Digital Investigation 20 (2017) S12–S19

Contents lists available at ScienceDirect

Digital Investigation

journal homepage: www.elsevier.com/locate/diin

DFRWS 2017 Europe — Proceedings of the Fourth Annual DFRWS Europe

Bit-errors as a source of forensic information in NAND-flash memory

Jan Peter van Zandwijk*

Netherlands Forensic Institute, Laan van Ypenburg 6, 2497 GB The Hague, The Netherlands

ARTICLE INFO

Article history:
Received 26 January 2017
Accepted 26 January 2017

Keywords:
NAND-flash reliability
P/E cycles
Retention bit-errors
Error-correcting codes
Hardware forensics

ABSTRACT

The value of bit-errors as a source of forensic information is investigated by experiments on isolated NAND-flash chips and USB thumb-drives. Experiments on isolated NAND-flash chips, programmed directly using specialized equipment, show detectable differences in retention bit-errors over forensically relevant time periods with the device used within manufacturer specifications. In experiments with USB thumb-drives, the controller is used to load files at different times onto the drives, some of which have been subjected to stress-cycling. Retention bit-error statistics of memory pages obtained by offline analysis of NAND-flash chips from the thumb-drives are to some extent linked to the time files are loaded onto the drives. Considerable variation between USB thumb-drives makes interpretation of bit-error statistics in absolute sense difficult, although in a relative sense bit-error statistics seems to have some potential as an independent side-channel of forensic information.

Introduction

Due to high storage density, low access latency and low costs, NAND-flash has become the medium of choice for non-volatile data storage. This holds for both low-end consumer electronics as well as for high-end server based storage environments. The latter increasingly use state-of-the-art solid state disks (SSDs) containing NAND-flash memory chips for persistent data storage. With increased proliferation of SSDs in server based storage environments, more attention has been devoted to reliability and endurance of the underlying NAND-flash memory chips (Meza et al., 2015). This all the more important, since increase of NAND-flash storage capacity goes hand in hand with a decrease in both NAND-flash reliability and endurance. Decrease in reliability means that bit-errors are more likely to occur in NAND-flash using smaller technology and in NAND-flash which stores multiple bits in a single cell (i.e. multi level cells, MLC, or triple level cells, TLC) than in NAND-flash which stores a single bit per cell (single level cells, SLC). Also, endurance of NAND-flash decreases with the size of the technology and the use of MLCs and TLCs. The number of times a cell can reliably be programmed and erased (usually called P/E cycles) typically decreases from 100.000 for SLCs to 5000 or less for MLCs and TLCs.

At the chip level, much research has been devoted to the characterization of mechanisms causing bit-errors in NAND-flash (see e.g. Cai et al., 2012). These mechanisms include retention time, read- and write-disturb and erase errors. Retention time bit-errors occur when information stored in NAND-flash memory cells changes over time. The mechanism responsible for this is the leakage of charge stored in the cells. Read and write-disturb errors occur when information in a cell is changed due to the fact that *neighbouring* cells are being read or programmed. Erase errors occur when an erase operation on a NAND-flash memory page fails to reset the page to the erased state. In that case, the page is marked as bad and discarded for further data storage. These error mechanisms all depend on the number of P/E cycles a cell has undergone. Due to repeated programming and erasing, structures in the NAND-flash memory cells tend to deteriorate, thereby increasing the susceptibility of that cell to one of the errors mechanisms and increasing the rate at which bit-errors develop. Besides P/E cycles, it is known that susceptibility of a cell to bit-errors is affected by the actual value of the data stored in the memory cell. As an example, fully programmed cells will tend to loose charge more quickly than cells which are in the erased state. Therefore, fully programmed cells are more prone to retention bit-errors than fully erased ones. In the literature, retention time is generally reported to be a larger source of bit-errors than the other error mechanisms mentioned above (e.g. Cai et al., 2012).

In order to circumvent problems associated with NAND-flash reliability, manufacturers use randomization techniques and error-correcting codes (ECCs). The former is aimed at reducing the number of errors occurring due to cell-to-cell interference (i.e.

* Fax: +31 70 8886559.
E-mail address: j.p.van.zandwijk@nfi.minvenj.nl.

http://dx.doi.org/10.1016/j.diin.2017.01.005

read- and write-disturb errors), while the latter provides a mechanism for correcting any errors that might have occurred in the data. To increase NAND-flash endurance, it is important to ensure that all memory pages have undergone about the same number of P/E cycles. Manufacturers employ sophisticated *wear leveling techniques* to accomplish this and to ensure an even spread of wear throughout the NAND-flash.

The extent to which a digital forensic practitioner is confronted with issues of NAND-flash reliability and endurance depends on the type of acquisition being performed. In many investigations, data is acquired from NAND-flash through the controller, a piece of hardware providing an interface between the NAND-flash and the host operating system. In that case, the controller takes care of error-correction by application of ECCs. It also takes care of data randomization and wear leveling, which therefore occur completely transparent to the user. In cases where acquisition through the controller is impossible, however, one must resort to low-level acquisition, i.e. reading data directly from NAND-flash memory using specialized equipment. In those cases, data is obtained which still includes randomization and possibly bit-errors. In a recent paper (Van Zandwijk, 2015), techniques are developed to process raw NAND-flash dumps and perform offline de-randomization and error-correction. As a by-product, these techniques yield detailed statistics of number of bit-errors per memory page. Given their origin, these bit-errors are related to processes that are of potential forensic interest, such as e.g. retention time and the number of times data has been read. This in turn suggests that statistics of bit-errors might be useful as a (noisy) source of forensic information and can potentially serve as an independent side-channel conveying for instance time information.

The purpose of this study is to assess the value of bit-error statistics in NAND-flash as a source of forensic information. For bit-error statistics to be of any use in a forensic context it needs to be established whether detectable changes in number of bit-errors occur over forensically relevant time periods and under normal use of the device. In evaluating NAND-flash reliability, researchers have subjected memory chips to variable number of P/E cycles and have baked chips at various temperatures to simulate different retention times. In these experiments, often the largest effects on bit-error rates are observed for experimental conditions well outside factory specifications. Therefore, it is important to see whether any detectable effects are present under less extreme conditions. For this purpose raw NAND-flash chips will be subjected to moderate data-retention and stress-cycling (i.e. inducing variable number of P/E cycles), while using specialized equipment to read and write data, thereby replacing the role of the controller.

Next, the value of bit-errors for forensic purposes will be investigated under experimental conditions mimicking more or less more normal use of a storage device. To this end, experiments are performed on USB thumb-drives, some of which are stress-cycled. In the experiments, files are added to the drives through the controller at different retention times and it is investigated to what extent these differences in retention times are reflected in the statistics of bit-errors present in memory pages of files on the underlying NAND-flash. As a first step, retention time experiments seem to be a suitable research topic since retention bit-errors are known to be the largest in magnitude and therefore presumably the most easy to detect.

Materials and methods

NAND-flash memory chips

Experiments are performed on raw multi-level cell NAND-flash chips (Hynix H27UAG8T2ATR) with a storage capacity of 2 GB, shipped in TSOP package. According to manufacturer documentation and ID information read from the chip, it is produced with 41 nm technology and contains memory-pages of 4320 bytes, consisting of 4096 data bytes and 224 spare bytes. For this device, the size of an erase block (i.e. the smallest number of pages which can be erased simultaneously) is 128 pages. Manufacturer documentation specifies data retention time of 10 years and a number of P/E cycles of 5000, when a 12 bit error-correction code is applied to each 512 bytes of data.

A second series of experiments is performed on cheap USB thumb-drives (Emtec Colormix) with a specified storage capacity of 4 GB. Initial research was performed on one test specimen to determine the properties of the NAND-flash contained in the thumb-drive. After opening, this test device was found to contain a Phison PS2251-68-5 controller and an unmarked NAND-flash in a TSOP package. Internet search revealed that ID-bytes read from the NAND-flash match those of a Toshiba TC58TEG6DDLTA00 multi-level cell device. Unfortunately, no factory documentation on this type of NAND-flash device was found and therefore information specifying technology size, retention time and number of P/E cycles is lacking.

In order to determine parameter values for data randomization and ECC added by the controller when storing data on the NAND-flash of the Emtec Colormix device, the test device was loaded with known data through the controller. Next, the NAND-flash was de-soldered from the printed circuit board and read using equipment developed at the Netherlands Forensic Institute (NFI). Using methods described in Van Zandwijk (2015), parameter values for data randomization and error-correcting code applied by the controller were reverse engineered from a raw dump of the content of the NAND-flash. For this controller, no LFSR-based randomization was found to be in use. However, on the basis of known data, it was possible to explicitly reconstruct the complete byte-sequence used for data randomization. This sequence was used in subsequent experiments for de-randomization of data. Table 1 summarizes parameters found for the device. The meaning of the symbols is briefly explained in the table. For a more extensive description, the reader is referred to Van Zandwijk (2015).

Experimental protocols

Raw NAND chips

Raw NAND-flash chips were programmed and read using equipment developed at the NFI. In order to investigate the magnitude of retention bit-errors on forensically relevant timescales, the following two protocols are applied to Hynix H27UAG8T2ATR NAND-flash chips.

Raw NAND data retention protocol. Two fresh NAND-flash chips were completely programmed with random data a single time and kept in Petri dishes, either at room temperature or at 70 °C on a heating stove. Baking chips at higher temperatures is a technique often adopted to simulate longer retention times (see e.g. Cai et al., 2012; Niset and Kuhn, 2005). Using the Arrhennius model, one can compute an acceleration factor (AF) to extrapolate retention times at the higher temperature back to retention times at e.g. room temperature. Using parameters from Niset and Kuhn (2005), the AF of 70 °C with respect to room temperature is calculated to be about 60.

Periodically, the two chips were read and number of bit-errors computed by comparing read data to the data originally written to the chip. Number of bit-errors is expressed as bit-error-rate (BER) = number of bit-errors/number of bits.

Raw NAND stress-cycle protocol. Two fresh NAND-flash chips were subjected to a stress-cycle protocol. This protocol is aimed at

Table 1
Parameters determined for NAND-flash present in Emtec Colormix USB thumb-drive.

Specified storage capacity	4 GB
Capacity (bytes)	9,261,023,232[a]
Pagesize (bytes)	17,664
Datachunksize (bytes)[b]	1024
Number of datachunks per page	16
Sparesize per datachunk (bytes)	74
Codeword formatting	[Spares in byte reversed order | scrambled data in byte reversed order]. Each byte in the codeword is bit-reversed.
Code parameters	BCH-code over GF (2^{14}). n = 8824,k = 8208,t = 44[c]

[a] Although the device is sold as a 4 GB USB-thumb drive, the batch used in this research is found to contain a 8.6 GiB NAND flash.
[b] Memory pages in NAND flash are subdivided into a number of smaller data areas, called datachunks here.
[c] n is the size of the codeword of the binary BCH-code found to be in use, measured in bits. k is the dimension of the code and t the number of biterrors the code is able to correct.

creating regions in the chip which have undergone a different number of P/E cycles. To this end, specified areas in the chip are repeatedly programmed with random data and erased directly afterwards. By varying the areas being programmed and erased, four regions in the chip are created, which have undergone respectively 0, 500, 1200 and 2500 P/E cycles. Note that these number of P/E cycles are all well below the manufacturer specified maximum number of 5000 P/E cycles.

After stress-cycling, the two chips were erased completely and completely reprogrammed with fresh random data. Next, chips were kept in Petri dishes, one at room temperature and the other at 70 °C, and periodically read. Again, BER is computed in the same way as in the raw NAND data retention protocol.

USB thumb-drives

USB thumb-drives were accessed through the controller. To this end, the thumb-drive was connected to a Virtual Machine running Ubuntu 9.10 and files containing random data were copied to and deleted from the drive using standard Linux commands. Care was taken to explicitly flush all filesystem buffers after each command. Thumb-drives were subjected to the following two protocols.

USB thumb-drive data retention protocol. Five files containing random data were copied onto two unused thumb-drives at times specified in Table 2. In between, drives were kept in Petri dishes, one at room temperature, the other at 70 °C on a heating stove.

USB thumb-drive stress-cycle protocol. As in the USB data retention protocol, initially three files of 1 GB containing random data are copied onto two unused thumb-drives at times specified in Table 2. As in the USB thumb-drive data retention protocol, one drive is stored at room temperature between addition of files, the other is kept at 70 °C on a heating stove. After addition of the first three files, a file of 628 MB is 200 times copied onto each drive and deleted directly after copying. Owing to the fact that 3 GB of each drive is occupied throughout experiments with the first 3 files, there is no space left for wear leveling and the controller is most likely forced to repeatedly use the same remaining space on the drive for storage of the 628 MB file, thereby inducing wear in that area of the drive. After stress-cycling, two files containing 314 MB random data are copied to the drives at times described in Table 2.

USB thumb-drive NAND-flash processing and data analysis

Directly after addition of the last file in both the data retention and the stress-cycle protocol, NAND-flash chips were desoldered from the printed circuit board of the thumb-drives. Subsequently, a raw dump of content of the NAND-flash was produced using NFI equipment. Raw dumps of NAND-flash chips were further processed offline. This entails to de-randomizing the content of memory pages and applying ECC information to correct bit-errors in the data. As a by-product, the latter produces statistical data on the number of bit-errors per memory page. In order to attribute the content of individual memory pages to files copied onto the thumb-drive, SHA1 hashes of 1 k datachunks in memory pages are computed after de-randomization and decoding. Subsequently, these hashes are compared to piecewise 1 k SHA1 hashes of the original files, similarly as described in Breeuwsma et al. (2007). A memory page is attributed to a specific file when the SHA1 of at least 14 of the 16 datachunks matches the SHA1 of a datachunk from that file.

By attributing the content of a memory page to a file copied onto the thumb-drive, one at the same time assigns statistical information on the number of bit-errors in that page to the file. Bit-error information from all memory pages assigned to a file can be aggregated to form the distribution of number of bit-errors for memory pages of that file. Comparing the distributions of bit-errors in memory pages of each of the five files copied onto the thumb-drives can be used to investigate the question to what extent it is possible to use the number of bit-errors in a given memory page as a (noisy) source of information to what file the page belongs. Indeed, if distributions of bit-errors of memory pages would be more or less non-overlapping for different files, it would be very well possible to use bit-error statistics to assign memory pages to a specific file. Conversely, if distributions of bit-errors of two different files are strongly overlapping, it would be impossible to distinguish between the two on the basis of bit-error information alone.

Table 2
Details of experimental protocols used for Emtec Colormix USB thumb-drives.

Number	Time (hours)	Data retention protocol	Stress-cycle protocol
1	0	Added file 1 (1 GB)	Added file 1 (1 GB)
2	22	Added file 2 (1 GB)	Added file 2 (1 GB)
		Added file 3 (1 GB)	Added file 3 (1 GB)
3	70	–	Start stress-cycle protocol. File 6 (628 MB) is copied and deleted 200 times.
4	77	Added file 4 (314M)	Added file 4 (314M)
5	101	Added file 5 (314 M)	Added file 5 (314 M)
6	101	End experiment, chip desoldered.	End experiment, chip desoldered.

Between additions of files, the thumb-drives are stored in glass Petri dishes at either room temperature or 70 °C.

Results

Raw NAND-flash chips

Fig. 1 shows results of applying the data retention protocol to raw Hynix H27UAG8T2ATR NAND-flash chips. This figure shows BER as a function of time for chips kept at room temperature and at 70 °C. Detectable changes in BER are found to occur over periods of days to weeks, which seem to be relevant timescales in a forensic investigation. Note the differences in rate at which BER increases over time between NAND-flash chips kept at room temperature and kept at 70 °C.

Results of the raw NAND-flash stress-cycle protocol are shown in Fig. 2 for chips kept at room temperature and kept at 70 °C. The figure shows the block averaged BER as a function of blocknumber for a number of specified retention times. See figure legend for details. Note that for both temperatures, differences in P/E cycles of specific regions in the NAND-flash chip lead to clearly detectable differences in the rate at which bit-errors develop over time after reprogramming with fresh random data.

USB-thumb drives

Table 3 presents statistical data on the number of memory pages which could be attributed to files copied to the USB thumb-drives after offline data processing. For both protocols, at least 82% of the data of files copied onto the thumb-drive could be attributed to memory pages on the basis of SHA1 hashes using offline derandomization and decoding of raw dumps produced from desoldered NAND-flash chips. Somewhat surprisingly, the batch of Emtec Colormix USB thumb-drives used in this study is found to contain 8 GB NAND-flash chips, while content of the thumb-drive is specified to be 4 GB. From analysis of raw dumps, it was found that a large portion of the memory pages in the NAND-flash remain empty throughout both protocols (see Table 3). The reason for this is not clear, but possibly some hardware constraint might be used to limit the size of the NAND-flash available to the controller for storage.

Note that in Table 3 the number of empty pages is roughly the same for all four thumb-drives investigated. Besides this, note that less than 20% of the file used in stress-cycling (file 6 in Table 3) is found back on the NAND-flash after processing. Both can be taken as an indirect indication that during the stress-cycling protocol pages have been repeatedly erased and reprogrammed. More research would be necessary to gain a more thorough understanding of the actual amount of P/E cycles pages have been subjected to (see also Section "Discussion and conclusions").

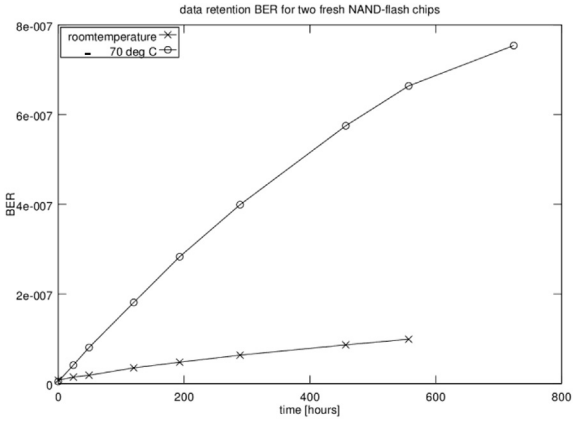

Fig. 1. Results of raw NAND-flash data retention protocol applied to two fresh Hynix H27UAG8T2ATR NAND-flash chips, programmed once with random data and periodically read. Number of bit-errors is expressed as BER.

Figs. 3 and 4 show bit-error statistics of memory pages for thumb-drives kept at respectively 70 °C and room temperature for the two protocols used in this study. In each figure, colour coding is used to designate files loaded onto the thumb-drive. Note that memory pages belonging to the same file can get spread out throughout the NAND-flash memory due to wear leveling techniques applied by the USB-controller when storing data onto NAND-flash. Comparing the top panel from Fig. 4 with other panels from Figs. 3 and 4, it can be seen that there is considerable variability between the areas in the NAND-flash chips used by the controller for data storage. For the drives shown in Fig. 3, the first three files added onto the drives appear to be stored in two separated areas of the NAND-flash, while only one area appears to be used for storage of these files in the top panel of Fig. 4. Possibly, this indicates the use of two separate planes or dies in the NAND-flash. Similarly, the averaged bit-error level per page also varies between the four USB thumb-drives studied (compare e.g. the two panels of Fig. 3). To some extent, the base-line bit-error level appears to be related to the temperature at which the drives are stored. For some files copied onto the drives, it can be seen from Figs. 3 and 4 that there is a definite interrelationship between the number of bit-errors in memory pages belonging to that file (e.g. file 5, which is added last, colour coded as purple), which holds for pages located in different areas of the chip. This interrelationship becomes more pronounced when bit-errors are averaged over blocks of 256 pages (i.e. the black dots in Figs. 3 and 4). On the basis of earlier experiments (see Section "Experimental protocols") we believe that 256 pages is the actual size of an erase block for this type of NAND-flash chip, but independent confirmation of this is lacking.

The effect of stress-cycling on bit-error rates can be judged by comparing bit-error statistics of file 4 (light blue) and file 5 (purple), which were both added onto the drives after stress-cycling. As can been seen from Figs. 3 and 4, memory pages of these two files appear to be stored more fragmented onto the two drives which have been subjected to stress-cycling than in the other drives. In the latter ones, these two files are stored more sequentially, as is the case for the three files which have been added first. Due to significant variation between behaviour of NAND-flash chips studied, it is difficult to judge the effect of stress-cycling on bit-error rate in absolute terms. Qualitatively, it seems to be somewhat higher in the stress-cycled drives.

Figs. 5 and 6 show the distribution of page-wise bit-errors for all five files copied onto the thumb-drives for the two drives which have been subjected to stress cycling, i.e. the top panels of Figs. 3 and 4. Top panels of Figs. 5 and 6 contain page-wise distributions of bit-errors for different files copied onto the drives, while bottom panels show bit-errors averaged over blocks of 256 pages. Note that the former distributions are broader and more overlapping than the latter ones, which have smaller widths and are more separated.

As explained earlier, if distributions are non-overlapping, number of bit-errors can be used as a means to discriminate between different files. Note that separation between distributions for individual files is larger when bit-errors are averaged over blocks of 256 pages, the presumed size of an erase block. The data shown in the bottom panels of Figs. 5 and 6 suggest, that it would be possible to select memory pages belonging to file 5 (last added, colour coded as purple) on the basis of block averaged bit-error statistics. For this purpose, the page-wise bit-error distributions, shown in the top panels, appear to be less suited because page-wise bit-errors are more noisy and, hence, lead to broader distributions.

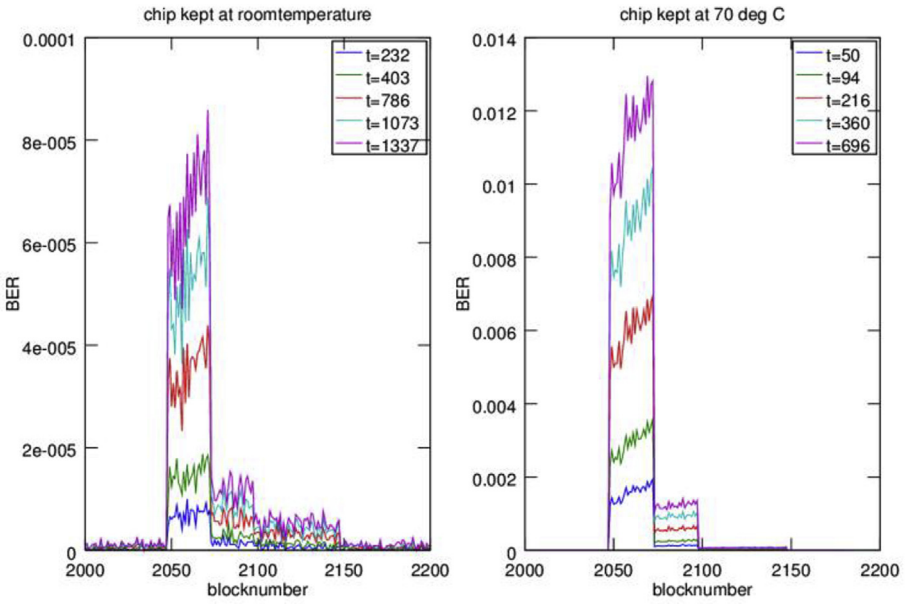

Fig. 2. Block averaged BER (i.e. number of bit-errors per erase block/number of bits in an erase block) for raw NAND-flash as a function of number of the corresponding erase block. An erase block contains 128 pages of 4320 bytes each. In both panels, time in hours corresponding to the traces is indicated in the legend. *Left panel:* data for chip kept at room temperature. *Right panel:* data for chip kept at 70 °C. Due to the stress-cycle protocol, blocks have been subjected to the following number of P/E cycles: Blocknumber <2047: 0 P/E cycles; Blocknumber 2048–2072: 2500 P/E cycles; Blocknumber 2073–2097: 1200 P/E cycles; Blocknumber 2098–2147: 500 P/E cycles; Blocknumber >2148: 0 P/E cycles.

Table 3
Number of memory pages attributed to files in USB thumb-drive stress-cycle and data retention protocols.

	Temperature = 70 °C		Room temperature	
	Stress-cycle	Data retention	Stress-cycle	Data retention
File 1 (1 GB)	59,648 (90%)	59,904 (90%)	55,040 (82%)	59,392 (89%)
File 2 (1 GB)	64,256 (97%)	64,256 (97%)	63,744 (96%)	64,768 (97%)
File 3 (1 GB)	64,768 (97%)	65,280 (98%)	65,280 (98%)	65,278 (98%)
File 4 (314 M)	17,920 (89%)	19,200 (96%)	17,024 (84%)	18,560 (92%)
File 5 (314 M)	18,304 (91%)	18,944 (94%)	17,535 (87%)	19,838 (99%)
File 6 (628 M)	6912 (17%)	0 (0%)	5896 (15%)	0 (0%)
# Empty pages	291,303	295,399	298,859	294,865

Memory pages in descrambled and decoded raw dumps of NAND-flash are attributed to files using SHA1 hashes. File numbers are the same as described in Table 2. Percentages in the table are computed with respect to the total number of memory pages present in the corresponding file (i.e. 66,536 for files 1–3, 20,096 for files 4–5 and 40,192 for file 6). Note that file 6 is only used for stress-cycling and is deleted from the drive afterwards. Therefore, no memory pages belonging to this file are found in dumps pertaining to the data retention protocol and only a small number of pages, which are remnants of the file, in the stress-cycle protocol.

Discussion and conclusions

In this study we set out to investigate the value of page-wise bit-error statistics as a source of forensic information in NAND-flash. For this purpose, retention bit-errors were induced in both isolated raw NAND-flash chips and in USB-thumb drives. As explained earlier, besides retention time, also reading and writing of data can introduce bit-errors. More research and additional experimentation is required to shed light on the relative contribution of these sources of bit-errors. This is all the more important because contributions of different bit-error mechanisms might depend on the specific use of the NAND-flash. For instance, USB thumb-drives and SSDs are primarily used for data storage, which means that typically only user-assisted writes are performed on the NAND-flash. Then, data is kept for some time and subsequently read a limited number of times. Therefore, it seems to be plausible that data retention is the primary source of bit-errors in these devices. On the other hand, the use of NAND-flash in tablet computers or smartphones might be different. Containing data belonging to the operating system of the device, the NAND-flash in these devices might be exposed to many automatic reads and writes. This different usage might have an effect on relative contribution of different sources of bit-errors in NAND-flash.

In experiments on raw NAND-flash chips, detectable changes in bit-error statistics were observed over forensically relevant time periods while using devices within manufacturer specification. This is encouraging, since in research specifically aimed at investigating NAND-flash reliability, chips are often tested at or beyond manufacturer specifications. The data from Fig. 2 shows that BER depends both on retention time and the number of P/E cycles a specific area of the chip has been subjected to. These results indicate that stress-cycling a NAND-flash by a limited number of P/E cycles can already lead to detectable differences in the rate at which bit-errors develop. From the viewpoint of forensic data extraction, this can be taken as a hint that fresh NAND-flash might be less susceptible to perturbations than NAND-flash that has been used for some time and therefore has presumably been exposed to some stress-cycling. This in turn could suggest that response of a NAND-flash chip from a newly bought reference device to potentially disturbing events such as X-raying or desoldering might be different from that of a NAND-flash chip from an exhibit that has been used for some time.

Experiments on four USB thumb-drives containing the same controller and an unmarked NAND-flash chip show considerable inter-chip variability (Figs. 3–6). This variability makes it, of course, very difficult to interpret the (block averaged) number of bit-errors in an absolute sense as a measure of the time data has been present in a certain memory page. In a relative sense, however, bit-error statistics seem to offer some perspective for use in a forensic context.

Data from Figs. 3–6 show the possibility to link, at least qualitatively, (block averaged) number of bit-errors to files copied onto the drives by attributing the corresponding data to that file. By using files containing random data, it is easy to unequivocally attribute data from memory pages to specific files. In real life

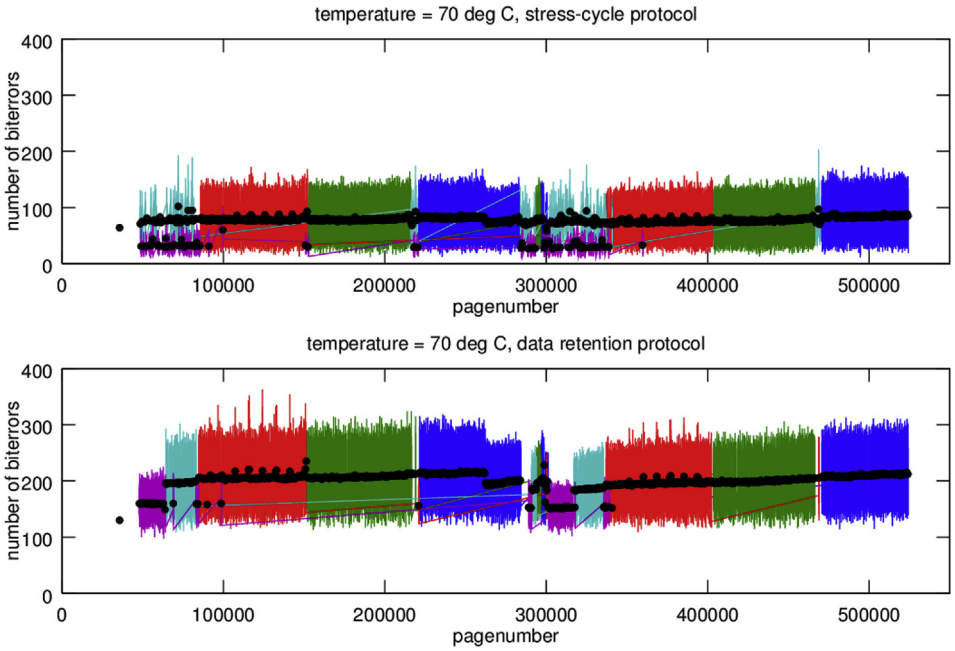

Fig. 3. Pagewise bit-error statistics for two USB thumb-drives stored at 70 °C for files copied onto drive in stress-cycle protocol (*top panel*) and data retention protocol (*bottom panel*). In each panel, files described in Table 2 are colour coded as follows: file1, file2, file3, file 4, file 5. Data pertaining to file 6 (used in stress-cycling) is not shown. Black dots indicate bit-errors averaged over blocks of 256 pages. For comparison with Figs. 1 and 2: 100 bit-errors per page corresponds to a BER of $7 \cdot 10^{-4}$.

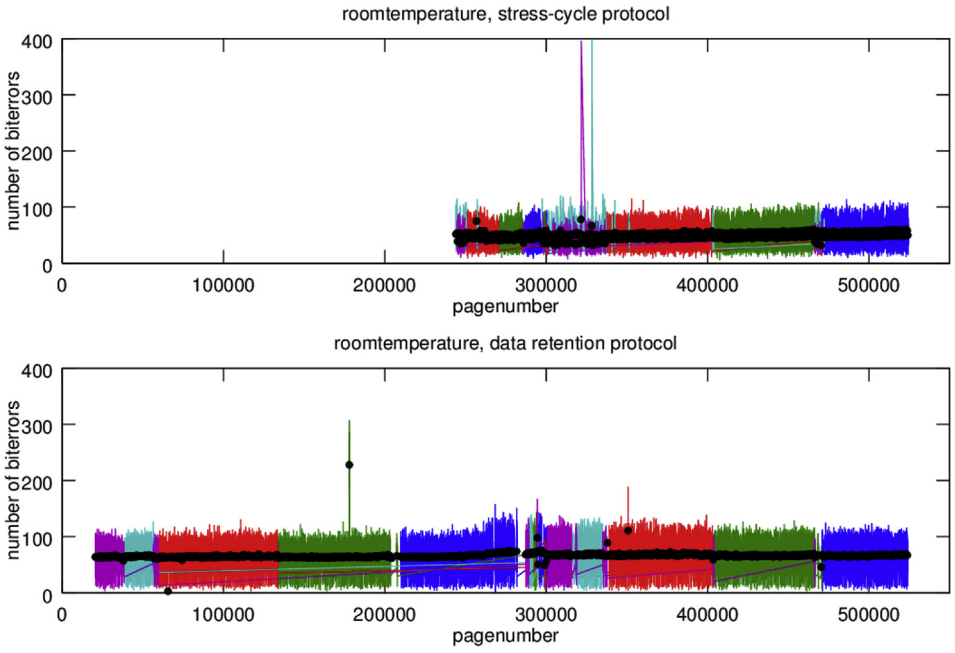

Fig. 4. Pagewise bit-error statistics for two USB thumb-drives stored at room temperature for files copied onto drive in stress-cycle protocol (*top panel*) and data retention protocol (*bottom panel*). Same colour coding of files as in Fig. 3. Block averaged data over blocks of 256 pages are shown as black dots.

situations, this might be more difficult, since pages might contain the same data or fragments of data of deleted files. At any rate, in a forensic context, the connection between bit-errors and files can potentially be used to help grouping memory pages belonging to the same file together, which have been spread across the NAND-flash due to wear leveling. This can help filesystem reconstruction and smart carving of data.

Current experiments can be taken as an indication that bit-error statistics contain some independent time information, since files that have been present at the NAND-flash for different amount of times show a different amount of retention bit-errors (cf. Figs. 3–4). Therefore, it would be tempting to use the (block averaged) number of bit-errors in memory pages of a file as a means of relative dating of files present on a NAND-flash device. At relatively dating,

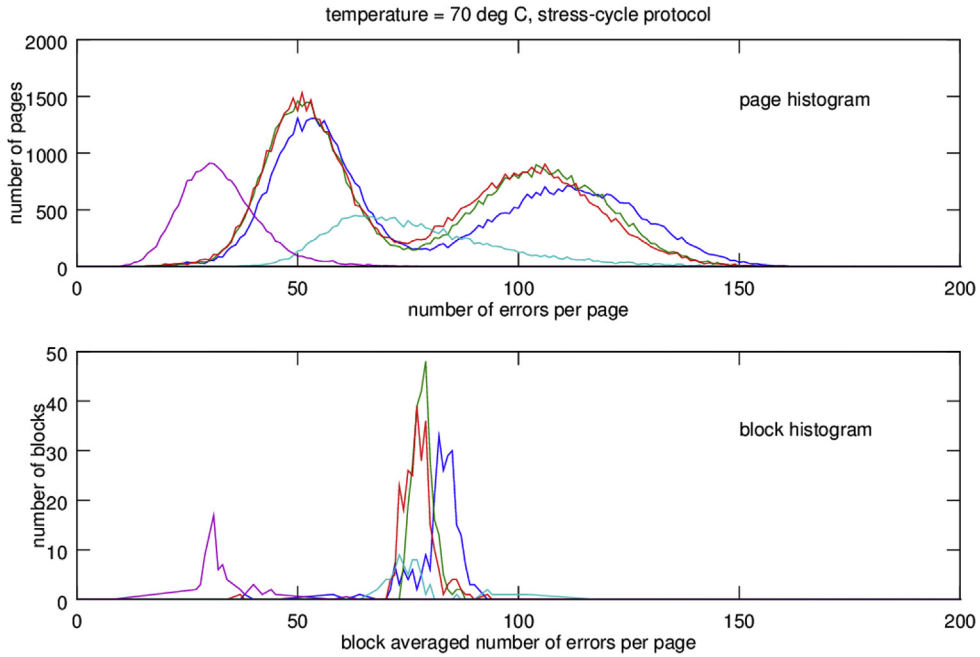

Fig. 5. Histograms of distribution of number of bit-errors per page for thumb-drive kept at 70 °C, stress-cycle protocol. Same colour coding of files as used in Figs. 3 and 4. *Top panel:* distributions of number of bit-errors per page. *Bottom panel:* Distribution of number of bit-errors per page, averaged over blocks of 256 pages.

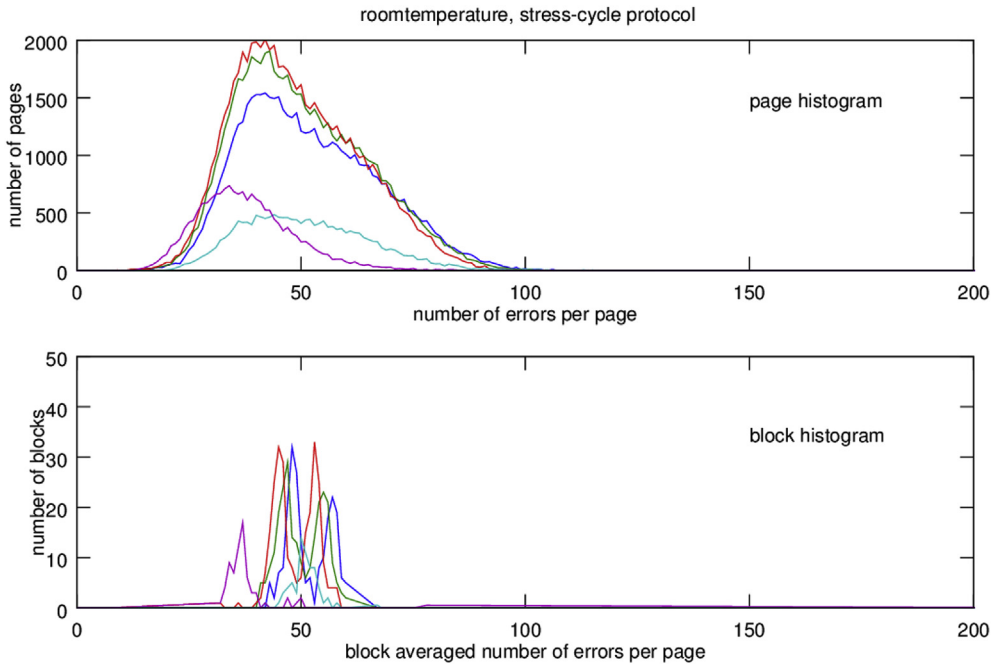

Fig. 6. Histograms of distribution of number of bit-errors per page for thumb-drive kept at room temperature, stress-cycle protocol. Same colour coding of files as used in Figs. 3 and 4. *Top panel:* distributions of number of bit-errors per page. *Bottom panel:* Distribution of number of bit-errors per page, averaged over blocks of 256 pages.

however, there is the problem that it is not possible to separate the contribution of retention time and P/E cycles on the occurrence of bit-errors from one another. This means that without additional information, no distinction can be made between a file that has been briefly on part of the NAND-flash which has undergone many P/E cycles and a file which has been on part of the NAND-flash with few P/E cycles for a long time. One idea to try and compensate for the effect of P/E cycles on rate of bit-error development would be to erase a NAND-flash chip entirely after analysis and reprogram it completely with random data. After that, the chip can be stored at a high temperature for some time to let retention bit-errors develop. If the chip is read out completely after that, one might observe that areas which have undergone the most P/E cycles, will contain the most bit-errors, similarly as the data shown in Fig. 2. This additional information could then be used to correct the original data for differences in P/E cycles.

Acknowledgements

I am indebted to Marcel Breeuwsma and Ronald van der Knijff for useful comments on an early draft of this paper.

References

Breeuwsma, M., de Jongh, M., Klaver, C., van der Knijff, R., Roeloffs, M., 2007. Forensic data recovery from flash memory. Small Scale Digit. Device Forensics 1, 1–17.

Cai, Y., Haratsch, E.F., Mutlu, O., Mai, K., 2012. Error patterns in MLC NAND flash memory: measurement, characterization and analysis. Date 521–526.

Meza, J., Wu, Q., Kumar, S., Mutlu, O. A large-scale study of flash memory failures in the field. Proceedings of the ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems 2015, 177–190.

Niset, M., Kuhn, P., 2005. Typical data retention for nonvolatile memory. Freescale Semicond. Eng. Bull. EB618/D. Rev 4.

Van Zandwijk, J.P., 2015. A mathematical approach to NAND flash-memory descrambling and decoding. Digit. Investig. 12, 41–52.