

What is NPU?

With NPUs supporting on-device AI technology across SoC (System on Chip) designs used in laptops, smartphones, and single-board PCs, it is important to understand and compare specifications, their techniques to optimize performance and make informed purchasing decisions.

With NPUs supporting on-device AI technology across SoC (System on Chip) designs used in laptops, smartphones, and single-board PCs, it is important to understand and compare their specifications, to optimize performance and make informed purchasing decisions.

What is NPU?

Neural Processing Unit (NPU) is a special type of processor that handles tasks related to artificial intelligence (AI) and Machine Learning. In devices such as smartphones and laptops, NPUs perform processor-intensive tasks such as image processing, real-time video filters, voice recognition, augmented reality, and object recognition, much more effectively.

You can find NPUs on some of the most popular handsets today. For example, Apple's latest iPhone and M-Series laptops are both integrated with their own NPU called ANE (Apple Neural Engine). Google uses a proprietary NPU called TPU (Tensor Processing Unit) on its Pixel devices and Coral prototyping kit, and Qualcomm is integrating ARM NPUs into its latest SoCs for smartphones and laptops.

NPUs are becoming more popular as we use AI on many devices, from smartphones to large data centers. As AI develops, we need hardware that can handle it better, and NPUs are the solution.

How is NPU different from CPU and GPU?

For modern handsets, wearables, and other mobile devices, the NPU, CPU, and GPU are integrated into a single SoC for faster data transfer, lower power consumption, and less footprint, more cumulative.

1. CPU is a general-purpose processor to run software.
2. GPUs are specialized parallel processors that assist the CPU in running graphics-intensive tasks.
3. The NPU is an even more specialized processor, supporting the GPU and CPU in Machine Learning and neural network processing tasks such as object detection, speech recognition, and AI processing.

Basically, the CPU is capable of handling many different applications well but is not optimized for graphics and AI tasks. GPUs are better for graphics and AI tasks than CPUs because they can multitask. NPUs excel at high-level parallel calculations like GPUs but with lower power requirements, making them ideal for bringing the demanding computations of AI processing to mobile devices.

How to compare NPU specifications

When comparing NPUs, it is important to understand the key specifications that affect their performance. Here are some of the most important specs to pay attention to:

1. **TOPS (Tera Operations Per Second)** : This is a measure of how many trillions of operations the NPU can perform per second. Higher TOPS values usually indicate better performance.
2. **Power efficiency** : Measured in TOPS per watt, this specification indicates how efficient the NPU is in terms of power consumption. Higher performance means longer battery life for mobile devices.
3. **Precision** : The NPU can handle different levels of precision, such as 8-bit, 16-bit, and 32-bit operations. Lower precision can speed up computation and save energy but can reduce accuracy. For many AI tasks, 8-bit precision is sufficient.
4. **Memory bandwidth** : This measures the amount of data the NPU can read or write per second. Higher memory bandwidth can help process large data sets faster.
5. **Supported frameworks** : NPU must support popular AI frameworks such as TensorFlow, PyTorch, and Caffe. This ensures that developers can easily deploy their AI models on the NPU.

But like most computer hardware, the higher the specs the better. A 40 TOPS capable NPU like Qualcomm's Snapdragon X Elite chip will have better AI capabilities than a 10 TOPS capable NPU like Intel's Meteor Lake chip.

How will NPU make a laptop or PC better?

Buying a laptop or PC with an NPU has several key benefits:

1. **Enhanced performance for AI tasks** : The NPU is specifically designed to handle AI tasks efficiently, meaning faster performance for applications that rely on Machine Learning and artificial neural networks.
2. **Better battery life** : The NPU is optimized for low power consumption, which can help extend the battery life of laptops and mobile devices.
3. **Improved user experience** : Features like real-time language translation, voice assistants, augmented reality apps, and advanced image processing can run more smoothly and efficiently with NPUs.
4. **Offloading work from the CPU/GPU** : By handling AI-specific tasks, the NPU frees up the CPU and GPU to perform other functions, leading to better overall system performance.
5. **Future-proofing** : As AI becomes more integrated into everyday applications, having an NPU ensures that the device is ready for future technology and software advancements.

NPUs are changing the way AI tasks are handled in modern devices, and their role in computing will only increase as AI becomes an essential part of everyday life. With NPUs becoming more widely adopted, understanding and comparing NPU specifications can help you make informed decisions to optimize performance and future-proof your devices.

You finished reading the article "**What is NPU?**" edited by the [TipsMake](#) team. We hope this article has provided you with many useful tech tips and tricks. You can search for similar articles on tips and guides. Thank you for reading and for following us regularly.