Microsoft Azure

From Lab to Data Center: Accelerating Silicon Readiness Through Reliability Innovation

Presenter: Sujata Paul

Agenda

01 **Introduction & Context:**
Cloud silicon and the need for reliability

02 **Quality vs. Reliability:**
Defining and distinguishing these concepts

03 **Intelligence-Based Qualification (IBQ):**
Targeted reliability testing approach

04 **Future Directions:**
Emerging challenges in reliability engineering

05 **Conclusion:**
Learnings and takeaways

Introduction – Reliability Engineering in Microsoft's Silicon Efforts

- Microsoft built silicon for Azure—Cobalt CPU and Maia AI accelerator
- Powers mission-critical services: AI inference, Azure VMs, and Teams
- Reliability is strategic: even minor hardware faults can disrupt cloud services at scale
- A single accelerator fault can trigger failovers, raise power draw, and drive costly swaps
- 24×7 SLOs (Service Level Objectives)demand near-zero failure rates across global fleets
- Reliability is engineered from day one—design through operations—to protect service integrity and availability

The Azure Maia 100 and Cobalt 100 chips mark Microsoft's debut into custom silicon tailored for its cloud infrastructure. / Microsoft

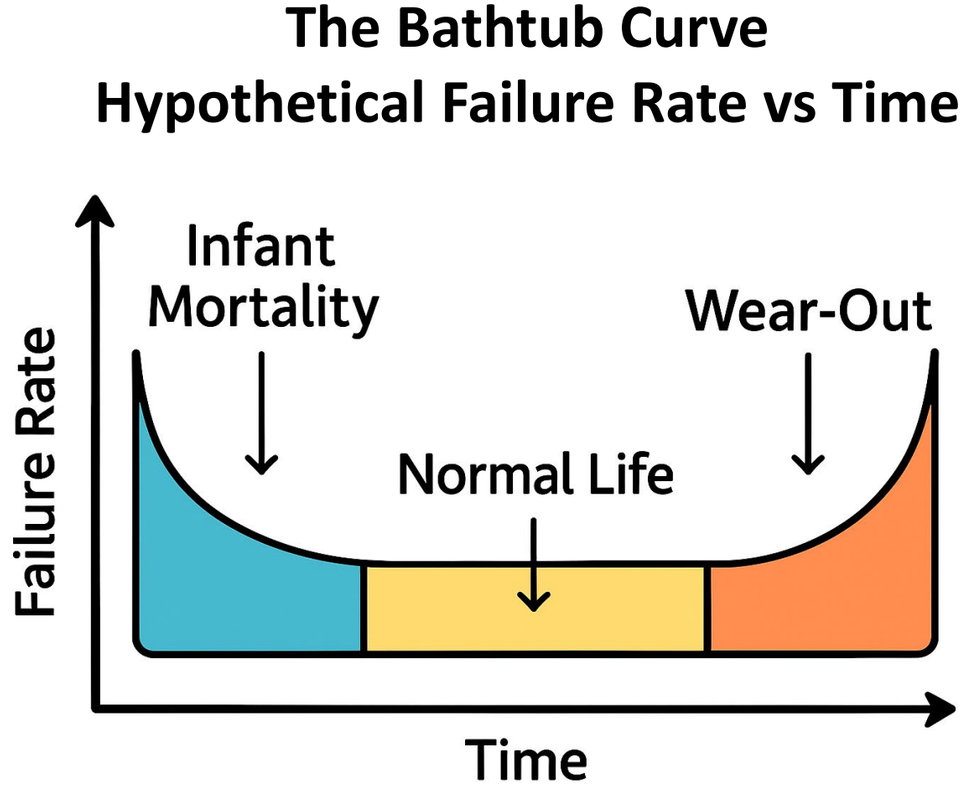
Quality vs. Reliability

- Quality (Time-Zero): Conformance to specs at deployment. Ensures a defect-free product at shipment
  - Focus: Catch manufacturing defects early (no “infant mortalities” in field)

- Reliability (Lifetime): Ability to perform within specs over the intended lifespan (5–10 years in data center)
  - Focus: Sustain performance over time under normal use, with minimal failures in later life

- Both are vital in the cloud:
  - High initial quality → fewer early failures in field
  - High reliability → low failure rate as years pass

- Summary: Quality is about delivering a flawless chip on day one; Reliability is about that chip staying flawless for years . We need to engineer both into our products



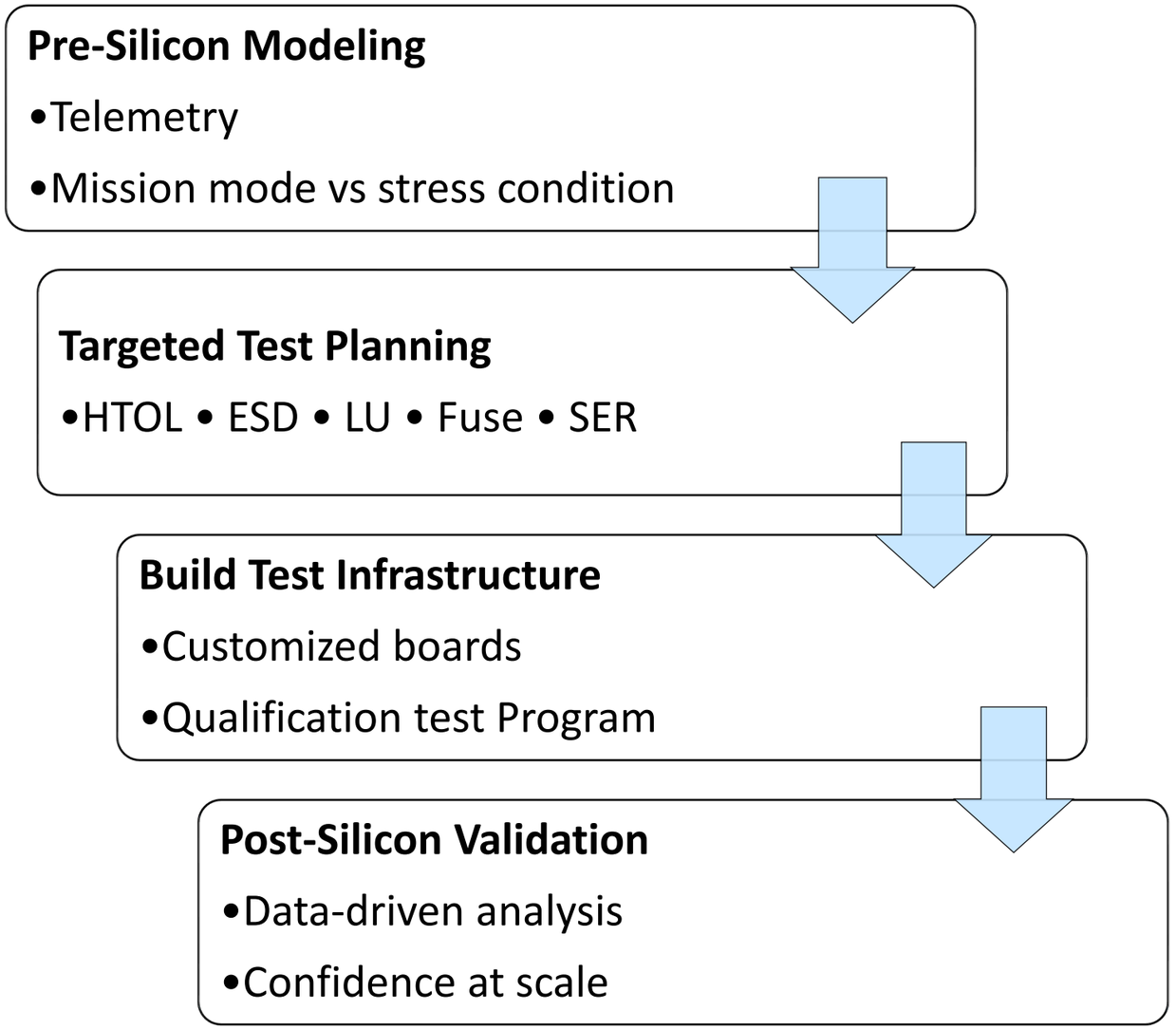
Intelligence-Based Qualification (IBQ)

Intelligence-Based Qualification (IBQ)

- Focus testing on highest-risk failure areas, not just blanket stress
- Use design analysis and field data to target weak points
- Avoid over-stressing chips; test smarter, not harder
- Outcome: Reliable chips, efficient qualification, high confidence

Key Reliability Tests:

- High Temperature Operating Life (HTOL)
- Electrostatic Discharge (ESD)
- Latch-Up (LU)
- Fuse Test
- Soft Error Rate (SER)



HTOL (High Temperature Operating Life)

- **Purpose:** Simulate years of chip aging in weeks
- **Test Method:** Run chips hot and biased; monitor for changes
- **Key Checks:** Look for defects, performance drift and reliability issues
- **Hardware Setup:** Chips mounted on custom burn-in boards inside ovens, kept active with stress patterns
- **Monitoring:** Live tracking of voltage, current, temperature; ATE(Automated Test Equipment) tests measure performance drift
- **Result:** Cobalt & Maia chips show minimal aging, zero failures
- **Impact:** Confirms robust design for heavy cloud workloads

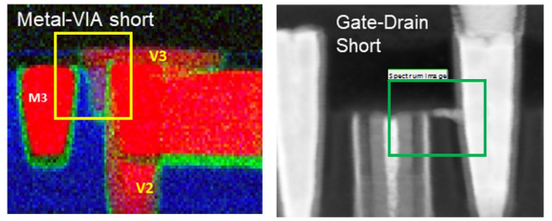
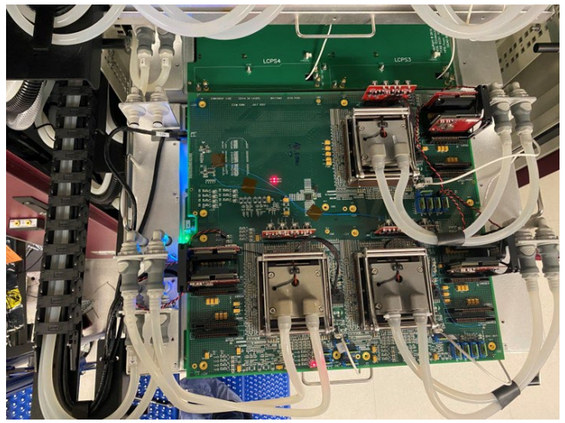
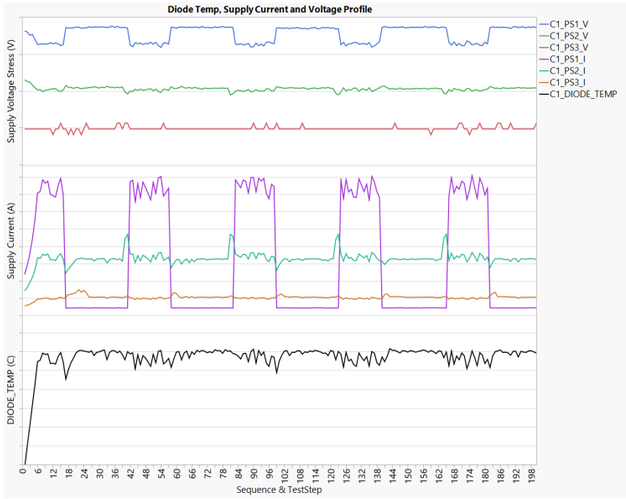


Image Source: from web

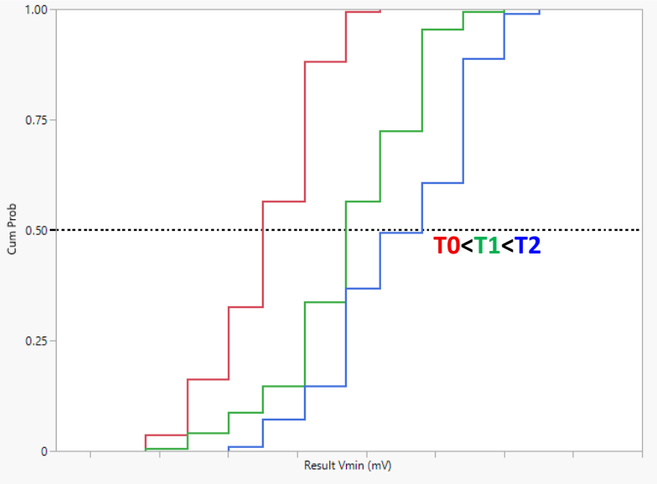
Example of defect



Hardware Setup
Burn-in boards in ovens



Monitoring — Live V/I/T

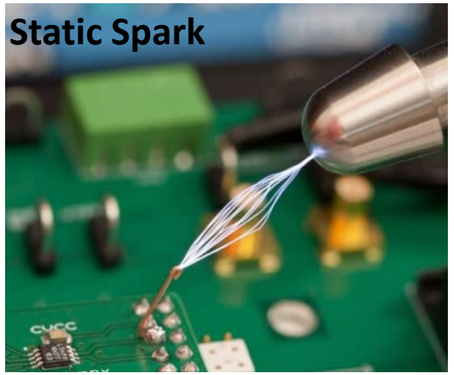


Results —Minimal aging

Microsoft Azure

ESD (Electrostatic Discharge)

- **ESD Risk:** kV static can damage nanometer devices
- **Protection/Design Tradeoff:** On-chip diodes & clamps at I/O pins shunt surge currents. Must balance robustness vs. added capacitance to preserve high-speed I/O performance
- **Test Standards:**
  - HBM (Human Body Model): Simulates human touch, ±1 kV Zap. Slow two Pin contact discharge to device
  - CDM (Charged Device Model): Simulates chip discharge, ±500 V Zap. Fast, high-current pulse
- **Test Process**:
  - Customized board design; Automated zaps (150V to 1kV) per pin
  - No leakage, parametric shifts, or functional failures after zaps
- **Results:** Cobalt & Maia passed; many pins exceeded spec
- **Impact:** Robust through fab, assembly, and field handling; no Signal Integrity penalty



Example of ESD Damage

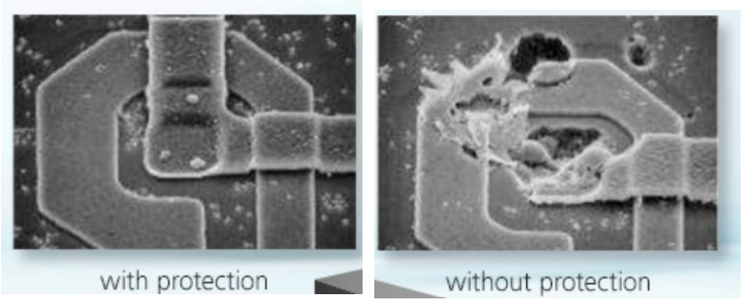
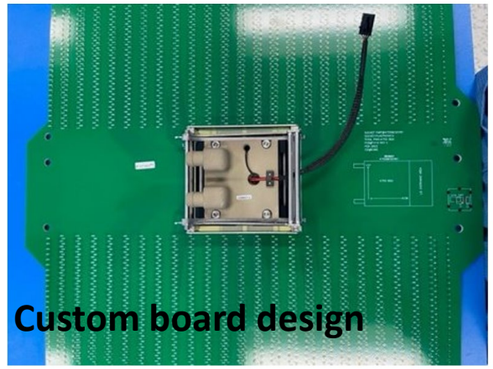


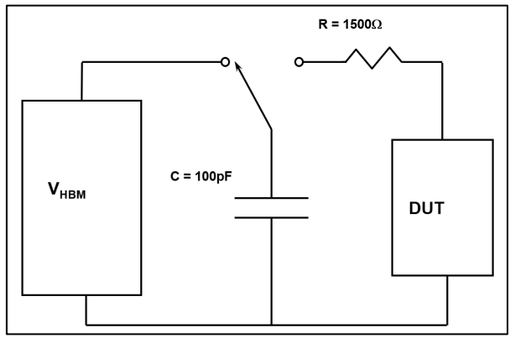
Image Source: www.semtech.com

Image Source: Analog Devices

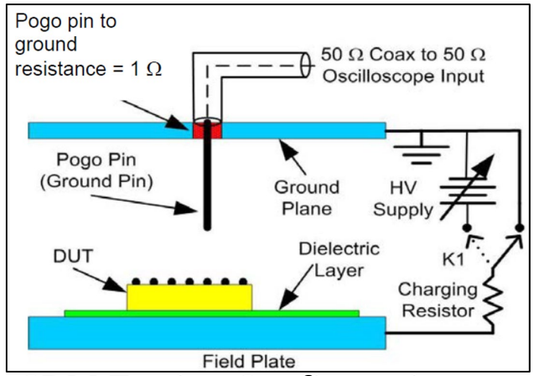


Automated Testers

Image Source: www.thermofisher.com



HBM Circuit Schematic



CDM Testing Configuration

LU (Latch Up)

- **What is Latch-Up?**
  Parasitic short between power & ground in CMOS chips can cause high current, risking damage

- **Prevention:**
  Robust design: increased spacing, guard rings, lower voltages.
  Industry-standard testing (JEDEC JESD78) ensures resilience

- **Test Process:**
  Over-Voltage: 1.5× max V on power pins at high temp
  Over-Current: ±100 mA into I/O pins under bias/high temp

- **Results:**
  Cobalt & Maia passed all conditions. Tolerated ≥1.5× V, ±100 mA. No leakage, no parametric shifts observed

- **Impact:**
  Reliable integration, immune to voltage spikes and miss plug events.
  Reduces field failure risk, simplifies system deployment

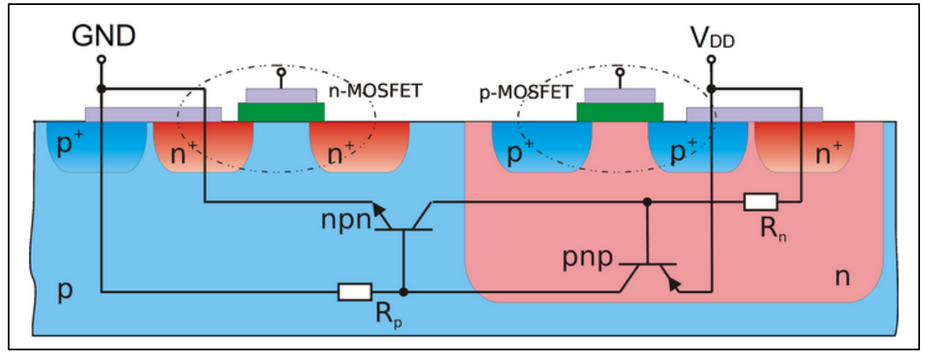


Image Source: from Web

Parasitic (PNP/NPN) Structures in CMOS

Image Source: www.advantest.com

LU set up in ATE 93K environment

Fuse Test

- **Purpose:** Modern chips have many one-time programmable (OTP) eFuses for configuration, calibration, security keys, feature enabling/disabling . Once “blown” (permanently open), they store a bit of data. Reliability of fuses is critical because they often control security or functionality that must persist for the device’s life

- **Test Process:**
  Read Disturb Test: Tens of thousands of reads under stress
  — no unintended fuse blows

  Fuse Regrowth Test: High-temp bake (150 °C, >1000 hrs)
  — no blown fuses reconnected

- **Result:** Cobalt & Maia passed all tests — secure, permanent fuse behavior

- **Impact:** Long-term reliability for security, ID, and config features

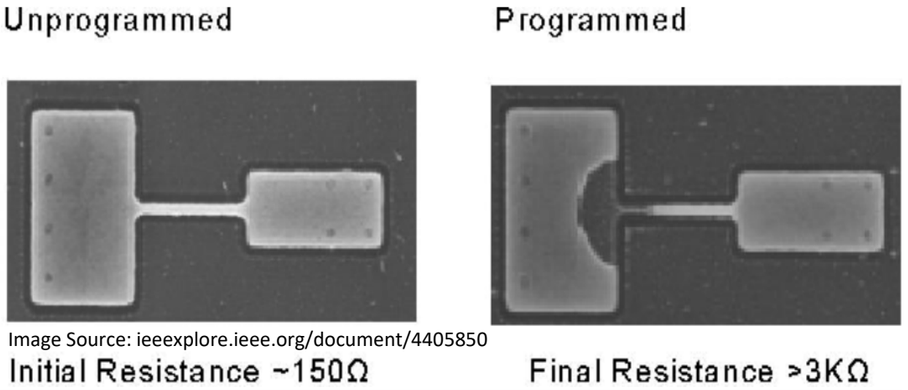


Image Source: ieeexplore.ieee.org/document/4405850

Example picture of Metal fuse

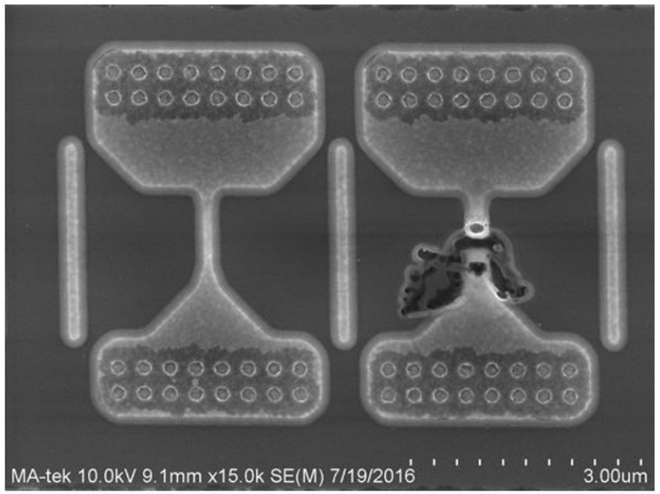


Image Source: semiengineering.com

Fuse regrowth

Soft Error Rate (SER) – Radiation Resilience

- **Soft Errors:**
  - Transient bit flips or logic glitches caused by external radiation (neutrons, alpha particles)
  - Do not damage hardware permanently but can corrupt data or computations

- **MSFT Testing Method**:
  - Neutron beam exposure simulates years of natural radiation in hours
  - Measured in FIT (Failures in Time)

- **Focus Areas/Mitigations**:
  - Memory-heavy blocks (e.g., Maia SRAMs)
  - ECC (Error Correction Code) and parity used for detection/correction in real time

- **Results:**

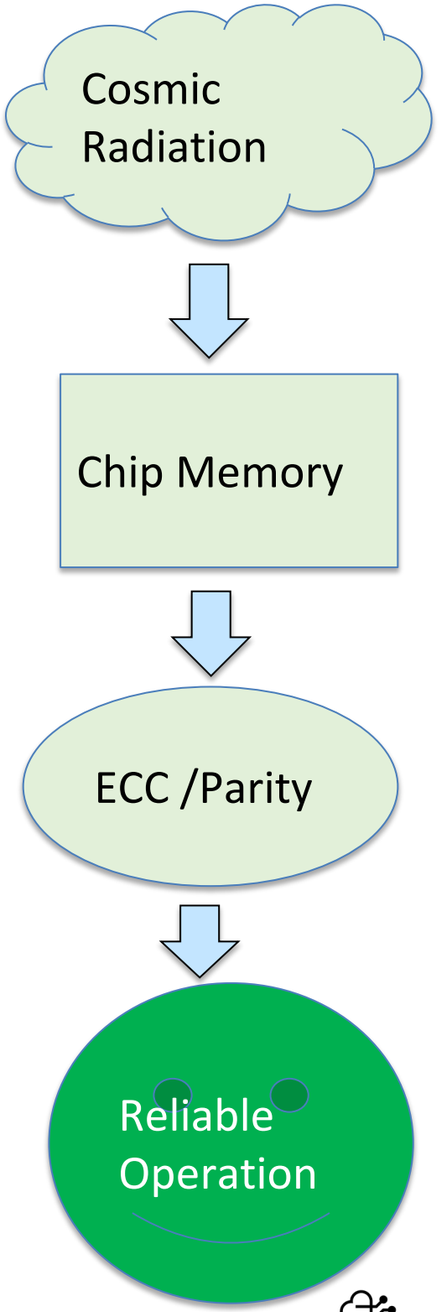
✅ Memory bit flips corrected by ECC or flagged by parity: no silent corruptions

📉 Uncorrected error rate- few hundred FIT/device: ~1 error per server in decades

🔍 No SDCs (Silent Data Corruption) observed: every radiation-induced event was detected or fixed

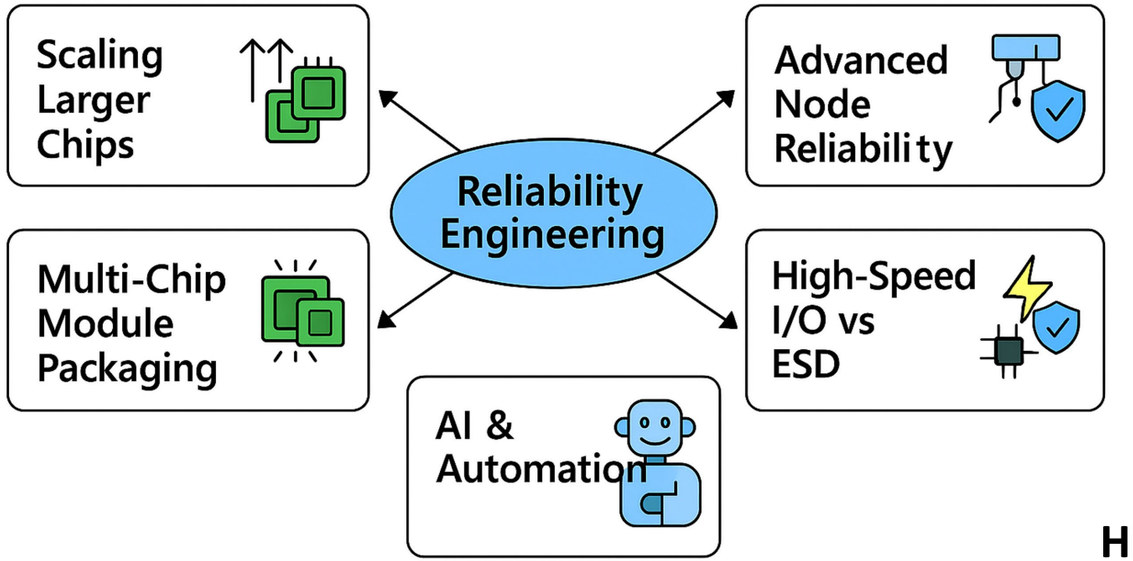
🔒 Logic SER tested: very low vulnerability due to protective design

- **Implication:** Chips show strong immunity to radiation-induced errors. ECC and design hardening ensures reliable operation in large-scale deployments



Future Directions in Reliability Engineering

Scaling Larger Chips
More transistors, higher power—need smarter designs and on-die monitors to catch localized issues.



Advanced Node SDC (Silent Data Corruption)
Shrinking transistors can cause rare logic errors missed by standard tests.

Multi-Chip Packaging
Stacked dies introduce package-level stress (cracking, delamination). MCM/Chiplets require new stress tests and robust connections.

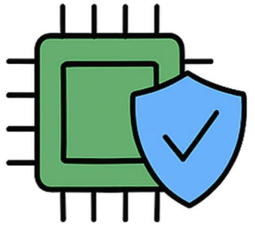
AI & Automation
Use machine learning on test program, test data and field telemetry to predict failures.

High-Speed I/O vs ESD
Traditional ESD protection compromises performance, explore “invisible” ESD devices that activate only during discharge.

Conclusion - Meeting Hyperscale Reliability Demands

- Cobalt and Maia chips meet or exceed the stringent reliability requirements for Azure’s hyperscale cloud .
- Intelligence-Based Qualification guided testing to critical areas, accelerating time-to-production while ensuring no major failure mode was overlooked.
- “From lab to data center” signifies a chip isn’t truly a success until it works flawlessly at scale in real-world use .
- Result: High confidence to deploy these chips at scale in mission-critical services.
- Reliability is a competitive advantage: By rigorously engineering reliability, Microsoft ensures high availability and reduced downtime in Azure services – a key expectation in the cloud market.

- Chips meet high standards for reliability and security



- Extensive testing ensures robust performance in the cloud

- Committed to continuous innovation in reliability

Acknowledgement

Silicon Reliability Team:
Ray Talacka, Georgia Modoran, Anis Rahman, Shridhar Dixit, Serena Wang, Joey Lee, Saurabh Agrawal

Thank you