9 Graphic Display Technology

9.1 Introduction

The human visual system responds to pictures more readily than words. Also, certain pictures may have universal meaning, so that communication across different languages is simplified. However the user interface of early computers was text based, due to cost and performance constraints. Graphical user interfaces for computers were developed in the 1970's but were expensive and required a lot of computing power. Today's computers provide high computing power at low cost. The demand for graphical interfaces to various equipment have driven down the cost of graphical interface devices. It is now quite feasible to have a low cost embedded system with graphical interface as can be seen in the various Portable Entertainment Devices available today. Earlier technologies worked with analogue video signals but currently, all-digital data video is coming commonplace. This chapter examines the hardware considerations in advanced embedded systems like the SoC's commonly available.

9.2 Considerations in graphical displays

We live in an age of mobile information technology which relies heavily on the visualization of multimedia data and user input. These tasks are often performed by high-resolution electronic displays equipped with a touch screen or a user input device like a keypad. Choosing a big or a small display, connecting it via the HDMI connector or via the GPIO pins, needing touch capabilities or not, these are only some of the decisions one will have to take depending on the end goal. Ultimately it will be a trade-off with what will the project requires and what's left on the platform to connect and drive a display.

9.3 The nature of a computer image

Because of the digital nature of computers, an image was constructed in the form of a grid. Each cell of a grid would be part of a picture, so it was a picture element or 'pixel'. Each pixel has a colour value, which also includes black for the absence of any colour information. Generally, pixels are square in shape. As in television, the number of pixels making up an image is known as the resolution. This is given as the number of horizontal pixels times the number of vertical pixels. The amount of data required to represent the colour is known as the colour depth. The entire digital image has to be stored in a frame buffer which may be part of system memory. The frame buffer data needs to be transferred to DACs which convert the digital data into analogue colour. The amount of data that needed to be computed and moved to provide the display can tax the resources of the hardware.

Most of today's colour displays are based on three primary colours – red (R), green (G) and blue (B). They can be superimposed in various ways so the human eye sees just the actual colour.

Computer images are displayed at a fixed resolution which specifies the number of pixels to be displayed horizontally and vertically. The colour at each pixel is represented by digital data. This quantity is known as the colour depth. Commonly used quantities are 1 bit, 4 bits, 8 bit gray and colour, 24 bit colour, and 32 bit colour. The easiest to store is 24 bit colour, where one pixel has three bytes stored. One each for its component colours. Of course this takes the most space.

The 8 bit colour scheme is a special type of image data storage. Colours can be represented as a combination of three 8 bit quantities. However, not all possible values of all colours (16 million - 2^{24}) are used in an image. It is possible to create a good approximation to the colours in a picture by just using 256 of the most common colours in *that* image. This process of quantization uses only 1/3 of the data needed, which is a great saving in file space. All this can be done by hardware at high speeds, so that the job of finding the most common 256 colours to represent a picture is done quickly. But it must be noted that often, with images like photographs, the effect of reducing the colours results in *posterization* where large patches of a single colour will replace a large area of smoothly changing colours. Thus using 256 colours should be used for artificially generated images like cartoons or animations. Each of the 256 colours are stored in the image as 3 byte entries. This structure is known as a palette. So each pixel of an 8-bit image points to a 3-byte entry in the palette. Because each image is different, a palette has to be stored with *every* image. This is not so with 8 bit gray scale images as each 8 bit image actually represents just the brightness information only.

Gray colours are often created by a linear combination of the RGB colours. By suitable thresholding of gray values, 1 bit, or black and white images may be obtained.

The following diagram gives an overview of how a pixel may be represented digitally.

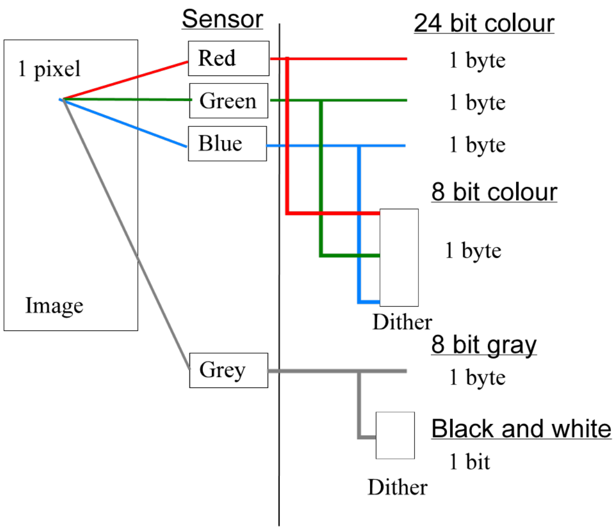

Image formation and colour depth options

To display an 8 bit colour image, the values of the palette are loaded into a hardware *look-up table*. This is used to drive the colour DACs which provide the signal for the appropriate colour.

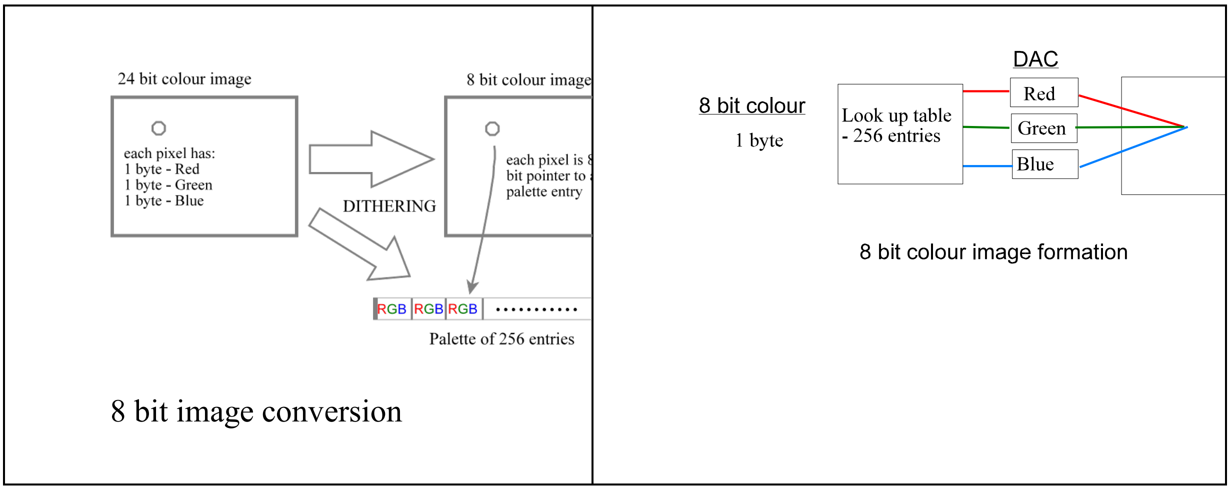

Fig 9.1 8-bit colour image formation

9.3.1 Display technology

There are many different technologies used to build a display. These technologies are typically known by the acronym that refers to the primary electro-optical effect that is used to emit and/or modulate visible light at a pixelated level, e.g CRT, LCD and OLED. This may be further divided into technologies which generate their own light – emissive, or those that modulate light generated separately (e.g via a backlight) – reflective or transmissive. Displays are also referred to as being direct view, virtual or projected to describe the manner in which the image is viewed by the user. These categories are used in Fig. 9.2 to provide an overview of the major technologies:

- Direct view displays - the user looks direct onto a display
- Projection displays - the user sees the image of a microdisplay projected onto a reflective screen
- Microdisplays - the user views an image which is typically displayed near the eye
- 3D displays - displays which can produce three-dimensional images

The different characteristics of emissive, transmissive, and reflective displays are presented in Fig. 9.3. As simplification, the principles are only visualized for black and white, with a summary of the corresponding display features in Table 9.1

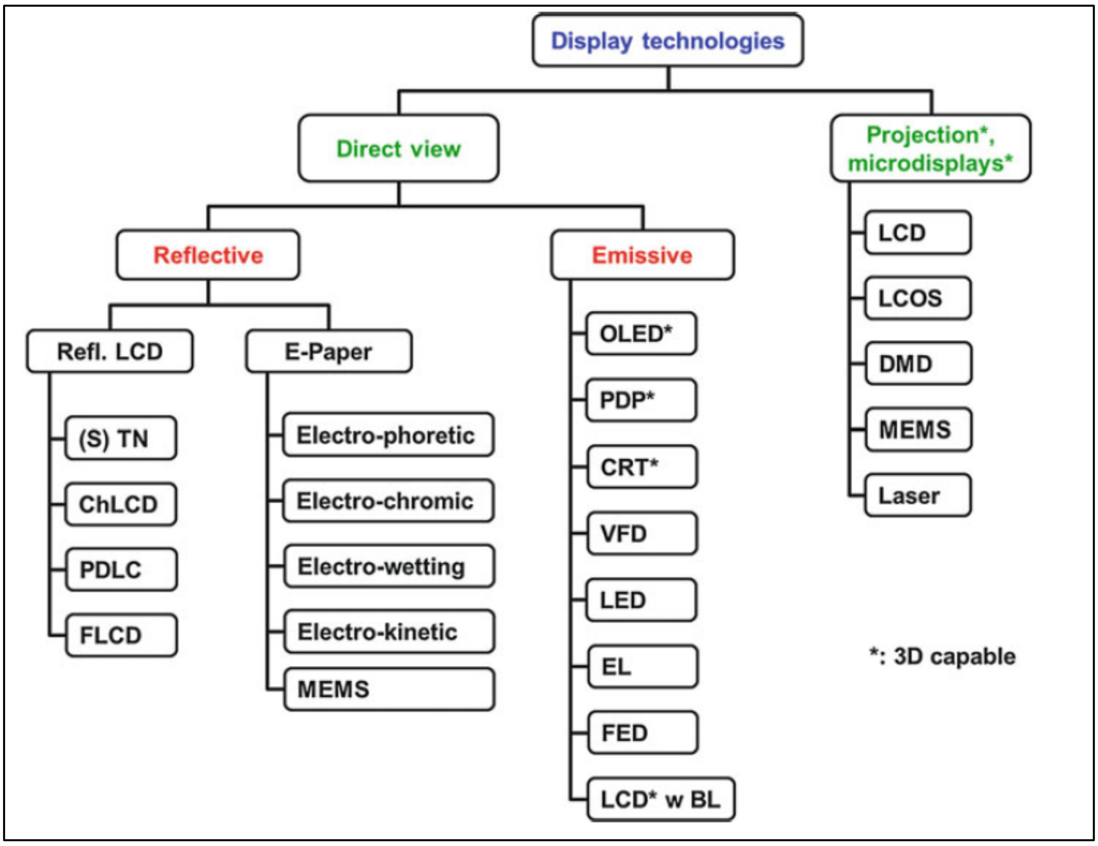

Fig 9.2 Overview of display technologies (see Appendix for definitions of the abbrevations)

- Emissive displays generate light by converting electrical power to light.
- Reflective displays modulate their reflectance, typically via a voltage-controlled effect.
- LCDs are labelled as both reflective (e.g., 7-segment display as used in a watch) and transmissive.

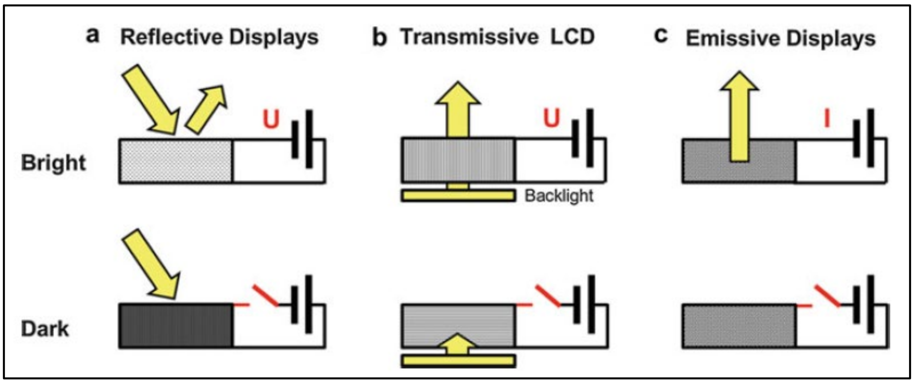

Fig 9.3 Visualization of fundamental electric-to-optic conversion for major flat panel display technologies

Topic	Reflective	Emissive/transmissive
Examples	Segmented LCD Passive matrix LCD E-paper	OLED Plasma Display Panel (PDP) Active matrix colour LCD (AM LCD) with backlight
Typical merits	Low power Sunlight readable	Multimedia Supply chan of AM LCDs
Typical shortcomings	No or limited colour Some with slow response time	Bright light performance High power consumption

Table 9.1 Techology dependant typical display features

- The reflection characteristic of reflective (non-emissive) displays is controlled by a voltage (U, low current, Fig. 9.3a) switching between black (absorption, low reflectance) and white (high reflectance). These displays are therefore readable in bright light but must be illuminated when dark. As ambient light is utilized, the power consumption is extremely low ("green displays"). Examples are segmented monochrome LCDs and e-paper displays. The latter ones are mostly bistable, which means that electrical power is only needed when a pixel should change its status. This is very significant in terms of power consumption, which facilitates very long battery life for e-readers based on this technology.
- Transmissive display (Fig. 9.3b) is the term used for color LCDs (mostly active matrix) which are equipped with a bright backlight. A voltage changes the transmission between a low (black) and a high value (white) of the liquid crystal layer (cell) to modulate the light from the backlight subsystem. So the power consumption of the basic liquid crystal cell is low but that of the display is high due to the necessary backlight. Because of the reliance upon an integrated light sources, transmissive LCDs can also be referred to as emissive displays, and even as "LED" displays, which actually only refers to the backlight.
- Emissive displays (Fig. 9.3c) generate light by converting electrical power into visible light. Note that "I" is used as the current symbol here, since these are often current driven, and, relative to reflective technologies, higher powers are typically required. For emissive display, the power consumption also depends on the actual data shown; hence, a black background is preferable in this context. Examples are OLED, PDP, CRT, VFD, FED, EL, and LED.
- Transmissive and emissive displays typically provide a high-quality optical performance under indoor ambient conditions, but their outdoor use is limited due to the need for a very high luminance to provide sufficient contrast. This requires higher power consumption, hence reducing battery life and impacting on longevity of the display system due to heat generation.

It is evident that no single display technology exists which is suitable for all applications. A key decision point is the choice between low-power monochrome displays and multimedia screens consuming a relatively large amount of power.

9.3.2 Display Driving Principles

After introducing pixels earlier, the next step toward understanding how a display works is to consider how the pixels are electrically driven, e.g., setting a gray level or switching between gray levels (the simplest case is black ↔ white). The two fundamental approaches are shown in Fig. 9.4 – direct drive (a) and matrix drive (b). For direct drive all segments (typically used instead of "pixel" for this case) are directly connected to the driving electronics. So there is full control over the voltage and the waveform of any segment. The arrangement and the shape of the segments can be individually designed for the application to be, for example, bars or icons, with the geometry defined by electrode structuring.

. Matrix drive can be classified into passive matrix drive and active matrix drive.

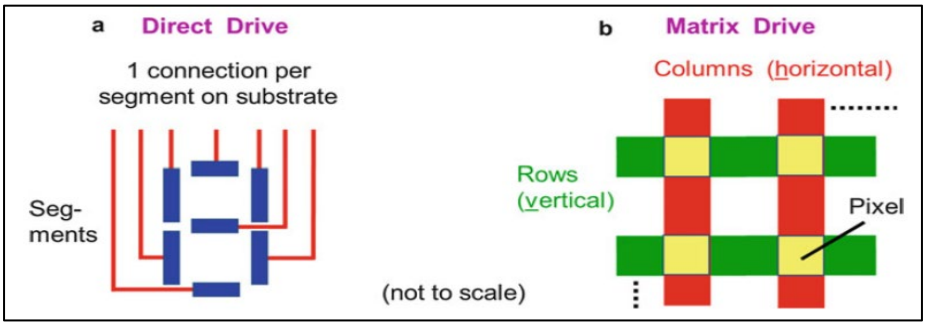

Fig. 9.4 Fundamental driving principles of pixels: Direct (a) and matrix (b) drive

9.3.2.1 Passive matrix

In a *passive matrix* display (Fig. 9.5a), pixels are addressed row by row, this is called time multiplexing. That means that all pixels on row 1 are updated first, then all pixels on row 2, etc meaning that for a display with three rows, each row is only addressed ⅓ of the total time. On retro displays (e.g CRT), it is sometimes possible to see this effect as a continuous sweeping across the screen. For LCDs, this reduces the contrast of the display which, in turn, limits the total number of rows possible. This method is often called *multiplex driving* for segmented displays. Passive matrix drive is a cost-effective method to drive displays as it doesn't require any additional hardware. However, just a few display technologies have the characteristics required for passive matrix drive.

9.3.2.2 Active matrix

The sophisticated display, that you most likely are looking at right now, is based on active matrix (Fig 9.5b) technology. In an active matrix, each pixel contains at least one transistor. More often, there are multiple transistors and capacitors. In an OLED for example, each pixel contains a quite sophisticated circuit with 5-10 transistors. By adding transistors to the pixel, it can more easily be controlled. This is partly because transistors offer a threshold voltage which is an important feature for a display matrix to function properly. A capacitor, on the other hand, functions as an energy storage when the pixel is not addressed. In this way, all pixels can maintain their state even for a large number of rows. The Apple iMac display, for example, can in this way achieve 2880 rows without a problem. The drawback with active matrix is the high price point since the fabrication requires expensive deposition processes. For that reason, active matrix is mainly suitable for high-end displays.

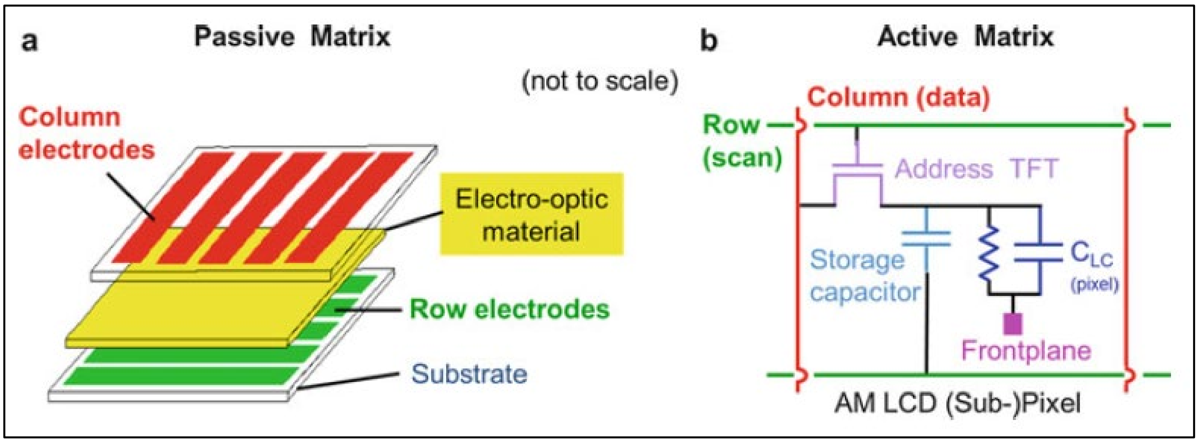

Fig. 9.5 Schematic drawing of passive (a) and active (b) matrix driving

9.4 Types of interfaces of display devices

Current embedded systems can be connected to a display in several ways, depending on the model. We look at some of the interface that is typically used in a project.

Composite Video

Despite the great success it has received the Raspberry Pi has humble origins and at its heart it was designed to be inexpensive. It was thought that providing Composite Video would have allowed in many countries that still rely on old TV set to use the Raspberry Pi without the need to purchase a dedicated monitor. A composite video connection transmits a basic analog video signal between devices. It sends standard-definition video only - and does not carry an audio signal. If you use a composite video signal for the picture, then you will need a separate connection to hear the audio.

DSI

DSI (Display serial interface) is a high-speed serial interface based on a number of (1GBits) data lanes. The total voltage swing of the data lines is only 200mV which is extremely small, and therefore the term low voltage differential signalling (LVDS) reflects this. The electromagnetic noise created and power consumed is very low.

HDMI

HDMI (High-Definition Multimedia Interface) is used for transmitting uncompressed video or digital audio data to the Computer Monitor, Digital TV, etc. Generally, this HDMI port helps to connect your current embedded design to the Digital television. In the vast majority of cases, simply plugging your HDMI-equipped monitor into for example, a Raspberry Pi using a standard HDMI cable will automatically lead to the Pi using the best resolution the monitor supports. You should connect any HDMI cable before turning on the Raspberry Pi.

GPIO

GPIO is the acronym for General Purpose Input Output and very often the GPIO pins are multiplexed with other functions so they need to be set up before use. So for example, you can reconfigure each pin as either an input or output but also deem them to provide a different type of interface for the various types of displays and boards in general. You can find displays that use DPI, SPI, I2C and even UART or an ad-hoc interface. Some Interfaces like DPI are particularly fast but will require lots of GPIO to interface with the screen, others like SPI and I2C need very few pins but won't be as fast as DPI. Ultimately choosing between them comes down to your projects requirements.

- DPI

  DPI stands for Display Parallel Interface (or Parallel Display interface). It allows you to use very cheap displays by driving them manually. However, it may not be right for every project.

  Pros:
  Very fast, easily driving display at 60hz.
  No complicated interface hardware.
  Pixel perfect output. Digital, not analog.
  Easy to understand protocol.
  No bulky connectors.
  Very inexpensive.
  Cons:
  Uses a lot of GPIO pins
  Ribbon cables break easily if you're not careful.
  Short range.

  The DPI interface allows displays to be attached to the Raspberry Pi GPIO either in RGB24 (8 bits for red, green and blue) or RGB666 (6 bits per colour) or RGB565 (5 bits red, 6 green, and 5 blue). If we wanted to drive the display at full 24 bit true color (RGB24) we would need 8 pins each for red, green, and blue as well as pins for signals like hsync, vsync, clock, and display enable. This is a total of 28 pins. The way to reduce pin count is to simply omit the two least-significant-bits on each color bus, giving us 18 bit high color (RGB666). Another option is to use RGB565. Obviously we lose some colour information but it's not as bad as one would expect. This is one way the SMI interface described in an earlier chapter can be used.

- SPI/I2C (or I^2C)

  The Serial Peripheral Interface (SPI) and Inter-Integrated-Circuit bus (I2C) are serial interface on the GPIO. With SPI the amount of pins required is much less than DPI, it only requires 4-5 pins. More than one device can be connected to SPI but it's generally not used for more than one device.

  I2C requires few pins, in fact it only needs 2 which makes this interface very light in terms of GPIO usage. One additional advantage is that it's very easy to connect more than one device to I2C bus making this option quite versatile.

One of the main downsides of these technologies is that you will end up with fairly small displays and very low resolutions.

- UART
  UART stands for Universal Asynchronous Receiver/Transmitter. A UART's main purpose is to transmit and receive serial data. Almost all microcontrollers have dedicated UART hardware built into their architecture. The main reason for integrating the UART hardware into microcontrollers is that it is a serial communication and requires only two wires (and ground) for communication.

Appendix:

ChLCD	Cholesteric Liquid Crystal Display
CRT	Cathode Ray Tube
DMD	Dot Matrix Display
EL	Electroluminescent
FED	Field Emission Display
FLCD	Ferroelectric Liquid Crystal Display
LCD w BL	LCD with Backlight
LCoS	Liquid Crystal on Silicon
MEMS	Microelectromechanical Systems
OLED	Organic LED
PDLC	Polymer-Dispersed Liquid Crystals
PDP	Plasma Display Panel
TN	Twisted Nematics
VFD	Vacuum Fluorescent Display