CERN-DD 78-23 add

CM-P00059812

LARGE COMPUTER SYSTEMS AND NEW ARCHITECTURES ADDENDUM

T. BLOCH

CERN, Geneva, Switzerland

This addendum brings the paper referenced up-to-date as of January 1981. The new market situation in the area of super computers is reviewed and the only really new machine, the CDC CYBER 205, is described.

The CRAY-1S series is briefly described and the FPS-164 is mentioned. The fact that Burroughs discontinued the BSP project is mentioned.

1. Introduction

The preceding paper was written late in 1978 and the continued evolution in this field of very large computers justifies a thorough review of the market situation as it has evolved since. A technical analysis of new supercomputers now available is also necessary.

2. The market situation, general purpose computers

At the very top end of the general purpose computers only IBM and IBM-compatible manufacturers have announced new machines. The Amdahl 470/V8 and the Fujitsu M-200 (alias Siemens 7.880) are now available and have been confirmed as being noticeably faster than the IBM 3033. The Hitachi M-200H and variants of it sold by NAS, BASF, Olivetti has also been delivered now both in Japan and in other countries and has proven to be even faster. Yet all of these machines in their single processor versions lie well below the CDC 7600 (and Cyber 176) in pure processing speed. Late in 1980 IBM announced the dual processor 3081 which is scheduled for delivery in late 1981. The basic speed of one CPU in this machine is very close to that of the IBM 3033 but by integrating two processors in the basic configuration and by using new and less expensive technology, IBM now has a very competitive offering in the very large computer market. The throughput

of the 3081 is expected to be comparable to that of the CDC 7600. A more interesting announcement, seen from the perspective of high speed processing, is the announcement from Amdahl of a single processor machine, the Amdahl 580 which matches the IBM 3081 and, presumably, the CDC 7600. The Amdahl 580 is not yet available for benchmarking (first deliveries are for early 1982) but this seems to be the first time that one expects a processor with IBM 370 architecture to match the yardstick speed of the CDC 7600 (the IBM 360-195 had 360 architecture). Since Amdahl machines traditionally do not push floating point performance as much as other important characteristics of their machines, we will still need some serious benchmarks before a proper judgement can be made.

3. The market situation, supercomputers and array processors

In early 1980 CDC announced the Cyber 205, a modern technology and very powerful vector processor based on the STAR-100 architecture. Deliveries are scheduled to start in early 1981. The Cyber 205 has a number of options in terms of memory sizes and number of pipelines but otherwise this is now the only offering from CDC in this field.

In 1979, CRI announced the CRAY-1S series which includes the original CRAY-1's as a subset

but which allows an expansion towards the high end by adding more memory (up to 4 Mwords) and a powerful I/O subsystem based on up to four peripheral processors and a large ($\frac{1}{2}$ –1 Mwords) MOS I/O buffer memory. The CPU remains unchanged. More than 20 CRAY-1's are now installed and the machine has proved to be reliable (over 100 H mean time between failures) and maintainable. Presently CRI manufacture about one system per month.

At the end of 1980 Burroughs announced that they would no longer pursue the BSP project. Although a restricted set of benchmarks were conducted by the middle of 1979 the BSP never made it to delivery probably partly because the CCD memory never worked out and had to be replaced with MOS memory.

Floating Point Systems have announced the replacement of the very successful AP 120B (more than 1000 installed), the FPS-164. The FPS-164 has a more flexible memory architecture, has an increased word length, allows bigger memories and operates with multiple users and interrupts – in short, it is much more autonomous than its predecessor. The basic CPU is unchanged, however, and still runs at 267 ns. FORTRAN will be available and first deliveries are expected early in 1981 (with VAX as host).

ICL delivered the 64×64 DAP to Queen Mary College in the summer of 1980 with one year's delay. More orders are now being made and at least one other DAP is already installed at British National Oil Corporation.

3.1. CDC Cyber 205 (figs. 1–5 and table 1)

The line of development chosen by CDC in the field of very large computers since the STAR 100 has been to update the technology without making fundamental changes to the architectural concepts. An intermediate machine, the Cyber 203, was announced in 1979 which had a bipolar memory and a new scalar processor which also handled instruction issue but which used the same technology as the STAR-100 for the vector unit. One Cyber 203 was sold to the Fleet Numeric Weather Service in Monterey and the NASA Langley Research Center chose to upgrade their STAR-100 to a Cyber 203.

Fig. 1. Cyber 205, CPU overview.

Lawrence Livermore Laboratory decided to scrap their STAR 100's instead of upgrading.

In early 1981 CDC plan to start delivering the new Cyber 205 system. This machine is, basically, a Cyber 203 with a vector processing unit redesigned and implemented in the same LSI technology (168 gates/chip) as the scalar processor of the Cyber 203. Comparing back to the STAR-100 the Cyber 205 has a fast bipolar memory, a new scalar processor and a redesigned vector processing unit allowing 1, 2 or 4 general vector pipelines. The I/O system has also been reviewed extensively to achieve higher effective rates and to take advantage of progress made in the field of high speed bus systems. Altogether this has resulted in a very modern technology machine where, out of approximately 250 instructions only a few tens of instructions are different from those of the STAR-100.

The Cyber 205 has a 20 ns CPU cycle and is

Fig. 2. Cyber 205, vector processor overview.

Fig. 3. Cyber 205, the ADD unit, an example of a pipeline functional unit.

Fig. 4. Cyber 205, instruction buffering.

available with 1, 2 or 4 million (64 bits plus 14 error correction bits) words of 80 ns semiconductor memory (4096 bits/chip). The scalar part of the CPU uses 4 independent pipelined functional units (and one not pipelined) to execute an instruction set based on 3-address instructions working on a register file of 256 general purpose registers. An instruction stack of 8×8 words with prefetch logic is used to fetch and decode scalar and vector instructions. The scalar processor part also deals with the virtual memory addressing system using an associative unit with 16 high speed associative registers as the top part of a virtual page address translation table. One or two 128 bit-wide vector pipelines are each capable of producing two 64 bit results or four 32-bit results (when using 32-bit floating point representation) in one CPU cycle using a three address memory-to-memory order code. This brings the peak result rate to between 100 MFLOPS (64 bit number, 1 double pipe) and 400 MFLOPS (32-bit numbers, 2 double pipes). In certain cases like $c_i = b_i * (s + a_i)$ or $c_i = a_i + s * b_i$ (linked triads) the pipelines are able to produce the complete result at the same speed thus achieving an effective rate of 200 to 800 MFLOPS instead. An entry level system with one 64-bit pipeline is also available which performs at half the speed of the lowest numbers above.

A most important feature of the Cyber 205 is the phenomenal bandwidth of the central memory.

Table 1
CDC Cyber 205

Development started:	mid 1970's, successor to the STAR-100
Delivery:	1st half 1981
orders:	UK Weather Service University of Bochum (Regional Centre)
168 gates/chip	"gate array" LSI
Technology for the CPU	
26 chip types, 50 LSI board types	in 65 total LSI boards
Low junction temperature	(60°C)
80 ns bipolar main memory	
CPU	
20 ns clock period	
Scalar unit	
256 64-bit registers	
4 fully pipelined specialized functional units	
1 divide, SQRT, convert unit	
50 MIPS maximum	
Vector unit	
1 string unit, 16-bits per clock	
1, 2 or 4 64-bit floating point pipeline units	
Each unit has four specialized functional paths	
For a two pipe system results rates are 100 MFLOPS (64 bits)	
Divides are 8 times slower	
Result rates are doubled for 32-bit operands	
Linked TRIADS can effectively double the rates giving 800 MFLOPS (32 bits) maximum	
Memory	
1048576 to 4194304 64-bit words (plus SECDED for half words)	
80 ns cycle time, 300 ns access time to scalar unit	
512 banks of 32-bit words (scalar access)	
or 8 banks of 512-bit words (vector access)	
per megaword of memory	
400 or 800 Mwords/s memory bandwidth to vector unit	
Virtual memory with 48-bit address	
Memory access protection via four keys and locks for each job/page, lockout codes for write/read/instruction fetch	
Virtual memory addressing	
16 associative Address Registers (AR) and space table in memory	
A hit in AR's brings that entry to the top (AR00)	
No hit in AR's but in space table brings AR15 to top of space table and hit to AR00	
A no hit generates an access interrupt and monitor decides whether to add the "missing" entry	
Reference and alter bookkeeping is done on a page by page basis (in virtual addressing mode only)	
All <i>small</i> pages in table must have the <i>same</i> size	
Instructions	
32 bit words or 64 bit words	
Scalar is 3-address register-register	

Table 1 (continued)

Vector is 3-address memory-memory (or 4-address for linked TRIADS)
Sparse vector instructions operand streams must consist of consecutive locations in memory
Scatter and gather instructions
Very similar to STAR-100 (15% instructions changed)
Delays
Scalar
Source operand and result register conflicts (with scalar or vector)
Register file result path conflict
No memory references (except instruction fetch permitted during vector operation)
Vector
Start up times (till first result) typically about 50 cycles, 50% longer for divide, scatter and gather and twice as long for vector macro instructions
Scatter and gather proceed at one result per 1.25 cycles
Divides take 8 times as long as other vector operations
Memory
Bandwidth high enough not normally to cause delays even with up to 400 Mbyte/s of I/O going in parallel
I/O
8 or 16 16-bit channels (having a 32-bit interface to the mainframe)
Maximum rates:
Any channel 12.5 Mbyte/s
Total memory bandwidth available for concurrent I/O: 400 Mbyte/s
I/O interface network (50 Mbyte/s coaxial bus)
Provides attachment for:
– Front end computers
– Tape controllers
– Disk controllers
Context switching
40 words of invisible package and 256 scalar registers are stored and reloaded
The invisible package contains partial results and microcode breakpoint status from interrupted vector instructions as well as normal information
Timing varies from 123 cycles to 190 cycles depending on direction and number of pipes

Each 1 Mword section is split into 16 phased bands of "superwords" (512 bits each, referred to as SWORDS) allowing a service rate to/from the vector processor of 320 Mbyte/s for a 1 Mword system and 640 Mbyte/s for larger system. This sort of bandwidth is necessary both in order to

MONITOR MODE AND I/O USE ABSOLUTE ADDRESSES

Fig. 5. Cyber 205, memory addressing.

cope with configurations with 4 pipes and in order to allow the crucial vector scatter and gather operations to proceed at a very high effective rate ($1\frac{1}{4}$ cycles per element). In addition to this a concurrent bandwidth of 400 Mbyte/s is available to the I/O subsystem split over a maximum of 16 channels of 25 Mbyte/s each. I/O has top priority to the memory but will not, normally, interfere with the streaming rates to the vector processor. Finally, the multiple bank system is so subdivided that the scalar processor accesses each Megaword of memory as 512 phased banks of 32-bit half-words.

3.2. CRAY-1S (fig. 6 and table 2)

The CRAY-1S line offers a memory of up to 4 Mwords (using 4 K bipolar memory circuits) and, on an optional basis, a much more powerful I/O

Fig. 6. CRAY-1, S/1200 and above.

Table 2
CRI CRAY-1S

Old models still available (renamed)
New models are identical except for:
– Up to 4 Mword memory
– Two, three or four I/O processors deal with all I/O
– All I/O goes from an IOP to a MOS buffer memory
– All I/O is channelled between the buffer memory and the central memory via the BIOP using a 106 Mbyte/s special channel
– IBM compatible block multiplexor channels available on the XIOP
– MCU replaced by MIOP
More orders for CRI CRAY-1:
Lawrence Livermore Laboratory
Ministry of Defense (UK)
Max Planck Institute for Plasma Physics
Bell Laboratories
Los Alamos National laboratory
Century Research Corporation
Mitsubishi Research Institute
Boeing Computer Services
Kirtland Air Force Base
AEA UK (Harwell)
CISI/EDF (Paris)
Production getting to 12 machines per year

subsystem. This has revealed itself as being necessary in some applications where extremely high I/O demands decreased the efficiency of the CPU because of insufficient data rates or, more com-

monly, because of CPU overhead in dealing with the interruptions caused by the I/O. Up to 5000 interrupts per second have been seen and the fact that the system copes with them without getting completely paralysed is impressive in itself.

The new I/O subsystem has two principal components: a MOS buffer memory of 0.5 or 1 Mwords and two, three or four input/output processors (IOP's). I/O (read) goes through the IOP's and into the buffer memory from where it goes into the main machine when a sufficient amount of data has been collected. Write is the same but in the other direction. Each IOP is a powerful mini with 12.5 ns cycle time, 16-bit word length, 65536 word memory of 50 ns cycle time and a system-oriented instruction set. The "first" one, the MIOP, acts as a maintenance control unit and has a tape unit, a small disk, a card reader and a line printer directly attached. It runs up to 3 front-ends and has a connection to the buffer memory (as all IOP's) plus a communications line to the mainframe for "dead-start" and diagnostic purposes. The BIOP is the other standard IOP and it runs the disk subsystem (48 or less 600 Mbyte disks in four streams). The BIOP has a very high speed connection to the CRAY-1 central memory (106 Mbyte/s) through which all I/O eventually passes. Two more optional IOP's can be installed, the XIOP for running IBM compatible equipment (tape units are supported now) on block multiplexor channels and the DIOP for running up to 32 more disks.