Imperial College London

14 – Technology of Vision

Prof Peter YK Cheung

Dyson School of Design Engineering

URL: www.ee.ic.ac.uk/pcheung/teaching/DE4_DVS/
E-mail: p.cheung@imperial.ac.uk

The materials in this Lecture is not based on any textbook, but from various industrial sources and device data.

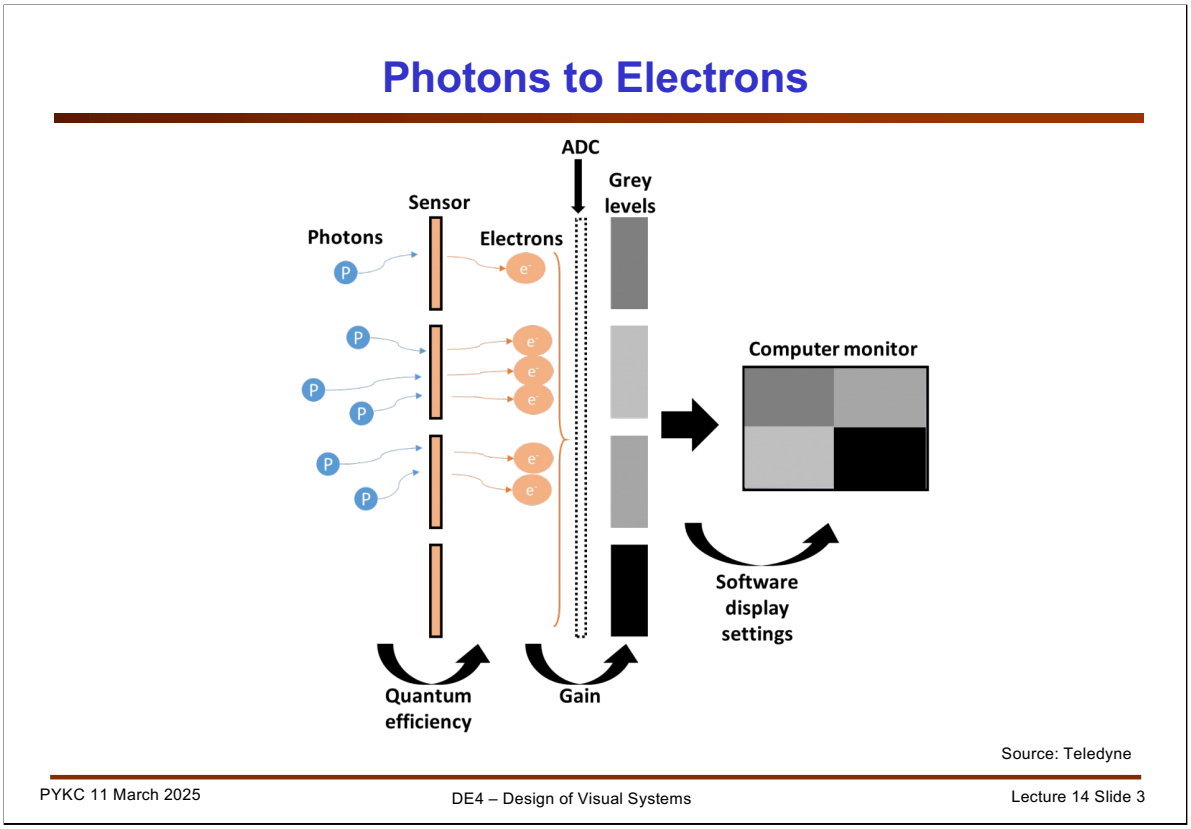
Three Categories of Vision Technology

- Sensors for vision
  - Visible light image sensors
  - IR image sensors
  - Time of Flight (ToF) image sensors
- Display Technology
  - Thin-film Transistor (TFT) Liquid Crystal Displays(LCD)
  - In-Plan Switch (IPS) Liquid Crystal Displays (LCD)
  - Organic LED (OLED) Displays
  - Digital Light Processing (DLP) Projectors
- Visual Data Processors
  - Graphics Processing Units (GPU)
  - AI/Neuro Processors

Three categories of technologies will be considered. They are:

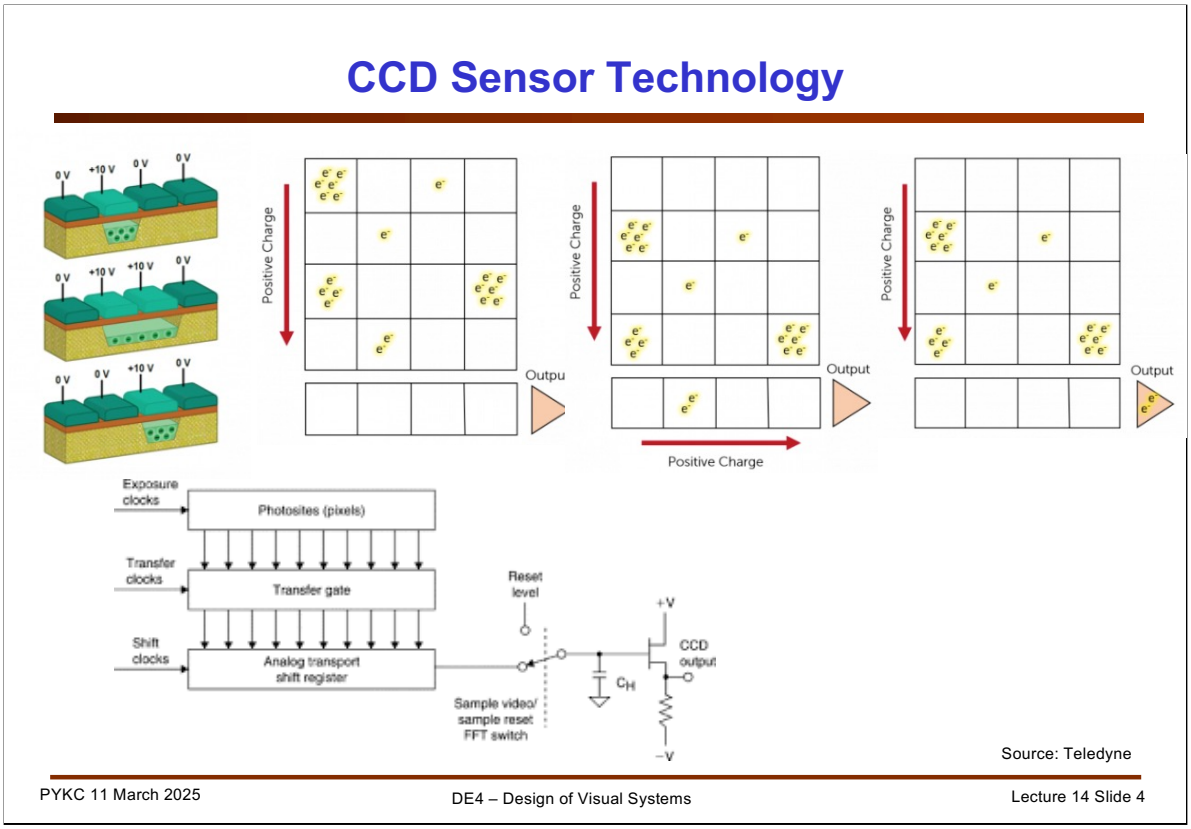
1. Image sensors;
2. Display technologies;
3. Integrated circuit processing technologies.

Due to the limitation of time, the coverage will not be comprehensive. However, I have chosen the more relevant technologies for Design Engineers to discuss.



There is a parallel between how image sensor works and how retina converts photons into electric impulses. When exposed to light, each pixel of the image sensor convert photons that hit the sensor into electrons. This provides a current proportional to the amount of light received. The steps are:

1. Photons hit the sensor are converted into electrons (called photoelectrons).
    - The rate of this conversion is known as quantum efficiency (QE). With a QE of 50%, only half of the photons will be converted to electrons and information will be lost.
2. The generated electrons are stored in a well in each pixel, giving a quantitative count of electrons per pixel
    - The maximum number of electrons that can be stored in the well is known as the full well capacity, which determines the dynamic range of the sensor.
3. The electron charge in each pixel's well is amplified into a voltage, this is the analogue signal.
4. The analogue signal is converted from a voltage into a digital signal with an analogue to digital converter (ADC). This arbitrary digital signal is the grey level of the image.
    - The rate of this conversion is known as gain. With a gain of 1.5, 100 electrons are converted to 150 grey levels.
5. The bit-depth of the camera determines how many grey levels are available for the signal to be converted into, a 12-bit camera has 4096 (2^{12}) available grey levels, a 16-bit camera has 65,536 (2^{16}).
    - The bit depth also determines the full well capacity and therefore the dynamic range.
6. The map of grey levels of the image is then processed by a computer with various algorithms and subsequently displayed on the computer monitor as the visual output.



A **Charge-Coupled Device** (CCD) is an imaging device that detects photons, converts them into photoelectrons and moves electrical charge. They are comprised of a silicon surface onto which an integrated circuit is etched. This etched surface forms an array of pixels which collect incoming photons, generating photoelectrons.

These photoelectrons have a negative charge, so can be shifted along the sensor to readout registers where they can be amplified and converted into digital grey levels. This process is called charge transfer. When the electrons are shifted down the sensor to the readout register, they are shifted horizontally off the register onto an output node. They are then transferred to a capacitor, an amplifier, and analog-to-digital converter, and finally the imaging software in which the image is processed.

The advantages of CCD sensors are:

1. **Excellent low-light sensitivity** – in this sensor, CCD is like rods. It has high quantum efficiency and generate less noise (than CMOS sensors).
2. **Better image fidelity** – CCD sensors use centralised ADC and therefore produce more consistent and uniform output and sharper images (than CMOS sensors).

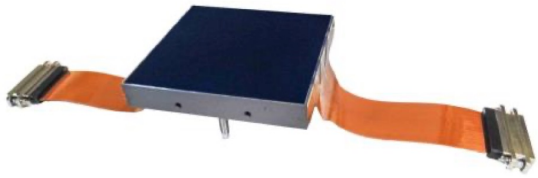
The disadvantages of CCD sensors ae:

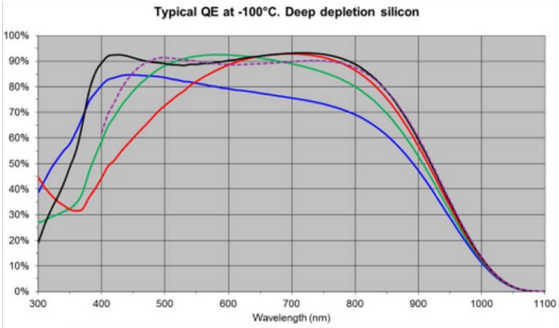
1. **Consumes more power** - they just do!
2. **Slower readout speed** – due to serial readout.
3. **Higher cost** – they involve more costly processes and larger devices.
4. **Susceptible to 'blooming' effect** – When overexposed, a pixel sensor spills over the neighbours and this distorting the image.

Because of these disadvantages, CCD camera are usually found in **scientific**, **medical** and **astronomical** imaging applications where the imagining environment is known and cost of sensor is less important than the quality of the image.

Example of a CCD Sensor

Teledyne CCD290-99 Scientific CCD Sensor, 9216 x 9232 Pixels





Number of pixels	9216 (H) × 9232 (V)
Pixel size	10 µm square
Image area	92.2 mm × 92.4 mm
Outputs	16
Package size	98.5 × 93.7 mm
Package format	Silicon carbide with two flexi connectors
Focal plane height, above base	20.0 mm
Connectors	Two 51-way micro-D
Flatness	20 µm (peak to valley)
Amplifier sensitivity	7.5 µV/e^-
Read-out noise	4 e^- at 0.5 MHz 2.5 e^- at 50 kHz
Maximum pixel data rate	3 MHz
Charge storage (pixel full well)	90,000 e^-
Dark signal	4 e^-/pixel/hour (at –100 °C)

Source: Teledyne

The Teledyne CCD290 scientific CCD sensors was been designed to provide a large image area for demanding astronomical and scientific imaging applications. It uses a technique known as “Back-illumination” to achieve both very low read-out noise and exceptional sensitivity.

The sensor has an image area having 9216 × 9232 pixels with registers at both top and bottom each with eight outputs for short read-out times. The pixel size is 10 µm square. The image area has two separately connected sections to allow full-frame or split full-frame read-out modes. Depending on the mode, the read-out can be through 8 or 16 of the output circuits. A fixed-barrier dump drain is also provided to allow fast dumping of unwanted data.

The output amplifier is designed to give very low noise at read-out rates of up to 3 MHz. The low output impedance simplifies the interface with external electronics and the optional dummy outputs are provided to facilitate common mode rejection.

The package provides a compact footprint with guaranteed flatness at cryogenic temperatures. Connections are made at the top and bottom of the device via two flexi connectors that also provide a thermal break. The sides may be close butted if needed.

Specifications are tested and guaranteed at 173K (–100 °C).

By changing the coating, the sensor can have different spectral responses.

(Taken from Teledyne’s datasheet.)

Example of a CCD Camera

DXM 1200 Digital Eclipse Hardware

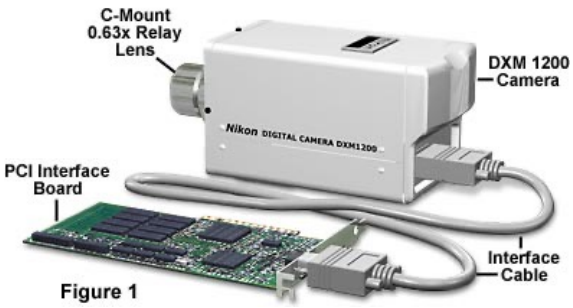
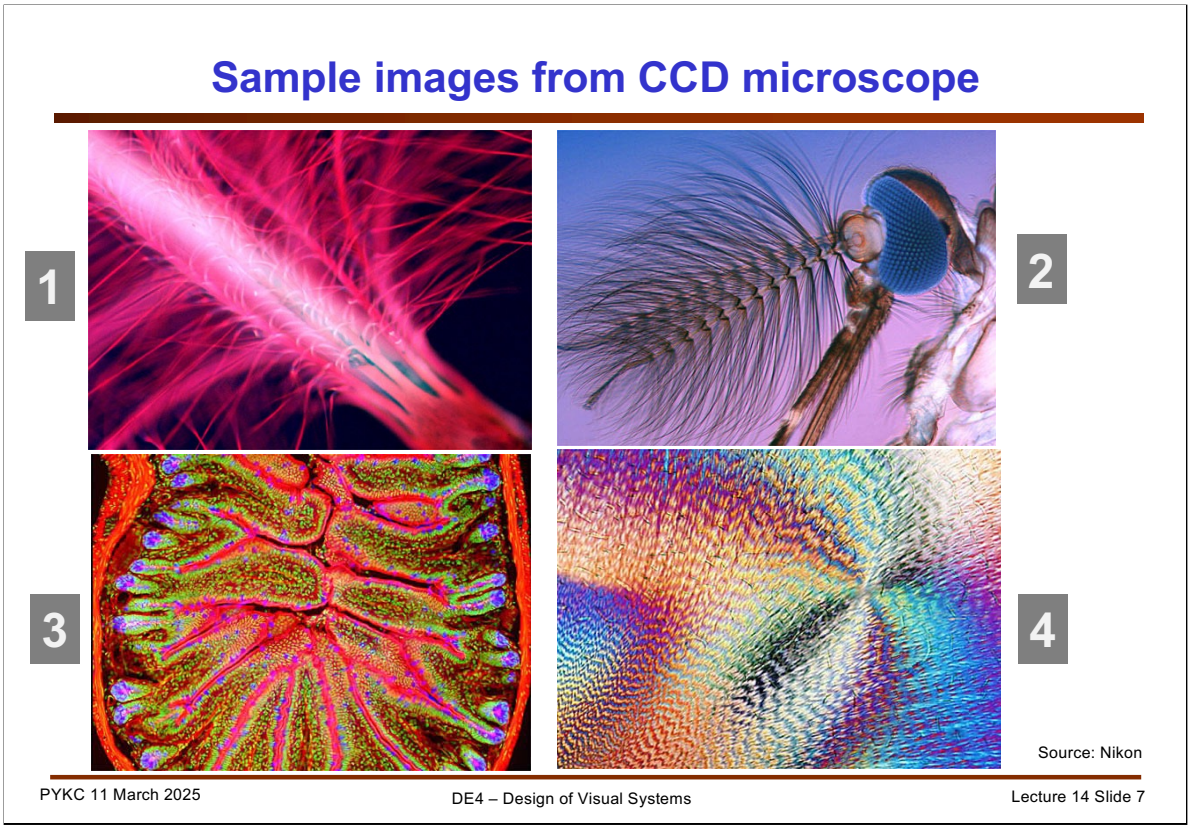


Figure 1

Specification	Nikon DXM 1200
Maximum Pixel Output	12 Million
CCD	Sony ICX085AK (Bayer Mosaic Filters)
Pixel Size (Shape)	6.7 Microns (Square)
CCD Chip Size	10.0 (H) x 8.7 (V) (Millimeters)
CCD Cooling	None
Fastest Shutter Speed	1/12000 Seconds
Slowest Shutter Speed	170 Seconds
Sensitivity Setting	3 Levels (Normal, High, Max
Exposure Control	Manual
Signal/Noise	50 dB
Dynamic Range	8 Bits (48 dB)
Dark Current	4 Electrons/Pixel/Second
Read Noise	16 Electrons/Pixel
Well Depth	16,000 Electrons
Resolution	1800 TV Lines
Live Image Display Rate	12 Frames/Second
Maximum Image Size	3840 x 3072 Pixels
Image Storage Format	TIF, BMP, JPG
Lens Mount	C-mount
Computer Interface	PCI BusMaster

Source: Nikon

This is another CCD based scientific camera by Nikon. It is not as high resolution as the Teledyne device, but it produces beautiful images as will be seen on the next slide.



These are four images taken with Nikon's Digital Eclipse Microscope equipped with Sony ICX085AK CCD sensor. (https://www.microscopyu.com/galleries/digital-imaging).

1. Down Feather

Feathers have an exquisite beauty and functionality that has captured the attention and imagination of people for centuries. They are specialized epidermal growths, formed by papillae that are composed of keratin, lipids, and pigments that give them their brilliant colors. Keratin is an ideal material for feathers because it is lightweight and flexible, yet strong enough to form a structure that can withstand the rigors of flight.

2. Mosquito Antennae

The male mosquito has large bushy antennae, which he uses to listen for the buzz of a potential mate. He responds only to the humming frequency given by a female of the same species and will fly in the direction of the sound to mate with her. Male mosquitoes do not bite, but feed on plant juices and flower nectar. Only female mosquitoes bite animals and require a blood meal.

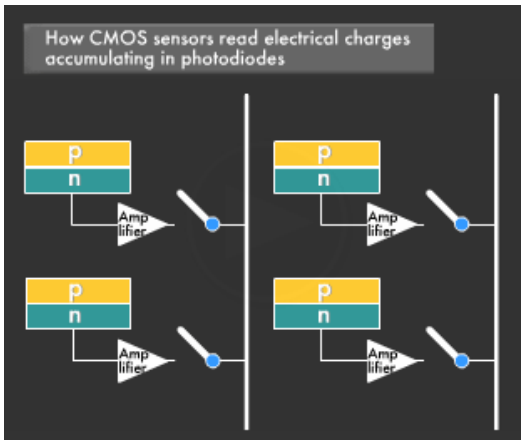
3. Mouse Intestine

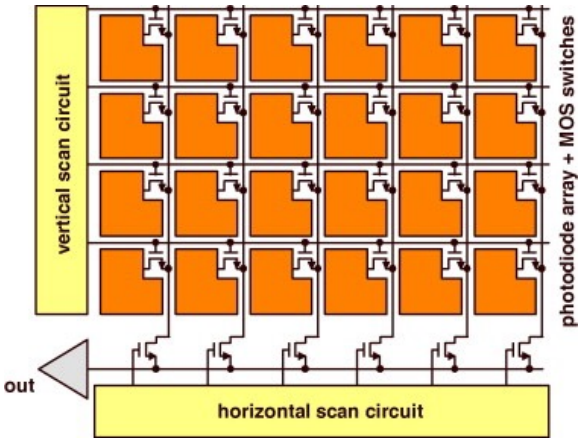
Laboratory mice are special breeds of house mice and are used in many scientific experiments because of their close mammalian relationship to humans. Compared to larger mammals, mice and other rodents are small, easy to handle, inexpensive to house, and breed quickly. During the late 20th Century (and on into the current century), scientists bred different strains of mice with genetic deficiencies in order to produce models for human diseases.

4. Vitamin C Crystallites

Ascorbic acid, also known by the chemical name L-ascorbic acid, is a water-soluble vitamin that functions as a powerful antioxidant. Although most animals can synthesize vitamin C, others — such as including humans and other primates and guinea pigs — obtain it only through their diets. Vitamin C is commonly found naturally in peppers, citrus fruits, tomatoes, melons, broccoli, and green leafy vegetables such as spinach, turnip, and mustard greens.

CMOS Sensor – photodiode array





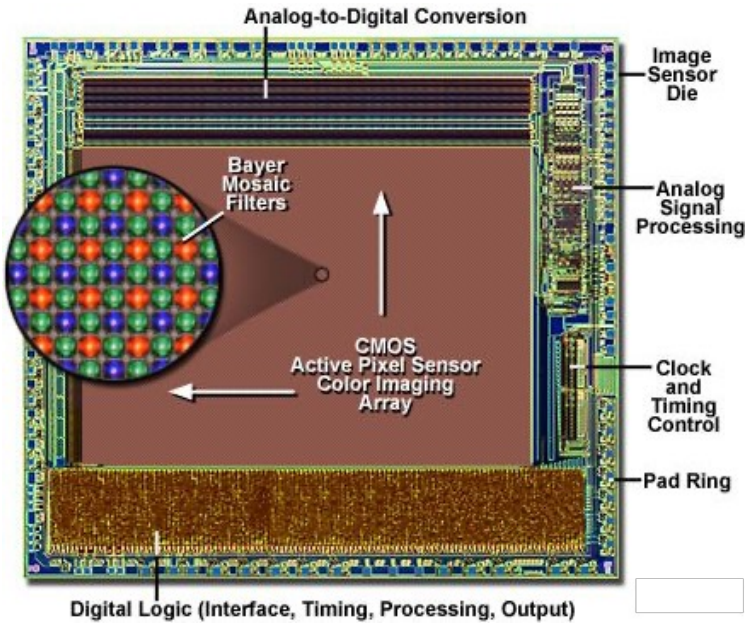
Source: Canon

The basic structure of CMOS (**Complementary Metal-Oxide Semiconductor**) sensor arrays consists of pixel cells of light sensitive photodiodes as shown in the slide. The p-n materials doped in silicon form a p-n junction photo diode. When photons fall on these pixel cells, charges are accumulated and amplified.

A CMOS image sensor structure is very much like that of a silicon SRAM – it has a 2D grid of pixel cells with row scanning circuit (vertical) and horizontal scanning circuits. Unlike CCD sensors, the pixel values can be read randomly although this is very rarely done. However, some sensors with high output data rate could read the pixel columns in parallel instead of serially.

CMOS sensors, in spite of its potential higher noise, is the predominant technology in image sensor today. It can be found in almost all types of applications, replacing CCD sensors.

Structure of a CMOS Sensor Device

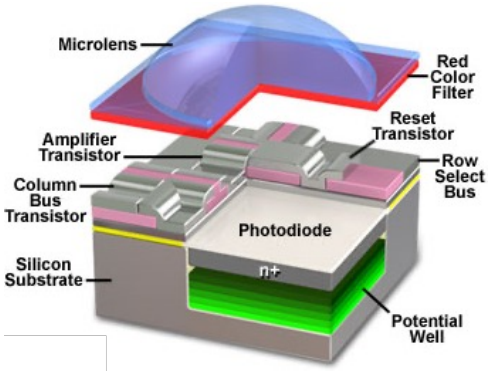


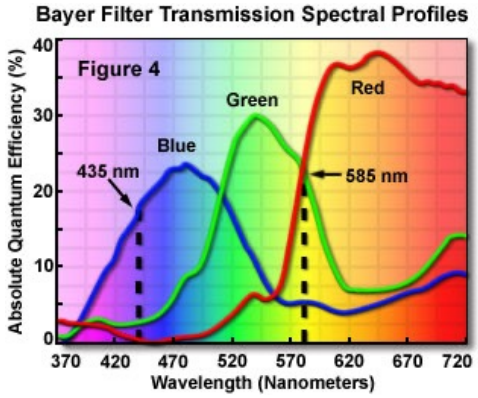
Source: Olympus

Here is a chip photograph of a typical CMOS sensor. CMOS sensors can detect colour by having built in colour lens put in front of the photodiodes. This is known as the Bayer Mosaic filter with one red, one blue and two green lens, forming a 2x2 grid.

While the photodiode array occupies the majority area of the silicon die, a typical CMOS sensor also has the built-in digital interface and timing logic, the analogue-to-digital converters converting the analogue voltages from the pixel array to digital numbers and various other logic cirucits.

Bayer Mosaic Filter arrangement





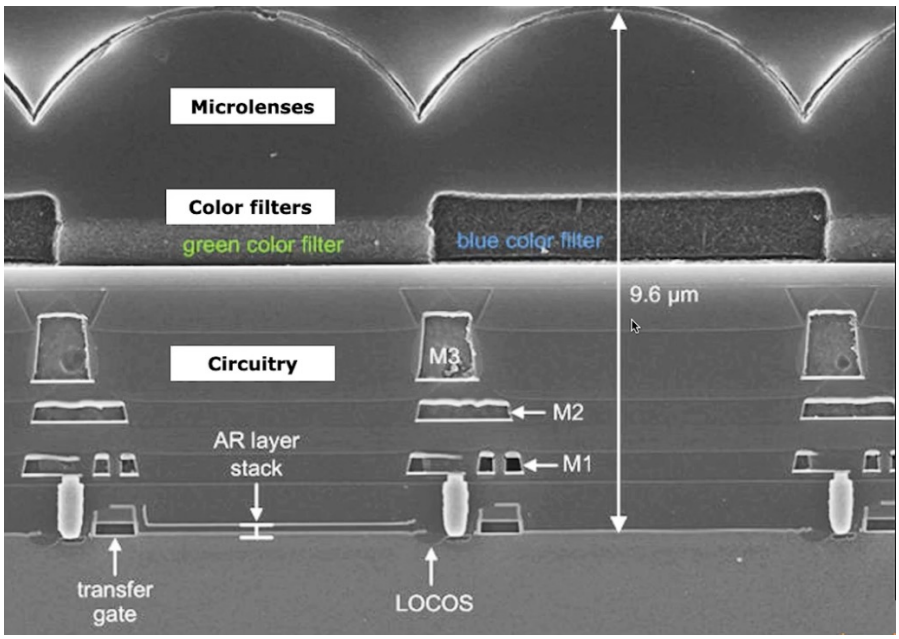
Source: Olympus

The colour sensing capability of a CMOS sensor is achieved with a 2x2 grid of colour lens as shown in the slide. The microlens is fabricated onto the appropriate colour filter with the semiconductor sensor and circuit underneath.

The 3D cross-section diagram here shows that the transistor circuits and wiring is sandwiched between the light sensing photodiode and the lens.

By changing the colour filter, it is possible to change the spectral sensitivity of each sub-pixel cell as shown in the plot above.

SEM image of Cross Section of a CMOS Sensor



This is a Scanning Electron Microscope image of the cross-section of a CMOS sensor showing a blue and a green pixel cell.

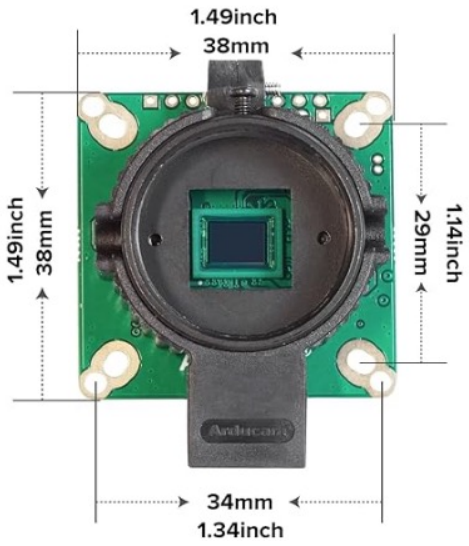
Low Cost Camera - Raspberry Pi

	Camera Module v1	Camera Module v2	Camera Module 3	Camera Module 3 Wide	HQ Camera	GS Camera
Net price	$25	$25	$25	$35	$50	$50
Size	Around 25 × 24 × 9 mm	Around 25 × 24 × 9 mm	Around 25 × 24 × 11.5 mm	Around 25 × 24 × 12.4 mm	38 x 38 x 18.4mm (excluding lens)	38 x 38 x 19.8mm (29.5mm with adaptor and dust cap)
Weight	3g	3g	4g	4g	30.4g	34g (41g with adaptor and dust cap)
Still resolution	5 Megapixels	8 Megapixels	11.9 Megapixels	11.9 Megapixels	12.3 Megapixels	1.58 Megapixels
Video modes	1080p30, 720p60 and 640 × 480p60/90	1080p47, 1640 × 1232p41 and 640 × 480p206	2304 × 1296p56, 2304 × 1296p30 HDR, 1536 × 864p120	2304 × 1296p56, 2304 × 1296p30 HDR, 1536 × 864p120	2028 × 1080p50, 2028 × 1520p40 and 1332 × 990p120	1456 x 1088p60
Sensor	OmniVision OV5647	Sony IMX219	Sony IMX708	Sony IMX708	Sony IMX477	Sony IMX296
Sensor resolution	2592 × 1944 pixels	3280 × 2464 pixels	4608 x 2592 pixels	4608 x 2592 pixels	4056 x 3040 pixels	1456 x 1088 pixels
Sensor image area	3.76 × 2.74 mm	3.68 x 2.76 mm (4.6 mm diagonal)	6.45 x 3.63mm (7.4mm diagonal)	6.45 x 3.63mm (7.4mm diagonal)	6.287mm x 4.712 mm (7.9mm diagonal)	6.3mm diagonal
Pixel size	1.4 µm x 1.4 µm	1.12 µm x 1.12 µm	1.4 µm x 1.4 µm	1.4 µm x 1.4 µm	1.55 µm x 1.55 µm	3.45 µm x 3.45 µm

This chart provides an overview of what is currently available in the low-cost sensor range sold to interface with Raspberry Pi. Note that Sony CMOS sensors dominate the range here with resolution between 1.6M pixels to 12M pixels, and the cost is around US$50 (not including the lens).

Let us examine one such camera, the HQ model with Sony IMX477 sensor in a bit more detail.

Low Cost Camera - Raspberry Pi



Resolution	4056 (H) x 3040 (V)
Megapixels	12.3MP
Supply Voltage	1.05 V (Digital), 1.8 V (Interface), 2.8 V (Analog)
Chroma	Color, RGB
Shutter Type	Rolling Shutter Only
Frame Rate	15 to 240 fps, depending on the video mode
ADC resolution	10-bit, 12-bit
Pixel Size	1.55 µm (H) x 1.55 µm (V)
Interface	MIPI 2-lane/4-lane, D-PHY V1.2
Data Rate	Max. 2.1Gbps/lane
Clock Frequency	6-27 MHz

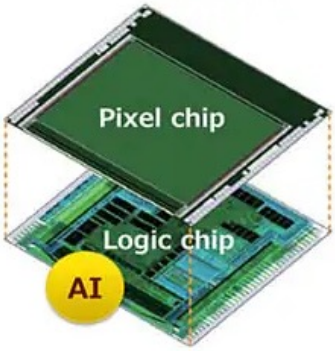
Source: Raspberry Pi Foundation

This is one of Sony's popular sensor with 12.3MP and with a native resolution of 4056 x 3040 resolution at 60 frames per second. The frame rate is between 15 to 240 frames per second. At maximum frame rate, the resolution is down to 1080p (or 1920 x 1080). According to Sony's datasheet, the frame rate and resolution (10-bit or 12-bit per pixel) is given by the following table.

Another aspect to note is that the sensor has built-in Camera Serial Interface 2 (CSI-2). This is also known as a 2-lane MIPI (mobile industry processor interface) standard. This is designed for fast data exchange between the camera and the microprocessor (such as the Raspberry Pi).

Drive mode	Number of active pixels	Maximum frame rate [frame/s]	Output interface
Full (4:3) (Normal)	4056 (H) × 3040 (V) approx. 12.33 M pixels	60	CSI-2
		40	CSI-2
Full (4:3) (DOL-HDR)	4056 (H) × 3040 (V) approx. 12.33 M pixels	DOL 2 frame : 30 DOL 3 frame : 15	CSI-2
Full (16:9) 4K2K (Normal)	4056 (H) × 2288 (V) approx. 9.28 M pixels	79	CSI-2
Full (16:9) 4K2K (DOL-HDR)	4056 (H) × 2288 (V) approx. 9.28 M pixels	DOL 2 frame : 39 DOL 3 frame : 19	CSI-2
Full (4:3) Binning (Normal)	2028 (H) × 1520 (V) approx. 3.08 M pixels	179	CSI-2
Full (16:9) Binning 1080P (Normal)	2028 (H) × 1128 (V) approx. 2.29 M pixels	240	CSI-2
Full (16:9) Binning 720P (Normal)	1348 (H) × 750 (V) approx. 1.01 M pixels	240	CSI-2
Full (16:9) Scaling 1080P (Normal)	2024 (H) × 1142 (V) approx. 2.31 M pixels	79	CSI-2
Full (16:9) Scaling 720P (Normal)	1348 (H) × 762 (V) approx. 1.03 M pixels	79	CSI-2

Sony's Latest Sensor



Intelligent vision sensor stacked configuration

<Main functions on the logic chip>

- ✓ Conventional image sensor operation circuit
- ✓ ISP which processes the image signal
- ✓ Original DSP dedicated to AI signal processing
- ✓ Memory for the AI model

⇒ Eliminates the need for high-performance processor or external memory

Source: Sony

Sony introduced an image sensor with AI processing functionality in May 2020. It realizes high-speed edge AI processing. The pixel chip is back-illuminated, and the logic chip has a number of features in addition to the conventional image sensor operation circuit, such as Sony's original DSP (Digital Signal Processor) and memory to store the AI model of users' choice. This configuration allows a single image sensor to handle everything from image capturing to AI processing without high-performance processors and external memory.

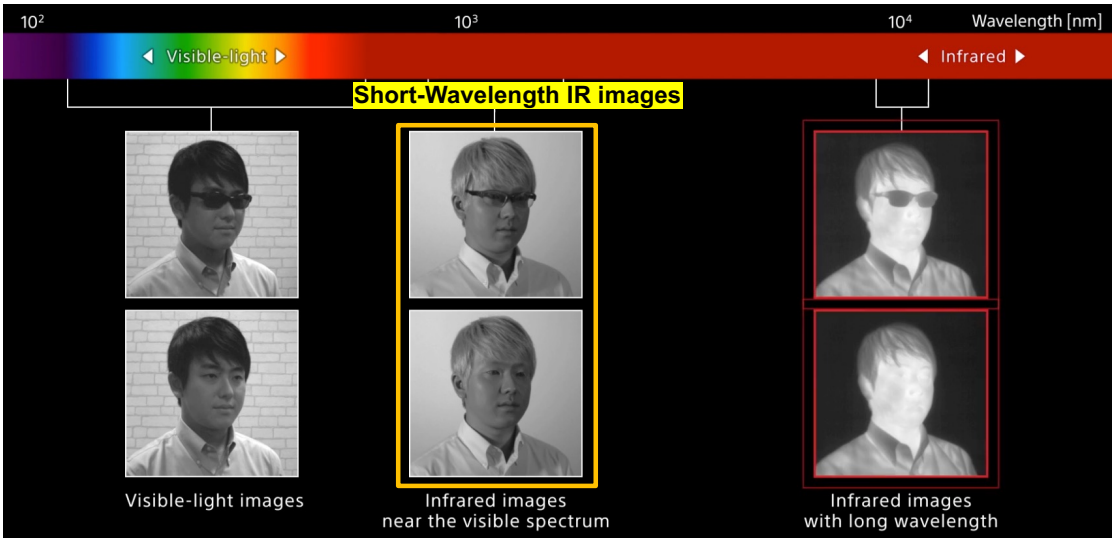
Integration of AI into an Image Sensor

Source: Sony

Integrating the image sensor onto logic processing chip has many advantages.

Extracting only the necessary data can reduce data transmission latency and power consumption of the camera and address privacy concerns when using cloud services. The reduced data usage of these image sensors will also help lower the power consumption of the cloud server. AI models can be rewritten and updated with the latest AI models to match various system environments and conditions. In addition to making the development of tiny cameras with AI functionality possible, these image sensors are expected to enable various applications in the retail and industrial equipment industries, and combined with cloud computing, will contribute to realizing optimal systems.

Infrared (IR) and SWIR Images



Source: Sony

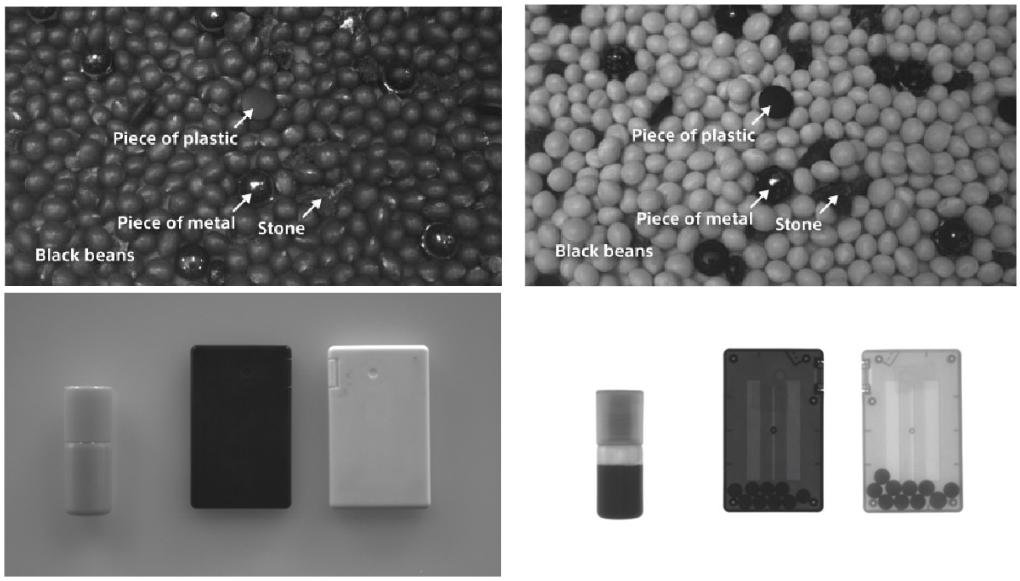
Generally, light with a wavelength of 400 nm to 780 nm is referred to as visible light, and light with a wavelength of 780 nm to 106 nm as infrared light. The wavelength band of SWIR is from 900 nm to 2,500 nm, which is the region of infrared light closest to visible light. Sony has developed what they called the SenSWIR technology with cameras capable of broad imaging over the range of 400 nm - 1,700 nm, including visible light as well as SWIR light.

SWIR (Short Wavelength Infra-Red) light penetrates and is absorbed by different substances other than visible light, so its attributes can be applied in a variety of different situations.

The slide above shows images from a conventional visible light camera, a conventional IR camera, and Sony's new SWIR camera.

Infrared (IR) and SWIR Images

Example of detecting contaminants in food



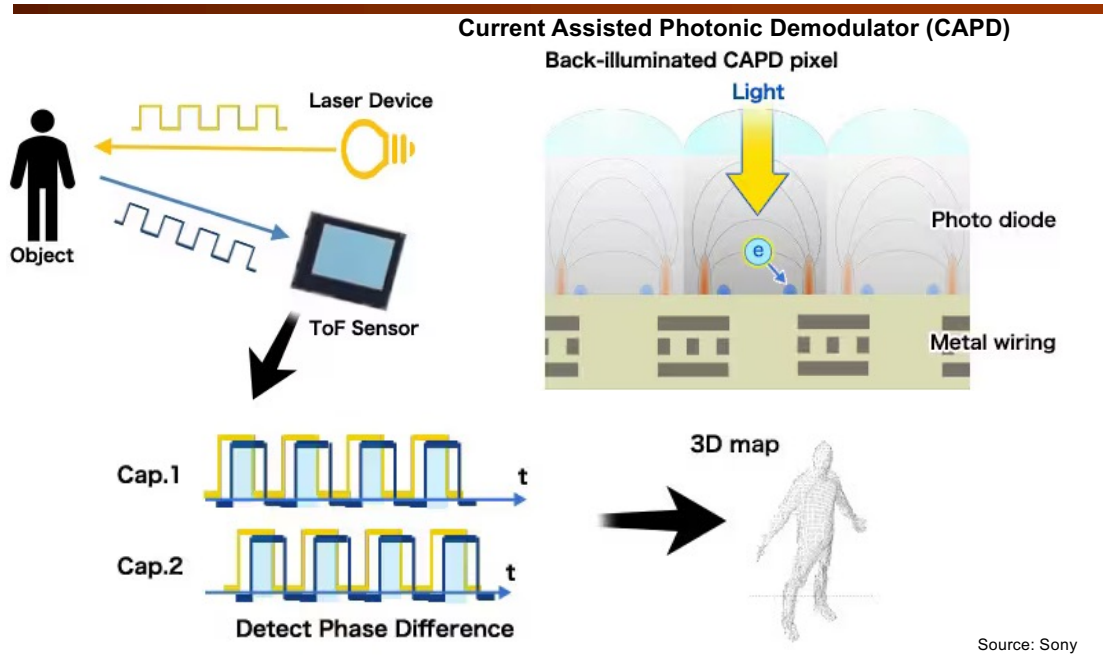
Source: Sony

There are many possible applications for a SWIR camera, particularly in the domain of automated inspection.

Above are two examples: pieces of plastics are detected inside a bowl of black beans. The bottom example is how the contents of containers are revealed under a SWIR camera. Here is a list of possible applications.

- Sorting fruits and vegetables
- Container content inspections
- Contaminant detection
- Sorting materials
- Positioning in Semiconductor Manufacturing
- Temperature monitoring
- Firefighting
- Remote monitoring
- Observation of agricultural lands

Time of Flight (ToF) Image Sensor



So far, we have only considered image sensors that sense visible light directly. There are other types of image sensors that produces different type of images.

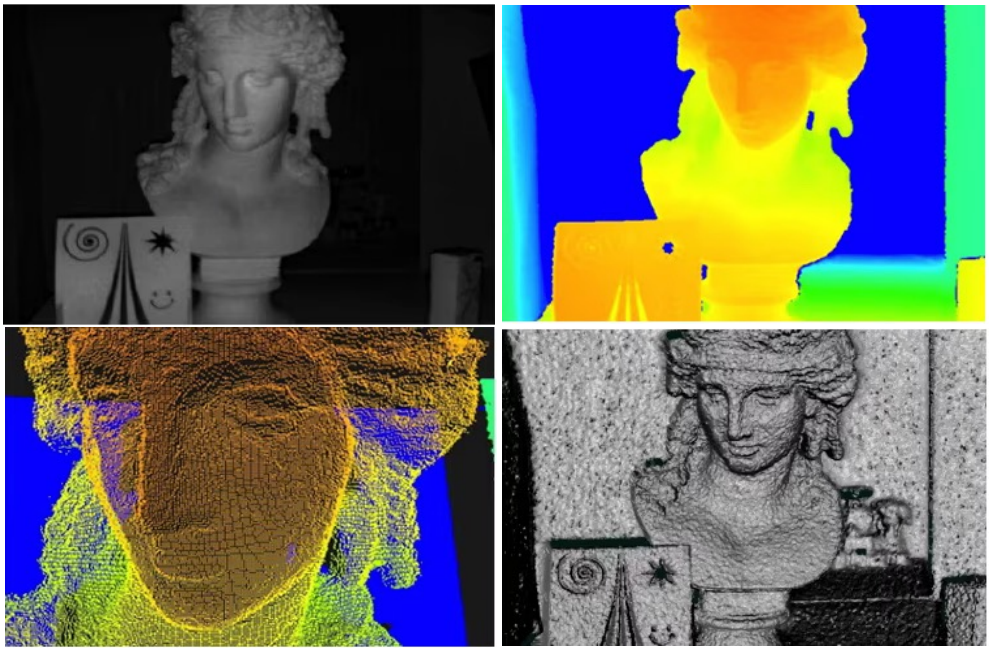
With ToF (**Time of Flight**) image sensors, depth information is acquired by measuring the time it takes for an emitted light pulse to reflect and return to the sensor surface. There are various types of ToF. Sony focused on developing an indirect-ToF (iToF) method that works by measuring the phase delay of returning light after it has been reflected by the target object.

To realize this sensor, Sony developed a new back-illuminated technology which they called "**Current Assisted Photonic Demodulator**" or CAPD.

Making the most of the back-illuminated structure, it efficiently converts light into electrons, enabling time detection of under 50 picoseconds and greatly improving distance resolution. It allows for miniaturization of 3D camera modules.

This technology can be applied to AR (augmented reality) and VR (virtual reality) through real-time acquisition of high-resolution 3D depth information. It is used for face recognition in smartphones, features related to AR and VR, and autonomous robots and drones.

Time of Flight (ToF) Image Sensor



Source: Sony

Here are the outputs generated from such a ToF sensor: Sony's IMX316, which has a resolution of 184 x 244 pixels.

Clockwise from top left:

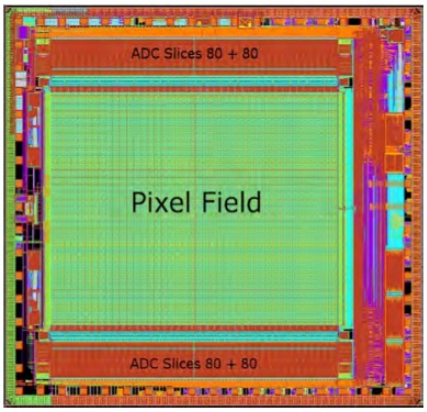
(1) Monochrome image from a conventional image sensor,

(2) Depth Heat Map with distance measured from the ToF sensor and colour coded (blue = furthest distance),

(3) Point cloud display with a collection of points of data plotted in a 3D space, where the points extracted from abrupt changes of distance,

(4) 3D display by combining the monochrome image with the ToF image to provide a 3D sense of depth.

Another ToF Sensor – epc660



epc660

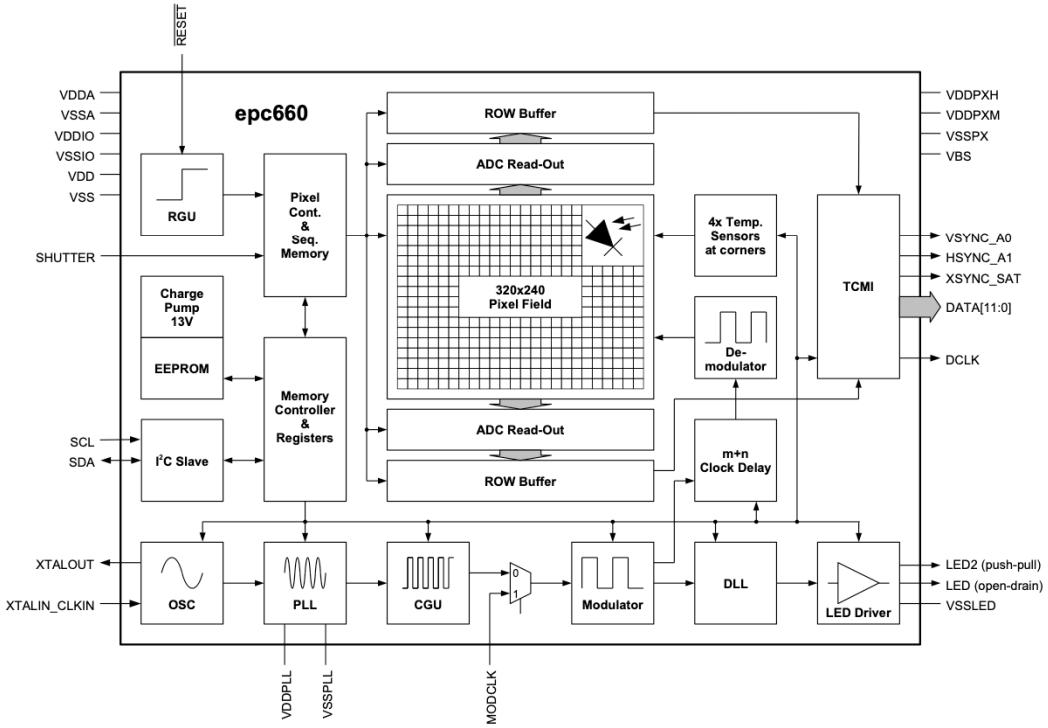
3D TOF imager 320 x 240 pixel

Main Features

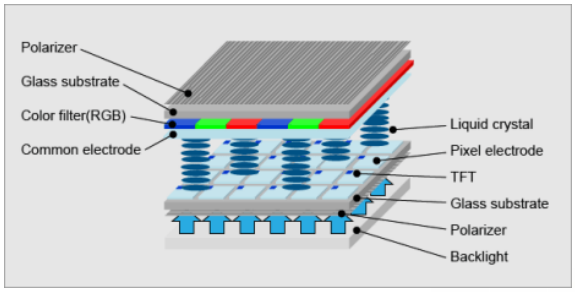
- **General**
  - 3D TOF imager in full monolithic design
  - 320 x 240 pixel-field, backside illuminated
  - QE >80% @ 850nm
  - Full well capacity 8'000 ke- (ambient and signal)
  - 39 fps full 3D TOF frame rate, single frame rate up to 158fps
  - Region of interest setting allows up to several kfps
  - 4 integrated temperature sensors
- **Measurement performance**
  - Absolute accuracy in the sub-centimeter range with appropriate setup and calibration
- **Integrated LED (or laser diode) driver**
  - Laser diode (LD) illumination possible
  - Open-drain LED output pad, up to 200mA drive
  - Push-pull LED2 output pad, up to 50mA drive
- **Parallel digital data interface TCMI**
  - 48MS/s max. data rate, 2.5/3.3V compatible
  - 12/8-bit parallel DATA output + XSYNC/SAT flag
  - VSYNC, HSYNC and DCLK outputs
- **I²C control interface (slave)**
  - 400kHz (FM) / 1MHz (FM+)
- **Integrated EEPROM 128 x 8-bit**
  - Calibration data and user programmable parameters
  - Unique chip ID

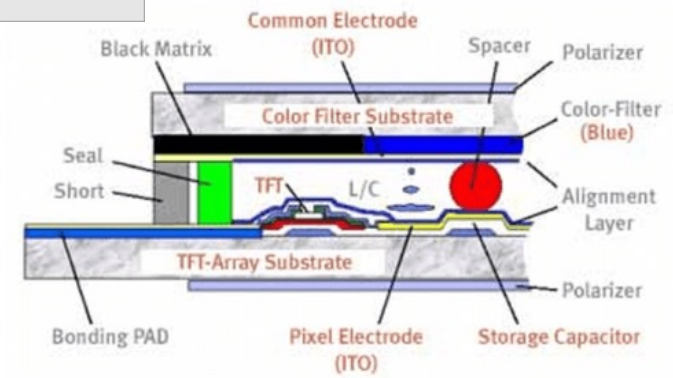
Source: EPC

The epc660 is a fully integrated 3D-TOF imager with a resolution of 320 x 240 pixels (QVGA) made by Espros Photonic Corporation (EPC) from Switzerland. It is a highly integrated system-on-chip camera system. Apart from the actual CCD pixel-field, it includes the complete control logic to operate the device. Data communication is done through a high-speed digital 12-bit parallel video interface. Even for mobile devices, only a few additional components are needed to integrate 3D camera capability. Depending on the system design, a resolution in the millimetre range for measurements up to 100 meters is feasible. 65 full frame TOF images are delivered in maximal configuration. By using the advanced operation modes, this can be boosted up to more than 1000 TOF images per second!



TFT LCD Technology





Source: Orient Display

One of the most common technology is the TFT LCD Display (Thin-Film-Transistor Liquid Crystal Display). This has a sandwich-like structure with liquid crystal material filled between two glass plates. Two polarizer filters, colour filters (RGB) and two alignment layers determine exactly the amount of light is allowed to pass and which colours are created.

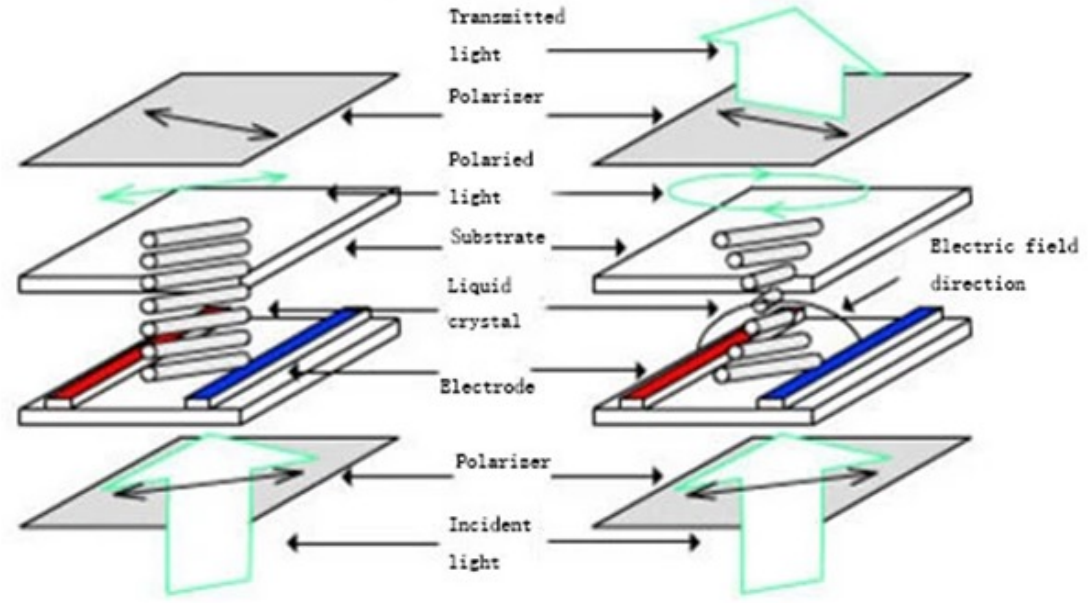
Each pixel in an active matrix is paired with a transistor that includes a capacitor which gives each sub-pixel the ability to retain its charge, instead of requiring an electrical charge sent each time it needed to be changed. The TFT layer controls light flow through a colour filter, displays the colour, and a top layer houses your visible screen.

Utilizing an electrical charge that causes the liquid crystal material to change their molecular structure allowing various wavelengths of backlight to “pass-through”. The active matrix of the TFT display is in constant flux and changes or refreshes rapidly depending upon the incoming signal from the control device.

The pixels of TFT displays are determined by the underlying density (resolution) of the colour matrix and TFT layout. The more pixels the higher detail is available. Available screen size, power consumption, resolution, interface (how to connect) define the TFT displays.

The TFT screen itself cannot emit light (unlike an OLED display), it has to be used with a back-light of white bright light to generate the picture. Newer panels utilize LED backlight (light emitting diodes) to generate their light and therefore utilize less power and require less depth by design.

In-plane-switching (IPS) LCD Technology

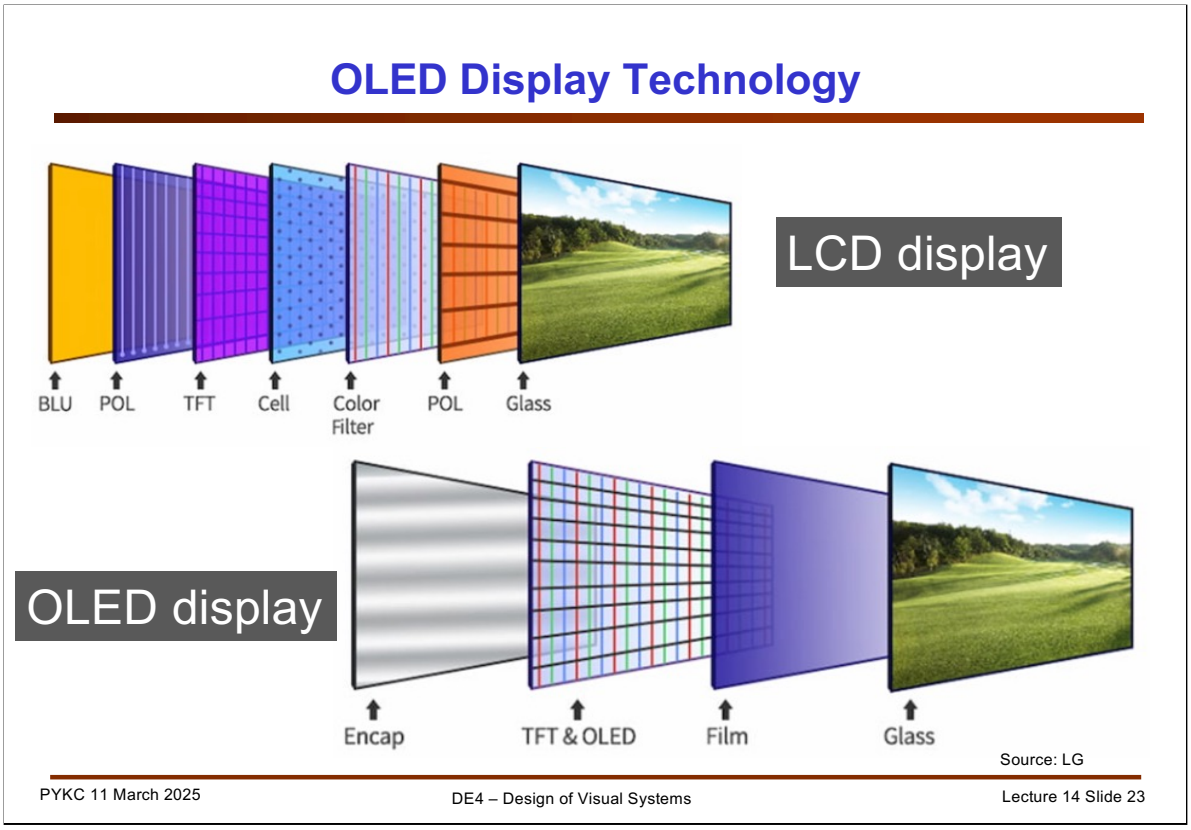


Source: Orient Display

IPS stands for **in-plane-switching technology**. This was introduced in 1992 by Hitachi, and is also one type of TFT LCD display.

The basic LCD structure is similar to the conventional TFT display but the inside display schematic is different. In an IPS LCD panel, when no electric field is applied to the liquid crystal cells, the cells naturally align liquid crystal cells in a horizontal direction between two glass substrates which blocks the transmission of light from the backlight. This makes the display dark and results in a black display screen. When an electric field is applied, the liquid crystal cells are able to rotate through 90° allowing light to pass through resulting in a white display screen. IPS panels have superior image quality, good contrast ratio and wide viewing angles of up to 170° .

IPS panels are well suited for graphics design and other applications which require accurate and consistent colour reproduction.



Organic Light-Emitting Diode (OLED is a display technology in which millions of **light emitting diodes** (LEDs) made with **organic molecules** each emit their own light, not requiring a separate light source.

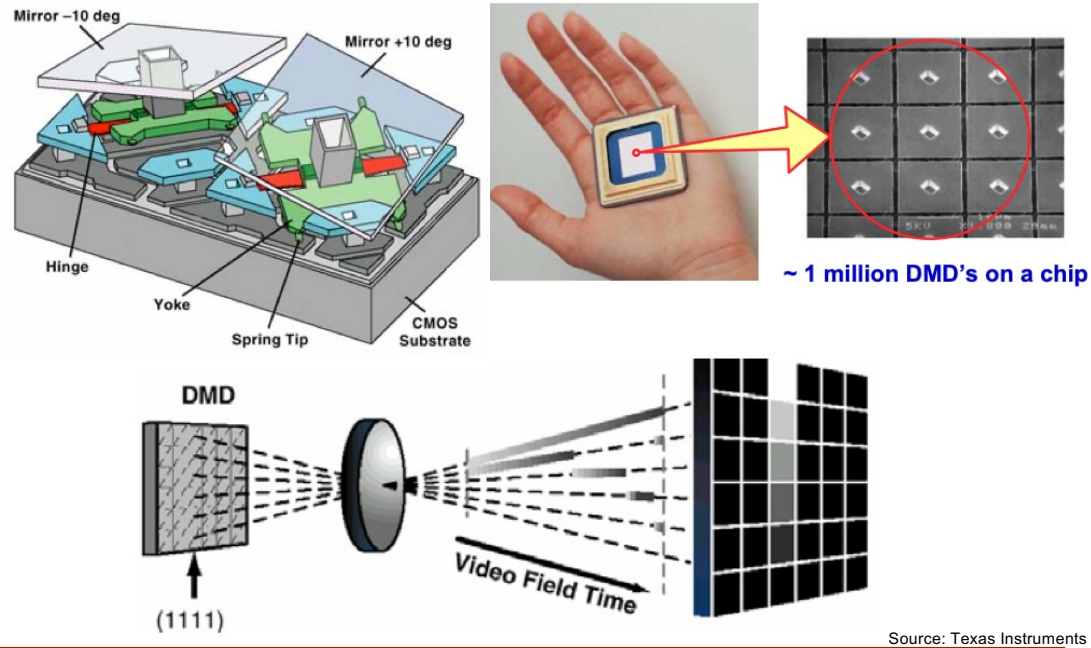
The structure of OLEDs and LCDs are very different. LCDs have a more complex structure than OLEDs and require more component layers. On the other hand, OLEDs have a relatively simple structure and require fewer component layers.

OLEDs require only one **polarizing plate** (POL) that plays a role in transmitting light, while LCDs require two. Since LCDs cannot **emit light on their own**, they require a light source called a backlight unit (BLU). Various other various sheets and components, including polarizing plates, are employed to efficiently use the light from the light source.

In contrast, OLEDs can emit and control light with each component, so they **do not require a backlight** unit or other components used in LCDs. As a result, OLED panels are much **thinner** and **lighter** than LCDs and can be implemented in **various shapes**, like being foldable or rollable. One can even implement **translucent** (i.e. see-through) displays with this technology.

The advantages of OLED are: 1) they sharper image and better contrast, allowing near-perfect blacks; 2) more accurate colour reproduction; 3) no flicker and therefore kinder to eyes; 4) can be made flexible or even transparent; 5) thinner and lighter.

Digital Micromirror Device (DMD) Technology



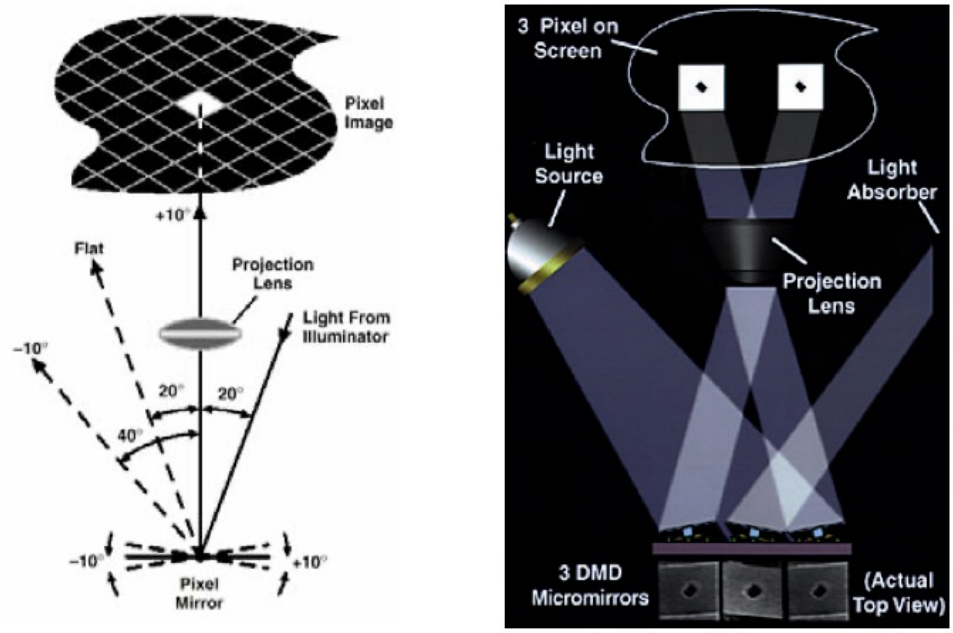
Source: Texas Instruments

In 1987, Texas Instruments invented the Digital Micromirror Device (DMD) that revolutionized the projection of images onto screens. The idea is to make very small rotating mirrors using a technology known as MEMS, or Micro-Electrical-Mechanical Systems. This is a method of making minuet mechanical structures using silicon integrated circuit processing methods onto a chip, often also with integrated electronics to drive the mechanical parts. This technology also gave rise the the low-cost accelerometers and gyroscopes (as used in car air-bag system and in drones).

The mirrors can be rotated at either + 10 degrees or -10 degrees on a hinge and defection is controlled by electronics circuit which switches between 0 and 1. Over a 1 million DMDs were integrated onto a single chip because the mirror was only 16 micron x 16 micron.

In the lower diagram below, the four pixels are turned on and the corresponding mirrors are rotated to light up the four pixels on the screen.

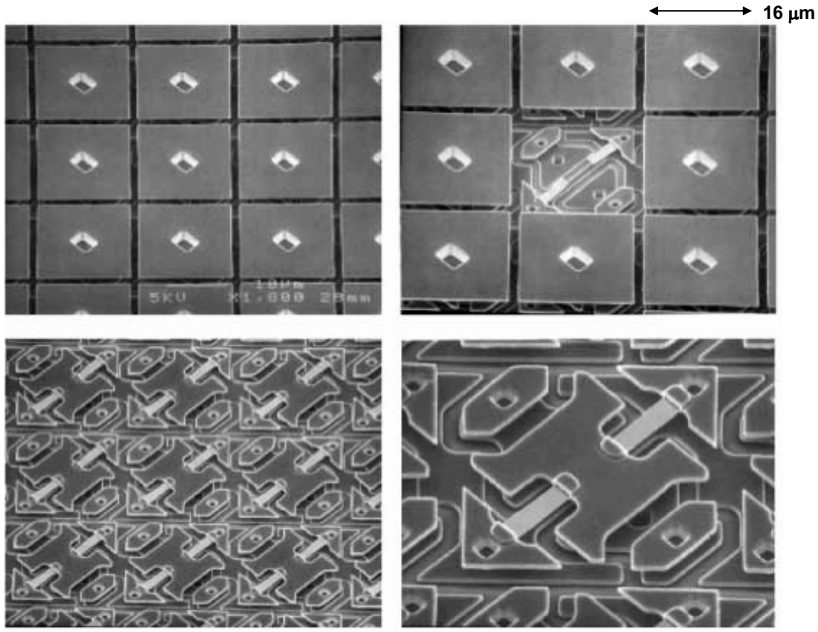
Principle of Projection using DMD



Source: Texas Instruments

Diagram above illustrates the optical switching action of the mirror. When the mirror rotates to its ON state (+10 degrees), light from a projection source is directed into the pupil of a projection lens and the pixel appears bright on a projection screen. When the mirror rotates to its OFF state (-10 degrees), light is directed out of the pupil of the projection lens and the pixel appears dark. Thus, the optical switching function is simply the rapid directing of light into or out of the pupil of the projection lens.

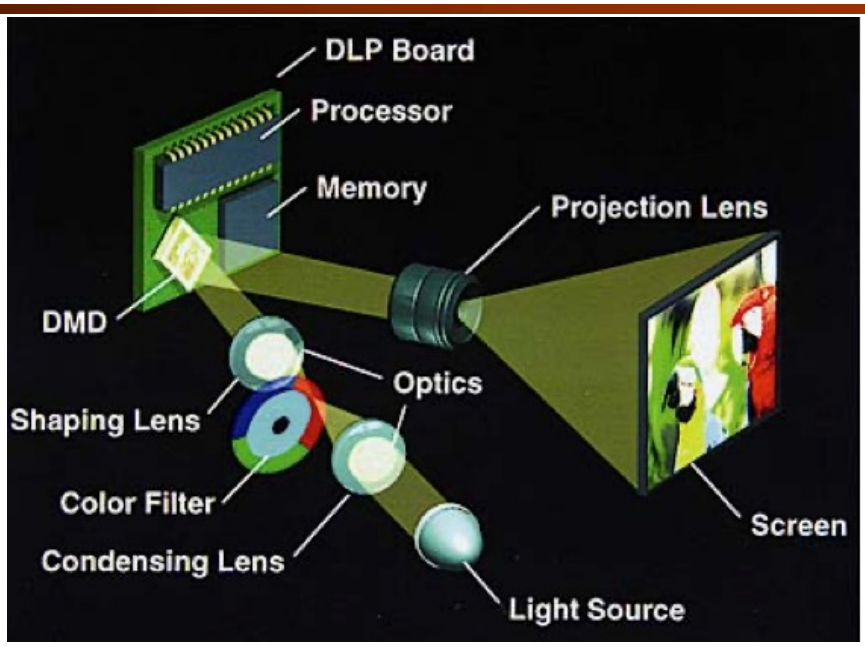
SEM images of DMD by TI



Source: Texas Instruments

This is the scanning electron microscope (SEM) images of the DMD array with the aluminum mirror and the MEMS mechanism underneath. The size of each mirror is around 16 microns (10^{-6} m) across.

Digital Micromirror Device (DMD) Technology



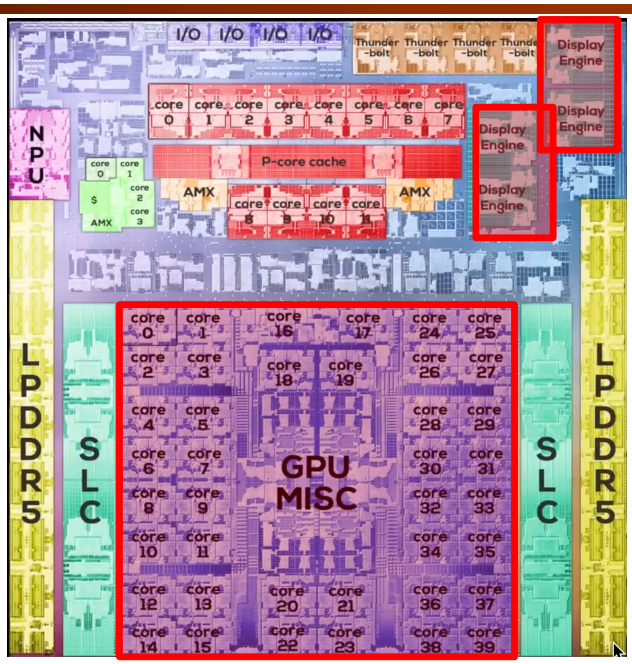
DLP – Digital Light Processing

Source: Texas Instruments

This slide shows a single chip DLP projection system using the DMD technology by TI. (DLP stands for Digital Light Processing.)

The light source is usually metal halide because of its greater luminous efficiency (lumens delivered per electrical watt dissipated), but recently OLED is also used. A condenser lens collects the light, which is imaged onto the surface of a transmissive colour wheel. A second lens collects the light that passes through the colour wheel and evenly illuminates the surface of the DMD. Depending on the rotational state of the mirror (+10 or -10 degrees), the light is directed either into the pupil of the projection lens (on) or away from the pupil of the projection lens (off). The projection lens has two functions: (1) to collect the light from each on-state mirror, and (2) to project an enlarged image of the mirror surface to a projection screen.

Apple Silicon M3 Pro Chip Photo



Source: Apple Computers

Shown here is the chip photograph of Apple's latest M3 Pro chip. The area shown in shaded RED are circuits devoted to visual processing. It occupies more than 50% of the silicon area.

There are two main types of visual processing units: 40 GPUs (Graphics Processing Units) and four Media Display Engines.

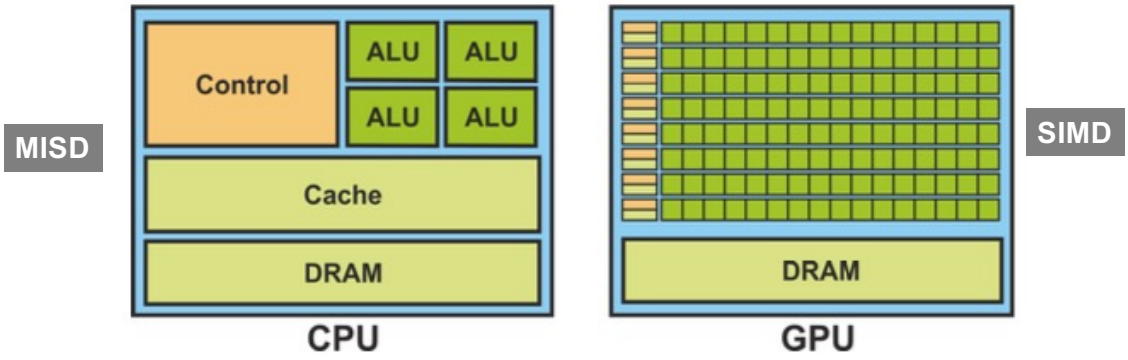
The GPUs are for accelerating general computer graphics algorithms, rendering and image processing. We will consider the architecture of GPU in the next few slides.

The Media Engines are for encoding and decoding videos based on different standards. Details are not relevant to this module, and will not be considered further.

	Video decode engines (H264 & HEVC)	Video encode engines (H264 & HEVC)	ProRes encode and decode engines	AV1 decode
Apple M3 chip	1	1	1	1
Apple M3 Pro chip	1	1	1	1
Apple M3 Max chip	1	2	2	1

What is a GPU?

- Based on SIMD architecture – Single Instruction, Multiple Data.
- Exploit data parallelism.
- High compute density due to massive degree of parallel processing.
- Initially developed as graphics accelerators for games.
- Now used for many other scientific processing and particularly for AI training.
- Software based on C/C++, and also on CUDA (Nvidia's own programming language).



What is a GPU? How does a GPU differs from a CPU?

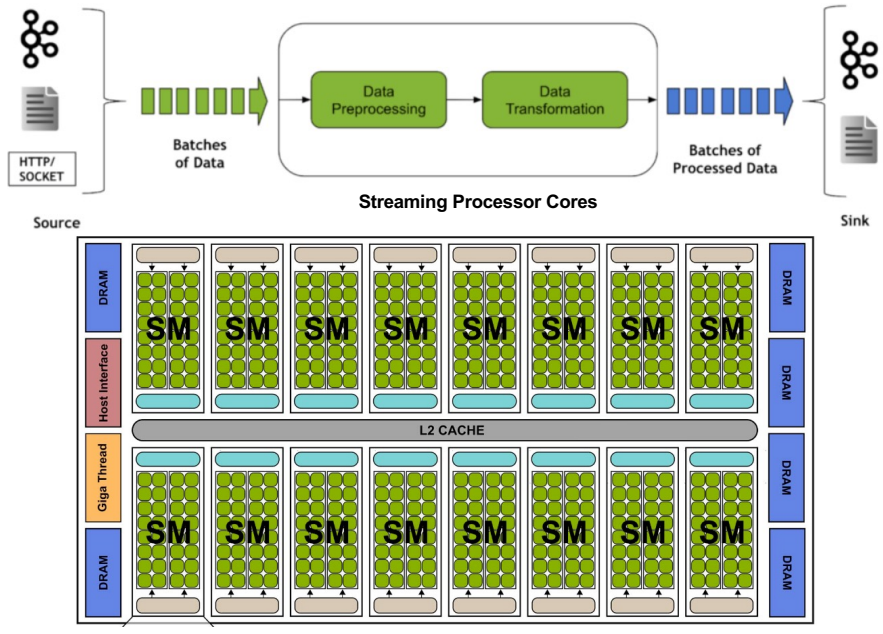
CPU is has a Multiple Instruction, Single Data (MISD) architecture. A modern CPU would have multiple processor cores (4 shown here), so that can execute multiple instruction streams. However, they tend to share a common memory through cache and DRAM.

A GPU generally is a Single-Instruction, Multiple-Data architecture where there are many parallel cores (hundreds or more) each running the same instruction (at least many cores are running the same code). However, the program operates on different part of an image or screen.

Below is a comparison between the two types of processors.

CPU	GPU
Central Processing Unit	Graphics Processing Unit
4-8 Cores	100s or 1000s of Cores
Low Latency	High Throughput
Good for Serial Processing	Good for Parallel Processing
Quickly Process Tasks That Require Interactivity	Breaks Jobs Into Separate Tasks To Process Simultaneously
Traditional Programming Are Written For CPU Sequential Execution	Requires Additional Software To Convert CPU Functions to GPU Functions for Parallel Execution

Typical Nvidia GPU Architecture



Source: NVIDIA

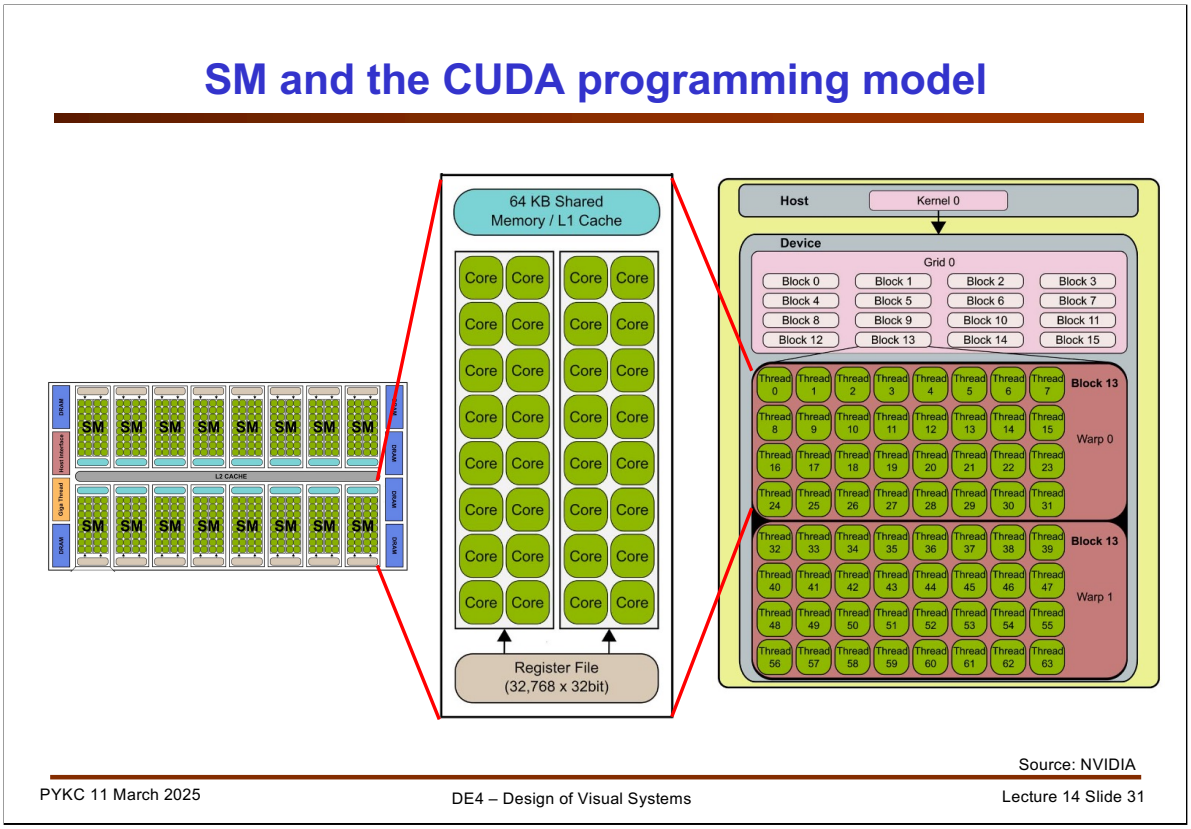
Nvidia is arguably the most popular (and successful) manufacturer of GPUs.

Each GPU consists of many streaming processors (SMs) as shown in the slide.

A stream is an unbounded sequence of events that go from producers to consumers. A lot of data is produced as a stream of events, for example financial transactions, sensor measurements, or web server logs.

Stream processing pipelines often involve multiple actions such as filters, aggregations, counting, analytics, transformations, enrichment, branching, joining, flow control, feedback into earlier stages, back pressure, and storage.

Data are pumped in at one end to the SM, and results are collected at the output. Generally speaking, one batch of output can be produced on each clock cycle, resulting very high throughput. However, the pipeline can be very long and therefore the latency (number of clock cycles needed for one operation) can be long.

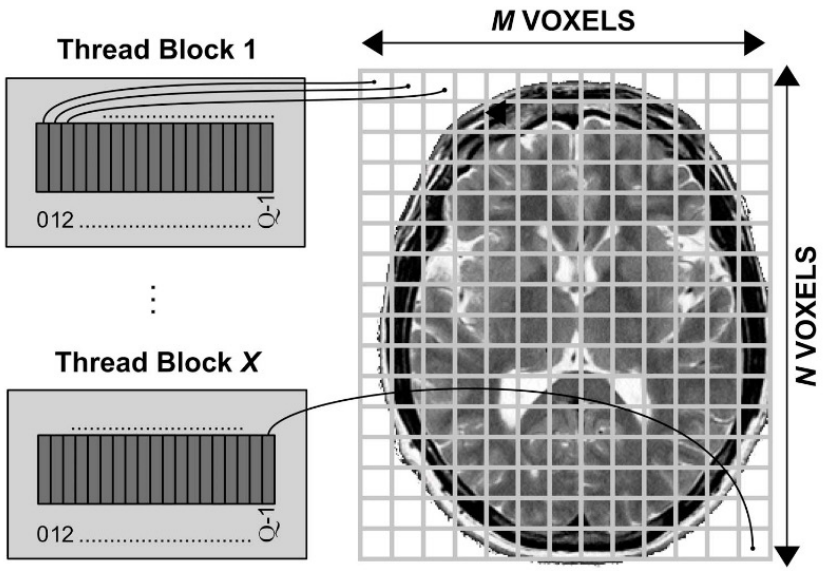


Each Nvidia GPU contains multiple SMs, which are individual processing units responsible for executing tasks in parallel. SMs consist of many processor cores (32 in this example), allowing the SM to perform mathematical operations, called **threads**, simultaneously. Each SM also share a very large **register file** (32k x 32 in this example) and a common level 1 **cache memory**.

The **CUDA programming language** has a software model as shown above. The top level is known as a “Kernel”. Each kernel consists of many ”Blocks”. A block is processed by a warp, which is mapped to a SM in this example. Each SM can **execute multiple threads concurrently**.

Detail of this is beyond the scope of this module but you should be aware that programming GPU is NOT like programming CPU due to its large degree of parallelism.

Nvidia CUDA Programming Model



Source: Moises Hernandez

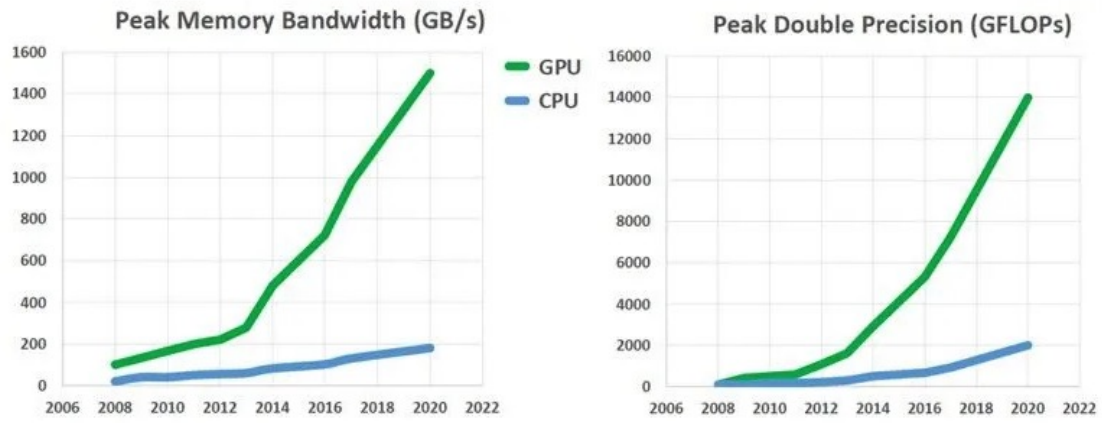
This slide shows how a GPU can process an image in parallel.

The brain MRI image is divided into M x N voxels. (A voxel is a measurement of volume in a structure that is to be imaged).

To process the image, each voxel is handled by the GPU as one thread, which is mapped to a warp which is excuted by one or more SMs.

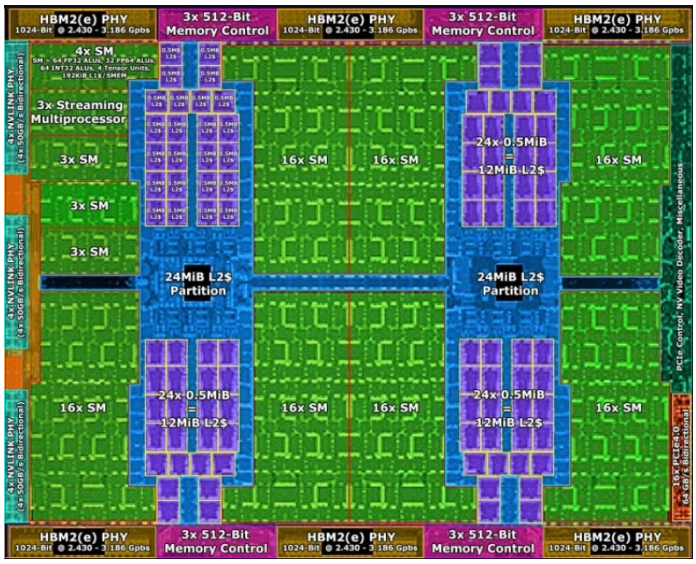
In the example above, the CUDA kernel performs some algorithm. Voxels are assigned to threads of CUDA blocks. Each CUDA block is comprised of Q threads and processes Q voxels each.

GPU vs CPU on Performance



The graphs above shows the acceleration of performance of GPUs since 2008. Not only are GPUs executing many more instructions per second (GFLOP = giga floating point operation, a giga is 10^9), it also boasts a much higher memory throughput, able to move up to 1.5TB per second of data to and from memory.

Nvidia Latest – Ada GPU AD102

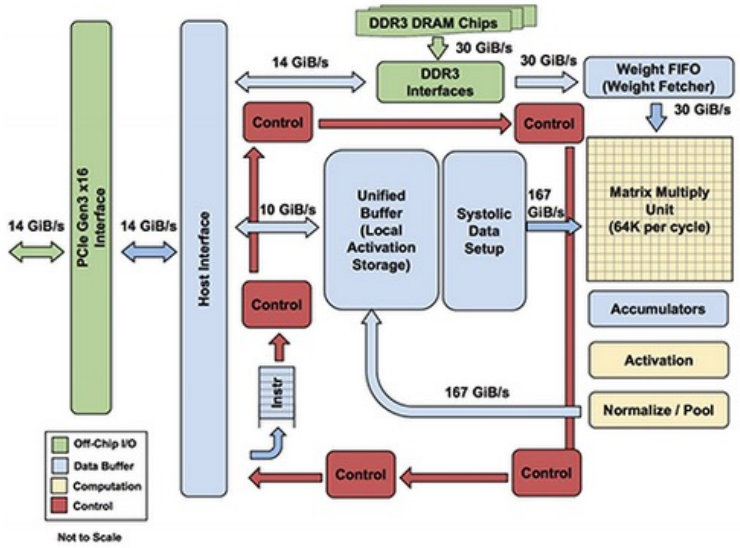


- 288 FP64 cores in
- 144 Streaming Multiprocessors (SMs)
- 12 Graphics Processing Clusters (GPCs)
- 72 Texture Processing Clusters (TPCs)
- 384-bit memory interface
- 12 32-bit memory controllers.

Source: NVIDIA

Here is one of the most up-to-date NNVIDIA GPU based on the Ada GPU family. It has an estimated peak performance of 100 teraFLOPS and 300GB/s memory bandwidth.

Google Tensor Processing Unit (TPU)



The TPU includes :

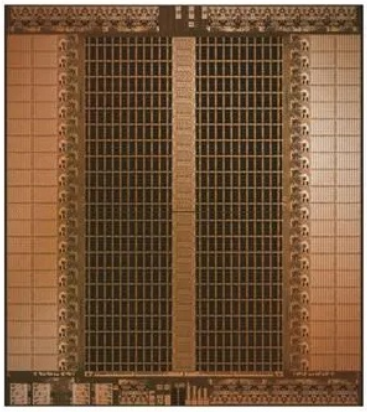
- **Matrix Multiplier Unit** (MXU): 65,536 8-bit multiply-and-add units for matrix operations
- **Unified Buffer** (UB): 24MB of SRAM that work as registers
- **Activation Unit** (AU): Hardwired activation functions

Source: Google

The final type of processing unit to consider is not specifically for visual data processing, but for general AI applications. Google started to design its first AI chip in 2015, and called this Tensor Processing Unit or TPU. Unlike GPU and CPU, TPUs are designed high density but low precision computation for implementing neural networks, particularly CNN (Convolutional NN) which has lots of convolution (i.e. multiply-add) operations. TPU is designed to support Google's own AI framework known as TensorFlow. The latest version (V5) has a peak performance of 393 Tera Operations per second. (Since TPU uses integer arithmetic, these are integer operations, not floating point operations as found in GPUs.)

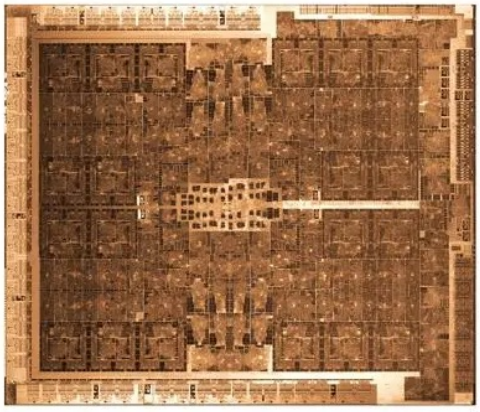
GroqChip 1 – Another AI inference engine

GroqChip™ 1



Simplified design enables **compute performance**

Typical GPU



Complex design for processing data results in **compute costs**

Source: Groq

Google is not alone in designing AI acceleration chip. Groq designed a chip that delivers predictable and repeatable performance with low latency and high throughput across the system called the tensor streaming processor (TSP).

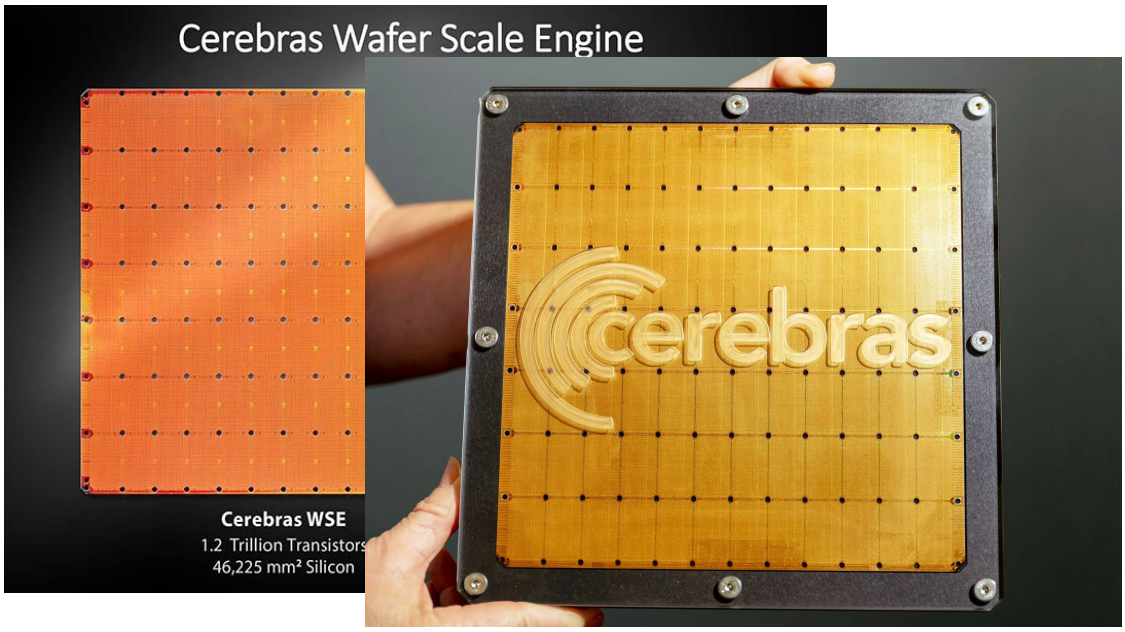
The new, simpler processing architecture is designed specifically for the performance requirements of machine learning applications and other compute-intensive workloads. They created their own AI interference system targeting large language model (LLM) as found in OpenAI's ChatGPT, Google's Gemini or Meta's Llama2. Groq LLM server implements the pre-trained Llama2 LLM model. (Try this out on groq.com.)

Details of their design is proprietary. However, the above chip photograph suggests that they are highly regular hardware layout encouraging extreme parallelism, even more so than GPUs.

Specifications

Item	Description
Availability	In production. Contact Groq support at info@groq.com
Process Node	14nm
Performance	Up to 750 TOPs, 188 TFLOPs (INT8, FP16 @900 MHz)
Memory	230 MB SRAM per chip Up to 80 TB/s on-die memory bandwidth
Chip Scaling	16 integrated RealScale™ chip-to-chip interconnects
I/O	Integrated PCIe Gen4 x16 controller
Numerics	INT8, INT16, INT32 & TruePoint™ technology MXM: FP32 VXM: FP16, FP32
Power	Max: 300W; TDP: 215W; Average: 185W

Cerebras 1.2 trillion transistor AI interference engine



This is currently the largest AI interference engine available in the world. Announced in 2024, the Cerebras wafer scale inference engine contains 1.2 trillion transistor. This promises to outperform all other inference chips in the market by a significant margin while reducing power consumption. Nvideo, the current chip leader in the AI space only have chip level products which are 50 times smaller and lower in performance than Cerebras offering.

Specifications

Item	Description
Availability	In production. Contact Groq support at info@groq.com
Process Node	14nm
Performance	Up to 750 TOPs, 188 TFLOPs (INT8, FP16 @900 MHz)
Memory	230 MB SRAM per chip Up to 80 TB/s on-die memory bandwidth
Chip Scaling	16 integrated RealScale™ chip-to-chip interconnects
I/O	Integrated PCIe Gen4 x16 controller
Numerics	INT8, INT16, INT32 & TruePoint™ technology MXM: FP32 VXM: FP16, FP32
Power	Max: 300W; TDP: 215W; Average: 185W