Non-volatile memory & Datapath component (3)

Prof. Usagi



Recap: Registers

- Register: a sequential component that can store multiple bits
- A basic register can be built simply by using multiple D-FFs



re multiple bits ultiple D-FFs

Recap: A Classical 6-T SRAM Cell bitline'









Recap: DRAM cell



- 1 transistor (rather than 6)
- Relies on large capacitor to store bit
 - Write: transistor conducts, data voltage level gets stored on top plate of capacitor
 - Read: look at the value of d
- Problem: Capacitor discharges over time
 - Must "refresh" regularly, by reading d and then writing it right back

her than 6) Capacitor to store



Recap: Latency of volatile memory

	Size (Transistors per bit)	Latency
Register	18T	~ 0.1 r
SRAM	6T	~ 0.5 r
DRAM	1T	50-100



(ns)

ns

ns

) ns

Recap: Thinking about programming

```
struct student_record
    int id;
    double homework;
    double midterm;
    double final;
};
int main(int argc, char **argv)
{
    int i,j;
    double midterm average=0.0;
    int number of records = 10000000;
    struct timeval time_start, time_end;
    struct student record *records;
    records = (struct
student record*)malloc(sizeof(struct
student_record)*number_of_records);
    init(number of records, records);
    for (j = 0; j < 100; j++)
        for (i = 0; i < number_of_records; i++)</pre>
            midterm_average+=records[i].midterm;
    printf("average: %lf\n",midterm_average/
number_of_records);
   free(records);
    return 0;
}
```

```
int main(int argc, char **argv)
    int i,j;
    double midterm_average=0.0;
    int number_of_records = 10000000;
    struct timeval time_start, time_end;
    id = (int*)malloc(sizeof(int)*number_of_records);
    init(number_of_records);
    for (j = 0; j < 100; j++)
        for (i = 0; i < number_of_records; i++)</pre>
            midterm_average+=midterm[i];
```

```
free(id);
free(midterm);
free(final);
free(homework);
return 0;
```

}

```
A. Left
B. Right
<sup>9</sup>C. About the same
```

midterm = (double*)malloc(sizeof(double)*number_of_records); final = (double*)malloc(sizeof(double)*number_of_records); homework = (double*)malloc(sizeof(double)*number of records);

More row buffer hits in the **DRAM, more SRAM hits**

Which side is faster in executing the for-loop?



Recap: Flash memory

- Floating gate made by polycrystalline silicon trap electrons
- The voltage level within the floating gate determines the value of the cell
- The floating gates will wear out eventually









- Non-volatile memory case study: flash memory
- Sequential Datapath Components



Programming in MLC









2nd Page Programming in MLC





Flash memory characteristics

- Regarding the following flash memory characteristics, please identify how many of the following statements are correct
 - ① Flash memory cells can only be programmed with limited times
 - The reading latency of flash memory cells can be largely different from (2) programming
 - The latency of programming different flash memory pages can be different (3)
 - The programmed cell cannot be reprogrammed again unless its charge level is (4)refilled to the top-level
 - A. 0
 - B. 1
 - C. 2
 - D. 3
 - E. 4





Fewer writes per cell





Similar relative performance for reads, writes and erases

Laura M. Grupp, Adrian M. Caulfield, Joel Coburn, Steven Swanson, Eitan Yaakobi, Paul H. Siegel, and Jack K. Wolf. Characterizing flash memory: anomalies, observations, and applications. In MICRO 2009.

Flash memory characteristics

- Regarding the following flash memory characteristics, please identify how many of the following statements are correct
 - ① Flash memory cells can only be programmed with limited times
 - The reading latency of flash memory cells can be largely different from (2) programming

 - ③ The latency of programming different flash memory pages can be different The programmed cell cannot be reprogrammed again unless its charge level is (4)refilled to the top-level
 - A. 0
 - B. 1
 - C. 2





Phase change memory

- The bit is stored in the crystal structure of a tiny spec of metal.
- To write, it melts the metal (650C)
 - let it cool quickly or slowly to set the value
 - Crystaline and amorphous states have different resistance







Spin-torque transfer

- Bits stored as magnetic orientation of a thin film
- Change the state using polarized electrons (!)
- Depending on polarization, resistance differs
- More complex cell structure
- Great promise potential DRAM replacement
 - Roughly the same speed, power, and bandwidth.
 - But it's durable!



Non-volatile memory technologies

	H.D.D	Flash	Optane	
Latency	~ 10-15 ms	~ 100 us (read) ~ 1 ms (write)	7 us (read) 18 us (write)	
Bandwidth	~200 MB/Sec	3.5 GB/sec (read) 2.1 GB/sec (write)	1.35 GB/sec (read) 290 MB/sec (write)	
Dollar/GB	0.0295	0.583	2.18	

Flash is still the most convincing technology for now





STT-MRAM

35 ns



If programmer doesn't know flash "features"

 Software designer should be aware of the characteristics of underlying hardware components

Spotify is writing massive amounts of junk data to storage drives

Streaming app used by 40 million writes hundreds of gigabytes per day.

DAN GOODIN - 11/10/2016, 7:00 PM



Spotify has been quietly killing your SSD's life for months





CLA v.s. Carry-ripple

- Size:
 - 32-bit CLA with 4-bit CLAs requires 8 of 4-bit CLA
 - Each requires 116 for the CLA $4^{*}(4^{*}6+8)$ for the A+B 244 gates
 - 1952 transistors
 - 32-bit CRA
 - 1600 transistors
- Delay
 - 32-bit CLA with 8 4-bit CLAs
 - 2 gates * 8 = 16 **Win**
 - 32-bit CRA
 - 64 gates

Area-Delay Trade-off!



Serial Adder



Feed a_i and b_i and generate s_i at time i. Where is c_i and c_{i+1} ?



The basic idea



Excitation Table of Serial Adder

a _i	bi	Ci	Ci+1	Si
0	0	0	0	0
0	0	1	0	1
0	1	0	0	1
0	1	1	1	0
1	0	0	0	1
1	0	1	1	0
1	1	0	1	0
1	1	1	1	1



Excitation Table of Serial Adder

a _i	bi	Ci	Ci+1	Si
0	0	0	0	0
0	0	1	0	1
0	1	0	0	1
0	1	1	1	0
1	0	0	0	1
1	0	1	1	0
1	1	0	1	0
1	1	1	1	1





Area/Delay of adders

- Consider the following adders?
 - ① 32-bit CLA made with 8 4-bit CLA adders
 - ② 32-bit CRA made with 32 full adders
 - ③ 32-bit serial adders made with 4-bit CLA adders
 - ④ 32-bit serial adders made with 1-bit full adders
 - A. Area: (1) > (2) > (3) > (4) Delay: (1) < (2) < (3) < (4)
 - B. Area: (1) > (3) > (2) > (4) Delay: (1) < (3) < (2) < (4)
 - C. Area: (1) > (3) > (4) > (2) Delay: (1) < (3) < (4) < (2)
 - D. Area: (1) > (2) > (3) > (4) Delay: (1) < (3) < (2) < (4)
 - E. Area: (1) > (3) > (2) > (4) Delay: (1) < (3) < (4) < (2)



Area/Delay of adders

- Consider the following adders?
 - ① 32-bit CLA made with 8 4-bit CLA adders
 - ② 32-bit CRA made with 32 full adders _____
 - ③ 32-bit serial adders made with 4-bit CLA adders
 Each CLA (3-gate delay + 2-gate delay)*8 cycles 5*8+1 = 41
 ④ 32-bit serial adders made with 1-bit full adders

 - Each CLA (2-gate delay + 2-gate delay)*32 cycles 4*32 = 128A. Area: (1) > (2) > (3) > (4) Delay: (1) < (2) < (3) < (4)
 - B. Area: (1) > (3) > (2) > (4) Delay: (1) < (3) < (2) < (4)
 - C. Area: (1) > (3) > (4) > (2) Delay: (1) < (3) < (4) < (2)
 - D. Area: (1) > (2) > (3) > (4) Delay: (1) < (3) < (2) < (4)
 - E. Area: (1) > (3) > (2) > (4) Delay: (1) < (3) < (4) < (2)



Each carry — 2-gate delay — 64

Frequency

- Consider the following adders. Assume each gate delay is 1ns and the delay in a register is 2ns. Please rank their maximum operating frequencies
 - ① 32-bit CLA made with 8 4-bit CLA adders
 - ② 32-bit CRA made with 32 full adders
 - ③ 32-bit serial adders made with 4-bit CLA adders
 - ④ 32-bit serial adders made with 1-bit full adders
 - A. (1) > (2) > (3) > (4)
 - B. (2) > (1) > (4) > (3)
 - C. (2) > (1) > (3) > (4)
 - D. (4) > (3) > (2) > (1)
 - E. (4) > (3) > (1) > (2)

Frequency

- Consider the following adders. Assume each gate delay is 1ns and the delay in a register is 2ns. Please rank their maximum operating frequencies

 - 3 32-bit serial adders made with 4-bit CLA adders
 ¹/_{5ns} = 200MHz

 32-bit serial adders made with 1-bit full adders
 ¹/_{4ns} = 250MHz

 - A. (1) > (2) > (3) > (4)
 - B. (2) > (1) > (4) > (3)
 - C. (2) > (1) > (3) > (4)

E. (4) > (3) > (1) > (2)

Announcement

- Assignment #4 due tonight Chapter 4.8-4.9 & 5.2-5.4
- Lab 5 is up due this Thursday
 - Watch the video and read the instruction BEFORE your session
 - There are links on both course webpage and iLearn lab section
 - Submit through iLearn > Labs
- Office Hours
 - All office hours share the same meeting instance if you have registered once, you cannot do it again.
 - Zoom does not resend registration confirmation and does not allow us to "re-approve" if you have registered
 - The only way is to dig out the e-mail from Zoom
- Last reading quiz due next Tuesday
- Check your grades in iLearn

Electrical Computer Science Engineering





