

ESE534: Computer Organization

Day 7: February 6, 2012
Memories



Previously...

- Arithmetic: addition, subtraction
- Reuse:
 - pipelining
 - bit-serial (vectorization)
- Area/Time Tradeoffs
- Latency and Throughput

Today

- Memory
 - features
 - Area and delay intuition and modeling
 - design
 - technology

Preclass 1

When is:

$$2000\sqrt{N} + N < 100N$$

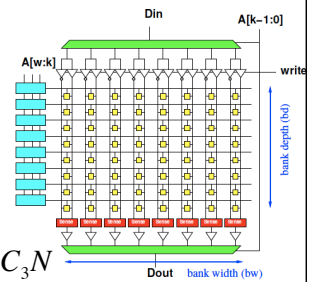
Preclass 2

Find m to minimize:

$$\frac{C_1 N \log_2(N)}{m} + C_2 m + C_3 N$$

Preclass 2

- How related?

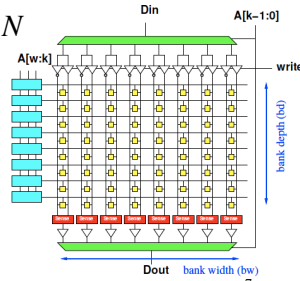


$$\frac{C_1 N \log_2(N)}{m} + C_2 m + C_3 N$$

Preclass 2

$$\frac{C_1 N \log_2(N)}{m} + C_2 m + C_3 N$$

- Related?
- $m = bw$
- $bw * bd = N$
– So $bd = N/m$

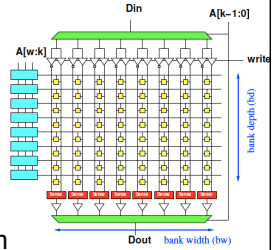


Penn ESE534 Spring2010 -- DeHon

Preclass 2

$$C_1 b d \log_2(N) + C_2 b w + C_3 N$$

- C_3 is for area of single memory cell
- $\log_2(n)$ is for decoder
– In red
- C_1 is for gates in decoder
- C_2 is for amps at bottom of row and drivers at top



Penn ESE534 Spring2010 -- DeHon

8

Preclass 1

$$\frac{C_1 N \log_2(N)}{m} + C_2 m + C_3 N$$

When we solved for N , we found m is proportional to \sqrt{N}

$$C_1 \sqrt{N} \log_2(N) + C_2 \sqrt{N} + C_3 N$$

If we approximate $\log_2(N)$ as a constant, we get

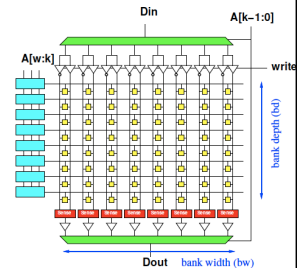
$$C_4 \sqrt{N} + C_3 N$$

$$2000 \sqrt{N} + N$$

Penn ESE534 Spring2010 -- DeHon

Preclass 3, 4

- Delay
 - Scale with bd ?
 - Scale with bw ?



Penn ESE534 Spring2012 -- DeHon

10

Memory

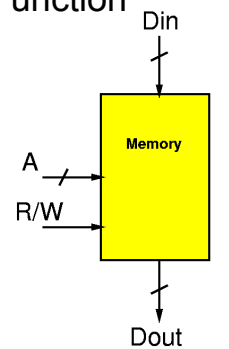
- What's a memory?
- What's special about a memory?

Penn ESE534 Spring2012 -- DeHon

11

Memory Function

- Typical:
 - Data Input Bus
 - Data Output Bus
 - Address
 - (location or name)
 - read/write control



Penn ESE534 Spring2012 -- DeHon

Memory

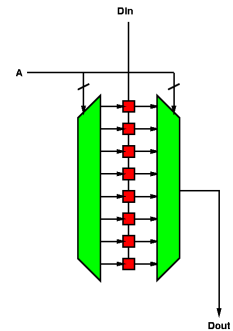
- Block for storing data for later retrieval
- State element

Penn ESE534 Spring2012 -- DeHon

13

Collection

- What's different between a memory and a collection of registers?



Penn ESE534 Spring2012 -- DeHon

Memory Uniqueness

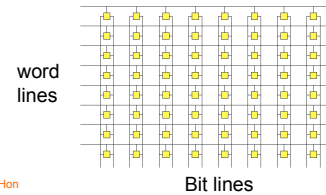
- **Cost**
- Compact state element
- Packs data very tightly
- At the expense of sequentializing access
- Example of Area-Time tradeoff
 - and a key enabler

Penn ESE534 Spring2012 -- DeHon

15

Memory Organization

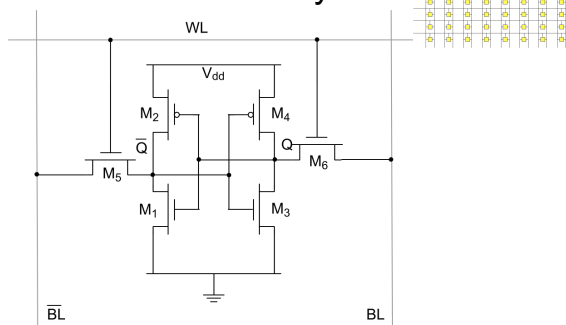
- **Key idea:** sharing
 - factor out common components among state elements
 - can have big elements if amortize costs
 - state element unique → small



Penn ESE534 Spring2012 -- DeHon

16

SRAM Memory bit



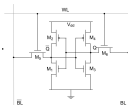
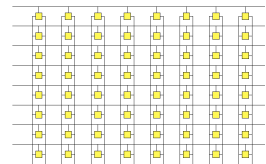
Source: <http://commons.wikimedia.org/wiki/File:6t-SRAM-cell.png>

Penn ESE534 Spring2012 -- DeHon

17

Memory Organization

- **Share:** Interconnect
 - Input bus
 - Output bus
 - Control routing
- **very** topology/wire cost aware design
- **Note:** local, abutment wiring



Penn ESE534 Spring2012 -- DeHon

18

Share Interconnect

- Input Sharing
 - wiring
 - drivers
- Output Sharing
 - wiring
 - sensing
 - driving

19

Penn ESE534 Spring2012 -- DeHon

Address/Control

- Addressing and Control
 - an overhead
 - paid to allow this sharing

20

Penn ESE534 Spring2012 -- DeHon

Memory Organization

21

Penn ESE534 Spring2012 -- DeHon

Dynamic RAM

- Goes a step further
- Share refresh/restoration logic as well
- Minimal storage is a capacitor
- “Feature” DRAM process is ability to make capacitors efficiently

22

Penn ESE534 Spring20

Some Numbers (memory)

- Unit of area = λ^2 (F=2 λ)
 - [more next week]
- Register as stand-alone element $\approx 4K\lambda^2$
 - e.g. as needed/used last lecture
- Static RAM cell $\approx 1K\lambda^2$
 - SRAM Memory (single ported)
- Dynamic RAM cell (DRAM process) $\approx 100\lambda^2$
- Dynamic RAM cell (SRAM process) $\approx 300\lambda^2$

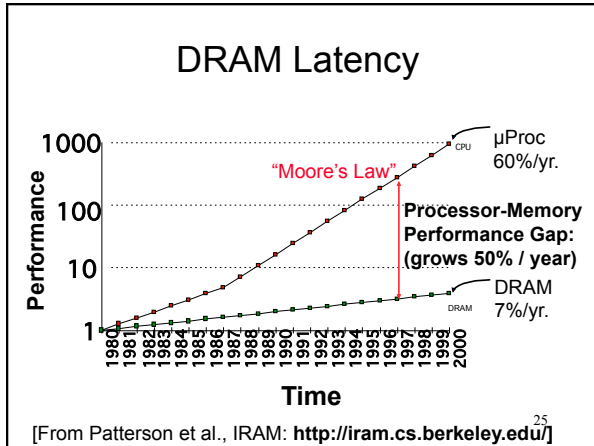
23

Penn ESE534 Spring2012 -- DeHon

DRAM Cell Layout

24

Penn ESE534 Spring2012 -- DeHon
Source: <http://techn.nikkeibp.co.jp/article/HONSHU/20071219/144399/>



Contemporary DRAM

- 1GB DDR3 SDRAM from Micron
 - <http://www.micron.com/products/dram/ddr3/>
 - 96 pin package
 - 16b datapath IO
 - Operate at 500+MHz
 - 37.5ns random access latency

Speed Grade	Data Rate (Mbits)	Target RCD/RP/CL	tRCD (ns)	tRP (ns)	CL (ns)
-25E	800	5-5-5	12.5	12.5	12.5
-25	800	6-6-6	15	15	15
-187E	1,066	7-7-7	13.1	13.1	13.1
-187	1,066	8-8-8	15	15	15
-15E	1,333	9-9-9	13.5	13.5	13.5
-15	1,333	10-10-10	15	15	15

Options

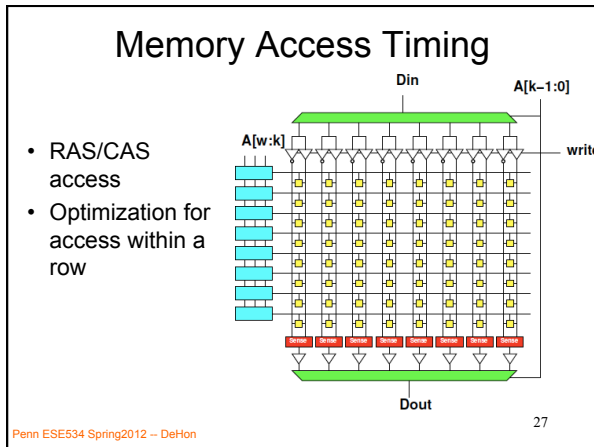
- Configuration
 - 64 Meg x 16 (8 Meg x 16 x 8 banks)
 - 128 Meg x 8 (16 Meg x 8 x 8 banks)
 - 256 Meg x 4 (32 Meg x 4 x 8 banks)
 - FBGA package (lead-free)
 - x4, x8, 96-ball FBGA (9mm x 15.5mm)
 - x16, 96-ball FBGA (9mm x 15.5mm)
- Timing - cycle time
 - 2.5ns @ CL = 6 (DDR3-800)
 - 2.5ns @ CL = 5 (DDR3-800)
 - 1.87ns @ CL = 8 (DDR3-1066)
 - 1.87ns @ CL = 7 (DDR3-1066)
 - 1.5ns @ CL = 10 (DDR3-1333)
 - 1.5ns @ CL = 9 (DDR3-1333)

Marking

- 64M16
- 128M8
- 256M4
- BY LA

Penn ESE534 Spring2012 -- DeHon

26



Contemporary DRAM

- 1GB DDR3 SDRAM from Micron
 - <http://www.micron.com/products/dram/ddr3/>
 - 96 pin package
 - 16b datapath IO
 - Operate at 500+MHz
 - 37.5ns random access latency

Speed Grade	Data Rate (Mbits)	Target RCD/RP/CL	tRCD (ns)	tRP (ns)	CL (ns)
-25E	800	5-5-5	12.5	12.5	12.5
-25	800	6-6-6	15	15	15
-187E	1,066	7-7-7	13.1	13.1	13.1
-187	1,066	8-8-8	15	15	15
-15E	1,333	9-9-9	13.5	13.5	13.5
-15	1,333	10-10-10	15	15	15

Options

- Configuration
 - 64 Meg x 16 (8 Meg x 16 x 8 banks)
 - 128 Meg x 8 (16 Meg x 8 x 8 banks)
 - 256 Meg x 4 (32 Meg x 4 x 8 banks)
 - FBGA package (lead-free)
 - x4, x8, 96-ball FBGA (9mm x 15.5mm)
 - x16, 96-ball FBGA (9mm x 15.5mm)
- Timing - cycle time
 - 2.5ns @ CL = 6 (DDR3-800)
 - 2.5ns @ CL = 5 (DDR3-800)
 - 1.87ns @ CL = 8 (DDR3-1066)
 - 1.87ns @ CL = 7 (DDR3-1066)
 - 1.5ns @ CL = 10 (DDR3-1333)
 - 1.5ns @ CL = 9 (DDR3-1333)

Marking

- 64M16
- 128M8
- 256M4
- BY LA

Penn ESE534 Spring2012 -- DeHon

28

1 Gigabit DDR2 SDRAM

[Source: <http://www.elpida.com/en/news/2004/11-18.html>]

Penn ESE534 Spring2012 -- DeHon

29

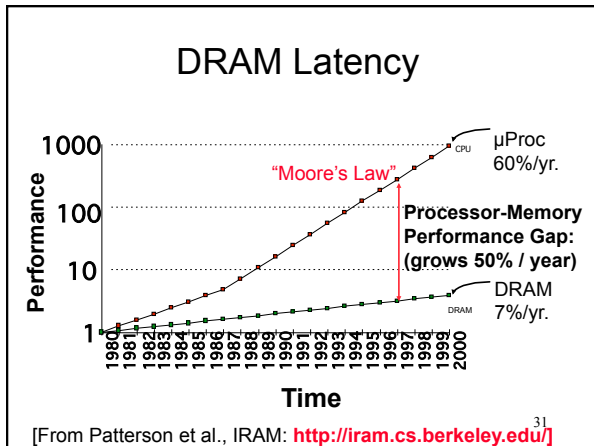
Contemporary DRAM

- 4GB DDR3 SDRAM from Micron
 - http://download.micron.com/pdf/datasheets/modules/ddr3/ksf16c256_512x64hz.pdf
 - Operate at 800MHz
 - 50ns random access latency

Speed Grade	Industry Nomenclature	Data Rate (MT/s)					tRCD (ns)	tRP (ns)	tRC (ns)		
		CL = 11	CL = 10	CL = 9	CL = 8	CL = 7					
-1G6	PC3-12800	1600	1333	1333	1066	1066	800	667	13,125	13,125	48,125
-1G4	PC3-10600	-	1333	1333	1066	1066	800	667	13,125	13,125	49,125
-1G1	PC3-8500	-	-	-	1066	1066	800	667	13,125	13,125	50,625
-1G0	PC3-8500	-	-	-	1066	-	800	667	15	15	52.5
-80C	PC3-6400	-	-	-	-	-	800	800	12.5	12.5	50
-80B	PC3-6400	-	-	-	-	-	800	667	15	15	52.5

Penn ESE534 Spring2012 -- DeHon

30



Basic Memory Design Space

- Width
- Depth
- Internal vs. External Width
- **Banking**
 - To come

External Width

Depth

Internal Width

External Width

Penn ESE534 Spring2012 -- DeHon

Depth

- What happens as make deeper?
 - Delay?
 - Energy?

External Width

Depth

Internal Width

External Width

Penn ESE534 Spring2012 -- D

Banking

- Tile Banks/memory blocks
- What does banking do for us? (E, D, A)?

A0-1 D0

A1-1 D1

A2-1 D2

A3-1 D3

Penn ESE534 Spring2012 -- DeHon

Independent Bank Access

Address/Din Xbar

Output Xbar

- Utility? Costs and benefits?

Penn ESE534 Spring2012 -- DeHon

Memory

- **Key Idea**
 - Memories hold state compactly
 - Do so by minimizing key state storage and amortizing rest of structure across large array

Penn ESE534 Spring2012 -- DeHon

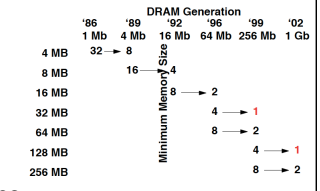
Yesterday vs. Today (Memory Technology)

- What's changed?

Yesterday vs. Today (Memory Technology)

- What's changed?

- Capacity
 - single chip
- Integration
 - memory and logic
 - dram and logic
 - embedded memories [Patterson et al., IEEE Micro April 1997]
- Room on chip for big memories
 - And many memories...
- Don't have to make a chip crossing to get to memory



Important Technology Cost

- IO between chips \ll IO on chip
 - pad spacing
 - area vs. perimeter (4s vs. s^2)
 - wiring technology
- **BIG** factor in multi-chip system designs
- Memories nice
 - very efficient with IO cost vs. internal area

On-Chip vs. Off-Chip BW

- Use Micron 1Gb DRAM as example

Table 1: Key Timing Parameters

Speed Grade	Data Rate (Mbits)	Target RCD/RP/CL	tRCD (ns)	tRP (ns)	CL (ns)
-25E	800	5-5-5	12.5	12.5	12.5
-25	800	6-6-6	15	15	15
-187E	1,066	7-7-7	13.1	13.1	13.1
-187	1,066	8-8-8	15	15	15
-15E	1,333	9-9-9	13.5	13.5	13.5
-15	1,333	10-10-10	15	15	15

Options

- Configuration
 - 64 Meg x 16 (8 Meg x 16 x 8 banks) 64M16
 - 128 Meg x 8 (16 Meg x 8 x 8 banks) 128M8
 - 256 Meg x 4 (32 Meg x 4 x 8 banks) 256M4
- FBGA package (lead-free)
 - x4, x8, 86-ball FBGA (9mm x 15.5mm) BY
 - x16, 96-ball FBGA (9mm x 15.5mm) LA
- Timing-cycle time
 - 2.5ns @ CL = 6 (DDR3-800) -25
 - 2.5ns @ CL = 5 (DDR3-800) -25E
 - 1.87ns @ CL = 8 (DDR3-1066) -187
 - 1.87ns @ CL = 7 (DDR3-1066) -187E
 - 1.5ns @ CL = 10 (DDR3-1333) -15
 - 1.5ns @ CL = 9 (DDR3-1333) -15E

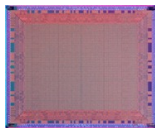
On Chip BW:

assume 8 1024b banks?

assume independent 1024x1024b banks?

Memory on a Virtex-6

- 40nm process node
 - Last generation, 28nm sampling now
- 300,000 programmable 6-input gates (6-LUTs)
- 1,064 memory banks \sim 512x72
 - Total \sim 38Mbits
- Operating at 600MHz
- On chip bandwidth?



[image from <http://www.electronicweekly.com/Articles/08/10/2009/47137/top-10-programmable-devices.html>]

Costs Change

- Design space changes when whole system goes on single chip
- Can afford
 - wider busses
 - more banks
 - memory tailored to application/architecture
- Beware of old (stale) answers
 - their cost model was different

What is Importance of Memory?

- **Radical Hypothesis:**
 - Memory is simply a very efficient organization which allows us to store data compactly
 - (at least, in the technologies we've seen to date)
 - A great engineering **trick** to optimize resources
- **Alternative:**
 - memory is a **primary**
- **State is a primary, but state≠memory**

Penn ESE534 Spring2012 -- DeHon

43

Admin

- HW1 and 2 graded – note feedback (even if received perfect score)
- HW4 out today
- Reading
 - Was for today
 - Nothing new for Wed., Mon.
- Office Hours: Tuesday 4:15pm

Penn ESE534 Spring2012 -- DeHon

44

Big Ideas [MSB Ideas]

- Memory: efficient way to hold state
- Resource sharing: key trick to reduce area
- Memories are a great example of resource sharing

Penn ESE534 Spring2012 -- DeHon

45

Big Ideas [MSB-1 Ideas]

- Tradeoffs in memory organization
- Changing cost of memory organization as we go to on-chip, embedded memories

Penn ESE534 Spring2012 -- DeHon

46