

ANÁLISIS DE ARQUITECTURAS MODERNAS DE MICROPROCESADORES UTILIZADOS EN CIRCUITOS INTEGRADOS (CISC, RISC, VLIW, EPIC), ASÍ COMO ARQUITECTURAS PARA EL PROCESAMIENTO MULTIMEDIA Y DIGITAL DE SEÑALES – DSP

La arquitectura de un microprocesador define tanto sus características y el modo en que está programado. La arquitectura especifica únicamente la manera en que se ejecutan las instrucciones, dadas por el procesador, el número y la organización de los registros de datos (es decir, las áreas de almacenamiento para datos durante la ejecución de la operación), y la configuración de los terminales de entrada y salida (es decir, los canales físicos a través de los cuales los datos son transferidos desde una parte del computador a otra).

Cuando la arquitectura ya esta definida, todas las características del procesador quedan fijadas (lo que conduce a definir el lenguaje de programación que corresponda a dicha arquitectura, donde cada bit de la instrucción corresponde a una disponibilidad material de un elemento lógico)

TIPOS BÁSICOS DE ARQUITECTURAS DE MICROPROCESADORES

Hoy podemos distinguir entre cuatro tipos de arquitecturas: arquitectura CISC (Complex Instruction Set Computer) o ejecución de una instrucción compleja, que forma la base de todos los procesadores x86 o compatibles con intel, arquitectura RISC (Reduced Instruction Set Computer) o que ejecuta un número limitado de instrucciones, arquitectura VLIW (Very Long Instruction Word), y la más cercana EPIC (Explicit Parallel Instruction Computing) o una máquina completamente2013/ paralela. Cabe señalar que algunos productos del mercado pueden combinar un número de estos modelos de arquitecturas.

La arquitectura CISC constituye un acercamiento histórico al concepto de microprocesador. Está todavía presente en la serie x86 (es decir, los procesadores fabricados por Intel, AMD o Cyrix), nacido junto con el Intel 8086 en 1978. En esa época la memoria del sistema era un bien escaso, costoso y lento (los mayores sistemas sólo tenían unos pocos megabytes de memoria, y los procesadores de los ordenadores personales unos pocos kilobytes). La arquitectura CISC fue diseñada para reducir el uso de estos recursos. Sin embargo, el número de transistores es incomparable con el actual número. El 8086 cuenta con 29000 transistores (el Pentium 4 HT tiene casi 6000 veces más, unos 167 millones). Entonces Intel eligió como solución de compromiso, que un cuarto de siglo más tarde, no cambiase profundamente la operación de los procesadores de esta larga serie: un pequeño número de registros, instrucciones con logitud variable, que son muy complejas y requieren un número considerable de ciclos para su ejecución por el procesador. Los procesadores, aprovechando la arquitectura CISC pueden procesar instrucciones complejas, directamente “grabadas” en sus circuitos electrónicos. Esto significa que algunas instrucciones, difíciles de cumplir por un microprograma con instrucciones básicas, se realizan por medio de hardware, es decir, son directamente programadas en el prpio chip y pueden ser llamadas por macro comandos con el fin de ganar velocidad en la ejecución.

Un microprocesador a nivel físico usa un limitado número de instrucciones microprogramadas (logica AND, lógica OR, cambio de registros), fabricados en silicio.

MP parallelism

Subscalar model

$$T_a = N(k \cdot \tau)$$

Instruction level

Parallelism

(pipeline)

$$T_b = [k + (N-1)] \cdot \tau$$

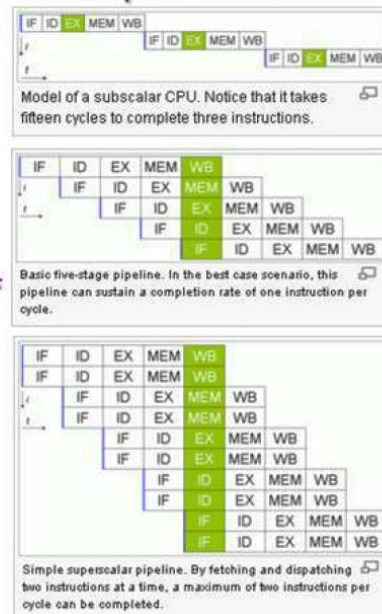
Simple

Superscalar

Model

(2 pipelines)

$$T_c \cong T_b/2$$



N - number of tasks (instructions)

k - number of phases

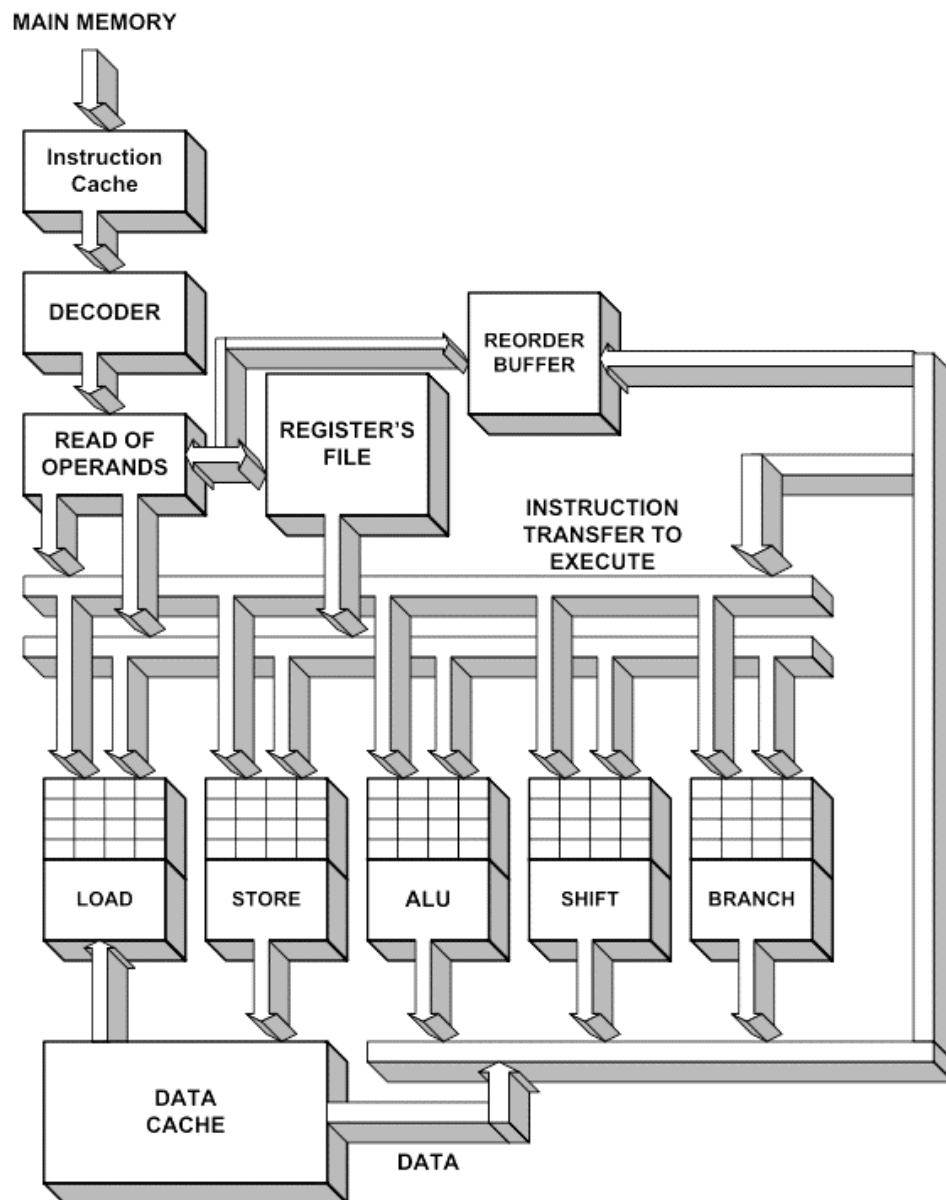
τ - execution time per one phase

Figura 1: 3 modelos de MP y el tiempo de ejecución

No es posible desarrollar todas las instrucciones máquina en el procesador. Actualmente el procesador ejecuta las instrucciones máquina, que constituyen un lenguaje de comunicación con el microprocesador como una secuencia de instrucciones microprogramadas.

A diferencia de la arquitectura CISC, un procesador basado en tecnología RISC no tiene incorporadas funciones avanzadas. La arquitectura, creada por John Cocke en 1960 recomienda acortar y simplificar las instrucciones con igual longitud y permitiendo una rápida ejecución, es decir, la realización en un ciclo de reloj. Las instrucciones en este caso están programadas de una manera síncrona en el bus de datos. El bus de datos funciona como una cinta transportadora de una planta, o como un músico, que en cada unidad de tiempo interpreta una sola nota de su partitura: a cada unidad de tiempo le corresponde una parte elemental, muy simple de ejecutar: buscar la instrucción en el registro, decodificación, es decir, interpretación de lo que se ha hecho, ejecutar y guardar los resultados (los modernos procesadores Pentium trabajan 14 ciclos de reloj).

Desde que la ejecución de las instrucciones es idéntica, es posible que estas instrucciones sean ejecutadas en paralelo, agrupando unos pocos bus de datos en una arquitectura común. Aquí hablamos sobre arquitecturas a gran escala (ver la figura). El bus de datos elemental tiene al menos dos niveles y opera a 4 ciclos de reloj. Así, el procesador ejecuta más instrucciones simultáneamente por cada ciclo de tiempo. Cada nivel puede intercambiar su resultado con otro nivel al final del ciclo de tiempo, y viceversa. Los programas se ejecutan más rápido. Cabe destacar que el procesador puede reorganizar las instrucciones en el camino durante el proceso, a fin de optimizar la secuencia: el procesador estudia el flujo de instrucciones de entrada, decide reorganizar el paquete de instrucciones de una manera óptima, teniendo en cuenta unas simples reglas, comprueba si la nueva secuencia no depende de una instrucción anterior. Llamamos a esto ejecución dinámica, que está presente en los modernos y complejos procesadores (ver las figuras siguientes).



El diagrama de bloques siguiente de un procesador Pentium muestra dos caminos de instrucciones, el camino U y el camino V. El camino U puede ejecutar todas las instrucciones de números enteros y en coma flotante. El camino V puede ejecutar simplemente instrucciones de números enteros e instrucciones en coma flotante FXCH.

Se muestran las cachés de códigos y datos separadas. La cache de datos tiene dos puertos, uno para cada camino (las etiquetas son triplemente presentadas para permitir ciclos de consulta simultáneos). La caché de datos tiene un TLB (Translation Lookaside Buffer) para traducir direcciones lineales a direcciones físicas usadas por la cache de datos.

La memoria caché de código, el Buffer de destino de datos (Branch Target Buffer) y los Buffer de datos (Prefetch Buffers) son los responsables de recibir instrucciones sin procesar en las unidades de ejecución del procesador Pentium. Las instrucciones se recuperan desde la

memoria caché de código o desde un bus externo. La filas de direcciones son recordadas por el Branch Target Buffer. La memoria caché de código TLB traduce direcciones lineales a direcciones físicas usadas por la memoria caché.

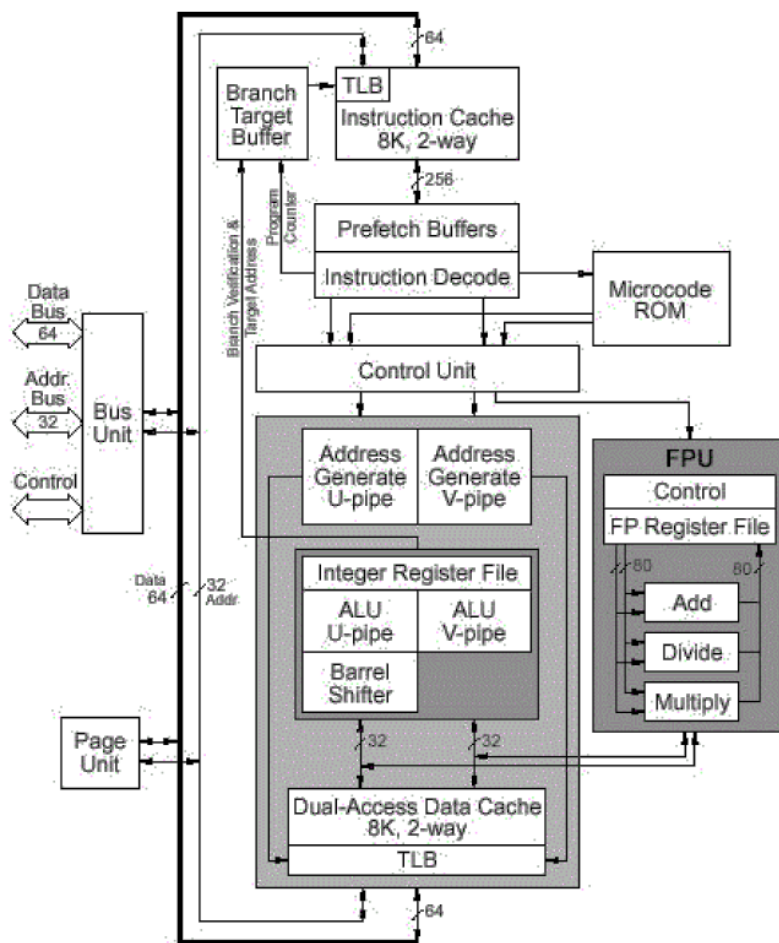


Figura 2: Diagrama de bloques de un procesador Pentium

Los procesadores más nuevos van aún más lejos en esta optimización, aplicando procesos de predicción al desarrollo del programa. De hecho, cuando un programa se encuentra con un ciclo condicional, es decir, una declaración del tipo “if, if not”, el proceso predice la forma en que el programa tendría mayor oportunidad de continuar y organizar la secuencia de instrucciones de esa manera. Tal predicción es, generalmente hablando, simple y basada en la cronología de secuencias. Si el procesador ha hecho una buena elección, es decir, lo que sucede estadísticamente entre el 80 y 90% de las veces, ya estaría por delante de algunas secuencias de instrucciones y habría ganado tiempo durante la ejecución con respecto al desarrollo normal. Si ha cometido un error, volvería sobre sus pasos, vaciando en el bus de datos paralelo la secuencia de instrucciones equivocadas y cargando las nuevas.

El precio a pagar por un gran error es tiempo perdido de ciclos de reloj y la gran longitud de buses utilizados, la más pesada de las penas. Los procesadores, usan buses largos, pueden operar a altas frecuencias, cuando los buses son largos y cuando las operaciones son finamente divididas, es decir, las instrucciones son más simples. Esta es también una forma de compromiso, que los constructores de microprocesadores tienen

que resolver.

Los microprocesadores modernos trabajan a frecuencias del orden de los Ghz, que significa que el procesador busca una nueva instrucción en unos pocos nanosegundos a pesar de que la memoria RAM, contiene el código del programa en ejecución, necesita más de 50 ns para extraer las instrucciones desde las celdas de memoria y transferirlas al procesador.

Con el fin de resolver este problema son utilizadas las memorias cash. Las memorias cash son bloques de memoria, que son muy rápidas y se localizan como memorias intermedias entre el procesador y el sistema de memoria y que almacenan con antelación porciones de datos e instrucciones para que puedan ser ofrecidas al procesador.

El microprocesador moderno tiene dos niveles de memoria cash, cash primaria (L.1 para el nivel uno), construida en el circuito de el microprocesador; cash secundaria (L.2) y la siguiente (L.3), de las cuales una parte también está construida en el microprocesador, pero cuya capacidad puede ser alcanzada por un circuito externo a la CPU. Cuando el proceso requiere instrucciones o datos, regresa primero a L.1 (la más rápida), a continuación a la más lenta L.2 y L.3 (que ofrece una capacidad más grande).

Debemos destacar que para la realización de memorias cash gran parte de los transistores en el microprocesador es usada. Así el Pentium 4 HT, equipado con 167 millones de transistores usa más de la mitad para memoria cash. El procesador Itanium 2 Medison 9M con 600 millones de transistores tiene para el nivel 3 de cash (L.3) 9 millones de transistores. Con el fin de hacer una diferencia entre el procesador y la memoria cash de un chip en un microprocesador llamamos Core a la parte del circuito que representa el corazón del procesador y Cash a las memorias intermedias.

La arquitectura EPIC (Explicit Parallel Instruction Computing) utiliza ambos principios de la arquitectura de super escala y el procesamiento en paralelo de datos. Los procesadores de 64 bits desde la nueva generación Itanium, producidos por Intel, han tomado prestado la arquitectura EPIC, más exitosa que la ILP (Instruction Level Parallelism), desarrolladas conjuntamente por HP e Intel.

A diferencia de la tecnología RISC, donde el procesador organiza las secuencias de instrucciones, en el modelo EPIC este es el compilador, que es responsable de optimizar el código a fin de tomar ventaja de la ejecución en paralelo.

En la predicción de un ramal, realizada por procesadores RISC se basa en un comportamiento estadístico, justificando la solución por control sobre la instrucción en la dirección correcta. El tamaño del error es de 10 a 20 %, lo que lleva a un aumento en el tiempo de ejecución de un programa y que puede ser reducido si la optimización se realiza según un algoritmo controlado por un compilador y teniendo en cuenta la ejecución previa como un conjunto en lugar de reaccionar basándose en el comportamiento pasado.

Así, los microprocesadores EPIC calculan en paralelo las diferentes posibilidades y la mejor es guardada después. Por lo tanto el controlador controla por separado el importante llamado flujo de datos (un proceso, permitiendo optimizar el orden de procesamiento de los datos) con el objetivo de controlar todos los cálculos en paralelo, sin pérdidas de velocidad.

La tecnología EPIC, cuya función es dominante hoy en día, ofrece al compilador nuevas posibilidades para encontrar paralelismo entre las instrucciones.

En muchas aplicaciones, especialmente en aplicaciones multimedia, se ejecutan un gran número de operaciones elementales, la clave del comportamiento de los microprocesadores está en lo que llamamos “paralelismo con un pequeño paso”. Esto significa ejecutar en paralelo las más posibles instrucciones elementales.

Los principios de transporte de muchos procesadores y máquinas de super escala en las arquitecturas RISC, obviamente, van en esta dirección.

El concepto VLIW (very long instruction word) introduce una cuarta idea: reagrupamiento de un gran número de operaciones elementales independientes en una palabra de instrucción, que va a ser leída y ejecutada en un sólo ciclo de reloj. La arquitectura VLIW está basada en la ejecución de instrucciones, más complejas que con la estructura CISC. Mientras la arquitectura RISC consiste en la simplificación de la ejecución de instrucciones por su contracción, el concepto VLIW usa justamente el método opuesto – codificación de cuatro o más instrucciones en una palabra, es decir, una sola operación de máquina.

El diagrama conceptual y el diagrama de bloques simplificado de VLIW se muestran a continuación:

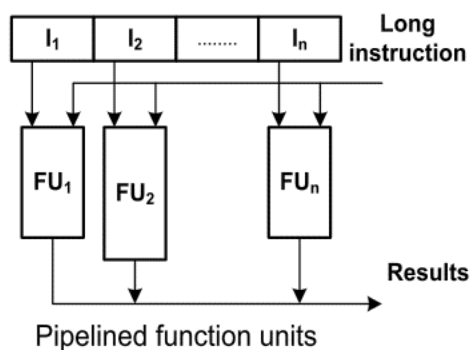


Figura 3: Diagrama conceptual de VLIW

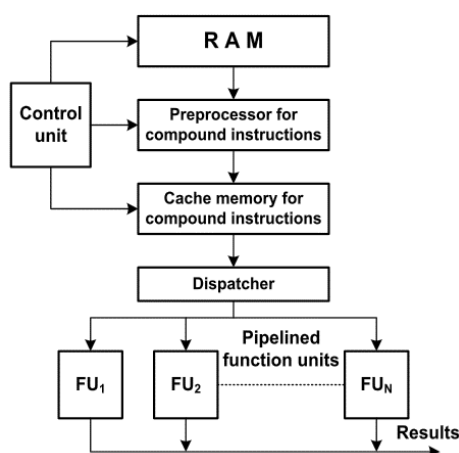


Figura 4: Diagrama de bloques simplificado de VLIW

Clasificación de Flynn

La clasificación de Flynn clasifica las arquitecturas de las computadoras, fue propuesta por Michael J. Flynn in 1966.

Las cuatro clasificaciones definidas por Flynn están basadas en el número de instrucciones concurrentes (o control) y los flujos de datos disponibles en la arquitectura. :

Una Instrucción, un dato (SISD)

Computador secuencial que no explota el paralelismo en las instrucciones ni en flujos de datos. Una unidad de control (CU) obtiene un flujo de instrucción (IS) desde memoria. La CU genera las señales apropiadas de control para dirigir un elemento de procesamiento individual (PE) para operar en un sólo flujo de datos (DS) es decir, una operación a la vez. Ejemplos de arquitecturas SISD son las máquinas con un sólo procesador o monoprocesador como el PC o los antiguos mainframe.

Una Instrucción, múltiples datos (SIMD)

Un computador que explota varios flujos de datos dentro de un único flujo de instrucciones para realizar operaciones que pueden ser paralelizadas de manera natural. Por ejemplo, un procesador vectorial.

Múltiples instrucciones, un dato (MISD)

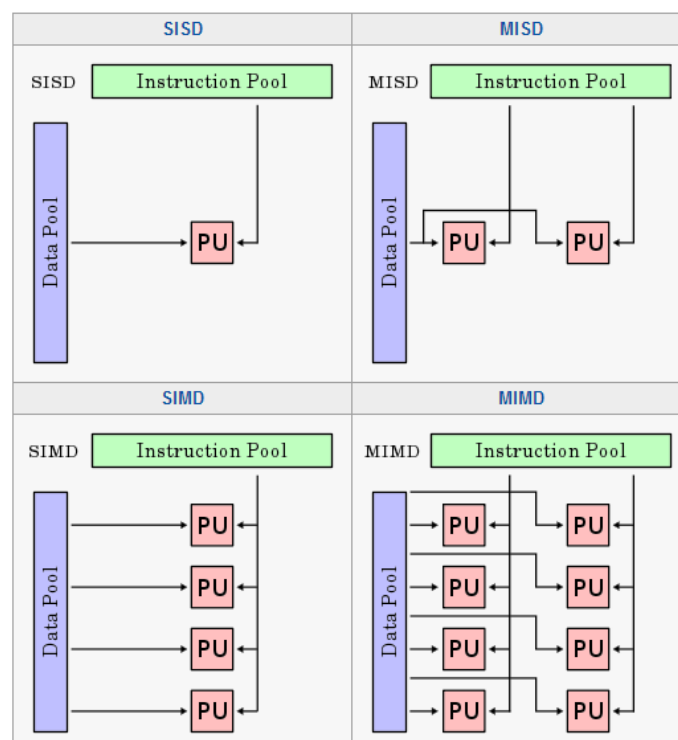
Múltiples instrucciones operan un único flujo de datos. Poco común debido al hecho de que la efectividad de los múltiples flujos de instrucciones suele precisar de múltiples flujos de datos. Sin embargo, este tipo se usa en situaciones de paralelismo redundante, como por ejemplo en navegación aérea, donde se necesitan varios sistemas de respaldo en caso de que uno falle. También se han propuesto algunas arquitecturas teóricas que hacen uso de MISD, pero ninguna llegó a producirse en masa.

Múltiples instrucciones, múltiples datos (MIMD)

Varios procesadores autónomos que ejecutan simultáneamente instrucciones diferentes sobre datos diferentes. Los sistemas distribuidos suelen clasificarse como arquitecturas MIMD; bien sea explotando un único espacio memoria, o uno compartido de distribuido.

Diagrama de clasificaciones

Estas cuatro muestran a donde cada "PU" una unidad de



comparación de las

arquitecturas se continuación en (processing unit) es procesamiento:

Procesamiento multimedia de datos

Las aplicaciones multimedia siempre requieren más potencia por problemas de computación, tales como compresión y descompresión en tiempo real de sonido o vídeo. Algunos procesadores modernos contienen instrucciones especializadas, llamadas multimedia y permiten incrementar el uso de estos dispositivos. Las aplicaciones multimedia manejan valores de 8, 12 ó 16 bits, desperdiciando así los recursos de los modernos procesadores que son capaces de operar con instrucciones de 32 o 64 bits. Las instrucciones multimedia eluden este problema, tratando dos de estos valores cortos en grandes registros. Esto puede verse en el procesamiento de vectores, típicamente tiene 4 ó 8 constituyentes. Por tanto este vector de tipo paralelo es la base de las instrucciones SIMD (Single Instruction Multiple Data) o más datos en una instrucción. Hoy en día los compiladores no son capaces aún de controlar eficazmente tales instrucciones e incluyen un manual con programas con códigos ensamblador y están limitados por bibliotecas predeterminadas. De todos modos, la búsqueda de fragmentos de código, capaces de extraer partes de estas instrucciones debe ser realizada por un programador.

Procesamiento y procesadores de señales digitales

Las señales eléctricas con las que trabajan las computadoras son valores binarios o lógicos, de hecho, tienen dos estados (comienzo y parada) 0 y 1. Los procesadores están diseñados y optimizados históricamente para trabajar con esos valores.

Las dimensiones físicas del mundo moderno (la velocidad de un coche, la temperatura de un horno, el volumen de un instrumento musical) son una gran gama de valores analógicos. La secuencia de valores discretos, representando la transformación de cantidades analógicas en un flujo de información binaria, entendible por el procesador digital, es una abstracción de la realidad, bastante pesada para procesar de lo que sería deseable.

Para el procesamiento de ciertos valores analógicos puede ser más eficiente el uso de dispositivos semiconductores, capaces de procesar señales analógicas directamente, en lugar

de convertirlos en valores binarios. Al no usar lógica binaria digital no se beneficia del progreso exponencial representado por la ley de Moore.

Los procesadores de señal digital (DSP) constituyen un compromiso de futuro. Especializados en el procesamiento en tiempo real de cantidades analógicas, materializan el mundo real, estos procesadores con arquitectura optimizada pueden llevar a cabo cálculos complejos en un solo ciclo de reloj. Esta peculiaridad permite que alcancen fácilmente idéntico número de entradas y salidas, tanto analógicas como digitales.

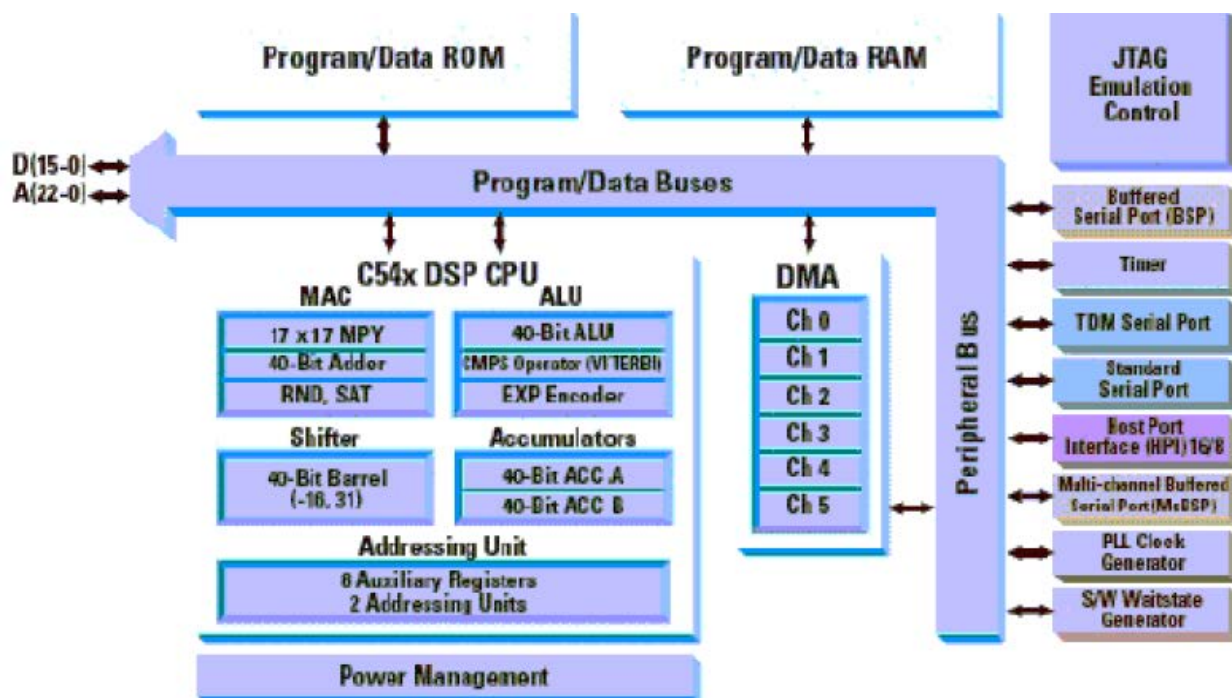
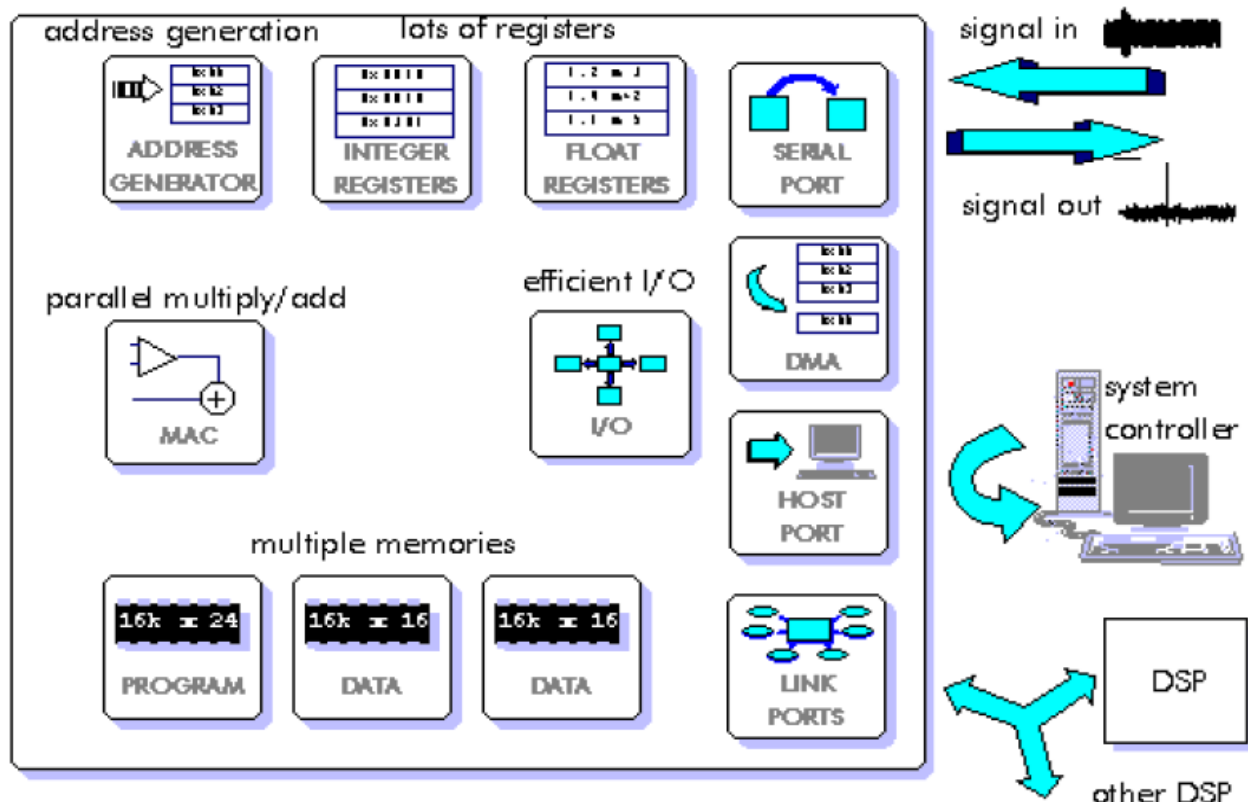
Hoy en día los DSPs son ampliamente utilizados en la mayoría de las aplicaciones para procesamiento digital de señales en tiempo real. Podemos verlos en modernos teléfonos móviles, y en equipamiento de audio y video, es decir, de lo que se trata es de cómo obtener una señal física compleja y su procesamiento. Esto concierne a una proporción precio y consumo equivalente, mucho más exitosa que la solución basada en microprocesadores universales, que requieren más tiempo de programación (hablamos de aceleración de material al resultado de asegurar flujo de información sobre los cálculos realizados de un problema con instrucciones especializadas, capaz de llevar a cabo operaciones de procesamiento complejo, contrariamente al camino de programación, donde convencionalmente el procesador sigue un programa de instrucciones generales).

La función principal usada en DSP es la función MAC (Multiply and Accumulate), es decir, multiplicación seguida de suma y entonces los resultados se guardan. Esta función es la clave de casi todos los cálculos para procesamiento y filtrado. Así el DSP usa sus mejores productos escalares o productos por convolución, así como las operaciones para asignar y sustituir en un ciclo de reloj.

Hoy en día los productores de DSP introducen más y más instrucciones en sus circuitos, que corresponden a operaciones completas para el procesamiento de señal. Desde el comienzo de 1990 estos circuitos han usado elevado paralelismo, mayor cantidad de memoria en el chip y el concepto VLIW, similar al primero, usado por Intel en sus nuevas generaciones de 64 bits. Por ejemplo, el procesador DSP TMS 320 C64x de Texas Instruments con tecnología 90nm se eleva a 1.1 Ghz y puede realizar 9 billones de operaciones MAC por segundo, es decir, 10 veces más que los procesadores DSP de la generación anterior.

Lo que trajo el progreso de la tecnología DSP a partir de los años 90 es la mayoría del equipamiento de las comunicaciones y especialmente el de procesamiento de la voz. La combinación entre telefonía móvil, teléfonos móviles y las aplicaciones VOIP han causado efectivamente la necesidad de este tipo de procesadores durante la dicha década.

La próxima generación con objetos capaces de interaccionar con el mundo físico, sólo tendrá que reforzar sólo hará que fortalecer el apetito por los procesadores digitales.



Datos históricos sobre los procesadores DSP

En 1978 Intel lanza su primer procesador con señales analógicas 2920. Hablando coloquialmente sobre un circuito, era la combinación en un único chip de un ADC, un DAC y un procesador con señales internas. El circuito, sin embargo, carecía de un multiplicador y no era competitivo. En 1979 AMI lanzó su S2811. Diseñado como un microprocesador periférico, tenía que ser inicializado por su anfitrión. Tampoco tuvo éxito.

El mismo año Laboratorios Bell presentó su procesador monolítico DSP – MAC4. Pero los procesadores DSP reales llegaron al mercado un año después:

μPD7720

AT&T, presentados en la conferencia internacional IEEE Solid State Circuit en 1980. Ambos procesadores fueron ampliamente aceptados para investigaciones llevadas a cabo en ese momento en el área de la telefonía. El procesador

μPD7720

4MHz y 128 bits de RAM, 512 bits de ROM y memoria de programa de 512 bit. Tenía un formato de instrucción de tipo VLIW, permitiendo llevar a cabo simultáneamente todo tipo de operaciones aritméticas y lógicas, con el objetivo de aumentar y reducir el número de registros y cambiando operaciones por ciclo de reloj.

En 1983 Texas Instruments presentó el TMS 32010, para demostrar que los circuitos DSP podían tener éxito comercial. Texas Instruments hoy en día, es un líder en el área de los circuitos DSP.

Tendencias en el desarrollo de microprocesadores ELIS

Desde el comienzo de los años 90, la producción de microprocesadores ha progresado lo que ha permitido embeber la arquitectura completa de un ordenador en un solo circuito. Los años siguientes se generalizó esta tendencia de un PC en un chip, es decir, integrar en el mismo sustrato de silicio, un microprocesador, un circuito para gráficos y un controlador de memoria. Tal proceso permitió reducir algunos canales aislantes entre el núcleo del procesador y el subsistema, así como la reducción del coste de producción. Las funciones añadidas al circuito del microprocesador, sin embargo, fueron menos satisfactorias que aquellos, el sistema tiene circuitos externos cuya magnitud está limitada. De este modo, en lo concerniente a pequeñas computadoras, como un todo, aplicaciones en comunicaciones móviles (Smartphone, PDA,...) o cierto tipo de plataformas.