Secure Erase of Disk Drive Data

Gordon Hughes and Tom Coughlin

Introduction

A cardinal rule of computer data storage devices is to protect user data at all costs. This includes protection against accidental erasure, using "recycle" folders and unerase commands. Drives use elaborate error detection and correction techniques to never return *incorrect* user data. All this implies that true file erasure is an abnormal situation.

Consequently, user data often remains stored on disk drives when they are discarded, transferred to another user, or returned off lease. Even if users delete their files, they can be recovered from "recycling" folders or by special programs such as Norton Unerase. Other special programs are available to more definitely erase user data, such as Norton WipeInfo, which offers a "Government Wipe." Because these are special programs, they are not widely known or used. Additionally, some user data can be unreachable by erasure programs, such as drive data blocks removed from use due to excessive errors. (These blocks are reassigned to other disk locations, and the defective record area is marked unusable in the drive's "g-list"- but the original data can remain in g-list sectors.) Norton WipeInfo only runs on older Windows OS, and wipes an individual folder.

A "secure erase" (SE) command is being added to computer disk drives. Secure Erase ("SE") is a positive, easy-use *data destroy* command, amounting to "Electronic data shredding," It will be easy to use, not require any special software, and completely erase all possible user data areas. SE is a simple addition to the existing "format drive" command currently present in computer operating systems and storage systems. SE commands have been added to the standard interface specifications for disk drives, ATA for desktop or personal computers, and SCSI for enterprise drives. These commands define a drive internal SE operation that erases all possible user accessible record areas by overwriting. Although SE add can be implemented easily by modifying the drive Format Unit firmware, while providing a feature of potential value to many drive customers, it is not yet widely implemented in disk drives.

Secure erasure capability will soon be required by the U.S. government for their disk drive purchases. Considering the additional security and features this capability provides to many users we expect there will be considerable commercial interest in this capability as well.

However, some engineering is necessary to incorporate SE into drives. Magnetic recording SE development is required, before drive firmware can be modified to carry out SE. These techniques have been studied at the Center for Magnetic Recording Research at the University of California San Diego.

This article discusses SE recording technology. It will be shown that a simple format drive command is effective, if low frequency (LF) overwriting frequencies are used, the g-list is also overwritten, and all possible user record areas are erased. Pseudo-random LF overwriting patterns are recommended for maximum security. Drive and spinstand experiments are shown, and a protocol of SE validation tests.

To illustrate the effectiveness of SE, exotic data recovery experiments will also be shown. These use digital scope signal averaging and computer correlation techniques to recover overwritten user data (erased beyond recovery by normal drive read channels.) The ATA drive spec defines two SE levels, normal and enhanced. A two-pass SE will be discussed for the latter, involving a combination of low frequency and random data writing, and positioning the heads ± 5-10% off-track. This defeats exotic recovery techniques, and also erases the track edge signals (which will be shown to be primarily transition noise in the tests here, not signal).

Background

Experience gained from supporting customers with computer security and information storage media issues have shown a tremendous need for the capability to reliably eradicate data from computer hard drives for security and privacy reasons. The need arises when:

Mainframes and storage networks:

- When a user releases storage, a drive moves to a new storage server, is removed for maintenance, or returned from lease.
- Storage devices are re-configured for other uses or users, for instance in expiring leased data storage facilities at an SSP or data center
- A RAID drive backs up data to a hot spare

Individual user PCs and workstations:

- A computer (and hard drive) is replaced by a newer machine and the older machine is discarded, or sold
- A project is completed and the data must be purged to protect "need to know" or to prepare the drives for new users or applications.
- When a user departs an organization and either leaves sensitive/personal data on the computer or may take the computer (and the organization's data) with them.
- When a drive is to be returned to a drive manufacturer or a drive repair facility after a drive failure or near failure (for instance upon a SMART drive replacement after imminent failure is determined.
- A virus has been detected and all possible traces of the offending code must be eliminated.
- An extreme virus or hacker attack where it is desirable to completely erase the data on some disks and reinstall back-up data

The elimination of unwanted data from a computer hard drive is not a simple task. Deleting a file merely removes its name from the directory structure; the data itself remains in the drive's data storage sectors where it can be retrieved until the sectors are overwritten. Reformatting a hard drive clears the file directory and severs the links between storage sectors, but the data can be recovered until the sectors are overwritten. Software utilities that overwrite individual data files (or an entire hard drive) are susceptible to error and require constant modifications to accommodate new hardware and evolving computer operating systems. As a consequence, computer users, system administrators, security personnel and service providers have spent considerable time in an endless game of technology catch-up while trying to develop solutions for the above problems.

Disk Drive Secure Erasure

In 1996 researchers at the University of California, Center for Magnetic Recording Research (CMRR) in San Diego and the federal government began an investigation into incorporating a secure overwrite feature into new hard disk drives. This secure overwrite feature when enabled would allow hard disk drives to automatically overwrite all stored data via a simple external command. In discussions with U.S. hard drive manufacturers, they expressed their support for the secure overwrite or erasure feature. Fundamental issues needing to be addressed included:

- Determining how to overwrite data for security purposes
- Confirming secure erasure was implemented correctly
- Confirming the effectiveness of secure erasure through the entire drive environmental specification
- Making a simple and safe user interface for secure interface
- Creating sufficient end-user demand to make secure erasure economically feasible.

Modern hard drives store information in sectors that contain addressing fields, timing signals, the user's data and appropriate error correction coding, and other features essential for the drive's proper operation. In order to accomplish secure erasure and continued operation of the disk drive, secure erasure must overwrite only the data portion of the sector without disturbing the addressing information or timing codes that would make the drive unusable.

Implementation of Secure Erase

Erasure can be at various fixed frequencies (including DC) and either on or off-track. The erasure can be performed once or several times. Spin-stand experiments were performed at CMRR at UCSD to determine the effectiveness of various erasure techniques.

For a constant frequency 20 MHz "signal" written on a data track it was found that an erasure resulting in the signal reduction of over -50 dB could be accomplished with a single on-track overwrite. This single on-track overwrite is more effective as the overwritten frequency is reduced (Figure 1)

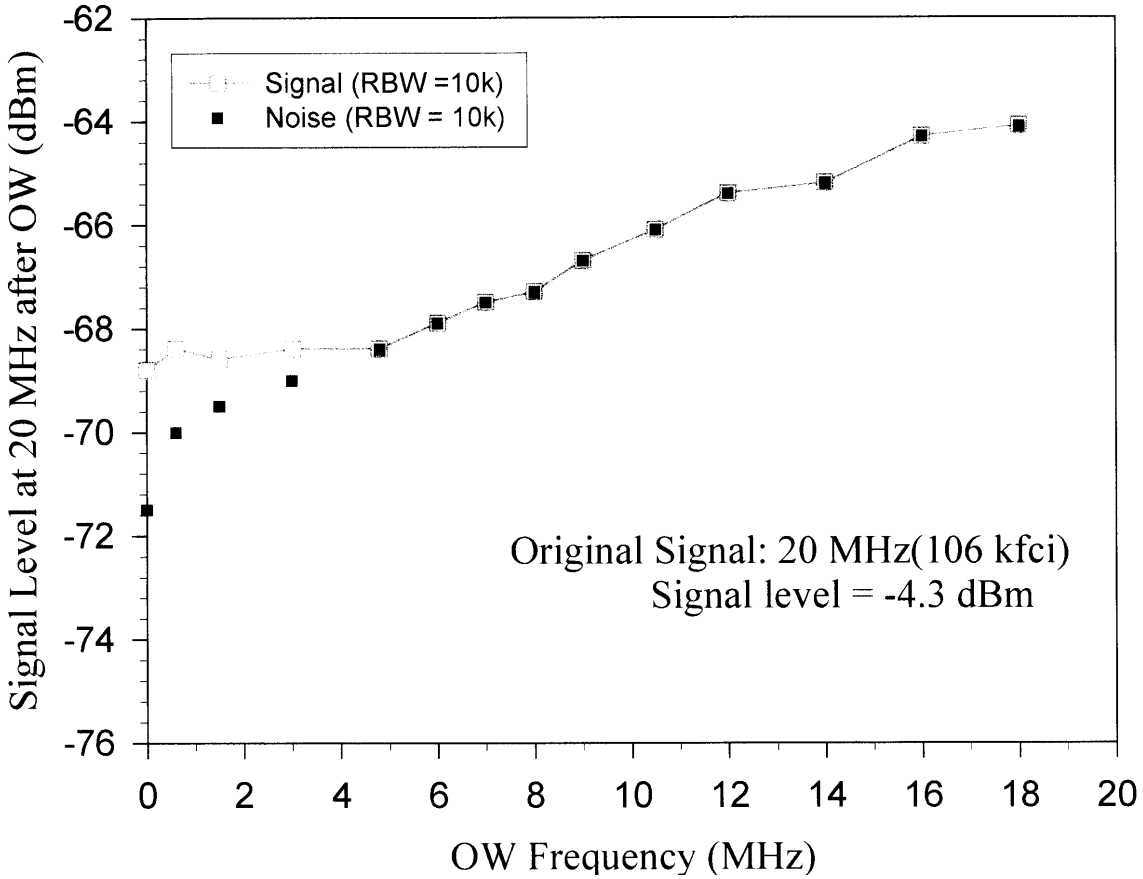

Figure 1. Test showing that LF gives best overwrite. Noise level is the overwrite limit at higher frequencies, because transition noise from overwriting signal dominates.

A further reduction of the original signal of a few dB is possible for a second on-track overwrite. Lower overwriting frequencies are more effective in overwriting old signals because they write a wider track than the original higher frequency signal. Thus the lower frequency overwritten track will more effectively erase the edges of the original signal, even with a single overwrite. For a single low frequency overwrite the SNR of the remaining 20 MHz signal is <10 dBm, this is only marginally worse with a second overwrite. Modern channels require the head playback SNR to be greater than 19 dB. Erased signal SNR of this level will be impossible to

recover using a disk drive channel. Thus a single on-track overwrite will usually be sufficient to erase the prior information to a level that would be impossible to recover within a disk drive, using its own channel.

It is possible that under extreme environmental differences between when the data was written and when the overwritten track is recorded the two written tracks will be significantly offset from each other. Under these conditions it is possible that some of the original track could remain. If the overwritten track is at a much lower frequency (usually as low as possible) the amount of original data remaining at the track edge will be significantly reduced.

Track edges can contain highly distorted original data as well as transition noise and are therefore difficult to recover data from. Figure 2 shows a crosstrack scan of the residual signal, when a swath of tracks is DC-erased (blue curve) and when a 20 MHz square wave is overwritten by 10.5 MHz (red curve), chosen so the overwriting signal doesn't have harmonics at 20 MHz. The scan does not change if the 20 MHz signal is not written at all (black curve), so it must be entirely due to the overwriting 10.5 MHz. This means that although the amplitude peaks 4-6 dB at the track edges, this cannot be unerased 20 MHz signal. A full spectral analysis (Figure 3) shows that these track edge peaks are primarily ac-erased transition noise (the noise floor at 20 MHz rises 4 dB above the DC erase level, when 10.5 MHz is recorded). The poor write field gradients off the head sides ac erases a higher level of transition noise than the well-written 10.5 MHz transitions in the main track.

One erasure pass appears to be sufficient to make old data unrecoverable. A two-pass erasure can provide an additional level of security. Writing LF helps erase the track edges, and then overwriting with random data defeats exotic techniques (see later section of this article). The two passes can be slightly off-track in the positive and negative direction in order to ensure elimination of the track edges.

Secure erasure can take some time and is likely to be an off-line operation. For example, a Seagate Barracuda 180 GB disk drive spinning at 7200 RPM with 24 heads and 24,247 tracks per disk surface a complete single-pass overwrite of all user blocks in about 40 minutes. A double-pass overwrite of all user blocks takes about 80 minutes.

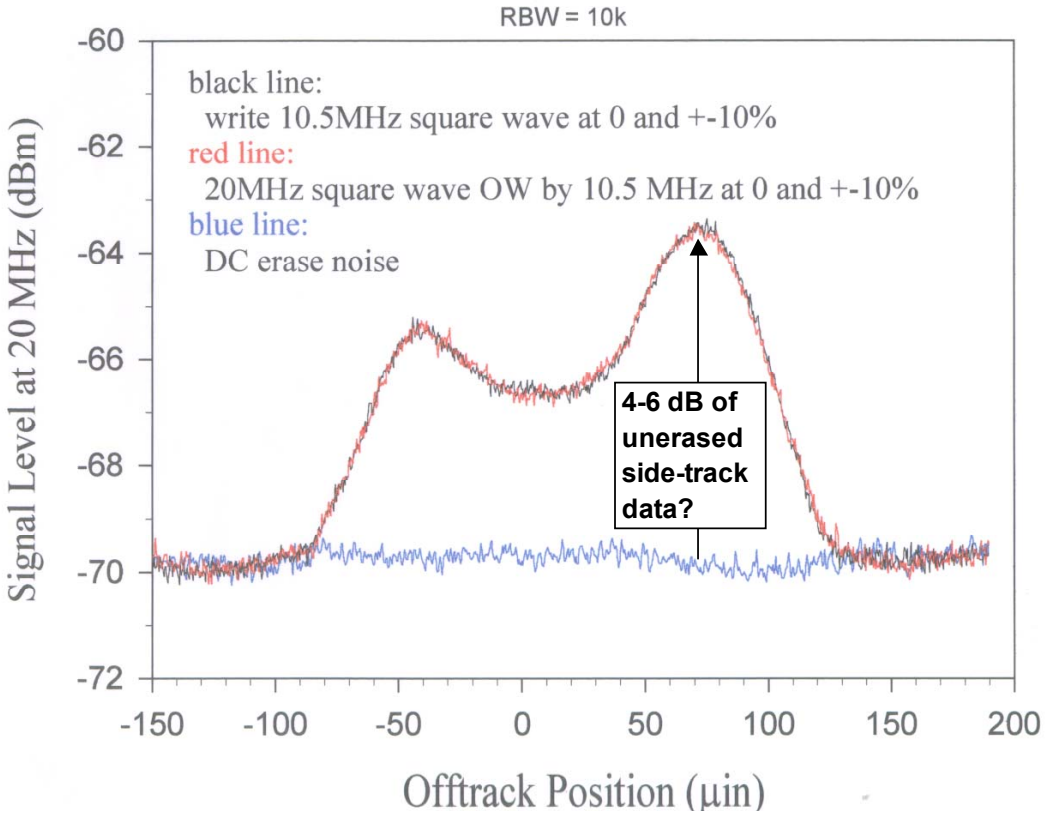

Figure 2: 20 MHz data overwritten by 10.5 MHz. Are the track edge peaks old 20 MHz data?

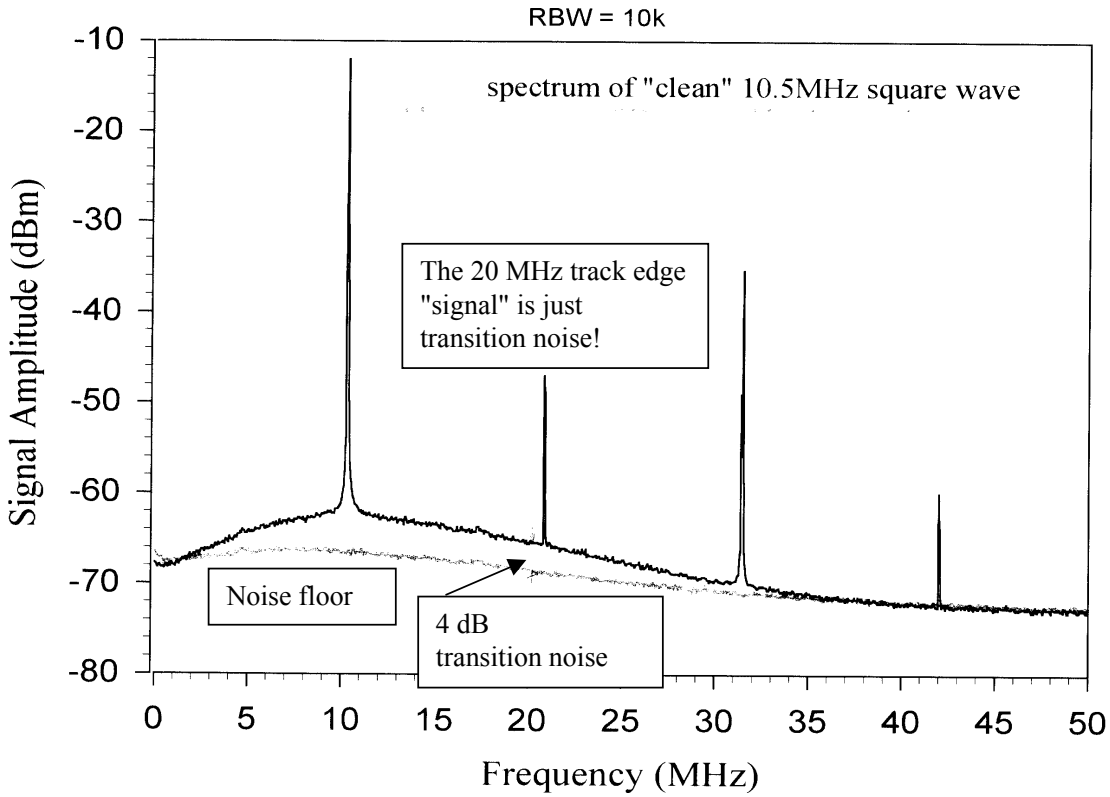

Figure 3. NO! The track edge peaks are primarily ac-erased transition noise

Secure Erasure Standards

A “secure erase” (SE) feature has been added to the standard specifications for disk drives, ATA for desktop or personal computers, and SCSI for enterprise drives. These require that an SE operation erase all possible user accessible record areas, by overwriting. This feature will be as easy to use as a “format drive” command, when SE support appears in drives and in computer operating systems

The SE feature “piggybacks” on the Format Unit command already in these specs, thus adding little or no cost to a drive, while providing a feature of potential value to many drive customers.

There will be engineering details for disk drive companies to implement the secure erasure protocols and computer operating system companies need to add the already-defined SE command to their format drive software. CMRR researchers are available to work with drive companies and their customers on implementation of single-pass (regular) as well as double-pass (enhanced) secure erasure.

Exotic Data Recovery

Drive information can sometimes be recovered that has been erased using a single erasure pass on-track. It should be first pointed out that single-frequency squarewave overwrite tests are not meaningful indicators of information recovery. A spectrum analyzer can see -60 dB overwritten signals but it can’t recover data. The CMRR technique requires reading a data block many times, computer averaging the playback waveforms, then erasing the block and re-recording the overwrite data to obtain its averaged playback waveform data, which is subtracted from the first waveform. The demonstration shown below merely means that it is possible, not that it is practical or will work on any drive. It requires knowing the data pattern being looked for, and also knowing the overwriting data pattern. So it “begs the question.” It can be defeated by using a random data overwrite pattern.

Data recovery techniques showed that SE overwritten user data is beyond recovery by normal drive read channels. But CMRR could recover overwritten user data by digital scope signal waveform averaging, software correlation techniques. These recording experiments were done on a spinstand using a dual stripe MR head 2.5 μm write width, 1.8 μm read width, 2500 Oe/0.65 Mrt disk. These had 38 dB overwrite and a playback noise floor -69 dB.

First a swath of tracks was DC erased, then arbitrary user bits “HELPHELP” were written, the playback waveform was averaged 100 times to improve SNR, and saved as waveform 1. Then a random bits overwrite, the same signal averaging, and saved as waveform 2. Another DC erase followed by overwriting with the same random bits and signal averaging, is saved as waveform 3.

Figure 4 shows the averaged initial playback waveform 1, and the residual wave 3 minus wave 2. Not much residual signal left!

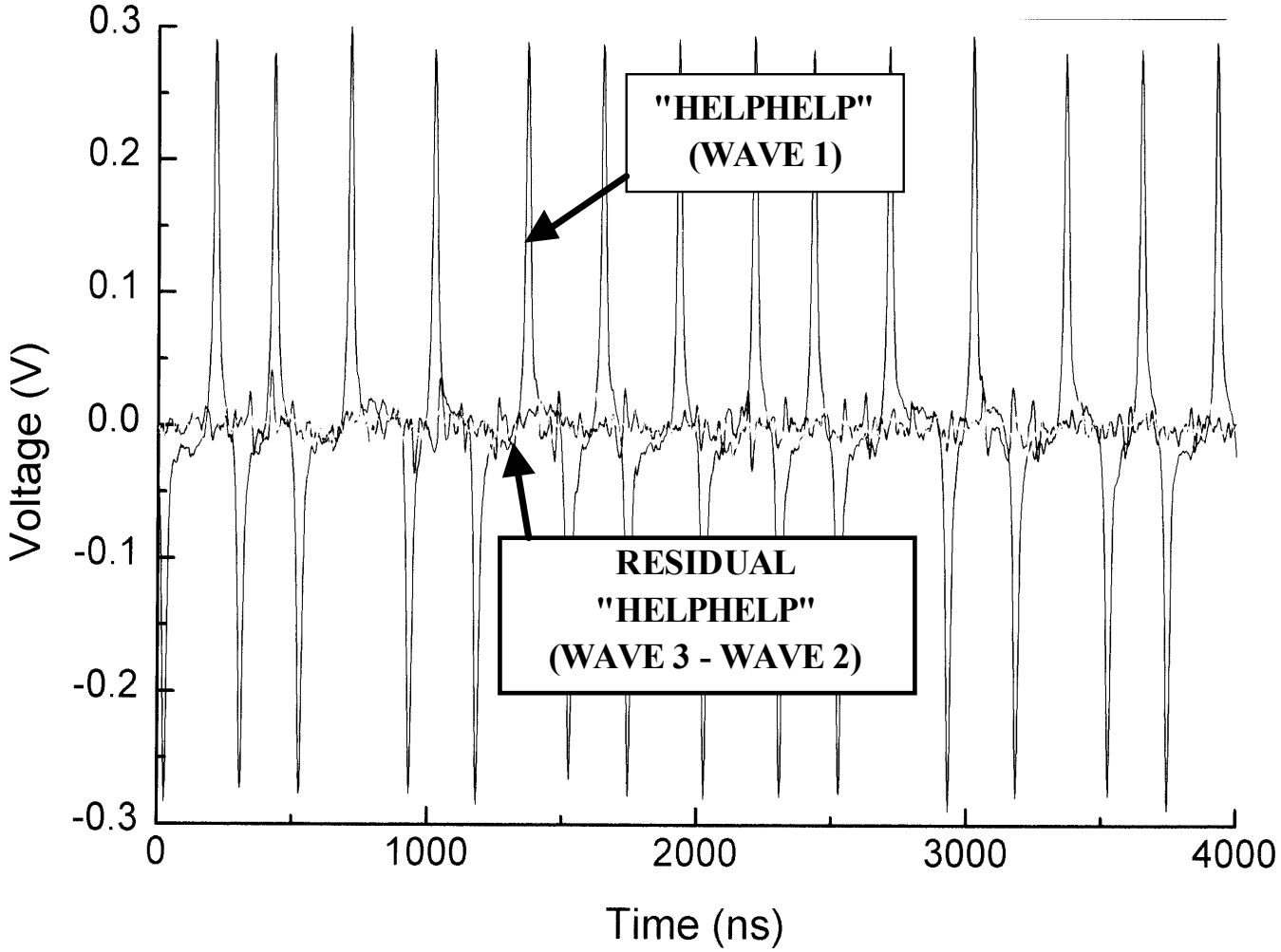

Figure 4The original “helphelp” waveform and the residual waveform 3 minus waveform 2

Figure 5 shows results from a correlation detector. The residual correlation has a correct peak at zero offset, which is about twice the second highest peak. This indicates that “helphelphelphelp” is correctly detected (but a 2:1 signal-to-“noise” ratio means high error rate). (When this was repeated with TWO random overwrites, correlation detection did not work.)

Beyond these data recovery techniques which use drive hardware, other exotic techniques can be proposed such as putting recorded discs into scanning magnetic force microscopes. It is easy to obtain pictures that appear to show unerased track edge data. But no one has shown complete recovery of a data sector, including the data synchronization preamble, bit de-randomizer, partial response and modulation codes, and error correction code.

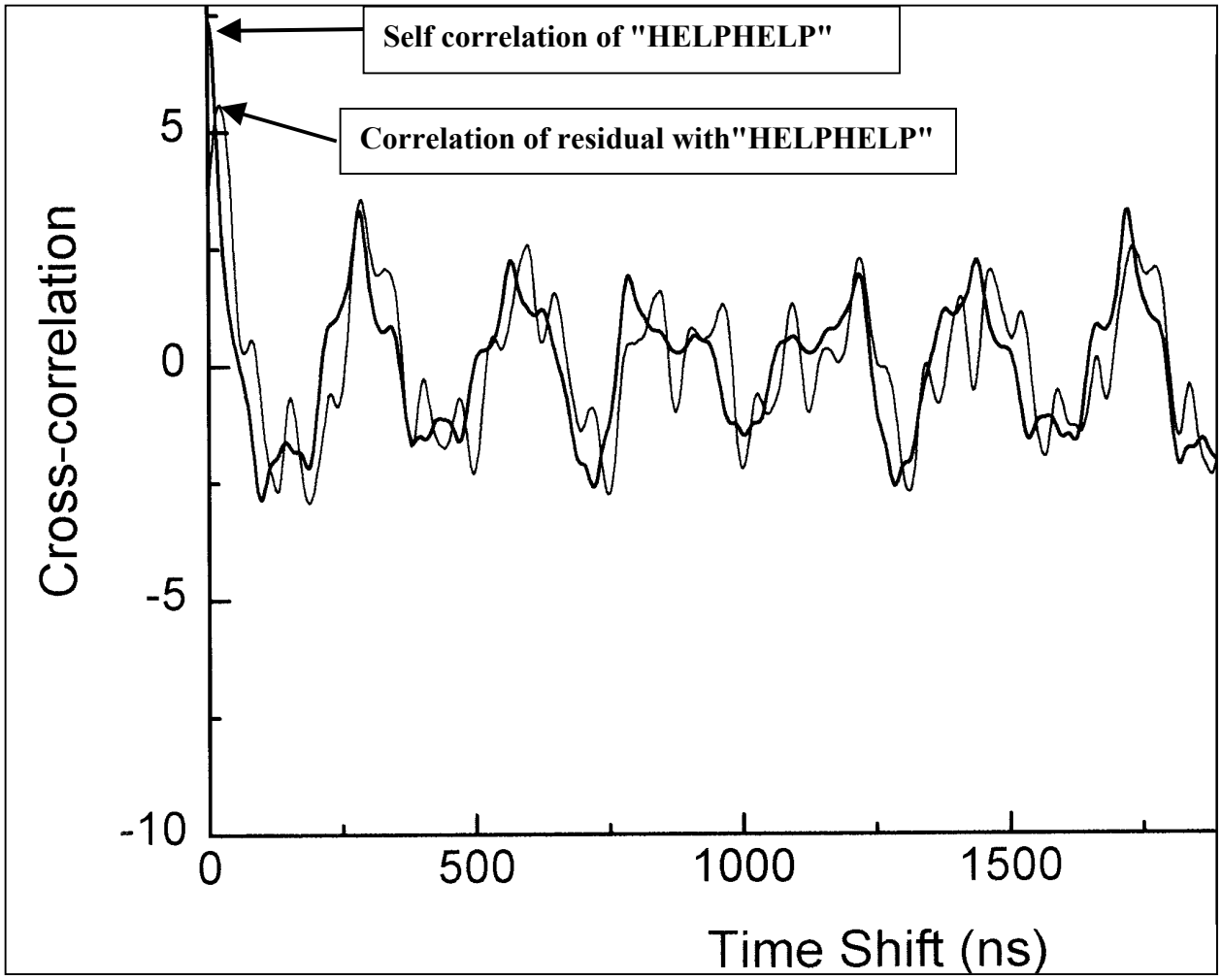

Figure 5. Correlation detector worked! Matches Highest peak

About the Authors

Gordon Hughes is the Associate Director of the Center for Magnetic Recording Research at the University of California, San Diego. He is the principal investigator of the UCSD SMART project on disk drive failure prediction. He received his BS in Physics and Ph.D. in Electrical Engineering from Cal Tech. He worked at Xerox PARC on magnetic recording research for disk drives, then joined Seagate Technology as Senior Director of Recording Technology. At Seagate he designed recording heads, disks, and systems, and was part of the development team that established sputtered thin film disk media as today's standard. He is a Fellow of the IEEE. Contact: 858-534-5317, gfhughes@ucsd.edu, UCSD, La Jolla CA 90293-0401

Tom Coughlin is President of Coughlin Associates, a data storage consulting firm. His current work encompasses technology and market analysis and project consulting in data storage from components to data storage systems. He held engineering and engineering management positions at several disk drive and drive component companies in the past such as Seagate, Maxtor, Micropolis, Nashua Computer Products,Syquest, and Ampex. He has a BS in Physics and a Masters in Electrical Engineering with a minor in Materials Science from the University of Minnesota. He is a senior member of the IEEE and Adjunct Professor at Santa Clara University. For more information see his web site www.tomcoughlin.com. Contact: 408-978-8184, tom@tomcoughlin.com.