Chapter 2-1: CPUs

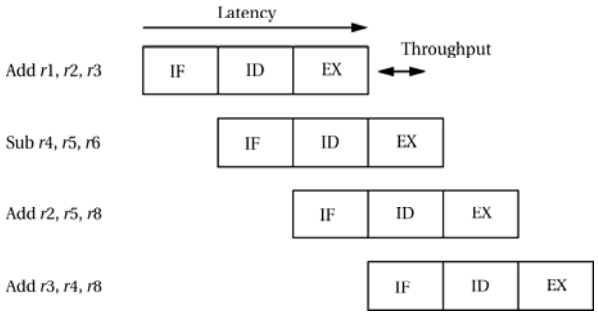
Soo-Ik Chae

Topics

- CPU metrics.
- Categories of CPUs.
- CPU mechanisms.

Performance as a design metric

- Performance = speed:
  - Latency.
  - Throughput.
- Average vs. peak performance.
- Worst-case and best-case performance.



Other metrics

- Cost (area).
- Energy and power.
- Predictability: important for embedded systems
  - Pipelining: branch penalty.
  - Memory system (Cache) : cache miss penalty
- Security: difficult to measure because of the fact that we do not know of a successful attack.

Flynn's taxonomy of processors

- Single-instruction single-data (SISD): RISC, etc.
- Single-instruction multiple-data (SIMD): all processors perform the same operations.
- Multiple-instruction multiple-data (MIMD): homogeneous or heterogeneous multiprocessor.
- Multiple-instruction multiple data (MISD).

Other axes of comparison

- RISC.
  - Emphasis on software
  - Single-cycle, simple instructions
  - Register to register: LOAD" and "STORE" are independent instructions
  - Low cycles per second,
  - Large code sizes
  - Spends more transistors on memory registers
- CISC.
  - Emphasis on hardware
  - multi-cycle, complex instructions
  - Memory-to-memory: LOAD" and "STORE" incorporated in instructions
  - High cycles per second
  - Small code sizes
  - Transistors used for storing complex instructions

RISC	CISC
1. 1-cycle simple instructions	1. multi-cycle complex instructions
2. only LD/ST can access memory	2. any instruction may access memory
3. designed around pipeline	3. designed around instn. set
4. instns. executed by h/w	4. instns interpreted by micro-program
5. Fixed format instns	5. variable format instns
6. Few instns and modes	6. Many instns and modes
7. Complexity in the compiler	7. Complexity in the micro-program
8. Multiple register sets	8. Single register set

Other axes of comparison

- Instruction issue width.
  - Single issue
  - Multiple issue: higher performance, high cost, increased power consumption
- Scheduling for multiple-issue machines.
  - Static scheduling: VLIW
  - Dynamic schedule: superscalar
- Vector processing: instruction for 1D or 2D arrays
- Multithreading: a fine-grained concurrency mechanism that allows the processor to quickly switch between several threads of execution

Embedded vs. general-purpose processors

- Embedded processors may be optimized for a category of applications.
  - Must be flexible
  - Customization may be narrow or broad.
  - Billions of 8-bit processors sold each year
  - 100s millions of 32-bit processors for embedded systems
- We may judge embedded processors using different metrics:
  - Code size.
  - Memory system performance.
  - Preditability.

ARM Processor Family

Processor family	# of pipeline stages	Memory organization	Clock Rate	MIPS/MHz
ARM6	3	Von Neumann	25 MHz	
ARM7	3	Von Neumann	66 MHz	0.9
ARM8	5	Von Neumann	72 MHz	1.2
ARM9	5	Harvard	200 MHz	1.1
ARM10	6	Harvard	400 MHz	1.25
StrongARM	5	Harvard	233 MHz	1.15
ARM11	8	Von Neumann/ Harvard	550 MHz	1.2

ARM Architecture Version Summary

Core	Version	Feature
ARM1020T	v5T	▪ Improved ARM/Thumb Interworking ▪ CLZ instruction for improved division
ARM9E-S, ARM10TDMI, ARM1020E	v5TE	▪ Extended multiplication and saturated maths for DSP-like functionality
ARM7EJ-S, ARM926EJ-S, ARM1026EJ-S	v5TEJ	▪ Jazelle Technology for Java acceleration
ARM11, ARM1136J-S,	v6	▪ Low power needed ▪ SIMD (Single Instruction Multiple Data) media processing extensions

J: Jazelle

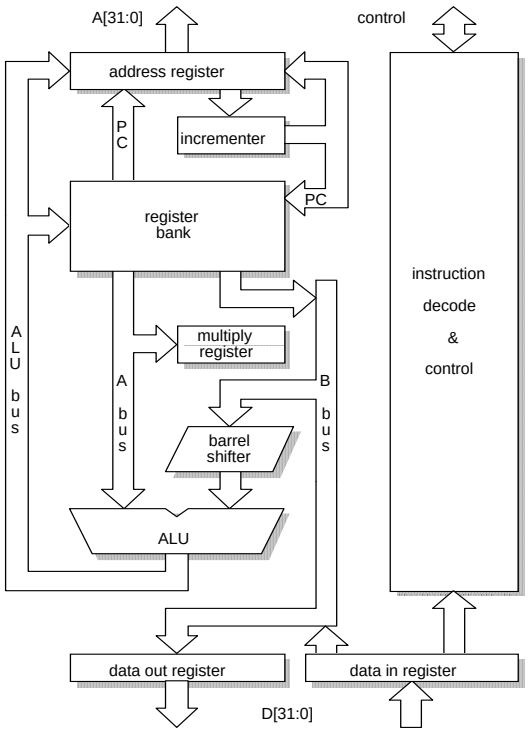
S: Synthesizable

E: Enhanced DSP instruction

F: integral vector floating point unit

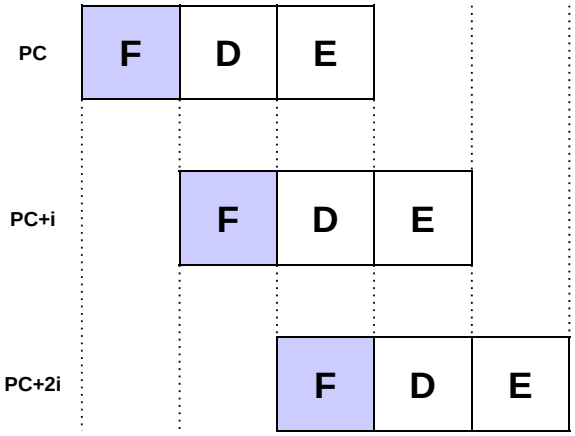
ARM7 3-stage pipeline organization

- Organizations
  - **Address generating block**
    - Address register
    - Incrementer
  - **Register bank**
    - 31-GPRs, 6-PSRs
    - 2 read, 1 write ports
    - Additional 1 read, 1 write port for PC
  - **Barrel shifter**
  - **ALU**
  - **Data register**
  - **Control logic**
    - External interface
    - Instruction decoder
    - Datapath control

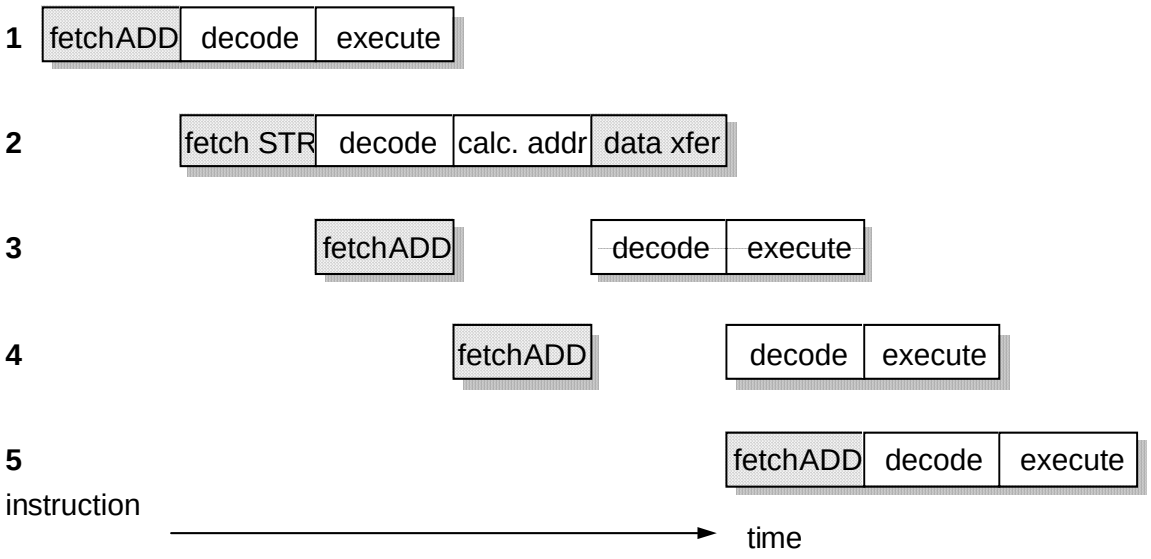


ARM7 3-stage Pipeline

- ARM7 family has 3 stage pipeline
- 3 stage pipeline
  - Fetch
    - Instruction fetch from memory
  - Decode
    - Instruction decoding
    - Datapath control signals for the next cycle
  - Execute
    - Reading registers
    - Shift and ALU operations
    - Writing back to the register bank

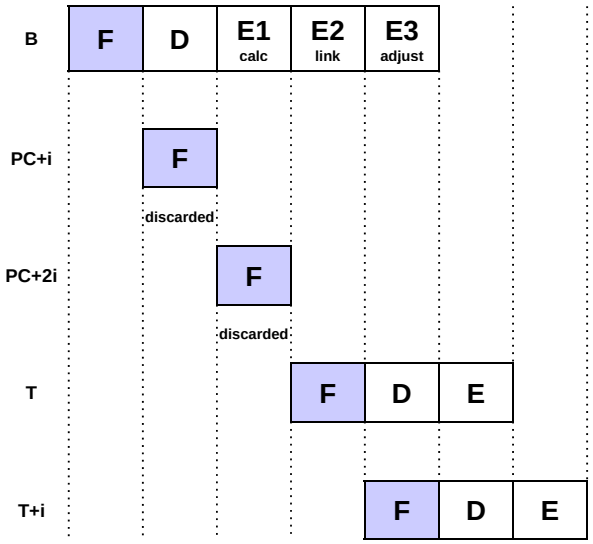


ARM7 multi-cycle instructions

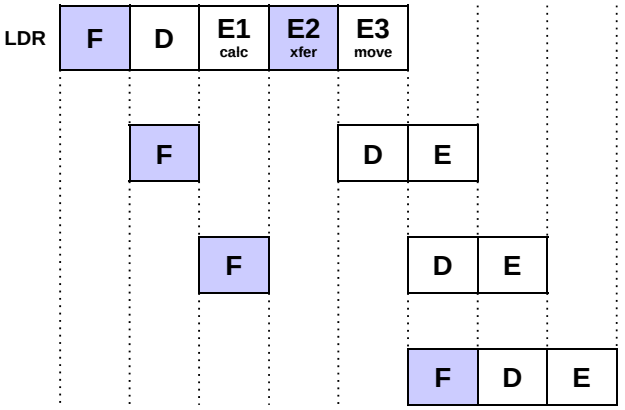


ARM7 multi-cycle instructions

- Branch

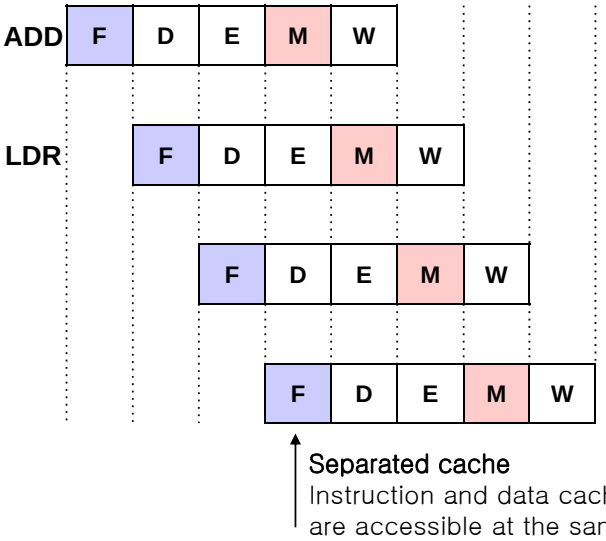


- LDR

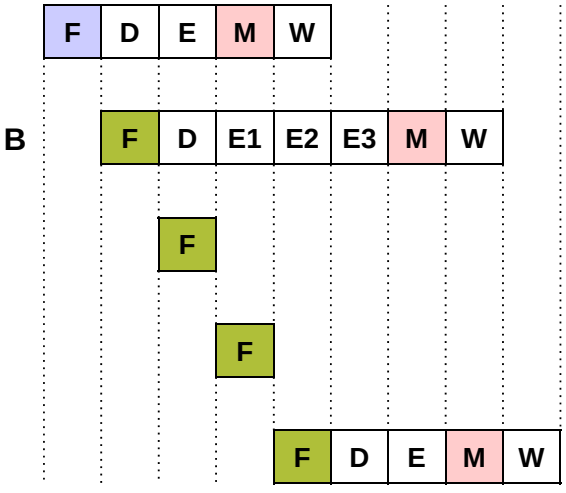


ARM9TDMI core

- LDR

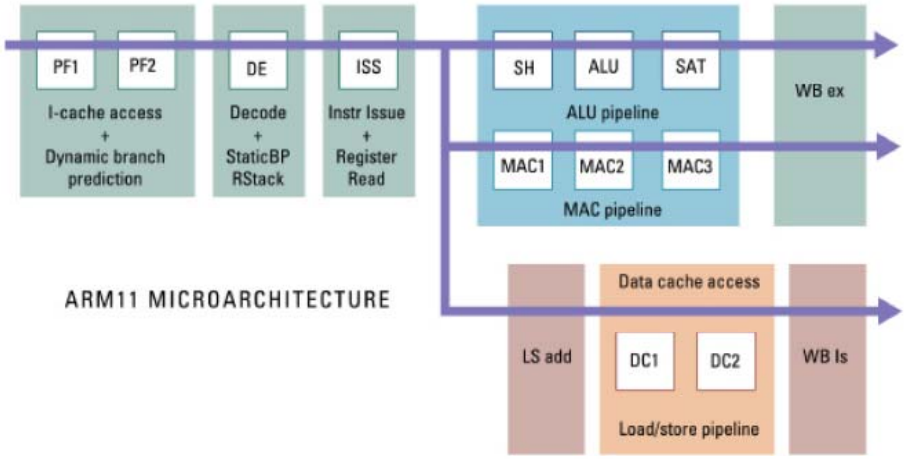


- Branch



ARM11

- 8 stage pipeline
  - Branch Prediction and Return Stack
  - Separate processing units for the ALU, MAC, and Load-Store (LS) instructions
    - although the pipeline is single issue



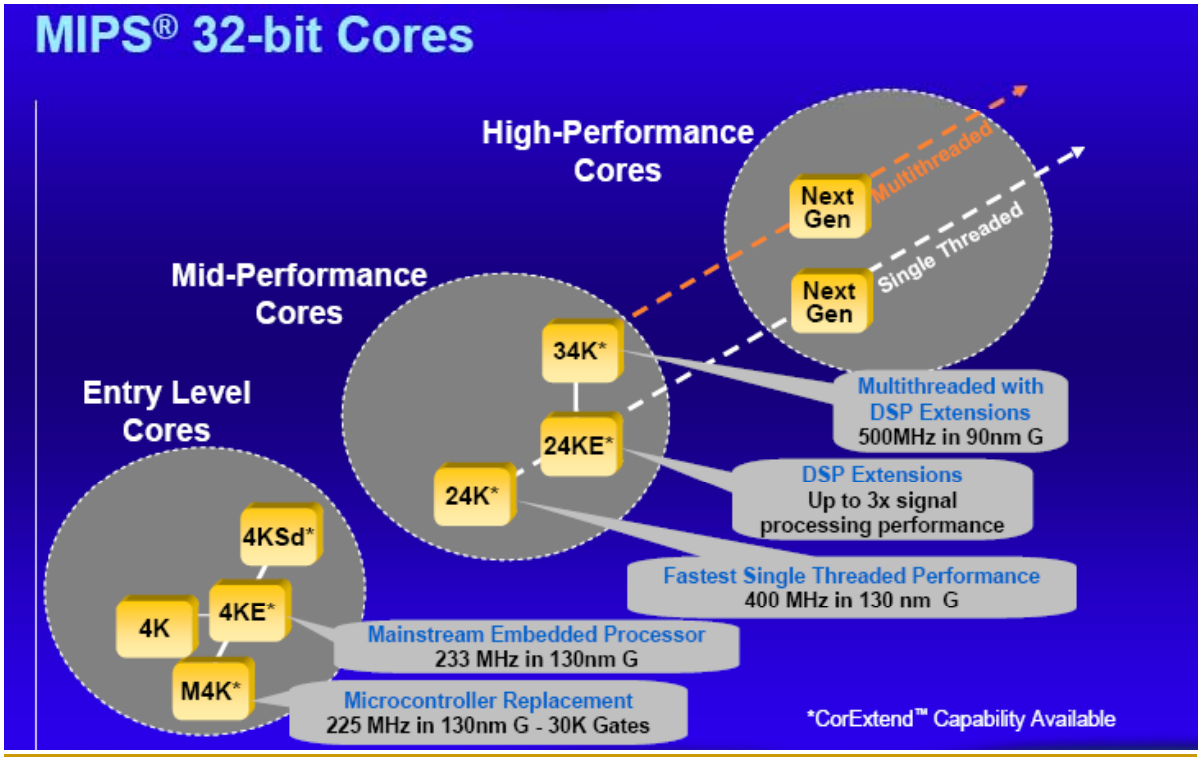
Feature Comparison

Feature	ARM9E	ARM10E	XScale	ARM11
Architecture	ARMv5TE(J)	ARMv5TE(J)	ARMv5TE	ARMv6
pipeline length	5	6	7~8	8
Java decode	(ARM926EJ)	(ARM1026EJ)	No	Yes
V6 SIMD instructions	No	No	No	Yes
MIA instructions	No	No	Yes	Available as coprocessor
Branch prediction	No	Static	Dynamic	Dynamic
Independent Load-store unit	No	Yes	Yes	Yes
Instruction issue	Scalar, in-order	Scalar, in-order	Scalar, in-order	Scalar, in-order
Concurrency	None	ALU/MAC, LSU	ALU, MAC, LSU	ALU, MAC, LSU
Out-of-order completion	No	Yes	Yes	Yes
Target implementation	Synthesizable	Synthesizable	Custom chip	Synthesizable and Hardmacro
Performance range	Up to 250MHz	Up to 325MHz	200MHz ~ > 1GHz	350MHz ~ > 1GHz

MIPS32 processor family

- MIPS: MIPS32 4K has 5-stage pipeline; 4KE family has DSP extension; 4KS is designed for security.

MIPS32 processor family



PowerPC processor family

- PowerPC: 400 series includes several embedded processors; MPD7410 is two-issue machine; 970FX has 16-stage pipeline.

PowerPC processor family

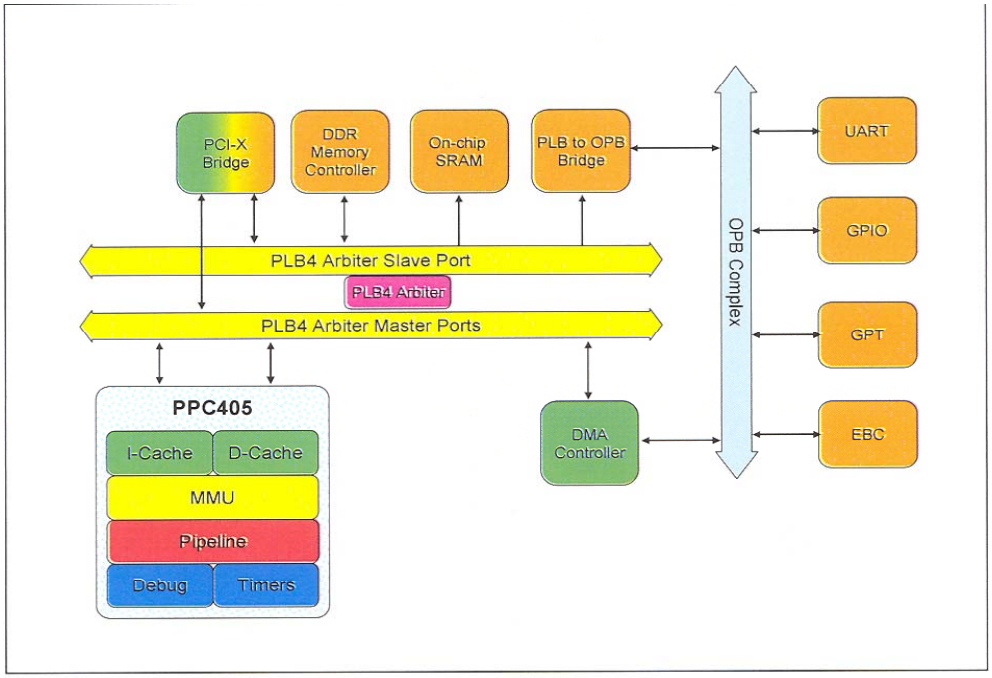


Figure 1: PPC405 CPU and CoreConnect based SOC Example

PowerPC processor family

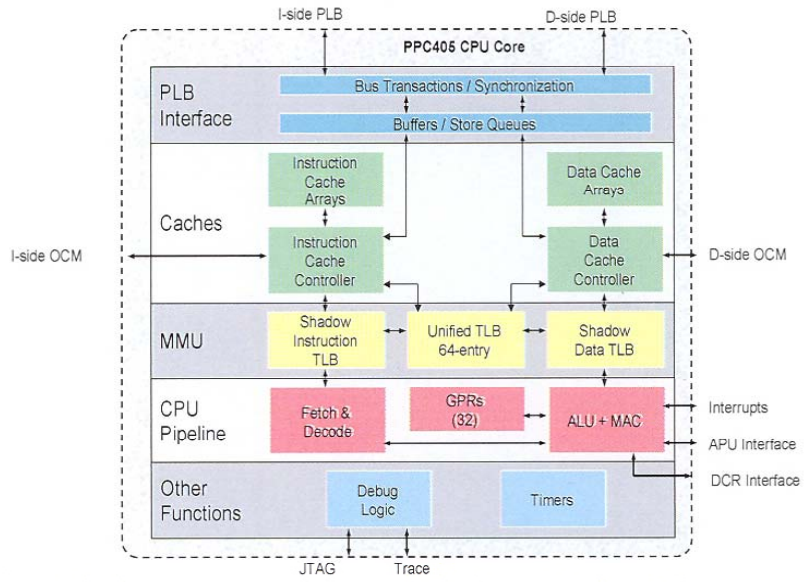


Figure 2: PPC405 CPU core block diagram.

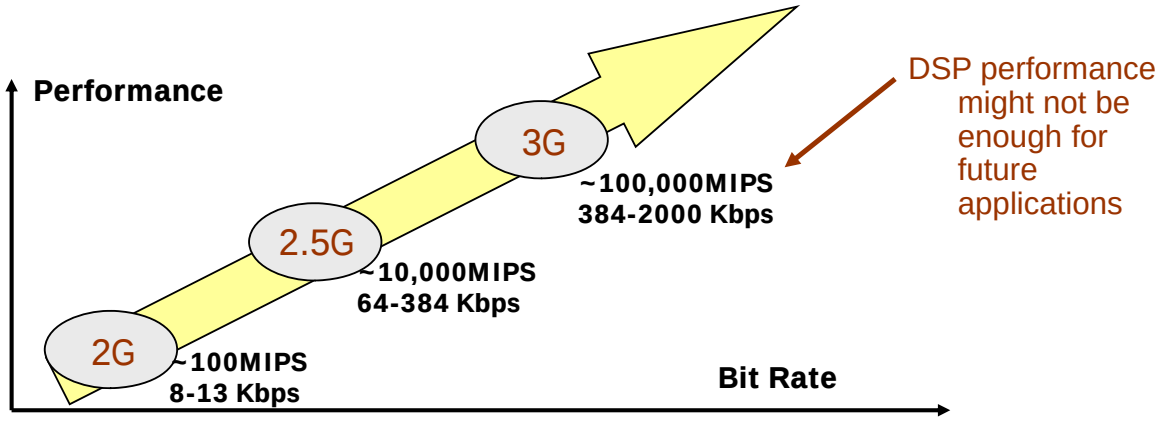
What is DSP?

- DSP = Digital Signal Processing
  OR
  DSP = Digital Signal Processor?

- DSP used to denote both
  - meaning can be deduced from the context in which the term DSP is used.
- What is a Digital Signal Processor (DSP)?
  - Microprocessor specifically designed to perform fast DSP operations (e.g., Fast Fourier Transforms, inner products, Multiply & Accumulate)

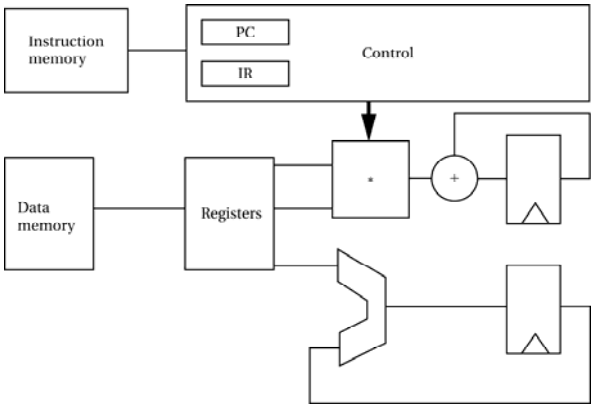
DSP performance

- Wireless Systems requires more and more high performance and higher bandwidth



Digital signal processors

- First DSP was AT&T DSP16:
  - Hardware multiply-accumulate unit.
  - Harvard architecture.
- Today, DSP is often used as a marketing term.
- Modern DSPs are heavily pipelined.



TMS320C55x ™ DSP Generation, 16-bit Fixed Point – Most Power Efficient DSP

Specifications	Features	Applications
• C55x™ DSP core delivers 300 MHz for up to 600-MIPS performance	• Advanced automatic power management	• Feature-rich, miniaturized personal and portable products
• 1.6-volt core and 3.3-volt peripherals	• Configurable idle domains to extend your battery life	• 2G, 2.5G and 3G cell phones and basestations
	• Shortened debug for faster time-to-market	• Digital audio players
	• 144-MHz/200-MHz clock rate	• Digital still cameras
	• 256-KB RAM, 64-KB ROM	• Electronic books
	• Three McBSPs, I 2 C, watchdog timer, general-purpose timers	• Voice recognition
	• USB 2.0 full-speed (12 Mbps)	• GPS receivers
	•10-bit ADC	• Fingerprint/Pattern recognition
	•real-time clock (RTC)	• Wireless modems
		• Headsets
		• Biometrics

TMS320C55x ™ DSP + RISC, 16-bit Fixed Point – OMAP Processor

Specifications	Features	Applications
• Dual CPU processor integrating a TMS320C55x™ DSP core and an ARM925TDMI™ RISC @150 MHz	▪150-MHz TI-enhanced ARM925	• Internet appliances
• 1.8-volt core and 1.8-volt peripherals	• 16 KB instruction cache and 8 KB data cache	• Applications processing
	• Data and instruction MMUs	• Enhanced gaming
	• 32-bit and 16-bit instruction sets	• Webpad
	150-MHz TMS320C55x™ DSP	• Point-of-sale
	• 12 KW (24 KB) instruction cache	• Medical devices
	• 80 KW (160 KB) SRAM	• Industry-specific PDAs
	• 16 KW (32 KB) ROM	• Telematics
	• Two 16-bit memory interfaces for SDRAM and flash	• Digital media processing
	• Nine-channel system DMA controller	• Military and government cellular
	• LCD controller	
	• USB 1.1 host and client	
	• MMC/SD card interface	
	• Seven serial ports plus three UARTs, Nine timers, Keyboard interface	
	• Less than 250 mW at 1.6 V	

TMS320C62x ™ DSP Generation, 16-bit Fixed Point – High Performance DSP

Specifications	Features	Applications
• 16-bit fixed-point DSPs	• C6000™ DSP Platform VelociTI™ advanced architecture	• Pooled modems
• Up to 2400 MIPS	• Up to eight 32-bit instructions executed each cycle	• Digital Subscriber Line (xDSL)
•Running at 300 Mhz	• Eight independent, multi-purpose functional units thirty-two 32-bit registers	• Wireless basestations
	• Industry's most advanced C compiler and Assembly Optimizer maximize efficiency and performance	• Central office switches
		• Private Branch Exchange (PBX)
		• Digital imaging
		• Call processing
		• 3D graphics
		• Speech recognition
		• Voice over packet

TMS320C67x ™ DSP Generation, 32-bit Floating Point – High Performance DSP

Specifications	Features	Applications
• 32-bit floating point DSPs	• C6000™ DSP Platform VelociTI™ advanced architecture	• Pooled modems
• Up to 1350 MFLOPS	• Up to eight 32-bit instructions executed each cycle	• Digital Subscriber Line (xDSL)
•Running at 225 Mhz	• Eight independent, multi-purpose functional units thirty-two 32-bit registers	• Wireless basestations
	• Industry's most advanced C compiler and Assembly Optimizer maximize efficiency and performance	• Central office switches
	• IEEE floating-point format	• Private Branch Exchange (PBX)
	• Up to 1350 MFLOPS at 225	• Digital imaging
	• Two new multi-channel serial ports (McASP) (C6713 DSP) can support up to stereo channels of I^2S (Inter IC Sound) and compatible with S/PDIF transmit protocol. Note I2S is a protocol for transmitting 2 channels of digital audio over a single serial connection	• Call processing
		• 3D graphics
		• Speech recognition
		• Voice over packet

TMS320C64x ™ DSP Generation, 16-bit Fixed Point – High Performance DSP

Specifications

•16-bit fixed point processor

TMS320C64x DSP high per-

formance core provides scalable

performance of up to 1.1 GHz

• The industry's fastest DSPs with

up to 600 MHz (4800 MIPS) performance

• C64x DSPs are software compatible with TI's C62x™ DSPs

Features

• C6000™ DSP Platform VelociTI™ advanced architecture

• Up to eight 32-bit instructions executed each cycle

• Eight independent, multi-purpose functional units thirty-two 32-bit registers

• Industry's most advanced C compiler and Assembly Optimizer maximize efficiency and performance

Applications

•DSL and pooled modems

•Basestation transceivers

•Wireless LAN

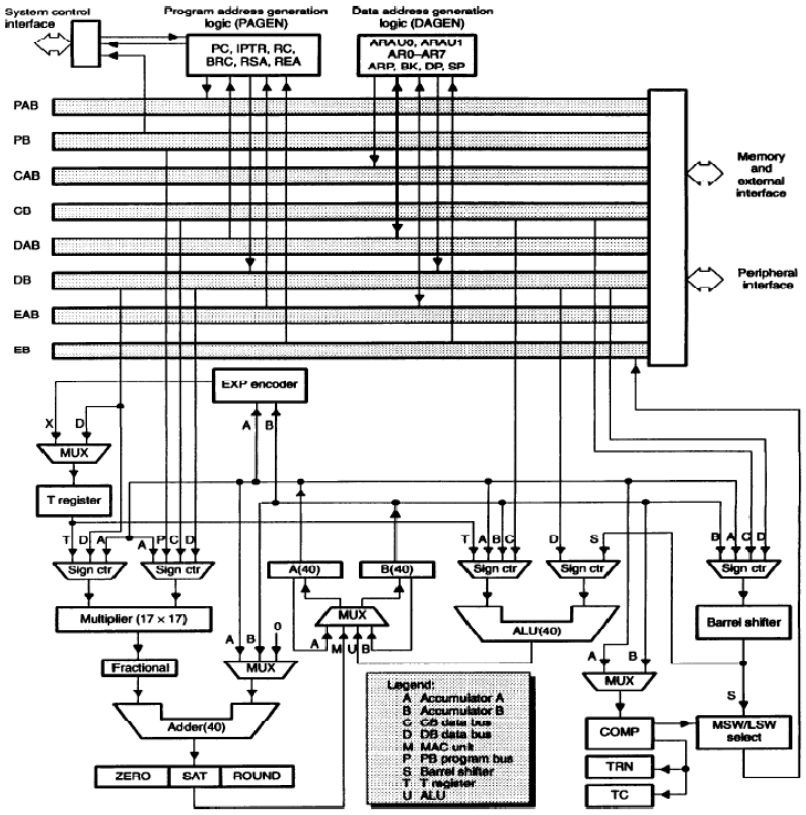
•Enterprise PBX

•Multimedia gateway

•Broadband video transcoders

•Streaming video servers and clients

•Highspeed raster image processing (RIP)



'C54x Block Diagram

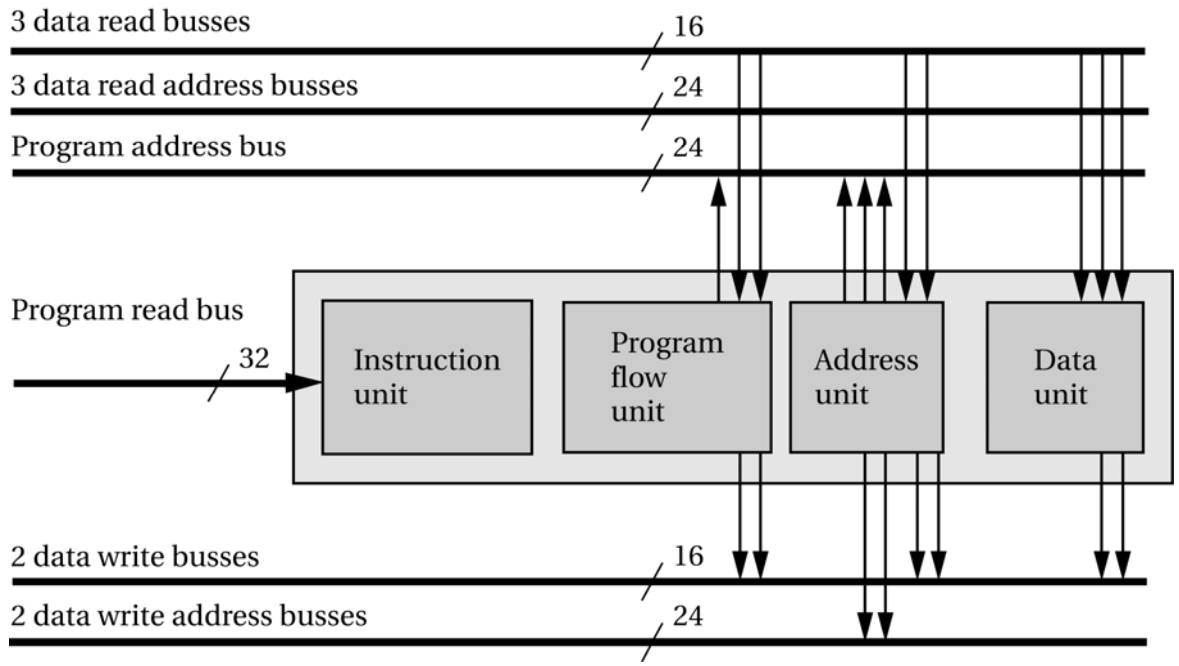
- 17x17 MAC Unit
- Saturation and Rounding Hardware
- Two 40-bit ACC's
- 40-bit ALU
- 40-bit Barrel Shifter
- Temporary Register
- Exponent Encoder
- Program and Data Address Generation Units
- Compare, Select and Store Unit
- 4 Internal Bus Pairs
- External Interface

DSP54.1 - 5

Example: TI C5x DSP

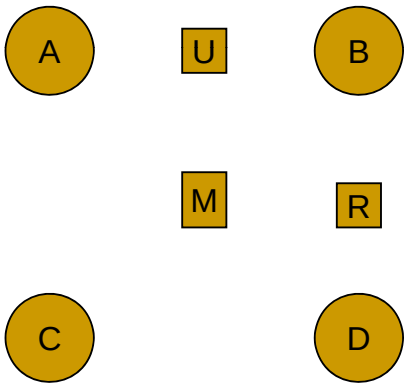
- 40-bit arithmetic unit
  - 32-bit values with 8 guard bits
- Barrel shifter.
- 17 x 17 multiplier.
- Comparison unit for Viterbi encoding/decoding.
- Single-cycle exponent encoder for wide-dynamic-range arithmetic.
- Two address generators.

TI C55x microarchitecture



TI C55x co-processors

- Designed to support
  - Pixel interpolation
  - Motion estimation
  - DCT/IDCT computation
- Interpolates U, M, R values given A, B, C, D pixels.



Fixed Point Vs Floating Point

Floating Point

Applications

- **Modems**
- **Digital Subscriber Line (DSL)**
- **Wireless Basestations**
- **Central Office Switches**
- **Private Branch Exchange (PBX)**
- **Digital Imaging**
- **3D Graphics**
- **Speech Recognition**
- **Voice over IP**

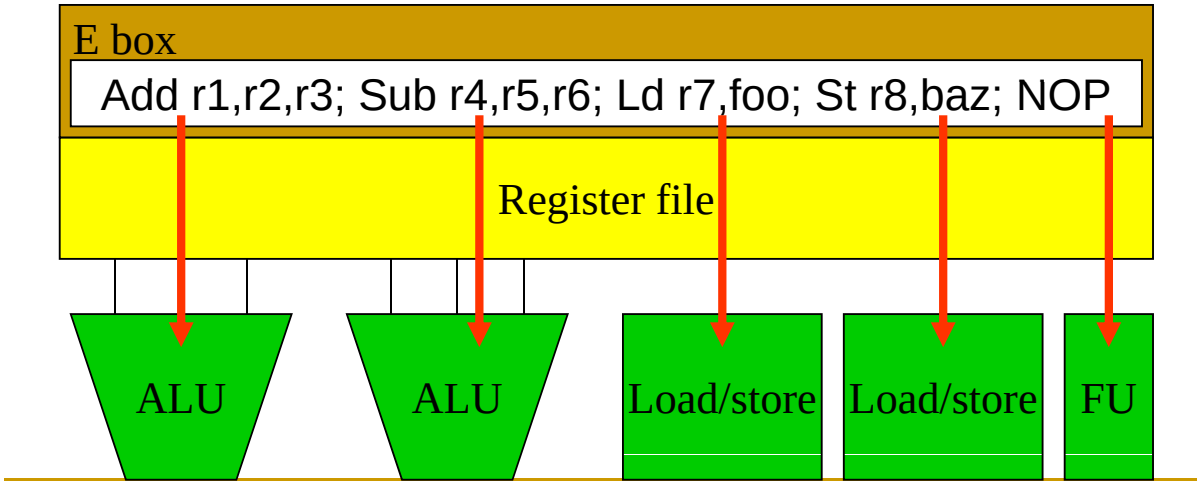
Fixed Point

Applications

- **Portable Products**
- **2G, 2.5G and 3G Cell Phones**
- **Digital Audio Players**
- **Digital Still Cameras**
- **Electronic Books**
- **Voice Recognition**
- **GPS Receivers**
- **Headsets**
- **Biometrics**
- **Fingerprint Recognition**

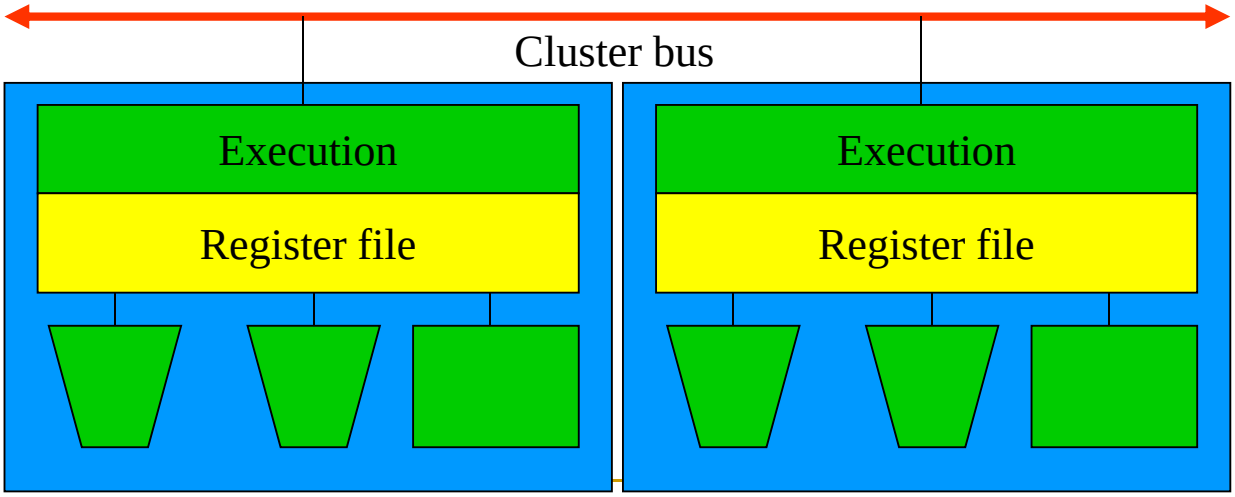
Simple VLIW architecture

- Powerful compiler
- A packet of instruction
- Large register file with multiple ports feeds multiple function units.



Clustered VLIW architecture

- Register file, function units divided into clusters.



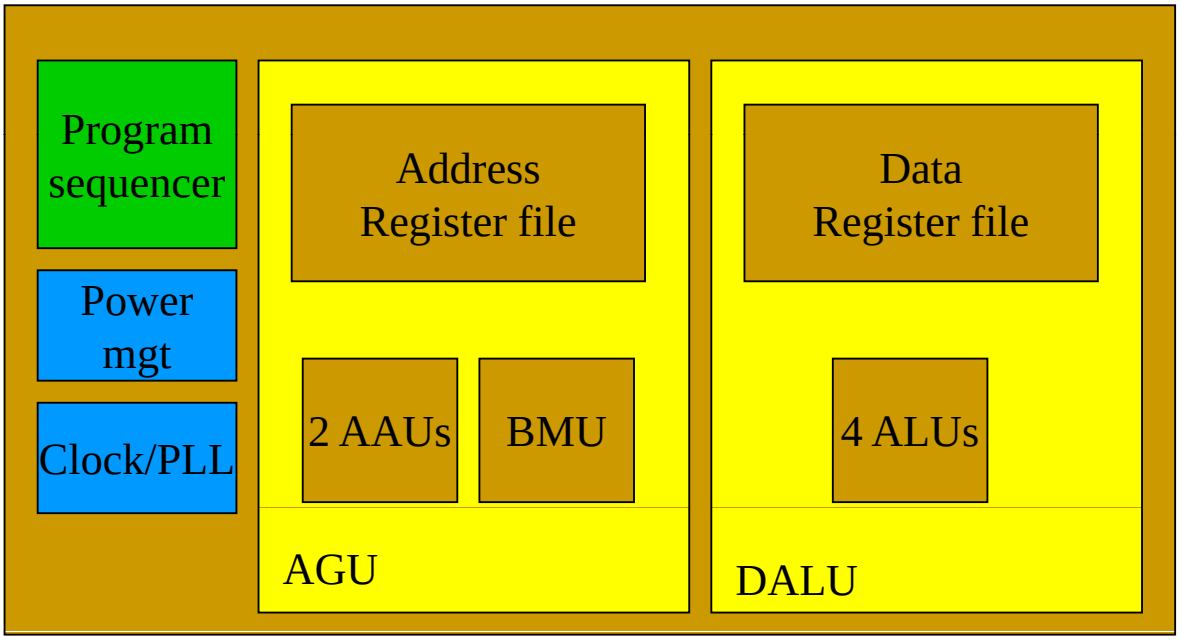
Parallelism extraction in VLIW

- Static:
  - Use compiler to analyze program.
  - Simpler CPU.
  - Can make use of high-level language constructs.
  - Can't depend on data values.
- Dynamic:
  - Use hardware to identify opportunities.
  - More complex CPU.
  - Can make use of data values.

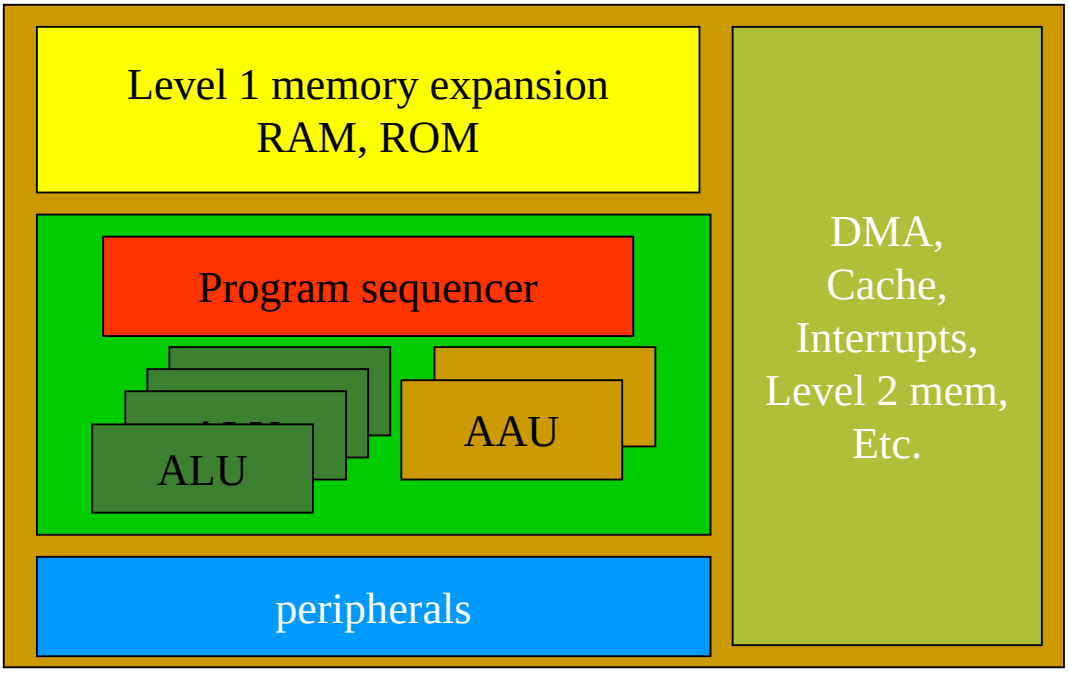
Motorola Starcore SC140

- DALU includes 4 ALUs, 1 register file.
- AGU includes 2 address arithmetic units (AAU), 1 address register file.
- Program sequencer and control unit (PSEQ).
- Performance:
  - 4 MACs per cycle.
  - 10 RISC MIPS per MHz clock.

SC140 Core

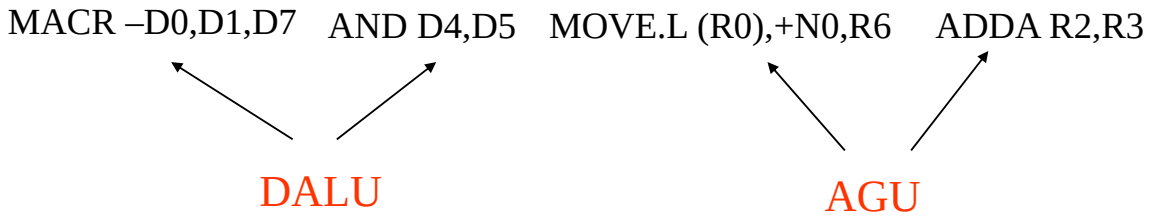


Typical SC140 configuration



Instruction format

- 16-bit instructions.
- Up to 6 instructions per cycle.
- Instructions are grouped to define allowable simultaneous operations.



AGU addressing

- Allowable addressing modes:
  - Linear: useful for general purpose addressing.
  - Modulo: useful for FIFO queues.
  - Reverse-carry: useful for FFT.
  - Automatic updating during register indirect.
  - Stack.
- Array addressing: base, offset, modifier registers.

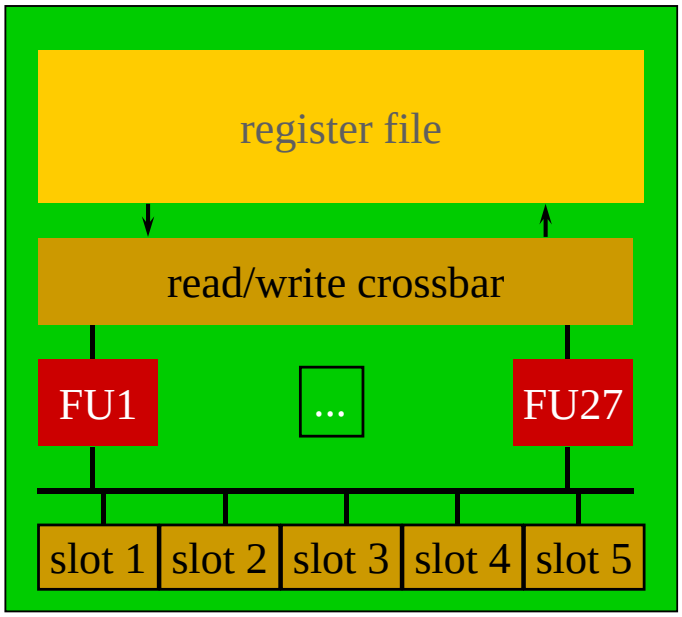
TM-1 characteristics

- Characteristics
  - 5 RISC operations/sec
  - Floating point support
  - Sub-word parallelism support
  - Guarded operation (If Conversion)
  - Additional custom operations

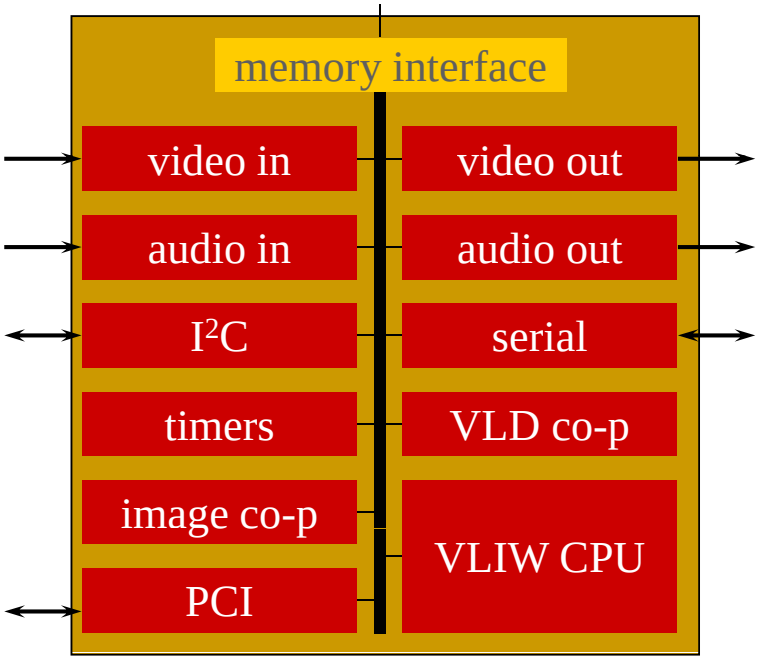
27 function units

#Functional Unit	Constant	5
	Integer ALU	5
	Load/Store	2
	DSP ALU	2
	DSPMUL	2
	Shifter	2
	Branch	3
	INT/Float MUL	2
	Float ALU	2
	Float Compare	1
	Float sqrt/div	1
#Register		128
Instruction cache		32KB, 8 way
Data cache		16KB, 8 way
#Operation slots/instruction		5

TM-1 VLIW CPU



Trimedia TM-1



Superscalar processors

- Instructions are dynamically scheduled.
  - Dependencies are checked at run time in hardware.
- Used to some extent in embedded processors.
  - Embedded Pentium is two-issue in-order.

SIMD and subword parallelism

- Many special-purpose SIMD machines.
- Subword parallelism is widely used for video.
  - ALU is divided into subwords for independent operations on small operands.
- Vector processing is widely used for integer values.

SIMD Extensions

Name		AltiVec	MAX-2	MDMX	MMX/3D Now	MMX/SIMD	VIS
Company		Motorola	HP	MIPS	AMD	Intel	Sun
Instruction set		Power PC	PA RISC 2.0	MIPS-V	IA32	IA32	SPARC V.9
Processor		MPC7400	PA RISC	R10000	K6-2	Pentium III	UltraSparc
Fixed-Point (Integer)	8-bit	16	NA	8	8	8	8
	16-bit	8	4	4	4	4	4
	32-bit	4	NA	NA	2	2	2
Floating-point	Single Precision	4	2	2	2	4	Na
Fixed-Point Register File	Size	32x128b	32x64b	32x64b	8x64b	8x64b	32x64b
	Shared with	Dedicated	Integer Reg.	FP Reg.	Dedicated	FP Reg.	FP Reg.

SIMD Extensions

GPP with Multimedia exten.								
ISA Name	AltiVec	MAX-1/2	MDMX	MMX/ 3DNow	VIS	MMX/ SIMD	SSE	SSE2
Company	Motorola	HP	MIPS	AMD	Sun	Intel	Intel	Intel
Instruction Set	Power PC	PARISC2	MIPS-V	IA32	P. V.9	IA32	IA64	IA64
Processor	MPC7400	PA RISC	R1000 PA8000	K6-2	Ultra Sparc	Pentium2	P.3	P.4
Date	1999	1995	1997	1999	1995	1997	1999	2000
Datapath	128-bit	64-bit	64-bit	64-bit	64-bit	64-bit	128-bit	128-bit
Size of Reg. File	32x128b	(31) /32x64b	32x64b	8x64b	32x64b	8x64b	8x128b	8x128b
Shared with	Dedicated	Int. Reg.	FP Reg.	Dedicated	FP Reg.	FP Reg.	Dedicated	Dedicated
Int. data types								
8-bit	16		8	8	8	8	16	16
16-bit	8	4	4	4	4	4	8	8
32-bit	4			2	2	2	4	4
64-bit							2	2

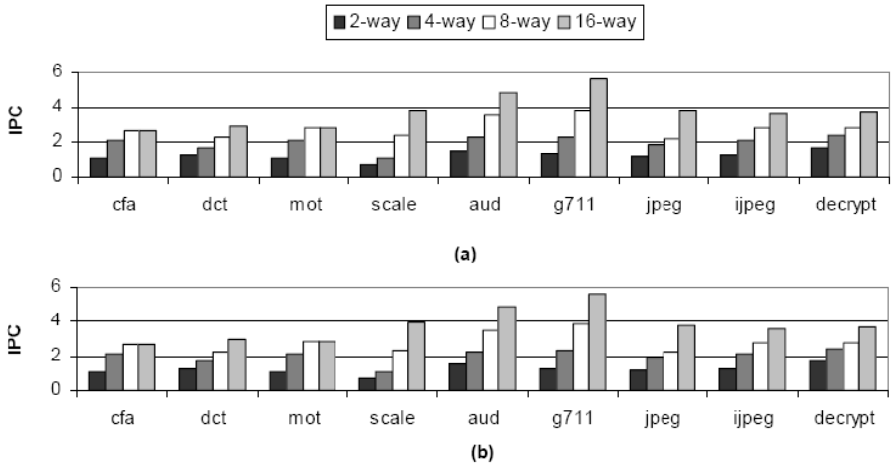
Multithreading

- Low-level parallelism mechanism.
- Hardware multithreading alternately fetches instructions from separate threads.
- Simultaneous multithreading (SMT) fetches instructions from several threads on each cycle.

Processor Resource Utilization

- Processor choice depends on program characteristics.
  - Leverage our knowledge of the core algorithms
- Many researchers assume that multimedia algorithms exhibit high levels of parallelism.
  - Experiments with SimpleScalar shows that this is not the case.
  - Most applications exhibit fewer than 4 IPC.

Available parallelism in multimedia applications (Talla et al.)



	cfa	dct	mot	scale	aud	g711	jpeg	ijpeg	decrypt
4-way	< 1 %	< 1 %	< 1 %	< 2 %	< 4 %	< 1 %	< 1 %	< 1 %	< 1 %
8-way	< 1 %	< 1 %	< 1 %	< 3 %	< 1 %	< 1 %	< 1 %	< 1 %	< 1 %
16-way	< 1 %	< 1 %	< 1 %	< 1 %	< 1 %	< 1 %	< 1 %	< 1 %	< 1 %

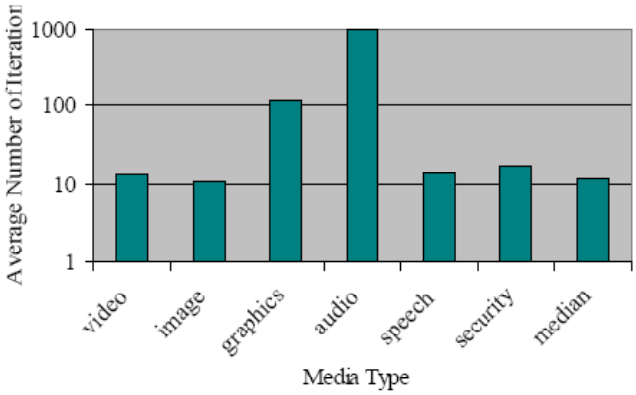
(c)

Fig. 1. (a) IPC with both the SIMD and non-SIMD resources scaled, (b) IPC with non-SIMD resources scaled, but SIMD resources are constant (same as 2-way processor configuration), and (c) performance improvement of (a) over (b)

Dynamic behavior of loops in MediaBench (Fritts)

- Path ratio
  - (instructions executed per iteration) / (total number of loop instructions).
- MediaBench shows small path ratio -> considerable conditional behavior in loops.

Operand characteristics in MediaBench

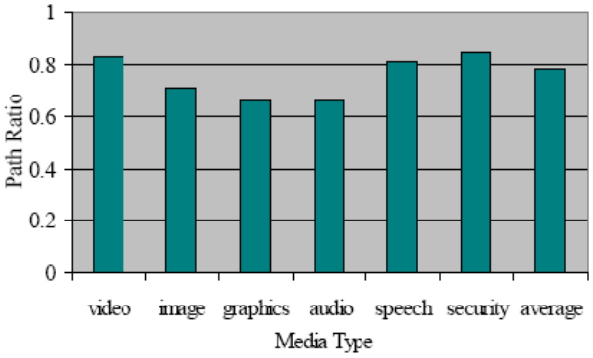


- average number of loop iterations

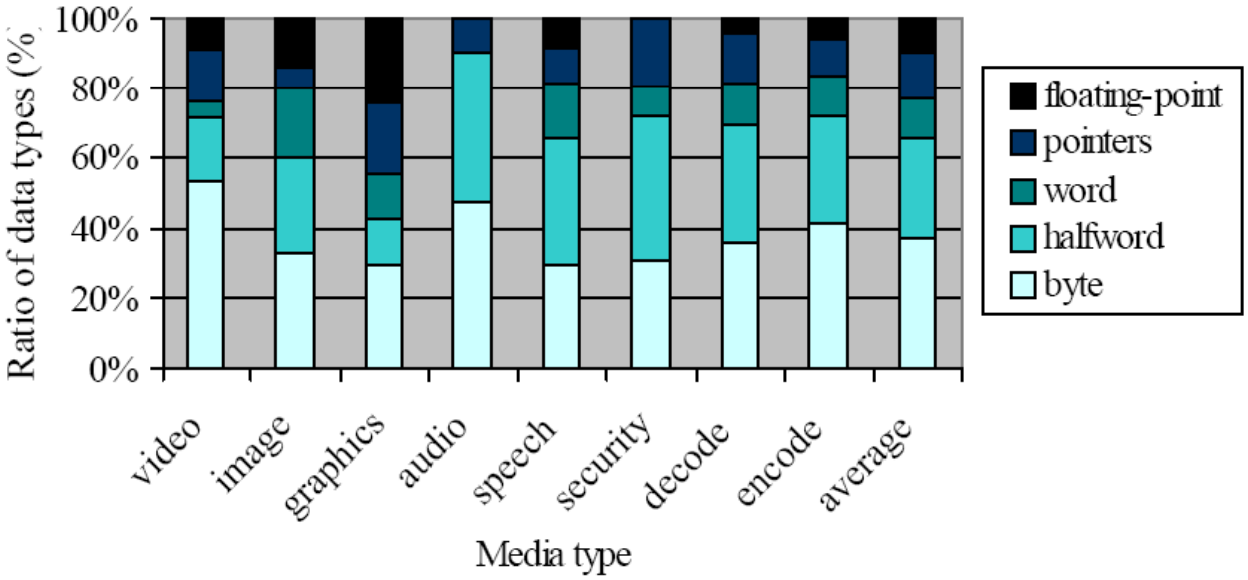
More than 10

- average path ratio

78%



Operand characteristics in MediaBench



Operand characteristics in Video Codecs

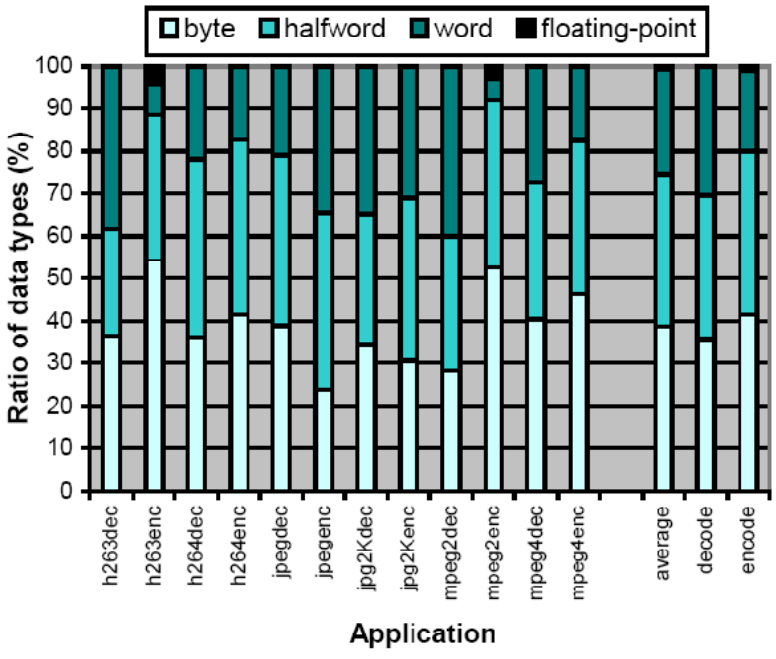
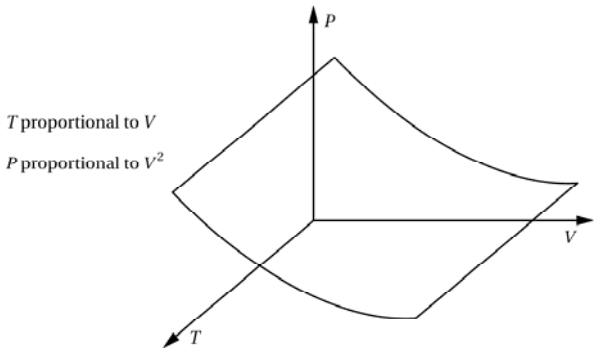


Figure 4. Ratio of data types according to video type.

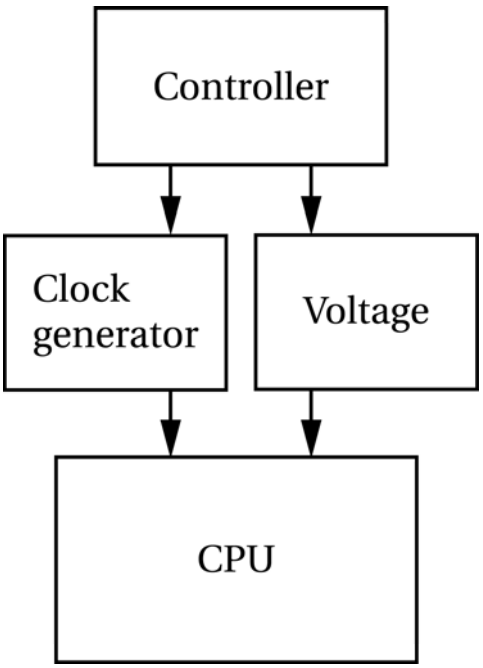
Dynamic voltage scaling (DVS)

- Power scales with V^2 while performance scales roughly as V.
- Reduce operating voltage, add parallel operating units to make up for lower clock speed.
- DVS doesn't work in high-leakage processors.



Dynamic voltage and frequency scaling (DVFS)

- Scale both voltage and clock frequency.
- Can use control algorithms to match performance to application, reduce power.



Razor architecture

- Used specialized latch to detect errors.
- Recovers only on errors, gains average-case performance.

Razor architecture

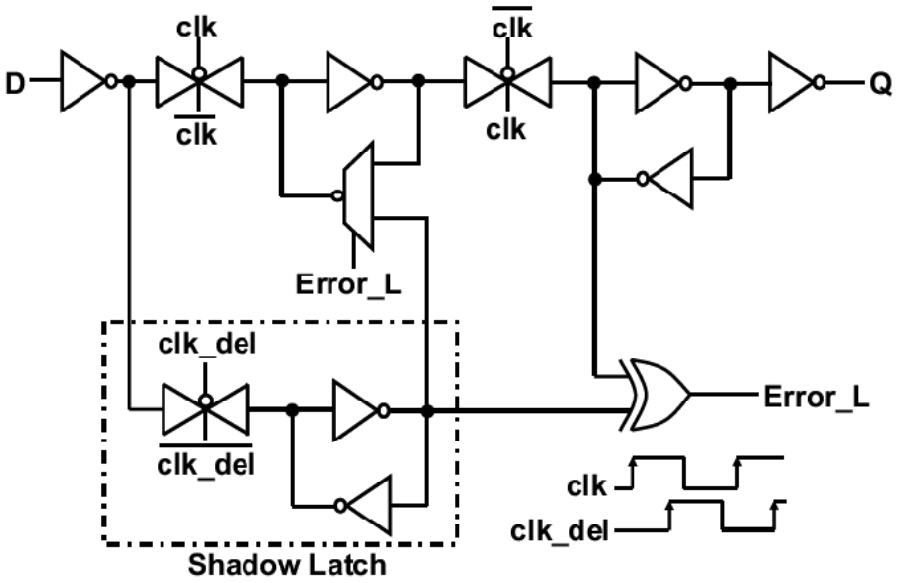


Fig. 2. Circuit schematic of double sampling flip-flop.