# A Computer Architecture Workshop: Visions for the Future Celebrating Yale@75

**September 19th, 2014**

Trevor Mudge
The University of Michigan, Ann Arbor

# Visions for the Future?

- We need to think out-of-the-box
- Get back to the basics of computing
- Do something radical
- Get rid of the von Neumann straight jacket
- Our abstractions don't expose the basic logical fabic
- I propose the following: Let's take a leaf from toys—LEGO
- There's a solution that's been with us a long time....

- It's a basic building block that digital systems can be constructed from
- No –not a NAND gate
- It has a higher level of abstraction

# The Universal Logic Module—ULM

A ULM of this type consists of a universal Boolean function $U(z_1, z_2, \cdots, z_m)$ and $s$ auxiliary functions $g_1(z_1, z_2, \cdots, z_n)$ , $\cdots, z_n)$; see Fig. 1. A Boolean $_m)$ is universal in $n$ variables $x_1$, Boolean function of $n$ variables realized by an appropriate substitu- e set $\mathfrak{z}$ for each $z_j$. In this model, the $\cdot, x_n, \bar{x}_n, 0, 1, g_1, g_2, \cdots g_s\}$. The $g_s$ and the first $n$ arguments of $U$ are sponding uncomplemented $x_i$; i.e., hat if $s = 0$, the model reduces to that

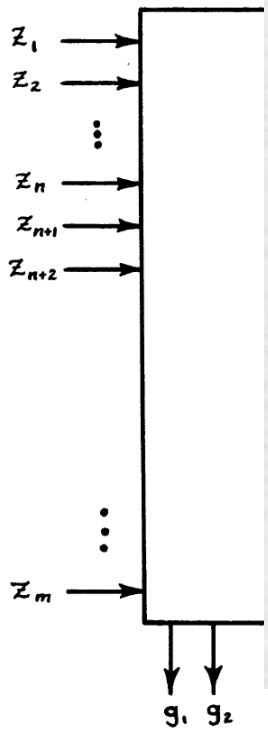



Fig. 1. General form of the ULM with interconnected external terminals.

# Golden Age—Dennard Scaling Worked

- About every 2 years we got:
- 2× area
- 40% increase in frequency
- 50% reduction in power
- No change in power density

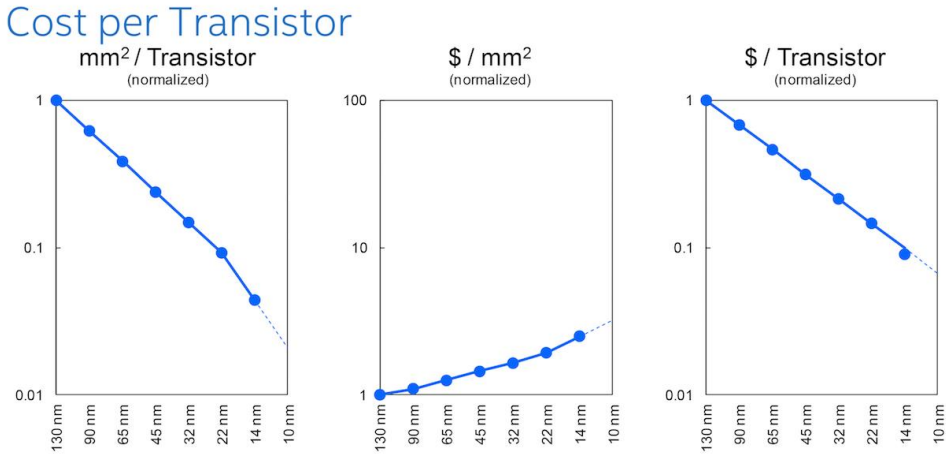

4

# Turn of the Century

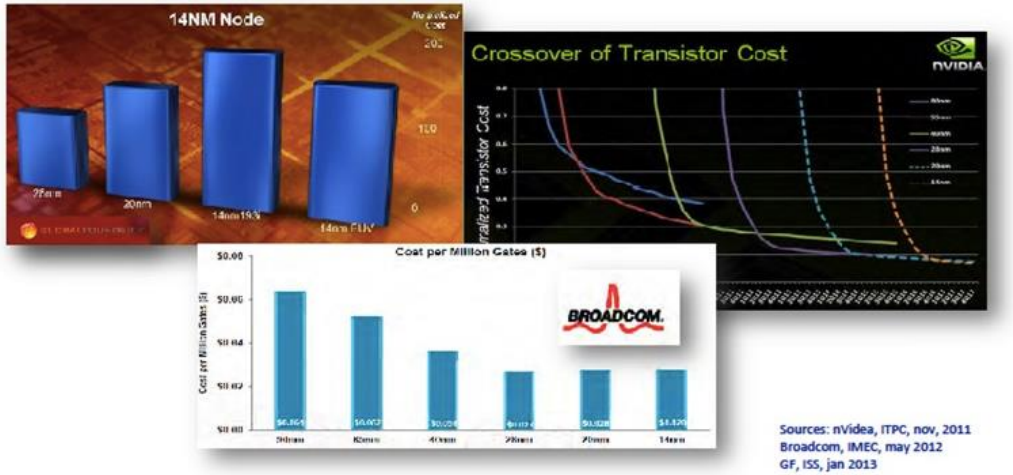
- Cracks appeared
- First crack: frequency scaling became a problem
  - Vdd got stuck & power density started to creep up
  - Clock frequency could no longer be a selling point
  - Core count was the new metric
- Later:
  - Each new tech node no longer showed power and frequency scaling
  - Node numbers became symbolic names —still followed the $1/\sqrt{2}\times$ sequence
  - Only area shrank with new nodes

# What About Cost?

- Two Views
- Intel




- Everyone else
  - cost per transistor reached a minimum at 28 nm



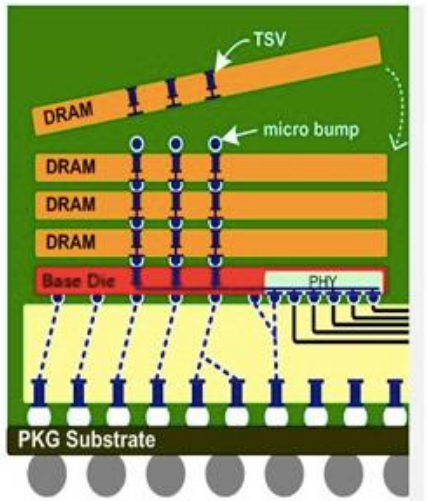
# Where Does That Leave Us

- No “new” technology that looks promising
  - Cheap
  - Reliable
  - Support density
- Can’t see much beyond 7 years
  - But where’s the capital investment?
- From a computer architect’s viewpoint
- Fast complex cores have been explored
- Multiple cores are being explore
  - Unlikely to reach large numbers except for data centers / HPC
- Application domain specific architectures look promising
  - Gpu led the way
  - Only so many applications that warrant specialization

# Some Opportunities

- Memory—new interesting technologies
  - 3D monolithic FLASH
  - Various NV-memory technologies
- Advanced packaging
  - 3D die stacking
  - Interposers
- Longer time horizon
- 3D monolithic ICs

  —see FLASH




| ITEM | TARGET |
|---|---|
| Burst Length | 2, 4 |
| Stack Density | 1GByte per stack (2Gbit per slice) |
| Channel / Slice | 2 |
| Banks / Channel | 8 |
| IO / Channel | 128 |
| Prefetch / Channel | 32B (128x2bit) |
| Channels / Stack | 8 |
| Total TSV Data IO Width | 1024 |
| Clock Speed | 500MHz |
| Peak Read BW / Stack | 128 GB/s |
| Page Size | 2KB |
| Data Parity | 1 bit / 32 bit |
| DRAM Core Voltage | 1.2V |
| Logic Buffer IO Voltage | 1.2V |

High Bandwidth Memory

# FIN