Quality & Reliability: Fundamentals

Technical Paper - Volume 5

Introduction

Intel manufactures over a quintillion transistors every year. That's a billion billion—a one with 18 zeroes. How does the company ensure this many transistors not only work correctly but keep working correctly for years to come? These are the key challenges of quality and reliability.

Understanding fundamental quality and reliability concepts is important not only for Intel, but also for its customers, partners, and suppliers. For Intel, quality means that its products work as specified at the time they are manufactured and integrated into systems. Reliability means that products work correctly over a defined period.

Table of Contents

Quality	Reliability
Performs according to specification *at time of shipment*	Performs according to specification *over a period of time*
A "quality escape" is a device that does not meet specification at the time of shipment	A "reliability event" is when a device no longer meets specification after some period of time

Table 1: Definitions of Quality and Reliability

A quality escape is a part that is not operating according to specification at the time of shipment. These can occur due to a manufacturing defect—something manufactured incorrectly—that was then not screened and contained by the supplier. Hence it "escaped" to the customer. Details about how Intel works to deliver the highest quality are discussed in the quality section of this paper, with a focus on manufacturing.

A reliability event is when a part that worked as expected at the time of shipment later fails to meet its specification. Reliability events can have several causes, such as transistor aging mechanisms. Details about how Intel works to deliver the highest reliability are discussed in the reliability section of this paper, with a focus on the customer experience of reliability over time.

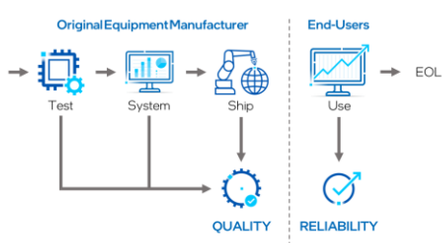

Figure 1: Illustrates original equipment manufacturers shipping products with quality and the importance of reliability for end-users.

Manufacturing for Quality

The highest level of semiconductor quality is achieved through a focus on and prioritization of quality across technology development, product architecture and design, and manufacturing. As a leading semiconductor supplier, Intel constantly strives to deliver the highest levels of quality.

This section will focus on Intel's manufacturing process to enable quality. For additional information about Intel technology development, product architecture, and design, please refer to Quality and Reliability Implications for the Connected World. Topics related to Intel® Xeon® Processor quality and leadership are explained in the recent data center leadership paper.

Intel Testing Process

Intel uses several different quality test steps to monitor and improve quality processes, which helps deliver the highest possible customer quality experience. By supplementing scan test vectors with functional tests, subtle defects that cannot be detected with conventional methods are thereby screened, resulting in significantly improved product quality. Intel develops functional tests for all phases of manufacturing, based on both pre-silicon metrics and post-silicon analysis. Advanced analytics and machine learning methods are used to optimize quality. Quality signals from processor testing are then fed back into the fabrication process.

Processors are initially screened before final packaging in what is referred to as sort testing. This is accomplished with machines that use thousands of probes to connect to the processors while running the test content. Processors that do not meet specifications are not allowed to continue in the manufacturing flow and are scrapped. The processors that pass sort testing are then prepared for package-level tests.

Package-level testing is a more robust process than pre-packaging testing, in both content and duration. The first main packaged test is often referred to as class testing. Class-test machines can run both types of scan tests: The first is where inputs are loaded into internal parts of the die and outputs are then verified for each small part. The second consists of at-speed tests which are used to simulate more complex content.

After class test, some processors undergo system-based testing (SBT). During SBT, multiple operating systems and applications are run to verify that no defective parts escaped earlier testing. SBT hardware is based on a reference board design and can include memory dual in-line memory modules (DIMMs) or input-output devices.

Intel uses software tools to calculate test coverage, which is the fraction of transistors a test will cover. Complete coverage—considered to be levels approaching or exceeding 99%—is not always achievable. This might seem like it would lead to the following, where X is a manufacturing parameter dependent on the product's die size, defect density, and transistor types:

$$\text{Quality Escapes} = X \cdot \underbrace{(1 - C)}_{\text{Test missed}} = 0.01 \approx 10{,}000 \text{ DPM}$$

This equation, however, does not capture reality. An important effect in integrated circuits, described in a 1984 article in the *Journal of Solid-State Circuits*[1] and since then referred to as the "JSSC model," is that one physical defect can cause several logical faults. Logical faults are circuit behavior errors that tests can catch. The distribution of physical defects per die (with most having 0 or 1) largely follows a regular Poisson distribution. Logical faults, however, follow a modified Poisson, where each die with a physical defect also has, on average, many logical faults.

[1] V. D. Agrawal, S. C. Seth, and P. Agrawal, "Fault Coverage Requirements in Production Testing of LSI Circuits," *IEEE Journal of Solid-State Circuits*, vol. SC-17, no. 1, pp. 57-61, February 1982.

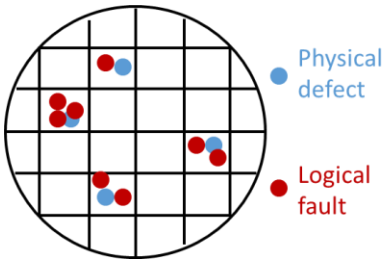

Figure 2: Diagrammatically showing how die that have 1 physical defect almost always have several logical faults, any of which could be correctly caught by a test.

Intel captured this effect in a model of escape-to-customer defects per million (DPM). This replaces the escapes expression above with the following, where n_0 is a measured process-specific parameter.

$$Escapes = X \cdot (1 - C) \cdot \underbrace{e^{-(n_0 - 1) \cdot C}}_{\text{JSSC correction}} \approx 0.0005 \approx 500\ DPM$$

Semiconductor Reliability

Reliability in the semiconductor industry is the understanding of how parts operate according to specification over time. Manufacturers need to understand the leading mechanisms that can cause a reliability event. Semiconductor reliability is dependent on use conditions which can influence different reliability mechanisms in different ways.

Intel focuses on achieving high reliability over the processor lifecycle, since its processors are used in many different conditions with increasing lifespans. This section describes several reliability mechanisms for semiconductor devices and how manufacturers can improve processor reliability for end-users.

Reliability is often divided into three categories based on the age or use time of the part. These classifications are early-life, mid-life, and wearout, as shown in a simplified bathtub curve illustration in Figure 3.

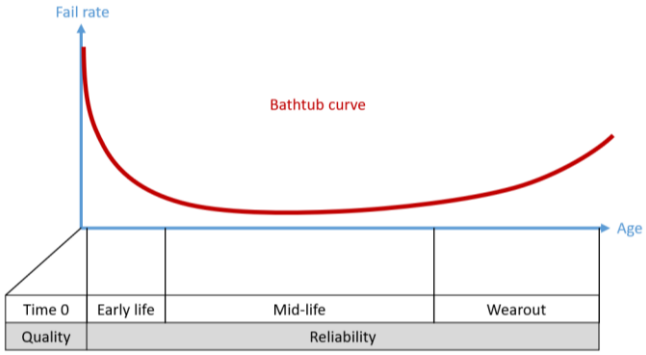

Figure 3: A plot of fail rate (fails per unit time) generically representative of semiconductor devices.

Early Life

Early-life failures (ELF) are historically driven by latent defects. These are defects that require some amount of device use (voltage/temperature/time) to activate. Prior to activation, they otherwise allow the device to operate according to specification.

Figure 4 shows a model of defects which is useful to articulate the concept of ELF mechanisms. As a simplified example, consider conductive defects on a conductive wire grid, where a short circuit causes a failure. The dark blue is the original printed grid width, and the light blue is the effective grid width after degradation. Defects come in a continuous distribution of sizes. If they are large enough, they will always cause a failure. If they are small enough, they never fail. There is a medium size

that, though it does not cause a failure immediately, will later after some amount of degradation. These yellow "ELF fail" defects do not short the dark blue time-0 lines to each other, but they do short the light blue after-degradation lines. In this simplified example, these latent defects cause early-life failures.

Intel optimizes the customer ELF experience through a combination of design and manufacturing optimizations as well as test and stress methodologies, including the test processes described earlier in the quality section. In addition, large area semiconductor devices, including products designed for at-scale data-center applications, often complete a device burn-in step. Burn-in is a tool where processors undergo elevated temperature and voltage stress to activate latent defects. When a latent defect is activated, that processor is scrapped, and thus a processor with a potential ELF event is prevented from shipping to the customer environment.

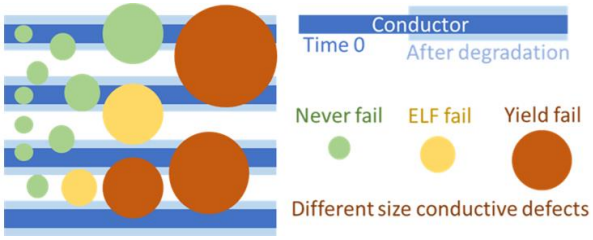

Figure 4: Illustrative visual model of conductive defects on a set of conductive lines. After degradation, ELF defects that are not wide enough to short two conductive lines initially will activate (conductive line widens), leading to a short and representing an ELF event. This type of particle defect model is an oversimplification. Today in advanced semiconductor manufacturing, ELF mechanisms are largely due to our scaling and are caused by transistor patterning variation and the spatial distribution of inherent materials across billions of transistors.

Midlife

Mid-life typically experiences a low and near constant fail rate, as shown in Figure 3. During this time, random failures can be experienced, such as those caused by uncorrected cosmic radiation events.

Wearout

Reliability wearout includes the well-known transistor reliability mechanisms of bias temperature instability (BTI) and hot carrier injection (HCI). Back-end-of-line mechanisms such as time-dependent dielectric breakdown (TDDB) and electromigration (EM), as well as package-related mechanisms driven by temperature cycling and humidity concerns, also contribute to reliability wearout.

Intel optimizes the mid-life and wearout customer reliability experience by the stress and test mechanisms previously described. Additionally, optimizations and quantifications of reliability are completed through a robust reliability verification process. This includes the accelerated testing discussed in the next section.

Accelerated Testing to Optimize Reliability

Reliability testing uses accelerated stress conditions to simulate aging in sampled parts. Accelerated testing allows engineers to understand reliability mechanisms much sooner than would be experienced under normal use conditions, as shown in Figure 5. Reliability test results allow a more complete understanding of reliability limiting mechanisms and thus facilitate high levels of customer reliability.

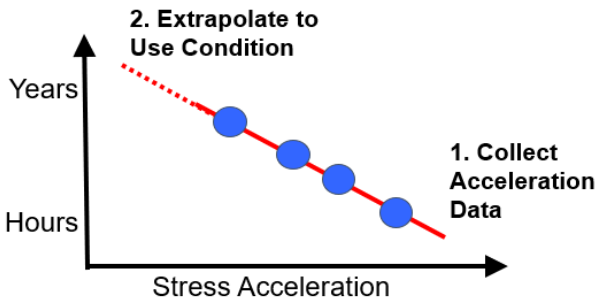

Figure 5: Applying stress (often voltage and/ or temperature acceleration) enables devices to fail in much shorter time periods than possible under normal use conditions. Accelerated testing allows to apply a lifetime's worth of stress to a device in a few hundred hours. This allows engineers the ability to evaluate a semiconductor device's reliability in a more reasonable time scale.

Figure 6 illustrates the distribution of times-to-fail that we might find for two different voltages, with the proper-use voltage shown in blue and a higher-stress voltage in red.

Figure 6: Voltage acceleration allows failure distribution to shift left as compared to normal use conditions.

The times-to-fail scales by an acceleration factor, which is the ratio of the use and stress durations shown in the following equation:

$$\mathrm{AF} = \frac{t_{\mathrm{use}}}{t_{\mathrm{stress}}} = \frac{100{,}000\ \mathrm{hr}}{1{,}000\ \mathrm{hr}} = 100$$

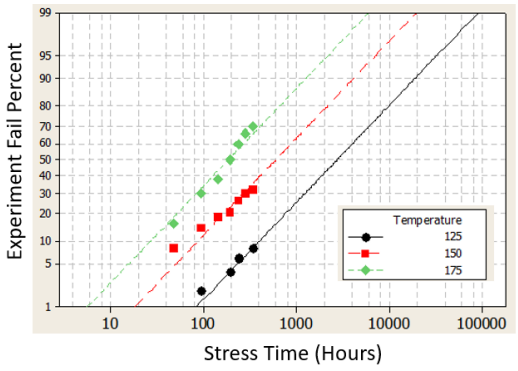

Figure 7: Example stress data with 3 different temperature accelerations tests completed at 125°C, 150°C, and 175°C. The stress experiment lasted 300 hours, with the highest temperature stress resulting in the fastest (and with the largest number of fails) as expected. The scales on the axes are such that a straight line indicates a lognormal distribution of times-to-fail, and all 3 of these datasets are well-fit by a lognormal distribution. Representative accelerated stress data for illustrative purposes only.

Different reliability mechanisms have different stressors. Most integrated circuit reliability mechanisms are stressed by increased voltage and temperature, but many other stresses are possible. These include humidity, current, mechanical stress, and temperature cycles. Example data illustrating temperature acceleration is show in Figure 7.

Before releasing a product, Intel tests, measures, and models many reliability mechanisms to ensure that the processors meet reliability expectations, known as the Intel quality and reliability verification (QRV) process. The QRV process starts by defining the correct requirements early in the product design. It involves comprehensive product design and technology

development integration, implementation of robust and controlled manufacturing processes and systems, and ultimately leads to QRV testing on the finished product.

During QRV testing, Intel uses both standards-based and knowledge-based qualification strategies to ensure that Intel products are shipped with the highest quality while considering the cost and practicality of these requirements. Intel uses industry accepted standards such as Joint Electron Device Engineering Council (JEDEC) and internally developed stress methodologies to ensure products conform to product quality and reliability requirements. QRV stress tests are conducted on units that have completed the full manufacturing flow.

QRV tests include electrical testing for all stages of the reliability bathtub curve with a focus on package/thermal mechanical testing. This includes temperature cycling, high temperature storage, and accelerated moisture resistance testing.

Quantifying Quality & Reliability

Intel estimates quality and reliability levels by placing a random sampling of processors through dedicated tests, which would not be feasible to complete on all processors. With random sampling, the DPM found with any one sample can vary somewhat from the population DPM.

As an illustration, Figure 8 shows the expected measured distributions for three different sample sizes, assuming an actual (or population) DPM of 500. With a sample size of 1 million parts, the resulting binomial distribution has a very small standard deviation. With a sample size of 24K parts, however, the standard deviation is quite a bit larger. Here, in 50% of tests run the sample DPM would fall within a range of 330 to 670. In general, the smaller the sample size the larger the possible variation from the true DPM.

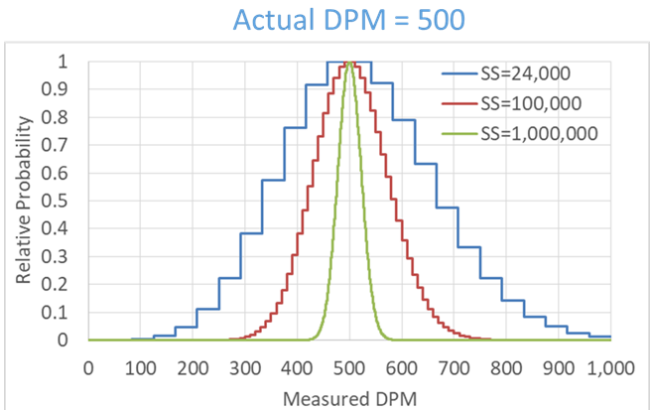

Figure 8: Variation in defects per million (DPM) measured in a subset of population can be described with binomial statistics. When measuring a subset of the population, one would expect to get a result described by a binomial distribution with smaller volume measured resulting in a larger possible variation from the true DPM.

When the true DPM level is not known it is possible to complete an experiment to predict quality levels. In Figure 9 we show an example of the accept quality level (AQL) and reject quality level (RQL). The AQL is a reference DPM we want a high probability of accepting (90% in this case). The RQL is a reference DPM we want a high probability of rejecting (90% in this case is shown as a 10% probability of accepting). In this example, to have confidence in a 500 DPM level, a sample size of 28K units and no more than 14 fails in the experiment would be required.

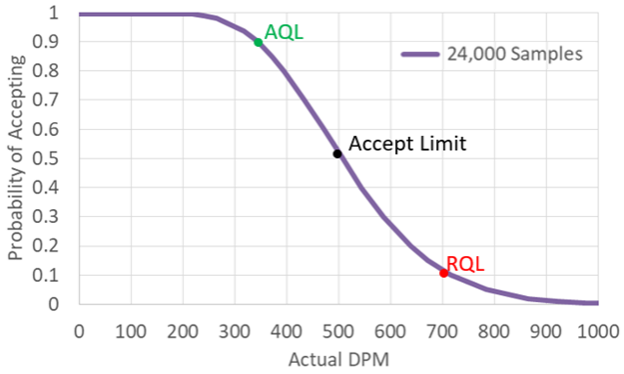

Figure 9: This graph shows the probability of a statistical sample passing a test as a function of the true actual DPM, which is not known.

Conclusion

This paper defined and reviewed quality and reliability concept fundamentals, with a focus on manufacturing for quality and how customers experience quality and reliability over time.

Intel follows a consistent and robust methodology for optimizing product quality and reliability. This process includes technology development, product architecture and design considerations, robust stress and test, and a quality and reliability validation effort.

It is important for Intel and its customers, partners, and suppliers to have a universal understanding of quality and reliability fundamentals. This helps facilitate productive discussions and joint efforts to improve results over time, thereby benefiting the ecosystem overall.