

Unidad I: Arquitecturas de cómputo

1.1 Modelos de arquitecturas de cómputo

1.1.1 Clásicas

Estas arquitecturas se desarrollaron en las primeras computadoras electromecánicas y de tubos de vacío. Aun son usadas en procesadores empotrados de gama baja y son la base de la mayoría de las arquitecturas modernas

Arquitectura Mauchly-Eckert (Von Newman)

La principal desventaja de esta arquitectura, es que el bus de datos y direcciones único se convierte en un cuello de botella por el cual debe pasar toda la información que se lee de o se escribe a la memoria, obligando a que todos los accesos a esta sean secuenciales. Esto limita el grado de paralelismo (acciones que se pueden realizar al mismo tiempo) y por lo tanto, el desempeño de la computadora. Este efecto se conoce como el cuello de botella de Von Newman

1.1.2 Segmentadas

Las arquitecturas segmentadas o con segmentación del cauce buscan mejorar el desempeño realizando paralelamente varias etapas del ciclo de instrucción al mismo tiempo. El procesador se divide en varias unidades funcionales independientes y se dividen entre ellas el procesamiento de las instrucciones.

1.1.3 De multiprocesamiento

Cuando se desea incrementar el desempeño más allá de lo que permite la técnica de segmentación del cauce (límite teórico de una instrucción por ciclo de reloj), se requiere utilizar más de un procesador para la ejecución del programa de aplicación.

Las CPU de multiprocesamiento:

SISO – (Single Instruction, Single Operand) computadoras independientes

SIMO – (Single Instruction, Multiple Operand) procesadores vectoriales

MISO – (Multiple Instruction, Single Operand) No implementado

MIMO – (Multiple Instruction, Multiple Operand) sistemas SMP, Clusters

1.2 Análisis de los componentes

1.2.1 CPU

La unidad central de procesamiento, UCP o CPU (por el acrónimo en inglés de central processing unit), o simplemente el procesador o microprocesador, es el componente del computador y otros dispositivos programables, que interpreta las instrucciones contenidas en los programas y procesa los datos. Los CPU proporcionan la característica fundamental de la computadora digital (la programabilidad) y son uno de los componentes necesarios encontrados en las computadoras de cualquier tiempo, junto con el almacenamiento primario y los dispositivos de entrada/salida. Se conoce como microprocesador el CPU que es manufacturado con circuitos integrados. Desde mediados de los años 1970, los microprocesadores de un solo chip han reemplazado casi totalmente todos los tipos de CPU, y hoy en día, el término "CPU" es aplicado usualmente a todos los microprocesadores.

La expresión "unidad central de proceso" es, en términos generales, una descripción de una cierta clase de máquinas de lógica que pueden ejecutar complejos programas de computadora. Esta amplia definición puede fácilmente ser aplicada a muchos de los primeros computadores que existieron mucho antes que el término "CPU" estuviera en amplio uso. Sin embargo, el término en sí mismo y su acrónimo han estado en uso en la industria de la informática por lo menos desde el principio de los años 1960. La forma, el diseño y la implementación de los CPU ha cambiado drásticamente desde los primeros ejemplos, pero su operación fundamental ha permanecido bastante similar.

Los primeros CPU fueron diseñados a la medida como parte de una computadora más grande, generalmente una computadora única en su especie. Sin embargo, este costoso método de diseñar los CPU a la medida, para una aplicación particular, ha desaparecido en gran parte y se ha sustituido por el desarrollo de clases de procesadores baratos y estandarizados adaptados para uno o muchos propósitos. Esta tendencia de estandarización comenzó generalmente en la era de los transistores discretos, computadoras centrales, y microcomputadoras, y fue acelerada rápidamente con la popularización del circuito integrado (IC), éste ha permitido que sean diseñados y fabricados CPU más complejos en espacios pequeños (en la orden de milímetros). Tanto la miniaturización como la estandarización de los CPU han aumentado la presencia de estos dispositivos digitales en la vida moderna mucho más allá de las aplicaciones limitadas de máquinas de computación dedicadas. Los microprocesadores modernos aparecen en todo, desde automóviles, televisores, neveras, calculadoras, aviones, hasta teléfonos móviles o celulares, juguetes, entre otros.

1.2.1.1 Arquitecturas

Tanto la miniaturización como la estandarización de los CPU han aumentado la presencia de estos dispositivos digitales en la vida moderna mucho más allá de las aplicaciones limitadas de máquinas de computación dedicadas. Los microprocesadores modernos aparecen en todo, desde automóviles, televisores, neveras, calculadoras, aviones, hasta teléfonos móviles o celulares, juguetes, entre otros.

1.2.1.2 Tipos

Los primeros CPU fueron diseñados a la medida como parte de una computadora más grande, generalmente una computadora única en su especie. Sin embargo, este costoso método de diseñar los CPU a la medida, para una aplicación particular, ha desaparecido en gran parte y se ha sustituido por el desarrollo de

clases de procesadores baratos y estandarizados adaptados para uno o muchos propósitos. Esta tendencia de estandarización comenzó generalmente en la era de los transistores discretos, computadoras centrales, y microcomputadoras, y fue acelerada rápidamente con la popularización del circuito integrado (IC), éste ha permitido que sean diseñados y fabricados CPU más complejos en espacios pequeños (en la orden de milímetros).

1.2.1.3 Características

Desde mediados de los años 1970, los microprocesadores de un solo chip han reemplazado casi totalmente todos los tipos de CPU, y hoy en día, el término "CPU" es aplicado usualmente a todos los microprocesadores.

1.2.2 Memoria

En informática, la memoria (también llamada almacenamiento) se refiere a parte de los componentes que forman parte de una computadora. Son dispositivos que retienen datos informáticos durante algún intervalo de tiempo. Las memorias de computadora proporcionan una de las principales funciones de la computación moderna, la retención o almacenamiento de información. Es uno de los componentes fundamentales de todas las computadoras modernas que, acoplados a una unidad central de procesamiento (CPU por su sigla en inglés, central processing unit), implementa lo fundamental del modelo de computadora de Arquitectura de von Neumann, usado desde los años 1940.

Dispositivo basado en circuitos que posibilitan el almacenamiento limitado de información y su posterior recuperación.

Las memorias suelen ser de rápido acceso, y pueden ser volátiles o no volátiles.

La clasificación principal de memorias son RAM y ROM. Estas memorias son utilizadas para almacenamiento primario.

1.2.2.1 Conceptos básicos del manejo de la memoria

1.2.2.2 Memoria principal semiconductor.

1.2.2.3 Memoria cache

1.2.3 Manejo de la entrada/salida

En computación, **entrada/salida**, también abreviado E/S o I/O (del original en inglés input/output), es la colección de interfaces que usan las distintas unidades funcionales (subsistemas) de un sistema de procesamiento de información para comunicarse unas con otras, o las señales (información) enviadas a través de esas interfaces. Las entradas son las señales recibidas por la unidad, mientras que las salidas son las señales enviadas por ésta.

El término puede ser usado para describir una acción; "realizar una entrada/salida" se refiere a ejecutar una operación de entrada o de salida. Los dispositivos de E/S los usa una persona u otro sistema para comunicarse con una computadora. De hecho, a los teclados y ratones se los considera dispositivos de entrada de una computadora, mientras que los monitores e impresoras son vistos como dispositivos de salida de una computadora. Los dispositivos típicos para la comunicación entre computadoras realizan las dos operaciones, tanto entrada como salida, y entre otros se encuentran los módems y tarjetas de red.

1.2.3.1 Módulos de entrada/salida

En primer termino hablaremos de los dispositivos de entrada, que como su nombre lo indica, sirven para introducir datos (información) a la computadora para su proceso. Los datos se leen de los dispositivos de entrada y se almacenan en la memoria central o interna.

Los dispositivos de entrada convierten la información en señales eléctricas que se almacenan en la memoria central. Los dispositivos de entrada típicos son los teclados, otros son: lápices ópticos, palancas de mando (joystick), CD-ROM, discos compactos (CD), etc. Hoy en día es muy frecuente que el usuario utilice un dispositivo de entrada llamado ratón que mueve un puntero electrónico sobre una pantalla que facilita la interacción usuario-máquina.

1.2.3.2 Entrada/salida programada

Los dispositivos de Entrada y Salida permiten la comunicación entre la computadora y el usuario.

1.2.3.3 Entrada/salida mediante interrupciones

El problema con E/S programada es que el CPU tiene que esperar un tiempo considerable a que el módulo de E/S en cuestión esté preparado para recibir o transmitir los datos. El CPU debe estar comprobando continuamente el estado del módulo de E/S. Se degrada el desempeño del sistema.

Una alternativa es que el CPU tras enviar una orden de E/S continúe realizando algún trabajo útil. El módulo de E/S interrumpirá al CPU para solicitar su servicio cuando esté preparado para intercambiar datos. El CPU ejecuta la transferencia de datos y después continúa con el procesamiento previo.

Se pueden distinguir dos tipos: E/S síncrona y E/S asíncrona

E/S Síncrona: cuando la operación de E/S finaliza, el control es retornado al proceso que la generó. La espera por E/S se lleva a cabo por medio de una instrucción wait que coloca al CPU en un estado ocioso hasta que ocurre otra interrupción. Aquellas máquinas que no tienen esta instrucción utilizan un loop. Este loop continua hasta que ocurre una interrupción transfiriendo el control a otra parte del sistema de operación. Sólo se atiende una solicitud de E/S por vez. El sistema de operación conoce exactamente que dispositivo está interrumpiendo. Esta alternativa excluye procesamiento simultáneo de E/S.

E/S Asíncrona: retorna al programa usuario sin esperar que la operación de E/S finalice. Se necesita una llamada al sistema que le permita al usuario esperar por la finalización de E/S (si es requerido). También es necesario llevar un control de las distintas solicitudes de E/S. Para ello el sistema de operación utiliza una tabla que contiene una entrada por cada dispositivo de E/S (Tabla de Estado de Dispositivos).

La ventaja de este tipo de E/S es el incremento de la eficiencia del sistema. Mientras se lleva a cabo E/S, el CPU puede ser usado para procesar o para

planificar otras E/S. Como la E/S puede ser bastante lenta comparada con la velocidad del CPU, el sistema hace un mejor uso de las facilidades.

1.2.3.4 Acceso directo a memoria

El acceso directo a memoria (DMA, del inglés *direct memory access*) permite a cierto tipo de componentes de una computadora acceder a la memoria del sistema para leer o escribir independientemente de la unidad central de procesamiento (CPU) principal. Muchos sistemas hardware utilizan DMA, incluyendo controladores de unidades de disco, tarjetas gráficas y tarjetas de sonido. DMA es una característica esencial en todos los ordenadores modernos, ya que permite a dispositivos de diferentes velocidades comunicarse sin someter a la CPU a una carga masiva de interrupciones.

Una transferencia DMA consiste principalmente en copiar un bloque de memoria de un dispositivo a otro. En lugar de que la CPU inicie la transferencia, la transferencia se lleva a cabo por el controlador DMA. Un ejemplo típico es mover un bloque de memoria desde una memoria externa a una interna más rápida. Tal operación no ocupa al procesador y, por ende, éste puede efectuar otras tareas. Las transferencias DMA son esenciales para aumentar el rendimiento de aplicaciones que requieran muchos recursos.

Cabe destacar que aunque no se necesite a la CPU para la transacción de datos, sí se necesita el bus del sistema (tanto bus de datos como bus de direcciones), por lo que existen diferentes estrategias para regular su uso, permitiendo así que no quede totalmente acaparado por el controlador DMA.

El acceso directo a memoria es simplemente eso, un acceso a memoria que se crea al particionar la memoria en bloques del mismo tamaño.

1.2.3.5 Canales y procesadores de entrada/salida

En las entradas y salidas que se ejecutan desde el procesador entran tres conceptos que se deben conocer.

El DMA

Los Canales

Y las instrucciones

DMA

Con un incremento moderado de la lógica asociada con el periférico se puede lograr transferir un bloque de información a o desde la memoria principal sin la intervención directa de la UCP. Esto requiere que los periféricos o los controladores de E/S sean capaces de generar direcciones de memoria y transferir datos a o desde el bus del sistema. La UCP es todavía la responsable de iniciar la transferencia de cada bloque. A partir de este punto el controlador de E/S realiza toda la operación sin que la UCP tenga que ejecutar ningún otro programa. La UCP y el controlador de E/S interaccionan sólo cuando la UCP debe dar el control del bus del sistema al controlador de E/S, como respuesta a una petición de este último. Este tipo de capacidad es lo que se conoce como Acceso Directo a Memoria (DMA).

La mayoría de los computadores de hoy día tienen la posibilidad de interrupción y de DMA. Un controlador de DMA posee un control parcial de la operación de E/S. La UCP puede liberarse totalmente si se introduce en el computador un Procesador de E/S (PE/S) o canal. Análogamente al DMA, el PE/S tiene acceso a la memoria principal y puede interrumpir a la UCP, sin embargo puede emplear un repertorio de instrucciones diferentes del de la UCP (ya que está orientado a operaciones de E/S).

CANALES

El canal de E/S es una extensión del concepto de DMA. Un canal de E/S tiene la capacidad de ejecutar instrucciones de E/S, lo que da un control total sobre las operaciones de E/S.

Las instrucciones de E/S se almacenan en la memoria principal y serán ejecutadas por un procesador de propósito específico en el mismo canal de E/S.

1.2.4 Buses

En arquitectura de computadores, el bus (o canal) es un sistema digital que transfiere datos entre los componentes de una computadora o entre computadoras. Está formado por cables o pistas en un circuito impreso, dispositivos como resistores y condensadores además de circuitos integrados.

En los primeros computadores electrónicos, todos los buses eran de tipo paralelo, de manera que la comunicación entre las partes del computador se hacía por medio de cintas o muchas pistas en el circuito impreso, en los cuales cada conductor tiene una función fija y la conexión es sencilla requiriendo únicamente puertos de entrada y de salida para cada dispositivo.

La tendencia en los últimos años se hacia uso de buses seriales como el USB, Firewire para comunicaciones con periféricos reemplazando los buses paralelos, incluyendo el caso como el del microprocesador con el chipset en la placa base. Esto a pesar de que el bus serial posee una lógica compleja (requiriendo mayor poder de cómputo que el bus paralelo) a cambio de velocidades y eficacias mayores.

Existen diversas especificaciones de que un bus se define en un conjunto de características mecánicas como conectores, cables y tarjetas, además de protocolos eléctricos y de señales.

1.2.4.1 Tipos de buses

Bus paralelo

Es un bus en el cual los datos son enviados por bytes al mismo tiempo, con la ayuda de varias líneas que tienen funciones fijas. La cantidad de datos enviada es bastante grande con una frecuencia moderada y es igual al ancho de los datos por la frecuencia de funcionamiento. En los computadores ha sido usado de manera intensiva, desde el bus del procesador, los buses de discos duros, tarjetas de expansión y de vídeo, hasta las impresoras.

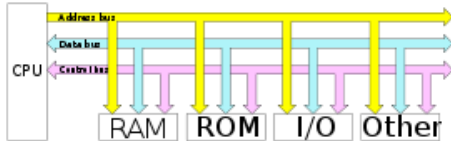


Diagrama de un Bus Backplane como extensión del bus de procesador.

El front-side bus de los procesadores Intel es un bus de este tipo y como cualquier bus presenta unas funciones en líneas dedicadas:

Las líneas de dirección son las encargadas de indicar la posición de memoria o el dispositivo con el que se desea establecer comunicación.

Las líneas de control son las encargadas de enviar señales de arbitraje entre los dispositivos. Entre las más importantes están las líneas de interrupción, DMA y los indicadores de estado.

Las líneas de datos transmiten los bits de forma aleatoria de manera que por lo general un bus tiene un ancho que es potencia de 2.

Un bus paralelo tiene conexiones físicas complejas, pero la lógica es sencilla, que lo hace útil en sistemas con poco poder de cómputo. En los primeros microcomputadores, el bus era simplemente la extensión del bus del procesador y los demás integrados "escuchan" la línea de direcciones, en espera de recibir instrucciones. En el PC IBM original, el diseño del bus fue determinante a la hora de elegir un procesador con I/O de 8 bits (Intel 8088), sobre uno de 16 (el 8086), porque era posible usar hardware diseñado para otros procesadores, abaratando el producto.

Bus serial

En este los datos son enviados, bit a bit y se reconstruyen por medio de registros o rutinas de software. Está formado por pocos conductores y su ancho de banda depende de la frecuencia. Es usado desde hace menos de 10 años en buses para discos duros, unidades de estado sólido, tarjetas de expansión y para el bus del procesador.

1.2.4.2 Estructura de los buses

Estructura de un bus:

- Líneas de datos:
- Líneas de dirección
- Líneas de control

Datos:

- Llevan datos y también comandos para los dispositivos de entrada / salida.
- Su amplitud influye en el rendimiento del bus
- Relacionado con el tamaño de palabra del sistema

Direcciones:

- Llevan direcciones de memoria en acceso a memoria, o permiten seleccionar un dispositivo conectado al bus.
- Su amplitud determina el espacio de direcciones tanto de memoria como entrada / salida

Control:

- Son señales de control de acceso y uso del bus.
- Arbitraje del bus
- Sincronización de las comunicaciones
- Reloj del sistema

1.2.4.3 Jerarquías de buses

Jerarquía

Tendencia a utilizar múltiples buses debido a que:

- degradación del rendimiento a medida que aumenta el número de dispositivos colgados al bus.
- "Cuello de botella" de las comunicaciones debido a los tiempos de espera para acceder al bus.

Por ellos se introducen múltiples buses basando la jerarquía en los requerimientos de comunicación de los dispositivos:

- Bus local: Conecta al procesador con la cache y con algún dispositivo e/s muy rápido.
- Bus del sistema: Conecta al procesador (a través de la cache) con la memoria del sistema y con un segundo nivel de dispositivos de velocidad media.
- Bus de expansión: Se conecta al bus del sistema y hace de interfaz entre este y los dispositivos más lentos

1.2.5. Interrupciones

En el contexto de la informática, una interrupción (del inglés Interrupt Request, también conocida como petición de interrupción) es una señal recibida por el procesador de un ordenador, indicando que debe "interrumpir" el curso de ejecución actual y pasar a ejecutar código específico para tratar esta situación.

Una interrupción es una suspensión temporal de la ejecución de un proceso, para pasar a ejecutar una subrutina de servicio de interrupción, la cual, por lo general, no forma parte del programa, sino que pertenece al sistema operativo o al BIOS). Una vez finalizada dicha subrutina, se reanuda la ejecución del programa.

Las interrupciones surgen de la necesidad que tienen los dispositivos periféricos de enviar información al procesador principal de un sistema informático.

La primera técnica que se empleó para esto fue el polling, que consistía en que el propio procesador se encargara de sondear los dispositivos periféricos cada cierto tiempo para averiguar si tenía pendiente alguna comunicación para él. Este método presentaba el inconveniente de ser muy ineficiente, ya que el procesador consumía constantemente tiempo y recursos en realizar estas instrucciones de sondeo.

El mecanismo de interrupciones fue la solución que permitió al procesador desentenderse de esta problemática, y delegar en el dispositivo periférico la responsabilidad de comunicarse con él cuando lo necesitara. El procesador, en este caso, no sondea a ningún dispositivo, sino que queda a la espera de que estos le avisen (le "interrumpan") cuando tengan algo que comunicarle (ya sea un evento, una transferencia de información, una condición de error, etc.).