ECE 225
Lecture 13
3-D Integrated Circuits

Prof. Kaustav Banerjee
Electrical and Computer Engineering
University of California, Santa Barbara
Outline

- Motivation for 3-D ICs
- Performance Analysis
- Thermal Analysis
- Technologies for 3-D ICs
- Applications
- Summary
Motivation for 3-D ICs

- Increasing Chip Size and Interconnect Delay
- Limitations of Cu/Low-k
- Limitations of Repeaters
- Effects on Design Process
- Heterogeneous Integration

Interconnect Delay is Increasing (ITRS '99)
Will better materials like copper and low-k dielectrics solve the interconnect problem?
Low-k Dielectrics

By Bruce Meyerson
Associated Press

NEW YORK — IBM Corp. has developed a method for insulating the tiny circuits on a computer chip so they can be placed even closer together, making room for more computing power on each semiconductor.

The new process, which coats the circuits with a chemical polymer instead of silicon dioxide, can either improve a chip's processing speed by as much as 30 percent or halve the amount of power it uses, IBM said yesterday.

The insulation cuts down on magnetic interference, or "crosstalk," between the electrical pulses that travel along the millions of copper circuits that a chip uses to crunch data.

Because the electrical pulses are better shielded from the magnetic fields generated by neighboring pulses, the distance between circuits can be reduced by more than a quarter to 0.13 microns — a measurement 600 times thinner than a human hair.

Thanks to the space savings, chipmakers can double the number of transistors they cram onto a piece of silicon to about 400 million, said Bijan Davari, a vice president of IBM's semiconductor research and development center in East Fishkill, N.Y.

Chips made with the new insulation are expected to be available next year, IBM said.

The first application will be to produce powerful microprocessors for major business systems like those used to run Web sites and communications networks.

Old dielectric SiO$_2$ $K = 4$
New dielectric Silk $K = 2.5$

The difference is not even a factor of 2!
Limit of Low-k Dielectrics (Air: $K = 1$)

Intel’s 14nm Logic Technology (2014 IEDM)

Air-Gap Interconnect Structure
Cu Resistivity: Effect of Scaling

- Effect of Cu diffusion Barrier
  - Barriers have higher resistivity
  - Barriers can’t be scaled below a minimum thickness

- Effect of Electron Scattering
  - e scattering from the surface
  - increases effective resistivity

Problem is worse than anticipated in the ITRS roadmap
Cu Resistivity: Effect of Scaling

$T = 100 \, ^0\text{C}$

- Effective Resistivity ($\mu\Omega\cdot\text{cm}$)
- Technology Node ($\mu$m)
- Year

Cu, $P=0.5$

Al

- $P=0$
- $P=0.5$
- $P=1$
Cu Resistivity: Effect of Scaling

Technology Node (μm)

Repeaters per Longest Global Line

- Ideal resistivity = 1.7 microohm-cm
- ALD Barrier
- IPVD Barrier
Scaling Effects: Maximum Chip Temperature

(S. Im and K. Banerjee, IEDM 2000)

![Graph showing scaling effects on maximum chip temperature and power density.]

- Negligible Change in Power Density (ITRS '99)
- Increase in $T_{\text{max}}$ Due to Joule Heating of Interconnects (FEM Simulation)
Effects on Design Process

- Increasing wire delay causes timing closure problem
- Prevents CAD methodologies to adopt higher levels of abstraction to simplify and accelerate the design process
Can we solve the problem by using more repeaters?
Integrated SoC to Sense, Process, and Collaborate

Difficult in Planar IC Technology
Novel Design Architectures Needed!

Single Layer

2 Layers

Replace horizontal by vertical interconnect

3-D ICs utilize the vertical dimension
3-D ICs: Multiple Active Si Layers

(Banerjee et al., Proc. IEEE, 2001)

- **Advantages**
  - Reduce Interconnect Length by Vertically Stacking Multiple Si Layers....**lower power**
  - Improve Chip Performance
  - Reduce Chip Area
  - Heterogeneous integration possible, e.g., memory, digital, analog, optical, etc.
Performance Analysis Strategy

- Estimate Chip Area
- Estimate Interconnect Delay

Chip Area Determined from Wiring Requirement
Used 50 nm ITRS Data
Wire-Length Distribution

Calculate # of Interconnects of length l

- Conservation of I/O’s
  \[ T_A + T_B + T_C = T_{A-to-B} + T_{A-to-C} + T_{B-to-C} + T_{ABC} \]
  \[ T_A + T_B = T_{A-to-B} + T_{AB} \]
  \[ T_B + T_C = T_{B-to-C} + T_{BC} \]

- Values of T within a block or collection of blocks are calculated using Rent’s rule, e.g.,
  \[ T_A = k (N_A)^p \]
  \[ T_{ABC} = k (N_A + N_B + N_C)^p \]

- Recursive use of Rent’s rule gives wire-length distribution for the whole chip

Rent’s Rule: \[ T = k N^p \]

\( k \) and \( p \) denote the average number of fan-out per gate and the degree of wiring complexity (with \( p = 1 \) representing the most complex wiring network), respectively, and are empirically derived as constants for a given generation of ICs.
Wire Length Distribution For 3-D ICs

- \( T = k N^p = T_1 + T_2 - T_{\text{int}} \)
- \( T_1 = T_2 = k (N/2)^p \)
- \( T_{\text{ext, i}} = T_i - T_{\text{int}}/2 = k 2^{p-1} (N/2)^p \)

\[ K_{\text{eff, int}} = k (1 - 2^{p-1}) \]
\[ K_{\text{eff, ext}} = k 2^{p-1} \]
Wire-Length Distribution

Vertical inter-layer connections reduce metal wiring requirement (50 nm ITRS)
Chip Area Estimation

- Placement of a wire in a tier is determined by some constraint, e.g., maximum allowed RC delay.

- Wiring Area = wire pitch x total length
  \[ A_{\text{req}} = p_{\text{loc}}L_{\text{tot_loc}} + p_{\text{semi}}L_{\text{tot_semi}} + p_{\text{glob}}L_{\text{tot_glob}} \]

- \( L_{\text{tot}} \) calculated from wire-length distribution

- Chip Area = \( \frac{A_{\text{req}}}{\# \text{ of Metal Layers}} \)
2 Active Layer Results

- constant chip frequency, $f_c$
- Less wiring needed
- Around 30% reduction in chip area

![Graph showing chip area comparison between 2D and 3D layers](image)
Improved Performance with Area

- Higher $f_c$ obtained from larger wire pitch
- Results in increase in chip area
- Cannot increase $f_c$ indefinitely
Performance Summary

2-Layer 3-D Chip
More Than Two Active Layers

- Simulation based on data from ITRS (50 nm)
- Further improvement in delay achieved
Thermal Issues

- Silicon
  - ILD1
  - ILD2
  - Gate
- Package
  - Heat Flow
  - Heat Sink
  - Heat Sink

Thermal Issues
Analytical Die Temperature Model

- Temperature Rise of the $j^{th}$ Active Layer in an $n$-layer 3-D Chip

$$\Delta T_j = \sum_{i=1}^{j} \left[ R_i \left( \sum_{k=i}^{n} \frac{P_k}{A} \right) \right]$$

- $n$ = Total Number of Active Layers
- $R_i$ = Thermal Resistance Between Adjacent Layers
- $P_k$ = Power Dissipation in the $k^{th}$ Layer

(Im and Banerjee, IEDM 2000)
Analytical Die Temperature Model

Temperature Rise of the Uppermost (n\textsuperscript{th}) Layer in an n-layer 3-D Chip

\[ \Delta T_j = \sum_{i=1}^{j} \left[ R_i \left( \sum_{k=i}^{n} \frac{P_k}{A} \right) \right] \]

\[ \Delta T_n = \left( \frac{P}{A} \right) \left[ \frac{R}{2} n^2 + \left( R_1 - \frac{R}{2} \right) n \right] \]

- R = Thermal Resistance Between Adjacent Layers
- \( R_1 \) = Package Thermal Resistance

Small n

\[ \Delta T \propto n \]

\[ R_1 >> R \]
3-D Fabrication Technologies

Epitaxial Lateral Overgrowth

Purdue
Solid Phase Crystallization of $\alpha$-Si

- Locally induce nucleation
- Grow laterally, inhibiting additional nucleation
- Build MOSFET in a single grain with SOI like performance

Stanford
Ni Seeded Lateral Crystallization

- Initially fabricate amorphous device
- Ni seeding for simultaneous crystallization and dopant activation
- Low thermal budget ($T_{\text{max}} < 500^\circ\text{C}$)
- Devices on top of a metal line
  - MOSFETs
  - Optical detectors
3-D Fabrication Technologies

Wafer Bonding

Bonding using Glue Layer
- ILD/Al
- Thinned Si
- Polymer

Thermocompression Bonding using Cu Pads
- ILD/Al
- Thinned Si
- Cu

MIT
3D ICs: Implications for Circuit Design

- **Critical Path Layout:** By vertical stacking, the distance between logic blocks on the critical path can be reduced to improve circuit performance.

- **Repeaters:** Some chip area can be saved by placing repeaters (~ 10,000 for high performance circuits) on the higher active layers.
3D ICs: Implications for Circuit Design

- Microprocessor Design: on-chip caches on the second active layer will reduce distance from the logic and computational blocks.

“3-D Shared Memory using Wafer Stacking”

K.W. Lee et al., (Tohoku Univ. Japan) IEDM 2000

- 3-D integration technology --wafer stacking

- Parallel processor system with three memory layers --multiport memory, overcomes bus bottleneck
3D ICs: Implications for Circuit Design

- **Integration of Disparate Technologies Easier**

- **RF and Mixed Signal ICs:** Substrate isolation between the digital and RF/analog components can be improved by dividing them among separate active layers - ideal for system on a chip design.

- **Optical Devices and I/Os** can be integrated on the top layer
3D ICs: Implications for Circuit Design

- **Neuromorphic Vision Chip**: Tohoku Univ.
  
  Achieved an image processing and pattern recognition system with parts of functions of the retina and visual cortex using Si

- **3-D Systems-on-a-Chip**: MIT, Lincoln Labs
  
  3D ring oscillators and back-illuminated 64X64 active pixel sensor arrays with fully parallel A/D conversion

- **3-D Assembly Technology**: North Corp., Japan
Compact Modeling of Through-Si Vias (TSV)


- Increase functionality, performance and integration density...
- “More-than-Moore” technologies

TSVs – Connections between layers
Top Cross-Sectional View of a TSV

- Top Cross-Sectional View of a TSV

---

**Si depletion region**

**Dielectric (SiO$_2$)**

**TSV metal**

- $r_{via}$
- $W_{dep}$
- $t_{ox}$

**Si Bulk (neutral) region**
Compact TSV Modeling for 3-D ICs

- High aspect ratio (>10)
- Long lengths (20-100 μm)

RLGC transmission line model: with consideration of depletion region, AC conduction and eddy current effect

- $C_1$ accounts for the electrostatics, $Y_2$ accounts for the conductance and capacitance of bulk Si, $Z_{metal}$ accounts for the inner impedance of the TSV-metal, $L_{outer}$ and $R_{sub}$ are due to the outer impedance of the TSV-pair
## Carbon Nanotubes for TSV Application

### Why Carbon Nanotubes (CNTs)?

<table>
<thead>
<tr>
<th>Property</th>
<th>Cu</th>
<th>SWCNT</th>
<th>MWCNT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max current density (A/cm(^2))</td>
<td>10(^7)</td>
<td>&gt;1x10(^9)</td>
<td>&gt;1x10(^9)</td>
</tr>
<tr>
<td>Thermal conductivity (×10(^3) W/m-K)</td>
<td>0.385</td>
<td>1.75-5.8</td>
<td>3.0</td>
</tr>
<tr>
<td>Electrical mean free path (nm)</td>
<td>40</td>
<td>&gt;1,000</td>
<td>25,000</td>
</tr>
</tbody>
</table>
Reduced skin effect due to the large kinetic inductance

Impedance (RL) result for Cu, W and CNT TSVs

- SWCNT: D = 1 nm, 1/3 Metallic
- MWCNT: D = 20 nm

Reduced skin effect in CNTs

![Graph showing impedance results for different materials and frequencies.](image)
Impedance (RL) result for Cu, W and CNT TSVs

- **SWCNT**: $D = 1 \text{ nm}$, $1/3$ metallic
- **MWCNT**: $D = 20 \text{ nm}$

### Parameters
- **Metal**: SiO$_2$
- **Si depletion region**: 0.7 nm
- **Metal thickness**: 2.5 nm
- **p-Si**: $10 \Omega \text{-cm}$
- **SWCNT**: 1/3 metallic
- **MWCNT**: 20 nm

### Inductance vs. Frequency

- **Cu**
- **W**
- **SWCNT**
- **MWCNT**

- Frequency (Hz): $10^9$ to $10^{11}$
- Inductance (pH/µm): 0.72 to 0.82

### Circuit Symbols
- $R_{sub}$
- $L_{outer}$
- $2Z_{metal}$
Scaling of TSVs According to ITRS

The dimension of TSVs from ITRS is smaller than the values provided by experimental references.
Comparative Performance Analysis

- Delay not quite dependent on TSV material
- Other properties of TSV materials (thermal, reliability, ...) are more important
- CNT – greater thermal conductivity and current carrying capability
Summary -- TSVs

Importance of MOS effect

Performance among TSV materials; possible application of MWCNTs; CG accuracy more important than RL accuracy

RLGC compact model

Graph: Delay ratio w.r.t. Cu

- W
- SWCNT (Fm=1/3)
- MWCNT

Inverter size

0 50 100 150
1.000 1.002 1.004 1.006