Машини с неунифициран достъп до паметта (NUMA). TERA T3D.

Проектирани са с цел отстраняване “тясното място”, което се полуава при достъп до паметта в UMA машините. Логически разделената (общата) памет е физически разпределена между процесорните възли в NUMA – това води до архитектури с разпределена обща памет. По този начин пааралелните компютри стават много гъвкави и мощни, но от друга страна възниква проблем при зареждането на данните в локалните памети. Достъпът до локалната памет в процесорния възел е по-бърз, отколкото достъпа до отдалечена локална памет. Структурата и проектирането на NUMA машините много приличат на тези в мултикомпютрите с разпределена памет. Основната разлика е в организацията на адресното пространство. При мултипроцесорите глобалното адресно пространство е неунифицирано видимо за всички процесори, т.е. всичките процесори могат явно да се обръщат към всички памети. За ралика от тях, при мултикомпютрите адресното пространство е размножено в ЛП-ти на процесорните елементи (ПЕ) и не е разрешен достъп до ЛП на друг ПЕ. Тази разлика също рефлектира и на SW ниво: мултикомпютрите с разпределена памет се прогламират въз основа на парадигмата*message-passing*, докато при NUMA машините – се използва принципа за адресно пространство (обща памет).

През последните години все по-трудно се прави разграничение между тези машини. Разликата става все по-малка и от факта, че използваната форма за достъп до ЛП-ти е еднаква и за двата класа MIMD компютри. Отдалечения достъп в NUMA машините е реализиран посредством *съобщения*, подобно на *message-passing* мултикомпютрите.

При NUMA машините, както и при мултикомпютрите, основното при проекти-рането е организацията на процесорните възли, мрежата и техниките за редуциране достъпа до отдалечени памети.

Представители на NUMA машини са BBN TC2000, IBM RP3, Cray T3D и мултипроцесора Hector.

CrayT3D е съвременна NUMA машина, проектирана с цел създаване на високо скаларен паралелен СК, обединяващ идеите на общата памет и на *message-passing* програмирането. При NUMA машините общата памет е разпределена между РЕ, за да се

избегне „тясното място”, възникващо при достъп до паметта и нямат HW поддръжка за кохерентност на кеша. Но се използва специален
SW пакет
и
модел на програмиране
, наречен
CRAFT
, управляващ кохерентността и гарантира цялостта на данните. HW структура на CrayT3D е разделена на две части:

- Микроархитектура,
- Макроархитектура.

Микроархитектурата е основана от Digital’s 21064 Alpha AXP микропроцесора, който като другите съвременни микропроцесора, има три основни недостатъка:

- Ограничени адресно пространство,
- Малка или latency-hiding възможност,
- Няколко или никакви синхронни основи.

Cray има над 128GB разпределена памет, която изисква поне 37 бита от физическия адрес. За да се увеличи броя на адресните битове след 34, които са осигурени от Alpha чип, то CrayT3D добавя 32 входно регистрово множество.

За да подобри механизма на Alpa чип за *latency hiding*, Cray въвежда 16-word FIFO, наречена *prefetch queue*, която позволява 16 извлечени инструкции да бъдат представени.

Хардуерно Cray T3D поддържа механизми за синхронизация. **Barrier** хардуер сравнява 16 паралелни логически AND дървета, които позволяват на много
barrier
да бъдат конвейрни. При достигане границата (barrier) на процесора, то тя трябва

асоциативни да установи своя бит в 1. И когато всички процесори достигнат своята граница

,

AND функцията се изпълнява и нулира хардуерно битовете на barrier на всеки процесор, след това сигнализира за продължение.

Cray T3D осигурява специално множество от регистри за реализиране на **fetch-and-incr ement**

хардуера. Съдържанието на тези регистри автоматично се увеличава, когато бъде прочетено. В паметта на PE-ти се поддържа

messaging

. Atomic swap регистрите се използват за размяна на данни между регистър и отдалечена клетка от паметта като неделима (безкрайно малка) операция. Латентността на atomic swap може да бъде скрита чрез използването на prefetch техника.

Макроархитектурата дефинира как да се свържат и интегрират възлите на паралелния компютър, докато микроархитектурата специфицира организацията в самия възел. Двете части на макроархитектурата са системата памет и мрежата за свързване. Системата памет използва разпределена обща памет, където PE-ти директно могат да адресират клетка от паметта на друг PE. Физическият адрес има два компонента: *номер* на PE и *отместване* в PE. Всеки PE съдържа 16 или 64MB локална DRAM памет. Достъпът до локалната памет е между 13 и 38 такта (87 до 253nsec). Латентността за достъп до отдалечена памет варира между 1 и 2 микросекунди. Кеша за данни е резидентен на Alpha AXP микропроцесора.

Топологията на е 3-d торус, който е бил избран въз основа на направените измервания върху глобалната латентност и глобалната дължина на реалните SW пакети. Мрежата премества данните в един или 4 64-разрядни пакета. Всяко преместване през мрежата изисква 1 такт (6.67nsec при 150Mhz). Всеки възел включва 2 PE. PE-ти са независими, имат отделна памет и даннови пътеки. Те споделт помежду си само мрежата. Сърцето на Cray T3D е 4x4x4 3-D торус, който е свързан с Cray Y-MP host (традицианна паралелна векторна система) и с I/O контролери, чрез специални I/O *gateways*. Тези *gate ways*

са поставени през 3-D торус и имат версии за ниска и висока скорост (LOSP, HISP:400MB/s full-duplex). Една конфигурация може да има 8x8x8=512 възела, всеки съдържащ по 2 PE-та и това прави 1024 PE-та общо.

Фиг.7 Структура на Cray T3D.