3D IC Design and CAD Challenges

Ruchir Puri
IBM T J Watson Research Center
Yorktown Heights, NY 10598
Precedent for 3D Integration: When Real Estate Becomes Pricey

Vertical Integration isn’t new!

NYC Office Inventory, Rent, and Skyscrapers

Data courtesy of Richard Persichetti
Grubb & Ellis, New York, NY
A fundamental shift in technology has occurred in terms of interconnects

- Mx resistance is increasing at an alarming rate
- High resistance drives repeater challenges
  - 130nm-2000, 90nm-20K, 65nm-193K, 45nm ~2-3M
- Costs us lots of power with buffers being the leakiest and accounting for > 50% of logic leakage.
3D Integration Impact on wire length

- Significantly reduced number of global long wires even if total wire length may not be dramatically reduced
  - Drives lower repeater count
    - Latency and Power improvements
Increasing area overhead to achieve performance

1) Architecture overhead increasing area of die

2) Accessible portion of chip over normalized cycle time is decreasing generation over generation

3) Deeper Pipes are decreasing delay per cycle
Trends: Uni-processor Cross-Chip Latency

- Perfect Storm: (a) Wire non-scaling; (b) die size growth; (c) Shorter FO4 stages
- Power Cost of Cross-Chip Latency increase

"Challenges for Computer Architects. Breaking the Abstraction Barriers", Saman Amarasinghe

From the SIA Roadmap

- From the SIA

- Accessible Die Area
- Fixed Die Area
- Scaled Area

- 2nd active layer ??

- Bernstein, IBM
Growing cache size to get performance

- Growing data sets will increasingly stress cache size
- Multi-core floor planning and SRAM concerns will slow cache size growth to maintain manageable chip size

Implications

- Cache size and bandwidth has significant impact on system performance
  - Bring as much cache memory on chip as possible
- Multiple core architecture permits high flexibility in power management
  - Transition from multiple all-purpose core design to heterogeneous many core architecture allows further flexibility in power performance trade-off
- With increasing number of cores and on-chip cache capacity bandwidth is even more important
  - 3D integration is the key future enablement technology for this paradigm
3D Design Enablement

- 3D Design needs 3D Place and Route and sophisticated 3D standard cell placement algorithms

- 3D total wire length advantages are debatable
  - Even in 2D the design is partitioned and the communication between those partitions is minimized.
    - Lot of total wire length comes from local communication

- 3D needs to be exploited at Architecture Level with Architecture Level Physical Planning
  - Research must focus in this area.
  - Co-design and Analysis of Chip Performance, Power, Thermal, and Physical View.
A Case for 3D Architectural Planning

- Mostly, total wire length advantages reported before do not consider the area impact of vertical vias.

- **Wire length isn’t the most important criterion**
  - As long as you can route the design.

- **Latency and power is**

- Timing Critical paths are between and inside “large objects”

- A simplistic flow would not be successful

- It’s an architectural issue – 3D floorplanning
Via Density and Application Space

Interlayer Via Density (number of vias/mm²)

- Transistor Level: 0.25u via
- Macro Level: 0.75u via
- Unit Level: 2.0u via
- Core Level: 4.0u via

3D Design partitioning Level
Is it getting hot in here? 3D Thermal Analysis

- Construct thermal profile based on thermal conductivity, thickness of layers
- Point power sources determined by unit activity, layout.
- Field Solver solution
- Instantaneous and average power density analysis for high performance and low power settings
Chip Performance is limited by global paths at core/unit level. For significant performance improvement, 3D integration at core or unit level is desirable.

Sweet Spot

Area efficiency (F₀/(F₀+contact))

12%  15%  18%  21%

Interlayer Via Density (number of vias/mm²)

100000  10000  1000

12%  15%  18%  21%

12%

Chip Performance is limited by global paths at core/unit level. For significant performance improvement, 3D integration at core or unit level is desirable.

Sweet Spot

Area efficiency (F₀/(F₀+contact))

12%  15%  18%  21%

Interlayer Via Density (number of vias/mm²)

100000  10000  1000

12%  15%  18%  21%

12%
3D Design Exploration

- 3D interconnects have significant impact on critical path latencies
  - Linking the tools for Architectural Analysis and 3D chip integration physical planning will be essential to study benefit of 3D for a given design
  - Impact of physical planning decisions on architecture CPI/power & vice-versa.

CPI analysis, Power analysis

Standard format to exchange path latency data

Micro-Architecture

Floorplan: Area
Interconnect, Power
3D Early Analysis Framework

Performance, Functional Model

Power Analysis

3D Chip Floorplan

Graphic Front-End

Interconnect Analysis

Block Diagram

Chip Integration

Implementation

External IP

Industry Standard Models

Thermal Analysis

3D Physical View

Accelerators

ASIC

L3

MC

IO

CPU-Tahoe

CPU-no Tahoe

Rx-Tahoe

Rx-no Tahoe

Packet size [bytes]

Power [W]

Number of connections

0 2 6 10 20 30 50 100

0 20 40 60 80 100

0.0 0.4 0.8 1.2 1.6 2.0

64 128 256 512 1024 2048 4096 8192

CPU-Tahoe

CPU-no Tahoe

Rx-Tahoe

Rx-no Tahoe

Industry Standard Models

SEMATECH/ACM Thermal and Design Issues in 3D ICs, 2007

© 2007 IBM Corporation
CAD Challenges, 3D IC Design Enablement

- **Early analysis tools for 3D IC designs**
  - To justify 3D cost overhead, essential to study benefit of 3D early in the design cycle
    - requires strong link between architectural level analysis tools and 3D physical planning tools.
  - early analysis tools study tradeoffs between layer count, power density, and performance.

- **Physical Design and Analysis tools for 3D IC designs**
  - placement and routing tools to be aware of 3D technology
    - Note: This is different than algorithms for 3D Place & Route
  - Increased circuit density of 3D IC leads to increased power density: may result in intermediate layer thermal problems
    - Development of “thermally aware design tools” at all levels
    - Partition design to place highly loaded, active components in layer close to the heat-sink
3D Test Accommodation

- 3D Assembly conceptually is a Multi-chip Module (MCM); each layer same as “Known-Good-Die”
- Conforming to MCM Design convention and logic partitioning simplifies 3D Test, assures coverage
- Can accommodate Memory (ABIST) and Logic (LSSD) on the same layer

3D Test Segments
1) Individual Chips / Levels
2) Vertical Connection
3) Host Carrier
4) Completed Assembly
3D Defect Limited Yield Response

Goes Quadratically with die size, linearly with chip count at given die size.
Direct benefit from 3D
  some yield loss for vertical via,
  some yield gain from density
Technology compatible with Known-Good-Die practices

Rule-of-thumb: Cutting die area by 2X improves productivity by 4X
3D Design and CAD Challenges: Summary

- 3D Integration postpones interconnect-related limitations to extend classic scaling
  - Substantial challenges to 3D include Heat Dissipation, Test, and Yield

- Success in 3D requires EDA enablement
  - Base tools such as extraction, LVS, DRC.
  - A simple view of standard cell based 3D Placement and Routing algorithms and tools will not be successful
  - Sweet spot of 3D partitioning lies at Unit level and beyond.

- 3D Architectural Level Physical Planning and Early analysis Tools are needed that can concurrently analyze the impact of 3D technology on performance, physical, power and thermal views of the design.
Acknowledgments

- The author wishes to recognize the following individuals for helpful inputs and discussions:

  Michael Ignatowski (IBM)
  Kerry Bernstein (IBM)
  Wilfried Haensch (IBM)
  Anna Topol (IBM)
  Albert Young (IBM)
  Koushik Das (IBM)
  Phil Emma (IBM)
  Jeff Burns (IBM)