

1. Information Representation

1.1 Data Representation

Applications of BCD and Hex

Binary Coded Decimal

1. Digital displays (calculator, clock display)
 - easier for a human to read large binary numbers than if the number is in pure binary
 - simpler to create hardware that only counts in decimal and uses basic calculations
2. Representing monetary values in computers
 - pure binary cannot present accurate fractions, due to the fractions not fitting perfectly within a binary base 2 system → floating point rounding errors
 - Binary-coded decimal removes the problem of floating point rounding errors

Hexadecimal

1. Memory dump: memory contexts that are output to a printer or monitor
2. HTML color code, error code → Easier to read, easier for computers to calculate

Character Sets

ASCII (American Standard Code for Information Interchange)

- **standard ASCII:** 7-bit codes, represent only English characters
- **extended ASCII:** 8-bit codes, allow for non-English characters and for drawing characters to be included

Unicode

- **UTF-8:** 1-4 bytes, **UTF-16:** 2, 4 bytes, **UTF-32:** 4 bytes
 - mainly UTF-8, which covers all the ASCII characters
- **Aims of Unicode:**
 - create a **universal standard** that covered all languages and writing systems
 - create **unambiguous encoding** where each 16-bit or 32-bit code represents the same character
 - ASCII code tables are not standardized and have different versions
 - reserve part of the code allowing users to assign codes for their own characters and symbols (e.g., for Chinese and Japanese)

1.2 Multimedia

Bitmap Images

Keywords

- **Pixel:** smallest picture element that makes up an image
- **Image Resolution:** number of pixels that make up an image
- **Color Depth:** number of bits used to represent the colors in a pixel

- e.g., 8-bit color depth can represent 2^8 colors
- **Screen Resolution:** number of horizontal and vertical pixels that make up a screen display
 - If the screen resolution is smaller than the image resolution, the image will need to be either cropped or lowered the resolution (resulting in lower image quality)
 - If the screen resolution is larger than the image resolution, the enlarged image will be **pixelated** (the number of pixels per square inch, **pixel density**, is smaller)
- **File Header:** a section that stores identifying information about the file and metadata with information about the file contents

Data Encoding

- Data for bitmapped images are encoded using a **grid of individual pixels**, where **each pixel represents a specific color** or shade
- These **pixels are organized in rows and columns** following the image resolution, forming a matrix that represents the image
- The color or shade of each pixel is stored as **binary data**, indicating the intensity of the red, green, and blue (RGP) components for colored images
- The binary data for each pixel is stored in the image file, usually in a **specific file format** such as JPEG, PNG, etc

File Size Calculation

$$\text{File Size} = \text{Image Width} \times \text{Image Height} \times \text{Color Depth}$$

Factors of Image Quality

1. **Image Resolution:** affecting image sharpness and file size
 - High image resolution → Better quality and larger file
2. **Colour Depth:** affecting the range of colors that can be represented in the file
 - High color depth → Better quality and larger file

Vector Graphics

Data Encoding

- **Drawing Objects:** basic elements (lines, shapes, and texts) in vector graphics
- **Property:** characteristics of drawing objects (color, size, and shape)
- **Drawing List:** an ordered list of drawing objects, defining the appearance of the vector graphics

Comparison with raster images

Vector graphic images	Bitmap images
made up of geometric shapes that require definition/attribute	made up of tiny pixels of different colors
to alter/edit the design, it is necessary to change each of the geometric shapes	possible to alter/edit each of the pixels to change the design of the image
do not require a large file size since it is made up of simple geometric shapes	generally, the file size is large because of the use of pixels
not usually very realistic as the number of geometric shapes is limited	realistic since images are built up pixel by pixel
file formats are usually .svg, .cgm, .odg	file formats are usually .jpeg, .bmp, .png

Vector graphic images	Bitmap images
ideal for tasks requiring scalability and crisp edges (logos and icons)	suitable for photographs and complex images with many details and color variations

Sound

Keywords

- **Analogue Data:** data represented using a continuous range of values.
- **Digital Data:** data represented using discrete values, usually in binary
- **Sampling:** the process of converting analogue data into digital data at regular intervals
- **Sampling Rate:** number of sound samples taken per second
- **Sampling Resolution:** number of bits used to represent sound amplitude (bit depth)

Data Encoding

- The analogue data of sound is digitalized using an Analogue-to-Digital Converter (ADC)
- The amplitude is recorded a set number of times a second
- Each amplitude is given a corresponding binary number
- The binary number is saved in sequence

Factors of Sound Quality

1. **Sampling rate:** increasing sample rate increases sound quality and file size
 - Sound is recorded more often
 - The digital **waveform** is closer to the analogue waveform
 - The quantization errors are smaller
2. **Sampling resolution:** increasing sample resolution increases sound quality and file size
 - A wider range of amplitudes can be stored
 - Each binary amplitude is closer to the analogue amplitude

1.3 Compression

Need for compression

- **less bandwidth** required
- **less storage space** required
- **less transmission time** required

Types for compression

	Lossy Compression	Lossless Compression
Definition	file compression method where parts of the original file cannot be recovered during decompression.	file compression method where the original file can be restored following decompression.
Quality Preservation	Quality loss due to data loss	No quality loss
File Size Reduction	Significant reduction in file size	Moderate reduction in file size
Reversibility	Irreversible , the data cannot be recovered once it is lost	Reversible , original data can be completely restored

	Lossy Compression	Lossless Compression
Processing Speed	Faster processing due to data loss and simplified algorithms	Slower processing due to the complexity of preserving all data
Use Cases	Multimedia applications where minor data quality loss is acceptable	Archiving, software distribution, and texts, where preserving original data is essential
Example Formats	JPEG, MP3, MP4	PNG, FLAC, ZIP

Compressions of different file types

Text

- **Lossless** compression
 - **Huffman Coding**: assigns shorter codes to more frequent symbols and longer codes to less frequent symbols
 - **Run-Length Encoding (RLE)**: Replaces consecutive identical elements with a single instance and a count

Sound

- Compressed file format: **MP3** (MPEG Audio Layer III)
 - **Lossy** file compression
 - Algorithm: **Perceptual music shaping**
 - sounds with frequencies outside the human hearing range will be discarded
 - if two sounds are played at the same time, only the **louder** sound can be heard by the human ear, so the softer sound is discarded

Bit-map Image

- **Lossy** compression **JPEG** (Joint Photographic Experts Group)
 1. JPEG downsamples the color channel (so that the photo contains less colors)
 2. JPEG breaks down the images into tiny 8×8 squares
 3. For each square, JPEG describes the square by its general format/texture
 4. JPEG discards those tiny details in the photo
- **Lossless** compression: similar to that of texts

Vector Graphic

- **Lossless file compression**
 - **Eliminate redundant information**: If there are overlapping or hidden objects, or if there are objects outside the viewable area, the objects are discarded
 - **Path simplification**: Complex paths with multiple points can sometimes be simplified to fewer points
 - **Use references and patterns**: if a particular shape, line style, or color gradient is used multiple times, it can be defined once by references and used throughout the algorithm (Huffman Coding)