Chapter 4

Power/EM I

In this chapter, we're going to learn about:

- Electric Engineering basics
- When & Where in an electronic circuit is power consumed
- How can we measure power & EM consumption?

4.1 Electronic Circuits 101

A basic electronic circuit

The most basic electronic circuit (Figure 4.1) consists of a power supply (Vdd) and some sort of electrical load (a component consuming electric power) connected to the power supply on the one hand and to the "ground" (the reference point from which electric potential or voltage is measured) on the other hand. As the electric current - a targeted flow of free electrons - flows through the load, the load does some kind of a "work". We can think of electric current as behaving like water - going from the hill (Vdd) to the valey (Ground) with rivers and obstacles (Loads/Resistors) trying to prevent it from flowing. The higher the resistance of the load, the less current will be able to flow through it.

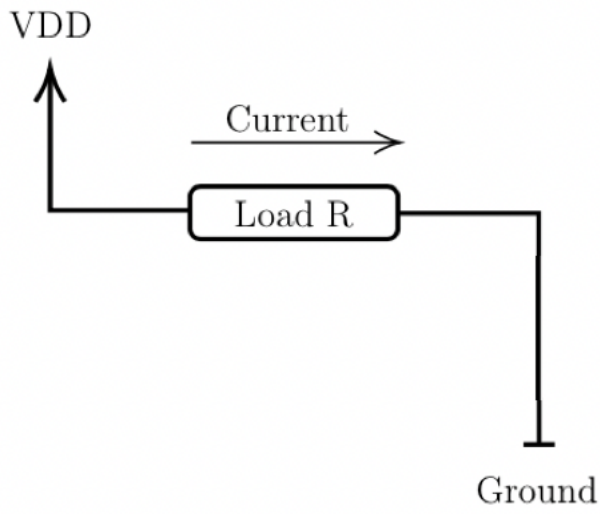

Figure 4.1: Basic electronic circuit.

There are two different ways to wire different loads on an electric circuit - "in series" and "in parallel":

- **In series** - electricity has to pass through one load to reach the other.
- **In parallel** - the loads are conected side by side and the electric current passes through them in parallel.

The electric potential difference between the power supply and the ground creates an electric current which flows through the load toward the ground. The difference in electric potential between two points is measured in Volts (usually denoted by $\boldsymbol{V}$). The magnitude of the current flowing through the circuit at a given time is measured in Amperes (denoted by $\boldsymbol{A}$). The electrical resistance of the load is a measure of its opposition to the flow of electric current through it. It is measured in Ohms (and denoted by $\boldsymbol{R}$).

Resistors

As the name implies, a resistor resists the flow of electrical current. The amount of resistance is measured in Ohms. A resistor is considered a passive component that consumes power that is dissipated as heat. The power rating of a resistor determines how much power it can consume without overheating.

Figure 4.2: Resistor Symbol in Circuit Diagrams.

Ohm's law

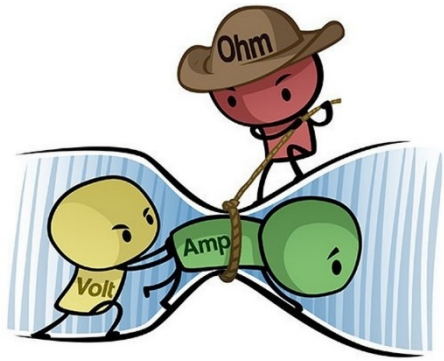

Figure 4.3: Ohm's Law.

Ohm's law defines the relationship between the Voltage, Current and Resistance in a circuit: The voltage is equal to the current multiplied by the resistance of the load.

$$V = I \cdot R$$

Since in most of the circuits we are using, the voltage is fixed (defined by the characteristics of the power supply), a change in the resistance of the circuit will cause an inverse change

in the current. A useful analogy for the relations between V, I and R is to imagine a fountain springing from a high mountain with water flowing down through a river to the sea. The difference in height between the fountain and the sea is the Voltage, the width of the river can be thought of as the resistance, and the flow of the water is the current.

Power

Power is the rate at which work is done by the circuit and is measured in Watts. Electricity bills are measured in units of Kilo-Watt x hour, i.e. 1 KWH is 1,000 watts used for an hour, and this is the energy that we used and we need to pay for.

$$Power = \frac{Work}{Time}$$

As an example we can consider a smartphone: the battery's capacity is measured in milli-Ampere hour (mAh). So, if a battery has 3000 mAh of capacity, and the phone's basline power consumption is 1A then we can use it for 3 hours without recharging. Of course, if we overload the phone by watching videos on youtube (requires cpu/gpu for decoding, wifi/mobile for downloading content, etc.) turning the flash-light on and so forth - then the power consumption would be higher and the battery will run out quicker.

Different types of work can be done using electricity

- Electromagnetic work (light a bulb, transmit a Wi-Fi signal)
- Thermal work (heating)
- Mechanical work (spin a motor, vibrate a speaker's diaphragm to play sounds)
- Chemical work (charging a battery)

- Computational work (store or load from memory, compute a value)

Power Consumption

When the current leaves the circuit to the ground then we consume it as power, but sometimes we need to be careful as there are cases where the current is not really leaving the circuit, like battery charging. In order to measure the power consumption we will connect our measurement device between the load and the ground. The power consumption of a device is the work it does divided by time. It is measured in Watts ($\boldsymbol{W}$). The power consumption can also be calculated as current ($\boldsymbol{I}$) multiplied by Voltage ($\boldsymbol{V}$): $P = I \cdot V$

Current and Voltage dividers

Before we take a look at two simple electronic circuits, we need to introduce two additional terms: A **short (closed) circuit** is a piece of wire with (almost) zero resistance. The circuit is in a closed state and there is electric current flowing through the circuit, simply stated - it "works" as normal. An **open circuit** is a circuit which doesn't allow any current to pass through it. The circuit is in an open state and there is no current in the circuit. That's to say - it doesn't "work".

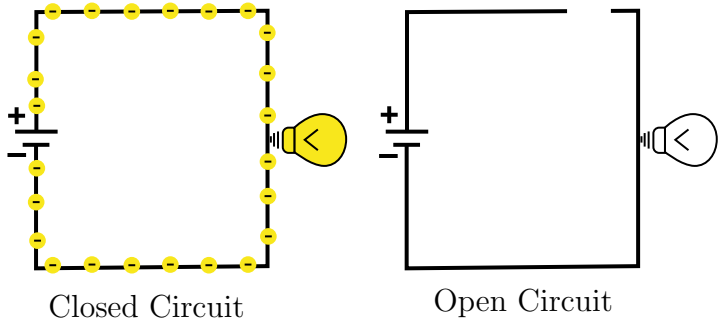

Figure 4.4: Open and closed circuits.

Connecting in serial

If we connect a short circuit between the load and the ground (Figure 4.5), it will have no influence on it since we basically just cut a cable and put another one instead.

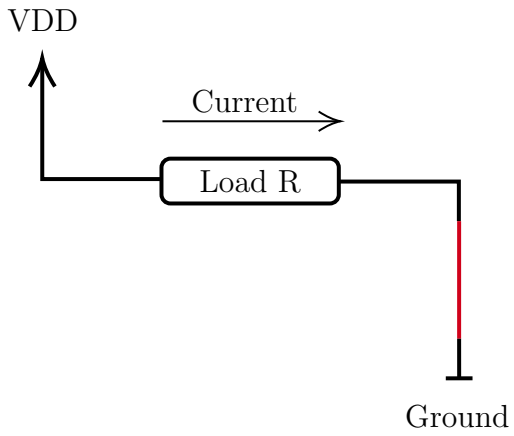

Figure 4.5: short(low resistance) circuit between the load and the ground.

If we connect an open circuit after the load (See Figure 4.6), it will increase the resistance to a very high value, causing the current to become effectively zero. And if the current is zero - the voltage is also zero (Ohm's Law).

Connecting in parallel

If we connect an open circuit in parallel to the load (See Figure 4.7), the current will flow only through the load's path, so the current on the open circuit will be 0. However, the voltage drop between both points of the open circuit will be the same as the drop between the load sides.

If we connect a short circuit in parallel to the load (See Figure 4.8), the current will "prefer" flowing through it rather than through the load, so the current through the load will be equal to zero, while the current through the short circuit will be very high - by Ohm's law, since the voltage stays the same as before.

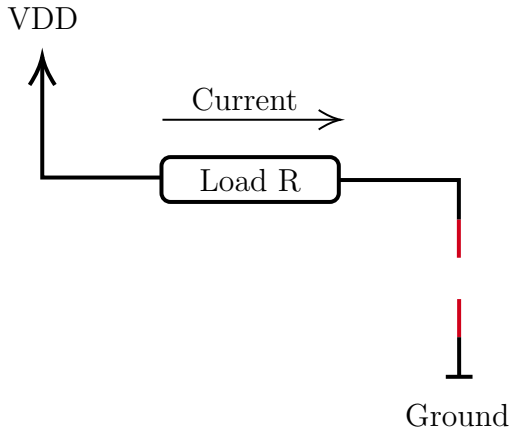

Figure 4.6: open circuit after the load.

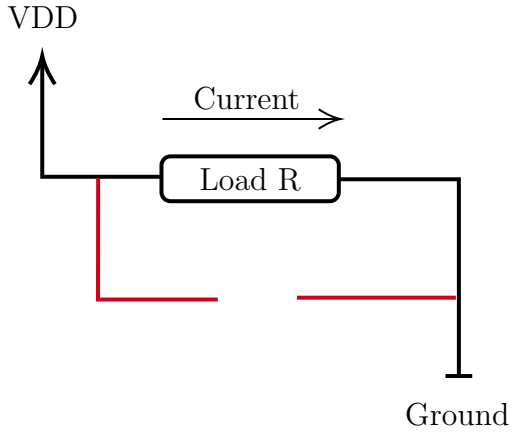

Figure 4.7: A close circuit in parallel to the load.

Since the cable is not a perfect conductor, some of the energy will be consumed in the form of thermal work, so the cable will heat up.

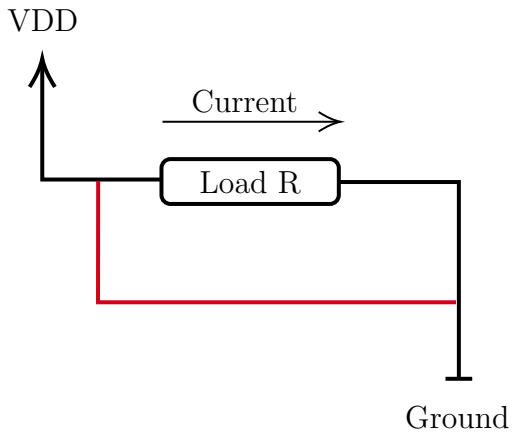

Figure 4.8: A short circuit in parallel to the load.

4.2 Measuring Power Consumption

As an attacker, we want to measure the power consumption of the load in order to deduce interesting insights about it. For this, we are going to use an Amperemeter.

Amperemeter

An Amperemeter is a device capable of measuring the amount of electric current going through it. It has a very low resistance, so it doesn't affect the circuit it connects to - since it is basically equivalent to an extra piece of conducting wire.

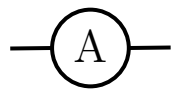

Figure 4.9: Amperemeter Symbol in Circuit Diagrams.

Using an Amperemeter to measure power consumption

To use an Amperemeter - we "cut" the wire connected to the load and connect both sides to the Amperemeter.(See

Figure 4.10)

Doing so causes all current flowing through the load to pass through the Amperemeter as well, so we are able to read the current at any given time. The resistance of the Amperemeter is very low so it doesn't affect the circuit's voltage.

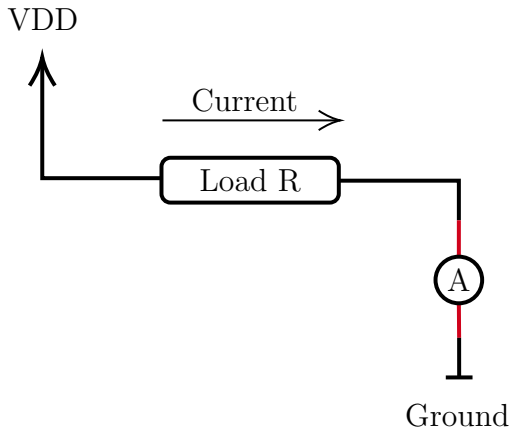

Figure 4.10: an Amperemeter connected in serial.

To measure the power consumption, we measure the current (I) and since the voltage (V) hasn't changed, we can compute the power consumption as $P = I \cdot V$. One major shortcoming of this method is that sometimes we don't want (or can't) cut the circuit to connect an Amperemeter. A potential alternative would be to connect the Amperemeter in parallel (See Figure 4.11)

However, connecting the Amperemeter in parallel would actually burn the Amperemeter as it has no resistance and so, all of the current will flow through it. So, instead of an Amperemeter we can use a Voltmeter as in Figure 4.12.

Voltmeter

The Voltmeter's resistance is very very high so the current will not go through it. The Voltmeter is measuring the voltage drop between one side of the load and the other side of

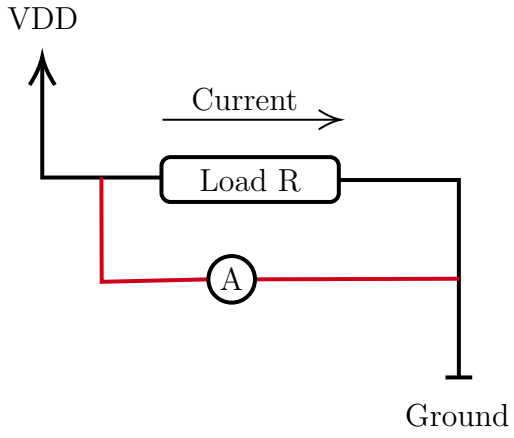

Figure 4.11: An Amperemeter connected in parallel to the load

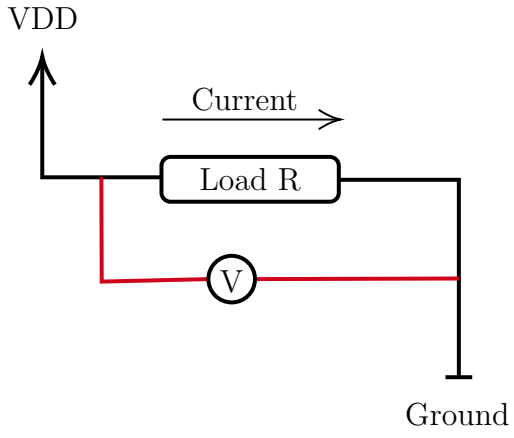

Figure 4.12: a Voltmeter connected in parallel to the load

it. To measure the current using a Voltmeter we take a load with a very small resistance and connect it as in Figure 4.13.

By connecting a Voltmeter in parallel with a very small and accurate resistor, we can measure the electric current using Ohm's law: $I = V/R$

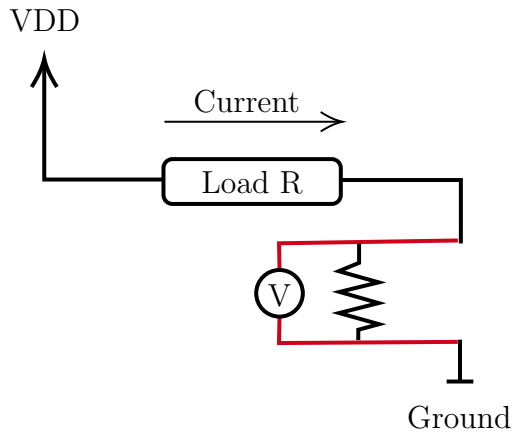

Figure 4.13: Voltmeter connected in parallel

Power consumption - what's next?

We saw what power consumption means and how we can measure it. It is very important to note that for every interesting enough circuit - **the power consumption varies with time!** Our goal, as an attacker, is to find a relationship between the secret information we want to extract and the power consumption. Then, we can exploit this relationship by measuring the device's/circuit's power consumption over time and using this information to extract secret info.

Types of electronic components

Generally speaking, there are two types of components in an electric circuit:

- **Passive devices**:
  - Resistor
  - Inductor
  - Capacitor
  - Diode
- **Active devices**:

- Transistor
- Amplifier
- Integrated circuit (IC)

For us, active devices are much more interesting as their power consumption characteristics change as a function of the state of the circuit.

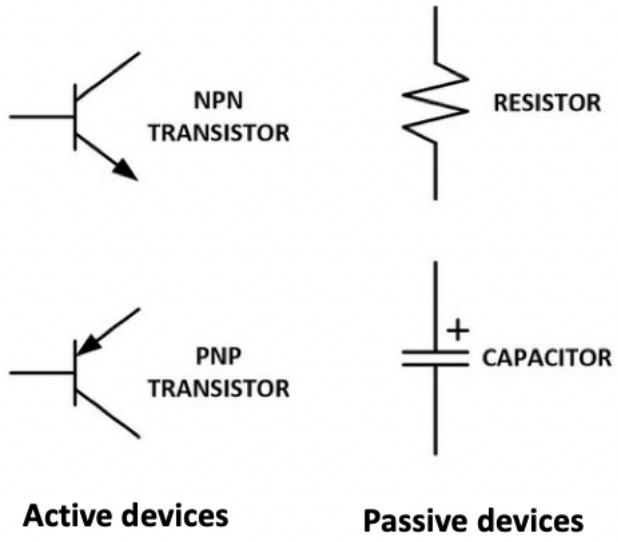

Figure 4.14: Examples of electronic components.

Specifically, of outmost interest for us is the transistor. In integrated circuits, there are lots transistors. Analyzing their power consumption behaiour can shed a lot of light on the data they're processing, There are various kinds of transistors and we will concentrate on understanding a specific type - namely the "Field-Effect Transistor" (FET).

Field-Effect Transistor

The Field-Effect Transistor (FET) is an electronic device which uses an electric field to control the flow of current. FETs are 3-terminalled devices, having a source, gate, and

drain terminals. FETs control the flow of current by the application of a voltage to the gate terminal, which in turn alters the conductivity between the drain and source terminals. In order to understand FETs, we first need to gain some understanding of semiconductors.

Semiconductors

We know that some materials, like copper or gold are good conductors, and others - like plastic or glass are very bad conductors or insulators.

A **semi**conductor is a substance, usually a solid chemical element or compound, that can conduct electricity under some conditions but not others, making it a good medium for the control of electrical current. Its conductance varies depending on the current or voltage applied to a control electrode, or on the intensity of irradiation by infrared (IR), visible light, ultraviolet (UV), or X rays.

Generally, an atom is built from three sub-particles:

- Neutrons - irrelevant for this discussion as they carry no electric charge
- Protons - heavy, positively charged particles forming the nucleus of the atom (together with the neutrons).
- Electrons - light and negatively charged particles orbiting the atom's nucleus in predefined orbits (this is basically Bohr's model of the atom which isn't quite accurate from the point of view of quantum mechanics, but is a good approximation in our context and in chemistry)

The **Silicon** atom has four electrons in its outer orbit (also known as *valence electrons*). In Silicon crystals, all of those outer orbit electrons are involved in the covalent bonds between the Silicon atoms. This implies that pure Silicon is a pretty bad conductor, since conductance relies on the flow of

free electros, which are very hard to find in Silicon crystals as all outer-orbit electrons are busy forming covalent bonds and aren't free to move around between atoms.

We can change the behavior of the silicon and turn it into a conductor by doping it - i.e. mixing a small amount of an impurity into the silicon crystal.

There are two types of such useful impurities we can introduce:

- N-type – where phosphorus or arsenic is added to the silicon in small quantities. They both have five outer orbit electrons, so one of them is out of place when they get into the silicon lattice. While having nothing to bond to, the fifth electron is free to move around. As electrons have a negative charge, this kind of impurity is called N-type (N = Negative).
- P-type - where boron or gallium is added to the silicon. They both have only three outer electrons. So, when we mix them into the silicon lattice, there will be "holes" in the lattice where a silicon electron has nothing to bond to. The hole is looking for an electron from a neighbor atom and when that happens the hole is "moving". As the absence of an electron creates the effect of a positive charge, this kind of impurity is called P-type (P = Positive).

How does the Field-Effect Transistor work?

In the Field-Effect Transistor there are two "n+" areas with N-type doping and a "p" area with P-type doping (see Figure 4.15). The "n+" areas contain a lot of free electrons and the "p" area contains a lot of "holes". Initially, the free electrons from the "n+" areas are moving into the holes in the "p" area and the transistor as a whole is not conducting any current form the Source to the Drain (i.e. is an open circuit) as there are no free electrons that can move around. When an electrical field is applied to the Gate it gets charged with

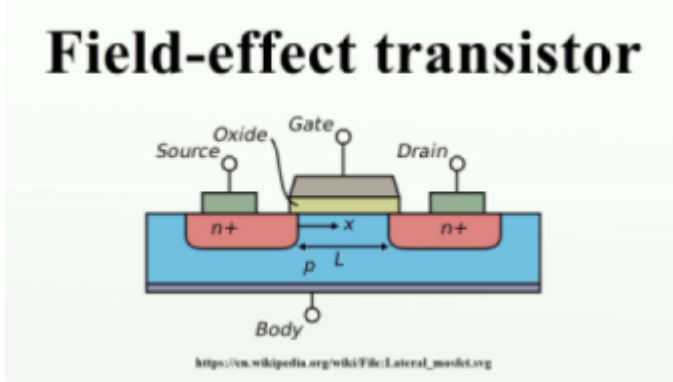

Figure 4.15: Field-Effect Transistor.

lots of free electrons. These electrons in the Gate can't move to the Silicon itself as there is an insulating oxide layer between them, but the repulsive force between the negatively charged electrons pushes the electrons in the part of the "p"" area between the "n+" areas to the bottom creating a "channel" allowing the flow of electrons from Source to the Drain through the holes left by the electrons the were pushed away - thus turning the transistor into a conductor (i.e. closed/short circuit).

Buliding an Amplifier using a FET

We can easily construct an Amplifier using a FET by connecting the input signal to the FET's Gate terminal and the power supply to the Source terminal. Then, the Drain terminal's output current will be an amplification of the input signal going into the Gate.

CMOS

Complementary Metal Oxide Semiconductor (CMOS) is a method for creating logical gates. Its core concept is based on using a "pull up network" (Vdd) denoting a logical 1 and a "pull down network" (Vss) denoting a logical 0. Then, a set of connections and transistors create a closed circuit of

either the pull up or the pull down network with the output terminal "Q", depending on the value of the input terminal(s). It is very important to design the circuit in such a way that for every input, exactly one of the networks will be close-circuited to the output, since if none of them is close-circuited the output will be undefined and if both of them are close-circuited to the output they will short circuit each other, potentially causing damage to the circuit.

Now, let's see a few examples for implementing logical gates using CMOS.

NOT Gate CMOS

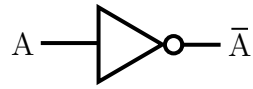

Figure 4.16: Logical NOT Gate

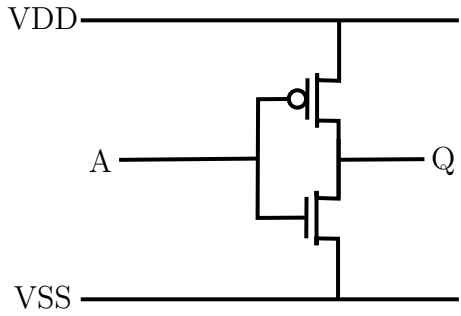

Figure 4.17: CMOS NOT gate

When the voltage of input A is low (A=0), the upper transistor's channel is closed (since this is a PMOS transistor that closes the circuit when it gets a 0 input signal) and we have a connection between Vdd(1) and Q, so Q = 1. When the voltage of input A is high (A=1), the lower transistor's circuit is closed (since it is an NMOS transistor) and we have a connection between Vss(0) and Q, so Q = 0. It is trivial to see that we've implemented a logical NOT operation with A being the input and Q being the output.

Table 4.1: NOT gate truth table.

input	A	0	1
output	Not A	1	0

When does a CMOS circuit consume power?

There is an interesting question though – when does this circuit consume power? As mentioned previously, power is consumed when eletric current from the power source (Vdd) reaches the ground (Vss). In our case there is never a connection between Vdd and Vss so it appears power is never consumed, thus defying the energy conservation principcal. The answer to this is that a bit of power is actually consumed when transistors switch between their two possible outputs - i.e. connections with either Vss or Vss. We will see later on, how these minor power consumptions can be used for our purposes.

AND Gate CMOS

Figure 4.18 is a logical AND gate.

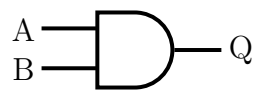

Figure 4.18: AND Gate

We will create a CMOS AND gate using 4 transistors:

Table 4.2: AND gate truth table.

input a	0	0	1	1
input b	0	1	0	1
output	0	0	0	1

First we want to build the Pull Up Network, so for input A and B, for A=B=1 then the output Q is 1. Then we will build the Pull Down Network to support the other combinations of inputs from the truth table to deliver 0 as the output Q.

Storage circuits

Now, let's see how we can use CMOS circuits to store data - i.e. create memory components. A flip-flop is a circuit, comprised of two latches that has two stable states and can be used to store a single bit of data.

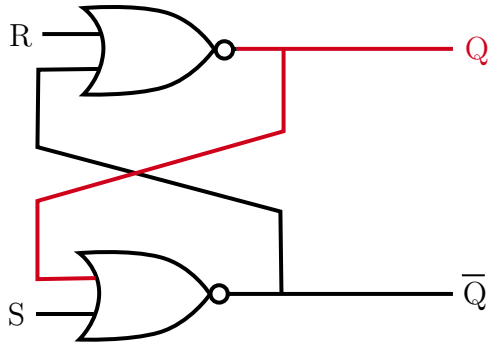

Figure 4.19: Flipflop

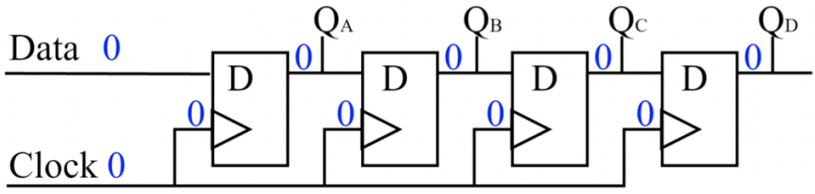

Figure 4.20: A series of flip-flops

The circuit can be made to change state by signals applied to one or more control inputs and will have one or two outputs. It is the basic storage element in sequential logic. Flip-flops and latches are fundamental building blocks of digital electronics systems used in computers, communications, and many other types of systems.

A flip-flop is a device which stores a single bit of data; one of its two states represents a "one" and the other represents a "zero". Such data storage can be used for storage of state, and such a circuit is described as sequential logic in electronics. When used in a finite-state machine, the output and next state depend not only on its current input, but also on its current state (and hence, previous inputs). It can also be used for counting of pulses, and for synchronizing variably-timed input signals to some reference timing signal.

Flip-flops can be either level-triggered (asynchronous, transparent or opaque) or edge-triggered (synchronous, or clocked). The term flip-flop has historically referred generically to both level-triggered and edge-triggered circuits that store a single bit of data using gates. We will refer to Flip-Flop as edge-triggered i.e. clock-synchronized.

It is both interesting and important to note that for each storage element in a circuit the data/output changes at the same time orchestrated by the clock signal and so we have a combined, amplified signal that we, as attackers, can monitor to extract secret information.

Core i7 chip

To demonstrate how combinations of CMOS gates scale up to form real-world chips, Figure 4.21 is an image of a modern Intel core i7 CPU, which is fundamentally a CMOS chip.

We can see a multitude of different components, most notably:

- The integrated GPU containing multiple repeating patterns denoting the GPU processing units
- The identically looking 4 CPU cores with the repeating yellow patterns of CPU cache memory

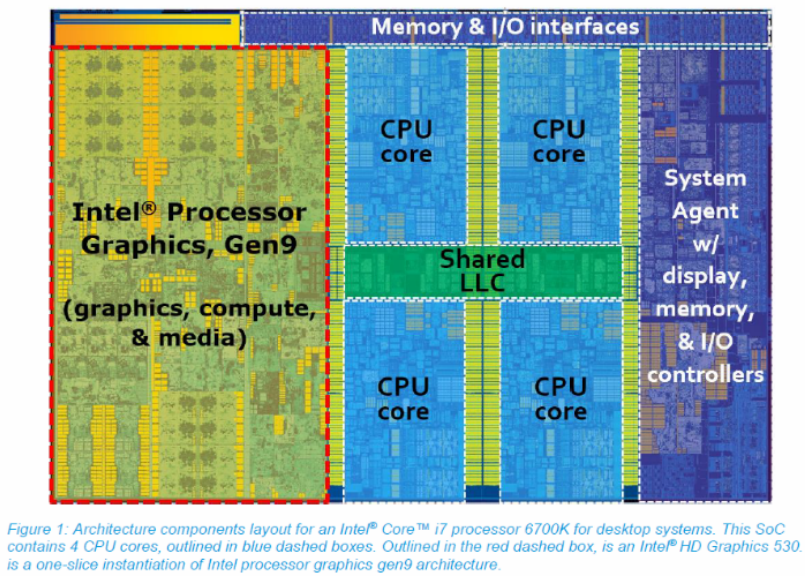

Figure 4.21: Intel i7 CPU

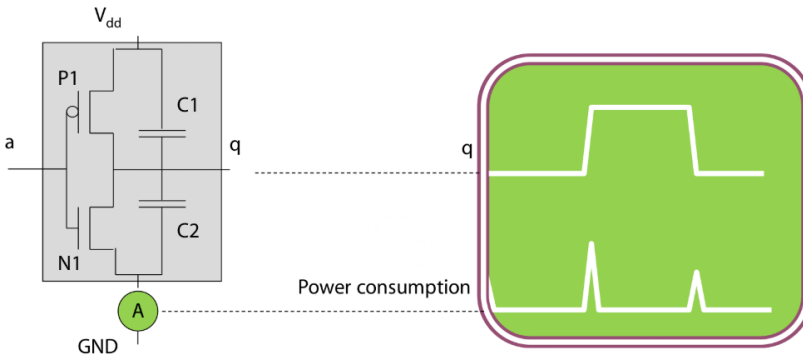

Figure 4.22: Not gate connected to an oscilloscope.

Power Consumption Variability

In Figure 4.22 we have a CMOS NOT gate, like the one we saw before, with an Amperemeter connected to it to measure the power consumption, and a plot of both the gate's output (q) and power consumption as would've been captured by an oscilloscope connected to the Amperemeter. C1 and C2 are capacitors that are required due to low-level electrical

characteristics which we won't describe in deail.

In the plot we see a few interesting phenomena:

- The power consumption is zero when the output (Q) is constant.
- There is a spike in power consumption each time the output (Q) changes.
- The power consumption spike goes higher when Q changes from 0 to 1 compared to when Q changes from 1 to 0.

More generally, power consumption is the sum of the static and the dynamic power consumption which changes with time:

$$P(t) = P_{stat} + P_{dyn}(t)$$

As an attacker, we don't really care about the static power consumption that depends on the device's supply voltage, transistor manufacturing technology, etc. The dynamic power consumption, on the other hand, depends on the clock rate, the circuit activity and input data. It can reveal a lot about the data flowing through the circuit, which could be very beneficial.

Theoretically, we can accurately measure the full power consumption of the device on the one hand and compute a full power consmuption simulation on the other hand and then try to extract the interesting data by comparing the results and solving the resulting equations to find the data manipulated during the measurement. However, making this accurate computation isn't feasible and we'll, instead, compromise for an approximate model of the device's power consmuption that will allow us to achive practical results subject to a set of assumptions we'll make about the device.

Hamming distance model

We'll assume the following about the device we wish to attack:

- This is a CMOS device.
- When it's static in terms of the outputs of its transistors - the static power is very low, and when it switches, there is a relatively high power consumption.
- This is synchronous circuit, which means it has a lot of transistors that all change at the same time.
- The power consumption is proportional to the amount of changes and the outputs of these transistors.

Count the number of changes in the output bits of the device's transistors brings us to the so called "Hamming distance model".

The *Hamming distance* between two vectors is the number of differing bits between the vectors. For example, the Hamming distance between the vectors 01101010 and 11011011 is 4, since they differ in exactly 4 indices, namely: 0,2,3,7.

Formally, assuming we have a vector:

$$X = x_n, x_{n-1}, ..., x_1, x_0$$

We'll first define the "Hamming weight" operator:

$$HW(X) = \sum x_i$$

Then, we'll define the "Hamming distance":

$$HD(X, Y) = HW(V_1 \bigoplus V_2)$$

For CMOS devices, we'll approximate the power consumption as proportional to the amount of bit transitions from one to zero or from zero to one - i.e. the hamming distance between the state vectors (the set of all CMOS transistor outputs) of the device across a time interval.

Power Consumption Noise

This is all good in a "perfectly spherical" world, but in the real world we still need to account for noise. There are a number different types of noise that affect our power consumption measurements:

- **Switching noise** - computations or some other power consumption occuring in the device except for the computation we want to monitor, that also affect the power consumption. This can be either correlated (i.e. happens every time our target computation is performed) or uncorreleated (i.e. happens in random relative to our target computation and so can be easily reduced by repeating the measurement a few times).
- **Measurement noise** - caused by mild inaccuracies in our measurement equipment and setup.
- **Thermal noise** - noise caused by electrons jumping arround, appearing, disappearing and creating radiation along the way. This noise is higher when the temperature of the device/circuit is higher.

To summarize, the actual measured power consumption can be modeled as:

$$P_{meas}(t) = P_{stat} + P_{dyn}(t) + N(t)$$

(N(t) - Noise)

A number of approaches exist to minimize the amount of noise, depending on its nature:

- Repeat the measurement to average out uncorrelated noise.
- Control the amount of thermal noise by running the experiment in very low temperatures.
- Prevent radiation originating noise by putting the device in a Faraday cage.

- Modify the device in a way that either increases the signal or decreases the nosie. For example - remove a noise generating module.

Another thing we can do is move from measuring power consumption to measuring electromagnetic radiation.

From Power to EM

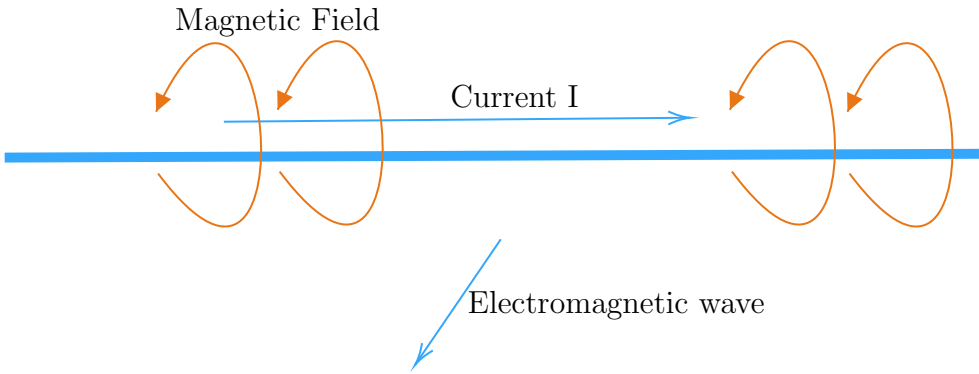

Figure 4.23: electromagnetic emission.

A fundamental law of physics states that moving charged particles create magnetic fields. Specifically, a directed electric current creates a magnetic field with magnitude and orientation descibred by the Biot-Savart law. The direction of the magnetic field can be found using the "left hand rule" (Figure 4.23). This is very good news for us, as this means that instead of measuring the power consumption, we can simply measure the magnetic field and from it deduce the magnitude of the electric current that is its source. One disadvantage of this method is that it is very localized, so we need to get VERY close to the device or alternatively point a very sensitive directional antenna to it to make measurements.

Measurement setups

We'll now take a look at a few examples of attacker setups.

Power measurement setup

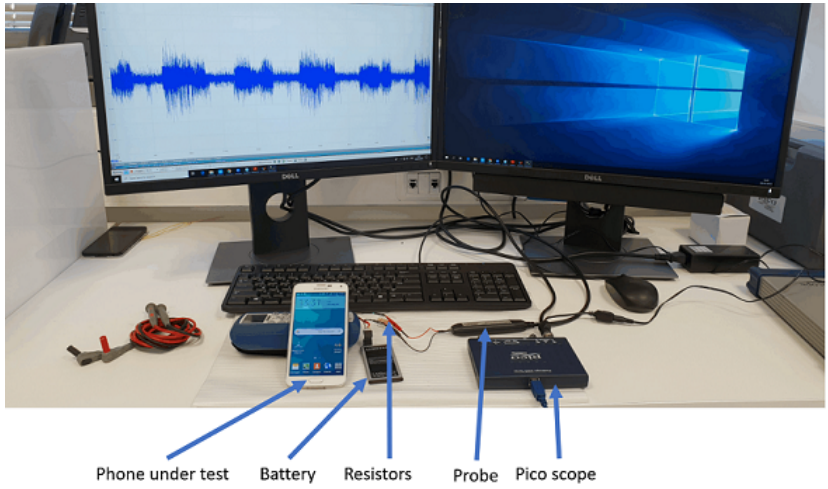

Figure 4.24: Power measurement setup

In Figure 4.24 we have:

- A Pico scope - a small oscilloscope that measures voltage over time.
- A Voltmeter probe that is connected to the battery of the phone under test using a bunch of small resistors connected in parallel - this is a way to measure the current, as we explained above.

EM measurement setup

In Figure 4.25 we want to measure the EM field of a micro controller chip. We place the whole setup on a computer-controlled XYZ stage that allow placing the board very close to the Magnetic field probe with very high (micron level) precision - this is crucial since electromagnetic emanations

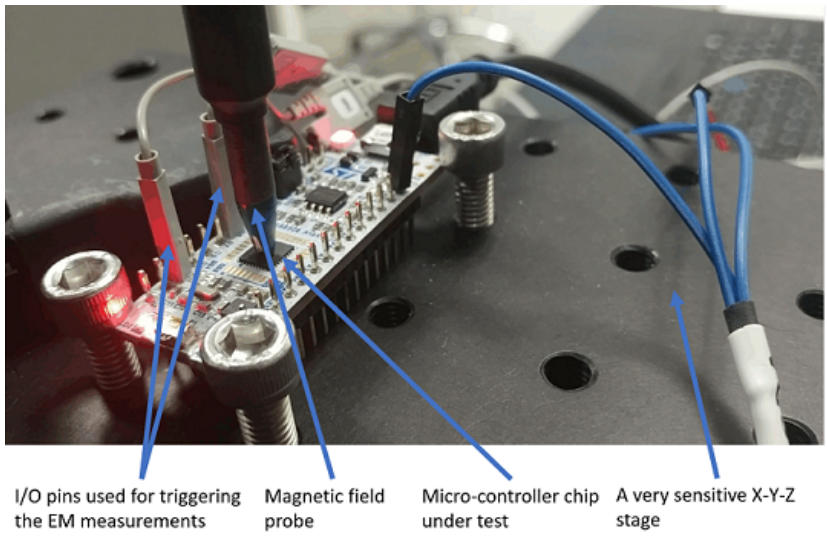

Figure 4.25: EM measurement setup

can only be recorded from a very close distance. Another important thing to note is the pair of gray cables connected to I/O pins on the board, used for triggering the beginning and end of the measurement - this is very beneficial because it allows us to accuretly measure the EM field for a very short time interval during which the computations we'd like to track occur. A short time interval allows us recording with a higher resolution, gaining more valuable info that will help analyse the signal to extract the secret data we want.

Generic attack setup schematic

Figure 4.26 describes a generic power-analysis attack setup schematic. We have the Device Under Test (DUT) connected to a very clean power supply (preferably a battery that gives a much cleaner supply of power than a standard AC/DC power supply). The "Stimulus Generator" is either a piece of software or hardware that allows us to automatically trigger the device's operation(s) we want to monitor. Then, we have the measurement front-end which is either a "Baseband Front End" i.e. a detour in the wiring conneced to a cur-

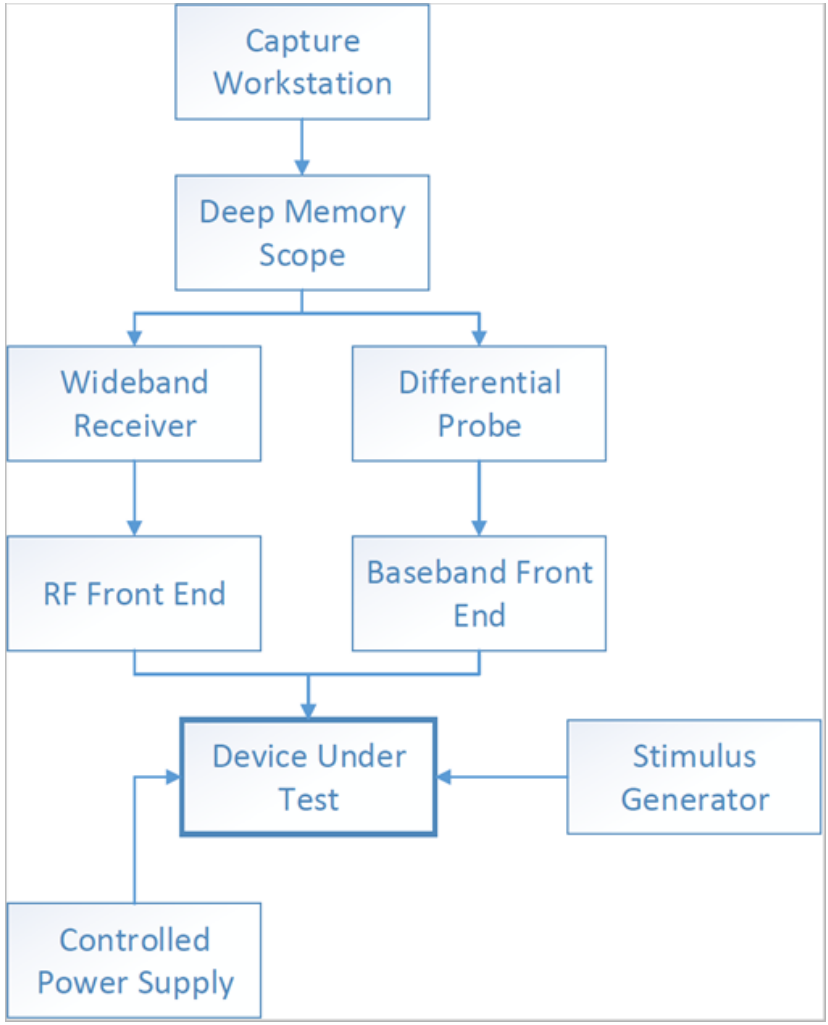

Figure 4.26: Generic attack setup schematic

rent/voltage proble or an "RF Front End" and a Wideband receiver for measuring the Electormagnetic field with good accuracy. The "Deep Memory Scope" is basically our oscilloscope that is capable of capturing lots of data very fast and transfer the digitized data to the "Capture Workstation" which is a very strong server or workstation that can handle the burden of thoroughly analysing the data to get the results we need. This is quite a costly setup with cost in the

range of hundreds of thousands of dollars, but once we have it in our lab - we can perform some very cool attacks!

Research Highlights

- A nice paper provides a short biographical sketch of Ohm and a discussion of his experimental and theoretical work in general and ohm's law in specific [31].
- Stefan Mangard on his paper [32] presents a simple power analysis attack on AES Implementation of the key expension. The attack, which performed on smart cards, exploit the infromation leaking during the AES key expansion, and utilizes it to substantially reduce the key space that needs to be considered in a brute-force search for the secret key.
- W. Shan [33] presented a countermeasure to AES attack. the Algorithm porposed, based on maching learning, Wishes to find out the best hamming distance redistribution mapping to compensate the probability of hamming distance of the intermediate data directly, thus, make it unable to be distinguished from correct and incorrect sub-key.
- Another Paper posted on 2015 [34] shows the weakneses of a very importanct component for running deep learning Algorithms, the GPU. The paper results, tested on NVIDIA TESLA GPU, shows that parallel computing hardware systems such as a GPU are highly vulnerable targets to power-based side-channel attacks, and survey some of the weakneses.