Migrating from Cray Y-MP to Cray T90 and T3E

Ulrich Detert, Central Institute for Applied Mathematics
Research Centre Juelich (KFA) D-52425 Juelich Germany

ABSTRACT: *The paper describes some aspects of application migration from Cray Y-MP systems to Cray T90 and T3E platforms at KFA. Major migration steps are outlined and performance data are given for all involved systems. This includes a comparison of T3E and T3D performance for selected Fortran kernels and application codes. Furthermore, communication bandwidth and latency are compared.*

1 Introduction

With the beginning of the year 1996 KFA has begun to upgrade its existing Cray infrastructure from two Y-MP systems (Y-MP/8, Y-MP M94) to a combination of T90, T3E, and J90 (Fig. 1, Fig. 2). The current status of the hardware upgrade is depicted in Fig. 3. The following paper outlines some aspects of the upgrade, especially with respect to application migration and application performance.

Due to the architectural similarities of Cray Y-MP and T90 little difficulties can be expected, when migrating applications between these two systems. The migration to the T3E system, however, may require much more effort. Namely, the message passing parallelization is a matter of major concern.

With respect to performance, two aspects are of importance: single CPU performance for Cray T90 and Cray T3E on the one hand, and scalability of shared or distributed memory parallel applications on the other hand. Furthermore, I/O performance may be crucial for a number of applications on both, PVP and MPP architectures. The following paper focuses mostly on single CPU performance for T90 and T3E. Furthermore, some performance figures for typical KFA application codes are added. As far as T3E is involved, these codes have been run on a moderate number of PEs. A comparison of T3E performance data with T3D data is also given.

I/O has not been investigated for the purpose of this paper, since the current I/O hardware configuration (and as far as T3E is concerned also the software environment) is still far from its final state (compare Fig. 3).

2 Application Migration

2.1 Migration from Y-MP to T90

Due to the upwards compatibility of Cray Y-MP and Cray T90, application migration for these systems is rather transparent to the user. The fact that Fortran 77 will be supported on T90 systems for only a limited time span, however, involved additional porting effort for KFA users. Even though Fortran 77 is a complete subset of Fortran 90, application codes will not always run without change when ported from the CF77 to the C90 compiler. Difficulties may arise, for instance, if the cpp preprocessor was used with CF77 and is now replaced by the gpp preprocessor. Also, optimization properties of CF77 and C90 may be completely different. This may require compiler-specific code changes, if ultimate performance is aimed at. For the near future, a major item of concern will be the migration to Cray T90 IEEE arithmetic. Two issues are of importance here: the conversion of existing numeric binary data sets to IEEE format and the preservation of arithmetic accuracy for numerically critical applications.

2.2 Migration to T3E

The migration of application codes to Cray T3E is more versatile than to T90. At KFA, mainly two platforms exist where application codes may be taken from to be ported to Cray T3E. One is the existing and still operational MPP system Intel Paragon, the other is the Cray PVP platform (Y-MP/8 and Y-MP M94). For the former, application migration in essence means the conversion of NX message passing to either MPI, PVM, or shmem put/get. At KFA, the strategy is to recommend MPI message passing for portability reasons.

The porting of PVP codes to the T3E platform is being carried out in two steps. For about one year KFA has an aggreement with the Konrad-Zuse-Zentrum fuer Informationstechnik in Berlin to utilize 32 PEs of their T3D system for the preparation of application codes for operation on T3E. T3D programs can then be migrated to the T3E system at KFA. The main advantage of this mode of operation was that application migration could start early before the availability of the T3E system at KFA. Main steps in the process of porting codes from T3D to T3E are, firstly, the adaptation of specific properties of the shmem get/put routines from T3D to T3E, secondly, platform-specific code optimizations like cache or stream buffer utilization, and finally in some cases, the conversion of PVM message passing calls into MPI calls for portability or performance reasons.

3 Performance

In the following, some performance data shall be given for T90, T3D, and T3E. The T90 performance data stem mostly from investigations that have been carried out for the KFA acceptance tests. The data for T3E and T3D have been taken from measurements for Fortran loops on the one hand and from KFA production codes on the other hand.

3.1 T90

Fig. 4 depicts the single-CPU T90 performance of three KFA application codes and a selection of Fortran loops. The multi-CPU comparison between T90 and Y-MP has been gained by extrapolating these results to 4, 8, and 12 CPUs, respectively.

3.2 T3D and T3E Fortran Kernels

Figures 5 to 8 reflect results of loop measurements on T3D, T3E and T90. On T3D and T3E all measurements have been done with a preceding cache flush operation and a repeat loop around the measured loop. Thus, an assessment of execution times with and without cache reuse is possible. The average values of the measurements are given as a horizontal bar in all diagrams. As a result we can see that the larger cache in the T3E leads to a ratio of about 4.6 between T3D and T3E for the cached loops, but only to a ratio of about 2.7 for the non-cached versions. For the comparison of T3E and T90 the result is that approximately 9 PEs of a T3E yield the performance of one T90 CPU for the cached loops, however, roughly 25 PEs per CPU are required for the non-cached versions.

Fig. 8 shows the performance effect of the T3E stream buffers. As expected, the performance gain is correlated with the vector length of the loops. On an average, the ratio for stream buffers switched on and off is about 1.8.

3.3 Communication

Fig. 9 and Fig. 10 reflect the communication bandwidth and latency for MPI and shmem put/get communication on T3D and T3E. The achievable bandwidth for the shmem routines on T3E is above 300 MB/s. Different from T3D there is no significant performance difference between the put and the get operation. MPI ssend (synchronized send) delivers over 250 MB/s on T3E - significantly more than MPI send with about 100 MB/s. However, with respect to latency, MPI send is faster than ssend. It is worth while to notice that the latencies of the shmem get/put operations have increased when moving from T3D to T3E whereas for the MPI routines they have decreased.

3.4 Application Code Performance

Fig. 11 shows the performance of a Car-Parrinello code for T90 and T3E on 16 PEs. Only the most significant routines are listed in the diagram. There is one routine (XCENER) that dominates the execution time on T90. This routine is hardly vectorizable but can be efficiently parallelized on T3E. As a consequence, the ratio of the overall execution time on T90 and T3E is 2.1. Thus, for this code 8 PEs of the T3E equal one CPU of T90 in performance. For the Car-Parrinello code the performance ratio between T3D and T3E is about 2.7 on 16 PEs and 3.0 on 32 PEs, respectively.

Other codes have been investigated for the comparison between T3D and T3E as well. A Crystal Growth Simulation code yields a performance ratio of 3.5 for T3E over T3D, for a QCD code the ratio is 3.0.

4 Conclusion

The process of migrating from Cray Y-MP to T90 and T3E has not yet finished at KFA. For T90, a significant step will be the move to IEEE arithmetic and GigaRing I/O hardware. The performance expectations for the T90 are fully met with respect to CPU performance. In the I/O area optimizations on the operating system level or the application level may be required for I/O-intensive application codes. This, however, can only be assessed when the final GigaRing I/O hardware is available.

For T3E the situation is similar. The performance expectations are well met, even though for certain application codes single PE performance optimizations may be helpful. The I/O hardware currently installed at KFA is very preliminary. Here, parallel I/O via multiple GigaRing interfaces is crucial for future application codes. A concluding assessment of the functionality and stability of the T3E system at KFA will only be possible after the installation of the fully featured system with 512 PEs, multiple GigaRings, and officially released operating system software.

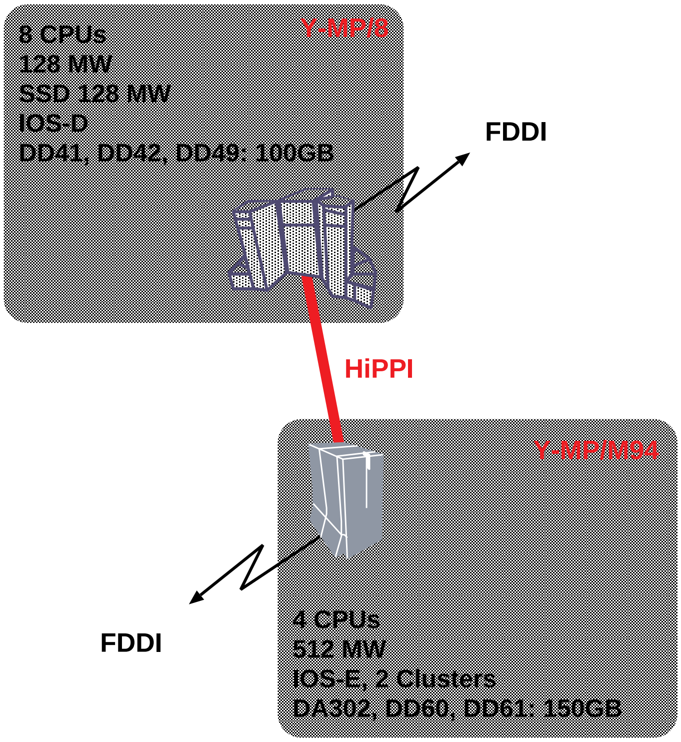


Figure 1: Cray configuration (past)

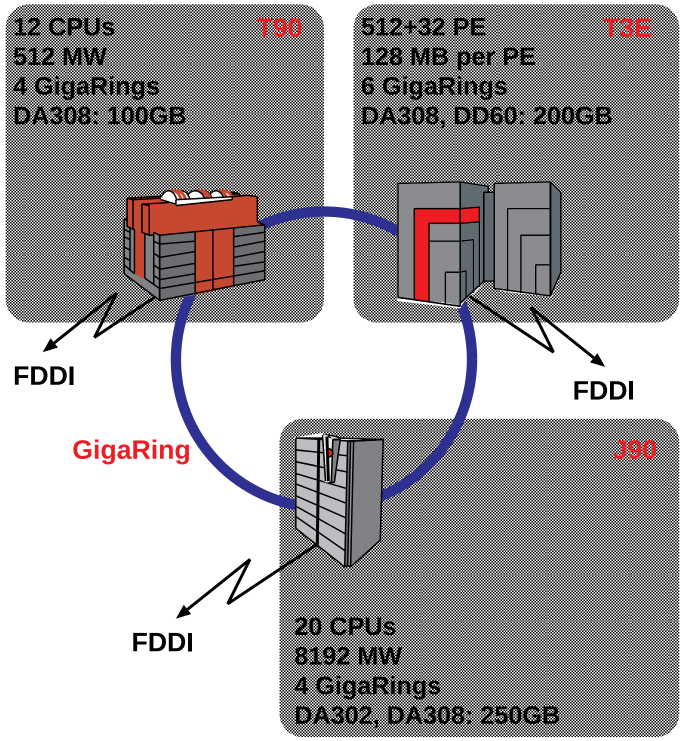


Figure 2: Planned Cray configuration

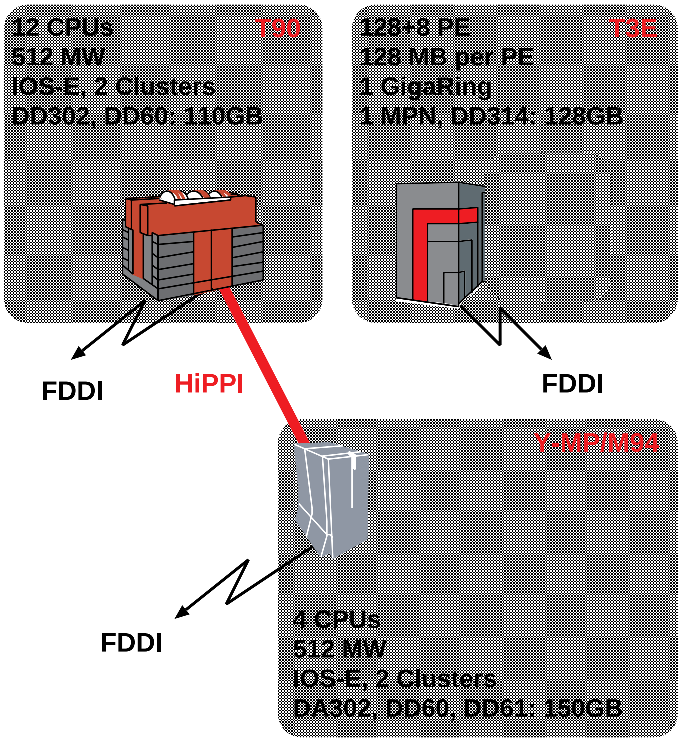
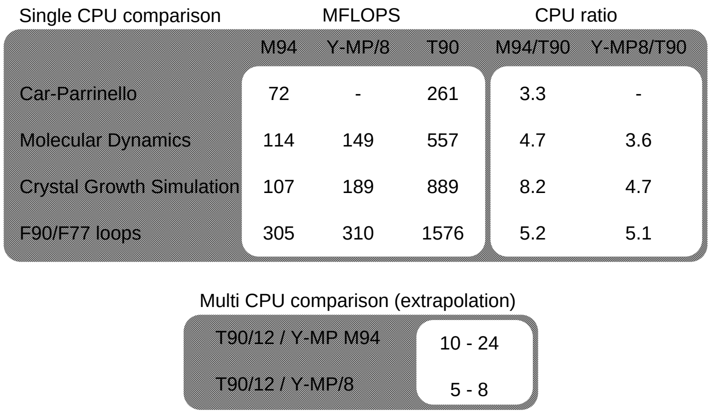


Figure 3: Current Cray configuration



Single CPU comparison	MFLOPS			CPU ratio	
	M94	Y-MP/8	T90	M94/T90	Y-MP8/T90
Car-Parrinello	72	-	261	3.3	-
Molecular Dynamics	114	149	557	4.7	3.6
Crystal Growth Simulation	107	189	889	8.2	4.7
F90/F77 loops	305	310	1576	5.2	5.1

Multi CPU comparison (extrapolation)	
T90/12 / Y-MP M94	10 - 24
T90/12 / Y-MP/8	5 - 8

Figure 4: Performance T90, Y-MP/8, Y-MP M94

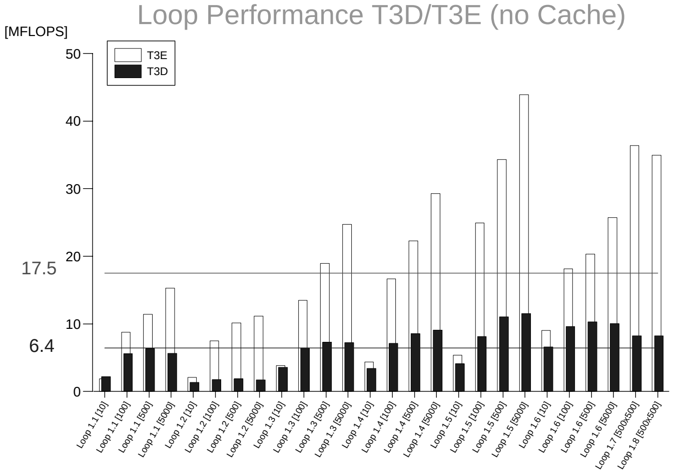


Figure 5: Fortran kernels (uncached data)

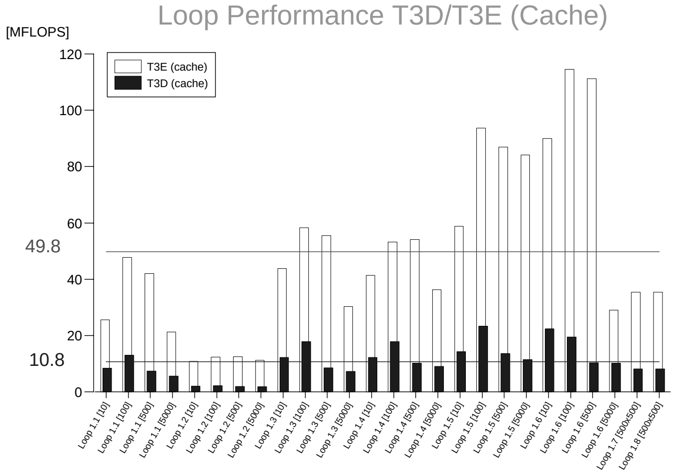


Figure 6: Fortran kernels (cached data)

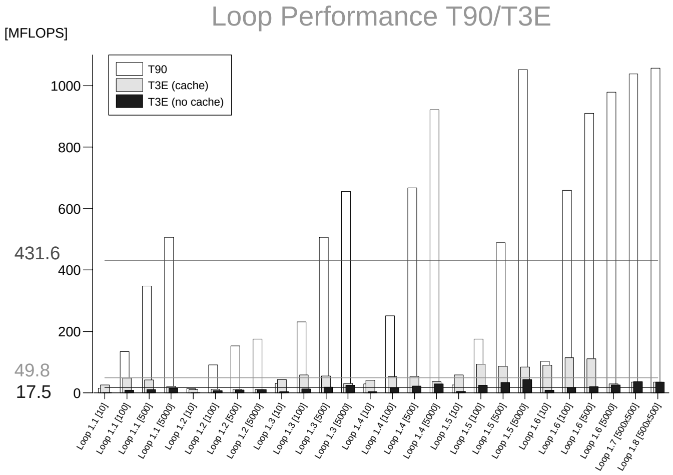


Figure 7: T90/T3E performance

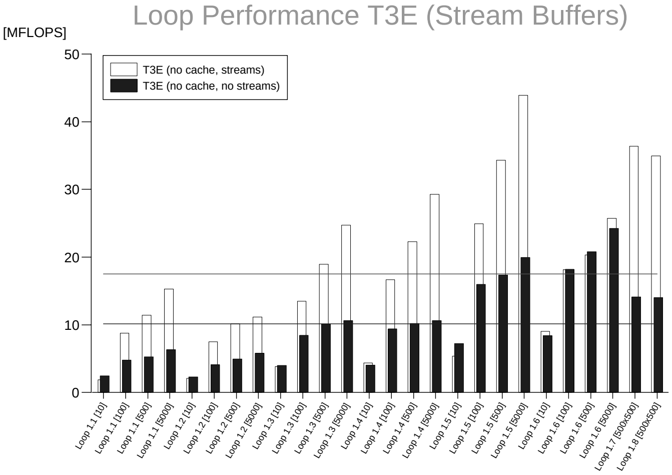


Figure 8: T3E stream buffer performance

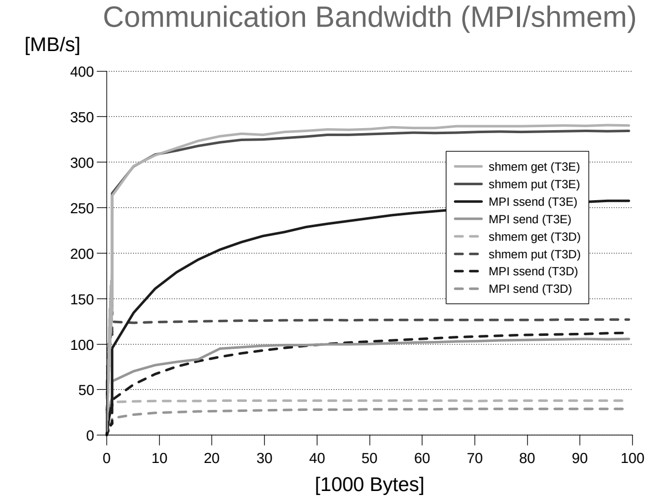


Figure 9: Communication bandwidth

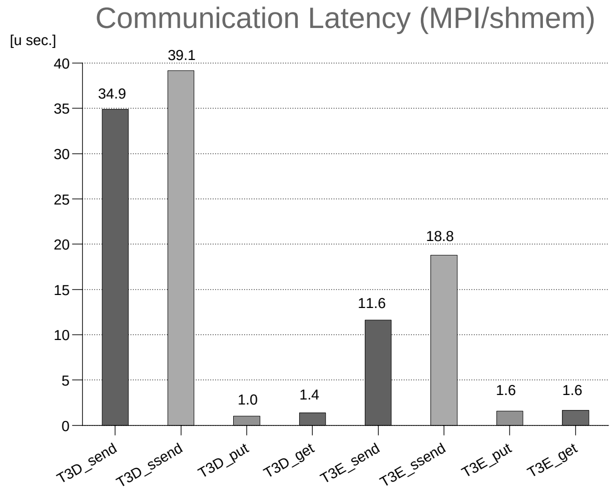


Figure 10: Communication latency

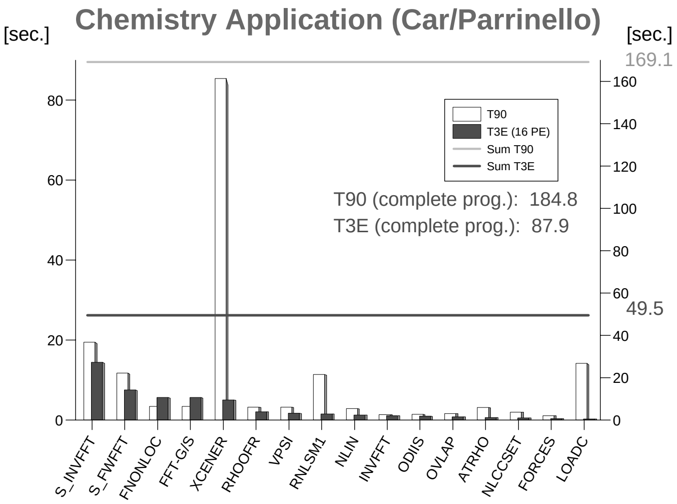


Figure 11: Application code performance