Failure Analysis in the AI Era: Ensuring Reliability at Scale

James Chambers, Ph.D.

VP, Silicon Engineering and Sourcing, NVIDIA

Turning Sand into Intelligence

Powered by NVIDIA Engineering

SAND

Silicon Dioxide

INTELLIGENCE

AI, ML, Accelerated Compute

Six Decades of Machine Learning

ML requires a substantial amount of compute resources

Arthur Samuel

Machine Learning Checkers

IBM 701

4K Vacuum Tubes / 9.0KB Memory

1959

AlexNet

Machine Learning Image Recognition

Two NVIDIA GTX 580 GPUs

6.0B Transistors / 6.0GB Memory

2012

Both applications required the most-advanced compute capabilities available at the time

One Decade of Machine Learning and Artificial Intelligence

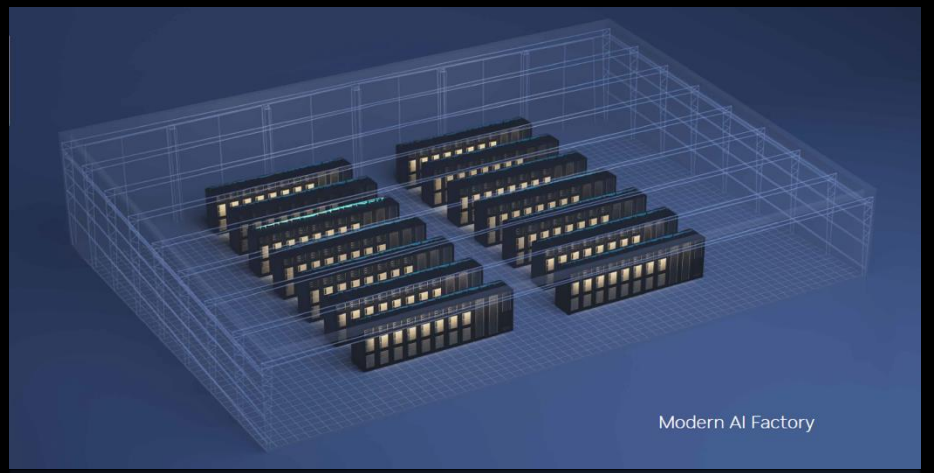
Demand for computational resources has grown significantly and shows no sign of slowing

AI/ML models are getting larger and larger

Not a Wave → A Revolution

Generative AI is the 5th Industrial Revolution

Coal



Tokens

1765	Steam Engines	→	Mechanical Work
1870	Internal Combustion Engines	→	Portable Work
1964	Mainframe Engines	→	Computational Work
1995	CPU Engines	→	Transactional Work
2022	GPU Engines	→	Intellectual Work

A NEW INDUSTRIAL REVOLUTION

AI FACTORY

TOKENS
"UNITS OF INTELLIGENCE"

$100T

SCIENCE

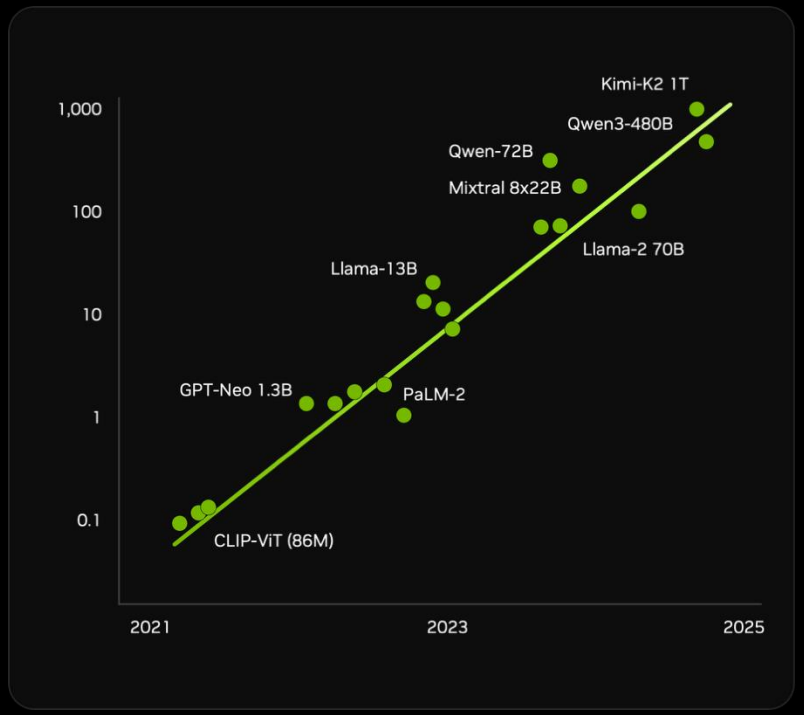
ENTERPRISE

HEALTHCARE

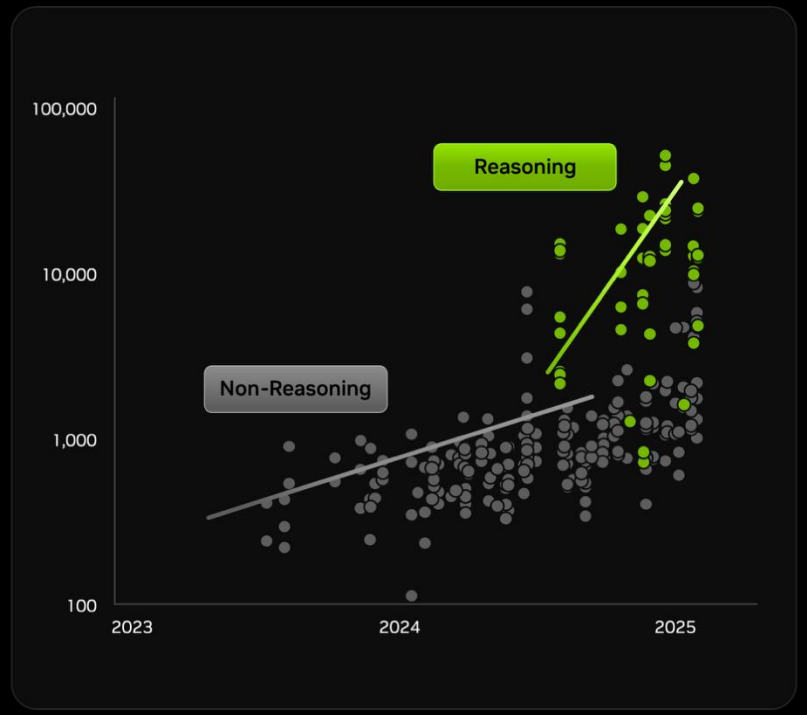
MANUFACTURING

Insane Demand for AI Computing

Model Size Growing
10X Parameters Per Year

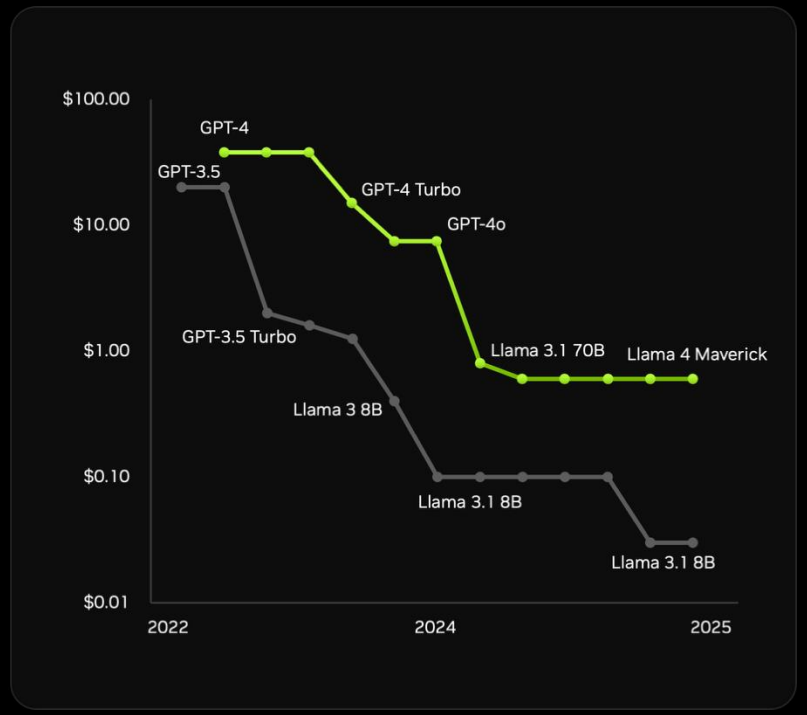


Test-Time Scaling "Thinking"
5X Tokens Per Year



Source : EPOCH AI

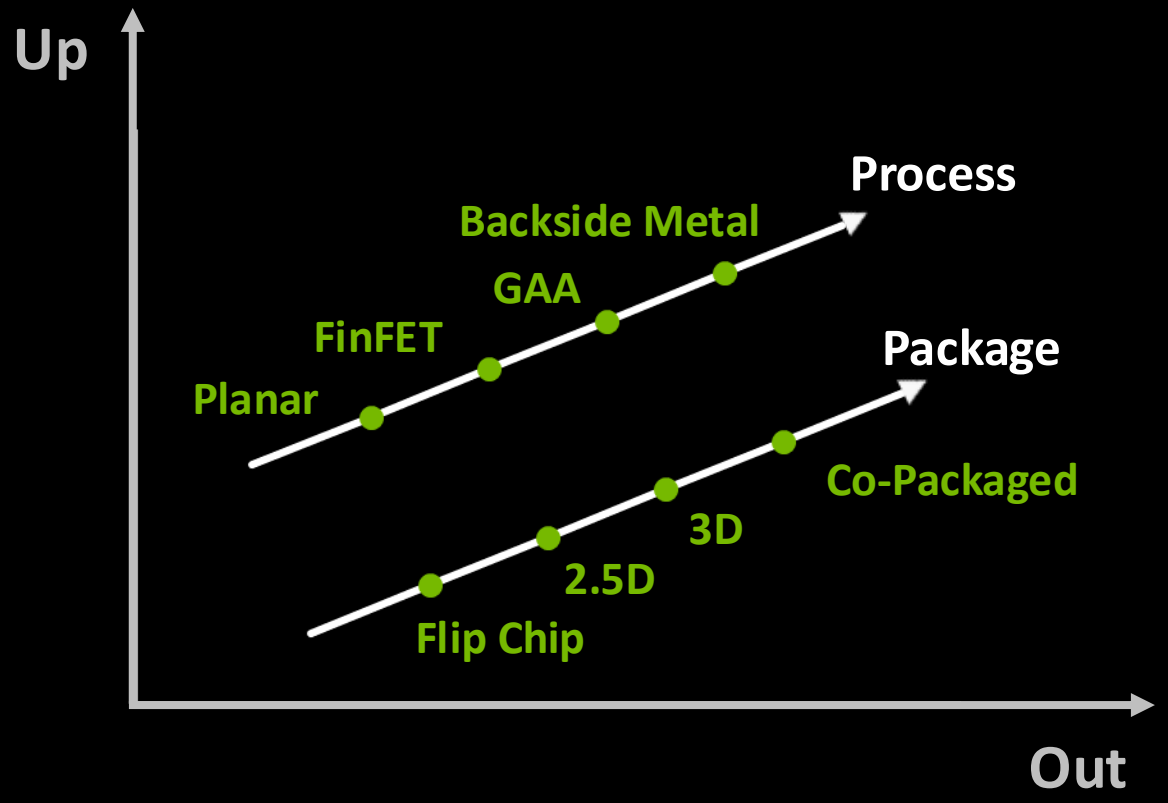
Token Cost
10X Cheaper Per Year



Source : Artificial Analysis

Process and Packaging Technologies Continue Their Advance

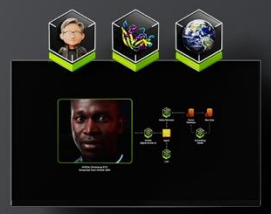
New materials, processes, and integrations adding complexity



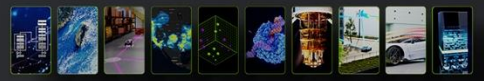
Continuous innovations needed from semiconductor ecosystem

NVIDIA Pioneering Accelerated Computing

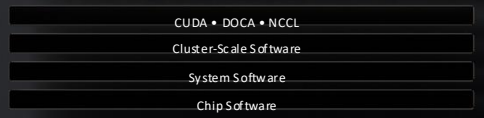
NIM
CUDA-Accelerated
Agentic AI Libraries



Omniverse
CUDA-Accelerated
Physical AI Libraries



CUDA-X Libraries

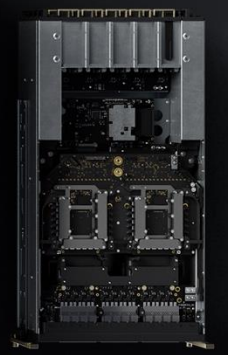


Accelerated
Software Stack

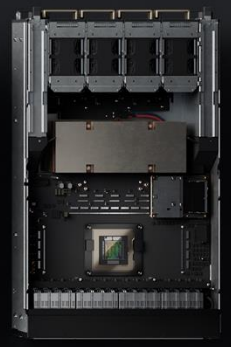
GB200 NVL72 SuperPOD

Grace Blackwell
MGX Node

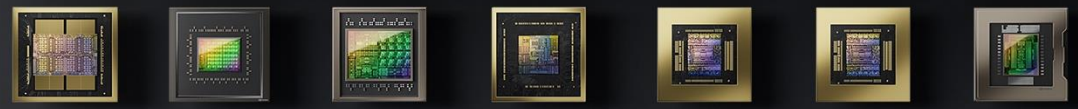
NVLink Switch



Quantum Switch



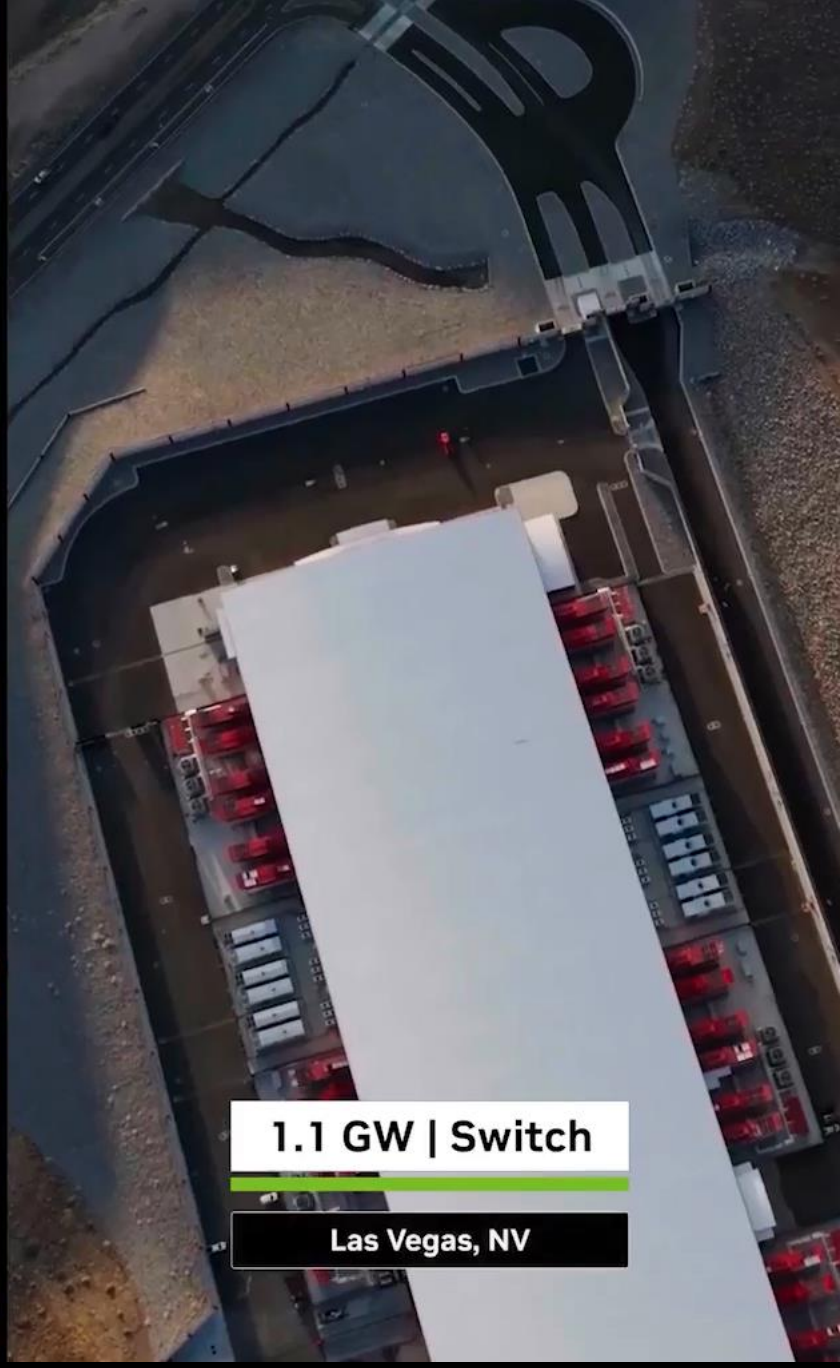
Spectrum-X Switch



Chips Purpose-Built for AI Supercomputing
GPU | CPU | DPU | NIC | NVLink Switch | IB Switch | ENET Switch


1.2 GW | Stargate
Abilene, TX

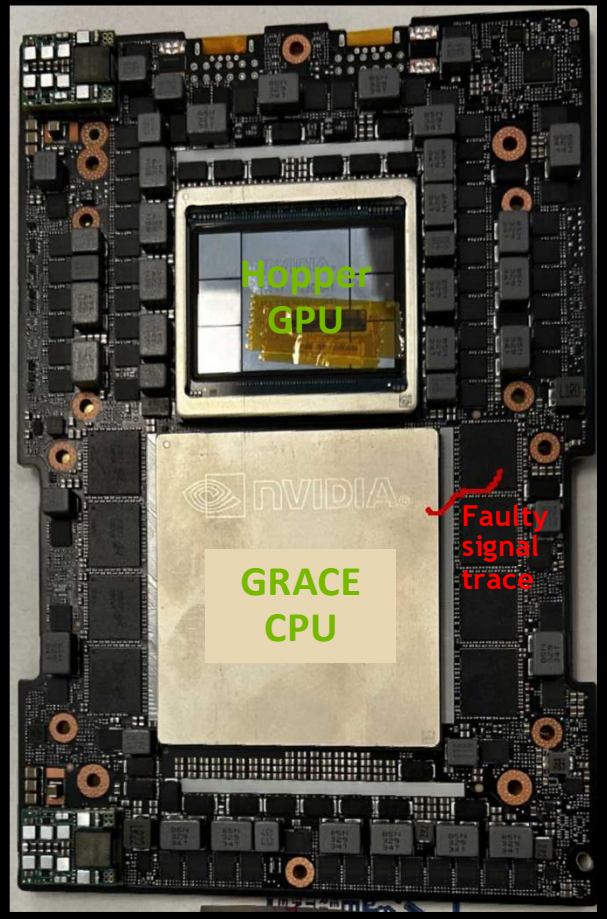

2 GW | Switch
Cartersville, GA


1.1 GW | Switch
Las Vegas, NV

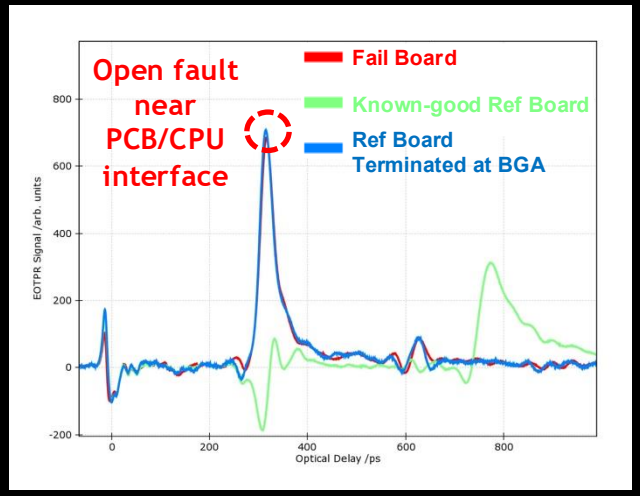
Advances in Process and Packaging Require Innovation in FA

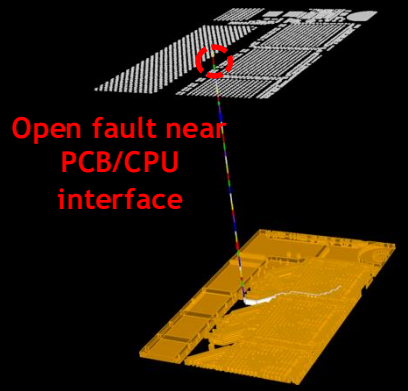
Signal traces across complex system poses challenges in FI/FA

Grace Hopper Superchip

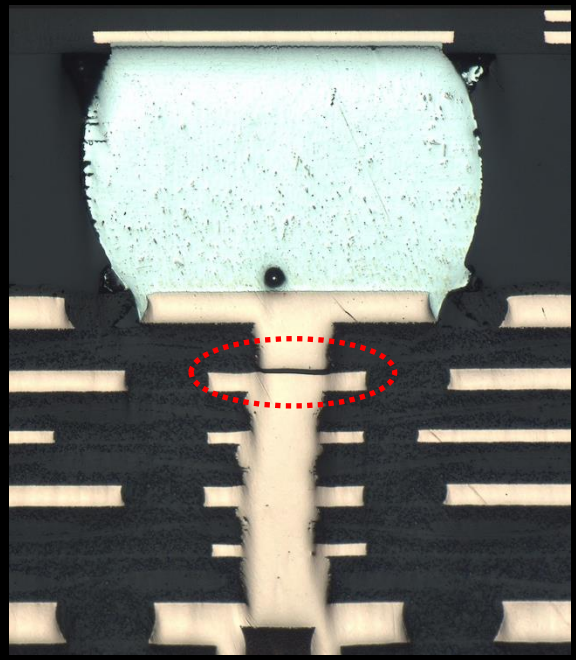


EOTPR Isolation





Physical Analysis

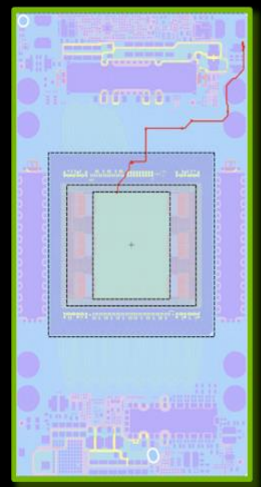


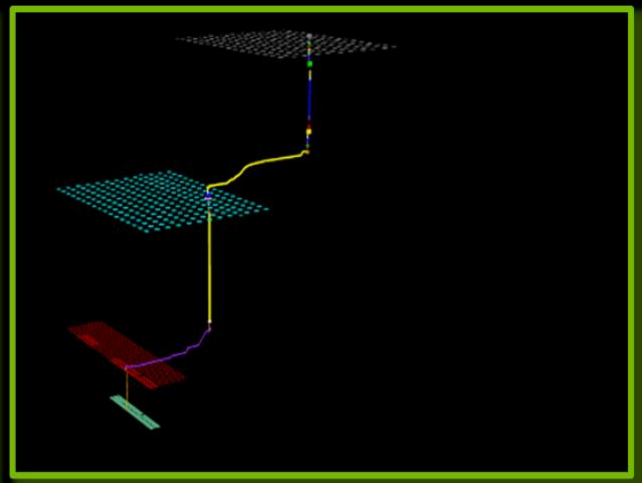
- Larger boards/systems with signal trace across complex system components require more accurate fault isolation

Advances in Process and Packaging Require Innovation in FA

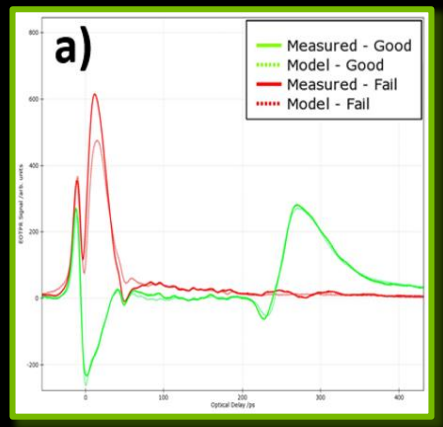
System-Level EOTPR

Full system traceability, model generation, 2D/3D Visualization

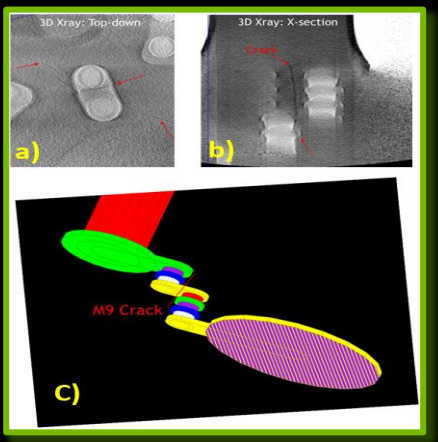




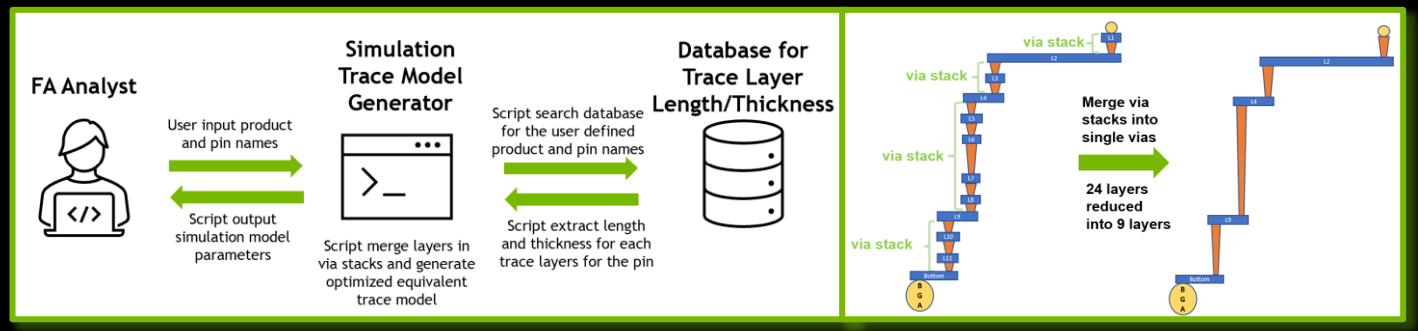
EOTPR Measured vs. Model



Root-Cause Analysis



Automated Trace Model Generation



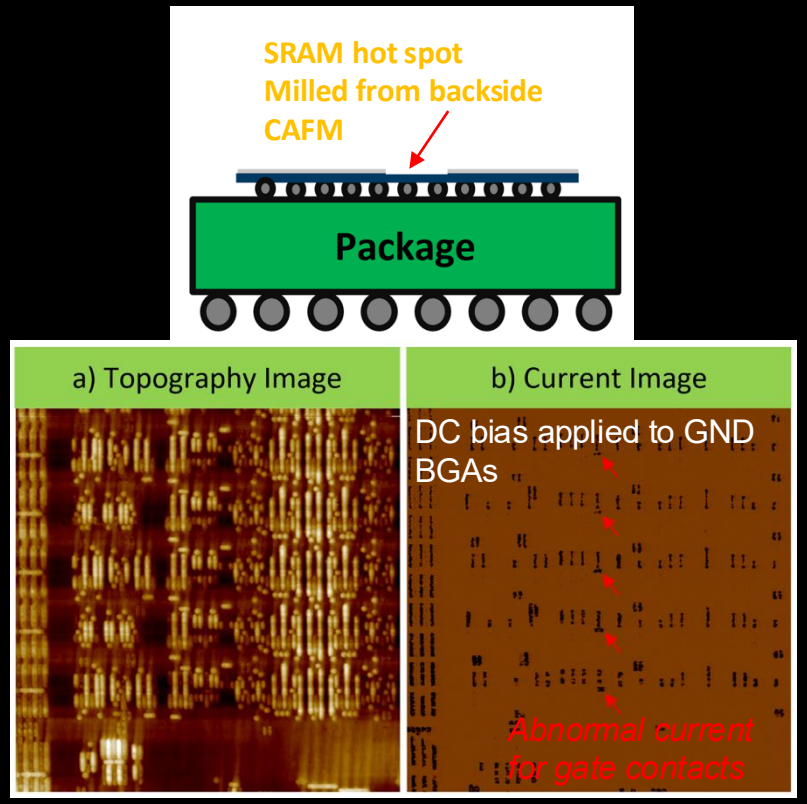
Advanced Package Fault Simulation: The Impact of Accelerated Trace Model Generation, Chuan Zhang, et al., ISTFA 2024
Hyper-Integrated CAD Database for One-Shot EOTPR Fault Localization in Heterogeneous Sip-on-Board Assemblies, Arpan Bhattacherjee, ISTFA 2025

- Emphasize non-destructive testing at both package-, board-, and system-levels

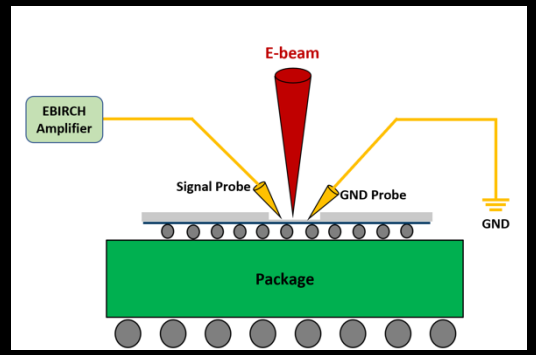
Advances in Process and Packaging Require Innovation in FA

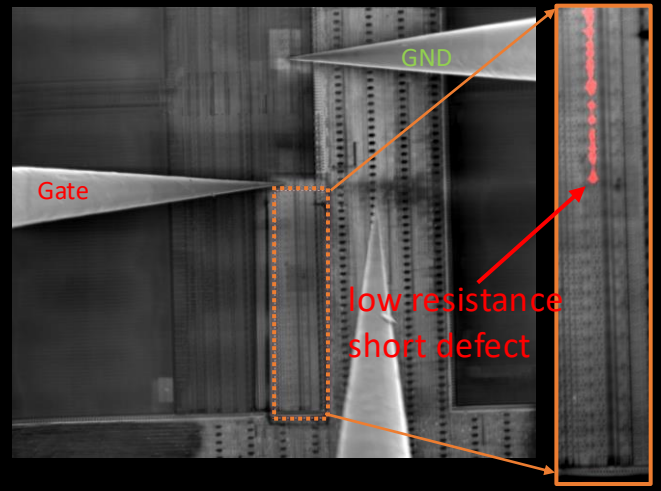
3D devices require 3D probing approaches for defect localization: Backside

Backside CAFM

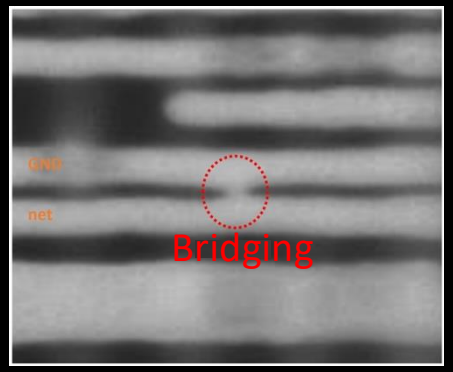


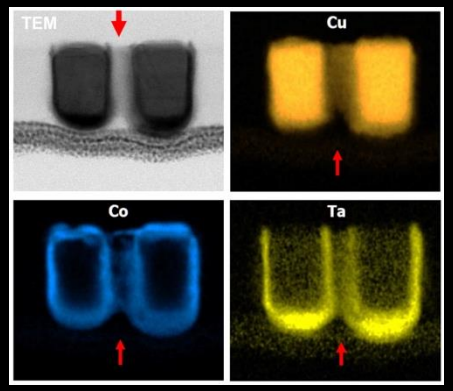
Backside EBIRCH (nanoprobing)





Backside SEM/TEM





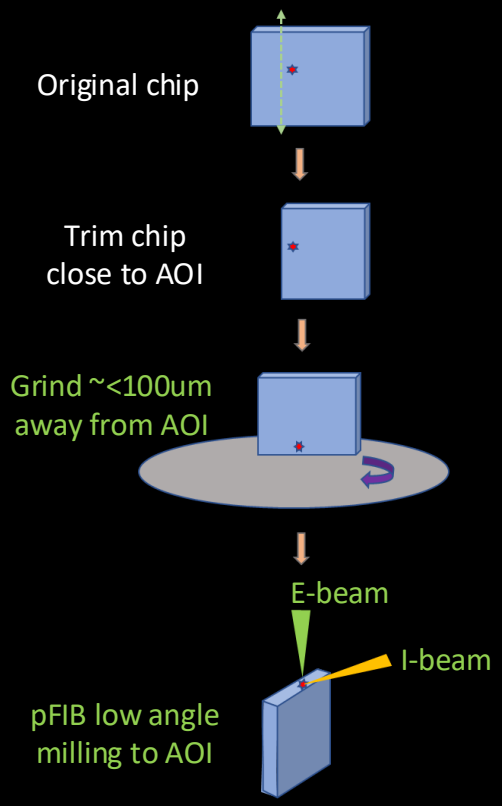
"Backside EBIRCH Defect Localization for Advanced Flip Chip Failure Analysis", Chuan Zhang, et. al., ISTFA, 2021

- CAFM/nanoprobing from different directions (back-side as opposed to conventional front-side)

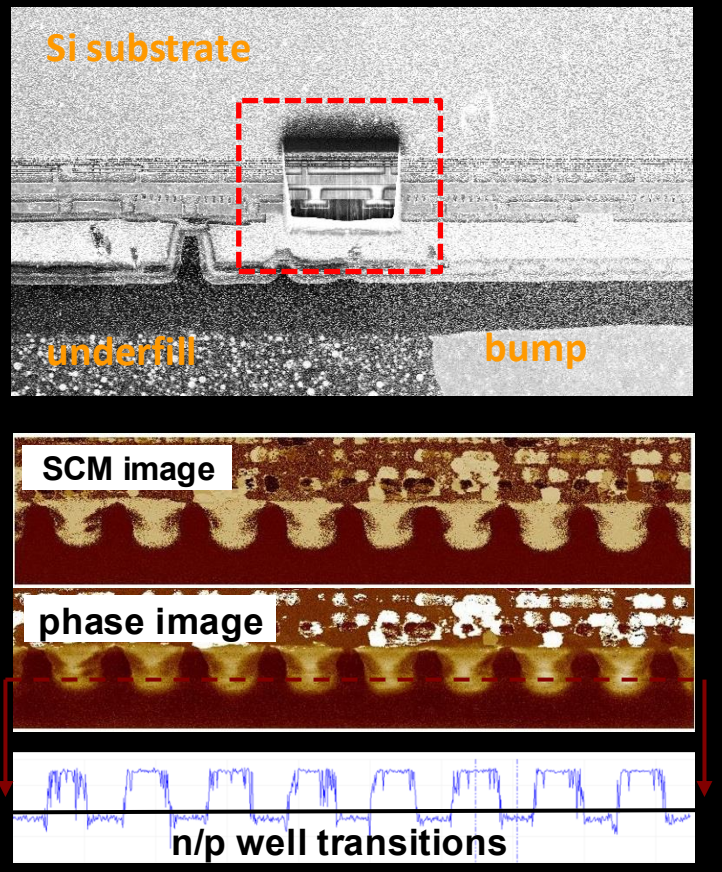
Advances in Process and Packaging Require Innovation in FA

3D devices require 3D probing approaches for defect localization: Side

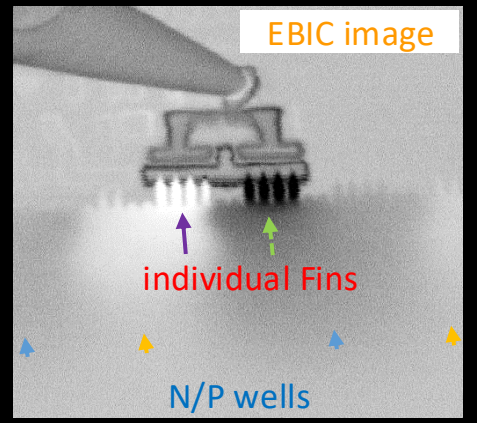
Site-Specific Cross-Sectioning

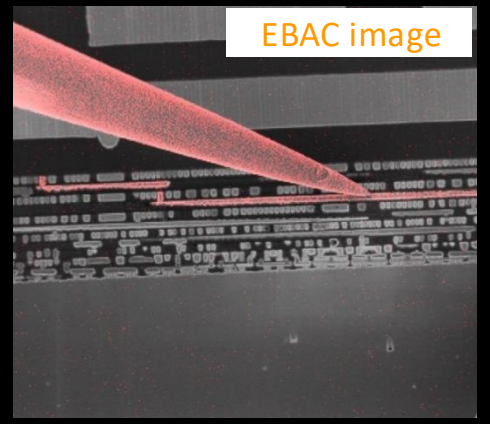


Cross-sectional AFM Analysis



Cross-Sectional Nanoprobing Analysis





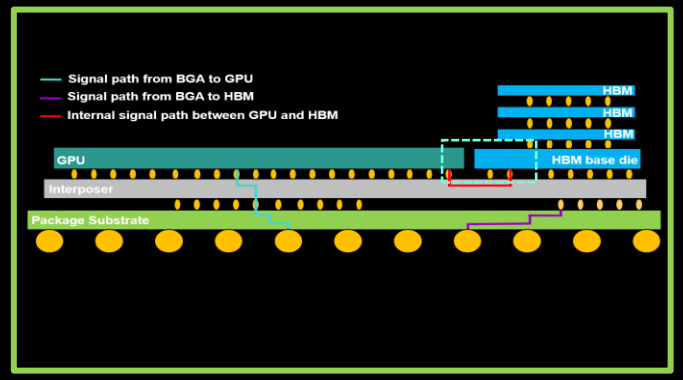
"Site-specific Low Angle Plasma FIB Milling for Cross-sectional Electrical Characterization", Chuan Zhang, et. al., ISTFA, 2019

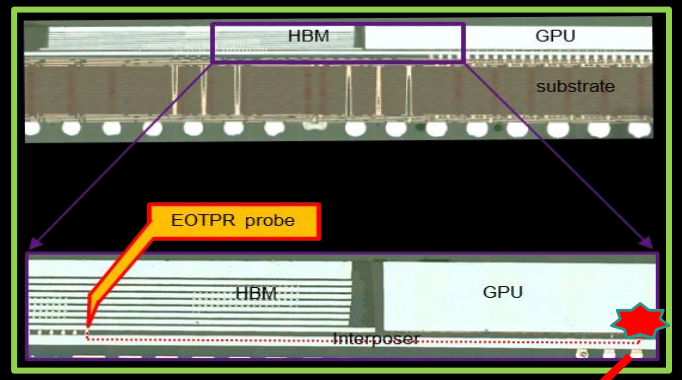
- AFM/nanoprobing from different directions (Front-side, back-side, cross-sectional probing)

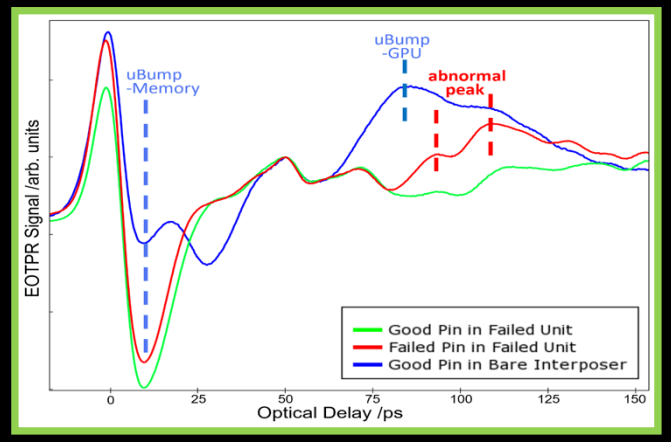
Advances in Process and Packaging Require Innovation in FA

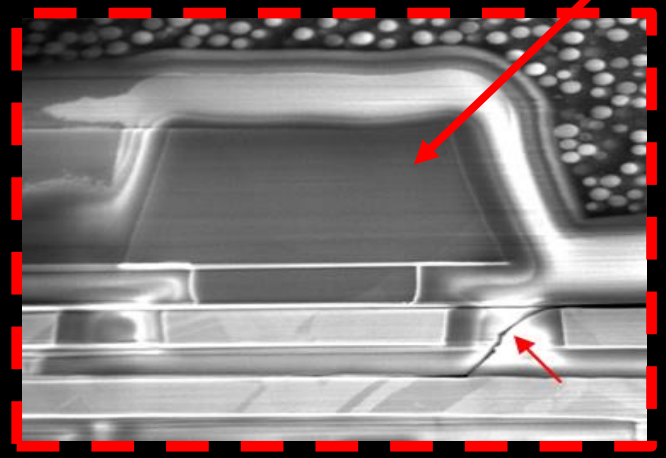
3D devices require 3D probing approaches for defect localization: Side

Cross-sectional EOTPR to localize defect in internal signal path between GPU and HBM









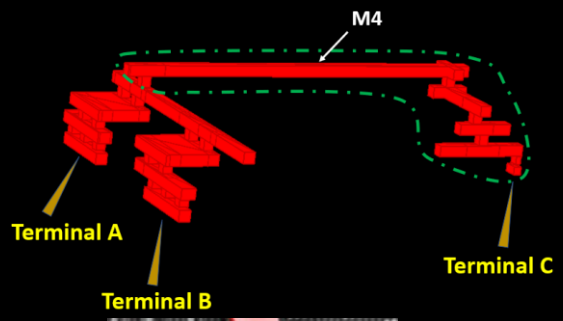
"Localization and Characterization of Defects for Advanced Packaging Using Novel EOTPR Probing Approach & Simulation, ISTFA, 2020", Chuan Zhang, et. al., ISTFA, 2020

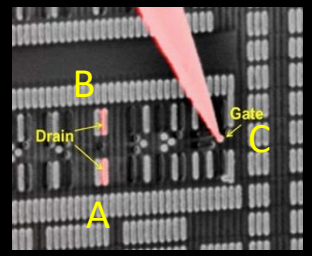
- 2.5D/3D packaging requires micro-probing/EOTPR probing from different directions (Front-side, back-side, cross-sectional probing

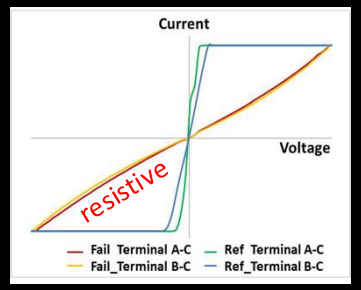
Advances in Process and Packaging Require Innovation in FA

3D devices require 3D probing approaches for defect localization: All Sides

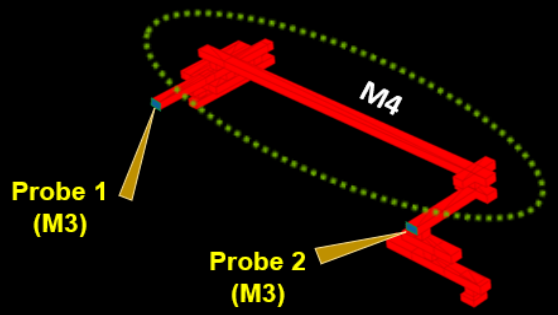
Backside nanoprobing

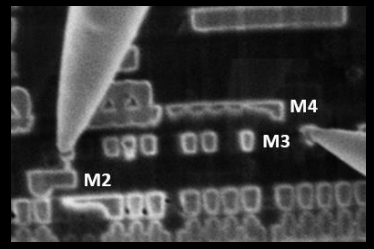


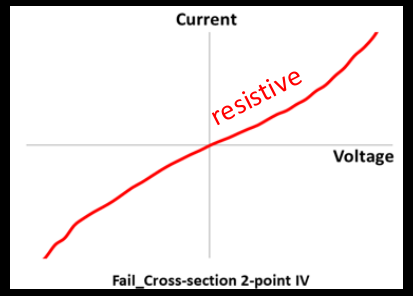




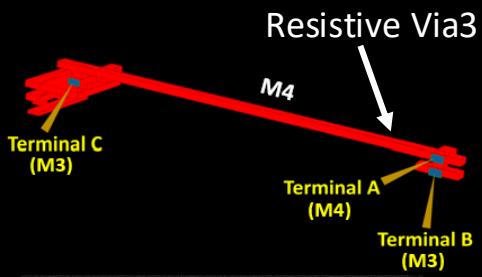
Cross-sectional nanoprobing

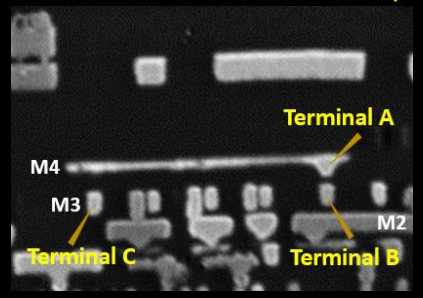


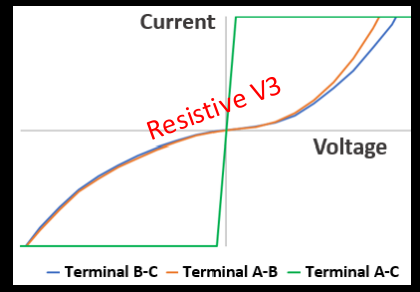


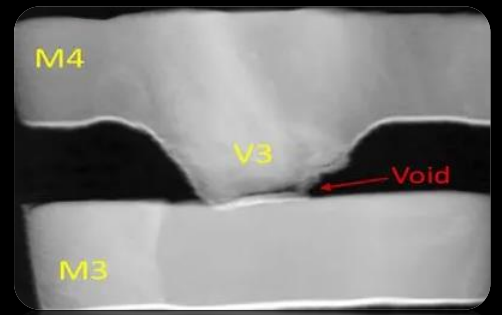


TEM Analysis









"Localizing IC Defect Using Nanoprobing: A 3D Approach", Jane Li, et. al., ISTFA, 2019

- Nanoprobing from different directions (Front-side, back-side, cross-sectional probing)

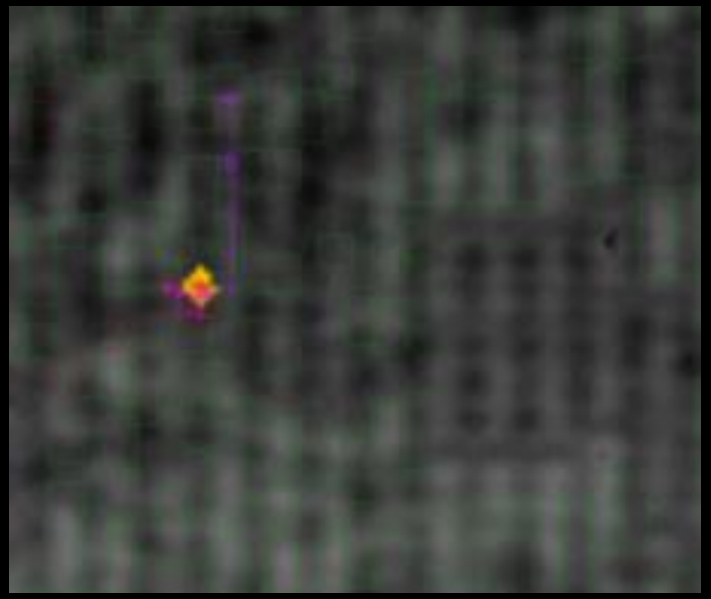
Golden Age of Optical Fault Isolation

Photon Emission, Laser Stimulation, and Laser Voltage Probing/Imaging

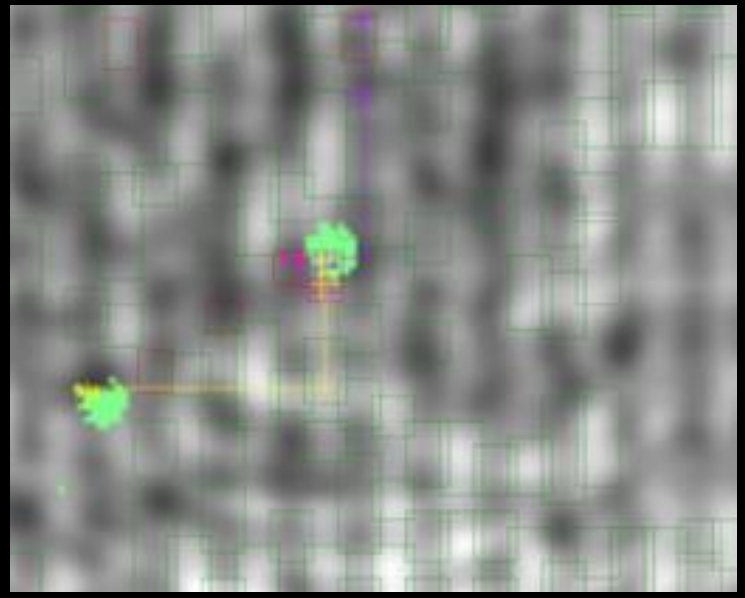
- Flip-Chip, backside access, and Si transparency in the NIR enabled Fault-Isolation capabilities for 20+ yrs
- Advancements in SIL technology, reduction in LVP wavelengths, and "alignment tricks" have kept up with process technology scaling

Advanced node GPU imaged using 3.3 NA broadband SIL

PEM (LN2 cooled InGaAs)

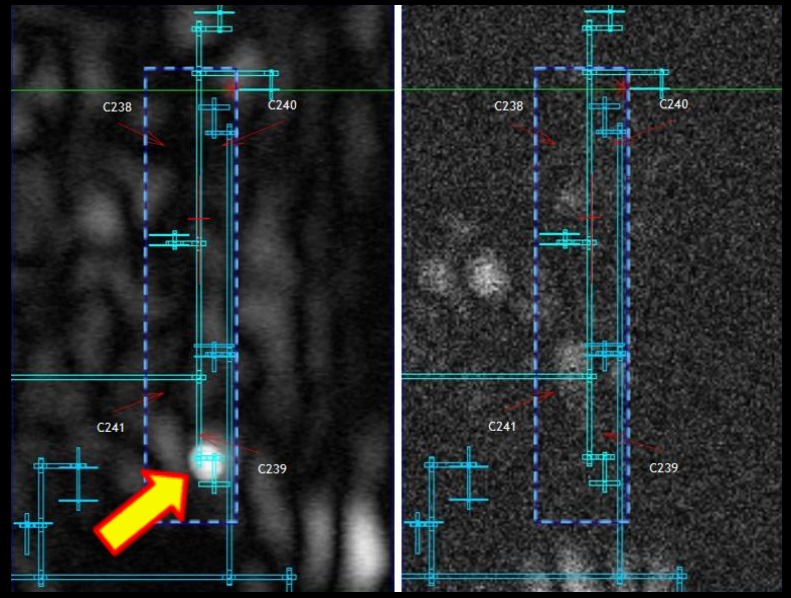


DLS 1064nm Laser, 1.5mW



1064nm LVi & LVP are still very useful!

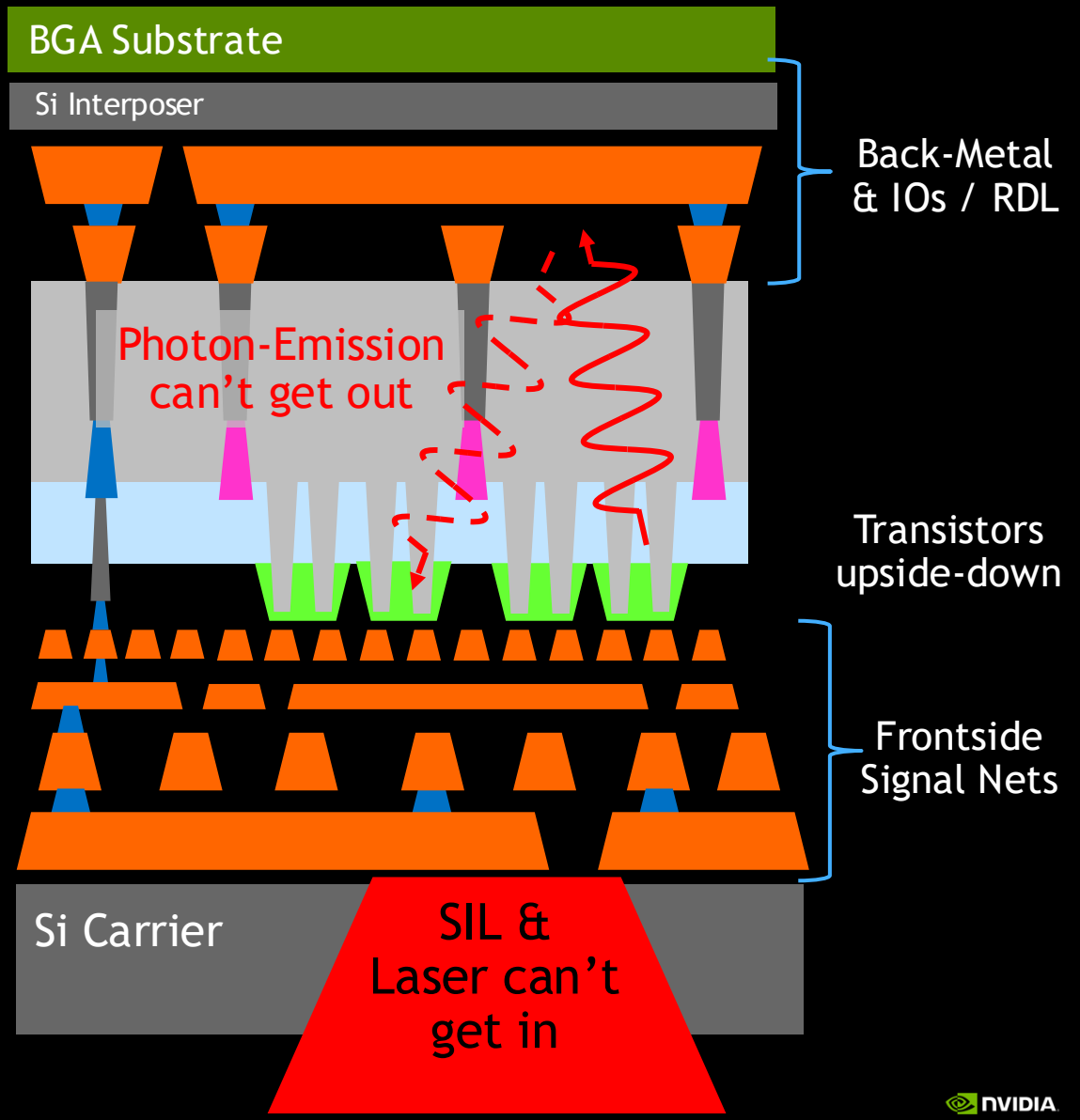
Data-Modulation Clk-Modulation



The Dawn of a New Age of Fault Isolation

Backside metal and die-stacking block optical access via SIL

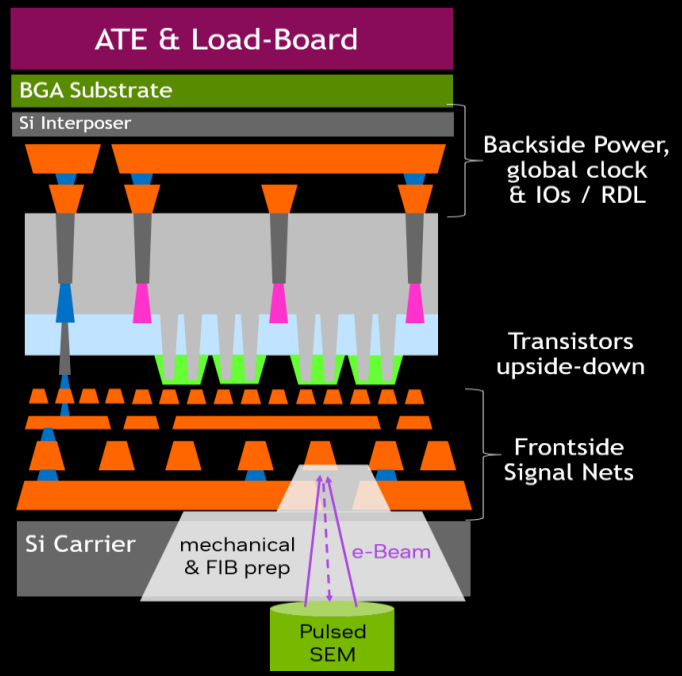
- Fault Isolation in the process of being re-invented
- Most promising techniques
  - Lock-in Thermography
  - Thermal-Laser-Stimulation
  - E-Beam Probing
  - Focused X-Ray Assisted Device Alteration



FI Challenge

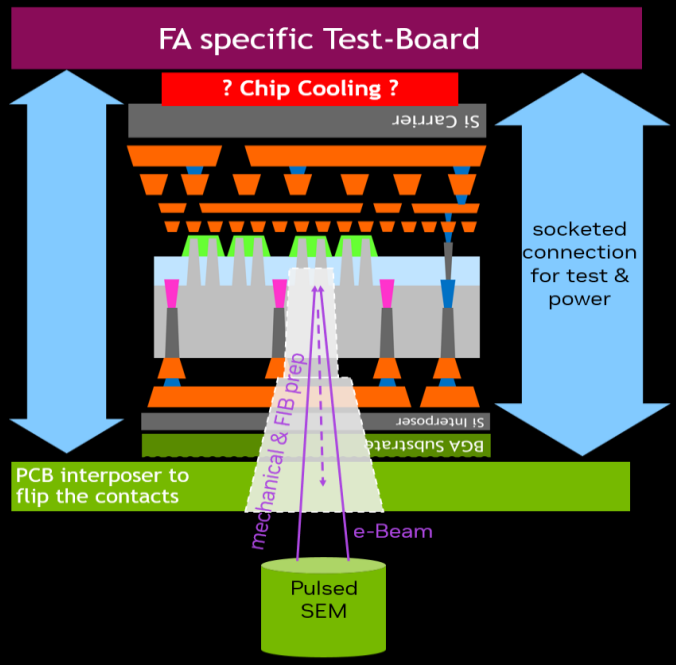
Meet, then beat the success rate and cycle time of optical FI

Front-side EBP



- Probing on signal nets is viable if access can be created
- Heavily dependent on FIB circuit edit
- Requires CAD development to direct access locations
- Need design help to improve diagnosability and ease signal access where possible (e.g. signal risers)

Back-side EBP



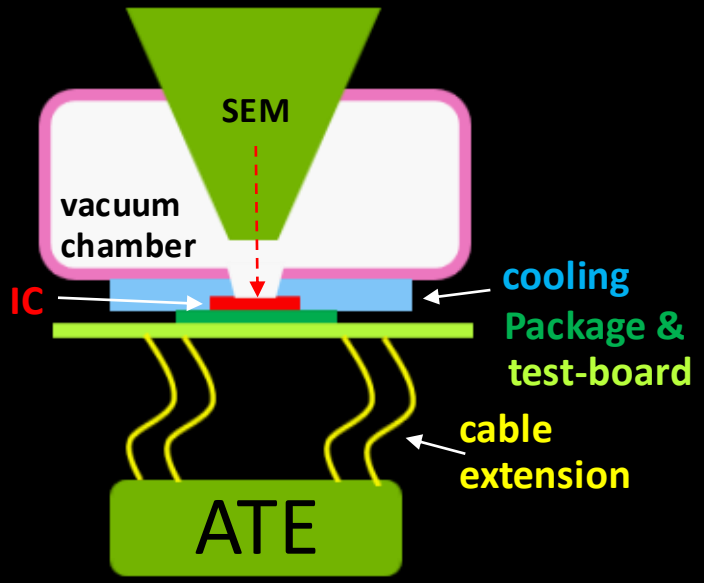
- Sufficient resolution for single FET logic state mapping
- FET access will be much more difficult & limited, requiring "re-packaging", complex mechanical, circuit edit
- Access may be blocked by vital signals on substrate, interposer, RDL, and back-side metallization

- Sample-preparation, DUT-stimulus, and in-vacuum DUT cooling require significant developments

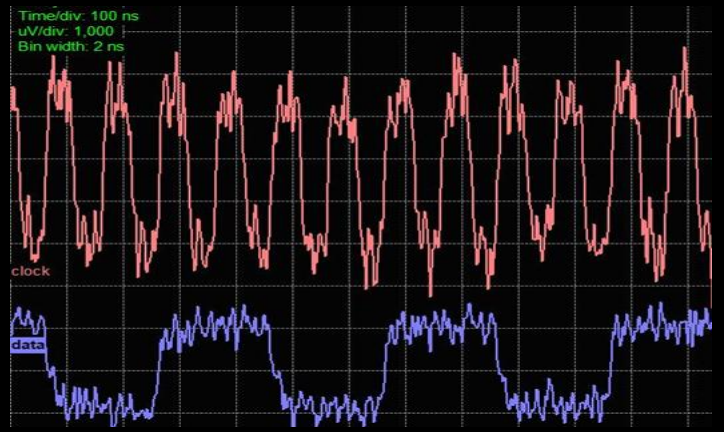
Logic Fault Isolation on Backside Power ICs

No more optical access to the FETs

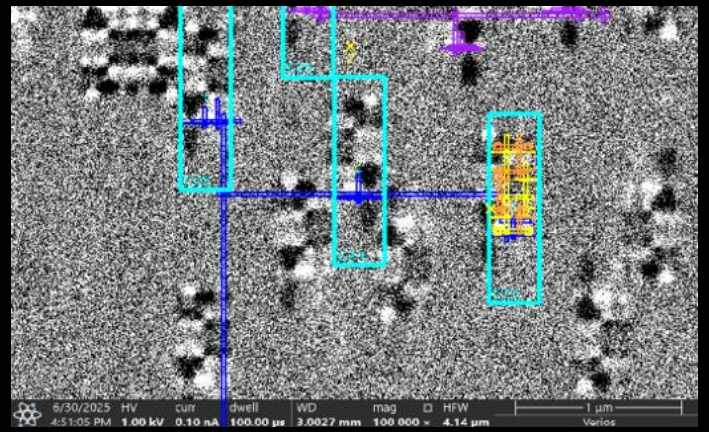
- Metal layers above and below FET layer block access for Laser-Voltage-Probing & Photon Emission
- E-Beam-Probing can partially replace LVP
  - Signal access ➔ small FIB holes without affecting the test of interest
  - DUT test & cooling ➔ ATE, test-board, IC & cooling layer dock to eBP vacuum chamber
  - Significant HW & workflow development ongoing



eBP Waveforms (like LVP)



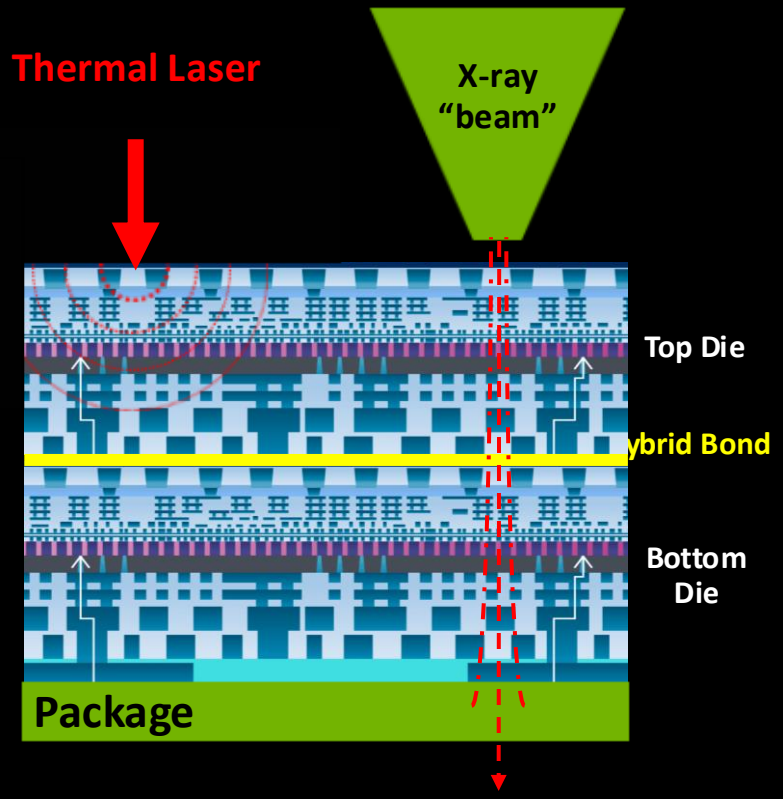
Frequency mapping on exposed bottom of FETs (die-down)

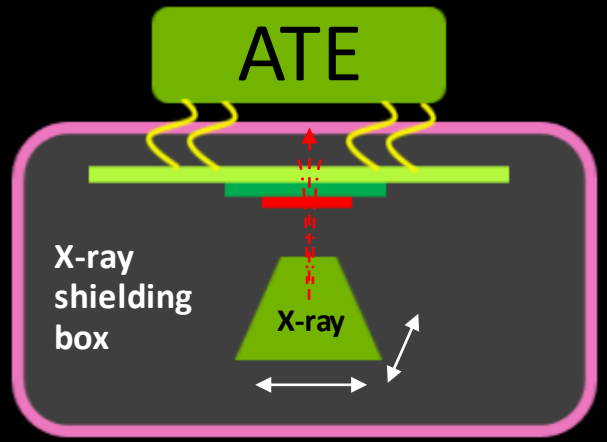


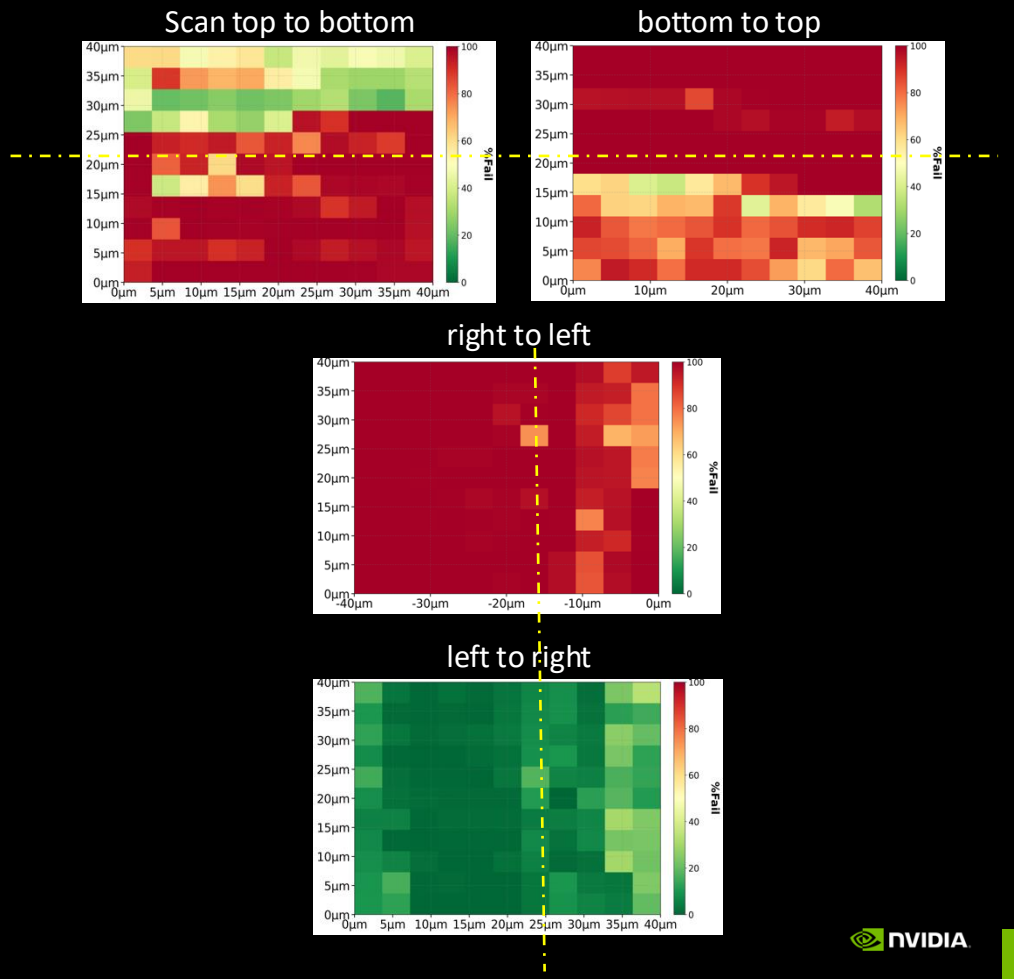
Logic Fault Isolation on Backside Power ICs

No more optical access to the FETs

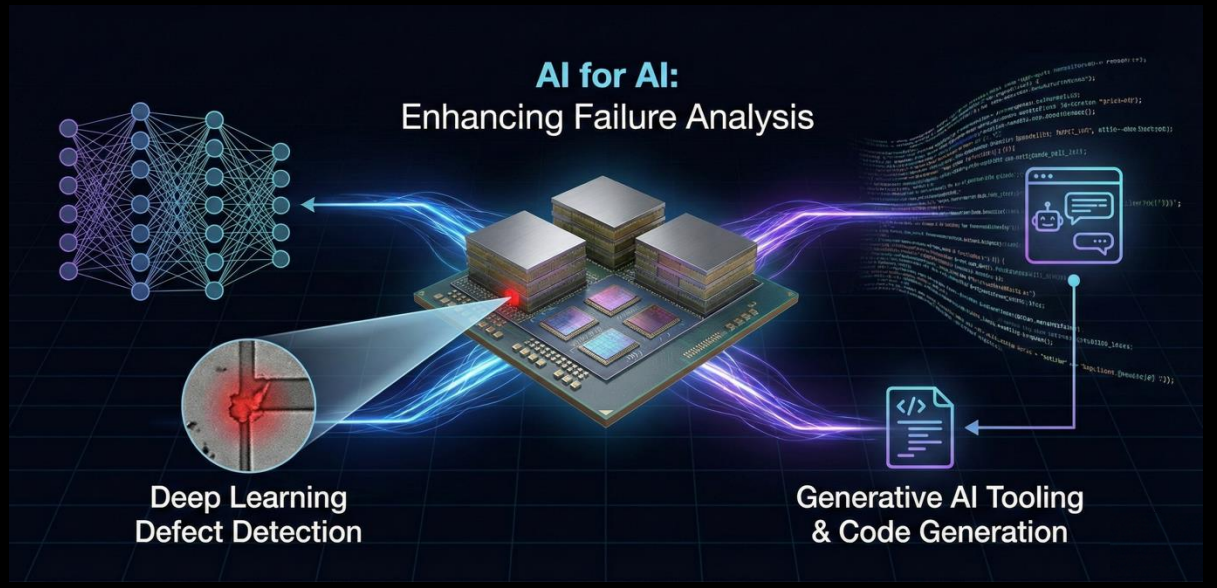
- Thermal Laser Stimulation success reduces with distance from top surface
- Scanning-Focused X-ray (aka XADA) promises µm level resolution even for stacked logic die
- Requires test inside X-ray shielding box
- Most of X-ray perturbation persists (unlike LADA/TADA)







AI Empowered Failure Analysis of AI Silicon



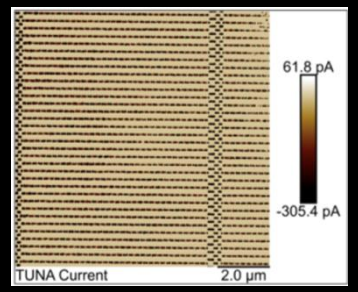
- The FA Challenge: Complexity at Scale
  - AI chips now scale to trillions of transistors using 2.5D/3D packaging
  - Finding tiny, subtle defects hidden among trillions of transistors defies traditional manual FA methods
  - Human-centric microscopy inspection is too slow and inefficient for AI data center scale chip FA

- The Solution: AI-Empowered FA
  - Deep Learning: Automates visual defect recognition (Computer Vision)
  - Generative AI: Autonomously writes the software apps to capture these defects

Compute-Intensive AI Workloads Uncover Even the Most Subtle of Silicon Defects

AI for Defect Detection

CAFM Current Image

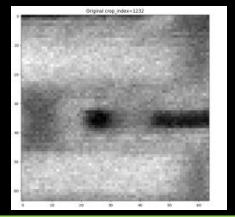


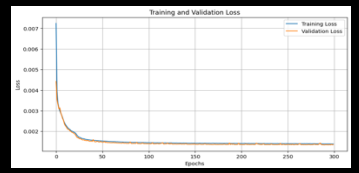
Autoencoder: Unsupervised deep learning algorithm

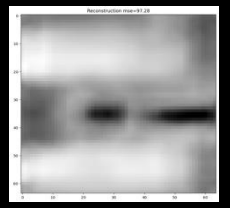
input ◄ Ideally X ≈ X' ► Reconstructed input

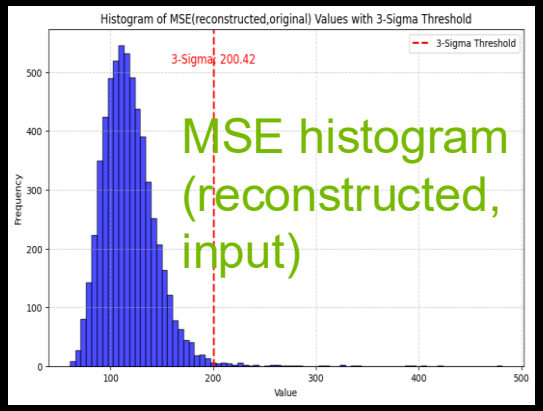
X → Encoder g_φ → LF → Decoder f_θ → X'

Latent Features
Compressed low dimensional representation of the input

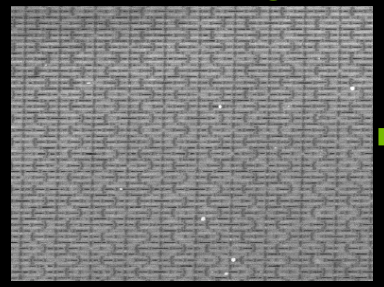






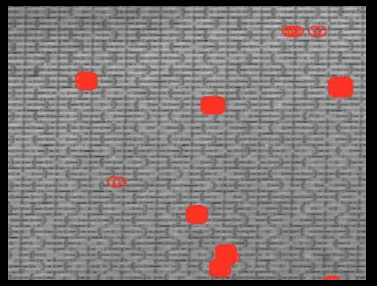


SEM Image



Small crops

Defects detected (>MSE threshold)



Application of the Attention-Guided Neural Network for the Defects Detection, Alexey Solovey, et al., ISTFA 2024
AI Applications for Failure Analysis - AI-Driven Anomaly Detection in Electron Microscopy Images of Periodic Circuit Structures, Alexey Solovey, et al., ISTFA 2025

- Explore defect detection for non-periodical structures: Layout and circuit aware

Deep-Learning Based Anomaly Detection App, Written by GEN-AI

1. Input (Natural Language)

What You Asked For:

- Automate defect detection in SEM/microscopy images with repeating patterns
- Combine pattern detection with AI-based quality analysis
- Make it production-ready and easy to deploy

2. GenAI Code Generation

What Cursor Delivered:

Complete Production System

- 3 usage methods: CLI tool, Python API, and Jupyter notebook
- Automated 7-step pipeline: detection → cropping → training → analysis → reporting
- 5 quality metrics for different pattern types (spots, lines, dots)

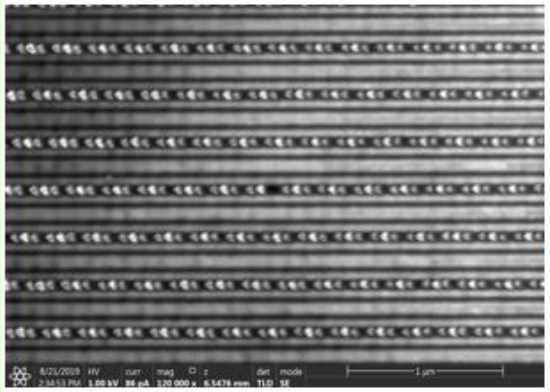
Deployment-Ready Package

- Command-line tool for batch automation (`run_defect_detection.py`)
- Comprehensive deployment guide with examples
- Requirements file with all dependencies
- Automated output generation (annotated images + CSV reports)

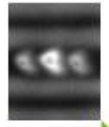
Production Features

- Configurable sensitivity and metrics
- Batch processing examples for multiple images
- Integration code for existing pipelines
- Troubleshooting guide and parameter tuning tips

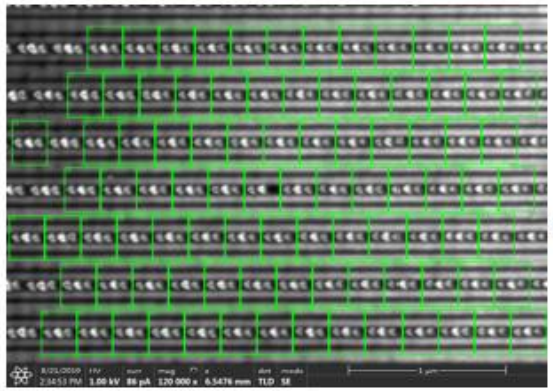
3. The Execution: SEM Abnormal VC Detection App



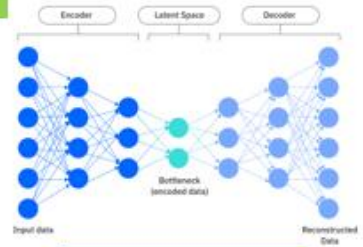
Raw SEM image



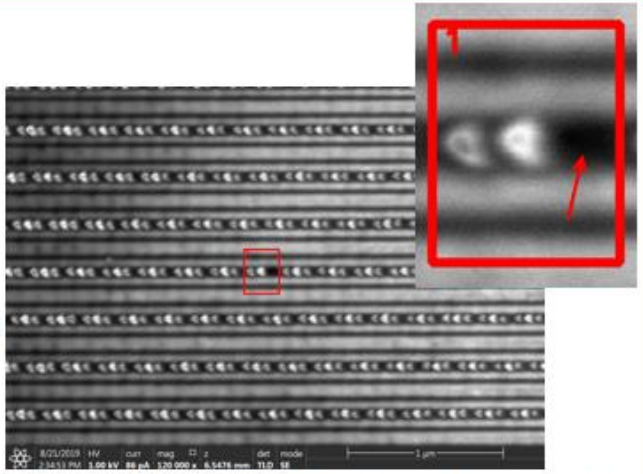
user select an example structure in SEM



Script auto- detected and labeled all similar structures



Script trained anomaly detection AI using these similar structures



AI detected and labeled an abnormal voltage contrast in SEM image

Call to Action

Failure Analysis Community

Expand deep collaboration with the fab, design, product, and system key stakeholders

Continue to drive innovation in the face of disruptive technology advancements

Meet/Beat today's FI performance

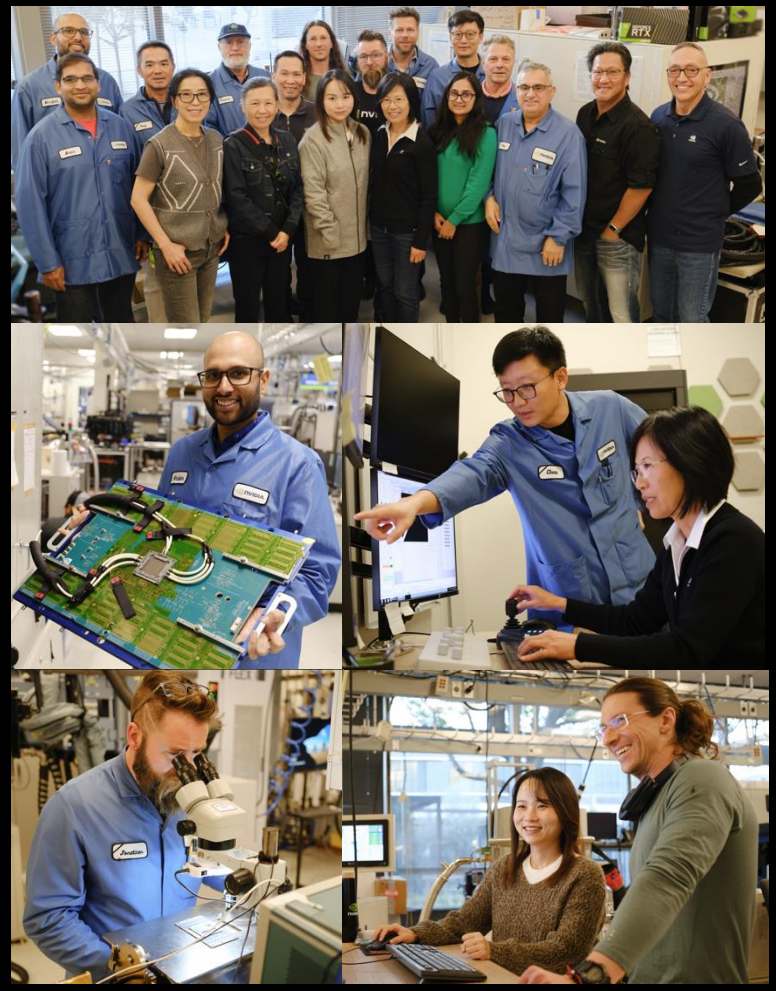
Drive defectivity reduction to the delight of customers

Use AI in your daily work

Create an image that captures the "essence of AI".

Acknowledgements

NVIDIA SiFA Lab



Key Collaborators

- Arpan Bhattacherjee
- Jane Li
- Joy Liao
- Rudolf Schlangen
- Christian Schmidt
- Alexey Solovey
- Chuan Zhang


NVIDIA