CS 152 Computer Architecture and Engineering

Lecture 7 - Memory Hierarchy-II

Krste Asanovic
Electrical Engineering and Computer Sciences
University of California at Berkeley

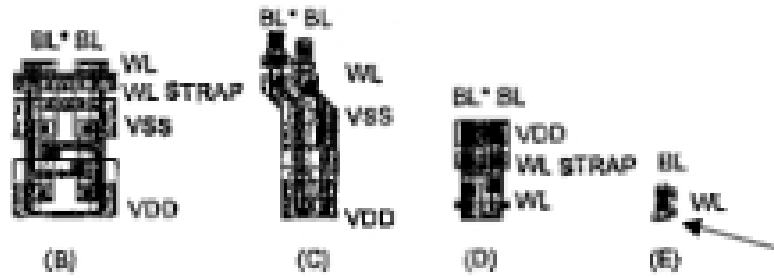
http://www.eecs.berkeley.edu/~krste
http://inst.eecs.berkeley.edu/~cs152

Last time in Lecture 6

- Dynamic RAM (DRAM) is main form of main memory storage in use today
  - Holds values on small capacitors, need refreshing (hence dynamic)
  - Slow multi-step access: precharge, read row, read column
- Static RAM (SRAM) is faster but more expensive
  - Used to build on-chip memory for caches
- Caches exploit two forms of predictability in memory reference streams
  - Temporal locality, same location likely to be accessed again soon
  - Spatial locality, neighboring location likely to be accessed soon
- Cache holds small set of values in fast memory (SRAM) close to processor
  - Need to develop search scheme to find values in cache, and replacement policy to make space for newly accessed locations

Relative Memory Cell Sizes

On-Chip SRAM in logic chip



DRAM on memory chip

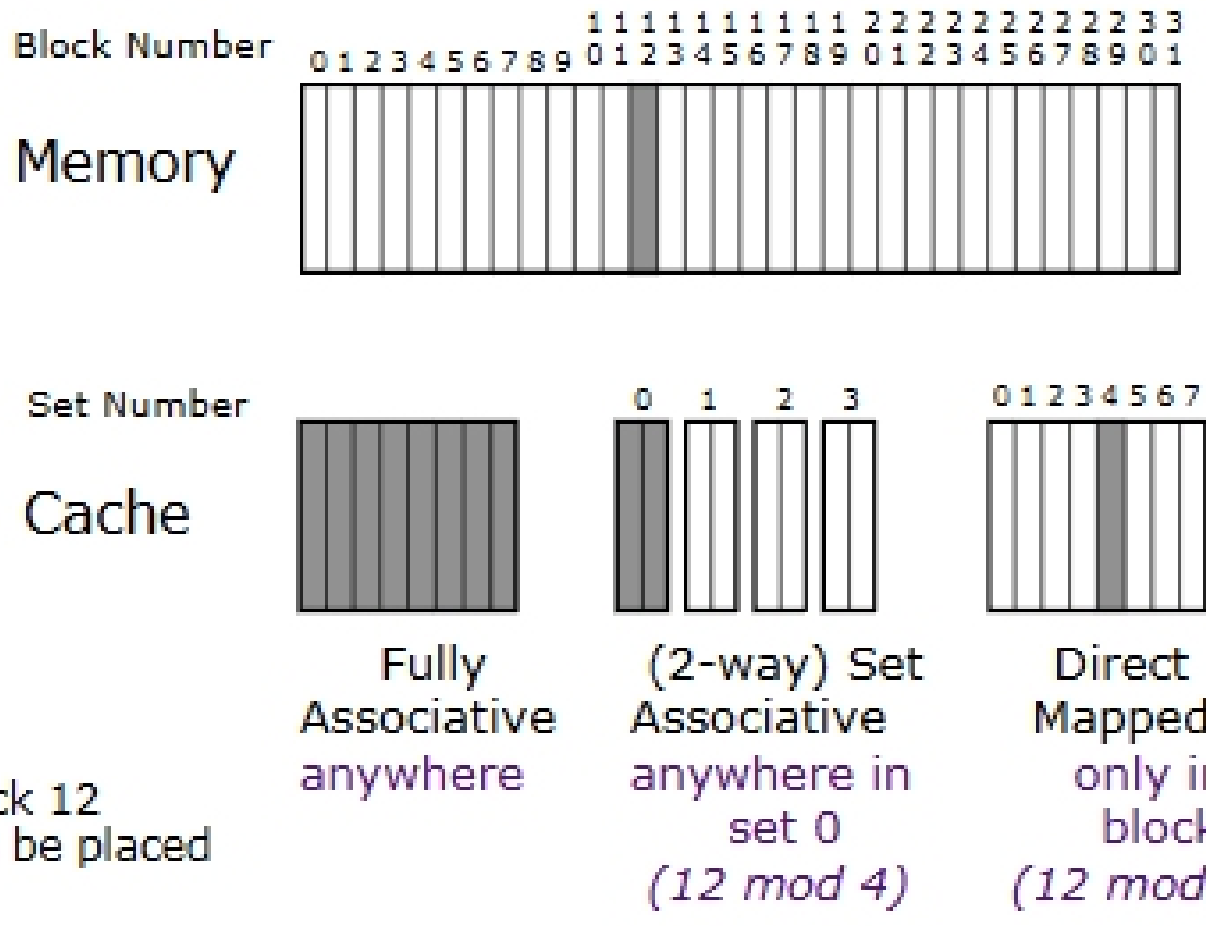
1 Memory cell in 0.5µm processes
a) Gate Array SRAM
b) Embedded SRAM
c) Standard SRAM (6T cell with local interconnect)
d) ASIC DRAM
e) Standard DRAM (stacked cell)

[Foss, "Implementing Application-Specific Memory", ISSCC 1996]

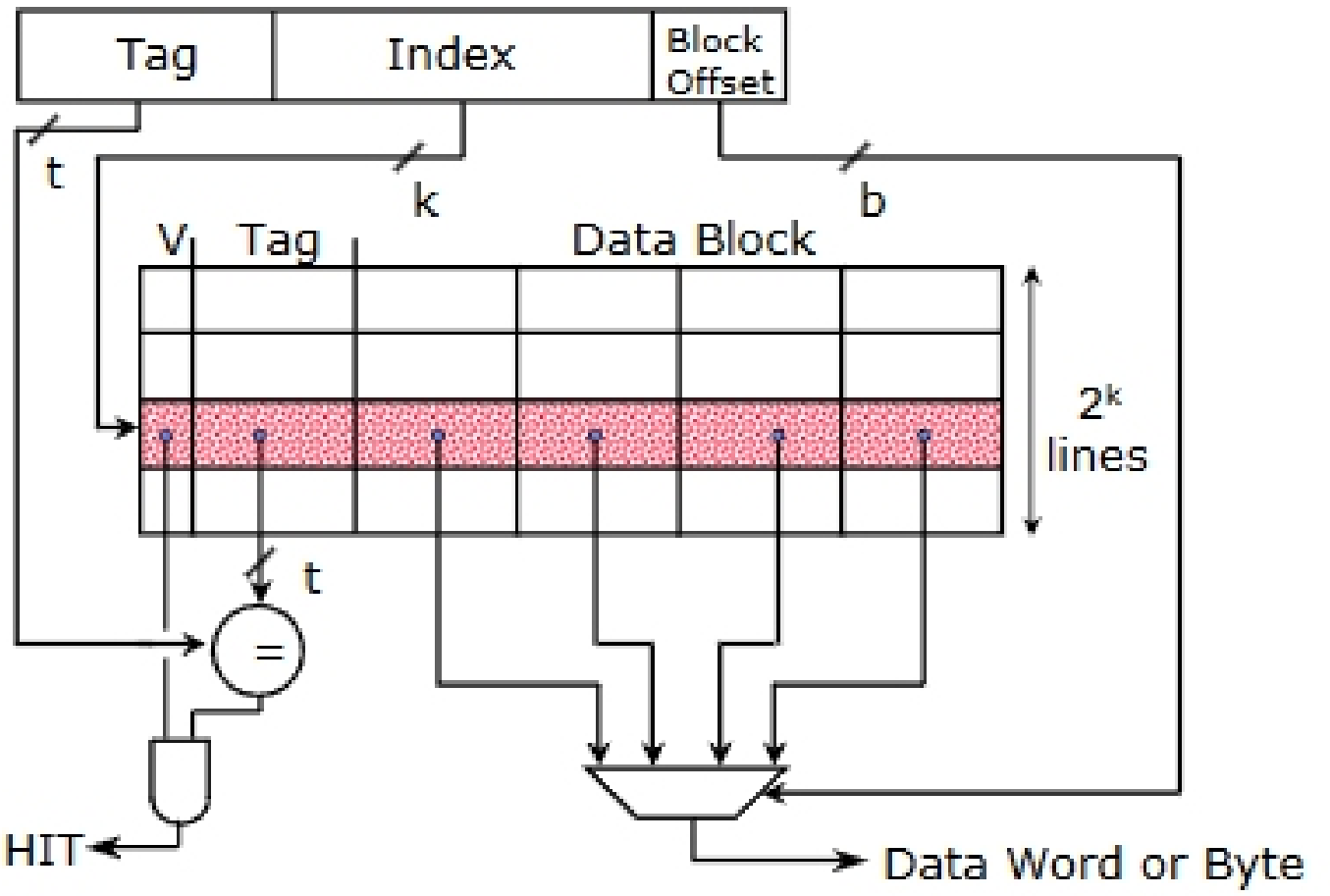
Memory	Process	Cell size (μm^2)	Cell efficiency	Bits in $100mm^2$ (10^6)	Gate size (μm^2)	Gate utilization	Gates in $100mm^2$ (10^6)
Gate array SRAM	3-metal ASIC	370	80%	216	185	70%	378
Embedded SRAM	3-metal ASIC	67	70%	1045	185	70%	378
Standard SRAM	2-metal 6T local int.	43	65%	1512	245	40%	163
Embedded ASIC-DRAM	3-metal ASIC	23	60%	2609	185	70%	378
Standard DRAM	2-metal stacked cell	3.2	50%	15625	413	40%	97

Table 1: Memory and logic density for a variety of 0.5µm implementations.

Placement Policy



Direct-Mapped Cache



2-Way Set-Associative Cache