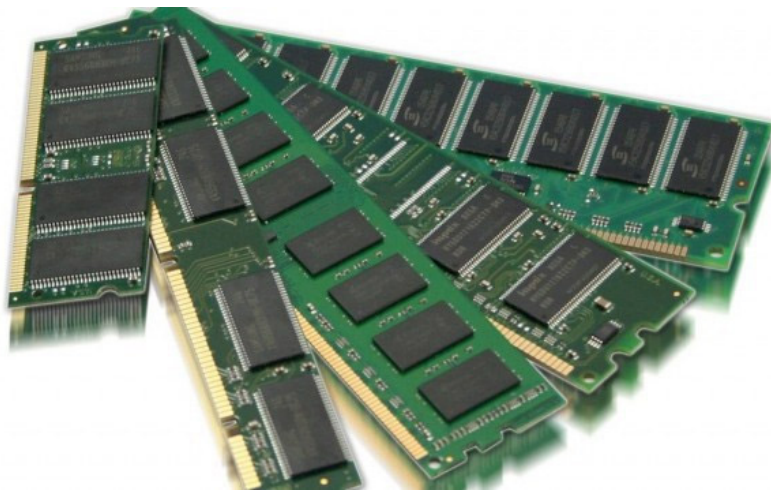


Why aren't CPU and RAM computers packed together to increase processing speed?

Why do computer designers do not put CPU and RAM right on the same die to reduce the distance between them and speed up the processing speed?

The RAM in the computer is responsible for giving data to the CPU to handle tasks. Therefore, CPU processing speed will be limited by the speed of reading and writing data from RAM.

On the computer's motherboard, we will see whether CPU and RAM are placed close to each other but there is a certain distance between them. This distance compared to the size of each transistor on the processor is very large. So why don't computer designers put CPU and RAM right on the same die to reduce the distance between them and speed up the processing speed?



Memory of CPU - cache

In fact, even within the CPU itself, there is a separate RAM area called cache but usually only a few MB capacity. Because it is much more expensive than RAM, it will cost the CPU price higher, depending on its memory capacity.

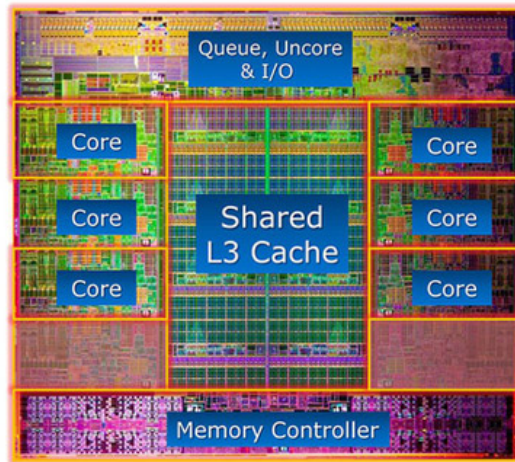
This cache helps the CPU minimize processing time while waiting for data from RAM to be transferred.

When you need to process a data, the CPU will first search the cache first to see if the cache has read / written that value recently. If the cache is available, the CPU will use the value there to get the fastest processing speed. If not, the new CPU finds RAM to get the value it needs.

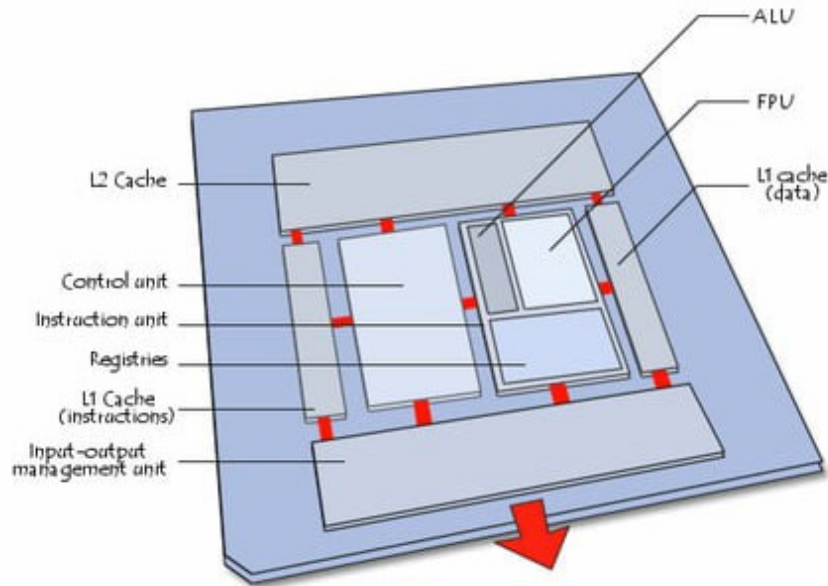
So, in order for the CPU to increase processing speed and simplify the design, why do chip designers not increase the cache capacity to the same amount of RAM?

In fact, for the CPU to achieve the high speed required, its cache uses static RAM or SRAM. However, even though SRAM is very fast, the amount of power it consumes is huge, not to mention it is quite cumbersome.

Intel® Core™ i7-3960X Processor Die Detail



To store a data "bit", SRAM needs 6 transistors. This means that 48 billion transistors are needed, in order to have an SRAM wall of 1 GigaByte capacity (GB). If compared to the number of transistors on the CPU, that would be a terrible number. Even Intel's Broadwell-E 6-core Core i7-3960X processor only has 2.27 billion transistors, and of course its cache is only 15 MegaByte (MB).



A description of the computer processor.

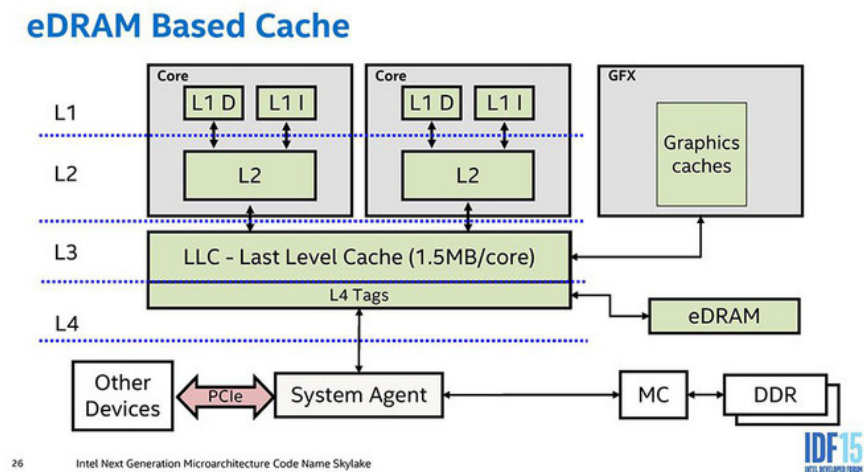
In the picture above, it is not the computer processors but the integrated circuits of the memory that are the most complicated and space-consuming components. So, if we increase the size of the on-chip cache, we won't have room to put the CPU. Of course, with a limited size, it is obvious that the processor cache has a capacity of several MB per chip.

One of the important issues preventing the increase in cache capacity is cost. The cache itself is made from expensive SRAM, increasing capacity will make the processor cost per CPU and cache increase.

Why not use other types of memory cheaper?

Today's computer RAM modules are made from DRAM, which is cheaper and larger in capacity. In addition, DRAM has a neat structure, only need a transistor for each bit of data. The downside of DRAM is that it is significantly slower than SRAM.

So, why do designers not put DRAM in the CPU to increase the speed and memory capacity for the processor? In fact, Intel has applied that architecture but cannot replace the role of regular RAM.

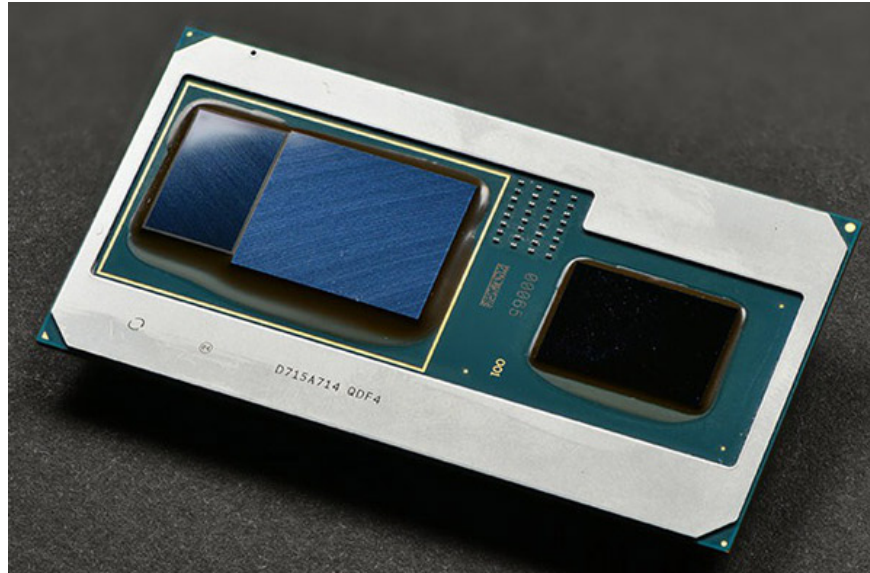


eDRAM's position in Intel's Haswell processor.

Starting from the Haswell generation chip, Intel has integrated a DRAM memory on the same die chip with the CPU, called eDRAM (or Embedded DRAM - embedded DRAM). On the Coffee Lake generation chip, Intel still uses this eDRAM.

However, although the new architecture has the advantage of performance, it still cannot replace both SRAM and conventional DRAM. The reason is speed, because even when integrated right on the die chip next to the CPU, it still has slower clock speed than the CPU clock, while SRAM has the same clock speed as the CPU. In addition, to not lose data DRAM needs to be constantly refreshed thousands of times per cycle, while that is not needed with SRAM. Therefore, the delay of DRAM or eDRAM is still much larger than that of SRAM and they are only used as L3 or L4 cache next to the CPU.

Currently, eDRAM on CPUs is only 128 MB, while the RAM bars now have capacity up to GigaByte. The huge disparity in capacity also makes eDRAM irreplaceable DRAM.



Intel Kaby Lake G Core i7-8705G is designed with processor with CPU, GPU and RAM attached on a die chip.

Until recently launched Kaby Lake G chips, Intel has just begun to use processors with CPU, GPU and large amount of RAM placed on a single die chip but still have the usual RAM .

Also, different ways of doing things are one of the reasons eDRAM is not a substitute for DRAM. On cache, the data is addressed the same and can be deleted at any time. In RAM, data is placed in separate locations for each application.

In addition, packing RAM and CPU together will make it difficult to upgrade. People may be forced to buy the CPU if they want to increase the amount of RAM, which causes unnecessary waste.

Overall, packing computer CPU and RAM into a single block provides less practical value than the current separation. The CPU itself has integrated many different types of internal memory that can help it increase processing speed significantly.

You finished reading the article "**Why aren't CPU and RAM computers packed together to increase processing speed?**" 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.