Abundant-Data Computing

The N3XT 1,000X

Subhasish Mitra

Department of EE & Department of CS
Stanford University
Solution: NanoSystems

Transform new nanotech into new systems enable new applications

New devices
New fabrication
New sensors
Abundant-Data Explosion

“Swimming in sensors, drowning in data”

- Mine, search, analyze data in near real-time
  - Data centers, mobile phones, robots
Abundant-Data Applications

Memory wall: processors, accelerators

Deep Learning Accelerators

- AlexNet (CNN)
  - Compute: 15%, Memory: 85%

- ResNet-152 (CNN)
  - Compute: 20%, Memory: 80%

- Language Model (LSTM)
  - Compute: 8%, Memory: 92%
Nano-Engineered Computing Systems Technology
N3XT NanoSystems

Computation immersed in memory

Increased functionality

Memory

Computing logic

Fine-grained, ultra-dense 3D

Impossible with today’s technologies
Carbon Nanotube FET (CNFET)

1. **Energy Delay Product**
   - ~10X benefit
   - Full-chip case studies
   - [IBM, IMEC, Stanford, others]

2. **First CNT computer**

[Shulaker Nature 13, ISSCC 13, IEDM 14]
Example: OpenSPARC T2

[Diagram showing the comparison of total energy per cycle vs. clock frequency for different types of FETs: FinFET, Nanowire FET, and CNFET. The CNFET shows a significant 9x EDP benefit compared to the other types, marked as preferred.]
Putting into Perspective

- Existing technology benchmarking + CNFETs

![Graph showing the comparison between different FET types such as Si-CMOS, "beyond" CMOS, and CNFETs, with axes for adder frequency (GHz) and adder energy per operation (fJ)].

32-bit adder [Nikonov & Young, 2013 & 2015]
3D Integration

- Massive ILV density $\Rightarrow$ TSV density

TSV (chip stacking)

Through silicon via (TSV)

Dense, e.g., monolithic

Nano-scale inter-layer vias (ILVs)
Device + Architecture Benefits

Naturally enabled

Emerging logic
Emerging memory
Monolithic 3D integration

- Low-temperature fabrication: $< 400 \, ^\circ\!\mathrm{C}$

[Wei IEDM 13, Shulaker VLSI Tech 14]
First 3D NanoSystem

>2 Million CNFETs, 1 Mbit RRAM

Abundant data: Terabytes / second

In-situ classification:
Extensive, accurate

[Shulaker Nature 17]
N3XT Simulation Framework

Joint technology, design & app. exploration

- Heterogeneous technologies
- System-level analysis
- Architecture exploration
  - Energy, exec. time
  - Thermal
- Physical design, yield, reliability
- Abundant-data apps
Massive Benefits: Deep Learning, Graph Analytics, ...

~1,000X benefits, existing software
chip stacking: 2-4x benefits

<table>
<thead>
<tr>
<th>Benefits</th>
<th>PageRank</th>
<th>Connected Components</th>
<th>Breadth-First Search</th>
<th>Linear Regression</th>
<th>Language model (LSTM Neural Network)</th>
<th>AlexNet (Neural Network)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy</td>
<td>851×</td>
<td>400×</td>
<td>510×</td>
<td>970×</td>
<td>1,950×</td>
<td>210×</td>
</tr>
<tr>
<td>Execution Time</td>
<td>100×</td>
<td>10×</td>
<td>1×</td>
<td>1×</td>
<td>1×</td>
<td>1×</td>
</tr>
</tbody>
</table>

IBM graph analytics  
DeepDive  
BigDataBench  
TensorFlow
More Opportunities
Accelerators
Brain-inspired
Technology innovations

Conclusion

- **Nanosystems today**

- **Game ON, to the era**

- **N3XT 1,000X**
  - Compute + memory + sensing
  - Densely interwoven