Overview of HALT (Highly Accelerated Limit Test) and HASS (Highly Accelerated Stress Screen) Methodology

What is HALT?

Highly Accelerated Life Testing (HALT) is the process of applying increasing levels of environmental stresses (mainly thermal and vibration) to electronic and electromechanical assemblies well above design and end-use stresses. The goal of HALT is to identify and address weaknesses up to the Fundamental Limit of the Technology (FLT) early in the design process. The FLT is the point at which improvement of margins would require exotic methods or materials. It is critical to have the system operational and monitored during the application of stresses. The ultimate goal is to build a robust assembly, based on standard technologies, with the highest probability of reliability throughout its life cycle. HALT is based on the same principle as the application of derating of components, that is, that more operating margin results in a higher probability of long-term field reliability. HALT is also based on the principle that failure modes that will appear in an electronic product after months or years of low-cycle stresses are the same as those that appear much more quickly under high-cycle fatigue. HALT is not meant to quantify the field life of electronics. The deliverables from HALT are operational and destruct limits, as well as failure mechanisms.

HALT methods have emerged from a new frame of reference, or paradigm shift, in electronics reliability. The frame of reference shifted from focusing on the worst-case stresses a part or assembly will be subjected to in the end-use environment to emphasizing the inherent strength and long life of today's components. HALT methods focus on the physics of failure and on factors that may shorten the inherent long life of electronics. Practitioners of HALT methods view reliability from the frame of reference of knowledge that, for most electronic assemblies, the end-use environment is benign relative to the components, electronic assembly strength, and inherent fatigue life. Past approaches to environmental stress testing were based on the worst-case life-cycle end-use environment. In the early days of electronics manufacturing, discrete components such as tubes and discrete transistors were relatively fragile, and it was a challenge to have them operate reliably in the worst-case end-use environmental conditions. The early discrete electronic components also had, in many cases, a relatively short inherent fatigue life. Wear-out modes in many components would occur during the system's technologically useful life. With the rapidly improving power and performance of IT hardware, technological obsolescence almost always occurs well before hardware fatigue becomes a reliability factor

Electronics components manufactured today have significantly higher thermal capabilities, mechanical strength, and fatigue life than will ever be needed in most applications. In fact, electronic components have doubled their inherent lifetimes, or average time to failure, every fifteen months[1]. Much of this improvement in strength resulted from changes in the requirements of volume circuit board manufacturing processes, such as SMT solder reflow.

Optimizing HALT requires adapting stresses to the limits of a technology. For instance, when applying a vibration to a portable computer, the hard disk drive assembly will have a much lower vibration operating limit than the motherboard. The same is true for the temperature limits. To optimize HALT on a product, the system should be divided, when possible, and the subsystems should have HALT applied separately. The hard disk drive (HDD) and power supply can be placed outside the chamber and connected to the system through extension cables. The system can alternatively be booted from the LAN or a USB key. Understanding the ultimate goals of HALT will help guide the user in the best adaptation of HALT to the technology.

We have a design verification test to show it meets the design specifications; why do we need to stress it beyond the design requirements?

The simple answer is that electronic materials are much stronger and more reliable today, and we have the opportunity to use that high level of strength to find the weaknesses that result in field failures. Electronics components' inherent lifetimes may be 15 or 40+ years, but the fact is, we will never really know. Nevertheless, they are much longer than technological obsolescence. Design validation tests are generally well below the strength of the materials, and do not significantly accelerate the aging or effect of fatigue on the materials. This is especially true for discovering latent defects or design problems that may occur many months after shipment to the customer. Almost all defect mechanisms in electronics that fail in time under low, life-cycle stresses will fail in a short time at much higher stress levels. By applying increasing stresses in HALT, those mechanisms that are the weakest link in an electronics assembly are found in a time frame that you can use to improve the design before hundreds or thousands are made and shipped.

---

[1]Yibin Zhang, Diganta Das, Asaf Katz and Michael Pecht , Örjan Halberg "Trends in Reliability Component Testing", *Semiconductor International*, September, 1999

Increasing the operational margins for a higher probability of field reliability follows the same principle that we use for derating guidelines on components, that is, by increasing operational margins, we increase the probability of field reliability. HALT can be viewed as an empirical system derating evaluation process, determining margins for an entire system and improving them where possible.

Another economic benefit of increasing the operating limits, margins, and strength is the capability to safely apply greater stress levels for very short periods to weed out early life failures or “infant mortality”. This is the basis for the efficiency of HASS (Highly Accelerated Stress Screens) or HASA (Highly Accelerated Stress Audits). The higher the operational margins, the higher the levels of combined stresses that can be safely applied without significantly reducing a system's life. Higher stress levels in HASS processes reduce the time required to detect latent defects in ongoing production. The shorter a HASS process is, the more units that can be processed in a given time, and the result is a more efficient and cost-effective process to ensure that throughout the manufacturing period, there are no hidden defect mechanisms being produced due to an error in component manufacturing or an assembly process going out of control.

If we break it and fix the “weak-link”, then repeat the process, where do we stop, and how do we know when we have made it “robust enough”?

Each limit should be investigated to determine the effort and expense required to increase its strength or capability. The majority of the time, a small change in component value, wattage, orientation, or fastening results in a significant increase in margin. Theoretically, we stop once we have improved the operation and destruct margins as much as possible using standard materials, components, and processes. This point is referred to as the fundamental limit of the technology. In practice, we stop when margins cannot be increased without a major cost or schedule impact, or when margins are equal to or better than those of previous designs that have proven reliable in the field.

So you break it with stress beyond what it will ever experience in its life cycle, how is that relevant to reliability?

To effectively answer these questions, we must step back to the fundamentals of reliability and examine the causes of unreliability. We must observe from historical field data.

- How things fail,
- Why they fail,
- And how variations affect those rates of failures

By understanding the fundamental causes of unreliability, we can most effectively focus on how we create the best possible reliability. Reliability in electronics is fundamentally probabilistic. There is a very wide variety of mechanisms and causes of failures, and very few commonalities between them. An example illustrating how improving operation and destruction margins affects reliability can be seen in the stress (load) and strength distributions.

As long as **STRENGTH > LOAD**, we're OK. We won't have failures.

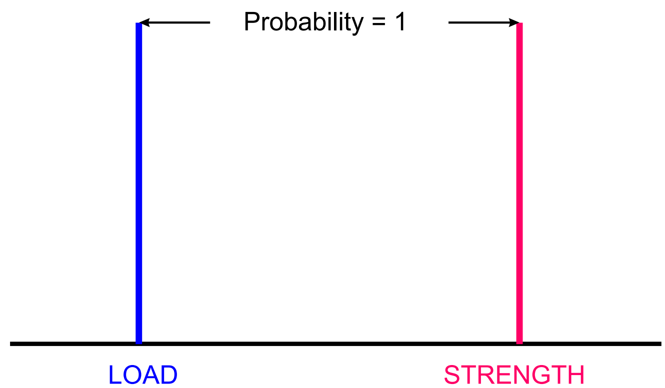

Figure 1 Strength > Load is OK

Of course, Figure 1 is overly simplistic. Right? We know that strength varies between

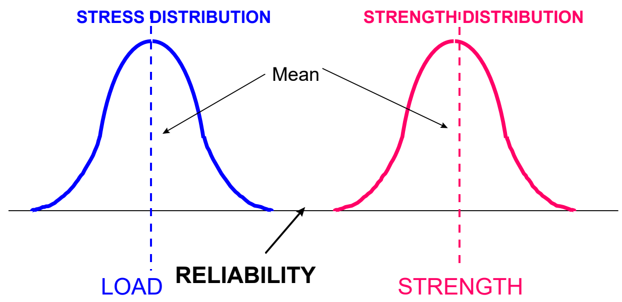

Figure 2 Properties are Distributed

individual products over time of production. Any measurement of a large sample of an electronic assembly's strength or operational parameter would probably result in a bell-shaped distribution without sharp boundaries. Some products are extra-strong. A few are extra-weak, but most fall around the intended design strength.

Similarly, loads vary with some bell-shaped distribution. A few loads will be extra-severe. A few will be extra-gentle.

As long as the **STRENGTH** of the **weakest** product exceeds the greatest **LOAD**, we're OK.

But if “something” brings the two distributions together, as in Figure 3,

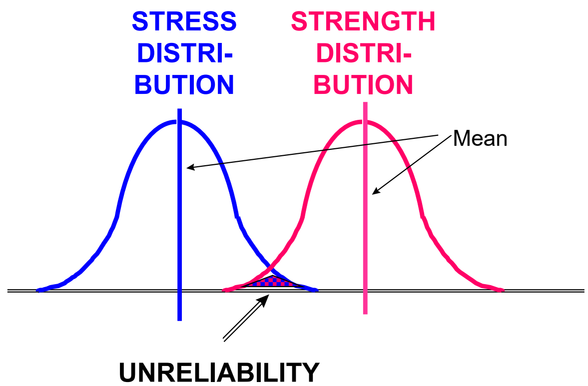

Figure 3 Occasional Max Load > Least Strength

A maximum **LOAD** event will exceed the **STRENGTH** of one of our products, and our customer will have a failure.

What could cause such a shift, such a weakening, such a loss of **STRENGTH** over time? Usually, there are multiple factors and changes that together cause a shift in the margin. It can be an engineering change, a change in machine operators, or a change in your vendors' manufacturing capabilities. Anyway, the **STRENGTH** distribution shifts to the

left, and some failures occur.

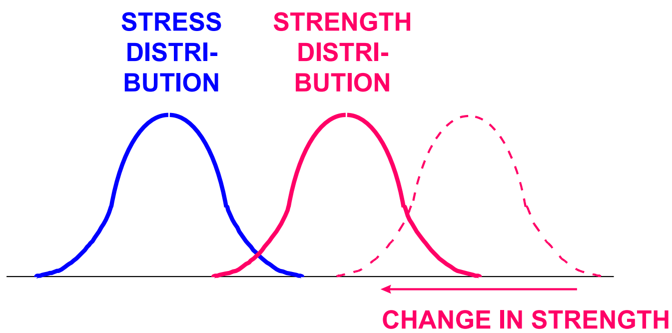

Figure 4 Lessening of Strength

What are we going to do about this? Well, we can simulate the STRENGTH-reducing

The effects of aging are caused by rapidly creating fatigue damage. We do this by applying LOADS or stresses that induce fatigue damage (the physical mechanism of materials aging) greater than normal and watch for any failures. Figure 5.

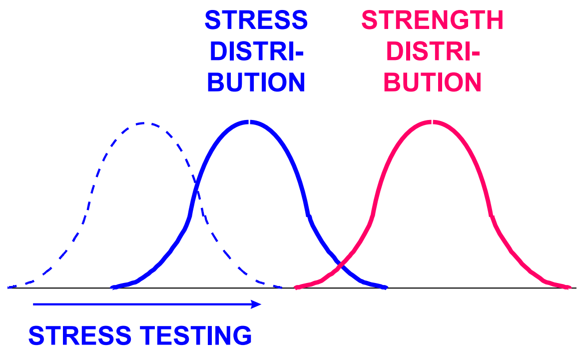

Figure 5 Stress Testing Simulates Aging

OK, having reviewed some basics of failures and reliability, let's observe that electronic equipment has become remarkably reliable. One result is that reliability tests have become very long, very expensive, and time-consuming.

Accelerated testing reduces test time by conducting tests at stress levels above operational conditions. **Thermal cycling** and **random vibration** are the tools most used to accelerate failure mechanisms.

Just a few rules regarding HALT evaluations.

- There are no absolute stress levels that the tested unit <u>must</u> pass.
- Our goal is to improve our product's reliability through improved operational capabilities (margins).
- We must never ship HALT units to a customer, but they can be used for additional in-house testing.

HALT is not a complicated process. Generally, one environmental stress at a time is used to make understanding the physics of the failure mechanism simpler. If time permits, combinations of stresses can be applied in increasing steps.

How many Units will we need?

More is better, and requesting 5 to 10 units is a good start. It is always difficult to get as many samples as we would like during the development stage, because each test and development group needs units to evaluate. It may be that you discover most of the weaknesses in the first few units, and you will need to make design changes before running more through the HALT. Three to five samples of the units to test are an acceptable number to start with. More units may be needed if a better understanding of the strength distribution of a failing mechanism is necessary.

What Stress do we start with?

The order of applied stresses should be from least to most destructive to obtain the most information about each UUT. Typically, this is a sequence of cold step stress, hot step stress, multi-axis vibration in increasing levels, and then combining these stresses to be applied simultaneously. For low-temperature stress, the product or sub-subsystem is placed in a HALT chamber, and the temperature is stepped down in 10 °C steps, holding each step long enough for the entire product to reach the set point. This is typically 10-15 minutes for most products, or as long as needed to run a diagnostic program. Then, decrease the temperature, continuing until it no longer operates. Generally, if the product is operational below -30C, the typical low-temperature operating limit for most electronics, little change is needed.

The second part is high-temperature step stressing in 10ºC increments to the operational limit. This is typically, but not always, a destruct limit. Generally, HALT users consider an upper operating limit of +100-110ºC to be sufficient for most electronics, with no changes necessary. However, it is important to recall that more margin means a higher probability of field reliability, and that higher stresses for shorter durations can be applied if a HASS process is used for an Ongoing Reliability Test (ORT). Thermal operating limits below +100ºC that are not caused by over-temperature protection circuits should be thoroughly investigated and improved. Generally, over-temperature protection circuits should be disabled to discover the true hardware margin capability during HALT. When these cannot be overridden, as in the internal thermal protection shutdown circuit of a CPU, other means should be investigated to suppress the engagement of these circuits, such as separately cooling the component while the rest of the motherboard is heated.

Don’t we have to redesign the products we just finished designing when we find a weakness in HALT?

Most failures found in HALT and relevant to field reliability do not, many times, require a redesign. Figure 6 (courtesy of C. Felkins, StorageTek) shows examples of how

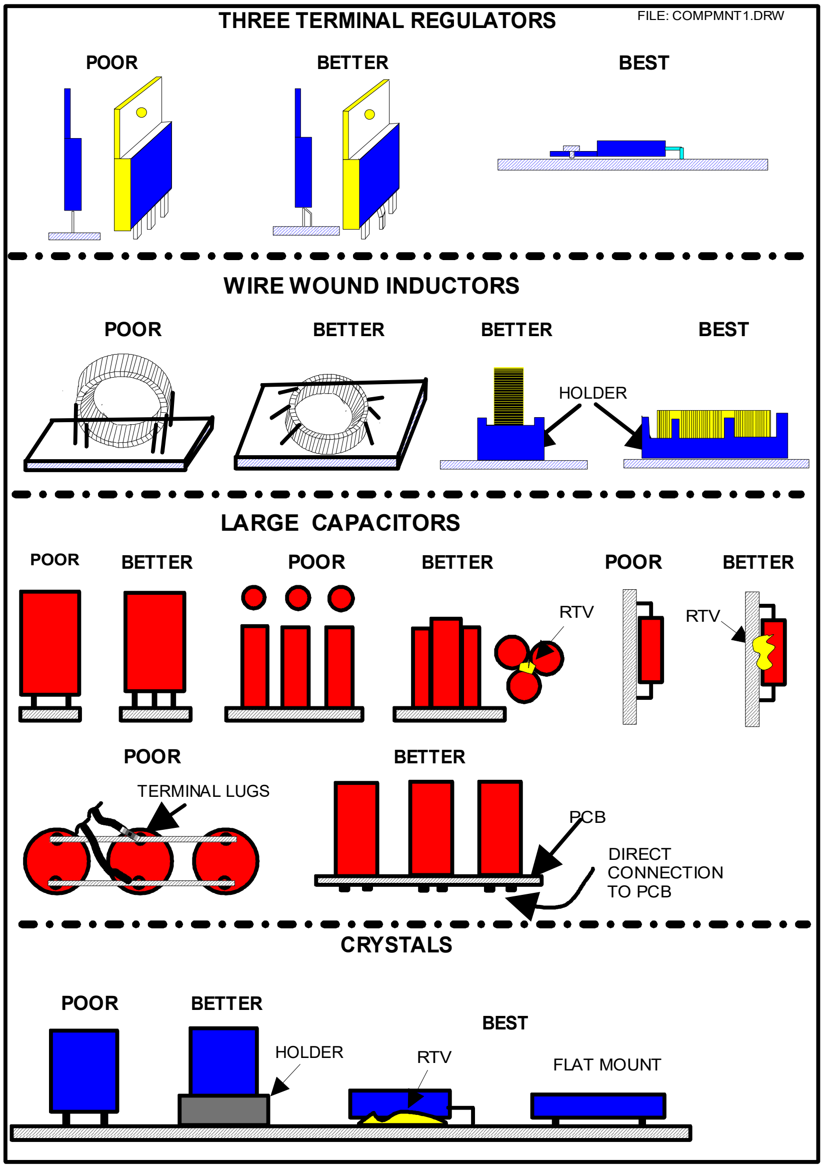

Figure 6. Component Mounting Improvements for improved strength

Changes to components' mechanical attachment can lead to large increases in mechanical strength.

FAQ's on HASS and HASA

What is HASS, and what is HASA?

HASS (Highly Accelerated Stress Screening) is the process of applying a combination of environmental stresses (mainly thermal cycling and random multi-axis vibration) to rapidly precipitate and detect latent defects in electronic products to be shipped. The stress levels in HASS are determined by derating the operation and the destruct limits found during the last HALT iteration. The term "HASS" means it is applied to 100% of the production population. HASA (Highly Accelerated Stress Audit) is the use of Highly Accelerated Stress Screens on a sample lot as an audit process to identify latent defect trends in the larger production population rapidly. Most HASS and HASA processes consist of 4 to 5 thermal cycles, with the last one to two hours of vibration. Thermal and vibration stress levels HASS and HASA are developed in the same way from the HALT evaluation. Unlike HALT, HASS/HASA has pass and failure criteria.

HASS is most cost-effective and beneficial for products with higher initial costs, small- to medium-volume production, and high reliability requirements. HASA is most beneficial when the manufacturing process is established and stable, and when the goal is to detect a major shift in strength or a defective lot during the sustaining manufacturing process.

HASA is primarily used for electronics that have lower initial costs, lower reliability requirements, and high-volume production. The efficiency of finding a defective population depends on the sampling rate.

So, after we finish with HALT and release the design to production, how do we determine the stress levels for the HASS or HASA process?

The HASS or HASA process for precipitating and detecting potential latent defects is derived from the final HALT iteration of the operation and destruct limits. A typical HASS/HASA process combines thermal cycling from 15º C below the maximum operational temperature to 10 to 15º C above the lowest operational temperature while simultaneously applying one-half the maximum vibration operational (or destruct) limit. Four to five thermal cycles are applied, with vibration applied at different levels, never exceeding half the destruct vibration level. A HASS/HASA process typically takes only one to two hours to complete. Depending on the size of the chamber and the required monitoring equipment, multiple units may be processed per HASS/HASA cycle. Another way to reduce HASS/HASA process costs is to use higher levels of stress. It has been demonstrated in the predecessor process, Environmental Stress Screening (ESS), used by military electronics suppliers that the higher the thermal delta, the higher the thermal change rate, and the higher the vibration applied during screening, the less time is needed to detect a hidden defect. Since the HASS limits are based on the maximum stress levels found in HALT, pushing thermal and vibration margins in HALT during the design phase also reduces the costs and time to apply HASS.

It is very important to have the HASS units powered and monitored while the stresses are applied. There are many types of latent defects, such as cracked solder joints that may only be detected momentarily when stress is applied, and are not detectable when stress is removed. HASS is applied in the same type of environmental chamber used to perform HALT. HALT/HASS chambers use liquid nitrogen to rapidly cool the UUT (Unit Under Test) and electrical resistive heating to rapidly heat it. These chambers are capable of thermal change rates of 60º C/minute on the UUT. In addition, multi-axis vibration is provided by pneumatically driven hammers on a vibration table that produces pseudo-random broadband vibration. Vibration produced in HALT/HASS chambers is sometimes referred to as Repetitive Shock (RS) vibration to distinguish it from the vibration produced by traditional single-axis electrodynamic shakers. RS systems have no control over the spectrum or axis intensity. They produce a three-dimensional “fuzz ball” of vibration. The goal is to rapidly induce mechanical resonance in all axes simultaneously. Mechanical resonance induces fatigue, which brings latent defects to a detectable state during operational monitoring.

Testing beyond specification seems risky, and it could damage good units. How can we be sure we are not shortening a system's total fatigue life or damaging good units that we ship?

Of course, we want to be sure that using a high level of combined stresses will not

- Induce flaws
- Leave cosmetic damage
- Remove significant fatigue life

Before a HASS process is applied to the manufactured product and shipped, a process called the Proof of Screen (POS) must be completed. A POS consists of taking new samples of the product and applying the HASS or HASA stress regimen multiple times, 10 to 20 times. The assumption is that if 10 screens remove all the fatigue life from the product, 1 application will still leave 90% of the life remaining. Typically, products can be subjected to much more than 10 applications of HASS, and some companies performing HASS apply 30–50 applications for the POS, for even more confidence that the screens are relatively benign for non-defective units.

The process of HALT is intended to find the margins beyond the specifications, and HASS is intended to monitor ongoing production to ensure that each unit or lot (using HASA) does not have a significant decrease from those margins. HALT and HASS reference the empirically observed strength of the system under test and do not rely on specifications or end-use environments. The manufacturer determines specifications for components and systems under the maximum conditions they will encounter in end-use environments to meet long-term operational specifications. Much of the strength in components comes from the need in the manufacturing process to heat the component to reflow solder. The ability to withstand high temperatures and rapid temperature changes during manufacturing is why we use much higher stresses than in end-use conditions to identify the components and assemblies that will eventually fail over time.

How do we know that an EC or component change does not make the product weaker or have less fatigue life?

That is a good question, and the basic answer is that we don’t know what effect a change in a component might have on a previously robust design. Often, if a change to the product reduces the overall strength, a HASS or HASA process will detect it. If the change to the product is suspected to be significant, the product HALT should be reapplied after the change to assess whether the margins have shifted. The amount of change required before re-running a HALT is an engineering judgment call. Most of the time, it is not done, and the effect of the change is seen or not seen in the HASS or HASA process. Most EC changes that have a reliability risk are detected in HASS and HASA.

When can we stop HASS or HASA?

Screening can be an effective and efficient way to find defects, especially in early production, when the learning curve is steeper, and the risk of reliability issues is higher. There is, of course, no way to know when a vendor or its component suppliers may create a latent defect in their products, and therefore, when a HASS or HASA will provide significant payback for catching the defect. Depending on the customer's cost of failure, HASS or HASA may remain a cost-effective process to prevent field failures throughout the sustaining manufacturing period.