

---

# Arquiteturas de Computadores

---

Fundamentos de Projetos de  
Computadores

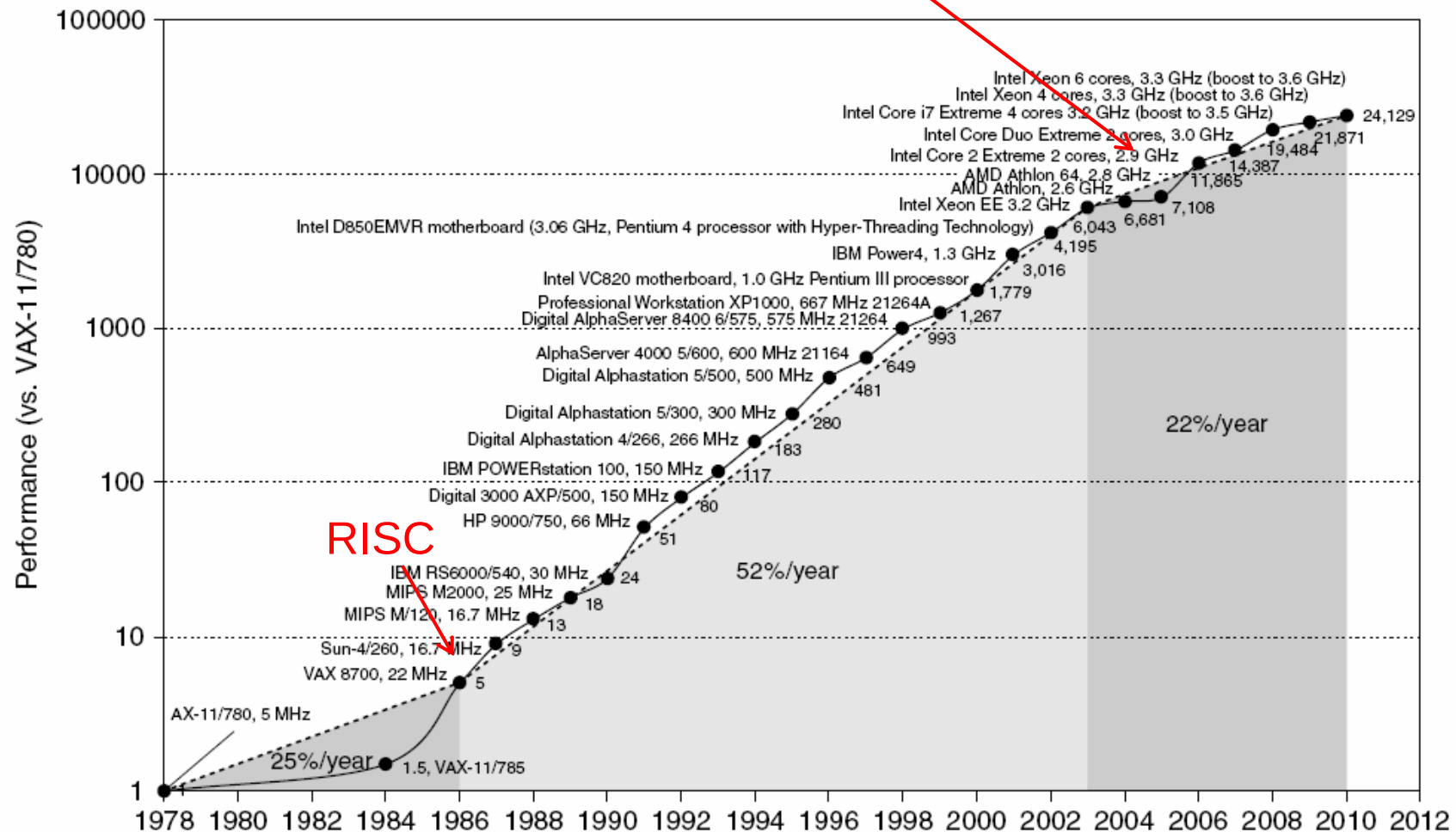
---

# Tecnologia

- Melhorias no desempenho:
    - Melhorias na tecnologia dos semicondutores
      - Menor tamanho, velocidade do relógio
    - Melhorias nas arquiteturas de computadores
      - Possibilitadas pelos compiladores de linguagem de alto nível, UNIX
      - Levou a arquiteturas RISC
-

# Desempenho de um único processador

Movimento para multiprocessador



---

# Tendências atuais em arquiteturas

- Não se pode continuar tentando somente o paralelismo a nível de instrução (ILP)
    - A melhoria de desempenho de um processador único acabou em 2003
  - Novos modelos para desempenho:
    - Paralelismo no nível de dados (DLP)
    - Paralelismo no nível de threads (TLP)
    - Paralelismo no nível de requisições (RLP)
  - Estes modelos requerem reestruturação explícita das aplicações
-

---

# Classes de computadores

- Dispositivo móvel pessoal (PMD)
    - e.g. smartphones, tablets
    - Ênfase em eficiência de energia e tempo-real
  - Computadores de desktop
    - Ênfase relação preço-desempenho
  - Servidores comerciais
    - Ênfase em disponibilidade, escalabilidade e vazão
  - Servidores Científicos
    - Ênfase em desempenho ponto-flutuante e redes internas rápidas
  - Computadores embutidos
    - Ênfase no preço
-

---

# Paralelismo

- Classes de paralelismo em aplicações:
    - ❑ Paralelismo no nível de dados (DLP)
    - ❑ Paralelismo no nível de tarefas (TLP)
  - Classes de paralelismo arquitetural:
    - ❑ Paralelismo no nível de instrução (ILP)
    - ❑ Arquiteturas vetoriais/Unidades de Processamento Gráfico(GPUs)
    - ❑ Paralelismo no nível de threads
    - ❑ Paralelismo no nível de requisição
-

---

# Definindo arquitetura de computadores

- Visão prévia de arquitetura de computadores:
    - Projeto do conjunto de instruções (ISA)
    - i.e. decisões visam definir:
      - registradores, endereçamento de memória, modos de endereçamento, operandos de instruções, operações disponíveis, instruções de controle de fluxo, codificação de instruções
  
  - Visão atual de arquiteturas de computadores:
    - Requerimentos específicos de uma determinada máquina
    - Projeto para maximizar desempenho com restrições de: custo, potência, e disponibilidade
    - Inclui ISA, microarquitetura, hardware
-

---

# Tendências em tecnologia

- Tecnologia de circuito integrado
    - Densidade de transistores: 35%/ano
    - Tamanho da pastilha: 10-20%/ano
    - Integração total: 40-55%/ano
  - Capacidade DRAM: 25-40%/ano (reduzindo)
  - Capacidade flash: 50-60%/ano
    - 15-20X mais barato/bit que DRAM
  - Tecnologia de disco magnético: 40%/ano
    - 15-25X mais barato/bit que Flash
    - 300-500X mais barato/bit que DRAM
-



---

# Largura de banda e latência

- Largura de banda ou vazão
    - ❑ Total de trabalho realizado em um determinado tempo
    - ❑ 10,000-25,000X de melhoria para processadores
    - ❑ 300-1200X de melhoria para memória e disco
  - Latência ou tempo de resposta
    - ❑ Tempo entre o início e conclusão de um evento
    - ❑ 30-80X de melhoria para processadores
    - ❑ 6-8X de melhoria para memória e disco
-

# Largura de banda e latência

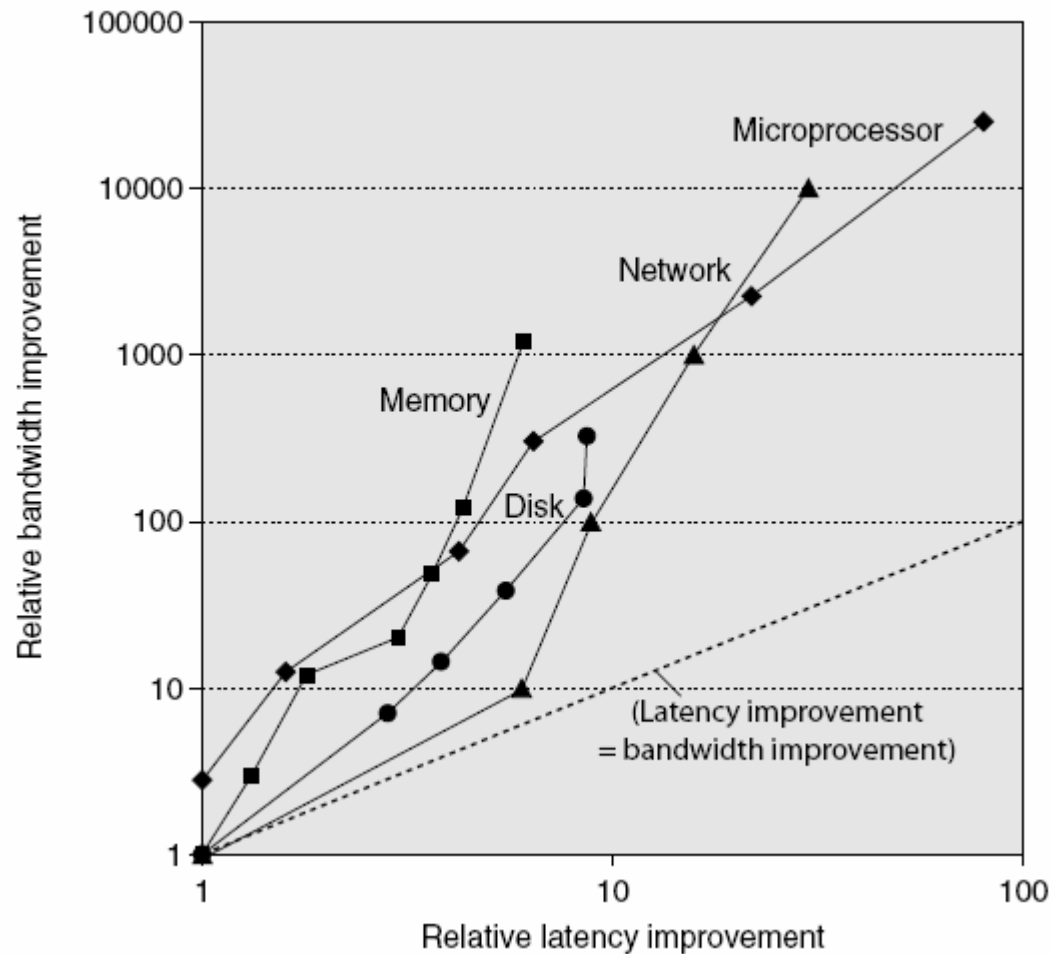


Gráfico log-log de latência e largura de banda

---

# Transistores e conexões

## ■ Tamanhos

- ❑ Tamanho mínimo do transistor ou conexão na dimensão x ou y
  - ❑ 10 microns em 1971 para .032 microns em 2011
  - ❑ Desempenho dos transistores cresce linearmente
    - O atraso nas conexões não melhora com a diminuição do tamanho
  - ❑ Número de transistores aumenta de forma quadrática
-

---

# Potência e energia

- Problema: Máxima potência requerida pelo processador
  - Thermal Design Power (TDP)
    - Caracteriza consumo de energia sustentável
    - Utilizada para determinar valores para geração de potência e sistema de resfriamento
    - Menor que a potência de pico, maior que o consumo médio de potência
  - A taxa do relógio pode ser reduzida dinamicamente para limitar consumo de potência
  - Energia por tarefa em geral é uma medida melhor
-

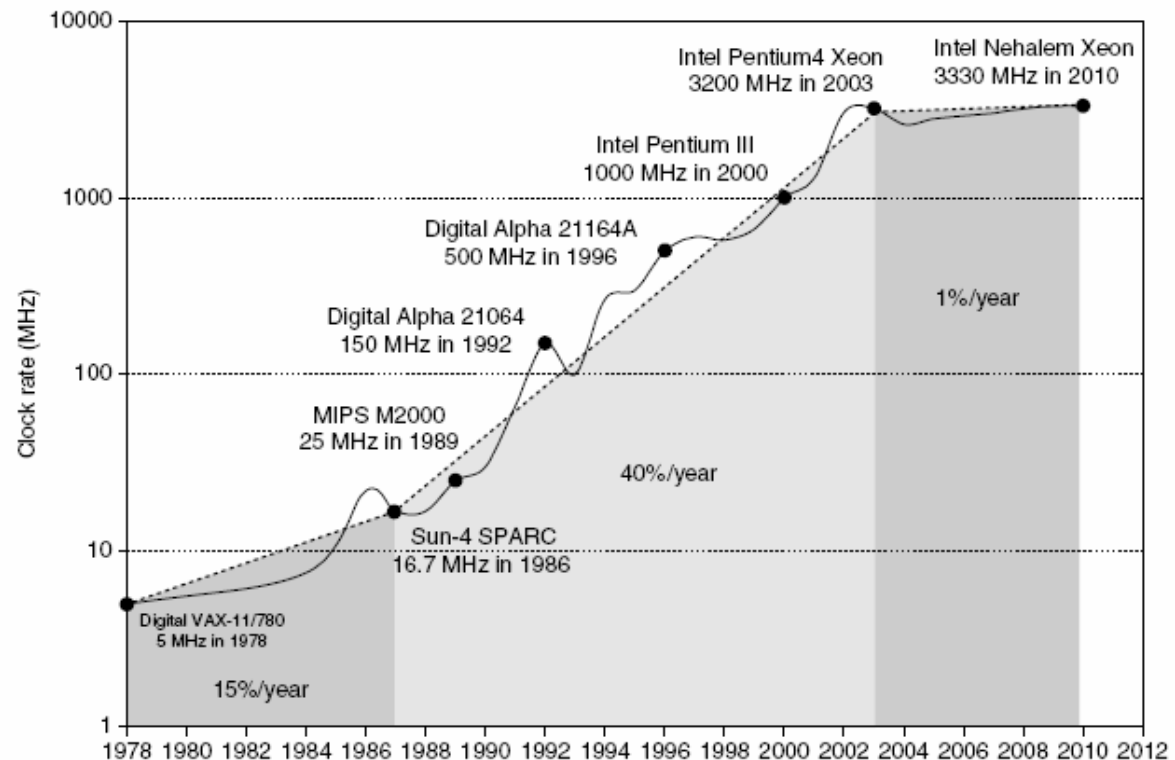
---

# Energia dinâmica e potência

- Energia dinâmica
    - Transistor chaveia de 0 -> 1 ou 1 -> 0
    - Em cada transição a energia é proporcional a  $\frac{1}{2} \times \text{Carga Capacitiva} \times \text{Voltagem}^2$
  - Potência dinâmica
    - $\frac{1}{2} \times \text{Carga capacitiva} \times \text{Voltagem}^2 \times \text{Frequência de transições}$
  - Reduzindo a taxa de relógio reduz a potência, não energia
-

# Energia dinâmica e potência

- Intel 80386 consumia ~ 2 W
- 3.3 GHz Intel Core i7 consome 130 W
- Calor tem que ser dissipado de um chip 1.5 x 1.5 cm
- Este é o limite que pode ser resfriado com ar



---

# Reduzindo a potência

- Desabilitar relógio de módulos inativos
  - Escalonamento dinâmico de voltagem-frequência
  - Colocar DRAMs e discos em modo baixa potência
  - Overclocking, desligando cores
-

---

# Potência estática

- Corrente<sub>estática</sub> x Voltagem
  - Aumenta com número transistores
  - Para reduzir: power gating (desliga a fonte de módulos inativos)
-



---

# Tendências para custo

- Custos caem com o tempo, principalmente para dispositivos que sobrevivem a testes
  - DRAM: preço segue o custo
  - Microprocessadores: preço depende do volume
    - 10% a menos para cada dobro de volume
  - Custos de confecção de circuitos integrados
-