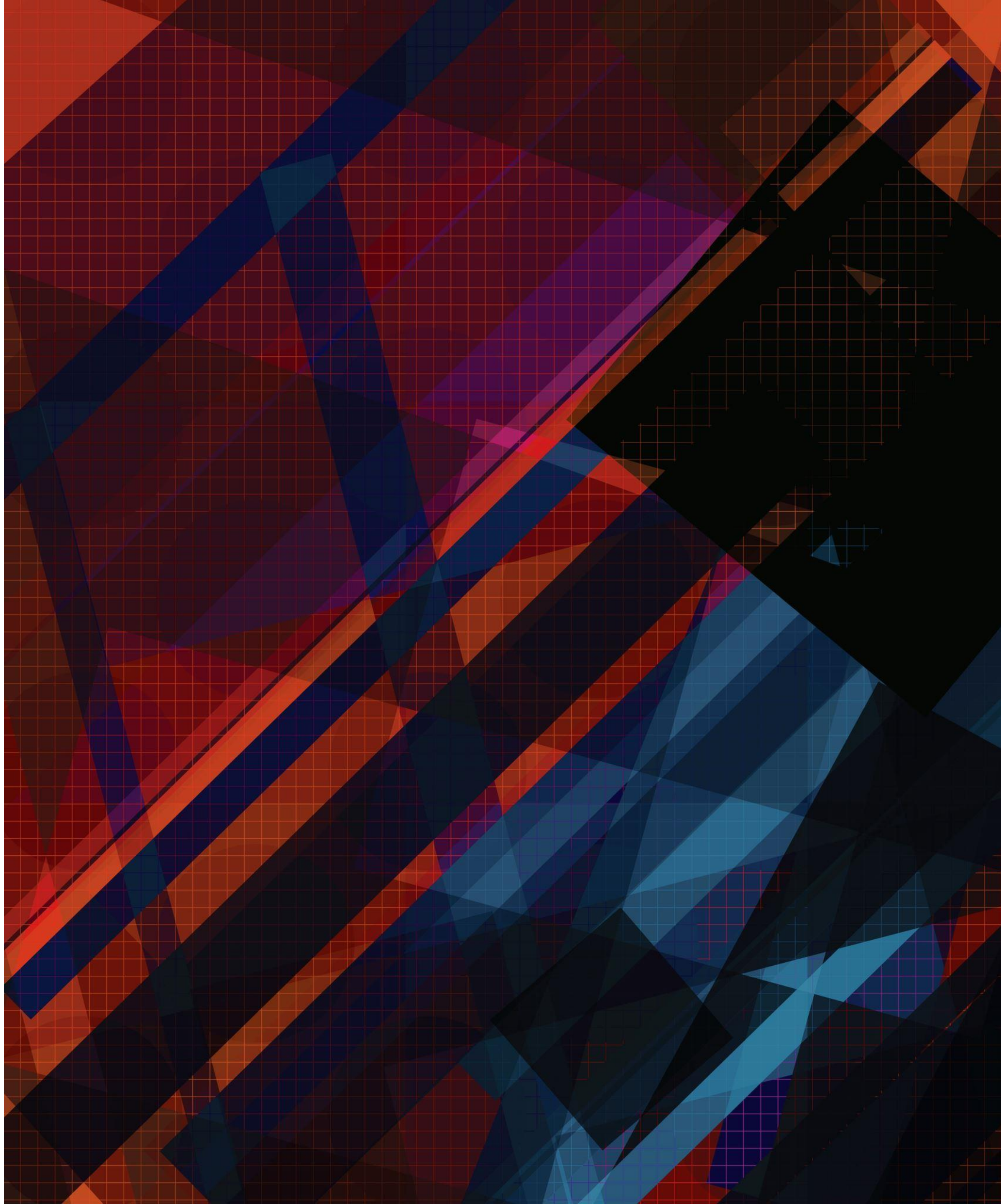
SIGNAL PROCESSING (FOR MACHINE LEARNING)



WHY IS **SIGNAL PROCESSING** NEEDED IN **AI** ?

WHEN SIGNAL PROCESSING IS NOT NEEDED IN AI

- Natural Langugae Processing (**NLP**) that deosn't involve the speech component
- Tabular structured data (**CSVs**) - various tasks
- Malware detection
- **Graph modeling** (social networks analysis, knowledge graph representations)

WHEN IS SINGAL PROCESSING REQUIRED IN AI?

- **Computer vision**
  - Object detection and classification
  - Image segmentation
  - Image generation
  - Satellite image analysis
- **Biomedical** signal processing
  - EEG
  - EMG
  - Medical images – MRI, X-Ray, CT
- **Audio** processing tasks
  - Speech synthesis (Test-to-Speech)
  - Automatic Speech Recognition (ASR)
  - Source separation

1D SIGNALS
TIME DOMAIN

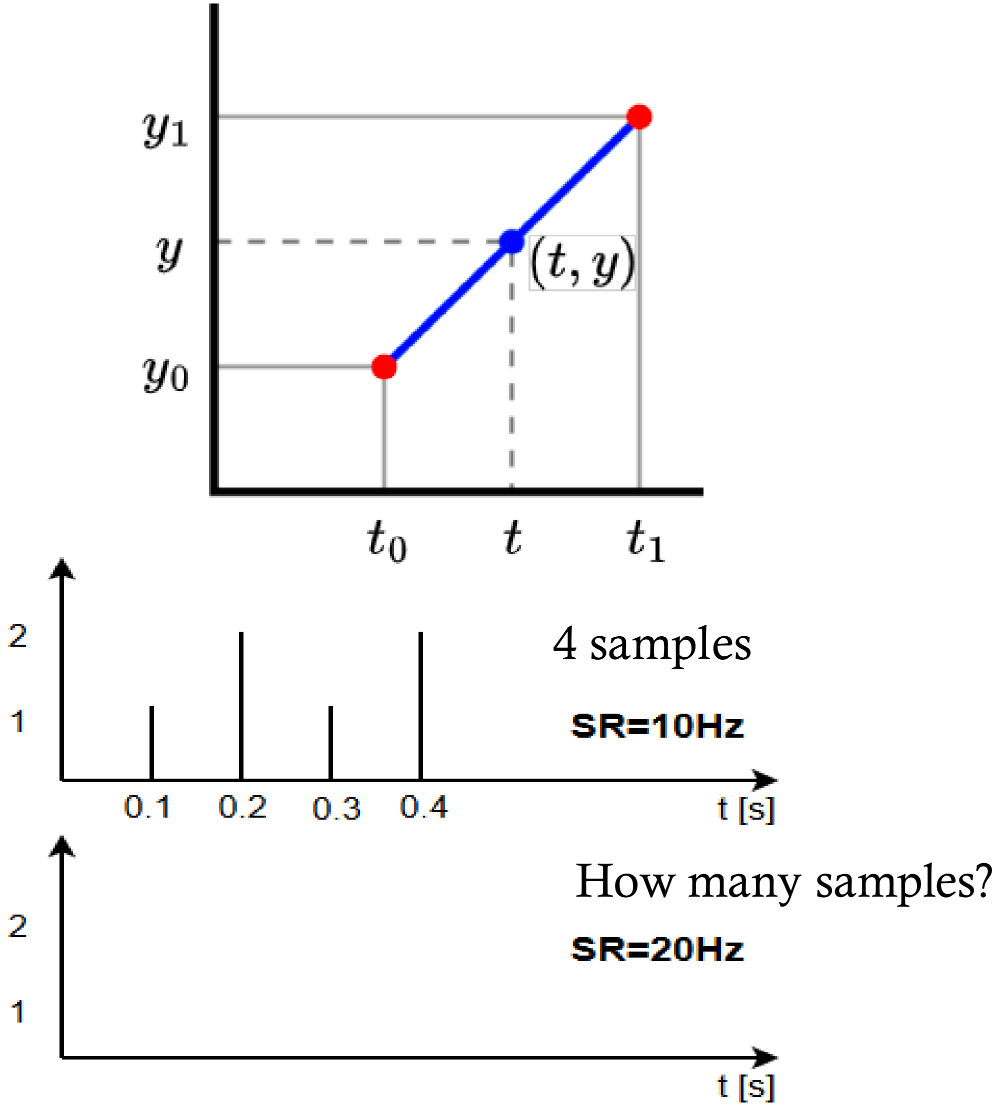
- Let's represent a signal in time domain:
    - What is **sample rate** (rata de eșantionare) ?
    - What is **quantization** (cuantizare) ?

1D SIGNALS RESAMPLING

- Example of method:
- linear interpolation between samples (t0, y0) and (t1, y1), at time t:

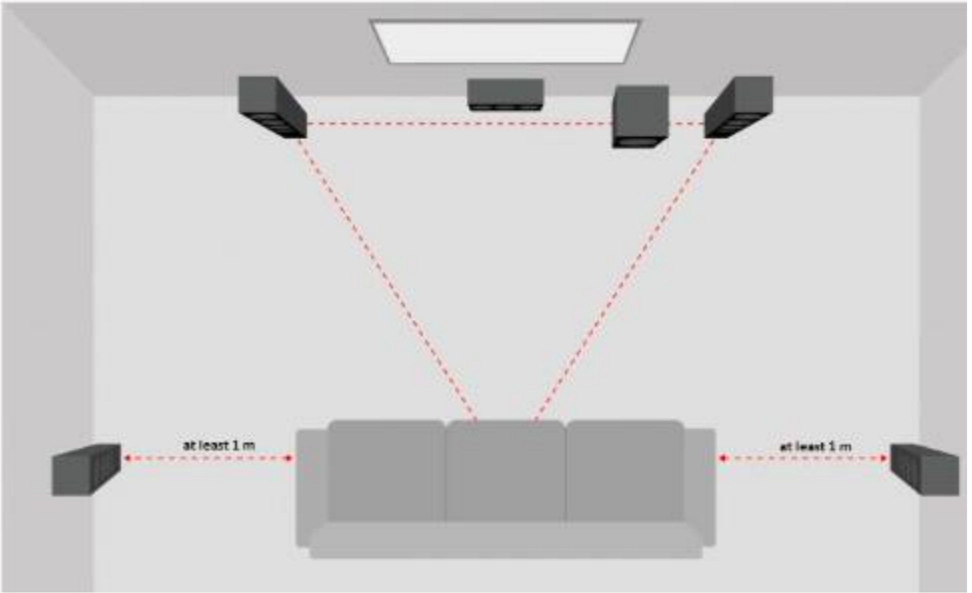
$$y(t) = y_0 + (y_1 - y_0)\frac{t - t_0}{t_1 - t_0}$$

- Simple scenario:
  - A) resample from 10 Hz to 20 Hz
  - B) resample from 20Hz to 10Hz
- **WHY IS SAMPLE RATE (SR) CRITICAL FOR MACHINE LEARNING?**



1D SIGNALS
CHANNEL NUMBER

- **Audio**
- Mono (1 source)
- Stereo/Binaural (2 sources)
- Home cinema (Ex: 5.1 - RL RR FL FR C SBW)
- **Other signals**
  - o Ex: EMG bracelet with 8 channels



Why does it matter?

1D SIGNALS
TIME DOMAIN FEATURES

- Used as **features for ML models**!
- ZCR (Zero crossing rate)
  - The times the signal changes its sign per second
- RMS Energy
  - Overall average 'power' of a signal
- Autocorrelation
  - Multiply the signal with itself – compute average
  - Apply for multiple values of the *time lag k*
- Computed on **SMALL WINDOWS** (**WHY?**)

$$ZCR = \frac{1}{N-1}\sum_{n=1}^{N-1} |s(n) \cdot s(n+1) < 0|$$

$$RMS = \sqrt{\frac{1}{N}\sum_{n=1}^{N} x(n)^2}$$

$$R_x(k) = \sum_{n=0}^{N-k-1} x(n) \cdot x(n+k)$$

1D SIGNALS FREQUENCY DOMAIN

- Fourier Transform
  - express the signal as a sum of cosines

$$X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x[n] e^{-j\omega n}$$

$$\cos(2\pi f t) - j\sin(2\pi f t)$$

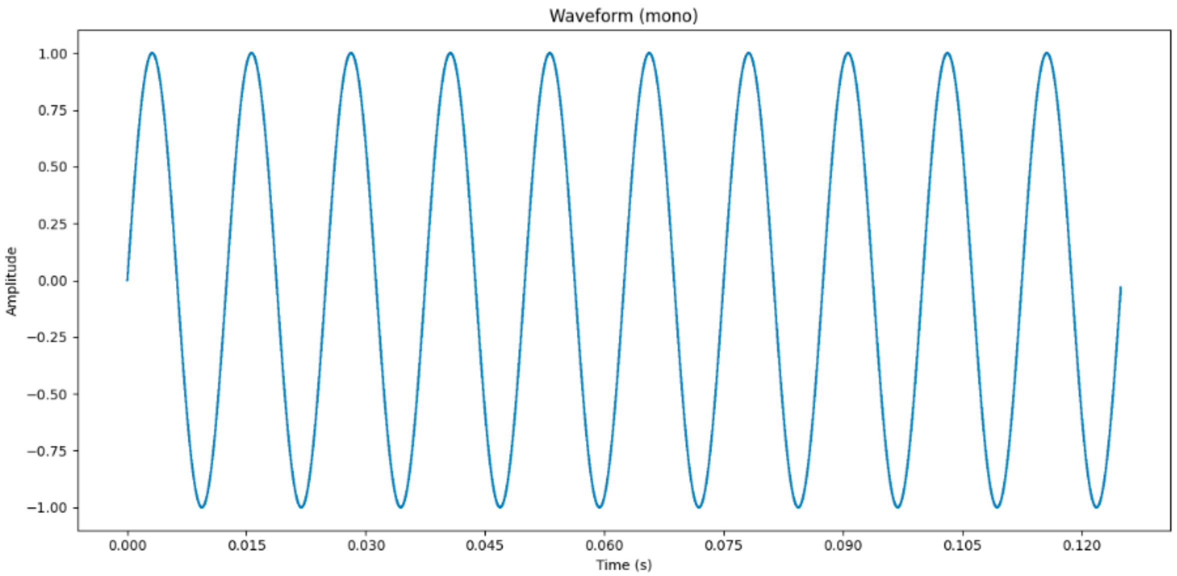
$$X(f) = a + jb$$

$$|X(f)| = \sqrt{a^2 + b^2}$$

$$\phi(f) = \text{atan2}(b, a)$$

1D SIGNALS
FOURIER

- Fourier Transform
  - express the signal as a sum of cosines which are:
    a) Scaled an amplitude $|X(f)|$
    b) Shifted in phase (delayed) by $\Phi(f)$



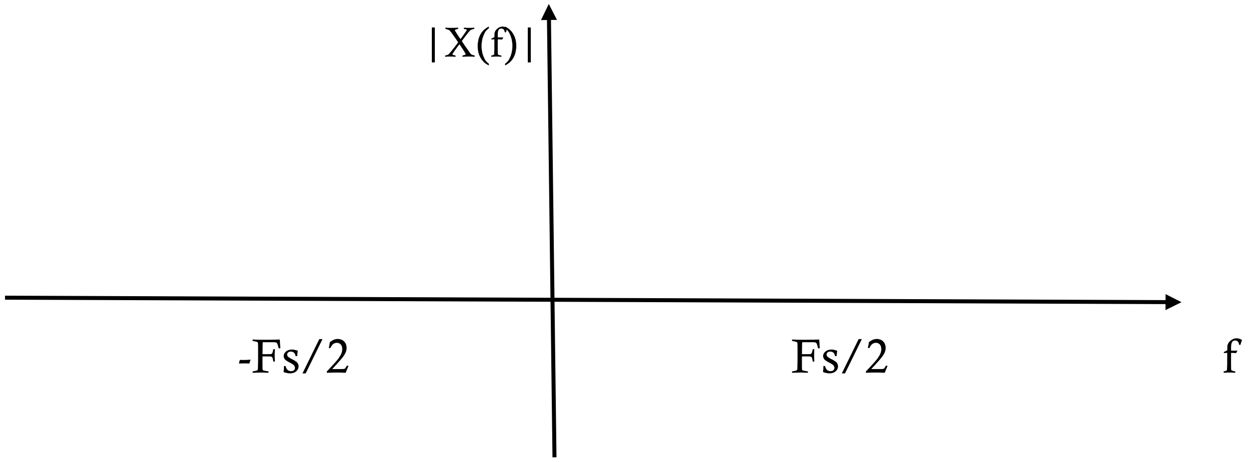
1D SIGNALS
FOURIER

- Some real-life signals may be estimated using this method
- But in most cases, an infinite number of terms are required for perfect stimation

1D SIGNALS
DFT. NIQUIST

$$X_s(f) = \frac{1}{T_s} \sum_{k=-\infty}^{\infty} X(f - kF_s)$$

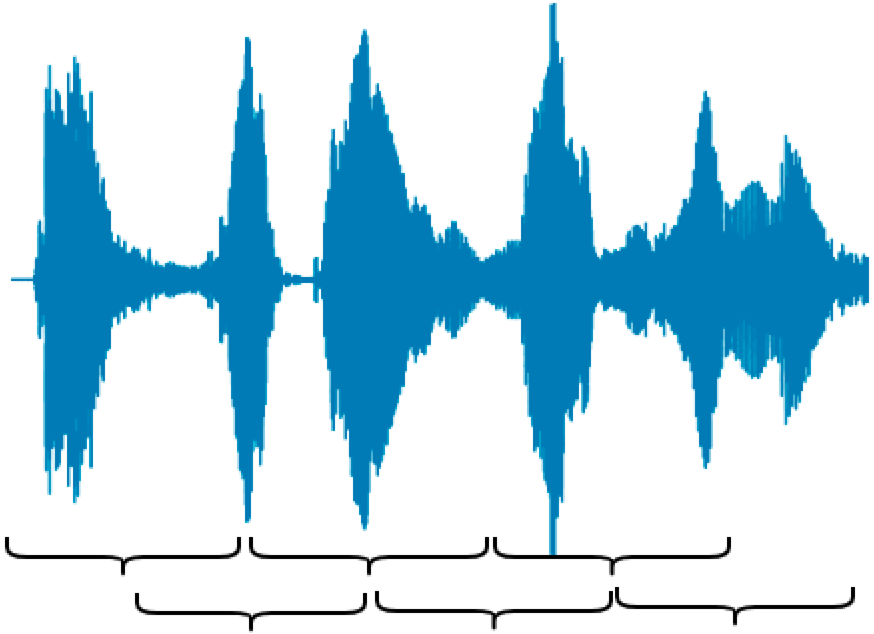
- DFT = Discrete Fourier Transform
- **Th. Niquist**: The sampling frequency has to be at least double the maximum frequency component of the signal.
  - Fsample >= 2 max{Fsignal}
- **But why?**
- The spectrum has **period Fs**



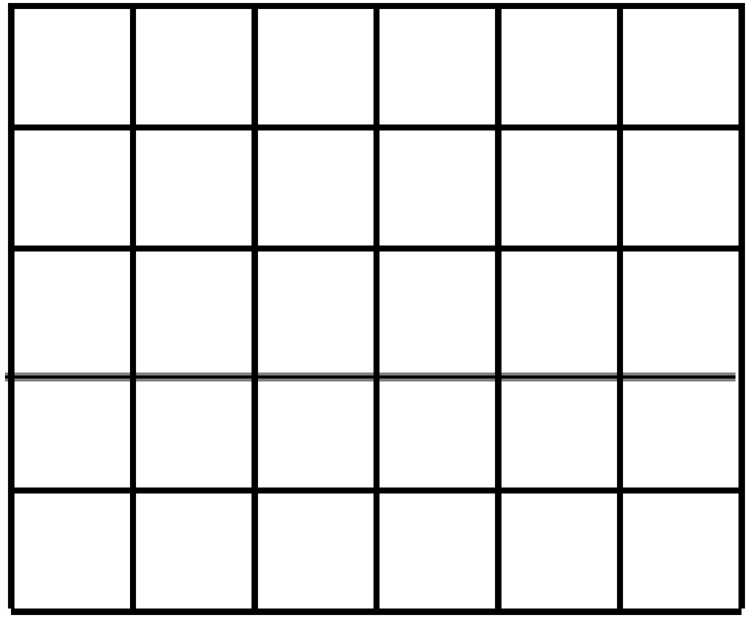
1D SIGNALS
TIME-FREQUENCY DOMAIN (SPECTROGRAMS)

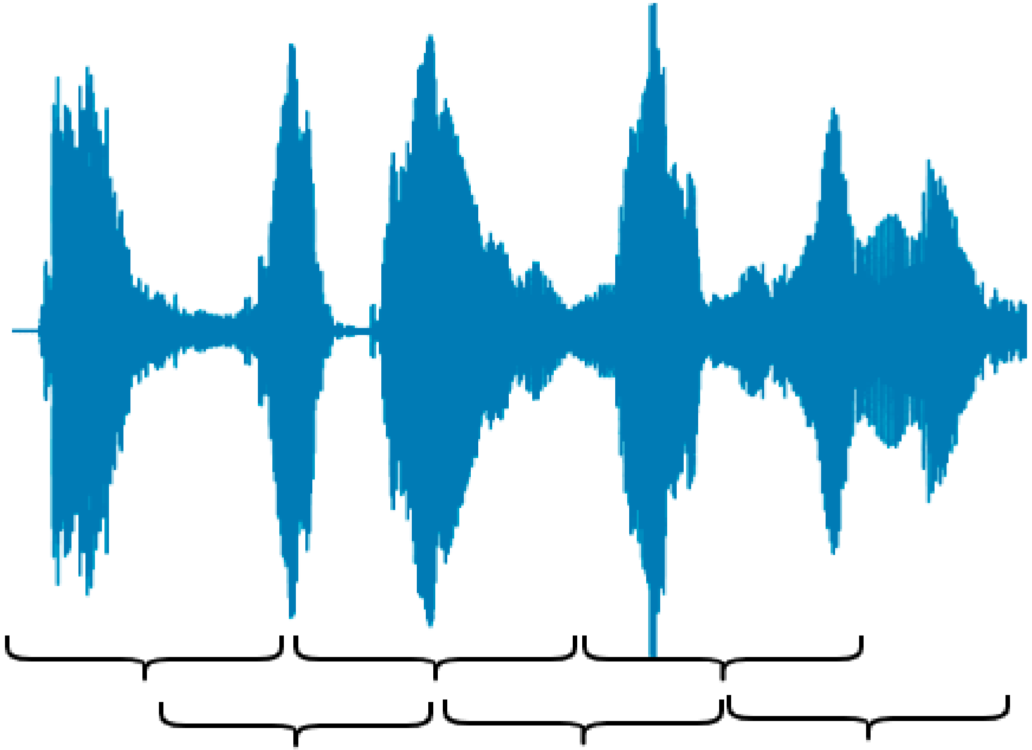
- **Is frequency domain enough?**

- Let's discuss:
  - Windowing
  - Window overlapping – **why?**
  - **FFT assumes the signal is infinite => sharp edges!** => Hamming or Han windows
  - Creating spectrograms



1D SIGNALS
SPECTROGRAMS



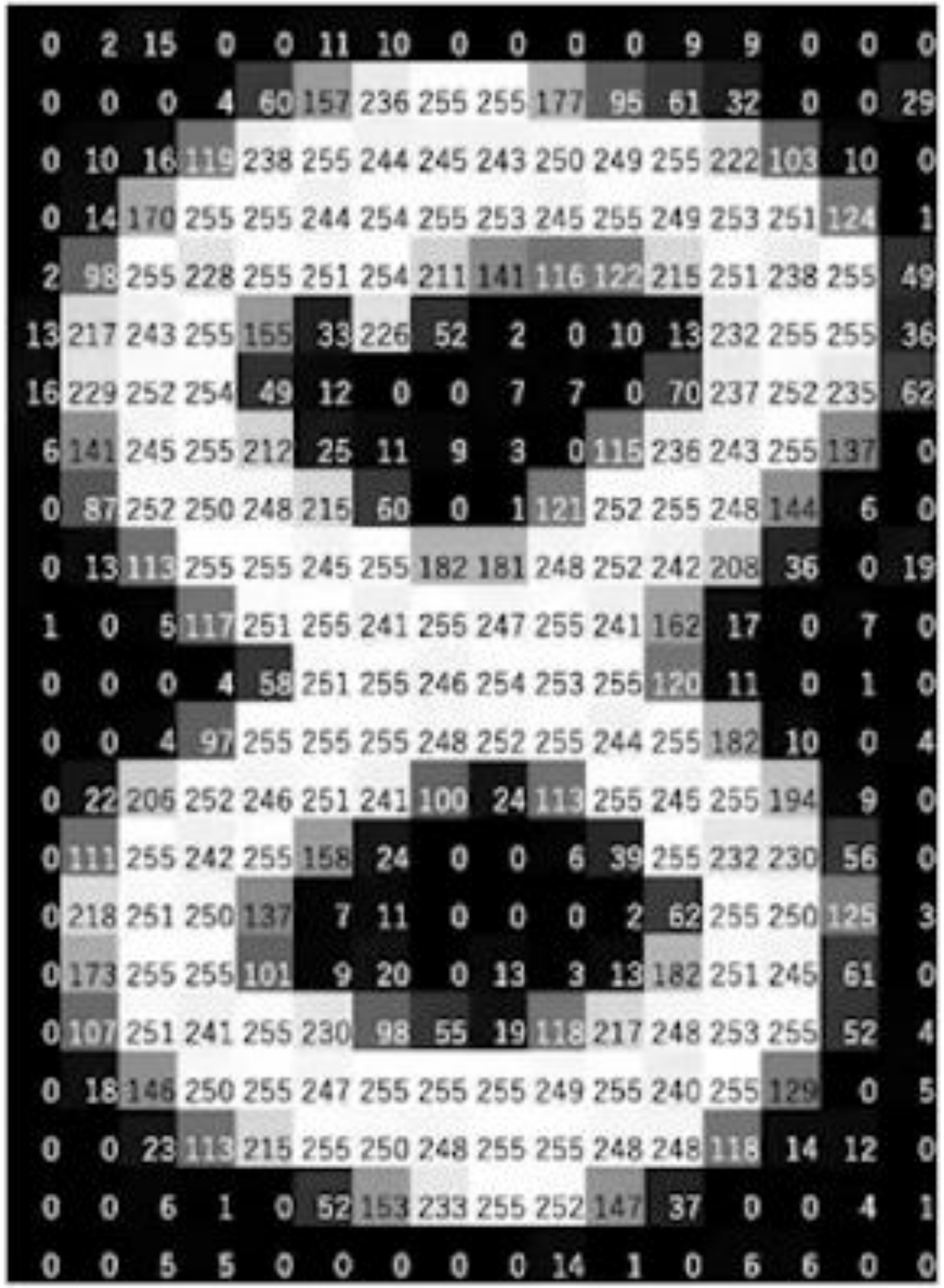


1D SIGNALS
PSD

- PSD = Power Spectral Density
- PROBLEM: FFT is a sum/integral
    - => the same sine, but LONGER (in time) will lead to higher values.
- SOLUTION: PSD shows the POWER of the frequency component
    - => is time duration agnostic.
- Remember: Power = Energy / Time
- PSD is frequently used in brain signal analysis (**EEG**)

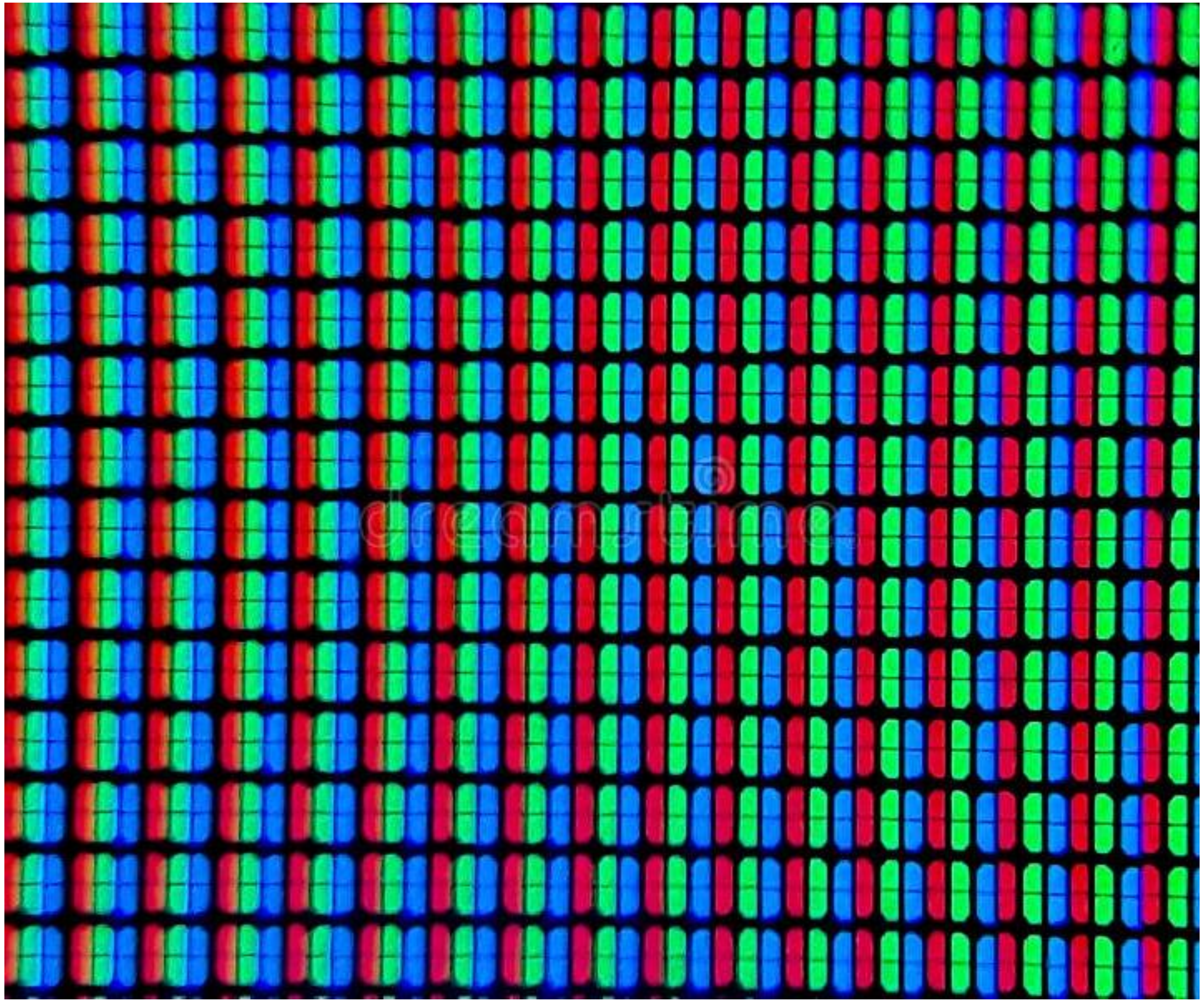
2D SIGNALS
IMAGE REPRESENTAION

- Vectorial images (svg)
- Raster images - **MATRICES**
  - o Grayscale (1channel)
  - o Color RGB (3 channels)
  - o Others (satellite, medical, IR)



2D SIGNALS
RGB CHANNELS

- Common color images have 3 values, for each pixel
  - R (Red)
  - G (Green)
  - B (Blue)



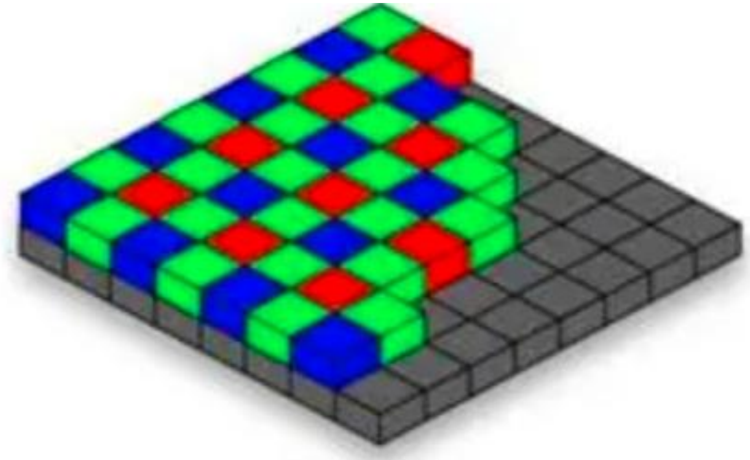
2D SIGNALS
BIT DEPTH

- An image = a **MATRIX**
- **Question 1**: for one individual element of the image, what are the values it can take? What range is it is?
- **Question 2**: does it matter if the total number of values is lower or higher?

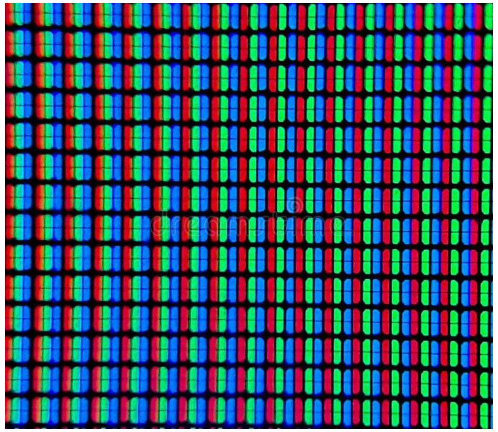
2D SIGNALS QUANTIZATION

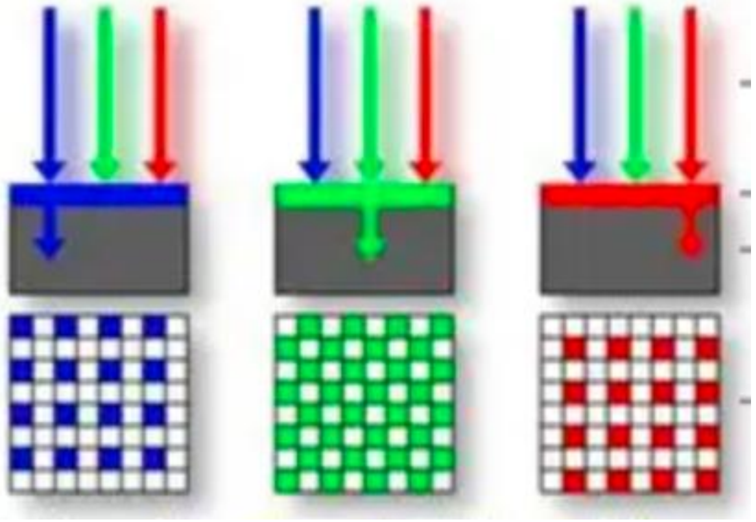
- Usually, images use 8 bits for every gray level
- **Question**: how many bits are required for an RGB image pixel?

2D SIGNALS
COLOR SPACES



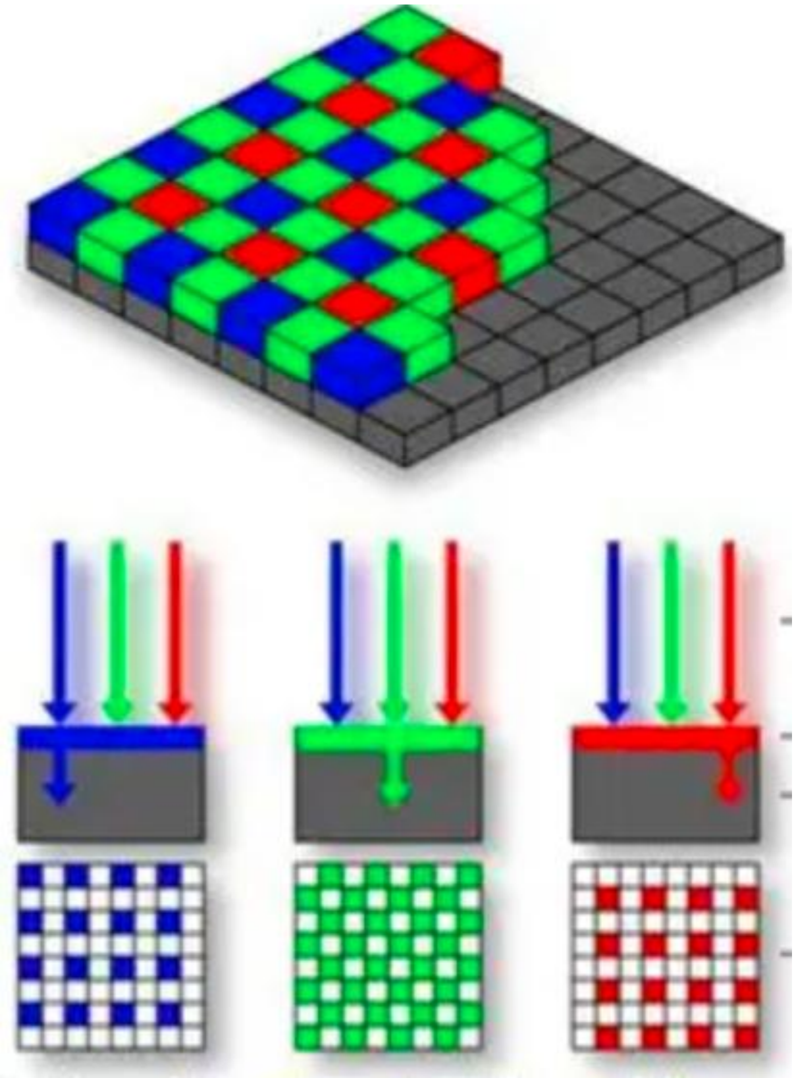
- Usually, we express images in RGB format, but **WHY**?
- **Is RGB the best color space?**





2D SIGNALS
THE BAYER GRID

- Why are there more **green** sensors in the Bayer grid?

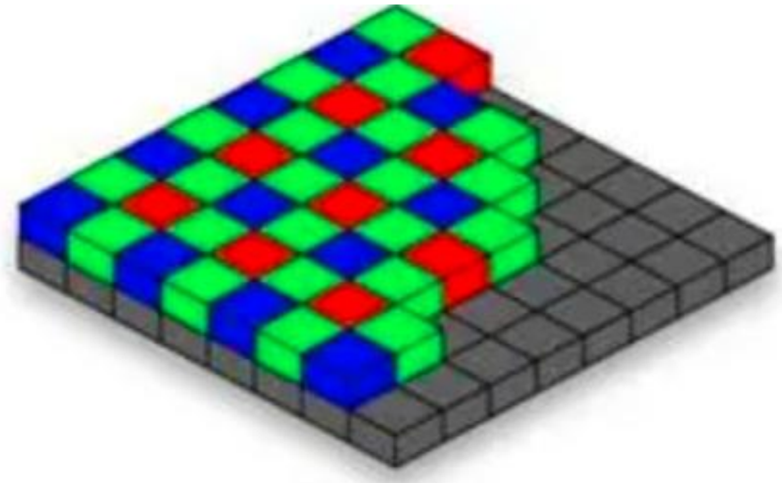


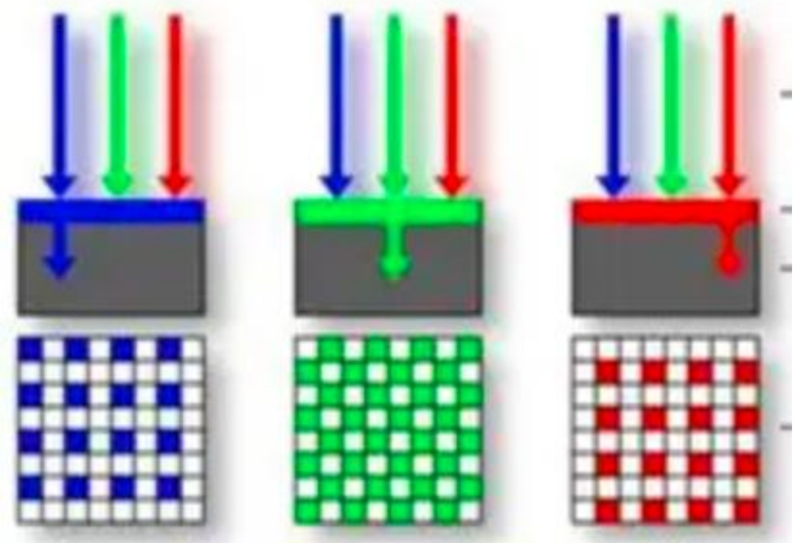
2D SIGNALS
THE BAYER GRID

- Why are there more **green** sensors in the Bayer grid? (compared to the others)
- **R: because of sensitivity**

- **(perceptual) Luminance** = perceptual brightness

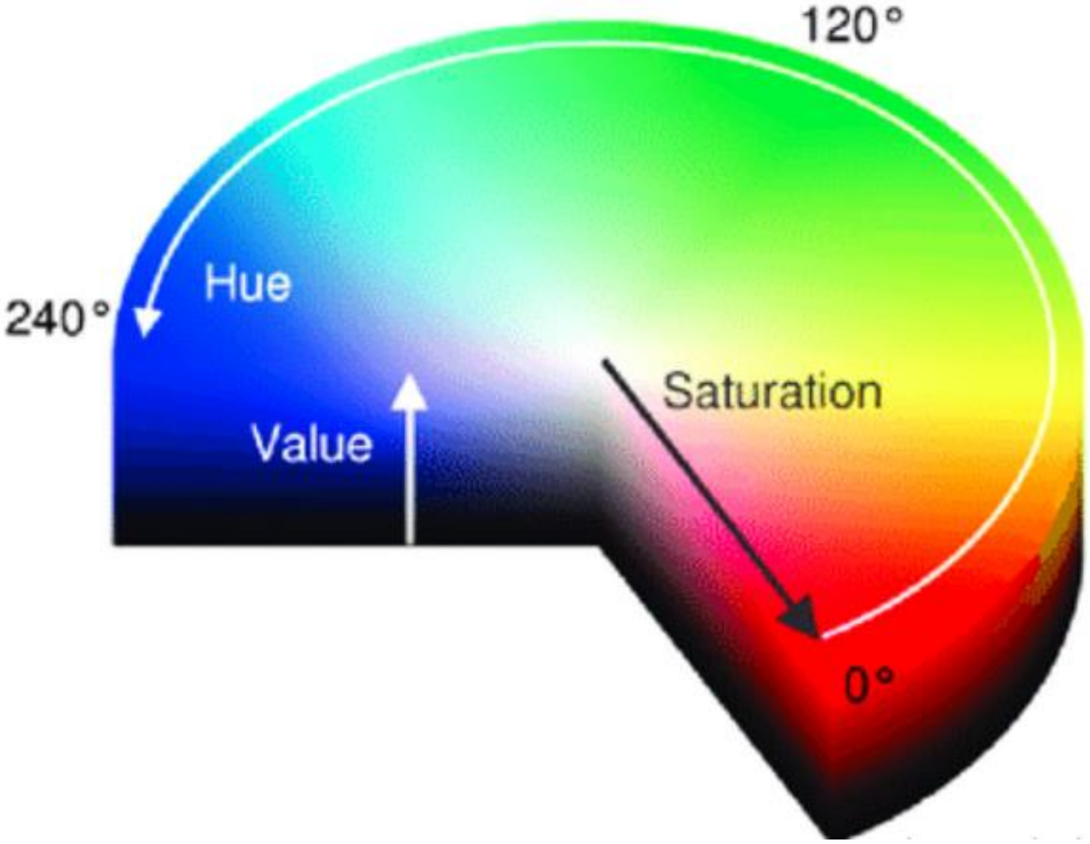
$$Y = 0.299 \cdot R + 0.587 \cdot G + 0.114 \cdot B$$





2D SIGNALS
OTHER COLOR SPACES

- HSV
  - **H**ue = "what kind of **colors** are used"
  - **S**aturation = "How much **pigment** is used" (distance from the gray axis)
  - **V**alue = "how light-dark it is"
- HSV is more **intuitive for humans**, but it is not perceptually uniform.
- HSV is sometimes used in ML for extracting handcrafted features when working with small datasets



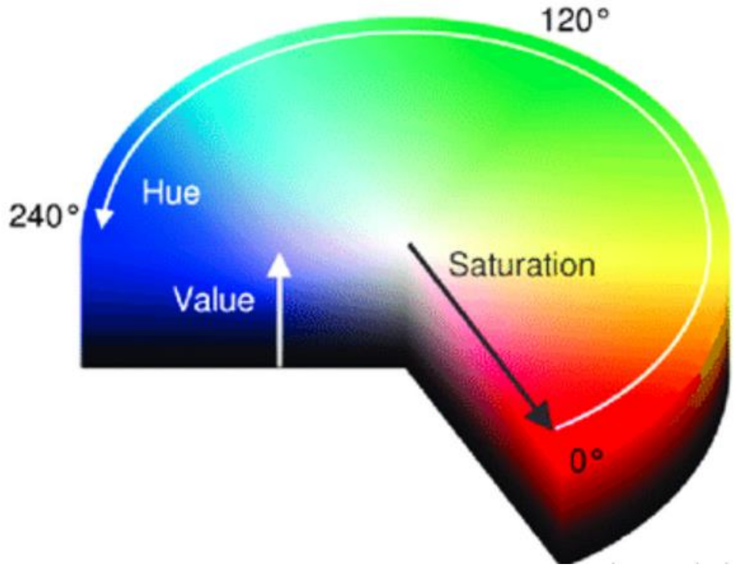
2D SIGNALS
HSV COLOR SPACE

- For each pixel, RGB-HSV transformation is defined by:

$$H = \begin{cases} \arccos\left\{\dfrac{(R-G)+(R-B)}{2\sqrt{(R-G)^2+(R-B)(G-B)}}\right\} & B \le G \\ 2\pi - \arccos\left\{\dfrac{(R-G)+(R-B)}{2\sqrt{(R-G)^2+(R-B)(G-B)}}\right\} & B > G \end{cases}$$

$$S = \frac{\max(R,G,B)-\min(R,G,B)}{\max(R,G,B)}$$

$$V = \frac{\max(R,G,B)}{255}$$

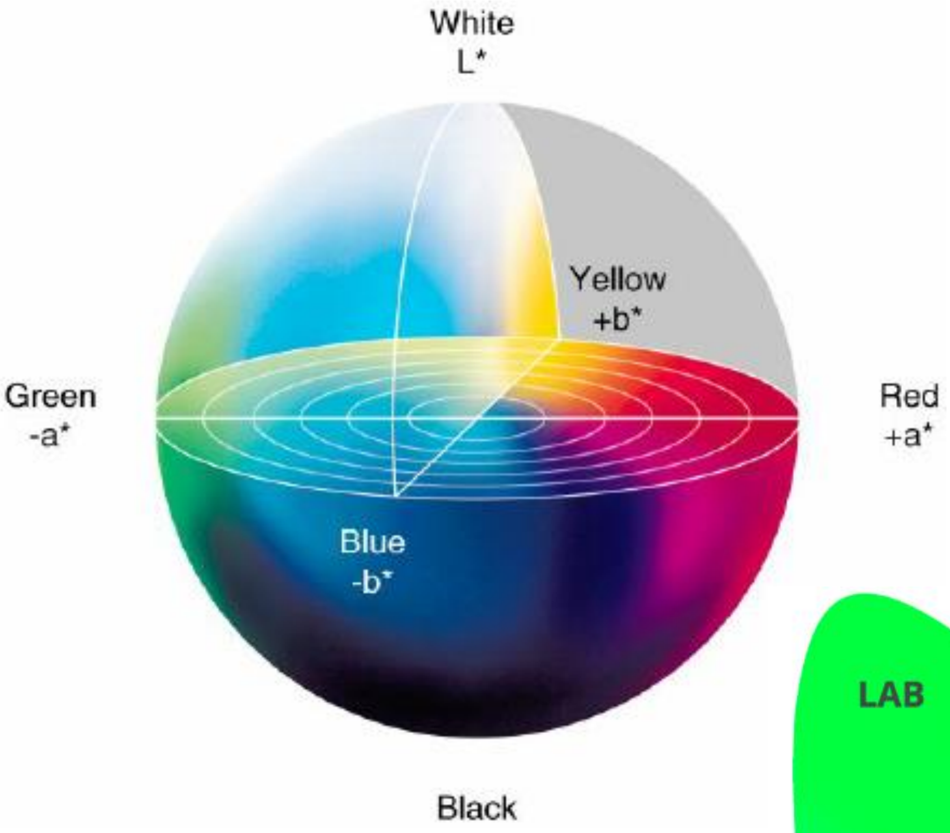


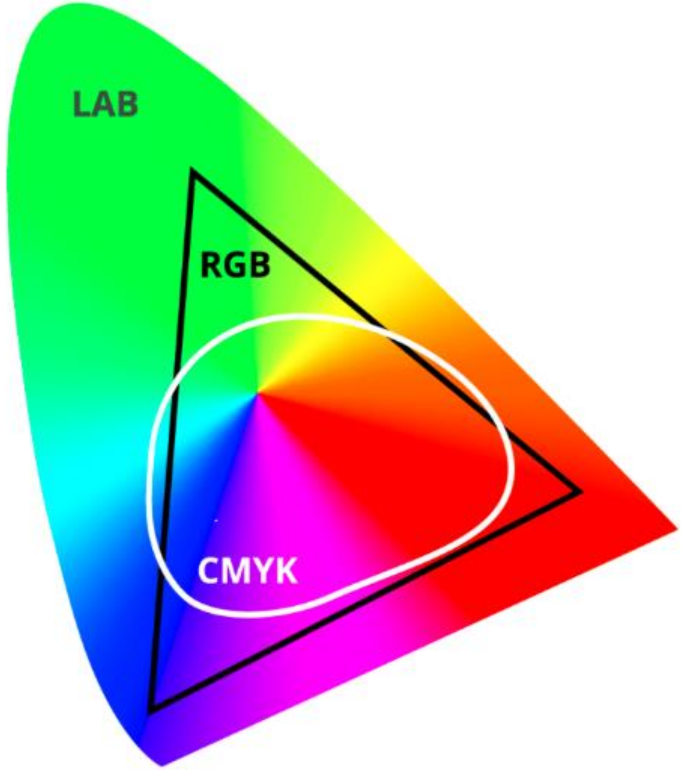
http://blog.csdn.net/

2D SIGNALS
LAB SPACE

- Lab:
    - o L = Luminance
    - o a= Green (-) - Red (+)
    - o B = Blue (-) - yellow (+)
- Lab is **perceptually uniform**

- **What is gamut?**
- **Gamut** = 'full range'





2D SIGNALS
HISTOGRAMS

BACK TO GRAYSCALE…

- A **histogram** shows how **gray values are distributed** in an image.

$$p(k) = \frac{h(k)}{N}$$

- h(k) = number of pixels with value k
- N = total number of pixels

0	0	1	1	1
0	2	2	3	3
2	2	2	3	3

2D SIGNALS CONVOLUTION

- A **filter slides** along the image
- For each location, the region from the image gets **multiplied elementwise** with the filter
- The output feature map gets assigned the **average of elements, after product**.

0	0	0	0	0
0	0	0	0	1
0	0	0	0	1
2	2	2	2	2
2	2	2	2	2
2	3	2	1	2

1	1	1
0	0	0
-1	-1	-1

Convolution result

- *convolution can be defined for every type of signal, **including 1D**, but it may be more intuitive in 2D

2D SIGNALS
FILTERING USING CONVOLUTION

- **HPF** (High Pass Filter)
- What is high frequency in an image?

- **LPF** (Low Pass Filter)
- What is low frequency in an image?

2D SIGNALS
FILTERING WITH CONVOLUTION

- **HPF** (High Pass Filter)
  - 'quick' variation in space
  - Details
  - Edges
- **LPF** (Low Pass Filter)
  - “Slow” variation in space
  - Smooth surfaces
  - Background / overall shapes

- **Derivation** filters
  - Coefficients sum up to 1
- **Smoothing** filters
  - Coefficients sum up to 1
  - All values positive

Sobel kernels

1	2	1
0	0	0
-1	-2	-1

1	0	-1
2	0	-2
1	0	-1

Averaging kernel

1/9	1/9	1/9
1/9	1/9	1/9
1/9	1/9	1/9

COMPRESSION

Why compression?

1. Lossless compression
2. Lossy compression

FILE FORMATS

🎵 **Audio**

- Wav
  - **Uncompressed**, raw waveform
  - Very large size
  - Usually used for training ML models
- flac
  - Compressed, **lossless**
  - **Smaller size** than wav
- MP3
  - Compressed, **lossy**
  - Much smaller size

🖼️ **Images**

- BMP
  - Uncompressed
  - Used for icons, thumbnails
- PNG
  - Compressed, **lossless**
  - **Larger size** compared to JPEG
  - Alpha channel
- JPG/JPEG
  - Compressed, **lossy => block artefacts**
  - **Smaller size** compared to JPEG

WHY DOES FILE FORMAT MATTER FOR AI?