A Software-Supported Methodology for Exploring Interconnection Architectures Targeting 3-D FPGAs

Kostas Siozios, Vasilis F. Pavlidis*, and Dimitrios Soudris

School of Electrical and Computer Engineering
National Technical University of Athens
Athens, Greece
{ksiop, dsoudris}@microlab.ntua.gr

*LSI-EPFL
1015 Lausanne, Switzerland
vasileios.pavlidis@epfl.ch
Presentation Outline

- Design problems of existing reconfigurable architectures
- 3D integration technology
- 3D reconfigurable architectures (3D-FPGAs)
- CAD tools for supporting 3D architectures
- Design more efficient 3-D FPGAs
- Conclusions
The interconnection problem
The interconnection problem

A Software-Supported Methodology for Exploring Interconnection Architectures Targeting 3-D FPGAs
Interconnection: A Big Challenge
Impact of interconnection to performance

Source: ITRS 2003

High-Performance Computing requires even more powerful architectures but....
Historical driving forces

- 1971: 4004 Processor, 2300 Transistors
- 1978: 8008 Processor, IBM PC
- 1985: i386 Processor, 32-bit
- 1993: Pentium Processor, 3.1M transistors
- 2006: 561M Transistors

A Software-Supported Methodology for Exploring Interconnection Architectures Targeting 3-D FPGAs
Projecting the future

A Software-Supported Methodology for Exploring Interconnection Architectures Targeting 3-D FPGAs
Existing way to design chips

2D system implementation (many chips)
Features of 3D architectures

- Integration of heterogeneous technologies
  - SRAM, DRAM, FLASH, RF technologies
  - Combination of sensors, battery, etc

- Alleviate problems related to interconnection

- Achieve performance improvements
  - More functionality can be implemented

- Reduce chip dimensions
  - $1\text{mm}^3 > 50\text{mm}^2$ silicon!!!

3D system implementation (ONE chip)

Source: IMEC
Three-Dimensional (3D) integration

Think Different
Design 3D
Open issues

- Design space exploration is essential for efficient 3D devices
- CAD tools that facilitate the design of 3D circuits are required
  - Up to date there are only a few academic approaches

- **Goal of this work:** Study & select the appropriate interlayer connectivity for target 3D FPGA devices

  Explore the number and spatial location of vertical interconnects on different layers
Modeling 3D FPGAs

- It is the result of applying 3D integration onto FPGAs
- 3D FPGAs employ multiple layers (logic, memory, etc) interconnected into three dimensions
- The interlayer communication is realized with extending SBs, in order to support connections to other layers
  - 2D SB: provide connections to the same layer
  - 3D SB: provide connections also to rest layers
Modeling 3D FPGAs

A Software-Supported Methodology for Exploring Interconnection Architectures Targeting 3-D FPGAs

Legend
- Logic block
- 3D SB
Modeling 3D FPGAs

This is the existing way for designing 3D FPGAs

A Software-Supported Methodology for Exploring Interconnection Architectures Targeting 3-D FPGAs
The methodology for designing 3D FPGAs

**Step 1**
- Application netlist
  - Partitioning
  - Assignment
  - Layer ordering

**Step 2**
- Acceptable partitioning?
  - Yes
    - Select bonding technology
  - No
    - Select distribution of interlayer connections (i.e., 3D SBs)

**Step 3**
- Acceptable P&R?
  - No
    - Floorplan
      - Placement
      - Routing
  - Yes

**Step 4**
- Performance, power, area metrics
  - Acceptable
    - Application netlist
      - Partitioning
      - Assignment
      - Layer ordering
  - No

---

A Software-Supported Methodology for Exploring Interconnection Architectures Targeting 3-D FPGAs
2D/3D MEANDER Framework

Application description (HDL)

Synthesis

Technology mapping

2D or 3D

P&R (EX-VPR)

Power (PowerModel)

Partitioning (3DPart)

P&R (3DPRO)

Power Estimation (3DPower)

2D MEANDER

3D MEANDER

Public available at:
http://proteas.microlab.ntua.gr

Qualitative comparison among P&R tools

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Measure delay</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Measure wirelength</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Measure power consumption</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Supported switch boxes</td>
<td>Subset Wilton</td>
<td>N/A</td>
<td>Designer defined</td>
</tr>
<tr>
<td></td>
<td>Universal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heterogeneous architectures</td>
<td>No</td>
<td>N/A</td>
<td>Yes</td>
</tr>
<tr>
<td>Architecture-level exploration</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Multiple 3D technologies</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Timing-aware P&amp;R</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Power-aware P&amp;R</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Part of a complete framework</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Public available</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Results of our methodology

- The tools rather than providing the optimal solution, they also give all the Pareto solutions that trade-off the design parameters.
- Supports flexible cost functions
- Based on them, the designer selects the one that meet the application’s constraints.
Search space exploration

"Optimal" architecture
Selected architecture with 3 layers
Architecture space with similar performance

Normalized Delay

Normalized Energy Consumption

2 Layers (Wirebonding)  2 Layers (TSV)  2 Layers (F2F)
3 Layers (Wirebonding)  3 Layers (TSV)
4 Layers (Wirebonding)  4 Layers (TSV)

CLMA benchmark
#4-LUT = 8073
#Network = 8135
Modeling 3D FPGAs

This is the existing way for designing 3D FPGAs

but...
Demand for interlayer connections

Layer 1
Region 1 (Low demand)
Region 2 (High demand)

Layer 2
Region 1 (Low demand)
Region 2 (High demand)

Layer 3
Region 1 (Low demand)
Region 2 (High demand)
Demand for interlayer connections

- There is no uniform demand for inter-layer connections
  - Higher demand on the center of each layer
  - Higher demand in the middle of the 3D stack
2D/3D Switch Boxes

- 2D and 3D SBs occupy different number of transistors
  - 2D SB: $6 \times W$ transistors
  - 3D SB: $15 \times W$ transistors
- Up and Down connections are 3D bonding wires
- Considerable performance, power, area and fabrication gains if we don’t waste them
Distribution of 3D SBs (proposed)

Legend
- Green: Logic block
- Yellow: 3D SB

Region 1
Region 2

Increased interlayer connectivity
Reduced interlayer connectivity
Distribution of 3D SBs (proposed)

Increased interlayer connectivity

Reduced interlayer connectivity

Legend
- Logic block
- 3D SB
- 2D SB

Region 1

Region 2
Search space exploration

CLMA benchmark
#4-LUT = 8073
#Network = 8135

A Software-Supported Methodology for Exploring Interconnection Architectures Targeting 3-D FPGAs

Normalized Delay

“Optimal” architecture
Selected architecture with 3 layers (K=30%)

Delay: +2% & Energy: +3 %

Architecture space with similar performance

Normalized Energy Consumption

2 Layers (Wirebonding)  2 Layers (TSV)  2 Layers (F2F)
3 Layers (Wirebonding)  3 Layers (TSV)
4 Layers (Wirebonding)  4 Layers (TSV)
Percentage of fabricated interlayer connections that are actually utilized

Almost similar percentages of utilized interlayer connections

A Software-Supported Methodology for Exploring Interconnection Architectures Targeting 3-D FPGAs
Comparison results in terms of EDP

- Gains from 3-D FPGAs as compared to 2-D architectures:
  - K=100% [4]: EDP reduction by 61%
  - Proposed (K=30%): EDP reduction by 55%

- But proposed architecture contains about 70% fewer TSVs
Conclusions

- We study constraints introduced by the spatial variation of interlayer connections.
- Our 3-D FPGAs have 70% fewer TSVs (as compared to existing design approaches) without penalty in delay, power consumption and silicon area.
- New CAD algorithms/tools were proposed to software support the new 3D architectures.
Thank you