Kap. 8: Vektor-maskiner og SIMD maskiner

- **Vektor-prosessering**
- **Lagersystemer for hurtig vektor-aksess**
  - C-aksess / S-aksess / C/S-aksess
- **Superdatamaskiner**
  - Eksempler
  - Ytelses-mål
- **(Sammensatt) vektor prosessering**
  - Sammensatte vektor-operasjoner
  - "Vektor-looping" og kjeding
- **"Multipipeline network"**
- **SIMD maskiner**
  - Arkitektur
    - » Realisering av lager
  - Eksempler
    - » CM-2
    - » MasPar MP-1
    - » CM-5

Superdatamaskiner

- **De raskeste "kommersielle" maskiner**
- **Fram t.o.m. i dag: multiprosessorer der hver prosessor er spesialkonstruert for hurtig vektor-prosessering**
  - = (*multi*)*vektor-maskiner*
  - flere funksjonelle enheter
  - med stor grad av pipelining
  - vektor-registre m.m.
- **"Behovet ligger alltid foran ytelse"**
- **Vektor-prosessering kan også utføres på SIMD-maskiner**
- **I nyere tid har vi fått noen MIMD-superdatamskiner**
  - Thinking Machines Corp. (TMC)
- **Hybride SIMD/MIMD-maskiner**
  - CM.5 fra TMC

Vektor-prosessering

- **Vektor**
  - samling skalare verdier
  - *"stride"*
    - » adresse-inkrement mellom etterfølgende elementer i en vektor
- **Vektor prosessor**
- **Vektorisering**
  - transformasjon fra skalare beregninger til vektorberegninger
  - vektoriserende kompilator
- **Vektor prosessering**
  - reduserer programvare overhead (løkker)
  - reduserer lageraksess konflikter
    - » strukturering av aksesser
    - » pipelining
  - skreddersydd for utførelse v.h.a. trinnvis prosessering (pipelining)
  - økte kostnader til maskinvare og kompilator

Vektor-instruksjoner

(Repeter side 184 [Hwa93])

- **To hovedtyper maskiner**
  - **register-register** maskiner (CRAY)
  - lager-lager maskiner
- **Fig. 8.1**
  - (register-register)
  - vektor-vektor instruksjon
  - vektor-skalar instruksjon
  - vektor-lager instruksjon
  - vektor-reduksjons instruksjon
    - » Eks.: max, min, sum, middel etc.
  - *Samle (Eng. gather) og spre (Eng. scatter)* instruksjoner
    - » **Fig. 8.2**
  - Maskerings-instruksjoner
    - » Fig 8.2. c)
- **"Gather/scatter" og "masking" instruksjoner er sentrale for effektiv prosessering av tynne (Eng. sparse) vektorer (og matriser)**

Lagersystemer for hurtig vektor-aksess

- **Transport av vektor-operander mellom lager og registre er vanligvis pipelined ("acces pipe") og parallel (dvs. flere aksess-veier)**
- **Vektor-operander**
  - base-adresse, "stride", lengde
- **Aksess-metoder**
  - **C-aksess** (Concurrent access)
    - » lavordens-fletting
    - » overlapping av *m* etterhverandreføIgende aksesser
    - » repeter Fig. 5.15 a) og 5.16 b)
    - » "stride"-verdi vil påvirke ytelse
  - **S-aksess** (Simultaneous access)
    - » lavordens fletting
    - » alle *m* lagermoduler aksesseres simultant
    - » Fig. 8.3
  - **C/S-aksess (fig. 8.4)**
    - » Kombinasjon. F.eks. *n* aksess-busser som hver har *m* lagermoduler med C-aksess. De *n* aksess-bussene brukes parallelt og simultant med S-aksess
    - » *m x n* ord kan hentes i hver "lager-modul-syklus" (= "major cycle")

Superdatamaskiner (vektor)

- **CRAY**
- **Japanske**
  - Fujitsu
  - NEC
  - Hitachi
- **Andre**
  - Control Data Corp. (CDC)
    - » STAR-100 (1973), Cyber 205 (1982)
    - » ETA-10
      - lager til lager arkitektur
      - tatt ut av produksjon
  - ...
- **Andre (ikke vektor)**
  - Intel Paragon m.fl.
  - Thinking Machines CM5
  - IBM SP
- **(Tabell 8.1)**

"Supercomputer vendors", from http://www.umiacs.umd.edu/~dbader/sites.html

Alta Technology
Amdahl Corporation
Chen Systems
Convex Computer Corporation
Cray Research, Inc.
Digital Equipment Corporation (DEC)
Fujitsu
Geli Engineering: Pentium Workstation Clusters
Hewlett-Packard
Hitachi Central Research Lab, Japan
IBM Corp.
Integrated Computing Engines, Inc. [ICE]
Intel Supercomputing Systems
Meiko Scientific
nCUBE
NEC Supercomputer Center, Japan
NTT Transmission Systems Labs, Japan
Parsytec Computer GmbH
Sequent Computer Corporation
Siemens Nixdorf and Pyramid Technology
Silicon Graphics, Inc.
Tandem
Tera Computer Co.
Thinking Machines Corporation

Superdatamaskiner fra CRAY

- **Cray Research Inc. (CRI)**
  - Stiftet av Seymour Cray i 1972
  - Ca. 400 maskiner solgt, (lite!?)
- **Cray 1 (1975)**
  - superdatamaskin for vektorprosessering
  - stor grad av pipelining
  - register til register arkitektur
- **Cray 1S (1976)**
  - ECL logikk
  - 12.5 ns klokke syklus
  - 10 stk. funksjonelle pipeline
  - en-bruker, batch-orientert
- **Cray X-MP (1983)**
  - Steve Chen
  - 8.5 ns klokke
  - multiprosessering
    - » 1-4 CRAY-1 prosessorer med felles lager
    - » delte register-"klynger" for inter-prosessor kommunikasjon/synkronisering
  - SSD ("Solid State Disk"), dvs. halvleder-disk
    - » utvidet delt lager
    - » mellom primær-lager og disk

Superdatamaskiner fra CRAY, forts.

- **CRAY Y-MP (1988)**
  - 6 ns klokke
  - opp til 8 prosessorer
  - **Fig. 8.9, Fig. 8.10** (og kap. 8.2.2.)
- **CRAY C-90 (1990)**
  - 4.2 ns klokke
  - 16 prosessorer
  - UNIX-basert OS (UNICOS)
  - Flere C90 maskiner kan koples i en klynge rundt en SSD (Solid State Disk), **Fig. 8.11**
- **CRAY-maskiner ved NTNU**
  - CRAY Y-MP3D/464 ("fases ut" etter J90)
    - » 4 prosessorer // 1.33 Gflops toppytelse
    - » 512 Mbyte primærlager //1 Gbyte SSD // 55 Gbyte masselager
  - **CRAY J90** (ankom des 95)
    - » 8 prosessorer
    - » se senere foil
  - **CRAY T3E** (kommer høst 96)
    - » 64 prosessorer
    - » se senere foil

Nye superdatamaskiner ved NTNU
(Utklipp/sammendrag fra news m.m.)

Norges forskningsråd (NFR) og Universitetet i Trondheim (UNIT) [nå **NTNU**] har inngått en avtale om tungregning i perioden **1996-98**. NTNUovertar prosjektansvaret for tungregneprosjektene som SINTEF har drevet for forskningsrådet i tilsammen 9 år. Avtalen gjelder fra 26 nov 95.

Parallelt med avtalen med forskningsrådet har UNIT inngått en avtale med **Cray Research** om levering av nytt utstyr. Den nye tekniske løsningen omfatter en **kombinasjon av vektor- og parallellteknologi**. På denne måten tas det sikte på å oppnå både kontinuitet og vekstmulighet.

Installasjonen vil bestå av en **8-prosessors CRAY J90** vektormaskin som installeres i desember, og en **64-prosessors CRAY T3E MPP-maskin** som kommer i 3. kvartal 1996. En moderne backupløsning med taperobot inngår også fra starten av.

For å sikre stabil drift under **byggeperioden** av det nye **Realfagbygget** ved UNIT, som blir liggende vegg i vegg med dagens maskinrom, vil nytt utstyr bli installert i universitetets lokaler på **Dragvoll**. **Intel Paragon-maskinen** i Trondheim vil også drives av UNIT, og vil bli flyttet til samme lokaliteter. CRAY Y-MP vil bli holdt i drift til J90 har passert akseptanseperioden for å sikre overgangen.

J90 er en mindre og billigere vektormaskin enn Y-MP, men med **samme hovedarkitektur**. Den kjører under samme operativsystem. Øvrig basis programvare er også den samme. Det blir derfor ingen vesentlige endringer i brukergrensesnittet. Installert applikasjonsprogramvare vil i hovedsak være som før. Det anbefales å kompilere om egne programmer. Den har en prosessorytelse på 200 Mflops, mot Y-MPs 333 Mflops. Den aktuelle konfigurasjonen har dobbelt så mange prosessorer som dagens Y-MP, slik at total teoretisk ytelse øker fra 1.3 til **1.6 Gflops**. Minnet vil være på 1024 Mbyte. Dette er det dobbelte av det dagens Y-MP har, slik at minnekapasiteten i gjennomsnitt per prosessor blir som i dag. I/O-ytelsen vil være lavere enn på Y-MP.

CRAY T3E er en MPP-maskin med distribuert minne. Anlegget vil få en konfigurasjon med 64 prosessorer, hver med 128 Mbyte minne. Teoretisk ytelse per prosessor er på 600 Mflops, som gir en total teoretisk peak-ytelse på 38.4 Glops. Den nye løsningen vil gi betydelig øket regnekraft for den som kan utnytte MPP-teknologien..

Kilde; bl.a. http://hpc.ntnu.no/

Superdatamaskiner fra CRAY, forts.

- **CRAY MPP (T3D) (Gammel foil basert på læreboka)**
  - ("CRAY må følge med motebølgen")
  - Massiv parallellitet med bruk av kraftige standard RISC mikroprosessorer (Alpha)
    - » (Fig. 8.12)
    - » prosessor omgis av spesial-HW for kommunikasjon
  - T3D, 3dimensjonal Torus sammenkoplingsnettverk
  - koples til en Y-MP C90
  - MIMD
    - » kan emulere SIMD / kan partisjoneres
  - globalt adresserbart, fysisk distribuert lager
    - » spesial-HW for
      - aksess av ikke-lokalt lager (Eng. remote)
      - synkronisering
  - Mach-basert mikro-kjerne (OS) i hver node
  - feiltoleranse via redundante prosessorer
    - » rekonfigurering i programvare
  - Vesentlig innsats på programvaresiden
    - » flaskehals i dag for denne type maskiner
    - » Mål (for mange): portabel effektiv programvare
    - » standarder
      - språk, bibliotek

CRAY J90-serien

- **Ankom i desember 1995**
- **J90 serien**
  - **"ideal for budget-conscious departments and organizations" !**
  - **god (lav) pris/ytelse**
- **Binær-kompatibel med Y-MP, dvs. tradisjonell vektormaskin**
- **0.8 - 6.4 Gflops topphastighet**
- **4- 32 prosessorer (CRAY typen)**
  - **CMOS-teknologi**
    - **» 10 ns klokke**
    - **» 200 Mflops pr. CPU**
    - **» billigere drift, bl.a. luftavkjølt, vanlig strøm**
- **1- 8 Gbyte primærlager**
- **Bare 1.1 m² "fotavtrykk", 635 Kg**
- **UNICOS, FORTRAN 90, C, C++, TCP/IP, X Window m.m.m.**
- **Ved NTNU:**
  - **CRAY J916 ("bure")**
    - **» 8 prosessorer, 1 Gbyte primærlager**

CRAY T3E serien (1/2)

- **T3E serien**
  - **Bygger på T3D-serien som ble introdusert i 1993**
  - **T3E AC-serien (Air-cooled)**
    - **» 16 -128 prosessorer**
  - **T3E LC-serien (Liquid-cooled)**
    - **» 64-2048 prosessorer**
  - **1024 prosessors T3D har mange rekorder**
- **DECchip 21164 (DEC Alpha EV5)**
  - **64 bit superskalar RISC (300 Mhz)**
    - **» 4-veis instruksjonsoppstart**
    - **» 2 flop pr. klokke**
  - **600 Mflop topphastighet**
  - **raskere versjoner vil kunne brukes**
- **3D torus**
  - **toveis linjer**
  - **6 ganger raskere enn i T3D**
  - **adaptiv ruting, kan rute rundt hot-spots**
- **9.6 Gflops - 1.2 Tflops**
  - **målsettning om 1Tflops opprettholdt ytelse ("sustained") før år 2000**
- **1Gbyte - 1.2 Tbyte primærlager**
- **1.0 - 20.8 m² "fotavtrykk"**

CRAY T3E serien (2/3)

- **MIMD parallell maskin**
  - **massivt parallel**
  - **spesial HW for hurtig synkronisering og SIMD-modus**
    - **» flere synkroniseringsmekanismer**
    - **» hurtig barriere-synkronisering**
      - **godt egnet for BSP**
- **UNICOS/mk**
  - **microkernel-based**
  - **bygger på POSIX**
  - **skalerbart OS**
    - **» funksjoner distribuert, ikke "replisert" (kopiert)**
      - **litt på alle PE'er**
    - **» noen dedikerte PE'er for kjøring av UNICOS**

CRAY T3E serien (3/3)

- **Programmeringsomgivelser**
  - **FORTRAN 90, C, C++, PVM, MPI, CRAY Shared Memory, HPF, CRAFT, analysis tools, debugger**
- **Feiltoleranse**
  - **i snitt 1 redundant PE for hver 128 bruker-**
    - **» utbytting ved "hot-swap", dvs. uten rebooting**
- **Ved NTNU: CRAY T3E AC64("huge")**
  - **kommer høsten 1996**
  - **64 prosessorer, 38.4 Gflops toppytelse, 8 Gbytes primærlager**

Superdatamaskiner ved NTNU oppsummering

- **CRAY Y-MP (cray.sintef.no)**
  - **1.3 Gflops toppytelse, tradisjonell vektormaskin, 4 (CRAY) prosessorer, 512 Mbyte primærlager, skiftes ut vinteren 95/96**
- **CRAY J90 (bure.ntnu.no)**
  - **1.6 Gflops toppytelse, tradisjonell vektormaskin, 8 (CRAY) prosessorer, 1024 Mbyte primærlager**
- **Intel Paragon (elle.ntnu.no)**
  - **2.8 Gflops toppytelse, MIMD parallel maskin, 56 (i860) prosessorer, 2 Gbyte primærlager**
- **CRAY T3E (huge.ntnu.no) (kommer)**
  - **38.4 Gflops toppytelse, MIMD parallel maskin, 64 (DEC Alpha) prosessorer, 8 Gbytes primærlager**
- **Filtjener (grane.ntnu.no)**
  - **Silicon Graphics 4D/320S**
- **Brukerstøtte;**
  - **Unni Hansen, tlf. 7359 0560, Epost: hpc@itea.ntnu.no**

Japanske superdatamaskiner

- **Har ofte hatt en bedre vektor-ytelse enn CRAY-maskiner**
  - ofte høyere klokkerate
  - bedre kompilatorer (?)
  - Fig. 8.7
- **NEC SX-X**
  - Den raskeste (opp til 1992) ?
    - » 22 Gflops
  - 2.9 ns klokke
  - Fig. 8.5
- **Fujitsu VP2000**
  - 1 el. 2 prosessorer
  - **Fig. 8.13**
    - » 3.2 ns klokke
  - Rekonfigurerbare vektor-registre
    - » For å unngå overhead ved såkalt "vektor-looping"
    - » Fig 8.14 (b)

Japanske superdatamaskiner, forts.

- **Fujitsu VPP 500**
  - Vector Parallel Processor
  - lansert sept. 1993
  - 7 - 222 prosessorer, MIMD
    - » hver på 1.6 Gflops
    - » GaAs
  - prosessorer og 1-2 kontroll-prosessorer er sammenkoplet via en stor crossbar-svitsj, Fig. 2.26b
  - globalt delt, fysisk distribuert lager ("virtual shared memory")
    - » "Data Transfer Unit" i hver PE
      - oversetter logiske adresser til fysiske adresser
      - spesial-HW for hurtig barriere synkronisering
  - kan brukes sammen med VP2000
  - UNIX-basert OS
  - **Fig. 8.15**

Ytelses-mål

r vektor / skalar hastighetsforhold

f " vektoriseringsandel "

P Hastighetsøkning ved vektorprossesering

$$P = \frac{1}{(1-f) + f/r} = \frac{r}{(1-f)r + f} \qquad (8.11)$$

- ***r* bestemmes av maskin-konstruksjon**
- ***f* påvirkes av anvendelses-programmer og kompilator**
  - Markedsområde bør påvirke konstruksjon
    - » Fig. 8.6
    - » IBMs generelle stormaskiner: $3 < r < 5$
    - » Superdatamaskiner: $10 < r < 25$
- **Balansert ytelse**
  - *vektor balanse punkt*
    - » Andel av vektorisert kode i et program som er nødv. for å oppnå lik utnyttelse av vektor og skalar enhet
    - » Typisk 0.9 -- 0.98 (se tabell 8.2)

Arkitektur og konstruksjon for god ytelse, sentrale mål

1) balansert vektor/skalar-ytelse
- i samsvar med "typisk anvendelse"

2) skalerbarhet
- arkitektur "tåler" økende # prosessorer
  - » opprettholde god aksess til delt lager
  - » opprettholde hurtig IPC

3) god lagerkapasitet og ytelse
- kort forsinkelse for skalare operander
- høy båndbredde for vektor-operander (og evt. parallell proseessering)
- stor kapasitet
- lager-hierarki
  - » vanlig med to-nivå hovedlager
  - » f.eks. hurtig statisk RAM og langsom dynamisk RAM

4) høy-ytelse I/O
- økt prosesseringskraft har medført at vi løser større problemer
  - » økte IO-krav
  - » (se evt. fig. 8.7 c)

- **Trend mot skalerbare systemer med massiv parallellitet (MPP)**

Stormaskiner og minisuperdatamaskiner

- **Kap. 8.2.4. (bare delvis pensum)**
  - Eksemplene VAX 9000 og Stardent 3000 utgår
- **mange ulike navn brukes på maskiner "like under" superdatamaskiner**
  - umoden og rotete terminologi
  - Eks. "high-end mainframe, minisupercomputer, near-supercomputer,supercomputing workstation"
    - » en minisuperdatamaskin kan typisk ha 5-15% av ytelsen og 3-10% av kostnaden til en superdatamaskin
    - » Eks. Convex, Alliant, Encore, Sequent
    - » åpen arkitektur (kjent standard), standard mikroprosessorer, UNIX
  - ofte laget for mer generelle anvendelser, administrativ databehandling og transaksjons-prosessering
  - typisk arkitektur
    - » 1-6 prosessorer
    - » to-nivå hovedlager
    - » opsjonell vektor-HW
    - » kraftig IO-system

LINPACK

- **Kjent benchmark**
  - Jack Dongarra
- **Løsning av lineære likninger**
- **Fortran**
- ***Indikerer* relativ ytelse for reelle problemer**
- **Tabell 8.5**
  - Kolonne-2; standard-kompilering, $n = 100$
  - Kolonne-3; "alle triks tillat", $n = 1000$
    - » kan komme ganske nære topphastighet
- **I praksis kan en oppnå i snitt ca. 10% av topphastighet**

(Sammensatt) vektor prosessering

- **Sammensatte vektor-operasjoner**
  - en sammensatt beregning på vektorer
  - Eks. 8.7, SAXPY
  - analyse kan avklare muligheter for hurtig parallell utførelse, f.eks. ved vektor-kjeding
- **"Vektor-looping"**
  - Når en vektor er lengre enn vektor-registrene må den hakkes opp og behandles i segmenter
  - Dette kalles "Vektor-looping" el. "Strip-mining"
  - Gjøres av maskinvare under programvare-kontroll og bør være usynlig for programmerer
- **Vektor-kjeding (Vector chaining)**
  - standard teknikk i dag
  - videreutvikling av intern foroversending (Eng. Internal forwarding)
  - kjeden må kunne koples opp uten å bryte ressurs eller data-avhengigheter
  - vektor-registre brukes som grensesnitt-registre mellom funksjonelle enheter
  - Eks. 8.8, **Fig. 8.18** og **Fig. 8.19**
  - kjedings-grad
    - » # forskjellige pipeline enheter som kan lenkes sammen
    - » begrenses av # funk. enheter og # vektor.regs.
    - » Tabell 8.6 (side 437)

Rekursiv vektorprosessering

- **Når resultatet fra en funksjonell pipeline sendes tilbake til en av dennes inngangsregistre**
  - benytter en komponent-teller knyttet til hvert vektor-register
    - » holder rede på hvor langt en har kommet
  - i hver pipeline-sykel benyttes dette registret som et skift-register på komponentnivå
- **Egnet for effektiv vektor-reduksjon**
  - f.eks. summasjon

"Multipipeline network"

- **Generalisering av vektor-kjeding**
- **Nettverk av flere samlebånd ("Pipenet")**
- **Inspirert av Systoliske array**
  - H.T. Kung (1978)
  - nettverk av "funksjonelle enheter"
    - » ofte enkle, dedikerte PE'er
    - » ofte 2D
    - » opererer synkront
    - » funksjon direkte kodet i VLSI
  - lokal kommunikasjon
    - » fast oppkopling
  - avbilde algoritme i HW
  - Anbefalt referanse: H.T. Kung, "*Why Systolic Architectures?",* IEEE Computer, Jan.1982, side 37-46.
- **"Pipenet"**
  - Systoliske array der en tillater dynamisk oppkopling mellom de ulike PE'er
  - "programmerbar systolisk array"
  - Alternativ: funksjon i PE kan endres (iWARP)

"Pipenet"

- **Hwang og Xu**
- **Funksjonelle enheter (FE) koplet sammen via buffrede crossbar-svitsjer (BCN)**
  - Fig. 8.20
  - BCN1 kopler dynamisk registre til FE'er
  - BCN2 setter opp dynamisk koplinger mellom FE'er
- **To nivåer pipelining**
  - av FE'er gjennom BCN
  - innen en funksjonell enhet
- **Flere ferdige oppkoplingsmønstre kan være forhåndslagret for kjente sammensatte funksjoner**
- **Transformering av programgrafer**
  - Fig. 8.20 er et enkelt eksempel
    - » ofte må programmerer eller kompilator gjøre en grundig analyse og evt. transformasjoner
  - midt på side 444 og ut 8.3.3. utgår.

SIMD maskiner

- **En samling (enkle) prosessorer (PE) styrt av en kontrollenhet**
  - Vertsmaskin ("host)
  - Synkron oppførsel av prosessorer ("lockstep")
    - » prosessor-array
  - Realisering av lager, Fig. 8.22
    - » distribuert
    - » felles lager
  - IPC-nettverk
    - » gjerne ulikt for ulike SIMD-maskiner
    - » .... spesial-maskiner for spesielle anvendelser
    - » 2D rutenett mest vanlig
  - Kjente eksempler
    - » Illiac IV
    - » Goodyear MPP
    - » ICL DAP (senere AMT DAP)
- **CM-2**
- **MasPar MP-1**
- **CM-5**

Mer om SIMD maskiner

- **SIMD instruksjoner**
  - vektor-operasjoner el. data-parallelle operasjoner
    - » aritmetikk & logikk
    - » ruting
      - kringkasting, shift, permutering m.m.
    - » maskering (enable, disable, m.m.)
- **Kan også utføre vektor-operasjoner**
  - ved "bredde-parallellitet" (Eng. spatial parallelism) istf. pipelining (Eng. temporal parallelism)
- **Vertsmaskin og I/O**
  - vanlig seriell IO utføres ofte av vertsmaskin
    - » f.eks. Ethernet
  - en del nyere maskiner har my spesial-HW for hurtig parallel IO til PE'ene
- **Beslektet med SIMD-maskiner:**
  - "assosiative maskiner"
    - » basert på assosiativt lager, også kalt innholds-adresserbart lager (CAM)
  - intelligent lager
    - » lager-kretser med logikk for enkle beregninger
  - ikke dekket av pensum
    - » dr.ing. fag Datamaskinarkitektur-2

CM-2 arkitektur

- **Connection Machine, CM-2**
  - Thinking Machines Corp. (TMC)
- **Massivt parallell maskin**
  - fra 4k til 64k 1-bits prosessorer
- **Hovedarkitektur**
  - koplet til vertsmaskin (front-end)
  - en kontrollenhet (sequencer)
  - distribuert lager
  - flere IPC-nettverk
  - Fig. 8.23
- **Kommunikasjon mellom prosessorer (IPC)**
  - Hyperkube & **NEWS** (North-East-West-South)
    - » hyperkube (og dermed 4x4 news) for 16 PE'er inne på en prosessor-krets
    - » egen **ruter** i hver prosessor-krets som kopler disse sammen i hyperkube
  - scanner (Sum-OR-net)
    - » spredning og reduksjon
  - kringkasting via instruksjonsbus

  (+parallel tilkopling til masselager
    - » Data vault)

Prosessor-node i CM-2

- **Prosessor-node**
  - Fig. 8.24
  - 16 + 16 bit-slice prosessorer (1-bits)
    - » heltalls operasjoner utføres bit-serielt
    - » PARIS - instruksjons-sett
  - opsjonell flyttallsenhet
    - » 32-bits buss mot denne
  - felles lager-modul
    - » kan brukes som 32 lokale lagre
  - 16 + 6 (ECC) bits databuss
  - "smart" grensesnitt mot lager for å kunne la de 32 prosessorene operere samlet som en 32-bits prosessor med flyttalls-ko-prosessor
    - » "Multi dimensional access" (MDA)-lager
  - Programvare kan styre operasjons-modus
    - » bit-seriell-modus (64k ,1-bits PE)
    - » ord-parallell-modus (2k, 32-bits PE)
- **Hyperkube**
  - 16 prosessor pr. brikke
  - 12-dimensjonal hyperkube av brikker (4096 stk)
  - spesial-HW for ruting
  - meldingsutveksling
    - » vanlig hyperkube-ruting (se eks. 8.10)

Prosessor-strukturer i CM-2

- **Ulike logiske PE-array strukturer kan koples opp gjennom to nivåer**
  - 16 prosessorer på en brikke kan organiseres i ulike NEWS-konfigurasjoner
    - » 4 x 4 / 2 x 8 / 1 x 16 / osv....
      - "flipper-nettverk"
  - Et utvalg av forbindelser i hyperkuben kan benyttes for å oppnå en tilsv. struktur mellom brikkene
- **I tillegg kan hver fysiske prosessor emulere et antall virtuelle prosessorer**
  - opp til 64 v.hj.a. tidsmultipleksing
  - kan kommunisere vilkårlig via delt lager
  - (det gir 3 logiske nivåer av samme struktur)
- **"scanning and spread"**
  - scanning (combining)
    - » spesial-HW for hurtig innsamling av data
      - sum, max, min, osv.
    - » kombinerer kommunikasjon og beregning
  - spread
    - » det motsatte (kringkasting (?))
    - » brukes det samme nettverket det andre veien?
  - mer om dette i.f.m. CM-5

Mer om CM-2 (kursorisk) (1)

- **Kilde**
  - Kap. 18 i boka *Computer Architecture, case studies*, av Robert R. Baron og Lee Higbie, Addison Wesley 1992.
- **Thinking Machines**
  - det "sprekeste" firmaet
  - utgangspunkt i Danny Hillis sin PhD
    - » visjonær
    - » kort
    - » lettlest
- **Connection Machine (CM)**
  - 64 k prosessorer
    - » (at disse er 1-bits bit-serielle "sto ofte langt bak med liten skrift")
- **Forbedringer fra CM-1**
  - større adresse rom
  - flyttalls støtte
  - virtuelle PE´er
  - kan være vanskelig å utnytte full parallellitet....
    - » implementasjon av muligheter for 2k 32-bits prosessorer er en innrømmelse
      - (alt er ikke egnet for 1-bits massiv parallellitet)

Mer om CM-2 (kursorisk) (2)

- **Virtuelle prosessorer**
  - Del hver PE sitt lokale lager inn i 2**N blokker
  - Hver PE kan simulere 2**N virtuelle prosessorer
  - Programmereren ser da en maskin med 65536 x (2**N) VP´e
  - Kan f.eks. løse større problemer med parallellitet av typen "et element pr. prosessor"
- **Paris = PARallel Instruction Set**
  - N-dimensjonale tabeller er sentralt
    - » Brukerdefinert datatype
      - a) element-type
      - b) #dimensjoner
      - c) # elementer langs hver dimensjon
    - » b) og c) utgjør *geometrien*
  - Abstrakt SIMD maskin med enkle, bit-serielle virtuelle prosessorer (VP)
  - • VP´ene organiseres i prosessor-set (VP-set)
    - » har geometri tilsv. tabeller
    - » Lagrer "hvert sitt element" i tabeller med lik geometri
  - instanser av tabeller allokeres sammen med et VP-set
  - vanlige ALU-instruksjoner + rikt utvalg av funksjoner for kommunisering av data, og for kombinering av data

MasPar MP-1

- **MasPar**
  - eneste vesentlige gjenlevende SIMD-leverandør
  - samarbeid med DEC
- **Univ. i Bergen har en slik maskin**
  - *parallab*
- **MP-1**
  - 1024, 4k, ... ,16k 4-bits prosessorer
  - X-net
    - » realiserer kommunikasjon med 8 naboer v.h.a. 4 nabolinjer og svitsje-punkt med tre tilstander
  - Hovedarkitektur
    - » Fig. 8.25a
    - » PE array
    - » kontroll-enhet (ACU)
    - » vertsmaskin
      - UNIX, standard IO, Ethernet
    - » høyhastighets IO grensesnitt
  - Prosessor-kort med 1024 PE'er
    - » Fig. 8.25b
    - » 64 stk. PE cluster (PEC)
      - 16 prosessorer i hver
      - internt X-net
      - felles linje mot andre kort

IPC og prosessor-oppbygging i MP-1

- **ACU-PE array kommunikasjon**
  - nettverk for kringkasting og reduksjon
- **X-net**
  - 2D rutenett, torus
  - NEWS + NE, NW, SE, SW
  - 8 naboer
  - Tilsv. Blitzen-nettverk
- **1024 PEC'er sammenkoplet via 1024 x 1024 flertrinns crossbar**
  - maks. 16 kort, hver på 64 PEC'er
  - en inn/ut linje fra hver PEC til S1/S3
  - S1, S2, S3 i Fig. 8.25b) realiserer 1024 x 1024 svitsj
- **Prosessor-oppbygging**
  - Fig. 8.26b.
  - 4 bits ALU
  - instruksjoner ankommer ferdig dekodet
  - register-basert RISC

CM-5, bakgrunn

- **CM-2 erfaringer**
  - SIMD programmering er ikke alltid naturlig
  - ytelse/pris forholdet er bedre for standard mikro-prosessorer enn bit-slice prosessorer
- **MIMD/SIMD diskusjonen**
  - MIMD har i praksis asynkrone prosessorer
  - SIMD har i praksis synkrone prosessorer
  - MIMD gir mest naturlig, fleksibel programmering
  - synkrone prosessorer gir lettere programmering
    - » bedre kontroll over prosessorer/prosesser
  - En kombinasjon er ønskelig
    - » Synkron MIMD-programmering [Natvig90]
      - PRAM pascal, BSP, m.m.....
    - » Connection Machine (1991)
      - synchronised MIMD structure

CM-5 arkitektur

- **SIMD/MIMD kombinasjon**
  - SIMD
  - multiple SIMD
  - synkronisert MIMD (?)
    - » tilsv. bulk synkron parallell (BSP) ???
- **Fra 32 til 16k SPARC-noder ++ !**
  - 2 Teraflops topphastighet (30m x 30m "footprint")
- **Hovedarkitektur**
  - Fig. 8.27
  - meget kraftige PE'er
    - » skal kunne erstattes av "siste nytt i mikroprosessorer" (dvs. generasjons-skalerbar)
  - flere kontrollenheter (CP)
  - flere nettverk
    - » data ("fat-tree")
      - vanlig datatrafikk mellom PE'er, punkt til punkt
      - høy-ytelse
      - variabel-lengde pakker
    - » kontroll (binær-tre)
      - enkelt-verdier og synkronisering
      - hurtig
    - » diagnose
      - oppdage og isolere feil
  - høyhastighets IO-grensesnitt

CM-5, bruker-partisjoner, operativsystem

- **Bruker-partisjoner**
  - En kontroll-prosessor + et knippe PE'er + tilsv. del av data og kontroll-nettverk
  - en PE kan v.hj.a. tidsmultipleksing ("time-sharing") være med i flere logiske bruker-partisjoner
  - dette kalles også "space-sharing"
  - kjører UNIX
  - **kan utveksle data med *andre* partisjoner og IO-enheter via system-kall**
  - **data og kontroll-nettverk for *egen partisjon* kan benyttes direkte uten å gå via system-kall**
    - » god ytelse
- **"operativsystem-bilde"**
  - Fig. 8.28
    - » ulike partisjoner og samarbeid med IO-enheter

CM-5's data-nettverk

- **"4-ary fat tree"**
  - Leiserson (1985)
    - » "binary fat tree" fig. 2.17c. side 82
  - "tykkere nærmere roten"
  - PE'er i løvnoder i et tre gir økt kommunikasjon nærmere rot-noden
  - interne noder er kommunikasjons-svitsjer
  - Fig. 8.29 er *prinsipp*-skisse
  - hver brukerpartisjon må tilsvare et *dedikert sub-tre*
  - "hierarkisk systembuss" der ulike partisjoner er uavhengige
  - en melding velger pseudo-tilfeldig blant ulike ledige alternative veier
    - » lastbalansering
  - "wormhole routing"
  - Realisering
    - » 2 nederste nivåer: bakplan
    - » derette kabler
      - 90% av lyshastighet

CM-5's kontroll- og diagnose-nettverk

- **Kontroll-nettverk**
  - Komplett binær-tre
  - prosessorer som løvnoder
  - brukerpartisjon = subtre
    - » kontrollprosessor koples til som en ekstra løvnode
  - 65-bits pakker
  - tilbyr effektiv utførelse av data-parallel-kode
    - » kringkasting/kombinering
    - » hurtig synkronisering
  - kan "kople rundt" feil
- **Diagnose-nettverk**
  - binær-tre (ikke nødvendigvis komplett)
  - en eller flere diagnose-prosessorer i rot-noden
  - selv-diagnose

CM-5: Kontroll-prosessorer og prosesserings-noder

- **Kontroll-prosessor**
  - vanlig arbeidsstasjon
  - CMOST
    - » UNIX-basert
    - » utvidelser for håndtering av parallelle ressurser
  - (Fig. 8.30)
- **Prosesserings node**
  - SPARC prosessor
    - » hurtig kontekst-bytte er viktig for tidsdeling mellom flere bruker-partisjoner
  - "memory-controller"
  - intern 64 bits databuss
  - Fig. 8.31
  - 4 Vektorenheter
    - » erstatter "memory controller"
    - » 128 Mflops (64 bits)
  - Fig. 8.32

CM-5 IPC

- **Utvidelse av CM-2 sine mekanismer for IPC**
- **4 klasser som kan anvendes på ulike datatyper**
  - tabell, matrise, lenket liste, irregulært mønster m.m.
  - CM-5 HW og SW støtter dette
  - Duplisering, Fig. 8.33
    - » utføres av kontroll-nettverket
    - » Anvendelser: nedlasting av kode og data, synkronisering, konfigurasjon av partisjoner
  - Reduksjon, Fig. 8.34
    - » utføres av kontroll-nettverket
    - » Eks: sum, max, min, OR, AND, etc...
  - Permutering, Fig. 8.35
    - » Anvendelser: matrise-beregninger, FFT, m.m.
    - » Eks. 8.11 utgår
  - Parallel prefiks, Fig. 8.36
    - » utføres av kontroll-nettverket
    - » også kalt "forward scan"
    - » kan utføres i parallel (n / log n) raskere enn serielt

Mer om CM5 (kursorisk)

- **(Fra GordonBell?)CM-5**
  - skalerbar programvare
  (dvs. SUN-arb.stasj, IBM stormaskin, CM super-
  datamaskin, samme FORTRAN-90 program)
-
- **Se og CACM meget fersk**