**TOPS”** — **Trillion operations per second**

It tells you how many computing operations an AI accelerator can handle in one second at 100% utilization. In essence, it looks at how many math operation problems an accelerator can solve in a very short period of time. TOPS = (number of MAC units) x (frequency of MAC operations) x 2.

**NPU, GPU & CPU**

The **CPU** is responsible for smooth switching of mobile applications, the **GPU** supports fast loading of game screens, and the **NPU** is specifically responsible for the implementation of AI computing and AI applications.

**What is Nano-meter “nm” technology in CPUs?**

To build a CPU first step is its architecture. CPU (processor) has billions and trillions of transistors. Transistor gate length is in nanometer or we can say size of transistor is in nanometer. E.g. 7nm, 10nm, 12nm, 14nm, 16nm, 28nm. Less the transistor size in “nm” more the powerful, efficient and better processing. If we have tiny transistor then we can pack more transistors in the CPU and more transistor means strong CPU performance. Other reason is, if transistor is tiny electrons don’t need to travel far, means it has less electric resistance. Small transistor also consume less energy or power. Small transistor also increase power efficiency. Lithography process in CPU has unit in nm. Some processors has 8nm lithography process.