Forensic Insights from Electromagnetic Radiation

Workshop at ICTer Conference 2023

Dr. Asanka P. Sayakkara

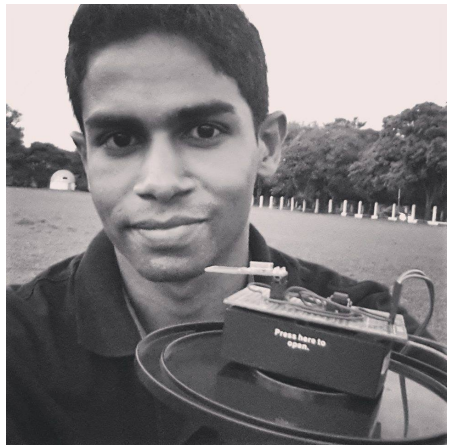
(asa@ucsc.cmb.ac.lk)

University of Colombo School of Computing
Sri Lanka.

10th November, 2023

Asanka P. Sayakkara

- BSc in Computer Science, University of Colombo School of Computing (UCSC), 2012.
- PhD in Computer Science from University College Dublin, Ireland, 2020.
- Forensic & Security Research (ForSec) group of University College Dublin, Ireland, 2017–2020.
- Senior lecturer at University of Colombo School of Computing (UCSC), Sri Lanka.
- Coordinator of the MCS/MSc in CS degree programs.



- Introduction to Computing (FoS), Digital Forensics, Embedded Systems, and Operating Systems II.
- Running *Signal Insights* research lab.
- `https://ucsc.cmb.ac.lk/profile/asa`
  `https://www.asayakkara.org`

Signal Insights Research Lab

- Potential of exploiting various signals originating from various sources.
- Signal sources: artificial, as well as biological sources (bioacoustics).
- Electromagnetic side-channels and covert channels.
- Radio tomographic imaging.
- Passive acoustic monitoring (of elephants).
- `https://www.asayakkara.org/signal-insights-lab.html`

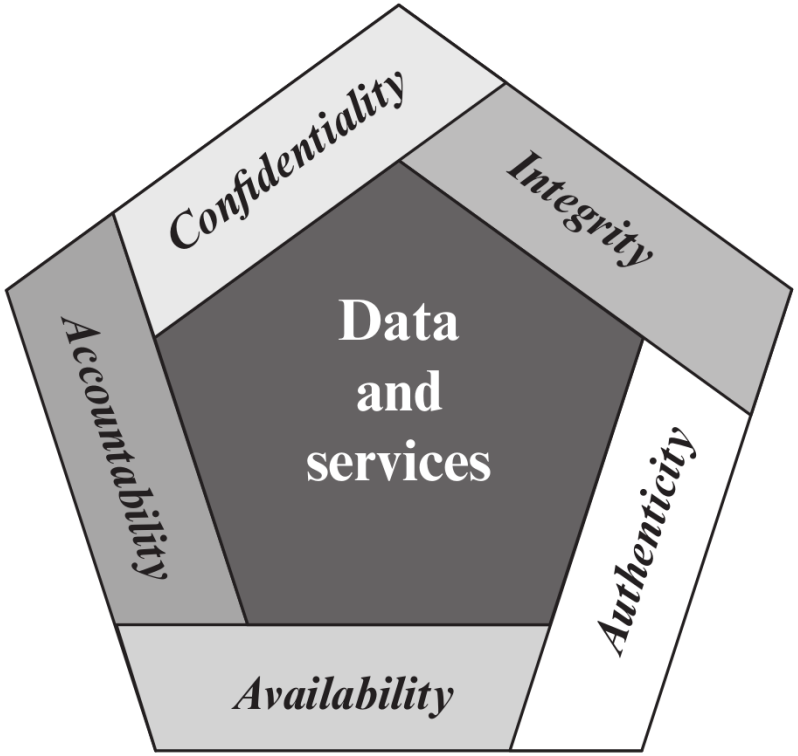
Workshop Agenda

- **Part 1:**
  - Hardware security.
  - Digital forensics.
  - Limitations of forensics.
  - EM-SCA for forensics.
- **Part 2:**
  - SDR hardware.
  - SDR software.
- **Part 3:**
  - EM trace data acquisition.
  - EM trace data processing.
- **Part 4:**
  - Exploring a large EM dataset.
  - Training ML models on EM data.

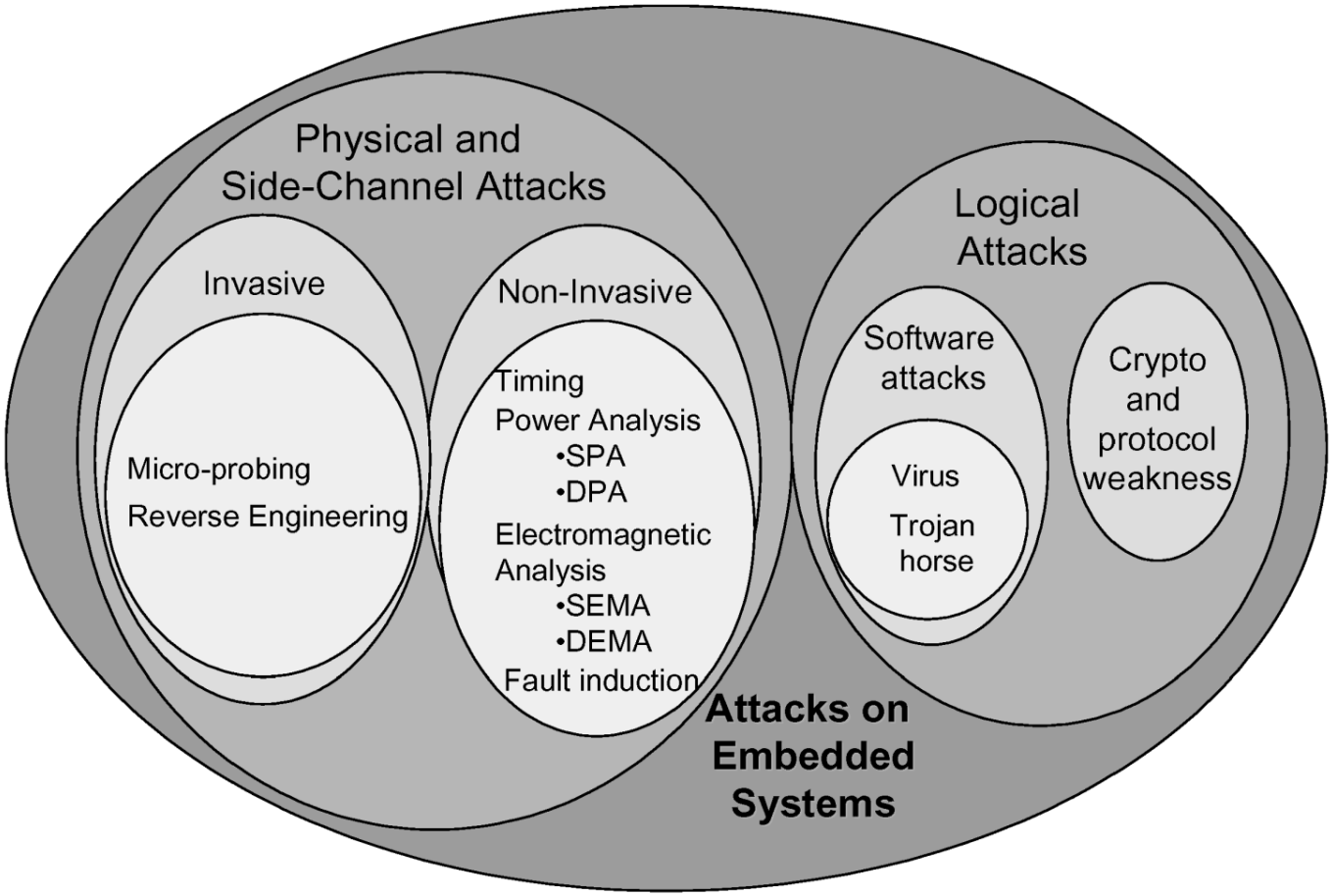
Part 1

Information Security

- Security is an essential element of modern computing systems.



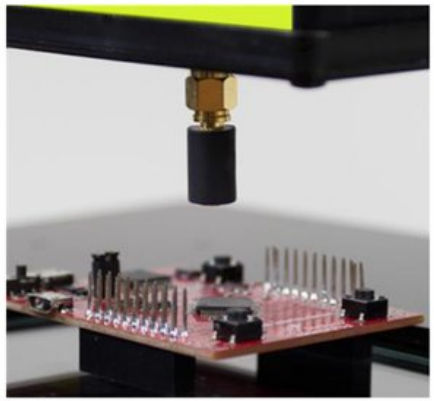
Hardware Security

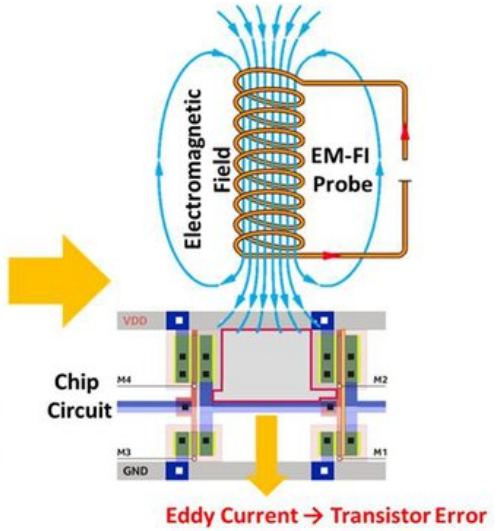


Fault Injection Attacks

- Modern ICs are designed to work within specific operating ranges.
- Faults arise whenever a deviation from the expected operating conditions occurs.
- Of particular importance to security researchers are the errors produced by faults that can be used to compromise the security of computing devices.

Technique	Accuracy (Spatial)	Accuracy (Temporal)	Cost	Risk (Damage)
Clock glitch	none	high	low	none
Voltage spike	none	high	low	low
Heat	low	none	low	low
EM Pulse	medium	medium	medium	medium
Laser beam	high	high	high	medium



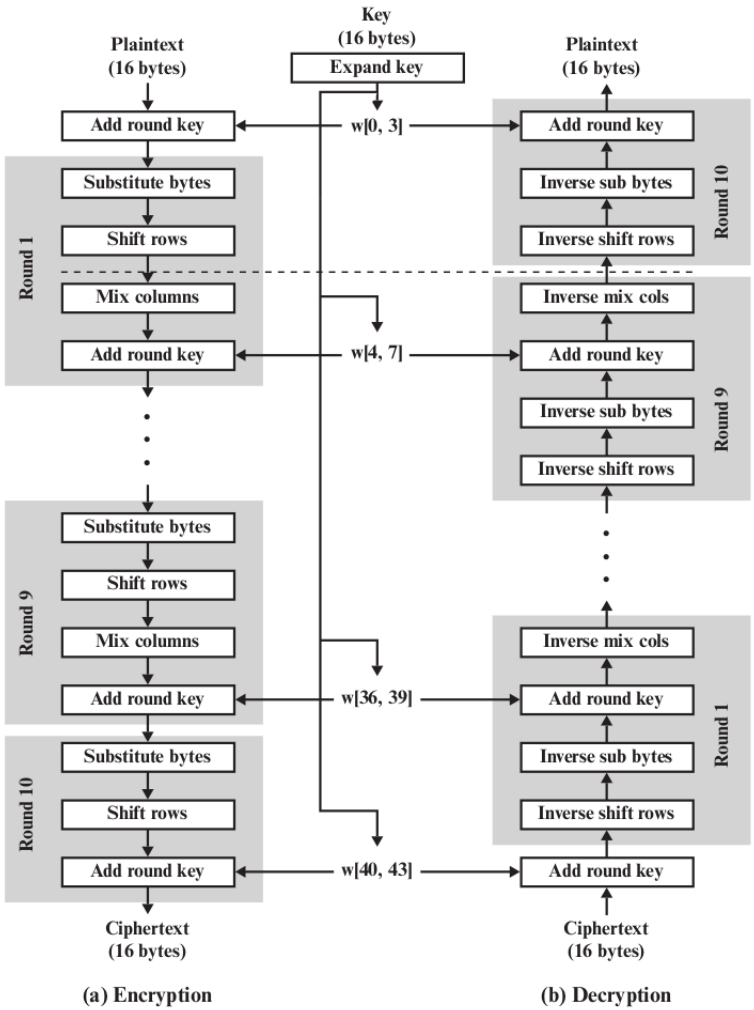


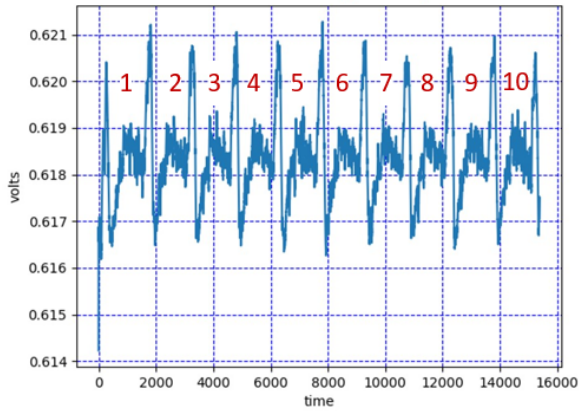
Side-Channel Analysis

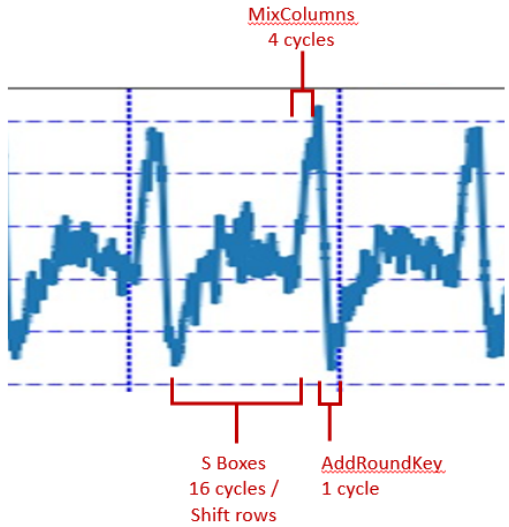
- Data encryption algorithms are designed assuming that the intermediate data they are handling will not be available to third parties.
- This assumption is not exactly correct for computers.
- Information leakes through unintended channels from computers:
  - Timing side-channel
  - Acoustic side-channel
  - Power side-channel
  - Electromagnetic side-channel
- By exploiting such side-channels, attacks can be mounted at computing systems to extract confidential data.

Simple Power Analysis – SPA

Identifying AES algorithm

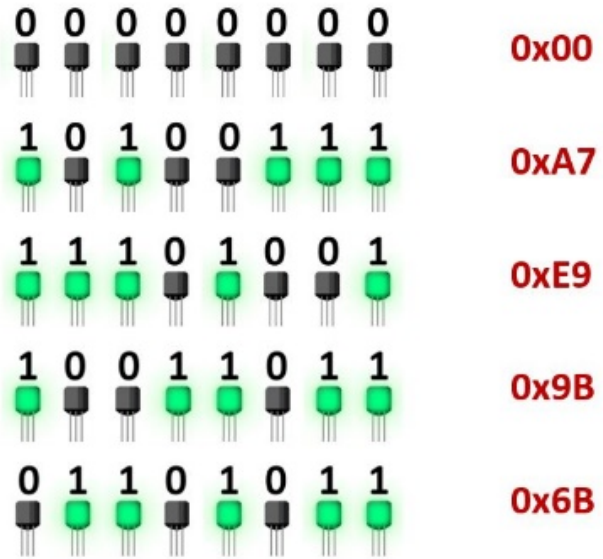




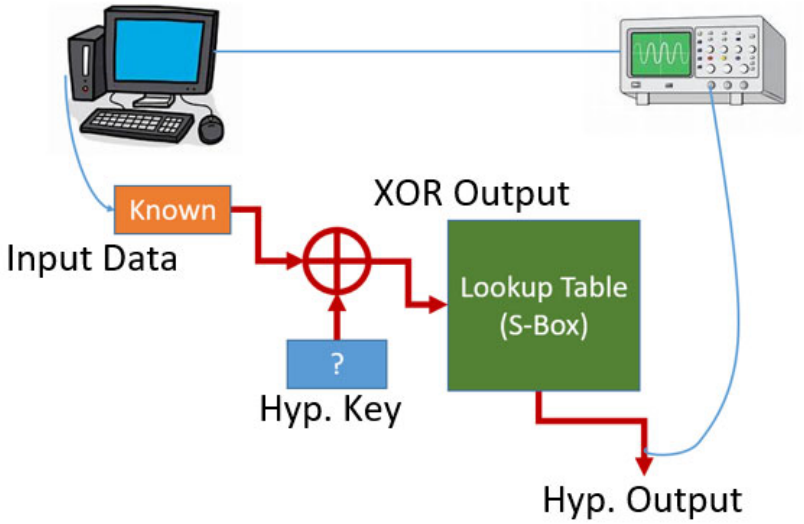


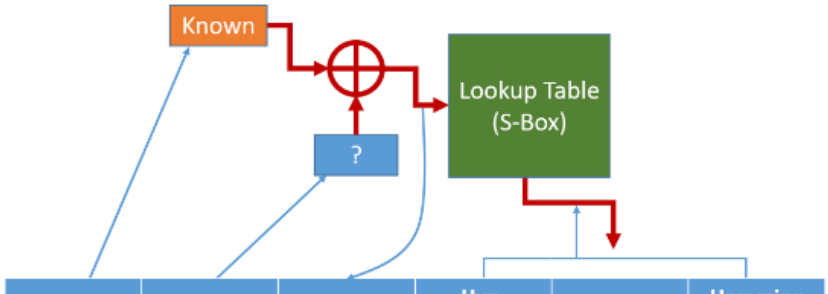
Correlation Power Analysis – CPA

- Power consumption to retain a logic level 1 consumes more energy than logic level 0.
- When a register holds some bit pattern, the *hamming weight* gets reflected in the energy consumption (and EM radiation too).
- The Correlation Power Analysis (CPA) exploits this phenomena to identify the unknown bit pattern of the secret key.



Correlation Power Analysis (cont.)





Input Data	Hyp. Key	XOR Output	Hyp. Output	Binary	Hamming Weight
0x68	0x00	0x68	0x45	01000101	3
0x4A	0x00	0x4A	0xD6	11010110	5
0xFF	0x00	0xFF	0x16	00010110	3
0xA5	0x00	0xA5			
0x21	0x00	0x21			

Power Measurement	0x00	0xA1	...	0xFF
	3	6		0
	5	5		0
	3	6		0
	3	5		8
	2	6		8
	6	6		7
	1	6		8
	8	6		5

Correlation Power Analysis (cont.)

Take a look at the simulated CPA attack in the Jupyter Notebook.

Electromagnetic Side-Channel Analysis

- Time-varying electrical currents are generating electromagnetic (EM) radiation.
- Our electronic equipment are a source of strong EM radiation.
- The EM radiation of computers (specifically, the processors) is shown to be correlating with the software running on them, i.e., the exact instructions and their execution pattern.
- EM Side-Channel Analysis (EM-SCA) is the exploitation of these radiation to eavesdrop on computers:
  - Software behaviour dection.
  - Malicious firmware modification detection.
  - Cryptographic key retrieval.

Electromagnetic Side-Channel Analysis (cont.)

A Few Demonstrations

- Radiation from the laptop screen/graphic card:
  `https://www.youtube.com/watch?v=YtolwTPDBwk`
- Remote servelliance of video displays:
  `http://www.youtube.com/watch?v=8OlkywZBJGU`
- Data exfiltration through EM covert channel on an Ethernet cable:
  `https://www.youtube.com/watch?v=ciM4M5h3q0w`
- A clock glitch attack on Arduino Uno:
  `https://www.youtube.com/watch?v=Me9Kf2Ga0vs`
- Intercepting smart card communication using HackRF One:
  `https://www.youtube.com/watch?v=HiUu3_3kQYY`

Old Incidents and Crimes

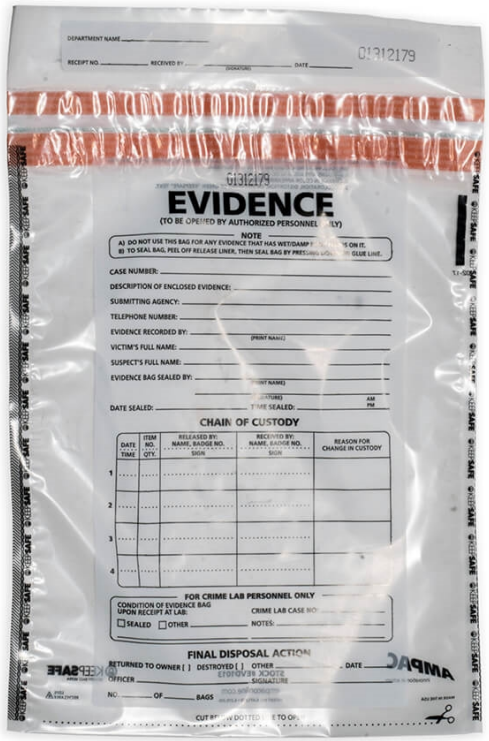
Fingerprints, bloodstains, hand-written notes, eye witness, etc.

Modern Incidents and Crimes

DNA analysis, biometrics, CCTV footage, facial recognition, etc.

Digital Forensics

Digital Evidence



Any data stored or transmitted using a computer that support or refute a theory of how an offense occurred or that address critical elements of the offense.

Digital Evidence Sources

- Open computer systems (laptops, desktops, smartphones, and their storage devices).
- Embedded computer systems (smart wearables, printers, access control systems, etc.).
- Communication systems (wired and wireless network data).

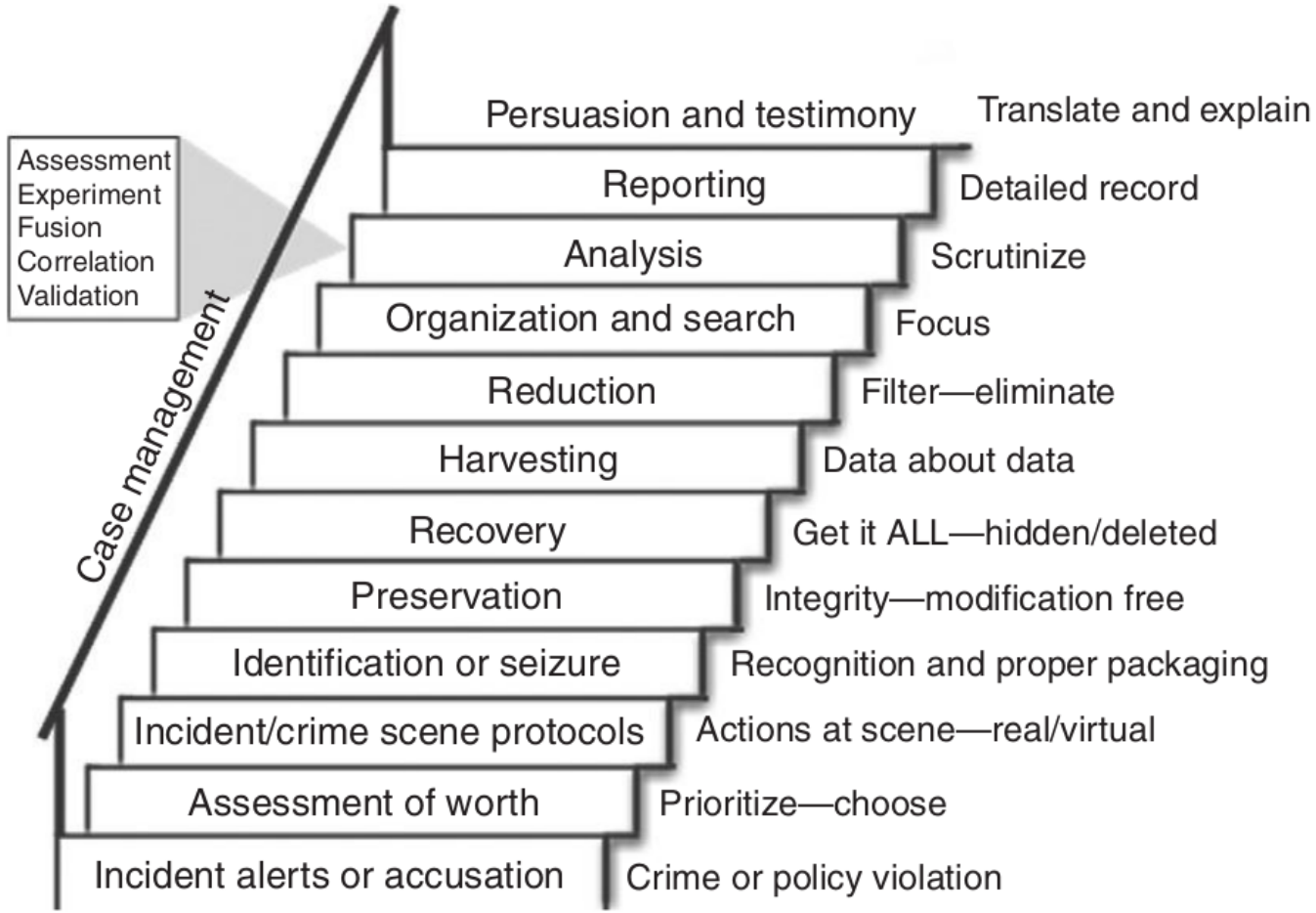
Chain of Custody

cmdLabs
Continuity of Possession Form

Case Number:	2010-05-27-00X	Client/Case Name:	Digifinger Intrusion
Evidence Type:	hard drive	Evidence Number:	0023
Details:	Mac storage <network share>		

Date of Transfer	Transferred From	Transferred To	Location of Transfer	Action Taken by Recipient
5/27/10	signature: Sam Spade print name: Sam Spade	signature: Phillipe Marlowe print name: Philip Marlowe	Digifinger HQ Linthicum MD	Collected evidence for examination
	signature print name	signature print name		

Digital Forensics Process



‘Science’ in Digital Forensics

- Forensics is a scientific discipline — so as digital forensics.
- A scietific discipline should follow the **scientific method**.
- Repeatability and Reproducibility
- Elimination of hypotheses.

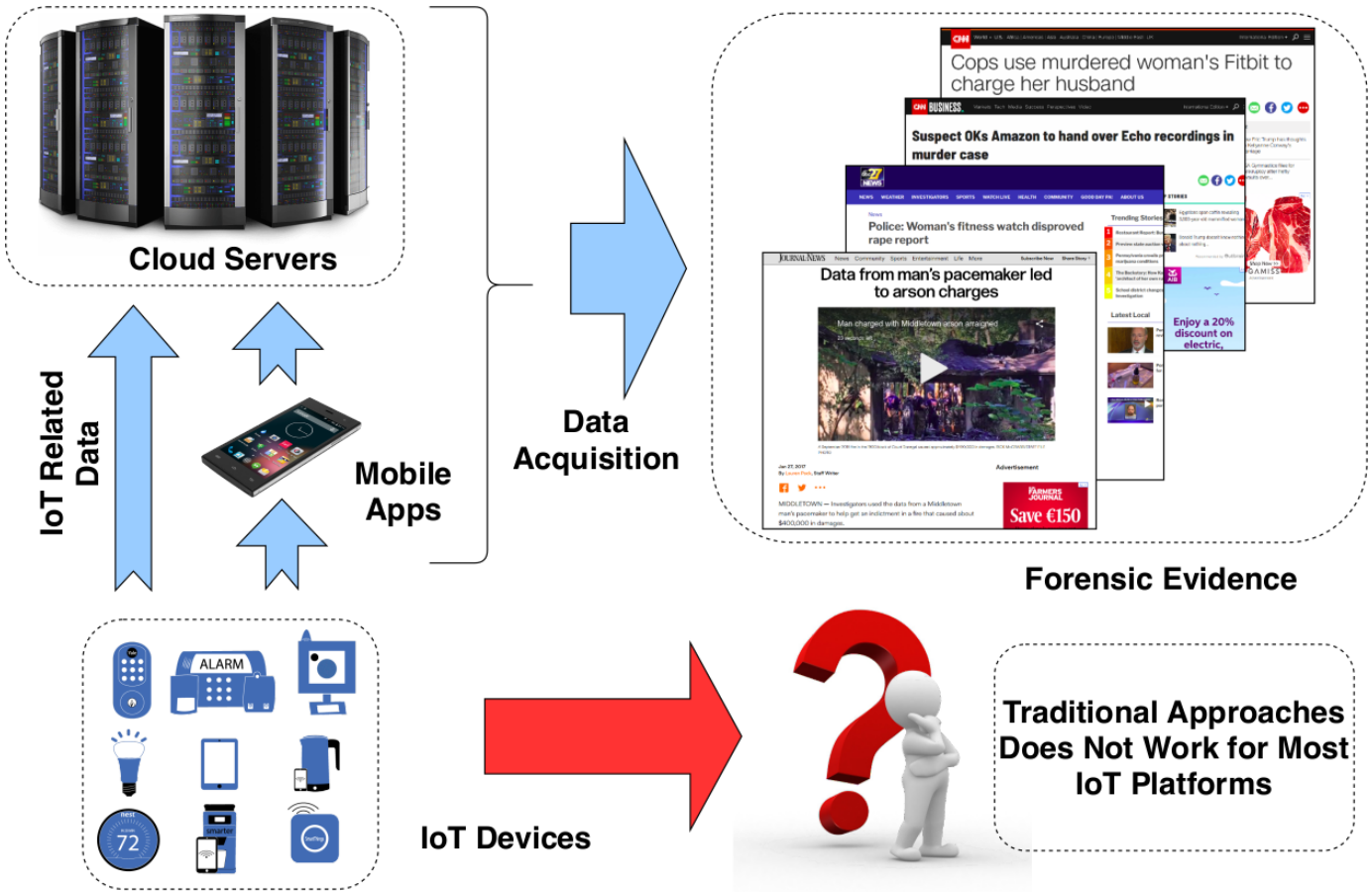
ACPO Guidelines

- If a digital crime scene is not handled properly, the entire operations afterwards will be futile.
- Various standards exists on how to handle a digital crime scene.
- The *Association for Chief Police Officers* (ACPO), has published a guide of good practices, which is highly recognised.
- These are not strict rules, but helpful guidelines — under certain conditions we may have to deviate from them.

ACPO Guidelines

- **Principle 1:** No action taken by law enforcement agencies or their agents should change data held on a computer or storage media which may subsequently be relied upon in court.
- **Principle 2:** In circumstances where a person finds it necessary to access original data held on a computer or on storage media, that person must be competent to do so and be able to give evidence explaining the relevance and the implications of their actions.
- **Principle 3:** An audit trail or other record of all processes applied to computer-based electronic evidence should be created and preserved. An independent third party should be able to examine those processes and achieve the same result.
- **Principle 4:** The person in charge of the investigation (the case officer) has overall responsibility for ensuring that the law and these principles are adhered to.

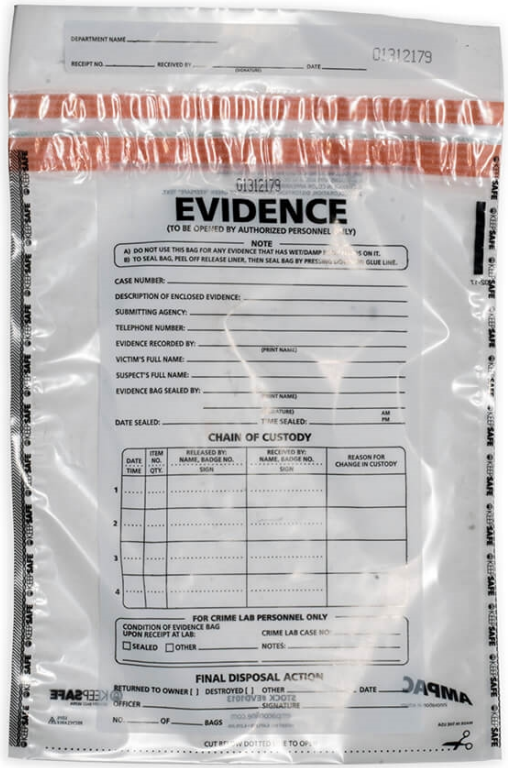
Challenge of IoT & Smart Device Forensics



Challenge of IoT & Smart Device Forensics (cont.)

- New devices are emerging in the market too frequently.
- The internal components, storage, and interfaces of the devices varies drastically.
- Analysing devices at too close to the hardware level — such as chip-off forensics — is susceptibe to irreversible mistakes that can destroy a device entirely.
- It would be ideal if we can inspect a device from a safe distance.

Evidence vs Insights

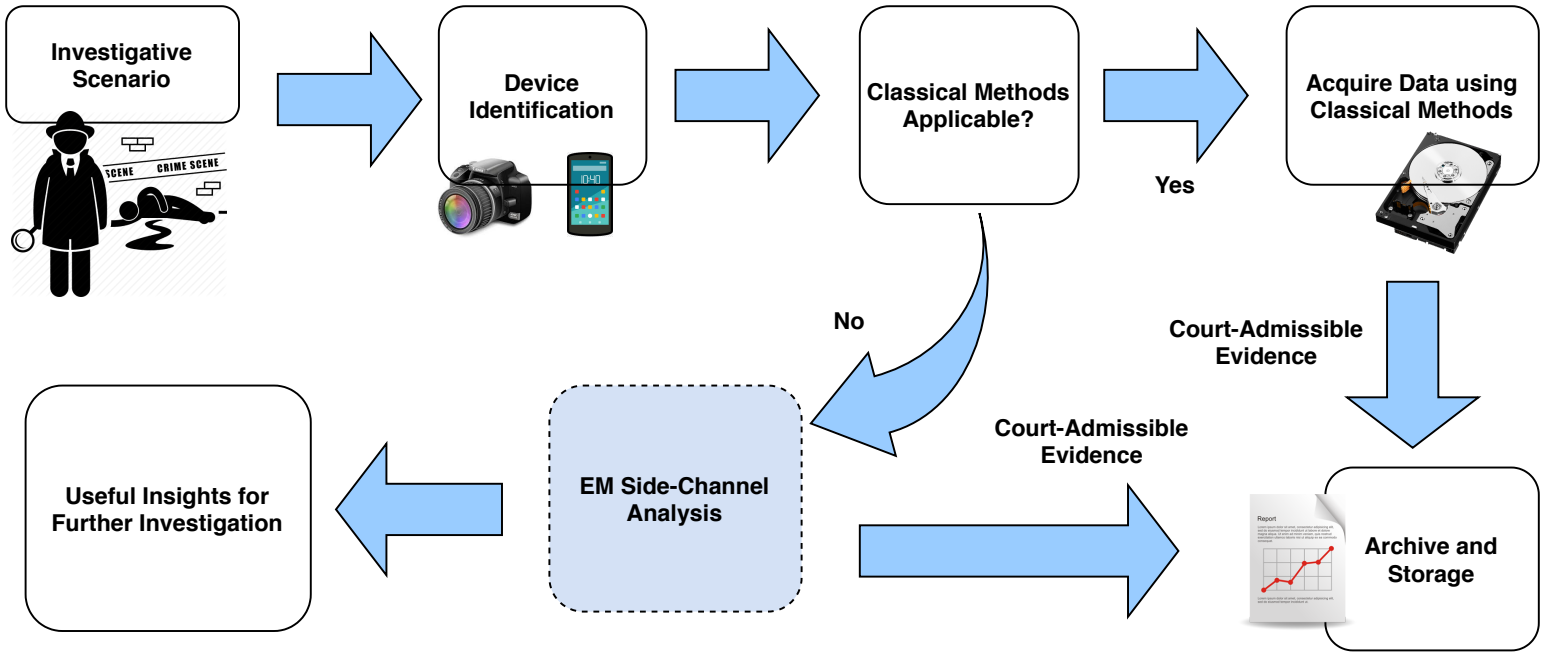


- Digital **evidence** are information that may be presented to a court of law.
- They need to be concrete enough to be relied upon at the courts.
- The field of digital forensics is aimed at providing this reliability as much as possibile.
- In some situations, where evidence are not available, some **insights** can a lifeline for an investigator.
- Insights are — most likely — not reproducible, but they can provide useful hints and directions to go and locate reliable evidence through other means.

Forensic Insights from IoT and Smart Devices

- Is this device running the official firmware from the manufacturer?
- Has a malware been injected to the memory of this device?
- Is this device doing something it is not supposed to be doing right now; such as wiping the storage or encrypting it, instead of shutting down?

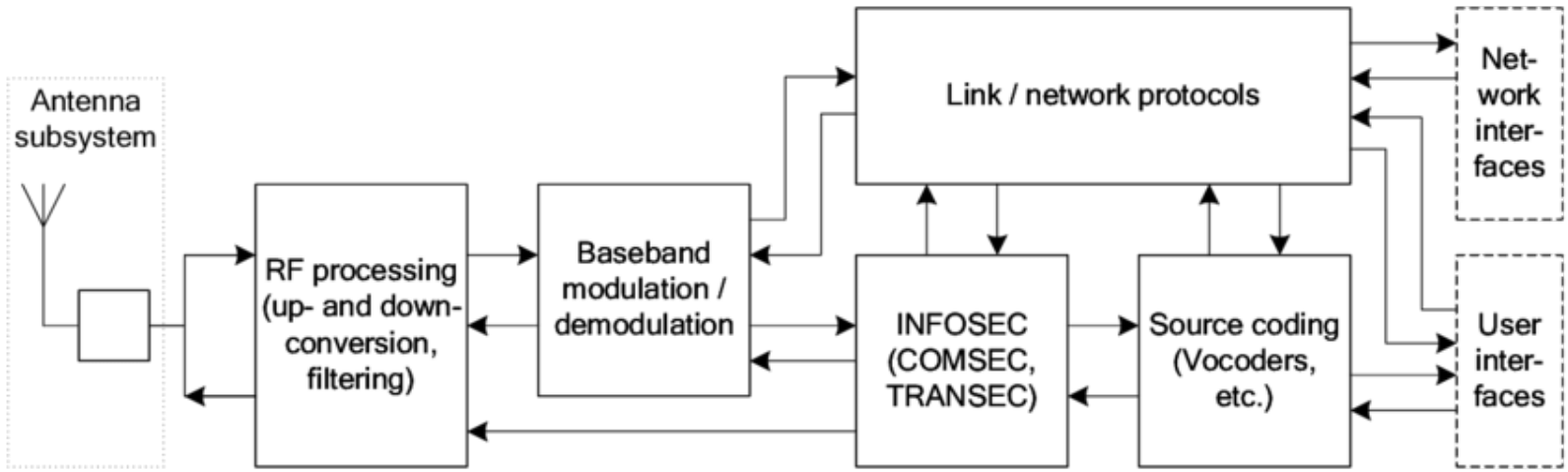
Forensic Insights through EM-SCA



Part 2

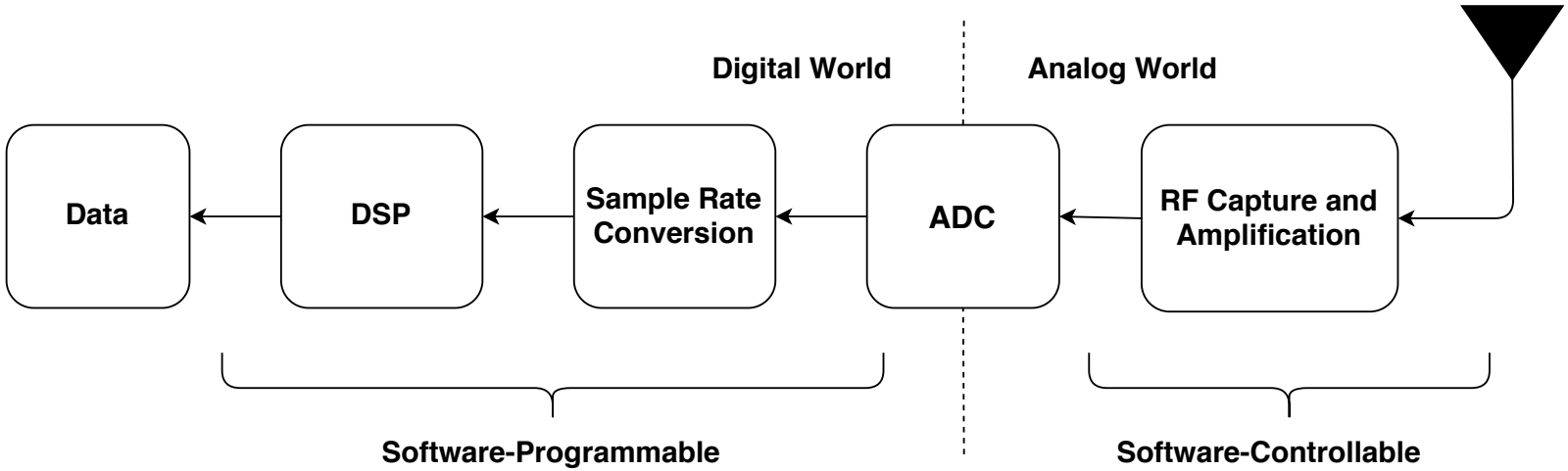
Radio/Wireless Devices

- Built using hardware components that deal with analog signals.
- Digitisation (if any) occurs in a very late stage.
- Application-specific hardware configurations.



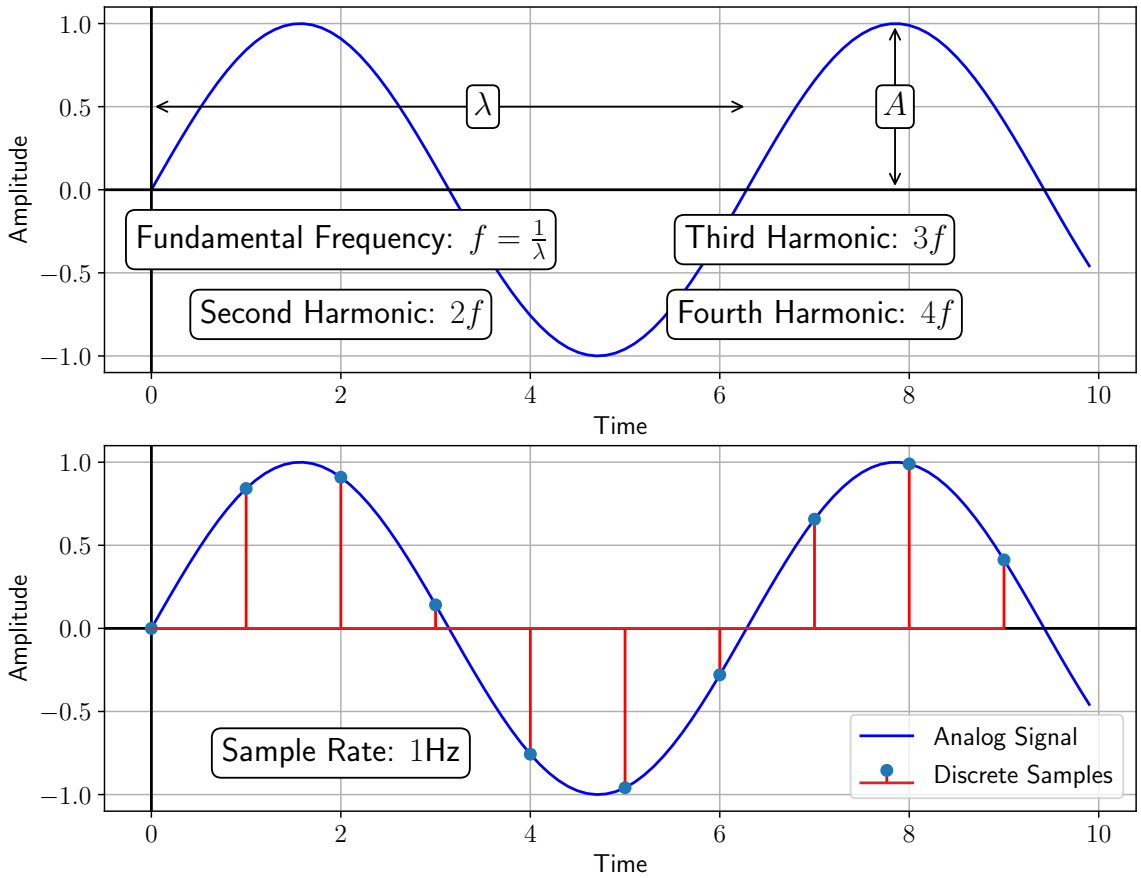
Software Defined Radios

- Moving most of radio functions from analog domain into the digital domain.
- Requires a generic hardware radio interface including a fast analog-to-digital converter (ADC).
- Involves a local oscillator (LO) with a mixer — not illustrated in the figure.
- Need sophisticated and optimised software implementations for digital signal processing (DSP).



Software Defined Radios (cont.)

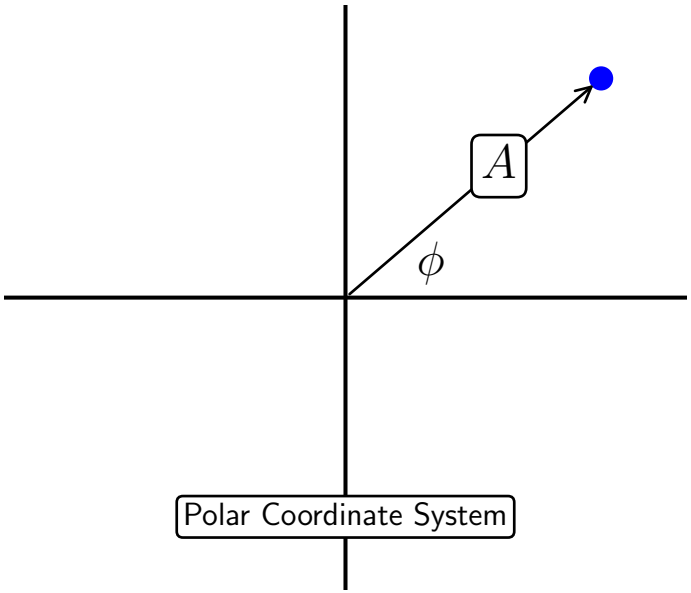
Converting analog signals to digital signals require **sampling** and **quantisation**.

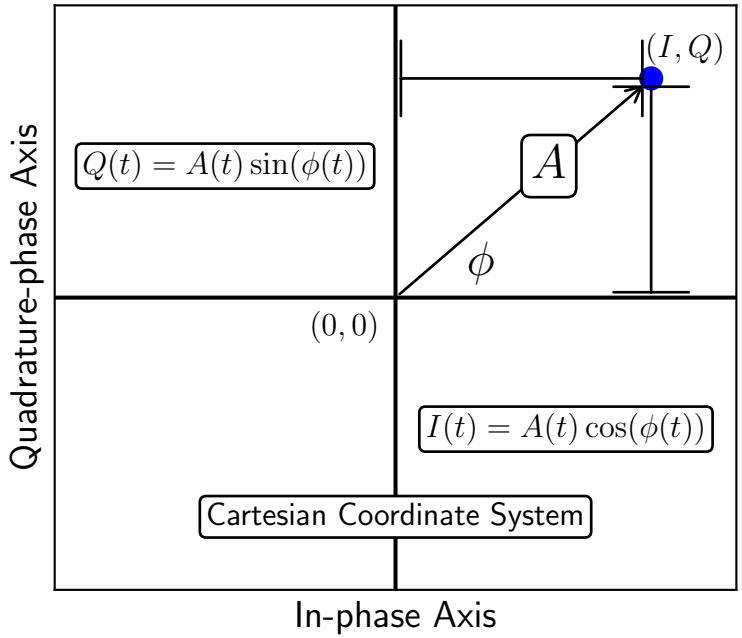


Real-valued sampling faces the Nyquist limit: difficult to capture high frequencies.

Software Defined Radios (cont.)

- SDRs are using complex **In-phase/Quadrature (I/Q)** sampling.
- Each sample taken at a given time instance consists of two values; hence, each sample is a complex number.
- A EM signal captured by an SDR is basically an array of complex numbers.
- Sampling rate is equal to the bandwidth of the captured data.





SDR Hardware

RTL-SDR

- A digital TV tuner repurposed as an SDR.
- *Realtek RTL2832U* and similar chips were discovered to be hackable.
- The cheapest possible SDR you may find.
- A wide variety of manufacturers; hence different variations of capabilities.
- Sample rate is about 3.2 MHz
- Tunable frequency range: 22 MHz – 1 GHz.
- Receive only; no transmission (simplex).
- Read more: `https://www.rtl-sdr.com`

SDR Hardware (cont.)

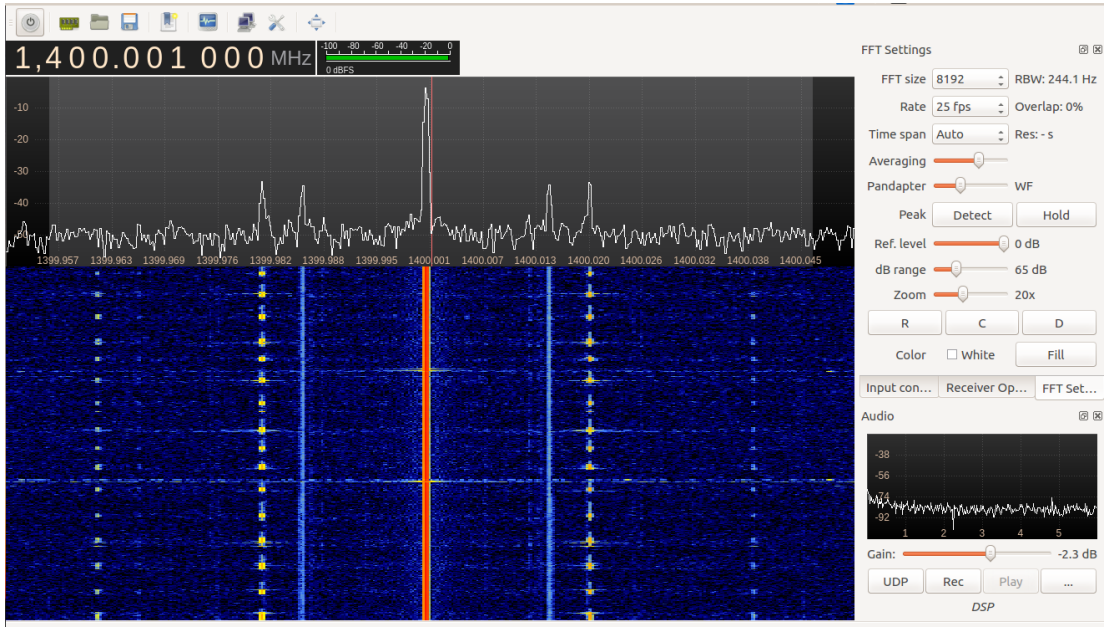
HackRF One

- A purpose-built SDR device with a mid-range price tag.
- Sample rate: upto 20 MHz.
- Tunable frequency range: 1 MHz – 6 GHz.
- A wide range of antennas can be connect through the SMA connector.
- Half-duplex: either transmit or receive at a given time.
- Possible to time-synchronise with another device through clock input or output.
- Read more: `https://greatscottgadgets.com/hackrf/one`

SDR Software

GQRX

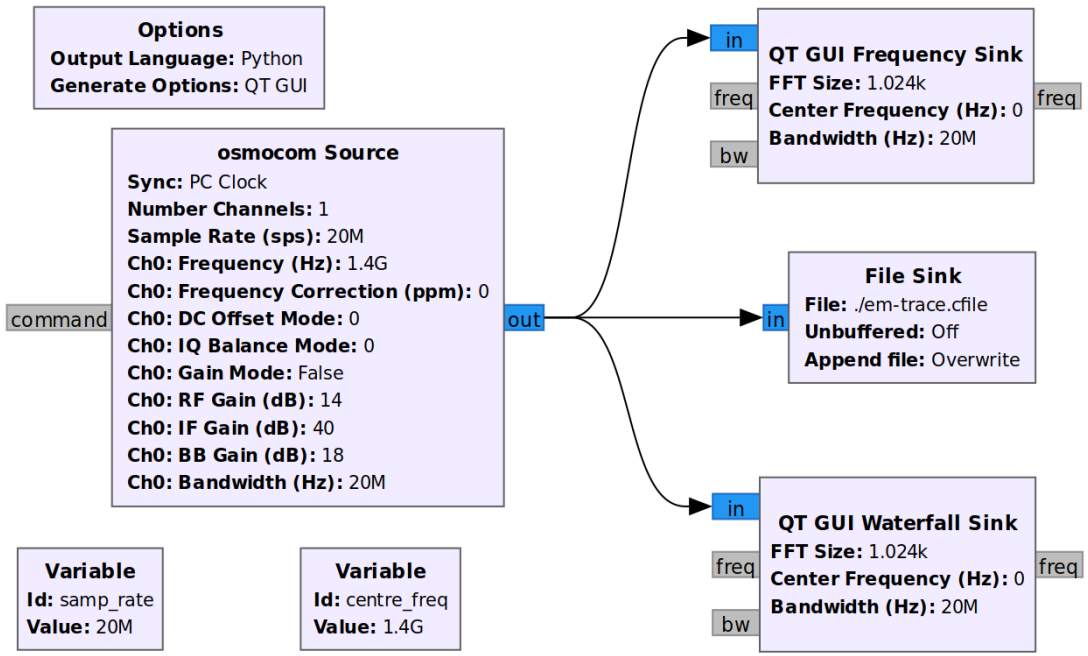
- Easily view signals at different frequencies.
- Facilitates live processing and saving observed signals.
- Uses *GNURadio* library underneath.
- Launch it using `gqrx` command from the terminal.



SDR Software (cont.)

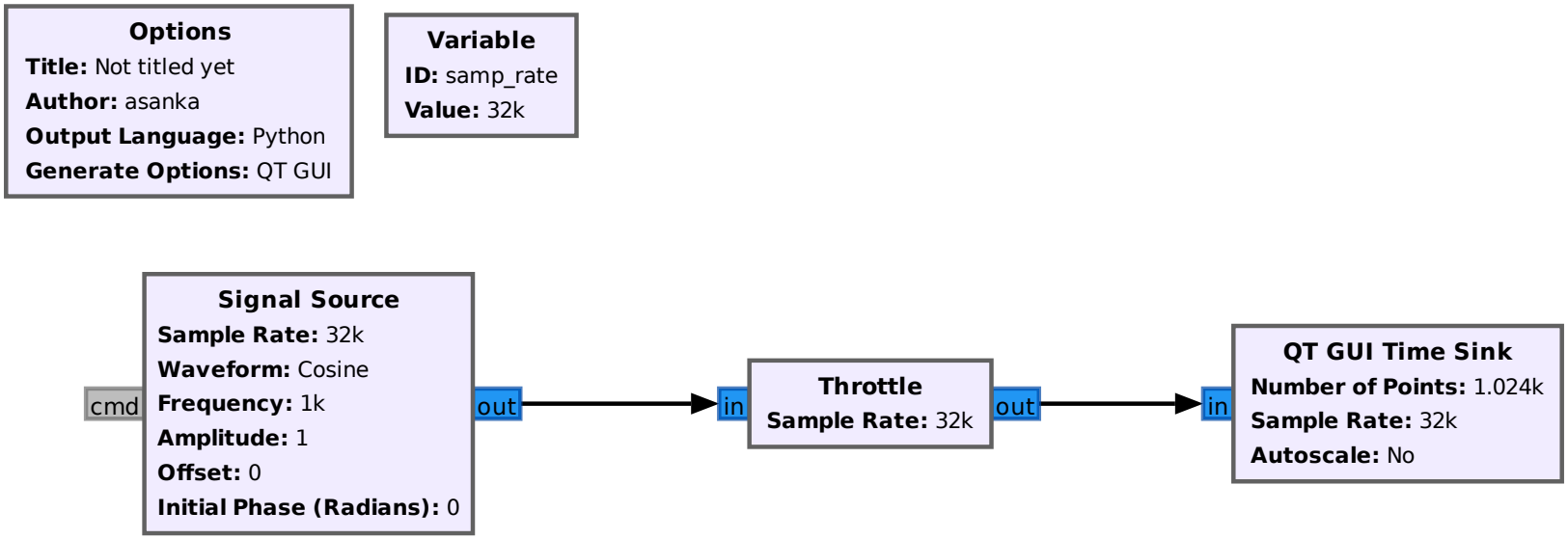
GNURadio Companion (GRC)

- Various signal processing blocks.
- Custom build any application by creating flow graph using blocks.
- Generates executable Python scripts for flow graphs, using *GNURadio* library.
- Launch it using `gnuradio-companion` command from the terminal.



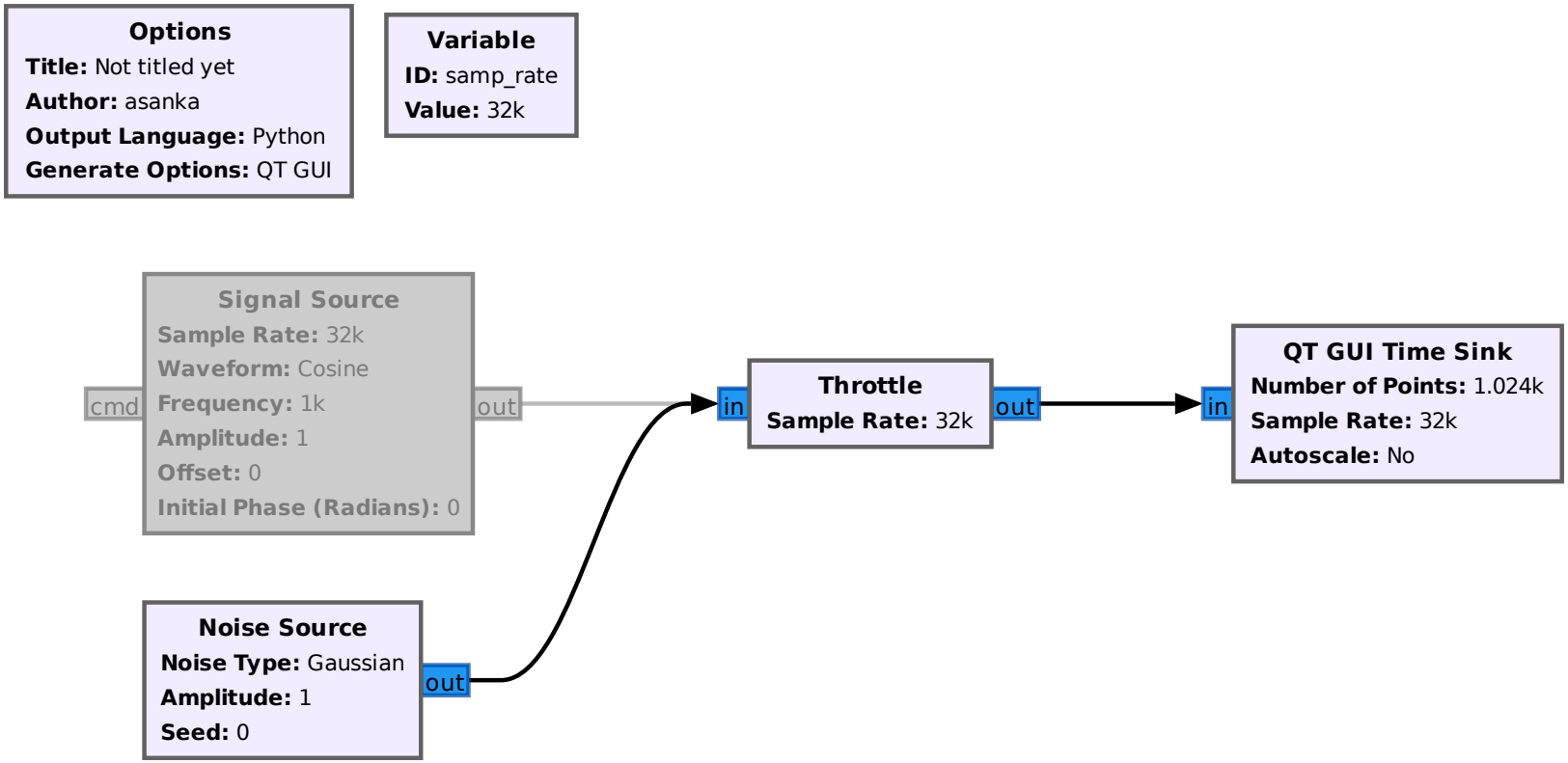
Exercise 1

Generating and visualising sine waves



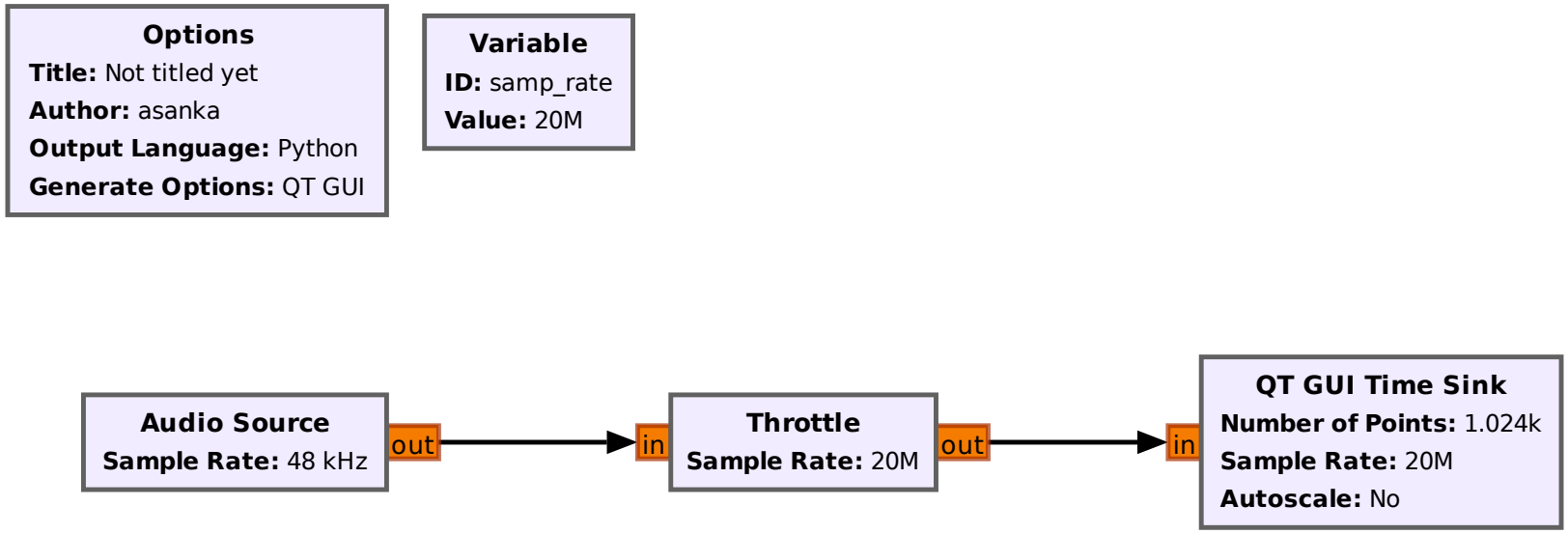
Exercise 2

Generating and visualising noise



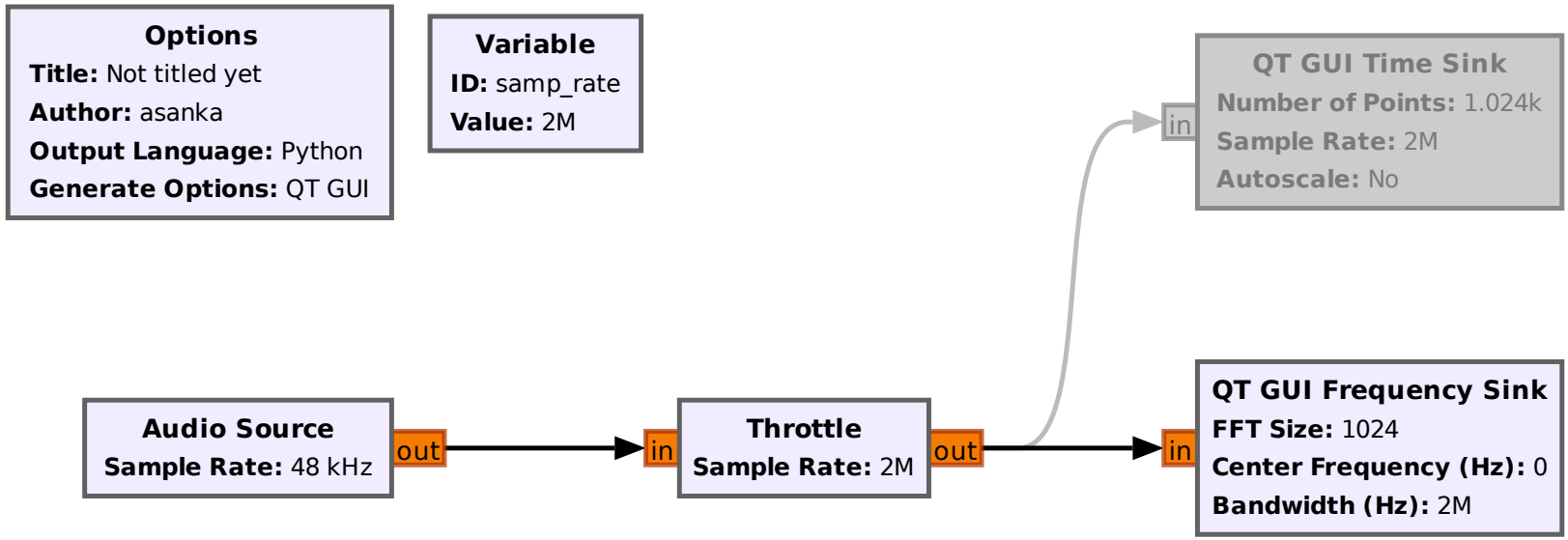
Exercise 3

Capturing audio



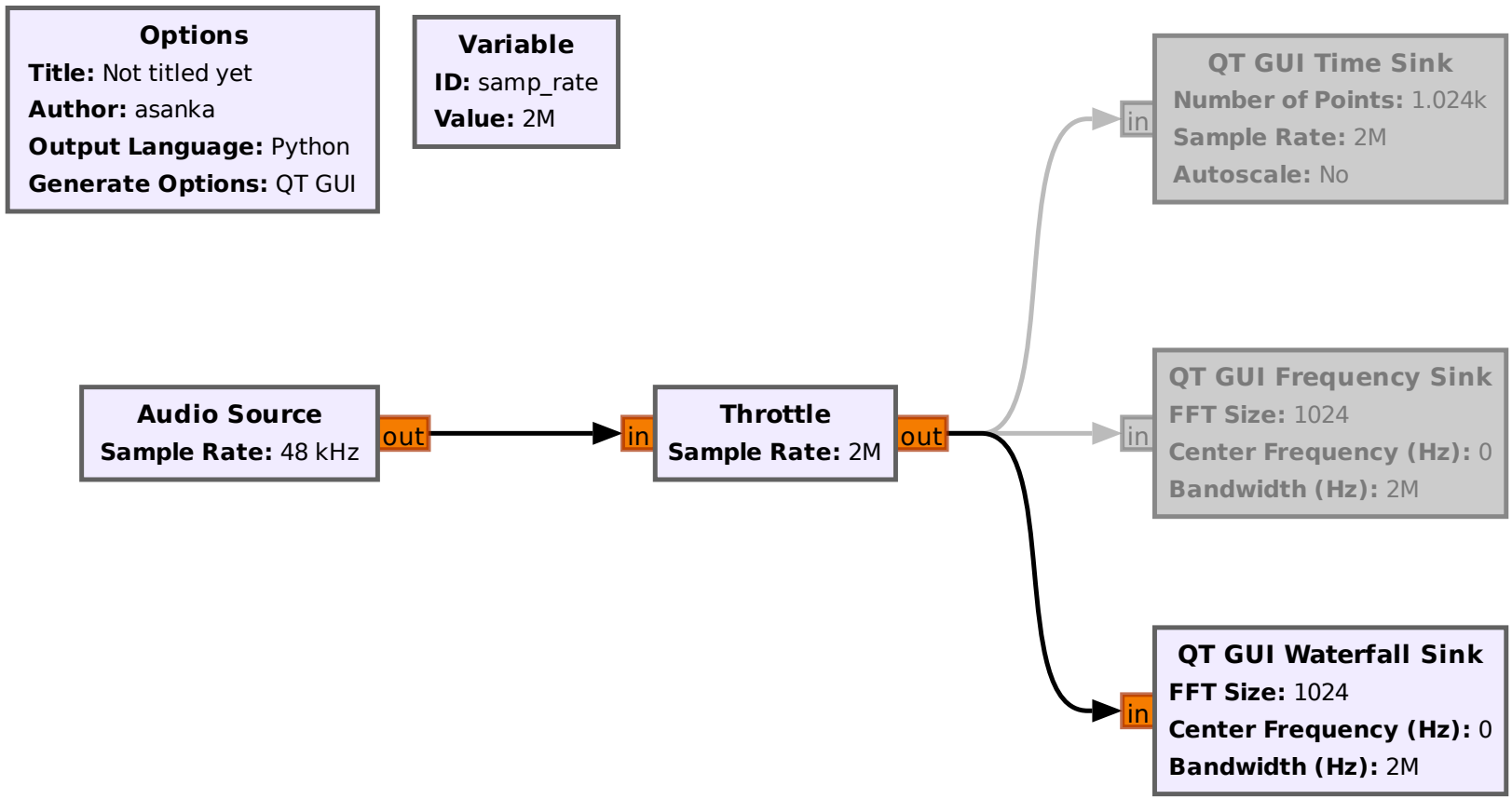
Exercise 4

Visualising frequency domain



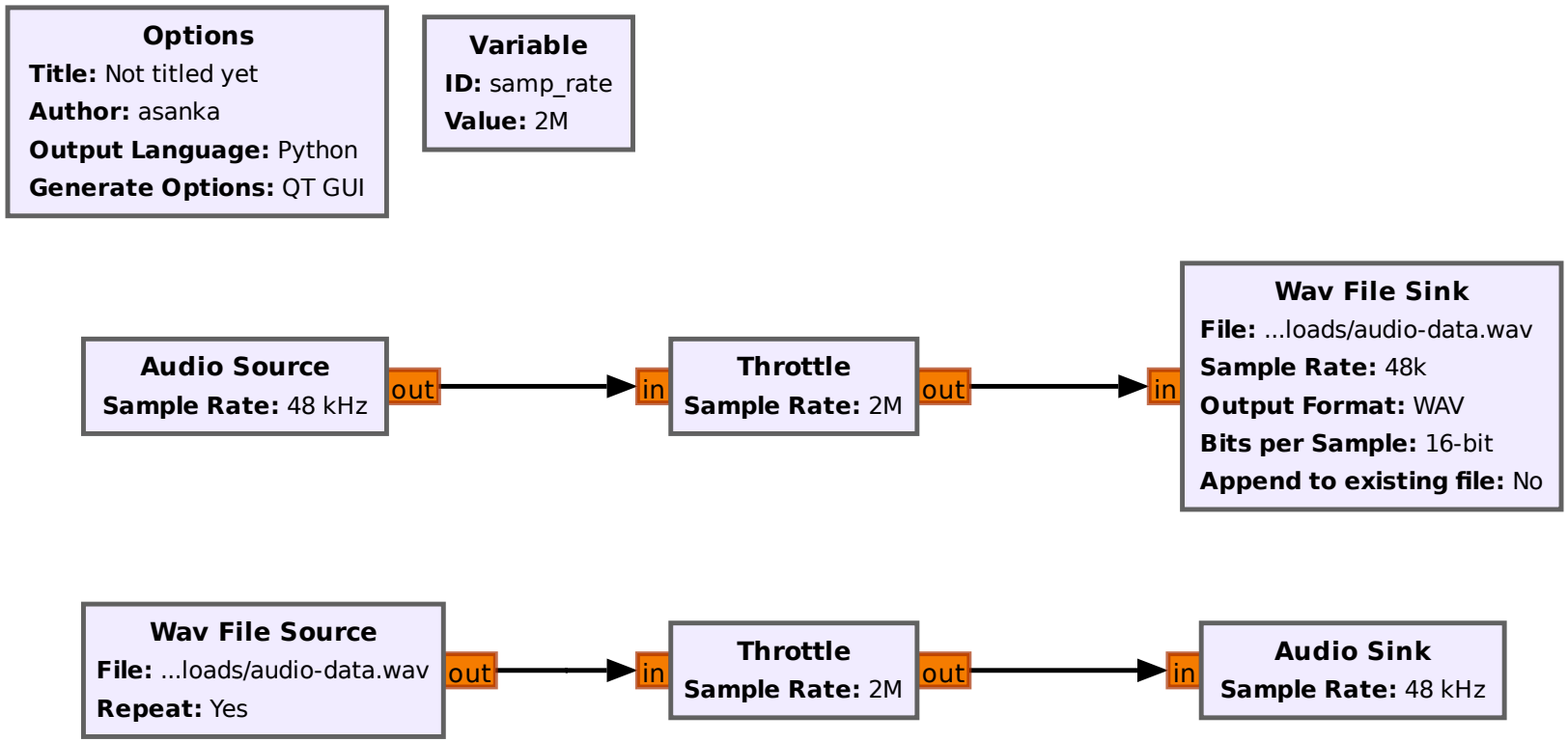
Exercise 5

Spectrogarm/waterfall plot



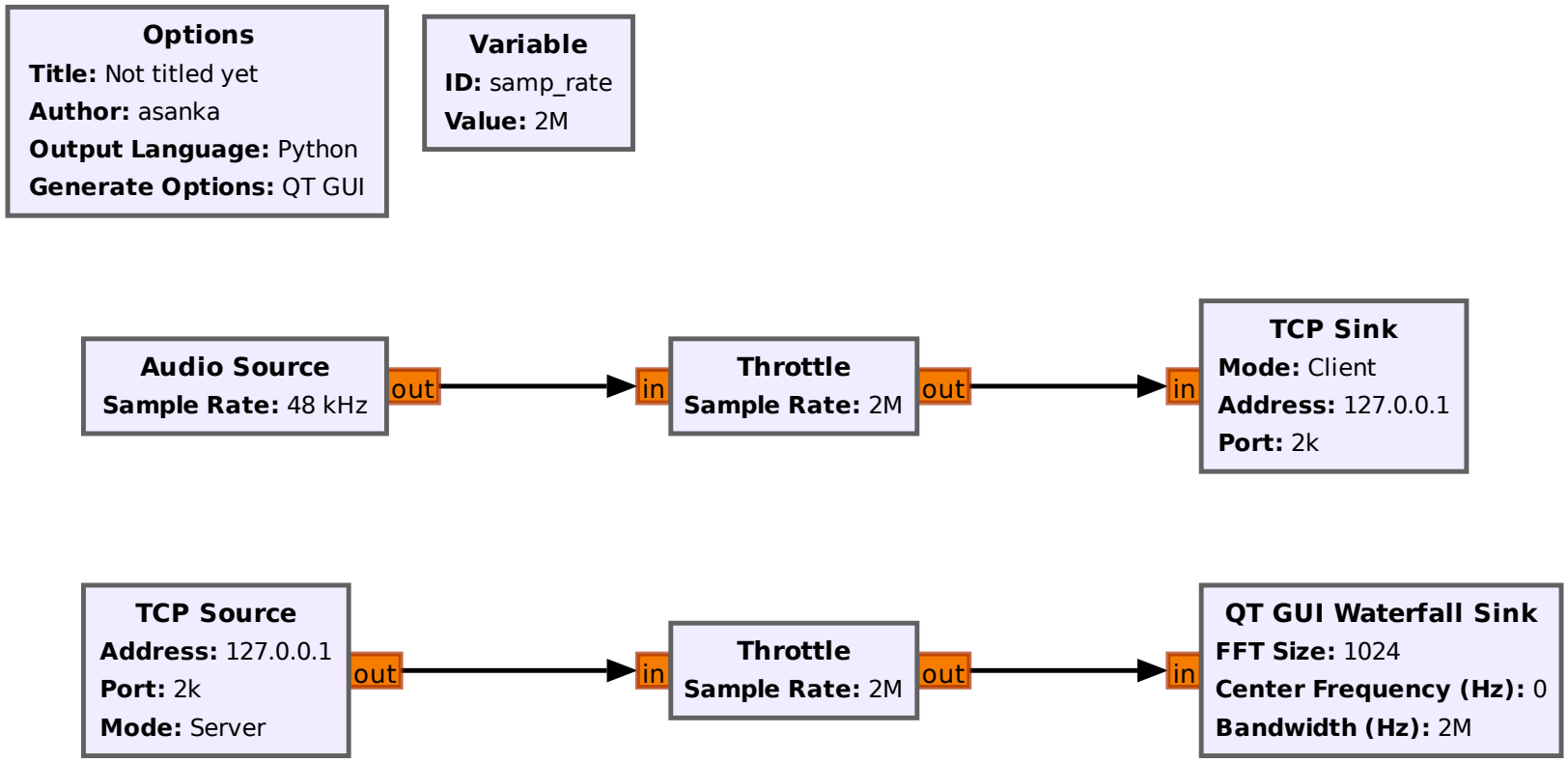
Exercise 6

Saving data to (wav) files



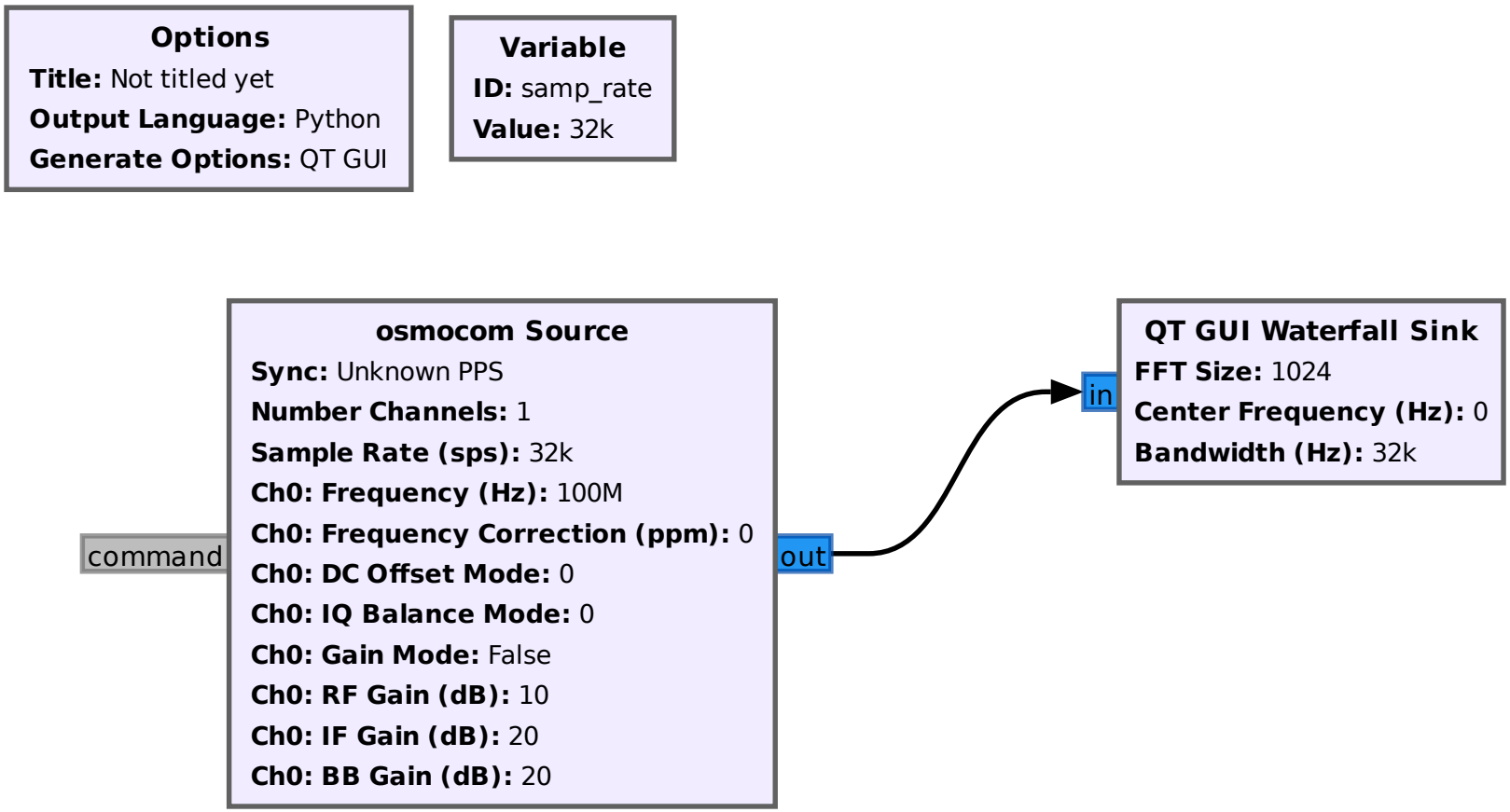
Exercise 7

TCP connections between flowgraphs



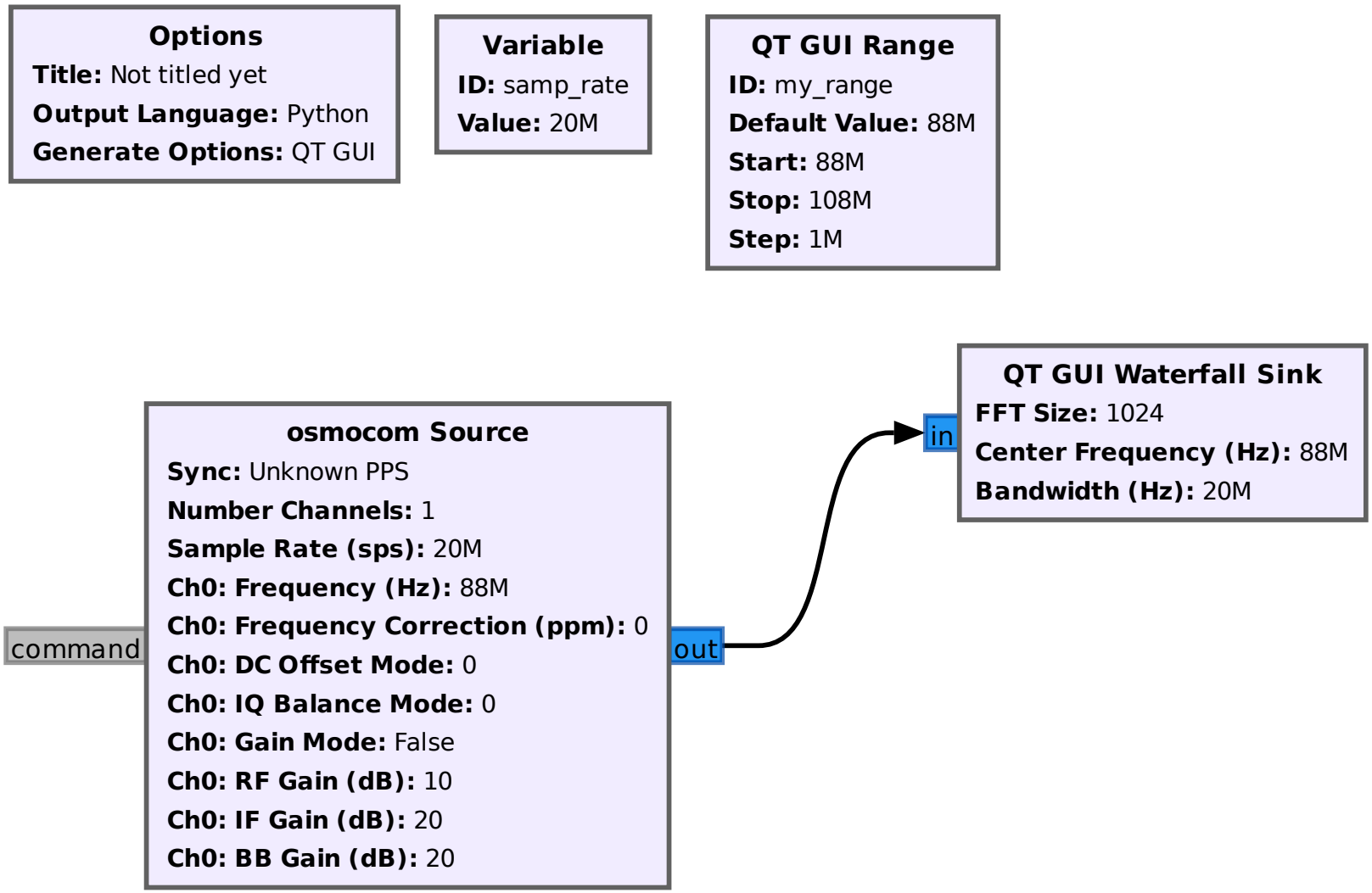
Exercise 8

Capturing data from HackRF



Exercise 9

Additional GUI components



Exercise 10

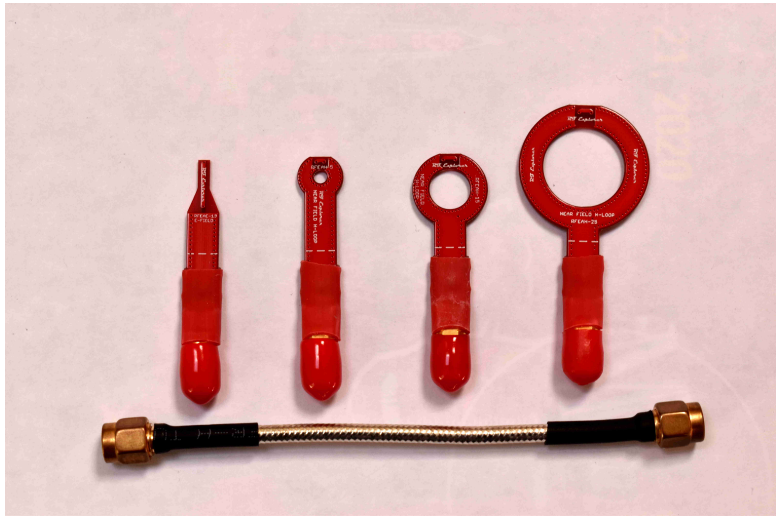
Let's link the HackRF connected to my computer with a GRC flow graph on your computer through TCP.

This setup will be useful for us to complete the exercises in the next segment of the workshop, i.e., Part 4.

Part 3

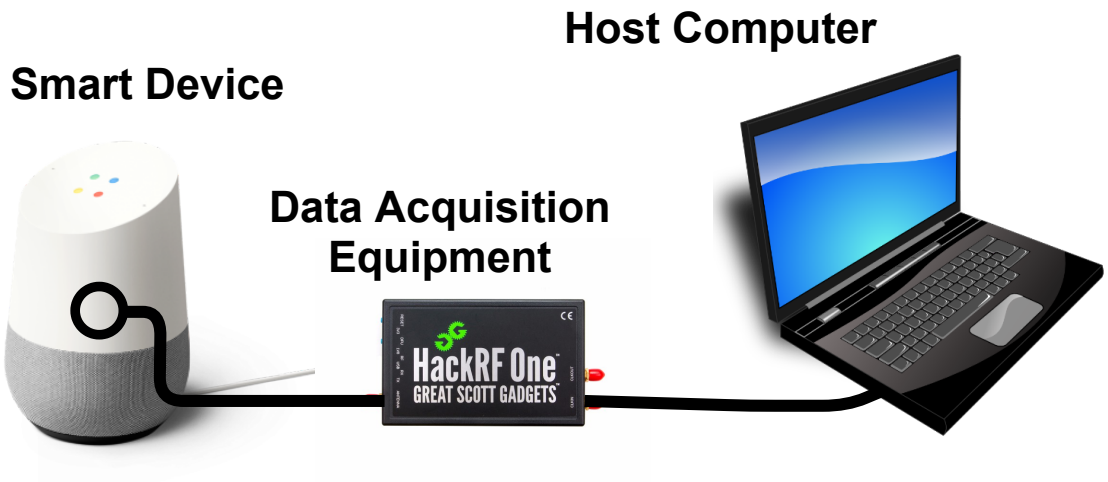
Capturing EM Side-Channel Radiation

- Our key focus is EM radiation from the processor/microcontroller/SoC.
- Strongest signals are in the clock frequency or its harmonics.
- Signal acquisition should be performed as closer to the target chip as possible.
- Magnetic H-loop antennas are more suitable for the job.



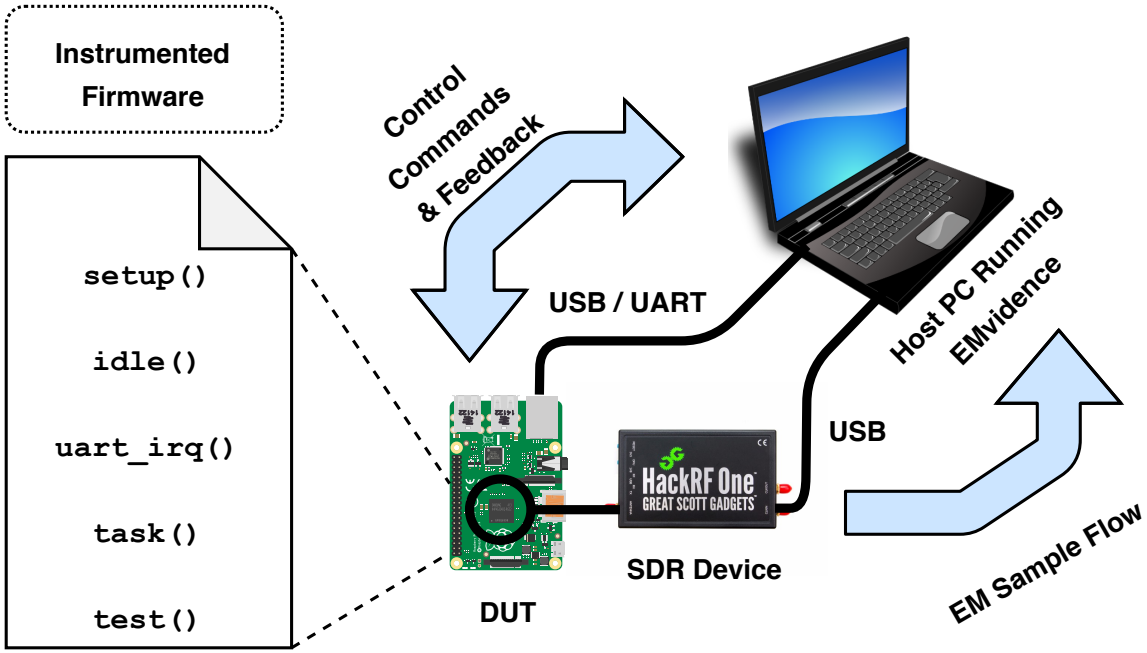
Capturing EM Side-Channel Radiation (cont.)

Passive acquisition:



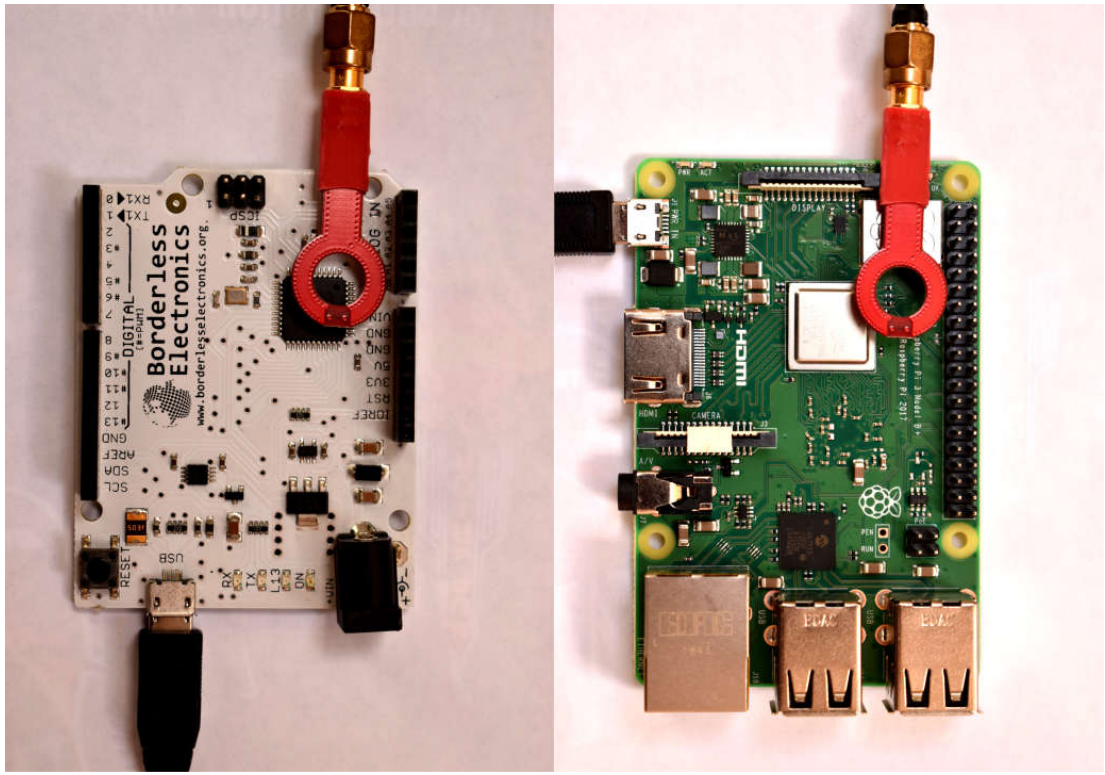
Capturing EM Side-Channel Radiation (cont.)

Instrumented acquisition:



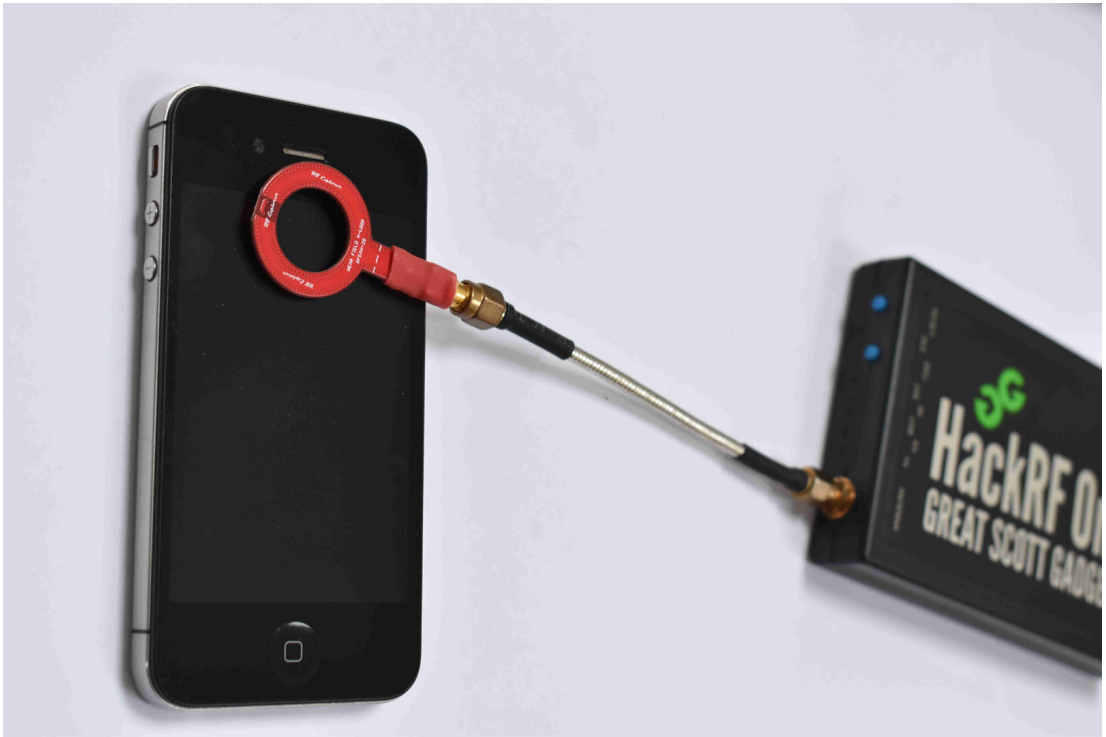
Capturing EM Side-Channel Radiation (cont.)

Arduino and Raspberry Pi



Capturing EM Side-Channel Radiation (cont.)

Smartphone



Processing EM Side-Channel Data

- Once the EM radiation from a target device — device-under-test (DUT) — is captured, we can process it.
- GNURadio Companion flowgraphs have blocks for signal processing, but it is not sufficient.
- The ideal way is to process data using a scientific computing language, such as Python.
- Once the EM data is loaded into a *numpy* array, the possibilities are endless.

Processing EM Side-Channel Data (cont.)

Loading I/Q data into Python:

```
import numpy as np
import matplotlib.pyplot as plt

def getData(cfileName):
    data = np.fromfile(cfileName, dtype="float32")
    data = data[0::2] + 1j*data[1::2]
    return data

data = getData("/home/asanka/Desktop/my-data.cfile")
```

Processing EM Side-Channel Data (cont.)

Plotting I/Q data in Python:

```
fig = plt.figure()
plt.psd(data, NFFT=2048, Fs=20e6)
plt.show()
```

```
fig = plt.figure()
pxx, freq, t, cax = plt.specgram(data, NFFT=1024, Fs=20e6,
        Fc=88e6, mode='magnitude')
fig.colorbar(cax).set_label('Intensity [dB]')
plt.xlabel("Time (s)")
plt.ylabel("Frequency (Hz)")
plt.show()
```

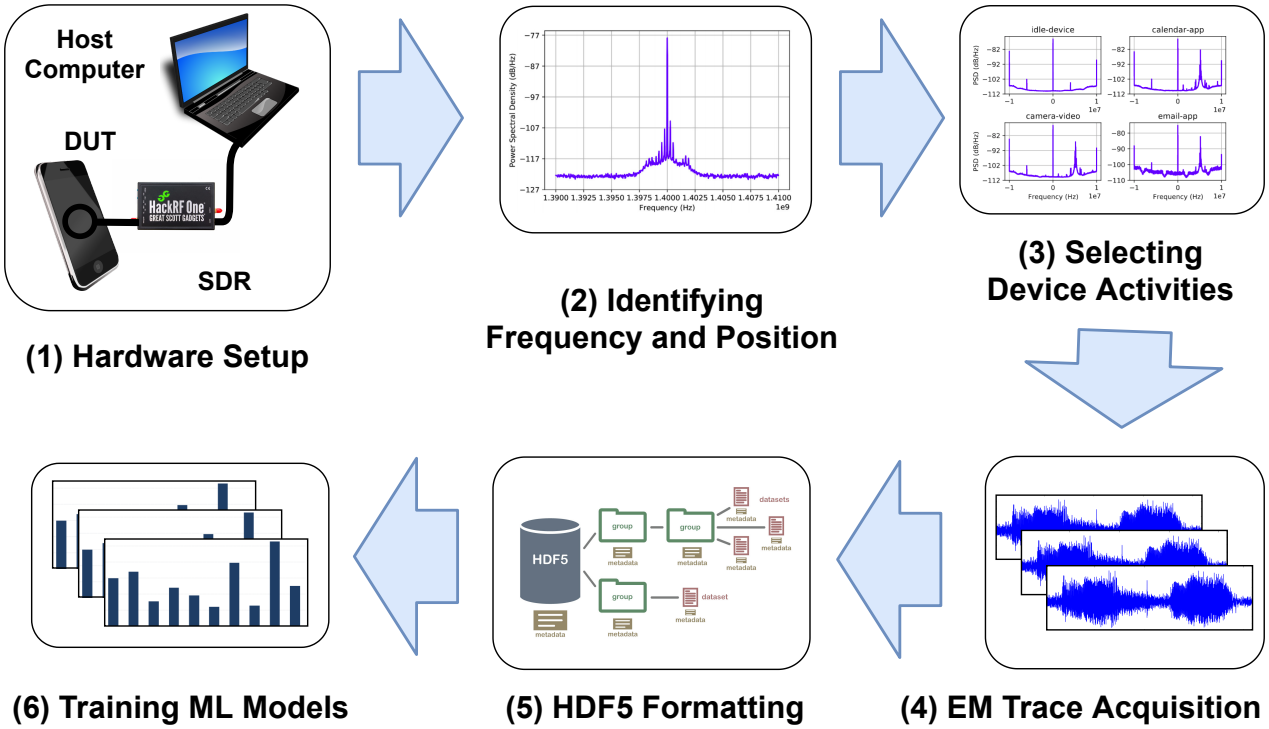
Activity

- Let's use EM side-channel emission of a DUT to determine whether the device is powered up or not.
- Our model DUT is an Arduino Uno.
- You can write a program to the Arduino Uno to perform some task.
- Collect EM trace data for the two cases: the Arduino is powered off and running your program.
- Visualise the two EM traces. Can you observe a difference?
- Explore whether you can automate the detection process programmatically. You can use various approaches, such as digital signal processing, statists, machine learning, etc.

Part 4

Analysing EM Dataset

The pipeline from capturing EM data to analysis...



Analysing EM Dataset

Huge Size of the Data Files

- In GNU Radio library, two 32 bit (4 byte) floating point values are used to represent a complex I/Q sample.
- Therefore, each EM data sample is a 8 bytes long complex value.
- Consider if we sampled data at the maximum sample rate of HackRF One device (i.e., 20 MHz) using GNURadio library to save data.
- Size of data per second $= 8$ bytes $\times 20 \times 10^6 = 160$ MB
- Size of data for 10 seconds $= 160$ MB $\times 10 = 1.6$ GB

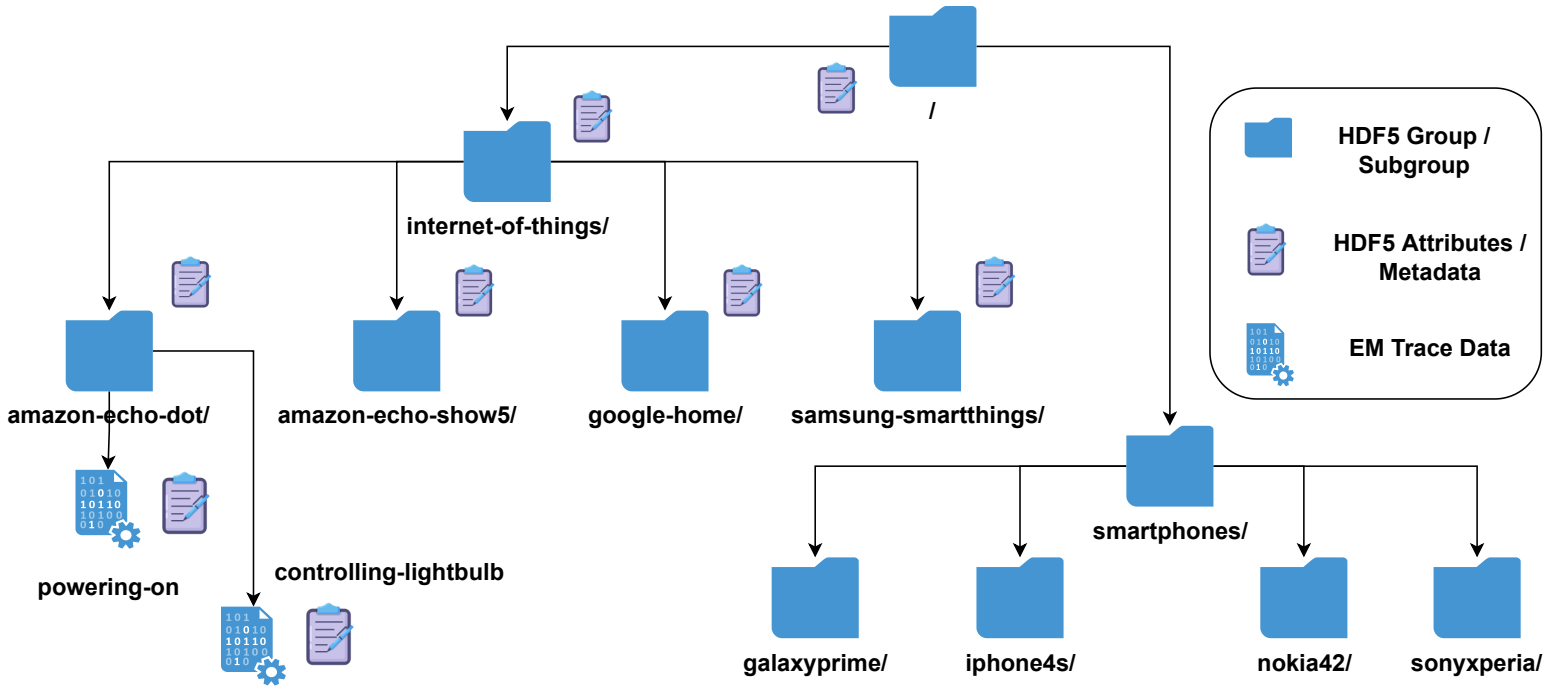
Analysing EM Dataset (cont.)

The specifications of devices in the dataset:

Smart Device	System-on-Chip	Architecture	CPU Frequency	Software Activities
Amazon Echo Show 5	MediaTek MT 8163	ARMv 8-A	1.5 GHz (4 cores)	(1) asking a definition, (2) asking for time, (3) asking to play radio, (4) controlling light-bulb, (5) device idle, (6) device resetting, (7) just wake up word, (8) powering off, (9) powering on.
Amazon Echo Dot (3rd Gen)	Mediatek MT 8516	ARMv 8-A	1.3 GHz (4 cores)	(1) asking a definition, (2) asking for time, (3) asking to play radio, (4) controlling light-bulb, (5) device idle, (6) device muted (7) device resetting, (8) just wakeup word, (9) powering on.
Google Home	Marvell 88DE3006 Armada 1500 Mini Plus	ARMv 7	1.2 GHz (2 cores)	(1) asking a definition, (2) asking for time, (3) asking to play radio, (4) controlling light bulb, (5) device idle, (6) device muted (7) device resetting, (8) just wake-up word, (9) powering on.
Samsung SmartThings Hub (v2)	MCIMX6L2DVN10AB	ARMv 7-A	1 GHz (1 core)	(1) controlling smart outlet, (2) device idle, (3) device powered off, (4) device powering on, (5) opening the app, (6) viewing arrival sensor, (7) viewing door sensor, (8) view motion sensor.
Apple iPhone 4S	Apple A5	ARMv 7-A	1 GHz (2 cores)	(1) calendar app, (2) camera photo, (3) camera video, (4) email app, (5) gallery app, (6) home screen, (7) device idle, (8) phone app, (9) SMS app, (10) web browser app.
Sony Xperia T	Qualcomm Snapdragon MSM8260A	ARM v7-A	1.5 GHz (2 cores)	(1) calendar app, (2) camera photo, (3) camera video, (4) email app, (5) gallery app, (6) home screen, (7) device idle, (8) phone app, (9) SMS app, (10) web browser app.
Samsung Galaxy Grand Prime	Qualcomm Snapdragon MSM8916	ARMv 8-A	1.2 GHz (4 cores)	(1) audio recording, (2) camera photo, (3) camera video, (4) email app, (5) gallery app, (6) home screen, (7) device idle, (8) phone app, (9) SMS app, (10) web browser app.
Nokia 4.2	Qualcomm Snapdragon SDM439	ARMv 8-A	1.95 GHz (4 cores), 1.45 GHz (4 cores)	(1) calendar app, (2) camera photo, (3) camera video, (4) email app, (5) gallery app, (6) home screen, (7) device idle, (8) phone app, (9) SMS app, (10) web browser app.

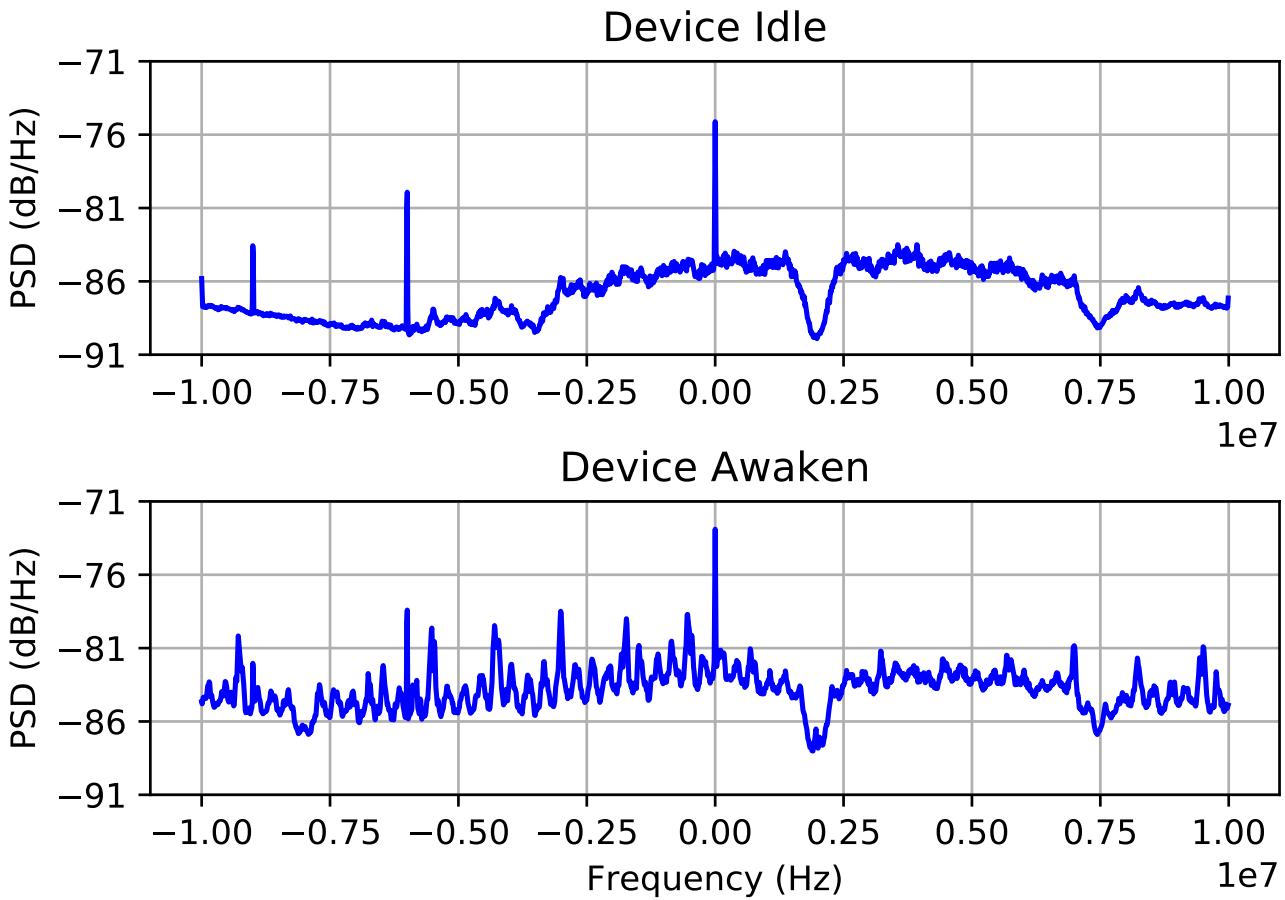
Analysing EM Dataset (cont.)

Structure of the dataset in HDF5 file format (em-dataset.h5):



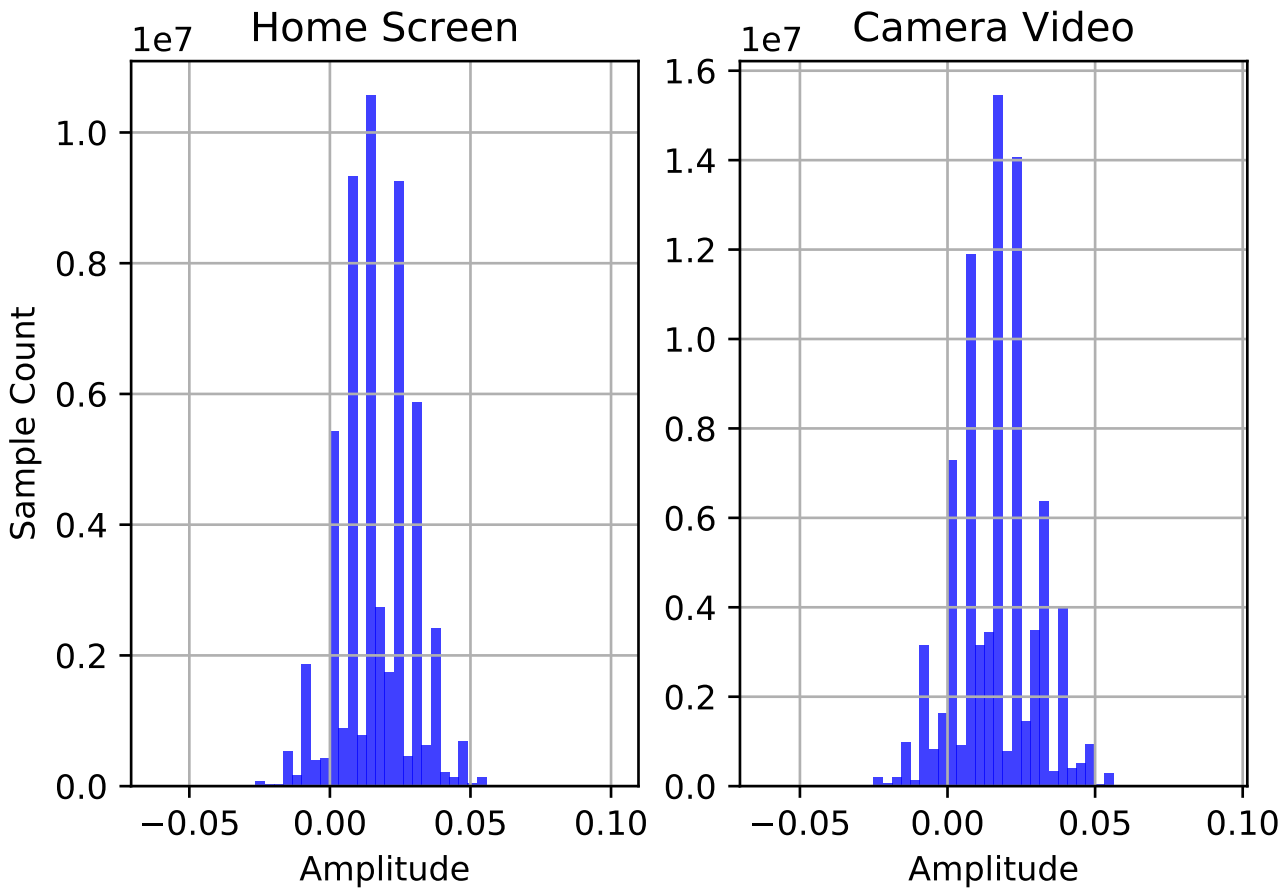
Analysing EM Dataset (cont.)

Amazon Echo Dot – Power Spectral Density (PSD) Plots



Analysing EM Dataset (cont.)

Nokia 4.2 – Histogram



Analysing EM Dataset (cont.)

- Go ahead and launch Jupyter Notebook inside the downloaded Git repository.
- We'll explore the code examples for the following things:
  1. Reading and basic visualisation of EM data.
  2. Preprocessing EM data for feature extraction.
  3. Simple machine learning classifiers to distinguish software behaviour.

Conclusion

- EM-SCA for digital forensic insight acquisition is still in its early days.
- Loads of technical and scientific problems remaining to be solves; great for research!
- *Cross-device portability* of trained models.
- No need to possess hardware equipment to conduct research in this area; datasets are available to work on (from our group and many others).
- Thank you for your participation. Feel free to get in touch:

  Asanka Sayakkara (`asa@ucsc.cmb.ac.lk`)

References

1. Asanka Sayakkara, Le-Khac, N-A., and Scanlon, M., "Electromagnetic Side-Channel Attacks: Potential for Progressing Hindered Digital Forensic Analysis", International Workshop on Speculative Side Channel Analysis (WoSSCA 2018), Amsterdam, Netherlands, July 2018.

2. Asanka Sayakkara, Le-Khac, N-A., and Scanlon, M., "A Survey of Electromagnetic Side-Channel Attacks and Discussion on their Case-Progressing Potential for Digital Forensics", Elsevier Digital Investigation, 2019.

3. Asanka Sayakkara, Le-Khac, N-A., and Scanlon, M., "Leveraging Electromagnetic Side-Channel Analysis for the Investigation of IoT Devices", DFRWS USA, Portland, OR, USA, July 2019.

4. Asanka Sayakkara and Nhien-An Le-Khac , "Electromagnetic Side-Channel Analysis for IoT Forensics: Challenges, Framework, and Datasets," in IEEE Access, vol 9, pp. 113585-113598, 2021.

Find more here: `https://www.asayakkara.org/publications.html`