The Economist

September 16, 2024

How to build more powerful chips without frying the data centre

Runaway energy consumption remains a problem

The Blackwell chip from Nvidia, shovel-maker for the artificial-intelligence (AI) gold rush, contains 208bn transistors spread over two "dies", pieces of silicon each about 800 square millimetres in area, that house the processor circuitry. The two dies are linked by a blazing 10 terabytes (ie, ten thousand gigabytes) per second chip-to-chip connection. Each die is flanked by four blocks of high-bandwidth memory (HBM) chips that together store 192 gigabytes of data. The advanced packaging methods used to build this megachip are now in the spotlight, with some speculating that they may lead to production delays.

Blackwell's putting together of different parts show the limits to which chipmakers must push their technology to boost computing power while keeping energy consumption in check. Over the course of a year one of these megachips, which cost $70,000, will consume 5.2MWhrs—about half the energy of an average American household. Adding more transistors is the best way to boost a chip's processing power: communication within a single chip may be a thousand times faster and use a hundred times less energy than shuttling data between chips. But since Dennard scaling hit a wall in the mid-2000s, shrinking transistors has not significantly improved energy efficiency. Gordon Moore suggested two other tricks to pack in more transistors: increase the die size (ie, make chips bigger) and use "device and circuit cleverness". In 1971 the 4004, an Intel processor, had a die size of 12 square mm. Current lithographic tools cannot build chips bigger than 800 square mm, about the size of each Blackwell die. Circuit cleverness is the only other path.

One clever idea is to use the chip area more efficiently. In profile, a semiconductor chip is like a multi-layer cake, with the layers which make up the transistors at the bottom and 10-20 layers of metal wiring stacked on top. Leading-edge chips can use almost 100km of tiny metal wires (the sum of connections between billions of transistors). This turns the chip into a crowded highway of data and power lines. At the top layer are thick metal lines that distribute power across the chip. Since these beefy power lines burrow down from the top layer all the way to the bottom layer to connect to the transistors, they can hog nearly a fifth of the area used for routing electrical signals.

Baking the cake

To create more space, leading foundries are now working on "backside power delivery" which moves the thick power lines to a layer below the transistors, known as the "back" side of the chip. This has three benefits. First, it frees up space, allowing all the area above the transistors to be used for routing data. With less congestion in the metal wires, groups of transistors, called gates, can be packed more tightly in the same space. Second, the power lines below the transistors need shorter connections to link up with them, and those connections can be thicker, which makes power delivery more energy efficient. Finally, these robust power lines are less susceptible to voltage spikes during demand surges, ensuring faster switching for the transistors. The result is a faster, more power-efficient chip without having to shrink the transistor size.

But moving the power wires below the transistor introduces additional steps in the manufacturing process. Normally, once the multi-layer cake of the chip has been built, the wafer is flipped and enclosed in a package that connects internal circuits to the outside world. With backside power delivery, the flipping has to be followed by polishing and the addition of the power lines. The additional steps are worth the hassle. In 2023 Intel used backside power delivery and the smarter placement of components which it allows to get improvements of 6% in transistor speed and 10% in packing density with components no smaller than the ones it was using before.

Another tactic is not to cram all functions onto a single chip, but to pick the best manufacturing technology for each task. Though the whizzy processors need the most advanced nodes, other parts of the chip, like the modules that communicate with the outside world, don't need the tiniest transistors. Breaking a chip into smaller blocks called "chiplets", and then packaging them together, lets chipmakers use the maximum area on a die for the processing units, because the rest of the circuitry is moved onto other chiplets. Communication between chiplets in the same package is much faster and uses less energy than with circuitry outside the package.

Packaging was long the poor relation of chipmaking. Foundries focused on producing wafers of silicon. But with the rise of chiplets, packaging is in the spotlight. Chiplets are assembled by placing multiple dies from different process technologies side by side on a layer called the interposer. These dies are then bonded to the interposer with "microbumps".

The interposer, which sits on a substrate typically made of organic resin, acts as a bridge, connecting the dies to each other with high-density wiring, and connecting them to the outside world through the package. This allows for fast data transfer between chiplets and improves pow-

er efficiency. Nvidia's Blackwell processor uses TSMC's version of this to combine its two dies and eight HBM chips into a superchip. TSMC plans to develop interposers six times the size of the largest dies, to host multiple processing chiplets and stacks of HBM.

Intel, meanwhile, has unveiled plans to ditch resin and stack its chiplets on glass instead. The ultra-flatness of glass is better for fine-pitch, high-density wiring and it has better thermal and mechanical stability, especially at larger sizes. The firm says this switch could increase connection density ten-fold over organic interposers.

The next big leap in chiplet technology is stacking dies directly on top of each other, slashing the distance between them. Memory-makers have been early pioneers in such 3d stacking. HBM, used in specialist AI chips, typically stacks eight to 12 memory chips connected by high-capacity routing lines, boosting bandwidth between the memory and processor.

Now AI chips are following suit. 3D packaging can provide 10,000 connections per square millimetre, compared with 25 for side-by-side packaging. More connections means smoother data traffic between chiplets. It is also more energy efficient, using less than 1% of the energy of the previous version to move each bit. The MI300X, a competitor to Nvidia's H100 made by AMD, another chip designer, stacks eight accelerator chiplets on four interposer dies, along with eight stacks of HBM, in one package.

These gains come at a cost. Samuel Naffziger of AMD notes that, because multiple chips need to be tested before being combined into a single package, 3d packaging adds time and complexity to the manufacturing process. Packing chiplets together also increases the heat density in the chip. To facilitate heat dissipation, chip designers locate the layers that make the most heat at the top of the stack, and place less heat-generating components, such as memory, below.

Training AI models requires huge systems in which hundreds of processors are linked together. Even after cramming chiplets into packages, vast amounts of data still need to zip between these separate processors and their connected memory chips. Copper wires, the usual method for connecting components on a motherboard, are sluggish and waste energy. So some firms are turning to light to speed things up.

Fibre-optic cables are the backbone of the internet, carrying 99% of intercontinental internet traffic. These cables are also used over short distances to connect racks in data centres. In both cases, equipment at each end of the fibre turns signals from electrical pulses into light and vice versa. Now optical communication is making its way to silicon as well. Ayar Labs, an American startup, is among the firms that is building a chiplet offering optical communication between processors. Its chiplets sit on the edge of a package and turn electrical signals into light, which is then sent through the fibre. At the receiving end, another set of chiplets converts light back into electrical signals and feeds data to processors in the package. The firm claims this improves chip-to-chip bandwidth up to tenfold with eight times more power efficiency.

Suburbs and skyscrapers

The ultimate leap in energy-efficient, high-speed chip-to-chip communication would be to collapse all the chiplets into one single chip, multiple layers of processors, memory and sensors. This packs even denser connections between different parts of the chip. Subhasish Mitra of Stanford University likens this to moving from a sprawling suburban layout to a towering skyscraper.

Just as a skyscraper has lifts shuttling people up and down, this megachip would need millions of connections whisking data between layers of compute and memory. These lifts could also help keep the chip cool, channelling heat to the bottom layers. Eventually, chipmakers could cluster these chip towers side by side. Mr Mitra believes that this could boost energy efficiency as much as a thousand-fold.

Making such a chip would be hard. Fabrication temperatures for conventional transistors can exceed 1,000°C. With transistors located in multiple layers, the metal lines criss-crossing the skyscraper transistors would melt. To realise the vertical-chip vision, the logic and memory technologies on upper layers must be fabricated at temperatures below 400°C. Newer transistor technologies like carbon nanotubes and 2d materials, which can be processed below 400°C, might be better suited to this than silicon. Mr Mitra has demonstrated a version of this chip on a 90nm node, built with layers of carbon nanotubes and memory.

Speaking in 2015, Gordon Moore admitted he was "amazed" by how long his prediction had held. But he felt that extending his namesake law for a few more decades would require "a lot of good engineering". TSMC believes that a combination of these approaches could yield a trillion transistors on a chip by 2030. Chipmakers are clearly not done yet. ■