SYNOPSYS®
Silicon to Software™

System-Technology Co-Optimization

Scouting the Landscape

Xi-Wei Lin

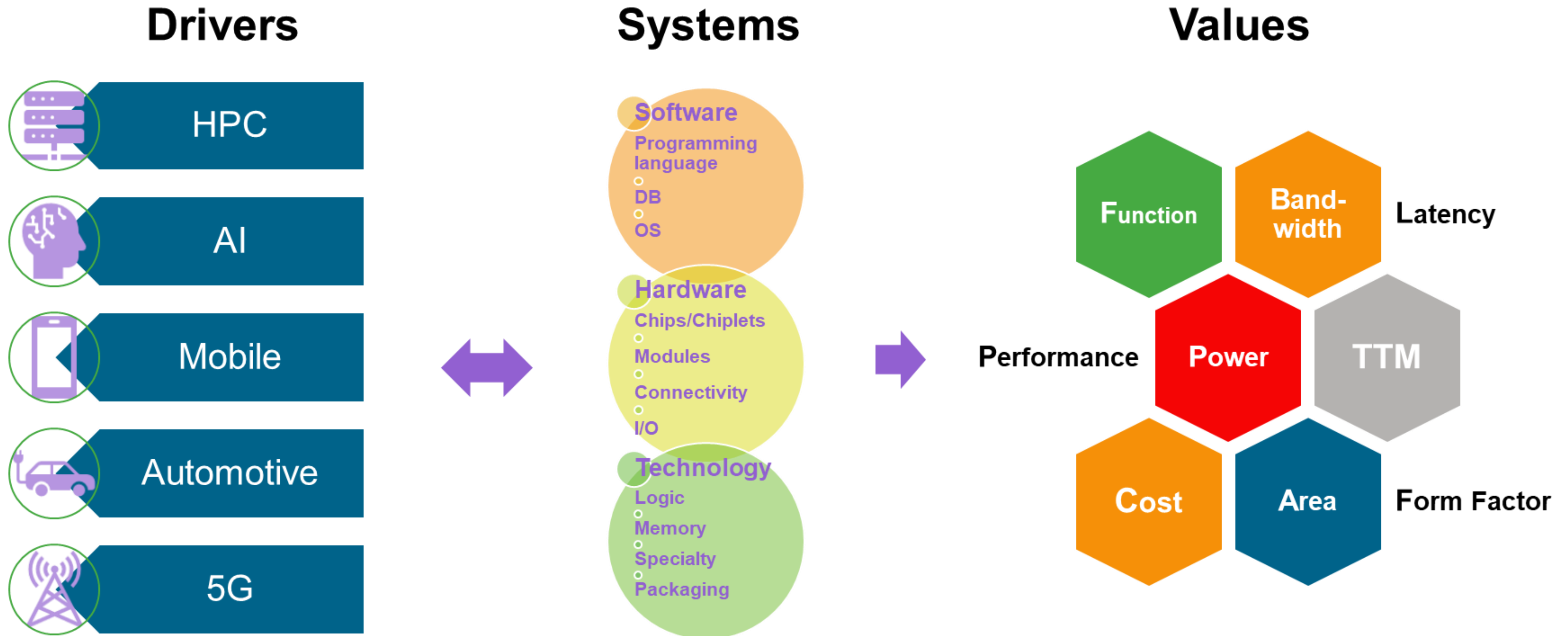
@IEEE EDPS, Milpitas, CA

October 6, 2022

Outline

- Value Proposition
- DTCO vs. STCO
- Case Studies
  - D2D connectivity, hybrid bonding, thermal and stress effects
- Cost Analysis
- Summary

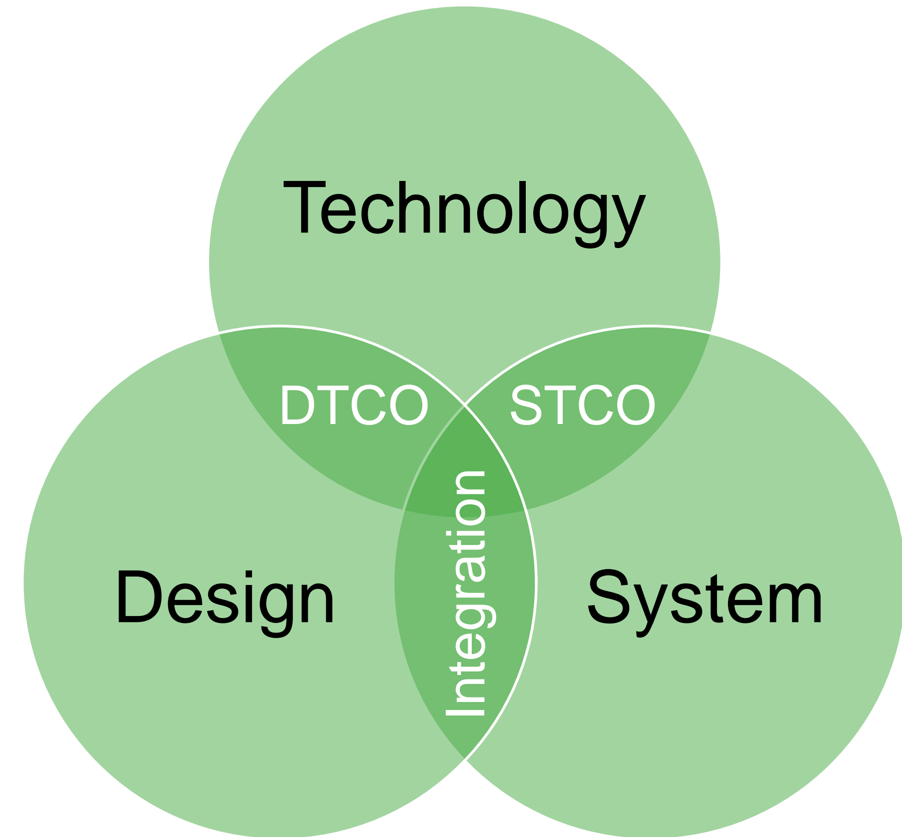
System-Level Innovation and Value Creation



- System level architectures and applications are fundamentally based on hardware and the underlying technologies, which in turn are driven by system requirements.

DTCO vs. STCO

- DTCO (Design-Technology Co-Opt)
  – Device level technology co-optimization with component design
- STCO (System-Technology Co-Opt)
  – Packaging or device level technology co-optimization with system architectures
- System
  – Heterogeneous integration of design or technology components with software stack



Value Proposition: a) Shift left for TTM and b) PPAC improvement

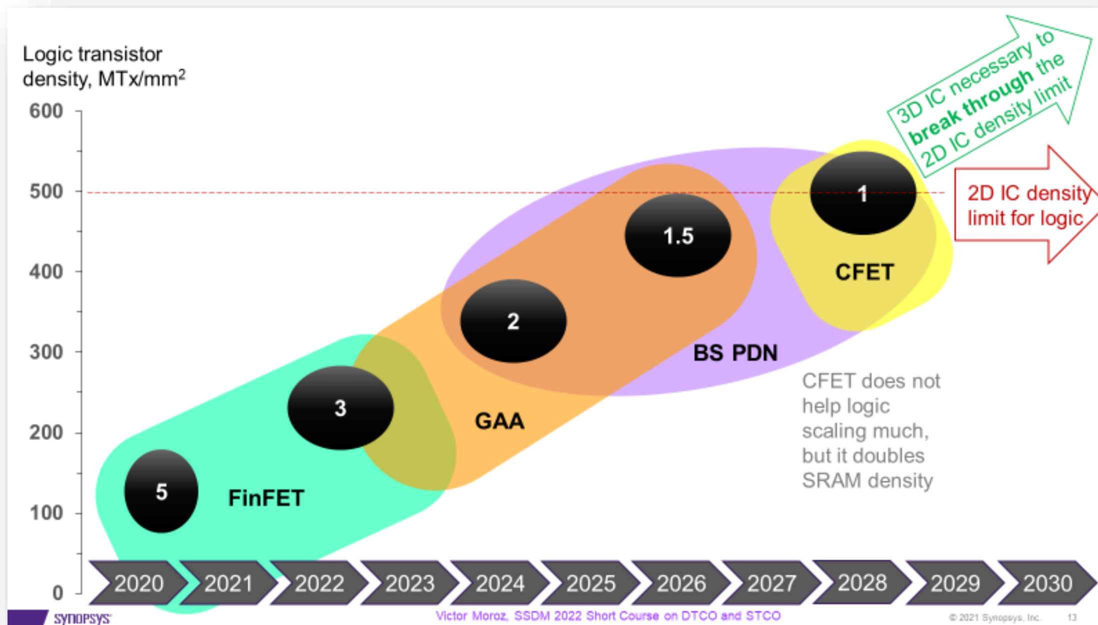
Mitigate Limitations

Limitations	System or Technology Solutions
"Power wall"	Multicore architectures @ same clock speed
Power efficiency	Dynamic voltage & frequency scaling (DVFS)
"Memory wall"	In- or near-memory compute (i.e., computational storage, e.g., FPGA in SSD) or bigger cache near processor via 3DIC
"Bandwidth wall"	Interconnect technology (TSV/μbump and 2.5D interposer/Si bridge) for high I/O density and chip stacking for HBM
Flash memory lifetime due to low endurance	Zoned Namespace (ZNS) spec to let the host manage data placement and garbage collection on the device

- Technology enables system, as much as constrains it, whereas system leverages technology, while mitigating its limitations

Shift Left: Align System Design to Future Technologies

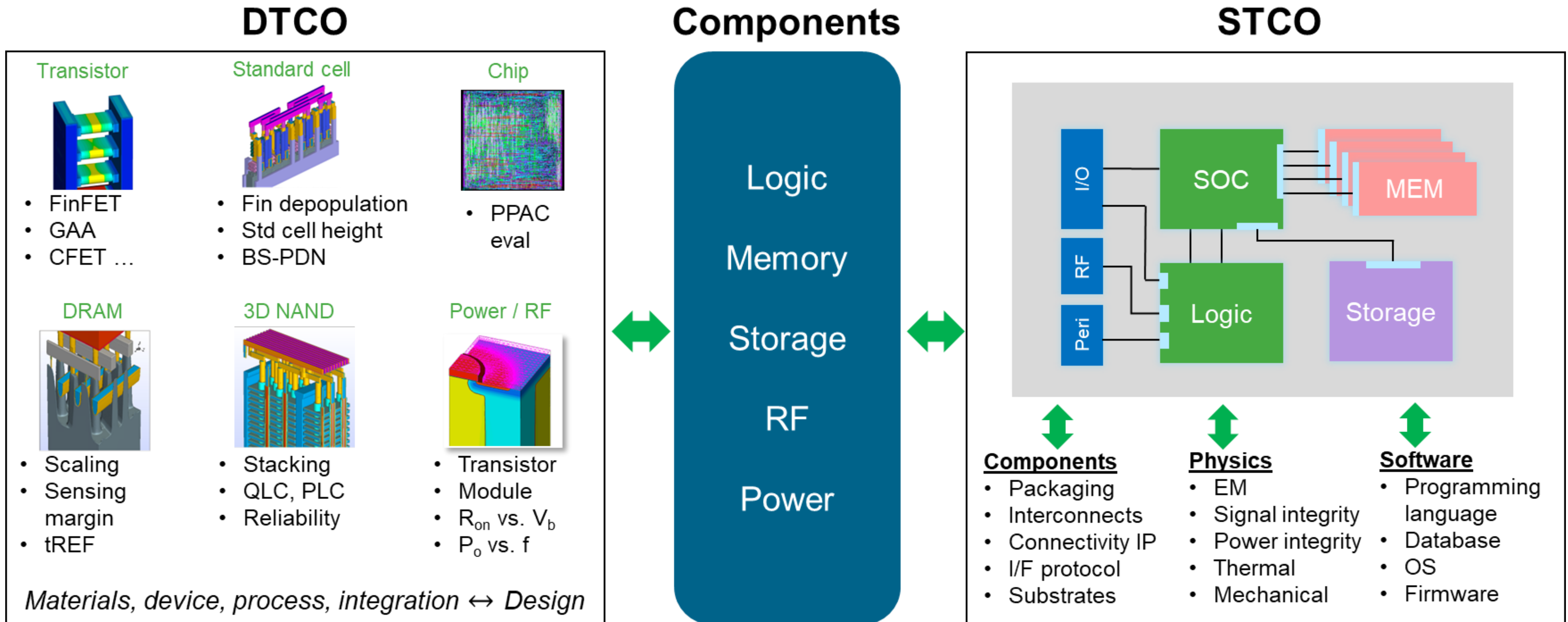
- Questions to Ask …
  - Technology trends and roadmap
  - Technology choice
  - Foundry choice
  - Risk of leading or following



- Impacts to System
  - Designers: Hyperscalers, IDMs, fabless
  - System life cycle is longer than that of technology
  - Application specific ICs are needed to boost system performance and power efficiency for specialized functions
  - The choice of technology and foundry in early stages carries potential advantages, as well as risks
    - PPAC, form factor, TTM, etc.

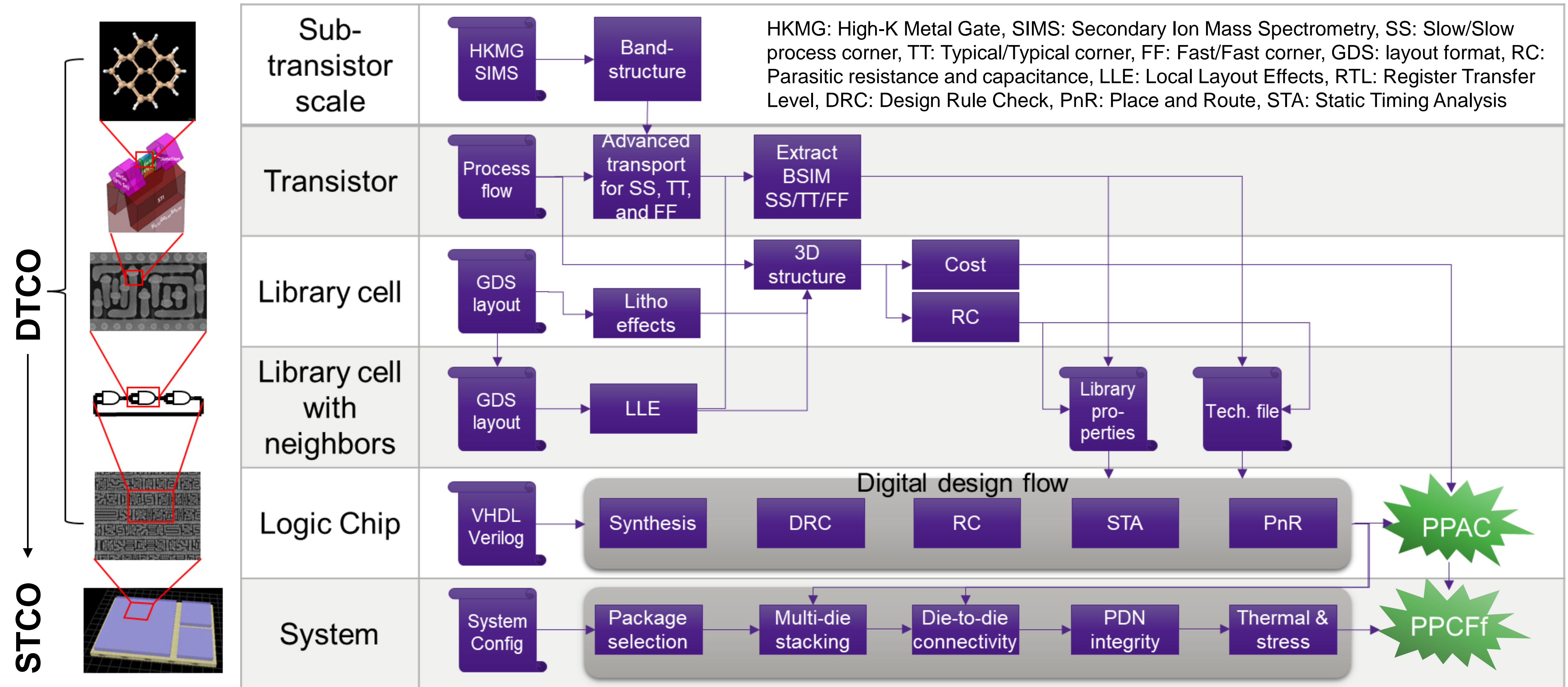
- STCO enables early assessment of technology and its impacts, and helps to shape product strategy

STCO as an Extension of DTCO



- DTCO feeds STCO with device optimization, while STCO adds system level components and physics

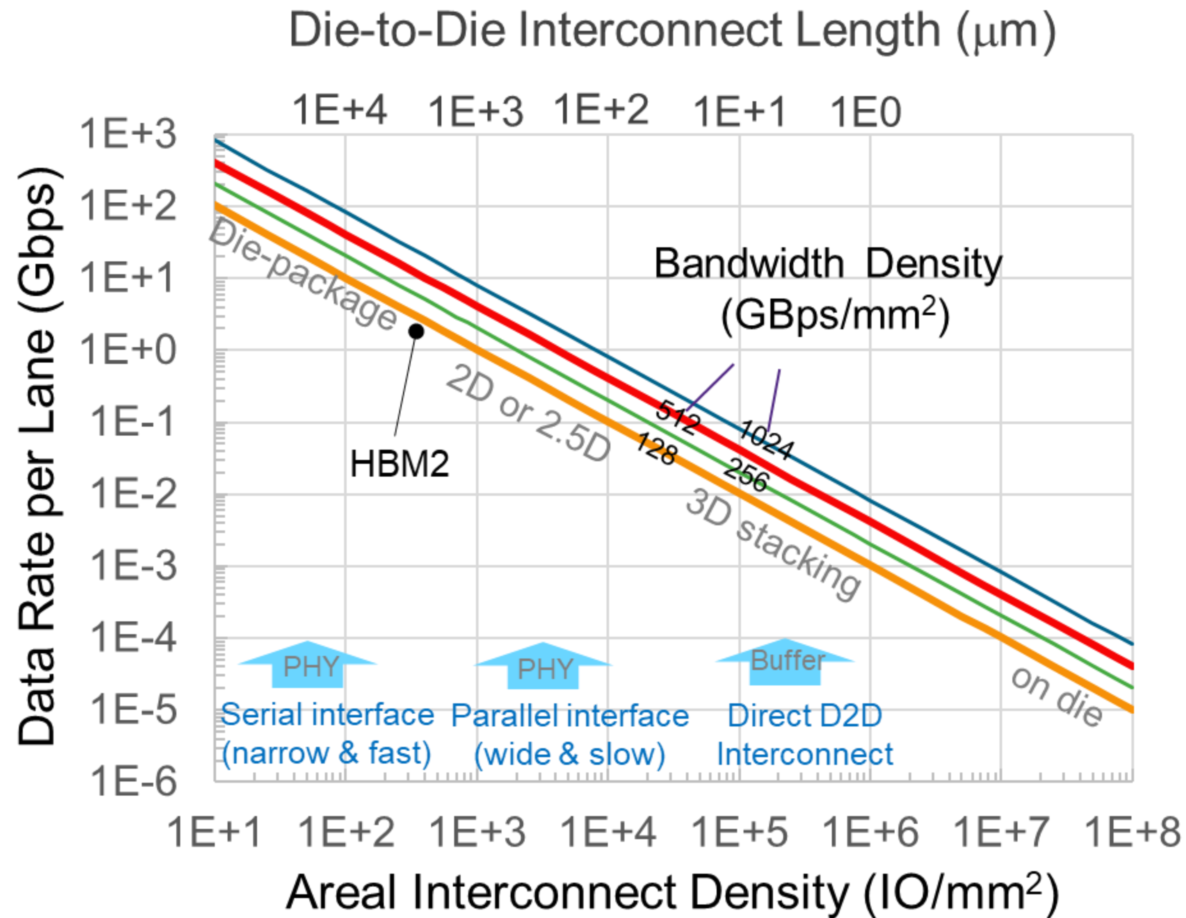
STCO Flow for Multi-Die System Design



PPCFf: performance, power, cost, form factor

Source: Victor Moroz, IEDM 2021 Short Course

Die-to-Die Connectivity Design Trade-offs

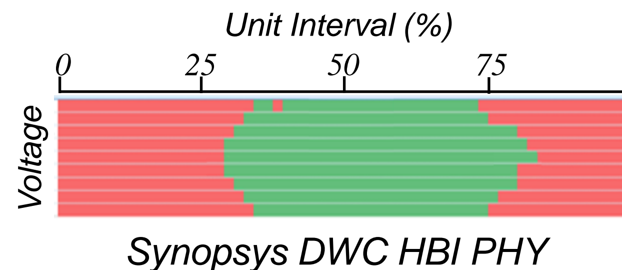


- **Bandwidth** = R_{data} x N_{IO}
  - A target bandwidth can be achieved by trade-off of data rate vs. number of IOs
- **Serial interface**
  - Narrow bus, but fast lane
- **Parallel interface**
  - Wide bus, but slow lane
  - *Lower data rate makes IP design simpler, resulting in lower power*
- **PHY** and interface **standards**
  - PHY and controller enable data communications by standard protocols, such as AIB, OpenHBI, UCIe, and CEI-112G-XSR
- **Direct D2D interconnect** by direct bonding
  - Drastically increases IO density, reduces interconnect length and LRC, thus
  - Enables direct data transmission by buffer insertion, in lieu of PHY.

Die-to-Die Interface PHY Example

Architecture	Parallel Interface	Serial Interface
Package	2.5D interposer	Organic substrate
Bump pitch	40 - 55 μm	130 – 150 μm
Interconnect density	10^2 – 10^3 IO/mm^2	10^1 IO/mm^2
Line space	>0.4 μm	> 10 μm
Interconnect length	<5 mm	<50 mm
Data rate/lane	2 – 8 Gbps	2.5 – 112 Gbps
BW density	2-3 Tbps/mm	1.6-2 Tbps/mm
Power	<0.5 pJ/bit	1.0-1.5 pJ/bit
Latency TX+RX	~4.5 ns	~5.5 ns
Bit error rate	<<1E-15	<1e-15 for NRZ
Standards	HBI, OpenHBI, AIB2.0	OIF, CEI 112G, USR/XSR

Receiver eye diagram, i.e., sampling threshold voltage vs. unit interval (UI), for a Synopsys DWC HBI PHY with a 4 Gbps interposer link. The eye opening is >50% of the UI.



Synopsys DWC HBI PHY

- System requirements:
  - Form factor, power efficiency, latency
- Technology enabler:
  - 2.5D interposer, μbump/TSV
- Options
  - Compared with organic substrate, 2.5D Si interposer makes parallel interface feasible and desirable, thanks to ~3X smaller bump pitch and 10X shorter channel length.
- Results
  - By consequence, the required data rate per lane (signaling speed) is reduced, making PHY design simpler and resulting in ~2X-3X **lower power and** 1ns **less latency**.

3DIC Interconnect Pitch Scaling

Die-to-Die Interconnect Length (μm)

10^4 10^3 10^2 10^1 10^0

Bump or TSV Pitch (μm)

10^{-1} 10^0 10^1 10^2 10^3

On-chip

~0.5-10μm

Direct bonding & μTSV, nTSV

3D stacking

~40-55μm
~10-40μm

TSV & μbump

2D, 2.5D interposer, embedded bridge, or 3D F2F stacking

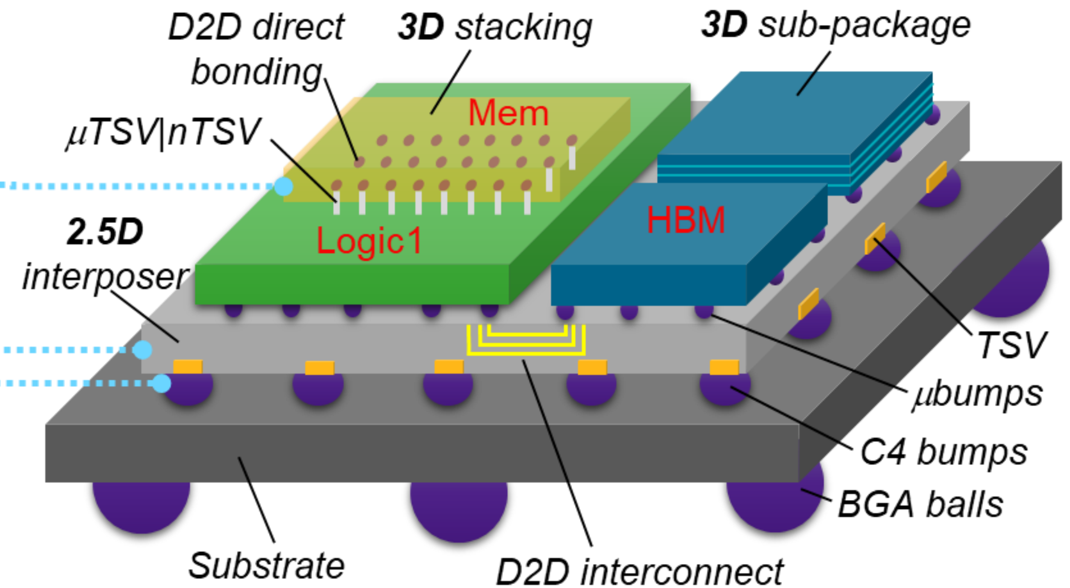
~100-150μm

C4 bump

flipchip

10^1 10^2 10^3 10^4 10^5 10^6 10^7 10^8

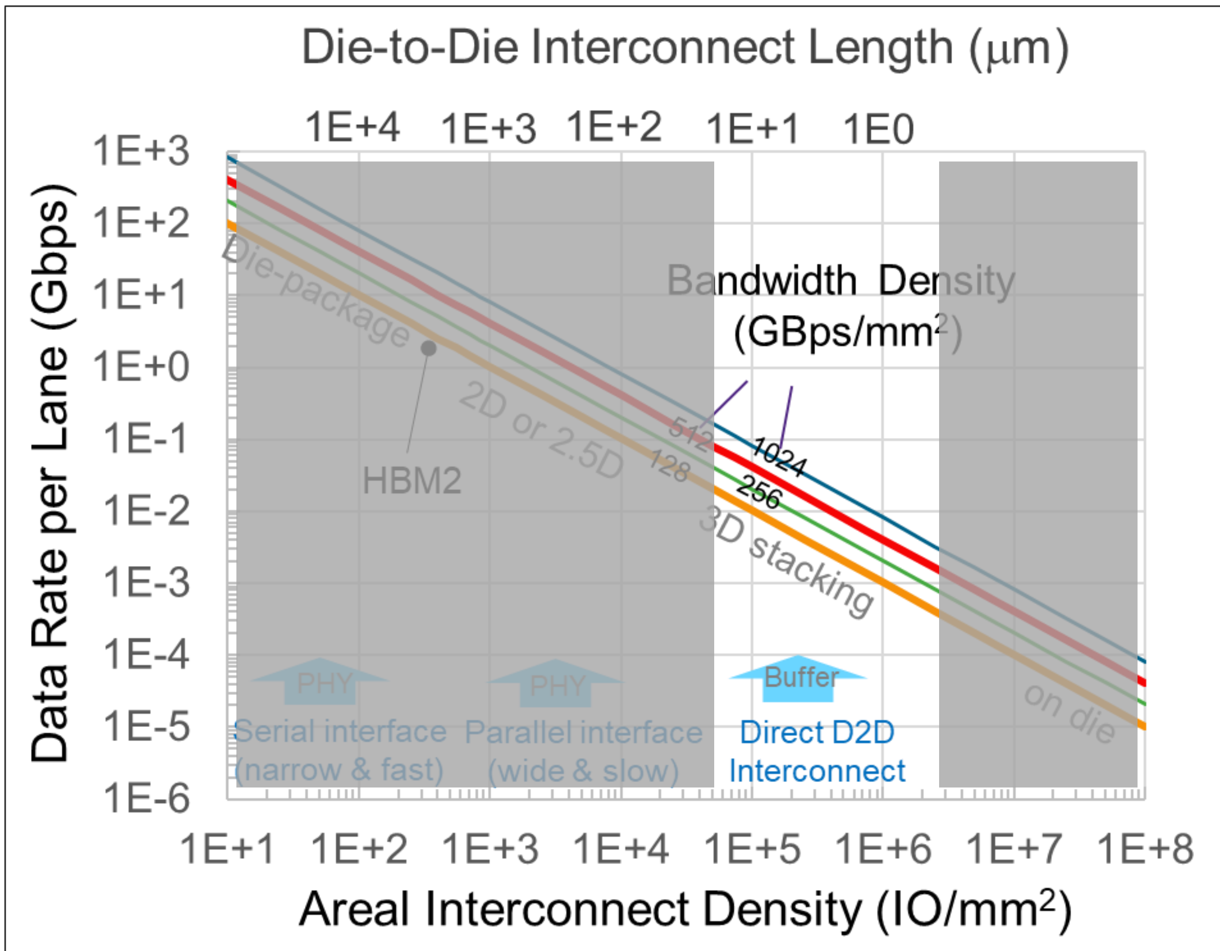
Areal Interconnect Density (IO/mm²)



- Multi-scale interconnects may co-exist in a package

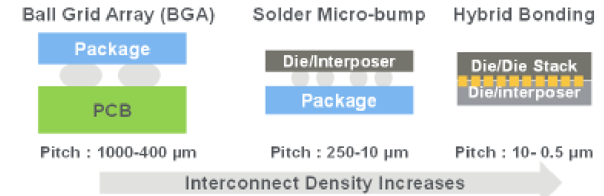
Source: Xi-Wei Lin et al., IEDM 2021

Hybrid Bonding for Fine-pitch Interconnect Scaling



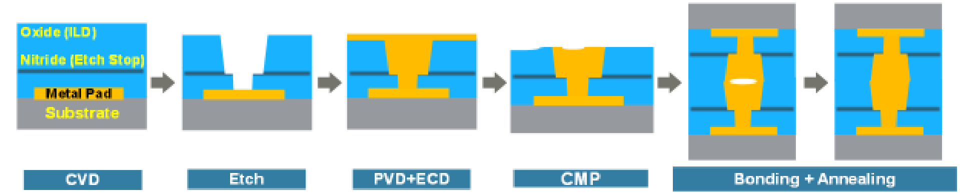
Hybrid Bonding Technology

Fine-pitch Interconnect Scaling



Hybrid bonding enables sub-um pitch

Hybrid Bonding Process Flow [1]

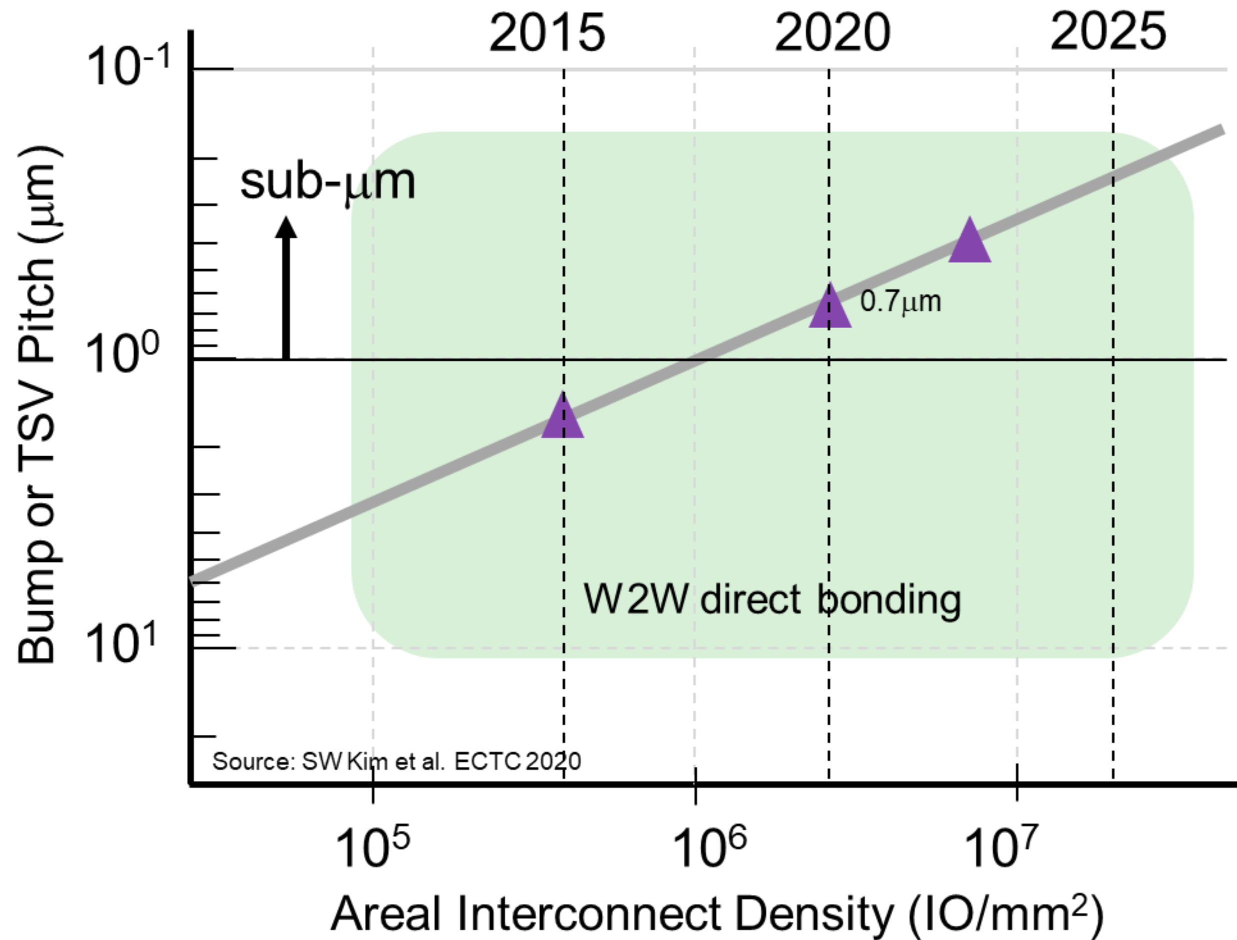


[1] Albert Lan, "TSV and Hybrid Bonding Solutions for 3D Heterogeneous Integration Packaging Applying in Next AI/HPC Era", 17th International Conference on Device Packaging, '21

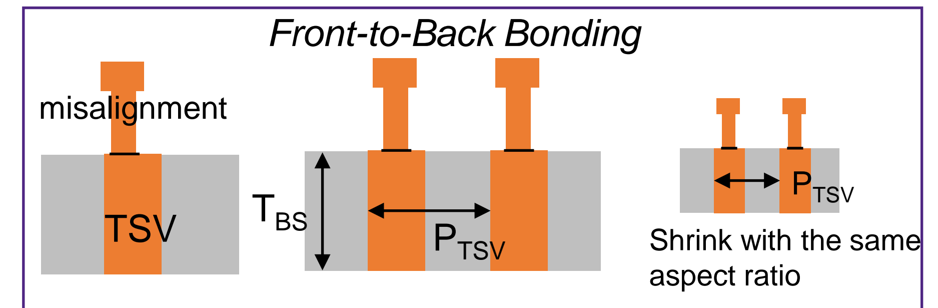
Source: L. Jiang et al., AMAT, ECTC 2022

synopsys © 2021 Synopsys, Inc. 14

HB Scaling Trend and Challenges

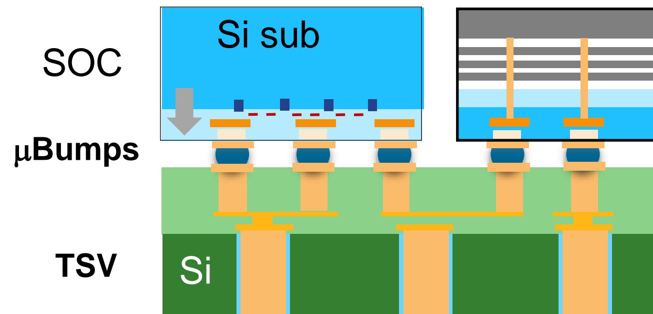


- System Impacts
  - Drastically increases IO density, reduces interconnect length and LRC
  - Choice of direct data transmission by buffer insertion versus PHY & interface
- Trends
  - Min W2W pitch reported in 2020 < 1μm
  - Min W2W pitch at production ~ 2μm
  - Min D2W pitch at production ~ 10μm
  - Further pitch scaling is expected
- Limits



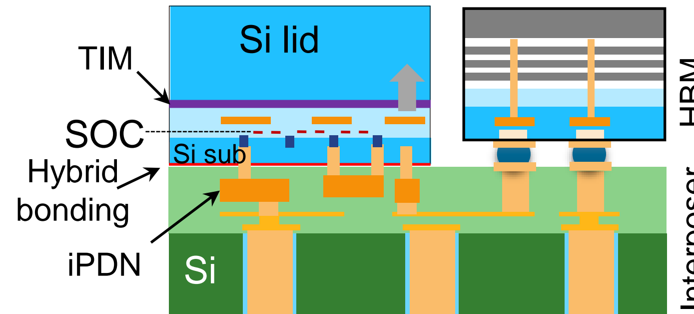
Thermal Analysis

Reference

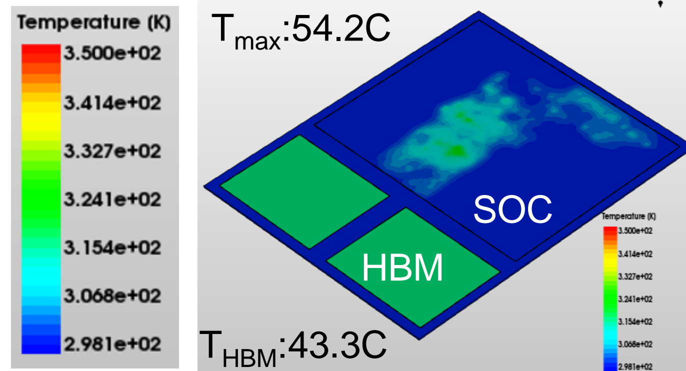


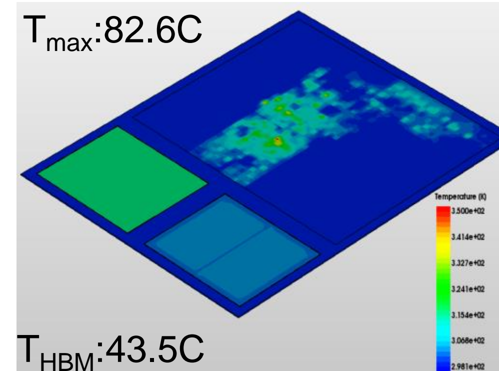
SOC face down, FS-PDN, μbumps

***Design 3**: Si lid with TIM*



***Novel Design**: SOC face up, BPR, backside nTSV, hybrid bonding, iPDN in interposer*

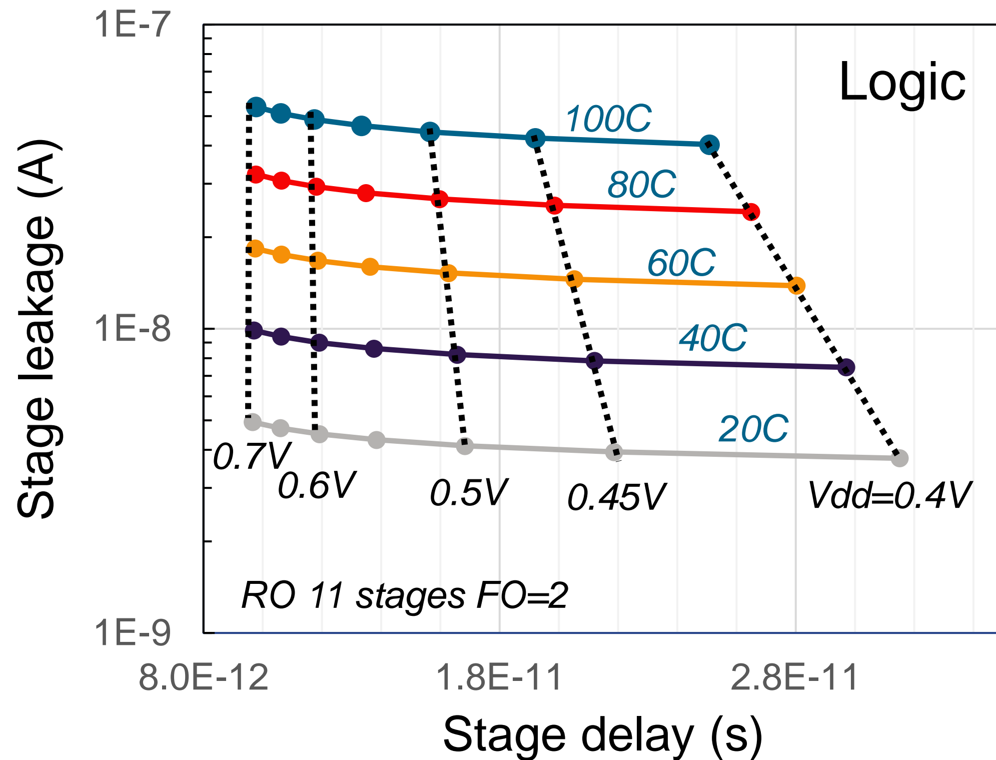


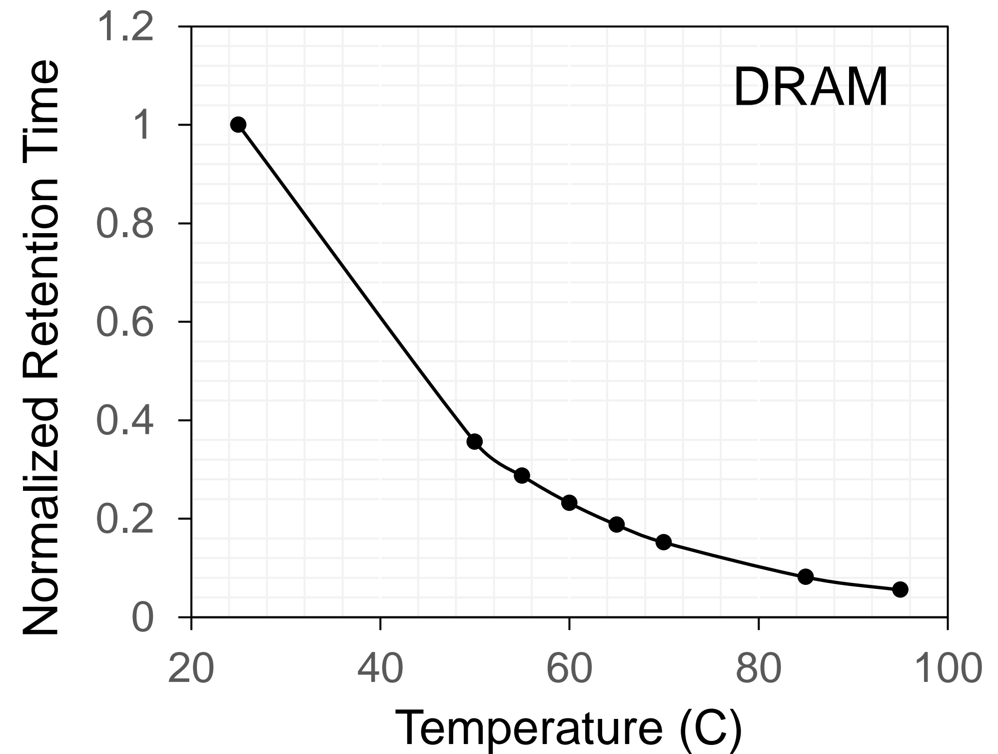


- Package level thermal analysis is done for power density map extracted from 2D SOC design.
- The use of thermal interface material (TIM) for Si lid (Design 3) leads to higher T_{max} (82.6°C), due to thicker material and inferior thermal conductivity.

Source: Xi-Wei Lin et al., IEDM 2021

Thermal Effects on Electrical Properties

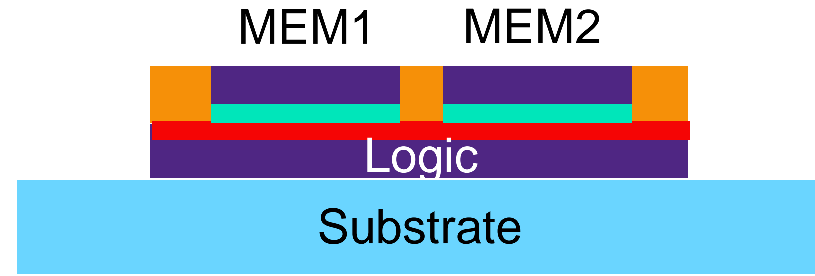




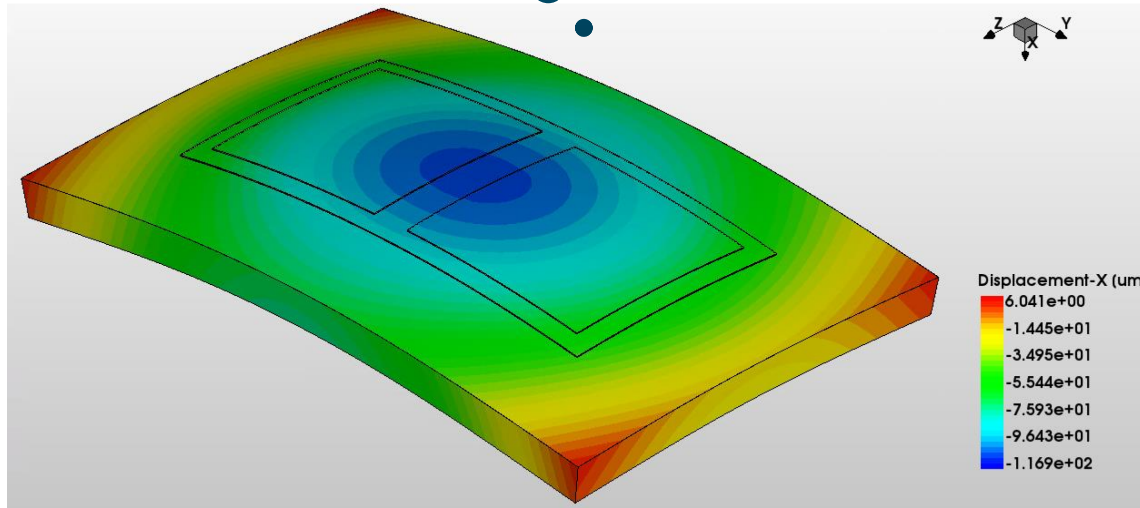
- For logic SOC, junction temperature has little effects on delay around the nominal Vdd=0.7V, but affects leakage exponentially.
- For DRAM, the retention time is sensitive to temperature.

Stress: Warpage of a Stack of 2 Chiplets on a Larger Chip

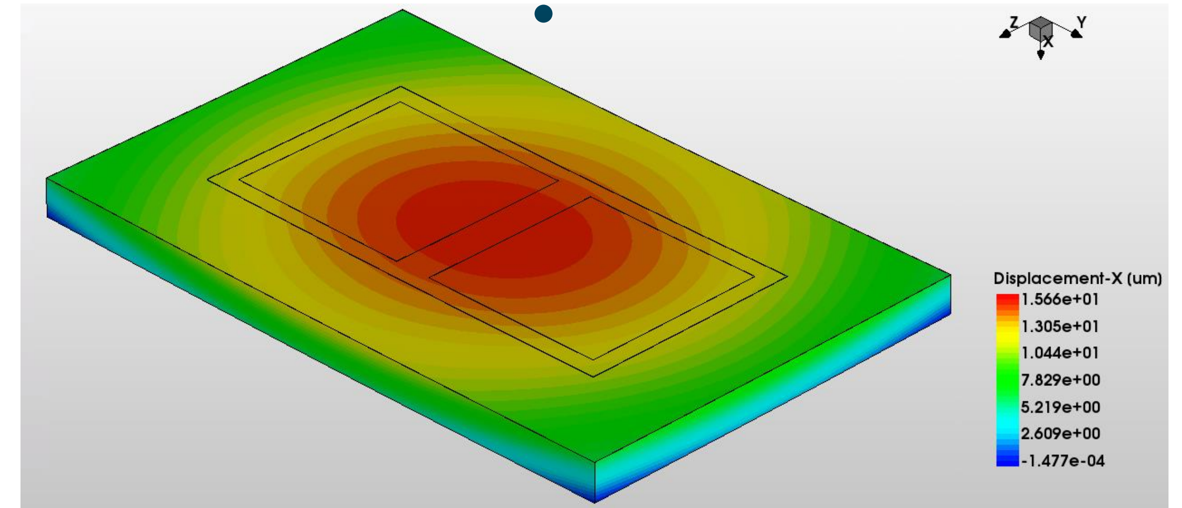
Strong warpage, but low Si stress



Weak warpage, but high Si stress



50 um thick Si dies

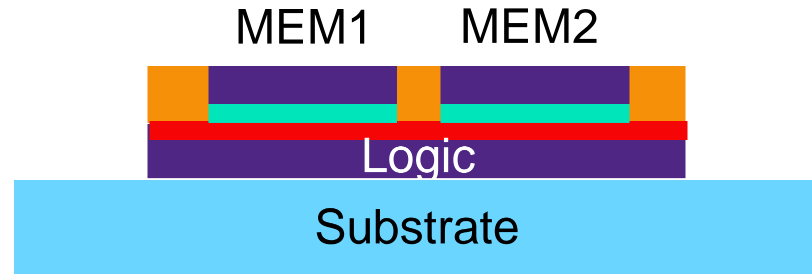


100 nm thin Si dies

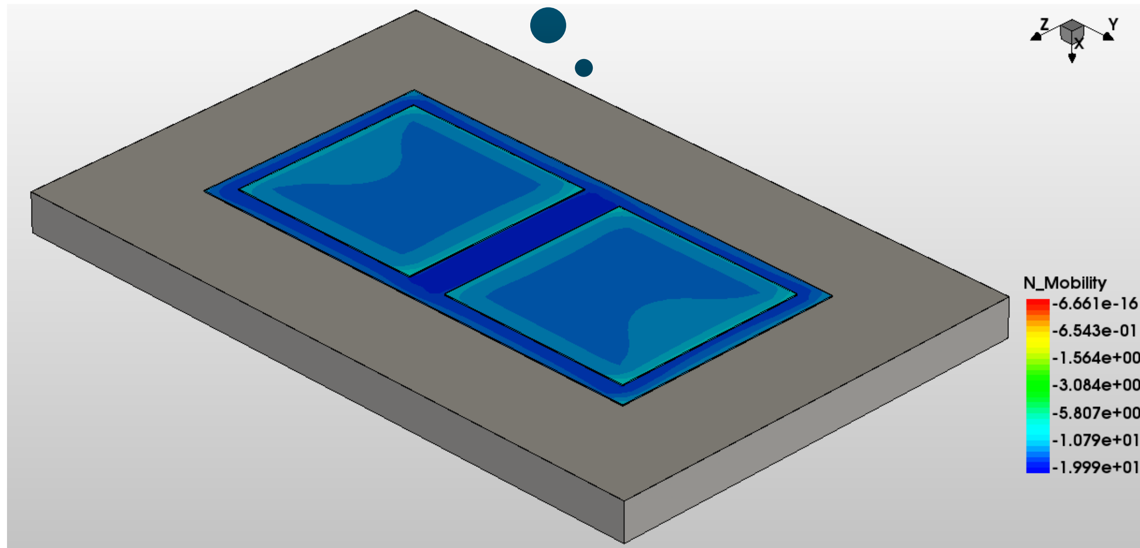
Source: Xiaopeng Xu, Synopsys

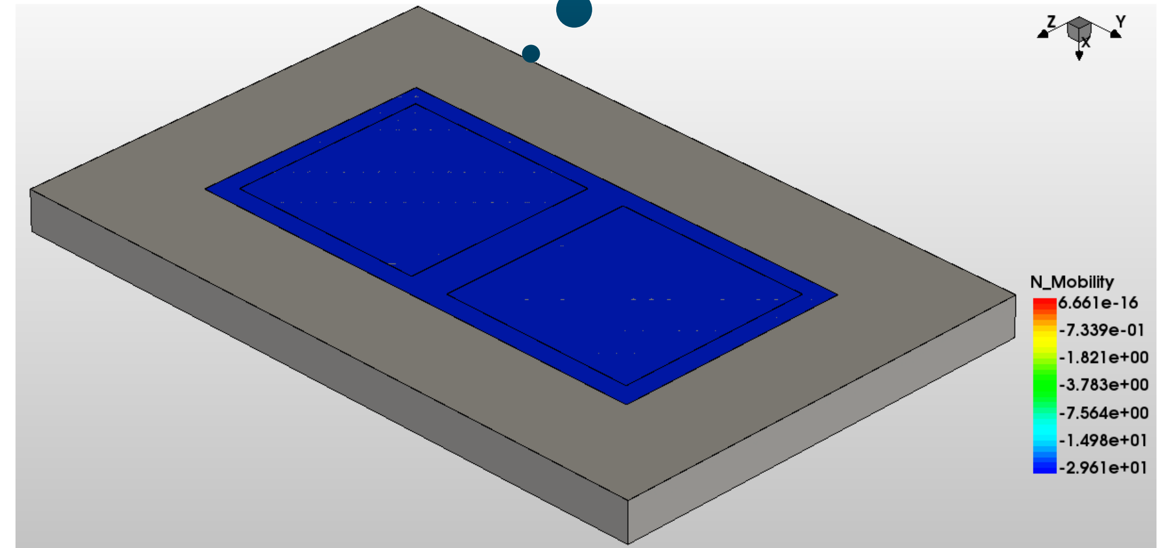
NMOS Idlin Map for a Stack of 2 Chiplets on a Larger Chip

10% to 20% NMOS degradation



Uniform 30% NMOS degradation





50 um thick Si dies

100 nm thin Si dies

Source: Xiaopeng Xu, Synopsys

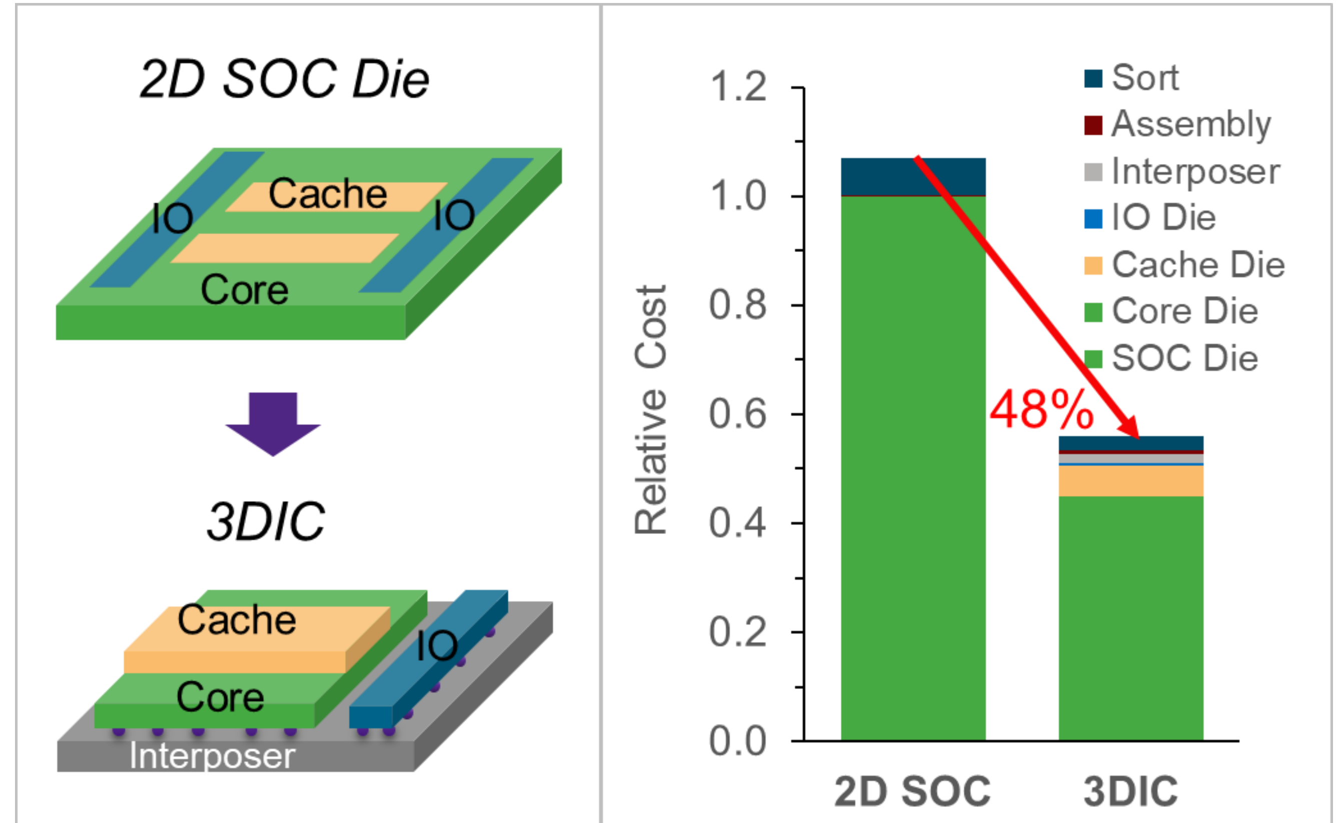
Cost Analysis: Example of Disaggregated HI Scenario

2D SOC

Monolithic	Tech Node	Area
SOC	2nm GAA 17ML	6.0 cm^2
- Core	Same	51%
- Cache	Same	34%
- IO	Same	15%

3DIC

Chiplet	Node	Interconn.	Area
Core + L1-2	2nm GAA 17ML	TSV + μbumps	3.88 cm^2
L3 Cache	2nm GAA 4ML	Direct bonding	1.22 cm^2
IO	90nm planar 7ML	μbumps	1.17 cm^2
Silicon interposer		6ML + TSV + μbumps	6.84 cm^2



- 3DIC lowers the cost by 48%, thanks to a) better yield due to smaller die size; b) simpler BEOL for L3 cache chiplet; and c) mature and cheaper node for IO
- They offset added cost due to sort and assembly and interposer

Source: Xi-Wei Lin et al., IEDM 2021

Summary

- STCO value proposition: a) shift left for TTM and b) optimization for PPAC and form factor.

- STCO is a natural extension of DTCO, with added system level components (e.g., interconnects, connectivity IP) and physics (e.g., EM, thermal, stress), as illustrated by examples.

- Cost analysis is critical to system exploration.

SYNOPSYS®

Silicon to Software™