3D Heterogeneous Systems: Design and Technologies

Muhammad S. Bakir
Nanotechnology Research Center

muhannad.bakir@mirc.gatech.edu
Acknowledgments to Team

• Calvin King, PhD
• James Yang, PhD
• Yue Zhang, PhD
• Ramasamy Ravindran, PhD
• Suzanne Bock, Engineer
• Mahavir Parekh, MS
• Parag Thadesar, PhD
• Ashish Dembla, PhD
• Paresa Modarres, PhD
• Li Zheng, PhD

Former Students:
• Jesal Zaveri, MS [Microsoft]
Outline

• Introduction: Overview of the Interconnect Problem

• 3D Stacking Technologies for Electronics
  – Accounting Signal, Power, and Thermal Interconnects

• 3D Stacking of CMOS and MEMS/sensors
  – Novel interface interconnects

• Conclusion
Turn of the Century Marked Many New Paradigms …

On-chip wires dominate:
- Latency & energy dissipation exceed transistor
- More masking levels

‘Simple’ scaling has ended:
- Lithography + strain + high-k
  (+ $V_{dd}/V_t$ scaling slowed)
- Non-planar CMOS next ???

Historic frequency scaling ended

Due to:
- Design complexity
- Energy efficiency
- But, no “free lunch”
Many-Core Processor Emergence

Aggregate off-chip bandwidth:
- Today: ~0.8 Tbps
- Soon: Several Tbps

Few challenges:
- Interconnect quality & density
- Latency (inches of wire)
- Power:
  - ~15-20% μP power used for signal I/Os

Some Challenges in Off-Chip Signaling

**Freq. Dependent Losses**

![Frequency vs. Loss](image)

- a: Total Loss
- b: Skin Loss
- c: Dielectric Loss

H. Dawei et al., IEEE J. Select Topics Quantum Electronics 2003

**Copper Surface Roughness**

![Copper Surface Roughness](image)

Card A- Ground Layer


Loss inc by 5.5%-49.5% at 5 GHz

**Not Enough Signal Pins**

![Image: T. Karnik et al., in Bakir, Meindl book, 2009](image)

**A Whole Lot of Discontinuities**

[*SiliconPipe website*]

B. Casper et al. CICC 2007
Energy Cost for Off-Chip Communication

<table>
<thead>
<tr>
<th>Operation</th>
<th>Energy (pJ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>64b Floating FMA (2 ops)</td>
<td>100</td>
</tr>
<tr>
<td>64b Integer Add</td>
<td>1</td>
</tr>
<tr>
<td>Write 64b DFF</td>
<td>0.5</td>
</tr>
<tr>
<td>Read 64b Register (64 x 32 bank)</td>
<td>3.5</td>
</tr>
<tr>
<td>Read 64b RAM (64 x 2K)</td>
<td>25</td>
</tr>
<tr>
<td>Read tags (24 x 2K)</td>
<td>8</td>
</tr>
<tr>
<td>Move 64b 1mm</td>
<td>6</td>
</tr>
<tr>
<td>Move 64b 20mm</td>
<td>120</td>
</tr>
<tr>
<td>Move 64b off chip</td>
<td>256</td>
</tr>
<tr>
<td>Read 64b from DRAM</td>
<td>2000</td>
</tr>
</tbody>
</table>

(Bill Dally, Advanced Computing Symposium, September 16, 2009)

• Lots of energy goes into communication
3D Stacking: Overview of Interconnect Benefits

- From inches to “gate pitch” length
  - 1,000x length reduction
- Lower latency
  - 0.1ns range for 3D
- Less power
  - 10-35 mW/Gbps today’s off-chip links
  - Better than 1 mW/Gbps for 3D
  - For reference, ~15-20% MUP power for I/Os
- Higher interconnect density
  - 100x easily
- Smaller system footprint

P. Emma et al. IBM J. R&D 2008
M. Bohr, ISSCC 2009 (Intel)
Low Energy & High Performance 3D Stacking Options

Wireless based 3D

Through-silicon via (TSV) based 3D

Miura et al, JSSC 2008

Katti et al, TED 2010
Outline

• Introduction: Overview of the Interconnect Problem

• 3D Stacking Technologies for Electronics
  – Accounting Signal, Power, and Thermal Interconnects

• 3D Stacking of CMOS and MEMS/sensors
  – Novel interface interconnects

• Conclusion
Impact of Novel Cooling on Power Dissipation

\begin{align*}
\text{(1)} \quad & P = \frac{1}{R_{\text{thermal}}} (T - T_{\text{ambient}}) \\
\text{(2)} \quad & P = aC_{\text{total}} \left[ V_{dd}(T) \right]^2 f + N_{\text{gates}} V_{dd}(T) I_{\text{leak}} e^{-\frac{V_t(T) + \Delta V_t}{nkT/q}}
\end{align*}

Sekar et al., IITC 2008

<table>
<thead>
<tr>
<th></th>
<th>Freq.</th>
<th>Power</th>
<th>Temp.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air cooling: ~0.6 °C/W</td>
<td>3 GHz</td>
<td>102 W</td>
<td>88 °C</td>
</tr>
<tr>
<td>Advanced Cooling: ~0.25 °C/W</td>
<td>3 GHz</td>
<td>83 W</td>
<td>47 °C</td>
</tr>
</tbody>
</table>

• Thermal ‘ancillary technologies’ are critical to minimizing power dissipation and increasing reliability and performance
Experimental Results

Microfluidic Channels

Electrical TSVs

X Ray Image

Optical Microscope Image

J. Zaveri et al., *IMAPS* 2009
Assembly of Air-Gap C4 I/Os

SEM

X-ray

Air-gap Fluidic

Si Substrate

Pressure
100g,

Heat
235 °C

X-ray image of C4 Fluidic I/Os after assembly

C. King, et al. ECTC 2010
Cross-sectional Images After Assembly

- 100 μm diameter fluidic TSVs
- 45 μm tall C4 fluidic I/Os
- 47 μm tall electrical I/Os

C. King et al., ECTC 2010
Fluidic Testing

• Flow rate measured up to 100ml/min

C. King et al., ECTC 2010
Yue Zhang et al.
Chip-Package Co-Design for Power Delivery

Many interdependencies and trade-offs for power delivery:

- **Package P/G plane metal allocation**
- **On-die decap insertion**
- **On-die P/G grid metal allocation**
- **P/G I/O pad allocation**
- **TSV allocation**
3D Problem Needs a 3D Solution: Use of a ‘Decap’ Die

“Decap” die: 100% decap.

$|V_{\text{noise}}| = 400 \text{ mV}$

$|V_{\text{noise}}| = 182 \text{ mV}$

$|V_{\text{noise}}| = 312 \text{ mV}$, 22% reduction

$|V_{\text{noise}}| = 256 \text{ mV}$, 36% reduction

→functions as a local high-frequency energy storage

G. Huang et al., EPEP 2007
But, there is more to the story …

3D Stacking of Electronics and MEMS/sensors
MEMS Market

“The number of different MEMS devices is large and steadily growing”

Table 3: Forecasted growth (in $bn) of MEMS devices

<table>
<thead>
<tr>
<th>Application</th>
<th>2005</th>
<th>2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pressure sensors</td>
<td>3.0</td>
<td>6.0</td>
</tr>
<tr>
<td><em>In vitro</em> diagnostics</td>
<td>0.01</td>
<td>5.0</td>
</tr>
<tr>
<td>Read/write heads</td>
<td>2.0</td>
<td>4.0</td>
</tr>
<tr>
<td>Ink jet print heads</td>
<td>2.0</td>
<td>3.5</td>
</tr>
<tr>
<td>Optical displays</td>
<td>1.0</td>
<td>3.0</td>
</tr>
<tr>
<td>Gyroscopes</td>
<td>0.1</td>
<td>2.0</td>
</tr>
<tr>
<td>Lab-on-a-chip</td>
<td>0.01</td>
<td>2.0</td>
</tr>
<tr>
<td>Drug delivery systems</td>
<td>0.0</td>
<td>1.5</td>
</tr>
<tr>
<td>Inertial sensors</td>
<td>0.2</td>
<td>1.5</td>
</tr>
<tr>
<td>Chemical sensors</td>
<td>0.1</td>
<td>1.0</td>
</tr>
<tr>
<td>Optical switches</td>
<td>0.1</td>
<td>1.0</td>
</tr>
<tr>
<td>RF devices</td>
<td>0.1</td>
<td>1.0</td>
</tr>
<tr>
<td>Microspectrometers</td>
<td>0.02</td>
<td>0.4</td>
</tr>
</tbody>
</table>

*T. Marinis, Strain 2009*
MEMS/Sensor and Need for Electronics

• MEMS/Sensors need electronics
  – Signal conditioning, amplification, analysis, device actuation, etc.

• Challenges for MEMS/electronics monolithic integration:
  – Most of the state-of-the-art foundry wary of preprocessed wafers
  – MEMS last approach gives limited window to MEMS designers
    – Limited processes, materials, and devices
  – Each monolithic process is unique
  – Increases the development time as well as NRE cost
  – Supply voltage
  – However, provides small electrical parasitics

H. S. Yang et al. ECTC 2010
Package Based / Hybrid Integration

• Sensor and CMOS can be manufactured independently
  – Shorter Time-to-Market
  – Less complex process
  – Lower Non-Recurring Expense

• Poor performance of a 2D integration
  – Low Density I/Os
    • Limits # of sensors
  – High Parasitic Loads

BioMEA by Charvet et al. Biosensors and Bioelectronics
Monolithic Integration

• Limits processes and materials for sensors
• Each monolithic process is unique
• Complex process

Performance Benefits
– Signal to noise ratio
– High Density Array
– Low Unit Cost

F. Heer et al., Biosensors and Bioelectronics, 2007
Our Approach: Heterogeneous 3D Integration

• Independent Fabrication of CMOS and Sensors

• Performance benefits of 3D integration

Advanced I/O & 3D bonding

Sensor Die

CMOS IC

R. Ravindran et al. ECTC 2010
H. S. Yang et al. ECTC 2010
Principle of SiNW Sensor Operation

• Nanowire (NW) sensors detect the charge induced due to the presence of charged species bound to their surface

• Surface charge leads to either an accumulation or depletion of carriers

• Analogous to surface potential in a FET controlling depletion depth and the onset of inversion

Figure retrieved from P. Nair et al., IEEE Trans. Elec. Devices 54, 3400 (2007).

Figure retrieved from I. Kimukin et al., Nanotechnology 17, S240 (2006).
Sensor / TSV Integration

- Integration with a high temperature (900°C) SiNW process demonstrated
- SiNW can be fabricated prior to TSV due to CMP-free planarization

H. S. Yang et al. IITC 2010
Mechanically Flexible Interconnects

Flexible interconnects allow low-force and low-resistance connection without damages to the pad or interconnect itself

R. Ravindran et al. ECTC 2010
Flexible Interconnect Fabrication

Mask #1

Photodefinable Sacrificial Polymer

Reflow

H. S. Yang et al. ECTC 2010
Flexible Interconnect Fabrication

Seed Layer Deposition

Copper Electroplating
(Mask #2)

Remove Sacrificial Polymer

H. S. Yang et al. ECTC 2010
Compliance Measurements

Hysistron Triboindenter

Indenter Tip

Displacement vs. Force Graph

MFIs

H. S. Yang et al. ECTC 2010
Compliance vs. Thickness

H. S. Yang et al. ECTC 2010
Elastic Range of Motion

- MFIs were indented 5µm, 7.5µm, 10µm and 20µm
- Stand off height remained 20µm

MFI bounces back even after it has been pressed flat against the substrate

H. S. Yang et al. ECTC 2010
Conclusion

1) Innovation in silicon technology without **REVOLUTIONARY** innovation in silicon ancillary technologies will yield progressively “performance crippled” electronic systems

2) 3D stacking & novel silicon ancillary technologies are key to
   • Advancing computing systems
   • Enabling heterogeneous integration of electronics & MEMS/sensors

3) **Looking forward to collaboration opportunities !!!**