23. Future Directions

Jacob Abraham

Department of Electrical and Computer Engineering
The University of Texas at Austin

VLSI Design
Fall 2020

November 19, 2020
From (Physically) BIG Computers . . .

ENIAC

[Images of ENIAC computer system]
... to Multicore Chips
Integrated Circuits are Everywhere

Orbits of GPS satellites
(source: http://www.eumetsat.int/)

Internet map
(source: Wikipedia)
Myriad of Intelligent systems

- Cost, power consumption constraints
- In critical applications, resiliency is very important

Example: self-driving cars

- 100 Million lines of code for software, sensing and actuation
- 64 TOPS for cognition and control functions
Application requirements
Software, hardware (function) partitions
Select processor cores
  - ARM, MIPS, Tensilica, DSP
Memory and interfaces
  - DRAM, SRAM, Flash, Rambus, etc.
System interfaces
  - USB, PCI, PCMCIA, Ethernet, 802.11, Firewire, Bluetooth, etc.
Glue ASICs
Characteristics of Building Blocks

- VLSI design involves the implementation of complex functions using simple building blocks
  - Logic building blocks
  - Analog transfer characteristics
  - Composition

- We should be able to deal with other types of building blocks
  - For example, the theory developed for relay computers is applicable to CMOS transistors

Hierarchically compose building blocks

- Systems include hardware and software

Apple iPhone board
Test and Verification are Still Problems

State-space explosion

- Need to check a very large number of states to find a system-level test or to uncover a bug
- Even combinational equivalence checking NP-complete

Problem: the number of protons in the universe is around $10^{80}$, which is less than the number of states for a system with 300 storage elements!

Hubble photo of Coma Cluster: thousands of galaxies in a spherical shape 20 million light years across
Six different types of Boolean relationships between pairs of genes taken from the Affymetrix U133 Plus 2.0 human dataset
The two axes correspond to the expression levels of two genes
Source: Sahoo et al., Genome Biology 2008
Serial Connectivity Trends

Source: Xilinx
5G Expected Timeline

Source: cpqd.com
The World is not all Digital: System-on-Chip Market Size ($ Billions)

SONY Corp & Market Estimates

MS-SOC Contribution to the SoC Market Size

World Wide Semiconductor Market Size

SoC Market Size
**Laser Impact on the Brain**

- Monochromatic light in the near-infrared wavelengths
- Modulates brain function
- Produces neurotherapeutic effects in a non-destructive and non-thermal manner

**Mechanism of Low-Level Light Therapy (LLLT or 3LT)**

- Based on bioenergetics (fundamentally different from electric or magnetic stimulation)
- LLLT modulates the function of neurons
- Involves the absorption of photons by specific molecules in neurons
- Part of the mitochondrial respiratory enzyme cytochrome oxidase
Effect of Stimulation on EEG Power Spectral Density

Normalized PSD (dB/Hz)
Second Generation FinFET (Tri-Gate) Transistors

22 nm 1\textsuperscript{st} Generation Tri-gate Transistor

14 nm 2\textsuperscript{nd} Generation Tri-gate Transistor

Source: Intel
Air Gaps in Low-K Dielectric Materials

Intel 22nm Transistor

Source: Intel/Micron

Intel/Micron 25nm NAND Flash technology
Possible Future Transistors

(a) FinFET
(b) Nanowire
(c) Nanosheet

Source: IBM
Gate All Around (GAA) Transistors

(STI: Shallow Trench Isolation)

Source: androidauthority.com
Next Generation Manufacturing – Extreme Ultra-Violet Lithography

Wavelength reduction & larger NA enable the Litho roadmap: More than 100x Gain in Resolution

(NA: Numerical Aperture)

Source: Zeiss
TSMC plans on using FinFET transistors for its 3nm mode before switching to GAAFET (gate all around) for 2nm chips (A. Friedman, Oct. 2020)

Source (figure): M. Tyson, Hexus, June 13, 2019
Resistive RAM – RRAM (Memristors)

Atomic force microscope images of 17 HP Labs non-linear devices in a row, each a pair of oxide layers between the single bottom wire and one of the top wires.
Applications Which Will Require Bigger Computers

HPC Needs Decades of Moore’s Law

- 1 ZFlops
- 100 EFlops
- 10 EFlops
- 1 EFlops
- 100 PFlops
- 10 PFlops
- 1 PFlop
- 10 TFlops
- 1 TFlop
- 1 GFlop
- 10 GFlops
- 100 GFlops
- 100 MFlops


Weather Prediction
Genomics Research
Medical Imaging
Climate Modeling Requirements

Resolution: 200 km
Resolution: 1.5 km
Source: Wehner et al., IEEE Spectrum, October 2009

<table>
<thead>
<tr>
<th>Processor</th>
<th>Clock Speed</th>
<th>GFlops/Core</th>
<th>Cores</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMD Opteron</td>
<td>2.8 GHz</td>
<td>5.6</td>
<td>1,700,000</td>
<td>179MW</td>
</tr>
<tr>
<td>Tensilica LX2</td>
<td>500MHz</td>
<td>1</td>
<td>10,000,000</td>
<td>3MW</td>
</tr>
</tbody>
</table>
### Physical Dimensions: limited by 300 mm wafer

- TSMC 7 nm technology
- 1.2 Trillion transistors (redundancy to deal with defects)
- 400,000 AI compute cores
- As a comparison, Joule supercomputer has 84000 CPU cores, and consumes 450 KW of power
- Cerebras CS-1 uses 20 KW of power
Carbon nanotubes are around 1 nm in diameter
Much stronger than steel, flexible
Can possibly conduct $10^9 \text{ A/cm}^2$
Carbon Nanotube Transistor

Source: IBM
Self-assembly techniques (proposals include use of DNA)
Good success with liftoff techniques
PROBLEM 1: Metallic tubes

**Problem:** Some nanotubes cannot switch on and off, thus won’t work as semiconductors.

**Solution:** Switch off good CNTs and heat defective ones with electricity until they vaporize into CO₂.

Defective tube is always on.

Heat vaporizes defective tube.

Transistors open or close to switch tubes on or off.

---

PROBLEM 2: Misaligned tubes

**Problem:** Two of four components are shorted out by misaligned tube.

**Solution:** Algorithm determines how to overcome misalignments to prevent disruption.

SOURCE: Stanford Electrical Engineering/Computer Science, Max Shulaker
First Carbon Nanotube Computer


One instruction computer (SUBNEG – subtract and branch if negative (Turing complete))
- Few experimental demos
- Transistors ≠ system
Limited “tricks”

Complexity \(\rightarrow\) design bugs
Improving Computing Performance: New Innovations

Target: 1,000X performance

New innovations required

Source: S. Mitra et al., Stanford University
Computation immersed in memory

Increased functionality

Memory

Fine-grained, ultra-dense 3D

Computing logic

Impossible with today’s technologies

Source: S. Mitra et al., Stanford University
N3XT Computation Immersed in Memory

3D Resistive RAM
Massive storage

1D CNFET, 2D FET
Compute, RAM access

STT MRAM
Quick access

1D CNFET, 2D FET
Compute, RAM access

1D CNFET, 2D FET
Compute, Power, Clock

Source: S. Mitra et al., Stanford University
RISC-V Micprocessor (RV 16X-NANO) Built With Complementary Carbon Nanotube Transistors

(a) Fabricated chip
(b) 3-D schematic of layout; CNFETs are physically located in the middle of the stack, with metal routing both above and below

Source: G. Hills, Nature, 2019
Are Things Really Changing Compared to the Past?

Exponential Growth of Computing for 110 Years
Moore's Law was the Fifth, not the First, Paradigm to Bring
Exponential Growth in Computing

Source: Kurzweil, updated by Jurvetson
Look at the number of neurons (in the cerebral cortex for mammals) in different species.
Can’t We Make VLSI Chips Smarter?

Approximately 100 transistors to emulate a neuron

Possible number of neurons per processor

- 100 Core Processor
  - Human: 100,000,000,000
  - Elephant: 200,000,000,000

- Pentium: 31,000
  - Sea Slug: 7,000
  - Zebra Fish: 10,000
  - Ant: 10,000
  - Fruit Fly: 4,000,000
  - Cockroach: 1,000,000
  - Frog: 16,000,000
  - Dog: 160,000,000
  - Chimpanzee: 6,200,000,000

Source: Intel, Wikipedia