

Digital Logic Design



CHIP Xor {	
IN a, b;	
OUT out;	
PARTS:	
Not(in=a, out=nota);	· ·
Not(in=b, out=notb);	
And(a=nota, b=b, out=w1);	
And $(a=a, b=notb, out=w2)$:	
Or(a=w1, b=w2, out=out);	
}	
,	



Lecture # 19-20

Registers, Memory and Counters



Slides of first half of the course are adapted from: https://www.nand2tetris.org Download s/w tools required for first half of the course from the following link: https://drive.google.com/file/d/0B9c0BdDJz6XpZUh3X2dPR1o0MUE/view



1110001100001000



Today's Agenda

- Review of Sequential Chips
- What are **Registers**?
- Design and HDL of 1-bit to 16-bit Registers
- Concept of Memory Hierarchy
- Multi-Byte Read/Write
- Design and HDL of Random Access Memory (8, 64, 512, 4K, 16K words)
- What are **Counters**?
- Why do we need Counter for our Hack Computer
- Concept of Program Counter
- Design and Implementation of PC for Hack Computer
- Demo on H/W Simulator



Review of Sequential Chips



- Sequential chips are capable of maintaining state, and, optionally acting on the state, and on the current input
- The simplest and most elementary sequential chip is DFF, which maintain a state, i.e., the value of the input from the previous time unit
- Using DFF we can design registers, and using registers we can design RAM, whose state is the current values of all its registers. Given an address, the RAM emits the value of the selected register
- All combinational chips are constructed from NAND gates, while all sequential chips are constructed from DFF gates, and combinational chips



CPU Registers





- A register is a small memory place inside the CPU that may hold data, memory address or instruction
- The size of registers in a 64-bit computer must be of 64 bits
- In our Hack computer is a 16 bit computer, so the registers we are going to design will be of 16 bits
- There are several different classes of CPU registers which works in coordination with the computer memory to run operations efficiently. (More on it later)



1-Bit Register





- A single-bit register, which we call Bit, or binary cell, is designed to store a single bit of information (0 or 1)
- The chip interface diagram shows that it has two input pins and one output pin. The input pin carries a data bit, the load pin enables the cell for writes, and an output pin that emits the current state of the cell
- When you read the out pin of the binary cell, you will always get whatever is the state of the binary cell
- To write the binary cell, we set the load bit to 1, now what ever is there on the input bit will be stored inside the binary cell and will be available on the out pin in the next clock cycle
- When the load bit is zero, the chip keep remembering the last input that was loaded into it for infinity until a new load operation is performed



- Goal: Remember an input bit forever, until requested to load a new value
- More accurately:
 - Stores a bit until...
 - Instructed to load, and store, another bit

Holling of Juggting

Sequential Chips: 1-Bit Register





































Computers Are Flexible





HDL for 1-bit Register



Bit.hdl

```
/** 1-bit register:

* If load[t] == 1 then. //load input in the ff

* out[t+1] = in[t]

* else //out does not change

*/ (out[t+1] = out[t])

CHIP Bit {

IN in, load;

OUT out;

PARTS:

Mux(a=sendBack, b=in, sel=load, out=MuxOut);

DFF(in=MuxOut, out=sendBack, out=out);
```







Multi-Bit Register



Multi-Bit Register



- A register is actually a group of flip-flops, each flip flop capable of storing one bit of information
- An n-bit register consists of a group of n flip-flops capable of storing n bits of binary information. In this course we will focus on designing of 16-bit registers for our computer
- A 16 bit register can be created from an array of 16 1-bit registers



16 Bit Register API

- The API of the 16 bit Register chip is essentially the same as the 1-bit register, except that the input and output pins are designed to handle multi-bit values
- The interface diagram and API of a 16-bit register is shown below
- The Bit and Register chips have exactly the same read/write behavior:
 - **Read:** To read the contents of a register, we simply probe its output
 - Write: To write a new data value **d** into a register, we put **d** in the **in** input and set the load input to 1. In the next clock cycle, the register commits to the new data value, and its output starts emitting **d**, and it will keep emitting this new value forever till the time we decide to write a new value in it



HDL for 16-bit Regist load load in out DFF Bit Bit out Bit in sendBack W **Register.hdl** /** * 16-bit register: * If load[t] == 1 then out[t+1] = in[t]* else out does not change CHIP Bit (IN in, load; */ OUT out; CHIP Register { PARTS: Mux(a=sendBack, b=in, sel=load, out=MuxOut); IN in[16], load; DFF(in=MuxOut, out=sendBack, out=out); OUT out[16]; PARTS: Bit(in=in[0], load=load, out=out[0]); Bit(in=in[1], load=load, out=out[1]); Bit(in=in[2], load=load, out=out[2]); Bit(in=in[3], load=load, out=out[3]); Bit(in=in[15], load=load, out=out[15]);

Instructor: Muhammad Arif Butt, Ph.D.







Class Quiz:









Memory Overview



Stored Program Concept





Memory Hierarchy

- Accessing a memory location is expensive, as we need to supply an address and then read/write the contents of that location of memory. Moreover, moving the memory contents into the CPU and vice versa also takes time
- Solution: Memory Hierarchy





Memory Unit	Example Size	Typical Speed
Registers	16, 64-bit registers	1 nanosecond
Cache memory	4 - 8 Megabytes (L1 and L2)	5-60 nanoseconds
Primary Storage	2 - 32+ Gigabytes	100-150 nanoseconds
Secondary Storage	500 Gigabytes – 4+ Terabytes	3-15 milliseconds



Memory Sizes/Capacity

Decimal Term	Abbreviation	Value	Binary Term	Abbreviation	Value	%Larger
kilobyte	KB	10 ³	kibibyte	KiB	2 ¹⁰	2%
megabyte	MB	10 ⁶	mebibyte	MiB	2 ²⁰	5%
Gigabyte	GB	10 ⁹	gibibyte	GiB	2 ³⁰	7%
terabyte	ТВ	1012	tebibyte	TiB	2 ⁴⁰	10%`
petabyte	PB	10 ¹⁵	pebibyte	PiB	2 ⁵⁰	13%
exabyte	EB	1018	exabibyte	EiB	2 ⁶⁰	15%
zettabyte	ZB	10 ²¹	zebibyte	ZiB	2 ⁷⁰	18%
yottabyte	YB	10 ²⁴	yobibyte	YiB	2 ⁸⁰	21%



Random Access Memory (RAM)

- The computer's main memory is also called the Random Access Memory, because irrespective of the RAM size, every word gets selected instantaneously, at more or less the same time
- It is also known as read/write memory as it allows CPU to read as well as write data and instructions into it
- RAM is a microchip implemented using semiconductors. There are two categories of RAM
 - Dynamic RAM (DRAM): It is made up of memory cells where each cell is composed of one capacitor and one transistor. DRAM must be refreshed continually to store information. The refresh operation occurs automatically thousands of times per second. DRAM is slower and less-expensive
 - Static RAM (SRAM): It retains the data as long as power is provided to the memory chip. It needs not be refreshed periodically. SRAM uses multiple transistors for each memory cell. It does not use capacitor. SRAM is often used as cache memory due to its high speed. SRAM is more expensive than DRAM



Multi-Byte Ordering

- All 32 bit machines load and store 32 bits of data (word) with each operation. The question is how are the bytes of a multi-byte variable ordered in memory?
- Consider a 32 bit variable having a value of 0x01234567, that needs to be stored at address 0x100
- There are two conventions that the h/w designers can follow:
 - Big Endian: Most significant byte is written at the lowest address byte (MSB first). Used by MIPS and Internet

0x100	0x101	0x102	0x103	
01	23	45	67	

Little Endian: Least significant byte is written at the lowest address byte (LSB first). Used by x86 and ARM

0x100	0x101	0x102	0x103	
67	45	23	01	



Designing Random Access Memory



Design of RAM

- RAM is an array of *n w*-bit registers, equipped with direct access circuitry. The number of registers (n) and the width of each register (w) are called the memory's size and width respectively
- In simple words, you can think of RAM as a sequence of n addressable registers with addresses 0 to n-1
- At any given time only one register in the RAM is selected. It is this register whose value is available on *out* during a read operation. Similarly, it is this register whose contents will be over written during a write operation
- Now to select a register we need its address. Address width varies with the number of registers/words in the RAM, e.g., for RAM8 the address size is 3 bits

 $\mathbf{k} = \text{Address bits} = \log_2 \mathbf{n}$





Read/Write Logic of RAM

- At any given point of time: one register in the RAM is selected, all the other registers are irrelevant
- Read: To read the contents of register number *i*,
 - Set address = i
 - Result: The RAM's output pin *out* emits the state ad of the register *i*. This is a combinational operation, independent of the clock
- Write: To write a new data value *d* into register number *i*,
 - Set address = i
 - Set in = d
 - Set *load* = 1
 - Result: The state of register *i* becomes *d* and form next clock cycle onwards, *out* emits *d*



Chip Name	Size (n) A	ddress bits (k)
RAM8	8	3
RAM64	64	6
RAM512	512	9
RAM4K	4096	12
RAM16K	16384	14



8-Register/words RAM



8-Register/words RAM

load

Implementation tips:

- Memory of **8** registers, each 16 bit-wide. *Out* holds the value stored at the memory location specified by address. If load==1, then the in value is loaded into the memory location specified by address (the loaded value will be emitted to out from the next time step onward)
- Feed the 16 bit *in* value to all the registers, simultaneously
- Use DMux8Way chip to select one of the eight registers specified by address
- Use Mux8Way16 chip to select the 16 bit contents of register specified by address on 16 bit *out* output

Mux8Way16(a=out0, b=out1, c=out2, d=out3, e=out4, f=out5, g=out6, h=out7, sel=address, out=out);

RAM8 Demo

Designing Larger Size RAM Chips

RAM512

Same technique can be used to implement RAM4K and RAM16K

64-Register/words RAM

Implementation tips:

- Memory of **64** registers, each 16 bit-wide. *Out* holds the value stored at the memory location specified by address. If load==1, then the in value is loaded into the memory location specified by address (the loaded value will be emitted to out from the next time step onward)
- Feed the 16 bit *in* value to all the registers, simultaneously
- Use DMux8Way chip to select one of the eight registers specified by address
- Use Mux8Way16 chip to select the 16 bit contents of register specified by address on 16 bit *out* output

```
CHIP RAM64 {
```

```
IN in[16], load, address[6];
OUT out[16];
PARTS:
DMux8Way(in=load, sel=address[3..5], a=load0, b=load1, c=load2, d=load3, e=load4, f=load5, g=load6, h=load7);
RAM8(in=in, load=load0, address=address[0..2], out=out0);
RAM8(in=in, load=load1, address=address[0..2], out=out1);
RAM8(in=in, load=load2, address=address[0..2], out=out1);
RAM8(in=in, load=load3, address=address[0..2], out=out2);
RAM8(in=in, load=load4, address=address[0..2], out=out3);
RAM8(in=in, load=load5, address=address[0..2], out=out4);
RAM8(in=in, load=load6, address=address[0..2], out=out6);
RAM8(in=in, load=load7, address=address[0..2], out=out7);
```

Mux8Way16(a=out0, b=out1, c=out2, d=out3, e=out4, f=out5, g=out6, h=out7, sel=address[3..5], out=out);

512-Register/words RAM

Implementation tips:

- Memory of **512** registers, each 16 bit-wide. *Out* holds the value stored at the memory location specified by address. If load==1, then the in value is loaded into the memory location specified by address (the loaded value will be emitted to out from the next time step onward)
- Feed the 16 bit *in* value to all the registers, simultaneously
 - Use DMux8Way chip to select one of the eight registers specified by address
 - Use Mux8Way16 chip to select the 16 bit contents of register specified by address on 16 bit *out* output

```
CHIP RAM512 {
```

```
IN in[16], load, address[9];
OUT out[16];
PARTS:
DMux8Way(in=load, sel=address[6..8], a=load0, b=load1, c=load2, d=load3, e=load4, f=load5, g=load6, h=load7);
RAM64(in=in, load=load0, address=address[0..5], out=out0);
RAM64(in=in, load=load1, address=address[0..5], out=out1);
RAM64(in=in, load=load2, address=address[0..5], out=out2);
RAM64(in=in, load=load3, address=address[0..5], out=out3);
RAM64(in=in, load=load4, address=address[0..5], out=out4);
RAM64(in=in, load=load5, address=address[0..5], out=out5);
RAM64(in=in, load=load6, address=address[0..5], out=out6);
```

```
RAM64(in=in, load=load7, address=address[0..5], out=out7);
```

Mux8Way16(a=out0, b=out1, c=out2, d=out3, e=out4, f=out5, g=out6, h=out7, sel=address[6..8], out=out);

4K-Register/words RAM

Implementation tips:

- Memory of **4K** registers, each 16 bit-wide. *Out* holds the value stored at the memory location specified by address. If load==1, then the in value is loaded into the memory location specified by address (the loaded value will be emitted to out from the next time step onward)
- Feed the 16 bit *in* value to all the registers, simultaneously
 - Use DMux8Way chip to select one of the eight registers specified by address
 - Use Mux8Way16 chip to select the 16 bit contents of register specified by address on 16 bit *out* output

```
CHIP RAM4K {
    IN in[16], load, address[12];
    OUT out[16];
    PARTS:
    DMux8Way(in=load, sel=address[9..11], a=load0, b=load1, c=load2, d=load3, e=load4, f=load5, g=load6, h=load7);
    RAM512(in=in, load=load0, address=address[0..8], out=out0);
    RAM512(in=in, load=load1, address=address[0..8], out=out1);
    RAM512(in=in, load=load2, address=address[0..8], out=out2);
    RAM512(in=in, load=load3, address=address[0..8], out=out3);
    RAM512(in=in, load=load4, address=address[0..8], out=out4);
    RAM512(in=in, load=load5, address=address[0..8], out=out5);
    RAM512(in=in, load=load6, address=address[0..8], out=out6);
    RAM512(in=in, load=load7, address=address[0..8], out=out7);
```

Mux8Way16(a=out0, b=out1, c=out2, d=out3, e=out4, f=out5, g=out6, h=out7, sel=address[9..11], out=out);

16K-Register/words RAM

Implementation tips:

- Memory of **16K** registers, each 16 bit-wide. *Out* holds the value stored at the memory location specified by address. If load==1, then the in value is loaded into the memory location specified by address (the loaded value will be emitted to out from the next time step onward)
- Feed the 16 bit *in* value to all the registers, simultaneously
- Use DMux8Way chip to select one of the eight registers specified by address
- Use Mux8Way16 chip to select the 16 bit contents of register specified by address on 16 bit *out* output

```
CHIP RAM16K {
    IN in[16], load, address[14];
    OUT out[16];
    PARTS:
    DMux4Way(in=load, sel=address[12..13], a=load0, b=load1, c=load2, d=load3);
    RAM4K(in=in, load=load0, address=address[0..11], out=out0);
    RAM4K(in=in, load=load1, address=address[0..11], out=out1);
    RAM4K(in=in, load=load2, address=address[0..11], out=out2);
    RAM4K(in=in, load=load3, address=address[0..11], out=out3);
    Mux4Way16(a=out0, b=out1, c=out2, d=out3, sel=address[12..13], out=out);
```


Counters

Overview of Counters

- A counter is a special type of register that goes through a pre-determined sequence of states upon the application of input pulses
- Counters are used for:
 - Counting the number of occurrences of an event
 - Keeping time or calculating amount of time between events
 - Baud rate generation
- A w-bit counter consists of two main elements:
 - A w-bit register to store a w-bit value
 - A combinational logic to
 - Compute the next value (according to a specific counting function)
 - Load a new value of user/programmer choice
 - Reset the counter to a default value
- Examples:
 - Simple Up/Down Binary Counters
 - BCD Counter(s)
 - Gray Code Counter
 - Ring Counter
 - Johnson Counter

Why we Need Counter Chip for Hack CPU

Computer System

Why we Need Counter Chip for Hack CPU

- Consider a counter chip designed to contain the address of the instruction that the computer should fetch and execute next
- In most cases, the counter has to simply increment itself by 1 in each clock cycle, thus causing the computer to fetch the next instruction in the program
- In other cases, we may want the program to *jump to an instruction at memory address n*, so the programmer want to set the counter to a value of *n*, rather than its default counting behavior with *n*+1, *n*+2, and so forth
- Finally, the program's execution can be restarted anytime by resetting the counter to 0, assuming that the address of the program's first instruction
- In short, we need a loadable and resettable counter

- Every computer has a special register called the Program Counter, normally called the PC, which keeps track of the instruction to be fetched and executed next
- The PC is designed to support three possible control operations:
 - **Reset:** Fetch the first instruction

Program Counter Register

- Next: Fetch the next instruction
- Goto: Fetch instruction at address **n**

n the nevt	instruction

PC++

50

Counter Abstraction

Counter Simulation

Instructor: Muhammad Arif Butt, Ph.D.

16 Bit Program Counter Implementation

16 Bit Program Counter Implementation

Program Counter Demo

Things To Do

• Perform testing of the chips designed in today's session on the h/w simulator. You can download the .hdl, .tst and .cmp files of above chips from the course bitbucket repository:

https://bitbucket.org/arifpucit/coal-repo/

• Interested students should also try to design, implement and simulate binary down counter, cascaded BCD counter, Gray Counter, Ring counter, and Johnson counter

Coming to office hours does NOT mean you are academically weak!