



## Немного истории

Во второй половине XX века слово Cray стало синонимом громоздкого слова «суперкомпьютер». Да и в общем, так остается и по сей день. Собственно, большие и мощные ЭВМ выпускали и выпускают многие. Однако именно компьютеры Cray остаются непревзойденными по соотношению цена/эффективность. Не говоря уже о том, что они невероятно долгое время оставались абсолютными лидерами по вычислительной мощности. А ведь все эти легендарные суперкомпьютеры – детища одного человека. Ветерана II Мировой войны Сеймура Крея.

### **Биография**

Сеймур Крей, создатель первого в мире суперкомпьютера, родился в городке Chippewa Falls, штат Висконсин, в 1925 году. Отец Крея в первые годы своей карьеры работал инженером в электрической компании Northern States Power Company. Позже он перевез семью в Chippewa Falls, где ему предложили должность городского инженера. Мать не работала, вела жизнь домохозяйки, обеспечивая успех карьеры мужа.

Сеймур Крей с самого детства увлекался тем же, чем и почти что каждый первый подросток – он возился с электроникой. (Надо помнить электронику того времени – все исключительно на лампах.) В школе его любимым кабинетом была электротехническая лаборатория. Сеймур разбирал и собирал обратно электрическое оборудование любого назначения: радио, электрические схемы, электродвигатели и все тому подобное.

Но в 1943 году, едва окончив школу, Сеймур Крей был призван в американскую армию. Повоевать ему пришлось изрядно. Военная судьба забросила его поочередно в два театра военных действий. Вначале высадка в Нормандии. С боями Крей прошел через Францию и Германию. И одним из первых пожал руку советским союзникам на Эльбе. Но войну он закончил на Филиппинских островах, воюя уже с японцами.

Вернувшись из армии, молодой Крей продолжил образование. В 1950 году Сеймур стал обладателем степени бакалавра университета штата Миннесота. Специализация – электрические технологии. Годом позже он получил ученую степень магистра, занимаясь прикладной математикой. Тогда же Сеймур Крей поступил на работу в фирму Engineering Research Associates (ERA).

В те годы ERA финансировалась военно-морским флотом США и занималась разработкой криптографического оборудования – самая что ни есть прикладная математика. К моменту прихода в компанию Крея ей еще не исполнилось и года. Размещалась ERA на территории бывшего завода. Судьба любит пошутить – именно этот завод производил во время войны деревянные планеры, на одном из которых Сеймур Крей когда-то десантировался в Нормандии.

## Создавая компьютер

В ERA перед Креем поставили «несложную» задачу – создать компьютер. Что это такое и как делается, в то время не знал вообще никто. В результате Сеймуру Крею пришлось потратить немало времени на изучение самого понятия «вычисления» («computing»), и на определение того, что же такое вообще «компьютер». Довелось ему и посещать лекции Фон Неймана. В итоге Крей понял: по большому счету никто ничего не знает. А значит, надо начинать самому.

И начал с нуля работу над компьютерами серии 1100. Это они впоследствии стали всемирно известны под названием UNIVAC. В качестве сотрудника ERA Крей поработал над созданием двух первых компьютеров компании. А затем взялся за собственный проект научного компьютера. На начальной стадии Крей работал в свободное от основной работы время. Вообще, как говорят его биографы, работа в неурочное время была для Крея абсолютно нормальной на протяжении всей жизни.

Создавая собственную ЭВМ, Сеймур Крей прежде всего учитывал критику UNIVAC со стороны заказчиков этого компьютера. Первой собственной разработкой Крея стал проект 1103. Работая над 1103, Сеймур поставил во главу угла принцип простоты. «Ничего лишнего» – любил говорить он. Интересно, что в те годы остальные разработчики шли диаметрально противоположным путем: они максимально усложняли свои вычислительные системы, «навешивая» на них различные схематические решения.

А вот Крей не создавал излишеств, зато начал использовать принципы RISC-технологии еще до того, как вообще появилось понятие RISC. Также Крей стремился к минимизации сроков разработки проектов. Для него время было решающим фактором в борьбе за клиента.

Через пару лет ERA была куплена компанией Remington Rend. Новый владелец постепенно начал уделять больше внимания разработке коммерческих компьютеров – напомним, что прежде заказчиками вычислительных систем в США были исключительно государственные организации. Вскоре Remington Rend урезала финансирование линии научных ЭВМ. Тогда Крей присоединился к коллеге Биллу Норису – вместе они создали фирму Control Data Corporation (CDC).

Сеймур Крей сумел убедить остальных основателей CDC в том, что производство коммерческих компьютеров, которым те хотели заняться, не принесет достойной прибыли. Удивительно, но Крей оказался прав. На тот момент рынок вычислительных систем для бизнеса еще не сформировался, зато ученые охотно покупали компьютеры. Начав под руководством Крея разработку и сборку больших машин, предназначенных для научно-исследовательских задач, CDC вскоре отобрала рынок мейнфреймов у Remington Rend. В первый же год работы CDC получила прибыль в \$600.000 – огромные деньги по тем временам. А Сеймур Крей просто наслаждался работой.

## Набирая обороты

Начиная каждый новый проект, Крей прежде всего собирал отзывы заказчиков о предыдущих моделях его компьютеров, изучал их критику и пожелания. То есть Сеймур Крей фактически создал еще и маркетинг в области IT. Но сам он хотел создать самый быстрый научный компьютер.

Тем временем компания CDC росла, диверсифицируя свой компьютерный бизнес для выпуска ЭВМ различного назначения. Крей, чтобы не отвлекаться на коммерческие вопросы, предпочел переехать в другой город, в Chippewa Falls, и создать там отдельную научно-исследовательскую лабораторию CDC. Именно там и был создан первый компьютер, полностью разработанный лично Креем – CDC 1604. Это была небольшая революция: CDC 1604 оказался первым и единственным в мире компьютером, собранным только на транзисторах.

Крей использовал транзисторы, хотя на тот момент, это была самая дорогая элементная база. CDC 1604 оказался самой успешной с коммерческой точки зрения большой ЭВМ. Крей своей разработкой доказал, что нестандартные компоненты (в данном случае – транзисторы) могут успешно использоваться для построения вычислительных машин. Далее последовали модели CDC 6600 и CDC 7600, с восторгом воспринятые научным сообществом.

К 1972 году стало ясно: рынок больших компьютеров, предназначенных для решения узкоспециализированных научных проблем, быстро сокращается. Будущее явно было за универсальными решениями. Руководство CDC решило перейти с рынка решений для науки и переориентации на коммерческие ЭВМ. Крей принял решение уйти из фирмы и основать свой бизнес. Так было положено начало компании Cray Research (Миннеаполис, штат Миннесота).

Тогда же началось проектирование будущей легенды – компьютера Cray 1. Сеймур Крей собирался впервые использовать в ЭВМ интегральные схемы и векторные регистры. Знаменитый S-образный корпус компьютера, за который Cray 1 получил прозвище «диван», преследовал вполне конкретную цель: значительно снизить длину проводников внутри ЭВМ. В таком корпусе не было ни одного провода длиннее метра. Более того, специально для Cray 1 была создана уникальная система охлаждения с использованием фреона – опять же, впервые в мире.

Первый Cray 1 был установлен в Национальной лаборатории в Лос-Аламосе. А затем начались коммерческие продажи, причем вполне успешные. На Cray 1 компания Cray Research заработала по-настоящему большие деньги, а Сеймур Крей – статус живой легенды.

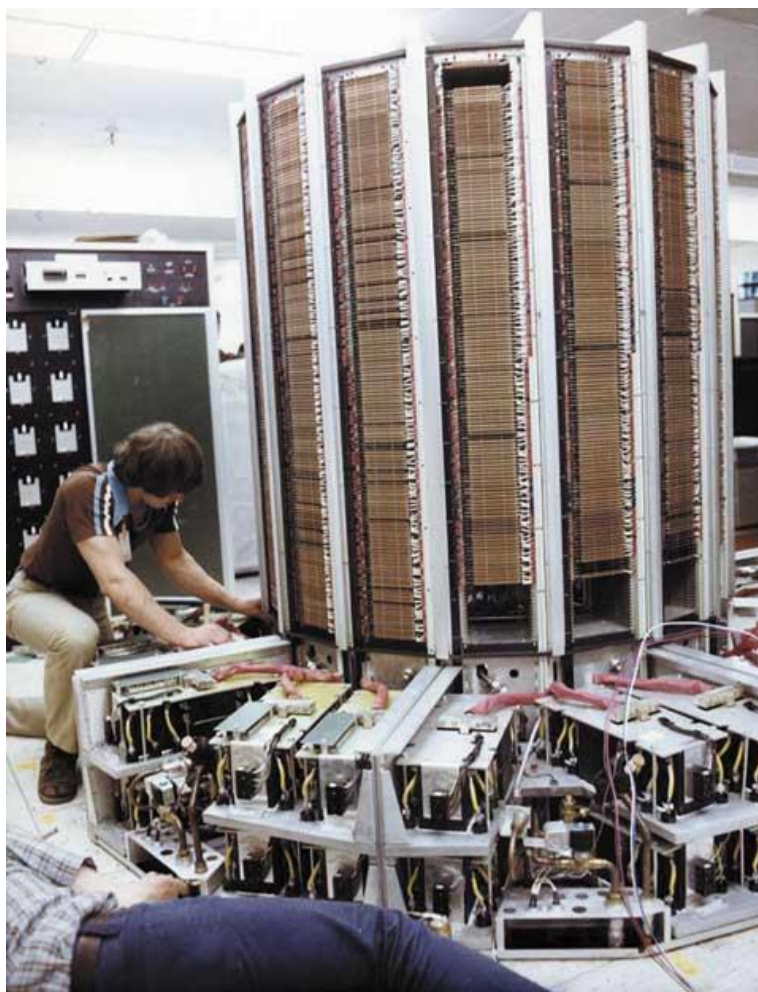
На волне успеха продаж Cray 1 Сеймур Крей создал не менее коммерчески успешную модель Cray 2. Тем временем конкуренция на рынке обострялась, и Крею пришлось сразу приступить к созданию модели Cray 3. Этот компьютер работал на невероятной по тем временам частоте – 500 МГц. Но это не самое удивительное. Используя собственные научные открытия в области полупроводников, Крей нашел замену кремнию – арсенид галлия. Это и позволило достичь максимальной скорости вычислений.

Но полоса везения кончилась. Не удалось продать ни одного экземпляра Cray 3. Модель Cray 4 работала уже на скорости один гигагерц. Но и ее не удалось не то что продать, даже довести до стадии коммерческой готовности. После серии неудач Сеймур Крей решил создать новую компанию – Cray Computer Corporation, но в 1996-м пришлось закрыть и ее.

Сеймур Крей не оставлял попыток возродить свою суперкомпьютерную компанию. Однако 22 сентября 1996 года в Колорадо он попал в тяжелейшую автокатастрофу и с многочисленными травмами и переломами был доставлен в госпиталь штата. За неделю врачам не удалось вывести его из критического состояния, и ночью 5 октября Сеймур Крей умер.



Сеймур Крей рядом с Cray-1



Процесс сборки Cray-1



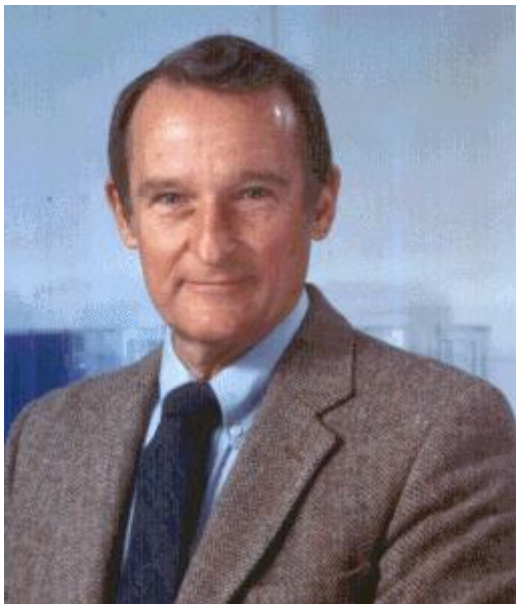
Система охлаждения Cray-1. Из-за огромного количества тепла, выделяемого системой, сотрудники Cray Research нередко использовали суперкомпьютер в качестве «обогревателя» в своих офисах в зимнее время



## Введение

Суперкомпьютер (англ. *supercomputer*, **СуперЭВМ**) — вычислительная машина, значительно превосходящая по своим техническим параметрам большинство существующих компьютеров. Как правило, современные суперкомпьютеры представляют собой большое число высокопроизводительных серверных компьютеров, соединённых друг с другом локальной высокоскоростной магистралью для достижения максимальной производительности в рамках подхода распараллеливания вычислительной задачи.

Началом эры суперкомпьютеров можно, пожалуй, назвать 1976 год, когда появилась первая векторная система Cray 1. Работая с ограниченным в то время набором приложений, Cray 1 показала настолько впечатляющие по сравнению с обычными системами результаты, что заслуженно получила название “суперкомпьютер” и определяла развитие всей индустрии высокопроизводительных вычислений еще долгие годы. Но более чем за два десятилетия совместной эволюции архитектур и программного обеспечения на рынке появлялись системы с кардинально различающимися характеристиками, поэтому само понятие “суперкомпьютер” стало многозначным и пересматривать его пришлось неоднократно.

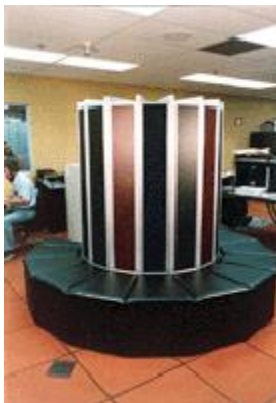


**Основатель Cray Research Сеймур Крэй.**

Корпорация Tera/Cray INC образована в марте 2000 года путем слияния Tera Computer Company и Cray Research. В новой компании со штаб-квартирой в Сиэтле занято около 900 человек. История корпорации началась в 1972 году, когда Сеймур Крэй (Seymour Cray) основал Cray Research. В 1987 году была основана Tera Computer Company.

Cray INC. проектирует, создаёт и продаёт высокопроизводительные MPP-системы, векторные процессоры и универсальные параллельные компьютерные системы такие, например, как серия Cray T90 — векторные процессорные системы; серия Cray SV1 — масштабируемые векторные процессорные системы, серия Cray T3E — масштабируемые параллельные суперкомпьютеры.

## Вычислительные системы



**CRAY-1.** Первый суперкомпьютер корпорации Cray. Выпущен в 1976 году. Пиковая производительность 133 Мфлопа. Первая система была установлена в Лос-Аламосской лаборатории.



**CRAY-2.** Векторно-параллельная система CRAY-2 выпускалась с 1982-го по 1992 годы. Эта система имела до 4-х процессоров и пиковую производительность 1,95 Гфлоп. В то время система CRAY-2 имела объем памяти в 2 048 Мбайт, который и в настоящее время считается достаточно большим.

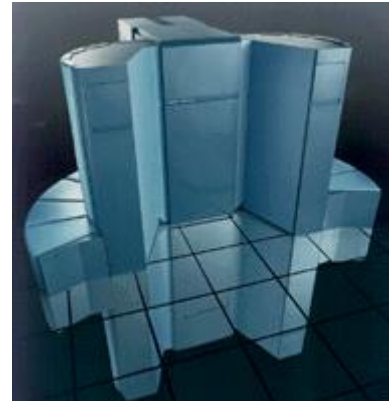
**CRAY-3.** Векторно-параллельная система выпущена в 1993 году. Число процессоров могло достигать 16 процессоров, пиковая производительность - до 16 Гфлоп/с. Выпуск завершён в 1995 году.



**CRAY T3D.** Это первая система из ряда масштабируемых параллельных суперкомпьютеров. Запущена в 1993 году.



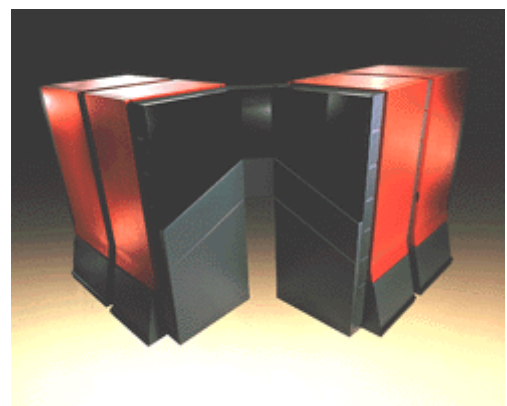
**CRAY X-MP.** Система выпущена в 1982 году и была первым многопроцессорным суперкомпьютером корпорации Cray. Производительность 500 Мфлоп. Впервые для суперкомпьютеров использовалась операционная система UNIX, UNICOS.



**CRAY Y-MP8D.** Система выпущена в 1988 году. Перешла барьер производительности в 1 Гфлоп. Система была стандартом, с которым обычно сравнивались другие суперкомпьютеры.



**CRAY Y-MP2E.** Первый суперкомпьютер Cray'я с воздушным охлаждением.



**CRAY Y-MP8E.** Как и система CRAY Y-MP8D, так и данная система содержала до 8-и векторных процессоров, но включала более быструю, улучшенную систему ввода/вывода, называемую Model E I/O.



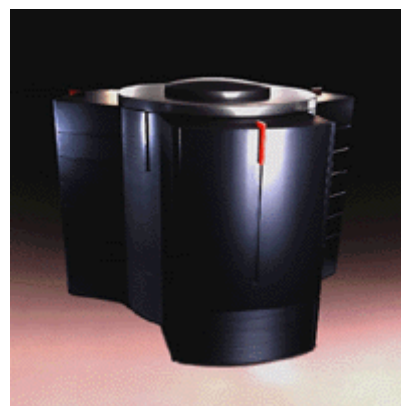


**CRAY EL.** Первый недорогой суперкомпьютер, являющийся предшественником серии CRAY J90.



**CRAY J90.** Векторно-параллельная система с общей памятью и воздушным охлаждением выпускалась с 1994-го по 1998 годы.

Максимальное число процессоров - 32.  
Пиковая производительность - 6,4 Гфлоп/с.



**CRAY C90.** Выпущенный в 1991 году CrayC90 стал первым высокопроизводительным векторно-параллельным суперкомпьютером компании. Число процессоров - до 16-и.  
Пиковая производительность - 16 Гфлоп/с.

## Краткие характеристики некоторых систем.

### **Cray Origin2000**

Производитель	Silicon Graphics
Класс архитектуры	Модульная система с общей памятью (cc-NUMA).
Процессор	64-разрядные RISC-процессоры MIPS R10000, R12000/300MHz
Модуль	Основной компонент системы - модуль Origin, включающий от 2 до 8 процессоров MIPS R10000 и до 16GB оперативной памяти.
Масштабируемость	Поставляются системы Origin2000, содержащие до 256 процессоров (т.е. до 512 модулей). Вся память системы (до 256GB) глобально адресуема, аппаратно поддерживается когерентность кэшей.
Коммутатор	Модули системы соединены с помощью сети CrayLink, построенной на маршрутизаторах MetaRouter.
Память	Система T932 содержит от 1GB до 8GB (до 1 GB в модели T94 и до 4GB в модели T916) оперативной памяти и обеспечивает скорость обменов с памятью до 800MB/sec.
Системное ПО	Используется операционная система SGI IRIX.
Средства программирования	Поставляется распараллеливающий компилятор Cray Fortran 90. Поддерживается стандарт OpenMP.

### **Cray T90**

Производитель	Cray Inc., Cray Research.
Класс архитектуры	Многопроцессорная векторная система (несколько векторных процессоров работают на общей памяти).
Предшественники	CRAY Y-MP C90, CRAY X-MP.
Модели	Серия T90 включает модели T94, T916 и T932.
Процессор	Системы серии T90 базируются на векторно-конвейерном процессоре Cray Research с пиковой производительностью 2GFlop/s.
Число процессоров	Система T932 может включать до 32 векторных процессоров (до 4-х в модели T94, до 16 модели T916), обеспечивая пиковую производительность более 60GFlop/s.
Масштабируемость	Возможно объединение нескольких T90 в MPP-системы.
Память	Система T932 содержит от 1GB до 8GB (до 1 GB в модели T94 и до 4GB в модели T916) оперативной памяти и обеспечивает скорость обменов с памятью до 800MB/sec.
Системное ПО	Используется операционная система UNICOS.

## Cray SV1

Производитель	Cray Inc.
Класс архитектуры	Масштабируемый векторный суперкомпьютер.
Процессор	Используются 8-конвейерные векторные процессоры MSP (Multi-Streaming Processor) с пиковой производительностью 4.8 GFLOP/sec; каждый MSP может быть подразделен на 4 стандартных 2-конвейерных процессора с пиковой производительностью 1.2 GFLOP/sec. Тактовая частота процессоров - 250MHz.
Число процессоров	Процессоры объединяются в SMP-узлы, каждый из которых может содержать 6 MSP и 8 стандартных процессоров. Система (кластер) может содержать до 32 таких узлов.
Память	SMP-узел может содержать от 2 до 16GB памяти. Система может содержать до 1TB памяти. Вся память глобально адресуема (архитектура DSM).
Системное ПО	Используется операционная система UNICOS.
Средства программирования	Поставляется векторизирующий и распараллеливающий компилятор CF90. Поддерживается также явное параллельное программирование с использованием интерфейсов MPI, OpenMP или Shmem.

## Cray X1

Производитель	Cray Inc.
Класс архитектуры	Масштабируемый векторный суперкомпьютер.
Процессор	Используются 16-конвейерные векторные процессоры с пиковой производительностью 12.8 GFLOP/sec. Тактовая частота процессоров - 800MHz.
Число процессоров	В максимальной конфигурации - до 4096.
Память	Каждый процессор может содержать до 16GB памяти. В максимальной конфигурации система может содержать до 64TB памяти. Вся память глобально адресуема (архитектура DSM). Максимальная скорость обмена с оперативной памятью составляет 34.1 Гбайт/сек. на процессор, скорость обмена с кэш-памятью 76.8 Гбайт/сек. на процессор.
Системное ПО	Используется операционная система UNICOS/mp.
Средства программирования	Реализованы компиляторы с языков Фортран и Си++, включающие возможности автоматической векторизации и распараллеливания, специальные оптимизированные библиотеки, интерактивный отладчик и средства для анализа производительности. Приложения могут писаться с использованием MPI, OpenMP, Co-array Fortran и Unified Parallel C (UPC).

## Cray XT3

Производитель	Cray Inc.
Класс архитектуры	Массивно-параллельный суперкомпьютер.
Процессор	Используются процессоры AMD Opteron.
Число процессоров	В максимальной конфигурации - до 30508.
Память	Каждый процессор может содержать от 1 до 8 Гбайт оперативной памяти. В максимальной конфигурации система может содержать до 239 Тбайт памяти.
Системное ПО	Используется операционная система UNICOS/lc.
Средства программирования	На компьютере устанавливаются компиляторы Fortran 77, 90, 95, C/C++, коммуникационные библиотеки MPI (с поддержкой стандарта MPI 2.0) и SHMEM, а также оптимизированные версии библиотек BLAS, FFTs, LAPACK,

	ScaLAPACK и SuperLU. Для анализа производительности системы устанавливается система Cray Apprentice <sup>2</sup> performance analysis tools.
--	--

## Cray XT4

Производитель	Cray Inc.
Класс архитектуры	Массивно-параллельный суперкомпьютер.
Процессор	Используются двухъядерные процессоры AMD Opteron.
Число процессоров	В максимальной конфигурации - до 30508.
Память	Каждый процессор может содержать от 1 до 8 Гбайт оперативной памяти. В максимальной конфигурации система может содержать до 239 Тбайт памяти.
Системное ПО	Используется операционная система UNICOS/lc.
Средства программирования	На компьютере устанавливаются компиляторы Fortran 77, 90, 95, C/C++, коммуникационные библиотеки MPI (с поддержкой стандарта MPI 2.0) и SHMEM, а также оптимизированные версии библиотек BLAS, FFTs, LAPACK, ScaLAPACK и SuperLU. Для анализа производительности системы устанавливается система Cray Apprentice <sup>2</sup> performance analysis tools.

## Cray T3E

Производитель	Cray Inc.
Класс архитектуры	Масштабируемая массивно-параллельная система, состоит из процессорных элементов (PE).
Предшественники	Cray T3D
Модификации	T3E-900, T3E-1200, T3E-1350
Процессорный элемент	PE состоит из процессора, блока памяти и устройства сопряжения с сетью. Используются процессоры Alpha 21164 (EV5) с тактовой частотой 450 MHz (T3E-900), 600 MHz (T3E-1200), 675 MHz (T3E-1350) пиковая производительность которых составляет 900, 1200, 1350 MFLOP/sec соответственно. Процессорный элемент располагает своей локальной памятью (DRAM) объемом от 256MB до 2GB.
Число процессоров	Системы T3E масштабируются до 2048 PE.
Коммутатор	Процессорные элементы связаны высокопроизводительной сетью GigaRing с топологией трехмерного тора и двунаправленными каналами. Скорость обменов по сети достигает 500MB/sec в каждом направлении.
Системное ПО	Используется операционная система UNICOS/mk.
Средства программирования	Поддерживается явное параллельное программирование с помощью пакета Message Passing Toolkit (MPT) - реализации интерфейсов передачи сообщений MPI, MPI-2 и PVM, библиотека Shmem. Для Фортран-программ возможно также неявное распараллеливание в моделях CRAFT и HPF. Среда разработки включает также набор визуальных средств для анализа и отладки параллельных программ.

## Подробное описание некоторых систем

### 1) CRAY T90



Векторно-параллельная серия Cray T90 включает в себя три модели: систему Cray T94, поставляемую с воздушным и жидкостным охлаждением и масштабируемую до 4-х процессоров; систему Cray T916 с жидкостным охлаждением и масштабируемую до 16-и процессоров; систему Cray T932 также с жидкостным охлаждением и масштабируемую до 32-х процессоров.

Система T90 выпускалась с 1995 по 1998 годы.

#### Аппаратное обеспечение

Наименование модели	Cray T94	Cray T916	Cray T932
Число процессоров	1 – 4	8 – 16	16 – 32
Пиковая производительность, Гфлоп	1,8 – 7,2	14 – 28	28 – 56
Оперативная память, Гбайт	0,5 – 2	2 – 8	4 – 16
Пиковая пропускная способность памяти, Гбайт/с	100	450	900
Пиковая пропускная способность ввода/вывода, Гбайт/с	8	16	32
Охлаждение	Воздушное или жидкостное	Жидкостное	Жидкостное

#### Основные характеристики

- Система основана на векторных процессорах Cray и предназначена для решения научных, технических и инженерных приложений.
- Система расширяется до 32-х процессоров и имеет до 60 Гфлоп пиковой производительности.
- Пропускная способность памяти (memory bandwidth) около 1 Тбайт суммарно.
- Центральная память (central memory) от 512 Мбайт до 16 384 Мб.
- Совокупная пропускная способность ввода/вывода (aggregate I/O bandwidth) около 30 Гбайт/с.
- 64-х разрядный IEEE формат данных с плавающей запятой.
- Двоичная совместимость с предыдущей системой Cray C90.
- Оперативная система UNICOS, основанная на UNIX System V.



## Система Cray T94

Компьютерная система Cray T94 — это суперкомпьютер с высокоскоростными процессорами, каждый из которых имеет пиковую производительность приблизительно 2 Гфлоп/с. Также как и в базовой системе Cray T90, система Cray T94 имеет до 4-х процессоров, 2 048 Мбайт центральной памяти и пиковую производительность приблизительно 8 Гфлоп.

Система Cray T94 включает в себя специально изготовленные высокоскоростные процессоры, быстродействующую статическую оперативную память и подсистему ввода/вывода с высокой пропускной способностью. Такая комбинация быстрых процессоров, быстрой памяти и быстрого ввода/вывода обеспечивает высокую производительность для большинства сложных и разнообразных по объёму работ.

Система Cray T94 поддерживает различные соединения вида ATM (Asynchronous Transfer Mode - асинхронный режим передачи), FDDI (Fiber Distributed Data Interface - распределённый интерфейс передачи данных по волоконно-оптическим каналам) и HIPPI (High-Performance Parallel Interface - высокоскоростной параллельный интерфейс). Поддержка технологии дисководов включает IPI дисководы (Intelligent Peripheral Interface - интеллектуальный интерфейс периферийных устройств) и, в ближайшем будущем, SCSI (small computer systems interface - интерфейс малых компьютерных систем) и оптические диски с максимальным объёмом памяти около 64 Тбайт. Система также поддерживает технологию жёстких дисков SSD (solid-state disk).

### Основные особенности системы Cray T94:

- Полная двоичная совместимость с серией Cray C90.
- От одного до четырёх процессоров.
- Приблизительно 8 Гфлоп пиковой производительности.
- От 512 до 2 048 Мбайт центральной памяти.
- Пропускная способность памяти около 100 Гбайт/с.
- Общая пропускная способность ввода/вывода около 8 Гбайт/с.
- Дополнительный твердотельный диск от 4 096 до 16 384 Мбайт.
- Оперативная система UNICOS, основанная на системе UNIX System V.

## Система Cray T916

Компьютерная система Cray T916 - это суперкомпьютер с высокоскоростными процессорами, каждый из которых имеет пиковую производительность приблизительно 2 Гфлоп/с. Cray T916 имеет до 16-и процессоров, 8 192 Мбайт центральной памяти и пиковую производительность приблизительно 32 Гфлоп.

Система Cray T94 поддерживает различные соединения вида ATM, FDDI и HIPPI. Поддержка технологии дисководов включает IPI дисководы и, в ближайшем будущем, SCSI и оптические диски с максимальным объёмом памяти около 128 Тбайт. Система также поддерживает технологию жёстких дисков SSD (solid-state disk).

### **Основные особенности системы Cray T916:**

- Полная двоичная совместимость с серией Cray C90.
- До 16 процессоров.
- Приблизительно 32 Гфлоп пиковой производительности.
- От 2 048 до 8 192 Мбайт центральной памяти.
- Пропускная способность памяти около 400 Гбайт/с.
- Общая пропускная способность ввода/вывода около 17 Гбайт/с.
- Дополнительный жёсткий диск от 4 096 до 16 384 Мбайт.
- Оперативная система UNICOS, основанная на системе UNIX System V.

### **Система Cray T932**

Компьютерная система Cray T932 - это суперкомпьютер с высокоскоростными процессорами, каждый из которых имеет пиковую производительность приблизительно 2 Гфлоп/с. Cray T932 имеет до 32-х процессоров, 16 384 Мбайт центральной памяти и пиковую производительность приблизительно 60 Гфлоп.

Система Cray T932 поддерживает различные соединения вида ATM, FDDI и HIPPI. Поддержка технологии дисководов включает IPI дисководы и, в ближайшем будущем, SCSI и оптические диски с максимальным объёмом памяти около 256 Тбайт. Система также поддерживает технологию жёстких дисков SSD (solid-state disk). Система Cray T932 предельно масштабируема: при добавлении большего числа процессоров производительность увеличивается почти линейным образом.

### **Основные особенности системы Cray T932:**

- Полная двоичная совместимость с серией Cray C90.
- 8, 16 или 32 процессора.
- Приблизительно 60 Гфлоп пиковой производительности.
- От 1 024 до 16 384 Мбайт центральной памяти.
- Пропускная способность памяти около 800 Гбайт/с.
- Общая пропускная способность ввода/вывода около 35 Гбайт/с.
- Дополнительный жёсткий диск от 4 096 до 16 384 Мбайт.
- Оперативная система UNICOS, основанная на системе UNIX System V.

## 2) Суперкомпьютер Cray X1



Суперкомпьютер Cray X1 сочетает в себе характеристики векторно-конвейерных и массивно-параллельных компьютеров, являясь наследником как линии компьютеров Cray T3E, так и Cray SV1. Он предоставляет исключительные возможности векторной обработки, высокие скорости обменов с памятью и межпроцессорные обмены с низкой латентностью. В 21-ю редакцию списка мощнейших компьютеров мира Top500 (июнь 2003 года) вошло 10 суперкомпьютеров Cray X1. Фирма Cray Inc. ставит целью достижение к 2010 году реальной производительности компьютера в 1 Pflops.

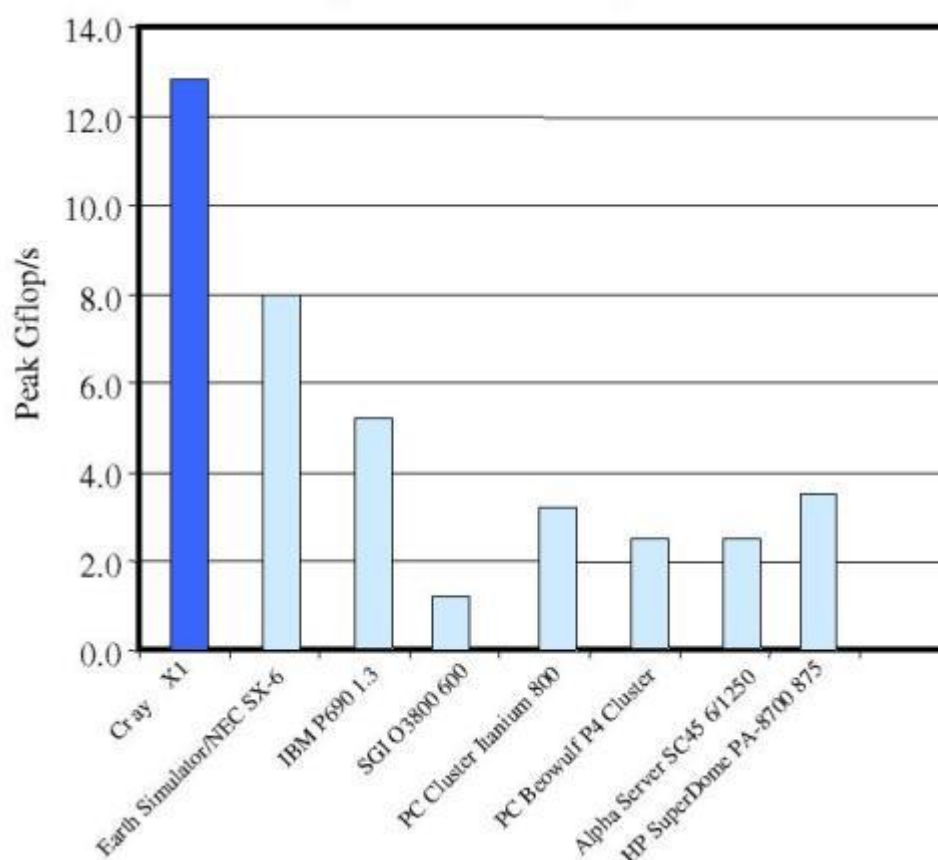
Cray X1 в максимальной конфигурации объединяет 4096 векторных процессоров с пиковой производительностью 52.4 Tflops и объемом оперативной памяти 65536 Гбайт. Стоимость минимальной конфигурации компьютера при начале выпуска (ноябрь 2002 года) составила \$2.5 млн. Возможные конфигурации Cray X1 приведены в следующей таблице:

Число стоек	Число процессоров	Объем памяти	Пиковая производительность
1*	16	64-256 Гбайт	204.8 Gflops
1	64	256-1024 Гбайт	819.0 Gflops
4	256	1024-4096 Гбайт	3.3 Tflops
8	512	2048-8192 Гбайт	6.6 Tflops
16	1024	4096-16384 Гбайт	13.1 Tflops
32	2048	8192-32768 Гбайт	26.2 Tflops
64	4096	16384-65536 Гбайт	52.4 Tflops

\* Конфигурация с воздушным охлаждением.

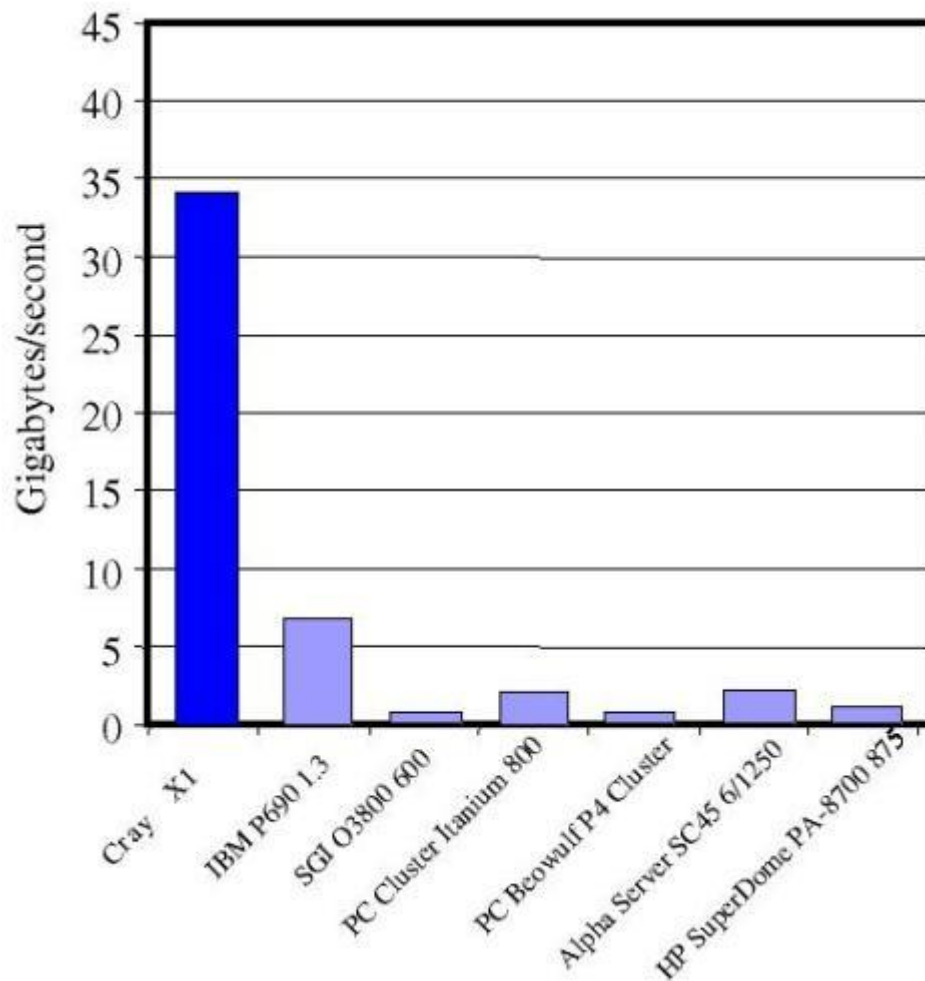
Каждый векторный процессор Cray X1 работает на тактовой частоте 800 МГц и способен выдавать до 16 результатов операций с плавающей точкой за такт, что дает в пике 12.8

Gflops на процессор. Таким образом, это один из наиболее мощных существующих на середину 2003 года процессоров:



Реальная производительность на тесте Linpack при использовании от 4 до 60 процессоров составила (в пересчете на один процессор) 11.55 Gflops, что составляет эффективность использования процессора около 90%.

Оперативная память компьютера Cray X1 является физически распределенной, но логически общей для всех процессоров системы. Таким образом, поддерживается модель DSM (Distributed Shared Memory). Реализован также эффективный протокол поддержки когерентности кэш-памяти. Максимальная скорость обмена с оперативной памятью составляет 34.1 Гбайт/сек. на процессор, скорость обмена с кэш-памятью 76.8 Гбайт/сек. на процессор, что намного превосходит показатели других современных компьютеров:



Топология связи узлов компьютера Cray X1 представляет собой модифицированный двумерный тор, общая пропускная способность 400 Гбайт/сек. для 64-процессорной конфигурации.

Cray X1 работает под управлением операционной системы UNICOS/mp, являющейся продолжением линии предыдущих ОС UNICOS и UNICOS/mk. UNICOS/mp может работать на системе, объединяющей до 4096 процессоров и администрируется с одного узла. Пользователи могут работать на компьютере как в интерактивном режиме, так и через специальную систему очередей.

Пользователям предоставляется возможность написания программ как в модели распределенной, так и общей памяти. Реализованы компиляторы с языков Фортран и Си++, включающие возможности автоматической векторизации и распараллеливания, специальные оптимизированные библиотеки, интерактивный отладчик и средства для анализа производительности. Приложения могут создаваться с использованием MPI, OpenMP, Co-array Fortran и Unified Parallel C (UPC).



### 3) Суперкомпьютер Cray XT4



#### Общая структура

Суперкомпьютер XT4 является дальнейшим развитием системы Cray XT3, отличаясь от неё в основном использованием двухъядерных процессоров AMD Opteron. Cray XT4 состоит из следующих компонентов:

- процессорные элементы, в каждом из которых есть процессор и память;
- коммуникационная сеть с малой латентностью и высокой пропускной способностью;
- оптимизированная версия операционной системы;
- специальные средства управления и мониторинга системы;
- высокоскоростные подсистемы ввода/вывода.

#### Варианты компоновки системы

Вычислительные узлы Cray XT4 компонуются в стойки (до 96 вычислительных узлов на стойку). Возможные конфигурации приводятся в следующей таблице:

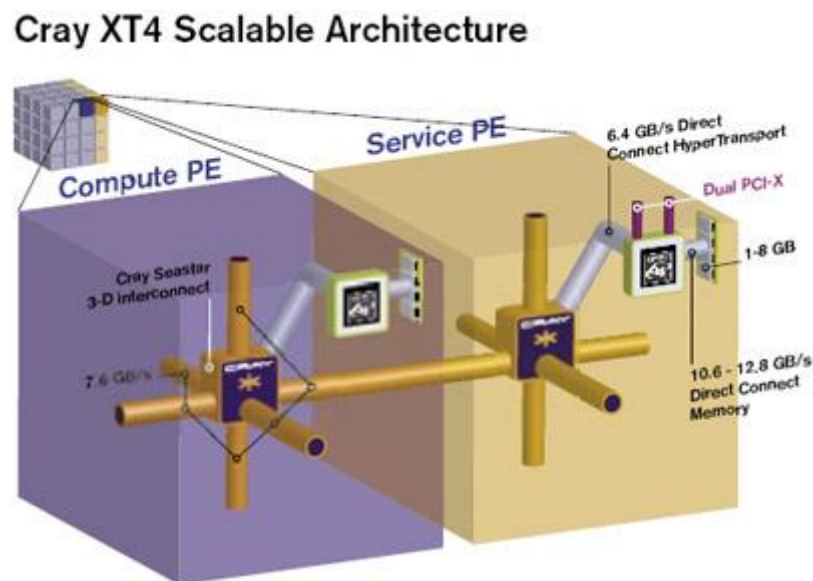
Число стоек	Число процессоров	Объем памяти	Пиковая производительность*	Топология
6	548	4.3 Тбайт	5.6 TFlop/s	6*12*8
24	2260	17.7 Тбайт	23.4 TFlop/s	12*12*16
96	9108	71.2 Тбайт	94.6 TFlop/s	24*16*24
320	30508	239 Тбайт	318 TFlop/s	40*32*24

\* Пиковая производительность рассчитывалась для конфигураций, использующих двухъядерный процессор AMD Opteron 2.6 ГГц.

## Масштабируемые процессорные элементы

Как и в предыдущих MPP-системах компании Cray, базовым элементом систем Cray XT4 является однопроцессорный узел (процессорный элемент, ПЭ). Каждый процессорный элемент состоит из одного микропроцессора AMD (одно-, двух- или четырехъядерного), имеющего собственную память и средства связи с другими ПЭ. Такая архитектура устраняет проблему асимметричности, которая есть в кластерах SMP-узлов, поскольку производительность приложений будет одинаковой вне зависимости от конкретного распределения процессов по процессорам, что важно для обеспечения масштабируемости.

В системах Cray XT4 есть два типа процессорных элементов: вычислительные и сервисные.



На вычислительных ПЭ работает легковесное ядро операционной системы UNICOS, имеющее минимальные накладные расходы, основной задачей которого является эффективная поддержка работы приложений. На сервисных ПЭ установлена ОС Linux, а сами сервисные ПЭ могут быть настроены для выполнения сетевых, системных функций, для идентификации пользователей, а также для выполнения функций ввода/вывода.

Конструктивно 4 вычислительных ПЭ объединяются в Cray XT4 в один вычислительный сервер-лезвие, что обеспечивает хорошую масштабируемость системы при очень небольшом объеме. Сервисные серверы-лезвия содержат по 2 сервисных ПЭ и поддерживают прямую связь с устройствами ввода/вывода.

## Процессоры AMD Opteron

Кэш-память данных с высокой степенью ассоциативности, расположенная на кристалле процессора AMD, поддерживает мощные алгоритмы внеочередного выполнения команд и может выдавать до 9 инструкций одновременно. Встроенный контроллер памяти устраняет необходимость в использовании отдельного чипа контроллера памяти типа Northbridge, и при этом обеспечивает очень низкую латентность при доступе к локальной памяти - менее 60 наносекунд, что является существенным подспорьем для эффективного выполнения программ, особенно для алгоритмов с неоднородным доступом к памяти. Контроллер памяти с шириной тракта 128 бит обеспечивает каждому процессору AMD

Opteron доступ к локальной памяти со скоростью от 10.6 до 12.8 ГБ/с, т.е. более одного байта на каждый Флоп. Такое соотношение позволяет получить высокую производительность на алгоритмах, для которых необходим интенсивный доступ к памяти.

Технология HyperTransport обеспечивает прямое соединение со скоростью 6.4 ГБ/с между процессором и коммуникационной сетью Cray XT4, устраняя распространенное узкое место, которым в большинстве сетей является интерфейс PCI.

## **Память**

Каждый ПЭ Cray XT4 может использовать от 1 до 8 ГБ оперативной памяти DDR2. Поскольку в вычислительных ПЭ память не буферизуется, задержка при доступе к ней минимальна.

## **Масштабируемая коммуникационная сеть**

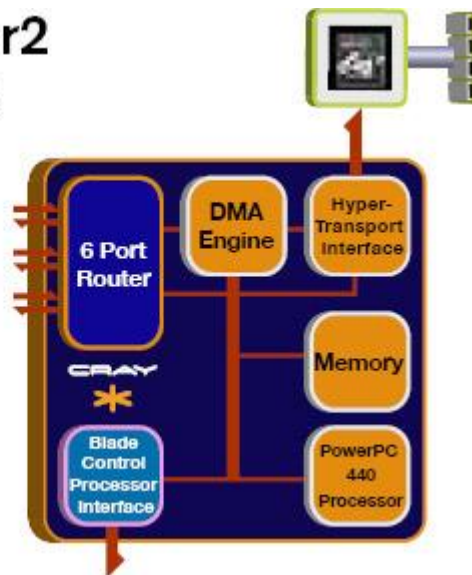
В системах Cray XT4 используется коммуникационная сеть с высокой пропускной способностью и малой латентностью. Она построена на основе микросхем Cray SeaStar2 и высокоскоростных соединений, использующих технологию HyperTransport.

Коммуникационная сеть соединяет все процессорные элементы согласно топологии трехмерного тора, исключая необходимость использования коммутаторов. Такой подход повышает надежность системы и позволяет без особых затрат увеличивать количество узлов до десятков тысяч, что намного превышает возможности архитектур с переключателями типа fat-tree. Поскольку эта сеть является основным коммуникационным каналом систем Cray XT4, она отвечает и за все проходящие сообщения, и за весь входящий/исходящий трафик, связанный с глобальной файловой системой.

## **Микросхема Cray SeaStar2**

Микросхема Cray SeaStar2 выполняет две функции: обработку данных, поступающих по каналам связи, и высокоскоростную маршрутизацию. Каждая микросхема состоит из канала HyperTransport, модуля прямого доступа к памяти (DMA), процессора передачи данных и управления, высокоскоростного маршрутизатора сети и сервисного порта.

# Cray SeaStar2 Architecture



Маршрутизатор микросхемы SeaStar2 создает 6 высокоскоростных соединений с шестью непосредственными соседями по топологии трехмерного тора. Пиковая пропускная способность каждого канала при двусторонней передаче данных - 7.6 ГБ/с. Маршрутизатор использует протокол связи, поддерживающий контроль и коррекцию ошибок, а также повторную передачу.

Микросхема Cray SeaStar2 включает в себя модуль DMA и связанный с ним процессор PowerPC 440. Их задачей является снятие с процессора функций подготовки и демультиплексирования сообщений, что позволяет сосредоточить все его ресурсы исключительно на вычислениях. Логика SeaStar2 соответствует логике отправки и получения MPI-сообщений, поэтому нет необходимости использовать большие буфера памяти, которые необходимы при работе обычных кластерных систем. Модуль DMA и операционная система Cray XT4 позволяют минимизировать латентность путем установления прямой связи между приложением и аппаратурой сети, минуя проход через ядро ОС и связанную с этим систему прерываний.

## Надежность коммуникационной сети

Каждый из шести каналов микросхемы использует протокол, поддерживающий вычисление контрольной суммы и автоматическую повторную передачу. При наличии неустойчивого соединения канал может быть запущен в режиме с сокращенным набором ресурсов, но при этом соединение не будет разорвано.

Микросхема Cray SeaStar2 содержит сервисный порт, который соединяет независимую сеть управления и локальную шину Cray SeaStar2. Этот порт обеспечивает системе управление доступом к памяти и ко всем регистрам компьютера, облегчает его начальную загрузку, поддержку текущей работы и мониторинг.

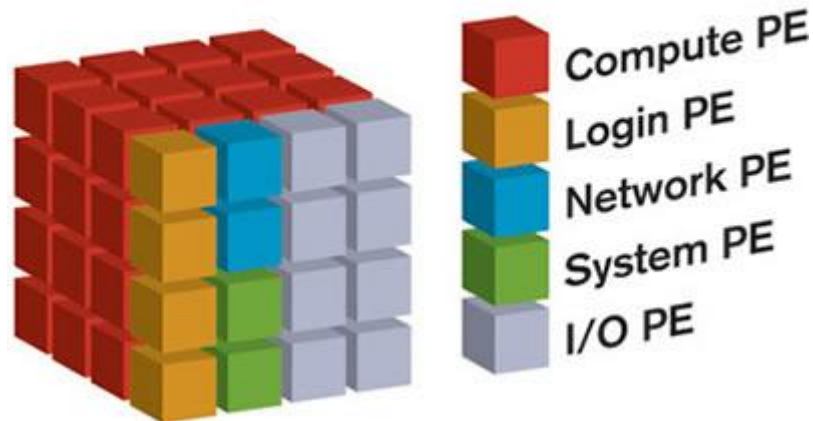
## Масштабируемая операционная система

Операционная система UNICOS/Ic систем Cray XT4 разработана для выполнения ресурсоемких приложений при масштабировании до 120000 процессоров. Как и в предыдущих поколениях MPP-систем компании Cray, UNICOS/Ic состоит из двух основных

компонентов - микроядра для вычислительных ПЭ и полнофункциональной операционной системы для сервисных ПЭ.

Микроядро взаимодействует с процессом приложения лишь ограниченным набором функций, такими как управление адресацией виртуальной памяти, предоставление средств безопасности и обеспечение базовых средств планирования. Микроядро является легковесным, причем никаких дополнительных слоев между приложением и аппаратурой нет.

## UNICOS/lc Architecture



На сервисных ПЭ, как правило, работает полнофункциональная версия ОС Linux. Процессорные элементы этого типа могут быть настроены на выполнение сетевых или системных служб, а также функций ввода/вывода и идентификации пользователей в системе.

ПЭ, предназначенные для идентификации, предоставляют программисту систему с полным доступом к среде разработки и ко всем стандартным утилитам, командам и консолям Linux. Это сделано для создания привычного окружения, удобства разработки нового и переносимости уже существующего ПО. Сетевые ПЭ обеспечивают высокоскоростное соединение с другими системами, а ПЭ ввода/вывода поддерживают масштабируемую связь с глобальной параллельной файловой системой. Системные ПЭ используются для поддержки работы глобальных системных служб, таких как работа с базами данных. Эти службы могут масштабироваться до размеров всей системы или лишь до того уровня, который необходим пользователям.

Приложения пользователей передаются на исполнение интерактивно через ПЭ идентификации с помощью явных команд запуска заданий Cray XT4 или с помощью системы пакетной обработки PBS Pro, тесно связанной с планировщиком системных ПЭ. Для выполнения каждого приложения на компьютере выделяется отдельный набор вычислительных ПЭ, не пересекающийся с наборами, выделенными для других приложений. Конфигурации разделов для пакетной обработки и интерактивного исполнения заданий определяются системным администратором.

Системы Cray XT4 поддерживают единую файловую систему с общим по всем узлам корневым разделом, поэтому нет необходимости пересылать изменения в каждый ПЭ отдельно, поскольку изменения становятся видны сразу во всей системе.



## Масштабируемая среда разработки

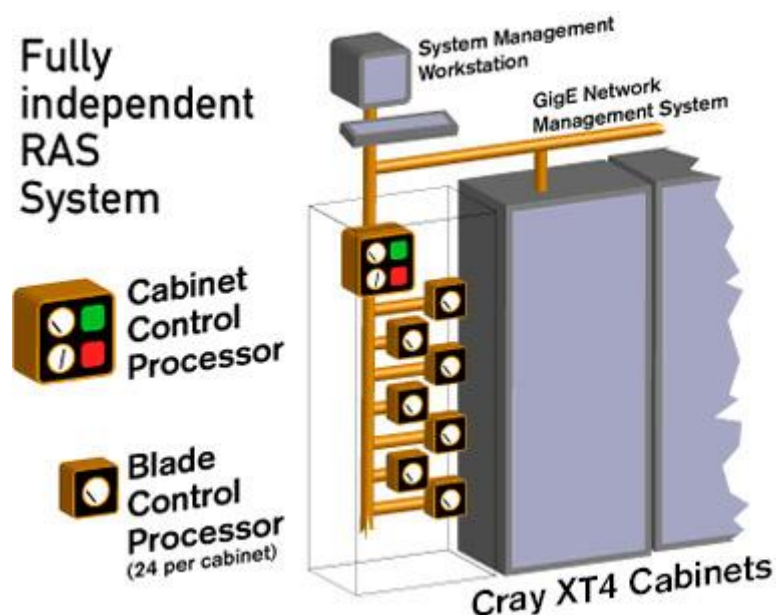
Среда разработки систем Cray XT4 включает в себя программные инструменты, которые хорошо дополняют друг друга, помогая в разработке масштабируемых приложений. Поскольку процессоры AMD поддерживают 32- и 64-битные приложения и полностью совместимы с линейкой x86-64, на системах Cray XT4 работает большинство известных компиляторов и библиотек, в том числе оптимизирующие компиляторы для C, C++, Fortran90, а также многие математические библиотеки, например, оптимизированные версии BLAS, FFTs, LAPACK, ScaLAPACK и SuperLU.

Библиотеки передач сообщений включают MPI и SHMEM. Реализация MPI соответствует стандарту 2.0 и оптимизирована под особенности коммуникационной сети систем Cray XT4. Пропускная способность по количеству проходящих сообщений может меняться в зависимости от размеров системы, достигающей десятков тысяч ПЭ. Библиотека SHMEM совместима с предыдущими версиями систем Cray и использует все возможности микросхемы Cray SeaStar2 для минимизации накладных расходов.

В составе штатного ПО систем Cray XT4 традиционно присутствуют средства анализа производительности семейства Cray Apprentice<sup>2</sup>. Они позволяют пользователям оценивать использование ресурсов компьютера их программами и помогают обнаружить многие узкие места, в частности, несбалансированность загрузки процессоров.

## Система управления и мониторинга

Система управления и мониторинга (CRMS) объединяет аппаратные и программные компоненты для мониторинга состояния компьютера, обнаружения неисправностей и восстановления после сбоев. Будучи независимой системой с собственными управляющими процессорами и выделенной сетью, CRMS следит и управляет всеми основными аппаратными и программными компонентами систем Cray XT4. Помимо функций восстановления после сбоев, CRMS контролирует процессы включения, выключения и начальной загрузки системы, управляет коммуникационной сетью и отображает состояние компьютера.



Все критические компоненты систем Cray XT4 продублированы, чтобы минимизировать количество сбоев. Если из-за неполадок перестанет работать ПЭ ввода/вывода, это не скажется на задании, использовавшем этот ПЭ. Могут выйти из строя процессор или локальная память некоторого ПЭ, но это никак не повлияет на задания, использующие этот узел для транзита сообщений. На системных платах нет ни одной двигающейся компоненты, что также повышает надежность системы в целом.

Процессор и платы ввода/вывода систем Cray XT4 используют компоненты, подключаемые через сокет, поэтому микросхемы SeaStar2 и DIMM, модули контроля напряжения и процессоры AMD в случае необходимости могут быть легко заменены и модернизированы.

### **Масштабируемые средства ввода/вывода**

Подсистема ввода/вывода компьютеров Cray XT4 хорошо масштабируется, подстраиваясь под требования самых тяжелых приложений. Архитектура подсистемы включает дисковые массивы, которые напрямую соединены с ПЭ ввода/вывода через высокоскоростную сеть. Файловая система Lustre управляет распределением операций с файлами по всему хранилищу. Масштабируемая архитектура ввода/вывода позволяет в каждой конфигурации компьютера получить требуемую пропускную способность путем подбора соответствующего числа массивов и сервисных ПЭ.

Для повышения производительности средств ввода/вывода модули Lustre встраиваются прямо в приложения и выполняются на системном микроядре. Данные передаются напрямую из приложений и серверов Lustre на процессорные элементы ввода/вывода, без необходимости создания промежуточной копии данных в самом ядре.

### **4) CRAY SV1**



Cray SV1 (или Tera/Cray SV1) — суперкомпьютер нового уровня по масштабируемости и

производительности. Процессор Cray SV1 имеет пиковую производительность 4,8 Гфлоп. Путем масштабируемости от отдельного процессора можно перейти к сотням процессоров с производительностью до Терафлоп.

Система Cray SV1 является преемником как Cray J90, основанным на CMOS-технологии, так и Cray T90, который основан на ECL-технологии.

Выпуск системы начался в 1998 году. Выпускаются модели SV1-A, SV1-1 и SV1 Supercluster.

## Основные характеристики

- Пиковая производительность до 1 Тфлоп и масштабируемость памяти до 1 Тбайта.
- Особенности двух типов процессоров: производительность многопоточного процессора (*Multi Streaming Processor* - MSP) - 4,8 Гфлоп; стандартного процессора (*The Single-streaming processor* - SSP) — 1,2 Гфлоп.
- В максимальной конфигурации отдельный узел может содержать 6 MSP и 8 SSP процессоров. Система масштабируется до 32-х узлов.
- Используются CMOS и DRAM (*Dynamic Random Access Memory*) технологии.
- Используется операционная система UNICOS (вариант Cray Unix).
- Транслятор CF90. Обеспечивает автоматическую скалярную, векторную и параллельную оптимизацию. Включает в себя библиотеки и инструментарий.
- Трансляторы Pascal, ADA, Cray C++ и C.

Каждый узел Cray SV1 может быть сконфигурирован с двумя типами векторных процессоров. Два типа процессора обеспечивают на рынке максимальную гибкость. Отдельный потоковый процессор (*Single-streaming processor* — SSP) с пиковой производительностью 1,2 Гфлопа удовлетворяет всем требованиям для прикладных и вычислительных задач. Многопоточный процессор (*Multi-Streaming Processor* - MSP), имеющий производительность 4,8 Гфлопа, может решать крупные задачи. С помощью комбинаций процессоров можно решать задачи разнообразными методами, не применявшимися ранее.

## Масштабируемость



CRAY SV1 CLUSTER

Суперкомпьютер Cray SV1 масштабируется до 32-х систем с пиковой производительностью до 1 Тфлоп и масштабируемости памяти до 1 Тбайта. Каждая система в кластере является UNICOS-системой с высокопроизводительной памятью со сплошной адресацией (*flat memory*). Внутри каждой системы процессоры объединяются в симметричную многопроцессорную конфигурацию. Все текущие приложения могут запускаться на отдельном узле без модификации.

## Аппаратное обеспечение

Модель	Cray SV1-1A	Cray SV1-1	Cray SV1-4	Cray SV1-8	Cray SV1-32
Число узлов	1	1	4	8	32
Пиковая производительность, Гфлоп	9,6 – 19	9,6 – 38	38 – 154	77 – 308	308 – 1229
Оперативная память, Гбайт	2 – 16	4 – 32	16 – 128	32 – 256	128 – 1024
Число процессоров: 4,8 Гфлоп +1,2 Гфлоп Все по 1 Гфлоп	До 3 4+ 9,6 – 19	До 6 8+ 9,6 – 38	до 24 32+ 38 – 154	до 48 64+ 77 – 308	до 192 256+ 308 – 1229
Тактовая частота, МГц	300	300	300	300	300
Память	DRAM	DRAM	DRAM	DRAM	DRAM
Охлаждение	Воздушное	Воздушное	Воздушное	Воздушное	Воздушное или водяное

Конфигурации в отдельных стойках представлены в двух моделях: SV1-A и SV1-1, которые имеют 4-х и 8-и процессорные платы, соответственно. Каждая процессорная плата содержит 4 процессора, которые могут давать пиковую скорость в четыре операции с плавающей запятой за цикл, доходя до пиковой производительности в 1,2 Гфлоп/с на процессор. Однако 4 процессора могут быть объединены в платы процессоров, соединенных попарно параллельно к так называемой форме MSP (*Multi Streaming Processor* - многопоточный процессор). В SV1 объединены скалярный и векторный кэши объемом в 256 Кбайт. Этот кэш является важным из-за того, что пропускная способность в 7,7 Гбайт/с на процессорной плате составляет 0,8 восьмибайтных операндов за цикл. Кэш может отправить 1 операнд за цикл на процессор. Как и в системе NEC SX-5 отдельные стойки могут быть объединены в форме кластера (в терминологии Cray'я - суперкластер) посредством так называемого интерфейса GigaRing. Если системы, представленные в стойках, являются системами SM-MIMD, то суперкластер, составленный из таких стоек, является системой DM-MIMD и может управляться только под параллельными языками программирования как MPI или HPF.

**Производительность системы** проиллюстрируем на примере перемножения матриц по тесту EuroBen Benchmark. При расчётах, проведённых на одном SSP-процессоре и с кэшем, пиковая скорость для матриц 300-го порядка составила 999 Мфлоп/с и уменьшилась до скорости в 666 Мфлоп/с для матриц порядка 1000. Без кэша скорость для матриц порядка  $n=300$  достигла 625 Мфлоп/с и слегка увеличилась до 650 Мфлоп/с для матриц порядка 1000. В моде MSP такое поведение подобно: с кэшем скорость для матриц 100-го порядка достигла 2,61 Гфлоп/с и уменьшилась до 1,41 Гфлоп/с для  $n=1000$ . Без кэша полученная скорость для  $n=100$  составила 1 Гфлоп/с и увеличилась до 1,4 Гфлоп/с при  $n=1000$ . Это означает, что для задач с умеренными размерностями кэш может повышать производительность с множителем 1,5-2. Эффективность MSP-моды является сейчас не слишком высокой: только около 50% в благоприятной ситуации. Ввиду того, что особенность MSP введена в употребление недавно, можно ожидать, что в ближайшем будущем её эффективность будет значительно увеличена.

## 5) CRAY SV2

В настоящее время идет работа по созданию новой суперкомпьютерной системы Cray SV2, о чём было объявлено в сентябре 1999 года. В ней будут значительно расширены возможности суперкомпьютеров: будет обеспечена исключительная пропускная способность памяти, внутренних соединений и производительность векторной обработки. Суперкомпьютер будет иметь мощные векторные процессоры с архитектурой cсNUMA. Планируется достижение пиковой производительности многих десятков Тфлоп.

В дополнение к государственным приложениям в критических технологиях в области высокопроизводительного компьютеринга Cray SV2 будет нацелен на решение задач в таких областях как автомобильный дизайн, аэрокосмическое проектирование, предсказание погоды и климата, академические исследования.

## 6) Cray MTA



### **CRAY MTA-2**

Cray MTA - это масштабируемая параллельная система с многопоточной архитектурой и однородной общей памятью. Программное окружение для распараллеливания задач использует трансляторы с языков Фортрана и Си: Fortran 77, C, C++. Идёт подготовка к развёртыванию транслятора Fortran 90.

### **Основные характеристики**

- Число процессоров от 16 до 256.
- 64-х разрядные данные, адресация и команды.
- До 128-и потоков на процессор.
- Параллельная (совместная) ссылка на ячейку памяти на поток.
- Арифметика с плавающей запятой в стандарте IEEE 754.
- Для кэша команд уровня 1 - 8 Кб, для уровня 2 - 2 Мб.
- Теоретическая пиковая производительность:

На процессор (64 разряда) - 1 Гфлоп/с;  
Максимальная (64 разряда) - 256 Гфлоп/с.

- Оперативная система Unix BSD4.4+.
- Автоматическое распараллеливание и векторизация.
- Трансляторы Fortran 77, Fortran 90 (в разработке), HPF, ANSI C и C++.
- Межпроцедурный анализ и оптимизация.
- Символическая отладка оптимизированного параллельного кода.
- Дисковый массив объёмом в 200-800 Гбайт со скоростью в 256 Мбайт/с.
- HIPPI (высокоскоростной параллельный интерфейс) по протоколу TCP/IP может взаимодействовать с протоколами FDDI и ATM.



- Многопользовательская поддержка больших и малых заданий.
- Водяное охлаждение мощностью 4 Кватт на процессор.

Хотя память в МТА физически распределённая, но система настойчиво представляется как машина с общей памятью (с неоднородным (распределённым) временем доступа).

Системы МТА построены из ресурсных модулей. Каждый ресурсный модуль содержит вычислительный процессор; процессор ввода/вывода, соединяемый с устройством ввода/вывода через HIPPI; устройства памяти. Каждый МТА-процессор имеет до 128 RISC-подобных виртуальных процессоров.

**Производительность** измерялась на 16-и процессорной системе в суперкомпьютерном центре Сан Диего (США). Эта система запускалась при тактовом цикле в 4,4 нс вместо планируемого в 3 нс. Поэтому пиковая производительность процессора составляла 450 Мфлоп/с. По тесту EuroBen Benchmark при перемножении матриц 800-го порядка производительность достигла 388 Мфлоп/с, что составляет эффективность в 86%. При одномерном быстром преобразования Фурье (FFT) одного миллиона элементов скорость в 106 Мфлоп/с была установлена для одного процессора, а для четырёх процессоров была получена близкая к ней скорость.

## 7) Серия CRAY T3E

Системы Cray T3E - это масштабируемые параллельные системы, которые используют DECchip 21164 (DEC Alpha EV5) RISK-процессоры с пиковой производительностью 600 Мфлоп и 21164A для машин Cray T3E-900 и Cray T3E-1200. Каждый процессорный элемент (ПЭ) Cray T3E имеет свою собственную DRAM-память объёмом от 64 Мбайт до 2 Гбайт. Модели T3E, T3E-900, T3E-1200, T3E-1350.

Каждый узел в системе содержит один процессорный элемент (ПЭ), включающий процессор, память и средство коммутации, которое осуществляет связь между ПЭ. Система конфигурируется до 2048 процессоров. Пиковая производительность составляет 2,4 Тфлоп.

Системы Cray T3E выполняют операции ввода/вывода через многочисленные порты на один или более каналы GigaRing.

**Оценка производительности** производилась при решении плотной линейной системы уравнений порядка 148800 на машине T3E-1200 с 1200 процессорами. Была достигнута скорость в 1,127 Тфлоп/с, что составляет 63% эффективности.

Машина **T3E Classic** выпускалась с 1996 по 1997 годы. Имела распределённую память и до 2048 процессоров. Теоретическая пиковая производительность достигала 1228 Гфлоп/с.

## Системы Cray T3E с воздушным охлаждением



Параллельные системы Cray T3E с воздушным охлаждением масштабируются от 16 до 128

процессоров с пиковой производительностью от 9,6 до 76,8 Гфлоп. Они имеют до 256 Гбайт общей адресуемой памяти с 2 Гбайт памяти на ПЭ.

Архитектура: MIMD (*Multiple Instruction Multiple Data*) с аппаратной поддержкой обработки отдельных команд нескольких данных — SIMD (*Single Instruction Multiple Data*).

Поддерживается 32-х и 64-х разрядная арифметика в стандарте IEEE.

Модели (наиболее популярные)	Число стоек	Число ПЭ	Основная память *, Гбайт	Пиковая производительность, Гфлоп	Максимальное число каналов ввода/вывода	Пропускная способность ввода/вывода, Гбайт/с
Отдельная стойка						
AC16	1	16	2,0	9,6	2	2
Несколько стоек						
AC32	2	32	4	19,2	4	6
AC64	3	64	8	38,4	8	8
AC80	4	80	10	48	10	10
AC128	6	128	16	76,8	16	16
* Конфигурация основной модели включает 128 Мбайт памяти на ПЭ. Доступны необязательные размеры памяти: 64, 256 или 512 Мбайт, или 1 или 2 Гбайт на ПЭ.						

В программное обеспечение входят:

- Оперативная система UNICOS/mk
- Cray CF09, CRAFT, HPF, Cray C/C++
- Высокооптимизированный пакет передачи сообщений
- PVM, MPI
- Оптимизированные научные библиотеки
- Отладчик
- Автоматическое конвертирование данных (Cray, IBM, CDC)
- Полная совместимость с UNIX (TCP/IP)

## Системы Cray T3E с жидкостным охлаждением



**Cray T3E-900**

процессоров с пиковой производительностью от 38,4 до 1,2 Тфлоп. Они имеют до 4 Тбайт общей адресуемой памяти с 2 Гбайт памяти на ПЭ.

Архитектура: MIMD (*Multiple Instruction Multiple Data*) с аппаратной поддержкой обработки отдельных команд нескольких данных — SIMD (*Single Instruction Multiple Data*).

Поддерживается 32-х и 64-х разрядная арифметика в стандарте IEEE.

Параллельные системы Cray T3E с жидкостным охлаждением масштабируются от 64 до 2048

Модели (наиболее популярные)	Число стоек	Число ПЭ	Основная память *, Гбайт	Пиковая производительность, Гфлоп	Максимальное число каналов ввода/вывода	Пропускная способность ввода/вывода, Гбайт/с **
LC64	1	64	8	38,4	4	4
LC128	1	128	16	76,8	8	8
LC256	1	256	32	150	16	16
LC384	2	384	48	230	24	24
LC512	2	512	64	300	32	32
LC1024	4	1024	128	600	64	64
LC2048	8	2048	256	1200	128	128

\* Конфигурация основной модели включает 128 Мбайт памяти на ПЭ. Доступны необязательные размеры памяти: 64, 256 или 512 Мбайт, или 1 или 2 Гбайт на ПЭ.

\*\* Максимальная пропускная способность ввода/вывода в системах с жидкостным охлаждением может быть удвоена с помощью специального размещения.

В программное обеспечение входят:

- Оперативная система UNICOS/mk
- Cray CF09, CRAFT, HPF, Cray C/C++
- Высокооптимизированный пакет передачи сообщений
- PVM, MPI
- Оптимизированные научные библиотеки
- Отладчик
- Автоматическое конвертирование данных (Cray, IBM, CDC)
- Полная совместимость с UNIX (TCP/IP)

## Система T3E-1200

Серия Cray T3E-1200 в два раза превышает производительность систем Cray T3E при уменьшенной вдвое стоимости за Мфлоп. Конфигурации в воздушно-жидкостном охлаждении имеют от 6-и процессоров, а в жидкостном - от 32 процессоров. Каждый процессор имеет производительность в 1,2 Тфлоп, для всей системы пиковая производительность меняется от 7,2 Гфлоп до 2,5 Тфлоп. Масштабируется до тысяч процессоров. Выпущена в 1997 году. Стартовая цена составляет 630 000 долларов США.

Система предназначена для наиболее важных научных и технических задач в аэрокосмической, автомобильной, финансовой, химико-фармацевтической, нефтяной и т.д. промышленности, также в широких областях прикладных исследований, включая химию, гидродинамику, предсказание погоды и сейсмические процессы.

Для поддержки масштабируемости используется оперативная система UNICOS/mk - масштабируемая версия UNICOSR. Системы T3E-1200 поддерживают как явное распараллеливание распределённой памяти посредством CF90 и C/C++ с передачей сообщений (MPI, MPI-2 и PVM) и передачу данных, так и неявное распараллеливание посредством возможностей HPF и Cray CRAFT.

Системы T3E выполняют операции ввода/вывода через многочисленные порты на один и более каналов посредством интерфейса GigaRing. Каждый канал сдвоенного кольца ввода/вывода, содержащий в двух кольцах данные, которые перемещаются в противоположных направлениях, передают данные ввода/вывода с высокой пропускной способностью и повышенной надёжностью. Все каналы ввода/вывода доступны и управляемы всеми процессорными элементами. На системах T3E каждый интерфейс GigaRing имеет максимальную пропускную способность в 500 Мбайт/с.

В дополнение к высокой производительности и пропускной способности процессорных элементов и высокой масштабируемости, системы Cray T3E-1200 имеют две уникальные особенности: STREAMS и E-Регистры. STREAMS доводят до максимума пропускную способность локальной памяти, позволяя микропроцессору запускать при полной скорости для ссылки для вектороподобных данных. E-Регистры предоставляют операции gather/scatter (соединение/вразброс) для ссылок на локальную и удалённую память, и используют полную пропускную способность внутреннего соединения для удалённого чтения и записи отдельного слова.

Число процессоров	6 - 128 32 - 2048
Тактовая частота процессора, МГц	600
Пиковая производительность, Тфлоп	2,4+
Размер памяти на процессор, Гбайт	0,256 - 2
Топология внутреннего соединения	3D двунаправленный торус (Torus)
Максимальная двоичная пропускная способность, Гбайт/с	122
Максимальное число каналов GIGARING	128
Пиковая пропускная способность ввода/вывода, Гбайт/с	128

## Система T3E-1350

Система Cray T3E-1350 является пятой в серии в линейке Cray T3E крупномасштабируемых массивно параллельных систем.

Система Cray T3E-1350 порождена объединением ультрабыстрых микропроцессоров (675 МГц) и межпроцессорными соединениями посредством масштабируемой оперативной системы и высокопроизводительной системой ввода/вывода. Процессоры Cray T3E-1350 являются сильносвязанными посредством внутреннего соединения: трёхмерный двунаправленный торус, обладающий очень низким временем ожидания и высокой пропускной способностью. Cray T3E-1350 имеет пиковую скорость соединений в 650 Мбайт/с.

Система управляется распределённой оперативной системой Cray UNICOS/mk (масштабируемая версия оперативной системы UNICOS) и масштабируемой системой ввода/вывода GigaRing.

Система Cray T3E-1350 поддерживает как явные, так и неявные методы. Явные методы включают распараллеливание распределённой памяти посредством CF90 и C/C++ с передачей сообщений (SHMEM, MPI, PVM) и моделями ко-массивов (Co-array) программирования на Фортране. Неявное распараллеливание поддерживается посредством HPF и Cray CRAFT.

В дополнение к высокой производительности, пропускной способности процессорных элементов (ПЭ) и масштабируемости система Cray T3E-1350 имеет две уникальные особенности: STREAMS и E-Registers. STREAMS максимально увеличивают пропускную способность локальной памяти. E-Registers обеспечивают операции gather/scatter (слияние/вразброс) для ссылок на локальную и удалённую память и используют полную пропускную способность внутреннего соединения для удалённого чтения и записи отдельного слова.

## Основные особенности

Пиковая производительность	(LC) от 54 Гфлоп до 3 Тфлоп
Архитектура	MIMD с аппаратной поддержкой для SIMD
Микропроцессор	21164A (EV5.6) 4-магистральный суперскалярный RISC 2 операции с плавающей запятой за цикл 32-х и 64-х разрядная арифметика в стандарте IEEE
Локальная память, Мбайт	256 или 512Мбайт
Защита ошибок в данных	SECCDED
Тактовая частота, МГц	675
Пиковая производительность, Мфлоп	1350 на ПЭ
Пиковая скорость выполнения команд	2700 MIPS (миллионы инструкций в секунду) на ПЭ
Пиковая пропускная способность памяти, Мбайт/с	1200 на ПЭ
Конструктивное оформление	8 ПЭ на модуль, жидкостное охлаждение
Число ПЭ на систему	от 40 до 2176 в шагах по 8 ПЭ
Технология памяти	64 Мбайт 50 нс DRAM
Архитектура памяти	Когерентный кэш, физически распределённая, coherent, physically distributed, с общей адресацией
Общая память системы	от 10 Гбайт до 1 Тбайт
Топология сети межсоединений	Трёхмерный двунаправленный свёрнутый торус
Пиковая производительность интерфейса ввода/вывода на модуль, Мбайт/с	500

## Cray СуперКластер



Система Alpha Linux SuperCluster запущена в середине 2001-го года. Она представляет собой крупномасштабный кластер, объединяющий в себе разрешение проблем скорости с увеличением технических возможностей информационного центра, с приемлемой ценой и популярной операционной системой Linux.

Система включает в себя следующие возможности:

- Более высокая производительность в широкой области сложности задач и объёма работ.
- Более высокие возможности информационного центра, включая высокую эффективность (динамическую реконфигурируемость, параллельный ремонт, "горячую" перезагрузку - warm boot и "горячую" замену - hot swap).
- Общий перезапуск с контрольной точки (в ситуации прерывании системы сохраняет результаты работы и без помех продолжает задание при восстановлении).
- Общее администрирование ресурсов.
- Отдельный образ системы даже при очень больших размерах системы (тысячи процессоров).
- Эффективное планирование заданий, установление приоритетов и учёт использования ресурсов.

## MPP-системы Cray T3E/CrayT3D

Cray Research выпускает как суперкомпьютеры векторной архитектуры серий Cray T90, Cray J90, так и MPP-системы Cray T3E. К этому следует добавить SMP-системы Cray CS6400 производства Cray Research Superservers. При существующей на сегодняшний день сложной обстановке на рынке суперкомпьютеров неизбежно встает вопрос о перспективах этих вычислительных систем. Векторные компьютеры испытывают усиливающееся давление со стороны систем, использующих RISC-процессоры. После слияния Cray Research и Silicon Graphics у объединенной компании окажется сразу два SMP-сервера на RISC-процессорах, CS6400 и POWER CHALLENGE. Вполне естественная в будущем ориентация на микропроцессоры MIPS/SGI означает, на взгляд автора, и неопределенность судьбы серверов CS6400, базирующихся на процессорах фирмы Sun. Памятуя о том, что прогнозы - вещь неблагодарная, рискну все же предположить, что линия компьютеров с архитектурой CrayT3E имеет достаточно надежные перспективы.

Одна из самых важных характеристик суперкомпьютеров MPP-архитектуры, имеющих физически распределенную память, - так называемый диаметр системы. Это понятие означает максимальную длину, которая может понадобиться посланному сообщению, при прохождении из одного узла MPP-компьютера в другой. Длина измеряется, грубо говоря, количеством промежуточных узлов, которые надо миновать на пути к конечному узлу. Если в MPP-системах Intel Paragon, являющихся сегодня мировыми лидерами [производительности](#), применяется топология двухмерной решетки, то в компьютерах Cray T3D, а затем и Cray T3E, использована более сложная топология трехмерного тора. Это позволяет уменьшить диаметр системы.

Передача [данных по](#) тору может происходить в обоих направлениях. Скорость обмена полезными данными (за вычетом управляющей информации) равна 160 Мбайт/с. Для оценки пропускной способности всей системы может использоваться "половинная" пропускная способность. Это максимальная скорость, с которой одна половина (узлов) компьютера может обмениваться с другой половиной. В Cray T3D она составляет 78,6 Гбайт/с для 2048 процессорных элементов, а в Cray T3E - свыше 122 Гбит/с для 512 процессорных элементов при скорости передачи полезных данных от узла к узлу 480 Мбайт/с (в 3 раза выше, чем в Cray T3E). Для сравнения, максимальная скорость передачи данных между двумя узлами системы в Intel Paragon равна 175 Мбайт/с.

Сеть, соединяющая узлы (interconnect network, в дальнейшем IN), включает в себя также и маршрутизаторы. В отличие от Cray T3D, в Cray T3E появилась адаптивная маршрутизация, позволяющая перемаршрутизировать сообщения с тем, чтобы обходить интенсивно используемые "узкие места". С точки зрения производительности, IN Cray T3E существенно превосходит как Intel Paragon, так и своего предшественника Cray T3D.

На рис. 1 представлена общая архитектурная схема компьютеров, содержащих Cray T3D. Дело в том, что Cray T3D представляет собой не отдельный компьютер, а может работать только в системе, содержащей векторно-конвейерную машину от Cray Research, например Cray C90. Это серьезный недостаток, поэтому Cray T3E был разработан уже как полностью автономная система.

Процессорные элементы в Cray T3D основаны на микропроцессорах DEC Alpha 21064 с пиковой производительностью 150 MFLOPS. В состав процессорного элемента в Cray T3D



входит также локальная память размером 16 или 64 Мбайт. Пара процессорных элементов образует узел системы, в котором они совместно используют аппаратуру подключения к IN, включая маршрутизатор. Поскольку в Cray T3D может быть включено до 2048 элементов, диапазоны возможных конфигураций следующие: от 32 до 2048 процессоров, пиковая производительность от 4,8 до 307,2 GFLOPS, оперативная память от 0,5 до 128 Гбайт. Очень важно то, что, хотя оперативная память физически распределена и каждый процессорный элемент имеет свою локальную память, вся память системы является глобально адресуемой и доступна из любого процессорного элемента, для любого микропроцессора.

Системы CrayT3E являются уже полностью автономными компьютерами, имеющими собственные средства ввода/вывода вместо каналов IOC векторных систем. В процессорных элементах Cray T3E применяются микропроцессоры DEC Alpha 21164 с тактовой частотой 300 МГц и пиковой производительностью 600 MFLOPS.

Локальная оперативная память в Cray T3E может иметь емкость от 64 Мбайт до 2 Гбайт. Пропускная способность локальной памяти Cray T3E составляет 1,2 Гбайт/с. Каждый процессорный элемент Cray T3E снабжен собственным маршрутизатором и является узлом IN.

Конструктивно в Cray T3E процессорные элементы вместе с их маршрутизаторами объединяются в модули (по четыре процессорных элемента на модуль в системах с воздушным охлаждением и по восемь - с жидкостным охлаждением). Модуль является минимальной единицей наращивания конфигурации.

В системах Cray T3E появилась также новая архитектура ввода/вывода GigaRing. Обмен данными с устройствами ввода/вывода в Cray T3E, в отличие от Cray T3D, осуществляется не через каналы IOC векторных компьютеров, а через собственные каналы Cray T3E, называемые GigaRing. Каждый канал GigaRing состоит из двух 32-разрядных колец, по которым данные передаются в противоположных направлениях. Скорость передачи по такому каналу, как видно из названия, составляет 1 Гбайт/с.

Канал передает данные между узлами GigaRing (не следует путать с их узлами IN!). Выделяют три класса узлов GigaRing: сетевые (подключенные посредством HIPPI/FDDI/ATM/Ethernet), подсистемы хранения данных (дисковые и ленточные) и системные.

В настоящее время в качестве системных может выступать процессорная часть Cray T3E вместе с IN. В планы компании Cray Research также входит обеспечение возможности работы векторных компьютеров Cray T90 и J90 в качестве системных узлов GigaRing.

Существует несколько типов узлов ввода/вывода. Узлы Fibre Channel поддерживают протокол Fibre Channel Arbitration Loop, обеспечивающий канал передачи данных с пиковой скоростью 100 Мбайт/с. Один узел обслуживает пять таких каналов, к каждому из которых можно подсоединить до 80 устройств.

Узлы ESCON и узлы блок-мультиплексных каналов позволяют подключать к Cray T3E ленточные накопители, работающие через каналы мэйнфреймов IBM соответствующих типов. Узел HIPPI обеспечивает поддержку этого сетевого стандарта на скоростях 100 и 200 Мбайт/с. Наконец, многоцелевой узел включает 2 шины SBus (традиционно используе-

мые в компьютерах Sun Microsystems), к каждой из которых подключается до 4 контроллеров. Эти контроллеры осуществляют поддержку сетевых технологий ATM, Ethernet (10BaseT), FDDI (одиночное и двойное подключение), а также интерфейса Fast&Wide SCSI-2 в дифференциальном варианте.

Системы Cray T3E с воздушным охлаждением занимают от 1 до 6 шкафов. Нарастивание конфигурации происходит модулями, включающими 4 процессорных элемента и 1 интерфейс GigaRing. Компьютеры с жидкостным охлаждением имеют до 8 шкафов. Их конфигурации нарастаются модулями, имеющими 8 процессорных элементов и 1 интерфейс GigaRing. Характеристики различных конфигураций Cray T3E приведены в таблице 1, где AC означает воздушное охлаждение, а LC - жидкостное.

Модели	Число ПЭ	Память	Пиковая производительность
<b>AC16</b>	6-128	1-256 Гбайт	9,6-76,8 GFLOPS
<b>LC64-LC2048</b>	64-2048	4 Гбайт-4 Тбайт	38,4 GFLOPS-1,2 TFLOPS

**Таблица**  
Конфигурации Cray T3E.

**1.**

Масштабируемость производительности, IN, системы ввода/вывода операционной системы - все это позволило руководителю компании Дж. Самперу сделать заявление, что Cray T3E - первый в мире реально масштабируемый суперкомпьютер. Суперкомпьютеры такого класса - это штучный товар. Крупнейшие суперкомпьютерные центры США оснащены компьютерами Cray T3D. Компания получила уже несколько заказов на системы Cray T3E, ориентированные на научно-технические применения. А один из спонсируемых NSF суперкомпьютерных центров в Питтсбурге ее уже инсталлировал. Стартовая цена меньше 1 млн. долл.