UFRR

UNIVERSIDADE FEDERAL DE RORAIMA

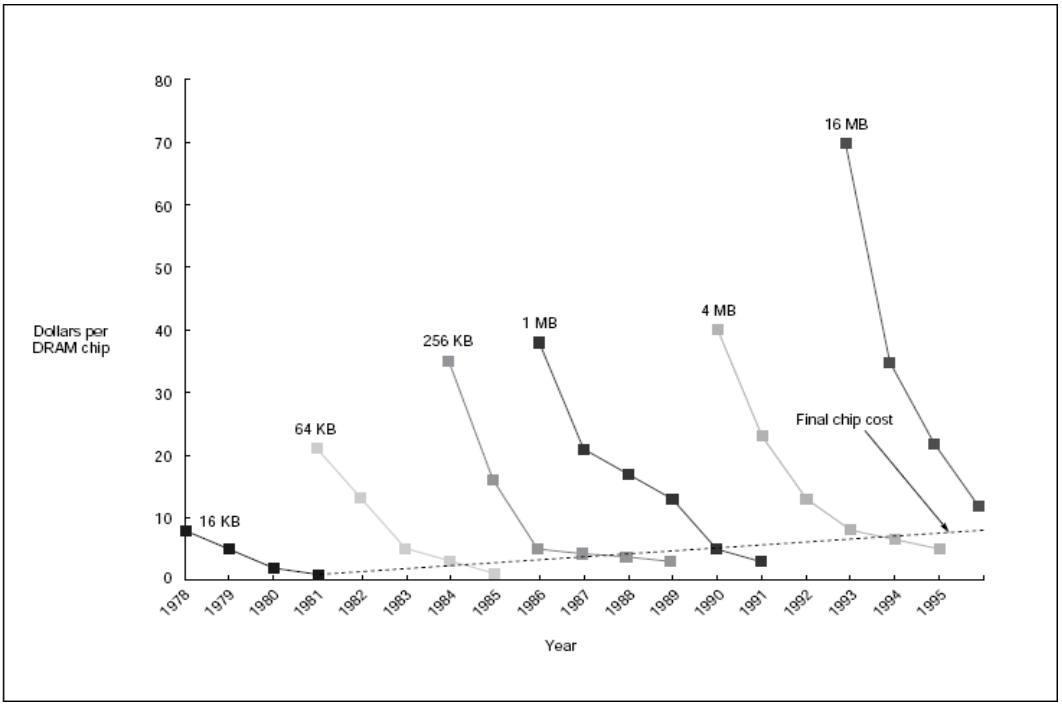
Custos

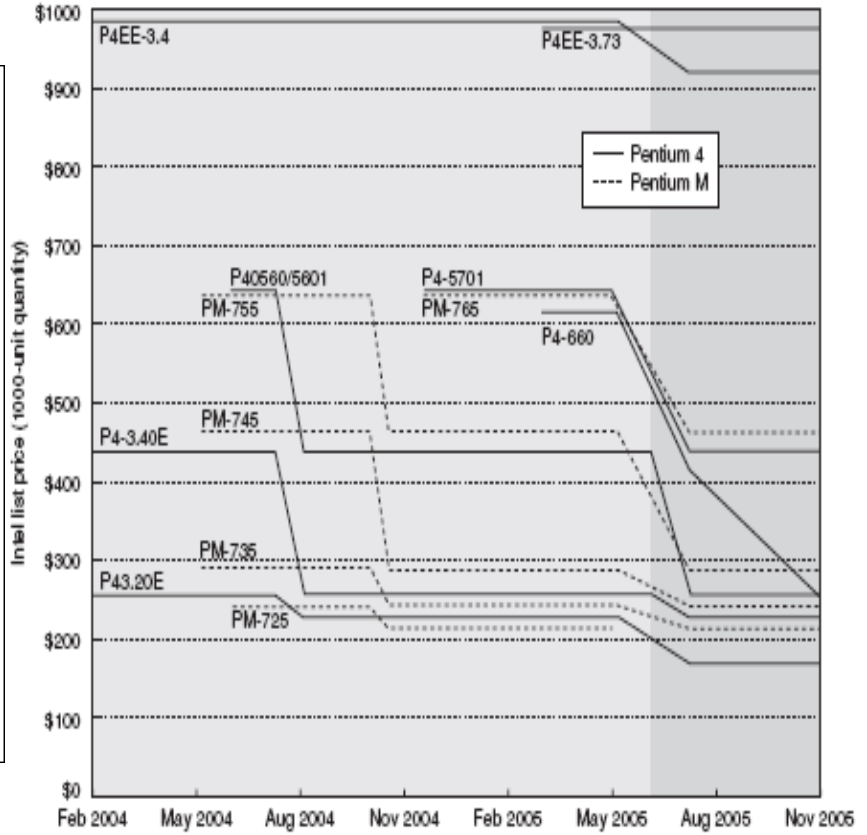
Custos

- Vários aspectos do projeto de um computador afetam em seu custo
- Entender esses fatores permitem a tomada de decisões melhores

- Fator TEMPO
  - Custos de manufatura tendem a diminuir no tempo
    - Aprendizado, melhorias de processo
  - Ex. Custo do MByte DRAM cai 40% por ano
  - Preços de processadores também caem, mas não tão uniformemente

Custos


Dollars per DRAM chip
80
70
60
50
40
30
20
10
0
16 KB
64 KB
256 KB
1 MB
4 MB
16 MB
Final chip cost
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
Year


Intel list price (1000-unit quantity)
$1000
$900
$800
$700
$600
$500
$400
$300
$200
$100
$0
P4EE-3.4
P4EE-3.73
Pentium 4
Pentium M
P4-5701
P4-660
P40560/5601
PM-755
PM-765
PM-745
P4-3.40E
PM-735
P43.20E
PM-725
Feb 2004
May 2004
Aug 2004
Nov 2004
Feb 2005
May 2005
Aug 2005
Nov 2005

Custos

- Fator Volume
  - Maior volume de vendas causa diminuição de preços
  - Amortização dos custos de projeto e testes pelo mercado de massa

- Fator Comódites
  - Mais pontos de venda e mais fabricantes tendem a causar queda de preços
  - Concorrência força eficiência no processo

Custo de Circuitos Integrados

- Boa parte dos custos de um computador se referem aos circuitos integrados
- Exemplo de distribuição de custos

System	Subsystem	Fraction of total
Cabinet	Sheet metal, plastic	1%
	Power supply, fans	2%
	Cables, nuts, bolts	1%
	Shipping box, manuals	0%
	Subtotal	**4%**
Processor board	Processor	6%
	DRAM (64 MB)	36%
	Video system	14%
	I/O system	3%
	Printed circuit board	1%
	Subtotal	**60%**
I/O devices	Keyboard and mouse	1%
	Monitor	22%
	Hard disk (1 GB)	7%
	DAT drive	6%
	Subtotal	**36%**

Custos

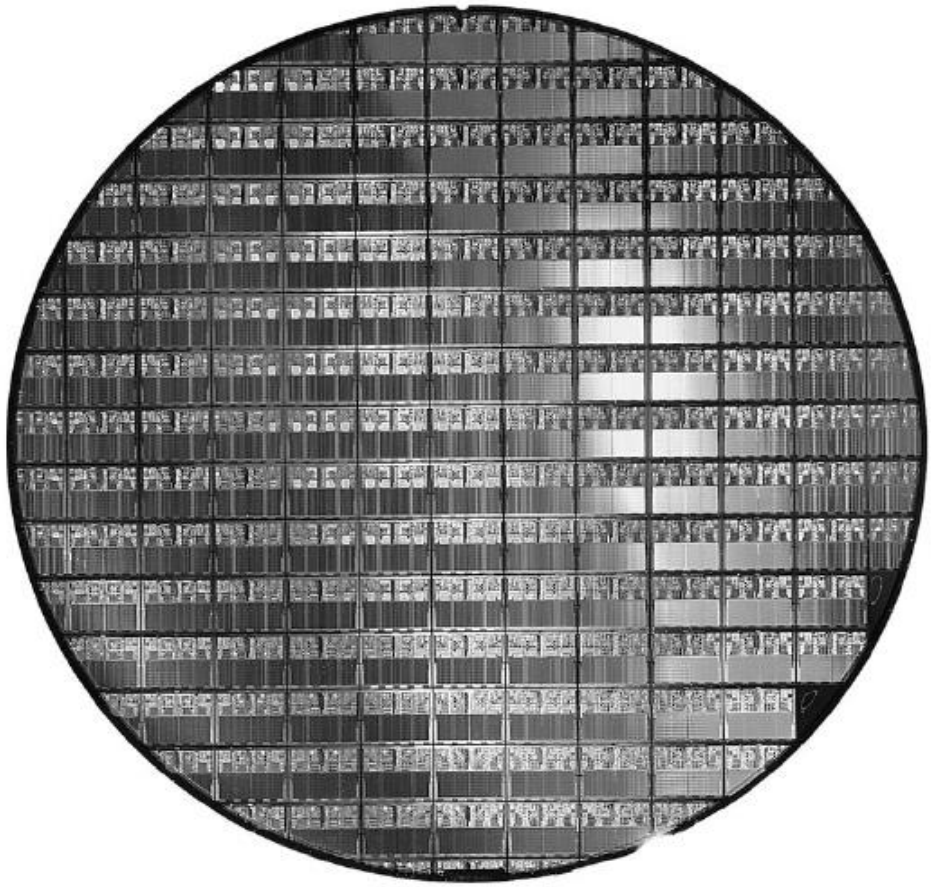
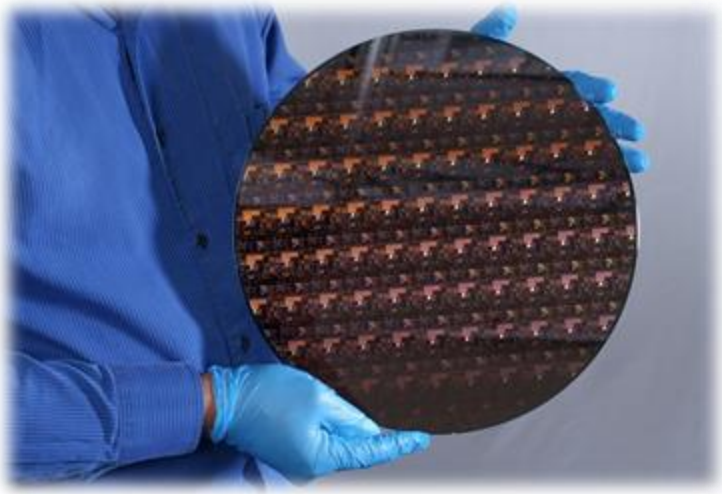
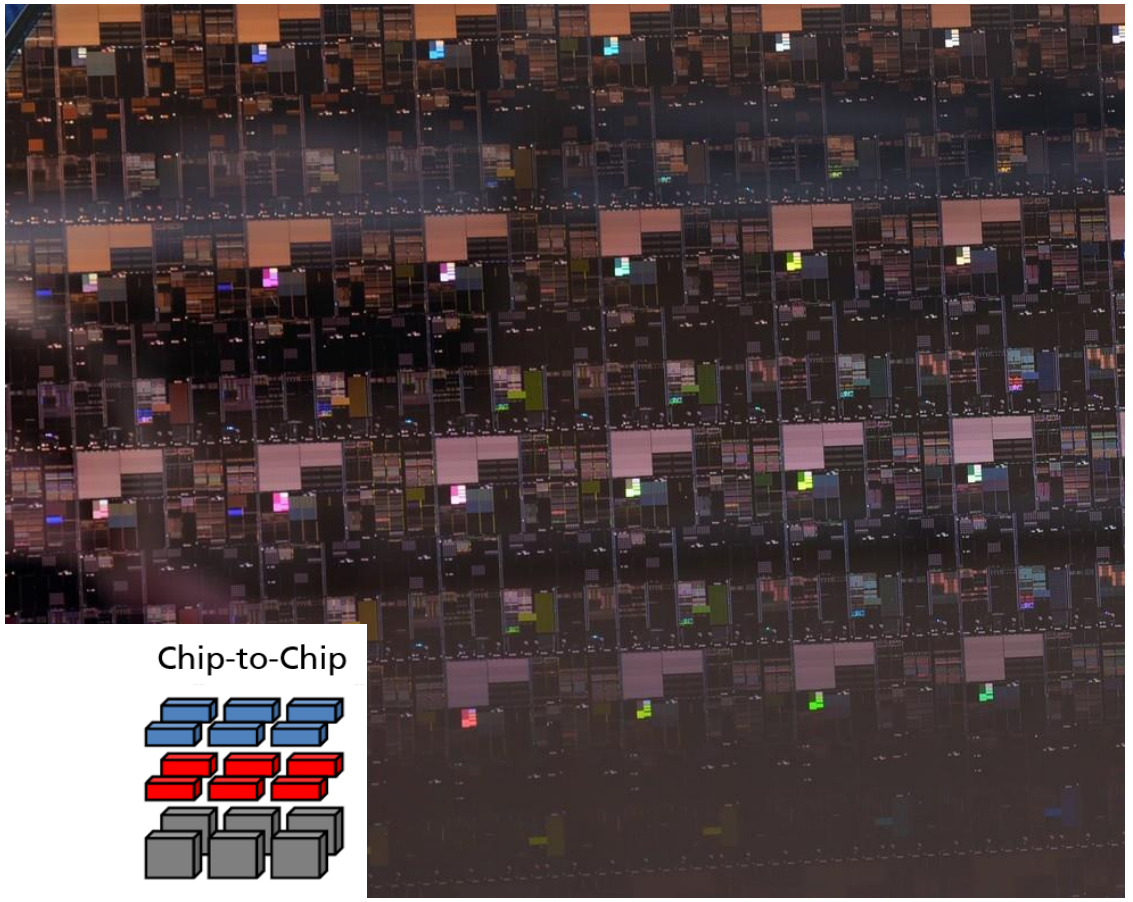
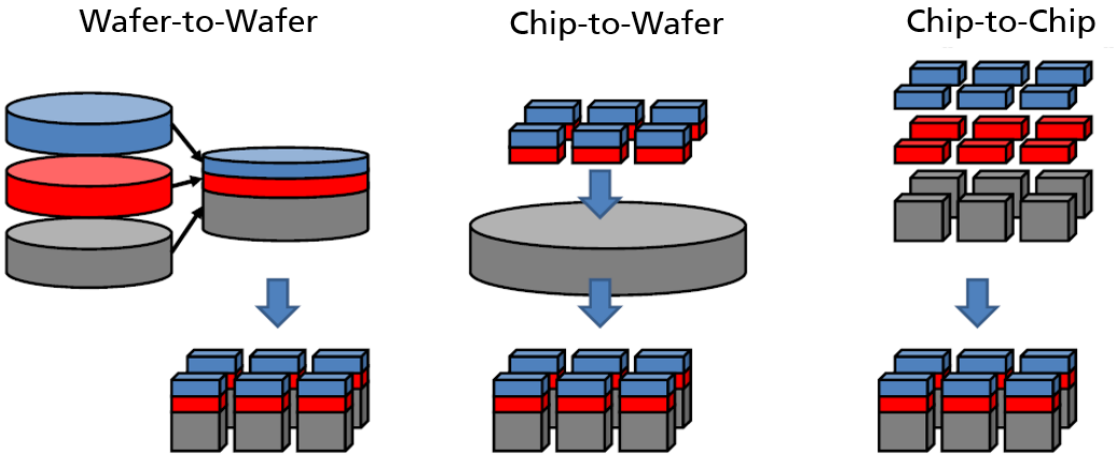


Figure 1.12 This 300mm wafer contains 117 AMD Opteron chips implemented in a 90 nm process. (Courtesy AMD.)

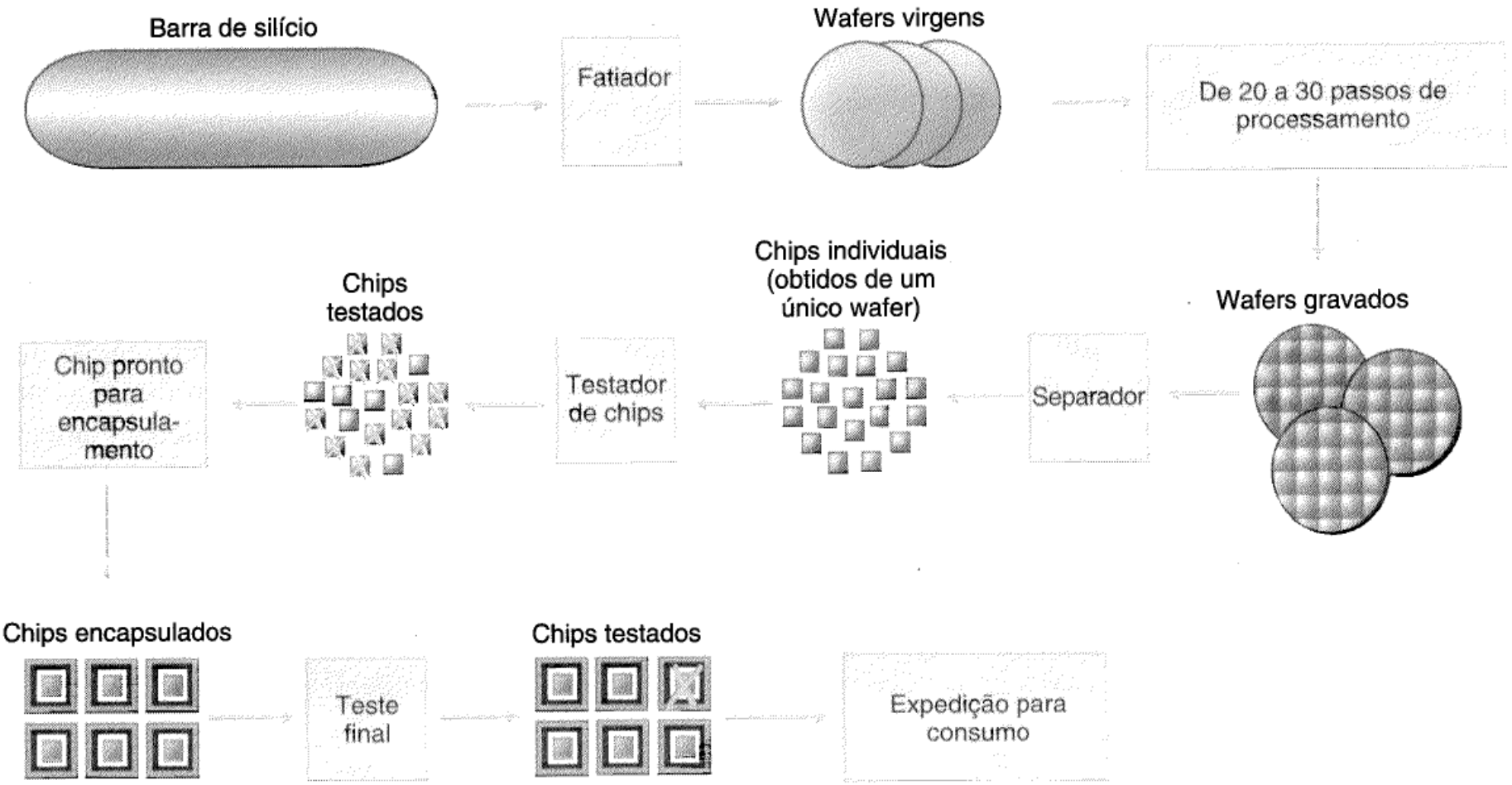
Custos







Processo de Fabricação de CIs



Fórmulas de custo de CIs

- A prática mostrou que há muitas perdas de CIs por defeitos em wafers.
  - Portanto, rendimento é levado em consideração.
  - Fórmula do rendimento obtida de forma empírica após anos de observação

$$\text{Custo por chip} = \frac{\text{Custo por wafer}}{\text{Chips por wafer} \times \text{rendimento}}$$

$$\text{Chips por Wafer} \approx \frac{\text{Área do wafer}}{\text{Área do chip}}$$

$$\text{Rendimento} = \frac{1}{\left(1 + \left(\text{Defeitos por área} \times \text{Área do chip}/2\right)\right)^2}$$

- Quantidade de chips é um aproximação, pois wafer são redondos e chips são quadrados.

Fórmulas de custo de CIs

- Ex. 1: Qual o rendimento de um Wafer para um chip de lado 1cm e para um chip de lado 2cm, assumindo um defeito por área de 0.5 por cm^2?

- Ex. 2: Qual o custo de cada um desses chips assumindo um wafer de 20 cm de diâmetro a $ 100?

Fórmulas de custo de CIs

- Ex. 1: Qual o rendimento de um Wafer para um chip de lado **1cm** e para um chip de lado **2cm**, assumindo um defeito por área de **0.5** por **cm²**?

$$\text{Rendimento} = \frac{1}{\left(1 + \left(\text{Defeitos por área} \times \text{Área do chip}/2\right)\right)^2}$$

$$R_{C1} = \frac{1}{\left(1 + \left(0{,}5 * 0{,}5\right)\right)^2}$$

$$R_{C2} = \frac{1}{\left(1 + \left(0{,}5 * 2\right)\right)^2}$$

Fórmulas de custo de CIs

- Ex. 1: Qual o rendimento de um Wafer para um chip de lado **1cm** e para um chip de lado **2cm**, assumindo um defeito por área de **0.5** por **cm²**?

$$\text{Rendimento} = \frac{1}{\left(1 + \left(\text{Defeitos por área} \times \text{Área do chip}/2\right)\right)^2}$$

$$R_{C1} = \frac{1}{(1+(0{,}5*0{,}5))^2} = 0{,}64$$

$$R_{C2} = \frac{1}{(1+(0{,}5*2))^2} = 0{,}25$$

Fórmulas de custo de CIs

- Ex. 2: Qual o custo de cada um desses chips assumindo um wafer de **20 cm de diâmetro** a **$ 100**?

$$A = \pi \cdot r^2$$

$$\text{Custo por chip} = \frac{\text{Custo por wafer}}{\text{Chips por wafer} \times \text{rendimento}}$$

$$\text{Chips por Wafer} \approx \frac{\text{Área do wafer}}{\text{Área do chip}}$$

Fórmulas de custo de CIs

- Ex. 2: Qual o custo de cada um desses chips assumindo um wafer de **20 cm de diâmetro** a **$ 100**?

$$CW_{C1} = \frac{314}{1} = 314$$

$$CW_{C2} = \frac{314}{4} = 78{,}5$$

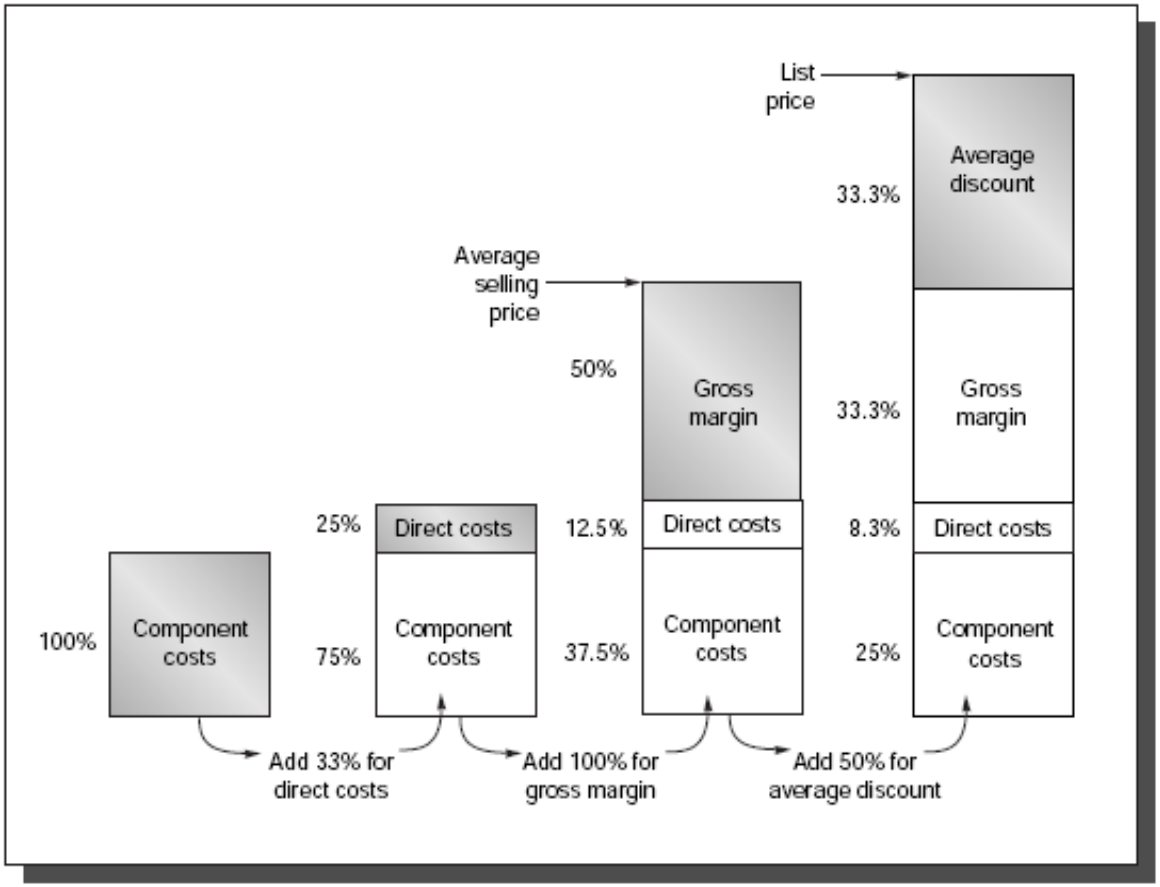
$$\text{Custo por chip} = \frac{\text{Custo por wafer}}{\text{Chips por wafer} \times \text{rendimento}}$$

$$CC_{C1} = \frac{100}{314 * 0{,}64} = 0{,}49$$

$$CC_{C2} = \frac{100}{78{,}5 * 0{,}25} = 5{,}09$$

Formação de preço

- Obviamente, o preço de um produto leva em consideração margens de lucro e outros fatores comerciais.



UNIVERSIDADE FEDERAL DE RORAIMA

Desempenho: Busca Eterna

Evolução de Processadores

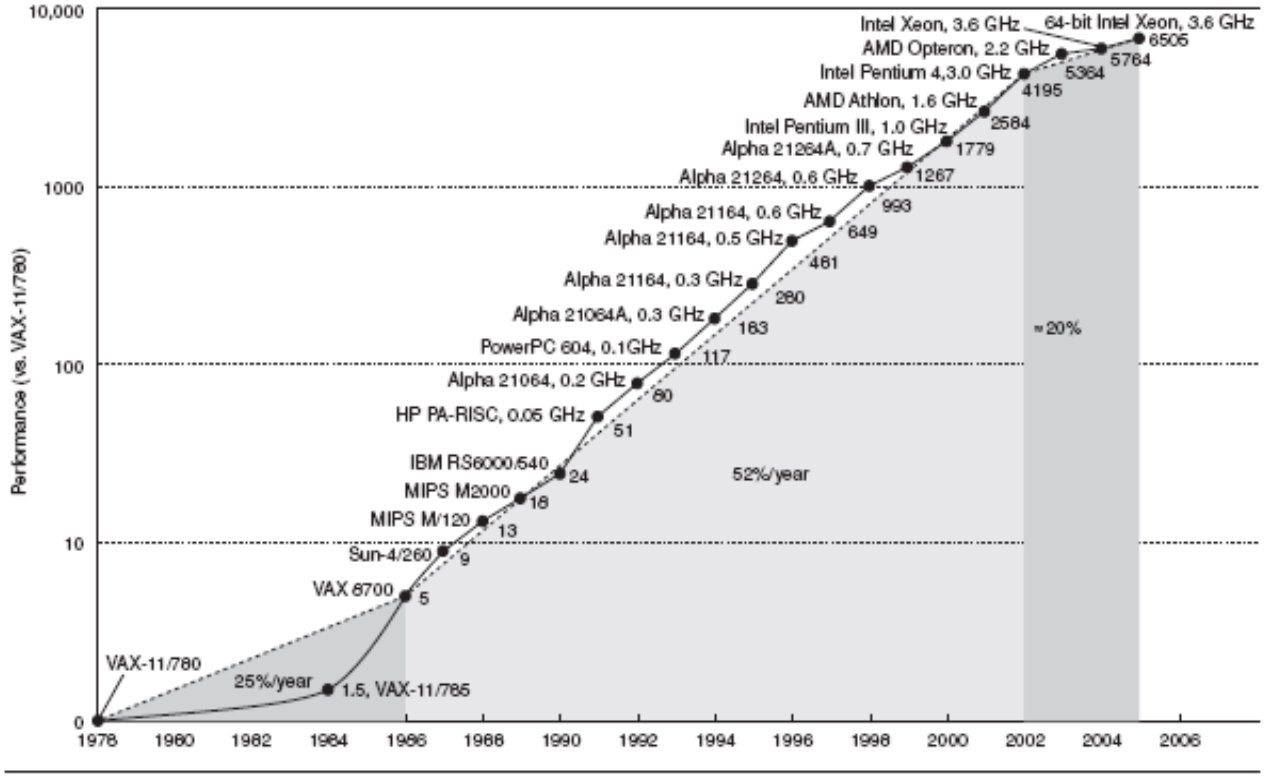
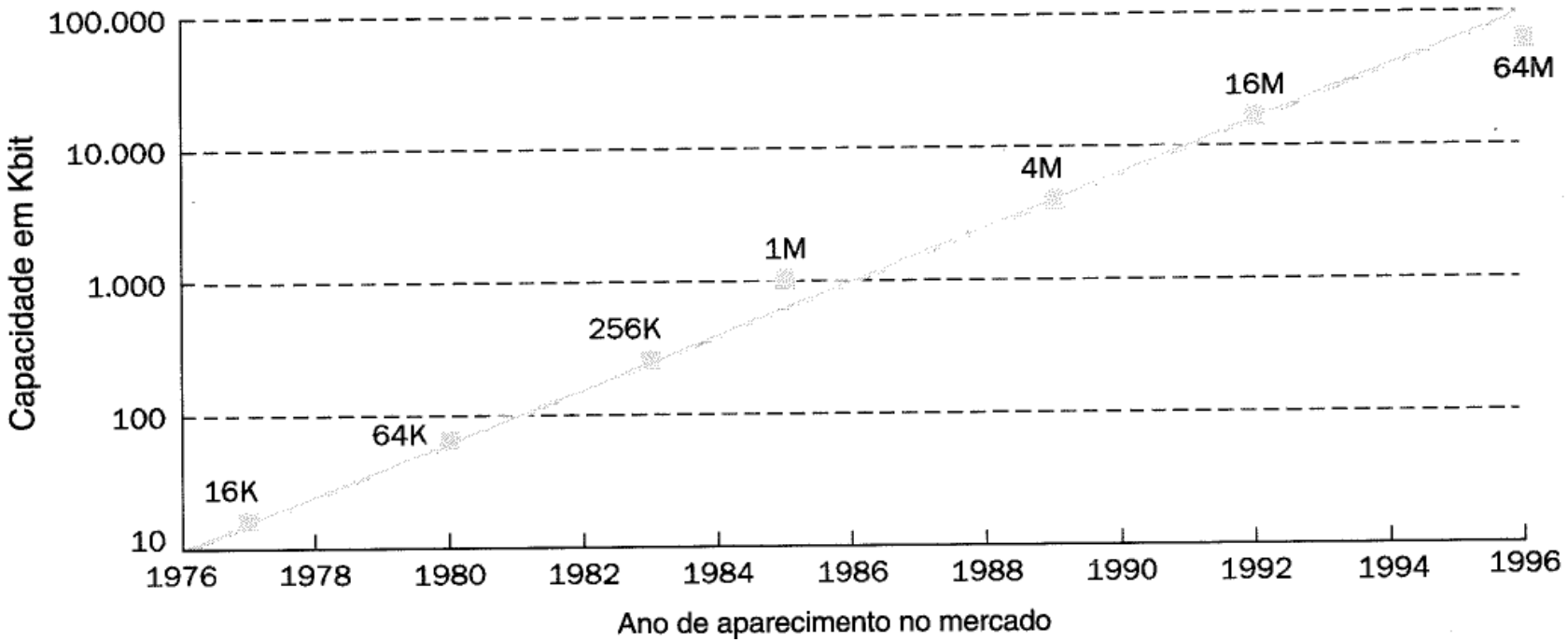


Figure 1.1 Growth in processor performance since the mid-1980s. This chart plots performance relative to the VAX 11/780 as measured by the SPECint benchmarks (see Section 1.8). Prior to the mid-1980s, processor performance growth was largely technology driven and averaged about 25% per year. The increase in growth to about 52% since then is attributable to more advanced architectural and organizational ideas. By 2002, this growth led to a difference in performance of about a factor of seven. Performance for floating-point-oriented calculations has increased even faster. Since 2002, the limits of power, available instruction-level parallelism, and long memory latency have slowed uniprocessor performance recently, to about 20% per year. Since SPEC has changed over the years, performance of newer machines is estimated by a scaling factor that relates the performance for two different versions of SPEC (e.g., SPEC92, SPEC95, and SPEC2000).

Evolução de Memórias



Introdução

- Como medir e informar desempenho?

- Tarefa não trivial

- Hardware moderno diverso e complexo
- Diversidade gigantesca de programas e aplicações
- Mas o desempenho é o principal fator de escolha de determinada arquitetura de computador

- Interesse mais amplo!

- Entender como determinado software se comporta
- Como implementar determinado conjunto de instruções
- Características do hardware que afetam significativamente o desempenho

Definição de Desempenho

- Depende da necessidade do "cliente"

- Ex. Aviões de diferentes tamanhos e velocidades

- Métricas importantes a computadores:

- Tempo de resposta: Tempo que um programa leva para ter sua execução concluída. Quanto menor melhor.
- Throughput: Quantidade de trabalho executado em um determinado intervalo de tempo. Quanto maior melhor.

- Ex. Processadores podem executar um programa por vez ou executar vários em esquema de tempo compartilhado.

Tempo de Execução

- Métrica mais básica
  - Quanto menor, melhor
- Se compararmos Máq. X e Y

$$\text{Performance}_X = \frac{1}{\text{Tempo de execução}_X}$$

$$\text{Performance}_X > \text{Performance}_Y$$

$$\frac{1}{\text{Tempo de execução}_X} > \frac{1}{\text{Tempo de execução}_Y}$$

$$\text{Tempo de execução}_Y > \text{Tempo de execução}_X$$

- Quantas vezes X é mais rápida que Y (n)?

$$\frac{\text{Performance}_X}{\text{Performance}_Y} = \frac{\text{Tempo de execução}_Y}{\text{Tempo de execução}_X} = n$$

Tempo de Execução

- Ex. A roda um programa em 10 s, e B em 15 s. Quantas vezes A é **mais rápida que** B?

- **O tempo de execução de um programa pode depender de vários outros fatores, como o tempo gasto pelo SO, o tempo gasto do processador com I/O ou tempo gasto com outros programas concorrentes.**
  - Uma medida do tempo de término de um programa pode mudar em uma mesma máquina. É impreciso.
  - Necessidade de outras métricas para focar o hardware.

Clocks/Ciclos

- Processadores executam suas atividades mais fundamentais em ciclos com períodos constantes.
- O período de um ciclo de clock é o tempo necessário para completá-lo.
- **Ex. Um processador de clock de 1 GHz, cujo período é de 1 ns.**
- Período é o inverso da frequência de clock.
- Relação com o tempo de execução:

$$\text{Tempo de execução no processador para o programa} = \text{Número de ciclos do clock do processador para o programa} \times \text{Ciclo do clock}$$

ou, alternativamente,

$$\text{Tempo de execução no processador para o programa} = \frac{\text{Número de ciclos do clock do processador para o programa}}{\text{Freqüência do clock}}$$

- > Desempenho: Menor ciclo ou menos ciclos por programa

Exemplo

Nosso programa favorito roda em 10 segundos em um computador A, que tem um clock de 400 MHz. Estamos tentando ajudar um projetista de computador em início de carreira a construir uma máquina B, que deverá rodar nosso programa em 6 segundos. O projetista já sabe que pode contar com a tecnologia para aumentar de modo significativo a freqüência do clock da sua máquina, mas este aumento vai provocar reflexos em outros parâmetros da performance, fazendo com que a máquina B precise de 1,2 vez mais ciclos do que a máquina A para executar o tal programa. Qual a freqüência do clock que nosso projetista deve implementar em sua máquina?

Exemplo

Nosso programa favorito roda em 10 segundos em um computador A, que tem um clock de 400 MHz. Estamos tentando ajudar um projetista de computador em início de carreira a construir uma máquina B, que deverá rodar nosso programa em 6 segundos. O projetista já sabe que pode contar com a tecnologia para aumentar de modo significativo a freqüência do clock da sua máquina, mas este aumento vai provocar reflexos em outros parâmetros da performance, fazendo com que a máquina B precise de 1,2 vez mais ciclos do que a máquina A para executar o tal programa. Qual a freqüência do clock que nosso projetista deve implementar em sua máquina?

Solução

- Número de ciclos em A:

$$\text{Tempo de processador}_A = \frac{\text{Número de ciclos de clock do processador}_A}{\text{Freqüência do clock}_A}$$

$$10 \text{ segundos} = \frac{\text{Número de ciclos de clock do processador}_A}{400 \times 10^6 \frac{\text{ciclos}}{\text{segundo}}}$$

$$\text{Número de ciclos de clock do processador}_A = 10 \text{ segundos} \times 400 \times 10^6 \frac{\text{ciclos}}{\text{segundo}} = 4000 \times 10^6 \text{ ciclos}$$

- Clock em B a partir do tempo de proc. Em B

$$\text{Tempo de processador}_B = \frac{1{,}2 \times \text{Número de ciclos de clock do processador}_A}{\text{Freqüência do clock}_B}$$

$$6 \text{ segundos} = \frac{1{,}2 \times 4000 \times 10^6 \text{ ciclos}}{\text{Freqüência do clock}_B}$$

$$\text{Freqüência do clock}_B = \frac{1{,}2 \times 4000 \times 10^6 \text{ ciclos}}{6 \text{ segundos}} = \frac{800 \times 10^6 \text{ ciclos}}{\text{segundo}} = 800 \text{ MHz}$$

Relação com Número de Instruções

- É importante relacionar desempenho do hardware com o software que eles executa.
- **Programas são um conjunto de instruções.**
- O tempo de execução de um programa é o tempo de execução de todas as suas instruções, uma a uma.
- Uma forma de relacionar a execução de instruções é considerar o tempo médio de execução destas
  - -Obs. Instruções diferentes podem levar diferentes nros de ciclos
- Em número de ciclos de clock:
  - Número médio de ciclos por instrução é chamado CPI

$$\begin{array}{l}\text{Número de}\\ \text{ciclos de clock}\\ \text{do processador}\end{array} = \begin{array}{l}\text{Número de}\\ \text{instruções de}\\ \text{um programa}\end{array} \times \begin{array}{l}\text{Número}\\ \text{médio de ciclos}\\ \text{por instrução}\end{array}$$

Exemplo

Suponha que temos duas implementações diferentes da mesma arquitetura do conjunto de instruções. A máquina A tem um ciclo de clock de 1 ns e uma CPI de 2,0 considerando um programa qualquer. A máquina B tem um ciclo de clock de 2 ns e uma CPI de 1,2, para o mesmo programa. Qual das duas máquinas executa esse programa mais rapidamente? Calcule também quanto uma é mais rápida que a outra.

Solução

- As duas máquinas executam o mesmo número de instruções (I), pois trata-se do mesmo programa

$$\text{Ciclos}_A = I \times 2{,}0$$
$$\text{Ciclos}_B = I \times 1{,}2$$

$$\text{Tempo}_A = \text{Ciclos}_A \times \text{Clock}_A = I \times 2{,}0 \times 1\ \text{ns} = 2 \times I\ \text{ns}$$

$$\text{Tempo}_B = \text{Ciclos}_B \times \text{Clock}_B = I \times 1{,}2 \times 2\ \text{ns} = 2{,}4 \times I\ \text{ns}$$

- A mais rápido que B, pelo fator de:

$$\frac{\text{Performance do processador}_A}{\text{Performance do processador}_B} = \frac{\text{Tempo de execução}_B}{\text{Tempo de execução}_A} = \frac{2{,}4 \times I\ \text{ns}}{2 \times I\ \text{ns}} = 1{,}2$$

Fórmula básica

- Relação dos principais fatores que afetam o desempenho de um computador:
  - Clock, CPI e Nro de Instruções

$$\text{Tempo de processador} = \text{Número de instruções} \times \text{CPI} \times \text{ciclo de clock}$$

ou

$$\text{Tempo de processador} = \frac{\text{Número de instruções} \times \text{CPI}}{\text{Freqüência de clock}}$$

CPI

- Como obter o CPI?

- Medição de programas

-Determinadas arquiteturas têm quantidades de ciclos diferentes para diferentes instruções

- CPI depende dos programas

- CPI pode ser levado em consideração individualmente ou em classes, conhecendo-se os diferentes tipos de instruções e suas frequências de execução (Ci = Nro de instruções de um tipo ou classe)

$$\text{Número de ciclos do processador} = \sum_{i=1}^{n} (\text{CPI}_i \times \text{C}_i)$$

Exemplo

Um projetista de compilador está tentando decidir entre duas seqüências de código para uma determinada máquina. Para tanto, obteve com a equipe de hardware os seguintes dados:

Classe de instrução	CPI para esta classe de instrução
A	1
B	2
C	3

Considerando o código a ser gerado para uma particular declaração de uma linguagem de alto nível, o responsável pelo projeto do compilador está considerando duas possíveis seqüências de código, com as seguintes contagens de instruções:

Seqüência de código	Número de instrução para a classe		
	A	B	C
1	2	1	2
2	4	1	1

Qual das duas seqüências executa mais instruções? Qual a mais rápida? Qual a CPI para cada seqüência?

Solução

- Seq. 1 executa 5 instruções. Seq. 2, 6 instruções.
- Total de ciclos para as seqs.:

$$\begin{aligned}\text{Ciclos de clock}_1 &= (2 \times 1) + (1 \times 2) + (2 \times 3) = \\ &= 2 + 2 + 6 = 10 \text{ ciclos} \\ \text{Ciclos de clock}_2 &= (4 \times 1) + (1 \times 2) + (1 \times 3) = \\ &= 4 + 2 + 3 = 9 \text{ ciclos}\end{aligned}$$

- Seq. 2 mais rápida com mais instruções!!
- CPIs:

$$\text{CPI} = \frac{\text{Ciclos de clock do processador}}{\text{Número de instruções}}$$

$$\text{CPI}_1 = \frac{\text{Ciclos de clock do processador}_1}{\text{Número de instruções}_1} = \frac{10}{5} = 2$$

$$\text{CPI}_2 = \frac{\text{Ciclos de clock do processador}_2}{\text{Número de instruções}_2} = \frac{9}{6} = 1{,}5$$

Programas especiais para análise de desempenho

- Workload: Conjunto de programas que um usuário usa para comparar computadores, bastaria rodar o mesmo workload nas máquinas
- **Mas usuários diferente têm workloads diferentes**
- **Alternativa: Bechmarks**
  - Conjunto de programas próprios para avaliar desempenho
  - Tentam representar sequências de instruções importantes e comuns em programas reais
  - Como são muito conhecidos, podem ser burlados
    - Otimizações podem ser feitas visando apenas trechos muito específicos do benchmark
    - Exemplo da Intel Pentium, em 92, que uso um compilador especial no teste matrix300 do SPEC
  - Sempre será melhor avaliar com base em programas reais

SPEC92 System Performance Evaluation Cooperative

- Benchmark baseado em alguns programas reais

Benchmark	Source	Lines of code	Description
espresso	C	13,500	Minimizes Boolean functions.
li	C	7,413	A lisp interpreter written in C that solves the 8-queens problem.
eqntott	C	3,376	Translates a Boolean equation into a truth table.
compress	C	1,503	Performs data compression on a 1-MB file using Lempel-Ziv coding.
sc	C	8,116	Performs computations within a UNIX spreadsheet.
gcc	C	83,589	Consists of the GNU C compiler converting preprocessed files into optimized Sun-3 machine code.
spice2g6	FORTRAN	18,476	Circuit simulation package that simulates a small circuit.
doduc	FORTRAN	5,334	A Monte Carlo simulation of a nuclear reactor component.
mdljdp2	FORTRAN	4,458	A chemical application that solves equations of motion for a model of 500 atoms. This is similar to modeling a structure of liquid argon.
wave5	FORTRAN	7,628	A two-dimensional electromagnetic particle-in-cell simulation used to study various plasma phenomena. Solves equations of motion on a mesh involving 500,000 particles on 50,000 grid points for 5 time steps.

Exemplo de Descrição de Procedimento de Análise

Hardware		Software	
Model number	Powerstation 590	O/S and version	AIX version 3.2.5
CPU	66.67 MHz POWER2	Compilers and version	C SET++ for AIX C/C++ version 2.1 XL FORTRAN/6000 version 3.1
FPU	Integrated	Other software	See below
Number of CPUs	1	File system type	AIX/JFS
Primary cache	32KBI+256KBD off chip	System state	Single user
Secondary cache	None		
Other cache	None		
Memory	128 MB		
Disk subsystem	2x2.0 GB		
Other hardware	None		

SPECbase_fp92 tuning parameters/notes/summary of changes:

FORTRAN flags: -O3 -qarch=pwrx -qhsflt -qnofold -bnso -BI:/lib/syscalss.exp

C flags: -O3 -qarch=pwrx -Q -qtune=pwrx -qhssngl -bnso -bI:/lib/syscalls.exp

Princípios para Melhora de Desempenho

- Faça os casos mais comuns melhores!
    - Quanto e onde gasta recursos para melhorias?

- Comparação de melhoria: Lei de Amdahl
    - Define quando é ganho com determinada melhoria

$$\text{Speedup} = \frac{\text{Performance for entire task using the enhancement when possible}}{\text{Performance for entire task without using the enhancement}}$$

Alternatively,

$$\text{Speedup} = \frac{\text{Execution time for entire task without using the enhancement}}{\text{Execution time for entire task using the enhancement when possible}}$$

Outras Métricas de Desempenho

- MIPS (Milhões de Instruções por segundo)
  - Muito comum para marketing e fácil de entender
  - Depende das instruções, dificultando comparação de diferentes máquinas
  - Varia de programa a programa
  - Pode variar de forma inversa ao desempenho se escolher um conjunto de interesse de instruções

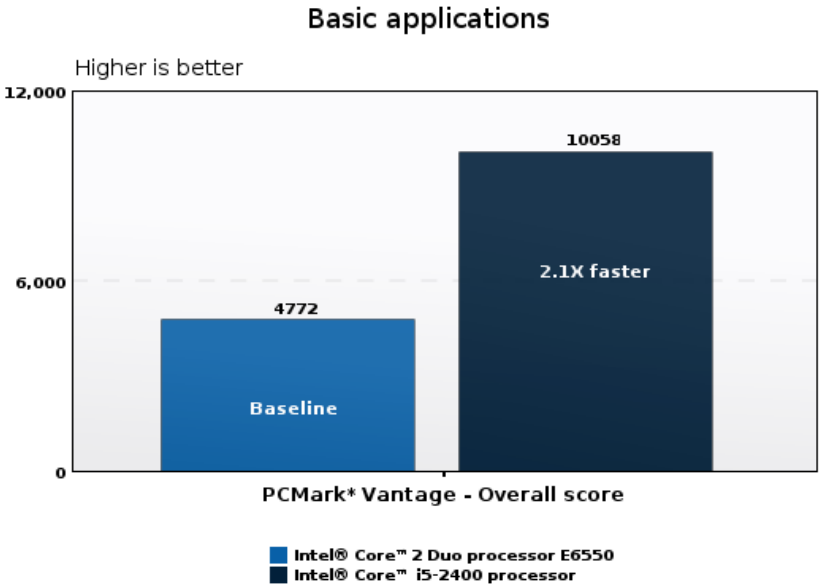
$$\text{MIPS} = \frac{\text{Instruction count}}{\text{Execution time} \times 10^6} = \frac{\text{Clock rate}}{\text{CPI} \times 10^6}$$

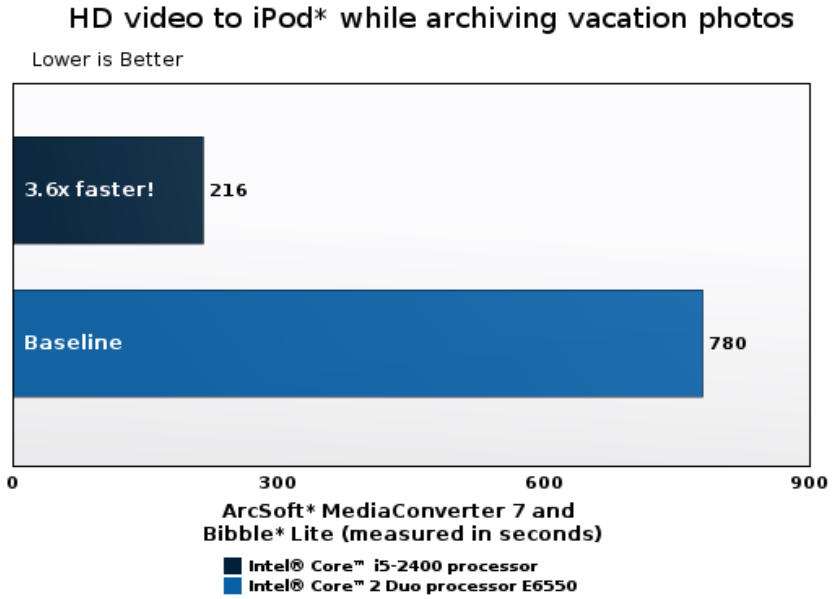
Outras Métricas de Desempenho

- MFLOPs (Milhões de Operações de Ponto Flutuante por segundo)
  - Também muito popular para divulgação
  - Sofre dos problemas anteriores. Ex. Compiladores quase nunca usam operações de ponto flutuante

$$\text{MFLOPS} = \frac{\text{Number of floating-point operations in a program}}{\text{Execution time in seconds} \times 10^6}$$

Exemplo


Basic applications
Higher is better
12,000
6,000
0
10058
2.1X faster
4772
Baseline
PCMark* Vantage - Overall score
Intel® Core™ 2 Duo processor E6550
Intel® Core™ i5-2400 processor


HD video to iPod* while archiving vacation photos
Lower is Better
3.6x faster!
216
Baseline
780
0
300
600
900
ArcSoft* MediaConverter 7 and
Bibble* Lite (measured in seconds)
Intel® Core™ i5-2400 processor
Intel® Core™ 2 Duo processor E6550

Exemplo: Top 10 Supercomputers

https://www.top500.org/lists/top500/list/2021/11/

Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	**Supercomputer Fugaku** - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu RIKEN Center for Computational Science Japan	7,630,848	442,010.0	537,212.0	29,899
2	**Summit** - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM DOE/SC/Oak Ridge National Laboratory United States	2,414,592	148,600.0	200,794.9	10,096
3	**Sierra** - IBM Power System AC922, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM / NVIDIA / Mellanox DOE/NNSA/LLNL United States	1,572,480	94,640.0	125,712.0	7,438
4	**Sunway TaihuLight** - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway, NRCPC National Supercomputing Center in Wuxi China	10,649,600	93,014.6	125,435.9	15,371
5	**Perlmutter** - HPE Cray EX235n, AMD EPYC 7763 64C 2.45GHz, NVIDIA A100 SXM4 40 GB, Slingshot-10, HPE DOE/SC/LBNL/NERSC United States	761,856	70,870.0	93,750.0	2,589

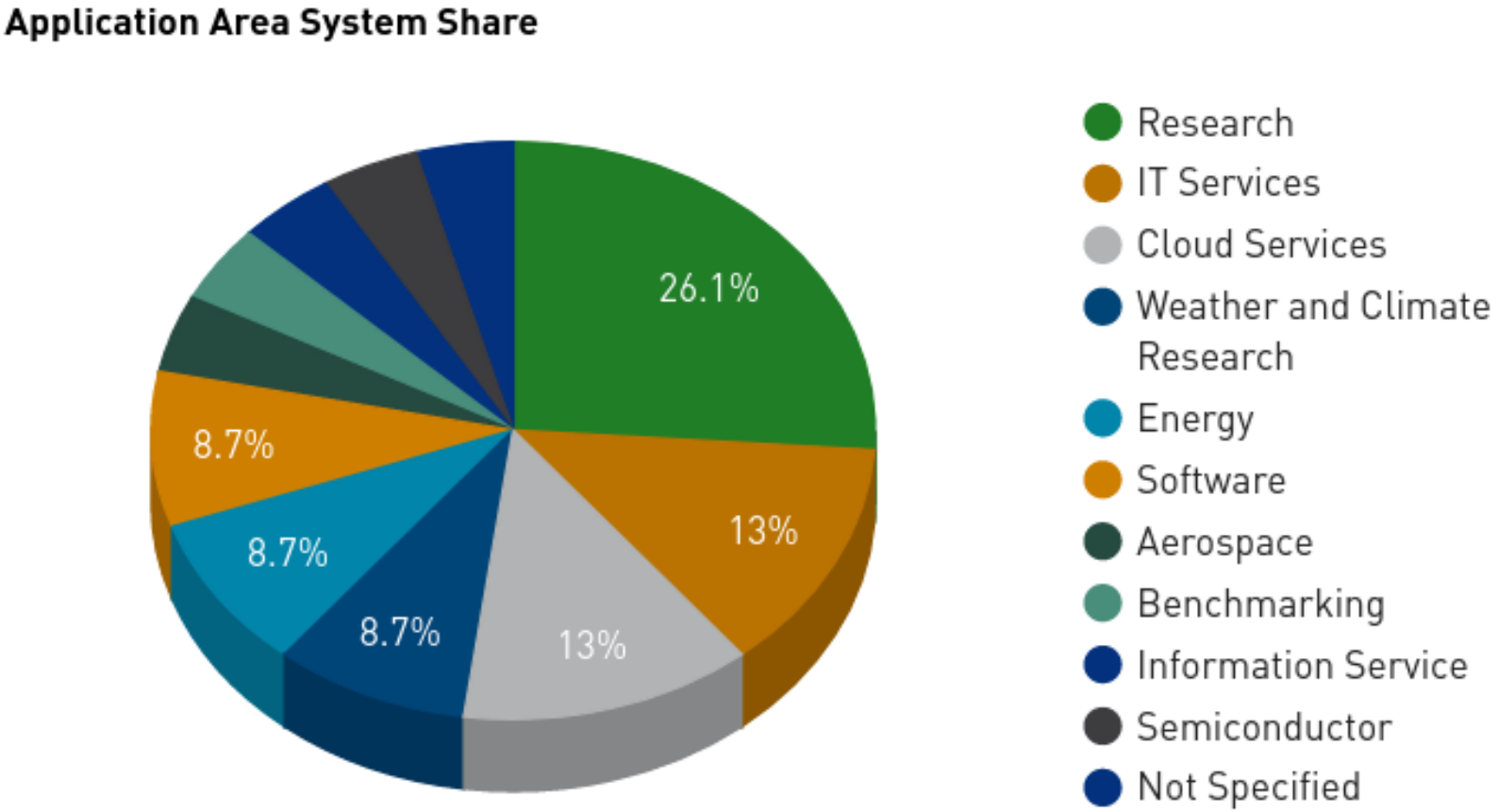
Exemplo: Top 10 Supercomputers

https://www.top500.org/lists/top500/list/2021/11/

Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
55	**Dragão** - Supermicro SYS-4029GP-TVRT, Xeon Gold 6230R 26C 2.1GHz, NVIDIA Tesla V100, Infiniband EDR, Atos Petróleo Brasileiro S.A Brazil	188,224	8,983.0	14,006.5	943

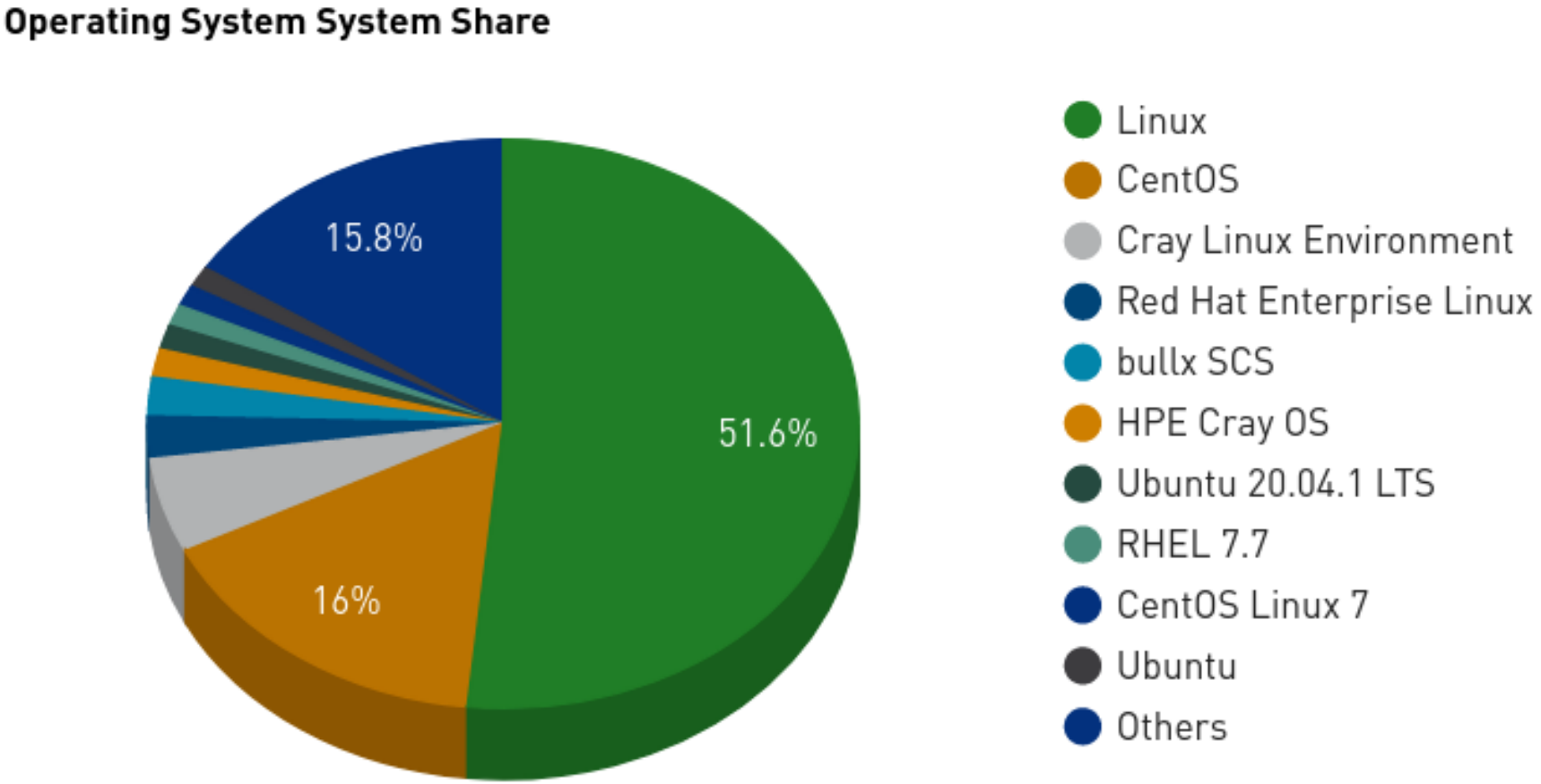
Exemplo: Top 10 Supercomputers

https://www.top500.org/statistics/list/



Exemplo: Top 10 Supercomputers

https://www.top500.org/statistics/list/



Como avançar no desempenho

- Esforços, até então, concentram-se em aumentar frequência de clock e construir arquiteturas que favoreçam CPI.
- Mas estamos chegando no limite do silício
- Clock alto aumenta aquecimento e consumo de energia
- Mais recente, o esforço passou a considerar mais CPUs ou núcleos, sendo o assunto relevante de P&D.

- Futuro?

-Spintrônica: Novos materiais com novas propriedades e com menor dissipação

- Bio-computação: Ex. Armazenamento em bactérias
- Nano-Computação: Nano-materiais e nano-máquinas
- ???