2008-07-14

Predicting the Longevity of DVDR Media by Periodic Analysis of Parity, Jitter, and ECC Performance Parameters

Daniel Patrick Wells
Brigham Young University - Provo

BYU ScholarsArchive Citation
Wells, Daniel Patrick, "Predicting the Longevity of DVDR Media by Periodic Analysis of Parity, Jitter, and ECC Performance Parameters" (2008). *Theses and Dissertations*. 1530.
https://scholarsarchive.byu.edu/etd/1530

PREDICTING THE LONGEVITY OF DVD-R MEDIA BY PERIODIC ANALYSIS OF PARITY, JITTER, AND ECC PERFORMANCE PARAMETERS

by

Daniel P. Wells

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

School of Technology
Brigham Young University
August 2008

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a thesis submitted by

Daniel P. Wells

This thesis has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

Date	Barry M. Lunt, Chair
Date	Joseph J. Ekstrom
Date	C. Richard Helps

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the thesis of Daniel P. Wells in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

Date

Barry M. Lunt
Chair, Graduate Committee

Accepted for the School

Val D. Hawks
Director, School of Technology

Accepted for the College

Alan R. Parkinson
Dean, Ira A. Fulton College of Engineering
and Technology

ABSTRACT

PREDICTING THE LONGEVITY OF DVD-R MEDIA BY PERIODIC ANALYSIS OF PARITY, JITTER, AND ECC PERFOMANCE PARAMETERS

Daniel P. Wells
School of Technology
Master of Science

For the last ten years, DVD-R media have played an important role in the storage of large amounts of digital data throughout the world. During this time it was assumed that the DVD-R was as long-lasting and stable as its predecessor, the CD-R. Several reports have surfaced over the last few years questioning the DVD-R's ability to maintain many of its claims regarding archival quality life spans. These reports have shown a wide range of longevity between the different brands. While some DVD-Rs may last a while, others may result in an early and unexpected failure. Compounding this problem is the lack of information available for consumers to know the quality of the media they own. While the industry works on devising a standard for labeling the quality of future media, it is currently up to the consumer to pay close attention to their own DVD-R

archives and work diligently to prevent data loss. This research shows that through accelerated aging and the use of logistic regression analysis on data collected through periodic monitoring of disc read-back errors it is possible to accurately predict unrecoverable failures in the test discs. This study analyzed various measurements of PIE errors, PIE8 Sum errors, POF errors and jitter data from three areas of the disc: the whole disc, the region of the disc where it first failed as well as the last half of the disc. From this data five unique predictive equations were produced, each with the ability to predict disc failure. In conclusion, the relative value of these equations for end-of-life predictions is discussed.

ACKNOWLEDGMENTS

While there are many who deserve acknowledgment for their part in making this paper possible, the first that come to mind are my committee chair, Dr. Barry Lunt, and the department for their support and understanding, not only throughout the program but more specifically during my absence as my family and I endured several personal trials.

I also want to thank Dr. Jaron Hansen from the Department of Chemistry and Biochemistry, as well as Dr. Dennis Eggett from the Statistics Department. Both of whom provided helpful instruction on various aspects of my research.

My wife and children deserve a very special thank you for their patience and help throughout this entire process. This would not have been possible without their support. And to my many other family members and friends whose support and assistance were invaluable, thank you very much.

Finally, I want to thank my father who instilled in me from a young age a love for learning and the aspiration for excellence.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Introduction

Since the advent of recordable DVD media (DVD-R) in 1997, manufacturers have been advertising their products as a valid solution for archiving digital content. Some of them have even claimed their disc had "life spans ranging from 30 to 100 years…" (Bennett 2004, 29). In 2004 the National Institute of Standards and Technology (NIST) released a study which analyzed the "reliability and longevity" of both commercial CD-R and DVD-R media. The test showed that while some DVD-Rs may be suitable for archiving, others were definitely not (Slattery et al. 2004). The results of the test caused great concern in the archival industry since there was no way for the consumers to determine which type of disc they might be buying. Presently, there are still no tests or standards in place that allow manufacturers to accurately and consistently rate the longevity of their particular brand of DVD-R.

If the point of failure of any burned DVD could be predicted by the consumer, it would alleviate one of the largest concerns of using the recordable DVD as an Archival Format. Our research will determine whether or not the error information provided by the parity, jitter, and ECC performance parameters, taken periodically over time, will enable us to reliably predict the failure point for any DVD without prior knowledge of the manufacturing process used to create it.

1.1 Recordable Optical Storage, a Brief History

Optical disc technology has changed the way people view portable digital media. It was first introduced to the consumer in the form of the Audio CD, released near the end of 1982 in Japan and Europe. Roughly two years later, in 1985, the CD-ROM was released which dramatically increased the possibilities of using optical discs as computer storage. The advent of these two forms of media produced an explosion of ideas and uses for optical technology.

Within a decade, manufacturers began to release various non-standardized optical disc formats that showcased the ability to increase the data capacity of the CD. The movie industry, recognizing the value of this increased capacity, approached the computer industry to initiate talks to discuss requirements for a technology designed to store and deliver high quality protected video. In January of 1995, two competing formats were proposed. Sony and Philips submitted their Multi Media Compact Disc (MMCD) while Toshiba and its partners backed the Super Density (SD) Format. After much deliberation, the industries joined forces and the Digital Versatile Disc (DVD) Consortium was born. By the end of 1995 the DVD format had been finalized (Purcell 2000, 1-11).

At first, very expensive equipment was required to master content to both the CD and the DVD. In 1988, Philips and Sony created the first standard for a recordable CD. There were several other proposed specifications but one, the CD-R, won out in the end. The introduction of the CD-R made mastering content to a CD much more affordable and therefore brought it into the home. Simply stated, a CD-R is created by using a laser to “burn” the data onto a layer of dye within the disc. These burned spots reflect laser light differently than the unburned portions. Using these changes in reflectivity, another laser would be able to read the data from the recorded CD (Purcell and Martin 1997, 23-28). It is important to note that the dye layer was organic in nature and therefore susceptible to degradation over time. Some dyes lasted longer than others and could better withstand the elements. Over time, these more resilient dyes became the standard. Currently, chemically stabilized phthalocyanine, cyanine and azo dyes are the standard for CD-R technology (Slattery et al. 2004, 520).

Recordable DVDs became available in 1997. Since this technology is relatively young when compared to CD-R technology, its manufacturing processes are not nearly as stable or as standardized. Due to this fact, the quality of the recordable DVD media can be very inconsistent and even unreliable (Slattery et al. 2004).

1.2 CD-R as an Archival Format

In the mid to early 1990’s, the majority of the CD-R media manufacturers spent their time testing consistent manufacturing processes and compatibility issues between readers, recorders and media. However, a few like Eastman Kodak and Mitsui Toatsu Chemicals (MTC), submitted very optimistic test results detailing the longevity of their media. Both companies used extreme heat and humidity to accelerate the aging process of the discs—a technique called Eyring Acceleration Modeling. With this approach, they could estimate how long their discs would last under normal conditions before failure occurred. Eastman Kodak predicted that 95% of their product would be reliably read after 217 years. MTC claimed that, on average, their media life was over 300 years.

Another conclusion from the MTC study, which contradicted common beliefs at the time, was that the CD-R had a significantly longer life that the CD-ROM (Leek 1995).

The archival industry became very excited about the CD-R as an archival storage format. Leek quotes Jerry McFaul, the chairman of the Special Interest Group on CD-ROM Applications and Technology Foundation (SIGCAT), during the organization's 1995 conference:

> CD-Recordable represents the technology that the federal government has been looking for to preserve billions of dollars worth of data collected at taxpayers' expense…The archiving user community is encouraged by the rigorous CD-R testing done by the media manufacturers, particularly Eastman Kodak, because CD-R longevity is an important issue and CD-R media testing lays the foundation for additional archiving uses of CD-R…Even if the media manufacturers' estimate of longevity is off by 50 percent, meaning that CD-R life may only be 50 years, CD-R media is still better than magnetic tape media…

It seemed clear that the CD-R was quickly on its way to becoming the standard for archiving digital media. Discs were easy to create, easy to store, easy to handle and now it appeared they would outperform the current archival storage mediums as well in regards to longevity.

1.3 DVD-R as an Archival Format

When the DVD-R appeared in 1997, the CD-R had become a viable and extremely convenient archival storage medium. It was only natural that, due to the significant increase of storage capacity, the DVD-R would quickly begin to replace the CD-R as the preferred archival storage medium. Even though the DVD-R was a very young technology compared to the CD-R, there seemed to be an inherent trust that DVD-Rs were just as reliable. However, in 2004, a study performed by NIST caused great alarm among those who had been using the DVD-R as their primary archival storage format. The test showed drastic stability differentiations between the various types of DVD-Rs tested. Some types performed well while others performed very poorly. It was suggested that the performance variations were caused by differences in the manufacturing processes, with particular emphasis on the organic dye layer. However, the lack of dependable methods for consumers to determine the quality of dye that might have been used in a particular sample, meant there was no real way to be sure which brand of DVD-R media could be reliably trusted for archival purposes (Slattery et al. 2004). This presented two major concerns in the archival community: how long will my current DVD-R archive last and which DVD-R media can I trust to use in the future?

In summary, the benefits gained from the larger capacity of a DVD-R are obvious. Nevertheless, there is currently extreme reluctance to use DVD-R’s for future digital archives as well as justifiable concern that existing DVD archives are at risk. Before the DVD-R can be trusted again as a valid archival medium, steps must be taken to ensure that data loss can be prevented. If the manufacturer-provided life expectancy for a batch of DVD-R media is unknown or not trusted, then the consumer needs their own method for determining when their media will fail. In other words, methods that allow consumers to monitor the degradation of real data in a DVD-R archive over time, for the purpose of predicting media failure, need to be devised. If periodic monitoring of the read- back errors on the media can reliably provide the needed data to predict failure, then prior knowledge of its life expectancy becomes less significant. Consumers will be able to make educated decisions about when to move their data to newer media mitigating the risk of data loss.

2 Review of Literature

2.1 Physical Description of a DVD

ECMA International, formally known as the European Computer Manufacturers Association, describes themselves on their webpage as an organization "dedicated to the standardization of Information and Communication Technology (ICT) and Consumer Electronics (CE)". In 1998 they published a paper, ECMA-279, which defined the international standard for the recordable DVD (DVD-R). In 2004 they updated this standard and released it as ECMA-359. This ECMA standard "specifies the mechanical, physical and optical characteristics of…[the] DVD Recordable disk to enable the interchange of such disks" (ECMA International 2004).

2.1.1 **Physical Attributes**

A one sided DVD-R disc, as specified by ECMA-359, consists of four functional layers of material: a substrate, a recording layer, an adhesive or bonding layer and a dummy substrate (see fig. 2.1). Their shape is circular and they are typically manufactured with either an 80 mm or 120 mm diameter. The substrate layers are made from plastics known as polycarbonates. The recording layer, where the data is stored, is made from various layers of materials.

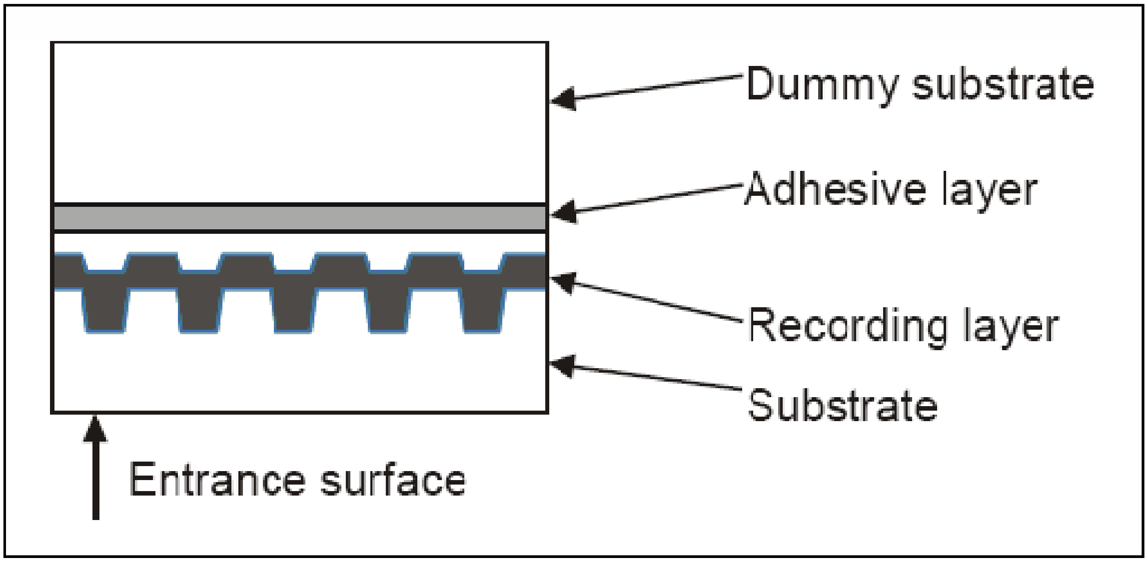


Figure 2.1 – The layers of a DVD-R. Source: Figure from ECMA 2004, adapted from Figure 1.

The recording layer is typically constructed of two distinct materials: a reflective metal and a specialized dye (see figure 2.2). The dye reacts to heat from the write laser to create "burnt" spots, known as *marks*. The portions left unmodified are known as *spaces*. This process alters the overall reflectivity of these areas on the recording layer. When the read laser passes back over the same area, it is able to use these differences in reflectivity to represent the binary data on the track. Figure 2.3 shows a magnified view of a CD-R disc that has been partially recorded. The left half of the image shows a section with recorded data while the right half shows the un-recorded section. The DVD-R uses a technique for recording data similar to the CD-R.

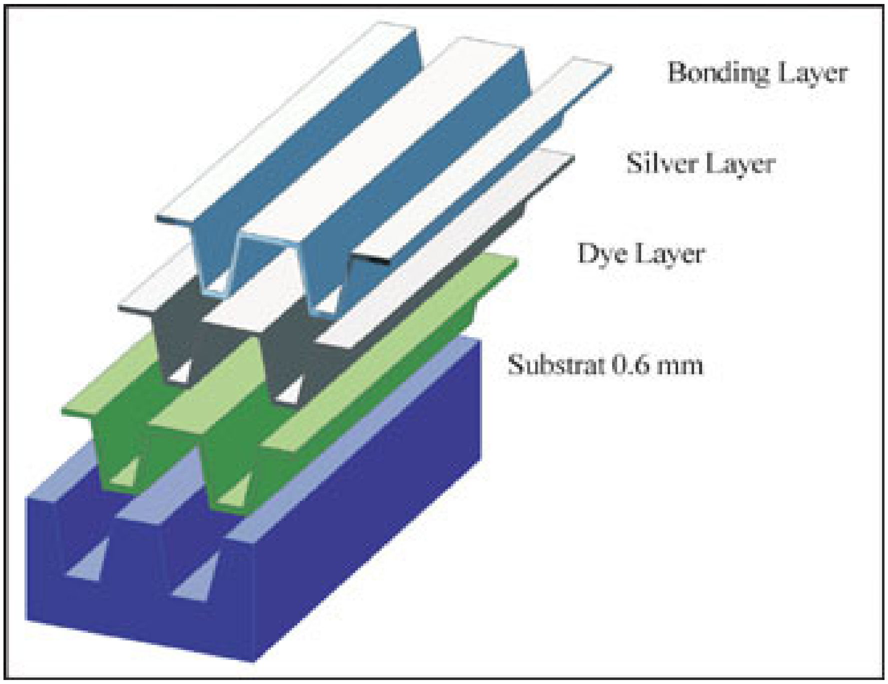


Figure 2.2 – A 3-D view of both the metallic and dye layers. Source: Worthington 2005a, slide 7.

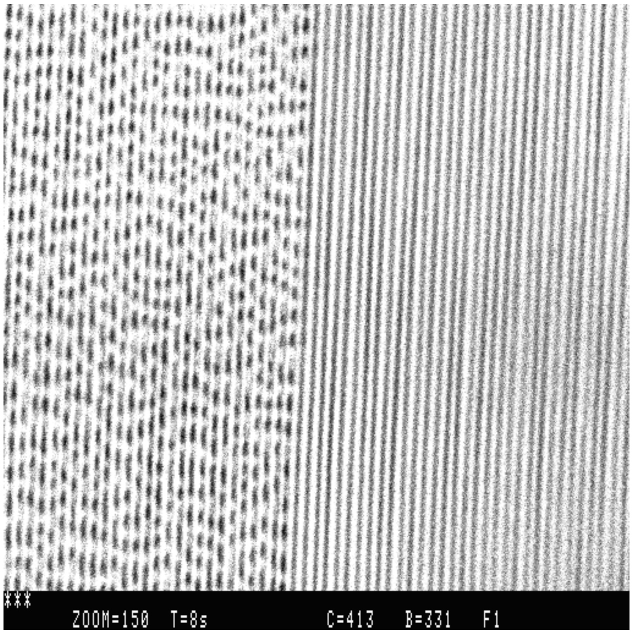


Figure 2.3 - Magnified view of CD-R taken by confocal microscope. Source: Lunt 2004, used with permission.

According to the *DVD Technical Guide* on the Pioneer Corporation website, the recording layer contains a single recording track which spirals out from the center of the disc, similar to a vinyl record. As with the vinyl record, the DVD-R track is a physical feature. As the DVD spins, the laser is guided inside the track and reads or writes data as the surface of the DVD-R moves beneath it. Figure 2.4 shows a graphical representation of these tracks, or grooves. As seen in the figure, these tracks have an intentionally serpentine shape to them. As the laser moves though the track, it senses this side to side motion. This motion is called a *wobble* and is designed to oscillate at a fixed frequency. The *wobbled groove* provides tracking and speed information as the laser moves through the track. The raised area between two neighboring tracks is called a *land*. The lands contain additional physical features called *land pre-pits.* These *land pre-pits* provide additional addressing information.

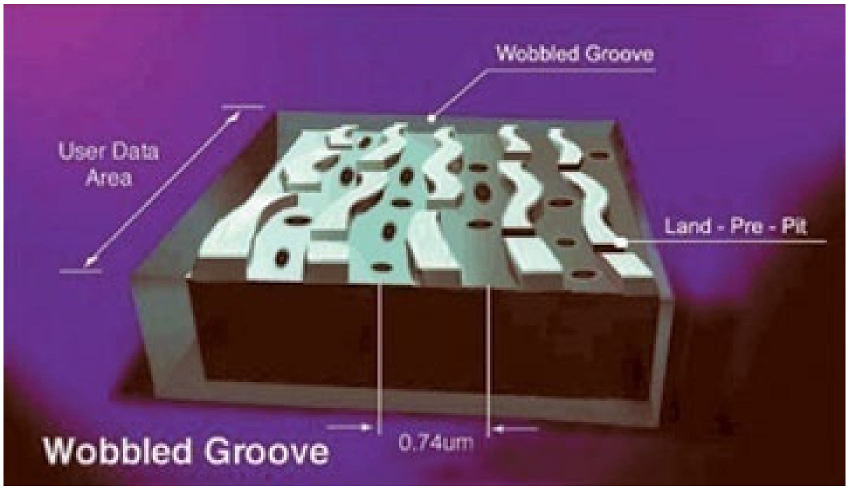


Figure 2.4 – A computer image of DVD-R track features. Source: Worthington 2005a, slide 6.

2.2 Logical Description of a DVD-R

As with all computer storage formats, the DVD-R data is stored in a fashion that can represent binary data. The application is quite simple; as the laser is reading the disc, it senses the transition as it passes from a mark to a space and vice versa. The following quote describes how these transitions are interpreted in a CD-ROM, where the marks and lands are actually physical differences in height know as pits and lands. “In a compact disc, every transition from pit to land is interpreted as ‘1.’ No transition means ‘0,’ and the length of each…segment represents the number of ‘0s’ in the data stream” (Khurshudov 2001). The same is true for the DVD-R. Figure 2.5 is a representation of multiple tracks containing marks and spaces with their binary values shown below each track.

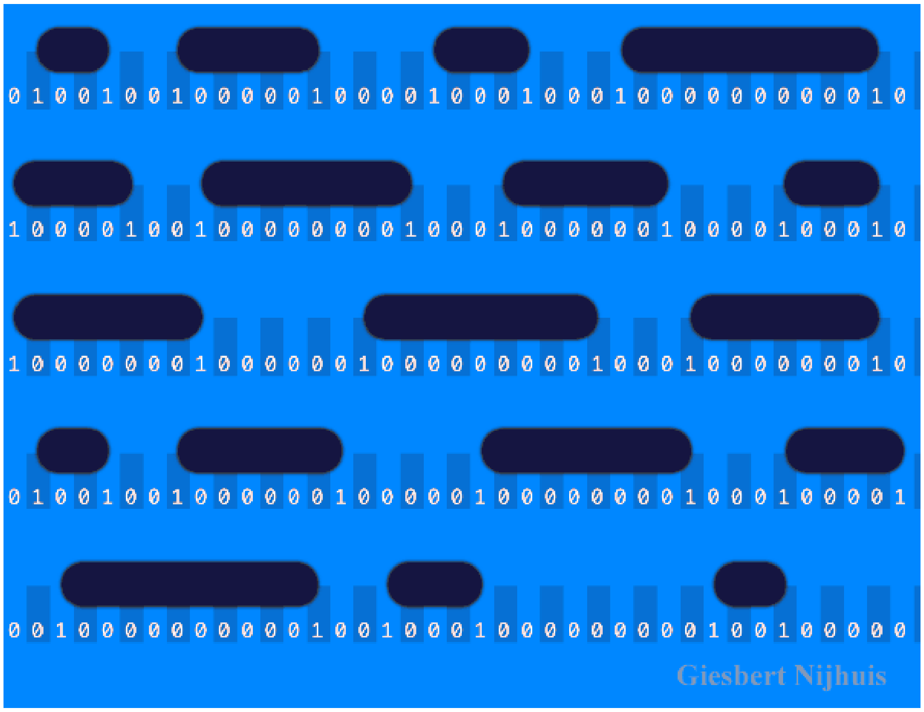


Figure 2.5 - Pit and lands and their binary equivalents. Source: adapted from Nijhuis 2007.

As with any digital storage or transmission system, some data will be corrupted as it travels from point A to point B. In a DVD, this corruption can be introduced during one or all of the three main stages of its life: during recording, during storage and during read back. The causes for data corruption during recording and read back can be very similar and are often the result of fluctuations in the system's electronics due to any number of reasons. During storage, the cause for data corruption is very different; it is due to chemical or mechanical changes which can be accelerated by environmental factors. Whatever the reason for corruption, data loss is not acceptable and steps should be taken to ensure that when, not if, data loss does occur, it can be rectified. This process of identifying when there is corrupted data and then calculating how to fix that data is known as error detection and correction (EDAC). While the entire breadth of EDAC lies outside the scope of this paper, understanding the basics of EDAC is fundamental to understanding the logical structure of a DVD and how the binary bits are ordered within the disc's data track.

2.2.1 **Basic Error Detection and Correction**

When transmitting any digital signal, "things like noise, power-line fluctuations, and imperfections in the media" can cause data to "be corrupted before it reaches the receiver" (Meyer 1990a, 42). Meyer continues:

> This is why error detection coding and error correction coding (EDC/ECC) techniques are incorporated in the digital transmission system. The data on the media is encoded at the transmitting end with the EDC/ECC bits and decoded at the receiving ends to correct for errors and to recover the original data. If there are errors, the EDC/ECC algorithms are designed to detect and correct a certain number of errors or to cause retransmission of data.

In short, EDC/ECC consists of adding redundant data, carefully calculated using the data, to the original digital stream. These additional redundant bits can be used not only to determine if data has been corrupted but also to verify which specific bits need to be corrected.

The simplest form of error detection is the parity bit. A basic parity bit can be calculated by performing an "exclusive or" (XOR) on all the information bits in a predefined set of bits. The result of the XOR is the parity bit. When the information bits are read back, the same XOR function is performed and the result is compared to the parity bit. If any single bit has changed, the parity bits will not match. This parity bit works well if the probability of two bits being incorrect is low. As shown in figure 2.6, if two bits are wrong then the errors go undetected.

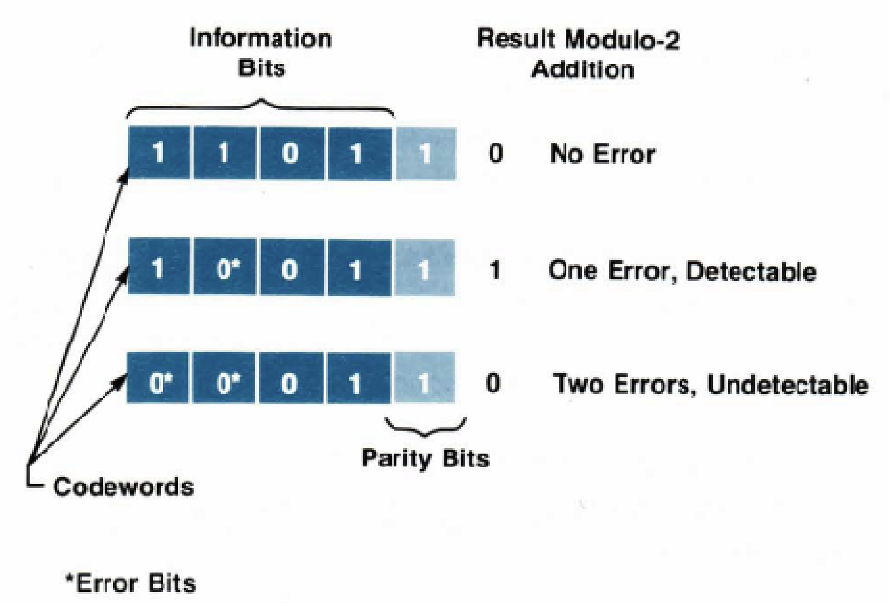


Figure 2.6 - Simple parity example. Source: Meyer 1990b. figure 1.

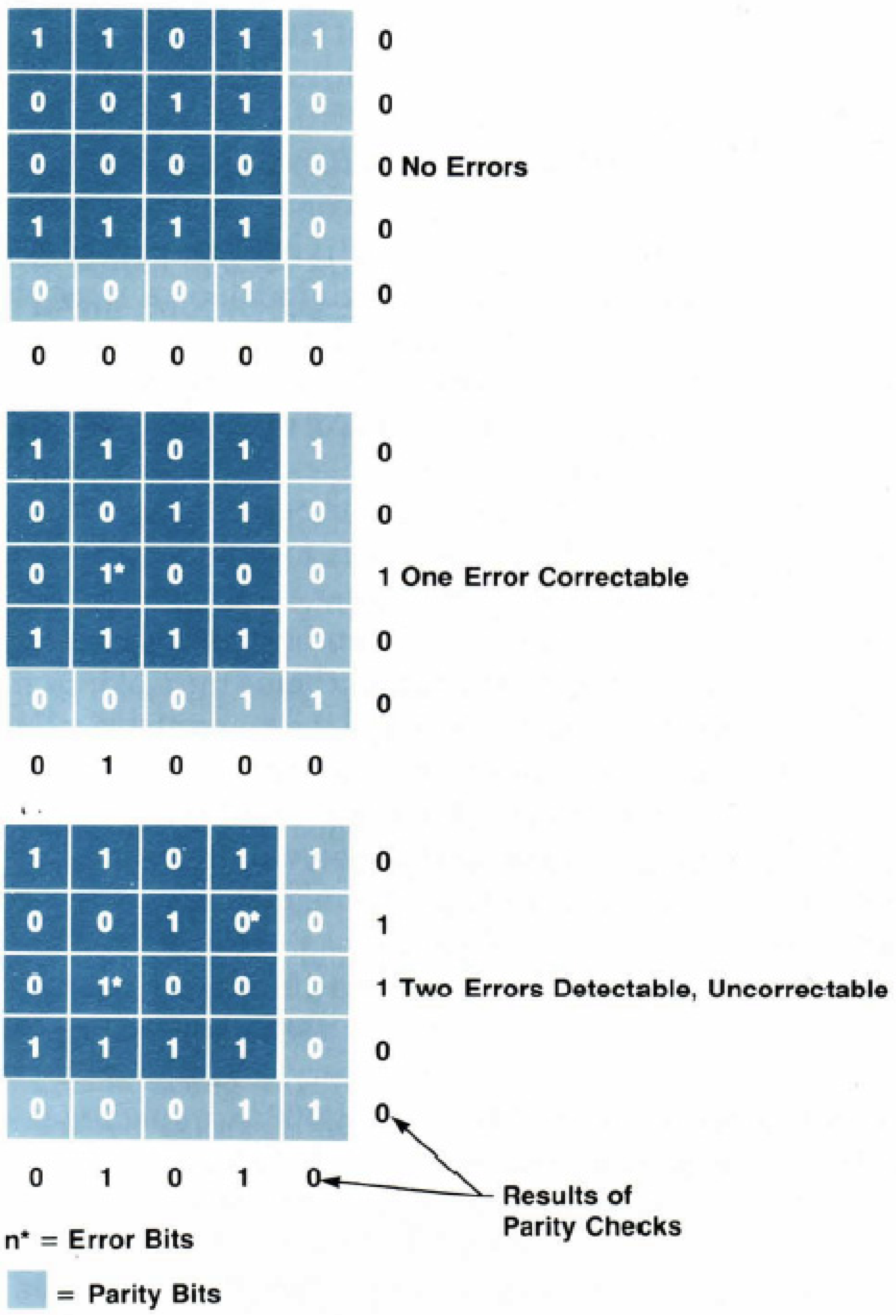


Figure 2.7 - Product code parity example. Source: Meyer 1990b, figure 2.

This example can be extended to show the error correction capabilities of using redundant parity bits. Looking at the information bits as an array of data instead of a single line, a parity bit can be calculated on both the row and the column of the array (see figure 2.7). With this configuration if an error is detected, the parity bit in both the end of the row and the column will not match. This will give the coordinate of the bit in error and it can be corrected by simply switching its current state.

The idea of using redundancy to perform error detection and correction was originally presented in a technical paper by C.E. Shannon in 1948. While his paper proved the importance of a well-designed redundancy algorithm, no real code was actually presented (Comtech AHA Corporation, 1). About this same time, the first practical EDC/ECC was introduced. It was called the Hamming Code. Instead of calculating a single parity bit for each row and column, it calculated multiple parity bits per row by performing an "exclusive or" operation (represented later as the symbol $\oplus$) on various combinations of the data bits. For example, a (7,4) Hamming code has 4 data bits (A,B,C and D) and 3 check bits (E, F and G) for a total of 7 bits per row. The 3 check bits are determined using the following calculations: $E = A \oplus B \oplus C$, $F = A \oplus B \oplus D$, and $G = A \oplus C \oplus D$. Since there are only 4 data bits, there can only be 16 valid combinations of the data resulting in 16 possible 7-bit words, for this particular Hamming code. Keeping in mind that a 7-bit word has 128 possible combinations, if a word is detected that does not match one of the known 16, an error is identified. The original "exclusive or" equations are calculated again on the word in error. If a bit is bad, some of the equations will not match the expected outcome and the equation will return a false response. By looking at the resulting combination of which equations return true and which ones return false, the bit in error can be determined (Comtech AHA Corporation, 2-3).

The Hamming Code belongs to the family of ECC known as *block codes*. "Block codes process the information on a block by block basis, treating each block of information bits independently from others" (Morelos-Zaragoza 2002, 3). Block codes are quite useful in digital transmission systems because they are good at correcting errors caused by channel noise (random errors) and can also handle the types of errors typically caused by defects in the storage media, known as burst errors (Comtech AHA Corporation, 5).

Since Shannon's 1948 paper, designing of EDC/ECC has evolved into a branch of science and mathematics that stands on its own. Mathematically the codes have become extremely complex and intricate as they attempt to improve robustness, efficiency and accuracy. One such group of codes is the Reed-Solomon codes. Another group in the block code family, they were introduced in 1960 by Irving Reed and Gustave Solomon. Variations of this code have been used for NASA space communications, cable modem communications, Terrestrial Digital HDTV transmission and many other applications (Morelos-Zaragoza 2002, 61). According to the Comtech AHA Corporation, an implementer of data encoding technologies, Reed Solomon codes "provide powerful correction, have high channel efficiency, and are very versatile". They further attribute the code's popularity "to standards compliance and economic implementations" (Comtech AHA Corporation, 1). These codes also happen to be the EDC/ECC algorithms of choice for the DVD formats (ECMA International 2001, 67).

Reed Solomon codes are similar to Hamming codes in that they use multiple combinations of parity bits for its EDC/ECC. The main difference is that Reed Solomon uses parity bytes instead of bits. This gives it the ability to not only detect and correct errors at the bit level but also at the byte level. Another significant way that Reed Solomon differs from the above Hamming code example, at least in its implementation in the DVD, is that parity bytes are not only calculated on the rows (Parity Inner) but also on the columns (Parity Outer), adding even more redundancy. The following sections, describing the DVD-R's logical data format, give a basic summary of the data and parity structure as defined by Section 4 of the ECMA 359 Standard.

2.2.2 **DVD-R Data Format**

EDC/ECC works best when the corrupted data is isolated. In other words, if the bits in error are few and are surrounded by numerous healthy bits, the EDC/ECC has a much better chance of accurately correcting the data. Because of this, EDC/ECC by itself works well for the occasional read/write error, known as random errors.

On the other hand, when large amounts of consecutive data are in error (an error burst) due to any number of potential external factors, the EDC/ECC is more likely to fail. Many EDC/ECC implementations rely on scrambling and/or interleaving the data to help minimize the negative effects of "burst" errors during transmission and storage. When data is scrambled before a transmission and a burst error occurs, a large section of bad data is left in the middle of the transmission. Unscrambling the data on the other end will distribute the bad data and turn the one large bad section in the middle into several smaller bad sections spread throughout the data. In other words, burst errors can be transformed into what would more closely resemble random error. This can give the EDC/ECC a better chance of succeeding.

In the case of optical media, these types of errors can arise while the data is stored on the disc and physical corruption of that disc occurs. The three most common causes of this damage are: environmental factors (i.e., heat and humidity), manufacturing flaws and physical damage from misuse. The specification for the DVD-R format requires multiple levels of EDC/ECC and interleaving in order to minimize the effects of burst and random errors. The following sections will provide a brief overview of the data format. For complete details of the DVD-R data format, Section 4 of ECMA standard 359 can be found in Appendix A.

2.2.2.1 **Data Frames**

To prepare user data (referred to as main data) for storage on a DVD-R, it is first broken down into sections of data 2048 bytes in length. Each section is further divided into twelve rows. The first row contains 160 bytes of main data and is pre-pended with three additional fields of data: the ID field (Information Data, 4 bytes), the IED field (ID Error Detection Code, 2 bytes) and an RSV field (Reserved for application use, 6 bytes). Following this first row are ten more rows each comprised entirely of 172 bytes of main data. The twelfth and final row contains 168 bytes of main data and 4 bytes of EDC data. The EDC field contains check bytes that were calculated over the preceding 2060 bytes (see figure 2.8). This entire block is known as the “data frame”.

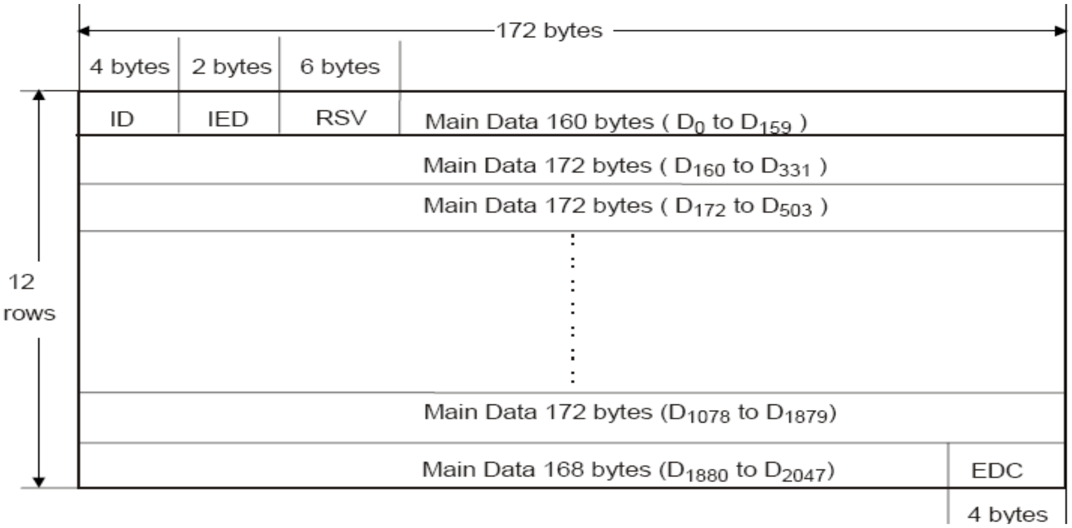


Figure 2.8 - DVD-R Data Frame. Source: Standard ECMA-359, figure 23.

It is in the data frame where we first see scrambling occur. Before the data frame is assembled, the 2048 bytes of main data are first scrambled. The exact method for scrambling this data can be found in Appendix A.

2.2.2.2 **The ECC Block**

Once the data frame is assembled, it is then used to piece together the next logical chunk of code. The "ECC Block" is started by stacking sixteen data frames on top of each other resulting in 192 rows of 172 bytes each. Parity bytes are then calculated on each of the rows and columns. The equations for calculating the parity bytes can be found in Appendix A. Parity of Inner Code (PI) bytes are added to the rows and Parity of Outer Code (PO) bytes are added to the columns. In all, 10 PI bytes are added to each row and 16 PO bytes are added to each column (see figure 2.9).

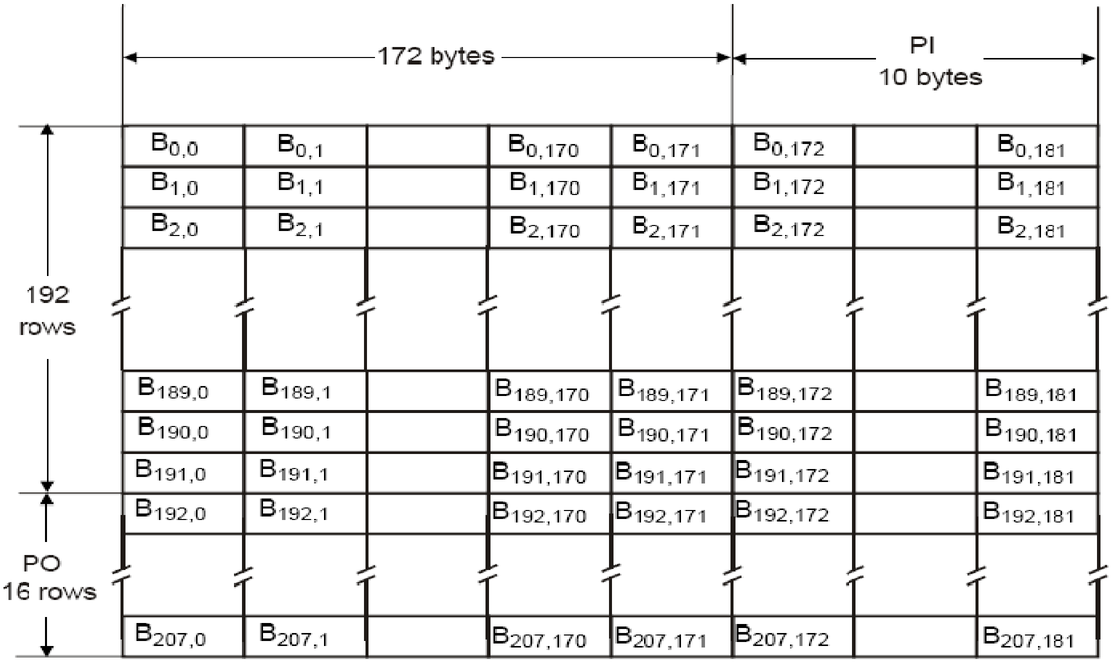


Figure 2.9 - DVD-R ECC Block. Source: Standard ECMA-359 figure 27.

2.2.2.3 **The Recording Frame**

From one ECC block, sixteen "Recording Frames" are created. This is accomplished by interleaving the sixteen rows of PO bytes back into the preceding rows of data. In short, a recording frame is made from twelve rows of data from the ECC block and one row of PO data (see figure 2.10).

Once the recording frame is complete, no more redundant EDC/ECC bytes will be added nor does the data go through any further scrambling. It does, however, go through a few more steps to convert it to the final signal that will be stored on the disc. These remaining steps are once again detailed in Appendix A.

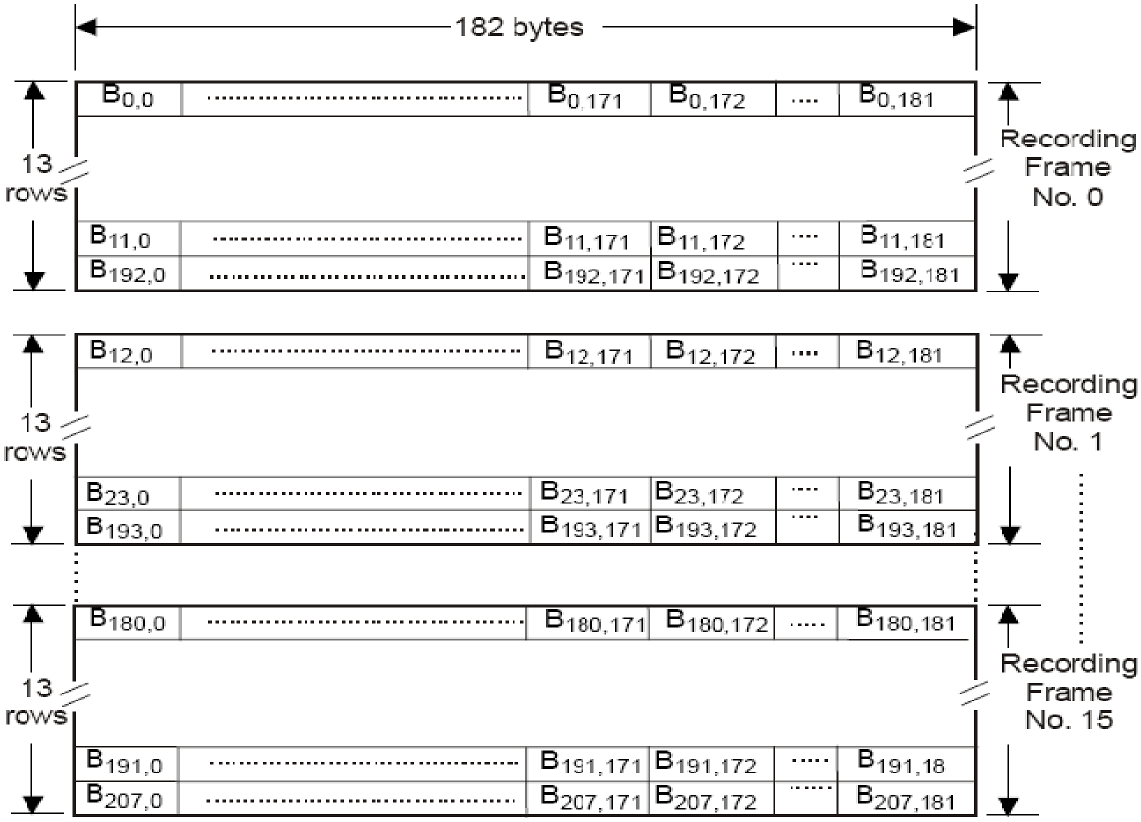


Figure 2.10 - DVD-R Recording Frame. Source: Standard ECMA- 359, figure 28.

2.3 The DVD-R's Life Expectancy

As materials age, their physical structures typically begin to break down. As their composition changes over time so do their physical properties. Plastics that were once resilient and soft may become hard and brittle. A metal that once had a shiny finish may become dull and corroded. When an object is manufactured from components that age, that object is at risk of becoming less useful over time and in fact, may lose its function altogether. Because of this, many manufacturers spend a great deal of time and money developing tests to determine the overall useful life expectancy of their products. One very common technique is called accelerated lifetime testing. In order to perform accelerated lifetime testing on a material, one must first discover which external stressors are the primary factors for the degradation of the test material. Once those stressors are determined, they can be magnified in a controlled environment causing the test material to degrade at an accelerated rate. The degradation is monitored closely and the results are used to project the average life expectancy of the test material while under normal circumstances.

2.3.1 Environmental Stressors Contributing to DVD-R Failure

As mentioned previously, a DVD-R is composed of four different types of unique materials: a polycarbonate, an adhesive, a metallic and an organic dye layer. Each of these materials may have different stressors as well as different rates of decay. In the case of the DVD-R, as with most products constructed from many materials, the failure of any one component can contribute to the overall loss of functionality.

"The polycarbonate substrate makes up most of the disc…Anything in or on the polycarbonate layer that interferes with the ability of the laser to focus on the data layer will result in the misreading of data" (Byers, National Institute of Standards and Technology (U.S.), and Council on Library and Information Resources 2003, 5). The polycarbonate layers can absorb any surrounding moisture including any contaminants the moisture may contain. While the absorbed moisture itself can interfere with the laser focus, it does not necessarily degrade the polycarbonate layer, especially when compared to the degradation seen in the others layers over the same period of time. Furthermore, when no additional contaminants are involved, removing the moisture from the disc will return it to its original functional state (Byers, National Institute of Standards and Technology (U.S.), and Council on Library and Information Resources 2003, 18). Of all the materials in a DVD-R, the polycarbonate layers seem to be the least susceptible to degradation due to environmental factors. They are, however, susceptible to damage from physical misuse—scratches, fingerprints and dirt being a few examples.

Evidence of damage to the adhesive layer has not been reported often. When this occurs the layers themselves may begin to separate; this will likely render the disc useless to a standard disc reader. This type of damage has not been documented occurring to DVD-R discs. But there has been anecdotal evidence of this occurring with pressed DVD-ROM's (Iraci 2005, 142).

The metal reflective layer has historically been an alloy of one of three different types of metals: gold, silver or aluminum. Aluminum is not only the least expensive, but also has the highest reflectivity. It would seem that aluminum would be the ideal choice. However, aluminum does oxidize quickly when exposed to oxygen, a highly abundant element in our atmosphere. In conditions of high humidity, the absorptive nature of the polycarbonate brings the oxygen right into the reflective layer. However, the result of aluminum reacting to oxygen is the creation of a protective and self-limiting oxide layer that covers the surface of the aluminum. Assuming that this oxide layer did not exist previously, the formation of this layer in an existing disc may affect the data, but it seems unlikely since it is thin enough to be optically transparent (typically about 25 Å). Gold is the most expensive and least reflective but is often valued because it does not corrode or oxidize. This leaves silver. It is cheaper than gold and is more reflective. However, it is at risk of corrosion and oxidation. It should be noted that silver does not oxidize easily but it does corrode through exposure to any sulfur compounds that might be in the air (i.e. sulfur dioxide). Nevertheless, steps can be taken to control the air quality to ensure that no traces of sulfur can be found in the air where the discs are being stored. As a further deterrent, “manufacturers use various silver alloys to help inhibit silver corrosion, and most [recordable] discs available today use a silver alloy reflective layer” (Byers, National Institute of Standards and Technology (U.S.), and Council on Library and Information Resources 2003, 14).

The last material used is an organic dye layer that coats the metal layer. Because of its organic nature, the dye has a natural tendency to degrade over time. Fred Byers of NIST states:

> High temperatures and humidity will accelerate the process. Prolonged exposure to UV light can degrade the dye properties and eventually make the data unreadable. Heat buildup within the disc, caused by sunlight or close proximity to heated light sources, will also accelerate dye degradation.

The organic dyes are combined with stabilizers that serve to improve their performance in the recording process. These stabilizers also attempt to prolong the lifetime of the dye. Even with these stabilizers, the organic dye layer still seems to be most susceptible to accelerated degradation due to fluctuating, or extreme environmental conditions. As mentioned previously, the most commonly used dyes are cyanine and azo mixed with various stabilizers.

In summary, there are three main environmental contributors to the degradation of a DVD-R disc over time: heat, humidity and exposure to UV light. The amount of heat and humidity will directly affect the organic dye’s rate of degradation as well as the metal layer’s potential for corrosion and/or oxidation. It is important to remember that the organic dyes were developed to have a specific chemical reaction when exposed to the “burning” laser. As such, any heat as well as the presence of light can also act to accelerate this chemical reaction. Prolonged exposure to both heat and light can lessen the distinction between the marks and lands representing the 0’s and 1’s on the disc, making it more difficult to read the data.

2.3.2 **Metrics for Determining the Quality of the Data on the Disc**

There are two main metrics for monitoring the quality of the data recorded on a DVD-R: ECC errors and jitter. When these measurements exceed certain values—which are discussed later—data is lost.

2.3.2.1 **Description of ECC Errors**

As mentioned in section 2.2.2.2, each ECC Block consists of 208 rows of data. Each of these rows contains inner parity data. If any row in an ECC block contains at least one error, a *Parity Inner Error* will occur. “If more than 5 (or 5 consecutive) bytes are in error in an ECC Block row we have a *PI Failure* (PIF)” (Worthington, 2005b. *emphasis added*). A PIF is still a correctable error. Its presence simply signifies that the error correction is transitioning from using the inner parity bits to using the outer parity bits. The outer parity bits are capable of correcting 8 bytes per column over the 172 columns per ECC block. If the errors exceed this, a *Parity Outer Error* is generated and the ECC block is considered uncorrectable. With an uncorrectable ECC block, data loss is usually the result. “However, most DVD players will expand their data matrix and retry an uncorrectable ECC before tagging it as a lost piece of information” (Worthington, 2005b).

The DVD-R specification uses the total number of PI errors across eight consecutive ECC blocks to determine the quality of the data on the disc. The *Parity Inner Sum Eight Error* (8PIE) count is calculated by taking the total number of PI errors in a block and summing it with the total number of PI errors in the previous seven ECC blocks. According the DVD-R spec, this number must not exceed 280 (ECMA International 2004, 22).

2.3.2.2 **Description of Jitter**

Another measurement used to analyze the quality of the data on a disc is *jitter*. As seen in section 2.2, the transitions between the marks and the spaces on the disc represent different patterns of data. The various lengths between these transitions are known as *symbols* and have a distinct and predictable duration. In a white paper from the company AudioDev, Ulf Wilhelmsson gives an excellent definition for these symbols:

> The digital information lies in how the lengths between the transitions are chosen. The transitions are also used to regenerate the bit clock with the clock period, T. By using this bit clock it is possible for the player to extract the data from the disc by identifying the discrete lengths of 3T, 4T up to 11T and 14T, called Symbols. These are the finest structures of the data on the disc (Wilhelmsson 2004, 3).

As the read laser runs over the marks and spaces, the differences between them are manifested by variations in intensity of the reflected light. These variations in intensity are read as an analog signal that fluctuates up and down and is called the HF-signal. The digital data is embedded in this analog signal. A circuit called a slicer, takes the HF-signal and calculates the transition points from the phase changes. Using these transition points, the slicer transforms the analog signal into a digital signal called the EFM, or eight to fourteen modulation. “From the resulting data stream (EFM), the bit-clock (P-Clock) is regenerated (Wilhelmsson 2004, 4). The P-Clock is then used to help identify the different symbols (see figure 2.11).

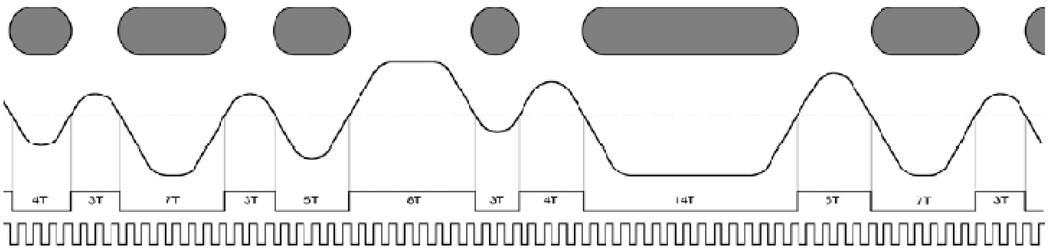


Figure 2.11 - Sequence of Marks, HF, EFM and P-Clock with the identified symbol types. Source: Wilhelmsson 2004, figure 2.

In order for the data to be readable, the beginning and end of each symbol (known as the transition points) must fall at very specific times in relation to the bit clock. Otherwise the reader can be confused about what type of a symbol it is trying to decipher. Variations in the symbol's expected location and size, in relation to the clock, are known as jitter.

There are two main types of jitter: data-to-data (DD) jitter and data-to-clock (DC) jitter. This study monitors DD jitter, which focuses on the readability of the data on the disc. This type of jitter looks specifically at the size (time length) and positions of the symbols in relation to the bit clock. Variations in the expected size are reported as jitter.

2.4 The Eyring Acceleration Model

In order to understand how a specific optical disc may degrade over time in normal ambient conditions, accelerated lifetime testing is often used. These tests speed up the degradation process so it can be observed under a reasonable test time frame. In order to understand degradation and how accelerated lifetime testing works, it is important to remember that the degradation of recordable media is due to chemical reactions that occur over time in the materials that make up the disc. These chemical reactions are the result of the material's natural tendency to break down into a less ordered state.

To accelerate chemical reactions, energy must be added in order for the reaction to occur. This is called the "activation energy". Heat is a source of energy and can often provide the needed energy to sustain a chemical reaction. The ambient heat in the environment is often enough to cause chemical reactions to occur, therefore causing natural degradation. Depending on the amount of heat in the environment, these reactions may occur at a faster or slower rate. The Arrhenius equation defines the relationship between that reaction rate and temperature. It can be written as

$$k = A\, e^{\left(\frac{-E_A}{RT}\right)} \tag{2.1}$$

"where A is the so-called 'pre-exponential factor' (which has the same units as the rate constant), E_A is the activation energy of the reaction (in units of Kj/mol), T is the temperature (in degrees K), and R is the gas constant (in this case, 8.3J/deg mol)" (Hites 2007, 81).

The Arrhenius model is useful in analyzing reactions where temperature is the main contributor to the reaction. Many reactions, however, are the result of multiple stressors. The Arrhenius model fails to accurately describe reactions "when stresses other than temperature are involved" (Tobias and Trindade 1995, 191). There is another model—the Eyring model—that does consider the contributions of additional stressors that affect the rate of a chemical reaction.

Elements of the Arrhenius equation can be seen in the Eyring equation that can be written as

$$t = Ae^{\frac{\Delta H}{kT}}e^{B\times RH} \tag{2.2}$$

where t is the time to failure, A is the pre-exponential time constant, ΔH is the activation energy per molecule, k is Bolzmann's constant (1.3807 × 10^{-23} J/molecule degree), T is the temperature in Kelvin, RH is the percentage of relative humidity and B is the RH exponential constant (ECMA International 2007, 9). The equation can be further simplified as

$$\ln(t) = (\ln A) + \left(\frac{\Delta H}{k} \times \frac{1}{T}\right) + (B \times RH) \tag{2.3}$$

2.4.1 **Applying the Eyring Acceleration Model**

The Eyring Acceleration Model is the most common method used to predict longevity of optical media. Using this model, the discs are subjected to higher than normal temperatures and humidity. The accelerated degradation of the discs under these conditions is observed over time and until the discs have reached their "end-of-life".

There are two published standards in particular which specifically outline the methods for using the Eyring Acceleration model to test the predicted lifetime of recordable optical discs. In 2002, the ISO released a standard for testing CD-R's (ISO 2002). In 2007, ECMA released a similar standard for testing DVD-R's (ECMA

International 2007). In both tests, the discs are divided into separate batches. Each batch is aged at a specific combination of temperature and relative humidity for a pre-determined amount of time. At set intervals, the discs are removed and their read-back errors are analyzed. For DVD-Rs, the end-of-life condition is met when the PIE8 sum error count in any section of the disc exceeds 280. The median time for disc failure is then calculated for each batch of discs. At this point, linear regression analysis can be used in conjunction with the collected data and the Eyring equation to calculate the coefficients that are specific to the test media. The custom Eyring equation can then be used to predict the lifetime of the discs under normal storage conditions (typically 25 degrees Celsius and 50% relative humidity). An example of this process, including simulated data, can be seen in Annex B of the document ECMA 379 (ECMA International 2007, 13).

2.4.2 **Eyring Acceleration Model Testing Assumptions**

In each of the previously mentioned standards, there is a section describing the assumptions that are made in order for the test to be considered valid. One of these assumptions is that the Eyring method itself accurately models the main failure mechanisms in the disc. At first glance, one might wonder why this is an assumption and not an assertion. If it is possible that the Eyring model does not accurately describe the degradation processes, how can we rely on the results we observe when using this model? It is important to remember that a recordable optical disc is constructed from various distinct materials. As discussed earlier, each of these materials has its own unique vulnerabilities to environmental stressors and its own rate of decay.

In the year 2000 David Nikles and John Wiest from the Center for Materials for Information Technology, wrote and presented a paper at the "Seventeenth IEEE Symposium on Mass Storage Systems" entitled *Accelerated Aging Studies and the Prediction of the Archival Lifetime of Optical Disk Media*. In it they share specific concerns regarding the application of current longevity testing procedures "for the complicated materials packages in optical data storage media" (Nikles and Wiest 2000, 31). The following excerpt clearly states just a few of their concerns about the assumptions made when using Eyring type models:

> However, these assumptions may be invalid depending on the underlying physical phenomena governing degradation or if the extrapolation passes through a phase transition, such as a glass transition in a polymeric component. Another problem is that many degradation processes, such as polymer hydrolysis or corrosion can be autocatalytic, i.e. the product of the degradation process can catalyze further degradation. A predictive model of archival lifetime must be based on an understanding of the chemical and physical processes leading to failure (Nikles and Wiest 2000, 33).

In a system as complex as a DVD-R, it can be quite difficult to identify all of the potential reactions that may occur as the environmental conditions change.

While it is important to know that these concerns exist, it is equally important to know that they may not matter. The temperature differences in these standardized tests most likely are not extreme enough for these concerns to present themselves. However the testing standards do take into account the possibility that the degradation process of some test discs may not be appropriately modeled by the Eyring equation. They require that before the Eyring analysis is performed, the data itself must conform to specific criteria to ensure that the data behaves as expected. The failure times of the discs in each group are normalized by ordering them by median rank. "The median rank of the specimens is calculated using the estimate (i - 0.5)/n, where i is the time-to-failure order and n is the total number of specimens at the stress condition" (ECMA International 2007, 15). This data is plotted against the actual failure times and a best-fit line is drawn through each group of data. If these best-fit lines are reasonably parallel to each other, then we can assume that the reaction rates at each stress level are equivalent. Thus it is reasonable to use the Eyring equation to model the life expectancy of the discs in question. If the lines are not parallel then the assumptions set forth in the standard may not hold true (ECMA International 2007, 19).

2.5 Recent Longevity Studies

In late 2004, the National Institute of Standards and Technology (NIST) performed their own study on optical media that differed slightly from previous studies. They were not trying to calculate a life expectancy for the discs in question. In fact due to the limited sample size, no acceleration models were performed at all to predict real world longevity. According to their research, "the results from these tests are to demonstrate, in terms of error rates, the ability of some DVD and CD media to maintain stability given these extreme conditions" (Slattery et al. 2004, 520). In addition to heat and humidity, they were also interested in studying the effects of light exposure in the degradation process. The results of their study showed some very interesting and useful facts about CD-Rs. For the DVD-R, however, the main consequence of their study was increased concern and doubt about its viability as an archival storage format. It was not because they proved that the DVD-Rs could not withstand the tests. It was because they showed that some performed very well while others performed very poorly (figure 2.12). Although it could not be confirmed, they believed that the DVD-Rs tested were manufactured using a stabilized cyanine dye. They explained that the differences in the results were based on what appeared "to be varying proprietary modifications made to the dye formulations, and perhaps different manufacturing processes and quality control procedures" (Slattery et al. 2004, 522). Statements like these are what seemed to cause the most concern about the use of the DVD-R as an archival medium. Consumers really had no way of knowing the quality of disc they might be buying.

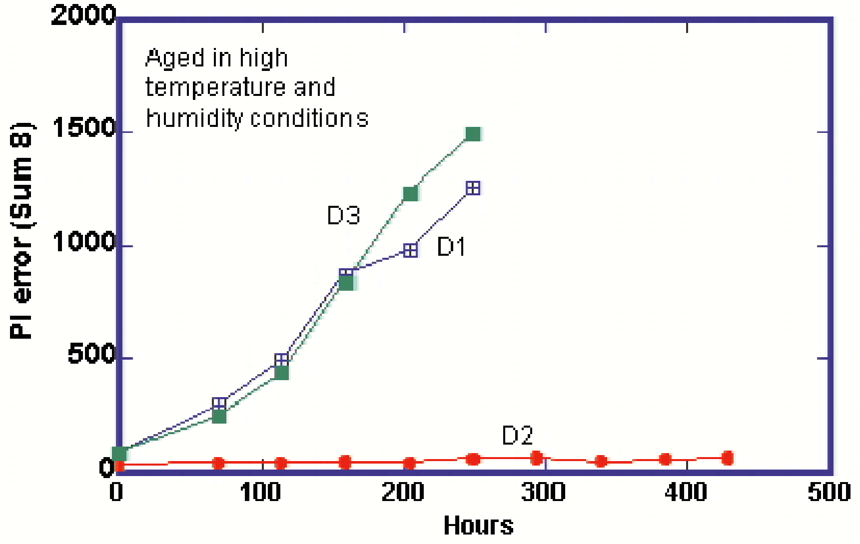


Figure 2.12 – PI (Sum 8) increase in DVD-R when exposed to extreme temperature and humidity. Source: NIST 2004, figure 5.

A year later another study was performed that also decided to forgo the traditional longevity studies in favor of relative comparisons. Joe Iraci, the author of this later study, collected a large variety of optical discs (Iraci 2005). Among them were forty DVD-R discs from ten manufacturers. The discs were aged at 85 degrees Celsius and 85% relative humidity for four 21 day periods. After the first period, 50% of the DVD-Rs had failed. At the end of the test, 92% of the discs had failed. This information alone doesn't tell us much. But when compared to the failure rate that was observed for CD-Rs made with phthalocyanine dyes, a distinct difference can be seen. Only 28% of the CD-R's had failed in the same period. For archivists, this study just confirmed that it might have been a bad idea to move over to DVD-R media.

In February of 2007, a paper was released in the journal *IEEE Transactions on Magnetics* that was written by M. Irie and Y. Okino. It was "intended to establish a standard technique to measure high-speed, high density optical disk life expectancy" (Irie and Okino 2007, 864). This paper showed that their DVD-R life estimation method—which seemed to be based on the methods set forth in the earlier ISO Standards for the CD-R and CD-ROM—was able to produce a lognormal distribution graph of failure times using PI error data. In other words, the Eyring Acceleration Model, in conjunction with proper statistical analysis, seemed to produce the type of results one would expect to see if the assumptions set forth in the model hold true. They also concluded that the expected lifetime of the media tested was about sixty seven years (Irie and Okino 2007, 866).

In the paper published from the 2004 NIST study, the organization made it clear that they intended to develop both a standard for determining the life expectancy of recordable DVD's as well as a grading mechanism that manufacturers would use to rate their media based on the results of the standardized testing. They asserted that "a comprehensive study [was] underway in a collaboration between NIST and the Library of Congress (LoC)" to create these tools (Slattery et al. 2004, 524). The first part of this was realized in 2007 with the publishing of a standard for testing DVD-R discs. This standard, which has been previously referenced in this paper, is known as ECMA 379.

2.6 The Need for Additional Test Methodologies

While accelerated lifetime testing should prove very beneficial for consumers in the future as manufacturers begin to adopt this standard form of testing, it provides little use to those with current DVD-R archives. Accelerated lifetime testing takes time and requires expensive equipment. Furthermore, it is a destructive process.

For ten years, consumers have been using potentially un-reliable media to store their digital files. Today there exist enormous collections of data stored on recordable DVD’s throughout the world. These collections continue to grow despite the fact that some media purchased today may not last into the next decade. It is unknown how long it will be before a standardized grading method is made available which will allow consumers to know how long the media they purchase tomorrow will survive.

In the meantime, tests that allow consumers to predict the end-of-life for existing archives of recordable DVD’s would be very beneficial. If a predictive correlation can be observed between the increase of certain read-back errors over time and a media’s end-of-life, this knowledge could be used by consumers to reduce the potential for data loss in their current archives. By periodically analyzing the increased read-back errors over time on media of unknown quality, an estimate of remaining life could be calculated. This would allow sufficient time for the consumer to move data to new discs in the case of impending failure.

It is the intention of the research presented in this paper to investigate the feasibility of such tests. If this research shows that specific patterns in the read-back errors—which present themselves over time—can predict data failure with reasonable accuracy, then a new method for archivists to circumvent data loss may be possible.

3 Test Methodology

The experiments conducted on optical discs to this point, have focused on two major objectives: the discovery of a standard method for determining the expected lifetime of a specific batch of media; and the comparison of relative stability for various types of currently available optical discs. While a standardized methodology for determining media longevity will indirectly benefit the consumer, it must first be adopted by the manufacturers. The studies on relative stability proved to be very interesting and helped to identify concerns in the current marketplace. They did not, however, provide any practical knowledge that consumers could use to protect their current and future DVD-R collections.

This study differed from previous studies in that it focused on providing a practical test methodology that could be useful directly to the consumer. This study focused on investigating the use of read-back errors, which are relatively inexpensive to collect, to predict the expected lifetime of discs already in use.

3.1 Experimental Setup

Briefly stated, the test consisted of exposing eighteen DVD-Rs to elevated heat and humidity to produce rapid aging. The discs were removed periodically to analyze any read-back errors that were present. The test was complete once all of the discs were classified as “un-usable’. A more precise description of the term un-usable, as defined by this study, is covered in depth at the end of section 3.1.2.

3.1.1 **Preparation of the Media**

For this test, it was assumed that all eighteen test discs were identical. It is asserted that since eighteen consecutive discs from the same spindle were used, they were considered virtually identical. This is based on the assumption that these discs were manufactured at roughly the same moment in time, on the same machinery under identical environmental circumstances. The discs in this study were rated for burning at 8X speed and were manufactured by a company with a positive reputation in the archival industry. The recordable side of the discs had a dark blue, almost bluish brown, color. Based on common DVD-R manufacturing processes we believe the discs had an aluminum or aluminum alloy reflective layer as well as an azo-based dye layer.

In preparation for the creation of the discs, a file-based DVD image was created in the form of an ISO file. The image was designed to fill each DVD-R to capacity with files ranging from three to seven megabytes. To attempt to isolate potential effects of varied writing strategies, the discs were divided into three groups of six each. Each group of discs represented a different write strategy. The image was burned onto the discs at 6X speed, 8X speed and MAX speed. It seems important to note that while our burner has a max speed of 16X, the actual observed speed for discs burned at full speed rarely exceeded 12X. They were then labeled with both an identifying number as well the speed at which they were burned, in the area known as the Clamping Zone of the disc (figure 3.1).

Figure 3.1 – Disc number nine, burned at 8X speed.

Figure 3.2 – BHD-203 test chamber from Associated Environmental Systems

3.1.2 Accelerated Aging Parameters and Equipment Set-up

The discs were exposed to elevated heat and humidity conditions using an environmental chamber manufactured by Associated Environmental Systems, model number BHD-203 (figure 3.2). The DVD's were placed on a custom-built rack designed specifically to house the eighteen discs within the chamber (figure 3.3). The rack was originally built to keep the discs at a very specific angle to allow for even exposure to the surface of the discs from an overhead light source. However, since no overhead light was used in this study and the angles on the rack were adjustable, the discs were mounted in a completely vertical position. In this configuration, the rack worked well as it allowed adequate separation of the DVD's in the chamber as well as placed them parallel to the air flow within the chamber.

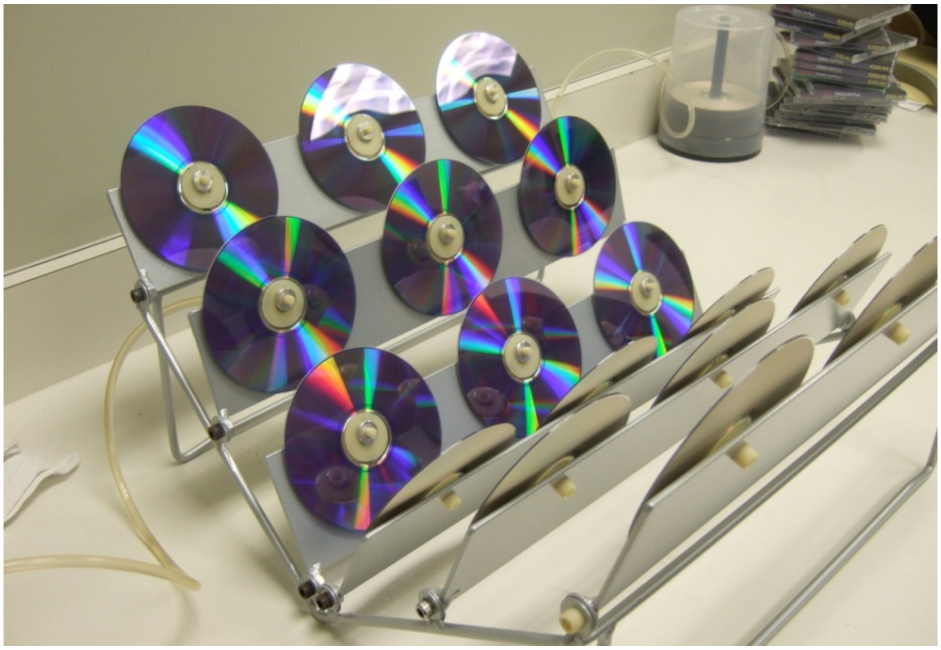


Figure 3.3 – DVD-Rs installed on the test rack. The angles were later adjusted to be completely vertical before placing the rack in the chamber.

The heat and humidity ramping profiles used for this test followed the recommendations outlined in the ECMA 379 specification (ECMA International 2007, 7-8). As seen in figure 3.4, the ramping profiles were divided into six main periods. The first period slowly increased the temperature from the ambient temperature to the test temperature. At the same time, the relative humidity was ramped to a specific intermediate relative humidity (see table 3.1). Once the first ramping period was complete, the relative humidity was then also slowly brought up to the test condition. The discs were soaked at these test conditions for the duration of the incubation period. After the incubation period, the percent relative humidity was lowered back to the intermediate setting and the discs were left in the chamber for a few hours to achieve equilibrium. The last step was to bring the discs back to ambient temperature and relative humidity.

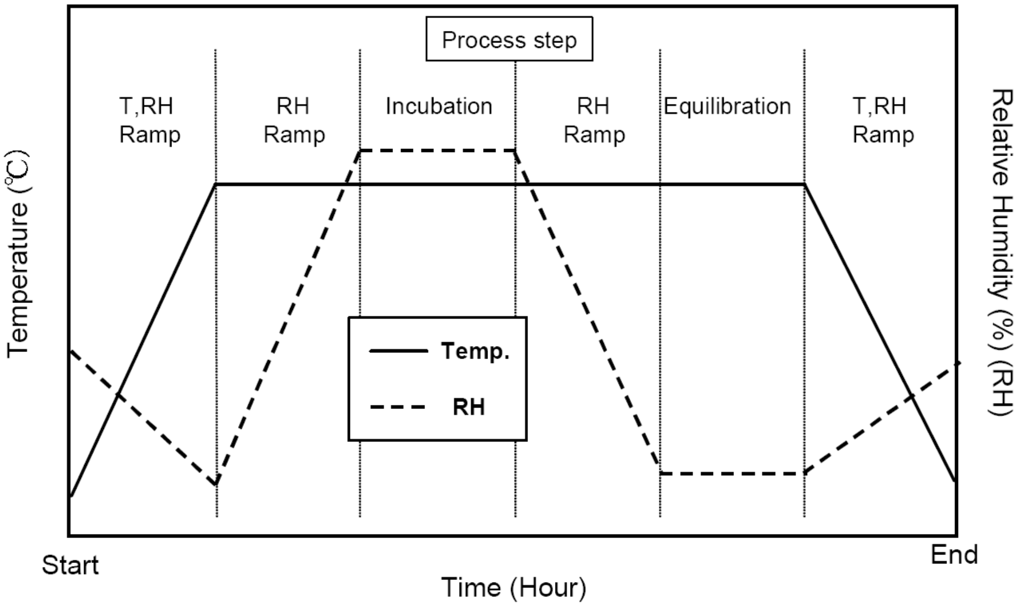


Figure 3.4 - Ramping profile template as specified by the ECMA 379 Document. Source: ECMA 2007, figure 1.

Table 3.1 - Actual conditions and times for ramping profile used in this study

Process Step	Temperature	Relative Humidity	Duration
	°C	%	hours
Start	25	25	—
T, RH ramp	80	30	1
RH ramp	80	80	1
Incubation	80	80	24
RH ramp	80	30	1
Equilibration	80	30	7
T, RH ramp	25	25	1
end	25	25	—

The ECMA 379 standard specified four separate test conditions in order to collect sufficient data for the Eyring calculations: 85 degrees Celsius at 85% relative humidity, 85 degrees Celsius at 70% relative humidity, 65 degrees Celsius at 85% relative humidity, and 65 degrees Celsius at 75% relative humidity. Preliminary testing showed that the 85 degrees Celsius and 85% relative humidity settings caused our media to degrade much too quickly. Because of this, the test parameters where changed to 80 degrees Celsius and 80% percent relative humidity to allow for the collection of more data in the same time period. However, this study adhered as closely as possible to the settings specified in the ECMA 379 standard for the 85 degrees Celsius and 85 percent humidity conditions since they were the closest fit for the conditions actually used in this study. All of the temperature and relative humidity settings as well as the time intervals used in this test can be seen in table 3.1. There was one particular parameter set forth in the standard that this study was not able to follow, the incubation period. The standard specified the incubation period at 250 hours for a total of 1000 hours. In preliminary testing the test discs all failed during the first 250 hour period. Because of this, to facilitate the collection of more data, an incubation time interval of 24 hours was chosen instead. More about this decision can be read in section 4.1.1.1.

Two separate criteria were used to determine the end-of-life of each disc. According to the DVD-R specification, the total number of parity inner errors (PIE) within any eight consecutive ECC blocks, or 8PIE, cannot exceed 280 (ECMA International 2004, 22).

Some studies have indicated that an 8PIE condition of 280 errors may not accurately indicate unrecoverable disc failure. These studies claim that most modern players can accurately recover data from discs with an 8PIE sum of 300 to 500 (Worthington, 2005b). Because of this—and since this study hoped to provide answers for existing DVD-R archives—an alternate "real world" test was performed to determine the end-of-life of the disc. The criteria of this test were based on a DVD reader's ability to successfully extract all of the data contained on the test disc. A standard DVD drive in a modern computer was used to attempt to copy all of the data off of the DVD-R disc to a computer's hard drive. In addition to a successful copy, an MD5 sum check was performed to ensure that the copied files matched the original files. A text file, which contained an MD5 hash for each of the original test files, was created by a software package called DigestIT 2004. To compare the list of hashes to the newly copied files, the software entitled TeraCopy 2.0 beta 2 was used. It is important to note that the MD5 hash comparison, in all instances, agreed with the results of the file copy. There was never an instance where a file that successfully copied from the disk subsequently failed the MD5 sum check.

3.2 DVD-R Analyzer

For the read-back error analysis we used an analyzer built by Optical Disc Technologies (ODT). It uses a drive made by Plextor (model PX-716A). The drive's firmware has been modified by ODT. The analysis software uses this modified firmware to get direct access to the disc's ECC data. The initial analysis of the test discs and each analysis thereafter collected data on the following errors: Parity Inner Errors (PIE), Parity Inner Sum Eight Errors (8PIE), Parity Inner Failures (PIF) and Parity Outer Failures (POF). The software collected the errors with a sampling rate of one sample per second of data (48 ECC Blocks / sample). The read speed for the drive was set to 3-8X Constant Angular Velocity (CAV).

The analyzer also provided various measurements of DD jitter. The analyzer took a sample of data from the beginning (1^{st}), the middle (2^{nd}) and the end (3^{rd}) of each disc. Our sample size was set to 640 KB, the default and recommended size. From that sample, the length of each symbol (3T, 4T, 5T, 6T, 7T, 8T, 9T, 10T, 11T and 14T) and its location were measured. Measurements for both marks and spaces were taken. Using this data, the analyzer calculated the jitter value as well as the deviation from the expected center of the symbol. For each symbol, four separate measurements were recorded: the count, the mean length, the average jitter, and the distance of the deviation from the expected center. For the remainder of this paper, specific symbol measurements will be referenced using four different descriptors: space or mark; 1^{st}, 2^{nd}, or 3^{rd} sample; symbol name; and type of measurement (count, mean, jitter, and dev). For example, to refer to the mark's deviation from the center when measured in the 3^{rd} sample for the symbol 9T, the term Mark-3^{rd}-9T-Dev will be used.

3.3 Analyzing the Data

Once all of the discs had died, the collected read-back errors and jitter data were input into a multiple logistic regression model to determine if any of the specific values (or combinations of the values) held any significance for predicting the data failure on the disc. A stepwise selection technique was used to determine which values predicted more accurately the end-of-life of the DVD-Rs.

A logistic regression uses predictor variables to calculate probabilities of the occurrence of a future event. The format for the model is

$$model = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \tag{3.1}$$

where β_0 is the intercept, β_1 through β_n are the predictor coefficients, and x_1 through x_n are the predictor variables. The probability can then be calculated using the following:

$$probability = \frac{e^{model}}{1+e^{model}} \tag{3.2}$$

Logistic regression modeling is most commonly seen in the field of medicine. It is often used to isolate risk factors for a specific disease or syndrome. After years of collecting data in patients, that information can be used to find commonalties between patients that might give clues to a future condition. For example the use of logistic regression modeling, on data collected over several years for a group of patients who have all suffered from heart attacks, has allowed researchers to identify things like age, gender, cholesterol level and blood pressure as risk factors for heart attacks. With enough information the logistic regression also determines the level of influence each risk factor, or predictor variable, contributes to the probability of having a heart attack. The final equation that is generated from the logistic regression allows medical professionals to make predictions such as this hypothetical example: a 60 year old man with high blood pressure and high cholesterol may have a 40% chance of having a heart attack sometime in the next five years.

A program called SAS (Statistical Analysis Software) was used to perform the logistic regression analysis on the collected read-back errors from the discs in this study. The analysis was intended to discover which read-back errors could be considered risk factors, or predictor variables, for the end-of-life condition of the discs. Once the predictor variables and coefficients were determined by the software, the resulting equation was then re-applied to the original sample data to determine how accurately the calculated model would have predicted the end-of-life of each disc in this study. An example of the SAS script used and the output can be seen in Appendix B.

As stated before, the future event to be predicted using the logistic regression was the end-of-life of a DVD-R. As mentioned in section 3.1.2, the original intent of this study was to use two end-of-life data points for the analysis: a PIE8 Sum count exceeding 280 and the point at which actual data loss occurred. However after the first incubation period, the PIE8 Sum count on every test disc was well over 280. The lack of data leading up to that point prevented us from using the PIE8 Sum metric as an end-of-life point measurement. This study therefore focused on the point of actual data loss as the only end-of-life measurement.

It should be noted that the test was performed at only one temperature and humidity combination. Therefore, not enough data was collected to be able to predict the media’s expected lifetime using the Eyring Acceleration Model. The logistic regression model was the only analysis performed on the data.

4 Data Analysis

While a great deal of interesting data was observed as this experiment was carried out, it should be recognized that since this test was performed on only one type of media in one environmental condition, additional tests are needed before any claim can be made that these results may be applied to DVD-R media in general.

Keeping this in mind, this section will begin by outlining a few general observations that were made about the discs as they went through the testing process. Conclusions may not be drawn from all of the general observations since insufficient data was available to explain some observations. However, these observations were too significant to be ignored and played an important role in how we prepared the data for the logistic regression analysis.

4.1 General Observations

There are two types of results that will be focused on in this section: general trends in the read-back errors, and physical damage.

4.1.1 Observed Trends in the Read-Back Errors

4.1.1.1 Basic Observations of the Parity Inner Sum 8 Error (PIE8 Sum)

The ECMA 379 life expectancy standard specifies a max PIE8 Sum error of 280 as the end-of-life point for their tests. It also dictates the incubation period for the test, which will vary depending on the specific test temperature and relative humidity setting used in each phase. As mentioned in section 3.1.2, the incubation period used in this study was much shorter than what was outlined in the ECMA 379 standard. In order for this study to have followed the ECMA 379 standard precisely, four incubation periods of 250 hours each would have been required. In order to collect enough data, the ECMA 379 standard assumes that the test discs would last at least 500 hours, or two incubation periods, before their max PIE8 Sum error count would reach the pre-determined 280 count.

As mentioned before, this was not the case with the discs in this study. After just 24 hours, every one of the test discs reported errors exceeding the 280 max PIE8 Sum count. In fact, the PIE8 Sum count per ECC block in these discs, averaged over the entire surface of the disc, not only exceeded the 280 value but was well into the 1,000s for most discs. In light of these numbers, the quality of the discs in this particular study was far below the minimum standard that was expected by the authors of the ECMA 379 standard. Not only did this give cause for a much shorter incubation period of 24 hours to be used, but also made it impossible to use the PIE8 Sum error count as an end-of-life metric. Therefore for this study, end-of-life was defined only as the point in time at which any part of the data failed to copy from the disc to another location.

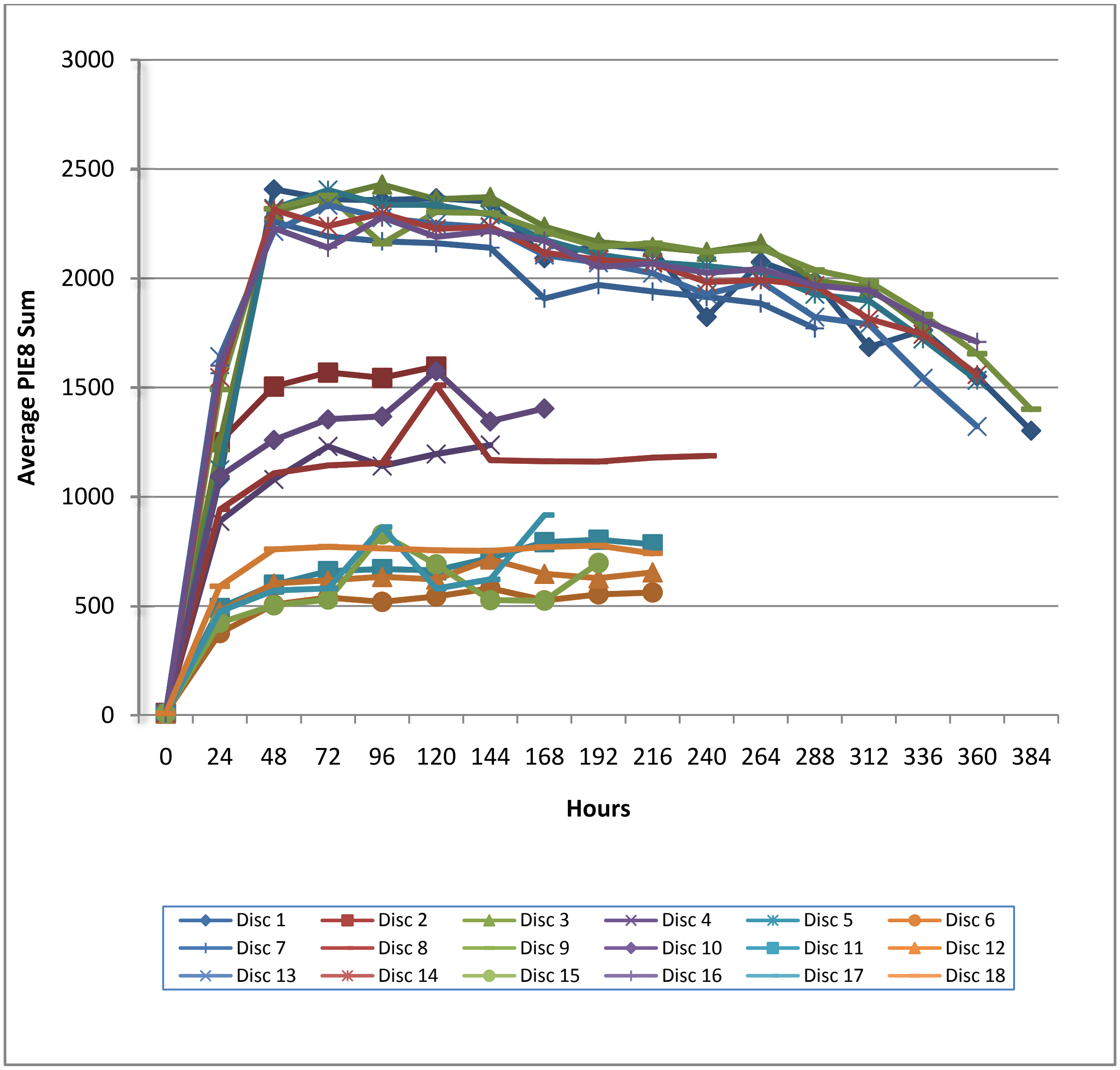


Figure 4.1 - Average PIE8 Sum value for each disc over the lifetime of the test

Figure 4.1 shows a summary view of how long each of the eighteen discs lasted and their average PIE8 Sum count throughout the test. The end-of-life, for each disc, is the point at which no further data is visible on the graph. Interestingly enough, the group with the highest overall error averages lasted the longest. Other than the fact that one group's average errors consistently seemed to improve over time—which will be discussed later—another interesting trend was noted upon examination of this data.

Based on looking at just the PIE8 Sum averages after the first incubation period, the discs were already neatly separated into three very distinct groups. Aside from the similarities in the overall PIE8 Sum averages, the discs also had grouped themselves together in another way after the first incubation period. Distinct patterns had emerged which highlighted significant differences in how the PIE8 Sum errors were increasing over the entire surface of the disc (figure 4.2).

In figure 4.2, two types of data can be seen. The green data, which corresponds to the drive's read speed, will be discussed later. The blue data, representing the PIE8 Sum error counts, is where the three previously mentioned patterns can be seen. These patterns are referred to in this paper as patterns 1, 2 and 3, from top to bottom. Pattern 1 showed high averages in the middle section of the data and was the first group to die. The second pattern, the next group to die, showed low averages in the first half to two thirds of the data and high averages the second half to one third of the data. The third pattern showed high averages in the first half to two thirds of the discs and dropped near the end. Although the third group had the highest PIE8 Sum averages overall, it was the last group to die.

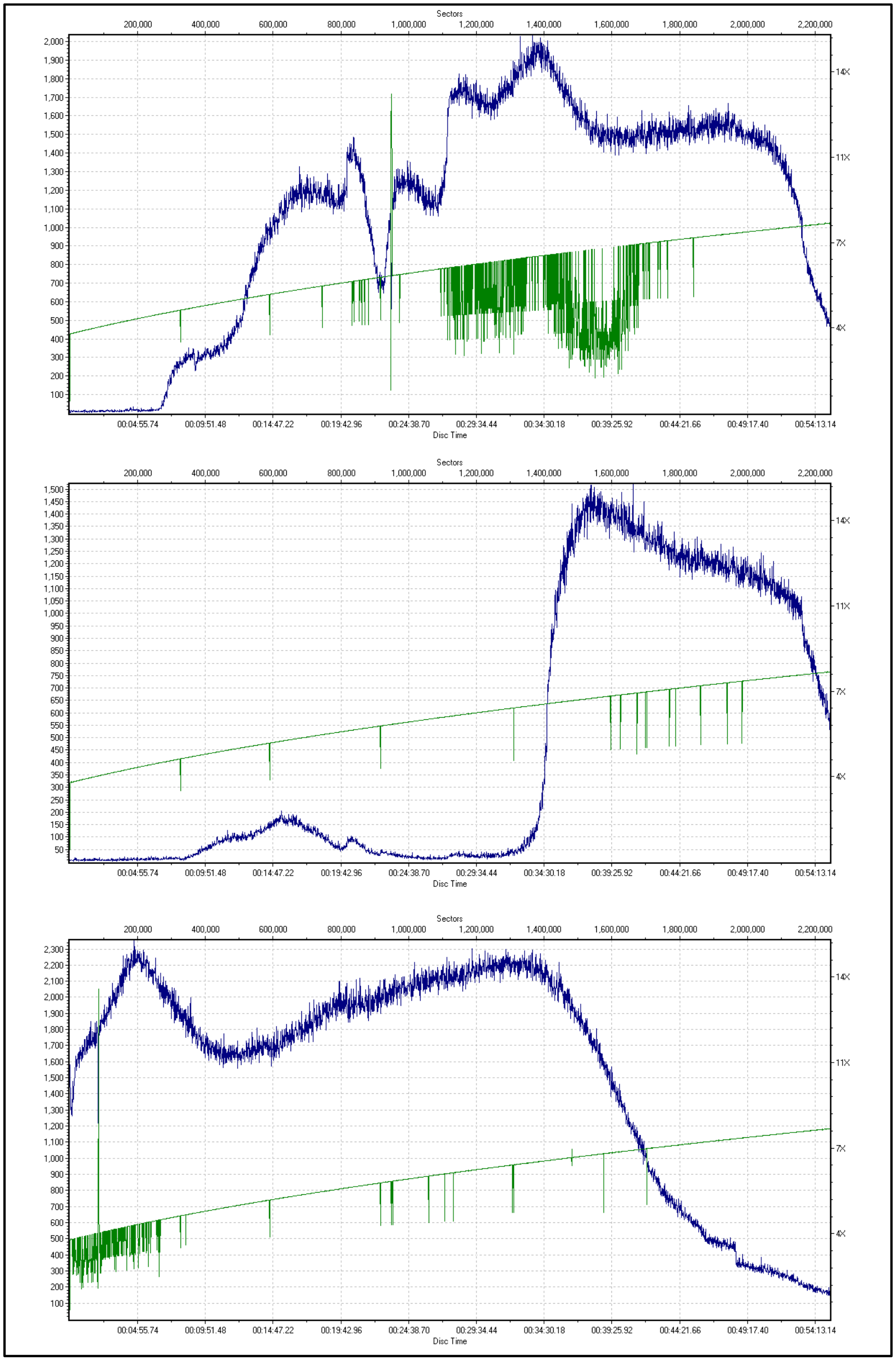


Figure 4.2 – Three disctinct PIE8 Sum patterns (blue) emerged after the first incubation period. From top to bottom are discs 10, 17 and 14. Source: ODT Analysis software.

Remember, that the discs were placed in the environmental chamber on a custom built rack (figure 3.3). This rack positioned the discs at three various levels within the chamber: the highest level defined as level three; the middle as level two, and the lowest as level one. It seemed plausible that the PIE8 Sum error patterns might have been caused by height differences within the chamber. Since after each incubation period the discs were returned to the same rack position, it would seem that if rack height were the cause of the PIE8 Sum patterns, an obvious association would be seen between the height level of the disc and the error pattern that emerged. However, the data did not confirm any such association. Pattern 1 appeared on one disc each from all three levels. There were two discs from level three, two discs from level two and three discs from level one that showed pattern 2. Pattern 3 emerged on three discs from level three, three discs from level two and two discs from level one. It was obvious from the even distribution of each error pattern over the three different heights that there was no association between the disc's height in the chamber and the error pattern that emerged. Summary data for height level, disc number and average PIE8 Sum pattern can be found in table 4.1.

Since these discs were burned using three different write speeds, it also seemed as though these patterns might be easily explained by the three differing write strategies used to burn the discs. However, this was also not the case. Table 4.1 contains data about each of the discs, including the write strategy used as well as the resulting PIE8 Sum error pattern. The data in this table shows that, for the test discs in this study, there was definitely not a link between write strategy and the pattern of PIE8 Sum errors that emerged.

Table 4.1 - Various disc information sorted by average PIE8 error pattern.

Disc	Failure Period	Pattern	Write Speed	Height Level
Disc 2	5th Period	1	MAX	3
Disc 4	6th Period	1	MAX	2
Disc 10	7th Period	1	8X	1
Disc 6	9th Period	2	MAX	2
Disc 8	10th Period	2	8X	1
Disc 11	9th Period	2	8X	1
Disc 12	9th Period	2	8X	1
Disc 15	8th Period	2	6X	2
Disc 17	7th Period	2	6X	3
Disc 18	9th Period	2	6X	3
Disc 1	16th Period	3	MAX	3
Disc 3	14th Period	3	MAX	3
Disc 5	15th Period	3	MAX	2
Disc 7	12th Period	3	8X	1
Disc 9	16th Period	3	8X	1
Disc 13	15th Period	3	6X	2
Disc 14	15th Period	3	6X	2
Disc 16	15th Period	3	6X	3

One clue was found that might explain the emergence of the distinct patterns. In the Test Methodology section, it was asserted that these discs were identical based on the fact that they all came from the same spool and had sequential batch numbers. Looking at only the incrementing portion of the batch numbers, they ranged from 373 to 393. Three numbers were missing from the sequence: 374, 391 and 392. Table 4.2 shows similar data as table 4.1, only this time the data is sorted by the batch number.

Table 4.2 - Various disc information sorted by batch number.

Disc	Failure Period	Pattern	Batch
12	9th Period	2	373
17	7th Period	2	375
6	9th Period	2	376
14	15th Period	3	377
18	9th Period	2	378
16	15th Period	3	379
15	8th Period	2	380
13	15th Period	3	381
10	7th Period	1	382
7	12th Period	3	383
11	9th Period	2	384
9	16th Period	3	385
8	10th Period	2	386
5	15th Period	3	387
4	6th Period	1	388
3	14th Period	3	389
2	5th Period	1	390
1	16th Period	3	393

On close inspection it can be seen that the discs which presented pattern number three, also known as the discs that lasted the longest, have alternating batch numbers starting with 377 and ending with 393. While nothing conclusive can be drawn from this observation without additional information, it seems possible that the different patterns might be explained by looking more closely at how these discs were manufactured and packaged.

While not much may be directly concluded from the appearance of these distinct patterns, the importance for recognizing their existence will become more apparent as we discuss the read-back error data that was collected and how it was organized for analysis.

4.1.1.2 **Basic Observations of the Parity Outer Failure (POF)**

Usually a POF error occurs only when a drive is unable to use the built-in error detection and error correction codes to recover a section of corrupted data. This did not, however, prove to be the case with the results from the ODT analyzer. It is important to remember that the ODT analyzer was only used to analyze the errors. A separate optical drive was used to attempt to copy the disc's data and test for failure to copy. The ODT analyzer showed high POF errors in all of the discs in this study long before the data actually failed. In fact all of the discs, in the last group to fail, showed a max POF value of 270 in multiple ECC blocks. That is one failure per column. One disc even showed an average POF value of 144 errors per ECC block. One explanation for this is that the ODT analyzer does not perform any of the extra read strategies employed by modern drives. Even though the analyzer showed signs of major data degradation on a disc, the modern drive used in this study was able to employ additional strategies to recover all of the data.

Another interesting observation was made regarding the analyzer's behavior as the number of POFs increased. When the analyzer began to encounter large numbers of POFs, its read speed began to slow. In the case of the last group of discs, which began to have very high POFs over the entire surface of the disc, the read speed dropped to below 1X for a large portion of the disc scan. Figure 4.3 shows the reaction of the analyzer's read speed to POF errors. Once again, the blue lines represent the errors while the green line shows the actual read speed of the drive. There are two types of errors shown in figure 4.3. The top graph shows the POF errors while the bottom shows the PIE8 Sum errors.

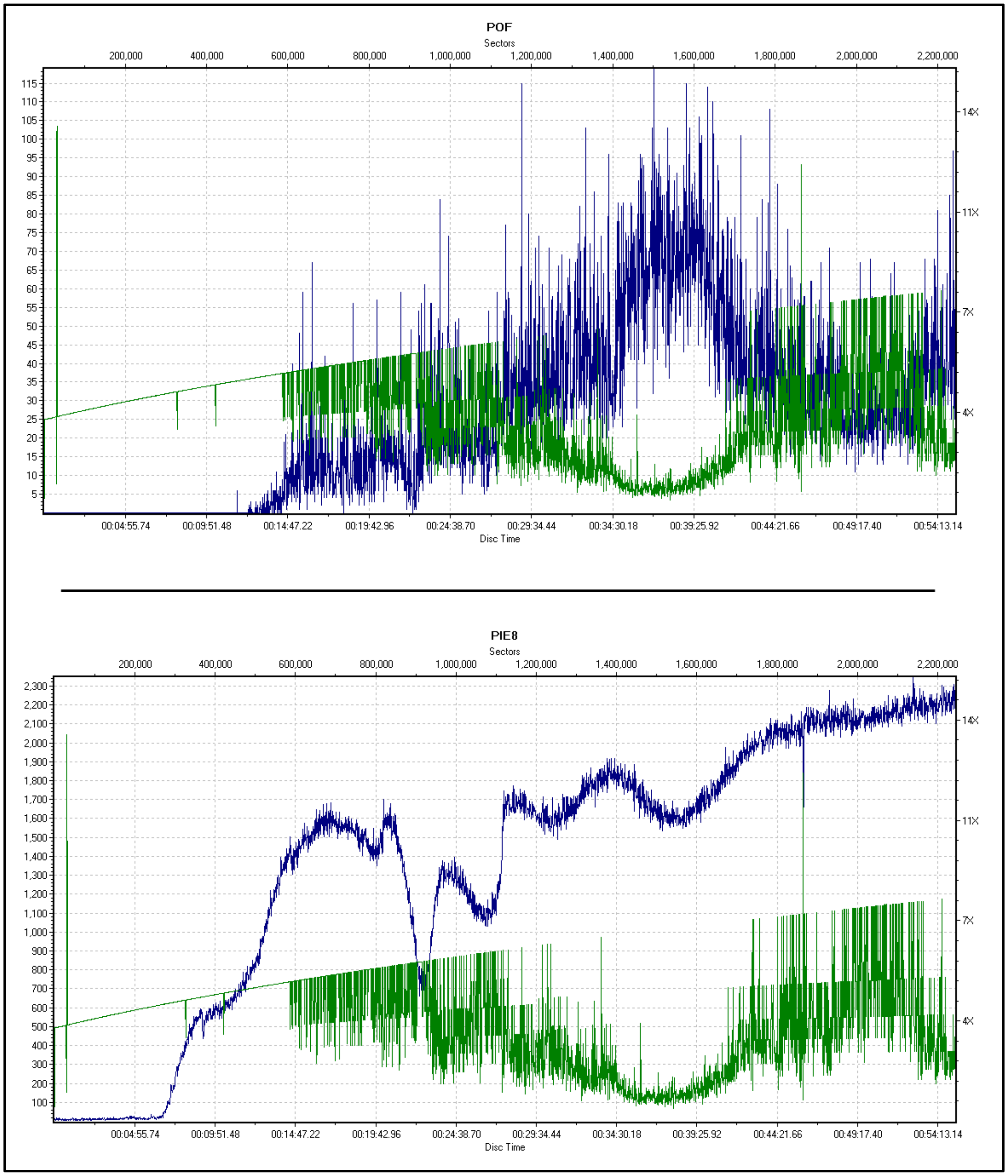


Figure 4.3- A POF graph (top) and PIE8 Sum graph (bottom) of errors for disc number 10 after the 4th incubation period. The left axis shows error count and its data is represented by the blue line. The green line shows the read speed and its values can be seen in the right axis. Source: ODT analysis software.

This decrease in read speed also affected the PIE counts. The PIE count is greatly influenced by the read speed of the drive. The same disc will see its PIE count increase as the read speed increases. If the read speed drops, the PIE count drops as well. Keeping this in mind, it was only natural to see a decrease in the PIE8 Sum average in regions of the disc where the POF count began to rise. Going back to the graph in figure 4.1, this trend can be seen represented in the data. Looking at the data for the group of discs that lasted the longest, the PIE8 Sum average dropped significantly before the discs finally died. This is because the POFs increased significantly, causing the average read speed to drop drastically. Figure 4.4 shows both the average PIE8 Sum errors as well as the average POF errors as seen on disc 13 over its entire test lifetime.

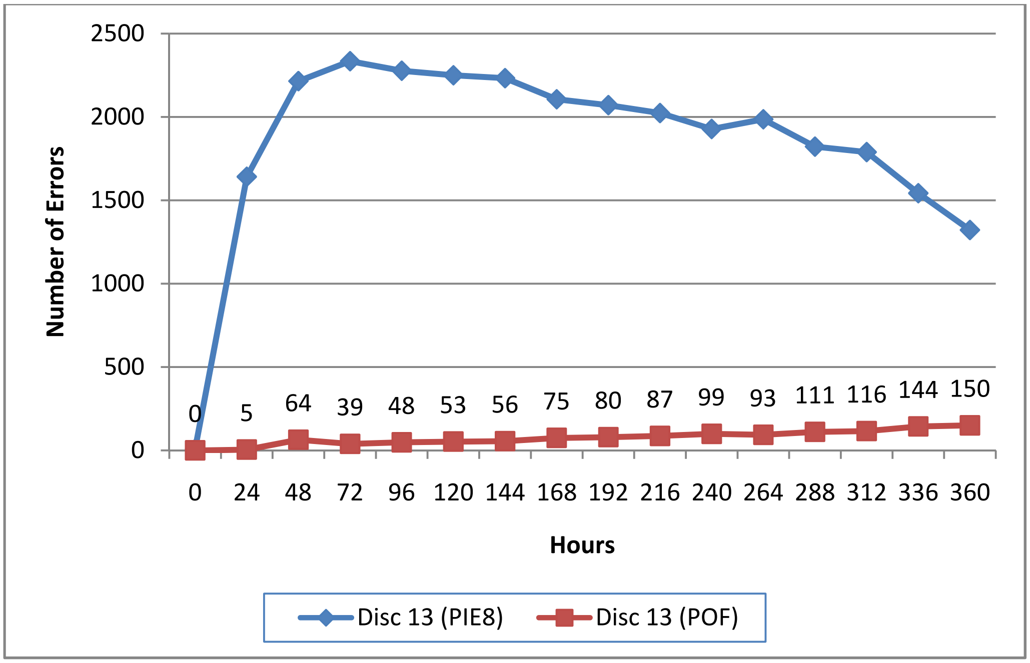


Figure 4.4 - Average POE and average PIE8 Sum for disc 13.

4.1.2 **Physical Damage**

The first physical change that became apparent began to appear within the first few incubation periods. Spots appeared on the label side of a few discs (Figure 4.5). These discs have an inkjet printable surface. These spots were most likely the result of moisture in the air being collected and absorbed by this surface. Since these spots appeared on the label side, it is unlikely that they affected the data in any way.

Figure 4.5 – Spots on label side of disc.

After several days in the chamber, a few of the discs started to show a type of damage that was visible along the outer edge of the discs. This damage can be seen in figure 4.6. At first glance it appeared that the label face itself was starting to peel from the back surface of the disc. However, on further inspection it became apparent that the damage might actually be resulting from separation of the top substrate from the metal layer. In other words, it would seem that the visible damage was caused by degradation of the adhesive layer.

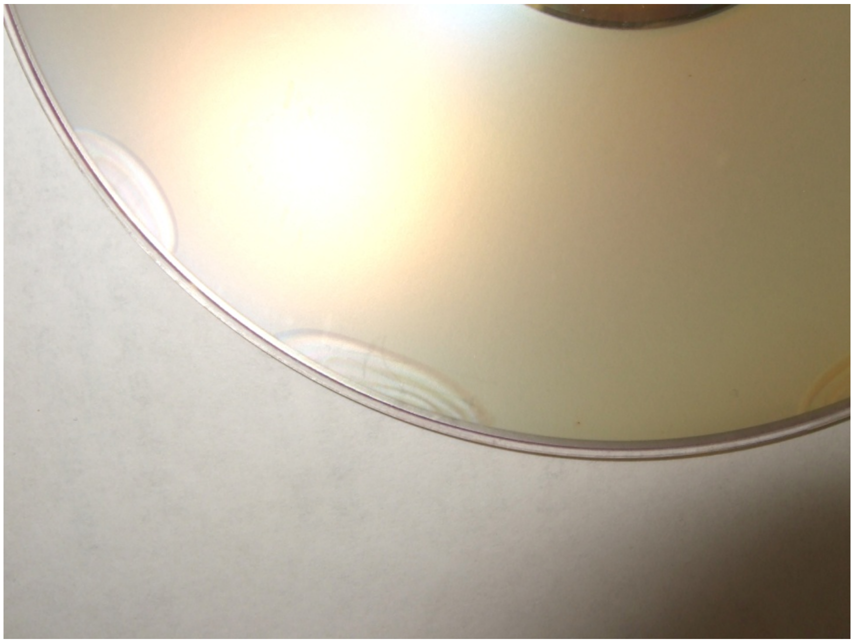


Figure 4.6 – Damage caused by deterioration of the adhesive layer.

In figure 4.7, further proof can be seen that this damage was occurring within the inner layers. It shows a purple substance appearing along the outer edge of the disc in the same area as the damage. This discoloration was most likely a result of the dye layer seeping out of the disc through the damaged area. To further validate that this damage was indeed being caused by a breakdown of the adhesive layer, attempts were made to manually separate the layers by prying them apart, starting in the area where this damage was visible. With very little effort, the layers began to separate. The top substrate, or dummy substrate, came free leaving the dye layer and reflective layer largely intact on the lower substrate (Figure 4.8). Upon close inspection of the image, residue of the purple dye layer was left behind and was visible on the lower substrate where a small section of reflective layer was removed with the dummy substrate.

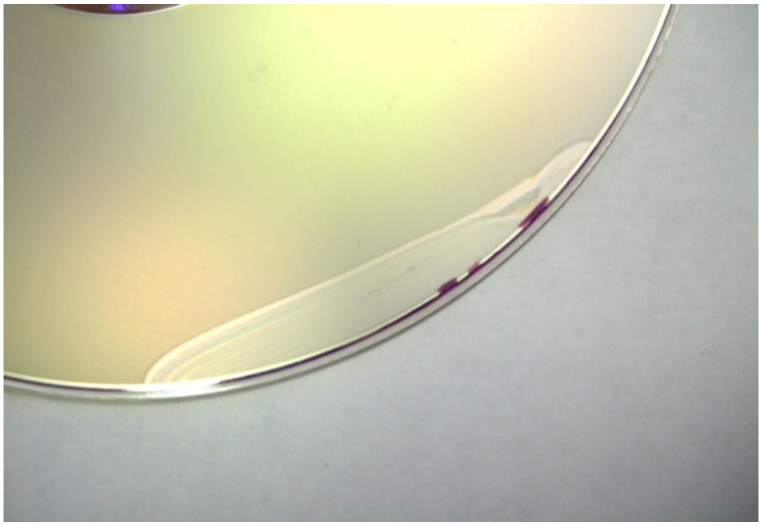


Figure 4.7 – Purple discoloration possibly due to leakage of the dye layer.

Figure 4.8 – Dummy substrate (left) separated from disc (right).

While fourteen of the eighteen discs showed various degrees of the breakdown of the adhesive layer, the purple discoloration was only visible on two of the discs. It is not clear whether or not this damage affected the data on the disc. None of this damage was visible to the naked eye from the data side of the disc. Furthermore, the damage only began to appear after the ninth and tenth incubation periods. At this point ten out of the eighteen discs had already failed. If this damage did contribute to failure, it would have only done so on the remaining eight discs.

4.2 Logistic Regression

While the general observations reveal some interesting things about the more obvious trends, it is difficult, and often erroneous, to claim concrete findings from simple observations. Furthermore, much of the data collected was either too abundant, or the differences were too subtle, to observe any trends without some form of non-biased analysis. For this purpose, logistic regression was used to analyze the data further to determine whether any of the observed trends truly held any predictive value for anticipating the end-of-life of the discs in this test. Additionally, this analysis looked at all of the data collected, compared the various values against each other and output which errors provided the greatest significance for predicting failure.

A logistic regression was run on three separate sets of data. The first set consisted of both the jitter data and the read-back errors averaged over the entire surface of the disc. This is the default output for the analyzer software. The analyzer software also had the ability to export the read-back error data at a resolution determined by the sampling frequency of the software. The sampling frequency used for the analysis in this study was 1 sample for every 48 ECC blocks, or roughly 1 sample for every 1.5 megabytes. This made it possible to calculate the errors manually for any portion of the disc chosen.

Since the exact area of failure for each disc was known—based on the sector in which the failed file resided—a twenty megabyte sample of the error data surrounding the failed portion was taken. This data was then averaged and a logistic regression analysis was performed on this second set of data. Unfortunately, due to the manner in which the jitter data was collected by the analyzer, it was not possible to include jitter data as part of the analysis specific to the failed portion.

The third and final set of data, which was analyzed using logistic regression, was the averaged data taken from only the last half of the disc. Since the point of failure for every disc in this study occurred somewhere in the last half of the disc, it seemed appropriate to focus on that area as well.

4.2.1 **Comparing the Prediction Equations**

In order to determine which of the prediction equations produced the most accurate results, they were compared using three parameters: number of false positives, number of false negatives, and the percentage of actual disc life maintained through the use of the prediction equation.

4.2.1.1 **Understanding the Prediction Equation**

In order to understand how the equations were compared, it is first crucial to understand how the equation is used. As discussed earlier, the regression analysis provides which metrics to use in the analysis as well as a coefficient specific to each metric. Refer back to section 3.3 for help in understanding the formulaic substitutions that follow. For example, assume that the regression analysis specified two read-back errors, the average POF (x_1) and the standard deviation (StdDev) of the PIEs (x_2). Along with this data it also specified an intercept (β_0), the average POF coefficient (β_1), and the PIE StdDev coefficient (β_2). Throughout the results section of this paper, the data returned by the logistic regression analysis will be presented in tabular form (Table 4.3).

Table 4.3 - Sample data returned by logistic regression.

Error Type	Coefficient
Intercept	-6.095
Average POF	-0.0308
PIE StdDev	0.0388

Following the steps outlined in section 3.3, the data from table 4.3 can be used to create a probability of failure equation. This particular equation would be:

$$probablity = \frac{e^{-6.095 - (AveragePOF \times 0.0308) + (PIEStdDev \times 0.0388)}}{1+e^{-6.095 - (AveragePOF \times 0.0308) + (PIEStdDev \times 0.0388)}} \tag{4.1}$$

The actual data from each incubation period is then entered back into this equation and a probability of survival after the next incubation period is calculated. Table 4.4 contains sample data from a disc and the results after entering in the collected data back into the above probability equation. This disc survived five incubation periods and failed during the sixth.

Table 4.4 - Sample data and results from regression analysis.

Time Period	Average POF	PIE StdDev	Model	Probability of Failure
Initial	0	0.867837	-6.06133	0%
1	9.460750853	45.28064	-4.04672	2%
2	27.02303754	47.63712	-3.41437	3%
3	31.14590444	57.09069	-2.92059	5%
4	38.48549488	55.42953	-2.75898	6%
5	**33.8447099**	**54.88446**	**-2.92307**	**5%**

In order to use this equation, a **probability of failure threshold** must be determined. This threshold is what is used to determine when a disc's calculated failure probability is high enough to warrant replacement. For the disc represented in table 4.4, using a 5% probability of failure threshold would have resulted in the disc being replaced before its actual end-of-life had been reached.

However, using a 5% probability of failure threshold would have marked the disc for replacement after period 3. This is known as a false positive and would have resulted in the disc being retired with only 50% of its actual life being used. This measurement will be referred to as the disc's **percent of actual lifetime**. While this may appear undesirable, it is important to remember that with false positives, data loss is still prevented. A good equation may have a high number of false positives. The key is to prolong the point in the disc's actual lifetime when those false positives occur. An effective equation may result in every disc having a false positive, especially if those false positives occur near the end of each discs actual lifetime.

A far worse outcome, however, would result in choosing a probability of failure threshold that is too high. If a 6% threshold had been chosen for this test disc, it would never have been replaced and data loss would have occurred without warning. This is known as a false negative and is a much worse result than a false positive.

In order to complete the analysis of this sample equation, the data from all of the test discs would be run back through the equation. This would create data, similar to that found in table 4.4, for every disc in the test. This data would also be analyzed to determine if a common probability of failure threshold could be chosen that would result in all of the discs being flagged for replacement before their actual end-of-life occurred. Once a threshold was discovered, the percent of actual lifetime would be determined for each disc by applying that threshold back to the data to determine at what point the discs would have been flagged for replacement. The percent of actual lifetime for each disc would then be averaged together to determine the **average percent of actual lifetime** for the equation itself.

Consequently, the most important measurement of quality, for the equations derived from the logistic analyses of the three datasets, was the ability to choose a **probability of failure threshold** that completely removed all false negatives. The quality could be further defined by identifying which equations were able to maximize the lifetime of the discs through either the lack of false positives or through the false positives occurring late in the disc's actual lifetime. In other words, an equation that resulted in a higher **average percent of actual lifetime** could be considered a better equation.

One last measurement that could be used to determine the quality of the equation is the **flexibility** of the probability of failure threshold. An equation that offered a wide range of options for the probability of failure threshold would be more versatile and could be adapted to suite a wider range of applications. With this type of equation, the lower probability of failure threshold would be used to ensure that no disc would reach its end-of-life prior to replacement. This choice, however, would also result in the lowest average percent of actual lifetime for the equation. To maximize the percent of actual lifetime, the highest probability of failure would be chosen. This, unfortunately, could increase the risk of a disc reaching its end-of-life before replacement.

And last of all, the **complexity** of the probability of failure equation could prove valuable to the overall determination of an equation's quality. An equation that is able to predict end-of-life with only two read-back errors as variables, may prove more valuable than one that requires ten.

4.2.2 Analyzing the Datasets

For the sake of comparison, a threshold of 5% was first used on all of the resulting equations for each data set. This value was chosen as it seemed to be the best value which worked for all datasets. To determine the flexibility of the equation, the upper limit of the threshold for each equation was also determined.

While the regression analysis was able to use the subtle interplay of the various errors to predict failure, certain interactions between specific errors might be missed by the analysis. If one of these interactions exists, it can be magnified by multiplying the errors in question and adding the product of that multiplication back as another metric to be analyzed using the logistic regression. Unfortunately the only way to catch them all is to represent every error combination within the study. Considering that the study looked at over 260 unique errors, this approach was not practical. However, based on the general observations and some of the preliminary regression analyses, there seemed to be an interaction between the PIE8 Sum and the POF errors. For the logistic regression we added two interactions: average PIE8 Sum multiplied by average POF, and average PIE8 multiplied by max POF. Regression analysis was performed twice on each data set, once with these additional interactions and once without them. The addition of these interactions proved useful in several of the final equations.

4.2.2.1 **Logistic Regression of the Entire Disc**

It may seem unlikely that data, averaged over the entire surface of the disc, would not prove useful for predicting failure in a small portion of the disc. However, the equation generated from the regression analysis performed well. Table 4.5 shows which errors were selected by the analysis and their corresponding coefficients. Note that the added interactions did not play a role in this analysis.

Table 4.5 - Logistic regression values for whole disc data

Error Type	Coefficient
Intercept	0.8524
Mark-2nd-9T-Dev	-2.0298
Mark-3rd-14T-Jitter	8.447
Space-3rd-3T-Jitter	-10.05
PIE8 Max	-0.00811
POF Max	0.0634

With a probability of failure threshold of 5%, the prediction equation for the whole disc data was able to successfully predict the end-of-life for all discs. Out of the eighteen discs, each one reported a false positive. The average percentage of actual lifetime for all eighteen discs was 26%. This equation was able to be improved by increasing the threshold to 96% probability of failure. At the new threshold, the number of false positives was reduced to 15. This new threshold also was able to increase the average percentage of actual lifetime to 55%. With the usable probability of failure thresholds ranging from 5% to 96%, this equation proved to be very flexible. It had five input variables giving it an average complexity score (figure 4.6).

Table 4.6 - Summary of quality metrics for the whole disc data probability equation

Equation Name	Probability of Failure Threshold / Flexibility	Complexity	Percent of actual life
Whole Disc Data	5% - 96%	5 Variables	26% - 55%

4.2.2.2 **Logistic Regression of the Failed Portion of the Disc**

Looking directly at the section of the disc that failed seemed like it would provide the most accurate information for predicting failure. Remember, this data set did not contain any jitter information. However two different equations were produced for this dataset, one with the previously mentioned interactions and one without. The output for both equations can be seen in table 4.7.

Table 4.7 - Logistic regression values for failed portion of disc

Equation 1	
Error Type	**Coefficient**
Intercept	-3.0665
PIE8 Avg	-0.00308
POF Max	0.0322
PIE8 StdDev	-0.0307
PIF StdDev	0.8053
Equation 2 (w/ interactions)	
Error Type	**Coefficient**
Intercept	-4.8473
POF Max	0.0618
PIE8 x Avg POF	-0.00003

The error data was entered back into equation 1 using a 5% probability of failure threshold. It resulted in 17 false positives and the end-of-life of all eighteen discs was accurately predicted. However the percent of actual lifetime with this equation was only 24%. In addition, the 5% threshold proved to be the only option for this equation. Lowering the threshold did not increase the average lifetime and raising it would cause false negatives. In other words, this equation received the lowest score possible for flexibility (Table 4.8).

The second equation, with the interaction, proved to increase the average lifetime of the discs to 50%, but introduced a false negative when the 5% probability of failure threshold was used. Lowering the threshold to 3%, however, did eliminate the false negative. The new parameters caused the average lifetime of the discs to drop to 42% of their actual life. While the second equation did improve upon the average percent of actual lifetime, given that 3% was its only choice for the probability of failure threshold, it also scored low for flexibility. However, since this equation only required two input variables, POF Max and PIE8/Avg POF interaction, it scored well for complexity (table 4.8).

Table 4.8 – Summary of quality metrics for the failed portion probability equation

Equation Name	Probability of Failure Threshold / Flexibility	Complexity	Percent of actual life
Failed Portion	5%	4 Variables	24%
Failed Portion w/interaction	3%	2 Variables	42%

4.2.2.3 **Logistic Regression of the Last Half of the Disc**

Since every one of the test discs failed at some point in the last half of their data, isolating and analyzing that section seemed important. Once again, with this section of data the same average PIE8 Sum interaction with the POF average data helped to improve the equation. Also, jitter data was included with this data set, but only from the 3rd sample which was taken from the last section of the disc (see section 3.2). Table 4.9 summarizes the results from analyzing this area of the disc.

Table 4.9 - Logistic regression values for the last half of the disc

Equation 1	
Error Type	**Coefficient**
Intercept	-2430.6
Mark-3rd-5T-Jitter	27.1225
Space-3rd-3T-Jitter	-23.6958
Space-3rd-5T-Mean	12.9953
POF Avg	-0.0675
PIE8 Max	-0.00948
Equation 2 (w/interactions)	
Error Type	**Coefficient**
Intercept	-3539
Mark-3rd-5T-Jitter	37.8798
Space-3rd-3T-Jitter	-33.5693
Space-3rd-5T-Mean	18.9425
PIE8 Max	-0.0133
PIE8 StdDev	-0.0138
POF StdDev	0.2837
PIE8 x POF	-0.00008

Applying the 5% probability of failure threshold, equation 1 from table 4.9 provided excellent results. It only had 15 false positives. There were no false negatives and the average percent of actual lifetime was 79%. Changing the probability of failure threshold to 7% dropped the false positives to 14 and increased the percent of actual lifetime slightly, to 80%. This equation had very high average percentage of actual life. Its probability of failure threshold had a useful range of 5% to 7%, giving it a mediocre score for flexibility. Its five input variables gave it an average score for complexity (Table 4.10).

Once again, adding the interactions to the equation also helped in the long run. With the new equation, there were only 9 false positives when using the 5% probability of failure threshold. This did come with the cost of one false negative making the 5% probability of failure threshold unusable. However, dropping the threshold just one percent eliminated the false negative. The value for average percent of actual lifetime for this configuration was an astounding 89%, the highest of all five equations. However the new equation had no flexibility due to 4% being its only option for the probability of failure threshold. With seven input variables, it also received the worst score for complexity (Table 4.10).

Table 4.10 – Summary of quality metrics for last half of the disc's probability equation

Equation Name	Probability of Failure Threshold / Flexibility	Complexity	Percent of actual life
Last 50% of Disc	5% - 7%	5 Variables	79% - 80%
Last 50% of Disc w/interactions	4%	7 Variables	89%

5 Conclusions and Recommendations

5.1 Research Summary

For the past ten years, the DVD-R has been a popular choice for storing digital content. The trust in the newer optical storage format was largely based on the success and stability of its predecessor, the CD-R. Transitioning to the larger capacity format seemed like a logical choice. In 2004, a study performed by NIST cast doubt that the DVD-R format was truly stable enough to be trusted as an archival format. Some types of DVD-Rs worked while others fell quite short of expectations. Further compounding the situation was the fact that the consumer had no practical way to tell the difference between discs that would last and those that would not (Slattery et al. 2004).

One of the outcomes of the NIST paper was a proposal to work with the Library of Congress and other standards organizations to publish a test methodology and ratings criteria to be used by disc manufacturers (Slattery et al. 2004, 524). These standards would provide a consistent grading scale that could be placed on the disc's packaging, allowing consumers to easily identify the certified life expectancy and quality of a specific type of DVD-R media. While this process still has not been completed, evidence that it is underway can be seen. One such piece of evidence was the standardized life expectancy test methodology for DVD-R media that was published in June of 2007 as the ECMA Standard 379 (ECMA 2007). This document outlines the use of accelerated lifetime testing and the Eyring Acceleration Model to calculate the expected lifetime of a specific line of DVD-R media. While these changes should benefit the industry greatly, their completion will do nothing for the countless DVD-R archives that currently exist today and will continue to be created between now and the time of implementation.

This research focused on testing whether or not information that is readily available to the consumers—disc read-back errors—could be used to predict the failure of a DVD-R. To test this, eighteen discs were aged in an environmental chamber at 80 degrees Celsius and 80 percent relative humidity. Every twenty- four hours, the discs were removed and each disc was analyzed to record the current state of its read-back errors. When actual data loss occurred in a disc, or when one or more files failed to copy from the disc, it was considered to have reached its end-of-life. Once all of the discs had reached their end-of-life, a logistic regression analysis was performed on various sets of the collected data. The logistic regression analyzed each of the read-back errors to determine which error, or combination of errors, held any significance for predicting the end-of-life of each disc. The end result of the logistic regression was the creation of several equations with varying levels of accuracy for predicting a disc's end-of-life. The test data was then entered back into each of the equations in order to test and compare its ability to predict a disc's end-of-life.

5.2 Conclusions

The logistic regression analysis on the three unique datasets produced five functional equations, each one capable of anticipating the end-of-life for each disc. The performance of each equation was further tuned by the selection of an appropriate probability of failure threshold. The optimum probability threshold was the one that provided the highest percent of actual disc life, minimized the number of false positives, and completely eliminated any false negatives. Table 5.1 summarizes the results from the five different equations. The equations are named for the dataset from which they were derived.

Table 5.1 - Results summary for the equations derived from the logistic regression

Equation Name	Probability of Failure Threshold / Flexibility	Complexity	Percent of actual life
Whole Disc Data	5% - 96%	5 Variables	26% - 55%
Failed Portion	5%	4 Variables	24%
Failed Portion w/interaction	3%	2 Variables	42%
Last 50% of Disc	5% - 7%	5 Variables	79% - 80%
Last 50% of Disc w/interactions	4%	7 Variables	89%

Based on the results from this study, the equation generated from the whole disc data seems to be the best overall equation. It was the only equation that offered any real flexibility for the probability of failure threshold. And while 26% to 55% of actual lifetime may not seem ideal, it is important to remember that current DVD-R quality tests use the maximum value of 280 PIE8 Sum as the end-of-life for a disc. At just 25% of actual life, our discs were showing average PIE8 Sum values in the thousands. In other words, in the case of the test discs in this study, 25% of actual life is much higher than they would have been given using the currently used quality metric of 280 PIE8 Sum.

That being said, which equation to use will be largely determined by which quality measurement proves most important. At this time, flexibility seems as though it would be the most useful. However, future studies may show that certain read-back errors cannot be generalized in a standardized probability of failure equation. For example, if at a later time jitter data, tied specifically to a certain symbol (i.e. 3T), proves to be too closely tied to the type of data on the disc, then a simple equation that does not include the jitter data would be more reliable. Therefore the equations calculated from the failed portion of the disc, which use no jitter data, would be the more effective equations.

In conclusion, the end result of this research showed that through the collection of read-back error data and logistic regression analysis, it was indeed possible to predict the end-of-life of the discs in this study. In fact, this study generated five separate equations from data taken from three distinct data sets from the discs: data from the whole disc, data from the failed portion of the disc, and data from the last 50% of the disc. All five of these equations were able to predict the end-of-life of each disc in this study.

5.3 Recommendations for Future Research

Because of the limited sample size, results derived in this study cannot yet be extended to DVD-R media in general. This research will need to be repeated under a variety of conditions to test if the results will be consistent. Below is a list of several additions that can be made to future studies to help solidify these results.

- Various grades of DVD-Rs from multiple disc manufacturers should be tested.

- Repeat the test using discs that allow for using Max PIE8 Sum of 280 as the end-of-life.
- The discs should be burned using several different burners.
    - The results of this study could be linked to specific write strategies implemented by the burner.
- The discs should be incubated at several different test temperatures and relative humidity settings.
- The discs should be burned with completely different sets and types of data.
    - This would eliminate any results that might have arisen from each sector of the discs containing identical bits.
- A method should be derived for representing the incubation period under normal operating conditions.
    - The linear regression predicts failure during the next incubation cycle.
    - Performing the Eyring acceleration testing in conjunction may provide the mechanism for making this conversion.
- Inclusion of additional read-back errors or combinations of read-back errors (interactions).

  - The jitter information was very limited in this study.
  - The analyzer provided a few proprietary read-back errors that were not included in this study due to lack of standardization.
    - Perhaps closer collaboration with the vendor of the analyzer to define and refine these additional read-back errors would result in even more accurate predictions.
- The use of multiple analyzers to verify and compare each analysis
  - If different brands of analyzers produce drastically different results, the produced equations would be specific to certain brands of analyzers.

Bibliography

Comtech AHA Corporation. AHA Application Note. Primer: Reed-Solomon Error Correction Codes (ECC). http://www.aha.com/show_pub.php?id=39. (accessed January 20, 2007).

Bennett, H. 2004. Understanding Recordable & Rewritable DVD First Edition: Optical Storage Technology Association (OSTA).

Byers, F. R., National Institute of Standards and Technology (U.S.), and Council on Library and Information Resources. 2003. *Care and handling of CDs and DVDs : a guide for librarians and archivists*. Gaithersburg, MD: Council on Library and Information Resources.

ECMA International. 2001. Standard ECMA-359 3rd Edition. Geneva: ECMA International.

ECMA International. 2004. Standard ECMA-359 1st Edition. Geneva: ECMA International.

ECMA International. 2007. Standard ECMA-379 1st Edition. Geneva: ECMA International.

ECMA International. What is ECMA International. http://www.ecma-international.org/memento/index.html (accessed May 31, 2007).

Hites, R. A. 2007. *Elements of environmental chemistry*. Hoboken, N.J.: Wiley-Interscience.

Iraci, J. 2005. The Relative Stabilities of Optical Disc Formats. *RESTAURATOR-International Journal for the Preservation of Library and Archival Material*:134-150.

Irie, M., and Y. Okino. 2007. Standardized Life Expectancy of High-Speed Recordable Optical Disks. *IEEE Transactions on Magnetics* 43:864-866.

ISO, International Organization for Standardization. 2002. ISO 18927. Imaging materials — Recordable compact disc systems — Method for estimating the life expectancy based on the effects of temperature and relative humidity.: International Organization for Standardization.

Khurshudov, A. 2001. *The essential guide to computer data storage : from floppy to DVD*. Upper Saddle River, NJ: Prentice Hall PTR.

Leek, M. 1995. A Look at CD-Recordable Media Longevity Studies - Will a Good Disc Last Forever. *CD-ROM Professional*:102-110.

Meyer, J. C. 1990a. Error Correction Implementation and performance in a CD-ROM Drive. *Hewlett-Packard Journal*:42-48.

Morelos-Zaragoza, R. H. 2002. *The art of error correcting coding*. Chichester: John Wiley & Sons.

Nijhuis, G. Optical Compact Disc Code. LaesieWorks. http://www.laesieworks.com/digicom/Storage_CD.html. (accessed July 18, 2007).

Purcell, L. 2000. *CD-R/DVD : disc recording demystified*, *McGraw-Hill video/audio professional*. New York: McGraw-Hill.

Purcell, L., and D. Martin. 1997. *The complete recordable-CD guide*. San Francisco, CA: SYBEX.

Nikles, D. E., and J. M. Wiest. 2000. Accelerated Aging Studies and the Prediction of the Archival Lifetime of Optical Disk Media. http://romulus.gsfc.nasa.gov/msst/conf2000/PAPERS/A03PA.PDF. (accessed January 20, 2008).

Pioneer Corporation. DVD Technical Guide. http://pioneer.jp/crdl/tech/index-e.html. (accessed July 12, 2007).

Slattery, O, RC Lu, J Zheng, F Byers, and X Tang. 2004. Stability comparison of recordable optical discs - A study of error rates in harsh conditions. *Journal of Research of the National Institute of Standards and Technology* 109 (5):517-524.

Tobias, P. A., and D. C. Trindade. 1995. *Applied reliability*. 2nd ed. New York: Van Nostrand Reinhold.

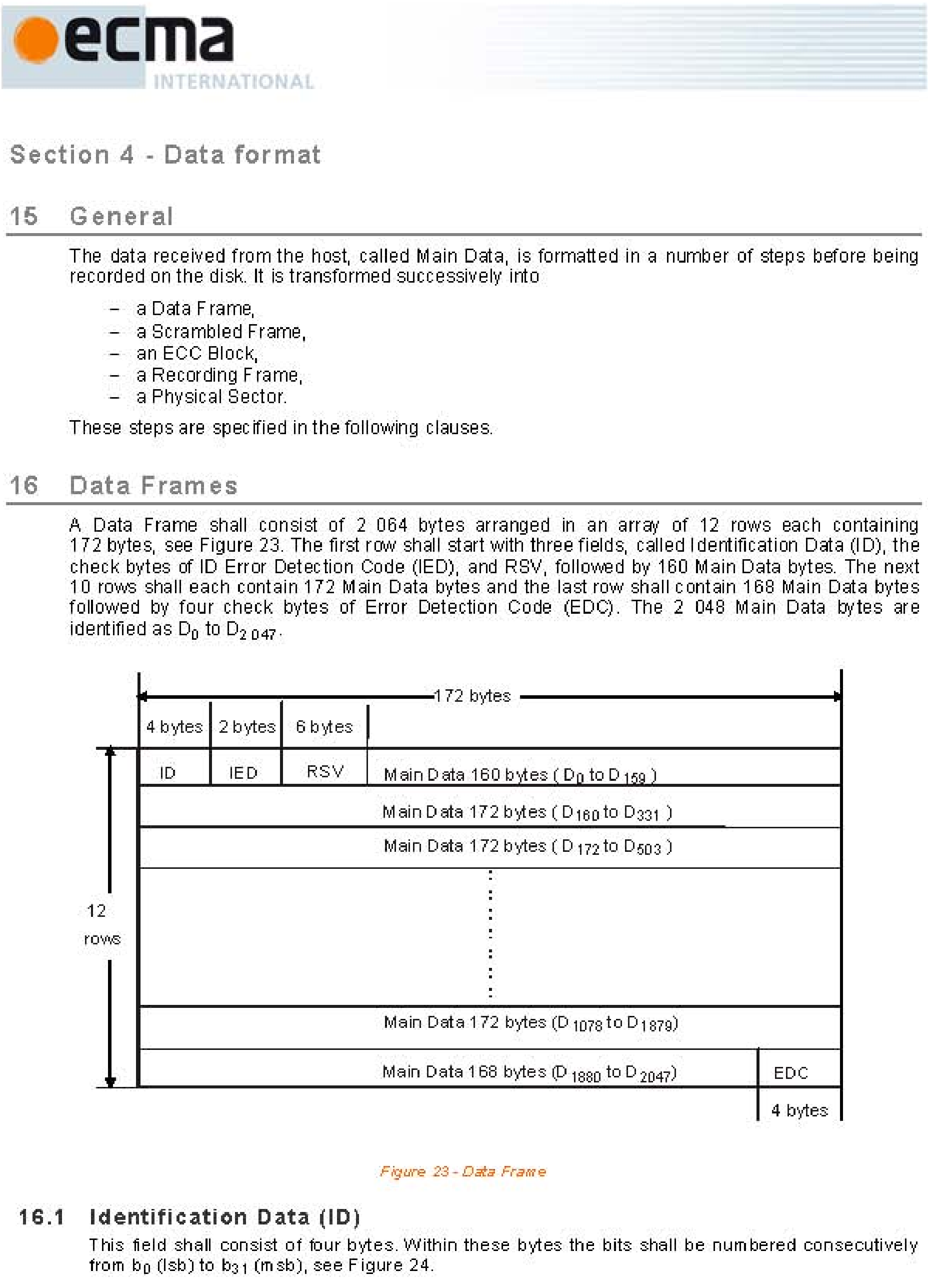
Wilhelmsson, U. 2004. "Inter Symbol Interference" AudioDev Technical Whitepaper. http://www.audiodev.com/filearchive/2/2094/AudioDev%20ISI%20White%20Paper.pdf (accessed January 23, 2008).

Worthington, M. Optical Disc Technology. CD, DVD, BD and Beyond. Presentation given to Brigham Young University, Provo, UT, 2005a.

Worthington, M. DVD Quality, "Strategies and Measurements". Revision 0.2. Paper presented during a lecture given at Brigham Young University, Provo, UT, 2005b.

APPENDICES

Appendix A – Section 4 from ECMA 359 (pp. 32-42)



ecma INTERNATIONAL

Section 4 - Data format

15 General

The data received from the host, called Main Data, is formatted in a number of steps before being recorded on the disk. It is transformed successively into

- a Data Frame,
- a Scrambled Frame,
- an ECC Block,
- a Recording Frame,
- a Physical Sector.

These steps are specified in the following clauses.

16 Data Frames

A Data Frame shall consist of 2 064 bytes arranged in an array of 12 rows each containing 172 bytes, see Figure 23. The first row shall start with three fields, called Identification Data (ID), the check bytes of ID Error Detection Code (IED), and RSV, followed by 160 Main Data bytes. The next 10 rows shall each contain 172 Main Data bytes and the last row shall contain 168 Main Data bytes followed by four check bytes of Error Detection Code (EDC). The 2 048 Main Data bytes are identified as D_0 to $D_{2\,047}$.

Figure 23 - Data Frame

16.1 Identification Data (ID)

This field shall consist of four bytes. Within these bytes the bits shall be numbered consecutively from b_0 (lsb) to b_{31} (msb), see Figure 24.

b_{31} b_{24}	b_{23} b_0
Sector Information	Sector Number

Figure 24 - Identification Data (ID)

b_{31}	b_{30}	b_{29}	b_{28}	b_{27} and b_{26}	b_{25}	b_{24}
Sector Format type	Tracking method	Reflectivity	Reserved	Zone type	Data type	Layer number

Figure 25 - Sector Information of the Identification Data (ID)

The least significant three bytes, bits b_0 to b_{23}, shall specify the sector number in binary notation. The sector number of the first sector of an ECC Block of 16 sectors shall be a multiple of 16.

The bits of the most significant byte shown in Figure 25, the Sector Information, shall be set as follows.

a) Sector format type	bit b_{31}	shall be set to ZERO, indicating the CLV format type specified for Read-only disk and Recordable disk.
b) Tracking method	bit b_{30}	shall be set to ZERO, indicating Differential Phase tracking.
c) Reflectivity	bit b_{29}	shall be set to ZERO, indicating the reflectivity is greater than 40%, measured with PBS PUH.
d) Reserved	bit b_{28}	shall be set to ZERO.
e) Zone type	bit b_{27} and bit b_{26}	shall be set to ZERO ZERO in the Data Zone. shall be set to ZERO ONE in the Lead-in Zone. shall be set to ONE ZERO in the Lead-out Zone.
f) Data type	bit b_{25}	shall be set to ZERO, indicating Read-Only data shall be set to ONE, indicating Linking data (see clause 23).
g) Layer number	bit b_{24}	shall be set to ZERO, indicating that through an entrance surface only one recording layer can be accessed.

Other settings are prohibited by this Ecma Standard.

16.2 ID Error Detection Code

When identifying all bytes of the array shown in Figure 23 as $C_{i,j}$ for i = 0 to 11 and j = 0 to 171, the check bytes for ID Error Detection code (IED) are represented by $C_{0,j}$ for j = 4 to 5. Their setting shall be obtained as follows.

$$IED(x) = \sum_{j=4}^{5} C_{0,j}\, x^{5-j} = I(x)\, x^2 \bmod G_E(x)$$

where

$$I(x) = \sum_{j=0}^{3} C_{0,j}\, x^{3-j}$$

$$G_E(x) = \prod_{k=0}^{1} (x + \alpha^k)$$

α represents the primitive root of the primitive polynomial

$P(x) = x^8 + x^4 + x^3 + x^2 + 1$

16.3 RSV

This field shall consist of 6 bytes. Their setting is application dependent, for instance a video application. If this setting is not specified by the application, the default setting shall be all ZEROs.

16.4 Error Detection Code

This field shall contain four check bytes of Error Detection Code (EDC) computed over the preceding 2 060 bytes of the Data Frame. Considering the Data Frame as a single bit field starting with the most significant bit of the first byte of the ID field and ending with the least significant bit of the EDC field, then this msb will be $b_{16\,511}$ and the lsb will be b_0. Each bit b_i of the EDC shall be as follows for i = 31 to 0

$$EDC(x) = \sum_{i=31}^{0} b_i x^i = I(x) \bmod G(x)$$

where:

$$I(x) = \sum_{i=16\,511}^{32} b_i x^i$$

$$G(x) = x^{32} + x^{31} + x^4 + 1$$

17 Scrambled Frames

The 2 048 Main Data bytes shall be scrambled by means of the circuit shown in Figure 26 which shall consist of a feedback bit shift register in which bits r_7 (msb) to r_0 (lsb) represent a scrambling byte at each 8-bit shift. At the beginning of the scrambling procedure of a Data Frame, positions r_{14} to r_0 shall be pre-set to the value(s) specified in Table 3. The same pre-set value shall be used for 16 consecutive Data Frames. After 16 groups of 16 Data Frames, the sequence is repeated. The initial pre-set number is equal to the value represented by bits b_7 (msb) to bit b_4 (lsb) of the ID field of the Data Frame. Table 4 specifies the initial pre-set value of the shift register corresponding to the 16 initial pre-set numbers.

Table 4 - Initial value of shift register

Initial pre-set number	Initial value	Initial pre-set number	Initial value
(0)	(0001)	(8)	(0010)
(1)	(5500)	(9)	(5000)
(2)	(0002)	(A)	(0020)
(3)	(2A00)	(B)	(2001)
(4)	(0004)	(C)	(0040)
(5)	(5400)	(D)	(4002)
(6)	(0008)	(E)	(0080)
(7)	(2800)	(F)	(0005)

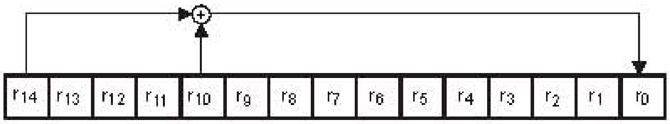


Figure 26 - Feedback shift register for generating scramble data

The part of the initial value of r_7 to r_0 is taken out as scrambling byte S_0. After that, 8-bit shift is repeated 2 047 times and the following 2 047 bytes shall be taken from r_7 to r_0 as scrambling bytes S_1 to S_{2047}. The Main Data bytes D_k of the Data Frame become scrambled bytes D'_k where

$D'_k = D_k \oplus S_k$ for k = 0 to 2 047

$\oplus$ stands for Exclusive OR.

18 ECC Block configuration

An ECC Block is formed by arranging 16 consecutive Scrambled Frames in an array of 192 rows of 172 bytes each, see Figure 27. To each of the 172 columns, 16 bytes of Parity of Outer Code are added, then, to each of the resulting 208 rows, 10 byte of Parity of Inner Code are added. Thus a complete ECC Block comprises 208 rows of 182 bytes each. The bytes of this array are identified as $B_{i,j}$ as follows, where i is the row number and j the column number.

$B_{i,j}$ for i = 0 to 191 and j = 0 to 171 are bytes from the Scrambled Frames

$B_{i,j}$ for i = 192 to 207 and j = 0 to 171 are bytes of the Parity of Outer Code

$B_{i,j}$ for i = 0 to 207 and j = 172 to 181 are bytes of the Parity of Inner Code

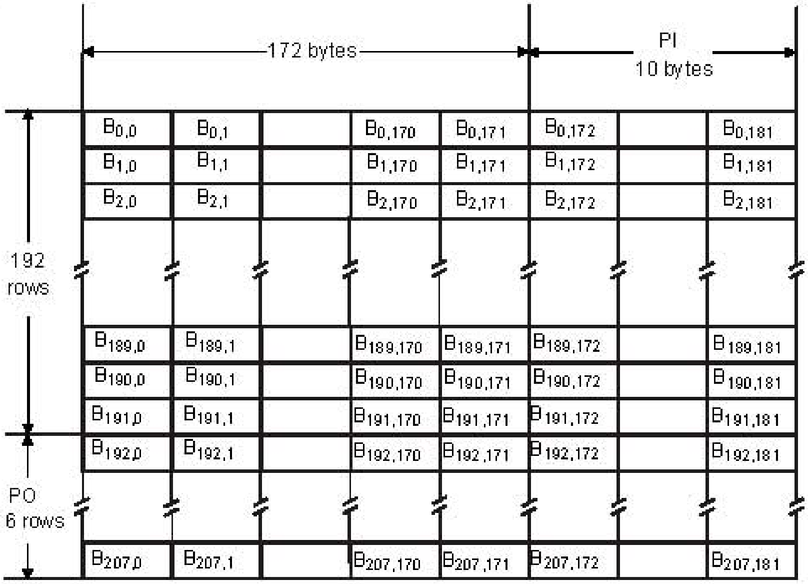


Figure 27 - ECC Block

The PO and PI bytes shall be obtained as follows.

In each of columns $j = 0$ to 171, the 16 PO bytes are defined by the remainder polynomial $R_j(x)$ to form the outer code RS (208,192,17).

$$R_j(x) = \sum_{i=192}^{207} B_{i,j}\, x^{207-i} = I_j(x)\, x^{16} \bmod G_{PO}(x)$$

where:

$$I_j(x) = \sum_{i=0}^{191} B_{i,j}\, x^{191-i}$$

$$G_{PO}(x) = \prod_{k=0}^{15} (x + \alpha^k)$$

In each of rows $i = 0$ to 207, the 10 PI bytes are defined by the remainder polynomial $R_i(x)$ to form the inner code RS (182,172,11).

$$R_i(x) = \sum_{j=172}^{181} B_{i,j}\, x^{181-j} = I_i(x)\, x^{10} \bmod G_{PI}(x)$$

where:

$$I_i(x) = \sum_{j=0}^{171} B_{i,j}\, x^{171-j}$$

$$G_{PI}(x) = \prod_{k=0}^{9} (x + \alpha^k)$$

α is the primitive root of the primitive polynomial $P(x) = x^8 + x^4 + x^3 + x^2 + 1$.

19 Recording Frames

Sixteen Recording Frames shall be obtained by interleaving one of the 16 PO rows at a time after every 12 rows of an ECC Block, see Figure 28. This is achieved by re-locating the bytes $B_{i,j}$ of the ECC Block as $B_{m,n}$ for

$m = i + \text{int}\,[i / 12]$ and $n = j$ for $i \leq 191$

$m = 13\,(i - 191) - 1$ and $n = j$ for $i \geq 192$

where int [x] represents the largest integer not greater than x.

Thus the 37 856 bytes of an ECC Block are re-arranged into 16 Recording Frames of 2 366 bytes. Each Recording Frame consists of an array of 13 rows of 182 bytes.

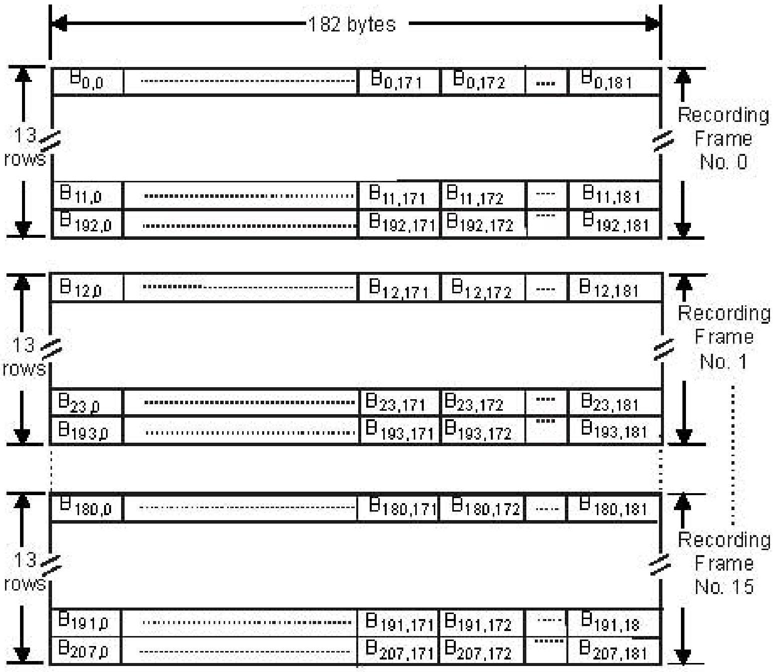


Figure 28 - Recording Frames obtained from an ECC Block

20 Modulation

The 8-bit bytes of each Recording Frame shall be transformed into 16-bit Code Words with the run length limitation that between 2 ONEs there shall be at least 2 ZEROs and at most 10 ZEROs (RLL 2,10). Annex G specifies the conversion Tables to be applied. The Main Conversion Table and the Substitution Table specify a 16-bit Code Word for each 8-bit bytes with one of 4 States. For each 8-bit byte, the Tables indicate the corresponding Code Word, as well as the State for the next 8-bit byte to be encoded.

The 16-bit Code Words shall be NRZI-converted into Channel bits before recording on the disk, see Figure 29.

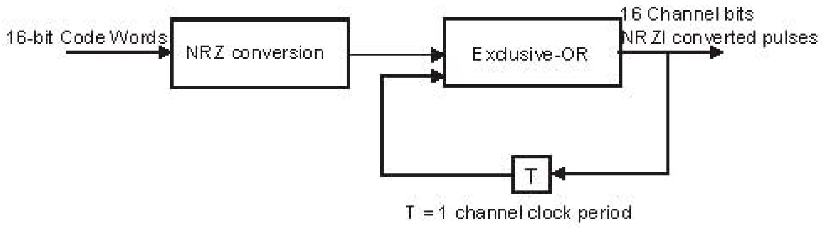


Figure 29 - NRZI conversion

21 Physical Sectors

The structure of a Physical Sector is shown in Figure 30. It shall consist of 13 rows, each comprising two Sync Frames. A Sync Frame shall consist of a SYNC Code from Table 4 and 1 456 Channel bits representing the first, respectively the second 91 8-bit bytes of a row of a Recording Frame. The first row of the Recording Frame is represented by the first row of the Physical Sector, the second by the second, and so on.

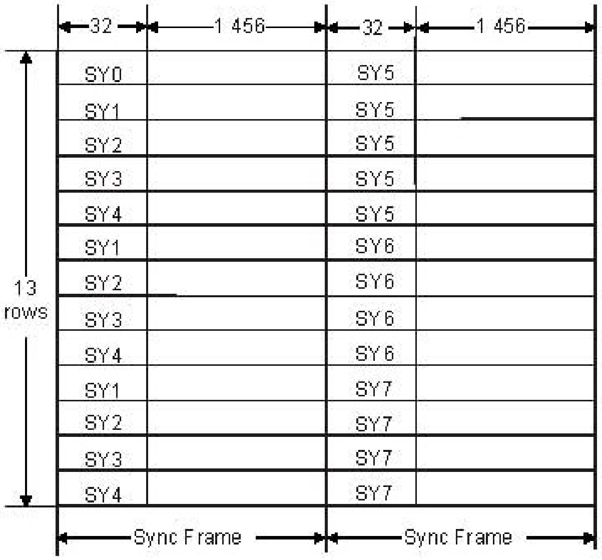


Figure 30 - Physical Sector

Recording shall start with the first Sync Frame of the first row, followed by the second Sync Frame of that row, and so on row-by-row.

Table 5 - SYNC Codes

State 1 and State 2		
	Primary SYNC Codes	Secondary SYNC Codes
	(msb) (lsb)	(msb) (lsb)
SY0 =	0001001001000100 0000000000010001 /	0001001000000100 0000000000010001
SY1 =	0000010000000100 0000000000010001 /	0000010001000100 0000000000010001
SY2 =	0001000000000100 0000000000010001 /	0001000001000100 0000000000010001
SY3 =	0000100000000100 0000000000010001 /	0000100001000100 0000000000010001
SY4 =	0010000000000100 0000000000010001 /	0010000001000100 0000000000010001
SY5 =	0010001001000100 0000000000010001 /	0010001000000100 0000000000010001
SY6 =	0010010010000100 0000000000010001 /	0010000010000100 0000000000010001
SY7 =	0010010001000100 0000000000010001 /	0010010000000100 0000000000010001
State 3 and State 4		
	Primary SYNC Codes	Secondary SYNC Codes
	(msb) (lsb)	(msb) (lsb)
SY0 =	1001001000000100 0000000000010001 /	1001001001000100 0000000000010001
SY1 =	1000010001000100 0000000000010001 /	1000010000000100 0000000000010001
SY2 =	1001000001000100 0000000000010001 /	1001000000000100 0000000000010001
SY3 =	1000001001000100 0000000000010001 /	1000001000000100 0000000000010001
SY4 =	1000100001000100 0000000000010001 /	1000100000000100 0000000000010001
SY5 =	1000100100000100 0000000000010001 /	1000000100000100 0000000000010001
SY6 =	1001000010000100 0000000000010001 /	1000000001000100 0000000000010001
SY7 =	1000100010000100 0000000000010001 /	1000000010000100 0000000000010001

The Physical Sector is a sector after the modulation by 8/16 conversion which adds a SYNC Code to the head of every 91 bytes in the Recording Frame.

22 Suppress control of the d.c. component

To ensure a reliable radial tracking and a reliable detection of the HF signals, the low frequency content of the stream of Channel bit patterns should be kept as low as possible. In order to achieve this, the Digital Sum Value (DSV, see 4.7) shall be kept as low as possible. At the beginning of the modulation, the DSV shall be set to 0.

The different ways of diminishing the current value of the DSV are as follows.

a) Choice of SYNC Codes between Primary or Secondary SYNC Codes.

b) For the 8-bit bytes in the range 0 to 87, the Substitution Table offers an alternative 16-bit Code Word for all States.

c) For the 8-bit bytes in the range 88 to 255, when the prescribed State is 1 or 4, then the 16-bit Code Word can be chosen either from State 1 or from State 4, so as to ensure that the RLL requirement is met.

In order to use these possibilities, two data streams, Stream 1 and Stream 2, are generated for each Sync Frame. Stream 1 shall start with the Primary SYNC Code and Stream 2 with the

Secondary SYNC Code of the same category of SYNC Codes. As both streams are modulated individually, they generate a different DSV because of the difference between the bit patterns of the Primary and Secondary SYNC Codes.

In the cases b) and c), there are two possibilities to represent an 8-bit byte. The DSV of each stream is computed up to the 8-bit byte preceding the 8-bit byte for which there is this choice. The stream with the lowest |DSV| is selected and duplicated to the other stream. Then, one of the representations of the next 8-bit byte is entered into Stream 1 and the other into Stream 2. This operation is repeated each time case b) or c) occurs.

Whilst case b) always occurs at the same pattern position in both streams, case c) may occur in one of the streams and not in the other because, for instance, the next State prescribed by the previous 8-bit byte can be 2 or 3 instead of 1 or 4. In that case the following 3-step procedure shall be applied.

1) Compare the |DSV|s of both streams.

2) If the |DSV| of the stream in which case c) occurs is smaller than that of the other stream, then the stream in which case c) has occurred is chosen and duplicated to the other stream. One of the representations of the next 8-bit byte is entered into this stream and the other into the other stream.

3) If the |DSV| of the stream in which case c) has occurred is larger than that of the other stream, then case c) is ignored and the 8-bit byte is represented according to the prescribed State.

In both cases b) and c), if the |DSV|s are equal, the decision to choose Stream 1 or Stream 2 is implementation-defined.

The procedure for case a) shall be as follows. At the end of a Sync Frame, whether or not case b) and or case c) have occurred, the DSV of the whole Sync Frame is computed and the stream with the lower |DSV| is selected. If this DSV is greater than + 63 or smaller than -64, then the SYNC Code at the beginning of the Sync Frame changed from Primary to Secondary or vice versa. If this yields a smaller |DSV|, the change is permanent, if the |DSV| is not smaller, the original SYNC Code is retained. During the DSV computation, the actual values of the DSV may vary between -1 000 and +1 000, thus it is recommended that the count range for the DSV be at least from -1 024 to +1 023.

23 Linking scheme

The linking scheme is specified for appending data in the Incremental recording mode. It consists of three types of linking methods named 2K-Link, 32K-Link and Lossless-Link.

23.1 Structure of linking

The appended data shall be recorded from or to the Linking sector, which is the first Physical Sector of the ECC Block and it contains the linking point.

On each linking operation, the data recording shall be terminated at the 16th byte in the first Sync Frame of the Linking sector and shall be started at the 15th to 17th byte in the first Sync Frame of Linking sector. When a disk is in the case of Figure 31 (b), Block SYNC Guard Area shall be located in the first ECC Block before linking and becomes a part of the Linking Loss Area after linking.

23.2 2K-Link and 32K-Link

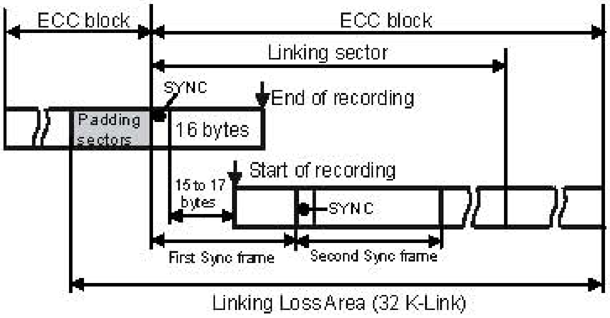
A Linking Loss Area shall be allocated in cases of 2K-Link and 32K-Link to prevent any degradation of the data reliability due to the influence of linking. It may contain padding sectors as shown in Figures 32 (2K-Link) and 33 (32K-Link) and shall have a minimum size of 2 048 bytes and 32 768 bytes respectively. All Main data in the Linking Loss Area shall be set to (00).

The Data type bit (see 16.1) of the sector followed by a sector belonging to the Linking Loss Area shall be set to ONE, but the Data type bit of the Linking sector is always set to ZERO. See Figures 32 and 33.

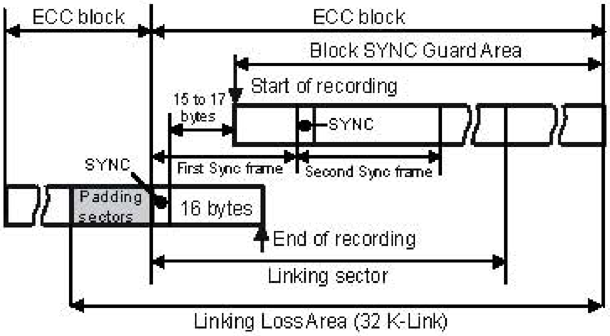
The last recorded sector in each RZone shall be recorded by using 2K-Link or 32K-Link and its Data type bit shall be set to ONE.

23.3 Lossless-Link

The linking without Linking Loss Area, as shown in Figure 34, is allowed and referred to as Lossless-Link. There is no sector which has the Data type bit of ONE in this linking scheme.



(a) Linking at just after the Recorded Area



(b) Linking at just before the Recorded Area

Figure 31 - Structure of Linking

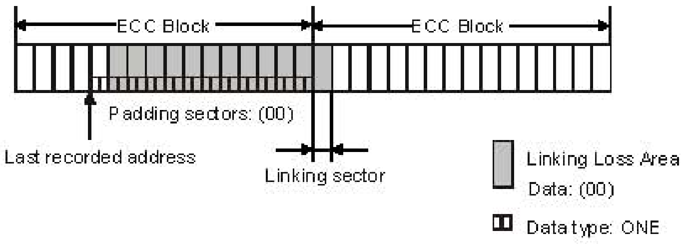


Figure 32 - Structure of ECC Block with Linking Loss Area of 2 048 bytes (2K-Link)

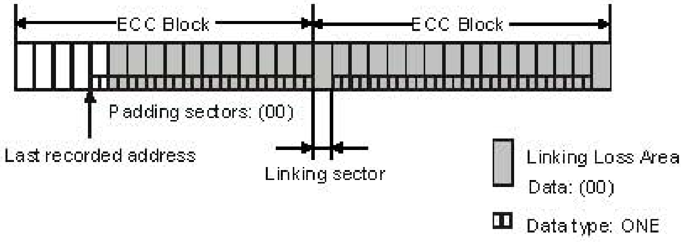


Figure 33 - Structure of ECC Block with Linking Loss Area of 32 768 bytes (32K-Link)

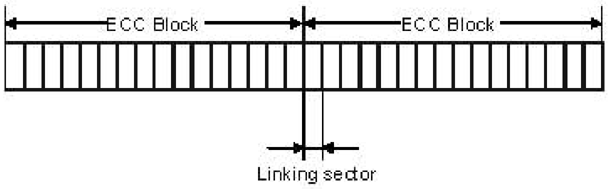


Figure 34- Structure of ECC Block without Linking Loss Area (Lossless-Link)

Appendix B – Sample SAS Script and Output

SAS script used for whole disc data

```
options ls=78 pageno=1;

PROC IMPORT OUT= WORK.jit

            DATAFILE= "C:\data\jitter-all.xls"

            DBMS=EXCEL REPLACE;

     SHEET="Sheet1$";

     GETNAMES=YES;

     MIXED=NO;

     SCANTEXT=YES;

     USEDATE=YES;

     SCANTIME=YES;

RUN;

PROC IMPORT OUT= WORK.ecc

            DATAFILE= "C:\data\errors-fulldisc-all.xls"
```

```
          DBMS=EXCEL REPLACE;

     SHEET="Sheet1$";

     GETNAMES=YES;

     MIXED=NO;

     SCANTEXT=YES;

     USEDATE=YES;

     SCANTIME=YES;

RUN;

proc sort data=jit;

by disc__ time_period;

proc sort data=ecc;

by disc__ time_period;

data both;merge jit ecc;

by disc__ time_period;

run;



proc logistic data=both descending;
```

```
model failed=Mark1st3TMean   Mark1st3TJitter Mark1st3TDev  Mark1st4TMean
        Mark1st4TJitter Mark1st4TDev


Mark1st5TMean           Mark1st5TJitter Mark1st5TDev            Mark1st6TMean
        Mark1st6TJitter Mark1st6TDev            Mark1st7TMean           Mark1st7TJitter
        Mark1st7TDev            Mark1st8TMean           Mark1st8TJitter Mark1st8TDev
        Mark1st9TMean           Mark1st9TJitter Mark1st9TDev
        Mark1st10TMean          Mark1st10TJitter        Mark1st10TDev


Mark1st11TMean          Mark1st11TJitter        Mark1st11TDev
        Mark1st14TMean          Mark1st14TJitter        Mark1st14TDev


Mark1stT        Mark2nd3TMean           Mark2nd3TJitter         Mark2nd3TDev
        Mark2nd4TMean           Mark2nd4TJitter


Mark2nd4TDev Mark2nd5TMean          Mark2nd5TJitter         Mark2nd5TDev
        Mark2nd6TMean           Mark2nd6TJitter


Mark2nd6TDev Mark2nd7TMean          Mark2nd7TJitter         Mark2nd7TDev
        Mark2nd8TMean           Mark2nd8TJitter


Mark2nd8TDev Mark2nd9TMean          Mark2nd9TJitter         Mark2nd9TDev
        Mark2nd10TMean          Mark2nd10TJitter


Mark2nd10TDev           Mark2nd11TMean          Mark2nd11TJitter
        Mark2nd11TDev           Mark2nd14TMean          Mark2nd14TJitter


Mark2nd14TDev           Mark2ndT        Mark3rd3TMean           Mark3rd3TJitter
        Mark3rd3TDev Mark3rd4TMean


Mark3rd4TJitter         Mark3rd4TDev Mark3rd5TMean          Mark3rd5TJitter
        Mark3rd5TDev Mark3rd6TMean


Mark3rd6TJitter         Mark3rd6TDev Mark3rd7TMean          Mark3rd7TJitter
        Mark3rd7TDev Mark3rd8TMean


Mark3rd8TJitter         Mark3rd8TDev Mark3rd9TMean          Mark3rd9TJitter
        Mark3rd9TDev Mark3rd10TMean
```

Mark3rd10TJitter Mark3rd10TDev Mark3rd11TMean
Mark3rd11TJitter Mark3rd11TDev Mark3rd14TMean

Mark3rd14TJitter Mark3rd14TDev Mark3rdT Space1st3TMean
Space1st3TJitter Space1st3TDev Space1st4TMean
Space1st4TJitter Space1st4TDev Space1st4TCount
Space1st5TMean Space1st5TJitter Space1st5TDev
Space1st5TCount

Space1st6TMean Space1st6TJitter Space1st6TDev Space1st7TMean
Space1st7TJitter Space1st7TDev

Space1st8TMean Space1st8TJitter Space1st8TDev Space1st9TMean
Space1st9TJitter Space1st9TDev

Space1st10TMean Space1st10TJitter Space1st10TDev
Space1st11TMean Space1st11TJitter Space1st11TDev

Space1st14TMean Space1st14TJitter Space1st14TDev Space1stT
Space2nd3TMean

Space2nd3TJitter Space2nd3TDev Space2nd4TMean
Space2nd4TJitter Space2nd4TDev Space2nd5TMean

Space2nd5TJitter Space2nd5TDev Space2nd6TMean
Space2nd6TJitter Space2nd6TDev Space2nd7TMean

Space2nd7TJitter Space2nd7TDev Space2nd8TMean
Space2nd8TJitter Space2nd8TDev Space2nd9TMean

Space2nd9TJitter Space2nd9TDev Space2nd10TMean
Space2nd10TJitter Space2nd10TDev

Space2nd11TMean Space2nd11TJitter Space2nd11TDev
Space2nd14TMean Space2nd14TJitter Space2nd14TDev

Space2ndT Space3rd3TMean Space3rd3TJitter Space3rd3TDev
Space3rd4TMean Space3rd4TJitter Space3rd4TDev
Space3rd5TMean Space3rd5TJitter Space3rd5TDev
Space3rd6TMean Space3rd6TJitter

```
Space3rd6TDevSpace3rd7TMean         Space3rd7TJitter          Space3rd7TDev
          Space3rd8TMean          Space3rd8TJitter


Space3rd8TDevSpace3rd9TMean         Space3rd9TJitter          Space3rd9TDev
          Space3rd10TMean         Space3rd10TJitter


Space3rd10TDev          Space3rd11TMean         Space3rd11TJitter       Space3rd11TDev
          Space3rd14TMean         Space3rd14TJitter          Space3rd14TDev
          Space3rdT         PIE8_Avg        PIE_Avg PIF_Avg POF_Avg     PIE8_Max
          PIE_Max           PIF_Max         POF_Max PIE8_StdDev         PIE_StdDev
          PIF_StdDev        POF_StdDev  PIEPOF /selection=stepwise;


run;
```

SAS output for whole disc data script

```
                           Model Information

                  Data Set                    WORK.BOTH
              Response Variable            Failed              Failed
                     Number of Response Levels     2
                     Model                         binary logit
                 Optimization Technique        Fisher's scoring


                  Number of Observations Read         203
                  Number of Observations Used         197


                            Response Profile

                         Ordered                  Total
                          Value       Failed    Frequency

                            1           1           18
                            2           0          179

                     Probability modeled is Failed=1.

      NOTE: 6 observations were deleted due to missing values for the response or
                              explanatory variables.
```

```
                Stepwise Selection Procedure


                 Step  0. Intercept entered:


                 Model Convergence Status

         Convergence criterion (GCONV=1E-8) satisfied.

                     -2 Log L = 120.445


                 Residual Chi-Square Test

              Chi-Square      DF     Pr > ChiSq

                165.4371     139        0.0625




                  The LOGISTIC Procedure

            Step  1. Effect Mark3rd14TJitter entered:


                 Model Convergence Status

         Convergence criterion (GCONV=1E-8) satisfied.


                   Model Fit Statistics

                              Intercept
                      Intercept         and
              Criterion          Only     Covariates

              AIC            122.445       110.166
              SC             125.728       116.732
             -2 Log L         120.445       106.166


           Testing Global Null Hypothesis: BETA=0

         Test               Chi-Square       DF     Pr > ChiSq

         Likelihood Ratio        14.2789        1         0.0002
           Score               14.0343        1        0.0002
```

```
           Wald                12.6745      1       0.0004


                    Residual Chi-Square Test

                   Chi-Square     DF    Pr > ChiSq

                     150.7144     138      0.2169


         NOTE: No effects for the model in Step 1 are removed.


                  Step  2. Effect PIF_StdDev entered:


                    Model Convergence Status

           Convergence criterion (GCONV=1E-8) satisfied.




                     The LOGISTIC Procedure

                      Model Fit Statistics

                                    Intercept
                        Intercept       and
                 Criterion       Only    Covariates

                 AIC           122.445      103.410
                 SC            125.728      113.259
                -2 Log L        120.445       97.410


             Testing Global Null Hypothesis: BETA=0

          Test              Chi-Square     DF    Pr > ChiSq

          Likelihood Ratio       23.0349      2        <.0001
            Score               19.8463      2       <.0001
            Wald                15.7787      2       0.0004


                    Residual Chi-Square Test

                   Chi-Square     DF    Pr > ChiSq
```

```
                 127.8674      137         0.6998


         NOTE: No effects for the model in Step 2 are removed.


              Step  3. Effect Space3rd3TJitter entered:


                    Model Convergence Status

            Convergence criterion (GCONV=1E-8) satisfied.




                      The LOGISTIC Procedure

                        Model Fit Statistics

                                       Intercept
                           Intercept         and
               Criterion          Only     Covariates

                AIC            122.445        96.260
               SC             125.728       109.392
              -2 Log L          120.445        88.260


             Testing Global Null Hypothesis: BETA=0

        Test                 Chi-Square       DF     Pr > ChiSq

        Likelihood Ratio        32.1851        3        <.0001
          Score                31.7449        3       <.0001
          Wald                 20.7706        3       0.0001


                    Residual Chi-Square Test

               Chi-Square       DF     Pr > ChiSq

                 173.5805      136         0.0163


         NOTE: No effects for the model in Step 3 are removed.


                 Step  4. Effect PIE8_Avg entered:
```

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

```
                 The LOGISTIC Procedure

                  Model Fit Statistics

                           Intercept
                Intercept          and
       Criterion        Only     Covariates

        AIC          122.445        89.871
        SC           125.728       106.287
       -2 Log L        120.445        79.871


          Testing Global Null Hypothesis: BETA=0

  Test               Chi-Square      DF     Pr > ChiSq

  Likelihood Ratio        40.5736       4        <.0001
    Score              42.5186       4        <.0001
    Wald               23.9279       4        <.0001


                Residual Chi-Square Test

             Chi-Square      DF     Pr > ChiSq

              120.8737      135        0.8025


 NOTE: No effects for the model in Step 4 are removed.


           Step  5. Effect POF_Max entered:


              Model Convergence Status

      Convergence criterion (GCONV=1E-8) satisfied.
```

```
                 The LOGISTIC Procedure

                  Model Fit Statistics

                           Intercept
                Intercept     and
     Criterion       Only    Covariates

      AIC         122.445      77.712
      SC          125.728      97.411
     -2 Log L      120.445      65.712


         Testing Global Null Hypothesis: BETA=0

Test             Chi-Square    DF    Pr > ChiSq

Likelihood Ratio     54.7330     5       <.0001
  Score             45.9573     5      <.0001
  Wald              22.8892     5      0.0004


               Residual Chi-Square Test

          Chi-Square     DF    Pr > ChiSq

            115.5891     134      0.8727


         Step  6. Effect PIF_StdDev is removed:


               Model Convergence Status

     Convergence criterion (GCONV=1E-8) satisfied.


                  Model Fit Statistics

                           Intercept
                Intercept     and
     Criterion       Only    Covariates

      AIC         122.445      77.222
      SC          125.728      93.638
     -2 Log L      120.445      67.222
```

```
                    The LOGISTIC Procedure

             Testing Global Null Hypothesis: BETA=0

     Test               Chi-Square       DF     Pr > ChiSq

     Likelihood Ratio       53.2224       4        <.0001
       Score              45.1693       4        <.0001
       Wald               23.8948       4        <.0001


                    Residual Chi-Square Test

               Chi-Square       DF     Pr > ChiSq

                 107.8305      135        0.9589


NOTE: No effects for the model in Step 6 are removed.


            Step  7. Effect Mark2nd9TDev entered:


                    Model Convergence Status

     Convergence criterion (GCONV=1E-8) satisfied.


                      Model Fit Statistics

                                   Intercept
                        Intercept        and
               Criterion          Only     Covariates

               AIC           122.445         68.046
               SC            125.728         87.745
              -2 Log L        120.445         56.046


             Testing Global Null Hypothesis: BETA=0

     Test               Chi-Square       DF     Pr > ChiSq

     Likelihood Ratio       64.3990       5        <.0001
       Score              46.3559       5        <.0001
       Wald               19.7913       5        0.0014
```

```
                          The LOGISTIC Procedure

                       Residual Chi-Square Test

                   Chi-Square       DF     Pr > ChiSq

                      60.5504      134         1.0000


          NOTE: No effects for the model in Step 7 are removed.


                  Step  8. Effect PIF_StdDev entered:


                       Model Convergence Status

              Convergence criterion (GCONV=1E-8) satisfied.


                         Model Fit Statistics

                                         Intercept
                              Intercept        and
                  Criterion        Only     Covariates

                  AIC           122.445         67.827
                  SC            125.728         90.809
                  -2 Log L      120.445         53.827


                Testing Global Null Hypothesis: BETA=0

        Test                 Chi-Square       DF     Pr > ChiSq

        Likelihood Ratio        66.6182        6         <.0001
        Score                   47.5286        6         <.0001
        Wald                    17.6382        6         0.0072


                       Residual Chi-Square Test

                   Chi-Square       DF     Pr > ChiSq

                      54.1478      133         1.0000


                 Step  9. Effect PIF_StdDev is removed:
```

```
                    The LOGISTIC Procedure

                  Model Convergence Status

          Convergence criterion (GCONV=1E-8) satisfied.


                     Model Fit Statistics

                              Intercept
                  Intercept         and
         Criterion       Only     Covariates

          AIC          122.445       68.046
          SC           125.728       87.745
         -2 Log L       120.445       56.046


          Testing Global Null Hypothesis: BETA=0

     Test              Chi-Square      DF     Pr > ChiSq

     Likelihood Ratio       64.3990       5        <.0001
       Score              46.3559       5       <.0001
       Wald               19.7913       5       0.0014


                  Residual Chi-Square Test

             Chi-Square      DF     Pr > ChiSq

               60.5504     134        1.0000


        NOTE: No effects for the model in Step 9 are removed.


NOTE: Model building terminates because the last effect entered is removed by
                    the Wald statistic criterion.
```

```
                                 The LOGISTIC Procedure

                             Summary of Stepwise Selection

                                    Effect                                  Number      Score
Step     Entered                    Removed                    DF        In      Chi-Square

1        Mark3rd14TJitter                                      1         1       14.0343
2        PIF_StdDev                                            1         2       12.5764
3        Space3rd3TJitter                                      1         3       9.0955
4        PIE8_Avg                                              1         4       9.1951
5        POF_Max                                               1         5       14.4519
6                                   PIF_StdDev                 1         4
7        Mark2nd9TDev                                          1         5       13.6277
8        PIF_StdDev                                            1         6       4.1996
9                                   PIF_StdDev                 1         5

                             Summary of Stepwise Selection

                      Wald                                Variable
                      Step      Chi-Square     Pr > ChiSq      Label

                      1                         0.0002         Mark3rd14TJitter
                      2                         0.0004         PIF StdDev
                      3                         0.0026         Space3rd3TJitter
                      4                         0.0024         PIE8 Avg
                      5                         0.0001         POF Max
                      6         2.3089          0.1286         PIF StdDev
                      7                         0.0002         Mark2nd9TDev
                      8                         0.0404         PIF StdDev
                      9         3.2522          0.0713         PIF StdDev


                        Analysis of Maximum Likelihood Estimates

                                                   Standard        Wald
    Parameter           DF       Estimate          Error           Chi-Square      Pr > ChiSq

    Intercept           1        0.8524            4.1920          0.0413          0.8389
    Mark2nd9TDev        1        -2.0298           0.7190          7.9692          0.0048
    Mark3rd14TJitter    1        8.4470            2.3437          12.9898         0.0003
    Space3rd3TJitter    1        -10.0500          2.9192          11.8522         0.0006
    PIE8_Avg            1        -0.00811          0.00222         13.3245         0.0003
    POF_Max             1        0.0634            0.0170          13.9456         0.0002
```

```
                         The LOGISTIC Procedure

                          Odds Ratio Estimates

                      Point              95% Wald
      Effect          Estimate       Confidence Limits

      Mark2nd9TDev      0.131        0.032        0.538
      Mark3rd14TJitter  >999.999     47.156       >999.999
      Space3rd3TJitter  <0.001       <0.001       0.013
      PIE8_Avg          0.992        0.988        0.996
      POF_Max           1.065        1.031        1.101


       Association of Predicted Probabilities and Observed Responses

  Percent Concordant   95.5   Somers' D   0.911
  Percent Discordant   4.5    Gamma       0.911
  Percent Tied         0.0    Tau-a       0.152
  Pairs                3222   c           0.955
```