

Infraestrutura de Hardware

Aula 04 - Desempenho Computacional

2016.2

João Marcelo Teixeira
joao.marceloteixeira@ufrpe.br

DEINFO - UFRPE

4 Desempenho Computacional

Desempenho é chave em decisões de projeto, assim como custo e consumo de potência

- É um dos fatores que têm impulsionado a inovação
- Não é o mesmo que velocidade (taxa de clock mais alta)

Tópicos neste Capítulo

4.1 Custo, Desempenho e Custo/Desempenho

4.2 Definição de Desempenho Computacional

4.3 Melhorando o Desempenho e a Lei de Amdahl

4.4 Medição de Desempenho vs Modelagem

4.5 Comparando o Desempenho Computacional

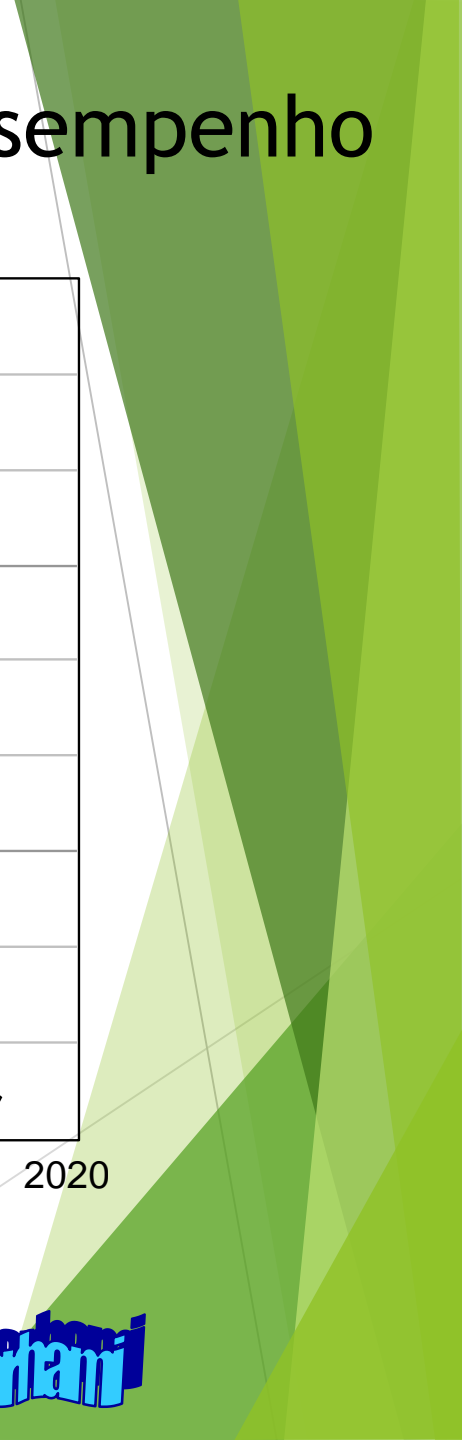
4.6 A Busca por mais Desempenho

Desempenho

2020

Desempenho

Desempenho



Custo/Desempenho

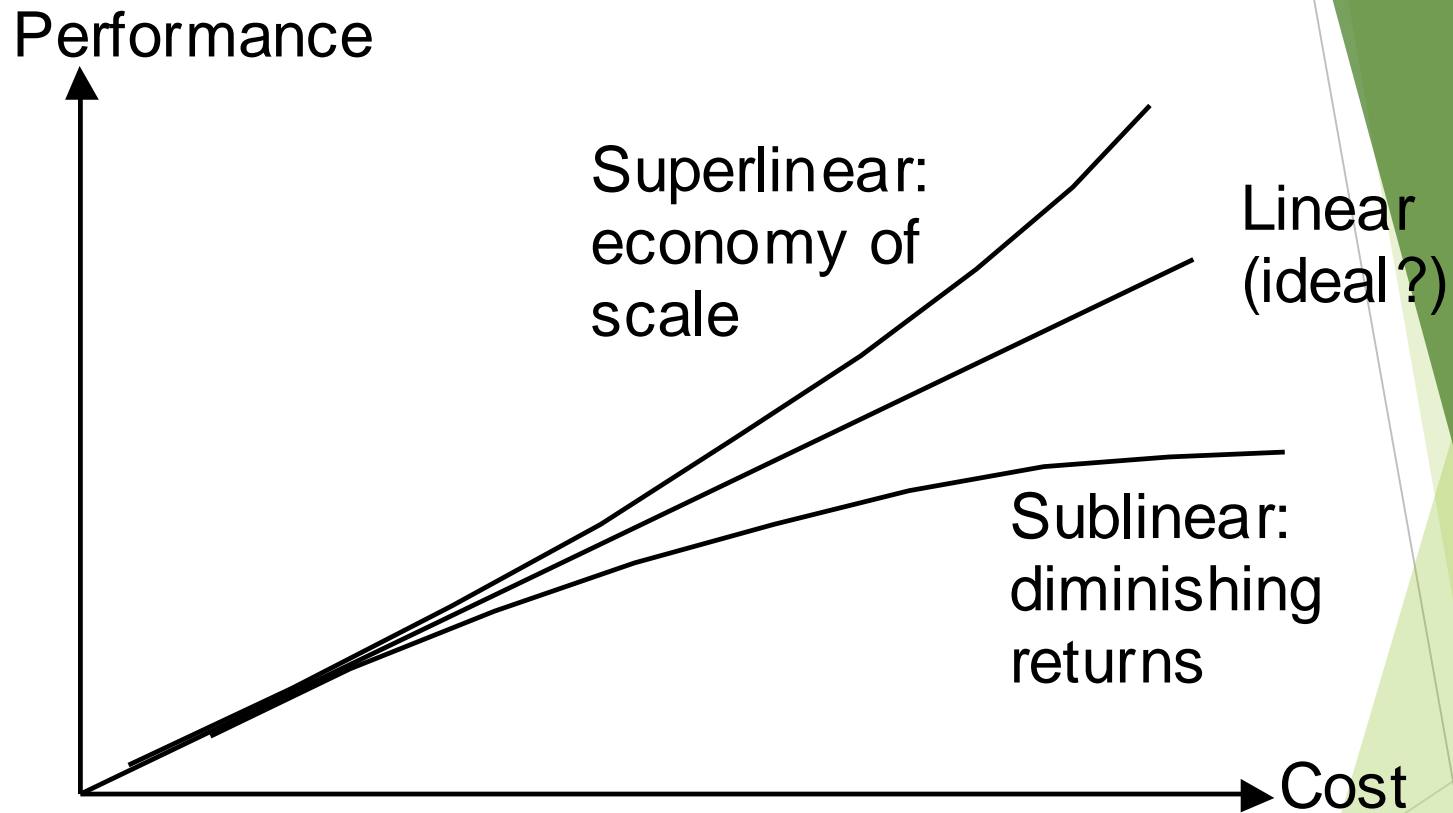


Figura 4.1 Melhoria no desempenho como uma função do custo.

4.2 Definição de Desempenho Computacional

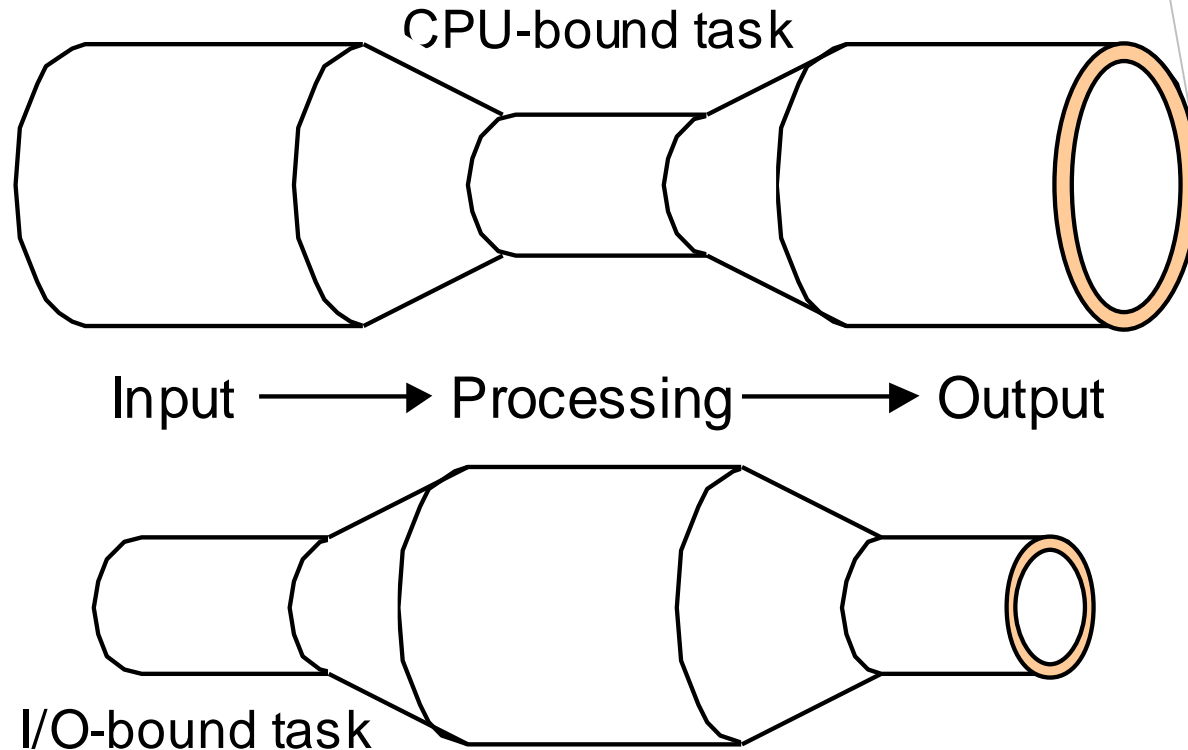


Figura 4.2 Os pipelines acima mostram que um desequilíbrio entre poder de processamento e capacidade de E/S gera um gargalo no desempenho.

Desempenho em uma Aeronave: Uma Analogia

Tabela 4.1 Características chave de seis aeronaves de passageiros: todos os valores são aproximados; alguns estão relacionados com um modelo/configuração específica da aeronave ou correspondem a uma média dos valores encontrados.

| Aeronave | Passageiros | Alcance (km) | Velocidade (km/h) | Valor (\$M) |
|-----------------|--------------------|-------------------------|------------------------------|------------------------|
| Airbus A310 | 250 | 8 300 | 895 | 120 |
| Boeing 747 | 470 | 6 700 | 980 | 200 |
| Boeing 767 | 250 | 12 300 | 885 | 120 |
| Boeing 777 | 375 | 7 450 | 980 | 180 |
| Concorde | 130 | 6 400 | 2 200 | 350 |
| DC-8-50 | 145 | 14 000 | 875 | 80 |

Diferentes Visões de Desempenho

Desempenho do ponto de vista de um passageiro: **Velocidade**

Note que, todavia, o tempo de vôo corresponde a uma parte do tempo total da viagem.

Além disso, se a distância a ser percorrida excede o **alcance** de uma aeronave mais rápida, uma aeronave mais lenta pode ser melhor por não precisar de pausas para reabastecimento.

Desempenho do ponto de vista de uma companhia aérea: **Vazão**

Medida em passageiro-km por hora (relevante se o valor da passagem fosse proporcional à distância viajada, o que não é verdade na realidade)

Airbus A310 $250 \times 895 = 0.224$ M passageiro-km/hr

Boeing 747 $470 \times 980 = 0.461$ M passageiro-km/hr

Boeing 767 $250 \times 885 = 0.221$ M passageiro-km/hr

Boeing 777 $375 \times 980 = 0.368$ M passageiro-km/hr

Concorde $130 \times 2200 = 0.286$ M passageiro-km/hr

DC-8-50 $145 \times 875 = 0.127$ M passageiro-km/hr

Desempenho do ponto de vista da ANAC: **Segurança**

Custo Efetivo: Custo/Desempenho

Tabela 4.1 Características chave de seis aeronaves de passageiros: todos os valores sÃo aproximados; alguns estÃo relacionados com um modelo/configuraÃÃo especÃfica da aeronave ou correspondem a uma media dos valores encontrados.

| Aeronave | Passageiros | Alcance (km) | Velocidade (km/h) | Valor (\$M) |
|----------|-------------|--------------|-------------------|-------------|
| A310 | 250 | 8 300 | 895 | 120 |
| B 747 | 470 | 6 700 | 980 | 200 |
| B 767 | 250 | 12 300 | 885 | 120 |
| B 777 | 375 | 7 450 | 980 | 180 |
| Concorde | 130 | 6 400 | 2 200 | 350 |
| DC-8-50 | 145 | 14 000 | 875 | 80 |

Quanto maior o valor, melhor

Quanto menor o valor, melhor

Vazão
(M P km/hr)

0.224

0.461

0.221

0.368

0.286

0.127

Custo /
Desempenho

536

434

543

489

1224

630

Definição de Desempenho e Speedup

Desempenho = $1 / \text{Tempo de execução}$  é simplificado para

Desempenho = $1 / \text{Tempo de exec. da CPU}$ 

$$\begin{aligned} (\text{Desempenho de } M_1) / (\text{Desempenho de } M_2) &= \text{Speedup de } M_1 \text{ sobre } M_2 \\ &= (\text{Tempo de Execução de } M_2) / (\text{Tempo de execução de } M_1) \end{aligned}$$

Terminologia: M_1 é x vezes **tão rápido quanto** M_2 (ex., 1.5 vezes tão rápido)
 M_1 é $100(x - 1)\%$ **mais rápido que** M_2 (ex., 50% mais rápido)

$$\begin{aligned} \text{Tempo de CPU} &= \text{Instruções} \times (\text{Ciclos por instrução}) \times (\text{Segundos por ciclo}) \\ &= \text{Instruções} \times \text{CPI} / (\text{Taxa de clock}) \end{aligned}$$

O número de instruções, CPU e taxa de clock não são completamente independentes, então melhorando um deles em determinado fator pode não levar a uma melhoria geral no mesmo fator.

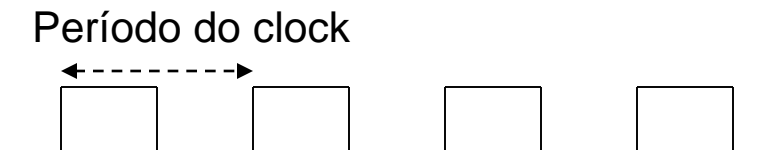
Elaboração da fórmula do tempo de CPU

Tempo de CPU = Instruções × (Ciclos por instrução) × (Segundos por ciclo) = Instruções × CPI médio / (taxa de clock)

Instruções: Número de instruções executadas, não o número de instruções no programa (contagem **dinâmica**)

CPI médio: É calculado baseando-se no conjunto dinâmico de instruções e no conhecimento de quantos ciclos de clock são necessários para executar várias instruções (ou classes de instruções)

Taxa de clock: 1 GHz = 10^9 ciclos / s (tempo de ciclo 10^{-9} s = 1 ns)
200 MHz = 200×10^6 ciclos / s (tempo de ciclo = 5 ns)



Clock mais rápido \neq Menor Tempo de Execução

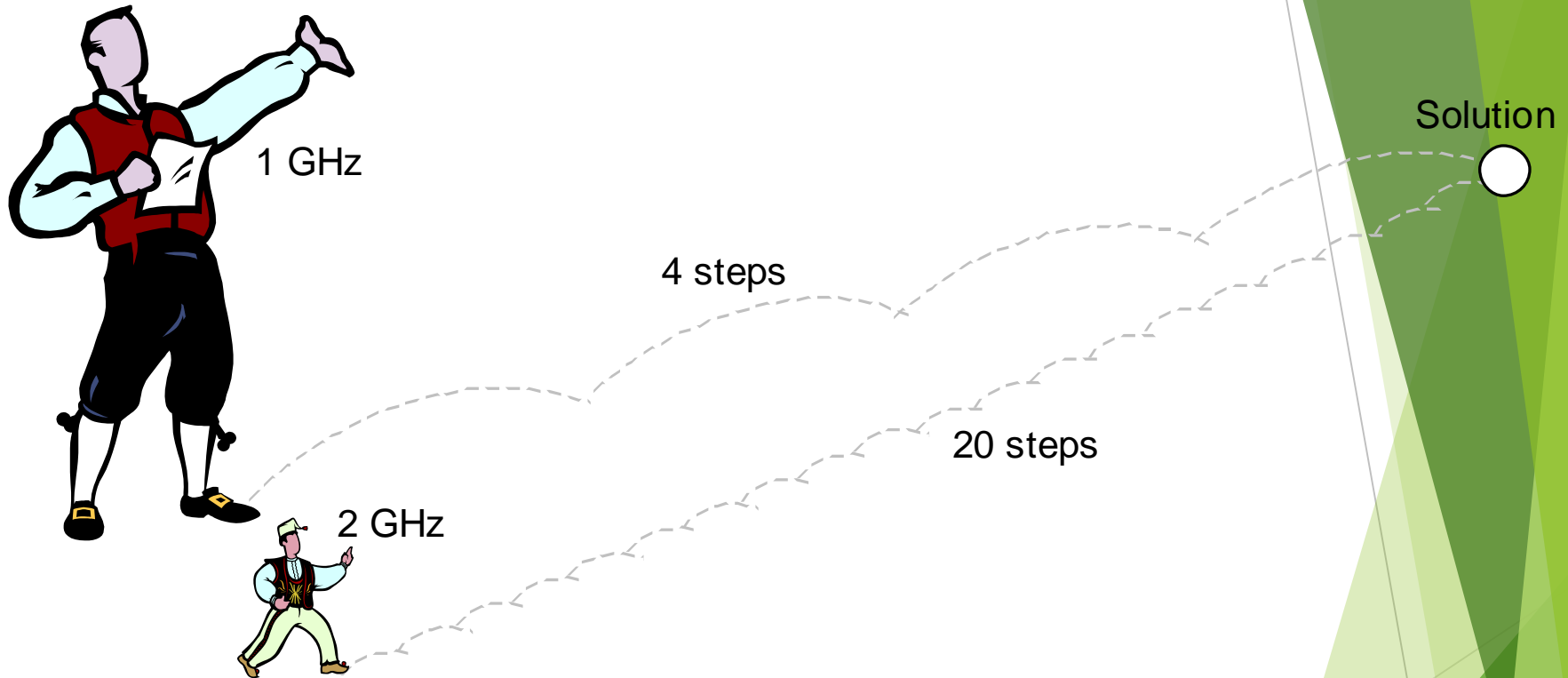
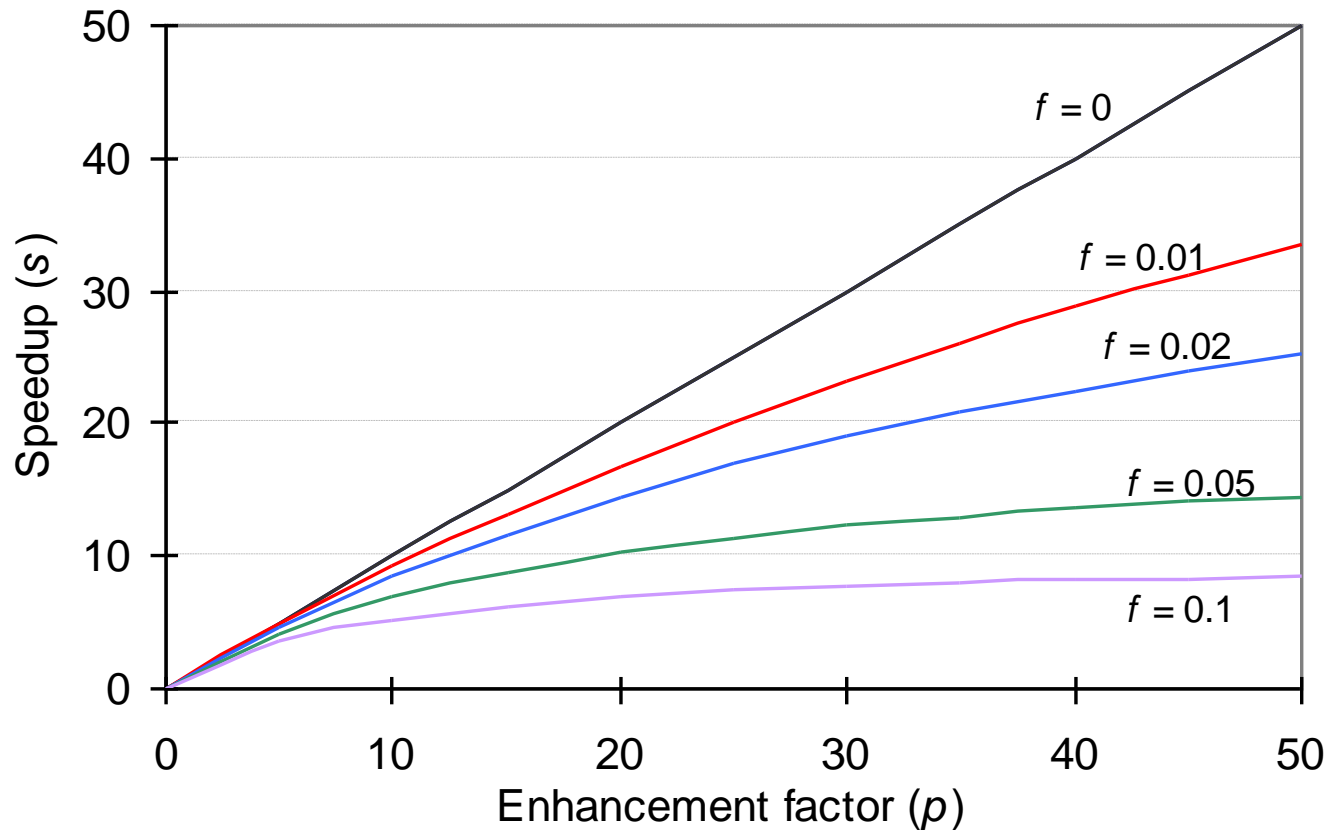


Figura 4.3 Passos mais rápidos não significam necessariamente um tempo menor de viagem.

4.3 Melhorando o Desempenho: a Lei de Amdahl



f = fração não afetada

p = speedup do restante

$$s = \frac{1}{f + (1-f)/p} \leq \min(p, 1/f)$$

Figura 4.4 Lei de Amdahl: speedup conseguido se uma fração f de uma tarefa não é afetada e a fração restante $(1 - f)$ executa p vezes mais rápido.

Uso da Lei de Amdahl no Projeto

Exemplo 4.1

Um processador gasta 30% de seu tempo com adição, 25% com multiplicação e 10% com divisão de ponto flutuante. Avalie as seguintes melhorias, todas elas custando o mesmo para serem implementadas:

- a. Redesign do somador para deixá-lo duas vezes mais rápido.
- b. Redesign do multiplicador para deixá-lo três vezes mais rápido.
- c. Redesign do divisor para deixá-lo 10 vezes mais rápido.

Solução

- a. Speedup do novo somador = $1 / [0.7 + 0.3 / 2] = 1.18$
- b. Speedup do novo multiplicador = $1 / [0.75 + 0.25 / 3] = 1.20$
- c. Speedup do novo divisor = $1 / [0.9 + 0.1 / 10] = 1.10$

E se ambos somador e multiplicador forem redesenhados?

Uso da Lei de Amdahl no Gerenciamento

Exemplo 4.2

Membros do grupo de pesquisa de uma universidade visitam a biblioteca com frequência.

Cada ida à biblioteca leva 20 minutos. O grupo decide assinar algumas publicações, as quais representam 90% das viagens realizadas à biblioteca; o tempo de acesso a essas publicações é reduzido para 2 minutos.

- Qual o speedup médio para acesso às publicações?
- Se o grupo possui 20 membros, cada um fazendo duas viagens semanais à biblioteca, qual o gasto justificável para a assinatura? Assuma 50 semanas de trabalho por ano e o valor da hora de um pesquisador é de 25 dólares.

Solução

- Speedup no tempo de acesso às publicações = $1 / [0.1 + 0.9 / 10] = 5.26$
- Tempo economizado = $20 \times 2 \times 50 \times 0.9 (20 - 2) = 32,400 \text{ min} = 540 \text{ h}$
Gasto recuperado = $540 \times \$25 = \$13,500 = \text{Valor máximo justificável}$

4.4 Medição de Desempenho vs Modelagem

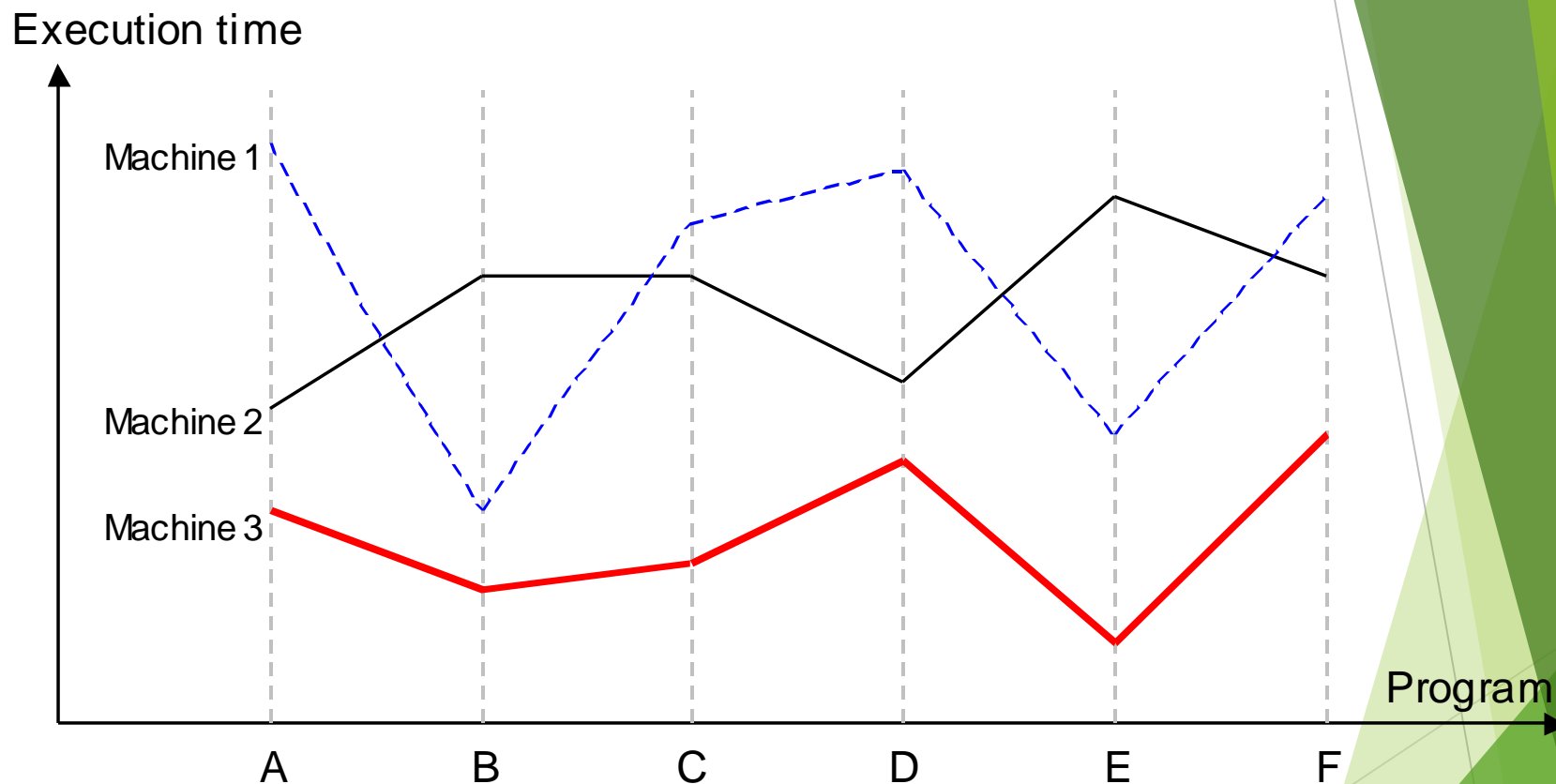


Figura 4.5 Tempo de execução de 6 programas em três máquinas.

Lei de Amdahl Generalizada

Tempo original de execução de um programa = 1 = $f_1 + f_2 + \dots + f_k$

Novo tempo de execução depois da fração f_i ser acelerada por um fator p_i

$$\frac{f_1}{p_1} + \frac{f_2}{p_2} + \dots + \frac{f_k}{p_k}$$

Fórmula do Speedup

$$S = \frac{1}{\frac{f_1}{p_1} + \frac{f_2}{p_2} + \dots + \frac{f_k}{p_k}}$$

Se uma fração em particular tornar-se mais lenta, deve-se usar $s_j f_j$ ao invés de f_j / p_j , onde $s_j > 1$ representa o fator de “lentidão”.

Referências de Desempenho

Exemplo 4.3

Você é um engenheiro da Outtel, uma startup que pretende competir com a Intel através do seu novo design de processador que ultrapassa o desempenho do ultimo processador da Intel em um fator de 2.5 em instruções de ponto flutuante. Esse nível de desempenho foi atingido com decisões de projeto que levaram a um aumento de 20% no tempo de execução de todas as outras instruções. Você está encarregado de definir referências de desempenho que irão ilustrar as vantagens do processador da Outtel.

- Qual é a menor fração f necessária de tempo gasto com instruções de ponto flutuante em um programa no processador da Intel de forma que se obtenha um speedup de 2 ou mais com o processador da Outtel?

Solução

- Nós usamos uma forma generalizada da formula de Amdahl, onde uma fração f é acelerada em um fator específico (2.5) e o restante é desacelerado por outro fator (1.2) :
- $1 / [1.2(1 - f) + f / 2.5] \geq 2 \Rightarrow f \geq 0.875$

Estimativa de Desempenho

$\text{CPI médio} = \sum_{\text{Todas as classes de instruções}} (\text{Fraction da classe-}i) \times (\text{CPI da classe-}i)$

$\text{Tempo de ciclo da máquina} = 1 / \text{taxa do clock}$

$\text{Tempo de execução da CPU} = \text{Instruções} \times (\text{CPI médio}) / (\text{taxa do clock})$

Tabela 4.3 Frequência de uso, em porcentagem, para várias classes de instruções em quatro aplicações representativas.

| Aplicação → Classe da instrução ↓ | Compressão de dados | Compilador de Linguagem C | Simulação de um reator | Modelagem de movimento atômico |
|--|--------------------------------|--------------------------------------|-----------------------------------|---|
| A: Load/Store | 25 | 37 | 32 | 37 |
| B: Inteiro | 32 | 28 | 17 | 5 |
| C: Deslocamento/Lógico | 16 | 13 | 2 | 1 |
| D: Float | 0 | 0 | 34 | 42 |
| E: Desvio | 19 | 13 | 9 | 10 |
| F: Todas as outras | 8 | 9 | 6 | 4 |

Cálculo de CPI e IPS

Exemplo 4.4 (parte 2 de 5)

Considere duas implementações M_1 (600 MHz) e M_2 (500 MHz) de um conjunto de instruções contendo três classes de instruções:

| <u>Classe</u> | <u>CPI de M_1</u> | <u>CPI de M_2</u> | <u>Descrição</u> |
|---------------|--------------------------------|--------------------------------|------------------------|
| F | 5.0 | 4.0 | Ponto-flutuante |
| I | 2.0 | 3.8 | Aritmética de inteiros |
| N | 2.4 | 2.0 | Não aritmética |

- Qual o pico de desempenho de M_1 e M_2 em MIPS (milhões de instruções por segundo)?
- Se 50% das instruções executadas são da classe N, com o restante dividido igualmente entre F e I, qual máquina seria mais rápida? O quão mais rápida?

Solução

- Pico de desempenho para $M_1 = 600 / 2.0 = 300$; for $M_2 = 500 / 2.0 = 250$
- CPI médio de $M_1 = 5.0 / 4 + 2.0 / 4 + 2.4 / 2 = 2.95$;
CPI médio de $M_2 = 4.0 / 4 + 3.8 / 4 + 2.0 / 2 = 2.95 \rightarrow M_1$ é mais rápido; 1.2 vezes

O número de MIPS pode enganar

Exemplo 4.5

Dois compiladores produzem código de máquina em uma máquina com duas classes de instruções. Segue o número de instruções:

| <u>Classe</u> | <u>CPI</u> | <u>Compilador 1</u> | <u>Compilador 2</u> |
|---------------|------------|---------------------|---------------------|
| A | 1 | 600M | 400M |
| B | 2 | 400M | 400M |

- Qual o tempo de execução dos dois programas quando se usa um clock de 1 GHz clock?
- Que compilador produz código mais rápido e o quão mais rápido ele é?
- Qual das saídas dos compiladores executa a uma maior taxa de MIPS?

Solução

- a. Tempo de execução 1 (2) = $(600M \times 1 + 400M \times 2) / 10^9 = 1.4 \text{ s}$ (1.2 s)
- b. A saída do compilador 2 roda $1.4 / 1.2 = 1.17$ vezes mais rápido
- c. A taxa de MIPS 1, CPI = 1.4 (2, CPI = 1.5) = $1000 / 1.4 = 714$ (667)

4.5 Comparando o Desempenho Computacional

Tabela 4.4 Tempos de execução estimados para três programas.

| | Tempo da máquina X | Tempo da máquina Y | Speedup de Y sobre X |
|------------|--------------------|--------------------|----------------------|
| Programa A | 20 | 200 | 0.1 |
| Programa B | 1000 | 100 | 10.0 |
| Programa C | 1500 | 150 | 10.0 |
| Todos | 2520 | 450 | 5.6 |

Analogia: Se um carro viaja para uma cidade a 100km de distância a 100km/h e retorna a 50km/h, a velocidade media não será $(100 + 50) / 2$, e sim calculada a partir do fato que o veículo viajou 200 km em 3 horas.

Comparando o Desempenho Total

Tabela 4.4 Tempos de execução estimados para três programas.

| | Tempo da máquina X | Tempo da máquina Y | Speedup de Y sobre X | Speedup de X sobre Y |
|------------|--------------------|--------------------|----------------------|----------------------|
| Programa A | 20 | 200 | 0.1 | 10 |
| Programa B | 1000 | 100 | 10.0 | 0.1 |
| Programa C | 1500 | 150 | 10.0 | 0.1 |
| | Média aritmética | | 6.7 | 3.4 |
| | Média geométrica | | 2.15 | 0.46 |

A média geométrica não representa uma medida de speedup global, mas fornece um indicador que aponta na direção correta.

4.6 A Busca por mais Desempenho

Poder computacional no início dos anos 2000:

Gigaflops em desktops

Teraflops em centros com supercomputadores

Petaflops como meta de projetos futuros

Atenção à terminologia (verificar Tabela 3.1)

Prefixos para unidades grandes:

Kilo = 10^3 , Mega = 10^6 , Giga = 10^9 , Tera = 10^{12} , Peta = 10^{15}

Para memória:

$$K = 2^{10} = 1024, \quad M = 2^{20}, \quad G = 2^{30}, \quad T = 2^{40}, \quad P = 2^{50}$$

Prefixos para unidades pequenas:

micro = 10^{-6} , nano = 10^{-9} , pico = 10^{-12} , femto = 10^{-15}

Tendências de Desempenho e Obsolescência

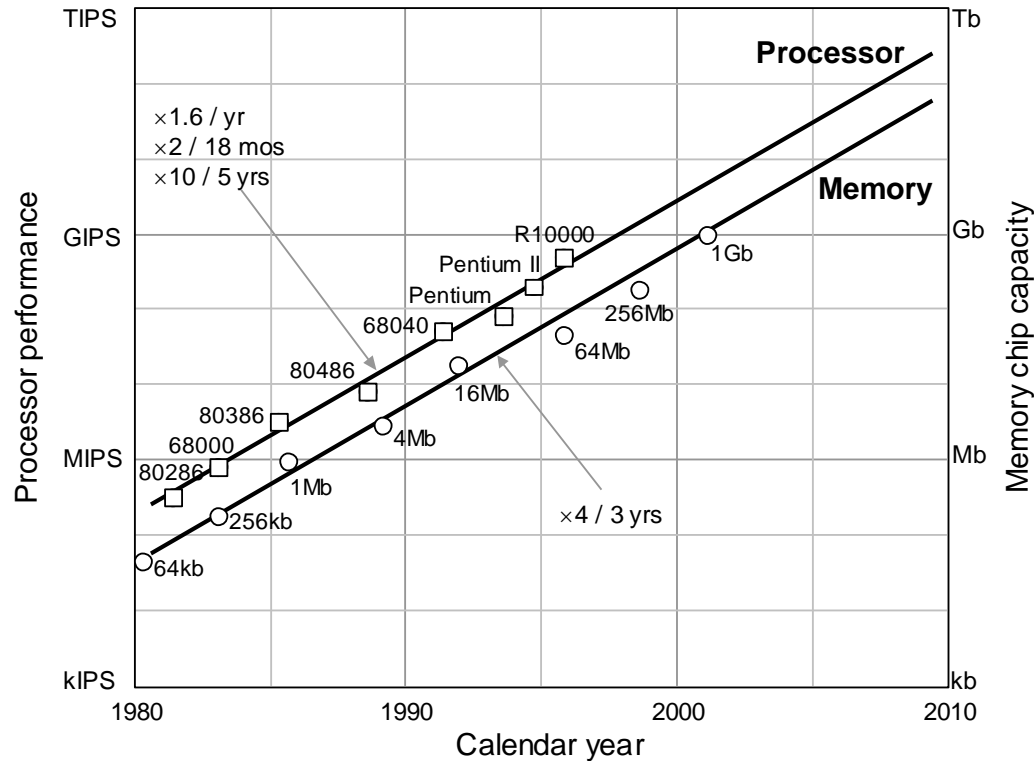


Figura 3.10 Tendências em desempenho de processadores e capacidade de memória (Lei de Moore).



“Posso retornar a ligação depois? Acabamos de comprar um computadores e estamos tentando configure-lo antes que ele fique obsoleto.”

Jan. 2007

Slide

25

B. Pariani

Supercomputadores

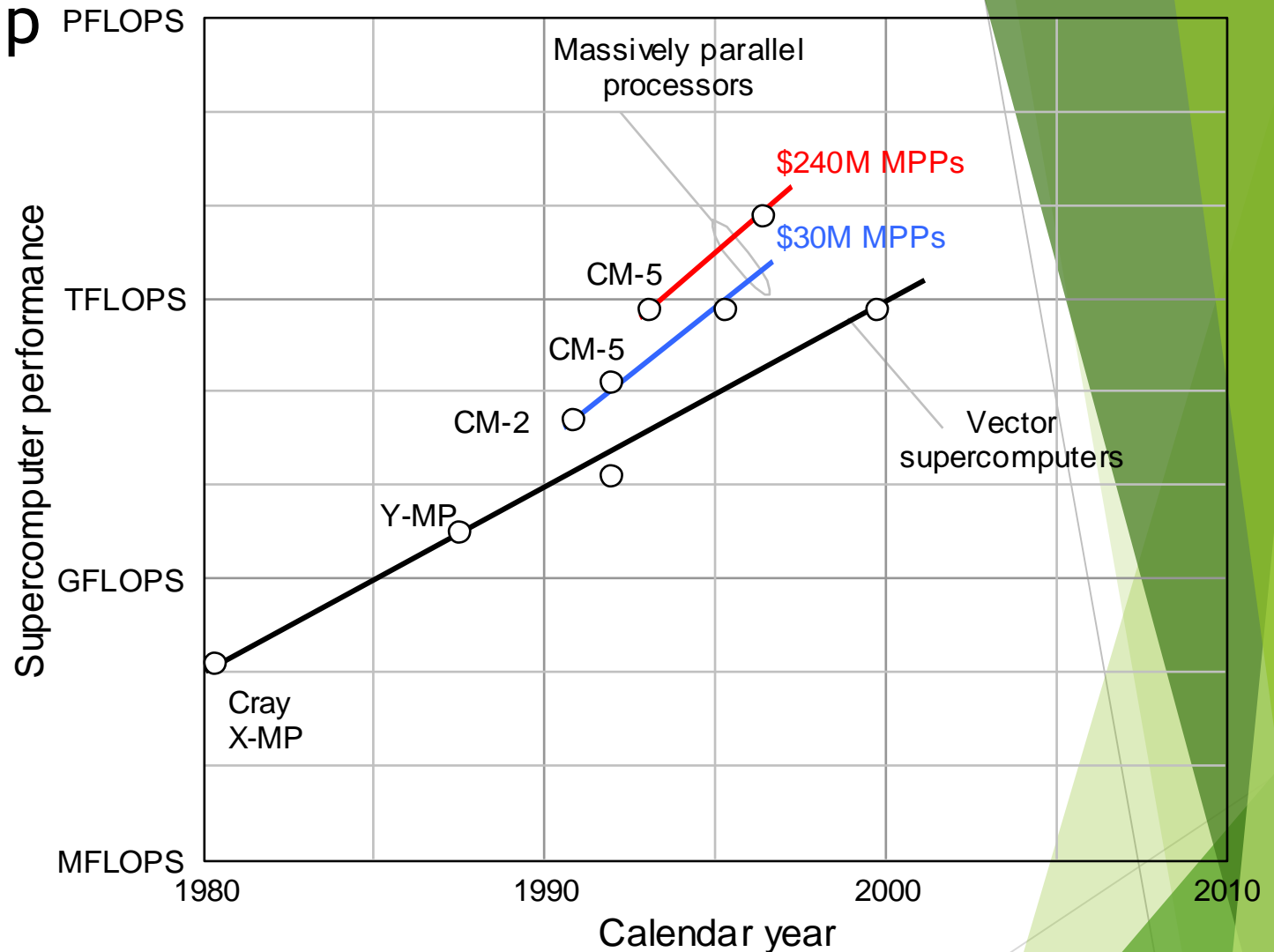


Figura 4.7 Crescimento exponencial do desempenho de supercomputadores.

Os Computadores Mais Poderosos

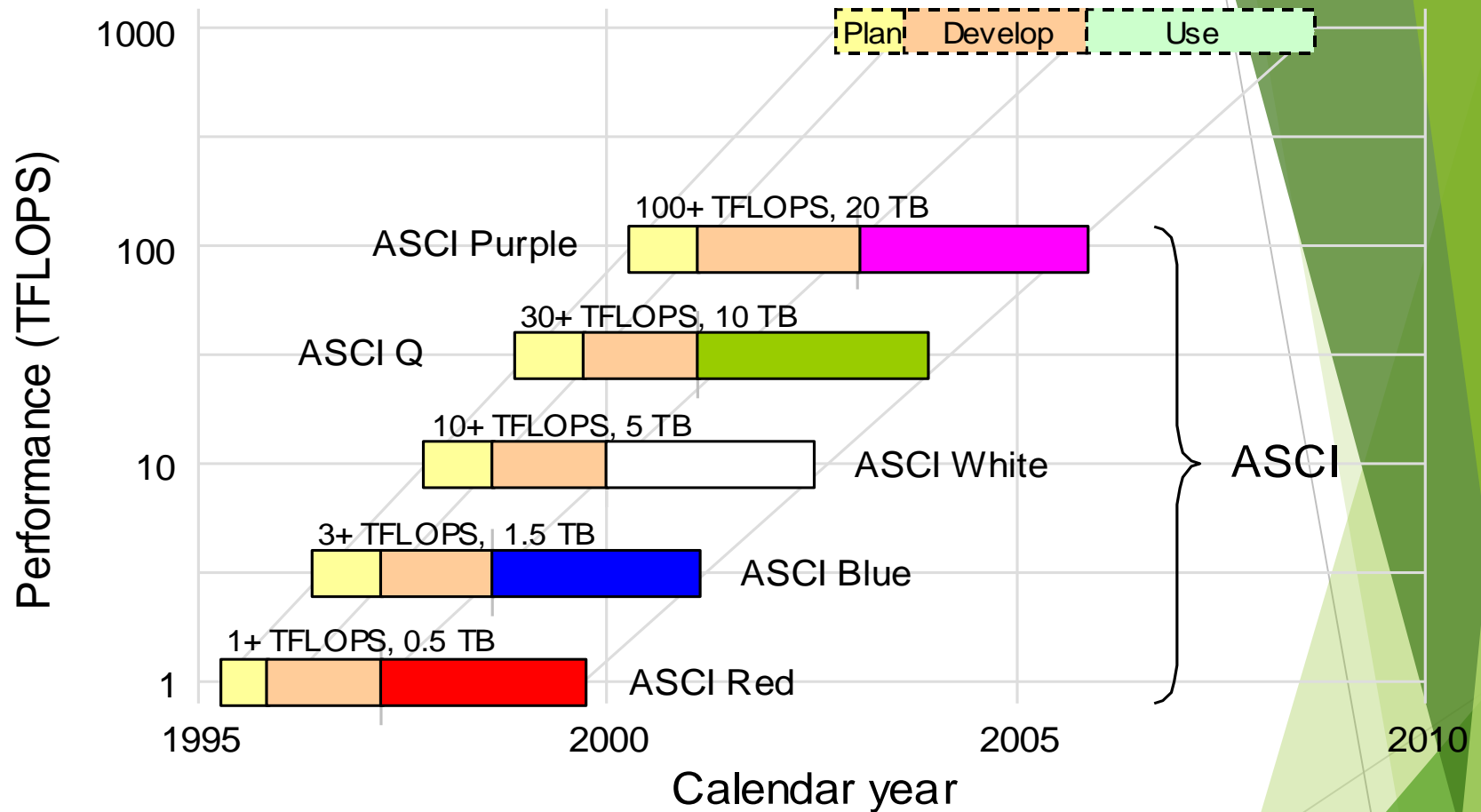


Figura 4.8 Milestones do programa da Iniciativa de Aceleração Estratégica de Computação (ASCI) da DOE extrapolando ao nível de PFLOPS.

Desempenho é Importante, mas não é Tudo

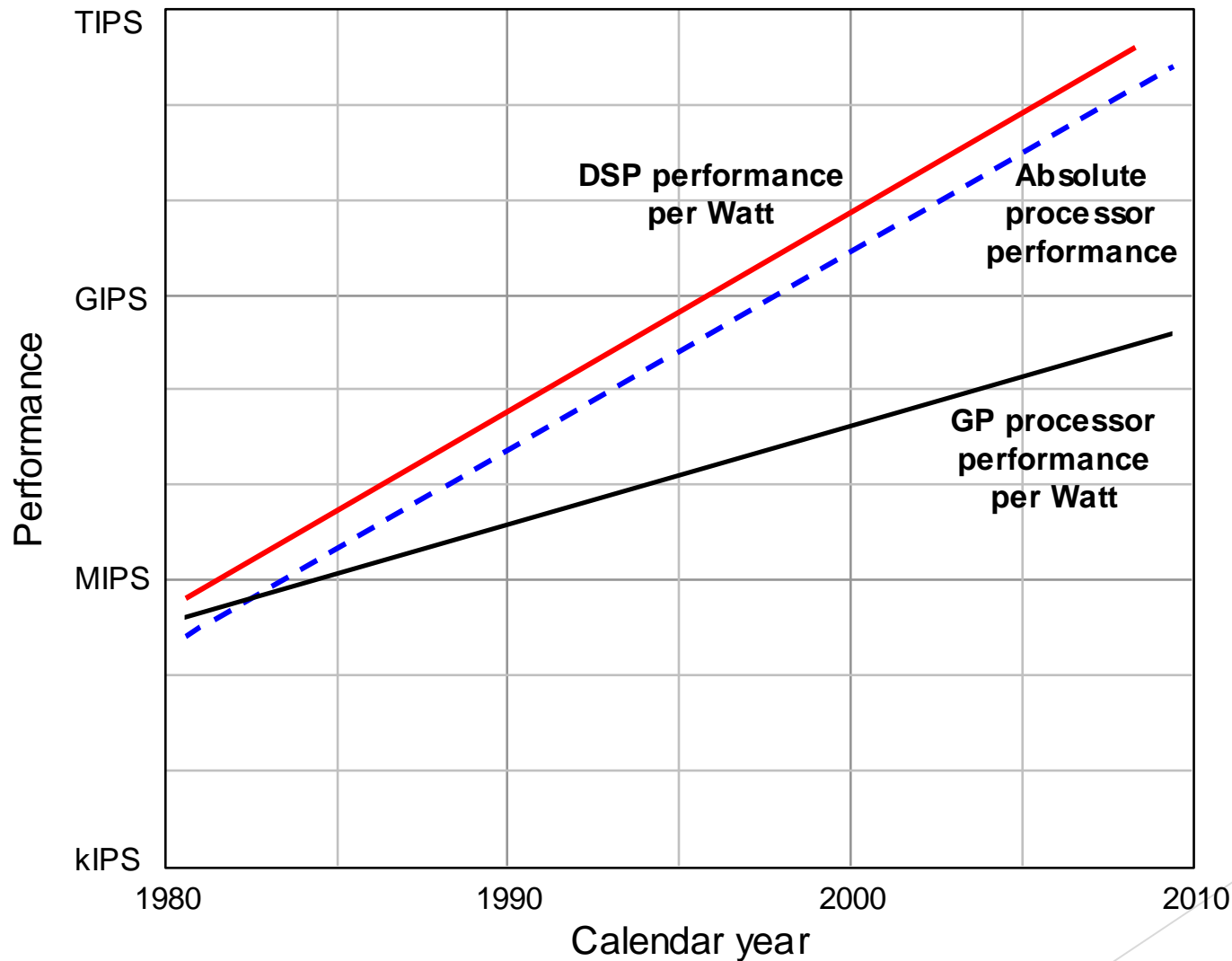


Figura 25.1
Tendência em desempenho computacional por watt de potência usada em computadores de propósito geral e DSPs.