Stacked Horizontal Nanowire based 3-D Integration for Future High Performance Computing

Mostafizur Rahman (rahmanmo@umkc.edu)
University of Missouri-Kansas City
Beyond CMOS Opportunities

- Classical/Quantum Computation
  - Quantum Gates
  - Quantum Algorithms
  - In Memory Circuits
  - Asymmetric Circuits
  - Memristor (TMO, 2-D insulators..)
  - PCM (GST, GES..)
  - Straintronics (MESH, MTJ)
  - Ferroelectric (NCFET, P-FET..)
  - TFET (III-V, gnTFET..)
  - Digital FETs
  - Analog FETs
  - Approximate gates
  - 3-D circuits

- General Purpose Computing
  - Boolean Gates
  - Majority Logic
  - Neuronal Networks
  - Threshold Gates
  - Binary Gates
  - Spin torque (SFET, ASL, STO..)
  - Spin (RRAM)
  - Neuron Logic
  - Synaptic Circuits
  - Majority Logic
  - Multivalued Circuits
  - Threshold Gates
  - Binary Gates
  - Spin torque (SFET, ASL, STO..)

- Machine Learning Algorithms
  - Implication Logic
  - Neural Nets
  - Synaptic Circuit
  - Memristor (TMO, 2-D insulators..)
  - PCM (GST, GES..)
  - Straintronics (MESH, MTJ)
  - Ferroelectric (NCFET, P-FET..)
  - TFET (III-V, gnTFET..)

- Classical/Probabilistic Computing
  - Approximate gates
  - 3-D circuits

- Classical/In Memory Computation
  - Implication Logic
  - Neural Nets
  - Synaptic Circuit
  - Memristor (TMO, 2-D insulators..)
  - PCM (GST, GES..)
  - Straintronics (MESH, MTJ)
  - Ferroelectric (NCFET, P-FET..)
  - TFET (III-V, gnTFET..)

- Classical/Boolean Computation
  - Boolean Gates
  - Majority Logic
  - Neuronal Networks
  - Threshold Gates
  - Binary Gates

- Energy Consumption & Performance (High to Low)
  - Classical/Probabilistic Computing
  - Classical/In Memory Computation
  - Classical/Boolean Computation
  - General Purpose Computing
  - Machine Learning Algorithms

Architecture/Application Layer
Circuit Layer
Device Layer

SRC: Intel BCB 2017, SRC NRI, ITRS
3-D IC Classification

- **Wafer-Wafer**
  - Single Wafer
  - Multi-Chip Planar
  - Wire Bonding
  - Wire and Pad Bonding
  - Silicon Interposer
  - Integrated TSV

- **Die-Die**
  - TSV Stacked Memory
  - Heterogenous Dies with TSVs

- **Device-Device**
  - Monolithic 3-D

**Logic Density**
Stacked Horizontal Nanowire based 3-D IC (SN3D) Fabric

- Architected components to address device, circuit, connectivity, heat management and manufacturing requirements in 3-D

- Stacked nanowires are building blocks
• SN3D uses nanowires in a single die based process

• Local connectivity through fabric features
Outline

• SN3D Core Components and Logic
• Benchmarking
• Thermal Management
• Cost Analysis
• Fabric assembly by Integration of core components
• Intrinsic thermal management – stark contrast to other 3-D direction
Full Adder Design in SN3D

- CMOS circuit style
  - Uses both n- and p-type V-GAA Junctionless transistors
- Fabric specific physical mapping
- Local interconnection, noise and delay mitigation through utilizing fabric features and circuit optimizations
SRAM Cell Design in SN3D

- Fabric specific physical mapping
- High drive strength transistors
- Stacked(series) Inverters – CC-CG
SN3D SRAM Array Organization and Benefits

- **Sharing FVs (BL, \(\overline{BL}\), WL, Vdd, Gnd)**—2 Effective FVs for each cell
- **3D Abutment of adjacent Cells**

![Two Adjacent Cells](image1.png)
![Top view](image2.png)
![Metal Routed Layout](image3.png)
![3D View of SRAM Array](image4.png)
Design Benefits

- 3:2:1 ratio for all designs
- TSVs add additional area overhead
- M3D is limited to two tier design, only 30% reduction in footprint
- M3D needs high precision alignment of inter-tier-vias
Benchmarking: Methodology

SN3D Circuit and Layout Design: Logic, SRAM etc.,

Density Evaluation

Interconnect Extraction (PTM Model)

RC calculations

SN3D TCAD Process Simulation

3-D TCAD Device I-V Simulation

3-D TCAD Device C-V Simulation

Device Modeling for HSPICE simulation

SN3D Circuit (Logic, SRAM) HSPICE Simulation

Functionality, Power and Performance Evaluation

Experimental Results

Design Rule

SN3D Design Rules

<table>
<thead>
<tr>
<th></th>
<th>Width - Z (nm)</th>
<th>Length - X (nm)</th>
<th>Thickness - Y (nm)</th>
<th>Spacing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transistor Channel</td>
<td>16(2λ)</td>
<td>16(2λ)</td>
<td>16(2λ)</td>
<td>39</td>
</tr>
<tr>
<td>Transistor Spacing</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>23</td>
</tr>
<tr>
<td>Gate Electrode</td>
<td>32(4λ)</td>
<td>16(λ)</td>
<td>34</td>
<td>-</td>
</tr>
<tr>
<td>Source/Drain contact</td>
<td>32(3λ)</td>
<td>24(3λ)</td>
<td>34</td>
<td>20</td>
</tr>
<tr>
<td>Bridge</td>
<td>24(3λ)</td>
<td>-</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Horizontal insulation</td>
<td>32(4 λ)</td>
<td>24 (3 λ)</td>
<td>5</td>
<td>-</td>
</tr>
</tbody>
</table>

UMKC
Benchmarking Results

- 6.4x performance/power, 67% area reduction for SRAM
- >10x area reduction, 19% and 18% performance and power improvements for logic

---

Outline

• SN3D Core Components and Logic
• Benchmarking
• Thermal Management
• Cost Analysis
• Self-heating exacerbates for vertical/SOI FETs
• Lack of heat dissipation paths in 3-D is a key problem for hotspot

Thermal Simulation Methodology

- The Conductive heat transfer in solids is obtained by:
  \[ \rho C_p \frac{\partial T}{\partial t} + \nabla \cdot (-k \nabla T) = Q \]
- Finite Element Based Method (FEM) for fine-grained modeling

### Device Materials and Their Dimensions

<table>
<thead>
<tr>
<th>Region</th>
<th>Material</th>
<th>Dimension (LxWxT)nm</th>
<th>Thermal Conductivity Wm(^{-1}) K(^{-1})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drain</td>
<td>Silicide</td>
<td>10 x 16 x 16</td>
<td>45.9</td>
</tr>
<tr>
<td>Drain Electrode</td>
<td>Ti</td>
<td>10 x 16 x 12</td>
<td>21</td>
</tr>
<tr>
<td>Channel</td>
<td>Doped Si</td>
<td>16 x 16 x 16</td>
<td>13</td>
</tr>
<tr>
<td>Source</td>
<td>Silicide</td>
<td>10 x 16 x 16</td>
<td>45.9</td>
</tr>
<tr>
<td>Gate Oxide</td>
<td>HfO(_2)</td>
<td>16 x 18 x 2</td>
<td>0.52</td>
</tr>
<tr>
<td>Gate Electrode</td>
<td>TiN</td>
<td>10 x 16 x 6</td>
<td>1.9</td>
</tr>
<tr>
<td>Spacer</td>
<td>Si(_3)N(_4)</td>
<td>10 x 16 x 16</td>
<td>1.5</td>
</tr>
</tbody>
</table>
Thermal Simulation Results (FEM)

- Temperature reduction by 53% to 375K from 700K
- Heat Pillar was effectively dissipating heat from heated region
- Heat Pillar can be shared among neighboring FETs
- Highest temperature was 375K for top transistors
Effectiveness of Heat Extraction Features

- SN3D’s connectivity features enable heat extraction
- Heat Junction and Heat Pillar reduces heat further
  - Can be customized to meet thermal requirements

SN3D circuit’s thermal profile

Outline

• SN3D Core Components and Logic
• Benchmarking
• Thermal Management
• Cost Analysis
Early Cost Estimation Methodology

Design phase information

Major Cost Dictating Aspects

Cost variables

Cost Model
Die Area Estimation

86%, 72% and 74% reduction in footprint compared to 2D, T3D and M3D respectively [7]

\[ A_{2D/\text{SN3D}} = \frac{N_G A_{G,2D/\text{SN3D}}}{N_G A_{G,3D/M3D} + N_{TSV/MIV} A_{TSV/MIV}} \]

Rent's Correlation for Terminal Count: \( T_{SN3D} = \left( \sum_{i=1}^{n} T_i \right) = nk \left( \frac{N_G}{n} \right)^p \)

Total Interconnect \( I_T = \alpha k N_G \left( 1 - N_G^p \right) \) \[2\]

Distribution \( i(l) = f(N_G, l, k, p) \) \[7\];

Parameterizing Process Steps

Arbitrary unit process constant:

$$k_c = k_{PL} + k_{DF} + k_{ET} + k_{DP} + k_{IM}$$

Relative Cost of the Major Process steps:

$$k_{PL} = 0.32k_c; k_{DF} = 0.22k_c; k_{ET} = 0.18k_c; k_{DP} = 0.16k_c; \text{and } k_{IM} = 0.12k_c$$

---

Quantifying Process Sequence

### Major Processes
- Photolithography
- Diffusion
- Etching
- Deposition
- Implantation

Typical Process steps Count:
- Photolithography: 9
- Diffusion: 4
- Implantation: 7
- Deposition: 5
- Etching: 5

$$c_p = 0.32n_{PL} + 0.22n_{DF} + 0.18n_{ET} + 0.16n_{DP} + 0.12n_{IM}$$

$$C_{SN3D} = 26.54k_c A_{SN3D} + (2n_{ET} + 1)k_c + 0.12n_{IM}k_c$$

---

Cost Estimation Results

Final Cost Models:

\[ C_{2D} = 6.26k_c A_{2D} + 2k_c n_m A_{2D} + C_{cooling} \]
\[ C_{3D}/C_{M3D} = 7.26k_c A_{3D} + 2k_c n_m A_{3D} + C_{bonding} + C_{cooling} \]
\[ C_{SN3D} = 26.54k_c A_{SN3D} + 2k_c n_m A_{SN3D} + C_{cooling} \]

- No Bonding Cost for SN3D
- Bonding Cost for TSV 3-D and M3-D are taken as a relative cost from \[11\]
- Cooling Cost: \[ C_{cooling} = K_c T + c \]

83% and 81% reduction in total cost compared 2-D CMOS Monolithic 3-D integration

SN3D Implementation Aspects


Summary

• Beyond CMOS opportunities span different application domains

• 3-D Integration is obvious choice for moving forward
  – Highest advantage with monolithic 3-D

• SN3D is a new stacked nanowire based 3-D IC technique
  – Integrated device, circuit, connectivity and heat management
  – Possibility of >10x density benefits
  – 83% cost reduction compared to 2-D CMOS
Acknowledgements

• Csaba Andras Moritz (Umass Amherst)
• Santosh Khasanvis (Bluerisc)
• Pritish Narayanan (IBM)
• Masud Chowdhury (UMKC)
• Kelin Kuhn (Cornell, prev. Intel)

Sponsors
NSF, Univ. of Missouri Research Fund