Future computing

– the power of hybrid architecture

ECE 758 paper reading
Student: Li Tang (217-70-152)
Professor: D. Misra

Views for future computing [1]

Famous Moore's Law makes a great contribution to the development of microchips. However as the limit of cmos technology is becoming the major obstacle for improving computer performance, more and more novel methods are proposed to address this problem.

Novel architectures for assembling components and ideal nano-meterials are both very suitable for the candidate solutions. In the other hand, new schemes for processing information such as quantum computing and bio-computing are also potential ways to address the problem of cmos limits.

As conventional solution, in other words the novel architecture, multi-core architecture based CPUs have been commercially produced such as INTEL i7 (4-core) and AMD Opteron (6-core). However the frequency of each core in the CPUs is nearly the same with the past ones. So the inter-communication plays a very important role for improving performance. Interconnection, cache coherence and multi-core's software support are all very urgent problems for fully exerting multi-core architecture.

It seems that nano-materials will help researchers to solve the problem. Nano-scale devices could be made much smaller which means bringing us much faster switching speed. Obviously in the same size chip more components and wires could be placed although some of them may be not reliable. So the prerequisite is finding a rational way to use nano-scale devices to provide reliable computing.

At last, quantum computing, optical computing and bio-computing are all beyond conventional computing paradigm. They change the styles of coding, storage and processing which indicates they will not come without a low cost. In addition the computing models are also very immature. All problems make these novel computing technologies will have a long way to go.

Challenges in nano-computing [2, 3]

Current nano-technology is not mature enough to produce reliable devices to replace all the conventional cmos components. In the other hand the lack of related prototypes, tools and evaluation models are all obstacles to use nano-computing. Therefore the researches are extremely necessary to be executed for solving these problems.

Firstly, completely new nano-chip without traditional components is not a wise way to access nano-computing. While a more gradually method - hybrid architecture that combines nano-scale components and micro-scale components may be more suitable.

Another problem is how to establish a reliable computing system through connecting unreliable components. With the size of components comes to nano-scale there is a percentage of the components will not work rightly. Therefore the key problem for the combination is finding a rational architecture to address the interconnection for improving reliability. Additionally temporal and hardware redundancy is very necessary to keep the devices work correctly.

In the high level, the engineers will also face many challenges to use nano-computing. Particularly the interfaces between different levels are very different with the current design. Understanding requirements among different groups before developing a well-defined interface is deeply needed. Further, a new design style bottom-up will be used which is very different from the methods today. Related CAD tools could not support the design, verification and optimization too much. At last, due to the lack of real nano-products and reasonable models, the engineers could not

estimate the parameters such as power and error-rate before the fabrication very precisely and understand some behaviors of nano-devices.

Hybrid architecture for higher reliability [4]

When the component size scales to less than 10 nanometers, the operation of transistors will be highly problematic. In the other hand, as the decreasing of the size, much more components will be contained in the same area chip. So it is very necessary to develop a hierarchy to connect the huge unreliable components to let the whole work rightly.

Crossbar is an easy way to address this problem, because crossbar could be implemented with a low cost and have a comparably high bandwidth. In conventional network applications and even the microprocessor such as SUN UltraSPARC T1, the crossbar was used to connect the components. However another problem also needs to be solved. That is keeping all the nano-scale components work well should cost too much, so about 10% potentially disable ones of all the components will be nonfunctioning during the manufacturing process. This process with the name nano-imprint lithography (NIL) could make the cost more reasonable and the nano-wires to reach 15nm wide.

Another vital technique is to use a thin layer of switching material and a second layer of nano-wires that formed on top of the switching layer by NIL. Chemical etching makes every top layer nano-wire is connected to a bottom layer nano-wire through a switch, then two states of open and closed could be set.

Regarding the case of memories, fault tolerant could be reached by software techniques from coding theory. In addition some errors could be masked by the inherent features of computer architecture which could be called soft errors. So the errors also could be avoided by the computer architecture, software or coding methods.

Another key point is to utilize a hybrid architecture to combine nano-scale and micro-scale components for higher density in the chip. Hybrid architecture field-programmable nano-wire interconnect (FPNI) [5–7] is the combination of

nano-scale interconnection and micro-scale function. Since nano-scale components are much smaller than the configurable resource which consumes a major part of the area in conventional FPGA, much higher density is achieved. While the traditional functional units implemented by using ASIC process is more reliable, therefore the combination of both two could make a reasonable balance between the performance and the fault-tolerant.

FPGA is a very good example of the hybrid architecture for leveraging the flexibility and performance. Take the Xilinx Virtex series FPGA for an example, it puts ASIC modules such as PowerPC, Block RAM and High Speed I/O inside the FPGA. They could be connected with the logic resource in the FPGA by using the routing resource. Therefore the design could feature high performance and low power by combining coarse-grained ASIC modules such as PowerPC and the flexible and high fault-tolerant FPGA resource. What we need to do during the design is just to instantiate the ASIC modules with a controlling wrapper in the Xilinx tools and then to utilize a rational interconnection to complete the final design.

ASIC modules in FPGA which could also be considered as the coarse-grained modules or conventional cmos components in hybrid cmos-nano devices. However the flexible FPGA logic resource could be dynamically reconfigured into new functional units without interrupting the running process. Additionally, by using the logic resource some redundant modules could be generated to keep the results correctly. The most abstractive feature of FPGA is the dynamic reconfiguration which could replace the faulty modules by only closing this module and reconfiguring this module with the same logic resource. Before benefitting from this hybrid architecture, the researcher must solve some problems such as interconnection, tool support and design process.

My opinions

As to me, several potential technologies may address the reliability problems. The first aspect is the hybrid architecture. I agree with the way of the combination of

different scales components very much for improving the reliability. In many fields with high requirements of reliability such as space applications, FPGA is the very ideal choice for implementing the function. Because the FPGA could easily implement redundant modules, these modules could also be used to check if the results are right. They could also be used as the backup components. Furthermore, FPGA could be reconfigured by downloading the configuration file which also means the engineer could modify the configuration file to upgrade functions or improve reliability. Each coin has two sides, for making up the lower performance for configurable logic, some pre-manufacturing ASIC modules such as PowerPC have to be embedded inside the FPGA. I think hybrid architecture could be pushed furthermore. In other words, three different scales components could be combined into one chip which is showed in figure 1. The ASIC modules with high performance and the configurable logic with high reliable both help nano-scale wires to reach higher density with fault-tolerant and high performance.

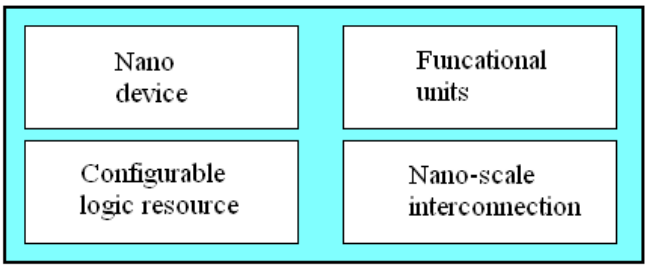


Figure 1. Components with different scales

Another trend is the 3D integrated circuit technique, decreasing size of the components and wires could make the chip have several layers with functional units. Different layers may use different processes to place specific components, and then vertical wires and horizontal wires are used to connect inter-layers and inter-components. This architecture could fully utilize the nano-scale advantages to obtain higher density. The same kind of components such as in terms of size are placed in the same layer could effectively avoid the faults caused by different sized components. The architecture is showed in figure 2.

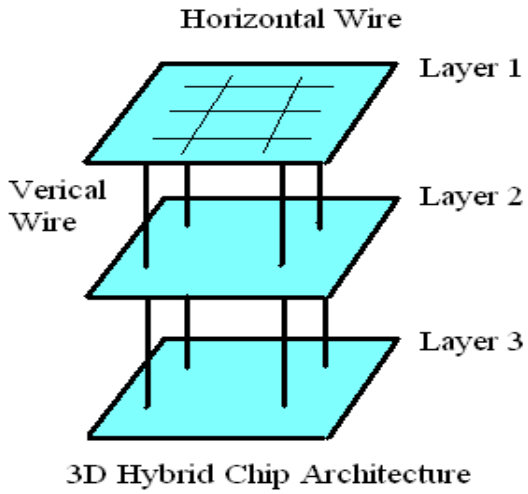


Figure 2. 3D hybrid architecture

Lastly, the support of design tools would be also very important for the use of nano-computing. The design process needs a solid support from the tool chains. For obtaining higher performance and lower error rate the CAD tools also should provide precisely evaluation models, effectively optimization and correct verification.

References

[1] The Next 20 Years of Microchips. SCIENTIFIC AMERICAN, January 2010, 83-89.

[2] R. Iris Bahar, Dan Hammerstrom,et al. Architectures for Silicon Nanoelectronics and Beyond. Computer. January 2007, 25-33.

[3] TOSHINORI MUNAKATA. BEYOND SILICON: NEW COMPUTING PARADIGMS. COMMUNICATIONS OF THE ACM. Vol. 50, No. 9. September 2007.

[4] Warren Robinett, Gregory S. Snider, et al., COMPUTING WITH A TRILLION CRUMMY COMPONENTS. COMMUNICATIONS OF THE ACM. Vol. 50, No. 9. September 2007.

[5] Snider, G.S. and Williams, R.S. Nano/CMOS architectures using field-programmable nanowire interconnect. Nanotechnology, 2006.

[6] Strukov, D.B. and Likharev, K.K. CMOL FPGA: A cell-based, reconfigurable architecture for hybrid digital circuits using two-terminal nanodevices. Nanotechnology; preprint available at rsfq1.physics.sunysb.edu/likharev/nano/FPGA05.pdf.

[7] Strukov, D.B. and Likharev, K.K. A reconfigurable architecture for hybrid CMOS/nanodevice circuits. In Proceedings of FCCM'05 (Napa Valley, CA, April 2005); preprint available at rsfq1.physics.sunysb.edu/likharev/nano/FCCM2005.pdf.