An Overview of Relay Integrated Circuits and Technology

Elad Alon (UC Berkeley)

in collaboration with

Tsu-Jae King Liu (UC Berkeley),
Dejan Marković (UCLA),
Vladimir Stojanović (MIT)
Since ~2000 supply voltages ($V_{DD}$) stuck at ~1V
- Leakage stops you from lowering threshold ($V_T$)

Leads to very poor power scaling... 1kW chips?
Parallelism allows slower, more efficient cores

- While maintaining overall throughput

Works well (if you can parallel program), but…
CMOS circuits have an absolute minimum energy

- Need to balance leakage and dynamic components

Parallelism doesn’t help if already at $E_{\text{min}}$…

Leakage: Game Over for CMOS
Relays To The Rescue?

- Mechanical relays don’t leak, turn on abruptly
  - Potential pathway to continued energy scaling

- Device/circuit co-design critical

Relay Structure and Operation

ON:

\[ |V_{gb}| > V_{pi} \text{ (pull-in)} \]

OFF:

\[ |V_{gb}| < V_{po} \text{ (pull-out)} \]

Poly-SiGe

Tungsten
NEM Relay as a Logic Element

- Mimics operation of CMOS transistors
  - Electrostatic actuation is ambipolar
- Unlike CMOS, non-inverting logic is possible
  - Switch state set only by gate-to-body voltage
Digital Circuit Design with NEM Relays

- **CMOS**: delay set by electrical time constant
  - Cascade simple gates to distribute fanout

- **Relays**: delay dominated by mechanical movement
  - So, want all to switch simultaneously
  - Implement logic as a single complex gate (1930’s)
Need to Compare at Block Level

- Single mechanical delay per block
  - Substantially mitigates perceived delay disadvantage

- Often fewer devices for same function
  - Comparable area despite larger individual devices

Example: 32-bit Relay Adder

- **Ripple carry configuration**
- **Cascade full adder cells to create larger complex gate**
- **Stack of 32 relays, still a single mechanical delay**
Scaled Relay vs. CMOS Adders

- **Compare vs. CMOS adder** in 90nm technology
- **For similar area:**
  - >9x lower E/op
  - >10x greater delay

* D. Patil *et al.*, “Robust Energy-Efficient Adder Topologies,” *ARITH 2007*
Parallelism

- Can extend energy benefit up to GOPS throughput
- As long as parallelism is available
Low contact R not critical

Good news for reliability…
Higher contact R, hard contact (W) improves reliability
- Limits power dissipation, material flow

Current endurance record: 65 billion cycles
- Theory/experiments predict $>10^{15}$ cycles @ 1V VDD

H. Kam et al., IEDM 2010
Relay Energy Limit

- Spring force must be able to overcome surface adhesion force $F_A$:
  - $E_{\text{min}} = 2F_A g_d \approx 2E_{\text{surf}}$

- For large contacts, $F_A$ scales with area

- Extracted surface adhesion energy $\sim 5 \mu\text{J/m}^2$
  - 90nm x 90nm: $E_{\text{surf}} = 0.04 \text{aJ}$!

- Ultimate energy limit set by # of contact bonds
  - $\sim 0.2 \text{aJ per bond}$

H. Kam et al., IEDM 2009
Conclusions

- Relay characteristics enable energy scaling beyond CMOS
  - Nearly ideal $I_{on}/I_{off}$
  - Need to adapt circuit design style

- Reliability improving
  - Circuit level insights critical (contact R)

- Potential for 10X or more lower E/op than CMOS
  - How to build complex relay-based systems?
  - Will this really work in practice?