Materials and Devices beyond CMOS Transistor for Energy Efficient Computing

Dmitri Nikonov

Exploratory IC, Components Research

Intel

Seminar at Princeton Plasma Physics Laboratory
Outline

- Moore’s Law scaling and the energy crisis
- Beyond-CMOS devices for lower energy
- Spintronics materials and devices
- Ferroelectric and multiferroic materials and devices
- Benchmarking of beyond-CMOS devices
Moore’s Law

Double the number of transistors on a chip every 2 years.

Moore’s Law is Alive and Well

Moore’s Law – The number of transistors on integrated circuit chips (1971-2018)

Moore’s law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years. This advancement is important as other aspects of technological progress – such as processing speed or the price of electronic products – are linked to Moore’s law.

The data visualization is available at OurWorldInData.org. There you find more visualizations and research on this topic. Licensed under CC-BY-SA by the author Max Roser.
CMOS Supply Voltage - Historical Trend

In the last 15 years voltage scaling is stalled

Dennard scaling

5.0V

0.7V
Semiconductor industry faced the power crisis before with bipolar transistors.

Source: Chen (IBM), ISS Europe, 2007.
Energy Crisis

- Exploding demand for computing due to datacenters, AI
- Required energy will approach a few % of world production by 2030
- CMOS business as usual will lead to stalling IT, deficit of computing
- Need more energy efficient devices to continue sustainable development, curb carbon emissions

SRC, Decadal Plan for Semiconductors, 2021
Collective States = Energy Efficiency

\[ E = e\Delta VN \sim 4000kT \]
Leakage determined by barrier

\[ I_{on}/I_{off} < \exp\left(\frac{e\Delta V}{kT}\right) \]

Generic Electronic Switch | Generic Spintronic Switch
--- | ---
Barrier | 20 kT (from Ion/Ioff) | 60 kT (non-volatile)
Voltage | 0.5 – 1 V | 10-100 mV
Particles | Ne = 200 electrons | Ns = 10000 spins
Sw. Energy Limit | 4000kT = Ne*20kT | 60 kT
Phenomenon | Non collective | Collective

\[ E = \frac{1}{2}\mu_0\mu_B N_s H_k \sim 60kT \]
Leakage not related to barrier
## 2 Collective States = Non-Volatility

<table>
<thead>
<tr>
<th>Class</th>
<th>Variables</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Charge</td>
<td>Q, I, V</td>
<td>CMOS, TFET</td>
</tr>
<tr>
<td>Electric Dipole</td>
<td>P</td>
<td>FeFET</td>
</tr>
<tr>
<td>Magnetic Dipole</td>
<td>M, I\text{spin}</td>
<td>ASL, SWD, NML</td>
</tr>
<tr>
<td>Orbital State</td>
<td>Orb, Bose condensate</td>
<td>BisFET</td>
</tr>
<tr>
<td>Strain</td>
<td>\sigma</td>
<td>PiezoFET</td>
</tr>
</tbody>
</table>

Can have non-volatile states at room temperature
Beyond-CMOS Devices, part 1

Electronic
Tunneling FET - multiple!!!

Ferroelectric
Negative Cap FET
FEFET
PiezoFET

Straintronic
MITFET

Graphene pn Junction
ITFET

Orbitronic
BisFET
Beyond-CMOS Devices, part 2

SpinFET

Domain Wall Logic

All Spin Logic

Spin-Torque Triad

Spin Torque Oscillator

Nano Magnet Logic

Spin Majority Gate

Charge-spin logic

Spin Wave Device
Tunneling Field-Effect Transistor

Tunnel FETs operate by tunneling through the S/D barrier rather than diffusion over the barrier.

Two required conditions:

- Thin enough barrier over a large enough area for effective (high current) tunneling.
- Sufficient density of states on both the transmission and receiving sides to provide energetic locations for the carriers.
TFET Sub-threshold Slope

Tunneling probability increases sharply at the onset of Source Valance Band and Channel Conduction Band overlap.
Magnetoresistance

**Parallel**

- **FM**
- **Ferromagnet (FM)**

**Anti-parallel**

- **Magnetization**
- **Current**

\[
R_{AP} > R_P \\
MR = \frac{R_{AP} - R_P}{R_P}
\]

- Resistance of the stack with anti-parallel magnetizations is higher
- Magnetoresistance definition
Spin transfer torque

M = magnetization of free layer
p = polarization of injected electrons from pinned layer

Electrons get transmitted and reflected at the barrier.
Each brings a unit of spin $\hbar/2$
Combined transfer of angular momentum is torque, which rotates magnetization.
Spin-Orbit Torque for Low-Power

- **Need:** Operate memory and logic at 0.1V supply.
- **Method:** Macrospin switching by spin-orbit effect + spin drift-diffusion.
- **Result:** Spin-orbit effect produces faster magnetization switching at much lower voltage and energy than Spin transfer torque.

![Diagram of magnetic layer stack with current and voltage](image)

Multiferroic BiFeO$_3$

- **Ferroelectric (FE)** below $T_C = 1100$ K
- Fe atoms shift to corner of the cube in E-field
- **Antiferromagnetic (AFM)** below $T_N = 640$ K
- Spins on Fe interchange in direction
- So far one of 3 room temperature multiferroics

Coupling of electric and magnetic above room temperature
Multiferroic BiFeO$_3$

- Ferroelectric (FE) below $T_C = 1100$ K
- Fe atoms shift to corner of the cube in E-field
- Antiferromagnetic (AFM) below $T_N = 640$ K
- Spins on Fe interchange in direction
- So far one of 3 room temperature multiferroics

Coupling of electric and magnetic above room temperature
Anti-Ferromagnetic Order, L

- Superexchange = electrons hop Fe – O – Fe
- Forbidden if spins are parallel
- Lower energy if spins are anti-parallel, two sub-lattices M1 and M2
- G-type anti-ferromagnetism = spin reverses along all cubic directions
- AFM vector L, along the line of spins
Canted Magnetization, Mc

- Oxygen octahedra are not straight (Jahn-Teller distortion) but tilted
- If an oxygen is shifted from the straight line, modified exchange = Dzyaloshinskii-Moriya interaction (DMI)
- The two neighbor spins are not exactly opposite. Resulting “canted magnetization” Mc
- P, L, and Mc perpendicular to each other, right triple
Magnetoelectric Switching

- Magnetoelectric effect = voltage-controlled switching of magnetization (charging a capacitor)
- More energy efficient than charge-controlled switching (spin torque)
- Magnetoelectric multiferroic, BiFeO3
Large spin orbit coupling and inverted valence and conduction band states result in spin momentum locked surface states which have large $\theta_{soc}$. 
Spin to Charge Conversion with Spin-Orbit

- High efficiency spin to charge conversion using spin orbit effects.
- Read off of the magnetization state.

ZPL90 in-situ deposition of Full MESO device stack: ME and SO Films

Multiferroic films (ME) in PLD chamber; magnet, heavy metal and TI films (SO) in PVD chamber
Sub-100mV Logic Device Research Based On Magneto-Electric and Spin-Orbit Effects (MESO)

MESO enables 4 CMOS nodes/generations of energy efficiency improvement same CMOS node.

Inputs to Benchmarking – Lower Voltage

Lower Voltage = Best Path for Low Energy

Tunnel FETs:
Lower E*D than CMOS.

Magnetoelectric Spintronics:
Slower, but lower energy; and Non-Volatile.
Throughput vs. Capped Power

Cap=10W/cm²

Throughput, TIOPS/cm²

Power Density, W/cm²

32bit ALU

CMOS Ref
Electronic
Spintronic
Ferroelectric
Orbitronic
Straintronic

Tunnel FETs:
Rival CMOS in throughput at lower power.

Magneto-electric Spintronic:
Very low power.

TIOPS = Tera Integer Operations Per Second

Seminar at Princeton Plasma Physics Laboratory
Take-Aways

- Moore’s Law scaling of integrated circuits give exponential improvement of computing capacity but leads to the energy crisis.
- Beyond-CMOS devices can switch at lower energy and promise the solution of the energy crisis.
- Spintronics devices are based on spin torques.
- Ferroelectric and multiferroic devices utilize lower energy switching of non-volatile order parameters.
- Benchmarking of beyond-CMOS devices was developed and used for identifying promising devices, such as MESO.
BACKUP
Nanomagnet Energy Barrier

Energy barrier not lowered = reason for non-volatility
CMOS Challenge With Energy

- As CMOS scales -> energy/op decreases.
- But energy/op not decreasing fast enough (for 2x increase transistors/cm²).
Power density approaches a Power Density constraint

Magnetoresistance and Its Uses

1000x capacity of hard drives

A. Fert  P. Grunberg

Nobel Prize 2007, physics
# Nomenclature of Beyond-CMOS Devices

<table>
<thead>
<tr>
<th>Device name</th>
<th>acronym</th>
<th>input(s)</th>
<th>control</th>
<th>int. state</th>
<th>output</th>
<th>class</th>
<th>subclass</th>
</tr>
</thead>
<tbody>
<tr>
<td>Si MOSFET high perf.</td>
<td>CMOS HP</td>
<td>V</td>
<td>Vg</td>
<td>Q</td>
<td>V</td>
<td>electronic</td>
<td>barrier</td>
</tr>
<tr>
<td>Si MOSFET low voltage</td>
<td>CMOS LV</td>
<td>V</td>
<td>Vg</td>
<td>Q</td>
<td>V</td>
<td>electronic</td>
<td>barrier</td>
</tr>
<tr>
<td>van der Walls FET</td>
<td>vdWFET</td>
<td>V</td>
<td>Vg</td>
<td>Q</td>
<td>V</td>
<td>electronic</td>
<td>barrier</td>
</tr>
<tr>
<td>Homojunction III-V TFET</td>
<td>HomJTFET</td>
<td>V</td>
<td>Vg</td>
<td>R</td>
<td>V</td>
<td>electronic</td>
<td>tunneling</td>
</tr>
<tr>
<td>Heterojunction III-V TFET</td>
<td>HeJTFET</td>
<td>V</td>
<td>Vg</td>
<td>R</td>
<td>V</td>
<td>electronic</td>
<td>tunneling</td>
</tr>
<tr>
<td>Graphene nanoribbon TFET</td>
<td>gnrFTET</td>
<td>V</td>
<td>Vg</td>
<td>R</td>
<td>V</td>
<td>electronic</td>
<td>tunneling</td>
</tr>
<tr>
<td>Interlayer tunneling FET</td>
<td>ITFET</td>
<td>V</td>
<td>Vg</td>
<td>R</td>
<td>V</td>
<td>electronic</td>
<td>tunneling</td>
</tr>
<tr>
<td>Two D Heterojunction Interlayer TFET</td>
<td>ThinFET</td>
<td>V</td>
<td>Vg</td>
<td>R</td>
<td>V</td>
<td>electronic</td>
<td>tunneling</td>
</tr>
<tr>
<td>GaN TFET</td>
<td>GaNFTET</td>
<td>V</td>
<td>Vg</td>
<td>R</td>
<td>V</td>
<td>electronic</td>
<td>tunneling</td>
</tr>
<tr>
<td>Transition Metal Dichalcogenide TFET</td>
<td>TMDFET</td>
<td>V</td>
<td>Vg</td>
<td>R</td>
<td>V</td>
<td>electronic</td>
<td>tunneling</td>
</tr>
<tr>
<td>Graphene pn-junction</td>
<td>GpnJ</td>
<td>V</td>
<td>Vg</td>
<td>R</td>
<td>V</td>
<td>electronic</td>
<td>refraction</td>
</tr>
<tr>
<td>Ferroelectric FET</td>
<td>FEFET</td>
<td>V</td>
<td>Vg</td>
<td>P</td>
<td>V</td>
<td>ferroelectric</td>
<td>hysteresis</td>
</tr>
<tr>
<td>Negative capacitance FET</td>
<td>NCFET</td>
<td>V</td>
<td>Vg</td>
<td>P</td>
<td>V</td>
<td>ferroelectric</td>
<td>non-hysteresis</td>
</tr>
<tr>
<td>Piezoelectric FET</td>
<td>PiezoFET</td>
<td>V</td>
<td>V</td>
<td>o</td>
<td>V</td>
<td>straintronic</td>
<td>polarization</td>
</tr>
<tr>
<td>Bilayer pseudospin FET</td>
<td>BisFET</td>
<td>V</td>
<td>Vg</td>
<td>BC</td>
<td>V</td>
<td>orbitronic</td>
<td>exciton</td>
</tr>
<tr>
<td>Excitonic FET</td>
<td>ExFET</td>
<td>V</td>
<td>Vg</td>
<td>BC</td>
<td>V</td>
<td>orbitronic</td>
<td>exciton</td>
</tr>
<tr>
<td>Metal-insulator transistor</td>
<td>MITFET</td>
<td>V</td>
<td>Vg</td>
<td>Orb</td>
<td>V</td>
<td>orbitronic</td>
<td>bandstructure</td>
</tr>
<tr>
<td>SpinFET (Sughara-Tanaka)</td>
<td>SpinFET</td>
<td>V</td>
<td>Vg, Vm</td>
<td>Q, M</td>
<td>V</td>
<td>spintronic</td>
<td>spin drift</td>
</tr>
<tr>
<td>All-spin logic</td>
<td>ASL</td>
<td>M</td>
<td>V</td>
<td>M</td>
<td>M</td>
<td>spintronic</td>
<td>spin diffusion</td>
</tr>
<tr>
<td>Charge-spin logic</td>
<td>CSL</td>
<td>I</td>
<td>V</td>
<td>M</td>
<td>I</td>
<td>spintronic</td>
<td>spin Hall</td>
</tr>
<tr>
<td>Spin torque domain wall</td>
<td>STT/DW</td>
<td>I</td>
<td>V</td>
<td>M</td>
<td>I</td>
<td>spintronic</td>
<td>domain wall</td>
</tr>
<tr>
<td>Spin majority gate</td>
<td>SMG</td>
<td>M</td>
<td>V</td>
<td>M</td>
<td>M</td>
<td>spintronic</td>
<td>domain wall</td>
</tr>
<tr>
<td>Spin torque oscillator</td>
<td>STO</td>
<td>I</td>
<td>V</td>
<td>M</td>
<td>I</td>
<td>spintronic</td>
<td>nanomagnet</td>
</tr>
<tr>
<td>Spin wave device</td>
<td>SWD</td>
<td>M</td>
<td>I or V</td>
<td>M</td>
<td>M</td>
<td>spintronic</td>
<td>spin wave</td>
</tr>
<tr>
<td>Nanomagnetic logic</td>
<td>NML</td>
<td>M</td>
<td>B or V</td>
<td>M</td>
<td>M</td>
<td>spintronic</td>
<td>nanomagnet</td>
</tr>
</tbody>
</table>
Majority Gates = More Efficient Compute

Adder = 28 transistors (at least)

... or just 2 majority gates (All Spin Logic)

... or just 3 majority gates (Nanomagnetic Logic)

... or just 1 majority gate (Spin Wave Devices)!

Fewer devices for same computing function
Full MESO Operation Animation
Beyond-CMOS devices require CMOS

A: Electronic

B: Ferroelectric

C: Magnetic: Current driven - Spin Torque

D: Magnetic: Voltage driven - Magneto-Electric

Contrary to concept of finding the next “switch” research
Electronic vs. Ferroelectric Circuits

Electronic

Switching time

\[ t_{el} \approx \frac{CV_{dd}}{I} \]

Switching energy

\[ E_{el} \approx CV_{dd}^2 \]

Ferroelectric

\[ Q = P_{fe}A + CV_{dd} \]

Charging, intrinsic time

\[ t_{ch} \approx \frac{Q}{I} \]

Switching energy

\[ E_{fe} \approx QV_{dd} \]

- Charging, intrinsic time
  - \[ t_{fe} \approx 70 \text{ ps} \]
Spintronic Writing Circuits

Current driven - spin torque

\[ U_b = K_u v_{nm} \]
\[ I_c = \frac{e a U_b}{h P} \]
\[ t_{stt} = \frac{e M_s v_{nm}}{g \mu_B P (3I_c - I_c)} \log \left( \frac{2\pi \sqrt{2k_B T}}{U_b} \right) \]
\[ E_{stt} = I_{dev} V_{dd} t_{stt} \]

Voltage driven - magnetoelectric

\[ P_{ms} = \varepsilon_0 \varepsilon_{ms} E_{ms} \]
\[ Q = P_s A + CV_{dd} \]
\[ t_{mag} = \frac{\pi}{2\gamma B_{me}} \]
\[ E_{me} \approx Q V_{dd} \]
Treatment of Interconnects

Electronic

\[ C_{ic} \approx l_{ic} \cdot 126aF / \mu m \]

\[ t_{ic} \approx 0.7C_{ic}V_{dd}/I \]

\[ E_{ic} \approx 0.5C_{ic}V_{dd}^2 \]

Neglecting resistance of wires

Spintronic

Cascaded nanomagnets for interconnects

\[ t_{ic} \approx t_{mag} + l_{ic}/c \]

Propagation delay

\[ E_{ic} \approx E_{mag} \]
Levels of Simulation

1\textsuperscript{st} principles

Many-body quantum mechanics. E.g. Density Functional Theory

Numerous tools

2\textsuperscript{nd} principles

Atomistic energies and coupling constants. E.g. Tight-binding

Prof. Iniguez (Luxembourg)

3\textsuperscript{rd} principles

Continuous medium. E.g. Landau-Khalatnikov eqs.

Intel

Today

Seminar at Princeton Plasma Physics Laboratory
Macrospin (i.e. no spatial variation, no exchange stiffness)
All \( m \) are unit vectors. The two sublattices equivalently described:

\[
L = \frac{\hat{m}_1 - \hat{m}_2}{2} \\
M_c = \hat{m}_1 + \hat{m}_2 \\
\hat{m}_1 = L + \frac{M_c}{2} \\
\hat{m}_2 = -L + \frac{M_c}{2}
\]
Exchange Bias and Exchange Coupling

Exchange bias
Acts as field along CANTED MAGNETIZATION (Mc)

Exchange coupling
Acts as easy axis anisotropy along ANTIFERROMAGNETIC (L)

\[ F_{FM-AFM} = M_{fm} H_{eb} (\hat{m}_c \cdot \hat{m}) - \frac{M_{fm} H_{ec}}{2} (l \cdot \hat{m})^2 \]

SEMIFERROIC, BFO

FM, CoFe
Magneto-Electric Spin-Orbital (MESO) Device

- The way to lower switching energy $E \sim CV^2$, is lowering voltage
- 12 years of research in the Semiconductor Research Corporation (SRC)
- Magnetization switching can be done at lower voltage (~0.1V)
- Non-volatility of logic = built-in registers and latches = added benefit

Insert’s effect on spin to charge conversion efficiency

Depending on the sign of spin orbit coupling of the new surface states they can enhance or reduce \( \theta_{SOC} \). Doping in TI needs careful study but is promising.
Spin Orbit Module – Reads Magnetization

Direction of the magnet controls the direction of the charge output
Direction of current determines the sign of input voltage for next stage → Cascading
Spin Orbit Module – Material Functionality

Ferromagnet

Spin orbit coupling material

Resistivity: $\rho_{SOC}$

Spin to Charge efficiency: $\theta_{SOC}$

Spin diffusion length: $\lambda_{sd}$

Geometric Factors

Topological Insulators have high $\rho_{SOC}$ and large $\theta_{SOC}$

\[
\frac{\Delta V_{SOC}}{I_{Supply}} = P_{FM} \rho_{SOC} \theta_{SOC} \frac{1}{t_{SOC} w_{SOC}} \tanh \left( \frac{t_{SOC}}{2\lambda_{sd}} \right)
\]

$I_{Supply} = 10 \mu A$
Measurement Results for Topological Insulators

Topological Insulators have large $\theta_{SOC}$
Optimizing Spin to Charge conversion in TI

Doping TI with insert layer can protect the surface states and also enhance $\theta_{SOC}$. 

Seminar at Princeton Plasma Physics Laboratory
Delay vs. Area

Spintronics is slower than electronics, but more compact.

If power per area exceed the cap (10W/cm²), effective area is rescaled to be larger.
Exchange Bias and Exchange Coupling

- MESO is >10x lower energy than high-performance CMOS
- At the expense of slower speed
- Went through this trade off around 1990: transition from bipolar to CMOS transistors

Exchange Bias and Exchange Coupling

- CMOS is limited by dissipated power density
- Exhibited as the capability to remove heat from the chip, but mostly power available to the data center
- MESO is not limited by power, can achieve higher computing throughput (!)

Legal Notices

This presentation contains information provided by Intel Corporation ("Intel"), and may refer to Intel’s plans and expectations for the future, which are provided for discussion purposes only and are subject to change without notice. Forward-looking statements involve a number of risks and uncertainties: Refer to Intel’s SEC filings for authoritative discussion of Intel’s results and plans.

This presentation imposes no obligation upon Intel to make any purchase, and Intel accepts no duty to update this presentation based on more current information. Intel is not liable for any damages, direct or indirect, consequential or otherwise, that may arise, directly or indirectly, from the use or misuse of the information in this presentation.

Copyright © 2021 Intel Corporation.
Intel and the Intel logo, are trademarks of Intel Corporation in the U.S. and/or other countries. Other names and brands may be claimed as the property of others.