Full-chip Inter-die Parasitic Extraction in Face-to-Face-Bonded 3D ICs

Yarui Peng\textsuperscript{1}, Taigon Song\textsuperscript{1}, Dusan Petranovic\textsuperscript{2}, and Sung Kyu Lim\textsuperscript{1}
Georgia Institute of Technology, Atlanta, GA, USA
\textsuperscript{1}School of ECE, Georgia Institute of Technology, Atlanta, GA, USA
\textsuperscript{2}Mentor Graphics, Fremont, CA, USA
Face-to-face Bonding Structure

- Inter-die capacitance becomes important when die-to-die distance is small, especially for face-to-face (F2F) bonded structures with direct copper bonding.

F2F-bonded 3D IC structure with interconnect parasitics
To analyze the trend in F2F structure, we build a test structure in Raphael with repeated pattern

- Wire dimensions are based on M4 dimensions in a 45nm technology
- Intra-die coupling: AB cap and CD cap
- Inter-die overlap coupling: AC cap and BD cap
- Inter-die fringe coupling: AD cap and BC cap
Die-to-die Distance Impact

- With a closer die-to-die distance:
  - Intra-die cap (AB Cap) decreases due to stronger E-field sharing
  - Inter-die cap increases significantly
  - Inter-die overlap cap (AC cap) increases much more than inter-die fringe cap (AD cap)
With a larger wire-to-wire distance

- Both intra-die coupling (AB cap) and total cap reduces
- Inter-die first increases with larger overlap cap (AC cap) due to weaker E-field sharing then slightly decrease due to smaller fridge cap (AD cap)
Three Ways of Full-chip F2F Extraction

- **Die-by-die extraction**
  - Extract dies separately

- **Holistic extraction**
  - Extract all layers simultaneously

- **In-context extraction**
  - Extract each die separately but aware of a few neighboring die layers
Die-by-die extraction is the straight-forward flow currently enabled by many CAD tools:

- Assumes each die can be extracted separately
- Ignores all parasitic between dies
- Accurate when dies are separated far or have a ground layer in between
- Holistic extraction takes all layers into consideration and it introduces more CAD and LVS complexity

Technology generation:
- 2D technology
- 3D F2F technology generator
- 3D technology
- Characterization
- 3D extraction rule
- 3D extraction
- 3D Holistic parasitics

Design generation:
- 2D Bot die
- 3D design convertor
- 3D Bot die
- Top-design generator
- 3D Top die
- 3D top-level design
- Assemble design
- 3D holistic design

Holistic extraction flow
Die-by-die vs. Holistic Extraction

- **Die-by-die uses the same metal stack as 2D technology**
  - Enables reuse of existing DRC, LVS and PEX rule decks
- **Holistic extraction needs to rebuild rule decks**
  - All original and derived layers and device renamed and remapped
  - Need technology recalibration

![Layer Structure Diagram]

- **Same as 2D Technology**
  - M1
  - M2
  - M3
  - M4

- **Die-by-die layer structure**

- **Holistic layer structure**
  - M1B
  - M2B
  - M3B
  - M4B
  - M1T
  - M2T
  - M3T
  - M4T
Holistic Design Example

- By assembling of individual dies, we are able to create a holistic design which contains all metal layers
In-context extraction takes in a few metal layers from the neighboring die as interface layers

- Keeps most of inter-die coupling and remains accurate
- Reduces CAD complexity and compatible with current tool flow

In-context Extraction Flow

Technology generation
- 2D technology
- 2D library
- In-context technology generator
- Characterization
- In-context extraction rule
- Bot-die extraction
- Double-counting and surface layer handling
- Bot in-context parasitics

Design generation
- 2D Bot die
- 2D Top die
- In-context design generator
- Bot in-context die
- Top in-context die
- Top-die extraction
- Bot in-context parasitics
• **In-context technology can be calibrated incrementally**
  – Base layer calibration results can be derived from existing rule decks
  – The surface layer in the in-context extraction is defined as the layer furthest from the substrate

In-context layer structure
(with one interface layer from each die)
In-context design only needs additional routing information from the neighbor die

- Enables much simpler rule deck generation
• With in-context extraction, capacitance on interface layers are double-counted
  – A simple solution is to halve all caps from interface layers in SPEF files
• Surface layer only sees one neighboring layer
  – Introduce large error with less E-field sharing
• Note each layer is not the surface layer in both in-context dies
  – E.g., M3T is the surface layer in bottom die but not in top die
• Surface layer correction based on weighted average
  – Use a weighted average for caps on interface layers
  – Larger weight for layers farther from the surface

Wrong ×

Correct √

Bottom die

Top die
For each layer, we define:
- D: distance to the surface
- R: ratio between D values in the bottom and top in-context die

Example (two interface layers per die)
The surface correction weight of a capacitor is the product of R ratios of both layers it connects to, normalized to 100%:

- The R ratio of ground layer is defined as 1:1
- A ground cap on M4T: $R_{M4T} \times R_{gnd} = 1:2 \times 1:1 = 33\%$ (bot) : 67\% (top)
- A cap between M4T and M4B: $R_{M4T} \times R_{M4T} = 1:2 \times 2:1 = 50\%$ (bot) : 50\% (top)
Sample FFT Design in F2F

- A 64-point FFT with 38K gates and 330 F2F vias is implemented
Die-by-die vs. Holistic Extraction

- With a 1um F2F via height, die-by-die extraction underestimates coupling capacitance significantly
  - Especially for layers close to the other die

<table>
<thead>
<tr>
<th>Layer</th>
<th>M1B</th>
<th>M2B</th>
<th>M3B</th>
<th>M4B</th>
<th>M4T</th>
<th>M3T</th>
<th>M2T</th>
<th>M1T</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Holi</td>
<td>26.2</td>
<td>949</td>
<td>1808</td>
<td>3703</td>
<td>3089</td>
<td>1755</td>
<td>1013</td>
<td>38.2</td>
<td>12381</td>
</tr>
<tr>
<td>D-D</td>
<td>20.1</td>
<td>856</td>
<td>1620</td>
<td>1955</td>
<td>1413</td>
<td>1399</td>
<td>747</td>
<td>21.2</td>
<td>8032</td>
</tr>
<tr>
<td>Err</td>
<td>-6.06</td>
<td>-93.4</td>
<td>-187</td>
<td>-1747</td>
<td>-1676</td>
<td>-356</td>
<td>-266</td>
<td>-17.0</td>
<td>-4349</td>
</tr>
<tr>
<td>Err%</td>
<td>-23%</td>
<td>-9.8%</td>
<td>-10%</td>
<td>-47%</td>
<td>-54%</td>
<td>-20%</td>
<td>-26%</td>
<td>-45%</td>
<td>-35%</td>
</tr>
</tbody>
</table>

Total coupling capacitance of each layer
• Inter-die coupling occupies a large portion of total coupling cap
  – Especially when dies are close and few metal layers are used

<table>
<thead>
<tr>
<th>Layer</th>
<th>M1B</th>
<th>M2B</th>
<th>M3B</th>
<th>M4B</th>
<th>M4T</th>
<th>M3T</th>
<th>M2T</th>
<th>M1T</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1B</td>
<td>5.76</td>
<td>3.03</td>
<td>17.1</td>
<td>0.13</td>
<td>0.03</td>
<td>0.14</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>M2B</td>
<td>3.03</td>
<td>381</td>
<td>147</td>
<td>396</td>
<td>18.6</td>
<td>0.69</td>
<td>2.58</td>
<td>0.01</td>
</tr>
<tr>
<td>M3B</td>
<td>17.1</td>
<td>147</td>
<td>1261</td>
<td>231</td>
<td>9.9</td>
<td>140</td>
<td>0.72</td>
<td>0.28</td>
</tr>
<tr>
<td>M4B</td>
<td>0.13</td>
<td>396</td>
<td>231</td>
<td>1826</td>
<td>1184</td>
<td>18.6</td>
<td>46.7</td>
<td>0.12</td>
</tr>
<tr>
<td>M4T</td>
<td>0.03</td>
<td>18.6</td>
<td>9.9</td>
<td>1184</td>
<td>1311</td>
<td>196</td>
<td>369</td>
<td>0.28</td>
</tr>
<tr>
<td>M3T</td>
<td>0.14</td>
<td>0.69</td>
<td>140</td>
<td>18.6</td>
<td>196</td>
<td>1226</td>
<td>148</td>
<td>25.3</td>
</tr>
<tr>
<td>M2T</td>
<td>0.00</td>
<td>2.58</td>
<td>0.72</td>
<td>46.7</td>
<td>369</td>
<td>148</td>
<td>442</td>
<td>4.63</td>
</tr>
<tr>
<td>M1T</td>
<td>0.00</td>
<td>0.01</td>
<td>0.28</td>
<td>0.12</td>
<td>0.28</td>
<td>25.3</td>
<td>4.63</td>
<td>7.54</td>
</tr>
</tbody>
</table>

Breakdown of holistic extraction
In-context vs. Holistic Extraction

- Our in-context extraction with double counting and surface layer corrections matches very well with holistic extraction
  - Using two interface layers from each die

<table>
<thead>
<tr>
<th>Layer</th>
<th>Ground capacitance</th>
<th>Coupling capacitance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M1B</td>
<td>M2B</td>
</tr>
<tr>
<td>Holi</td>
<td>1136</td>
<td>6588</td>
</tr>
<tr>
<td>In-C</td>
<td>1137</td>
<td>6583</td>
</tr>
<tr>
<td>Err</td>
<td>1.10</td>
<td>-4.20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Layer</th>
<th>Coupling capacitance</th>
<th>Coupling capacitance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M1B</td>
<td>M2B</td>
</tr>
<tr>
<td>Holi</td>
<td>26.2</td>
<td>949</td>
</tr>
<tr>
<td>In-C</td>
<td>26.3</td>
<td>950</td>
</tr>
<tr>
<td>Err</td>
<td>0.15</td>
<td>0.81</td>
</tr>
</tbody>
</table>
Interface Layer Impact

- More interface layers helps improve accuracy
  - With two interface layers per die gives a good tradeoff

<table>
<thead>
<tr>
<th>Layer</th>
<th>M1B</th>
<th>M2B</th>
<th>M3B</th>
<th>M4B</th>
<th>M4T</th>
<th>M3T</th>
<th>M2T</th>
<th>M1T</th>
<th>Total</th>
<th>Err%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Holistic</td>
<td>26.2</td>
<td>949</td>
<td>1808</td>
<td>3703</td>
<td>3089</td>
<td>1755</td>
<td>1013</td>
<td>38.2</td>
<td>12381</td>
<td>-</td>
</tr>
<tr>
<td>In-C:1</td>
<td>26.1</td>
<td>953</td>
<td>1701</td>
<td>3708</td>
<td>2994</td>
<td>1604</td>
<td>994</td>
<td>37.8</td>
<td>12018</td>
<td>-2.93%</td>
</tr>
<tr>
<td>In-C:2</td>
<td>26.3</td>
<td>950</td>
<td>1803</td>
<td>3679</td>
<td>3058</td>
<td>1734</td>
<td>1001</td>
<td>38.0</td>
<td>12287</td>
<td>-0.76%</td>
</tr>
<tr>
<td>In-C:3</td>
<td>26.2</td>
<td>949</td>
<td>1794</td>
<td>3671</td>
<td>3057</td>
<td>1745</td>
<td>1012</td>
<td>38.2</td>
<td>12292</td>
<td>-0.72%</td>
</tr>
</tbody>
</table>

- Our weighted methods improves in-context extraction accuracy

<table>
<thead>
<tr>
<th>Layer</th>
<th>M3B</th>
<th>M4B</th>
<th>M4T</th>
<th>M3T</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Holi</td>
<td>1808</td>
<td>3703</td>
<td>3089</td>
<td>1755</td>
<td>10354</td>
</tr>
<tr>
<td>Original</td>
<td>3069</td>
<td>6779</td>
<td>5781</td>
<td>3522</td>
<td>19151</td>
</tr>
<tr>
<td>Halved</td>
<td>1618</td>
<td>3611</td>
<td>3082</td>
<td>1849</td>
<td>10159</td>
</tr>
<tr>
<td>Weighted</td>
<td>1803</td>
<td>3679</td>
<td>3058</td>
<td>1734</td>
<td>10272</td>
</tr>
</tbody>
</table>
In-context extraction captures inter-die aggressors, provides better accuracy in full-chip analysis

- Especially for 3D nets which communicates across dies
Full-chip Analysis Results

- Full-chip analysis also shows non-negligible impact from inter-die capacitance, especially on noise results and 3D nets
  - Die-by-die extraction underestimates delay, power and noise
  - In-context extraction gives much more accurate results

<table>
<thead>
<tr>
<th>Primetime measurement</th>
<th>Holi</th>
<th>D-D</th>
<th>Err%</th>
<th>In-C</th>
<th>Err%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Longest path delay (ns)</td>
<td>3.90</td>
<td>3.66</td>
<td>-6.2%</td>
<td>3.81</td>
<td>-2.3%</td>
</tr>
<tr>
<td>3D nets switching power (mW)</td>
<td>1.05</td>
<td>1.01</td>
<td>-3.5%</td>
<td>1.04</td>
<td>-0.5%</td>
</tr>
<tr>
<td>Total switching power (mW)</td>
<td>12.1</td>
<td>11.9</td>
<td>-1.7%</td>
<td>12.0</td>
<td>-0.8%</td>
</tr>
<tr>
<td>Total coupling cap on 3D nets (fF)</td>
<td>4.37</td>
<td>2.96</td>
<td>-32%</td>
<td>4.19</td>
<td>-4.1%</td>
</tr>
<tr>
<td>Total wire cap on 3D nets (fF)</td>
<td>10.8</td>
<td>9.35</td>
<td>-13%</td>
<td>10.6</td>
<td>-1.8%</td>
</tr>
<tr>
<td>Avg aggressor # on 3D nets</td>
<td>285</td>
<td>200</td>
<td>-30%</td>
<td>277</td>
<td>-2.8%</td>
</tr>
<tr>
<td>Max noise on 3D nets (mV)</td>
<td>41.3</td>
<td>30.40</td>
<td>-26%</td>
<td>38.8</td>
<td>-6.1%</td>
</tr>
</tbody>
</table>
• We studied impacts of E-field sharing in F2F structure
• We showed inter-die coupling cannot be ignored in F2F-bonded 3D ICs, especially with few metal layers and close die-to-die distance
• We implemented and compared three extraction methods with full-chip analysis results
  – Die-by-die extraction underestimates total coupling capacitance
  – Holistic extraction is able to capture all inter-die coupling at the cost of high complexity
  – Our first-of-its-kind in-context extraction is highly accurate, and captures most E-field interactions across dies