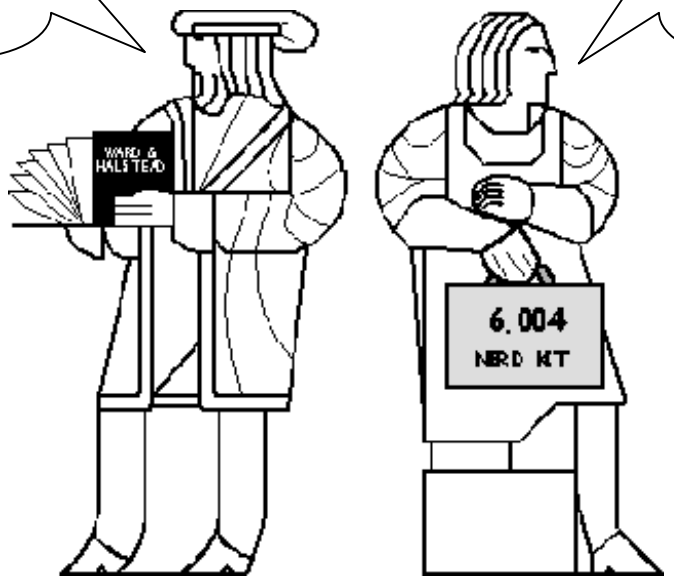
Computer Architecture: Exciting Times Ahead!

Prediction is very difficult, especially about the future.
Neils Bohr

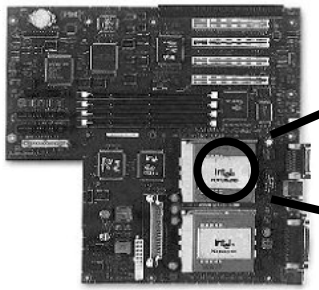
The best way to predict the future is to invent it.
Alan Kay



Handouts: Lecture Slides, MPR article

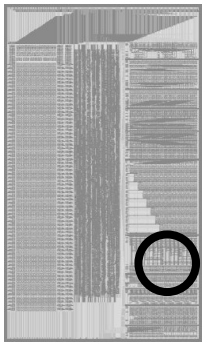
What Do You See?

Personal Computer:
Hardware & Software



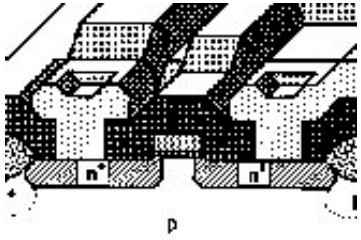
Circuit Board:
≈8 / system
1-2G devices

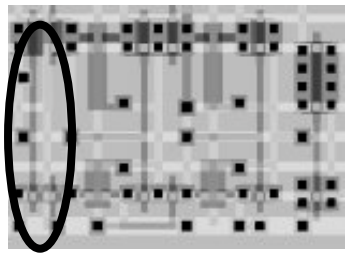
Integrated Circuit:
≈8-16 / PCB
.25M-16M devices



Module:
≈8-16 / IC
100K devices

MOSFET



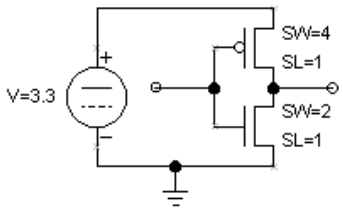


Cell:
≈1K-10K / Module
16-64 devices

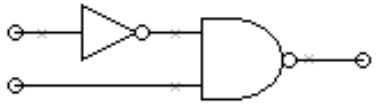
Scheme for
representing
information

Gate:
≈2-16 / Cell
8 devices

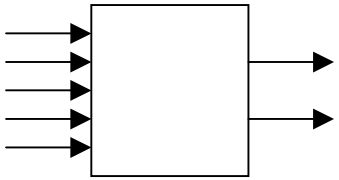
6.004 Roadmap



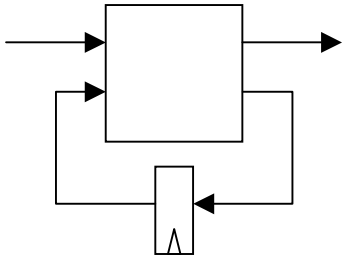
Fets & voltages



Logic gates



Combinational logic circuits



Sequential logic

Combinational contract:

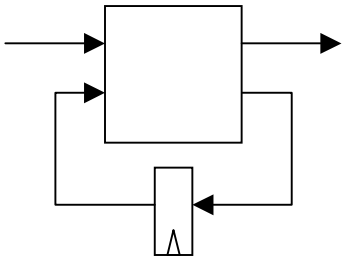
- discrete-valued inputs
- complete in/out spec.
- static discipline

Acyclic connections
Summary specification
Design:

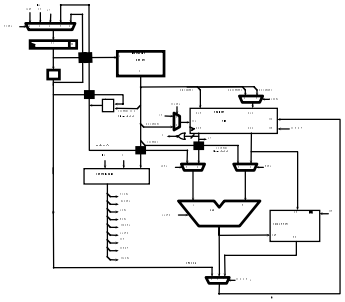
- sum-of-products
- simplification
- muxes, ROMs, PLAs

Storage & state
Dynamic discipline
Finite-state machines
Metastability
Throughput & latency
Pipelining

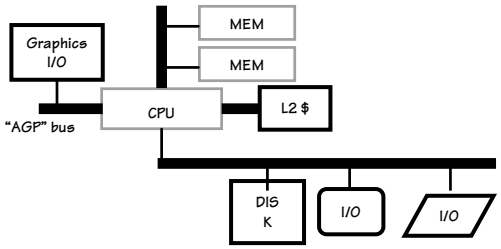
6.004 Roadmap (cont'd.)



Sequential logic



CPU Architecture



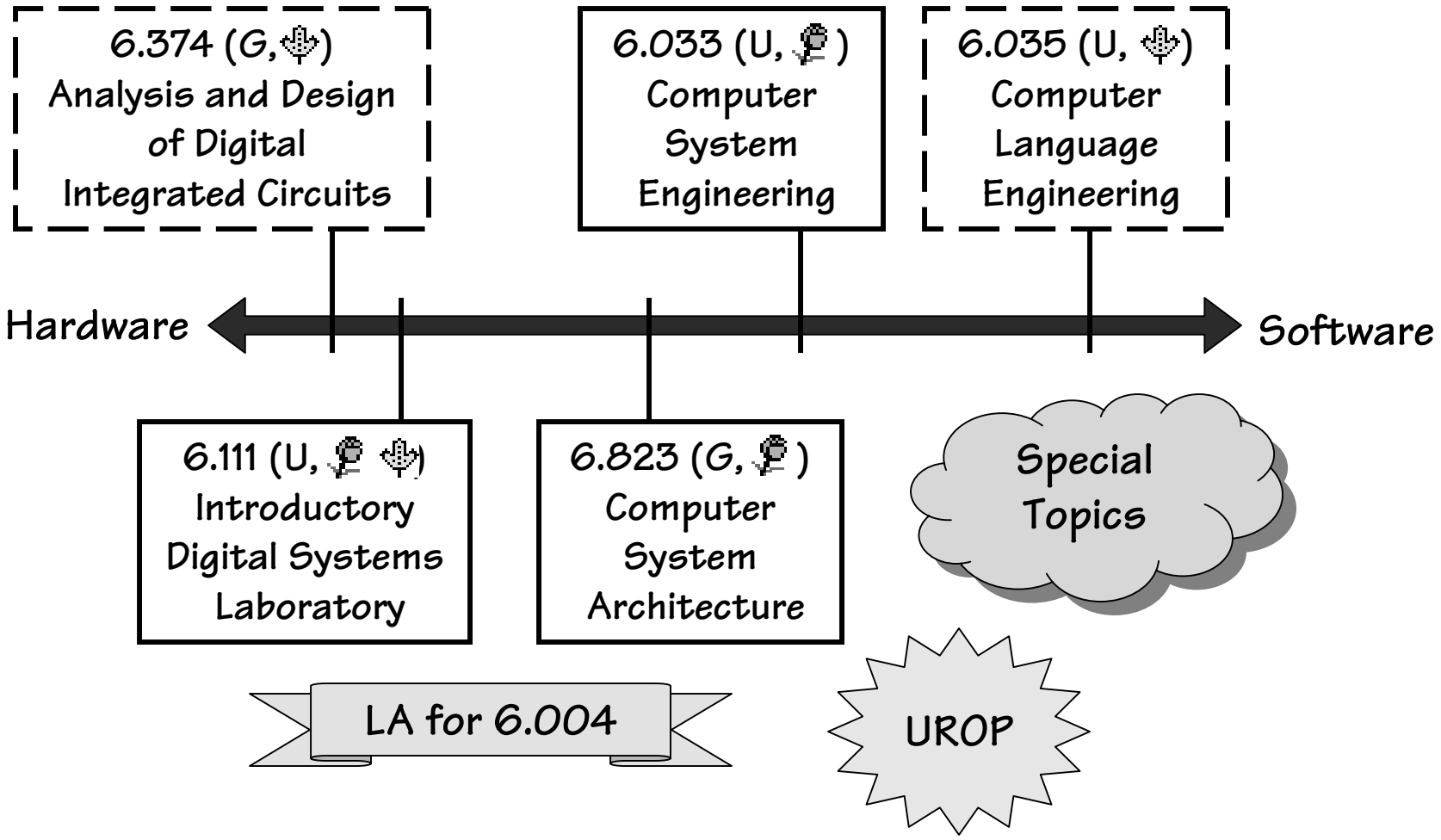
Computer Systems

?

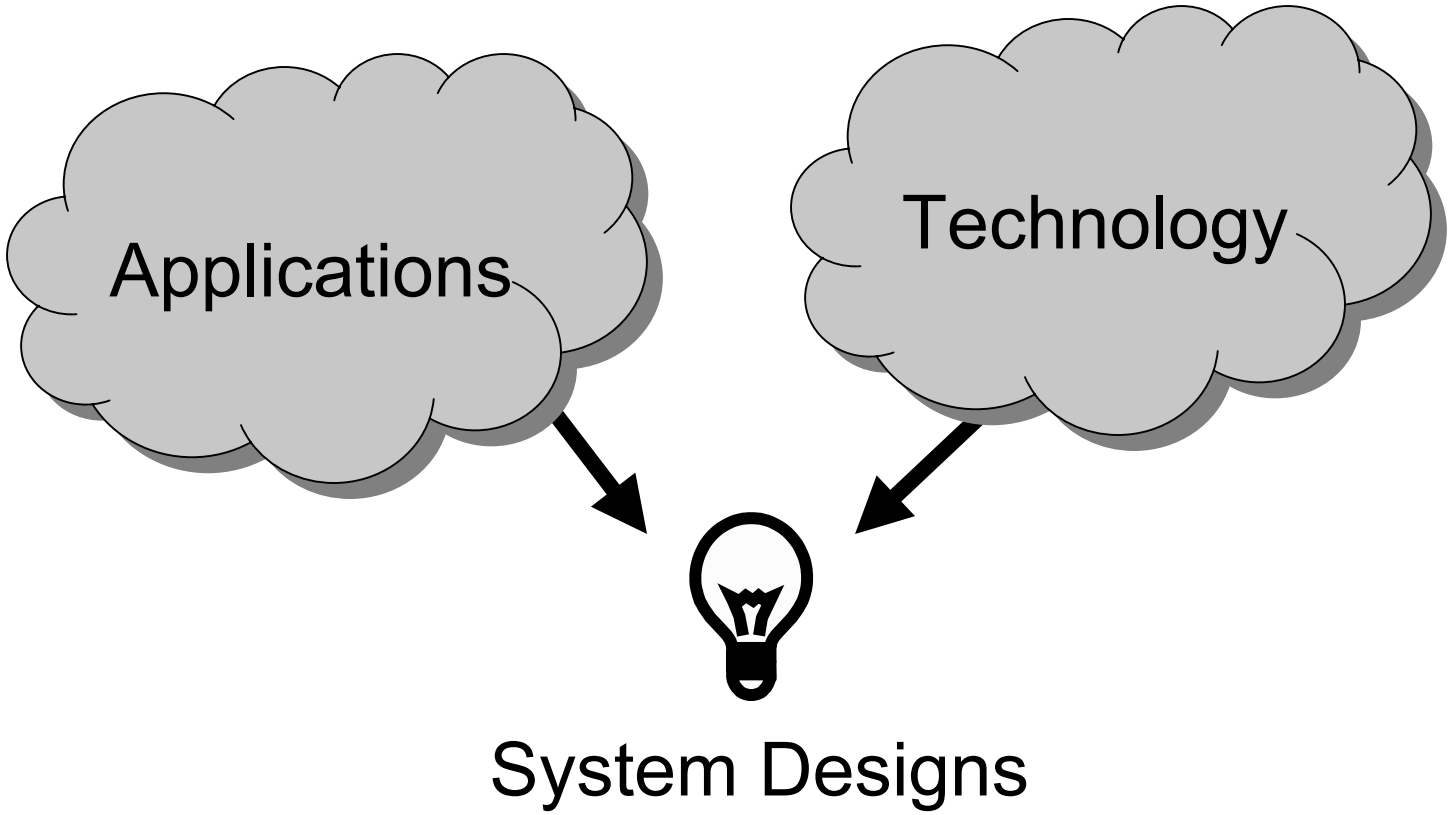
Computing Theory
Instruction Set Architectures
Beta implementation
Pipelined Beta
Software conventions
Memory architectures

Interconnect
Virtual machines
Interprocess communication
Operating Systems
Real time, Interrupts
Parallel Processing

Follow-on Courses



What is Computer Architecture?



Designing Across Trends

2001:
MS Office,E-commerce, WWW, Digital Cameras, MP3, DVD, Games, Wireless,

Killer apps in 2010?

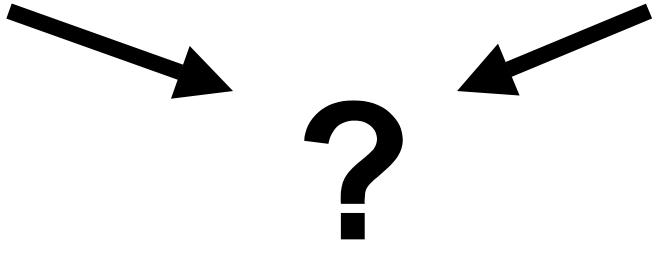
2001:

100 million transistors/chip
10x transistors every 5 years
1% performance/week!

CMOS 2010:

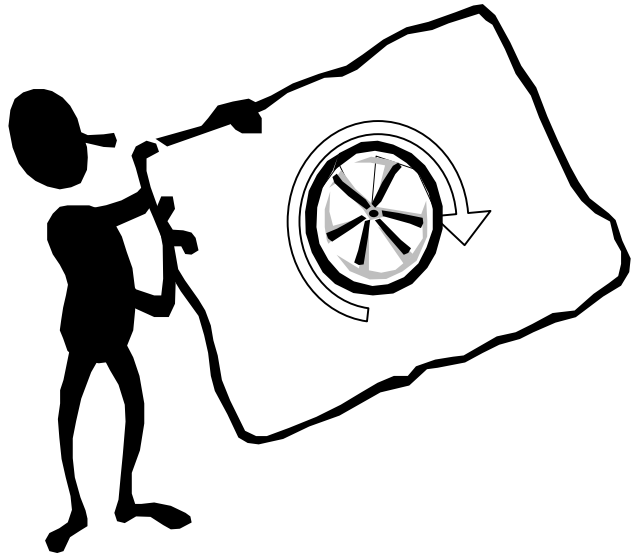
10 billion transistors

10 GHz clock



Modern Processor

- Execute seven instructions per cycle
- Out-of-order execution
- Register renaming
- Speculative execution
- Predicated execution
- Multi-way branches
- 10ns cycle time



… circa 1965-69!!!
IBM Advanced Computing Systems project
(http://www.cs.clemson.edu/~mark/acs.html)

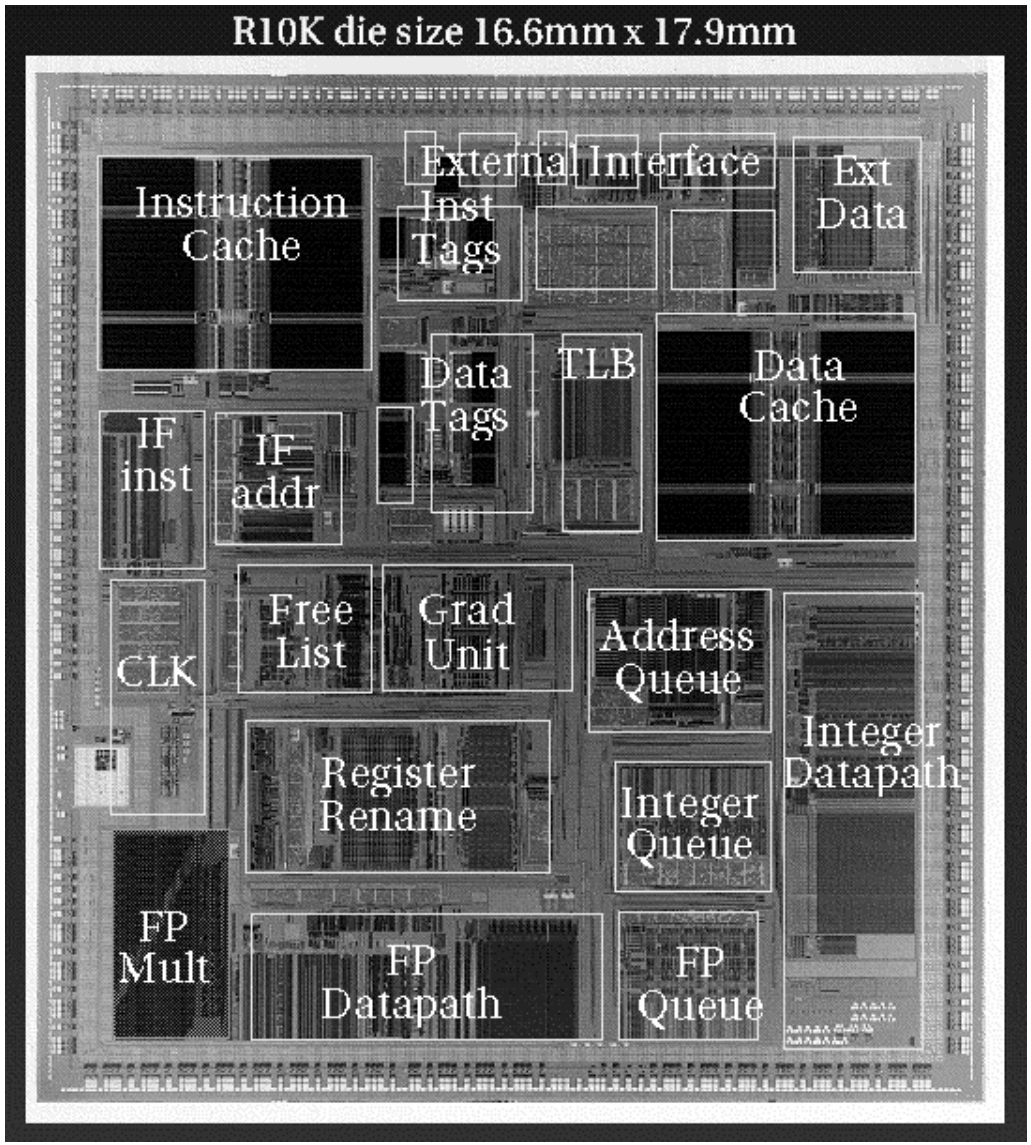
An Exciting Time For Architects!

- Monstrous transistor budgets
  - ~100 Pentium-IIIs on one die!
  - ~1000 RISC cores on one die!
- New application areas open to new ISAs
  - Not just x86 running MS Office
- No more supercomputer tricks left to steal
  $\rightarrow$ *need new architectural ideas!*

Future Directions

- The Giant Uniprocessor
- Simultaneous Multithreading
- Tiled Architectures
- Intelligent RAM
- Low-Power Architectures
- Reconfigurable Computers

MIPS R10000



- 0.35µm CMOS, 4 metal layers
- Four instructions per cycle
- Out-of-order execution
- Register renaming
- Speculative execution past 4 branches
- On-chip 32KB/32KB split I/D cache, 2-way set-associative
- Off-chip L2 cache
- Non-blocking caches

Compare with 6.004 5-stage pipe

- ~1.6x performance SPECint95
- ~5x CPU logic area
- ~10x design effort

Giant Uniprocessor Scaling

- Difficult to find more instructions to execute in parallel
- Circuitry needed to execute N instructions in parallel grows faster than N^2
- Design/verification complexity growing exponentially
  - Intel: >300 engineer-years to verify P6

Simultaneous Multithreading

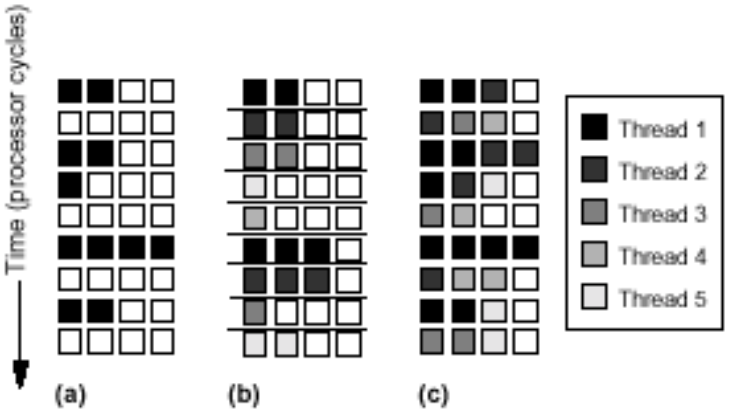
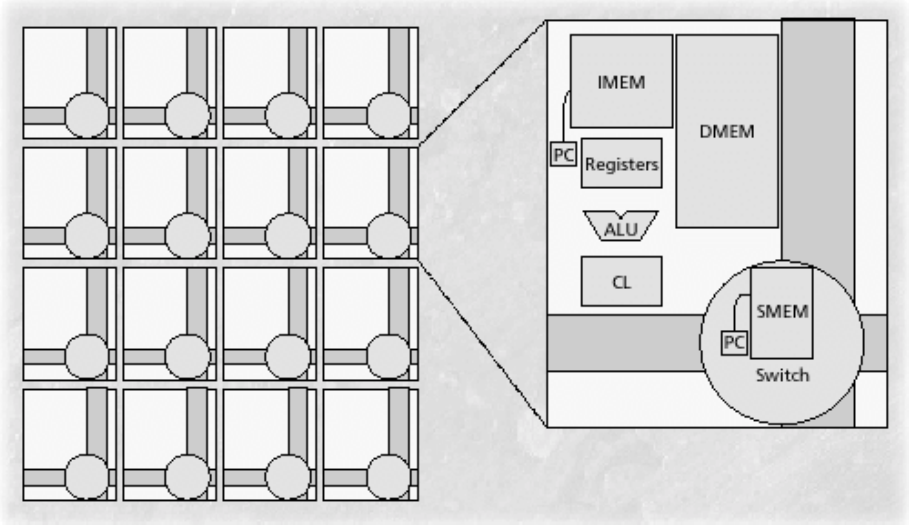


Figure 1. How architectures partition issue slots (functional units): a superscalar (a), a fine-grained multithreaded superscalar (b), and a simultaneous multithreaded processor (c). The rows of squares represent issue slots. The processor either finds an instruction to execute (filled box) or the slot goes unused (empty box).

- Use resources of wide superscalar to execute multiple threads at same time
- High single thread performance plus good hardware utilization with multiple threads
- Commercial implementation coming, Alpha 21464
  - 8-way out-of-order, speculative, superscalar execution engine
  - our hardware threads per C U share execution engine
- Disadvantages: Complexity, Area

(http://www.cs.washington.edu/research/smt/papers/ieee_micro.pdf)

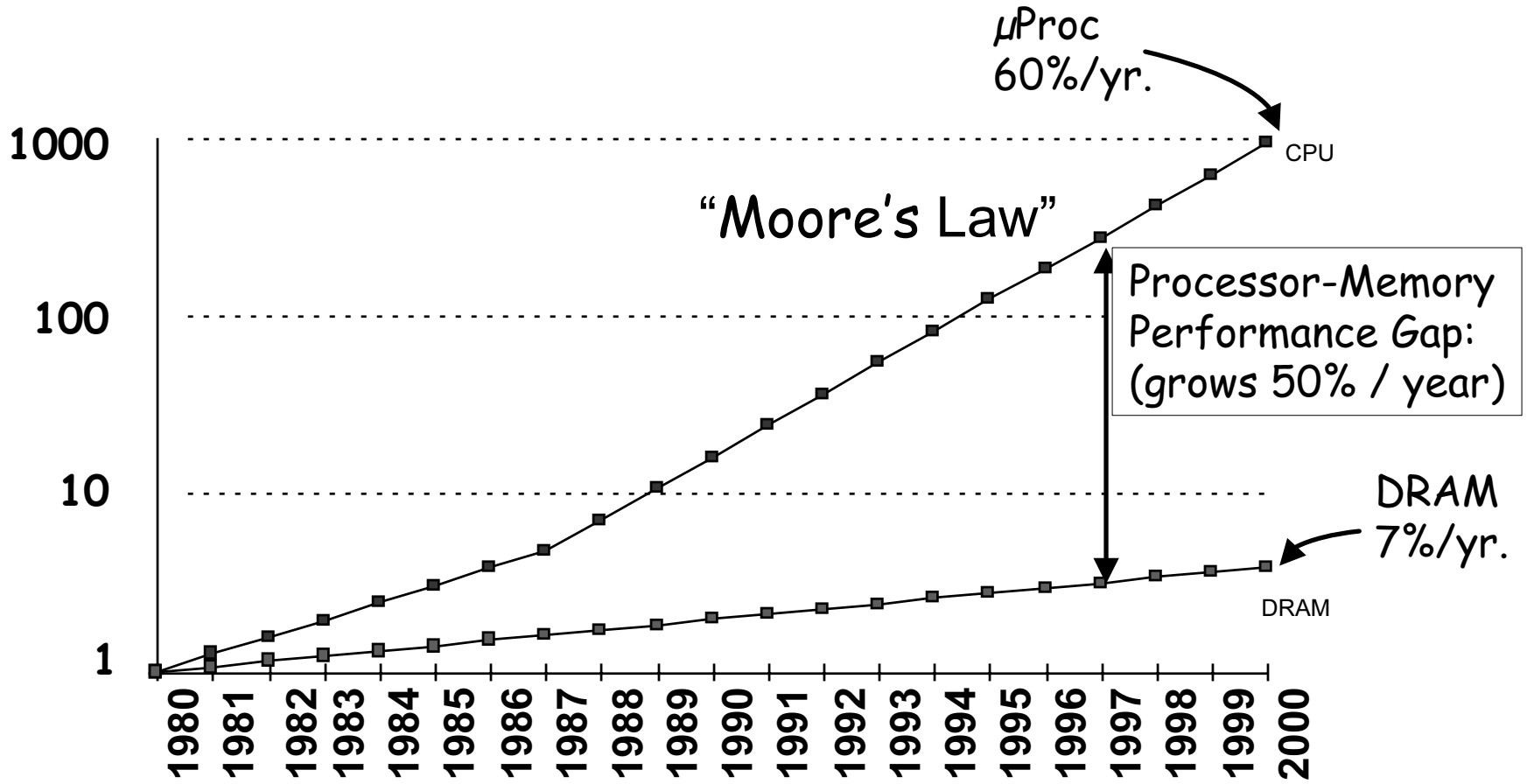
MIT RAW Processor



- Divide silicon die into 2D array of processing "tiles"
  - each tile has simple processor+memory+network switch
  - simple VLSI design process, layout one tile then replicate
- Compiler manages communication between tiles
- Can build virtual "giant" CPUs in software
- Disadvantage: Compiler complexity, brittle performance

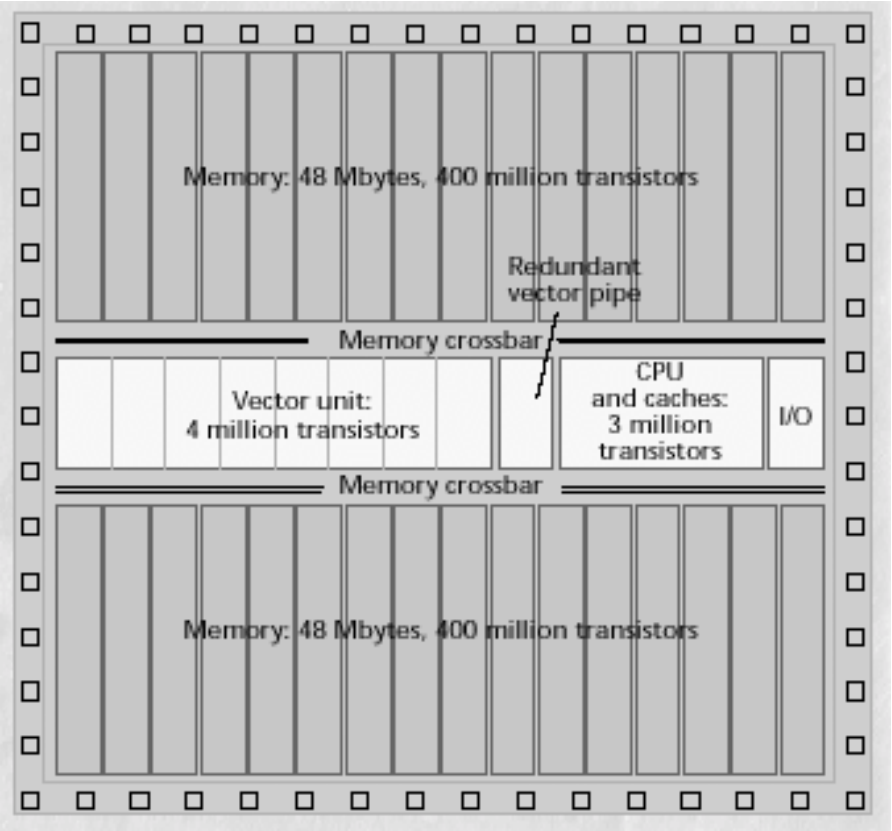
(Prof. Agarwal, http://www.cag.lcs.mit.edu/raw/)

Processor-DRAM Gap (latency)



[From David Patterson, UC Berkeley]

U.C. Berkeley IRAM



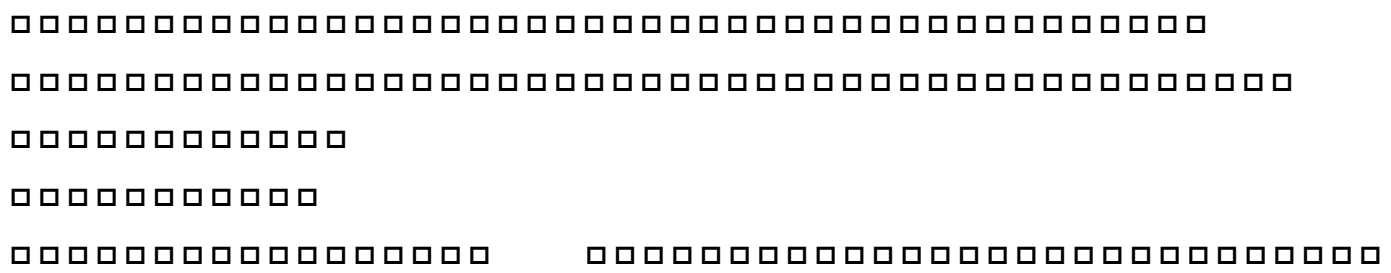
Microprocessor & DRAM on a single chip:

- on-chip memory latency 5-10X, bandwidth 50-100X
- improve energy efficiency 2X-4X (no off-chip busi
- serial I/O 5-10X v. buses
- smaller board area/volume
- adjustable memory size/width

(http://iram.cs.berkeley.edu/)

Low-Power Architectures

- Improve energy efficiency of programmable processors by re-examining the hardware-software interface
- Goal: reward compile-time knowledge with run-time energy savings



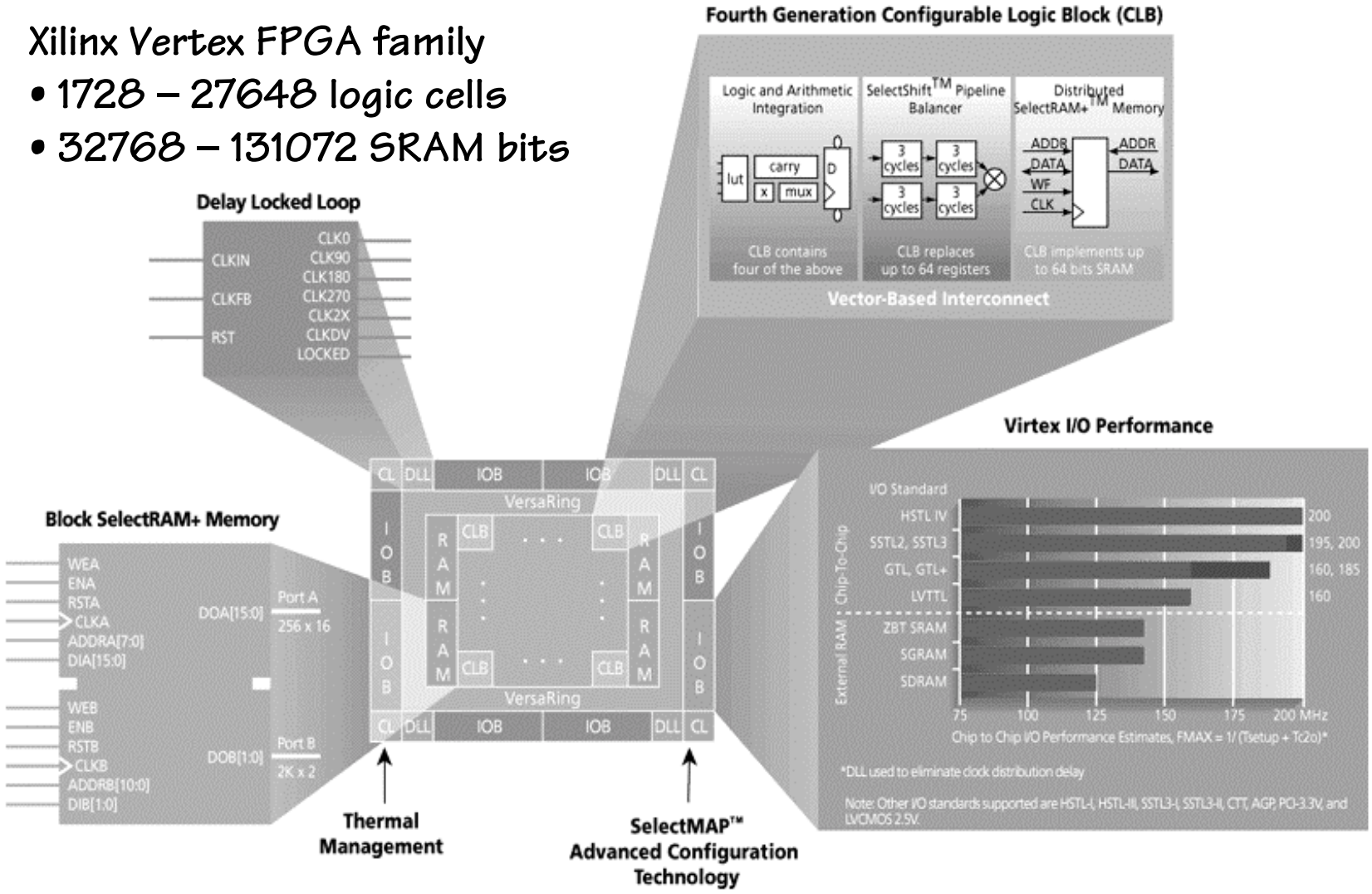
- Energy reductions:
  - no tag RAM read/compare or no tag CAM search
  - only low order address bits need to be computed
  - no TLB lookup for physically tagged caches

  → Reduces cache access energy to just RAM read

(Prof. Asanovic: http://www.cag.lcs.mit.edu/scale)

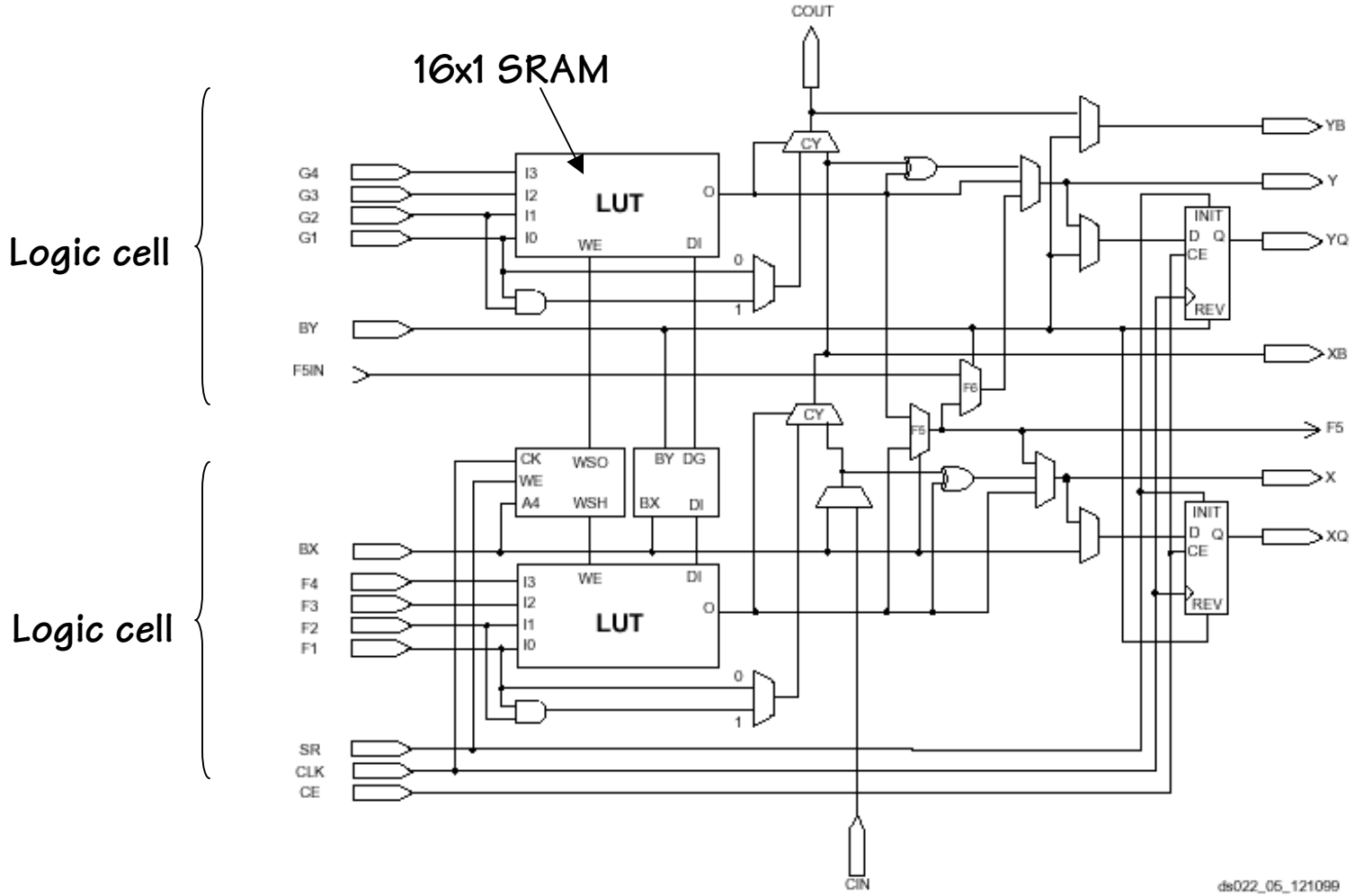
Reconfigurable Computing

Xilinx Vertex FPGA family
- 1728 – 27648 logic cells
- 32768 – 131072 SRAM bits



Xilinx Vertex CLB

1 Configurable Logic Block = 4 Logic Cells (organized in pairs)



Technology Trends

One process generation (defined as 30% linear shrink of feature size):

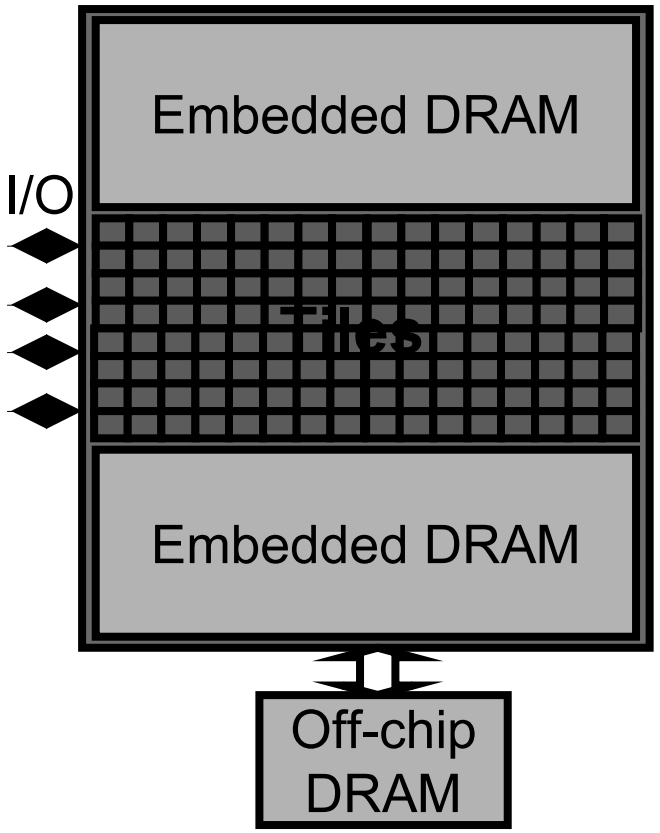
- halves die size (or doubles transistor budget for same size die)
- lowers manufacturing costs
  - more die/ wafer, yield improves exponentially with die size
- improves intrinsic speed (CV/I) of transistors by 30-50%
- 25% reduction in power supply voltage, cuts CV^2F power in half

Tough problems:

- increased design/verification time
- interconnect delay
  - lower C by using insulators with low dielectric constants
  - lower R with thicker metal layers or changing Al to Cu
- heat dissipation, external connections
  - improved packages
    - plastic vs. ceramic vs. organic
    - pins vs. ball grid

2010 Architecture?

Giant uniprocessors (maybe with SMT) remain popular in markets where software is the main expense.



Tiled (VLIW/reconfigurable/vector) machines become popular in systems with resource constraints (hardware cost, low power, hard real time)

- 10 GHz processor clock
- 5 GHz network clock
- 128 processing tiles
- >5 TFLOPS peak (32b FLOPS)
- >40 TOPS peak (8b OPS)
- 1GB on-chip DRAM
- 100 GB/s off-chip DRAM interface
- 100 GB/s I/O
- 25x25mm^2 in 0.050μm CMOS

Think Outside the Box

Don't be IMPRISONED by 6.004 abstractions! Consider

Alternatives to State:

- Precise state of complex, asynchronous systems is intractable;
- Lessons from engineering revolutions: feedback, equilibria.

Alternatives to Logic:

- Approximate behavioral models rather than truth tables?
- RAM with READ, PUNISH operations?

Alternatives to Programming:

- Goal synthesis as engineering tool
- Learning, Training, Evolution

Alternatives to these alternatives! That's your job....

THE END!

Pens, pencils, paper
they attempt to solve problems
that teachers set forth.

The only problem
with Haiku is that you just
get started and then