NIST Special Publication 500-198

Computer Systems Technology

U.S. DEPARTMENT OF COMMERCE
National Institute of Standards and Technology

NIST



# Monitoring and Reporting Techniques for Error Rate and Error Distribution in Optical Disk Systems

Fernando L. Podio

NIST Special Publication 500-198

# Monitoring and Reporting Techniques for Error Rate and Error Distribution in Optical Disk Systems

Fernando L. Podio, Editor

Proceedings of a Workshop
Held in Colorado Springs, Colorado
on August 5, 1991

Advanced Systems Division
Computer Systems Laboratory
National Institute of Standards and Technology
Gaithersburg, MD 20899

October 1991

**U.S. DEPARTMENT OF COMMERCE**
**Robert A. Mosbacher, Secretary**
NATIONAL INSTITUTE OF STANDARDS
AND TECHNOLOGY
John W. Lyons, Director

## Reports on Computer Systems Technology

The National Institute of Standards and Technology (NIST) has a unique responsibility for computer systems technology within the Federal government. NIST's Computer Systems Laboratory (CSL) develops standards and guidelines, provides technical assistance, and conducts research for computers and related telecommunications systems to achieve more effective utilization of Federal information technology resources. CSL's responsibilities include development of technical, management, physical, and administrative standards and guidelines for the cost-effective security and privacy of sensitive unclassified information processed in Federal computers. CSL assists agencies in developing security plans and in improving computer security awareness training. This Special Publication 500 series reports CSL research and guidelines to Federal agencies as well as to organizations in industry, government, and academia.

**National Institute of Standards and Technology Special Publication 500-198**
**Natl. Inst. Stand. Technol. Spec. Publ. 500-198, 89 pages (Oct. 1991)**
**CODEN: NSPUE2**

**U.S. GOVERNMENT PRINTING OFFICE**
**WASHINGTON: 1991**

For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, DC 20402

# Preface

This report constitutes the proceedings the workshop on Monitoring and Reporting Techniques for Error Rate and Error Distribution in Optical Disk Systems held on August 5, 1991, in Colorado Springs, Colorado. The workshop was the first step in a Computer Systems Laboratory of the National Institute of Standards and Technology (CSL/NIST) program of work to identify Federal Government needs and the state of the art on error rate monitoring and reporting techniques of optical disk systems. This program is aimed towards promoting discussions on possible future optical disk drive implementations and to promote consensus among drive manufacturers in the way of reporting error rate information through an interface.

Thirty-five participants from Government and industry attended the workshop. Three members of different agencies in the Federal Government gave presentations. Industry speakers included two representatives from error correction and detection (ECC) chip manufacturers and two representatives from drive manufacturers. The workshop included discussions on (a) Government needs on error detection and correction, (b) error measurement capabilities on current logic chips, (c) error management strategies, (d) error recovery techniques, (e) error reporting capabilities of current drives, and (f) plans for a Government/industry working group.

There was a panel discussion during the afternoon session. A preliminary set of user requirements was identified. These are as follows: (a) Report on the consumption of spare sectors and the rate of change, (b) Report correction (on record base) rate above some threshold (corrections/capacity) at the codeword level, (c) Capability to display the uncorrected sector content at the host level, (d) Capability of reporting all soft errors on demand, (e) Capability of reporting all hard errors on demand, and (f) Report on errors encountered reading header information.

Discussions took place on how to reach consensus on reporting error rate information to the host to satisfy the user's requirements. It was proposed that a working group or similar structure sponsored by CSL/NIST be organized. This idea was well received by the participants. It was also suggested that the approach to achieve consensus on standardized method(s) of reporting error rate information to the host should be the documentation of a Small Computer Systems Interface (SCSI) common command set for error rate monitoring and reporting. This suggestion was also well received by the majority of the participants.

Specific vendors and commercial products were cited during the workshop by speakers and participants. The inclusion or omission of a particular company or product does not imply either endorsement or criticism by NIST.

I express appreciation to the speakers for their input to the workshop. I gratefully acknowledge also the assistance of all those who made the workshop a success.

Fernando L. Podio, Editor

# Contents



(*) Reproduced with permission from SPIE - The International Society for Optical Engineering. Presentation given by W. Brown.

# Monitoring and Reporting Techniques for Error Rate and Error Distribution in Optical Disk Systems


## Abstract

This report constitutes the proceedings of the workshop on Monitoring and Reporting Techniques for Error Rate and Error Distribution in Optical Disk Systems held on August 5, 1991, in Colorado Springs, Colorado. The objectives of this workshop were to identify the state of the art on error rate monitoring and reporting techniques in optical disk systems and to promote discussions on possible future implementations.

The workshop presentations included the description of a Computer Systems Laboratory of the National Institute of Standards and Technology (CSL/NIST) program for investigating error reporting capabilities of optical disk drives, Federal Government needs on error detection, correction and reporting, and the state of the art on error reporting capabilities in current generation drives. Presentations also included the description of the capabilities of error correction and detection current chips and discussions on error management strategies.

The participants noted the need for tools for reporting error rate activity through interfaces with the host and emphasized the importance of this subject to data managers in the Federal Government.

During a discussion panel, the participants identified a preliminary set of user requirements and it was suggested that a Government/industry working group be organized in order to document a consensus position between the the Federal Government and industry in error reporting capabilities for future generation drives.



Keywords: error detection and correction chips; error recovery techniques; optical disks, error detection and correction; optical disks, error distribution; optical disks, error management techniques; optical disks, error rate.

# The NIST Program for Investigating Error Reporting Capabilities of Optical Disk Drives

Fernando L. Podio

Computer Systems Laboratory
National Institute of Standards and Technology

Federal Government data managers and many private companies are already using optical digital data disks for long term storage of valuable data. Optical disk drives are designed with strong but not unlimited error corrections capabilities. If the level of errors overcome the error detection and corrections mechanisms implemented in the optical disk drive controllers, uncorrectable errors may occur. Most of the current optical disk systems do not provide the host with sufficient error rate information. Federal Government data managers are interested in being able to monitor error rate activity in optical disk drive systems. This could be achieved through the interface, if the drives would provide information on correctable error rate activity, such as maximum number of correctable errors, maximum number of errors per interleave, and the location of those errors. Reports on errors encountered while reading header information is also necessary. This information would provide data managers with a better understanding of the status of their data and would allow them to design more efficient recopying policies to copy the data to similar or different media in a timely and economic fashion.

The Computer Systems Laboratory of the National Institute of Standards and Technology (CSL/NIST) organized a program of work to identify Federal Government needs and the state of the art on error rate monitoring and reporting techniques of optical disk systems to the host. The first stage of this program was to organize a workshop to identify the state of the art on error rate monitoring and reporting techniques and promote discussions on possible future implementations. CSL/NIST also proposes the organization of a working group to discuss user requirements, to achieve consensus between the Federal Government and industry on standardized method(s) of reporting error rate information to the host through an interface, and to document these methods. Another stage of the CSL/NIST program is to implement a demonstration platform at NIST to show the Federal Government users the state of the art on error rate monitoring and reporting techniques as well as the interface capabilities. The platform will consist of two hosts, a SUN SPARCstation IPC and a DOS engine. When drives become available with the recommendations achieved during the working group, the demonstration platform will be adapted to accommodate this next generation of drives with enhanced error reporting methodologies. Any information that the working group may develop that may be used for the next generation standards (such as the Small Computer Systems Interface (SCSI) standard will be provided to the relevant standards committees.

# The NIST Program for Investigating Error Reporting Capabilities of Optical Disk Drives

Fernando Podio
Project Leader for
Optical Storage Research
301/975-2947

National Institute of Standards and Technology
Computer Systems Laboratory

# Optical Disk Drives Error Reporting Considerations

* Most optical disk systems do not report sufficient information to the host on error rates, error distributions, etc.

* Error rate information would include parameters such as the total number of correctable errors, maximum number of errors per interleave, and the location of these errors.

* This information would provide data managers with a better understanding of the status of their data.

* Monitoring the status of data recorded on optical media is particularly important because of its anticipated use for the long-term storage of valuable data.

# Specifics of the NIST Program

* Workshop
* Government/industry working group
* Demonstration platform
* Participation in the relevant standards committees

# NIST Workshop

* Federal Government needs.
* Identify the state of the art on error rate monitoring and reporting techniques.
* Promote discussions on possible future implementations.
* Organization of a Government/industry working group.

## Government/Industry Working Group

* Technical representatives from Government and industry.
* Document standardized method(s) of reporting error rate information (NIST guideline/standard).
* Contributions to relevant standards committees.
* Meetings concurrently with TC X3B11 (Optical Digital Data Disks).
* First meeting: October 7, 1991, Phoenix, AZ.
* Target date for NIST publication: end of FY92.

# Demonstration Platform

* Demonstration of the state of the art on error rate monitoring and reporting techniques of optical disk drives to U.S. Government managers.

* Two hosts:

  SUN SPARCstation IPC
  Compaq 386/25e

* Optical disk drives

* Adapt the demonstration platform to accommodate next generation drives with enhanced error reporting methodologies.

# Standards Commitees

* TC X3B11 (Optical Digital Data Disks)
* TC X3B11.1 (Label and File Structures)
* TC X3T9.2 (Lower Level Interface)

# DATA MANAGEMENT IN NOAA

William M. Callicott

August 5, 1991

ABSTRACT

NOAA has 11 terabytes of digital data stored on 240,000 computer tapes. There are an additional 100 terabytes (TB) of geostationary satellite data stored in digital form on specially configured SONY U-Matic video tapes at the University of Wisconsin. There are over 90,000,000 non-digital form records in manuscript, film, printed and chart form which are not easily accessible. The three NOAA Data Centers service 6,000 requests per year and publish 5,000 bulletins which are distributed to 40,000 subscribers. Seventeen CD-ROMs have been produced. Thirty thousand computer tapes containing polar satellite data are being copied to 12 inch WORM optical disks for research applications. The present annual data accumulation rate of 10 TB will grow to 30 TB in 1994 and to 100 TB by the year 2000. The present storage and distribution technologies with their attendant support systems will be overwhelmed by these increases if not improved. Increased user sophistication coupled with more precise measurement technologies will demand better quality control mechanisms, especially for those data maintained in an indefinite archive. There is optimism that the future will offer improved media technologies to accommodate the volumes of data. With the advanced technologies, storage and performance monitoring tools will be pivotal to the successful long-term management of data and information.

TABLE OF CONTENTS

## 1.0. Data Management in NOAA

Management of environmental data and information resources is becoming an increasingly visible and important issue for the scientific community. This is of particular importance to the National Oceanic and Atmospheric Administration (NOAA), which routinely measures and collects large amounts of environmental data and information in its own work, and is also officially charged with maintaining environmental records for the Nation. Through its activities over time, NOAA has become the steward of a treasury of Earth systems data and information--the most comprehensive, long-term, and up-to-date environmental description of the Earth that exists today. This treasury contains answers to urgent environmental questions facing the Nation. The success of all NOAA's scientific work, and the national priorities which it supports, depends on the accountability and accessibility of environmental data and information. These data and information must be accurate, complete, stable and fully-integrated across the spectrum of NOAA's organizational functions, and they must be made easily accessible, in a timely and cost-efficient way. Throughout this effort to provide resources to meet the NOAA missions of managing data for global change research and for enhancing warning and forecast services, there will be a continual inherent process to migrate and protect data held in the NOAA archives.

## 2.0. NOAA Data Management Operations:

The NOAA Centers respond to requests from a broad community of research, legal, engineering, individuals, insurance, business, consultants, and manufacturing. About 6,000 requests per year are received for digital data. Within the next two years as global change research increases, this should grow to 9,000. The current data files stored on computer readable media have a volume of 11 terabytes stored primarily on 240,000 computer tapes. There are 100 TB of GOES data at the University of Wisconsin copied in digital form on specially recorded SONY U-matic video tapes. There are analog, i.e., non-digital, holdings having a volume equivalent to over 50 TB in a digital domain. With the conversion of some of the analog data to digital format, and with new sources of digital data, the digital holdings will grow by 1996 to about 35 TB not counting source level data from GOES. The rapidly expanding quantity of data will force a different approach to managing data and information within the centers. The use of an integrated mass storage systems with appropriate file management will be essential to manage the archive, storage hierarchy and data migration processes. New mass storage hardware and software technologies will have to be developed and adopted and the current archives copied. If a mass storage system is not developed, the data centers will require immense tape storage areas and reduce the access mechanism from data granules to physical volumes of data.

In accordance with the NASA/NOAA Memorandum of Understanding for remotely sensed earth observations, data processing, distribution, archiving and related science support, EOS data is intended to be archived at the NOAA data centers. The EOS data from prototype operational instruments used for NOAA operational purposes, will be an inherent part of NOAA's data archives. As part of the EOS pathfinder activity, NOAA is migrating the environmental satellite data from the operational polar orbiting satellites from computer tapes to 12 inch SONY optical write once, read many (WORM) disks. Also, there are selected special sensor microwave data from the Air Force Defense Meteorological Satellite Program (DMSP) of interest to EOS scientists. Initially, 8 terabytes of data will be migrated to optical disk.

The computing capacity at the centers is provided by mainframes, workstations, and personal computers (PCs). In the aggregate, the systems do not have the capacity to handle the anticipated archive, the growth in data ingest and dissemination, and expanded quality control, analysis and reprocessing requirements in the coming years. The computational capacity will need to grow from 10 MFLOPS at the beginning of 1991 to over 300 MFLOPS by 1996, i.e., a factor of 30:1. On-line disk storage should grow from 60 GB to over 500 GB during this period. The configurations are evolving from a central processor surrounded by dedicated terminals to a fully distributed client-server architecture which can expand in response to workload demands.

2.0. NOAA Data Centers: There are three National Data Centers and over a dozen centers of data in NOAA. The data centers are structured as formal archive centers and serve at this Nation's world data centers for their respective disciplines. The following provides a description of the centers and their activities.

2.1. The National Climate Data Center (NCDC) in Asheville, North Carolina was established in 1950. The National Archives and Records Service, in compliance with the Federal Records Act of 1950, specified that NCDC's climatological records be permanently retained. It has been designated as a World Data Center (WDC)-A for Meteorology. It also operates the Satellite Data Service Division which manages the high volume satellite data. The Center is responsible for ingest, quality control, archiving, managing, providing user access, and performing analysis of data which describes the global climate system. The NCDC also supports major new programs such as the National Weather Service Modernization, Climate and Global Change, the Coast Watch Initiative, and the Level-0 Earth Observing Systems (EOS) path finder effort. In 1992, new data sources from foreign satellites will be introduced. The center has a staff of 290 full time employees (FTEs), 100 contract and 190 federal employees. Each working day results in 160 orders from 360 user contacts. Annually, about 5,000 bulletins are prepared and supplied to 40,000 subscribers. Much of NCDC's data (by physical volume) is in the form of paper records such as ship logs, and although manually accessible, is not readily usable, and they appear to be deteriorating. There are fifty thousand cubic feet of paper records at NCDC. There are also film and other non-machine readable information stored in the National Archives.

NOAA has a contract with the University of Wisconsin to collect and archive data from the NOAA geostationary operational satellites (GOES). The GOES data collected from 1978 to the present, is recorded and stored at the University of Wisconsin's Space Science and Engineering Center. To date, the GOES data are stored on 25,000 Sony U-matic beta video (19mm commercial video standard) video cassettes in 4 GB increments. Access to the data is not highly efficient because some of the cartridges are offsite in the state records office, and because the data must be reingested as a satellite readout. The GOES archive represents the largest amount of data anywhere in the NOAA system to be rescued and be made more readily accessible. To improve access, the center is engaged in a pathfinder study on mechanisms to improve the access to the data.

2.2. The National Geophysical Data Center (NGDC) in Boulder, Colorado was established in 1972. Its mission is to manage solid earth and marine geophysical data as well as ionospheric, solar and other space environmental data; and to provide facilities for World Data Center-A for Geophysics which encompasses Solid Earth Geophysics, Solar-Terrestrial Physics, Marine Geology and Geophysics, and Glaciology. The center has a staff of 60 FTEs.

NGDC has over 300 databases including 54 million (M) Ionograms, 2.5M magnetograms, 12 million flight miles of aeromagnetic data and 10 million miles of ship track data. There are 25,000 magnetic tapes partly at NGDC in Boulder and partly in Asheville at NCDC. About 2000 tapes per year arrive from originators outside of NGDC. Their goal is to keep all of the ingest tapes and maintain two additional copies for use in normal center activities (3 copies total). About 14,000 tapes have no backup. There is a requirement to copy 12,000 tapes a year for routine migration. NGDC relies on the error checking provided by the tape copying system and supplements this with printouts of beginning/end of record data and record counts.

NGDC began the NOAA CD-ROM program four years ago. Its first CD-ROM title was "Geophysics of North America". The center now has a total of 12 CD-ROMs completed or underway. This effort should change the distribution system for data from its tape orientation and probably deflect some use of the network to obtain data. During 1990, distribution of data by CD-ROMs has increased both the number of requests for digital data and the annual amount of data distributed by the center. The distribution of data by CD-ROM puts the data into a form highly convenient and useful to PCs and workstations.

The National Snow and Ice Data Center (NSIDC) under contract to NGDC operates the World Data Center-A for Glaciology. The role of NSIDC is to acquire, archive and disseminate data relating to all forms of snow and ice. It provides data to about 500 requesters per year from a digital archive data base of about 15 GB (300 standard tapes); 7GB are from the NIMBUS-7 Scanning Multi-channel Microwave Radiometer, and 3 GB are Special Sensor Microwave/Image (SSM/I) data. Many of the NSIDC datasets are redundantly held at other NOAA data centers. A daily volume of 1 GB of data from the Defense Meteorological Satellite Program (DMSP) Operational Linescan System (managed by NGDC) will be forwarded from the DMSP readout site to NSDIC on EXABYTE tape. A similar means is being developed to distribute a weekly data volume of 800MB between the Joint Ice Center (JIC), in Suitland and the National Snow and Ice Data Center (NSIDC) in Boulder. The NSIDC has been designated as an EOSDIS Distributed Active Archive Center (DAAC). As such, it will build up its computational and archival ability to meet EOSDIS defined mission goals. As a DAAC, the NSIDC will relocate to the University of Colorado campus.

2.3. The National Oceanographic Data Center (NODC) has been in operation since 1961 as an interagency, facility under the U.S. Navy, and became a part of NOAA in 1970. Its mission is to manage oceanographic data. It has operated as a World Data Center-A (WDC-A) for Oceanography since 1962. NODC has a staff of 85 FTEs. NODC's files include data collected by U.S. federal agencies; state and local government agencies; universities and research institutions; foreign government agencies and institutions; and private industry. Currently, NODC maintains a digital archive of both in-situ and satellite-sensed ocean data in excess of 30 GB. A potential equivalent of 10 GB of digital data are currently maintained in analog form such as data reports, manuscripts, and analog instrument recordings. With the establishment of NODC data management responsibilities for ocean observing satellites, including non-NOAA geostationary and orbiting platforms; new global collection efforts, including the Global Ocean Flux Study (GOFS), the World Ocean Climate Experiment (WOCE), and the Climate and Global Change Project; and new U. S. coastal ocean studies, including CoastWatch and the Coastal Ocean Program, the archive is expected to increase twenty fold between FY90 and FY95.

NODC has a large amount of analog data. A tablet digitizer is used to annually process 10,000 expendable bathythermograph traces (XBT) to 2 MB of data. NODC has a contract with the Navy to annually process 100,000 similar traces. Mechanical bathythermographs (MBT) on glass slides (300,000) await conversion and are expected to result in 1.5 GB of data. Acoustic Doppler Profiling, done from University and NOAA ships, is expected to be a new source of data. There are perhaps 20 ships that may be equipped with these profilers. At the present, only a few are capturing the data for archiving purposes. Each profiler should provide 10 MB per ship month. In the future, 100 ship months per year of these data could be archived.

The NOAA Coastwatch Data Management, Archive, and Access System (NCAAS) now under development will result in expansion of the archive, related quality control, and retrieval and distribution activities based on SONY 12 inch WORM Optical Disks. NODC also is responsible for the NOAA Library and there is interest in digitizing some of its data and metadata holdings.

NODC is responsible for the NOAA Earth System Data Directory, and interfacing it with the larger NASA based master directory effort. This is part of the catalog interoperability effort which is underway among other government agencies and foreign data centers. The master directory is also needed to fit with the EOSDIS version 0 effort where the catalog interoperability will perform relevant IMS (Information Management System) functions. NOAA's master directory is VAX based, with specially written software, and the ORACLE Data Base Management System.

A prototype database system has been developed to provide NODC users with direct access to an on-line, interactive data archive. It maintains a data base of over 23 million marine observations from 310,000 ocean stations. Access to the data is obtainable through spatial or temporal searches with arbitrary combinations of instruments, platforms, and parameters. By FY 1993, NODC plans to add all of its vertical profile data (Nansen, Bathythermograph, C/STD, etc.) to the POSEIDON system.

2.4. NOAA Centers of Data include those data collection and operations elements performing observations and monitoring services as part of NOAA's recurring mission responsibilities. Listed below are the principal centers:

| Discipline | Title | Location |
|---|---|---|
| Bathymetry, Nautical charts, Geodesy | Charting and Geodetic Services | Rockville, Maryland |
| Climate | Climate Analysis Center | Camp Springs, Maryland |
| Fisheries | National Marine Fisheries Service | Seattle, WA; Woods Hole, MA; Miami,FL; Bay St. Louis,MS; San Diego, CA |
| Ice | Navy Joint Ice Center | Suitland, Maryland |
| Lake Data | Great Lakes Environmental Research Lab | Ann Arbor, MI |
| Oceanography | Center for Ocean Analysis and Prediction | Monterey, CA |
| Oceanography | Ocean Products Center | Suitland, Maryland |

| | | |
|---|---|---|
| Pacific Ocean Data | Equatorial Pacific Information Collection | Seattle, WA |
| Particle Deposition Data | Air Resources Lab | Silver Spring, MD |
| Sea Level | University of Hawaii | Honolulu, HI |
| Snow and Ice | National Snow and Ice Data Center | Boulder, CO |
| Tides | National Tide and and Water Level Data Base | Rockville, MD |
| Trace Gases | Global Monitoring for Climate Change | Boulder, CO |

## 3.0. Digital Data Request History and Projection

The support for global change by the NSF has increased about 35% per year since 1987 and is expected to increase for FY 1992. NOAA's global change funding has roughly doubled each year since 1989. Other agencies are also increasing their global change funding. The overall funding for all agencies has increased seven times from FY89 to 1991. From this, one could expect a large increase in the number of data requests at the centers. However, in the aggregate, there has only been a modest increase in the number of requests for each of the last two years ('89 and '90), and the projections are, therefore, based on this modest rate of increase. Another view is that the secondary distribution of data from scientist to scientist may be on the increase because of the ease of transmission over networks, coupled with a desire to obtain a dataset that has had use in a familiar science project. Possibly this secondary distribution masks the size of the actual data dissemination.

The biggest impact on the number of requests and volume of data distributed has been from the introduction of CD-ROMs. This indicates that the increased use of CD-ROMs for data distribution provides a means for rapid deployment of the data among members of the research community. Secondary distribution of data from scientist to scientist may be on the increase because of the ease of transmission over networks, coupled with a desire to obtain a dataset that has had use in a familiar science project.

## 4.0. Mass Storage/File Management Requirements

As the amount of data increases and the NOAA mission focuses on improving accessibility of data for global change research, there is an urgent requirement to develop mass storage systems with file management software at the centers to improve archive management, provide vastly improved access mechanisms, and to reduce the amount of media and associated space requirements. Moreover, the mass storage system is the heart of a file management system. For the immediate purpose at hand, a mass storage system should include the following:

### 4.1. Hierarchical File Director

A hierarchical file directory is needed that permits, as a minimum, the acceptance of UNIX file names. The directory needs to maintain the access and update history information for the file. The directory should allow for handling mixed media within a single search, for cross indexing between devices, and for recording data set utilization records for future knowledge based system applications. This directory must interoperate with a number of different data base systems passing query information through during interoperable data searches.

### 4.2. A hierarchy of storage levels

The mass storage system should support a hierarchy of storage levels with increasing physical capacity and decreasing performance at the bottom, and decreasing physical capacity and high transfer rates at the top (as viewed from the user client processes). At the top, this permits the evolution to direct access electronic storage, so that the mass storage becomes a truly integrated part of a computing environment.

### 4.3. File Management Software

The file management software should offer options for data compression. It should permit the use of checksums as an overall error control mechanism. Data conversion software should be available to migrate the data from one physical media to another, as generic files, without disturbing the data content. The software would sample files on a statistical basis reading them to verify that they were still intact and that the media had not deteriorated beyond the point where only soft data checks were obtained. In the event that sufficient degradation was detected during this sampling process, the files would be migrated to new media with a corresponding directory update. Migration would also be triggered during normal accesses whenever too many soft errors occur.

Migration of files from archival or working media to a buffer storage area would occur following the initial access to permit data to be more rapidly retrieved from the faster devices in the storage hierarchy. The migration and actual location of the data should be presented to the user/application programs in a transparent manner including, as an option, presentation to client processes in a manner simulating direct retrieval from the ingest media if desired. To accomplish this transparency, the file management software should provide for the ingest of data from existing media and distributed to: standard half-inch magnetic tapes in all densities and formats, EXABYTE, DAT, CD-ROMs, optical disks, video cassette recording technologies, digital optical media, etc. An encapsulation of the ingested media's data should record the presence and location of permanent data errors, physical record lengths in bytes, the presence and location of ingest media specific flags, such as tape marks, end of tape flags, etc., so that upon access, the data can be handed to processing programs that need to be aware of the different media. The file management software should be able to handle any of the existing data formats and to invoke conversion routines to a standard if one is adopted. The ability to move files from the mass storage to a requester's media in the original format should be provided.

## 4.4. Network Access and Networking

The NOAA centers should be coupled to the internet, at internet backbone rates, and eventually to the National Research and Education Network (NREN). With 740 universities, laboratories and industrial sites now on the network, and 75 more expected in the 2nd half of 1991, the internet is widely available to the scientists involved in global change and EOS. There are 16 NSFnet backbone sites. Two of these, the University of Maryland and NCAR, are in close proximity to NOAA data centers. Where large data volume data transfers are required, conventional conveyance services would probably suffice with economic considerations determining the mode of conveyance.

## 5.0. Considerations

The system life under the NOAA mandate to manage data for long-term global change research purposes is open ended. The value of data increases with age for use in performing long term environmental change research. Global patterns are known to be subtle over time, even when viewed in a rapidly changing environment. Today's collection of environmental data is pitifully small and of too short a duration compared to the amount required to filter out the statistical noise over an extended time domain.

The operational requirements are influenced by incremental science requirements established as the knowledge of the relationships between instrument responses and conversion to physical units became better known from the results of research and development of more sophisticated processing algorithms. Because the development of sufficient knowledge to fully understand the earth observation responses is an accretive and repetitive process, the entire data set will require repeated reprocessing to adequately describe the data for long-term documentation and preservation.

Aggregation of similar but different and sometimes disparate data types is also an important feature to include. A well understood aggregation principle will allow for compartmenting the data across the media domain in a "most" convenient form for vastly improving the access mechanisms. This will become increasingly important as the longevity of the archive increases. Aggregation implies some degree of redundancy, but in a positive sense, in that redundancy of particularly critical data sets reduces the risk of data loss over time.

Volume management may require compression mechanisms to reduce the over-volume of data as it ages. Critical data and information will require the application of lossless data compression where data sets are reduced in volume. When permitted, other means can be used to reduce data volume with controlled data loss as achieved through sampling, or through small-loss data compression techniques or a combination of the techniques to yield much higher compression ratios. This may particularly attractive for managing very high volume image data sampled in the visible spectra.

The technological gallop of the last several years continues to accelerate and new storage and processing technologies are obviating the need to consider destruction of cumbersome data through full scene sampling, scan sampling or reduction to gross descriptive parameters which describe the sum of the parts in abbreviated form. In order to take advantage of improved and less costly storage technologies, there is a plan to migrate the data periodically as the volumes dictate and the technology allows and with each migration to yet developed capabilities, it becomes even more feasible to consider placing all of the data in a near-line access environment. The migration process will require content processing to re-develop the cross reference inventory information to include additional content description information as part of the inventory metadata file to increasingly document the data as it ages. Data migration is an essential element in developing a never-ending data life for the sake of offering an extended time baseline data set essential for detecting global scale changes.

A wide variety of media will be used for distributing data and information to users. The large number of formats used by the NOAA data centers means that many conversion procedures will be needed. It would be better not to convert the data in the archives themselves, but to write procedures which can be invoked in a demand fashion. In this way, the data can be left in its original form while confidence is gained in the accuracy of the conversion. If any problems arise in the converted data, the original data will not be contaminated. The problem becomes one of reworking the conversion routine and alerting previous users of the defect rather than trying to fix a partially scrambled dataset. The downside is that there will be an additional processing cost when the data is requested. Another problem with the standard data format concept is that many researchers who submit data to the data centers will probably never conform to a standard format. Insistence on a format may become an impediment to releasing the data to a center for distribution.

## 6.0. Assumptions and Constraints

Factoring today's technology advancement timescale for the purpose of being both realistic and conservative, the period of migration to exploit new technologies and avoid system obsolescence to extend the validity of the data and information is established at no less than every 10 years. A suggestion was recently made that the migration rate be a function of the expected half life of the medium used. For the 3480 tapes, the manufacturers agree that 10 years of full performance life should be expected, thus the half life for migration purposes would be five years. The criteria for accepting a new technology as a candidate for data migration is; improved archival qualities, the per data byte storage cost must be one half the previous, the physical storage requirements be at least five times less, and the data transfer rate to move the data from the media be no slower than that of the older media. And, finally, the data migration step will include data processing to derive or extend content description data to be used for the purpose of validating the preservation state of the data and for reinventoring the data to add content information developed through the accretion of user knowledge and experience.

Another assumption is that all of the data will be reprocessed three times during a 25 year cycle. This reprocessing cycle may coincide with a data migration step since all of the migration will include a content review during the passage of data from one medium to another. The development of reprocessing algorithms will not be charged as a data management system task but will require that the data management system be able to put significant quantities of data on-line or at-hand for "live" ingest mode processing.

## 7.0. Conclusion

To continue managing data as we do today would eventually require a facility to accommodate an enormous number of media units to hold the data volumes projected for the future. As the new data continues, the added function of migrating data from degenerating media (from a systems as well as physical viewpoint), will compound the storage requirements as the annual data volume accumulates by the hundreds of trillions of data samples each year. If acceptable mass store systems are not continually developed to match the data growth and data management requirements, the logistics would be overwhelmed and the system would fall apart never to be recovered again because of the enormous cost to recover an inevitable backlog.

The data only has value to the research community if it is conveniently and efficiently accessible. If the data were placed in a warehouse environment, which would ultimately have to happen if nothing was done, it would soon lose its value and possibly its identity because of the cost to acquire it and eventually would be lost because of the huge cost to locate, ship back, copy and return the data copies as the data volume grows beyond manageable proportions. This is aside to the issue of data loss due to media deterioration. The only acceptable solution would be through development of a system capability to provide highly efficient and sophisticated data management capabilities which would accommodate online data sources. In order for this to happen, advanced media technologies have to be employed along with advanced sophisticated data management software to eliminate the manual interfaces where possible to provide the data in a ready mode for user interaction at the subsetting level. The data value increases dramatically when placed in this type of environment as the access to the system can provide instant gratification and encourages repeated and expanded data query activities. This, in turn, accelerates the research progress and enhances the research results thus broadening the value of the data to the advancement of science and knowledge.

To physically compress the data through the implementation of high capacity media coupled with the systems capability to control and index these data in an online or near-line environment offers a significant reduction in the requirements to house the archives both in terms of physical space and recurring energy and labor expenses. The closer on line the data are placed, the less labor is required to physically mount data either in the appropriate archive slot or onto the processing system. As electronic access become more fully integrated into the system, direct labor service categories will be eliminated. A major cost avoidance to be reckoned with is the cost of adding increasing large physical facilities as the data volume grows at the projected rates. The pace of implementation of new technologies should allow shrinking of the space requirements to match the increase of data accumulating in the facilities. An indirect benefit of space compression through improved storage technologies would be realized form compressing the facilities requirements sufficiently to consider replicating the data in distributed locations as a risk reduction measure.

The broadest benefits are in terms of what the value of the data is to the world of environmental change. Without a responsibly managed data record of scientific measurements over time, there would be no baseline to objectively determine if the environment we live in is really changing, how much, and at what rate. Without these data, economies would be based on subjective opinions and in some cases, hear say. Public policy would more often than not be misguided and consequences of enormous proportions could occur to our physical well being either through economic collapse or through direct physical changes. Even today, global change scenarios portend potential devastating effects to our coastal cities and this country's agro-economies. But, do we build dikes and seek alternate water sources if we are not really sure what, if any, impacts there are? Without the data, no one knows, so any investment in mitigating a potential problem is an economic risk. The other question is; even if we know, can we afford to take avoidance action? Or better yet, is the cause due to environmental causes or due to a much broader cyclic processes. One thing is certain, there is a great potential for change based on the knowledge at hand today, and sound economic planning based on knowledge may be sufficient to avoid economic collapse. In a Nation with a trillion dollar economy, the risks are too great not to invest an insignificantly small percent of this economy to acquire the maximum amount of knowledge possible and establish this knowledge base as soon as possible. In the case of environmental data, data is knowledge; there can never be enough data, and the data record can never be long enough. But where there is data, it must be systematically managed to be of any value at all.

# DATA MANAGEMENT

WILLIAM M. CALLICOTT

OFFICE OF SYSTEMS DEVELOPMENT
DATA MANAGEMENT SYSTEMS DIVISION
NATIONAL ENVIRONMENTAL SATELLITE,
DATA AND INFORMATION SERVICE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

WORKSHOP ON
MONITORING AND REPORTING TECHNIQUES FOR ERROR
RATES AND ERROR DISTRIBUTIONS IN OPTICAL DISK SYSTEMS

COLORADO SPRINGS, COLORADO
AUGUST 5, 1991

- Ingest
- Quality Control
- Catalog
- Access
- Preservation

# INGEST

- o Three discipline centers and U.S. World Data Centers:

  National Climate Data Center - Asheville, North Carolina
  National Oceanographic Data Center - Washington, D.C.
  National Geophysical Data Center - Boulder, Colorado

- o Sixteen centers of data:

  Satellite Data Processing and Distribution - Suitland, MD
  National Weather Service - Silver Spring, MD
  National Meteorological Center - Camp Springs, MD
  Charting and Geodetic Services - Rockville, MD
  Climate Analysis Center - Camp Springs, MD
  National Marine Fisheries Service - Seattle, WA; Woods Hole, MA; Miami, FL; Bay St Louis, MS; San Diego, CA
  Navy Joint Ice Center - Suitland, MD
  Great Lakes Environmental Research Lab - Ann Arbor, MI
  Center for Ocean Analysis and Prediction - Monterey, CA
  Ocean Products Center - Suitland, MD
  Equatorial Pacific Information Collection - Seattle, WA
  Air Resources Laboratory - Silver Spring, MD
  University of Hawaii - Honolulu, HI
  National Snow and Ice Data Center - Boulder, CO
  National Tide and Water Level Data Base - Rockville, MD
  Global Monitoring for Global Change - Boulder, CO

# NOAA CENTERS

- o DIGITAL HOLDINGS INCLUDE: 11 TB ON 240,000 COMPUTER TAPES
- o IN-SITU DATA INCLUDED IN ABOVE: 2 TB
- o GOES DATA HELD AT THE UNIVERSITY OF WISCONSIN: 100 TB ON 25,000 SONY U-MATIC BETA TAPES
- o SERVICE OVER 7,000 REQUESTS FOR DATA AND INFORMATION PER YEAR
- o ANNUALLY PRODUCE 5,000 BULLETINS TO 40,000 SUBSCRIBERS
- o OVER 90,000,000 PAGES OF NON-DIGITAL DATA AND INFORMATION

## IMMEDIATE DIGITAL VOLUME GROWTH
(Volumes in Billions of Bytes)

| | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 |
|---|---|---|---|---|---|---|
| Climate Data Center | 520 | 570 | 840 | 1,300 | 2,130 | 3,420 |
| Oceanographic Data Center | 30 | 160 | 200 | 260 | 280 | 370 |
| Geophysical Data Center | 450 | 540 | 650 | 780 | 930 | 1,110 |
| Satellite Data Services | 10,400 | 14,200 | 18,000 | 21,800 | 25,600 | 29,400 |
| GOES Data Archive | 107,000 | 113,000 | 120,000 | 134,600 | 149,200 | 164,000 |
| Accumulative Total: | 118,400 | 128,470 | 139,690 | 158,740 | 178,140 | 198,300 |

## PRESERVATION

o NOAA MISSION TO MANAGE THE ARCHIVES ON A PERMANENT BASIS AS A NATIONAL TRUST

o ALTERNATIVES FOR PRESERVING DATA ON AN INDEFINITE BASES:

- FIND A MEDIA THAT IS INDELIBLE INDEFINITELY

  ... Manage access system indefinitely
  ... Ensure media longevity (entomb media and site)
  ... Recurring Quality control
  ... Lifetime system maintenance
  ... Migrate on demand
  ... Operate archive center(s) as deep archive

- ASSUME A 10 YEAR SYSTEM AND TECHNOLOGY CYCLE

  ... Recurring migration of 10 year old media continually looking at developing technology advantages
  ... Keep two copies, one entire data set entombed, one in active archives at distributed discipline archive centers
  ... Maintain portions of the data at the active centers on-line, the rest near-line
  ... The master copy kept near-lie with sufficient on-line capability to service access requests and for migration processing

## ARCHIVE INTEGRITY

- MEDIA PERFORMANCE CONTINUALLY MONITORED THROUGH ERROR DETECTION AND CORRECTION INFORMATION PASS BACK
  - PROCESSING ON-DEMAND
  - SCHEDULED MEDIA MAINTENANCE
- MIGRATION OFFERS OPPORTUNITIES TO:
  - UP-TO-DATA CATALOG FACILITIES
  - INCLUDE LOW-LEVEL DATA INVENTORY DESCRIPTIONS WITHIN CATALOG INVENTORY FILE
  - IMPROVE DATA AGGREGATION TO MEET CURRENT SCIENCE REQUIREMENTS
  - COMPACT STORAGE AND DATA TRANSFER THROUGH THE USE OF ADVANCED TECHNOLOGIES
  - REGENERATION OF SYSTEMS AND DATA EXERCISES THE DATA FOR ITS HEALTH
  - ENABLES INCREASED ON-LINE ACCESS FACILITIES
  - OPENS NEW DOORS FOR DATA ACCESS AND TRANSFER
- QUALITY CONTROL:
  - ANALYZE DATA TO ENSURE CREDIBILITY DURING INGEST
  - MONITOR MEDIA PERFORMANCE TO ENSURE RELIABILITY OVER TIME
  - MAINTAIN LOG OF ACCESS ACTIVITIES TO BUILD DECISION HISTORY

A Reprint from the

# PROCEEDINGS

Of SPIE-The International Society for Optical Engineering

Volume 899

Optical Storage Technology and Applications

12-15 January 1988
Los Angeles, California

## Determining error management strategies

**Winton E. Brown, Allen M. Earman**
Eastman Kodak Company, Rochester, New York 14650

Determining error management strategies

Winton E. Brown and Allen M. Earman

Eastman Kodak Company
Rochester, New York 14650

ABSTRACT

A statistical model has been designed to interpret experimental burst error data in determining an optimal error management strategy for optical media. The model is based on the theory of a Non-homogeneous Poisson Point Process. The assumptions for the evaluation of defects capable of causing error bursts in a user's data stream (burst starts and burst lengths) will be discussed. With this model it is possible to determine the optimal error management strategy to achieve any desired system performance, including depth of interleaving, error correcting capability and certification fencing for excessively long error bursts.

## 1. Summary

A statistical-theoretical model for a clustered error burst process has been designed based on the theory of Non-homogeneous Compound Poisson Point Processes. This model is the first step in the design of statistical control algorithms for the purpose of quality control and certification of production disks. Detailed empirical error burst-gap data have been gathered from a variety of optical disks. At this stage, we use the model to assess the performance of several possible EDAC strategies in conjunction with the available empirical data (7, 8). Part I (Discussion, Appendix) discusses the theoretical development of the model. The purpose here is to evaluate the Mean Corrected Byte Error Rate for five correction codes with available empirical burst length distributions where we allow the Certification Fencing Limit and the mean cluster burst frequency to vary for different levels of clustering.

Note Figure #1, a plot of the frequency distribution of error burst data of available empirical data. The data is presented on a double log axis. The y-axis is $\log_{10}$

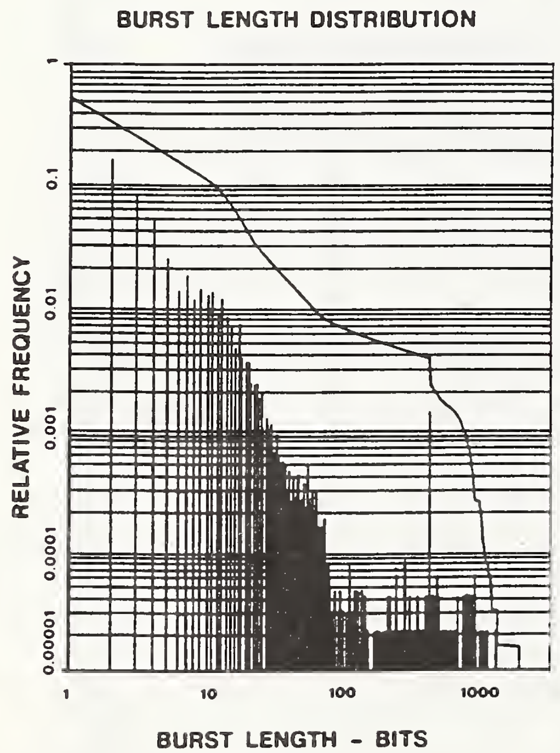



Figure 1

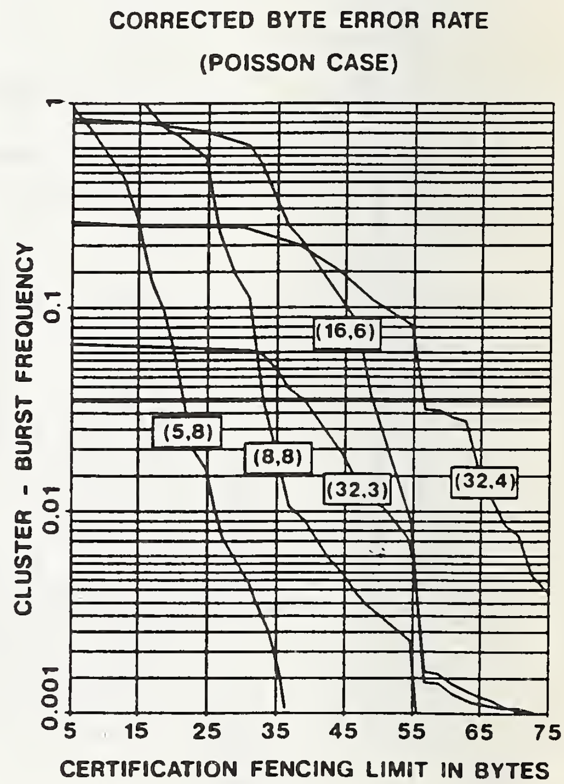



Figure 2

probability from a base line of .00001 to 1. The x-axis is $\log_{10}$ bits from 1 to 3200. The continuous curve is a plot of one minus the cumulative distribution. Ninety percent of all bursts are of size 11 bits or less, 99 percent are of size 58 bits or less. The relatively large pulse at 448 is a sampling artifact. Although not readily evident from this plot, the distribution is actually highly skewed toward burst sizes less than 10. The mean burst is 6.98 bits, the variance is 1871.1 bits$^2$.

Figure #2 shows the plot of contours for a Corrected Byte Error Rate equal to $10^{-12}$ under the assumption that burst starts are a purely random phenomenon (Poisson Process) for each of the five codes under consideration in this report:

(5,8)* code
(32,3) code
(32,4) code
(16,6) code
(8,8) code

The y-axis displays the mean frequency of burst starts per 1024 byte data block from $10^{-3}$ to $10^0$. The x-axis is the Certification Fencing Limit in bytes from 5 to 75 and refers to the maximum length of bursts that will be allowed on the disk after Certification. Blocks which have bursts longer than the Certification Fencing Limit will be removed from user accessibility in the Certification process. The heavy horizontal line at $\hat{\Lambda}_M$ = .03484 is the mean frequency of burst starts per 1024 byte data block as estimated from available data.

Notice that the (5,8) code requires a minimum Certification Fencing Limit of 22 bytes in order to achieve a $10^{-12}$ byte error rate at the available estimate of mean burst start frequency (intersection point of the $10^{-12}$ contour line with the heavy horizontal line at $\hat{\Lambda}_M$ = .03484) whereas the (32,3) code requires a minimum Certification Fencing Limit of 38 bytes. This is due to the 6.4X greater interleaving of the (32,3) code as opposed to the (5,8) code; i.e., long bursts are chopped up into a greater number of independent code blocks, thus the chance of a single burst affecting more than two to three bytes per code block is considerably reduced. The increased interleaving of the (32,3) code, however, is traded for a significantly lower code rate: 0.842 (ratio of user bytes to total bytes per code word); because the (32,3) code has more than twice as many parity bytes per block as the (5,8) code which has a code rate of 0.927.

The other disadvantage of the (32,3) code evident from the plot is that it has a threshold in burst frequency at about $\Lambda$ = .064 at which point no further decrease in Certification Fencing Limit will offer an improvement in Corrected Byte Error Rate. This is due to the reduced error correcting capability of the (32,3) code. The random error correcting capability of these codes increases geometrically by the number of bytes the code will correct. The (5,8) code will achieve a $10^{-12}$ Corrected Byte Error Rate up to 0.9 mean burst start frequency provided we are willing to fence out all blocks with bursts in excess of 5 bytes.

Figure #2 also displays the $10^{-12}$ contour for three other codes which achieve various combinations of interleaving and error correcting power between the above mentioned two. The (32,4) code with a code rate of 0.80 will require a minimum Certification Fencing Limit of 57 bytes at the available estimate of mean burst frequency ($\hat{\Lambda}_M$ = .03484), but has a burst frequency threshold at $\Lambda$ = 0.25. The (16,6) code with a code rate of 0.842 has a minimum Certification Fencing Limit of 48 bytes and a mean burst start frequency threshold at $\Lambda$ = 0.82. The (8,8) code has a minimum Certification Fencing Limit of 33 bytes and a mean burst start frequency threshold in excess of $\Lambda$ = 1.0. Figures 3, 4, and 5 show the progressive reduction of minimum Certification Fencing Limit and mean cluster burst start frequency threshold in these five codes as we allow for greater degrees of clustering of the bursts over the disk surface, k = 2, 5, 10 respectively, i.e., some data blocks have higher rates of error bursts than the mean cluster burst start frequency specified, whereas most blocks have logarithmically smaller rates of error burst starts than the mean. The proportion of blocks with various levels of mean block burst start frequency are

* These are non-standard code designations which the authors have adopted to describe the salient statistical features of these codes as they would be implemented on a 1024 byte block of user data. The first number, is the degree of interleaving, i.e., the extent to which the codes are 'wrapped' into the data block which is done to protect against the possibility that a long burst could exceed the error correcting power of any one of the codes. The second number is the error correcting power of the codes. For instance the (5,8) designation means 5-fold interleaved, 8-byte correcting. For a Reed-Solomon error correcting code this designation translates to a ($n = k + 2 \cdot t$, $k = 1025/nL$) where nL = 5, the degree of interleaving, and t = 8, the error correcting power of the code (9). All of the codes considered here use the Reed-Solomon algorithm approach. For the (5,8) code this designation corresponds to: ((n,k): (221,205)) for each of five interleaved codewords (2).

represented by a Gamma distribution where the reciprocal mean square to variance ratio is given by k, i.e., increasing k signifies a greater degree of clustering. See Figure #6.

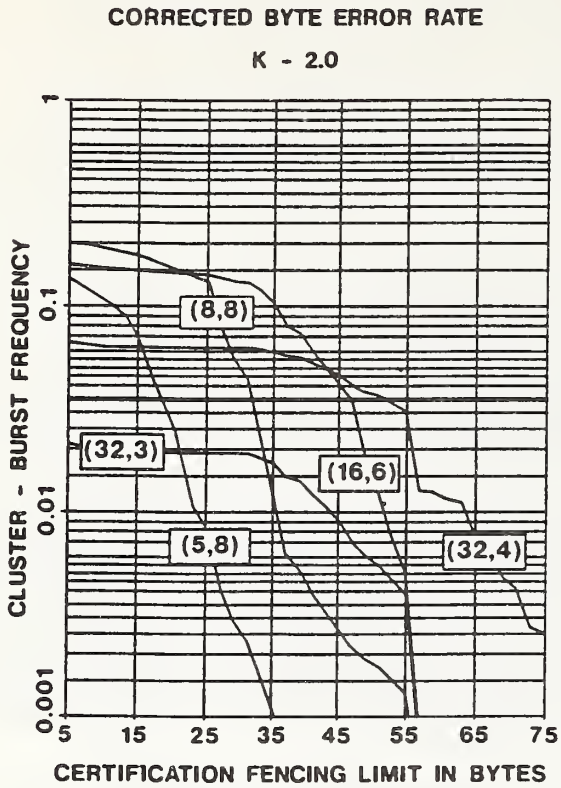



Figure 3

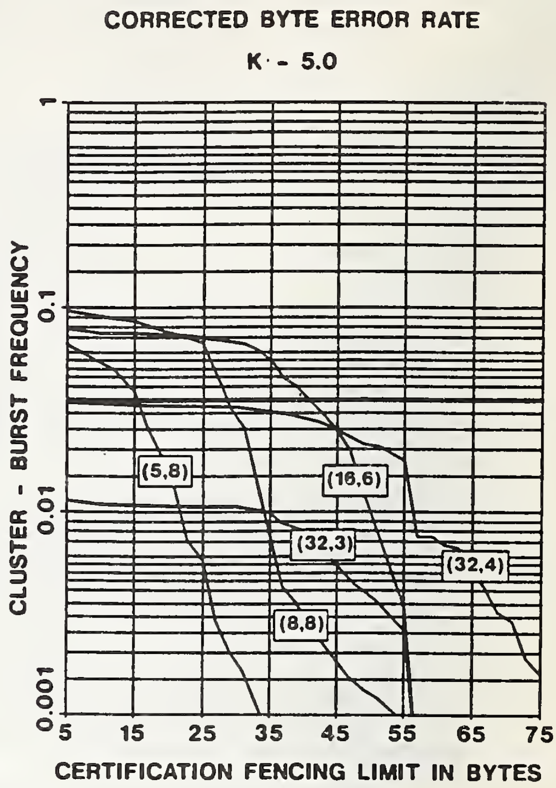



Figure 4

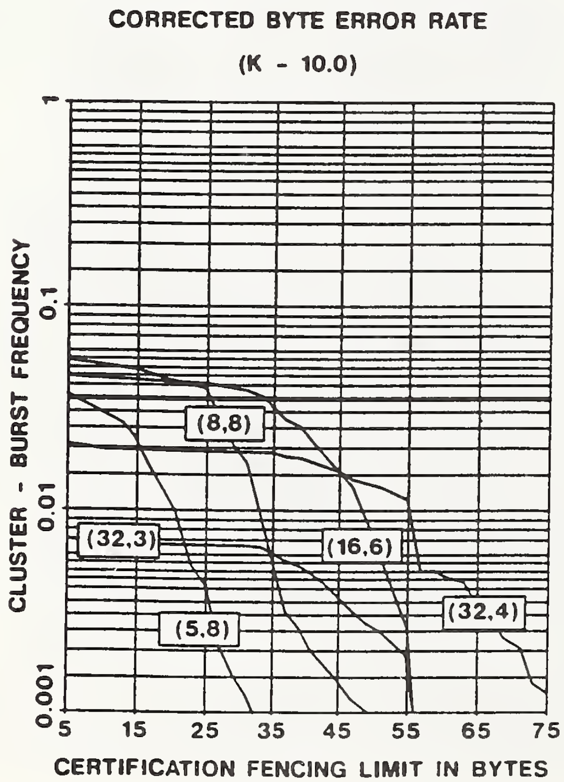



Figure 5

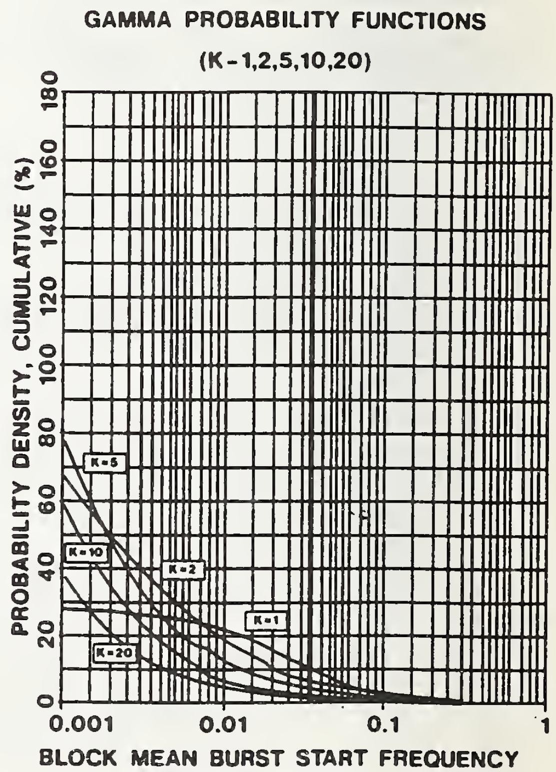



Figure 6

Table I summarizes the minimum Certification Fencing Limit at the available estimate of mean frequency of burst starts (1) and the mean cluster burst start frequency threshold (2), for the five codes considered:

Table I

| Code | Poisson Case | | k = 2 | | k = 5 | | k = 10 | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (1) | (2) | (1) | (2) | (1) | (2) |
| (5,8) | 22 | .9 | 18 | .14 | 16 | .067 | 7 | .037 |
| (32,3) | 38 | .064 | -- | .02 | -- | .012 | -- | .0071 |
| (32,4) | 57 | .25 | 52 | .065 | 5 | .034 | -- | .021 |
| (16,6) | 48 | .82 | 47 | .16 | 41 | .08 | 34 | .045 |
| (8,8) | 33 | 1.0 | 32 | .20 | 28 | .098 | 26 | .055 |

We estimate that the clustering on our current disks is represented by k between 2 and 5. At these levels of clustering of the burst start frequency parameter, the (32,3) code simply has insufficient burst correcting capability to achieve a $10^{-12}$ Corrected Byte Error Rate at the available estimate of mean burst start frequency for <u>any</u> level of Certification Fencing Limit. The (32,4) code is marginally acceptable provided we assume that production disks will in <u>every</u> case have a mean burst start frequency better than the available estimate of mean burst start frequency $\hat{\Lambda}_M = .03484$. The (16,6) code and the (8,8) code will achieve a minimum Certification Fencing Limit of 41 and 28 respectively. These two codes also maintain reasonable minimum Certification Fencing Limits at mean burst start frequencies up to 2.6 times greater than the available estimate of mean burst start frequency at clustering levels as high as that represented by k = 5.0. The (5,8) code will achieve a threshold in mean cluster burst start frequency of $\Lambda = .067$ (1.9 times the available estimate of mean burst start frequency) but requires Certification down to 16 bytes.

## 2. Conclusions

The two-stage Compound Poisson Point process model utilizes a burst-error density data-set generated by empirical evaluation of prototypical Kodak Optical Disks. This model is a dynamic instrument for the development and refinement of a high-reliability error-management strategy for Kodak Optical recording products and provides updated results and conclusions as the data-set is revised.

The system performance of our optical disk as perceived by the user after coding/decoding through these EDAC strategies depends very much on the statistical structure of the error burst process on production disks as modified by Certification Fencing, i.e., the distribution of burst length and the cluster distribution of block mean cluster burst start frequency. However, unless we can demonstrate that production disks have a radically different burst length distribution or the block burst start frequency parameter has a mean value above 1.5 times the available estimated mean frequency of burst starts or is clustered to an extent greater than that represented herein by k=2.0 i.e. $\sigma_\Lambda{}^2 = 2.0 \cdot \mu_\Lambda{}^2$ then the following conclusions are warranted:

a. The (5,8) code will no doubt work. Note however, that it has a narrow Certification Fencing Window requiring Certification possibly down to five bytes or less depending on clustering level k and the mean cluster burst start frequency.

b. The (32,3) code does not have sufficient error correcting capability to handle disks with a mean cluster burst start frequency equal to the available estimate of mean burst start frequency, where k, the clustering coefficient = 2.0.

c. The (32,4) code will work for disks with a mean cluster burst start frequency equal to the available estimate of mean burst start frequency, where k, the clustering coefficient = 2.0 and has a fencing limit of 52. It should be noted however, that if the mean block burst start frequency is on the order of 3-5 times that of the available estimate or has a burst clustering greater than k = 2.0, the (32,4) code will not have sufficient error correcting capability.

d. The (16,6) code recommended by the authors has a fencing limit of 47 at mean cluster burst start frequencies equal to the available estimate of mean burst start frequency where k, the clustering coefficient, = 2.0. The code also exhibits a sufficient Certification Fencing Window at high levels of mean cluster burst start frequency: 30 bytes at 3.8 times the available estimate of mean burst start frequency (however the Certification Fencing Window falls off rapidly to 1 byte at mean frequencies above this to five times the available estimate of mean burst start frequency).

e. A close second would be the (8,8) code which has a minimum Certification Fencing Limit of 32 at mean cluster burst start frequencies equal to the available estimate of mean burst start frequency where k, the clustering coefficient, = 2.0. It also has the advantage of protecting against higher mean block burst start frequencies up to 5.7 times the available estimate of mean burst start frequency. Notice that the Certification Fencing Window is greater than that of the (5,8) code but less than that of the (16,6) code.

The above conclusions are based on the available estimate of burst length distribution and mean frequency of burst starts and under the assumption that the degree of clustering of the block mean burst start frequency is adequately represented by a Gamma distribution for which the reciprocal mean square to variance ratio = 2.0; however, they do not necessarily represent what may be produced on a daily basis. Accordingly it will be necessary to gather additional data from the Certification development experiments and from the spin-stand test equipment that represent the actual manufacturing process characteristics in order to evaluate the ultimate code performance.

If production disks are significantly better than indicated by the preliminary data (the disks measured in (7, 8) were not produced under strict quality control procedures), any of these EDAC strategies may work satisfactorily; the issue then would focus on how much interleaving of the code block should be done to insure a suitably large minimum Certification Fencing Window, i.e., a tradeoff in cost of Certification vs. percentage of disk committed to parity bytes. It should be noted, however, that there are other sources of error that must be handled by the EDAC system. These may include future contamination or degradation to the disk. Because of these one may need to provide additional margin in Uncorrected Byte Error Rate to cover for a possible two-fold increase in mean cluster burst frequency and possible increase in burst length to insure meeting end of life quality specifications. This capability could be achieved with tighter Certification Fencing Limits, higher interleaved codes or increased power in each code.

## 3. Discussion:

3.1. Theory The error model used for these analyses is a 2-stage Non-homogeneous Poisson Point Process model, essentially the error burst process embedded in a Poisson process (see 10). To follow this development, we need first to distinguish between a shower of defects on the disk surface, presumably measured by reflecting laser light off the surface, and a measurable burst of errors in a user's data stream wherein there is at least one bit in error for every byte interval of user's data. Under this scenario, it would be possible, for instance, to have two defect showers which overlap resulting in one long burst of errors in a user's data stream. This model conceptually differs from the 2-stage burst-gap renewal model discussed in (11). It is difficult to assess the occasion of overlap in the defect showers from detected errors in a user's data stream alone, because an overlap of two defect showers would cause a succession of error bursts which would be counted as one long error burst under reasonable measurement and estimation procedures. However, there are important reasons for dealing with the error burst problem in this manner in spite of the difficulties in measurement:

3.1.1. A model built around the defect shower problem as opposed to the error burst problem is in principle capable of 'bridging the conceptual gap' between quality control of materials and the manufacturing process, to certification of finished disks, to optimal EDAC strategy and to user perceived performance over the lifetime of the product (longevity).

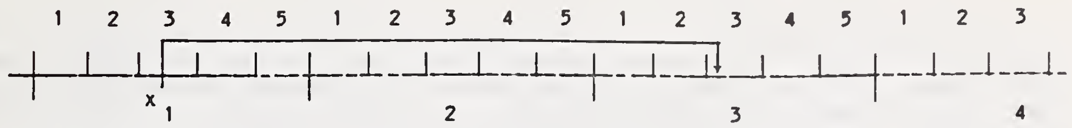
3.1.2. The parameters of such a model are statistically interpretable; i.e., confidence limits can be placed around key parameter estimates and Bayesian sampling statistics, for instance, can be incorporated with stated risks.

3.1.3. The explication of the model is mathematically and computationally tractable; i.e., the mathematics and computations can be done with a minimum of simplifying assumptions. The one or two assumptions are at a conceptual level rather than at an algorithmic level so we can readily ascertain their maximal effect.

The development here proceeds with the (5,8) code interleaving structure and parity byte structure but these results are easily generalized to the other codes. In the 5-fold interleaved 8-byte correcting code, there are 5 interleaved code blocks, with 5 bytes = 40 bits per interleaved set. The t = 8 byte correcting code will detect and correct 8 bytes in error in each of the 5 interleaved code blocks. For a Reed-Solomon code, there are $2 \cdot t$ bytes (16 bytes for the (5,8) code) of parity for each code block (3). A 1025 byte data block divided by 5 = 205 bytes in each user code block. (A user data block is of size

1024 bytes; the 1025th byte is made blank). 205 bytes per user block plus 16 bytes of parity = 221 bytes for each of 5 code blocks = 1105 bytes total. The rate for the code = 0.927.

error code block #




interleaved set #

Figure 7

Figure 7 shows the first three and one-half interleaved sets of a typical data block set up for the (5,8) code. The bytes are each numbered according to the error code block to which they belong: 1-5. We also show a typical defect shower which starts at x = 19 bits and is capable of causing an error burst (at least one defective bit in every user byte) to bit number 98. The defect shower is z - x + 1 = 80 bits long.

Suppose we pick one of these 221 interleaved sets, say the first, and assume that a single shower of physical defects starts 'randomly' in this interval distributed Poisson with a $\lambda_{Bk}$ parameter equal to the mean burst start frequency $\hat{\Lambda}_M$ = .03484 per 1024 byte block ($\lambda$ per byte = .000034023, $\lambda T$ = .000017013 per 5-byte interleaved set, T = 5). The arrival of a single shower of physical defects in an interleaved set of T = 5 bytes has a probability of $\lambda T e^{-\lambda T}$ = .00017010. Now, if no other defect showers arrive on this interleaved set (an event which has probability $1.447 * 10^{-8}$), we can show (by integrating the exponential distribution) that the marginal distribution of the start time (t) of the shower in [0,T] is uniform over [0,T]. Note: We require that the second event together with all the ensuing events arrive after T.

$$\int_0^t \int_T^\infty \frac{\lambda e^{-\lambda x_1} \lambda e^{-\lambda(x_2-x_1)}}{\lambda T e^{-\lambda T}}\, dx_2 dx_1$$

$$\int_0^t \frac{\lambda e^{-\lambda T}}{\lambda T e^{-\lambda T}}\, dx_1 = \int_0^t \frac{1}{T}\, dx_1 \tag{1}$$

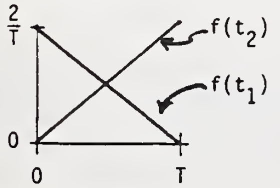
If we permit more than one shower to start in this interval of T = 5 byte interleaved set, then we are immediately faced with our first conceptual difficulty in the development of this model for which an assumption greatly simplifies further progress. The essential question addressed in its simplest form is: If two defect showers arrive in the first interleaved set, which one came first? An application of the theory of Order Statistics in this context will show that the first defect shower out of two has a starting time distribution of:

$$\int_0^{t_1} \frac{2\cdot(T-s)}{T^2} \cdot ds \tag{2}$$

and the 2nd defect shower has a starting time distribution of:

$$\int_0^{t_2} 2 \cdot \frac{s}{T^2}\, ds \tag{3}$$

These may be seen as simple straight line functions over [0,T]:



In general, if we have n defect showers, starting in the 1st interleaved set [0,T], the starting time of the $j^{th}$ one is Beta distributed (See (4)):

$$f(t_j) = \frac{n!}{(n-j)!\,\Gamma(j)} \cdot t_j^{j-1} \frac{(T-t_j)^{n-j}}{T^n} \quad (4)$$

All of this is necessary if we are to concern ourselves with overlap of the defect showers which might start within our interleaved set. If we choose to ignore the issue of defect shower overlap by making the assumption of uniformly distributed defect shower start times, we in effect permit our estimates to be biased by double counting which for mean shower start frequencies on the order of $\Lambda$ = .000170 happens infrequently.*

Assume a single shower of defects starts at location x somewhere in the first interleaving set and proceeds along with sufficient density to involve at least one bit in every user byte and then stops z - x + 1 bits later (see Figure #7). Then the conditional probability density of such a defect shower capable of causing an error burst of length z - x + 1 bits given that there is exactly one shower starting somewhere in the first interleaved set is:

$$\frac{1}{40} f_M(z - x + 1) \quad (5)$$

where $f_M$ represents the distribution of defect shower lengths as measured by available burst length statistics (7).

The event that the shower misses the first and second byte of the first error code block, (i.e., it must miss byte #1 and #6 completely) requires that it starts after bit #8 and stops on or before bit #40. The probability of this event is the sum of the joint probability of all possible events (subject to the constraints) of defect showers starting at bit x and ending at bit z of length z - x + 1.

$$g_1(0) = \frac{1}{40} \sum_{z=9}^{40} \sum_{x=9}^{z} f_M(z-x+1) \quad (6)$$

for the available burst length distributional data fenced to 18 bytes, $g_1(0)$ has a value of $g_1(0)$ = 0.72861 see Table II. This can be rewritten as:

$$g_1(0) = \frac{1}{40} \sum_{z=9}^{40} \sum_{x=1}^{z} f_M(z-x+1) - \frac{1}{40} \sum_{z=9}^{40} \sum_{x=1}^{8} f_M(z-x+1) \quad (7)$$

The first term is the probability that the burst stops after bit #8 and is contained within the first interleaved set which has a probability of .77259 and the latter term is the probability that the shower starts in the 1st byte of the error code block and stops between bit #9 and bit #40 (.04398).

In general, we can show that if the burst starts in the first interleaved set, the probability of it involving j of the bytes in the $m^{th}$ code block, m = 1, 2, 3, 4 or 5 is:

(nL = number of bytes in an interleaved set, i.e. nL = 5)

$$g_m(j) = \frac{1}{nL\cdot 8} \sum_{z=((j-1)\cdot nL+m-1)\cdot 8+1}^{(j\cdot nL+m-1)\cdot 8} \; \sum_{x=1}^{\min(z,8\cdot m)} f_M(z-x+1)$$

* We believe that the overlap probability is small (approximately one percent) due to the relatively low frequency of burst starts and in the case of evaluating the EDAC strategies it is insignificant. Presumably in Certification of the discs, the ordering of defect shower start times and possible overlap will be important. The next phase of this development will be to challenge the assumption that order of arrival and possible overlap is unimportant.

$$+\frac{1}{nL\cdot 8}\sum_{z=(j\cdot nL+m-1)\cdot 8+1}^{((j+1)\cdot nL+m-1)\cdot 8}\ \sum_{x=1}^{nL\cdot 8} f_M(z-x+1)$$

$$-\frac{1}{nL\cdot 8}\sum_{z=(j\cdot nL+m-1)\cdot 8+1}^{((j+1)\cdot nL+m-1)\cdot 8}\ \sum_{x=1}^{\min(z,8\cdot m)} f_M(z-x+1) \tag{8}$$

This expression includes all defect showers which cover the $m^{th}$ byte of the 1st interleaved set and those which do not cover this byte but cover an additional $m^{th}$ byte in the $(j + 1)$st interleaved set down the block.

The first term includes all events where the $m^{th}$ byte of the first interleaved set is covered by the defect shower and $j - 1$ $m^{th}$ bytes of the $j - 1$ succeeding interleaved sets are covered by the same defect shower. The second term includes all events where the $j$ $m^{th}$ bytes of the $j$ succeeding interleaved sets are covered by the same defect shower. The third term subtracts out of the second term the events which cover the $m^{th}$ byte of the first interleaved set.

To compute these sums, we merely need to change variable to $y = z - x + 1$ and collect all factors of common terms being careful to respect the summation limits which also undergo a change of variable. Table II lists $g_m(j)$ where $m = 1$ and for the burst length variable fenced to 18 bytes for the 5-fold interleaved code. Notice that since the bursts are fenced to 18 bytes the first error code block cannot be affected by this one burst by more than four bytes, i.e., the probability of $g_1(j) = 0$, for $j \geq 5$.

Ignoring the issue of defect shower overlap (essentially double counting the overlap) we can now show that if the defect showers arrive as a clustered Poisson process over the data block where $\lambda$ , the block frequency of defect shower arrivals is chosen from a Gamma distribution, the probability of $j$ bytes of error code block $m$ being affected over the entire 1105 byte block is:

$$h(j) = \sum_{n=1}^{\infty}\frac{T^n}{n!}\cdot g_m^{n*}(j)\int_{\lambda=0}^{\infty}\frac{\lambda^{n+\alpha-1}\, e^{-\lambda(T+\frac{1}{\beta})}}{\Gamma(\alpha)\beta^{\alpha}}\,d\lambda \tag{9}$$

$\alpha$ , $\beta$ are parameters of the Gamma distribution
$T$ is the length of the block: 1105 bytes in the case of the (5,8) code.

The expression under the integral sign permits us to deal with the clustering aspects of the defect shower frequency variable. We assume that the defect shower frequency for each block is distributed according to a Gamma distribution. The integral expression is recognized as $Z_G^{(n)}(-T)$, the $n^{th}$ derivative of the moment generating function of a Gamma distribution. $g_m^{n*}(j)$ is the $n^{th}$ fold convolution of the $g_m(j)$ distribution derived in equation (6).

Computing this expression requires that we prove three lemmas which will permit us to generate recursively the $h(j)$ (see Appendix). $h(j)$ is listed in Table III along with $\mathrm{Log}_{10}(1.0 - \sum_{i=1}^{j} h(j))$ in the case where $g_m(j)$ is fenced to 18 bytes and where the block defect shower frequency parameter $\Lambda$ has a mean $\mu_\Lambda$ equal to the available mean burst start frequency estimate $\hat{\Lambda}_M = .03484$ and a variance of $\sigma_\Lambda^2 = (2\cdot\mu_\Lambda^2) = .0024279$.

The Expected Value of the Corrected Byte Error Rate (per byte of user data) is computed under the following assumption:

When the number of byte errors ($n$) in the code word exceeds the byte correcting capability ($t$) of the error correcting code, i.e., $n \geq t$, the resulting number of incorrect bytes after decoding will be twice the byte correcting capability $(2\cdot t)$ plus the number of byte errors in excess of the byte correcting capability of the code $(n - t)$ which yields $(t + n)$ bytes (6).

This result occurs for $n \geq t$ because the Error Correcting Code syndrome calculator is forced to solve $t$-dependent equations for $n$ variables which results in erroneous solutions. The n + t resulting byte errors is the maximum value for miscorrection; yet, the probability of maximum miscorrection is relatively high given the total number of bytes to correct in a user data block.

This discussion has developed a two-stage compound Poisson process model for a defect burst error process for a set of nL independent interleaved Reed-Solomon code blocks divided equally into a 1024 byte data block where the mean frequency of burst starts for each data block can be considered as an independent sample from a Gamma burst start frequency clustering distribution. We have provided three lemmas (see Appendix) necessary to compute the system performance probability distribution. Certain expectations on this probability function, i.e., the expected number of errors per block, the mean corrected byte error rate (per byte of user data) and the mean ECC processing rate are readily calculable from these algorithms.

Table II

Probability of N bytes of error code block #1 becoming affected by a defect shower which starts in the first interleaved set. The shower is distributed as an empirical burst length variable fenced to 18 bytes. The code used is 5-fold interleaved, 8-byte correcting.

| N | PROB(N) |
|---|---|
| 0 | 0.728614777372148897933942O535 |
| 1 | 0.2649306530624980491300045521 |
| 2 | 0.0052848762831913630579725813 |
| 3 | 0.0009641486473836936936719594 |
| 4 | 0.0002055446347779961844088538 |
| 5 | 0.0000000000000000000000000000* |
| 6 | 0.0000000000000000000000000000 |
| 7 | 0.0000000000000000000000000000 |
| 8 | 0.0000000000000000000000000000 |

* Note that the fencing limit does not allow bursts which would yield an $N \geq 5$.

## 5. APPENDIX

1. The $n^{th}$ derivative of the moment generating function of the Gamma distribution can be generated recursively as:

$$Z_{G_\Lambda}^{(o)}(-T) = \frac{1}{(1+T\beta)^\alpha}$$

$$Z_{G_\Lambda}^{(n)}(-T) = Z_{G_\Lambda}^{(n-1)}(-T) \cdot \frac{(\alpha+n-1)\cdot\beta}{(1+T\beta)} \tag{10}$$

2. By redefining our $g_m(j)$, $j = 0, 1, 2, 3 \ldots$ function on the set of positive integers:

$$\hat{g}_m(j) = g_m(j)/(1-g(0)) \qquad j = 1, 2, 3 \ldots \tag{11}$$

Then we can show that (see 5)

$$\text{a.} \quad h(o) = Z_{G_\Lambda}^{(o)}(-T(1-g_m(o))) = \frac{1}{(1+T(1-g_m(o))\cdot\beta)^\alpha} \qquad j = 0$$

$$\text{b.} \quad h(j) = \sum_{n=1}^{j} \frac{(T(1-g_m(o)))^n}{n!} \cdot \hat{g}_m^{n*}(j)\; Z_{G_\Lambda}^{(j)}(-T(1-g_m(o))) \qquad j \neq 0 \tag{12}$$

Table III

Probability of N bytes in error code block #1 becoming affected by multiple defect showers which might arrive in a 1105 byte block of data plus parity. Each shower is distributed independently as an empirical burst length variable fenced to 18 bytes. The shower frequency parameter lambda is distributed gamma with mean = .0348416 and variance = .0024279 i.e. (K=2.0). The code used is 5-fold interleaved, 8-byte correcting.

| N | PROB(N) | LOG10(1.-CUM(PROB(N))) |
|---|---|---|
| 0 | 0.98995011630604467493780041940314661 | -1.998 |
| 1 | 0.00966345781106285064293065937782599 | -3.413 |
| 2 | 0.00033426374144657952718457643475778 | -4.283 |
| 3 | 0.00004311491313984440234272608359284 | -5.043 |
| 4 | 0.00000876059727872845980474679151195 | -6.543 |
| 5 | 0.00000027181089490662170544772931222 | -7.829 |
| 6 | 0.00000001338039851114222550672923814 | -8.842 |
| 7 | 0.00000000128882820328157349391510545 | -9.821 |
| 8 | 0.00000000014336175737891506707100674 | -11.122 |
| 9 | 0.00000000000708394424190102616795798 | -12.337 |
| 10 | 0.00000000000042160673156896789140834 | -13.416 |
| 11 | 0.00000000000003513411029005161853348 | -14.487 |
| 12 | 0.00000000000000305782532566343318285 | -15.697 |
| 13 | 0.00000000000000018732327105918611220 | -16.874 |
| 14 | 0.00000000000000001233487254232789104 | -17.984 |
| 15 | 0.00000000000000000095762468561016906 | -19.093 |
| 16 | 0.00000000000000000007520359010464914 | -20.263 |
| 17 | 0.00000000000000000000507864908309931 | -21.419 |
| 18 | 0.00000000000000000000035250207989653 | -22.543 |
| 19 | 0.00000000000000000000002647414723923 | -23.669 |
| 20 | 0.00000000000000000000000199331697798 | -24.820 |
| 21 | 0.00000000000000000000000014066418431 | -25.965 |
| 22 | 0.00000000000000000000000001003144314 | -27.096 |
| 23 | 0.00000000000000000000000000074250636 | -28.227 |
| 24 | 0.00000000000000000000000000005498398 | -29.370 |
| 25 | 0.00000000000000000000000000000395973 | -30.509 |
| 26 | 0.00000000000000000000000000000028638 | -31.629 |
| 27 | 0.00000000000000000000000000000002108 | -32.601 |
| 28 | 0.00000000000000000000000000000000155 | -33.016 |
| 29 | 0.00000000000000000000000000000000011 | -33.016 |
| 30 | 0.00000000000000000000000000000000001 | -33.016 |
| 31 | 0.00000000000000000000000000000000000 | -33.316 |

Corrected Byte error rate (log10) = − 12.202

3. This form of our expression now allows us to compute $h(j)$ recursively using Brown's adaptation (1982) of Adelson's routine (1966) for recursive generation of Compound Poisson Probabilities under the case of Gamma variation in the Poisson Parameter (see 4, 1).

$$\text{let } C = \frac{T\cdot(1-g_m(o))\cdot\beta}{1+\beta T(1-g_m(o))}$$

Then:

$$h(0) = \frac{1}{(1+\beta T(1-g_m(o)))^{\alpha}}$$

$$h(1) = C\cdot\alpha\cdot\hat{g}_m(1)\cdot h(o)$$

$$h(2) = C\cdot(1+\frac{(\alpha-1)}{2})\cdot \hat{g}_m(1)\cdot h(1) + C\cdot\alpha\cdot\hat{g}_m(2)\cdot h(o)$$

$$\vdots$$

$$h(j) = C\cdot(1+\frac{(\alpha-1)}{j})\cdot\hat{g}_m(1)\cdot h(j-1) + C\cdot(1+2\cdot\frac{(\alpha-1)}{j})\cdot\hat{g}_m(2)\cdot h(j-2)$$

$$+ \ldots\ldots + C\cdot(1+\frac{(j-1)\cdot(\alpha-1)}{j})\cdot\hat{g}_m(j-1)h(1)$$

$$+ C\cdot\alpha\cdot\hat{g}_m(j)\cdot h(o) \qquad (13)$$

## 4. Bibliography


1. R. M. Adelson, "Compound Poisson Distributions," Operations Research Quarterly, Vol. 17 pp. 73-75, (March 1966).
2. E. R. Berlekamp, "Final Report on VLSI Plans for 25 mips/channel EDAC, Expandable for up to 5 Channels," Report from Cyclotomic's Inc., unpublished.
3. E. R. Berlekamp, "Algebraic Coding Theory - Revised 1984 Edition," Aegian Park Press, California, 1984.
4. W. E. Brown, "Low Order Rate Inventory Control," Paper presented to TIMS Conference, Detroit, MI, April 1982.
5. W. E. Brown, "Notes on a Revision of Adelson's Routine for Computing Compound Poisson Probabilities for Functions Defined on the Set of Non-negative Integers," unpublished.
6. A. Marchant, "Preliminary Analysis of PRO-200 Error Rates," Letter to J. Wheeler, unpublished.
7. M. Melchie, "Notes," Regarding measured burst gap data on PRO-200 discs, unpublished.
8. M. Melchie, "Preliminary," Regarding measured burst and gap data for PRO 200 discs, unpublished.
9. W. W. Peterson and E. J. Weldon, Jr., "Error Correcting Codes - Second Edition," MIT Press, Massachusetts, 1972.
10. M. Thomas, "A Generalization of Poisson's Binomial Limit for Use in Ecology," Biometrica, pg. 18, (1948).
11. E. J. Weldon, Jr., "Selection and Implementation of an Interleaved Reed-Solomon Code for Use with an Optical Disk," unpublished.

# Error Measurement and Recovery Capabilities on Cirrus Logic Optical Disk Controller Chip-set


Nicos Syrimis
Cirrus Logic, Inc
Tel: 415-226 2153    Fax: 415-226 2170




# Agenda

1. Optical Disk Controller Chip-set Overview
2. Error Recovery/Reporting Capabilities
3. Defect Management
4. Software Reporting Schemes

1. Optical Disk Controller Chip-set Overview

2. Error Recovery/Reporting Capabilities

3. Defect Management

4. Software Reporting Schemes





## Optical Disk Controller - Summary

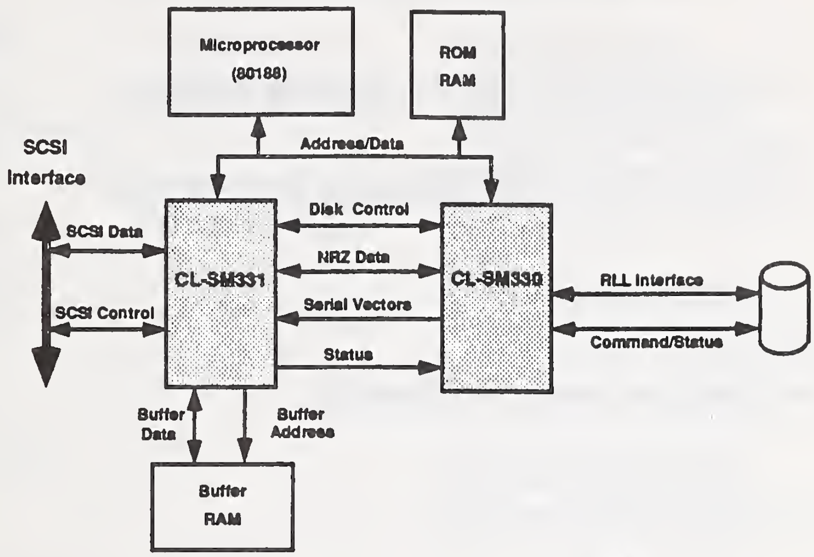
- High integration - two chip solution

CL-SM330 Optical Disk ENDEC/ECC
- ENDEC/Formatter
- ECC

CL-SM331 SCSI Optical Disk Controller
- SCSI interface
- Buffer manager
- Disk formatter

- 100 pin QFP packaging
- CMOS technology

## Optical Disk Controller - System Block Diagram






## Optical Disk Controller Implementation

- SM330/331 controller controls the datapath between host and drive
  - Data transfers between host and drive, via buffer
  - Error tolerance and recovery on header and data

- High level control performed by software in μC
  - Command decode and interpretation
  - Logical to physical address translation
  - Defect management (sector slipping, replacement, retirement)
  - Error recovery control

- Error measurement/logging
  - Under software control in μC
  - Use hardware logging/recovery capabilities of SM330/331

1. Optical Disk Controller Chip-set Overview

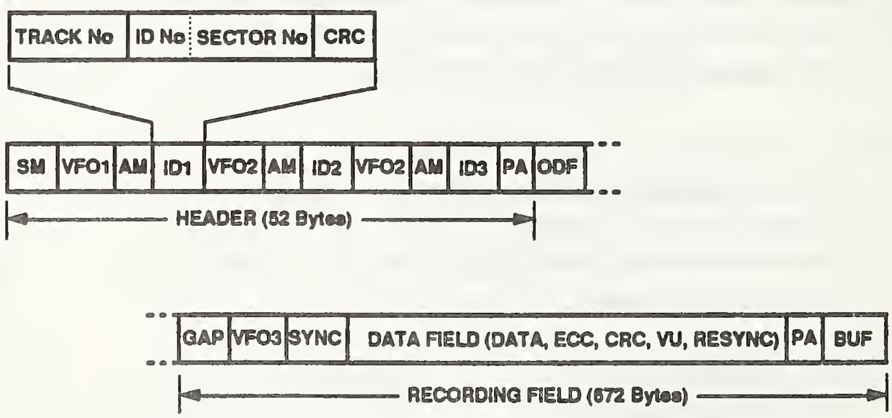
2. Error Recovery/Reporting Capabilities

3. Defect Management

4. Software Reporting Schemes

Error Measurement NS Rev 1.0 8/2/91-7 CIRRUS LOGIC

## CCS Sector Format - 90mm, 512 Bytes/sector




Error Measurement NS Rev 1.0 8/2/91-8 CIRRUS LOGIC

## Error Tolerance on Optical Format

| Field | Error Tolerance | SM330/331 Capability |
| --- | --- | --- |
| SM<br>(Sector Mark) | Redundant Pattern | Voting on Marks/Spaces<br>Asymmetry/Relaxation control<br>Pseudo Sector Mark generation |
| IDs | Redundant (triple) | Majority voting<br>ID skip |
| SYNC | Redundant Pattern<br>(12 nibbles) | Programmable voting threshold<br>Automatic sync on 1st/2nd ReSync<br>Recovery: sync on 3rd/4th ReSync |
| ReSync | N/A | Resynchronize to ReSync<br>Automatic window control<br>Protection on false ReSync detect<br>Missing ReSync threshold |

Error Measurement NS Rev 1.0 8/2/91-9

CIRRUS LOGIC

## Error Tolerance on Optical Format - ECC

| Field | Error Tolerance | SM330/331 Capability |
| --- | --- | --- |
| Data<br>Field | 512 Byte Sectors<br>5 ECC Interleaves<br>(max corr. 5x8=40 bytes) | <br>On-the-fly error correction<br>Extended recovery: 16 bytes/IV<br>(max corr. 5x16=80 bytes) |
| | 1024 Byte Sectors<br>10 ECC Interleaves<br>(max corr. 10x8=80 bytes) | <br>On-the-fly error correction<br>Extended recovery: 16 bytes/IV<br>(max corr. 10x16=160 bytes) |

Error Measurement NS Rev 1.0 8/2/91-10

CIRRUS LOGIC

## Error Reporting Capability

- SM330/331 controller reports error conditions for logging
  - SM not found (Pseudo Sector Mark Generated)
  - ID voting threshold not met - details of bad AM/CRC per ID
  - Data SYNC error
  - ReSync threshold exceeded
  - Error/Interleave ECC correction threshold exceeded
  - Total errors/sector ECC correction threshold exceeded
- SM330/331 ECC error monitoring capability - use for error log
  - Interrupt generated at the end of every sector read/corrected
  - Report available on:

    Total errors/sector detected

    Maximum errors/interleave detected

Error Measurement NS Rev 1.0 8/2/91-11

CIRRUS LOGIC

1. Optical Disk Controller Chip-set Overview
2. Error Recovery/Reporting Capabilities
3. Defect Management
4. Software Reporting Schemes

Error Measurement NS Rev 1.0 8/2/91-12

CIRRUS LOGIC

## Defect Management

- Defect management defined by ANSI/ISO/ECMA standards
- Two methods are available
  1. Sector Slipping
     - Media needs certification
     - Optional
     - Primary Defect List (P-List - PDL)
  2. Linear Replacement
     - Compulsory
     - Grown Defect List (G-List - SDL)
- Defect management algorithms performed by software in $\mu C$
  - SM330/331 controller supports efficient sector slipping operation

Error Measurement NS Rev 1.0 8/2/91-13 CIRRUS LOGIC

1. Optical Disk Controller Chip-set Overview
2. Error Recovery/Reporting Capabilities
3. Defect Management
4. Software Reporting Schemes

Error Measurement NS Rev 1.0 8/2/91-14 CIRRUS LOGIC

## SCSI-2 Error Logging/Measurement Capabilities

- Read/Write Error Recovery Parameters
  - Allows setting of error thresholds, correction spans, etc
- Read Defect Data
  - Allows reading of P-List (PDL) and G-List (SDL)
- Mode Sense Command
  - Read Error Log (Standard Header, Vendor Unique data)
  - Buffer overflow/underflow condition
  - Error counter information
  - Last "n" error events page
  - Non-media error page
  - Vendor Specific information



CIRRUS LOGIC

Optical Disk System Error Detection/Correction:
A Records Management Perspective

August 5, 1991

Charles Obermeyer II
Office of Archival Research and Evaluation
National Archives and Records Administration
Washington, DC 20408

Error detection and correction mechanisms are designed to provide data integrity so that information stored on media can be retrieved with a 100% confidence level. Mechanisms used today are very good at achieving this within the limits of error correctability. As a consequence, users are lulled into a false sense of security regarding the permanence and retrievability of their data.

A weakness in the system is that errors can be introduced later as a function of system operation and storage environment. Some of these errors are temporary because they are caused by components in need of maintenance. For instance, disk drives have moving parts that wear out, optical lasers drift out of calibration, and atmospheric contaminants can cause media read errors. Other errors are permanent such as those which cause destructive deterioration of the media. For instance, environmental conditions can cause ageing artifacts to appear in the media which destroy the readability of the disk in isolated regions. There is a point of no return from which the optical information storage and retrieval system can no longer correct all errors that were detected for a particular block of data. Current systems generally inform the user "after-the-fact" that the point of no return has been reached. This is equivalent to a doctor telling you on your death bed that you are going to die. The information is explanatory but it doesn't really help the situation.

Records managers and archivists in both the federal and private sectors are concerned about the integrity of their optical disk systems and the permanence of the data written onto media by these systems. It is important for operations managers to have early warning signs which indicate that equipment is behaving in an abnormal way even if the equipment is taking care of errors that are being detected. Early problem detection in operational systems allows corrective action to be taken during non-production hours.

A similar situation exists with optical media. The system that writes and reads optical media takes care of correcting errors that are detected. However, all media are subject to degradation over time. Records managers and archivists must be able to determine the permanence of data recorded on optical media as a function of time in order to detect contamination or degradation. Contamination can often be corrected by cleaning the disk and determining/eliminating the source of the contamination product. Degradation over time can be corrected by a copy operation to a new media. The copy operation must be done before the media reaches the point of non-correctability.

Clearly, analytical tools are needed to provide managers and archivists with early warning mechanisms for detection of problem areas in both optical drive subsystems and in optical media. Such tools could be used to create profiles which could serve as a device or media specific "health" record. It is even more important that these tools be standardized in terms of the information captured so that reporting utilities can have the flexibility needed to provide useful as well as informative reports.

Such tools require at least two components. First, the optical storage and retrieval subsystem must be able to capture, store, and make available to the I/O bus standardized information relating to errors detected in its own subsystem and errors detected from the media. Second, a set of software utilities should be developed to capture and report this information in a standardized way so that meaningful subsystem and media profiles can be established.

Industry, the user community, and the government can work together to define and develop the analytical tools needed and to stimulate the creation of utilities that will provide both short and long term benefit to the optical disk user and provider communities. Systems operations will clearly benefit from the tools that provide early warning of systems failures. The concept of "permanence of records" can be strengthened by utilities that can monitor the "health" of deployed optical media. These tools can strengthen the acceptance of optical disk technology as a permanent records management methodology.

We need to take a leading role in this area if optical disk technology is to become the permanent electronic record's tool of choice. Right now, a small but crucial part of the optical disk technology highway is in the dark. The darkness is affecting the comfort factor of the user community and could become a drawback to the wider acceptance and use of the technology if not corrected. The definition and development of optical subsystem and optical media analytical tools will help illuminate the path of progress for optical technology.

# AMD

# ABEP

# Advanced Burst Error Processor

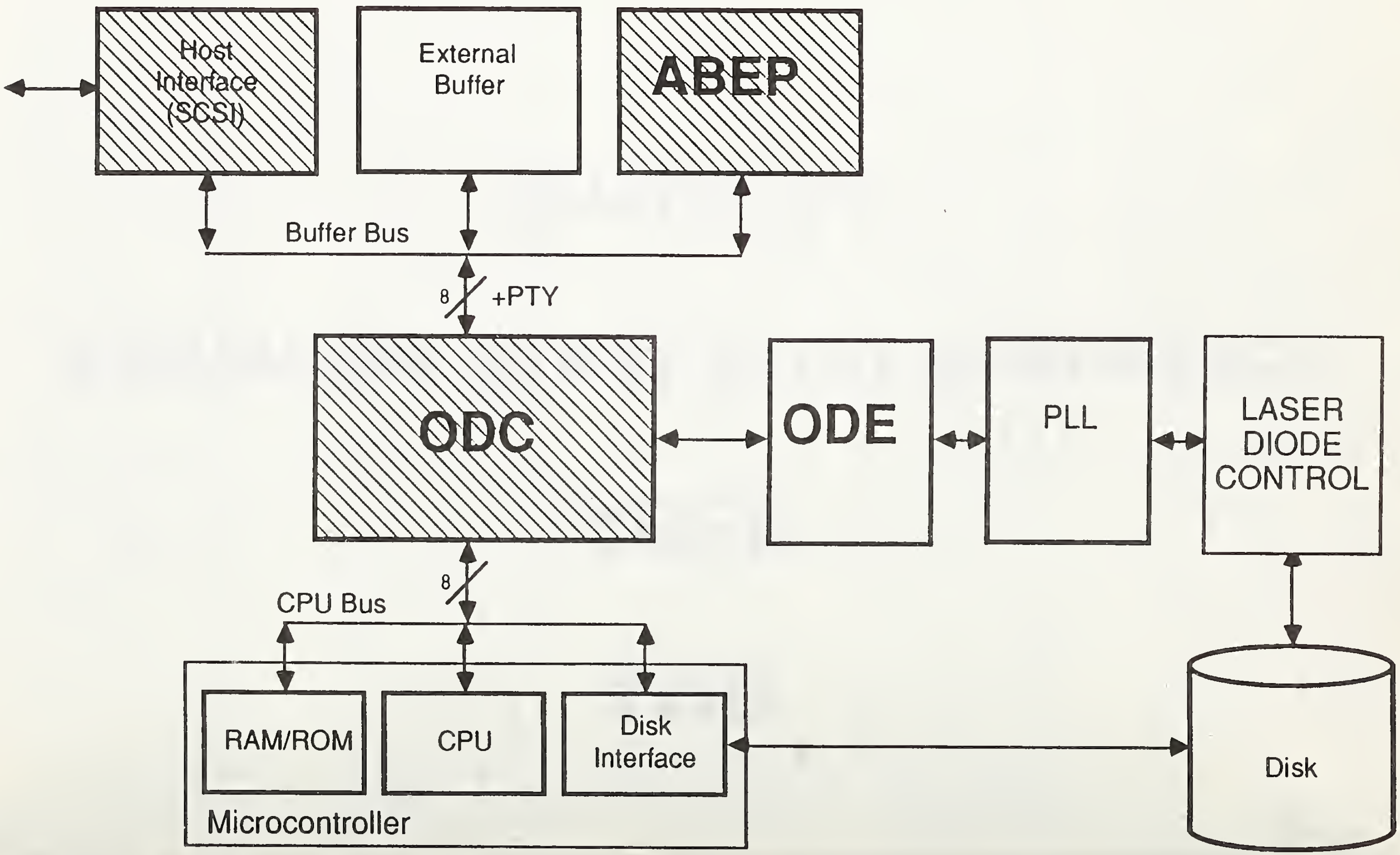
# Am95C94

Ben Chang

AMD

7/30/91 1

# ODC System Application




AMD

7/30/91 2

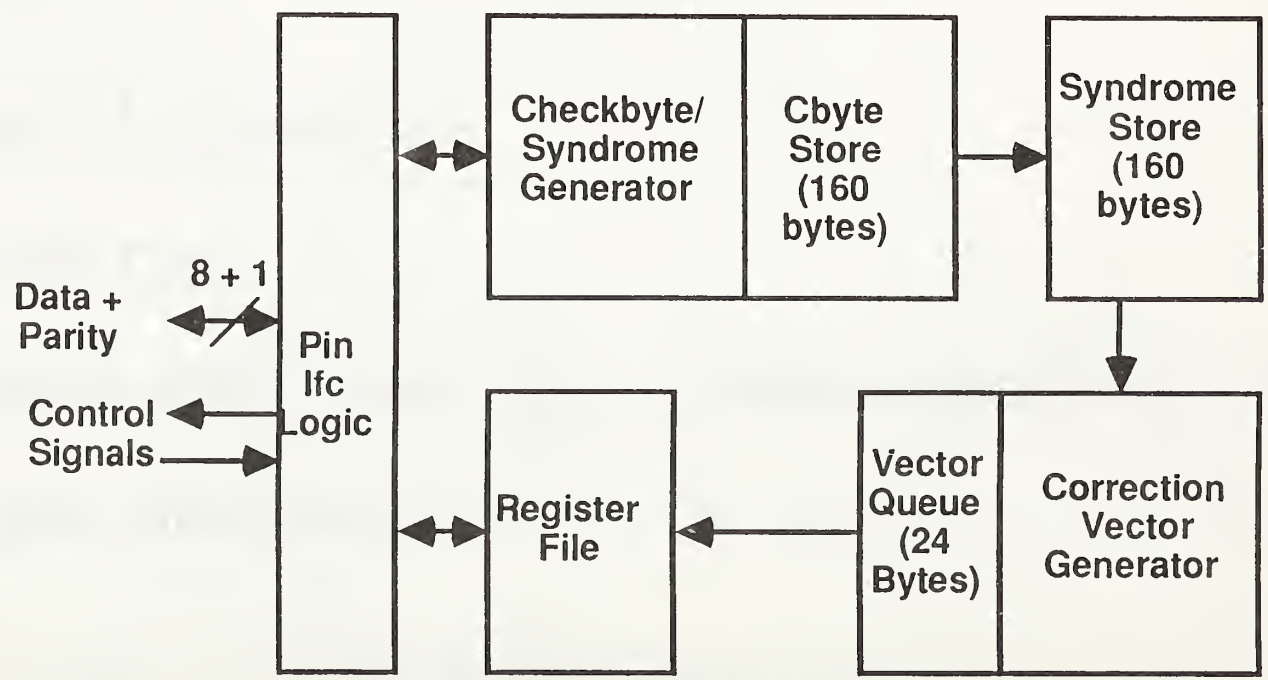
# Key Features

- 5/10 way interleaving
- complete support of X3B11 optical standards
- Layered CRC
- On the fly correction

AMD

7/30/91 3

# ABEP block diagram




AMD

7/30/91 4

# High Performance Architecture

- Buffer bus decoupled from CPU

- Buffer bus designed for a sustained bandwidth of 8MBytes/sec
  - Allows 4 to 5MBytes/sec host transfers
  - Simultaneous 24 to 32MBits/sec NRZ serial disk transfer

- Error correction does not affect performance or data transfer bandwidth - worst case error correction situation:
  - Correcting 1 Byte/10 microseconds
  - Need 4 Byte transfers on buffer bus/10 microseconds
  - Bandwidth required is 0.4 MBytes/sec

- Programmable SCSI threshold - allows data to be accumulated in the buffer before SCSI or disk transfers are initiated

AMD

7/30/91 5

# Hardware Error Correction

- Correction times:

  - ABEP: 7 microseconds to generate a correction vector

  - ODC: 10 microseconds to read the vector, compute the physical address and apply the correction (e.g. 400 microseconds to correct 40 Bytes in error).

  - Correction is transparent to the CPU (no S/W overhead)

  - Minimum buffer bus bandwidth consumed (a burst of 3 reads + 1 read-modify-write cycle)

AMD

7/30/91 6

# Hardware Error Correction

- Error correction done in hardware, in real time by the ABEP/ODC
- ABEP implements the X3B11-CCS Reed-Solomon polynomial
  - 5-way interleaving, sector size 128 to 1275 bytes (data + ECC)
  - 10-way interleaving, sector size 128 to 2550 bytes (data + ECC)
- ABEP has dual syndrome RAMs - one sector can be in the process of being corrected while the next one is being read
- Optional CPU interrupt when programmable error threshold is reached in any one interleave
- Layered CRC generation and verification implemented in hardware

AMD

7/30/91 7

# *Error Handling and Reporting*

# *In The*

# *Kodak 6800 Optical Recording System*

*By Scott A. Gerger*
*Small Format Optical Media Development Group*
*Eastman Kodak Company*
*Rochester, New York*

*5 August 1991*

. . . or

*"What in the world is a*

***'184', '185', '216', or '204'?"***

## ***Outline:***

*I. Problem Statement*

*II. Decision of Management and Engineering*

*III. System 6800 Error Reporting*

*IV. EDAC System Messages*

*V. Behind the EDAC System Messages*

*VI. Questions?*

## ***Problem Statement:***

***Given that errors occur during the operation of any complex opto-mecha-tronic system, how much information regarding errors should be reported to the host system ?***

## *Decision of Management and Engineering*

*"All of it!"*

***The Kodak 6800 System should make available to the host as much information regarding the operation of the drive as possible .***

## *System 6800 Error Reporting*

****There are 210 message types reported via the SCSI Interface.*

****There are multiple error logs for examining the recent history and frequency of errors which have occurred.*

****There is a special error log for monitoring the focus and tracking processor.*

## *Error reporting in the System 6800*

*Errors are grouped in the following manner:*

* *128 Hardware Errors*
* *40 Media Errors*
* *10 Unit Attention Errors (strictly speaking not errors)*
* *5 Aborted Command Errors*
* *7 Illegal Request Errors*
* *20 Miscellaneous Messages*

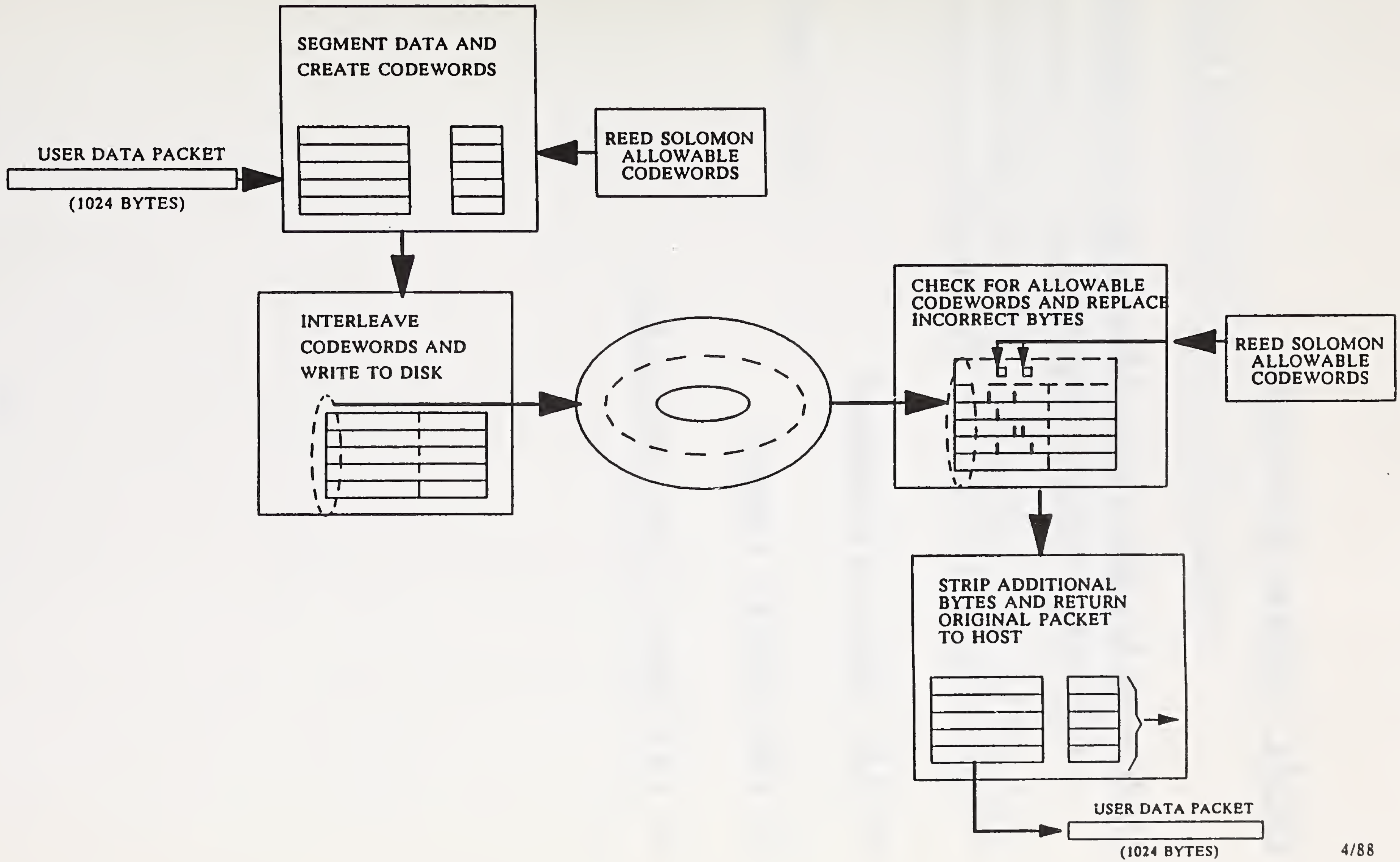
## *<u>EDAC System</u>*

***** Certification a key step in data integrity assurance***

***** Reed-Solomon Coding Scheme***

***** Interleaving to level 5 with 16 bytes of redundant check bytes per subcode (a total of 80 bytes per 1024 byte sector)***

# ERROR DETECTION AND CORRECTION OPERATION




4/88

## *EDAC System Messages*

***Of the 40 Media Error Messages, 4 are related to EDAC system operation.***

***** 204 Marginal Sector Detected***

***** 216 Retry Attempted***

***** 184 EDAC Pause***

***** 185 EDAC Uncorrectable***

## ***Behind the EDAC System Messages***

***Question:** Is there a way to monitor the activity of the EDAC system prior to receiving the 184, 185, 204, and 216 error messages?*

***Answer:** Yes*

## ***Behind the EDAC System Messages***

***Question: Is there a way to monitor the activity of the EDAC system prior to receiving the 184, 185, 204, and 216 error messages?***

*Answer: No*

***Standard Answers:***

*1. Let's check with Steve about that...*
*(Steve Pope, Kodak Berkley Research)*

*2. Give me your card and I'll* [get] *back to you on that one.*

*3. You really need to talk to Dan Nelson, Optical Drive Software Engineering, about that (716-588-0662).*

*4. I'm sure that Fred Rakvica, Optical Drive EDAC Systems Engineering, can answer your question (716-588-0590).*

# Error Reporting in the LD4100/LF4500 SCSI Event Log

Laser Magnetic Storage
Randy Glissmann

## ERROR REPORTING GOALS:

1. TRACK MEDIA AND DRIVE DEGRADATION
2. REPORT MEDIA AND DRIVE FAILURES

DRIVE WRITE OPERATION STEPS:

1. RECEIVE DATA TO STORED
    - SCSI PARITY ERROR
    - BUFFER PARITY ERROR

2. CALCULATE AND ADD ERROR CORRECTION
    - EDAC PARITY ERROR

3. SEEK TO REQUESTED PHYSICAL LOCATION
    - SEEK ERROR
    - HEADER ERROR
    - TRACK CROSSING ERROR
    - NO TRACK CAPTURE ERROR
    - SEEK FAILED AFTER MAXIMUM RETRIES

4. VERIFY BLANK SECTOR
    - OVERWRITE ERROR

5. Write data

- Servo Defect Error
- Write Power Error
- Write Timing Error
- Sync Error
- Tracking Error
- Focus Error
- Spares Area Status

WRITE VERIFICATION CHOICES:

1. READ VERIFY
2. SINGLE - PASS WRITE

## Read Verify

1. Data written on track is read
   Read Errors
2. Data rewritten if in error
   Write Errors

## Single - Pass Write: No read performed

1. Media Defects Monitored
2. Reflected Write Power Level Monitored
3. Data rewritten if number of errors exceeds recovery threshold

DRIVE READ OPERATION STEPS:

1. SEEK TO REQUESTED PHYSICAL LOCATION
   - SEEK ERROR
   - HEADER ERROR
   - TRACK CROSSING ERROR
   - NO TRACK CAPTURE ERROR
   - SEEK FAILED AFTER MAXIMUM RETRIES
2. READ DATA
   - SERVO DEFECT ERROR
   - SYNC ERROR
   - TRACKING ERROR
   - FOCUS ERROR
   - UNCORRECTABLE DATA ERROR
   - READ RETRIES
3. TRANSFER DATA TO HOST
   - BUFFER PARITY ERROR
   - SCSI PARITY ERROR

## Logged Correction Activity

- Number of sectors with X bytes in error
- Number of sectors with X codewords in error
- Number of codewords requiring X bytes correction
- Number of Read Amplitude Failures
- Number of Read Retries
- Number of Relocated Sectors Read
- Number of Read Commands Executed

| NIST-114A (REV. 3-90) | U.S. DEPARTMENT OF COMMERCE NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY | 1. PUBLICATION OR REPORT NUMBER NIST/SP-500/198 |
|---|---|---|
| | | 2. PERFORMING ORGANIZATION REPORT NUMBER |
| | **BIBLIOGRAPHIC DATA SHEET** | 3. PUBLICATION DATE October 1991 |

**4. TITLE AND SUBTITLE**

Monitoring and Reporting Techniques for Error Rate and Error Distribution in Optical Disk Systems

**5. AUTHOR(S)**

Fernando L. Podio, Editor

**6. PERFORMING ORGANIZATION (IF JOINT OR OTHER THAN NIST, SEE INSTRUCTIONS)**

U.S. DEPARTMENT OF COMMERCE
NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
GAITHERSBURG, MD 20899

**7. CONTRACT/GRANT NUMBER**

**8. TYPE OF REPORT AND PERIOD COVERED**

Final

**9. SPONSORING ORGANIZATION NAME AND COMPLETE ADDRESS (STREET, CITY, STATE, ZIP)**

National Institute of Standards and Technology
Building 225, Room A61
Gaithersburg, MD 20899

**10. SUPPLEMENTARY NOTES**

**11. ABSTRACT (A 200-WORD OR LESS FACTUAL SUMMARY OF MOST SIGNIFICANT INFORMATION. IF DOCUMENT INCLUDES A SIGNIFICANT BIBLIOGRAPHY OR LITERATURE SURVEY, MENTION IT HERE.)**

This report constitutes the proceedings of the workshop on Monitoring and Reporting Techniques for Error Rate and Error Distribution in Optical Disk Systems held on August 5, 1991, in Colorado Springs, Colorado. The objectives of this workshop were to identify the state of the art on error rate monitoring and reporting techniques in optical disk systems and to promote discussions on possible future implementations.

The workshop presentations included the description of a Computer Systems Laboratory of the National Institute of Standards and Technology (CSL/NIST) program for investigating error reporting capabilities of optical disk drives, Federal Government needs on error detection, correction and reporting, and the state of the art on error reporting capabilities in current generation drives. Presentations also included the description of the capabilities of error correction and detection current chips and discussions on error management strategies.

The participants noted the need for tools for reporting error rate activity through interfaces with the host and emphasized the importance of this subject to data managers in the Federal Government.

During a discussion panel, the participants identified a preliminary set of user requirements and it was suggested that a Government/industry working group be organized in order to document a consensus position between the Federal Government and industry in error reporting capabilities for future generation drives.

**12. KEY WORDS (6 TO 12 ENTRIES; ALPHABETICAL ORDER; CAPITALIZE ONLY PROPER NAMES; AND SEPARATE KEY WORDS BY SEMICOLONS)**

error detection and correction chips; error recovery techniques; optical disks, error detection and correction; optical disks, error distribution; optical disks, error management techniques; optical disks, error rate.

**13. AVAILABILITY**

| | |
|---|---|
| X | UNLIMITED |
| | FOR OFFICIAL DISTRIBUTION. DO NOT RELEASE TO NATIONAL TECHNICAL INFORMATION SERVICE (NTIS). |
| X | ORDER FROM SUPERINTENDENT OF DOCUMENTS, U.S. GOVERNMENT PRINTING OFFICE, WASHINGTON, DC 20402. |
| X | ORDER FROM NATIONAL TECHNICAL INFORMATION SERVICE (NTIS), SPRINGFIELD, VA 22161. |

**14. NUMBER OF PRINTED PAGES**

89

**15. PRICE**

ELECTRONIC FORM

* U.S. G.P.O.:1991-311-891:40054

## ANNOUNCEMENT OF NEW PUBLICATIONS ON COMPUTER SYSTEMS TECHNOLOGY

Superintendent of Documents
Government Printing Office
Washington, DC 20402

Dear Sir:

Please add my name to the announcement list of new publications to be issued in the series: National Institute of Standards and Technology Special Publication 500-.

Name ______________________________________________

Company ____________________________________________

Address ____________________________________________

City ________________ State ____________ Zip Code __________

**(Notification key N-503)**

# *NIST Technical Publications*

## *Periodical*

**Journal of Research of the National Institute of Standards and Technology**—Reports NIST research and development in those disciplines of the physical and engineering sciences in which the Institute is active. These include physics, chemistry, engineering, mathematics, and computer sciences.
Papers cover a broad range of subjects, with major emphasis on measurement methodology and the basic technology underlying standardization. Also included from time to time are survey articles on topics closely related to the Institute's technical and scientific programs. Issued six times a year.

## *Nonperiodicals*

**Monographs**—Major contributions to the technical literature on various subjects related to the Institute's scientific and technical activities.

**Handbooks**—Recommended codes of engineering and industrial practice (including safety codes) developed in cooperation with interested industries, professional organizations, and regulatory bodies.

**Special Publications**—Include proceedings of conferences sponsored by NIST, NIST annual reports, and other special publications appropriate to this grouping such as wall charts, pocket cards, and bibliographies.

**Applied Mathematics Series**—Mathematical tables, manuals, and studies of special interest to physicists, engineers, chemists, biologists, mathematicians, computer programmers, and others engaged in scientific and technical work.

**National Standard Reference Data Series**—Provides quantitative data on the physical and chemical properties of materials, compiled from the world's literature and critically evaluated. Developed under a worldwide program coordinated by NIST under the authority of the National Standard Data Act (Public Law 90-396). NOTE: The Journal of Physical and Chemical Reference Data (JPCRD) is published bi-monthly for NIST by the American Chemical Society (ACS) and the American Institute of Physics (AIP). Subscriptions, reprints, and supplements are available from ACS, 1155 Sixteenth St., NW., Washington, DC 20056.

**Building Science Series**—Disseminates technical information developed at the Institute on building materials, components, systems, and whole structures. The series presents research results, test methods, and performance criteria related to the structural and environmental functions and the durability and safety characteristics of building elements and systems.

**Technical Notes**—Studies or reports which are complete in themselves but restrictive in their treatment of a subject. Analogous to monographs but not so comprehensive in scope or definitive in treatment of the subject area. Often serve as a vehicle for final reports of work performed at NIST under the sponsorship of other government agencies.

**Voluntary Product Standards**—Developed under procedures published by the Department of Commerce in Part 10, Title 15, of the Code of Federal Regulations. The standards establish nationally recognized requirements for products, and provide all concerned interests with a basis for common understanding of the characteristics of the products. NIST administers this program as a supplement to the activities of the private sector standardizing organizations.

**Consumer Information Series**—Practical information, based on NIST research and experience, covering areas of interest to the consumer. Easily understandable language and illustrations provide useful background knowledge for shopping in today's technological marketplace.

*Order the* **above** *NIST publications from: Superintendent of Documents, Government Printing Office, Washington, DC 20402.*

*Order the* **following** *NIST publications—FIPS and NISTIRs—from the National Technical Information Service, Springfield, VA 22161.*

**Federal Information Processing Standards Publications (FIPS PUB)**—Publications in this series collectively constitute the Federal Information Processing Standards Register. The Register serves as the official source of information in the Federal Government regarding standards issued by NIST pursuant to the Federal Property and Administrative Services Act of 1949 as amended, Public Law 89-306 (79 Stat. 1127), and as implemented by Executive Order 11717 (38 FR 12315, dated May 11, 1973) and Part 6 of Title 15 CFR (Code of Federal Regulations).

**NIST Interagency Reports (NISTIR)**—A special series of interim or final reports on work performed by NIST for outside sponsors (both government and non-government). In general, initial distribution is handled by the sponsor; public distribution is by the National Technical Information Service, Springfield, VA 22161, in paper copy or microfiche form.

**U.S. Department of Commerce**
National Institute of Standards and Technology
Gaithersburg, MD 20899

Official Business
Penalty for Private Use $300