

Sistemas HPC e Construção de Chips

CES-25 – Arquiteturas para Alto Desempenho

Prof. Paulo André Castro

pauloac@ita.br

Sala 110 – Prédio da Computação

www.comp.ita.br/~pauloac

IEC - ITA

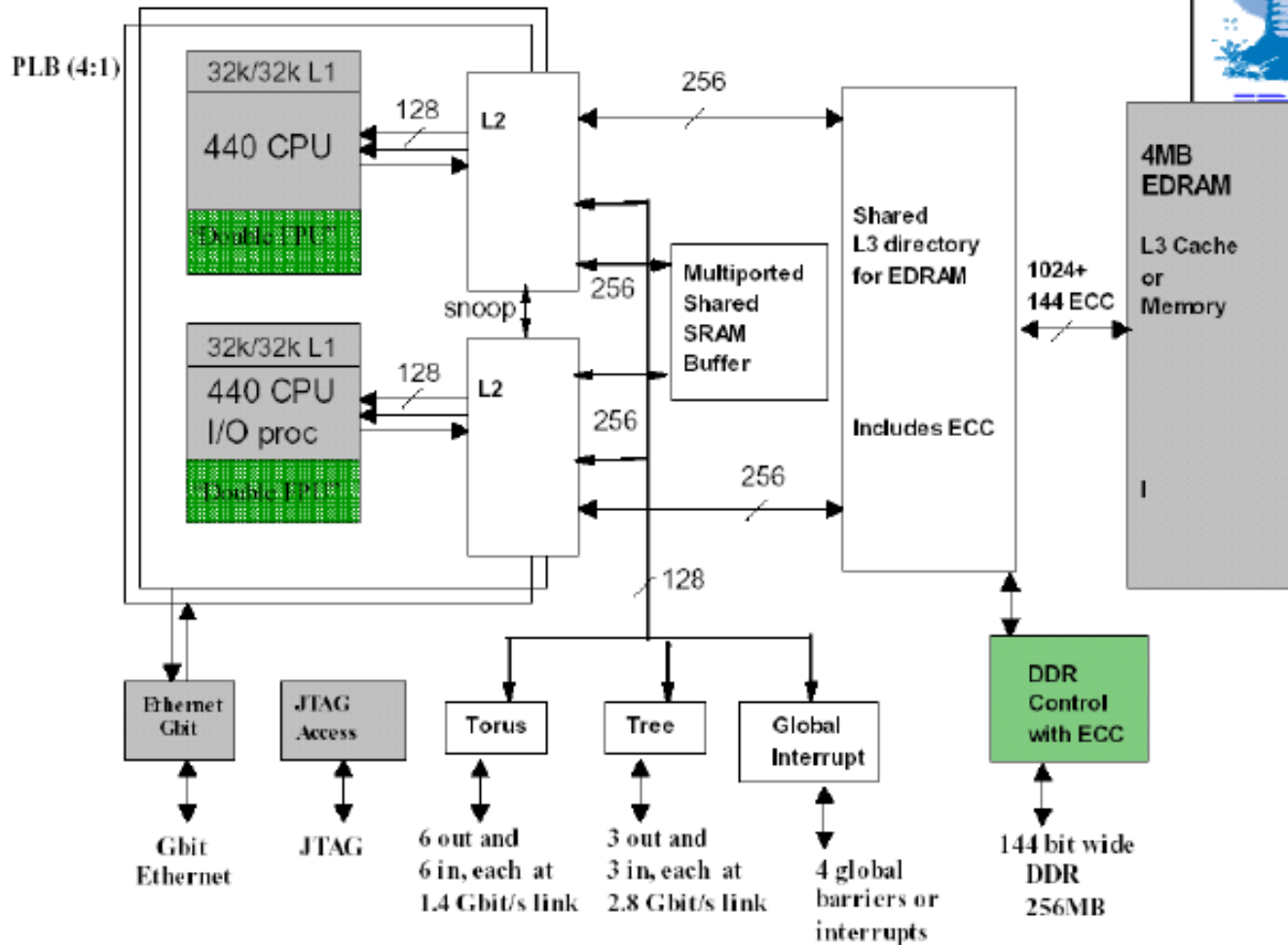
Aplicações de Sistemas HPC

- Sistemas de Computação de Alto Desempenho (HPC) são muitas vezes utilizados para realizar simulações numéricas de processos físicos. Como por exemplo:
 - Simulação de efeitos de explosões nucleares
 - Simulação da evolução de marés e movimento dos oceanos
 - Simulação atmosférica/metereológica
 - escoamento de fluídos
 - Mapeamento de Sub-solo através de cálculo de reflexão de ondas. etc.
- Outra área que demanda alto poder computacional são servidores com **grande** volume de transações processadas por unidade de tempo

Sistemas Computacionais de Alto Desempenho (HPC)

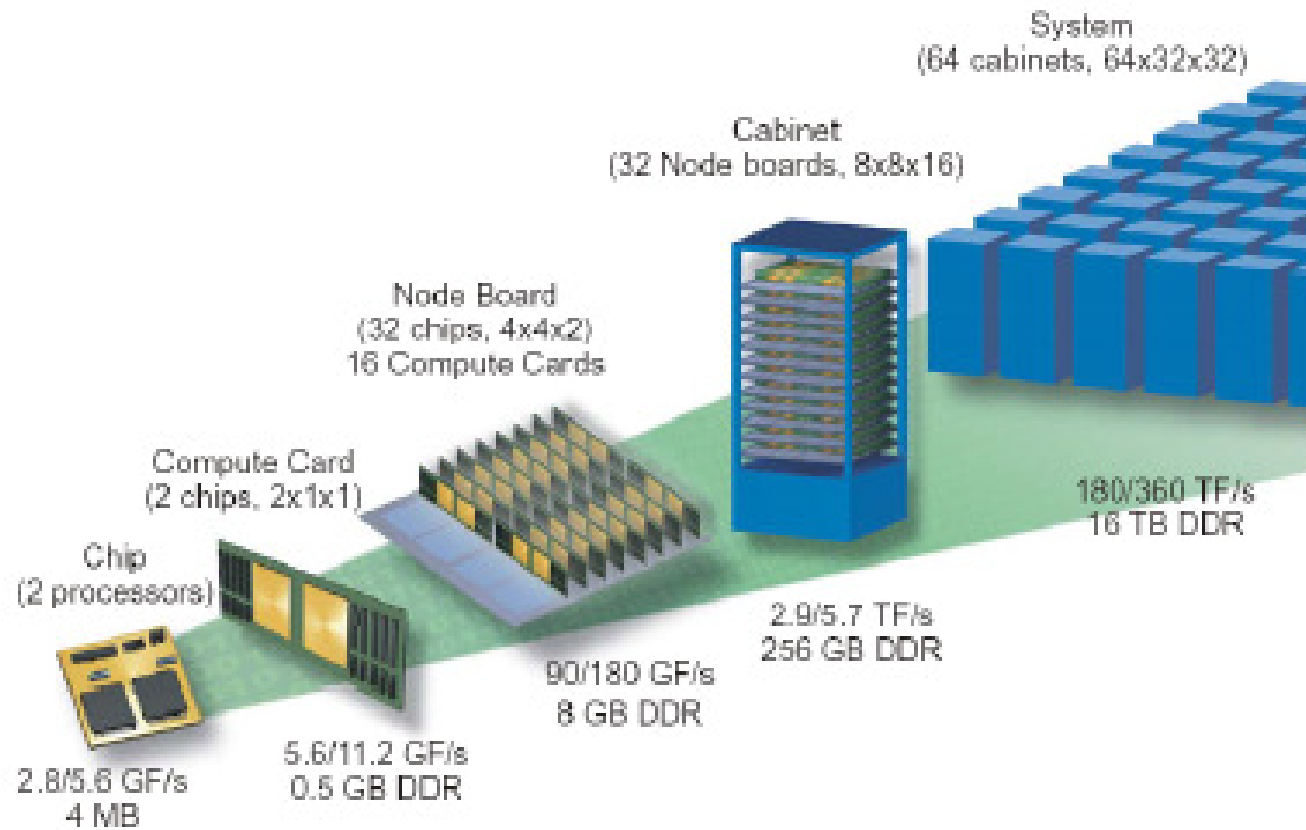
- **Computadores Massivamente paralelos** são computadores especialmente projetados para trabalhar com centenas ou até milhares de processadores e que utilizam mecanismos de memória distribuída e geralmente são utilizados em **simulações matemáticas**. Exs.:
 - Projeto **Blue Gene** (IBM)
- **Clusters** são vários computadores conectados de modo a funcionar como um único sistema, e podem atingir alto poder de processamento. Em geral, utilizam middleware (software) específico para que os computadores troquem informações de modo eficiente: MPI (Message-Passing Interface), PVM (Parallel Virtual Machine), etc. São frequentemente utilizados para atender grande **volume de transações**. Ex.:
 - **Cluster Google**

BlueGene/L Compute ASIC





BlueGene/L



Números do BlueGene/L

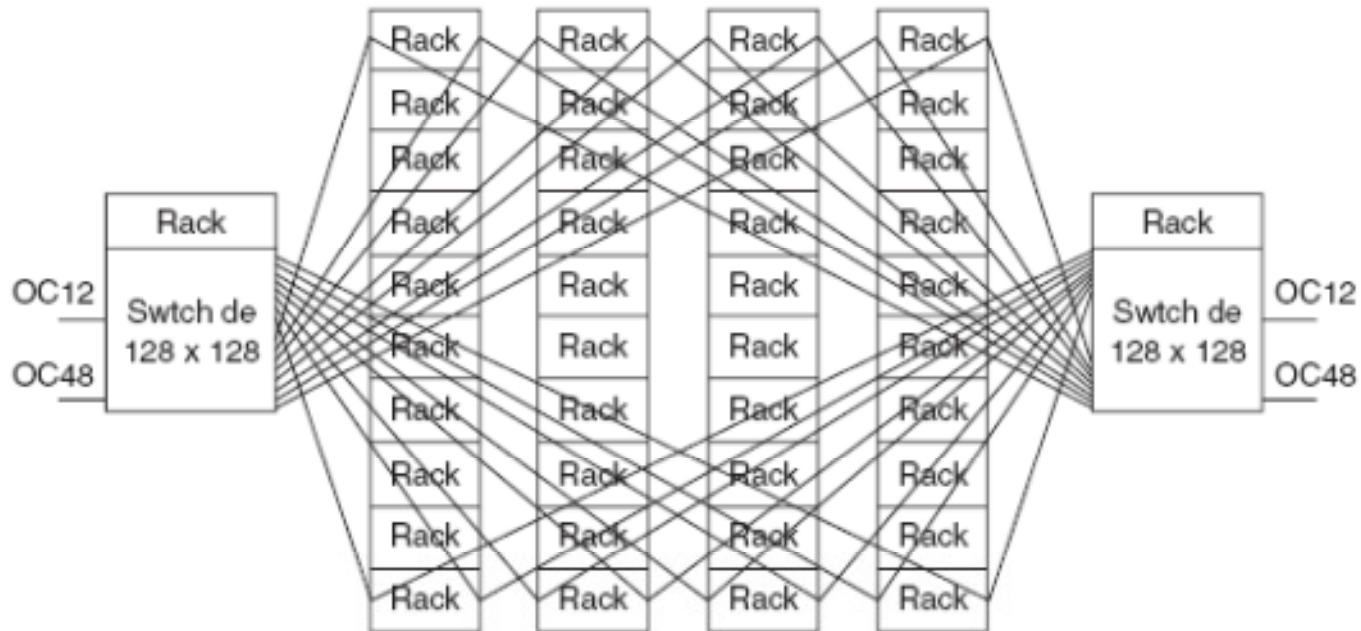
- Cada Gabinete
 - 2048 processadores
 - 128GB de memória SDRAM DDR
 - 512 MB por Computer Card (extensível até 2GB)
- Sistema
 - 128K processadores
 - 16TB de memória SDRAM DDR
- Poder de processamento
 - 180 a 360 Teraflops
- Alto consumo de energia
 - Considerando 50 W por processador, 102 KW
 - Considerando 10h/dia, 20 dias/mês, 20MWh / mês

Cluster Google

- 4 sites
 - Dois no Vale do Silício
 - Dois na Virgínia
- O tempo de vôo para mensagens por meio dos Estados Unidos é de aproximadamente 0,1 segundo;
- então, é importante para a Europa ser servida por sites da Virgínia e para a Califórnia ser servida por sites do Vale do Silício.
- Para tentar atingir a meta de 0,5 segundo de latência, o software do Google normalmente estima de onde a busca se origina a fim de reduzir os atrasos de tempo de vôo

Fonte: (Hennessy, Patterson, 2003)

Cluster Google



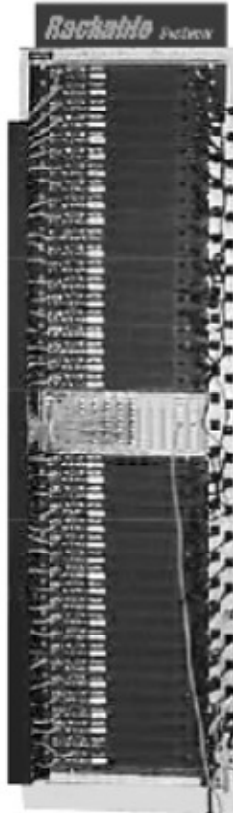
40 racks / máximo de 64 racks

Linhas

OC12:622 Mbits/s

OC48: 2.4 Gbit/s

Rack

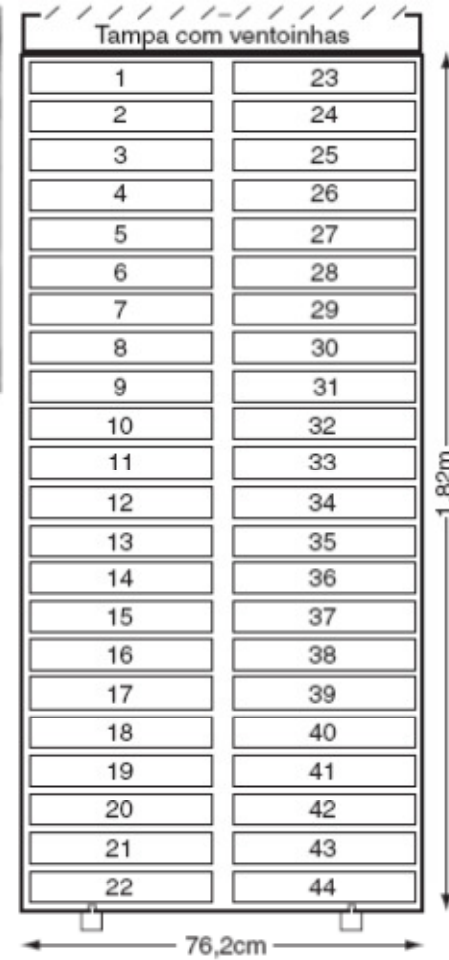


Visão frontal (também visão de trás)

48,26cm



Visão em close dos PCs de 1 RU



Rack Google

- 80 processadores por Rack e 2 switches
 - Parte frontal e traseira
 - Cada parte, 1 switch, 20 PCs acima e 20 PCs abaixo
- Rack Unit (RU)
 - Essa unidade contém duas unidades IDE de 5.400RPM, 256MB de SDRAM de 100MHz, uma placa-
 - Motherboard, uma única fonte de alimentação e um microprocessador Intel.
 - Cada PC executa versões 2.2.16 ou 2.2.17 do kernel Linux em uma distribuição Red Hat ligeiramente modificada

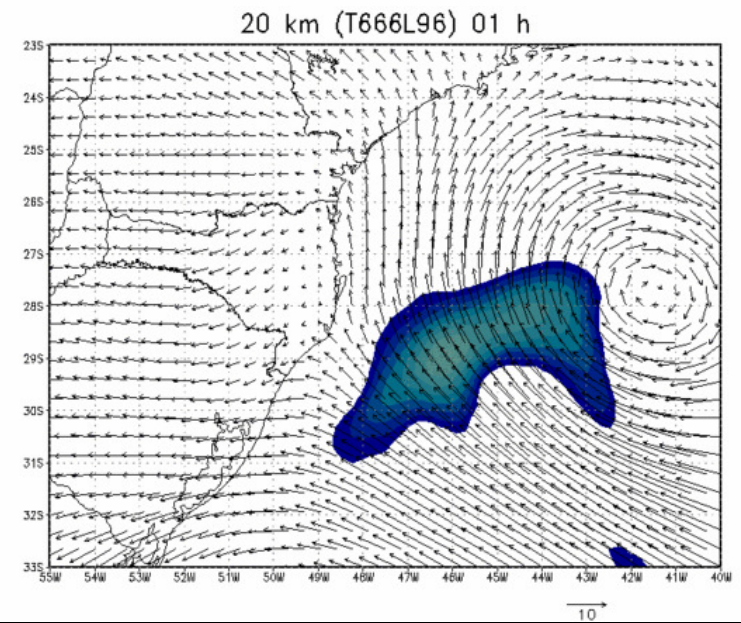
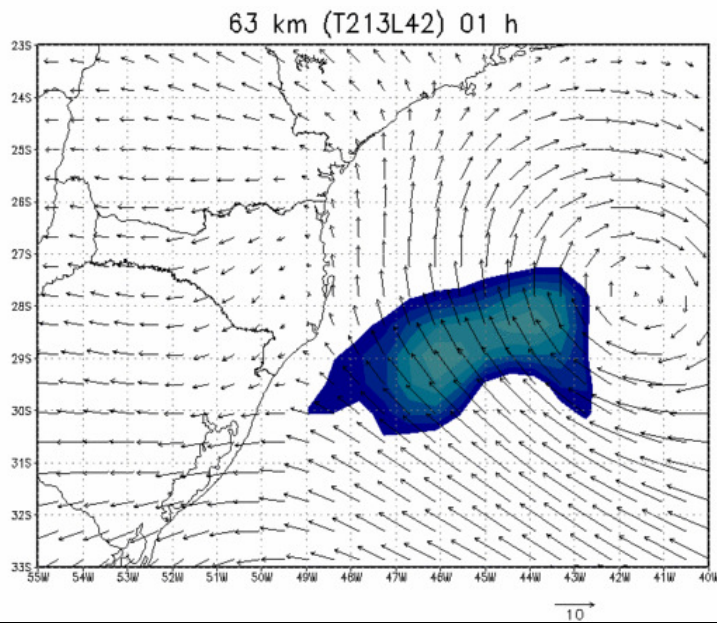
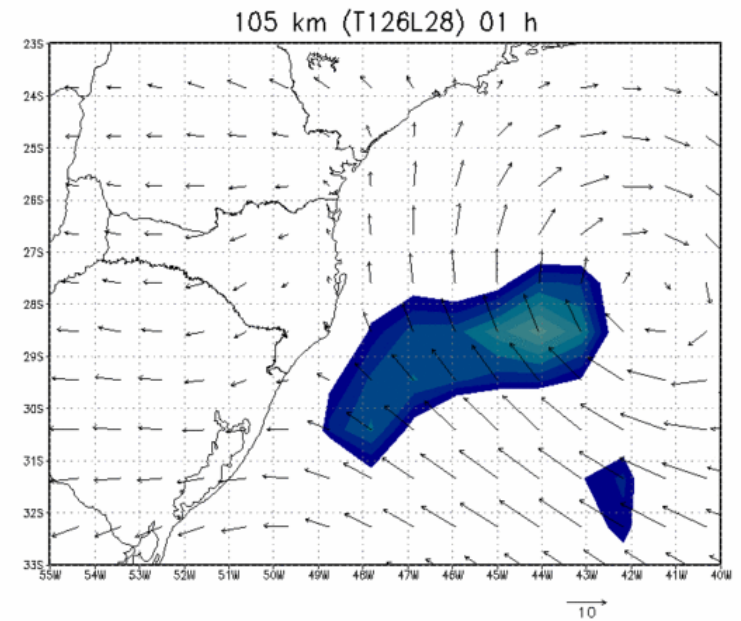
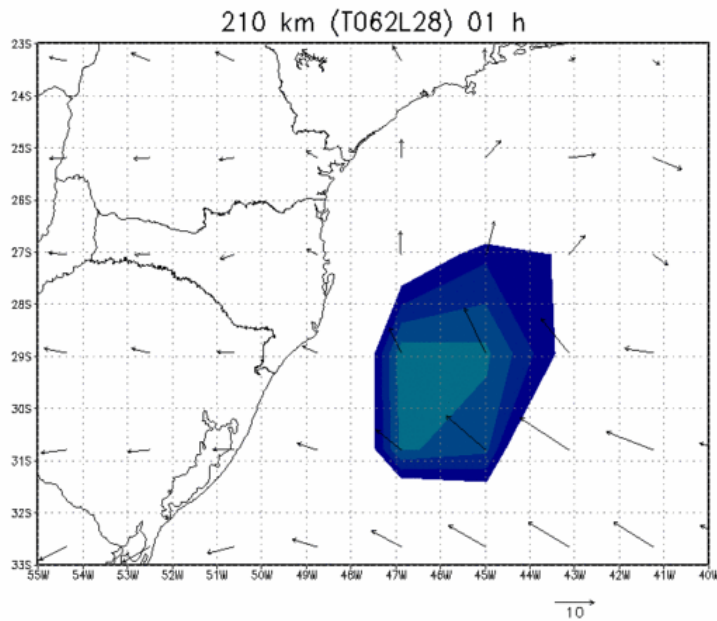
Processamento de Alto Desempenho (PAD) no Brasil

- Alguns Exemplos Brasileiros: (cortesia do Prof. Jairo Panneta/ CPTEC e Petrobrás)
 - CPTEC /INPE: Previsão Meteorológica
 - Petrobrás: Estudos geológicos

PAD - CPTEC

- **Previsão Numérica Operacional de Tempo e Clima**
- **Domínio tecnológico da aplicação**
- **Busca incessantemente melhorar a qualidade e a acuidade da previsão**
 - **Computadores cada vez mais potentes**
 - **Paralelismo é essencial**

PAD - CPTEC



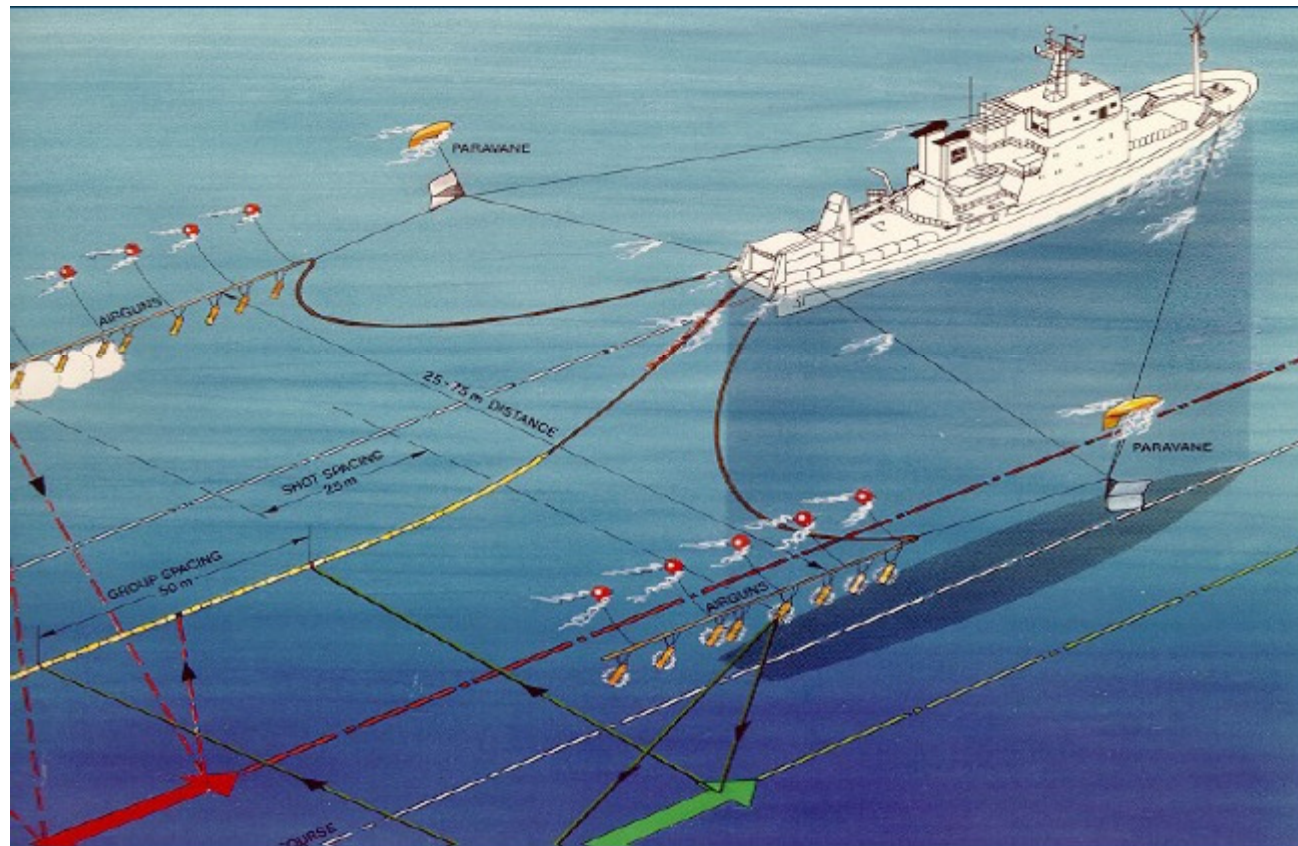
PAD - CPTEC

	1994	1998	2004	2007
Máquina	NEC SX3	NEC SX4	NEC SX6	NEC SUN
Processadores	1	8	96	1100
Vel. Máxima (GFlops)	3	16	768	5748

PAD - Petrobrás

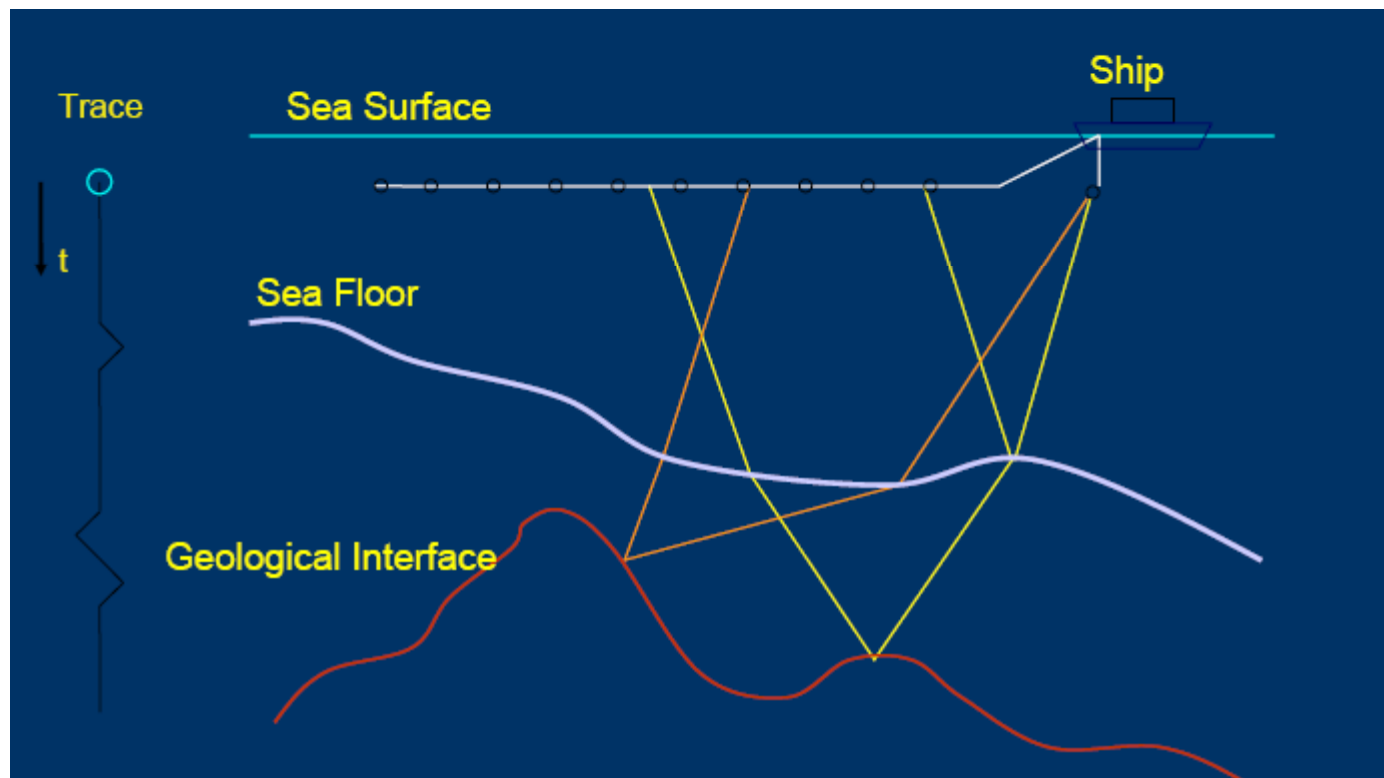
- **Processamento Sísmico posiciona poços**
- **Domínio tecnológico da aplicação**
 - **Migração sísmica desenvolvida na casa**
- **Busca incessantemente melhorar a qualidade e a acuidade da imagem**
 - **Computadores cada vez mais potentes**
 - **Paralelismo é essencial**

Petrobrás – Método Sísmico



Petrobrás – Método Sísmico

- Método de Migração de Kirchhoff



PAD – Pesquisa Petrobrás 1998

76 PCs

Desenvolvimento de Software



Cluster Petrobrás - 2006

- 5.000 núcleos de CPU



Cluster Petrobrás - 2006

- 5.000 núcleos de CPU
- 1000 jobs de Migração de Kirchhoff

Jobs mais caros	Núcleos / job (média)	Dias / job (média)
10	787	22
100	656	10
1000	185	2

PAD – Produção Petrobrás 2008

- **14.000 processadores**
- **3.000 Tarefas (jobs)**



Conclusões

- Sistemas Computacionais de Alto Desempenho tem diversas aplicações em vários problemas do mundo real
- Técnicas estudadas em Arquiteturas de Computadores podem ser usadas para otimizar o desempenho também em sistemas de software



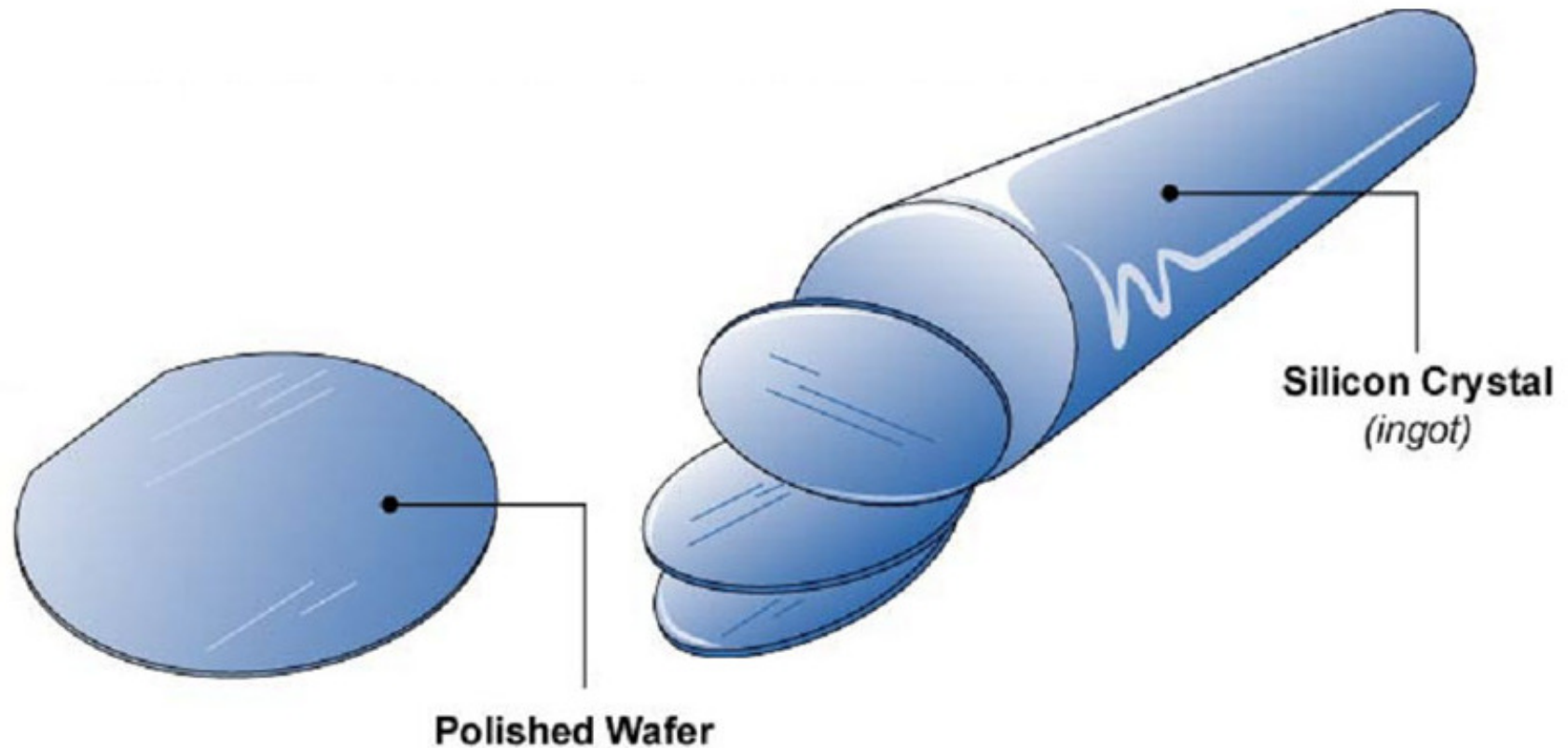
O chip do processador é fundamental para o bom **Desempenho** de qualquer computador.

Mas de onde vem os chips??

Silicon

- The chips are built on Wafers. Raw wafers are made of silicon, which comes from beach sand.
- They are created through a method called Czochralski process, where a seed crystal (a piece of silicon crystal) is mounted on a rod and then dipped into molten silicon.
- The rod is pulled upwards and rotated at the same time, making a big cylindrical piece of silicon crystal, also known as ingot.

Cilindros de Silício são cortados em Wafers

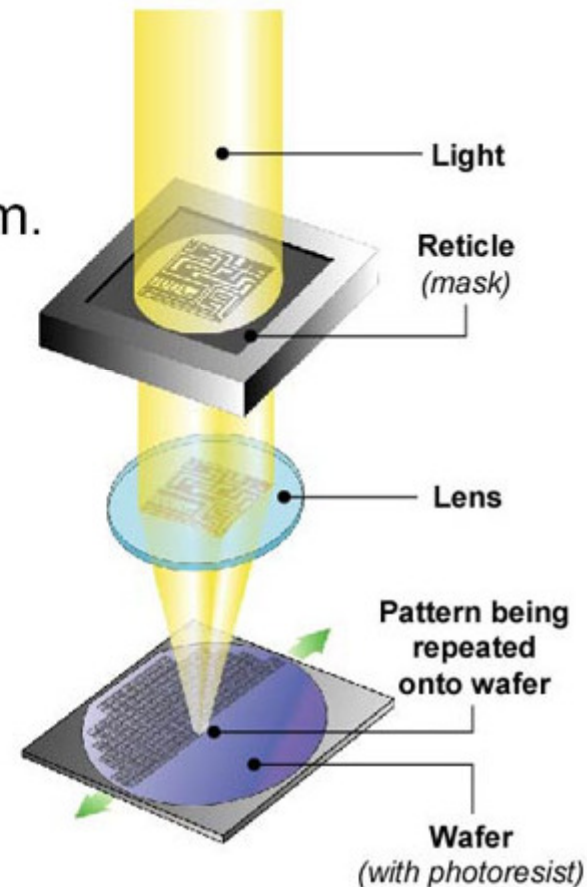


Wafer

- The ingot resulted from this process measures from three to six feet (one to two meters) long and can have up to 12 inches (300 mm) in diameter (this is where terms like 12-inch or 300-mm wafers come from).
- The ingot is then sliced into wafers. These wafers are polished and sent to the chip manufacturers. The wafers usually are less than 1mm thick
- These raw (“virgin”) wafers are where the chips will be manufactured on.

Photolithography (fotolitografia)

- We grow layers one on top of the other.
 - With an insulator between them.
- We use masks to create the shapes.



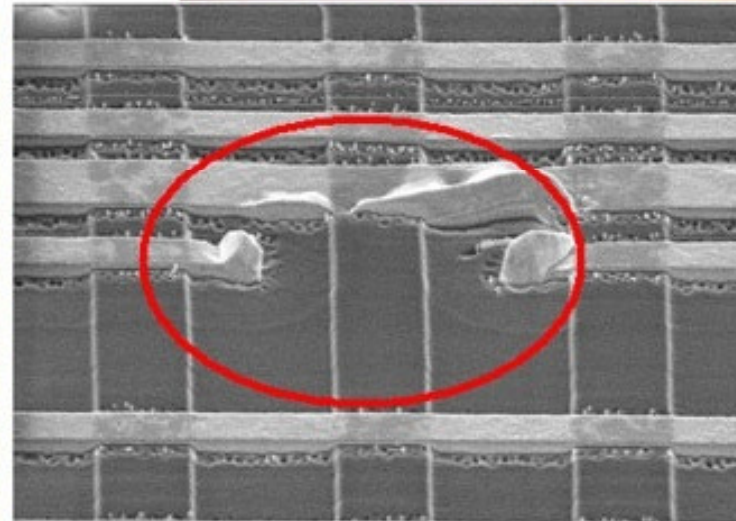
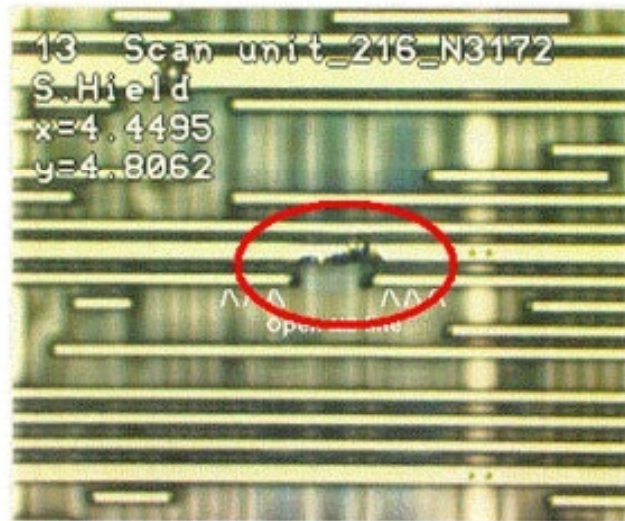
Several Layers and Masks

- Chips are mounted on the wafer through a process called **photolithography**.
- Under this process, chemicals sensitive to ultraviolet light are used. When exposed to ultraviolet light, they can become “soft” or “hard”.
- Basically this process consists in blocking the ultraviolet light from the chemicals applied to the wafer using masks created by the designers, removing the “soft” parts and then repeating the process again with another mask, until the chip is finished.

Several Layers and Masks - 2

- Of course each mask has a different pattern and they tell how the transistors and wires inside the chip will be manufactured.
- The numbers of masks used vary depending on the project. A Pentium 4 processor, for example, uses 26 masks.
- After this process, the transistors and circuit is built on the wafer.

Any small (even dust) particle
can ruin the silicon...

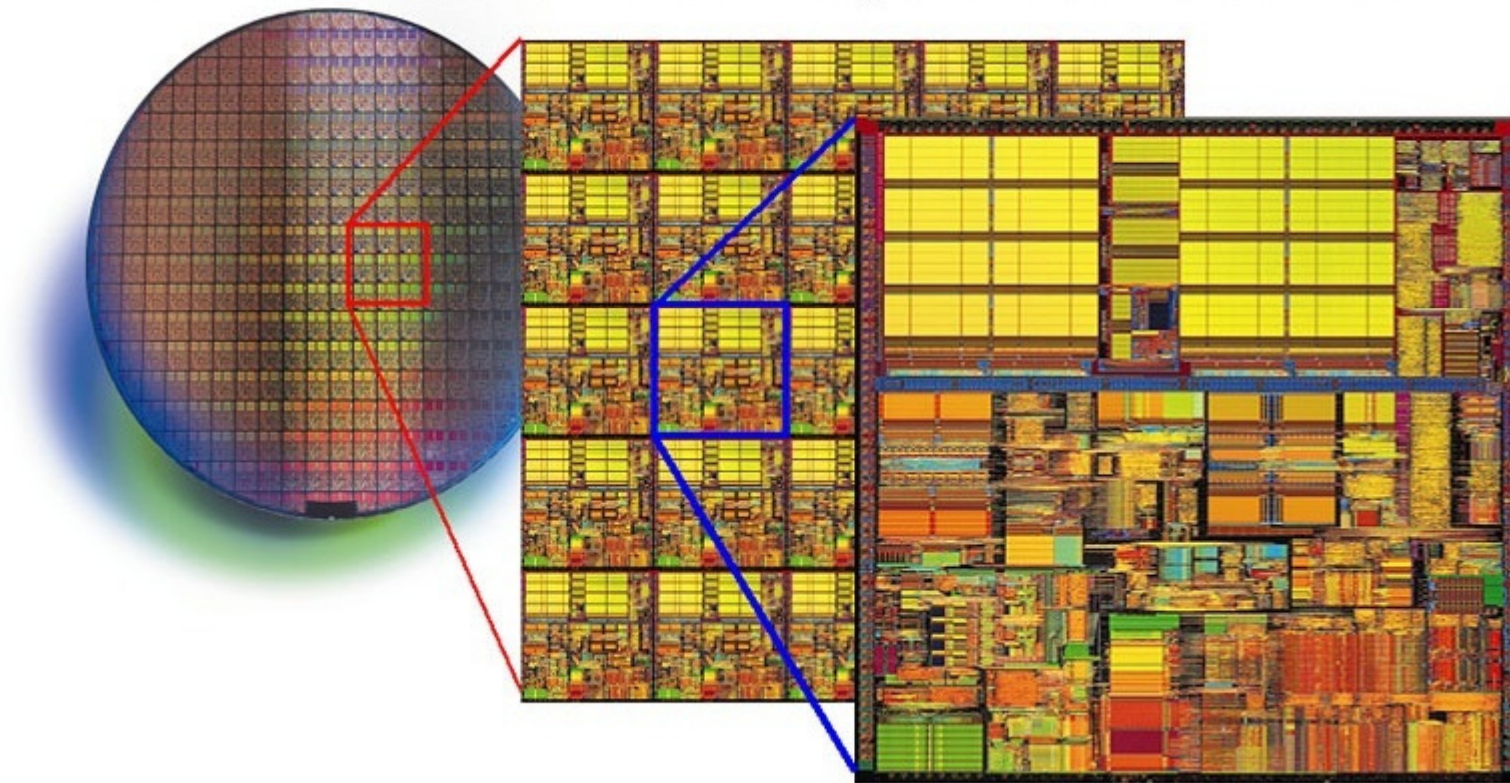


- What about vibrations on the wafer? Or the light emitter ? Masks? Lens?

That is the reason why Chip factories need Clean rooms



From Wafer to Die



- The chips on the wafer are then tested and the wafer is sent to the next step...
- where the chips are cut from the wafer, have their terminals attached and are packed.
- After that they are tested, labeled and sold.



Some Questions..

- How many chips can I built on one wafer?
 - Wafer Area/Die Area minus all the dies on the wafer edge.

$$\text{Dies per wafer} = \frac{\pi \times (\text{Wafer diameter}/2)^2}{\text{Die area}} - \frac{\pi \times \text{Wafer diameter}}{\sqrt{2} \times \text{Die area}}$$

- **Example**

Find the number of dies per 300 mm (30 cm) wafer for a die that is 1.5 cm on a side.

The die area is 2.25 cm². Thus

$$\text{Dies per wafer} = \frac{\pi \times (30/2)^2}{2.25} - \frac{\pi \times 30}{\sqrt{2} \times 2.25} = \frac{706.9}{2.25} - \frac{94.2}{2.12} = 270$$

Some Questions...

- How many of those chips will really work ?
 - It doesn't matter how good is the manufacturing process, some dies will not work.
 - More masks and layers mean more errors
 - A formula was developed empirically by looking the yield of many manufacturing lines.

$$\text{Die yield} = \text{Wafer yield} \times \left(1 + \frac{\text{Defects per unit area} \times \text{Die area}}{\alpha} \right)^{-\alpha}$$

Example assuming Wafer yield equals to 100%

Find the die yield for dies that are 1.5 cm on a side and 1.0 cm on a side, assuming a defect density of 0.4 per cm² and α is 4.

The total die areas are 2.25 cm² and 1.00 cm². For the larger die, the yield is

$$\text{Die yield} = \left(1 + \frac{0.4 \times 2.25}{4.0}\right)^{-4} = 0.44$$

For the smaller die, it is $\text{Die yield} = \left(1 + \frac{0.4 \times 1.00}{4.0}\right)^{-4} = 0.68$

That is, less than half of all the large die are good but more than two-thirds of the small die are good.

- Many 32 and 64 bits microprocessors in a 90nm technology fall between these two sizes.
- For the 2.25cm² die in a 300mm wafer, it means that only 120 (118.8) from 270 possible die will really work.

**Foto de um
Wafer com 117
chips AMD
Opteron
implementados
com tecnologia
90nm**

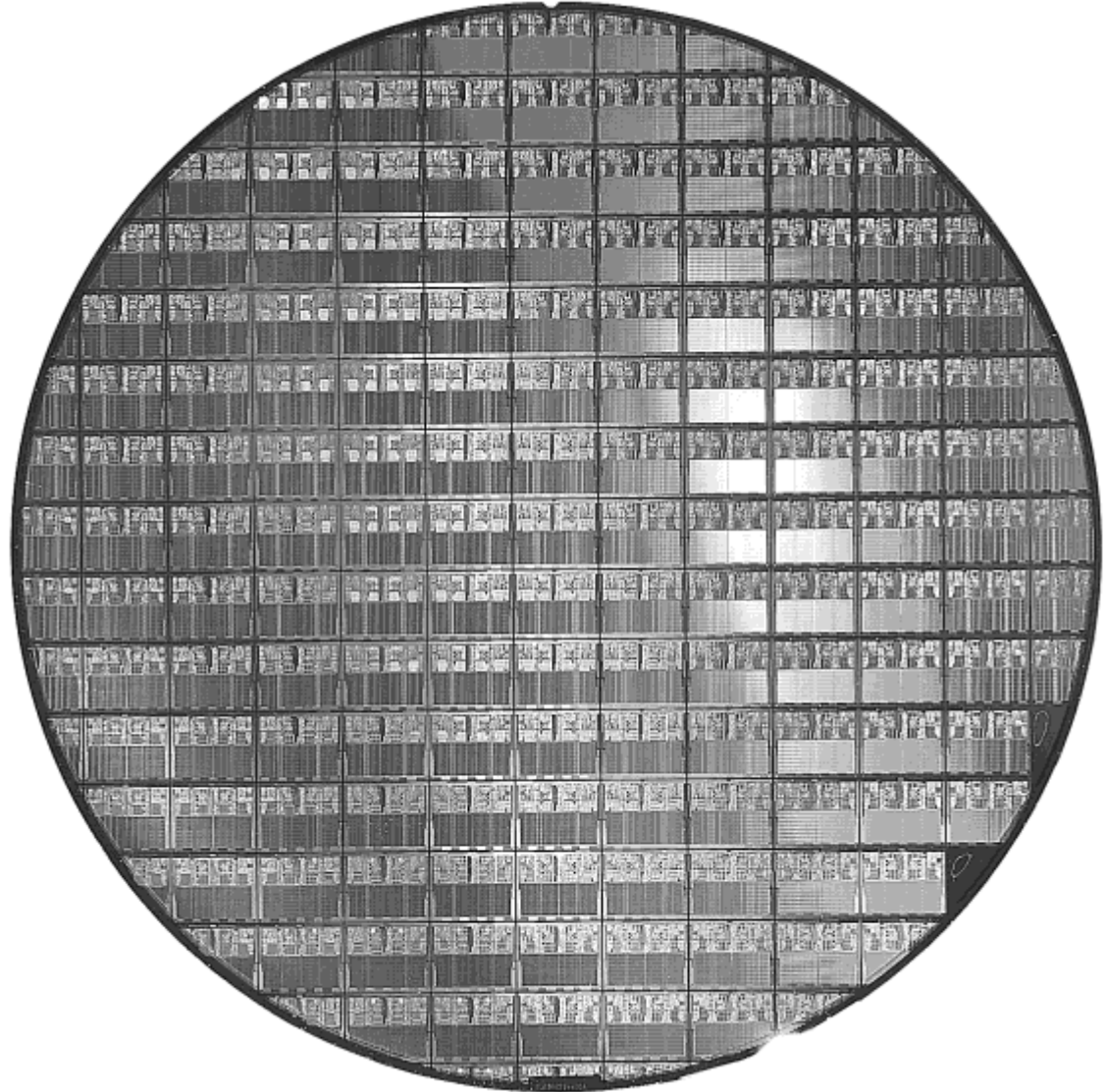
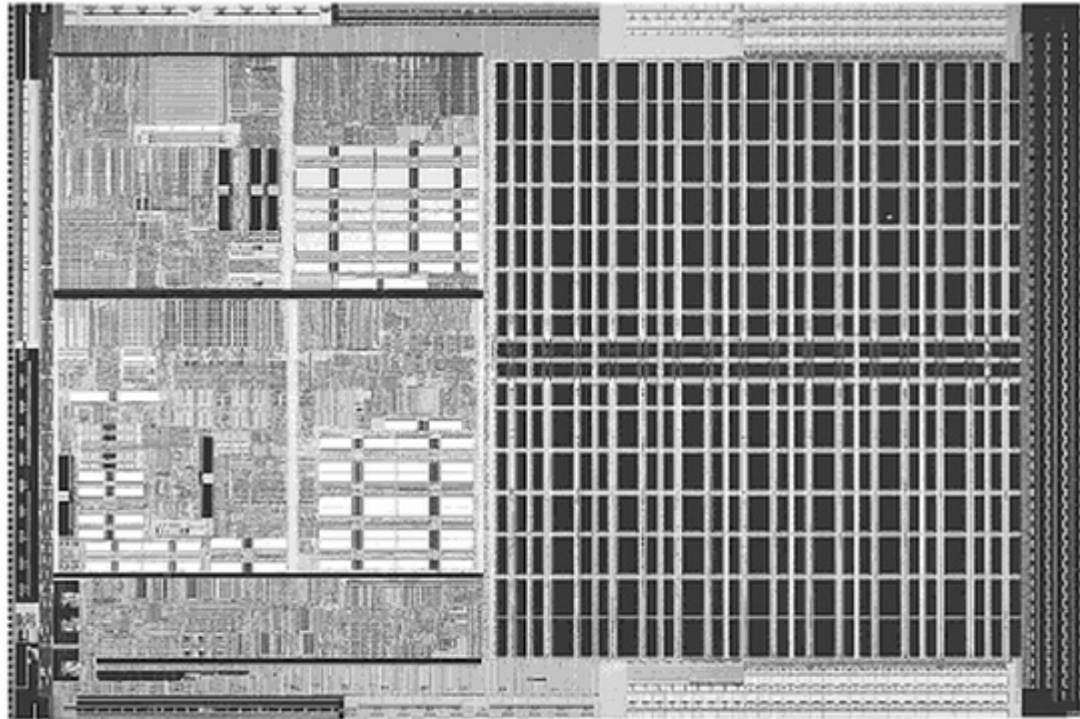


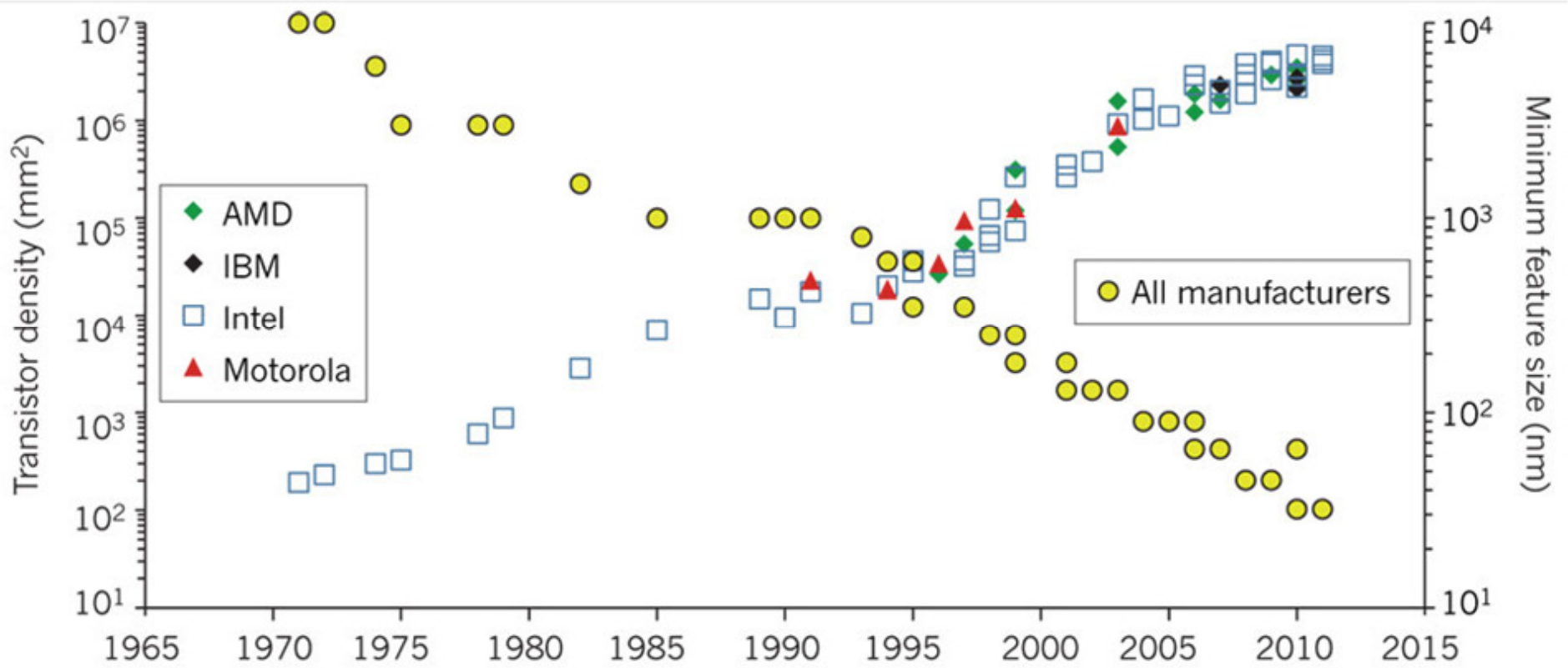
Foto de um die (matriz ou molde) AMD Opteron



Dados de alguns processadores Fonte: (Hennesy, Patterson, 2007)

Chip	Die size (cm ²)	Estimated defect rate (per cm ²)	Feature size(nm)	Transistors (millions)
IBM Power 5	3.89	0.3	130	276
Sun Niagara	3.80	0.75	90	279
AMD Opteron	1.99	.75	90	233

A evolução da densidade de transistores por area e da “Tecnologia” de construção(Transistor gate length or feature size)



Obs.: Dados de densidade apresentados por fabricante

Dados de feature size apresentados para todos os fabricantes de forma unificada (All Manufacturers)

Construção de Chips no Brasil

- CEITEC: Empresa pública federal vinculada ao Ministério da Ciência, Tecnologia e Inovação (MCTI) focada no desenvolvimento e produção de circuitos integrados (CIs) para RFID (identificação por radiofrequência) e aplicações específicas. Localizada em Porto Alegre/RS. Foi fundada em 2008.
 - Assinou em 2011 contrato de transferência de tecnologia 600nm CMOS com empresa alemã X-Fab
 - Primeiras vendas de Chips em 2012. Chips de RFID para identificação de bovinos. O produto é conhecido como Chip do Boi
- Unitec Semicondutores (antiga Six) foi incorporada pelo grupo argentino Corporación América. A unidade industrial será instalada em Ribeirão das Neves (MG). Foi fundada em 2012 e terá investimento da ordem de 1Bi R\$.
 - The plant will manufacture custom chips for niche markets, including industrial and medical applications, generating 300 direct jobs.
 - A planta irá trabalhar com tecnologia de 100nm licenciada da IBM