

Review Paper of 21st Century Computer Architecture

Chao Chen

1. ABSTRACT

This review paper gives a brief summary and some review points of the paper 21st Century Computer Architecture. The following chapter 2 introduces this article in brief. The summary and review of this topic are addressed through different aspects of computer architecture research include new technologies, emerging requirements, cross-cutting interfaces and energy efficiency in chapters 3 and 4. The chapter 5 concludes this topic and the chapter 6 lists some references adopted by this article.

2. INTRODUCTION

21st Century Computer Architecture was released as a white paper by Computing Community Consortium (CCC) at May 25, 2012. The object of the paper is to give some guidance of research direction on the computer architecture. The targets of computer architecture research are performance optimizing and energy efficiency. In this article, the challenges and new opportunities of the semi-conductor industry are stated in the first chapter: instruction and summary, and the following chapter introduces four different research fields of computer architecture. The last chapter makes a conclusion to summary the points of this paper.

The review and thoughts about this topic, from my own understanding, are introduced in the following chapters.

3. SUMMARY

3.1 CHALLENGES AND OPPORTUNITIES

In the first chapter of the paper, the two aspects that lead development of Information and Communication Technology (ICT) are introduced: semiconductor technology and computer architecture. The main technologies are also mentioned: the Moore's Law does not work so better as before; the end of Dennard Scaling and difficulty in hiding unreliability of transistor as before; the cost which become more expensive on communication than on computation in many new applications and the shortcoming of non-recurring engineering of semiconductor which make R&D make the cost become expensive to amortize. With CMOS technology scaling weakening as an enabler of ICT innovation, computer architects must step up their role even further. So computer architecture should steer its directions to: Architecture as infrastructure-from sensors to clouds, energy first and new technologies in semi-conductors. Another is new cross-cutting implication, which should break current layer with new interfaces.

3.2 RESEARCH DIRECTIONS

3.2.1 ARCHITECTURE AS INFRASTRUCTURE: FROM SPANNING SENSORS TO CLOUDS

From the Appendix A of the paper, the author conclude that the new emerging applications demands a rich ecosystem, which enables ubiquitous embedded sensing/compute devices feed data to warehouse facilities and cloud servers, where such massive data was processed and return to edge devices. So the new architecture research should focus on this cross-environment co-design, which includes: smart sensing and computing, portable edge devices and the infrastructure-cloud servers.

3.2.2 ENERGY FIRST

From the Appendix A of the paper, the author conclude that the new emerging applications demands a rich ecosystem, which enables ubiquitous embedded sensing/compute devices feed data to warehouse facilities

and cloud servers, where such massive data was processed and return to edge devices. So the new architecture research should focus on this cross-environment co-design, which includes: smart sensing and computing, portable edge devices and the infrastructure-cloud servers.

3.2.3 TECHNOLOGY IMPACTS ON ARCHITECTURE

From the Appendix A of the paper, the author conclude that the new emerging applications demands a rich ecosystem, which enables ubiquitous embedded sensing/compute devices feed data to warehouse facilities and cloud servers, where such massive data was processed and return to edge devices. So the new architecture research should focus on this cross-environment co-design, which includes: smart sensing and computing, portable edge devices and the infrastructure-cloud servers.

3.2.4 CROSS-CUTTING ISSUES AND INTERFACES

Besides making computers being faster and more energy efficient, making a computer system better means much more, such as security, programmability, reliability, verifiability and more.

Although traditional interfaces had helped foster decades of evolutionary architectural, they need to be changed at this technology crossroads. What we need in the new ecosystem required by new emerging applications are better interfaces for High-Level Information, for Parallelism, for Abstracting Heterogeneity, for Orchestrating Communication, for security and reliability.

4. DISCUSSION

4.1 EVOLUTION OF HARDWARE

As the most famous law in the ICT industry, Moore's law's prediction is valid for several decades. But with the improvement of clock speed and logic density, some problems became more apparent. Such as power and RC Delay. Intel's former chief architect-Bob Colwell concluded that Moore's law will be dead within a decade. We could also get some information from the processors delivered to market: the speed of CPU frequency was not improve drastically, which is fixed to no more than 4GHz, instead, the IC manufactures emphasize on multi-cores. Without big innovation on semi-conductors, the speed of processor could not improve anymore. This bottleneck gives a good chance for us to re-think our computation model: we should not just focus on improve speed on one processor; in fact, we have some other choice, such as SMP or DMP. But for the traditional programmer, the multi-thread programming is a small challenge. While comparing with performance improvement, it is worth.

In this area, the most urgent problems to solve are new storage technology such as memory/storage stack and chip stacking technology such as 3D integration. Only these new evolutions could give fundamental solution to current problems. Since the R&D current IC is a non-recurring engineering, with the improvement of technology and process level, the cost is tremendous. Once some problems were not found until the IC tape out, it will cause huge loss and even make one company go bankrupt. So the research on new EDA tools is very urgent.

4.2 NEW APPLICATION REQUIREMENTS

Considering the new emerging applications, the most conspicuous thing is that we have entered into on big data era. With the boom of social network and E-commerce in these years, big data sizes are a constantly moving target, as of 2012 ranging from a few dozen terabytes to many petabytes of data in a

single data set. As more and more smart devices are connected, nearly everything could be data source, from the sensors for security in your home to smart watch for health care, from smart Google glass to huge amount of cameras used for smart transportation, the amount of data will increase explosively.

Data is everywhere, but to make it useful for us, we need to classify, process and store it. It is impossible to handle the data by one huge servers or some cluster, then there come the data center. Many big Internet companies have setup their own data centers, there are huge amount servers in these data centers, which consume tremendous energy and need robust data computation model to handle big data. So the main task in this area is to setup one robust and high efficient big data computation model.

Besides the smart sensors and data centers, we also need some portable edge device to control and monitor these applications. These devices include laptop, smartphone, pad or some other monitors. These devices are interacting with human directly, so we should bring human factors into computer design.

4.3 BIG DATA

As a forerunner in big data area, Google had proposed one paper with name of *MapReduce: Simplified Data Processing on Large Clusters and implement it*, which open the gate of programming model of big data. MapReduce is a programming model for processing large data sets with a parallel, distributed algorithm on a cluster. MapReduce is useful in a wide range of applications, including distributed pattern-based searching, distributed sorting, web link-graph reversal, term-vector per host, web access log stats, inverted index construction, document clustering, machine learning, and statistical machine translation. Moreover, the MapReduce model has been adapted to several computing environments like multi-core and many-core systems, desktop grids, volunteer computing environments, dynamic cloud environments, and mobile environments.

At Google, MapReduce was used to completely regenerate Google's index of the World Wide Web. It is a good example of Applications-Focused Architecture. Apache implemented the MapReduce idea in its Hadoop project and release the source codes. Basing on Hadoop project, some other data processing and mining software such as HBase and Hive are also released. Many other big companies, such as Yahoo, Facebook, Intel and Microsoft etc. are all working on Hadoop and try to improve it.

In short, MapReduce is a successful attempt to handle big data. But there are so many issues to handle the big data, as stated in the paper, we need more deep research and attempt on exploiting parallelism to enable future applications.

4.4 ENERGY EFFICIENCY

There has been an enormous increase in the global demand for energy in recent years as a result of industrial development and population growth. Since the early 2000s the demand for energy, especially from liquid fuels, and limits on the rate of fuel production has created such a bottleneck leading to the current energy crisis. Energy consumption in ICT is huge, which could be seen from the so many data centers of big companies, the power supply and cooling system are most cost for these companies. From the small chips to big cluster, the power efficiency is the first problem need to handle, otherwise, the stability and reliability is out of question. The energy problem is connected with everyone; although iPhone is a very cool device, but the battery problem is still nightmare for someone who has frequent calls. In fact, there are so many idle resources but consume energy, and we need to improve our computation model to maximize the effect of every computer and minimize their power consumption. For example, the AWS of Amazon is a good one in this aspect: everyone could request service based on requirement. You need not to know location of the servers and just submit you job, then get your result. The architecture of AWS has several properties stated in the paper: power efficiency, security, reliability and scalability.

5. CONCLUSION

This white paper introduced some most important factors for the next generation architecture. The challenges and opportunities needs huge effort from and will benefit all participators in this industry: government entities who are in charge of policy orientation of, the research institution which focus on new technology and new architecture and the companies who focus on R&D and sales and service of new products. So 'It was the best of times, it was the worst of times', and there is a new era waiting for us.

In the past several decades, the technology got more attention in ICT. We have to admit, without the development of technology, we could not make such wonderful products and get such achievement. But one great company name Apple, give us some better products such as iPod, iPhone and iPad other than technology. We could get some design ideas on Apple's product, which are also stated in the paper: such devices motivate ideas that bring human factors to computer design, security and programmability and reliability, and the most important thing is better human-machine interfaces. In brief, we should pay more focus on human necessary, not just the technology itself.

6. REFERENCE

21st Century Computer Architecture, A community white paper, May 25, 2012

https://en.wikipedia.org/wiki/Computer_architecture

<http://www.extremetech.com/computing/165331-intels-former-chief-architect-moores-law-will-be-dead-within-a-decade>

http://en.wikipedia.org/wiki/Big_data

http://en.wikipedia.org/wiki/Energy_crisis