# CS3521 : Computer Systems Architecture 

## Lecture 5: MIPS Integer ALU

September 10, 2009

## Integer Addition

■ Example: $7+6$


■ Overflow if result out of range

- Adding +ve and -ve operands, no overflow
- Adding two +ve operands
- Overflow if result sign is 1
- Adding two - ve operands

■ Overflow if result sign is 0

## Integer Subtraction

- Add negation of second operand
- Example: $7-6=7+(-6)$

| $+7:$ | $00000000 \ldots 00000111$ |
| :--- | :--- |
| $-6:$ | $\underline{11111111 \ldots 11111010}$ |
| $+1:$ | $00000000 \ldots 00000001$ |

- Overflow if result out of range
- Subtracting two +ve or two -ve operands, no overflow
- Subtracting +ve from -ve operand
- Overflow if result sign is 0
- Subtracting -ve from + ve operand
- Overflow if result sign is 1


## Dealing with Overflow

- Some languages (e.g., C) ignore overflow
- Use MIPS addu, addui, subu instructions

■ Other languages (e.g., Ada, Fortran) require raising an exception
■ Use MIPS add, addi, sub instructions
■ On overflow, invoke exception handler
$\square$ Save PC in exception program counter (EPC) register
■ Jump to predefined handler address
$\square \mathrm{mfc}$ ( (move from coprocessor reg) instruction can retrieve EPC value, to return after corrective action

## Throw Hardware At It: Money is no Object!

Using $\mathrm{c}_{\mathrm{i}}$ for Carry $\mathrm{In}_{\mathrm{i}}$

$$
c_{2}=b_{1} c_{1}+a_{1} c_{1}+a_{1} b_{1}
$$

and

$$
\mathrm{c}_{1}=\mathrm{b}_{0} \mathrm{c}_{0}+\mathrm{a}_{0} \mathrm{c}_{0}+\mathrm{a}_{0} \mathrm{~b}_{0}
$$

Substituting for $\mathrm{c}_{1}$, we get:

$$
c_{2}=a_{1} a_{0} b_{0}+a_{1} a_{0} c_{0}+a_{1} b_{0} c_{0}+b_{1} a_{0} b_{0}+b_{1} a_{0} c_{0}+b_{1} b_{0} c_{0}+a_{1} b_{1}
$$

Continuing this to 32 bits yields a fast, but unreasonably expensive adder
Just how fast?
Assume all gate delays are the same regardless of fan-in

## Carry-Lookahead Adders

- The basic formula can be rewritten:
$\square c_{i+1}=b_{i} c_{i}+a_{i} c_{i}+a_{i} b_{i}$
- $\mathrm{c}_{\mathrm{i}+1}=\mathrm{a}_{\mathrm{i}} \mathrm{b}_{\mathrm{i}}+\left(\mathrm{a}_{\mathrm{i}}+\mathrm{b}_{\mathrm{i}}\right) \mathrm{c}_{\mathrm{i}}$
- Applying it to $\mathrm{c}_{2}$, we get:
$\square \mathrm{c}_{2}=\mathrm{a}_{1} \mathrm{~b}_{1}+\left(\mathrm{a}_{1}+\mathrm{b}_{1}\right)\left(\mathrm{a}_{0} \mathrm{~b}_{0}+\left(\mathrm{a}_{0}+\mathrm{b}_{0}\right) \mathrm{c}_{0}\right)$
■ Define two "signals" or abstractions:
- Generate: $\mathrm{g}_{\mathrm{i}}=\mathrm{a}_{\mathrm{i}} * \mathrm{~b}_{\mathrm{i}}$
- Propagate: $\mathrm{p}_{\mathrm{i}}=\mathrm{a}_{\mathrm{i}}+\mathrm{b}_{\mathrm{i}}$

■ Redefine $\mathrm{c}_{\mathrm{i}+1}$ as:
■ $c_{i+1}=g_{i}+p_{i} * c_{i}$
$\square$ So $c_{i+1}=1$ if

- $g_{i}=1$ (generate) or
- $\mathrm{p}_{\mathrm{i}}=1$ and $\mathrm{c}_{\mathrm{i}}=1$ (propagate)


## Carry-Lookahead Adders

- Our logic equations are simpler:

■ $\mathrm{c}_{1}=\mathrm{g}_{0}+\mathrm{p}_{0} \mathrm{c}_{0}$
■ $\mathrm{c}_{2}=\mathrm{g}_{1}+\mathrm{p}_{1} \mathrm{~g}_{0}+\mathrm{p}_{1} \mathrm{p}_{0} \mathrm{c}_{0}$

- $\mathrm{c}_{3}=\mathrm{g}_{2}+\mathrm{p}_{2} \mathrm{~g}_{1}+\mathrm{p}_{2} \mathrm{p}_{1} \mathrm{~g}_{0}+\mathrm{p}_{2} \mathrm{p}_{1} \mathrm{p}_{0} \mathrm{c}_{0}$
$\square \mathrm{c}_{4}=\mathrm{g}_{3}+\mathrm{p}_{3} \mathrm{~g}_{2}+\mathrm{p}_{3} \mathrm{p}_{2} \mathrm{~g}_{1}+\mathrm{p}_{3} \mathrm{p}_{2} \mathrm{p}_{1} \mathrm{~g}_{0}+\mathrm{p}_{3} \mathrm{p}_{2} \mathrm{p}_{1} \mathrm{p}_{0} \mathrm{c}_{0}$


## Carry-Lookahead Adders



University of Texas at Austin CS352H - Computer Systems Architecture Fall 2009 Don Fussell 8

## Carry-Lookahead Adders

■ How much better (16-bit adder)?
■ Ripple-carry: $16 * \mathrm{~T}_{\text {add }}=16 * 2=32$ gate delays
■ Carry-lookahead: $\mathrm{T}_{\text {add }}+\max \left(\mathrm{p}_{\mathrm{i}}, \mathrm{g}_{\mathrm{i}}\right)=2+2=4$

- Much better, but still too profligate

■ What if we apply another level of this abstraction?

- Use the four-bit adder on the previous slide as a building block
- Define P and G signals
- $\mathrm{P}_{0}=\mathrm{p}_{3} \mathrm{p}_{2} \mathrm{p}_{1} \mathrm{p}_{0}$
$\square \mathrm{G}_{0}=\mathrm{g}_{3}+\mathrm{p}_{3} \mathrm{~g}_{2}+\mathrm{p}_{3} \mathrm{p}_{2} \mathrm{~g}_{1}+\mathrm{p}_{3} \mathrm{p}_{2} \mathrm{p}_{1} \mathrm{~g}_{0}$
- Similarly for $\mathrm{P}_{1}-\mathrm{P}_{3}$ and $\mathrm{G}_{1}-\mathrm{G}_{3}$
- Derive equations for $\mathrm{C}_{1}-\mathrm{C}_{4}$
$\square \mathrm{C}_{1}=\mathrm{G}_{0}+\mathrm{P}_{0} \mathrm{c}_{0}$
$\square \mathrm{C}_{2}=\mathrm{G}_{1}+\mathrm{P}_{1} \mathrm{G}_{0}+\mathrm{P}_{1} \mathrm{P}_{0} \mathrm{c}_{0}$, etc.
- See discussion in Appendix C. 6


## Carry-Lookahead Adders



16-bit adder performance $=\mathrm{T}_{\text {add }}+\max \left(\mathrm{P}_{\mathrm{i}}, \mathrm{G}_{\mathrm{i}}\right)=2+2+1=5$ (with thrifty hardware)

## Arithmetic for Multimedia

- Graphics and media processing operates on vectors of 8-bit and 16-bit data

■ Use 64-bit adder, with partitioned carry chain
$■$ Operate on $8 \times 8$-bit, $4 \times 16$-bit, or $2 \times 32$-bit vectors

- SIMD (single-instruction, multiple-data)

■ Saturating operations
■ On overflow, result is largest representable value
$\square$ c.f. 2 s -complement modulo arithmetic

- E.g., clipping in audio, saturation in video


## Shifters

■ Two kinds:

- Logical: value shifted in is always "0"
$" 0 " \longrightarrow$ msb $\quad$ Isb $\longleftarrow " 0 "$

■ Arithmetic: sign-extend on right shifts
$\longrightarrow \mathrm{msb} \quad \mathrm{Isb} \longleftarrow \mathrm{O}=0$

■ What about n-bit, rather than 1-bit, shifts?
Want a fast shifter

## Combinatorial Shifter from MUXes



## Unsigned Multiplication

■ Paper and pencil example (unsigned):

- Multiplicand
- Multiplier

|  |  | 1000 |
| :---: | :---: | :---: |
|  |  | 1001 |
|  |  | 1000 |
|  |  | 000 |
|  | 0 |  |
| 10 | 0 |  |
| 010 | 0 | 1000 |

- m bits x n bits $=\mathrm{m}+\mathrm{n}$ bit product

■ Binary makes it easy:

- 0: place 0
(0 x multiplicand)
■ 1: place copy
( $1 \times$ multiplicand)


## Unsigned Combinatorial Multiplier



Stage i accumulates $A$ * $2^{i}$ if $B_{i}==1$

## How Does it Work?



- At each stage shift A left (x2)
- Use next bit of B to determine whether to add in shifted multiplicand
- Accumulate 2 n bit partial product at each stage


## Sequential Multiplication Hardware



University of Texas at Austin CS352H - Computer Systems Architecture Fall 2009 Don Fussell

## Observations

■ One clock per multiply cycle

- $~ 32$ clock cycles per integer multiply

■ Vs. one cycle for an add/subtract
■ Half of the bits in the multiplicand are always zero

- 64-bit adder is wasted
- Zeros inserted in left of multiplicand as shifted

■ Least significant bits of product unchanged once formed

■ Instead of shifting multiplicand to left, shift product to right!

## Optimized Multiplier

■ Perform steps in parallel: add/shift


■ One cycle per partial-product addition

- That's ok, if frequency of multiplications is low


## Faster Multiplier

■ Uses multiple adders

- Cost/performance tradeoff

- Can be pipelined
- Several multiplication performed in parallel


## MIPS Multiplication

- Two 32-bit registers for product
- HI: most-significant 32 bits
- LO: least-significant 32-bits
- Instructions
- mult rs, rt / multu rs, rt
- 64-bit product in HI/LO
- mfhi rd / mflo rd
- Move from HI/LO to rd
- Can test HI value to see if product overflows 32 bits
- mul rd, rs, rt
- Least-significant 32 bits of product -> rd


## Division


$n$-bit operands yield $n$-bit quotient and remainder

## Division Hardware



## Optimized Divider



- One cycle per partial-remainder subtraction
- Looks a lot like a multiplier!
- Same hardware can be used for both


## Faster Division

- Can't use parallel hardware as in multiplier
- Subtraction is conditional on sign of remainder

■ Faster dividers (e.g. SRT devision) generate multiple quotient bits per step

- Still require multiple steps


## MIPS Division

■ Use HI/LO registers for result
■ HI: 32-bit remainder
■ LO: 32-bit quotient
■ Instructions
■ div rs, rt / divu rs, rt
■ No overflow or divide-by-0 checking
$■$ Software must perform checks if required
■ Use mfhi, mflo to access result

## Next Lecture

■ Floating point

- Rest of Chapter 3

