Overwriting Hard Drive Data: The Great Wiping Controversy

Conference Paper *in* Lecture Notes in Computer Science · December 2008

DOI: 10.1007/978-3-540-89862-7_21 · Source: DBLP

CITATIONS
63

READS
13,051

3 authors, including:

Craig Wright
Association for Computing Machinery
27 PUBLICATIONS **453** CITATIONS

SEE PROFILE

Shyaam Sundhar
American Military University
2 PUBLICATIONS **161** CITATIONS

SEE PROFILE

Overwriting Hard Drive Data: The Great Wiping Controversy

Craig Wright[1], Dave Kleiman[2], and Shyaam Sundhar R.S.[3]

[1] BDO Kendalls, Sydney, Australia
Craig.Wright@bdo.com.au
[2] ComputerForensicExaminer.com, Florida, US
dave@davekleiman.com
[3] Symantec, USA
shyaam@gmail.com

Abstract. Often we hear controversial opinions in digital forensics on the required or desired number of passes to utilize for properly overwriting, sometimes referred to as wiping or erasing, a modern hard drive. The controversy has caused much misconception, with persons commonly quoting that data can be recovered if it has only been overwritten once or twice. Moreover, 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) passes to securely overwrite the previous data. One of the chief controversies is that if a head positioning system is not exact enough, new data written to a drive may not be written back to the precise location of the original data. We demonstrate that the controversy surrounding this topic is unfounded.

Keywords: Digital Forensics, Data Wipe, Secure Wipe, Format.

1 Introduction

Often we hear controversial opinions on the required or desired number of passes to utilize for properly overwriting, sometimes referred to as wiping or erasing, a modern hard drive. The controversy has caused much misconception, with persons commonly quoting that data can be recovered if it has only been overwritten once or twice. Moreover, 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.

One of the chief controversies is that if a head positioning system is not exact enough, new data written to a drive may not be written back to the precise location of the original data. This track misalignment is argued to make possible the process of identifying traces of data from earlier magnetic patterns alongside the current track.

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.

R. Sekar and A.K. Pujari (Eds.): ICISS 2008, LNCS 5352, pp. 243–257, 2008.

The basis of this belief is a presupposition is 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 we can show is false and that in fact, there is a distribution based on the density plots that supports the contention that the differential in write patterns is too great to allow for the recovery of overwritten data.

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.

This is of prime importance to forensic analysts and security personal. The time needed to run a single wipe of a hard drive is economically expensive. The requirements to have up to 35 wipes [12] of a hard drive before disposal become all the more costly when considering large organisations with tens of thousands of hosts. With a single wipe process taking up to a day to run per host through software and around an hour with dedicated equipment, the use of multiple wipes has created a situation where many organisations ignore the issue all together – resulting in data leaks and loss.

The inability to recover data forensically following a single wipe makes the use of data wiping more feasible. As forensic and information security professionals face these issues on a daily basis, the knowledge that a single wipe is sufficient to remove trace data and stop forensic recovery will remove a great deal of uncertainty from the industry and allow practitioners to focus on the real issues.

1.1 What Is Magnetic Force Microscopy[1]

Magnetic force microscopy (MFM) images the spatial variation of magnetic forces on a sample surface. The tip of the microscope is coated with a ferromagnetic thin film. The system operates in non-contact mode, detecting changes in the resonant frequency of the cantilever induced by the magnetic field's dependence on tip-to-sample separation. A MFM can be used to image naturally occurring and deliberately written domain structures in magnetic materials. This allows the device to create a field density map of the device.

1.2 MFM Imagery of Overwritten Hard Disk Tracks

The magnetic field topography (Fig. 2A below) was imaged with an MFM to measure the magnetic force density. This image was captured using the MFM in Lift Mode (lift height 35 nm). This results in the mapping of the shift in the cantilever resonant frequency.

[1] The MFM senses the stray magnetic field above the surface of a sample. A magnetic tip is brought into close proximity with the surface and a small cantilever is used to detect the force between the tip and the sample. The tip is scanned over the surface to reveal the magnetic domain structure of the sample at up to 50 nm resolution.

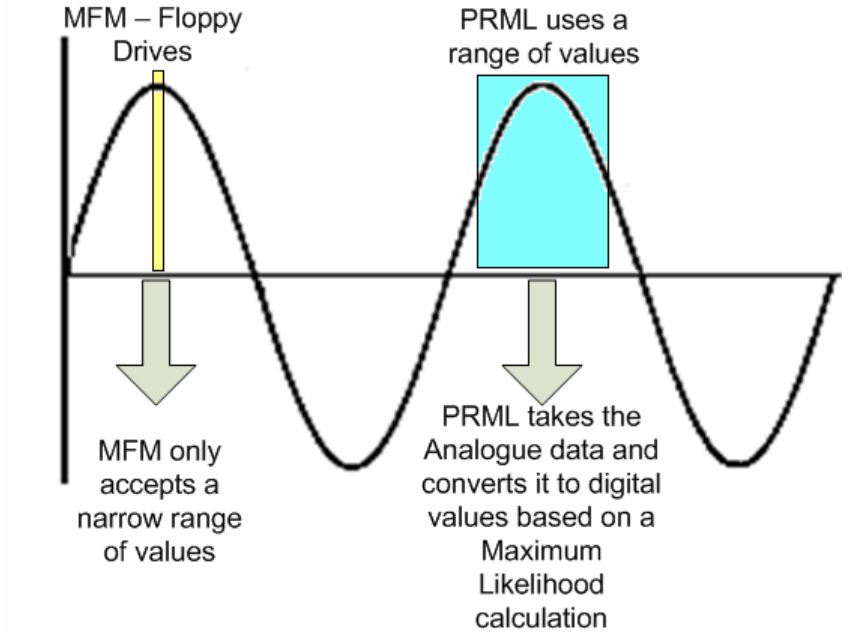
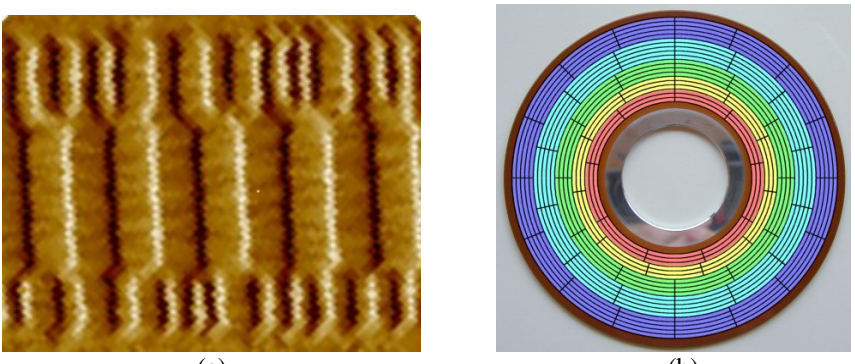


Fig. 1. The concepts of how Partial Response Maximum Likelihood (PRML) (a method for converting the weak analog signal from the head of a magnetic disk or tape drive into a digital signal) (and newer Extended Partial Response Maximum Likelihood (EPRML) drive) encoding is implemented on a hard drive. The MFM reads the unprocessed analog value. Complex statistical digital processing algorithms are used to determine the "maximum likelihood" value associated with the individual reads.

The acquisition time for 1 byte is about 4 minutes (this would improve with newer machines). The image displays the:

- track width and skew,
- transition irregularities, and
- the difference between written and overwritten areas of the drive.

Because of the misconception, created by much older technologies (such as floppy drives) with far lower densities, many believe that the use of an electron microscope will allow for the recovery of forensically usable data. The fact is, with modern drives (even going as far back as 1990) that this entire process is mostly a guessing game that fails significantly when tested. Older technologies used a different method of reading and interpreting bits than modern hard called *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 (Fig. 3), 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.



(a) (b)

Fig. 2. (a). This image was captured and reconstructed at a 25-µm scan from an Atomic Force Microscope [15]. The image displays the residual from overwrites and alignment. (b). This image from PCGuide.com displays the configuration of a 20 track hard drive. These tracks are separated into five zones which are displayed in a separate color as follows: 5 x 16 sector tracks in the blue zone, 5 x 14 sector tracks in the cyan zone, 4 x 12 sector tracks in the green zone, 3 x 11 sectors tracks in the yellow zone, and 3 x 9 sector tracks in the red.

Additionally, hard disk drives use zoned bit recording (Fig 2b) which differs from floppy drives and similar technologies. Older technologies (including floppy disks) used a single zone with a write density that is several orders of magnitude larger than that used with hard disks. We have not tested recovery from a floppy disk using these methods, but it would be expected that the recovery rate would be significantly greater than with respect of that of a hard disk platter - although still stochastically distributed.

The fact is many people believe that this is a physical impression in the drive that can belie the age of the impression. This misconception is commonly held as to the process used to measure the magnetic field strength. Using the MFM in Tapping Mode[2], we get a topography image that represents the physical surface of the drive platter.

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. This results in differences in the expansion and contraction rates across the drive platters. This differential can result in misalignments. Thermal recalibration is used on modern drives to minimize this variance, but this is still results in an analogue pattern of magnetic flux density.

One of ways used to minimize the resultant error has come through the introduction of more advanced encoding schemes (such as PRML mentioned previously). Rather than relying on differentiating the individual peaks at digital maxima,

[2] Tapping mode can also be called Dynamic Force mode, intermittent contact mode, non-contact mode, wave mode, and acoustic AC mode by various microscope vendors. When operating in tapping mode the cantilever is driven to oscillate up and down at near its resonance frequency by a small piezoelectric element.

magnetic flux reversals are measured by hard drive heads and processed using an encoding process (PRML or EPRML) that is based on determining maximum likelihood for the flux value using a digital signal sampling process. Complex statistically based detection algorithms are employed to process the analog data stream as it is read the disk. This is the "partial response" component. This stochastic distribution of data not only varies on each read, but also over time and with temperature differentials. This issue has only grown as drive densities increase.

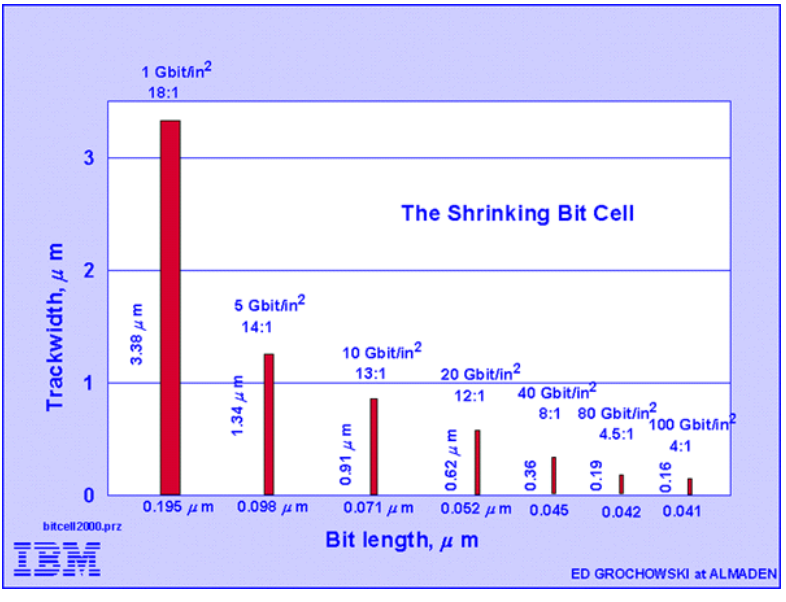


Fig. 3. This graph from IBM demonstrates how the bit size used with modern hard drives is shrinking. This has resulted in a dramatic increase in the density of hard disks which has resulted in the error rate from movement and temperature remaining an issue even with the improvements in compensating technologies.

A Track is a concentric set of magnetic bits on the disk. Each track is commonly divided into 512 bytes sectors. The drive sector is the part of each track defined with magnetic marking and an ID number. Sectors have a sector header and an error correction code (ECC).

A Cylinder is a group of tracks with the same radius.

Data addressing occurs within the two methods for data addressing:

- CHS (cylinder-head-sector) and
- LBA (logical block address).

The issue from Guttmann's paper [12] is that we can recover data with foreknowledge of the previous values, but not with any level of accuracy. The issues with this are twofold. First, to have any chance of recovery it is necessary to have perfect

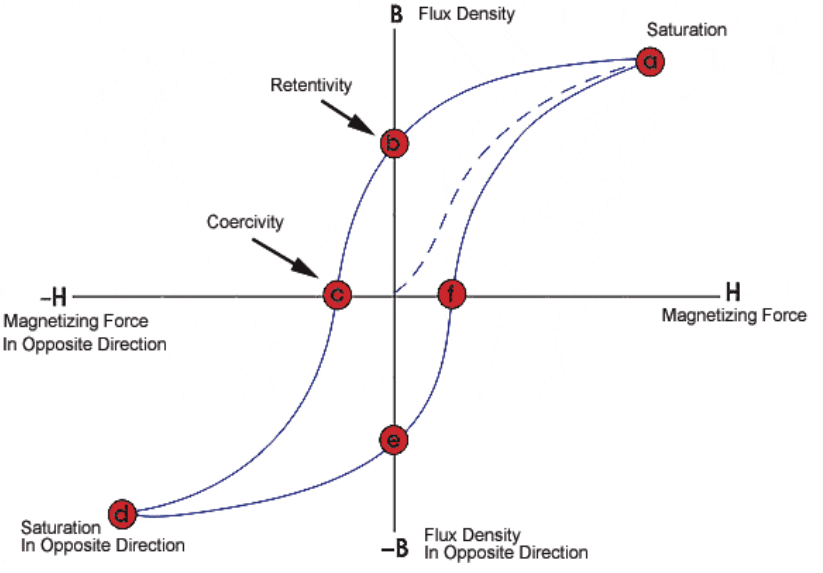


Fig. 4. The hysteresis loop[3] demonstrates the relationship between the induced magnetic flux density (B) and the magnetizing force (H). It is often referred to as the B-H loop. This function varies with a number of prevalent conditions including temperature.

knowledge of what was previously written to the drive. This situation most often never occurs in a digital forensic investigation. In fact, if a perfect copy of the data existed, there would be no reason to recover data from the wiped drive. Next, the level of recovery when presented with a perfect image is too low to be of use even on a low density pristine drive (which does not exist in any actual environment). Carroll and Pecora (1993a, 1993b) demonstrated this effect and how stochastic noise results in a level of controlled chaos. The Guttmann preposition [12] is true based on a Bayesian a-prior model assuming that we have the original data and the pattern from the overwrite, but of course this defeats the purpose of the recovery process and as noted is still not sufficiently accurate to be of any use. Stating that we can recover data with a high level of accuracy, given that we have the original data, is a tautology, and there would be no reason to do the recovery.

The previously mentioned paper uses the determination that the magnetic field strength is larger or smaller than that which would be expected from a write suggests the prior overwritten value. This is that a factored magnetic field strength of 0.90 or 1.10 (where 1.0 is a "clean" write with no prior information) would represent the previous information written to the drive that has been overwritten. This is postulated to be a means through which the use of an electron microscope could be deployed to recover data from a drive that has been wiped. The problem with this theory is that there are both small write errors on an unwritten sector and remnant magnetic field densities from prior use of the drive sector.

[3] Image sourced from Iowa's State University Center for Nondestructive Evaluation NDT (Non Destructive Testing).

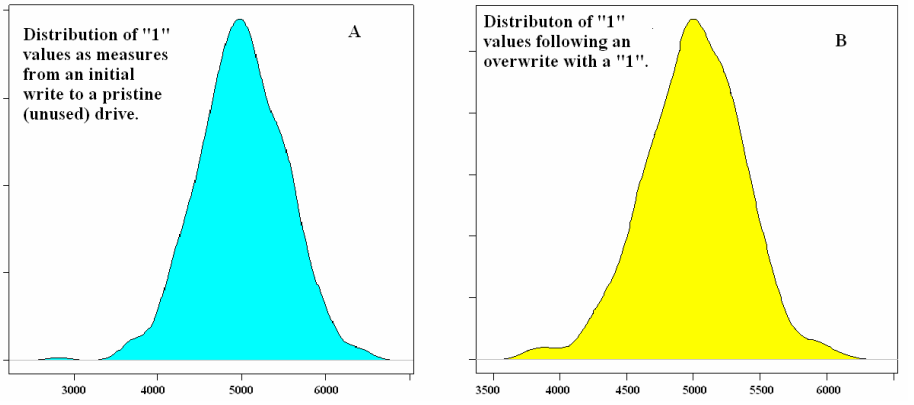


Fig. 5. This example displays the experimentally derived magnetic field density functions for hard drive rewrites where "A" displays the measured distribution of binary "1" values on initial copy. "B" displays the distribution of values associated with a binary "1" value following an overwrite with another binary "1".

Magnetic signatures are not time-stamped, accordingly there is no "unerase" capability [15]. Figure 4 displays the B-H loop for magnetic flux. Starting at a zero flux density for a drive platter that has not been previously magnetized, the induced flux density created when the drive head writes to the platter follows the dashed line (displayed in Fig. 4) as the magnetizing force is increased. Due to a combination of power constraints, timing issues and write density, modern hard drives do not saturate the magnetic flux on the drive to point "a". Rather, they use sophisticated statistical measures (PRML and EPRML) to determine the maximum likelihood of the value stored on the drive. In demagnetizing a drive (reducing H to zero) the curve moves from point "a" to point "b" on Figure 4. Some residue from the prior magnetic flux remains in the material even though the magnetizing force is zero. This phenomenon is known as remanence. The retentivity of disk platter will not reach the maxima (defined by points "b" and "d" in figure 4) as the drive heads do not reach saturation. Further, fluctuations in temperature, movement and prior writes influence the permeability[4] of the platter. 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.

Consequently, the B-H curve does not go back to the originating point when the magnetic flux is rewritten and the B-H curve will vary with use due to temperature fluctuations. On subsequent writes, the hysteresis curve follows a separate path from position "f" in Figure 4. As drive heads do not cause the hard drive platter to reach the saturation point, the resultant B-H loop will vary on each write.

[4] 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 μ is the permeability, B is the flux density and H is the magnetizing force).

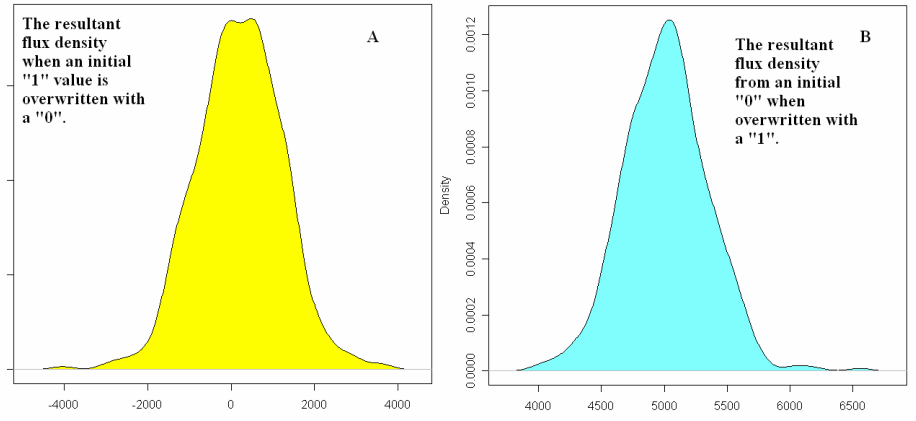


Fig. 6 (A and B). This example displays the magnetic field density functions that were experimentally obtained following a rewrite (wipe) of the prior binary unit on the hard drive. Plot "A" displays the density distribution associated with "1" values following an overwrite with a "0". Plot "B" displays the density function for an initial "0" value that has been overwritten with a "1".

A common misconception concerning the writing of data to a hard drive arises as many people believe that a digital write is a digital operation. As was demonstrated above, this is a fallacy, drive writes are analogue with a probabilistic output [6], [8], [10]. It is unlikely that an individual write will be a digital +1.00000 (1). Rather - there is a set range, a normative confidence interval that the bit will be in [15].

What this means is that there is generally a 95% likelihood that the +1 will exist in the range of (0.95, 1.05) there is then a 99% likelihood that it will exist in the range (0.90, 1.10) for instance. This leaves a negligible probability (1 bit in every 100,000 billion or so) that the actual potential will be less than 60% of the full +1 value. This error is the non-recoverable error rating for a drive using a single pass wipe [19].

As a result, there is no difference to the drive of a 0.90 or 1.10 factor of the magnetic potential. What this means is that due to temperature fluctuations, humidity, etc the value will most likely vary on ***each*** and ***every*** pass of a write. The distributions of these reads are displayed as histograms in Fig. 5. The distribution is marginally different to the original, but we cannot predict values. From Fig. 6 it is simple to see that even with the prior data from the initial write we gain little benefit. These images display the differences in the voltage readings of the drives (which are determined through the magnetic field strength). Clearly, some values that are more distantly distributed than would be expected in the differenced results (Fig. 6 B) with voltage values that are significantly greater then are expected. The problem is that the number of such readings is far lower than the numbers that result through shear probability alone.

Resultantly, there is no way to even determine if a "1.06" is due to a prior write or a temperature fluctuation. Over time, the issue of magnetic decay would also come into play. The magnetic flux on a drive decays slowly over time. This further skews the results and raises the level of uncertainty of data recovery.

Consequently, we can categorically state that there is a minimal (less than a 0.01% chance) of recovering any data on a NEW and unused drive that has a single raw wipe pass (not even a low-level format). In the cases where a drive has been used (even being formatted for use) it is not possible to recover the information – there is a small chance of bit recovery, but the odds of obtaining a whole word are small.

The improvement in technology with electron microscopes will do little to change these results. The error from microscope readings was minimal compared to the drive error and as such, the issue is based on drive head alignment and not the method used for testing.

1.3 Read Error Severities and Error Management Logic

A sequence of intricate procedures are performed by the hard drive controller in order to minimise the errors that occur when either writing data to or reading data for a drive. These processes vary with each hard drive producer implementing their own process. Some of the most common error management processes have been listed below.

ECC Error Detection: A drive sector is read by the head. An error detection algorithm is used to determine the likelihood of a read error. In the event that an error state is considered to be unlikely, the sector is processed and the read operation is considered as having been concluded successfully.

ECC Error Correction: The controller uses the ECC codes that it has interpreted for the sector in order to try and correct the error. A read error can be corrected very quickly at this level and is usually deemed to be an "automatic correction".

Automatic Retry: The next phase involves waiting until the drive platter has completed a full spin before attempting to read the data again. Stray magnetic field variances are a common occurrence leading to drive read error. These fluctuations may result due to sudden movement and temperature variations. If the error is corrected following a retry, most drives will judge the error condition to be "corrected after retry".

Advanced Error Correction: Many drives will, on subsequent retries after the first, invoke more advanced error correction algorithms that are slower and more complex than the regular correction protocols, but have an increased chance of success. These errors are "recovered after multiple reads" or "recovered after advanced correction".

Failure: In the event that the drive is incapable of reading the sector, a signal is sent to the drive controller noting a read error. This type of failure is an unrecoverable read error.

Modern encoding schemes (PRML and EPRML) have a wide tolerance range allowing the analogue values that the drive head reads from and writes to a hard disk to vary significantly without loss of data integrity. Consequently, the determination of a prior write value is also a stochastic process.

2 Data and Method

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 also divided into a second category in order to test a number of write patterns. 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 (from an older Quantum 1 GB drive to current drives dated to 2006). The data patterns where written to each drive in all possible combinations.

1. The data write was a 1 kb file (1024 bits).
2. Both drive skew and the bit was read.
3. The process was repeated 5 times for an analysis of 76,800 data points.

Table 1. Table of Probability Distributions for the older model drives. Note that a "used" drive has only a marginally better chance of any recovery than tossing a coin. The Pristine drive is the optimal case based on an early Seagate 1Gb drive.

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[5]	0.51321887	0.009672
16 bits	0.26339361	9.35E-05
32 bits	**0.06937619**	**8.75E-09**
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

[5] This represents one (1) ASCII character.

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. This presents this method with an advantage to recovering the data that would not be found when conducting a forensic examination and recovery of a drive.

Even on a single write, the overlap at best gives a probability of 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.

Resultantly, if there is less than a 1% chance of determining each character to be recovered correctly, the chance of a complete 5-character word being recovered drops exponentially to 8.463E-11 (or less on a used drive and who uses a new raw drive format). This results in a probability of less than 1 chance in 10Exp50 of recovering any useful data. So close to zero for all intents and definitely not within the realm of use for forensic presentation to a court.

Table 1 below, shows the mapped out results of probable recovery with a pristine drive of a similar make and model[6] to that which would have been used in the paper by Dr. Gutmann. This drive had never been used and was had raw data written to it for the first time in this test. The other drive was a newer drive[7] that has been used (I used this for my daily operations for 6 months) prior to the wiping procedure. A total of 17 variety of drives dated from 1994 to 2006 of both the SCSI and IDE category where tested for this process. A total of 56 drives where tested. On average only one (1) drive in four[8] (4) was found to function when the platter had been returned after an initial reading with the MFM.

3 Data Relationships

The only discernable relationship of note is between an initial write of a “1” on a pristine drive that is overwritten with a “0”. This is a function of the drive write head and has no correlation to data recovery, so this is a just point of interest and noting to aid in data extraction from a forensic perspective. All other combinations of wipes displayed comparative distributions of data that where suggestive of random white noise.

3.1 Distributions of Data

The tables used in this section display the probabilities of recovery for each of the drives tested. Although the chances of recovering any single bit from a drive are relatively good, the aim in any forensic engagement is to recover usable data that can be presented in court.

---

[6] SEAGATE: ST51080N MEDAL.1080 1080MB 3.5"/SL SCSI2 FAST.
[7] Western Digital WD1200JS.
[8] 23.5% of drives where able to be used for an overwrite following an initial MFM scan.

These tests where run as a series of 4 tests on each of 17 types of drives. The reported (Table 1) recovery rate of 92% this was the optimal rate (which was itself stochastically distributed). The results were in distributed over a wide range of values with the use of the drive impacting on the capacity to recover data.

This clearly shows that any data recovery is minimal and that no forensically sound recovery is possible. The recovery of a single 32 bit value (such as an IP address) is highly unlikely. It has been stated[9], that the smallest fragment of usable digital forensic evidence is a 32 bit field (the IP address). To be used in a Civil court case, the evidence needs to be subjected to the balance of probability (usually 51%). In a criminal matter, the preponderance is set at between 95% and 99% to account for all reasonable doubt. The rate at which evidence may be recovered using this technique is too low to be useful. In fact, with the optimal recovery under 7% for a single IP address on an older drive. This is an event that cannot occur outside the lab.

The bit-by-bit chance of recovery lies between 0.92 (+/- 0.15)[10] and 0.54 (+/- 0.16)[11]. We have used the higher probability in the calculations to add an additional level of confidence in our conclusions. This demonstrates that the chances of recovering a single 8-bit character on the pristine drive are 51.32%. The recovery rate of a 32-bit word is 0.06937619 (just under 7%). As such, the chances of finding a single 4 letter word correctly from a Microsoft Word document file is 2.3166E-05 (0.00002317%)

Table 2 below is a table that further illustrates the wiping fallacy. We tested this by completing a single pass wipe, to simulate minimal use we repeated the process.

Once again, we can see the data recovery is minimal.

Table 2. Table of Probability Distributions for the "new" 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
32 bits	**0.011604223**	**6.2771E-07**
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	5.8096E-100
1024 bits	1.16873E-62	3.3752E-199

The standard daily use of the drive makes recovery even more difficult, without even considering a wipe, just ***prior*** use. In this case, the 3 former wipes are used to simulate use (though minimal and real use is far more intensive). The chances of

[9] Rob Lee, SANS Forensics 508.
[10] For the optimal recovery on an old drive.
[11] On a used "new" drive.

recovering a single 8-bit word (a single character) are 0.0281475 (or 2.8%) – which is actually lower than randomly selecting the character.

The calculated probability of recovering data from any used drive that uses a newer encoding scheme (EPRML) and high density was indistinguishable from a random guess. When recovering data from the 2006 model drive, the best determination of the prior write value was 49.18% (+/- 0.11)[12] from the "all 0's" pattern when overwritten with the "all 1's" pattern. 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%. In many instances, using a MFM to determine the prior value written to the hard drive was less successful than a simple coin toss.

3.2 Distribution of Recovered Data

The following is a retrieval pattern from the drive. Where the 8-bit word is correctly read, a "1" is listed. Where the value did not match the correct pattern that was written to the drive, a "0" is displayed.

```
 [1]   0 0 1 0 1 0 0 1 0 0 1 1 1 0 1 0 1 1 1 1 1 0 0 1 0 0 1 0 1 0 1 1 1 0 1 0 0 0 0 0 0 0 1 0 1 1 1
[48]  0 1 0 1 0 0 0 1 0 1 1 1 1 1 1 0 0 0 0 0 0 1 0 0 0 0 1 1 0 1 1 1 1 0 0 0 1 0 1 0 1 1 1 1 0 1 0 0 0
[95]  0 1 1 0 0 1 0 0 1 1 0 1 0 0 1 0 1 0 0 0 1 0 1 1 1 0 0 1 0 0 1 1 0 0 0 0 1 0 1 0 1 0 1 0 0 1 1
[142] 1 1 1 1 1 1 1 1 0 0 1 1 1 1 0 0 0 0 0 0 1 0 0 1 0 0 0 1 0 0 1 0 1 1 0 1 1 1 0 0 1 1 0 0 0 1 0 0
[189] 1 0 1 1 0 1 0 1 0 0 1 1 1 0 1 0 1 1 0 1 0 1 1 1 1 0 1 1 0 0 1 0 0 1 0 1 0 0 1 1 0 0 0 0 1 0 1
[236] 0 0 1 0 1 1 1 0 1 1 0 0 1 1 1 1 1 1 1 1 0 0 1 1 1 0 0 1 0 1 0 1 0 0 1 0 0 0 0 0 0 0 0 1 0 1 0 1
[283] 0 1 0 1 0 1 1 0 0 0 1 1 0 0 0 1 0 0 1 0 0 1 1 1 0 0 1 1 1 1 1 1 0 0 1 1 0 0 1 1 1 1 1 0 0 0 0
[330] 0 0 0 0 0 0 0 1 1 0 1 0 1 1 1 0 1 0 1 1 0 0 0 1 1 0 0 0 1 1 0 1 0 1 0 0 0 0 0 1 1 0 0 1 0 1 0
[377] 1 1 1 0 1 0 1 0 0 1 0 0 0 0 0 0 0 1 0 0 0 1 0 0 1 1 1 1 1 1 0 1 1 0 1 1 0 1 0 1 1 1 1 0 1 0 0 0 1
[424] 1 1 0 0 1 1 1 0 0 0 0 1 1 1 0 1 0 0 0 0 1 0 0 1 0 0 0 1 1 0 0 0 0 1 0 0 1 1 1 1 0 1 1 1 1 0 0 0 1
[471] 1 1 1 1 0 1 1 1 1 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 1 0 1 1 0 0 1 0 1 1 1 1 1 1 1 1 0 1 0 1 0 0 0 0
[518] 1 0 0 1 0 1 1 1 1 1 0 0 1 0 1 1 0 1 0 1 1 1 1 1 0 0 1 0 0 0 1 0 0 0 0 1 1 0 0 0 0 1 1 0 0 0 0
[565] 1 1 1 0 0 0 0 0 0 1 0 0 1 0 0 1 0 0 1 1 1 1 0 0 0 0 0 1 0 1 1 1 1 1 1 1 0 1 1 1 1 1 0 0 1 0 0 0
[612] 1 1 0 1 1 1 1 0 1 1 1 1 0 0 0 1 1 0 0 0 1 1 0 0 1 0 1 1 0 0 1 0 0 0 0 0 1 0 1 0 1 0 0 0 0 1 0 0
[659] 1 1 0 1 0 1 0 1 1 0 0 0 0 0 0 1 1 0 1 0 0 1 1 1 0 0 0 1 1 1 0 1 1 0 0 0 0 0 1 1 1 1 1 0 0 1 1
[706] 1 0 0 1 1 1 0 1 0 1 0 1 1 0 1 0 0 0 0 0 0 1 1 0 1 1 1 0 1 0 0 1 0 1 0 1 0 0 1 1 0 0 0 0 0 0 0 1
[753] 1 0 0 1 1 1 0 0 1 0 1 1 0 0 0 0 1 1 1 1 1 1 0 0 0 1 0 0 1 1 1 0 0 0 0 1 0 1 0 0 0 1 0 1 1 0 0 0
[800] 0 0 1 0 0 0 1 0 0 0 0 0 0 0 1 0 1 1 1 1 0 1 0 1 0 0 0 0 0 1 0 1 1 1 1 1 1 0 0 1 0 0 1 0 0 0 0 0 0 0
[847] 0 0 0 0 1 0 1 0 0 0 0 0 0 0 1 0 1 1 0 1 1 1 0 1 1 0 0 0 1 1 1 1 1 1 1 0 0 1 0 0 0 0 0 1 1 1 0 0 1 0
[894] 1 1 1 0 1 1 0 0 1 0 0 0 0 0 1 1 0 1 1 1 1 1 0 1 1 0 0 1 0 0 1 0 0 1 1 1 0 1 1 0 0 1 0 0 1 0 0 1 1
[941] 1 0 0 0 1 0 1 0 1 0 1 1 1 1 1 1 0 1 1 0 0 0 1 0 0 1 1 1 1 0 0 1 0 1 1 1 1 1 1 1 0 0 1 1 1 0 1 0
[988] 0 0 1 0 1 1 1 0 1 0 0 1 1 1 0 0 1 1 0 0 1 0 0 0 0 0 1 1 1 1 1 1 0 1 0 0 1 1 0
```

As an example, the following is the start of the paper by Peter Gutmann [12], first displayed accurately, and next at an optimal retrieval level.

3.2.1 Correct Display

Secure deletion of data - Peter Gutmann - 1996
Abstract
With the use of increasingly sophisticated encryption systems, an attacker wishing to gain access to sensitive data is forced to look elsewhere for information. One avenue of attack is the recovery of supposedly erased data from magnetic media or random-access memory.

3.2.2 Display from Recovery (Optimal)

‰o
'cKræ}d8Œeti²n•of0daÊI0Ptr0G§tWÇîï_¼Á1u960eb8tÈñutW00000Dç•Ã#Ì0
Hf$00\000%£z0\0ã0000á0áä«it\tþÛ0u³e•Fƒ°i™%\eàsinqTyøîopÚ"Ë:i†aze0

[12] This is reported at a 99% confidence level.

®Mcryption0sîÙtems?DKtA""cÐÏ0+¢sinŒ0toK–ai2z÷c(ns~0tü0;e
½iti)e""daÆa>s0foôce¸ÑtÒÍl2o–
ìelI¶˜$eöe›Ÿr""inf¬rm‰ion.0OnRïavem>egoN0-¨ïtRÁ"1i
läßh±0"eÛoie=y0Cz-
*su•¨s/ˆlÜ{era'Jd0dataF¨ro>•magne³;&£õãÈáã%or*r%ndoª-Qcc«ÇŸ0mà*
@ryl00000000000000000

Although on the perfect drive some words could be recovered, there is little of forensic value.

3.2.3 Display from Recovery (Expected)

¡ÄuÜtÞdM@ª""îFnFã:à•ÅÒÌ¾‘¨L‘¿ôPÙ!#¯-×LˆÙÆ!mC
2´³„‡·}NŽýñêZØ^›l©þì®·äÖŒv¿^œº0TÏ[ªHÝBš¸ð
7zô|»òëÖ/""º[ýÀ†,kR¿xt¸÷\Í2$Iå""•ÑU%TóÁ'ØoxÈ$i
Wï^™oËS²Œ,Ê%ñ ÖeS» eüB®Èk‹|YrÍÈ¶=ÏÌSÃ¡öp¥D
ôÈŽ"|ûÚA6¸æ÷U•$µM¢;Òæe•ÏÏMÀùæç]#•Q
————————————Á¹Ù""—OX“h
ÍýïÉûË Ã""W$5Ä=rB+5•ö–GßÜä9ïõNë-ß¨Ya“–ì%×Ó¿Ô[Mãü
·†Î,ƒ,…[Ä,KDnFJˆ·×ÅŒ¿êäd¬sPÖí8'v0æ#!)YÐúÆ©
k-‹HÃˆø$°•Ø°Ïm/Wîc@Û»Ì"„zbíp00000000000000000

On the drive that had been wiped 3 times (prior) to the data being written and then added, the results are worse. What needs to be noted is that small errors in the calculations lead to wide discrepancies in the data that is recovered. Further, it needs to be noted that any drive recovered is not likely to be in a pristine state. The daily use of a drive reduces the chances of recovery to a level that is truly insignificant.

4 Conclusion

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.

References

1. Abramowitz, M., Stegun, I.A.: Handbook of Mathematical Functions. Dover, New York (1965)
2. Amit, D.J.: Field Theory. In: The Renormalization Group and Critical Phenomena. World Scientific, Singapore (1984)
3. Braun, H.B.: Fluctuations and instabilities of ferromagnetic domain-wall pairs in an external magnetic field. Phys. Rev. B. 50, 16485–16500 (1994)
4. Brown, G., Novotny, M.A., Rikvold, P.A.: Thermal magnetization reversal in arrays of nanoparticles. J. Appl. Phys. 89, 7588–7590 (2001)
5. Bulsara, A., Chillemi, S., Kiss, L., McClintock, P.V.E., Mannella, R., Marchesoni, F., Nicolis, G., Wiesenfeld, K. (eds.): International Workshop on Fluctuations in Physics and Biology: Stochastic Resonance, Signal Processing and Related Phenomena, p. 653. Nuovo Cimento 17D (1995)
6. Carroll, T.L., Pecora, L.M.: Phys. Rev. Lett. 70, 576 (1993a)
7. Carroll, T.L., Pecora, L.M.: Phys. Rev. E 47, 3941 (1993b)
8. Gomez, R., Adly, A., Mayergoyz, I., Burke, E.: Magnetic Force Scanning Tunnelling Microscope Imaging of Overwritten Data. IEEE Transactions on Magnetics 28(5), 3141 (1992)
9. Gammaitoni, L., Hänggi, P., Jung, P., Marchesoni, F.: Stochastic resonance. Reviews of Modern Physics 70(1) (January 1998)
10. Gomez, R., Burke, E., Adly, A., Mayergoyz, I., Gorczyca, J.: Microscopic Investigations of Overwritten Data. Journal of Applied Physics 73(10), 6001 (1993)
11. Grinstein, G., Koch, R.H.: Switching probabilities for single-domain magnetic particles. Phys. Rev. B 71, 184427 (2005)
12. Gutmann, P.: Secure Deletion of Data from Magnetic and Solid-State Memory. In: Proceedings of the Sixth USENIX Security Symposium, San Jose, CA, July 22-25, pp. 77–90 (1996), http://www.cs.auckland.ac.nz/~pgut001/pubs/secure_del.html
13. Hänggi, P., Bartussek, R.: In: Parisi, J., Müller, S.C., Zimmermann, W. (eds.) Nonlinear Physics of Complex Systems: Current Status and Future Trends. Lecture Note in Physics, vol. 476, p. 294. Springer, Berlin (1991)
14. Liu, D.: Topics in the Analysis and Computation of Stochastic Differential Equations, Ph. D. thesis, Princeton University (2003)
15. Mayergoyza, I.D., Tse, C., Krafft, C., Gomez, R.D.: Spin-stand imaging of overwritten data and its comparison with magnetic force microscopy. Journal Of Applied Physics 89(11) (2001)
16. Moss, F.: In: Weiss, G.H. (ed.) Contemporary Problems in Statistical Physics, pp. 205–253. SIAM, Philadelphia (1994)
17. Ren, W.E., Vanden-Eijnden, E.: Energy landscape and thermally activated switching of submicron-size ferromagnetic elements. J. Appl. Phys. 93, 2275–2282 (2003)
18. Reznikoff, M.G.: Rare Events in Finite and Infinite Dimensions, Ph. D. thesis, New York University (2004)
19. Rugar, D.H.M., Guenther, P., Lambert, S., Stern, J., McFadyen, I., Yogi, T.: Magnetic Force Microscopy: General Principles and Application to Longitudinal Recording Media. Journal of Applied Physics 68(3), 1169 (1990)
20. Tesla, N.: The Great Radio Controversy, http://en.wikipedia.org/wiki/Invention_of_radio
21. Jiles, David: Introduction to magnetism and magnetic materials, 2nd edn. Chapman & Hall, Boca Raton (1998)