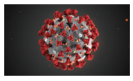
inst.eecs.berkeley.edu/~ee241b

# EE241B : Advanced Digital Circuits

## Lecture 17 – Power-Performance Tradeoffs

**Borivoje Nikolić**

March 16, hpcwire.com: Folding@home Turns Its Massive Crowdsourced Computer Network Against COVID-19





## Announcements

- Project midterm reports postponed until Tuesday, March 31
- Assignment 3 postponed until Thursday, April 2.
- Reading – req'd
  - Markovic et al, Methods for true energy-performance optimization, IEEE Journal of Solid-State Circuits, vol. 39, no.8, pp. 1281-1293, August 2004.
  - Chandrakasan and Brodersen, Low power CMOS digital design, IEEE Journal of Solid-State Circuits, vol. 27, no. 4, pp. 473-484, Apr. 1995.
- Recommended
  - Zyuban et al, Integrated Analysis of Power and Performance for Pipelined Microprocessors, IEEE Trans. on Computers, vol.53, no. 8, August 2004.



## Outline

- Module 4
  - SRAM alternatives
- Module 5
  - Low-power design
  - Power-performance tradeoffs

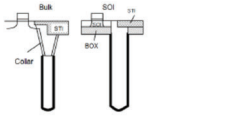


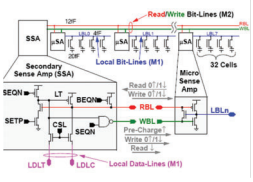
## 4.J SRAM Alternatives

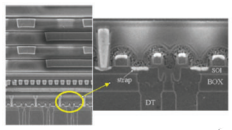


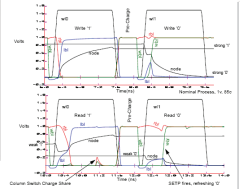
## eDRAM

- Process cost: Added trench capacitor



Barth, ISSCC'07, Wang, IEDM'06



## Crosspoint Memories

- Barrett, IRE Trans. Comp. 1961.

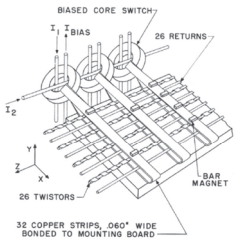
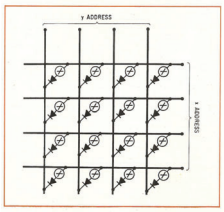



Fig. 2—Memory structure. $I_1$ and $I_2$ are access drive currents to core-selection switch. Presence or absence of a magnet over a twistor-strip solenoid crosspoint yields a "zero" or "one." Signals observed between twistor and return wire.



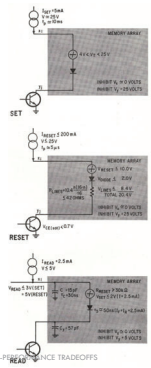
## Crosspoint Memories


- Neale, Nelson, Moore, Electronics'70
  - 16 x 16 array (256b) of 'read-mostly memory'



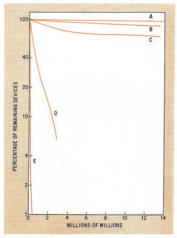



## Crosspoint Memory



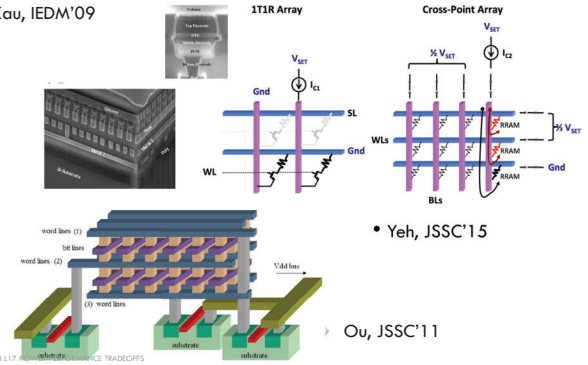

- Four modes
  - Form
  - Set
  - Reset
  - Read

**Endurance**



## 3D Crosspoint Arrays

- Kau, IEDM'09




- Yeh, JSSC'15

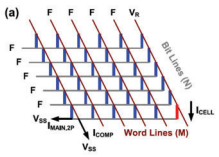


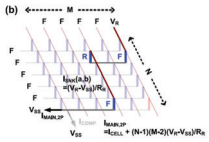

- Ou, JSSC'11



## Crosspoint Arrays

- Read and sneak currents



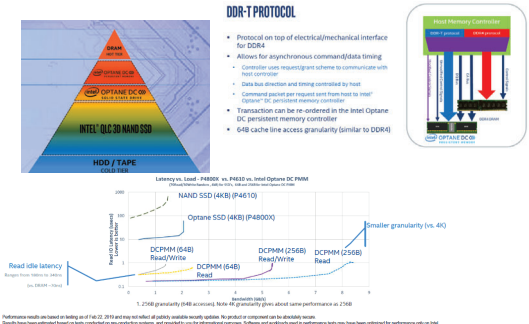




Bae, TED 4/17



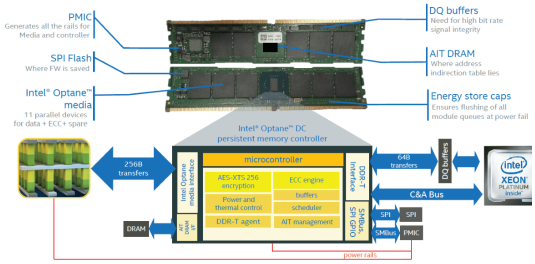
## In the News…

- Intel Optane = 3D XPoint






## Optane DDR






## 5. Low-Power Design



## Importance of Power Awareness

- Energy: Crucial for Portable Applications
  - Determines battery lifetime
  - Amount of computation that can be performed
  - Performance is what sells products
- Power: Crucial for High-Performance Applications
  - Determines cooling and energy costs
  - Most designs today are power limited
  - Still need maximum performance



## The Old Design Philosophy

- Maximum performance is primary goal
  - Minimum delay at circuit level
- Architecture implements the required function with target throughput, latency
- At circuit level, supplies, thresholds set to achieve maximum performance, subject to reliability constraints
- Performance achieved through optimum sizing, logic mapping, architectural transformations



## Constant Field Scaling Model

Traditional scaling model

$$\text{If } V_{DD} = 0.7, \text{ and } \text{Freq} = \left(\frac{1}{0.7}\right),$$
$$\text{Power} = CV_{DD}^2 f = \left(\frac{1}{0.7} \times 1.14^2\right) \times (0.7^2) \times \left(\frac{1}{0.7}\right) = 1.3$$

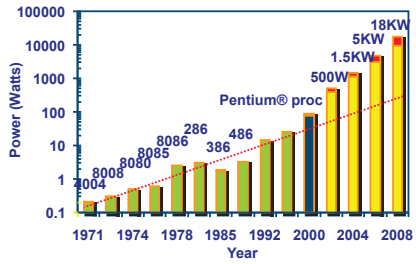
Maintaining the frequency scaling model of 1990s

$$\text{If } V_{DD} = 0.7, \text{ and } \text{Freq} = 2,$$
$$\text{Power} = CV_{DD}^2 f = \left(\frac{1}{0.7} \times 1.14^2\right) \times (0.7^2) \times (2) = 1.8$$

While slowing down voltage scaling

$$\text{If } V_{DD} = 0.85, \text{ and } \text{Freq} = 2,$$
$$\text{Power} = CV_{DD}^2 f = \left(\frac{1}{0.7} \times 1.14^2\right) \times (0.85^2) \times (2) = 2.7$$

## 2001 Picture: Power As a Problem




S. Borkar

**Power delivery and dissipation will be prohibitive**



## The New Design Philosophy

- Maximum performance is too power-hungry, and/or not even practically achievable
- Extract maximum performance under a power/energy envelope
- Excess performance (as offered by technology) to be used for energy/power reduction

**Trading off speed for power**



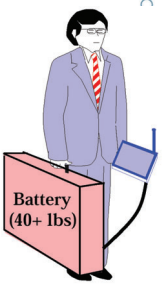
## 5.A Power and Energy Basics



## Portability: Battery Limits

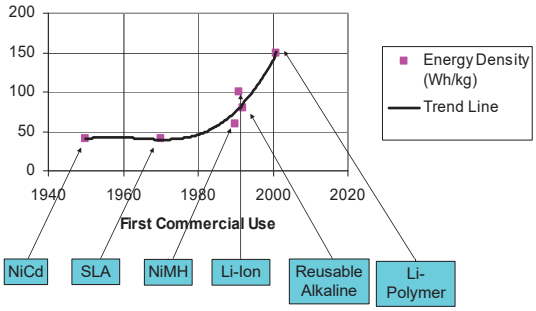
- Little change in basic technology
  - store energy using a chemical reaction
- Battery capacity doubles every 10 years
  - Has slowed down
- Energy density/size, safe handling are limiting factor

| *Energy density of material* | *KWH/kg* |
|---|---|
| Gasoline | 14 |
| Lead-Acid | 0.04 |
| Li polymer | 0.15 |

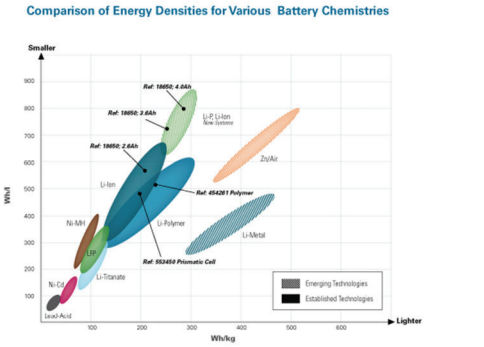





## Battery Progress



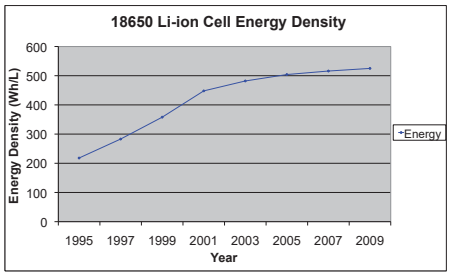



ICCNEXERGY



## Battery Technology Saturating




Battery capacity naturally plateaus as systems develop

[Courtesy: M. Doyle, Dupont]



## 5.B Power-Performance Tradeoffs

## Know Your Enemy

- Where does power go in CMOS?
- Switching (dynamic) power
  - Charging capacitors
- Leakage power
  - Transistors are imperfect switches
- Short-circuit power
  - Both pull-up and pull-down on during transition
- Static currents
  - Biasing currents



## Summary of Power Dissipation Sources

$$P \sim \alpha \cdot (C_L + C_{CS}) \cdot V_{swing} \cdot V_{DD} \cdot f + (I_{DC} + I_{Leak}) \cdot V_{DD}$$

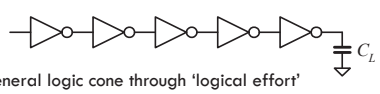
- $\alpha$ – switching activity
- $C_L$ – load capacitance
- $C_{CS}$ – short-circuit "capacitance"
- $V_{swing}$ – voltage swing
- $f$ – frequency
- $I_{DC}$ – static current
- $I_{leak}$ – leakage current

$$P = \frac{energy}{operation} \times rate + static\ power$$

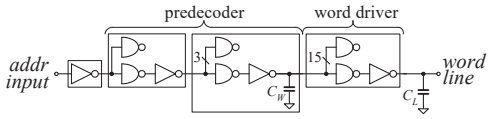


## CMOS Performance Optimization

- Reminder - sizing: Optimal performance with equal fanout per stage



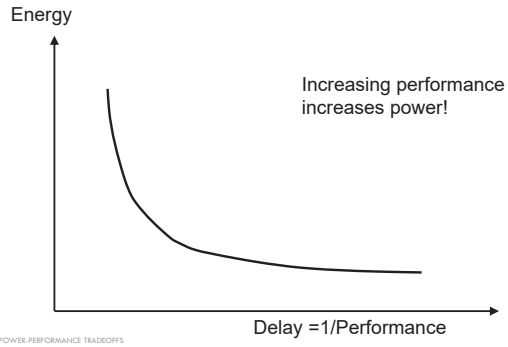

- Extendable to general logic cone through 'logical effort'
- Equal effective fanouts ($g_i C_{i+1}/C_i$) per stage
- Optimal fanout is around 4




[Ref: I. Sutherland, Morgan-Kaufman'98]

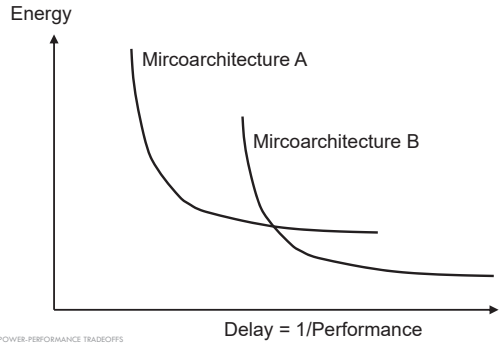


## Performance Optimization






## Performance Optimization






## Next Lecture

- Spring break
- Low-power design
  - Power-performance tradeoffs