The Trends of Supercomputers

Mitsuru Maruyama

Technical Institute, Century Research Center Corporation
3-6-2, Nihombashi Honcho, Chuo-ku, Tokyo 103, Japan

Abstract
Since later 1970s, the advent of supercomputers have been realizing scientists and engineers dreams on thier computing needs in thier problem solving. Rapid innovation in computer technologies produces more powerful supercomputers as time goes. In this report, brief history, definition, arcchitecture, characteristics, applications, and the future trends of supercomputers are discussed.

KEY WORDS: (Supercomputer)(Computer Architecture)(Vector Procssing)(Scalar Processing)(Pipeline Computer) (Parallel Computer)(Array Processor)(Vectorization) (Multi-processor)(Multi-tsking)(Mini-supercomputer)

1. A Brief History of Supercomputers

In earlier stages, computers were designed for scientific and commercial usages separately. Since announcement of IBM System/360, these two streams of design concept were combined into a single concept which can handle both scientific and commercial applications. Beside this trends, some computer manufacturers such as Control Data still aimed to produce much powerfull computers in scientific and engineering applications in which floatingpoint arithmetic operations are much important while character handling type of operations are dominant in commercial applications.

In 1964, Control Data delivered CDC 6600 computer under contract with AEC which might be the first supercomputer in the world(it was 3 MIPS machine that is now work station level performance!). It was called "Number cruncher" since it bites and digests huge amount of numerical data. Successor of 6600 was 7600 which had about 15 MIPS performance.

Beside CDCs efforts on high performance scientific and engineering purpose computers, there had been several works in large scale computing field during late 1960s. University of Illinois and Burroughs built Illiac IV which originally designed to have 256 processing elements but finally only 64, for partial diffirential equations solving applications, Texas Instruments developped "Advanced Scientific Computer" for seismic processing application, and also CDC had worked to achieve 100 MIPS performance in thier STAR-100 project.

It should be said that those machines were not commercially successful since the machines shown certain performance on "long vector" operations but on "scalar" operations those were slower than medium scale general purpose computers. In special applications for military radar image processing, some large scale machines such as Goodyear's STARAN and SDC's PEPE were built but those were not placed in commercial market.

It should be noted that there were some failures in the development of supercomputers. Burroughs' BSP(Burroughs Scientific Processor) which has sixteen processing elements and was implemented with charge coupled device(CCD) and ICL's DAP(Distributed Array Processor) which was a array of 4096 single bit micro-processors were challengeable machines from technical viewpoint. Also noted is these two machines were parallel architecture later mentioned.

Cray Research delivered the first Cray-1 computer to Los Alamos National Laboratory in 1976. Difference in the design architecture of Cray-1 compared with former machines was it shown high performance on scalar and "short vector" operations. Cray Research said it is a second generation supercomputer compared with above machines to be called the first generation supercomputers. Cray's success triggered the second round of supercomputer match in the market. In earlier 1980s, new challengers from Japanese computer manufacturers have entered to the match, and now we are seeing the third round in which the third generation supercomputer will appear.

2. Definition of Supercomputer

It may be difficult to define what supercomputer is in short, followings are some characteristics of supercomputers should have:

- high-precision floatingpoint arithmetic operation capability
- more than ten times faster than the fastest general purpose computers (current supercomputers have preformance ranging 100-1000 MIPS/100 MFLOPS -10 GFLOPS)
- vecter or array processing capability to achieve above computational speed
- designed to solve large scale scientific and engineering problems(this may not be true since supercomputers are just penetrating into financial and securities field!)

Dr. Eiichi Goto who is the inventor of Parametron circuitry element mentioned that supercomputer is a computer which is designed to be the fastest computer in the world at the time as the designers intention.

3. Architecture of Supercomputer

Architecture of the current supercomputer is divided into two categories:

- pipeline architecture: designed by an assembly line concept.
- parallel architecture: has a number of processing elements or processors concurrently operate.

According to general categorizing manner of computer architecture, a computer may fall into one of the following four categories;

- SISD(Single Instruction Sigle Data stream), a single instruction stream causes a single data stream in the machine as in general purpose computers and pipeline machines.
- SIMD(Single Instruction Multiple Data stream), a single instruction causes multiple data streams in multiple processing elements as in Illiac IV.
- MISD(Multiple Instruction Single Data stream), multiple instruction streams cause a single data stream. None is found in this category but vector operation chaining in a pipeline machine may show this scheme.
- MIMD(Multiple Instruction Multiple Data stream), in a multi-processor configuration, multiple instruction streams drive multiple data streams in the processors.

In a pipeline computer, the arithmetic and logical units are segmented into several steps and each segmented step process a part of specified operation, then the processed partial result send to the next segment. For example, suppose floatingpoint addition. The addition may be divided into four to six steps of processes; adjust of exponent, shift of mantissa, addition of mantissa, normalization of result, and readjust of expornent, etc. Normally, each step needs one clock period so that operands can be fed in each clock period, result are output each clock period, and virtually four to six additions are processed in the unit. Also devised is to have multiple arithmetic/logical units which independently operate with each other.

From more generalized viewpoint, a pipeline machine is thought it has a pipe which includs arithmetc and logical units from memory to memory, then data continuously flow through the pipe. The efficiency of the pipeline machine is determined how much of degree the pipe is filled with data.

Parallel computers have a number of processing elements or processors depending on whether SIMD or MIMD architecture. Each processing element or processor has own local memory beside shared/grobal memory, and are connected to communicate with each other, mostly with adjacent ones. Mode of the connection may be dynamicaly changed according to problems to be solved. In other words, a parallel computer is an array of processing elements or processors to achieve high performance by the parallelism.

The most of commercially available supercomputers today are pipeline machines, since algorithms of accumulated FORTRAN programs need not be substaintially changed to run on the machines. Parallel architecture is suited some special application such as solution of partial diffirential equations, image data processing, etc., but to run existent progams effectively on parallel machines substaintial amount of works would be needed. Pipeline and parallel architecture are not exclusive so that those may becombined in a single machine. A multi-processor configuration of pipeline machines may be a good example.

4. Scalar and Vector Processing

A scalar operation is defined as an indivisual operation is done on indivisual operand(s) with an instruction, while in a vector operation same repeatitive operations are done on a set of operands with a few number of instructions.

Though supercomputers show thier powerful computing capability on the vector processing, it should be noted that actual jobs consist of scalar and vector operations. When we evaluate performance of a supercomputer for specific jobs, ratio between scalar and vector operatopns become an important factor. Performance improvement utilizing the accelerated vector processing is calculated by following formula:

$$P=\frac{1}{(1-V)+V/a}$$

where: P = performance improvement
V = ratio of vector operation
a = accelerate factor

Processing speed on vector operands also depends on the architecture of the machine; some machines need long vector length to gain the power of the machine while some machines work effectively even on short vectors.

In utilizing vector processing capability of supercomputers, role of compiler is very important. To represent vector operations in programs, subroutine method, language extension, and automatic vectorization are used. Automatic vectorization may be most desirable to avoid machine dependency. Also programming tools for vectorization and program tunning are useful.

From our experience, we categorize application progams as follows:

- programs needed simple conversion for specific machine(if necssary)
- programs needed some tunning works such as DO loop review
- programs needed program re-design and data re-structure
- programs are suitable to scalar machines rather than vector processors

For programs in each category we determined what level of tunning has to be done by considering on the repeatitive frequency of the program.

5. Applications

Applications of supercomputers have been diversified in various fields of science and engineering. Here listed are only some major applications;

- Nuclear engineering
- Weather forecasting
- Weapon R&D/Defense analysis
- Structual analysis
- Seismic analysis/Reservoir simulation
- Aerodynamics/Fluiddynamics
- Circuitry analysis/design
- Computer graphics/Image processing
- Quantum chemistry/Chemical engineering
- Molecular design
- Biomedical research
- Artificial intelligence
- Econometric analysis

It should be noted that the utilization of supercomputers has spreaded from R&D to design phase, now toward actual production phase.

6. The Future Trends of Supercomputers

Because of the limit on propagation speed of pulse in wire, it will be difficult to have much improvement on single processor performance. Multi-processor configuration is one direction of the futue trends. Currently, leading edge systems have eight processors each rating close to one GFLOPS. Alternative is to connect more number(say, 256, 512, or more) of less performance processors(processing elements), which would be just parallel archtecture. Floating Point Systms recently announced a connection of 16,000 processors which would be rating at 262 GFLOPS.

Increasing of the number of processors arise multi-tasking problems on programming, which impose compiler developers and progammers to resolve the problems. The problems are; how to assign the tasks to each processor, how to synchronize the distributed tasks and how to schedule the tasks to let the processors be busy. There are four levels of the multi-tasking; job level, job step level, program level, and micro-tasking. Also parallel machines would ask us following problems; parallelism in application problems, parallel algorithm to solve the problems, and programming language to express the algorithm. Effectiveness of parallel machines depend on the "logical depth" of the applications.

Improvement on circuit speed also will take place by using new elements such as GaAs, HEMT, etc. with VLSI technology, and the recent discovery of high-temperature superconductive materials give us hope of substaintial circuit speed improvement in the future. Circuit boards of ETA-10 of ETA Systems are dipped into liquid nitrogen to inclease the circuit speed even elements are not superconductive.

Cooling of supercomputers may be interesting topics. To gain high speed, physical size of computers must be as smaller as possible to shorten pulse propagation time in the circuits. Power dissipation in the high density circuits, that is comparable to electric range in kitchen, cause serious cooling problems. Air, water, freon gas, and liquid immersion cooling techniques have been devised. The cooling is a source of trouble for supercomputer designers.

Cray Research has a project so called 200X of which target is to realize 200 times faster machine than the current leading edge supercomputer. What 200X faster does mean is; not thinkable problems, i. e., one year+ machine time would be needed to solve them, will become benchmark stage, and large productions which take 10 hours will take only 3 minutes that is just pre/post production range. Jobs currently processed in batch mode will be handled in interactive mode that may much enhance problem solving power of scientists and engineers who are expecting advent of more powerful machines at all times.

Concurrent with computational speed, larger memory space will be needed to handle larger scale problems. Gigabyte range of physical memory size is now commonly implemented. In addition to main memory, external/extended semi-conductor memory should be attached to avoid accesses to slower mechanical storage devices such as disk drive.

Supercomputers have been operated in batch environment having large scale front-end computers. Increasing supercomputer capability and the spread of personal computers and engineering work stations, the environment is moving to interactive. In Apple Computer, hundreds of Macintosh workstations are connected to Cray X-MP via local area networks. Now, UNIX is becoming standard operating system of many supercomputers.

To build an effective supercomputing environment, we have to integrate heterogeneous computers and equipments by network facility as a single system. we call the activities "system integration." Within the integrated system, user friendly man-machine interface is much important while the role of the supercompter is merely "computing server."

The word "Mini" or "Near" supercomputer is popular today. A catch phrase of those computers is that at one tenth of price of supercomputers, one fourth of computationalpower. Distributed processing with network facility became our standard operational environment, in which we handle pre- and post- processing on our desk top work stations interactively and heavy batch processing would be sent to back-end supercomputers. As personal or departmental work stations, the role of mini-supercomputers will be much important in the near future.

7. Conclusions

To respond to needs on numerically intensive computation and to satisfy requirements from scientific and engineering fields, the innovation of the supercomputer will be accelerated. It is beared up by the rapid electronics technology innovation.

The direction of the architecture innovation is toward multi-processors or parallel computers. This will impose us new considerations of the parallel argorithm and parallel programming in our problem solving.

On the other hands, user friendly man-machine interface mechanism should be much emphasised to increase human productivity, based on which integrated systems have to be built around supercomputers.

There will be no end to dreams of scientists and engineers on the supercomputer.