

ABSTRACT

Title of dissertation: **PROCESSING OF EX-SITU ACQUIRED SIGNALS FROM MAGNETIC DISKS**

Patrick McAvoy, Doctor of Philosophy, 2008

Dissertation directed by: Professor I. D. Mayergoyz
Department of Electrical and Computer Engineering

The ubiquity and high performance of hard disk drives for nonvolatile digital data storage cannot be denied. As the magnetic recording industry continues to develop new techniques for increasing storage density and reducing cost per bit, diagnostic and forensic tools for characterizing and interpreting the magnetic patterns recorded onto disk drive media become increasingly important. Therefore, this dissertation presents developments to the uniquely suitable spin-stand-based method of imaging magnetization patterns on media extracted from commercial hard disk drives. The emphasis of the presented research is placed on the following three areas: microscopy enhancement techniques for longitudinal magnetic recording media, “drive-independent” characterization and reconstruction of disk data, and the exploration of spin-stand microscopy in the novel context of perpendicular magnetic recording.

First, it is known that, while the spin-stand microscopy technique offers high-speed and massive scale imaging capabilities, the images obtained are corrupted by distortion due to the non-local sensing or finite spatial resolution of the imaging

sensor. Two techniques for mitigating this distortion, one based on characterizing the head by means of its linear response function, and a new method based on spatial Hilbert transforms, are described and demonstrated. Furthermore, a two-dimensional extension of the Hilbert transform in the context of magnetic recording is derived based on physical arguments and its application to spin-stand imaging is demonstrated. Second, although magnetic media imaging is interesting in its own right, an extension of this capability is the identification of commercial hard disk drive write channels and the subsequent reconstruction of the data written to the associated disks in a “drive-independent” manner on the spin-stand. For fundamental and practical reasons, a multilayered encoding process is performed on digital data before it is written to the disk; the presented work details the theoretical and experimental results obtained in characterizing and reversing these codes. Finally, because perpendicular recording technology has recently come on the market in consumer disk drives, the spin-stand microscopy technique is extended to imaging the media employing this new mode of recording. In particular, the novel aspects of perpendicular recording are discussed and their impact on spin-stand microscopy is demonstrated.

PROCESSING OF EX-SITU ACQUIRED SIGNALS
FROM MAGNETIC DISKS

by

Patrick McAvoy

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2008

Advisory Committee:
Professor Isaak D. Mayergoyz, Chair/Advisor
Professor Thomas E. Murphy
Professor Adrian Papamarcou
Professor Romel D. Gomez
Professor Howard C. Elman

© Copyright by
Patrick McAvoy
2008

ACKNOWLEDGMENTS

In writing the section on ECC, I am indebted to Mr. Sergiy Tkachuk for his work on this topic. Dr. Chun-Yang Tseng's work on positioning recording heads and centering disk platters on the spin-stand was vital to much of the experimental work in the present dissertation. Collaboration with Dr. Chun Tse, his previous work, and his tutorials on spin-stand operation were indispensable. Dr. Charles Krafft contributed a wide range of insight and practical support at the Laboratory for Physical Sciences, at which the experiments described in this dissertation were performed. I also extend my thanks to the advisory committee for their comments and suggestions.

Finally, I am the most highly indebted to Professor Isaak D. Mayergoyz, under whose direction this dissertation's research was performed, for his immense breadth and depth of knowledge, creativity, and guidance.

Table of Contents

List of Figures	v
List of Abbreviations	x
1 Introduction	1
1.1 Overview	1
1.2 The Spin-stand	2
1.3 Outline	4
2 Spin-stand Magnetic Microscopy Techniques for Longitudinal Recording	6
2.1 Overview of Longitudinal Magnetic Recording	6
2.2 Convolution Model of the Read Process	10
2.2.1 The Head Response Function	10
2.2.2 Solution Techniques	18
2.2.3 Experimental Results	24
2.3 Conclusions	32
3 The Hilbert Transform in Magnetic Recording	33
3.1 The 1D Hilbert Transform	33
3.1.1 Theory	33
3.1.2 Experimental Results	37
3.2 The 2D Hilbert Transform	39
3.2.1 Derivation and Theory	39
3.2.2 Experimental Results	48
3.3 Conclusions	53
4 Identification of the Hard Disk Drive Write Channel and Digital Data Re- construction	54
4.1 Overview	54
4.1.1 Data Organization on the Disk	56
4.1.2 Detection	57
4.2 Run-Length-Limited (RLL) Coding	67
4.2.1 Background	67
4.2.2 Identification Techniques	69
4.3 Scrambling	73
4.3.1 Background	73
4.3.2 Identification Techniques	78
4.4 Experimental Results	81
4.5 Conclusions	83

5	Error Correcting Code (ECC) Identification	84
5.1	Overview	84
5.2	Identification Techniques	90
5.3	Experimental Results	93
5.4	Conclusions	95
6	Spin-stand Magnetic Microscopy Techniques for Perpendicular Recording	96
6.1	Background and Overview	96
6.2	Response Function Characterization for the Read Head in Perpendicular Recording	102
6.2.1	Background	103
6.2.2	Definition of the Response Function	106
6.2.3	Measurement of the Response Function	110
6.2.4	Experimental Results	116
6.3	ISI Removal in Perpendicular Recording	117
6.4	Perpendicular Recording Image Characteristics	120
6.4.1	DC values and AC Erasure	120
6.4.2	The Hilbert Transform in Perpendicular Recording	128
6.5	Conclusions	130
7	Conclusions and Future Directions	132
7.1	Summary and Conclusions	132
7.2	Future Directions	136
7.2.1	DC Response	136
7.2.2	Write circuit and ISI experiments	137
7.2.3	Coding schemes	138
	Bibliography	139

List of Figures

1.1	Illustration of the along-track and across-track scanning directions on a hard disk platter, with the motion of the head above the disk denoted by a dashed line.	3
2.1	Longitudinal recording write head.	7
2.2	Recording plane with magnetization \vec{M} , magnetic charge σ_m (small circles), and the external magnetic field \vec{H} above the medium.	9
2.3	Simulation of ISI for the tribit pattern, with T fixed and PW_{50} of the head response function varying, yielding different densities D	12
2.4	DC erasure trimming to produce “spot” of magnetic charge described in [1, 5].	13
2.5	Spin-stand images of raw and averaged untrimmed isolated transitions written by $0.39 \mu m$ write head.	14
2.6	Spin-stand images of raw and averaged trimmed isolated transitions written by $0.39 \mu m$ write head.	15
2.7	Spin-stand image of measured 2D head response function for a $0.39 \mu m$ write head.	15
2.8	Numerical Simulation of Head Response Measurement: Cases A (top) and B (bottom) (described in the text).	16
2.9	Numerical Simulation of Head Response Measurement: Cases C (top) and D (bottom) (described in the text).	17
2.10	ISI removed from a periodic tribit pattern, with good parameter choice ($\lambda = 10^{-3}$, $\alpha = 10^4$).	25
2.11	Periodic tribit readback voltage and its “denoised” reconstruction.	26
2.12	ISI removed from a periodic tribit pattern, λ too small ($\lambda = 10^{-5}$).	26
2.13	ISI removed from a periodic tribit pattern, λ too large ($\lambda = 10$). The recovered charge locations in this case are actually further apart than the original readback peaks with ISI.	27
2.14	Periodic Tribit readback voltage and its “denoised” reconstruction with λ too large ($\lambda = 10$).	28

2.15	ISI Removed from a periodic tribit pattern, with the head response too wide (top) and too narrow (bottom). Here, the too wide response is $2PW_{50}$ and the too narrow is $PW_{50}/4$ where PW_{50} is the width of the response used in figure 2.10.	29
2.16	ISI Removed from a periodic tribit pattern, with α too small. The top plot of M_x shows how the regions of zero charge are corrupted by noise. The transition peaks are located at the zero crossings of the magnetization.	29
2.17	Image taken from a commercial hard disk from 2002, before and after ISI removal.	30
2.18	Along-track signal taken from a commercial hard disk from 2002, before and after ISI removal.	30
2.19	Image taken a commercial hard disk from 2003, before and after ISI removal.	31
2.20	Along-track signal taken from a commercial hard disk from 2003, before and after ISI removal.	31
3.1	Magnitude as a function of spatial frequency for the “filter” represented by Hilbert transformation and subsequent differentiation . . .	37
3.2	ISI Removed from a periodic “F9” (11111001) pattern using the 1D Hilbert Transform method with $\alpha = 10^4$	38
3.3	ISI Removed from a periodic “F9” (11111001) pattern using the 1D Hilbert Transform method with $\alpha = 10^4$, showing false transitions. . .	39
3.4	ISI Removed from a periodic “F9” (11111001) pattern using the 1D Hilbert Transform method with insufficient saturation, $\alpha = 1$	40
3.5	ISI Removed from a year 2000 commercial disk ata region using the 1D Hilbert Transform method, $\alpha = 2000$	40
3.6	The 3D schematic with media plane and the magnetic fields emanating from it.	42
3.7	The surface magnetic charge representation of the media with the transverse (normal) components of the resultant magnetic field. . . .	43
3.8	The surface virtual electric current representation of the media with the longitudinal (tangential) components of the resultant magnetic field.	44

3.9	The simulated readback voltage image of $H_z(x, y)$ overlaid with its reconstructed vectorial tangential field (top); the reconstructed across-track component $H_y(x, y)$ (middle); and the reconstructed along-track component $H_x(x, y)$ (bottom).	49
3.10	The simulated readback voltage image of $H_z(x, y)$ (top), and the 2D Hilbert transform reconstructed charge $\sigma_m(x, y)$ without arctangent saturation (bottom).	50
3.11	The simulated readback voltage image of $H_z(x, y)$ (top), and the 2D Hilbert transform reconstructed charge $\sigma_m(x, y)$ with arctangent saturation ($\alpha = 5 \cdot 10^{-2}$) (bottom).	51
3.12	The readback voltage image of $H_z(x, y)$ (top) from a commercial disk with 60 kilotracks per inch, and the 2D Hilbert transform reconstructed charge $\sigma_m(x, y)$ with arctangent saturation ($\alpha = 15 \cdot 10^{-2}$) (bottom).	51
4.1	The format of the data sector.	57
4.2	An example of NRZ and NRZI representations of bits.	59
4.3	Block diagram of the two-layered coding process in write (top) and read (bottom) modes.	65
4.4	The effects of a two-layered coding scheme on a part of a sector from a JPEG file.	66
4.5	The effects of the same two-layered coding scheme as figure 4.4 on a constant all “0” bit sequence.	66
4.6	RLL 8/9 (0, 4/4) histograms of the zero run lengths for the total encoded sequence (top), odd interleave (middle), and the even interleave (bottom).	71
4.7	RLL 8/9 (0, 3/6) histograms of the zero run lengths for the total encoded sequence (top), odd interleave (middle), and the even interleave (bottom).	71
4.8	Autocorrelation of a pseudo-noise sequence produced by a 10th order scrambler.	76
4.9	Block diagram of the two-layered scrambling process in write (top) and read (bottom) modes.	78

5.1	Interleaving of ECC codewords, the example here shown with three interleaves.	90
6.1	Sketch of the longitudinal recording media, showing the intensification of the in-plane (horizontal) component of the external magnetic field as the longitudinal magnetization charge spatial density increases.	98
6.2	The recording plane (x - z) with perpendicular magnetization	99
6.3	Perpendicular recording write head and medium design with soft underlayer.	100
6.4	Perpendicular recording write head employing the side shielding design described in [64].	101
6.5	(left) Perpendicular recording media with soft underlayer and (right) its effect on the charge distribution.	105
6.6	(a) The unit step, and its (b), (c) superposition and translation to produce (d) a small region of charge.	108
6.7	Charge distributions for the measurement of the impulse response function on a perpendicular medium via superposition, with blue denoting negative charge regions, red denoting positive charge, and (c) showing a double positive charge.	111
6.8	Numerical adjustment for the DC filtering effect of the spin-stand preamplifier on a spin-stand image obtained of various low-frequency perpendicular disk tracks.	115
6.9	Measurement of 1D step response on a perpendicular medium with its corresponding equivalent 1D impulse response.	117
6.10	Measurement of 2D step response on a perpendicular medium, with the numerical DC-level adjustment described in equation 6.20 applied.	118
6.11	Simulation of perpendicular and longitudinal readback voltages of the same bit pattern with their spectra, highlighting the DC content in perpendicular readback and lack thereof in longitudinal recording readback.	122
6.12	Spin-stand image of an AC-erased perpendicular medium.	125
6.13	Spin-stand Image of F6 overwritten on F9 on an AC-erased perpendicular medium.	127

6.14 Spin-stand Image of F6 overwritten on F9 on an AC-erased perpendicular medium.	127
6.15 Spin-stand Image of Tracks written on a DC-erased perpendicular medium.	129
6.16 Readback waveform at 0.9 μm across-track distance in figure 6.15 of a track written on a DC-erased perpendicular medium.	129
6.17 The spin-stand image of low-frequency patterns on perpendicular media (top) with their H_x (middle) and H_y (bottom) components computed via the 2D Hilbert transform.	131
6.18 A detail of figure 6.17 with the readback at 1.7 μm across-track distance.	131

List of Abbreviations

σ_m	magnetic charge
PW_{50}	Pulse full width at half maximum
BER	Bit Error Rate
GMR	Giant Magnetoresistive
HDD	Hard Disk Drive
ISI	Intersymbol Interference
MFM	Magnetic Force Microscopy
MR	Magnetoresistive
NLTS	Nonlinear Transition Shift
PRML	Partial Response Maximum Likelihood
RLL	Run-Length Limited
TMR	Tunneling Magnetoresistive

List of Publications

- 1.) I.D. Mayergoyz, C. Tse, C. Krafft, and P. McAvoy, "A novel approach to removing intersymbol interference from spin-stand images," *IEEE Trans. Magn.*, **40**, 2197 (2004).
- 2.) C. Tseng, I.D. Mayergoyz, C. Tse, P. McAvoy, and C. Krafft, "Dynamic track-following of off-centered hard disks in spin-stand imaging," *J. Appl. Phys.*, **97**, 10R301 (2005).
- 3.) C. Tse, C. Tseng, P. McAvoy, C. Krafft, and I.D. Mayergoyz, "Whole-track imaging and diagnostics of hard disk data using the spin-stand imaging technique," *J. Appl. Phys.*, **97**, 10P104 (2005).
- 4.) C. Tse, C. Krafft, I. Mayergoyz, P. McAvoy, and C. Y. Tseng, "Forensic recovery of hard disk data by using the spin-stand imaging technique," *Proc. SPIE*, vol. 5778, pp. 595-606, (May 2005).
- 5.) I.D. Mayergoyz, P. McAvoy, C. Tse, C. Krafft, and C. Tseng, "The 2-D Hilbert Transform in Magnetic Recording," *IEEE Trans. Magn.*, **42**, 2282 (2006).
- 6.) C. Tseng, I. Mayergoyz, P. McAvoy, and C. Krafft, "Iterative compensation for hysteresis effects in positioning and tracking problems," *J. Appl. Phys.*, **103**, 07D902 (2008)
- 7.) P. McAvoy, C. Tseng, I. D. Mayergoyz, and C. Krafft, "Spin-stand imaging of perpendicularly recorded data," *IEEE Trans Magn.*, to appear November, 2008.

Chapter 1

Introduction

1.1 Overview

It has been previously demonstrated that a spin-stand can be used to image the recorded magnetization patterns on media extracted from hard disk drives [1]-[7]. This capability is valuable for its diagnostic utility in enabling the analysis of the phenomena actually occurring on the disk media in a more direct and accessible manner than is afforded by simple (one-dimensional) readback waveform analysis or magnetic force microscopy (MFM). In the latter technique, truly large-scale imaging is impractical due to the nature of the MFM scanning process, whereas the spin-stand technique by its very nature incorporates rapid imaging capability to enable meaningfully large scanning areas. Magnetic recording has been the dominant means for large-scale, random access, non-volatile digital data storage for decades and will remain so for the foreseeable future due to its technical maturity, low cost, and continual innovation in storage density. Most recently, perpendicular recording products have entered the consumer market, and these mark a new direction for the magnetic storage industry. Therefore, this dissertation presents developments to the spin-stand imaging technique in both the traditional longitudinal and novel perpendicular modes of magnetic recording.

1.2 The Spin-stand

The principal piece of equipment used to perform the experimental research in this dissertation is the magnetic recording spin-stand. The spin-stand is a device used in the magnetic recording industry for the testing and evaluation of magnetic read and write heads, recording media, and channel designs. As such, it possesses a construction which enables it to emulate several aspects of hard disk functionality. First, the spin-stand employs a high-quality air-bearing spindle motor on which hard disk platters can be mounted and programmed to stably rotate at any angular velocity typical for hard disk operation (e.g., 3600RPM to 10000RPM). Second, the spin-stand can accept various magnetic recording heads mounted to a positioning actuator, which places the heads in proximity to the disk for the reading and writing of magnetization patterns. This positioning actuator can be programmed to make displacements in the radial direction of the mounted disk on the order of nanometers. The particular model of spin-stand used in the presented research is the Guzik 1701MP.

It was first suggested in [2] to use these two features of the spin-stand's design for two-dimensional imaging of magnetization patterns recorded on hard disk platters. By virtue of its construction, the spin-stand exhibits the ability to move an imaging sensor (read head) in two orthogonal directions above and with respect to a flat target (a hard disk platter; see figure 1.1). One of these directions is "along-track," or in the direction defined by circular movement about the center of the disk. Scanning in this direction is effected by the rotation of the disk on the spindle

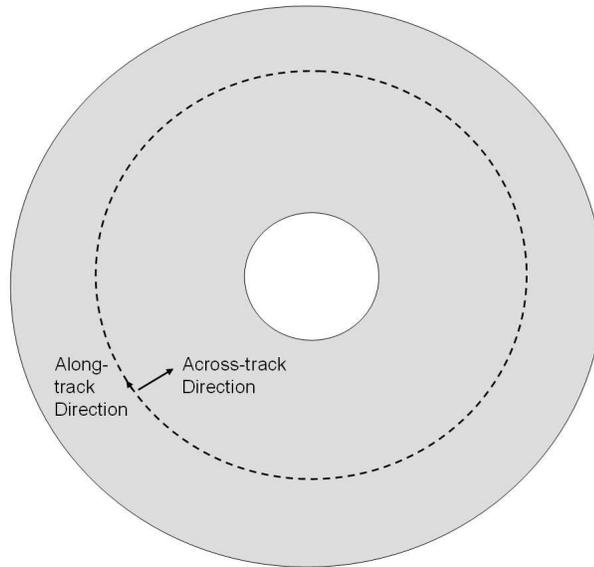


Figure 1.1: Illustration of the along-track and across-track scanning directions on a hard disk platter, with the motion of the head above the disk denoted by a dashed line.

with respect to the read head. The second scanning direction (perpendicular to the first) is the “across-track” direction, defined by the radial displacement from the center of the disk, and achieved by the positioning actuator to which the read/write head is attached. Thus, it is seen that the relative motion of sensor and target is made in two different ways. In the along-track direction, the disk moves, while in the across-track direction, the head moves. Putting these two scanning directions together, it is readily apparent that the entire disk can be scanned by progressively stepping the head in the across-track direction while at each step acquiring the read sensor signal for one disk revolution.

The chief advantages of this imaging set-up are speed of acquisition and adaptability to different imaging requirements (e.g., different disks). Because the spin-

stand “spins” the disk, scanning in the along-track direction is rapid (proportional to the speed of rotation). The mechanism for this speed of acquisition derives from the fact that the read head acts as a transducer of spatial magnetic flux to temporal voltage. Therefore, a digitizing oscilloscope or acquisition board with sufficient bandwidth and sampling capabilities suffices for acquiring image data for a given across-track position. Additionally, the ability to mount different heads and adjust parameters like skew angle (the in-plane angle of the head with respect to the track) allows different disks to be imaged. Due to considerations both mechanical (disk surface roughness, thickness) and electromagnetic (external magnetic field strength, areal density), properly matching the read/write heads and the disk media to be imaged is necessary. Finally, the high rate of image acquisition coupled with developed techniques for accurate repeatability of positioning (e.g., [9], [10], [12]) allows for temporal averaging. This averaging capability boosts the image quality by eliminating zero-mean noise. Overall, these abilities contrast favorably with those of slower, more traditional microscopy techniques like MFM, which typically raster scan over small, rectangular regions.

1.3 Outline

This dissertation consists of six topics, each forming a separate chapter. The prior work has shown that the images obtained via the spin-stand technique are distorted by the finite spatial resolution (local averaging) of the imaging sensor (a shielded magnetoresistive readback head), which creates intersymbol interference

(ISI) [8]. In chapter 2, the present work investigates the prior developed technique of head response function deconvolution, especially in terms of its numerical implementation. Additionally, in chapter 3, a new technique is described that involves the use of the Hilbert transform for one-dimensional ISI removal. This Hilbert transform technique is then expanded in scope and derived from physical arguments in a multi-dimensional vector form [11], whose utility is then demonstrated for two-dimensional images.

Next, although magnetic media imaging is interesting in its own right, a natural extension of this capability is in the identification of commercial hard disk drive write channels and the subsequent reconstruction of the digital data written to the disks in a “drive-independent” manner. This topic is explored in detail in chapters 4 and 5. For fundamental and practical reasons that will be discussed in these chapters, a multilayered process of encoding is performed on digital data before it is written to disk, and this present work details the theory and experimental results obtained in identifying and inverting these coding schemes (run-length-limited encoding, scrambling, and error correction).

Subsequently, the timely subject of perpendicular recording, having recently come to the market in consumer storage products, is explored in the context of extensions to the spin-stand imaging techniques in chapter 6. Finally, a summary with conclusions is made and a discussion of areas for future research is offered in chapter 7.

Chapter 2

Spin-stand Magnetic Microscopy Techniques for Longitudinal Recording

Building upon the prior work in [1], this chapter describes and demonstrates the processing of spin-stand images of longitudinal recording media. This work illustrates the utility of the spin-stand imaging technique for diagnostic and forensic analysis of the recording media present in the most commonly encountered hard disk drives.

2.1 Overview of Longitudinal Magnetic Recording

The historically dominant form of nonvolatile random-access digital data storage is longitudinal magnetic recording in hard disk drives [14]. In this technique, digital bits are stored as alternating regions of remanent magnetization oriented in the plane of a thin-film, hard ferromagnetic medium. These regions of magnetization are written in circular tracks onto a disk on whose surface the thin-film storage medium lies, and a host of electromagnetic and electromechanical machinery operates to access and process them. The present work will focus on aspects of the electromagnetic components, and this section will give a simplified explanation of writing and reading the magnetization in order to introduce the necessary terminology for this chapter.

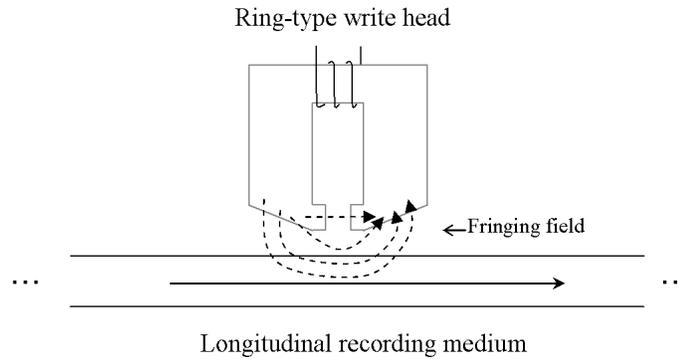


Figure 2.1: Longitudinal recording write head.

In order to write bits, a “write head” produces a magnetic field proportional to a controlling “write current” while moving above the storage medium at a nominally constant velocity and fixed vertical displacement (“flying height”). The fringing part of this write field impinges upon the medium and interacts with it, as shown in figure 2.1. If the write field magnitude is higher than the medium coercivity H_c (the magnetic field necessary to switch the magnetization direction) [14], a region of uniform magnetization is left behind with orientation parallel to the medium. The direction of this magnetization along the line of motion depends on the polarity of the square-wave write current, thus effecting binary recording. As the polarity of the write current is switched in time (or, proportionally, position, since velocity is constant), so is the direction of the write field and thus the remanent longitudinal magnetization that follows it. In this sense, the write head is an electrical current to magnetic field transducer and forms the “transmitter” of the storage communication system.

The “receiver” of the system is the read head. Historically, this was an induc-

tive device, which by Faraday's law senses the time derivative of the longitudinal magnetic flux emanating from the recording medium. Much higher performance magnetoresistive (MR) sensors became dominant a decade ago in consumer hard disk products and their continual improvement has been a principle force behind the tremendous rise in storage densities since their introduction. The principle subsequent developments in commercial magnetoresistive sensors for magnetic storage involve the giant and tunnelling magnetoresistive (GMR and TMR, respectively) effects [15, 16]. Depending on their design, these read heads exploit a magnetoresistive effect to act as transducers of magnetic field to electrical resistivity and thus voltage. In order to maximize their spatial resolution, MR sensors are shielded by high magnetic permeability structures, which guide flux that does not originate from directly under the head away from the sensor. By this means, the spacing between the shields and the distance between the head and the medium determines the degree of local sensing. Therefore, in the ideal limit of both zero shield gap and head-medium spacing, the shielded MR sensor directly senses only the *vertical* component of the external magnetic field above the medium that arises from the recorded magnetization.

The magnetic charge formulation of the fields emanating from the magnetic media is useful in understanding the read process. In figure 2.2, an idealized schematic of the infinitely thin recording medium is shown in two dimensions called the *recording plane* x - z , which assumes no variation in the cross track (y) direction. In general, on recording medium of uniform thickness h , the magnetic charge is

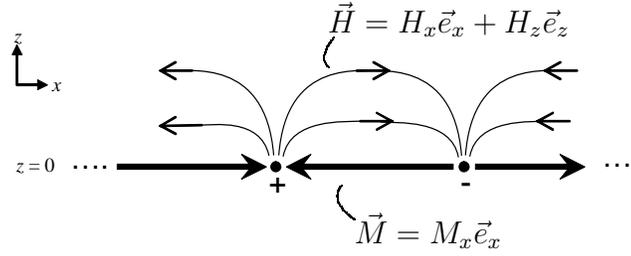


Figure 2.2: Recording plane with magnetization \vec{M} , magnetic charge σ_m (small circles), and the external magnetic field \vec{H} above the medium.

defined as

$$\sigma_m = -\mu_0 h \nabla \cdot \vec{M}. \quad (2.1)$$

Therefore, in this 1D context, it is

$$\sigma_m(x) = -\mu_0 h \frac{dM_x(x)}{dx}, \quad (2.2)$$

as depicted in figure 2.2. In this context, charge marks a transition in magnetization. Furthermore, the figure illustrates qualitatively that the magnitude and sign of the vertical component of the external magnetic field, H_z , resembles the charge. Therefore, because the MR sensor picks up H_z , when it is brought in proximity with the recording medium and scanned along the x direction, the read head senses the transitions in magnetization.

2.2 Convolution Model of the Read Process

2.2.1 The Head Response Function

Most generally, the readback voltage obtained from scanning hard disk media is the product of a nonlinear system that comprises the write and read processes. The magnetoresistive (MR) sensor itself (“the read head”) together with the nature of the magnetic fields that it senses from the recorded bits both contribute to this fact. First, the read process in magnetic recording is, strictly defined, inherently nonlinear. The very physical nature of the magnetoresistive effect is nonlinear, with the resistance (and thus voltage) varying according to $\cos^2(\theta)$ (anisotropic MR, or AMR) or $\sin(\theta)$ (giant MR, or GMR), where θ is the angle between the magnetization of the MR element and the bias current [17]. Second, the write process is inherently nonlinear by virtue of the hysteretic relationship of magnetization versus applied field. Nonlinear transition shift (NLTS) occurs in the write process if, (in the simplest case) two transitions are too closely spaced. This causes the external magnetic field of the first written transition to interfere with the writing of the second transition, which thereby shifts in location. Further nonlinear effects are seen during the write process due to overwriting previous magnetization with that of the opposite direction (hard shift and partial erasure) [18].

Nevertheless, in practice, sufficient care is taken both while writing and reading such that a linear approximation holds. To compensate for the head nonlinearity, a proper MR bias current is chosen to bias the sensor in the linear part of its voltage-magnetic field response curve. When restricted to a small variation about

this bias point (i.e., sensed magnetic fields small enough), the read head performs an approximately linear measurement. To compensate for the predictable effects of NLTS, so-called write current “precompensation” is applied, which pre-shifts the transitions so that they will be written approximately linearly. (Hard shift and partial erasure cannot be compensated, but this can be neglected and the effects considered as part of the “media noise.”)

It is useful, therefore, to consider the regime in which the readback image voltage V produced by the spin-stand imaging technique is the output of a linear time-invariant system whose impulse response is defined as the head response function [1, 4, 5, 7, 8]. The read head, in reality, does not sense just the normal field directly beneath it, but a small region that contributes to the sensed voltage. By the principles of linear superposition and translation invariance, the image voltage is proportional to the two-dimensional convolution of the surface magnetic charge σ_m with the head response kernel R :

$$V(x, y) = \int_S \sigma_m(x', y') R(x - x', y - y') dx' dy'. \quad (2.3)$$

Or equivalently, the voltage is a mapping of the charge by operator $\mathcal{R}\{\}$, the convolution with $R(x, y)$ on the domain:

$$V(x, y) = \mathcal{R}\{\sigma_m(x, y)\} \quad (2.4)$$

By this model, any readback voltage is equal to the linear superposition of the media surface magnetic charge under the local averaging effect of the head. Note that, obviously, a similar expression holds for the one-dimensional case by setting

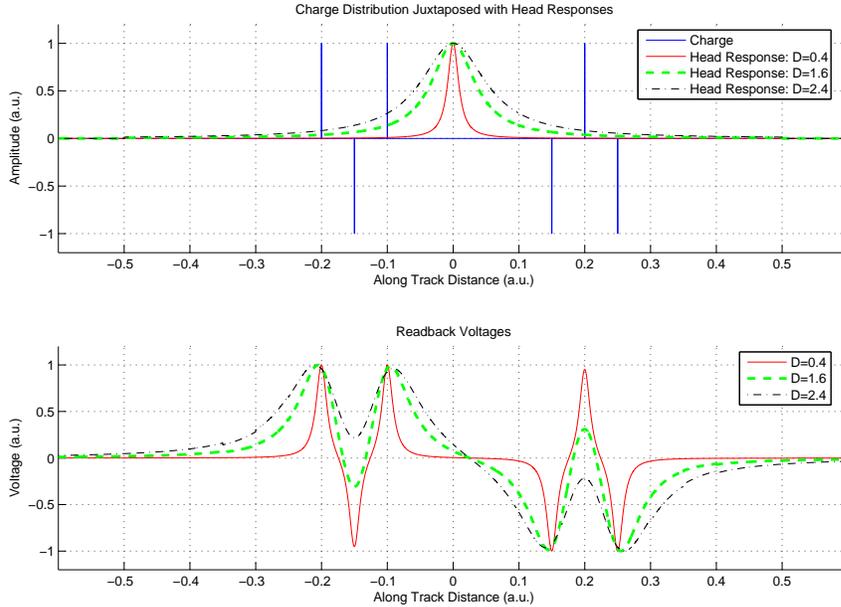


Figure 2.3: Simulation of ISI for the tribit pattern, with T fixed and PW_{50} of the head response function varying, yielding different densities D .

the y variations in equation 2.3 to zero; i.e.,

$$V(x) = \int_L \sigma_m(x') R(x - x') dx'. \quad (2.5)$$

This linear model directly characterizes and quantifies the dominant form of distortion in spin-stand imaging and hard disk drive readback voltage, intersymbol interference (ISI). ISI results when the head response function overlaps two or more adjacent transitions at the same time, causing destructive interference, and is a function of the density $D = \frac{PW_{50}}{T}$ (where PW_{50} denotes the full width at half maximum of the response function and T denotes the bit period). Figure 2.3 illustrates in one dimension this effect on a pair of so-called “tribit” patterns of three adjacent transitions. Neighboring transitions interfere destructively due to the fact that they

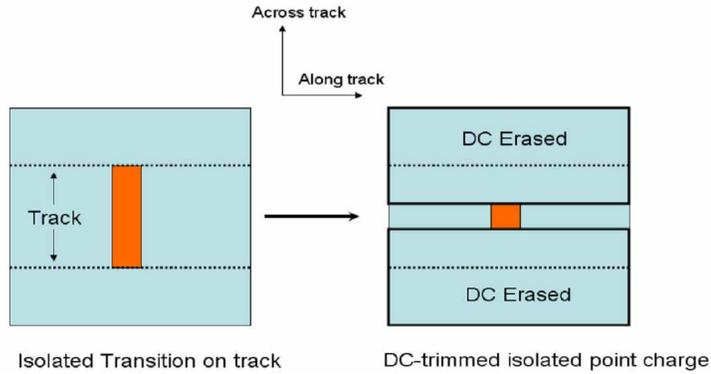


Figure 2.4: DC erasure trimming to produce “spot” of magnetic charge described in [1, 5].

are always opposite in polarity. The effect of ISI is twofold. First, readback pulses are broadened and reduced in magnitude, and second, the local peaks are shifted with respect to their true positions.

Due to the importance of the head response function in determining the image characteristics, its measurement is important. Nominally, to obtain the head response function requires writing a 2D (or 1D) delta function to the disk media and reading it with the MR head; the read voltage then is thus the head response kernel. Prior work [1, 5] has shown that an approximation to an impulse can be obtained by DC erasing the vicinity of an isolated transition and taking the simple average of successive scans to remove zero-mean noise. Figure 2.4 illustrates the general process of DC erase trimming to produce a spot of magnetic charge. An example of this experimental procedure is shown in figures 2.5 and 2.6. In the top of figure 2.5, two isolated transitions in a train of alternating polarity isolated transitions are depicted in this spin-stand image. These were written by the spin-stand on an

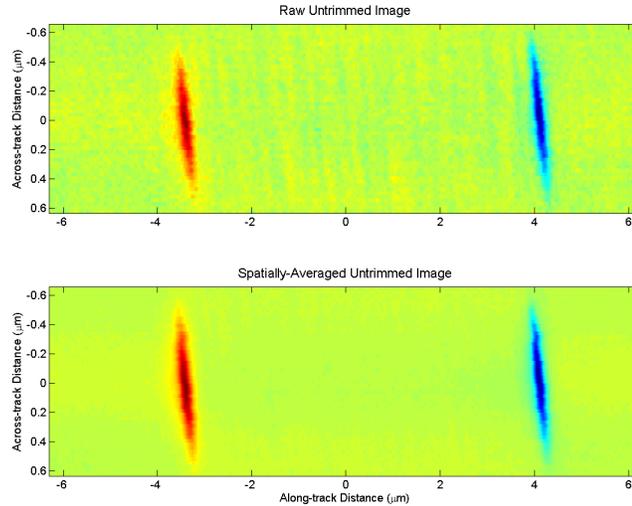


Figure 2.5: Spin-stand images of raw and averaged untrimmed isolated transitions written by $0.39 \mu\text{m}$ write head.

already DC-erased part of the disk medium by a $0.39 \mu\text{m}$ write head. In the bottom of figure 2.5, the result of averaging in space (i.e., averaging among the train of pairs of transitions) is depicted. In figure 2.6, the spin-stand images of the resulting DC-trimmed isolated transitions is shown. In the top of figure 2.6, the raw image of the isolated charge appears dominated by noise due to fact that the significantly trimmed transitions are reduced in magnitude compared to their untrimmed counterparts. When temporal averaging (middle plot) and spatial averaging (bottom plot) are performed, the zero-mean media and electronic noise sources are removed, giving a measurement of the head response function. Figure 2.7 shows the final measured head response function from this particular experiment.

In practice, true impulses cannot be written. Nevertheless, this fact does not matter, as illustrated by the numerical simulation results in figures 2.8 and 2.9, which

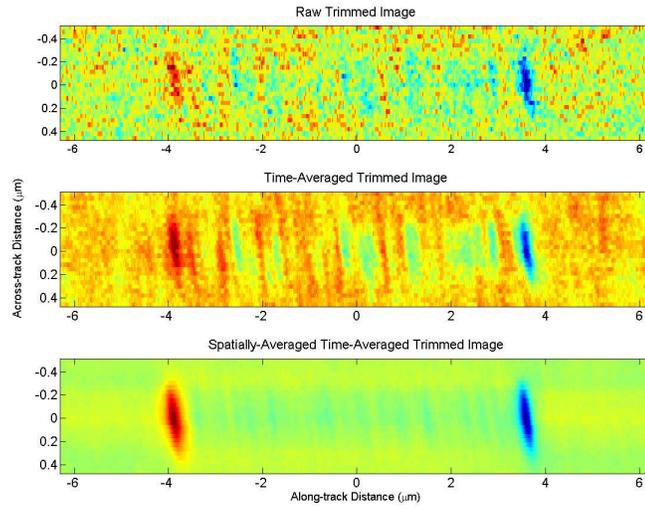


Figure 2.6: Spin-stand images of raw and averaged trimmed isolated transitions written by $0.39 \mu m$ write head.

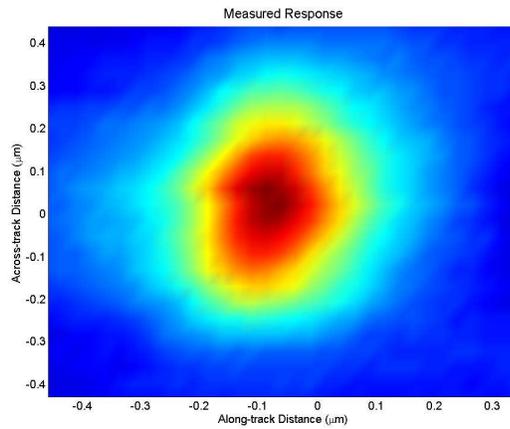


Figure 2.7: Spin-stand image of measured 2D head response function for a $0.39 \mu m$ write head.

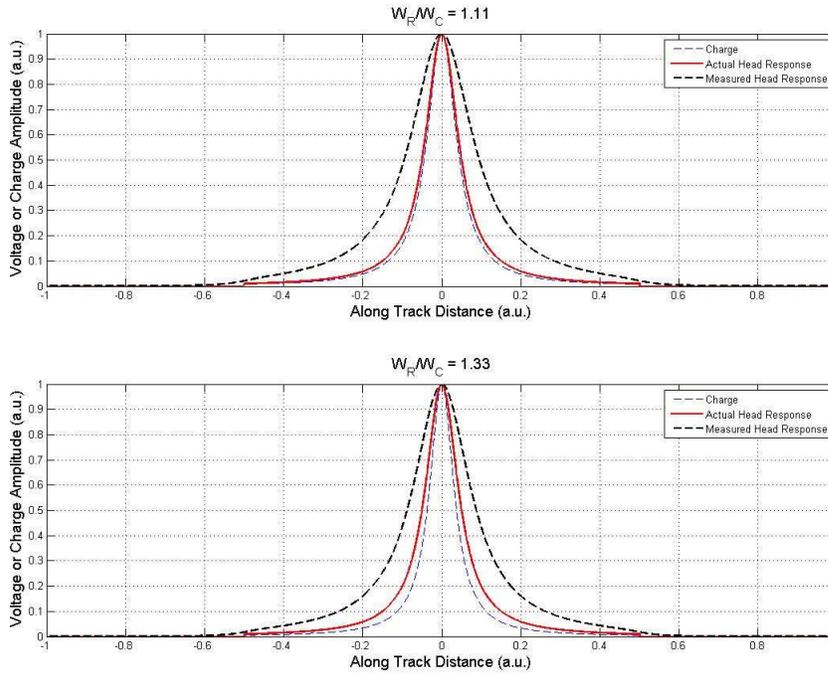


Figure 2.8: Numerical Simulation of Head Response Measurement: Cases A (top) and B (bottom) (described in the text).

show the convolution of a fixed head response with varying charge profiles. The plots show the charge distribution (blue dashed line), the actual head response function (red solid line) and the resulting readback voltage (the “measured” response, black dashed line). These four one-dimensional cases show that as W_R , the response PW_{50} increases with respect to W_C , the charge PW_{50} , the “measured” response approaches the “actual” head response, as expected. In case A, $W_R/W_C = 1.11$; in case B, $W_R/W_C = 1.33$; in case C, $W_R/W_C = 2$; and in case D, $W_R/W_C = 5$.

These simulations assume an arctangent transition in magnetization, a typical assumption, which therefore gives a Lorentzian shaped readback pulse. Interestingly,

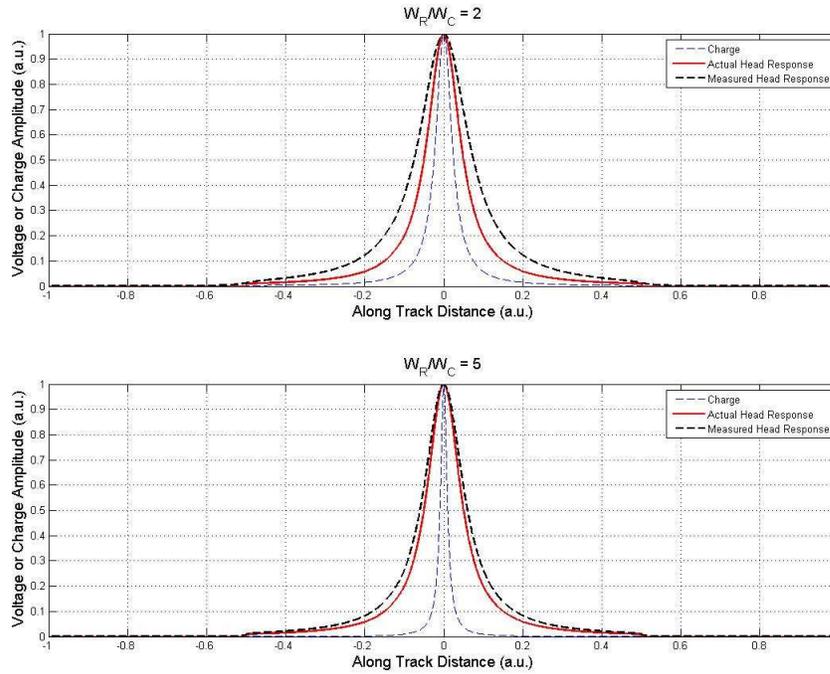


Figure 2.9: Numerical Simulation of Head Response Measurement: Cases C (top) and D (bottom) (described in the text).

to measurably approximate a delta function with this pulse shape does not require a very small relative width; in fact, a ratio under an order of magnitude is sufficient, as seen in case D. However, even if a large enough ratio W_R/W_C cannot be achieved in practice, a useful head response may still be obtained by means of numerical scaling [7]. Indeed, in practice, it is useful to numerically scale the head response function width as desired for different performance effects in the presence of noise.

2.2.2 Solution Techniques

Having the head response function characterized allows its unwanted effect on the readback image, ISI, to be remedied. To remove ISI is to solve for the charge in the integral equation 2.3, given a known response function R and readback voltage V . That is,

$$\sigma_m(x, y) = \mathcal{R}^{-1}\{V(x, y)\} \quad (2.6)$$

Solving for the charge in equation 2.3 is, however, an instance of a Fredholm integral equation of the first kind, which is an ill-posed problem [19]. It is well known that in such integral equations of the first kind a small perturbation in V will lead to a large perturbation in σ_m , implying instability. One way of seeing this fact is by looking at the 2D Fourier transform representation of equation 2.3,

$$\tilde{V}(k_x, k_y) = \tilde{\sigma}_m(k_x, k_y)\tilde{R}(k_x, k_y), \quad (2.7)$$

which gives the following solution in spatial frequency domain:

$$\tilde{\sigma}_m(k_x, k_y) = \tilde{V}(k_x, k_y)/\tilde{R}(k_x, k_y). \quad (2.8)$$

Obviously, if $\tilde{R}(k_x, k_y)$ is zero anywhere within the support of $\tilde{V}(k_x, k_y)$, a singularity in spatial frequency appears in the charge $\tilde{\sigma}_m(k_x, k_y)$ and the method is not stable. In effect, the singularities represent frequency content that has been “lost,” which gives rise to the ill-posed nature of the problem. This situation is indeed what happens in the context of imaging in magnetic recording. This fact is due to the nature of the head response function, which, as it describes the non-local sensing properties of the head, blurs or low-pass filters the charge distribution. Inversion,

therefore, requires high-pass filtering, and the inherent instability of the process immediately implies that attempting to exactly solve this problem is not desirable. This situation is well known in various contexts. An equivalent viewpoint is to use a least-squares method of solution on equation 2.3. In the discrete form of the problem, this corresponds to replacing \mathcal{R}^{-1} with the corresponding pseudoinverse \mathcal{R}^\dagger (e.g., by singular value decomposition) [20]. Here, however, the conditioning of the operator/matrix \mathcal{R} is generally so poor that the least-squares solution is also totally corrupted by unwanted large variations.

Thus, a method of *regularizing* — “correcting” the ill-posed nature and improving the conditioning — of the problem is necessary. There are numerous methods of regularization, with Tikhonov regularization being a common and simple one [19]. In this technique (here applied to equation 2.3), the related inexact problem of finding the approximate charge distribution σ'_m such that

$$\mathcal{A} + \lambda\mathcal{B} = \|\mathcal{R}\{\sigma'_m(x, y)\} - V(x, y)\|^2 + \lambda \|\sigma'_m(x, y)\|^2 \quad (2.9)$$

is minimized, given a regularization constant λ [19]. In this notation, the functional \mathcal{A} represents the accuracy of the approximate solution σ'_m with respect to the ideal solution σ_m , whereas \mathcal{B} represents the smoothing of large norm perturbations in the approximate solution. The particular formulation given here is known as zeroth-order regularization. In the general case, \mathcal{B} is some functional of $\|\sigma'_m\|^2$ that is determined by *a priori* information about the smoothness of the expected solution [19], and in general is not the identity functional, as it is here. The least squares

solution to this minimization problem yields, in spatial frequency domain,

$$\tilde{\sigma}'_m(k_x, k_y) = \frac{\tilde{V}(k_x, k_y)\tilde{R}^*(k_x, k_y)}{|\tilde{R}(k_x, k_y)|^2 + \lambda}, \quad (2.10)$$

with * denoting complex conjugation.

Experience has shown that the regularization of equation 2.3 alone as described in equation 2.10 is not sufficient in the spin-stand imaging environment for longitudinal recording, and the deconvolution process can be improved further still. Another means of making the solution of the deconvolution problem more tractable lies in exploiting the relationship between the magnetization on the media and its magnetic charge representation [8]. Recall that the charge is defined as being proportional to the negative divergence of the magnetization,

$$\sigma_m(x, y) = -\mu_0 h \nabla \cdot \vec{M} = -\mu_0 h \left(\frac{\partial M_x}{\partial x} + \frac{\partial M_y}{\partial y} \right). \quad (2.11)$$

Using the relationship between magnetization and charge allows equation 2.3 to be rewritten as a more stable integral-differential equation,

$$V(x, y) = \int_S -\mu_0 h \left(\frac{\partial M_x}{\partial x} + \frac{\partial M_y}{\partial y} \right) R(x - x', y - y') dx' dy'. \quad (2.12)$$

Additionally, it is known that only the curl-free component of the magnetization produces the field that can be sensed by the head [1]. Looking at the z -component of $\nabla \times \vec{M} = 0$ yields

$$\frac{\partial M_y}{\partial x} = \frac{\partial M_x}{\partial y}. \quad (2.13)$$

Taking Fourier transforms of equations 2.1 and 2.13 gives an algebraic system of two equations and two unknowns,

$$k_x \tilde{M}_x(k_x, k_y) + k_y \tilde{M}_y(k_x, k_y) = \frac{j \tilde{\sigma}_m(k_x, k_y)}{\mu_0 h}, \quad (2.14)$$

$$k_y \tilde{M}_x(k_x, k_y) = k_x \tilde{M}_y(k_x, k_y), \quad (2.15)$$

which yields the relations

$$\tilde{M}_x(k_x, k_y) = \frac{jk_x \tilde{\sigma}_m(k_x, k_y)}{\mu_0 h (k_x^2 + k_y^2)}, \quad (2.16)$$

$$\tilde{M}_y(k_x, k_y) = \frac{jk_y \tilde{\sigma}_m(k_x, k_y)}{\mu_0 h (k_x^2 + k_y^2)}. \quad (2.17)$$

In the 1D case, the relation between magnetization and charge becomes simply

$$\tilde{M}_x(k_x) = \frac{j \tilde{\sigma}_m(k_x)}{\mu_0 h k_x}. \quad (2.18)$$

By using equations 2.16 and 2.20 and substituting equation 2.8 for $\tilde{\sigma}_m$, a solution can be found that is more stable than that found by using 2.8 alone. However, in practice, this is usually not enough to yield a sufficiently smooth solution. Thus, zeroth order regularization can be applied by substituting equation 2.10 for $\tilde{\sigma}'_m$ in 2.16 and 2.20 in order to obtain the approximate magnetization solutions in spatial frequency domain:

$$\tilde{M}'_x(k_x, k_y) = \frac{jk_x \tilde{V}(k_x, k_y) \tilde{R}^*(k_x, k_y)}{\mu_0 h (k_x^2 + k_y^2) \left(|\tilde{R}(k_x, k_y)|^2 + \lambda \right)}, \quad (2.19)$$

$$\tilde{M}'_y(k_x, k_y) = \frac{jk_y \tilde{V}(k_x, k_y) \tilde{R}^*(k_x, k_y)}{\mu_0 h (k_x^2 + k_y^2) \left(|\tilde{R}(k_x, k_y)|^2 + \lambda \right)}. \quad (2.20)$$

To gain intuition into the effects of the regularization here it is useful to look at the one-dimensional version of the problem. In 1D, equation 2.19 becomes

$$\tilde{M}'_x(k_x) = \frac{j \tilde{V}(k_x) \tilde{R}^*(k_x)}{\mu_0 h k_x \left(|\tilde{R}(k_x)|^2 + \lambda \right)}. \quad (2.21)$$

After simple algebraic transformations, this can be rewritten as

$$\tilde{M}'_x(k_x) = \left(\frac{j \tilde{V}(k_x)}{\mu_0 h k_x \tilde{R}(k_x)} \right) \left(\frac{1}{1 + \frac{\lambda}{|\tilde{R}(k_x)|^2}} \right). \quad (2.22)$$

Recalling equations 2.8 and 2.18, this in turn can be rewritten in the form

$$\tilde{M}'_x(k_x) = \tilde{M}_x(k_x)H\{\tilde{R}(k_x)\}, \quad (2.23)$$

where

$$H\{\tilde{R}(k_x)\} = \left(1 + \frac{\lambda}{|\tilde{R}(k_x)|^2}\right)^{-1}. \quad (2.24)$$

It is readily apparent that $H\{\tilde{R}(k_x)\}$ then represents a special filtering term in the regularized solution. It suppresses the non-regularized solution at the spatial frequencies at which the response function goes to zero, and the regularization constant λ acts as a weight to control the degree to which small $|\tilde{R}(k_x)|^2$ is suppressed. By this means, the instability is compensated and the deconvolution made possible.

The formulation in equation 2.23 is very useful and suggests that the “filter” $H\{\tilde{R}(k_x)\}$ can be modified to suit some desired performance abilities. One possible modification to the filtering term is to not only suppress small \tilde{R} , but to penalize its high frequency nulls even more. One way to do this is the following:

$$H\{\tilde{R}(k_x)\} = \left(1 + \frac{\lambda k_x}{|\tilde{R}(k_x)|^2}\right)^{-1}. \quad (2.25)$$

Here, the regularization factor is effectively λk_x , which means that higher spatial frequency components of R will go to zero faster than lower frequencies. Interestingly, another option that works very well is

$$H\{\tilde{R}(k_x)\} = \left(1 + \frac{\lambda}{k_x |\tilde{R}(k_x)|^2}\right)^{-1}, \quad (2.26)$$

which has the opposite effect. This linear frequency weighting works because the Fourier components of the response function die off roughly exponentially with frequency on their own due to the Lorentzian-like shape of the response function and

energy conservation. The modified filter in equation 2.26 can thus be interpreted as compensating for this in an attempt to limit the distortion induced by the regularization process itself.

Finally, further enhancement can be obtained by exploiting the fact that digital magnetic recording is saturation recording. As previously mentioned, bits are represented in longitudinal recording by one of two directions of the longitudinal magnetization, M_x . This implies that, in the ideal case with no ISI and infinitely steep transitions (delta function charges), the magnitude of a bit in terms of magnetization is constant and the direction is represented by one of two levels (e.g., 1 and -1). Therefore, any deviations from a square-wave-like longitudinal magnetization correspond to either non-localized charge distributions or unwanted distortions. The latter are dominant and include ISI and noise. Hence, another technique to improve the quality of the reconstructed charge is to apply a saturating post-processing step to the the magnetization after regularized deconvolution [21]. Arctangent (with a scaled argument) is one particularly useful saturation operator \mathcal{S} with a controllable amount of saturation,

$$\mathcal{S}\{M_x\} = \tan^{-1}\{\alpha M_x\}. \quad (2.27)$$

Note that in the limit $\alpha \rightarrow \infty$, the arctangent approaches a step function. Controlling α adjusts the degree to which small perturbations in the components of M_x are suppressed and large, slowly varying regions are strengthened. A large α gives more saturation, with the caveat that any low frequency noise present is amplified. The constant α furthermore has the effect of controlling the widths of the postprocessed

readback pulses (the divergence of M_x), for at the limit of a true square wave in M_x , the transition widths are zero, which results in delta-function-like charges.

2.2.3 Experimental Results

In practice, it is found that for the best ISI removal performance, all three of the aforementioned techniques — regularization, solving for magnetization instead of charge, and saturating post-processing — should be used in conjunction. Doing so provides several parameters that can be tweaked: the regularization constant λ , the arctangent scale α , and the response width. In the following examples, equation 2.23 for 1D (or its analogue for 2D) is used with the regularization filter in equation 2.26. The recovered charge is obtained from the magnetization by equation 2.1.

As an example of the different results that may be obtained by the deconvolution algorithm when changing these parameters, it is instructive to look at a 1D example taken from spin-stand readback of a 2600 Oe disk with $0.39\mu\text{m}$ head on which a periodic tribit pattern (“...000111000111000...”) has been written. The density $D = \frac{PW_{50}}{T}$ upon readback is approximately 1.5. Assuming ideal localized charges are the source of the readback, the algorithm yields figure 2.10 as its result with all parameters tuned properly. It is readily observable that the recovered (ISI-removed) charge is localized and has proper magnitude and locations restored. The middle transition of the tribit is no longer heavily reduced, and the outer transitions are shifted slightly inward from the peaks in the signal with ISI. An interesting side effect of the deconvolution algorithm is the ability to produce a reconstructed, “de-

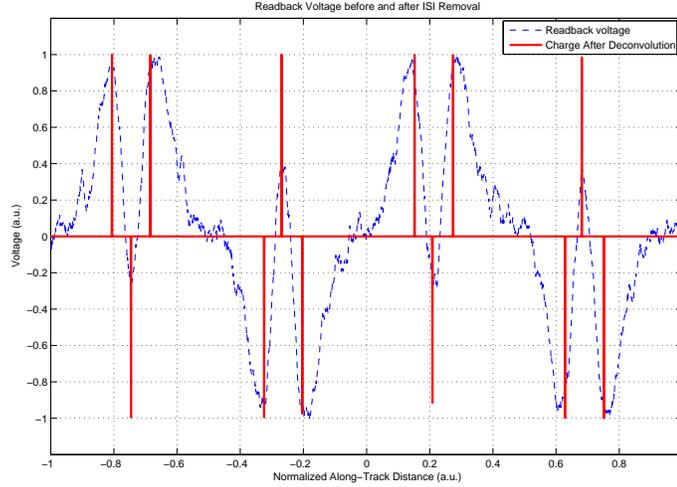


Figure 2.10: ISI removed from a periodic tribit pattern, with good parameter choice ($\lambda = 10^{-3}$, $\alpha = 10^4$).

noised” version of the ISI-corrupted input by convolving the response function with the recovered charge (figure 2.11).

If, on the other hand, the regularization constant λ is not correct, the effect is one of two possibilities. First, if the constant is too small, the algorithm will be unstable and spurious transitions (figure 2.12) or, in the extreme case, high frequency oscillations will result and obliterate the desired solution. Recall that the reason for regularization in the first place was for stability of the solution. Second, if the constant is too large, the recovered charge will be corrupted in position. As λ increases, the relative weight of the smoothing functional \mathcal{B} in the minimization problem 2.9 increases while that of the accuracy functional \mathcal{A} decreases. Thus, the regularized output of the deconvolution algorithm, as the solution to this minimization problem, reflects more of the underlying smooth (low-frequency) characteristics of the

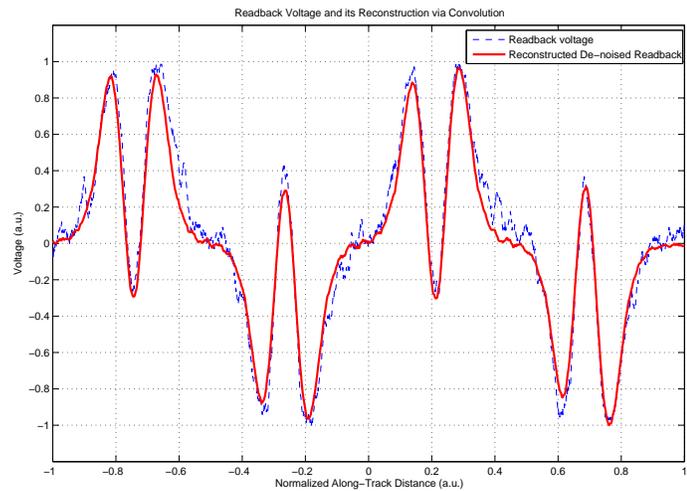


Figure 2.11: Periodic tribit readback voltage and its “denoised” reconstruction.

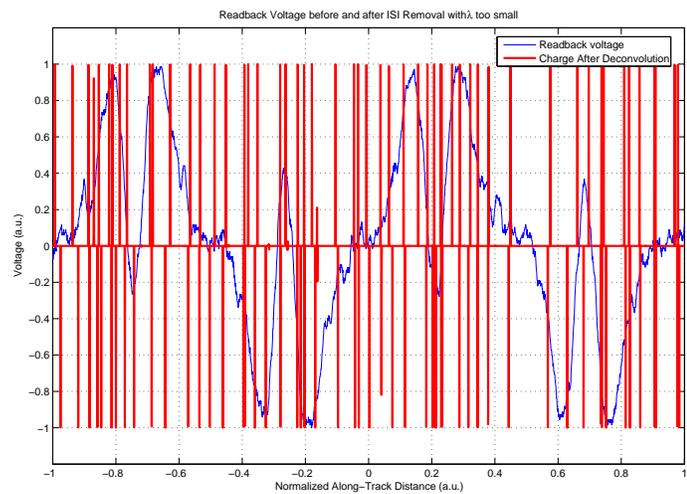


Figure 2.12: ISI removed from a periodic tribit pattern, λ too small ($\lambda = 10^{-5}$).

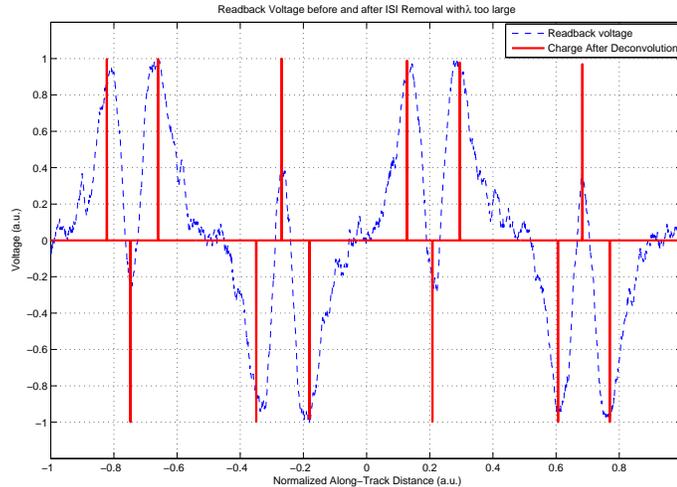


Figure 2.13: ISI removed from a periodic tribit pattern, λ too large ($\lambda = 10$). The recovered charge locations in this case are actually further apart than the original readback peaks with ISI.

input at the expense of accuracy as λ increases. For this reason, for λ too large, the ISI-removed charge locations appear shifted away from each other and not closer to each other, as they should (figure 2.13). In other words, with the other algorithm parameters correct, the result of deconvolution with a too-large regularization constant is only peak magnitude restoration and not linear bit shift compensation. As further proof, figure 2.14 shows the reconstructed denoised readback when the recovered charge has been corrupted by large λ ; it is quite dissimilar to the original readback.

Another important parameter that must be correct is the response function width (or in two dimensions, widths along the major and minor axes). As previously mentioned, a measured head response function can be numerically scaled and often

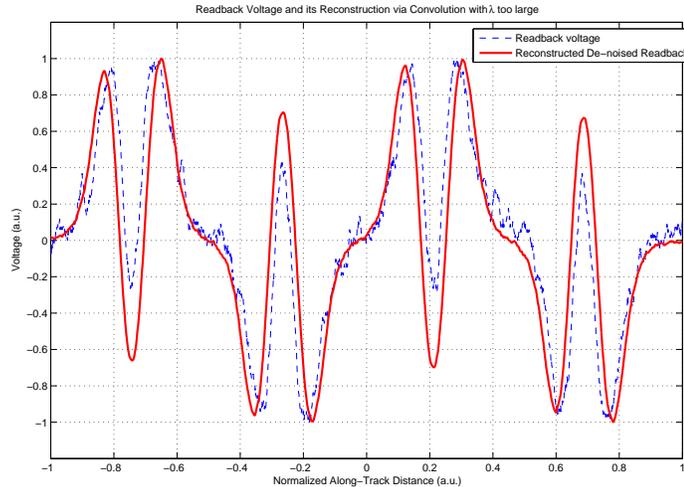


Figure 2.14: Periodic Tribit readback voltage and its “denoised” reconstruction with λ too large ($\lambda = 10$).

must be for optimal deconvolution performance. The effect of the head response width on deconvolution can be quite intuitively understood; too wide will completely miss small desired features, whereas too small will overemphasize small undesirable features (noise). These two cases are illustrated in figure 2.15.

The last parameter in the deconvolution algorithm is the arctangent scaling factor α . When α is too small, high frequency noise corrupts the recovered magnetization. To recover the charge requires the divergence of this magnetization, or in 1D, differentiation. Differentiating high frequency noise magnifies that noise, as seen in figure 2.16.

Lastly, to illustrate the utility of the regularized deconvolution algorithm with arctangent saturation on 2D images from modern hard disks, figure 2.17 shows the results of the regularized deconvolution algorithm with arctangent saturation

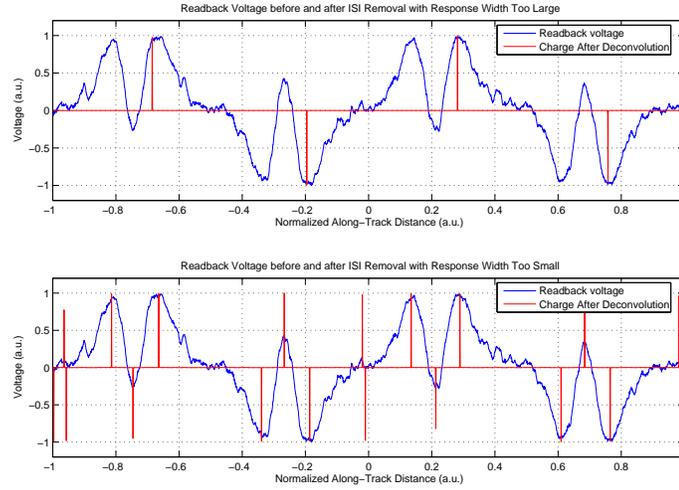


Figure 2.15: ISI Removed from a periodic tribit pattern, with the head response too wide (top) and too narrow (bottom). Here, the too wide response is $2PW_{50}$ and the too narrow is $PW_{50}/4$ where PW_{50} is the width of the response used in figure 2.10.

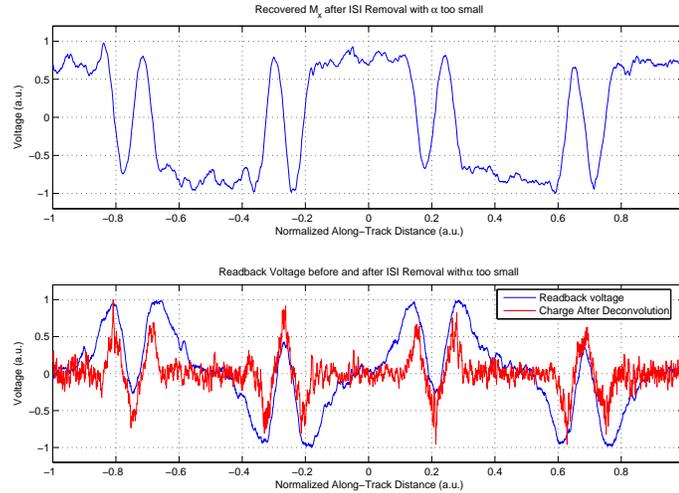


Figure 2.16: ISI Removed from a periodic tribit pattern, with α too small. The top plot of M_x shows how the regions of zero charge are corrupted by noise. The transition peaks are located at the zero crossings of the magnetization.

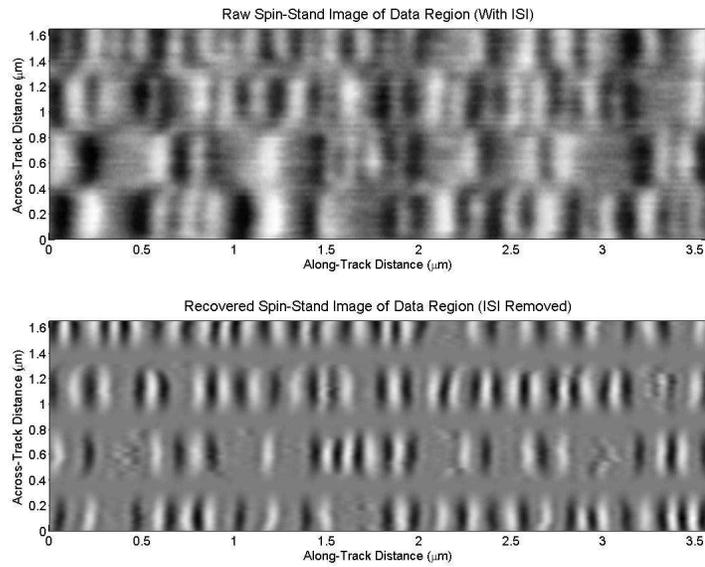


Figure 2.17: Image taken from a commercial hard disk from 2002, before and after ISI removal.

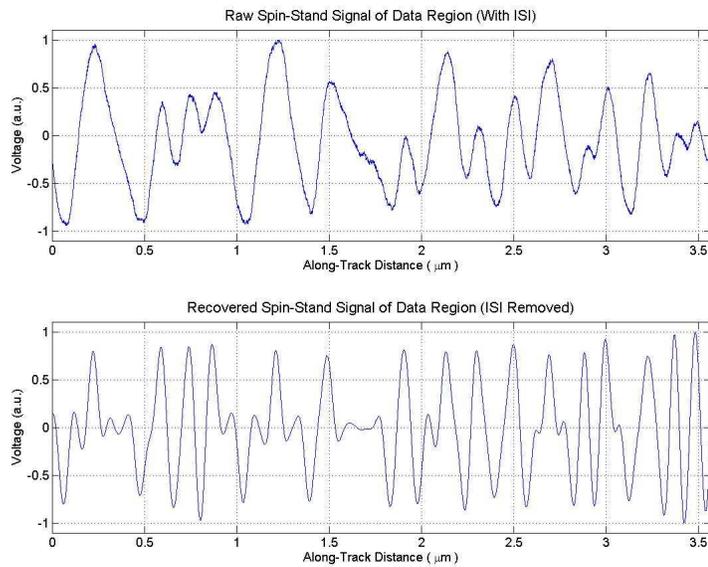


Figure 2.18: Along-track signal taken from a commercial hard disk from 2002, before and after ISI removal.

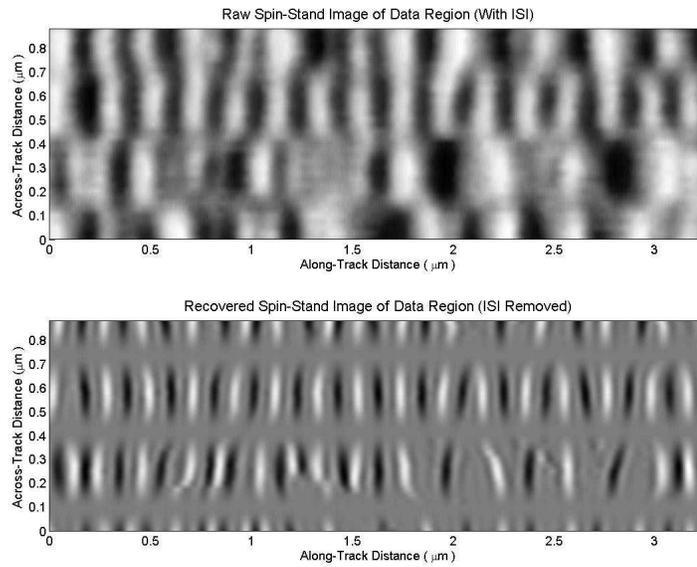


Figure 2.19: Image taken a commercial hard disk from 2003, before and after ISI removal.

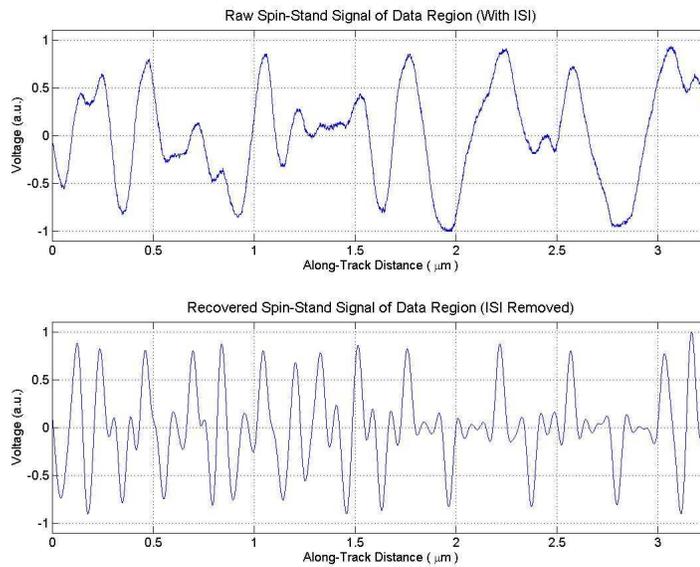


Figure 2.20: Along-track signal taken from a commercial hard disk from 2003, before and after ISI removal.

applied to a spin-stand scan of a disk from the year 2002 with a $0.19\mu m$ width read head. Here, an arctangent processing step has also been applied in the cross-track direction to accentuate the guard bands, the area of DC-erased media between the tracks. Figure 2.18 shows the resulting ISI-removed 1D readback signal from the center of the bottommost track in 1.15.

Figures 2.19 and 2.20 show the results of the deconvolution for a disk from the year 2003, again with a $0.19\mu m$ read head (and also with an arctangent processing step in the cross-track direction to accentuate the guard bands). Figure 2.20 shows the 1D signal taken from the image at approximately $0.3\mu m$ cross-track distance.

2.3 Conclusions

A technique for 1D and 2D ISI removal in spin-stand images of longitudinal recording disks has been described. This distortion is due to the non-local sensing properties of the head, and is quantified by the head response function, whose measurement was described and depicted. The implementation of the distortion removal algorithm was shown to depend also on regularization and saturation techniques. Finally, a series of results have been shown that illustrate the ISI removal technique's effectiveness in removing this distortion.

Chapter 3

The Hilbert Transform in Magnetic Recording

It is well known that the Hilbert transform (also familiar in communications theory) is applicable to magnetic recording in describing the relationship between in-plane and out-of-plane magnetic field components [22]-[26]. This chapter describes the utility of this relationship in the context of spin-stand imaging and additionally derives a multidimensional extension of the Hilbert transform from physical arguments.

3.1 The 1D Hilbert Transform

3.1.1 Theory

The literature states that, in the two-dimensional recording plane approximation of the magnetic recording medium (x - z plane, e.g., figure 2.2), the transverse (H_z) and longitudinal (H_x) components of the external magnetic field emanating from the media are related by the one-dimensional Hilbert transform [22]-[26]. The Hilbert transform $\mathcal{H}\{\}$ is an integral transform that performs a $\frac{\pi}{2}$ phase shift and is given by

$$H_x(x) = \mathcal{H}\{H_z(x)\} = \frac{1}{\pi} PV \int_{-\infty}^{\infty} \frac{H_z(x')}{x - x'} dx', \quad (3.1)$$

where PV represents that the integral is understood in the sense of its Cauchy principal value. Likewise, the transverse component is given by

$$H_z(x) = -\mathcal{H}\{H_x(x)\} = -\frac{1}{\pi}PV \int_{-\infty}^{\infty} \frac{H_x(x')}{x-x'} dx', \quad (3.2)$$

This one-dimensional transform follows directly from the one-sided, half-plane nature of the two-dimensional geometry. That is, the source produces the fields on only one side of the infinite flat recording surface (perpendicular to the recording plane) at $z = 0$. Minuhin's derivation [24, 25] uses a formulation based on the complex scalar potential produced in this situation and exploits the properties of analytic functions in the half plane. Equivalently, Mallinson [22] and Bertram [26] look at solution of Laplace's equation for the scalar potential given the one-sided (half-plane) boundary conditions.

The Fourier transform representation of the Hilbert transform is of great computational utility. By ready inspection it can be seen that the Hilbert transform integral given in equation 3.1 is the convolution

$$H_x(x) = H_z(x) * \frac{1}{\pi x}. \quad (3.3)$$

It then follows from rudimentary Fourier analysis that the transform of the Hilbert transform kernel $\frac{1}{\pi x}$ is $-j\text{sgn}(k_x)$, where $\text{sgn}()$ denotes the signum function. Therefore, equations 3.1 and 3.2 are given in spatial frequency domain as

$$\tilde{H}_x(k_x) = -j\text{sgn}(k_x)\tilde{H}_z(k_x), \quad (3.4)$$

$$\tilde{H}_z(k_x) = j\text{sgn}(k_x)\tilde{H}_x(k_x). \quad (3.5)$$

The Hilbert transform has the interesting property of scaling invariance, and is the only nontrivial linear integral transform for which this property holds [27]. That is, if a scaling in space is performed in H_z ($H_z(x) \rightarrow H_z(ax)$) then the same scaling is obtained in H_x ($H_x(x) \rightarrow H_x(ax)$), and vice versa. This fact is expressed most directly in the Fourier transform representation of the Hilbert transform. From the scaling property of the Fourier transform, it is known that

$$H_x(ax) \implies \frac{1}{|a|} \tilde{H}_x \left(\frac{k_x}{a} \right). \quad (3.6)$$

Therefore, substituting the Hilbert transform relation 3.4 gives

$$H_x(ax) \implies -j \operatorname{sgn} \left(\frac{k_x}{a} \right) \frac{1}{|a|} \tilde{H}_z \left(\frac{k_x}{a} \right). \quad (3.7)$$

The only non-constant part of the kernel's transform is at the origin, where it is zero. Therefore, any scaling in space does not affect the Hilbert transform process itself, and $\operatorname{sgn}(k_x) = \operatorname{sgn}(k_x/a)$. Thus,

$$H_x(ax) \implies -j \operatorname{sgn}(k_x) \frac{1}{|a|} \tilde{H}_z \left(\frac{k_x}{a} \right), \quad (3.8)$$

yielding

$$H_x(ax) = \mathcal{H}\{H_z(ax)\}. \quad (3.9)$$

The Hilbert transform appears to be merely descriptive but is in fact usable for ISI removal for longitudinal recording readback. The MR head readback voltage is, as discussed in section 2.2, the media magnetic charge after blurring by the head response function. As mentioned in section 2.1, the shielded MR head ideally senses only the vertical component of the medium external field H_z and by symmetry, the

surface charge σ_m is directly proportional to the vertical component of the field. Thus, up to a constant, the head readback is a blurred representation of H_z . It follows then that the 1D Hilbert transform of the readback voltage is equivalent to a blurred form of H_x , which follows the along-track magnetization M_x up to a sign ($H_x \propto -M_x$).

Interestingly, it turns out that, at moderate levels, the ISI in H_z is typically more pronounced in terms of peak amplitude loss than it is in H_x . Once H_x is obtained, it can be saturated by means of arctangent or any of the family of odd-symmetric step-like functions, and subsequently differentiated to obtain a “cleaned” reconstruction of H_z (or σ_m). The degree of saturation, determined by the derivative of the step function near the origin, controls the degree of pulse slimming and spurious transition (noise) amplification, as discussed previously in the context of deconvolution enhancement. Using a nonlinear operation to approximate deconvolution is a concept known in image processing in general, e.g., the work on shock filters in nonlinear PDE-based deconvolution and denoising [29, 30].

Even in the limit of no saturation, however, this technique still has ISI-removal properties that can be ascertained by Fourier analysis. Using equations 2.2 and 3.1 and assuming $V(x) = H_z(x)$ and $H_x = -M_x$, the recovered charge is given by

$$\sigma_m(x) = \frac{d}{dx} \mathcal{H}\{V(x)\}. \quad (3.10)$$

Taking the Fourier transform of this gives

$$\tilde{\sigma}_m(k_x) = -jk_x \text{sgn}(k_x) \tilde{V}(k_x). \quad (3.11)$$

The magnitude of the term multiplying $\tilde{V}(k_x)$ looks like an absolute value function

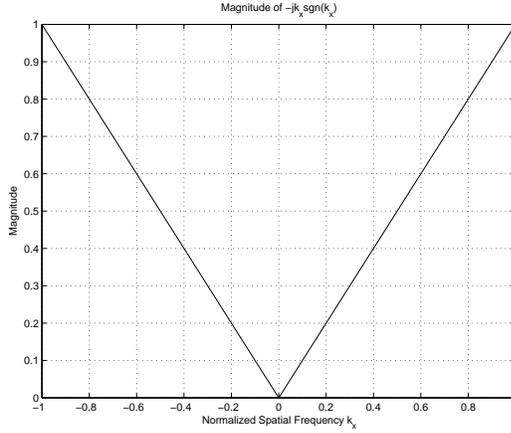


Figure 3.1: Magnitude as a function of spatial frequency for the “filter” represented by Hilbert transformation and subsequent differentiation

of spatial frequency as shown in figure 3.1. Therefore, higher frequency component magnitudes are emphasized in comparison to lower frequency components, which has the effect of pulse slimming; i.e., differentiation is a high-pass filtering operation. Neighboring slim pulses do not interfere with each other to the same extent as wide pulses, and thus, a degree of ISI removal takes place. In practice, however, due to the presence of noise, differentiation cannot be done without a saturation step preceding it. With saturation, the total 1D Hilbert transform based ISI-removal algorithm is

$$\sigma_m(x) = \frac{d}{dx} \tan^{-1}(\alpha \mathcal{H}\{V(x)\}). \quad (3.12)$$

3.1.2 Experimental Results

Unlike the head response deconvolution technique for ISI removal, calibrating the 1D Hilbert transform technique requires only adjustment of the saturation (arc-

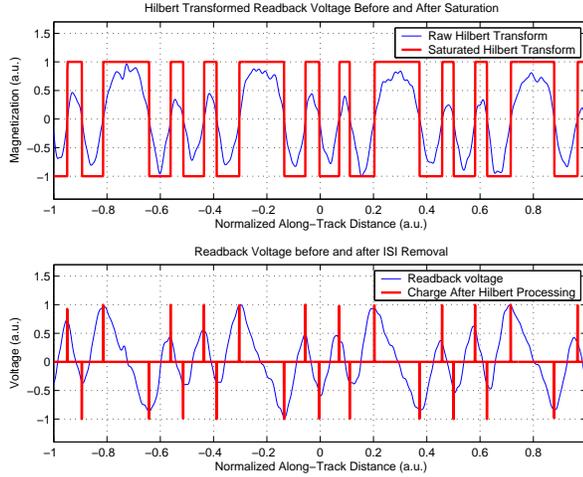


Figure 3.2: ISI Removed from a periodic “F9” (11111001) pattern using the 1D Hilbert Transform method with $\alpha = 10^4$.

tangent) scaling factor. This is beneficial in the sense that optimizing the algorithm is not a multidimensional problem, but drawbacks exist in terms of flexibility and range of performance in comparison. The following examples show the surprising capabilities of this technique along with its limitations.

Figure 3.2 shows the successful result of the Hilbert transform ISI-removal method from spin-stand readback of a “F9” pattern written on a 2600 Oe disk and read with a $0.39\mu\text{m}$ head, resulting in a density of 1.5. The readback was subsequently low-pass filtered to remove high-frequency noise.

In order for the Hilbert transform technique to accurately remove ISI, the readback voltage must be relatively smooth so that spurious transitions do not result. Without low-pass filtering the data in the previous plot, the detail in figure 3.3 is obtained, showing false transitions. False transitions occur when M_x , the Hilbert transform of the readback voltage, has zero crossings that should not be present, but

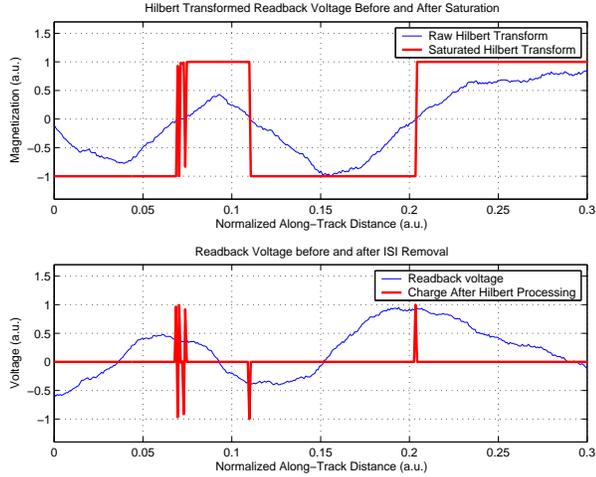


Figure 3.3: ISI Removed from a periodic “F9” (11111001) pattern using the 1D Hilbert Transform method with $\alpha = 10^4$, showing false transitions.

are present due to noise. For α too large, the arctangent saturation subsequently blows these up, and they appear in the recovered charge after differentiation. If, on the other hand, α is too small, then the differentiation step will magnify the noise, as seen in figure 3.4 with $\alpha = 1$. Finally, figure 3.5 depicts the 1D Hilbert transform method’s abilities of tackling ISI from readback from a commercial hard disk from the year 2000.

3.2 The 2D Hilbert Transform

3.2.1 Derivation and Theory

The one-dimensional Hilbert transform relates the scalar longitudinal and transverse field components H_x and H_z under the assumption that the 2D recording plane x - z is a valid representation of the media recording geometry. In actual-

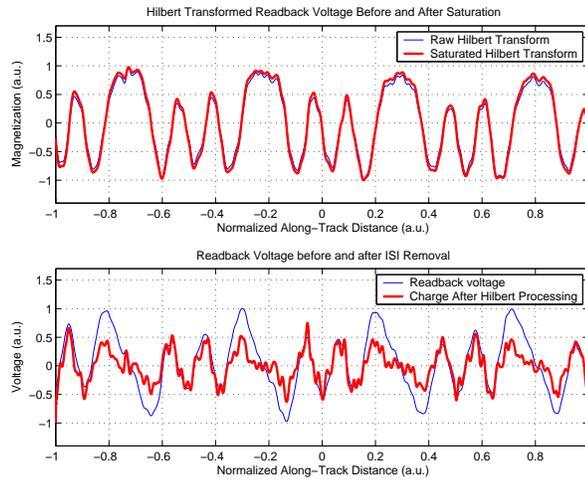


Figure 3.4: ISI Removed from a periodic “F9” (11111001) pattern using the 1D Hilbert Transform method with insufficient saturation, $\alpha = 1$.

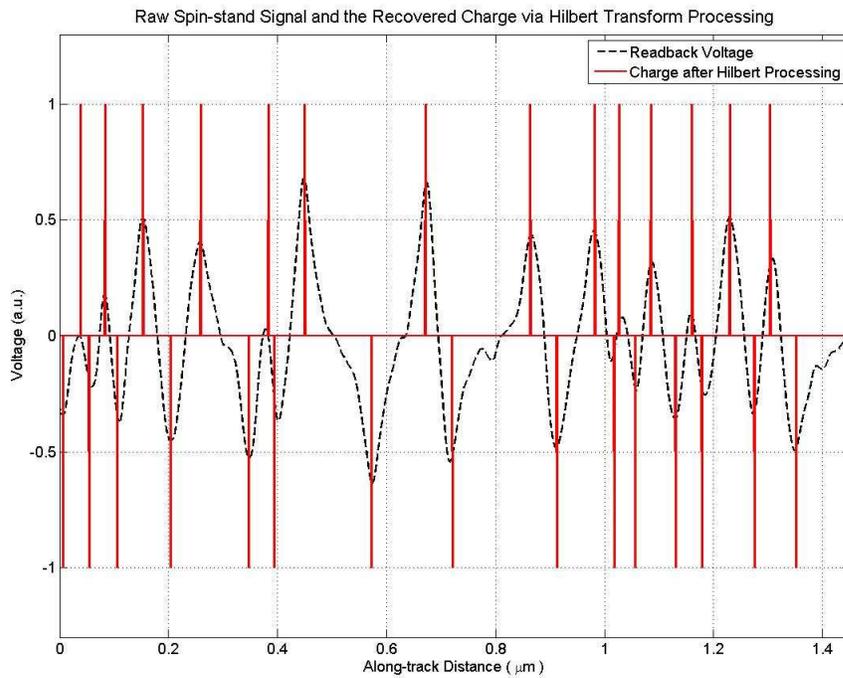


Figure 3.5: ISI Removed from a year 2000 commercial disk ata region using the 1D Hilbert Transform method, $\alpha = 2000$.

ity, as track densities increase, the assumption that the recording media is uniform in the across-track (y) direction becomes simply incorrect. Thus, the “standard” one-dimensional Hilbert transform does not apply to this situation.

To deal with this fact, the one-dimensional Hilbert transform can be extended to two dimensions by physical arguments [11], giving an integral transform that relates the scalar transverse component H_z , normal to the media plane, to the vectorial longitudinal component $H_x\vec{e}_x + H_y\vec{e}_y$ parallel to the media plane (figure 3.6). (Note that the media plane x - y and the recording plane x - z are orthogonal.) It is worth noting that a derivation different from the one that follows can be made in a different context (geomagnetics), which yields similar results (although not called a 2D Hilbert transform) [28].

To derive a relation for the 2D Hilbert transform in magnetic recording, first consider the medium source in its surface magnetic charge form, as introduced previously. The magnetic charge is located at $z = 0$ on the zero-thickness x - y media plane, as shown on figure 3.7. The transverse component of the magnetic field is therefore given by

$$H_z^- - H_z^+ = \frac{\sigma_m}{\mu_0}. \quad (3.13)$$

Due to the symmetry of the flat media plane, it is readily apparent that

$$H_z^- = -H_z^+, \quad (3.14)$$

which in turn implies that

$$H_z^- = \frac{\sigma_m}{2\mu_0}. \quad (3.15)$$

Naturally, the magnetic scalar potential due to this surface magnetic charge is given

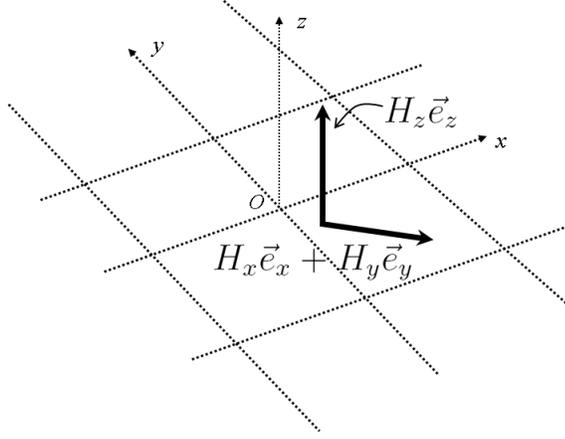


Figure 3.6: The 3D schematic with media plane and the magnetic fields emanating from it.

by

$$\phi_m(x, y, z) = \frac{1}{4\pi\mu_0} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\sigma_m(x', y')}{((x - x')^2 + (y - y')^2 + z^2)^{1/2}} dx' dy', \quad (3.16)$$

which can be related back to the field via its gradient,

$$\vec{H} = -\nabla\phi_m. \quad (3.17)$$

Applying equation 3.17 to equation 3.16 gives the longitudinal components H_x and H_y in terms of the charge σ_m :

$$H_x(x, y, z) = -\frac{\partial\phi_m}{\partial x} = \frac{1}{4\pi\mu_0} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{(x - x')\sigma_m(x', y')}{((x - x')^2 + (y - y')^2 + z^2)^{3/2}} dx' dy', \quad (3.18)$$

$$H_y(x, y, z) = -\frac{\partial\phi_m}{\partial y} = \frac{1}{4\pi\mu_0} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{(y - y')\sigma_m(x', y')}{((x - x')^2 + (y - y')^2 + z^2)^{3/2}} dx' dy'. \quad (3.19)$$

Substituting equation 3.15 into these expressions and evaluating at the media surface $z = 0$ gives the following 2D Hilbert transform:

$$H_x(x, y, 0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{(x - x')H_z(x', y', 0)}{((x - x')^2 + (y - y')^2)^{3/2}} dx' dy', \quad (3.20)$$

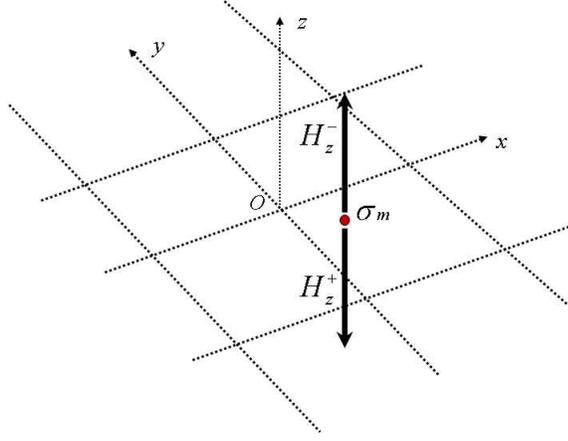


Figure 3.7: The surface magnetic charge representation of the media with the transverse (normal) components of the resultant magnetic field.

$$H_y(x, y, 0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{(y - y')H_z(x', y', 0)}{((x - x')^2 + (y - y')^2)^{3/2}} dx' dy'. \quad (3.21)$$

Thus, we now have the integral transform that gives the longitudinal components of the magnetic field while knowing the transverse component of the field at the surface.

To derive the analogous “inverse” transform, giving H_z while knowing the longitudinal vectorial components, requires the use of another equivalent representation of the recording medium. Here it is useful to consider the source of the media magnetic fields as virtual electric surface current \vec{i} at $z = 0$, as shown in figure 3.8.

The virtual current is related to the field it produces by the relation

$$\vec{e}_z \times (\vec{H}^- - \vec{H}^+) = \vec{i}. \quad (3.22)$$

The symmetry induced by the flat media plane again simplifies the problem by yielding the relation

$$\vec{e}_z \times \vec{H}^- = -\vec{e}_z \times \vec{H}^+, \quad (3.23)$$

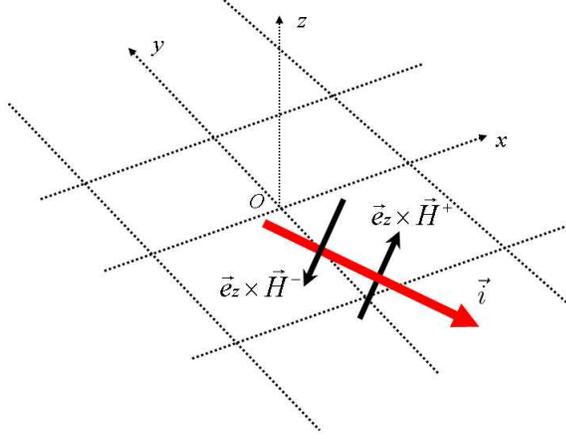


Figure 3.8: The surface virtual electric current representation of the media with the longitudinal (tangential) components of the resultant magnetic field.

which with equation 3.22 implies that

$$\vec{e}_z \times \vec{H}^- = \frac{\vec{i}}{2}. \quad (3.24)$$

At the same time, the magnetic field can be related to the virtual current via the Biot-Savart law,

$$\vec{H}(x, y, z) = \frac{1}{4\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\vec{i}(x', y', z') \times ((x - x')\vec{e}_x + (y - y')\vec{e}_y + z\vec{e}_z)}{((x - x')^2 + (y - y')^2 + z^2)^{3/2}} dx' dy'. \quad (3.25)$$

After substituting equation 3.24 into equation 3.25 and employing the definition $H_z = \vec{H} \cdot \vec{e}_z$, the following expression for H_z at the media surface is obtained:

$$H_z(x, y, 0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{[(\vec{e}_z \times \vec{H}(x', y', 0)) \times ((x - x')\vec{e}_x + (y - y')\vec{e}_y)] \cdot \vec{e}_z}{((x - x')^2 + (y - y')^2)^{3/2}} dx' dy'. \quad (3.26)$$

Straightforward vector algebra simplifies the numerator of the integrand. First, note that

$$\vec{e}_z \times \vec{H}(x', y', 0) = \vec{e}_y H_x(x', y', 0) - \vec{e}_x H_y(x', y', 0). \quad (3.27)$$

Defining the numerator of the integrand as

$$N = \left[(\vec{e}_z \times \vec{H}(x', y', 0)) \times ((x - x')\vec{e}_x + (y - y')\vec{e}_y) \right] \cdot \vec{e}_z, \quad (3.28)$$

substituting equation 3.27 into N and distributing yields the numerator

$$\begin{aligned} N = & \vec{e}_z \cdot (\vec{e}_y \times \vec{e}_x) H_x(x', y', 0)(x - x') + \vec{e}_z \cdot (\vec{e}_y \times \vec{e}_y) H_x(x', y', 0)(y - y') \\ & - \vec{e}_z \cdot (\vec{e}_x \times \vec{e}_x) H_y(x', y', 0)(x - x') - \vec{e}_z \cdot (\vec{e}_x \times \vec{e}_y) H_y(x', y', 0)(y - y'), \end{aligned} \quad (3.29)$$

which simplifies to

$$N = - [H_x(x', y', 0)(x - x') + H_y(x', y', 0)(y - y')]. \quad (3.30)$$

Thus, the “inverse” 2D Hilbert transform has been derived, yielding the transverse (normal) component H_z given the longitudinal (tangential) magnetic field components on the media surface:

$$H_z(x, y, 0) = -\frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{H_x(x', y', 0)(x - x') + H_y(x', y', 0)(y - y')}{((x - x')^2 + (y - y')^2)^{3/2}} dx' dy'. \quad (3.31)$$

The integral expressions for the 2D Hilbert transform in both directions (longitudinal to transverse and transverse to longitudinal) are, as in their single dimensional case, understood in the sense of Cauchy principle value due to their singular nature. Note also in comparing the just derived 2D generalization with the 1D case that the signs of the integrals as well as their general form is quite similar. To see further similar properties requires the derivation of the 2D Hilbert transform’s Fourier transform representation, which is useful in itself.

Solution of the boundary value problem of Laplace’s equation for the scalar magnetic potential in the region above the medium yields the Fourier transform

representations of equations 3.20, 3.21, and 3.31. As a scalar potential, $\phi_m(x, y, z)$ satisfies Laplace's equation in three dimensions in the region $z > 0$,

$$\nabla^2 \phi_m(x, y, z) = 0, \quad (3.32)$$

with the mixed boundary conditions

$$\frac{\partial \phi_m(x, y, 0)}{\partial z} = -H_z(x, y, 0), \quad (3.33)$$

$$\phi_m(x, y, \infty) = 0. \quad (3.34)$$

This PDE is reduced to an ODE by means of the tangential 2D Fourier transform of the scalar potential,

$$\tilde{\phi}_m(k_x, k_y, z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi_m(x, y, z) e^{-j(k_x x + k_y y)} dx dy. \quad (3.35)$$

This yields the 1D analogous boundary value problem

$$\frac{d^2 \tilde{\phi}_m(k_x, k_y, z)}{dz^2} = - (k_x^2 + k_y^2) \tilde{\phi}_m(k_x, k_y, z) = 0, \quad (3.36)$$

$$\left. \frac{d\tilde{\phi}_m}{dz} \right|_{z=0} = -\tilde{H}_z, \quad (3.37)$$

$$\tilde{\phi}_m(k_x, k_y, \infty) = 0. \quad (3.38)$$

The solution of this problem is simple, giving

$$\tilde{\phi}_m(k_x, k_y, z) = \tilde{H}_z \frac{e^{-z\sqrt{k_x^2 + k_y^2}}}{\sqrt{k_x^2 + k_y^2}} = \tilde{H}_z \frac{e^{-z|\vec{k}|}}{|\vec{k}|}. \quad (3.39)$$

Evaluating the above solution at $z = 0$ and recalling that $H_x = -\frac{d\phi_m}{dx}$ and $H_y = -\frac{d\phi_m}{dy}$ (from equation 3.17) yields the Fourier transform representation of the 2D Hilbert transform:

$$\tilde{H}_x(k_x, k_y, 0) = \frac{-jk_x}{|\vec{k}|} \tilde{H}_z(k_x, k_y, 0), \quad (3.40)$$

$$\tilde{H}_y(k_x, k_y, 0) = \frac{-jk_y}{|\vec{k}|} \tilde{H}_z(k_x, k_y, 0). \quad (3.41)$$

Equations 3.40 and 3.41 together imply that $k_y \tilde{H}_x = k_x \tilde{H}_y$ and give the Fourier transform representation of the inverse transform:

$$\tilde{H}_z(k_x, k_y, 0) = \frac{j(\tilde{H}_x k_x + \tilde{H}_y k_y)}{|\vec{k}|}. \quad (3.42)$$

These Fourier representations give a couple interesting properties of the 2D Hilbert transform. First, like the 1D Hilbert transform, scaling invariance exists by virtue of the scaling property of the Fourier transform; i.e, if $H_x(x, y) \rightarrow H_x(ax, ay)$ and $H_y(x, y) \rightarrow H_y(ax, ay)$, then $H_z(x, y) \rightarrow H_z(ax, ay)$, and vice versa. Second, Parseval's relation together with equations 3.40 and 3.41 give the relation

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [H_x^2(x, y) + H_y^2(x, y)] dx' dy' = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H_z^2(x, y) dx' dy', \quad (3.43)$$

which can be rewritten by substituting equations 3.15 and 3.24 as

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\vec{i}(x, y)|^2 dx dy = \frac{1}{\mu_0} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sigma_m^2(x, y) dx dy, \quad (3.44)$$

relating the virtual surface electric current to its corresponding surface magnetic charge.

Thus, it can be readily seen that this is indeed a 2D generalization of the Hilbert transform in the context of magnetic recording. First, the 1D transform can be easily recovered from the 2D transform by neglecting variation in the across-track direction, i.e., by setting $dy = 0$. Second, the properties of rate independence manifested in the equations by scaling invariance is present in both 1D and the 2D transforms due to their similar Fourier representations. Likewise, both 1D and 2D

transforms preserve L_2 norms. Finally and most fundamentally, they both have the same physical interpretation, that of relating the normal and tangential components of the magnetic field emanating from the medium at the surface of the same medium. The major distinction between the 1D and 2D transforms is that the 2D transform is a vector-to-scalar relationship (and vice versa), whereas the 1D transform relates scalar quantities.

3.2.2 Experimental Results

Although interesting in its own right, the 2D Hilbert transform has further utility in that it can be used analogously as the 1D Hilbert transform for ISI removal. Thus, in the manner of equation 3.12, the 2D algorithm with saturation is

$$\sigma_m(x, y) = \nabla \cdot \tan^{-1}(\alpha \mathcal{H}\{V(x, y)\}), \quad (3.45)$$

where $\mathcal{H}\{\}$ here is understood as the 2D Hilbert transform just derived and the other quantities are as previously defined. Also, in writing the divergence, the assumption is tacitly made that $\frac{\partial M_x}{\partial x} = -\frac{\partial H_x}{\partial x}$ and $\frac{\partial M_y}{\partial y} = -\frac{\partial H_y}{\partial y}$.

The first example, figure 3.9, presents the result of applying the 2D Hilbert transform to a numerical simulation of two adjacent readback tracks with 2D ISI. To produce the ISI-removed charge requires applying the divergence to this recovered vectorial field quantity. Doing so yields the reconstructed charge seen in figure 3.10, where arctangent saturation has *not* been applied before taking the divergence. As discussed in the previous section on the 1D Hilbert transform, the Hilbert transform *itself* has ISI removal capabilities which are obvious in the low noise environment of

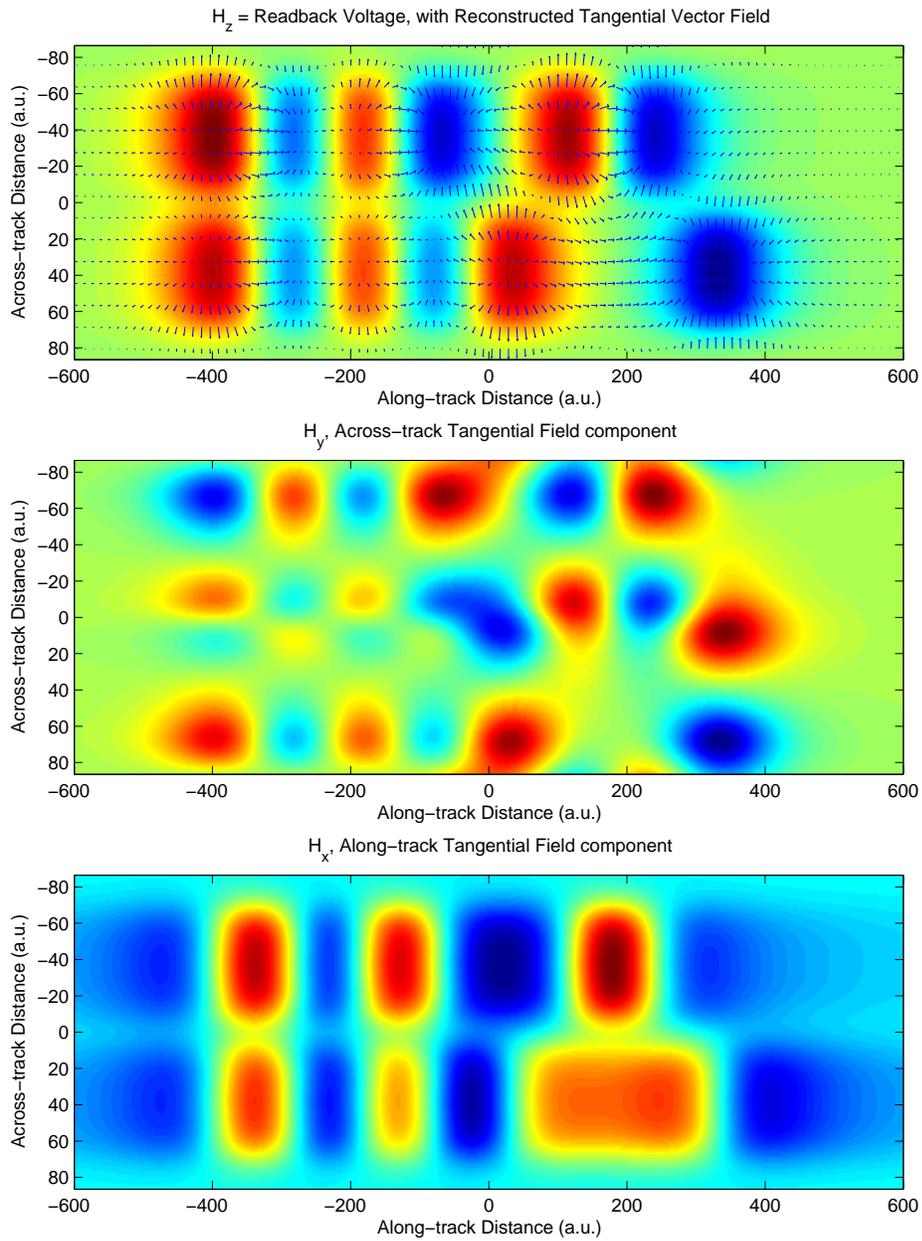


Figure 3.9: The simulated readback voltage image of $H_z(x, y)$ overlaid with its reconstructed vectorial tangential field (top); the reconstructed across-track component $H_y(x, y)$ (middle); and the reconstructed along-track component $H_x(x, y)$ (bottom).

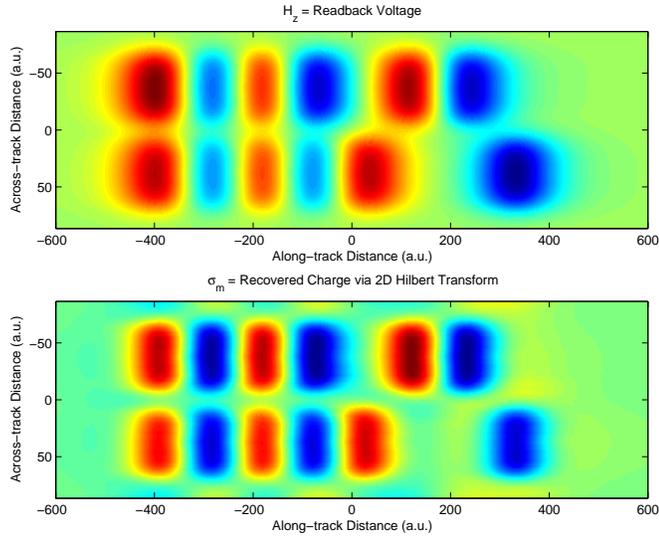


Figure 3.10: The simulated readback voltage image of $H_z(x, y)$ (top), and the 2D Hilbert transform reconstructed charge $\sigma_m(x, y)$ without arctangent saturation (bottom).

this simulation. Applying arctangent saturation to both field components H_x and H_y does further narrowing, as seen in figure 3.11. The next example, figure 3.12, is the result of employing the algorithm on an image taken from a commercial hard disk platter from the year 2003.

While the approximations $\frac{\partial M_x}{\partial x} = -\frac{\partial H_x}{\partial x}$ and $\frac{\partial M_y}{\partial y} = -\frac{\partial H_y}{\partial y}$ are valid, the magnetization can be more rigorously solved in the following manner. Because $\nabla \cdot \vec{B} = 0$ and $\vec{B} = \mu_0(\vec{H} + \vec{M})$, we know that

$$\nabla \cdot \vec{H} = -\nabla \cdot \vec{M}. \quad (3.46)$$

Furthermore, because of the flat symmetry property of the ideally thin medium of

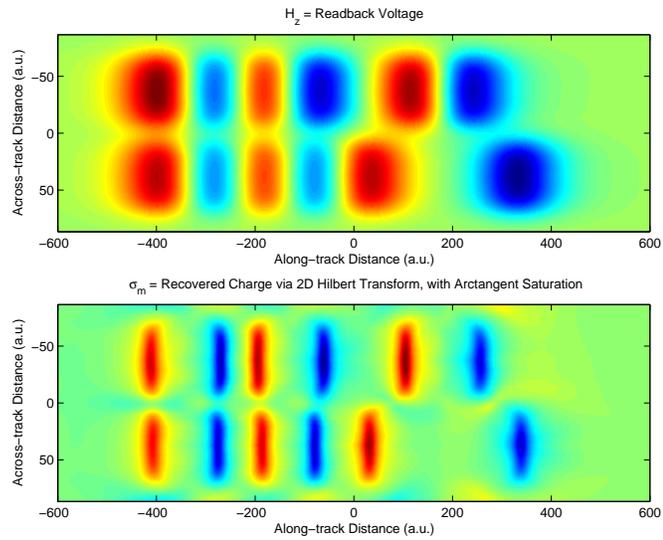


Figure 3.11: The simulated readback voltage image of $H_z(x, y)$ (top), and the 2D Hilbert transform reconstructed charge $\sigma_m(x, y)$ with arctangent saturation ($\alpha = 5 \cdot 10^{-2}$) (bottom).

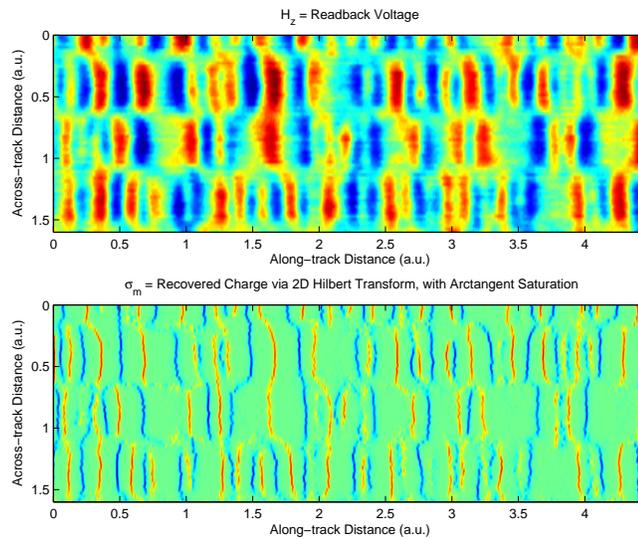


Figure 3.12: The readback voltage image of $H_z(x, y)$ (top) from a commercial disk with 60 kilotracks per inch, and the 2D Hilbert transform reconstructed charge $\sigma_m(x, y)$ with arctangent saturation ($\alpha = 15 \cdot 10^{-2}$) (bottom).

thickness h , the z -derivative of H_z can be approximated as

$$\frac{\partial H_z(x, y, z)}{\partial z} = \frac{H_z(x, y, h) - H_z(x, y, 0)}{h}. \quad (3.47)$$

Since the medium possesses flat symmetry, we know that $H_z(h) = -H_z(0)$. Therefore, at the surface of the medium,

$$\frac{\partial H_z(x, y)}{\partial z} = \frac{2H_z(x, y)}{h}. \quad (3.48)$$

Substituting this expression into equation 3.46, it can be rewritten as

$$-\nabla \cdot \vec{M} = \nabla \cdot \vec{H}_t + \frac{2H_z(x, y)}{h}, \quad (3.49)$$

where $\vec{H}_t = H_x \vec{e}_x + H_y \vec{e}_y$. Additionally, as previously discussed, only the curl-free component of magnetization produces the magnetic field sensed by the read head.

That is,

$$\vec{e}_z \cdot (\nabla \times \vec{M}) = 0. \quad (3.50)$$

Together, equations 3.49 and 3.50 form a system of two linear equations for the two unknown x and y components of $\vec{M} = M_x \vec{e}_x + M_y \vec{e}_y$ in the plane. Rewriting these equations in terms of their 2D Fourier transforms yields the following system of equations:

$$-(k_x \tilde{M}_x + k_y \tilde{M}_y) = k_x \tilde{H}_x + k_y \tilde{H}_y + \frac{2\tilde{H}_z}{h}, \quad (3.51)$$

$$k_x \tilde{M}_y - k_y \tilde{M}_x = 0. \quad (3.52)$$

Solving for \tilde{M}_x and \tilde{M}_y yields

$$\tilde{M}_x = -\frac{k_x^2 \tilde{H}_x + k_x k_y \tilde{H}_y + \frac{2k_x}{h} \tilde{H}_z}{k_x^2 + k_y^2}, \quad (3.53)$$

$$\tilde{M}_y = -\frac{k_x k_y \tilde{H}_x + k_y^2 \tilde{H}_y + \frac{2k_x}{h} \tilde{H}_z}{k_x^2 + k_y^2}. \quad (3.54)$$

It is readily seen that equations 3.53 and 3.54 give the 2D Fourier transforms of the in-plane magnetization components M_x and M_y in terms of the known quantities H_z (measured), H_x and H_y (computed from H_z via the 2D Hilbert transform).

3.3 Conclusions

The theory of the 1D Hilbert transform's application to magnetic recording has been given. Furthermore, a generalization of the Hilbert transform in two dimensions has been derived that exhibits similar properties. This 2D Hilbert transform is of interest in its own right, and examples of its use in processing spin-stand images have been given.

Chapter 4

Identification of the Hard Disk Drive Write Channel and Digital Data Reconstruction

While imaging techniques described in chapters 2 and 3 are of great use in exploring what is actually happening at the physical level of the magnetization recorded on a hard disk platter, it is also worthwhile to explore what exactly these imaged patterns of magnetization represent. This chapter, therefore, presents the developed techniques of drive-independent analysis of hard disk recording channels, with an emphasis on reversing the effects of coding and reading the actual binary data stored on the disk. These results of these techniques are then given for several disk drive products.

4.1 Overview

The modern hard disk drive (HDD) employs a large collection of integrated circuits to control the electromechanical (spindle motor, head actuator) and electromagnetic (write and read heads) devices that compose it, and to process the analog and digital electrical communication signals that are produced (coding, detection, logic interface).

This chapter will focus on the internal algorithms implemented in the mixed-signal processor generally known as the “read channel” chip, but which acts in

both the read and write processes. In the write process, this chip preconditions the digital binary stream of data by encoding it before passing it to the write current driver, which controls the write head. Because the magnetization on the disk follows the write current, this preconditioning is designed to make more tractable the process of reading and reconstituting the digital data from the readback voltage signal. The path of preconditioning the write current signal is termed the “write channel,” whereas the “read channel” is the process that effects its inverse. When operating in reading mode, the chip first translates the analog voltage to digital form (“detection”), and then removes the write channel preconditioning (“decoding”).

Each HDD or family of HDD products (e.g., a certain model line of a manufacturer) has a different read channel chip, which forms a crucial component of the total product’s performance. As such, due to the competitiveness of the industry, neither the documentation on the internal algorithms of particular chips nor their basic external connectivity (pin-out) is readily available. Therefore, in the spin-stand environment used in the present research, techniques have been developed to first identify the write channel that produced a given disk’s readback, and then, via software, to reconstruct the recorded digital data. By this means and in conjunction with centering and track following techniques [9, 10], a HDD can be opened, its disk platters extracted and placed on the spin-stand, and the recorded data subsequently deciphered.

4.1.1 Data Organization on the Disk

As a digital storage device that records on the order of hundreds of billions of bits, the modern hard disk drive (HDD) organizes its data in a hierarchical structure formed by historical precedent and practical necessity. First, almost all modern computers are *byte-addressable*; i.e., CPU instructions and memory locations are identified by means of an integral number of 8-bit bytes, which makes the byte the fundamental native unit of data manipulation and thus storage. This is a seemingly mundane and obvious fact, but its influence will be ubiquitous throughout the sophisticated signal processing discussed in this chapter that takes place in the write channel of the HDD.

Once arranged in bytes, the user data is grouped into sectors of 512 bytes, appended by any error correction code (ECC) bytes that may have been computed by the disk controller, as depicted in figure 4.1. In practice, there are always ECC bytes present to correct misdetected bits due to noise. After this total stream of bytes is encoded by the preconditioning process sketched previously, it is prefixed by a periodic sequence of bits called a “preamble” and a sequence of “synchronization” bytes. The former exist to aid detection, while the latter are necessary for decoding, as will be discussed. Together, this whole sequence forms a data sector and is written to disk. Each sector is then written in sequence on a track on the disk, together with occasional periodically spaced regions for head positioning (servo areas) containing no user data and outside the scope of the present work.

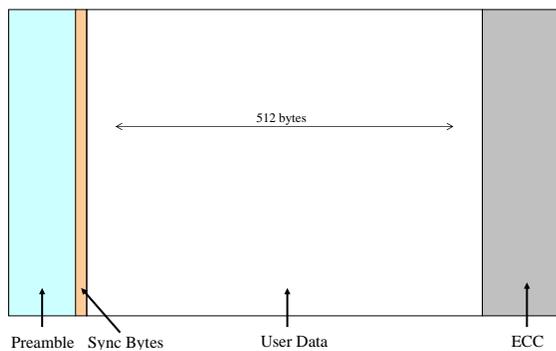


Figure 4.1: The format of the data sector.

4.1.2 Detection

Detection is the process of converting the voltage readback signal to the digital binary signal that originally produced the magnetization on the disk. Historically, this was accomplished by so-called “peak detection,” a simple threshold/zero-crossing-based presence/absence detection scheme. For peak detection to produce a low bit error rate (BER), the signal-to-noise ratio (SNR) must be high ($\geq 20dB$) and the density $D = \frac{PW_{50}}{T}$ must be low enough that ISI is negligible [31]. (Recall that PW_{50} is the full width at half maximum of the head response function — the isolated pulse voltage — and T is the bit period.)

However, cost per bit, which is correlated to area per bit, drives the industry. Therefore, HDD manufacturers want to use the highest bit density that their head and media technology will allow. For a given read head, linear density (bits per along-track distance) is proportional to D , which can be increased by decreasing the bit period T . Peak detection, however, is severely limiting in terms of density

and requires roughly $D \leq 1$ for good bit error rate BER performance [31].

Hence, modern drives (starting from the early 1990s and becoming totally dominant in the second half of the decade) use an implementation of a family of more sophisticated sampled-amplitude detection schemes called PRML, for partial-response (PR) maximum-likelihood (ML) detection [36]. In PRML, ISI is an inherent feature of the detection scheme, which allows for higher densities D . In effect, a wider head for a given bit spacing may be used, or equivalently, a smaller bit spacing may be used for a given head. The price of PRML schemes is vastly increased complexity in the design and implementation of the circuits, but the advances in the semiconductor integrated circuit industry (smaller, cheaper transistors) have made this cost feasible. In this sense, the continual innovation of the semiconductor industry has formed a fortunate synergy with the development of MR and GMR heads and together they have made possible the exponential growth of storage densities from 1990 to the present.

Before insight may be gained into how PRML works, some basic definitions of how bits are represented are in order. The two main useful representations are NRZ (non-return-to-zero) and NRZI (non-return-to-zero inverted, or “modified”). NRZ has a form like the write current or the recorded magnetization, whereas NRZI has a form like ISI-free readback voltage (the transitions in magnetization), as illustrated in figure 4.2. NRZ and NRZI are simply linear transformations of each other and are effectively equivalent. If the NRZ bits form a sequence $a_{nrz}(n)$ and the NRZI bits a sequence $a_{nrzi}(n)$, the two are related by the modulo-2 equations $a_{nrzi}(n) = a_{nrz}(n) + a_{nrz}(n - 1)$ and $a_{nrz}(n) = a_{nrzi}(n) + a_{nrz}(n - 1)$ (addition and subtraction

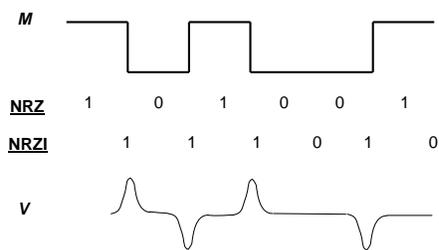


Figure 4.2: An example of NRZ and NRZI representations of bits.

are equivalent, modulo-2).

Now consider sampling the raw readback voltage signal at integer multiples of a known bit period T . If ISI corrupts the signal, then the sample value $s(n)$ obtained at each time $t_n = nT$ ($n = 0, 1, 2, \dots$) is a linear function of the past, present, and future NRZI bits due to the averaging of the head [31]. The manner in which the bits interfere with each other is determined precisely by the head response function, whose shifted linear superposition in time produces the readback voltage as discussed in chapter 1. There is a finite number of possible sample values that can be obtained by this superposition, and the exact values depend on the particular finite-length sequence of bits seen by the head at each instant. Therefore, if the head response function is accurately known, the underlying NRZI bits can be ascertained by the sequence of readback samples. If the readback samples are corrupted by additive white Gaussian noise, a ML detector in the form of the Viterbi algorithm is the optimal means to estimate the bit sequence from the readback samples [45]. This is the most basic picture of sampled amplitude detection and is useful for schematic purposes, but it is not readily usable by itself.

An equivalent and easier to use (causal) relationship between the bit pattern and the sample values is given by the *channel response function* G , which relates the present sample value to the present and L previous NRZ bits. With no ISI, the present sample value is proportional to the present bit; with ISI, only a contiguous sequence of $L + 1$ bits affects the present sample value [32, 33]. The channel response function is related to the head (“impulse”) response function as the dibit response, i.e., the head response function superimposed with itself at an adjacent bit period. Thus, the channel response function can be thought of as the voltage response in time of the system to the NRZ bits “...000010000...” or to the NRZI bits “...0000110000...”. (The head response, on the other hand, is the voltage response in time to the NRZI bits “...000010000...”). For instance, for a head response function $R(t) = \text{sinc}(t) + \text{sinc}(t - T)$, the channel response function is $G(t) = \text{sinc}(t) - \text{sinc}(t - 2T)$. These are the responses for the “PR4” shape [46], and are special (due to the orthogonality of the sinc function) in that they are zero at most $t = nT$. For real heads, the responses look more like $R(t) = \frac{1}{a^2+t^2}$ and $G(t) = \frac{1}{a^2+t^2} - \frac{1}{a^2+(t-T)^2}$, whose samples are not so compact in support.

The crucial fact alluded to here is that the actual channel response function produced by the read head may be too long (L too large) to be computationally usable [33]. The complexity of the Viterbi algorithm, which performs the ML detection and is matched to the response, goes like 2^L . Furthermore, generally as L increases, so does the number of possible sample values, which makes the necessary timing and gain recovery processes during sampling more difficult. An especially important subtlety lies in the fact that L is not just a function of the degree of ISI

(linear density) but the *shape* of the head and channel responses, as evidenced by the sinc and Lorentzian examples just given above.

Therefore, the central idea of partial response signaling is to equalize or map the actual channel response function (and head response) into a target or desired response function that has ISI spanning a more tractable number of bit periods $m < L$ [33]. In fact, this is the meaning of the term “partial” in PR; the full (actual) response is not used, but rather one that is similar to it yet having only “part” of the nonzero channel response samples. Kretzmer’s 1966 paper [46] contains several partial response target choices. The fourth item in the table looks roughly like the longitudinal magnetic recording channel response, hence the name and choice of the PR4 family in magnetic recording. PR4 has a density of 1.65 and $m = 2$. Simple extensions exist to PR4, which incorporate more ISI and therefore have higher densities (EPR4 has density of 2 and $m = 3$, and EEPR4 has density of 2.31 and $m = 4$), at the cost of higher ML detector complexity [31]. Recently, conceptually similar “generalized” partial response targets with better performance in the presence of noise have entered use [32, 69].

The diverse topics pertaining to PRML schemes and their implementation — including choice of the partial response target and how equalization is performed by means of a frequency-selective filter with gain — fill volumes and are not germane to the present discussion. Nevertheless, there are two salient features of PRML in general that directly affect the write channel. These are timing and gain recovery and the Viterbi detector path memory.

First, the necessity of timing and gain recovery during sampling puts con-

straints on the recorded binary sequence and prohibits certain patterns. In the actual HDD environment, the read process deviates significantly from the ideal picture of the head flying at a constant speed and a uniform distance above a perfectly flat medium, all in the absence of noise. The velocity of the head is controlled by the spindle motor, which for reasons of cost and mass production is not the highest quality motor available and even with feedback control still exhibits speed fluctuations. Due to the aerodynamic structure of the head, a change in velocity alone will affect the flying height, but this is further exacerbated by the roughness of the medium and contaminants such as dust, which perturb the head trajectory. Although several particle filters are placed to isolate the HDD enclosure from the outside, actual collisions of the head with the medium or contaminants are frequent enough that read channel chips are designed with purpose-built thermal asperity detection capability [34, 35]. (Thermal asperity is an effect in the MR readback voltage when such a collision occurs.) Flying height variations and thermal asperity affect the amplitude and baseline (DC level) of the readback signal. Finally, electronic, media, and head noise corrupts the voltage readback to some extent, which results in the readback deviating from the perfect convolution model by the addition of a series of noise terms.

Due to the combination of these deviations, the readback signal as a function of time is altered from one that has known predictable values at a fixed series of sampling instants. Instead, a clock must be generated from the readback signal itself, and the slowly varying phase and frequency deviations of this clock must be tracked and compensated with a phase-locked loop (PLL). The preamble written at

the beginning of every sector exists precisely for the purpose of deriving the initial sampling phase. Like all communication systems, this timing (or “clock”) recovery system is necessary so that synchronization is maintained and sampling occurs at the correct positions in the signal. With timing recovery, samples are obtained at the time instants $t_n = nT + \tau_n$, where τ_n is a phase correction computed by the PLL from the previous timing error. Due to the varying strength of the signal, gain control must also be performed. This scales the samples $s(n)$ by a computed factor γ_n based on the previous sample error. The particular algorithms used for timing and gain recovery (e.g., [47, 48, 49]), are not important here, but the feature shared by all such schemes is that zero-value samples yield no information to the PLL. From the timing perspective, a time-shifted zero sample is still zero. From the gain perspective, a scaled zero sample is still zero. For this reason, to maintain the correct sampling and gain, non-zero samples must occur frequently, where the minimum frequency of occurrence is a channel design parameter. It is quite simple to create a pattern that has an insufficient number of transitions. The trivial case is a sector containing all zeros; there is no way to “count” the number of zeros without externally provided timing information, which is unavailable. For this reason, a mapping of the user bits is necessary to artificially introduce more non-zero content.

The second feature of PRML detection schemes that affects the write channel is the finite survivor-path memory in hardware implementations of the Viterbi ML detector. ML detection is also known as “sequence” detection because it estimates the most likely sequence of bits based on a sequence of received samples. This estimation is performed on a trellis, which describes every possible sequence of bits in

terms of the corresponding sequence of sample values [31]. By comparing the actual samples received with all of the possible “ideal” sample values given by the channel response function, several competing “most likely” bit sequences are considered. (There are 2^m different competing sequences for a given channel response.) For each possible competing sequence, an accumulated error is computed as successive samples are received. When the most recent estimated bits of two competing sequences are identical, which can only occur when the most recent sample is *non-zero*, the sequences are said to have “converged.” When this occurs, the accumulated errors for each sequence are compared and the sequence with the larger accumulated error is eliminated. As samples are received, competing bit sequences converge frequently enough that, given enough time, only one bit sequence remains at some delayed number of bits in the past (while the most recent bits are still “undecided”). This means that this part of the bit sequence has been fully detected and therefore the Viterbi detector needs to store it in its path memory no longer. For this reason and for circuit implementation considerations (cost, finite chip area), a finite-length memory is built into the Viterbi detector. (The size of this memory is dependent upon m ; larger m require a longer path memory.) However, competing sequences will not converge within this path memory if too many consecutive zero samples are received. If the number of zero samples exceeds the path memory, then the detector catastrophically fails and the bits are lost. The detector will also fail given certain rare periodic input patterns which similarly prevent convergence of competing sequences [31]. Therefore, further restrictions of the written pattern are needed to combat these inherent issues with the Viterbi detector.

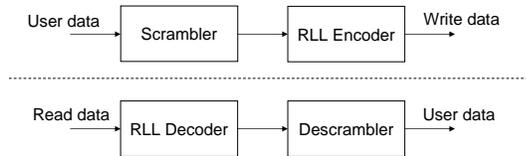


Figure 4.3: Block diagram of the two-layered coding process in write (top) and read (bottom) modes.

These features of the PRML detection process thus require preconditioning to the binary data before it is written to the disk, and this preconditioning is effected by two layers of coding which will be discussed in the following sections of this chapter. The first of these is run-length-limited (RLL) coding, and the second is scrambling. In the write process, scrambling is first applied to the data, and then RLL encoding maps the scrambled sequence into another one before passing it to the write driver. The read process reverses the order, with RLL decoding coming first and scrambling proceeding thereafter, as depicted in figure 4.3.

An example of the effects of these two layers of coding in a particular commercial hard disk product is shown in figures 4.4 and 4.5, which depict their effects on two very different user data patterns, bits from a JPEG image (left) and a pattern of all “0” bits. In the top plots, the channel bits (resulting from both scrambling and RLL coding) are shown. In the middle plots, the RLL coding is removed and only the effects of scrambling are shown. The bottom plots show the original user data bit sequences. It is obvious that the scrambling and RLL encoding steps successfully transform these different deterministic data sequences into similar patterns with particular statistical properties.

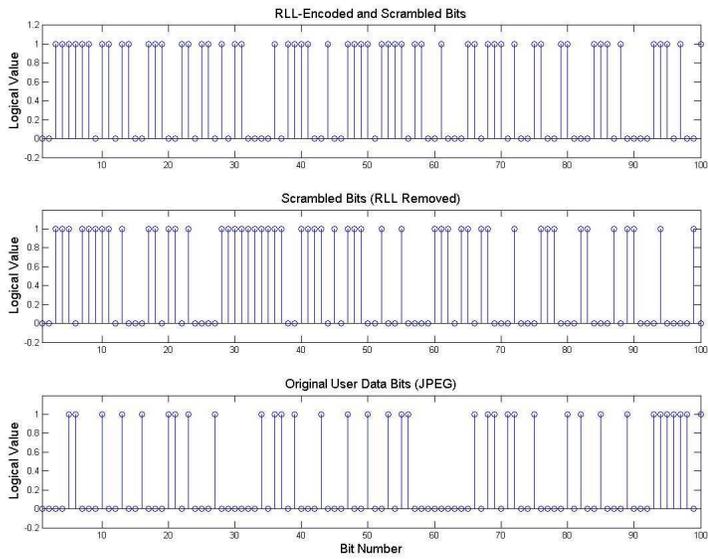


Figure 4.4: The effects of a two-layered coding scheme on a part of a sector from a JPEG file.

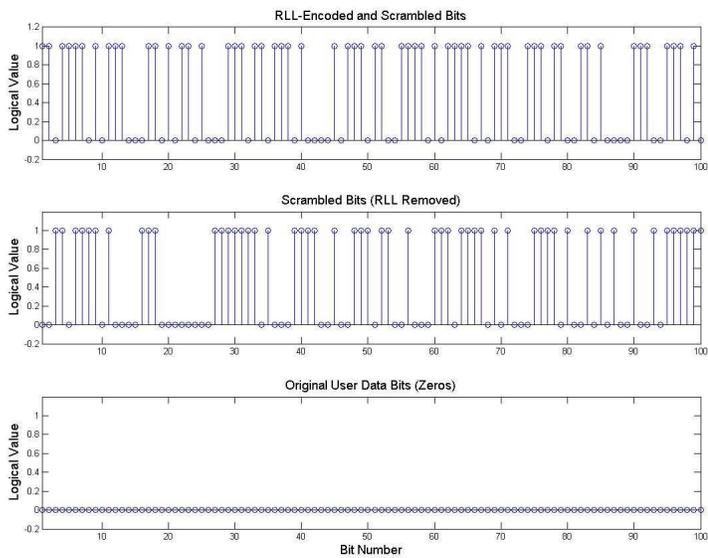


Figure 4.5: The effects of the same two-layered coding scheme as figure 4.4 on a constant all “0” bit sequence.

4.2 Run-Length-Limited (RLL) Coding

4.2.1 Background

RLL coding maps an arbitrary binary sequence of length m to a binary sequence of length n with a controlled number of logical zeros between each logical one (in the NRZI sense). For this reason, it is known as a *constrained* code and is defined by its rate m/n and its constraint (d, k) , where d denotes the minimum number of zeros between two consecutive ones, and k denotes the maximum number of zeros between consecutive ones. For information theoretic reasons beyond the scope of this work, the rate m/n must be less than unity, which implies that the output of an RLL encoder is longer in length than the input. The subsequent effective loss of recording density is the price paid for the necessity of introducing run length constraints. RLL codes are also, by nature, nonlinear codes. However, their implementation is straightforward and is achieved by means of digital combinatorial logic or by table look-up for both encoding and decoding.

The minimum run length constraint d was used in the past to limit ISI in the readback. Clearly, with more zeros between neighboring transitions, there will be less interference between them. This led to the widespread use of the now obsolete codes like $1/2$ (2,7) and $2/3$ (1,7), used in read channel chips with peak detection [39, 40]. With more modern PRML schemes, the need for limiting ISI is eliminated, and thus almost every RLL code in common use for the last decade has used $d = 0$; i.e., there is no longer a minimum run length constraint. Another modification that PRML has imposed on RLL codes relates to a simplification of the Viterbi detector

hardware implementation. It can be shown that in the PR4 Viterbi detector, the even and odd samples/bits form independent sub-sequences and therefore can be processed (detected) separately [36]. This “deinterleaved” implementation is more efficient than processing the sequence in the standard manner, and similar simplifications can frequently be made for higher order PR targets like EPR4 [37, 38]. Therefore, the maximum run length constraint k is typically split into k_G/k_I , where k_G is the global run length constraint and k_I is the run length constraint for each (odd, even) interleave. The reason for the constraint on the interleaves’ zero run lengths is as discussed previously. Since the interleaves are processed independently through separate Viterbi detectors, there must be non-zero samples occurring frequently enough so that the competing estimated interleaved bit sequences converge. The global run length constraint on the total bit sequence exists for clock and gain recovery, which has nothing to do with the interleaved detector.

For instance, in the 8/9 (0, 4/4) RLL scheme popular in the mid-1990s [41], the input bit sequence is split into 8-bit words, which are successively mapped to 9-bit words. The resulting bit sequence is 9/8 times longer, and there are no more than 4 zeros between successive ones. Furthermore, when the odd and even interleaves are formed, the number of zeros between successive ones in each interleave does not exceed 4. Examples of other codes include 8/9 (0, 3/6) [41], 16/17 (0, 6/8) [42], 16/17 (0, 12/8) [43], and 32/34 (0, 9/9) [44], among many others. Modern RLL schemes typically take an integer multiple of bytes as input (in increments of 1, 2, and 4 bytes in the examples just given) because the data itself is stored as an integer multiple of bytes.

In order to tell the read channel when the RLL encoded bytes start in the detected data, the write channel employs synchronization or “sync” bytes. These mark the beginning of RLL encoding and are not generally known but must be deciphered if sectors are to be decoded. An important characteristic of these sync bytes is that they are RLL encoded but *not* scrambled.

4.2.2 Identification Techniques

Identification of the RLL encoding scheme entails determining the mapping of the unconstrained bit sequence to the constrained bit sequence. Because RLL codes are nonlinear, it is not possible to characterize their operation by means of a linear transformation (matrix equation, convolution, Fourier transform, etc.). This fact alone makes analysis much more difficult than it would be if RLL were linear. However, it might seem possible that a series of known user data patterns could be passed to an unknown “black box” RLL encoder and the subsequent outputs analyzed and tabulated to form at least a partial codebook by brute force. Unfortunately, this (very expensive) possible technique is frustrated by the fact that a second layer of coding (scrambling) exists underneath the RLL encoding that is *a priori* unknown and whose own identification requires flawless RLL identification beforehand; thus, the actual input to the RLL encoder is not generally known even if the underlying user data is.

Therefore, the best technique to identify the RLL encoding scheme without recourse to high-power cryptographic machinery is statistical analysis and subse-

quent correlation with published specifications. By analyzing the statistics of an RLL-encoded bit sequence, an accurate estimation of both the code rate m/n and the constraints $(d, k_G/k_I)$ can be made and doing so does not require knowledge of the input to the RLL encoder.

Having accurately detected the raw encoded bits, the statistical analysis of the constraints is a straightforward process of computing a histogram for the number of zeros between consecutive ones. By looking at the maximum number of zeros obtained in a number of different sectors containing different underlying user data, a lower bound on the value of k_G is obtained. Likewise, the sequence of encoded bits can be deinterleaved and the odd and even interleaves analyzed likewise to obtain a lower bound on k_I . Because the input to the RLL encoder for each sector is scrambled, it is approximately random (as will be discussed later). For this reason and because each sector contains on the order of several hundred codewords (256 data codewords for a 16/17 scheme, 512 for a 8/9 scheme), the aforementioned lower bounds quickly converge to an accurate estimate of the actual maximum run length constraint as the number of sectors analyzed increases. In practice, looking at less than five sectors is sufficient to obtain a good estimate of the constraints. For example, figure 4.4 shows the histograms obtained for a sector of 4096 random bits which have been encoded with the 8/9 (0, 4/4) RLL scheme to produce a constrained bit sequence of length 4608. This histogram shows how frequently different runs of zeros occur (on a common log scale). If the same random data sequence is encoded with an 8/9 (0, 3/6) encoder, different statistics are produced (figure 4.5).

On the other hand, the determination of the code rate m/n is less precise

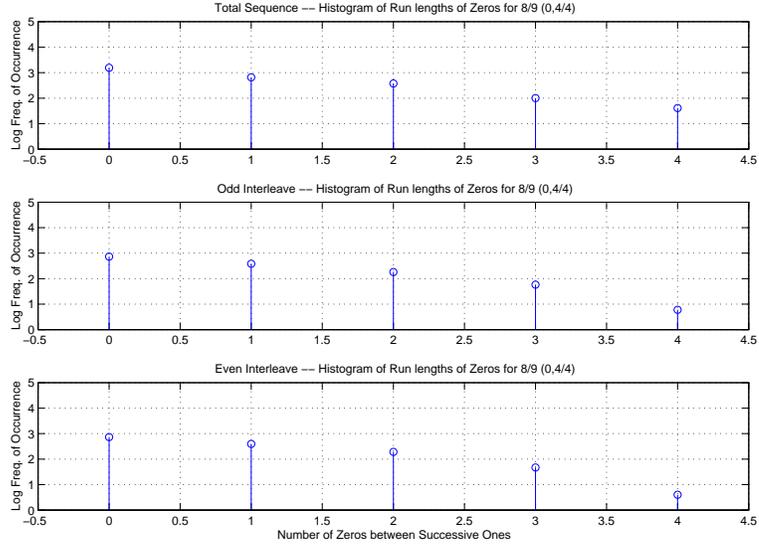


Figure 4.6: RLL 8/9 (0, 4/4) histograms of the zero run lengths for the total encoded sequence (top), odd interleave (middle), and the even interleave (bottom).

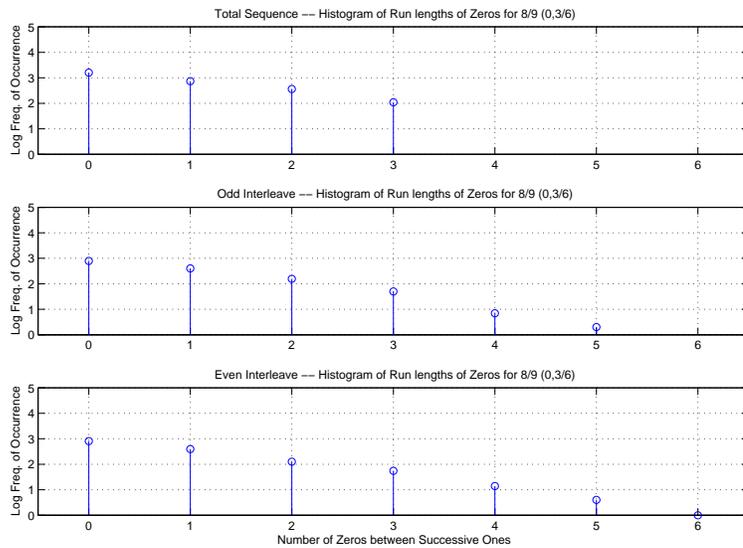


Figure 4.7: RLL 8/9 (0, 3/6) histograms of the zero run lengths for the total encoded sequence (top), odd interleave (middle), and the even interleave (bottom).

because the exact structure of the sector is not known beforehand. Therefore, M , the length of the unencoded bit sequence, must be an estimate. The length of the encoded bits, N , is readily obtained as the length of the detected bits minus the length of the preamble (readily identifiable by its periodicity). However, whereas the unencoded data section is known to be 4096 bits (512 bytes) in length, the length of the synchronization bytes and the length of the ECC bytes (figure 4.1) are not known until both the RLL and scrambler have been fully identified.

Nevertheless, a good estimate of the code rate is possible in practice. Consider a sector in which there are an estimated 540 unencoded bytes (28 ECC and sync bytes), or 4320 bits. If an 8/9 encoder were used, the length of the encoded sequence would be 4860 bits, whereas if a 16/17 encoder were used, the length of the encoded sequence would be 4590 bits. This is a difference of 270 bits or approximately 34 bytes, which is a significantly large difference in lengths, and allows a likewise large range of misestimation of the unencoded sequence length. Generally, the code rate can be narrowed down to a small number of possibilities based on the date of the HDD's manufacture; e.g., there are no 8/9 disks in 2001, and there are no 48/52 disks in 1996. Likewise, the number of ECC bytes tends to increase with year (as densities increase and SNRs drop), so the number of non-data bytes in the sector can be estimated to the requisite precision based on the year of the drive's manufacture.

Once a code rate and constraints have been estimated, the catalog of documented RLL schemes can be consulted. Because an RLL encoding scheme is such a crucial component of the HDD write channel, when a scheme is developed (by a drive or channel maker) it is typically registered with the US, European, or Japanese

patent offices. Sometimes containing ambiguous information, a patent on an RLL scheme is nevertheless crucial because it gives the logic equations and mapping tables necessary for implementing the encoder and decoder. With this information, the RLL decoder can be implemented in software and applied to the detected bits. It is not known for certain if the RLL scheme has been correctly identified, however, until the next layer of coding, the scrambler, has been identified as well. Furthermore, sometimes several different codes exist with the same $m/n(d, k)$ due to the non-uniqueness of the mapping between uncoded and coded words. When this is the case (e.g., $2/3(1,7)$ has several different realizations), proper identification requires iteration through the possibilities until the correct scheme is found.

4.3 Scrambling

4.3.1 Background

In order to prevent periodic bit patterns (and readback samples) from interfering with the detection process, the HDD write channel employs linear scrambling to randomize bit sequences before RLL encoding. Scrambling is a means of producing a pseudo-random sequence based on linear feedback and has long been an attractive randomization method in communications devices due to its simplicity of implementation [50, 51]. In hardware, a shift register (which shifts its contents in one direction each clock cycle) with the proper feedback connections is all that is needed. Unlike RLL, scrambling does not increase the length of the bit sequence, but merely maps it into a different sequence of the same length. Although scram-

bling does randomize bit sequences, thereby eliminating most long runs of zeros (and ones), it nevertheless does not give the same level of control over run lengths that RLL gives. Therefore, RLL encoding is typically still needed and both layers are used in HDD write channels.

The linear feedback in a scrambler is defined by a modulo-2 polynomial

$$h(D) = 1 + \sum_{k=1}^N h_k D^k \quad (4.1)$$

where D is a unit delay and the coefficients h_k are binary [51]. (In z-transform notation, $D = z^{-1}$.) Equivalently, the feedback is defined by the sequence h_k for $k = 1, 2, \dots, N$ (the polynomial coefficients), where N is the order of the scrambler. With the scrambling polynomial defined, the output scrambled bits y_n are related to the input unscrambled bits x_n by the modulo-2 equation

$$y_n = x_n + \sum_{k=1}^N h_k y_{n-k}, \quad (4.2)$$

for $n = 0, 1, 2, \dots, L$ where L is the length of the bit sequence to be scrambled. Because the current output is a function of the current input and the output at the previous N times, there must be an “initial state” $\{y_{-1}, y_{-2}, \dots, y_{-N}\}$ defined in order to compute first N outputs y_0, y_1, \dots, y_{N-1} . The basic principle of operation of the scrambler is simple: the previous output bits are summed in a special way such that when they are then added to the current input, successive output bits are uncorrelated. Such an output bit sequence is called “pseudo-random.” In order to see why it is merely “pseudo” random requires further discussion of the scrambling polynomial $h(D)$.

In order to transform the input bits into a pseudo-random bit sequence, the polynomial $h(D)$ must be “primitive” [51]. A primitive polynomial is an irreducible polynomial (i.e., cannot be factored), and has an odd number of nonzero coefficients (including the mandatory D^0 term in equation 4.1). Additionally, a primitive polynomial is one that produces the longest possible scrambled output given a constant input (all zeros or all ones) before repeating. It can be readily observed from equation 2.2 that if x_n is constant for all n , the output y_n is only a function of y_{n-k} for $k = 1$ to N . There are only 2^N different combinations of y_{n-k} ; thus, the output sequence *must* repeat itself eventually. This occurs when y_{n-k} becomes the initial state again at some $n = M$. When $h(D)$ is primitive, $M = 2^N - 1$ (the “maximal length”), which means y_{n-k} cycles through every possible combination of bits [51]. This is a unique property and holds only for primitive polynomials, which exist for every order n [51]. For example, two tenth-order primitive polynomials are $h(D) = 1 + D^3 + D^{10}$ and $h(D) = 1 + D^7 + D^{10}$, which yield $M = 1023$.

Nevertheless, the fact remains that M is finite and given a constant input, the output sequence will repeat eventually. For this reason, the scrambling is deemed “pseudo” random and the result of scrambling a constant input is called a “pseudo-noise” sequence. This periodic structure, however, provides a useful property. The autocorrelation $R(n)$ of a pseudo-noise sequence is a constant C at all values except at $n = aM$ (a an integer), where it has a peak with value $2C$ [50, 51]. This is illustrated for a scrambled all-zeros sequence with $h(D) = 1 + D^3 + D^{10}$ in figure 4.6.

Descrambling the scrambled bit sequence is effected by simple rearrangement

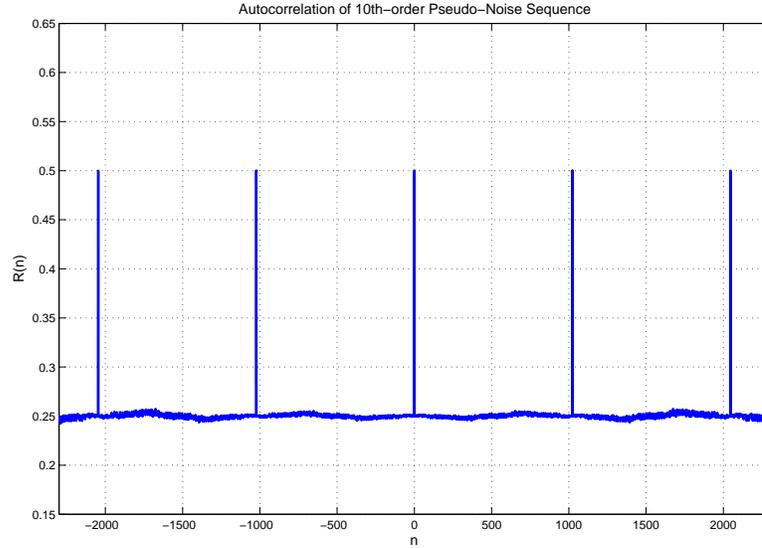


Figure 4.8: Autocorrelation of a pseudo-noise sequence produced by a 10th order scrambler.

of equation 4.2. Because addition and subtraction are identical modulo 2, the descrambling relation is

$$x_n = y_n + \sum_{k=1}^N h_k y_{n-k}, \quad (4.3)$$

where the quantities are defined as before. This descrambler relation is called “self-synchronizing” because the initial state $\{y_{-1}, y_{-2}, \dots, y_{-N}\}$ affects only the first N descrambled output bits. Thus, if the first N bits are not needed, the initial state of the scrambler need not be known for descrambling the rest of the bits. This fact is important for identifying an unknown scrambler, but ultimately, the initial state must be obtained to descramble all the data in a HDD sector (as will be discussed in the next section).

The preceding discussion has been phrased with an implicitly understood

structure of scrambling and descrambling a sequence of consecutive bits (bit 1, bit2, etc.). In actuality, the HDD write channel typically uses an 8-bit interleaved structure. That is, bits 1, 9, 17, 25, etc. form one interleave, while bits 2, 10, 18, 26, etc. form another interleave, and so forth for a total of eight interleaves. Each interleave is scrambled/descrambled separately with the same polynomial, in parallel. Equivalently, this means the write channel scrambles/descrambles by *bytes*. Each byte can be thought of as a vector of 8 bits and the above equations can be applied with the understanding that the arithmetic operations are performed element-wise and the indices refer to bytes and not bits.

Research has shown that when this byte-wise scrambler is used in a HDD write channel, it is furthermore augmented by a second layer of processing. In the scrambling process, the user data is first *descrambled* (with initial state all ones) before being scrambled (with another initial state). Likewise, in the descrambling process, the scrambled data is first descrambled, and then scrambled (with initial state all ones). The block diagrams for these steps are shown in figure 4.8. An important feature of this second stage is that all constant-valued interleaves are unaffected beyond the first N bytes, and interleaves containing all ones are completely unaffected for all n . Substituting all ones for the initial state and the descrambler input y_n into equation 4.3 yields the following relation for all n :

$$x_n = 1 + \sum_{k=1}^N h_k = 1 + 0 = 1 \quad (4.4)$$

where the fact that h_k is binary and has an even number of nonzero values from $k = 1$ to N was used. A similar relation can be shown for a constant all zeros

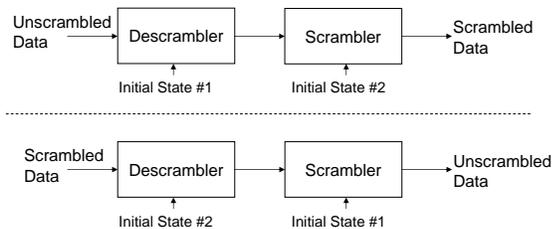


Figure 4.9: Block diagram of the two-layered scrambling process in write (top) and read (bottom) modes.

pattern, $x_n = 0$ for $n > N$ (regardless of the descrambler initial state). This fact is important for identifying an unknown scrambler, as will be discussed in the next section.

4.3.2 Identification Techniques

With the assumption that the RLL encoding has been correctly characterized and removed, to identify the scrambler used in a given unknown write channel requires first solving for the polynomial $h(D)$ and subsequently the initial state. With knowledge of the values of both the scrambled and unscrambled bytes and exploiting the knowledge that they are related through a linear mapping, the polynomial is obtained. Due to the two-stage structure described above which is typically found in HDD write channels, a constant input (all zeros or all ones) is chosen as the known input data for identification. Choosing a constant-valued input also assists in gaining information about the order of the scrambler polynomial, for if it is small enough that its period N is achieved in one sector, the autocorrelation of the result-

ing pseudonoise sequence shows the peaks at multiples of N seen in figure 4.7.

Solution of the scrambling polynomial proceeds as follows. Rewriting the scrambler equation 4.2 for $n = 0, 1, \dots, N - 1$ gives the following:

$$\begin{aligned}
 y_0 &= x_0 + h_1 y_{-1} + h_2 y_{-2} + \dots + h_N y_{-N} \\
 y_1 &= x_1 + h_1 y_0 + h_2 y_{-1} + \dots + h_N y_{-(N+1)} \\
 &\vdots \\
 y_{N-1} &= x_{N-1} + h_1 y_{N-2} + h_2 y_{N-3} + \dots + h_N y_{-1}
 \end{aligned} \tag{4.5}$$

Again, in the HDD write channel, these relations hold true in a byte-wise sense. To simplify the following discussion, consider the quantities in 4.5 to be the bits of a single interleave of the byte sequence (e.g., the first bit of each byte). The same scrambler polynomial holds for each interleave, so values from any can be used in solving for h .

Due to the finite memory of the scrambler, quantified by its order N , any $2N$ consecutive values of scrambled data y_n can be used to solve for h . Therefore, since the initial state is unknown at this stage, values from the middle of the sector can be used because the effect of the initial state is present only on the first N positions of the scrambled output. For the sake of discussion and without loss of generality, the assumption can be made that y_n is known for $n = -N$ to $N - 1$ and x_n is known

for $n = 0$ to $N - 1$. Rewriting 4.5 in matrix form yields

$$\begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_{N-1} \end{pmatrix} = \begin{pmatrix} y_{-1} & y_{-2} & \cdots & y_{-N} \\ y_0 & y_{-1} & \cdots & y_{-(N-1)} \\ \vdots & \vdots & \ddots & \vdots \\ y_{N-2} & y_{N-3} & \cdots & y_{-1} \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_N \end{pmatrix} + \begin{pmatrix} x_0 \\ x_1 \\ \vdots \\ x_{N-1} \end{pmatrix}, \quad (4.6)$$

or

$$\vec{y} = \hat{Y}\vec{h} + \vec{x}. \quad (4.7)$$

Thus, obtaining the polynomial coefficients h_k amounts to solving this modulo-2 matrix equation, where the matrix \hat{Y} is Toeplitz; i.e., $\vec{h} = \hat{Y}^{-1}(\vec{y} + \vec{x})$. There exist several techniques for modulo-2 matrix inversion (e.g., [53]).

Once the scrambler polynomial has been obtained, the initial state can be directly solved. By definition, $h_0 = h_N = 1$. Furthermore, assuming that x_n , and y_n are known for $n = 0, 1, \dots, N - 1$, simple rearrangement of equation 4.2 yields

$$\begin{aligned} y_{-1} &= x_{N-1} + \sum_{k=1}^{N-1} h_k y_{(N-k-1)} \\ y_{-2} &= x_{N-2} + \sum_{k=1}^{N-1} h_k y_{(N-k-2)} \\ &\vdots \\ y_{-N} &= x_0 + \sum_{k=1}^{N-1} h_k y_{(N-k-N)}, \end{aligned} \quad (4.8)$$

which is the initial state. Unlike the solution of the polynomial, this process must be repeated for each bit within the byte, i.e., for each interleave, to obtain the entire byte-wise initial state.

4.4 Experimental Results

Once both the RLL and scrambler have been determined, the digital user data can be reconstructed. Additionally, once the full encoding is understood, the structure of the sector can be fully ascertained as well. This includes the value and number of the sync bytes and the ECC bytes. Several generations of HDD write channels from different manufacturers have been fully identified and the data from their disks deciphered.

The simplest of these cases is a drive from Manufacturer A produced in 1993. This is a pre-PRML drive using peak detection, and as such, there is no scrambling and only RLL. The RLL scheme was found to be a $2/3$ (1,7) scheme by IBM [40]. The sync bytes, which are RLL encoded and precede each sector, were found to have decimal values 165 and 182 (hexadecimal A5 and B6). Four ECC bytes were found to be appended to each sector (which will be discussed further in the following chapter).

The earliest PRML write channels that were identified were from a family of drives by Manufacturer B from 1996/1997 employing the IBM 8/9 (0,4/4) RLL code [41] and a 10th order scrambler of the structure described in figure 4.8. The scrambler polynomial was found to be $h(D) = 1 + D^3 + D^{10}$, with initial state bytes (in decimal) {63, 158, 8, 27, 28, 65, 89, 54, 253, 247}. Employing the same read channel chip but operating in a different mode is a drive from Manufacturer C of the same era. The RLL and scrambling are the same as in the Manufacturer B drives, but the sync bytes are different. In the disks from Manufacturer B, the sync

bytes (in decimal) are {147, 147, 147, 147, 105}, whereas the Manufacturer C drive's sync bytes are {147, 31, 147, 105}. Frequently, read channel producers produce read channels with features that can be tuned to the needs of particular customers (HDD manufacturers), and this appears to be the case with this read channel. The number of ECC bytes used in the Manufacturer B HDDs of this family is 18.

Another channel identification from a drive of the same era but different manufacturer was a Manufacturer D HDD (from 1997). This drive uses a 16/17 (0, 12/8) RLL scheme [43] and a 12th order scrambler with polynomial $h(D) = 1 + D^5 + D^8 + D^9 + D^{12}$. This scrambler has the structure depicted in figure 4.8 and uses the initial state bytes (in decimal) {222, 69, 244, 3, 102, 216, 254, 239, 140, 48, 226, 250}. The sector sync bytes (in decimal) were found to be {112, 129, 182}, and 28 ECC bytes are appended to each sector.

A more modern family of HDDs from Manufacturer B (1998/1999 and 2000) have been found to employ the 16/17 (0, 6/8) RLL scheme [42] and use the same 10th order scrambler polynomial as in the 1996/1997 products, $h(D) = 1 + D^3 + D^{10}$. However, they employ a different scrambler initial state, {127, 60, 16, 55, 56, 131, 178, 108, 251, 239} (in decimal). The sync bytes were found to be {145, 160, 145, 177, 145, 160, 145, 3} and a total of 24 ECC bytes were found.

In total, this successful identification of a variety of write channels is a thorough proof of concept for the identification techniques described in this chapter. Coupled with software-implemented PRML detection and the ability to place the spin-stand read head accurately on the recorded disk tracks [9, 10], this identification allows the spin-stand to be used to reconstruct large quantities of digital data (files) from

hard disk platters removed from their native environment.

4.5 Conclusions

The layers of coding (RLL, scrambling) present in the HDD write channel exist to render detection possible. The methods for completely characterizing the particular implementations used in a given write channel have been presented and some results have been given to illustrate their effectiveness. Although the particular manifestations of the write channel continue to change as recording densities increase, the fundamental structure (namely, the necessity of breaking up periodicity in written bit sequences) does not vary significantly and the described identification techniques cover a wide range of drive families.

Chapter 5

Error Correcting Code (ECC) Identification

5.1 Overview

Due to the presence of noise in the readback signal and the sheer volume of data stored on a modern HDD, the detection process cannot statistically guarantee a bit-error rate (BER) sufficient for large scale recording use. Typically, $BER = 10^{-6}$ at the output of the detector (roughly 1 bit error per 100 kilobytes), but $BER = 10^{-12}$ (at the worst) is desired for greater reliability (roughly 1 bit error per 100 gigabytes). ECC is necessary to achieve this, and it does so by appending redundant data bytes which are intimately connected to the user data. Unlike RLL coding and scrambling, which are performed by the read channel chip and which directly modify the user data, the ECC is computed outside the read channel (typically at the disk controller) and merely adds additional data to the end of each sector. (This method of appending the error correcting bytes is known as “systematic” ECC structure.) As far as the read channel is concerned, ECC bytes are data and as such are scrambled and RLL encoded before being written to disk. When operating in read mode, the read channel first detects the data, removes RLL encoding, descrambles, and then passes the user data and ECC back to the disk controller, which checks for errors and corrects any that exist (if possible).

The objective of any ECC scheme is to increase the Hamming distance d

between different data sequences (effectively, the number of differing symbols) by cleverly selecting the values of these redundant bytes. Depending on the actual coding scheme used, small deviations from the actual user data may be detected (quantified by a small d), and sometimes fully corrected with the ECC. The error detection theorem states that for $d < d_{min}$, the minimum distance inherent to the ECC scheme, errors can be detected, and the error correction theorem states that for $d < d_{min}/2$, errors can be corrected [52]. Different codes exist with varying detection and correction abilities and different levels of overhead (degree of redundancy).

The ECC schemes used in hard disk drives are a subset of linear block codes called cyclic codes. Cyclic codes are valuable because they, like the pseudo-random scramblers discussed before, can be easily implemented with binary shift register circuits in hardware. Historically, Fire codes and computer-generated codes were used that operated on the user data as a single block of 4096 bits [54]. Their more powerful cousins, Reed-Solomon codes, have been predominant since the early 1990s. In HDDs, Reed-Solomon ECC in HDDs operates on the user data as several interleaved sub-sequences of *bytes* to provide detection and correction for multiple bursts of errors. Bursts, or connected strings of errors, are the most common error phenomenon in the HDD read process for two reasons. First, errors most frequently occur due to phenomena like media noise, thermal asperity, clock cycle slips, etc., which affect short contiguous lengths of the readback and not sporadic bits throughout. Second, the RLL decoding process has a tendency to make single bit detection errors into single byte decoding errors (although RLL schemes are typically designed to prevent single bit errors from ballooning to multiple bytes). By operating on bytes and not

bits, a single-byte correcting Reed-Solomon code can nominally correct a burst of 8 bit errors. By interleaving, the length of the correctable error is multiplied by the number of interleaves, as will be discussed.

Before discussing the particular details of Reed-Solomon codes, some definitions are required for understanding cyclic codes in general. Sequences of symbols (bits, or bytes for Reed-Solomon) are known as “words.” The “message” word of length k is the input data for which the ECC is computed. The “codeword” of length n is the message word appended with the ECC symbols, giving the nomenclature “ (n, k) ” to describe a cyclic code. A defining feature of cyclic codes is that any cyclic shift of a valid codeword for a given cyclic ECC scheme produces another valid codeword for that particular scheme. A word may be represented by a vector or sequence of symbols, but an equivalent form is to represent it with a polynomial in X , where the power of X denotes the position of a symbol in the word and the coefficients are the symbols. For example, in binary, the word “10011” may be represented as the polynomial $X^4 + X + 1$. (This notation is, in fact, identical to the polynomials in D mentioned in the section on scramblers but with X instead.)

Another feature of cyclic codes is that a generator polynomial $g(X)$ of order $r = (n - k)$ exists that produces the ECC symbols from the message word $m(X)$. The ECC symbols represented by $c(X)$ (also order r) are the remainder of the polynomial division $\frac{X^r m(X)}{g(X)}$. With **mod** denoting modulus, this is equivalently rewritten

$$c(X) = X^r m(X) \mathbf{mod} g(X). \quad (5.1)$$

These r ECC symbols are appended to the end of the sector data, or in polynomial

notation, the codeword $w(X) = c(X) + X^r m(X)$. An important note is that in hardware, the ECC symbols are produced by shift registers whose initial contents must be non-zero; the content of the shift registers before encoding is the initial state. In effect, the initial state is some arbitrary non-user data that is encoded before the message and which must be known by the ECC decoder.

In general, the capabilities of a given code vary with the properties of the generator polynomial and symbol size; typically, a code is created for a particular codeword length and desired error correction length. For instance, a particular Fire code polynomial operating on bit sequences is $g(X) = (X^{13} + 1)(X^{35} + X^{23} + X^8 + X^2 + 1)$, with degree 48, giving 48 ECC bits and correcting error bursts up to 7 bits long [54].

Reed-Solomon codes, on the other hand, have a particular structure and well defined set of capabilities. They have $d_{min} = r + 1$, the largest minimum distance possible for cyclic codes, which means that with r (even) ECC symbols, r errors can be detected and $r/2$ errors can be corrected in a codeword [52]. Reed-Solomon codes in HDDs operate on byte-sized symbols, but in general, they may operate on symbols in any binary Galois field extension $GF(2^m)$, where $m \geq 1$ ($m = 8$ for byte-sized symbols), and have a maximum codeword length of $2^m - 1$ symbols. A Galois field is a finite set of elements (symbols) that is closed under standard algebraic operations (addition, subtraction, multiplication, division) and for which the additive and multiplicative identities exist [55]. The simplest binary Galois field is that of binary numbers, $GF(2)$, with elements 0 and 1. In $GF(2^8)$, the elements are bytes, which may be written in “tuple” representation as their corresponding

bits; e.g., decimal 255 is 11111111.

Equivalently, the symbols of a Galois field can be represented as powers of the “primitive element,” denoted α , along with the zero and unit elements; i.e., the field is the set $\{0, 1, \alpha, \alpha^2, \dots, \alpha^{2^m-2}\}$. Because a Galois field $\text{GF}(2^m)$ has only $2^m - 1$ non-zero elements, the field can be constructed by a binary primitive polynomial $p(X)$ of order m (see section 4.3.1). The primitive element α is defined as a root of this constructing polynomial. All other elements of the field are produced by addition and multiplication by α and 1, using the fact that $p(\alpha) = 0$. An important detail is that the realization of a Galois field is not unique — e.g, several different primitive polynomials typically exist for a given m — but each of these realizations is related to the others by an isomorphism and is effectively equivalent. Furthermore, due to the finite number of elements and the periodicity inherent to the primitive polynomial that connects them, there is a periodicity in the powers of α . For example, in $\text{GF}(2^8)$, the relations $\alpha^{-1} = \alpha^{254}$, $\alpha^0 = \alpha^{255} = 1$, $\alpha^1 = \alpha^{256}$, etc. hold.

To explain how arithmetic in a Galois field works, consider constructing $\text{GF}(2^8)$ with the particular 8th order primitive polynomial $p(X) = X^8 + X^4 + X^3 + X^2 + 1$. In any realization of $\text{GF}(2^8)$, the zero element or additive identity is defined as “0” (00000000 in tuple representation) and the unity element or multiplicative identity is defined as “1” or α^{255} (00000001 in tuple representation). Addition and subtraction are performed modulo-2, element-wise in the tuple representation. Let α in tuple representation be given as 00000010, α^2 as 00000100, and so on, up to α^7 as 10000000. Then by using the fact that $p(\alpha) = 0$, it is known that $\alpha^8 + \alpha^4 + \alpha^3 + \alpha^2 + 1 = 0$, which gives $\alpha^8 = \alpha^4 + \alpha^3 + \alpha^2 + 1$ (00011101). The

rest of the elements may be generated by multiplying further by α and substituting previously known elements where appropriate, e.g., $\alpha^9 = \alpha\alpha^8 = \alpha^5 + \alpha^4 + \alpha^3 + \alpha$ (00111010). Element-wise addition frequently simplifies the expressions, e.g., $\alpha^{12} = \alpha^4\alpha^8 = \alpha^8 + \alpha^7 + \alpha^6 + \alpha^4$, which in tuple form gives $00011101 + 10000000 + 01000000 + 00010000 = 11001101$. This process continues up until α^{254} , at which point the entire Galois field has been constructed.

The Reed-Solomon generator polynomial $g(X)$ for HDDs exists in $\text{GF}(2^8)$ and produces the ECC bytes with equation 5.1 understood in $\text{GF}(2^8)$ arithmetic. The structure of $g(X)$ for Reed-Solomon codes has the general form

$$g(X) = (X - \alpha^j)(X - \alpha^{j+1}) \cdots (X - \alpha^{j+(r-1)}), \quad (5.2)$$

where $r = n - k$ is the number of ECC symbols as defined before and j is some arbitrary integer. Thus, the generator polynomial's roots are consecutive powers of the primitive element α . The initial state of the ECC encoder also exists in $\text{GF}(2^8)$, i.e., as bytes.

As briefly mentioned previously, the capability of Reed-Solomon codes is enhanced in HDDs by employing an interleaved structure. Interleaving means that several separate Reed-Solomon codewords are formed from the sector data, and the bytes of these words are not consecutive but interleaved, as shown in figure 5.1. Because these codewords are separate, an error spanning two consecutive symbols is, in effect, two independent single errors and not a double error in one codeword. Thus if r error bytes can be detected in a single interleave, Nr consecutive error bytes can be detected with N interleaves. This structure further increases the burst

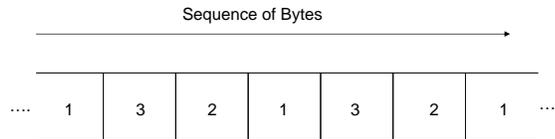


Figure 5.1: Interleaving of ECC codewords, the example here shown with three interleaves.

detection and correction abilities of the ECC scheme. Furthermore, a Reed-Solomon $GF(2^8)$ codeword is limited to a length of 255 bytes, whereas HDD data sectors have 512 bytes of data (not counting the ECC symbols). Since at least three codewords are necessary to cover a length of more than 512 bytes, if given the choice between a consecutive and interleaved structure, interleaving is clearly preferable for its greater performance.

5.2 Identification Techniques

The identification of a ECC scheme is, like scrambler identification, greatly facilitated by the linearity of the code. As a linear code, any linear combination of valid codewords is also a valid codeword. Furthermore, like the scrambler, the main features to identify are the (generator) polynomial and initial state. However, unlike the scrambler, the entire codeword must be ascertained first because the ECC symbols are the remainder of the division of the whole codeword. (Recall that for the scrambler, only a small subset of the whole scrambled sequence was needed to identify the polynomial and initial state.)

The central idea used in identifying the generator polynomial is the fact that

$$g(X) = X^r \bmod g(X) + X^r, \quad (5.3)$$

where $m(X) = 1$ has been substituted into equation 5.1. This relation holds for any linear cyclic code, regardless of the size of the symbols. (Thus, $m(X) = 1$ is understood for Reed-Solomon codes to be the multiplicative identity element of $\text{GF}(2^8)$.) The message word $m(X) = 1$ implies that all zeros with a “1” at the end should be written to the data sector. The ECC symbols, then, are the coefficients of the generator polynomial corrupted by the initial state. To recover the polynomial then requires removing the effect of the initial state.

The problem of removing the initial state without knowing the polynomial is solved by performing a two-step process. First, a sector with $m(X) = 0$ (all zeros data) is written. If the initial state were zero, by the linearity of the encoding process, the ECC symbols would also be zero, yet this is never the case. Thus, the ECC symbols produced by an all-zeros data sector are the initial state itself. Second, to thus remove the corrupting effect, the initial state is subtracted from the read ECC symbols for $m(X) = 1$ to give $g(X)$.

A variation on this technique is necessary when $m(X) = 1$ is not directly obtainable. This is the case when there are other non-data, non-ECC bytes in the sector not directly under the user’s control, such as cyclic redundancy check (CRC) bytes or unknown address information. In this variation, it is possible to find two codewords whose difference gives the desired pattern $m(X) = 1$. This is useful because the linear combination of any two codewords is itself a codeword.

Performing this subtraction has the further effect of removing the effect of the initial state because it is common to both sectors. Thus, when two sectors with the same ECC initial state are subtracted and the message $m(X) = 1$ is produced, the ECC symbols thus resulting are precisely the generator polynomial $g(X)$. To then recover the actual (non-zero) initial state, the message word of any sector is encoded with equation 5.1, and this result subtracted from the actual sector codeword's ECC symbols.

For Reed-Solomon codes, which in HDDs are used in the interleaved structure previously discussed, the identification techniques must be used on the interleaves. Each interleave forms its own codeword and has the same generator polynomial. Therefore, any interleave may be used to solve for $g(X)$, but in general the initial states of the interleaves are different and must be solved separately. The fact that there is an interleave structure furthermore implies that the number of interleaves itself must be identified. As previously mentioned, there must be at least three interleaves, and the actual number N used can be easily guessed from the total length of the sector, which must be evenly divisible by N , and the total number of ECC symbols.

Once the polynomial and initial state have been identified, the ECC scheme may be used to detect and correct errors. In particular, Reed-Solomon decoders are well known and several ready-to-use implementations exist in software, e.g., in Octave Forge [56] and Matlab's communications toolbox [57]. For non-Reed-Solomon codes, general decoding algorithms exist such as "error-trapping" [55].

5.3 Experimental Results

To illustrate the effectiveness of the identification techniques, schemes from drives of several different years have been successfully characterized.

The first scheme successfully identified is from the disk by Manufacturer A from 1993. This drive could have been one of the last mainstream HDDs in production to still use a bit-wise cyclic code, a 32-bit computer-generated code with polynomial $g(X) = X^{32} + X^{28} + X^{26} + X^{19} + X^{17} + X^{10} + X^6 + X^2 + 1$. Subsequent research has found that this is called the “Glover 140A0443” code [39]. In terms of bits, the initial state was found to be $\{1, 1, 1, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0\}$. Grouping these 32 bits into bytes gives the initial state (in decimal) $\{240, 69, 10, 192\}$.

The next scheme identified is a Reed-Solomon code from the 1996/1997 family of Manufacturer B’s HDDs. In these drives, the sector has three interleaves. Using the realization of $\text{GF}(2^8)$ constructed by the primitive polynomial $p(X) = X^8 + X^4 + X^3 + X^2 + 1$, the 6th order generator polynomial (in tuple form) $g(X) = (00000001)X^6 + (10010001)X^5 + (01000111)X^4 + (01011100)X^3 + (00000010)X^2 + (10110011)X + (00111010)$ was found. In α notation, $g(X) = X^6 + \alpha^{165}X^5 + \alpha^{253}X^4 + \alpha^{131}X^3 + \alpha X^2 + \alpha^{171}X + \alpha^9 = (X - \alpha^{-1})(X - \alpha^0)(X - \alpha^1)(X - \alpha^2)(X - \alpha^3)(X - \alpha^4)$. Thus, it is seen that the generator polynomial has roots that are consecutive powers of the primitive element, as expected. The initial state for the first interleave was found to be $\{\alpha^{224}, \alpha^{75}, \alpha^{27}, \alpha^{223}, \alpha^{26}, \alpha^{25}\}$, the second interleave $\{\alpha^{100}, \alpha^{199}, \alpha^{238}, \alpha^3, \alpha^{50}, \alpha\}$, and the third interleave $\{\alpha^4, \alpha^{104}, \alpha^{51}, \alpha^{198}, \alpha^2, 1\}$. In

tuple notation, with the tuple bits represented as a decimal number, this gives the following incremental-looking initial states for the three respective interleaves: $\{18, 15, 12, 9, 6, 3\}$, $\{17, 14, 11, 8, 5, 2\}$, $\{16, 13, 10, 7, 4, 1\}$. This ECC scheme can detect 6 errors per interleave (an 18 byte error burst across all three interleaves) and correct error bursts as long as 9 bytes using all three interleaves.

Finally, another Reed-Solomon scheme was deciphered from the 1998/1999 and 2000 drive families of Manufacturer B. Again, there are three interleaves in the sector and using $p(X) = X^8 + X^4 + X^3 + X^2 + 1$ to construct $\text{GF}(2^8)$, the 8th order generator polynomial $g(X) = (00000001)X^8 + (01111011)X^7 + (11110111)X^6 + (00001110)X^5 + (10001011)X^4 + (00011100)X^3 + (11111011)X^2 + (11111111)X + (00010000)$ was found. In α notation, $g(X) = X^8 + \alpha^{172}X^7 + \alpha^{232}X^6 + \alpha^{199}X^5 + \alpha^{237}X^4 + \alpha^{200}X^3 + \alpha^{234}X^2 + \alpha^{175}X + \alpha^4 = (X - \alpha^{-3})(X - \alpha^{-2})(X - \alpha^{-1})(X - \alpha^0)(X - \alpha^1)(X - \alpha^2)(X - \alpha^3)(X - \alpha^4)$. Therefore, again, the generator polynomial has roots that are consecutive powers of the primitive element. In fact, the polynomial is identical to the previous case but with the addition of two more consecutive powers. The initial state for the first interleave was found to be $\{\alpha^{28}, \alpha^{141}, \alpha^{224}, \alpha^{75}, \alpha^{27}, \alpha^{223}, \alpha^{26}, \alpha^{25}\}$, the second interleave $\{\alpha^{129}, \alpha^{52}, \alpha^{100}, \alpha^{199}, \alpha^{238}, \alpha^3, \alpha^{50}, \alpha\}$, and the third interleave $\{\alpha^{239}, \alpha^{14}, \alpha^4, \alpha^{104}, \alpha^{51}, \alpha^{198}, \alpha^2, 1\}$. In tuple notation, with the tuple bits represented as a decimal number, this gives the previous case with an extension of two elements for each of the three respective interleaves, $\{24, 21, 18, 15, 12, 9, 6, 3\}$, $\{23, 20, 17, 14, 11, 8, 5, 2\}$, $\{22, 19, 16, 13, 10, 7, 4, 1\}$. This ECC scheme can detect 8 errors per interleave (an 24 byte error burst across all three interleaves) and correct error bursts as long as 12 bytes using all three interleaves.

5.4 Conclusions

Being able to use the ECC bytes that are appended to each data sector is useful in the context of large-scale digital data reconstruction. Without using the ECC redundancy, reading any sizeable file without errors would be statistically improbable. Therefore, techniques were developed to recover information about the ECC schemes used in hard disk drives. In particular, the identification of an ECC scheme is greatly assisted by its linear structure.

Chapter 6

Spin-stand Magnetic Microscopy Techniques for Perpendicular Recording

In this chapter, the techniques involved in spin-stand imaging will be described for the area of perpendicular recording and their results will be exhibited. The specific novelty of perpendicular recording will be described and the resulting new problems for spin-stand imaging will be presented with their solutions.

6.1 Background and Overview

The basic idea of perpendicular recording is not new and has been understood for decades [58, 59, 76]. Nevertheless, consumer products have not employed perpendicular recording until recently, with (for instance) the appearance of new commercial perpendicular drives in 2005. Previously, it was more cost effective for the industry to leverage its existing manufacturing and design experience with longitudinal recording to increase storage density than to solve the myriad engineering issues related to bringing perpendicular recording out of the laboratory. These problems include significant novelties such as new write head and media designs as well as modifications to the channel detection and decoding implementations [61]-[64], [67]-[76].

However, the leap has been taken into commercial perpendicular magnetic

recording because longitudinal recording densities have increased recently to such an extent that the fundamental superparamagnetic limit has been approached. The exact definition of this limit is dependent upon material parameters (e.g., composition, thickness) and signal processing (detection) SNR performance, but 100 gigabits per square inch (Gb/in²) is the nominal value understood as the fundamental areal density limit for longitudinal recording. The superparamagnetic effect states that if a magnetic grain has a small enough volume, its orientation may be perturbed by room temperature thermal energy [60]. Thus, at the superparamagnetic limit, bits can become no smaller, lest they lose their stability and randomly flip polarity on a short time scale due to random thermal fluctuations. Such random fluctuations in bit values clearly destroy the medium's ability to accurately store data. To achieve greater storage density therefore requires a different means of storing the recorded magnetization, and the means chosen by the magnetic recording industry is perpendicular recording. Perpendicular recording is also limited by the superparamagnetic effect, but its projected maximum areal density is 500 Gb/in² to 1 Tb/in² [61]- [63], five to ten times the longitudinal limit.

Furthermore, practical considerations make such high density longitudinal recording less attractive compared with the perpendicular alternative. As density increases in longitudinal recording, the locations of magnetic charges become closer together, which makes the demagnetizing field produced between the charges ($\propto 1/r^2$) greater in magnitude, as illustrated in figure 6.1. Since the demagnetizing field opposes the remanent magnetization, a recording medium with a higher coercivity H_c is required for remanence. A higher coercivity, in turn, requires a

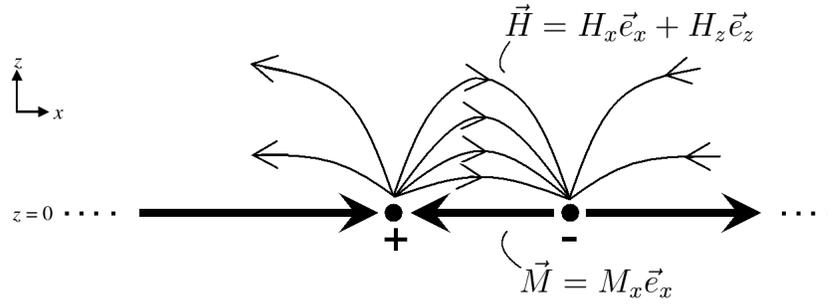


Figure 6.1: Sketch of the longitudinal recording media, showing the intensification of the in-plane (horizontal) component of the external magnetic field as the longitudinal magnetization charge spatial density increases.

larger write field to record the magnetization, which implies a larger write current. Furthermore, this larger write current must be switched faster in time to achieve the desired higher linear spatial density (assuming the same rotational velocity), thereby significantly increasing the power requirement for writing to disk. The difficulty of writing in longitudinal recording is further exacerbated by the fact that it is only the fringing component of the field produced by the ring-like write head (recall figure 2.1) that interacts with the longitudinally oriented medium, while the main direct component remains in the head’s air gap. Thus, it is desirable if these issues can be avoided or otherwise mitigated.

In contrast, in perpendicular recording, magnetization is written not parallel and along the medium (“longitudinally”) but oriented normal (“perpendicularly”) to the medium surface. Recalling equation 2.1 (defining magnetic charge as the negative divergence of magnetization), it is obvious that when magnetization is oriented perpendicularly, the charge appears on the upper and lower boundaries of the

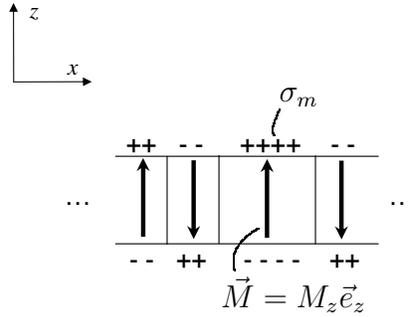


Figure 6.2: The recording plane (x - z) with perpendicular magnetization

recording medium. The resulting magnetic charges are thus arranged such that the medium thickness and not the density controls the magnitude of the demagnetizing field, as shown in figure 6.2. In order to write magnetization perpendicular to the medium, a different write head design employing a direct write field (more efficient than a fringing field) is required [63]. This write head design, known as a single-pole head, has a structure shown in figure 6.3 [65, 66]. The salient features of this design are two in number: a narrow write pole which directs the magnetic flux created by the write-current coil, and a wide return pole, which gathers the flux returning to the head, thereby completing the magnetic circuit. The single write pole itself is compact in order to increase the magnetic write flux density under the point where writing occurs, while the return pole has a large area to decrease the flux density where the flux returns to the head, so as not to disturb the underlying previously recorded regions of magnetization. Additionally, the write flux gradient can be increased by the use of side shields of high magnetic permeability, as described in [64] and illustrated in figure 6.4. Due to their proximity and high permeability, such shielding around the narrow write pole captures those components of the write flux

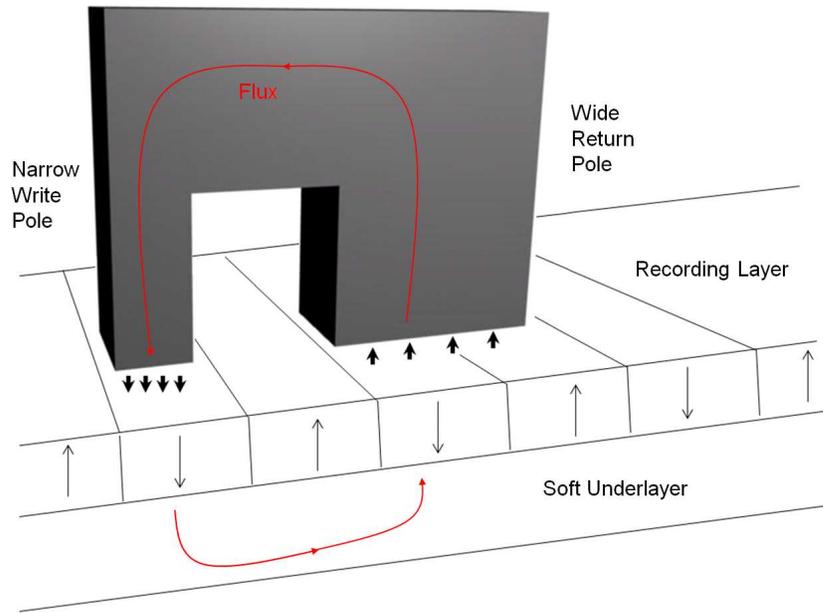


Figure 6.3: Perpendicular recording write head and medium design with soft underlayer.

not directed perpendicular to and directly under the pole face. Since an objective of write head design in perpendicular recording is to maximize the magnetic write flux in the gap between the pole face and the recording medium, this shielding provides a crucial performance enhancement.

Together with this new head design is a complementary new media design that incorporates a magnetically soft layer underneath the magnetically hard recording layer itself (see figure 6.3). While the layer of the medium that exhibits magnetic remanence is desired to be magnetically hard (having a large coercivity H_c), the underlayer possesses both a high magnetic permeability μ and a small coercivity H_c so as to act as an efficient guider of magnetic flux. Therefore, the soft underlayer provides a low-reluctance return path for the write flux during the write process.

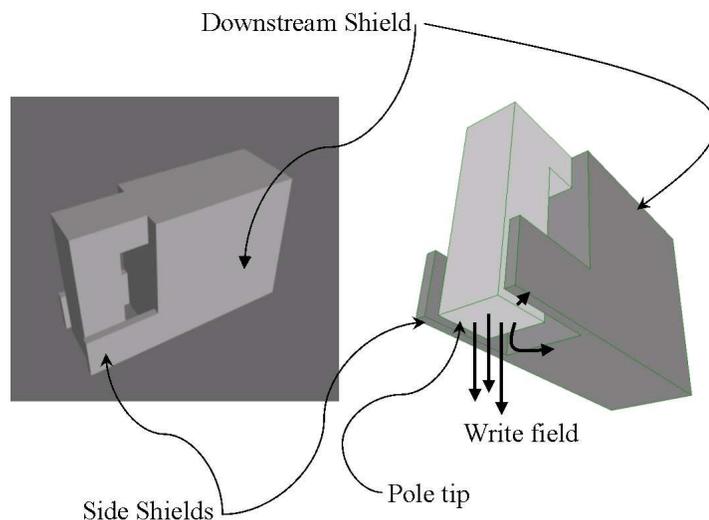


Figure 6.4: Perpendicular recording write head employing the side shielding design described in [64].

Flux leaves the narrow pole of the write head, impinges upon the perpendicular recording medium, which orients itself according to the polarity of the perpendicular field, and then reaches the soft underlayer, which channels the flux back to the wide return pole. Furthermore, the presence of the soft underlayer increases the theoretical maximum perpendicular write field over that in longitudinal recording. This can be explained by the fact that the high magnetic permeability of the underlayer allows its boundary to be considered as a zero equipotential. Therefore, the medium sees the presence of an “image” write head positioned on the opposite side of the underlayer, which adds its equal effective flux to that of the actual write head. (Although, in practice, the enhancement is less than a factor of two [76]). Nevertheless, in addition to the use of the direct write field is employed for recording, the soft underlayer in perpendicular magnetic recording medium further enhances

perpendicular recording's writing efficiency over that of longitudinal recording.

Regardless of these differences in write head and recording media design, the read head employed in perpendicular recording is still the GMR sensor (or a functionally similar advancement like a tunneling MR [TMR] sensor) that has been used in longitudinal recording. Since a shielded MR readback head senses the vertical component of the demagnetizing field, in perpendicular recording the actual magnetization is sensed (up to a sign) and not its transitions, as in longitudinal recording. This fact has obvious implications for imaging perpendicular media. Due to the novelty of perpendicular recording in commercial products, it is of great interest to pursue imaging via the spin-stand technique in this context [13], [77], [78].

6.2 Response Function Characterization for the Read Head in Perpendicular Recording

Just as in longitudinal recording, the notion of the read head's response function is important for understanding readback and imaging in perpendicular recording. This function (as discussed in chapter 2) describes the manner and extent to which the read head senses not just charge directly underneath it, but neighboring regions of charge as well. Although the read head sensor itself is the same in perpendicular recording as it is in longitudinal recording, the manner in which the read head can be characterized presents itself differently due to the several novel features of the different recording mode. While the impulse response function characterization obtained for a head in a longitudinal recording context (see chapter 2)

is still descriptive of the head's non-local sensing abilities, it does not *fully* describe its performance in the context of reading perpendicular media. Additionally, the method of measuring the head response function is not the same as in perpendicular recording for similar reasons. In essence, the definition of the head response function differs between longitudinal and perpendicular recording because the nature of the recorded magnetic charge, the resulting magnetic field emanating from the medium, and therefore the signals produced by the read head in perpendicular recording are all significantly different. These novel characteristics make a direct impulse response characterization of the read head impossible, suggest the importance of the step response function, and alter the definition of head response itself.

6.2.1 Background

The media design and mode of perpendicular recording imply particular characteristics for magnetic charge distributions. As discussed in chapter 2, the read head senses the vertical component of the external magnetic field emanating from the recording medium surface. In the idealized situation of a perfectly shielded, point-like head with negligible vertical separation from the medium surface (i.e., perfect local sensing) and the existence of *only* surface charge, the read head exactly senses the surface charge and the readback signal perfectly reproduces the sensed charge. In figure 6.2, depicting the distribution of charge for perpendicularly oriented magnetization, it is obvious that the head will sense only two different charge values, either positive or negative magnetic charge. There are only two possible

directions of magnetization in the perpendicular recording medium, up and down, which produce these regions of positive and negative magnetic charge, respectively, on the top surface. As in longitudinal recording, these charges produce a demagnetizing magnetic field with a polarity opposing the direction of magnetization.

However, in longitudinal recording, the magnetization is oriented in the plane of recording, which produces charge only at the changes (transitions) in magnetization direction and therefore three different charge situations – positive charge (negative transition), negative charge (positive transition), and no charge (no transition). In longitudinal recording, an isolated charge is read back by the head as an isolated voltage pulse on a zero voltage background. In contrast, scanning the head in the context of the idealized perpendicular recording read process yields a two-level signal (a square wave in 1D), with the level of the read voltage signal denoting either positive or negative charge (or magnetization). Therefore, unlike in longitudinal recording, it is impossible to read a region of isolated charge on the perpendicular medium because the medium itself does not allow isolated surface charge. Since the idea of measuring the read head impulse response is based on reading this isolated charge (an approximate “impulse”), the impulse response cannot be directly measured from perpendicularly recorded media in the same way that it was done for longitudinal recording.

Furthermore, it is notable that in perpendicular recording, a second layer of charges is present that contributes to the field sensed by the read head. These charges, opposite in polarity to the surface layer, exist by virtue of the normal orientation of the magnetization (equation 2.1). The presence of the high permeability

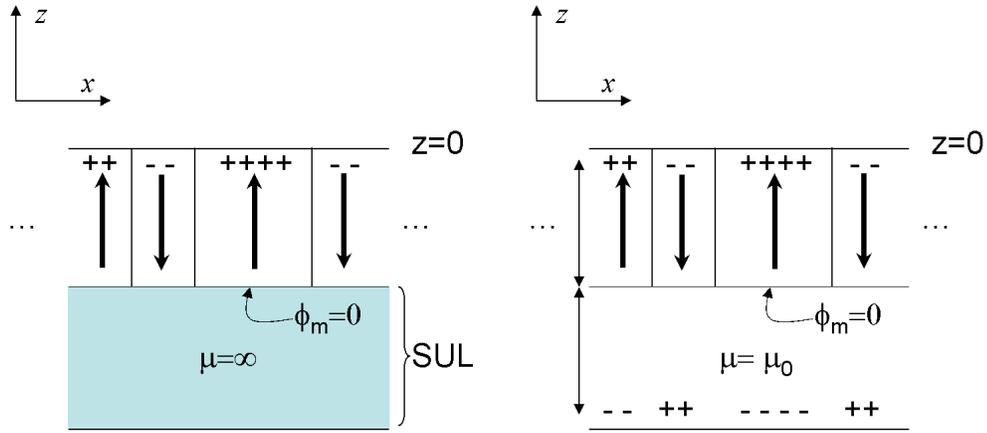


Figure 6.5: (left) Perpendicular recording media with soft underlayer and (right) its effect on the charge distribution.

soft underlayer modifies the distribution of these charges so that they form an image charge layer on the opposite side of the underlayer boundary (see figure 6.5) because, as noted in the previous section, this boundary can be treated as a zero equipotential. As an image of the surface charge distribution, this layer is opposite in sign but otherwise identically distributed as the surface layer such that the zero potential boundary condition is maintained. Therefore, due to this geometry of the perpendicular charge distribution, an “isolated” charge in perpendicular recording also does not exist in this sense, because every surface magnetic charge is accompanied by its image below the soft underlayer.

Hence, the nature of the perpendicular recording medium must modify what is meant by “response function.” While in longitudinal recording an isolated charge could be imaged and called an impulse response, in perpendicular recording, all isolated charges come in pairs. The proximity of the read head to the surface

charge layer implies its dominance in defining the readback, but the second layer of charges contributes to the readback as well. Therefore, any “response function” in perpendicular recording is understood to be the response to both layers of charge.

6.2.2 Definition of the Response Function

Understanding the perpendicular response function as the function that measures the non-local sensing characteristics of the read head due to both the surface charge and its image below the soft underlayer, a definition of the impulse response can be made that is similar to longitudinal recording. That is, by the ideas of linear superposition and translational invariance, the 2D spin-stand perpendicular readback voltage image is defined as

$$V(x, y) = \int_S \sigma_m^S(x', y') R_\delta(x - x', y - y') dx' dy', \quad (6.1)$$

where $V(x, y)$ is the raw image, $\sigma_m^S(x, y)$ is the surface charge distribution, and $R_\delta(x, y)$ is the impulse response due to both the surface and lower layers of charge. This definition holds due to the linear properties of the readback image and can be shown as follows. The image can be defined as a linear superposition of both the surface and lower charge distributions as

$$V(x, y) = V_S(x, y) + V_L(x, y), \quad (6.2)$$

in which

$$V_S(x, y) = \int_S \sigma_m^S(x', y') R_S(x - x', y - y') dx' dy', \quad (6.3)$$

and

$$V_L(x, y) = \int_S \sigma_m^L(x', y') R_L(x - x', y - y') dx' dy', \quad (6.4)$$

where $\sigma_m^S(x, y)$ is the surface charge distribution, $\sigma_m^L(x, y)$ is the lower charge distribution, $\sigma_m^S(x, y) = -\sigma_m^L(x, y)$, and $R_S(x, y)$ and $R_L(x, y)$ are the impulse responses to isolated charges at the surface $z = 0$ and the lower layer $z = -L$, respectively. Rewriting equation 6.2 with the knowledge that the surface and lower charge distributions are opposite and equal yields the equation

$$V(x, y) = \int_S \sigma_m^S(x', y') (R_S(x - x', y - y') - R_L(x - x', y - y')) dx' dy'. \quad (6.5)$$

Therefore, an effective response due to both the upper and lower charge distributions can be defined,

$$R_\delta(x, y) = R_S(x, y) - R_L(x, y). \quad (6.6)$$

Substituting equation 6.6 into equation 6.5 gives the definition in equation 6.1; therefore, for a given read head, this function fully describes the non-local sensing of the head. The perpendicular impulse response $R_\delta(x, y)$ is thus the readback image $V(x, y)$ when the upper charge layer is a two-dimensional impulse function,

$$\delta(x, y) = \delta(x)\delta(y). \quad (6.7)$$

An alternate method of characterizing the response function of the head for spin-stand images of perpendicular media is via the step response. It can be shown that the raw image can be described as

$$V(x, y) = \int_S \sigma_m^S(x', y') \frac{\partial^2}{\partial x \partial y} R_u(x - x', y - y') dx' dy', \quad (6.8)$$

where $R_u(x, y)$ is understood as the step response function [13]. That is, $R_u(x, y)$ is the readback image when the surface layer of charge is a 2D unit step $u(x, y)$ (and

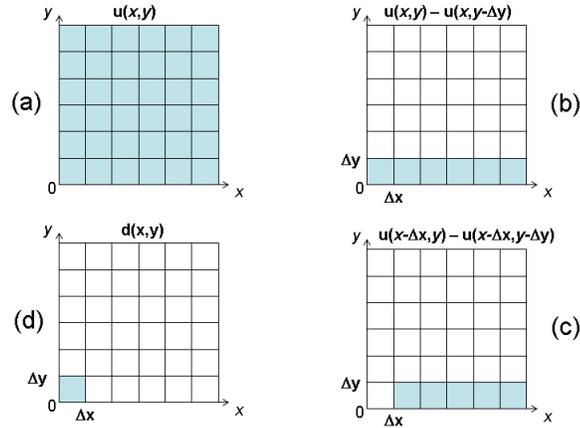


Figure 6.6: (a) The unit step, and its (b), (c) superposition and translation to produce (d) a small region of charge.

the lower layer of charge is a negative step),

$$u(x, y) = u(x)u(y), \quad (6.9)$$

which is unity for positive x and y and zero elsewhere. That is,

$$R_u(x, y) = \int_S u(x', y') R_\delta(x - x', y - y') dx' dy'. \quad (6.10)$$

Equation 6.8 and the link between the perpendicular impulse response function and step response function via the second order mixed partial derivative can be derived by considering the region of charge shown in figure 6.6d. This small region of charge of area ΔA and value σ_m is defined as

$$d(x, y) = \begin{cases} \sigma_m & \text{if } (x, y) \in \Delta A \\ 0 & \text{otherwise.} \end{cases} \quad (6.11)$$

As shown by figures 6.6a, 6.6b, and 6.6c, equation 6.11 can be rewritten in terms of

the linear superposition of translations of the unit step as

$$d(x, y) = \sigma_m ([u(x, y) - u(x, y - \Delta y)] - [u(x - \Delta x, y) - u(x - \Delta x, y - \Delta y)]). \quad (6.12)$$

The superposition and translational invariance properties of the readback process along with equation 6.10 give the readback image due to $d(x, y)$ (positioned at the origin) as

$$\begin{aligned} \Delta V_{0,0}(x, y) = \\ \sigma_m (R(x, y) - R(x, y - \Delta y)) - \\ \sigma_m (R(x - \Delta x, y) - R(x - \Delta x, y - \Delta y)). \end{aligned} \quad (6.13)$$

If the small region of charge is placed at some arbitrary point (x_i, y_j) instead of the origin, equation 6.13 is more generally written as

$$\begin{aligned} \Delta V_{i,j}(x, y) = \\ \sigma_m (R(x - x_i, y - y_j) - R(x - x_i, y - y_j - \Delta y)) - \\ \sigma_m (R(x - x_i - \Delta x, y - y_j) - R(x - x_i - \Delta x, y - y_j - \Delta y)). \end{aligned} \quad (6.14)$$

By again using the properties of superposition and translational invariance, a general image composed of N small regions of charge distributed at various points (x_i, y_j) is described by the summation of equation 6.14 over those points,

$$V_N(x, y) = \sum_{j=1}^N \sum_{i=1}^N \Delta V_{i,j}(x, y). \quad (6.15)$$

If equation 6.15 is subsequently rewritten in terms of $R_u(x, y)$ by substituting equation 6.14 and rearranging terms, then the image $V_N(x, y)$ is given as

$$V_N(x, y) =$$

$$\sum_{j=1}^N \sum_{i=1}^N \sigma_m \left(\frac{R(x - x_i, y - y_j) - R(x - x_i, y - y_j - \Delta y)}{\Delta x \Delta y} - \frac{R(x - x_i - \Delta x, y - y_j) - R(x - x_i - \Delta x, y - y_j - \Delta y)}{\Delta x \Delta y} \right) \Delta x \Delta y. \quad (6.16)$$

Clearly, if the regions of charge described by equation 6.11 become smaller in area, then equation 6.16 in the limit that $\Delta A = \Delta x \Delta y$ approaches zero becomes

$$\begin{aligned} V(x, y) &= \lim_{\Delta A \rightarrow 0} V_N(x, y) \\ &= \int_S \sigma_m(x', y') \frac{\partial^2}{\partial x \partial y} R_u(x - x', y - y') dx' dy'. \end{aligned} \quad (6.17)$$

With $\sigma_m(x, y)$ identified as the surface charge distribution $\sigma_m^S(x, y)$, equation 6.17 indeed proves that the raw voltage image $V(x, y)$ can be described in terms of the step response function $R_u(x, y)$, and that $R_\delta(x, y) = \frac{\partial^2 R_u(x, y)}{\partial x \partial y}$, as expected.

6.2.3 Measurement of the Response Function

As mentioned in the previous sections, only regions of positive and negative charge are produced on perpendicular recording media. Because regions lacking charge cannot be created, charge distributions like the unit impulse or step functions cannot be directly recorded by a write head because they contain large regions of “zero” charge. This issue is especially notable in measuring the impulse response function, for which the charge distribution is an isolated region of positive charge surrounded by an (approximately) infinite region of no charge.

On the other hand, an indirect measurement of the impulse response function can be performed by using the linear superposition property of the readback imaging process (equation 6.1). In this technique, two different regions of charge are written,

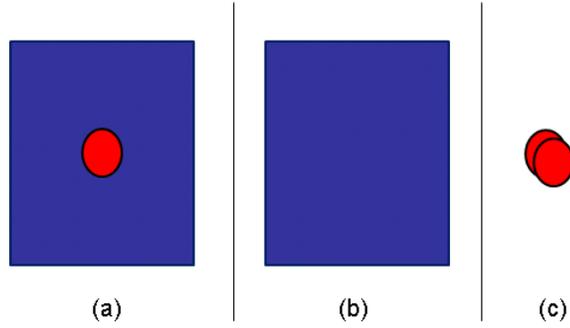


Figure 6.7: Charge distributions for the measurement of the impulse response function on a perpendicular medium via superposition, with blue denoting negative charge regions, red denoting positive charge, and (c) showing a double positive charge.

imaged, and subsequently superimposed to produce the desired charge distribution [13]. Figure 6.7 shows one such set of charge distributions. In figure 6.7a, an isolated positive charge is surrounded by a larger region of negative charge. Such a charge distribution could be produced in the same manner as the isolated charge region written for the longitudinal impulse response function measurement described in section 2.2 (with DC-erased trimming in the across-track direction). It is clear that the charge in figure 6.7a can be described as the sum of the charge distributions in 6.7b, a uniform region of negative charge, and 6.7c, an isolated region containing a double positive charge and surrounded by a zero charge region. The image $V_c(x, y)$ due to distribution 6.7c is clearly twice the impulse response function, and it can be computed by subtracting the image $V_b(x, y)$ due to 6.7b from the image $V_a(x, y)$

due to 6.7a. That is,

$$V_c(x, y) = V_a(x, y) - V_b(x, y) = 2R_\delta(x, y). \quad (6.18)$$

In practice, however, measuring the image produced by a large uniform region of charge as in figure 6.7b is difficult. It is important to note that while no region of “zero charge” can be written on the perpendicular medium, regions of perpendicular media producing no significant external magnetic field can be written. An infinite or very large region of uniform surface charge, accompanied by its opposite polarity image layer, will produce no appreciable external field sensible by the read head. Instead, only an internal demagnetizing field exists, as the external field due to the negative surface charges cancels with that due to the positive charges in the lower layer. (An approximate analogy can be drawn to electrostatics in the case of an infinite parallel plate capacitor, with polarization replacing the magnetization in the magnetostatic picture of the perpendicular recording medium.) For this reason, it is desirable to avoid measurement of the impulse response function.

The step response function $R_u(x, y)$, on the other hand, lends itself well to measurement. Strictly understood, a similar process of measurement via superposition would have to be performed to measure the true step response function. While a measured step function image would be two-level, the lower level of this two-level charge distribution would not be zero. Thus, it would require a level shift by means of superposition with a region of constant charge. In practice, however, the second-order derivative in the definition of the readback image in terms of $R_u(x, y)$ (equation 6.8) renders null the addition of any constant term to $R_u(x, y)$. Therefore,

it is useful to consider the measured step response function $R_u^1(x, y)$, since

$$\frac{\partial^2 R_u^1(x, y)}{\partial x \partial y} = \frac{\partial^2 R_u(x, y)}{\partial x \partial y} = R_\delta(x, y). \quad (6.19)$$

Measurement of such a step response function is effected by writing a low-frequency periodic pattern (e.g., “111111000000111111000000...”) on a perpendicular medium with the spin-stand on a DC-erased background (e.g., “111111111111”) and subsequently imaging it with the spin-stand. That is, a region of a disk platter is first uniformly written with one value of magnetization in both the along-track and across-track directions, and subsequently in this region a periodic track pattern whose period is much larger than the response width of the head (approximating infinite extent) is written. The small region around the transition from one polarity to the other in the magnetization/charge/readback image is therefore the region of interest containing the measured step response function.

However, the experimental measurement of the 2D step function on the spin-stand is exacerbated by the spin-stand read preamplifier’s limitations in conjunction with the nature of spin-stand imaging itself. The preamplifier, which amplifies the small signals (in the μV range) produced by the read head so that they may be usable (in the mV range), is AC-coupled (like many hard disk head preamplifiers) and thereby rejects the DC component of the readback signal. Recall from chapter 1 that the process of spin-stand imaging builds up the scanned image by discrete steps of the head in the across-track direction. The fact that the 2D image is composed of a concatenation of 1D traces means that any large 1D slice of constant magnetization is filtered out by the preamplifier and appears as zero voltage. Such narrow

regions of constant magnetization are both charge containing and external field producing. Note that this is a different phenomenon from that seen for the essentially zero external field in the case of an “infinite” 2D region of constant magnetization. Therefore, although an entire “track” of a given radius on a perpendicular disk may be recorded with a given direction of magnetization, it will appear – regardless of the magnetization’s polarity – indistinguishable from any other constant value. This phenomenon is illustrated in figure 6.8, which shows images of low frequency patterns written and imaged on the spin-stand before and after applying a simple DC adjustment algorithm. The zero-valued “guard bands” present in the top image of figure 6.8 are not actually there, but a manifestation of this preamplifier distortion; their values are shifted properly in the bottom image. The algorithm employed for this DC level adjustment is a simple normalization of peak-to-peak amplitudes for the 1D measurement slice of the image present at each across-track step. First, the peak-to-peak voltage range of each slice is computed. The differences between each of these ranges and the “full” desired peak-to-peak range known from non-constant slices (e.g., the track at $1 \mu m$ in figure 6.8) is computed. Then, the difference for a given across-track slice is added to the entire measurement at that across-track step. Therefore, for those measurement slices requiring no level adjustment, none is made, but those requiring the full adjustment (the zero-valued “guard bands”) receive it. That is, for each measurement slice i , the following operation is performed:

$$V_{adjusted}(x, y_i) = V(x, y_i) + 0.5(V_{desired}^{p-p} - V_i^{p-p}), \quad (6.20)$$

where $V(x, y_i)$ is the 1D measurement at across-track step i , $V_{desired}^{p-p}$ is the desired

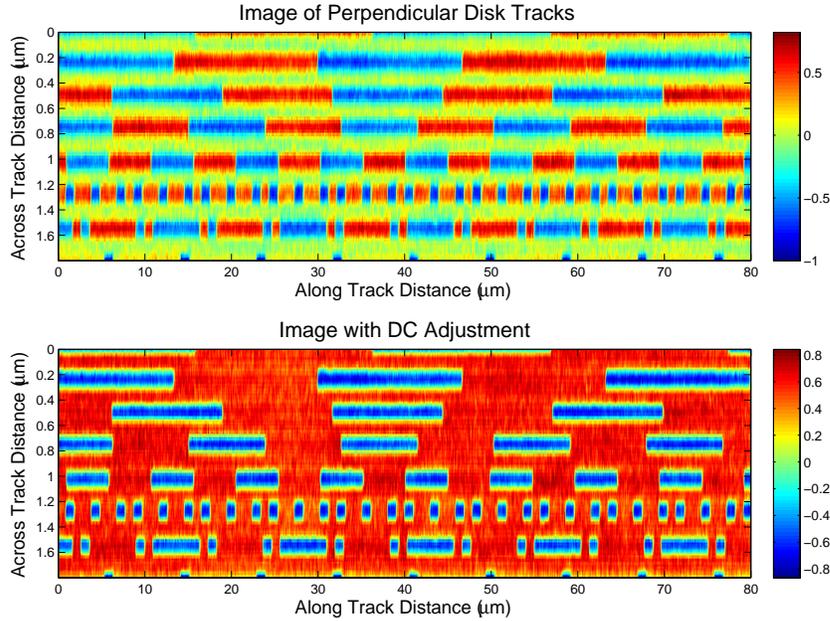


Figure 6.8: Numerical adjustment for the DC filtering effect of the spin-stand preamplifier on a spin-stand image obtained of various low-frequency perpendicular disk tracks.

full peak-to-peak amplitude, and V_i^{p-p} is the peak-to-peak amplitude of $V(x, y_i)$.

A further difficulty in measurement of the 2D response function presents itself in the fact that only tracks of finite width may be written by the spin-stand’s write head and only one track may be written per revolution of the disk. This is not a problem when creating large uniform regions of magnetization, because the region may be extended in the across-track direction with subsequent across-track steps of the head. However, finite track width makes approximating an infinite “boundary” between two levels of magnetization, like that present in the 2D step transition, difficult. To do so requires exactly matching the timing in writing the

transition in magnetization between subsequent adjacent writing iterations (across-track steps), which necessitates sophisticated real-time triggering such as that used in servo writing but unavailable in our spin-stand environment. In practice, however, the principle of “write wide, read narrow” in the design of write/read heads comes into play to limit the effect of the finite track width. Because the read head’s sensing ability in the across-track direction is smaller than the companion write head’s track write width, an “infinite” extent in the across-track direction can be approximated by truncating the image before the edge of the track.

6.2.4 Experimental Results

In one dimension, a step response is produced by writing a long region of one direction of magnetization followed by a second long region of the other direction of magnetization and subsequently reading it back. The length of the uniform regions of magnetization guarantees the step transition’s isolation from linear superposition effects, in the same way that isolation from other charge allows the measurement of the impulse response in longitudinal recording. In this sense, it is the transition in magnetization that is the truly important actor in determining the head response, which manifests itself variously as either charge (longitudinal impulse) or transition in charge polarity (perpendicular step).

The result of such an experiment for a particular read head with a read width of 94 nm is shown in figure 6.9. Also shown is a scaled (for display) derivative of the measured step response, thereby illustrating an equivalent 1D impulse response for

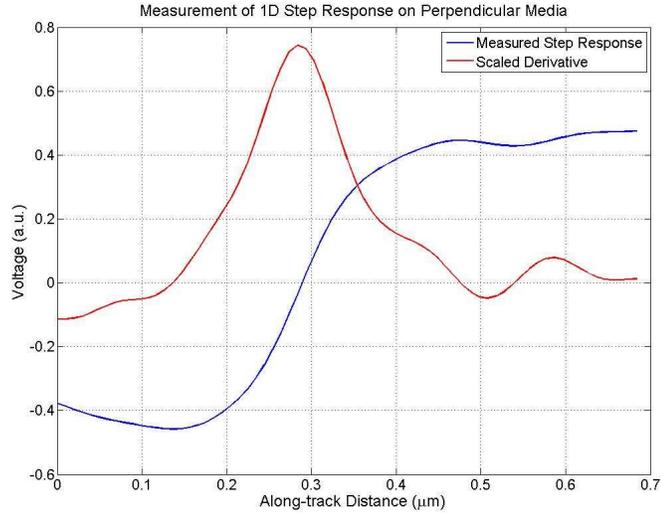


Figure 6.9: Measurement of 1D step response on a perpendicular medium with its corresponding equivalent 1D impulse response.

this head. In 2D, the step response function for the 94 nm wide read head measured from a track written by a write head with a write width of 127 nm is shown in figure 6.10. It is useful to note that the DC normalization technique described in equation 6.20 has been applied to the image, which manifests itself most strongly as a level shift in the region from $0.00 \mu\text{m}$ to $\approx 0.04 \mu\text{m}$ in the across-track direction to its appropriate constant value.

6.3 ISI Removal in Perpendicular Recording

Like readback images obtained from longitudinal magnetization, those obtained from perpendicular recording media are corrupted by ISI. However, due to the different natures of the readback voltages, the ISI is manifested differently even though the same read head technology may be used, as described in section 6.2.

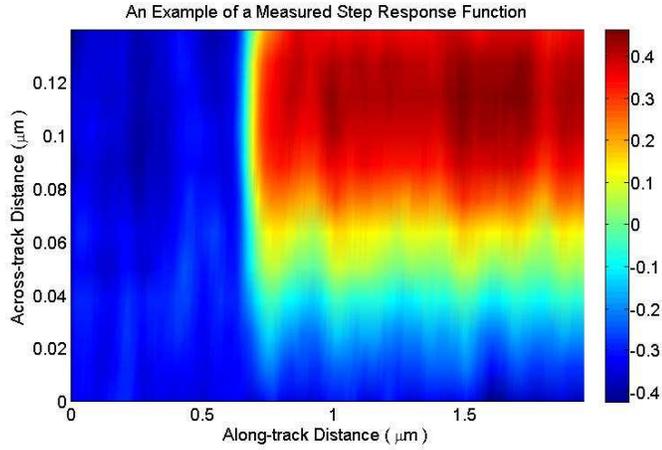


Figure 6.10: Measurement of 2D step response on a perpendicular medium, with the numerical DC-level adjustment described in equation 6.20 applied.

Recall the relations for the readback image in terms of the head response function ($R_\delta(x, y)$, the impulse response, or $R_u(x, y)$, the step response) and the surface charge distribution $\sigma_m^S(x, y)$ given in equations 6.1 and 6.8.

An important difference between the equation 6.1 in terms of the impulse response for perpendicular recording and equation 2.3 for longitudinal recording is that there is the means to stabilize the ISI deconvolution problem by transforming it into an integral-differential equation takes a different form. In the longitudinal case, the in-plane magnetization components were found instead of the charge (see section 2.2.2), exploiting the definition of the charge as the negative divergence of the magnetization. Because the fringing component of the magnetic field produced by this magnetization is sensed in longitudinal recording, the charge and magnetization as functions of the along-track distance have an integral/derivative relationship (i.e., charge at the center of a longitudinal track is the derivative of the in-plane, along-

track component of the magnetization). However, as discussed above, charge present on a perpendicular recording medium follows the magnetization itself because the magnetization is oriented normal to the recording medium surface. Therefore, the divergence of the magnetization on the outer surface of the medium plane looks functionally much like the magnetization itself.

Hence, the use of the step response function suggests itself again. Not only is its measurement more readily obtainable, the readback in terms of the step response function (equation 6.8) possesses a second-order derivative in the convolution integral. Writing the 2D Fourier transforms of equations 6.1 and 6.8 illustrates the greater utility of the step response function. The readback image in spatial frequency (k_x, k_y) domain in terms of the impulse response function is given by

$$\tilde{V}(k_x, k_y) = \tilde{R}_\delta(k_x, k_y)\tilde{\sigma}_m(k_x, k_y), \quad (6.21)$$

whereas the same given by the step response function is

$$\tilde{V}(k_x, k_y) = -k_x k_y \tilde{R}_u(k_x, k_y)\tilde{\sigma}_m(k_x, k_y), \quad (6.22)$$

with tilde denoting the transform of the corresponding function. Solving for the charge $\tilde{\sigma}_m(k_x, k_y)$ in each case gives

$$\tilde{\sigma}_m(k_x, k_y) = \frac{\tilde{V}(k_x, k_y)}{\tilde{R}_\delta(k_x, k_y)}, \quad (6.23)$$

and

$$\tilde{\sigma}_m(k_x, k_y) = -\frac{\tilde{V}(k_x, k_y)}{k_x k_y \tilde{R}_u(k_x, k_y)}. \quad (6.24)$$

Because equation 6.24 contains a product of the spatial frequency terms in the denominator, high frequency terms in the raw readback image $\tilde{V}(k_x, k_y)$ are suppressed

and the stability of the solution is potentially increased. Therefore, it seems desirable to employ equation 6.24 over equation 6.23 in solving for the undistorted charge distribution. (Naturally, the charge in normal space is obtained by the inverse 2D Fourier transform of equation 6.24.)

Much like the longitudinal case, it is desirable to further stabilize the deconvolution solution by means of regularization. With simple zeroth order regularization, the charge in spatial frequency domain is given by

$$\tilde{\sigma}_m(k_x, k_y) = -\frac{\tilde{V}(k_x, k_y)}{k_x k_y \tilde{R}_u(k_x, k_y) + \lambda}, \quad (6.25)$$

where λ is the constant of regularization.

6.4 Perpendicular Recording Image Characteristics

Since the utility of spin-stand imaging is for microscopy, diagnostics, and forensics of hard disk platters, the novelty of perpendicular recording in general suggests that several features of commercial perpendicular disks are of interest.

6.4.1 DC values and AC Erasure

As alluded above in the context of step response measurement, one of the biggest differences mathematically between longitudinal and perpendicular charge distributions and readback signals is the possible presence of a DC component in the case of perpendicular recording. Since our spin-stand’s preamplifier (and the preamplifier used in commercial hard disks) filters this component, this has side effects on the imaging process in the form of the phenomena of “baseline wander”

and track-edge interference effects. Baseline wander occurs because the preamplifier effectively band-pass filters the perpendicular readback signal, and the attenuation of any low-frequency components causes movement in the preamplified signal's baseline value. This feature of perpendicular recording affects commercial hard disks as well [69]-[71]. Furthermore, track-edge distortions can occur in perpendicular recording readback if the read track is near a region of constant DC-valued magnetization, such as a DC-erased guardband [73]-[75]. A particular useful technique (AC erasure) that has been developed to help combat its effects is relevant to spin-stand imaging of perpendicular media. AC erasure works by greatly reducing the DC value of the readback signal obtained from the background "unrecorded" regions of the disk.

The previous section has discussed the trivial case of a perpendicular charge or readback signal containing only DC, i.e., a constant value. In one case, when a large 2D region possesses one value of magnetization, the equal and opposite charges present at the surface and below the soft underlayer counteract one another, all of the magnetic field remains inside the medium, and no appreciable external magnetic field is detected. In another case, a narrow region extended in the along-track direction is surrounded by varying distributions of magnetization, but is still read by the spin-stand as "zero," because the pre-amplifier filters out the DC components of the read signal. It is this latter case that is applicable to commercial hard disk media and of interest to spin-stand imaging.

Numerical simulations of more general readback that represents the same underlying pseudo-randomly generated recorded magnetization (bits) on both perpendicular and longitudinal disks are shown in figure 6.11. It is readily apparent

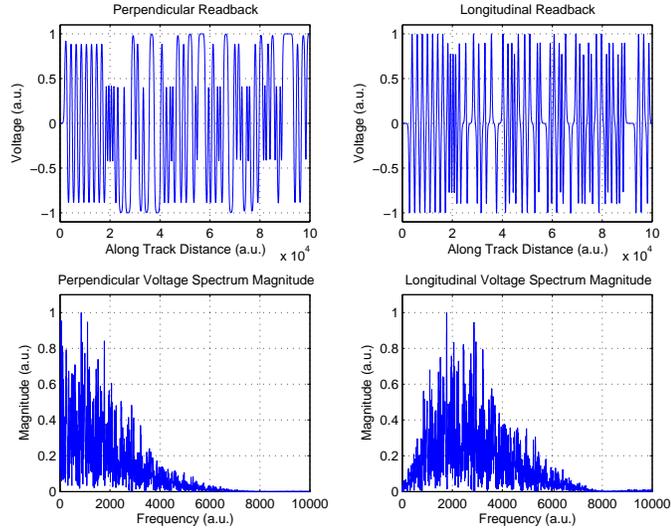


Figure 6.11: Simulation of perpendicular and longitudinal readback voltages of the same bit pattern with their spectra, highlighting the DC content in perpendicular readback and lack thereof in longitudinal recording readback.

that, for the same underlying magnetization, these two readback waveforms exhibit quite different spectra. The perpendicular readback on the left of figure 6.11 clearly contains more power at lower frequencies and at DC itself, while the longitudinal readback spectrum on the right goes to zero at DC. In longitudinal recording, where magnetic charge is present at the transitions of in-plane magnetization, a complete system of surface magnetic charges is sensed by the read head. Because $\nabla \cdot \vec{B} = 0$ (there are no isolated magnetic charges in nature), an equal number of positive and negative charges exist. Tracks of recorded magnetization are circular, and every positive transition (negative charge) must have a corresponding negative transition in longitudinal magnetization (positive charge). This means that, in the longitudinal readback voltage, each positive voltage pulse has a corresponding negative voltage

pulse, giving a total full-track voltage with zero mean.

On the other hand, in perpendicular recording, magnetization is oriented normal to the plane of the recording medium. As previously discussed, magnetic charge is present in two layers, one at the surface of the medium (which predominantly contributes to the readback signal) and the other effectively beneath the medium due to the presence of the soft underlayer. The system of charges is still complete when both layers are considered, and both layers are sensed by the read head. However, because the surface and lower layers of charge do not equally contribute to the field sensed by the read head, the necessity to have an equal number of positive and negative surface charges (and their accompanying negative and positive lower layer charges) that exists in the longitudinal recording is not present. Nothing, in principle, forces there to be an equal number of recorded logical ones and zeros and, therefore, an equal number of positive and negative surface charges (even though scrambling and coding may cause a tendency toward this situation in practice). Hence, any “imbalance” between the spatial regions of positive and negative perpendicular magnetization will give a non-zero mean to the total surface charge distribution and, thus, a DC component to the readback signal. In actuality, the non-ideal properties of the read head and media (e.g., finite permeability of the read head shields and soft underlayer) can reduce the magnitude of the DC response [69], [70], but the non-zero DC response is still an important feature of perpendicular recording that is missing from longitudinal recording.

For this reason, several techniques have been developed to compensate or eliminate the DC component in read head signals. Some techniques merely compensate

with signal processing, and therefore, involve coding and detection channel design [70]-[72]. Indeed, as discussed in chapter 4, the scrambling and coding techniques that generally have been used in hard disk drive channels tend to produce approximately equal numbers of logical ones and zeros written on the disk. Nevertheless, a more fundamental technique that is germane to spin-stand imaging is the technique of AC erasure, which modifies the magnetization on the disk itself [73],[74].

The perpendicular medium is said to be “AC-erased” when the readback voltage from the medium is approximately zero due to magnetic domain sizes (or bit sizes) of alternating polarity that are too small for the read head to fully resolve. In effect, AC erasure exploits intentional and extreme ISI with a high frequency periodic (nominally DC-free) bit pattern. The intent is to create small “bits” such that the precision of the read head cannot resolve any in particular, and the alternating polarity of the small domains causes the averaging effect of the readback process to sense them as nearly zero. Furthermore, any small residual (non-zero) value in the readback of the AC-erased region is less important than the fact that its periodicity gives the residual a DC value of ideally zero, therefore minimizing DC distortion in the readback. In practice, the process of AC erasure can be performed in two ways. A universally and in-situ performable method is to simply write a very high frequency periodic pattern continually at a range of across-track positions. A faster but less targeted method, which AC erases entire platters, is claimed in [74]. The method described in [74] employs a large DC magnetic field oriented parallel to the media surface for a precessional switching effect that causes the media to form small alternating polarity domains when the field is removed.

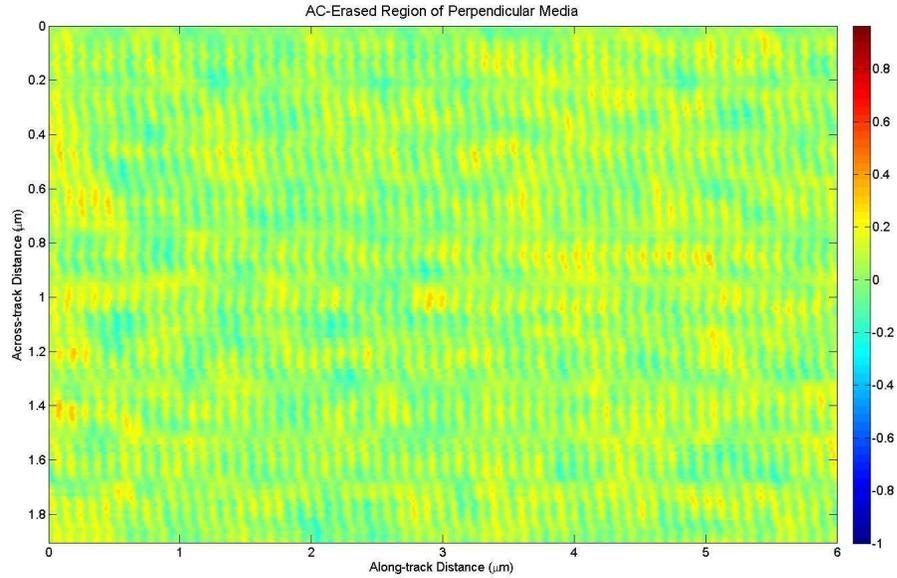


Figure 6.12: Spin-stand image of an AC-erased perpendicular medium.

Figure 6.12 shows an AC-erased region of a commercial perpendicular disk platter from 2007. It is most likely the AC erasure was performed by a high frequency write pattern, and the period seen in the image is approximately 100 nm, comparable to the spin-stand’s read head width of 94 nm. (Therefore, it is possible that aliasing could be occurring and the actual AC erase frequency is much higher.)

An example of the necessity of AC erasure in perpendicular recording is seen in a particular experiment involving overwritten data containing an unequal number of logical ones and zeros (see figures 6.13 and 6.14). In this experiment, a periodic (NRZI modulated) data pattern of hexadecimal F9 (“...11111001...”) is written by the Guzik spin-stand on an AC-erased area of a commercial perpendicular disk medium. Then, another periodic data pattern of F6 (“...11110110...”) is written on top of the previously written F9, but with a small across-track offset (≈ 50 nm),

and the resulting overwritten pattern of magnetization is imaged. As the readback waveforms at the across-track positions noted in figure 6.14 indicate and the image in figure 6.13 displays, the edge of the overwritten F9 data is still readable. (Recall from chapter 4 that NRZI modulation places the bits at the transitions in magnetization.) While the AC-erased background is visible and some minor interference between that background and the edges of the track is evident, significant distortion of the track edge is not present. Furthermore, variation in the baseline of the readback is small (only approximately 5% of its peak-to-peak amplitude). This may partially be due to the nature of the recorded pattern, but may also illustrate that AC erasure mitigates baseline wander to a degree as well. (It is reasonable to assume that any residual baseline shift in a commercial drive's readback voltage would be compensated in the detection scheme.)

The previous example contrasts with the case of writing on DC erased media, as shown in figures 6.15 and 6.16. Here, both baseline wander and across-track interference are apparent. It is easily seen in figure 6.15 that the tracks are distorted variously in the across-track direction according to the polarity of the magnetization (charge), as it either constructively or destructively interferes with the constant-valued DC erased guard band of the same or opposite polarity. It can be seen in figure 6.15 that the positive (red) regions of charge tend to be wider than the negative (blue) regions. Although the guard band is imaged as zero due to the DC-rejection of the pre-amplifier, its actual positive value is readily apparent from this image by noting the corruption of the widths of the negative magnetization (charge) regions. Figure 6.16 shows the readback waveform at the center of the track at across-track

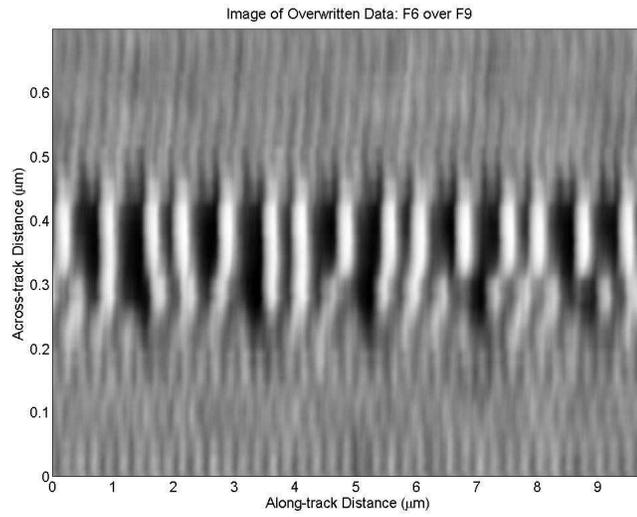


Figure 6.13: Spin-stand Image of F6 overwritten on F9 on an AC-erased perpendicular medium.

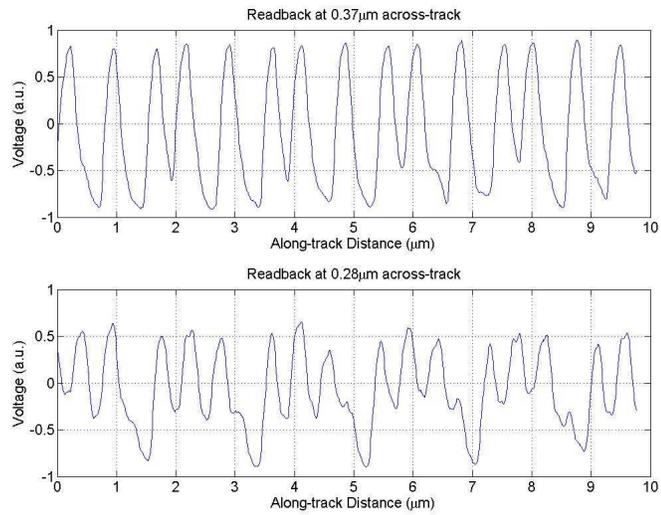


Figure 6.14: Spin-stand Image of F6 overwritten on F9 on an AC-erased perpendicular medium.

distance $0.9 \mu\text{m}$ in figure 6.15 over a much longer distance. At this scale, it exhibits very obvious baseline wander (approximately 60% of its peak-to-peak amplitude) due to the non-zero mean of the pattern (NRZI bits "...00011000...").

6.4.2 The Hilbert Transform in Perpendicular Recording

In chapter 3, the Hilbert transform was discussed in the context of relating the longitudinal magnetic field to the transverse magnetic field in two dimensions, and more generally in three dimensions with the derived extension. Another implication of this relation in two dimensions (the 1D Hilbert transform) is that it connects the voltage readback between longitudinal and perpendicular recording. If the same bit pattern is written on both a longitudinal medium and a perpendicular medium and both are read with the same head, the readback voltage signals in the two cases are related through the 1D Hilbert transform [26]. Furthermore, the 2D Hilbert transform derived in chapter 3 can be used to calculate equivalent longitudinal 2D images from measured images of perpendicular media, which reduces problems to the previous case already dealt with in longitudinal recording. While this relation does not take into account the different natures of charge distribution between longitudinal and perpendicular recording, in particular the presence of two layers of charge producing the field emanating from perpendicular media, it nevertheless gives meaningful results.

In figures 6.17 and 6.18, a series of low-frequency periodic patterns are shown with their 2D Hilbert transform components H_x and H_y (with the typical assumption

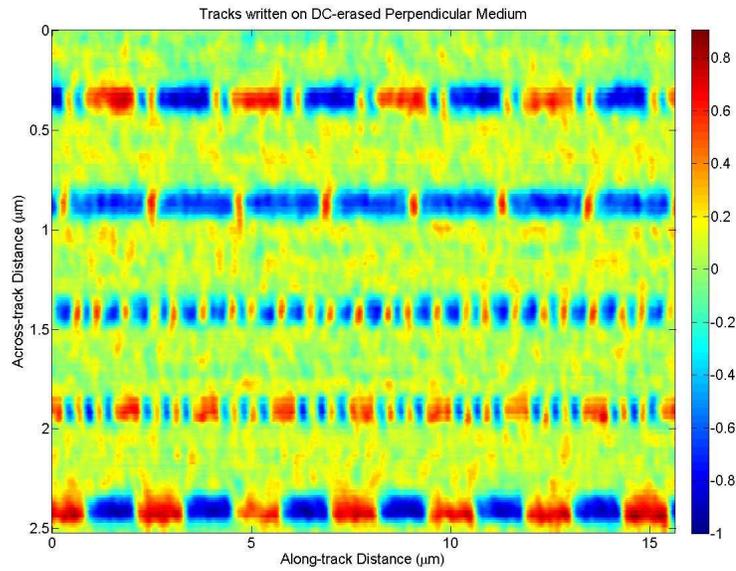


Figure 6.15: Spin-stand Image of Tracks written on a DC-erased perpendicular medium.

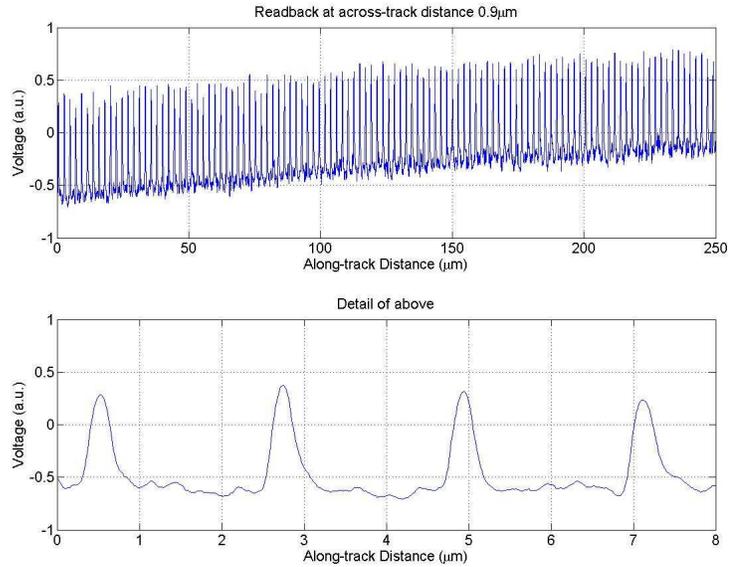


Figure 6.16: Readback waveform at 0.9 μm across-track distance in figure 6.15 of a track written on a DC-erased perpendicular medium.

that the measured image is H_z). These are obtained by applying the formulas 3.18 and 3.19 computed via the Fourier representations given in 3.40 and 3.41. It can be clearly seen that while the perpendicular image patterns have square-wave-like characteristics, the H_x component computed via the 2D Hilbert transform appears as a series of pulses alternating in polarity, as would be expected when reading low-frequency patterns on a longitudinal disk.

6.5 Conclusions

It has been shown that perpendicular recording is both a new direction for the magnetic recording industry and spin-stand imaging. The novelties present in perpendicular recording, arising from a completely different recording geometry and head/media design, range from two-level signals and non-zero DC components to a different sense of defining the response function of the read head. The methods to understand and measure the perpendicular recording response function in the form of either an impulse response or a step response have been presented. The use of the Hilbert transform to process spin-stand images to their approximate longitudinal equivalents has been described.

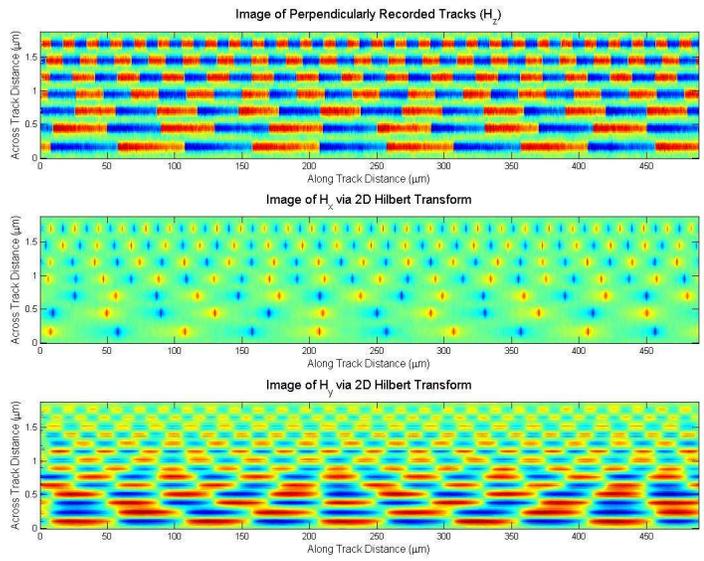


Figure 6.17: The spin-stand image of low-frequency patterns on perpendicular media (top) with their H_x (middle) and H_y (bottom) components computed via the 2D Hilbert transform.

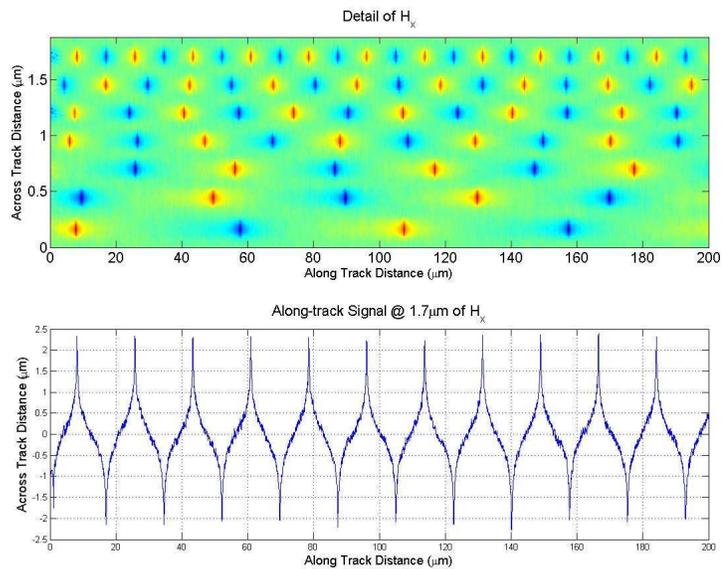


Figure 6.18: A detail of figure 6.17 with the readback at $1.7 \mu\text{m}$ across-track distance.

Chapter 7

Conclusions and Future Directions

7.1 Summary and Conclusions

A number of topics have been explored pertaining to the development of spin-stand-based methods for diagnostic and forensic analysis of hard disk media. These range from microscopy techniques for longitudinal and perpendicular recording media to the “drive-independent” characterization of write channels, and together they represent a complementary set of tools. As magnetic hard disk drives remain ubiquitous for the foreseeable future in large-scale, random-access storage applications, the presented spin-stand-based techniques provide a useful means to study the state of the art in commercial products. Indeed, by building on their current utility, the tools and methods described in this dissertation offer ample opportunity to grow with the continually progressing level of hard disk drive technology and the state of the magnetic recording industry in general.

Techniques have been discussed and demonstrated pertaining to the removal of sensor-based distortion in images obtained from longitudinal recording media. This distortion, or ISI, is due to the lack of perfectly local sensing in the imaging sensor employed. On one hand, the spin-stand offers tremendously fast and large-scale imaging capability by virtue of its specifically disk-oriented construction, having a high-quality spindle motor and precision sensor positioning abilities. At the same

time, the sensor employed is merely the same GMR or TMR head used in hard disks themselves. For all its tremendous capabilities in sensing small magnetic fields while flying at mere nanometers above an only approximately smooth disk surface, the read head is not perfect. As such, it is desirable for precision microscopy applications to compensate for the distortion introduced by the imprecision of the head. One direct way to do this was described that employs a characterization of the head by means of its response function, as described in [1, 5]. With this characterization, the readback voltage due to the recording medium is defined as the convolution of the head response function with the underlying charge distribution. By writing a small spot of charge on the platter and subsequently imaging it and making some arguments about scaling [7], the read head response function can be measured. Having this characterization, a technique to solve for the magnetization, and therefore the undistorted charge, is given. Chapter 1 of this dissertation focused on the implementation and results of such an algorithm with respect to its implementation with regularization of the solution.

Another technique useful in enhancing the spin-stand microscopy of longitudinal media was given in chapter 2 with the Hilbert transform. It was discussed that the Hilbert transform is well known in its 1D form in magnetic recording to describe in two dimensions the relationship between the in- and out-of-plane components of the magnetic field emanating from the medium. It turns out that it can also be used as a component in a saturation-based algorithm that enhances readback obtained by the spin-stand by directly and easily solving for the longitudinal magnetization given the readback voltage. This argument was based on the fact that the exter-

nal component of the magnetic field emanating from the medium mimics up to a sign the underlying magnetization that produces it. Subsequently, it was shown that a multidimensional extension to the Hilbert transform can be defined based on physical arguments pertaining to the geometry of magnetic recording. Due to the “one-sided” nature of the recording medium, a form relating the normal component of the magnetic field H_z to its vectorial tangential or in-plane components H_x and H_y can be derived. This is of interest in itself, but is also of great utility in the processing of spin-stand images. Just as the 1D Hilbert transform was shown to be useful for readback enhancement, the 2D Hilbert transform’s usefulness in enhancing images was illustrated.

With these methods for image enhancement understood, techniques for making sense of commercial hard disk data were discussed. Although with microscopy alone we can see the magnetization and charge patterns on the surface of the disk, this alone does not interpret what these patterns means. To do so involves understanding the several layers of encoding applied to binary digital data before it is modulated onto the write-current waveform and subsequently recorded magnetization. The three layers of encoding studied were scrambling, RLL, and ECC. Scrambling and RLL exist in the commercial hard disk drive write channel in order to precondition the data bits such that the magnetization on the disk is easily readable. In general, without this preconditioning, the synchronous detection system employed in the read channel cannot function with the high degree of reliability (low BER) required for a consumer product. That said, RLL by definition is a non-linear code and requires statistical analysis correlated with published specifications (e.g., in patents

and papers) in order to identify it. Scrambling, on the other hand, is linear and techniques were described to solve for the polynomial and initial states involved in defining this layer of “coding.” In practice, the process of characterizing the RLL and scrambling schemes used in a commercial disk drive must be done together due to their layered structure as described in chapter 4. ECC, the error correction code symbols appended to each sector of data, can be treated separately. ECC also is linear, and the method for solving for the polynomial and initial state have been given in chapter 5.

Finally, perpendicular recording, the novel and important new direction for the recording industry, was investigated in the context of spin-stand microscopy. It was discussed that perpendicular recording offers greater performance in terms of density, but at the cost of very different write head and media design and very different readback signals. Likewise, the process of spin-stand imaging is understood in a different way due to these novelties of perpendicular recording. Because of the presence of a soft underlayer beneath the perpendicular recording media to complete the flux path during writing, the reading process is different than in longitudinal recording due to the presence of a second layer of image charge beneath the media surface. This fact furthermore implies that any head response employed in the context of perpendicular recording must take this layer into effect. However, just as in longitudinal recording, the readback voltage image can be modeled as the convolution between the head response function and the surface magnetic charge distribution. It was shown that response function cannot be directly measured in the perpendicular recording context due to the nature of the medium, which only

supports positive and negative (but not zero) charge distributions. Furthermore, it was shown that while the head response may be characterized as either an impulse response function or a step response function, the step response function offers some advantages with regard to ease of measureability and in ISI removal performance. Lastly, some interesting properties of the perpendicular recording used in commercial disks were shown. In particular, the use of AC recording was demonstrated and the utility of spin-stand imaging in characterizing its effects on recorded magnetization. Also, it was shown that the Hilbert transform may be used to compute effective longitudinal readback and images from measured perpendicular images.

7.2 Future Directions

Although significant progress has been made in developing the aforementioned techniques, there are several areas in which further progress can be made. The primary direction lies in upgrading fundamental spin-stand capabilities, but there is also interest in expanding the research in general.

7.2.1 DC Response

Besides ISI, which is an unavoidable aspect of imaging, the biggest source of corruption in spin-stand images is the undesirable filtering of DC voltage in the spin-stand's read head pre-amplifier. While this pre-amplifier is a crucial part of the read circuit (just as it is in hard disk drives) because it amplifies the very small output voltage of the read head into a more usable range (mV rather than μV), its

standard interface is differential and AC-coupled for common mode noise rejection purposes. It would be greatly desirable to either integrate a pre-amplifier without this DC rejection, or to find a robust algorithm that would compensate for this. A simple post-processing technique to normalize levels of DC charge in spin-stand images was offered in chapter 6, but it surely can be improved. Having a full DC response would allow direct measurement of the perpendicular read head response and ensure more accurate images of perpendicular media.

7.2.2 Write circuit and ISI experiments

Another greatly desirable enhancement to the spin-stand is the integration of a higher frequency write circuit. The particular spin-stand used in the research in this dissertation is limited to 166 Megaflux per second (MFlux/s), or 83 MHz, in its write frequency. This greatly limits the minimum distance achievable between adjacent transitions written on media mounted on the spindstand. Indeed, this write frequency is certainly below the maximum achievable by the most recent heads that can be mounted on the spin-stand used in the presented research. While greater density can be achieved by writing with a larger electrical frequency while the disk rotation speed is lower, there is a very limited extent to which rotation speeds can be altered and the heads still function properly. (Changing speeds is also lacking in consistency, because the flying height of the head depends on the aerodynamic design of the head slider and the disk surface roughness, among other factors.) For this reason, higher write frequencies are necessary to perform high-precision ISI

experiments, wherein the distance between neighboring transitions can be varied in some controllable range. The variation in the distance between transitions thereby controls the varying amount of ISI in the readback image, which in turn allows the study of ISI removal performance. This is particularly necessary in the case of perpendicular recording ISI experiments, in which the spin-stand's present write driver cannot write at sufficiently high densities to induce significant ISI in the subsequent readback voltage.

7.2.3 Coding schemes

While a variety of hard disk drives were studied in chapters 4 and 5 in the context of write channel and ECC identification and digital data reconstruction, it is naturally interesting to continue to do so for each succeeding generation of drives. As densities increase, SNRs must fall, which necessitates ever more sophisticated coding and detection techniques to maintain the BERs appropriate for commercial devices. For instance, coding techniques like low-density parity check coding are areas for commercial drive performance improvements [79, 80]. Performing “drive-independent” study and implementation of such techniques would be desirable.

Bibliography

- [1] C. Tse, *Magnetic imaging of hard disk drives on a spin-stand*, Ph.D. Thesis, University of Maryland College Park (2003).
- [2] I.D. Mayergoyz, C. Serpico, C. Krafft, and C. Tse, "Magnetic imaging on a spin-stand," *J. Appl. Phys.*, **87**, 6824 (2000).
- [3] I.D. Mayergoyz, C. Tse, C. Krafft, and R.D Gomez, "Spin-stand imaging of overwritten data and its comparison with magnetic force microscopy," *J. Appl. Phys.*, **89**, 6772 (2001).
- [4] A.A. Companieh, R. Eaton, R. Indeck, and M. Moser, "In situ ultra-high resolution magnetic imaging," *IEEE Trans. Magn.*, **37**, 1257 (2001).
- [5] I.D. Mayergoyz, C. Tse, C. Krafft, D.I. Mircea, and P. Andrei, "Extraction of the response function of GMR head for spin-stand imaging," *IEEE Trans. Magn.*, **38**, 2453 (2002).
- [6] C. Tse, C. Krafft, I.D. Mayergoyz, and D.I. Mircea, "High-speed massive imaging of hard disk data by using the spin-stand imaging technique," *J. Appl. Phys.*, **93**, 6578 (2003).
- [7] I.D. Mayergoyz, C. Tse, C. Krafft, and D.I. Mircea, "Scaling of head response function in the spin-stand imaging technique," *J. Appl. Phys.*, **93**, 6581 (2003).
- [8] I.D. Mayergoyz, C. Tse, C. Krafft, and P. McAvoy, "A novel approach to removing intersymbol interference from spin-stand images," *IEEE Trans. Magn.*, **40**, 2197 (2004).
- [9] C. Tseng, I.D. Mayergoyz, C. Tse, P. McAvoy, and C. Krafft, "Dynamic track-following of off-centered hard disks in spin-stand imaging," *J. Appl. Phys.*, **97**, 10R301 (2005).
- [10] C. Tse, C. Tseng, P. McAvoy, C. Krafft, and I.D. Mayergoyz, "Whole-track imaging and diagnostics of hard disk data using the spin-stand imaging technique," *J. Appl. Phys.*, **97**, 10P104 (2005).
- [11] I.D. Mayergoyz, P. McAvoy, C. Tse, C. Krafft, and C. Tseng, "The 2-D Hilbert Transform in Magnetic Recording," *IEEE Trans. Magn.*, **42**, 2282 (2006).

- [12] C. Tseng, I. Mayergoyz, P. McAvoy, and C. Krafft, "Iterative compensation for hysteresis effects in positioning and tracking problems," *J. Appl. Phys.*, **103**, 07D902 (2008).
- [13] P. McAvoy, C. Tseng, I. D. Mayergoyz, and C. Krafft, "Spin-stand imaging of perpendicularly recorded data," *IEEE Trans. Magn.*, to appear Nov. 2008.
- [14] Shan X. Wang, Alexander M. Taratorin, *Magnetic Information Storage Technology* (Academic Press, San Diego, CA, 1999), Chap. 1.
- [15] J. M. Daughton, P.A. Bade; M. L. Jenson, M. M. M. Rahmati, "Giant magnetoresistance in narrow stripes," *IEEE Trans. Magn.*, **28**, 2488 (1992).
- [16] S. Mao, Y. Chen, F. Liu, X. Chen, B. Xu, P. Lu, M. Patwari, H. Xi, C. Chang, B. Miller, D. Menard, B. Pant, J. Loven, K. Duxstad, S. Li, Z. Zhang, A. Johnston, R. Lamberton, M. Gubbins, T. McLaughlin, J. Gadbois, J. Ding, B. Cross, S. Xue, P. Ryan, "Commercial TMR heads for hard disk drives: characterization and extendibility at 300 Gbit/in²," *IEEE Trans. Magn.*, **42**, 97 (2006).
- [17] Shan X. Wang, Alexander M. Taratorin, *Magnetic Information Storage Technology* (Academic Press, San Diego, CA, 1999), Chap. 6.
- [18] Shan X. Wang, Alexander M. Taratorin, *Magnetic Information Storage Technology* (Academic Press, San Diego, CA, 1999), Chap. 4.
- [19] W. Press, B. Flannery, S. Teukolsky, and W. Vetterling, *Numerical Recipes in C: the Art of Scientific Computing (Second Edition)* (Cambridge University Press, Cambridge, 1992), Chap. 18.
- [20] W. Press, B. Flannery, S. Teukolsky, and W. Vetterling, *Numerical Recipes in C: the Art of Scientific Computing (Second Edition)* (Cambridge University Press, Cambridge, 1992), Chap. 2
- [21] I.D. Mayergoyz, C. Tse, C. Krafft, "Method for Intersymbol Interference Removal in Data Recovery," US Patent 7,002,762 (2006).
- [22] J. C. Mallinson, "One-sided fluxes – a magnetic curiosity?," *IEEE Trans. Magn.*, **9**, 678 (1973).
- [23] J. C. Mallinson, "On the properties of two-dimensional dipoles and magnetized bodies," *IEEE Trans. Magn.*, **17**, 2453 (1981).

- [24] V. B. Minuhim, "Hilbert transform and phase distortions of signals," *Radio Eng. Electron. Phys.*, **18**, 1189 (1973).
- [25] V. B. Minuhim, "Phase distortions of signals in magnetic recording equipment," *Telecom. Radio Eng.*, **29/30**, no. 1 (1975).
- [26] H. N. Bertram, *Theory of Magnetic Recording* (Cambridge University Press, Cambridge, UK, 1994).
- [27] M. Brokate, A. Pokrovskii, D. Rachinskii and O. Rasskazov, "Differential equations with hysteresis via a canonical example," *The Science of Hysteresis*, (Academic Press, New York, 2005), Vol. 1, pp. 125-291.
- [28] J. T. Weaver, "On the separation of local geomagnetic fields into external and inner parts," *Zeitschrift für Geophysik*, **30**, 29 (1964).
- [29] L. Alvarez and L. Mazorra, "Signal and image restoration using shock filters and anisotropic diffusion," *SIAM Journal on Numerical Analysis*, **31**, 590 (1994).
- [30] G. Gilboa, N.A. Sochen, and Y.Y. Zeevi, "Regularized shock filters and complex diffusion," *Proceedings of ECCV '02*, 399 (2002).
- [31] Shan X. Wang, Alexander M. Taratorin, *Magnetic Information Storage Technology* (Academic Press, San Diego, CA, 1999), Chap. 11-12.
- [32] J. Moon and W. Zeng, "Equalization for maximum likelihood detectors," *IEEE Trans. Magn.*, **31**, 1083 (1995).
- [33] J. Fitzpatrick, J. K. Wolf, and L. Barbosa "New Equalizer Targets for Sampled Magnetic Recording Systems," *1991 Conference Record of the Twenty-Fifth Asilomar Conference on Signals, Systems and Computers*, **1**, 30 (1991).
- [34] K.B. Klaassen. "Magnetic recording channel front ends," *IEEE Trans. Magn.*, **27**, 4503 (1991).
- [35] B.E. Bloodworth, P.P. Siniscalchi, G.A. De Veirman, A. Jezdic, R. Pierson, and R. Sundararaman. "A 450-Mb/s analog front end for PRML read channels," *IEEE J. of Solid State Circuits.*, **34**, 1661 (1999).
- [36] R.D. Cideciyan, F. Dolivo, R. Hermann, W. Hirt, and W. Schott, "A PRML system for digital magnetic recording," *IEEE Journal on Selected Areas in Communications*, **10**, 38 (1992).

- [37] A.A. Friedmann, J.K. Wolf “Simplified EPR4 Detection,” *IEEE Trans. Magn.*, **34**, 129 (1998).
- [38] K.K. Fitzpatrick “A reduced complexity EPR4 post-processor,” *IEEE Trans. Magn.*, **34**, 135 (1998).
- [39] J. Cecil, R. Gopalan, W. Llewellyn, S. Masood, P. Tucci, and L. Wakeman, “Disk interface design guide and user’s manual: National Semiconductor Application Note 413,” National Semiconductor, (1986)
- [40] P.A. Franaszek, “Apparatus for encoding unconstrained data onto a (1,7) format with rate 2/3,” US Patent 4,488,142 (1984).
- [41] J.S. Eggenberger and A.M. Patel, “Method and apparatus for implementing optimum PRML codes,” US Patent 4,707,681 (1987).
- [42] R. Kuki and K. Saeki, “Encoder/decoder system with suppressed error propagation,” US Patent 6,097,320 (2000).
- [43] J.L. Sonntag, “Apparatus and method for increasing density of run length limited block codes without increasing error propagation,” US Patent 5,604,497 (1997).
- [44] S.A. Altekhar and S.M. Shih, “Rate 32/34 (D=0,G=9/I=9) modulation code with parity for a recording channel,” US Patent 6,229,458 B1 (2001).
- [45] G.D. Forney, “Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference,” *IEEE Trans. Inform. Theory*, **IT-18**, 363 (1972).
- [46] E.R. Kretzmer, “Generalization of a technique for binary data communication,” *IEEE Trans. Commun. Technol.*, **COM-14**, 67 (1966).
- [47] M. Mueller and M. Müller, “Timing recovery in digital synchronous data receivers,” *IEEE Trans. Commun.*, **COM-24**, 516 (1976).
- [48] F. Dolivo, W. Schott, and G. Ungerböck, “Fast timing recovery for partial-response signaling systems,” *Proc. IEEE Int. Conf. on Communications*, **1**, 573 (1989).
- [49] P. Roo, R.R. Spencer, P.J. Hurt, “Analog timing recovery architectures for PRML detectors,” *Proc. Globecom ’95 Conf.*, **1**, 571 (1995).

- [50] S. Golomb, *Shift Register Sequences* (Holden-Day, San Francisco, 1967).
- [51] S.A. Tretter, *Communication System Design Using DSP Algorithms* (Plenum Press, New York, 1995), Chap. 9.
- [52] Shan X. Wang, Alexander M. Taratorin, *Magnetic Information Storage Technology* (Academic Press, San Diego, CA, 1999), Chap. 8.
- [53] R.M. Neal, *Software for Low Density Parity Check Codes*, available at www.cs.toronto.edu/~radford/ftp/LDPC-2006-02-08/index.html.
- [54] A. Patel, "Signal and error-control coding," *Magnetic Storage Handbook (Second Edition)* (C.D. Mee and E.D. Daniel eds., McGraw-Hill, New York, 1996).
- [55] S. Lin and D.J. Costello Jr., *Error Control Coding: Fundamentals and Applications, Second Edition* (Pearson Prentice-Hall, Inc., Upper Saddle River, NJ, 2004), Chap. 1.
- [56] J.W. Eaton, "GNU Octave," available at www.octave.org.
- [57] The Mathworks, "Matlab," available at www.mathworks.com.
- [58] S. Iwasaki, Y. Nakamura, and K. Ouchi, "Perpendicular magnetic recording with a composite anisotropy film," *IEEE Trans. Magn.*, **15**, 1456 (1979).
- [59] R.I. Potter and I.A. Beardsley, "Self-consistent computer calculations for perpendicular magnetic recording," *IEEE Trans. Magn.*, **16**, 967 (1980).
- [60] Shan X. Wang, Alexander M. Taratorin, *Magnetic Information Storage Technology* (Academic Press, San Diego, CA, 1999), Chap. 16.
- [61] R. Wood, "The feasibility of magnetic recording at 1 terabit per square inch," *IEEE Trans. Magn.*, **36**, 36 (2000).
- [62] H.N. Bertram and M. Williams, "SNR and density limit estimates: a comparison of longitudinal and perpendicular recording," *IEEE Trans. Magn.*, **36**, 4 (2000).
- [63] M. Mallery, A. Torabi, and M. Benakli, "One terabit per square inch perpendicular recording conceptual design," *IEEE Trans. Magn.*, **38**, 1719 (2002).
- [64] M. Mallery, "Vertical Magnetic Recording Arrangement," US Patent 4,656,546 (1987).

- [65] I. Mayergoyz and C. Tse, *Spin-stand Microscopy of Hard Disk Data*, (Elsevier, Amsterdam, 2007), Appendix A.
- [66] Shan X. Wang, Alexander M. Taratorin, *Magnetic Information Storage Technology* (Academic Press, San Diego, CA, 1999), Chap. 7.
- [67] S. Gopalaswamy and P. McEwen, “Read channel issues in perpendicular magnetic recording,” *IEEE Trans. Magn.*, **37**, 1929 (2001).
- [68] H. Sawaguchi, Y. Nishida, H. Takano, and H. Aoi, “Performance analysis of modified PRML channels for perpendicular recording systems,” *J. of Magnetism and Magnetic Materials*, **235**, 265 (2001).
- [69] R.D. Cideciyan, E. Eleftheriou, and T. Mittelholzer, “Perpendicular and longitudinal recording: a signal-processing and coding perspective,” *IEEE Trans. Magn.*, **38**, 1698 (2002).
- [70] A. Patapoutian, “Baseline wander compensation for the perpendicular magnetic recording channel,” *IEEE Trans. Magn.*, **40**, 235 (2004).
- [71] M. Fatih Erden and E. M. Kurtas, “Baseline wander compensation for perpendicular recording,” *IEEE Trans. Magn.*, **40**, 3114 (2004).
- [72] X. Hu and B.V.K.V. Kumar, “Evaluation of low-density parity-check codes on perpendicular magnetic recording model,” *IEEE Trans. Magn.*, **43**, 727 (2007).
- [73] E. N. Abarra, P. Gill, B. R. Acharya, J. Zhou, M. Zheng, G. Choe, and B. Demczyk, “Bulk AC-erasure technique for perpendicular recording media: effect of exchange coupling,” *IEEE Trans. Magn.*, **41**, 3127 (2005).
- [74] E. N. Abarra, P. Gill, M. Zheng, J. N. Zhou, B. R. Acharya, and Gunn Choe, “Preconditioning, write width, and recording properties of Co-Cr-Pt-O perpendicular media with various underlayer designs,” *IEEE Trans. Magn.*, **41**, 581 (2005).
- [75] W. Jiang, G. Khera, R. Wood, M. Williams, N. Smith, and Y. Ikeda, “Cross-track noise profile measurement for adjacent-track interference study and write-current optimization in perpendicular recording,” *J. Appl. Phys.*, **93**, 6754 (2003).
- [76] H. J. Richter, “The transition from longitudinal to perpendicular recording,” *J. Phys. D: Appl. Phys.*, **40**, R149 (2007).

- [77] A. S. Chekanov, E. N. Abarra and G. Choe, “Perpendicular recording media imaging using dynamic read-back microscopy,” *IEEE Trans. Magn.*, **42**, 2345 (2006).
- [78] M. Hashimoto, M. Salo, Y. Ikeda, A. Moser, R. Wood, and H. Muraoka, “Analysis of written transition curvature in perpendicular magnetic recording from spin-stand testing,” *IEEE Trans. Magn.*, **43**, 3315 (2007).
- [79] B. M. Kurlowski, P. H. Siegel, J. K. Wolf, “Joint message-passing decoding of LDPC codes and partial-response channels,” *IEEE Trans. Inform. Theory*, **48**, 1410 (2002).
- [80] W. E. Ryan, F. Wang, R. Wood, Y. Li, “Optimal code rates for the Lorentzian channel: Shannon codes and LDPC codes,” *IEEE Trans. Magn.*, **40**, 3559 (2004).