Nanomechanical Switches for Ultra-Low-Power Computation and Memory

Tsu-Jae King Liu

Department of Electrical Engineering and Computer Sciences
University of California, Berkeley, CA 94720-1770 USA

June 22, 2014
A Vision of the Future

Emergence of Ambient Intelligence:
Sense/monitor, communicate and react to the environment

→ Ultra-low-power, robust computing technology required!
Outline

• Introduction
  – CMOS energy efficiency
  – Why mechanical switches?

• Nanomechanical switches for computing

• Nanomechanical memory

• Summary
Improving CMOS Energy Efficiency

- Parallelism is the main technique to improve system performance under a power density constraint.

![CMOS Energy vs. Delay Graph](image-url)

- Operate at a lower energy point.
- Run in parallel to recoup performance.
CMOS Energy Efficient Limit

- Leakage & sub-threshold slope define minimum Energy/op

**Graph:**
- **Dynamic Energy:** $E_{\text{total}} = \alpha L_d f CV_{DD}^2 + L_d f I_{\text{OFF}} V_{DD} t_{\text{delay}}$
- **Leakage Energy:**

**Diagram:**
- $t_{\text{delay}} = \frac{L_d f CV_{DD}}{2I_{\text{ON}}}$
- $\alpha$: Activity Factor
- $L_d$: Logic Depth
- $f$: Fanout
- $C$: Capacitance per Stage
Future Logic Switch Requirements

- Zero leakage → Zero standby power
- Simple design → Low cost
- Robust → Wide applicability
- Reconfigurable → Versatility
- Low operating voltage → Low active power
Micro-Electro-Mechanical Switch

OFF State (as fabricated):  NORMALLY OFF

ON State:

- Zero OFF-state current ($I_{OFF}$); abrupt switching
  - Turns on by electrostatic force when $|V_{GS}| \geq V_{PI}$
  - Turns off by spring restoring force when $|V_{GS}| \leq V_{RL}$
Surface Micromachining Process

Cross-sectional View

- Mechanical structures can be made using conventional microfabrication techniques
- Structures are freed by selective removal of sacrificial layer(s)
The semiconductor industry has developed air-gap back-end-of-line (BEOL) processes which can be leveraged to make MEM switches.
Outline

• Introduction

• Nanomechanical switches for computing
  – Logic switch designs
  – Relay-based integrated circuit design
  – Relay scaling limit

• Nanomechanical memory

• Summary
First Micro-Relay


- 4-terminal (4-T) micro-electro-mechanical switch demonstrated
  0.35 \( \mu \text{m} \) thick \( \times \) 76 \( \mu \text{m} \) long SiO\(_2\) membrane; 7 \( \mu \text{m} \) actuation gap
  60 V switching voltage; 40 \( \mu \text{s} \) switching delay

Structure:

Device Operation:
Scaled 4-T Electro-Mechanical Relay

R. Nathanael et al., IEDM 2009

Isometric View

- Voltage applied between the gate and movable body brings the channel into contact with source & drain
  - Oxide layer insulates body

AA’ Cross-section

- Body Dielectric (Al₂O₃)
- Channel (W)
- BODY (p+ poly-Si₀.₄Ge₀.₆)
- Insulator (Al₂O₃)
- SOURCE (W)
- GATE (W)
- DRAIN (W)
4-T MEM Relay Process Flow

R. Nathanael et al., IEDM 2009

- Deposit Al₂O₃ substrate insulator
  - ALD at 300°C

- Deposit & pattern W electrodes
  - DC magnetron sputtering

- Deposit 1ˢᵗ sacrificial LTO
  - LPCVD at 400°C

- Define contact regions

- Deposit 2ⁿᵈ sacrificial LTO
  - \( t_{\text{dimple}} = \frac{t_{\text{gap}}}{2} \)

- Deposit & pattern W channel

- Deposit Al₂O₃ gate oxide
Process Flow (cont’d)

R. Nathanael et al., IEDM 2009

- Deposit p+ poly-Si\textsubscript{0.4}Ge\textsubscript{0.6} gate
  - LPCVD at 410°C

- Pattern gate & gate oxide layers using LTO as a hard mask

- Release in HF vapor

- Coat with ultra-thin (~0.3nm) TiO\textsubscript{2}
  - ALD at 300°C
\[ I_D-V_G \] Characteristics

R. Nathanael et al., IEDM 2009

- NMOS- or PMOS-like operation is achieved by applying an appropriate body bias voltage:

- Zero off-state leakage; abrupt switching (SS < 0.1 mV/dec)
Contact Endurance

Y. Chen et al., to be published

- Endurance increases exponentially with decreasing $V_{DD}$
  - projected to reach $10^{15}$ cycles at 1 Volt, for W contacts
Switching Delay

R. Nathanael et al., IEDM 2009

• Turn-on delay improves with gate overdrive, saturates at ~200 ns for $V_B = 0V$.

• Turn-on delay improves w/ body biasing to reduce $V_{pl}$ → 100 ns turn-ON delay
Relay Scaling

- Scaling has similar benefits for relays as for MOSFETs:

$$k_{\text{eff}} \propto \frac{EWt^3}{l^3}$$

### Pull-in Voltage

$$V_{PI} \propto \sqrt{\frac{k_{\text{eff}} g^3_{\text{actuation}}}{\varepsilon_0 A}}$$

### Pull-in Delay

$$t_{PI} \propto \sqrt{\frac{mg_{\text{contact}}}{k_{\text{eff}} g_{\text{actuation}}}} \left( \frac{V_{PI}}{V_{DD}} \right)$$

<table>
<thead>
<tr>
<th>Relay Parameter</th>
<th>Scaling Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spring constant, $k_{\text{eff}}$</td>
<td>$1 / \kappa$</td>
</tr>
<tr>
<td>Mass</td>
<td>$1 / \kappa^3$</td>
</tr>
<tr>
<td>Pull-in voltage, $V_{PI}$</td>
<td>$1 / \kappa$</td>
</tr>
<tr>
<td>Pull-in delay, $t_{PI}$</td>
<td>$1 / \kappa$</td>
</tr>
<tr>
<td>Switching energy</td>
<td>$1 / \kappa^3$</td>
</tr>
<tr>
<td>Device density</td>
<td>$\kappa^2$</td>
</tr>
<tr>
<td>Power density</td>
<td>$1$</td>
</tr>
</tbody>
</table>

### Scaled Relay Design

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actuation Area, $A$</td>
<td>$1 \ \mu m^2$</td>
</tr>
<tr>
<td>Actuation Gap, $t_{\text{gap}}$</td>
<td>$4.3 \ \text{nm}$</td>
</tr>
<tr>
<td>Dimple Gap, $t_{\text{dimple}}$</td>
<td>$3 \ \text{nm}$</td>
</tr>
<tr>
<td>Pull-in voltage, $V_{PI}$</td>
<td>$9.4 \ \text{mV}$</td>
</tr>
<tr>
<td>Release voltage, $V_{RL}$</td>
<td>$6 \ \text{mV*}$</td>
</tr>
<tr>
<td>^Pull-in delay, $t_{PI}$</td>
<td>$25 \ \text{ns}$</td>
</tr>
</tbody>
</table>

*Assumes negligible contact adhesion

Beam length = 3.5 \ \mu m, thickness = 40 \ \text{nm}, width = 83 \ \text{nm}
Digital Circuit Design with Relays
F. Chen et al., ICCAD 2008

• **CMOS**: delay is set by electrical time constant
  – Quadratic delay penalty for stacking devices
  → Buffer & distribute logical/electrical effort over many stages

• **Relays**: delay is dominated by mechanical movement
  – Can stack ~100 devices before $t_{d,elec} \approx t_{d,mech}$
  → Implement relay logic as a single complex gate
Single-Gate, Dual-Source/Drain Relay

I-R. Chen et al., ECS Spring Meeting 2012 and Transducers 2013

Plan-View SEM

Temperature Dependence:

Measured I-V Characteristics

\[ I_{DS1}, I_{DS2}, I_G \]

\[ V_B = -7.1V \]

\[ V_B = 0V \] with body bias
Relay Multiplier Sub-Ckt: (7:3) Compressor

Fariborzi et al., ASSCC 2011

- Most complex scaled MEM-relay integrated circuit reported to date (46 relays)
Energy-Delay Comparison with CMOS

M. Spencer et al., JSSC 2011; H. Fariborzi et al., ESSCIRC 2011

- 90nm relay vs. CMOS adders and multipliers:
  >2-100× energy savings @ 3-100× higher delay

Adder Comparison

Multiplier Comparison
Single-Gate Relay Inverter/Buffer

R. Nathanael et al., VLSI-TSA 2012

Voltage levels:
- \( V_{DD} = 1\text{V} \)
- \( V_{B\text{HIGH}} = 13\text{V} \)
- \( V_{B\text{LOW}} = -12\text{V} \)
- \( GND = 0\text{V} \)

**Graphs:**
- Input voltage vs. time
- Output voltage vs. time

**Diagram:**
- Circuit schematic
  - Input \( V_{IN} \)
  - Output \( V_{OUT} \)
  - Buffer (BUF)
  - Inverter (INV)

**Photograph:**
- SEM image of the integrated circuit with dimensions labeled 50 \( \mu\text{m} \).
Dual-Gate, Dual Source/Drain Relay

R. Nathanael et al., VLSI-TSA 2012

- Gate electrodes are interdigitated to ensure that each gate has equal influence on the movable body.
Measured $V_{PI}$ & $V_{RL}$ of Dual-Gate Relay

R. Nathanael et al., VLSI-TSA 2012

- "1"≡$V_G$
- Each gate has equal influence
- Depending on $V_B$, relay can be actuated using one or two gate electrodes
Dual-Gate Relay Circuit: AND/NAND

R. Nathanael et al., VLSI-TSA 2012

V_{DD} = 8V

V_{B\_LOW} = -4V

V_{B\_HIGH} = 15V

GND = 0V

V_{IN1} 0 1 1 1 0

V_{IN2} 0 1 0 1 0

V_{OUT} (AND) 0 0 1 0 0

V_{OUT} (NAND) 0 1 0 1 0

Time (s)

0 0.1 0.2 0.3
Dual-Gate Relay Circuit: OR/NOR

R. Nathanael et al., VLSI-TSA 2012

- \( V_{DD} = 8\text{V} \)
- \( V_{B\_LOW} = -6\text{V} \)
- \( V_{B\_HIGH} = 12\text{V} \)
- \( V_{\text{OUT (OR)}} \)
- \( V_{\text{OUT (NOR)}} \)
- \( V_{\text{IN1}} \)
- \( V_{\text{IN2}} \)
- \( V_{\text{OUT}} \)
- GND = 0\text{V}
See-Saw (Active Turn-Off) Relay


- Perfectly complementary switching, symmetric about $V_{DD}/2$

Scanning Electron Micrographs:

$I_D$-$V_G$ characteristics

Switching Voltages vs. $V_{DD}$

$V_{ON_LEFT} = V_{OFF_LEFT} = 3.16V$

$V_{ON_RIGHT} = V_{OFF_RIGHT} = 7.14V$

$L_A = 42\mu m$

$W_A = 40\mu m$

$V_{BR}/2 = V_{DD}/2$

$V_{ON_LEFT}$

$V_{ON_RIGHT}$
See-Saw Relay Logic Gates


Circuit Symbol

Relay Configurations

<table>
<thead>
<tr>
<th>Function</th>
<th>$V_G$</th>
<th>$V_{BL}$</th>
<th>$V_{SL}$</th>
<th>$V_{BR}$</th>
<th>$V_{SR}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BUFFER</td>
<td>$V_{IN}$</td>
<td>LO</td>
<td>HI</td>
<td>HI</td>
<td>LO</td>
</tr>
<tr>
<td>INVERTER</td>
<td>$V_{IN}$</td>
<td>LO</td>
<td>LO</td>
<td>HI</td>
<td>HI</td>
</tr>
<tr>
<td>AND</td>
<td>$V_{IN1}$</td>
<td>LO</td>
<td>$V_{IN2}$</td>
<td>HI</td>
<td>LO</td>
</tr>
<tr>
<td>OR</td>
<td>$V_{IN1}$</td>
<td>LO</td>
<td>HI</td>
<td>HI</td>
<td>$V_{IN2}$</td>
</tr>
</tbody>
</table>

Demonstrated Single-Input Operations

- **BUFFER**
  - $V_{IN}$ (0, 1)
  - $V_{OUT}$ (0, 1)

- **INVERTER**
  - $V_{IN}$ (1)
  - $V_{OUT}$ (0)

- **INPUT**
  - $V_{IN}$ (0, 1)

Demonstrated Dual-Input Operations

- **AND**
  - $V_{IN1}$ (00, 01, 11, 10)
  - $V_{IN2}$ (00, 01, 11, 11)
  - $V_{OUT}$ (0, 0, 1, 0, 1, 1)

- **OR**
  - $V_{IN1}$ (00, 01, 11, 10)
  - $V_{IN2}$ (00, 01, 11, 11)
  - $V_{OUT}$ (0, 1, 1, 1, 1, 1)
Stiction: Relay Scaling Limiter (?)

In order for a conventional electrostatic switch to turn off properly,

\[ F_{\text{spring}} > F_{\text{adhesion}} \]

\[ F_{\text{adhesion}} = \frac{H_c}{6\pi g^3} A_c \]

- \( H_c \) = Hamaker constant [zJ]
- \( g \) = separation between contacting surfaces [nm]
- \( A_c \) = real contact area [nm²]

To turn on the relay: electrostatic force \( F_{\text{elec}} > F_{\text{spring}} > F_{\text{adhesion}} \)

\( \rightarrow \) Limits scaling of actuation area \( A \) and/or \( V_{\text{pl}} \)
Contact Adhesive Force Scaling


Schematic cross-section of contact dimple:

- Model indicates that adhesion is due to van der Waals force
  - $F_A$ should scale with area, $0.02 \text{nN/nm}^2$ for W-W contact

- Adhesive force is lowered by TiO$_2$ coating
4-T piezoelectric relay

**I-V Characteristics**

- Small hysteresis is indicative of low surface adhesion

**Input/Output Waveforms**

- ~23 aJ projected switching energy for 5 mV operation
In contrast to logic switches, NVM devices can have:

- Long write/erase times (>1 us)
- Modest endurance (< $10^6$ cycles)
- Large operating voltages (>1 V)
Outline

• Introduction

• Nanomechanical switches for computing

• Nanomechanical memory
  – Early designs
  – Recent design for cross-point array architecture

• Summary
First MEMS NVM Cell


- Bistable buckled beam
  - Switched with electrostatic force
  - Immune to radiation, shock

- Too large (>100um²) for reasonable storage capacities...

Fig. 1. Schematic drawing of a micro-electro-mechanical nonvolatile memory cell based on a bistable bridge (B) on a spacer (S) on the substrate (SUB) with lateral electrodes (L). The materials used are indicated, and the symbols for the dimensions are defined.

Fig. 6. Experimentally observed and calculated switching voltages $V$ as a function of the length $l_0$ of a 300-Å-thick bridge.
Nanomech™ Technology

R. Gaddi et al. (Cavendish Kinetics), Microelectronics Reliability Vol. 50, pp. 1593-1598, 2010

TEM cross-section

Endurance test results

• The beam can be pulled out of contact by biasing the cap.
• A select device (e.g. a transistor) is needed for each MEM switch in the memory array.
Cross-Point Array Architecture

- Most compact architecture (4F² cell size)
- Requires selector device at each cross-point to reduce “sneak” leakage current during a read operation
  - e.g. 2-terminal diode

![Diagram of Cross-Point Array Architecture]

Electro-Mechanical Diode NVM Cell


- Achieve bi-stable operation of an electrostatic gap-closing actuator by leveraging the built-in electric field of a diode:

\[ F_{\text{adhesion}} + F_{\text{elec}} > F_{\text{spring}} \]

**CROSS-SECTIONAL ILLUSTRATIONS**

**Reset State**
- WL (p-type)
- BL (n-type)
- Isolation Dielectric

**Set State**
- WL (p-type)
- BL (n-type)
- Isolation Dielectric

**Parameters:**
- \( V_{g} \)
- \( V_{RL} \)
- \( V_{PI} \)
- \( V_{Read} \)
- 2 stable states
- pull-in
- release

**Diode Bias:**
- Forward diode bias
- Reverse diode bias
First Prototype MEM Diode


- Cross-point arrays of MEM diodes were fabricated using conventional planar processing techniques.
  - The WL (100-nm p-type poly-Si$_{0.4}$Ge$_{0.6}$) is supported by SiN$_x$ spacers formed along the sidewalls of the BL (100 nm n-type poly-Si).
  - The air-gap (~13 nm thick) between the WL and BL is formed by selectively removing a sacrificial layer of LTO using HF vapor.

Bird’s-eye view of memory array

Cross-sectional view of a bit-cell
Set Operation


- A sudden increase in current is seen at the voltage when the WL is pulled in to the BL.
  - $V_{\text{Set}} \approx 6 \text{ V}$, $t_{\text{Set}} \approx 2 \text{ us}$

- Since it is an applied voltage (not current) that is required to actuate the WL, the Set current can be lowered by inserting a current-limiting resistor.
**I-V Characteristics**


- **Set/Reset current ratio > $10^6$**
  - Leakage through SiN$_x$ spacers can be reduced with process improvements.

- **For $V_{\text{Read}} = 1.2$ V, the ratio of Set cell current to sneak path leakage current is ~100 for the prototype device**
  - This rectification ratio can be increased with process improvements.
To Reset a memory cell, a voltage pulse is applied to counteract the built-in field of the p-n diode.

- forward bias $\rightarrow$ current flow

A sudden decrease in current is seen at the voltage when the WL comes out of contact with the BL.

- $V_{\text{Reset}} \approx -6 \, \text{V}$, $t_{\text{Reset}} \approx 100 \, \text{ms}$
Retention Behavior


![Graph showing retention behavior over time](image-url)
Endurance Characteristics


P = 1 atm; T = 25°C

Set: $V_{BL}=6V$, $V_{WL}=0V$, 1msec
Reset: $V_{BL}=-15V$, $V_{WL}=0V$, 100msec
Read: $V_{BL}=-1.2V$
## Scaled NVM Technology Comparison

**W. Kwon et al. (UC Berkeley), 2012 International Memory Workshop**

<table>
<thead>
<tr>
<th>Technology</th>
<th>NAND Flash</th>
<th>PCM</th>
<th>STT-MRAM</th>
<th>Redox RRAM</th>
<th>Electro-mechanical diode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell Size</td>
<td>2.5 F²</td>
<td>6 F²</td>
<td>20-40 F²</td>
<td>5-8 F²</td>
<td>4-6 F²</td>
</tr>
<tr>
<td>Scaling Limit</td>
<td>16 nm</td>
<td>5-10 nm</td>
<td>7-10 nm</td>
<td>5-10 nm</td>
<td>5 nm W x 20 nm L</td>
</tr>
<tr>
<td>Storage Mechanism</td>
<td>F-N Tunneling</td>
<td>Phase change by Joule heating</td>
<td>Electron spin torque transfer</td>
<td>Ion transport and redox reaction</td>
<td>Mechanical gap closing actuation</td>
</tr>
<tr>
<td>Write/Erase Voltage</td>
<td>18-20 V</td>
<td>&lt; 3 V</td>
<td>&lt; 1.8 V</td>
<td>&lt; 0.5 V</td>
<td>&lt; 3 V</td>
</tr>
<tr>
<td>Write time</td>
<td>&gt;10 us</td>
<td>50-120 ns</td>
<td>&lt; 100 ns</td>
<td>&lt; 5 ns</td>
<td>&lt; 1 ns</td>
</tr>
<tr>
<td>Endurance</td>
<td>10⁴ - 10⁵</td>
<td>10¹⁵</td>
<td>10¹²</td>
<td>10¹⁶</td>
<td>&gt;10¹⁰</td>
</tr>
<tr>
<td>Retention</td>
<td>10 yrs</td>
<td>10 yrs</td>
<td>10 yrs</td>
<td>10 yrs</td>
<td>&gt; 10 yrs @ 200 °C</td>
</tr>
<tr>
<td>Ease of Integration</td>
<td>10 Masks</td>
<td>2-3 Masks to BEOL</td>
<td>3-4 Masks to BEOL</td>
<td>2-3 Masks to BEOL</td>
<td>2 Masks</td>
</tr>
<tr>
<td>Write Energy per Bit</td>
<td>&gt; 1 fJ</td>
<td>&lt; 2 pJ</td>
<td>&lt; 4 pJ</td>
<td>1 fJ</td>
<td>&lt; 0.1 fJ</td>
</tr>
</tbody>
</table>


- NEMory technology potentially offers the best performance and lowest energy consumption!
3-D Integration

- Conventional cross-point memory array technologies require 2 lithography steps per memory layer

- NEMory technology requires only 2 lithography steps to define the 3-D memory array
  - built-in redundancy (2 WL/cell)
  - allows for longer beam length

Unity Semiconductor Corp.

Outline

• Introduction

• Nanomechanical switches for computing

• Nanomechanical memory

• Summary
Summary

• Electronic devices which enable more energy-efficient computing and information storage, at ever lower cost per function, will be required to realize the vision of ambient intelligence.

• Mechanical devices show promise in this regard:
  ✓ Ideal switching behavior
  ✓ Simple fabrication process (lower cost)
  ✓ Enhanced device functionality
  ✓ Adequate reliability

Contact adhesive forces must be well controlled in order to fully realize this promise.
Acknowledgements

- **Faculty:** Elad Alon, Vladimir Stojanović; Dejan Marković (UCLA)
- **Former post-docs:** Vincent Pott (LSI Logic), Louis Hutin (CEA-LETI)
- **Former students:** Hei Kam, Jaeseok Jeon, Rhesa Nathanael, Jack Yaung, Wookhyun Kwon
  @ MIT: Fred Chen, Sumit Datta, Hossein Fariborzi
  @ UCLA: Kevin Dwan, Cheng C. Wang, Mingzhe Jiang
- **Students:** I-Ru Chen, Yenhao Chen, Chuang Qian, Matt Spencer
- **SEMatech:** Muhammad Hussain, Casey Smith, Chanro Park, Weize Xiong, Rinus Lee
- **Research Funding:**
  - DARPA/MTO NEMS Program
  - DARPA/MARCO Focus Center Research Program
  - NSF Center of Integrated Nanomechanical Systems (COINS)
  - NSF Center for Energy Efficient Electronics Science (E3S)
- **UC Berkeley Micro/Nanofabrication Laboratory**