Parallel Processing and Circuit Design with Nano-Electro-Mechanical Relays

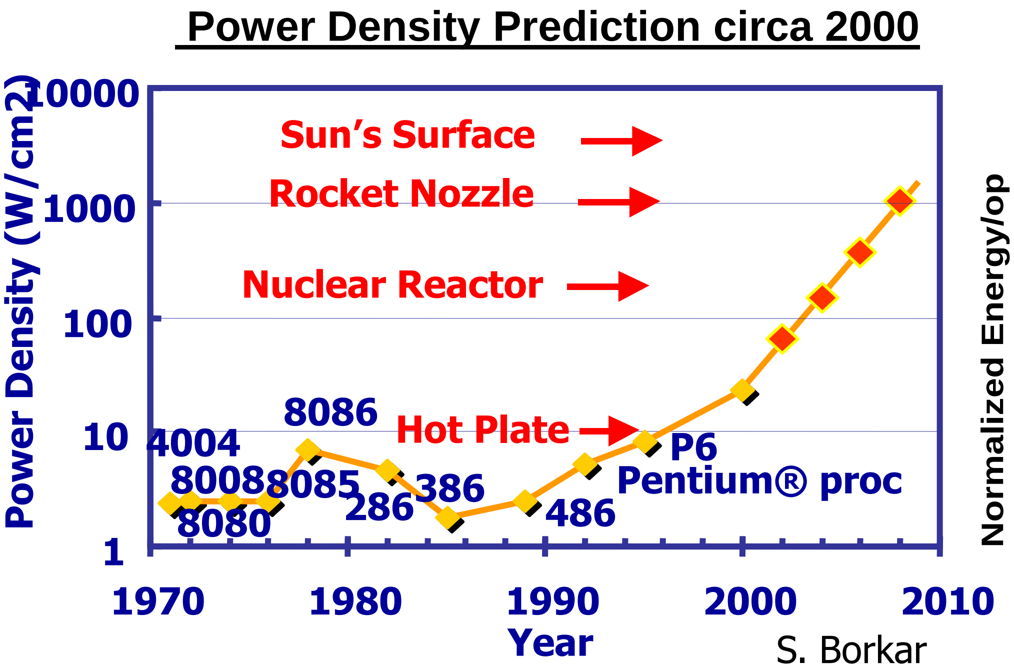
Elad Alon[1], Tsu-Jae King Liu[1], Vladimir Stojanovic[2], Dejan Markovic[3]

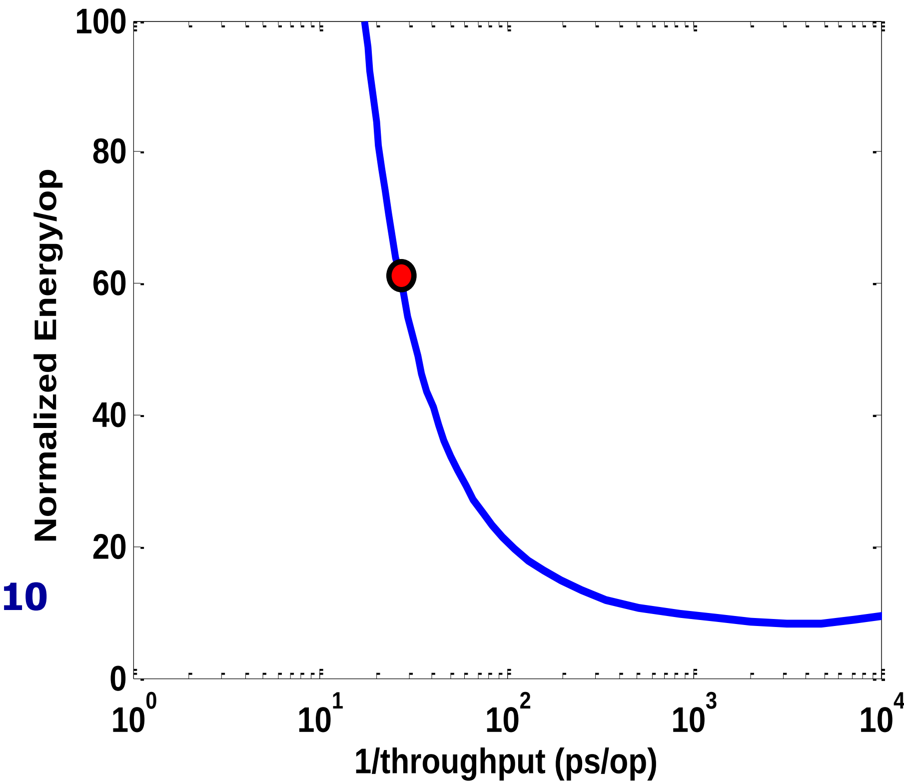
[1]University of California, Berkeley

[2]Massachusetts Institute of Technology

[3]University of California, Los Angeles

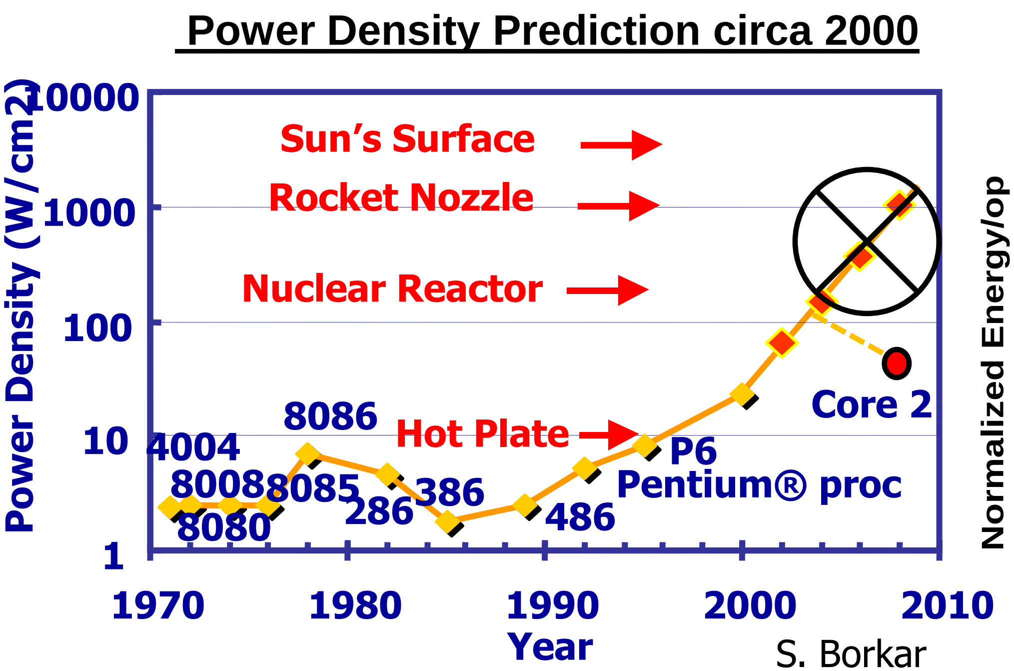
CMOS is Scaling, Power Density is Not

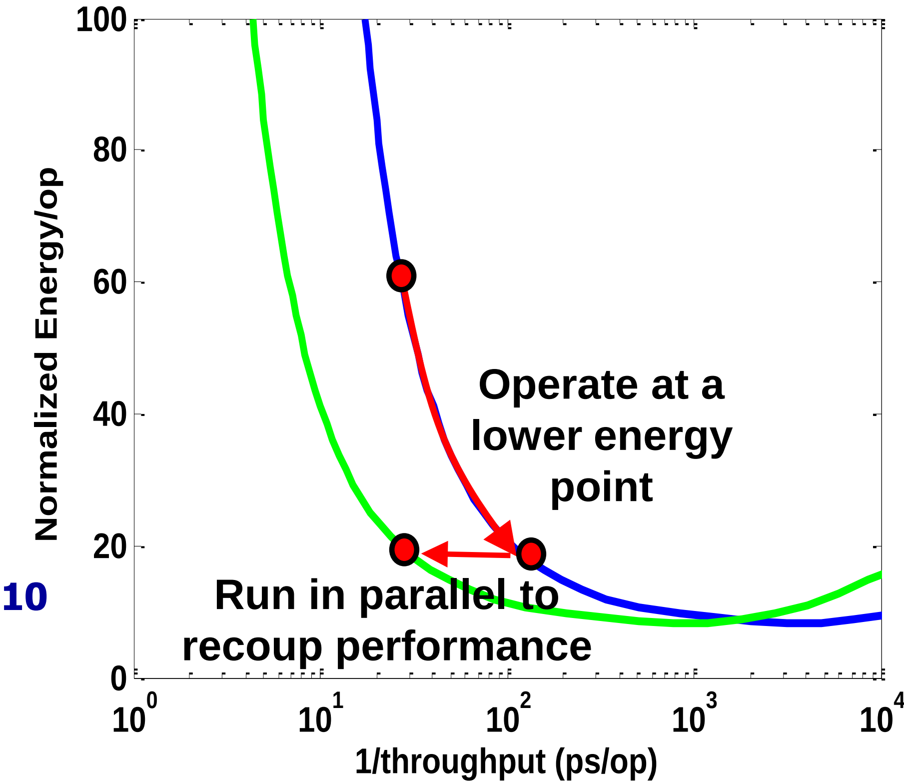




- V_{dd} and V_t not scaling well → power/area not scaling

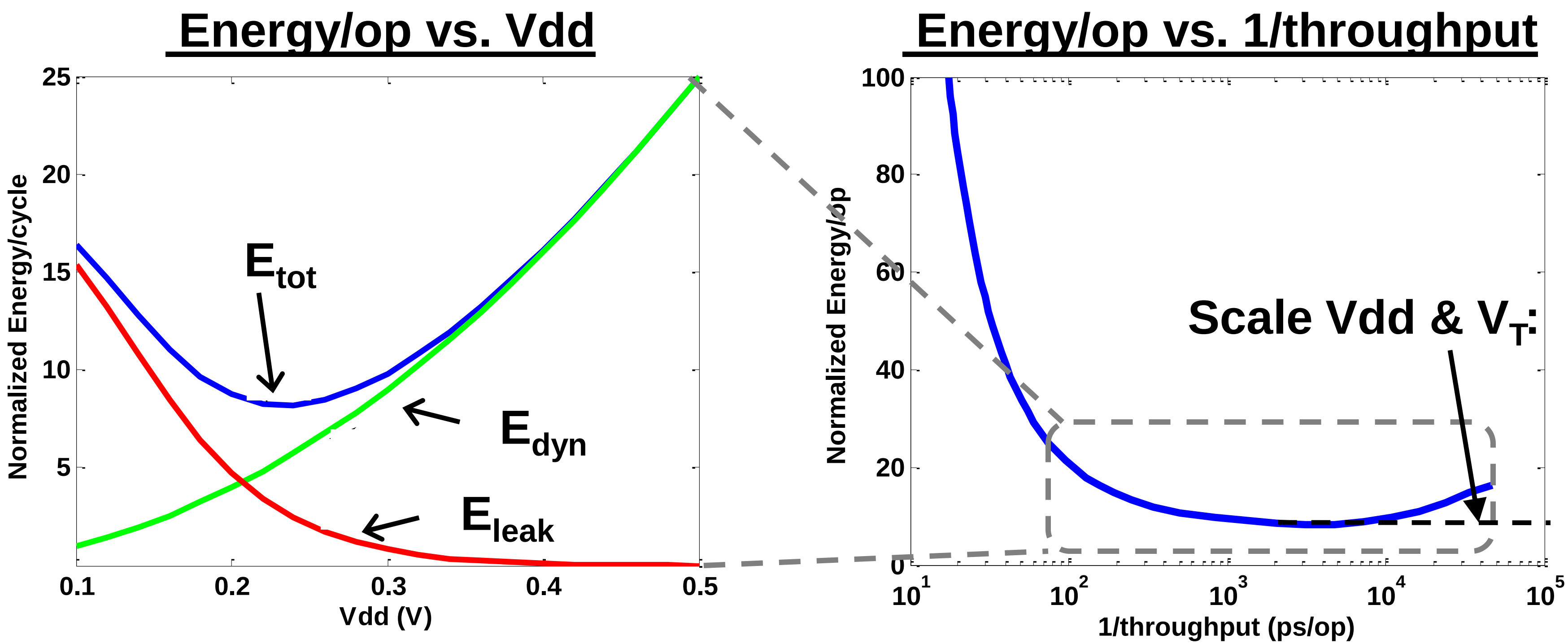
CMOS is Scaling, Power Density is Not





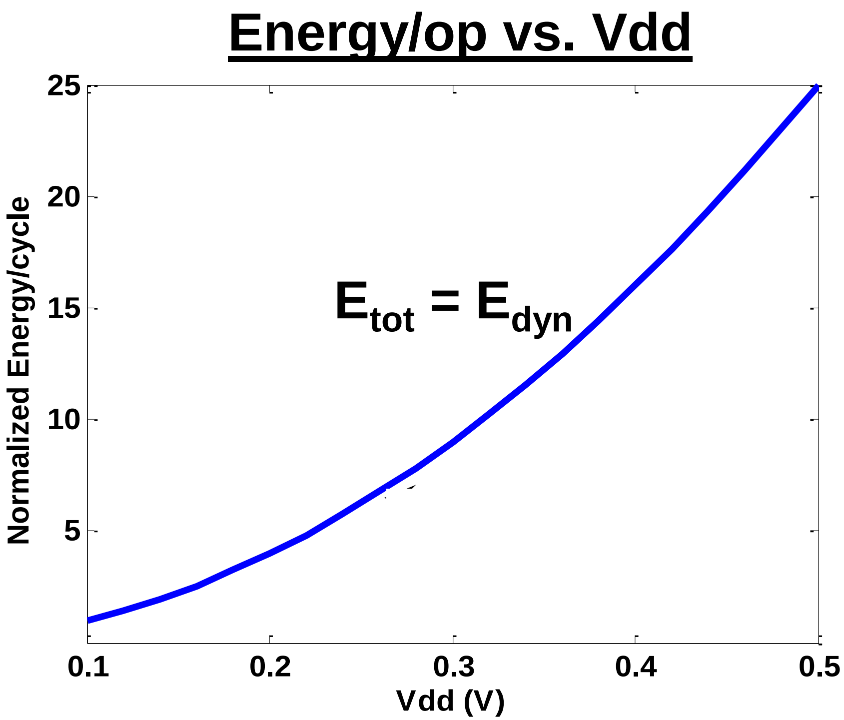
- V_{dd} and V_t not scaling well → power/area not scaling
- Parallelism to improve throughput within power budget

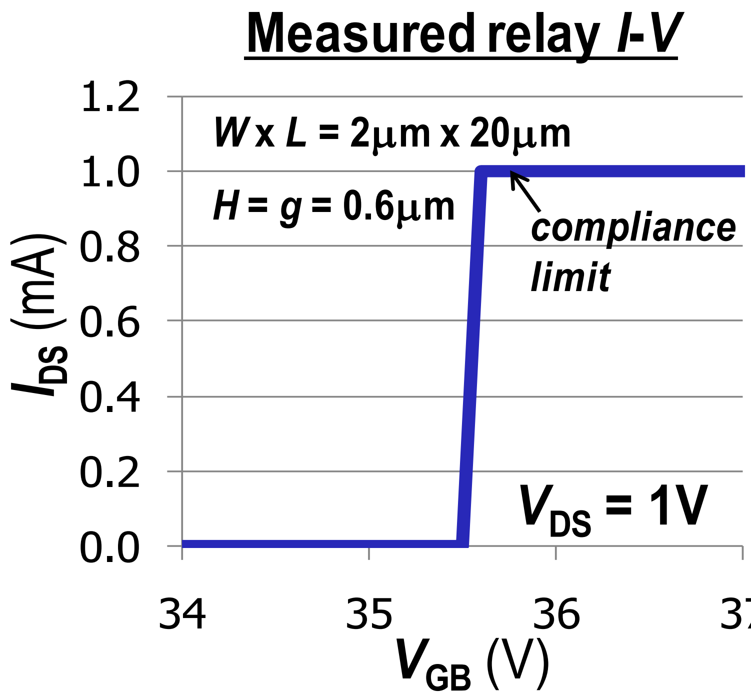
Where Parallelism Doesn't Help



- CMOS circuits have well-defined minimum energy
  - Caused by leakage and finite sub-threshold swing
  - Need to balance leakage and active energy
- Limits energy-efficiency, no matter how slow the circuit runs

What if There Was No Leakage?





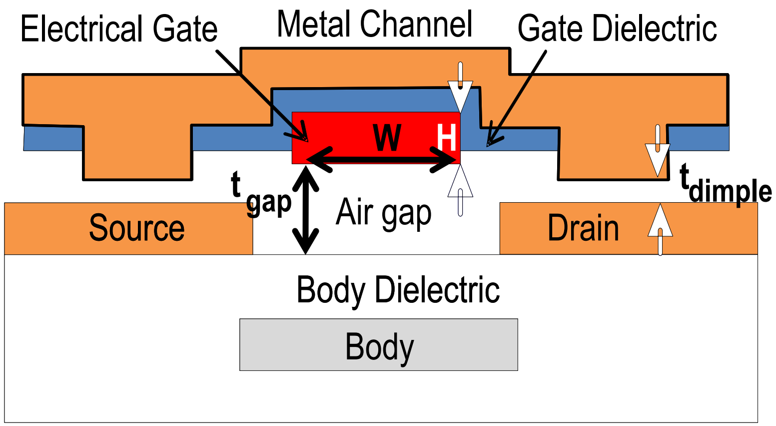
- No longer need to balance increasing component of energy as Vdd decreases
- Relay (mechanical switch) offers near infinite sub-threshold slope and no leakage current

Outline

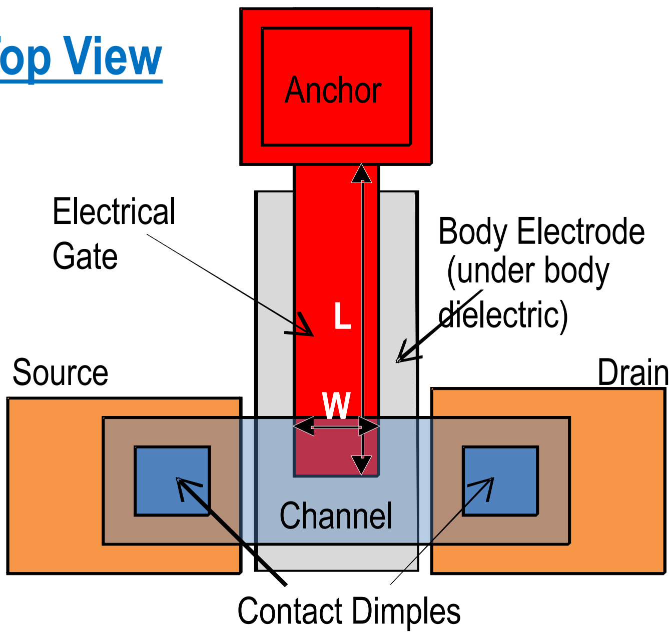
- If we could reliably build relays, would relay circuits offer superior energy-efficiency?
- NEM Relay Structure and Model
- Digital Circuit Design with NEM Relays
- Comparisons to CMOS

NEM Relay Structure & Operation

Cross-Section

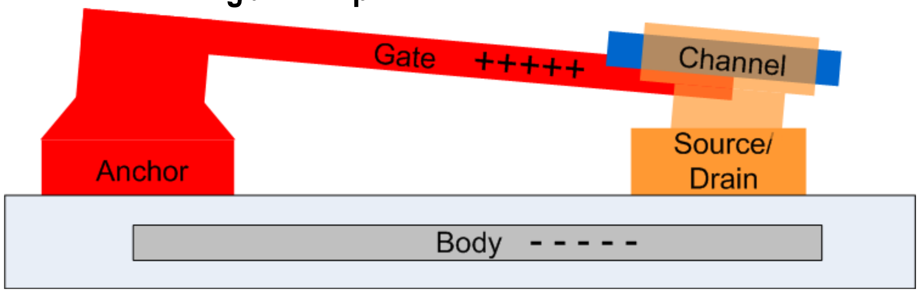


Top View



Closed Relay ("ON"):

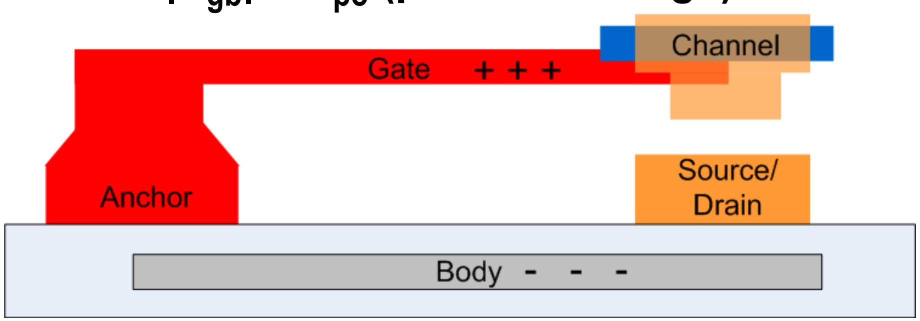
$|V_{gb}| > V_{pi}$ (pull-in voltage)



- On-resistance depends on materials, pressure ($|V_{gb}|$)

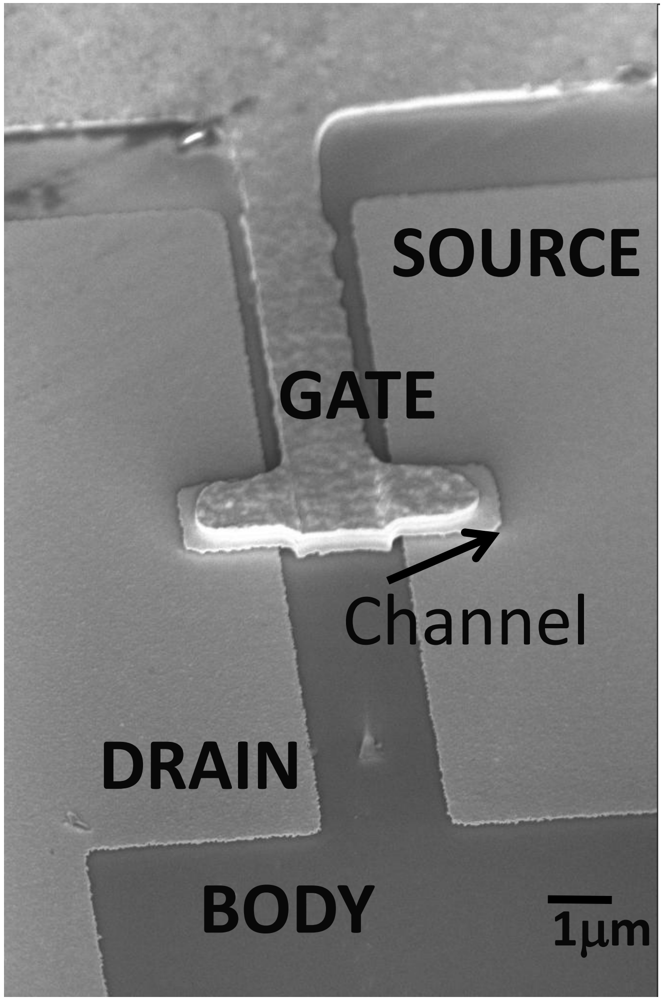
Open Relay ("OFF"):

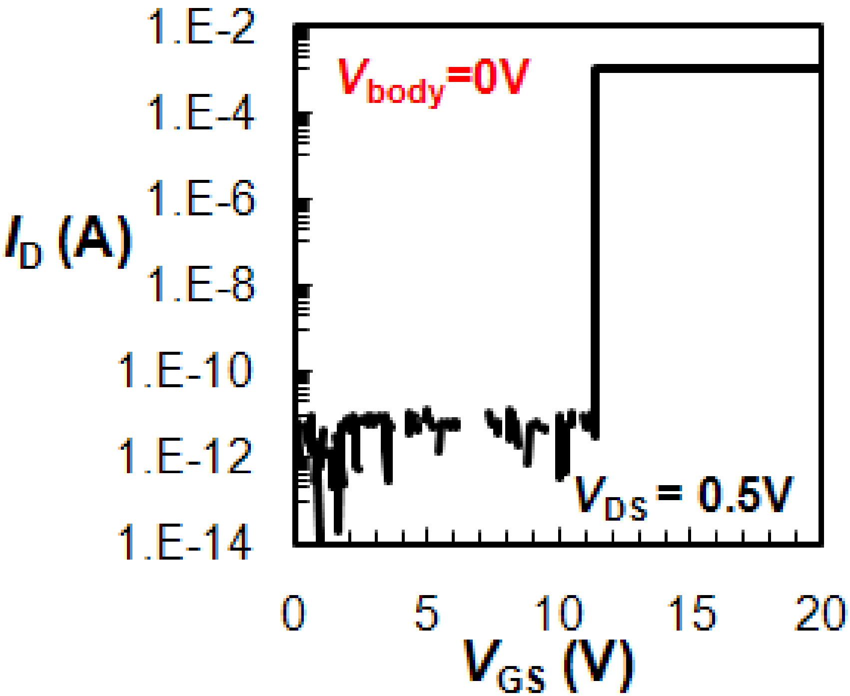
$|V_{gb}| < V_{po}$ (pull-out voltage)



- Infinite off-resistance → zero leakage

Preliminary Fabricated Relay



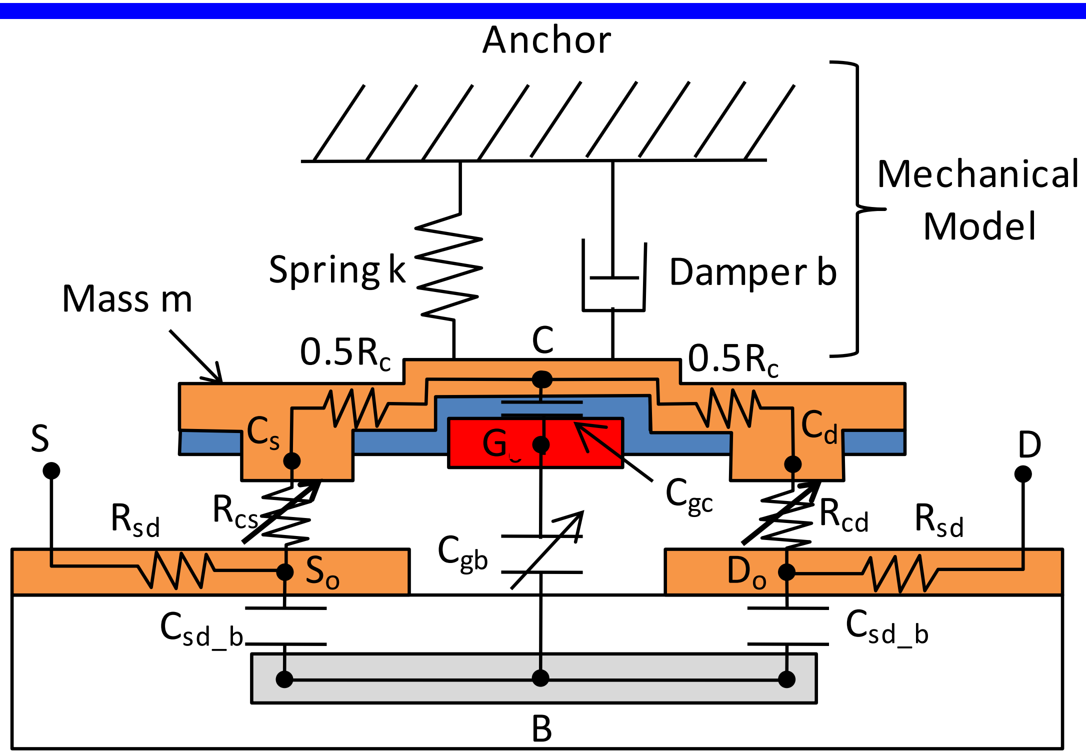


<u>Measured *I-V*</u>

Cantilever *W*,*L*,*H* = 2µm,20µm,200nm

Actuation gap thickness = 400nm

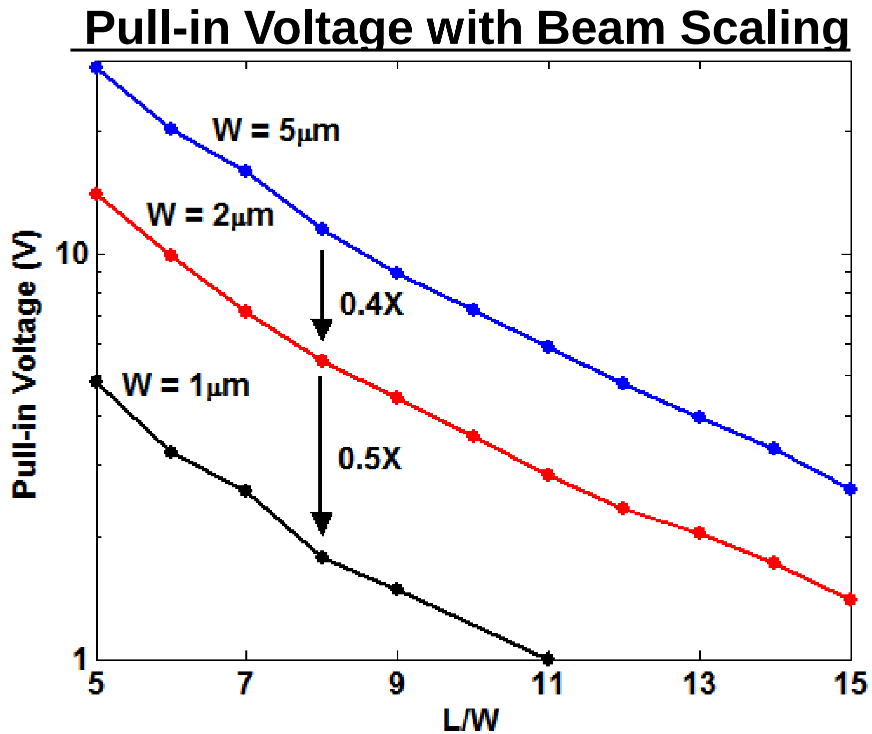
NEM Relay Model



- Compact Verilog-A model for circuit simulation:
  - Mechanical dynamics: spring (k), damper (b), mass (m)
  - Electrical parasitics: non-linear gate-body (C_{gb}), gate-channel (C_{gc}), and source/drain-body cap (C_{sd_b}), contact ($R_{cs,d}$)

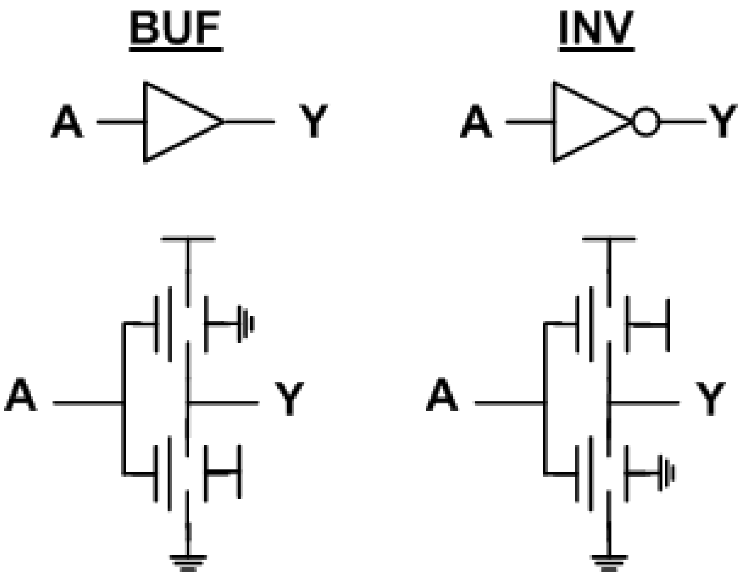
NEM Relay Scaling

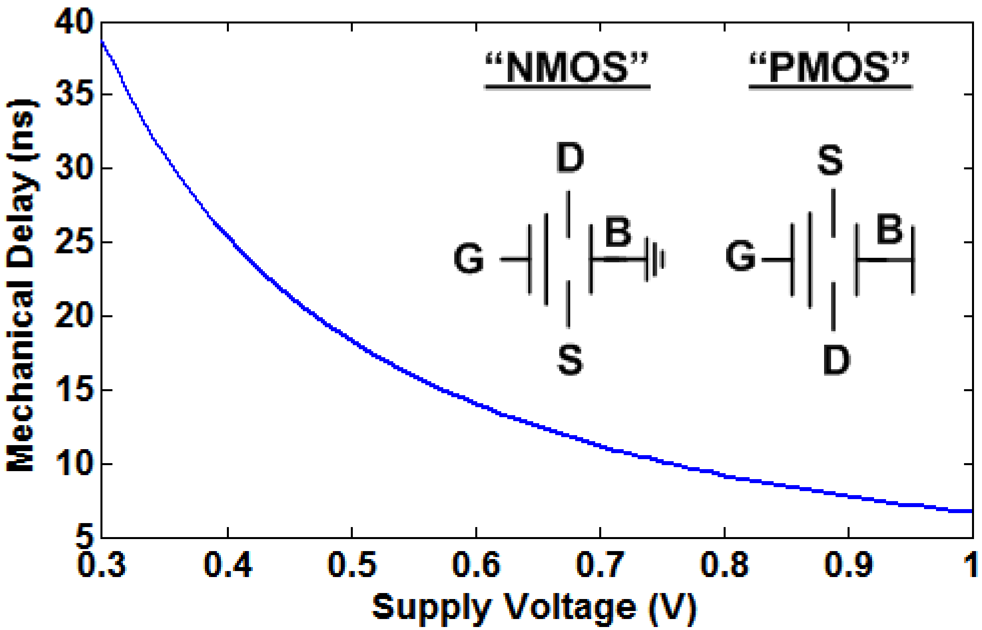
- Scaling (e.g., constant-E) benefits similar to CMOS



- Measured pull-in voltages scale linearly
  - If overcome reliability issues & surface forces scale:
    $\{W, L, \mathrm{H}, \mathrm{t}_{gap}\} = \{90\mathrm{nm}, 2.3\mathrm{um}, 90\mathrm{nm}, 10\mathrm{nm}\} \rightarrow V_{\mathrm{pi}} = 200\mathrm{mV}$
- Mechanical delay also scales linearly (~10ns @90nm)

NEM Relay as a Logic Element

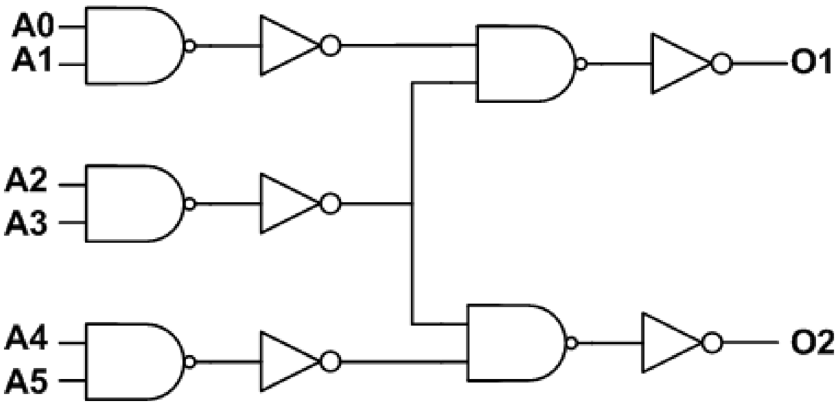




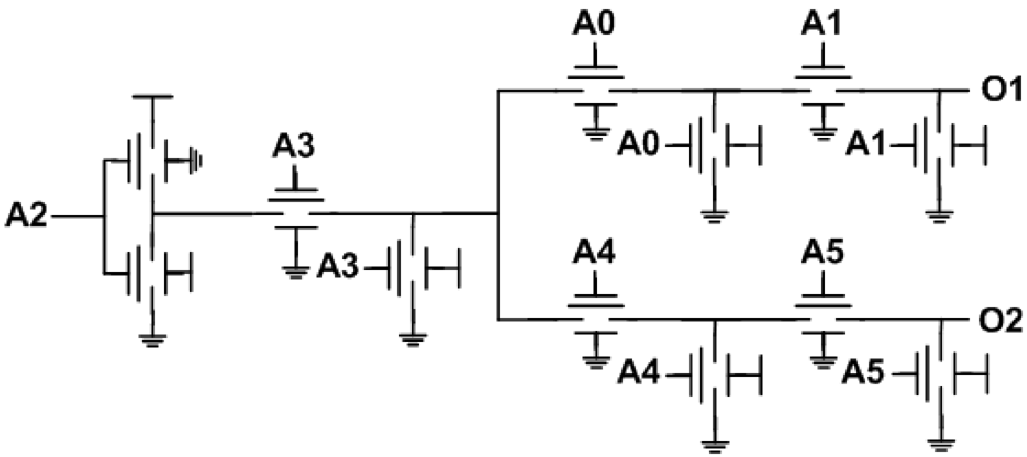
- 4-terminal design mimics MOSFET operation
  - Electrostatic actuation is ambipolar
  - Non-inverting logic possible: actuation independent of drain/source voltages
- Mechanical delay (~10ns) >> electrical τ (~1ps)
  - Can stack 200 devices (with 1kΩ on-resistance) before electrical delay is comparable to mechanical delay

Digital Circuit Design with Relays

CMOS: 30 transistors

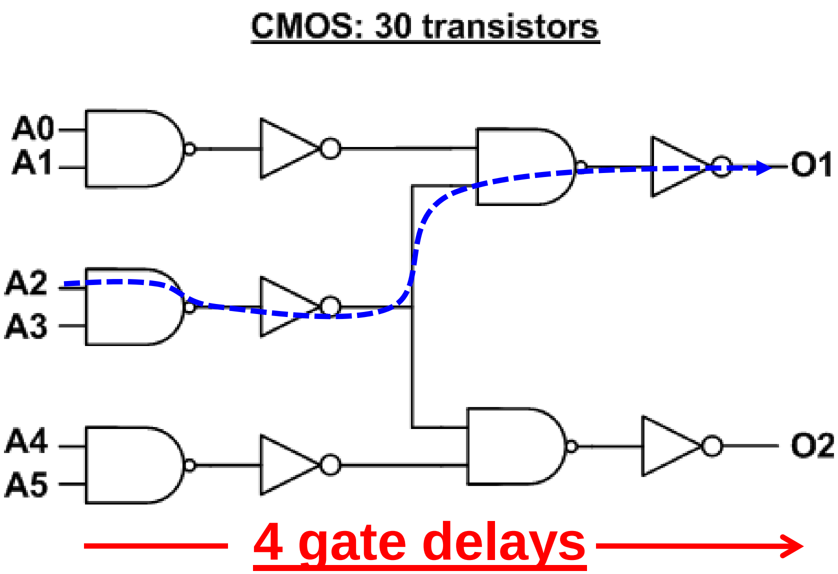


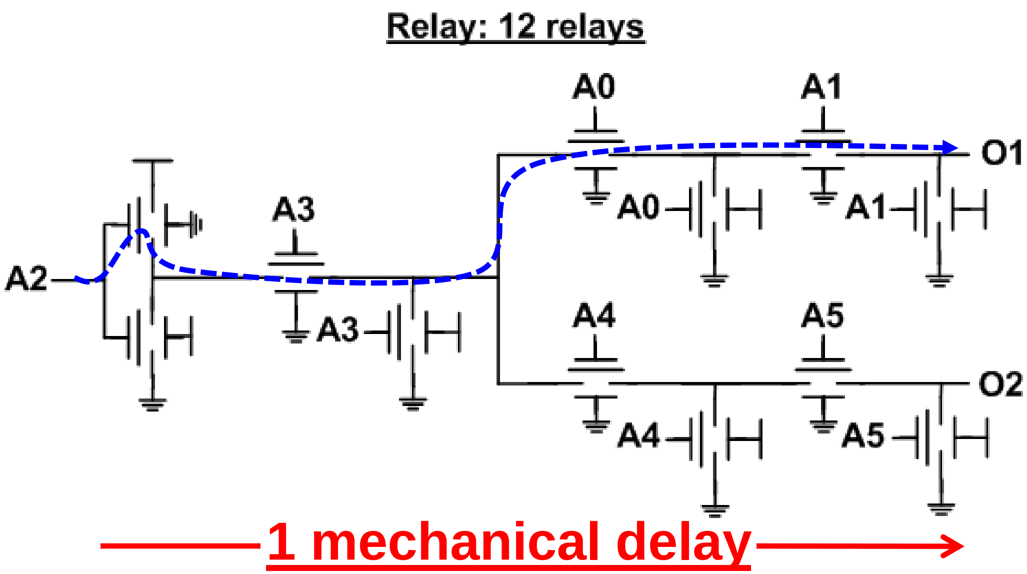
Relay: 12 relays



- CMOS: delay set by electrical time constant
  - Quadratic delay penalty for stacking devices
  - Buffer & distribute logical/electrical effort over many stages
- Relays: delay dominated by mechanical movement
  - Want all relays to switch simultaneously
  - Implement logic as a single complex gate

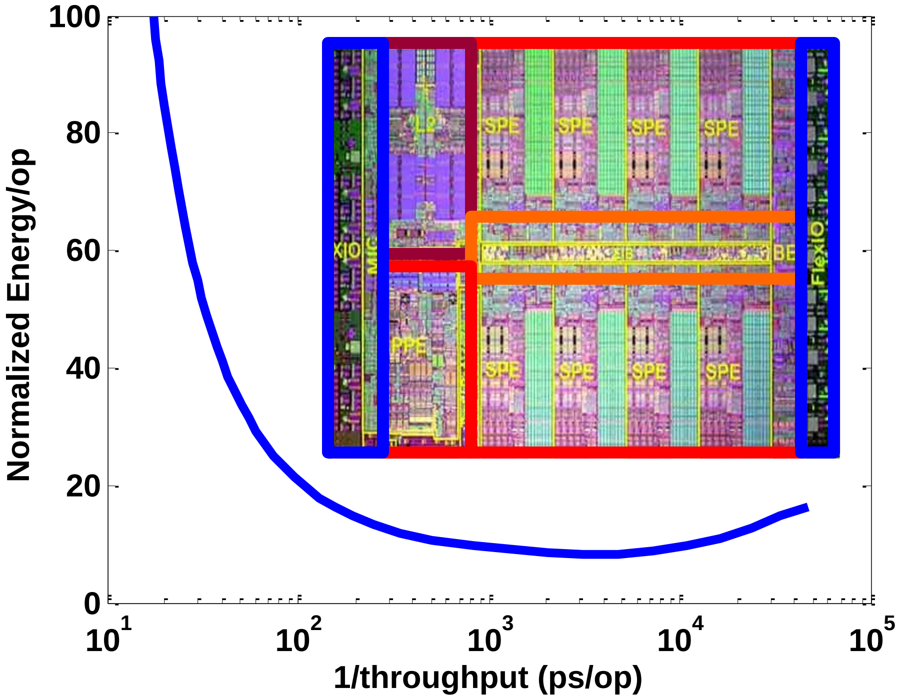
Digital Circuit Design with Relays

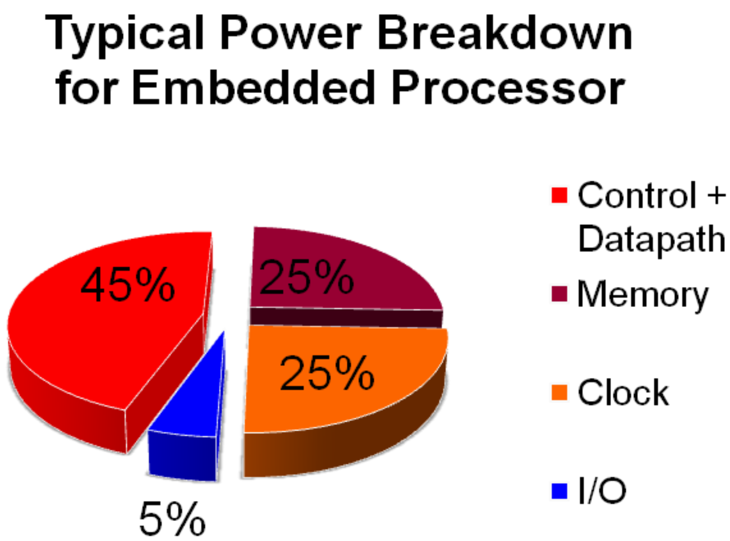




- **Delay Comparison vs. CMOS**
  - **Single mechanical delay vs. several electrical gate delays**
  - **For reasonable load, relay delay unaffected by fan-out/fan-in**
- **Area Comparison vs. CMOS**
  - **Larger individual devices**
  - **Fewer devices needed to implement the same logic function**

CMOS vs. Relays: Digital Logic

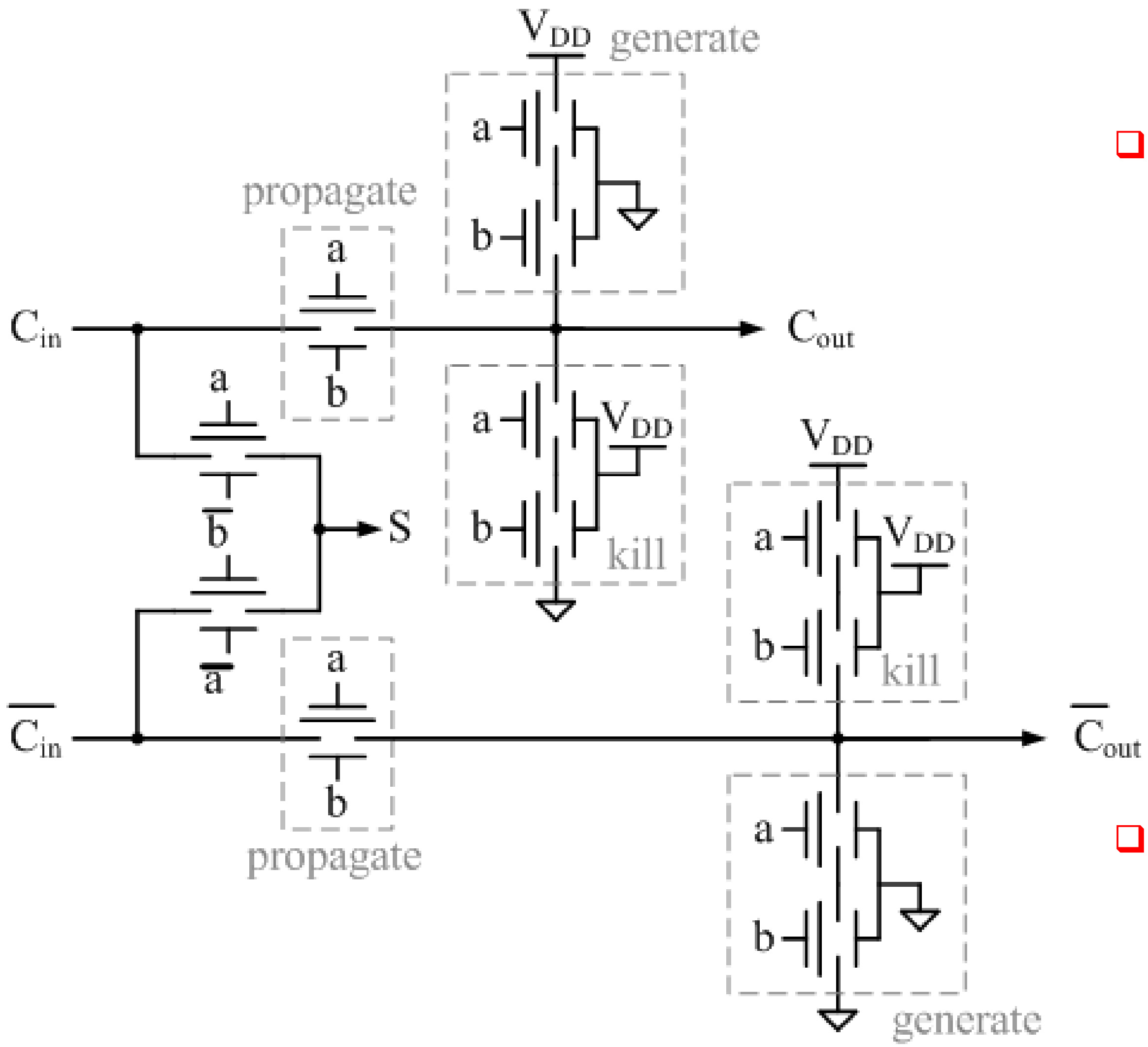




- Most processor components exhibit the energy vs. performance tradeoff of static CMOS
  - Control, Datapath, Clock

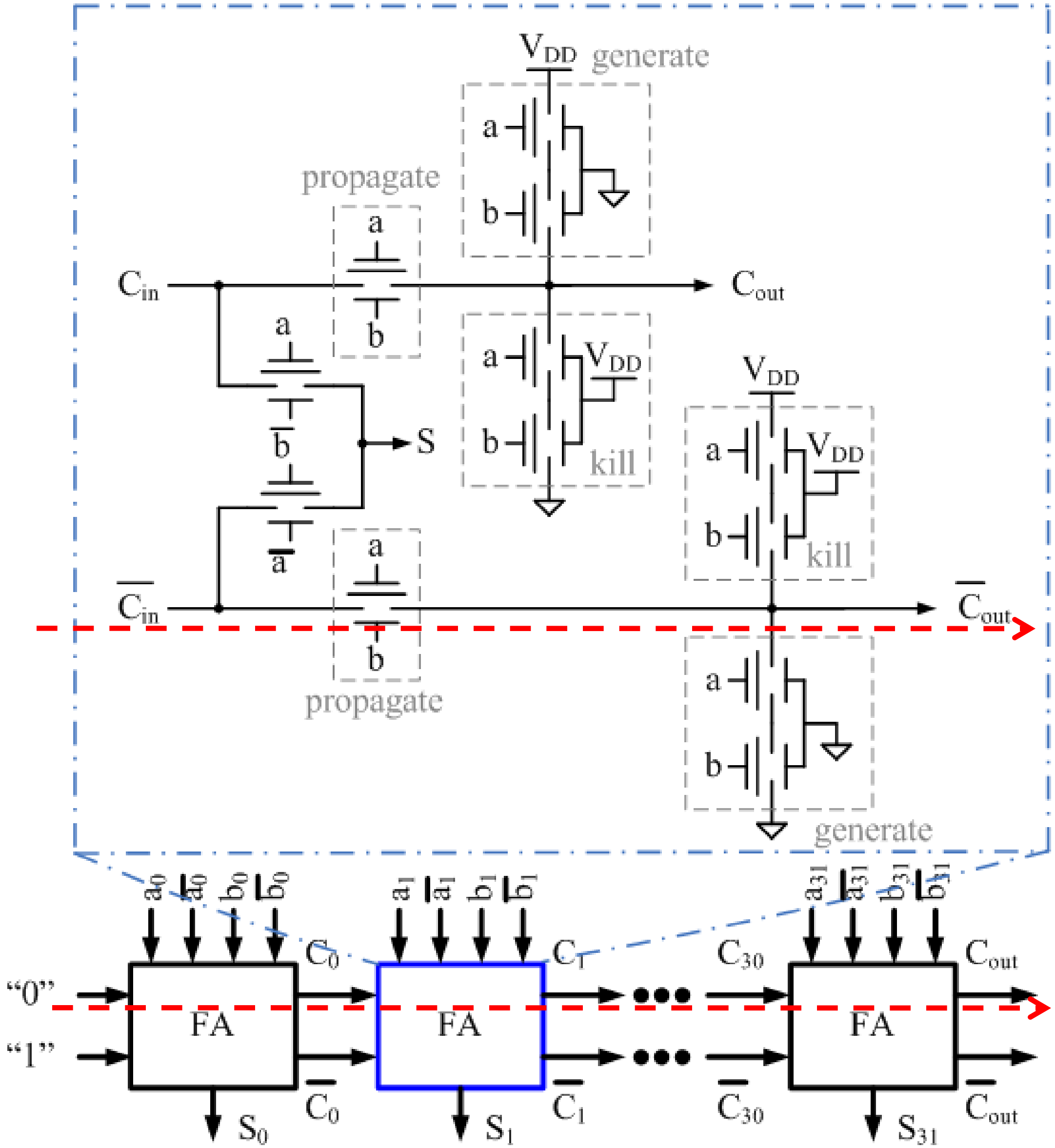
- Adder energy performance tradeoff is representative

Relay-based Adder



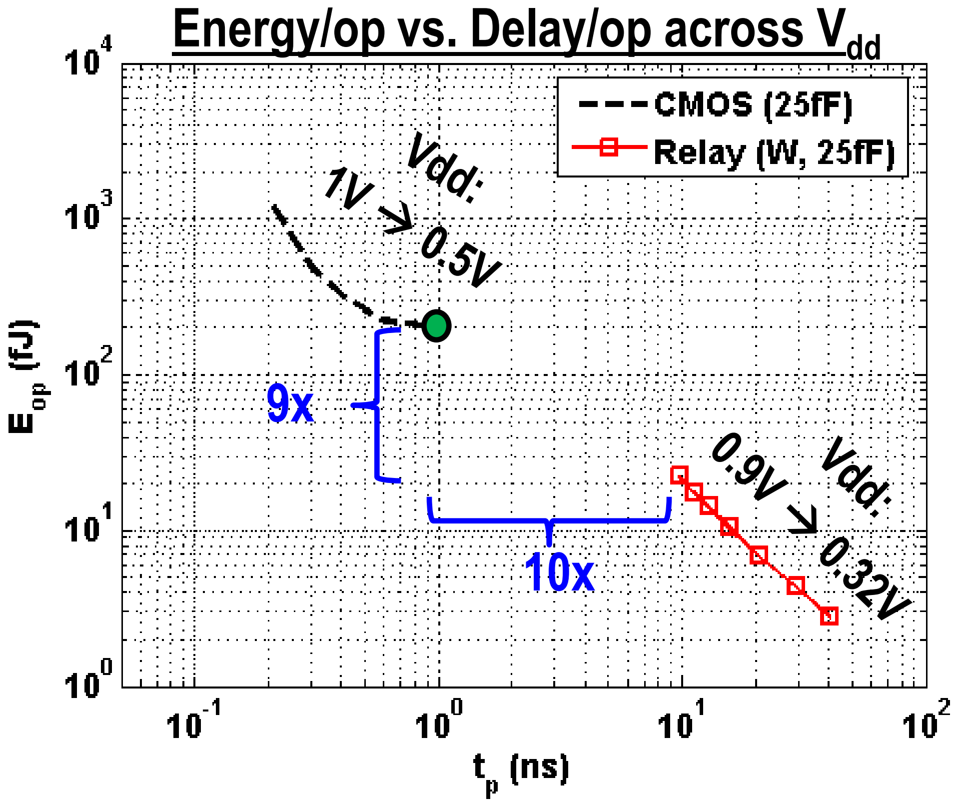
- Full adder cell:
  - 12 relays vs. 24 transistors
  - XOR "free"
  - Complementary signals avoid extra mechanical delay (to invert)
- Relays all sized minimally

N-bit Relay-Based Adder



- Ripple carry configuration
- Cascade full adder cells to create larger complex gate
- Stack of N relays, but still single mechanical delay

Snapshot: NEM Relays vs. CMOS



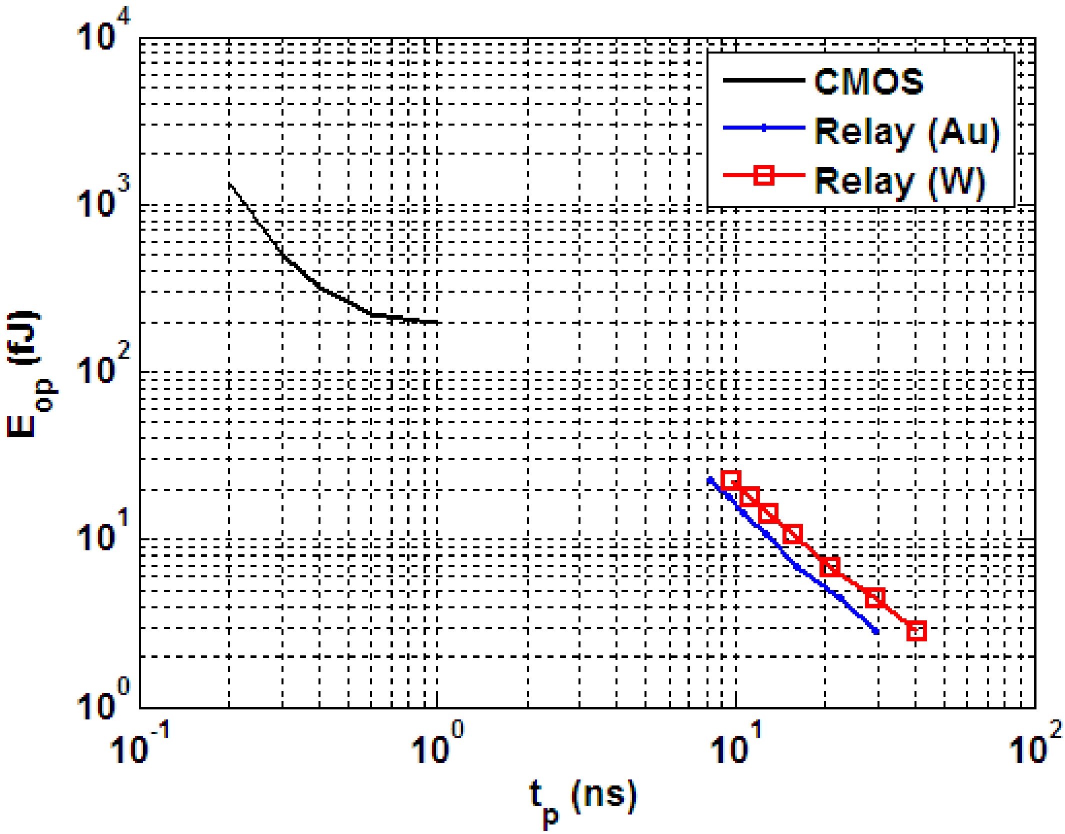
32-bit adder	CMOS	Relay
Supply Voltage	0.5 V	0.32 - 0.9 V
Load Cap per Output	25 fF	25 fF
Total Gate Cap	4.0 pF	125 fF
Area	600 μm²	480 μm²

Energy is for adder only

- Compare vs. Sklansky CMOS adder[1]
- For similar area: >9x lower E/op, >10x greater delay

[1]D. Patil et. al., "Robust Energy-Efficient Adder Topologies," in Proc. 18th IEEE Symp. on Computer Arithmetic (ARITH'07).

Energy-Delay Results: Contact R

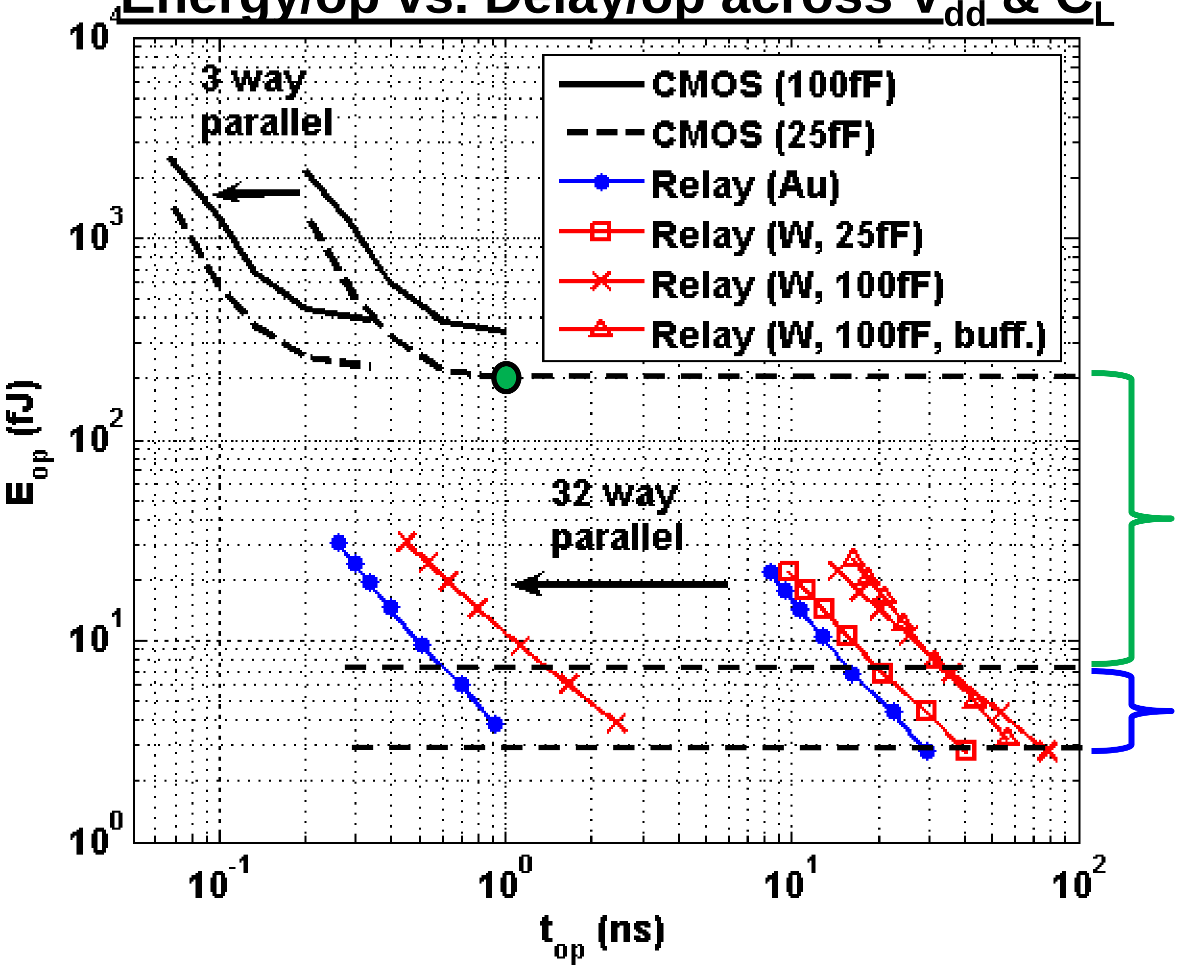


- Low contact R not critical
  - Enables low force, hard contact
- Good news for reliability…
  - More later

Energy-Delay Comparison

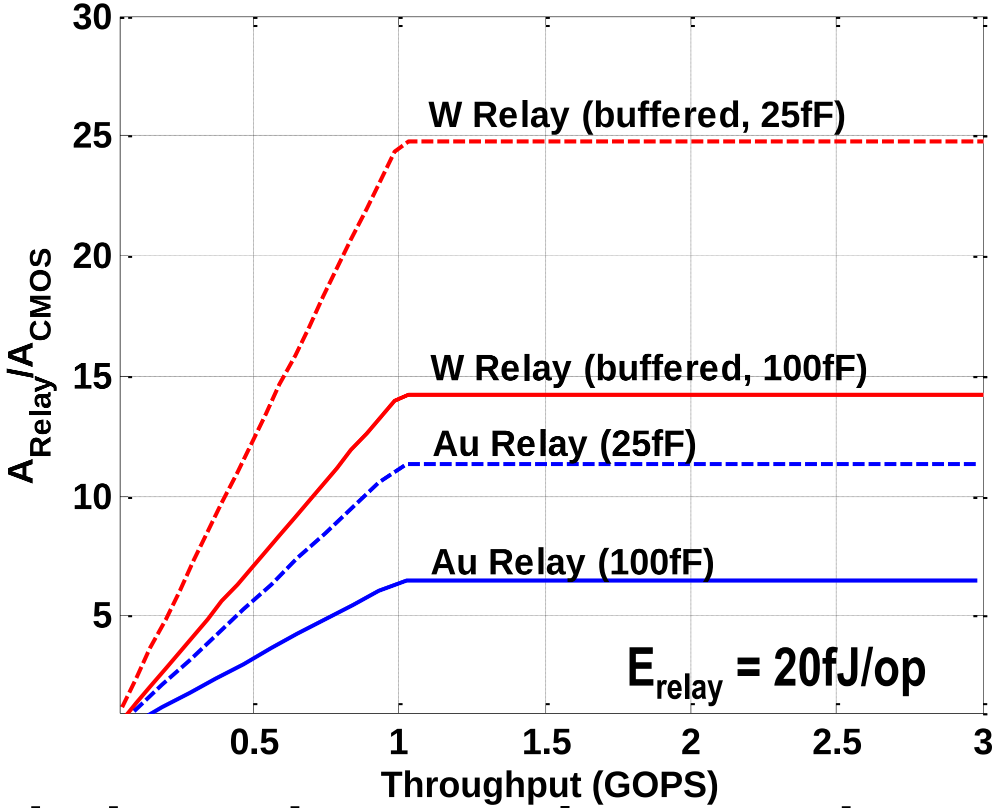
- Relays always more energy efficient
  - >10x better, even in the GOPS range

Energy/op vs. Delay/op across V_{dd} & C_L



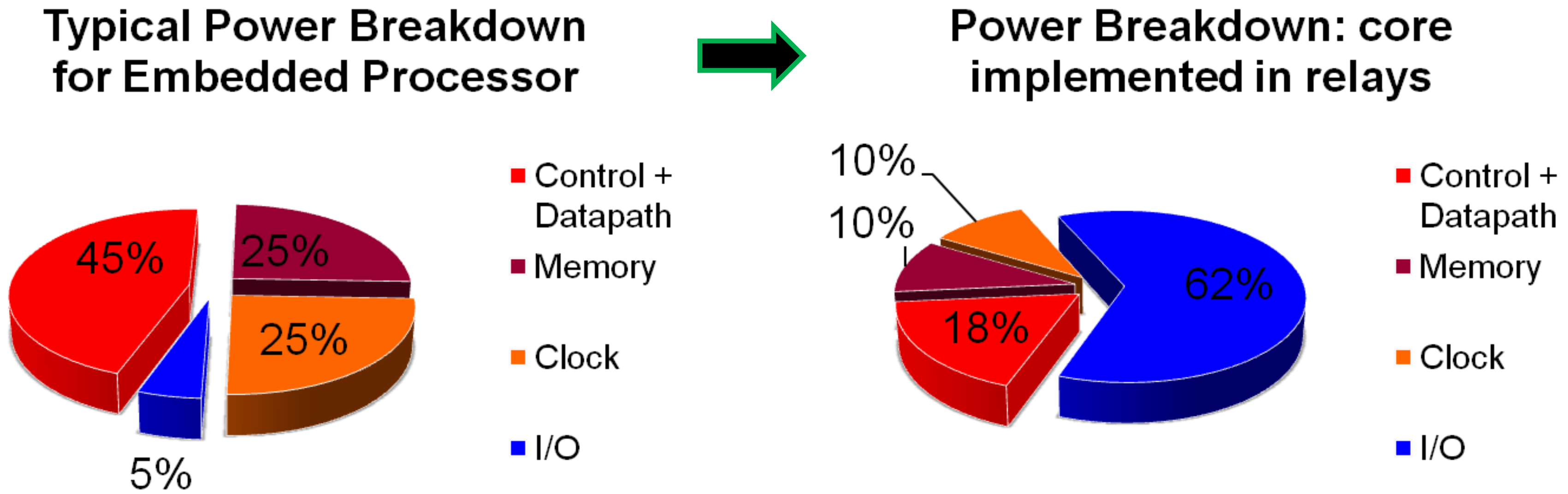
- 30x cap. gain
  - Lower device C_g, C_d
  - Fewer devices
- 2.4x Vdd gain
  - No leakage energy
  - Limited by E_{surf}

Area vs. Throughput



- For high throughputs, relays require area overhead to achieve similar throughputs as CMOS
- Area overhead is bounded (max. 6x-25x) :
  - CMOS needs parallelism to maintain E_{min} for throughputs over 1GOPS

Revised Power Breakdown



- I/O energy dominant if core is ~30x more efficient

- Need to explore processing of analog signals too…
  - No or limited "linear gain" in relays – rely on switching[1]

*F. Chen et. al., "Integrated Circuit Design with NEM Relays," *IEEE ICCAD*, Nov. 2008.

Conclusions

- NEM relays offer unique characteristics
  - Nearly ideal I_{on}/I_{off}
  - Significantly lower C_g than CMOS
  - Switching delay largely independent of electrical τ
- Relay circuits show potential for order of magnitude better energy efficiency
  - Examined 32-bit adders
  - Exploring memories, multipliers, complete microcontroller
- Key challenges are reliability and scaling
  - Circuit level insights critical: engineer contact R

Acknowledgements

- Hei Kam, Fred Chen, Hossein Fariborzi, Abhinav Gupta, Jaeseok Jeon, Rhesa Nathanael, Vincent Pott, Matthew Spencer, Cheng Wang

- Berkeley Wireless Research Center
- NSF
- DARPA
- FCRP (C2S2, MSD)
- Intel Foundation Graduate Fellowship