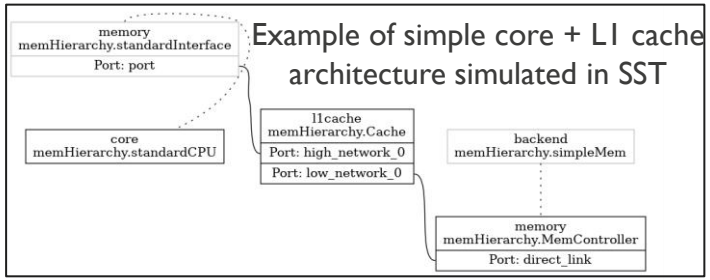
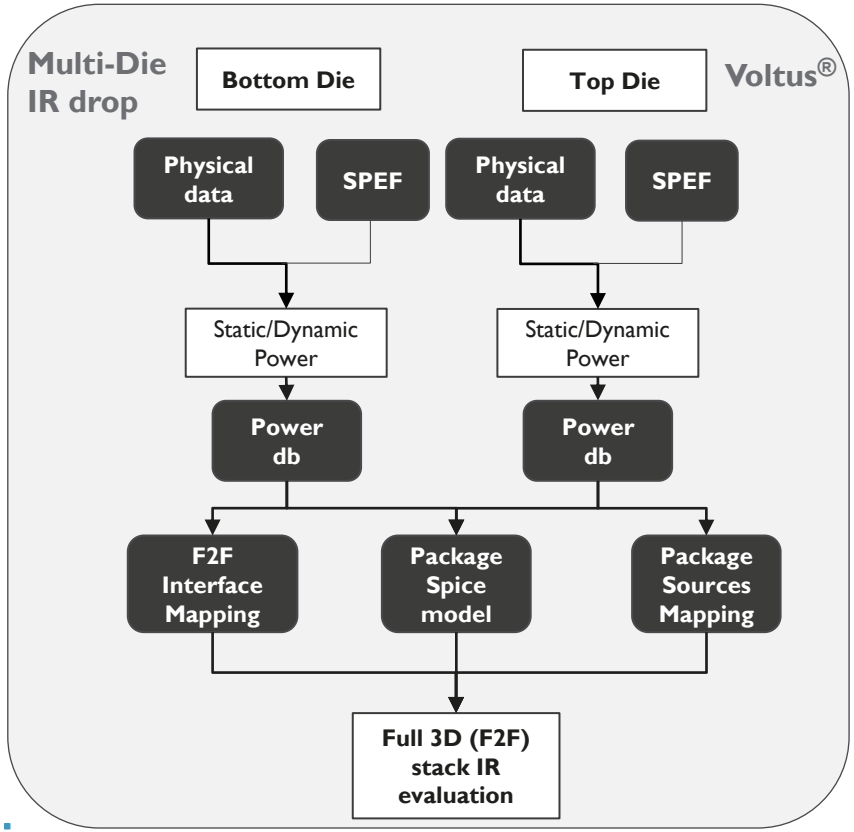
Summary

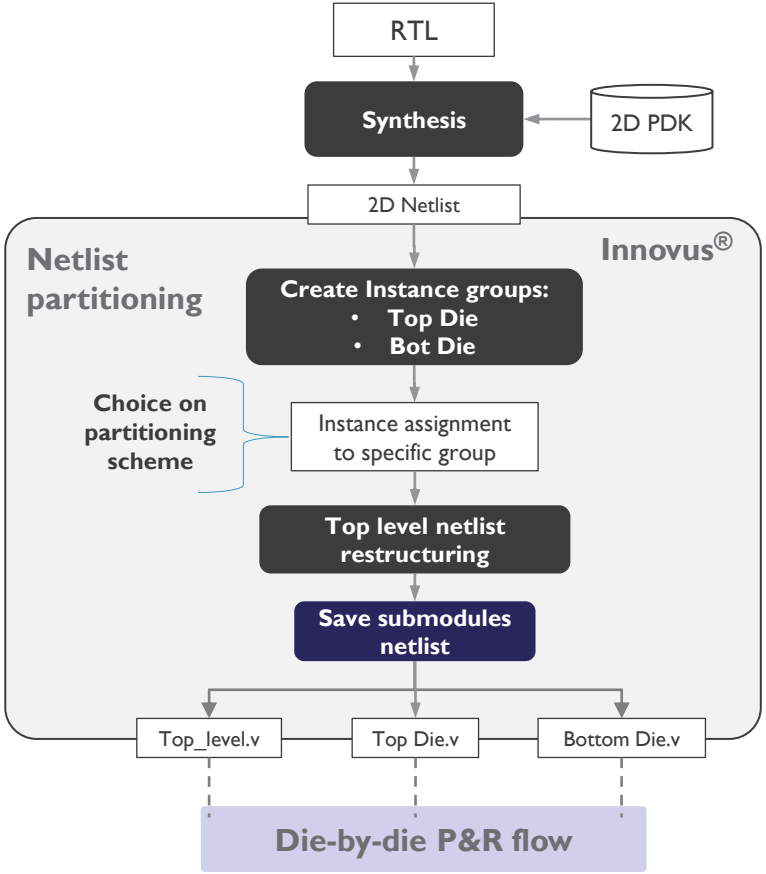
- Worked on novel methodologies for 3D-IC block level design and signoff
  - Fine-Pitch 3D integration with Hybrid bonding on HPC SoC (April 2021 PTW – J107) and AI SoC (April 2022 PTW – D116)
- Evaluation of Backside PDN for High-Performance Computing
  - Block-level Evaluation and Optimization of Backside PDN for High-Performance Computing at the A14 node | IEEE Conference Publication | IEEE Xplore
- PPA comparison at SoC level with different CMOS technologies and DTCO/STCO scaling boosters
  - Nanosheet and Forksheet width modulation (October 2021 PTW – Y112)
  - Backside signal routing (April 2021 PTW – Y106)
- Computer architecture and system modeling and exploration
  - SST Simulator, Cadence Helium
- Several supervised PhD and Master student projects
  - Application-Aware SoC optimization
  - Chiplets-based SoC cost model (VLSI 2024 submitted)
- RTL Logic design for a compute block (current project)



Examples of developed design methodologies for 3D ICs

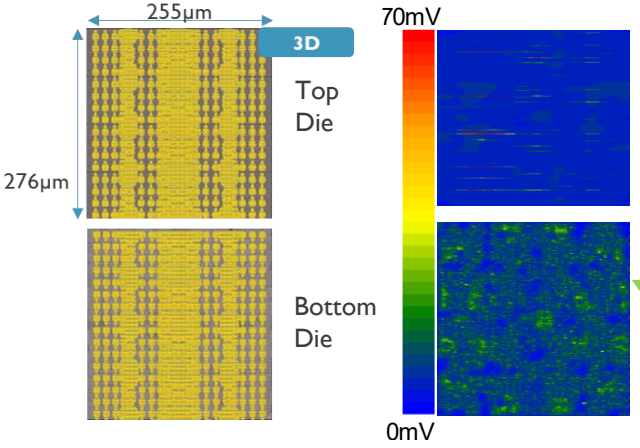
All built within commercial EDA tools



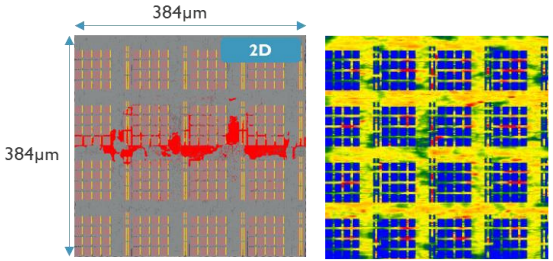


Methodologies application on a memory-dominated SoC for AI

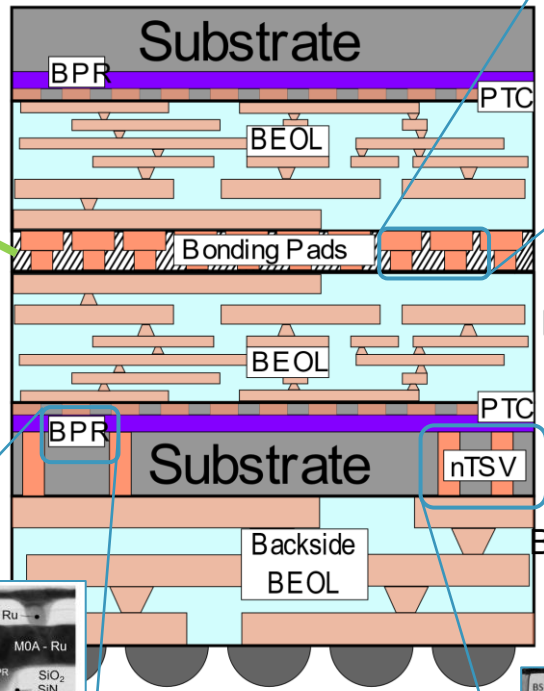
System-level evaluation of 3D power delivery network at 2nm node (spiedigitallibrary.org)

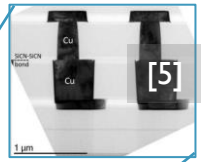


From 3D PNR to 3D IR drop, compared with standard 2D



Full 3D stack

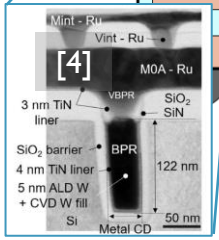


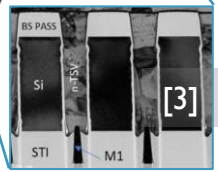


Top Die

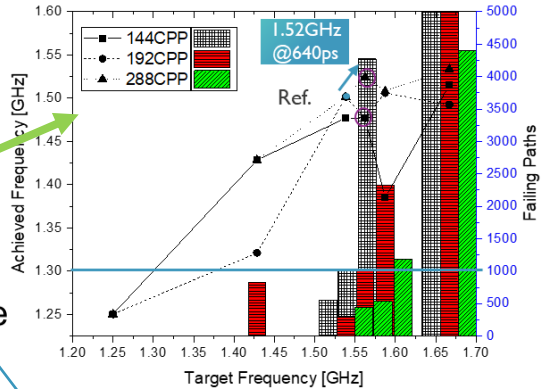
Bottom Die

Backside PDN





How the n-TSV pitch influences the system achieved frequency (line) and number of failing paths (columns)

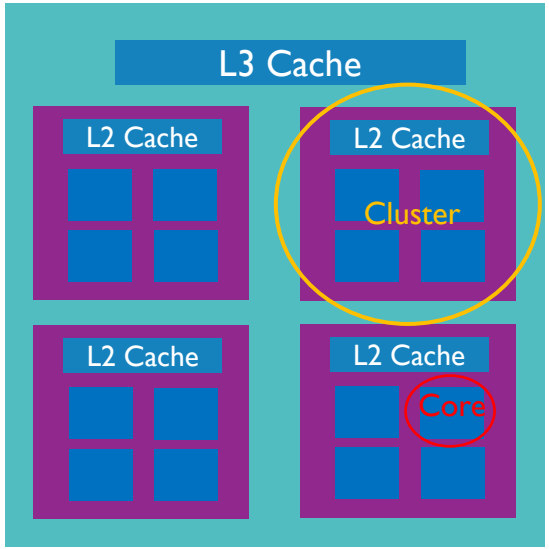


Apr 2022 PTW - D116

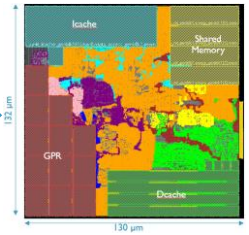
Application-Aware SoC Optimization

Proposed project, ongoing with 2nd back-to-back student

Vortex Platform OpenSource GP-GPU (gatech.edu)

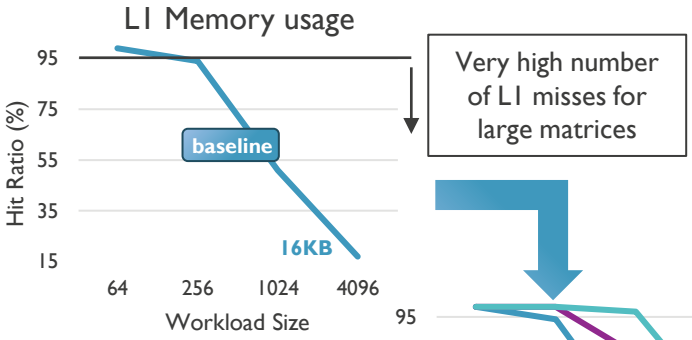


Basic configuration
- 1 cluster with 1 core
- 4 warps and 4 Threads
- 16KB L1 cache
- No L2 or L3 caches



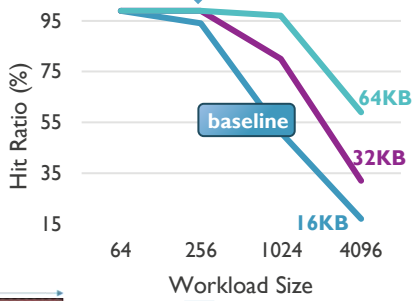
We have the PNR of a block with logic and memory, but does it make sense from the application point of view?

Representative Matrix-Matrix multiplication workload and RTL simulation



Very high number of L1 misses for large matrices

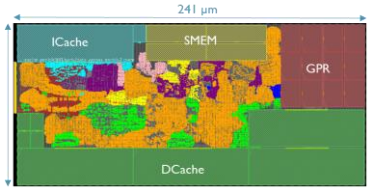
Increasing the L1 memory eliminates the memory access bottleneck



Co-Optimization options
- Optimize Fplan to reduce M2 utilization
- Adjust Icache capacity for area reduction

...

Now we have an Application-aware PD implementation!



imec