Digital and Biological Storage Systems – a Quantitative Comparison[1]

Tomasz Bilski
Poznan University of Technology
Pl. Sklodowskiej-Curie 5
60-965 Poznan, Poland
Phone: 0 48 61 66 53 554
tomasz.bilski@put.poznan.pl

ABSTRACT

The paper presents a quantitative comparison of digital/electronic and biological storage systems. Two biological storage systems are included: DNA and brain memory. First we will show some examples of digital-biological systems integration. In the main part of the paper, we discuss different storage aspects, mostly quantitative, such as: organization, functionality, data density, capacity, power consumption, redundancy, integrity, access time and data transfer rate. Numerous analogies between biological and electronic storage systems are pointed out. Finally, we will try to answer the question: which digital storage systems and media are the best equivalents for brain and DNA storage? The analysis of the storage systems resemblances and differences may facilitate to carry bioinformatics research.

Keywords

Data storage systems and media, redundancy, integrity, bioinformatics.

1. INTRODUCTION

For many years people have tried to look for relationship and correspondence between living creatures and machines (including computers).

The developments in biology and computer science influence one another. Bioinformatics is a far-reaching research area. Designers of digital storage look for the ideas within the biological systems – for example DNA storage and DNA computing projects. On the other hand medical research is carried on to build electronic prosthesis for damaged brain parts. The world's first electronic prosthesis of hippocampus has been recently tested by researchers at the University of Southern California. It is assumed that the digital circuit connected with other parts of the brain will perform the same processes as the damaged part of the brain it is replacing [7].

Bionetics'07, December 10-13, 2007, Budapest, Hungary.

Many other laboratories throughout the world work on brain-computer interface systems that will help people with severe motor disabilities [22]. The theory of information fundamentals are used to find the neural information coding schemes [5].

The digital and biological storage systems are already integrated and certainly will be more integrated in the future. The brain associative mode of operation is an archetype for holographic memory that is able to associatively search the entire data volume in a parallel way.

The first example of computation (Hamiltonian path problem solution) with DNA data encoding has been presented by L. Adleman in 1994 [1]. At the same time first concepts of DNA storage were discussed. In a multi-wavelength DNA storage concept the DNA polymers were used as fluorescent light sources for volumetric three-dimensional optical storage [9]. Widely publicized experiment has been carried in 2003 by P. Wong et al. [28]. The research team encoded data as artificial DNA and stored within the genome of multiplying bacteria. The DNA messages, each about 150 nucleotides long, were inserted into bacteria such as Escherichia coli and Deinococcus radiodurans. The beginning and end of each message had special DNA tags to stop the bacteria from identifying the message as an invading virus and destroying. The scientists were able to retrieve the DNA message after a hundred bacterial generations. Currently scientists face a problem with provision of a suitable environment for bacteria.

Quantitative relationships between human brain and computer have been discussed by scientists for years. Both A. Turing [26] and J. von Neumann [20] studied the topic. R. Merkle in his paper from 1989 [16] presented some approximations of the brain's computational power. He used three different approaches, based on: number of synapse operations per second, total nerve-impulse-distance per second and the total energy consumption. He concluded that the brain has a computational power $10^{13}\div10^{16}$ operations per second. The same issue had been studied by H. Moravec in 1998 [17]. He estimated the processing power and memory capacity necessary to match general intelectual performance of the human brain. He concluded that the required hardware will be available in cheap computers in 2020s.

[1] The research project is scheduled for years 2005÷2008 and partially supported by scientific grant from the Polish Ministry of Education and Science.

In the next sections of the paper we will reveal some analogies between the biological and electronic/digital storage systems. The following quantitative aspects of storage systems will be presented:

- data organization and functionality,
- coding and physical representation,
- data density and capacity,
- access time and transfer rate,
- power consumption,
- data integrity and redundancy.

It must be quickly clarified that for several reasons it is possible to compare the digital and biological storage media only in a vague way.

2. DATA ORGANIZATION AND FUNCTIONALITY

2.1 Computer

Computer memory is multilayer and hierarchical. It consists of, connected locally and remotely: registers, caches, operating memory, on-line memory, off-line memory. Data are stored in blocks (e.g. disk sectors). Blocks such as sectors are built from several fields: synchronization bytes, actual data, address, control data, and error detection/correction field. On the lowest level of hierarchy, stored blocks usually have constant length (for example: 512, 2048, 4096 bytes). Consecutive physical blocks are grouped into clusters. Some amount of clusters forms a file. Files have different sizes. They are categorized and organized into folders/directories and volumes. The conventional file classification is based on some diverse criteria, such as: nature of data (text, image, audio, ...), owner, application, age and usage rate. The volumes are stored on physical media such as disks or tapes, which may be organized into arrays or libraries. The sizes of the different blocks vary from several hundreds bytes (hard disk sector) to petabytes (tape library). Each sector, cluster, file or tape library is discretely addressable and the numbers of the data blocks on some levels of hierarchy are practically unlimited.

Generally all data operations are possible: read, write, modify, delete and copy. The operations may be performed with local or remote access to storage media and system. Some digital storage media (e.g. ROM, CD-R) limit the possible sorts of operations. Each operation is performed with a discrete address of data block (e.g. sector number, SCSI address, unique file name). Typically each storage element is dedicated to a given task (e.g. disk for online data, tape for backup). However, there are compound storage systems which may dynamically adapt themselves to the user and application requirements – for example, smart cells in the storage grids may efficiently change their functions according to the current application.

2.2 Brain

The brain is multilayer and has three interconnected types of memory [3]: sensory, working and long-term.

The sensory memories are equivalent to the computer registers or input buffers. They consist of the separate buffers for stimuli received through the three senses (sight, hearing and touch). The buffers are connected to the working memory. Attended and filtered data from the sensory buffer are transmitted to the working memory.

Working (short-term) memory is more or less corresponding to computer operating memory (more precisely to cache). It is assumed that data in the short-term memory vanish quickly (in $0.1 \div 10$ s) [20] as a result of decay and interference (we may say that data are replaced in short-term memory according to LFU (least frequently used) cache replacement algorithm). On the other hand the repeated exposure to a stimulus or the rehearsal of a given piece of data transfers it into long-term memory. The contents of working memory may be also at the output side of long-term memory.

As the name indicates the long-term memory is used for storage over a long time. It is organized into three broad divisions (equivalent to disk folders): episodic, semantic and procedural. Episodic memory stores events and experiences of one's life in a more or less serial form. Semantic memory is a structured, hierarchical record of facts, texts and concepts (the semantic hierarchy is similar to computer directory tree). Procedural memory (human executable files) deals with our skills, such as bicycle riding. The processes and organization inside the long-term memory have associative behavior.

The long-term memory is able to perform two main acitivities: storage (write) and retrieval (read). The information may be retrieved from the memory by two processes: recall or recognition. Recall is a reading from the memory with the associative addressing. It can be supported by the provision of retrieval cues (associative address) which enable to access the information in memory. Recognition process let us provide the knowledge that the information currently buffered in a short-term memory has been seen/heard before. Recognition is of lesser complexity, as auxiliary information is provided as a prompt.

The question is, if the data deletion process is performed by the long-term memory. Some neurobiologists claim that the forgetfulness is due to the fact that it becomes difficult to access certain items. In other words the associative addresses are not available [24]. This may be compared to the situation in which the file is still in some disk sectors but the operating system doesn't know the sectors addresses and hence can't access the file. A new concept based on the most recent research states that a special PP1 enzyme actively suppresses the memories [23].

There is a vague analogy between the computer storage grid and brain. The brain is composed of many complex functional networks, which may adapt themselves to the existing external requirements.

2.3 DNA

DNA storage is hierarchical and multilayer. We may say – it consists of files grouped into folders which are organized into volumes of data.

Human genome is stored inside 23 chromosomes (comparable to volumes of data in computer system). The chromosomes are linear structures and have different lengths. The smallest one have 33 400 000 pairs of nucleotides or bases, the largest – about five times more. A chromosome consists of two types of nucleotide sequences: repetitive and unique. It is assumed that repetitive DNA doesn't perform any function. Unique DNA sequences contain genes (folders) and other non genetic strings. Each chromosome contains several hundreds genes interspersed by repetitive DNA. Genes are built of exons and introns. Exons (files) are DNA sequences coding proteins. Introns don't code proteins. Exons have length from a dozen to several thousand nucleotides. Average human gene consists of multiple small exons separated by introns that are 10 or 100 times larger. The middle length of the human gene is about 30 000 nucleotides. The exact number of genes in the genome is unknown. The estimated value is 20÷25 thousand genes [27].

For replication process chromosomes are divided into many replicons (data clusters), which may be copied simultaneously. Average replicon size in human chromosome is about 45 000 pairs of nucleotides. Replicons are not equivalent to exons or introns. The DNA copying process may be compared to the block-based backup rather than to the file-based one.

DNA data is used primarily to copy the information for next generations. But other types of operations inside the helix are possible. The so called transposons are DNA sequences that can move around to different positions within genome of a single cell. Transposons may perform insertions, deletions and translocations within the genome.

3. CODING AND PHYSICAL REPRESENTATION

3.1 Computer

Computer storage processes are simultaneously digital and analog. They have digital character at higher levels of memory structure and analog at the lowest level. The processes are based on the changes of the medium states (states of transistors, states of regions with some magnetic or optical properties). Usually (e.g. in run length limited codes) in magnetic and optical storage the change of the state is interpreted as a channel code binary one and the absence of such change – binary zero. The changes are stimulated by the write element and are sensed by the read element. In different digital media diverse physical phenomena are used to represent data: electrical charge in semiconductor flash memory, refractive index in optical memory or magnetic remanence in magnetic memory.

3.2 Brain

The fundamental bricks of the brain are neurons. Each neuron has an axon and thousands tiny synapses. The brain's billions of neurons connect with one another in complex networks. The networks form groups (known also as cliques) of neurons and communication within a given group is achieved with a use of synchronization.

It is assumed that a memory trace is a sequence of biochemical changes inside the synapses between neurons. New memory traces are formed by development and modification of synapses. According to the theory presented by D. Hebb in 1949 [8], each synapse possesses a variable firing threshold which is reduced as the given neuron is repeatedly activated. Since Galvani experiments in 1791 we know that the transmitted signals have electrochemical nature. A given neuron may both receive and send signals to other neurons. It uses short and uniform electrical pulses with amplitude less than 100 mV to transmit information. A pulse is generated when the membrane potential of a neuron reaches a threshold value. The signal travels toward synapses terminating at postsynaptic neurons where it initiates or inhibits new signals. The synaptic gap between a synapse and next neuron is about 10 nm (incidentally this is roughly the same as the gap between hard disk head and disk surface). Many neurons fire bursts which are intrinsically generated patterns of closely spaced pulses. A burst may last up to 25 ms. It is assumed that bursts increase reliability of communication between neurons [13].

A sequence, of such processes may contain information based on diverse coding schemes. The details of the data encoding process are so far unknown. Few concepts exist: timing code, rate code and neural population code. In timing code it is assumed that data are represented by different time delays between the consecutive signals (roughly similar to RLL (run length limited) channel code in which the separation between two consecutive "1" bits determines the number of "0" bits), in rate code – by different number of signals per unit time (a kind of frequency modulation coding), and in neural population code – by nature and strength of correlations among neurons. Some experiments (for example [11]) confirmed that the data encoding process is performed in the brain part called hippocampus.

Some scientists suggest that it is possible to convert neural clique assemblies to strings of binary digits that would permit universal categorizations of internal brain representations [12].

3.3 DNA

The basic elements of DNA are four nucleotides called bases (we may also call them DNA bits). Since there are 4 different nucleotides we may assume that a single nucleotide is equivalent to two regular bits of data. The basic element of the genetic code is codon (let us call it DNA byte). Each codon consists of three nucleotides and encodes one of the 20 amino acids used in the synthesis of proteins. A "DNA sector" which encodes single protein is limited by particular start and stop codons. The information storage and copy processes are based on chemical bindings and reactions.

4. DATA DENSITY AND CAPACITY

4.1 Computer

The data density and capacity of different storage media show great divergence. Semiconductor flash modules have the capacity order of gigabytes, optical disks capacity is limited to tens gigabytes, hard disks and magnetic tapes have capacity measured in hundreds of gigabytes (and terabytes in near future). Single tape library may contain petabytes of data.

Approximately 10÷20% of the total medium or system data capacity is used to store metainformation, such as: sector addresses, file location data, spare blocks of data, servo data and so on.

A single bit on a hard disk surface uses area of thousands square nanometers. A single transistor in a flash module uses about 50 000 nm^2. Approximated areal densities of some storage media are as follows: LTO2 magnetic tape – 0.3 $Mbit/mm^2$, Blu-ray optical disk – 15 $Mbit/mm^2$, hard disk – 150 $Mbit/mm^2$. The highest volumetric storage density is currently available with a use of holographic storage (10 $Gbit/mm^2$). It is estimated that the limits of holography are near 10 $Tbit/cm^3$ [2].

Finally it must be noted that the total computer storage capacity is virtually unlimited since the available number of local and remote storage devices and systems is practically unrestricted. Furthermore let us remind that data compression techniques are extensively use to improve storage utilization.

4.2 Brain

Primarily, it must be noted that many brain capacity factors (such as coding, compression) are not known so the exact numbers which we present in this section are very rough.

The short-term memory has very limited capacity which may be improved when data are organized into separate blocks. It is assumed that the short-term memory may simultaneously contain a small number (4÷7) of pieces of information such as letters, numbers, words.

Total long-term memory capacity is an interesting topic discussed by scientists for many years. In 1937 Alan Turing suggested that human memory has limited capacity [25]. Later he estimated the brain data capacity at about 10 Gbit. Subsequently John von Neumann [20] calculated approximately that the total data stream provided to the brain throughout the lifetime is about 35 EB. More recent approximations are based on the number of neurons, synapses or on the human learning capabilities.

It is assumed that there are about 10^{11} neurons and $10^{13}÷10^{15}$ synapses. If we take for granted that each synapse is equivalent to a few bits of data (it has been found that the locust synapse transmits about 20 discrete signal levels [24]), the approximated memory capacity is in the range of 0.01÷1 PB [14, 16]. On the other hand the experimental research [10] has shown that a man is able to memorize information at the mean rate of about 2 bits/s. So, assuming that human live for 75 years ($\approx 2.4\times10^9$ s) the total amount of remembered data is much lower, about 600 MB.

Neurons have diameter from about 4 to 100 µm. Axons which connect neurons may have a length from a fraction of 1 mm to above 1 m. A single synapse at the end of an axon is several nanometers wide. On the surface of the cerebral cortex the density of neurons is near 150 000/mm^2 [6]. If we assume that a synapse is equivalent to 10 bits and a neuron has 1000 synapses then each neuron is equivalent to 10 000 bits and the cortex data areal density is about 1.5 Gbit/mm^2.

The average brain size of modern human tends to vary between sources, but typical value is near 1400 cm^3. If we assume that the brain has data capacity near 0.1 PB then its volumetric data density is about 70 MB/mm^3.

Recent experiments by Lin et. al. [12] revealed that memory coding units are functionally organized in a categorical and hierarchical manner, suggesting some forms of data compression. Brain representations of external events are not based on the exact details of the events. They have a form of selective pictures based on some associations which enables the brain to acquire large storage capacity.

Finally it must be noted that some scientists claim that the brain's capacity for storing information is virtually unlimited – no human is able to utilize the whole capacity of his brain [21]. On the other hand diminished memory capacity is a well known concomitant of aging in humans.

4.3 DNA

The DNA double helix has a diameter near 2 nm. Adjacent nucleotides are separated by 0.34 nm along the helix axis, so we may calculate that the linear density is about 5.9 Mbit/mm. In a cell during mitosis the chromosome is crammed into a more dense structure about 5 µm long. A portion of the human DNA that occupies 500 µm^3 consists of about 2×10^8 nucleotides. So the volumetric density is near 1 Mbit/µm^3.

Human genome contains about 3.1 billions nucleotides [27], so its total capacity is about 6.2 Gbit or 775 MB. It must be recalled that there is a low ratio of useful capacity to total capacity. Average human gene consists of multiple small exons separated by introns that are 10 or 100 times larger. About 98% of human DNA chain consists of repetitive DNA, transpozons and introns which do not code proteins (their functions are only partially known to DNA researchers). Since only about 2% of the genome is used to code proteins we may assume that the human genome useful capacity is about 15 MB.

5. ACCESS TIME AND TRANSFER RATE

5.1 Computer

The access time of different storage media shows a large discrepancy. It is apparent that semiconductor flash modules have the shortest access time, it is of about 0.01 ms. Hard disks have access time of about 10 ms. Magnetic tape access time is measured in 10s seconds.

Similarly internal data transfer rate of diverse storage media are different. Representative hard disk and streamer have data transfer rate at 100s Mbit/s (hard disks are about two times faster than tapes). Flash memory transfer rate approaches 100 Mbit/s.

Usually data are written in tracks and blocks/sectors and the operations are performed sequentially. On the other hand there are storage systems that use some forms of parallel operations, e.g. SDLT streamer reads 8 tracks, holographic devices let us read or write complete pages of thousands bits.

In some storage systems, like HSM (Hierarchical Storage Management) or ILM (Information Lifecycle Management) tiered hierarchy organization models are used. All data are categorized according to their age and usage rate. The tiers of storage differ in access time and cost per megabyte. Frequently used data are stored "near" the system in fast, expensive and easily accesible storage devices (such as hard disk). Old and rarely used data are stored "far" from the system in the devices that are cheap and slow (magnetic tape).

5.2 Brain

The experiments have shown that in order to write the data into the sensory memory of a human brain the stimuli should be available for at least 50 ms and the exact write process takes about 100 ms. Similarly the read process is done in milliseconds. For example the experiments have shown that 70÷80 ms after neural activity reaches the primary visual cortex the visible faces are recognized [19]. This high recognition speed is incomparable with very low speed of memorization – some experiments [10] suggest the mean rate of only several bits/s.

The electrical signals are propagated through the neurons with a speed order of 10÷100 m/s and frequency of 250÷2000 Hz. Neurons need about one thousandth of a second to return to normal state after firing an electrical pulse. The brain operations are massively parallel. According to the neural population coding concept data are represented by nature and strength of correlations among neurons. Some quantitative features describing synapse data throughput are already available. P. Simmons et al. [24] estimated that single locust synapse can transmit data at rates up to 450 bit/s. The number of discrete signal levels the synapse transmits is about 20.

There is a correspondence between the brain and HSM system. The data that are frequently used are "better fixed" inside the human memory and easier to recall (we may say: they are at high level of the HSM). On the other hand data rarely used and not used at all by a human may deteriorate and become harder to recall (they are at low level of the HSM).

5.3 DNA

During a cell replication process chromosomes are copied. The process rate is linked to the type of a cell. A single chain of nucleotide pairs in Escherichia coli (prokaryotic) cell is copied with a speed of about 500 base pairs/s, in a human (eukaryotic) cell – with a lower speed of about 50 base pairs/s.

The data copying in the eukaryotic cell is parallel. All the chromosomes inside the cell are copied simultaneously. Additionally, the process of a single chromosome replication is parallel. Serial chromosomes are divided into about thousand subunits called replicons which can be replicated in parallel, rather than serially. The complete DNA copying process in a human cell takes approximately 7 hours – this is long time in the digital world but entire terabyte tape read/write operation may also take few hours.

This parallel DNA copying process may be compared to the tape library operation. Tape library (cell) may have many drives (chromosomes) working simultaneously. Each drive, with multichannel head, may read simultaneously many tracks (replicons) from the tape with multichannel head.

6. POWER CONSUMPTION

6.1 Computer

From the power consumption point of view, we may distinguish two forms of digital storage media. First category consists of media that are able to preserve data without power consumption (e.g. flash semiconductor EEPROM, magnetic and optical disks). Energy supply is needed only during read, write or erase processes and active mode of operation. Second category consists of media that are incapable to keep data without constant energy supply (e.g. semiconductor RAM).

Modern hard disks have three basic modes of operation: active, idle and standby. The modes differ in power consumption. The representative disk power consumption in these modes is respectively 2 W, 1 W and 0.2 W. Correspondingly, semiconductor flash memory power consumption is relevant to the current operation: read/write/erase – 100 mW, standby – 20 mW. Representative streamer uses 10÷50 W during operation (all the values are approximated).

6.2 Brain

Brain performs its functions only with a constant energy supply. It is estimated that the total human brain power consumption is about 20÷25 W. The energy is used for many brain functions, including memory. In 1989 R. Merkle [15] estimated that about 10 W (like in tape drives) is used for brain data processing, which includes memory operations. Obviously there are modes of brain operation (for example active or sleep) that differ in power consumption. In some stages of a sleep the brain waves have lower frequency than in an active mode. However brain as a living organ never completely shuts down.

6.3 DNA

The quantitative data for DNA power consumption are not available. We may try to estimate the value. The human body consists of about $10^{13}\div10^{14}$ cells and total human power consumption is estimated at about 100 W. So a single cell uses roughly $10^{-8}\div10^{-9}$ mW of power. We may assume that DNA data storage process inside a cell uses not more then this level of power.

7. DATA INTEGRITY AND REDUNDANCY

7.1 Computer

Computer storage hardware consists of media and devices. There are many factors that affect media and devices stability and reliability: temperature, humidity, electromagnetic radiation, electrostatic field, ultraviolet radiation, mechanical force and so on. The medium states (e.g. magnetic remanence) that represent the data degrade very slowly and gradually with time. The comprehensive computer media (tapes, CD's) stability studies were carried by J. von Bogarth in the US National Media Laboratory [4]. The experiments have shown that in the appropriate environment storage media stability reaches hundreds years. On the other hand in unsuitable conditions (high humidity, high temperature) read errors may appear after a single month of storage.

Storage devices reliability is measured in MTBF (mean time between failures). The modern devices (disks, streamers) have MTBF equal roughly to $10^5\div10^6$ hours. Other important factor is error rate – the data read error rate of the modern storage devices is at the level of 10^{-13} to 10^{-17}.

Redundancy is a vital element of electronic storage media and systems. It is used to detect, correct and tolerate a subset of all possible errors. There are many forms of redundancy: redundant codes, CRC, ECC, redundant hardware. Redundancy may be: active, passive, cold and hot standby sparing.

Storage redundancy is multilayer. At the lowest organization layer, there are redundant codes such as EFM (eight to fourteen modulation) used in optical disks. At the data sector layer, classic forms of redundancy are CRC codes and spare sectors on hard disks. If a particular hard disk (or flash) sector is damaged then a spare sector in a different location is used. At the higher organization layer disks may form RAIDs (redundant array of inexpensive disks), data may be backed up and replicated. Occasionally complete information systems are duplicated.

7.2 Brain

The research on synapses operation indicates that they are very prone to errors. Available research data [13] point out that single neuron pulses often fail to cross synapses; in fact, the probability of synapse proper operation can be less than 0.1.

In the natural course of life human memories grow weaker with time. It is assumed that the process is a vital element of healthy brain. Without it brain would soon become overwhelmed by the remembered data. Additionally there are many causes and many sorts of amnesia and other memory disorders: death of brain cells (during an average human life about 20% of neurons die naturally), strokes of brain, mechanical brain damage. It must be observed that studying the disorders helps to enrich our knowledge of brain functions.

The cerebral redundancy is multilayer. Some forms of coding redundancy have been found, for example E. Mukamel [18] found redundancy in the retinal code used to transfer visual data to the brain. It is assumed that the neuron bursting is used to make synapses highly reliable. Single pulses often fail to cross synapses. Bursts ensure a large degree of neurotransmitter release and higher probability of subsequent transmission [13]. Redundancy is used at the neuron organization level – if a given neuron dies the brain automatically undergoes a reconfiguration in which new connections are made to get around useless neuron (similarly to sector replacement in a hard disk). Some years ago scientist discovered that brains in adult and old people undergo neurogenesis. The new neurons appear in the hippocampus and connect themselves with existing circuits.

Redundancy is also available at brain parts stratum – when one part of the brain is damaged the brain reorganizes and the functions previously performed by this part are transferred to the other parts of the brain. The particular brain regions may alter their functions. For example, in people who lose their sight at a young age, the visual cortex processes touch or sound. These cerebral forms of redundancy are correlated to hardware standby sparing.

7.3 DNA

The DNA is as stable as its living carrier (bacteria, plant or animal cell). In some cases after the cell death the DNA data may survive. In suitable conditions the DNA may hold its data for years. On the other hand the DNA is vulnerable to hydrolisis and oxydation. Many factors lead to mutations in the genes. In mutations, letters of the genetic code can be changed and stretches of DNA can be deleted. The mutation rate of DNA replication process of Escherichia coli is near 10^{-9}.

Some biological cells are more resistant to environmental conditions than others. For example bacteria such as Deinococcus radiodurans are especially good at surviving extreme conditions. They can tolerate high temperatures, desiccation and ultraviolet and ionising radiation doses about 1000 times higher than fatal ones to humans [28].

The DNA data protection mechanism has two basic layers. The genetic code is to some extent redundant and the genetic information is massively replicated. Thanks to the redundancy of the genetic code some copy errors are tolerated. Most of the amino acids are encoded by more than one codon (in total 61 codons are used to code 20 amino acids), for example alanine is coded by four different codons. On the other hand there are amino acids (such as methionine) which are coded by single codons. Furthermore, the same DNA information is replicated in every cell of the particular living organism.

8. SUMMARY AND CONCLUSIONS

It is possible to quantitatively compare three storage media only in a vague way. Primarily, the actual values of parameters are dependent on many factors. Secondly, the brain memory features are not already precisely specified. Finally, the attributes of the three media differ to some extent. On the other hand there are functional analogies between the systems. In the following tables the storage comparison is presented. In the case of quantitative features orders of values are included.

Table 1. Digital storage parameters

Parameter	Comments
Basic element	transistor, magneti/optical domain
Basic element size	$1000 \div 10000$ nm^2
Addressing scheme and organization	discrete, hierarchical
Redundancy	multilayer, many different forms
Capacity	1 GB÷1 TB, virtually unlimited
Linear density	10 kbit/mm (HD)
Areal density	100 Mbit/mm^2 (HD) 1 Gbit/mm^2 (hologr.)
Volumetric density	10 Gbit/mm^3 (holography limits)
Access time	0.01ms (flash) 10 000 ms (tape)
Internal data transfer	100 Mbit/s (HD)
Power consumption	10^4 mW (streamer) $10^1 \div 10^2$ mW (flash)
Operation voltage	10^4 mW (streamer) $10^1 \div 10^2$ mW (flash)
Error rate	$10^{-13} \div 10^{-17}$

Table 2. Brain storage parametrs (assumed values in italics)

Parameter	Comments
Basic element	neuron/synapse
Basic element size	10 nm^2 (synapse)
Addressing scheme and organization	associative, hierarchical
Redundancy	redundant code, standby sparing
Capacity	unlimited for human needs (?)
Linear density	-
Areal density	*1 Mbit/mm^2* (cortex density)
Volumetric density	*0.1 Gbit/mm^3*
Access time	10 ms
Internal data transfer	100 bit/s (single synapse)
Power consumption	10^4 mW (brain)
Operation voltage	10 mV (single axon)
Error rate	10^{-1} (single pulse/synapse)

Table 3. DNA storage parameters

Parameter	Comments
Basic element	nucleotide
Basic element size	1 nm^3
Addressing scheme and organization	sequential, parallel copying
Redundancy	redundant code, data replication
Capacity	0.1 GB (human genome)
Linear density	1 Mbit/mm
Areal denisty	-
Volumetric density	1 $Pbit/mm^3$
Access time	-
Internal data transfer	100 kbit/s (DNA copying)
Power consumption	$10^{-8}\div10^{-9}$ mW (single cell)
Operation voltage	not applicable
Error rate	10^{-9}

A question is, if it is possible to digitally simulate biological storage and what kind of storage devices and systems are the best approximations/equivalents of biological storage? Brain storage system is very complex and only partially known. It seems that the best candidate for brain storage model is digital storage grid with dense interconnection network and associative (holographic) storage devices in the grid nodes. We assume that the grid operations are massively parallel. Perhaps this model together with neural network based processing will help to simulate all brain functions. Sequential structure with parallel copying is an important DNA feature. Similar sequential structure and parallel read and write are available in tape libraries with many drives and multichannel heads.

In 1998 Moravec [17] estimated that the required hardware to match intelectual performance of the human brain will be available in cheap computers not earlier than in 2020s. Now we have 2007 and contemporary, cheap computers use storage media that are, from many points of view, better than human memory. They use less power, they have superior data density, higher internal transfer and shorter access time. Some human brain properties such as associative addressing turn out to be available for computer systems with a holographic storage in near future. There are still tasks (e.g. image recognition) in which computers are worse than brain which has been optimized by thousands years of evolution.

9. REFERENCES

[1] Adleman L. Molecular Computation of Solutions to Combinatorial Problem. *Science*, 266., 1994, 1021-1024.

[2] Ashley J., Bernal M.-P., Burr G. W., Coufal H., Guenther H., Hoffnagle J. A., Jefferson C. M., Marcus B., Macfarlane R. M., Shelby R. M., Sincerbox G. T. Holographic Data Storage. *IBM Journal of Research and Development*, vol. 44, number 3., 2000, 341-368.

[3] Atkinson R., Shiffrin R. Human memory: A proposed system and its control processes. *The psychology of learning and motivation. Advances in research and theory*, Vol. 2. Academic Press, 1968.

[4] Bogart van J. W.C. *1996 Media Stability Studies*. NML Final Report. National Media Laboratory, 1996.

[5] Borst A., Theunissen F.E. Information theory and neural coding. *Nature Neuroscience* 2, 1999, 947-957.

[6] Calvin W.H. *How Brains Think*. Orion Publishing Group, 1996.

[7] Graham-Rowe D. World's first brain prosthesis revealed. *NewScientist*, 12 March 2003. http://www.newscientist.com/news/news.jsp?id=ns99993488

[8] Hebb, D.O. *The organization of behavior*. Wiley, New York, 1949.

[9] Kann J.L., Canfield B.W., Jamberdino A.A., Clarke B.J., Daniszewski E., Sunada G. Mass Storage and Retrieval at Rome Laboratory. *Papers from the 5th NASA Goddard Conference on Mass Storage Systems and Technologies*, 1996.

[10] Landauer T.K. How Much Do People Remember? Some Estimates of the Quantity of Learned Information in Long-term Memory. *Cognitive Science* 10, 1986, 477-493.

[11] Lin L., Chen G., Kuang H., Wang D., Tsien J.Z. Neural Encoding of the Concept of Nest in the Mouse Brain. *Proceedings of the National Academy of Sciences USA*, t. 104, No. 14, 2007, 6066-6071.

[12] Lin L., Osan R., Tsien J.Z. Organizing Principles of Real-Time Memory Encoding: Neural Clique Assemblies and Universal Neural Codes. *Trends in Neuroscience* 29, 2006, 48-57.

[13] Lisman J.E. Bursts as a unit of neural information: making unreliable synapses reliable. *Trends in Neuroscience* 20, 1997, 38-43.

[14] Matick R.E., *Computer Storage Systems and Technology*, John Wiley & Sons, New York, 1977.

[15] Merkle R.C. Energy limits to the computational power of the human brain. *Foresight Update* No. 6. August 1989.

[16] Merkle R.C. How Many Bytes in Human Memory. *Foresight Update*, No. 4. October 1988.

[17] Moravec H. When will computer hardware match the human brain. *Journal of Evolution and Technology*, vol. 1, 1998, http://www.jetpress.org/volume1/moravec.pdf.

[18] Mukamel E.A., Schnitzer M.J. Retinal coding of visual scenes – repetitive and redundant too? *Neuron*, May 5; 46 (3), 2005, 493-504.

[19] Mundel T., Towle V.L., Dimitrov A., Wilson H.W., Pelizzare C., Torres I.J., Spire J.-P., Milton J.G. Human face perception in transient prosopagnosia. *Neuroimage* 7, 1998, 349.

[20] Neumann von J. *The computer and the brain*. Yale University Press, New Haven, 1958.

[21] Rathus S.A. *Psychology in the New Millenium*. Wadsworth, 1999.

[22] Schalk G., McFarland D.J., Hinterberger T., Birbaumer N., Wolpaw J.R. BCI2000: A General-Purpose Brain-Computer Interface (BCI) System. *IEEE Transactions On Biomedical Engineering*, Vol. 51, No. 6. June 2004, 1034-1043.

[23] Silva A.J., Josselyn S.A., Cognitive neuroscience: The molecules of forgetfulness, *Nature* 418, 29 Aug. 2002, 929-930.

[24] Simmons P.J., de Ruyter van Steveninck R. Reliability of Signal Transfer at a Tonically Transmitting, Graded Potential Synapse of the Locust Ocellar Pathway. *The Journal of Neuroscience*, 25(33). August 17 2005, 7529-7537.

[25] Tulving, E. Cue-dependent forgetting. *American Scientist*, (62), 1974 74-82.

[26] Turing A. On Computable Numbers, with an Application to the Entscheidungsproblem. *Proc. Lond. Math. Soc*. (2), 42, 1937.

[27] Watson J.D., Berry A. *DNA: The Secret of Life*. Knopf, 2003.

[28] Wong P., Wong K., Foote H. Organic data memory using the DNA approach. *Communications of the ACM*, New York, Vol. 46, Iss. 1. Jan. 2003, 95.