UDC 004.6

D. Yu. Manko, Ie. V. Beliak and Ye. M. Morozov
Institute for Information recording NAS of Ukraine
Shpak, 2 str., 03113 Kiev, Ukraine
email: dmitriy.manko@gmail.com

Problems of long-term data storage

This paper is devoted to analysis of problems of long-term information storage. A short historical overview of the methods of long-term information storage is presented. The overview briefly covers a period from ancient times up to nowadays. The fundamental problems faced by both researchers and private companies working in the field of long-term data storage are highlighted. The main problems come along this long-term data storage, known as digital dark ages are shown. Different ways of overcoming them are revealed. It is shown that many scientific facilities successfully resolved the task of data corruption due to degradation over time and external influence. It is noted that the Institute for information recording, NAS of Ukraine is a research facility proposed and realized a way of overcoming the longevity of data carrier problem. The analysis of the principles of the work of «fast» memory is made, its variants are presented. The possibility of introducing the principles of «fast» visual memory on optical media is considered and discussed.

Key words: *visual memory, digital dark ages, optical media, QR code.*

Introduction

Being based on achieved performances and rather long exploitation experience, the main area of optical disc use can be considered as long-term information storage. Just this direction is meant when developing the new types of optical discs [1].

Thus, the optical long-term disc in a future can be used only as archive data carrier. Authors of work [1] made the same conclusion. However, the existing optical discs cannot provide the required level of reliability and data storage time due to the low stability of the polycarbonate substrate, and insufficiently strong adhesion of a metal layer with a polycarbonate substrate [2, 3]. To solve the problem of long-term storage is proposed to use a variety of technical solutions which are based on the using highly stable materials for the recording media and the media substrate [4, 5].

The most progress is achieved in the area connected with physical data corruption due to external influence. The idea is to use durable, highly stable material and use it for design of data carrier. The data carrier usually performed as an optical disk in a popular storage format, or alternatively creation a new storage format. The main difference in design of optical data carrier consists in choice of material and data format of data carriers. In optical write-once discs for long-term storage, so-called M-drives offered used cermet as recording medium [6]. In 2012, Hitachi proposed using of silica glass [7]. Syylex Ltd. [8] and the Institute for information recording, NAS of Ukraine (IIR) [5] proposed using of sodium glass as a durable material. Artificial sapphire was proposed by IIR [2–4] and University of Southampton [9]. IIR proposes to utilize the conventional CD and DVD formats [2–4], and University of Southampton to put forward their own [9].

Digital Dark Age problem

Information can be reliably presented in two major ways: analogue and digital. Both ways have advantages and disadvantages. Let's consider digital way. Before doing that, we wish briefly describe the problem of digital dark ages (DDA). The very idea of DDA is based on following facts [10–12]:

— Data corruption due to degradation over time and external influence;

— Media obsolescence — hundreds of previously popular storage formats are now unreadable for practical purposes because the format is obsolete (5,25-inch floppy discs, Laser discs, Jaz Drives);

— The software used to create or access the data uses proprietary formats or becomes obsolete;

— Inadequate metadata (data explaining what it is and how to read and understand it).

According to the museum of obsolete media [13] near 170 kinds of data carriers are obsolete up to date, including well-known 8-inch floppy disk, 5,25-inch mini-floppy disk, HD-DVD-R, Sega Mega Drive, Memory Stick, etc.

Research centers around the world are working hard to resolve above mentioned problems. Recently, one can watch a tendency of shortening in use of optical discs. For example, in 2012, the market of this production was lowered by 10 % [10–12]. In 2013–2015 this tendency keeps continuing [10].

This situation is related with many reasons [10–12]:

— change in approaches to data storage based on cloud technology for data processing;

— availability of high-speed telecommunication channels caused drop in the necessity to create proper archives of multimedia information, computer soft is not bought on discs but is loaded from the web;

— storage size of optical carriers does not allow to provide recording the currently increasing data volumes in scientific investigations using a unique equipment as well as data of meteorological investigations;

— writable CDs are now cut out by flash memory due to exploitation convenience;

— notebook manufacturers abandon optical drives that increase the weight of notebooks and prefer convenient small-size flash memory;

— widely used is the practice to pre-load an operation system on a hidden part of hard disc drive, which allows re-loading without optical discs.

Historical overview of long-term memory

The amount of information obtained from analysis of analogues VM is astonished. The history itself is based on interpreting of analogues VM sources from bygone era. One of the good examples of it is the Phaistos disc (Fig. 1). The disc is a one of fired clay that was found by Luigi Pernier, an Italian archaeologist in 1908 in the Minoan palace of Phaistos on the Greek island of Crete. It probably dates from the middle or late Minoan Bronze Age (2000 to 1000 BC). It is about 16 cm in diameter and covered on both sides with a spiral of stamped symbols in closed sections or partitions. Forty-five symbols are used in sixty-one compartments (both sides) for a total ideographic count of 241 representations [14].

Fig. 1. The Phaistos Disc

The Phaistos disc is a good example of long-term data storage. But, unfortunately this data carrier has not encrypted so far [14], so we cannot consider data of this disc as information fully.

The nowadays example of analogues VM is a dog tag (identification tags worn by military personnel). The tags are primarily used for the identification of dead and wounded soldiers; they have personal info about the soldiers and convey essential basic medical information, such as blood type [15] and history of inoculations. The tags often indicate religious preference as well. Dog tags are usually fabricated from a corrosion-resistant metal [15]. The very idea of dog tag dates back from the ancient Rome, there it was known as a «signaculum» [16]. Signaculum was made of leather pouch and carried by Roman soldiers around their neck.

«Quick» memory

Thus, the idea of storage information using VM emerged. Since its introduction in 1994, the QR code has gained wide acceptance in such diverse industries as manufacturing, warehousing and logistics, retailing, healthcare, life sciences, transportation and office automation.

The QR (Quick Response) code is a two-dimensional (2D) matrix code that belongs to a larger set of machine — readable codes, all of which are often referred to as barcodes, regardless of whether they are made up of bars , squares or other — shaped elements. Compared with 1D codes, 2D codes can hold a larger amount of data in a smaller space, and compared with other 2D codes, the QR code can hold much more data still. In addition, an advanced error — correction method and other unique characteristics allow the QR code to be read more reliably and at higher speeds than other codes. Like written language, barcodes are visual representations of information. Unlike language, however, which humans can read, barcodes are designed to be read and understood (decoded) by computers, using machine — vision system consists of optical laser scanners or cameras and barcode — interpreting software. The rules with which a barcode is constructed (its grammar) and the character set it uses (its alphabet) are called its symbology.

There are a lot of different quick codes, developed for different purposes. The main among them and their characteristics are presented in Table.

ISO/IEC Standardized symbols

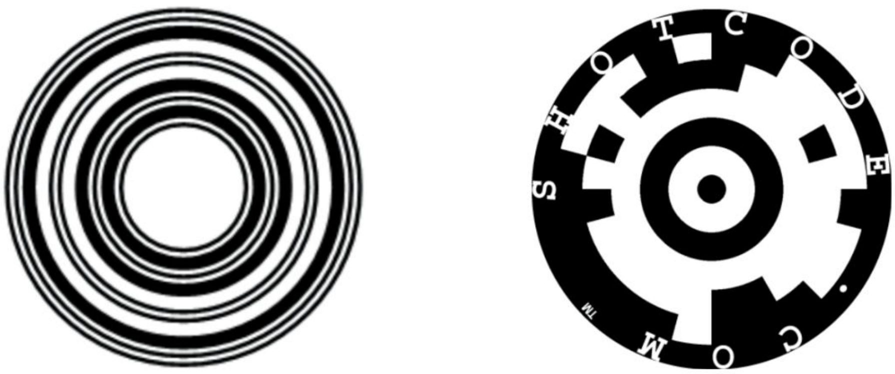
	PDF417	Data matrix	Maxi code	QR code	Aztec code
Developer (Country)	Symbol (USA)	CI Matrix (USA)	UPS (USA)	DENSO (Japan)	Hand Held Product (USA)
Code type	Multi-low	Matrix	Matrix	Matrix	Matrix
Data size (Alpha-numeric)	1850	2355	93	4296	3067
Characte-ristics	High capacity	High capacity, small space	Fast reading	High capacity, small space, fast reading	High capacity
Main market	Identity documents, Tickets	Food administration, Medical	Logistics	All industries	Airline, railroad
Standard	AIMI ISO	AIMI ISO	AIMI ISO	AIMI ISO JIS	AIMI

The implementation was the quick response (QR) code [17] which can be easily decoded by today's smartphones (Fig. 2). The level of QR code containing the largest amount of information can lose up to 7 % of the data before the code becomes unreadable. For the encoding of the final disk, it is likely that a coding scheme would be required which focuses on being easy decodable.

By keeping the size of the QR code low, it is possible to read out the disk by an optical microscope. For the demonstration, the entire disk was covered with a centimeter sized QR code. Each pixel of the code consists of a set of much smaller QR codes with pixels of only a few micrometers in size. The initial attempt to create a medium containing embedded data which is able to survive for 1 million years is promising. The optical readable data in the form of QR codes was able to survive the temperature up to 713 K [18].

Fig. 2: Optical microscope images of the same QRcode

Readout of bar-code or QR code from optical disk surface proves that there are applications where circular barcode structures are more preferable. Circular 1D barcodes were widely used for tagging CD/DVD items [19]. Structure of 1D circular barcode (Fig. 3,*a*) consists of series of concentric circles and typically based on a standard barcode symbology. This type of barcode is readable by the devices used to read traditional barcodes.



a) b)

Fig. 3. Circular barcode structure: a) Circular 1D barcode; b) ShotCode

Photoluminescent circular 1D barcode was one of the elements of multilayer optical disk's photoluminescent protective layer, together with photoluminescent image, microtext, additional data sectors and corrupted sectors area [20, 21]. Multilayer disk driver's objective lens is supposed to have a 2 mm vertical shift range which allows reading and recording circular 1D barcode at the same mode as a disk data reading out and recording (Fig. 4).

ShotCode (Fig. 3,*b*), in other hand, was a first circular 2D barcodes [22]. It was developed to be read by low resolution cameras (mobile phones or webcams). ShotCode had limited payload (5 to 6 bytes), so nowadays it has been widely replaced by QR codes. But researchers still have an interest in circular barcodes and develop new formats for codes with a data density comparable to QR codes. One of the most prominent projects was shown at [23]. Researchers decided to start from the require-ments of minimum payload value of 25 alphanumeric characters (200 bits according to ASCII characters code). It was supposed that the barcode design has to support different sizes (1 mm – 10 cm) and be stable to image distortion or noise. Three design types of the circular 2D barcodes differ only in the number of repetitions of the start marker and ra-dial zebra patterns (Fig. 5).

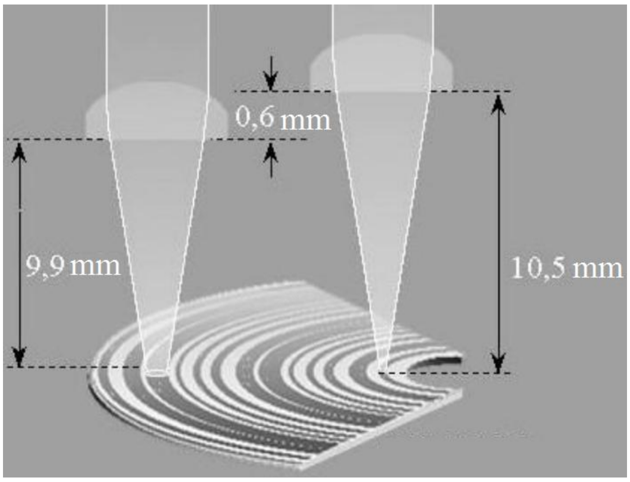


Fig. 4. Photoluminescent circular 1D barcode recording process

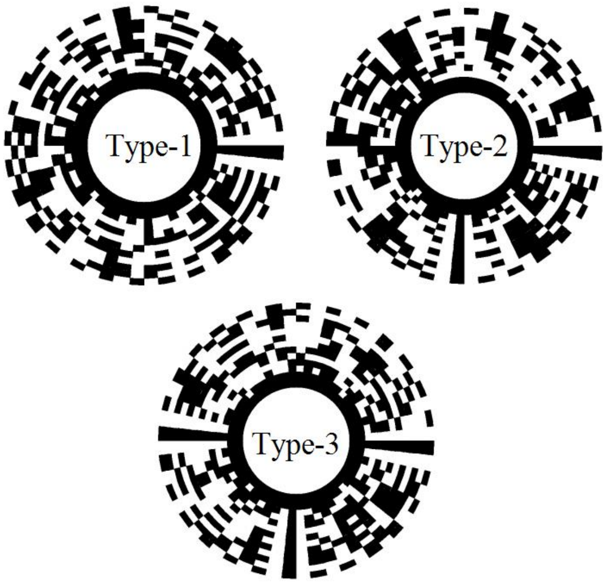


Fig. 5. Three types of generic circular 2D barcodes [b5]

The barcode includes:
— solid black ring: for locating the barcode in an image;
— white circle which is surrounded by black ring;

— square markers for a rough perspective correction;

— start marker which defines the beginning and determine the outer radius of the barcode;

— angular zebra pattern to determine the angles and achieve accurate perspective correction;

— radial zebra pattern which determines the radial timing of the rings;

— data zone which stores useful data and error correction codes.

It was shown that decoder is able to decode 5 mm 300 dpi circular 2D barcodes with a small average bit error rate that can be corrected with an error correction code.

Conclusions

The analysis of some aspects of long-term information storage has been performed. A short historical overview revealed that long-term storage of information was successfully fulfilled from the bygone era since 1000 BC. It is shown that the methods of long-term data storage using highly stable materials are actual nowadays. Methods of resolving the problems of digital dark ages are discussed. It is highlighted, that the Institute for Information recording, NAS of Ukraine successfully resolved the problem of data carrier degradation by proposing sodium glass and artificial sapphire as material for data carriers. The analysis of the principles of the work of «fast» visual memory is made, its variants are presented. The possibility of data storage using QR codes on surface of high stable materials is revealed. It is discussed that the readout of bar-code or QR code from optical disk surface proves that there are applications where circular barcode structures are more preferable.

1. Kryuchyn A.A., Petrov V.V., Kostyukevych S.O., Kostyukevych K.V., Kudryavtsev A.A. , Moskalenko N.L. Is there any future of optical discs? *Semiconductor Physics, Quantum Electronics & Optoelectronics*. 2013. Vol. 16. N 4. P. 362–365.

2. *Petrov Viacheslav, Andriy Kryuchyn and Ivan Gorbov.* High-density optical disks for long-term information storage. *Proc. SPIE.* 2011. **8011**. P. 80112J–80112J-7.

3. Gorbov I.V. The using of accelerated ageing of the optical media for estimating life expectancy of it. *Data recording, storage and processing*. 2009. Vol. 10. N 2. P. 3–12.

4. Kryuchyn A.A., Petrov V.V., Shanoilo S.M., Lapchuk A.S., Morozov Y.M. Sapphire optical discs for long term data storage. *SPIE Optical Engineering+ Applications*. 2014. P. 92010C–92010C-9.

5. Gorbov I.V., Belyakovsky V.O. Optical discs carriers for long-term data storage. *Data Recording, Storage & Processing*. 2007. Vol. 9. N 3. P. 73–87 (in Ukrainian).

6. Toppin E. Setting a new standart in permanent archival storage, Digital 2Disc, 2010.- P.42-44.

7. CD диск на основі кварцового скла. Розробка технології і відтворення запису (японською). URL: http://www.hitachi.co.jp/New/cnews/month/2012/09/0924.html

8. Durability of recordable DVD±R and DVD made of glass (Syylex) at elevated temperature and humidity. URL: https://www.lne.fr/publications/guides-documents-techniques/syylex-glass-dvd-accelerated-aging-report.pdf

9. Jingyu Zhang. 5D Data Storage by Ultrafast Laser Nanostructuring in Glass. URL: http://www.orc.soton.ac.uk/fileadmin/downloads/5D_Data_Storage_by_Ultrafast_Laser_Nanostructuring_in_Glass.pdf.

10. Michael Kriss. Handbook of Digital Imaging, John Wiley & Sons, 2015. 1824 p.

11. Melissa Terras. Digital Images for the Information Professional. Routledge, 2016. 264 p.

12. Ross Harvey Preserving Digital Materials. Walter de Gruyter, 2012. 261 p.

13. Museum Of Obsolete Media. URL: http://www.obsoletemedia.org/

14. John R. Elliott, Stephen Baxter. The DISC Quotient. *Acta Astronautica*. 2012. V. 78. P. 20–25.

15. Paul F. Braddock. Dog Tags: A History of the American Military Identification Tag, 1861-2002. 2003. 169 p.

16. Stephanie Lynn Budin, Jean Macintosh Turfa. Women in Antiquity: Real Women across the Ancient World. Routledge, 2016. 1110 p.

17. *Celalettin Aktaş.* The Evolution and Emergence of QR Codes. Cambridge Scholars Publisher, 2017. 139 p.

18. Tungsten optical disc can store data for 1 billion years. URL: https://www.extremetech.com/extreme/168548-tungsten-optical-disc-can-store-data-for-1-billion-years

19. Barcode software and components. URL: http://www.technoriversoft.com

20. Беляк Є.В., Єгупова Л.І., Крючин А.А., Петров В.В., Шанойло С.М., Кравець В.Г., Косяк І.В. Фізико-технічні методи верифікації та ідентифікації оптичних дисків. Київ: Ін-т пробл. реєстрації інформації НАН України, 2007. 86 с. (Препринт. НАН України, Ін-т пробл. реєстрації інформації).

21. Беляк Є.В. Технологія нанесення люмінесцентних ідентифікаційних елементів на поверхню оптичного диска. *Реєстрація, зберігання і оброб. даних*. 2006. Т. 8. № 3. С. 3–10.

22. Using visual tags to bypass Bluetooth device discovery. *Mobile Computing and Communications Review*. 2005. Vol. 9. N 1.

23. IMVS Fokus Report 2015 «Annular Barcodes». URL: http://www.fhnw.ch/technik/imvs/publikationen/artikel-2015/fokusreport-stamm

Recieved 19.06.2017