Supply Chain Aware Chip Architecture

August Ning, Georgios Tziantzioulis, David Wentzlaff

PRINCETON UNIVERSITY

Semiconductor and Foundry Demand is Growing

- Semiconductors are the linchpin for crucial control and compute logic in all electronics
- Global semiconductor sales are projected to exceed $600 billion in 2022
- More firms are designing in-house silicon for stronger hardware/software integration

Chip Shortage

- The 2020-present chip shortage has been widely discussed spanning semiconductor engineering, manufacturing, supply chain management, economics, and policy
- The chip shortage has led to widespread production delays across multiple industries

Semiconductor Supply Chains are Vulnerable to Disruptions

- The specialized fabrication process makes it difficult to source alternative suppliers
- It takes multiple years new foundries to be built and reach full production capacity
- Complex internal assembly lines prevent foundries from quickly restarting production

Can chip architectures be more resilient to manufacturing challenges?

Chip Creation Process: Challenges and Modeling

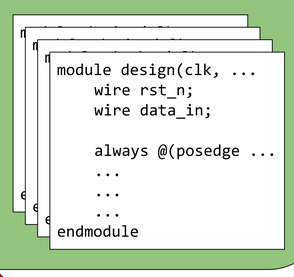
- Designers are locked into a firm and a process node for many design iterations
- Tapeout tools may take hours to days to run
- Advanced process nodes require more DRC and LVS rule checks
- Tapeout timing issues need to be fixed with ECO iterations

- Only three firms (Intel, TSMC, Samsung) produce at "sub 10 nm" processes
- Wafer etching can introduce defects and reduce wafer yield rates
- More demand from more firms and hoarding increase queuing times
- All component dies in the final chip have to arrive before packaging can begin

Challenges

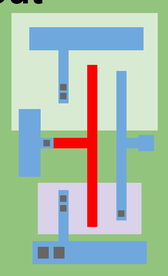
Design

- Functionality
- Design Constraints
- Process Node
- Soft IP



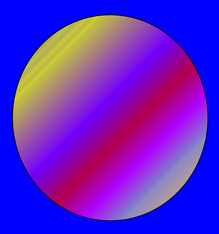
Tapeout

- Synthesis
- Layout
- Design Rules
- Verification



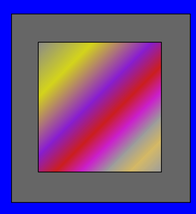
Fabrication

- Mask Creation
- Photolithography
- Silicon Etching
- Metal Layer Routing



Packaging

- Die Testing
- Assembly
- Packaging

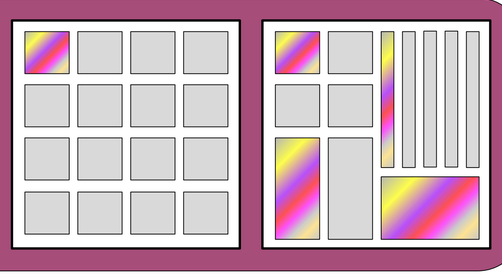


Design Defect

ECO

Time-to-Market Modeling

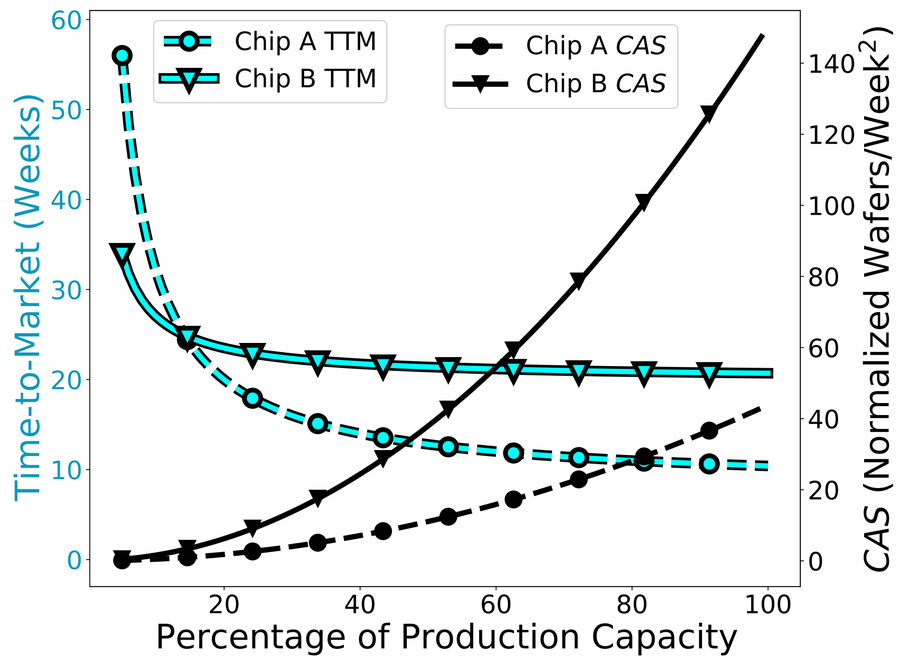
- Block hierarchy design means only unique transistors need to be taped out
- Tapeout time depends on process node choice



- Wafer production rate depends on process node
- Wafer yield rate depends on process node and die area
- Internal foundry assembly lines means each wafer lot (~25 wafers) takes 12-20 weeks to complete the fabrication process

- Packaging firms need to test more dies than are packaged due to yield loss
- Multiple dies may be packaged in a final chip (ex. "chiplet" architectures)
- Internal packaging assembly lines take on average 6 weeks to complete

Chip Agility Score

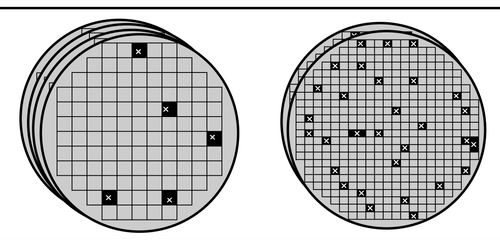


- Computer architects have the least control over **foundry production capacity *μW***
- The time-to-market ***TTM*** of chip design ***d*** depends on current market conditions ***c***, the number of final chips ***n***, and the process nodes ***p*** used in the design
- **Chip Agility Score (*CAS*)** characterizes the *resiliency* of architectures against manufacturing disruptions
- Chips that use multiple process nodes are vulnerable to disruptions on their constituent process nodes which can delay packaging
- Designs with the lowest time-to-market do not necessarily have the lowest ***CAS***

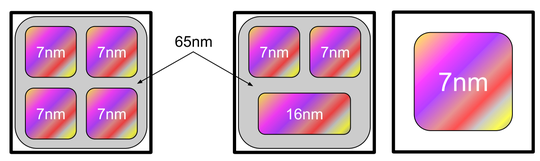
$$CAS = \left(\sum_{p_i \in d} \left| \frac{\partial TTM(c, d, n, p_i)}{\partial \mu_W(p_i)} \right| \right)^{-1}$$

Potential Evaluation

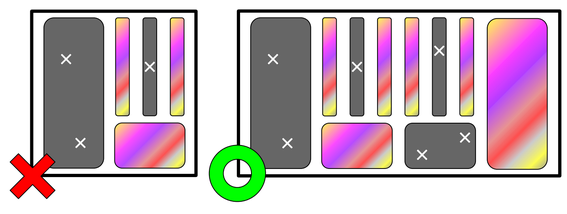
Architects can estimate on which process node their design will have the fastest time-to-market and under what conditions to switch processes (future 3 nm)



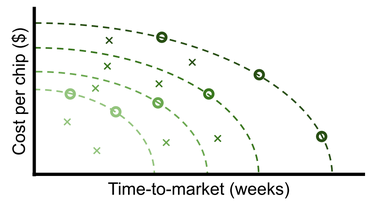
Evaluate tradeoffs between multi-die and mixed process node packaging
(ex. chiplet vs monolithic architectures)



Identify optimal binning/hardware redundancy configurations for fastest time-to-market and/or highest ***CAS***



Combine time-to-market and ***CAS*** modeling with cost analysis, performance analysis, etc. to find pareto-optimal designs



Acknowledgment

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-2039656. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation

aning@princeton.edu
augustning.com

Princeton Parallel Group