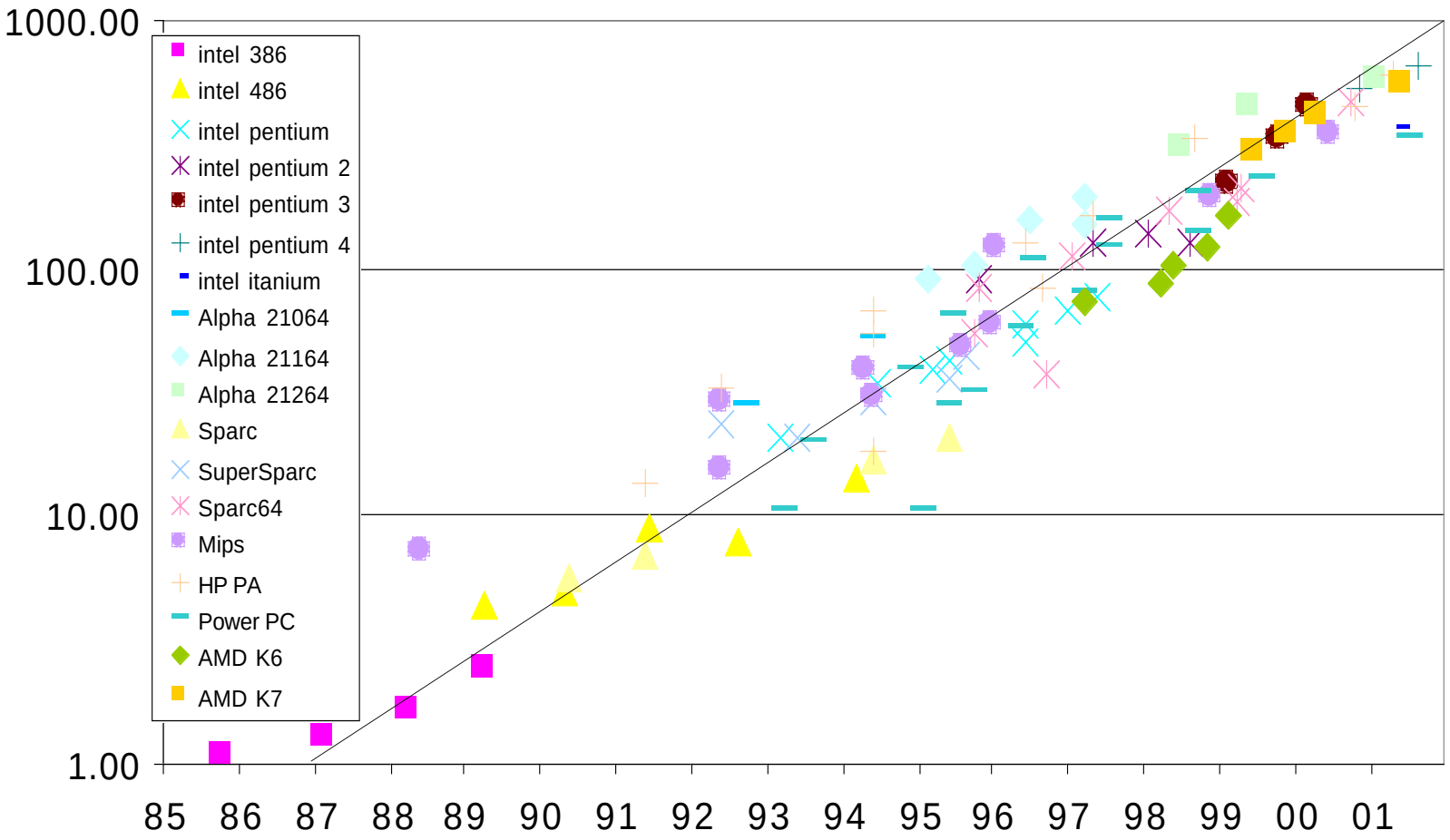
Lecture 2 continued

Wire Modeling

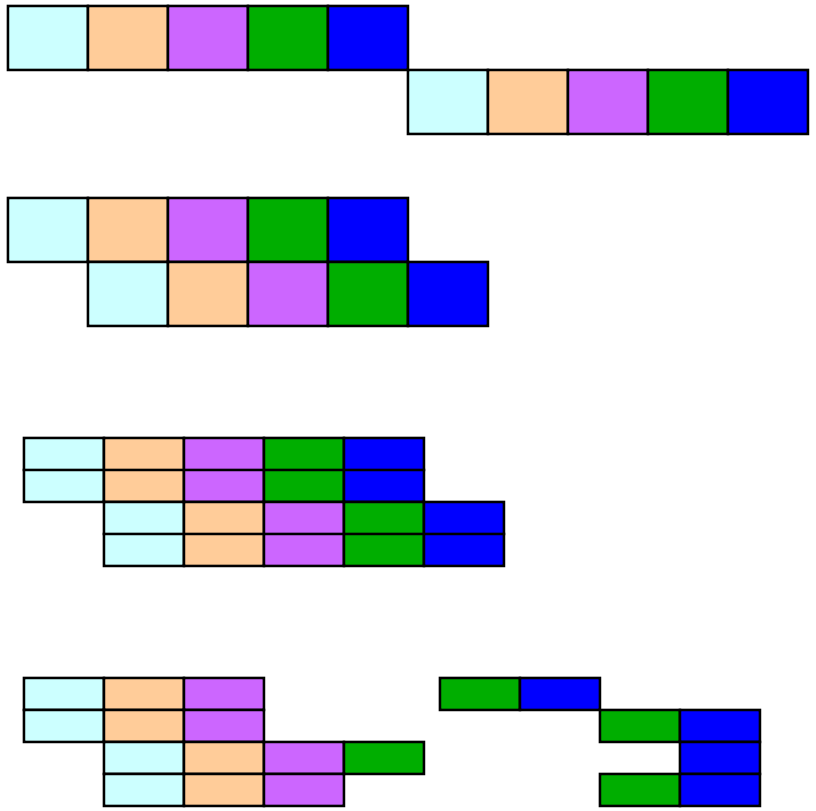
Mark Horowitz
Computer Systems Laboratory
Stanford University
horowitz@stanford.edu

Fruits of Scaling – SpecInt 2000



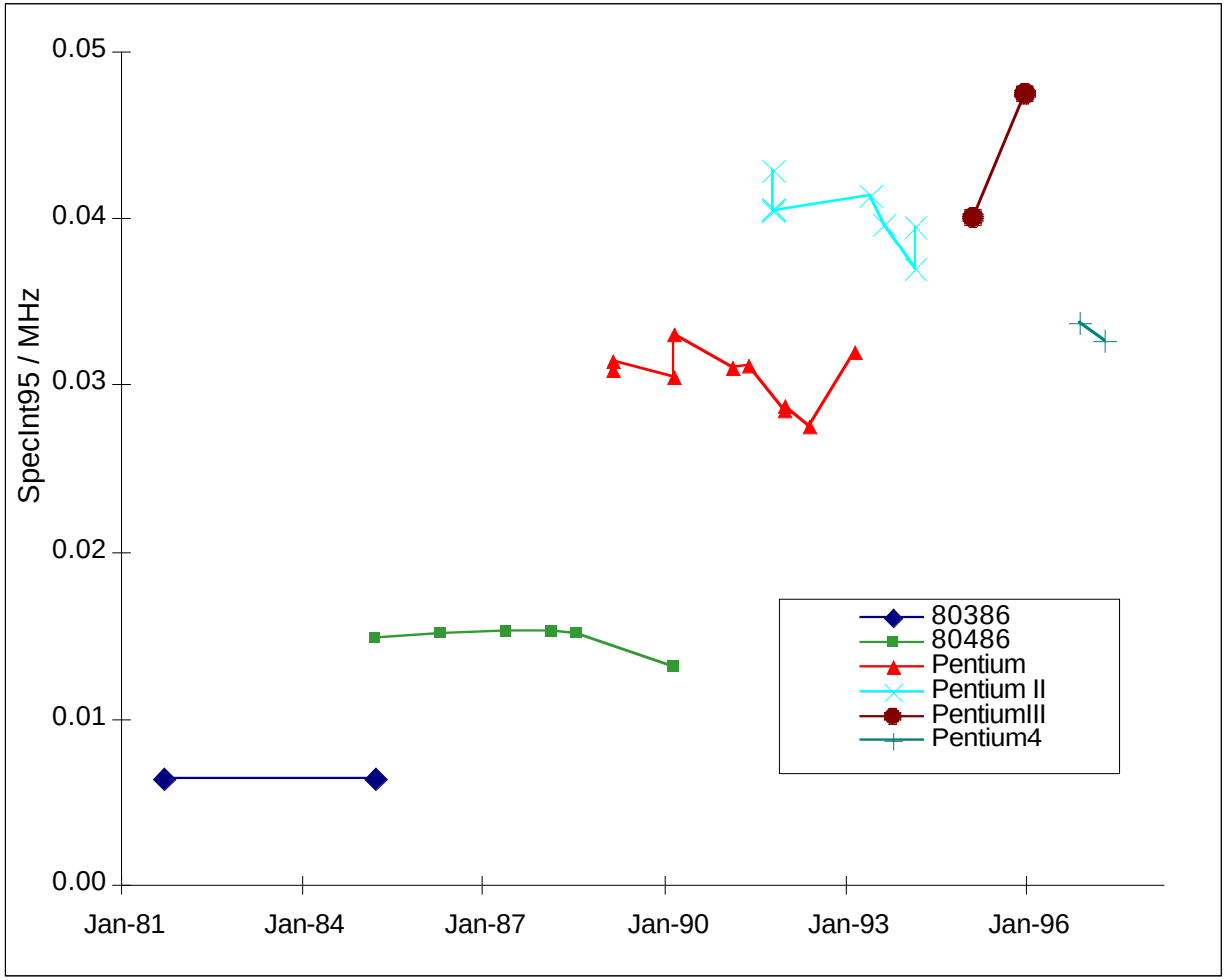
Computer Architect's Job

- Convert transistors to performance
- Use transistors to
  - Exploit parallelism
  - Or create it (speculate)
- Processor generations
  - Simple machine
    - Reuse hardware
  - Pipelined
    - Separate hardware for each stage
  - Super-scalar
    - Multiple port mems, function units
  - Out-of-order
    - Mega-ports, complex scheduling
  - Speculation
- Each design has more logic to accomplish same task (but faster)

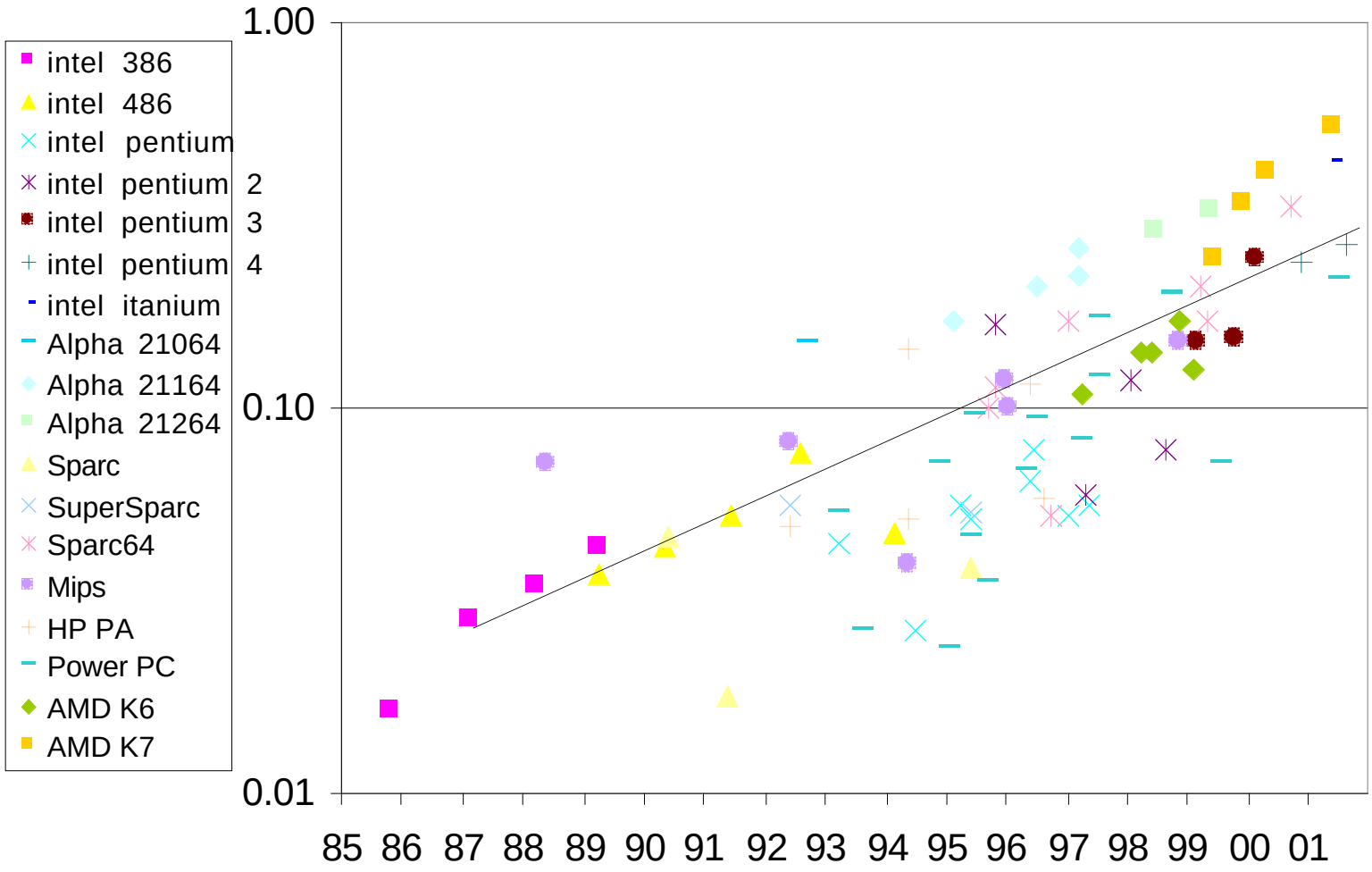


Architecture Scaling

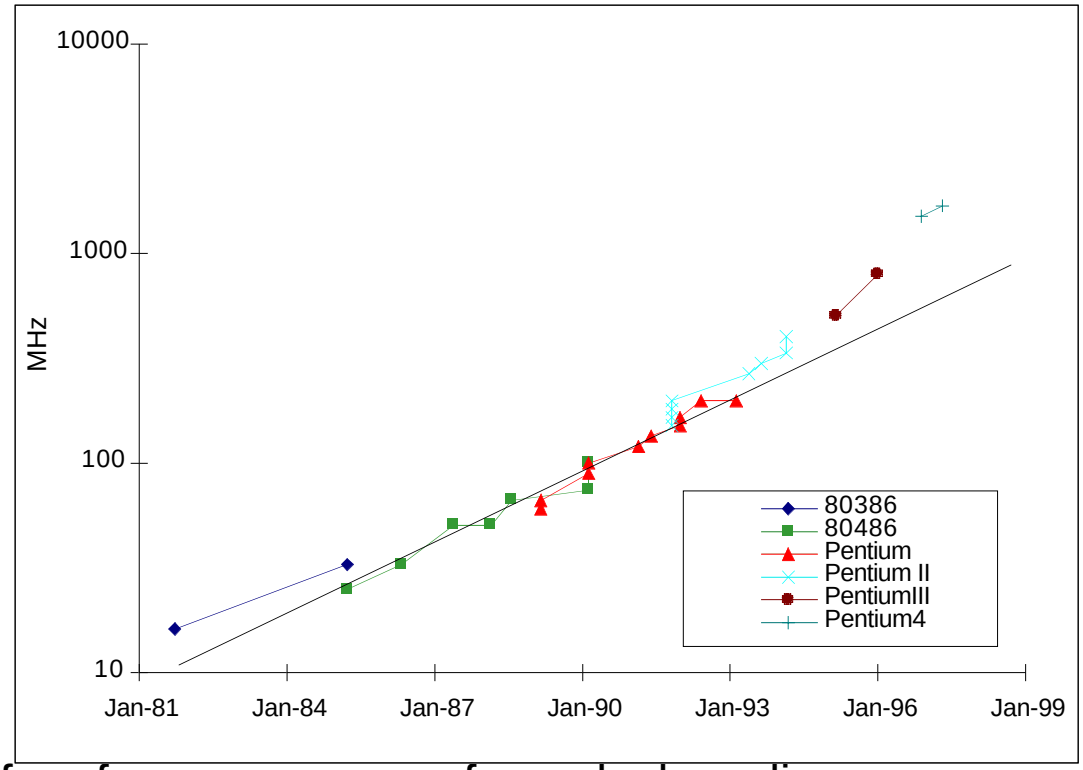
- Plot of IPC
  - Compiler + IPC
  - 1.5x / generation
  - Until PIII, now falling
- There is a lot of hardware to make this happen
  - It costs in power



SpecInt/MHz

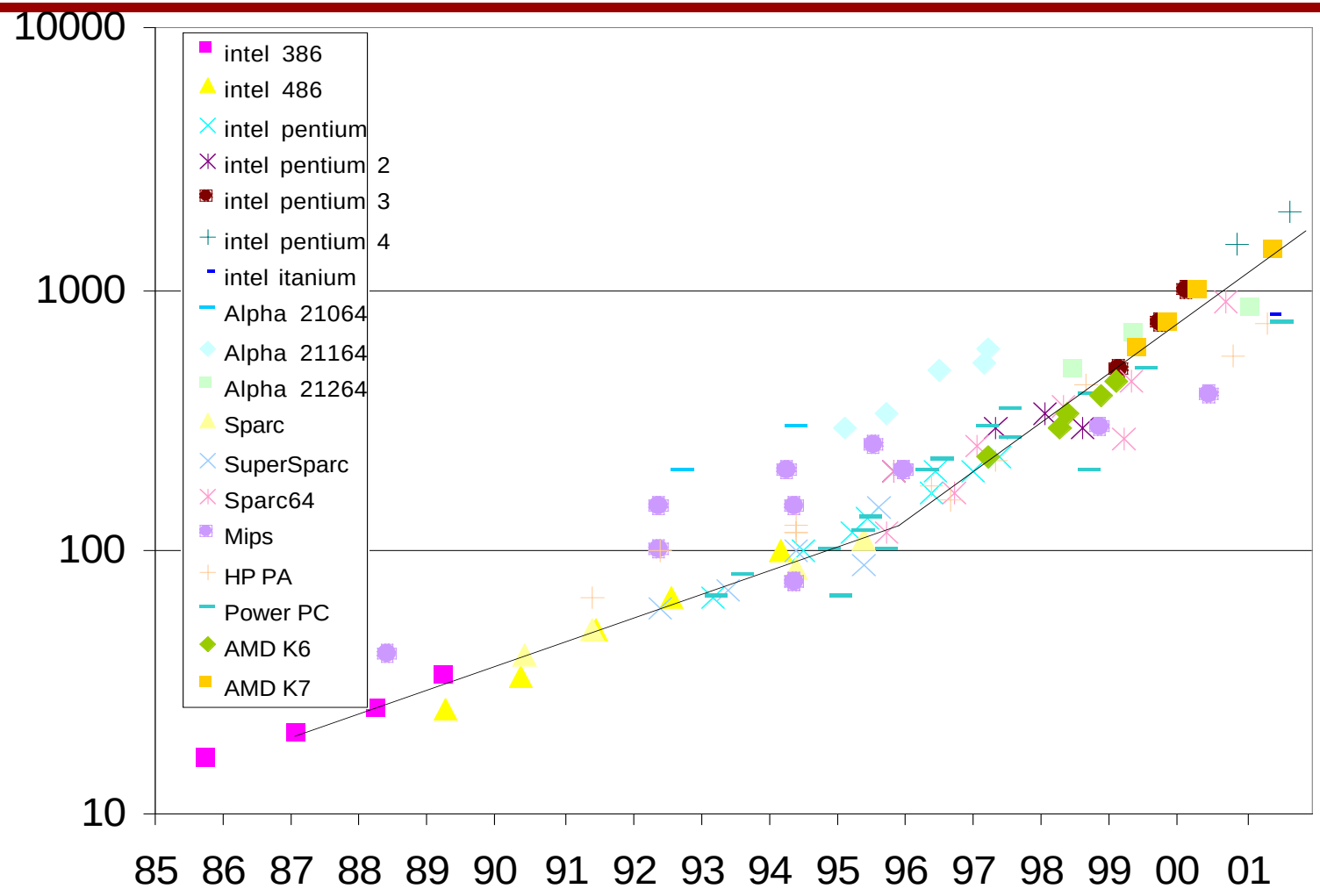

intel 386
intel 486
intel pentium
intel pentium 2
intel pentium 3
intel pentium 4
intel itanium
Alpha 21064
Alpha 21164
Alpha 21264
Sparc
SuperSparc
Sparc64
Mips
HP PA
Power PC
AMD K6
AMD K7
1.00
0.10
0.01
85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 00 01

Clock Frequency



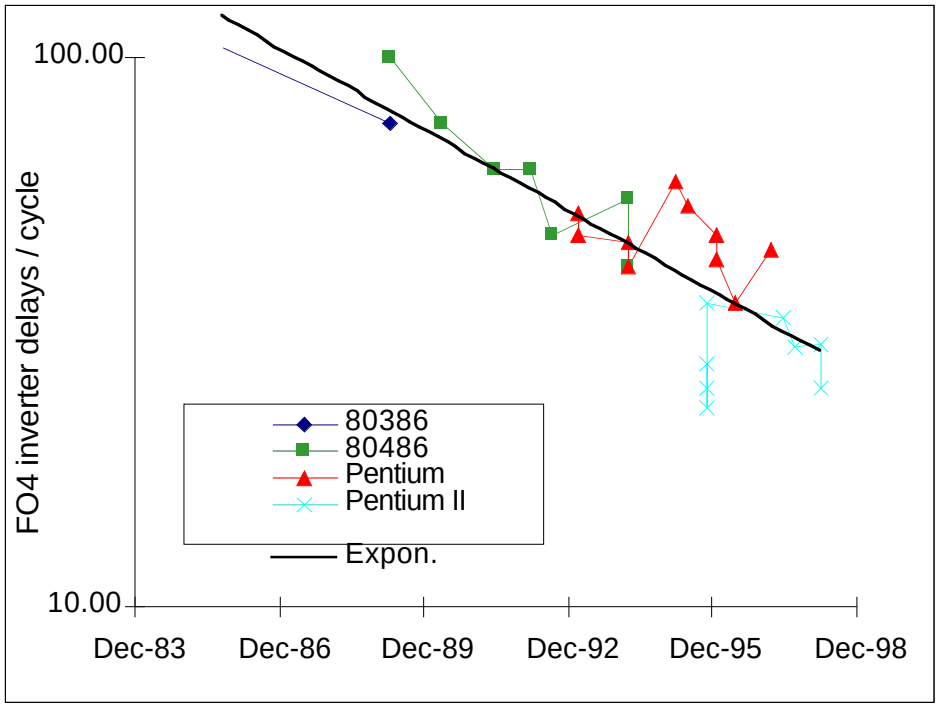
- Most of performance comes from clock scaling
  - Clock frequency double each generation
- Two factors contribute: technology (1.4x/gen), circuit design

Clock Frequency Scaling



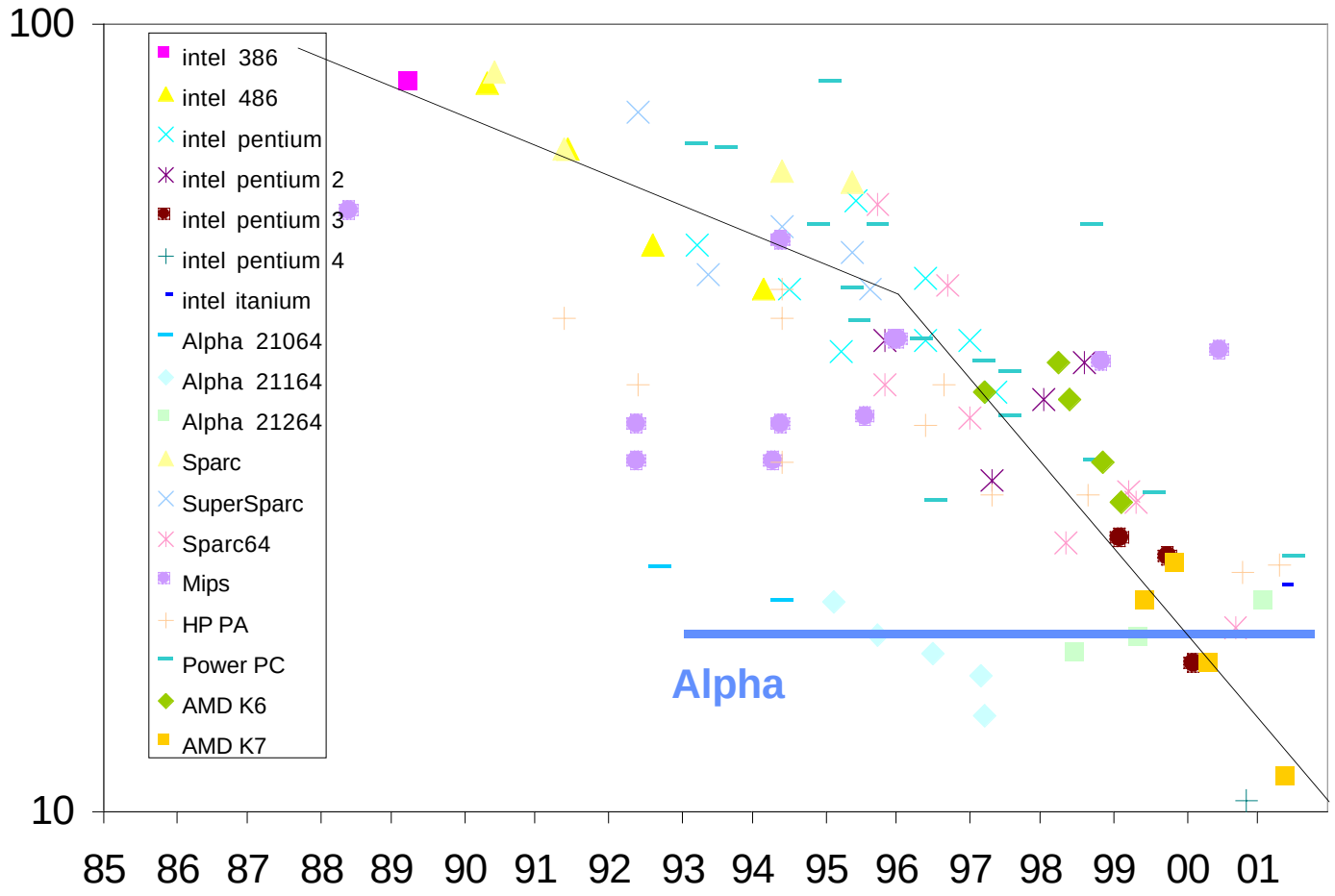
Clock Cycle in ‘FO4’

- Clock speed has been scaling faster than base technology
- Number of FO4 delays in a cycle has been falling



- Number of gates decrease 1.4x each generation
- Caused by:
  - Faster circuit families (dynamic logic)
  - Better optimization
  - Better micro-architecture
  - Better adder/mem arch
- Implies increased power

Clock Cycle in ‘FO4’



Note that the points at around 10 FO4 are not correct. The FO4 for these technologies is about _ my simple formula

Wire Layers

Not all wiring layers have the same characteristics

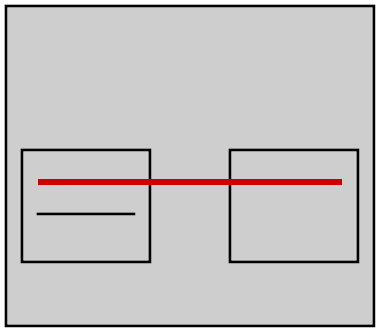
- Today have three types of levels
  - M1, sometimes M2
    - Finest pitch, highest resistance, local interconnection in a cell
  - M2-4?
    - Normal routing level, most of the wires
  - M5+
    - Thick coarse metal, used for global wires
    - When scaling forces thinner metal, create new top layer

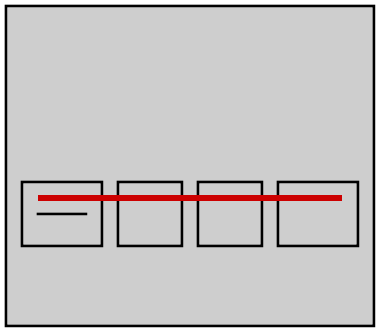
Noise Issues

- Two main noise sources for wires
  - Capacitance coupling
  - Inductive coupling
- Capacitance coupling is mostly a nearest neighbor issue
  - High aspect ratio wires make this worse
  - Real push for low-κ dielectric between wires
- Inductive coupling
  - Is much more complex to analyze
    - Depends on where the return currents flow
  - Reduce these problem by design constraints
    - Gnd returns in buses, power and gnd planes (e.g. 21264)

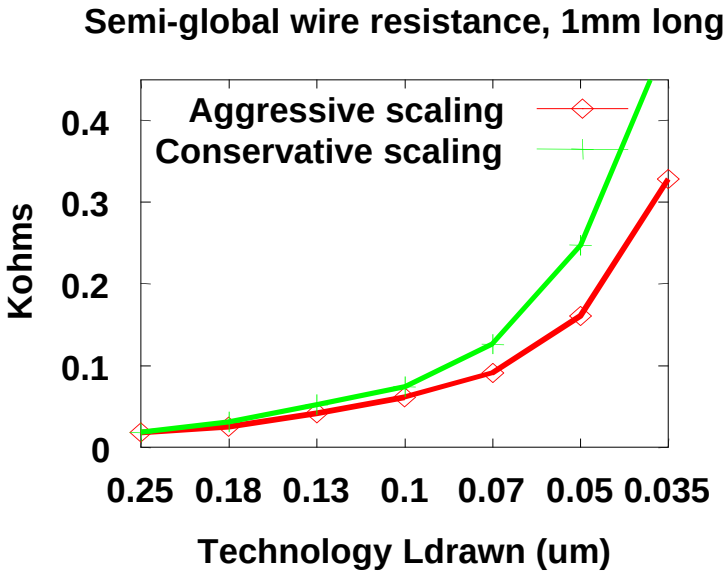
This is, and will continue to be, managed by design rules

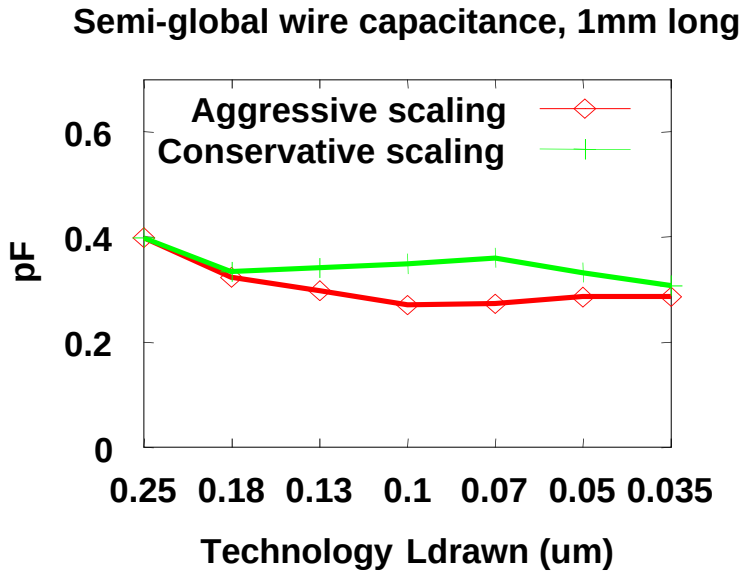
Scaling Global Wires



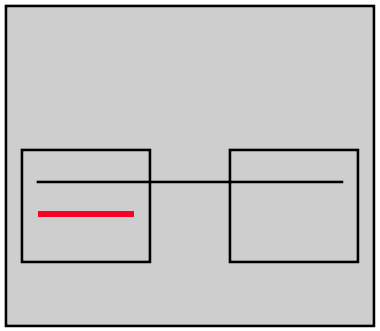


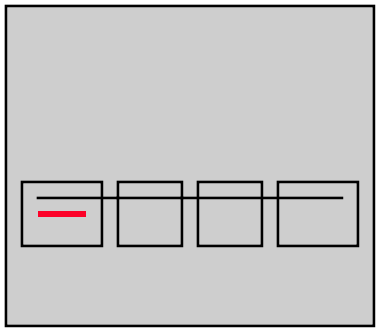
- R gets quite a bit worse with scaling; C basically constant



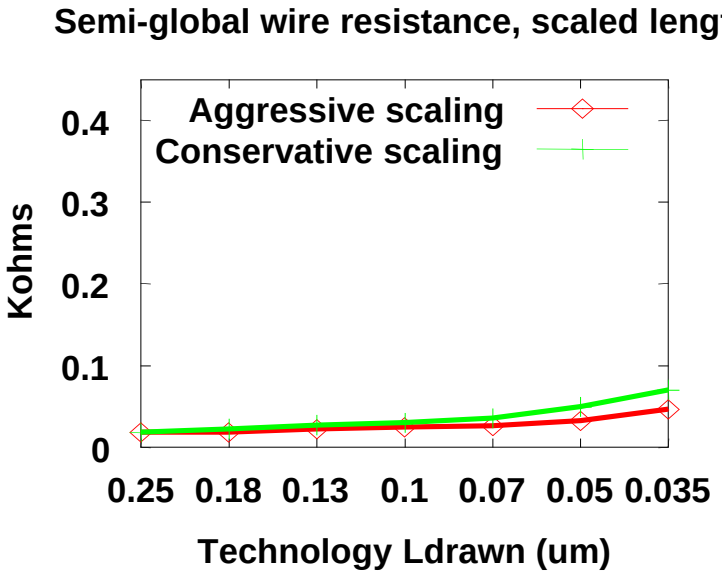


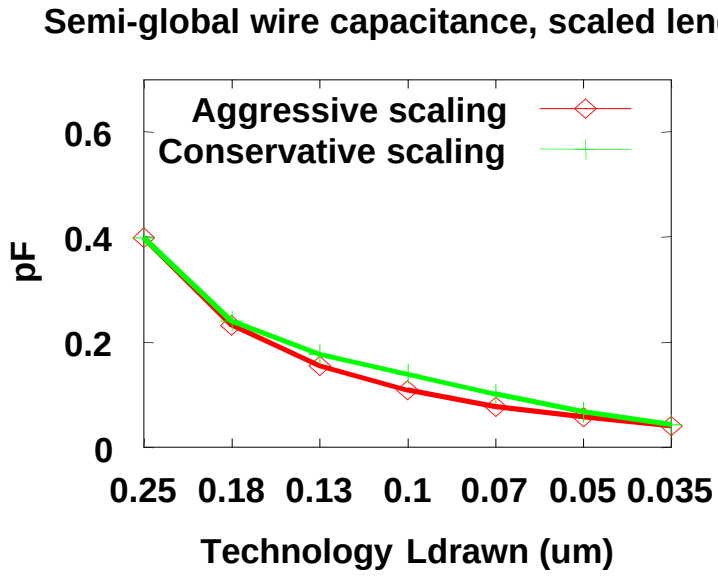
Scaling Module Wires





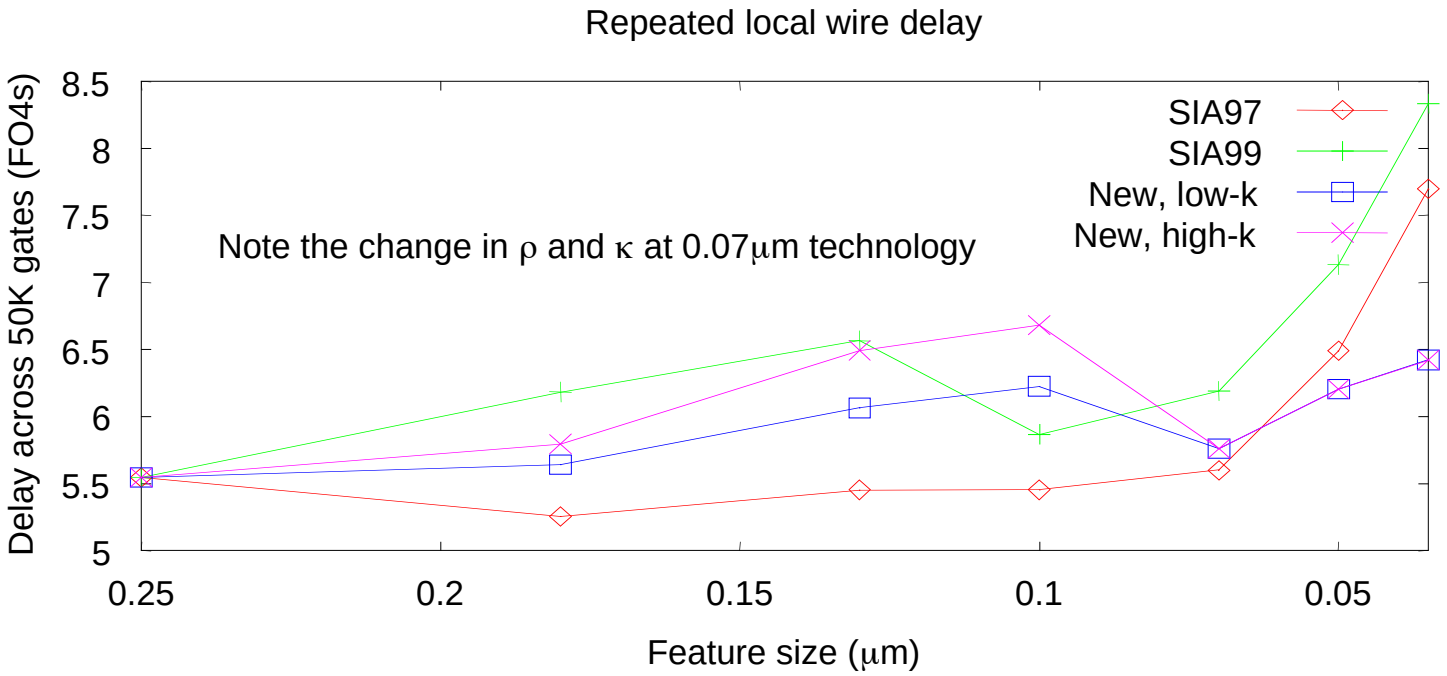
- R is basically constant, and C falls linearly with scaling





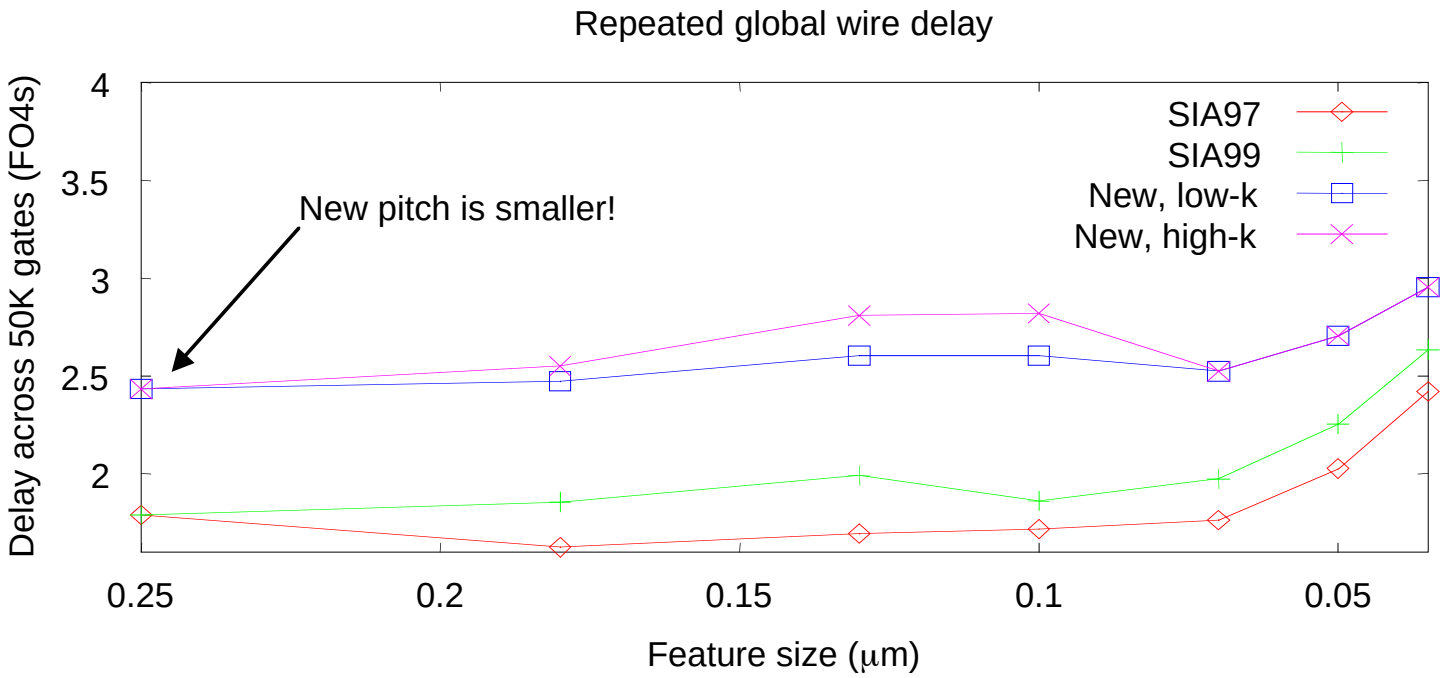
Scaling Local Wires

- Normalize a local wire to an inverter loaded with fanout=4
  - This wire spans the edge of a block of 50K gates
- Bandwidth across a 50K gate block scales with technology
  - $1.6/L_{drawn}$ Tb/s (6.4 Tb/s in 0.25μm; ~ 30 Tb/s in 0.05μm)



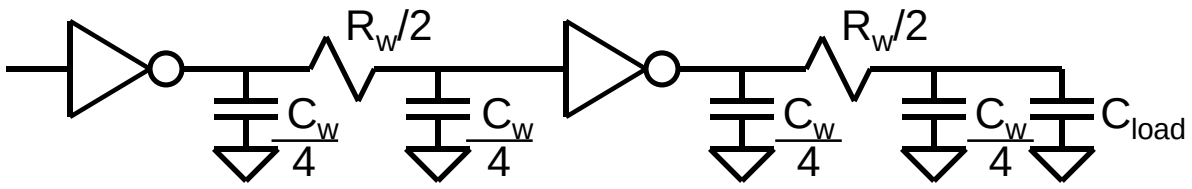
Local Wires on Top Metal

- Also look at global wires that span edge of 50K gate blocks
  - Nobody would really route in local metal, anyway



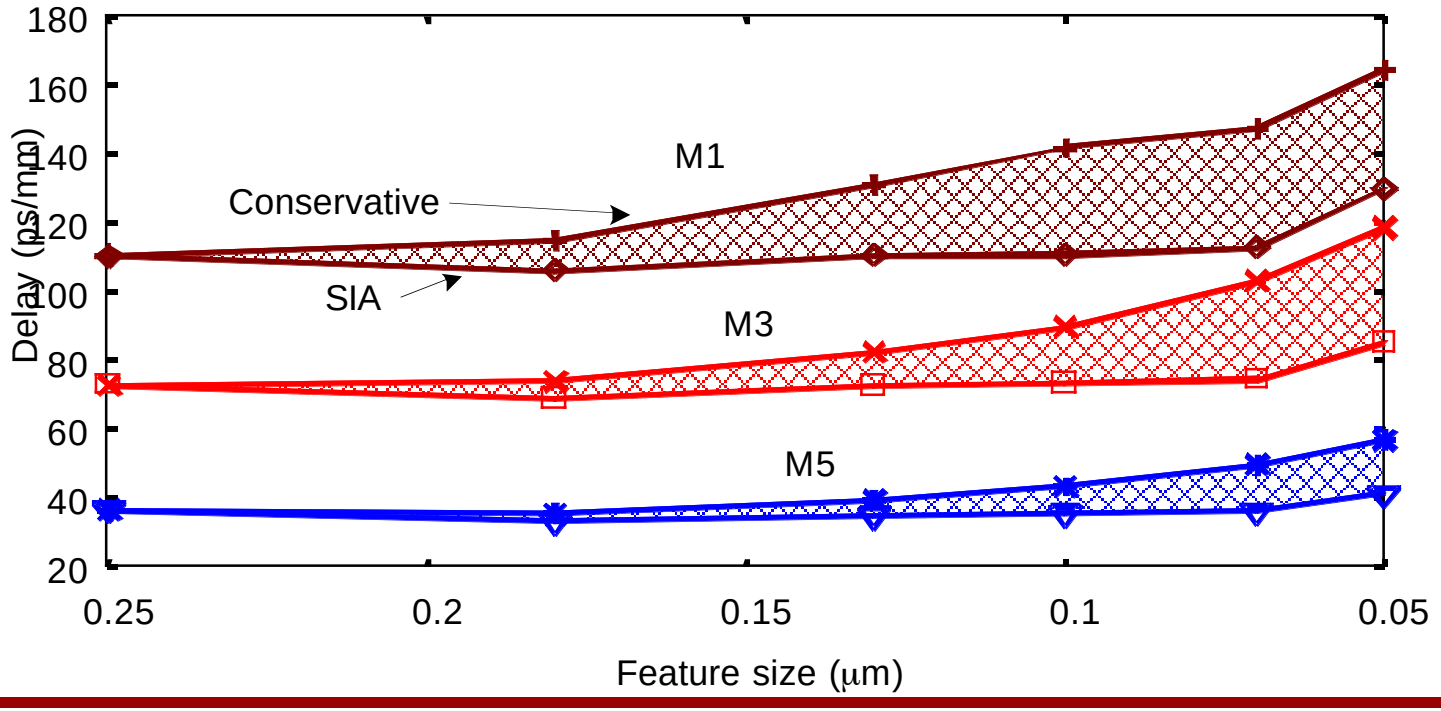
Wires Become Active

- Since the wire delay increases with length
  - Use regeneration to speed up long wires
- Velocity of wire become geometric mean = k (FO4 R_w C_w)$^{1/2}$
  - Compensate for decreasing wire performance
  - Repeater density increases w/ technology
  - But roughly constant in grids
- Note that wires are now active
  - Can think of doing some interesting stuff in the wires
    - More on that later



Signal Velocity for Repeated wires

- Under SIA scaling, pretty constant over many generations
- Under conservative scaling, slow change at sub-0.1μm techs
  - Makes wire delay increase slowly



Caveat:

- There is some concern that these wire models are optimistic
  - Overhead of barrier layers
  - Scattering changes when the wire width < mean free path
- Potential for conductivity of Cu to fall at small dimensions
  - Surface design will be critical for low resistance wires