

Definición de Arquitectura

Arquitectura de los Computadores Personales



Luis J. Saiz Adalid



Contenidos sujetos a una Licencia Creative Commons 2.5:
Reconocimiento – No comercial – Sin Obra Derivada

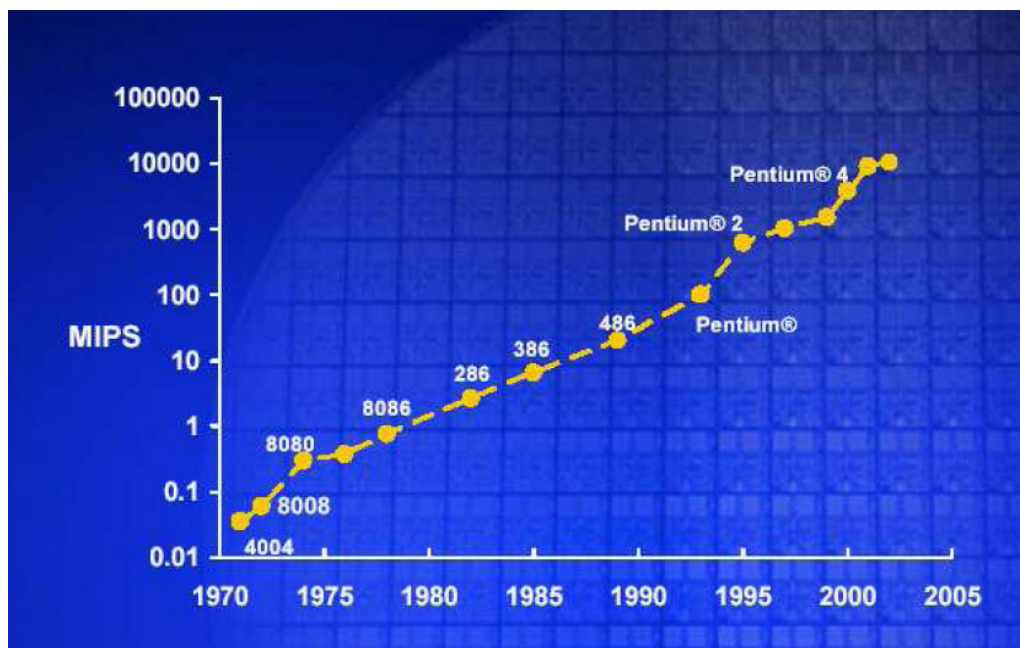


1



- Evolución del rendimiento de los computadores
- Concepto de arquitectura
- Tendencias
- Análisis de prestaciones

- Antes de los 70: 1,3 anual
 - Mejoras tecnológicas
 - Innovaciones en el diseño
- Año 70-80: 1,35 anual
 - Utilización del microprocesador
 - Mejoras tecnológicas de los circuitos integrados
- Año 80-actualidad
 - Arquitecturas RISC (Reduced-Instruction Set Computers)
 - Mejoras arquitectónicas: paralelismo a nivel de instrucciones y caches
 - Mejoras tecnológicas



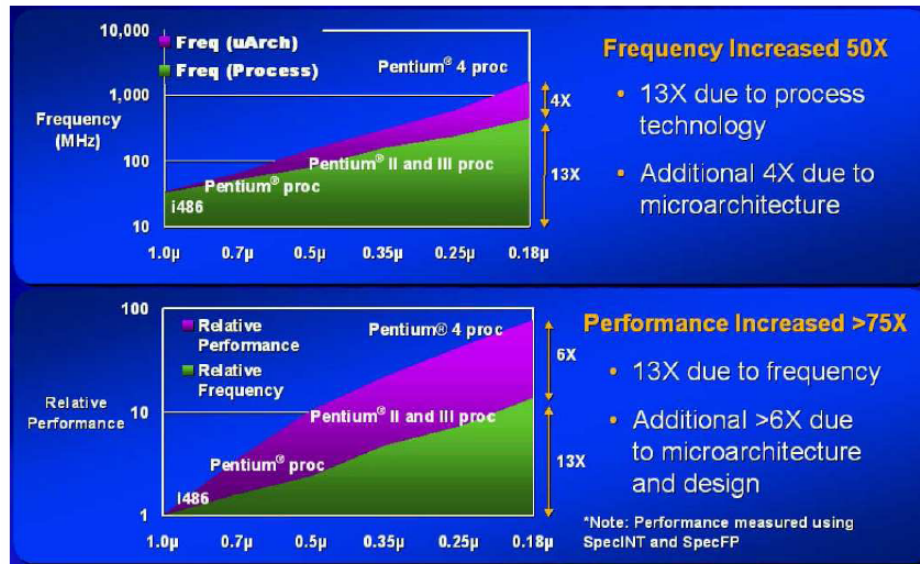
- CISC
 - Complex instruction set computers
 - Computadores con un conjunto de instrucciones complejo
- RISC
 - Reduced instruction set computer
 - Computadores con un conjunto de instrucciones reducido.

- Los términos complejo y reducido, expresan la característica distintiva de cada arquitectura, pero NO deben tomarse sólo como referencia las instrucciones, sino también la complejidad del hardware del procesador
- Con tecnologías de semiconductores comparables e igual frecuencia de reloj
 - Un procesador RISC típico tiene una capacidad de procesamiento de 2 a 4 veces mayor que la de un CISC
 - La estructura del hardware del RISC es tan simple, que se puede realizar en una fracción de la superficie ocupada por el circuito integrado de un procesador CISC

- Aunque los microprocesadores RISCs puedan parecer como sustitutos de los CISC, esto no es tan simple:
 - Para aplicar una determinada arquitectura de microprocesador son decisivas las condiciones de realización técnica y sobre todo la rentabilidad, incluyendo los costos de software
 - Existían y existen razones de compatibilidad para desarrollar y utilizar procesadores de estructura compleja así como un extenso conjunto de instrucciones

- La gran capacidad de integración ha permitido que viejos diseños (ejemplo, x86) hayan integrado muchas de las innovaciones originalmente diseñadas en los RISCs
- En los PCs (arquitecturas compatibles x86), hay un módulo que decodifica instrucciones x86 y las traduce a un repertorio interno de tipo RISC

- El aumento del rendimiento se debe a la combinación de mejoras tecnológicas y arquitectónicas



- Computadores de sobremesa
- Computadores empotrados
- Servidores
- Tarea del diseñador de computadores
 - Determinar los atributos importantes
 - Diseñar el sistema para maximizar el rendimiento
 - Considerar limitaciones de coste y consumo

- Juego de instrucciones
- Organización funcional
- Diseño lógico e implementación

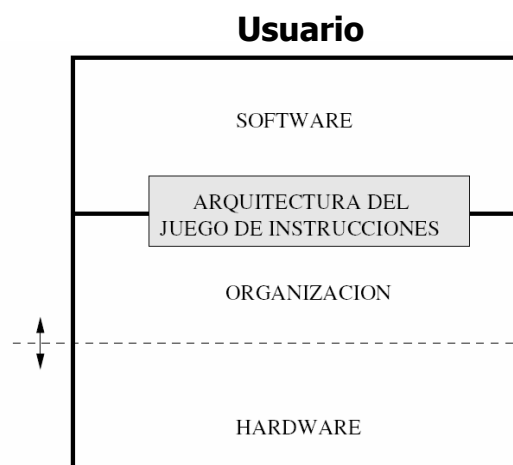
- Por tanto, optimizar el diseño requiere conocer un amplio conjunto de tecnologías:
 - Compiladores, sistemas operativos, diseño de circuitos integrados, consumo de energía, refrigeración, ...

- La arquitectura de un computador comprende tanto el juego de instrucciones como su implementación
- Aspectos que abarca
 - Arquitectura del juego de instrucciones
 - Organización
 - Realización
- Los tres aspectos NO son independientes: las decisiones tomadas en cada uno de ellos pueden influir sobre el resto

- Realiza una descripción del comportamiento del computador, tal como la ven los programadores de sistema
- Sinónimos:
 - ISA (Instruction Set Architecture)
 - Exo-arquitectura
 - Arquitectura externa
 - Arquitectura lógica

- Las instrucciones se ejecutan por medio del *hardware*. La organización de un computador describe el hardware utilizando distintos niveles de abstracción
- Define las capacidades y características de los componentes principales (operadores, subsistema de memoria, ...), su interconexión (buses, multiplexores, ...) y control, para satisfacer las especificaciones de la arquitectura del juego de instrucciones
- Sinónimos: endo-arquitectura, arquitectura interna, arquitectura física, estructura, microarquitectura

- Comprende el diseño lógico detallado y los aspectos específicos de implementación
- Debe satisfacer los requerimientos definidos en la organización



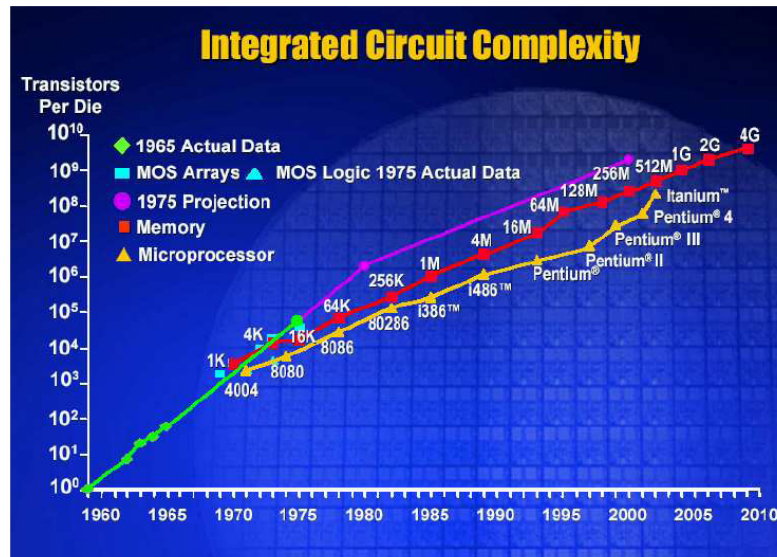
- Requisitos funcionales
- Selección del mejor diseño
- Conocimiento de tendencias en la tecnología y en el uso de los computadores

- **Ámbito funcional**
 - Propósito general, científico, servidor, empotrado, ...
- **Nivel de compatibilidad *software***
 - Código fuente (nuevos compiladores) o binario (no se necesita nuevo software)
- **Requisitos del sistema operativo**
 - Espacio de direccionamiento, protección, gestión de memoria (paginación, segmentación), gestión de procesos (multitarea)
- **Estándares requeridos por el mercado**
 - E/S (P/SATA, PCI/-X, Ultra SCSI, USB, etc.), redes (Ethernet, Infiniband, etc.), coma flotante (IEEE 754), lenguajes, sistemas operativos, ...

- Tres entornos de diseño actuales:
 - Computador de sobremesa: optimizar precio/prestaciones.
 - Servidor: disponibilidad, escalabilidad y productividad
 - Computadores empotrados: alcanzar las prestaciones requeridas con un bajo precio, memoria pequeña, consumo reducido, tiempo real...
- Acciones realizadas:
 - PC sobremesa: desarrollo de técnicas avanzadas en el procesador; gráficos y E/S bien integradas con el procesador.
 - Servidor: multiprocesadores basados en microprocesador; subsistemas de E/S escalables.
 - Empotrado: adaptación de las técnicas empleadas en los procesadores de gama alta para obtener un buen equilibrio entre prestaciones, precio y consumo reducidos.

- Tendencias en tecnología
 - Microprocesadores
 - Memorias DRAM
 - Discos
 - Interfaces e interconexiones
- Tendencias en el uso de los computadores
- Tendencias en el coste

- Ley de Moore: El número de transistores por chip aumenta un 58% cada año (2X cada 18 meses)

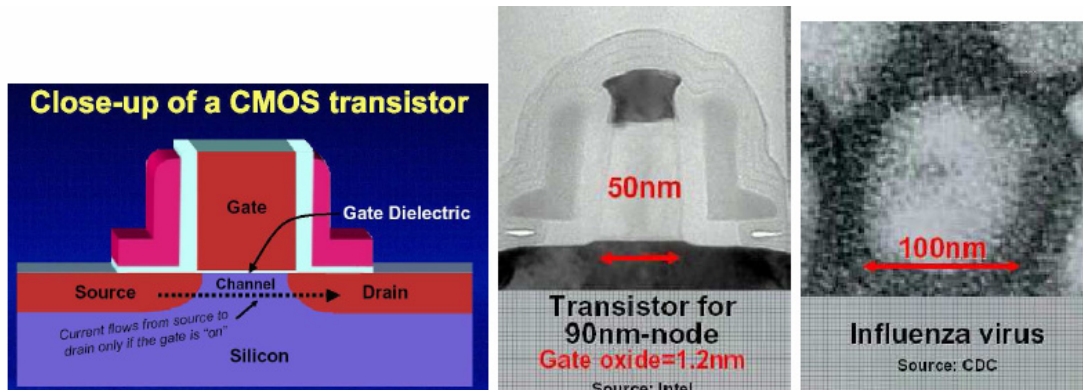


Contenidos sujetos a una Licencia Creative Commons 2.5:
Reconocimiento – No comercial – Sin Obra Derivada



21

- Tamaño del transistor (*Feature size*) más pequeño
 - El n° de transistores crece cuadráticamente con la reducción del tamaño de transistor

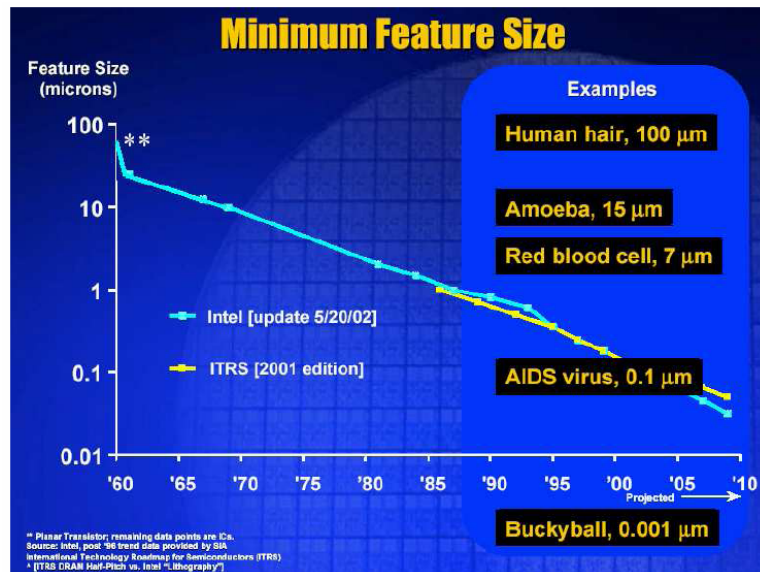


Contenidos sujetos a una Licencia Creative Commons 2.5:
Reconocimiento – No comercial – Sin Obra Derivada

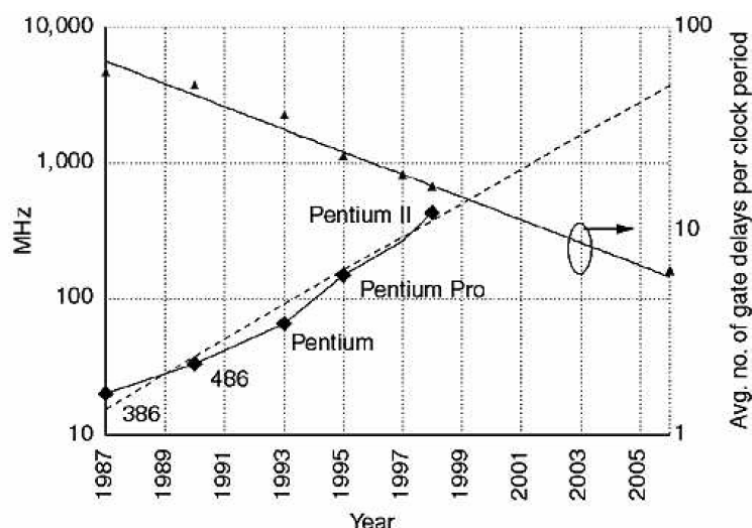


22

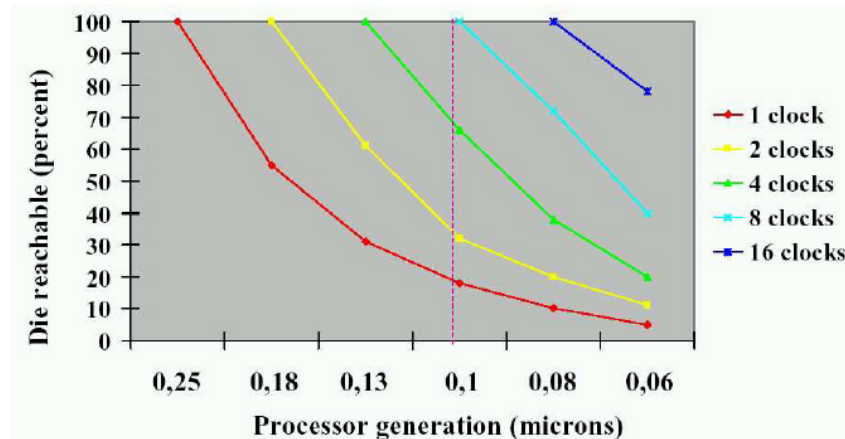
- De $10\ \mu\text{m}$ ($10 \cdot 10^{-6}$) en 1971 a $0.045\ \mu\text{m}$ (45 nm) en 2007



- La velocidad de los transistores aumenta al reducir el su tamaño
 - La frecuencia de reloj se duplica cada dos años (2X en 2 años)
 - También se reduce el número de puertas por ciclo de reloj

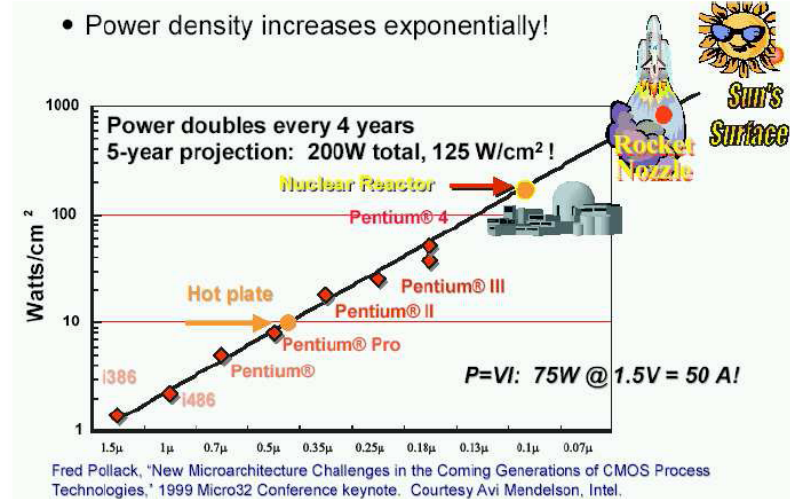


- Aumento del retardo relativo de las interconexiones
 - Retardo depende de R.C del cable → Al reducir el tamaño de transistor se reduce el tamaño del cable, pero aumenta R y C.
- La fracción de chip accesible en un ciclo de reloj se va reduciendo: 100% con 0.25m a 5% con 0.06m.



- Consumo:
 - Energía por transistor $\cong C.f.V^2 + I_{fuga} \cdot V$
- A menor tamaño de transistor, $\uparrow I_{fuga}$
- $T^3 \cdot P = cte$!
 - Reducir a la mitad el tiempo de ejecución requerir 8 veces más de potencia
 - De 0.01 W en los 1^{os} micros a 100 W en P4

- Aspectos importantes:
 - Distribución de la corriente al microprocesador
 - Evacuación del calor generado



Contenidos sujetos a una Licencia Creative Commons 2.5:
Reconocimiento – No comercial – Sin Obra Derivada



27

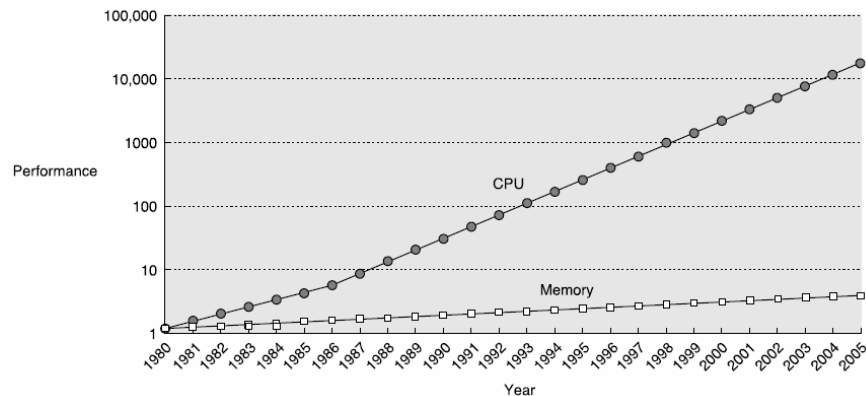
- Tiempo de diseño
 - Aproximadamente 6 años para disponer del diseño lógico más otros 2 años para tenerlo en producción
 - Muchas veces hay que diseñar para la siguiente tecnología disponible

Contenidos sujetos a una Licencia Creative Commons 2.5:
Reconocimiento – No comercial – Sin Obra Derivada



28

- La densidad de integración aumenta en un 60% cada año
- El tiempo de acceso se reduce en un 7% cada año (2X cada 10 años)
- Las mejoras en el interfaz con la memoria han contribuido considerablemente en aumentar el ancho de banda: 1 GB/s (PC133, 1996), 6.4 GB/s (PC6400, 2003), 17 GB/s (PC3-17000, 2008)

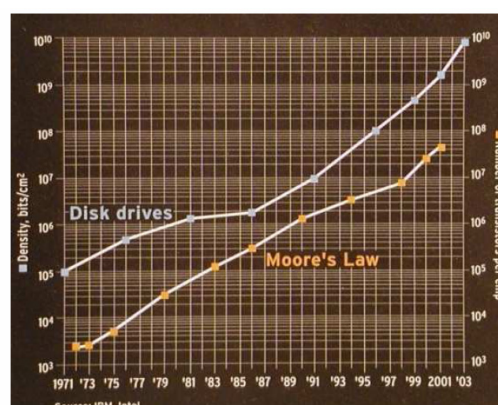


Contenidos sujetos a una Licencia Creative Commons 2.5:
Reconocimiento – No comercial – Sin Obra Derivada

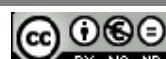


29

- Hasta 1990, la densidad de integración aumentaba en un 30% cada año. Desde entonces, se duplica cada año (2X cada año)
- El tiempo de acceso se reduce en un 30% cada 10 años (1.5X cada 10 años)

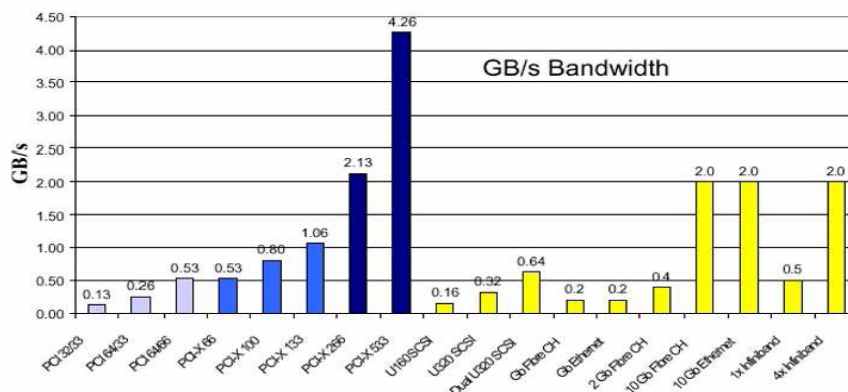


Contenidos sujetos a una Licencia Creative Commons 2.5:
Reconocimiento – No comercial – Sin Obra Derivada



30

- Hasta hace poco tiempo, evolución lenta: 10 años para pasar de Ethernet 10 a 100
- Actualmente, mejora de la tendencia: 4X en 4 años.
- Aparición de nuevas tecnologías de interconexión de altas prestaciones



Contenidos sujetos a una Licencia Creative Commons 2.5:
Reconocimiento – No comercial – Sin Obra Derivada



31

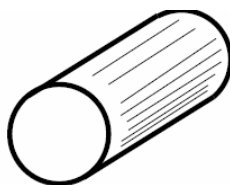
- Determinadas arquitecturas demandan cierta tecnología mientras que la tecnología disponible en un momento dado puede dictar ciertas decisiones arquitectónicas
- Ejemplos:
 - Microprocesador de 32 bits. Cuando la tecnología MOS fue capaz de albergar de 25.000 a 50.000 transistores en un solo chip a principios de los 80
 - Caches. La diferencia de velocidad entre el procesador y la memoria motivó la aparición de las memorias cache. El continuo crecimiento de esta diferencia ha sugerido el empleo de varios niveles (L1, L2, L3) de cache en los diseños

- Etapas para propagar señales. El retardo creciente de los cables motivó la inclusión en el Pentium 4 de dos etapas de segmentación destinadas únicamente a la propagación de la señales a través de los cables.
- Multiprocesadores en un chip. Un procesador sofisticado que funciona a mucha frecuencia consume mucho. En su lugar, podemos poner varios procesadores más sencillos que funcionan a menor frecuencia.

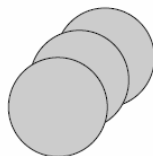
- Una arquitectura que se tarda n años en ser diseñada debe tener al menos un tiempo de vida de n años

- Requisitos de memoria de las aplicaciones: cada año, los programas y sus datos necesitan de 1.5 a 2 veces más memoria → consumo de 1 bit del bus de direcciones por año
- Lenguajes: los lenguajes de alto nivel han sustituido al lenguaje ensamblador para la programación de la máquina → El compilador es el usuario principal de la máquina

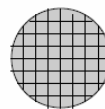
- Coste de un circuito integrado



BARRA DE SILICIO



OBLEA



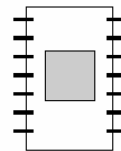
DADOS



DADO
(defectos)



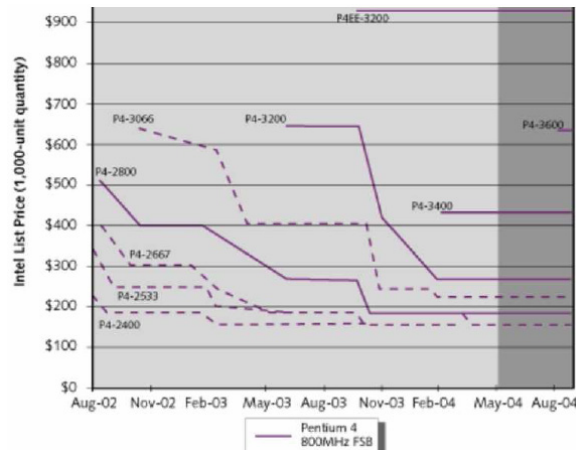
DADO



CIRCUITO INTEGRADO

- Coste final $\cong f$ (superficie dado⁴)
- Superficie dado = f (complejidad del diseño)

- Curva de “aprendizaje”
 - El coste de un componente disminuye con el tiempo, al aumentar la productividad (disminuir la tasa de componentes defectuosos)

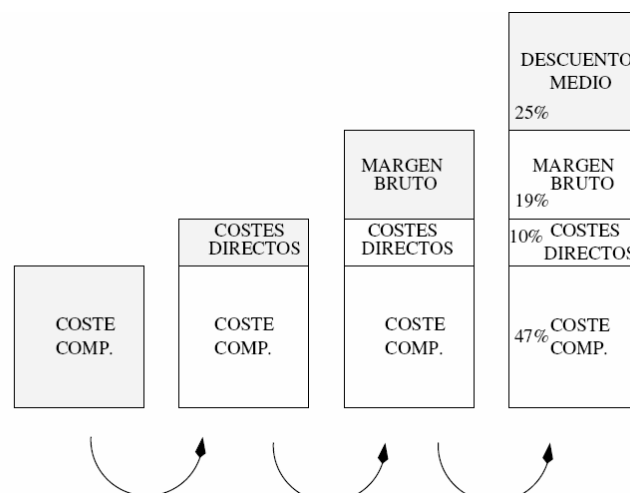


Contenidos sujetos a una Licencia Creative Commons 2.5:
Reconocimiento – No comercial – Sin Obra Derivada



37

- Utilización de componentes “universales”: DRAMs, discos, monitores
 - La competencia baja los precios
- Coste frente a precio



Contenidos sujetos a una Licencia Creative Commons 2.5:
Reconocimiento – No comercial – Sin Obra Derivada



38

- Vídeo:
 - The 5Ghz project
(http://www.tomshardware.com/2003/12/30/5_ghz_project)
- **Actividad 2:** acceder al enlace, leer las explicaciones y descargar y visualizar el vídeo

- Puntos de vista:
 - Usuario: quiere acabar pronto (▼ tiempo de ejecución)
 - Administrador: mucho trabajo hecho (▲ productividad)
 - Relación: $\text{Productividad} = 1 / \text{Tiempo de ejecución}$
- ¿Por qué medir las prestaciones y el coste?
 - Diseñadores: seleccionar cuantitativamente entre alternativas
 - Usuarios: ayuda a elegir

- “El computador que realice la misma cantidad de trabajo en el menor tiempo es el más rápido”
- Enfoque actual: utilizar directamente el tiempo de ejecución de los programas como medida de las prestaciones, indicando cuáles son esos programas.
 - Programas reales, benchmarks, ...
- Otras medidas: MIPS, MFLOPS

- Ecuación del tiempo de ejecución:

$$T_{\text{ejecucion}} = \frac{\text{seg}}{\text{programa}} = \frac{\text{num. instr.}}{\text{programa}} \times \frac{\text{ciclos}}{\text{num. instr.}} \times \frac{\text{seg}}{\text{ciclo}} = I * CPI * T$$

- Los tres parámetros están relacionados:
 - $I = f(\text{arq. del juego de instrucciones, compilador})$
 - $CPI = f(\text{arq. del juego de instrucciones, organización})$
 - $T = f(\text{tecnología, organización})$

- Hacer rápido el caso frecuente: acelerar las partes que se usan con mayor frecuencia
- Ley de Amdhal: ¿cómo afecta una mejora parcial en un sistema a su funcionamiento global?

$$S' = \frac{t}{t'} = \frac{1}{(1 - F) + \frac{F}{S}}$$

- **Actividad 3:** estudiar y completar el boletín