**HELSINGIN YLIOPISTO HELSINGFORS UNIVERSITET UNIVERSITY OF HELSINKI** 

#### **CPU Examples & RISC**

Ch 12.5-6 [Sta10]

x86/ARM

Ch 13 [Sta10] Instruction analysis RISC vs. CISC Register use Load 1A M IC - IA + IB

400P

Add Store Branch X NOOP

MERC

Lecture 8



# X86 architecture (e.g., Pentium)



#### (Sta10 Table 12.2)

	(a) Integer Unit in 32-bit Mode						
Туре	Number	Length (bits)	Purpose ESP, EBP, ESI, EDI				
General	8	32	General-purpose user registers				
Segment	6	16	Contain segment selectors CS, SS, DS,				
EFLAGS	1	32	Status and control bits ES, FS, GS				
Instruction Pointer	1	32	Instruction pointer EFLAGS				
			EIP				

#### (b) Integer Unit in 64-bit Mode

Туре	Number	Length (bits)	Purpose
General	16	32	General-purpose user registers
Segment	6	16	Contain segment selectors
RFLAGS	1	64	Status and control bits
Instruction Pointer	1	64	Instruction pointer



### **X86 Processor Registers**

#### (Sta10 Table 12.2)

	(0	c) Floating-Point	Unit	Functions as a FP stack, or store MMX values
Туре	Number	Length (bits)		Purpose
Numeric	8	80	Hold floating	-point numbers /
Control	1	16	Control bits	round, precis, int disable
Status	1	16	Status bits	FP sp, cc, exceptions
Tag Word	1	16	Specifies con	tents of numeric
			registers	fp, mmx, emmx
Instruction Pointer	1	48	Points to inst	ruction interrupted For
			by exception	exception
Data Pointer	1	48	Points to oper	rand interrupted by handler
			exception	support

# ч<u>г</u>і.

## Pentium: FP / MMX Registers (Sta10 Fig 12.24)

Aliasing

- FP regs used as stack
  - Intel 8087 coprocessor (1980) <u>http://en.wikipedia.org/wiki/Intel\_8087</u>
- MMX multimedia instructions use the same registers, but use them directly
- MMX-usage: bits 64-79 are set to 1 → NaN
- FP Tag (word) indicate which usage is current
  - First MMX instr. set
  - EMMS (Empty MMX State) instruction reset





### **Pentium: EFLAGS Register**

#### (Sta10 Fig 12.22)

31	/21					16	/15												0
	Ι	V	V	A	V	R		Ν	ю	0	D	Ι	Τ	S	Ζ	Α	P	(	7
	D	P	F	С	М	F		T	PL	F	F	F	F	F	F	F	F	I	2/

- ID Identification flag VIP Virtual interrupt pending = Virtual interrupt flag VIF AC = Alignment check = Virtual 8086 mode VM RF = Resume flag NT Nested task flag IOPL = I/O privilege level Overflow flag OF
- Condition of the processor: carry, parity, auxiliary, zero, sign, and overflow
  - Used in conditional branches



### **Pentium: Control Registers**





### **Pentium: Interrupts**

Calling interrupt handler; atomic hardware functionality! If not in privileged mode (*etuoikeutettu tila*)

PUSH(SS) stack segment selector to stack

PUSH(ESP) stack pointer to stack

PUSH(EFLAGS) status register to stack

EFLAGS.IOPL ← 00 set privileged mode

as subroutine call

EFLAGS.IF ← 0 disable interrupts (*keskeytys*)

EFLAGS.TF ← 0 disable exceptions (*poikkeus*)

PUSH(CS) code segment selector to stack

PUSH(EIP) instruction pointer to stack (käskyosoitin)

PUSH(error code) if needed

number ← interrupt controller / INT-instruction / status register

CS ← interrupt vector [number].CS

EIP ← interrupt vector [number].EIP

Return

- Privileged IRET-instruction
- POP everything from stack to their places

Address translation: Segment number and offset from interrupt vector => Address of the interrupt handler

Vector Number	Description
0	Divide error; division overflow or division by zero
1	Debug exception; includes various faults and traps related to debugging
2	NMI pin interrupt; signal on NMI pin Nonmaskable interrupt
3	Breakpoint; caused by INT 3 instruction, which is a 1-byte instruction useful for debugging
4	INTO-detected overflow; occurs when the processor executes INTO with the OF flag set
5	BOUND range exceeded; the BOUND instruction compares a register with boundaries stored in memory and generates an interrupt if the contents of the register is out of bounds.
6	Undefined opcode
7	Device not available; attempt to use ESC or WAIT instruction fails due to lack of external device
8	Double fault; two interrupts occur during the same instruction and cannot be handled serially
9	Reserved
10	Invalid task state segment; segment describing a requested task is not initialized or not valid
11	Segment not present; required segment not present
12	Stack fault; limit of stack segment exceeded or stack segment not present
13	General protection; protection violation that does not cause another exception (e.g., writing to a read-only segment)
14	Page fault
15	Reserved
16	Floating-point error; generated by a floating-point arithmetic instruction
17	Alignment check; access to a word stored at an odd byte address or a doubleword stored at an address not a multiple of 4
18	Machine check; model specific
19-31	Reserved
32-255	User interrupt vectors; provided when INTR signal is activated Maskable interrupt

Unshaded: exceptions Shaded: interrupts Computer Organization II, Autumn 2010, Teemu Kerola

(Sta10 Table 12.3)





## ARM (Ch 12.6 Sta10)



### **ARM features**

- Array of uniform registers (moderate number)
- Fixed length (32 bit) instruction (Thumb 16 bit)
- Load/Store architecture
- Small number of addressing modes (reg + instr. field)
- Autoincrement addressing mode (for program loops)
- Data processing instructions allow shift or rotate to preprocess one of source regs
  - Separate ALU and shifter for this purpose (avoid structural dependency or hazard)
- Conditional execution of instructions
  - Fewer conditional branches, improves pipeline efficiency





### **Processor execution modes**

- User mode
  - No access to protected system resources, can cause exception
- Supervisor mode
  - For OS, starts with <u>software interrupt</u> instruction
  - Abort mode due to memory faults
  - Undefined mode instruction not supported

Exception modes —

- Fast interrupt mode
  - Interrupt from designated fast interrupt source
  - Not interruptable, can interrupt normal interrupt
- Interrupt mode
  - Any other interrupt signal, can be interrupted by fast interrupt
- System mode
  - Only for certain priviledged OS tasks

	Modes									
				Privilege	d modes					
Register			Exception modes							
organization	User	System	Supervisor	Abort	Undefined	Interrupt	Fast Interrupt			
SP – stack pointer	R0	R0	R0	R0	R0	R0	R0			
LR – <u>link register</u>	R1	R1	R1	R1	R1	R1	R1			
(return address&mode)	R2	R2	R2	R2	R2	R2	R2			
PC – program counter	R3	R3	R3	R3	R3	R3	R3			
	R4	R4	R4	R4	R4	R4	R4			
CPSR – current	R5	R5	R5	R5	R5	R5	R5			
program status register	R6	R6	R6	R6	R6	R6	R6			
SPSR – saved	<b>R</b> 7	<b>R</b> 7	<b>R</b> 7	<b>R</b> 7	<b>R</b> 7	<b>R</b> 7	<b>R</b> 7			
program status register	R8	R8	R8	R8	R8	R8	R8_fiq			
	R9	R9	R9	R9	R9	R9	R9_fiq			
	R10	R10	R10	R10	R10	R10	R10_fiq			
Shaded regs replaced	R11	R11	R11	R11	R11	R11	R11_fiq			
in exception modes!	R12	R12	R12	R12	R12	<b>R1</b> 2	R12_fiq			
-	R13 (SP)	R13 (SP)	R13_svc	R13_abt	R13_und	R13_irq	R13_fiq			
	R14 (LR)	R14 (LR)	R14_svc	R14_abt	R14_und	R14_irq	R14_fiq			
	R15 (PC)	R15 (PC)	R15 (PC)	R15 (PC)	R15 (PC)	R15 (PC)	R15 (PC)			

CPSR	CPSR	CPSR	CPSR	CPSR	CPSR	CPSR
		SPSR_svc	SPSR_abt	SPSR_und	SPSR_irq	SPSR_fiq



Discussion? 24.11.2010 14



31 30 29 28 27 26 25 24 23 22 21 20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 0



User flags

- N,Z,C,V condition code
- Q overflow or saturation in SIMD-orient. instr.
- J Jazelle instruction in use
  - "Java byte code mode"
- GE[3:0] for SIMD as greater than or equal flags for individual bytes or halfwords of the result

System control flags

- E endianness in load/store
- A,I,F interrupt disable bits (A imprecise data aborts, I – normal IRQ, F – fast FIQ)
- T normal / Thumb instr.
- M[4:0] processor mode



ARM Interrupt vector

Table lists the exception types and the address in interrupt vector for that type.

The vector contains the start addresses of the interrupt handlers.

Exception type	Processor Mode	Normal entry address	Description
Reset	Supervisor	0x0000000x0	Occurs when the system is initialized.
Data abort	Abort	0x0000010	Occurs when an invalid memory address has been accessed, such as if there is no physical memory for an address or the correct access permission is lacking.
FIQ (fast interrupt)	FIQ	0x000001C	Occurs when an external device asserts the FIQ pin on the processor. An interrupt cannot be interrupted except by an FIQ. FIQ is designed to support a data transfer or channel process, and has sufficient private registers to remove the need for register saving in such applications, therefore minimizing the overhead of context switching. A fast interrupt cannot be interrupted.
IRQ (interrupt)	IRQ	0x0000018	Occurs when an external device asserts the IRQ pin on the processor. An interrupt cannot be interrupted except by an FIQ.
Prefetch abort	Abort	0x000000C	Occurs when an attempt to fetch an instruction results in a memory fault. The exception is raised when the instruction enters the execute stage of the pipeline.
Undefined instructions	Undefined	0x00000004	Occurs when an instruction not in the instruction set reaches the execute stage of the pipeline.
Software interrupt (Sta10 Table 12.4	Supervisor	0x0000008	Generally used to allow user mode programs to call the OS. The user program executes a SWI instruction with an argument that identifies the function the user wishes to perform.



# RISCarchitecture



#### Ch 13 [Sta10]

- Instructions
- RISC vs. CISC
- Register allocation





### **CISC - Complex Instruction Set Computer**

- Goal: Shrink the **semantic gap** (semanttinen kuilu) between high-level language and machine instruction set
  - Expressiveness of high-level languages had increased
  - Wanted "simple" compilations
    - Language structures match nicely with instructions
  - Lots of different instructions for different purposes
  - Lots of different data types (int, float, char, boolean, ...)
  - Lots of different addressing modes
  - Complex tasks performed in hardware by control unit (single instruction), not in the machine code level (multiple instructions)
    - Less instructions in one program (shorter code)
    - Efficient (just a few instructions) execution of complex tasks

### Which Operations and Operands Are Used?

Year 1982, computers VAX, PDP-11, Motorola 68000

Observe dynamic execution time behaviour

	Dynamic O	Occurrence	Machine-I Weig	nstruction ted	Memory-Reference Weighted			
	Pascal	С	Pascal	С	Pascal	С		
ASSIGN	45%	38%	13%	13%	14%	15%		
LOOP	5%	3%	42%	32%	33%	26%		
CALL	15%	12%	31%	33%	44%	45%		
IF	29%	43%	11%	21%	7%	13%		
GOTO	_	3%	_	_	_	_		
OTHER	6%	1%	3%	1%	2%	1%		

Weighted Relative Dynamic Frequency of HLL Operations [PATT82a]

				(HLL=High Level La	nguage)
	Pascal	С	Average	(IIIII IIIgii Level La	
					80% of references
Integer Constant	16%	23%	20%	Dynamia Danaantaga	to local variables
Scalar Variable	58%	53%	55%	of Operands	to <u>local</u> variables
Array/Structure	26%	24%	25%	()	Sta10 Table 12 2 12 2

(Sta10 Table 13.2, 13.3)



### Subroutine (procedure, function) calls?

- Lots of subroutine calls
- Calls rarely have many parameters
- Nested (sisäkkäinen) calls are rare

(Sta10 Table 13.4)

Percentage of Executed Procedure Calls With	Compiler, Interpreter, and Typesetter	Small Nonnumeric Programs
>3 arguments	0–7%	0–5%
>5 arguments	0–3%	0%
>8 words of arguments and local scalars	1–20%	0–6%
>12 words of arguments and local scalars	1-6%	0–3%

#### **Procedure Arguments and Local Scalar Variables**

98% less than 6 parameters

92% less than 6 local variables

How to use the information?



### **Observations from Real Programs**

- Most operands are simple
- Many jumps and branches
- Compilers do not always use the complex instructions
  - They use only a subset of the instruction set
  - Easier to do? Faster?
- Conclusion?

Occam's razor (Occamin partaveitsi)

"Entia non sunt multiplicanda praeter necessitatem" ("Entities should not be multiplied more than necessary") William Of Occam (1300-1349) English monk, philosopher

"It is vain to do with more that which can be done with less"



### **Optimize for Execution Speed**

- Optimize the parts that consume most of the time
  - Procedure calls, loops, memory references, addressing, ...
- Avoid optimizing rare events
  - Rarely used (10%) floating point instructions improved to run 2x:

$$\boxed{\begin{array}{c} \text{No speedup} \\ \text{ExTime}_{\text{new}} = \text{ExTime}_{\text{old}} * (0.9 * 1.0 + 0.1 * 0.5) \\ = 0.95 \text{ x ExTime}_{\text{old}} \end{array}}$$

**Speedup** =  $ExTime_{old} / ExTime_{new} = 1 / 0.95 = 1.053 << 2$ 

#### Amdahl's law

Speedup due to an enhancement is proportional to the fraction of the time (in the original system) that the enhancement can be used.



Gene Amdahl



### **RISC Approach**

- Optimize design for <u>execution speed</u>, instead of ease of compilation
  - Compilers are good, machines are efficient
    - Compiler can and has time to do the optimization
  - Do most important, common things in hardware and fast
    - E.g. 1-dim array reference
    - One machine instruction
  - And the rest in software (and slow)
    - E.g. multidimensional arrays, string processing, ...
    - Library routines for these
    - Many machine instructions

#### ⇒ **RISC architecture** (Reduced Instruction Set Computer)



### **RISC** architecture

- Plenty of registers (minimum 32)
  - Compilers optimize register usage
- LOAD / STORE architecture
  - Only LOAD and STORE do memory referencing
- Small set of simple instructions
- Simple, fixed-length instruction format (32b)
  - Instruction fetch and decoding simple and efficient
- Small selection of simple address references
  - No indirect memory reference
  - Fast address translation
- Limited set of different operands
  - 32b integers, floating-point
- One or more instructions are <u>completed</u> on each cycle



### **RISC** architecture

- CPU easier to implement
  - Pipeline control and optimization simpler
  - Hardwired (*langoitettu*) control
- Smaller chip (*lastu*) size
  - More chips per die (*kiekko*)
  - Smaller waste%
- Cheaper manufacturing
- Faster marketing
  - Design-to-market time



http://www.gadget-paradise.com/0127/samsung/computer/64gb-flash-memory/



![](_page_26_Picture_0.jpeg)

### **RISC vs. CISC**

	Comp (C	lex Instructi ISC)Comput	on Set ter	Reduced In Set (RISC)	nstruction Computer		Superscalar		
Characteristic	IBM 370/168	VAX 11/780	Inte1 80486	SPARC	MIPS R4000	PowerPC	Ultra SPARC	MIPS R10000	
Year developed	1973	1978	1989	1987	1991	1993	1996	1996	
Number of instructions	208	303	235	69	94	225			
Instruction size (bytes)	2–6	2–57	1–11	4	4	4	4	4	
Addressing modes	4	22	11	1	1	2	1	1	
Number of general- purpose registers	16	16	8	40 - 520	32	32	40 - 520	32	
Control memory size (Kbits)	420	480	246	-	_	-	—	_	
Cache size (KBytes)	64	64	8	32	128	16-32	32	64	

Characteristics of Some CISCs, RISCs, and Superscalar Processors

(Sta10 Table 13.1)

![](_page_27_Picture_0.jpeg)

### RISC (dark) vs. CISC (white background)

Processor	Number of instruc- tion sizes	Ma <b>x</b> instruc- tion size in bytes	Number of addressing modes	Indirect addressing	Load/store combined with arithmetic	Ma <b>x</b> number of memory operands	Unaligned addressing allowed	Ma <b>x</b> Number of MMU uses	Number of bits for integer register specifier	Number of bits for FP register specifier
AMD29000	1	4	1	no	no	1	no	1	8	3°
MIPS R2000	1	4	1	no	no	1	no	1	5	4
SPARC	1	4	2	no	no	1	no	1	5	4
MC88000	1	4	3	no	no	1	no	1	5	4
HP PA	1	4	10 "	no	no	1	no	1	5	4
IBM RT/PC	2ª	4	1	no	no	1	no	1	4 <sup>a</sup>	3 °
IBM RS/6000	1	4	4	no	no	1	yes	1	5	5
Intel i860	1	4	4	no	no	1	no	1	5	4
IBM 3090	4	8	2 <sup>b</sup>	no <sup>b</sup>	yes	2	yes	4	4	2
Intel 80486	12	12	15	no <sup>b</sup>	yes	2	yes	4	3	3
NSC 32016	21	21	23	yes	yes	2	yes	4	3	3
MC68040	11	22	44	yes	yes	2	yes	8	4	3
VAX	56	56	22	yes	yes	6	yes	24	4	0
Clipper	4ª	8 ª	9ª	no	no	1	0	2	4 ª	3°
Intel 80960	2ª	8 ª	9ª	no	no	1	yes <sup>a</sup>	-	5	3 <i>°</i>

a RISC that does not conform to this characteristic.

b CISC that does not conform to this characteristic.

(Sta10 Table 13.7)

![](_page_28_Picture_0.jpeg)

# **Register Files**

![](_page_29_Picture_0.jpeg)

### **Register Window to Register File**

- More physical registers than addressable in the instruction
  - E.g., SPARC has just 5 bits for register number → 0.. 31, but the processor has 40 to 540 registers
- Small subset of registers available for each instruction in register window
  - In the window references to register r0-r31
  - CPU maps them to actual (true) registers r0-r539

![](_page_29_Figure_7.jpeg)

![](_page_29_Figure_8.jpeg)

![](_page_30_Picture_0.jpeg)

### **Overlapped Register Windows**

Procedure parameters passed in registers (not in stack)

- Fixed number of registers for parameters, local variables, and return value passed via overlapped register window
- Overlapping area to allow parameter passing to the next procedure and back to caller

![](_page_30_Figure_5.jpeg)

![](_page_31_Picture_0.jpeg)

### **Circural Buffer for Overlapped Register Window**

(Sta10 Fig 13.2) Too many nested calls? Restore Save Most recent calls in registers B.in B.loc A.loc Older activations A.in saved to memory C.in Saved w0 window Restore when nesting pointer w5 w2 C.loc (F) depth decreases w4 w3 Overlap only when needed D.in **Global variables?** 11111 (E) D.loc 11111 In memory or own register window 111111 ////// Current SPARC window Real register names pointer r0-r7 global var. Call r8-r15 parameters (in caller) Virtual Return r16-r23 local variables register names r24-r31 parameters (to called) Discussion?

![](_page_32_Picture_0.jpeg)

### **Register File vs. Cache**

#### (Sta10 Table 13.5)

Large Register File	Cache			
All local scalars	Recently-used local scalars			
Individual variables	Blocks of memory			
Compiler-assigned global variables	Recently-used global variables			
Save/Restore based on procedure nesting depth	Save/Restore based on cache replacement algorithm			
Register addressing Number of bits	Memory addressing			

- The register file acts like a small, fast buffer (as cache?)
  - Register is faster, needs less bits in addressing, but
- It is difficult for compiler to determine in advance,
  - which of the global variables to place in registers
- Cache decides this issue dynamically
  - Most used and referenced data stay in cache

![](_page_33_Picture_0.jpeg)

### **Compiler-based register optimization** (allocation of registers)

- Problem: Graph coloring
  - Minimize the number of different colors, while adjacent nodes have different color
- Difficult problem (NP-complete)

![](_page_33_Picture_5.jpeg)

See course on Models of Computation

- Form a network of symbolic registers based on the program code
  - Symbolic register~ any program quantity that could be in register
  - The edges of the graph join together program quantities that are used in the same code fragment
- Allocate real registers based on the graph
  - Two symbolic registers that are not used at the same time (no edge between them) can be allocated to the same real register (use the same color)
  - If there are no more free registers, use memory addresses

![](_page_34_Picture_0.jpeg)

### Allocation of registers (compiler-based register optimization)

- Node (solmu) = symbolic register
- Edge (särmä) = symbolic registers used at the same time
- *n* colors = *n* registers

![](_page_34_Figure_5.jpeg)

(a) This sequence of acure use of registers (b) Register interference g

So, use the same physical register for A and D, and for C and E.

![](_page_35_Figure_0.jpeg)

![](_page_36_Picture_0.jpeg)

### **RISC-pipeline, Delayed Branch**

-	1 1		2	4	I	6	7	1	Forget dependency			
		2	3 D	4	2	0	/		problem here,			
100 LOAD X, rA	1	E	D						annoantrata on jumn			
101 ADD 1, rA		Ι	E	Bubb	le? ←				concentrate on jump!			
102 JUMP 105			Ι	E								
103 ADD rA, rB				Ι	E			Traditional pipeline clear pipeline				
105 STORE rA, Z					Ι	E	D					
				•				•				
100 LOAD X, rA	Ι	E	D									
101 ADD 1, rA		Ι	Е	Bubb	le?			RISC with inserted NOOP				
102 JUMP 106			Ι	E				Two port MEM				
103 NOOP				Ι	E			No need to clear pipeline (NOOP)				
106 STORE rA, Z					Ι	Е	D					
1001.040.8	-			_			-	RIS	C with reversed instructions			
100 LOAD X, Ar	1	E	E D						l lse of delay slot			
101 JUMP 105		Ι	E				R					
102 ADD 1, rA			Ι	E			] `	What	hat if conditional branch?			
105 STORE rA, Z				Ι	E	D		JZEK	<i>CO 105, rA</i>			
	L					-		(need	ADD 1,rA result before			
								comp	parison, cannot use delay slot)			

#### (Sta10 Fig 13.7)

Extra gain: Dependency problem also solved!

![](_page_37_Picture_0.jpeg)

### **RISC & CISC United?**

Pentium, CISC

'compilation' at every execution

 Each 1–11 byte-length CISC-instruction is 'translated' by hardware to 1-4 118-bit micro-operations (stored in L1 instruction cache)

http://www.cs.clemson.edu/~mark/330/colwell/pentium.gif

- Lower levels (including control unit) as RISC
- Lots of work registers, visible only to hardware
- Crusoe (Transmeta)

Just in time (JIT) compilation

- Emulate Intel architecture with simpler HW architecture
- Outside looks like Intel CISC-architecture
- Group of instructions 'translated' by software, just before execution, to fixed-length micro-operations; these can be optimized before execution
  - VLIW (very long instruction word, 128 bits)
  - 4 µops/VLIW-instruction
- Lower levels as RISC

'compilation' just once per group

![](_page_38_Picture_0.jpeg)

### Summary

- X86 and ARM processor implementation examples
  - Registers, addressing modes, instruction sets
- What is CISC? What is "wrong" with CISC
- What is RISC? What is "good" with RISC?
  - Lots of registers, load-store arch
  - Small set of simple instructions with just a few operand types
  - Simple instruction formats and addressing formats
- How to get more from HW registers?
  - Register windows to register file
  - Overlapping register windows
  - Register file vs. cache?
  - Register allocation problem and its solution

Combine RISC with CISC?

![](_page_39_Picture_0.jpeg)

- Main features and characteristics of RISC-architecture?
- What makes RISC RISC?
- Which addressing format is not RISC?
- Which operation type is not RISC?
- Which instruction format is not RISC?
- Which operand type is not RISC?
- Why would large L1 cache be better than large register file?
- How are register windows used?
  - When would *n* overlapped registers be enough?
  - What happens if *n* overlapped registers is not enough?