Breaking Moore's Law: Pushing PC's to Blistering New Levels

Saketh Kasibatla

June 7, 2013

The rate of increase of computer performance is not what it was a few years ago. Moore's law, which states that the number of transistors on a chip doubles every eighteen months, is breaking down, as the size of the transistor approaches the atomic scale. This breakdown of Moore's law has huge implications for the computing industry as much of the incredible progress and new developments we have seen are due to the breakneck speed of improvement of the computing power of processors over the years. Without Moore's law, we would not have smartphones or mobile computing. We would not be able to solve a large number of the scientific problems we have seen incredible progress in, such as the sequencing of the human genome, or the experiments being conducted at CERN using the LHC. Thus, chipmaking companies such as Intel, AMD, and NVIDIA are using new and varied techniques to maintain the blistering upward trend in performance that we have seen in the past decades.

In order to increase the processing speeds, large changes have been made to the main component of modern processors - the transistor. Under CMOS (Common Metal-Oxide Semiconductor), a process for constructing chips, a transistor consists of a source, a drain and a gate. The source and drain are made of doped silicon, or silicon with impurities added to make it conduct well. The two terminals are separated by a section of undoped silicon, on which rests the gate, which is made of metal separated by an efficient (high-k) insulator. When a large enough voltage is applied to the gate, electrons flow freely through the insulator and create a channel for electrons to flow from the source to the drain[6] (Fig. 1). When designing transistors, the designer wants maximum current flow when the transistor is on, and zero

current flow when it is off. Any current that flows while the transistor is off is considered leakage. Transistors must also switch quickly between on and off[2].

The problems with Moore's law have become more and more evident as transistors have shrunk from 45 nm (in Intel's Penryn chipset) to 22 nm (in Intel's Ivy Bridge). Leakage and idle consumption of power has grown due to the quantum size of each transistor. Some electrons can tunnel their way from source to drain regardless of whether the transistor is on or off. Modern chips use about 50 percent of their power to such leakage when they are idle[4]. This power, let out as heat, can melt the chip if transistors are made too small. To shrink the transistors, Intel has introduced several new fabrication methods and transistor designs to minimize this leakage. In the Penryn chipset, Intel introduced the high-k metal gate to insulate the gate better from the silicon. In the Ivy Bridge chipset, Intel introduced the 3-d transistor (Fig. 2), which has the gate surround the path for electron flow in order to reduce leakage even further[1].

In the face of a disintegrating Moore's law, to keep computers and software speeding up, another approach tho chipmaking that has recently taken hold is parallel computing. Intel and many other CPU makers have encouraged parallel computing with their newer chips which have anywhere from two to eight cores. AMD has pushed data parallellism on the GPU along with task parallelism on the CPU with its Accelerated Processing Units (APU), which contain a multi core CPU and a GPU on the same chip for easy parallelism, useful for graphics, security and much more. Hardware companies such as NVIDIA and Intel are already encouraging developers to take advantage of this new parallel hardware through libraries like OpenCL, CUDA, and OpenMP[1].

The current optimizations in hardware and software that the computing industry has been working to perpetuate, such as 3-d transistors, high-k insulation, and parallel computing should be sufficient to continue Moore's law for a few more years. But, the computing industry will need new ideas to keep pushing the speed of this trend, keep customers buying, and keep improving the computing industry. Currently, the most viable option to increase speed is to change the material of chips from silicon to gallium arsenide. This switch in materials will be relatively easy to execute as the current CMOS fabrication methods can easily be applied to gallium arsenide. According to OPEL technologies, a company experimenting with gallium arsenide, 200 nm transistors made of gallium arsenide can run faster than

the current silicon transistors on half the power, making this switch extremely attractive to companies and the consumer[1].

Another option is to switch materials from silicon to graphene, which conducts electrons extremely well and makes processors much faster than those of silicon. According to research done by IBM, graphene transistors are extremely fast, maxing out at a 155 GHz clock speed, and could be produced quite cheaply using current methods, but are unfortunately currently not very good for processing discrete digital signals, but could be useful for "applications such as networking that require communications at fast speeds and high frequencies"[5].

In the distant future, we could even see the adoption of such ideas as molecular transistors and quantum computing. Today, researchers have been able to create both molecular transistors and quantum bits (qbits), but each has its own problem. With molecular transistors, the problem lies in connecting many transistors as it is currently difficult to produce the contacts needed to connect transistors, as well as the wires to actually connect the contacts, so while we have produced individual molecular transistors, a chip with them is very far off, though, in the future, we may see massively parallel computers containing billions of these transistors[3]. Quantum computers offer a great deal of computing power in a small package, as a qbit is the smallest transistor that it is physically possible to make. These bits offer huge advantages for massively parallel computing as they can probabilistically do all possible versions of a calculation using just a few qbits, making cracking keys and the like extremely simple. The main problem with qbits currently is that the control necessary to make one work requires temperatures close to absolute zero. Both molecular computing and quantum computing hold a great deal of possibility for the distant future, but neither is viable in the near future.

Even though Moore's law is slowing down, and it is becoming more and more difficult to maintain as transistors shrink and the impending power wall looms, the computing industry still has many tools such as newly designed transistors and parallelism to continue Moore's law for a few more years. Beyond that, there are many technologies that are currently being honed such as gallium arsenide and graphene based chips, molecular, and even quantum computers. With these technologies, we can be assured that our computers will continue to become faster, and that the field of computers will continue to grow, unlocking new areas for us to explore.

References

[1] Brad Chacos. Breaking moore's law: How chipmakers are pushing PCs to blistering new levels, April 2013.

[2] channelintel. Video animation: Mark bohr gets small: 22nm explained, May 2011.

[3] Priya Ganapati. First functional molecular transistor comes alive.

[4] N.S. Kim, T. Austin, D. Baauw, T. Mudge, K. Flautner, J.S. Hu, M.J. Irwin, M. Kandemir, and V. Narayanan. Leakage current: Moore's law meets static power. *Computer*, 36(12):68–75, 2003.

[5] Agam Shah. IBM shows smallest, fastest graphene processor, April 2011.

[6] Jr. Watrous, Willis G. MOSFET transistor and method of fabrication, October 1976. U.S. Classification: 438/187; 29/271; 148/DIG.40; 148/DIG.60; 148/DIG.133; 148/DIG.167; 257/E21.433; 257/E29.04; 438/305; 438/978 International Classification: H01L 2122.
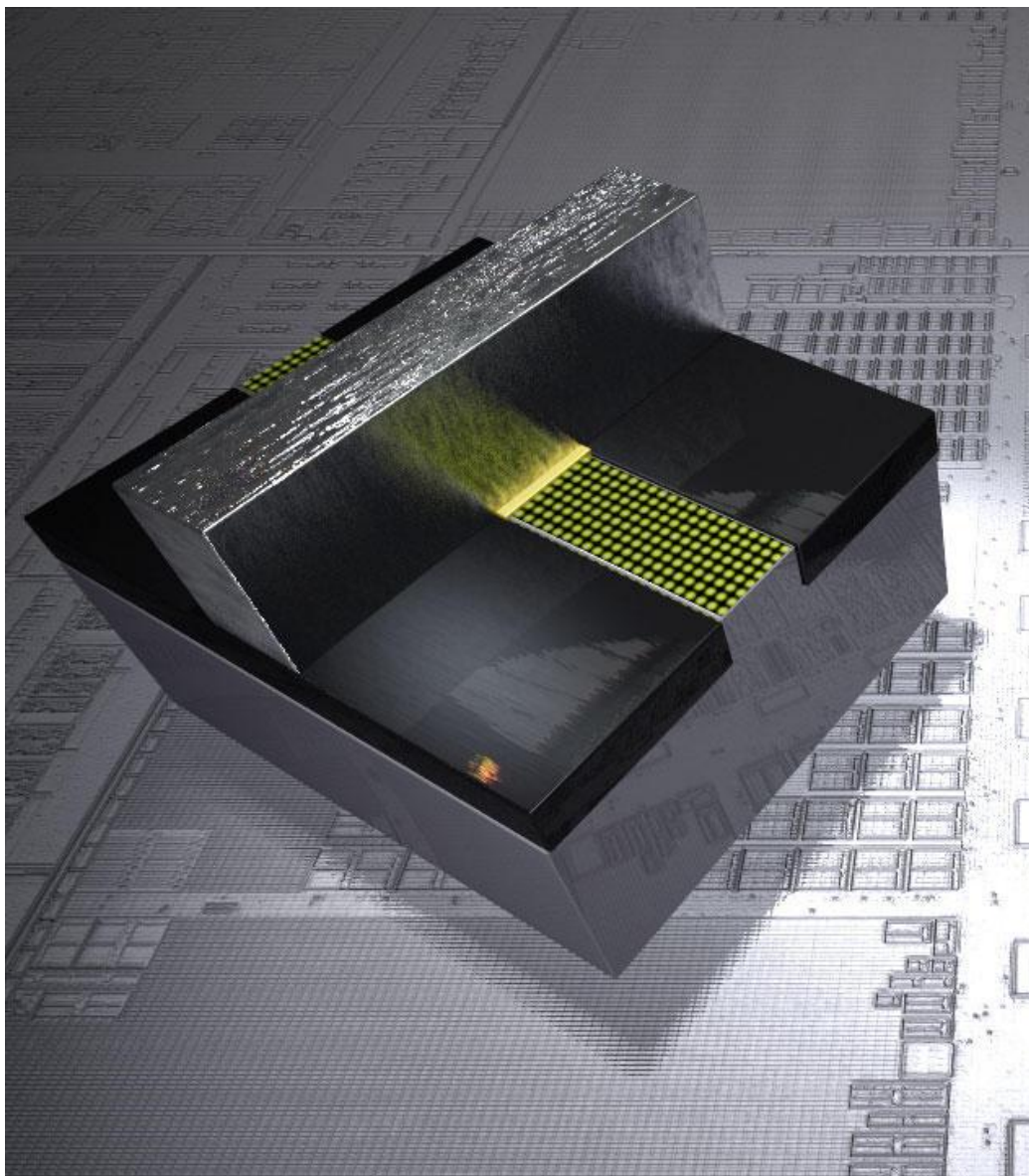
Figure 1: A diagram of a transistor. The yellow electrons flow from source to drain when the gate, sitting in between them, has sufficient voltage applied to it.

Figure 2: A 3-d transistor. The gate surrounds the path for electron flow to reduce leakage.