AMPERE

Touch

Primary Focus

Test, Repair and Reliability Challenges of Chip-let Interconnects

Dr. Sreejit Chakravarty, IEEE Fellow, Distinguished Engineer

Ampere Computing

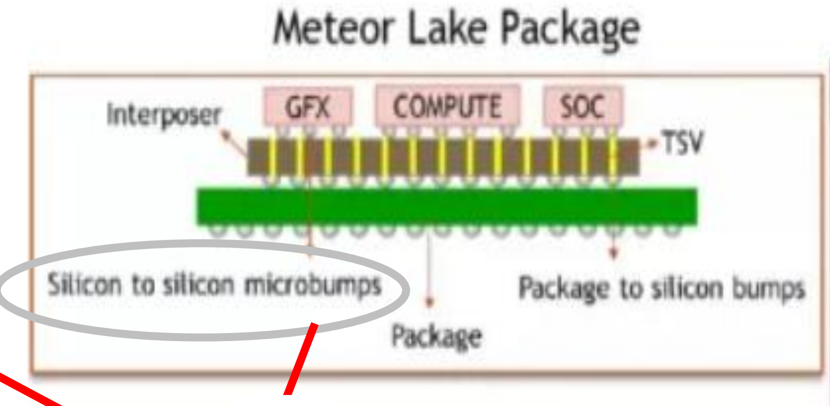
Santa Clara, CA

Agenda

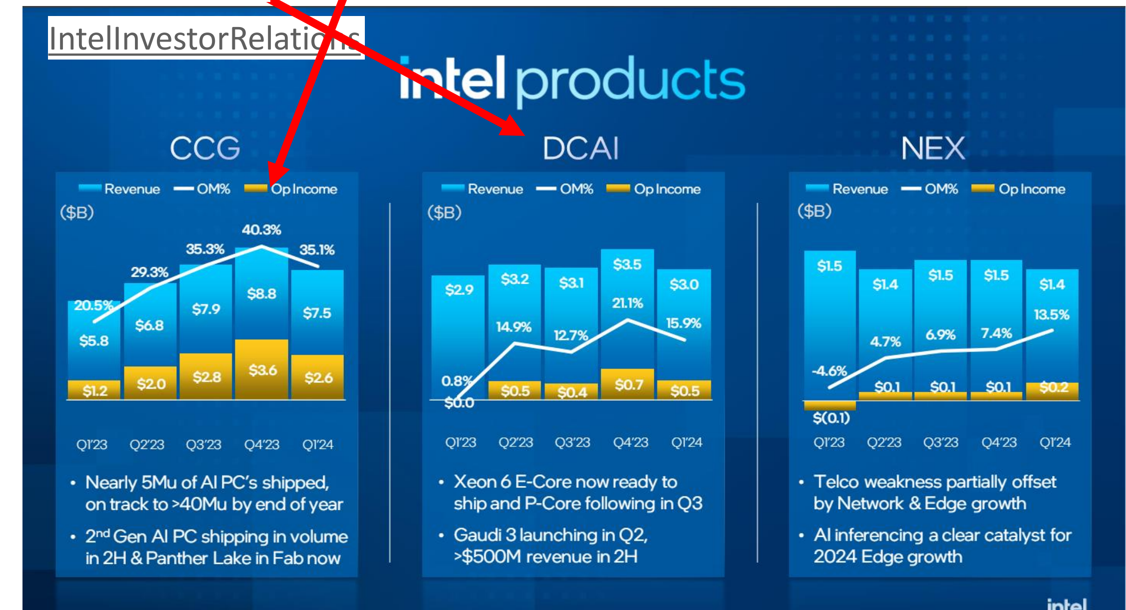
- Background and Motivation
- Packaging Technology and Chip-Let Interconnect Defects
- Chip-Let Interconnect Reliability Issues

CHIPLET BASED SoC EXAMPLES

IntelMeteorLake



3.0B 7.5B



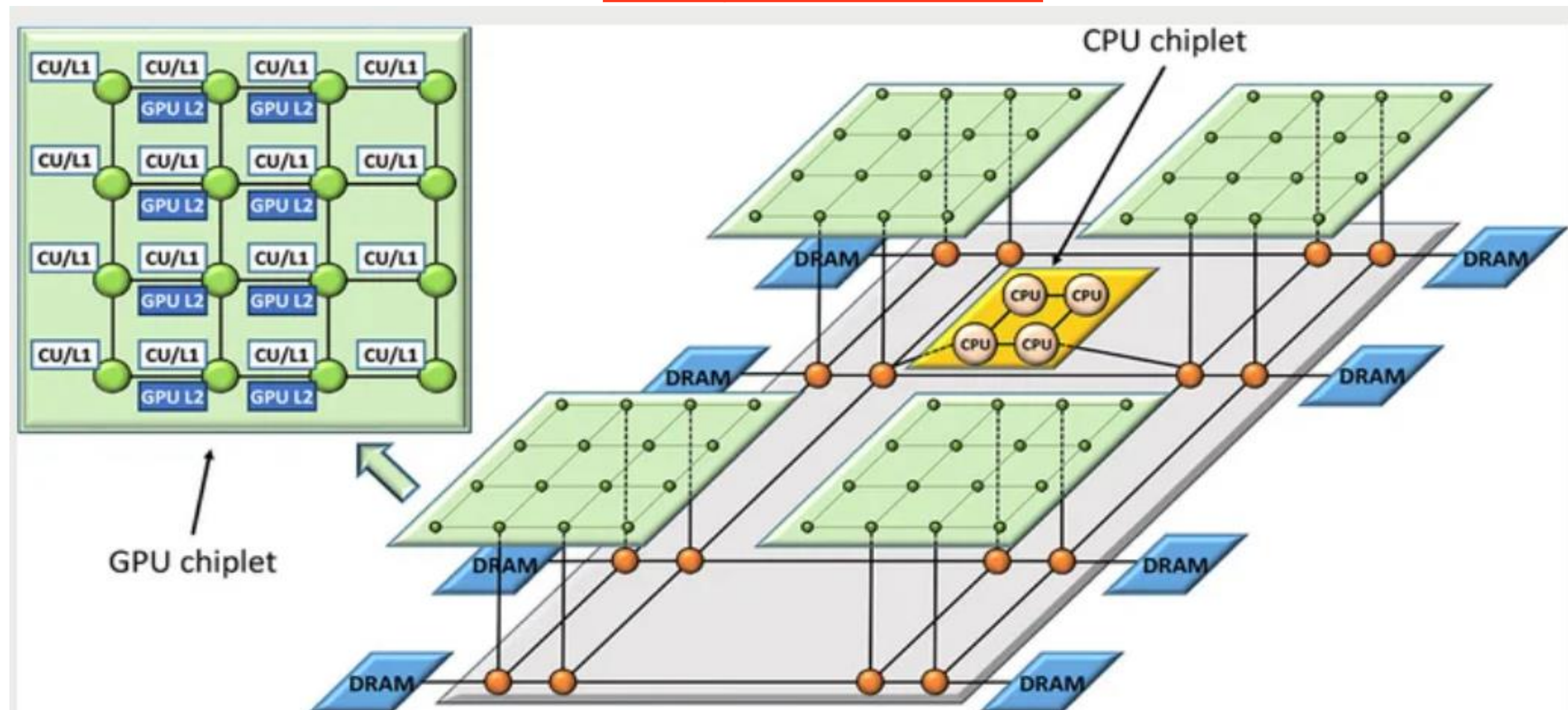
SoC Example using multiple interconnect technology

- **Chip-let based design**
  - **spans multiple market segments**
  - **Uses multiple packaging technology**

AI and Chip-Let Based Design

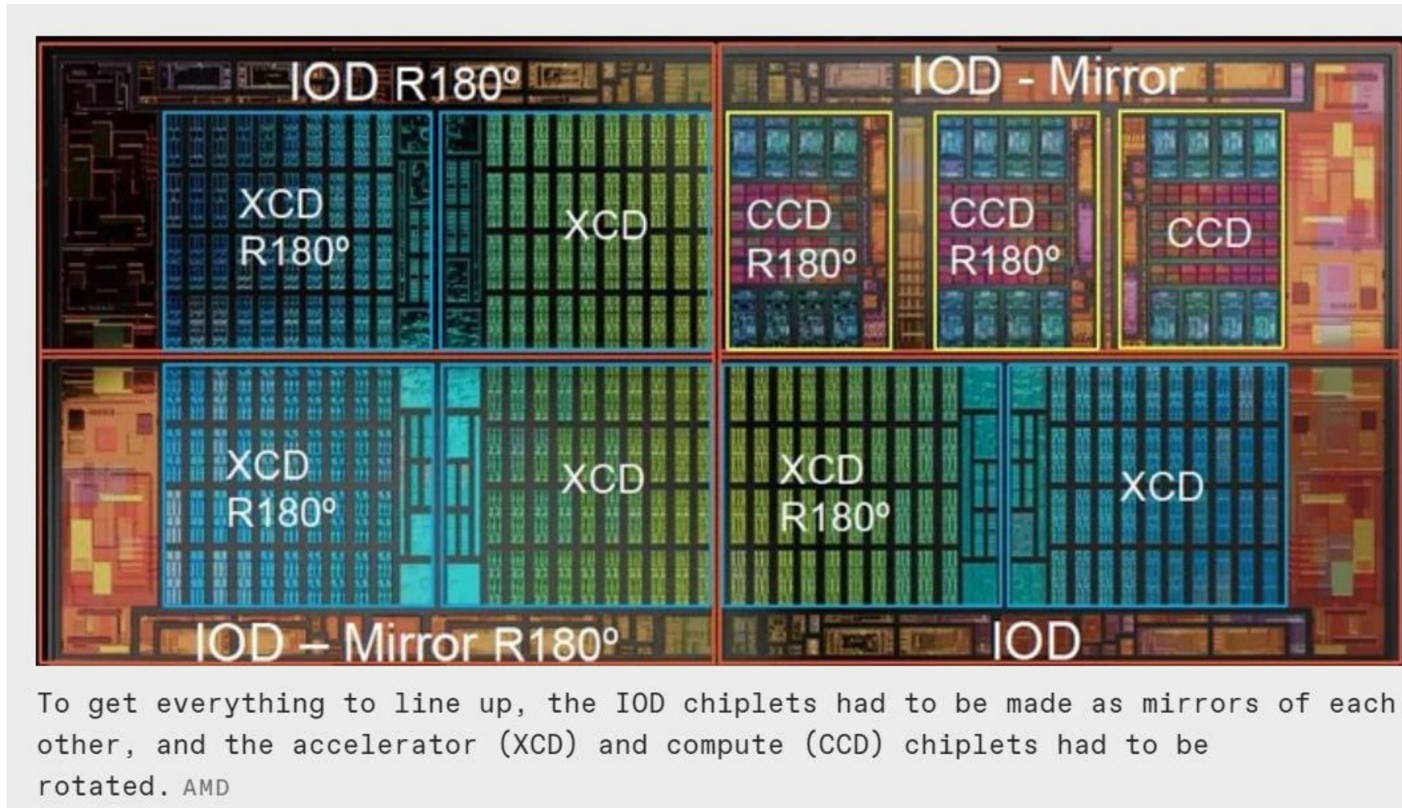
Mandatory AI Slide!

IEEESpectrumArticle



AI and Chip-Let Based Design

IEEESpectrumAMD_MI300



To get everything to line up, the IOD chiplets had to be made as mirrors of each other, and the accelerator (XCD) and compute (CCD) chiplets had to be rotated. AMD

CHIPLET INTERCONNECT TREND: Test Implication

Feature	Feature Trend	Test Implication
Density	100µm → 50µm → ≤ 10µm [35-25µm in the market]	• Higher defect level • DPPM impact and increasing repair need
Technology Change	µBump to Hybrid (direct) and combinations there-off	• Defect profile change • Repair support change
Chiplet interconnect count per SoC	100s to 200,000 and counting [Increasing very rapidly]	• DFT insertion and verification complexity increase • Test time increase • Greater need for repair to improve yield • Need for infield repair • Repair cost (muxing logic, repair fusing cost)
Frequency	100s of Mhz to 3.5+ Ghz and counting	• Greater need to test for marginalities, in addition to defects

Agenda

- ✓ Background and Motivation
- ➢ Packaging Technology and Chip-Let Interconnect Defects
- Chip-Let Interconnect Reliability Issues

Advanced Packaging Intro [Source: UCIe 1.1]

- Packaging used is based on power, form factor, size, cost etc.

Figure 1-4. Advanced Package interface: Example 1

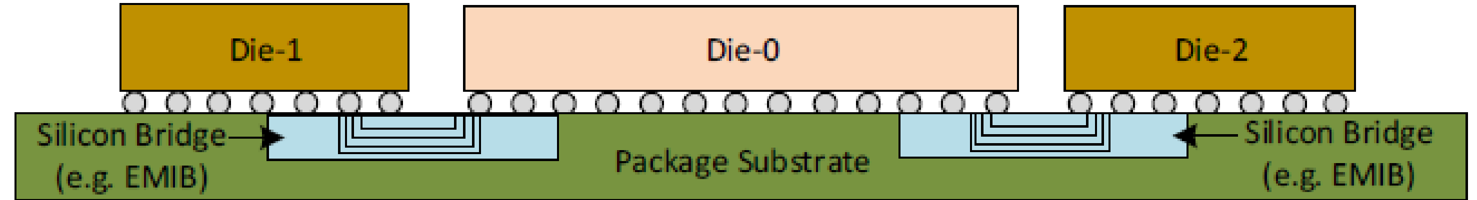


Figure 1-5. Advanced Package interface: Example 2

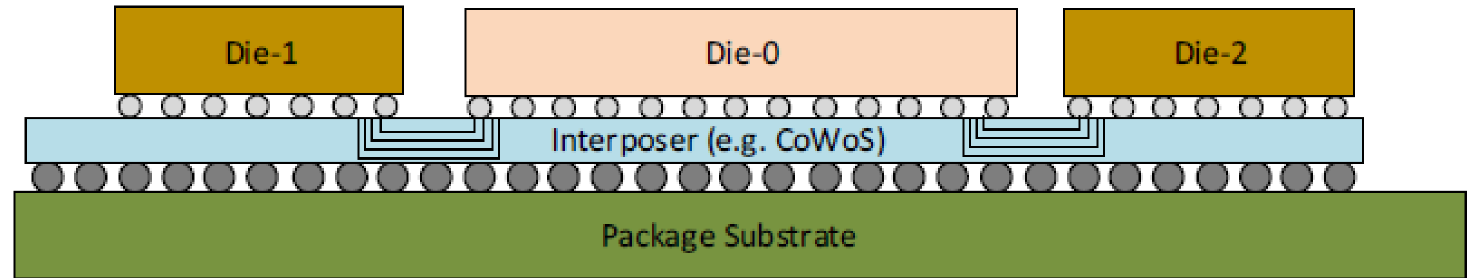
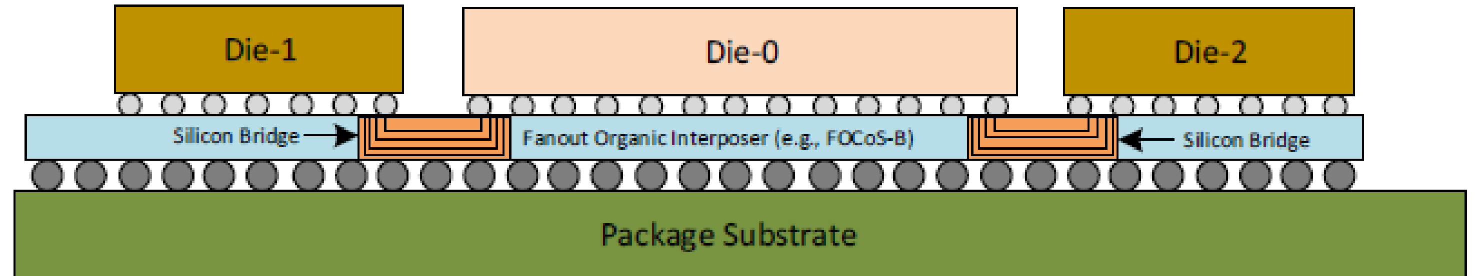
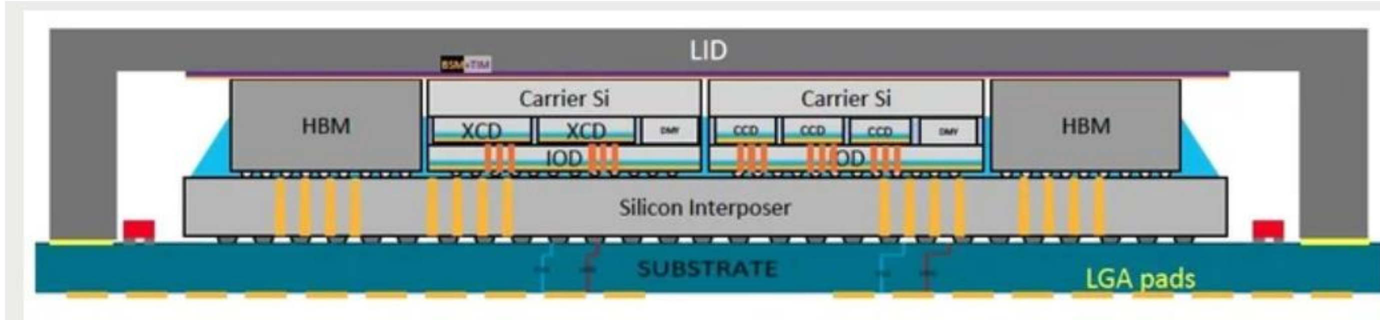


Figure 1-6. Advanced Package interface: Example 3



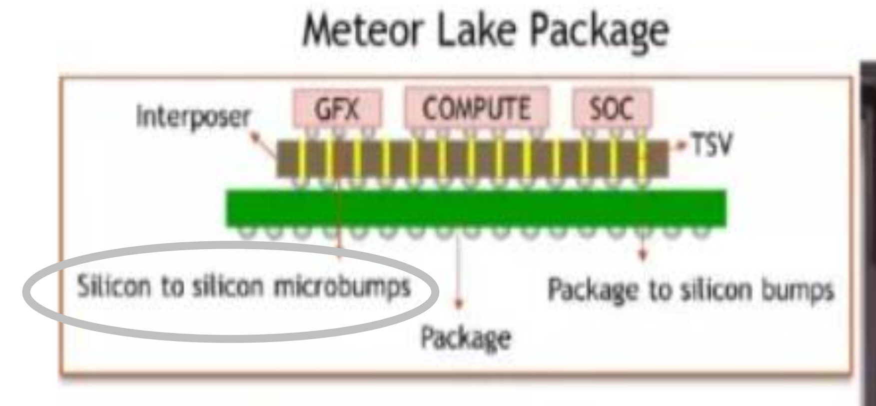
Silicon Interposer Usage Examples



Compute and AI chiplets are stacked on top of I/O and cache chiplets in the MI300a. AMD

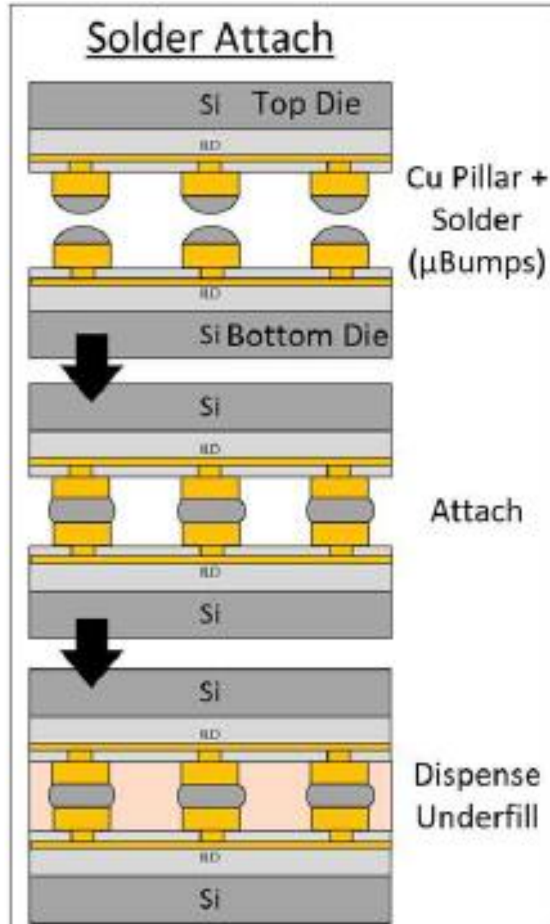
IEEESpectrumAMD_MI300

IntelMeteorLake



µBump based Interconnects

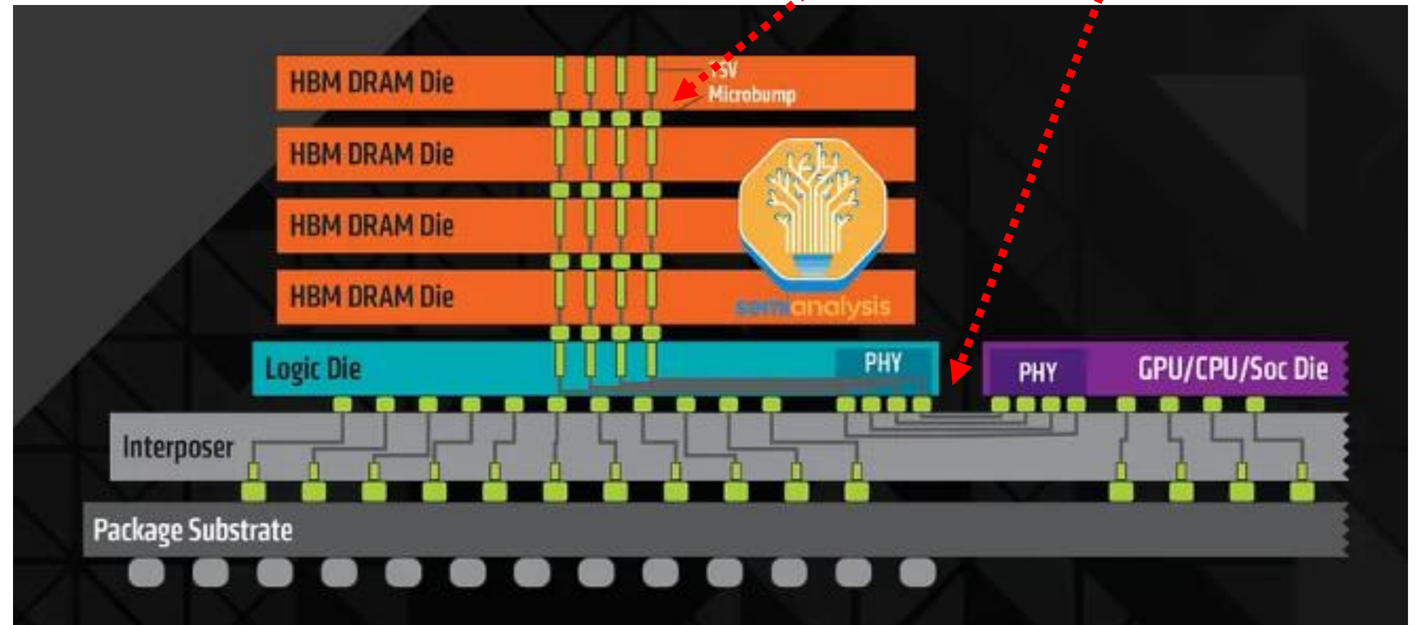
Source: EDM 2021 Intel Paper



semianalysis.com:AMD Radeon

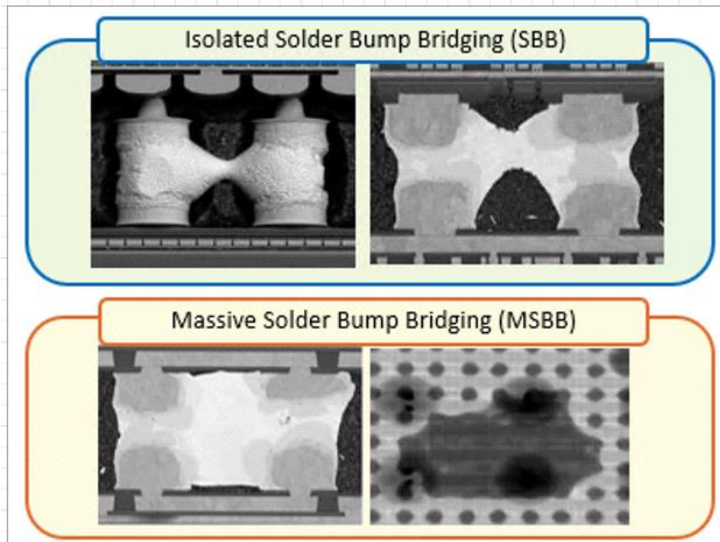
3D Usage

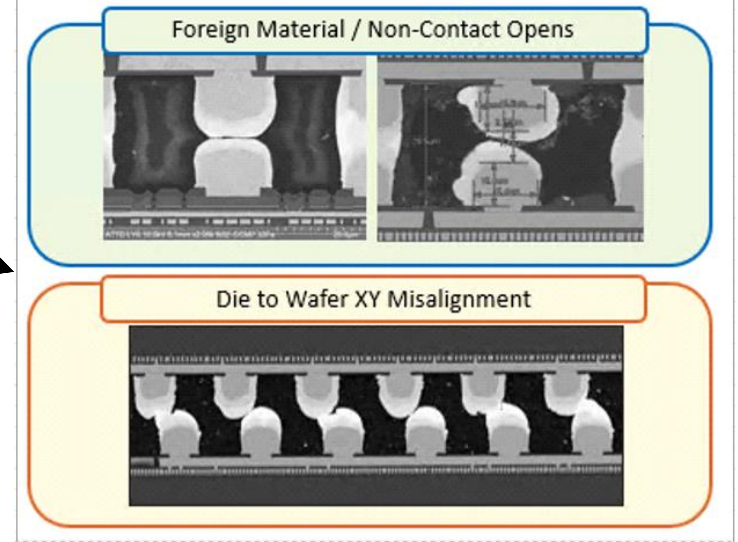
2.5D Usage



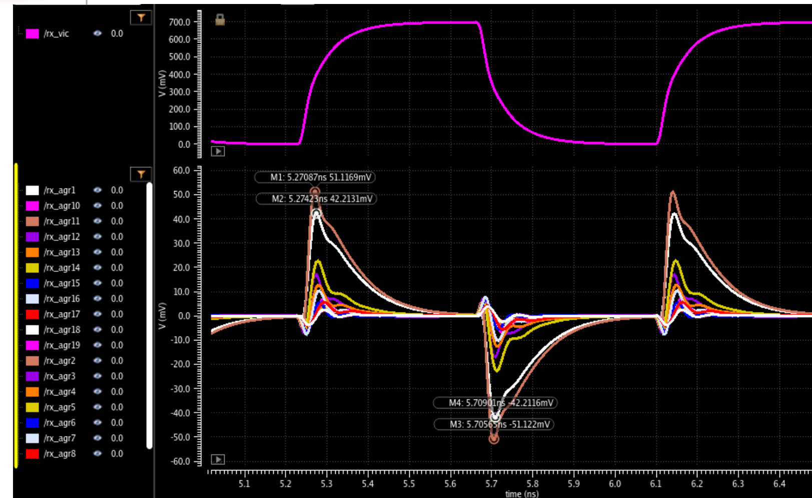
µBUMP DEFECT PROFILE

PHYSICAL: DEFECTS
[X-Ray]





COUPLING
[Spice Simulation]



μBUMP DEFECT PROFILE

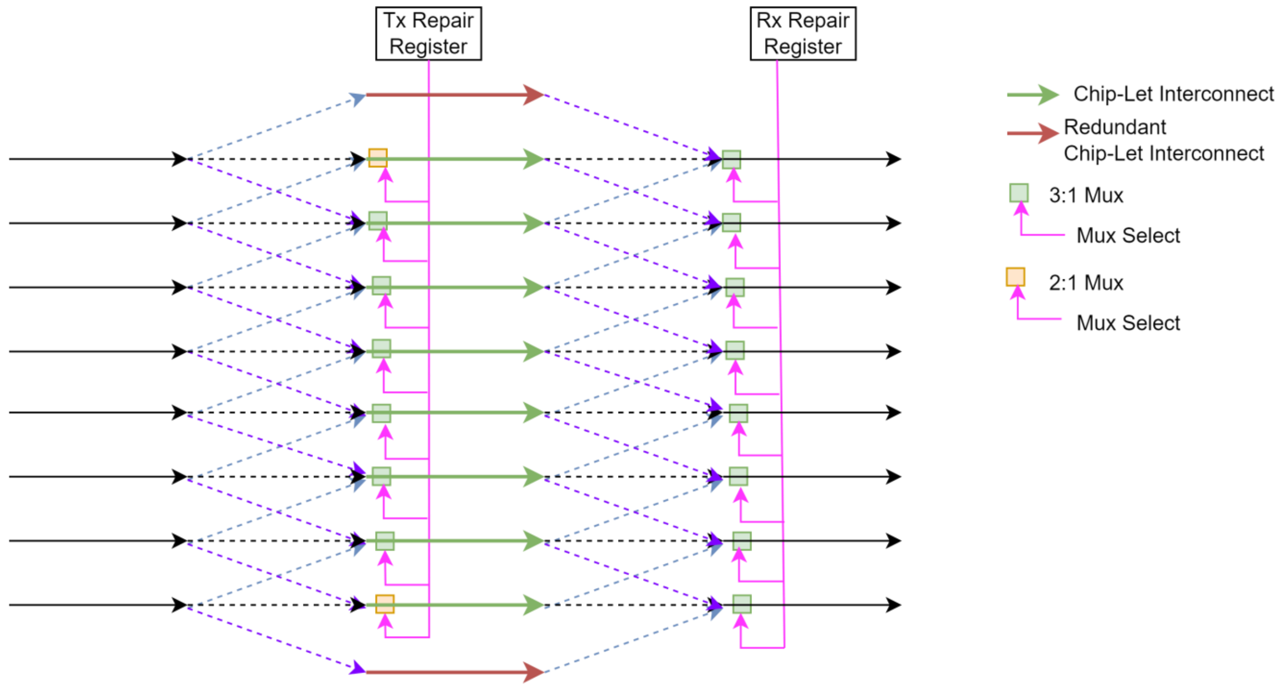
- Vcc/Vss shorts quite significant
  - Cannot be addressed during high volume manufacturing
  - Addressed through physical design rules
- Defects targetable during High Volume Manufacturing (HVM) Tests
  - ~1/3 Shorts between signal and Vcc/Vss
  - ~1/3 Shorts between 2-signal line
  - Slightly less than 1/3 multiple line shorts
  - ~5% opens
- Short between clock and signal
- Shorts across clusters
- Coupling Failures

Not covered by UCIe

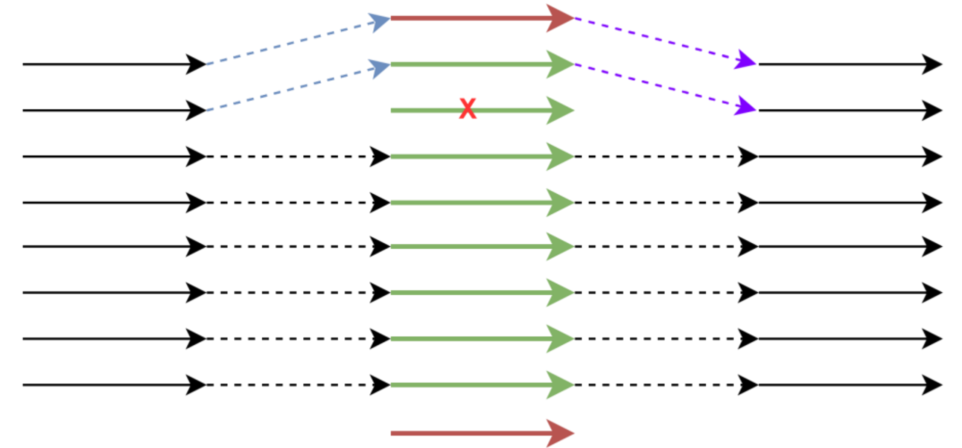
Not covered by HBM

Potential source of Test Escapes

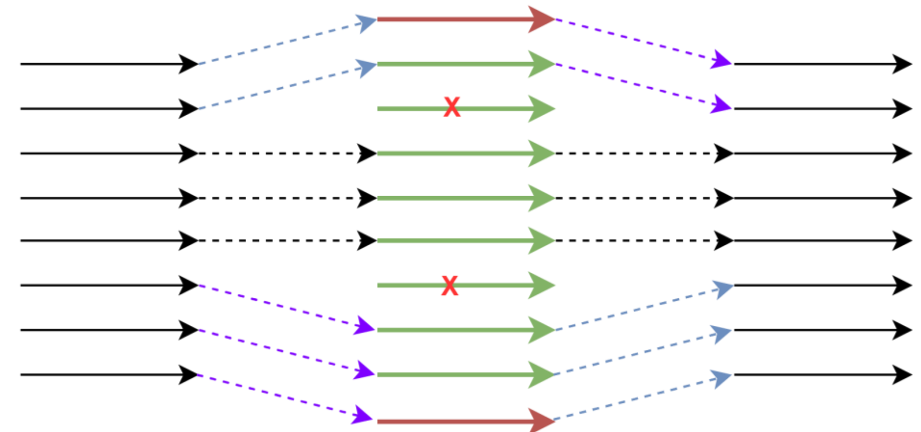
µBump Repair Idea, Two Redundant Chip-Let Interconnects



Recovering from 1 defective lane



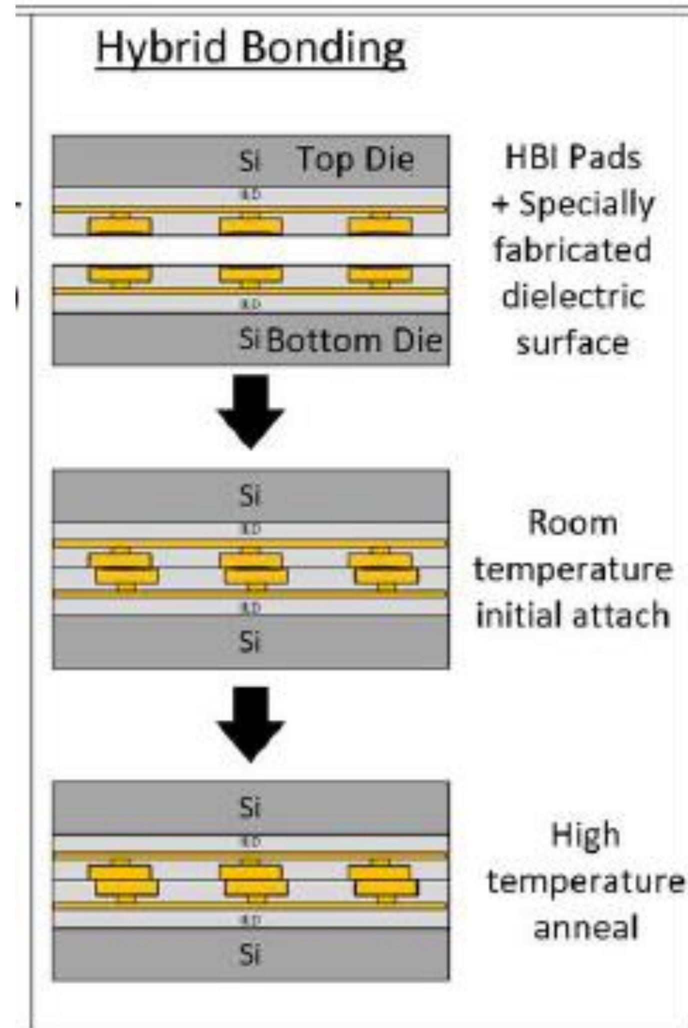
Recovering from 2 defective lane



- Alternate placement of redundant lanes are possible
- Clock and signal lanes can be shared/not-shared

Hybrid Bonding

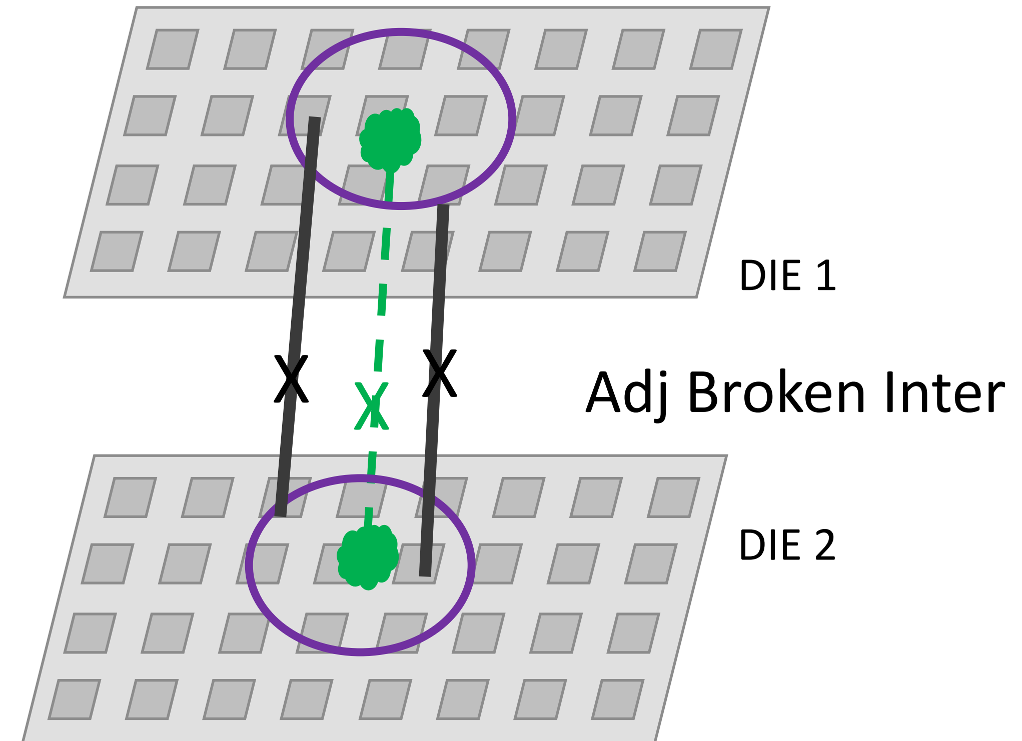
Source: EDM 2021 Intel Paper



AMPERE

HYBRID BONDING DEFECT PROFILE

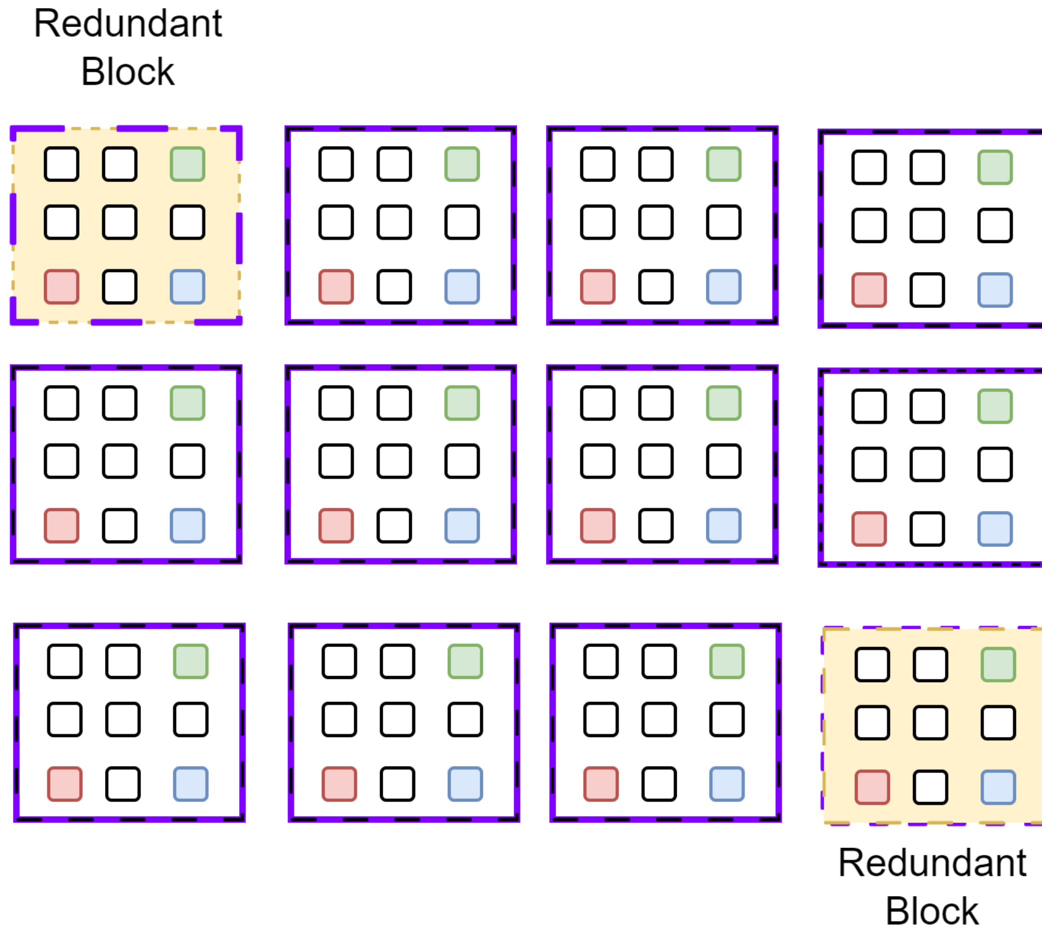
- Contaminants are the primary defect source
  - Surface near contaminant is concave/convex



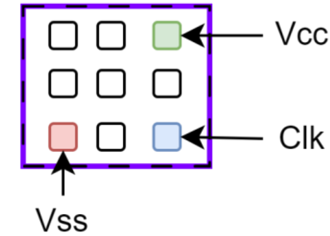
- Clustered Open Defects
- Not addressed by either UCIe or HBM

AMPERE

Hybrid Repair Idea for Clustered Opens



Block Definition



Each block is a collection of signals, Vcc, Vdd

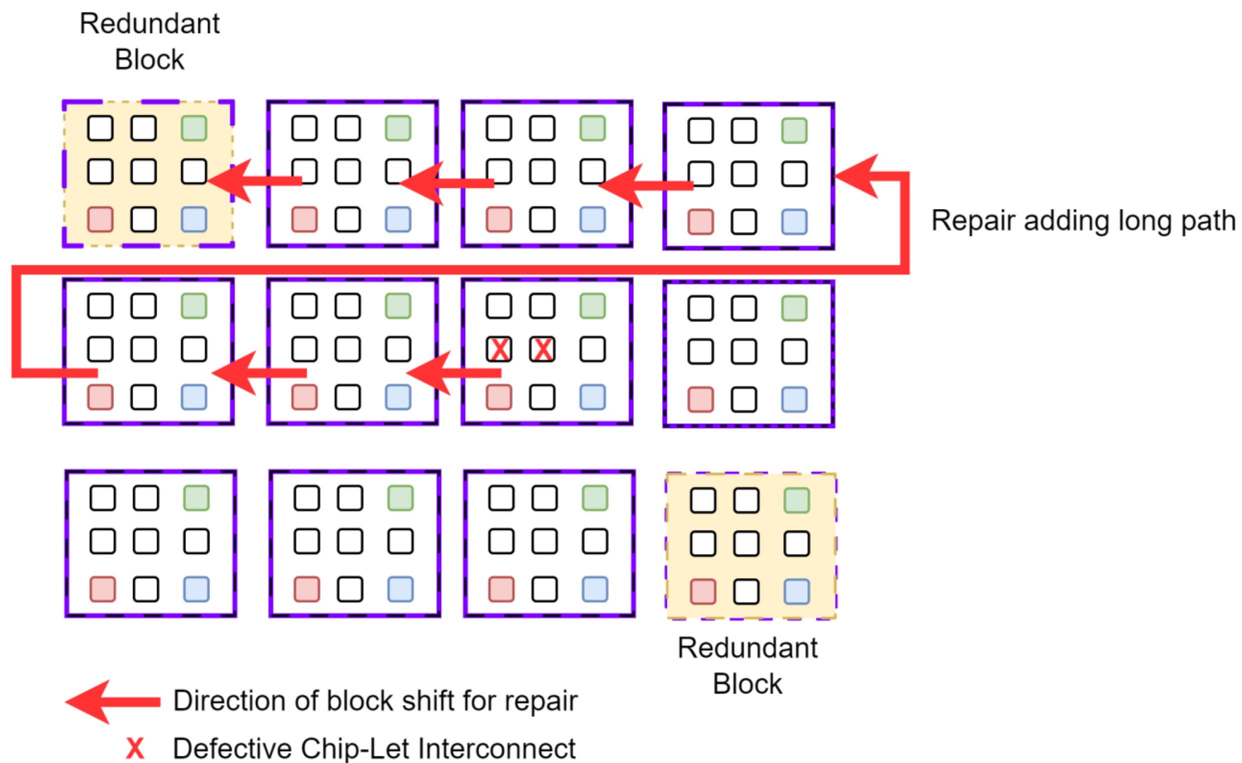
Redundnacy built into the layout
Primarily for Vcc and Vdd, sometimes for clocks

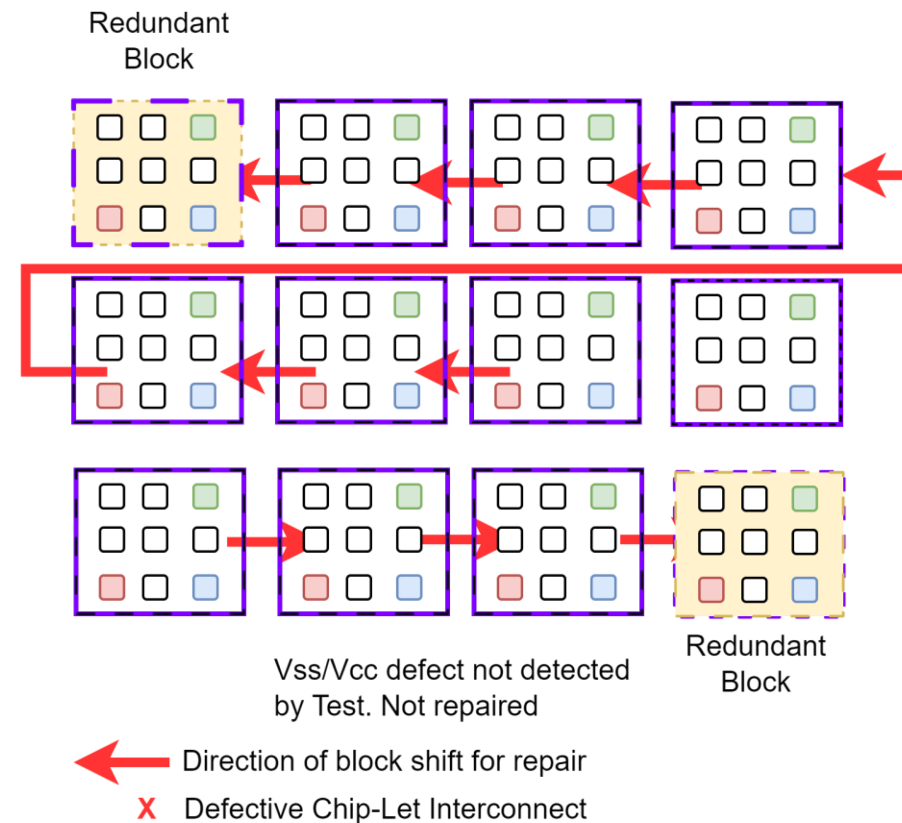
Don't add repair muxing logic to Vcc, Vdd, Clk
[Saves area]

- Assume: Simple Buffer Tx/Rx
- Can SerDes Tx/Rx be used for
very high-density interconnects? [<10µm]
- High repair-logic cost. Contained by reducing
block size and sacrificing yield.

Hybrid Repair Idea for Clustered Opens

Challenges



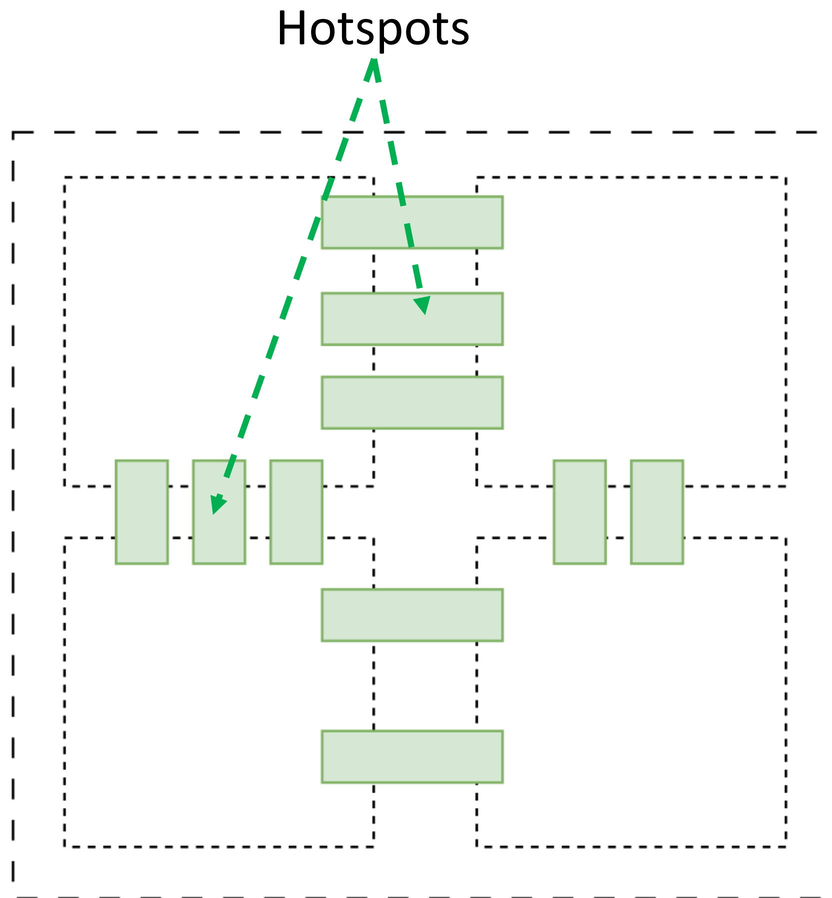


AMPERE

Agenda

- ✓Background and Motivation
- ✓Packaging Technology and Chip-Let Interconnect Defects
- ➢Chip-Let Interconnect Reliability Issues

Chip-let Interconnect Reliability Issues



- Typically, Chip-Let Interconnect regions, show up as hot spots in the SoC thermal map
- Why?
  - Power dissipation per unit area is high
    - Chiplet Tx/Rx run at a higher voltage
    - Capacitive load is high
  - Die shoreline area density is very high
- Impact of chiplet interconnects running hot
  - Chip-Let Interconnect defect aging
  - Device aging
- Aging leads to field failures

Marginal Defects and Defect Aging: µBump

High Temp Solder Migration

"Marginally bridge" → "Fully formed bridge"

Thermal Expansion of the contaminent and surface warping

"Marginally Open" → "Fully Open"

Passes Manufacturing testing

Results in field failures over time

Marginal Defects and Defect Aging: Hybrid Bonding

Redundant Block

Redundant Block

Direction of block shift for repair

Contaminent

Defective ("fully" open) Chip-Let Interconnect

Marginally defective (partially open) Chip-Let Interconnect

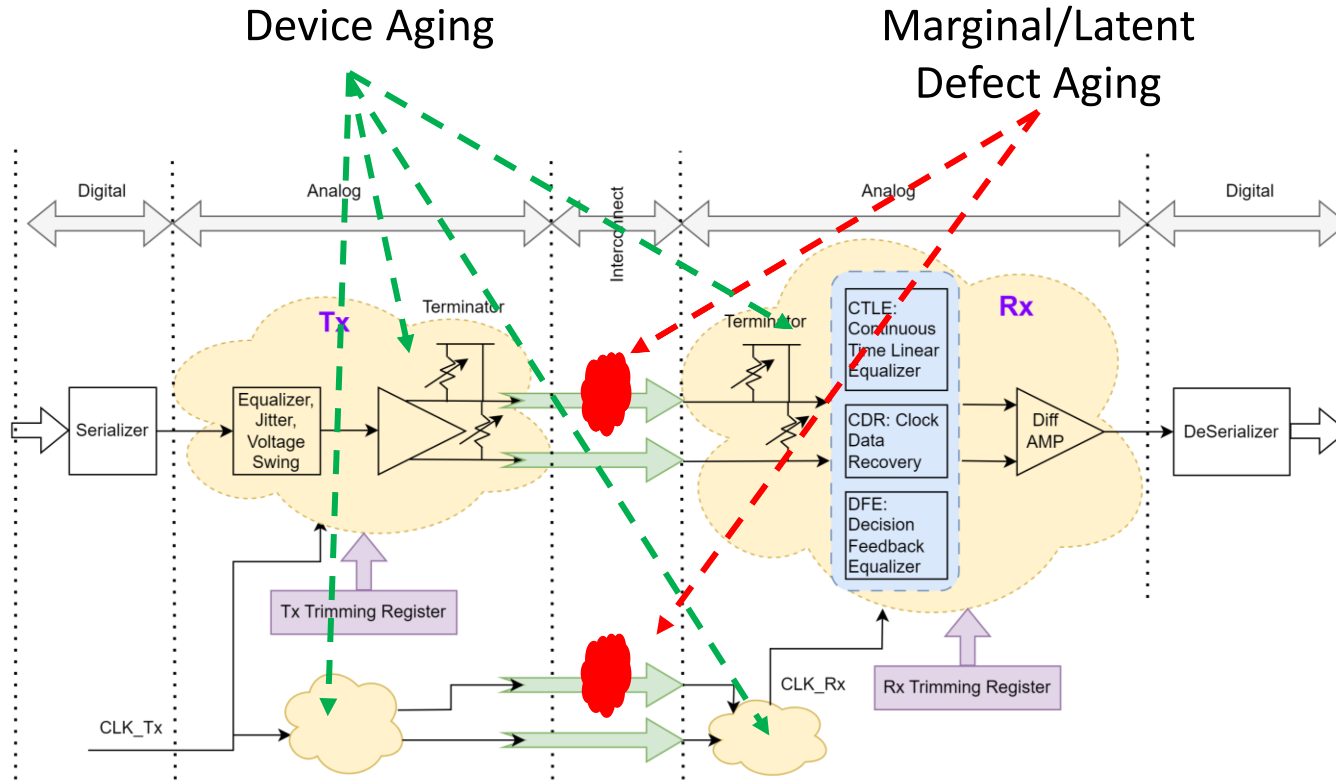
Latent Defects: Not detected at time-0 but will fail over time

Thermal Expansion of the contaminent and surface warping

"Marginally Open"

"Fully Open"

Device Aging of Chip-Let Interconnects



- Running Hot leads to device and/or defect aging
- Over time, latent defects will show up as field failures

Potential Mitigation Techniques

- Frequency adjustment using Temperature Monitors
  - Slow down clock frequency to cool down the interconnect area
  - Very challenging to add Temp Monitors in such dense area of the die
- Monitoring Signal Parameters
  - To identify if the interconnect is reaching its performance limits
  - Clock-jitter and jitter compensation
  - Clock duty cycle
  - Clock-data skews
  - These IPs have their own area overhead
  - Predicting failure from the collected data needs a data analysis infrastructure

Potential Mitigation Techniques

- Test and Repair on Bootup/powerup/power-down
  - Feasible for products that are turned on and off frequently: Automotive applications, smaller devices
    - The number of interconnects are relatively small
  - Not effective for large servers and AI engines used in data centers, AI model build etc.
    - These systems are used for very long running jobs
    - Large time-interval between successive bootup
    - The number of interconnects are large
- Periodic Test and Repair
  - Periodically pause data transfer across chip-let
  - Exercise the test and repair infrastructure to fix any issues or impending issues
  - Impacts performance

Thank you!
Q&A