1. What is a NPU and a microNPU?

A **neural processor**, a **neural processing unit** (**NPU**), or simply an AI Accelerator is a well-partitioned circuit that comprises all the control and arithmetic logic components necessary to execute [machine learning](https://en.wikichip.org/w/index.php?title=machine_learning&action=edit&redlink=1) algorithms. NPUs are designed to accelerate the performance of common machine learning tasks such as image classification, machine translation, object detection, and various other predictive models. **NPUs may be part of a large SoC,** a plurality of NPUs may be instantiated on a single chip, or they may be part of a dedicated neural-network accelerator.

NPUs sometimes go by similar names such as a *tensor processing unit* (*TPU*), *neural network processor* (*NNP*) and *intelligence processing unit* (*IPU*) as well as *vision processing unit* (*VPU*) and *graph processing unit* (*GPU*).

Executing [deep neural networks](https://en.wikichip.org/w/index.php?title=deep_neural_networks&action=edit&redlink=1) such as [convolutional neural networks](https://en.wikichip.org/w/index.php?title=convolutional_neural_networks&action=edit&redlink=1) means performing a very large amount of [multiply-accumulate operations](https://en.wikichip.org/w/index.php?title=multiply-accumulate_operations&action=edit&redlink=1), typically in the billions and trillions of iterations. The large number of iterations comes from the fact that for each given input (e.g., image), a single convolution comprises of iterating over every channel, and then every pixel, and then performing a very large number of MAC operations. Many such convolutions are found in a single model and the model itself must be executed on each new input (e.g., every camera frame capture).

Unlike traditional [central processing units](https://en.wikichip.org/wiki/central_processing_units) which are great at processing highly serialized instruction streams, machine learning workloads tend to be highly parallelizable, much like a [graphics processing unit](https://en.wikichip.org/w/index.php?title=graphics_processing_unit&action=edit&redlink=1). Moreover, unlike a GPU, NPUs can benefit from vastly simpler logic because their workloads tend to exhibit high regularity in the computational patterns of [deep neural networks](https://en.wikichip.org/w/index.php?title=deep_neural_networks&action=edit&redlink=1). For those reasons, many custom-designed dedicated neural processors have been developed.

1. What is a SOC ?

**It is same as a PROCESSOR.** A system on a chip or system-on-chip is an [integrated circuit](https://en.wikipedia.org/wiki/Integrated_circuit) that integrates most or all components of a computer or other [electronic system](https://en.wikipedia.org/wiki/Electronics). These components almost always include a [central processing unit](https://en.wikipedia.org/wiki/Central_processing_unit) (CPU), [memory](https://en.wikipedia.org/wiki/Computer_memory) interfaces, on-chip [input/outpu t](https://en.wikipedia.org/wiki/Input/output) devices,  [input/output](https://en.wikipedia.org/wiki/Input/output) interfaces, and [secondary storage](https://en.wikipedia.org/wiki/Computer_data_storage#Secondary_storage) interfaces, often alongside other components such as [radio modems](https://en.wikipedia.org/wiki/Radio_modem) and a [graphics processing unit](https://en.wikipedia.org/wiki/Graphics_processing_unit) (GPU) – all on a single [substrate](https://en.wikipedia.org/wiki/Wafer_(electronics)) or microchip.[[1]](https://en.wikipedia.org/wiki/System_on_a_chip#cite_note-1) It may contain [digital](https://en.wikipedia.org/wiki/Digital_signal_(electronics)), and also [analog](https://en.wikipedia.org/wiki/Analog_signal), [mixed-signal](https://en.wikipedia.org/wiki/Mixed-signal_integrated_circuit), and often [radio frequency](https://en.wikipedia.org/wiki/Radio_frequency) [signal processing](https://en.wikipedia.org/wiki/Signal_processing) functions (otherwise it may be considered only an application processor).

What is MAC/MAD operations

In [computing](https://en.wikipedia.org/wiki/Computing), especially [digital signal processing](https://en.wikipedia.org/wiki/Digital_signal_processing), the **multiply–accumulate** (**MAC**) or **multiply-add** (**MAD**) operation is a common step that computes the product of two numbers and adds that product to an [accumulator](https://en.wikipedia.org/wiki/Accumulator_(computing)). The hardware unit that performs the operation is known as a **multiplier–accumulator** (**MAC unit**); the operation itself is also often called a MAC or a MAD operation.

1. **Different types of processor architectures?**

An important aspect of an SoC is not only which components or blocks it houses, but also how they interconnect. Processor Architecture is a solution for the blocks to interface with each other.

**The two popular companies(Intel and ARM) design their own CPUs/processors, components of Processors and architectures(instruction set) and licenses them to other companies,** who design their own products around these CPUs/processors & that implement one or more of those architectures, including [system on a chip](https://en.wikipedia.org/wiki/System_on_a_chip) (SoC) and [system on module](https://en.wikipedia.org/wiki/System_on_module) (SOM) designs, that incorporate different components such as memory, interfaces, and [radios](https://en.wikipedia.org/wiki/Radio#Radio_communication). **These companies (INTEL and ARM) also designs**[**cores**](https://en.wikipedia.org/wiki/Semiconductor_intellectual_property_core) **(same as processors and CPU)**that implement these [instruction set architectures](https://en.wikipedia.org/wiki/Instruction_set_architecture) and licenses these designs to many companies that incorporate those core designs into their own products.

A core is **a small CPU or processor built into a big CPU or CPU socket**. It can independently perform or process all computational tasks. From this perspective, we can consider a core to be a smaller CPU or a smaller processor. ex: ARM Cortex M55 , Cortex A9 etc.

**Example: Intel Architectures v/s ARM artictectures**

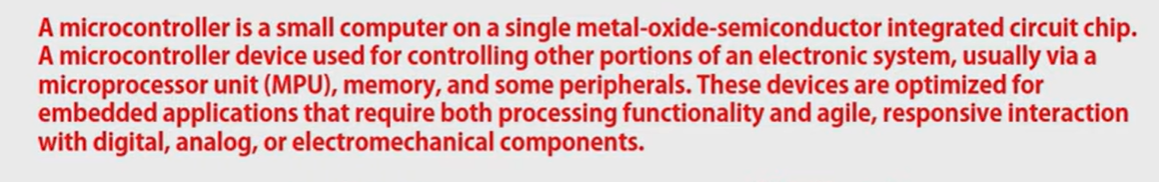
Intel processors/architectures are very complex. This type of processor is known as CISC (Complex Instruction Set Computer) architecture. ARM’s architecture/processors are a lot simpler, known as a RISC (Reduced Instruction Set Computer) architecture. The most powerful Intel chip has over 2.5 billion transistors. The ARM processor core only has around 35,000 That means the cost of production of an Intel chip is much higher and it uses more electrical power. The result of that is that ARM is immensely popular in applications where low cost and low power requirements are important.

**What is an ARM processor?**

ARM processors are a family of central processing units (CPUs) based on a reduced instruction set computer (RISC) architecture. ARM stands for Advanced RISC Machine. ARM architectures represent a different approach to how the hardware for a system is designed when compared to more familiar server architectures like x86.

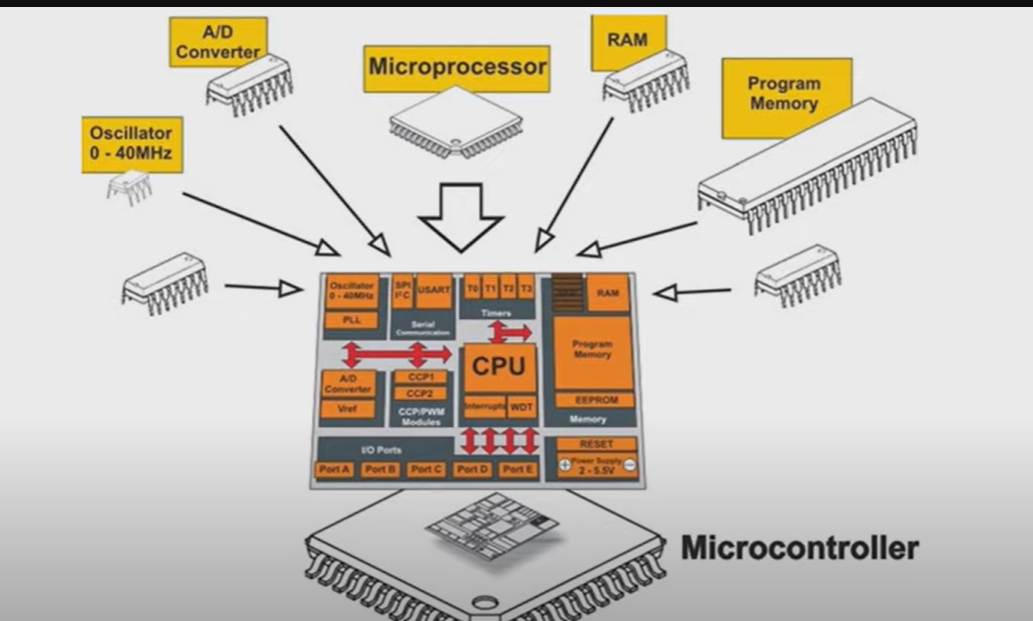
However, **ARM itself does NOT manufacture these CPUs themselves, instead companies like Apple, Qualcomm etc. source the licenses of these CPU and manufactures them (sometimes even these companies may again outsource their manufacturing to fabs in Taiwan(TSMC) or Korea because they are busy designing their SOCs). These CPUs manufactured by anyone is referred to as an ARM chip .**

**What is a MCU or a Microcontroller Unit?**



**Note : A MCU contains similar items that a computer contains: like a CPU(or a microprocessor), i/p - o/p devices, Ram, A/D (A to D converter) etc. (see below) placed on a single chip.**

**NOTE: CPU is same as a Processor or a microprocessor.**

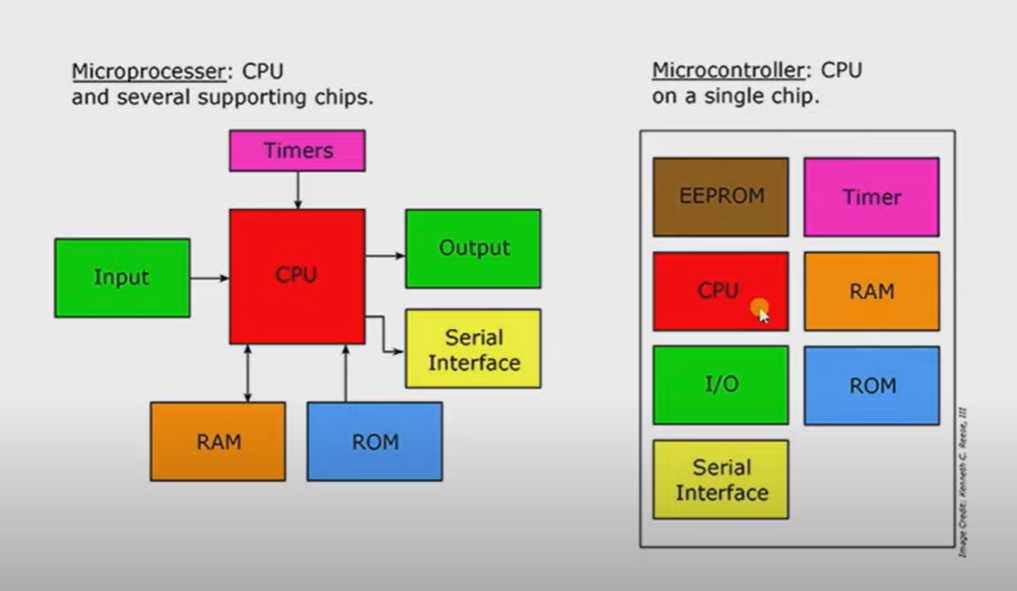


Applications of a MCU: include ATM, Printers, almost all electronic or electrical components, IOT devices like routers etc.

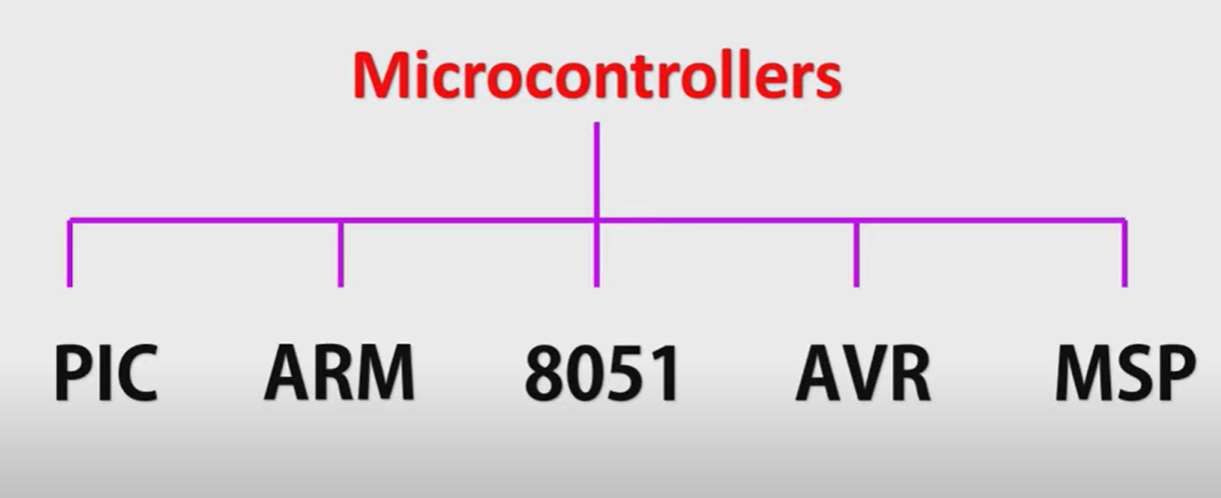


**Difference b/w Microcontrollers and microprocessors**: Simply microprocessors are CPUs with some other basic features like i/p port like USB port etc.

Microcontrollers in general have both program and data memory inside a single chip, along with various peripherals, whereas microprocessors use external memory and peripherals (e.g. the Intel or AMD processor in your desktop or laptop).

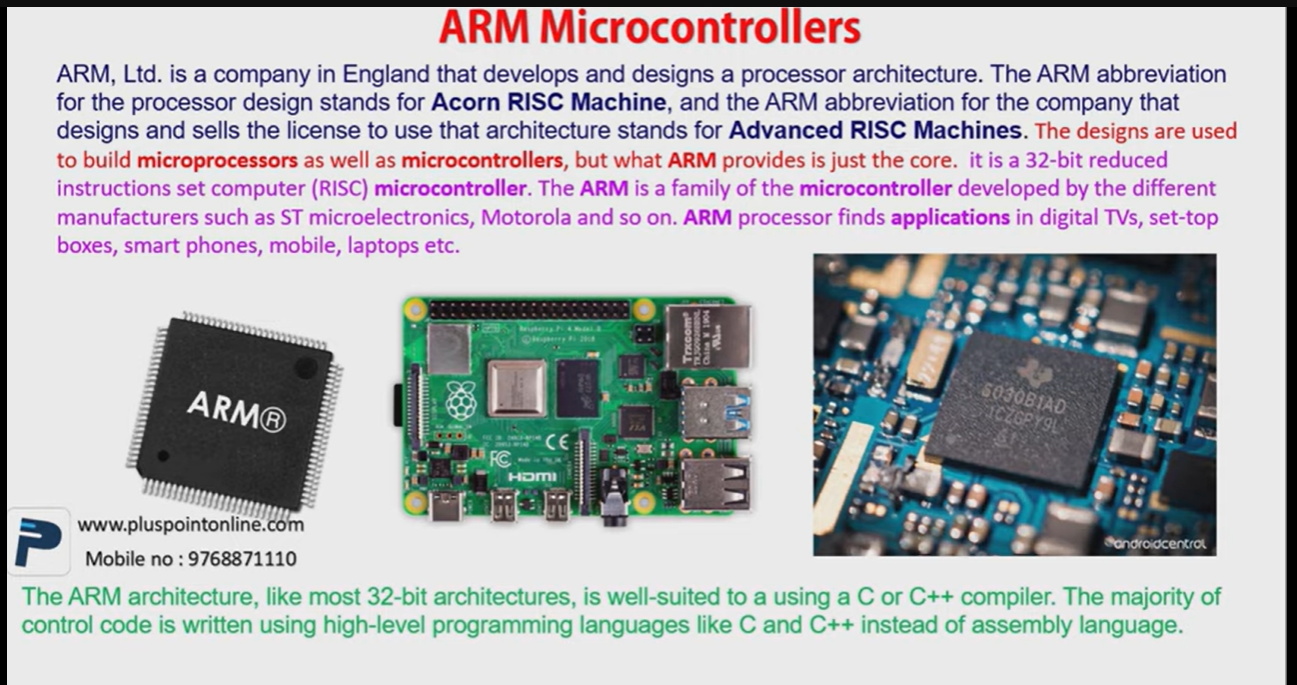


Types of Micro Controllers:



**Note : ARM Cortex M series are processors which find their wide applications in MCUs.**

**NOTE: ARM makes both these Processors ((or microProcessors or CPUs that go inside the MCUs)) and the microcontrollers(MCUs) . However they only design these usually using their RISC architecture to design these.**

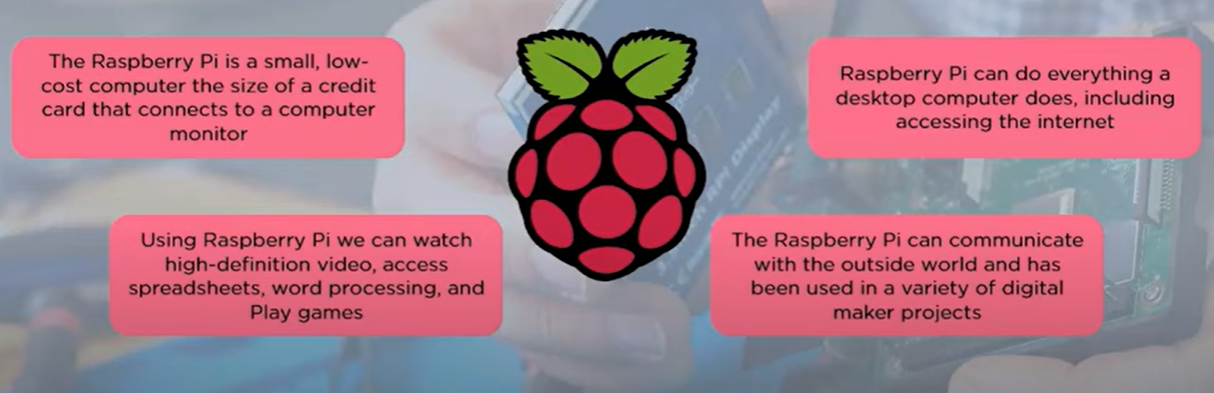


Note: Usually RaspberryPie use ARM MCUs

**WHAT IS A RASPBERRY PI ?**

**It is a mini-computer**. In fact, it was invented for Poor People or students who cannot afford a computer but still want learn and practice programming or basic computer functions. It has ***ports***which allows it to be connected to a keyboard, a mouse and a desktop screen (where you can see the output of your code).

The Raspberry Pi itself is an embedded computer, or also an SBC (single board computer). **The *processor*on the Raspberry Pi is somewhat of a hybrid between a microprocessor and microcontroller. It is actually a SoC (system on chip)**, **containing multiple dies stacked on top of each other**, with a Broadcom BCM 2835/6/7 ARM CPU, a Broadcom VideoCore GPU (graphics processing unit) and RAM all tied together.



SOME USE CASE OF Raspberry Pi:

