CRAY T90 Series Differences: Tutorial

Michael Burwell, Cray Research, Inc., Eagan, Minnesota, USA

ABSTRACT: *One of the flagship products of Cray Research, the CRAY T90 series of computer systems represents a different hardware architecture from other mainframe offerings. Differences also exist in the way interprocess communication is handled on CRAY T90 systems. Additionally, these systems use a variety of system configuration tools that are different from those operating on the rest of the Cray Research line of mainframe products. The tutorial summarized in this article provides an overview of these hardware architecture, interprocess communication, and software configuration differences. The presentation concludes with a summary of user-level differences, particularly as compared with CRAY C90 series systems.*

Introduction

The CRAY T90 series of computer systems is the current generation of CRAY Research's high-performance parallel vector processing (PVP) systems. They are the successors to the CRAY C90 series of high-end supercomputer systems. Because of the importance of the CRAY T90 series among the suite of products offered by Cray Research, coupled with the fact that they are still relatively recent offerings, this tutorial is provided.

Among the salient characteristics of CRAY T90 series systems are their compact design, larger number of processors, scalable architecture, shared memory feature, faster clock speed, advanced optical clock system, dual mode of operation (either "native" T90 mode or C90 mode), sanity codes, and easy and cost-effective maintainability meant to reduce the mean time to repair (MTTR). With a maximum of 32 processors, clock speed of approximately 2.2 ns, and a per-CPU performance of around 1.8 GFLOPS, these machines represent a fourfold to fivefold price/performance increase over CRAY C90 series systems.

One of the most noteworthy hardware features of CRAY T90 series machines is the relative absence of wires. A system interconnect board replaces the traditional wire mat, therefore effectively replacing what has been the single greatest source of failures in installed systems. By contrast, a typical CRAY C90 series machine has approximately 36 miles (58 km) of wiring.

Another unique capability is the use of sanity codes. Sanity code is a control mechanism that ensures that a logical path exists to all configured portions of the system. If the target module successfully returns the 6-bit sanity code pattern sent to it, the module is powered on logically. If the code is not returned successfully, the module causes that module to go into a master-clear state so that all output from that module is ignored.

Architecture Overview

The major hardware components of a CRAY T90 series system are as follows:

- Mainframe cabinet
- Input/output subsystem (IOS) / solid-state storage device (SSD) cabinet
- Heat exchange unit (HEU)
- AC/DC converter (HVDC-40)
- Support system (OWS, MWS, support system chassis, printer, modems)

All three basic models available (CRAY T94, CRAY T916, and CRAY T932) reflect a relatively small footprint (for example, only 37.5 ft^2 [3.45 m^2] for the largest system). (The mainframe cabinet footprint on a CRAY C916 system is 81 ft^2 [7.5 m^2].) The maximum number of CPUs is 32. Maximum memory ranges from 64 Mwords in the CRAY T94 to 1024 Mwords in the CRAY T932.

In CRAY T932 systems, the two halves of the chassis have separate power and cooling connections. This enables personnel to power off one-half of the chassis and drain the dielectric coolant, while the other half continues to operate.

The CRAY T94 system has a peak performance of 7.2 GFLOPS. The CRAY T94 mainframe is 5 ft wide, 3 ft deep, and 4 ft high (1.5 m x 1 m x 1.5 m). It has a dry weight of 2000 lbs (909 kg). The CRAY T916 system has a peak performance of about 28.8 GFLOPS. The CRAY T916 mainframe is 7.5 ft wide, 5 ft deep, and 5 ft high (2.3 m x 1.5 m x 1.5 m). It has a dry weight of 6000 lbs (2727 kg). The CRAY T932 system has a peak performance of approximately 57.6 GFLOPS. The CRAY T932 mainframe cabinet has the same footprint and appears the same as the CRAY T916 mainframe cabinet, and has a dry weight of 10,000 lbs (4545 kg). By comparison, the CRAY C90 mainframe cabinet has a weight of 13,500 lbs (6136 kg).

Unlike previous Cray Research mainframes, such as the CRAY C90 series, CRAY T90 series mainframes do not require rotating motor generators. The systems are powered by a static AC/DC converter. The converter uses standard commercial utility power. The high-voltage DC is statically converted in the mainframe to low voltages used by the circuitry. These converters are immersed in the cooling tank.

An uninterruptible power supply (UPS) is available on the CRAY T90 series. All UPSs are sized to provide power not only to the mainframe, but to the entire system, including the IOS, SSD, and peripherals.

Each system comes with an input/output subsystem (IOS) and also can be configured with a solid-state storage device (SSD). On a CRAY T90 series machine, the SSD and the IOS are combined into a single chassis. Several models of IOS/SSD cabinets are available. The largest SSD has 4096 Mwords of memory.

CRAY T90 System Integrated Circuit and Printed Circuit Board Technology

You can gain a further appreciation of the positive differences offered by CRAY T90 series systems by looking at the integrated circuit (IC) technology. Specifically, the following features are important to consider:

- Density and power consumption
- Printed circuit module types
- Boundary scan

Chip Density and Power Consumption

The integrated circuit technology selected for the CRAY T90 series was developed in conjunction with Motorola. The Emitter-Collector-Dotted (ECD) chip can have up to 50,000 gates, about half of which are usable in the actual chip design for processor logic. This technology makes possible approximately a fourfold increase in the density of gates per logic element over the CRAY C90 series. The chips operate at about 450 MHz.

The ECD chips also provide a programmable power feature. Critical paths within chips are run at full power, but gates on noncritical paths within an IC can run slower on lower power. Using this feature, each CRAY T90 series CPU consumes about half the power of a CRAY C90 series system while providing about four times more performance.

Module Types

The tutorial discusses six basic types of printed circuit board modules on a CRAY T90 series system:

- Central processor (CP) module
- Central memory (CM) module
- I/O module
- Shared (SR) module
- Network module
- Boundary scan module

Each module is smaller than a standard 8.5-by-11 in. sheet of paper and can have 52 to 54 separate layers. Each module consists of three boards that are laminated together with a direct solder to reduce connections. Over 1 mile of interconnect can be embedded within a single module. The two outer boards are 22 or 23 layers each and carry signals, while the inner 8 layers carry power and ground. Integrated circuits are mounted on both sides of the modules producing circuit densities of over 800,000 gates within 2 ns of each other. In addition, one-third of the chip-to-chip interconnects are eliminated, because the silicon is directly mounted on the PC module. This technique improves reliability.

It is particularly significant to note that modules are connected to each other using electrically activated zero-insertion-force (EZIF) connectors. Each connector supports 400 signal contacts. One feature of this type of connector is extremely tight impedance control. Each EZIF connector replaces 800 interconnect wires. Modules can be installed or removed easily by applying low voltage to operate the heat-controlled spring on each connector. The only wires that remain in the CRAY T90 series are for external I/O connections.

Each *central processor (CP) module* contains one CPU. A CRAY T90 series system can contain from 1 to 32 CPUs, depending on the chassis type. CP modules are arranged in a stack of up to four modules in a CRAY T94 chassis, and in stacks of up to eight modules in a CRAY T916 and CRAY T932 chassis. Each CPU has 8 physical ports to memory. The CPU is instruction compatible with the CRAY C90 series instruction set and can run in C90 mode.

The *central memory (CM) module* contains central memory that is common to all CP modules. Like the CP modules, the memory modules are arranged in stacks. The CRAY T94 chassis contains one stack of up to four CM modules. Each CM module consists of a single printed circuit board with memory stacks on one surface and logic options on the other surface. Each CM module has 16 memory stacks. Each memory stack consists of two printed circuit boards that contain forty 4-Mbit bipolar complementary metal oxide semiconductor (BiCMOS) static random-access memory (SRAM) chips having an access time of 15 ns. The stack has two columns of 20 chips each. Nineteen chips in each column store data. The bottom chip in each column functions as the spare chip for the column. This method of stacking memory chips represents a key difference between the CRAY T90 and the CRAY C90 series. When fully populated with 4-Mbit chips, a CM module contains 32 Mwords of memory.

The *I/O module* provides an interface between CP modules and external channels. In CRAY T916 and CRAY T932 systems, the I/O module connects directly to four of the eight modules in a CP stack and makes them “I/O capable.” Each I/O module can handle transfers for 8 LOSP channels, 8 HISP channels, 4 VHISP channels, and a variety of special channels, including those connected to the support system. CRAY T932 systems can contain up to four I/O modules. CRAY T916 systems can contain two I/O modules, and CRAY T94 systems contains only one.

The *shared (SR) module* functions as a central point for I/O command and status routing. Because a direct connection does not always exist between the CPU making a request and the I/O module receiving it, all I/O commands are always sent first to the SR module. The SR module forwards the request information to the proper I/O module that services the requested channel. CRAY T94 and CRAY T916 systems each contain one shared module; CRAY T932 systems contain two shared modules.

The *network module* is responsible for connecting the CPUs to the memory modules on CRAY T916 and CRAY T932 systems. Network modules are not needed on CRAY T94 systems.

The *boundary scan module* serves as an interface for all boundary scan operations. One important function of the boundary scan module is to detect interruptions, caused by shorts, in the continuity line for each module (this is a long, continuous metal line that runs near all of the options on the module).

Boundary scan testing

This tutorial provides a brief overview of boundary scan testing. This capability consists of a set of programs that makes it possible for field personnel to test most connections within a module, as well as connections between the modules, without having to perform diagnostic tests. Boundary scan testing can be run whenever a failure occurs that shuts down the system, and it is run to verify the integrity of the system after a repair procedure has been completed. Boundary scan cannot test memory chips or I/O channel connections because no boundary scan capabilities are built into those connections.

CPU Architectural Enhancements

CRAY T90 series systems reflect several CPU enhancements that mostly have to do with expanded registers. New CPU features pertain to registers, scalar cache, functional units, the exchange package, memory control, and logical address translation (LAT) tables. Some registers (for example, the A [address] and B [saved address] registers) were expanded to 64 bits, and overlapping gather/scatter operations were introduced, thereby reducing the time it takes to load the vector registers. Moreover, a 1024-word scalar cache was added to each CPU for A and S register load operations, new vector functional units have been added, and the exchange package was expanded to 32 words.

In addition, the memory control scheme, formerly a parallel access system (whereby all CPUs were involved in every memory reference), was replaced by a network access system. In this type of memory, the requesting CPU arbitrates the access only in its own ports, thereby reducing overhead. The CRAY T90 memory management scheme uses of a new method to reference logical addresses called *logical address translation tables* (LATs). LATs replace base and limit addresses, which were limited to 32-bit addresses and were not flexible enough. LAT tables enable use of multiple address spaces and perform an important function in the management of shared memory. LAT tables enable direct access to a maximum of 32 Gwords of memory. Unlike previous systems, there is no physical base address in the CRAY T90 memory management scheme.

Support System

The support system is the point of first-level access, control, and status of a CRAY T90 series system. The support system consists of the following components:

- Support system chassis
- OWS and MWS workstations
- CRAY T90 support multiplexer (TSM)
- HP LaserJet 4ML printer

Cray Research uses Sun Microsystems, Inc., SPARCstation 5 workstations as the operator workstation (OWS) and the maintenance workstation (MWS) with CRAY T90 series systems. The OWS and the MWS are physically identical. Cray Research upgrades each workstation with the addition of the following components:

- 1.05-Gbyte Seagate Technology, Inc. small computer system interface (SCSI-2) compatible disk drive
- An internal Network Peripherals, Inc. (NPI) FDDI adapter (Sun X1015A)
- 644-Mbyte internal compact disk read-only memory (CD-ROM). For customers concerned about security, Cray Research also offers a removable disk option for the workstations.
- External 5-Gbyte 4-mm tape drive

Unlike with the CRAY C90 series, there is no Ethernet connection between the OWS and the MWS. Workstation-to-workstation communication occurs through an FDDI connection (a 100-Mbit/s fiber-optic local area network) in the system support cabinet. Customers cannot connect their network to the FDDI ring. Instead, they can use the Ethernet port on the OWS to connect to the network.

Interprocess Communication

Prior to the introduction of CRAY T90 series systems, Cray Research system architectures supported several well-known forms of interprocess communication including files, signals, pipes, and named pipes. UNICOS 9.0 running on a CRAY T90 series system supports three additional types of interprocess communication known as System V IPC. These new types are shared memory, semaphores, and message queues. *Shared memory* allows two or more processes to share a specific region of memory. *Semaphores* are counters that are used to provide access to a shared data object for multiple processes. *Message queues* are linked lists of messages stored within the kernel and identified by a message queue identifier. These three new interprocess communication facilities are accessible only from the system-call level. Whereas semaphores and message queues are now available to other Cray PVP systems running the UNICOS 9.0 operating system, only CRAY T90 series systems use the shared memory feature. Shared memory provides the fastest means of exchanging data between processes.

CRAY T90 Series Configuration Tools

This tutorial supplies an overview of the operation of three main configuration tools available on the support system associated with CRAY T90 systems. These tools, listed as follows, are intended to aid administrators in configuring both hardware and software:

- `pact`(8) utility. This tool is used to configure the software environment. A graphical user interface (GUI), `pact` creates a software environment file and an associated parameter file. These files are used by the operator workstation (OWS) during the boot of a system environment into a CRAY T90 series machine.
- `fsct`(8) utility. This GUI tool is used to configure mainframe file systems by defining disk partitions (slices) and associating them with logical disk device names. Two useful features of `fsct` are a hardware tree and a software tree, which permit administrators to visualize the configuration as they define it.
- `tconfig`(8) utility. This tool is used to create and display machine environment files, which contain both hardware and software definitions. The hardware definitions are assembled from a program called the system configuration environment (SCE) program (also known as FastT). The software "map" is assembled from the system environment file (created by `pact`) and the UNICOS parameter file. `tconfig` also checks the consistency of the hardware and software maps. Finally, `tconfig` informs the boot software of its choices (through the machine environment file).

The `pact` and `tconfig` utilities reflect the fact that a CRAY T90 series system configuration is very "soft"; that is, almost all of the configuration mechanisms are controlled by software. There are no default memory configurations, number of CPUs, preassigned I/O clusters, and so on.

Additional CRAY T90 and CRAY C90 Series Differences

There are several differences that might affect users who are migrating from a CRAY Y-MP platform, such as a CRAY C90 system, to a CRAY T90 series system. These differences were outlined in the January 1996 issue of the *Cray Research Service Bulletin* and are summarized here as follows:

- Cray TotalView debugger and CDBX debugger. The TotalView debugger is the only debugger supported on CRAY T90 series systems.
- Fortran-callable system interface routines. Fortran-callable system interface routines that conform to the POSIX FORTRAN 77 Language Interfaces standard are available in both CRAY C90 and CRAY T90 mode.
- CF90 compiler. The CF90 compiler is the recommended Fortran compiler for CRAY T90 systems.
- CRAY T3D emulator. The CRAY T3D emulator is not available on CRAY T90 systems.
- IEEE floating point. Selected CRAY T90 series systems are being built with IEEE floating point. This standard is being adopted in order to ease the movement of data between CRAY T90 mainframes and other computational entities such as workstations and graphic displays. The only drawback to IEEE floating point as compared with Cray floating point is that the exponent range is reduced. In most cases, however, this difference should not be noticeable.
- IRIS Distributed Graphics Library (DGL). DGL is not supported on CRAY T90 systems.

In addition to these differences, users are reminded that most CRAY T90 systems (that is, those that do not support IEEE floating point) provide a CRAY C90 compatible mode, allowing CRAY C90 executable files to run on CRAY T90 systems. Applications also can be compiled and linked in CRAY T90 mode, which allows the system software to use several CRAY T90 enhancements that are not available on CRAY C90 system.

More information about changes in the UNICOS 9.0 release to support CRAY T90 systems is available in the *UNICOS 9.0 Release Overview*, publication RO-5000 9.0. Videotapes and related materials based on the CRAY T90 Series Workshop (T90WS) also are available from Software Education Services. In addition, the UNICOS System Administration (USA) course provides a CRAY T90 system overview and gives instruction in using the `pact`, `fsct`, and `tconfig` utilities.

For more information about these and other software courses, you can access the software training catalog by specifying the following URL on the World Wide Web:

```
http://www.cray.com/education
```


Acknowledgments

I am indebted to numerous people for the information found in this tutorial: to the hardware and software developers of the CRAY T90 series systems; to the hardware documentation and training staff of Service Publications and Training in Chippewa Falls; and to the software documentation and training staff. I especially appreciate the support given by the managers and the editing and production staff of Software Education Services.