Why Specify Workload?

by George Tyndall, Ph.D.
Senior Director, Advanced Reliability Engineering

Executive Summary

The reliability of hard disk drives has been quantified historically by a mean time to failure (MTTF), or an annualized failure rate (AFR), defined at a specified operating temperature, and an assumed functional duty cycle. In the following, we provide justification for replacing the ambiguous concept of duty cycle with the readily quantifiable "workload", which is defined as the total amount of data read from or written to the drive per unit time. This relatively subtle change provides the end-user with far greater clarity into the expected field reliability of HDDs in a variety of applications.

Hard disk drives (HDDs) are complex electromechanical devices that integrate many diverse technologies in their subassemblies. Given the number of constituent components characteristic of modern HDDs, there are several mechanisms of failure that can occur during different stages of the product lifetime. The time-dependence of the HDD failure rates has most often been described using a classical "bathtub curve" characterized by three distinct time regimes:

a) an early life (infant mortality) period where the failure rate drops with increasing time
b) the normal service life where a constant failure rate exists
c) the "wear-out" region where the failure rate increases with increasing time.

Modeling such a complex time dependence is difficult, and more importantly, largely unjustified.

Over the serviceable life of a well-designed HDD, the time dependent failure rate can be adequately treated utilizing a two (or three) parameter Weibull model. The general expression for the cumulative failures as a function of time, F(t), is given in this model by:

$$F(t) = 1 - exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right] \qquad (1)$$

where t= power-on-hours (POH), η is the time constant, β is the shape parameter. At low failure rates (nominally < 10%), the expression for the cumulative failures simplifies to

$$F(t) = \left(\frac{t}{\eta}\right)^{\beta} \propto (POH)^{\beta} \qquad (2)$$

As indicated by equation (2), the cumulative failure rate scales with POH as a simple power law. The shape parameter, β, is always less than unity over the serviceable life of an HDD, and generally lies in the range of 0.6 ± 0.2. Consequently, the HDD failure rate (or number of HDD failures per unit time) will decrease with increasing time, and the cumulative number of failure vs. time will display a concave curvature. A representative example of a field return profile over the product warranty period (3 years in this case) is shown in Figure 1.

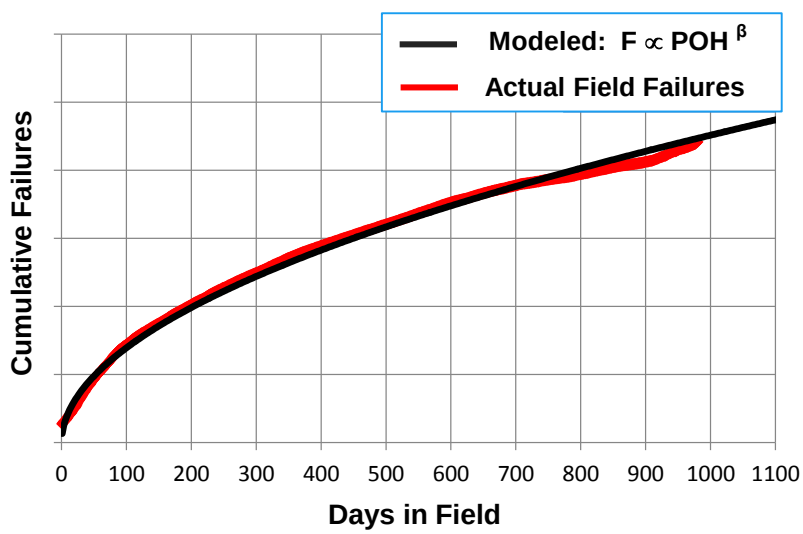

Figure 1. Comparison of the observed field failures with the modeling based on a power law dependence of failure rate vs power-on-hours (POH). The value of the shape parameter used to fit the data was β = 0.6.

A time-centric metric, the mean time to failure (MTTF), has been universally used to quantify the reliability of HDDs since power-on-hours is thought to be the critical parameter. In addition to MTTF, the annualized failure rate (AFR) is often used to characterize HDD reliability. The AFR

couples the intrinsic reliability of the drive as defined by the MTTF together with the actual POH in any given field application to derive an anticipated field failure rate. The annualized failure rate and the mean time to failure are simply related by:

$AFR = \frac{POH}{MTTF}$, for the mathematically simplifying case of β = 1. For other values of β, this expression will be slightly modified.

The reliability of disk drives also depends on both environmental and usage conditions. At extremes of either temperature or usage, the reliability of the HDD will degrade. The impact of temperature on HDD failure rates has been the topic of numerous studies and it is well established that HDD failure rates increase at elevated temperatures. The functional form of the increase in failure rate with increasing temperature is consistent with the well-known Arrhenius equation. Namely, the temperature-dependent failure rate, F(T) can be expressed as

$$F(T) \propto exp\left\{\frac{-E_A}{kT}\right\} \qquad (3)$$

where E_A is the activation energy, and k is the Boltzmann constant. HDD manufacturers typically define the quoted MTTF specification as the reliability that can be expected if the drive is operated at the nominal field temperature. This temperature is application specific, but is typically assumed to be on the order of 40°C as measured on the HDD base casting. For continual operation at temperatures in excess of this nominal, the field reliability will be compromised. Using a typical value of the activation energy, E_A =0.4 ± 0.1 eV, the failure rate will approximately double for every 15 – 20 C increase in temperature. The dependence of failure rate at low temperatures has not been studied as extensively.

The failure rate of HDDs also depends on the extent to which a drive is used during the power-on-time. The HDD industry has historically invoked the concept of functional duty cycle to account for this effect. Duty cycle is most often defined as the percentage of the power-on time that the drive is performing a host-initiated operation. A power law dependence of the failure rate on duty cycle (DC) is usually assumed. When coupled with the time dependence discussed above, the failure rate can be written as,

$$F \propto (DC)^x \; x \; (POH)^\beta \qquad (4)$$

where x ≤ 1. While incorporating a drive usage term into the reliability model is an intuitively appealing concept, duty cycle has been of only limited success in accounting for the absolute magnitude of failure rates in various field applications.

The impact of duty cycle on test failure rates was recently investigated at WD. Reliability Demonstration Testing (RDT) was conducted on nominally identical populations of HDDs using two distinct test scripts. Both test scripts operated at 97 - 98% duty cycle, but differed by 33% in the amount of data transferred per unit test time. These RDT test scripts were created by changing the mix of sequential vs. random operations, and the block sizes transferred. The cumulative failure rates vs. time for both tests are shown in the Weibull plot of Figure 2. As is apparent from this figure, the drives tested using the 180 GB/hr test script displays an absolute failure rate that is higher than that of the population tested at the lower workload of 120 GB/hr. This is true for all test times. These results contradict the expectations based on equation (4). In particular, since both test scripts used operated at the same duty cycle, equation (4) predicts that both populations should display identical failure rates. Perhaps even more disconcerting is the fact that two distinct MTTF values are obtained depending upon the manner in which the HDDs were tested. These results clearly illustrate that subtleties of drive usage can greatly impact reliability, and that these effects are not captured by the currently-used duty cycle treatment.

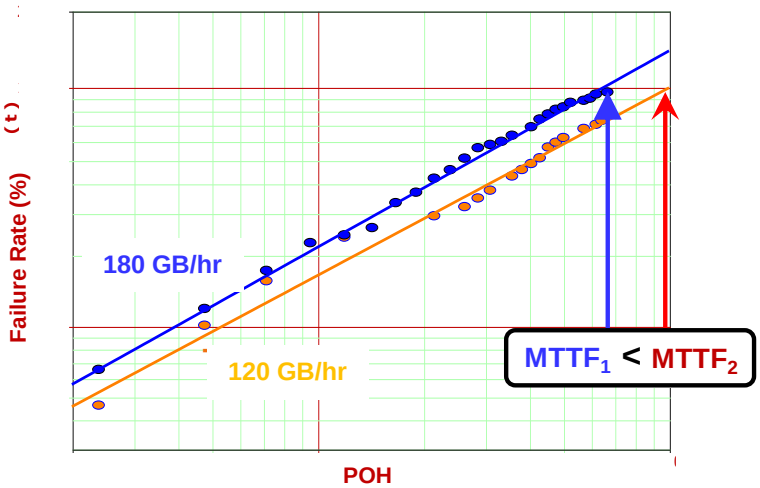

Figure 2. Weibull plot of Cumulative Failure Rates vs POH hrs for the same drive tested with two RDT scripts. The shape parameter for both data sets is identical, β = 0.8. The analysis illustrates that the MTTF is not uniquely defined, but rather depends on the specifics of the RDT test.

In order to explain the above discrepancies, and thereby develop the framework for a more rigorous model for HDD reliability, we briefly discuss the physics of failure for the dominant HDD failure modes, and how these failures are impacted by drive usage. Historically, the head-disk interface (HDI) has been key to achieving HDD reliability. The probability of HDI-related failures is directly related to the physical spacing between the magnetic recording head and the rotating disk. This spacing, or clearance as it is known in the industry, has steadily decreased as the magnetic recording density (capacity) of HDDs has increased. Today's HDDs

operate at head-disk clearances where it becomes increasingly difficult to completely eliminate intermittent head-disk contacts that can lead to a myriad of HDI-related failure modes including:

a) creation of disk defects / scratches
b) poor writes and off-track writes
c) modulation in the written signal
d) degradation of the electrical and magnetic performance of the head.

A relatively simple tolerance model can be used to qualitatively understand the probability of failure at the HDI as a function of clearance. This model assumes that if a significant number of head-disk contacts occur at any given interface, then the HDD will eventually suffer an HDI-related failure. The number of interfaces prone to contact in any given product design is given by the ratio of the mean clearance (μ) between the head and the disk to the standard deviation in clearance (σ). This quantity is known as the Z-factor in statistics and tables are readily available that allow one to determine the fraction of interfaces that will be prone to failure due to insufficient clearance. In order to design a robust head-disk interface with an acceptable failure rate, the value of the Z factor should be on the order of $Z \geq 3$. (Note: $Z = 3$ yields an HDI failure rate of 0.14%). As mentioned above, much of the aerial density gain in HDDs has been achieved by reducing the mean head-disk clearance. To maintain a comparable HDI failure rate, this has required that the standard deviation in clearance drop proportionately. Given that today's mean clearances are on the order of 1 – 2nm, it follows that the standard deviation in clearance must be on the order of 0.3nm – 0.6nm. To put this into perspective, the Van der Waals diameter of a carbon atom is 0.34nm. Controlling the clearance to atomic dimensions is a major technological challenge that will require continued improvements in the design of the head, media and drive features to achieve the necessary reliability.

In order to improve the robustness of the HDI, all HDD manufactures have in the recent past implemented a technology that lessens the stress at this interface. Without delving into the details of this technology, the basic concept is to limit the time in which the magnetic elements of the head are in close proximity to the disk. In previous products, the head-disk clearance was held constant during the entire HDD power-on time. With this new technology, however, the head operates at a clearance of >10nm during seek and idle, and is only reduced to the requisite clearance of 1 – 2nm during the reading or writing of data. Since the number of head-disk interactions becomes vanishingly small at a spacing of 10nm, the probability of experiencing an HDI-related failure will be proportional to the time spent reading or writing at the lower clearance level. The fact that all power-on-time should not be treated equivalently has not been previously discussed in the context of HDD reliability modeling.

The physics of failure indicates that the time spent with the pole-tip of the magnetic recording head in close proximity to the disk is more important to HDD reliability than the total power-on hours. The time spent reading and writing data is linearly proportional to the product of the total power-on hours and the workload per unit time (WL). In this context, the workload per unit time is defined as the total host-to-drive data transfer and includes both reads and writes. The most convenient measure of WL for the purpose of this discussion is TeraBytes per year (TB/yr). If we replace the total power-on-time in equation (3) by the time spent reading and writing data, and assume one year of operation, then the following expression for the annualized failure rate is obtained,

$$AFR \propto (WL \; x \; POH)^{\beta} \propto TB^{\beta} \qquad (5)$$

where TB (TeraBytes) is the total data transferred. It immediately follows from comparison of equations (2) and (5) and the fact that $\beta < 1$ in both cases, that the cumulative failures vs. TB transferred in the field will display a convex curvature analogous to that of Figure 1. That is, the number of HDD failures per unit time will decrease as the total amount of data transferred increases.

The major implication of equation (5) is that the critical reliability parameter is the total data transfer, and not the total power-on-hours. The validity of this expression is readily demonstrated using the same data presented above in Figure 2. A Weibull analysis of the failure rates obtained with the two different workload scripts was performed by replacing the total power-on time with TB transferred as the dependent variable. The results are shown in Figure 3. As is apparent from this plot, the cumulative failures vs TB transferred dependence is identical for the two populations. This confirms that TB transferred, or equivalently the time spent reading and writing data, is the critical parameter governing HDD reliability.

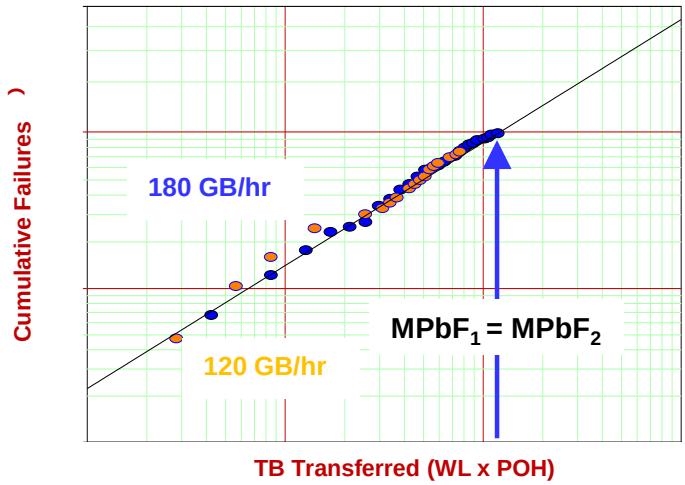

Figure 3. Weibull plot of Cumulative Failure Rates vs TeraBytes transferred for the same data presented in Figure 2. The shape parameter determined from the solid line fit, β = 0.8, is the same as that determined from the data when plotted vs POH.

The above discussion and the results shown in Figure 2 illustrate that MTTF is woefully inadequate in describing the intrinsic reliability of HDDs. Since the dominant failure modes limiting HDD reliability scale with the total data transfer, it follows that the Mean Petabytes to Failure (MPbF) would be the metric of choice. However, until the industry as a whole accepts MPbF as the preferred measure of product quality, WD will take the interim approach of specifying both the MTTF and the maximum workload at which the product meets the MTTF requirement.

In order to illustrate the concept of the added workload specification, we use data taken from the RDT testing of a recent product. Figure 4 shows a Weibull plot of the failure rate vs TB transferred where the solid circles represent the test failures. A least squares fit to the experimental data, shown as the solid blue line, yields a shape parameter, β = 0.6, indicative of a failure rate that decreases with increasing TB transferred. The RDT test data is then converted into a projected field failure rate, shown as the red line, by accounting for the difference in test vs expected field temperature (40C) using the Arrhenius equation (2). We stress that the projected field reliability will not be met if the HDD is continually operated at temperatures in excess of the 40C nominal temperature. Now, assume: a) the product shown in Figure 4 is destined for use in a 24 x 7 application (8760 POH/year), b) is specified to have an MTTF of 1.2M hrs, and c) is warranted for 5 years. In order to determine the maximum workload that will ensure the MTTF specification is met, we first determine the maximum number of drive failures that are allowable during the warranty period.

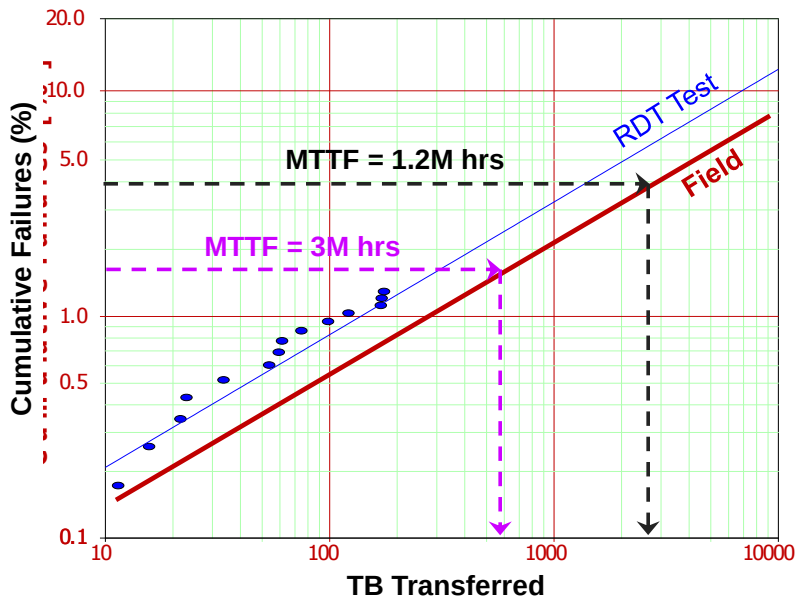

Figure 4. Weibull plot of Cumulative Failure Rates vs TeraBytes (TB) transferred for a recent WD product. The solid circles represent the failures observed during RDT testing. The solid blue line is least squares fit to the data and the solid red line indicates the projected failure rate in the field vs the total data transfer. Dashed lines are used to illustrate the maximum workload at which the MTTF of the product would be met.

Using the approximation that AFR = $\frac{POH}{8760}$, we obtain a maximum total failure rate of: 5 x $\left(\frac{8760\,hrs}{1.2M\ hrs}\right)$ = 3.65%. This total failure rate level is indicated by the horizontal dashed line denoted by MTTF = 1.2M hrs in Figure 4. The maximum allowable data transferred during the 5 year warranty period can then be read directly from the plot and is 2.8 PetaBytes, or 560 TB/year.

It is worth reiterating that both MTTF and the data transfer specifications must be taken together when ascertaining the relative reliability of an HDD. If the HDD is operated at data transfer levels that are less than the specified maximum, then one can expect higher levels of reliability. As an example, assume that the drive in a particular field application only transfers 120TB/year for each of 5 years (600 TB total). The total failure rate expected after 5 years will drop to nominally 1.5% and the corresponding MTTF will increase to 3M hrs. Conversely, if the HDD is continuously run at total data transfers in excess of the maximum product specification, then the reliability will be degraded. The above examples serve to highlight that MTTF values quoted in the absence of a specified workload are meaningless at best, and often times misleading.

A final comment is warranted on the workload specifications discussed herein for HDDs and the TB written limitation (or Endurance specification) of Solid State Drives (SSD). These two specifications, and the reasons underlying their use, are completely unrelated. In the case of HDDs, the workload defines the failure rate growth curve over the normal serviceable life of the drive, and is used in conjunction with MTTF to unambiguously define the HDD reliability. It is in no way related to the lifetime of the HDD. In contrast, the Endurance specification in SSDs does indeed define the lifetime of the device. When the TB written in an SSD exceeds the specification, the product enters the classical "wear-out" regime where the failure probability increases rapidly, and the reliability drops precipitously.

In conclusion, MTTF alone is insufficient to fully describe the field reliability of HDDs. This results from the fact that HDD failure rates are more tightly coupled to the total amount of data transferred rather than the total power-on-time. In order to more fully characterize the quality level of HDDs, WD will now specify both the MTTF and the maximum workload at which the MTTF will be met in the field. The additional workload specification together with WDs expanded product portfolio will ensure that customers can select and utilize HDDs with the requisite quality levels to meet the data transfer demands of any specific application.

For service and literature:

http://support.wd.com
www.wd.com

800.ASK.4WDC	North America
800.832.4778	Latin America (Spanish)
+800.6008.6008	Asia Pacific
+800.ASK.4WDEU	Europe (toll free where available)
+31.88.0062100	Europe, Middle East, Africa

Western Digital
3355 Michelson Drive, Suite 100
Irvine, California 92612
U.S.A.

2579-772003-A00 Jun 2013