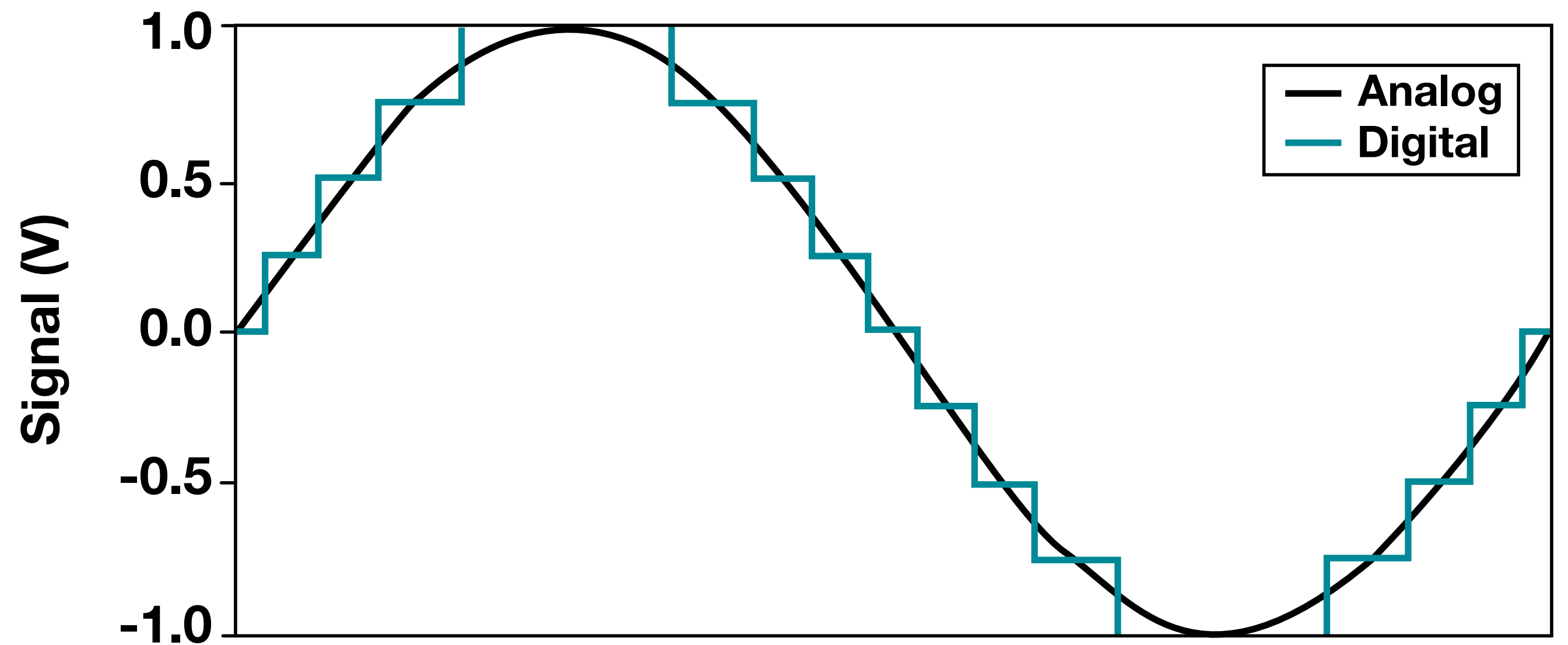
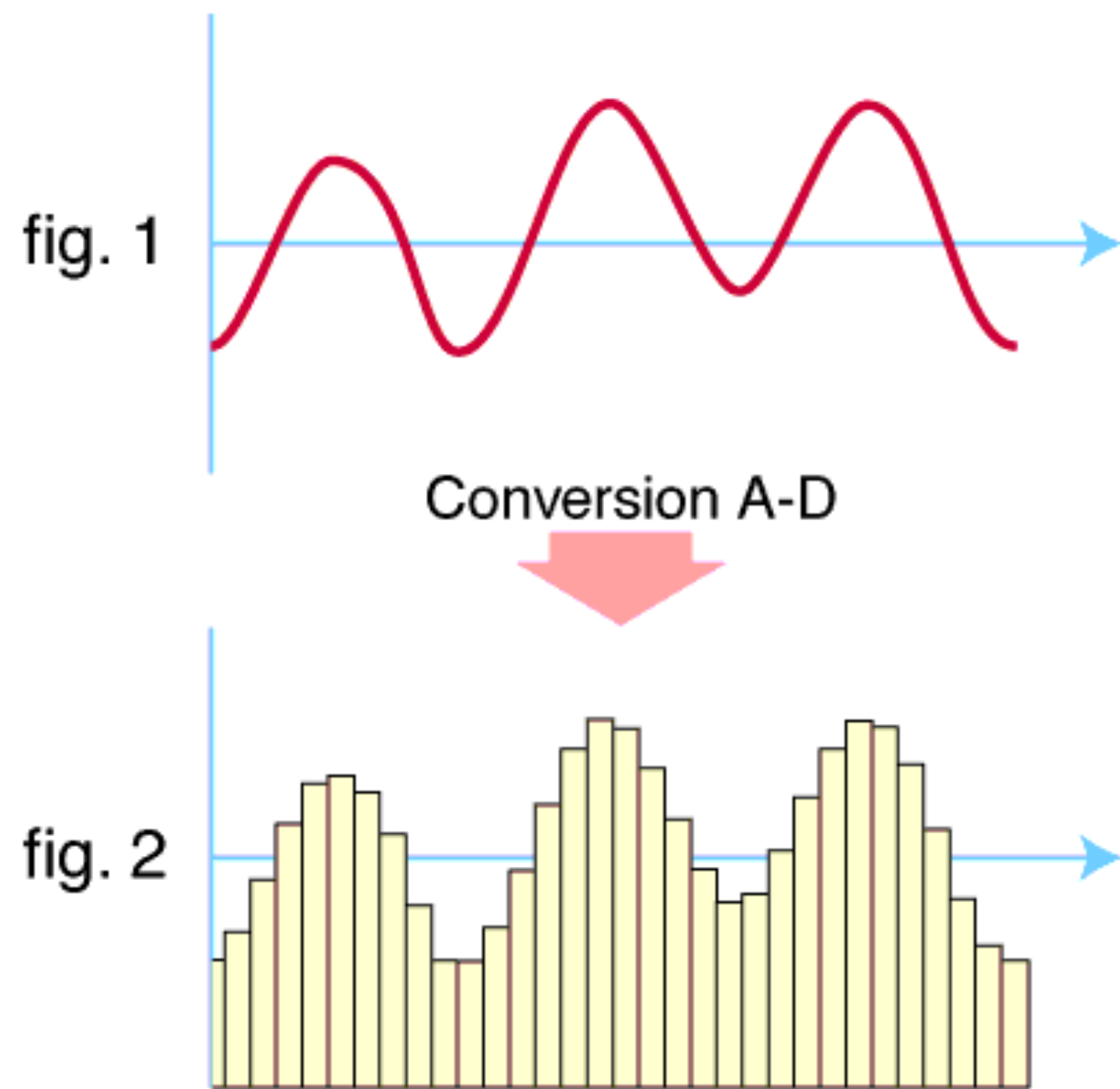


Conceptually,  $\Delta V$  is used to generate raw pixel values; implementations vary:

- With Correlated Double Sampling (CDS; next lecture), both  $V_{\text{transfer}}$  and  $V_{\text{reset}}$  are read, and the difference  $\Delta V$  is quantized by ADCs. CDS reduces noises.
- Without CDS,  $V_{\text{transfer}}$  directly enters the ADC, whose design needs to, effectively (by properly setting its reference voltage), subtract  $V_{\text{reset}}$ .

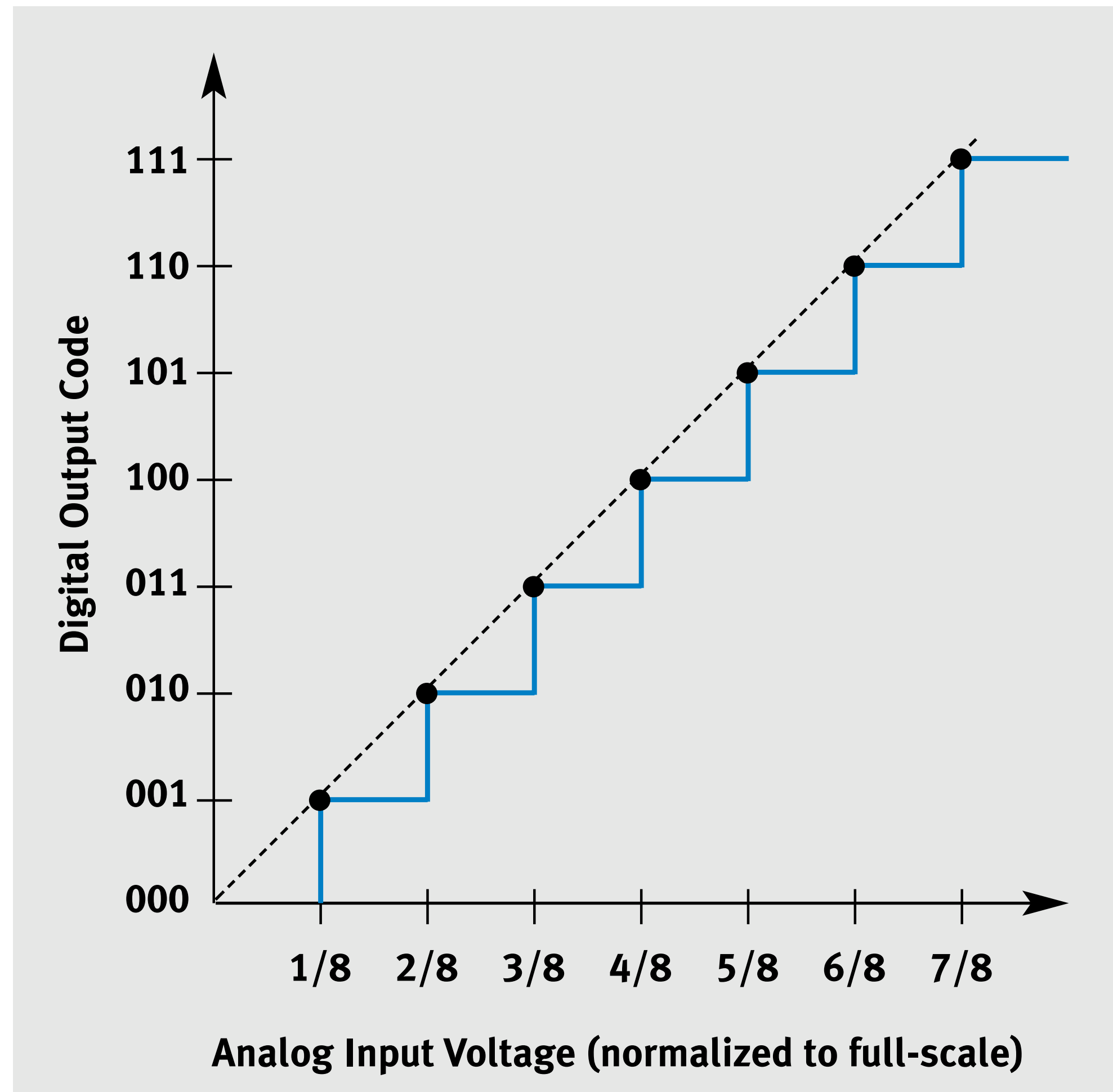


# Analog to Digital Conversion (ADC)

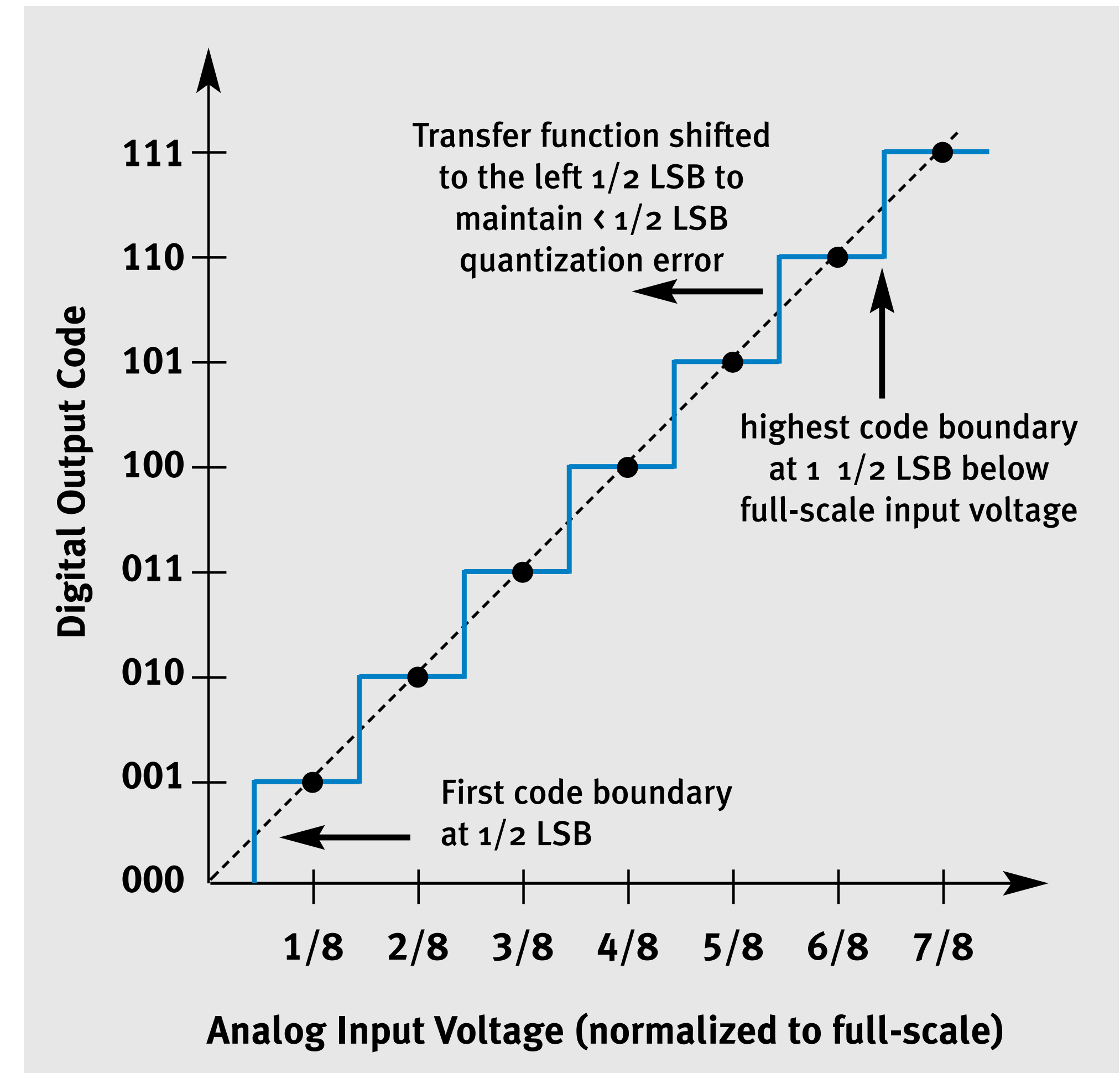


# Transfer Function

Ideal Transfer Function of a 3-bit ADC

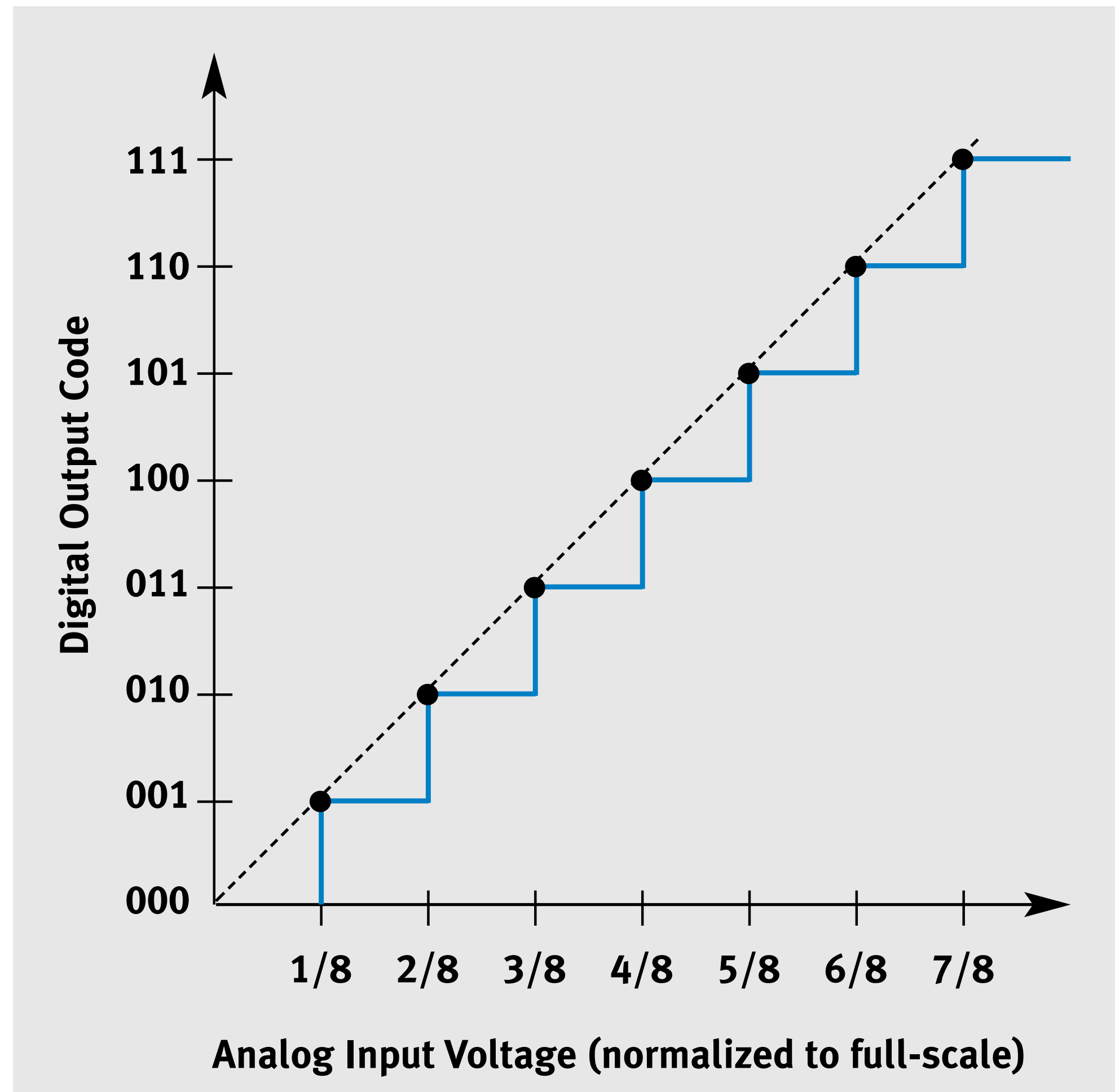


Another (more common) Transfer Function



# Analytical Model

Ideal Transfer Function of a 3-bit ADC



Full-scale range

$$\text{FSR} = V_{\text{max}} - V_{\text{min}}$$

Upper extreme      Lower extreme  
extreme              (usually 0)

Output digital value

$$n = \left\lfloor \frac{V_{in} - V_{min}}{\text{FSR}} (2^N - 1) \right\rfloor$$

# Contouring From Aggressive Quantization

Lower ADC resolution;  
smaller image size (in theory)

Higher ADC resolution; larger image size



# A Monochromatic, Noise-Free Sensor Model

$$Q(x, y) = \int_t \int_\lambda \int_y^{y+v} \int_x^{x+u} Y(x, y, \lambda, t) QE(\lambda) q \, dx dy d\lambda dt$$

# of incident photons at pixel (x, y);  
varies with  $\lambda$  and t.

Charge of an electron  
(elementary charge)

$$\Delta V(x, y) = \frac{Q(x, y)}{C} g$$

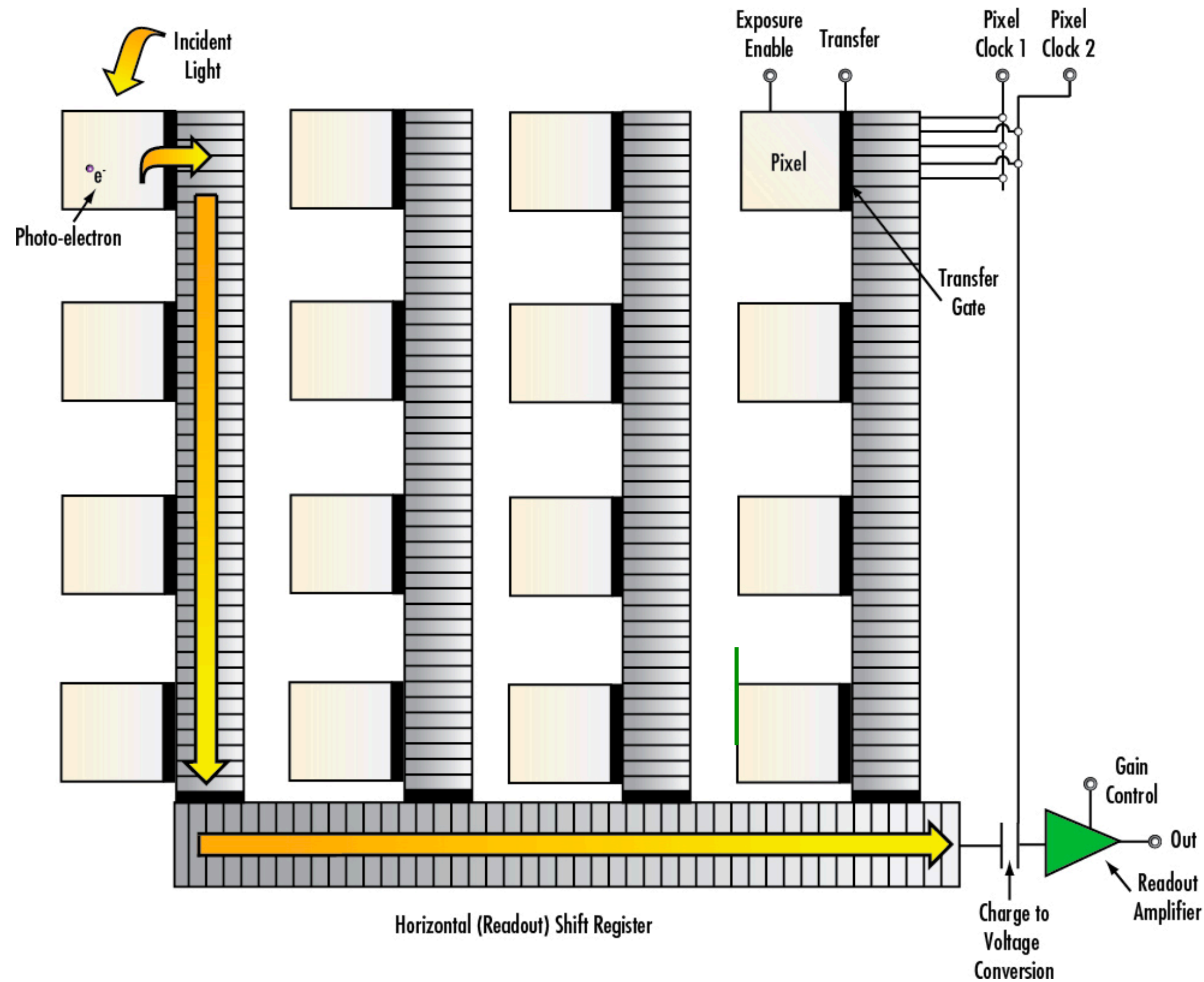
$$n(x, y) = \left[ \frac{\Delta V(x, y)}{V_{max}} (2^N - 1) \right]$$

Raw pixel value

Max voltage corresponding to  
max raw value (assuming  $V_{min} = 0$ )

# Different Sensor Designs

# CCD Sensor



Exposure all pixels simultaneously.

Shifting charges row by row. When shifting, naturally no pixel can be exposed, a.k.a., global shutter.

- A good introduction of CCD [here](#).

Convert charges to voltage and then to digital values in the end.

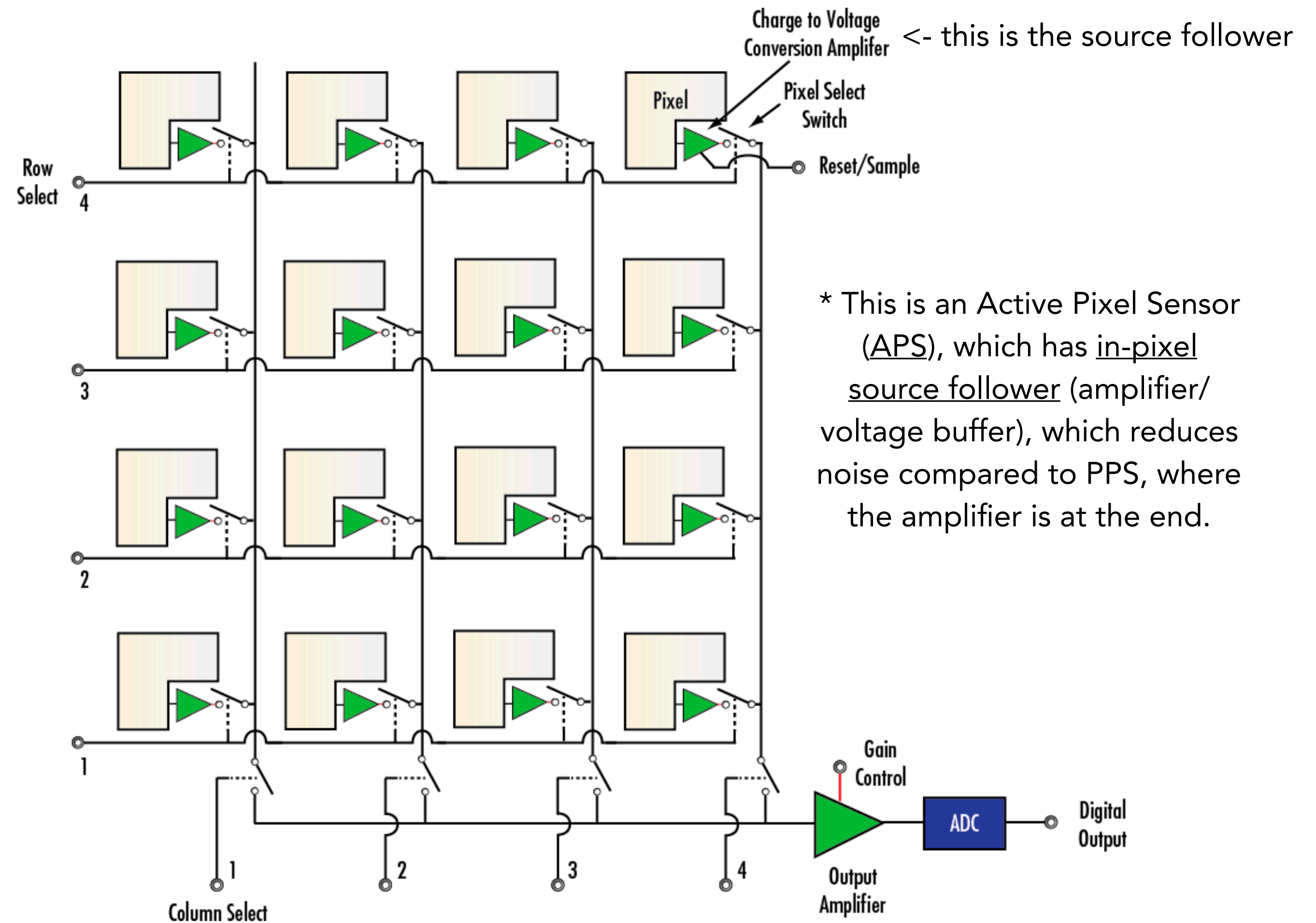
# CMOS Sensor

Charge to voltage conversion takes place within the pixel. Instead of shifting charges, directly read the voltages.

- but still row by row.

Pixels are now exposed row by row, a.k.a., rolling shutter.

- because voltages in each row must be read out immediately (no buffer).
- an artifact; not something we want.

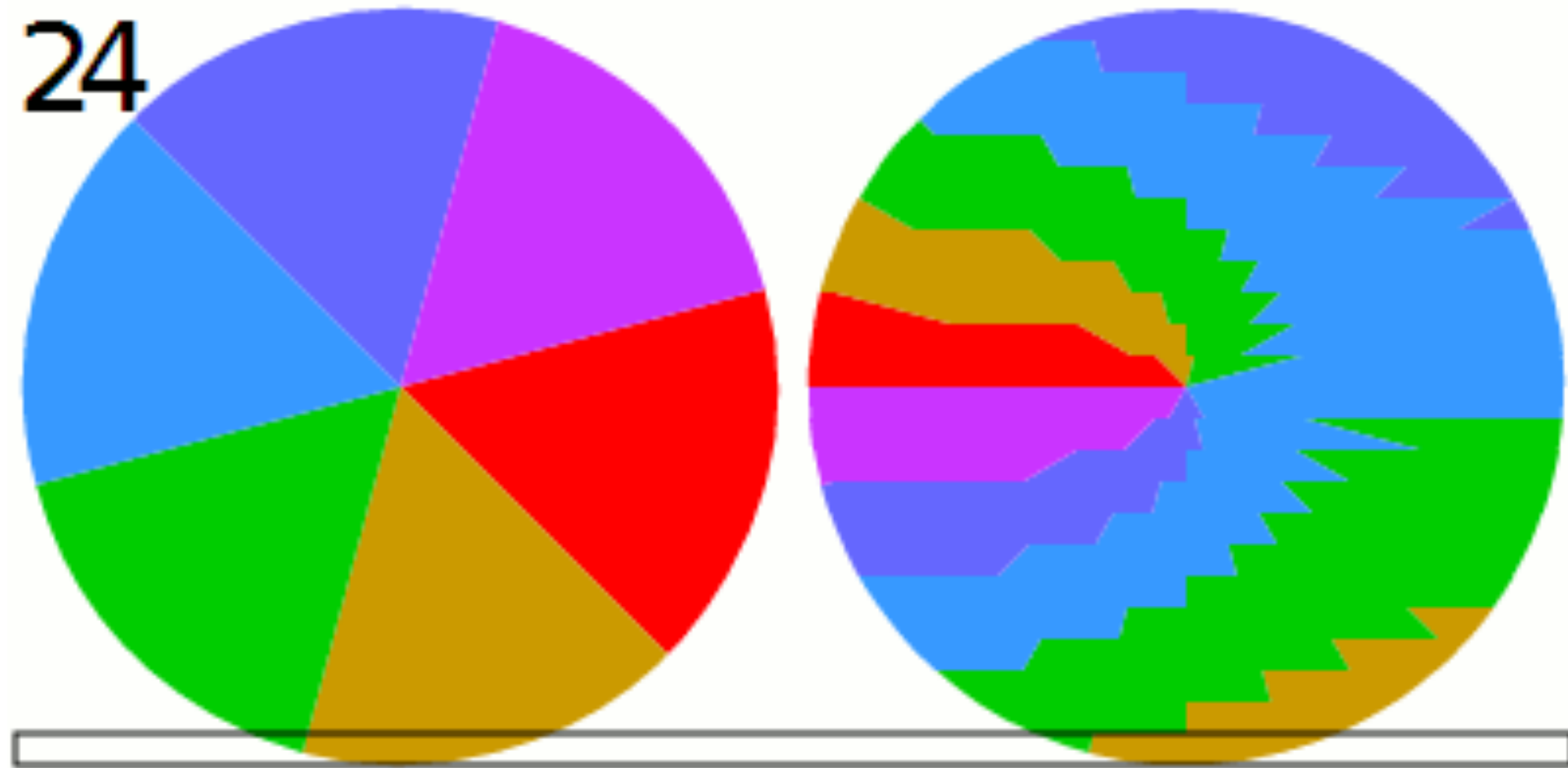




# Rolling Shutter Operation



# Rolling Shutter Artifact



[https://en.wikipedia.org/wiki/Rolling\\_shutter](https://en.wikipedia.org/wiki/Rolling_shutter)

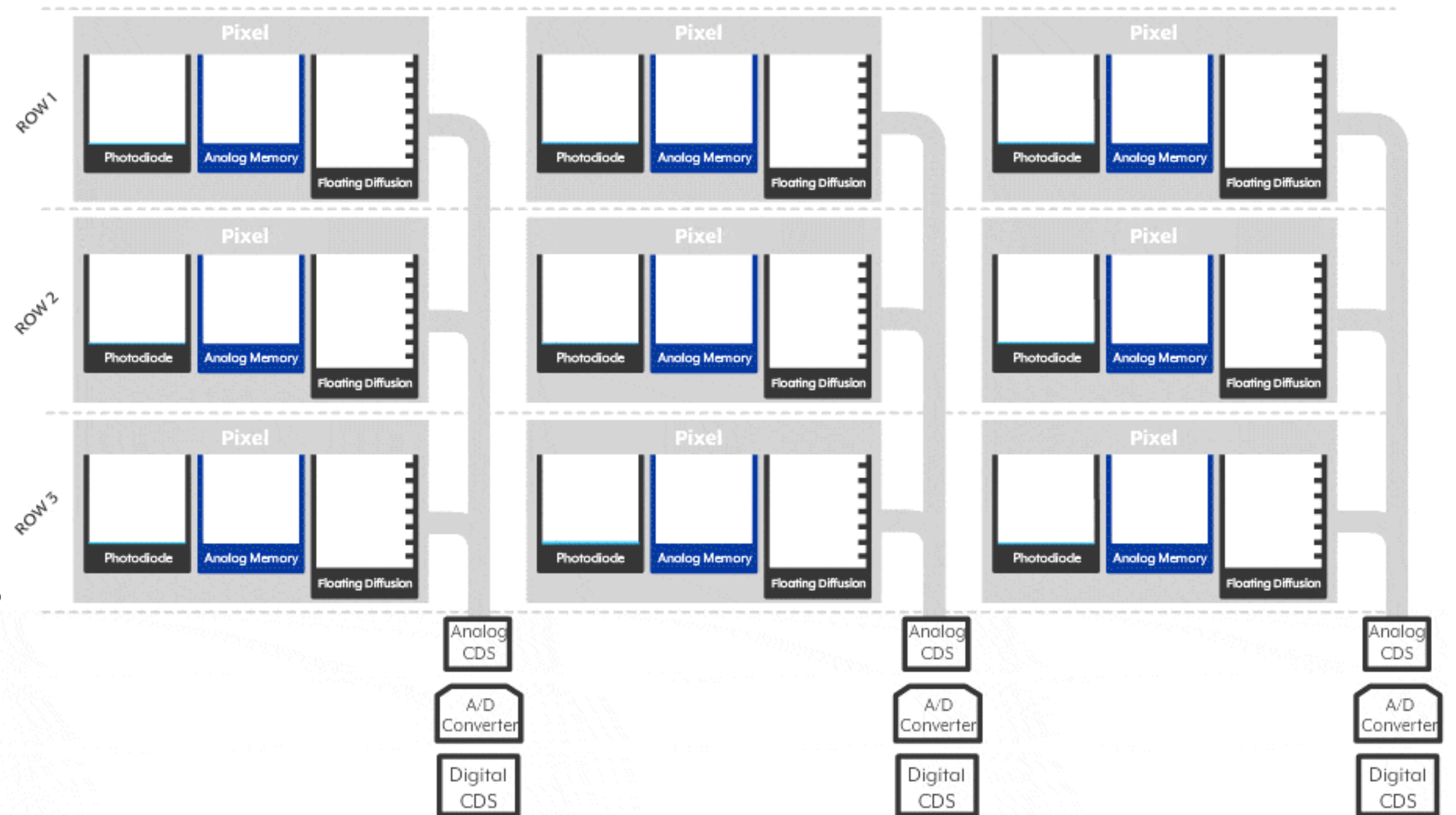
[https://www.reddit.com/r/interestingasfuck/comments/en3lap/rolling\\_shutter\\_effect\\_of\\_chopper\\_blades/](https://www.reddit.com/r/interestingasfuck/comments/en3lap/rolling_shutter_effect_of_chopper_blades/)

<https://www.bhphotovideo.com/explora/video/tips-and-solutions/rolling-shutter-versus-global-shutter>

# Global Shutter

CCD naturally uses global shutter, since no pixel can be exposed during read-out.

To implement global shutter for CMOS sensors, each pixel needs to have a temporary storage space (analog voltage buffer).



# CMOS Global Shutter Read-Out

Expose all pixels at the same time, but readout is still line by line.

Why not read all the pixels at the same time?

- Lots of wires.
- Unless you stack the read-out circuit with the pixel array in different layers!

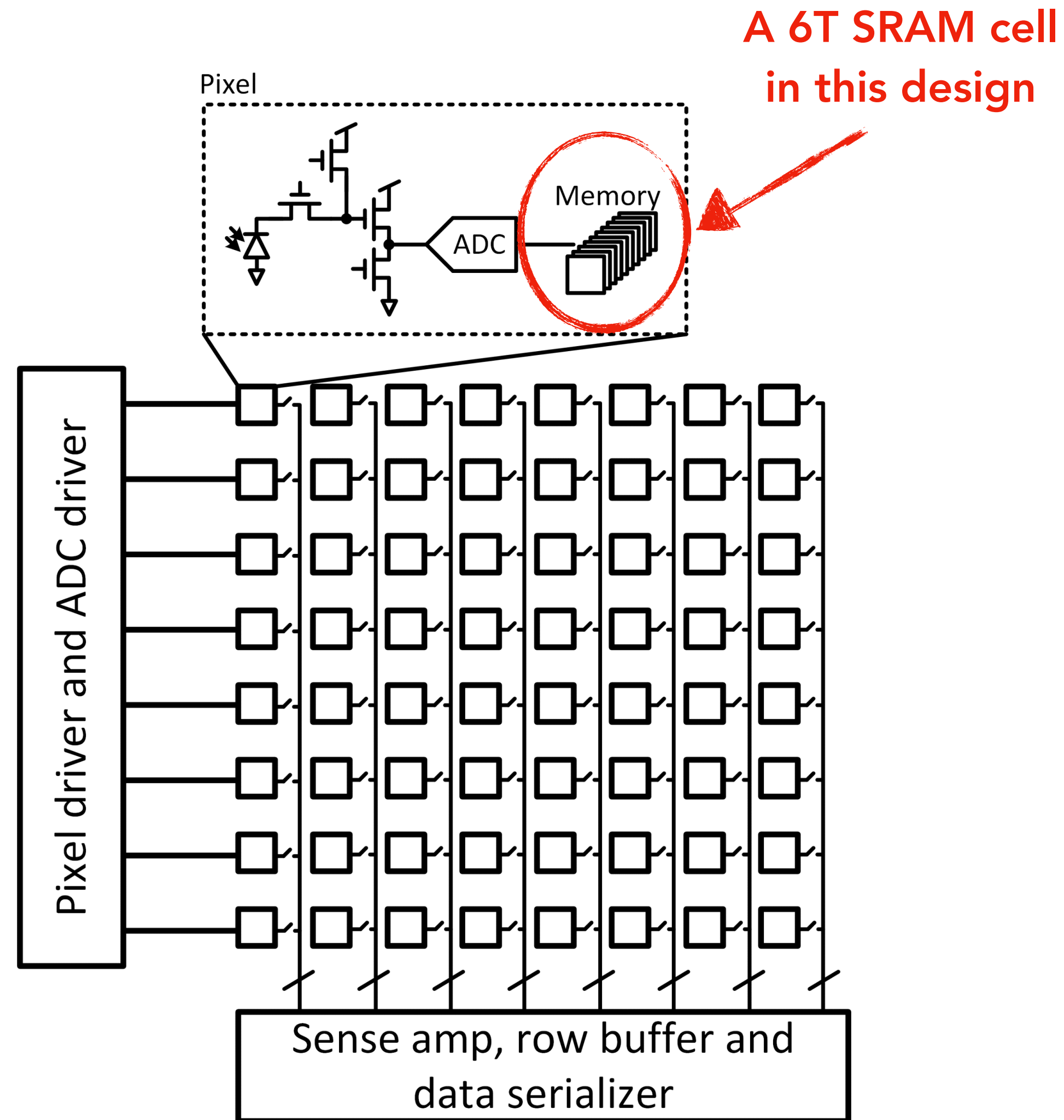
Therefore, need analog memory to store charges before read-out.

- A “solution” that avoids analog memory in global shutter is to expose later lines longer and still read line by line, but spatially non-uniform exposure is clearly undesirable.

Does global shutter avoid motion blur?

- No

# Digital Pixel Sensors



- Each pixel contains an ADC and digital memory
- Pixel data is stored in digital memory and readout as digital bits
- ADC circuit has to be small enough to keep pixel size reasonable
- Now you can treat the pixel array as your frame buffer memory!
- DPS opens new possibilities, we will demonstrate one example in next section

Usually the ADC and digital memory is on a separate layer different from the pixel array layer (later)

Slide credit: Chiao Liu (Meta)

# CCD vs. CMOS

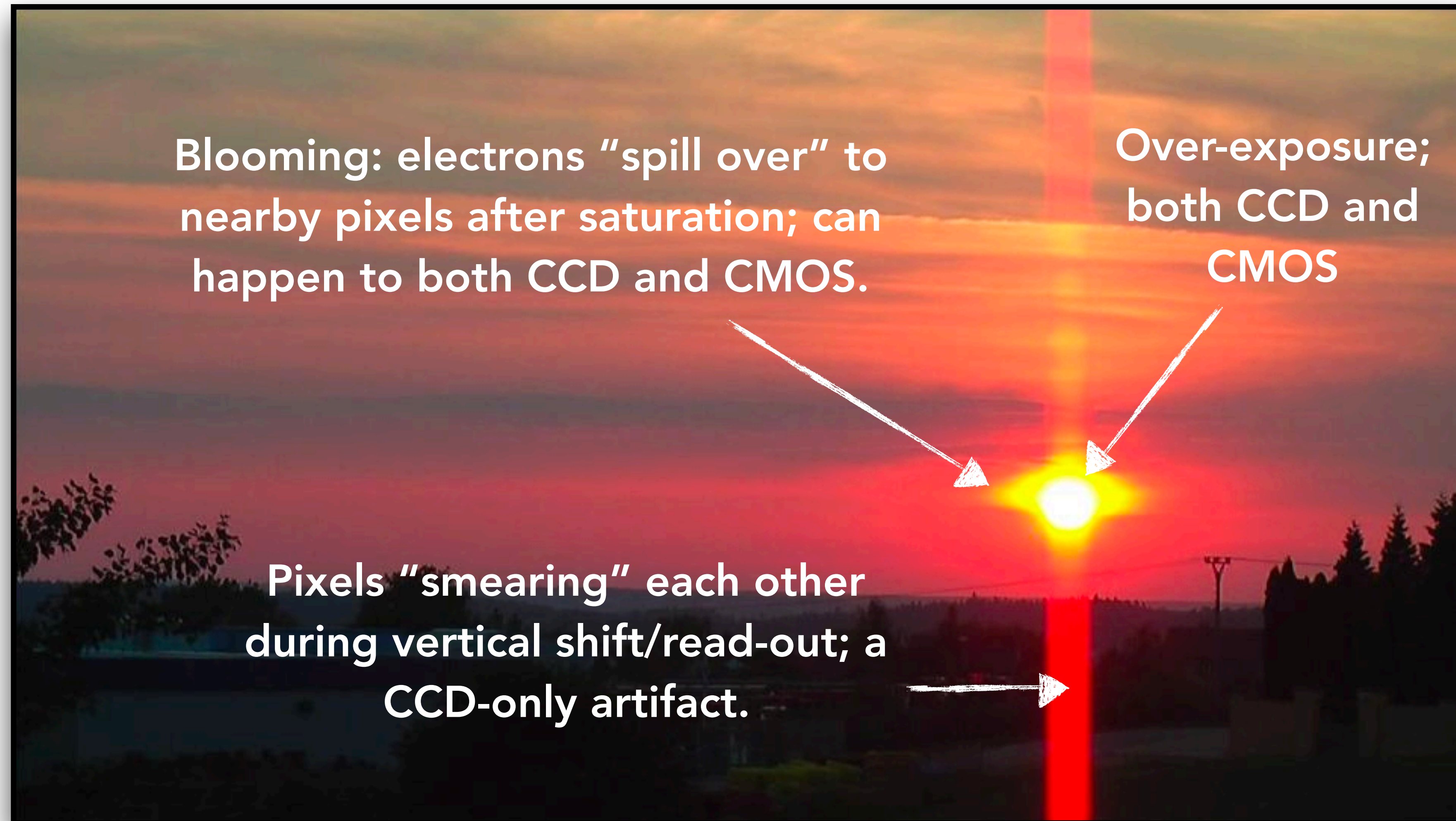
## CCD Sensors

- Higher quality (FF higher), expensive
- Slower, higher power (circuitry needs precise timing control)
- Astrophysical imaging (e.g., telescopes) still prefers CCD sensors

## CMOS Sensors

- Dominate the consumer photography market now, cheaper
- Same fabrication technology as digital microprocessors, enabling "camera on a chip"
- Faster and lower-power (can easily read just a small window of pixels)
- More noisy (e.g., each pixel has amplifier and/or ADC), but getting better

# Over-Exposure, Blooming, Smearing



# Time Delay Integration (TDI) CCD Sensor

CCD uses global shutter, but still doesn't avoid motion blur.

Long exposure introduces motion blur, but many application scenarios require a long exposure time, e.g., astrophysical imaging.

TDI is a clever sensor design that mitigates motion blur under long exposure.

Conventional CCD sensor: shift pixel rows after exposure.

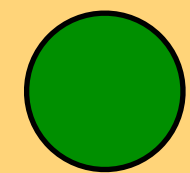
**TDI CCD:** shift rows during exposure and accumulate charges for each shift.



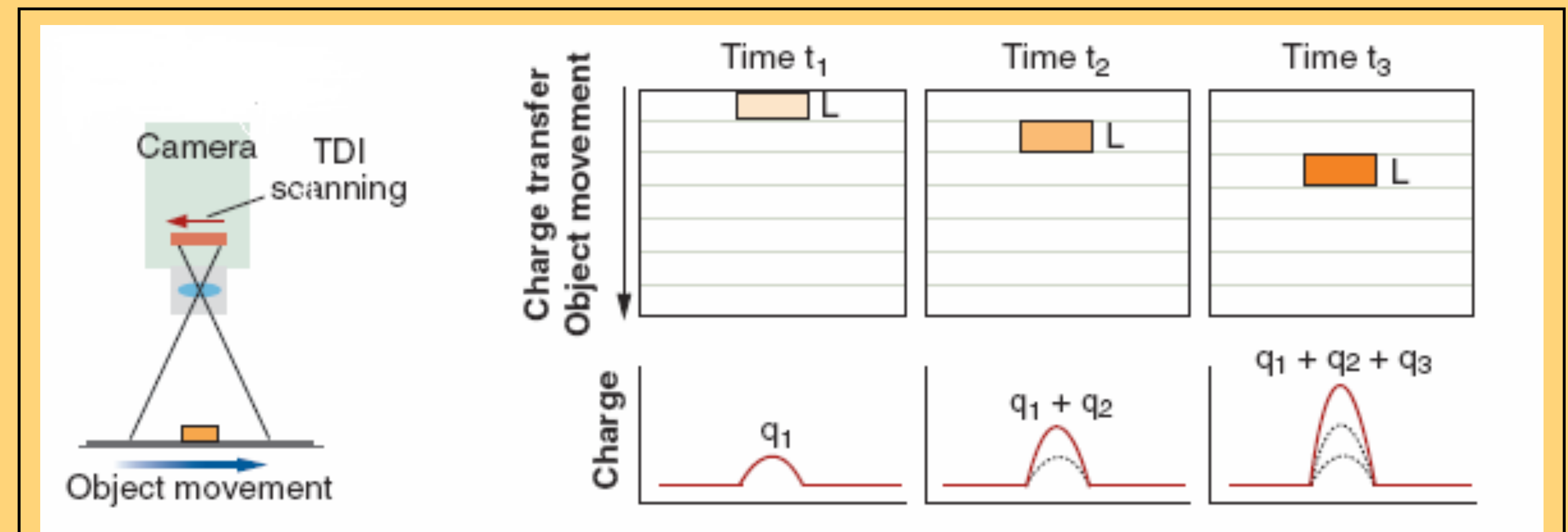
# TDI Operation

The shift is in sync with object movement. Effectively capture a moving object N times, once per row, increasing the effective exposure time.

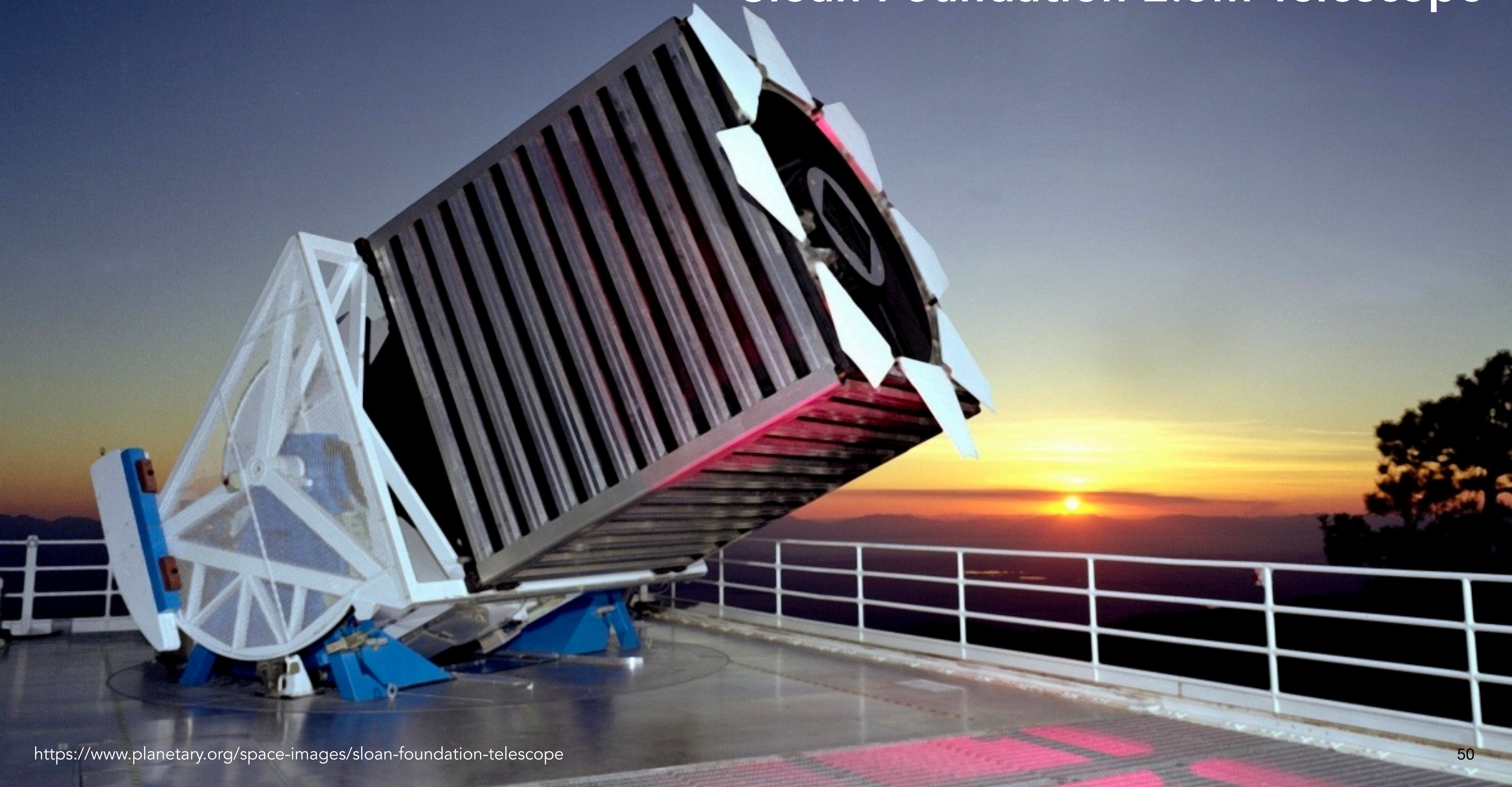
Object movement within a pixel doesn't matter (pixels integrate anyways).



9
+ 8
+ 13
+ 7

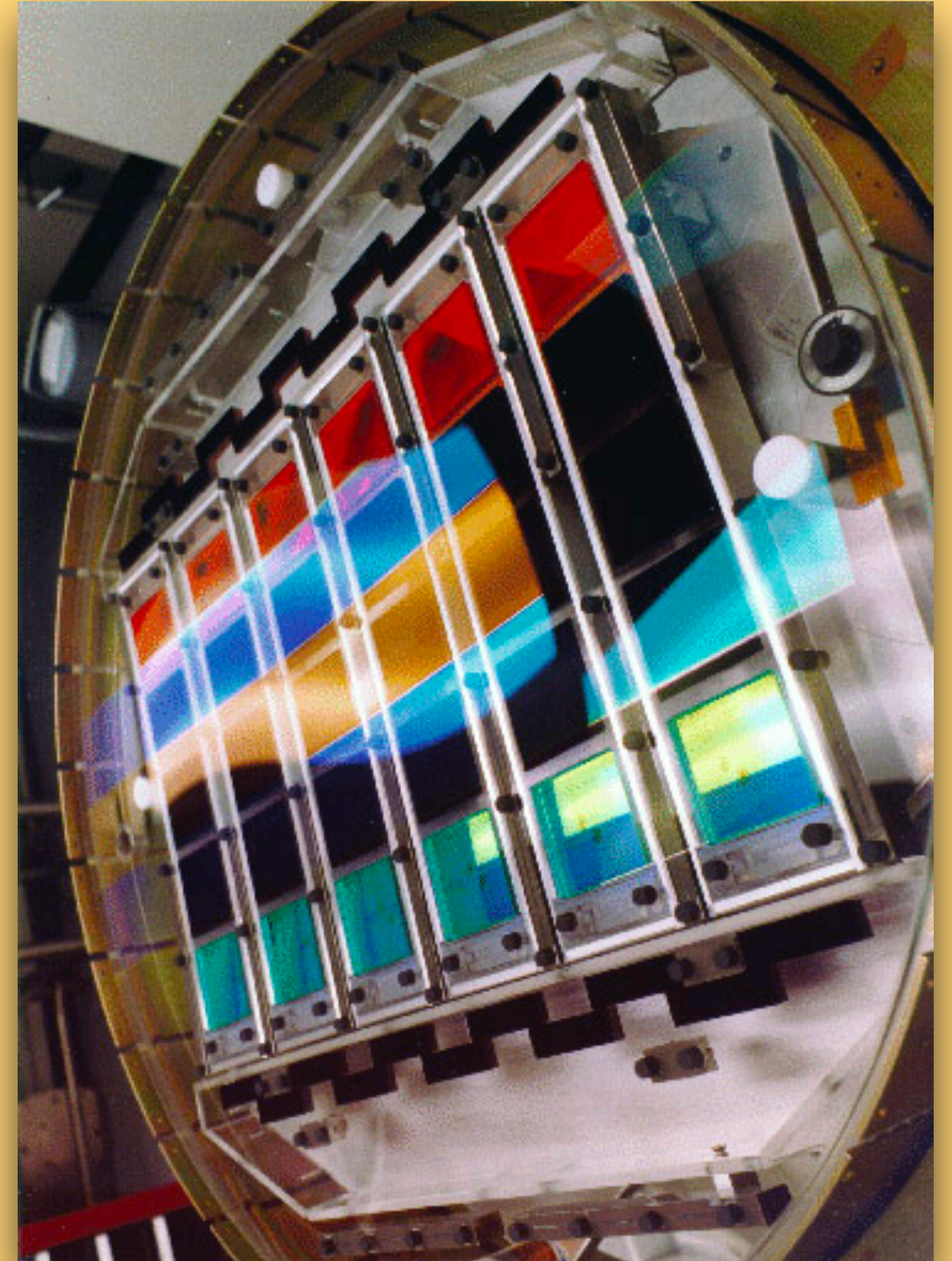
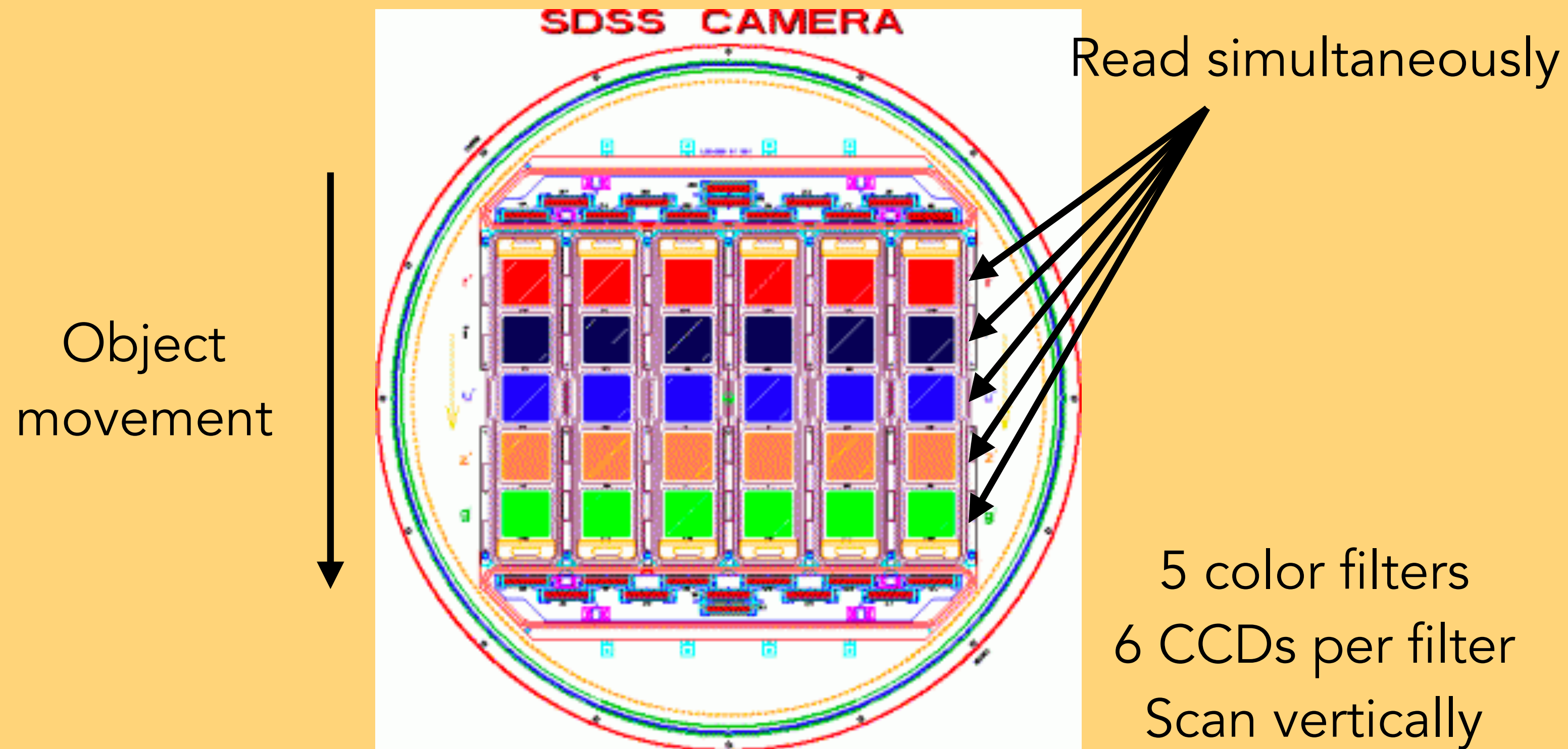


# Sloan Foundation 2.5m Telescope



# TDI in Astrophysics

Can estimate the TDI camera shift rate, since the star's movement speed is known.



# TDI Used in Sloan Digital Sky Survey

Images of objects move along the columns of the CCDs at the same rate the CCDs are being read.

The camera produces five images of a given object, all from the same column of CCDs, one from each CCD in that column.

It takes an object 54 seconds to move from the beginning of a CCD to the end, so the effective exposure time in each filter is 54 seconds.



# Stacked Sensor

# Why Stacking?

Enable in-sensor computing to reduce data transmission overhead

Save area:

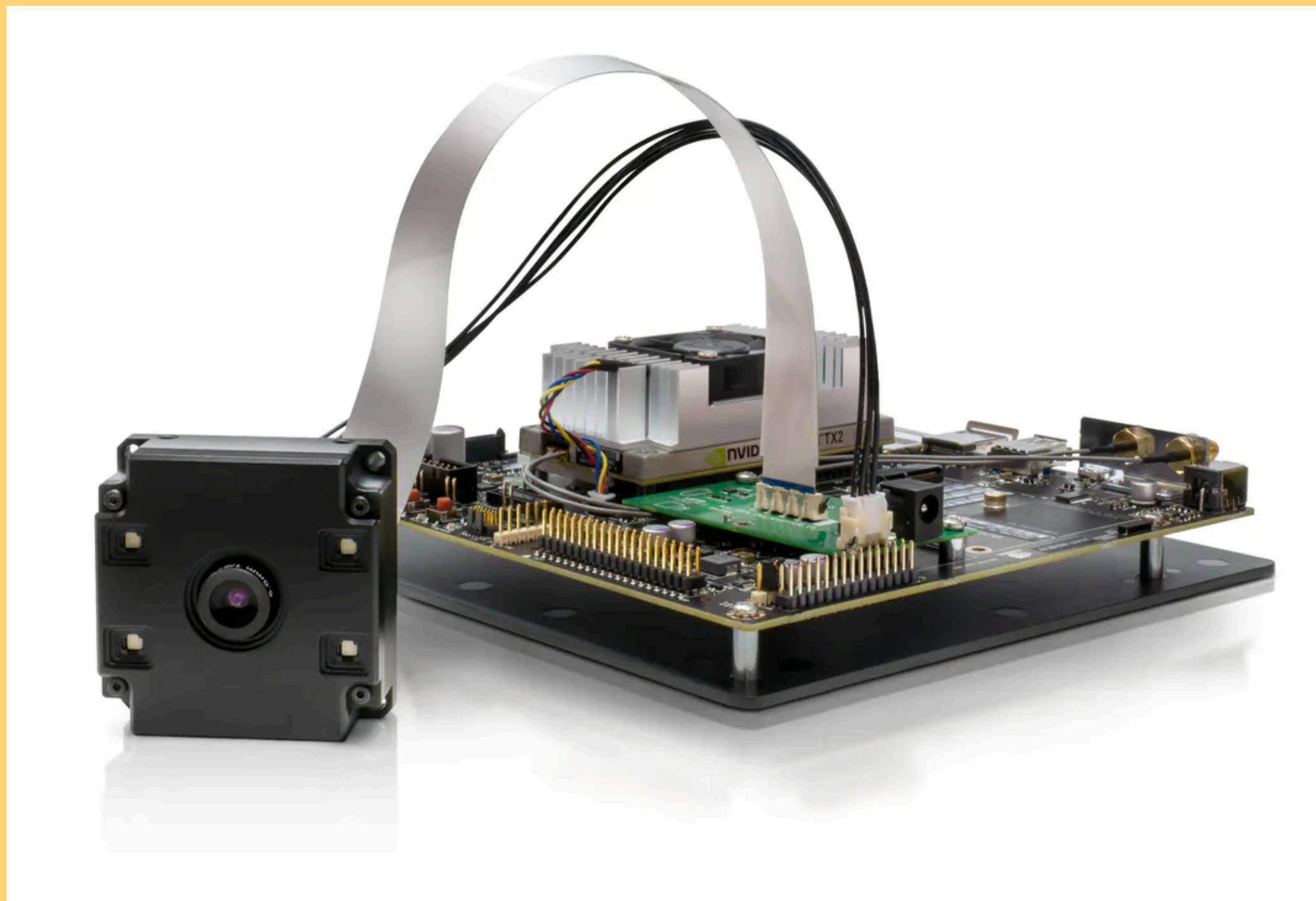
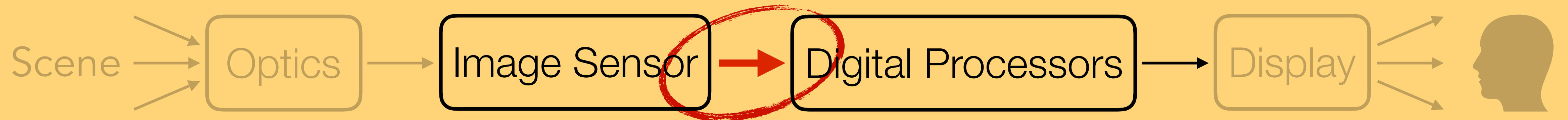
- Pixels have become more complicated, taking more area.
- Grow vertically rather than horizontally to reduce sensor footprint.

Can even stack memory (e.g., DRAM/SRAM) and processors with the pixel array in one sensor.

- Enable ultra-high frame rate
- Better support in-sensor processing (imaging you have a cache in sensor!)

# Data Transmission Bottleneck

## MIPI CSI Interface



Vision System Design

## Energy

1 B MAC  
Compute



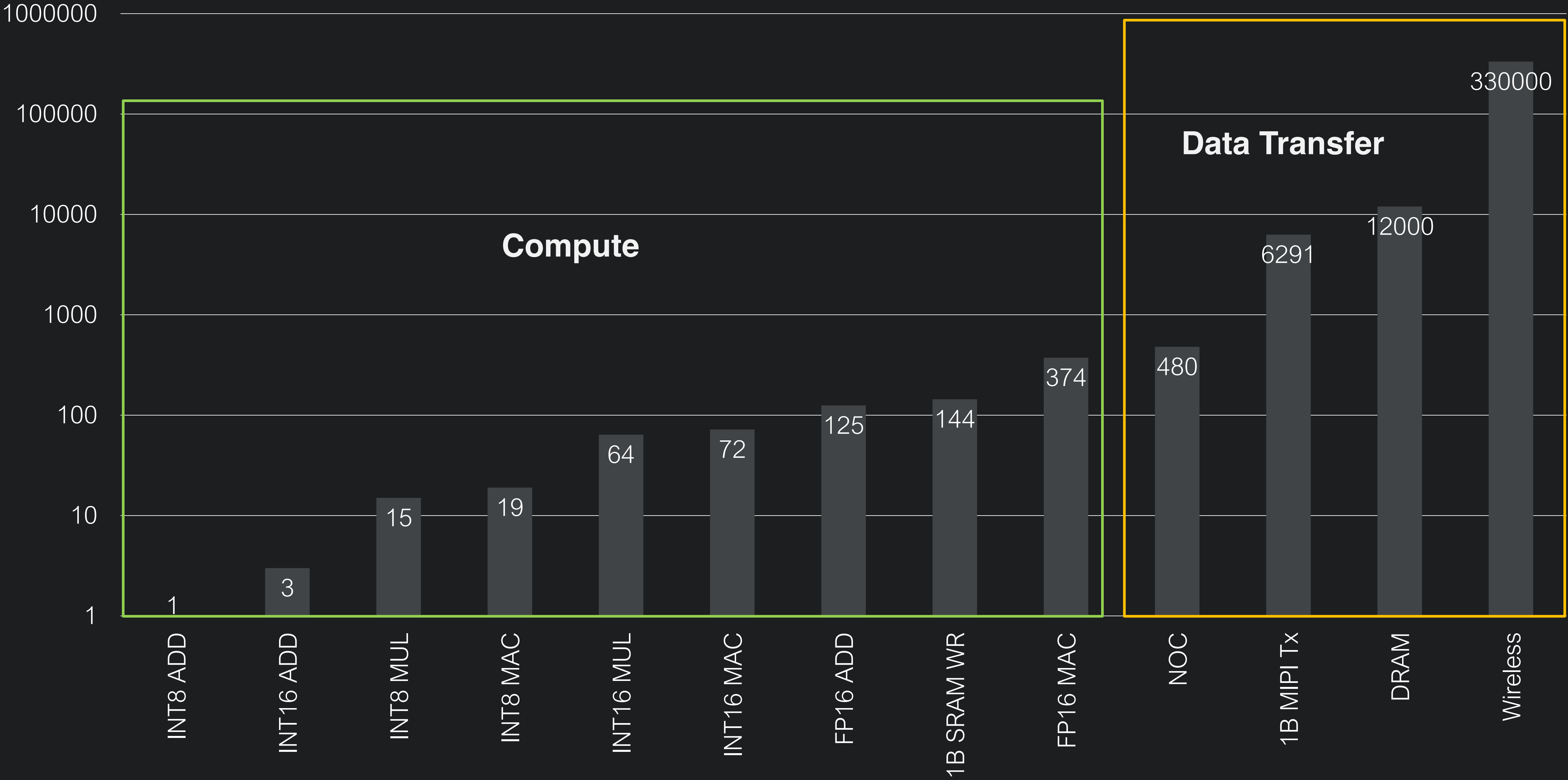
1 B MIPI  
CSI-2 Tx



1 B WiFi

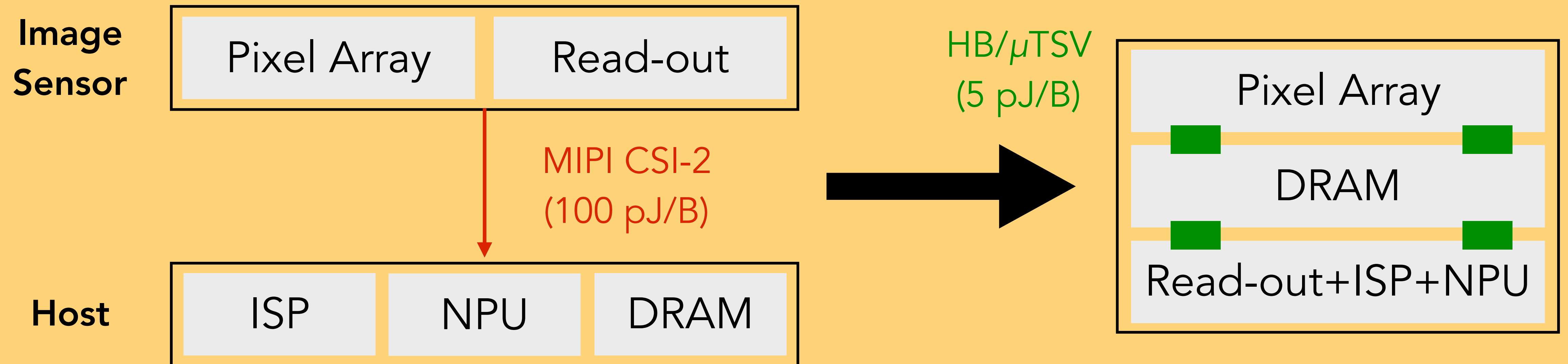


# Normalized Energy



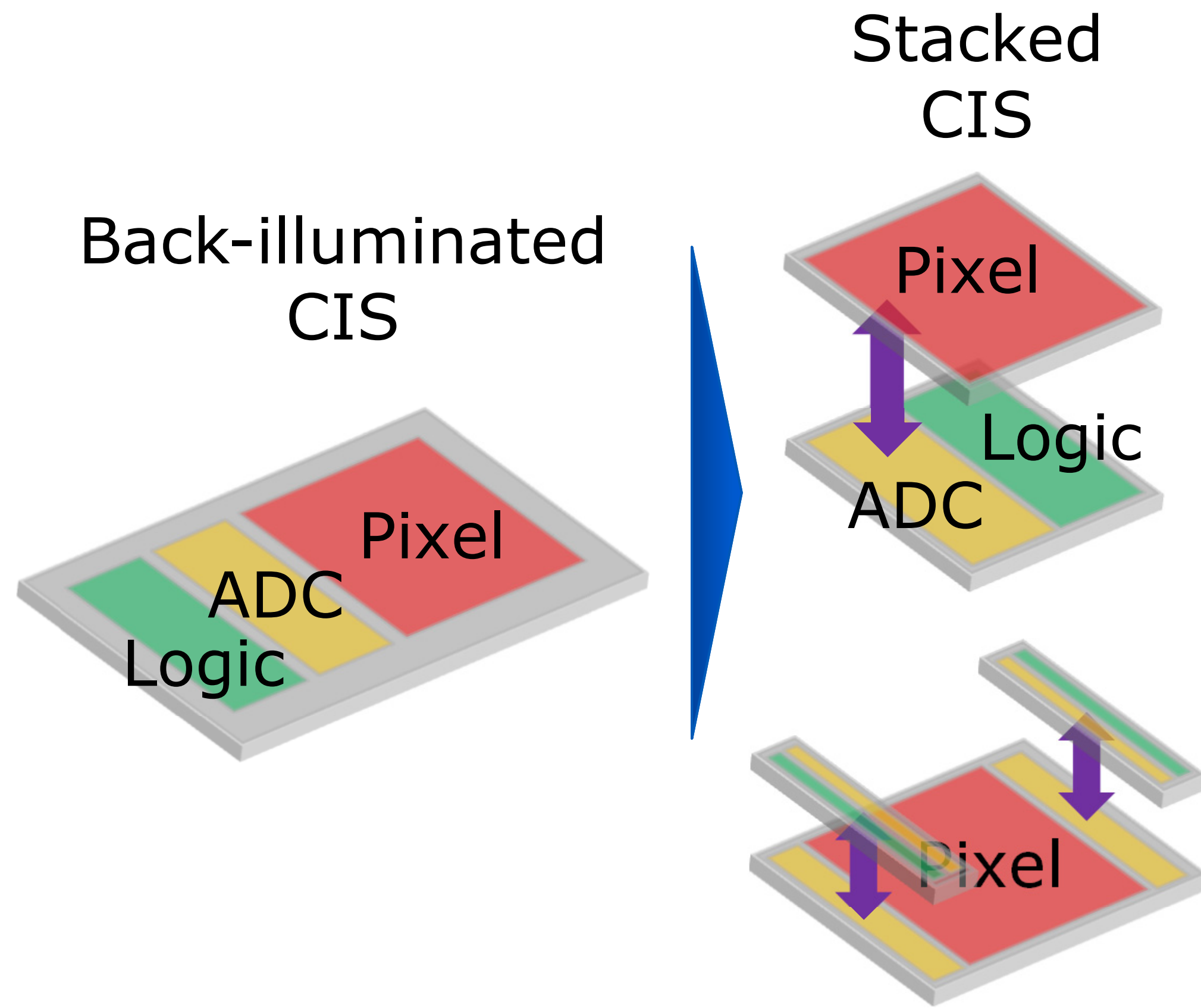


# Moving Computation Into Sensor



# Stacked CMOS Image Sensor

- Performance improvement accelerated using 3D configuration

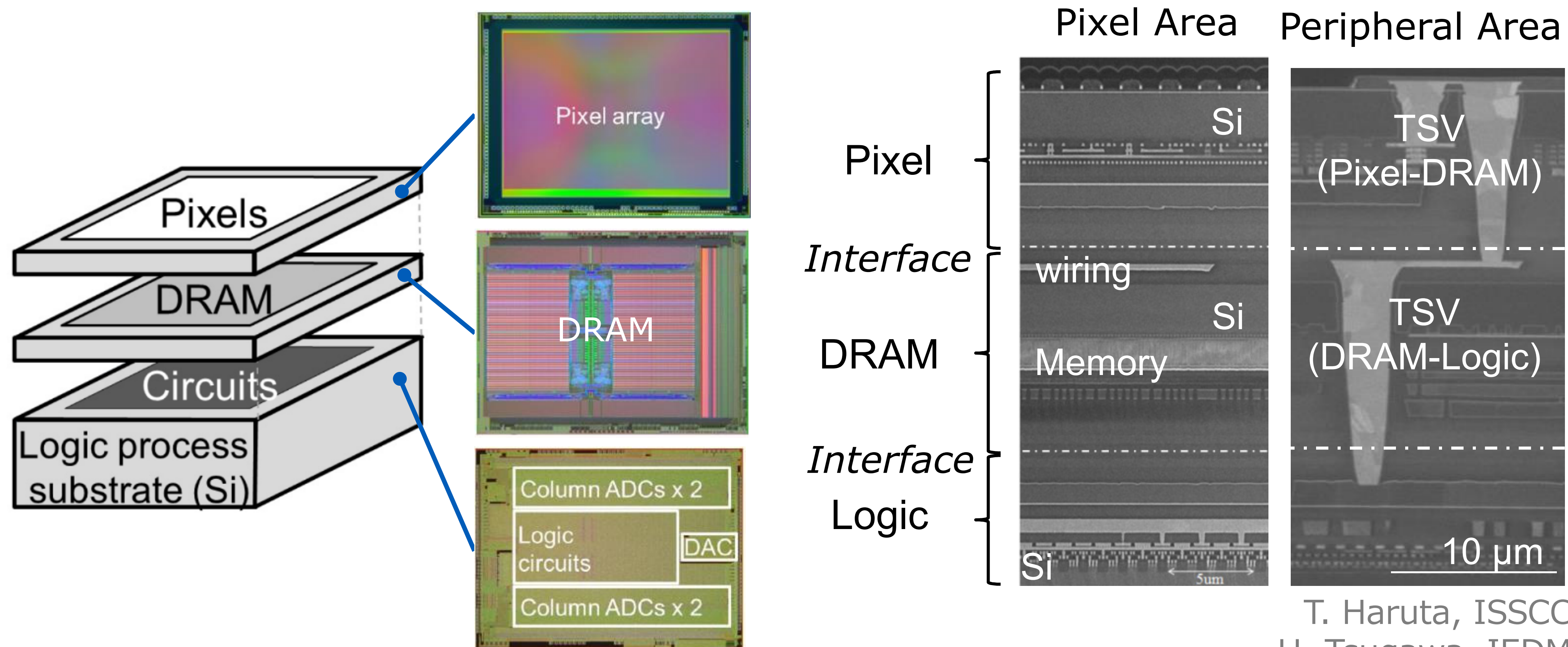


## Highly parallel ADCs & Optimized process technologies

- Double parallel ADCs
- Chip-on-chip stacked CIS
- Three-layer stacked CIS

# Three-layer Stacked CIS with DRAM

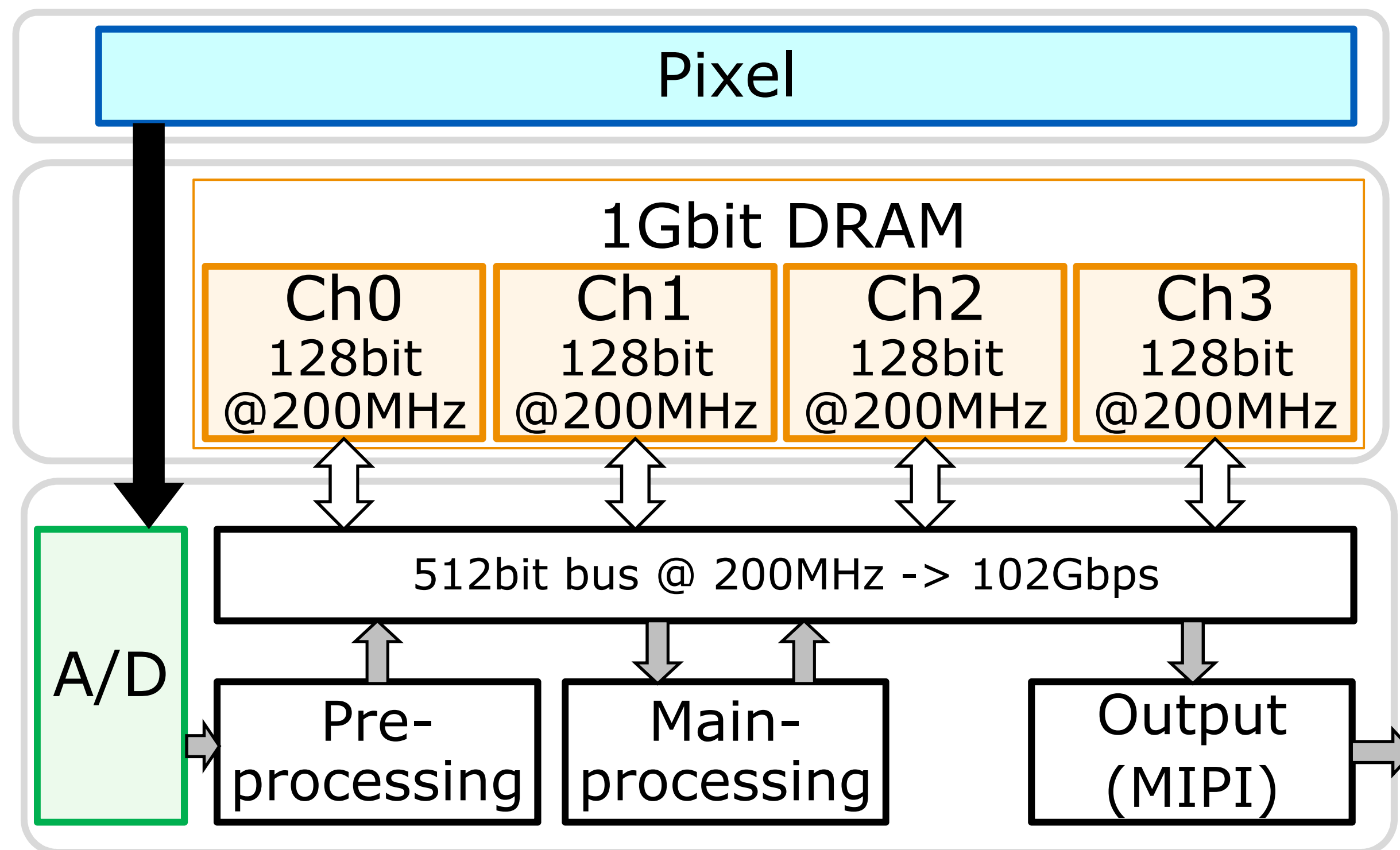
- DRAM buffer having wide data bandwidth for slow-motion capture



T. Haruta, ISSCC 2017  
H. Tsugawa, IEDM 2017

# Sony Develops the Industry's First<sup>\*1</sup> 3-Layer Stacked CMOS Image Sensor with DRAM for Smartphones

High-Speed Data Readout Minimizes Distortion<sup>\*2</sup> in Still Images, and Enables Super Slow Motion Movie Shooting



Data read-out is slow, limiting the frame rate to be ~ 100 FPS max.

In-sensor DRAM stores high-speed captures to enable high speed capture (~1,000 FPS)

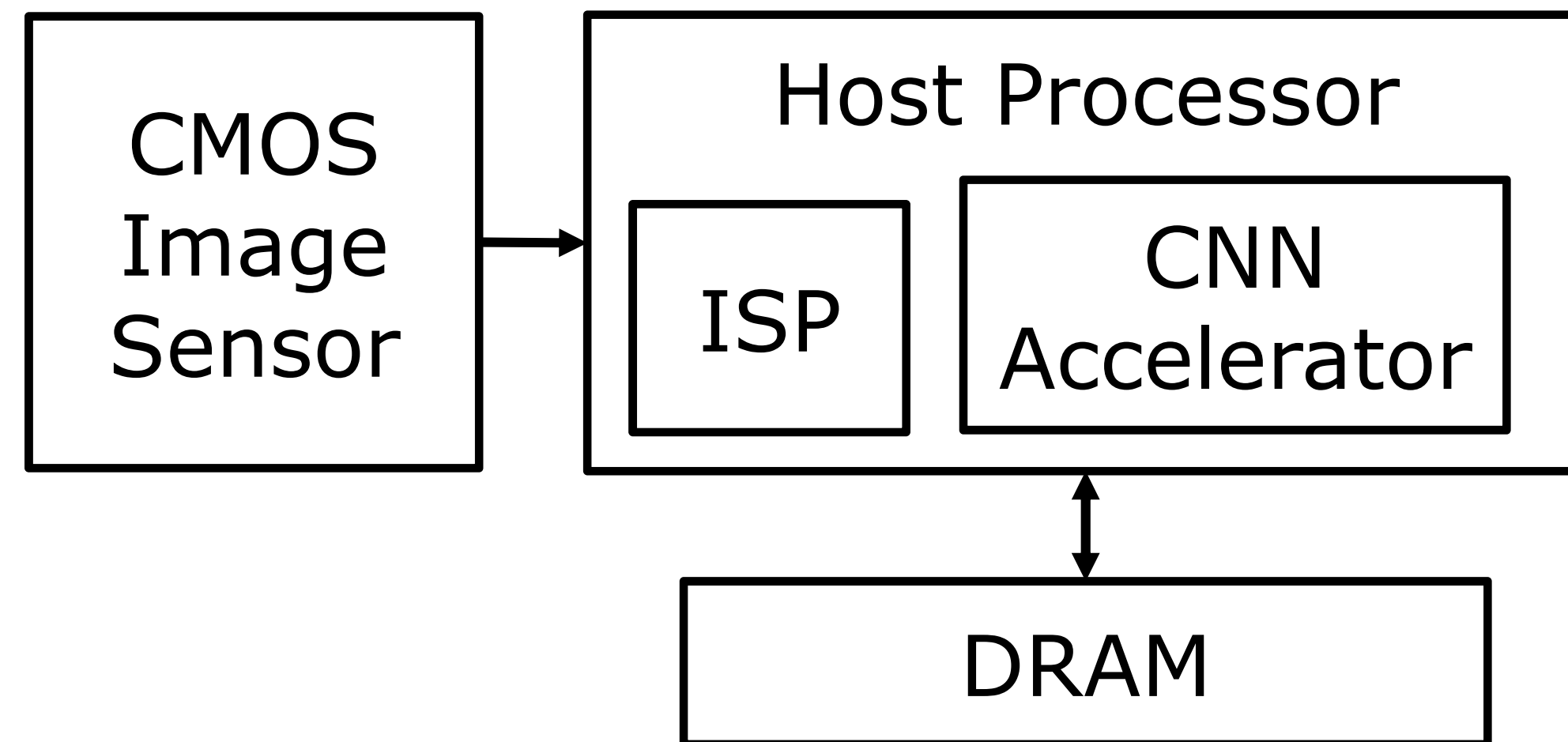
**25 Mbps to CPU**

T. Haruta, ISSCC 2017

# Sony to Release World's First Intelligent Vision Sensors with AI Processing Functionality\*1

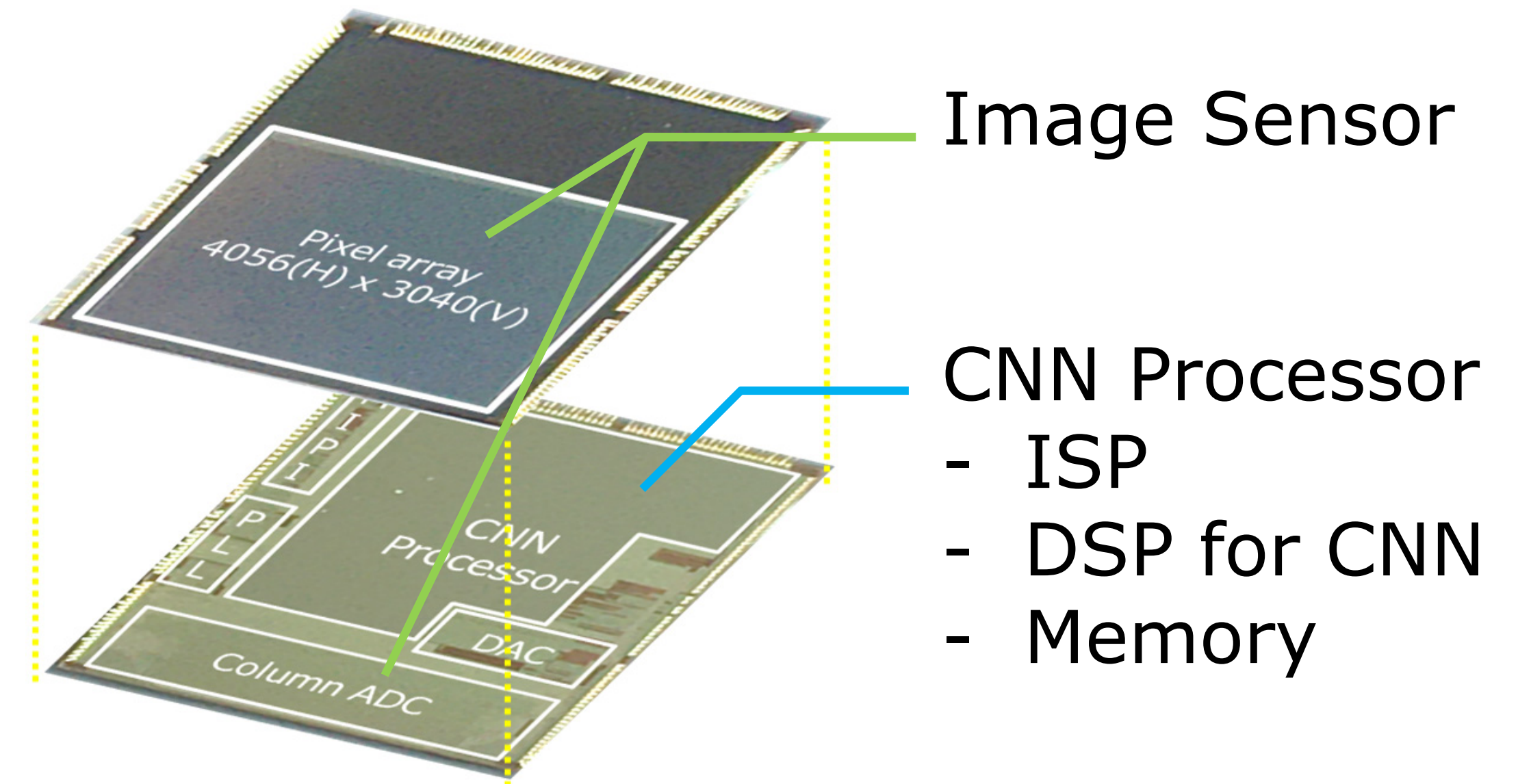
Enabling High-Speed Edge AI Processing and Contributing to Building of Optimal Systems Linked with the Cloud

## Conventional Configuration



Larger, higher power and more costly

## Intelligent Vision Sensor



All functions are implemented on chip

To be presented by R. Eki, ISSCC 2021