# Software Technology of Internet of Things

## Sampling Physical Phenomena

Aslak Johansen asjo@mmmi.sdu.dk

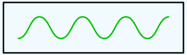
March 11, 2025

# Part 1:
# High-Level Overview

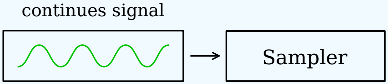
# General Data Collection Pipeline

# General Data Collection Pipeline

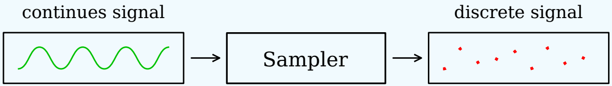
continues signal



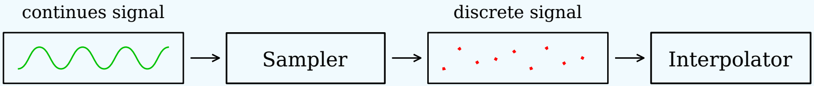
# General Data Collection Pipeline



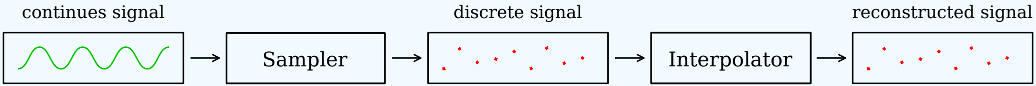
# General Data Collection Pipeline



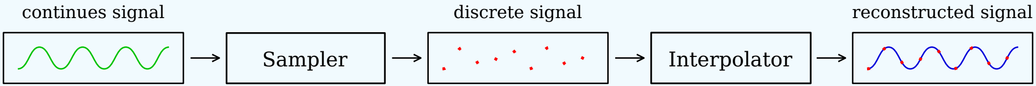
# General Data Collection Pipeline



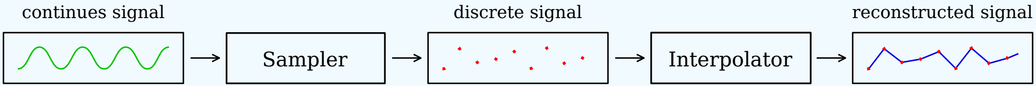
# General Data Collection Pipeline



# General Data Collection Pipeline



# General Data Collection Pipeline



# Part 2:
# The Temporal Aspect

# Sampling

We *can* take a single sample ... but usually we want more.

**Definition:** The *sampling interval* or *sampling period* $T$ is the time between two samples.

**Definition:** The *sampling frequency* or *sampling rate* $f_s$ is the average number of samples obtained within 1 second.

$$f_s = \frac{1}{T}$$

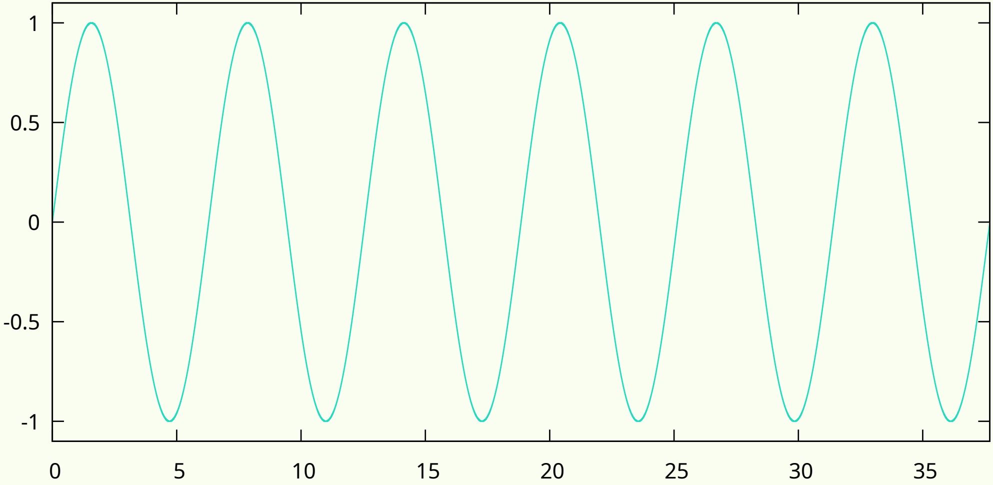
# Sampling Theorem

**Question:** How frequent do you have to sample to be guaranteed to be able to reconstruct a signal?

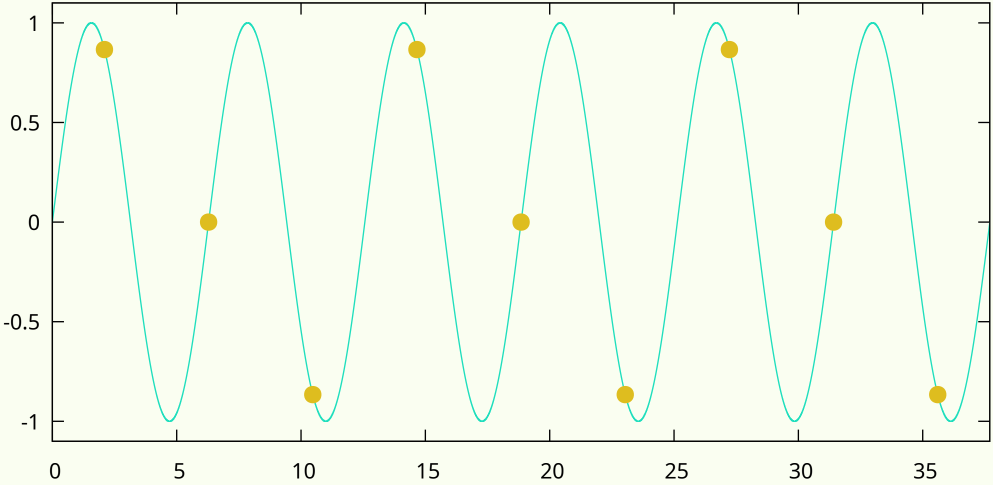
**Answer:** Twice as fast as the fastest frequency contained in the original signal.

**More info:** Look up Nyquist-Shannon sampling theorem.

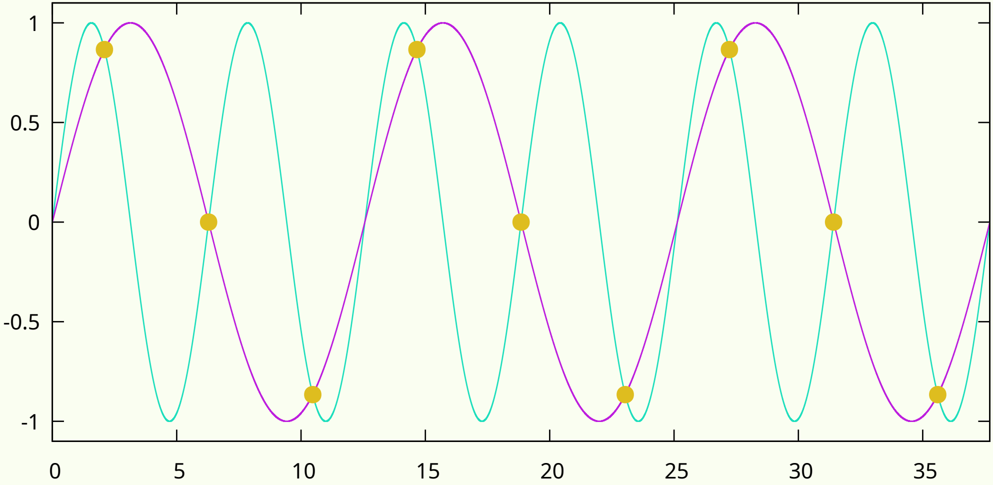
# **Sampling Theorem** Original Signal



# **Sampling Theorem** Samples



# **Sampling Theorem** Reconstructed Signal



# Nyquist Frequency Violations

**In Movies:** Too low a sampling rate (relative to the rotational speed of [the spokes of] a wheel) can make it appear to rotate the wrong way.

**In Photos:** When an image of fabric is scaled (without some form of softening) Moiré patterns tend to appear. The same can happen when the pattern of the fabric iterferes with the pattern of the sensors color filter array.

**In Audio:** The similar effect is called aliasing.

# Resampling

**Definition:** A *timeseries* is a mapping from time to values (e.g., samples):
$\mathit{time} \mapsto \mathit{value}$

At times the frequency of a timeseries need to be adapted to a specific purpose.

This operation is referred to as *resampling*:

- ***Downsampling*** By generating a new timeseries representing the same signal using less samples.
- ***Upsampling*** By generating a new timeseries representing the same signal using more samples.

Resampling introduces errors, including (but not limited to)

- Cutting off extreme values.
- Removing fluctuations.

# Part 3:
# Analog to Digital Conversion

# Sample and Hold Method

1. ***Flush*** Drain the charge of the capacitor by connecting it to ground.
2. ***Wait*** Make sure that input is not connected to ground.
3. ***Sample*** Connect the capacitor to the input. Over time this will charge the capacitor.
4. ***Hold*** Given a large enough sample time the capacitor is charged to a voltage matching that of the input.
5. ***Compare*** This charge can then be compared to multiple known values, once for each output bit.

# Compare Step

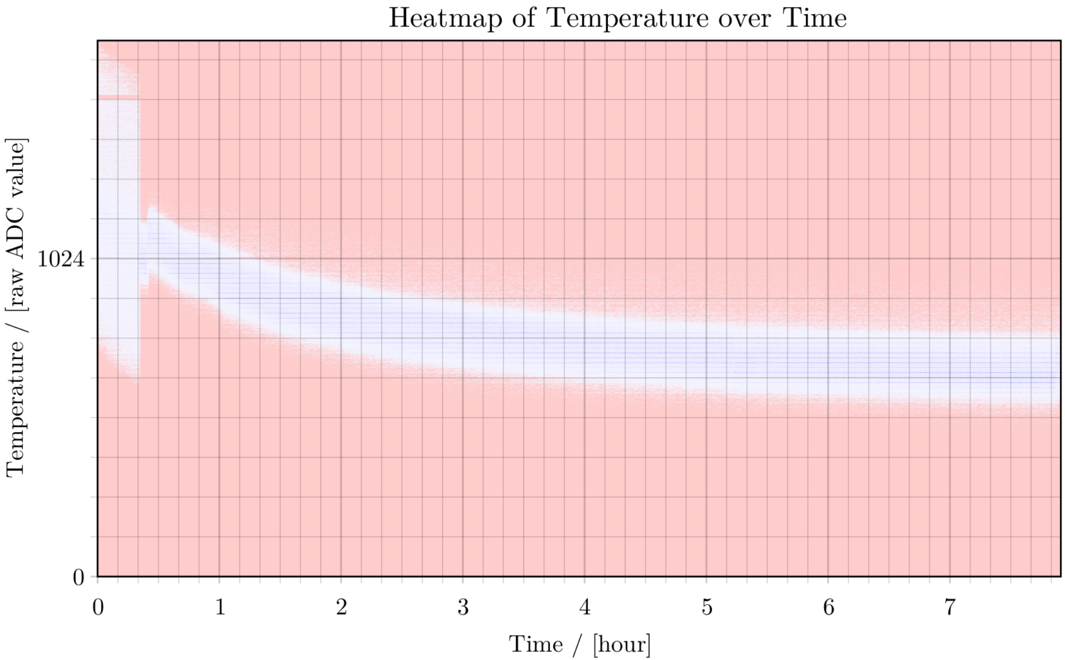
There are different methods for constructing ADCs.

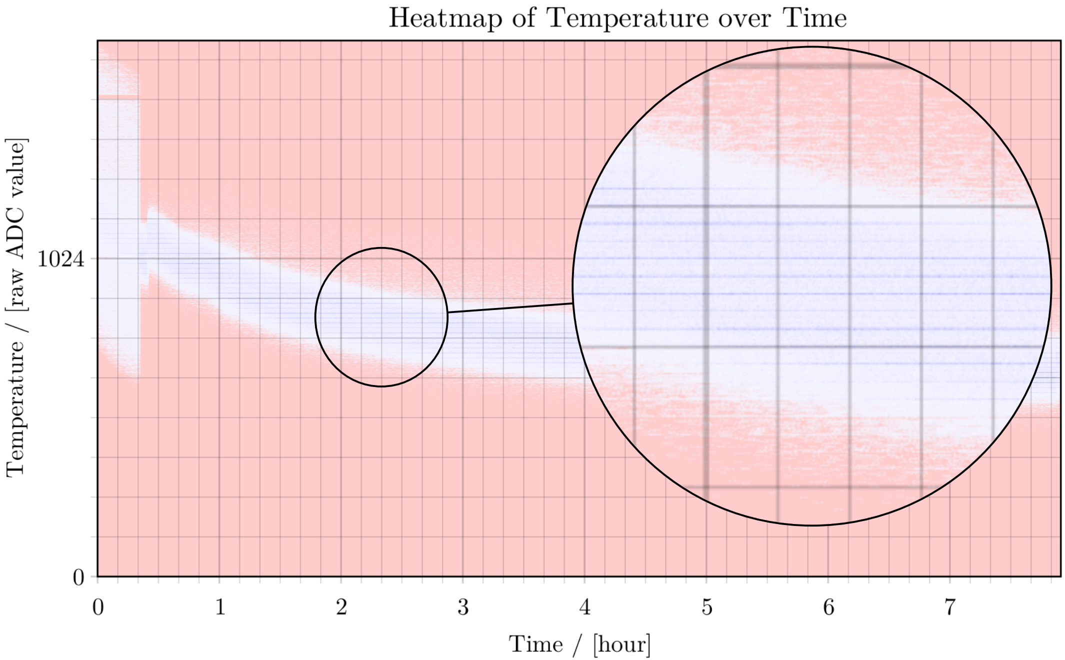
One comparison is made per bit. Designs spans the spectrum from all-parallel to all-sequential.

The known values (representing bits) may be laid out in either a linear (a *linear* ADC) or a logarithmic (a *logarithmic* ADC) pattern.

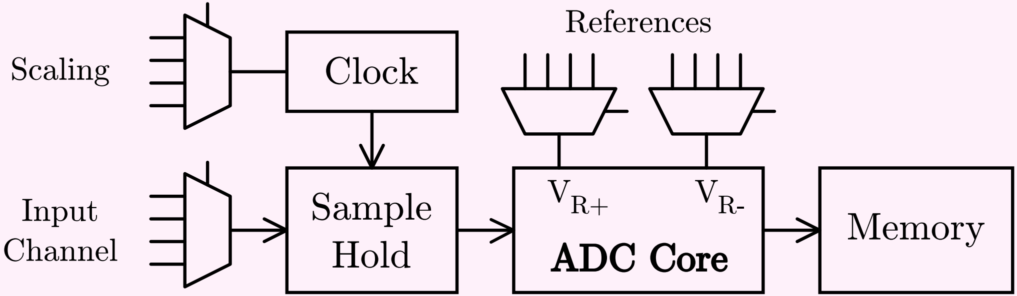
Why is this?

All designs have (more or less) common errors.


Heatmap of Temperature over Time
Temperature / [raw ADC value]
1024
0
0
1
2
3
4
5
6
7
Time / [hour]


Heatmap of Temperature over Time
Temperature / [raw ADC value]
1024
0
Time / [hour]
0
1
2
3
4
5
6
7

# Analog to Digital Converter



# Part 4:
# Imperfections

# Noise and SNR

Noise has many sources:

- Often the sensor itself produce a noisy signal
- Analog electrical signals pick up noise from electromagnetical signals
- AD conversion introduces error

The inherent noise in a given situation is known as the *noise floor*.

The *signal to noise ratio* (SNR) is a measurement that tells us how large a fraction of a signal consists of noise.

What happens if you average two samples?

1. The signal is doubled.
2. Many types of noise will cancel out(ish).
3. Accordingly, the SNR goes up ☺.

# Outliers

What is a outlier?

What makes a sample an outlier?

**Definition:** An outlier is an observation (aka sample) which lies *"far"* away from the brunt of the distribution.

**Note:** This is a fairly vague definition.

In order to avoid giving credence to such outliers by assigning them weight of more than their (expected) rate of occurrence would suggest these samples are often removed in a process known as *outlier removal*.

# When Outliers Become Black Swans

What happens if you do analysis on a dataset where removed sample(s) turns out to be significant?

Something may happen that your models simply don't account for; something that they say is impossible.

Such events are known as *black swans*, and they do occur.

Often with serious consequences.

Examples:

- 2020 COVID-19 pandemic
- 2011 Fukushima nuclear accident
- 2008 Financial crisis
- 2001 Dotcom bubble

# Part 5:
# Offloading

# Sampling Bandwidth

| | 1Hz | 10Hz | 100Hz | 1.0kHz | 10.0kHz | 100.0kHz | 1.0MHz |
|---|---|---|---|---|---|---|---|
| 6b | 6bps | 60bps | 600bps | 6.0kbps | 60.0kbps | 600.0kbps | 6.0Mbps |
| 7b | 7bps | 70bps | 700bps | 7.0kbps | 70.0kbps | 700.0kbps | 7.0Mbps |
| 8b | 8bps | 80bps | 800bps | 8.0kbps | 80.0kbps | 800.0kbps | 8.0Mbps |
| 9b | 9bps | 90bps | 900bps | 9.0kbps | 90.0kbps | 900.0kbps | 9.0Mbps |
| 10b | 10bps | 100bps | 1.0kbps | 10.0kbps | 100.0kbps | 1.0Mbps | 10.0Mbps |
| 11b | 11bps | 110bps | 1.1kbps | 11.0kbps | 110.0kbps | 1.1Mbps | 11.0Mbps |
| 12b | 12bps | 120bps | 1.2kbps | 12.0kbps | 120.0kbps | 1.2Mbps | 12.0Mbps |
| 13b | 13bps | 130bps | 1.3kbps | 13.0kbps | 130.0kbps | 1.3Mbps | 13.0Mbps |
| 14b | 14bps | 140bps | 1.4kbps | 14.0kbps | 140.0kbps | 1.4Mbps | 14.0Mbps |
| 15b | 15bps | 150bps | 1.5kbps | 15.0kbps | 150.0kbps | 1.5Mbps | 15.0Mbps |
| 16b | 16bps | 160bps | 1.6kbps | 16.0kbps | 160.0kbps | 1.6Mbps | 16.0Mbps |

Required bandwidth as a function of sampling rate and sample depth.
Colored according to what (9600, 38400, 115200, 576k and 5M)b/s can handle.

# Streaming Samples

**Question:** Given a 12 bit ADC with a 2 bit noise floor sampling at 8 kHz, how large a bandwidth is needed for offloading over a UART?

Theoretical answer:

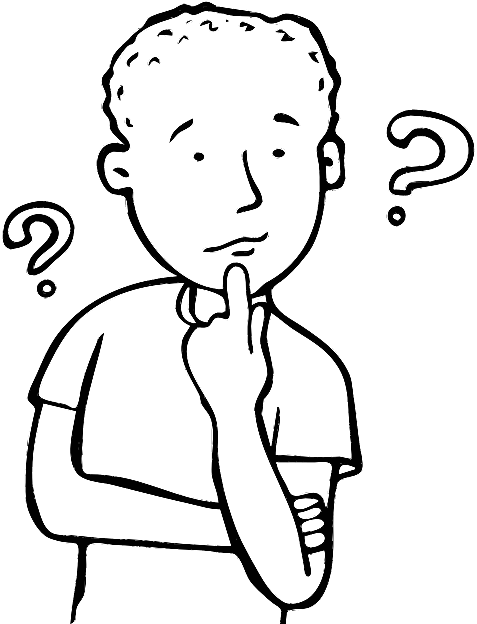
1. Each sample contains $(12 - 2)\text{b} = 10\text{b}$ information.
2. This gives us a required bandwidth of $10\text{b} \cdot 8\text{kHz} = 10\text{b} \cdot 8\text{k/s} = 80\text{kb/s}$.

Practical answer (UART edition):

1. If we assume a frame format of 1 start bit, 8 data bits, 0 parity bit and 2 stop bits, then for each $8\text{b}$ of information we will need to add $3\text{b}$ of control data.
2. That is a $37.5\%$ overhead resulting in a practical requirement of $80\text{kb/s} \cdot 1.375 = 110\text{kb/s}$.

Given what we have learnt about bit errors on serial lines, what is wrong with this?

# Questions?



https://openclipart.org/detail/238687/boy-thinking-of-question