Coding and signal processing for high density granular magnetic recording systems

Sari Shafidah Shafi'ee

2016

Sari Shafidah Shafi'ee. (2016). Coding and signal processing for high density granular magnetic recording systems. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/68645>

<https://doi.org/10.32657/10356/68645>

NANYANG
TECHNOLOGICAL
UNIVERSITY

**CODING AND SIGNAL PROCESSING FOR HIGH DENSITY
GRANULAR MAGNETIC RECORDING SYSTEMS**

SARI SHAFIDAH BTE SHAFI'EE

SCHOOL OF ELECTRICAL AND ELECTRONIC ENGINEERING

2016

**CODING AND SIGNAL PROCESSING FOR HIGH DENSITY
GRANULAR MAGNETIC RECORDING SYSTEMS**

SARI SHAFIDAH BTE SHAFI'EE

School of Electrical and Electronic Engineering

A thesis submitted to the Nanyang Technological University in partial
fulfillment of the requirement for the degree of Doctor of Philosophy

2016

Acknowledgements

I would like to express my deepest gratitude to my advisors, Dr Chan Kheong Sann and Professor Goh Wang Ling. I am sincerely thankful for your patience, guidance and scientific advices. The time and effort each of you devoted throughout the course of my PhD has molded my learning curve positively and allowed me to grow as a researcher. This dissertation stands as a testament to your tremendous dedication and support.

I would like to express my special appreciation and thanks to Kheong for being a great Project Leader in Data Storage Institute (DSI). I have travelled a long journey under your guidance. The knowledge and skills you imparted have shaped my career positively. Thank you for always ensuring that my work is aligned with my thesis. Pursuing a part-time PhD would otherwise have been doubly hard.

I would also like to thank the Agency of Science, Technology and Research (A*STAR) for the PhD sponsorship. Thank you to DSI for the resources and administrative assistance rendered throughout the course of my study.

Last but not least, I am truly grateful to my family, friends and colleagues for making me believe in myself and all that I am. Each and every one of you played a vital role in turning my cant's into cans and my dreams into plans. This dissertation would not have been possible without this circle of support.

Abstract

The Hard Disk Drive (HDD) industry has, in recent years, struggled with attempting to attain continual growth in storage capacity to meet the needs of a rapidly growing storage demand in the information age. Conventional Perpendicular Magnetic Recording (PMR) is already hitting its limit with areal density growth slowing down considerably in recent years. In moving towards higher densities, PMR faces the problem of simultaneously shrinking the size of bits and ensuring the media remains writable without compromising the thermal stability of its magnetic grains. To overcome this, various novel recording architectures have been proposed, namely Heat Assisted Magnetic Recording (HAMR), Bit Patterned Magnetic Recording (BPMR), Shingled Magnetic Recording (SMR) and Two-Dimensional Magnetic Recording (TDMR).

HAMR overcomes the trilemma by injecting thermal energy into a high coercivity and thermally stable media (needed for high density) that would otherwise require an excessively strong head field to write. BPMR organizes magnetic grains into a defined lattice and stores bit on single grains known as islands. SMR targets for a higher track density by writing data tracks using a shingling technique. TDMR uses the same writing mechanism as in SMR. In TDMR however, information is read simultaneously from multiple tracks through the use multiple reader heads.

To attain massive storage improvements in the near future, the industry needs to invent new ideas across the entire HDD platform. Development of new recording architectures, concurrent with the development of new trends in signal processing are essential in paving the way to new and intriguing prospects of significant performance gains. As such, the first part of the thesis focuses on a system level study of the recently proposed SMR and HAMR channel. The formulation and evaluation of theoretical models for these emerging recording technologies is essential in enabling an accurate prediction of its feasibility. Magnetic researchers often employ

micromagnetic channel models for this purpose. Despite its high accuracy, the micromagnetic model is computationally slow in reproducing the magnetization of grains. This renders it unsuitable for error performance studies. The thesis proposes a Grain Flipping Probability (GFP) channel model as an alternative to the micromagnetic model. The GFP consists of a multi-dimensional look-up table (LUT) that is characterized from micromagnetic simulations. It is highly accurate and is able to reproduce grain magnetizations multiple times faster than the micromagnetic model. With the GFP channel model developed, a novel study of the feasibility of SMR and HAMR is carried out in this thesis from a system's perspective.

Developments of new magnetic recording channels can be synergized with advancements in signal processing techniques to bring about an impactful gain in storage capacity in the near future. The remaining part of the thesis therefore focuses on novel signal processing schemes for magnetic recording channels. The iterative detector-decoder is the standard used in today's HDD. It comprises of a detector block and decoder block that aims to recover the stored information on a media through repetitive exchange of mutual information between the blocks. The detector operates on a trellis structure while the decoder operates on a factor graph. As the iterative detection-decoding scheme is sub-optimal, recent trend in signal processing research is inclined towards the area of joint detection and decoding where the detector and decoder are regarded as one entity that simultaneously detects and decodes the stored information. In this thesis, two novel joint detection/decoding techniques are proposed – 1) Joint Viterbi Detector Decoder (JVDD) 2) Joint Factor Graph Detector Decoder (JFGDD). The JVDD performs detection and decoding jointly in a single step over a channel trellis that consists of pre-defined parity check nodes. While JVDD has shown to perform well in the simulations carried out in this PhD work, its computational complexity can be a drawback in some circumstances. To manage this, novel error-correction codes coined the JVDD class of codes are proposed for the JVDD. The proposed JVDD codes are analytically optimized in this thesis so as to attain the best performance and lowest complexity for the JVDD.

The proposed JFGDD scheme on the other hand, operates on a super factor graph structure that is constructed from a combination of a channel factor graph and a code factor graph. As

the conventional detection algorithm is executed on a channel trellis, the thesis proposes an alternative implementation of the algorithm on a channel factor graph instead. The performance of the proposed JFGDD is susceptible to the presence of short cycles in its graph. As such, two novel methods of mitigating short cycles are proposed in this thesis. The first method proposes a non-binary adaptation of the algorithm that will minimize the number of connections and thus cycles in the factor graph. The second method proposes the design of constrained General Partial Response (GPR) targets.

Advances in signal processing for HDD should be made not only in the area of detection and decoding but also in the area of equalization. As such, the last part of the thesis proposes novel 2D equalization schemes for SMR and TDMR. The shingling technique of writing data in an SMR or TDMR channel causes the read-back signal to be severely corrupted with inter-track interference (ITI) that arises as a result of the overlapping of adjacent tracks. On the contrary, ITI in the existing PMR channel is negligible and are often ignored. To advance SMR/TDMR as a viable future recording architecture for the industry, equalization techniques should be formulated to exploit the ITI present and enhance the reliability of SMR/TDMR channels. The thesis therefore proposes novel 2D equalization schemes to manage the ITI in both channels.

Contents

A	List of Abbreviations	xi
B	List of Symbols	xiii
C	List of Tables	xvi
D	List of Figures	xviii
1	Introduction	1
1.1	Background	2
1.2	Motivations and Objectives	3
1.3	Major Contributions of Thesis	6
1.4	Organization of Thesis	10
2	Literature Review	12
2.1	Overview of Existing Magnetic Recording Architecture	13
2.1.1	Perpendicular Magnetic Recording	13
2.1.2	Media Trilemma	14
2.2	Overview of Emerging Magnetic Recording Architectures	16
2.2.1	Heat Assisted Magnetic Recording	16
2.2.2	Shingled Magnetic Recording	17
2.2.3	Two-Dimensional Magnetic Recording	18
2.3	Overview of Signal Processing Techniques.	19
2.3.1	Iterative Detector-decoder	21
2.3.2	Generalized Partial Response Equalizer	26

3	System Level Study of Future Recording Channels	30
3.1	Channel Model	31
3.1.1	Micromagnetic Model	32
3.1.2	Grain Flipping Probability Model	35
3.2	Shingled Magnetic Recording Channel	39
3.2.1	Varying Track Pitch	40
3.2.2	Varying Bit Length	45
3.3	Heat Assisted Magnetic Recording Channel	48
3.3.1	Varying Peak Temperature of Thermal Spot	50
3.3.2	Varying Diameter of Thermal Spot	54
3.3.3	Varying Alignment Between Thermal Spot and Magnetic Write Head .	56
3.4	Summary	58
4	Joint Viterbi Detector Decoder	61
4.1	JVDD	62
4.1.1	Algorithm	63
4.1.2	2D-JVDD	70
4.1.3	Managing Complexity of JVDD	73
4.2	JVDD Class of Codes	75
4.2.1	Gaussian Distribution Linear Diagonal Code.	76
4.2.1.1	Impact of GDLD Codes on JVDD	78
4.2.1.2	Simulations and Results for GDLD Codes	80
4.2.1.3	Optimization of GDLD Codes.	88
4.2.2	Proportionate Row Weight Codes	93
4.2.2.1	Optimization of Proportionate Row Weight Codes	94
4.3	Performance and Complexity Analysis of JVDD	96
4.3.1	JVDD Over a 1D ISI Channel	97
4.3.1.1	Performance	97

4.3.1.2	Complexity	99
4.3.2	JVDD Over a 1D GFP Channel	104
4.3.3	JVDD Over a 2D ISI Channel	106
4.4	Summary	108
5	Joint Factor Graph Detector Decoder	112
5.1	Algorithm	113
5.1.1	Factor Graph-Based Detector	115
5.1.2	Joint Factor Graph Detector Decoder	120
5.2	Mitigating Cycles in a Channel Factor Graph	124
5.2.1	Constraining Generalized Partial Response Target	124
5.2.2	Non-Binary Factor Graph-Based Detector	126
5.3	Simulations and Results	129
5.3.1	Factor Graph-Based Detector Over Uncoded 1D ISI Channel.....	130
5.3.2	JFGDD Over Coded 1D ISI Channel	132
5.3.3	JFGDD Over 1D Magnetic Recording Channel	134
5.4	Summary	138
6	Two-Dimensional Equalization Schemes for Magnetic Recording Channels	142
6.1	Equalization for Shingled Magnetic Recording Channel	144
6.2	Equalization for Two-Dimensional Magnetic Recording Channel	147
6.2.1	Symmetric TDMR Equalizer	148
6.2.2	Asymmetric TDMR Equalizer	152
6.2.3	TDMR Equalizer	155
6.3	Simulations and Results	158
6.3.1	Optimal Reader Pitch and Read Offset	159
6.3.2	Performance of Proposed Equalization Schemes	163
6.4	Summary	167

7	Conclusion and Future Work	170
7.1	Conclusion.	170
7.2	Future Work	174
7.2.1	Mitigation of Colored Noise	174
7.2.2	Hardware Implementation	175
7.2.3	Joint Detection and Decoding with Non-Binary Codes.	176
E	Publications and Patent	177
F	References	179

List of Abbreviations

1D	: One-Dimensional
2D	: Two-Dimensional
4D	: Four-Dimensional
ATE	: Adjacent Track Erasure
APP	: A Posteriori Probability
AWGN	: Additive White Gaussian Noise
BAR	: Bit Aspect Ratio
BCJR	: Bahl-Cocke-Jelinek-Raviv
BER	: Bit Error Rate
BPI	: Bits per inch
BPMR	: Bit Patterned Magnetic Recording
DDNP	: Data Dependent Noise Predictor
DOE	: Design of Experiment
ECC	: Error Correction Code
FER	: Frame Error Rate
FPGA	: Field Programmable Gate Array
GDLD	: Gaussian Distribution Linear Diagonal
GFP	: Grain Flipping Probability
GPR	: Generalized Partial Response
GPU	: General Processing Unit
HAMR	: Heat Assisted Magnetic Recording
HDD	: Hard Disk Drive
ISI	: Inter-Symbol Interference

ITI	: Inter-Track Interference
JFGDD	: Joint Factor Graph Detector Decoder
JVDD	: Joint Viterbi Detector Decoder
LDPC	: Low-Density Parity Check
LLR	: Log Likelihood Ratio
LMR	: Longitudinal Magnetic Recording
LMS	: Least Mean Square
LUT	: Look-up Table
MAP	: Maximum a Posteriori
ML	: Maximum Likelihood
MM	: Minimum Metric
MMLC	: Minimum Metric Legal Codeword
MMSE	: Minimum Mean Squared Error
MSE	: Mean Squared Error
NFT	: Near-Field Transducer
PDF	: Probability Density Function
PMR	: Perpendicular Magnetic Recording
PDNP	: Pattern-Dependent Noise Predictor
PR	: Partial-Response
PSD	: Power Spectral Density
RSM	: Response Surface Methodology
RHS	: Read Head Sensitivity
SMR	: Shingled Magnetic Recording
SNR	: Signal-to-Noise Ratio
SOVA	: Soft-Output Viterbi Algorithm
SPA	: Sum-Product Algorithm
SSD	: Solid State Drive
TPI	: Tracks per inch
TDMR	: Two-Dimensional Magnetic Recording

List of Symbols

λ	: Lagrange multiplier
σ	: sigma of JVDD code
σ_{grain}	: standard deviation of grain diameter
σ_{noise}	: sigma of AWGN noise
T	: JVDD threshold
ΔT	: temperature change in K
$\dot{y}$	: gradient of parabola of GDPD code at the bottom right corner
BL	: bit length
d_{min}	: minimum distance
d_x	: horizontal offset of main diagonal of GDLD code
d_y	: vertical offset of main diagonal of GDLD code
dis_x	: horizontal offset distance between thermal hot-spot and magnetic field
E_b	: user bit energy
E_c	: coded bit energy
$\mathbf{g}$	: target response
$\mathbf{h}$	: channel response
$\mathbf{H}$	: parity check matrix
$\mathbf{H}_{channel}$	: channel matrix
H_k	: coercive field
h_k	: channel coefficient
H_{write}	: effective writing field
l_{global}	: number of global iterations between detector and decoder
l_{dec}	: number of decoder iterations

k_B	: Boltzmann constant
K_o	: crystalline anisotropy of the FePt media at $T = 0$ K
K_u	: crystalline anisotropy
L_{bit}	: bit length
L_{mem}	: channel memory
M	: number of bits per symbol in non-binary factor graph
M_s	: saturation magnetization
M_{so}	: saturation magnetization of the media at $T = 0$ K
N	: codeword length
N_b	: number of bits per symbol in 2D-JVDD
N_{branch}	: number of branches emerging from a trellis node
N_{state}	: number of trellis states
N_{track}	: number of tracks
N_g	: target length
N_h	: channel length
N_w	: equalizer length
R	: code rate
R_{opt}	: operating code rate
RO	: reader offset
RP	: reader pitch
RW	: reader width
S_{max}	: JVDD maximum number of survivors
t	: error-correcting capability
T	: temperature in K
T_c	: Curie temperature
T_{peak}	: peak temperature of thermal hot-spot
T_{sig}	: width of thermal hot-spot
TP	: track pitch
w	: equalizer response

w_{col} : column weight
 w_{row} : row weight
 z_{pos} : position of constrained coefficient in a target

List of Tables

3.1	Comparison of computation time for a micromagnetic and GFP channel model. . .	39
3.2	Constant parameters used in SMR study	41
3.3	Target Response and SNR for various TP values	43
3.4	Error performance simulation parameters used for SMR study	43
3.5	Channel and storage density for varying TP	44
3.6	Target Response and SNR for various BL values	46
3.7	Channel and storage density for varying TP	48
3.8	Constant parameters used in HAMR study	51
3.9	Target Response and SNR for various T_{peak} values	51
3.10	Error performance simulation parameters used for HAMR study	53
3.11	Target Response and SNR for various T_{sig} values	55
3.12	Target Response and SNR for various dis_x values.	57
4.1	Constant parameters used in experimental study of GDLD codes	80
4.2	Variable parameters used in experimental study of GDLD codes	80
4.3	DOE table consisting of control and response parameter sets	90
4.4	Range of control parameters used in Phase 1 of DOE	91
4.5	Optimum set of control parameters obtained in Phase 1 of DOE	91
4.6	Range of control parameters used in Phase 4 of DOE	92
4.7	Optimum set of control parameters obtained in Phase 4 of DOE	92
4.8	Parameters used in JVDD study over 1D ISI channel	97
4.9	Variable parameters used in JVDD study over PMR channel	104
4.10	Constant parameters used in JVDD study over PMR channel.	105
4.11	Constant parameters used in 2D JVDD study.	106

4.12	Variable parameters used in 2D JVDD study.	107
5.1	Variable parameters used in study on Factor Graph-based detector over uncoded 1D ISI channel.	130
5.2	Constant parameters used in study on Factor Graph-based detector over uncoded 1D ISI channel.	130
5.3	Comparison of BCJR and Factor Graph detector complexity	131
5.4	Constant parameters used in JFGDD study over ISI channel	132
5.5	Variable parameters used in JFGDD study over ISI channel	132
5.6	Constant parameters used in JFGDD study with constrained targets over ISI channel	134
5.7	Variable parameters used in JFGDD study with constrained targets over ISI channel	134
5.8	Values of Z_{pos} constraints and the corresponding number of length-4 cycles present in the channel factor graph.	135
6.1	Constant parameters used in optimal RP and RO study	160
6.2	Variable parameters used in optimal RP and RO study	160
6.3	Optimum RP and RO for various TP for SMR channel	161
6.4	Optimum RP and RO for various TP for TDMR channel	161
6.5	Constant parameters used in performance comparison of the proposed equalization schemes	164
6.6	Variable parameters used in performance comparison of the proposed equalization schemes.	164
6.7	MSE values of various equalizers	165

List of Figures

1.1	Communication block diagram depicting the need for concurrent development of new recording architectures with new trends in signal processing.	2
2.1	Perpendicular magnetic recording	14
2.2	Areal density of HDD	15
2.3	Media trilemma involving SNR, thermal stability and ease of writing.	16
2.4	Writing mechanisms in HAMR	16
2.5	Tracks in (a) conventional PMR and (b) emerging SMR technology.	18
2.6	A block diagram of a magnetic recording channel.	20
2.7	An iterative detection/decoding scheme described in	21
2.8	A BCJR trellis.	22
2.9	An LDPC factor graph.	24
2.10	Illustration of length (a) 4 and (b) 6 cycles in a factor graph	26
2.11	A general partial response equalizer block diagram	27
3.1	Voronoi media	33
3.2	Mason William's and (b) Kanai's magnetic write head profile.	33
3.3	Micromagnetic output showing bits of alternating polarity written on a DC background.	34
3.4	Micromagnetic output for a (a) HAMR and (b) SMR channel consisting of pseudo-random data written on AC background.	35
3.5	Inclusion of surrounding bit patterns in the GFP LUT.	36
3.6	Grain flipping probability as a function of H_k	37
3.7	A (a) denominator array and a (b) numerator array that has been summed over the dimensions of H_k and bit patterns, resulting in a (c) GFP footprint as a function	38

of (x,y).	
3.8 Micromagnetic output for $BL = 15$ nm and $TP =$ (a) 26 nm (b) 42 nm.	41
3.9 GFP footprints for $BL = 15$ nm and $TP =$ (a) 26 nm (b) 42 nm	42
3.10 Impact of iterative detector-decoder iterations on system performance.	42
3.11 Error performance simulation results for varying TP and $BL = 15$ nm.	45
3.12 Micromagnetic output for $TP = 34$ nm and $BL =$ (a) 13 nm (b) 17 nm.	46
3.13 GFP footprints for $TP = 34$ nm and $BL =$ (a) 13 nm (b) 17 nm.	46
3.14 Error performance simulation results for varying BL and $TP = 34$ nm	47
3.15 Comparison of (a) media density and (b) storage density for various TP and BL values of an SMR channel.	48
3.16 Thermal profile of hot spot in a HAMR system.	50
3.17 Variation of (a) M_s and (b) K_u of magnetic grains with temperature [43].	50
3.18 Micromagnetic output for (a) $T_{peak} = 600$ K (b) $T_{peak} = 699$ K	52
3.19 GFP footprints for (a) $T_{peak} = 600$ K (b) $T_{peak} = 699$ K	52
3.20 Channel performance for various T_{peak} values.	53
3.21 Micromagnetic output for (a) $T_{sig} = 17$ nm (b) $T_{sig} = 29$ nm.	54
3.22 GFP footprints for (a) $T_{sig} = 17$ nm (b) $T_{sig} = 29$ nm.	54
3.23 Channel performance for various T_{sig} values.	55
3.24 Micromagnetic output for (a) $dis_x = 0$ nm (b) $dis_x = 9$ nm.	56
3.25 GFP footprints for (a) $dis_x = 0$ nm (b) $dis_x = 9$ nm.	56
3.26 Channel performance for various dis_x values.	58
4.1 Block diagram depicting difference between iterative detector-decoder and JVDD.	62
4.2 Computation of branch and path metric in the Viterbi algorithm.	66
4.3 Metric thresholding in the JVDD algorithm.	67
4.4 Effect of T on: (a) performance and complexity measured of the JVDD measured in terms of (b) average number of survivors and (c) total number of survivors.	67
4.5 The execution of the JVDD algorithm on a trellis consists of two parts – ‘metric thresholding’ at all nodes and ‘parity checking’ at designated parity check nodes.	68
4.6 Assigning parity check nodes in a JVDD trellis.	69

4.7	Metric thresholding and parity checking steps in the JVDD algorithm.	70
4.8	Block diagram depicting non-binary sequences for a 2D channel.	70
4.9	Number of survivors in a JVDD trellis for a short random LDPC code.	74
4.10	Effect of S_{max} on: (a) performance and computational complexity measured in terms of (b) average number of survivors and (c) total number of survivors in the trellis.	74
4.11	H matrix of a random LDPC code	75
4.12	Construction of H matrix of a GDLD code	77
4.13	Comparison of number of survivors for JVDD with LDPC and GDLD code each at length $N =$ (a) 512 and (b) 1024.	77
4.14	Comparison of a (a) sparse and (b) dense matrix of GDLD code.	78
4.15	Comparison of a (a) small and (b) large <i>offset</i> value of GDLD code.	80
4.16	Performance plots of GDLD codes with varying w_{row} and $\sigma =$ a) 50 b) 150 and c) 250.	81
4.17	A distribution plot of w_{col} of GDLD codes with $\sigma = 150$, <i>offset</i> = 150 and various w_{row}	82
4.18	Band-like structure in a deterministic GDLD H matrix.	82
4.19	Plots of d_{min} of GDLD codes with varying w_{row} and $\sigma =$ a) 50 b) 150 and c) 250. ...	83
4.20	Complexity plots of GDLD codes with varying w_{row} and $\sigma =$ a) 50 b) 150 and c) 250.	83
4.21	Performance plots of GDLD codes with varying σ and $w_{row} =$ (a) 50 (b) 150 (c) 250.	84
4.22	Plots of d_{min} of GDLD codes for various σ and $w_{row} =$ (a) 50 (b) 150 (c) 250.	84
4.23	Complexity plots of GDLD codes with varying σ and $w_{row} =$ (a) 50 (b) 150 (c) 250.	85
4.24	Performance plots of GDLD codes with varying <i>offset</i> and $w_{row} =$ (a) 50 (b) 150 and (c) 250.	85
4.25	Distribution plot of w_{col} in the H matrices of GDLD codes with $w_{row} = 150$, $\sigma = 200$ and various <i>offset</i>	86
4.26	Plot of d_{min} of GDLD codes for various <i>offset</i> and $w_{row} =$ (a) 50 (b) 150 and (c) 250.	86
4.27	Complexity plots of GDLD codes with varying <i>offset</i> and $w_{row} =$ (a) 50 (b) 150 and (c) 250.	87

4.28	Plot of number of JVDD survivors for GDLD code having parameters $w_{row} = 150$, $\sigma = 200$ and various <i>offset</i>	87
4.29	DOE technique attempts to optimize a response parameter given a set of control parameters.	88
4.30	Comparison of experimental and analytical values obtained in (a) Phase 1 and (2) Phase 4 of DOE	92
4.31	FER vs Complexity plot of optimum GDLD code obtained analytically from DOE against other code design parameters.	93
4.32	H matrix of a proportionate row weight JVDD code.	94
4.33	FER vs Complexity plots of optimum proportionate row weight code at $R = 0.90$ and $N =$ (a) 512 and (b) 2048	94
4.34	H matrices of optimum proportionate row weight code at (a) $N = 512$ and $N =$ (a) 512 and (b) 2048.	95
4.35	Comparison of optimized JVDD codes in terms of (a) FER vs Complexity at a fixed channel SNR and (b) FER vs SNR at a fixed complexity.	95
4.36	Performance comparison between JVDD and iterative detector over an ISI channel at various SNR, R and $N =$ (a) 512 (b) 1024, (c) 2048 and (d) 4096.	98
4.37	Number of multiplications in the JVDD and Iterative Detector for (a) $N = 512$ and (b) $N = 2048$	100
4.38	Average number of survivors in the JVDD trellis at (a) $N = 512$ and (b) $N = 2048$. . .	101
4.39	Number of additions in the JVDD and Iterative Detector for (a) $N = 512$ and (b) $N = 2048$	102
4.40	Computational time of the JVDD and Iterative Detector for (a) $N = 512$ and (b) $N = 2048$	103
4.41	(i) Performance and (ii) complexity of JVDD over a PMR channel at various R and (a) $N = 512$ (b) $N = 024$ and (c) $N = 2048$	105
4.42	Block diagram depicting comparison between 2D-JVDD and 2D iterative detector-decoder.	106
4.43	Performance of 2D-JVDD over a 2D ISI at various SNR and $N =$ (a) 64 and (b) 128. . .	107

4.44	Complexity of 2D-JVDD over a 2D ISI measured in terms of (i) average number of survivors and (ii) total number of survivors for $N =$ (a) 64 and (b) 128.	108
5.1	Block diagram depicting difference between conventional iterative detector-decoder, JVDD and JFGDD	113
5.2	Conversion of a channel trellis into a channel factor graph and subsequent integration of the channel factor graph with a code factor graph to develop a super factor graph structure required by a JFGDD	114
5.3	A channel factor graph consisting of channel and variable nodes.	115
5.4	Message passing from (a) variable node to channel node and from (b) channel node to variable node in a channel factor graph.	117
5.5	Varying number of iterations in a binary factor graph detector.	119
5.6	A super factor graph structure consisting of channel, variable and check nodes. .	120
5.7	Message passing in JFGDD from (a) variable node to channel node and from (b) variable node to check node in the JFGDD.	121
5.8	Detecting presence of length-4 cycles in a channel matrix.	123
5.9	Detecting presence of length-6 cycles in a channel matrix.	123
5.10	Channel factor graph free of length-4 cycles as a result of constraining the third and fifth coefficient of its designed target to 0.	125
5.11	Channel factor graph free of length-4 cycles as a result of constraining the second and fourth coefficient of its designed target to 0.	126
5.12	Channel factor graph of a non-binary factor graph-based detector designed with $M = 2$	127
5.13	Channel factor graph of a non-binary factor graph-based detector designed with $M = 3$	128
5.14	Comparison of performance among various detectors over an uncoded channel.	131
5.15	Comparison of performance among various detector-decoder systems over a coded 1D ISI channel	133
5.16	Comparison of performance among various detector decoder.	136
5.17	PSD plots of noise arising from various constraints.	137

5.18	Comparison of performance among various detector decoder systems over a 1D magnetic recording channel.	136
5.19	BSC channel model as a theoretical approximation to an uncoded magnetic recording system.	137
5.20	Plot depicting the theoretical approximation of the achievable user density of a conventional PMR system when used with various detection and decoding algorithms	138
6.1	Overview of equalization schemes for SMR and TDMR channels.	143
6.2	Block diagram depicting an SMR equalizer.	145
6.3	Block diagram depicting a Symmetric TDMR equalizer.	149
6.4	Block diagram depicting an Asymmetric TDMR equalizer.	153
6.5	Block diagram depicting a TDMR equalizer.	155
6.6	Illustration of reader pitch and read offset for a triple reader head configuration.	159
6.7	Contour plots depicting the relationship between RP , RO and SNR for an SMR equalizer at $TP =$ (a) 26nm (b) 30 nm (c) 34 nm (d) 38 nm.	161
6.8	Contour plots depicting the relationship between RP , RO and SNR for a 1D Symmetric TDMR equalizer at $TP =$ (a) 26nm (b) 30 nm (c) 34 nm (d) 38 nm.	162
6.9	FER plots comparing the performance of the (a) conventional PMR equalizer with the proposed (b) SMR, (c) Symmetric TDMR and (d) TDMR equalizers integrated with the iterative detector-decoder and JVDD at $N =$ (i) 1024 and (ii) 2048.	167
7.1	Auto-regressive channel model for noise whitening [58].	175

Chapter 1

Introduction

The demand for Hard Disk Drive (HDD) storage has been increasing rapidly in recent years due to various trends such as the vast growth of graphics and multimedia applications, and the massive increase in computer storage requirements of network data bases [1][2].

The HDD industry has, however, experienced a slowing in growth in storage capacity over the past few years. New technologies such as Solid State Drives (SSDs) have brought about significant competition to this industry although HDD remains as the least expensive large capacity storage option to date [3].

The increase in storage capacity of HDDs in the past has been mainly attributed to an increase in areal density (the amount of data that can be physically stored in a square inch of a disk platter [4]). The ability to continue to scale conventional magnetic recording technology, known as Perpendicular Magnetic Recording (PMR), to higher areal densities is limited by the media trilemma problem [4][5][6].

This chapter presents an overview of the challenges faced by the magnetic recording industry. The motivation and objectives of this thesis are then discussed, followed by a summary of the major contributions of the thesis. The chapter concludes with an outline of the thesis.

1.1 Background

A practical HDD system is made up of a magnetic recording channel which writes and reads information on a storage media as well as various signal processing components that aim to cancel out distortions introduced by the channel (Figure 1.1). The recording channel has made a huge leap in storage growth since 1956 [3]. Likewise, signal processing for HDD has made significant advances since 1990s [7]. Breakthroughs such as iterative detection schemes played a crucial role in improving the performance (and therefore reliability) and storage capacity of HDD [7].

Figure 1.1: Communication block diagram depicting the need for concurrent development of new recording architectures with new trends in signal processing.

The HDD is however approaching its practical limit in areal density in the near future [5]. The challenge in bringing about significant areal density gains lies in the ability of the existing PMR channel to retain the magnetization of bits despite thermal fluctuations. The limitations of the state-of-the-art signal processing technologies [7][8][9] such as sub-optimality in performance hinder it from assisting in a new breakthrough in storage capacities of the HDD.

The magnetic recording industry needs to invent new ideas across the entire HDD platform to attain massive storage improvements. Developments of new recording architectures, concurrent with the development of new trends in signal processing are essential in paving the way to new and intriguing prospects of significant performance gains.

1.2 Motivations and Objectives

To enable a new breakthrough in areal density, the magnetic recording industry has to leverage on the synergy between new magnetic recording architectures and advanced signal processing algorithms. New magnetic recording technologies such as Heat Assisted Magnetic Recording (HAMR), Shingled Magnetic Recording (SMR), Two-Dimensional Magnetic Recording (TDMR) and Bit Patterned Magnetic Recording (BPMR) are recently proposed in the Literature in recent years [10][11][12][13]. This thesis aims to study the feasibility of advancing the recently proposed magnetic recording technologies as a next-generation high density recording channel. In addition to that, in view of the present limitations of the state-of-the-art detection and equalization schemes, this thesis aims to propose novel signal processing algorithms for the HDD. The motivations and objectives of this thesis are described in greater detail as follows:

Feasibility of new recording architectures

Developing a realistic channel model is essential in enabling a comprehensive analysis of the feasibility of implementing a newly proposed recording architecture. It can address a broad set of parameters, spanning various components of the magnetic recording system such as the writer, reader and media. Micromagnetic modeling is one of the tools used by many researchers to predict the behavior of magnetic particles when subjected to a magnetic field. This tool has been extensively applied to magnetic recording scenario and much work has been done by others on micromagnetic simulation of various magnetic recording systems [14][15][16][17]. In [14] and [15] for example, the authors use micromagnetic simulations to estimate the attainable areal density for SMR using various new writer and reader heads. In these work, the authors also performed an optimization of their writer/reader heads based on the magnetic phenomena captured by the model. In [16], micromagnetic was used to study the magnetization reversal in a thin film magnetic recording media that is used by the HDD industry today. While the micromagnetic model is generally considered accurate, it is highly time consuming. To this end, the micromagnetic model is only useful for studying the magnetic

behavior of an isolated channel. It cannot be used for a system level study (known as error performance simulations) that integrates coding and signal processing techniques with the channel. Error performance simulations enable the analysis of Frame Error Rate (FER) or Bit Error Rate (BER) and are necessary for advancing HAMR, SMR and TDMR as a viable future recording technology. That aside, the inadequacy of the micromagnetic model for error performance simulations also hinders the evaluation of new signal processing techniques that are integrated with a given channel.

To enable error performance studies, researchers commonly use a generic mathematical model in place of a micromagnetic model, at the expense of accuracy. In [18] and [19] for example, a mathematical channel model was adapted to study and test signal processing algorithms at a system level for BPMR and TDMR respectively. In [20], a mathematical modeling of the channel was carried out to enable the characterization of magnetic noise in a system that is integrated with multiple reader arrays. Little progress has been made in developing alternative models. In [21] and [22], novel channel models specific for TDMR such as the Four-Rectangular-Grain Model model have been proposed in place of a micromagnetic or simplified mathematical model. In this thesis, a less time consuming alternative to the micromagnetic model that has a much higher accuracy than a mathematical model is proposed in this thesis. This is known as the Grain Flipping Probability (GFP) model. The novel GFP model can be applied to any recording channel. Using this model, a novel system level study of the SMR and HAMR recording is carried out. Various gains and limitations of the said channels were addressed with great accuracy from a system's perspectives.

Development of new signal processing techniques

The development of new signal processing techniques alongside the evolvement of new magnetic recording technologies is critical in bringing about powerful gains. The iterative detector-decoder is the standard used in HDD industry today [7][8][9]. In this scheme, detection and decoding are performed iteratively, passing information from the detector to the decoder back and forth. As the detector has knowledge of the channel while the decoder has

information about the code, the detector and decoder exchanges mutual information repetitively in order to recover encoded information written on the media. The detector operates on a trellis while the decoder works on a factor graph. The LDPC decoder that is typically used is sub-optimal due to the presence of cycles in its factor graph. The iterative nature of this scheme further asserts a lack of optimality.

In view of this, researches have explored the possibility of joint detection and decoding [23][24][25][26]. In a joint detector-decoder system, the detector and decoder are represented as one unit instead of two separate entities. The joint detector-decoder unit has knowledge of both the channel and the code. It is therefore able to jointly detect and decode bits with better performance. Various novel joint detection and decoding algorithms have been proposed in the literature. In [23], a soft-input David Mackay inner decoder was concatenated with a BCJR detector to form a detector block that is subsequently integrated with an outer LDPC decoder to simultaneously detect and decode read-back signal over a BPMR channel. In [24], modifications were made to the decoder factor graph to enable joint detection and decoding over a non-binary LDPC coded ISI channel. A more popular approach in the area of joint detection and decoding has been to design a super factor graph [24][25] which represents both the channel and the code and employing the message passing algorithm (MPA) for simultaneous detection and decoding over the super factor graph. Factor graph based algorithm for joint detection and decoding can generally be devised in various ways. For example, in [25], a joint detector-decoder system that estimates a posteriori probabilities of noiseless channel output symbols rather than binary channel inputs was designed and found to perform significantly better than the conventional iterative detector-decoder at the expense of greater computations. In this thesis, an alternative implementation of factor graph based detection and decoding is proposed. This is coined the Joint Factor Graph Detector Decoder (JFGDD). The factor graph based techniques proposed in the literature and in this thesis faced a common challenge of performance degradation posed by the presence of short cycles in the factor graph. In [26], a 'stretching' technique was introduced to transform short cycles into longer cycles, and a resulting performance convergence with the conventional system was reported. This thesis proposes two alternative methods of reducing the short cycles – 1)

constraining Generalized Partial Response (GPR) targets and 2) implementation of a non-binary factor graph based on symbol probabilities.

In addition to that, in view of the factor graph approach commonly taken by many researchers, this thesis explores an alternative method of performing joint detection and decoding. That is, a novel trellis based algorithm that simultaneously detects and decodes bits in a single step is proposed. This is coined the Joint Viterbi Detector Decoder (JVDD). To study the effectiveness of the two proposed joint detection and decoding approach, the GFP channel model was employed for a system level analysis of the proposed schemes.

Aside from detection and decoding, advances in equalization techniques can also be made. The GPR equalizer used in today's recording system is only applicable to one dimensional (1D) channels [27]. Little advancements have been made in recent years on new two-dimensional (2D) equalization schemes for 2D channels. In [28], the authors introduce a neural-network equalization approach for TDMR channels. No two-dimensional (2D) GPR equalization scheme has been devised at present as the existing PMR channel is a 1D channel. As such, this thesis proposes various 2D GPR equalization schemes that are appropriate for the next generation 2D recording channels, namely SMR and TDMR.

1.3 Major Contributions of Thesis

The main contributions of this thesis are summarized as follow:

1. System level study of future recording channels

In view of the shortcomings of the micromagnetic model discussed in Section 1.2, a novel GFP channel model was developed in this chapter in place of the micromagnetic model. The GFP model uses micromagnetic simulations to estimate the probability of grains flipping in a given circumstance. The circumstance is characterized by the grains' relative position to the write field, the surrounding bit patterns and magnetic coercivity. The model characterizes a look-up

table (LUT) of grain probabilities from micromagnetic simulations. It can then be used to produce signals that accurately model the noise characteristics of granular media, the dominant noise source in ultra-high density magnetic much faster than via micromagnetic simulations. Error performance studies would otherwise have been impossible. The ability to carry out error performance studies using an accurate model in a reasonable amount of time enables a comprehensive analysis of HAMR and SMR recording channels and the signal processing components associated with it. The GFP model was also used in the rest of the thesis to allow signal processing algorithms that are proposed to be evaluated.

The use of the GFP channel model for a system level study of the HAMR channel was presented in the following conference and published in the following journal:

- S. S. Shafi'ee, E. M. Rachid, K. S. Chan, H. T. Wang and E. Kwaku, "Application of the grain flipping probability model to heat assisted magnetic recording", 56th Conference on Magnetism and Magnetic Material, Arizona, USA, pp. CF-11, October–November 2011.
- S. S. Shafi'ee, E. M. Rachid, K. S. Chan, H. T. Wang and E. Kwaku, "Application of the grain flipping probability model to heat assisted magnetic recording", J. Appl. Phys., vol. 111, no. 7, pp. 07B714 - 07B714-3, April 2011.

2. Joint Viterbi Detector Decoder

In view of the sub-optimality of the state-of-the-art iterative detection/decoding technique, two joint detection and decoding techniques are proposed in this thesis. The first proposed technique involves the JVDD. In this technique, detection and decoding are carried out simultaneously in one step, over a channel trellis. The proposed JVDD is optimal over a coded ISI channel. Its high computational complexity is however a drawback. To manage this, a new class of codes known as the JVDD class of codes was proposed specifically for the JVDD. By employing this class of codes in this thesis, the JVDD reaped a significant reduction in complexity. The choice of the design parameters of the codes was however crucial in

determining the performance and complexity outcome of the JVDD. As such, extensive simulations were carried out together with theoretical analysis such as free distance computations to study the impact of the code design parameters on JVDD. The study indicated that an optimal set of design parameters of the codes exists and it gave JVDD its best performance and lowest complexity. The proposed codes were thus analytically optimized using a Response Surface Methodology (RSM). The key results for the study on JVDD codes were presented at the following conferences:

- S. S. Shafi'ee, K. S. Chan and Y. L. Guan, "A Performance Study of Joint Viterbi Detector Decoder (JVDD) with GDLD codes", International Conference on Computational Intelligence, Computing and Signal Processing (CICSP), Hong Kong, China, pp. 339 – 343, 19 - 20 October 2014.
- S. S. Shafi'ee, K. S. Chan and W. L. Goh, "Optimization of Codes for the Joint Viterbi Detector Decoder", accepted for publication in International Conference on Computing, Networking and Communications (ICNC) 2016.

Using the optimized codes, the performance of the JVDD was benchmarked against the state-of-the-art iterative detector-decoder through intensive computer simulations over a 1D ISI channel as well as a 1D PMR channel using the GFP channel model, at various channel conditions, codeword length and code rate. A modification of the JVDD algorithm was subsequently proposed for application to 2D channels. The proposed 2D scheme was coined the 2D-JVDD. Its performance and complexity was investigated over a 2D ISI channel. The study on JVDD was presented in the following conferences:

- S. S. Shafi'ee, K. S. Chan, Y. L. Guan and E. M. Rachid, "Joint Viterbi Detector Decoder for Grain Flipping Probability Model", 24th Magnetic Recording Conference (TMRC 2013), Tokyo, Japan, pp. 14, 20 -22 August 2013.
- K. S. Chan, S. S. Shafi'ee and Y. L. Guan, "Optimal Joint Viterbi Detector Decoder (JVDD) over ISI Channels", International Conference on Computing, Networking and Communications (ICNC) 2014, Hawaii, USA, pp. 282 - 286, 3 – 6 February 2014.

3. Joint Factor Graph Detector Decoder

The second joint detection and decoding technique proposed in this thesis involves the JFGDD. In this technique, detection over a channel factor graph and decoding over a code factor graph are simultaneously carried out to recover transmitted bits. Despite having the advantage of performing detection and decoding jointly, the performance of the JFGDD is susceptible to the presence of short cycles in the factor graphs. Existing Error Correction Code (ECC) technique was used to remove cycles in the code factor graph. Two methods were proposed to remove the cycles in the channel factor graph. First, a non-binary transformation of JFGDD was proposed by grouping consecutive bits in a stream so as to reduce the number of edges and therefore cycles in the channel factor graph. Second, designing a constrained target using a GPR equalizer was proposed to arrive at a channel factor graph with fewer or no cycles. Simulations were extensively carried out over a 1D ISI channel as well as PMR channel to evaluate the performance of the JFGDD, non-binary JFGDD and JFGDD with GPR equalizer against the state-of-art system. The study on JFGDD was presented in the following conference and journal:

- S. S. Shafi'ee, E. M. Rachid, K. S. Chan and Y. L. Guan, "Application and Optimization of Factor Graph-Based Detector on 1D ISI Magnetic Recording Channel", Asia-Pacific Magnetic Recording Conference 2012, Singapore, 31 October – 2 November 2012.
- S. S. Shafi'ee, E. M. Rachid, K. S. Chan and Y. L. Guan, "Application and Optimization of Factor Graph-Based Detector on 1D ISI Magnetic Recording Channel", IEEE Transactions on Magnetics, vol. 49, no. 6, pp. 2500 - 2503, June 2013.

4. Equalization/Detection Schemes for 2D Channels

In this thesis, various 2D GPR equalization and detection schemes were proposed for 2D channels. The 2D equalizers which consist of an SMR equalizer, symmetric and asymmetric TDMR equalizer, and 2D TDMR equalizer were designed with the aim of either equalizing away the effect of Inter-Track Interference (ITI) and employing 1D detection for ISI cancellation or making use of 2D detection techniques to simultaneously undo the effects of ITI and ISI. In this thesis, each of the proposed equalizers was integrated with the JVDD and the conventional

iterative detector-decoder in different set-ups. The equalization/detection schemes were studied and their performance for a triple reader head system was analyzed over a 2D GFP channel model through a large number of simulations for various code and channel conditions. The results for the equalization/detection schemes for 2D channels were presented in the following conference and journal:

- S. S. Shafi'ee, K. S. Chan, Y. L. Guan, "Novel Joint Detection Schemes for TDMR Channels", 25th The Magnetic Recording Conference (TMRC 2014), California, USA, 11 - 13 August 2014.
- S. S. Shafi'ee, K. S. Chan, Y. L. Guan, "Novel Joint Detection Schemes for TDMR Channels", accepted for publication in IEEE Trans. on Magn., vol. 51, no. 4, pp. 0018-9464, April 2015.

1.5 Organization of Thesis

The rest of the thesis is organized as follow:

Chapter 2 provides a literature review on the existing PMR technology and the media trilemma problem that it faces. An overview of emerging magnetic recording architectures as well as the state-of-art signal processing technologies is then presented.

Chapter 3 proposes a novel GFP channel model as an alternative to the commonly used micromagnetic model. The GFP model was integrated with the state-of-the-art signal processing components to realistically analyze a recently proposed SMR and HAMR technology from a system's perspective. The GFP channel model was employed in subsequent parts of the thesis for integration and evaluation of proposed coding and signal processing techniques.

Chapter 4, 5 and 6 focuses on advanced signal processing techniques that can be synergized with the development of new recording architectures.

In Chapter 4, a novel joint detection and decoding scheme known as the JVDD is proposed as an alternative to the state-of-the-art iterative detector-decoder. While the JVDD has shown to perform well, its computational complexity needs to be properly managed. The chapter proposes various ways of managing JVDD's complexity. This includes proposing a new class of error-correction codes well suited for JVDD and optimizing the design parameters of the code through an analytical technique. As the JVDD is only applicable to 1D channels, a proposed non-binary adaptation of the algorithm for 2D channels is presented in this chapter. The performance of the JVDD algorithm was benchmarked against the state-of-the-art iterative detector-decoder over 1D ISI, 1D PMR and 2D ISI channels through the use of extensive computer simulations.

Chapter 5 proposes an alternative joint detection and decoding scheme – JFGDD. Unlike the JVDD which operates on a trellis, the JFGDD operates on a factor graph. By so doing, its performance is susceptible to the presence of short cycles. Two methods of mitigating these cycles are proposed in this chapter. These include a non-binary classification of information that results in fewer edges and therefore fewer cycles in the factor graph, as well as designing a constrained target using a GPR equalizer to arrive at a graph with fewer cycles. Computer simulations were carried out to benchmark the performance and complexity of the JFGDD against the state-of-the-art iterative detector-decoder over a 1D ISI and PMR channel.

Chapter 6 presents novel equalization schemes for 2D magnetic recording channels. In this chapter, various equalizer designs for handling ITI were proposed for emerging SMR and TDMR channels. The relative performances of the proposed schemes were then compared through simulations.

The thesis concludes with Chapter 7 where a summary of the major contributions is provided alongside the key results of the work undertaken in various chapters. A future work plan is laid out to address possible upcoming research work in view of the shortcomings of some of the proposed ideas/results in this thesis.

Chapter 2

Literature Review

The existing Perpendicular Magnetic Recording (PMR) technology is approaching its maximum storage limit in the near future. New magnetic recording architectures need to be developed to sustain a continual growth in areal density of Hard Disk Drives (HDDs). The Heat Assisted Magnetic Recording (HAMR), Shingled Magnetic Recording (SMR), Two-Dimensional Magnetic Recording (TDMR) and Bit Patterned Magnetic Recording (BPMR) which were proposed in [10][11][12][13] as an alternative to conventional PMR have gathered much interest in recent years. This chapter provides a literature review of the traditional PMR as well as the emerging magnetic recording technologies. The scope of the PhD research work undertaken involves the PMR, HAMR, SMR and TDMR channel. BPMR has been perceived as a technology that has more challenges and is generally taken to be a longer-term technology than HAMR, SMR and TDMR. Therefore in this thesis, focus is given on nearer-term technology.

Advanced signal processing algorithms need to be formulated in parallel with the development of new magnetic recording channels in order to bring about impactful gains in storage density in the future. A literature review of the state-of-the-art signal processing techniques and their limitations is presented in the second part of this chapter. Novel signal processing algorithms that are developed in this PhD work will be presented in Chapter 4, 5 and 6.

2.1 Overview of Existing Magnetic Recording Architecture

This section provides a literature overview of the PMR technology used in today's HDDs. The challenge faced by the PMR technology in achieving higher storage capabilities, known as the media trilemma problem, is explained in this section.

2.1.1 Perpendicular Magnetic Recording

PMR is an existing technology used in a HDD in the market today. PMR technology showed an extraordinary advantage over an older recording technology known as Longitudinal Magnetic Recording (LMR) [29]. In a PMR channel, information bits are stored by magnetizing grains on a media in different directions through the use of a magnetic write head that is driven by a current source. The amplitude of the current should be sufficiently large to magnetize the media to saturation and the direction of the applied current determines the resulting polarity of bits written. The magnetization of bits on the media points in a direction perpendicular to the surface of the disk (this is in contrast to an LMR channel where the direction of magnetization is parallel to the disk surface) [30][31]. An illustration of the writing mechanism is shown in Figure 2.1 [32]. Information bits are written sequentially on the tracks of a spinning disk. The speed of rotation of the disk determines the spacing between bits. During a read-back process, a reader head which is positioned over the spinning disk is used to sense the magnetization patterns on the media. The read-back signal will often be distorted by various noise and interferences. The noise fundamentally arises from the writer and reader head, the electronic circuitry and media noise.

In the earlier years of HDD manufacturing, the PMR technology showed an unprecedented gain over LMR [29]. The HDD roadmap has however approached a critical point where the existing PMR architecture is no longer capable of sustaining a continual growth in storage capacity. The challenge faced by the existing PMR technology is known as the media trilemma problem. The next section explains the media trilemma in greater detail.

Figure 2.1. Perpendicular magnetic recording [32].

2.1.2 Media Trilemma

The storage capacity of a HDD is measured in terms of its areal density. The areal density refers to the amount of data that can be physically stored in a square inch of a disk platter. It is measured in terms of the number of bits packed per inch (bit density) and the number of tracks per inch (track density). Figure 2.2 provides an illustration of the areal density in a HDD (Note that the densities shown in this figure are for illustrative purposes only. Densities in real HDDs are much higher than depicted). Achieving significant gains in areal density requires the downsizing of bits as well as tightening of tracks, as depicted in Figure 2.2. The challenge in doing so arises from the difficulty in simultaneously attaining higher areal densities, ensuring thermal stability of grains and ease of writing data. This is known as the media trilemma problem [4][5][6].

A reduction in bit size requires a corresponding reduction in the grain size on the media in order to maintain a good media Signal-to-Noise Ratio (SNR) ($\text{SNR} \propto \sqrt{\text{number of grains per bit}}$) [6]. However, as grain sizes are reduced, their magnetic energy barrier, given by $K_u V$ (where K_u is the magnetic anisotropy energy holding the magnetization in its easy axis direction [30] and V is the volume of grains), decreases as well. In order to maintain thermal stability, the energy required to flip a grain needs to be significantly higher than the thermal energy in the system. Eventually, as the grains become so small, their magnetic energy approaches the thermal energy, given by $k_B T$ (where k_B is the Boltzmann constant and T is the absolute temperature).

When this occurs, the direction of magnetization of the grains may flip spontaneously due to ambient temperature, causing data written on magnetic disk to be lost. Ideally, the ratio of magnetic energy of grains to thermal energy should be greater than 60 to ensure data longevity on the order of 10 years [34]. In order to prevent magnetic instability of grains, the media anisotropy and therefore coercivity has to be increased.

Figure 2.2: Areal density of HDD.

A media with a higher magnetic coercivity however requires a stronger head field to flip grains during the write process. Unfortunately at present, smaller writer heads that are needed for higher density produce smaller fields. The maximum available head field is 2.4 T and no materials are available at present to further increase the strength of the write field [34]. Figure 2.3 summarizes the media trilemma illustratively.

The existing PMR technology is predicted to hit its limit of about 1 Tb/in^2 in the not too distant future [35][36][37]. The HDD industry needs to transition to a new approach that will allow HDD manufacturers to continue on the path of massive storage improvements. New magnetic recording architectures such as HAMR, SMR and TDMR have been proposed [10][11][12][13] to support this growth. A literature overview of these emerging technologies is presented in the next section.

Figure 2.3: Media trilemma involving SNR, thermal stability and ease of writing.

Figure 2.4: Writing mechanisms in HAMR [40].

2.2 Overview of Emerging Magnetic Recording Architectures

In this section, an overview of three emerging recording technologies (HAMR, SMR and TDMR) is presented. HAMR targets for an increase in bit density by coupling thermal field and magnetic field together. SMR targets for an increase in track density by devising a new method of writing data on a conventional disk. TDMR addresses the media trilemma from a coding and signal processing perspective.

2.2.1 Heat Assisted Magnetic Recording

HAMR technology aims to increase the density by addressing the media trilemma problem. It requires the conventional magnetic media to be replaced with one that has a high anisotropy such as Iron Platinum (FePt) alloy [38]. A HAMR media will thus have a high thermal stability, even at small grain sizes. This enables HAMR to meet the SNR criterion as the number of grains per bit will not drastically reduce and the bit size can reduce in proportion with the grain size.

As conventional heads cannot be used to write on a high anisotropy media, HAMR injects thermal energy into the grains during writing via a laser. In other words, a thermal hot spot is

produced in a HAMR system by means of a near-field transducer (NFT) that couples light from the laser with the media to produce thermal spots on the order of a few tens of nanometers in diameter, needed for writing very small bits. The presence of thermal energy helps to lower the coercivity of the media. At the instance of reduced coercivity, the magnetic field from an ordinary write head becomes sufficient to magnetize the grains on the media during the write process. Optimal writing requires that the laser operates near the Curie temperature of the media, where it will effectively lose much of its coercivity [39]. Upon completion of writing, the disk is rotated away from the laser beam. As the laser is only applied to a small section of the disk, rapid cooling occurs and the coercivity of the media rises. The recorded bits then stabilize at the storage temperature. The heating and cooling process that occurs during and after writing respectively is summarized in Figure 2.4 (adapted from [40]).

The newly proposed recording technique has been forecasted to extend the areal density of HDDs well beyond 1 Tb/in² [41][42][43]. Manufacturing reliable media composed of much smaller grains with higher coercivity and recording heads that can withstand the high temperature used during the write process would, however, be a daunting task. Getting the right intensity and location for the laser beam and its alignment with the magnetic field would pose significant challenges too.

2.2.2 Shingled Magnetic Recording

In PMR, magnetic tracks are written parallel to each other and are separated by an inter-track spacing (Figure 2.5a) to prevent adjacent track erasure (ATE) and inter-track interference. This reduces the areal density as some of the disk surface area is used only to buffer one track from the next. SMR technology was thus proposed in view of this.

In SMR technology, the number of tracks per inch on a disk is increased by overlapping the tracks. SMR writes a new track of data that overlaps with a segment of a previously written track as shown in Figure 2.5b. The overlapping of tracks is akin to shingles on a roof, hence the name shingled recording. With the exception of the last written track (known as the ‘fat track’),

this recording technique results in thinner written tracks, granting a higher areal density [44]. The main advantage of SMR lies with the ability of using larger writers to write narrow tracks. Wide writers however tend to have less steep gradients that would degrade the quality of the previously written track or bit [45]. Designing wide writers with sharp gradient is thus one of the challenges for SMR [46][47].

Figure 2.5: Tracks in (a) conventional PMR and (b) emerging SMR technology.

Conventional read-back in PMR systems or TDMR read-back can be utilized for SMR architecture. For read-back using ordinary PMR technique, written bits are recovered in the form of a waveform that is picked up by a reader head placed along the center of a track of interest. Read-back using TDMR technique will be explained in the next section. Modern reader heads require only a fraction of a written track width in order to reliably retrieve stored data. Although an SMR write process leaves thinner written tracks, the width of the exposed track is sufficient for the reader to pick up signals. Conventional reader heads used in PMR can thus be used in SMR.

2.2.3 Two-Dimensional Magnetic Recording

The SMR technique presented in the previous sub-section is a new writing technology where the read-back process can be done either through one-dimensional (1D) read-back as in ordinary PMR systems or 2D read-back. The technique of obtaining and processing a 2D read-back is known as TDMR [44]. Unlike conventional PMR systems which gather 1D read-back waveform from individual tracks, the TDMR reading technique involves building up a 2D signal

from a shingled block of data. In other words, information is read and processed one track at a time for PMR and multiple tracks at a time for TDMR.

For a 1D read-back signal with 1D signal processing, any interference from adjacent tracks, known as Inter-Track Interference (ITI) is inevitably destructive and will result in relatively severe performance degradation. Having 2D read-back data however enables a more complete picture to be extracted from multiple tracks. The ITI picked up during the read process contains information about the written data on adjacent tracks which can be used in the detection of that data using advanced coding and signal processing techniques such as in [48][49].

Gathering of 2D read-back signal during the read process in TDMR can be done in two ways – 1) place multiple reader heads on a slider to read a 2D block in one pass 2) spin the disk for multiple revolutions and collect the signal with a single reader head at each revolution. The former method would require construction of multiple reader elements while the latter results in increased reading latency proportional to the number of tracks read [50][51].

One advantage of TDMR is that the codeword length can be much longer than that for single track read-back by wrapping around the longer codeword onto multiple tracks. Longer codes are known to perform better [34]. In addition to that, TDMR enables reading by a wide reader as ITI will be mitigated by 2D detectors. It has been forecasted that the TDMR technology could bring about areal density gains of up to 2 Tb/in² [50][51][52].

2.3 Overview of Signal Processing Techniques

Increase in recording density of HDDs can be supported by advances in signal processing techniques. More specifically, advanced detection, decoding and equalization techniques can be introduced into HDDs to improve error-rate performance and reduce cost of manufacturing in ultra-high density magnetic recording.

A magnetic recording system can typically be represented by a block diagram in Figure 2.6. User information in the form of binary bits, u_k , are first encoded by an Error Correction Coding (ECC) encoder where redundant bits are added to generate a coded sequence, c . The redundant bits enable a decoder to detect and correct errors in the data stream. The coded sequence is then written (stored) onto the media. When the stored information needs to be retrieved, a reader is scanned over the medium and a signal is picked up. This signal is known as a read-back signal and the process of retrieving this signal is called the read-back process. Additive noise, media noise and Inter Symbol Interference (ISI) are generally present in read-back waveforms of a storage system. Additive noise is generated by electronic components in the system and by the reader while media noise, which is the major source of noise, arises from shifts in the location of bit transitions due to grains on the media. ISI occurs as the reader is sensitive to a certain region below it and multiple consecutive written bits impact the read-back signal at any time instance. At high track densities, the read-back signal also suffers from ITI. Electronic noise in a recording system can be modeled as Additive White Gaussian Noise (AWGN) with two sided Power Spectral Density (PSD) $= \frac{N_o}{2}$ [53]. Media noise, on the other hand, is captured by the micromagnetic and Grain Flipping Probability (GFP) models adopted in this thesis. Following read-back, the equalizer then shapes the channel to a target response of shorter length to allow for practical implementation of a detector. The detector attempts to undo the effects of the ISI channel and recover the coded sequence previously written onto the media. Following this, the decoder tries to estimate the binary data bits, $\hat{u}_k$.

Figure 2.6: A block diagram of a magnetic recording channel.

ECC codes that are typically used in a magnetic recording system include Low-Density Parity Check (LDPC) [54][55] and Quasi-Cyclic LDPC [56][57] codes. For detection and decoding, an iterative detector-decoder scheme is the standard used today [7][8]. This scheme has shown

significant coding gains over a scheme that uses a single step of detection and decoding [58] depicted in Figure 2.6. As for equalization, a General Partial Response (GPR) equalizer is commonly used [9][27]. The following sections review the state-of-the-art iterative detector-decoder as well as the GPR equalizer.

2.3.1 Iterative Detector-decoder

Figure 2.7: An iterative detection/decoding scheme described in [39].

An iterative detector-decoder system refers to a detection scheme where a detector and decoder exchange soft information, in the form of Log Likelihood Ratio (LLR), about the transmitted bits in an iterative manner. Typically, a Bahl-Cocke-Jelinek-Raviv (BCJR) [59] or Soft output Viterbi algorithm (SOVA) detector [60] is used together with an LDPC decoder, as shown in Figure 2.7. For each detected bit, the BCJR detector calculates a posteriori LLR value, L_{app_det} , based on the Maximum a Posteriori (MAP) algorithm. This soft information is passed to the LDPC decoder in the form of extrinsic LLR, L_{ext_det} , given by

$$L_{ext_det} = L_{app_det} - L_{a_det} \quad (2.1)$$

$$L_{ext}(c_k|\mathbf{y}) = L(c_k|\mathbf{y}) - L(c_k) \quad (2.2)$$

where c_k is the input bit to the channel and $\mathbf{y}$ is the received sequence. Extrinsic information from the detector then becomes a priori LLRs for the decoder, L_{a_dec} . The LDPC decoder uses a Sum-Product Algorithm (SPA) [61] to compute its own extrinsic LLR, L_{ext_dec} . Subsequently,

L_{ext_dec} becomes a priori information for the detector, L_{a_det} , and is used to update L_{app_det} of the individual bits, as shown in Figure 2.7. The feedback loop in the iterative detector-decoder improves the performance of the system greatly.

The MAP and SPA algorithm employed in the BCJR detector and LDPC decoder respectively is described in detail as follow:

MAP algorithm

The MAP algorithm operates on a trellis structure such as shown in Figure 2.8. A trellis graphically combines information represented on a state diagram with the evolution of time. The trellis nodes are sliced into vertical columns where each column represents a time index and each node represents a distinct state at a given time. A trellis branch depicts the transition from one state to another state in a given time index. For an information sequence involving binary bits, two branches emerge from a node at a given time - each depicting bit '0' and '1' respectively.

Figure 2.8 A BCJR trellis.

In the BCJR detector, the MAP algorithm aims to identify the bit that was most likely transmitted at each time index. It does so by minimizing the probability of bit error using the read-back sequence [59]. The MAP soft decision is given by

$$L_{app_det} = L(c_k|\mathbf{y}) = \ln \left(\frac{P(c_k = +1|\mathbf{y})}{P(c_k = -1|\mathbf{y})} \right) \quad (2.3)$$

Given that a trellis state of the BCJR, s_{k+1} , at time $k+1$ can be reached from state s_k at time k , L_{app_det} can be expressed in terms of the BCJR trellis as follow:

$$\begin{aligned}
L_{app_det} &= \frac{\sum_{(s',s) \in S^+} P(s_k = s', s_{k+1} = s | \mathbf{y})}{\sum_{(s',s) \in S^-} P(s_k = s', s_{k+1} = s | \mathbf{y})} \\
&= \frac{\sum_{(s',s) \in S^+} P(s_k = s', s_{k+1} = s, \mathbf{y})}{\sum_{(s',s) \in S^-} P(s_k = s', s_{k+1} = s, \mathbf{y})}
\end{aligned} \tag{2.4}$$

Using the chain rule for joint probabilities, $P(s_k = s', s_{k+1} = s, \mathbf{y})$ can be decomposed into

$$\begin{aligned}
P(s_k = s', s_{k+1} = s, \mathbf{y}) &= P(s_k = s', y_1, \dots, y_{k-1}) \cdot P(s_{k+1} = s, y_k | s_k = s') \cdot \\
&\quad P(y_{k+1}, \dots, y_N | s_{k+1} = s) \\
&= \alpha_k(s_k) \cdot \gamma_k(s_k, s_{k+1}) \cdot \beta_{k+1}(s_{k+1})
\end{aligned} \tag{2.5}$$

The terms $\alpha_k(s)$ and $\beta_k(s)$ can be recursively computed as follow:

$$\alpha_k(s) = \sum_{s' \in S} \alpha_{k-1}(s') \gamma_{k-1}(s', s) \tag{2.6}$$

$$\beta_k(s) = \sum_{s' \in S} \beta_{k+1}(s') \gamma_k(s, s') \tag{2.7}$$

The term $\gamma_k(s_k, s_{k+1})$ can be further decomposed into

$$\gamma_k(s_k, s_{k+1}) = P(s_{k+1} | s_k) \cdot P(y_k | s_k, s_{k+1}) \tag{2.8}$$

From the equations above, it follows that

$$L_{app_det} = \ln \frac{\sum_{(s',s) \in S^+} \alpha_k(s') \cdot \gamma_k(s', s) \cdot \beta_{k+1}(s)}{\sum_{(s',s) \in S^-} \alpha_k(s') \cdot \gamma_k(s', s) \cdot \beta_{k+1}(s)} \tag{2.9}$$

At this stage of the MAP algorithm, the BCJR detector can pass this soft information to the LDPC decoder in the form of L_{ext_det} . At the same time, it can also provide a hard decision of whether a transmitted bit is a 0 or 1 based on the sign of the computed L_{app_det} value. The decision rule can be written as

$$\hat{c}_k = \begin{cases} 0, & L(c_k | \mathbf{y}) < 0 \\ 1, & L(c_k | \mathbf{y}) \geq 0 \end{cases} \tag{2.10}$$

SPA algorithm

In an LDPC decoder, the SPA algorithm operates on a factor graph structure shown in Figure 2.9. The factor graph consists of n variable nodes and $(n-k)$ check nodes where n refers to the length of the code and k is the length of information bits. In Figure 2.9, check nodes are represented as triangles and variable nodes as circles. Each variable node, x , represents a code symbol and each check node, c , represents a function which is essentially a parity check equation. An edge connects a variable node to a check node if that variable is involved in that check node's parity equation.

Figure 2.9: An LDPC factor graph.

Decoding via SPA is carried out through the passing of messages along the edges of the factor graph. Messages are passed in the form of conditional probabilities that the received bit is a '1' or '0' given the read-back sequence $\mathbf{z}$. Each check node makes use of the probabilities it received from its neighbors in the previous iteration step to make a new estimate at the present iteration step for the bits involved in the parity check equation. This new estimate is then sent to the variable nodes. A message sent from a check node c_k to variable node x_i is denoted as r_{ki} and can be represented mathematically as:

$$r_{ki}(0) = \frac{1}{2} + \frac{1}{2} \prod_{i' \in N(k) \setminus i} (1 - 2q_{i'k}(1)) \quad (2.11)$$

$$r_{ki}(1) = 1 - r_{ki}(0) \quad (2.12)$$

where $N(k)$ is the set of variable nodes connected to check node c_k by an edge. Two messages are sent along each edge - $r_{ki}(0)$ which represents the conditional probability that the

received bit is a '0' and $r_{ki}(1)$ for the conditional probability that the received bit is a '1'. The message, q_{ik} , sent from variable node x_i to check node c_k can be expressed as:

$$q_{ik}(0) = m_{ik}(1 - P_i) \prod_{k' \in N(i) \setminus k} r_{k'i} \quad (2.13)$$

$$q_{ik}(1) = m_{ik}P_i \prod_{k' \in N(i) \setminus k} r_{k'i} \quad (2.14)$$

where $P_i = P[c_i = 1|z]$, $N(i)$ is the set of check nodes connected to variable node x_i and m_{ik} is a constant chosen such that

$$q_{ik}(0) + q_{ik}(1) = 1 \quad (2.15)$$

The equations above suggest that only extrinsic messages are passed between the nodes of the LDPC factor graph. The messages passed between the nodes can also be represented in terms of logarithmic likelihood ratio as follow:

$$\begin{aligned} \overline{q_{ik}} &= \ln \frac{q_{ik}(0)}{q_{ik}(1)} \\ &= \ln \left[\frac{1 - P_i}{P_i} \prod_{k' \in N(i) \setminus k} \frac{r_{j'i}(0)}{r_{j'i}(1)} \right] \\ &= \ln \frac{1 - P_i}{P_i} + \sum_{k' \in N(i) \setminus k} \overline{r_{k'i}} \end{aligned} \quad (2.16)$$

and

$$\begin{aligned} \overline{r_{ki}} &= \ln \frac{r_{ki}(0)}{r_{ki}(1)} \\ &= \ln \frac{\frac{1}{2} + \frac{1}{2} \prod_{i' \in N(k) \setminus i} (1 - 2q_{i'k}(1))}{\frac{1}{2} - \frac{1}{2} \prod_{i' \in N(k) \setminus i} (1 - 2q_{i'k}(1))} \\ &= \ln \frac{1 + \prod_{i' \in N(k) \setminus i} \tanh(\frac{\overline{q_{i'k}}}{2})}{1 - \prod_{i' \in N(k) \setminus i} \tanh(\frac{\overline{q_{i'k}}}{2})} \end{aligned} \quad (2.17)$$

The exchange of messages between variable nodes and check nodes are repeated numerous times until ultimately a final probability is computed at all variable nodes. At this point, a hard decision for each bit is then made by comparing the final probability with a threshold value.

Figure 2.10 Illustration of length (a) 4 and (b) 6 cycles in a factor graph.

In the LDPC factor graph, short cycles of length 4 and 6 commonly exist (Figure 2.10). Cycles of these lengths are known to significantly degrade the performance of the LDPC decoder. While a few measures can be implemented to reduce the presence of such cycles, a complete mitigation of these cycles (especially at long codeword length) is not possible and causes sub-optimality of the LDPC decoder. Aside from the sub-optimal LDPC decoder, the iterative detector-decoder is sub-optimal as a block. This can be justified from the fact that detection and decoding is done in an iterative manner and that the performance of the entire block improves with increasing number of iterations. An infinite number of iterations which gives rise to the best performance are in practice not possible.

2.3.2 Generalized Partial Response Equalizer

In GPR equalization, an equalizer, $\mathbf{w}$, of length, $2N_w+1$, and a target, $\mathbf{g}$, of length N_g , are jointly designed by minimizing the mean squared error (MSE) of the error signal, e_k , as illustrated in Figure 2.11. The designed target has a much shorter length than the channel, $\mathbf{h}$. This helps to reduce the complexity of the detector used after the equalization step. For a BCJR or Soft Output Viterbi Algorithm (SOVA) detector for instance, a short target results in reduced number of trellis states and thus fewer numbers of computations.

Figure 2.11: A general partial response equalizer block diagram.

The MSE of e_k can be expressed as

$$E[e_k^2] = E[d_k^2 - 2d_k y_k + y_k^2] \quad (2.18)$$

where $E[\cdot]$ denotes the expectation operator, and d_k and y_k are the output signal of $\mathbf{g}$ and $\mathbf{w}$ respectively at time instant k . d_k and y_k can be written as

$$d_k = \sum_{n=0}^{N_g-1} g_n c_{k-n} = \mathbf{g}^T \mathbf{c}_k \quad (2.19)$$

$$y_k = \sum_{n=-N_w}^{N_w} w_n r_{k-n} = \mathbf{w}^T \mathbf{r}_k \quad (2.20)$$

where $\mathbf{g} = [g_0, g_1, \dots, g_{N_g-2}, g_{N_g-1}]^T$ is the target vector, $\mathbf{c}_k = [c_k, c_{k-1}, \dots, c_{k-N_g+2}, c_{k-N_g+1}]^T$ is the coded information vector, $\mathbf{w} = [w_{-N_w}, w_{-N_w+1}, \dots, w_{N_w-1}, w_{N_w}]^T$ is the equalizer coefficients vector and $\mathbf{r}_k = [r_{k+N_w}, r_{k+N_w-1}, \dots, r_{k-N_w+1}, r_{k-N_w}]^T$ is the equalizer input vector. Equation (2.18) can then be represented as

$$E[e_k^2] = \mathbf{g}^T \mathbf{A} \mathbf{g} - 2\mathbf{g}^T \mathbf{P} \mathbf{w} + \mathbf{w}^T \mathbf{R} \mathbf{w} \quad (2.21)$$

where $\mathbf{A}$ is the $N_g \times N_g$ auto-correlation matrix of c_k , $\mathbf{R}$ is the $(2N_w+1) \times (2N_w+1)$ auto-correlation matrix of r_k and $\mathbf{P}$ is the $(N_g) \times (2N_w+1)$ cross-correlation matrix between c_k and r_k . These matrices are given by:

$$\mathbf{A} = E[\mathbf{c}_k \mathbf{c}_k^T] = E \begin{bmatrix} c_k c_k & c_k c_{k-1} & \cdots & c_k c_{k-Ng+1} \\ c_{k-1} c_k & c_{k-1} c_{k-1} & \cdots & c_{k-1} c_{k-Ng+1} \\ \vdots & \vdots & \ddots & \vdots \\ c_{k-Ng+1} c_k & c_{k-Ng+1} c_{k-1} & \cdots & c_{k-Ng+1} c_{k-Ng+1} \end{bmatrix} \quad (2.22)$$

$$\mathbf{R} = E[\mathbf{r}_k \mathbf{r}_k^T] = E \begin{bmatrix} r_{k+N_w} r_{k+N_w} & r_{k+N_w} r_{k+N_w-1} & \cdots & r_{k+N_w} r_{k-Nw} \\ r_{k+N_w-1} r_{k+N_w} & r_{k+N_w-1} r_{k+N_w-1} & \cdots & r_{k+N_w-1} r_{k-Nw} \\ \vdots & \vdots & \ddots & \vdots \\ r_{k-Nw} r_{k+N_w} & r_{k-Nw} r_{k+N_w-1} & \cdots & r_{k-Nw} r_{k-Nw} \end{bmatrix} \quad (2.23)$$

$$\mathbf{P} = E[\mathbf{c}_k \mathbf{r}_k^T] = E \begin{bmatrix} c_k r_{k+N_w} & c_k r_{k+N_w-1} & \cdots & c_k r_{k-Nw} \\ c_{k-1} r_{k+N_w} & c_{k-1} r_{k+N_w-1} & \cdots & c_{k-1} r_{k-Nw} \\ \vdots & \vdots & \ddots & \vdots \\ c_{k-Ng+1} r_{k+N_w} & c_{k-Ng+1} r_{k+N_w-1} & \cdots & c_{k-Ng+1} r_{k-Nw} \end{bmatrix} \quad (2.24)$$

The MSE cost function, J , can be expressed as

$$J = \mathbf{g}^T \mathbf{A} \mathbf{g} - 2 \mathbf{g}^T \mathbf{P} \mathbf{w} + \mathbf{w}^T \mathbf{R} \mathbf{w} \quad (2.25)$$

The cost function can be solved by equating the partial derivatives of J , $\nabla_{\mathbf{g}} J$ and $\nabla_{\mathbf{w}} J$, to 0. This will however result in a trivial solution where $\mathbf{g}$ and $\mathbf{w}$ are both 0. To avoid this, Lagrange multiplier [62], λ , can be introduced to find the MSE of e_k subject to the constraint

$$\mathbf{E}^T \mathbf{g} = \mathbf{i} \quad (2.26)$$

where $\mathbf{i}^T$ is a constraint vector of length N_i and $\mathbf{E}^T$ is an $N_i \times N_g$ matrix. To implement a monic constraint on a 5-tap target for example, $N_i = 1$, $N_g = 5$ and $\mathbf{E}^T$ (reduces to a vector) and $\mathbf{g}$ can be defined as

$$\mathbf{E}^T = [1 \quad 0 \quad 0 \quad 0 \quad 0]$$

$$\mathbf{g} = [g_0 \quad g_1 \quad g_2 \quad g_3 \quad g_4]^T$$

such that $\mathbf{i} = 1$. The cost function can then be re-written as

$$J = \mathbf{g}^T \mathbf{A} \mathbf{g} - 2 \mathbf{g}^T \mathbf{P} \mathbf{w} + \mathbf{w}^T \mathbf{R} \mathbf{w} - 2 \lambda (\mathbf{E}^T \mathbf{g} - \mathbf{i}) \quad (2.27)$$

J is thus minimized by setting its partial derivatives to 0 as follow:

$$\nabla_g J = 2\mathbf{A}\mathbf{g} - 2\mathbf{P}\mathbf{w} - 2\lambda\mathbf{E}^T = 0 \quad (2.28)$$

$$\nabla_w J = -2\mathbf{P}^T\mathbf{g} + 2\mathbf{R}\mathbf{w} = 0 \quad (2.29)$$

$$\nabla_\lambda J = \mathbf{E}^T\mathbf{g} - \mathbf{i} = 0 \quad (2.30)$$

From these 3 equations, the optimum $\mathbf{w}$ and $\mathbf{g}$ can be solved simultaneously and are given by

$$\lambda = \frac{1}{\mathbf{E}^T(\mathbf{A} - \mathbf{P}\mathbf{R}^{-1}\mathbf{P}^T)^{-1}\mathbf{E}} \quad (2.31)$$

$$\mathbf{w} = \mathbf{R}^{-1}\mathbf{P}^T \quad (2.32)$$

$$\mathbf{g} = \lambda(\mathbf{A} - \mathbf{P}\mathbf{R}^{-1}\mathbf{P}^T)^{-1}\mathbf{i} \quad (2.33)$$

A monic constraint, where $\mathbf{E} = [1 \ 0 \ \dots \ 0 \ 0]^T$, has been shown to work best [63][64], and is used in this thesis.

The GPR equalizer presented above is ideal for a one-dimensional (1D) magnetic recording channel. This equalizer system will however not perform well when applied to two-dimensional (2D) channels as its design involves the minimization of the MSE between a 1D equalizer output and a 1D target output. Novel equalizers therefore need to be implemented for 2D channels such as SMR and TDMR. This is proposed in Chapter 6 of the thesis.

Chapter 3

System Level Study of Future Recording Channels

Conventional magnetic recording technique known as Perpendicular Magnetic Recording (PMR) is approaching its areal density limit in the near future [35][36]. As such, new magnetic recording architectures such as Heat Assisted Magnetic Recording (HAMR), Shingled Magnetic Recording (SMR) and Two-Dimensional Magnetic Recording (TDMR) have been proposed in the literature in recent years [10][11][12][13]. The development and evaluation of theoretical models for these emerging recording technologies is essential in enabling an accurate prediction of its feasibility. Magnetic researchers often employ micromagnetic channel models for this purpose [15][16][17]. While this model provides an in-depth prediction of magnetic behaviors and interactions within an isolated channel, its high computational time makes it unsuitable for a system level study of a new magnetic recording scenario. System level study, known as error performance simulation, computes error rates associated with the performance of an integrated system which consists of an encoder, channel, equalizer, detector and decoder components. To this end, an approach commonly adopted in the literature involves using a generic mathematical model in place of a micromagnetic model for error performance simulations. The inaccuracy of a mathematical model could however give rise to a lack of reliability associated with predictions on the practicality of new recording systems.

In view of the impracticality in employing micromagnetic model for error performance simulations and the inadequacy of a mathematical model, this chapter of this thesis proposes a novel statistical model that models a magnetic recording channel with great accuracy and yet can be used for error performance studies with a reasonable amount of computation time. This model is known as the Grain Flipping Probability (GFP) model. With the GFP model, a novel system level study of two emerging magnetic recording channels, SMR and HAMR, is carried out with conventional signal processing techniques on the receiver end. The characterization of SMR and HAMR channels from a system's perspective is examined in this chapter.

3.1 Channel Model

A theoretical magnetic recording channel model captures the phenomenon associated with the write and read process in a Hard Disk Drive (HDD). It generates a read-back signal from the channel that can be processed with signal processing techniques on the receiver end. In micromagnetic models that are commonly used by magnetic researchers, read-back signals are computed through the use of complex differential equations. Despite its high accuracy in mimicking the magnetization behavior of grains subjected to an external magnetic field, the numerical solving of differential equations is computationally intensive and slow. This section of the thesis proposes a novel GFP channel model that generates read-back waveforms using a statistical method that is 3 to 4 orders of magnitude faster than a micromagnetic model.

The proposed GFP model consists of a look-up table (LUT) of probabilities that represent the likelihood of a magnetic grain flipping given a set of circumstances per bit interval. The LUT of probabilities are populated from a mass number of simulations using a micromagnetic channel model. As micromagnetic models can be formulated in various forms, this section first presents the specific micromagnetic model that is used to train the LUT of the proposed GFP model. The micromagnetic model was developed by magnetic researchers in Data Storage Institute (DSI). As micromagnetic constitutes a broad and independent field of study in the area of magnetism, the development of a novel micromagnetic channel model is outside the scope of this thesis.

Instead, existing micromagnetic models specific to SMR and HAMR that were developed in DSI are leveraged on in this thesis to enable the development of the GFP model that is proposed in section 3.1.2 and subsequently used in sections 3.2 and 3.3 of this chapter for a novel system level study of SMR and HAMR.

3.1.1 Micromagnetic Model

A channel model typically consists of a magnetic media and a write head that induces magnetization of grains on the media. In the micromagnetic model that is employed for developing the proposed GFP model, a magnetic disk is modeled using a Voronoi media as illustrated in Figure 3.1. As the grains on the media of a practical HDD system is of a continuous granular form, the Voronoi media constitutes a good depiction of the actual geometry of the grains. A Voronoi media is created by partitioning the plane into Voronoi cells such that each cell contains a nucleus, known as its Voronoi nucleus. The cells are partitioned in a way such that every point in the cell is closer to the nucleus of that cell than to the nucleus of any other Voronoi cell. Voronoi cell represents a magnetic grain. The diameter of the grains on the media is a random process characterized by the standard deviation, σ_{grain} . In Figure 3.1, the Voronoi media has dimensions 75 nm by 200 nm. It is made up of 306 grains, each of which has a diameter of 7 nm and $\sigma_{grain} = 20\%$. Magnetic properties such as saturation magnetization, M_s , and crystalline anisotropy, K_u , values are assigned to the grains depending on the material of the media that is of interest in any particular study. The relationship between M_s and K_u are defined as follow:

$$M_s = \frac{2K_u}{H_k} \quad (3.1)$$

where H_k refers to the coercive field.

Figure 3.1 Voronoi media.

In addition to the Voronoi media, the micromagnetic model that is used in the characterization of the proposed GFP model consists of two magnetic head profiles. One of the magnetic write head profiles is an unshielded square pole computed by Mason Williams and used during Information Storage Industry Consortium's (INSIC) initial Extreme High Density Recording (EHDR) program. The head field is modeled as a square pole of dimensions 50nm by 50nm and has a peak field of 1.4×10^6 A/m. The magnetic profile of the perpendicular component of the head field is shown in Figure 3.2(a). Another magnetic write head profile employed in this thesis is shown in Figure 3.2(b). The head profile is designed by Professor Kanai who is a collaborator at Niigata Institute of Technology (NIIT) and has generated head fields and shared them during our collaboration.

Figure 3.2: (a) Mason William's and (b) Kanai's magnetic write head profile.

The prediction of the behavior of magnetic particles on the media when subjected to an external magnetic field is formulated in the micromagnetic model using the Landau-Lifshitz Gilbert (LLG) [65] equation. The LLG equation is a differential equation describing the evolution of the magnetization of particles with time in the presence of a magnetic field. The magnetization of grains in a micromagnetic simulation is illustrated in Figure 3.3. In this figure, grains with their magnetization pointing out of the page are blue while grains with their magnetization pointing into the page are red. Figure 3.3(a) shows the magnetic grains in their initial DC-erased state. 1 bit-cell was then written causing grains under the influence of the applied magnetic field from the write head to change polarity, as illustrated by the red region in Figure 3.3(b). As the write head moves in a down-track direction, the write current is modulated, creating an alternating field of plus and minus writing the bit pattern of alternating polarity onto the media. The data is stored in the footprint left behind (Figure 3.3(c)).

Figure 3.3: Micromagnetic output showing bits of alternating polarity written on a DC background.

In this thesis, a vast number of micromagnetic simulations were carried out for a range of bit patterns to accurately populate the LUT of the GFP model under various circumstances. Figure 3.4 provides an illustration of granular magnetization that is yielded from a single micromagnetic simulation. In Figure 3.4(a), a micromagnetic output for a HAMR channel with pseudo-random data written on a single track is shown. Figure 3.4(b) shows a micromagnetic output for a SMR system where 4 tracks of data are written using a shingling technique. In the micromagnetic simulations carried out in this thesis, magnetic grains having an initial DC-erased

state (Figure 3.3) is not used. Instead, the grains on the media are in a random initial state as reflected in Figure 3.4. The statistics gathered from each micromagnetic simulation are then used to characterize the probabilities in a multi-dimensional LUT of the GFP channel model.

Figure 3.4: Micromagnetic output for a (a) HAMR and (b) SMR channel consisting of pseudo-random data written on AC background.

3.1.2 Grain Flipping Probability Model

Solving the LLG equation numerically in a micromagnetic channel model gives an accurate prediction of the magnetization phenomenon associated with the writing and reading of information bits. The reproduction of read-back signals from this process is however slow and computationally intensive. The GFP channel model aims to reproduce read-back signals via statistical means instead to ensure a significant amount of time savings and therefore practicality.

The GFP channel model consists of a multi-dimensional LUT of probabilities that represent the likelihood of a grain flipping given a set of circumstances per bit interval. The areal density of the characterized GFP model is determined by the parameters chosen in the micromagnetic simulations that are used in populating the GFP's LUT in the training phase. In the GFP model, a region of interest (ROI) is first defined for each nominal bit location. The ROI is selected such that it is wide enough to capture grains within the vicinity of the magnetic head field that have

a non-zero probability of flipping in a given bit interval. Every grain in the ROI has a probability of flipping conditioned on four parameters.

The first parameter for inclusion in the LUT is the anisotropy constant, H_k , of the magnetic grains. The second and third parameters are the position of the center of the grain, (x,y) , relative to the write head. The last parameter involves the surrounding bit patterns. The surrounding bit patterns are defined as the current bit, the previous bit written on the current track and two other bits written on the previous track. This is illustrated in Figure 3.5. The current bit's magnetization largely influences the flipping of grains in the ROI. The remaining bit patterns determine the magnetization of grains within the vicinity of the ROI and affect the likelihood of grains in the ROI flipping through exchange coupling and demagnetization field. In order for the GFP model to provide reliable statistics, these four parameters of the LUT need to be filled with sufficient data.

Figure 3.5 Inclusion of surrounding bit patterns in the GFP LUT.

The LUT of the GFP channel model is stored as two four-dimensional (4D) arrays, known as the numerator and denominator array. The denominator array counts the number of grains that have the capability to flip at every index position in the same bit interval and under the same set of circumstances. This is essentially a count of the grains within the ROI that are opposite in magnetization to the applied field. The numerator array, on the other hand, counts the number of grains that flipped under the influence of the magnetic field (as well as thermal field for a HAMR system) at every index position in a given bit interval. The grain flipping probabilities at each position on the Voronoi media can be obtained from the division of the elements of the numerator array by that of the denominator array.

As it is not possible to plot a figure of 4D arrays, the numerator and denominator arrays can be represented as a lower dimensional probability tables as a function of fewer parameters. For example, the arrays can be summed over the bit pattern and (x,y) bins, and the resulting arrays divided to arrive at a grain flipping probability as a function of H_k . A one-dimensional (1D) plot of the grain flipping probability as a function of H_k is shown in Figure 3.6. Similarly, the arrays can be summed over the dimensions of H_k and bit pattern, resulting in a grain flipping probability as a function of (x,y) . This gives a probability distribution as a function of the grain position within the ROI. Figure 3.7 shows the numerator and denominator arrays in this case and the resulting probability footprint. The probability footprint takes on the shape of the head field with grains in the middle region having a higher tendency to flip than those further out. The footprint of Figure 3.7(c) was obtained with Mason William's head field for a HAMR channel (Figure 3.2(a)). It can be noticed that there is a slight reduction in the probability of grains flipping in the center of the footprint. This phenomenon is observed because the angle of the applied field is coming down almost vertically near the center of the footprint, while it is coming down at an angle towards the edge. Grains that experience fields 180° to their easy axis, have a harder time flipping than grains that experience fields at a slight angle. This phenomenon is captured in the micromagnetic simulations and characterized in the GFP LUT.

Figure 3.6 Grain flipping probability as a function of H_k .

Figure 3.7: A (a) denominator array and a (b) numerator array that has been summed over the dimensions of H_k and bit patterns, resulting in a (c) GFP footprint as a function of (x,y) .

With a characterized LUT, a random number generator can be used to flip grains in a short interval of time. Table 3.1 compares the computation time required between a micromagnetic and GFP model, for a SMR and HAMR channel under a fixed set of circumstances. A time savings of approximately 1400 times and 2000 times is reaped for the HAMR and SMR channel respectively with the use of the GFP model. In Table 3.1, the number of bits written per frame varies from 15 to 45. In actual error performance studies, the number of bits written typically

range between 512 to 4096 bits per frame, and the number of frames simulated vary from 100 to 50 000.

The ability of the GFP model to mass produce grain magnetizations and therefore read-back signals accurately and in a reasonable amount of time makes it a good model for error performance studies of a fully integrated system. The remaining sections of this chapter present a novel system level study of SMR and HAMR. In these studies, the proposed GFP model is applied to each channel to generate reliable read-back waveforms that are processed with conventional signal processing techniques.

Table 3.1: Comparison of computation time for a micromagnetic and GFP channel model

Channel	No. of bits written	Time taken (s)	
		Micromagnetic	GFP
HAMR	1 x 15	1561	1.130
	1 x 25	2532	1.776
SMR	3 x 15	5841	3.777
	3 x 25	8601	4.238

3.2 Shingled Magnetic Recording Channel

In a SMR channel, a new track of data is written such that it overlaps with a portion of the previously written track. This is known as shingled writing, as described in Chapter 2. Such writing mechanism causes the distance between the centers of two adjacent tracks, known as the track pitch, TP , to be much smaller than that of a conventional PMR channel. As a result, severe inter-track interference (ITI) in the cross-track direction is present in an SMR channel whereas the ITI in PMR channels is negligible. In addition to the ITI, information sequence written using the SMR technique also suffers from distortions in the down-track direction due to the presence of adjacent bits. This is known as the inter-symbol interference (ISI). The ISI is

directly impacted by the distance between the centers of adjacent bits, known as the bit length, BL .

In general, an SMR channel permits the storage of more data on the media than a conventional PMR channel. SMR channel however introduces more degradation between adjacent tracks and bits. Conventional signal processing techniques can be introduced to partially cancel the effects of these distortions. To fully comprehend the relationship between the areal densities that SMR can achieve over PMR and the effectiveness of existing signal processing techniques in tackling the higher degree of distortions present in an SMR channel, this section studies the effect of varying TP at a constant BL and the effect of varying BL at a constant TP from a system's perspective. In each of these studies, the GFP model is developed specifically for an SMR channel and is characterized using a vast number of micromagnetic simulations for various values of TP and BL . The GFP model for the SMR channel is subsequently integrated with a Low-Density Parity Check (LDPC) encoder, Generalized Partial Response (GPR) equalizer and iterative detector-decoder to form a fully integrated SMR system.

3.2.1 Varying Track Pitch

In this section, the effect of varying TP (at a constant BL) on the performance of an SMR system is studied. Micromagnetic simulations are first carried out to populate GFP LUT's for various TP values and constant BL . Table 3.2 lists the set of constant parameters used. A Voronoi media with dimensions of 500 nm x 1000 nm and consisting of 7 nm grains were employed. For this study, a magnetic write head profile designed by Professor Kanai is used. Its profile is shown in Figure 3.2(b). A large number of micromagnetic simulations were run in parallel, writing three tracks of twenty pseudo-random information bits on each track with shingling technique in each simulation. The BL was kept constant at 15 nm while TP was varied between 26 to 42 nm.

The micromagnetic output for $TP = 26$ and 42 nm at a constant $BL = 15$ nm are shown in Figure 3.8 (a) and (b) respectively. The black triangles in the plots indicate the center position of a written bit and its upward and downward direction represents different bit polarity. In Figure

3.8(a), the shingled tracks appear to be more tightly squeezed in the cross-track direction as compared to that in Figure 3.8(b). For $TP = 26\text{nm}$, the tracks appear to overlap previously written track by more than half of their width. It can be inferred that the quality of the read-back signal on the shingled tracks (excluding the ‘fat track’) deteriorates greatly when TP is reduced by about half. The micromagnetic simulations outputs are used to characterize the GFP LUT’s by computing numerator and denominator arrays as described in section 3.1.2. Figure 3.9 (a) and (b) shows the corresponding GFP footprints obtained from the micromagnetic model at $BL = 15\text{nm}$ and $TP = 26$ and 42 nm respectively. The populated GFP LUT’s are subsequently used to evaluate the error performance of the SMR technology on a system level.

Table 3.2: Constant parameters used in SMR study

media	dimension	500 nm x 1000 nm	thickness	20 nm
	no. of grains	10 204	grain size	7 nm
	grain boundary	7 nm	M_s	4.8×10^5 A/m
data	tracks	3	no. of bits	20
	BL	15 nm

Figure 3.8: Micromagnetic output for $BL = 15$ nm and $TP =$ (a) 26 nm (b) 42 nm.

Figure 3.9: GFP footprints for $BL = 15$ nm and $TP =$ (a) 26 nm (b) 42 nm.

The GFP model is integrated with a GPR equalizer and a state-of-the-art iterative detector-decoder that consists of a BCJR detector concatenated with a LDPC decoder. The number of global iterations, I_{global} , between the BCJR detector and LDPC decoder of the iterative detector-decoder impacts the performance of the system. This is depicted in Figure 3.10(a) where lower Frame Error Rate (FER) values are seen to occur with larger I_{global} , especially at low SNR values. A larger I_{global} however gives rise to a higher computational complexity of the iterative system, as reflected in the plot of time taken in Figure 3.10(b). To balance this trade-off, $I_{global} = 6$ is chosen as the operating point of the iterative detector-decoder in this thesis, as indicated in Figure 3.10.

Figure 3.10: Impact of the number of iterations within the iterative detector-decoder on (a) performance of the system and (b) computation time taken.

In this section, as the TP value is varied for the SMR channel, the channel response of the system varies accordingly. The equivalent target response of the GPR equalizer system that is

designed in each case and the resulting equalizer SNR values are shown in Table 3.3. The deterioration in read-back signal at shorter TP is quantified by the SNR values shown.

Table 3.3 Target Response and SNR for various TP values

TP (nm)	Target Response	SNR (dB)
28	[0.9541,0.2991,0.0134]	8.314
30	[0.9498,0.3112,0.0324]	9.345
34	[0.9620,0.2725,0.0166]	11.093
42	[0.9542,0.2988,0.0166]	13.573

Table 3.4: Error performance simulation parameters used for SMR study

Media		Code		Detector-Decoder	
size	588 nm x 20000 nm	type	LDPC	I_{global}	6
no. of grains	2 400 000	length	4096	I_{dec}	150
grain size	7 nm	code rate	0.88-0.985		

With the GFP LUTs that are populated for various TP values and GPR equalizer designed for each case, error performance simulations were then carried out using random LDPC code of codeword length, $N = 4096$ and of varying code rate, R (a measure of the amount of overhead introduced by the code). Conventional read-back technique where a single reader head gathers a 1D read-back signal from the center data track is used. The signal is then subsequently passed to an iterative detector-decoder. Table 3.4 lists the parameters used in the error performance simulation part of the study.

The LDPC coded error performance simulation results are shown in Figure 3.11. For every curve, as the amount of overhead (R) increases, it is seen that the Frame Error Rate (FER) suddenly decreases rapidly after a certain point. This point, typically referred to as the waterfall region, denotes the R at which the system transitions from a failed system, with $FER = 1$, to a working system with low FER. In Figure 3.11, the performance of the SMR channel at $BL = 15$ nm is seen

to deteriorate with decreasing TP (increasing media density). This is attributed to the fact that as the tracks are squeezed more tightly together in the cross-track direction, the distortions caused by ITI become more severe. The major source of the deteriorating performance, however, lies with the fact that at smaller TP , the grains per bit are lower and therefore SNR is poorer (section 1.1). A lower R and therefore more overhead is needed to reduce the error-rates. For example, to achieve an FER of 0.01, R approximately equal to 0.915 and 0.985 are needed for $TP = 28$ nm and 42 nm respectively. While $TP = 28$ nm permits the storage of more data than for $TP = 42$ nm, it requires a much greater overhead to achieve the same performance. To determine the best operating point, the storage density is computed for each TP value as follow:

$$\begin{aligned} \text{Storage density} &= \text{Media density} * R_{op} \\ &= \text{tracks per inch} * \text{bits per inch} * R_{op} \end{aligned} \quad (3.2)$$

where R_{op} is defined as the operating code rate. In this section of the thesis, the operating code rate is defined as the R value at which the system has FER of at most $1e^{-4}$. It is obtained by interpolating each curve in Figure 3.11 (ignoring the edge of each curve which falls vertically down to an FER = 0). The storage density for each TP tested at $BL = 15$ nm is computed and shown in Table 3.5. From the table, it can be concluded that although a higher R_{op} is required for small TP ($TP = 28$ nm and 30 nm), the storage density is approximately larger than 0.4 Tb/in² as compared to that for $TP = 42$ nm. As such, $TP = 28$ nm and 30 nm is preferred at $BL = 15$ nm.

Table 3.5 Channel and storage density for varying TP

TP (nm)	Media density (Tb/in²)	R_{op}	Storage density (Tb/in²)
28	1.5361	0.9050	1.3902
30	1.4337	0.9375	1.3441
34	1.2650	0.9500	1.2018
42	1.0241	0.9650	0.9882

Figure 3.11: Error performance simulation results for varying TP and $BL = 15$ nm.

3.2.2 Varying Bit Length

In the next set of study, BL is varied between 10 to 17 nm while TP is kept unchanged at 34 nm. The media and magnetic write head set-up used in the section 3.2.1 (Table 3.2) are replicated in this study.

The micromagnetic and GFP output obtained for $BL = 13$ and 17 nm are shown in Figure 3.12 and 3.13 respectively. Based on the micromagnetic plots, the written tracks for the case of $BL = 13$ nm are seen to be squeezed more tightly together in the down-track direction as compared to that for $BL = 17$ nm. The GFP footprint for $BL = 17$ nm is slightly longer than for $BL = 13$ nm as the writer has been “painting” the media for 17 nm instead of 13 nm. Analyzing the plots visually, it can be inferred that for a fixed TP value, a read-back signal of better quality can be obtained with a larger BL but at the expense of media density. This can be quantified by computing the corresponding target response and thus SNR values for each case, as shown in Table 3.6. For example, in this table, an SNR drop of around 3 dB is shown when BL decreases from 17nm to 12nm at a fixed TP value.

Figure 3.12: Micromagnetic output for $TP = 34$ nm and $BL =$ (a) 13 nm (b) 17 nm.

Figure 3.13: GFP footprints for $TP = 34$ nm and $BL =$ (a) 13 nm (b) 17 nm.

Table 3.6 Target Response and SNR for various BL values

BL (nm)	Target Response	SNR (dB)
12	[0.9363,0.3495,0.0349]	9.269
13	[0.9367,0.3484,0.0339]	9.730
15	[0.9620,0.2725,0.0166]	11.093
17	[0.9673,0.2520,0.0292]	12.369

For the error performance simulations part of this study, the parameters in Table 3.4 are used. The FER plot obtained for various BL values is shown in Figure 3.14. From the plot, it can be noted that the performance of the SMR channel worsens with decreasing BL as ISI distortions

become more severe when bits are written closer together in the down-track direction. The main attributing factor however lies with the fact that for a fixed TP , decreasing BL implies decreasing number of grains per bit and hence poorer SNR. Lower R values are thus needed to get a better performance, as depicted by the curves for $BL = 13$ and 17 nm (where a difference of approximately 0.03 in R is seen to achieve an FER of 0.001). To determine the best operating point, the storage density for each of the BL value in Figure 3.14 is computed and presented in Table 3.7. It can be inferred from the table that although more overhead is required at short BL ($BL = 12$ nm), the storage density attained is greater than that of larger BL ($BL = 17$ nm) by around 0.4 Tb/in^2 .

Figure 3.14: Error performance simulation results for varying BL and $TP = 34\text{nm}$.

The media density and storage density were also computed for all values of TP and BL used in the study. Figure 3.15 presents an individual value plot that relates the media density and storage density to the range of values of TP and BL tested. While the plot of media density (Figure 3.15(a)) depicts the analytical relationship between media density and TP/BL , the plot of storage density (Figure 3.15(b)) provides an indication on a good operating region for the SMR channel. From the plot, an operating region where $TP = 28 - 30$ nm and $BL = 12 - 13$ nm is proposed for the SMR channel under the simulated conditions listed in Tables 3.3 and 3.4.

The use of various GFP LUTs for a range of TP and BL values for an SMR channel in this study enable a system level analysis of the compromise required between media density and coding

overhead. The study gives an insight on a good operating region that can be put into practice for an SMR system in future.

Table 3.7 Channel and storage density for varying TP

BL (nm)	Media density (Tb/in ²)	R_{opt}	Storage density (Tb/in ²)
12	1.5813	0.915	1.4469
13	1.4596	0.935	1.3648
15	1.2650	0.950	1.2018
17	1.1162	0.960	1.0716

(a)

(b)

Figure 3.15: Comparison of (a) media density and (b) storage density for various TP and BL values of an SMR channel.

3.3 Heat Assisted Magnetic Recording Channel

Unlike a traditional PMR channel that makes use of magnetic field to store information on a media, a HAMR channel relies on the presence of both a magnetic field and a thermal field to assist in the flipping of grains on a high coercivity media. In other words, HAMR technology makes use of a magnetic writer and a thermal hot spot to store information at a higher density. To advance HAMR as a viable upcoming magnetic recording technology, the impact of the thermal field on the performance of a HAMR system must first be properly understood. This is examined in this section.

In order to accurately represent a HAMR channel, a thermal hot spot and a magnetic head is incorporated into the micromagnetic HAMR model. In this thesis, Mason William's head profile (Figure 3.2(a)) is used for studies involving the HAMR architecture. The thermal spot can be modelled in various forms. In this thesis, the thermal spot in the micromagnetic model is designed as a symmetrical two-dimensional (2D) Gaussian profile with a peak temperature, T_{peak} , and equal width, T_{sig} , in the down-track and cross-track direction as shown in Figure 3.16. Mathematically, this can be expressed as (3.3) and (3.4). The thermal spot is placed such that it is at a down-track offset position of dis_x relative to the magnetic head.

$$T(x) = 300 + (T_{peak} - 300)e^{-\frac{x^2}{2\sigma_T^2}} \quad (3.3)$$

$$T_{sig} = \sqrt{2}\sigma_T \quad (3.4)$$

The presence of thermal field from the hot spot alters the magnetic properties of magnetic grains on a media. In the micromagnetic model used in this section, an Iron (III) Platinum (FePt) media is employed. The variation of its M_s and K_u with temperature is shown in Figure 3.17 [66]. Mathematically, the variation of these properties with respect to temperature was formulated from [66] to depict a practical scenario and is as follow:

$$K_u(T) = \frac{1}{2} K_o \left[1 - \tanh \frac{(T - T_c)^2}{\Delta_T^2} \right] \quad (3.5)$$

$$M_s(T) = M_{s_o} \left[1 - \frac{T}{T_c} \right]^{\beta_m} \quad (3.6)$$

where K_o refers to the crystalline anisotropy of the FePt media at $T = 0$ K and has a constant value of 1.5×10^6 J/m³. Δ_T denotes temperature change in K. M_{s_o} represents the saturation magnetization of the media at $T = 0$ K and has a constant value of 0.6283 T. T_c refers to the Curie temperature (temperature at which a magnetic material loses its ability to magnetize) of the FePt media and is equal to 660 K. β_m is a constant with value 0.3.

In the micromagnetic and thus GFP model for HAMR, T_{peak} and T_{sig} of the hot spot define the strength of the thermal field and the diameter of the thermal profile respectively. In other

words, these parameters determine the thermal gradient used in the writing process and thus the ease of writing and the error rates associated with it. In addition to that, the alignment of thermal and magnetic fields (dis_x) in the HAMR model is also critical in influencing the flipping of grains. This section studies the impact of each of the hot spot properties (T_{peak} , T_{sig} and dis_x) on the overall performance of the system using the GFP model for HAMR.

Figure 3.16: Thermal profile of hot spot in a HAMR system.

Figure 3.17: Variation of (a) M_s and (b) K_u of magnetic grains with temperature [66].

3.3.1 Varying Peak Temperature of Thermal Hot Spot

In this sub-section, the effect of varying T_{peak} of the hot spot is investigated while holding T_{sig} and dis_x of the hot spot, magnetic field and media constant. The constant parameters used in this study are displayed in Table 3.8. These parameters were obtained from [66]. During the micromagnetic simulations for a HAMR channel, the K_u and M_s of the grains are adjusted with T according to the profiles shown in Figure 3.17. A Voronoi media consisting of 4.5 nm grains with σ_{sigma} of 19.6 % and grain boundary of 0.5 were created for this study. A damping constant of 0.15 and M_s of 500×10^3 A/m were used. Mason William's head field described in section 3.1.1 was used to create the magnetic field. The hot spot's T_{sig} value was kept fixed at 14.8 nm in the down-track and cross-track direction. Pseudo-random bits were written with BL of 9 nm, at a nominal TP of 32 nm (corresponding to a channel areal density of 2 Tbps). T_{peak} of the hot-spot was varied from 600 - 699 K.

A large number of micromagnetic simulations were run in parallel, writing single track of 20 random bits each. Figure 3.18 (a) and (b) show the micromagnetic plots obtained at $T_{peak} = 600$ K and 699 K respectively. The white triangles in the plots indicate the center position of a written bit and its upward and downward direction represents different bit polarity. Comparing the number of magnetized grains per bit in the micromagnetic plots, it is apparent that the

Table 3.8: Constant parameters used in HAMR study

media	dimension	75 nm x 400 nm	thickness	15 nm
	no. of grains	1481	grain size	4 nm
	grain boundary	0.5 nm	K_u	8.18×10^5 J/K
	M_s	500×10^3 A/m		
write head	Peak H_k	1.4×10^6 A/m	velocity	9 m/s
hot spot	T_{sig}	14.8 nm	dis_x	0 nm
data	tracks	1	bits	20

Table 3.9 Target Response and SNR for various T_{peak} values

T_{peak} (K)	Target Response	SNR (dB)
699	[0.9842,0.1751,0.0264]	10.580
675	[0.9777,0.2097,0.0105]	10.486
650	[0.9743,0.2234,0.0279]	9.690
600	[0.9775,0.2077,0.0367]	9.590

region occupied by writing a bit pattern is larger when $T_{peak} = 699$ K than when $T_{peak} = 600$ K. This implies that writing at a higher temperature results in more grains flipping for a written bit.

The micromagnetic simulations outputs are used to characterize the GFP LUT's for various T_{peak} values. The GFP footprints showing the probability of a grain flipping at a given (x,y) position are

shown in Figure 3.19. At $T_{peak} = 699$ K, the area occupied by the colored region is larger than that for $T_{peak} = 600$ K. This also shows that more grains have flipped at $T_{peak} = 699$ K.

Figure 3.18: Micromagnetic output for (a) $T_{peak} = 600$ K (b) $T_{peak} = 699$ K.

Figure 3.19: GFP footprints for (a) $T_{peak} = 600$ K (b) $T_{peak} = 699$ K.

To further evaluate the T_{peak} property of HAMR at a system level, error performance simulations employing the GFP model needs to be carried out. To do so, a GPR equalizer and target are first jointly designed for the combined GFP and RHS channel model that corresponds to various T_{peak} values. Table 3.8 shows the equivalent GPR target response and SNR computed for each T_{peak} . The data in the table showed a trend of increasing SNR with T_{peak} . In the error performance simulations, a larger Voronoi media with the same grain dimensions was used to enable more bits to be written onto it. Random LDPC codes with $N = 4096$ are used with varying R . The target responses in Table 3.9 and a state-of-the-art iterative detector-decoder were used on the read-back end. Table 3.10 summarizes the parameters used.

Table 3.10: Error performance simulation parameters used for HAMR study

Media		Code		Detector-Decoder	
size	450 nm x 70000 nm	type	LDPC	I_{global}	6
no. of grains	1 555 050	length	4096	I_{dec}	150
grain size	4 nm	code rate	0.88-0.985

Figure 3.20: Channel performance for various T_{peak} values.

Figure 3.20 shows the result obtained in the error performance study carried out for varying T_{peak} . The impact on the channel performance shows that in general, the HAMR channel performs better with higher T_{peak} . For example, at a fixed R of 0.97, the FER obtained when $T_{peak} = 675$ K is approximately 0.004 whereas that obtained for $T_{peak} = 650$ K and 600 K are around 0.8 and 1 respectively. The result observed is due to the fact that larger T_{peak} gives rise to a larger temperature gradient and therefore sharper bit transitions on the media. Sharper bit transitions results in lesser media noise [53] and therefore higher SNR and lower FER. Aside from a sharper thermal gradient, a larger T_{peak} also causes the magnetization of more grains on the media. This contributes to a higher media SNR and better FER. Grains which are far from the Curie temperature, T_c , will not be as easy to flip as grains near or at the T_c . In fact, the M_s and H_k of the magnetic grains decrease sharply when T_c is approached [67]. Ideally, a T_{peak} value that is as close as possible to T_c should be chosen for a HAMR channel. For this set of study, this corresponds to $T_{peak} = 675$ K or 699 K.

3.3.2 Varying Diameter of Thermal Hot Spot

In this set of study, the effect of varying T_{sig} is investigated while holding the remaining simulation conditions constant. The same media and magnetic write head used in section 3.3.1 are employed here (Table 3.7). T_{peak} and dis_x of the hot spot is kept fixed at 600 K and 0 nm respectively. T_{sig} of the Gaussian hot spot is varied between 15 nm to 35 nm.

Figure 3.21: Micromagnetic output for (a) $T_{sig} = 17$ nm (b) $T_{sig} = 29$ nm.

Figure 3.22: GFP footprints for (a) $T_{sig} = 17$ nm (b) $T_{sig} = 29$ nm.

The micromagnetic output for $T_{sig} = 17$ nm and 29 nm, are shown in Figure 3.21, while the probability footprints for the same cases from the GFP LUT's are shown in Figure 3.22. From these plots, it is clear that changing T_{sig} effectively changes the width of the written tracks (or the track density, as well as the cross-track and down-track gradients). Hence the track density and Bit Aspect Ratio (BAR) are changing as well. The computed target response and SNR corresponding to the GFP and RHS for the case of varying T_{sig} is shown in Table 3.11. A trend of

increasing SNR with T_{sig} is seen from the table. To further validate this trend, error performance simulations were carried out using the GFP LUT and the designed target response for each case.

Table 3.11 Target Response and SNR for various T_{sig} values

T_{sig} (nm)	Target Response	SNR (dB)
17	[0.9765,0.2145,0.0237]	10.388
29	[0.9655,0.2590,-0.0276]	11.153
33	[0.9757,0.2185,0.0198]	11.189
35	[0.9904,0.1312,0.0433]	12.567

Figure 3.23: Channel performance for various T_{sig} values.

In the error performance evaluation, the same set of parameters used in section 3.3.1 (Table 3.9) was replicated here. Figure 3.23 shows the FER plot for various T_{sig} . The overall interaction between the various effects of T_{sig} , track density and BAR is fairly complex, but it can be seen from Figure 3.23 that generally, reducing T_{sig} has the effect of shifting the FER waterfalls to the left, i.e. poorer performance. This can be explained by observing that as T_{sig} reduces, the track density increases. The SNR drops as a result and the FER degrades.

3.3.3 Varying Alignment Between Thermal Hot Spot and Magnetic Write Head

In sections 3.3.1 and 3.3.2, the hot spot and magnetic head profiles were aligned at their centers ($dis_x = 0$ nm). In this sub-section, the effect of changing this alignment ($dis_x \neq 0$ nm) is studied while holding the remaining simulation conditions unchanged (Table 3.7). T_{peak} and T_{sig} were kept fixed at 699 K and 21 nm respectively. Values of dis_x between 0 to 10 nm were tested in the micromagnetic simulations and GFP LUT's were characterized for each case. Conditions for the error performance simulations remained unchanged as in sections 3.3.1 and 3.3.2 (Table 3.9).

Figure 3.24: Micromagnetic output for (a) $dis_x = 0$ nm (b) $dis_x = 9$ nm.

Figure 3.25: GFP footprints for (a) $dis_x = 0$ nm (b) $dis_x = 9$ nm.

Figure 3.24 and Figure 3.25 show the respective micromagnetic and GFP output for $dis_x = 0$ nm and 9 nm. The computed target responses and equalizer SNR values for each dis_x value is shown in Table 3.12. The SNR values in the table indicate a trend of initial deterioration in signal

quality as dis_x increases and a subsequent improvement as dis_x increases further. The FER plots obtained from error performance simulations are shown in Figure 3.26. From this plot, it can be noted that the performance of the HAMR channel first deteriorates when $dis_x = 1$ nm and $dis_x = 3$ nm are introduced between the thermal and magnetic fields. Increasing the offset distance further showed an improved performance with the best curve obtained at $dis_x = 9$ nm.

Table 3.12 Target Response and SNR for various dis_x values

dis_x (nm)	Target Response	SNR (dB)
0	[0.9839,0.1778,0.0197]	10.949
1	[0.9776,0.2106,0.0059]	10.836
3	[0.9719,0.2350,0.0161]	9.972
5	[0.9792,0.2019,0.0221]	10.711
9	[0.9807,0.1954,0.0002]	11.063

The results obtained can be attributed to the alignment of the down-track thermal and magnetic gradients. When the two gradients optimally align, the combined gradient results in the sharpest written transition and therefore giving rise to the best SNR and best performance. Mathematically, this phenomenon can be explained as follow: The effective writing field gradient, $\frac{dH_{write}}{dx}$, is given by the sum of effective thermal gradient, $\frac{dH_T}{dx}$, and the head field gradient, $\frac{dH_w}{dx}$ as stated in (3.7).

$$\frac{dH_{write}}{dx} = \frac{dH_T}{dx} + \frac{dH_w}{dx} \quad (3.7)$$

The effective thermal gradient is defined as follow:

$$\frac{dH_T}{dx} = \frac{dH_k}{dT} \times \frac{dT}{dx} \quad (3.8)$$

where $\frac{dH_k}{dT}$ refers to the gradient of the FePt media's $H_k(T)$ and $\frac{dT}{dx}$ is the temperature gradient of the media at the write temperature [68].

Figure 3.26: Channel performance for various dis_x values.

3.4 Summary

Developing an accurate and robust theoretical channel model is essential in assessing the feasibility of proposed magnetic recording technologies as well as signal processing techniques. The micromagnetic model is widely used by magnetic researchers to study the magnetization of particles for various recording architectures. It involves modeling of the physical components of the channel such as the media and the writer as well as the modeling of the magnetization of the grains. Although this modeling tool is highly accurate, it is very time consuming. This renders it unsuitable for evaluating error performance of a fully integrated system that consist of the channel and signal processing components. The micromagnetic model is therefore only useful in analyzing an isolated channel from a physics aspect. This alone would not suffice in analyzing and drawing valid conclusions on a recording system which comprises of not only the channel but a set of encoder and decoder, equalizer and detector. Computation of SNR values and evaluation of FER curves from error performance study would constitute a more robust and in-depth analysis of a channel at a system level. As such, a mathematical model that generically represents a channel response is commonly used in the literature. While this approach significantly reduces the time taken to generate a read-back signal and thereby making it

suitable for use in error performance simulations, a generic mathematical model is often largely inaccurate and fails to capture crucial phenomena such as the bit transition effects.

This chapter presents a novel GFP channel model as an alternative to the micromagnetic model. The GFP model consists of a multi-dimensional LUT of grain flipping probabilities that is populated using a mass number of micromagnetic simulations in a training phase. The characterized model was shown to reproduce magnetization of grains approximately 1400 times and 2000 times faster than a micromagnetic model for a HAMR and SMR channel respectively. Given its speed and accuracy, the GFP model is a key to performing error performance simulations.

In this chapter, the GFP channel model is employed in two novel studies involving the SMR and HAMR at system level. The impact of varying TP and BL of an SMR channel and the effect of T_{peak} , T_{sig} and dis_x of a thermal hot spot in a HAMR channel were studied using a GFP model developed specifically for each channel. The GFP model in each case is subsequently integrated with an LDPC encoder, GPR equalizer and a state-of-the-art iterative detector-decoder to develop a comprehensive system model. The model built permits the simulation of error performance with great accuracy.

The study conducted for an SMR channel quantifies the trade-off between media density (TP or BL) and coding overhead. An ideal operating region was identified for the set of parameters used that balances this trade-off. For example, $TP = 28\text{nm}$ and $BL = 15\text{nm}$ corresponding to media storage density and user storage density of 1.54 Tb/in^2 and 1.39 Tb/in^2 respectively was found to be the best. For the HAMR channel, a trend of decreasing FER was seen with increasing T_{peak} and T_{sig} of the thermal hot spot. For example, for an FePt media, T_{peak} that is close to T_c (700 K) showed an SNR gap of 1 dB and difference of 0.07 in R when compared to $T_{peak} = 600\text{ K}$. In addition to that, dis_x of the hot spot displayed an optimum trend with FER. The optimum $dis_{x_{was}}$ found to be around 9nm for the simulated conditions.

Subsequent chapters of this thesis focus on the development of novel signal processing techniques for magnetic recording systems. The GFP channel model presented in this chapter is

extensively used in the remaining part of the thesis for integration with proposed joint detector decoder (Chapter 4 and 5) and equalizer (Chapter 6) to enable a comprehensive error performance evaluation at a system level.

Chapter 4

Joint Viterbi Detector Decoder

Developments of new magnetic recording channels can be synergized with advancements in signal processing techniques to bring about an impactful gain in storage capacity in the near future. In today's Perpendicular Magnetic Recording (PMR) systems, the iterative detector-decoder and the Generalized Partial Response (GPR) equalizer constitutes the state-of-the-art signal processing technology on the receiver end [7][8][9]. The sub-optimality of the iterative detector-decoder [58][69] and the inadequacy of the GPR equalizer for emerging two-dimensional (2D) recording channels signify a need for better signal processing algorithms. As such, novel signal processing techniques are proposed in this PhD work and are presented in the remaining part of this thesis. Chapter 3 and 4 each presents an alternative method of performing detection and decoding. Chapter 5 proposes various equalization schemes that are applicable to emerging 2D recording channels.

The state-of-the-art iterative detector-decoder consists of two separate detector and decoder blocks that work together in an iterative manner. The algorithm was formulated such that local optimality is attained in every iteration step. While this gives rise to a high computational efficiency, it however does not necessarily ensure a globally optimum solution. In fact, the iterative detection and decoding scheme is sub-optimal for cyclic graphs. Cycles in the decoder graph, especially short ones, often have a damaging effect on its performance. As such, advancements in detection and decoding have been inclined towards joint detection and decoding schemes [23][24][25] where a detector and decoder are represented as a single entity

and the process of detection and decoding are performed simultaneously instead of sequentially and repetitively. In this chapter, a novel joint detector and decoder system coined the Joint Viterbi Detector Decoder (JVDD) is presented.

4.1 JVDD

The state-of-the-art iterative detector-decoder system typically consists of a Bahl-Cocke-Jelinek-Raviv (BCJR) detector [59] and Low-Density Parity Check (LDPC) decoder [54] as shown in Figure 4.1. The BCJR detector has information of the channel and attempts to cancel the effect of Inter-Symbol Interference (ISI) by executing the maximum a posteriori algorithm (MAP) on a channel trellis. The LDPC decoder, on the other hand, has information of the code and recovers encoded bits from the detected stream using the Sum Product Algorithm (SPA) over a decoder factor graph as described in Chapter 2. As the detector and decoder each has different knowledge and function, they work by exchanging extrinsic information between them repetitively before a conclusion is made on the most likely stored information on the media. Some deterioration in performance is generally induced through the separation of the intrinsically dependent detection and decoding tasks [58][69]. If the detector and decoder can both be made to operate on the same structure, these two blocks can be combined into a single entity. When this occurs, detection and decoding can be performed jointly.

Figure 4.1: Block diagram depicting difference between iterative detector-decoder and JVDD.

JVDD is a novel joint detection and decoding scheme proposed in this chapter as an alternative to the state-of-the-art iterative detector-decoder. In this scheme, the task of detection and decoding are combined and carried out jointly over a channel trellis, as illustrated in the block diagram of Figure 4.1. While the proposed scheme gives rise to a good performance, the computational complexity of the algorithm surges rapidly as the trellis length grows. Various effective measures can however be introduced to lower the number of computations and reduce the computation time required by the algorithm. The JVDD can be applied to a magnetic recording channel or any coded communication channels that suffer from ISI in general. In this thesis, JVDD is first applied to a one-dimensional (1D) ISI channel, followed by a PMR channel using the Grain Flipping Probability (GFP) channel model presented in Chapter 3. As the JVDD algorithm is only applicable to 1D channels, a non-binary adaptation of the scheme, coined as the 2D-JVDD, was proposed for 2D channels and tested over a 2D ISI channel.

4.1.1 Algorithm

The JVDD operates on a channel trellis and performs detection and decoding simultaneously by executing two key tasks in its algorithm – ‘metric thresholding’ and ‘parity checking’. The ‘metric thresholding’ part of the algorithm aims to find a path through the trellis that has the minimum metric (MM) in the Viterbi [70] sense whereas the ‘parity checking’ segment of the JVDD algorithm performs the role of identifying sequences that constitute legal codewords through error correction coding (ECC) technique [71][72]. With these two key parts of the algorithm, a path through the trellis that has a minimum metric in the Euclidean distance sense and yet belongs to a legal codeword will satisfy both the detection and decoding criteria jointly. This path has been coined as the minimum metric legal codeword (MMLC) and is the desired output of the JVDD. A detailed explanation of the two key steps in the JVDD algorithm is presented as follow:

Metric thresholding

The ‘metric thresholding’ step in the JVDD algorithm aims to search for a path through the channel trellis that has the smallest metric in the Euclidean distance sense with regards to a read-back sequence obtained from the channel. In search of the MM path, the Viterbi algorithm that is widely known in the literature [70] is modified and adopted in the ‘metric thresholding’ task of the JVDD. The Viterbi algorithm is a MAP estimator (section 2.2.1) in its most general form. Given an encoded signal $\mathbf{c}$ transmitted over an ISI channel, the received signal $\mathbf{y}$ at time instant k can be expressed as

$$y_k = \sum_{l=0}^{N_h-1} h_l c_{k-l} + n_k \quad (4.1)$$

where $\mathbf{n}$ refers to the Additive White Gaussian Noise (AWGN) noise with Power Spectral Density (PSD) $= \frac{N_0}{2}$, $\mathbf{h}$ is the response of the channel of length N_h . The algorithm makes an estimate of the transmitted sequence, $\hat{\mathbf{c}}$, by maximizing the probability $P(\mathbf{c}|\mathbf{y})$. Mathematically, this can be expressed as

$$\hat{\mathbf{c}} = \underset{\mathbf{c}}{\operatorname{argmax}} P(\mathbf{c}|\mathbf{y}) \quad (4.2)$$

By applying Bayes’ theorem, equation (4.2) can be written as

$$\hat{\mathbf{c}} = \underset{\mathbf{c}}{\operatorname{argmax}} \frac{P(\mathbf{y}|\mathbf{c})P(\mathbf{c})}{P(\mathbf{y})} \quad (4.3)$$

Equation (4.3) constitutes the MAP detector presented in section 2.2.1. In this equation, if $P(\mathbf{c})$ and $P(\mathbf{y})$ are both constant (ie. if the vectors $\mathbf{c}$ are equiprobable), then the MAP detector reduces to a Maximum Likelihood (ML) detector [73][74] where

$$\begin{aligned} \hat{\mathbf{c}} &\propto \underset{\mathbf{c}}{\operatorname{argmax}} P(\mathbf{y}|\mathbf{c}) \\ &\propto \prod_k p(y_k | c_k, \dots, c_{k-N_h+1}) \end{aligned} \quad (4.4)$$

For an AWGN channel, $P(y_k | c_k, \dots, c_{k-N_h+1})$ is given by [75]:

$$P(y_k | c_k, \dots, c_{k-N_h+1}) = \frac{1}{\sqrt{\pi N_0}} e^{-\frac{|y_k - \sum_{l=0}^{N_h-1} h_l c_{k-l}|^2}{N_0}}$$

$$\ln P(y_k | c_k, \dots, c_{k-N_h+1}) = -\ln(\sqrt{\pi N_0}) - \frac{|y_k - \sum_{l=0}^{N_h-1} h_l c_{k-l}|^2}{N_0} \quad (4.5)$$

In equation (4.5), $\ln P(y_k | c_k, \dots, c_{k-N_h+1})$ can be maximized by minimizing the Euclidean distance between the received sequence and the noiseless sequence as follow:

$$\hat{c} \propto \prod_k \left| y_k - \sum_{l=0}^{N_h-1} h_l c_{k-l} \right|^2 \quad (4.6)$$

The Viterbi algorithm therefore searches for the most likely transmitted sequence. This is in contrast to the BCJR detector used in the state-of-the-art iterative detector-decoder which searches for the most likely bit transmitted at every time interval. The Viterbi algorithm executes on a trellis structure that is defined in the same manner as that used by the BCJR detector explained in section 2.2.1. The Viterbi algorithm however computes a metric for every branch that emerges from a trellis node by finding the Euclidean distance between the noisy channel output and noiseless channel output. This is known as the branch metric and is illustrated in Figure 4.2. In the Viterbi trellis, a path is formed from the concatenation of consecutive branches. Consequently, a path metric is assigned to each path by summing the branch metrics for all the branches that make up a given path, as depicted in Figure 4.2.

At any time instant, whenever two paths merge into a node on the trellis, a process of splitting and elimination takes place. In other words, the Viterbi algorithm compares the path metric of each pair of paths entering a node and the larger of the two is discarded. The process of splitting and eliminating larger metric paths is repeated until the end of the trellis is reached. If the trellis is properly terminated with a few known bits, the survivor paths in the trellis merge to a single survivor. This is essentially the MM survivor for the given waveform.

Figure 4.2. Computation of branch and path metric in the Viterbi algorithm.

This Viterbi concept is employed in the JVDD algorithm. In the conventional Viterbi algorithm, an output MM sequence may not necessarily satisfy the MMLC criteria for the JVDD. In fact, the branch elimination technique that is used in the Viterbi algorithm will result in a high likelihood of discarding the correct MMLC sequence at some point in the trellis because its path metric happens to be the larger one out of two paths compared. When this happens, the JVDD will not be able to return the MMLC. To prevent this, ideally, all paths that emerge from the splitting step should be retained to eliminate the possibility of discarding the MMLC. This is however, akin to the brute force search and is impractical. As such, modification to the Viterbi algorithm has to be made in order to adopt it as a detection tool in the JVDD algorithm. The proposed modification involves the introduction of a threshold parameter, $\mathcal{T}$. During the Viterbi splitting and elimination process, instead of comparing the relative metrics of each pair of paths entering a node and discarding the larger metric path, the conventional Viterbi algorithm can be modified such that path metrics are compared and discarded based on a maximum allowable metric. The maximum allowable metric, m_{max} , is given by

$$m_{max} = mm + \mathcal{T} \quad (4.7)$$

where mm refers to the minimum metric of all paths evaluated at each node. A path in the trellis is discarded if its path metric is larger than m_{max} . This splitting and elimination technique in the modified Viterbi algorithm can be coined as the ‘metric thresholding’ step and is employed in the first phase of the JVDD algorithm. With the proposed modification to the

conventional Viterbi algorithm, the number of paths that are discarded at each node can now range between 0 to 2 (for binary input sequence). Figure 4.3 summarizes the ‘metric thresholding’ step in the JVDD algorithm.

Figure 4.3: Metric thresholding in the JVDD algorithm.

Figure 4.4: Effect of τ on: (a) performance and complexity measured of the JVDD measured in terms of (b) average number of survivors and (c) total number of survivors.

The choice of value of τ is critical in determining the performance and computational complexity of the JVDD algorithm. A large value of τ lowers the probability of incorrectly eliminating the MMLC. This claim can be supported with a performance plot shown in Figure 4.4(a). The plot is obtained with random LDPC code of length $N = 512$ and code rate, $R, = 0.90$ over an ISI channel. The curves indicate a decreasing trend of Frame Error Rate (FER) with increasing τ . The improved performance seen with large τ values however come at the expense of computational complexity. Figure 4.4(b) shows an increase in the average number of survivor paths in the JVDD trellis with increasing τ . To better understand the computational complexity of JVDD, Figure 4.4(c) illustrates the growth of survivor paths over the trellis for the case of Signal-to-Noise ratio, SNR = 8 dB. A greater overall accumulation of survivors is seen to occur

with larger τ . On the contrary, when τ is small, there is a high likelihood that the MMLC will be discarded in the ‘metric thresholding’ phase. In other words, there will be no legitimate codes amongst the survivors left in the trellis and the JVDD returns a null output. A trade-off between performance and complexity is therefore necessary in the ‘metric thresholding’ step of the JVDD algorithm.

Parity checking

In addition to the ‘metric thresholding’ step, the JVDD algorithm performs ‘parity checking’ in order to ensure that paths that are retained in the channel trellis of the JVDD constitute that of a legal codeword. The joint execution of these two key tasks in the JVDD algorithm ensures that a MMLC output is produced. Unlike the ‘metric thresholding’ task that is carried out at every node in the trellis from time index 0 to $N-1$ (N refers to the codeword length), the ‘parity checking’ step is only executed at trellis nodes that are designated as parity check nodes. This is illustrated in the JVDD trellis diagram of Figure 4.5

Figure 4.5: The execution of the JVDD algorithm on a trellis consists of two parts – ‘metric thresholding’ at all nodes and ‘parity checking’ at designated parity check nodes.

The designated parity check nodes where the ‘parity checking’ step is carried out are defined as the trellis nodes that correspond to a column which contains the last ‘1’ in any row of the parity check matrix, $\mathbf{H}$, used to encode the information sequence. Figure 4.6 illustrates this with an example. In the figure, a section of the $\mathbf{H}$ matrix of a short random LDPC code having $N = 64$ is shown. The last ‘1’ in each row of $\mathbf{H}$ occurs in column index 61, 63 and 64. Correspondingly, the

parity check nodes in the trellis of the JVDD will be defined as the trellis nodes at time index 61, 63 and 64.

Figure 4.6 Assigning parity check nodes in a JVDD trellis.

The JVDD algorithm performs ‘parity checking’ at the parity check nodes of the trellis by conducting syndrome checks on survivor paths. The H matrix of a linear block code, C , describes the relations that the coded bits must satisfy in order to be validated as a legal codeword of C . The syndrome check can be expressed mathematically as

$$cH^T = 0 \quad (4.8)$$

where c refers to a coded sequence. Each of the M rows of H represent a parity check equation. In each row, the bits are the coefficients of the parity check equation. They show how the modulo-2 addition of certain bits of each codeword adds up to 0. Through the syndrome checks, the ‘parity checking’ step in the JVDD algorithm discards survivor paths that fail the parity check and thus are not valid codewords. Figure 4.7 provides an illustration on the ‘metric thresholding’ task that is performed at all trellis nodes and an additional ‘parity checking’ task that is executed at the parity check nodes.

By combining the intrinsically dependent task of detection and decoding in a single step, the JVDD algorithm avoids the need for numerous iterations that is required in today’s state-of-the-art iterative detector-decoder. By doing so, it tackles the problem of sub-optimality associated with the iterative process. Aside from that, the modified Viterbi algorithm employed in the ‘metric thresholding’ step of the JVDD algorithm is optimal over an uncoded ISI channel [74]. The JVDD algorithm is therefore optimal over a coded ISI channel as a result. The JVDD,

however, suffers from a large accumulation of survivor paths in its trellis. In the ‘metric thresholding’ segment of the algorithm for example, the number of survivor paths can grow to a very large number if a large T is used. A large T is however necessary for a good performance. Various methods of managing the growth in survivors will be proposed in section 4.1.3.

Figure 4.7: Metric thresholding and parity checking steps in the JVDD algorithm.

4.1.2 2D-JVDD

The JVDD algorithm proposed in section 4.1.1 is only applicable to 1D channels. In order to apply this joint detection and decoding scheme to 2D magnetic recording channels, a non-binary adaptation of the algorithm needs to be implemented. This section proposes the modifications needed to transition from a binary JVDD to a non-binary JVDD (2D-JVDD).

Figure 4.8: Block diagram depicting non-binary sequences for a 2D channel.

Given a block diagram for a 2D system as shown in Figure 4.8, the input to the 2D channel consists of multiple tracks of binary sequences. These input sequences can be represented in a non-binary form by grouping the bits together in the cross-track direction to form a symbol

vector as illustrated in the figure. This can be done likewise for the read-back sequence. In general, the symbol vectors for the channel input, $\mathbf{c}_k$, and the read-back, $\mathbf{y}_k$, at any time instant k can be expressed as:

$$\mathbf{c}_k = [c_{0,k} \ c_{1,k} \ \dots \ c_{N_b,k}]^T \quad (4.9)$$

$$\mathbf{y}_k = [y_{0,k} \ y_{1,k} \ \dots \ y_{N_b,k}]^T \quad (4.10)$$

where N_b is the number of bits per symbol. For a two-dimensional magnetic recording channel (SMR or TDMR channel), N_b can be defined as

$$N_b = N_{track} \quad (4.11)$$

where N_{track} refers to the number of tracks written in a block. Each input symbol can take on 2^{N_b} values denoted by

$$\mathbf{c}_k \in \mathcal{C} \quad \text{where } \mathcal{C} = \{\hat{c}^{(0)}, \hat{c}^{(1)}, \dots, \hat{c}^{(2^{N_b}-1)}\} \quad (4.12)$$

As a state diagram for 2D channels in general differs from that of 1D channels, the structure of the channel trellis in which the 2D-JVDD algorithm executes on is therefore defined differently from that of 1D-JVDD. For the 2D-JVDD, the number of trellis states, N_{state} , and the number of branches emerging from any trellis node (or merging into any trellis node), N_{branch} , is defined by

$$N_{state} = 2^{L_{mem} \cdot N_{track}} \quad (4.13)$$

$$N_{branch} = 2^{N_b} \quad (4.14)$$

where L_{mem} refers to the memory of the channel. The channel trellis for the 2D-JVDD therefore has a larger number of states than that for 1D-JVDD and its branches represent a symbol instead of a bit. This gives an initial indication that a greater number of survivor paths and therefore higher complexity is expected in the 2D-JVDD as compared to 1D-JVDD.

Aside from the need to re-construct the channel trellis for a 2D case, the shift from a 1D to a 2D channel implies a transition from a 1D convolution of a binary input sequence with the channel response to a 2D convolution. As such, the computation of branch metrics in the ‘metric thresholding’ section needs to be formulated accordingly for the 2D-JVDD. In other words, the

MLSD for 2D detection would involve the computation of conditional pdf for a sequence of symbols. This is in contrast to 1D detection where the MLSD calculates the conditional pdf for a sequence of bits. This can be derived mathematically as follow – Given a 2D ISI channel, its 2D read-back signal, $\mathbf{y}$, can be expressed in terms of the 2D transmitted signal, $\mathbf{c}$, as:

$$y_{j,k} = \sum_{m=0}^{M_h-1} \sum_{l=0}^{N_h-1} h_{m,l} c_{j-m,k-l} + n_{j,k} \quad (4.15)$$

where $\mathbf{h}$ is the 2D channel response of size $M_h \times N_h$ and $\mathbf{n}$ is the AWGN noise. The estimate of the 2D transmitted sequence, $\hat{\mathbf{c}}$, can be obtained by maximizing the probability $P(\mathbf{y}|\mathbf{c})$ as shown in (4.16).

$$\begin{aligned} \hat{\mathbf{c}} &\propto \arg \max_{\mathbf{c}} P(\mathbf{y}|\mathbf{c}) \\ &\propto \prod P(y_{j,k} | c_{j,k}, \dots, c_{j-M_h+1,k-N_h+1}) \end{aligned} \quad (4.16)$$

Since the channel noise is AWGN in nature, $p(y_{j,k} | c_{j,k}, \dots, c_{j-M_h+1,k-N_h+1})$ is given by:

$$P(y_{j,k} | c_{j,k}, \dots, c_{j-M_h+1,k-N_h+1}) = \frac{1}{\sqrt{\pi N_0}} e^{-\frac{|y_{j,k} - \sum_{m=0}^{M_h-1} \sum_{l=0}^{N_h-1} h_{m,l} c_{j-m,k-l}|^2}{N_0}} \quad (4.17)$$

The 2D-JVDD makes an estimate of the written 2D sequence, $\hat{\mathbf{c}}$, by summing the Euclidean distance between the 2D read-back sequence and the 2D noiseless output sequence for every track (4.18).

$$\hat{\mathbf{c}} \propto \prod \left| y_{j,k} - \sum_{m=0}^{M_h-1} \sum_{l=0}^{N_h-1} h_{m,l} c_{j-m,k-l} \right|^2 \quad (4.18)$$

The 2D-JVDD algorithm not only performs detection but it also carries out the task of simultaneous decoding, similar to its 1D counterpart. For 2D systems that consist of multiple 1D encoders and a 2D channel such as depicted in Figure 4.8, the 2D-JVDD algorithm can be formulated such that its ‘metric thresholding’ step is performed in the 2D domain while the ‘parity checking’ step is carried out in 1D. This essentially requires the conversion of survivor

paths representing symbol sequences to binary bits prior to every ‘parity checking’ step in the 2D-JVDD.

4.1.3 Managing Complexity of JVDD

The ‘metric thresholding’ step of the JVDD algorithm can result in a large accumulation of survivor paths in the trellis, with the growth in number of paths being correlated with T as explained in section 4.1.1. In contrast, the ‘parity checking’ segment of the algorithm culls survivor paths drastically. Each parity checking step effectively halves the number of paths on average. To advance the JVDD as a viable alternative to the state-of-the-art iterative detector-decoder, the computational complexity of the algorithm (measured in terms of the number of survivor paths in the trellis) has to be effectively managed. This section proposes various ways of reducing the number of survivors in the JVDD algorithm (note that the proposed methods can be directly applied to the 2D-JVDD).

An example of the rapid accumulation of survivor paths in the JVDD trellis can be shown in the plot of Figure 4.9. The plot is obtained using a short binary LDPC codeword with $N = 256$ and code rate, $R = 0.90$. The figure illustrates an exponential growth in survivor paths over a large section of the trellis followed by a sharp drop in survivor count (at $\sim$ bit index = 230) upon the onset of the ‘parity checking’ steps in the JVDD. For longer codeword lengths, the region in Figure 4.9 where survivors accumulate will expand, thereby allowing survivor paths to accumulate for a longer period of time before elimination due to parity checks takes effect. Two methods can be proposed to curb this growth.

Firstly, a limit on the maximum number of allowable survivors, S_{max} , in the trellis can be imposed. To implement this, the survivors are stored in an ordered list and only the S_{max} smallest metric survivors are retained. This ensures that the growth in survivor paths is contained and computational complexity of the JVDD algorithm remains manageable. Imposing such limit to lower the complexity however comes with the trade-off of performance loss as the

MMLC undergoes a greater risk of being discarded during the ‘metric thresholding’ part of the algorithm. This claim can be supported with an FER plot shown in Figure 4.10(a) where the performance of the JVDD with random LDPC code having $N = 512$ and $R = 0.90$ is seen to improve with larger S_{max} , up to a saturation point beyond which no further performance improvement can be attained. The corresponding complexity plots of the Figure 4.10(a) illustrate the growth in average number of survivors (Figure 4.10(b)) as well as total number of survivors in the trellis (Figure 4.10(c)). In Figure 4.10(c), at SNR = 8 dB, the growth in survivors is cropped over a large part of the trellis resulting in high error rate when S_{max} is small ($S_{max} = 100$). No cropping occurs for $S_{max} = 9000$ which thus leads to a lower FER seen in Figure 4.10(a).

Figure 4.9: Number of survivors in a JVDD trellis for a short binary LDPC code.

Figure 4.10: Effect of S_{max} on: (a) performance and computational complexity measured in terms of (b) average number of survivors and (c) total number of survivors in the trellis.

In addition to introducing S_{max} , the computational complexity of JVDD can be reduced more significantly by designing a novel class of ECC codes specifically for the JVDD algorithm. The state-of-the-art iterative detector-decoder system at present makes use of random LDPC or

Quasi-Cyclic LDPC codes. Both codes are known to perform well with the SPA that is used in the decoder component of the iterative detector-decoder [55][56]. The use of either of these codes with the proposed JVDD however results in an exceptionally large number of survivor paths in the trellis. With the introduction of the S_{max} limit, deterioration in performance of the JVDD will occur.

Figure 4.11: $\mathbf{H}$ matrix of a random LDPC code.

For a random LDPC code for example, its $\mathbf{H}$ matrix can be represented pictorially as shown in Figure 4.11 where black and white pixels of the figure correspond to bit '1' and '0' respectively. The last '1' in each row of $\mathbf{H}$ are shown in red. In the diagram, the clustering of the last '1' bit in every parity check row towards the end of $\mathbf{H}$ is evident. As a consequence, when such code is used with the JVDD algorithm, the 'parity checking' step occurs only when the trellis is approaching its end. The number of survivors from the 'metric thresholding' step is therefore allowed to accumulate for a large section of the trellis before the bulk of them eventually get eliminated by the parity checks towards the end. If the parity checks can be made to occur at an earlier part of the trellis and at a more regular interval, the exponential growth in survivors can be curbed. In view of that, a novel class of codes is proposed specifically for the JVDD and is presented in the next section. This class of codes has been coined the "JVDD codes".

4.2 JVDD Class of Codes

In this section, two classes of JVDD codes are proposed – 1) Gaussian Distribution Linear Diagonal (GDLD) and 2) proportionate row weight codes. Both codes are proposed specifically for the JVDD. They are designed with the objective of ensuring that the last parity bit in every row of $\mathbf{H}$ is spread across the matrix. This is not a requirement for codes used in other detection and decoding systems including the state-of-the-art iterative detector-decoder. As such, the

JVDD codes are expected to perform well only when used with the JVDD over an ISI channel, and not for other detection/decoding schemes.

4.2.1 Gaussian Distribution Linear Diagonal Codes

A GDLD code is designed with a linear diagonal stretching across the matrix. A '1' bit is first placed on the diagonal that extends from the top left corner of the matrix to its bottom right corner, as illustrated in Figure 4.12(a). These bits represent the last parity bit of each row in $\mathbf{H}$. The remaining parity bits of $\mathbf{H}$ are placed to the left of this diagonal, as depicted in Figure 4.12(b). The exact number of such bits in each row of $\mathbf{H}$ is defined by the row weight of the code, w_{row} . In every row, the positions of each of these parity bits are randomly assigned using a Gaussian distribution - a sigma, σ , is defined for every $\mathbf{H}$ designed such that the remaining parity bits are placed within the σ of the Gaussian distribution. When such allocation of parity bits are performed on every row, a $\mathbf{H}$ matrix such as that in Figure 4.12(c) results. This $\mathbf{H}$ matrix is however not ideal as the presence of a linear diagonal that ends at the bottom right corner of the matrix will result in the last few bits of $\mathbf{c}$ being included in an insufficient number of parity checks. As matrices with low column weights, w_{col} (the number of '1's in each column), are known to perform poorly [72], this should be avoided for JVDD codes. As such, a third parameter is introduced in the design of GDLD code – *offset*. This parameter defines the horizontal offset of the diagonal from its origin (top corner of $\mathbf{H}$). Figure 4.12(d) shows the final $\mathbf{H}$ matrix of a GDLD code having $N = 512$, $R = 0.90$, $w_{row} = 150$, $\sigma = 200$ and *offset* = 50.

The introduction of GDLD codes significantly reduces the computational complexity of JVDD. Placing the last parity bit of each row on the diagonal restricts survivor paths from multiplying rapidly. Figure 4.13 compares the total number of survivor paths in the JVDD when the detector is used with random LDPC and GDLD codes, for $N = 512$ and 1024, and $R = 0.90$. The parameters of the GDLD codes in the plots are as follow – $w_{row} = 150$, $\sigma = 200$ and *offset* = 100 for $N = 512$ (Figure 4.13(a)) and $w_{row} = 200$, $\sigma = 250$, and *offset* = 160 for $N = 1024$ (Figure 4.13(b)). The

significant reduction in number of survivors in the figure is apparent. For $N = 1024$ for example, an average reduction in complexity of about 85% is reaped with GDLD code.

Figure 4.12: Construction of $\mathbf{H}$ matrix of a GDLD code.

The combination of the design parameters chosen for the GDLD codes can impact the performance and complexity of the JVDD in varying ways. A well selected combination of parameters can substantially reduce the computational complexity of the algorithm and enhance its performance at the same time. The next section examines the design considerations of GDLD codes.

Figure 4.13: Comparison of number of survivors for JVDD with LDPC and GDLD code each at length $N =$ (a) 512 and (b) 1024.

4.2.1.1 Impact of GDLD Codes on JVDD

The w_{col} of $\mathbf{H}$, the minimum distance, d_{min} , and the position of the linear diagonal of GDLD codes can directly affect the performance and computational complexity of JVDD. These factors have to be taken into account when generating GDLD codes. A code which has weakly protected columns is known to result in poor performance. In the construction of GDLD codes, the presence of weakly protected columns may arise due to two reasons. Firstly, designing a sparse parity check matrix may cause some columns in $\mathbf{H}$ to be inadequately protected. A sparse $\mathbf{H}$ can arise as a consequence of the choice of w_{row} and σ used during code construction. More specifically, when w_{row} is much smaller than σ , parity bits of $\mathbf{H}$ are spread across a larger width, leaving more empty spaces between them (Figure 4.14(a)). With a denser $\mathbf{H}$ (Figure 4.14(b)), these gaps are filled up, reducing the number of columns having small w_{col} .

Figure 4.14: Comparison of a (a) sparse and (b) dense matrix of GDLD code.

Secondly, weakly protected columns of GDLD codes can occur due to the choice of values for the *offset* parameter. This parameter shifts the diagonal of $\mathbf{H}$ to the right from the top left corner of $\mathbf{H}$ (Figure 4.12). When this horizontal shift is insufficient, as shown in Figure 4.15(a), the last few columns in $\mathbf{H}$ will be weakly protected. Some columns may even have no protection at all. This can be overcome with a larger *offset* value. However, if *offset* becomes too large, the last parity bits which are positioned on the diagonal get pushed towards the end of $\mathbf{H}$ as shown in Figure 4.15(b), and cause the ‘parity checking’ in the JVDD algorithm to start later. This in turn would cause the complexity of JVDD to increase.

Aside from the distribution of w_{col} , the performance of the JVDD will be directly affected by the d_{min} of the GDLD codes. The distance, d , between two codewords, $\mathbf{c}_1$ and $\mathbf{c}_2$, of any linear block

code refers to the number of positions in which the two codewords differ. Mathematically, this can be expressed as

$$d(\mathbf{c}_1, \mathbf{c}_2) = w(\mathbf{c}_1 \oplus \mathbf{c}_2) \quad (4.19)$$

where w is the Hamming weight. Mathematically, d_{min} refers to the minimum value of $d(\mathbf{c}_1, \mathbf{c}_2)$ over all codewords $\mathbf{c}_1$ and $\mathbf{c}_2$. Given that GDLD codes are linear block codes and using the theory of ECC, the sum of any two GDLD codewords is therefore another codeword. To this end, the all-zero vector can also constitute a codeword. Given these, d_{min} of GDLD can hence be expressed as

$$\begin{aligned} d_{min} &= \min d(\mathbf{c}_1, \mathbf{c}_2) \quad \text{where } \mathbf{c}_1 \neq \mathbf{c}_2 \text{ and } \mathbf{c}_1, \mathbf{c}_2 \in \mathcal{C} \\ &= \min d(\mathbf{c}, \mathbf{0}) \quad \text{where } \mathbf{c} \neq \mathbf{0} \text{ and } \mathbf{c} \in \mathcal{C} \\ &= \min w(\mathbf{c}) \quad \text{where } \mathbf{c} \neq \mathbf{0} \text{ and } \mathbf{c} \in \mathcal{C} \end{aligned} \quad (4.20)$$

When constructing GDLD codes, the values of the design parameters chosen will impact $w(\mathbf{c})$ in different ways. For a chosen set of design parameters, d_{min} therefore needs to be computed to give a prediction on the potential outcome with JVDD. In general, the error correcting capability, t , of a GDLD code can be computed as

$$t = \left\lfloor \frac{d_{min} - 1}{2} \right\rfloor \quad (4.21)$$

Based on equation (4.21), a combination of GDLD design parameters that leads to a small d_{min} value should be avoided in order to ensure a better error correcting capability and thus error rates of JVDD.

In summary, the overall combination of design parameters of the GDLD has a direct impact on the distribution of w_{col} and d_{min} , which in turn affects the performance of the JVDD. The complexity of the JVDD, on the other hand, is sensitive to the position of the linear diagonal of the code. To further comprehend the characteristics associated with GDLD codes, the effect of individual design parameter on JVDD is studied experimentally through the use of extensive computer simulations and presented in the subsequent sub-section.

Figure 4.15: Comparison of a (a) small and (b) large *offset* value of GDLD code.

4.2.1.2 Simulations and Results for GDLD Codes

In this section, the impact of each design parameter of GDLD code on the performance and computational complexity of JVDD is thoroughly investigated through the use of computer simulations. Coded simulations were run over an ISI channel with constant SNR, N and R . Three sets of experimental studies were carried out – varying w_{row} , σ and *offset* of GDLD codes. For each set of experiment, only one variable is varied while the rest are kept constant. The constant and variable parameters used are listed in Table 4.1 and 4.2 respectively.

Table 4.1: Constant parameters used in experimental study of GDLD codes

channel	$\mathbf{h}_k$	[0.1667,0.3333,0.1667]
	SNR	8 dB
code	N	512
	R	0.9
JVDD	S_{max}	10 000
	T	2

Table 4.2: Variable parameters used in experimental study of GDLD codes

w_{row}	50, 100, 150, 200, 250
σ	50, 100, 150, 200, 250, 300
<i>offset</i>	0, 50, 100, 150, 200, 250

For the simulations conducted in this section as well as in remaining parts of the thesis, the following definition of SNR is used:

$$SNR = 10 \log_{10} \frac{E_b}{N_0} = 10 \log_{10} \frac{E_c}{2R\sigma_{noise}^2} \quad (4.22)$$

where E_b is the user bit energy, $E_c = RE_b$ is the coded bit energy, σ_{noise}^2 refers to the 2-sided PSD of the Additive White Gaussian Noise (AWGN). The simulations results are presented as follow:

1. Effect of w_{row}

Figure 4.16 shows FER plots obtained for various values of w_{row} and with $offset$ and σ held fixed at a few different values. The curves suggest that a small σ (Figure 4.16(a)) results in poor performance (FER ~ 0.5), regardless of the value of w_{row} and $offset$. In general, the error-rate of JVDD is high when w_{row} is considerably smaller than σ , regardless of the $offset$ value. The trends of Figure 4.16(b) and 4.16(c) suggest that in general an optimal w_{row} value exists. For example, in Figure 4.16(b), the optimal w_{row} is approximately 100 for $offset = 150$. Comparing Figure 4.16(b) and 4.16(c), it appears that the optimal value of w_{row} increases with σ . In fact, it correlates with σ and is slightly smaller than it. The optimal w_{row} is however, relatively independent of $offset$, as seen in these plots. In addition to that, these figures also indicate that JVDD performs poorly when w_{row} is much larger than σ .

Figure 4.16: Performance plots of GDLD codes with varying w_{row} and $\sigma =$ a) 50 b) 150 and c) 250.

The simulation results obtained above can be explained as follow: When w_{row} is considerably smaller than σ , the parity bits in $\mathbf{H}$ are sparsely distributed to the left of the diagonal causing some columns in $\mathbf{H}$ to be inadequately protected, as discussed earlier in section 4.2.1.1. This can be supported by a plot of the distribution of w_{col} in $\mathbf{H}$ (Figure 4.17). For $w_{row} = 50$, it can be seen that w_{col} at some indexes is 0 or close to 0. This is not the case with larger w_{row} at the same σ value. Sufficient column protection is ensured when $w_{row} = 150$ as seen in Figure 4.17. This

justifies the improvement in FER seen in Figure 4.16. Having too much column protection, ($w_{row} = 250$), could however be redundant. In fact, when w_{row} becomes relatively larger than σ , $\mathbf{H}$ approaches a more deterministic matrix with large areas densely packed with '1's. The parity bits in $\mathbf{H}$ form a band-like structure as shown in Figure 4.18 and are the cause of the high FER values obtained.

Figure 4.17: A distribution plot of w_{col} of GDLD codes with $\sigma=150$, $offset=150$ and various w_{row} .

Figure 4.18: Band-like structure in a deterministic GDLD $\mathbf{H}$ matrix.

The dependency of w_{row} with σ can be further supported by analyzing the d_{min} values of the codes. The d_{min} values were obtained by finding the minimum non-zero weight of all possible codewords of each code used. Figure 4.19 shows plots of d_{min} corresponding to the performance results obtained in Figure 4.16. The d_{min} plots affirm the FER trends seen in Figure 4.16 (note that the red curves are less visible as they are overlapped by the green curves). That is, a code with a larger d_{min} performs better. For example, in Figure 4.19(b), a w_{row} value between 100 to 150 for $offset = 150$ results in a region of lowest FER (~ 0.0001) in Figure 4.16(b).

For the performance plots obtained in Figure 4.16, its corresponding computational complexity, measured in terms of the average number of survivors in the trellis, is shown in Figure 4.20. A common trend can be deduced - the complexity of JVDD first decreases with increasing w_{row} ,

and increases thereafter for all values of $offset$. It seems to be less affected by σ . In addition to that, the complexity curve for $offset = 150$ appears higher than any other $offset$ values. This will be addressed in the last part of this sub-section when the $offset$ parameter is presented.

Figure 4.19: Plots of d_{min} of GDLD codes with varying w_{row} and $\sigma =$ a) 50 b) 150 and c) 250.

Figure 4.20: Complexity plots of GDLD codes with varying w_{row} and $\sigma =$ a) 50 b) 150 and c) 250.

2. Effect of σ

The effect of σ on the performance and complexity of JVDD is studied next. FER plots representing the variation in σ for fixed sets of w_{row} and $offset$ are shown in Figure 4.21. These plots suggest that the performance of the GDLD codes depend jointly on σ and w_{row} . This is evident in the plots where larger FER values are seen when σ is much smaller than w_{row} . Figures 4.21(b) and 4.21(c) suggest that increasing σ relative to w_{row} leads to an improvement in performance (for a given w_{row} and $offset$), up to a saturation point beyond which no further gain can be reaped.

The dependency of σ with w_{row} can be explained with the same reasoning for the w_{row} study in part 1. That is, when σ is smaller than w_{row} , a resulting $\mathbf{H}$ matrix that is deterministic with a large dense region of '1's around the diagonal is a poor code that gives rise to bad error rates. Plots of d_{min} shown in Figure 4.22 further support the results obtained. The d_{min} plots in Figure 4.22 correspond to the codes used in the performance plots of Figure 4.21 and are seen to complement the performance outcome obtained (note that the red curves are less visible as they are overlapped by the green curves). For example, codes with $d_{min} = 2$ causes a higher FER that range between 0.001 and 0.1 while the reverse is seen for codes with larger d_{min} values. In Figure 4.22(b) and 4.22(c), when σ is larger than 100 and 200 respectively, d_{min} remains constantly large and thereby gave rise to a performance saturation seen in Figure 4.21(b) and 4.21(c).

Figure 4.21: Performance plots of GDLD codes with varying σ and w_{row} = (a) 50 (b) 150 (c) 250.

Figure 4.22: Plots of d_{min} of GDLD codes for various σ and w_{row} = (a) 50 (b) 150 (c) 250.

The corresponding complexity plots for JVDD are presented in Figure 4.23. Scanning across these plots, the trends in computational complexity seen are relatively the same for all simulation sets – an initial decrease in complexity with increasing σ before it levels off. The

complexity curves for larger *offset* are seen to be higher than the other *offset* values. This will be addressed in the next set of results.

Figure 4.23: Complexity plots of GDLD codes with varying σ and w_{row} = (a) 50 (b) 150 (c) 250.

3. Effect of *offset*

The effects on performance and complexity of varying *offset* of GDLD are next analyzed. In Figure 4.24, FER plots for various *offset* are shown. These plots indicate that increasing *offset* generally brings about an improvement in performance, up to a point where further increase in *offset* causes performance to deteriorate. The poorest performance is seen when *offset* = 0.

Figure 4.24: Performance plots of GDLD codes with varying *offset* and w_{row} = (a) 50 (b) 150 and (c) 250.

The results obtained can be attributed to the distribution of w_{col} of the codes. For *offset* = 0, the diagonal in $\mathbf{H}$ stretches from the top left corner of the matrix to its bottom right corner without any horizontal offset. As such, the last few columns in $\mathbf{H}$ are inadequately protected and are the source of poor performance. When *offset* increases, w_{col} of the last few columns increases as well. The bits are now better protected, giving rise to improved FER. However, when *offset* gets too large, the diagonal of $\mathbf{H}$ is shifted to the right by too much. While this increases w_{col} for the

last few columns, it lowers w_{col} of the first few columns significantly and causes a performance loss. This can be supported with a plot of the distribution of w_{col} in $\mathbf{H}$ shown in Figure 4.25. This justifies the optimal trend displayed in the results of Figure 4.24.

Figure 4.25: Distribution plot of w_{col} in the $\mathbf{H}$ matrices of GDLD codes with $w_{row} = 150$, $\sigma = 200$ and various *offset*.

Figure 4.26: Plot of d_{min} of GDLD codes for various *offset* and $w_{row} =$ (a) 50 (b) 150 and (c) 250.

The d_{min} plots for the GDLD codes shown in Figure 4.26 support the performance trends seen in Figure 4.24. Varying *offset* also has its impact on the complexity of JVDD. Increasing *offset* generally brings about an increase in complexity, regardless of the values of w_{row} and σ (Figure 4.27). This is due to the fact that with larger *offset*, parity checking in the JVDD algorithm starts at a later part of the trellis, allowing survivors to accumulate in the trellis before the bulk of them get culled during ‘parity checking’. When *offset* gets too large, the number or accumulated survivors becomes significantly large. When this number approaches the S_{max} limit (set at 10 000 in this study) of JVDD, cropping of survivors takes place as illustrated in Figure

4.28. At this stage, the MMLC is likely to get discarded. This explains the performance lost seen in Figure 4.24 when *offset* is too large.

Figure 4.28: Plot of number of JVDD survivors for GDLD code having parameters $w_{row} = 150$, $\sigma = 200$ and various *offset*.

In summary, an optimum combination of w_{row} , σ and *offset* of a GDLD code at a given N and R value would give the JVDD its best performance and with the lowest computational complexity. To identify the optimum combination of design parameters, the next section proposes the adoption of an analytical optimization tool for this means.

4.2.1.3 Optimization of GDLD Codes

An optimum GDLD code is one that results in the lowest error-rates and computational complexity for the JVDD system. An analytical optimization method can be employed to obtain an optimum set of design parameters for a GDLD code at a fixed N and R . This analytical method is known as Response Surface Methodology (RSM) [76].

Figure 4.29: DOE technique attempts to optimize a response parameter given a set of control parameters.

The RSM is a DOE technique that establishes a relationship between a set of control parameters, x_i , and response parameter, y , of a complex system as depicted in Figure 4.29 [77]. It attempts to find a best-fit second order model to a set of experimental data and subsequently an optimum solution that maximizes/minimizes the target. The best-fit second order model of RSM can be expressed mathematically in terms of its control parameters as follow:

$$y(x_0, x_1, \dots, x_{L-1}) = A_0 + A_1x_0 + A_2x_1 + A_3x_2 + \dots + A_Lx_{L-1} + A_{L+1}x_0x_0 + A_{L+2}x_0x_1 + \dots + A_{M-1}x_{L-1}^2 \quad (4.23)$$

The second-order model in equation (4.20) can be re-written in vector form given by (4.24).

$$y(\mathbf{x}) = \sum_k A_k x'_k = \mathbf{a}^T \dot{\mathbf{x}} \quad (4.24)$$

where $y(\mathbf{x}) = y(x_0, x_1, \dots, x_{L-1})$, $\mathbf{a}^T = [A_0 \ A_1 \ A_2 \ \dots \ A_{M-1}]$ and $\dot{\mathbf{x}} = [1 \ x_0 \ x_1 \ \dots \ x_{L-1}^2]^T$.

The best-fit second order RSM model is one which minimizes the cost function, J , as follow:

$$\begin{aligned}
J &= \sum_{i=0}^{P-1} (y(\mathbf{x}_i) - y_i)^2 \\
&= \sum_{i=0}^{P-1} \mathbf{x}_i^T \mathbf{a} \mathbf{a}^T \mathbf{x}_i - 2 \mathbf{a}^T \mathbf{x}_i y_i + y_i^2 \\
&= \sum_{i=0}^{P-1} \mathbf{x}_i^T \mathbf{A} \mathbf{x}_i - 2 \mathbf{a}^T \mathbf{x}'_i y_i + y_i^2
\end{aligned} \tag{4.25}$$

where $\mathbf{A}$ is an $M \times M$ matrix defined as:

$$\mathbf{A} = \begin{bmatrix} A_0 A_0 & A_0 A_1 & \dots & A_0 A_{M-2} & A_0 A_{M-1} \\ A_1 A_0 & A_1 A_1 & \dots & A_1 A_{M-2} & A_1 A_{M-1} \\ \vdots & \vdots & \dots & \vdots & \vdots \\ A_{M-1} A_0 & A_{M-1} A_1 & \dots & A_{M-1} A_{M-2} & A_{M-1} A_{M-1} \end{bmatrix}$$

Alternatively, equation (4.25) can be fully expressed in terms of the vector and matrix elements as:

$$\begin{aligned}
J &= \sum_{i=0}^{P-1} [1 \quad x_0 \quad \dots \quad x_{L-1}]_i \begin{bmatrix} A_0 A_0 & \dots & A_0 A_{M-1} \\ A_1 A_0 & \ddots & A_1 A_{M-1} \\ \vdots & & \vdots \\ A_{M-1} A_0 & \dots & A_{M-1} A_{M-1} \end{bmatrix} \begin{bmatrix} 1 \\ x_0 \\ \vdots \\ x_{L-1} \end{bmatrix}_i \\
&\quad - 2 [A_0 \quad A_1 \quad \dots \quad A_{M-1}] \begin{bmatrix} 1 \\ x_0 \\ \vdots \\ x_{L-1} \end{bmatrix}_i y_i + y_i^2 \\
&= \sum_{i=0}^{P-1} [1 \quad x_0 \quad \dots \quad x_{L-1}]_i \begin{bmatrix} A_0 A_0 + A_0 A_1 x_0 + \dots + A_0 A_{M-1} x_{L-1} \\ A_1 A_0 + A_1 A_1 x_0 + \dots + A_1 A_{M-1} x_{L-1} \\ \vdots \\ A_{M-1} A_0 + A_{M-1} A_1 x_0 + \dots + A_{M-1} A_{M-1} x_{L-1} \end{bmatrix}_i \\
&\quad - 2(A_0 + A_1 x_0 + \dots + A_{M-1} x_{L-1})_i y_i + y_i^2 \\
&= \sum_{i=0}^{P-1} (A_0 A_0 + \dots + A_0 A_{M-1} x_{L-1} + \dots + A_1 A_{M-1} x_0 x_{L-1} + \dots + A_{M-1} A_{M-1} x_{L-1}^2)_i \\
&\quad - 2(A_0 + A_1 x_0 + \dots + A_{M-1} x_{L-1})_i y_i + y_i^2
\end{aligned} \tag{4.26}$$

The partial derivative of J with respect to A_k can be computed as follow:

$$\frac{\partial J}{\partial A_k} = \sum_{i=0}^{P-1} 2 \mathbf{a}^T \cdot \check{\mathbf{x}}_i - 2x_{k-1}y_i \quad (4.27)$$

where $\check{\mathbf{x}}_i = [x_{k-1} \quad x_{k-1}x_0 \quad \dots \quad x_{k-1}x_{k-1} \quad \dots \quad x_{k-1}x_{L-1}]_i^T$. Solving for A_k values in (4.27) requires equating the partial derivative of J with respect to $\mathbf{a}$ to 0. This solves for the optimum A_k values in the best fit second order model. To solve for the optimum set of control parameters of the complex system in Figure 4.29, the best-fit value of A_k can be substituted back into (4.27). The optimum x_i values, x_i^* , are computed by equating the partial derivative of $y(\mathbf{x})$ with respect to each x_i to 0 and subsequently solving L equations with L unknowns.

Table 4.3: DOE table consisting of control and response parameter sets

Run	<i>offset</i>	<i>w_{row}</i>	σ	<i>FC</i>
0	0	50	100	1.458
1	50	250	250	1.177
2	50	100	250	0.312
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
P-1	100	50	50	0.059

To establish a relationship between the design parameters of GDLD code and the performance and complexity of JVDD, *offset*, *w_{row}* and σ of the GDLD are used as control parameters. As we wish to obtain a set of code parameters that gives JVDD its best performance (lowest FER) and lowest complexity (smallest average number of survivors), the response parameter is defined as a metric, $\log FC$, in this optimization problem. This metric is shown in equation (4.28). The logarithmic function is used in the metric definition to avoid arriving at an invalid solution where the FER or average number of survivors is < 0 .

$$\log FC = \log(FER * \text{Average no. of survivors}) \quad (4.28)$$

Table 4.3 shows a sample set of control and response parameters used in the DOE for optimizing GDLD code at $N = 512$ and $R = 0.90$. The DOE was carried out in multiple phases. In

Phase 1, the DOE was performed over a wide range of control parameters given by Table 4.4. The optimum values of x_0 , x_1 and x_2 are obtained by computing the partial derivative of y with respect to x to 0 as shown in (4.27). The mathematical expression is then equated to 0 and the optimum set of control parameters are obtained as shown in (4.29).

Table 4.4 Range of control parameters used in Phase 1 of DOE

x_0 (<i>offset</i>)	0, 100, 200
x_1 (<i>w_{row}</i>)	50, 100, 200, 300
x_2 (σ)	50, 100, 200, 300

Table 4.5 Optimum set of control parameters obtained in Phase 1 of DOE

$*x_0$ (<i>*offset</i>)	96.36
$*x_1$ (<i>*w_{row}</i>)	74.22
$*x_2$ (<i>*σ</i>)	225.42

$$\nabla_x y = \begin{bmatrix} \frac{\partial y}{\partial x_0} \\ \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \end{bmatrix} = \begin{bmatrix} A_1 + 2A_4x_0 + A_5x_1 + A_6x_2 \\ A_2 + A_5x_0 + 2A_7x_1 + A_8x_2 \\ A_3 + A_6x_0 + A_8x_1 + 2A_9x_2 \end{bmatrix} \quad (4.29)$$

$$\begin{bmatrix} x_0^* \\ x_1^* \\ x_2^* \end{bmatrix} = \begin{bmatrix} 2A_4x_0 + A_5x_1 + A_6x_2 \\ A_5x_0 + 2A_7x_1 + A_8x_2 \\ A_6x_0 + A_8x_1 + 2A_9x_2 \end{bmatrix}^{-1} \begin{bmatrix} -A_1 \\ -A_2 \\ -A_3 \end{bmatrix} \quad (4.30)$$

Solving (4.30), the optimum set of control parameters in Phase 1 of the DOE is given by Table 4.5. A plot comparing the analytical values of the response parameter computed from Phase 1 of the DOE and experimental values gathered from computer simulations is shown in Figure 4.30(a). A close fit between the two sets of values is evident. The solution of the DOE can be further refined by improving the range of the control parameters based on the optimum solution obtained in Phase 1. The DOE can be repeated in multiple phases until eventually the best solution is obtained. For the example illustrated in this section, the DOE was carried out until Phase 4. The improved range for the control parameters used in this phase is shown in Table 4.6 and the corresponding set of optimum design parameters of GDLD code is presented in Table 4.7. A comparison of the analytical and experimental values for Phase 4 of the DOE is shown in Figure 4.30(b). Comparing the plots of Figure 4.30, a better fit is seen between the

analytical and experimental curves in Phase 4 than in Phase 1. In Phase 4, J was also found to be lower than that in Phase 1 ($J = 6.806$ in Phase I and $J = 0.772$ in Phase 4).

Table 4.6 Range of control parameters used in Phase 4 of DOE

x_0 (offset)	0, 100, 125
x_1 (w_{row})	50, 100, 150, 200
x_2 (σ)	100, 200, 225

Table 4.7 Optimum set of control parameters obtained in Phase 4 of DOE

$*x_0$ ($*offset$)	89.85
$*x_1$ ($*w_{row}$)	98.65
$*x_2$ ($*\sigma$)	162.28

Figure 4.30: Comparison of experimental and analytical values obtained in (a) Phase 1 and (2) Phase 4 of DOE.

To confirm that the optimum code found in Phase 4 of the DOE is indeed an optimum one for a fixed N and R , its performance and complexity was compared against other good performing GDLD codes (found experimentally) having a different set of design parameters. This is shown in an FER vs complexity plot in Figure 4.31. This plot shows the relationship between the performance and complexity of the JVDD when used with a fixed code. The performance-complexity relationship is obtained by varying T of the JVDD. On this plot, the curve that is closest to the origin signifies the best curve as it achieves the best performance with the lowest complexity. Therefore from Figure 4.31, it can be concluded that the GDLD code that has been optimized analytically performs best. This code optimization technique was subsequently used to optimize GDLD codes at various N and R values.

Figure 4.31: FER vs Complexity plot of optimum GDLD code obtained analytically from DOE against other code design parameters.

4.2.2 Proportionate Row Weight Codes

In view of some of the shortcomings of the GDLD codes discussed in section 4.2.1.1, another type of JVDD code known as the proportionate row weight code is proposed specifically for the JVDD. The $\mathbf{H}$ matrix of this class of code is similar to that of GDLD codes – it has a linear diagonal with ‘1’ bits placed randomly to the left of the diagonal. For proportionate row weight codes however, the top corner of the diagonal in $\mathbf{H}$ can be displaced horizontally to the right by a percentage d_x of N , and its bottom right corner can be shifted vertically upwards by a percentage d_y of M . In other words, proportionate row weight codes consist of a variable gradient, unlike GDLD code whose gradient is fixed. Such design is proposed with the aim of eliminating the need for a performance and complexity trade-off that is unavoidable with a fixed gradient. In addition to that, instead of defining an absolute value of w_{row} as in GDLD codes, the number of ‘1’s placed in every row of the $\mathbf{H}$ matrix of a proportionate row weight code is defined as 50% of the row width, L_{row} . This is to avoid the existence of any localized dense region in $\mathbf{H}$ which has been shown in section 4.2.1.1 to cause poorer performance. Figure 4.32 represents the $\mathbf{H}$ matrix of a proportionate row weight code with the following properties: $N = 512$, $R = 0.90$, $d_x = 5\%$, $d_y = 10\%$ and $\sigma = 200$.

Figure 4.32: $\mathbf{H}$ matrix of a proportionate row weight JVDD code.

Similar to the case of GDLD codes, an optimum combination of the design parameters of proportionate row weight codes has to be identified at every N and R value in order to attain the best performance and lowest complexity for the JVDD. The analytical optimization technique that is presented for GDLD codes in section 4.2.1.2 can similarly be adopted here. The subsequent section discusses this.

4.2.2.1 Optimization of Proportionate Row Weight Codes

The three design parameters of proportionate row weight codes (d_x , d_y and σ) can be jointly optimized to give JVDD its best performance with the lowest complexity. The RSM optimization for this class of codes involves three control parameters (the design parameters) and one response parameter ($\log FC$).

Figure 4.33: FER vs Complexity plots of optimum proportionate row weight code at $R = 0.90$ and $N =$ (a) 512 and (b) 2048.

Figure 4.34: $\mathbf{H}$ matrices of optimum proportionate row weight code at $N =$ (a) 512 and (b) 2048.

The proportionate row weight codes were analytically optimized at various N and R values. In Figure 4.33, FER vs Complexity plot of the optimum proportionate row weight code at $N = 512$ and 2048 and $R = 0.90$ are shown. On these plots, the optimum code found analytically using RSM gives the lowest FER and complexity, as compared to other good proportionate row weight codes that were identified through computer simulations. The $\mathbf{H}$ matrices of the optimum codes of Figure 4.33 are illustrated in Figure 4.34. A good spread of '1's and a gentle gradient across the matrices signifies a good performing code.

Figure 4.35: Comparison of optimized JVDD codes in terms of (a) FER vs Complexity at a fixed channel SNR and (b) FER vs SNR at a fixed complexity.

The performance of the optimum proportionate row weight codes can be compared against that of the optimum GDLD codes found at a fixed N and R . A comparison of these codes on a FER vs Complexity plot (for a fixed channel SNR) shown in Figure 4.35(a) depicts a more favorable outcome from the proportionate row weight code than the GDLD code. For example, to achieve an FER of at most 0.01, the optimum proportionate row weight code at $R = 0.90$ requires a complexity of approximately 200 times lesser than an optimum GDLD code. In Figure 4.35(b), the performance of the JVDD over various channel SNR is compared for both codes, at the same level of complexity. The figure indicates that the proportionate row weight code

results in a better overall performance for JVDD than the GDLD code, with an approximate margin of about 0.5 dB.

4.3 Performance and Complexity Analysis of JVDD

In section 4.2, a class of JVDD codes was proposed specifically for the JVDD and the relationship between the design parameters of these codes with the performance and complexity of the JVDD was established. The codes were then optimized analytically using an RSM technique for a fixed channel, N and R . In this section, the performance and complexity of JVDD when used with the optimized JVDD codes is studied in greater detail over varying channel types, channel noise level, N and R . To access the advantage gain from the novel technique of joint detection and decoding over a channel trellis, the performance of the JVDD is benchmarked against the state-of-the-art iterative detector-decoder used in the industry today [7][8]. The iterative system is widely used in the literature. In [48] for example, magnetic recording researchers employ the iterative detector-decoder on the receiver end to prediction the future of emerging channel technology. Random LDPC codes are typically used with the iterative detector-decoder over magnetic channels [6][7][55]. In [78], the performance of random LDPC code having column weight, $w_{col} = 4$ was shown to perform well with the iterative detector-decoder. This code is thus used for the iterative detector-decoder discussed in this thesis.

Extensive computer simulations over 3 different channels were carried out for the proposed JVDD and the conventional iterative detector-decoder. Firstly, a comparison between the JVDD and the iterative detector-decoder over a simple 1D ISI channel representative of a magnetic recording system was performed. Secondly, the comparison of the two detection and decoding schemes was extended to a more realistic magnetic recording channel using the GFP model presented in Chapter 2. Unlike the ISI channel used in the first set of study, the GFP channel captures the noise coloration that arises from media irregularities and thus, it can better represent a practical recording channel than a 1D ISI channel. Lastly, the performance and complexity of 2D-JVDD were evaluated over a 2D ISI channel.

4.3.1 JVDD Over a 1D ISI Channel

In this set of study, channel simulations were carried out over a 1D ISI channel with the objective of evaluating the performance and complexity of JVDD against the state-of-the-art iterative detector-decoder over the said channel. Random LDPC codes with $w_{col} = 4$ are used for the iterative detector-decoder while optimum GDLD codes found in section 4.2 are used for the JVDD. Each simulation was carried out over an ISI channel with varying noise and with code properties (N and R) kept fixed. The parameters used are listed in Table 4.8. A performance comparison between the JVDD and iterative detector-decoder is first presented (section 4.3.1.1), followed by a comparison of complexity (section 4.3.1.2).

Table 4.8: Parameters used in JVDD study over 1D ISI channel

Detector	JVDD, Iterative detector-decoder		Code	JVDD code (JVDD) random LDPC code (Iterative),	
N	512, 1024, 2048, 4096		R	0.95, 0.85	
h_k	[0.1667, 0.3333, 0.1667]		SNR	5 – 13 dB	
$JVDD$	T	2	$Iterative$	I_{global}	5
	S_{max}	10 K, 30 K		I_{LDPC}	150

4.3.1.1 Performance

Figure 4.36 presents performance plots, in terms of ‘FER vs SNR’, which were gathered from the computer simulations carried out in this study. The plots for $N = 512 - 4096$ generally displayed a performance gain of JVDD over the conventional iterative detector-decoder, at both low and high R and SNR values. For example, at $N = 512$ (Figure 4.36(a)), the JVDD showed an SNR gain of 15 % and 17 % for $R = 0.85$ and 0.95 respectively. In terms of FER, a gain factor of about 10 orders was seen. Figure 4.36 depict similar gains for $N = 1024$ and 2048. The performance gap between the 2 systems however reduces as N increases. For example, at $N = 4096$ (Figure 4.36(d)), the SNR gain between both systems drops to 1 % for $R = 0.95$. The performance gap shrinks further as the channel gets noisier and as R decreases, until eventually the conventional

iterative detector-decoder starts to outdo the JVDD. At $N = 4096$ and $R = 0.85$ (Figure 4.36(d)) for example, JVDD suffered an SNR loss of about 4 % and FER loss by almost a factor of 10.

The trend of decreasing performance gap with increasing N that is displayed by the 2 detector-decoder systems can be attributed to the fact that the LDPC decoder in the iterative detector-decoder performs better at larger N values [72]. In contrast, the performance of JVDD worsens as N increases or as the channel SNR drops. The phenomenon seen will be explained through a complexity analysis of JVDD in the subsequent section.

Figure 4.36: Performance comparison between JVDD and iterative detector decoder over an ISI channel at various SNR and R , and (a) $N = 512$ (b) $N = 1024$ (c) $N = 2048$ and (d) $N = 4096$.

4.3.1.2 Complexity

Following the performance comparison of the proposed JVDD and the conventional iterative detector-decoder in section 4.3.1.1, this section presents a complexity comparison for both systems. In this section, the number of multiplication, additions, memory consumption and computational time are used as complexity indicators.

i) Multiplications

In the JVDD algorithm, the multiplication steps occur during the computation of Euclidean distance between the received signal and noise-free sequence, in the ‘metric thresholding’ phase (4.6). The total number of multiplications in the algorithm, $JVDD_N_{mult}$, is dependent only on the channel and code length, and can be expressed as:

$$\begin{aligned} JVDD_N_{mult} &= N_{NFM} N_{trellis} \\ &= 2^{N_h-1} N_b (N + N_{gb}) \end{aligned} \quad (4.31)$$

where N_{NFM} and $N_{trellis}$ refers to the number of noise-free metrics and the trellis length respectively. N_{gb} refers to the number of guard band bits that are needed in order to properly terminate the trellis at state 0.

For the Iterative system, the multiplication steps in the BCJR detector occur during the computation of the α , β and γ metrics in its MAP algorithm (2.6)–(2.9). For the LDPC decoder, multiplications are largely used in the computation of likelihood messages between the check and parity nodes of the factor graph (2.17). The exact number of multiplications varies according to the sparseness of the $\mathbf{H}$ matrix and the convergence of the SPA. These can be mathematically expressed as (4.32) and (4.33). Consequently, the total number of multiplications in the Iterative algorithm is given by (4.34).

$$\begin{aligned} BCJR_N_{mult} &= N_\alpha + N_\beta + N_\gamma + N_{prob} \\ &= 2^{N_h-1} N_b N_{trellis} (4 - 2^{N_h} N_b) \end{aligned} \quad (4.32)$$

$$LDPC_N_{mult} \leq 2 I_{ldpc} M (N_{edge} - 1) \quad (4.33)$$

$$Iter_N_{mult} \leq I_{global} (BCJR_N_{mult} + LDPC_N_{mult}) \quad (4.34)$$

where N_α , N_β , N_γ and N_{prob} are the number of multiplications involved in computing the metrics and joint probabilities of the MAP algorithm. N_{edge} refers to the maximum number of edges between a check and variable node of the LDPC factor graph.

Figure 4.37 depicts the total number of multiplication used in both algorithms. While the channel SNR and R do not have any impact on the JVDD and BCJR, these factors largely influence the number of multiplications in the LDPC decoder, and thus the Iterative algorithm. In general, the JVDD has a much lesser multiplication steps than the Iterative, regardless of the channel noise and ECC code. On this end, the proposed JVDD is least costly to implement.

Figure 4.37: Number of multiplications in the JVDD and Iterative Detector-Decoder for (a) $N = 512$ and (b) $N = 2048$.

ii) Additions

Unlike multiplications, addition steps occur in both the ‘metric thresholding’ and ‘parity checking’ phases of the JVDD. In the ‘metric thresholding’ phase, the splitting of survivors at every node and the subsequent computation of branch/path metrics in (4.6) and thresholding in (4.7) employs numerous additions, as equated in (4.35). In the ‘parity checking’ phase, the number of additions are accumulated when syndrome checks are performed on trellis survivors in (4.8). The number of additions involved in this phase is largely dependent on the sparseness

and number of parities in the $\mathbf{H}$ matrix of the code used, as shown in (4.36). When an irregular row-weight code such as the JVDD code is used, (4.36) can alternatively be expressed as (4.37).

$$\text{No. of adds (thresholding)} = N_{NFM} + N_{NFM} N_{trellis} + N_{trellis} + N_{trellis} S_{avg} N_b \quad (4.35)$$

$$\text{No. of adds (parity checking)} = w_{row} M \quad (4.36)$$

$$\text{No. of adds (parity checking)} < \max(w_{row}) M \quad (4.37)$$

$$JVDD_N_{add} < N_{NFM} + N_{trellis} (N_{NFM} + 1 + S_{avg} N_b) + w_{row} N (1 - R) \quad (4.38)$$

The total number of additions in the JVDD algorithm when used with the proposed JVDD code is therefore given by (4.38). In this equation, S_{avg} refers to the average number of trellis survivors. Figure 4.38 shows a plot of S_{avg} for the JVDD performance in Figure 4.36. S_{avg} generally decreases with increasing SNR as there is a lesser likelihood of discarding the MMLC. S_{avg} however, increases with increasing N as a longer trellis section allows greater survivor accumulation. These channel and code effects on S_{avg} will in turn have a similar impact on $JVDD_N_{add}$ in (4.38).

Figure 4.38: Average number of survivors in the JVDD trellis for (a) $N = 512$ and (b) $N = 2048$.

On the contrary, for the iterative system, the addition steps in the BCJR arise from summation of branch metrics across the trellis (2.9). For the LDPC, the additions are attributed to the computation of logarithmic likelihood messages from the variable nodes to check nodes of the factor graph (2.16). These message computations could alternatively be carried out using multiplication steps, as in (2.13)-(2.14), but this would result in a more costly hardware

implementation. The total number of additions in the Iterative algorithm is shown in (4.41) where N_{ext} is the number of additions arising from the exchange of extrinsic information (2.1).

$$BCJR_N_{add} = 2^{N_h}(N_b - 1) N_{trellis} + 2(N_{trellis} - 1) \quad (4.39)$$

$$LDPC_N_{add} \leq 2 I_{ldpc} N (N_{edge} - 1) \quad (4.40)$$

$$Iter_N_{add} \leq I_{global}(BCJR_N_{add} + LDPC_N_{add} + N_{ext}) \quad (4.41)$$

Figure 4.39: Number of additions in the JVDD and Iterative Detector-Decoder for (a) $N = 512$ and (b) $N = 2048$.

Figure 4.39 compares the number of additions in both algorithms when the channel noise and code are varying. While the additions in both algorithms depict the same dependency on SNR, the JVDD requires lesser additions when N is small (regardless of SNR and R). At larger N however, JVDD requires lesser additions only when R is high. The sensitivity of JVDD to the code and channel noise is directly attributed to the S_{avg} in (4.38), as explained earlier. On this end, the JVDD is ideal for applications that operate at short N or long N but low coding overhead.

iii) Memory

The memory consumption of JVDD is largely dependent on S_{avg} and its associated path metric and state transitions. A small amount of memory is used for storing the noise-free metrics across the states. In general, the memory usage of the algorithm can be written as (4.42). In (4.42), the JVDD algorithm has a dynamic memory consumption due to the variation of S_{avg} with SNR, N and R . On the contrary, the Iterative system uses a fixed amount of memory (for a

given N and R) to store the metrics in the BCJR trellis ((2.6) - (2.8)) and messages in the LDPC graph (2.7) - (2.9). Its overall memory consumption can be expressed as (4.43).

$$JVDD_{mem} = \text{size of}(3 S_{avg} + 2^{N_h-1} N_b) \quad (4.42)$$

$$Iter_{mem} = \text{size of}(2^{N_h-1} N_{trellis}(2 + N_b) + N_{edge}(N + M)) \quad (4.43)$$

From (4.42) and (4.43), it can be inferred that the channel SNR and R which have a significant impact on S_{avg} (as seen in Figure 4.38), will dynamically affect the memory of the JVDD algorithm but not the Iterative. In other words, the JVDD algorithm is more memory intensive than the Iterative, especially at low SNR and large N . This could translate into several hardware constraints such as the requirement for larger device sizes.

iv) Time

Figure 4.40: Computational time of the JVDD and Iterative Detector-Decoder for (a) $N = 512$ and (b) $N = 2048$.

In Figure 4.40, the computation time of the JVDD and the Iterative Detector-Decoder algorithms are compared through simulations. Time comparisons are however only useful for a 1st order complexity analysis as it is highly dependent on the software implementation. Figure 4.40 suggests that JVDD is as time efficient as the Iterative for small N . When N is large, S_{avg} grows dynamically and ultimately runs at 10 times slower than the Iterative system in the worst case.

In summary, the proposed JVDD scheme showed performance gains over the conventional Iterative Detector-Decoder over an ISI channel. The performance gap between the 2 schemes however narrows down as N increases and/or SNR decreases. Complexity analysis suggests that

the JVDD requires lesser number of multiplications and additions, and similar memory consumption and computation time than the Iterative system for applications that operate at small block lengths and over a less-noisy channel. The dynamic behavior of S_{avg} in the JVDD amplifies its complexity when the block size is large and channel noise is excessive. Curbing S_{avg} to a manageable complexity level would result in a performance loss from the trade-off.

4.3.2 JVDD Over a 1D GFP Channel

Having studied the performance of JVDD over an ISI channel, the same study is repeated over a more realistic magnetic recording channel. The complexity of JVDD will not be analyzed in this section as the analysis done in section 4.3.1.2 is applicable for this study too. In this section, a PMR channel employing the GFP channel model (Chapter 2) is used. Simulations were carried out for various R and N values of the code, while holding the media and channel parameters (bit length, BL , write field and reader width, RW) constant. As the channel length used in this study will result in a significantly large number of trellis states for JVDD and BCJR detector in the iterative system, a GPR equalizer is used to characterize a shorter length target to reduce the number of trellis states and therefore complexity (section 2.2.2). The complete set of variable and constant parameters used is represented in Tables 4.9 and 4.10 respectively.

Table 4.9: Variable parameters used in JVDD study over PMR channel

Detector	JVDD, Iterative detector-decoder
N	512, 1024, 2048
R	0.75-0.99 (in steps of 0.01)

Examining the FER plots obtained in Figure 4.41(i), the JVDD better performance than the iterative system over the GFP channel for the simulated parameters. This gain is however less prominent for shorter BL (poorer SNR). For larger BL (better SNR), JVDD displayed a steeper drop in FER than the iterative detector-decoder. For example, in Figure 4.41(i)(b), a gain of about 4% is seen for $BL = 14$ as opposed to a smaller gain of 2% for $BL = 11$ (for a fixed R). The observed trend can be explained as in section 4.3.1 – In a noisier channel, the presence of more

contending paths leads to a larger average number of survivors. This can be supported with a complexity plot depicting the average number of survivors over the trellis at different values of BL as shown in Figure 4.39(ii). The performance gap reduces further as N increase. In Figure 4.39(i)(c), at $BL = 11$, the performance of JVDD and the iterative detector-decoder is identical. If N increases beyond 2048 or if BL is reduced further, the JVDD will eventually perform worse than the state-of-the-art system.

Table 4.10: Constant parameters used in JVDD study over PMR channel

Media	Dimension	700nm x 70 000nm	Equalizer length	N_w	51
	Grain size	7nm	Target length	N_g	5
	Hk_{mean}	1 061 033	Iterative detector/decoder	I_{global}	5
Reader width	RW	30 nm		I_{dpc}	150
Channel bit length	BL	11, 14 nm	JVDD	S_{max}	10000, 30000

Figure 4.41: Performance of JVDD over a GFP channel at various R and $N =$ (a)512 (b) 1024 and (c)2048.

4.3.3 JVDD Over a 2D ISI Channel

The JVDD algorithm studied thus far in sections 4.4.2 and 4.4.3 is only applicable to 1D channels. In section 4.1.2, the JVDD algorithm was formulated in a non-binary form in order to apply the join detection and decoding technique to 2D channels. The resulting algorithm was coined the 2D-JVDD. In this section, the performance and complexity of the 2D-JVDD is evaluated over a 2D ISI channel through extensive computer simulations. A state-of-the-art 2D iterative detector-decoder is used as a benchmark in this study and is implemented using a 2D BCJR detector and multiple 1D LDPC decoders. A block diagram that illustrates the architecture of the 2D iterative detector-decoder and 2D-JVDD used is shown in Figure 4.42. Given these architectures, conversions from symbol LLR to bit LLR and vice-versa have to be performed prior to decoding and after decoding for both systems. Coded simulations were carried out over various channel SNRs while keeping the code properties unchanged. Tables 4.11 and 4.12 list the constant and variable parameters used.

Figure 4.42: Block diagram depicting comparison between 2D-JVDD and 2D iterative detector-decoder.

Table 4.11: Constant parameters used in 2D JVDD study

Code	R	0.85	Iterative detector- decoder	I_{global}	5
Channel	h_k	[0.05 0.5 0.05; 0.1 1 0.1; 0.05 0.5 0.05]		I_{dec}	150
JVDD code	$N=64$	$d_x=2, d_y=20$	JVDD	S_{max}	5000
	$N=128$	$d_x=8, d_y=38$			10000
LDPC code	w_{col}	4		T	1.5

Table 4.12: Variable parameters used in 2D JVDD study

Detector	JVDD, Iterative detector-decoder
N	64, 128
SNR	6-10 dB (in steps of 1)

The performance plots of 2D-JVDD over the said channel are shown in Figure 4.43. In the figure, the 2D-JVDD outperforms the 2D iterative detector-decoder at $N = 64$ and 128 by 2 dB and 1 dB respectively in the best case. At a fixed SNR value, the JVDD displayed a performance gain of at least 10 times. This gain however drops by about 60% when N is doubled. For shorter codes ($N = 64$), a small τ and S_{max} value of 1.5 and 5000 respectively are sufficient for the 2D-JVDD to perform well. At $N = 128$ however, a larger τ and S_{max} are required before it can outperform the 2D iterative system. Analyzing the corresponding complexity plots in Figure 4.44, the number of survivors in the 2D JVDD trellis accumulates to a large value for the case of $N = 64$ and even more for $N = 128$. In Figure 4.44(ii)(b) for example, the total number of survivors for $N = 128$ is so large that they start to get cropped by the S_{max} ceiling for a large part of the trellis. This implies that for $N > 128$, a significantly larger number of survivors will be cropped by S_{max} until eventually the 2D-JVDD suffers a performance loss as the chances of the MMLC being discarded increases.

Figure 4.43: Performance of 2D-JVDD over a 2D ISI at various SNR and $N =$ (a) 64 and (b) 128.

In view of the outcome of this study, it can be concluded that the 2D-JVDD algorithm is computationally complex given that there are a large number of states and branches in a 2D

channel trellis and therefore a large average number of survivors. To this end, the 2D-JVDD is more practical for applications that require short operating codeword length such as in communication channels. Before it can be practically applied to 2D magnetic recording channels which require long codeword lengths, new measures should be introduced to drastically reduce the number of survivors. Future work on the 2D-JVDD should center on modifications to the existing 2D-JVDD algorithm as well as through new code designs. This will be discussed further in the future work section of Chapter 7.

Figure 4.44: Complexity of 2D-JVDD over a 2D ISI measured in terms of (i) average number of survivors and (ii) total number of survivors for $N =$ (a) 64 and (b) 128.

4.4 Summary

The iterative detector-decoder constitutes the state-of-the-art detection and decoding technology used in the magnetic recording industry today. In order to achieve a significant

breakthrough in storage capacity of HDD, advancements in signal processing techniques have to be made. The sub-optimality of the existing iterative detector-decoder can be tackled to achieve this. One of the causes of the sub-optimality of the iterative detector-decoder lies in the iterative nature of the detection and decoding algorithms. This problem can be eradicated if the inherently dependent task of detection and decoding are carried out simultaneously instead. This chapter therefore proposes a novel joint detection and decoding scheme that has been coined the JVDD.

In this chapter, the JVDD is presented as an alternative to the state-of-the-art iterative detector-decoder. Unlike the iterative detector-decoder that performs detection over a channel trellis and decoding over a code factor graph separately and iteratively, the JVDD carries out the task of detection and decoding simultaneously in a single step, over a channel trellis. This simultaneous execution of these tasks are carried out through two key steps in the JVDD algorithm - ‘metric thresholding’ which searches for a path through the trellis that is closest to the transmitted sequence in the Euclidean distance sense and ‘parity checking’ which validates if paths belong to a legal codeword. The proposed JVDD algorithm is only applicable to 1D coded ISI channels. To extend its application to 2D channels, a non-binary adaptation of the algorithm was proposed in section 4.1.2. This was coined the 2D-JVDD.

The JVDD/2D-JVDD suffers a drawback in terms of high computational complexity (measured in terms of the number of survivors) of the algorithm. An exponential accumulation of survivor paths is often seen in the JVDD trellis, even at small N values of 512. To manage the complexity of JVDD, an S_{max} parameter was proposed for the algorithm to limit the maximum number of survivors allowed in the trellis. The S_{max} however gives rise to the need for a trade-off between performance and complexity of the JVDD. To curb the rapid growth of survivors in the trellis more effectively, a new class of linear block codes specific for the JVDD was proposed in section 4.2. This has been coined the JVDD class of codes and it consists of GDLD and proportionate row weight codes. The JVDD displayed a significant reduction in complexity when used with these codes. For example, complexity savings of about 85% was seen at $N = 1024$ and $R = 0.90$. The JVDD codes are defined by a combination of design parameters, each of which impacts the

performance and complexity of the JVDD in varying ways. This was thoroughly investigated in Section 4.2 through computer simulations and theoretical analysis of the minimum distance and column weight distribution of the codes. The proposed codes were shown to have an optimal combination of design parameters that minimizes the FER and complexity of the JVDD. To identify the optimum codes, an analytical optimization technique that involves the RSM was adopted and optimization of the codes at various N and R values and channel conditions was carried out.

With the optimum codes found, the performance of JVDD was thoroughly evaluated against the state-of-the-art iterative detector-decoder over varying channel types and conditions, N and R values. Three different types of channels were particularly studied – 1) a 1D ISI channel 2) a 1D magnetic recording channel that employs the GFP channel model presented in Chapter 2 3) a 2D ISI channel. The outcome of the studies indicated a general performance gain of the JVDD over the conventional iterative detector-decoder. For 1D ISI and GFP channels, a performance gain of up to 20 % was displayed by the JVDD for a fixed channel SNR or BL . The performance gap between the two detection and decoding schemes however gets narrower as N becomes larger. For example, a shift from $N = 512$ to 2048 causes the performance gap to decrease from 10 % to 5 % and complexity to increase by 20% over an ISI channel with SNR = 7 dB and $R = 0.95$. This was attributed to the presence of a longer trellis in the JVDD algorithm and thus greater accumulation of survivors. In addition to that, the LDPC decoder of the iterative system has been known to perform better at larger N . The study also revealed a poorer performance displayed by the JVDD at lower SNR. For GFP channels for example, a smaller performance gap of 2 % was seen at $BL = 11$ nm as compared to the 4 % gap seen at $BL = 14$ nm for $N = 1024$ and $R = 0.95$. The effect of both low SNR and large N proved to be unfavorable for the JVDD. At $N = 2048$ and $BL = 11$ nm for example, the JVDD's performance was identical to the iterative detector-decoder. Increasing N or lowering the SNR beyond this will cause the JVDD to lose out in performance over the conventional iterative detector-decoder. The complexity of the proposed JVDD is measured in terms of the number of survivors in its trellis. The complexity of the iterative detector-decoder cannot however be measured this way. As such, the computation time of both detection schemes were compared to give a first order estimate on

the complexity of both algorithms. The JVDD showed an identical computational time as the iterative detector-decoder at intermediate N values. It however operates up to 10 times longer than the iterative detector-decoder at large N values. A larger gap is seen when the channel SNR is bad. The difference in computation time however decreases with decreasing R . This is attributed to a higher number of parity checks and therefore a more frequent elimination of paths in the JVDD trellis.

For 2D channels, the 2D-JVDD displayed a good performance with reasonable complexity only at small N values. At $N = 64$ and $R = 0.90$ for example, a performance improvement of 10 times was reaped by the 2D-JVDD over the 2D iterative detector-decoder at SNR = 7 dB and a 2 dB gain was seen at a fixed FER value. As N approaches 128, the number of survivor paths in the 2D channel trellis grew significantly and as a result was cropped out by the S_{max} limit of the JVDD. A poor performance results. The high computational complexity of the 2D-JVDD hinders it from successfully operating at intermediate or large N values. As such, modifications to the 2D-JVDD algorithm or introduction of novel code designs for the 2D-JVDD are necessary before it can be practically applied to magnetic recording channels.

Chapter 5

Joint Factor Graph Detector Decoder

The sub-optimality of the state-of-the-art iterative detector-decoder has sparked a recent interest in the area of joint detection and decoding with the aim of achieving better performance gains for magnetic recording channels and communication channels alike [23][24][25]. Enabling joint detection and decoding involves the integration of the inherently dependent task of detection and decoding into a single process that can be carried out simultaneously.

A detector plays the role of predicting the most likely transmitted sequence by undoing the effects of Inter Symbol Interference (ISI) introduced by a channel. A decoder, on the other hand, aims to recover encoder bits from the detected stream using Error Correction Codes (ECC). To do so, the detector requires information of the channel while the decoder needs information of the code. In the state-of-the-art iterative detector-decoder, its detector component operates on a trellis structure which is derived based on the channel. The trellis is therefore known as the channel trellis. Its decoder component, on the other hand, operates on a factor graph structure that is derived based on the code. The factor graph is thus known as the code factor graph. This is depicted in Figure 5.1. Due to the different structure upon which each of its detection and decoding algorithm executes on, the detector and decoder in the existing iterative system constitute two separate entities.

Combining the detector and decoder into a single entity that is capable of simultaneously undoing the effect of the channel and the code can be achieved in various ways. This thesis proposes two methods of doing so. Firstly, the detector and decoder blocks can be integrated into a single component which operates on a channel trellis, as illustrated in Figure 5.1. Information about the code can be integrated into the channel trellis through nodes known as parity check nodes. The Joint Viterbi Detector Decoder (JVDD) achieves this and was proposed in Chapter 4. Secondly, a combined detector and decoder block can be made to operate on a super factor graph structure that is derived from the combination of a channel factor graph with a code factor graph (Figure 5.1). One method of achieving this is through the Joint Factor Graph Detector Decoder (JFGDD). This chapter proposes the JFGDD as a second alternative to the state-of-the-art iterative detector-decoder.

Figure 5.1: Block diagram depicting difference between conventional iterative detector-decoder, JVDD and JFGDD.

5.1 Algorithm

The proposed JFGDD requires knowledge about the channel as well as the code in order to perform joint detection and decoding. The factor graph upon which the JFGDD operates on should therefore be derived from both the channel and the code. As the state-of-the-art iterative detector decoder consists of the Bahl-Cocke-Jelinek-Raviv (BCJR) detector [59] that

executes on a channel trellis and a Low-Density Parity Check (LDPC) decoder [54] that operates on a code factor graph, the JFGDD can be constructed by retaining the code factor graph and converting the channel trellis into a channel factor graph. The channel factor graph can subsequently be combined with the code factor graph to obtain a super factor graph structure that represents both the channel and the code. This is illustrated in Figure 5.2.

Figure 5.2: Conversion of a channel trellis into a channel factor graph and subsequent integration of the channel factor graph with a code factor graph to develop a super factor graph structure required by a JFGDD.

In the state-of-the-art iterative detector-decoder, its LDPC decoder executes the Sum-Product Algorithm (SPA) on its code factor graph as described in Chapter 2. As the JFGDD retains the structure of the LDPC decoder, enabling joint detection and decoding requires the implementation of the SPA on its channel factor graph. The channel factor graph of the detector can subsequently be integrated with the code factor graph of the decoder to arrive at

a super factor graph for the JFGDD. The SPA can then be executed concurrently in the super factor graph, thereby achieving a joint detection and decoding process.

This section of the thesis first presents a proposed implementation of the SPA on a channel factor graph for an isolated detector block coined the factor graph based detector. The proposed integration of the channel and code factor graphs and the execution of the SPA on the super factor graph block are subsequently presented.

5.1.1 Factor Graph-Based Detector

This section proposes the implementation of the SPA over a channel factor graph in a factor graph-based detector. The factor-graph based detection algorithm is formulated through a factorization of a global function that represents a conditional probability density function (PDF) used in making an estimate of transmitted sequences. The detection algorithm is as follows:

Given a coded sequence $\mathbf{x}$ and a channel with coefficients h_k and length N_h , the readback sequence $\mathbf{z}$ can be expressed as

$$z_k = \sum_{l=0}^{N_h-1} h_l x_{k-l} + w_k \quad (5.1)$$

where $\mathbf{w}$ is a noise vector. For the simplest system, the noise vector is usually modeled as Additive White Gaussian Noise (AWGN) with zero mean and variance σ_{noise}^2 . The joint A Posteriori Probability (APP) of the written sequence, $P(\mathbf{x}|\mathbf{z})$, can be expressed using Bayes' Theorem as

$$\begin{aligned} P(\mathbf{x}|\mathbf{z}) &= \frac{p(\mathbf{z}|\mathbf{x})P(\mathbf{x})}{P(\mathbf{z})} \\ &\propto p(\mathbf{z}|\mathbf{x})P(\mathbf{x}) \end{aligned} \quad (5.2)$$

By applying the chain rule and assuming that $\mathbf{z}$ conditioned on $\mathbf{x}$ are mutually independent, $P(\mathbf{x}|\mathbf{z})$ can be written as

$$\begin{aligned} P(\mathbf{x}|\mathbf{z}) &\propto P(\mathbf{x}) \prod_i p(z_i|x_i, \dots, x_{i-N_h+1}) \\ &\propto P(\mathbf{x}) \prod_i g_i(x_i, \dots, x_{i-N_h+1}) \end{aligned} \quad (5.3)$$

where $P(\mathbf{x})$ is the a priori distribution of the coded sequence, $p(\mathbf{z}|\mathbf{x})$ is the conditional PDF for $\mathbf{z}$ when $\mathbf{x}$ was transmitted, and $g_i(x_i, \dots, x_{i-N_h+1})$ represents local functions that make up the factors of the global function in equation (5.3). These local functions can be expressed as

$$g_i(x_i, \dots, x_{i-N_h+1}) = p(z_i|x_i, \dots, x_{i-N_h+1}) \quad (5.4)$$

For a channel with AWGN noise, equation (5.4) can be written as

$$\begin{aligned} g_i(x_i, \dots, x_{i-N_h+1}) &= \frac{1}{\sigma_{noise}\sqrt{\pi}} \exp\left\{-\frac{|z_i - \sum_{l=0}^{N_h-1} h_l x_{i-l}|^2}{2\sigma_{noise}^2}\right\} \\ &\propto \exp\left\{-\frac{|z_i - \sum_{l=0}^{N_h-1} h_l x_{i-l}|^2}{2\sigma_{noise}^2}\right\} \end{aligned} \quad (5.5)$$

The structure of the factorization of the global function into several local functions g_i can be expressed using a factor graph. In Figure 5.3, a channel factor graph that corresponds to the global function in equation (5.2) for an ISI channel with $N_h = 4$ and non-zero channel coefficients is shown. Functions g_i in the equations above are represented as channel nodes, g , on a factor graph, as illustrated by the square nodes of Figure 5.3. Variable nodes, x , are represented by circles and corresponds to the channel input sequence. An edge connects a variable node to a channel node if and only if the variable is an argument of that function. Short cycles of length-4 commonly occur in factor graphs, as illustrated by the red loop in the figure below, and are known to give rise to performance degradation. This will be addressed in greater detail in section 5.2.

Figure 5.3: A channel factor graph consisting of channel and variable nodes.

To implement the MAP symbol detection over a factor graph, marginal APPs in the form of $P(x_n|\mathbf{z})$ needs to be computed. This can be done by applying the sum product rule to the factor graph. The SPA operates based on the following rule:

The Sum-Product Update Rule

The message sent from a node v on an edge e is the product of the local function at v with all messages received at v on edges other than e , summed up for the variable associated with e [79].

Using this rule, the SPA computes two messages for each edge in the channel factor graph of the detector. One message is sent from a variable node to a channel node and the second message is sent from the channel node to the variable node. A message sent from a variable node, x_i , to a channel node, g_j , is denoted as u_{ij} and can be computed as follow:

$$u_{ij} = \prod_{j' \in N(i) \setminus j} v_{j'i} \quad (5.6)$$

In the equation above, $N(i)$ represents the set of nodes that are connected by an edge to variable node x_i . Here, the message u_{ij} sent is derived by gathering messages received from the variable node's neighbors in the previous iteration step. This however excludes the message received from channel node g_j , as illustrated in Figure 5.4(a).

Denoting the message sent from a channel node to a variable node as v_{ji} , its computation can be expressed mathematically as

$$v_{ji} = \sum_{\sim i} \left[g(I) \prod_{i' \in N(j) \setminus i} u_{i'j} \right] \quad (5.7)$$

where $N(j)$ represents the set of nodes that are connected by an edge to channel node g_j , I is the set of variables that make up the argument of function g and $\sum_{\sim i}$ is a summation operator representing a summation over all variables of $g(I)$ excluding i . As depicted in Figure 5.4(b), the message sent from channel node g_j to variable node x_i comprises of messages gathered from the neighbors of j , excluding that received by node i in the previous iteration step. In other words, only extrinsic information is exchanged between the channel and variable nodes of the channel factor graph.

Figure 5.4: Message passing from (a) variable node to channel node and from (b) channel node to variable node in a channel factor graph.

The multiplications involved in the equations above would be costly to implement in hardware. Converting the multiplications to additions through the use of logarithms is therefore more favorable. Computation of the messages in logarithmic domain can be represented as follows:

$$\overline{u_{ij}} = \ln u_{ij} = \sum_{j' \in N(i) \setminus j} v_{ji'} \quad (5.8)$$

$$\overline{v_{ji}} = \ln v_{ji} = \ln \left\{ \sum_{\sim i} \exp \left[\ln g(I) + \sum_{i' \in N(j) \setminus i} \overline{u_{ji'}} \right] \right\} \quad (5.9)$$

Message passing between the channel and variable nodes are repeated for a pre-defined number of detector iterations, I_{det} . The performance of the factor graph detector improves with increasing I_{det} and eventually reaches a saturation point, as can be seen in the BER plot of Figure 5.5.

Figure 5.5: Varying number of iterations in a binary factor graph detector.

With the implementation of the SPA over a channel factor graph, a factor-graph based detector was formulated. To enable joint detection and decoding, information on the channel and the code has to be represented on a single structure. As such, an integration of the channel factor graph of the factor-graph based detector with the code factor graph of the LDPC decoder is required. The SPA can then be applied on the resulting super factor graph to enable a joint detection and decoding operation of the JFGDD. The next section proposes the joint detection and decoding algorithm for the JFGDD.

5.1.2 Joint Factor Graph Detector Decoder

Integrating the detector and decoder in a JFGDD system involves integrating the variable nodes from the channel factor graph and the code factor graph to obtain a super factor graph structure as shown in Figure 5.6. A JFGDD is therefore made up of three types of nodes – n channel nodes, n variable nodes and $(n - k)$ check nodes. With this structure, JFGDD now has information of both the code as well as the channel.

Figure 5.6: A super factor graph structure consisting of channel, variable and check nodes.

At each iteration step of the message passing, a variable node x_i sends a message u_{ij} to channel node g_j based on the messages it receives in the previous iteration step from the channel nodes as well as check nodes that are connected to its edges. In the computation of the message u_{ij} however, variable node x_i ignores the message v_{ji} that was sent to it by channel node g_j in the previous step. This is illustrated in Figure 5.7(a). Mathematically, this can be expressed as

$$u_{ij} = \sum_{j' \in N_g(i) \setminus j} v_{ji'} + \sum_{k \in N_c(i)} r_{ki} \quad (5.10)$$

where $N_g(i)$ and $N_c(i)$ denotes the set of channel nodes and check nodes that are neighbors of variable node x_i respectively. Likewise, when the variable node sends a message q_{ik} to a check node, it ignores the message r_{ki} that was sent to it by the check node, as depicted in Figure 5.7(b). The message sent can be formulated as

$$u_{ik} = \sum_{j \in N_g(i)} v_{ji} + \sum_{k' \in N_c(i) \setminus k} r_{k'i} \quad (5.11)$$

At the end of every iteration step, the LLR is given by the summation of messages received at the variable nodes. Message passing is carried out on the super factor graph of the JFGDD until the parity check equations are satisfied or a pre-defined number of iterations has been executed.

Figure 5.7: Message passing in JFGDD from (a) variable node to channel node and from (b) variable node to check node in the JFGDD.

Performing detection and decoding jointly in this manner avoids any increase in complexity that is associated with the separate task of detection and decoding currently used in the state-of-the-art iterative detector-decoder. It also eliminates the sub-optimality associated with the iterative process of message passing between the detector and decoder. The performance of JFGDD however, is susceptible to the presence of short cycles (length-4 and length-6). Length-4 cycle for instance is present in the channel factor graph of the detector as highlighted in Figure 5.3. It is commonly present in the code factor graph of the decoder as well as explained in section 2.3.1. The existence of such short cycles in factor graphs has been widely known in the literature to significantly degrade the performance of the system [71][79]. For the factor graph-based detector for example, the computation of messages was carried out based on the condition that the extrinsic information exchanged between successive iterations should be independent. The presence of short cycles renders this condition as invalid and as a result the computation of marginal APPs represents approximated values instead of exact ones.

Factor graphs should therefore be inspected for short cycles and measures to reduce its occurrence should be put in place. In general, the number of length-4 cycles in any factor graph can be computed mathematically, similar to the existing technique commonly used in a SPA decoder. To check for short cycles in a channel factor graph, the channel can first be expressed as a matrix, $\mathbf{H}_{\text{channel}}$, such that

$$\mathbf{y} = \mathbf{c} \cdot \mathbf{H}_{\text{channel}} + \mathbf{w} \quad (5.12)$$

where $\mathbf{c}$ and $\mathbf{y}$ are the channel input and output vectors respectively and $\mathbf{w}$ is the noise vector associated with the channel. In the simplest system, $\mathbf{w}$ is modeled as AWGN with mean 0 and variance σ_{noise}^2 .

4-cycles are formed when bits of '1' in matrix $\mathbf{H}_{\text{channel}}$ form a closed loop on two consecutive rows and column as shown in Figure 5.8. This can be inspected by summing up the bits in every two consecutive rows in the matrix as illustrated in the figure. 4-cycles are present if consecutive '2' are obtained in any of the summation. Subsequently, computing the exact

number of length-4 cycles involves finding the summation of number of 4-cycles in all possible combinations of two rows of $\mathbf{H}_{\text{channel}}$.

Figure 5.8: Detecting presence of length-4 cycles in a channel matrix.

Length-6 cycles on the other hand occur when there are six '1' bits in $\mathbf{H}_{\text{channel}}$ that occupy three consecutive rows and columns, as depicted in Figure 5.9. In general, length-6 cycle exists in a factor graph if the differences between the indexes of the non-zero channel coefficients, h_k , are all different.

Figure 5.9: Detecting presence of length-6 cycles in a channel matrix.

To prevent significant performance deterioration of JFGDD, short cycles should be mitigated in the channel factor graph as well as the code factor graph. As length-4 cycles are known to degrade performance of a system more severely than length-6 cycles, this thesis focuses on the elimination of length-4 cycles. Mitigating length-4 cycles in a code factor graph can be carried out using existing ECC technique such as that implemented by Neal Radford [79]. The Neal Radford's method is adopted in this thesis for reducing short cycles in the code factor graph of the JFGDD. To reduce the occurrence of length-4 cycles in the channel factor graph, this thesis proposes two novel methods of doing so. This is presented in the next section.

5.2 Mitigating Cycles in a Channel Factor Graph

Two methods are proposed to mitigate or completely remove length-4 cycles in the channel factor graph of the factor-graph based detector. The first method involves the use of a Generalized Partial Response (GPR) equalizer to design a constrained target which will result in a channel factor graph with zero or reduced number of short cycles. The second method involves in a non-binary implementation of the detection algorithm where consecutive nodes in the channel factor graph are grouped together, resulting in fewer edges and therefore cycles in the channel factor graph. The proposed methods are presented in greater detail in this section.

5.2.1 Constraining Generalized Partial Response Target

In a channel factor graph, the edges that connect the channel and variable nodes of the graph are constructed based on the position of the non-zero coefficients of the channel in which it represents (as presented in section 5.1.1). Designing a channel factor graph with reduced number of length-4 cycles along its edges can therefore be obtained by tweaking these coefficients and their positions. This can be achieved through the use of a conventional GPR equalizer (section 2.3.2). A GPR equalizer can be employed to reshape the channel response into a target response that will result in a channel factor graph with fewer short cycles or completely free of short cycles. Such targets can be achieved by establishing specific constraints during the design process of the GPR equalizer.

As presented in section 2.3.2, a GPR equalizer and GPR target are jointly designed by minimizing the mean squared error (MSE) of the error signal between the equalizer output and target output. This can be expressed mathematically as

$$E[e_k^2] = \mathbf{g}^T \mathbf{A} \mathbf{g} - 2\mathbf{g}^T \mathbf{P} \mathbf{w} + \mathbf{w}^T \mathbf{R} \mathbf{w} \quad (5.13)$$

where e_k refers to the error signal, $\mathbf{A}$ is the auto-correlation matrix of x_k , $\mathbf{R}$ is the auto-correlation matrix of r_k , $\mathbf{P}$ is the cross-correlation matrix between x_k and r_k , $\mathbf{w}$ is the equalizer response and $\mathbf{g}$ is the target response.

In the design process, constraints can be imposed on the coefficients of $\mathbf{g}$. For example, when designing a GPR target of length, $N_g = 6$, a constraint can be imposed such that the fourth coefficient of the target is restricted to be 0. In this case, the following constraint is specified:

$$\mathbf{E}^T \mathbf{g} = \mathbf{i} \quad (5.14)$$

where

$$\mathbf{E}^T = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

$$\mathbf{g} = [g_0 \quad g_1 \quad g_2 \quad g_3 \quad g_4 \quad g_5]^T$$

$$\mathbf{i} = [1 \quad 0]^T$$

The cost function can then be minimized according to the description provided in section 2.3.2.

The channel factor graph is subsequently designed based on the GPR target designed.

Figure 5.10: Channel factor graph free of length-4 cycles as a result of constraining the third and fifth coefficient of its designed target to 0.

Two examples can be used to illustrate the effect of constraining designed targets on the channel factor graph. Firstly, given a GPR target with $N_g = 6$, constraining its third and fifth coefficient to 0 (ie. $g_2 = 0$ and $g_4 = 0$) will result in a channel factor graph free of length-4 cycles as shown in Figure 5.10. Similarly, constraining the second and fourth coefficient of the

target to 0 (ie. $g_1 = 0$ and $g_3 = 0$) instead gives rise to a channel without any 4-cycles present (Figure 5.11). The constraints imposed have resulted in the elimination of length-4 cycles in both examples. Despite its effectiveness in mitigating or completely removing length-4 cycles in the channel factor graph, the use of a GPR equalizer will, in practice, result in coloration of noise. As such, a trade-off between the number of short cycles removed and the extent of noise coloration is necessary.

Figure 5.11: Channel factor graph free of length-4 cycles as a result of constraining the second and fourth coefficient of its designed target to 0.

5.2.2 Non-Binary Factor Graph-Based Detector

An alternative method of reducing the number of length-4 cycles in the channel factor graph of the JFGDD involves an implementation of a non-binary factor-graph based detector. This section presents the proposed method.

The factor graph-based detector proposed in section 5.1.1 computes and pass messages that represent bit probabilities within its nodes. As such, this factor-graph based detector can be classified as a binary factor graph-based detector. The binary factor graph-based detector can be implemented in a non-binary form where messages are represented in the form of symbol probabilities instead of bit probabilities. This is coined the non-binary factor graph-based detector. In this non-binary detector, symbols are defined by grouping M consecutive coded

bits or read-back samples. These symbols can be represented as vectors $\mathbf{X}_n$ and $\mathbf{Z}_n$ respectively where

$$\mathbf{X}_n = [x_{M \cdot n}, x_{M \cdot n + 1}, \dots, x_{M \cdot (n+1) - 1}]^T \quad (5.15)$$

$$\mathbf{Z}_n = [z_{M \cdot n}, z_{M \cdot n + 1}, \dots, z_{M \cdot (n+1) - 1}]^T \quad (5.16)$$

Each symbol can take on 2^M values denoted by $\hat{x}^{(0)}, \hat{x}^{(1)}, \dots, \hat{x}^{(2^M-1)}$. The messages passed along the edges of a non-binary factor graph-based detector are thus not representative of individual bits but rather each possible combination of variables listed above. As such, in the non-binary implementation, 2^M messages are passed between the channel and variable nodes (in each direction) of the channel factor graph at each iteration step.

Figure 5.12: Channel factor graph of a non-binary factor graph-based detector designed with $M = 2$.

The proposed grouping of consecutive bit sequences will lead to a significant reduction in the number of short cycles of the channel factor graph. For $M = 2$ for example, the resulting non-binary factor graph for a channel with memory, $L_{mem} = 3$ and $h_k \neq 0$ is shown in Figure 5.12. Each variable node now represents the vector $\mathbf{X}_i$ instead of the bit x_i and each channel node represents the grouping of several factors g_j . In Figure 5.12, length-4 cycles are still present in the non-binary factor graph but occur less frequently than in a binary factor graph of the same channel shown in Figure 5.3.

Figure 5.13: Channel factor graph of a non-binary factor graph-based detector designed with $M = 3$.

A non-binary factor graph-based detector for the same channel is represented in Figure 5.13 when $M = 3$. The absence of closed loops in this channel factor graph is apparent. In general, a completely cycle-free factor graph can be obtained when $M = L_{mem}$.

As the variable nodes in a channel factor graph of a non-binary factor graph-based detector represent symbols instead of bits, the messages passed between the nodes of the graph will be computed differently from that of a binary factor graph-based detector. The messages, $\mathbf{V}_{ji}$, passed from a channel node g_j to a variable node $\mathbf{X}_i$ can be expressed as the following vector:

$$\mathbf{V}_{ji} = [\hat{v}_{ji}^{(0)}, \hat{v}_{ji}^{(1)}, \dots, \hat{v}_{ji}^{(2^M-1)}]^T \quad (5.17)$$

where each vector element represents

$$\hat{v}_{ji}^{(l)} = P(\mathbf{X}_i = \hat{x}^{(l)} | \mathbf{Y}_j) \quad (5.18)$$

The computation of messages previously presented in the binary factor graph-based detector can be re-formulated for its cycle-free non-binary counterpart. In equation (5.6), the summation is applied to all possible combinations of symbol vectors for the non-binary factor graph detector and is not limited to binary values. The message vector, $\mathbf{U}_{ij}$, passed from a variable node $\mathbf{X}_i$ to a channel node g_j can be expressed similarly as follow:

$$\mathbf{U}_{ij} = [\dot{u}_{ij}^{(0)}, \dot{u}_{ij}^{(1)}, \dots, \dot{u}_{ij}^{(2^M-1)}]^T \quad (5.19)$$

where each vector element represents a-priori probabilities used in subsequent iteration step.

This can be written as

$$\dot{u}_{ij}^{(l)} = P(\mathbf{X}_i = \mathbf{x}^{(l)}) \quad (5.20)$$

For a cycle-free non-binary factor graph as in Figure 5.11, forwarding of messages can be done simply as follow:

$$\mathbf{U}_{(i-1)i} = \mathbf{V}_{(i-1)(i-1)} \quad (5.21)$$

$$\mathbf{U}_{ii} = \mathbf{V}_{(i+1)i} \quad (5.22)$$

5.3 Simulations and Results

Extensive computer simulations were carried out to evaluate the performance of the JFGDD detector against the state-of-the-art iterative detector decoder. Three studies were carried out. Firstly, the binary and non-binary factor graph-based detectors were each isolated from the decoder component and their performances were benchmarked against the BCJR detector that is used in the state-of-the-art iterative system. This study was carried out over an uncoded one-dimensional (1D) ISI channel. In the second study, the factor graph-based detectors were each integrated with an LDPC decoder to form a binary JFGDD and a non-binary JFGDD. The error rates obtained for both systems were compared to that of the iterative detector-decoder over a coded 1D ISI channel. The third study was carried out over a magnetic recording channel using the Grain Flipping Probability (GFP) channel model presented in Chapter 3. In this study, the GPR equalizer is used to design constrained targets for the binary JFGDD. The performance of the resulting binary JFGDD was compared to that of the non-binary JFGDD and the state-of-the-art iterative detector-decoder. The results obtained from the three studies are presented in this section.

5.3.1 Factor Graph-Based Detector Over Uncoded 1D ISI Channel

The performance of the binary and non-binary factor graph detectors was first studied over an uncoded 1D ISI channel. In the study, the channel noise and number of detector iterations, I_{det} , were varied for the binary and non-binary factor graph-based detectors and their performances were compared to that of the BCJR detector used in the state-of-the-art iterative system (section 2.3.1). The constant and variable parameters used are shown in Tables 5.1 and 5.2 respectively.

Table 5.1: Variable parameters used in study on Factor Graph-based detector over uncoded 1D ISI channel

detector	binary factor graph-based detector, non-binary factor graph-based detector, BCJR
SNR (dB)	4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14
I_{det}	2, 4, 6, 8, 10

Table 5.2: Constant parameters used in study on Factor Graph-based detector over uncoded 1D ISI channel

h_k	[0.6319, 0.7098, 0.3066, 0.0, -0.0499, -0.0186]
data length	4096

Figure 5.14 presents a Bit Error Rate (BER) against SNR plot which shows the relative performance of the three detector systems. Performance of the binary factor graph-based detector is seen to improve greatly when I_{det} is initially increased. For example, a reduction in BER of at least 10 times occurs when I_{det} increases from 2 to 6 for SNR value of at least 8 dB. The performance of this detector however saturates at $I_{det} = 6$, with no further performance gain seen by increasing I_{det} further. At its saturation point, the binary factor graph-based detector suffers a performance loss of about 1 dB in comparison to the BCJR detector. In contrast, the non-binary factor graph-based detector performs much better than its binary counterpart. A small number of iterations ($I_{det} = 2$) is sufficient for the detector to perform as well as the BCJR detector. The performance of the non-binary factor graph-based detector however saturates at this value. Further increase in I_{det} does not result in any further performance improvement, thereby preventing the non-binary detector to outperform the BCJR detector.

The poor performance displayed by the binary factor graph detector can be attributed to the presence of short cycles in its factor graph. The existence of such cycles is known to result in degradation of system performance, as discussed in section 5.2. On the other hand, in a non-binary factor graph detector, the number of length-4 cycles has been greatly reduced. This thus explains the performance gain seen in Figure 5.14, over the binary detector. That aside, the fact that more short cycles exist in a binary factor graph detector than in a non-binary one suggest that more iterations are required before saturation in performance can be attained. This trend can be identified in Figure 5.14.

Figure 5.14: Comparison of performance among various detectors over an uncoded channel.

Table 5.3: Comparison of BCJR and Factor Graph detector complexity

Detector	Number of multiplications	
	In general	For $L_{mem} = 5$
BCJR	$2^{L_{mem}+1}$	2^6
Binary	$I_{det}(L_{mem} + 1)(L_{mem} * 2^{L_{mem}+1} - 1)$	$36(5 * 2^6 - 1)$ (for $I_{det} = 6$)
Non-binary	$I_{det} * 2^{L_{mem}+2}$	2^8 (for $I_{det} = 2$)

A comparison of complexity of each of the detectors in terms of number of multiplications is shown in Table 5.3. The binary factor-graph based detector has the highest complexity amongst

the three detectors simulated. The complexity of the non-binary detector is slightly higher than that of BCJR. The higher complexity of the binary factor-graph based detector over its non-binary counterpart is attributed to the fact that the presence of much more cycles in a binary factor graph increases the number of messages being passed along the edges of the graph. Although the BCJR has a lower complexity than the factor-graph based detectors, in hardware implementation, the factor graph-based detectors are parallelizable whereas this is not possible for the BCJR detector due to its serial structure.

5.3.2 JFGDD Over Coded 1D ISI Channel

In this study, the binary and non-binary factor-graph based detectors are each integrated with an LDPC decoder to form a binary and non-binary JFGDD respectively. The performance of JFGDD in its binary and non-binary form with respect to the state-of-the-art iterative detector-decoder was investigated over a coded 1D ISI channel. The channel coefficients, code used and decoder iterations, I_{dec} , were kept unchanged throughout the study while the noise level and I_{det} were varied. Tables 5.4 and 5.5 present the constant and variable parameters used.

Table 5.4: Constant parameters used in JFGDD study over ISI channel

channel	h_k	[0.6319, 0.7098, 0.3066, 0.0, -0.0499, -0.0186]
code	type	random LDPC
	R	0.90
	w_{col}	4
decoder	I_{dec}	150

Table 5.5: Variable parameters used in JFGDD study over ISI channel

detector/decoder	Binary JFGDD, Non-binary JFGDD, iterative detector-decoder
SNR (dB)	6-9 (in steps of 0.25)
I_{det}	2, 4, 6, 8, 10

A Frame Error Rate (FER) plot depicting the performance of the systems studied is shown in Figure 5.15. Analyzing each of the FER curves, it can be concluded that the binary JFGDD

performs poorly under the simulated conditions. Even for a large number of detector iterations ($I_{det} = 10$), a performance loss of close to 1 dB over the iterative detector-decoder is displayed. Increasing I_{det} further will only bring about marginal performance improvement, with a cost of higher complexity. The non-binary JFGDD on the other hand, matches up in performance to the iterative detector-decoder. This is achieved with just a small number of detector iterations ($I_{det} = 2$). The non-binary JFGDD is however seen to reach its optimum performance at this value of I_{det} . At this point, it is not able to outperform the state-of-the-art iterative detector-decoder system.

The contrast in error rates obtained for the proposed binary and non-binary JFGDD systems can once again be attributed to the presence of length-4 cycles in the binary JFGDD. Such cycles occur less frequently in the non-binary JFGDD and thereby prevent significant degradation in performance that is encountered by the binary JFGDD. This accounts for the performance gap seen between the two systems in Figure 5.15. The effective removal of short cycles is therefore crucial in enabling the JFGDD system to function well.

Figure 5.15: Comparison of performance among various detector-decoder systems over a coded 1D ISI channel.

5.3.3 JFGDD Over 1D Magnetic Recording Channel

In this study, the binary and non-binary JFGDD were implemented over a magnetic recording channel using a GFP channel model (presented in Chapter 3). The state-of-the-art iterative detector-decoder is used as a performance benchmark for the JFGDD systems. The channel used in this study is long and would translate into a large complexity for the detectors. As such, a GPR equalizer is employed in this study to reshape the channel response into a shorter target response. For the binary JFGDD, the GPR equalizer is jointly designed with constrained targets. This was carried out to ensure a lower presence of length-4 cycles in the channel factor graph as proposed in section 5.2.1. For the non-binary JFGDD, short cycles in its channel factor graph have already been reduced from the grouping of consecutive variable nodes. A monic constraint which has been shown to perform well (section 2.3.1) is therefore used for the GPR equalizer. The monic constraint is also chosen for the GPR equalizer that is used with the iterative detector-decoder.

Table 5.6: Constant parameters used in JFGDD study with constrained targets over ISI channel

channel	N_h	25
equalizer	N_w	51
target	N_g	5
decoder	I_{dec}	150
code	type	LDPC
	R	0.90
	N	4096
	w_{col}	4

Table 5.7: Variable parameters used in JFGDD study with constrained targets over ISI channel

detector/decoder	type	Binary JFGDD, BCJR-LDPC
	I_{det}	2, 4, 6, 8, 10
target	Z_{pos}	1, 2, 3, 4, 12, 13, 24, 23, 34, 123, 124, 134, 234
channel	SNR (dB)	6-9 (in steps of 0.25)

In this study, the channel coefficients and code parameters were kept unchanged. For the binary JFGDD, targets designed using various constraints are studied. The constraints are specified by restricting the coefficients at various indexes of the target to '0'. For example, $Z_{pos} = 12$ indicates that the second and third coefficient of the target is constrained to 0 during the

equalizer design process and the remaining coefficients of the target will be determined by the MSE optimization process. The study was carried out for different sets of constraints and over a channel with varying SNR. Tables 5.6 and 5.7 tabulate the constant and variable parameters used respectively. The number of length-4 cycles present in the channel factor graph as a result of the various constraints imposed on the target during the design of the equalizer is computed and shown in Table 5.8 (in this table, N refers to the codeword length).

Table 5.8: Values of Z_{pos} constraints and the corresponding number of length-4 cycles present in the channel factor graph

Z_{pos}	no. of length-4 cycles present	Z_{pos}	no. of length-4 cycles present	Z_{pos}	no. of length-4 cycles present
1	$N-4$	2	$N-4$	3	$2N-6$
4	$2N-5$	12	0	13	$N-4$
23	0	24	0	34	$N-3$
123	0	124	0	134	0
234	0				

Figure 5.16 presents an FER against SNR plot for the simulations carried out. To avoid cluttering of the figure, curves representing only some of the constraints listed in Table 5.8 are shown. In the figure, the impact of the constraints on the performance of the binary JFGDD varies greatly and is in the following order: $Z_{pos} = 3$, $Z_{pos} = 2$, $Z_{pos} = 134$ and $Z_{pos} = 124$ (from best to worst). The best performing constraint ($Z_{pos} = 3$), however, still suffers from a performance loss (of approximately 1.5 dB) of the binary JFGDD over the state-of-the-art iterative detector-decoder. The introduction of the constrained targets for the channel factor graph has however given rise to a narrower performance gap between the two systems than that observed in section 5.3.2, in the best case.

Comparing the number of 4-cycles computed for each constraint in Table 5.8 with the performance plot obtained in Figure 5.16, the results gathered do not fully support the claim that the performance of binary JFGDD will be enhanced by the reduction of more length-4 cycles. For example, in Table 5.8, $Z_{pos} = 134$ and $Z_{pos} = 124$ results in a factor graph having no

length-4 cycles whereas the graph of $Z_{pos} = 3$ and $Z_{pos} = 2$ has $2N-6$ and $N-4$ length-4 cycles respectively. Their corresponding performance plots in Figure 5.16 however indicates a poorer performance displayed by $Z_{pos} = 134$ and $Z_{pos} = 124$ than $Z_{pos} = 3$ and $Z_{pos} = 2$. The unexpected outcome of the simulations carried out can be explained by analyzing the noise coloration caused by the equalizer in each set-up.

Figure 5.16: Comparison of performance among various detector-decoder.

In Figure 5.17, power spectral density (PSD) plots of the equalizer noise are shown for various constraints. The PSD plots for $Z_{pos} = 134$ and $Z_{pos} = 124$ displayed significant noise coloration whereas that for $Z_{pos} = 3$ and $Z_{pos} = 2$ are relatively flat. Despite giving rise to a factor graph free of length-4 cycles theoretically, the significant noise coloration caused by the equalizer that results from the characterization based on constraints $Z_{pos} = 134$ and $Z_{pos} = 124$ causes the performance to deteriorate greatly. Likewise, for $Z_{pos} = 2$, the less severe coloration of noise shown in Figure 5.17 accounts for a better performance attained in Figure 5.16, in comparison to that for $Z_{pos} = 134$ and $Z_{pos} = 124$, although theoretically $N-4$ length-4 cycles exist when $Z_{pos} = 2$. Similarly for $Z_{pos} = 3$, the flat PSD shown validates the best performing curve of Figure 5.16. In general, the removal of more length-4 cycles are removed by the equalizer, the greater is the coloration of noise.

The detection algorithm for the binary factor-graph based detector proposed in section 5.1.1 was formulated for channels that are corrupted with white noise. The algorithm is bound to

perform poorly in the presence of colored noise. In general, the use of constrained targets will lead to the presence of fewer length-4 cycles in the channel factor graph. The equalizer used however causes a coloration of the white noise present in the channel. The effect of performance loss due to the presence of colored noise could outweigh the performance gain from having lesser length-4 cycles in the channel factor graph. A trade-off between removal of cycles and noise coloration should therefore be carefully selected.

Figure 5.17: PSD plots of noise arising from various constraints.

The performance of the binary JFGDD with the best constrained target found ($Z_{pos} = 3$) is then compared against the non-binary JFGDD and the state-of-the-art iterative detector-decoder. The monic constraint was used in designing the GPR equalizer for the non-binary JFGDD and the iterative detector-decoder.

A performance plot obtained for the comparison on the relative performance of each of the detector-decoder systems over a magnetic recording channel is shown in Figure 5.18. These

plots indicate poorer error rates displayed by the binary JFGDD with constraint $Z_{pos} = 3$ (best constraint found in previous section), in contrast to the error rates displayed by the non-binary JFGDD and iterative detector-decoder. For example, even with a large I_{det} , the binary JFGDD with the best constrained target showed a performance loss of around 1.25 dB and 1 dB over the iterative detector-decoder and non-binary JFGDD respectively. The non-binary JFGDD, on the other hand, showed a slight performance loss of 0.25 dB over the conventional system when I_{det} is small (ie. $I_{det} = 2$). With larger I_{det} however, its performance converges to the iterative detector-decoder at all SNR values.

Figure 5.18: Comparison of performance among various detector-decoder systems over a 1D magnetic recording channel.

5.4 Summary

In this chapter, the JFGDD was proposed as a second alternative to the state-of-the-art iterative detector decoder. The JFGDD operates on a super factor graph structure that is integrated from a channel factor graph and a code factor graph. As the state-of-the-art iterative detector decoder consists of the BCJR detector that executes on a channel trellis and an LDPC decoder that operates on a code factor graph, the proposed JFGDD was constructed by retaining the

code factor graph and converting the channel trellis into a channel factor graph. The SPA that executes on the code factor graph therefore has to be implemented on the channel factor graph to enable a joint detection and decoding process. Section 5.1.1 therefore proposes the implementation of the SPA on the channel factor graph for the task of detection. The resulting detector was coined the factor-graph based detector. The integration of the channel and code factor graph and the joint execution of the SPA on the super factor graph of the JFGDD were subsequently proposed in Section 5.1.2.

Despite introducing a joint detection and decoding process and thereby eliminating the sub-optimality present in the iterative process of the state-of-the-art system, the JFGDD has a shortcoming in terms of the presence of short cycles in its channel and code factor graph. As short cycles such as length-4 cycles are known to cause performance degradation, these cycles need to be reduced or completely eliminated in order for the JFGDD to perform well. The existing Neal Ranford's technique of eliminating short cycles in the code factor graph was adopted. Two methods are proposed to mitigate the presence of short cycles in the channel factor graph. The first method involves the use of a GPR equalizer to design targets with constrained coefficients. The constraints are imposed such that the channel factor graph of the resulting target has fewer or no length-4 cycles. The second proposed method involves the modification of the binary factor graph-based detector into a non-binary one which computes and exchange probabilities in terms of symbols instead of bits. The grouping of consecutive variable nodes in this method will result in fewer or no length-4 cycles in the non-binary channel factor graph.

Extensive computer simulations were carried out to evaluate the performance of the JFGDD against the state-of-the-art iterative detector-decoder. Three studies were carried out. In the first study, the performance of the binary and non-binary factor graph-based detector performances were benchmarked against the BCJR detector (used in the iterative system) over an uncoded 1D ISI channel. The binary factor graph-based detector suffered a performance loss of 1 dB at its saturation point over the iterative detector-decoder. The non-binary factor graph-based detector on the other hand performs as well as the iterative system at its saturation

point. In the second study, the factor graph-based detectors were each integrated with an LDPC decoder to form a binary and non-binary JFGDD. Their performances were compared to the iterative detector-decoder over a coded 1D ISI channel. The binary JFGDD displayed a saturation in performance with a large number of iterations ($I_{det} = 10$). At this point, it suffers a loss of approximately 1 dB over the iterative detector-decoder. The non-binary JFGDD on the other hand reaches its saturation point at $I_{det} = 2$. At this point, its performance is identical to the iterative detector-decoder. The third study involves a performance comparison between the binary and non-binary JFGDD and the iterative detector-decoder over a magnetic recording channel using the GFP channel model (proposed in Chapter 3). In this study, the GPR equalizer is used to design various constrained targets for the binary JFGDD whereas the monic constraint is used for the other two detector-decoder systems.

For the binary JFGDD, simulation results showed a contradicting outcome where constraints leading to graphs with fewer or no 4-cycles (eg. $Z_{pos} = 124$) performed worse than those whose graphs contain more cycles (eg. $Z_{pos} = 3$). This was attributed to the fact that in practice, the use of the GPR equalizer causes the equalized signal to be corrupted with correlated noise. In fact, the more effective an equalizer is in removing short cycles, the stronger is the coloration of noise and hence the greater is the deterioration in performance. This performance deterioration has a stronger effect than the performance gain reaped with fewer graph cycles. The JFGDD with the best constrained found ($Z_{pos} = 3$) showed a performance loss of 1.25 dB and 1 dB over the non-binary JFGDD and iterative system. The non-binary JFGDD suffered a performance loss of approximately 0.25 dB for small values of I_{det} when benchmarked against the iterative detector-decoder. At its saturation point which occurs when $I_{det} = 6$, it achieves an identical performance.

The JFGDD proposed in this chapter can achieve an identical performance to the state-of-the-art iterative detector-decoder at best, through a non-binary implementation. Despite its inability to outperform the iterative system, the proposed JFGDD can be implemented in a parallelizable structure in hardware implementation. This is as oppose to the iterative detector-decoder which has a serial hardware implementation structure. In the future work, the binary

JFGDD system can be integrated with a Pattern Dependent Noise Predictor (PDNP) model which cancels out the effect of noise coloration by the equalizer. With this system, it is very likely that the simulation results for constraining GPR targets obtained matches with theoretical predictions and a better comparison of the two methods of mitigating short cycles in the channel factor graph can be made.

Comparing the results of the joint detection and decoding schemes proposed in this chapter and in Chapter 4, the JVDD system is clearly a more promising joint detector and decoder system than the JFGDD. It has a higher potential of replacing the state-of-the-art iterative detector-decoder system. As such, in the next chapter of the thesis, only the JVDD system is implemented and examined for equalization/detection of 2D magnetic recording channels.

Chapter 6

Two-Dimensional Equalization Schemes for Magnetic Recording Channels

Signal processing plays an important role in enhancing the performance of a magnetic recording channel. The signal processing blocks used on the receiver end of a magnetic recording system typically consists of an equalizer, detector and decoder (section 2.3). As such, advancements in signal processing techniques to support higher areal density growth should be made in the area of detection, decoding and equalization. In Chapters 4 and 5, novel joint detector decoder algorithms were proposed as alternatives to the sub-optimal state-of-the-art iterative detector-decoder. This chapter proposes novel equalization schemes to support the development of the recently proposed Shingled Magnetic Recording (SMR) and Two-Dimensional Magnetic Recording (TDMR) architecture [10][11][12].

Read-back signals from a magnetic recording channel are often corrupted with Inter-Symbol Interference (ISI) in the down-track direction and Inter-Track Interference (ITI) in the cross-track direction. In the present Perpendicular Magnetic Recording (PMR) channel, the effects of ITI are negligible as track pitches (TP) are large. As we shift towards a higher density recording channel such as SMR and TDMR, the increase in track densities of these systems causes the impact of ITI

distortions on performance to be more severe. Advances in equalization schemes can be introduced in order to manage this.

Figure 6.1: Overview of equalization schemes for SMR and TDMR channels.

The newly proposed SMR architecture targets a higher track density by writing tracks in a shingled manner (section 2.2). In other words, a new track is written by overlapping it with a segment of a previously written track. Read-back waveforms from an SMR channel is one-dimensional (1D) in nature and are obtained by reading information track by track using a single reader head. As such, the SMR channel was proposed in the Literature with existing 1D signal processing techniques that are used in today's PMR. Such conventional read-back technique causes the severe distortions of ITI in the SMR channel to be ignored. This chapter of the thesis proposes the use of multiple reader heads instead for gathering a two-dimensional (2D) read-back signal from an SMR channel. A novel 2D GPR equalization technique is proposed for the SMR channel where ITI picked up by multiple readers are exploited by equalizing the response from multiple tracks of data into a single track response. Figure 6.1 illustrates this using a block diagram.

In a TDMR channel, information bits are also written using a shingling technique, as in an SMR channel. A TDMR channel however gives a 2D read-back signal obtained by reading multiple

tracks of data at once using an array of multiple reader heads (section 2.2), in contrast to an SMR channel whose read-back is typically 1D in nature. At present, no 2D equalization schemes exist for TDMR as it is a recently proposed magnetic recording architecture and vast research is still ongoing before TDMR can be practically applied. As such, the second part of this chapter proposes novel 2D equalization schemes which equalize signals from multiple tracks of a TDMR channel into a 2D signal of various form (Figure 6.1).

6.1 Equalization for Shingled Magnetic Recording Channel

For the SMR channel that has been widely proposed in the Literature [10][11][12], the 1D signal gathered from the reader is processed with existing 1D GPR equalization followed by a 1D detector and decoder on the receiver end, in an attempt to decipher the information sequence that was written on the track. The shingling writing technique used in the SMR architecture introduces significant distortion in the cross-track direction, known as ITI, due to the overlapping of bits on adjacent tracks. This is in addition to the down-track interference, known as ISI, which arises due to overlapping of adjacent bits on the same track. The use of 1D equalizer and 1D detector cancels the effect of ISI in an SMR channel. The effect of ITI is however ignored. A gain in performance can be reaped if the ITI is addressed and measures are taken to cancel its effect. One way of achieving this is through a 2D equalization technique.

This section proposes the use of multiple reader heads for an SMR channel such that information of the ITI on a center data track can be exploited. With an array of multiple reader heads, a 2D signal is gathered during the read-back process. The signal will be corrupted with ITI as well as ISI. A novel 2D equalizer that employs the GPR technique is thus proposed in this section to equalize away the effects of ITI on the center data track. A 1D signal, representative of the center data track, results and is subsequently processed with a 1D detector for ISI cancellation. An estimate of the most likely written sequence on the center track of data is then made. The proposed 2D equalizer for an SMR channel with multiple readers is coined the “SMR

equalizer". Figure 6.2 provides an illustration of the SMR equalization with a 2D read-back signal. In the figure, triple reader heads are used to read three tracks of data simultaneously and a 2D equalization and 1D detection/decoding are performed to make an estimate of the information written on the center data track.

Figure 6.2: Block diagram depicting an SMR equalizer.

The proposed SMR equalizer reshapes a 2D channel response into a 1D target (coined the “SMR target”) having a much shorter length. The SMR equalizer and SMR target are designed jointly by minimizing the mean squared error (MSE) of the signals from the center data track. Mathematically, this can be expressed as

$$E[e_k^2] = E[d_k^2 - 2d_k y_{1,k} + y_{1,k}^2] \quad (6.1)$$

where $E[\cdot]$ denotes the expectation operator, d_k is the output signal of the SMR target (obtained by convolving signal $c_{1,k}$ with the 1D SMR target response), $y_{1,k}$ is the output signal of the 2D SMR equalizer and e_k is the error signal that is obtained from the difference between d_k and $y_{1,k}$. In general, d_k and $y_{1,k}$ can be written as:

$$d_k = \sum_{n=0}^{N_g-1} g_n c_{k-n} = \mathbf{g}^T \mathbf{c}_k \quad (6.2)$$

$$y_{1,k} = \sum_{m=-M_w}^{M_w} \sum_{n=-N_w}^{N_w} w_{m,n} r_{1-m,k-n} = \mathbf{w}^T \mathbf{r}_{1,k} \quad (6.3)$$

where $\mathbf{w} = [w_{-M_w, -N_w}, w_{-M_w, -N_w+1}, \dots, w_{M_w, N_w-1}, w_{M_w, N_w}]^T$ is the 1D representation of the 2D SMR equalizer, $\mathbf{g} = [g_0, g_1, \dots, g_{N_g-2}, g_{N_g-1}]^T$ represents the SMR target vector,

$\mathbf{c}_k = [c_k, c_{k-1}, \dots, c_{k-N_g+2}, c_{k-N_g+1}]^T$ refers to the coded sequence and $\mathbf{r}_{1,k} = [r_{1+M_w, k+N_w}, r_{2+M_w, k+N_w-1}, \dots, r_{2-M_w, k-N_w-1}, r_{1-M_w, k-N_w}]^T$ denotes input vector to the SMR equalizer.

The MSE can be minimized as follow:

$$\begin{aligned} E[e_k^2] &= E[d_k^2 - 2d_k y_{1,k} + y_{1,k}^2] \\ &= E[d_k^2] - E[2d_k y_{1,k}] + E[y_{1,k}^2] \\ &= E[\mathbf{g}^T \mathbf{c}_k \mathbf{c}_k^T \mathbf{g}] - E[2\mathbf{g}^T \mathbf{c}_k \mathbf{r}_{1,k}^T \mathbf{w}] + E[\mathbf{w}^T \mathbf{r}_{1,k} \mathbf{r}_{1,k}^T \mathbf{w}] \end{aligned} \quad (6.4)$$

The cost function can thus be expressed as

$$J = \mathbf{g}^T \mathbf{C} \mathbf{g} - 2\mathbf{g}^T \mathbf{P} \mathbf{w} + \mathbf{w}^T \mathbf{R} \mathbf{w} - 2\lambda(\mathbf{E}^T \mathbf{g} - \mathbf{i}) \quad (6.5)$$

where $\mathbf{i}$ is a constraint vector of length N_g and λ is the Lagrange multiplier that is added to the cost function equation in order to avoid a trivial solution (as presented in section 2.3.2). Matrices $\mathbf{C}$, $\mathbf{P}$ and $\mathbf{R}$ are given by:

$$\mathbf{C} = E[\mathbf{c}_k \mathbf{c}_k^T] = E \begin{bmatrix} c_k c_k & c_k c_{k-1} & \dots & c_k c_{k-N_g+1} \\ c_{k-1} c_k & c_{k-1} c_{k-1} & \dots & c_{k-1} c_{k-N_g+1} \\ \vdots & \vdots & \ddots & \vdots \\ c_{k-N_g+1} c_k & c_{k-N_g+1} c_{k-1} & \dots & c_{k-N_g+1} c_{k-N_g+1} \end{bmatrix} \quad (6.6)$$

$$\mathbf{P} = E[\mathbf{c}_k \mathbf{r}_{1,k}^T] = E \begin{bmatrix} c_k r_{1,k+N_w} & c_k r_{1,k+N_w-1} & \dots & c_k r_{1,k-N_w} \\ c_{k-1} r_{1,k+N_w} & c_{k-1} r_{1,k+N_w-1} & \dots & c_{k-1} r_{1,k-N_w} \\ \vdots & \vdots & \ddots & \vdots \\ c_{k-N_g+1} r_{1,k+N_w} & c_{k-N_g+1} r_{1,k+N_w-1} & \dots & c_{k-N_g+1} r_{1,k-N_w} \end{bmatrix} \quad (6.7)$$

$$\mathbf{R} = E[\mathbf{r}_{1,k} \mathbf{r}_{1,k}^T] = E \begin{bmatrix} r_{1,k+N_w} r_{1,k+N_w} & r_{1,k+N_w} r_{1,k+N_w-1} & \dots & r_{1,k+N_w} r_{1,k-N_w} \\ r_{1,k+N_w-1} r_{1,k+N_w} & r_{1,k+N_w-1} r_{1,k+N_w-1} & \dots & r_{1,k+N_w-1} r_{1,k-N_w} \\ \vdots & \vdots & \ddots & \vdots \\ r_{1,k-N_w} r_{1,k+N_w} & r_{1,k-N_w} r_{1,k+N_w-1} & \dots & r_{1,k-N_w} r_{1,k-N_w} \end{bmatrix} \quad (6.8)$$

J is minimized by setting its partial derivatives to 0. This yields the following solution:

$$\mathbf{w} = \mathbf{R}^{-1} \mathbf{P}^T \mathbf{g} \quad (6.9)$$

$$\mathbf{g} = \lambda(\mathbf{C} - \mathbf{P} \mathbf{R}^{-1} \mathbf{P}^T)^{-1} \mathbf{i} \quad (6.10)$$

$$\lambda = \frac{1}{\mathbf{i}^T (\mathbf{C} - \mathbf{P}\mathbf{R}^{-1}\mathbf{P}^T)^{-1} \mathbf{i}} \quad (6.11)$$

With the proposed SMR equalizer design, the effects of ITI on the center data track can be equalized away, leaving the distortions of ISI on the same track to be cancelled out using a detection technique. With a subsequent decoding step, an estimate of the information sequence that was most likely written on the center track is then made. To this end, the state-of-the-art iterative detector-decoder, the Joint Viterbi Detector (JVDD) (proposed in Chapter 4) and the Joint Factor Graph Detector Decoder (JFGDD) (proposed in Chapter 5) can be used in conjunction with the proposed SMR equalizer.

By exploiting the effects of ITI, the SMR equalizer should theoretically result in a better performance than the conventional 1D GPR equalizer. This will be verified later in the Simulations and Results section (section 6.3). The drawback of the proposed scheme however lies in the need for a multiple reader head element, as opposed to single reader head that is used in the existing PMR channel.

6.2 Equalization for Two-Dimensional Magnetic Recording Channel

The TDMR channel was proposed with the idea of using advanced signal processing on a 2D read-back waveform gathered from the channel [9][10][11][29]. The 2D signal in a TDMR channel is obtained by reading multiple tracks of data at once using an array of multiple reader heads. At present, no 2D equalization schemes exist for a magnetic recording channel. The conventional GPR equalizer used at present and other equalizers such as the Least Mean Square (LMS) equalizer [53] that were used in the past were designed specifically for a 1D signal. These 1D equalizers are sufficient for the existing PMR channel as the effects of ITI in this channel is negligible. The extent of distortion in a TDMR channel is however not negligible

because of the presence of shingled tracks. As such, the use of 2D equalizers is necessary to manage the ITI distortion.

This section proposes three novel 2D GPR equalization schemes for the TDMR channel. A 2D equalizer can be used to equalize the effect of ITI in the cross-track direction by reshaping the 2D channel response into multiple 1D target responses. The interference of the TDMR channel is thus treated as ISI in the down-track direction of each data track. Multiple 1D detector can then be used to cancel the effects of ISI on each track independently. In order to perform equalization in this manner, two 2D GPR equalizer designs are proposed. The first is coined the “Symmetric TDMR equalizer” and is presented in section 6.2.1. Its design involves reshaping of a 2D channel response into multiple 1D target responses, all of which are identical. The second proposed equalizer is coined the “Asymmetric TDMR equalizer” (section 6.2.2) where the multiple 1D target responses derived from a 2D channel response are not identical. Aside from equalizing the effects of ITI and treating the interference as 1D, equalization for a TDMR channel can also be carried out by treating the ITI and ISI interferences as a 2D interference. In other words, the 2D channel response is reshaped into a much shorter 2D target response and a 2D detector is employed to simultaneously undo the effect of ITI and ISI. The proposed equalizer design in this case is presented in section 6.2.3 and is coined the “TDMR equalizer”.

6.2.1 Symmetric TDMR Equalizer

The symmetric TDMR equalizer adopts the GPR equalization technique where an equalizer and a target are jointly designed using the minimum mean squared error (MMSE) criterion (section 2.3.2). The symmetric TDMR equalizer aims to reshape the 2D response of a TDMR channel into multiple symmetric 1D target responses that are much shorter in length as compared to that of the channel response. This is illustrated in Figure 6.3 where the symmetric TDMR equalizer is integrated with a TDMR channel that uses triple reader heads to read data on three adjacent tracks concurrently.

Figure 6.3: Block diagram depicting a Symmetric TDMR equalizer

The symmetric TDMR equalizer is designed by minimizing the MSE over the outputs of the three tracks, as depicted in Figure 6.3. In this case, the MSE of the error signal, $e_{j,k}$ can be expressed as

$$\begin{aligned}
 E[e_{j,k}^2] &= E[e_{0,k}^2] + E[e_{1,k}^2] + E[e_{2,k}^2] \\
 &= E[d_{0,k}^2 - 2d_{0,k}y_{0,k} + y_{0,k}^2] + E[d_{1,k}^2 - 2d_{1,k}y_{1,k} + y_{1,k}^2] + \\
 &\quad E[d_{2,k}^2 - 2d_{2,k}y_{2,k} + y_{2,k}^2]
 \end{aligned} \tag{6.12}$$

where $E[\cdot]$ denotes the expectation operator, and $d_{j,k}$ and $y_{j,k}$ are the output signals of the symmetric target and symmetric TDMR equalizer respectively. In general, $d_{j,k}$ and $y_{j,k}$ can be written as

$$d_{j,k} = \sum_{n=0}^{N_g-1} g_n c_{j,k-n} = \mathbf{g}^T \mathbf{c}_{j,k} \tag{6.13}$$

$$y_{j,k} = \sum_{m=-M_w}^{M_w} \sum_{n=-N_w}^{N_w} w_{m,n} r_{j-m,k-n} = \mathbf{w}^T \mathbf{r}_{j,k} \tag{6.14}$$

where $\mathbf{g} = [g_0, g_1, \dots, g_{N_g-2}, g_{N_g-1}]^T$ refers to the target vector, $\mathbf{c}_{j,k} = [c_{j,k}, c_{j,k-1}, \dots, c_{j,k-N_g+2}, c_{j,k-N_g+1}]^T$ is the coded information vector, $\mathbf{w} = [w_{-M_w, -N_w}, w_{-M_w, -N_w+1}, \dots, w_{0,0}, \dots, w_{M_w, N_w-1}, w_{M_w, N_w}]^T$ is the equalizer coefficients

vector and $\mathbf{r}_{j,k} = [r_{j+M_w,k+N_w}, r_{j+M_w,k+N_w-1}, \dots, r_{j-M_w,k-N_w+1}, r_{j-M_w,k-N_w}]^T$ is the equalizer input vector. Here, $\mathbf{w}$ and $\mathbf{r}_{j,k}$ are represented as 1D vectors for ease of mathematical derivation. From equation (6.12), J can be written as:

$$\begin{aligned} J = & \mathbf{g}^T \mathbf{C}_0 \mathbf{g} - 2 \mathbf{g}^T \mathbf{P}_0 \mathbf{w} + \mathbf{w}^T \mathbf{R}_0 \mathbf{w} + \\ & \mathbf{g}^T \mathbf{C}_1 \mathbf{g} - 2 \mathbf{g}^T \mathbf{P}_1 \mathbf{w} + \mathbf{w}^T \mathbf{R}_1 \mathbf{w} + \\ & \mathbf{g}^T \mathbf{C}_2 \mathbf{g} - 2 \mathbf{g}^T \mathbf{P}_2 \mathbf{w} + \mathbf{w}^T \mathbf{R}_2 \mathbf{w} \end{aligned} \quad (6.15)$$

In equation (6.15), $\mathbf{C}_j$ refers to the $N_g \times N_g$ auto-correlation matrix of $\mathbf{c}_{j,k}$, $\mathbf{R}_j$ is the $(2M_w + 1) \times (2M_w + 1)$ auto-correlation matrix of $\mathbf{r}_{j,k}$ and $\mathbf{P}_j$ is the $(N_g) \times (2M_w + 1)$ cross-correlation matrix between $\mathbf{c}_{j,k}$ and $\mathbf{r}_{j,k}$. These matrices are given by

$$\mathbf{C}_j = E[\mathbf{c}_{j,k} \mathbf{c}_{j,k}^T] = E \begin{bmatrix} c_{j,k} c_{j,k} & c_{j,k} c_{j,k-1} & \dots & c_{j,k} c_{j,k-N_g+1} \\ c_{j,k-1} c_{j,k} & c_{j,k-1} c_{j,k-1} & \dots & c_{j,k-1} c_{j,k-N_g+1} \\ \vdots & \vdots & \ddots & \vdots \\ c_{j,k-N_g+1} c_{j,k} & c_{j,k-N_g+1} c_{j,k-1} & \dots & c_{j,k-N_g+1} c_{j,k-N_g+1} \end{bmatrix} \quad (6.16)$$

$$\mathbf{R}_j = E[\mathbf{r}_{j,k} \mathbf{r}_{j,k}^T] = \begin{bmatrix} \mathbf{R}_{-M_w, -M_w} & \mathbf{R}_{-M_w, -M_w+1} & \dots & \mathbf{R}_{-M_w, M_w} \\ \mathbf{R}_{-M_w+1, -M_w} & \mathbf{R}_{-M_w+1, -M_w+1} & \dots & \mathbf{R}_{-M_w+1, M_w} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{M_w, -M_w} & \mathbf{R}_{M_w, -M_w+1} & \dots & \mathbf{R}_{M_w, M_w} \end{bmatrix} \quad (6.17)$$

$$\mathbf{P}_j = E[\mathbf{c}_{j,k} \mathbf{r}_{j,k}^T] = E \begin{bmatrix} c_k r_{k+N_w} & c_k r_{k+N_w-1} & \dots & c_k r_{k-N_w} \\ c_{k-1} r_{k+N_w} & c_{k-1} r_{k+N_w-1} & \dots & c_{k-1} r_{k+N_w} \\ \vdots & \vdots & \ddots & \vdots \\ c_{k-N_g+1} r_{k+N_w} & c_{k-N_g+1} r_{k+N_w-1} & \dots & c_{k-N_g+1} r_{k-N_w} \end{bmatrix} \quad (6.18)$$

In equation (6.17), $\mathbf{R}_{\alpha,\beta}$ refers to a sub-matrix in $\mathbf{R}_j$, and is given by:

$$\mathbf{R}_{\alpha,\beta} = E[\mathbf{r}_{j-\alpha} \mathbf{r}_{j-\beta}^T] \quad (6.19)$$

where $\mathbf{r}_{j-x}$ is defined as

$$\mathbf{r}_{j-x} = [r_{j-x,k+N_w}, r_{j-x,k+N_w-1}, \dots, r_{j-x,k-N_w}] \quad (6.20)$$

Using the above, $\mathbf{R}_j$ can be re-written as:

$$\mathbf{R}_j = \begin{bmatrix} E[\mathbf{r}_{j+M_w}\mathbf{r}_{j+M_w}^T] & E[\mathbf{r}_{j+M_w}\mathbf{r}_{j+M_w-1}^T] & \cdots & E[\mathbf{r}_{j+M_w}\mathbf{r}_{j-M_w}^T] \\ E[\mathbf{r}_{j+M_w-1}\mathbf{r}_{j+M_w}^T] & E[\mathbf{r}_{j+M_w-1}\mathbf{r}_{j+M_w-1}^T] & \cdots & E[\mathbf{r}_{j+M_w-1}\mathbf{r}_{j-M_w}^T] \\ \vdots & \vdots & \ddots & \vdots \\ E[\mathbf{r}_{j-M_w}\mathbf{r}_{j+M_w}^T] & E[\mathbf{r}_{j-M_w}\mathbf{r}_{j+M_w-1}^T] & \cdots & E[\mathbf{r}_{j-M_w}\mathbf{r}_{j-M_w}^T] \end{bmatrix} \quad (6.21)$$

The cost function in equation (6.15) is solved by equating the partial derivative of J to 0. Applying this to the equation will however result in a trivial solution where $\mathbf{g}$ and $\mathbf{w}$ are both 0. To avoid this, λ is introduced to find the MSE of $e_{j,k}$ subject to the constraint

$$\mathbf{E}^T \mathbf{g} = \mathbf{i} \quad (6.22)$$

where $\mathbf{i} = [1 \ 0 \ \dots \ 0]$ is a monic constraint vector of length N_g . The cost function with monic constraint can then be re-written as

$$\begin{aligned} J = & \mathbf{g}^T \mathbf{C}_0 \mathbf{g} - 2\mathbf{g}^T \mathbf{P}_0 \mathbf{w} + \mathbf{w}^T \mathbf{R}_0 \mathbf{w} + \\ & \mathbf{g}^T \mathbf{C}_1 \mathbf{g} - 2\mathbf{g}^T \mathbf{P}_1 \mathbf{w} + \mathbf{w}^T \mathbf{R}_1 \mathbf{w} + \\ & \mathbf{g}^T \mathbf{C}_2 \mathbf{g} - 2\mathbf{g}^T \mathbf{P}_2 \mathbf{w} + \mathbf{w}^T \mathbf{R}_2 \mathbf{w} - 2\lambda(\mathbf{i}^T \mathbf{g} - 1) \end{aligned} \quad (6.23)$$

The cost function with Lagrange multiplier in equation (6.12) is then minimized by setting its partial derivatives to 0. This yields the following:

$$\nabla_g J = 2(\mathbf{C}_0 + \mathbf{C}_1 + \mathbf{C}_2) \mathbf{g} - 2(\mathbf{P}_0 + \mathbf{P}_1 + \mathbf{P}_2) \mathbf{w} - 2\lambda \mathbf{i} = 0 \quad (6.24)$$

$$\nabla_w J = 2(\mathbf{R}_0 + \mathbf{R}_1 + \mathbf{R}_2) \mathbf{w} - 2(\mathbf{P}_0^T + \mathbf{P}_1^T + \mathbf{P}_2^T) \mathbf{g} = 0 \quad (6.25)$$

$$\nabla_\lambda J = \mathbf{i}^T \mathbf{g} - 1 \quad (6.26)$$

The optimum $\mathbf{w}$ and $\mathbf{g}$ can subsequently be found by solving equations (6.24) - (6.26) simultaneously. This gives the following solution:

$$\mathbf{w} = \bar{\mathbf{R}}^{-1} \bar{\mathbf{P}}^T \mathbf{g} \quad (6.27)$$

$$\mathbf{g} = \lambda(\bar{\mathbf{C}} - \bar{\mathbf{P}} \bar{\mathbf{R}}^{-1} \bar{\mathbf{P}}^T)^{-1} \mathbf{i} \quad (6.28)$$

$$\lambda = \frac{1}{\mathbf{i}^T (\bar{\mathbf{C}} - \bar{\mathbf{P}}\mathbf{R}^{-1}\bar{\mathbf{P}}^T)^{-1}\mathbf{i}} \quad (6.29)$$

where

$$\bar{\mathbf{C}} = \mathbf{C}_0 + \mathbf{C}_1 + \mathbf{C}_2 \quad (6.30)$$

$$\bar{\mathbf{R}} = \mathbf{R}_0 + \mathbf{R}_1 + \mathbf{R}_2 \quad (6.31)$$

$$\bar{\mathbf{P}} = \mathbf{P}_0 + \mathbf{P}_1 + \mathbf{P}_2 \quad (6.32)$$

The proposed design of the Symmetric TDMR equalizer applies to a TDMR channel that employ triple reader heads to read three adjacent tracks of data simultaneously. The proposed design can be extended to TDMR channels that use a different number of reader heads by modifying the MSE criterion in equation (6.12). The number of auto-correlation and cross-correlation matrices and their definitions in equation (6.16) – (6.18) will be different in this case.

With the proposed Symmetric TDMR equalizer design, the effects of ITI of the channel are equalized resulting in only ISI interference on every track of data. The ISI interference can be cancelled out using a 1D detector independently for each track. For a coded channel, 1D decoding can subsequently be applied to each track to make an estimate of the most likely written information sequence on the media.

6.2.2 Asymmetric TDMR Equalizer

In section 6.2.1, the symmetric TDMR equalizer was formulated such that multiple symmetrical target responses are characterized for a TDMR channel. This section proposes an asymmetric TDMR equalizer where the 2D channel response is reshaped into multiple asymmetrical 1D target responses as depicted in Figure 6.4. In other words, the GPR target responses are designed to have different coefficients -

$$\mathbf{g}_0 \neq \mathbf{g}_1 \neq \mathbf{g}_2 \quad (6.33)$$

where $\mathbf{g}_0$, $\mathbf{g}_1$ and $\mathbf{g}_2$ refers to each of the three asymmetrical targets for a TDMR channel with triple reader heads as illustrated in Figure 6.4.

Figure 6.4: Block diagram depicting an Asymmetric TDMR equalizer

The mathematical design of the proposed Asymmetric TDMR equalizer is much more complex than that for the Symmetric TDMR equalizer. The cost function in equation (6.15) of the symmetric TDMR equalizer can be reformulated for the asymmetric TDMR equalizer as follow:

$$\begin{aligned}
 J &= E[e_{0,k}^2] + E[e_{1,k}^2] + E[e_{2,k}^2] - 2\lambda_0(\mathbf{i}^T \mathbf{g}_0 - 1) - \\
 &\quad 2\lambda_1(\mathbf{i}^T \mathbf{g}_1 - 1) - 2\lambda_2(\mathbf{i}^T \mathbf{g}_2 - 1) \\
 &= \mathbf{g}_0^T \mathbf{C}_0 \mathbf{g}_0 - 2\mathbf{g}_0^T \mathbf{P}_0 \mathbf{w} + \mathbf{w}^T \mathbf{R}_0 \mathbf{w} - 2\lambda_0(\mathbf{i}^T \mathbf{g}_0 - 1) \\
 &\quad \mathbf{g}_1^T \mathbf{C}_1 \mathbf{g}_1 - 2\mathbf{g}_1^T \mathbf{P}_1 \mathbf{w} + \mathbf{w}^T \mathbf{R}_1 \mathbf{w} - 2\lambda_1(\mathbf{i}^T \mathbf{g}_1 - 1) \\
 &\quad \mathbf{g}_2^T \mathbf{C}_2 \mathbf{g}_2 - 2\mathbf{g}_2^T \mathbf{P}_2 \mathbf{w} + \mathbf{w}^T \mathbf{R}_2 \mathbf{w} - 2\lambda_2(\mathbf{i}^T \mathbf{g}_2 - 1)
 \end{aligned} \tag{6.34}$$

where matrices $\mathbf{C}_0$, $\mathbf{C}_1$, $\mathbf{C}_2$, $\mathbf{P}_0$, $\mathbf{P}_1$, $\mathbf{P}_2$, $\mathbf{R}_0$, $\mathbf{R}_1$, $\mathbf{R}_2$ and $\mathbf{w}$ are defined in the same manner as that for symmetric TDMR equalizer. In equation (6.34), λ_0 , λ_1 and λ_2 are the Lagrange multipliers for the respective tracks. These are necessary to avoid a trivial solution of the cost function.

Minimizing the cost function in equation (6.34) results in the following equations:

$$\nabla_{\mathbf{g}_0} J = 2\mathbf{C}_0 \mathbf{g}_0 - 2\mathbf{P}_0 \mathbf{w} - 2\lambda_0 \mathbf{i} = 0 \tag{6.35}$$

$$\nabla_{\mathbf{g}_1} J = 2\mathbf{C}_1 \mathbf{g}_1 - 2\mathbf{P}_1 \mathbf{w} - 2\lambda_1 \mathbf{i} = 0 \tag{6.36}$$

$$\nabla_{\mathbf{g}_2} J = 2\mathbf{C}_2 \mathbf{g}_2 - 2\mathbf{P}_2 \mathbf{w} - 2\lambda_2 \mathbf{i} = 0 \tag{6.37}$$

$$\nabla_w J = 2(\mathbf{R}_0 + \mathbf{R}_1 + \mathbf{R}_2)\mathbf{w} - 2(\mathbf{P}_0^T \mathbf{g}_0 + \mathbf{P}_1^T \mathbf{g}_1 + \mathbf{P}_2^T \mathbf{g}_2) = 0 \quad (6.38)$$

$$\nabla_{\lambda_0} J = \mathbf{i}^T \mathbf{g}_0 - 1 \quad (6.39)$$

$$\nabla_{\lambda_1} J = \mathbf{i}^T \mathbf{g}_1 - 1 \quad (6.40)$$

$$\nabla_{\lambda_2} J = \mathbf{i}^T \mathbf{g}_2 - 1 \quad (6.41)$$

From equation (6.38), the optimum $\mathbf{w}$ can be expressed as

$$\mathbf{w} = \mathbf{R}^{-1}(\mathbf{P}_0^T \mathbf{g}_0 + \mathbf{P}_1^T \mathbf{g}_1 + \mathbf{P}_2^T \mathbf{g}_2) \quad (6.42)$$

Substituting $\mathbf{w}$ into equations (6.35), (6.36) and (6.37) results in the following:

$$\mathbf{C}_{00}\mathbf{g}_0 + \mathbf{C}_{01}\mathbf{g}_1 + \mathbf{C}_{02}\mathbf{g}_2 = \lambda_0 \mathbf{i} \quad (6.43)$$

$$\mathbf{C}_{10}\mathbf{g}_0 + \mathbf{C}_{11}\mathbf{g}_1 + \mathbf{C}_{12}\mathbf{g}_2 = \lambda_1 \mathbf{i} \quad (6.44)$$

$$\mathbf{C}_{20}\mathbf{g}_0 + \mathbf{C}_{21}\mathbf{g}_1 + \mathbf{C}_{22}\mathbf{g}_2 = \lambda_2 \mathbf{i} \quad (6.45)$$

where $\mathbf{C}_{00}$ - $\mathbf{C}_{22}$ are matrices which are defined as follow:

$\mathbf{C}_{00} = (\mathbf{C}_0 - \mathbf{P}_0 \mathbf{R}^{-1} \mathbf{P}_0^T),$	$\mathbf{C}_{00} = -\mathbf{P}_0 \mathbf{R}^{-1} \mathbf{P}_1^T,$	$\mathbf{C}_{00} = -\mathbf{P}_0 \mathbf{R}^{-1} \mathbf{P}_2^T$
$\mathbf{C}_{10} = -\mathbf{P}_1 \mathbf{R}^{-1} \mathbf{P}_0^T,$	$\mathbf{C}_{10} = (\mathbf{C}_1 - \mathbf{P}_1 \mathbf{R}^{-1} \mathbf{P}_1^T),$	$\mathbf{C}_{10} = -\mathbf{P}_1 \mathbf{R}^{-1} \mathbf{P}_2^T$
$\mathbf{C}_{20} = -\mathbf{P}_2 \mathbf{R}^{-1} \mathbf{P}_0^T,$	$\mathbf{C}_{20} = -\mathbf{P}_2 \mathbf{R}^{-1} \mathbf{P}_1^T,$	$\mathbf{C}_{20} = (\mathbf{C}_2 - \mathbf{P}_2 \mathbf{R}^{-1} \mathbf{P}_2^T$

From equations (6.43)-(6.45), $\mathbf{g}_0$, $\mathbf{g}_1$ and $\mathbf{g}_2$ can be solved using the Gaussian elimination technique [81]. Using the Gaussian elimination, the optimum $\mathbf{g}_0$, $\mathbf{g}_1$ and $\mathbf{g}_2$ can be solved and expressed as:

$$\mathbf{g}_0 = (\mathbf{D}_{21}^{-1} \mathbf{D}_{22} - \mathbf{D}_{11}^{-1} \mathbf{D}_{12})^{-1} (\mathbf{D}_{21}^{-1} \mathbf{d}_2 - \mathbf{D}_{11}^{-1} \mathbf{d}_1) \quad (6.46)$$

$$\mathbf{g}_1 = (\mathbf{E}_{21}^{-1} \mathbf{E}_{22} - \mathbf{E}_{11}^{-1} \mathbf{E}_{12})^{-1} (\mathbf{E}_{21}^{-1} \mathbf{e}_2 - \mathbf{E}_{11}^{-1} \mathbf{e}_1) \quad (6.47)$$

$$\mathbf{g}_2 = (\mathbf{B}_{21}^{-1} \mathbf{B}_{22} - \mathbf{B}_{11}^{-1} \mathbf{B}_{12})^{-1} (\mathbf{B}_{21}^{-1} \mathbf{b}_2 - \mathbf{B}_{11}^{-1} \mathbf{b}_1) \quad (6.48)$$

where matrices $\mathbf{B}_{11}$ - $\mathbf{B}_{22}$, $\mathbf{E}_{11}$ - $\mathbf{E}_{22}$ and $\mathbf{D}_{11}$ - $\mathbf{D}_{22}$ are defined as follow:

$\mathbf{B}_{11} = \mathbf{C}_{10}^{-1} \mathbf{C}_{11} - \mathbf{C}_{00}^{-1} \mathbf{C}_{01}$	$\mathbf{E}_{11} = \mathbf{C}_{10}^{-1} \mathbf{C}_{12} - \mathbf{C}_{00}^{-1} \mathbf{C}_{02}$	$\mathbf{D}_{11} = \mathbf{C}_{12}^{-1} \mathbf{C}_{11} - \mathbf{C}_{02}^{-1} \mathbf{C}_{01}$
---	---	---

$\mathbf{B}_{12} = \mathbf{C}_{10}^{-1}\mathbf{C}_{12} - \mathbf{C}_{00}^{-1}\mathbf{C}_{02}$	$\mathbf{E}_{12} = \mathbf{C}_{10}^{-1}\mathbf{C}_{11} - \mathbf{C}_{00}^{-1}\mathbf{C}_{01}$	$\mathbf{D}_{12} = \mathbf{C}_{12}^{-1}\mathbf{C}_{10} - \mathbf{C}_{02}^{-1}\mathbf{C}_{00}$
$\mathbf{B}_{21} = \mathbf{C}_{20}^{-1}\mathbf{C}_{21} - \mathbf{C}_{00}^{-1}\mathbf{C}_{01}$	$\mathbf{E}_{21} = \mathbf{C}_{20}^{-1}\mathbf{C}_{22} - \mathbf{C}_{00}^{-1}\mathbf{C}_{02}$	$\mathbf{D}_{21} = \mathbf{C}_{22}^{-1}\mathbf{C}_{21} - \mathbf{C}_{02}^{-1}\mathbf{C}_{01}$
$\mathbf{B}_{22} = \mathbf{C}_{20}^{-1}\mathbf{C}_{22} - \mathbf{C}_{00}^{-1}\mathbf{C}_{02}$	$\mathbf{E}_{22} = \mathbf{C}_{20}^{-1}\mathbf{C}_{21} - \mathbf{C}_{00}^{-1}\mathbf{C}_{01}$	$\mathbf{D}_{22} = \mathbf{C}_{22}^{-1}\mathbf{C}_{20} - \mathbf{C}_{02}^{-1}\mathbf{C}_{00}$

Similar to the system with symmetric TDMR equalizer, the multiple 1D read-back sequences from each track of data can be passed to multiple 1D detectors for independent ISI cancellation as shown in Figure 6.4.

6.2.3 TDMR Equalizer

In sections 6.2.1 and 6.2.2, the Symmetric TDMR equalizer and Asymmetric TDMR equalizer was proposed to equalize the effects of ITI in the cross-track direction, leaving the ISI in the down-track direction of each track to be cancelled out independently using multiple 1D detectors. This section proposes an alternative equalization method for TDMR channels. The proposed equalizer has been coined the TDMR equalizer. The equalizer is proposed such that the ITI and ISI of the channel are treated as a 2D interference which requires a 2D detection to simultaneously undo the effects of ITI and ISI. In other words, the TDMR equalizer reshapes the 2D channel response into a 2D target response that has a much shorter length, as illustrated in Figure 6.5.

Figure 6.5: Block diagram depicting anTDMR equalizer

The 2D TDMR equalizer is designed by minimizing the MSE of the error signals across all data tracks. Unlike the design of the Symmetric and Asymmetric TDMR Equalizer presented in sections 6.2.1. and 6.2.2 respectively, the target output, $d_{j,k}$, in a 2D TDMR equalizer system is obtained via a 2D convolution of a 2D encoded sequence with a 2D target of size $M_w \times N_w$. Mathematically, this is expressed as

$$d_{j,k} = \sum_{m=0}^{M_g-1} \sum_{n=0}^{N_g-1} g_{m,n} c_{j-m,k-n} = \mathbf{g}^T \mathbf{c}_{j,k} \quad (6.49)$$

where $\mathbf{c}_{j,k} = [c_{j,k}, c_{j,k-1}, \dots, c_{j-M_g+1,k-N_g}, c_{j-M_g+1,k-N_g+1}]$ is a 1D representation of the encoded block of data and $\mathbf{g} = [g_{0,0}, g_{0,1}, \dots, g_{M_g-1,N_g-2}, g_{M_g-1,N_g-2}]$ is a 1D representation of the 2D TDMR target. Here, $\mathbf{g}$ and $\mathbf{c}_{j,k}$ are represented in 1D format for ease of mathematical derivation. Note that the equalizer signal output, $y_{j,k}$, in this case is the same as that defined for the Symmetric and Asymmetric TDMR equalizer (equation (6.14)). The cost function for the 2D TDMR equalizer with monic constraint can be written as:

$$\begin{aligned} J = & \mathbf{g}^T \mathbf{C}_0 \mathbf{g} - 2\mathbf{g}^T \mathbf{P}_0 \mathbf{w} + \mathbf{w}^T \mathbf{R}_0 \mathbf{w} + \\ & \mathbf{g}^T \mathbf{C}_1 \mathbf{g} - 2\mathbf{g}^T \mathbf{P}_1 \mathbf{w} + \mathbf{w}^T \mathbf{R}_1 \mathbf{w} + \\ & \mathbf{g}^T \mathbf{C}_2 \mathbf{g} - 2\mathbf{g}^T \mathbf{P}_2 \mathbf{w} + \mathbf{w}^T \mathbf{R}_2 \mathbf{w} - 2\lambda(\mathbf{i}^T \mathbf{g} - 1) \end{aligned} \quad (6.50)$$

The cost function in equation (6.50) resembles that in equation (6.15) for Symmetric TDMR equalizer. In equation (6.50) however, the size and computation of matrices $\mathbf{C}_x$ and $\mathbf{P}_x$ are different from that in the symmetric equalizer case. Here, $\mathbf{C}_0$, $\mathbf{C}_1$ and $\mathbf{C}_2$ refers to the $(M_g N_g) \times (M_g N_g)$ auto-correlation matrices of $\mathbf{c}_{j,k}$ while $\mathbf{P}_0$, $\mathbf{P}_1$ and $\mathbf{P}_2$ are the $(M_g N_g) \times (2M_w + 1)(2N_w + 1)$ cross-correlation matrices of $\mathbf{c}_{j,k}$ and $\mathbf{r}_{j,k}$. These matrices can be computed as follow:

$$\mathbf{C}_j = E[\mathbf{c}_{j,k} \mathbf{c}_{j,k}^T] = \begin{bmatrix} \mathbf{C}_{0,0} & \mathbf{C}_{0,1} & \cdots & \mathbf{C}_{0,M_g-1} \\ \mathbf{C}_{1,0} & \mathbf{C}_{1,1} & \cdots & \mathbf{C}_{1,M_g-1} \\ \vdots & \vdots & & \vdots \\ \mathbf{C}_{M_g-1,0} & \mathbf{C}_{M_g-1,1} & \cdots & \mathbf{C}_{M_g-1,M_g-1} \end{bmatrix} \quad (6.51)$$

$$\mathbf{P}_j = E[\mathbf{c}_{j,k} \mathbf{r}_{j,k}^T] = \begin{bmatrix} \mathbf{P}_{0,-M_w} & \mathbf{P}_{0,-M_w+1} & \cdots & \mathbf{P}_{0,M_w} \\ \mathbf{P}_{1,-M_w} & \mathbf{P}_{1,-M_w+1} & \cdots & \mathbf{P}_{1,M_w} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{P}_{M_g-1,-M_w} & \mathbf{P}_{M_g-1,-M_w+1} & \cdots & \mathbf{P}_{M_g-1,M_w} \end{bmatrix} \quad (6.52)$$

In equation (6.51), $\mathbf{C}_{\alpha,\beta}$ refers to a sub-matrix in $\mathbf{C}_j$, and is given by:

$$\mathbf{C}_{\alpha,\beta} = E[\mathbf{c}_{j-\alpha} \mathbf{c}_{j-\beta}^T] \quad (6.53)$$

where $\mathbf{c}_{j-x}$ (where $x \in \{\alpha, \beta\}$) is defined as

$$\mathbf{c}_{j-x} = [c_{j-x,k}, c_{j-x,k-1}, \dots, c_{j-x,k-N_g+1}] \quad (6.54)$$

In equation (6.52), $\mathbf{P}_{\alpha,\beta}$ refers to a sub-matrix in $\mathbf{P}_j$, and is given by:

$$\mathbf{P}_{\alpha,\beta} = E[\mathbf{c}_{j-\alpha} \mathbf{r}_{j-\beta}^T] \quad (6.55)$$

where $\mathbf{r}_{j-\beta}$ is defined as

$$\mathbf{r}_{j-\beta} = [r_{j-\beta,k+N_w}, r_{j-\beta,k+N_w-1}, \dots, r_{j-\beta,k-N_w}] \quad (6.56)$$

The optimum $\mathbf{w}$ and $\mathbf{g}$ are then solved using the same mathematical steps involved in the design of Symmetric TDMR equalizer and therefore shall not be repeated here.

The proposed TDMR equalizer presents the ITI in the cross-track direction and ISI in the down-track direction as a 2D interference that requires a 2D detection technique to simultaneously undo the effects of ITI and ISI. For a coded channel (which is used in practice), a subsequent decoding step is necessary to undo the correlations introduced by the error correction codes before a prediction of the original block of information written can be made. As the signal that emerges from the detection step is 2D in nature, a 2D decoding or multiple independent 1D decoding (that matches the encoder used) can be carried out for each track of data. In the former, a conversion of symbol LLRs from the 2D detector to bit LLRs for the 1D decoders is necessary if the state-of-the-art iterative detector-decoder is used on the receiver end.

Similarly, the bit LLRs from the decoders are converted back into symbol LLRs for the 2D detector to complete an iteration step.

6.3 Simulations and Results

The performance of the proposed SMR equalizer (section 6.1), Symmetric TDMR equalizer and TDMR equalizer (section 6.2) was evaluated through extensive computer simulations. The conventional 1D GPR equalizer was used as a performance benchmark for the proposed equalizers. Each of the equalizers were implemented and integrated with a channel that writes multiple tracks of information using the shingling technique and reads data using triple reader heads. The Grain Flipping Probability channel model that was presented in Chapter 3 is used in this study. On the detection and decoding end, two different detection and decoding schemes were used for each equalizer set-up – 1) the state-of-the-art iterative detector-decoder 2) the Joint Viterbi Detector Decoder (JVDD) that was proposed as an alternative to the iterative detector-decoder in Chapter 4. The JVDD displayed a better performance than the state-of-the-art iterative detector-decoder when used with a conventional 1D GPR equalizer over a PMR channel in Chapter 4. The JVDD is employed again in this study to evaluate its performance when used with the 2D equalizers proposed in this chapter for SMR/TDMR channel.

The triple reader heads employed for gathering read-back signals in this study are defined by a separation distance between the centers of any two consecutive readers, known as reader pitch, RP . The three reader heads are vertically aligned, with the center reader positioned at an offset position, RO , relative to the middle point of the center data track. This is illustrated in Figure 6.6. The values defined for RP and RO determines the quality of the read-back waveforms gathered from the channel, and as such has an impact on the performance of the system. In order to ensure an accurate comparison of the proposed equalization schemes, a preliminary study was carried out to identify the optimum RP and RO values for each of the experimental set-ups used. This is presented in section 6.3.1. With the optimum RP and RO

values obtained, the performances of the proposed equalizers were then evaluated using the GFP channel model. This is presented in section 6.3.2.

Figure 6.6: Illustration of reader pitch and read offset for a triple reader head configuration.

6.3.1 Optimal Reader Pitch and Read Offset

The performance of the proposed equalizers will be evaluated in section 6.3.2 for various track pitch, TP , values and a constant bit length, BL of an SMR and TDMR channel with triple reader heads. As such, the optimum values of RP and RO for the readers are first identified by characterizing each of the proposed equalizers over a wide range of RP and RO values and selecting the best combination of RP and RO based on the Signal-to-Noise Ratio (SNR) of the equalizers. The search for the optimum RP and RO of the readers was carried out using the GFP channel model with TP varied but BL kept fixed at 15 nm. For each channel TP , three tracks of random data were written using the shingling technique and are subsequently read by the triple reader heads with varying RP and RO . The width of the reader, RW , was kept constant at 30 nm. The complete set of constant and variable parameters used in this study is listed in Tables 6.1 and 6.2 respectively.

The SMR equalizer proposed in section 6.1 was implemented and characterized at varying channel TP and reader RP and RO . Each characterization was performed over a number of

iterations for an uncoded triple track data block so as to obtain auto-correlation and cross-correlation matrices (equations (6.6) – (6.8)) that adequately represent the signals. In each SMR equalizer characterization, the MSE of the error signal between the equalizer output and the target output is computed. With the MSE value, the SNR at the equalizer is then calculated by normalizing the signal energy to 1.

Table 6.1: Constant parameters used in optimal *RP* and *RO* study

Media	Dimension	700nm x 70 000nm	Reader	<i>RW</i>	30 nm
	Grain size	7 nm	SMR Equalizer	Dimension	1 x 15
Channel	Type	SMR, TDMR	TDMR Equalizer	Dimension	3 x 15
	<i>BL</i>	15 nm			

Table 6.2: Variable parameters used in optimal *RP* and *RO* study

Channel	<i>TP</i>	30, 34, 38, 42, 46, 50 nm
Reader	<i>RP</i>	8 to 62 nm (in steps of 4nm)
	<i>RO</i>	-20 to 20 nm (in steps of 5nm)

Figure 6.7 presents contour plots that depict the relationship between the SNR of the characterized SMR equalizer for various *RP* and *RO* of the triple reader heads and at a constant channel *TP*. The contour plots suggest that the proposed SMR equalizer is not very sensitive to *RP* and is able to tolerate small variations in *RO* without comprising on SNR. For example, for *TP* = 30 nm and *BL* = 15 nm (Figure 6.7 (b)), the *RP* and *RO* value that will lead to a good SNR (> 14 dB) ranges from 10 to 40 nm and -5 to 5 nm respectively. The ability of the SMR equalizer to tolerate such variations in *RO* and *RP* can be attributed to the fact that the equalizer characterizes a target for the center data track only. In other words, the responses of the two other tracks adjacent to the center tracks are ignored. A good performance can therefore be attained even if the readers are positioned close together over the center track.

Figure 6.7: Contour plots depicting the relationship between RP , RO and SNR for an SMR equalizer at $TP =$ (a) 26nm (b) 30 nm (c) 34 nm (d) 38 nm.

Table 6.3: Optimum RP and RO for various TP for SMR channel

TP (nm)	RP (nm)	RO (nm)
30	10	-4
34	10	-6
42	10	0
50	10	0

Table 6.4: Optimum RP and RO for various TP for TDMR channel

TP (nm)	RP (nm)	RO (nm)
30	26	-4
34	34	-2
42	42	-4
50	48	-4

Next, the equalizers proposed for TDMR channels in section 6.2 are each characterized using the same experimental set-up shown in Table 6.1 and Table 6.2. Figure 6.8 presents contour plots that illustrate the relationship between the SNR of a characterized Symmetric TDMR

equalizer for various RP and RO of the triple reader heads. The plots suggest that the Symmetric TDMR equalizer is more sensitive to variations in RP and RO , unlike the SMR equalizer. For a fixed TP value, the region in the contours where a good equalizer SNR is attained tends to occur when RP is close to TP and when RO is 0 or deviates slightly from 0. This can be attributed to the fact that unlike the SMR equalizer, the Symmetric TDMR equalizer tries to optimize the MSE of the error signal over three tracks of data. Hence if any of the three reader heads are positioned too far off from its assigned track, a poor MSE and thus SNR results. The same conclusion can be drawn for the Asymmetric TDMR equalizer and TDMR equalizer.

Figure 6.8: Contour plots depicting the relationship between RP , RO and SNR for a 1D Symmetric TDMR equalizer at TP = (a) 26nm (b) 30 nm (c) 34 nm (d) 38 nm.

The optimum RP and RO values found for each channel type and channel TP are given in Tables 6.3 and 6.4. These values are used in the subsequent study (section 6.3.2) which compares the performance of the various equalization/detection schemes proposed.

6.3.2 Performance of Proposed Equalization Schemes

The optimum RP and RO found in section 6.3.1 gives each of the proposed equalizers its best performance. With the optimum values obtained for the triple reader heads, the performance of each of the proposed equalizer designs are benchmarked against the conventional 1D GPR equalizer (identified as the PMR equalizer in this section). Note that the conventional 1D GPR equalizer uses a single reader and as such do not require any optimization of RP nor RO . The performance comparison was carried out using the GFP channel model where three tracks of encoded sequences were written at various TP and fixed BL of 15 nm. On the receiver end, each of the proposed equalizers was integrated with the state-of-the-art iterative detector-decoder as well as the JVDD in a separate set-up. Random Low-Density Parity Check (LDPC) codes with column weight, $w_{col} = 4$ are used for the iterative detector-decoder while JVDD codes proposed in Chapter 4 are used for the JVDD. For each experimental set-up, the length of the code, N , and the code rate, R , are varied to study the effect that the codes have on the entire system. Table 6.5 and 6.6 list the constant and variable parameters used respectively.

For each of the four equalization schemes (PMR, SMR, Symmetric TDMR and TDMR), the MSE of the error signal is computed during the equalizer design phase. The optimal MSE value computed for each equalization scheme at various TP values are presented in Table 6.7.

Figure 6.9 presents plots of Frame Error Rate (FER) against R for the four equalizers at $N = 1024$ and 2048. Comparing the performance curves for each of the proposed equalizers when integrated with the state-of-the-art iterative detector-decoder (the solid lines in the plots), the TDMR equalizer performs best overall. This gain in performance of the TDMR equalizer over the other proposed equalizer and the conventional equalizer is more apparent at a larger N value

(Figure 6.9(ii)). Despite the good performance displayed by the TDMR equalizer, the equalization scheme translates into a much higher complexity than the other schemes. The TDMR equalizer reshapes the 2D channel response into a 2D target which subsequently requires a 2D detector to simultaneously undo the ITI and ISI interferences. The Symmetric TDMR equalizer on the other hand requires three independent 1D detectors that can be executed in parallel while the SMR and PMR equalizer both require a single 1D detector. The complexity associated with a 2D detector is much higher than that for 1D detector. For the state-of-the-art iterative detector-decoder and JVDD for example, a 2D detector translates into a much higher number of trellis states and branches than its 1D counterpart.

Table 6.5: Constant parameters used in the performance comparison of the proposed equalization schemes

Media	Dimension	700nmx70000nm	SMR equalizer	Equalizer dimension	3x15
	Grain size	7nm		Target dimension	1x5
Channel	BL	15nm	Symmetric	Equalizer dimension	3x15
Reader	RW	30nm	TDMR equalizer	Target dimension	1x5
Code	Type	LDPC	2D TDMR equalizer	Equalizer dimension	3x15
	w_{col}	4		Target dimension	3x5
JVDD	S_{max}	10000, 30000	GDLD code	$w_{row} = 100, \sigma = 200, d_x = 100$	
	T	2		$w_{row} = 100, \sigma = 180, d_x = 100$	
Iterative	I_{global}	5		$w_{row} = 200, \sigma = 300, d_x = 200$	
det-dec	I_{det}	150	LDPC code	$w_{col} = 4$	

Table 6.6: Variable parameters used in the performance comparison of the proposed equalization schemes

Code	N	512, 1024, 2048
	R	0.80 – 0.985 (in steps of 0.2)
Channel	TP	30, 34, 42, 50 nm

Table 6.7: MSE values of various equalizers

	$TP = 30$ nm	$TP = 34$ nm	$TP = 42$ nm	$TP = 50$ nm
SMR	0.3086	0.2738	0.2270	0.2021
Symmetric TDMR	0.3088	0.2726	0.2262	0.2052
2D TDMR	0.2881	0.2599	0.2202	0.1980

The plots of Figure 6.9 also suggest that for channels employing 1D detection, the proposed SMR equalizer and Symmetric TDMR equalizer performs better than the conventional PMR equalizer. The SMR equalizer also displayed a slight performance gain over the Symmetric TDMR equalizer. Despite the gain, the SMR equalizer has a throughput three times lesser than the Symmetric TDMR equalizer as it makes use of read-back signals from three tracks of data to make an estimate of the information sequence on the center track. On the contrary, the Symmetric TDMR equalizer uses read-back waveforms from three data tracks to simultaneously decipher information written on all three tracks.

Comparing the performance curves of each equalizer set-up when integrated with the JVDD instead of the state-of-the-art iterative detector-decoder (dotted curves in Figure 6.9), the conventional PMR and proposed SMR, Symmetric TDMR and TDMR equalizers all displayed a better performance when integrated with the JVDD instead of the iterative detector-decoder on the receiver end. This is only valid when the SNR of the channel is good (ie. $TP \geq 34$ nm). For example, in Figure 6.9(a) and 6.9(b), the JVDD outperforms the iterative detector-decoder by about 30 % and 10 % respectively for $TP = 50$ nm and $BL = 15$ nm. The JVDD however performs worse than the iterative detector-decoder at lower channel SNR values. In Figure 6.9(c) for example, a performance loss of about 20 % is shown for $TP = 30$ nm and $BL = 15$ nm for the Symmetric TDMR equalizer set-up.

From the simulations results obtained in this section, it can be concluded that the proposed TDMR equalizer, Symmetric TDMR equalizer and SMR equalizer all displayed a better performance than the conventional PMR equalizer. The proposed equalizers however require multiple reader heads to be used unlike the conventional PMR equalizer that requires only a

single reader head. Various trade-offs has to be weighed in choosing the best proposed equalizer for high density recording. The TDMR equalizer showed the best performance out of all the proposed equalizer designs, at the expense of higher complexity associated with 2D detection. The SMR equalizer displayed a slight performance gain over the Symmetric TDMR equalizer at the expense of lower throughput. These required trade-offs are specific to the equalizers and are independent on the choice of detector and decoder component. When the integration of each of the proposed equalizers with a detector and decoder component is considered, the JVDD will give the overall system a better performance than the state-of-the-art iterative detector-decoder when the SNR of the channel is good. Otherwise, the iterative detector-decoder would be an ideal choice.

Figure 6.9: FER plots comparing the performance of the (a) conventional PMR equalizer with the proposed (b) SMR, (c) Symmetric TDMR and (d) TDMR equalizers integrated with the iterative detector/decoder and JVDD at $N =$ (i) 1024 and (ii) 2048.

6.4 Summary

Waveforms gathered from a read-back process of a magnetic recording channel are typically corrupted with ISI in the down-track direction and ITI in the cross-track direction. In the present PMR channel, the effects of ITI are negligible as the spacing between adjacent tracks is large. As we shift towards a higher density recording channel such as SMR and TDMR, tracks are written such that they overlap with a segment of a previously written track. While this enables a higher track density, such writing mechanism causes severe ITI degradation. Advances in equalization

schemes are required to manage the ITI such that the reliability of the channel is not compromised with higher track densities.

The SMR architecture that has been widely proposed in the Literature employs the conventional read-back process that is currently used for PMR. In other words, it uses a single reader head to read information on one track of data and employs 1D equalization and 1D detection on its receiver end to estimate the bits written on that track. When the read-back and signal processing is carried out in this manner, the effects of ITI on the signal are ignored. A better performance can be reaped if the ITI distortions are tackled. This chapter thus proposes the use of multiple reader heads to gather a 2D read-back signal from an SMR channel. A novel equalizer that has been coined the “SMR equalizer” is proposed such that the ITI picked up by the multiple readers are equalized away for the center data track, leaving ISI on the center track as the only interference that can be cancelled out with a 1D detector.

The TDMR channel adopts the same writing mechanism as the SMR channel. The TDMR channel however generates a 2D read-back signal by reading multiple tracks of data at once using an array of multiple reader heads. At present, no 2D equalization schemes exist for TDMR. As such, the second part of this thesis proposes novel 2D equalization schemes which manage the ITI distortions in a TDMR channel in various ways. A 2D equalizer that has been coined the “Symmetric TDMR equalizer” was proposed such that it equalizes away the effect of ITI in the cross-track direction by reshaping the 2D channel response into multiple 1D target responses that are all identical. The interference of the TDMR channel is thus treated as ISI in the down-track direction of each data track. Multiple 1D detectors can then be used to cancel the effects of ISI on each track independently. The second proposed equalizer was coined the “Asymmetric TDMR equalizer”. It shares the same objective as the Symmetric TDMR equalizer but it is designed with multiple 1D target responses that are not identical instead. The third proposed equalizer was coined the “TDMR equalizer” where the ITI and ISI interferences are treated as a 2D interference. In other words, the TDMR equalizer reshapes the 2D channel response into a much shorter 2D target response and a 2D detector is employed to simultaneously undo the effect of ITI and ISI.

Each of the proposed equalizer designs was implemented and their performances were benchmarked against the conventional 1D GPR equalizer used in PMR (identified as PMR equalizer in this chapter). The GFP channel model developed in Chapter 3 was employed in the experiments conducted where multiple tracks of random data are written at various TP and BL and a triple reader head array is used to gather read-back signals. The RP and RO of the reader heads result in varying performance outcome of the proposed equalizers. As such, a study was first carried out to identify the optimum RP and RO for the equalizer. From the study, the proposed SMR equalizer showed less sensitivity to RP and RO whereas the remaining proposed equalizers can only tolerate small variations in RP and RO of the readers. This was attributed to the fact that the SMR equalizer characterizes a target for the center data track only, ignoring the responses of the other tracks adjacent to the center tracks. The proposed Symmetric TDMR equalizer and TDMR equalizer on the other optimizes the MSE of the error signal over all data tracks. On the receiver end, each of the equalizers were integrated with the state-of-the-art iterative detector as well as the JVDD (proposed in Chapter 4) in a separate set-up.

From the extensive simulations carried out, results showed that the proposed TDMR equalizer, Symmetric TDMR equalizer and SMR equalizer all outperformed the conventional PMR equalizer. The proposed equalizers however require multiple reader heads unlike the PMR equalizer which operates with just a single reader. The cost associated with multiple reader heads and the complexity of the design should be weighed when considering the proposed equalizers as alternatives to the conventional PMR equalizer. Out of all the proposed equalizer designs, the TDMR equalizer performed the best, at the expense of higher complexity associated with 2D detection. The SMR equalizer displayed a slight performance gain over the Symmetric TDMR equalizer at the expense of lower throughput. These trade-offs are specific to the equalizers and are independent of the choice of detector and decoder component. When a complete integration of each of the proposed equalizers with a detector and decoder component is considered, the JVDD proposed in Chapter 4 resulted in a better performance than the state-of-the-art iterative detector-decoder when the channel is less noisy. The iterative detector-decoder is an ideal choice for channels with poor SNR.

Chapter 7

Conclusion and Future Work

This chapter summarizes the major contributions in this thesis as well as the conclusions drawn from the work undertaken in this thesis. An overview of the future work to be embarked on is also included in this chapter.

7.1 Conclusion

The Grain Flipping Probability (GPF) channel model was proposed in Chapter 3 as an alternative to the highly accurate but time consuming micromagnetic channel model commonly used by magnetic researchers. The GFP model which consists of a multi-dimensional look-up table (LUT) populated from the micromagnetic model showed its ability to reproduce grain magnetizations about 100 times faster than the micromagnetic model on average. While the conventional micromagnetic model is useful in accurately studying the magnetic phenomenon associated with the read and write process of an isolated magnetic recording channel, the proposed GFP model is useful for studying a channel integrated with signal processing components from a system's perspective. The development of this model also serves as a pre-requisite for developing new signal processing techniques in other parts of the thesis. In Chapter 4, the GFP model is used in two novel system level studies for the emerging Shingled Magnetic Recording (SMR) and Heat Assisted Magnetic Recording (HAMR) channel.

The performance of the SMR channel integrated with the state-of-the-art iterative detector-decoder and Generalized Partial Response (GPR) equalizer showed a high sensitivity to variations in the track pitch, TP , and bit length, BL , of the channel. This was attributed to the inter track interference (ITI) and inter symbol interference (ISI). A large TP and BL was shown to give a good channel Signal-to-Noise Ratio (SNR) but low media density. On the contrary a small TP and BL give a poorer SNR but higher media density. As the operating code rate, R_{op} , in the high SNR region is higher than that in the low SNR region, the storage density at large TP and BL tends to be higher than that at small TP and BL . From the analysis of the storage densities across a range of TP and BL values, an operating region where $TP = 28 - 30$ nm and $BL = 12 - 13$ nm is proposed in this thesis for the SMR channel under the simulated conditions.

The HAMR channel showed sensitivity in performance to three key parameters of its thermal hot-spot. Firstly, the peak temperature of the hot-spot, T_{peak} , determines the ease of grains flipping and the ideal operating T_{peak} value was found to be close to the Curie temperature of the media. Secondly, the width of the thermal hot-spot, T_{sig} , was shown to affect the number of grains that are magnetized within the vicinity of the applied thermal field and concurrently affects the bit-aspect-ratio (BAR) of the bits written. The interaction between the effects of varying T_{sig} , is fairly complex but in general, a large T_{sig} results in better channel performance. Lastly, the horizontal offset distance between the write field and thermal field, dis_x , affects the alignment of the gradients of both fields. An optimal dis_x was identified for a HAMR channel under the simulated conditions and was found to correspond to the down-track location where the gradient of both the thermal and magnetic field align maximally.

The Joint Viterbi Detector Decoder (JVDD) was proposed in Chapter 4 as an alternative to the sub-optimal state-of-the art iterative detector-decoder. The JVDD initially showed a high computational complexity that would not be feasible in a practical system. A class of JVDD codes was thus proposed specifically for the JVDD. They consists of the Gaussian Distribution Linear Diagonal (GDLD) and proportionate row weight codes. The proposed codes effectively reduced the computational complexity of JVDD by about 70% on average. The minimum distance, d_{min} , and column weight distribution, w_{col} , of the codes are however sensitive to the

design parameters chosen for the code. As such, the proposed codes were analytically optimized using the Response Surface Methodology (RSM).

With the optimized codes, the JVDD showed a performance gain of up to 2 dB over the state-of-the-art iterative detector for an ISI channel as well as a magnetic recording channel that uses the GFP model. The complexity of the JVDD and the iterative detector-decoder measured in terms of computation time were found to be identical at short codeword length, N . The computational time of the JVDD however grew significantly larger as N increases and SNR decreases. In the worst case scenario, a difference of up to 10 times was shown. The proposed JVDD is thus suitable for practical applications that operate at short or intermediate N and good channel SNR . The JVDD was also implemented in a non-binary form coined as 2D-JVDD for application to 2D channels. The study indicated a performance gain over the conventional 2D iterative detector-decoder at short N value of < 128 . The exceptionally high complexity of the 2D-JVDD hinders it from being potentially applied in a real application.

The Joint Factor Graph Detector Decoder (JFGDD) was proposed in Chapter 5 as a second alternative to the state-of-the-art iterative detector-decoder. The JFGDD displayed a performance loss of about 1 dB (at its saturation point) over the iterative detector-decoder. The presence of short length-4 cycles in the factor graph of JFGDD causes a significant deterioration in its performance. Two measures were proposed to mitigate or completely remove the short cycles in the channel factor graphs. Firstly, a non-binary implementation of the algorithm was proposed where the grouping of symbols instead of bits resulted in lesser edges and therefore lesser cycles in the graph. The non-binary JFGDD showed an improvement in performance over the binary JFGDD. It attains an identical performance as the iterative detector-decoder over an ISI channel and a magnetic recording channel with the number of detector iterations, $I_{det} = 6$. Despite not outperforming the iterative system like the JVDD, the non-binary JFGDD has an advantage over the iterative detector-decoder in hardware implementation as it is parallelizable. Secondly, the design of constrained GPR targets was proposed and showed a marginal improvement in error rates for JFGDD. An overall performance loss of about 1.25 dB to the iterative detector-decoder was discovered. Despite theoretical predictions that suggest

this technique could enhance the performance of JFGDD, the contrary was observed in the simulations. This was attributed to the fact that in practice, although an equalizer can be effectively used in reducing short cycles in a channel graph, it results in coloration of noise which causes performance deterioration. Comparing the JFGDD and the JVDD schemes, the JVDD proves to be more promising and has a higher potential of replacing the conventional iterative detector-decoder.

In Chapter 6, four novel equalization schemes were proposed for 2D magnetic recording channels to counter the effects of ITI and ISI introduced by such channels. They consist of the SMR equalizer, Two-Dimensional Magnetic Recording (TDMR) equalizer, Symmetric TDMR equalizer and Asymmetric TDMR equalizer. All four proposed schemes displayed a better performance than the conventional Perpendicular Magnetic Recording (PMR) equalizer used today. In the best case scenario, an approximate 9% performance gain was shown. The four proposed equalizers however require multiple reader heads unlike the PMR equalizer which operates with just a single reader. The cost associated with multiple reader heads and the complexity of the design should be weighed when considering the proposed equalizers as alternatives to the conventional PMR equalizer. Out of all the proposed equalizer designs, the TDMR equalizer performed best, at the expense of higher complexity associated with 2D detection. The SMR equalizer displayed a slight performance gain of 0.25 dB over the Symmetric TDMR equalizer at the expense of lower throughput. The difference in performance displayed was attributed to the Mean Square Error (MSE) obtained in each equalizer design. When a complete integration of each of the proposed equalizers with a detector and decoder component is considered, the combination of the proposed equalizers with the JVDD resulted in a better system performance than when combined with the state-of-the-art iterative detector-decoder over a channel with good SNR. The integration of the proposed equalizers with the iterative detector-decoder is an ideal choice for channels with poor SNR.

7.2 Future Work

In view of the scope of the work that is not fully explored in this thesis, this section proposes several relevant ideas that can be considered as a sequel for future research.

7.2.1 Mitigation of Colored Noise

In Chapters 4 and 5, the performance of the proposed JVDD and JFGDD were first compared with that of a state-of-the-art iterative detector-decoder over an ISI channel and subsequently a PMR channel using the GFP model. The JVDD and JFGDD were both implemented based on the fact that the noise in the channel is white. The noise in a PMR channel is, however in actual fact, colored. The integration of a Data Dependent Noise Predictor (DDNP) with the JVDD and JFGDD is thus proposed as a recommended future work to effectively whiten the colored and data dependent noise in a realistic magnetic recording channel. This may enable the JVDD and JFGDD to obtain an improved performance over colored noisy channels. In a recently published work [83], the authors reported a 4 dB performance improvement over conventional systems when DDNP is integrated with their proposed equalizer and detector systems. Similarly, in [83], performance loss was mitigated for an SMR system through the implementation of DDNP in forward and backward recursions (bi-directional DDNP) of the conventional BCJR detection algorithm.

The DDNP can be implemented as an auto-regressive channel model as illustrated in Figure 7.1. Noise whitening can be performed by passing the read-back signal through the inverse auto-regressive model [84][85]. With the integration of DDNP, the computation of branch metric in the JVDD and the number of states in its channel trellis requires modification. Similarly, for the JFGDD, the connections between nodes in the channel factor graph and the computation of messages passed along the nodes of the JFGDD have to be implemented in a different manner. The required modifications for both detection schemes and the evaluation of their noise whitening performance are recommended as a future work of this thesis.

Figure 7.1 Auto-regressive channel model for noise whitening [85].

7.2.2 Hardware Implementation

The channel model and algorithms proposed earlier in this thesis were implemented and simulated using clusters of servers. While computer simulations gives a good first indication on error rates and complexity, an extension of the work to hardware simulation environment would be a useful next stage of analysis. For example, the implementation of the proposed JVDD and JFGDD algorithms on hardware such as field-programmable gate array (FPGA) or graphics processing unit (GPU) would give a more accurate benchmark of computational time and complexity of the proposed algorithms with the state-of-the-art iterative detector-decoder. An LDPC coded magnetic recording system can be evaluated down to a frame error rate (FER) of 10^{-8} through hardware implementation [86][87]. Such low level of error floor is not attainable when simulations are carried out using central processing unit (CPU). Hardware implementation and analysis of the ideas proposed in this thesis are useful for a more accurate assessment of the algorithms as a viable future signal processing scheme. Knowledge on hardware implementation is however categorized under a different field of expertise and research, and the cost of acquiring hardware resources to embark on the proposed future research work could be costly.

7.2.3 Joint Detection and Decoding with Non-Binary Codes

In Chapters 4 and 5, the JVDD and JFGDD algorithms were proposed for a binary coded ISI channel. The proposed joint detection and decoding schemes can also be extended to a non-binary coded ISI channel (1D and 2D) by introducing several key modifications that takes into account the encoding and decoding using a non-binary error correction code and the integration of a binary (for 1D ISI channel) or non-binary (for 2D ISI channel) detector with the non-binary decoder. The introduction of non-binary LDPC coded signals to various joint/iterative detection and decoding have resulted in significant performance improvement in a few notable recent works [87][88][89]. The suggested future work extension can be especially useful when the algorithms are applied to a 2D ISI channel such as a TDMR or SMR channel. This would, however, come at the expense of increased computational complexity. For example, the application of non-binary LDPC codes to JFGDD would result in a super-factor graph with considerably greater number of variable, channel and parity nodes, each of which would be involved in the computation of symbol probabilities instead of bit probabilities (as in the case for binary LDPC code). Novel measures for managing the trade-off between increased complexity and better performance is a possible future area of research that is closely linked to the exploration of detection algorithms with non-binary codes suggested here.

Publications and Patent

Publications:

- [1] S. S. Shafi'ee, K. S. Chan, E. M. Rachid, H. T. Wang and E. Kwaku, "Application of the grain flipping probability model to heat assisted magnetic recording", 56th Conference on Magnetism and Magnetic Material, Arizona, USA, pp. CF-11, October–November 2011.
- [2] S. S. Shafi'ee, K. S. Chan, E. M. Rachid, H. T. Wang and E. Kwaku, "Application of the grain flipping probability model to heat assisted magnetic recording", J. Appl. Phys., vol. 111, no. 7, pp. 07B714 - 07B714-3, April 2011.
- [3] S. S. Shafi'ee, K. S. Chan, Y. L. Guan and E. M. Rachid, "Application and Optimization of Factor Graph-Based Detector on 1D ISI Magnetic Recording Channel", Asia-Pacific Magnetic Recording Conference 2012, Singapore, 31 October – 2 November 2012.
- [4] S. S. Shafi'ee, K. S. Chan, Y. L. Guan and E. M. Rachid, "Application and Optimization of Factor Graph-Based Detector on 1D ISI Magnetic Recording Channel", IEEE Trans. Magn., vol. 49, no. 6, pp. 2500 - 2503, June 2013.
- [5] S. S. Shafi'ee, K. S. Chan, Y. L. Guan and E. M. Rachid, "Joint Viterbi Detector Decoder for a Grain Flipping Probability Model", 24th Magnetic Recording Conference (TMRC 2013), Tokyo, Japan, 20 -22 August 2013.
- [6] S. S. Shafi'ee, K. S. Chan and Y. L. Guan, "Novel Joint Detection Schemes for TDMR

- Channels”, 25th The Magnetic Recording Conference (TMRC 2014), California, USA, 11 - 13 August 2014.
- [7] S. S. Shafi’ee, K. S. Chan and Y. L. Guan, “Novel Joint Detection Schemes for TDMR Channels”, IEEE Trans. on Magn., vol. 51, no. 4, pp. 0018-9464, April 2015.
- [8] S. S. Shafi’ee, K. S. Chan and Y. L. Guan, “A Performance Study of Joint Viterbi Detector Decoder (JVDD) with GDLD codes”, International Conference on Computational Intelligence, Computing and Signal Processing (CICSP), Hong Kong, China, pp. 339 – 343, 19 - 20 October 2014.
- [9] K. S. Chan, S. S. Shafi’ee and Y. L. Guan, “Optimal Joint Viterbi Detector Decoder (JVDD) over AWGN/ISI Channels”, International Conference on Computing, Networking and Communications (ICNC) 2014, Hawaii, USA, pp. 282 - 286, 3 – 6 February 2014.
- [10] A. James, K. S. Chan and S. S. Shafi’ee, “Sparse Construction of Joint Viterbi Detector Decoder (JVDD) Codes”, The Eleventh Advanced International Conference on Telecommunications (AICT 2015), Brussels, Belgium, pp. 29 – 33, 21 -26 June 2015.
- [11] S. S. Shafi’ee, K. S. Chan and W. L. Goh, “Optimization of Codes for the Joint Viterbi Detector Decoder”, International Conference on Computing, Networking and Communications (ICNC) 2014, pp. 282 - 286, Hawaii, USA, 3 – 6 February 2014.

Patent:

- [1] Joint detector/ decoder devices and joint detection/ decoding methods, 2013077953, Chan Kheong Sann, Elidrissi Moulay Rachid and Sari Shafidah Shafi’ee.

References

- [1] C. McLellan, "Storage in 2014: An Overview", accessed on 22 December 2014, available at: <http://www.zdnet.com/article/storage-in-2014-an-overview>.
- [2] J. Gantz and D. Reinsel, "THE DIGITAL UNIVERSE IN 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East", accessed on 22 December 2014, available at: <http://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-200.pdf>.
- [3] A. Baxter, "SSD vs HDD", accessed on 22 December 2014, available at: http://www.storagereview.com/ssd_vs_hdd.
- [4] R. E. Fontana, S. R. Hetzler and G. Decad, "Technology Roadmap Comparisons for TAPE, HDD, and NAND Flash: Implications for Data Storage Applications", IEEE Trans. Magn., vol. 48, no. 5, pp. 1692 - 1696, May 2012.
- [5] S. Takenoiri, S. Matsuo and T. Fujihira, "Magnetic Recording Media: Technical Trends and Future Outlook", Fuji Electric Review, vol. 57, no. 2, pp. 32-36, 2011.
- [6] M. L. Plumer, J. V. Ek and W. C. Cain, "New Paradigms in Magnetic Recording", Physics in Canada, vol. 67, no. 1, pp. 25-29, January 2011.
- [7] R. Galbraith and T. Oenning, "Iterative Detection Read Channel Technology in Hard Disk Drives", HGST White paper, November 2008.
- [8] A. Kavcic and A. Patapoutian, "The read channel," Proceedings of the IEEE, vol. 96, no.

- 11, pp. 1761-1774, November 2008.
- [9] P. Kovintavewat, I. Ozgunes, E. Kurtas, J. R. Barry, and S. W. McLaughlin, "Generalized partial-response targets for perpendicular recording with jitter noise," *IEEE Trans. Magn.*, vol. 38, no. 5, pp. 2340-2342, September 2002.
 - [10] M. H. Kryder, "Future trends in magnetic storage technology", *Digest of Technical Papers Joint NAPMRC 2003*, California, USA, 6 – 8 January 2003.
 - [11] S. N. Piramanayagam and T. C. Chong, "Developments in Data Storage – Materials Perspective", A John Wiley & Sons Inc. Publication, 2012.
 - [12] R. Wood, R. Galbraith and J. Coker, "2-D Magnetic Recording: Progress and Evolution", *IEEE Trans. Magn.*, vol. 51, no. 4, pp. 0018-9464, April 2015.
 - [13] X. Wang, K. Gao, H. Zhou and A. Itagi, "HAMR Recording Limitations and Extendibility", *IEEE Trans. Magn.*, vol. 49, no.2, pp. 686-692, February 2013.
 - [14] Y. Wang, M. F. Erden and R. H. Victora, "Study of Two-Dimensional Magnetic Recording System Including Micromagnetic Writer", *IEEE Trans. Magn.*, vol. 50, no. 11, pp. 0018-9464, November 2014.
 - [15] J. Park, B. Lengsfeld, R. Galbraith, R. Wood and F. Sidi, "Optimization of Magnetic Read Widths in Two-Dimensional Magnetic Recording Based on Micromagnetic Simulations", *IEEE Trans. Magn.*, vol. 51, no. 11, pp. 0018-9464, July 2015.
 - [16] M. R. Elidrissi, S. Bastrukov, H. T. Wang and J. Y. Khoo, "Micromagnetic Dynamics of Single-Domain Grain in Thin-Film Magnetic Recording Media", *IEEE Trans. Magn.*, vol. 49, no. 6, pp. 2610 – 2613, June 2013.
 - [17] Y. Kanai, S. J. Greaves, K. Yoshida and H. Muraoka, "Micromagnetic model analysis of high frequency heat-assisted magnetic recording", *J. Appl. Phys.*, vol. 117, no. 17, pp. 17C506, 17 February 2015.

- [18] Wu Tong, "Signal Processing for Bit-Patterned Media Recording", Ph.D. dissertation, Electrical and Computer Eng. Dept., National University of Singapore, 2014.
- [19] M. Mehrnough, B. J. Belzer, K. Sivakumar and R. Wood, "Signal processing for two dimensional magnetic recording using Voronoi model averaged statistics", 49th Annual Conference on Information Sciences and Systems (CISS), Baltimore, Maryland, pp. 1-6, 18 -20 March 2015.
- [20] Z. J. Liu, Z. M. Yuan, C. L. Ong, H. Guchang and S. Ang, "Analysis of magnetic noises in two-dimensional magnetic recording readers", J. Appl. Phys., vol. 117, no. 17, pp. 17A911, 27 March 2015.
- [21] M. Carosino, J. Y. Y. Chen, M. Mehrnough, B. J. Belzer, K. Sivakumar, R. Wood, J. Murray and P. Wettin, "Iterative Detection and Decoding for TDMR With 2-D Intersymbol Interference Using the Four-Rectangular-Grain Model", IEEE Trans. Magn., vol. 51, no. 7, pp. 0018-9464, February 2015.
- [22] S. M. Khatami and B. Vasic, "Detection for two-dimensional magnetic recording systems", International Conference on Computing, Networking and Communications (ICNC) 2013, San Diego, California, pp. 535 – 539, 28-31 January 2013.
- [23] W. Tong and M. A. Armand, "Joint and Separate Detection-Decoding on BPMP Channels", IEEE Trans. Magn., vol. 49, no.7, pp. 3779 - 3782, July 2013.
- [24] T. Sopon, L. M. M. Myint, P. Supnithi and K. Vichienchom, "Modified Graph-Based Detection Methods for Two-Dimensional Interference Channels", IEEE Trans. Magn., vol. 48, no. 11, pp. 4618 – 4621, November 2012.
- [25] R. Radhakrishnan and B. Vasic, "Joint Message-Passing Symbol-Decoding of LDPC Coded Signals over Partial-Response Channels", Proceedings of the 2009 IEEE international conference on Communications (ICC), Dresden, Germany, pp. 2650-2654, 14 – 18 June 2009.

- [26] G. Colavolpe and G. Germe, "On the Application of Factor Graphs and the Sum-Product Algorithm to ISI Channels", IEEE Trans. Comm., vol. 53, no. 5, pp. 818-825, May 2005.
- [27] N. M. F. Garrido and F. A. B. Cercas, "Equalization in hard disk drive read channels", 21st International Conference on Telecommunications (ICT) 2014, Lisbon, pp. 390 - 394 , 4-7 May 2014.
- [28] M. Yamashita, H. Osawa, Y. Okamoto, Y. Nakamura, Y. Suzuki, K. Miura and H. Muraoka, "Read/Write Channel Modeling and Two-Dimensional Neural Network Equalization for Two-Dimensional Magnetic Recording", IEEE Trans. Magn., vol. 47, no. 10, pp. 3558 - 3561, October 2011.
- [29] H. N. Bertram, H. Zhou, and R. Gustafson, "Signal to noise ratio scaling and density limit estimates in longitudinal magnetic recording," IEEE Trans. Magn., vol. 34, no. 4, pp. 1845-1847, Jul. 1998.
- [30] H. N. Bertram, "Theory of Magnetic Recording", Cambridge, U.K.: Cambridge University Press, 17 March 1994.
- [31] K. Sakhrat and L. Dmitri, "Perpendicular Magnetic Recording", Kluwer Academic Publisher, 2004th edition, 2004.
- [32] HGST, "Perpendicular Magnetic Recording Technology", accessed on 1 December 2014, available at: https://www.hgst.com/sites/default/files/resources/PMR_white_paper_final.pdf.
- [33] K. H. J. Buschow, "Handbook of Magnetic Materials", vol. 21, Newnes, January 2013.
- [34] H. J. Richter and S. D. Harkness, "Media for Magnetic Recording Beyond 100Gbit/in²", Materials Research Society Bulletin, vol. 31, no. 5, pp. 384-388, May 2006.
- [35] R. H. Victora, X. Jianhua and M. Patwari, "Areal density limits for perpendicular magnetic recording", IEEE Trans. Magn., vol. 38, no. 5, pp. 1886 - 1891, September 2002.

- [36] Seagate, "Strategic Update", accessed on 20 February 2015, available at: <http://www.seagate.com/www-content/about/investors/en-us/docs/2014-strategic-update-presentation.pdf>.
- [37] H. J. Richter, N. Supper, B. Wilson and B. D. Terris, "Magnetic Recording Near the Grain Size Limit", IEEE Trans. Magn., vol. 48, no. 11, pp. 3887 - 3890, November 2012.
- [38] J. P. Wang, "FePt Magnetic Nanoparticles and Their Assembly for Future Magnetic Media", IEEE Trans. Magn., vol. 96, no. 11, pp. 1847 – 1863, November 2008.
- [39] J. G. Zhu, H. Li, "Understanding Signal and Noise in Heat Assisted Magnetic Recording", IEEE Trans. Magn., vol. 49, no.2, pp. 765-772, February 2013.
- [40] M. H. Kryder, E. C. Gage, T. W. McDaniel, W. A. Challener, R. E. Rottmayer, Ganping Ju, Hsia Yiao-Tee and M. F. Erden, "Heat Assisted Magnetic Recording", Proc. of IEEE, vol. 96, no. 11, pp. 1810-1835, November 2008.
- [41] G. Bertero, M. Alex, B. Valcu and R. Eaton, "HAMR recording challenges at high linear densities", IEEE Magnetics Conference (INTERMAG) 2015, Beijing, 11 – 15 May 2015.
- [42] G. Parker, O. Mosendz, E. Champion, B. Stipe, X. Wang, T. Klemmer, G. Ju and A. Ajan, "A HAMR Media Technology Roadmap to an Areal Density of 4 Tb/in²", IEEE Trans. Magn., vol. 50, no. 1, pp. 0018-9464, 6 September 2013.
- [43] S. J. Greaves, Y. Kanai and H. Muraoka, "Thermally Assisted Magnetic Recording at 4 Tbit/in²", IEEE Trans. Magn., vol. 49, no.6, pp. 2665 - 2670, June 2013.
- [44] Brian Beeler, "What is Shingled Magnetic Recording (SMR)?", accessed on 21 January 2016, available at: http://www.storagereview.com/what_is_shingled_magnetic_recording_smr.
- [45] R. Galbraith, W. Hanson, B. Lengsfeld, T. Oenning, J. Park and M. Salo, "Noise Characterization in High-Areal Density Perpendicular and Shingled Magnetic Recording",

- IEEE Trans. Magn., vol. 50, no. 3, pp. 0018-9464, March 2014.
- [46] M. Salo, T. Olson, R. Galbraith, R. Brockie, B. Lengsfeld, H. Katada, and Y. Nishida, "The Structure of Shingled Magnetic Recording Tracks", IEEE Trans. Magn., vol. 50, no. 3, pp. 0018-9464, March 2014.
 - [47] H. Nobuhara, Y. Okamoto, Y. Nakamura, K. Takada, M. Yamashita, H. Osawa and H. Muraoka, "Influence of Writing ITI Effects in Shingled Magnetic Recording", IEEE Trans. Magn., vol. 49, no. 7, pp. 3814 - 3817, July 2013.
 - [48] E. Hwang, R. Negi, and B. V. K. Vijaya Kumar, "Signal Processing for Near 10 Tbit/in² Density in Two-Dimensional Magnetic Recording (TDMR)", IEEE Trans. Magn., vol. 46, no. 6, pp. 1813-1816, June 2010.
 - [49] M. Khatami and B. Vasic, "Constrained coding and detection for TDMR using generalized belief propagation", 2014 IEEE International Conference on Communications (ICC), Sydney, Australia, pp. 3889 – 3895, 10-14 June 2014.
 - [50] G. Mathew, E. Hwang, J. Park and G. Garfunkel, "Capacity Advantage of Array-Reader-Based Magnetic Recording (ARMR) for Next Generation Hard Disk Drives", IEEE Trans. Magn., vol. 50, no.3, pp. 330-337, March 2014.
 - [51] K. S. Chan, "A review study on the benefits of multi-reader heads in SMR and TDMR systems", International Symposium on Information Theory and its Applications (ISITA) 2014, Melbourne, Australia, pp. 684 - 688, 26-29 October 2014.
 - [52] R. Radhakrishnan, N. Varnica, and M. Oberg, "Estimation of areal density gains of TDMR system with 2D detector", International Symposium on Information Theory and its Applications (ISITA) 2014, Melbourne, Australia, pp. 674 -678, 26-29 October 2014.
 - [53] B. Vasic and E. M. Kurtas, "Coding and Signal Processing for Magnetic Recording Systems", CRC Press, 2005.

- [54] R.G. Gallager, "Low-Density Parity Check Codes", Cambridge, MA: MIT Press, 1963.
- [55] H. Song and B. V. K. Vijaya Kumar, "Low-density parity-check codes for partial response channels," IEEE Signal Processing Magazine, vol. 21, no. 1, pp. 56-66, January 2004.
- [56] H. Zhong, T. Zhang and E. F. Haratsch, "Quasi-Cyclic LDPC Codes for the Magnetic Recording Channel: Code Design and VLSI Implementation", IEEE Trans. Magn., vol. 43, no. 3, pp. 1118 – 1123, March 2007.
- [57] H. Zhong, T. Zhang and E. F. Haratsch, "High-rate quasi-cyclic LDPC codes for magnetic recording channel with low error floor", Proceeding of IEEE ISCAS 2006, March 2006.
- [58] R. Koetter, A. C. Singer and M. Tüchler, "Turbo Equalization", IEEE Signal Processing Magazine, pp. 67-80, January 2004.
- [59] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal Decoding of Linear Codes for minimizing symbol error rate", IEEE Trans. Inf. Theory, vol. IT-20(2), pp.284-287, March 1974.
- [60] A. Kavcic and J. M. F. Moura, "The Viterbi algorithm and Markov noise memory", IEEE Trans. Inf. Theory, vol. 46, no. 1, pp. 291-301, January 2000.
- [61] F. R. Kschichang, B. J. Frey and H. A. Loeliger, "Factor Graphs and the Sum-Product Algorithm", IEEE Trans. Inf. Theory, vol. 47, no. 2, pp. 498-519, February 2001.
- [62] Vapnyarskii, "Lagrange multipliers", Hazewinkel, Michiel, Encyclopedia of Mathematics, Springer, 2001.
- [63] J. Moon and W. Zeng, "Equalization for maximum likelihood detectors," IEEE Trans. Magn., vol. 31, no. 2, pp. 1083-1088, March 1995.
- [64] R. Venkataramani and S. Sankaranarayanan, "Optimal Channel Shortening Equalization for MIMO ISI Channels", IEEE Global Telecommunications Conference 2008, pp.1-5, 30

November - 4 December 2008.

- [65] M. Lakshmanan, "The Fascinating World of Landau-Lifshitz-Gilbert Equation: An Overview", *Phil. Trans. R. Soc. A*, vol. 369, pp. 1280-1300, 2011.
- [66] J. T. Thiele, K.R. Coffey, M. F. Toney, J. A. Hedstrom and A. J. Kellock, "Temperature Dependent Magnetic Properties of Highly Chemically Ordered Fe₅₅-xNiPt₄₅L₁₀ Films", *Jour. App. Phy.*, vol. 91, no. 10, pp. 6595-6600, 15 May 2002.
- [67] G. Wiederrecht, "Handbook of Nanoscale Optics and Electronics", Academic Press Publisher, 25 May 2010.
- [68] H. J. Richter, C. C. Poon, G. Parker, M. Staffaroni, O. Mosendz, R. Zakai, and B.C. Stipe, "Direct Measurement of the Thermal Gradient in Heat Assisted Magnetic Recording", *IEEE Trans. Magn.*, vol. 49, no. 10, pp. 5378 – 538, May 2013.
- [69] K. Chugg, A. Anastasopoulos and X. Chen, "Iterative Detection, Adaptivity, Complexity Reduction and Applications", Kluwer Academic Publishers, 2001.
- [70] J. Hagenauer and P. Hoeher, "A Viterbi algorithm with soft-decision outputs and its applications", *IEEE Global Telecomm. Conf.*, San Diego, USA, pp. 1680-1686, 1989.
- [71] A. Houghton, "Error Coding for Engineers", Kluwer Academic Publishers, 2001.
- [72] M. Y. Rhee, "Error Correcting Coding Theory", Mc-Graw Hill International Editions, 1989.
- [73] G. D. Forney, "Maximum likelihood sequence estimation of digital sequences in the presence of inter-symbol interference", *IEEE Trans. Inf. Theory*, vol. IT-18, pp. 363-378, May 1972.
- [74] J. G. Proakis, "Digital Communications Third Edition", Mc-Graw Hill International Edition, Electrical Engineering Series, 1995.
- [75] R. G. Gallager, "MIT Handout: Detection", accessed on 2 February 2015, available at:

<http://www.nari.ee.ethz.ch/teaching/wirelessIT/handouts/detection.pdf>.

- [76] R. H. Myers, C. Douglas and M. Christine, "Process and Product Optimization Using Designed Experiments", John Wiley & Sons, 2009.
- [77] L. Eriksson, "Design of Experiments: Principles and Applications", Umetrics Academy, 2008.
- [78] E. M. Kurtas, A. V. Kuznetsov, I. Djurdjevic, "System perspectives for the application of structured LDPC codes to data storage devices," IEEE Trans. Magn., vol.42, no.2, pp. 200-207, February 2006.
- [79] T. K. Moon, "Error Correction Coding: Mathematical Methods and Algorithms", A John Wiley & Sons, Inc., Publication, May 2005.
- [80] T. M. Cover and J. A. Thomas, "Elements of Information Theory", Wiley, 1991.
- [81] Wikibooks, "Linear Algebra/Gauss' Method", accessed on 18 April 2015, available at: https://en.wikibooks.org/wiki/Linear_Algebra/Gauss%27_Method.
- [82] C. K. Matcha and S. G. Srinivasa, "Generalized Partial Response Equalization and Data-Dependent Noise Predictive Signal Detection Over Media Models for TDMR", IEEE Trans. Magn., Vol. 51, no.10, pp. 0018-9464, June 2015.
- [83] Y. Ng, S. Zhang, K. Cai, Z. Qin, C. L. Ong, S. Ang, and Z. Yuan, "Performance of perpendicular magnetic recording with track squeeze using bidirectional pattern-dependent noise prediction detector", J. Appl. Phys., vol. 177, no. 17, pp. 17E315, 18 March 2015.
- [84] J. Moon and J. Park, "Pattern-Dependent Noise Prediction in Signal-Dependent Noise", IEEE Journal on Selected Areas in Communications, vol. 19, no.4, pp.730-743, April 2001.
- [85] A. Kavcic and A. Patapoutian, "A signal-dependent autoregressive channel model", IEEE Trans. Magn., vol. 35, no. 5, September 1999.

- [86] X. Hu, B. V. Kumar, L. Sun, and J. Xie, "FPGA-based LDPC Code Evaluation using an Advanced Magnetic Recording Channel Model", International Magnetism Conference (Intermag) 2006, San Diego, California, pp. 792, May 2006.
- [87] Y. Cai, S. Jen, K. Mai, B.V.K.V. Kumar, "Highly Parallel FPGA Emulation for LDPC Error Floor Characterization in Perpendicular Magnetic Recording Channel", IEEE Trans. Magn., vol. 45, no.10, pp. 3761 - 3764, October 2009.
- [88] S. Zhao, Z. Lu, X. Ma and B. Bai, "Joint Detection/Decoding Algorithms for Nonbinary LDPC Codes over ISI Channels", IET Communications, vol. 7, no. 14, p. 1522 – 1531, 24 September 2013.
- [89] A. Marinoni, P. Savazzi and P. Gamba, "Efficient detection and decoding of q-ary LDPC coded signals over partial response channels", EURASIP Journal on Wireless Communications and Networking 2013, pp. 2013:18, January 2013.