Trends in Supercomputing

Dean Michael Blasco Ancajas

***Abstract*— Perspectives are given on how supercomputers have evolved thru time. From the vector processors that Seymour Cray popularized to the modern grid computing that uses off-the-shelf processors. The architectures are presented in chronological order. The author then gives insights about the future of the supercomputing industry.**

***Index Terms*— supercomputing trends**

I. Introduction

SUPERCOMPUTERS have played a key role in the advancement [not just] in computer architecture but in science and engineering as a whole. Simply put, a supercomputer is the fastest machine in the world in terms of processing power, a supercomputer could turn a processor-bound problem into an I/O-bound one. Applications requiring the use of supercomputers include military applications,scientific research such as protein folding, weather forecasting, oil exploration and industrial designs–car crash simulation(a cheap alternative to actually banging cars to the walls). Protein simulation is one of the prime applications of IBM when it announced its Blue Gene programme on 1999 and computer simulation is the most promising basis for the study of protein folding although the computational effort required is simply enormous. A petaFLOP(10^{15}) machine, approximately 3.56x the performance of IBM Blue Gene/L–the fastest supercomputer since 2005–could take 3 years just to simulate 100 us of folding [6].

II. It All Started With Seymour Cray

The Cray-1 is not the first vector machine. During the release of Cray-1 there were currently three kinds of vector processors, the Star 100 processor of Control Data Corporation(CDC), the ILLIAC IV of University of Illinois and the Advanced Scientific Computer(ASC) of Texas Instruments(TI). The most powerful at that time was the ILLIAC IV[7]. Vector processors operate on vectors–linear arrays of numbers, a typical vector instruction might add two 64-element, floating-point vectors to obtain a single 64-element vector result. Although not apparent, the vector instruction just presented could represent an entire scalar loop. Compared to scalar code, vector codes could minimize pipelining hazards by as much as 90%. The Cray-1 was built using only four types of chip and it was the first computer to use transistor memory(apart from the usual magnetic core memory with high latency), with this transistor memory, the ratio of vector to scalar execution speed was brought down between 5 and 10(compared to the STAR's value of 256)[8]. This was a technology leap as vector computers will not be that slow when it encountered scalar codes in the program. Aside from the transistor memory feature Cray-1 introduced vector-register architecture to significantly lower start-up overhead of vector functional units and reduce memory bandwidth requirements. This properties combined with a fast scalar processor are one of the reasons Cray-1 was such a success. The Cray-1 was the fastest computer from 1976 to 1981.

III. Massively Parallel Processing

The easing of the cold war in the 1980s saw a decrease in the federal funding of supercomputers received from the US government. And with the advent of diminishing returns in IC design supercomputer companies were forced to look for other alternatives that could replace the uber expensive vector architectures. This alternative would come in the form of using multiple processors in a single computer called Massively Parallel Multiprocessing(MPP). MPP uses many processors with its own memory, running in parallel and linked with high-speed buses in the motherboard. Early examples of MPP are the Cray T3D and the Connection Machines(CM-2 and CM-5) at the University of Minnesota. The programming model of the Connection Machines presents one global memory, hence the memory in each of the processors are shared. Woodward and Co. ran a piecewise parabolic code and found out that the machines seamed to halt when whenever a processor accesses a memory location that it doesn't own. Contrast to Cray T3D which has low latency when accessing inter-processor memory [8].

IV. Cluster Computing

A cheaper approach that has recently surfaced is the use of clusters of computers connected by high-speed network. In the most recent list of the top500.org–a website that lists the top 500 fastest supercomputers in the world–72.80% of the computers are clusters while only 19.8% uses MPP.

Cluster computers are a group of loosely-coupled computers that work together and act as a single computer to solve a particular problem. The biggest thing that happened to supercomputing is the advent of cluster computers, companies such as Dell, IBM and HP are now supplying cluster packages. So even if you have the money, you don't have to build a supercomputer in order to have it, you can just buy one and save lots of money. Let your scientists do their jobs whilst it is the job of the companies to provide you with supercomputing power.

V. USA versus Japan

Aside from the United States of America, Japan is the only country that has a bonafide supercomputing industry. In fact Japan is home of the world's 500 or so vector computers. Japan's makers of supercomputers are NEC Corp., Fujitsu and Hitachi. Japan's claim to fame is the NEC Earth Simulator which ran at a peak performance of 40.96 TeraFLOPS and was the fastest computer from 2002 to 2004. Earth simulator is used to run holistic simulations of global climates in both the atmosphere and the oceans. The NEC Earth Simulator is composed of 640 nodes, each node is composed of 8 vector processors runnning at a peak performance of 8 GigaFLOPS. Simulations show that the sustainable performance is 26.58 TeraFLOPS.

USA reclaimed the lead in supercomputing in 2004 with the SGI Project Columbia. The performance of the said computer is 51.87 TeraFLOPS. IBM then surpassed the Columbia with 70.72 TeraFLOPS of performance. See Table I below for more details on the Top 3 supercomputers. Contrast to the Earth Simulator's building block which is a vector processor, the basic block of the IBM Blue Gene/L is a dual core PowerPC running at a low frequency of 500 MHz. The low frequency of the microchip was chosen so as to minimize the power requirement per processor hence enabling to incorporate more processors in a design. This design decision was in fact a success for IBM, at 1/50 the size and 1/14 the power required by the Earth Simulator, IBM Blue Gene has 6.4x as many processors and is twice as fast.

System	Applications	Builder
Blue Gene/L	Materials Science, Nuclear Stockpiles Simulation	IBM
Columbia	Aerospace Engineering, simulation of space missions, climate research	SGI
Earth Simulator	Atmospheric, oceanic, and earth sciences	NEC

TABLE I

The Top Three Supercomputers

Architecture	Performance	$ M
32768 processors; 8 TBytes of memory; Linux and custom OS	70.72 TFLOPS	100
10240; 20 Tbytes of memory;Linux OS	51.87 TFLOPS	50
5120 processors;10 TB of memory; Unix-based OS	35.86 TFLOPS	350-500

TABLE II

Top Three Supercomputers(continuation)

VI. Ubiquitous Computing

Unless we have a workforce for supercomputers, they will not be ubiquitous. A one of a kind event that occured in April 3 2004 in the University of San Francisco gymnasium is the gathering of a flash mob(a group of strangers organized over the internet) and their attempt to build a supercomputer that will earn a place in the world's top 500 fastest computers, the target speed is 403 GigaFLOPS[4]. They have 669 computers hooked up but one omnipresent problem of cluster computing is the reliability of the interconnects between computers, at the end only 256 computers were included in the cluster and the speed reached was 180 GigaFLOPS in the LinPack benchmark, less than half of what was desired. Although the effort fell short nonetheless most of the people were not dissapointed. Most of the attendees were young geeks who want to see the buzz behind supercomputing. Aside from young geeks, big names in the industry such as Gordon Bell of Digital Equipment Corporation and computer architecture superstar Gene Amdahl was present to witness the event.

VII. Future Developments

I have compiled data from www.Top500.org, a website that lists the Top 500 fastest computers in the world and have plotted them to predict future trends in supercomputing. First, if we assume performance of supercomputers to double every 14 months (as depicted in 1) then the petaFLOP barrier would be surpassed around 2009 to 2010.

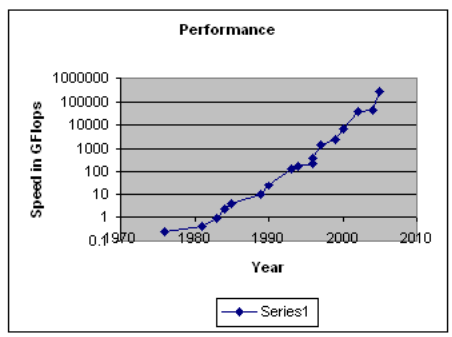

Fig. 1. Logarithmic Graph of Performance

Parallelism

Second observation would be that parallelism(MPP or Cluster) would still remain a mainstream among supercomputers. Parallelism is one of the few ways to cope up with Moore's Law and increase the overall performance of a computer. In the future, performance would be derived by increasing the number of processors in a design just like the Blue Gene/L. In terms of numbers, cluster computing would dominate over MPP since it is cheaper and software people will come up with new ways to make cluster computing more desirable. In terms of speed MPP would still remain in the upper echelon due to its fast interprocessor connectivity.

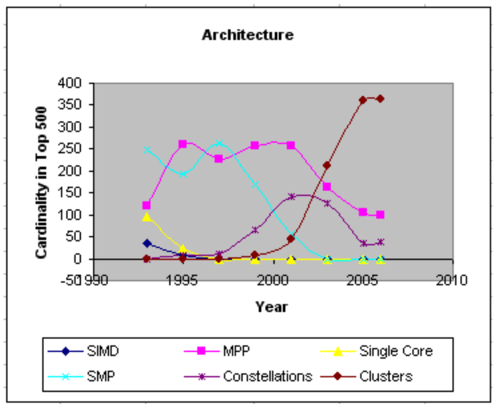

Fig. 2. Architecture

Vector Processors

A few vector processors would still be present but it is unlikely that they will become dominant. Another option would be to include cache memories on vector processors. If studies and simulations could prove an increase in the performance of vector processors with cache memory, then it might be worth retaining the vector parallel programming model.

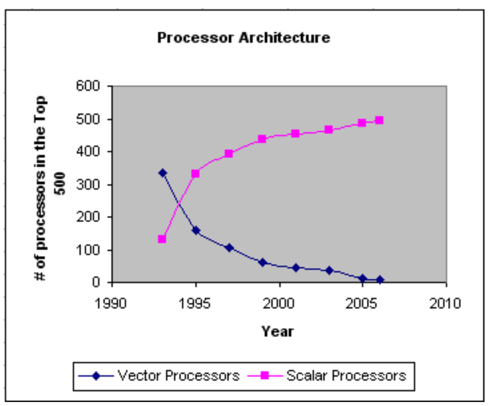

Fig. 3. Processor Architecture

SMP

SMP processors would continue to lie dormant unless it will be incorporated in a cluster. This incorporation would mean additional hardware functionalities. It must also mean that memory bandwidth would also be increased to utilize SMP. This might be a less desirable option for hardware people but can be chosen if there are no other alternatives.

Software Portability

As new hardware designs emerge software people must find ways to increase portability of codes(much like the Fortran compiler of the Cray-1). More parallel programmers will emerge and parallel programming could be a rising division of the software industry. It will only be a matter of time before Universities would make parallel programming a regular subject in the curriculum of Electrical Engineering and Computer Science.

References

[1] Bouknight. W.J. et al. “The ILLIAC IV System” IEEE Invited Paper, Vol. 60, No. 4, April 1972

[2] Kolodzey, J.S. “CRAY-1 Computer Technology” IEEE Transactions on Components, Hybrids, and Manufacturing Technology, Vol CHMT-4, NO. 2, June 1981 pp 181-186

[3] Patterson D., Hennesy J. ”Computer Architecture: A Quantitative Approach” 4th beta edition.

[4] Perry, T. “Building a Supercomputer in a Flash” IEEE Spectrum, June 2004 pp. 24-25

[5] Vaughan-Nichols, S.J. ”New Trends Revive Supercomputing Industry” Computer , vol.37, no.2 pp. 10- 13, Feb 2004

[6] Pulleyblank, W. “How to build a Supercomputer” Special Report Computing pp. 48-52

[7] Richard, R.M. “The Cray-1 Computer System”

[8] Woodward, P. “Perspectives on Supercomputing: Three Decades of Change” October 1996

[9] Yokokawa, M. “Present Status of the development of the Earth Simulator”

PLACE PHOTO HERE

Dean Michael Ancajas is an undergrad student of the Department of Electrical and Electronics Engineering, U of the Philippines, Diliman. Email at dbancajas<at>gmail<dot>com