AIPS MEMO NO. 58

AIPS Benchmarks on the CLSC and PSC Cray X-MPs

Kerry C. Hilldrup

National Radio Astronomy Observatory
Edgemont Road
Charlottesville, VA 22901 USA
(804)296-0211

25 January 1989

ABSTRACT

AIPS has been ported to two Cray X-MP systems under the COS operating system and DDT ("Dirty Dozen Test") benchmarks have been run on both systems, one in stand-alone mode. Comparisons between the DDT timings from the two X-MPs as well as the NRAO-CV Convex C1 have been tabulated. On some DDT problems, the CPU times suggest that the two X-MP systems are very nearly equal. However, for most of the DDT problems the CPU times differ substantially. On the Pittsburgh Supercomputer Center system, which uses a heavily modified version of Cray's permanent dataset management, the CPU times can be greater by a factor of 10 or more (depending on the size and verbosity of the problem). The best real time performance on the X-MPs is better than the C1 by about a factor of 8, but more typically only by a factor of 2-3. The real times at Pittsburgh were degraded by other users, but, even taking that into account, were not significantly better than those of a C1.

1 INTRODUCTION

Under a grant from the NSF supercomputer initiative, AIPS was ported to the Cray X-MP/48 at the Pittsburgh Supercomputing Center (PSC). Afterwards, as part of a cooperative effort with the Department of Astronomy, AIPS was also installed on the Cray X-MP/24 of the Center for Large Scale Computing (CLSC) at the University of Toronto. In so far as AIPS is concerned, the only major differences between the CLSC and the PSC systems are (1) the version of permanent dataset management used and (2) the processing of gather/scatter operations.

2 PERMANENT DATASET MANAGEMENT

The operating system used by the CLSC on their X-MP/24 is "vanilla" COS (i.e., as distributed by Cray). The PSC also refers to the operating system used on their X-MP/48 as COS; however, it is actually a heavily modified version of COS. The PSC contracted Westinghouse to house, maintain and operate their Cray X-MP and, as part of the agreement, the PSC elected to use a version of COS developed and modified over the years by Westinghouse. These modifications have evolved for historical reasons. The differences are primarily in the tape archive system, which is tightly coupled to the way permanent datasets are managed. Cray refers to COS permanent dataset management as PDM. In the PSC/Westinghouse version, it is referred to as PDM-II.

There are a large number of grammatical and syntactical differences between PDM and PDM-II, but perhaps the greatest difference is in terms of functionality. The enumeration of all the differences between PDM and PDM-II is beyond the scope of this memo. Most can be coded around, but some PDM features are entirely missing from PDM-II. For example, PDM gives the user the option of immediately returning (IR) from an attempt to ACCESS a permanent dataset which is for some reason not available. The dataset may not be available because some other job has it ACCESSed for exclusive use or the dataset has been "transparently" archived to tape and needs to be restored to disk first. Under PDM-II, the option of returning immediately simply does NOT exist. Instead, PDM-II automatically suspends the job for 5 minutes and tries again, after having first submitted a job to restore the archived dataset, if necessary. This would introduce intolerable delays during interactive AIPS sessions. This and other deficiencies of PDM-II precipitated the development of a "bridge" library whereby PDM grammar and syntax as well as missing PDM functionalities could, at least in part, be mapped into PDM-II calls. This library was developed by Jerry Kennedy of Westinghouse with input and debugging effort from the author.

In the final analysis, the bridge library was only useful in that it restored the crucial ability to ACCESS a permanent dataset with the immediate return option. We were still obliged to use the CATALOG and PURGE routines of PDM-II rather than the SAVE and DELETE routines of PDM. The error codes returned by the bridge library were also

different from that of PDM and varied depending on whether the job was interactive or batch. We were also still obliged to use PDM-II grammar and syntax in our AIPS programming and execution JCL, all in order to preserve the functionality of the local tape archival storage scheme, the design of which is probably less than optimal for frequently accessed, modified and saved files. The net result of all this extra baggage is excess CPU and real times for jobs that access files frequently. In the DDT, AIPS itself frequently accesses the user's message file and suffered substantially degraded performance as an apparent result of PDM-II.

3 GATHER/SCATTER OPERATIONS

The other major area of difference between the CLSC and PSC systems is that the PSC X-MP is configured with hardware gather/scatter and the CLSC system processes gather/scatter operations in software. All other conditions being equal, we would normally expect the performance of the "clean" algorithm on the PSC system to be superior in both CPU and real time. However, any such superiority is masked in the PSC results by the additional overhead of PDM-II and implies that the extra overhead is probably worse than it may appear.

4 PROCESSOR AND MEMORY REQUIREMENTS

The fact that the PSC machine has 4 processors and the CLSC machine has only 2 was inconsequential since AIPS is coded to run on only one processor at a time. Similarly, the additional memory configured on each processor of the PSC system (8 versus 4 MWord per processor) was not a factor. The Q-routines for COS make use of dynamic memory management to allocate and deallocate the memory required for a 64 KWord pseudo AP. The maximum field length (i.e., memory size) for the DDT jobs was always less than 1 MWord.

5 SOLID STATE DISK (SSD)

Both X-MP systems are configured with a Solid State Disk (SSD), 32 MWord at the CLSC and 128 MWord at the PSC. AIPS, however, was not implemented on either system to make use of the SSD. There is little doubt that the implementation of an SSD for scratch files would improve AIPS real-time performance dramatically; however, the policy at both the CLSC and PSC was to mount the SSD as a device restricted to "temporary" datasets. COS temporary datasets disappear automatically either when they are RELEASEd or when the job ends. However, AIPS jobs first create all the files needed by a given job, reserving the required disk space in the process, then RELEASEs them to be ACCESSed again later (sometimes much later) when the job is ready to actually write data to them. Since the creation of files in AIPS is entirely separate from their actual use the AIPS design does not lend itself to

the use of COS temporary files. Furthermore, scratch files may be ACCESSed and RELEASEd many times in the course of a job. In short, AIPS expects the files it creates to be permanent (scratch files included) until it, not the operating system, is ready to delete them. The bottom line is that it "may" be possible to treat AIPS scratch files as COS temporary datasets, but it would involve a rather substantial development effort (something similar to what would be required if an attempt is ever made to implement "memory resident" files).

6 DDT RESULTS

The attached tables represent comparisons between the timings from the execution of the small, medium and large DDT problems on the CLSC and PSC Cray X-MPs as well as on the NRAO-CV Convex C1 (NRAO1).

1) The NRAO-CV Convex C1 (NRAO1) timings were generated on 9-30-88 using the 15OCT88 release of AIPS. The CLSC timings were generated on 4-18-88 using the 15JUL88 release of AIPS. The PSC timings were generated over the period 10-11-88 to 10-13-88 using the 15OCT88 release of AIPS. In so far as the DDT is concerned, any differences between the 15JUL88 and 15OCT88 releases are not significant. All the DDT executions used the master images from the 15OCT87 version of the DDT tape as input.

2) All the NRAO1 timings were generated in stand-alone mode. Such timings on NRAO1 have been known to vary by as much as 5% for successive executions of the same problem. All the CLSC timings, except those for IMLOD and UVLOD, are also generated in stand-alone mode. None of the PSC timings were generated in stand-alone mode, thus no sensible comparisons can be made with the PSC real times.

3) On NRAO1, "RUN DDTLOAD" took 51 CPU versus 68 real seconds (6 disk AIPS implementation on NRAO1 increases startup and exit somewhat due to new catalog creation and empty catalog deletion). This process compiles and installs the POPS procedures used to execute the DDT problems. In the process, it generates hundreds of messages and therefore accesses the user's message file at high frequency. On the CLSC system, the same execution required 6.95 CPU versus 92 real seconds in stand-alone mode. On a loaded PSC system, the same execution took 63.91 CPU versus 3148 real seconds or more than 9 times the CPU time on the CLSC system. If MSGKIL=TRUE, which means messages are not written to the user's message file on disk, but just to the user's terminal, the required CPU time on the PSC system is considerably less and more in line with that of the CLSC under the same condition. This lends support to the notion that the excess CPU times witnessed on the PSC system can be attributed to PDM-II and/or "bridge" library overhead.

4) All DDT executions on NRAO1 were made with DDISK=MDISK=TDISK=2 and BADDISK set such that all AIPS map I/O was restricted to a single, 4-way striped disk partition with 64Kb block and fragment sizes and

using synchronous I/O. Both X-MPs were configured with multiple DD-39 disks ganged together (NOT striped) as a single generic resource. Asynchronous, word-addressable disk I/O is used for AIPS map I/O in the COS implementation.

5) IMLOD and UVLOD represent the tape oriented, scalar problems of the DDT. NRAO1 tape I/O was performed on a 125 ips drive using Convex asynchronous I/O. The IMLOD and UVLOD timings for the X-MPs are from executions where FITS disk files were used as input. The FITS disk files were stored on the front-end VAX. As part of the IMLOD and UVLOD executions, the FITS disk files were first FETCH'ed across to the main frame's disk, then read using BUFFER IN/OUT, asynchronous I/O.

6) COMB, SUBIM, UVDIF and UVSRT represent the problems of the DDT that DO NOT lend themselves to vector or parallel execution.

7) APCLN, APRES, ASCAL, MXMAP, MXCLN, UVMAP and VTESS represent the problems of the DDT that DO lend themselves to vector and/or parallel execution.

8) Timings for UVLOD, IMLOD, COMB, SUBIM, UVDIF and UVSRT are averages for 3, 8, 7, 2, 2 and 2 executions, respectively. One of the PSC executions of IMLOD involving a medium DDT master image apparently got suspended by the system, perhaps because some archived dataset needed to be restored from tape first. This may explain the large average real time for the IMLOD executions of the medium DDT problem.

9) The APCLN executions on the X-MPs produced bad results. It is not known whether the problem is due to internal compiler (CFT) errors or errors in the AIPS code.

10) As of the 15JUL88 release, the VTESS problem of the DDT no longer converges and will not reproduce the master image from the DDT tape on any system. This is due to a change in VTESS. Nevertheless, timings may still be valid since the DDT problem for VTESS is designed to perform 10 iterations.

SMALL DDT

DDT PROBLEM	NRAO1 (C1-XP) CPU	REAL	C/R	CLSC (X-MP/24) CPU	REAL	C/R	CLSC/NRAO1 CPU	REAL	PSC (X-MP/48) CPU	REAL	C/R	PSC/NRAO1 CPU	REAL	PSC/CLSC CPU
IMLOD [8]	4.90	11.75	.42	1.46	8.50	.17	.30	.72	4.22	206.75	.02	.86	17.60	2.89
UVLOD [3]	7.22	14.00	.52	1.33	10.00	.13	.18	.71	3.44	123.00	.03	.48	8.79	2.59
COMB [8]	3.05	6.88	.44	.30	3.25	.09	.10	.47	2.69	64.88	.04	.88	9.43	8.97
SUBIM [2]	2.80	7.50	.37	.25	3.50	.07	.09	.47	3.16	96.50	.03	1.13	12.87	12.64
UVDIF [2]	4.25	8.50	.50	.46	4.50	.10	.11	.53	1.51	36.00	.04	.36	4.24	3.28
UVSRT [2]	8.56	22.00	.39	.74	12.50	.06	.09	.57	7.15	195.50	.04	.84	8.89	9.66
APCLN [1]	37.51	57.00	.66	6.84	18.00	.38	.18	.32	14.71	1395.00	.01	.39	24.47	2.15
APRES [1]	8.31	21.00	.40	.66	9.00	.07	.08	.43	8.07	220.00	.04	.97	10.48	12.23
ASCAL [1]	85.93	107.00	.80	12.05	25.00	.48	.14	.23	23.43	313.00	.07	.27	2.93	1.94
MXCLN [1]	81.16	110.00	.74	12.48	55.00	.23	.15	.50	35.53	546.00	.07	.44	4.96	2.85
MXMAP [1]	14.58	30.00	.49	1.42	14.00	.10	.10	.47	11.09	227.00	.05	.76	7.57	7.81
UVMAP [1]	11.93	27.00	.44	1.05	12.00	.09	.09	.44	9.49	378.00	.03	.80	14.00	9.04
VTESS [1]	70.10	127.00	.55	5.39	66.00	.08	.08	.52	54.12	1260.00	.04	.77	9.92	10.04

	MEDIUM DDT												
DDT PROBLEM	NRAO1 (C1-XP) CPU	REAL	C/R	CLSC (X-MP/24) CPU	REAL	C/R	CLSC/NRAO1 CPU	REAL	PSC (X-MP/48) CPU	REAL	C/R	PSC/NRAO1 CPU	REAL	PSC/CLSC CPU
IMLOD [8]	7.64	16.88	.45	2.54	16.75	.15	.33	.99	3.91	1836.13	.00	.51	*****	1.54
UVLOD [3]	8.44	16.00	.53	1.37	11.67	.12	.16	.73	3.10	113.33	.03	.37	7.08	2.26
COMB [8]	5.37	11.50	.47	.32	4.25	.07	.06	.37	2.74	89.75	.03	.51	7.80	8.56
SUBIM [2]	5.14	12.00	.43	.43	4.00	.11	.08	.33	3.39	97.00	.03	.66	8.08	7.88
UVDIF [2]	6.64	11.50	.58	.74	4.50	.16	.11	.39	1.96	37.00	.05	.30	3.22	2.65
UVSRT [2]	13.05	33.00	.40	1.07	16.50	.06	.08	.50	7.55	255.00	.03	.58	7.73	7.06
APCLN [1]	187.44	244.00	.77	28.91	68.00	.43	.15	.28	39.42	2292.00	.02	.21	9.39	1.36
APRES [1]	23.01	44.00	.52	2.08	15.00	.14	.09	.34	9.99	333.00	.03	.43	7.57	4.80
ASCAL [1]	382.88	426.00	.90	48.85	59.00	.78	.12	.14	62.83	655.00	.10	.16	1.54	1.29
MXCLN [1]	271.73	348.00	.78	44.47	113.00	.39	.16	.32	60.78	935.00	.07	.22	2.69	1.37
MXMAP [1]	29.70	55.00	.54	2.86	22.00	.13	.10	.40	12.84	512.00	.03	.43	9.31	4.49
UVMAP [1]	27.19	55.00	.49	2.06	22.00	.09	.08	.40	10.64	396.00	.03	.39	7.20	5.17
VTESS [1]	201.39	311.00	.65	16.99	152.00	.11	.08	.49	59.23	2193.00	.03	.29	7.05	3.49

LARGE DDT

DDT PROBLEM	NRAO1 (C1-XP) CPU	REAL	C/R	CLSC (X-MP/24) CPU	REAL	C/R	CLSC/NRAO1 CPU	REAL	PSC (X-MP/48) CPU	REAL	C/R	PSC/NRAO1 CPU	REAL	PSC/CLSC CPU
IMLOD [8]	20.31	34.00	.60	6.78	49.50	.14	.33	1.46	5.72	112.88	.05	.28	3.32	.84
UVLOD [3]	23.57	36.00	.65	3.06	37.33	.08	.13	1.04	5.13	90.00	.06	.22	2.50	1.68
COMB [8]	15.94	27.13	.59	.57	11.13	.05	.04	.41	3.02	148.88	.02	.19	5.49	5.30
SUBIM [2]	14.60	27.00	.54	1.24	9.00	.14	.08	.33	4.22	129.00	.03	.29	4.78	3.40
UVDIF [2]	25.69	33.00	.78	2.94	11.50	.26	.11	.35	4.22	80.50	.05	.16	2.44	1.44
UVSRT [2]	48.24	93.00	.52	4.36	69.00	.06	.09	.74	11.12	637.50	.02	.23	6.85	2.55
APCLN [1]	764.59	1029.00	.74	93.88	303.00	.31	.12	.29	106.37	2750.00	.04	.14	2.67	1.13
APRES [1]	94.58	139.00	.68	9.64	56.00	.17	.10	.40	18.35	498.00	.04	.19	3.58	1.90
ASCAL [1]	4617.42	4840.00	.95	527.14	602.00	.88	.11	.12	617.14	1686.00	.37	.13	.35	1.17
MXCLN [1]	1521.10	2183.00	.70	198.36	588.00	.34	.13	.27	232.64	3709.00	.06	.15	1.70	1.17
MXMAP [1]	98.85	179.00	.55	9.36	66.00	.14	.09	.37	20.26	736.00	.03	.20	4.11	2.16
UVMAP [1]	89.34	157.00	.57	7.05	73.00	.10	.08	.46	17.02	3860.00	.00	.19	24.59	2.41
VTESS [1]	813.90	1381.00	.59	58.43	448.00	.13	.07	.32	108.32	3677.00	.03	.13	2.66	1.85