Intel Paper

Quality and Reliability Implications for the Connected World

Volume 1

Intel® Corporate Quality Network

Table of Contents

Abstract

Quality and reliability have been key considerations for every advancement in compute technology. Innovations in materials, integration, and fabrication techniques have maintained or improved component failure rates generationally, dramatically improving reliability per transistor. Increases in ecosystem scale and the growing use of such technologies for critical infrastructure are driving even higher expectations for quality, reliability, data integrity, and security. Meeting these evolving needs requires capabilities well beyond component technology alone, including advances in manufacturing test, architectural resilience, as well as a comprehensive approach to in-field test and maintenance. This article discusses this evolving landscape as well as Intel's integrated response. Intel is committed to engineering solutions across its IDM portfolio to address these challenges.

Introduction

Guided by Moore's Law, the building blocks of semiconductor technology have generationally decreased in size, thereby dependably increasing transistor density and improving CPU performance. The resulting scaling and technological advances have driven localized performance improvements across an increasingly distributed ecosystem. Advanced packaging technologies and modular design have pushed these boundaries even further, enabling feature integration beyond silicon density and wafer patterning limitations.

In parallel, the larger distributed scale of compute continues to grow at a dramatic rate. The advent of user-aware and smart-AI (artificial intelligence) applications, as well as a steady increase in the world's connected devices, is driving the immense scale of cloud computing and the resulting exponential growth of data. Ecosystem architecture is evolving to respond to the growing data required to satisfy the needs of local applications, needs at the edge, and the needs of machine-to-machine communications, helping to enable a seamless response for all user experiences.

As described by Intel CEO Pat Gelsinger, this larger distributed scale of compute is driven by the technology "superpowers" of compute, connectivity, infrastructure, AI, and sensing. These superpowers combine, amplify, and reinforce one another, and as they become more ubiquitous, they in turn unlock even more powerful new possibilities. These vectors are in turn enabled by an increasing amount of underlying infrastructure and compute performance in the form of networking devices, accelerators, cloud storage/compute, and encryption, as well as the software and application innovations that support them.

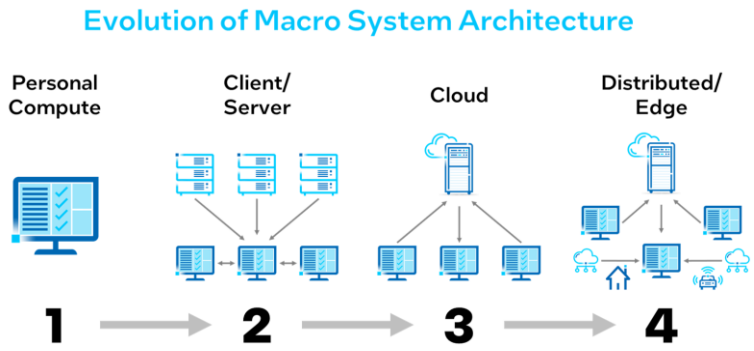

Figure 1: The Evolution of Macro System Architecture

The immense aggregate scale of this ecosystem creates high expectations for all aspects of component quality and reliability, especially data integrity and availability. Both quality escape rates in manufacturing and failure rates in operation—once satisfactory for historical levels of scale and applications—are today no longer acceptable in a world where data centers have millions of CPU cores potentially managing safety-critical infrastructure. New approaches and solutions, from silicon and package technologies to platform and system resiliency features, are required to deliver to on these growing expectations.

Given this range of challenges, there is no one solution that will deliver on all evolving quality expectations. Instead, we need to focus on a holistic set of innovations. Intel's IDM capability uniquely positions us to deliver quality solutions end-to-end, including technology advances, manufacturing leadership, roadmap for architecture resiliency improvements, as well as a comprehensive approach to customer support that includes in-field test and repair capabilities. In Volume 1 of this paper, we speak to some of the traditional defect and reliability challenges. In future volumes we hope to expand the scope of the quality discussion to include other attributes, including validation quality associated with functional convergence, along with methods deployed to deliver world-class software quality.

Q&R Challenges Based on UX and Scale

The above trends bring new meaning to the word "scale". At such levels of integration even the smallest quality or reliability issue may become visible to the end-user. A single bad encryption key can now impact a wealth of associated data and any single compute error, system crash, or hard failure can result in a loss of compute resources. As an IDM leader, quality in this larger sense becomes the quality of the overall resulting experience, from validation to system manufacturing, to final deployment and the end-user experience. Table 1 below defines some of the most important quality attributes for users.

Customer Quality Attribute	***Definition***	***Customer/User Impact***
System Manufacturing Quality	Failures during system manufacturing	Customer manufacturing cost and throughput
Validation Quality or New Product Introduction (NPI)	Commonly referred to as "performance to specification" (i.e., fully featured and meeting all specified performance and power specifications)	Systemic issues such as bugs or marginalities that escape validation can delay customer launch and slow product ramp
User Experience/ Deployed Application	Failures observed by the end-user or deployed application, many times referred to as "field reliability"	Failures during use can manifest in many ways, the most common of which is a loss of compute availability in the form of a system reboot or hard failure replacement

Table 1: Customer Quality Definition and Failure Impact

Failures during manufacturing can be related to a range of different causes (several of which are described in Table 2 below). These have a direct impact on customer manufacturing cost, both in the form of rework cost and lost capacity due to volume throughput implications. Increased package complexity and the integration of multiple devices can compound this risk, placing even more stringent requirements on incoming supplier quality. The cost implications of quality yield loss become more severe with a high-volume ramp. Before the product can launch and ramp, however, it must be fully validated.

Customer development during the NPI (new product introduction) phase is accelerated by a fully featured device operating within specified performance and power implications, without functional bugs or marginalities. Hardware or software bugs can delay customer launch and slow the pace of volume ramp. Once the platform launches, quality experience is measured by the end-user application. Quality requirements can vary across segments, but availability—the continuous operation of the compute device without any errors that cause a crash or require system reboot—is nevertheless a consistent metric across most applications.

As stated above, in some cases failures can lead to a permanent loss of data or, in even rarer cases, the propagation of silent data errors (SDE) without system or user notification. Even though such cases are very rare, the scale of compute is increasing the visibility of even very low failure rates. For example, in a cloud installation a component failure rate of 1 FIT (1 fail in 1 billion device hours)—extremely low by historical standards—may in a state-of-the-art data center now result in a failure within 30 days.

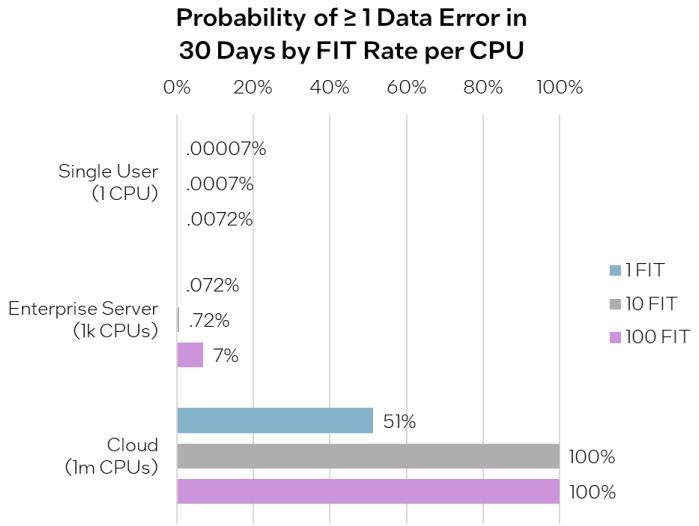

Figure 2: Probability of Error within 30 Days by FIT Rate

As the scale of compute increases, the complexity of system integration is also driving more intrinsic quality risks. The shrinking of physical feature sizes introduces a variety of challenges. The materials used may experience higher intrinsic stresses in operation, and smaller dimensions are more susceptible to defects, variation in manufacturing, and atomic-level effects. CPUs and memory chips now contain billions of transistors, compounded by enormous compute installations. Quality leadership at this new scale of compute must address all potential quality risks across the full range of failure modes represented by design bugs, silicon, and/or package defects, as well as intrinsic reliability performance. To better understand these risks, we traditionally model the relationships between failure mode to physical fault to the final customer manifestation. The table below defines some of the most common failure modes.

Failure Modes	*Definition*
Functional Bugs or Design Marginality	Software or hardware bugs that escape our validation process: The complexity of the integrated system compounds this risk further due to the interoperability and compatibility challenges associated with the wide range of deployed peripheral devices, usage environments, and software applications
Manufacturing Defect Escape	Silicon or package defects that escape the component manufacturing test process and are propagated to the customer's manufacturing or final field deployment
Latent Reliability Defect	Latent silicon or package defects that are accelerated during use: Traditionally defined as early life defects (defects per million or DPM); the level of field DPM can be managed with accelerated stress tests during manufacturing
Intrinsic Reliability (Wearout)	Failures induced by fundamental degradation or the breakdown of materials or interfaces as the component ages; can be silicon or package related
Transient or EOS Modes	Transient modes include soft-error failures due to cosmic or alpha rays and electrical overstress modes related electro-static-discharge (ESD) or latchup

Table 2: Failure Modes

Whether a failure occurs in a large data center installation or a home computer, the implications to the resulting user experience can be similar. The most familiar is the dreaded reboot, a fault that leads to a hang or crash and requires the system to be rebooted, leading to a loss of compute availability. (The larger context of faults and system failure manifestation is described in Figure 3 below.) A critical aspect of the overall quality experience is defined by how each failure manifests in the usage application: Are they persistent, intermittent, or non-recurring in the system operation? Do they produce a detected error, resulting in the termination of a program or system (DUE)? Or are they "silent", resulting in the propagation of incorrect data not observable by the system or user (SDE) or an unattributed execution error (UOE)?

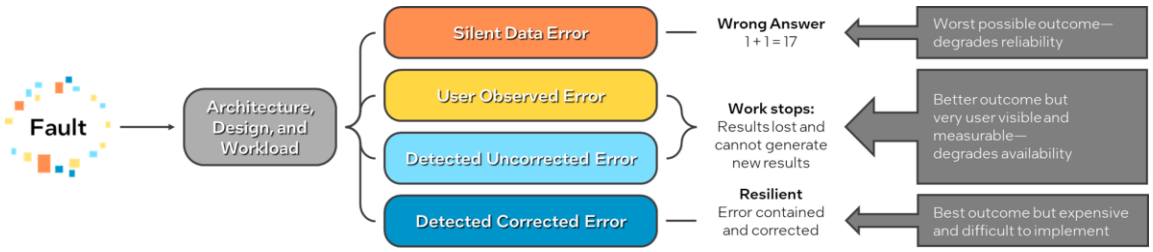

Figure 3: Faults and Data Error Taxonomy

The manifestation of a fault at the system level is highly dependent on where it occurs within the component architecture, as well as the workload being executed. Since some faults only impact operation under rarely occurring conditions, this confounds the traditional distinction between "quality escapes"—which are assumed to be experienced by the user at initial operation—and "reliability modes". At the system level, different physical failure modes can produce identical manifestations. The overall system mitigation for these faults is traditionally defined by the RAS architecture (reliability, availability, and serviceability).

Quality Leadership Response

As we have seen, the breadth of potential quality concerns, from silicon to package to platform complexity, are immense. Given this range of challenges, there is no single solution that will deliver on all growing quality expectations. What is needed is a focus on a holistic set of innovations. Intel's unique IDM capability well positions it to deliver such end-to-end quality solutions, providing improved process and defect controls, better coverage and manufacturing test screens, innovative architectural resiliency roadmap, and new capabilities, including in-field test and repair.

Figure 4: Structure and Flow for the Quality Leadership Response

Technology: Silicon and Package

Required component lifetimes now necessitate a comprehensive understanding and modeling of failure and degradation modes, combined with the right design tools to ensure electrical circuit implementations are consistent with technology and product applications. For MOS transistors, the incorporation of optimized Hik dielectrics is enabling dielectric scaling while providing extended reliability without a significant increase in leakage (SILC), breakdown (TDDB), or excessive shifts in transistor thresholds (BTI). Hot carrier injection (HCI) is being managed through the engineering of channel electrostatics and design choices. Interconnect resistance degradation (electromigration) has also been a fundamentally important mode from the beginning of IC technology.

Generational increases in current densities have driven the evolution of metallurgy from aluminum to copper to refractory materials. Process reliability engineering must continuously balance achieving required component lifetimes with maximizing performance (frequency, throughput, etc.). A key consideration is the conditions under which a component will operate. Components utilized in harsher operating environments, or in systems with long lifetime requirements, necessitate adjustments to design and performance.

Despite the engineering challenges of intrinsic reliability modes, latent defects have remained the dominant source of failures in semiconductor components. These are primarily shorts which develop between die features, due to particles or variations in the patterning, resulting in insufficient spacing. Defects may similarly result in marginal connections between features which then degrade in use. As with intrinsic fail modes, the manifestation of a failure is largely dependent upon where it occurs within the design.

The control of latent defects has continuously improved, both through process engineering to reduce the density of defects and through the refinement of accelerated and predictive screens to weed out these defects in the manufacturing flow. Most potential fails are removed with these techniques, though keeping up with transistor density, component silicon area growth, and sheer system scale poses a difficult challenge for conventional defect-reliability mitigations.

Intel has broken the constraints of 2D package-level integration with innovations in modular design. These innovations in disaggregated package technologies have further increased the scale of features and compute engines (cores) that can be integrated into a single package. This has increased performance and improved operating characteristics while removing the reticle size limitation of the total silicon area.

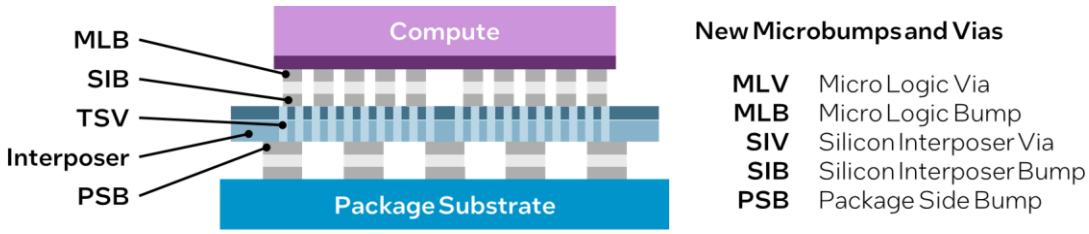

Figure 5: Intel Foveros Technology

Intel's Foveros technology incorporates an active silicon interposer and, alternatively, can utilize a passive interposer for lower cost and quicker time-to-market. The number of top compute dies are still limited by the base die or interposer. As shown in the figure below, EMIB technology removes base die size restrictions and incorporates another level of compute scale.

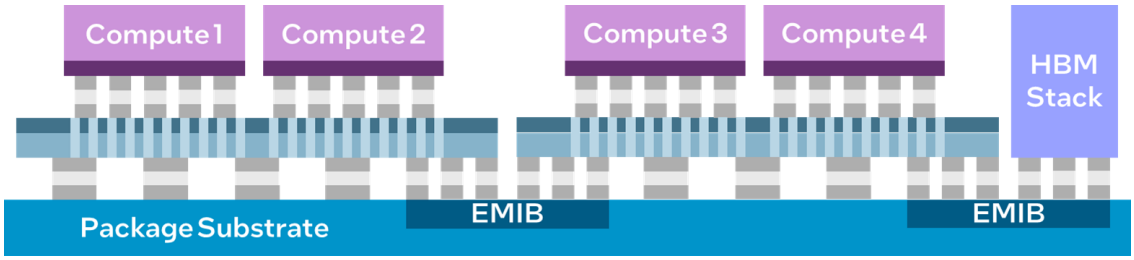

Figure 6: EMIB (Embedded Multi-Die Interconnect Bridge) Technology

These innovations have extended Moore's Law while increasing package-level complexity. The complex manufacturing process associated with integrating multiple chiplets, as well as the increasing number of material types and interfaces, has led to additional quality and reliability risks that must be addressed during development.

Manufacturing

As the area of silicon in a single component increases, the requirements for defect screening in manufacturing must tighten to achieve the same level of quality as before. For example, when a typical semiconductor SoC consisted of less than 1 cm^2 of silicon area, it may have been sufficient to screen 90% of manufacturing defects to deliver the quality that customers expected. Today, some highly integrated devices contain over 10 cm^2 of silicon area spread across multiple tiles or chiplets. As a result, screening 99% of defects might be inadequate to achieve historical levels of quality. For these reasons, multiple layers of test are now required to meet the quality expectations of customers.

Test capabilities embedded within the die, known as "design for test" (DFT) features, are the foundation of manufacturing test. An example is scan DFT, which utilizes dedicated test circuitry to shift specific patterns of 1s and 0s into logic gates using automated test equipment (ATE). The ATE tools propagate these patterns into the device and compare the outputs against the expected results. To achieve a high rate of defect detection, the scan circuitry must be inserted throughout the design. Advanced techniques, such as high-speed scan test, are required to detect more subtle defects.

The effectiveness of manufacturing test is described by "coverage", which measures the ability to screen defects in the silicon process before devices are shipped to customers. Traditionally, coverage has had two basic metrics: the percentage of easy-to-detect defects that can be screened ("stuck-at" coverage) and the fraction of defects only detectable with full speed testing that can be screened ("at-speed" coverage). Today, Intel is approaching levels in the range of 98% stuck-at coverage and 90% at-speed coverage, with a path to industry standard coverage of 99.5% stuck-at and 95% at-speed.

Given the amount of silicon area scaling and the complexity of compute interfaces, we have reached the point where high levels of ATE stuck-at and at-speed coverage may be insufficient to screen all defects. In these cases, system level test is utilized to cover the more complex interfaces and the most difficult-to-detect defects. As described previously, specific defects can be observed as a visible failure or, in rare cases, can manifest as a silent data error (SDE).

Because of the expectation for very low levels of SDE, there is particular attention on test algorithms which effectively screen these defects. Some SDE only occur under a very specific combination of conditions. These criteria can include the precise data and machine instruction sequence, specific operating voltage, frequency, and temperature conditions.

Tests that effectively screen for SDE utilize algorithms that explicitly check for the correct results across all cores in the processor. They run multiple loops of code that span the vast data, address, and instruction space of a modern processor. Examples of functional test types that are effective at screening SDE include:

1. Golden-Value Tests: Results are compared to pre- computed correct answers.
2. Application Library Snippet Tests: Examples include tests based on cryptography and Eigen linear algebra libraries. These tests are run in parallel across different threads, with the results compared at the end of each iteration.
3. Instruction Set Architecture (ISA) Targeted Stress Tests: These tests are run in parallel across different threads, with the results compared at the end of each iteration.
4. Software Stress Tests: Tests based on real-world software, including inverse transformations such as compression/decompression and encryption/decryption.
5. Randomized Stress Tests: These tests utilize machine generated assembly test sequences and randomized assembly traces. These tests are run in parallel across different threads, with the results compared at the end of each iteration.

While the combination of these manufacturing techniques delivers extremely high levels of quality, they cannot achieve a defect rate of zero. As discussed, latent reliability defect fails will only manifest after a period of operation. Additional approaches to defect management are therefore being developed, including enhanced architectural resiliency and in-field system mitigations.

For a more complete discussion of tests used for the detection of defects that manifest as SDE and the need to manage these defects.[1]

Architecture

For any application, the connected scope of data transfer extends well beyond the localized compute instance. As the scale of the compute ecosystem continues to grow, this means the probability of a failure will no longer be determined at the localized compute instance. Even with world-class coverage and defect screening techniques, the escape rate will unavoidably be above zero. The contribution of latent reliability defects, radiation, and potential wearout-induced mechanisms will similarly drive a non-zero contribution to field error events.

Addressing these new challenges requires new architectural capabilities to ensure RAS (reliability of execution, availability, and serviceability). Many of these concepts and techniques are not new, having been developed over decades to mitigate radiation-induced transient effects (SER). For example, parity and in-line error correction (ECC) were developed to detect, and where possible correct, transient errors transparently to the user. These detection and correction features, as well as more advanced techniques for detecting errors in computations, are increasingly being leveraged to mitigate persistent faults, with a particular focus on preventing SDE.

From user touchpoints to the connectivity of the edge and the cloud, data integrity and availability require a holistic approach to RAS. If future applications demand a trusted compute level with 100% data integrity, then a larger breadth of mitigations must be deployed. Features such as hardware or software lockstep, or logic residue, deliver real-time mitigation for errors, but come with significant performance loss due to the replication and self-checking nature of the compute mitigation.

When difficult-to-detect errors require specific workloads and rare code strings to manifest, periodic maintenance techniques could enable the detection of an issue prior to a field error event. The breadth of applications as well as the varying importance of data integrity across these workloads calls for an "on-demand" approach. Maintenance-based solutions can play a valuable role, especially where MTBF is longer in duration. Methods such as periodic in-field testing, in-field repair, and in-field prognosis will play a vital role in achieving the data integrity requirements of the compute ecosystem.

Another innovation, the Intel® Data Center Diagnostic Tool,[2,3] is a suite of tests targeted at SoC functionality, including each of the logic cores. This innovative approach to in-field diagnostics can be expanded to include generational improvements in coverage and the full breadth of DFT. In-field scan testing has been enabled on the latest Intel® server products to allow periodic background self-testing. This will identify "latent" faults without disrupting other activities, with minimal impact to overall performance.

Future opportunities include allowing cores to be periodically paired to operate concurrently (lockstep) to uncover subtle faults. This capability can also be selectively exploited for mission-critical functions, such as security encryption to guarantee error-free execution. Component and system hardware architectures can be optimized along with software to provide failure recovery by minimizing the impact of an isolated failure on other circuitry (i.e., blast radius). This is done by incorporating redundant elements and enabling dynamic reconfiguration to either completely repair the component or disable a failed element, allowing for continued functionality at reduced capacity.

The breadth of RAS architecture options is wide, but to enable the level of data integrity needed for some applications, a holistic approach across the compute hardware and overall system software is required.

Conclusion

Advances in silicon and package technology have enabled amazing growth in the scale of compute and its increasingly pervasive utilization for critical applications. With these successes comes increased expectations for quality and reliability, and safety and security, far above what has been required historically.

At today's scale of compute, a single failure that might have historically gone unnoticed is now apparent and can impact a wealth of data and critical functionality. Meeting these expectations requires capabilities beyond the component technology alone, including advances in manufacturing test, resilient architecture, and a comprehensive approach to in-field test and maintenance.

Intel® is delivering on evolving quality expectations by incorporating capabilities well beyond technology alone. These include the development of advances in manufacturing test, resilient architecture, and a comprehensive approach to in-field test and maintenance.

As an IDM, Intel® is engaged in the entire development cycle and is committed to customers to improve quality with a full range of innovations. Intel® is, and will continue to be, an industry leader in delivering high-quality products and solutions well into the future.

References

1. Lerner, D., Inkley, B., Sahasrabudhe, S.H., Hansen, E., Rojas Munoz, L.D., & van de Ven, A. (2022). Optimization of tests for managing silicon defects in data centers. *IEEE International Test Conference Proceedings*. Not yet available online.

2. Intel® Data Center Diagnostic Tool for Intel® Xeon® Processors. (2021). *Intel.com*. Retrieved on October 25, 2022 from: https://www.intel.com/content/www/us/en/support/articles/000058107/processors/intel-xeon-processors.html.

3. van de Ven, A. (2021). High reliability and availability require the right tools to manage system health. *Intel.com*. Retrieved on October 25, 2022 from: https://www.intel.com/content/www/us/en/developer/articles/tool/intel-opens-beta-for-datacenter-maintenance-tool.html.