

**A team of GIAC Certified Forensic Analysts (GCFA) and their thoughts on Digital Forensic and Incident Response techniques and trends.**

## Overwriting Hard Drive Data

By Dr. Craig Wright  
GIAC GSE (Compliance & Malware)

This post is based on a paper I published in December last year; "Overwriting Hard Drive Data: The Great Wiping Controversy" by Craig Wright, Dave Kleiman and Shyaam Sundhar R.S. as presented at ICISS2008 and published in the Springer Verlag Lecture Notes in Computer Science (LNCS) series.

## Background

Opinions on the required or desired number of passes to correctly overwrite (wipe) a Hard Disk Drive are controversial, and have remained so even with organizations such as NIST stating that only a single drive wipe pass is needed to delete data such that it can not be recovered (that is a wipe of the data).

The controversy has caused much misconception. This was the reason for this project.

It is common to see people quoting that data can be recovered if it has only been overwritten once, many times referencing that it actually takes up to ten, and even as many as 35 (referred to as the Gutmann scheme because of the 1996 Secure Deletion of Data from Magnetic and Solid-State Memory published paper by Peter Gutmann, [12]) passes to securely overwrite the previous data.

To answer this once and for all, a project was started in 2007 to actually test whether or not data can be recovered from a wiped drive if one uses an electron microscope. For the full details, you will need to actually read the published paper, though this post offers a synopsis. In subsequent communications with Prof. Fred Cohen I have come to realize that there are certain other uses for the methods I have used in this effort. The recovery of data from damaged drives is possible. Further, using the mathematical methods employed in the experiment (Bayesian statistics) one can recover data from damaged drives with far simpler means than through the use of a MFM (magnetic force microscope).

On top of this, I have to note (thanks to Prof. Cohen) that many larger modern drives are not overwritten in the course of use in many sectors due to size.

## Where did the controversy begin?

The basis of this belief that data can be recovered from a wiped drive is based on a presupposition that when a one (1) is written to disk the actual effect is closer to obtaining a 0.95 when a zero (0) is overwritten with one (1), and a 1.05 when one (1) is overwritten with one (1).

This can be demonstrated to be false.

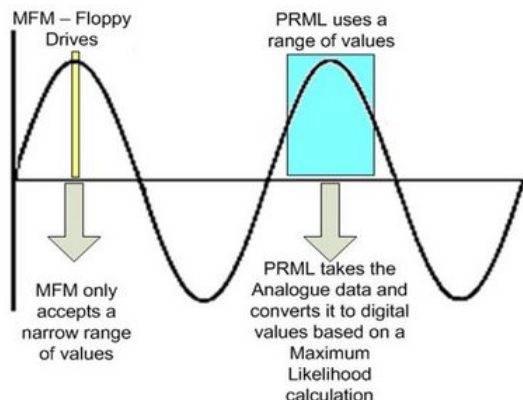
This was the case with high capacity floppy diskette drives, which have a rudimentary position mechanism. This was at the bit level and testing did not consider the accumulated error. The argument arises from the statement that "each track contains an image of everything ever written to it, but that the contribution from each "layer" gets progressively smaller the further back it was made. This is a misunderstanding of the physics of drive functions and magneto-resonance. There is in fact no time component and the image is not layered. It is rather a density plot.

## MFM - Magnetic Force Microscopy

To test this theory, we used a MFM. Magnetic force microscopy (MFM) images the spatial variation of magnetic forces on a sample surface. A MFM is a variety of what most people simply term an electron microscope.

### Partial Response Maximum Likelihood (PRML)

The concepts of how Partial Response Maximum Likelihood (PRML), a method for converting the weak analogue signal from the head of a magnetic disk or tape drive into a digital signal, and newer Extended Partial Response Maximum Likelihood (EPRML) drive, explain how encoding is implemented on a hard drive. The MFM reads the unprocessed analogue value. Complex statistical digital processing algorithms are used to determine the "maximum likelihood" value associated with the individual reads.



### SANS Forensic Blog Feed



#### Top Posts

- P2P Usage Leads To Presidential Security Breach
- Overwriting Hard Drive Data
- Nevada bill would make some security research a felony
- Windows Physical Memory: Finding the Right Tool for the Job
- Digital Forensic SIFT'ing: Registry and Filesystem Timeline Creation
- Memory Forensic Acquisition and Analysis 101

#### Blogroll

- SANS Computer Forensics Training and Courses

SANS

2009

Our Most Comprehensive Event of the Year

35 Technical Courses

Walt Disney World

March 2-9, 2009

Orlando

MORE INFO

#### Archives

- March 2009 (3)
- February 2009 (20)
- January 2009 (18)
- December 2008 (22)
- November 2008 (9)
- October 2008 (13)
- September 2008 (11)
- August 2008 (3)

Older technologies used a different method of reading and interpreting bits than modern hard drives that is known as peak detection. This method is satisfactory while the peaks in magnetic flux sufficiently exceed the background signal noise. With the increase in the write density of hard drives, encoding schemes based on peak detection (such as Modified Frequency Modulation or MFM) that are still used with floppy disks have been replaced in hard drive technologies. The encoding of hard disks is provided using PRML and EPRML encoding technologies that have allowed the write density on the hard disk to be increased by a full 30-40% over that granted by standard peak detection encoding.

### Common misconceptions

Drive writes are magnetic field alterations, the belief that a physical impression in the drive that can belie the age of the impression is wrong. The magnetic flux density follows a function known as the hysteresis loop. The magnetic flux levels written to the hard drive platter vary in a stochastic manner with variations in the magnetic flux related to head positioning, temperature and random error.

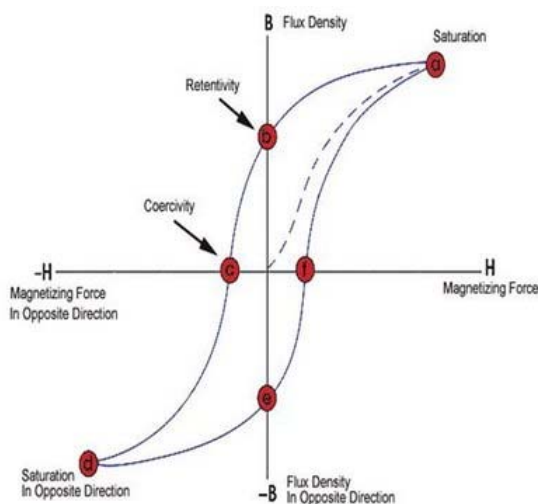
The surfaces of the drive platters can have differing temperatures at different points and may vary from the read/write head. As a consequence, there are many problems with the belief that data is recoverable following a wipe. The differences in the expansion and contraction rates across the drive platters uses a stochastically derived thermal recalibration algorithm. All modern drives use this technology to minimize variance. However, even with this algorithm, the data written to the drive is done in what is in effect an analogue pattern of magnetic flux density.

### A stochastic distribution

Complex statistically based detection algorithms are employed to process the analog data stream as data is read from the disk. This is the "partial response" component mentioned previously. A stochastic distribution of data not only varies on each read, but also over time and with temperature differentials. Worse, there is the hysteresis effect to be considered.

### Hysteresis

Stochastic noise results in a level of controlled chaos [Carroll and Pecora (1993a, 1993b)]. When looking at the effects of a magnetically based data write process, the hysteresis effect ensures that data does not return to a starting point.



As such, you can never go to the original starting point. On each occasion that you add and delete data from a drive, the resulting value that is written to the disk varies. This begins with a low level format, changes whenever any data is written to the drive and fluctuates on each attempt to zero the platter. What in effect you get is a random walk that never quite makes it back to the original starting point (save exposure to a powerful magnetic force or annealing process).

### Magnetic signatures are not time-stamped

There is no "unerase" capability [15] on a hard drive due to magnetic resonance. There are no "layers" of written data. The value of the magnetic field does vary on each write to the drive, but it does so due to other factors and influences. These include:

- fluctuations in temperature,
  - movement of the head,
  - prior writes to the drive...
- All and any of these effects will influence the permeability of the platter in a statistically significant manner.  
Variability in magnetic force

Jiles [21] notes that in the event that the temperature of a drive platter is increased from, 20 to 80 centigrade then a typical ferrite can become subject to a 25% reduction in the in permeability of the platter. Within a drive, the temperature of "normal" operation can vary significantly. With constant use, a drive can easily exceed 80 degrees centigrade internally. The system may not get this hot, but all that is required is a segment of the platter, and this is common.

The permeability is a material property that is used to measure how much effort is required to induce a magnetic flux within a material. Permeability is defined the ratio of the flux density to the magnetizing force. This may be displayed with the formula:  $\mu = B/H$  (where  $\mu$  is the permeability,  $B$  is the flux density and  $H$  is the magnetizing

Search

Find »

### Blogroll

- [SANS Computer Forensics Training and Courses](#)

### RSS Feeds

- [All posts](#)
- [All comments](#)

### Meta

- [Log in](#)

Blog at WordPress.com.  
Sandbox

force). Due to the changes experienced by a drive, MFM techniques (detailed above) as used with floppy drives do not work in modern hard drives.

### The hypothesis and the experiment

To test the hypothesis, a number of drives of various ages and types and from several vendors were tested. In order to completely validate all possible scenarios, a total of 15 data types were used in 2 categories.

Category A divided the experiment into testing the raw drive (this is a pristine drive that has never been used), formatted drive (a single format was completed in Windows using NTFS with the standard sector sizes) and a simulated used drive (a new drive was overwritten 32 times with random data from /dev/random on a Linux host before being overwritten with all 0's to clear any residual data).

The experiment was conducted in order to test a number of write patterns. There are infinitely many possible ways to write data, so not all can be tested. The idea was to ensure that no particular pattern was significantly better or worse than another.

Category B consisted of the write pattern used both for the initial write and for the subsequent overwrites.

This category consisted of 5 dimensions:

- all 0's,
- all 1's,
- a "01010101 pattern,
- a "00110011" pattern, and
- a "00001111" pattern.

The Linux utility "dd" was used to write these patterns with a default block size of 512 (bs=512). A selection of 17 models of hard drive where tested. These varied from an older Quantum 1 GB drive to current drives (at the time the test started) dated to 2006.

The data patterns where written to each drive in all possible combinations. Each data write was a 1 kb file (1024 bits). It was necessary to carefully choose a size and location. Finding a segment on a drive without prior knowledge is like looking for the proverbial needle in the haystack. To do this, the following steps where taken:

Both drive skew and the bit was read.

The process was repeated 5 times for an analysis of 76,800 data points.

The likelihood calculations were completed for each of the 76,800 points with the distributions being analyzed for distribution density and distance.

This calculation was based on the Bayesian likelihood where the prior distribution was known.

As has been noted, in real forensic engagements, the prior distribution is unknown. When you are trying to recover data from a drive, you generally do not have an image of what you are seeking to recover. Without this forensic image, the experiment would have been exponentially more difficult. What we found from this is that even on a single write the overlap at best gives a probability of as low as just over 50% of choosing a prior bit (the best read being a little over 56%).

This caused the issue to arise, that there is no way to determine if the bit was correctly chosen or not.

Therefore, there is a chance of correctly choosing any bit in a selected byte (8-bits) – but this equates a probability around 0.9% (or less) with a small confidence interval either side for error.

### The Results of the Tests

The calculated values are listed below for the various drives. Not all data is presented here, but it is clear to see that use of the drive impacts the values obtained (through the hysteresis effect and residuals). The other issue is that all recovery is statistically independent (for all practical purposes). The probability of obtaining two bits is thus multiplied.

Probability of recovery	Pristine drive	Used Drive (ideal)
1 bit	0.92	0.56
2 bit	0.8464	0.3136
4 bit	0.71639296	0.098345
8 bits <sup>§</sup>	0.51321887	0.009672
16 bits	0.26339361	9.35E-05
<b>32 bits</b>	<b>0.06937619</b>	<b>8.75E-09</b>
64 bits	0.00481306	7.66E-17
128 bits	2.3166E-05	5.86E-33
256 bits	5.3664E-10	3.44E-65
512 bits	2.8798E-19	1.2E-129
1024 bits	8.2934E-38	1.4E-258

Table of Probability Distributions for the older model drives.

Probability of recovery	Pristine drive (plus 1 wipe)	Pristine drive (plus 3 wipe)
1 bit	0.87	0.64
2 bit	0.7569	0.4096
4 bit	0.57289761	0.16777216
8 bits	0.328211672	0.028147498
16 bits	0.107722901	0.000792282
<b>32 bits</b>	<b>0.011604223</b>	<b>6.2771E-07</b>
64 bits	0.000134658	3.9402E-13
128 bits	1.81328E-08	1.55252E-25
256 bits	3.28798E-16	2.41031E-50
512 bits	1.08108E-31	5.8006E-100

Table of Probability Distributions for the "new" (ePRML) model drives.

What we see is that it quickly becomes practically impossible to recover anything \*and this is not even taking the time to read data using a MFM into account).

#### What this means

The other overwrite patterns actually produced results as low as 36.08% (+/- 0.24). Being that the distribution is based on a binomial choice, the chance of guessing the prior value is 50%. That is, if you toss a coin, you have a 50% chance of correctly choosing the value. In many instances, using a MFM to determine the prior value written to the hard drive was less successful than a simple coin toss.

The purpose of this paper was a categorical settlement to the controversy surrounding the misconceptions involving the belief that data can be recovered following a wipe procedure. This study has demonstrated that correctly wiped data cannot reasonably be retrieved even if it is of a small size or found only over small parts of the hard drive. Not even with the use of a MFM or other known methods. The belief that a tool can be developed to retrieve gigabytes or terabytes of information from a wiped drive is in error.

Although there is a good chance of recovery for any individual bit from a drive, the chances of recovery of any amount of data from a drive using an electron microscope are negligible. Even speculating on the possible recovery of an old drive, there is no likelihood that any data would be recoverable from the drive. The forensic recovery of data using electron microscopy is infeasible. This was true both on old drives and has become more difficult over time. Further, there is a need for the data to have been written and then wiped on a raw unused drive for there to be any hope of any level of recovery even at the bit level, which does not reflect real situations. It is unlikely that a recovered drive will have not been used for a period of time and the interaction of defragmentation, file copies and general use that overwrites data areas negates any chance of data recovery. The fallacy that data can be forensically recovered using an electron microscope or related means needs to be put to rest.

*Craig Wright, GCFA Gold #0265, is an author, auditor and forensic analyst. He has nearly 30 GIAC certifications, several post-graduate degrees and is one of a very small number of people who have successfully completed the GSE exam.*

#### References

- Abramowitz, M., & Stegun, I. A. (1965), "Handbook of Mathematical Functions" (Dover, New York).
- Amit, D. J., (1984), "Field Theory", The Renormalization Group and Critical Phenomena (World Scientific, Singapore).
- Braun, H. B. (1994) "Fluctuations and instabilities of ferromagnetic domain-wall pairs in an external magnetic field", Phys. Rev. B. 50 (1994), 16485-16500
- Brown, G., Novotny, M. A. & Rikvold, P. A. (2001) "Thermal magnetization reversal in arrays of nanoparticles", J. Appl. Phys. 89, 7588-7590.
- Bulsara, A., S. Chillemi, L. Kiss, P. V. E. McClintock, R. Mannella, F. Marchesoni, G. Nicolis, and K. Wiesenfeld, (1995), Eds., "International Workshop on Fluctuations in Physics and Biology: Stochastic Resonance, Signal Processing and Related Phenomena", published in Nuovo Cimento 17D, 653.
- Carroll, T. L., & Pecora, L. M. (1993a), Phys. Rev. Lett. 70, 576.
- Carroll, T. L., & Pecora, L. M. (1993b), Phys. Rev. E 47, 3941.
- Gomez, R., A. Adly, I. Mayergoyz, E. Burke. (1992). "Magnetic Force Scanning Tunnelling Microscope Imaging of Overwritten Data", IEEE Transactions on Magnetics, (28:5), 3141.
- Gammaitoni L., Ha"nggi P., Jung P., & Marchesoni F. (1998 ) "Stochastic resonance", Reviews of Modern Physics, Vol. 70, No. 1, January 1998, The American Physical Society
- Gomez, R., E. Burke, A. Adly, I. Mayergoyz, J. Gorczyca. (1993). "Microscopic Investigations of Overwritten Data", Journal of Applied Physics, (73:10), 6001.
- Grinstein G. & Koch, R. H. (2005) "Switching probabilities for single-domain magnetic particles, to appear" Phys. Rev. B 71, 184427, USA
- Gutmann, P. (1996) "Secure Deletion of Data from Magnetic and Solid-State Memory", Proceedings of the Sixth USENIX Security Symposium. July 22-25, San Jose, CA., 77-90. ([http://www.cs.auckland.ac.nz/~pgut001/pubs/secure\\_del.html](http://www.cs.auckland.ac.nz/~pgut001/pubs/secure_del.html))
- Ha"nggi, P., and R. Bartussek, (1996), in Nonlinear Physics of Complex Systems: "Current Status and Future Trends", edited by J. Parisi, S. C. Mu"ller, and W. Zimmermann, Lecture Notes in Physics 476 (Springer, Berlin, New York), p. 294.
- Liu, D. (2003) "Topics in the Analysis and Computation of Stochastic Differential Equations", Ph. D. thesis, Princeton University.
- Mayergoyza, I.D., Tse, C., Krafft, & C., Gomez, R. D. (2001) "Spin-stand imaging of overwritten data and its comparison with magnetic force microscopy", JOURNAL OF APPLIED PHYSICS VOLUME 89, NUMBER 11.
- Moss, F., (1994), in Contemporary Problems in Statistical Physics, edited by G. H. Weiss (SIAM, Philadelphia), pp. 205-253.
- Ren, W. E. W. & Vanden-Eijnden, E. (2003) "Energy landscape and thermally activated switching of submicron-size ferromagnetic elements", J. Appl. Phys. 93, 2275-2282.

Reznikoff, M. G. (2004) "Rare Events in Finite and Infinite Dimensions", Ph. D. thesis, New York University.

Rugar, D. H. Mamin, P. Guenther, S. Lambert, J. Stern, I. McFadyen, and T. Yogi. (1990). "Magnetic Force Microscopy: General Principles and Application to Longitudinal Recording Media", Journal of Applied Physics, (68:3), 1169

Nikola Tesla

"The Great Radio Controversy". [http://en.wikipedia.org/wiki/Invention\\_of\\_radio](http://en.wikipedia.org/wiki/Invention_of_radio)

Jiles, David, (1998 ) "Introduction to magnetism and magnetic materials", 2nd. Ed., Chapman & Hall

---

#### Possibly related posts: (automatically generated)

[Spin-Stand Microscopy of Hard Disk Data](#)

[Seagate Hard Drives Problems in Macbooks](#)

[A hard disk failure puts critical business information at risk](#)

[Hard drives getting bigger and smaller](#)

*This entry was written by [craigswright](#) and posted on January 15, 2009 at 8:00 am and filed under [Computer Forensics](#). Bookmark the [permalink](#). Follow any comments here with the [RSS feed for this post](#). Trackbacks are closed, but you can [post a comment](#).*

## 5 Comments

rsreese

Posted January 15, 2009 at 1:08 pm [Permalink](#)

Very cool findings! Thanks for the research...

simsong

Posted January 16, 2009 at 4:45 pm [Permalink](#)

I'm confused by this. I read the paper. I didn't see:

- \* That the authors validated their recovery approaches by attempting to read valid data.
- \* How the sectors on the disk were found.
- \* How the authors translate from the proprietary coding used by the vendors was translated to ascii.

Can someone point to where this is in the paper?

craigswright

Posted January 16, 2009 at 5:07 pm [Permalink](#)

To answer your query.

- \* That the authors validated their recovery approaches by attempting to read valid data.

You should note that a read was conducted on initial writes with known data. This means that the initial pattern was known and hence would be recovered in a manner that would lead to this point being validated. This is detailed in the paper.

- \* How the sectors on the disk were found.

Contiguous writes. All areas of the disk not checked left blank. thus the only data was our data. Knowing where the data was written was simple. No magic to this.

- \* How the authors translate from the proprietary coding used by the vendors was translated to ascii.

Not particularly difficult. The various PRML and ePRML schemes are published and available. Clock cycles and patterns are simple. It is not as much of an art as you seem to believe.

Regards,  
Craig

[ralienpp](#)

Posted February 17, 2009 at 11:42 am [Permalink](#)

This is an interesting article; everything seems clear, the figures support your statements.

I just checked out NIST's "Updated DSS Clearing and Sanitization Matrix AS OF: June 28, 2007", and they indeed propose a much more simple procedure for hard drives.

Can you recommend some articles by the "opposing team", where they provide a rationale for the multi-pass approach? If they still have something to say, I'd like to hear it.

craigswright

Posted February 25, 2009 at 2:17 pm [Permalink](#)

Other than the original paper by Peter Gutmann there is little on the subject.

Even then I have found little science on the subject. Science requires proof. What starts as a hypothesis needs to be tested before it becomes science. I have found nothing along these lines.

### Post a Comment

You must be [logged in](#) to post a comment.