

# Supercomputernes historie

- de til enhver tid hurtigste computere

Ole Holm Nielsen

PhD i fysik

IT-chef, Institut for Fysik

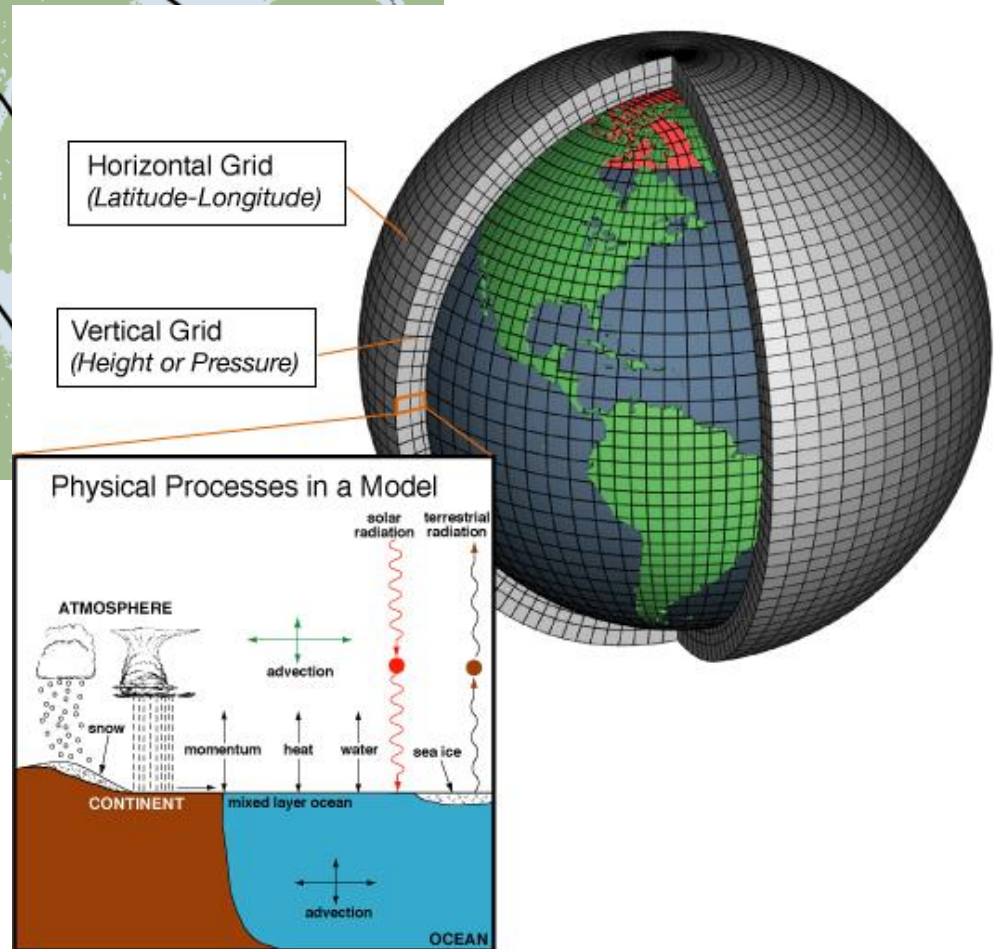
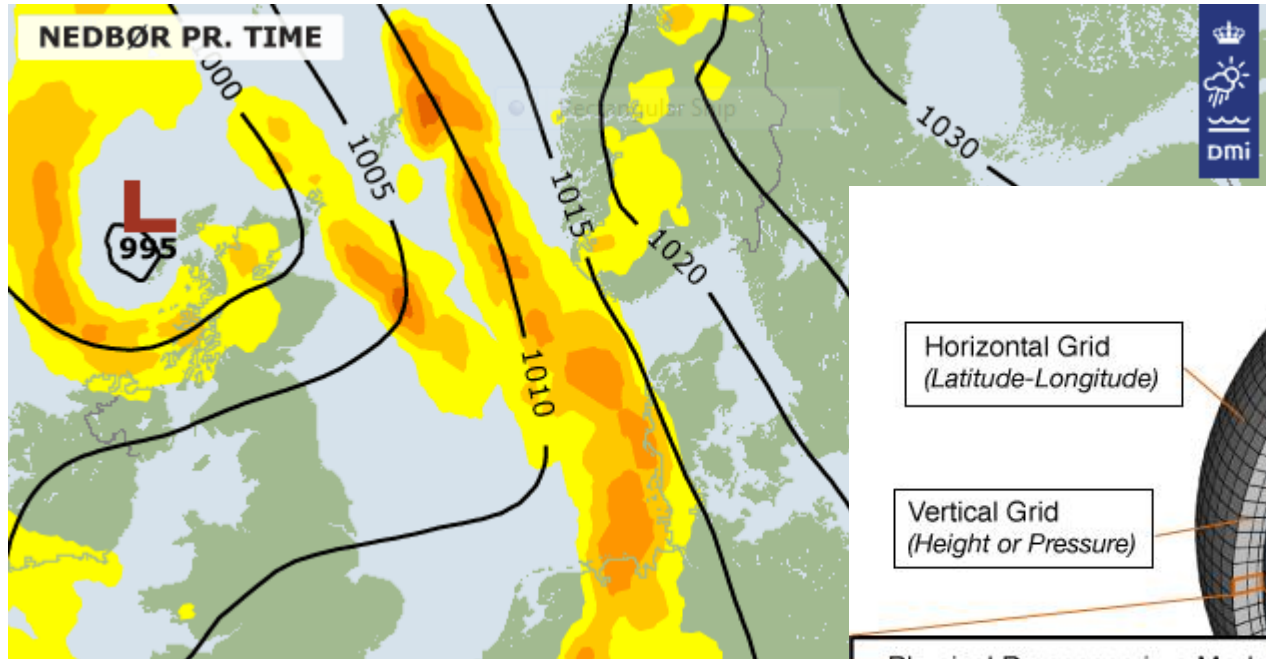
Danmarks Tekniske Universitet

Kongens Lyngby

# Hvorfor supercomputere?

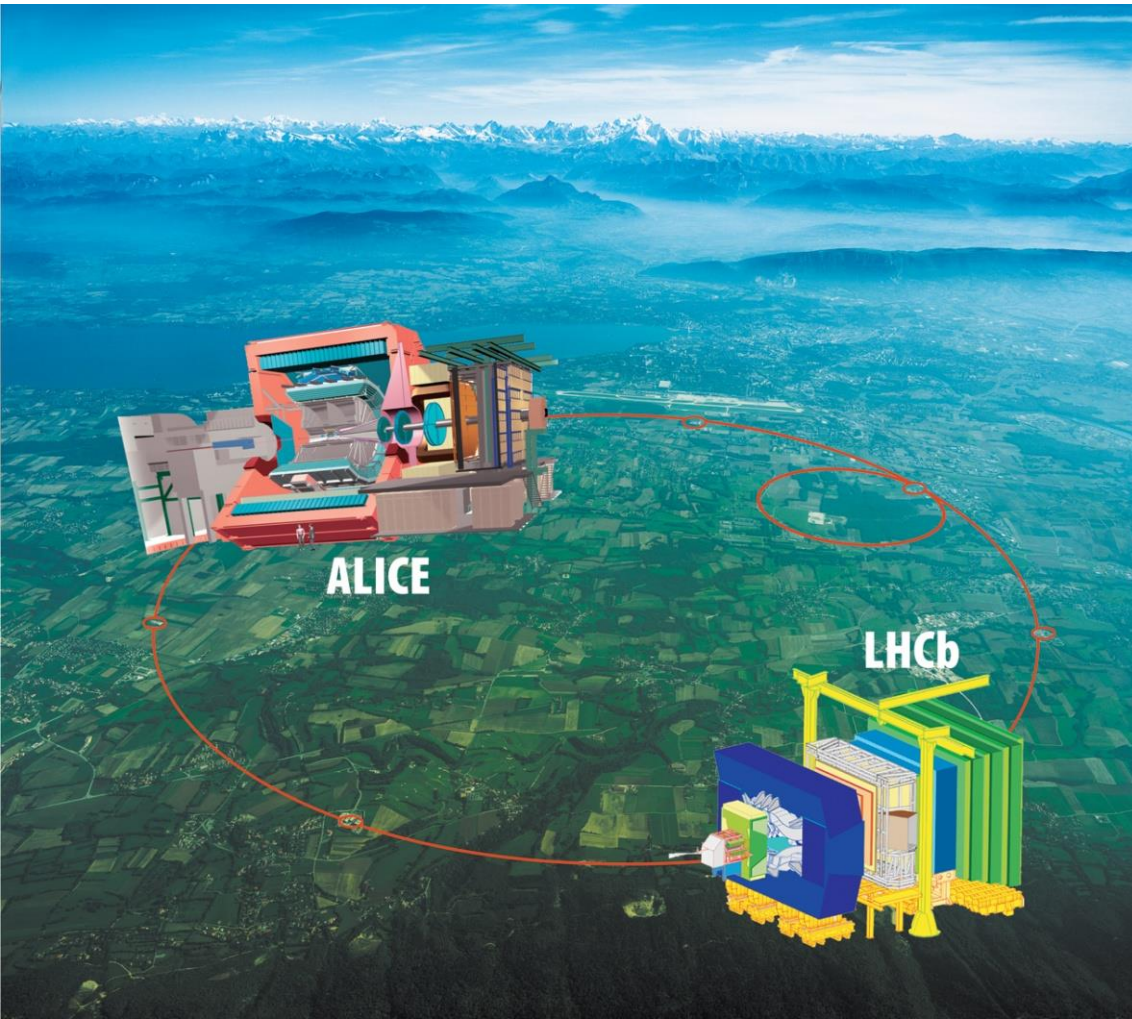
- Behov for at løse **komplekse matematiske modeller** med computerberegninger: F.eks. naturvidenskab, ingeniørvidenskab, produktdesign:
  - Løsning af kvantemekanikkens Schrödinger-ligning (kemi, materialefysik).
  - Simulering af plasma-dynamikken i stjerner eller universet.
  - Vejrudsigter (luft og vand i atmosfæren).
  - Luftstrømning (f.eks. vindmøller, fly, biler).
  - Mekaniske konstruktioner (f.eks. broer, biler).
- Behandling af **meget store mængder data**: Seismiske eller genetiske undersøgelser, billedbehandling, meteorologiske data, forretningsdata:
  - Seismisk undersøgelse af jordlag (olie-efterforskning).
  - Bio-informatik (gen- og protein-sekvens databaser).
  - Højenergi-fysik (CERN's LHC accelerator).
  - Satellitfotos (vejret, jord-ressourcer, forurening).
  - Analyse af økonomiske transaktioner (kreditkort, aktiehandler).
  - Overvågning af mennesker, analyse af trafik på Internettet, spionage.

# Eksempel: vejrudsigter



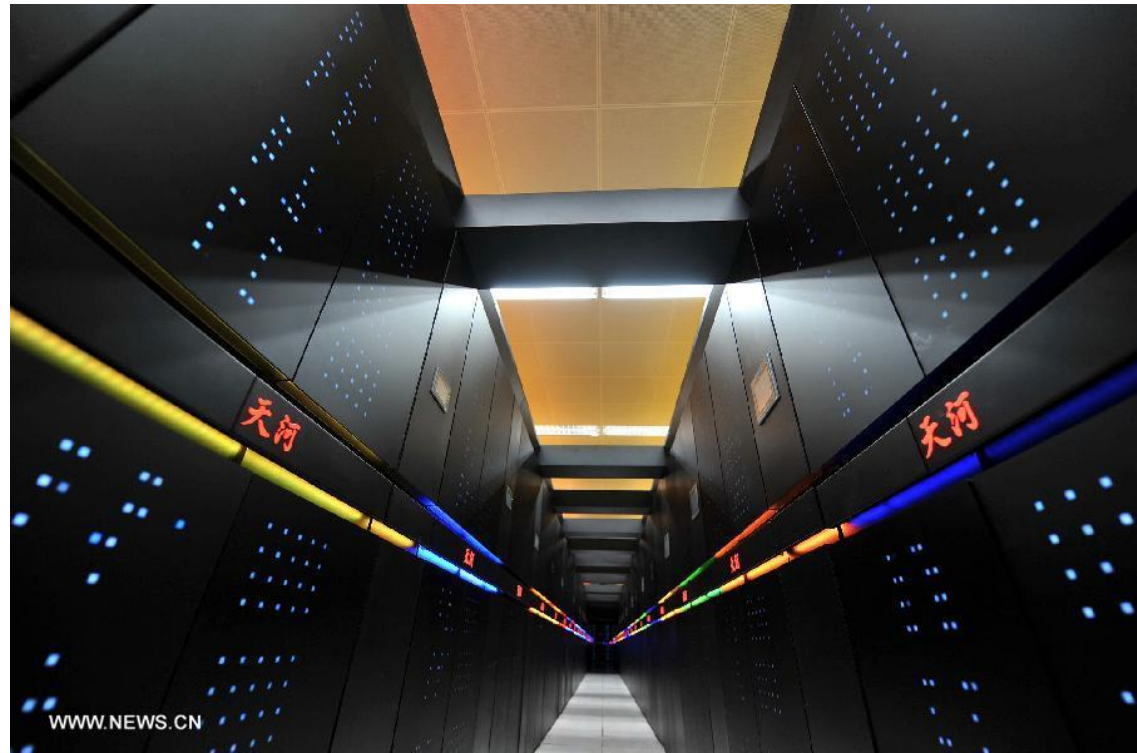
# Eksempel: CERN LHC datacenter

*Large Hadron Collider: 27 km accelerator*  
genererer ca. 30 PetaBytes data/år



# Verdens største supercomputer

- En supercomputer er en computer, som har **meget større regnekraft** end "almindelige computere" – men få år senere bliver den samme maskine reduceret til en "almindelig computer" pga. den tekniske udvikling (er nok kasseret forinden).
- Udviklingen af computerteknologi går rivende hurtigt!
- To gange om året udgives TOP500 listen over verdens største supercomputere ([www.top500.org](http://www.top500.org)).
- Verdensrekorden (6/2015):  
Tianhe-2 (Mælkevej-2)  
i Kina:
  - 3.120.000 CPU-kerner:  
32.000 Intel Xeon CPUer +  
48.000 57-kerne Xeon-Phi
  - 1.024.000 GB RAM +  
384.000 GB RAM (Xeon-Phi)
  - Ydeevne: 33,862 PetaFLOPS  
(=  $33.9 \cdot 10^{15}$  operationer/sekund)  
Teoretisk peak: 54.9 PetaFLOPS.
  - Strømforbrug: 18 MW
  - Pris: ca. 2.5 mia DKK




# Supercomputer TOP500 listen

Juni 2015



## System-typer:

- 87% er Linux (UNIX) klynger (clusters).
- 13% er Linux MPP (single system image).
- 1 system (nr. 436) kører Microsoft Windows HPC2008.
- 0 systemer med enkelt-CPU eller multi-CPU (SMP).
-  Nr. 161 og 243.

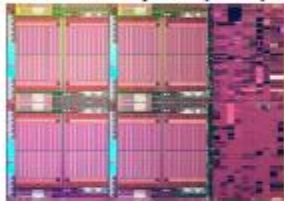
# Typisk cluster-node i TOP500 computer



## Commodity plus Accelerator Today

### Commodity

Intel Xeon  
8 cores  
3 GHz  
8\*4 ops/cycle  
96 Gflop/s (DP)

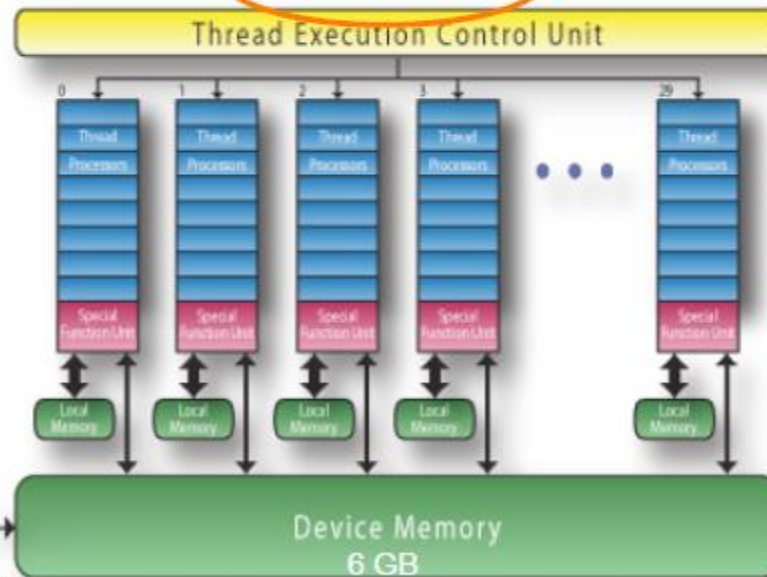


Interconnect  
PCI-X 16 lane  
64 Gb/s (8 GB/s)  
1 GW/s

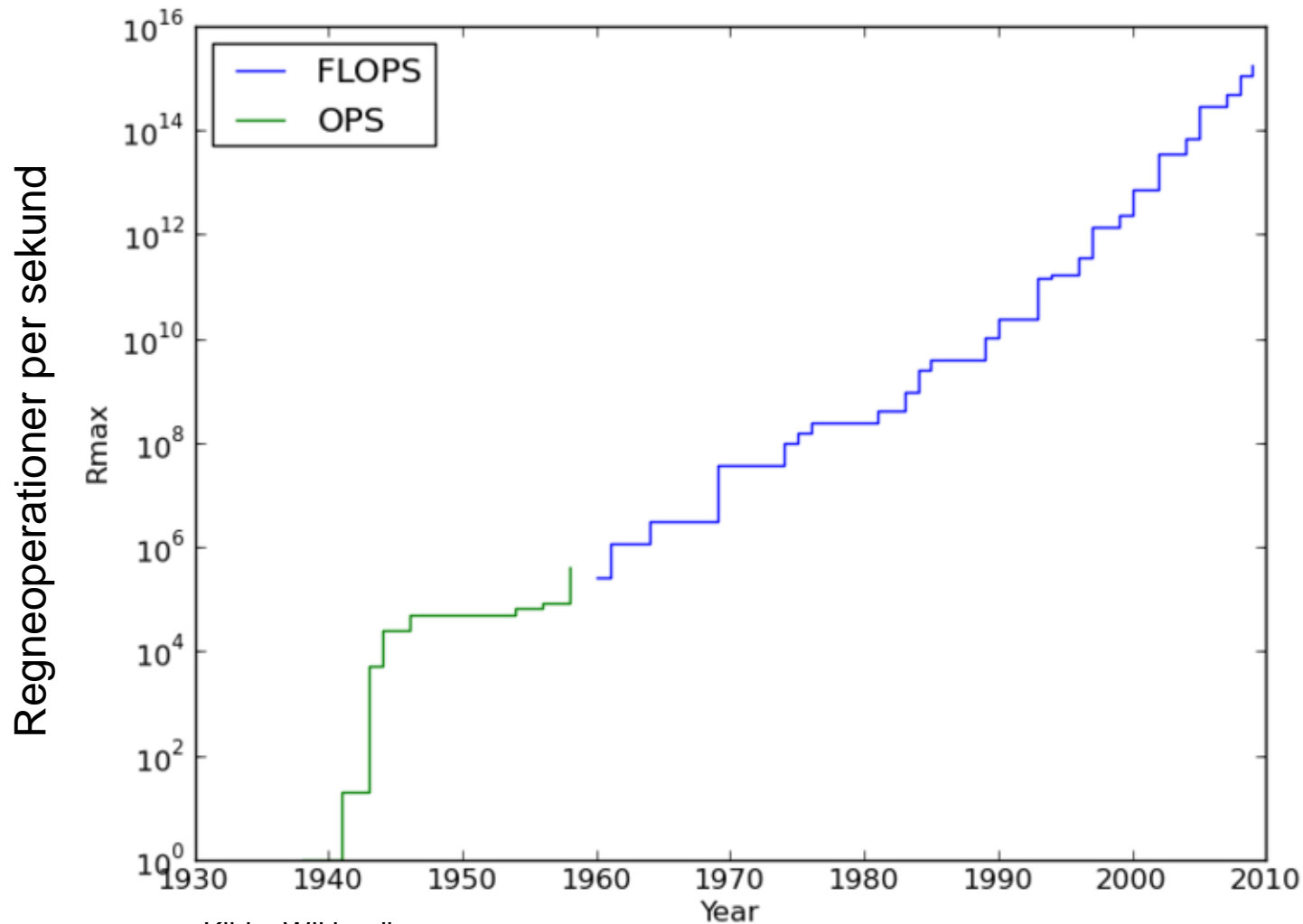
### Accelerator (GPU)

Nvidia K20X "Kepler"  
2688 "Cuda cores"  
.732 GHz  
2688\*2/3 ops/cycle  
1.31 Tflop/s (DP)

192 Cuda cores/SMX



# Historisk perspektiv: Computer regnehastighed over tid



Kilde: Wikipedia



# Menneskelige computere

- Ordet "computer" brugtes på engelsk siden 1613. Computeren skal udføre beregninger efter strikse forskrifter.
- New York Times (1892): Jobopslag tilbyder tjenestemandstilling som computer.

## *A COMPUTER WANTED.*

WASHINGTON, May 1.—A civil service examination will be held May 18 in Washington, and, if necessary, in other cities, to secure eligibles for the position of computer in the Nautical Almanac Office, where two vacancies exist—one at \$1,000, the other at \$1,400.

The examination will include the subjects of algebra, geometry, trigonometry, and astronomy. Application blanks may be obtained of the United States Civil Service Commission.

**The New York Times**

Published: May 2, 1892

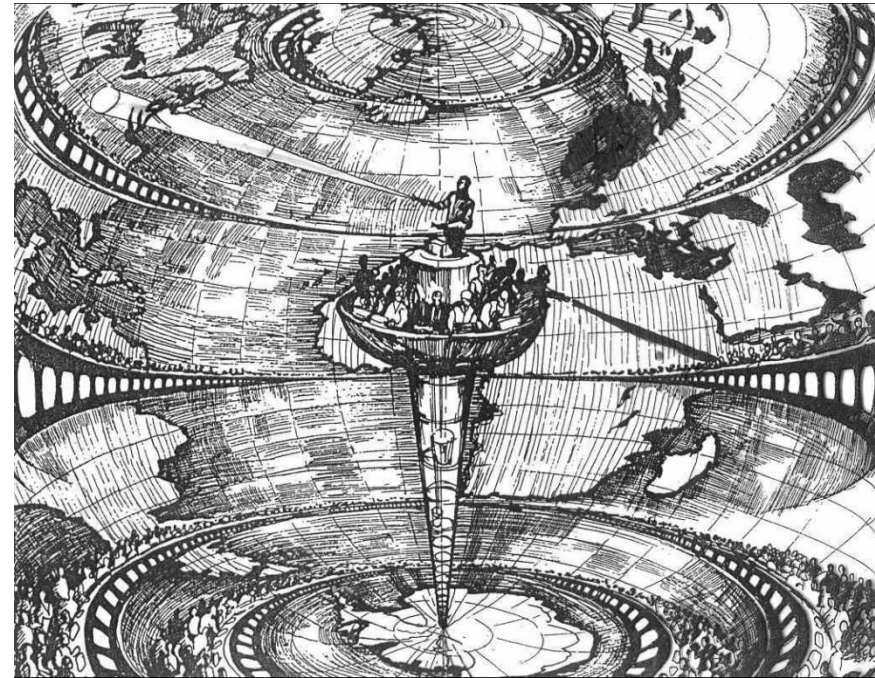
Copyright © The New York Times



Library of Congress (USA), foto ca. 1920

# Supercomputernes historie

- Vejrforudsigelser: Lewis Fry Richardson, *Weather Prediction by Numerical Process* (1922):  
After so much hard reasoning, may one play with a fantasy? Imagine a large hall like a theatre, except that the circles and galleries go right round through the space usually occupied by the stage. The walls of this chamber are painted to form a map of the globe. The ceiling represents the north polar regions, England is in the gallery, the tropics in the upper circle, Australia on the dress circle and the Antarctic in the pit. A myriad **computers** are at work upon the weather of the part of the map where each sits, but each **computer** attends only to one equation or part of an equation. The work of each region is coordinated by an official of higher rank.



# Computere på *Manhattan* projektet

- *Manhattan* projektet på *Los Alamos* national-laboratoriet i New Mexico, USA fra 1943-1945. Computere benyttes til at beregne detonation af atombomber.
- Menneskelige computere versus hulkort-maskiner:  
To perform some of these repetitive calculations, a group of scientists' wives were recruited to form a central computing pool. These "computers" included Stanley Frankel's wife, Mary; Josephine Elliott; Beatrice Langer; Mici Teller; Jean Bacher; and Betty Inglis.
- The new IBM punched-card machines were devoted to calculations to simulate implosion, and Metropolis and Feynman organized a race between them and the hand-computing group. "We set up a room with girls in it. Each one had a *Marchant*. But one was the multiplier, and another was the adder, and this one cubed, and all she did was cube this number and send it to the next one," said Feynman.  
For one day, the hand computers kept up: "The only difference was that the IBM machines didn't get tired and could work three shifts. But the girls got tired after a while."



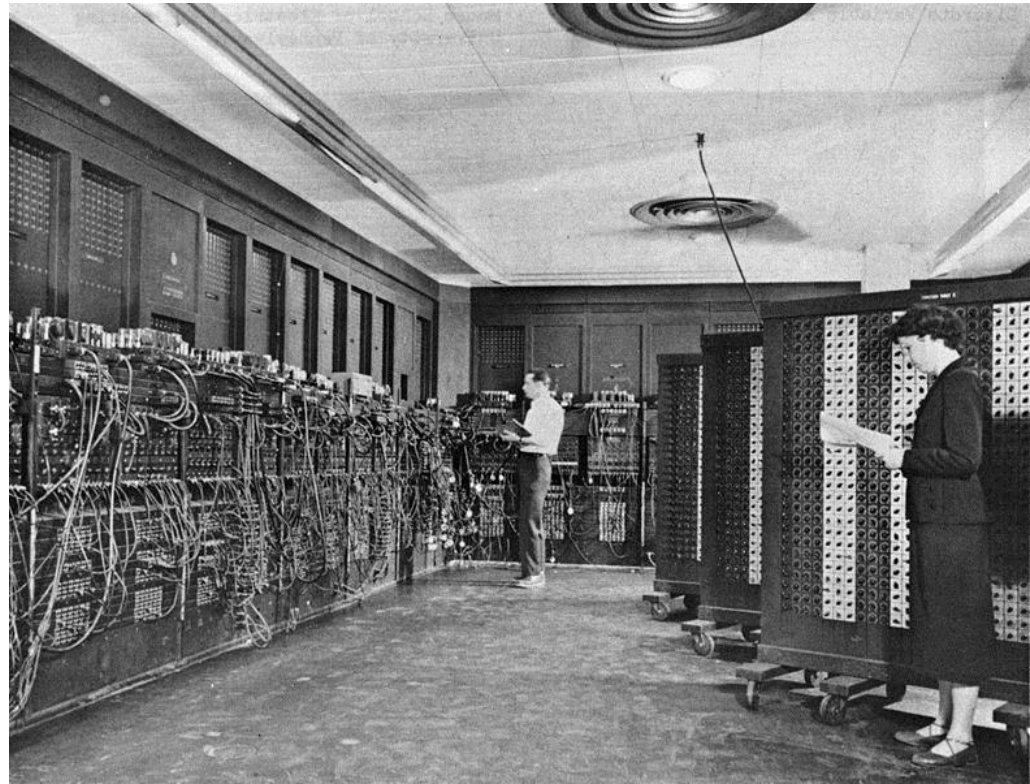
# ENIAC (ca. 1946)

Electronic Numerical Integrator And Computer (US Army)

ENIAC radiorør



Programmering af ENIAC med kabler og kontakter



## IBM annonce (1952)

An IBM Electronic Calculator speeds through thousands of intricate computations so quickly that on many complex problems it's like having 150 EXTRA Engineers.



### 150 Extra Engineers

An IBM Electronic Calculator speeds through thousands of intricate computations so quickly that on many complex problems it's like having 150 EXTRA Engineers.

No longer must valuable engineering personnel . . . now in critical shortage . . . spend priceless creative time at routine repetitive figuring.

Thousands of IBM Electronic Business Machines . . . vital to our nation's defense . . . are at work for science, industry, and the armed forces, in laboratories, factories, and offices, helping to meet urgent demands for greater production.

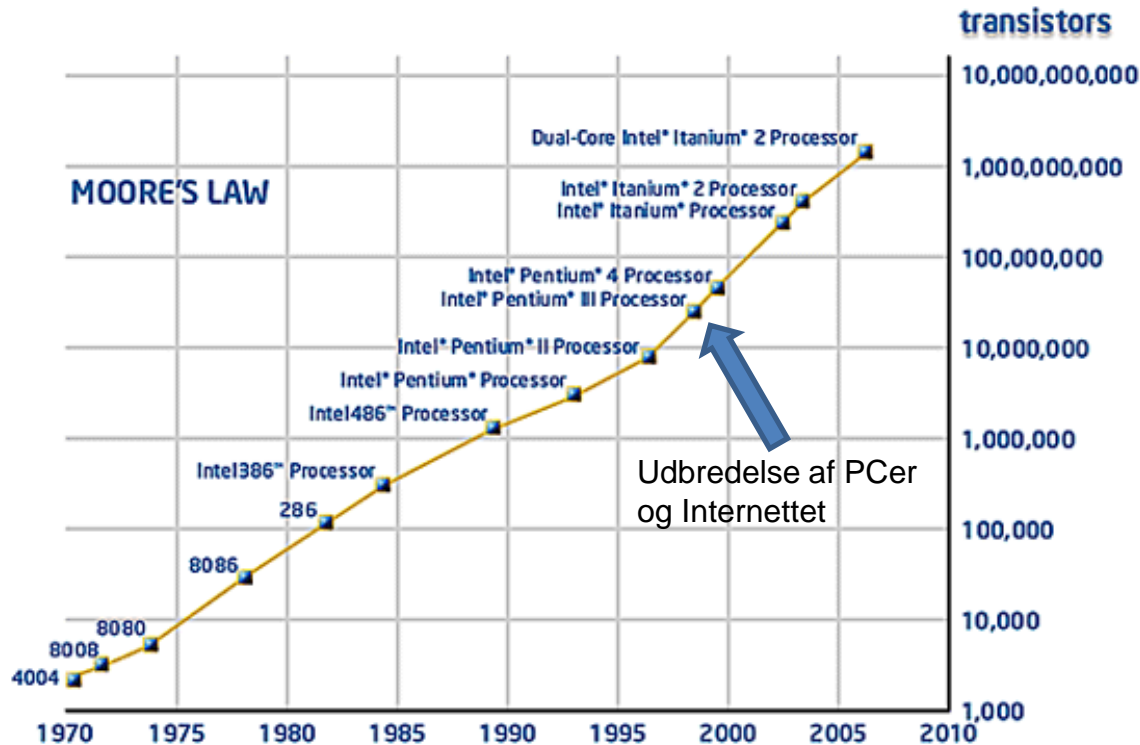
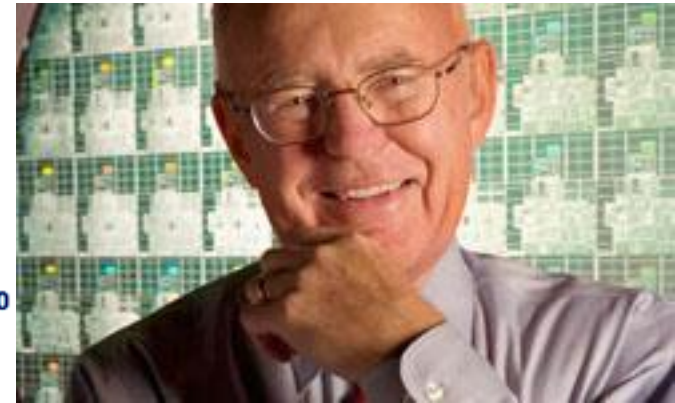


INTERNATIONAL BUSINESS MACHINES

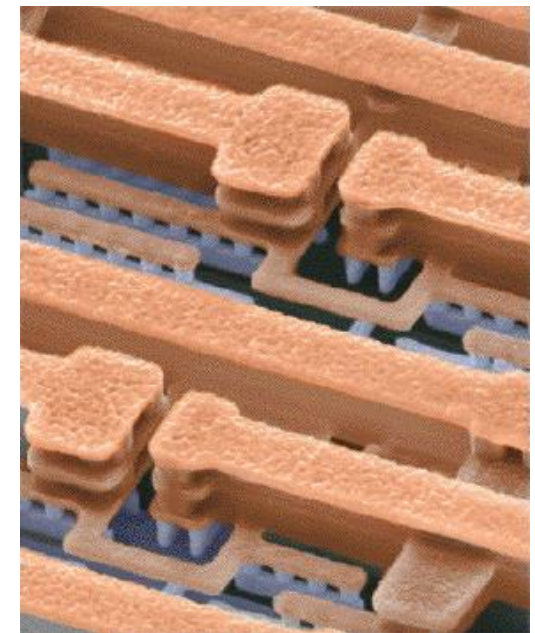
# Computerteknologiens rivende udvikling

- **Moore's lov** (1965): Silicium-chips får dobbelt så mange transistorer hver 18-24 måneder.
- **Mikroprocessor-revolutionen:**  
Den tredje industrielle revolution?

Gordon Moore

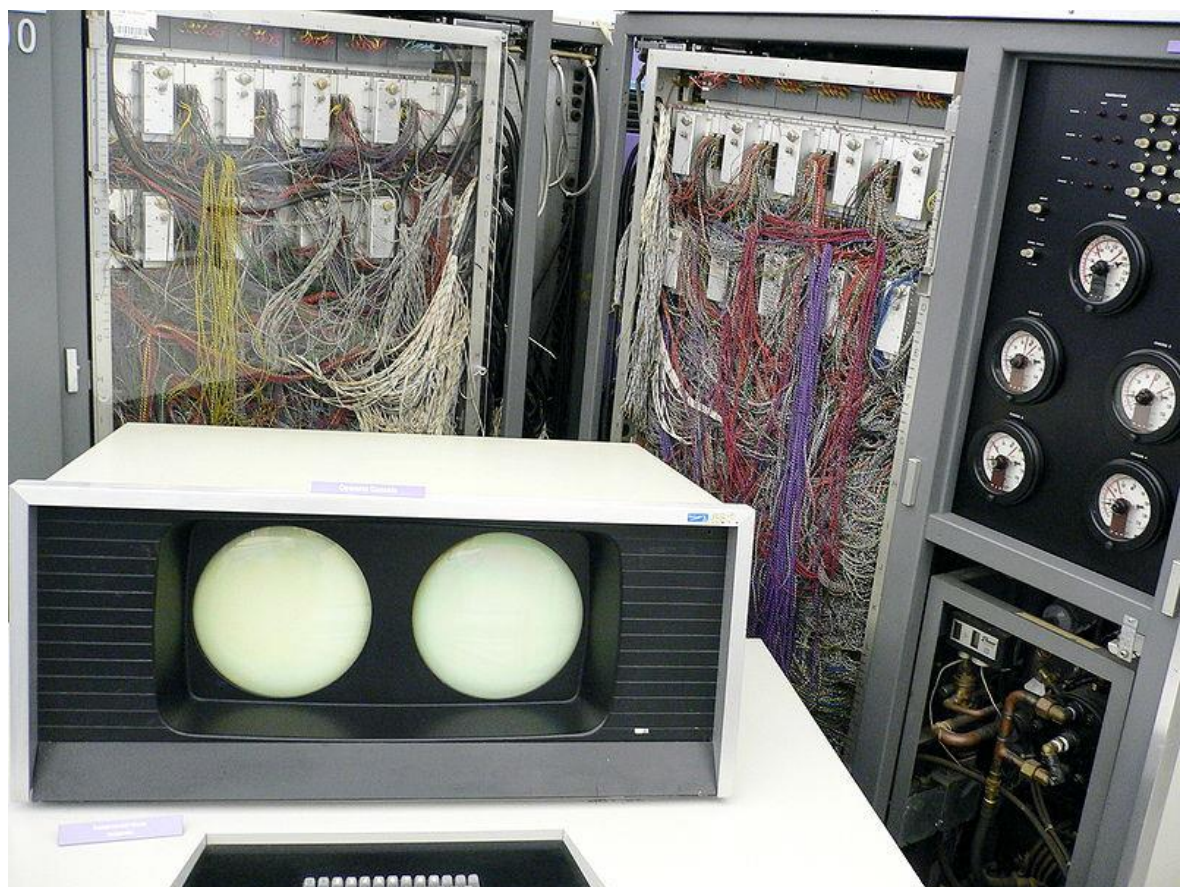


Kilde: Intel



# Supercomputer: CDC 6600 (1964)

- Ydeevne: 1 mill. FLOPS.
- RAM: 64 kiloBytes.
- Verdens hurtigste fra 1964 til 1969.
- Bygget med Silicium-transistorer.
- **Parallelisme:**  
10 funktionsenheder.
- CDC 6400 installeret i 1971 på det *Regionale Edb-Center ved Aarhus Universitet (RECAU)*.  
Noget langsommere end CDC 6600.



# Supercomputer: CDC 7600 (1969)

- Ydeevne: 10 mill. FLOPS.
- Verdens hurtigste 1969 til 1975.
- RAM: 512 kiloBytes + 4 MegaBytes sekundær RAM.
- **Parallel pipelining** af instruktioner forøger regnehastigheden.
- Foredragsholderen brugte i 1973-74 *University of California at Berkeley's* CDC 7600 som studenterprogrammør på et forskningsprojekt i højenergifysik på *CalTech*.

CDC 7600 på *Lawrence Livermore National Laboratory* (ca. 1971)





# Seymour Cray (1925-1996)



- **Seymour Cray:** Berømt computer-arkitekt fra USA. Cray designede verdens hurtigste supercomputere fra 1964 og fremefter.
- Ansat i *Control Data Corporation (CDC)*, 1950. Designede CDC 6600 og CDC 7600.
- Grundlagde *Cray Research, Inc.* i 1972.
- Leverer CRAY-1 (1976) og mange efterfølgende modeller.
- Død 1996 i trafikuheld.
- Firmaet CRAY leverer fortsat nogle af de største supercomputere i verden.

# Vektor supercomputere

- Seymour Cray's **CRAY-1** introducerede **parallele beregninger på vektor-registre** i CPUen.
- Hvis CPUen skal udføre samme beregning på en lang liste af tal (en *vektor*), så kan man effektivisere beregningerne "på samlebånd":
  - Læsning af tallene fra fortløbende positioner i RAM-lageret til CPU'ens vektor-registre.
  - Udføre samme aritmetiske operation på alle tallene i en pipeline.
  - Gemme resultaterne i fortløbende positioner i RAM-lageret.
- Eksempel på computer programkode (i *Fortran*) med vektorer:

```
do j=1,1000
  do i=1,1000
    c(i,j)=c(i,j)+a(i)*b(j)
  end do
end do
```
- Gevinst: **Mange** flere regneoperationer pr. sekund!  
Programmer kan blive **10-20 gange hurtigere**, hvis man benytter vektor-beregninger med parallelle, identiske beregninger.

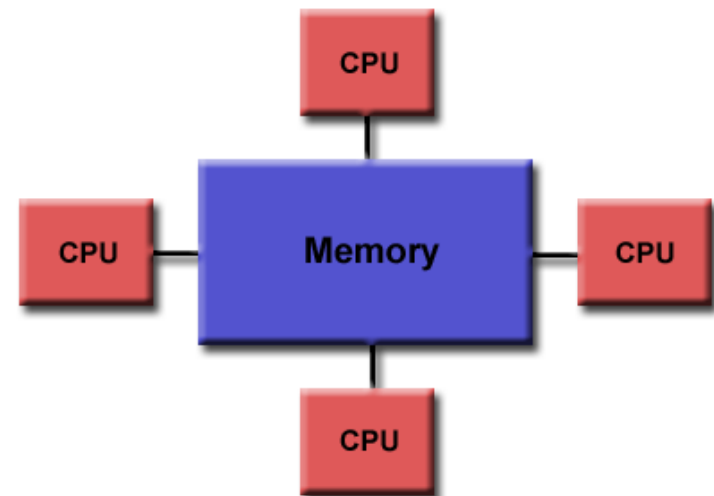
# Supercomputer: CRAY-1 (1975)

- **Vektor** arkitektur.
- Ydeevne: Op til 250 mill. FLOPS.
- RAM: Op til 32 MegaBytes.
- 64 vektor-registre med 64-bit tal.
- Særdeles populær supercomputer i den vestlige verden (>80 solgt).
- Pris: I omegnen af 10 mill. US\$.
- Foredragsholderen benyttede CRAY-1 hos SAAB i Linköping i 1984 til forskningsprojekt ved *NORDITA* (København).



# Shared Memory supercomputere

- **Erfaring:** Man løb ind i grænsen for, hvor hurtigt 1 CPU kan regne.
- **Løsning:** Hvis nu **N CPUer** ( $N=2,4,8,16$ ) kan arbejde med det samme, fælles RAM datalager, kan computeren så regne **N gange hurtigere i parallel?**
  - Ja, hvis en meget stor beregning kan deles op i mindre portioner, som kun har få data til fælles, så kan hver portion i mange tilfælde beregnes separat.
- **Udfordring:** Programmøren skal instruere de **N CPUer**, så de arbejder uafhængigt og med fuld hastighed. Men en gang imellem bliver CPUerne nødt til at udveksle data med hinanden, før beregningen kan fortsætte.
- **Ulemper:**
  - Datatrafik til datalageret vokser som  $N^2$ .
  - Prisen på hardware vokser som  $N^2$ .
  - Alle CPUer skal kende indholdet i de andres **RAM cache** (hvis en sådan benyttes), og synkronisere adgang til RAM.
- Nyt begreb: **RAM cache.**



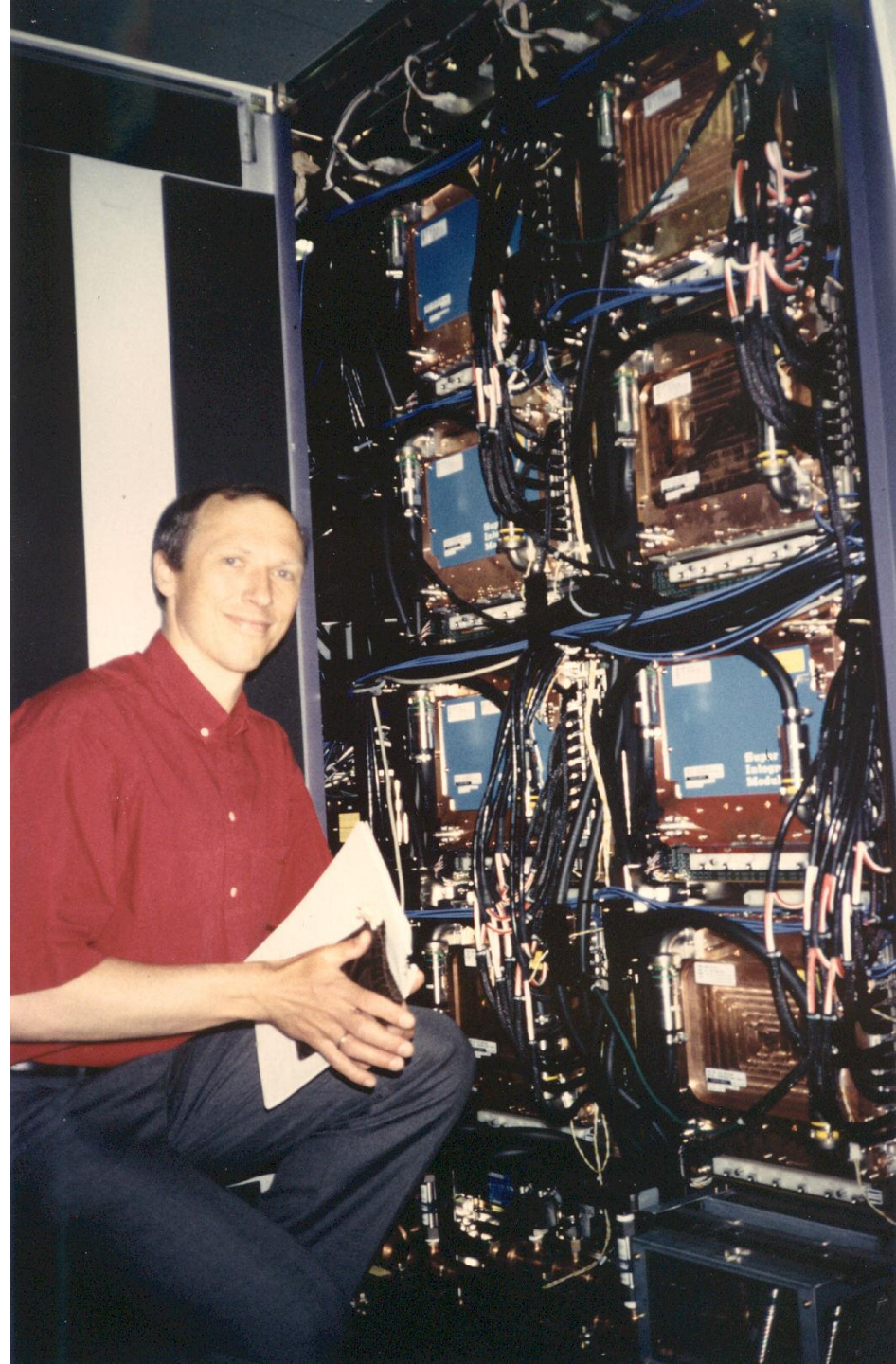
# Supercomputer: CRAY-2 (1985)

- *Shared Memory*, **parallel vektor** arkitektur.
- Ydeevne: 488 mill. FLOPS pr. CPU.
- 4-8 CPU-enheder.
- RAM: 4 GigaBytes.
- Køling af elektronik med "*Fluorinert*" væske.
- Foredragsholderen benyttede CRAY-2 ved *École Polytechnique* (nær Paris) i 1986 til forskningsprojekt ved *NORDITA*.



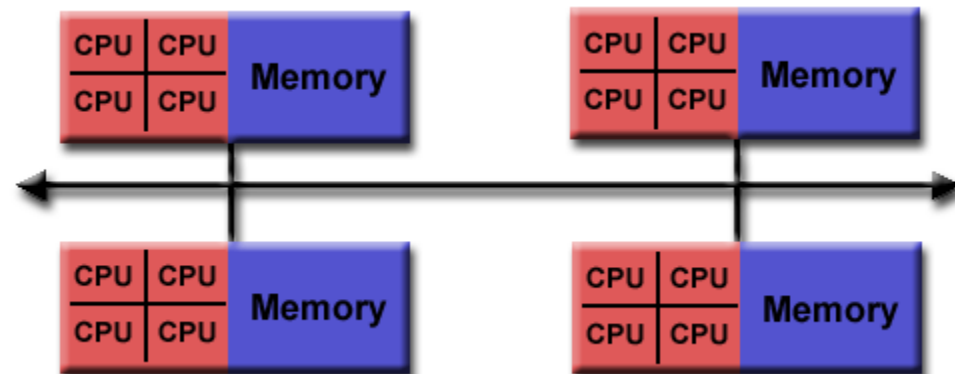
# Fujitsu: Japansk supercomputer

- Foredragsholderen ved 32-CPU *Shared Memory* supercomputer **Fujitsu VPP500/32**.
- Verdens 7. hurtigste computer i 1993.
- 46.1 GigaFLOPs.
- Pris >100 mill. DKK.
- Omprogrammering af stort computerprogram fra *DTU* til at udnytte mange CPUer i **parallel**.



# *Distributed Memory* supercomputere

- **Erfaring:** *Shared Memory* computere bliver meget komplekse og dyre at fremstille, når mere end ca. 20-30 CPUer sættes sammen. Komplexitet stiger som  $N^2$ . Problemer med *cache coherency*.
- **Løsning:** Mindre computere og hurtige data-netværk kan nu **masseproduceres**, så man laver bare **én stor parallel computer ud af mange små!** *Distributed Memory* betyder uafhængige computere.
- **Fordele:** Hvis netværket kan gøres meget stort og hurtigt, kan man bygge supercomputere med 100.000-vis (eller millioner!) af CPUer.
- **Udfordring:** Programmøren skal styre et stort antal CPUer samtidigt, og sørge for at CPUerne får de rigtige data tilsendt via netværket.
- **Ulemper:**
  - Programmøren skal selv sørge for at sende data de rigtige steder hen.
  - Fjerntliggende data tager meget længere tid at hente end lokale data.



# Massivt parallelle supercomputere

- **Connection Machine CM-5.**
- *MIMD Distributed Memory* arkitektur, single OS image MPP.
- Antal **CPUer**: Op til 1024
- **Netværk**: *Fat Tree network* forbinder alle CPUerne uden flaskehalse.
- Ydeevne: Op til 128 GigaFLOPS
- RAM: Op til 32 GigaBytes
- Foredragsholderen benyttede CM-5 ved *Joint Research Center for Atom Technology (JRCAT)* i Tsukuba, Japan fra 1992-1994 i et samarbejde med *DTU*.

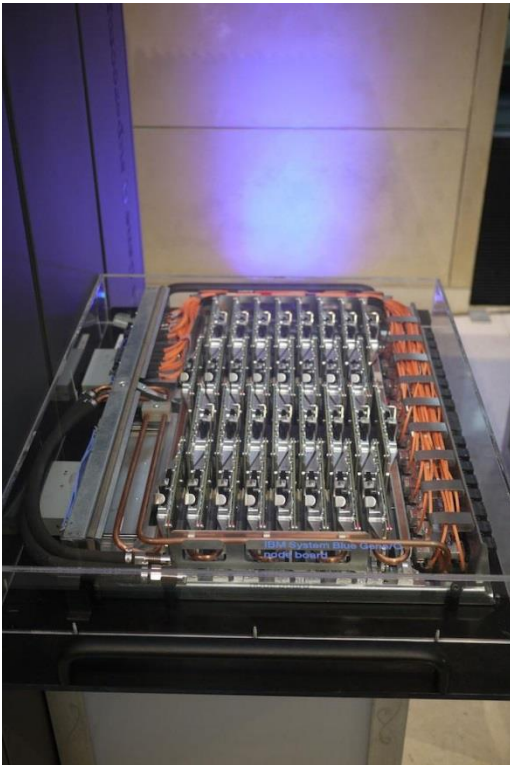
*Thinking Machines Corp. CM-5*





# IBM *BlueGene*

- Største BlueGene/Q er installeret 2012 hos *Lawrence Livermore Nat'l Lab*.
- 1.572.864 CPUer. Fem forskellige netværk forbinder alle noder.
- Ydeevne: 17,173 PetaFLOPS.
- *CAMD* forskningscentret på *DTU* har benyttet en IBM BlueGene/P i USA med 163.000 CPUer.



# Grafik-processorer (GPUer)

- **Erfaring:** Udviklingen af computerspil til PCer har ført til special-designede chips, der kan udføre store antal beregninger i **parallel**.
- En **GPU** (*Graphical Processing Unit*) med mange små regne-kerner *kan* være meget hurtigere end en CPU.
- **Løsning:** Sæt en GPU i computeren, og lad den udføre de data-parallele beregninger, som den er bedst til.
- **Gevinst:** Måske 10-20 gange hurtigere (og måske negativ gevinst!!).
- **Udfordringer:**
  - GPUer er kun god til mindre datamængder, og der skal være 1000-vis af beregninger i kø til kernerne hele tiden.
  - Mange større opgaver passer overhovedet ikke til GPUer.
  - Programmeringen af GPUer er ganske besværlig.



# Linux cluster med GPUer

*"Stampede"* maskinen ved *Texas Advanced Computing Center (TACC)*, 2014.

- Nr. 8 på TOP500 listen.
- Performance 5,168 PetaFLOPS (8,5 peak)
- 6400 servere à 16 CPU kerner.
- 6880 Intel Xeon Phi 61-kerne *"accelerator"* CPUer.
- 4,5 MW strømforbrug.



# Parallelisme i historisk perspektiv

Regnekraften er dels forøget pga. flere og hurtigere transistorer (**Moore's lov** om eksponentiel vækst), dels pga. mange former for **parallelisme**:

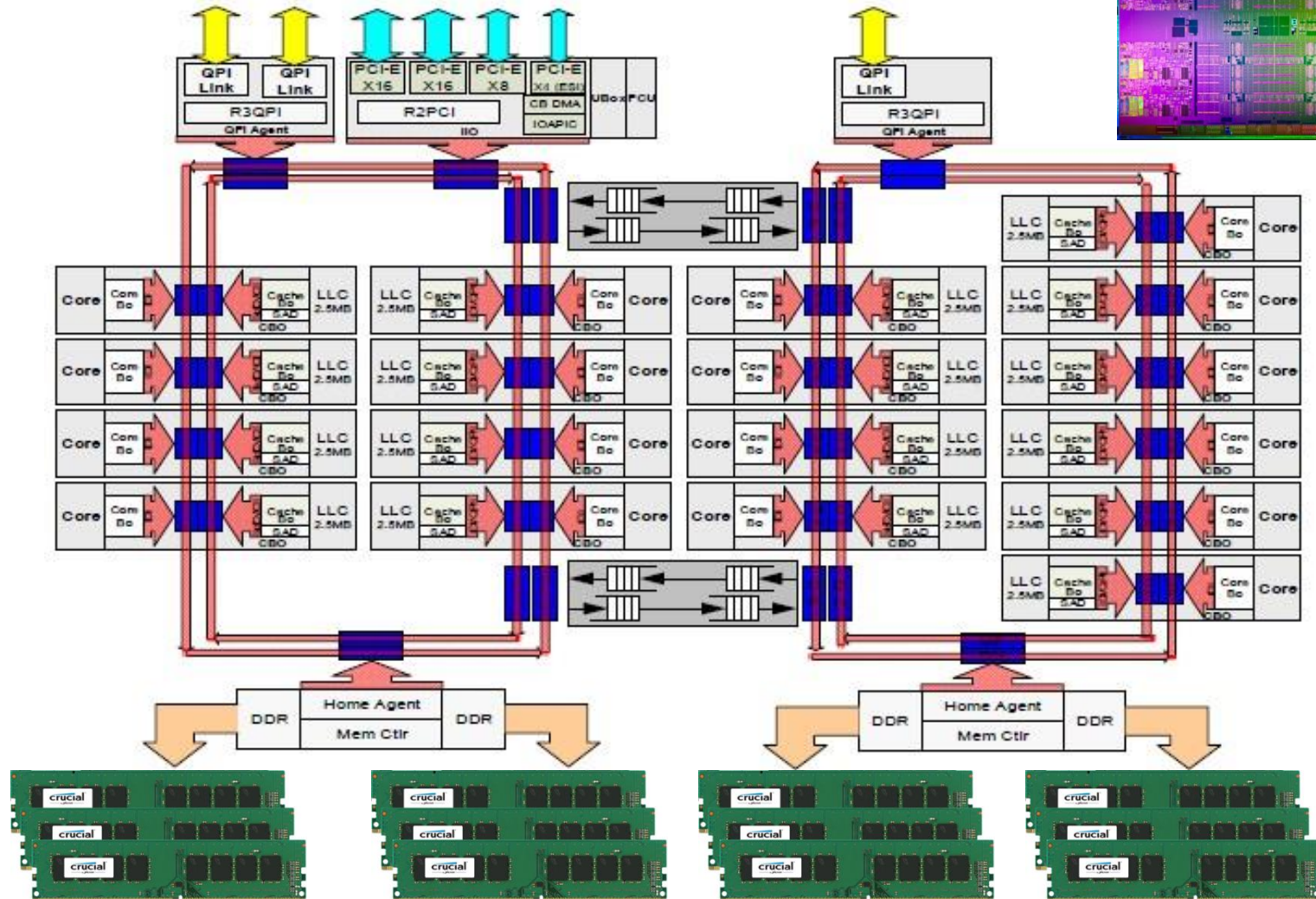
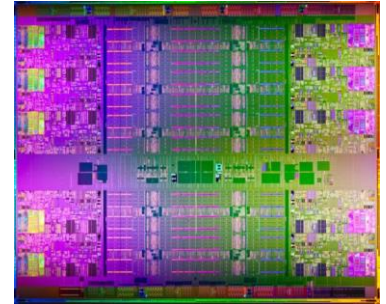
- **Instruktions-parallelisme:** CPUen laver flere operationer samtidigt, bl.a. dynamiske instruktions-køer og uafhængige eksekverings-tråde (*threads*).
- **Vektor-beregninger:** Mange ens beregninger på samlebånd. *Streaming SIMD Extensions (SSE)* på moderne CPUer.
- **SMP-CPUer:** Op til 10-tals CPUer regner på samme fælles RAM-lager. Hver CPU har normalt sit eget lokale *cache* lager.
- **Mange CPUer** ( $10^2$  til  $10^3$ ) på én chip (f.eks. GPUer) kan løse visse **data-parallele** opgaver meget hurtigt.
- **Netværks-parallelisme:** Uafhængige compute-noder ( $10^2$  til  $10^6$ ) arbejder i parallel. Lokalt Linux OS. Data udveksles via netværket.

**Alle disse former for parallelisme findes samtidigt i de moderne supercomputere!**

**Numeriske algoritmer og programmerings-metoder skal udvikles, så beregningerne udnytter parallelismen så optimalt som muligt.**

# Eksempel: Intel Xeon E7 v3 "Haswell" CPU

(18 CPU-kerner, 5.7 mia. transistorer, 662 mm<sup>2</sup>)



# Supercomputere i Danmark

I 2015 er der adskillige supercomputere i Danmark.

De fleste er installeret på universiteterne:

- DTU-CBS/DeIC *Computerome* Life Science cluster på DTU/Risø: 15.120 CPUer, 410 TeraFLOPS. TOP500 #161
- SDU/DeIC *Abacus-2.0* cluster i Odense: 9.360 CPUer, 288 TeraFLOPS. TOP500 #243
- Vestas i Århus: 15.672 CPUer, 162 TeraFLOPS.
- *Grendel* på AU: 7.200 CPUer (84 TeraFLOPS) + 140 GPUer (152 TeraFLOPS).
- *Steno* på KU: ca. 10.000 CPUer + 200 GPUer.
- *Niflheim* på DTU: 7.592 CPUer, 80 TeraFLOPS.
- DMI: 2 stk. Cray-XT5 à 18 TeraFLOPS (fra 2008; udskiftes i 2016).

# Dansk eksempel: *Niflheim*

- DTU og SDU byggede i 2002 store parallelle supercomputere i DK med ca. 500 kontor-PCer koblet sammen med *Ethernet* netværk.
- *Niflheim* ydeevne: 2.2 TeraFLOPS.
- Navnet: *Niflheim* er "landet med is og tåge" i nordisk mytologi.



# *Niflheim 2010*

- Danmarks hurtigste supercomputer 2009-2010.
- Antal compute-noder i netværket: 864.
- 5600 CPU kerner.
- Ydeevne: 57 TeraFLOPS.
- Strømforbrug:  
ca. 150 kW





# Opsummering

- Siden omkring 1940 har computere spillet en stor, og senere helt afgørende, rolle for forskning, teknologi, og utallige andre menneskelige aktiviteter – især efter at *Internettet* (1969) og *World Wide Web* (1989) blev opfundet.
- **Mikroprocessor-revolutionen** (dvs. transistor-kredsløb lagt ned på en silicium-skive) startede omkring 1965 og er vokset **eksponentielt** siden da. Antal fordoblinger af transistorer er ca. 25, dvs.  $2^{25}$  gange flere end i 1965.
- **Parallelisme** i mange former har været nøglen til at forbedre regnekraften, og parallelle numeriske algoritmer og programmer er blevet udviklet.
- **Supercomputere** blev frem til ca. 1990 fremstillet helt specielt og i kun få, meget dyre eksemplarer. Siden ca. 2000 har masseproducerede mikrochips helt overtaget også dette område af teknologien.
- De største **supercomputere** indeholder i dag op til **millioner af CPUer**, som alle er forbundet med **hurtigt netværk**. Regnekraft på mere end  $10^{16}$  operationer pr. sekund er nu muligt.

# Supercomputere i 2020?

- **Målsætning:** 1 ExaFLOPS = 1.000.000 TeraFLOPS =  $10^{18}$  operationer/sek.
- Hvordan vil en sådan supercomputer være bygget i 2020?  
Eller snarere i 2023...?
- En ExaFLOPS supercomputer vil skulle indeholde millioner af compute-noder, som hver består af tusinder af CPU-kerner.

# Udfordringer til Exa-scale

- Strømforbruget bliver den primære design-parameter!  
De største supercomputere skal holdes under ca. 20 MW af praktiske og økonomiske grunde.  
Dette vil kræve en 100-fold reduktion af det nuværende energiforbrug pr. CPU!
- Fokus på **optimering af regneoperationer** (FLOPs) erstattes af **fokus på data-lokalitet**.  
Flytning af data koster 100-1000 gange så meget energi som regne-FLOPs!
- Strømbesparelser i CPUer betyder, at identiske CPU-kerner kører ved forskellige hastigheder.
- **Hybride** CPU/GPU/Accelerator compute-noder bliver nødvendige.
- Der kræves helt nye måder at designe matematiske algoritmer, kommunikationsbiblioteker og programmer på, når man skal holde styr på millioner til milliarder af heterogene CPU-kerner.
- Tolerance over for **fejlende compute-noder** skal bygges ind i designet .
- **Konklusion:** Hardware og software skal sam-designes!