Signal Processing Using Short Word-Length

A Thesis Submitted in Fulfillment of the Requirements

for the Degree of Doctor of Philosophy

Amin Z. Sadik

School of Electrical and Computer Engineering

Science, Engineering and Technology Portfolio

RMIT University

July 2006

Declaration

I certify that except where due acknowledgement has been made, the work is that of the author alone; the work has not been submitted previously, in whole or in part, to qualify for any other academic award; the content of the thesis is the result of work which has been carried out since the official commencement date of the approved research program; and, any editorial work, paid or unpaid, carried out by a third party is acknowledged.

Amin Z. Sadik

July 2006

This dissertation is dedicated to my dearest wife Asma and my lovely daughters Samar, Sara, Tara, and Heifa

Acknowledgements

I would like to thank my family for their endless support and understanding.

Special thanks are to my supervisor, Associate Professor Zahir M. Hussain for his generosity and support during my PhD program.

I am also thankful to Associate Professor Peter O'Shea (from Queensland University of Technology, Brisbane) for his support.

Last but not least, I am grateful for the financial support of the Australian Research Council (ARC), which was funding my study and supporting my living expenses during my PhD candidature.

Publications and Awards

Below are the publications and the awards in conjunction with the author's PhD candidacy:

Journal Publications

1. A. Z. Sadik, Z. M. Hussain, and P. O'Shea, "An adaptive algorithm for ternary filtering," *IEE Electronics Letters*, vol. 42, issue 7, pp. 420-421, March 2006.
2. Amin Z. Sadik, Zahir M. Hussain, Xinghuo Yu, and Peter O'Shea, "An approach for stability analysis of a single-bit high-order digital sigma-delta modulator," *Elsevier Journal on Digital Signal Processing*, in press, 2007.
3. A. Z. Sadik, Z. M. Hussain, and P. O'Shea, "A Multiplierless DC-blocker for single-bit sigma-delta modulated signals," *EURASIP Journal on Applied Signal Processing*, in press, 2007.
4. A. Z. Sadik, Z. M. Hussain, and P. O'Shea, "DC cancellation in the single-bit domain using ternary filtering," *Elsevier Journal on Signal Processing*, Submitted, 2006.
5. Amin Z. Sadik, Zahir M. Hussain, Peter O'Shea, and Xinghuo Yu, "Approximate analysis for the dynamics of a third-order sigma-delta modulator," *Signal Processing Journal*, Submitted, 2006.
6. A. Z. Sadik and Z. M. Hussain,"A single-bit adaptive LMS-like filtering," *IEEE Transactions on Signal Processing*, Submitted, 2006.

Refereed Conference Publications

1. A. Z. Sadik, Z. M. Hussain, and P. O'Shea, "Structures for single-bit digital comb filtering," *IEEE Asia-Pacific Conference on Communications (APCC 2005)*, pp. 545-548, Perth, Oct. 2005.
2. A. Z. Sadik, Z. M. Hussain, and P. O'Shea, "A single-bit digital DC-blocker using ternary filtering," *IEEE TENCON 2005*, pp. 1793-1798, Melbourne, Nov. 2005.
3. A. Z. Sadik, Z. M. Hussain, and P. O'Shea, "Efficient structure for single-bit digital comb filters and resonators," *IEEE TENCON 2005*, pp. 2061-2065, Melbourne, Nov. 2005.
4. A. Z. Sadik, Z. M. Hussain, and P. O'Shea, "Adaptive LMS ternary filtering," *IEEE TENCON 2005*, pp. 2004-2006, Melbourne, Nov. 2005.
5. Amin Z. Sadik, Zahir M. Hussain, and Xinghuo Yu, "Stability analysis of a third-order digital sigma-delta modulator," *APCC 2006*, Accepted, 2006.
6. Amin Z. Sadik and Zahir M. Hussain, "Limit cycle investigation in a ternary structure," *IEEE TENCON 2006*, Accepted, 2006.
7. Amin Z. Sadik and Zahir M. Hussain, "New DSP Using Short Word-Length," International DSP Creative Design Contest (DSPCDC'2006), Southern Taiwan University of Technology (STUT), Ministry of Education, R.O.C., Nov 20-22 2006.
8. Amin Z. Sadik and Zahir M. Hussain, "Short Word-Length LMS Filtering," *International Symposium on Signal Processing and its Applications (ISSPA 2007)*, Accepted, UAE, Sharjah, February, 2007

Awards

1. Best Paper Award in IEEE TENCON 2005 (Melbourne, Australia, Nov. 2005) for my paper "A Single-Bit Digital DC-Blocker Using Ternary Filtering."

2. Best Paper Award in the International DSP Creative Design Contest (DSPCDC'2006), Southern Taiwan University of Technology (STUT), Ministry of Education, R.O.C., Nov 20-22 2006 for my paper "New DSP Using Short Word-Length"
3. The Technical Award of the International DSP Creative Design Contest (DSPCDC'2006), Southern Taiwan University of Technology (STUT), Ministry of Education, R.O.C., Nov 20-22 2006 for the creativity in the proposed 1b/2b systems as an alternative to existing multibit DSP systems.

Keywords

Ternary filtering, single-bit sigma-delta modulator, comb filtering, DC blocker, Digital Phased-locked loop (DPLL), circle map, limit cycle, stability, least-mean-square algorithm (LMS), block-LMS (BLMS), ternary adaptive filter, single-bit adaptivity, 2-bit adaptive filter.

Preface

The well-known multi-bit digital signal processing (DSP) suffers mainly from the complexity of the multipliers and the inefficient chip area utilization in VLSI technology. In the last two decades, single-bit and ternary processing systems, based on sigma-delta modulating (SDM), have been presented as potential alternatives to the conventional DSP. The increased effective speed expected for the new short word-length techniques should translate into massive cost savings and increased flexibility for many electronic systems. Unfortunately, there are many issues in the above alternatives that are unresolved.

This thesis is primarily concerned with the development of an efficient DSP using ternary and single-bit techniques which would hopefully be equivalent to the conventional DSP in future. It is expected that developments in this area would result in VLSI chip economy and reduced cost for electronics consumers.

I hope that this work will help researchers working in DSP, communications, and related topics and inspire further research in these fields.

Amin Z. Sadik

Melbourne

July 2006

Contents

Declaration **i**

Acknowledgements **iii**

Publications and Awards **iv**

Keywords **vii**

Preface **viii**

List of Acronyms and Principal Symbols **xxi**

Abstract **xxiii**

1 Introduction **1**

1.1 The Conventional Multi-Bit DSP 1

1.2 Thesis Objectives 2

1.2.1 Research Questions 2

1.2.2 Research Aims 3

1.3 Original Contributions 4

1.4 Thesis Organization 4

2 Literature Review **9**

2.1 Introduction 9

2.2 Sigma-Delta Modulation 11

2.2.1 The Limit Cycle Behavior in ΣΔM 18
2.2.2 Problems with ΣΔM Analysis 20
2.2.3 The Alternative Analysis Approaches 21
2.2.4 Adaptive Scaling Schemes in ΣΔM's 22
2.3 Efficient Filters 24
2.3.1 Fast FIR Filters 26
2.3.2 Single-Bit Filtering Techniques 31
2.4 Summary 34

3 Single-Bit Ternary Filtering Using Sigma-Delta Modulation 35
3.1 Introduction 35
3.2 Ternary FIR Filter 36
3.3 Summary 41

4 DSP Applications Using Single-Bit Filtering: Comb Filtering 43
4.1 Introduction 43
4.2 A Proposed Ternary-Sigma-Delta Comb Filter 44
4.2.1 Design and Simulation of Ternary Sigma-Delta Comb Filter 47
4.3 A Proposed Sigma-Delta Comb Filter 50
4.4 A Proposed Design Approach 52
4.4.1 Theory and Design 55
4.4.2 The Proposed Structure 57
4.4.3 Simulation and Discussion 59
4.5 Stability of the Proposed Structures 60
4.6 Summary 61

5 DSP Applications Using Single-Bit Filtering: DC Blocking 66
5.1 Introduction 66
5.2 ΣΔ-Ternary DC Blocker: System Design 68
5.2.1 The Ternary Filtering Stage 68
5.2.2 The ΣΔ Modulator Stage 71

5.2.3 The DC Blocker 72
5.2.4 Simulation and Discussion 74
5.3 A Proposed Structure for DC Canceling in Single-Bit Domain . 82
5.3.1 Design and Analysis 82
5.3.2 Simulation and Discussion 86
5.3.3 Stability 89
5.4 Summary 95

6 Limit Cycle Behavior in Ternary Structures 96
6.1 Introduction 96
6.2 Analysis of a Third-Order ΣΔ Topology 97
6.3 The High-Order ΣΔ Topology 101
6.4 Behavior of the System's Limit Cycles 102
6.4.1 Limit Cycle Notation 102
6.4.2 Zero-Input Limit Cycles 105
6.4.3 Limit Cycles for Non-Zero DC Input 108
6.5 Conclusion 111

7 A Stability of Sigma-Delta Modulators in Ternary Structures 113
7.1 Introduction 113
7.2 Stability Analysis of the Third-Order Topology 114
7.2.1 Non-Linear Dynamics Modeling 116
7.2.2 Traditional Stabilizing Design Approach 117
7.2.3 Fixed Point Approximation: An Analogy with DPLL . . 118
7.3 Summary 124

8 Short-Word Length LMS-Like Adaptive Filtering 125
8.1 Introduction 125
8.2 An Adaptive Ternary Algorithm: 128
8.2.1 Simulation and Discussion: 131
8.2.2 Discussion: 132
8.3 A Single-Bit Adaptive Approach 135

8.3.1 Gradient Approximation 136
8.3.2 System Design 139
8.4 Stability of SBLL 143
8.4.1 Dynamic Range of the SD Modulator 144
8.4.2 The Updating Step-Size 144
8.5 Simulation and Discussion 145
8.5.1 Learning Curves 146
8.5.2 Signal-to-Noise Ratio (SNR) 148
8.5.3 Non-Stationary Inputs 150
8.6 A Two-Bit LMS-Like Adaptive Filtering 152
8.6.1 Performance Comparison 153
8.7 Summary 154

9 Conclusions and Future Work **158**
9.1 Conclusions 158
9.2 Future Work 160

A Recursive Equation of Third-Order ΣΔ Topology **162**

B The Equivalent Function $f(k)$ **166**

C Difference Equation of M^{th}-Order ΣΔ System **169**

Bibliography **171**

VITA **187**

List of Figures

2.1 Block diagram of the basic ΣΔM. 11
2.2 Linear model of a quantizer. 15
2.3 Block diagram of the first-order ΣΔM with a linear model for the quantizer. 16
2.4 A comparison between the NTF for a first- and second-order ΣΔM. . 17
2.5 Limit cycles consist of discrete lines in the frequency spectrum. . . . 18
2.6 Idle tones are peaks in the frequency spectrum but superimposed on a noise background. 19
2.7 Adaptation schemes used in AΣΔ: Input scaling. 23
2.8 Adaptation schemes used in AΣΔ: output scaling. 24
2.9 Block diagram of a traditional FIR filter. 25
2.10 Block diagram of an IIR direct form II filter. 26
2.11 Block diagram of the error feedback ΣΔM. 28
2.12 Block diagram of the FIR filter with ΣΔ modulated filter coefficients. 29
2.13 Block diagram of the decoder used in FIR filter with ΣΔ modulated filter coefficients and with ΣΔ modulated input signal. 29
2.14 Block diagram of the FIR filter with ΣΔ modulated input signal. . . 30
2.15 Block diagram of the single-bit FIR filter. 32
2.16 Block diagram of the first order single-bit IIR filter. 33

3.1 Block diagram of a ternary FIR filter. 36
3.2 Block diagram of the 2nd-order ΣΔ modulator. 37

3.3 Block diagram of the digital ΣΔ FIR-like bit-stream filter proposed in [4]. 38
3.4 Simulated frequency response of the single-bit IIR filter for different values of α = 0.1, 0.01, and 0.001. 41
3.5 Theoretical frequency response of the single-bit IIR filter with α = 0.1, 0,01, and 0.001. 42

4.1 Block diagram of the ternary ΣΔ filter. 44
4.2 Block diagram of a ternary FIR filter. 45
4.3 Block diagram of a second-order ΣΔM with ternary quantizer $Q_T[.]$. 45
4.4 Frequency response of the ternary filter in the proposed comb structure. 48
4.5 Frequency Response of the proposed ternary-ΣΔ single-bit comb filter with OSR = 128. 49
4.6 Phase response of the proposed single-bit ternary-ΣΔ comb filter. . . 49
4.7 Block diagram of the proposed ΣΔM structure. 51
4.8 Noise and signal transfer functions of the proposed ΣΔ single-bit comb filter. 51
4.9 Frequency Response of the proposed ΣΔ single-bit comb filter. . . . 52
4.10 Phase Response of the proposed ΣΔ single-bit comb filter. 53
4.11 Block diagram of a second-order ΣΔM. 54
4.12 Block diagram of the proposed single-bit ΣΔ based digital comb filter. 58
4.13 Block diagram of the designed M-period resonator. 63
4.14 Noise transfer function, NTF($e^{j\Omega}$) and signal transfer function, STF($e^{j\Omega}$) for the designed single-bit M-period resonator 63
4.15 Frequency Response of the proposed single-bit M-period resonator filter order $M = 10$ and OSR, $R = 64$. 64
4.16 SQNR against OSR with input signal amplitude of 0.5. 64
4.17 SQNR of the proposed single-bit comb filter versus the input amplitude for different values of OSR. 65

5.1 Frequency response of ideal multi-bit DC blocker for different values of the gain parameter α=0.5, 0.9, and 0.995. 69
5.2 Structure of the proposed single-bit ternary DC blocker. 74
5.3 A block diagram of the linear model for a second-order $\Sigma\Delta$ modulator. 75
5.4 Frequency response of the theoretical STF (dotted) and NTF (solid) for b=10 and a=0.001. 75
5.5 Frequency response of the ternary filter stage using Remez technique, compared with the target response (dashed line). 76
5.6 Frequency response of the ternary filter stage using IFIR technique. . 76
5.7 Signal attenuation in ternary filter stage against the OSR for $b = 1.6$ and $a = .01$. (*) Remez. (o) IFIR. 78
5.8 The overall frequency response of the ternary-$\Sigma\Delta$M single-bit DC blocker. 79
5.9 The overall phase response of the ternary-$\Sigma\Delta$M single-bit DC blocker (solid), compared with the phase response of the target impulse response (dashed). 79
5.10 Spectra of the single-bit input and the single-bit output of the proposed DC blocker. Above: noisy sinusoid input spectrum. Below: output spectrum. 80
5.11 Spectra of the single-bit input FM and the single-bit output of the proposed DC blocker. Above: input spectrum. Below: output spectrum. 80
5.12 Spectra of the single-bit input AM-FM and the single-bit output of the proposed DC blocker. Above: input spectrum. Below: output spectrum. 81
5.13 Reconstructed sawtooth DC-biased input signal from the single-bit DC blocker. 81
5.14 The proposed DC-blocker. 83
5.15 Linear model approximation of the system quantizers. 83

5.16 Signal and noise transfer functions, STF($e^{j\Omega}$) and NTF$_y$($e^{j\Omega}$), of the DC blocker using first-order SDM with $\alpha = 0.0205$ and $\beta = 0.2705$. . 85

5.17 The ratio $\rho = \text{SNR}_{\text{ovo}}/\text{SNR}_{\text{ovi}}$ (in dB) versus the gain parameters α and β using 10-bit resolution. 87

5.18 The ratio ρ versus the gain parameters α and β (10-bit resolution) of the DC blocker using a second-order SDM. 88

5.19 Frequency response of the simulated DC-blocker: (solid) using second-order SDM ($\alpha_m = 0.0127$, $\beta_m = 0.0508$); (dashed) using fist-order SDM ($\alpha_m = 0.0205$, $\beta_m = 0.2705$). 89

5.20 Input and output spectra of the DC blocker (using a second-order SDM) for an FM input. 90

5.21 Input and output spectra of the DC blocker (using a second-order SDM) for an FM input. 90

5.22 Input and output spectra of the DC blocker for a noisy AM-FM input. 91

5.23 Input and output spectra of the DC blocker for a noisy sawtooth input. 91

5.24 Multiplication of a single-bit signal by a multi-bit constant. 92

5.25 Root-locus of the proposed DC blocker with β=0.8. 94

5.26 Pole-zero plot of the DC-blocker with first-order SDM at $\rho = \rho_m$ using α=0.0205 and β=0.2705. 94

6.1 Structure of the third-order $\Sigma\Delta$ modulator under consideration. . . 98

6.2 Structure of the M^{th}-order $\Sigma\Delta$ modulator under investigation. . . . 102

6.3 The autocorrelation function $R_{yy}(n)$ of the structure output for zero input. The initial conditions are: $u_o = .2$, $u_1 = .4$, and $u_2 = 0$, and $\alpha = 0.1$. 103

6.4 The output frequency spectrum under same parameters. 103

6.5 The autocorrelation function $R_{QQ}(n)$ representing the number of transitions within the limit cycle period. 104

6.6 Phase-plane portrait of the third-order structure for $x = 0$. The initial conditions are: $u_o = .2$, $u_1 = .4$, and $u_2 = 0$, and $\alpha = 0.1$ (the diagonal straight line represents $y = x$). 104

6.7 The system phase portrait with $\alpha = 0.1$ and initial conditions: $u_o = 1.1$, $u_1 = 1.11$, $u_2 = 1.2$. 107

6.8 The system phase portrait with $\alpha = 0.2$ and initial conditions: $u_o = 1.1$, $u_1 = 1.11$, $u_2 = 1.2$. 107

6.9 The phase portrait for $\alpha = 1/6$ and initial conditions of: $u_o = 1.1$, $u_1 = 1.11$, $u_2 = 1.2$. 108

6.10 The parameter α versus the maximum threshold dc input beyond which no stability is guaranteed. 109

6.11 The maximum limit cycle length $L_{\max}$ as a function of the dynamic range input under fixed initial conditions and for $\alpha = 0.1$. 110

6.12 The phase plane for x=$1/14$, $\alpha = 0.1$ with initial conditions $u_o = 0.5$, $u_2 = 0.8$, and $u_3 = 0.8$. 111

7.1 An attractor of third-order $\Sigma\Delta$ system with x=$1/20$, α=0.1, and initial condition set (0,0,-0.3) 115

7.2 An attractor of third-order $\Sigma\Delta$ system with x=$1/50$, α=0.1, and initial condition set (0.7,0.9,1) 115

7.3 Structure of the single-bit third-order $\Sigma\Delta$ modulator. 116

7.4 The theoretical boundary of the gain parameter α versus the average output according to (7.23). 123

7.5 Stability region (shaded) of the third-order structure (for zero average output). 124

8.1 Adaptive multi-bit noise cancelling. 127

8.2 Structure of the Adaptive ternary filter. Note that "D" represents a single-bit delay element. 130

8.3 SNR improvement using adaptive filtering: ternary (solid) versus LMS (dashed) using a noisy sinusoid. 132

8.4 The step-size μ versus the mean-square error of the LMS Wiener algorithm. 133
8.5 The tracking response of the adaptive ternary filter. (solid):estimated output, (dotted):received input. 133
8.6 The frequency spectra of the received signal (upper), and the estimated signal (lower). 134
8.7 Learning curve of the proposed structure for a noisy sinusoid. . . . 134
8.8 A proposed block-diagram for single-bit LMS-like adaptive filtering. . 136
8.9 The proposed single-bit block LMS-like (SBLL) adaptive structure. . 143
8.10 The second-order SD modulator used in Fig.(8.9). 145
8.11 A comparison between the (undecimated) learning curves of the single-bit adaptive filter SBLL and the conventional LMS for a noisy sinusoidal input with m=20, SNR_i = 24.6 dB, and noise power -27.6 dB for both cases. 147
8.12 A comparison between decimated learning curves of the single-bit adaptive filter SBLL (solid) and the conventional LMS (dashed) with m=20, SNR_i = 24.6 dB, and noise power -27.6 dB for both cases. . . 147
8.13 A comparison in SNR improvement (ρ) between the SBLL (solid) and the corresponding standard infinite-precision LMS algorithm (dashed) (m=20, OSR = 128). 148
8.14 A comparison between the the original analog noisy sinusoid, i.e., before SD modulation (above) with SNR = 10 dB, and the output of the SBLL filter (below). 149
8.16 Tracking response of the adaptive single-bit filter: (dark) estimated output; (light) received AM-FM input; SNR_i = 10 dB. 150
8.15 SNR improvement using the SBLL as a function of OSR: (solid) OSR = 256, (dashed) OSR = 128, (dotted) OSR = 64. 150
8.17 Power spectra of the AM-FM input (above, with SNR_i = 10 dB) and the output (estimation) signal (below). 151

8.18 Spectrum of the estimated signal using single-bit adaptive filter SBLL (solid) in response to a single-bit AM-FM input (dotted). 151

8.19 The proposed 2-bit block LMS-like (SBLL) adaptive structure. . . . 154

8.20 A comparison between the (undecimated) learning curves of the single-bit adaptive filter SBLL and the conventional LMS for a noisy sinusoidal input with m=20, SNR_i = 24 dB, and noise power -25.2 dB for both cases. 155

8.21 A comparison between decimated learning curves of the single-bit adaptive filter SBLL (solid) and the conventional LMS (dashed) with m=20, SNR_i = 24 dB, and noise power -25.2 dB for both cases. . . . 155

8.22 A comparison among the proposed LMS-Like adaptive filters and the conventional LMS algorithm (in terms of improvement in SNR represented by ρ). 156

8.23 SNR improvement using the proposed 2-bit adaptive filter as a function of OSR: (solid) OSR = 256, (dashed) OSR = 128, (dotted) OSR = 64. 156

8.24 Power spectra of a sinusoid input (above, with SNR_i = 10.4 dB) and the corresponding output (estimation) signal (below). 157

List of Tables

5.1 A Comparison between Remez and IFIR techniques (OSR=32). 77

7.1 Routh-Hurwitz array 121

A.1 Coefficients of the initial conditions 163

A.2 Coefficients of the signum terms 164

List of Acronyms and Principal Symbols

α	Gain parameter
$(.)^T$	Transposition
$(.)^*$	Conjugation
$(.)^H$	Hermitian transposition
Ω	Normalized Radian Frequency
μ	Step-size of algorithm adaptation
σ	Variance of a signal
ρ	SNR improvement
M	Length of an oversampled FIR filter
m	Length of Nyquist rate FIR filter
$\mathbf{h}(n)$	Vector of an FIR impulse response Coefficients at sample time n
$\mathbf{w}(n)$	Vector of an adaptive FIR filter coefficients at ample time n

ADC	Analog-to-Digital Converter
ΣΔM, SDM	Sigma-Delta Modulator
BLMS	Block Least-Mean Square
f	Frequency
f_s	Sampling Frequency
FPGA	Field Programmable Gate Array
FIR	Finite-Impulse Response
FIRb	Single-bit FIR filter
IIR	Infinite-Impulse Response
mse	Mean-Square Error
NTF	Noise Transfer Function
PCM	Pulse Code Modulation
DSP	Digital Signal Processing
PLC	Power On-Line Communication
PLL	Phase-Locked Loop
OSR	Oversampling ratio
SBLL	Single-Bit LMS
SNR	Signal-to-noise ratio
SQNR	Signal-to-noise and quantization ratio
STF	Signal-Transfer Function
$\mathrm{SNR_i}$	Input in-band SNR
$\mathrm{SNR_o}$	output in-band SNR
SNR	Signal-to-noise ratio

Abstract

Recently, short word-length (often single-bit) processing has become a very promising technique as it can implement many important DSP tasks with significant efficiency. The increased effective speed expected for the new short word-length techniques should translate into massive cost savings and increased flexibility for many electronic systems. Short word-length systems have already made a huge impact on industry. The core element in these systems is the single-bit sigma-delta modulator (SDM). Sigma-delta devices are based on oversampling techniques and have the capability of quantization noise shaping.

Despite the large body of work that has been done so far, there are many ill-understood and unresolved issues in sigma-delta modulation, and consequently in single-bit systems. These issues hindered the full adoption of single-bit techniques in industry. Among these problems are the stability of high-order modulators and their limit cycle behaviour. More importantly, there is *no* adaptive LMS structure of any kind for short-word length (ternary or single-bit) systems. The challenge in this problem is the harsh quantization that prevents straightforward LMS application.

In this thesis, the focus has been made upon three axes, namely, designing new single-bit DSP applications, proposing novel approaches for stability analysis, and tackling the unresolved problem of single-bit and short-word length adaptive filtering.

Two structures for single-bit digital comb filtering are proposed. The first

structure is based on ternary filtering, however, the output of the filter is in single-bit format. The second structure is based on second-order sigma-delta modulation SDM. These filters can be utilized in a wide range of promising applications.

Another design technique for single-bit digital comb filter is presented. The proposed filter response and performance are assessed in terms of signal-to-quantization-noise ratio (SQNR) and stability. It is found that the comb filter possesses a distinct frequency response in broadband signal applications. The same technique is utilized to design and simulate a single-bit N-period digital resonator. Feedback loop filters can be used to tune the frequency response of the sigma-delta modulators.

The DC content in single-bit domain is both undesirable and hard to remove. A ternary DC blocker structure is presented. This type of filtering is useful in practice to improve the stability and dynamic range of single-bit systems. The DC blocker is essentially a ternary filtering structure whose input and output are both assumed to have single-bit format. Performance is tested for different kinds of input signals, including sinusoidal, FM, and AM-FM signals.

We also proposed a single-bit multiplierless DC-blocking structure. The input is assumed to be a sigma-delta modulated bitstream. This DC-blocker is designed using a delta modulator topology with sigma-delta modulation embedded in its feedback path. Its performance is investigated in terms of the overall signal-to-noise ratio, the effectiveness of DC removal and the stability.

The above proposed structures would be very efficient to realize (as they contain no multi-bit multiplication) in hardware and can easily be implemented with FPGA.

On the second axis of this thesis, we considered the stability of a single-bit high-order sigma-delta modulator under dc input. A new approach for stability analysis is proposed. A nonlinear circle map is suggested to model the dynamics of the modulator. An analogy between the dynamics of the sigma-

delta modulator and the sinusoidal digital phase-locked loop (DPLL) is studied and an approximate fixed point solution is presented with stability criteria. Suggestions for designing stabilized high-order systems are also presented.

Despite their major advantage of hardware simplicity, ternary and single-bit systems have limited useability in practice due to their unresolved problem of adaptivity. The conventional LMS family of adaptive algorithms fail to converge if translated to the single-bit domain.

On the third axis of this work we tackled this challenging problem by introducing three short-word length LMS-Like adaptive filtering schemes.

First, an adaptive ternary LMS-like algorithm is proposed. Performance assessment using a sinusoidal input distorted by additive white Gaussian noise showed that the proposed algorithm is comparable to the traditional multi-bit Wiener-Widrow LMS algorithm.

Second, a single-bit-domain LMS adaptive filtering structure for noise cancelling is proposed, where all input, output, and filter coefficients are in single-bit format. The proposed structure is designed and analyzed, and its performance has been evaluated (and compared to the conventional Widrow-Hoff multi-bit LMS algorithm) in terms of convergence properties, signal-to-noise improvement, and computational complexity. Simulation results showed that the proposed adaptive structure exhibits performance that is equivalent to the infinite-precision LMS algorithm.

Finally, a 2-bit LMS-Like structure is introduced and its performance is compared with the ternary and single-bit adaptive algorithms. The reason behind presenting this structure is to find out the optimal word-length in the tradeoff between complexity and performance. As long as noise-cancelling adaptive filtering is concerned, the 2-bit adaptive filter outperforms the other algorithms. We expect that these adaptive algorithms will open the door for short word-length systems to be ready as a practical alternative for multi-bit signal processing systems.

Twelve papers have been published/ submitted during this candidature.

Chapter 1

Introduction

1.1 The Conventional Multi-Bit DSP

In traditional PCM technique, the input analog signal is sampled at the Nyquist rate and then represented by a multi-bit word through multi-bit quantization process (8 bit, 16 bit, or more). This technique, however suffers mainly from the complexity of the multipliers and the inefficient chip area utilization in VLSI technology. In the last two decades, sigma-delta modulation ($\Sigma\Delta$ M) technique has been presented as an alternative approach to conventional PCM techniques. In $\Sigma\Delta$ approach, the input analog signal is oversampled (many times greater than Nyquist rate) and coarsely quantized to short-length word, often single-bit. The output of the $\Sigma\Delta$M is a high-rate bit-stream -1, +1 and can be decimated and filtered to extract a good approximation to the input.

$\Sigma\Delta$M is an efficient technique to quantize an analog signal and has been used recently in a growing number of DSP applications. However, a comprehensive understanding of $\Sigma\Delta$M behavior is not achieved yet due to the presence of a non-linear element within its structure, i.e., *the quantizer*.

A ternary filter, which is an FIR filter with it's coefficients confined to -1, 0, +1, has been presented, as well, to increase the efficiency of the hardware implementation and power consumption. Single-bit digital systems based on $\Sigma\Delta$ modulation and ternary filtering have been found to be extremely efficient from the hardware implementation viewpoint. Unfortunately, there are many

issues in ΣΔM and ternary filters that are considered unresolved till now. These issues will be pointed out in the next Sections.

1.2 Thesis Objectives

Despite the revolutionary progress in digital systems in the past two decades, there is currently a limit to the applicability of digital systems. They can only be used where digital processing is able to "keep up with" the required tasks. Processing of very high frequency wideband signals, for example, is typically out of range of the conventional digital processing. To increase the range of applications that can be implemented digitally, it is crucial to increase the effective speed of digital processing. Short word-length (often single-bit) processing is a very promising technique in this regard, firstly because it lends itself well to parallel processing realizations, and secondly because short word-length operations can implement many important DSP tasks with remarkable efficiency. The increased effective speed expected for the new short word-length techniques should translate into massive cost savings and increased flexibility for many electronic systems. Short word-length system implementations have already made a huge impact on industry. For instance, short word-length A/D and D/A converters, and increasingly, digital audio systems using short word-length amplifiers, are common. Very promising applications based on single-bit ΣΔM systems have already been presented. Examples of such applications are video A/D conversion [1], wideband applications [2, 97], and ultrasonic beamforming [46]. Research on these systems is therefore critically important. Outcomes in this area will result in reduced costs for electronics consumers and increase the quality of life. It is the aim of this thesis to contribute to the body of knowledge in this direction.

1.2.1 Research Questions

The proposed PhD program attempts to answer the following questions:

1. Can single-bit systems be designed to perform or approximate the functions of existing multi-bit systems?

   Initial attempts are successful but an extensive research is required.
2. Does a ternary filter possess limit cycles similar to the case of $\Sigma\Delta$M? If yes, what laws would the limit cycles follow?
3. Is it possible to utilize LMS adaptive techniques (e.g., for communication channel noise canceling) in ternary filtering?

   Adaptivity in ternary and single-bit systems is an unresolved problem and represents the major practical obstacle towards their wide spread usage in communications.
4. Can ternary filtering be utilized efficiently in broadband and other communication applications?
5. What is the optimal word-length that makes these systems capable of replacing existing multi-bit systems?

1.2.2 Research Aims

The specific objectives arising from these questions have been addressed. These can be summarized as follows:

1. Designing new single-bit (or short word-length) systems using both ternary filtering and $\Sigma\Delta$ modulation.
2. Investigating the occurrence of limit cycles in ternary filters.
3. Investigating the LMS adaptation of ternary and single-bit filtering.
4. Investigating the optimal (shortest) word-length that suits various DSP applications.
5. Investigating the stability issues of the designed short word-length systems.

1.3 Original Contributions

This thesis makes many original contributions to the body of signal processing knowledge both in theory and in implementation. A number of novel algorithms and techniques to model and design new short word-length systems have been presented.

The main contributions of this dissertation are summarized below:

1. Designing new single-bit and ternary DSP applications. A single-bit ternary and $\Sigma\Delta$M-based comb filtering and DC-Blockers are proposed. The work led to publications in Eurasip Journal on Applied Signal Processing, as well as in APCC 2005 and TENCON 2005 Conferences.
2. The limit cycle behavior in ternary structures has been explored and shown to exist. This led to a publication in TENCON 2006.
3. A novel approach in the stability analysis of the ternary structure is proposed, which invokes the analogy between the operation of the $\Sigma\Delta$ and the digital phase-Locked Loop (DPLL) systems. This approach can be expanded to include higher (>3) $\Sigma\Delta$ modulators. This led to publication in Digital Signal Processing Journal.
4. The unresolved issue of LMS adaptivity in short-word length digital filtering has been addressed. Impressive results are obtained and adaptive structures (ternary and single-bit) are designed, analyzed and simulated. It is expected that this achievement would open the door for the short-word length techniques to replace the traditional multi-bit PCM counterparts in the near future. This led to a publication in IEE Electronics Letters.

1.4 Thesis Organization

This thesis is comprised of nine chapters, which can be divided in three parts. The first part includes development of new single-bit DSP applications as in Chapters 4 and 5. The second part contains the study of limit cycle behavior

and stability analysis of ternary topology and higher-order ΣΔ systems in Chapters 6 and 7. The third part consists of a proposed short-word length adaptive filtering approach as explained in chapter 8.

The dissertation is organized as follows:

Chapter 2: Literature Review

Literature survey on single-bit processing techniques is made in this chapter. As the single-bit processing techniques almost entirely involve the utilization of sigma-delta modulation at some stage, the single-bit format is inherently related to sigma-delta modulators (ΣΔM). The main obstacles that hindered the full adoption of these techniques in industry and life are addressed as well. Emphasis has been put on the inherent problems regarding ΣΔMs behavior. Issues such as limit cycle behavior and stability of high-order ΣΔ systems are considered as not fully understandable. The topic of single-bit adaptivity is quite a challenging task, both in theory and implementation, and is regarded as an unresolved problem.

Chapter 3:Single-Bit Ternary Filtering Using Sigma-Delta Modulation

In this chapter a bit-stream filtering structure is introduced. It consists of a ternary FIR filter cascaded with an IIR ΣΔM structure. This structure is being the basis of many single-bit DSP applications. Since many of the ternary filter tap values are zero and each non-zero tap requires only very simple multiplication hardware, the system is very resource efficient and fast, as no complex mathematical operations are required. Performance enhancement is possible through increasing the oversampling ratio, however, this requires increasing the number of taps and the sampling rate of the system, hence, there is an inherent trade-off between hardware efficiency and performance.

Chapter 4: DSP Applications Using Single-Bit Filtering: Comb Filtering

In this chapter, two structures for single-bit output comb filtering are pro-

posed and simulated. The first structure is a combination of a ternary filtering stage and a $\Sigma\Delta$M. The second structure which is based on a second-order $\Sigma\Delta$M, is designed and its performance is evaluated in terms of signal-to-quantization noise ratio (SQNR), the dynamic range (input signal level), and stability. Moreover, it is shown that the same design technique can be used for other single-bit systems, where we used it to design a multi-period resonator. It was shown that the proposed filters lend themselves very well to broadband input signals and can be utilized in emerging technologies such as the Broad-Band Power-line Communication (BPL).

Chapter 5: DSP Applications Using single-Bit Filtering: DC Blocking

In this Chapter, two efficient multiplierless structures for DC-canceling in the single-bit domain has been proposed. The first consists of a ternary filtering stage followed by a sigma-delta modulator stage. Two design techniques were utilized to generate the ternary taps. For each technique, the associated ternary filer stage was assessed in terms of DC attenuation and hardware efficiency. The simulated system response has been studied through the application of various DC-biased, noisy signals. The DC content was removed completely from all kinds of input signals.

The second is a novel single-bit domain DC canceling structure. It is evaluated in terms of the overall SNR and the magnitude of DC attenuation. The role of the gain parameters is investigated and optimal performance has been reached. The system is examined using different types of signals.

Chapter 6: Limit Cycle Behavior in Ternary Structures

The difference equation and the iterative solution that describe its operation are developed in this chapter. It is shown that the system exhibits limit cycle behavior under certain conditions of the system parameters. The M^{th}-order difference equation of similar $\Sigma\Delta$ topologies are also developed. Moreover, a general formula for obtaining the average output of these systems

is derived. The system was then simulated extensively and a random search method is utilized to discover and extract the limit cycles and identify their features. It seemed that this topology, which is a third-order $\Sigma\Delta$ modulator, possesses a highly non-linear behavior.

Chapter 7: A Stability Analysis Approach for Sigma-Delta Modulators in Ternary Structures

In this chapter, we attempt to set out a comprehensive analysis to the third-order $\Sigma\Delta$ topology utilized in ternary filters, both mathematically and by simulation. This is done by utilizing the circle map dynamics to accurately model the operation of the $\Sigma\Delta$ structure, which is treated as a third-order sinusoidal digital phase-locked loop system. Accordingly, the stability topic is addressed using the fixed point techniques. This analysis would be of remarkable importance to other higher-order $\Sigma\Delta$ structures after some appropriate modifications.

Chapter 8: Short-Word Length LMS Adaptive Filtering

In this chapter we tackle the unresolved problem of ternary and single-bit LMS adaptive filtering. We propose an approach for LMS adaptive ternary filtering. Despite the simple structure, simulation results showed that the proposed algorithm is parallel in performance to the standard multi-bit LMS algorithm. We expect that this approach will open the door for a wide range of applications for ternary systems.

In addition, a single-bit-domain LMS adaptive filtering structure for noise canceling is presented, where all input, output, and filter coefficients are in single-bit format. The proposed structure is analyzed and its performance is evaluated (in comparison to the conventional Widrow-Hoff multi-bit LMS algorithm) in terms of convergence properties, signal-to-noise ratio improvement, and computational complexity. Simulation results showed that the proposed adaptive structure exhibits performance that is equivalent to the infinite-precision LMS algorithm.

Chapter 9: Conclusions and Future Work

This chapter summarizes the main conclusions of this dissertation and presents possible future directions.

Chapter 2

Literature Review

2.1 Introduction

The short word-length (often single-bit) format generated by sigma-delta modulator ($\Sigma\Delta$M) makes for greatly simplified arithmetic processing. For hardware implementation, this simplified processing implies reduced silicon space and reduced power consumption [4].

Processing tasks which are rich in multiplications are particularly strong beneficiaries of the use of single-bit signal representation. This is so because multi-bit multiplications require complex hardware implementation, whereas in the single-bit domain, multiplications can simply be implemented using a couple of gates or a very simple look-up table [5]. An efficient hardware implementation of $\Sigma\Delta$ systems can be attained if both the input signal and the transverse FIR filter impulse response representations are in binary or ternary format [6]. Both $\Sigma\Delta$M and ternary filters use coarse quantization to enable simple hardware implementation. Ternary filters have an architecture similar to FIR transversal filters, however, the tap values are limited to {-1, 0, +1}.

Although there are several algorithms presented to design ternary filters [7, 8, 9, 10, 11], the design techniques are particularly difficult to implement. Moreover, techniques to predict the performance of a ternary filter are not often presented. Developing an easily implemented, optimal ternary filter de-

sign algorithm will allow signal processing designers to get the best possible performance from ternary filters, causing them to become more widely used and offering more possibilities for increased hardware efficiency [12].

Unlike ternary filters, ΣΔMs have been extensively analyzed using different techniques. Yet, ΣΔM system understanding is far from complete [13]. Issues related to ΣΔM such as noise performance, instability, integrator spans, idle tones, limit cycle behavior, chaos, and adaptation have been often addressed. Ternary filters, on the other hand, have undergone very limited analysis and there are many unresolved issues that should be addressed. Some of these issues, which are to be investigated in this research, are listed below. First of all, it is unknown whether ternary systems have limit cycles similar to those in ΣΔM, and if they have, what is the law that these limit cycles may follow? Second, adaptive ternary filtering is an unresolved issue. Adaptive filtering is a vital topic in modern signal processing and digital systems. However, due to the short word length nature of single-bit systems, this issue is quite a challenging task, both in theory and implementation. Third, there is a need to investigate the possibilities of designing new single-bit ΣΔM ternary systems suitable for broadband applications such as RF and the promising technology of Broadband Power-Line Communication (BPLC). This is a promising avenue because ternary filters lend themselves well to low frequency applications.

In this chapter, we attempt to conduct a comprehensive literature survey on the single-bit processing techniques and, as the single-bit processing techniques almost entirely involve the utilization of sigma-delta modulation at some stage, the single-bit format inherently related to sigma-delta modulators (ΣΔM). The main obstacles that hindered the full adoption of these techniques in industry and theory are addressed as well.

2.2 Sigma-Delta Modulation

Oversampled ΣΔ modulators are becoming a standard high-resolution data conversion element [14]. These oversampled data convertors have several advantages over conventional Nyquist-rate convertors, including insensitivity to analog component imperfections [15], their high linearity, reduced complexity of the anti-aliasing filter, and lower cost of implementation. ΣΔMs have fast become one of the dominant data conversion elements in the low frequency range of the market. They come in one of two varieties: digital to analog (DAC) and analog to digital (DAC) conversion elements. A typical trademark that provides evidence of a ΣΔ data converter in consumer audio equipment is the "1-bit" advertisement.

Typically, ΣΔM's are used to convert a signal from multi-bit resolution to a single-bit resolution with little or no loss of dynamic range. This conversion, or modulation, is achieved through oversampling and noise shaping techniques [16]. The ΣΔM trades resolution in time for resolution in amplitude. Since these modulators can convert multi-bit signals to single-bit signals, they are at the cornerstone of single-bit digital signal processing [17, 18]. The general structure of a basic ΣΔM is shown in Fig.(2.1).

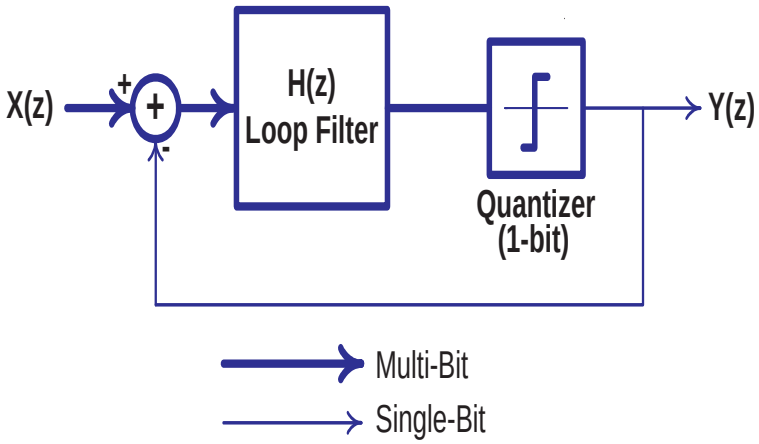

Figure 2.1: Block diagram of the basic ΣΔM.

This structure must be operated at an oversampled rate. The oversampling

ratio R is defined as $R = \frac{f_s}{2f_B}$, where f_s is the sampling frequency and f_B represents the input signal bandwidth. Oversampling decreases the *non-shaped* in-band noise by 3 dB for every doubling of the sampling frequency (f_s) [19]. By increasing the sampling rate from f_s to $f_{s1} = 2f_s$, the in-band quantization noise, previously spread over $[-\frac{f_s}{2}, \frac{f_s}{2}]$, is now spread over a larger frequency range $[-\frac{f_{s1}}{2}, \frac{f_{s1}}{2}]$, reducing the noise power spectral density to half the previous value. In addition to this reduction, there is a more significant attenuation in the in-band quantization noise power due to the inherent noise-shaping filtering action in ΣΔM. This is obvious from the in-band noise approximation which results from the white-noise assumption [20], that is

$$S_B \approx (\frac{\pi^{2M}}{2M+1})(\frac{1}{R^{2M+1}})\frac{\Delta^2}{12} \tag{2.1}$$

where S_B is the power of the in-band noise, M is the order of modulator, R is the oversampling ratio, and Δ is the quantization step. Equation (2.1) reveals that the attenuation of the in-band quantization noise due to increasing he oversampling ratio is exponentially improved as the modulator order increases. It is noteworthy that the above approximation has a practical limitation as it suggests that the precision can be improved indefinitely by increasing the order of noise shaping [20].

Given that ΣΔMs can only provide a relatively small bandwidth in comparison to the sampling frequency, noise shaping can result in significant reductions in quantization noise over the bandwidth of the interest. This is evident as ΣΔMs are typically operated with oversampling ratios ranging from 32 to 256. However, high oversampling ratios are a major obstacle towards the utilization of ΣΔ systems in broadband applications. Three different approaches have been proposed for obtaining good noise attenuation with low oversampling ratios [20]. The first approach is based on adopting higher-order transfer functions to increase the order of noise shaping (see equation 2.1). Utilizing a multibit quantizer is the second approach, while the third approach is the

cascading of multiple stages.

There are many ΣΔMs architectures for different orders; a good summary of these architectures can be found in [16] and [21].

In general, the quantizer utilized within ΣΔMs is a single-bit quantizer. Hence, the only possible outputs of such a quantizer are { 1,-1}. This single-bit quantizer has superior linearity qualities as compared to multi-bit quantizers, making it an extremely desirable quantization element in ΣΔM's [16]. This highly linear quantizer does not come without limitations. As only two output levels are possible, the quantization noise introduced is usually large.

The use of negative feedback in ΣΔM's is central to their operation. ΣΔM's typically operate by using negative feedback to suppress the quantization errors in the region of the loop filters passband. This negative feedback has also been found to provide some insensitivity to manufacturing imperfections within ADC's, unlike traditional multi-bit ADC's [16, 22].

The loop filter for low frequency applications such as audio have lowpass functions. Bandpass modulators can be created also through manipulation of the loop filter [22, 23]. Increasing the loop filter order in ΣΔM's can significantly improve the noise shaping capabilities. Such increases can lead to higher dynamic ranges at lower oversampling ratios and wider passband widths [16, 21].

Since the single-bit quantizer is a non-linear element within a negative feedback loop, the design and subsequent stability analysis of ΣΔM's is inherently complicated. Analysis and design of these structures is further complicated by increasing the order of the loop filter. This is evidenced by the large body of literature concerning design rules of thumb and stability issues (e.g. [24]). Systems with high-order loop filters and non-linear elements (such as the single-bit quantizer) in the feedback loop can be unstable.

ΣΔM design and analysis can be significantly simplified by linearizing the quantizer. While this linearization does not properly model the signal-dependent quantization noise, it has been found insightful [16, 22]. The 1-

bit quantizer is modelled by an additive white noise source with variance $\sigma_q = \Delta^2/12$, where Δ represents the quantization interval. This white noise model introduces the quantization error signal $q(n)$ and adds this to the quantizer input as shown in Fig.(2.2). Assuming a uniform distribution between -1 and +1 for quantization noise, the traditional linear model for M^{th}-order modulator relates the output spectrum $S_y(e^{j\Omega})$ to the input spectrum $S_x(e^{j\Omega})$ according to [25]

$$S_y(e^{j\Omega}) = S_x(e^{j\Omega}) + \frac{1}{3}[2\sin(\frac{\Omega}{2})]^{2M} \tag{2.2}$$

Assuming that the quantization noise is highly uncorrelated from one sample to the other and statistically independent of the signal, the ideal in-band SNR, SNR_{in}, achieved by an M^{th}-order $\Sigma\Delta$M, can be calculated as [21]

$$\begin{aligned} \text{SNR}_{\text{in}} &= 10\log_{10}(\sigma_{xy}^2) - 10\log_{10}(\sigma_{qy}^2) - 10\log_{10}(\frac{\pi^{2M}}{2M+1}) + \\ &\quad (20M+10)\log_{10}(\frac{f_s}{2f_B})(\text{dB}) \end{aligned} \tag{2.3}$$

where σ_{xy}^2 is the signal power (variance) at the output and σ_{qy}^2 is the in-band noise power at the output assuming zero mean. As the signal power is assumed to occur over the signal band only, it will not be subjected to any modification, and the signal power at the output σ_{xy}^2 is the same as the input signal power σ_x^2. The achieved SNR_{in} depends on the noise-shaping function of the modulator, which can be described in terms of z-domain poles p_i and zeros z_i. For an M^{th}-order modulator, the noise-transfer function NTF can be expressed as

$$\text{NTF}(z) = \frac{\prod_{i=1}^{M}(z - z_i)}{\prod_{i=1}^{M}(z - p_i)} = \frac{1}{1 + H(z)}. \tag{2.4}$$

A well-known choice for the noise-shaping pole locations is to arrange them in Butterworth configuration, whereas improved SNRs are achieved if noise shaping zeros are distributed across the baseband in conjugation pairs [26]. A $\Sigma\Delta$ structure with controllable SNR has been reported in [27, 28].

If we now return to Fig.(2.1) to mathematically describe the general $\Sigma\Delta$M

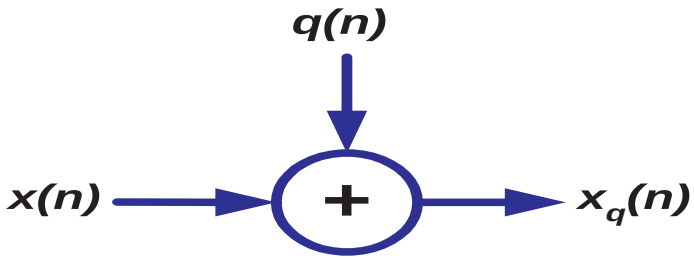

Figure 2.2: Linear model of a quantizer.

output using the linear model in the z-domain, the system can be described by:

$$Y(z) = \frac{H(z)}{1 + H(z)} X(z) + \frac{1}{1 + H(z)} Q(z) \tag{2.5}$$

where $X(z)$, $Y(z)$, $H(z)$ and $Q(z)$ represent input, output, loop filter and quantization signals and functions in the z-domain, respectively.

From equation 2.5, we can obtain signal and noise transfer functions (abbreviated as STF and NTF) as shown below.

$$\text{STF}(z) = \frac{H(z)}{1 + H(z)} \tag{2.6}$$

$$\text{NTF}(z) = \frac{1}{1 + H(z)} \tag{2.7}$$

It is clear from equations 2.5 and 2.7 that the quantization error $Q(z)$ is spectrally filtered. To illustrate this spectral filtering we will now analyze the first-order $\Sigma\Delta$M shown in Fig.(2.3). For convenience we have replaced the quantizer with its equivalent linear model.

The loop filter $H(z)$ for the first-order $\Sigma\Delta$M is given by:

$$H(z) = \frac{z^{-1}}{1 - z^{-1}} \tag{2.8}$$

which is an integrator. If we then substitute the integrator of equation (2.8) into the STF and NTF of equations (2.6) and (2.7), we obtain the first-order

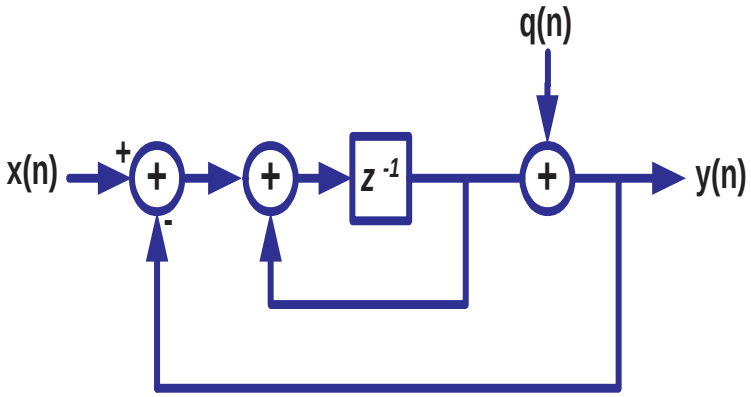

Figure 2.3: Block diagram of the first-order ΣΔM with a linear model for the quantizer.

ΣΔM transfer functions:

$$\mathrm{STF}(z)_{1^{st}} = z^{-1} \tag{2.9}$$

$$\mathrm{NTF}(z)_{1^{st}} = 1 - z^{-1} \tag{2.10}$$

where $\mathrm{STF}(z)_{1^{st}}$ is a pure delay that doesn't change the form of the input signal, whereas the quantization noise is filtered with the differentiator $1-z^{-1}$. As the differentiator has a high-pass frequency response as shown in Fig. 2.4, the quantization error or noise will be shaped away from the low frequency region. Hence, if the input signal is in the low frequency region, then it will be modulated into the single-bit format with reduced quantization error. It can be shown [29] that, for every doubling of the oversampling ratio in a first-order ΣΔM, SQNR improves by 9 dB (or equivalently, the resolution improves by 1.5 bits), where a 3-dB is due to the reduction in the power spectral density of the quantization noise and an extra 6-dB due to the noise shaping characteristic.

Increasing the number of integrals in the analog part of the modulator will improve the noise shaping performance and, consequently, give a higher resolution for the overall system. This is evident since the the noise-transfer

function for M^{th}-order modulator will be given as

$$\text{NTF}_M(z) = (1 - z^{-1})^M \tag{2.11}$$

For instance, the noise transfer function in the frequency domain for first- and second-order $\Sigma\Delta$ systems is shown in Fig.(2.4). The vertical line illustrates the band limit of a signal, where $f_B = 0.02 f_s$.

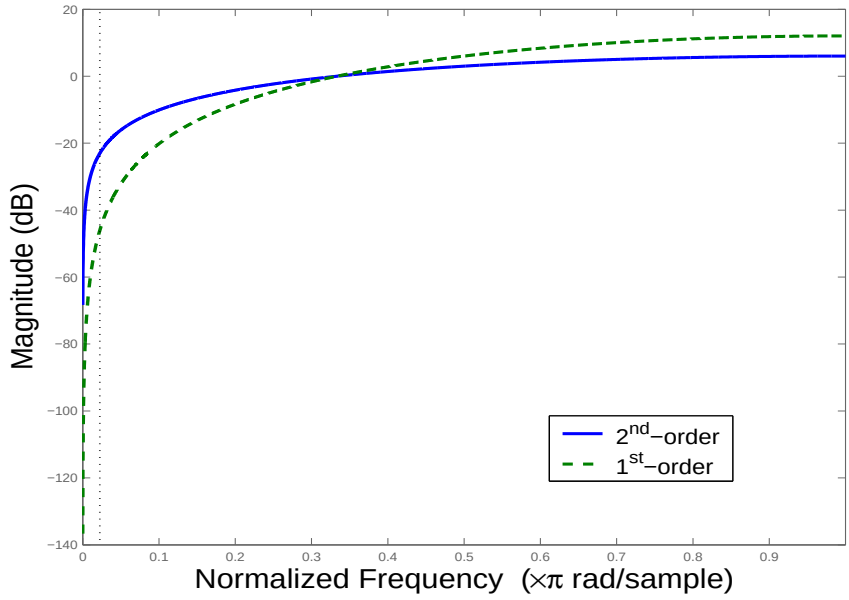

Figure 2.4: A comparison between the NTF for a first- and second-order $\Sigma\Delta$M.

Of course, the output signal from the modulator is in the single-bit format. Currently, the philosophy is to filter this single-bit signal to remove the quantization noise and then to sample decimate to the Nyquist frequency. This produces the pulse-code-modulated (PCM) format that is traditionally provided at the output of a ADC.

The traditional IIR and FIR filtering operations inherently produce multi-bit outputs. Hence, once the $\Sigma\Delta$M output is filtered using one of these traditional techniques, the signal is again in a multi-bit format.

2.2.1 The Limit Cycle Behavior in ΣΔM

ΣΔM's are known to exhibit spurious oscillations (tones) in the single-bit output sequence, called limit cycles, and are present for both low and high order modulators [30].

Limit cycles oscillations in sigma-delta modulated signals are usually regarded as a performance-degrading feature [32].

With conventional linear filtering, such limit cycles produce idle tones that may be audible to the listener when ΣΔM's are used for audio signal processing. However, one should recognize the difference between limit cycles and *idle tones* [33]. A Limit cycle is a sequence of L output bits which repeat itself indefinitely, while an idle tone represents a discrete peak in the frequency spectrum of the output of a ΣΔM such that it superimposed on a background of noise as illustrated in Fig.(2.5)and Fig.(2.6).

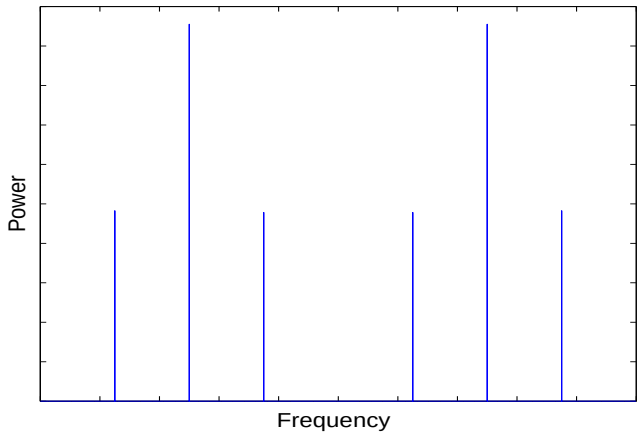

Figure 2.5: Limit cycles consist of discrete lines in the frequency spectrum.

Various approaches have been proposed to address the limit cycle phenomenon in nonlinear systems. For example, the describing function approach for computing the limit cycle points of uncertain nonlinear systems has recently attracted the attention of researchers. An algorithm proposed in [34] that makes use of the describing function analysis technique and tools of interval analysis for predicting the limit cycle behavior.

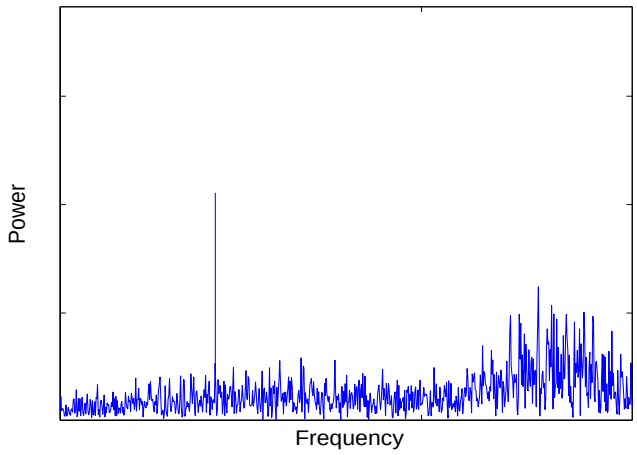

Figure 2.6: Idle tones are peaks in the frequency spectrum but superimposed on a noise background.

Considerable work has been done on the identification of limit cycle in first- and second-order ΣΔM's and on their elimination, (e.g. [35]-[39]), there exist no comprehensive analysis procedure for higher-order systems which might allow a designer to anticipate the spectral and power distributions of possible limit cycles. However, useful attempts have been published recently which were successful in gaining more insight into the limit cycle behavior in higher-order (> 3) ΣΔ systems. For instance, in [40], a framework for the analysis of a family of high-order ΣΔ systems was proposed. This analysis had facilitated the stability analysis through the reduction of a large number of high-order architectures to a diagonal form. That is, by transforming the state space into a coordinate system where the state variables interact only through the quantizer function.

In [41], state space matrices were utilized to describe high-order single-stage single-bit ΣΔM operated under the condition of a constant zero input signal. In this work a procedure was established for characterizing and validating potential limit cycles. A more generalized approach to describe 1-bit feedforward ΣΔ with constant input can be found in [33] which proved that, under almost all circumstances limit cycle behavior is observed in the output *if and only if* a limit cycle occurs in state space.

A common way of linearizing the modulator is to dither the quantizer with an independent noise signal [42]. Other remedies are: making the modulator chaotic (to break up simple oscillations [43, 44]) and using wavelet decomposition [45].

Almost all works regarding the limit cycle issue are done using constant input signal which is a highly relevant case. In actual practice, ΣΔM's are used for A/D conversion of band-limited signals and not with dc signals. Constant input signals are apparently the worst case for such modulators, and engineering practice recommends adding a high-frequency dither signals to make the input vary [14]. The dynamic behavior of ΣΔM under noisy or periodic inputs is still an unresolved topic [33].

2.2.2 Problems with ΣΔM Analysis

ΣΔM is an inherently non-linear system. The non-linearity is wholly contained within the representation of the quantizer element. However, sigma-delta modulators are often analyzed using linear techniques [47]. In these approaches, the quantizer is replaced by an additive white noise source and then a standard linear system analysis is applied. This approach can provide a good estimate of the noise performance, but it is unable to explain many aspects in the behavior of ΣΔM's, especially such phenomena as instability, integrator spans, quantizer switching frequency, idle tones, strong limit cycle behavior and chaos, which are inherently non-linear.

The non-linear nature of the ΣΔ system makes the stability analysis difficult [48]. Numerous methods have been attempted to solve this problem. For instance, Tsypkin's method [49], norm technique [50], and describing function analysis [51]. However, all of these techniques can only achieve a limited success and each suffers from limitations and/or deficiencies.

On the other hand, modeling the modulator is a challenging task as there are inherent sources of errors which arise from the modulator itself, mainly from the discrete-time integrator of each first-order loop [52]. This is so because

the integrators are the modulator stages where the analog signals exist and therefore, the component values have the most significant impact on the signal distortion. Practically, these discrete-time integrators are mainly based on switched-capacitor circuits acting as resistors. These switched-capacitors based integrators suffer from five error sources, namely: limited bandwidth, finite open-loop gain, limited output slew rate, mismatched capacitors, and voltage dependent capacitors [52].

2.2.3 The Alternative Analysis Approaches

To address these problems, it will be necessary to utilize some of the many non-linear analysis techniques available. A full rigorous non-linear analysis of these systems using any one technique would be very difficult, if not impossible [53]. An alternative approach is to identify specific problems, such as stability, and apply the most suitable non-linear analysis technique to that issue. Generally, three main approaches have been applied to the non-linear analysis of ΣΔMs: spectral analysis (noise and signal performance), geometric analysis (stability and integrator spans), and non-linear dynamics (limit cycle behavior).

1. *Spectral Analysis.* In this form of analysis [54, 55, 56, 57], equivalence is established between first-order ΣΔM (with constant input) and the circle map. The circle map is a well-known function in non-linear dynamics and ergodic theory. Using this equivalence it would be possible to explain some basic aspects of the system behavior, for instance, the fact that rational inputs lead to limit cycles.
2. *Geometrical Analysis.* In this approach the trajectories of the integrator outputs are analyzed. This approach is primarily concerned with stability and identifying the integrator spans of second-order systems [58, 49], and some third-order systems [57]. Extension on this approach has been made by including the class of inputs consisting a constant input of magnitude less than 1 and of an arbitrary sum of finite amplitude sinusoids [59]. This approach is further extended to consider the

trajectories as discrete points along a parabolic curve [60].

3. *Non-Linear Dynamics.* This approach has found an increasing interest. The strength of non-linear dynamics is that it provides insightful information into the behavior of non-linear systems and thus can give useful information about limit cycle effect and idle tones [61, 62]. It has been shown that a single ellipse or a finite number of ellipses are equivalent to the limit cycle behavior in the low-pass modulators [36].

2.2.4 Adaptive Scaling Schemes in ΣΔM's

Two significant advantages of high-order low oversampling ΣΔM's over lower-order ones. Firstly, the amplitudes of the idle tones and noise modulation are minimized. Secondly, when reducing the oversampling ratio, better performance will be achieved as compared to lower-order systems [63]. Unfortunately, higher-order ΣΔM's are only stable for relatively low *maximum stable input* beyond which the system becomes unstable. one approach to eliminate this drawback is to make the modulator capable of adaptive scaling.

One of the essential issues of ΣΔM design is to properly scale the integrators to avoid clipping, which may cause information loss and hence severe degradation in signal-to-noise ratio (SNR) [64]. Adaptive ΣΔM (AΣΔM) attempts to increase the dynamic range of sigma-delta modulators while keeping the quantization noise as low as possible [65]. AΣΔM achieves this objective by scaling either the input signal or the step-size of the quantizer through an estimation of the input signal strength. This estimation can be done from the input signal itself or from the modulator output as shown in Fig.(2.7) and Fig.(2.8), respectively. Using the input signal to perform the estimation is known as 'forward estimation' while using output signal is known as 'backward estimation'. Adaptation can be done continuously or sporadically in time. Several adaptation techniques have been investigated in the literature [66, 67, 68, 69, 70, 73]. Another scheme based on estimating the amplitude of the quantizer input instead of the input signal itself has also been proposed

[71, 72]. For instance, in [73], a method for improving the SNR of ΣΔM with single-bit quantization is introduced. However, this is done at the expense of two drawbacks. First, a moderate slew rate limitation of the input signal occurs. Second, the feedback signal is a multilevel sequence, and hence a multibit D/A converter is required in the feedback loop with particular demands on the linearity.

An adaptive scheme has been presented in [74] which reduces the order of the loop filter of high-order single-bit ΣΔM in order to stabilize them and also to improve their performance in the unstable region. The apparent drawback of this adaptive technique is the additional number of comparators and digital logic circuitry which are needed for its hardware implementation.

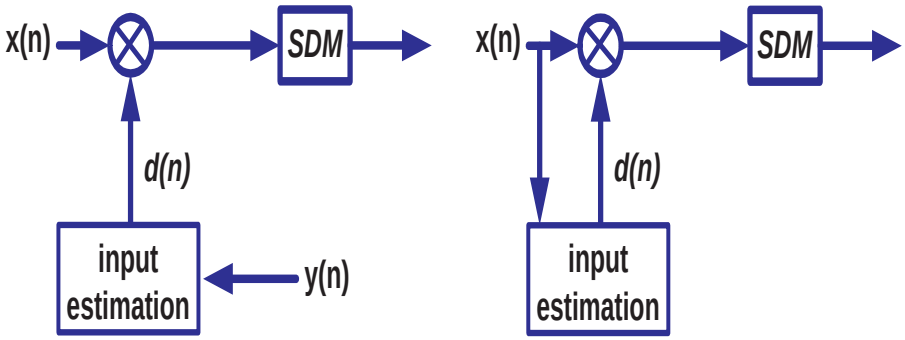

Figure 2.7: Adaptation schemes used in AΣΔ: Input scaling.

Although the discussion has already concentrated on the ΣΔM's, it can be generalized to include ternary filtering systems. The ΣΔM shares many features with other discrete-time processes in digital signal processing, for example, digital filters and digital phase-lock loops [75], etc. No one approach will yield all the required results, so it is important to have an understanding of which approach is useful for the specific problem under investigation. As the ternary filter suffers from several unresolved issues, one would think that these issues can be addressed through extension and/or modification of the approaches already utilized with the ΣΔM to tailor them to the case of

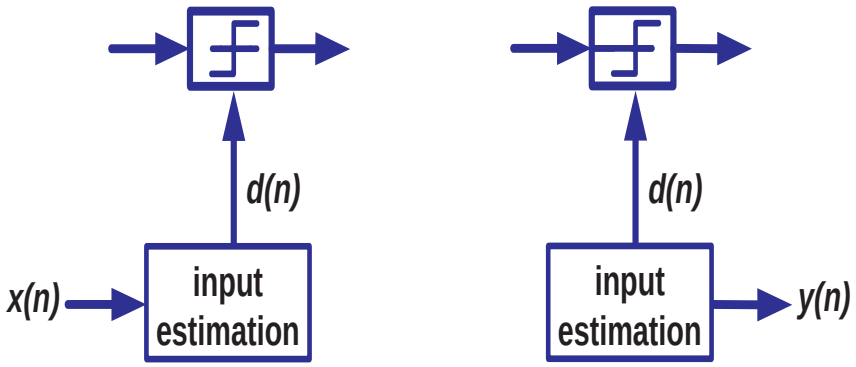

Figure 2.8: Adaptation schemes used in AΣΔ: output scaling.

ternary filter. However, as ΣΔM understanding is still far from complete, this task seems to be quite challenging.

2.3 Efficient Filters

Many signal processing tasks can be done via a microprocessor or a digital signal processor. Typically these hardware devices contain built-in multiplication circuits. Within digital signal processors it is not uncommon for several multiply-and-accumulate (MAC's) to be implemented in the integrated circuit. Such hardware can provide significant data throughput increases in digital filters as both the FIR and IIR structures shown in Fig.(2.9) and Fig.(2.10) require many multiply-and-accumulation operations per sampling period.

An alternative growing in popularity is to use programmable logic devices such as field programmable gate arrays (FPGA's) to undertake the digital filtering tasks. These devices have the advantage that operations can occur in parallel. This parallelism greatly increases the data throughput of digital filters, however, this speed increase comes at the cost of requiring large amounts of gates as compared to serial implementations. Again it is not uncommon for such FPGA devices to contain many built-in multipliers, however, these multipliers still require large amounts of silicon space within the FPGA.

As the speed of the processing elements and FPGA devices increases, the

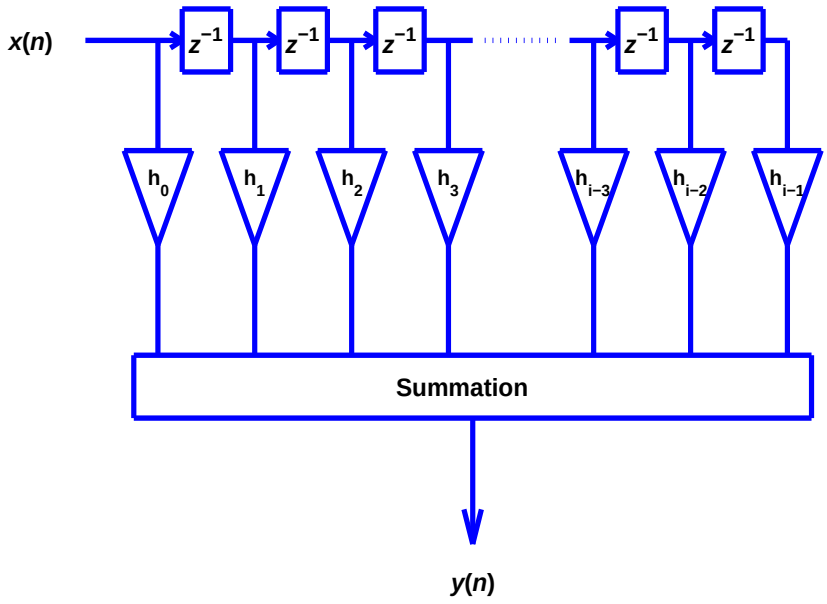

Figure 2.9: Block diagram of a traditional FIR filter.

applications that require increased filter complexity and speed are created. Limitations of processing speed dictate the maximum data throughput, and hence the maximum sampling rate that a given processing element can operate at [76]. Some obvious applications that require fast and efficient digital filters are decimation filters in $\Sigma\Delta$M [16], audio filter banks, charge-coupled-device filters, and software radio.

Each of the above applications is required to provide filtering with high data throughput and in some cases at high speed. For instance, software radio applications require complex hardware to be able to realize filters with different bandwidths and stopband attenuations. Audio filter banks require many different channels of varying complexity, and charge-coupled devices that operate in digital image capturing equipment require vast amounts of filtering. Each of these items would benefit from fast and efficient filtering techniques.

To achieve efficient and fast implementations, many techniques have been proposed. The theme in many of these techniques is to try to reduce the complexity of the multiplication operation so that simple and fast filter implementations can be achieved.

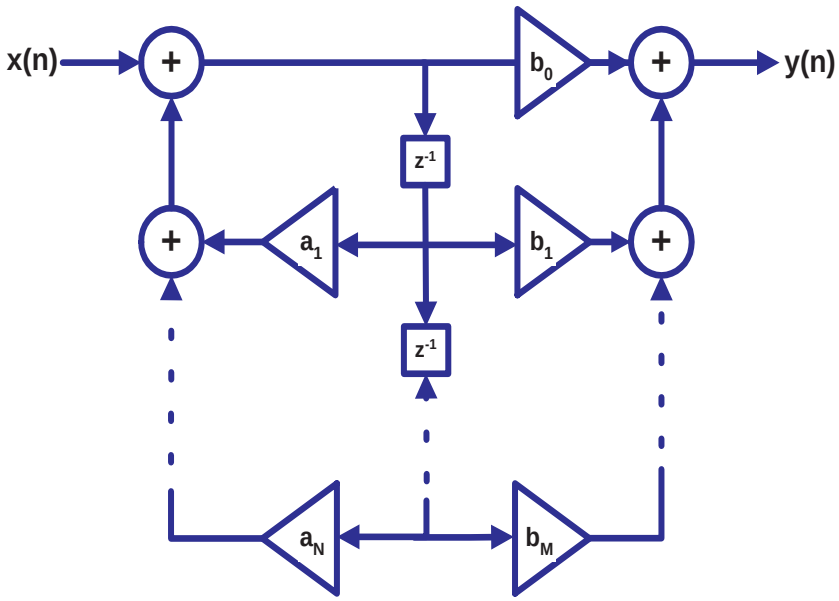

Figure 2.10: Block diagram of an IIR direct form II filter.

One popular method of reducing the complexity of the multiplication operation is to reduce the *word length* in both the input and the filter coefficients. The preferred method of word reduction is generally sigma-delta modulation, hence, this thesis will focus on these methods.

There are many techniques that use some form of sigma-delta modulation or the like to improve the efficiency of the digital filtering operations. An example of such techniques were reported in [11, 6, 5, 77, 78].

2.3.1 Fast FIR Filters

Fast and efficient filters generally fall into two categories: sigma-delta based and optimization techniques. Interestingly enough, all of these techniques use oversampling and single-bit or ternary formats to reduce either the input signal word length or the filter coefficients word length or both.

The optimization techniques generally use either dynamic programming [8, 9] or mini-max techniques [12, 79] to reduce the filter coefficients word length to the ternary format. The ternary library {+1,0,-1} is used because it adds almost no more complexity to the multiplication operation than the

single-bit word but it provides an improved stopband attenuation [12].

These optimization techniques focus on oversampling the signal and using ternary filter coefficients. While both of these techniques can provide some useful filters, they require the use of a reconstruction filter to completely define the spectral filtering response and to remove the quantization noise introduced by the harsh ternary filter coefficients. The filter coefficient generation methods were found to be very complex. They also require many iterations of converging algorithms that can take long periods of time to converge to a solution, if they converge at all.

There are many published works on sigma-delta based FIR filtering techniques, amongst which [80, 18, 81, 77, 82] are just a few. Most of these techniques use some form of sigma-delta modulation to reduce the word length of the filter coefficients or the input signal or both to either single-bit format or ternary format.

In [80] and [81] efficient FPGA narrowband filter implementation is discussed. The authors recognize that the precision, or number of bits, in the MAC operation can be reduced if the input signal to the FIR filter is reduced in precision. This filtering operation requires the input signal to be oversampled. This input signal is then resampled using a digital error feedback $\Sigma\Delta$M. The error feedback $\Sigma\Delta$M is shown in Fig.(2.11). This modulator uses a prediction filter in the negative feedback path. This prediction filter has a flat passband over the bandwidth of interest and is generally implemented with an FIR filter. The paper also discusses optimum prediction filter design based on statistics and a minimum-mean-squared error calculations.

Once the input signal has been resampled down to three or four bits, a full precision digital FIR filter is used to filter the signal. Overall, this resampling operation has shown over 50% reduction in logic resources as compared to traditional FIR filter implementation using FPGA. This filter shows a great promise for FIR filter implementations, however, further reduction in complexity can be achieved through harsher requantization to lower precision words.

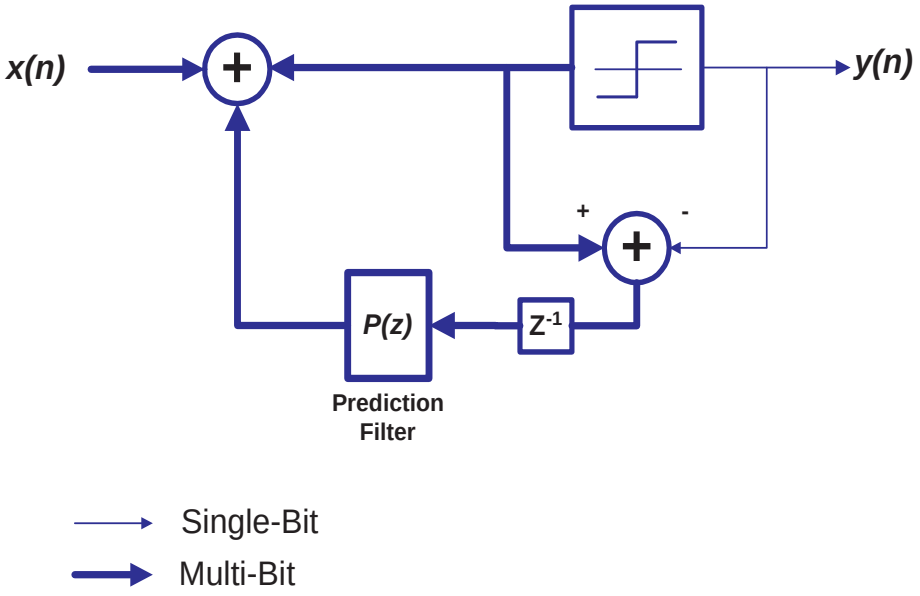

Figure 2.11: Block diagram of the error feedback ΣΔM.

In [77] and [6], fast and efficient digital filters are presented. These publications present two slightly different variations on sigma-delta filtering. In the first case the filter coefficients are modulated into a single-bit format. As a result, the input signal must be interpolated and zero-padded to R times the sampling frequency. The block diagram of this filter is shown in Fig.(2.12).

The decoder for this filter is used to reconstruct the signal by resampling to the Nyquist rate and filtering out the quantization noise. As discussed in [83], the use of two cascaded comb filters makes for simple implementation whilst removing any alias introduced into the system from the FIR filter. A block diagram of the decoder used is shown in Fig.(2.13).

The second structure that was proposed in [77] and [6] is shown in Fig.(2.14). This structure makes use of sigma-delta modulation of the input signal. No interpolation is required in this setup as the signal will be oversampled at the input of the modulator as the signal has already passed through a ΣΔM.

To perform filtering, this structure uses a zero-padded FIR filter with full precision filter coefficients. The filter is zero-padded R times to match the oversampling ratio of the ΣΔM. This eliminates many taps from the filtering

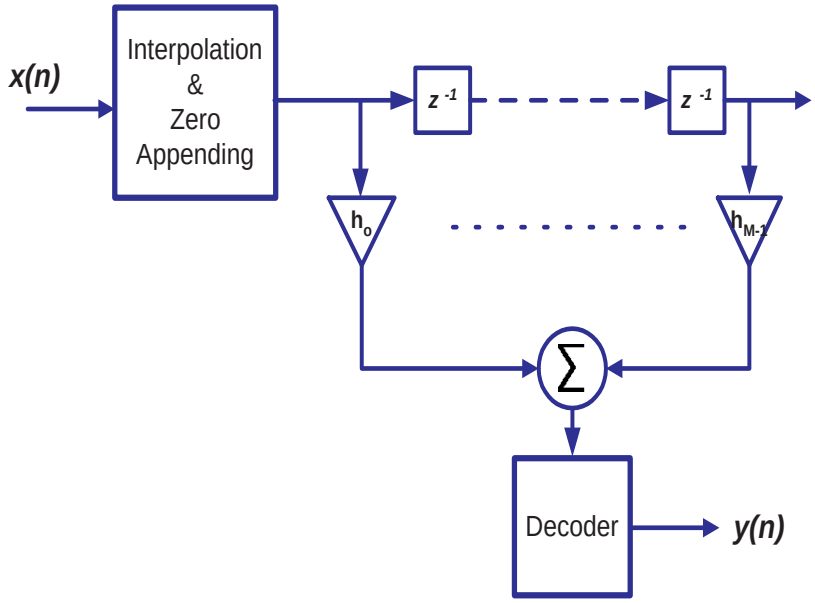

Figure 2.12: Block diagram of the FIR filter with ΣΔ modulated filter coefficients.

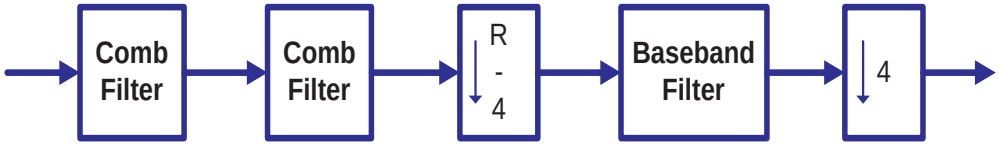

Figure 2.13: Block diagram of the decoder used in FIR filter with ΣΔ modulated filter coefficients and with ΣΔ modulated input signal.

operation, however, R aliases are generated from the FIR filter. As discussed earlier, the cascaded comb filters and the baseband filter effectively work together to remove these aliases and quantization noise. The outputs of this decoder and the FIR filters in these schemes are in a multi-bit format.

These publications (i.e.[77] and [6]) also discuss the use of a fully sigma-delta-encoded FIR filter. In this instance both the input signal and the filter coefficients are modulated into single-bit format. A similar structure as in Fig.(2.12) was utilized except that the interpolater was replaced with a sigma-delta modulator. This structure was found to further reduce the complexity of the filter.

Finally, in [6] ternary modulators were used in ΣΔM's, where an extra

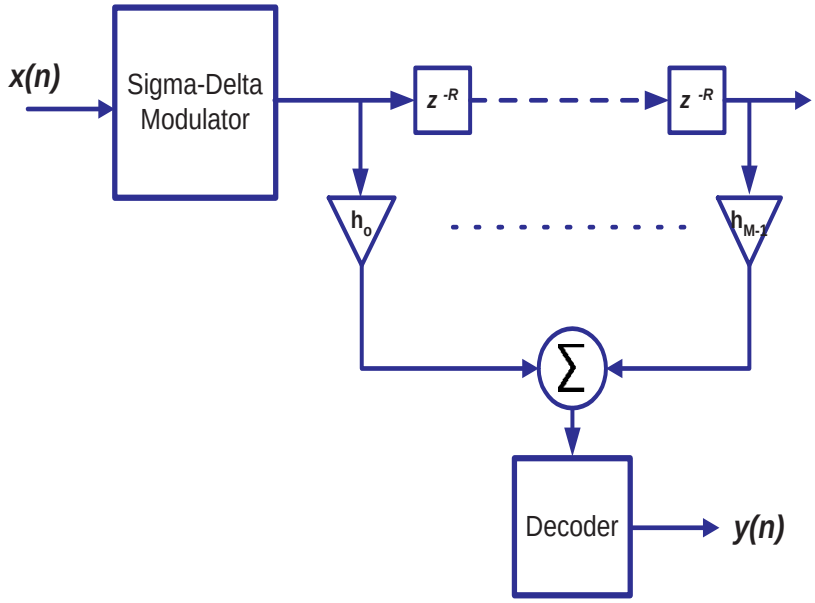

Figure 2.14: Block diagram of the FIR filter with ΣΔ modulated input signal.

symbol is added to the filters and input signals alphabets. This was found to improve the stopband attenuation of the filter and also the dynamic range of the ΣΔM. These works show a great promise; in particular, the benefit of ternary encoded filter taps may be useful. The use of a decoder though, is not conducive of single-bit processing because it produces a multi-bit output that requires complex hardware to process.

A few remaining efficient and fast filter implementations use modified quantizers within sigma-delta structures or modified sigma-delta structures to encode the filter coefficients into varying formats.

In [11], a look ahead decision feedback (LADF) sigma-delta architecture is used to encode the filter coefficients into a single-bit format. The authors found that cascaded ΣΔM architectures provide overall lower quantization noise than LADF, but argue that the LADF structure has far more simpler and less complex implementation and provides lower quantization noise than other sigma-delta architectures. Therefore, it is desirable for filter coefficient encoding. However, the LADF architecture and quantizer are a great deal more complex than simple single-bit quantizers and there associated ΣΔM

structures.

The last group of fast and efficient filters use canonic signed digit (CSD) quantizers. In [84] and [85], CSD or singed powers of two were used as the output of a quantization element within a sigma-delta modulator. The CSD output obtained from the ΣΔM can be used as FIR filter coefficients. In this case the multiplication operations become simple shifts. Another similar scheme, however, more promising, is in [86] which uses a slightly more complex system but essentially the same technique.

2.3.2 Single-Bit Filtering Techniques

As the name suggests, single-bit filters produce single-bit outputs. Despite the large number of publications on filtering methods that involve ΣΔM's, only a few (e.g., [87, 88]) of these publications involve the development of a single-bit output. In this section we introduce and describe the techniques that have been used to filter whilst maintaining a single-bit output.

In all publications on single-bit filtering, the authors have only found one method for single-bit FIR filtering [78, 82] and one method for single-bit IIR filtering [89, 90].

The single-bit FIR filtering technique is similar to that in [77] and [6]. However, the decoder in [6] has been replaced by a ΣΔM that has a lowpass signal transfer function. The single-bit FIR filter as proposed in [78] and [82] is shown in Fig.(2.15).

In this method the filter input is assumed to be in a single-bit format, while full precision filter coefficients are generated at the Nyquist rate. This newly generated impulse response is then interpolated R times, where R is the oversampling ratio of the input signal, via zero-interleaving. This zero-interleaving interpolation introduces R aliases that in previous works were removed by a decoder containing a cascade of comb filters and a baseband filter.

In the single-bit works of [78] and [82], a ΣΔM is utilized in place of the

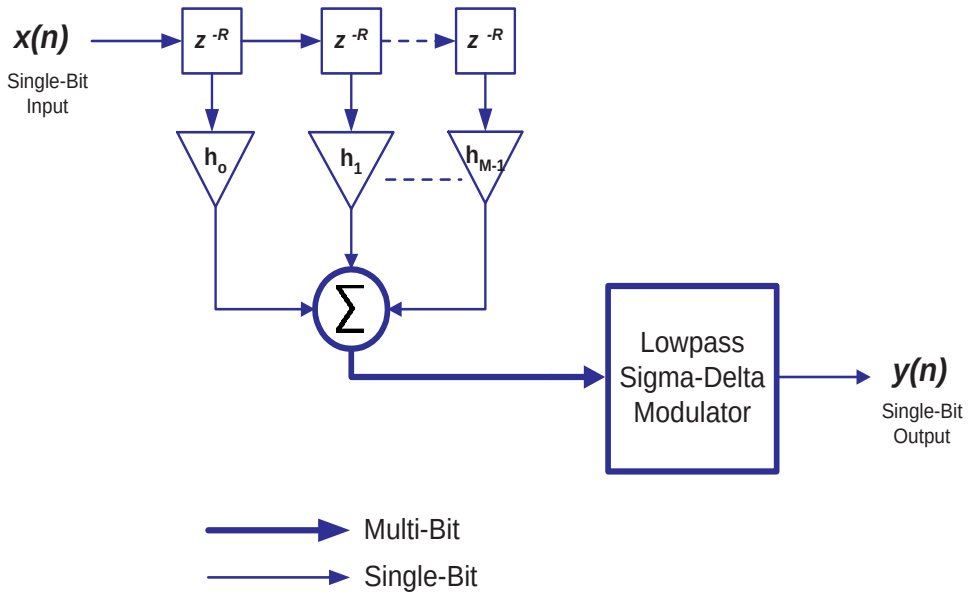

Figure 2.15: Block diagram of the single-bit FIR filter.

comb and baseband filters. This ΣΔM does two things: firstly, the ΣΔM used has a lowpass signal transfer function that is capable of removing the aliases created by the zero-interleaving process, secondly, the ΣΔM remodulates the multi-bit output signal from the FIR filter back into the single-bit domain.

This structure is claimed to be more efficient in silicon resources that a PCM digital filter up to about 80 taps. The structure still has the complexity of a full precision filter coefficients, this can also increase the word length of the FIR filter output.

The remodulation ΣΔM complexity is discussed by the same authors in [91]. Digital ΣΔM that have lowpass responses are typically not easy to find in the literature. Hence, the authors created their own digital ΣΔM that could be used in the single-bit filter. This digital ΣΔM has a fourth-order architecture and various powers of two multiplications. In [91] even more complex lowpass modulators are presented for the single-bit filter.

The final structure that we look at in this survey is a single-bit IIR filter, first presented in [89, 90]. In these works IIR filters were used to remodulate internal filter states to the single-bit format before multiplication operations

take place. This greatly simplifies the multiplication operation. The structure of the first order IIR filter is shown in Fig.(2.16).

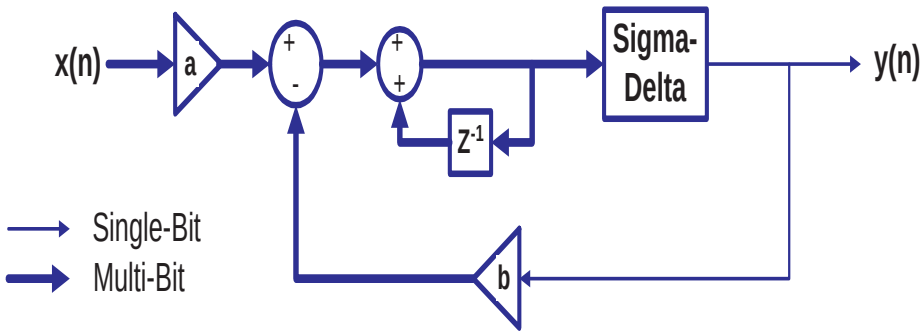

Figure 2.16: Block diagram of the first order single-bit IIR filter.

Originally, the integrator (accumulator) within the feed-forward loop was not present in the system. The ΣΔM is acting as a unit-delay element, hence, the system is a basic first-order recursive filter. However, the system was found to produce only noise in this configuration. The addition of the integrator, as shown in Fig.(2.16), reduces the quantization noise gain in the system while maintaining the same STF and NTF [89, 90].

This system was not further studied by its creators as its spectral shaping capabilities are rather limited and the stopband attenuation is poor. Instead, a new structure of IIR filter was required, a structure that could provide better spectral shaping abilities at high oversampling ratios (OSR's). This was found in the form of a quasi orthonormal state-space structure by the same authors as outlined in [92].

This quasi orthonormal state space IIR architecture was shown to have good filtering abilities with good stopband attenuations. The downside of this structure is that it requires N ΣΔM for an N^{th}-order IIR filter and the structure becomes very complex as the number of ΣΔM increases. This proliferation of ΣΔM only adds to the quantization noise in the band of interest and makes any stability analysis very difficult [89].

2.4 Summary

A brief literature survey is conducted in this chapter on the single-bit and ternary digital filtering. One of the recent trends in DSP systems is to replace the conventional multi-bit PCM systems with short-word length (usually ternary or single-bit) counterparts. Short-word length filtering techniques inevitably include a $\Sigma\Delta$ modulation at some of their stages. The main features of $\Sigma\Delta$M are presented and discussed. Unfortunately, $\Sigma\Delta$ modulators are inherently *non-linear* structures and suffer from many intriguing and unresolved problems. This is mainly due to the harsh quantization process at their outputs.

In this chapter, an emphasis has been made on these drawbacks, especially, the limit cycle phenomenon and stability. Despite the large body of work that has already been done, neither of these topics is well-understood and till now cannot be put in a closed-form. On the other hand (and more importantly), the LMS-like adaptivity in short-word length systems is an unresolved issue. In fact, the adaptivity is the major reason behind the abortion of the attempts to present the single-bit filtering techniques as an alternative to the existing multi-bit ones.

Our task is quite challenging, but this will not forbid us from trying innovative approaches and algorithms to tackle these problems.

Chapter 3

Single-Bit Ternary Filtering Using Sigma-Delta Modulation

3.1 Introduction

Recently a number of techniques for single-bit processing of ΣΔM single-bit streams have been presented [6, 90, 78]. In [78], the author makes use of a fourth-order ΣΔM and a zero-interleaved multi-bit FIR filter. However, this is not as efficient as the ΣΔM based IIR filter in [90]. The latter technique needs only multiplexors, without the parallel multi-bit multipliers that are required by the former technique. On the other hand, the IIR based filter suffers from the disadvantages that the phase is no longer linear, and that the filter is much more vulnerable to coefficient quantization errors than standard FIR filters. To alleviate problems due to the IIR filter coefficient quantization it is proposed in [90] that higher order IIR filters be implemented with quasi-orthonormal structures. These structures require N ΣΔMs if an N^{th}-order IIR filter is to be realized. This proliferation of ΣΔMs greatly reduces the implementation efficiency. In addition, increasing the number of modulators adds to the in-band noise in these structures because, the modulators are the main source of noise in these filters.

3.2 Ternary FIR Filter

The ternary filter is a FIR filter with ternary taps (i.e., +1, -1, or 0). The ternary nature of the taps allows a simple implementation of the FIR filter. This filter is extremely efficient when the input signal to the filter is in single-bit format; each multiplication in the FIR filtering operation can be then implemented in hardware with either a couple of logic gates or a very simple look-up table [6]. The structure of the ternary filter is shown in Figure 3.1.

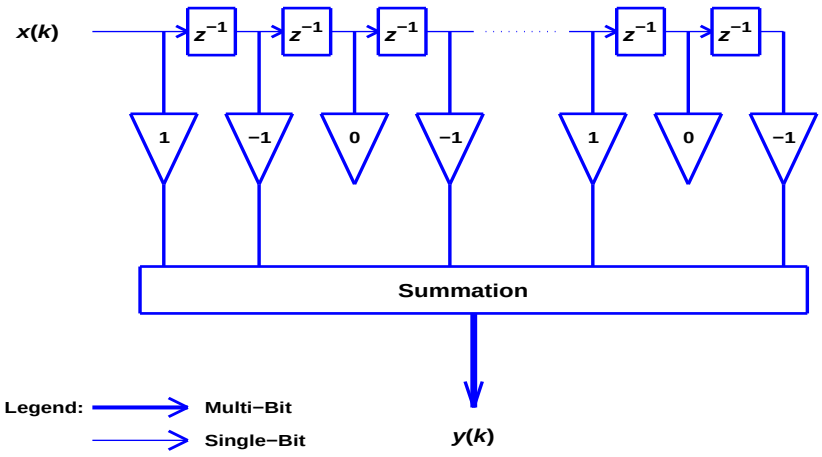

Figure 3.1: Block diagram of a ternary FIR filter.

The FIR filter output $y(k)$ can be described by a convolution of the ternary taps h_i and the input signal $x(k)$. If M is the order of the filter, the output of the filter is:

$$y(k) = \sum_{i=0}^{M} h_i x_{k-i}, \qquad h_i \in \{1, 0, -1\} \tag{3.1}$$

The tap values are generated via $\Sigma\Delta$ modulation of a target impulse response [11, 77, 6], or by using optimization techniques discussed in [9] and [5]. These various methods have their advantages and disadvantageous. Throughout this thesis, it will be assumed that $\Sigma\Delta$ modulation is adopted.

The first step in generation of ternary encoded taps is to generate a *tar-*

get impulse response. The target impulse response should have a low-pass frequency response. Standard FIR coefficient techniques such as the Remez exchange method [19, 96] can be used to attain such an impulse response.

Before a target impulse response is encoded to a ternary format, it must be scaled so that the maximum input to the ΣΔM is operating at its maximum signal-to-quantization-noise ratio. This scaling produces a magnification of the input signal, but this magnification can easily be removed, as will be discussed later.

The digital ΣΔM used to generate the ternary filter taps must meet two criteria. Firstly, a ternary quantizer is required to generate a tri-level output; this has the advantage of higher signal-to-noise ratio (SNR) than the common single-bit quantizer [6]. The second criteria is that the ΣΔM have a flat signal frequency response over the bandwidth f_B of the signal. That is, the ΣΔM should not unduly modify the shape of the impulse response; it should only add quantization noise which is largely confined to the out-of-band region.

The ternary filter requires operation at an oversampled rate (OSR), a requirement that will be met since the input signal is assumed to be a ΣΔ-modulated bit-stream. The structure of the typical second order ΣΔM which can be used to encode the ternary taps is shown in Figure 3.2.

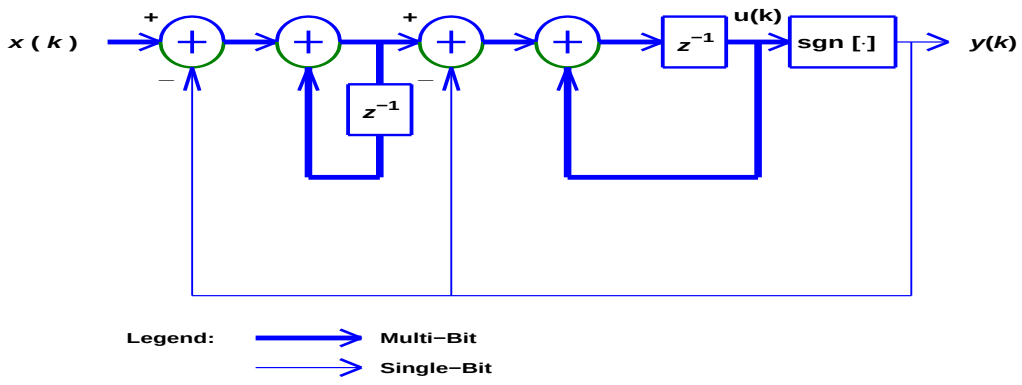

Figure 3.2: Block diagram of the 2nd-order ΣΔ modulator.

The z-domain analysis of the linear system model, the output of the ΣΔM shown in Figure 3.2 is given by:

$$Y(z) = X(z)z^{-1} + Q(z)(1 - z^{-1})^2 \tag{3.2}$$

where $X(z)$ represents the target impulse response and $Q(z)$ represents the quantization noise transfer functions. The noise shaping effect of the ΣΔM is evident from the presence of the filtering term, $(1 - z^{-1})^2$, acting on the noise term, $Q(z)$.The frequency response of the above ΣΔM is given by:

$$H_{\Sigma\Delta T}(e^{j\Omega}) = X(e^{j\Omega})e^{-j\Omega} + Q(e^{j\Omega})(1 - e^{-j\Omega})^2 \tag{3.3}$$

where $\Omega = 2\pi\frac{f}{f_s}$ is the normalized radian frequency.

One advantage of a low-bit resolution system is that the coefficient quantization noise falls in the same spectral region outside f_B as the input signal quantization noise and the remodulating filter quantization noise.

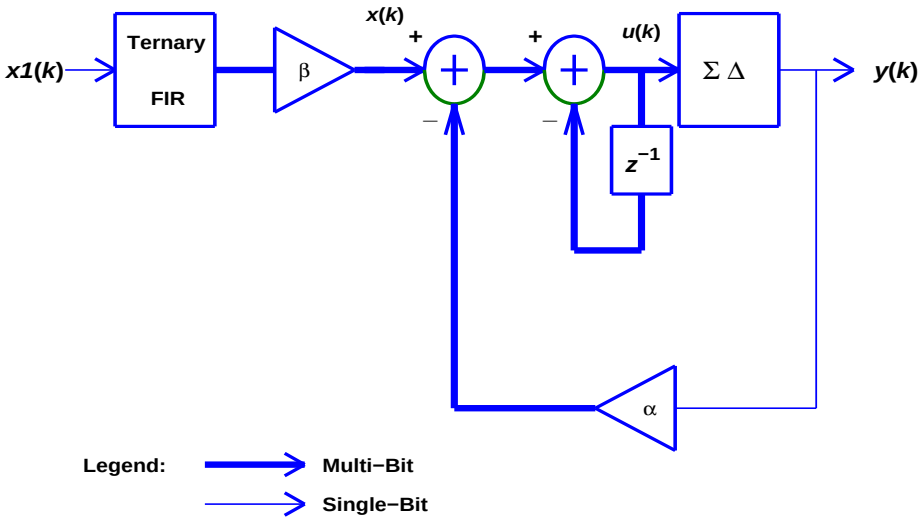

Figure 3.3: Block diagram of the digital ΣΔ FIR-like bit-stream filter proposed in [4].

The ternary FIR filter suffers from two disadvantages. It still contains some high frequency noise due to the coarse quantization of both the impulse response and the input signal. Second it also produces a multi-bit output. Such outputs are not as conducive to efficient hardware processing as single-bit output. To put the output in single-bit format, and to reduce some of the

high frequency noise, a recursive remodulating filter is used as shown in Figure 3.3 [93].

The tasks of remodulating the output of the ternary filter and reducing the high frequency noise cannot by achieved efficiently by the using conventional digital ΣΔM. These modulators generally have an all-pass signal frequency response, hence these modulators may be vulnerable to stability problems caused by high frequency components at the input. This is due to the fact that the high frequency energy components increase the quantizers input variance. As a result, the AC loop gain and stability margin may be reduced [78]. To overcome this difficulty, a structure with a *low-pass* signal transfer function and a single-bit output are required.

In [90], several remodulating structures are proposed. These structures contain an IIR filter with embedded ΣΔM. The simplest such a recursive filter has a first-order IIR structure. The digital ΣΔM used in this filter must only introduce a single delay throughout the system. This arises because the ΣΔM is used as a delay element in the IIR filter, and as such this limits the selection of ΣΔM's. The best choice of ΣΔM should provide good noise shaping at low OSRs. The requirement for a relatively low OSR stems from the fact that, as the OSR increases, the number of ternary taps (i.e., the order of the FIR ternary filter) should be increased to maintain the same frequency response. Hence, a second-order multiple feedback ΣΔM is suited to the task of re-modulation in the IIR ΣΔM filter.

Figure 3.2 shows a second-order ΣΔM used in this filter. This ΣΔM has the same structure as the modulator used to encode the impulse response except that it utilizes a single-bit quantizer. The transfer function of the IIR ΣΔ filter is given below:

$$H_{IIR}(z) = H_{IIRS}(z) + H_{IIRN}(z) \tag{3.4}$$

where H_{IIRS} is:

$$H_{IIRS}(z) = \frac{\beta z^{-1}}{1-(1-\alpha)z^{-1}} \tag{3.5}$$

and H_{IIRN} is:

$$H_{IIRN}(z) = \frac{(1-z^{-1})^3}{1-(1-\alpha)z^{-1}} \tag{3.6}$$

Note that H_{IIRS} and H_{IIRN} represent the signal and noise transfer functions respectively.

The IIR filter coefficient α was set so as to give the transfer function $H_{IIRs}(z)$ a cut-off frequency corresponding to the desired cut-off frequency of the system. The coefficient β is a gain parameter and should be set so that the overall filtering system has a gain of one. Recall that a ternary filter has a gain factor due to the scaling of the impulse response before modulation. This method of determining the IIR filter coefficients is extremely simple. A more accurate (but more complex) method of obtaining the coefficients of the ternary and the IIR filters can be found by optimization (in a least-square sense) to closely approximate a desired frequency response (see [5, 9]).

To determine the spectral filtering abilities of the single-bit IIR filter, a plot at various feedback gains were simulated. A white Gaussian noise signal was input to the first order single-bit IIR filter. The resulting data at the filters output was recorded. A 8192 point fast Fourier transform (FFT) of this output was performed and recorded for three gain "α" values of 0.1, 0.01 and 0.001. An estimate of the filters frequency response was calculated by taking the average of the FFT's for 1000 realizations. The results are shown in Fig.(3.4). Whereas Fig.(3.5) illustrates the theoretical filter frequency responses for the same values of α (i.e., 0.1, 0.01, and 0.001).

As expected, the stopband attenuations deviate from the theoretical response. This is due to the coarse quantization introduced by the double loop $\Sigma\Delta$M. The stopband attenuation is reduced in the simulated results because it becomes swamped by the shaped quantization noise. This reduction in stopband attenuation can also be seen as the filter passband is increased.

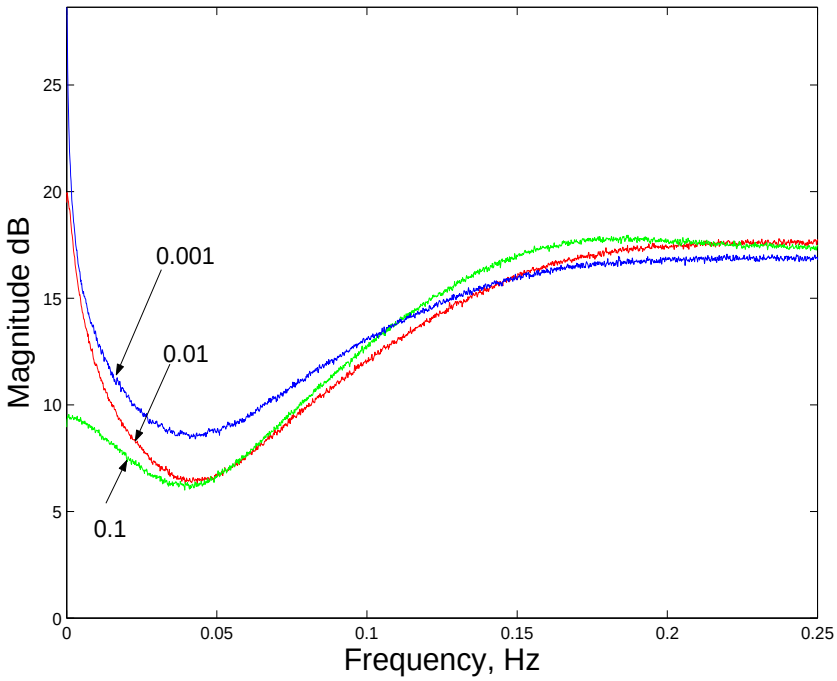

Figure 3.4: Simulated frequency response of the single-bit IIR filter for different values of $\alpha = 0.1$, 0.01, and 0.001.

The only parameter that can be changed in the filtering system is the feedback gain parameter α. This parameter controls the filters passband. Larger values of this gain parameter provide larger passbands. However, the stopband attenuation is reduced as the filters passband is increased; this is again caused by the shaped quantization noise from the $\Sigma\Delta$M swamping the filters transition and stop bands.

3.3 Summary

In this chapter a bit-stream filtering structure is introduced. It consists of a ternary FIR filter cascaded with an IIR $\Sigma\Delta$M structure. This structure is being the basis of many single-bit DSP applications. Since many of the ternary filter tap values are zero and each non-zero tap requires only very simple multiplication hardware, the system is very resource efficient and fast as no complex math operations are required. Performance enhancement is possible through

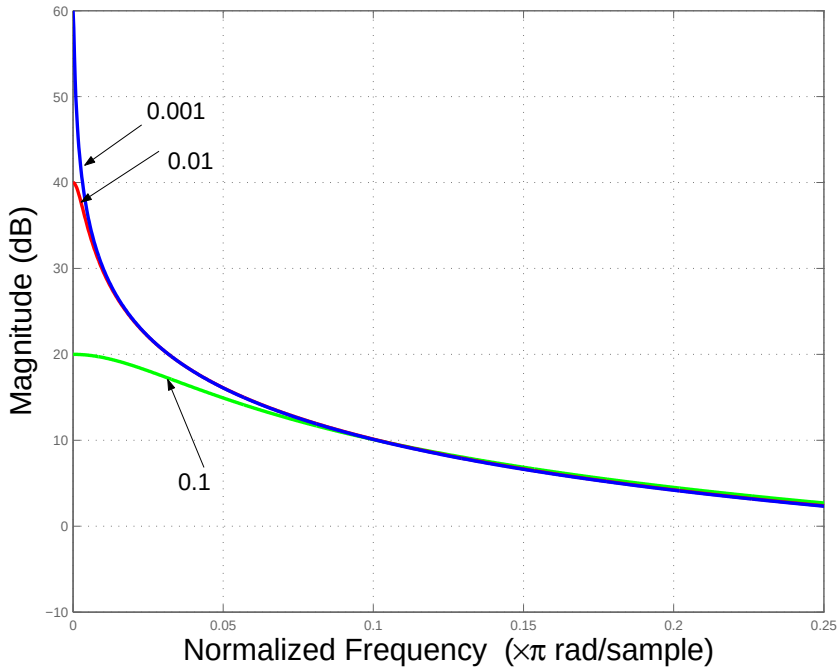

Figure 3.5: Theoretical frequency response of the single-bit IIR filter with $\alpha = 0.1$, 0,01, and 0.001.

increasing the oversampling ratio, however, this requires increasing the number of taps and the sampling rate of the system, hence, there is an inherent trade-off between hardware efficiency and performance.

Chapter 4

DSP Applications Using Single-Bit Filtering: Comb Filtering

4.1 Introduction

Comb filters found a wide range of applications such as the suppression of clutter from fixed objects in moving-target indicator radars and the rejection of power-line harmonics in the promising technology of using the power line communication (PLC) as a third pipe to deliver broadband access to home and business. Comb filters are usually constructed using multi-bit architectures.

Single-bit $\Sigma\Delta$modulation have recently received increased attention because of their good performances and efficient VLSI implementation.

In this chapter, two structures for single-bit digital comb filtering are proposed. The first structure is based on ternary filtering, however, the output of the filter is in single-bit format. The second structure is based on second-order sigma-delta modulation ($\Sigma\Delta$M).

A design method based on the standard $\Sigma\Delta$ topology has been then presented and used to construct efficient filters in the sense of improved quantization noise reduction. This is done by introducing a gain parameters in the feedback loop, i.e., introducing poles in the noise transfer function. This method is used to design a $\Sigma\Delta$-based comb filter. All of the presented filtering structures contain no multi-bit multiplication, making the comb filters efficient for implementation. These filters can be utilized in a wide range of promising

applications.

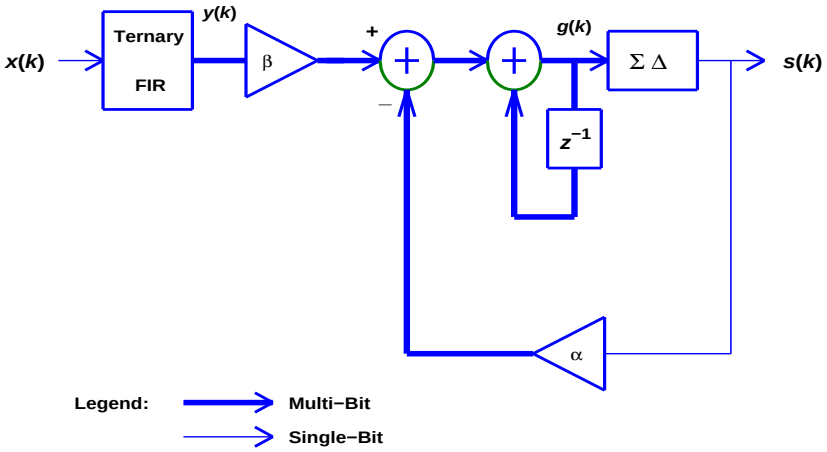

Figure 4.1: Block diagram of the ternary ΣΔ filter.

4.2 A Proposed Ternary-Sigma-Delta Comb Filter

The single-bit comb filter is designed according to the configuration shown in Fig.(4.1), which can accept both multi-bit and single-bit input formats. The ternary filter is an FIR filter with ternary taps (i.e., +1, 0, -1) [9]. This ternary format allows a simple implementation of the FIR filter; it is most efficient when the input signal is in single-bit format. The structure of the ternary filter is shown in Fig.(4.2). The ternary filter output $y(k)$ is given by the convolution of ternary taps $\{h(i)\}$ (or simply $\{h_i\}$) and the input signal $\{x(k)\}$ as follows:

$$y(k) = \sum_{i=0}^{M} h_i x_{k-i} \tag{4.1}$$

where M is the order of the filter. The tap values are generated via ΣΔ modulation of a target impulse response. The digital ΣΔM used for this purpose must have tri-level output, and must have a flat signal frequency response over the bandwidth of interest [87]. The ternary filter requires operation at

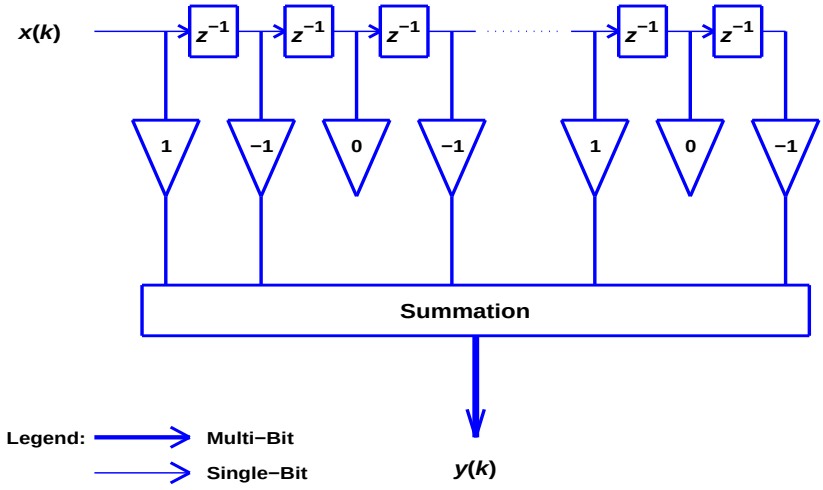

Figure 4.2: Block diagram of a ternary FIR filter.

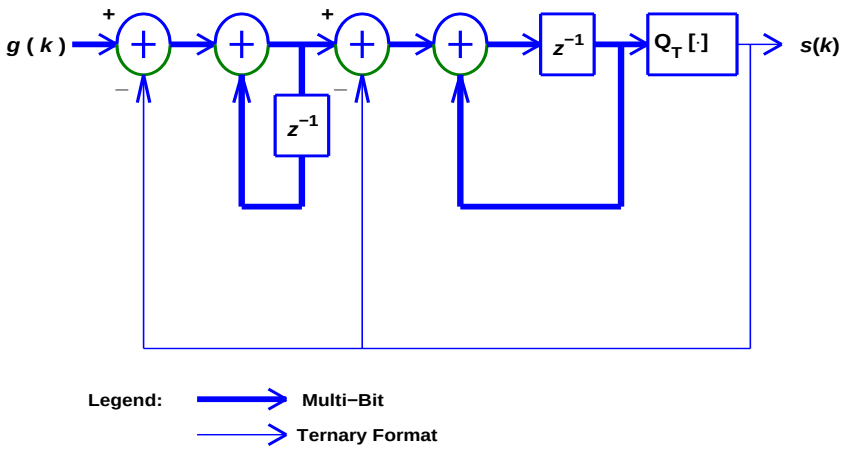

Figure 4.3: Block diagram of a second-order ΣΔM with ternary quantizer $Q_T[.]$.

an oversampled rate (OSR), a requirement that will be met since the input signal is assumed here to be a ΣΔ modulated bit-stream. The structure of the typical second-order ΣΔM which can be used to encode the ternary taps is shown in Fig.(4.3). To analyze the performance of ΣΔM, an approximate quantization noise model, referred to as the input-independent additive white noise approximation, is normally used. In this linear model, the quantization noise $q(n)$ is assumed to be uniformly distributed between $-\delta$ and δ (where δ is the quantization step-size) in the band of interest ($-f_s/2 \leq f \leq f_s/2$). The larger the OSR is, the better this assumption will be. Theoretically, in

second-order modulator, each doubling of sampling rate achieves a signal-to-quantization-noise ratio (SQNR) improvement of about 15 dB [88].

The z-domain transfer function of this ΣΔM model is given by:

$$H(z) = G(z)z^{-1} + Q(z)(1 - z^{-1})^2 \tag{4.2}$$

where $G(z)$ and $Q(z)$ represent the signal and the quantization noise transfer functions, respectively. The noise shaping filter, $(1 - z^{-1})^2$, attenuates the quantization noise in the signal band and amplifies it outside the signal band (higher frequencies). These high-frequency noise components can be eliminated by a subsequent digital filtering that also decimates the sample rate.

From (4.2), the frequency response is given by:

$$H(e^{j\Omega}) = G(e^{j\Omega})e^{-j\Omega} + E(e^{j\Omega})(1 - e^{-j\Omega})^2 \tag{4.3}$$

where $\Omega = 2\pi f/f_s$ is the normalized radian frequency.

The response of the overall system H_{ov} will be the combination of the frequency response of the ternary filter $H_T(e^{j\Omega})$ and the frequency response of the IIR-ΣΔM filter $H_{\text{IIR}}(e^{j\Omega})$ as follows [93]:

$$H_{\text{ov}}(e^{j\Omega}) = H_T(e^{j\Omega}) \cdot H_{\text{IIR}}(e^{j\Omega}). \tag{4.4}$$

From (4.3) and 4.4 we get:

$$H_{\text{ov}}(e^{j\Omega}) = H_T(e^{j\Omega}).[H_{\text{IIRS}}(e^{j\Omega}) + H_{\text{IIRN}}(e^{j\Omega})] \tag{4.5}$$

where $H_{\text{IIRS}}(e^{j\Omega})$ and $H_{\text{IIRN}}(e^{j\Omega})$ are the signal and noise parts of $H_{\text{IIR}}(e^{j\Omega})$, respectively. Now $H_{\text{ov}}(e^{j\Omega})$ can be expressed as follows:

$$H_{\text{ov}}(e^{j\Omega}) = \frac{G(e^{j\Omega})\ K(e^{j\Omega})}{D(e^{j\Omega})} \ + \ \frac{E(e^{j\Omega})\ P(e^{j\Omega})}{D(e^{j\Omega})} \tag{4.6}$$

where

$$K(e^{j\Omega}) = e^{-j\Omega} + e^{-2j\Omega}(\beta - 3) + 3e^{-3j\Omega} - e^{-4j\Omega} \quad (4.7)$$

$$D(e^{j\Omega}) = 1 - (1 - \alpha)e^{-j\Omega} \quad (4.8)$$

$$P(e^{j\Omega}) = e^{-j\Omega}(\beta - 5) + e^{-2j\Omega}(10 - 2\beta) + e^{-3j\Omega}(\beta - 10) + 5e^{-4j\Omega} - e^{-5j\Omega} \quad (4.9)$$

noting that α and β are the multiplication constants shown in Fig.(4.1).

Experimental results in analog-to-digital conversion (ADC) indicate that the signal-to-quantization-noise ratio (SQNR) can be improved by using a ternary quantizer in the feed-back loop. The extent of improvement depends on the quantizer characteristics, the thresholds, and the output level [94].

The IIR-based filter, however, suffers from the disadvantages that the phase is no longer linear, and that the filter is more vulnerable to coefficient quantization errors than standard FIR filter.

4.2.1 Design and Simulation of Ternary Sigma-Delta Comb Filter

The steps to design a ternary filter can be summarized as follows [95]:

1. Generate the FIR filter coefficients that satisfy the required specifications using Remez Exchange algorithm.
2. Interpolate the FIR filter coefficients by a factor of OSR (to oversample the target impulse response to the desired OSR). Several techniques can be used, such as spline, FFT, linear, and cubic. Spline method is adopted here.
3. Ternary encode the filter coefficients, where a set of ternary-valued coefficients are generated from the interpolated filter coefficients.
4. Upsample the input signal by a factor of OSR.
5. Remodulate the output of the ternary filter to single-bit format using $\Sigma\Delta$ M.

The ternary filter requires operation at an oversampled rate (OSR), and this will be met as the input signal is assumed to be a $\Sigma\Delta$ modulated bit-stream. The number of taps, M, is usually the same as the upsampling ratio, OSR, times the Nyquist rate filter order. However, a higher value of M could also be selected at the expense of increasing the delay of the filter, which is inversely related to the bandwidth. If M is too large, the high frequency contents of the signal will be attenuated.

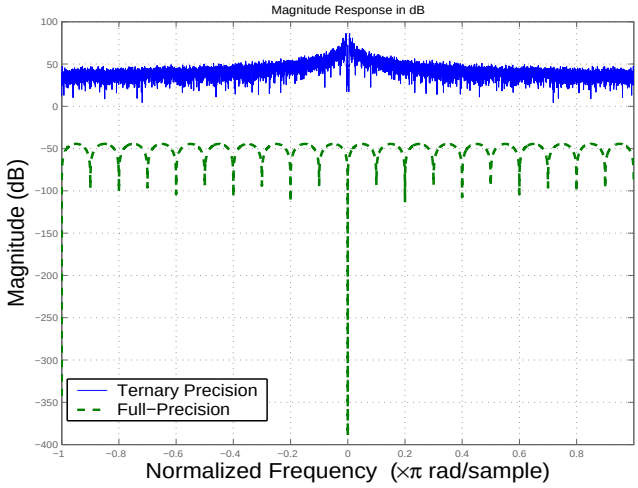

Figure 4.4: Frequency response of the ternary filter in the proposed comb structure.

In our simulation, the proposed single-bit comb filter is designed, as an example, to attenuate the effect of the 10^{th}-order harmonics in narrow-band signals transmitted over power-lines. A direct form FIR filter is designed for this purpose and used as a target impulse response.

Fig.(4.4) shows the frequency response of the FIR filter with full precision and the oversampled ternary precision coefficients (OSR=128). The simulated frequency response of the single-bit comb filter is shown in Fig.(4.5). The gain factor β is out of our interest here and is assumed to be constant equal to 0.001. The phase response of the filter system can be seen in Fig.(4.6). The non-linear effect of the $\Sigma\Delta$M stage on the phase performance of the overall combination is apparent.

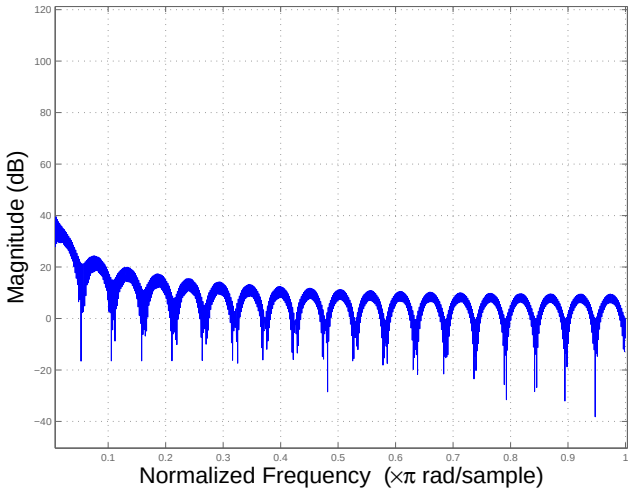

Figure 4.5: Frequency Response of the proposed ternary-ΣΔ single-bit comb filter with OSR = 128.

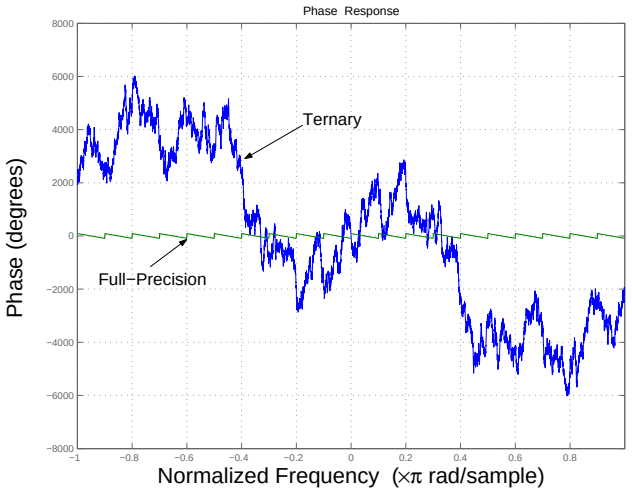

Figure 4.6: Phase response of the proposed single-bit ternary-ΣΔ comb filter.

4.3 A Proposed Sigma-Delta Comb Filter

A comb filter can be created by taking an FIR filter with the following system function [96]

$$H(z) = \sum_{k=0}^{M} h(k) z^{-k} \tag{4.10}$$

then replacing z by z^p, where p is a positive integer. Thus, the new FIR filter has a system function as follows:

$$H_c(z) = \sum_{k=0}^{M} h(k) z^{-kp}. \tag{4.11}$$

If the frequency response of the original FIR filter is $H(e^{j\omega})$, the frequency response of the new FIR in (4.11) is given by

$$H_c(e^{j\omega}) = \sum_{k=0}^{M} h(k) z^{-jkp\omega} \tag{4.12}$$

i.e.,

$$H_c(e^{j\omega}) = H(e^{jp\omega}). \tag{4.13}$$

Consequently, the frequency response $H_c(e^{j\omega})$ is simply p-order repetition of $H(e^{j\omega})$ in the range $0 \leq \omega \leq 2\pi$. Hence, if we replace z^{-1} by z^{-M} directly in the second-order $\Sigma\Delta$M shown in the Fig. (4.7), the transfer function of the corresponding discrete-time linear model can be given as follows [97]

$$H(z) = X(z) + Q(z)(1 - z^{-M})^2. \tag{4.14}$$

The noise-shaping filter, $(1 - z^{-M})^2$, is a comb filter with notches at frequencies $2\pi k/M$, where $k = 0, 1, 2, \cdots, M-1$. This filter can be used for any signal that has narrow-band frequency components in these locations. Fig. (4.8) shows the signal transfer function, $X(z)$, and the noise transfer function, $Q(z)$, according to (4.14).

The simulated frequency response of this structure (using the same sam-

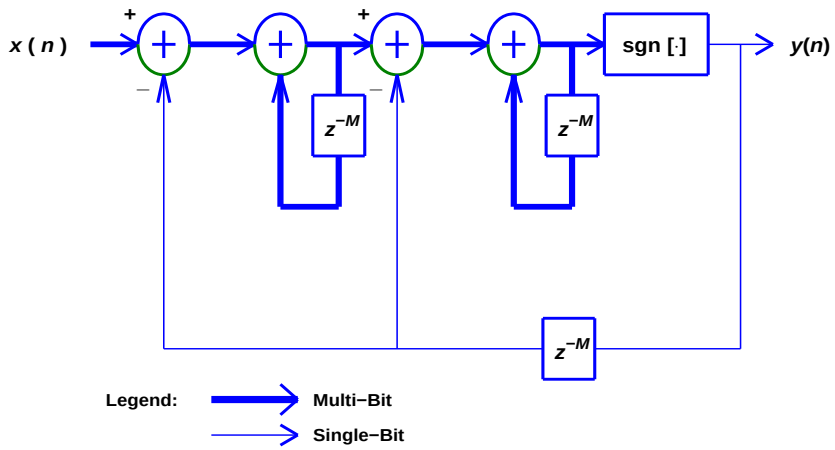

Figure 4.7: Block diagram of the proposed ΣΔM structure.

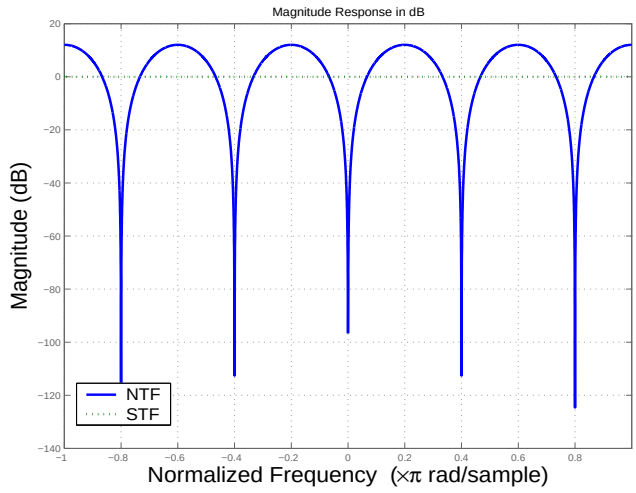

Figure 4.8: Noise and signal transfer functions of the proposed ΣΔ single-bit comb filter.

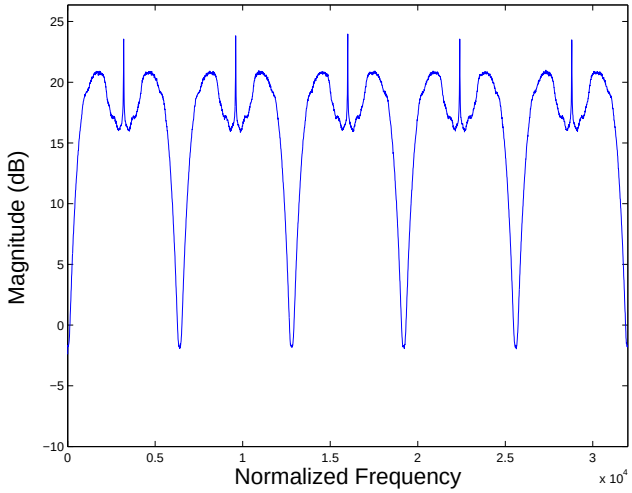

Figure 4.9: Frequency Response of the proposed ΣΔ single-bit comb filter.

pling rate and design requirements as in the ternary-ΣΔ structure) is shown in Fig. (4.9). Fig. (4.10) depicts the phase response of this structure.

The proposed structures lend themselves well in broad-band applications such as PLC, as they are efficient in hardware implementation with high performance.

From a hardware viewpoint, both of the proposed structures for digital comb filtering have the advantage of simple implementation, as there are no multi-bit addition or multiplication operations in their structure, however, the second structure is even simpler to construct.

4.4 A Proposed Design Approach

In this section, an alternative design method is introduced to design ΣΔ-based comb filter with optimized noise-shaping effect. The main idea behind this technique is to introduce poles into the noise transfer function in such away to improve SQNR. It should be emphasized that the proposed method is a general approach and can be used to design various single-bit ΣΔ-based system as well.

Single-bit processing has been attracting interest due to the promise of ef-

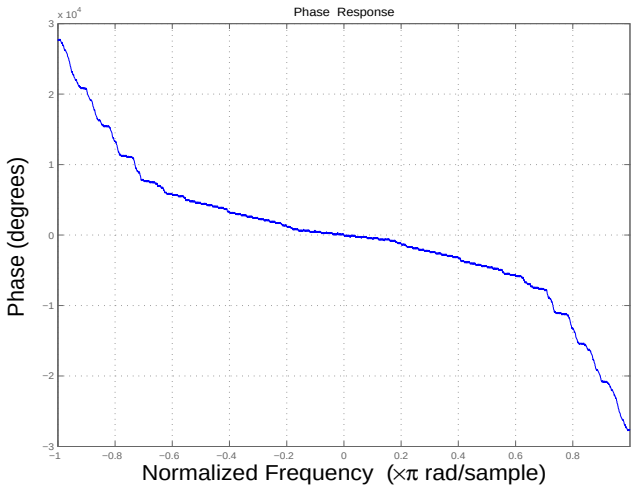

Figure 4.10: Phase Response of the proposed ΣΔ single-bit comb filter.

ficient and simple implementation. Sigma-delta modulators (ΣΔM's) are the main single-bit modulators or analog-to-digital converters (ADC's). However, a drawback of the ΣΔM system is that high resolution can only be obtained for low to medium bandwidths [2]. This is so because the oversampling ratio (OSR) should be high for better resolution (might be several orders of magnitude higher than the Nyquist rate). Hence, It is difficult to handle broadband applications, such as broadband power-line communication (BPC), using an ordinary ΣΔM.

The signal-to-quantization-noise (SQNR) can be improved by either increasing the order of the ΣΔM or increasing the OSR. Therefore, one approach to alleviate the low bandwidth bottleneck is to use a higher order ΣΔM with a reduced OSR. However, this will introduce the problem of unpredictable instability that is seemingly inherent in such high-order ΣΔ systems [98].

When designing a ΣΔM, the noise-transfer function (NTF) should be given significant consideration. In general, it is designed as a high-pass function so that the quantization noise can be moved to higher frequency bands. In general, NTF can be classified into two classes:

1. Pure differentiation of order M: In this case the noise-shaping function (noise transfer function) can be expressed as follows:

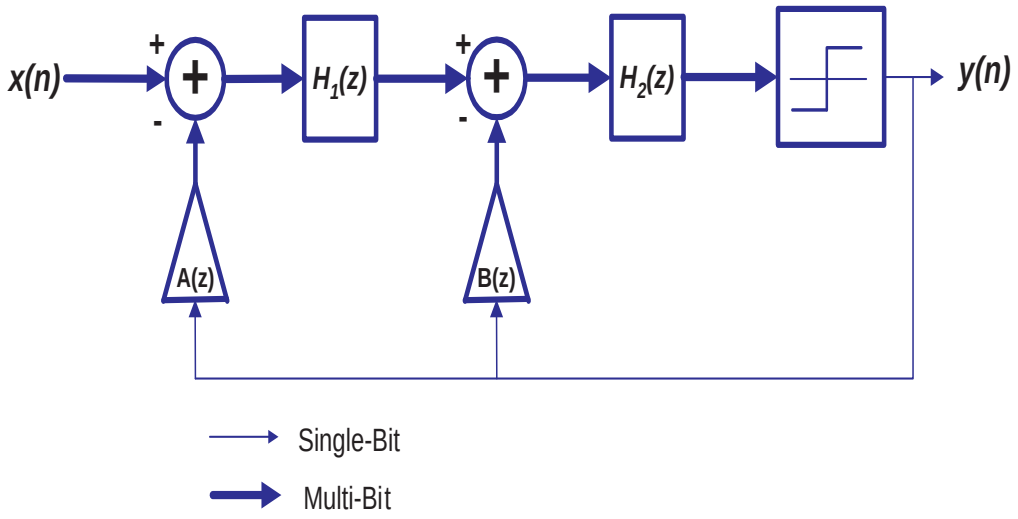

Figure 4.11: Block diagram of a second-order ΣΔM.

$$N(z) = (1 - z^{-1})^M. \tag{4.15}$$

As the order M increases, more noise power will move to high frequency bands, hence, noise in the low frequency bands will be reduce and, consequently, SNQR in the baseband is increased. Moreover, SQNR can be improved by increasing OSR. Due to the characteristics of this type of NTF, its usage is usually limited in mid and low bandwidth applications such as audio applications.

2. non-monotonic transfer function: Pure differentiation response can be modified by introducing poles into NTF as follows:

$$N(z) = \frac{(z-1)^M}{D(z)}. \tag{4.16}$$

In this case the NTF is an M^{th}-order polynomial with a leading coefficient of 1. It is simply a high-pass function, where the coefficients of the ΣΔM can be designed using analog filter techniques.

4.4.1 Theory and Design

The design of single-bit ΣΔM is a non-trivial task. Many works in the literature have reported methods to estimate the performance of ΣΔM analytically [54]. However, these methods only approximate the actual behavior of ΣΔM.
To characterize the modulator it is common to look at the STF and NTF. The STF describes how the modulator alters the original input signal spectrum and it is ideally, STF is unity. In a similar manner the NTF describes how the modulator shapes noise away from the center frequency, f_c. For a low-pass modulator, $f_c = 0$ Hz (DC), and for a band-pass modulator f_c is often equal to $f_s/4$ for simpler design.

The NTF is the main design task which determines the amount of baseband noise shaping performed by the modulator.

Fig.(4.11) shows a general second-order ΣΔM which contains two M^{th}-order FIR filters (can be assumed of different order as well)in its feedback loop to tune its response. Based on the linear model of ΣΔM, the z-transfer function of the above system can be found as follows:

$$Y(z) = \frac{X(z) + \frac{Q(z)}{H_1(z)H_2(z)}}{D(z)} \tag{4.17}$$

where $D(z)$ is given by:

$$D(z) = A(z) + B(z)\frac{1}{H_1(z)} + \frac{1}{H_1(z)H_2(z)} \tag{4.18}$$

with $A(z)$ and $B(z)$ being the transfer functions of the FIR filters (whose coefficients are $\{a_i | i = 0, \cdots, M\}$ and $\{b_i | i = 0, \cdots, M\}$ as follows:

$$A(z) = \sum_{i=0}^{M} a_i z^{-i} \tag{4.19}$$

$$B(z) = \sum_{i=0}^{M} b_i z^{-i}. \tag{4.20}$$

From above the signal and noise transfer functions can be expressed respectively as follows:

$$S(z) = \frac{1}{D(z)} \tag{4.21}$$

$$N(z) = \frac{1}{H_1(z)H_2(z)D(z)}. \tag{4.22}$$

Now, depending on the corresponding topology required for the standard second-order ΣΔM, the transfer functions $H_1(z)$ and $H_2(z)$ can take any of the following forms:

$$H_1(z) = H_2(z) = \frac{z^{-1}}{1 - z^{-1}} \tag{4.23}$$

$$H_1(z) = H_2(z) = \frac{1}{1 - z^{-1}} \tag{4.24}$$

$$H_1(z) = \frac{1}{1 - z^{-1}}; \quad H_2(z) = \frac{z^{-1}}{1 - z^{-1}}. \tag{4.25}$$

These topologies were found to have identical characteristics regarding noise shaping [99]. The only difference among them is the delay factor (z^{-1}) and the scaling gain. For instance, if we adopt the first form, $D(z)$ will be given as follows [25]:

$$D(z) = 1 + (a_o - 2)z^{-1} + (1 - b_o + b_1 + a_o)z^{-2} + G(z) \tag{4.26}$$

where $G(z)$ is given by:

$$G(z) = \sum_{i=1}^{M-1} b_{i+1}z^{-i-2} + \sum_{i=1}^{M}(a_i - b_i)z^{-i-2}. \tag{4.27}$$

For $a_o = 1$ and $b_o = 2$, the structure will be reduced to the standard ΣΔ topology, i.e., $D(z) = 1$, which implies two poles at $z = 0$. For coefficient values other than $a_o = 1$ and $b_o = 2$, $D(z)$ will be a second-order polynomial in z^{-1}, providing M equations with $2M$ unknown coefficients. These coefficients can be found using different approaches [100]. However, $D(z)$ will not increase

the order of noise shaping in the transfer function of the ΣΔM, but it may improve the stability of the system if it is well-designed.

4.4.2 The Proposed Structure

To improve system performance, $D(z)$ should be designed as an FIR low-pass filter to reduce the height of voltage steps at the output of the integrators [25], i.e., the input signal to $D(z)$ should not be attenuated at low frequency (when $z \rightarrow 1$). If we assume $D(z)$ as an $(M+1)^{\text{st}}$-order FIR filter with coefficients $\{d_i | i = 0, \cdots, M\}$ as follows:

$$D(z) = \sum_{i=0}^{M} d_i z^{-i}, \tag{4.28}$$

Then, (in a semi-digital implementation, where the coefficients are implemented by analog means) a comb filter can be produced if the coefficients are equal, i.e., $d_0 = d_1 = \cdots = d_M = \frac{1}{M+1}$. This means that we impose the following condition on the coefficients of $D(z)$:

$$\sum_{i=0}^{M} d_i = 1. \tag{4.29}$$

In this section, our main intention is to design a single-bit digital comb filter for the purpose of efficient hardware implementation .
Now, the next step is to select proper functions to represent $H_1(z)$ and $H_2(z)$. As the z-transfer function of the comb filter is basically composed of equally spaced zeros around the unit circle circumference, then refereing to eqn.4.17 we should chose $H_1(z)$ and $H_2(z)$ such that they match the required frequency response of the NTF as follows:

$$H_1(z) = H_2(z) = \frac{1}{(1 - z^{-M})}. \tag{4.30}$$

We choose $H_1(z)$ and $H_2(z)$ to have the same transfer function for conve-

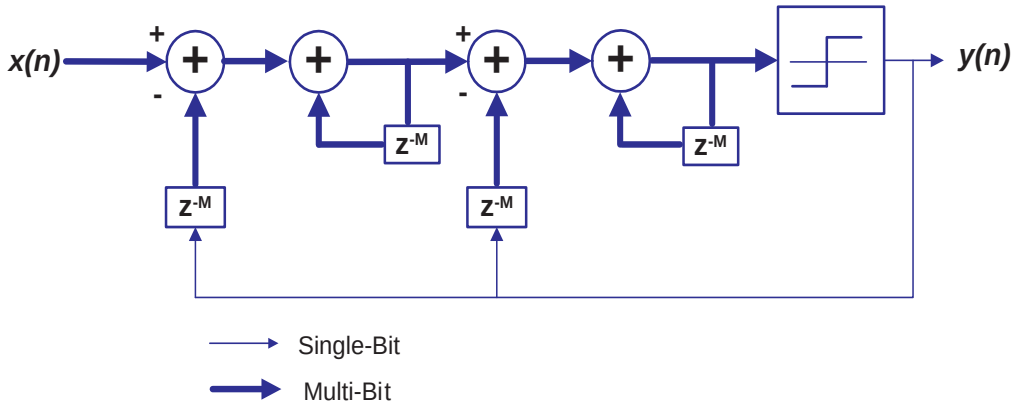

Figure 4.12: Block diagram of the proposed single-bit ΣΔ based digital comb filter.

nience but not necessarily. Then, the output of the ΣΔM will be as follows:

$$Y(z) = \frac{X(z) + Q(z)(1 - z^{-M})^2}{D(z)} \tag{4.31}$$

where $D(z)$ is now given by:

$$D(z) = A(z) + B(z)(1 - z^{-M}) + (1 - z^{-M})^2 \tag{4.32}$$

with $A(z)$ and $B(z)$ as defined earlier. For simplicity, we propose $A(z) = B(z)$, but of course this is not the case always, it depends on the application to be achieved. Therefore, if $\sum_{i=0}^{M} d_i = 1$ as we proposed earlier, then after a few arithmetical manipulations we can find $A(z)$ as follows:

$$A(z) = B(z) = z^{-M}. \tag{4.33}$$

This implies that in this case both $\sum_{i=0}^{M} a_i = 1$ *and* $\sum_{i=0}^{M} b_i = 1$ and these functions are represented by a pure M-delay line.

Now, we re-design $D(z)$ such that it introduces some poles into the noise transfer function to comply with the non-monotonic noise transfer function type mentioned above (item-2). This can be carried out simply by putting $A(z) = B(z) = 1$ as is the case with a standard ΣΔM, where $D(z)$ will be

given by:

$$D(z) = 3 - z^{-M} + z^{-2M}. \tag{4.34}$$

In this case $D(z)$ will introduce $2M$ poles into the NTF. These poles are distributed uniformly as conjugate pairs around a circle inside the unit circle on the z-plane. The radius r of this circle is $r = 0.577$. Therefore, it is expected that $D(z)$ will contribute to shaping the noise as well as to changing the STF in such away that converts the frequency response into M-period resonator.

The designed single-bit digital comb filter is shown in Fig.(4.12) which is identical to that previously designed structure in Section 4.2. While the structure of the designed M-period resonator is shown in Fig.(4.13)

Fig.(4.8) shows the same signal and noise frequency transfer functions STF($e^{j\Omega}$) and NTF($e^{j\Omega}$) obtained from equations (4.31) and (4.33) with $M = 10$ as compared to those obtained from equations (4.31) and (4.34) with $M = 10$ which can be seen at Fig.(4.14). From these two figures we can expect the role that the NTF can play in tuning the $\Sigma\Delta$ system response.

4.4.3 Simulation and Discussion

To verify the above analytic results, the proposed single-bit comb filter and multi-period resonator are simulated using MATLAB. To avoid redundancy, the frequency response of the single-bit $\Sigma\Delta$M based comb filter for $M = 10$ and OSR, $R = 64$ can be seen in Fig.(4.9).

Fig.(4.15) depicts the frequency response of the M-period single-bit resonator for $M = 10$ and OSR, $R = 64$.

The signal-to-quantization-noise ratio (SQNR) is an essential performance measure for $\Sigma\Delta$M. The in-band SQNR is given in [101] as follows:

$$\mathrm{SQNR}_{\mathrm{in-band}} = \frac{2\displaystyle\int_{0}^{0.5} |X(e^{j2\pi v})|^2 dv}{\displaystyle\int_{-1/(2\,\mathrm{R})}^{1/(2\,\mathrm{R})} |\mathrm{N}(e^{j2\pi v})|^2 dv} \tag{4.35}$$

where $X(e^{j2\pi v})$ is the Forier transform of the (oversampled) input signal $x(i)$. The SQNR can be estimated empirically. To do this, first the input signal spectrum must be removed from the output and replaced by interpolating the end points. Second the actual noise transfer function should be found to evaluate the SQNR as given by the expression above using Hannining-windowed FFT's. Sinusoidal inputs are used in this test. The input signal spectrum is chosen such that its spectral energy lies within a single FFT bin. Fig.(4.16) shows the simulated SQNR as a function of OSR for sinusoidal inputs. As expected [80], doubling the sampling frequency reduces the noise power, theoretically, by about 15 dB, of which 3 dB is due to the reduction in power spectral density of the quantization noise, with the additional 12 dB due to the action of the NTF.

This ΣΔM topology lends itself well to broadband frequency applications, such as Broad-Band Power-line Communication (BPL). This also suggests that the proposed single-bit ΣΔ comb filter can be utilized with relatively low OSR if the input frequency is high. In [97], we noticed that a similar structure has been proposed for UWB-OFDM applications.

Fig.(4.17) shows the SQNR as a function to the amplitude of the input signal for different OSRs. It can be seen clearly that the SQNR collapses at an absolute input level less than 0.3dB. This gives a boundary of the input dynamic range for stable operation.

4.5 Stability of the Proposed Structures

Linear analysis has been used to model the quantization noise in ΣΔ systems [6]. Though useful, the linear model is unable to model the system well enough to predict the stability and performance for a given design. This is due to the non-linear behavior of the ΣΔ systems. However, attempts to better predict the behavior of ΣΔM using non-linear analysis techniques have produced promising results [102]. Initial simulation results showed that the proposed

structures are stable, however, full analysis based on non-linear analysis will be handled in chapter 7

The stability of the system is decided by the poles in its transfer function. The single-bit comb filter does not contain prominent poles since $D(z) = 1$ implies two trivial poles at the center. Moreover, all zeros lies on the unit circle. On the other hand, the M-period resonator possesses $2M$ poles and all these poles are located inside the unit circle, in addition to the same zeros as in the NTF of the comb filter.

The stability of the modulator is assessed by looking at the quantizer input $x_q(n)$ Knee plot proposed in [103]. These were used to find which input values would result in the divergence of the quantizer input towards infinity. From which we may expect that this ΣΔM comb filter is to be stable as long as the input signal amplitude is limited by $|x_q| < 2$. Fig.(4.17) reveals this situation.

4.6 Summary

Two structures for single-bit digital comb filtering are proposed and simulated. In the first structure (Section 4.2), a comb filter is designed based on ternary filtering such that both the input signal and the target impulse response are encoded using a ΣΔ modulator. The second structure (Section 4.3) is based on a second-order ΣΔ modulator. The frequency response obtained in both cases is very near to the required response of a comb filter. The proposed filters can be built using simple hardware, and hence they are potentially suitable for VLSI implementation. They are also suitable for broadband applications such as power-line communications.

In Section 4.4 a design technique for single-bit systems using a feedback path filter to tune the response of the ΣΔ modulator was proposed. A single-bit digital comb filter is designed and its performance is evaluated in terms of signal-to-quantization noise ratio (SQNR), the dynamic range (input signal level), and stability. Moreover, we showed that the same design technique can

be used for other single-bit systems, where we used it to design a multi-period resonator. It was shown that the proposed filters lend themselves very well to broadband input signals and can be utilized in emerging technologies such as the Broad-Band Power-line Communication (BPL).

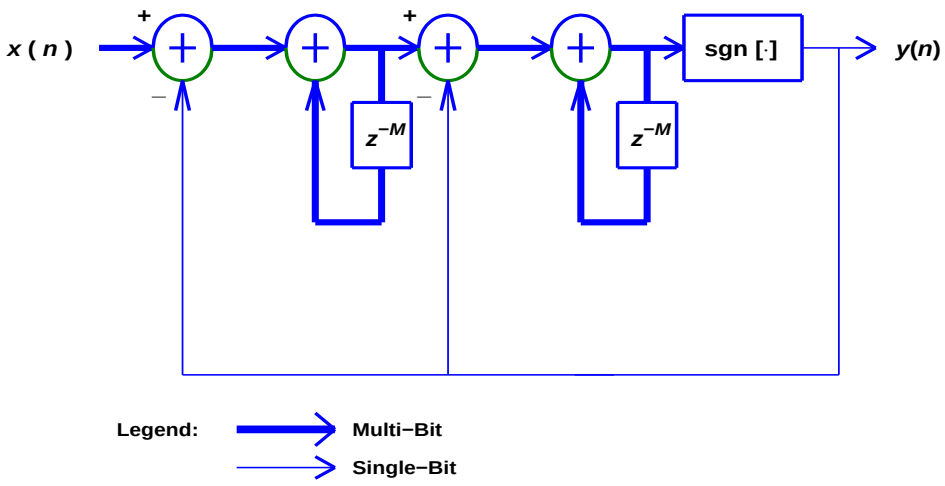

Figure 4.13: Block diagram of the designed M-period resonator.

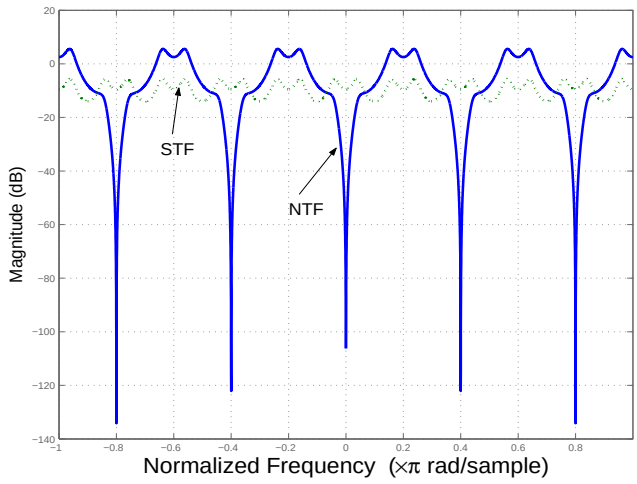

Figure 4.14: Noise transfer function, NTF($e^{j\Omega}$) and signal transfer function, STF($e^{j\Omega}$) for the designed single-bit M-period resonator

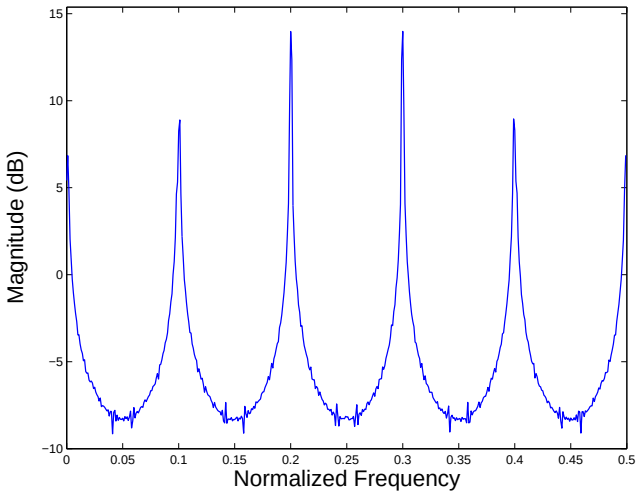

Figure 4.15: Frequency Response of the proposed single-bit M-period resonator filter order $M = 10$ and OSR, $R = 64$.

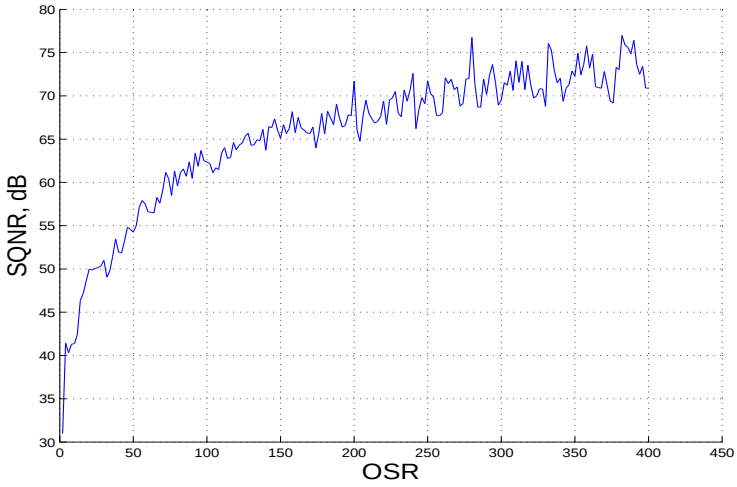

Figure 4.16: SQNR against OSR with input signal amplitude of 0.5.

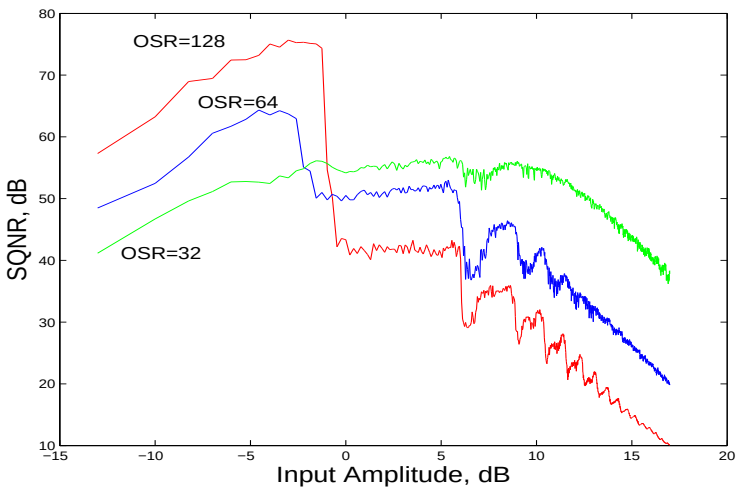

Figure 4.17: SQNR of the proposed single-bit comb filter versus the input amplitude for different values of OSR.

Chapter 5

DSP Applications Using Single-Bit Filtering: DC Blocking

5.1 Introduction

Single-bit hardware implementations have found increasing application in audio and other digital signal processing (DSP) systems as single-bit systems possess very attractive properties as compared to their multi-bit counterparts. The rising popularity of these single bit implementations is due to i) the fact that they can shape unwanted quantization noise out of the spectral region of interest and ii) their hardware resource efficiency (as compared to multi-bit systems). There are, however, a number of complex design issues associated with single-bit systems. One of the issues complicating the design is the fact that $\Sigma\Delta$ modulators (which are critical components in many single bit systems) deteriorate in performance when they are driven by DC-biased signals. The DC-bias affects the dynamic range of $\Sigma\Delta$ modulators and enhances unwanted limit cycles. The overall system stability is thus compromised.

It should be acknowledged that some of these limit cycles do not adversely affect operation of $\Sigma\Delta$ modulator. These "benign" limit cycles are known as idle patterns, which usually correspond to frequencies located faraway from the baseband, and can therefore easily be disturbed by applying an input to the system [32]. On the other hand, there are other large signal limit cycles which have steady or diverging amplitudes at frequencies often located near

the baseband. Therefore, it is very difficult, if not impossible, to disturb these cycles, which are audible in audio applications [43]. Of course, the DC content can easily be removed from multi-bit signals by using traditional DC cancellers, i.e., before encoding in single-bit format. Subsequent processing may, however, re-introduce DC components, and it is often critical to remove these.

Unfortunately, the design in the single-bit domain has been suffering from two obstacles. First, there are still several unresolved problems such as adaptivity and stability. Second, the design itself is not straightforward as in multi-bit techniques. However, we do expect that, ultimately, these pitfalls would be tackled in no far future, and the single-bit or at least the short word-length signal processing (DSP) systems would become very popular.

Several relevant previous works have studied and proposed different single-bit structures (e.g., [87],[104]). Among the problems that one might face practically is the development of DC component at various stages in the signal bitstream. A DC biased bitstream has a highly undesired impact on the performance of the single-bit system, as the DC content bears no information and enhances unwanted limit cycles (which may in turn affect system stability).

This chapter introduces two different approaches to eliminate DC content from a $\Sigma\Delta$ modulated bitstream. In the first approach, a single-bit digital ternary filtering DC-blocker is designed, simulated and evaluated via simulations. The evaluation is performed with respect to the effectiveness of DC removal and computational complexity.

A novel DC canceller structure is presented in the second approach. Both the input and the output are assumed to be in single-bit format. The proposed structure contains no multi-bit multipliers and would be very simple to implemented in Field Programmable Gate Arrays.

Both types of blocker are useful in practice to improve the stability and dynamic range of single-bit systems. Their performance is tested for different kinds of input signals including sinusoidal, FM, and AM-FM signals.

5.2 ΣΔ-Ternary DC Blocker: System Design

5.2.1 The Ternary Filtering Stage

The DC blocker proposed in this Section is essentially a "ternary filter". This filter is an FIR structure as shown in Fig.(4.2), with coefficients confined to the ternary set: {-1, 0, +1}.The simplicity of the coefficients means that the multiplications within the filter can be implemented with great efficiency in hardware. As the input to the ternary filter is in single-bit format and the coefficients are in ternary format, each multiplication operation (or scaling) can be implemented with either a couple of logic gates or a simple look-up table [104].

There are a number of algorithms which can be used to generate suitable ternary tap coefficients. Typically one starts with a multi-bit "target" impulse response designed in a standard fashion (say with the Remez Exchange Algorithm). Then a design procedure is applied so as to obtain a ternary filter impulse response whose transfer function closely matches that of the target filter in the spectral band of interest. One can use design procedures based on dynamic programming, mini-max optimization or ΣΔM of the target impulse response [9], [12]. In this work, the tap values are designed using ΣΔ modulation of a target impulse response. The digital ΣΔM used for this purpose must satisfy two conditions. Firstly, it must have a tri-level output {-1, 0, +1}. Secondly, the ΣΔ modulator must have a flat signal frequency response over the bandwidth of interest [4]. This implies that the ΣΔ modulator should not modify the specifications of the target impulse response in the band of interest.

There are three important points to note about the ternary filter shown in Fig.(4.2). Firstly, the filter requires operation at an oversampled rate (OSR). This requirement is not unduly restrictive as the input signal is assumed to be in single-bit format and to have been generated by a ΣΔ modulator. Such single-bit signals almost always have a substantial OSR. Secondly, the output is in multi-bit format. To restore the output to the same single bit format

as the input, a re-modulator must be applied. The advantage of having a single bit format for the output is that further processing (including digital to analog conversion) can be done efficiently. Thirdly, the ternary filter output has significant noise levels because of the harsh quantization of both the input signal and the target impulse response. If one uses an appropriate re-modulator this output noise can be shaped away from the spectral band of interest. An example of a suitable re-modulator is given in [56].

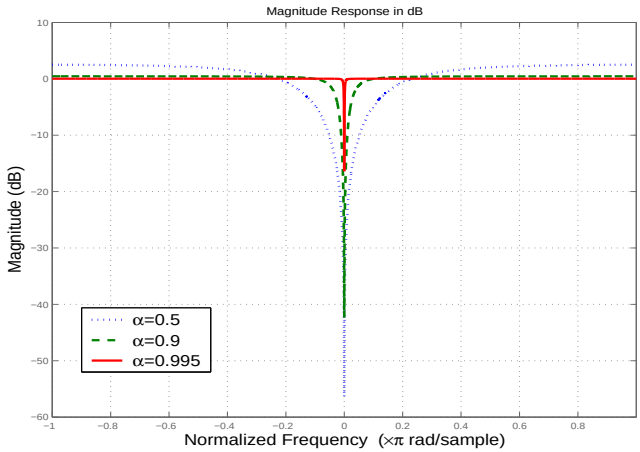

Figure 5.1: Frequency response of ideal multi-bit DC blocker for different values of the gain parameter α=0.5, 0.9, and 0.995.

Ternary filter design: The transfer function of a standard infinite precision IIR DC-blocking filter is [105]:

$$H(z) = \frac{1 - z^{-1}}{(1 - \alpha z^{-1})} \tag{5.1}$$

DC blocking occurs with the above transfer function by virtue of the zero in the above transfer function at $z = 1$ (0 Hz). The pole at $z = \alpha$ controls the system bandwidth, and therefore the system transit response. Fig.(5.1) depicts the frequency response based on eq.(5.1) for various values of the gain parameter α. As $\alpha \rightarrow 1$, the notch at DC gets narrower. This may seem ideal, however, there is a drawback. When $\alpha \rightarrow 1$, the impulse response duration will increase. Fortunately, as the end of the impulse response gets longer, its

initial amplitude decreases. When $\alpha = 1$, however, the pole and zero cancel each other at all frequencies, and hence, the impulse response shrinks to an impulse and the notch disappears. In practice, therefore, α cannot be made equal to 1.

A DC blocker can also be realized with an FIR filter. Although these filters are usually less efficient than IIR filters, they tend to have better phase response and are less vulnerable to instability [19]. Various different FIR design techniques can be employed to generate full-precision impulse responses for a suitable target filter. The resulting filter will have a series of dominant zeros in the vicinity of the 0 Hz axis in the z-plane. The full-precision target filter impulse response can then be interpolated by a factor of OSR (using either zero interlacing/ low-pass filtering or FFT techniques [96])and finally passed through a $\Sigma\Delta$ modulator. The output from this $\Sigma\Delta$ modulation process is the sequence of ternary filter coefficients (taps), $h(i)$, i.e., the ternary filter coefficients (taps), which are the ternary-quantized and encoded version of the interpolated impulse response. The ternary filter output $u(k)$ is given by the convolution of the ternary taps $h(i)$ (or simply h_i) with the input signal $x(k)$:

$$u(k) = \sum_{i=0}^{m*OSR} h_i x_{k-i} \tag{5.2}$$

where $m + 1$ is the order of the full-precision filter (i.e., at Nyquist rate).

It is important to realize that the design in the single-bit oversampling domain is not as straightforward as in the multi-bit domain. For instance, in our case where a DC blocker is intended to be designed, the problem of accurate discrimination between the DC content and the time-varying signal component will arise. This is due to the large oversampling ratio (32 or more) that practically pushes the time-varying signal spectrum to the vicinity of 0 Hz in the normalized frequency domain. Ultimate care should be given (by choosing the appropriate filter-order and design approach) to alleviate strong attenuation and to avoid the possible differentiation of the varying-

time signal itself by the DC blocker. Consequently, two ternary filter designs are investigated to allow for performance comparison.

From an efficiency perspective, it is important to design the filter impulse response to be as short as possible. A number of possible design techniques are feasible, and two will be compared in this Section. The first method used will be the well-known Remez Exchange technique. It will be used to generate a full-precision, linear phase target FIR filter of order $m = 21$, with coefficient symmetry $b(k) = b(m+1-k)$. This symmetry will give rise to good hardware efficiency. The second method used will be the so-called interpolated FIR (IFIR) approach [86]. The basis of this method is the fact that according to multi-rate processing theory, oversampling a signal compresses its spectrum. Hence, instead of directly designing a filter that satisfies the transition band specifications (i.e., f_{pass} and f_{stop}, as is the case with the Remez technique), OSR-times-stretched specifications (i.e., $OSR * f_{\text{pass}}$ and $OSR * f_{\text{stop}}$) used for designing a filter for an oversampled signal[86]. Using this approach, the filter order required to meet the new "stretched band-edges" is much lower than that designed to meet the original transition bands. A comparison between the two design techniques terms of filter performance and hardware efficiency will be conducted in Section 5.2.4.

5.2.2 The $\Sigma\Delta$ Modulator Stage

Spectral analysis of single-bit $\Sigma\Delta$ modulators with DC input has been addressed extensively, e.g., [106],[87]. It has been shown that when the input is a steady-state sinusoid (including a 0 Hz sinusoid), the quantization noise is not white. Rather it is highly colored. By adopting the linear-model approximation to represent the $\Sigma\Delta$ modulator, the power spectrum output corresponding to a DC input η is [56]:

$$S_{yy}(f) = (\frac{1}{3})[2\sin(\pi f)]^{2r} + \eta^2\delta(f) \tag{5.3}$$

where f represents the normalized frequency and r is the order of the modulator. The second term on the righthand side of eq. (5.3) represents the input signal (assumed to be a DC signal here), whereas the first term represents the quantization noise introduced by the modulator. The above result indicates that the $\Sigma\Delta$ modulator transfers the original input DC value to its output (by regulating the rate at which the output pulses occur, attempting to keep the average output equal to the average input) along with some highly colored quantization noise. This is true as long as the input signal is within the modulator dynamic range (and hence the system stability is maintained).

5.2.3 The DC Blocker

Our objective is to design a structure to eliminate the DC content in a time-varying input signal. Fig.(5.2) depicts the proposed structure, firstly utilized in [87], to carry out this task. The structure consists of a ternary filter stage cascaded with an IIR-$\Sigma\Delta$ re-modulator stage. The input to this structure, which is assumed to be a DC-biased sinusoidal signal, is assumed to be in single-bit format. The re-modulator stage re-encodes the multi-bit output of the ternary filter to single-bit format. The ternary filter stage consists of $(m+1) * OSR$ ternary taps.

Fig.(5.3) shows the linear model that represents the second-order $\Sigma\Delta$ modulator. The z-transform of the ternary filter output $Y_T(z)$), is given as:

$$Y_T(z) = X(z)z^{-1} + Q(z)(1 - z^{-1})^2 \tag{5.4}$$

where $X(z)$ and $Q(z)$ represent the z-transform of the signal and quantization noise, respectively.

From (5.4), the signal and noise frequency spectra can be obtained:

$$Y_T(e^{j\Omega}) = X(e^{j\Omega})e^{-j\Omega} + Q(e^{j\Omega})(1 - e^{-j\Omega})^2 \tag{5.5}$$

where $\Omega = 2\pi f/f_s$ is the normalized radian frequency.

The z-domain transfer function of the IIR stage that follows the ternary filter [see Fig.(5.3)] is given as follows:

$$H_{IIR}(z) = \frac{bz^{-1}}{1-(1-a)z^{-1}} + \frac{(1-z^{-1})^3}{1-(1-a)z^{-1}} \tag{5.6}$$

The output response of the overall system $Y_{\rm ov}$ will be the combined responses of the ternary filter $Y_T(e^{j\Omega})$ and that of the IIR-$\Sigma\Delta$ modulator filter $H_{\rm IIR}(e^{j\Omega})$:

$$Y_{\rm ov}(e^{j\Omega}) = Y_T(e^{j\Omega}) \cdot H_{\rm IIR}(e^{j\Omega}). \tag{5.7}$$

From equations 5.6 and 5.7 we get:

$$Y_{\rm ov}(e^{j\Omega}) = Y_T(e^{j\Omega}).[H_{\rm IIRS}(e^{j\Omega}) + H_{\rm IIRN}(e^{j\Omega})] \tag{5.8}$$

where $H_{\rm IIRS}(e^{j\Omega}) = bz^{-1}/[1-(1-a)z^{-1}]$ and $H_{\rm IIRN}(e^{j\Omega}) = (1-z^{-1})^3/[1-(1-a)z^{-1}]$ are the signal and noise parts of $H_{\rm IIR}(e^{j\Omega})$, respectively. The noise shaping filter, $H_{\rm IIRN}$, attenuates the quantization noise in the signal band and amplifies it outside the signal band. These high-frequency noise components can be eliminated by subsequent digital filtering. Once the quantization noise is filtered the signal can also be decimated.

Now, the overall output response, $Y_{\rm ov}(e^{j\Omega})$, can be expressed as follows:

$$Y_{\rm ov}(e^{j\Omega}) = \frac{X(e^{j\Omega})\ K(e^{j\Omega})}{D(e^{j\Omega})} \ + \ \frac{Q(e^{j\Omega})\ P(e^{j\Omega})}{D(e^{j\Omega})} \tag{5.9}$$

where

$$\begin{aligned} K(e^{j\Omega}) &= e^{-j\Omega} + e^{-2j\Omega}(b-2) + e^{-3j\Omega} && (5.10)\\ D(e^{j\Omega}) &= 1 - ae^{-j\Omega} && (5.11)\\ P(e^{j\Omega}) &= e^{-j\Omega}(b-4) + e^{-2j\Omega}(6-2b) \\ &\quad + e^{-3j\Omega}(b-4) + e^{-4j\Omega}. && (5.12) \end{aligned}$$

Note that a and b are multiplicative constants. The parameter a can be used

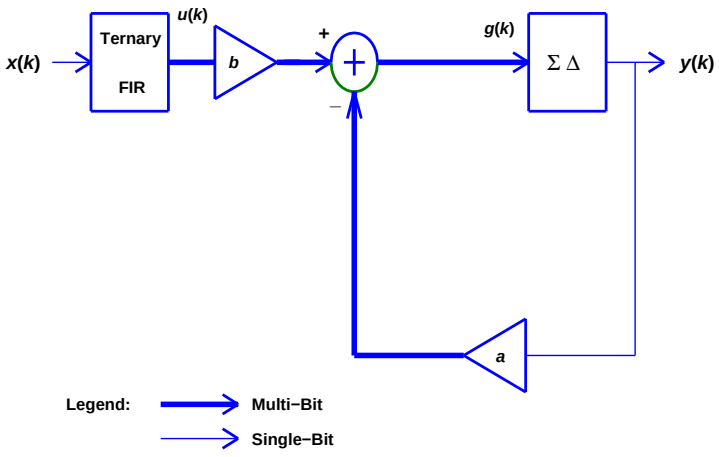

Figure 5.2: Structure of the proposed single-bit ternary DC blocker.

to control the location of the overall transfer function poles along the real z-axis and should be carefully set to insure correct operation and stability. The choice of a should ensure that the pole is located nearer to the origin than the zero. Also, the pole should not be faraway from the zero to ensure that there is a sharp cut-off and that suitable gain is obtained.

The parameter b controls the amplitude of the input signal and can be adjusted to get maximum SNR. Fig.(5.4) depicts the theoretical frequency response of the signal-transfer function ($STF = K(e^{j\Omega})/D(e^{j\Omega})$) and the noise-transfer function ($NTF = P(e^{j\Omega})/D(e^{j\Omega})$) of the overall structure calculated from eq.(5.9).

5.2.4 Simulation and Discussion

MATLAB is utilized to simulate the proposed DC blocker. As mentioned in Section 5.2.1, the ternary filter stage has been designed using two approaches. Figures (5.5) and (5.6) show the simulated frequency response of the ternary stage, in comparison with the calculated full-precision target response, using the Remez and IFIR techniques, respectively.

Note that the signal band of interest is Ω=0 to 0.016π for the OSR of 32 used in this simulation.

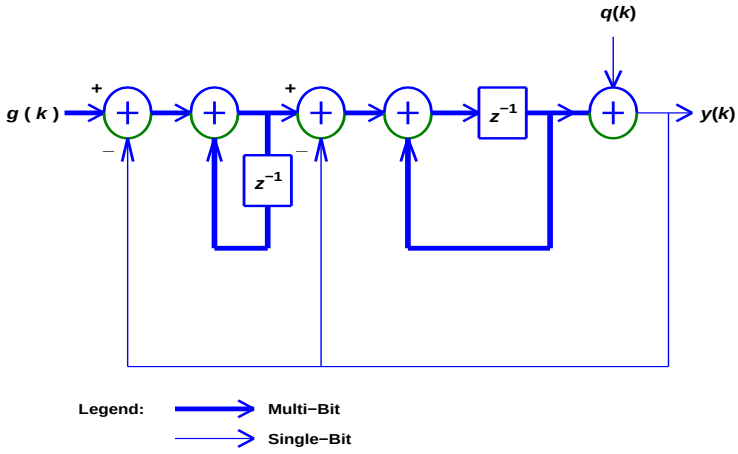

Figure 5.3: A block diagram of the linear model for a second-order ΣΔ modulator.

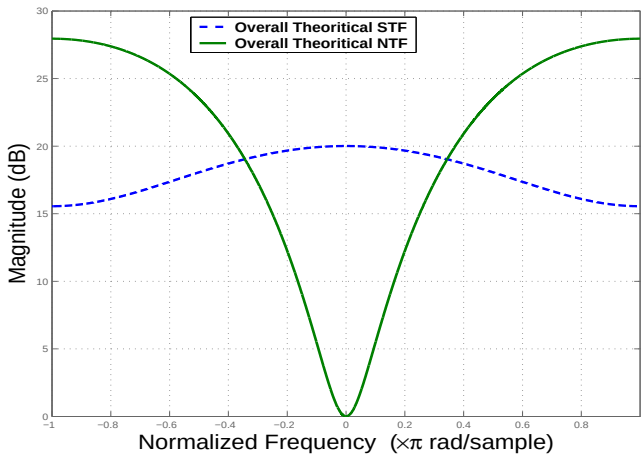

Figure 5.4: Frequency response of the theoretical STF (dotted) and NTF (solid) for b=10 and a=0.001.

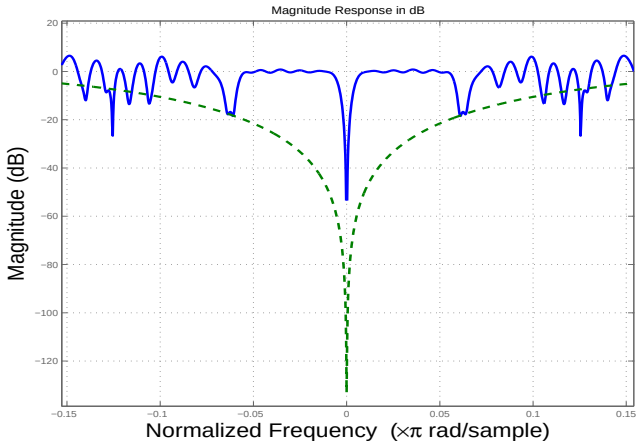

Figure 5.5: Frequency response of the ternary filter stage using Remez technique, compared with the target response (dashed line).

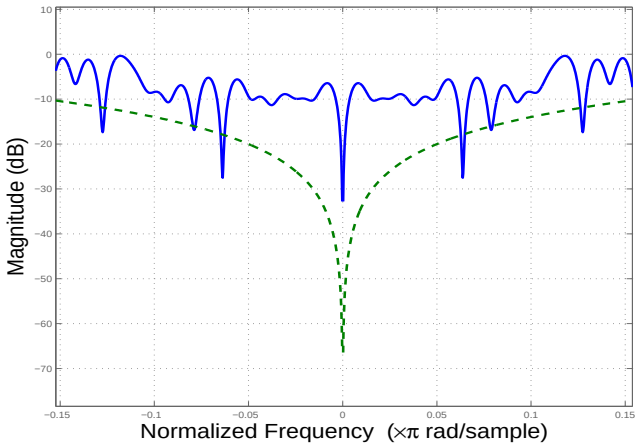

Figure 5.6: Frequency response of the ternary filter stage using IFIR technique.

Table 5.1: A Comparison between Remez and IFIR techniques (OSR=32).

Filtering Technique	Filter Order	Pass-band Ripple (dB)	DC Atten. (dB)	No. of Ternary Taps	Non-Zero Ternary Taps	Percent. of zero Taps	Phase Response
Remez	21	2	50	672	212	68%	Linear
IFIR	6	5	35	193	84	56%	Piece-wise Linear

Table 5.1 compares the performance of these ternary filters in terms of signal and DC-content attenuation as well as the number of non-zero filter taps required (OSR=32). It is interesting to find out that the required transition band of the DC-blocker can be realized using only 84 ternary taps when the IFIR design method is utilized. The price paid for this hardware simplicity is the increase in pass-band ripple. It is worth noting that the zero-valued taps (no hardware connection) constitutes the majority of the total number of taps (56%-68%) for both filter design methods.

The resulting ternary filter has an anti-symmetric impulse response. This anti-symmetry gives a linear phase response; also it can be utilized to halve the number of the coefficient multipliers [96]. Although ternary multipliers are simple in structure, the reduced hardware requirement is very pleasing.

It is important to note that as OSR increases, the signal spectrum approaches the frontier of the DC. In this case, the need for a larger order FIR filter becomes a vital demand. This can be deduced from Fig.(5.7), which shows the attenuation versus the OSR.

Fig.(5.8) shows the simulated overall filter frequency response as compared to the target response, using the Remez ternary filter. It is clear that the proposed structure presents good DC-blocking characteristics. Fig.(5.9) depicts the overall phase response along with that of the target phase response, where the overall phase response is deformed. This is expected due to the nonlinear

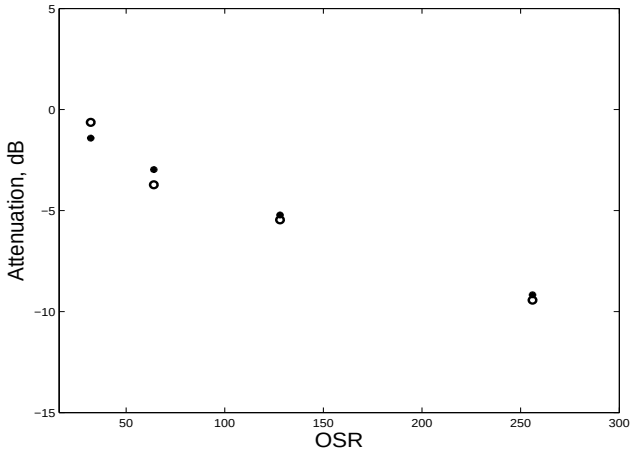

Figure 5.7: Signal attenuation in ternary filter stage against the OSR for $b = 1.6$ and $a = .01$. (*) Remez. (o) IFIR.

behavior of the $\Sigma\Delta$ modulator stage [107]. However, in the baseband, the phase response seems almost linear (apart from modulo 2π jumps).

In Fig.(5.10), the input and the output spectra of the DC-blocker are shown. It is evident that the DC component in the input signal is removed. The input is taken as: $A_{\rm DC} + A\sin(\omega_o t) + \nu(t)$, where $A_{\rm DC} = 0.5$, $A = 0.5$, $\omega_o = 8192\pi$ rad/sec (chosen to be in the audio band), and $\nu(t)$ is an additive white Gaussian noise (AWGN). Hence, the input signal contains a DC power twice in magnitude as the sinusoidal power. To meet the minimum requirement for audio applications, the signal-to-noise ratio (SNR) is made 20 dB. Several different input types has also been used in testing the DC-blocker, such as sawtooth, FM, and AM-FM as can be shown in Fig.(5.11) and Fig.(5.12), respectively. In all cases, the responses are comparable to those shown for the sinusoidal input.

The time-domain reconstructed DC-biased signal is shown in Fig.(5.13) for sawtooth input. The reason behind utilizing the sawtooth input signal is to check for correctness of the system response by ensuring that differentiating of the input signal has not been taken place.

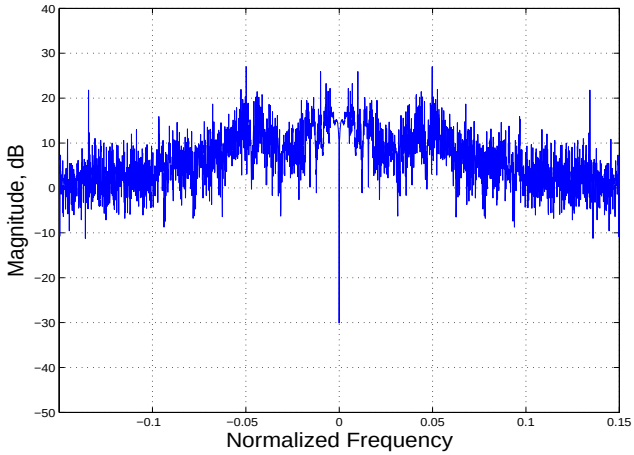

Figure 5.8: The overall frequency response of the ternary-ΣΔM single-bit DC blocker.

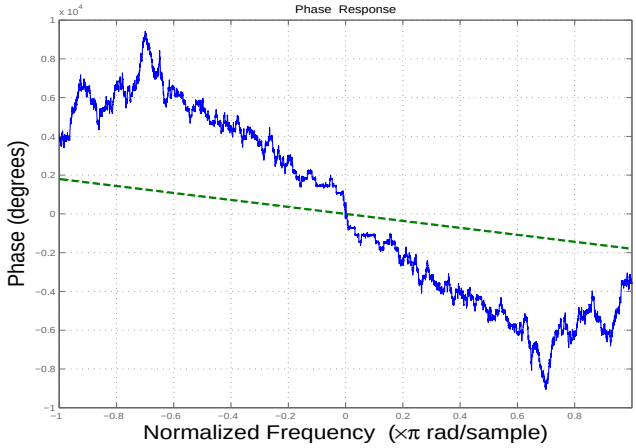

Figure 5.9: The overall phase response of the ternary-ΣΔM single-bit DC blocker (solid), compared with the phase response of the target impulse response (dashed).

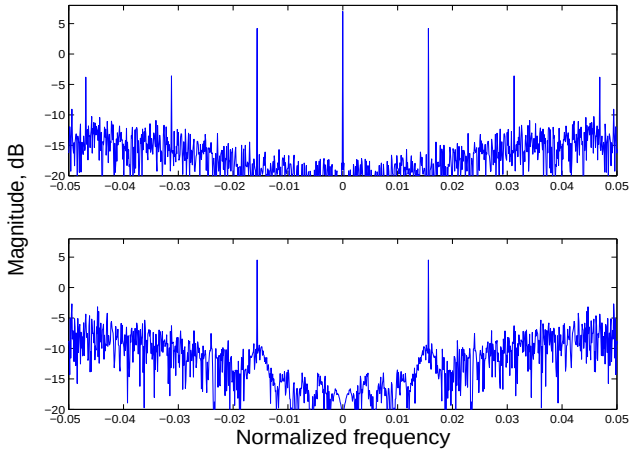

Figure 5.10: Spectra of the single-bit input and the single-bit output of the proposed DC blocker. Above: noisy sinusoid input spectrum. Below: output spectrum.

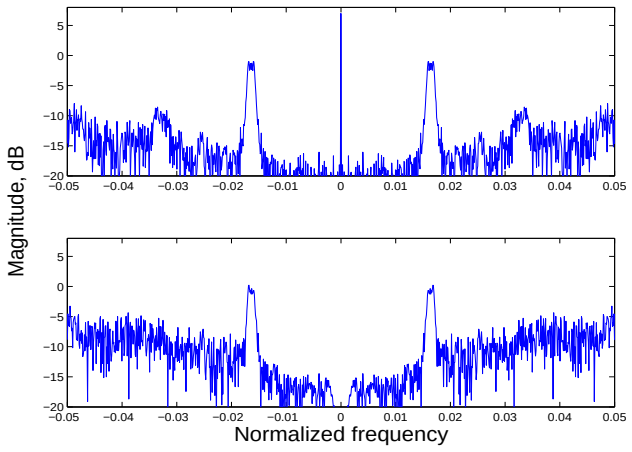

Figure 5.11: Spectra of the single-bit input FM and the single-bit output of the proposed DC blocker. Above: input spectrum. Below: output spectrum.

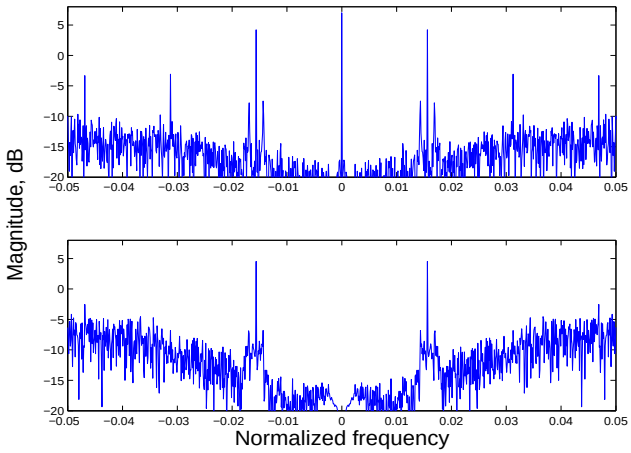

Figure 5.12: Spectra of the single-bit input AM-FM and the single-bit output of the proposed DC blocker. Above: input spectrum. Below: output spectrum.

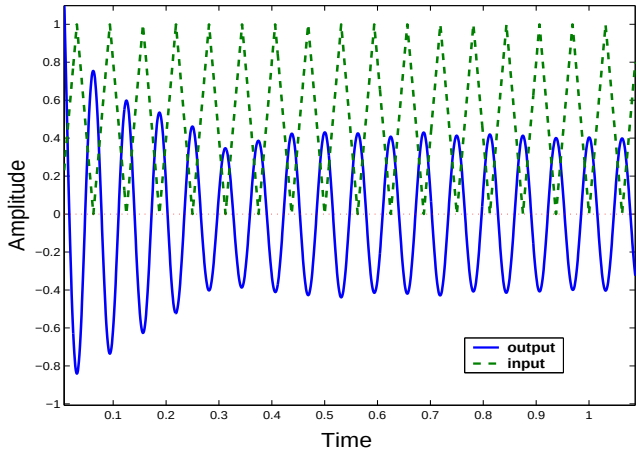

Figure 5.13: Reconstructed sawtooth DC-biased input signal from the single-bit DC blocker.

5.3 A Proposed Structure for DC Canceling in Single-Bit Domain

5.3.1 Design and Analysis

A simple multi-bit DC-Blocker can be seen in [80]. The transfer function of a traditional infinite precision IIR DC-blocking filter is given in eq. (5.1).

The DC cancelation is due to the transfer function having zero at $z = 1$ (0 Hz). The pole at $z = 1 - \alpha$ adjusts the system bandwidth. Our objective is to design a structure that eliminates the DC content which is encoded in single-bit format along with a time-varying input signal. The design of single-bit systems is a non-trivial task. To characterize single-bit systems, it is common to look at both the signal transfer function (STF) and the noise-transfer function (NTF). The STF describes how the modulator alters the original input signal spectrum, and for the DC blocking application must be a a high-pass function. The NTF indicates how effectively the modulator shapes noise away from the signal band of interest. The NTF is the main design task which determines the amount of baseband noise shaping performed by the modulator. In general, the NTF is designed to be one of two types; either a pure M^{th} order differentiator, [NTF$(z) = (1 - z^{-1})^M$], or "non-monotonic" transfer function which has poles in addition to zeros, NTF [NTF$(z) = (z - 1)^M / D(z)$]. For either type, as the order M increases, more noise power typically moves to unwanted frequency bands and noise in the wanted frequency bands is reduced. Consequently, signal-to-quantization-noise ratio (SQNR) in the band of interest is increased.

Fig.(5.14) shows the proposed single-bit DC-blocker system. This structure is comprised basically of a delta-modulator with sigma-delta modulation embedded in its feedback loop. We denote $x(n)$ as the single-bit input, $y(n)$ as the bitstream output, $u(n)$ as the input of the signal path quantizer [$\mathrm{P}_1(.)$], $s(n)$ as the feedback signal, and $v(n)$ as the input of the feedback path quantizer

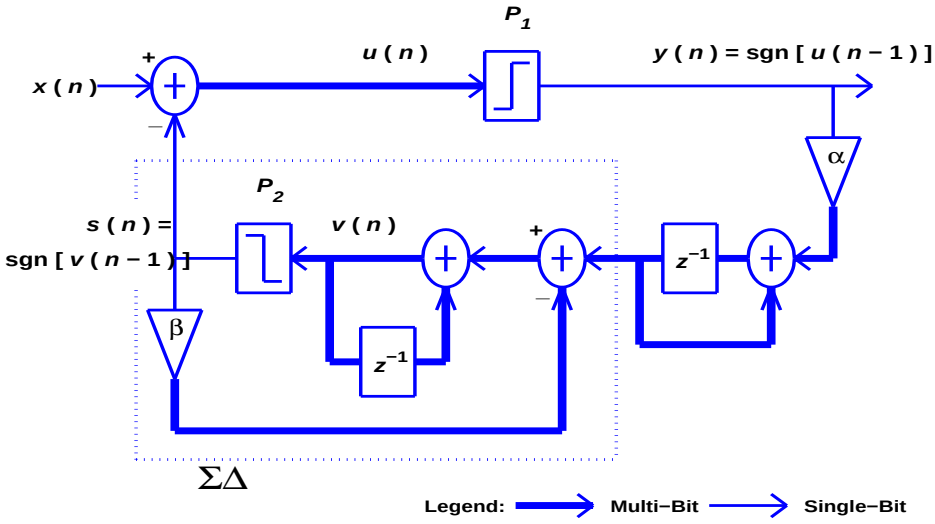

Figure 5.14: The proposed DC-blocker.

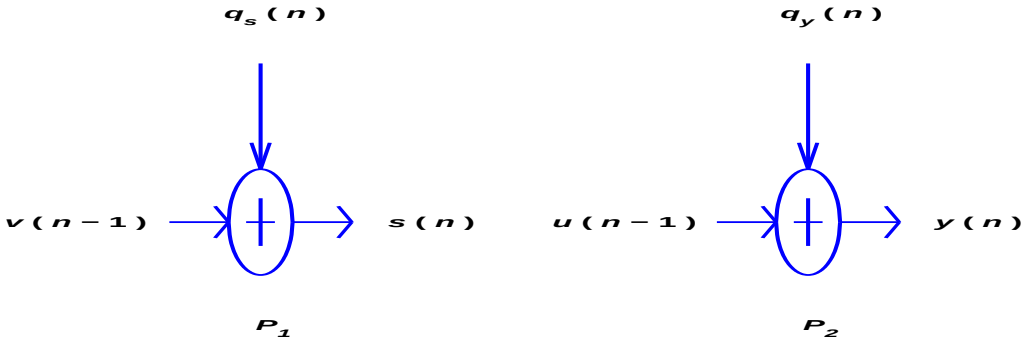

Figure 5.15: Linear model approximation of the system quantizers.

$[P_2(.)]$. In this case $y(n)$ and $s(n)$ are given as follows:

$$y(n) = \begin{cases} +1 & \text{for } u(n) \geq 0 \\ -1 & \text{for } u(n) < 0 \end{cases} \tag{5.13}$$

$$s(n) = \begin{cases} +1 & \text{for } v(n) \geq 0 \\ -1 & \text{for } v(n) < 0 \end{cases} \tag{5.14}$$

We adopt the well-known linear system approximation approach [16], in which the 1-bit quantization process is represented by a unity gain summing element, and the quantization noise is modeled as an additive, white, and

signal-independent noise source with variance $\sigma^2 = \delta^2/12$ (δ represents the quantization interval). Fig.(5.15) shows the linear models of the proposed system quantizers, where $q_y(n)$ and $q_s(n)$ represent the quantization noise of the quantizers $\mathrm{P}_1(.)$ and $\mathrm{P}_2(.)$, respectively. It is worth mentioning that the linear model approach is unable to explain many aspects of the SDM behavior such as integrator spans, stability, limit cycles, and chaos [107]. Nonetheless, the linear approach provides a good approximation to the noise performance of 1-bit systems.

Now the system function $H(z)$ can be represented as a linear combination of the signal transfer function [STF(z)] and the noise transfer function [NTF(z)], i.e., $H(z) = \mathrm{STF}(z) + \mathrm{NTF}(z)$. From Fig.(5.14), the z-transform of the output $Y(z)$ can be described as follows:

$$Y(z) = X(z)\frac{B(z)}{D(z)} + Q_s(z)\frac{z^{-1}(1-z^{-1})^2}{D(z)} + Q_y(z)\frac{B(z)}{D(z)} \tag{5.15}$$

where,

$$B(z) = (1-z^{-1})[1-(1-\beta)z^{-1}] \tag{5.16}$$

$$D(z) = [1-(2-\beta)z^{-1}+(1-\beta+\alpha)z^{-2}] \tag{5.17}$$

α and β being gain parameters. From (5.15) we have $\mathrm{STF}(z) = B(z)/D(z)$, whereas two separate noise-shaping functions are in effect: $\mathrm{NTF}_s(z)=z^{-1}(1-z^{-1})^2/D(z)$ and $\mathrm{NTF}_y(z)=B(z)/D(z)$, which high-pass filter the quantization noise Q_s and Q_y, respectively.

In order to remove the DC content from the bitstream input, the STF of the system should operate as a high-pass filter. Based on equation (5.15), Fig.(5.16) depicts the theoretical frequency response of $\mathrm{STF}(e^{j\Omega})$ and $\mathrm{NTF}_y(e^{j\Omega})$.

It is obvious from (5.15) that the system function, $Y(z)$, contains two zeros

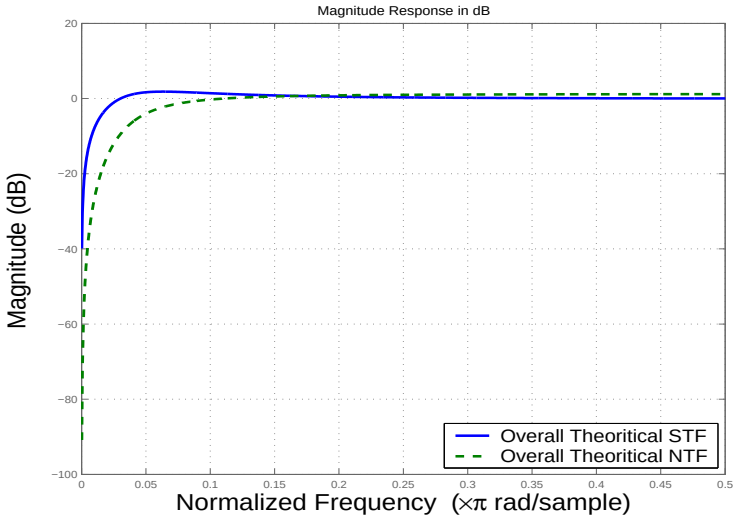

Figure 5.16: Signal and noise transfer functions, $\mathrm{STF}(e^{j\Omega})$ and $\mathrm{NTF}_y(e^{j\Omega})$, of the DC blocker using first-order SDM with $\alpha = 0.0205$ and $\beta = 0.2705$.

$z_{1,2}$ and two poles $p_{1,2}$ as follows:

$$z_1 = 1, \quad z_2 = 1 - \beta \tag{5.18}$$

$$p_{1,2} = (1 - 0.5\beta) \mp \sqrt{(1 - 0.5\beta)^2 - (1 - \beta + \alpha)}. \tag{5.19}$$

The above poles will take real values for $\beta \geq 2\sqrt{\alpha}$, and form a conjugate pair when $\beta < 2\sqrt{\alpha}$. The gain parameters α and β play an important role in characterizing the performance of the DC-blocker through the control of pole-zero locations. Accordingly, their combination will specify the system bandwidth. However, it should be noted that for α=0, the three poles occupy the locations of the three zeros, and hence cancel each other. For that α is the critical parameter in this regard, as its value determines how much the poles and zeros are separated.

The performance of the proposed structure can be evaluated in terms of the overall signal-to-noise ratio (in the band of interest) plus the signal-to-quantization-noise ratio, $\mathrm{SNR}_{\mathrm{ov}}$. From (5.15), $\mathrm{SNR}_{\mathrm{ov}}$ can be calculated as [14] [106]:

$$\mathrm{SNR}_{\mathrm{ov}} = \int_{-\Omega_B}^{\Omega_B} \left|\frac{X(e^{j\Omega})\mathrm{STF}(e^{j\Omega})}{G(e^{j\Omega})}\right|^2 \, d\Omega \tag{5.20}$$

where $\Omega_B \in (0, \pi)$ denotes the normalized desired signal bandwidth ($\Omega = \pi$ corresponds to half the sampling rate, Ω_s), and $G(e^{j\Omega})$ is given by:

$$G(e^{j\Omega}) = Q_y(e^{j\Omega})\mathrm{NTF}_y(e^{j\Omega}) + Q_s(e^{j\Omega})\mathrm{NTF}_s(e^{j\Omega}). \tag{5.21}$$

Note that $X(e^{j\Omega})$ represents the input bit-stream spectrum, assumed to contain quantisation noise as a result of a previous SDM encoding process in addition to white Gaussian noise.

5.3.2 Simulation and Discussion

MATLAB is utilized to simulate the proposed structure. We denote by $\mathrm{SNR}_{\mathrm{ovi}}$ the overall input SNR. To meet the standard audio specifications, we suggest $\mathrm{SNR}_{\mathrm{ovi}} = 20$ dB. To assess the performance of the DC-blocker, we define the parameter $\rho = 10\log_{10}(\mathrm{SNR}_{\mathrm{ovo}}/\mathrm{SNR}_{\mathrm{ovi}})$, where $\mathrm{SNR}_{\mathrm{ovo}}$ stands for the output $\mathrm{SNR}_{\mathrm{ov}}$. The optimal values for the gain parameters α and β are specified in the sense of maximum attainable $\mathrm{SNR}_{\mathrm{ovo}}$, i.e., maximum ρ (or ρ_m). Fig.(5.17) illustrates ρ as a function of the gain parameters α and β such that both span the interval (0,0.1]in a step-size of 2^{-10}. Simulation shows that the optimum operating point $\rho_m(\alpha_m, \beta_m)$ for the DC blocker in Fig.(5.14) occurs when $\alpha_m = 0.0205$ and $\beta_m = 0.2705$ such that maximum ρ equals about -1.51 dB. The resolution of the multi-bit region in the DC blocker is assumed to be 10-bit in this simulation. The resolution can be changed according to the application requirements.

The degradation in $\mathrm{SNR}_{\mathrm{ovo}}$ can be removed by replacing the first-order SDM in the feedback path of the DC-blocker with a higher-order one. Fig.(5.18) depicts the improvement in ρ_m when a second-order SDM stage is embedded in the proposed structure. In this case ρ_m=3.6 dB for α_m=0.0127 and β_m=0.0508, where significant improvement in the performance is achieved over the first-order SDM-based DC blocker.

A comparison between the simulated frequency response curves of the DC-

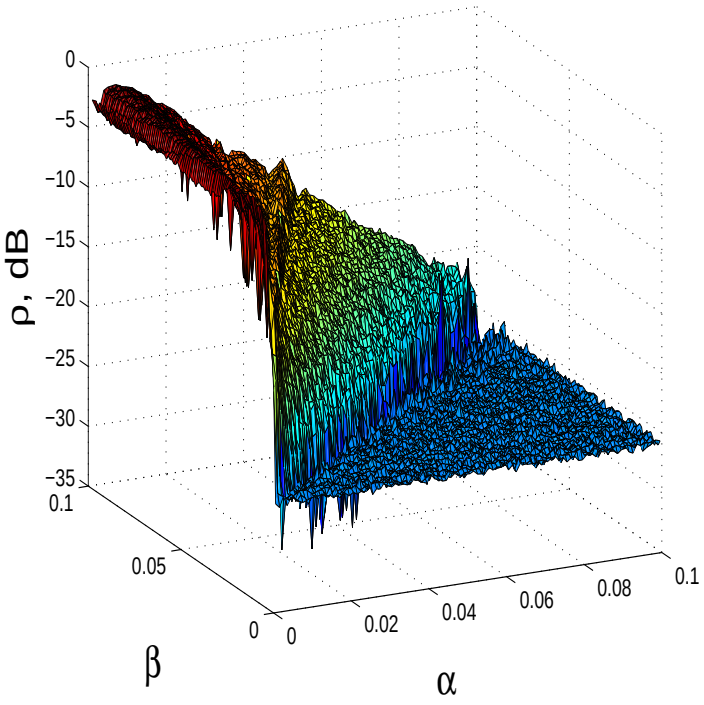

Figure 5.17: The ratio $\rho = \mathrm{SNR_{ovo}}/\mathrm{SNR_{ovi}}$ (in dB) versus the gain parameters α and β using 10-bit resolution.

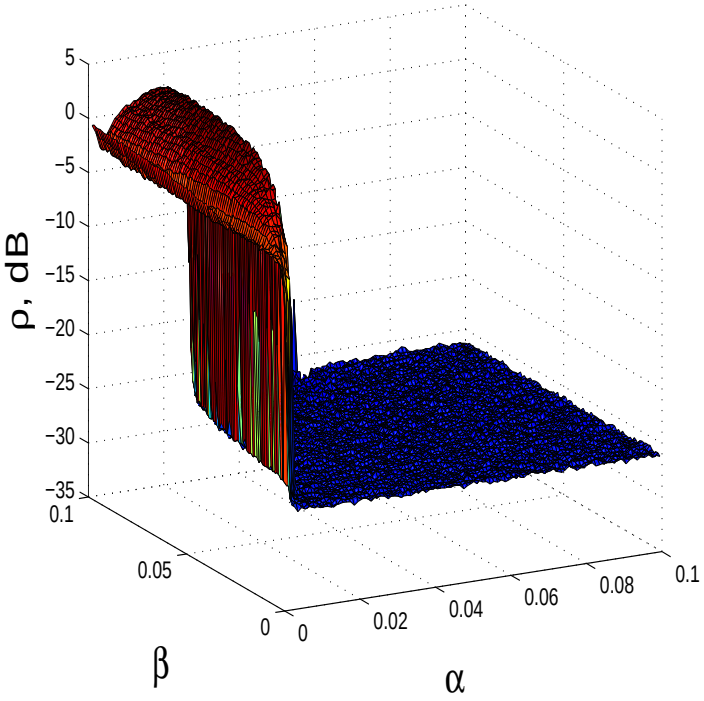

Figure 5.18: The ratio ρ versus the gain parameters α and β (10-bit resolution) of the DC blocker using a second-order SDM.

blocker for first- and second-order SDM stages for optimum ρ is depicted in Fig.(5.19). The dotted vertical lines indicate the desired signal band for OSR = 32.

In Fig.(5.20), the input and the output spectra of the DC-blocker are shown. It is evident that the DC component in the input signal is removed. Moreover, an improvement in the SNR of more than 2 dB is obtained. The input is taken as: $A_{\rm DC} + A\sin(\omega_o t) + \nu(t)$, where $A_{\rm DC} = 0.5$, $A = 0.5$, $\omega_o = 8192\pi$ rad/sec (chosen to be in the audio band), and $\nu(t)$ is an additive white Gaussian noise (AWGN) process. Hence, the input signal contains a DC component that is twice in magnitude as the sinusoidal component. To meet the minimum requirement for audio applications, the overall signal-to-noise ratio ($\rm SNR_{ovi}$) is made as 20 dB. Different input types have also been used in this test, including sawtooth, FM, and AM-FM signals. In all cases, the response curves are comparable to those shown for the sinusoidal input, as can be seen in

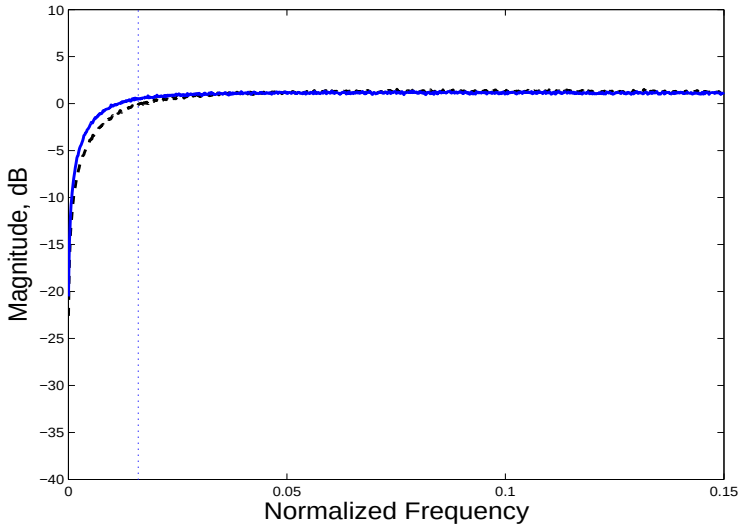

Figure 5.19: Frequency response of the simulated DC-blocker: (solid) using second-order SDM ($\alpha_m = 0.0127$, $\beta_m = 0.0508$); (dashed) using fist-order SDM ($\alpha_m = 0.0205$, $\beta_m = 0.2705$).

Fig.(5.21) and Fig.(5.22) for FM and AM-FM input, respectively. To check the performance of the DC-blocker against introducing undesirable differentiation, Fig.(5.23) shows the output spectrum in response to a noisy sawtooth input (with SNR=20dB).

From hardware implementation viewpoint, the proposed DC blocker is very simple, as it contains no multi-bit multipliers. The gain parameters α and β can be realized using conventional voltage dividers or simple digital scalers. For FPGA implementation, these two gains can be achieved by using two multiplexers, each of them multiplexes two fixed multi-bit numbers (that represent α and -α or β and -β), where the multiplexer output is dependent on the quantiser output as shown in Fig.(5.24).

5.3.3 Stability

The proposed structure is a linear system except for the single-bit quantizers which are non-linear elements. As mentioned earlier, using the linear approximation is inadequate to model this system accurately. However, the linear model does reveal some valuable analytical results when using a low order (≤ 2)

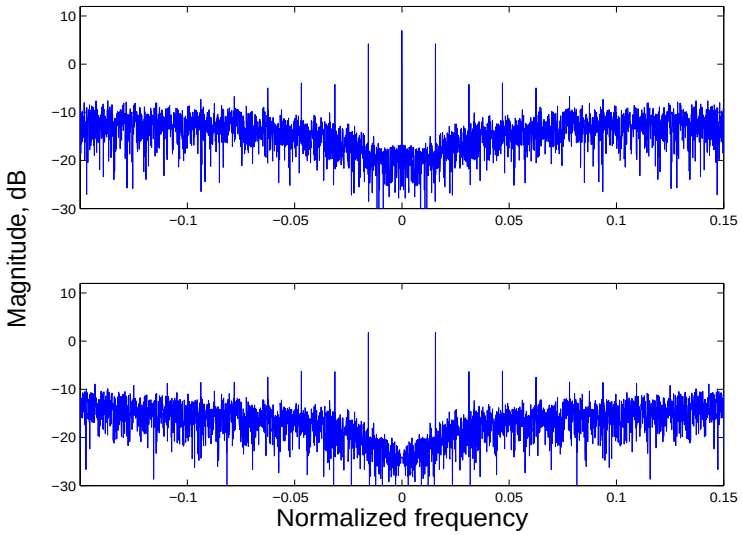

Figure 5.20: Input and output spectra of the DC blocker (using a second-order SDM) for an FM input.

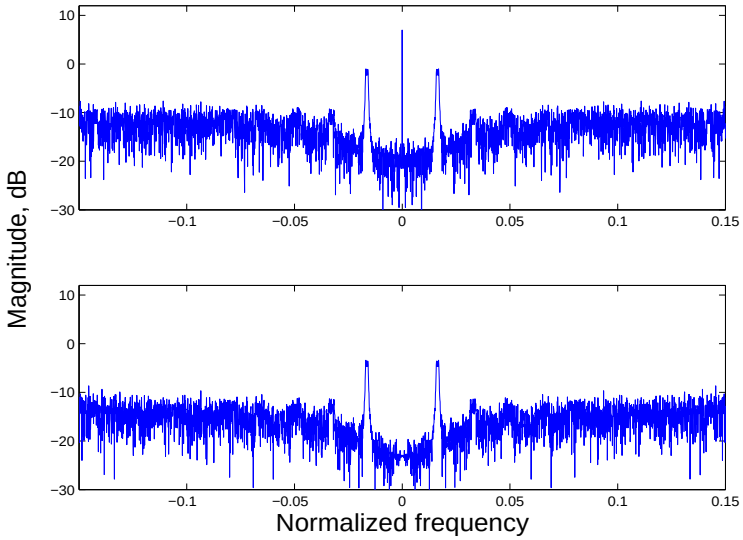

Figure 5.21: Input and output spectra of the DC blocker (using a second-order SDM) for an FM input.

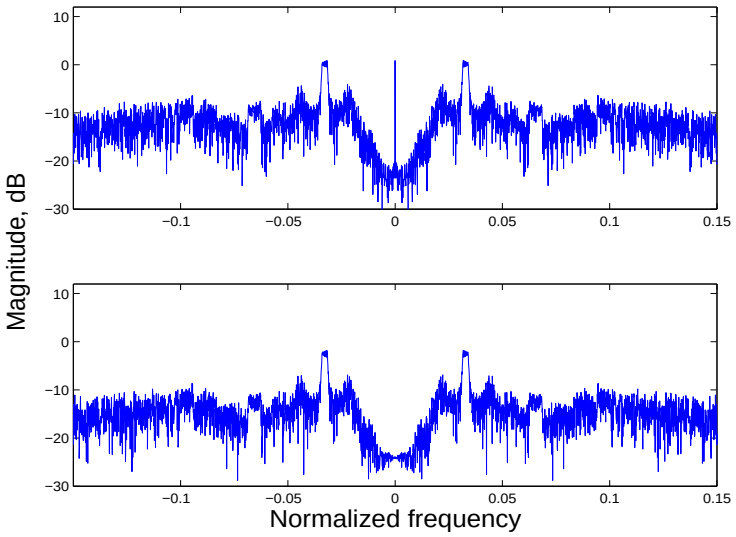

Figure 5.22: Input and output spectra of the DC blocker for a noisy AM-FM input.

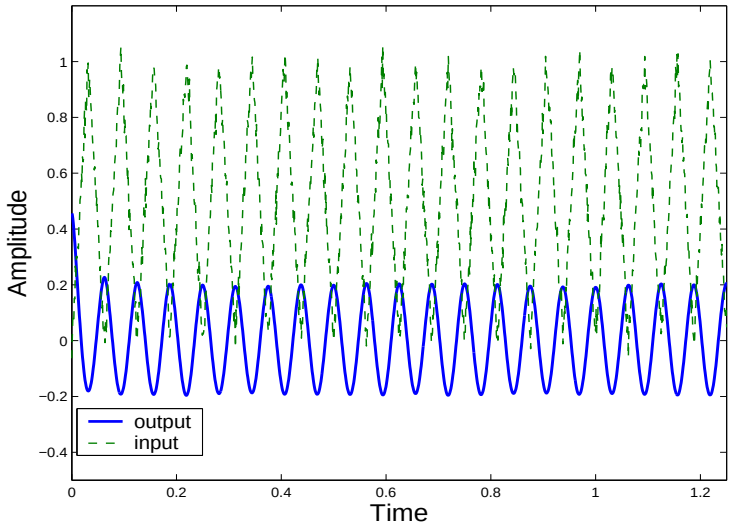

Figure 5.23: Input and output spectra of the DC blocker for a noisy sawtooth input.

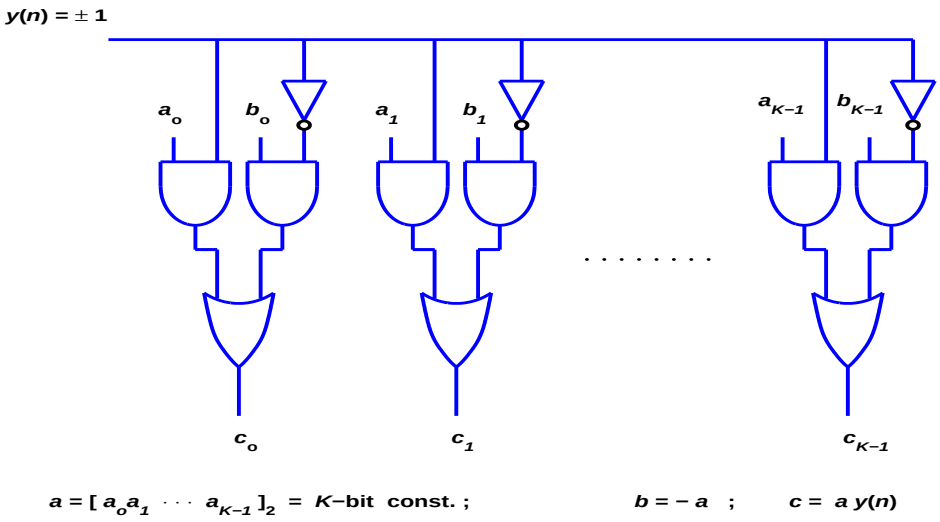

Figure 5.24: Multiplication of a single-bit signal by a multi-bit constant.

SDM stage in the feedback loop. The stability problem would be complicated when using a third (or higher)- order SDM as these high-order topologies are prone to the instability problem [98]. A detailed non-linear stability analysis is beyond the scope of this chapter, however, investigating the root-locus of the system would be useful to approximate its stability criteria.

Considering the system with first-order SDM as shown in Fig.(5.14) with zeros and poles as given by eq.(5.18) and eq.(5.19), the parameter β determines the locations of one zero (z_2) and the two poles $p_{1,2}$ (noting that for α=0, zeros and poles will cancel each other), while α controls the pole-zero separation in each pole-zero pair. Fig. (5.25) can be used to clarify the root-locus behavior of this system as follows when $\beta = 0.8$. Starting with $\alpha = 0$, each pole will occupy (cancel) a zero, i.e., $p_1 = z_1 = 1$ and $p_2 = z_2 = 1 - \beta$. The distance between these two initial poles is β. Let the mid point between the two initial poles be represented by p_c, then $p_c = 0.5(p_1 + p_2) = 1 - 0.5\beta$. As α increases, the two poles will travel horizontally in opposite directions on the real axis until they meet at p_c when $\alpha = (0.5\beta)^2$. The point p_c will remain as mid point between the two poles as they move. Further increase in α beyond $(0.5\beta)^2$ will drive the two poles to be a complex conjugate pair tracing a vertical line

centered at p_c (where the right pole moves upwards, while the other moves downwards). The intersection points between the unit circle and this vertical line will reveal the stability criteria of the system. Denoting the real axis as μ and the imaginary axis as ν, the intersection of the root locus with the unit circle occurs at

$$\nu = \sqrt{\beta - \frac{\beta^2}{4}}. \tag{5.22}$$

Moreover, the poles equation (5.19) will give

$$\left| (1 - \frac{\beta}{2}) \mp \sqrt{(1 - \frac{\beta}{2})^2 - (1 - \beta + \alpha)} \right| = 1 \tag{5.23}$$

Now using (5.22) and (5.23), the following conditions for the complex conjugate poles can be reached:

$$(\frac{\beta}{2})^2 < \alpha < \beta < 2 \tag{5.24}$$

while for real poles the following conditions are obtained:

$$\alpha < \min\{(\frac{\beta}{2})^2,\ \beta\}\ ; \qquad \beta < 2 \tag{5.25}$$

which can be reduced to

$$\alpha < (\frac{\beta}{2})^2 < 1. \tag{5.26}$$

The poles will exit the unit circle circumference if $\alpha > \beta$.

Fig.(5.26) shows the pole-zero plot of the system shown in Fig.(5.14) at the optimum operating point ρ_m, with α=0.0205 and β=0.2705. From an LTI system viewpoint, this plot confirms that the designed DC-blocker is always stable. This is so because all poles are located within the unit circle in the z-domain. However, since our system is nonlinear, this condition from linear analysis is considered sufficient for stability but not necessary [108]. Simulation results confirms this claim.

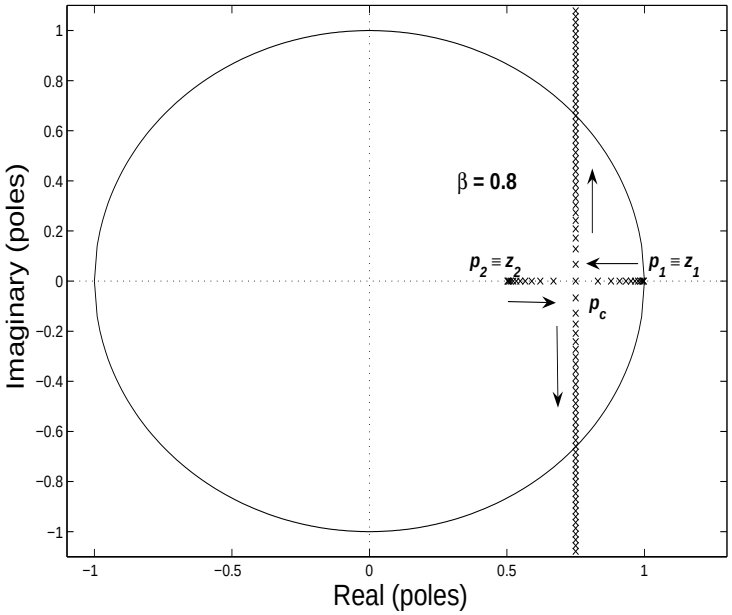

Figure 5.25: Root-locus of the proposed DC blocker with β=0.8.

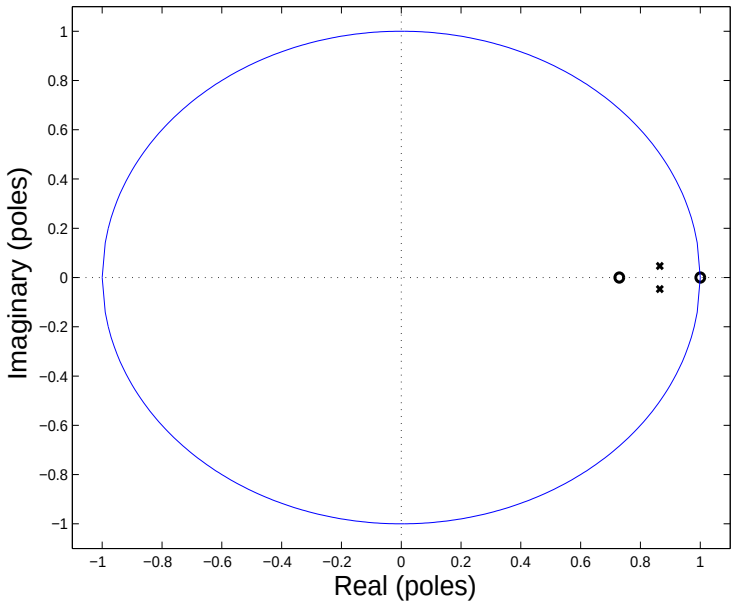

Figure 5.26: Pole-zero plot of the DC-blocker with first-order SDM at $\rho = \rho_m$ using α=0.0205 and β=0.2705.

5.4 Summary

In this Chapter, two efficient structures for DC-blocking in the single-bit domain has been proposed. The first consists of a ternary filtering stage followed by a sigma-delta modulator stage. Two design techniques were utilized to generate the ternary taps. For each technique, the associated ternary filter stage was assessed in terms of DC attenuation and hardware efficiency. The simulated system response has been studied through the application of various DC-biased, noisy signals. The DC content was removed completely from all kinds of input signals.

The second is a single-bit multiplierless DC-blocker. The structure is comprised of a delta-modulator structure with a sigma-delta modulating (SDM) stage in its feedback loop. The proposed system is evaluated in terms of the overall SNR and the magnitude of DC attenuation. It is shown that using a second-order SDM improves the overall system SNR as compared to using a first-order one. However, using higher-order SDM (> 2) would complicate the stability issue as higher-order SDM topologies inherently suffer from instability problems. The role of the gain parameters is investigated and optimal performance has been reached assuming 10-bit resolution. Stability criteria have been derived. The system is examined using different types of signals.

Both of these proposed structures are very efficient in hardware realization and can be easily implemented with FPGA.

Chapter 6

Limit Cycle Behavior in Ternary Structures

6.1 Introduction

One of the intriguing aspects in the behavior of ΣΔ modulators (ΣΔM's) is the generation of periodic patterns (or limit cycles) at its output when the input is a dc signal. The existence of those periodic modes (limit cycles) is due to the nonlinear nature of ΣΔMs [31]. Such a cyclic output produces discrete noise components. Studies have been made on the limit cycle nature of the first- and second-order ΣΔ modulators and on the appropriate techniques to eliminate them [32][42][109]. Similar behavior has been reported in digital IIR filters [110]. Despite the various attempts to identify the limit cycle mode of higher order (more than 2) ΣΔ systems [40] [33], a comprehensive analysis has not been achieved yet. Understanding the limit cycle phenomenon in higher order systems is becoming more crucial because of their improved performance at oversampling ratios similar to those ratios utilized by the first and second-order systems, which makes them attractive to the ΣΔ system designers. Moreover, due to their instability problem (which is the main drawback in these topologies), the limit cycle behavior should be thoroughly investigated as it has a strong impact on the issue of instability [111].

Recent works have shown that ternary filtering structures (which utilize a finite-impulse-response (FIR) filter, whose coefficients $\in \{\pm 1, 0\}$, followed by

a $\Sigma\Delta$ system) are promising in performing significant digital signal processing (DSP) tasks. In these structures, a third-order $\Sigma\Delta$ topology has been successfully utilized (see, for example, [87][93][104]). As limit cycles may occur in the $\Sigma\Delta$ part of the system, we expect that ternary filters will consequently experience limit cycles as well.

The focus of this chapter will be on a third-order $\Sigma\Delta$ system. However, the same approach can be easily extended to analyze higher-order systems (> 3).

The chapter is organized as follows. In Section-6.2, the difference equation of an 3^{rd} order $\Sigma\Delta$ modulator is developed, and a solution to the difference equation of a third-order system is obtained along with a general expression for the average output of this third order system is determined. In Section 6.3 the M^{th}-order difference equation is developed and a discussion of how to generalize the results from third order to higher order systems is presented. The system's limit cycle behavior is elaborated in Section 6.4. Conclusions are presented in Section 6.5.

6.2 Analysis of a Third-Order $\Sigma\Delta$ Topology

Fig.(6.1) illustrates the topology for a third-order $\Sigma\Delta$ modulator which has been considered in several works (e.g., [87][104][121]).

Assuming the input to be a dc signal of amplitude x, $u(k)$ to be the final integrator output (which is also the quantizer input), and $y(k)$ is the quantizer output such that $y(k) = \text{sgn}[u(k)]$ and is given by:

$$y(k) = \begin{cases} +1 & \text{for } u(k) \geq 0 \\ -1 & \text{for } u(k) < 0 \end{cases} \tag{6.1}$$

This system shown in Fig.(6.1), can be described using a third-order dif-

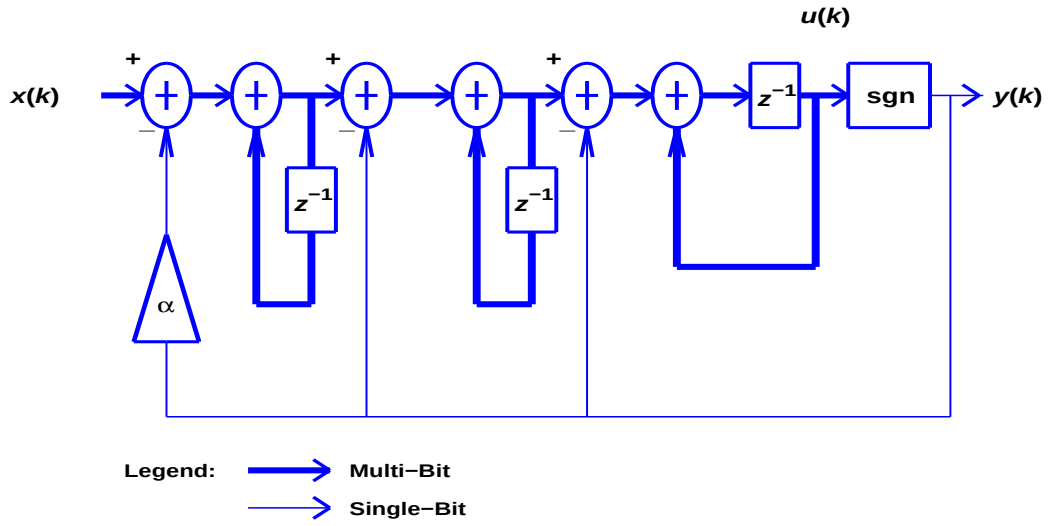

Figure 6.1: Structure of the third-order ΣΔ modulator under consideration.

ference equation as follows:

$$u(k) = 3u(k-1) - 3u(k-2) + u(k-3) - (\alpha+2)y(k-1) + 3y(k-2) - y(k-3) + x. \tag{6.2}$$

The above equation may be re-expressed recursively, with the right hand side using just the initial conditions and the DC input value as follows:

$$u(3) = 3u_2 - 3u_1 + u_o - (\alpha+2)y_2 + 3y_1 - y_o + x. \tag{6.3}$$

$$u(4) = 6u_2 - 8u_1 + 3u_o - 3(\alpha+1)y_2 + 8y_1 - 3y_o - (\alpha+2)y(3) + 4x. \tag{6.4}$$

and so on. This re-expression is derived in Appendix A to be:

$$\begin{aligned} u(k) \;=\; & \frac{1}{2}k(k-1)u_2 - k(k-2)u_1 + b(k)u_o - b(k)y_o + \\ & k(k-2)y_1 - [(k-1) - \alpha b(k)]y_2 - g(k,\alpha) + d(k)x \end{aligned} \tag{6.5}$$

where $u_o = u(1)$, $u_1 = u(2)$, and $u_2 = u(2)$ are the initial conditions of the final integrator output, with y_o, y_1, and y_2 being the corresponding quantizer output values [i.e., $y_i = \text{sgn}(u_i) | i \in \{0,1,2\}$]. The functions $b(k)$, $g(k)$, and

$d(k)$ are given as follows:

$$b(k) = \frac{(k-1)(k-2)}{2} \tag{6.6}$$

$$g(k,\alpha) = \sum_{n=1}^{k-3}[(\frac{\alpha}{2}n+1)(n+1)]y(k-n) \tag{6.7}$$

$$d(k) = \frac{kb(k)}{3}. \tag{6.8}$$

An asymptotic solution for the $\Sigma\Delta$ dynamical system is now considered. If (6.5) is divided by $d(k)$ and then the limit as $k \to \infty$ is taken, one obtains:

$$x \leftarrow g(k,\alpha)/d(k) = \frac{6}{k(k-1)(k-2)} \sum_{n=1}^{k-3}(\frac{\alpha}{2}n+1)(n+1)\text{sgn}(u_{k-n}), \quad \text{as} \;\; k \to \infty \tag{6.9}$$

It is difficult to find an analytic solution for this equation, largely because of the signum term. However, one can find an asymptotic solution (as $k \to \infty$) by replacing the k - dependent term $1/[(k-1)(k-2)]$ outside the summation by an n - dependent term inside the summation, i.e.,

$$\frac{1}{(k-1)(k-2)} \sum_{n=1}^{k-3} a(n) \to \sum_{n=1}^{k-3} \frac{a(n)}{f(n)} \;\; \text{as} \;\; k \to \infty \tag{6.10}$$

where $a(n) = (\frac{\alpha}{2}n+1)(n+1)\text{sgn}(u_{k-n})$. The function $f(n)$ can be given as (see Appendix B)

$$f(n) \to 3(n+\frac{2}{\alpha})(n+1) \;\; \text{as} \;\; k \to \infty. \tag{6.11}$$

Now (6.9) can be written as:

$$x \leftarrow \frac{\alpha}{k}\sum_{n=1}^{k}\text{sgn}(u_{k-n}) + \frac{2}{k}\sum_{n=1}^{k}\frac{\text{sgn}(u_{k-n})}{(n+2/\alpha)} \;\; \text{as} \;\; k \to \infty. \tag{6.12}$$

It can be proved that the second term on the right-hand side of (6.12) tends to zero since we have:

$$\frac{2}{k}\sum_{n=1}^{k}\frac{1}{(n+2/\alpha)}\text{sgn}(u_{k-n}) < \frac{2}{k}\sum_{n=1}^{k}\frac{1}{(n+2/\alpha)} \tag{6.13}$$

knowing that the signum function $\text{sgn}(u_{k-n}) \in \{1, -1\}$. Now, since the limit of the right-hand side of the inequality (6.13) is zero as $k \to \infty$, then the left-hand side will go to zero more rapidly (it is a transient term that decides the rate at which the system converges to the steady-state). Accordingly, for stable operation, the sequence $g(k, \alpha)/d(k)$ in (6.12) converges to x as $k \to \infty$, and in fact it is the average output if the equation is divided by α, i.e., average $= \frac{x}{\alpha} = \frac{1}{k}\sum_{i=1}^{k}\text{sgn}(u_i)$. We re-arrange this equation as follows:

$$\frac{x}{\alpha}k = \sum_{i=1}^{k}\text{sgn}(u_i), \quad k \to \infty. \tag{6.14}$$

Since the left-hand term is always a fraction and the right-hand term is always an integer, one expects (as it is in fact the case) that the system has no fixed-point or an equilibrium steady-state solution. Alternatively, this dynamical system can be characterized by time-varying states, i.e., by a periodic solution.

A periodic solution is a dynamical solution that is characterized by one basic frequency f_1. The spectrum of a periodic signal consists of a possible spike at zero frequency and spikes at integer multiples of the fundamental frequency f_1. The amplitudes of some of the harmonic frequency components may be zero. A periodic solution is called a limit cycle if there are no other periodic solutions sufficiently close to it. In other words, a limit cycle is an isolated periodic solution and corresponds to an isolated closed orbit in state space. Every trajectory near a limit cycle will approach it as $k \to \infty$. Consequently, (6.14) applies quite well when the system traps into stable limit cycles. Hence, the average output over any limit cycle can be given as:

$$\frac{x}{\alpha} = \frac{1}{L}\sum_{i=1}^{L}\text{sgn}(u_i) \tag{6.15}$$

where L is the period length of the stable limit cycle.

6.3 The High-Order $\Sigma\Delta$ Topology

In this Section, the focus is on high-order (> 2) $\Sigma\Delta$ structures as shown in Fig.(6.2). The general difference equation that describes the operation of the M^{th}-order topology depicted in Fig.(6.2)is derived in Appendix C to be:

$$u(k) = \sum_{n=1}^{M}(-1)^{n+1}\binom{M}{n}u(k-n) + \sum_{i=1}^{M}\sum_{n=0}^{M-1}(-1)^{i}\binom{n}{i-1}\alpha_n y(k-i) + x(k-1) \tag{6.16}$$

where $\{\alpha_n | n = 0, 1, 2, ..., M-1\}$ are the feedback parameters.

Adopting the same approach for calculating the average as above, we find that, for any order of the $\Sigma\Delta$ topology under investigation, the average output is:

$$\text{Average} = \frac{x}{\sum_{i=1}^{M} c_i} \tag{6.17}$$

where M is an integer that denotes the system order, and $\{c_i\}$ are the coefficients of the signum terms that appear in the system difference equation.

A few authors has confirmed a strong link between the limit cycle analysis and $\Sigma\Delta$M stability analysis. For example, [111, 24, 112] link saddle points in state trajectories (which constitute limit cycles) to system stability through a Poincaré map.

There are also hints in the literature that the non-linear dynamics of the *first-order* $\Sigma\Delta$ operation can be modelled using the dynamics of the standard map, and more specifically, the dynamics of the circle map [106, 53]. In this work, a comprehensive analysis of the third-order system under consideration is introduced based on circle map modelling and its approximation using fixed point analysis.

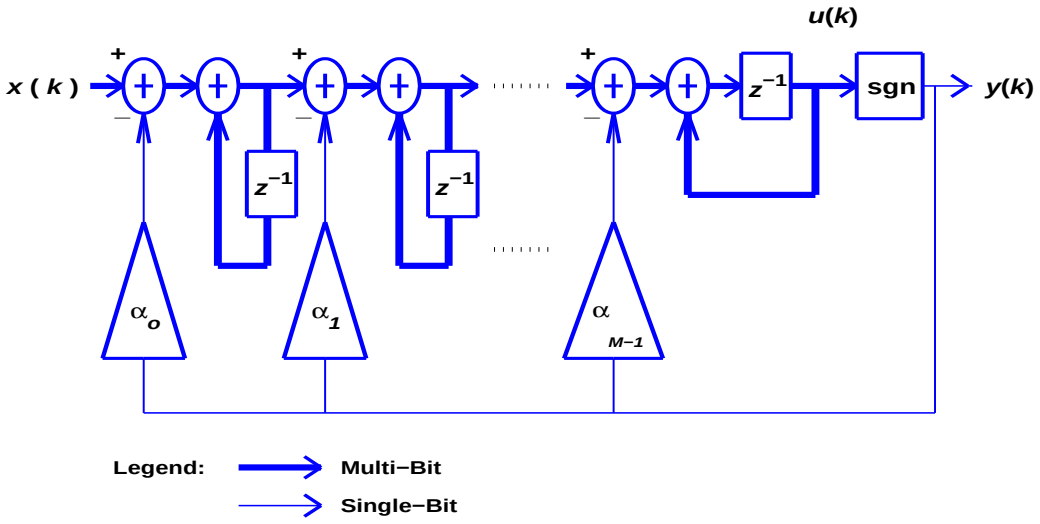

Figure 6.2: Structure of the M^{th}-order $\Sigma\Delta$ modulator under investigation.

6.4 Behavior of the System's Limit Cycles

In order to reveal the behavior of the system under investigation, MATLAB is utilized to develop a random search procedure to detect and extract the limit cycles from the system output. This is done in both the frequency domain, using high-resolution FFT, and time domain using autocorrelation function with variable lag lengths. The results can be compared and confirmed by inspection to discover the longest limit cycle sequence under specific operating conditions.

6.4.1 Limit Cycle Notation

Throughout this chapter, the cyclic sequences are described for optimum clarity as follows [41]:

$Q_{(i),j} = [q_1^+, q_1^-, ...q_{i-1}^+, q_{i-1}^-, q_i^+, q_i^-]$ where i denotes the number of transitions from +1 to -1 (or -1 to +1) within the limit cycle period, while the subscript j represents any integer number. The values between the brackets represent the number of successive outputs that constitute one cycle, that is the value q_i^+ represents the number of consecutive +1s , whereas q_i^- represents the number

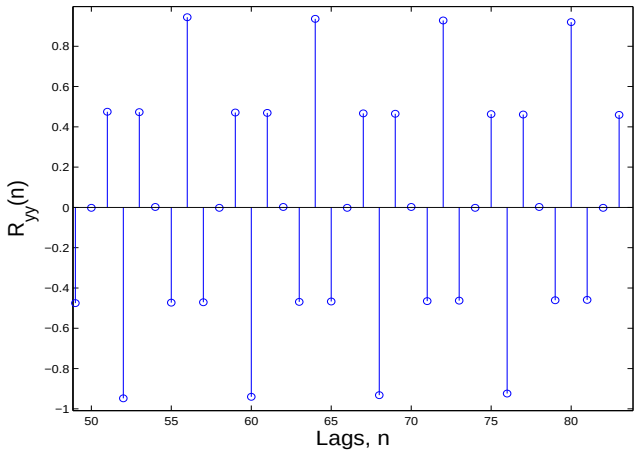

Figure 6.3: The autocorrelation function $R_{yy}(n)$ of the structure output for zero input. The initial conditions are: $u_o = .2$, $u_1 = .4$, and $u_2 = 0$, and $\alpha = 0.1$.

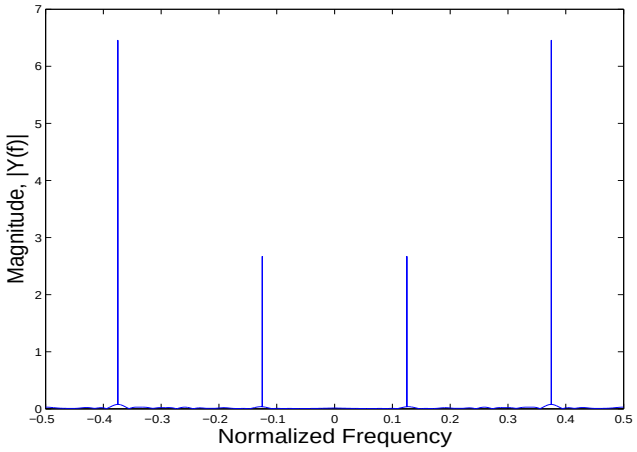

Figure 6.4: The output frequency spectrum under same parameters.

of consecutive -1s, both at the i^{th} transition. Therefore, the number of values within the brackets must be even and equals to $2i$ (i value for each positive and negative), as the number of positive outputs should equal to the number of negative ones.

The length of the limit cycle with a maximum number of transitions i (and hence a maximum length), is termed as $L_{\max}$. This is to distinguish the largest cycle from other limit cycles included inside it. Of course this does not mean that all possible limit cycles belong to a longer one, as there may exist some independent cycles.

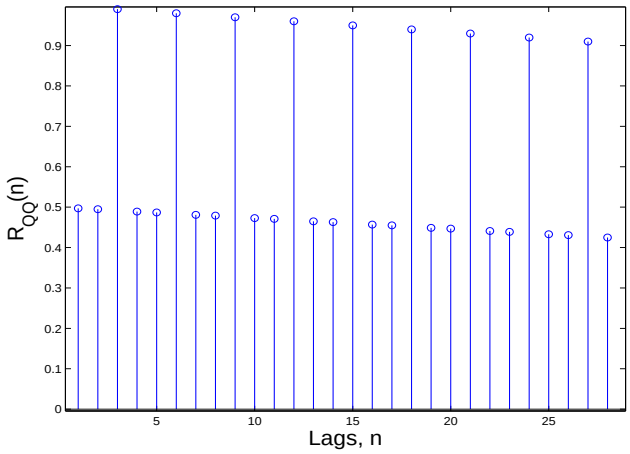

Figure 6.5: The autocorrelation function $R_{QQ}(n)$ representing the number of transitions within the limit cycle period.

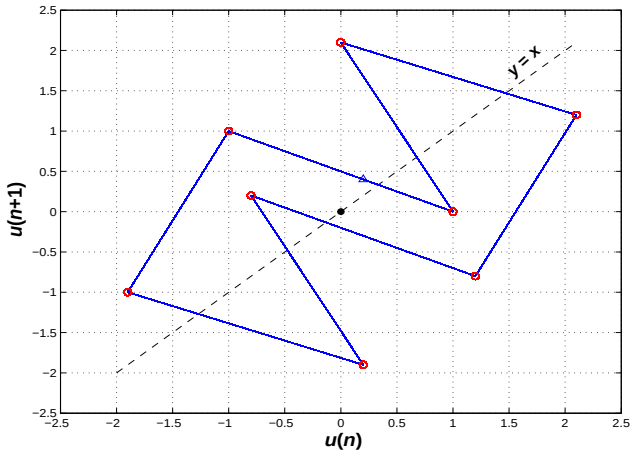

Figure 6.6: Phase-plane portrait of the third-order structure for $x = 0$. The initial conditions are: $u_o = .2$, $u_1 = .4$, and $u_2 = 0$, and $\alpha = 0.1$ (the diagonal straight line represents $y = x$).

6.4.2 Zero-Input Limit Cycles

The search program carries out the following:

1. Searching the output frequency spectrum $|Y(f)|$ to locate the tones that are above the quantization noise level. These tones correspond to limit cycles. The fundamental frequency and its harmonics are located and then $L_{\max}$ can be calculated as follows:

$$L_{\max} = f_s/f_o. \tag{6.18}$$

   where f_s is the sampling frequency, and f_o stands for the fundamental limit cycle frequency. The other frequency components are just multiples of f_o, i.e., mf_o, where $m = 1, 2, 3, ..$. In other words the term $L_{\max} = mL$ is associated with the fundamental frequency. However, one should be aware, as this is true only for stationary signals.
2. Constructing a sequence array Q by putting the bits of the output y with similar sign into groups, for example, if $y = 1, 1, 1, 1, -1, , 1, -1, ...$ then $Q_{(i),j} = [4^+, 1^-, 1^+, 1^-, ...]$. The autocorrelation function of Q, $R_{QQ}(n)$, gives us information about the transition number i within a limit cycle.

The following figures illustrate the approach utilized to discover and confirm the periodic patterns in the single-bit output by inspection. Fig.(6.3) shows the autocorrelation function $R_{yy}(n)$ of the single-bit output $y(k)$ under the initial conditions: $u_o = 0.2$, $u_1 = 0.4$, $u_2 = 0$, and $\alpha = 0.1$. The number of lags n (clock periods) in the autocorrelation function should be increased each time up to several hundreds to insure that no longer cyclic periods exist. The figure reveals the maximum limit cycle length, $L_{\max} = 8$, for zero input and under the specified parameters, in the third-order $\Sigma\Delta$ topology shown in Fig.(6.1). The same result can be obtained from the frequency domain as shown in Fig.(6.4). Furthermore, the number of transitions i within the limit cycle is calculated and can be found also by inspection as shown in Fig.(6.5).

It is obvious that $i = 3$. The main sequence of the limit cycles can now be easily found as:

$$Q_{(3),1} = [2^+, 1^-, 1^+, 2^-, 1^+, 1^-]$$

The system phase portrait with the same parameters is calculated using (6.5) and is depicted in Fig.(6.6). Exactly the same phase portrait is obtained by simulating the system equation (see (6.1)).

Now we explore the role of the initial conditions and the constant gain parameter α on the system operation. The constant parameter α has a major effect on the limit cycle behavior, as shown in Fig.(6.7) for $\alpha = 0.1$ and Fig.(6.8) for $\alpha = 0.2$ for new different initial conditions. It is obvious that $L_{\max} = 20$ and number of transitions = 14 in Fig.(6.7) with the following sequence:

$$Q_{(14),1} = [2^+, 3^-, 3^+, 2^-, 1^+, 1^-, 1^+, 1^-, 1^+, 1^-, \\ 1^+, 1^-, 1^+, 1^-],$$

while in Fig.(6.8) we have $L_{\max} = 40$, number of transitions = 22, and a mother sequence of:

$$Q_{(22),1} = [2^+, 3^-, 3^+, 1^-, 1^+, 3^-, 3^+, 2^-, 2^+, 3^-, 3^+, 2^-, \\ 1^+, 1^-, 2^+, 2^-, 1^+, 1^-, 1^+, 1^-, 1^+, 1^-].$$

To better understand the variation in the limit cycle behavior of the system that can take place as a result of changing the parameter α, Fig.(6.9) shows the limit cycle phase plane for $\alpha = 1/6$. In this case $L_{\max} = 200$. We found that for large values of $L_{\max}$, there is a likelihood that the system limit cycle contains a number of shorter limit cycles with length L such that L is a divisor of $L_{\max}$. Consequently crowded tones in output frequency spectrum will appear.

From the previous phase plane figures, it is evident that for zero-input, the state trajectories of the system are constructing straight lines in parallel with the diagonal $y = x$ line and they are critically dependent on the initial conditions.

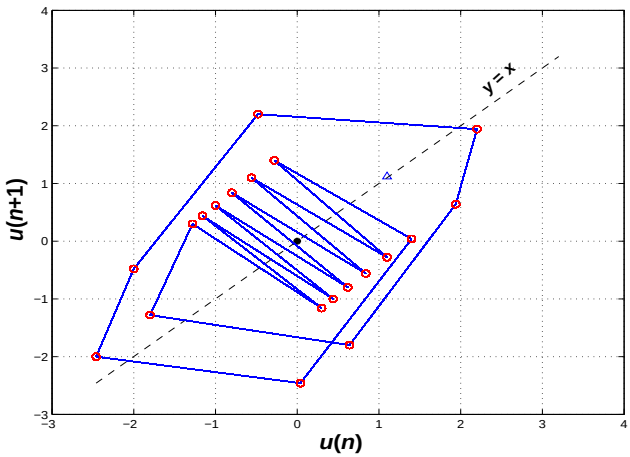

Figure 6.7: The system phase portrait with $\alpha = 0.1$ and initial conditions: $u_o = 1.1$, $u_1 = 1.11$, $u_2 = 1.2$.

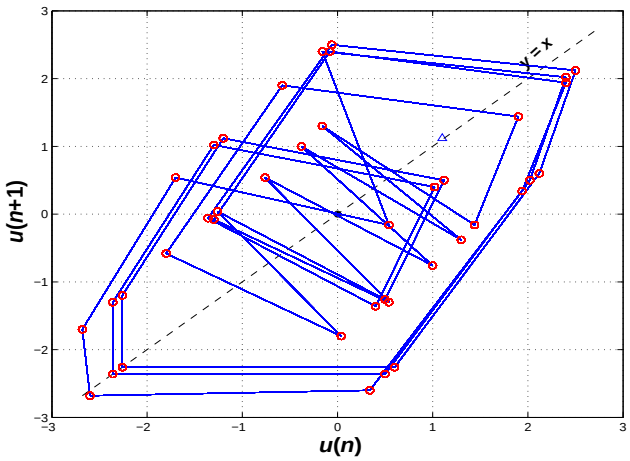

Figure 6.8: The system phase portrait with $\alpha = 0.2$ and initial conditions: $u_o = 1.1$, $u_1 = 1.11$, $u_2 = 1.2$.

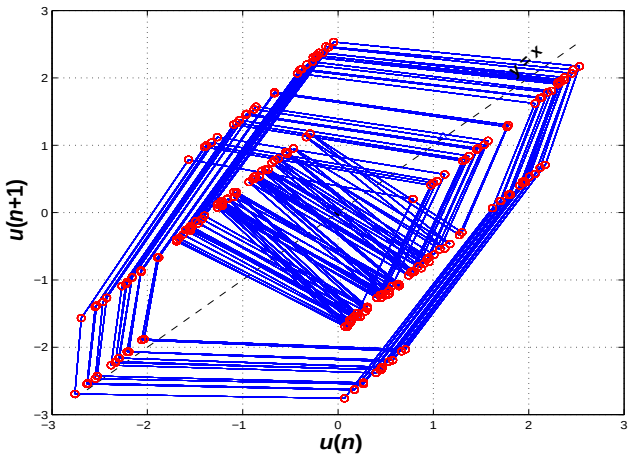

Figure 6.9: The phase portrait for $\alpha = 1/6$ and initial conditions of: $u_o = 1.1$, $u_1 = 1.11$, $u_2 = 1.2$.

6.4.3 Limit Cycles for Non-Zero DC Input

This third-order $\Sigma\Delta$ filter exhibits a highly non-linear behavior. For non-zero input, the parameter α possesses an important role in determining the system dynamics. It is evident during simulation that the system stability is extremely sensitive to the value of α. For certain initial conditions, α controls the input dynamic range upon which stability is maintained. Consequently there is always a threshold dc input value $x_{\max}$ beyond which unstable operation occurs. This can be seen in Fig.(6.10).

On the other hand, as anticipated (see (6.16)), α may alter the limit cycle behavior through varying both the transient and the steady state conditions of the system. This alteration extends to include the quantization noise structure as well.

To take a closer look at the nonlinear limit cycles behavior of the system, Fig.(6.11) depicts the maximum length of the limit cycles $L_{\max}$(that corresponds to the fundamental frequency) as a function of the dc input value for α=0.1 and under specific fixed initial conditions. The frequencies of these patterns normally reside in the baseband region, however, their power is relatively very low. This is due to the noise shaping effect of the $\Sigma\Delta$ modulator in this band of frequencies. While, on the contrary, shorter cycles that in fact

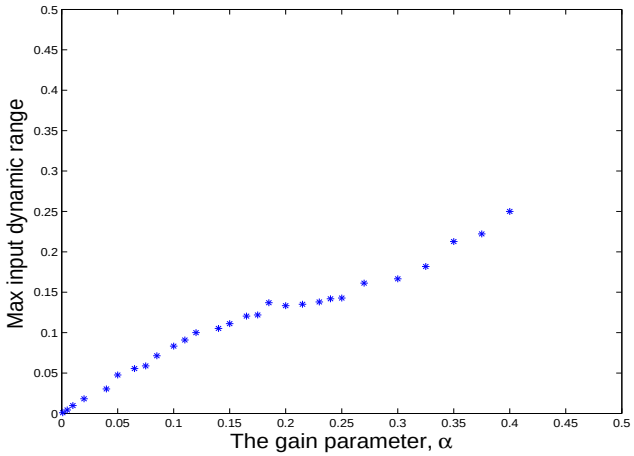

Figure 6.10: The parameter α versus the maximum threshold dc input beyond which no stability is guaranteed.

constitute a fraction of $L_{\max}$ ($L = L_{\max}/m$, where m is an integer), and consequently located in a higher frequency band, suffer several orders of magnitude less from attenuation and therefore, introduce the most sever contributions to the problem of instability. The value $L_{\max}$ increases according to the complexity of the fractions that represent the values of the initial conditions and/or the dc input. For simple fractions (e.g., 1, $\frac{1}{2}$,$\frac{1}{3}$, ...), $L_{\max}$ takes relatively small values, while for complex fractions (e.g., $\frac{111}{297}$), it will take larger values.

For non-zero input, the state space trajectories are tending to converge towards the diagonal line at their upper ends forming a semi-ellipsoids, and therefore they are no longer residing on straight lines as is the case for zero-input. This can be seen in Fig.(6.12). The sequence of the limit cycle is: $Q_{(8),1} = [10^+, 1^-, 1^+, 1^-, 12^+, 1^-, 1^+, 1^-]$

Now we summarize our findings regarding the limit cycles in the third-order structure under investigation:

1. The structure may contain long limit cycles (e.g., $L_{\max}$=200, 300, 500, ...) and this depends primarily on how complex the fraction of the initial conditions, α, and the input values are.
2. We found that the average value of the output when the system traps

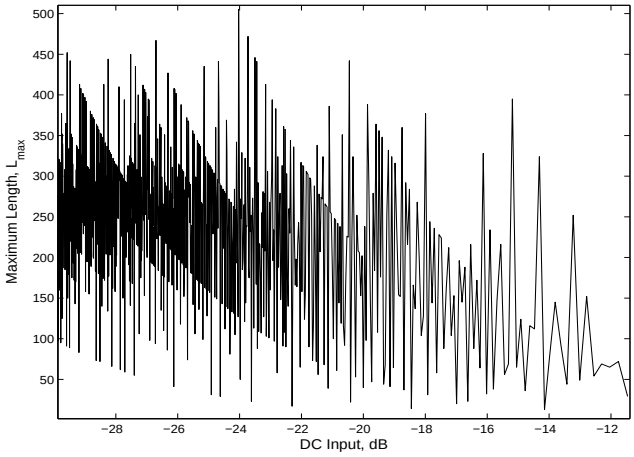

Figure 6.11: The maximum limit cycle length $L_{\max}$ as a function of the dynamic range input under fixed initial conditions and for $\alpha = 0.1$.

into a limit cycle is exactly equal to $\frac{x}{\alpha}$, and can be written as follows:

$$\frac{x}{\alpha}L = \sum_{k=0}^{L-1} \operatorname{sgn}(u_k) \tag{6.19}$$

This equation explains the reason behind the conclusion we just inferred in item 1 above, as on the left-hand side, $\frac{x}{\alpha}$ is always a fraction, while the right-hand side is always an integer. Hence, L must adjust the left side to an integer. So if $x = \frac{p}{q}$, where p and q are relatively prime, then we expect that L must be a integer multiple of $q\alpha$ (since α is a fraction as we will see later), i.e., $L = \alpha q \zeta$. Then, $p\zeta = \sum_{k=0}^{L-1} \operatorname{sgn}(u_k)$, where p and ζ are integers.

3. Regardless of the value of the initial conditions and the dc input, it is evident that the third-order $\Sigma\Delta$ system is driven into instability for the gain parameter $\alpha > 0.5$.

It is not the aim here to do an exhaustive limit cycle search. Consequently, the largest limit cycle $L_{\max}$ is then dependent on the system parameters and it is not the absolute maximum length. Our motive for specifying $L_{\max}$ is to find out the complete states that constitute the orbit of the system under specific parameters for the benefit of stability issue. A few authors has confirmed a

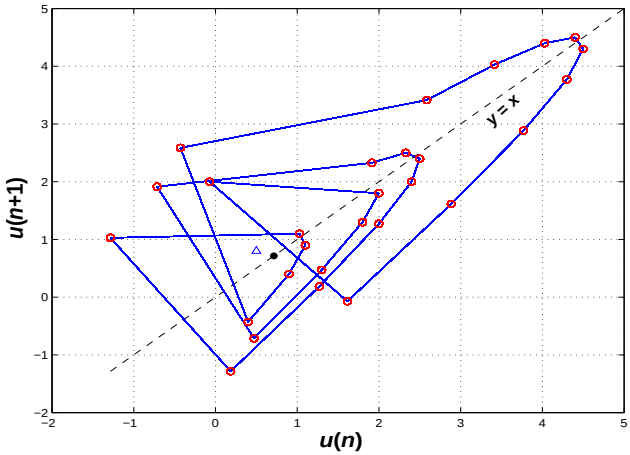

Figure 6.12: The phase plane for x=1/14, $\alpha = 0.1$ with initial conditions $u_o = 0.5$, $u_2 = 0.8$, and $u_3 = 0.8$.

strong link between the limit cycle analysis and ΣΔM stability analysis.

There are hints in the literature that the non-linear dynamics of the *first-order* ΣΔ operation can be modeled using the dynamics of the standard map, and more specifically, the dynamics of the circle map [106, 53]. In the next chapter, we shall introduce a comprehensive analysis of the third-order system under consideration based on circle map modeling and its approximation using fixed point analysis.

6.5 Conclusion

The third-order ΣΔ structure which is the core of the ternary structure was analyzed under dc input. The difference equation and the iterative solution that describe its operation are developed. It is shown that the system exhibits limit cycle behavior under certain conditions of the system parameters. The M^{th}-order difference equation of similar ΣΔ topologies are also developed. Moreover, a general formula for obtaining the average output of these systems is derived. The system was then simulated extensively and a random search method is utilized to discover and extract the limit cycles and identify their features. It seemed that this topology, which is a third-order ΣΔ modulator, possesses a highly non-linear behavior.

The work that has been done in this chapter will be of significant importance to address the stability issue of the ternary structure (and can be extended to the stability analysis of higher-order $\Sigma\Delta$ modulators as well) as will be seen in the next chapter.

Chapter 7

A Stability of Sigma-Delta Modulators in Ternary Structures

7.1 Introduction

Higher-order (> 2) single-bit $\Sigma\Delta$ modulators ($\Sigma\Delta$M's) are of increasing importance in many applications due to their improved performance as compared to the first- and second-order structures [40]. A comparison between low-order and high-order $\Sigma\Delta$ structures is addressed in [113, 80]. However, the stability of higher-order $\Sigma\Delta$ modulators can be an obstacle to their adoption in digital signal processing (DSP) applications. Despite the large body of work that has already been done, the stability issue is still not fully resolved. The approaches utilized to address the issue fall into one of two categories. The first category is the linear system approximation approach (e.g., in [114]). This approach suffers from inevitable drawbacks as it is unable to explain important phenomena such as limit cycles and chaos [112]. The second category incorporates all the truly non-linear analysis techniques attempts to better model the behavior of $\Sigma\Delta$ modulators have adopted nonlinear analysis techniques (e.g., in [107] and [115]). In [115], for example, a first-order $\Sigma\Delta$ modulator along with a bang-bang phase-locked loop (PLL) system was modelled using the maps of driven interval shifts.

In this chapter, we attempt to set out a comprehensive analysis to the third-order $\Sigma\Delta$ topology utilized in ternary filters, both mathematically and

by simulation. In Subsection 7.2.1, we utilize the circle map dynamics to accurately model the operation of the $\Sigma\Delta$ structure, which is treated as a third-order sinusoidal digital phase-locked loop system. Accordingly, the stability topic is addressed using the fixed point techniques in subsection 7.2.3. This analysis would be of great importance to other higher-order $\Sigma\Delta$ structures after some appropriate modifications.

7.2 Stability Analysis of the Third-Order Topology

In general, an orbit $O(u_o)$ of a discrete dynamical system F: $\mathbf{R}^n \rightarrow \mathbf{R}^n$ is said to be *stable* if for every $r > 0$ there exists $d > 0$ such that the Euclidian distance between the system's state variables $\mathbf{u}$ and $\mathbf{y}$, $\|y_o - u_o\| \leq d$ implies $\|y_n - u_n\| \leq r \ \forall \ n \geq 1$, where $\mathbf{u,y} \in \mathbf{R}^n$. An orbit that is not stable is called *unstable* [116]. In other words, $O(u_o)$ is unstable if there exists $r(u_o) > 0$ such that for every positive number d one can find an initial state y_o, $\|y_o - u_o\| \leq d$ whose orbit is not contained in the closed ball $D(u_o, r(u_o))$. Fig.(7.1) and Fig.(7.2) depict a stable set of limit cycle points (for different values of x under certain parameters) which is revealed by the bounded-orbit or the *attractor* to which the system evolves after a sufficiently long time.

However, higher-order (> 2) $\Sigma\Delta$ modulators (including the system considered earlier) suffer from well-known stability problems [111]. In simulation it was found that for such systems to attain stability, the integrators should be leaky in the sense that “sub-unity" integration gains should be introduced as shown in Fig.(7.3). Hence, for the stability analysis, the system in Fig.(7.3) will be considered, and dynamical system analysis will be utilized to model its structure with controllable values for d_1, d_2, and d_3 inside $(0, 1]$. Then the stability criteria will be determined and there will be an attempt to extend the stable region of operation by adjusting the state trajectories of the system integrators.

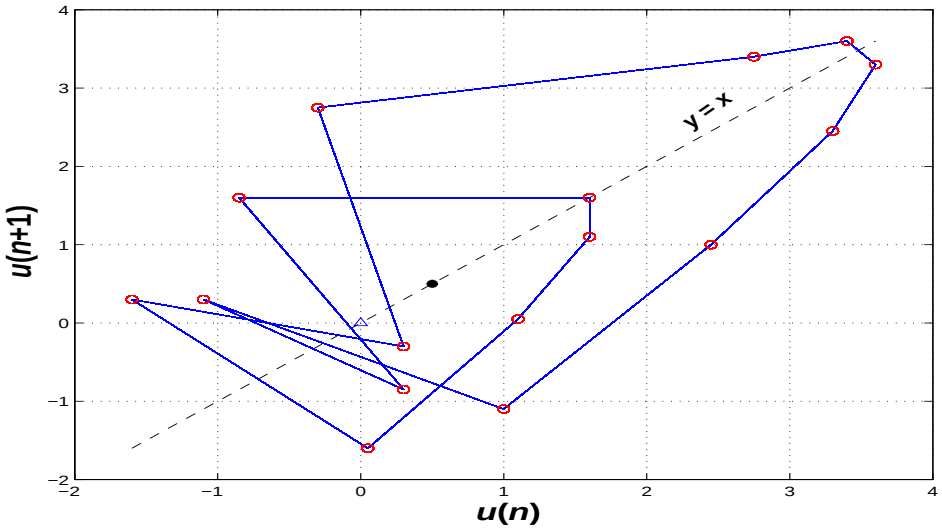

Figure 7.1: An attractor of third-order ΣΔ system with x=1/20, α=0.1, and initial condition set (0,0,-0.3) .

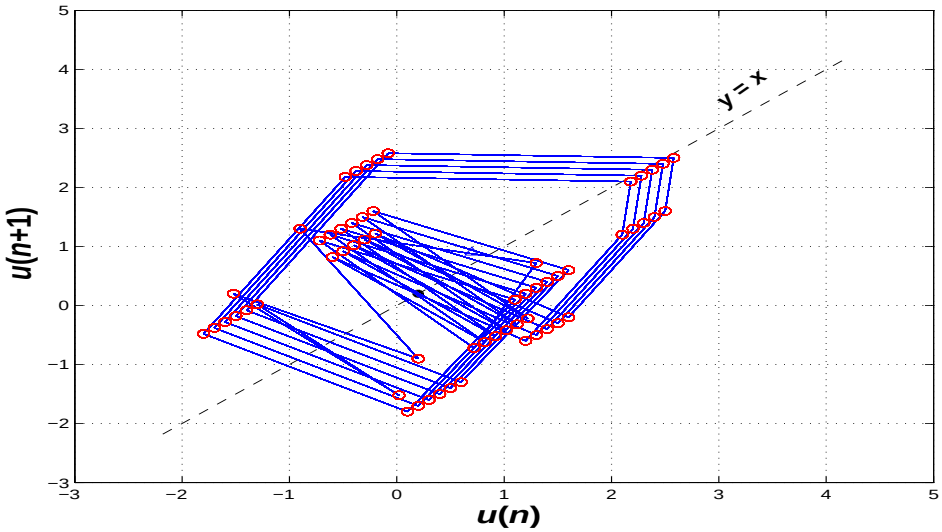

Figure 7.2: An attractor of third-order ΣΔ system with x=1/50, α=0.1, and initial condition set (0.7,0.9,1) .

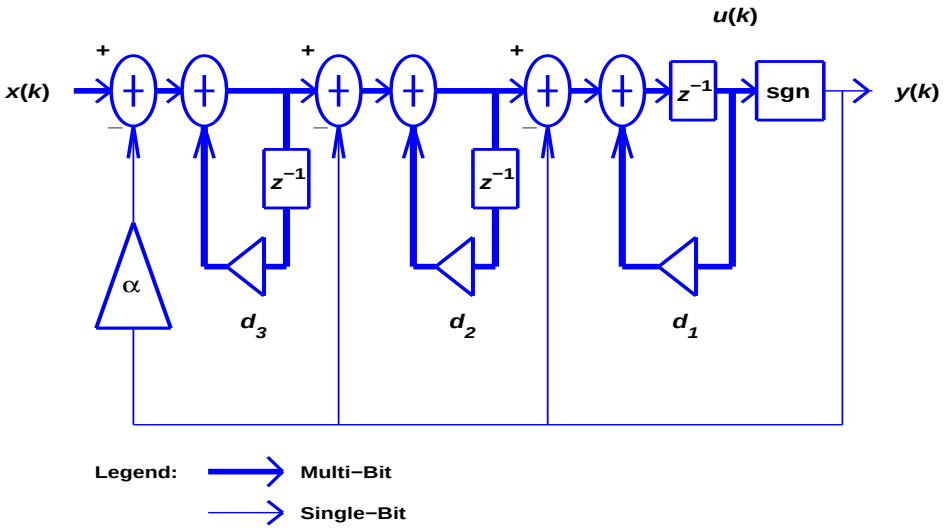

Figure 7.3: Structure of the single-bit third-order ΣΔ modulator.

7.2.1 Non-Linear Dynamics Modeling

The non-linear dynamics of the ΣΔ operation can be modelled using the dynamics of the standard map, and more specifically, the dynamics of the circle map. Here, the exact circle map is introduced that corresponds to the modified (arbitrary values for the d's) third-order system shown in Fig.(7.3). The circle map, also known as the *sine map*, is given by [117]:

$$u_{n+1} = F(u_n) = [u_n + \Omega - \frac{K}{2\pi}\sin(2\pi u_n)] \mod 1 \tag{7.1}$$

where K and Ω are the map parameters. The term Ω is confined to the interval $[0, 1]$. This map is a special case of the two-dimensional standard map. The state F maps the interval $[0, 1)$ onto itself when the circle map is confined to the interval $[0, 1)$ by using the mod 1 function. The circle map becomes piecewise linear when $K = 0$ and nonlinear when $K \neq 0$. For $1 > K \geq 0$, the circle map is an orientation preserving diffeomorphism. At $K = 1$, the map is a homeomorphism. For $K > 1$, the circle map becomes noninvertible (since it is not one-to-one, which implies the coexistence of different periodic oscillations) and critically dependent on the initial conditions.

Inspired by (6.2) and the circle map above, the $\Sigma\Delta$ system dynamics are formulated by the following non-linear circle map, with the parameters K_1, K_2, and K_3 taken from 6.2) as follows:

$$\begin{aligned} u_{k+1} &= K_1 u_n - (\alpha+2)\sin(m\gamma_k) - K_2 u_{k-1} + (d_2+2d_3)\sin(m\gamma_{k-1}) + \\ & K_3 u_{k-2} - (d_2 d_3)\sin(m\gamma_{k-2}) + x \end{aligned} \tag{7.2}$$

where $\gamma_k = \tan^{-1}(u_k)$, $\gamma_{k-1} = \tan^{-1}(u_{k-1})$, and $\gamma_{k-2} = \tan^{-1}(u_{k-2})$ are the phase angles that correspond to the integrator states u_k, u_{k-1}, and u_{k-2}, respectively, while m is an integer and $K_1 = (d_1+d_2+d_3)$, $K_2 = (d_1d_2+d_1d_3+d_2d_3)$, and $K_3 = (d_1d_2d_3)$. The d's are the gain parameters of the system integrators. As m increases, the behavior of the map approaches the dynamics of the $\Sigma\Delta$ modulator.

7.2.2 Traditional Stabilizing Design Approach

If a vector is defined as $X_k = (x_k\ y_k\ z_k)^T$, then (7.2) can be written alternatively as:

$$X_{k+1} = AX_k + B_k \tag{7.3}$$

where

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ K_3 & -K_2 & K_1 \end{pmatrix} \tag{7.4}$$

$$B_k = \begin{pmatrix} 0 \\ 0 \\ b_k \end{pmatrix}; \tag{7.5}$$

with $b_k = -(\alpha+2)\sin\{\tan^{-1}(z_k)\}+(d_1+d_2)\sin\{\tan^{-1}(y_k)\}-(d_1d_2)\sin\{\tan^{-1}(w_k)\}+x$. Taking a Euclidean norm of (7.3) yields

$$\|X_{k+1}\| \leq \|AX_k\| + \|B_k\| \leq \|A\|\|X_k\| + \|B_k\|.$$

First, it is obvious that $\|B_k\| < M_B$ (where $M_B > 0$ is a constant) given the boundedness of b_k.

If $\|A\| < 1$ (hence, $d_i < 1 \forall i$), then

$$\begin{aligned} \|X_{k+1}\| &\leq \|A\|^{k+1}\|X_0\| + \sum_{i=0}^{k} \|A\|^i \|B_{k-i}\| && (7.6)\\ &\leq \|A\|^{k+1}\|X_0\| + M_B \sum_{i=0}^{k} \|A\|^i && (7.7)\\ &= \|A\|^{k+1}\|X_0\| + M_B \frac{1-\|A\|^{k+1}}{1-\|A\|}. && (7.8) \end{aligned}$$

Obviously, for $\|A\| < 1$ we have

$$\lim_{k\to\infty} \|X_{k+1}\| \leq \frac{M_B}{1-\|A\|}. \quad (7.9)$$

That means the trajectory will converge within the boundaries constrained by (7.9) in the state space.

Note that if the matrix A is Hurwitz (that is, *all* its eigenvalues are located within the unit circle), then $\|A\| < 1$. Since $K_1,\ K_2,\ K_3$ are constant, they can easily be chosen to make $\|A\| < 1$. One way to achieve this result is through the application of the Routh-Hurwitz stability criteria. This in fact means that we can stabilize the system by adjusting the location of the system's poles, i.e., confining them within the unit circle.

7.2.3 Fixed Point Approximation: An Analogy with DPLL

We propose here the techniques adopted in analyzing the sinusoidal digital phase-locked loop (DPLL) to study the stability issues of the high-order $\Sigma\Delta$

modulator. From this point of view, the IIR loop operates on the principle of "tracking" the quantizer output, as the DPLL tracks the input frequency. As the stability of a periodic orbit of a continuous-time system may be determined by examining the stability of a fixed point of the associated map [117], the first step in this approach is to choose a suitable fixed point solution for our system. Intuitively, this would be the average output of the third-order $\Sigma\Delta$ modulator. This means that, under stable operation, the state trajectories are attracted to this point in an oscillatory behavior. Recalling (6.17), the proposed fixed point u^* is given as:

$$u^* = \lim_{k\to\infty} u_k = \tan^{-1}\left(\frac{x}{2+\alpha-(d_1+2d_2)+d_2d_3}\right). \tag{7.10}$$

Now, that the fixed point solution is obtained, it is necessary to find the range of filter parameters to meet the conditions that are necessary for the iterates of equation (6.5) to converge locally to the solution given by (7.10). For that, Ostrowski's theorem can be applied [118][119][120] if the function $F(u_k)$, which is given by (7.2) to be tested is continuously differentiable at the fixed point u^*. In this case, Ostrowski's theorem says that $\lim_{k\to\infty} u_k = u^*$ if:

$$\rho[F'(u^*)] < 1 \tag{7.11}$$

where $F'(u)$ is the partial derivative of the $n \times n$ matrix $F(u)$, $\rho(.)$ is the spectral radius of the matrix and is defined as follows:

$$\rho[F'(u^*)] = \max|\lambda_i|, \quad \lambda_i \equiv \text{Eigenvalues of } F'. \tag{7.12}$$

It is worth noting that in the case of nonlinear mappings, the condition $\rho[F'(u^*)] < 1$ is sufficient, but not necessary for convergence. While in the case of linear mappings, $\rho[F'(u^*)] < 1$ is both necessary and sufficient [118].

Now, reconsider (7.2), which models the dynamics of the structure shown in Fig.(7.3). For convenience, (7.2), which is a third-order equation, is trans-

formed into a system of three first-order equations in the following form:

$$u_{k+1} = F(u_k). \tag{7.13}$$

Let $w_k = u_k$, $y_k = u_{k+1}$, $z_k = u_{k+2}$. Therefore, (7.2) can be re-written in a matrix form as follows:

$$\begin{pmatrix} w_{k+1} \\ y_{k+1} \\ z_{k+1} \end{pmatrix} = \begin{pmatrix} y_k \\ z_k \\ F(z_k) \end{pmatrix} \tag{7.14}$$

where
$F(z_k) = K_1 z_k - (2+\alpha)\sin\{\tan^{-1}(z_k)\} - K_2 y_k + (d_2 + 2d_3)\sin\{\tan^{-1}(y_k)\} + K_3 w_k - (d_2 d_3)\sin\{\tan^{-1}(w_k)\} + x.$

To define a region of stability for the ternary-$\Sigma\Delta$ topology, consider (7.14). If $F(z_k)$ and $F'(z_k)$ are assumed to be continuous, then the Jacobian matrix of $F(z_k)$ is given by:

$$F'(z) = \begin{pmatrix} \frac{\partial f_1}{\partial w} & \frac{\partial f_1}{\partial y} & \frac{\partial f_1}{\partial z} \\ \frac{\partial f_2}{\partial w} & \frac{\partial f_2}{\partial y} & \frac{\partial f_2}{\partial z} \\ \frac{\partial f_3}{\partial w} & \frac{\partial f_3}{\partial y} & \frac{\partial f_3}{\partial z} \end{pmatrix}$$

$$\text{hence, } F'(z) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ h_1(w) & h_2(y) & h_3(z) \end{pmatrix} \tag{7.15}$$

where $h_1(w) = K_3[1 - \cos(w)]$, $h_2(y) = -K_2[1 - \cos(y)]$, and $h_3(z) = K_1[1 - K_1' \cos(z)]$. If we assume the fixed point is $\beta = u^* = \tan^{-1}[\frac{x}{(2+\alpha)-(d_1+2d_2)+(d_2 d_3)}]$, then at this point all the eignvalues must be less than one, i.e, $|\lambda_i| < 1$, where

$\lambda_i|i \in 1,2,3|$ should satisfy the characteristic equation: $|F'(\beta) - \lambda I| = 0$. Solving for λ, the characteristic equation is given by:

$$\lambda^3 - h_3(z)\lambda^2 - h_2(y)\lambda - h_1(w) = 0. \tag{7.16}$$

To extract the stability bounds from the characteristic equation, the following bilinear transformation [118]that maps the interior of the unit circle to the left-half plane (one-to-one map), is used:

$$\lambda = \frac{\psi + 1}{\psi - 1}. \tag{7.17}$$

Hence, (7.16) will be transformed as follows:

$$\psi^3(1-h_3-h_2-h_1)+\psi^2(3-h_3+h_2+3h_1)+\psi(3+h_3+h_2-3h_1)+(1+h_3-h_2+h_1) = 0. \tag{7.18}$$

It is now possible to apply the Routh-Hurwitz stability criteria, which allows a check for stability without computing the roots of the characteristic equation and can be used to determine the range of parameters that guarantees stability. One starts by building the Routh-Hurwitz array as shown in Table-7.1.

Table 7.1: Routh-Hurwitz array

Column-1	Column-2
$A = 1 - h_3 - h_2 - h_1$	$C = 3 + h_3 + h_2 - 3h_1$
$B = 3 - h_3 + h_2 + 3h_1$	$D = 1 + h_3 - h_2 + h_1$
$E = -(AD - BC)/B$	0
D	0

As the number of roots with positive real parts is equal to the number of sign changes in the first column, the elements of column-1 in the above array

should be all positive to ensure stability of the system, that is:

$$\begin{aligned}(1 - h_3 - h_2 - h_1) &> 0\\ (3 - h_3 + h_2 + 3h_1) &> 0\\ E &> 0\\ (1 + h_3 - h_2 + h_1) &> 0.\end{aligned} \tag{7.19}$$

Generally, useful conditions can be obtained from these inequalities. However, the fourth inequality ($D > 0$) is of particular interest. It provides an important criterion, that is:

$$\cos(\beta) < \frac{1 + K_1 + K_2 + K_3}{(\alpha + 2) + (d_2 + 2d_3) + (d_2 d_3)} \tag{7.20}$$

where $K_3 = (d_1 + d_2 + d_3)$, $K_2 = (d_1 d_2 + d_1 d_3 + d_2 d_3)$, and $K_1 = (d_1 d_2 d_3)$. This equation imposes a condition on the input dynamic range x in terms of the gain parameters (α, d_1, d_2, d_3) such that system stability can be preserved.

It is worth noting that, (7.20) can be generalized to represent any order of $\Sigma\Delta$ modulators when rewritten as follows:

$$\text{Average} = \frac{|x|}{\sum_{i=1}^{M} c_i} < \tan\{\cos^{-1}(\frac{1 + \sum_{i=1}^{M} |a_i|}{\sum_{i=1}^{M} |c_i|})\} \tag{7.21}$$

where M stands for the system order, $\{a_i\}$ is the set of the coefficients of the state space variables (u_i), and $\{c_i\}$ is the set of coefficients of their corresponding signum functions $[\text{sgn}(u_i)]$.

The stable input dynamic range for the third-order $\Sigma\Delta$ modulator shown

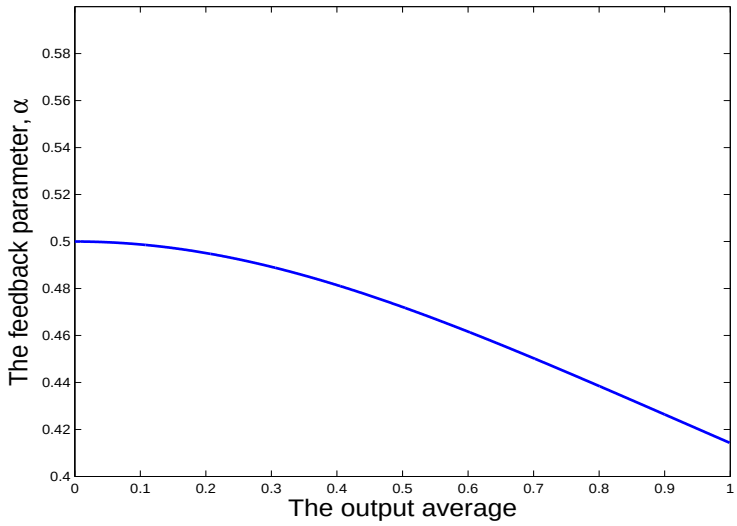

Figure 7.4: The theoretical boundary of the gain parameter α versus the average output according to (7.23).

in Fig.(7.3) with $d_1 = d_2 = d_3 = 1$ is given by:

$$x < \alpha \ \tan[\cos^{-1}(\frac{8}{6+\alpha})]$$

i.e.,

$$|x| < \alpha \sqrt{(\frac{8}{6+\alpha})^2 - 1} \tag{7.22}$$

while the stable feedback parameter range is confined to the interval (0,0.5) since:

$$\alpha < \frac{\cos(\beta)}{1+\cos(\beta)} \tag{7.23}$$

Fig.(7.4) shows the theoretical boundary of the feedback parameter α versus the average output (x/α) of the system. Fig.(7.5) illustrates the theoretical stability region (the shaded region) imposed by the intersection of the conditions obtained in (7.22) and (7.23). The boundary of this region is compared with the simulated boundary. As such we have shown that the fixed-point approximation that we suggested earlier lines up closely with simulation results.

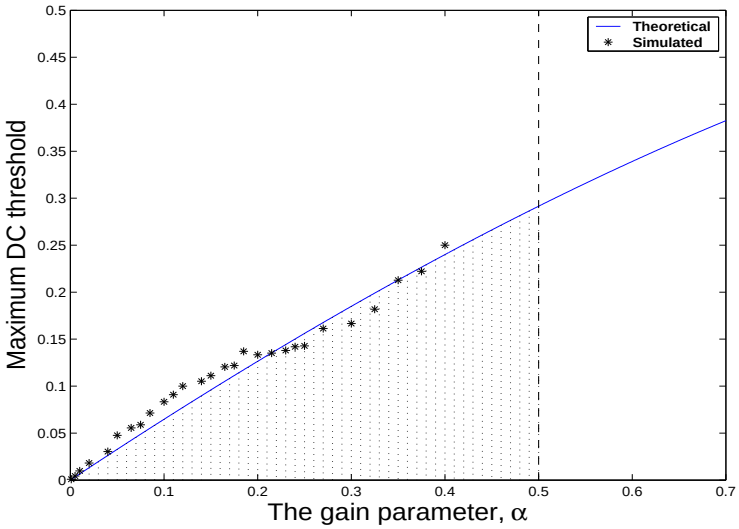

Figure 7.5: Stability region (shaded) of the third-order structure (for zero average output).

7.3 Summary

The behavior of a third-order $\Sigma\Delta$ structure was analyzed under dc input. The stability problem was addressed using an analogy between the dynamics of the $\Sigma\Delta$ structure and the sinusoidal digital PLL system. An approximate fixed point analysis was presented and a stability criteria was derived. Simulation results were in accord with the theoretical expectations. This analysis can be extended to any higher-order Sigma-Delta topology.

Chapter 8

Short-Word Length LMS-Like Adaptive Filtering

8.1 Introduction

Several recent works have made the theory of ternary filtering nearly mature and ready for application, e.g. [80, 87, 121]. However, for a new filtering theory to be of large scale application as a substitute for the traditional multi-bit DSP systems, efficient adaptive structures are inevitable. This is so because most applications are challenged by noise, distortion, and time-varying conditions. In fact, one of the major drawbacks that hindered analog signal processing (ASP) for decades was the lack of adaptivity.

Now days the demand for adaptive filtering can be found in nearly all applications, especially in communication systems.

Unfortunately, there is no adaptive LMS structure of any kind for short-word length (ternary or single-bit) filtering. The challenge in this problem is the harsh quantization that prevents straightforward LMS application.

The conventional infinite-precision LMS adaptive approach has proved to be efficient in finding optimal minimum mean-square solution for a wide variety of linear estimation problems. That approach is based on the recursive application of the steepest descent principle to direct the weight-vector towards the optimal solution predicted by Wiener-Hopf equations [123]-[125]. However, two sources of noise errors can be distinguished here, which degrade

the performance. First, the adaptation error (weight update misadjustment) is inherent to this approach due to the crude approximations adopted in calculating the instantaneous gradient. Second, the finite-precision noise, which arises in the practical (digital) implementation of the algorithm. The input data and internal calculations are all quantized to a finite-precision which is determined by design and cost considerations [126]. Accordingly, the quantization process causes the performance of a digital implementation to deviate from the corresponding theoretical design. The effects of these two sources of noise errors differ in their impact on the performance of the LMS filtering algorithms. For instance, gradient noise errors are more considerable than finite-precision errors during the transient stage; whereas the finite-precision errors become more significant during steady-state (as the adaptation errors get smaller) and consequently cause performance degradation in the form of excess mean-square error [127][128].

Unfortunately, when the LMS adaptive algorithm is applied to a short-word-length system (single-bit or ternary), the effect of the harsh quantization process (which is a nonlinear process) has a severe impact on the operation of the standard LMS algorithm and its variations, where they fail to converge to the Wiener solution. Single-bit systems enjoy very attractive properties as compared to their multi-bit counterparts. The single-bit implementation produces a relatively higher performance with lower hardware complexity; however, their useability in practice (e.g., in communication systems) is very limited due to their unresolved adaptivity problem.

To compensate for the reduced number of bits used in the quantization process, the short-word length systems require operation at an oversampled rate. The oversampling ratio (OSR) is a key parameter in these systems and is defined as ratio of the actual sampling rate to the Nyquist rate (normally OSR>>1). Therefore, although their filter order needs to be interpolated by a factor of OSR, short-word systems are extremely efficient from hardware implementation viewpoint [4],[80]. The core element in these systems is the

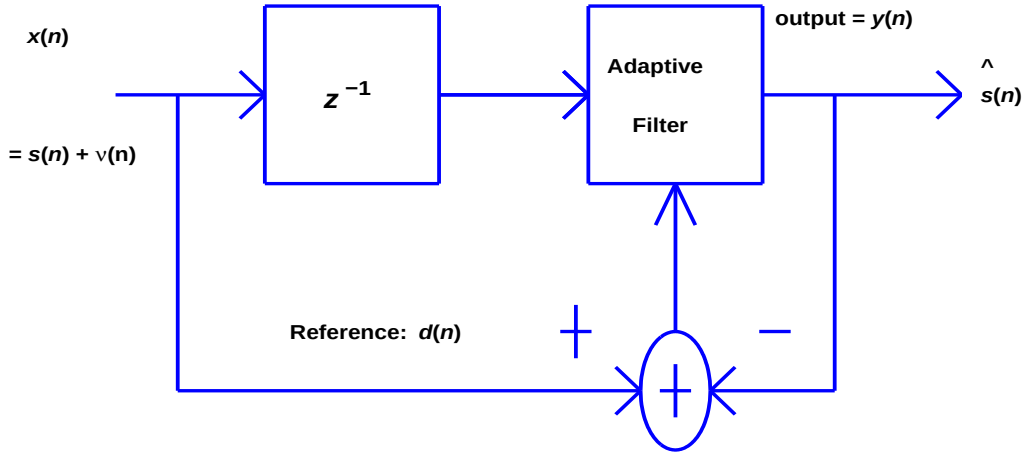

Figure 8.1: Adaptive multi-bit noise cancelling.

the sigma-delta (SD) modulator. Several works have addressed the topic of adaptivity in the single-bit domain (e.g., [129], [63], [73], [130], [74]), either by attempting to optimize the performance (SNR, stability, etc) or to increase the dynamic range. However, there has been no attempt towards the topic of single-bit adaptive filtering. In [131], a SD structure for adaptive LMS filtering which circumvent the interpolation and decimation requirements is introduced; however, the LMS algorithm is still achieved using multibit representation. Moreover, the oversampled SD structure requires more multibit arithmetic operations per time unit as compared to a PCM system.

A key application of adaptive systems is noise cancelling. For estimation of narrow-band signals with unknown frequency content, the adaptive realization of a Wiener filter (using Widrow-Hoff LMS algorithm) can be utilized with the desired (reference) signal $d(k)$ chosen to be the noisy input sequence $x(n) = s(n) + \nu(n)$ itself, while a delayed version of $x(n)$ is chosen as the filter input as shown in Fig.(8.1) (n being the sampling index). This structure is based on the fact that noise ν and original signal s are uncorrelated [132].

In this chapter we attempt to solve this problem by proposing an efficient structures for adaptive noise reduction. This application is of major significance in many applications, such as communication channel equalization. The

results were quite astonishing as the performance of these structures are comparable to that of multi-bit LMS Wiener algorithm. Two short-word length adaptive structures are proposed, namely, *ternary* and *single-bit* adaptive filters.

In Section 8.2 an adaptive ternary LMS-like algorithm is introduced. Performance assessment using a sinusoidal input distorted by additive white Gaussian noise showed that the proposed algorithm is comparable to the traditional multi-bit Wiener LMS algorithm. We expect that this approach will open the door for ternary systems to be ready for replacing multi-bit signal processing systems.

A single-bit adaptive LMS-like filtering is introduced , analyzed, and simulated in Section 8.3. The input (noisy) signal, the output (estimated) signal, and the FIR filter coefficients are all in single-bit (ternary) format; as such the system would be simple to implement using FPGA technology. The need for decimators, interpolators, and multibit multipliers is now eliminated. The proposed adaptive structure is shown to converge in the LMS sense. As the algorithm processes blocks of input data, it will be called SBLL (Single-Bit Block LMS-Like) to highlight the similarities with the standard block LMS. Subsection-8.3.1 outlines the main features of the proposed adaptive filtering and an approximation to the gradient function is addressed. In Subsection-8.3.2 a structure is set up and analyzed. In Section-8.4, the convergence properties are discussed and compared to the conventional infinite-precision LMS algorithm. Performance evaluation in terms of SNR, stability, and response to non-stationary inputs is addressed.

8.2 An Adaptive Ternary Algorithm:

To the best of our knowledge, the issue of ternary adaptive algorithm has not been addressed yet, and is considered as an unresolved problem.

We propose a structure inspired from the well-known LMS adaptive tech-

niques. Fig.(8.2) illustrates the ternary structure that carries out the proposed adaptive algorithm. We assume that the received (observed) signal, $r(n)$, is in single-bit format that represents the digitized original signal $x(t)$ distorted by white Gaussian noise $\eta(t) \in \mathcal{N}(\sigma^2, 0)$. The same scenario in Fig.(8.2) can be used to represent a baseband version of a digital single-bit communication system with bandpass modulation [122].

The symbol $\hat{x}(i)$ stands for the multibit estimated signal, and $y(i)$ is the *estimated* signal in single-bit format. The operation of this adaptive structure can be described as follows. The ternary system is comprised of M adaptable taps and operates at an oversampling rate R ($R = 64, 128, \cdots$). This requirement has already been met as the input signal is assumed to be $\Sigma\Delta$ modulated. The single-bit estimated signal $y(i)$ is loaded sequentially into a shift register of length M, where the register content can be expressed by the vector:

$$\mathbf{y}(i) = [y(i), \cdots, y(i - M - 1)].$$

Likewise, the regressor vector of the received single-bit signal will be: $\mathbf{r}(i) = [r(i), \cdots, r(i - M - 1)]$, which is assumed to be in the form $\mathbf{r} = \mathbf{x} + \eta$, where $\mathbf{x}$ is the original signal vector in single-bit, and η is the single-bit noise vector.

This structure updates the ternary coefficients (taps) $\{h(j)|j = 0, 1, 2.., M-1\}$ once every Δ samples, where Δ is dependent on the oversampling ratio R, i.e., at Nyquist rate. This weights updating can be expressed as follows:

$$\mathbf{h}_n = \frac{1}{2}(\mathbf{r}_n - \mathbf{y}_n). \tag{8.1}$$

where $n = i \bmod \Delta$, and ubscripts are used for time indexing instead of the brackets. The multi-bit estimated signal $\hat{x}$ at any instant i is given by:

$$\hat{x}_i = \alpha \mathbf{h}_{n-1}^T \mathbf{r}_i \tag{8.2}$$

where α is a small positive parameter.

As $\hat{x}$ is in multi-bit format, a second-order standard $\Sigma\Delta$ modulator is used

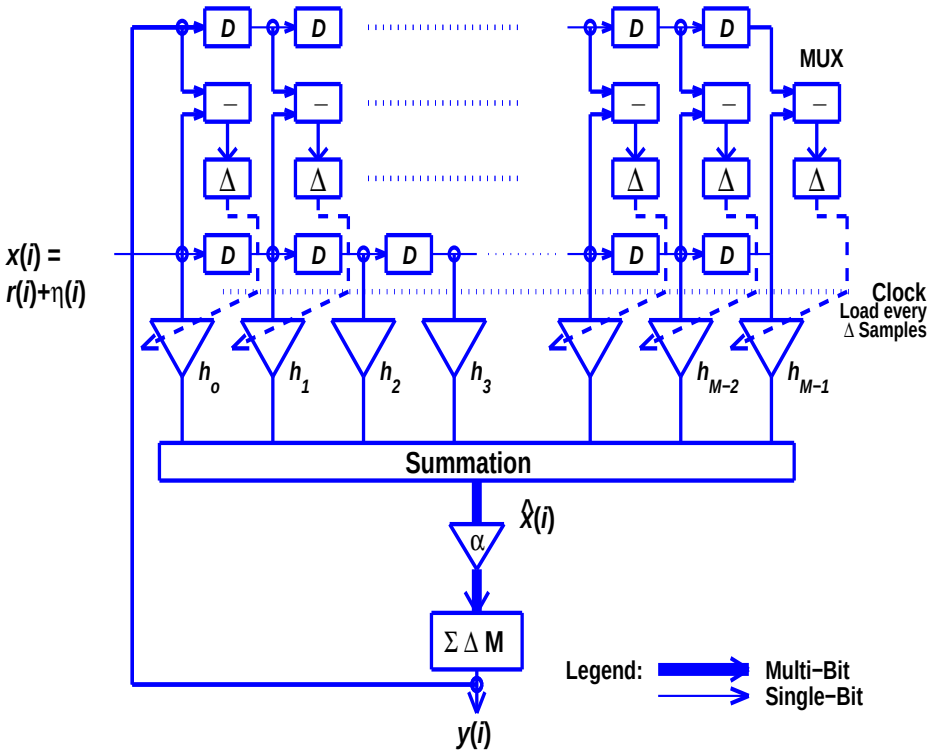

Figure 8.2: Structure of the Adaptive ternary filter. Note that "D" represents a single-bit delay element.

to convert $\hat{x}$ to a bit-stream (re-modulation). This $\Sigma\Delta$ stage should have a flat frequency response in the band of interest such that the information in $\hat{x}$ is maintained. However, this stage will inevitably introduce noise to the output y_i due to the quantization error Q_i. In addition, to preserve stability of the system, the value of $\hat{x}$ should be maintained within the dynamic range of the $\Sigma\Delta$ modulator. This can be guaranteed by introducing the gain parameter α, which is a small positive number. The parameter α is dependent on the oversampling ratio (R)and the filter order at Nyquist rate N. This is so because the number of taps is proportional to R. For simplicity of implementation, α takes on negative powers of 2.

The instantaneous single-bit estimated output y_i is given as follows:

$$y_i = \mathrm{sgn}(u_i) \tag{8.3}$$

where u_i is the quantizer input of the $\Sigma\Delta$ given by:

$$u_i = 2u_{i-1} - u_{i-2} - 2y_{i-1} + y_{i-2} + \hat{x}_{i-1} \tag{8.4}$$

From eq.(8.1), the filter coefficients vector can be given as:

$$\mathbf{h}_n = \frac{1}{2}[\mathbf{r}_n - \mathrm{sgn}(\mathbf{u}_n)] \tag{8.5}$$

From eq.(8.5), it is evident that elements of $\mathbf{h}_n \in \{0, +1, -1\}$.

8.2.1 Simulation and Discussion:

To assess the performance of the proposed adaptive ternary structure in terms of the improvement in the SNR, we attempt to compare it with that of a traditional LMS adaptive algorithm under similar circumstances. Fig.(8.3) illustrates the improvement in terms of the ratio $\rho = \mathrm{SNR_o}/\mathrm{SNR_i}$ versus the $\mathrm{SNR_i}$, where $\mathrm{SNR_o}$ and $\mathrm{SNR_i}$ denote the signal-to-noise ratio (SNR) at the output and at the input of the system, respectively. The oversampling ratio is chosen as $R = 128$, and the number of ternary coefficients is $M = 2560$. The observed signal (input) r_i is assumed to be the single-bit digitized version of the original sinusoid $x(t)$ which is distorted by additive white Gaussian noise $\eta(t)$. The sinusoid has an amplitude $A = 0.5$ and a frequency $f_o = 2000$ Hz.

We assume an adaptive LMS FIR filter with $N = M/R = 20$ coefficients, operating on the same input signal. To be in the safe side, we assume the filter sampling rate as 4× Nyquist rate with infinite bit resolution. Moreover, the optimum μ (that gives minimum MSE) is used in this comparison as shown in Fig.(8.4)(μ=0.0003 in this case).

It is obvious from Fig.(8.3) that the adaptive ternary filter shows superior response (better ρ) when the input SNR, $\mathrm{SNR_i}$, is less than 12 dB. On the other hand, the performance of the adaptive ternary algorithm deteriorates for $\mathrm{SNR_i} > 22$ dB as compared to the multi-bit LMS algorithm which exhibits better ρ until $\mathrm{SNR_i} = 28$ dB.

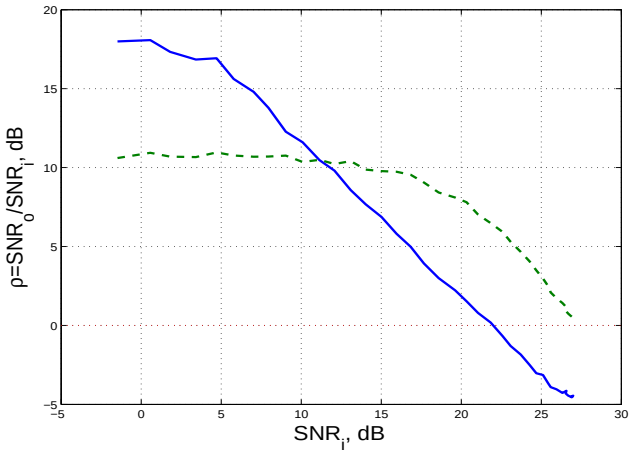

Figure 8.3: SNR improvement using adaptive filtering: ternary (solid) versus LMS (dashed) using a noisy sinusoid.

In Fig.(8.5), the tracking response of the adaptive ternary filter to the noisy sinusoid with $\mathrm{SNR_i} = 10$ dB is depicted. Fig.(8.6) shows the corresponding spectra of both the received and estimated signals. It is evident that the $\mathrm{SNR_o} = 24.28$ dB which means that an improvement of ρ=14.28 dB has been achieved.

This adaptive ternary structure is very efficient from hardware implementation point of view, as the ternary taps can be realized by using simple multiplexers. Moreover, the updating rate Δ can be achieved through the use of a conventional counter.

8.2.2 Discussion:

As per LMS, filter coefficients are updated based on a weighted difference between the filter output (which is an estimation of the original signal) and the reference signal (chosen her to be a delayed version of the input signal itself). This is based on the assumption that noise and signal are uncorrelated; this is true for sinusoids and all narrow-band signals. The estimation error (versus iterations) converge in LMS sense as in Fig.(8.7), similar to the conventional LMS. The system is stable as long as SDM is stable.

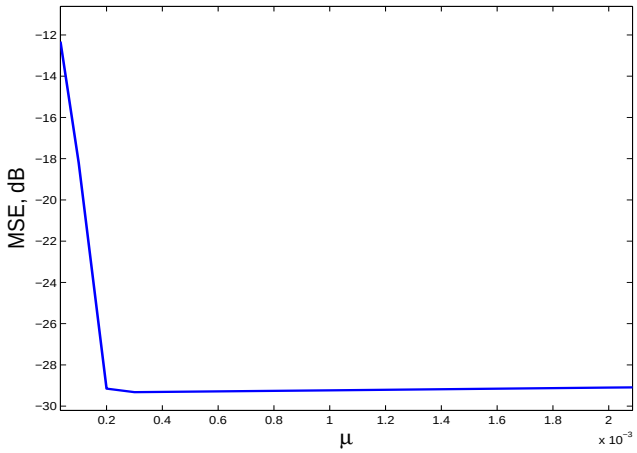

Figure 8.4: The step-size μ versus the mean-square error of the LMS Wiener algorithm.

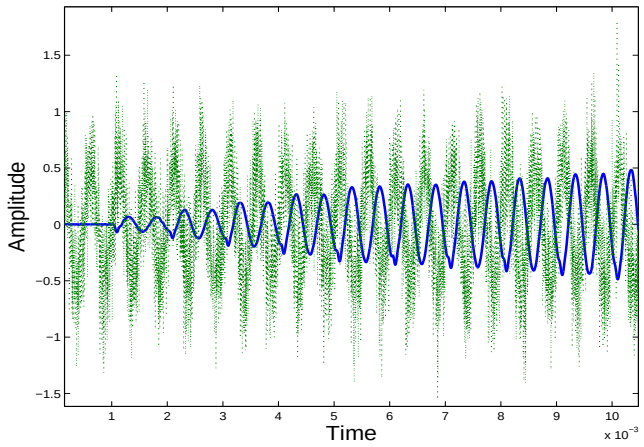

Figure 8.5: The tracking response of the adaptive ternary filter. (solid):estimated output, (dotted):received input.

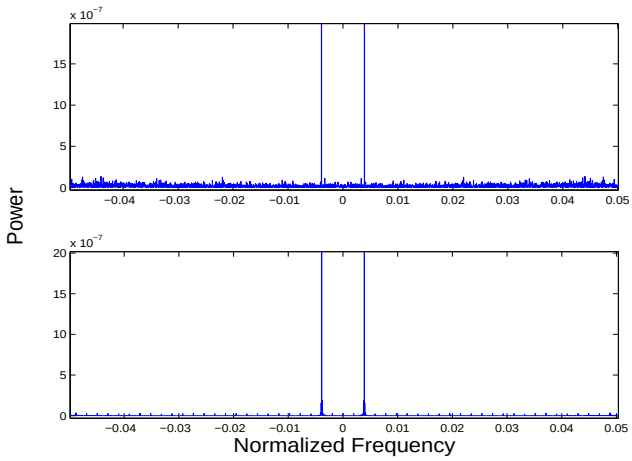

Figure 8.6: The frequency spectra of the received signal (upper), and the estimated signal (lower).

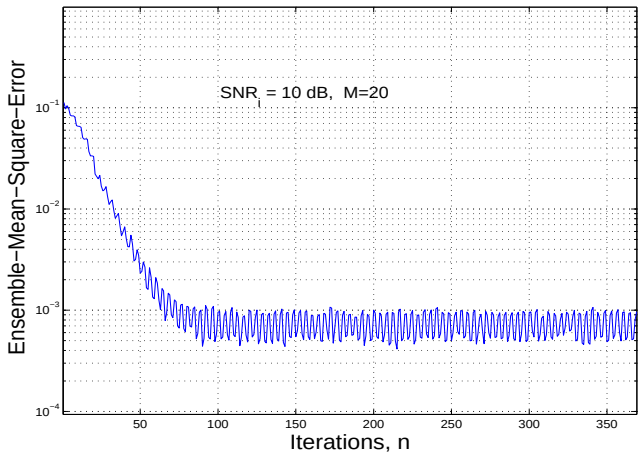

Figure 8.7: Learning curve of the proposed structure for a noisy sinusoid.

8.3 A Single-Bit Adaptive Approach

The proposed adaptive filtering structure is to be entirely achieved in single-bit domain and thus has to operate at an oversampled rate. Intuitively, due to the harsh quantization in the single-bit domain, one might exclude the sample-by-sample scheme of operation as per the standard LMS algorithm. This suggests using blocks of samples instead (the LMS algorithm can be viewed as a special case of the block-LMS with block-length = 1, [133]). Inspired by the conventional block LMS (BLMS) structure and the noise-cancellation principle in Fig.(8.1), Fig.(8.8) illustrates a block diagram of the suggested structure. The input signal is assumed to be a single-bit sigma-delta (SD) modulated noisy sinusoid shifting into a single-bit delay line of the FIR filter, referred to here as FIRb to highlight the fact that its coefficients are all in single-bit format. The FIRb filter order has to be an interpolated version (by a factor of OSR) of its equivalent Nyquist rate order (m), i.e., $M = m \times OSR$. The FIRb filter output (estimated signal) will be in multi-bit format; accordingly, it should be scaled and re-modulated into single-bit format by utilizing a SD modulator. The single-bit estimated signal is then synchronously shifted into a separate shift register (output delay line) with same length as that of the FIRb filter (M). In order to end up with approximated Wiener solution, the adaptive algorithm has to comply somehow with the conventional LMS algorithm to adjust the single-bit tap-weights. The weights are updated once per block of input samples, so they are updated at a rate much lower than the input signal sampling rate.

The main obstacle to be faced here is that the approximated instantaneous gradient will be a switching function since $x(n) \in \{-1, +1\}$. If this case happens in the conventional LMS algorithm, it will definitely lead to unsuccessful convergence to the optimal Wiener solution. This is so because The LMS algorithm relies on a noisy instantaneous estimate for the gradient vector, with the result that the weight-vector estimate for large number of iterations

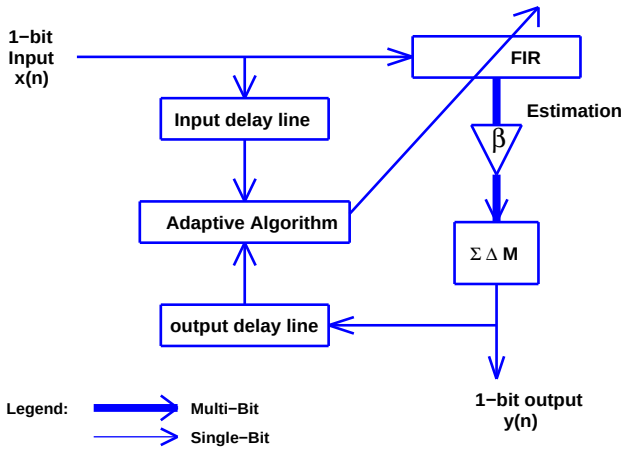

Figure 8.8: A proposed block-diagram for single-bit LMS-like adaptive filtering.

can only fluctuate around the optimum value **w*** in Brownian-motion manner (equivalent to the discrete-time version of the Langevin equation) [126]).

8.3.1 Gradient Approximation

In the standard adaptive FIR algorithm, following the same notations, the output $\mathbf{y}(k)$ is equal to the inner product:

$$\mathbf{y}(k) = \mathbf{x}^T(k)\mathbf{w}(k) \tag{8.6}$$

The error $e(k)$ is defined as

$$e(k) = d(k) - \mathbf{x}^T(k)\mathbf{w}(k) \tag{8.7}$$

where $d(k)$ is the desired response. The purpose of the adaptive algorithm is to adjust the tap weights to optimize the filter response in the sense of minimum mean-square-error (mse). Assuming the that the observed signal and the desired response are wide-sense stationary processes, then by squaring and expanding eq.(8.7) the mean-square error in terms of tap weight values is

given by:

$$e^2(k) = d^2(k) - 2d(k)\mathbf{x}^T(k)\mathbf{w}(k) + \mathbf{w}^T(k)\mathbf{x}(k)\mathbf{x}^T(k)\mathbf{w}(k) \tag{8.8}$$

Taking the expectation to both sides will give

$$E[e^2(k)] = E[d^2(k)] - 2E[d(k)\mathbf{x}^T(k)]\mathbf{w}(k) + \mathbf{w}^T(k)E[\mathbf{x}(k)\mathbf{x}^T(k)]\mathbf{w}(k). \tag{8.9}$$

From this equation, one can define $\mathbf{P}$ as the cross-correlation vector between the desired response and the observed matrix, i.e, $\mathbf{P} = E[d(k)\mathbf{x}]$. Moreover, the input autocorrelation matrix $\mathbf{R}$ is defined as $\mathbf{R} = E[\mathbf{x}\ \mathbf{x}^T(k)]$. Thus, the mean-square error can be described as

$$E[e^2(k)] = E[d^2(k)] - 2\mathbf{P}^T\mathbf{w}(k) + \mathbf{w}(k)^T\mathbf{R}\ \mathbf{w}(k). \tag{8.10}$$

From eq.(8.10), the mean-square error function can be viewed as a concave hyperparaboloidal surface which never goes negative [134]. This function can be represented using gradient method by differentiating eq.(8.10):

$$\blacktriangledown = \partial E[e^2]/\partial \mathbf{w} = -2\mathbf{P} + 2\mathbf{R}\ \mathbf{w}. \tag{8.11}$$

The Wiener-Hopf equations is obtained (in matrix form) by setting the gradient $\bigtriangledown$ to zero, thus, the optimal tap weight vector $\mathbf{w}^*$ is expressed as

$$\mathbf{w}^* = \mathbf{R}^{-1}\mathbf{P} \tag{8.12}$$

Practically, $\mathbf{R}^{-1}\mathbf{P}$ can not be found because of the lack of knowledge of the statistics of both $\mathbf{R}$ and $\mathbf{P}$. The LMS adaptive algorithm has overcome this problem by finding an approximate solutions to (eq.8.12) [124].

The LMS algorithm is an application of the steepest descent method, that

is, the tap weight vector $\mathbf{w}$ is iteratively updated as follows:

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu\blacktriangledown(n) \tag{8.13}$$

where μ is the step-size that controls the rate of convergence and $\bigtriangledown(n)$ is the gradient vector at time n. According to eq.(8.13), the change in the weight vector is proportional to the negative gradient. The LMS algorithm performs an instantaneous gradient descent estimation of the weight vector by assuming $E[e^2(k)] \rightarrow e^2(k)$ then differentiating $e^2(k)$ w.r.t $\mathbf{w}(k)$ (differentiating the instantaneous squared-error w.r.t weight components) as follows:

$$\blacktriangledown(k) = 2e(k)[\partial e(k)/\partial w_o, \cdots, \partial e(k)/\partial w_{M-1}]. \tag{8.14}$$

Thus, the instantaneous gradient will equal to $-2e(k)\mathbf{x}(k)$. Substituting into eq.(8.13), the well-known Widrow-Hoff LMS algorithm [124] is given by:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + 2\mu e(k)\mathbf{x}(k). \tag{8.15}$$

Now applying the same analysis used in the above standard LMS algorithm to a single-bit system (with both the input and the output in single-bit format, i.e., $\mathbf{x}(k)$ and $\mathbf{y}(k) \in \{+1, -1\}$) would definitely drive it to divergence due to the harsh quantization of the input, which will produce a switching instantaneous gradient function $(\partial e^2(k)/\partial w_i)$ that jumps around large quantities; a situation that cannot be tolerated by the adaptive algorithm.

In order to maintain the change of the weight vector (in single-bit adaptation) in a minimal manner sense [124], it is necessary to suggest a feasible solution which must manifest the principle of minimum perturbation [127] which has already been utilized by the existing adaptive algorithms. In addition to the approximation made by Widrow (i.e., $E[e^2(k)] \rightarrow e^2(k)$), it should be taken into account that the error is no longer a continuous variable, in fact $e(k) \in \{-1, 0, 1\}$ hence if we define the function $\gamma(k) \in \{-1, 0, +1\}$, the effect

of $\partial e(k)/\partial w_i$ would be replaced by $\gamma(k)$, noting that $w_i \in \{-1, 1\}$.

To comply with the steepest descent method given in eq.(8.13), the partial differentiation equivalent function $\gamma(k)$ should undergo a minimum change during successive iterations. Thus, the only non-zero choice is to use the approximation $\gamma(k) = 1$. The single-bit gradient function will be given as

$$\bigtriangledown(k) = -2e(k). \tag{8.16}$$

The above formula will be utilized to reach an approximation to the optimal Wiener solution **w*** in the single-bit-domain adaptive filter as will be seen in the next Subsection.

8.3.2 System Design

The input signal is assumed to be a single-bit sigma-delta modulated gaussian noise corrupted sinusoid. Let the $M \times 1$ single-bit input signal vector at time index n be expressed as

$$\mathbf{x}(n) = [x(n), x(n-1), \cdots, x(n-M+1)]^T \tag{8.17}$$

where $[.]^T$ indicates transposition. Let Δ denotes the block length ($1/\Delta$ represents the updating rate), and M represents the length of the interpolated single-bit FIRb filter (i.e., $M = m \times$ OSR, where m is the multi-bit Nyquist rate equivalent filter order). As per conventional BLMS, one may assume any of the following cases: $\Delta < M$, $\Delta = M$, or $\Delta > M$. However, the first two cases ($\Delta \leq M$) are probably preferred in most applications [133] and hence adopted throughout this Section. In addition, for a simpler implementation, it is better to choose power-of-2 values for M and Δ.

Let the single-bit coefficients vector of the FIRb filter at time index n be denoted by

$$\mathbf{w}(n) = [w_o(n), w_1(n), \cdots, w_{M-1}(n)]^T \tag{8.18}$$

and the single-bit estimation output vector at time n be denoted as

$$\mathbf{y}(n) = [y(n), y(n-1), \cdots, y(n-M+1)]^T. \tag{8.19}$$

To proceed in terms of block notation, let k refer to the block index which is related to the original sampling index n as follows

$$n = k\Delta + i \bmod \Delta, \qquad k = 1, 2, 3, ... \tag{8.20}$$

The single-bit input data for block k is therefore defined by the set $\{x(k\Delta + i)\}_{i=0}^{i=\Delta-1}$, which can be expressed in matrix form as

$$\mathbf{A}(k) = [\mathbf{x}(k\Delta), \mathbf{x}(k\Delta+1), \cdots, \mathbf{x}(k\Delta+\Delta-1)]. \tag{8.21}$$

Over this block of input data, the tap-weight vector of the filter is held constant at the value $\mathbf{w}(k)$.

The estimation output, $\hat{x}(k\Delta + i)$, produced by the FIRb filter in response to the input signal vector $\mathbf{x}(k\Delta + i)$ is given by

$$\hat{x}(k\Delta + i) = \mathbf{w}^T(k)\ \mathbf{x}(k\Delta + i). \tag{8.22}$$

This signal is in multi-bit format and should be remodulated into single-bit representation. Practically, this can be done by introducing a sigma-delta modulation stage. This $\Sigma\Delta$ modulator must have a flat signal frequency response over the bandwidth of interest. This implies that the $\Sigma\Delta$ modulator should not modify the specifications of the estimation signal, moreover, it requires operation at an oversampled rate (OSR). This requirement will be satisfied as the input signal has already been assumed here to be a $\Sigma\Delta$ modulated bit-stream. The single-bit version of the estimation output $y(k\Delta + i)$ can be then described as

$$y(k\Delta + i) = \mathrm{sgn}[\alpha\hat{x}(k\Delta + i)]. \tag{8.23}$$

where α is a gain parameter. Using the well-known linear approximation to model the behavior of the sigma-delta modulator [80][128], the output can thus be given as

$$y(k\Delta + i) = \alpha\hat{x}(k\Delta + i) + Q_y(k\Delta + i) \tag{8.24}$$

where $Q_y(k\Delta + i)$ represents the (shaped) quantization noise due to the modulation effect; given by the following convolution

$$Q_y(k\Delta + i) = \alpha \sum_{j=0}^{\Delta-1} h_j q(k\Delta + i - j). \tag{8.25}$$

where h_j characterizes the impulse response coefficients of the noise transfer function of the sigma-delta modulator (note that the term h_o is always unity) and $q(k\Delta+i)$ is the quantization noise. Assuming $q(k\Delta+i)$ is an i.i.d random process and also independent of the input signal $\hat{x}(k\Delta + i)$. Let the $1 \times \Delta$ quantization noise vector is defined as

$$\mathbf{Q}_y(k) = [Q_y(k\Delta), Q_y(k\Delta + 1), \cdots, Q_y(k\Delta + \Delta - 1)]^T. \tag{8.26}$$

Substituting eq.(8.22) into eq.(8.24), the single-bit output is

$$y(k\Delta + i) = \alpha\mathbf{w}^T(k)\ \mathbf{x}(k\Delta + i) + Q_y(k\Delta + i). \tag{8.27}$$

or, in matrix form,

$$\mathbf{y}(k) = \alpha\mathbf{A}^T(k)\ \mathbf{w}(k) + \mathbf{Q}_y(k), \tag{8.28}$$

where $\mathbf{A}(k)$ is an $M \times \Delta$ matrix defined in eq.(8.21).

To develop an adjustment formula for the tap weights vector, we start with defining the error signal. Recalling eq.(8.16) and taking into consideration the block nature of operation of the proposed SBLL, the error vector at any time instant is given by

$$\mathbf{e}(k\Delta + i) = \mathbf{x}(k\Delta + i) - \mathbf{y}(k\Delta + i), \tag{8.29}$$

and is defined at block k as

$$\mathbf{e}(k) = \mathbf{x}(k) - \mathbf{y}(k). \tag{8.30}$$

The tap-weights must be restricted to single-bit format, i.e., $w_j \in \{-1, +1\}$. This task can be carried out by using a single-bit quantizer. Thus, the updating formula may be described as

$$\mathbf{w}(k+1) = \text{sgn}[\mathbf{w}(k) + \mu\mathbf{e}(k)] \tag{8.31}$$

where μ is a proposed step-size parameter of the SBLL filtering. To this end, eq.(8.31) can be utilized to construct the proposed SBLL adaptive structure which is illustrated in Fig.(8.9).

Again, using the linear approach to model the quantizer, the quantizer output can be represented as a combination of the quantizer input and a white quantization noise. Thus, by substituting eq.(8.29) into eq.(8.31), the updating formula can be approximated as

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu[\mathbf{x}(k) - \mathbf{y}(k)] + \mathbf{Q}_w(k). \tag{8.32}$$

where $\mathbf{Q}_w(k)$ is an M-by-1 vector which represents the tap-weight quantization noise. Now, substituting eq.(8.28) into eq.(8.32) yields the final updating formula

$$\begin{aligned} \mathbf{w}(k+1) &= \mathbf{w}(k) + \mu[\mathbf{x}(k) - \\ &\quad \alpha\mathbf{A}^T(k)\mathbf{w}(k) + \mathbf{Q}_y(k)] + \mathbf{Q}_w(k). \end{aligned} \tag{8.33}$$

It is convenient here to recall the updating equation of the conventional BLMS using the same notation as above:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \frac{\mu}{\Delta}\sum_{i=0}^{\Delta-1}\mathbf{x}(k\Delta + i)e(k\Delta + i). \tag{8.34}$$

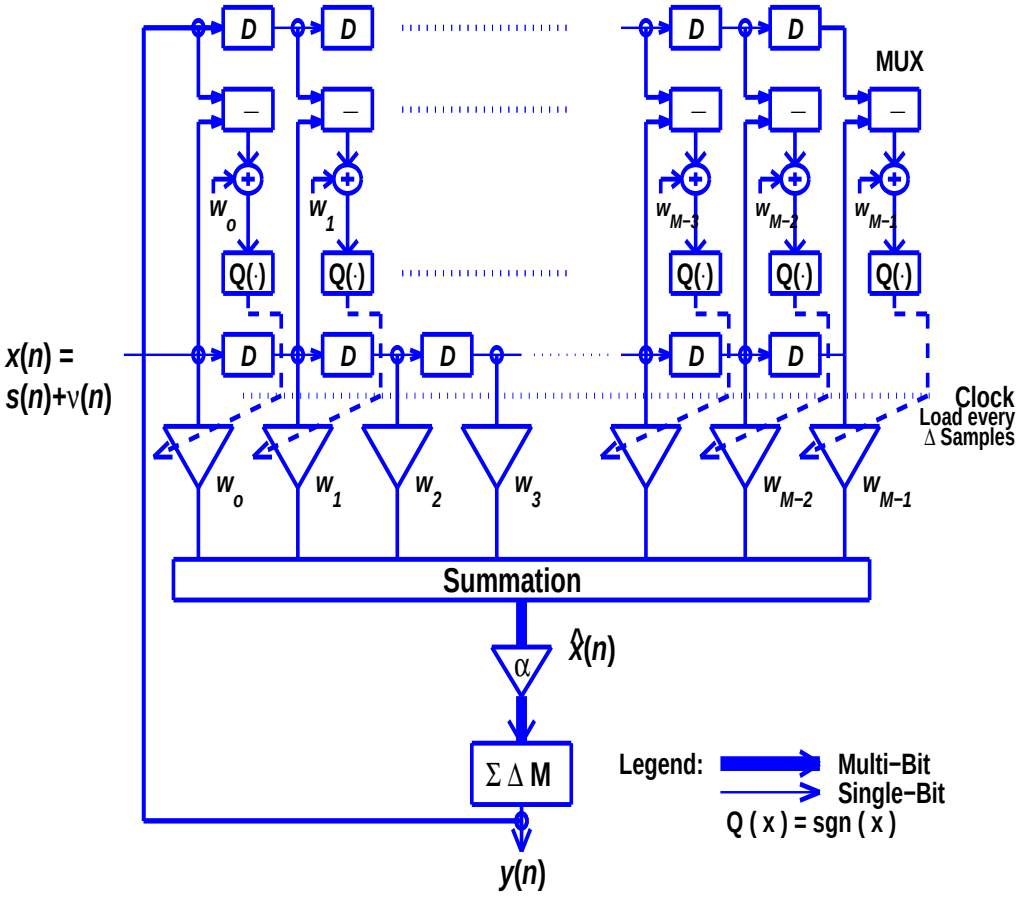

Figure 8.9: The proposed single-bit block LMS-like (SBLL) adaptive structure.

The second term of the right-hand side of eq.(8.34) represents the block gradient, which is a linear correlation between the error signal and the input vector. That is, the error signal is a single sample (produced by averaging the most recent Δ error samples). On the other hand, according to eq.(8.31), the block gradient in SBLL is represented by a single-bit quantized error vector which is described in eq.(8.33) as $\mathbf{x}(k) - \alpha \mathbf{A}^T(k)\mathbf{w}(k) + \mathbf{Q}_y(k)$.

8.4 Stability of SBLL

The convergence properties of the SBLL filtering are the required bounds on the convergence constant (μ), rate of convergence, and the misadjustment.

These properties are equivalent to that of the standard multi-bit BLMS and can be found in several works (e.g., [135][136][133]). However, it is expected that the convergence accuracy (misadjustment) in the SBLL case would be relatively more noisy because of the crude approximation adopted in the iterative calculation of the gradient function. In addition, convergence of the SBLL is affected by two additional issues: the intolerable harsh quantization errors imposed on the recursive weight updating formula (see eq.(8.33)), and the optimum input dynamic range of the SD modulator stage in the sense of maximum attainable SNR.

8.4.1 Dynamic Range of the SD Modulator

As shown in Fig.(8.9), the input to the SD modulator stage $\hat{x}(k)$ is the convolution between the input $x(k)$ and the FIRb filter coefficients (both are interpolated by a factor of OSR). The gain parameter α is introduced to ensure the stability of the SD modulator and is chosen such that it provides maximum SNR. This depends on the SD design parameters as well as the block length (Δ), that is,

$$\alpha = \frac{1}{\Delta}. \tag{8.35}$$

This may suggest that the performance of the SBLL would be improved further in terms of SNR when a suitable adaptive SD modulator scheme is utilized. Adaptive dynamic range single-bit SD modulators can be found in several works [129]-[73]. The second-order SD modulator shown in Fig.(8.10) is used in this work.

8.4.2 The Updating Step-Size

The effective step-size $\hat{\mu}$ for the second term of the gradient estimate on the right-hand side of the updating equation given in eq.(8.33) can be expressed as

$$\hat{\mu} = \alpha\ \mu = \frac{\mu}{\Delta}. \tag{8.36}$$

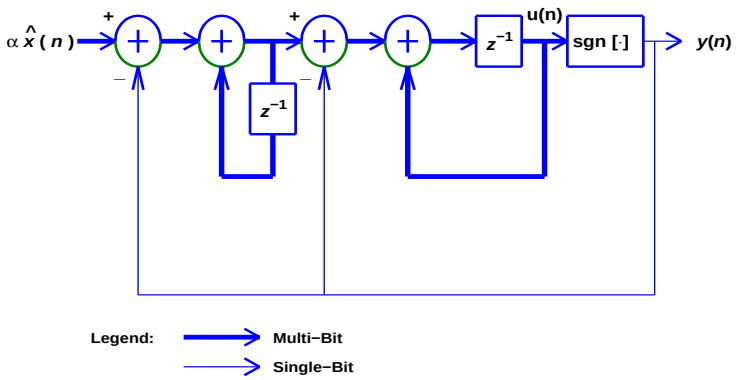

Figure 8.10: The second-order SD modulator used in Fig.(8.9).

This expression conforms to that utilized in the conventional BLMS. The convergence properties of the BLMS and LMS algorithms depend heavily on the eigenvalues of the input autocorrelation matrix. Moreover, the maximum and minimum eigenvalues can be related to the maximum and minimum values of the power spectrum. A necessary (but not sufficient) stability bound on the step-size parameter μ of the LMS filter for large FIR length (M) is given in [130] as follows

$$\hat{\mu} = \frac{2}{MS_{\max}}, \quad \text{or} \quad \mu = \frac{2}{mS_{\max}} \tag{8.37}$$

where $S_{\max}$ denotes the maximum value of the power spectral density $S(\Omega)$.

However, for a single-bit signal, the eigenvalues of the input autocorrelation matrix would have differen values and a different meaning. It is evident according to simulation results that the proposed adaptive filter is stable for $0 < \mu < 0.5$ (for instance, the SBLL is always stable for $\mu = 0.4999$).

8.5 Simulation and Discussion

The SBLL performance has been verified using MATLAB simulation. The input signal $x(n)$ is assumed, throughout the simulation unless otherwise is

stated, to be the single-bit (oversampled) digitized version of the original sinusoid $s(t)$ which is distorted by additive white Gaussian noise $\nu(t) \in \mathcal{N}(\sigma^2, 0)$. The sinusoid has an amplitude $A = 0.5$ and a frequency $f_o = 2000$ Hz. The Nyquist rate FIR filter order is assumed as m=20, and the oversampling ratio is chosen as OSR = 128, therefore, the number of single-bit coefficients is $M = 2560$.

8.5.1 Learning Curves

To assess the performance of the proposed single-bit adaptive algorithm, it is necessary to construct its ensemble-average learning curve, which is defined as [127]:

$$J(k\Delta + i) = E|d(k\Delta + i) - y(k\Delta + i)|^2 \tag{8.38}$$

where E denotes the expectation operator. The ensemble-average learning curve over the interval $0 \leq k \leq N$ is defined as the average over the L realizations as:

$$\hat{J}(k\Delta + i) = \frac{1}{L} \sum_{l=1}^{L} |e^{(l)}(k\Delta + i)|^2 \tag{8.39}$$

where $\hat{J}(k\Delta + i)$ is the sample-average approximation of the actual learning curve. The desired response used here is represented by a delayed version of the input signal $x(k\Delta + i)$.

To evaluate the convergence properties of the single-bit adaptive filtering SBLL, Fig.(8.11) depicts a comparison between learning curves of the oversampled SBLL ($M = 20\times$ OSR, where OSR = 128) and the conventional infinite-precision LMS algorithms for a noisy sinusoidal input with Nyquist rate FIR filter order $m = 20$ with additive Gaussian noise power of -27.5 dB and $\text{SNR}_i = 24.6$ dB. Whereas Fig.(8.12) compares the learning curve (decimated by a factor of OSR to return to Nyquist rate) of the SBLL with that of the standard infinite-precision LMS using same parameters as in Fig.(8.11).

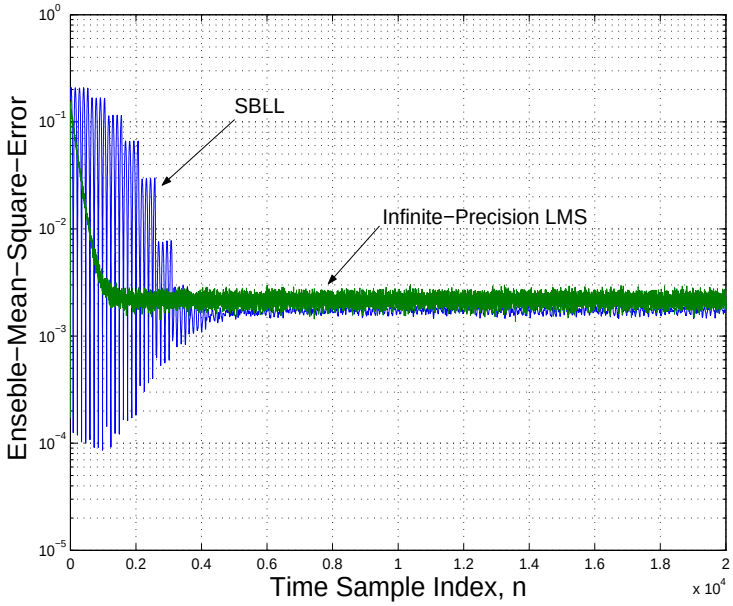

Figure 8.11: A comparison between the (undecimated) learning curves of the single-bit adaptive filter SBLL and the conventional LMS for a noisy sinusoidal input with m=20, SNR_i = 24.6 dB, and noise power -27.6 dB for both cases.

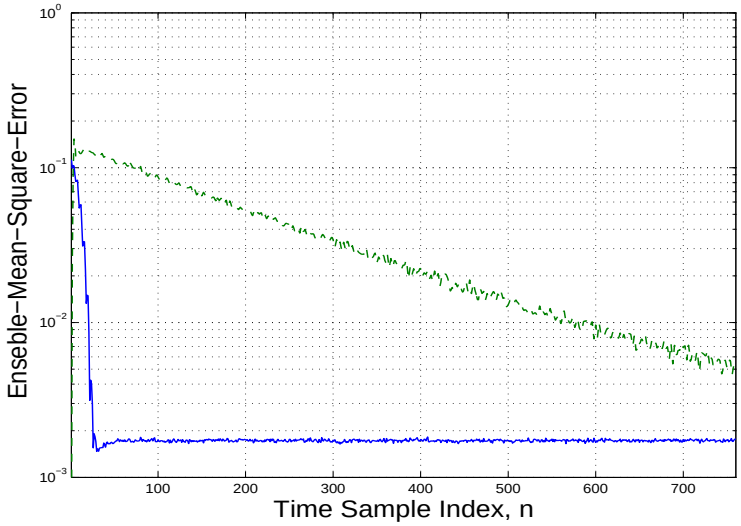

Figure 8.12: A comparison between decimated learning curves of the single-bit adaptive filter SBLL (solid) and the conventional LMS (dashed) with m=20, SNR_i = 24.6 dB, and noise power -27.6 dB for both cases.

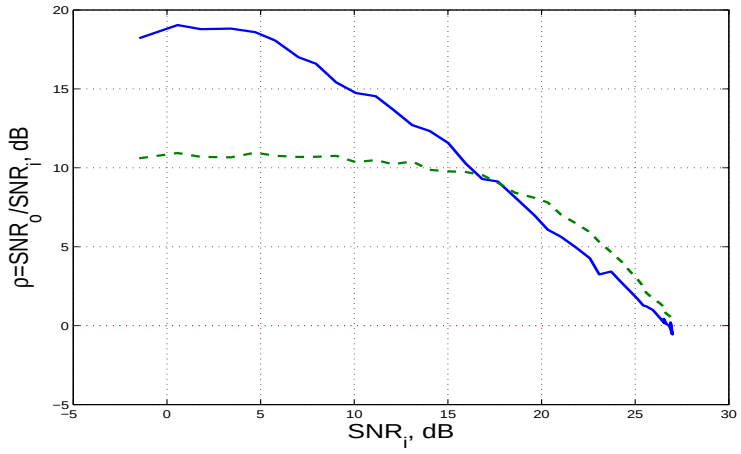

Figure 8.13: A comparison in SNR improvement (ρ) between the SBLL (solid) and the corresponding standard infinite-precision LMS algorithm (dashed) (m=20, OSR = 128).

8.5.2 Signal-to-Noise Ratio (SNR)

This Section is concerned with noise cancellation from narrow-band single-bit input signals. In order to assess the improvement in the output SNR (SNR$_o$) as a function of the SNR at the input (SNR$_i$), a performance parameter ρ is defined as ρ = SNR$_o$(dB)/SNR$_i$(dB). It is noteworthy here that these SNR terms refer to *in-band* signal-to-noise ratios, as we are not interested in frequency bands outside it. Fig.(8.13) shows a performance comparison (in terms of ρ) between the SBLL and the infinite-precision LMS. It is clear that the SBLL outperforms LMS for SNR$_i$ $<$ 16 dB. According to eq.(8.33), this phenomenon would be attributed to dithering effects, as the low SNR$_i$ would be compensated by the uncorrelated white noise due to the harsh quantization as discussed in Sections II and III. On the other hand, the converse occurs for SNR$_i$ $>$ 16. This is expected, using the same argument. However, both algorithms deteriorate (i.e., $\rho < 0$ dB) at almost the same value of SNR$_i$.

To emphasize the noise cancelling effect of the SBLL on the original noisy sinusoid analog input signal (before SD modulation), Fig.(8.14) shows this input with f_o=2 kHz and SNR=10 dB (noise power=-19 dB), along with the demodulated estimation signal (i.e., the decimated and low-pass-filtered ver-

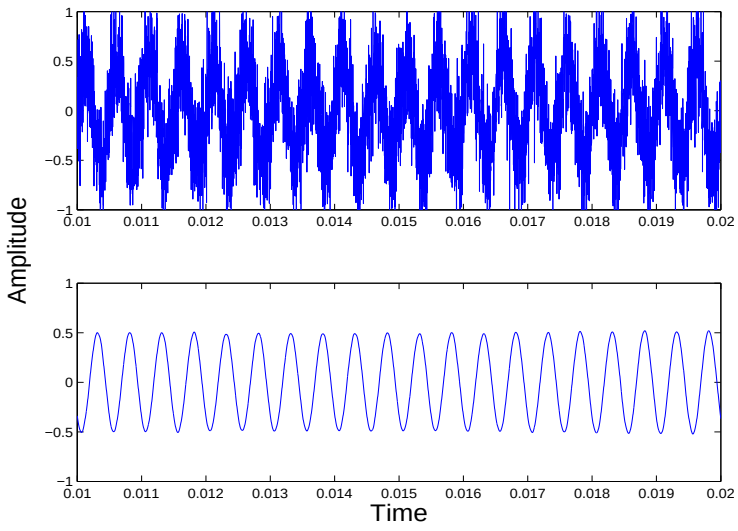

Figure 8.14: A comparison between the the original analog noisy sinusoid, i.e., before SD modulation (above) with SNR = 10 dB, and the output of the SBLL filter (below).

sion of the estimation signal).

On the other hand, the ρ is also affected by the oversampling ratio (OSR) which is a decisive design parameter in single-bit systems. Fig.(8.15) depicts the impact of different OSR (OSR = 64, 128, and 256) on the SNR performance of the proposed adaptive filter. This result is expected, as increasing OSR improves correlation, on which this de-noising is based. However, this will increase the FIRb length M, and therefore, selecting an appropriate OSR becomes a matter of tradeoff between hardware implementation simplicity and design requirements.

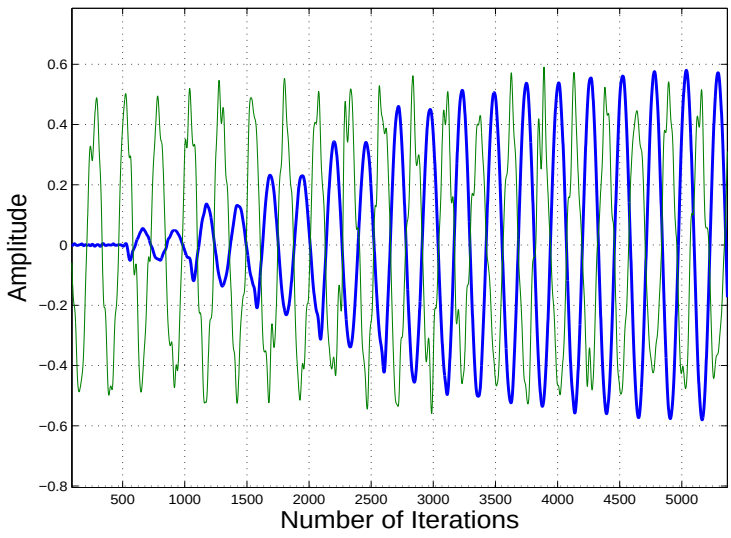

Figure 8.16: Tracking response of the adaptive single-bit filter: (dark) estimated output; (light) received AM-FM input; $\text{SNR}_i = 10$ dB.

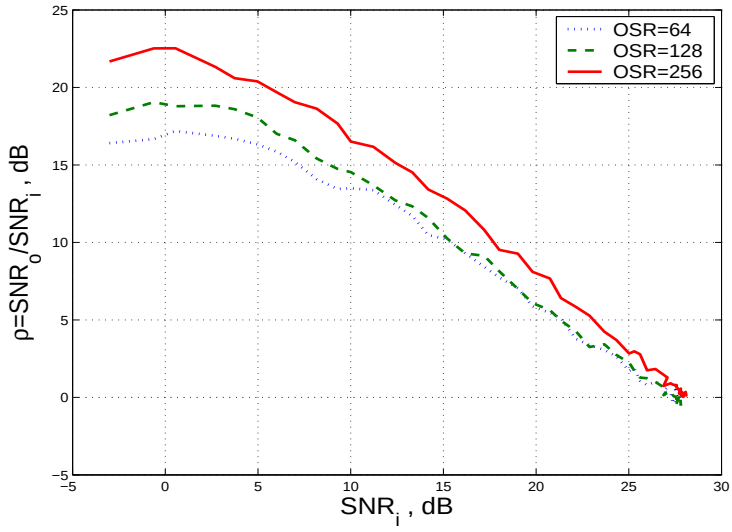

Figure 8.15: SNR improvement using the SBLL as a function of OSR: (solid) OSR = 256, (dashed) OSR = 128, (dotted) OSR = 64.

8.5.3 Non-Stationary Inputs

The SBLL is tested using narrow-band time-varying input signals. Simulations indicate a comparable performance to that of LMS. Fig.(8.16) depicts the the tracking response of the reconstructed (low-pass filtered) SBLL output

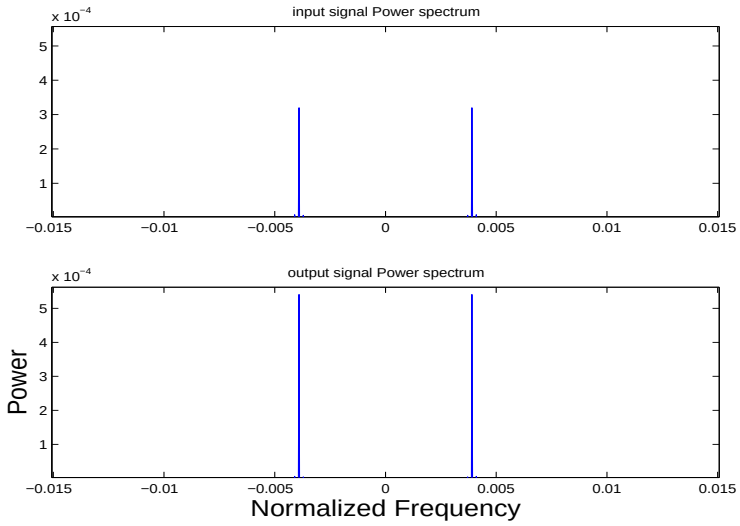

Figure 8.17: Power spectra of the AM-FM input (above, with $\mathrm{SNR}_i = 10$ dB) and the output (estimation) signal (below).

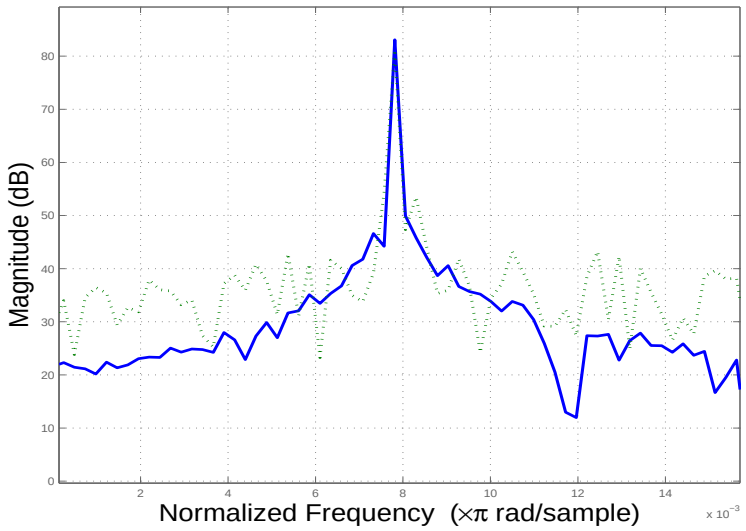

Figure 8.18: Spectrum of the estimated signal using single-bit adaptive filter SBLL (solid) in response to a single-bit AM-FM input (dotted).

to an oversampled single-bit non-stationary input signal which is a digitized version of the AM-FM signal $x = A\cos(2\pi f_1 t)\cos[2\pi(f_o + \beta\sin(2\pi f_2))]$, with modulation index β=0.05. The in-band power spectra of this signal along with the adaptive filter output are shown in Fig.(8.17), whereas Fig.(8.18) illustrates the spectrum of the estimated signal using the SBLL filter in response to the same single-bit AM-FM input.

8.6 A Two-Bit LMS-Like Adaptive Filtering

One of the thesis objectives is to address the wondering about the the optimum word length that might be adopted in short-word length filtering. Of course, single-bit digital filtering represents the ultimate limit of hardware implementation simplicity. This feature applies as well to the ternary structure as long as fixed (non-adaptive) filtering is concerned. This is so because the zero coefficients are considered as "don't care" (i.e., no connections are required). However, this is not the case if adaptive ternary filtering is considered, as the adaptive coefficients vary with time (nonstationary). Hence, adaptive ternary filtering should use two wires to represent the binary set {-1,0,+1}, which is in fact an inefficient usage of the hardware connections. Therefore, moving to 2-bit option seems to be an optimum exploitation to the VLSI chip area as the resolution will increase for the same hardware complexity.

On the other hand, a comparison between single-bit and 2-bit systems is a matter of compromise between the cost and performance, since single-bit adaptive filtering is definitely simpler in hardware realization as compared to its 2-bit counterpart, whereas 2-bit adaptive filtering is expected to show a superior performance.

It has been found that the same single-bit adaptive topology shown in Fig.(8.9) can be utilized after replacing the single-bit quantizers and the connections by their 2-bit counterparts as depicted in Fig.(8.19). That is, the internal $\Sigma\Delta$M is now with a four-level internal quantizer. The input (ob-

served) and the estimation output of the adaptive filter are assumed in 2-bit format, specifically, the set {-1,-0.5,+1,+0.5}.

8.6.1 Performance Comparison

Apart from the 2-bit format, the same analysis technique used in the case of single-bit adaptive filtering in Section-8.3 can be applied to the 2-bit adaptive structure shown in Fig.(8.19).

Learning Curve: Fig.(8.20) shows a comparison between the oversampled learning curve version of the 2-bit adaptivity and the traditional infinite-precision LMS one. While Fig.(8.21) shows the same comparison using the decimated learning curve version. It is evident from these two figures that the conversion rate in the 2-bit LMS-like adaptive filter is comparable to the ternary and single-bit ones.

SNR Improvement: Simulation shows a remarkable improvement in ρ in the case of 2-bit adaptivity relative to the other cases, especially in the large values region of input SNR (SNR_i). Fig.(8.22) shows a comparison among these proposed adaptive schemes. It is quite impressive to find out that the improvement in SNR achieved in the 2-bit LMS-Like adaptive filter exceeds that of the conventional infinite-precision LMS algorithm for all values of SNR_i. In the above figure, at SNR_i=20 dB point $\rho = 14$ dB for 2-bit case, $\rho = 9$ dB for LMS case, $\rho = 6.5$ dB for single-bit case, while $\rho = 2$ dB for the ternary adaptive filter case. The moderate performance of the proposed ternary adaptive structure (compared to that of the single-bit adaptive structure) is due to the filter taps updating method, which is relatively less optimized in the sense of LMS algorithm (see 8.5) compared to that of the single-bit adaptive case.

In Fig. (8.23), the effect of OSR on ρ for 2-bit adaptivity is depicted. Three different values of OSR (64, 128, and 256) are used.

To demonstrate the performance efficiency of the 2-bit adaptive filtering, Fig. (8.24) shows the input-output power spectra for noisy sinusoid input with SNR_i=10.4 dB. It is evident that SNR_o=31.5 dB has been achieved, hence

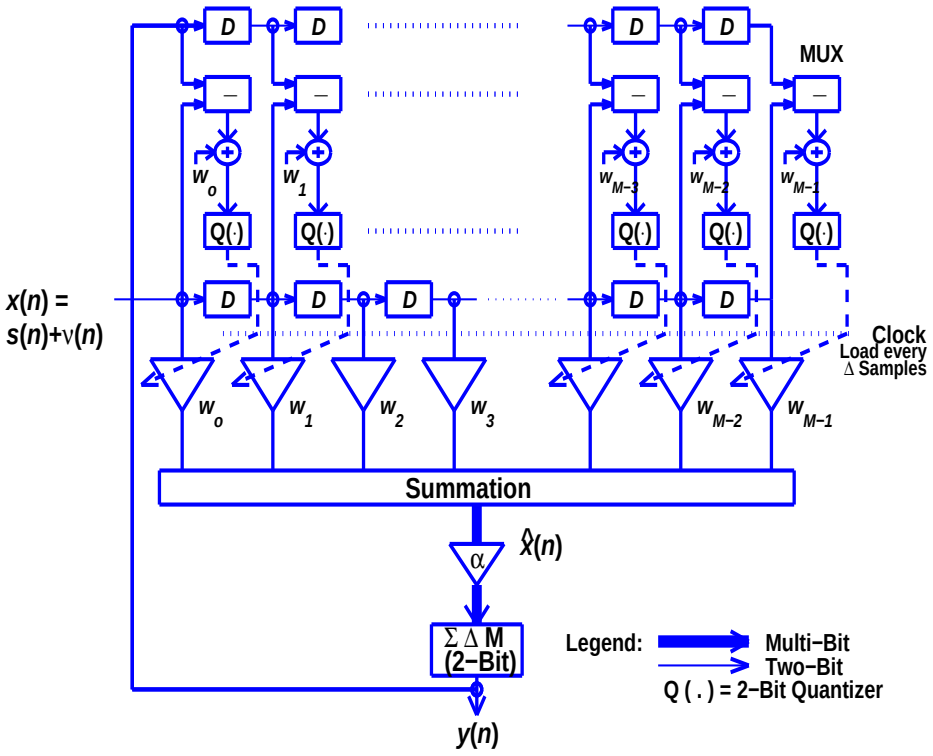

Figure 8.19: The proposed 2-bit block LMS-like (SBLL) adaptive structure.

ρ=21.1 dB.

8.7 Summary

The conventional LMS family of adaptive algorithms fail to converge if translated to the single-bit (ternary) domain. As such the distinctive advantage of short-word length systems, namely, the hardware implementation simplicity, has not been put into effect.

In this chapter we introduced an approach for adaptive ternary filtering. Despite the simple structure, simulation results showed that the proposed algorithm is parallel in performance to the standard multi-bit LMS algorithm. we expect that this approach will open the door for a wide range of applications for ternary systems.

In addition, a single-bit-domain LMS adaptive filtering structure for noise cancelling where all input, output, and filter coefficients are in single-bit for-

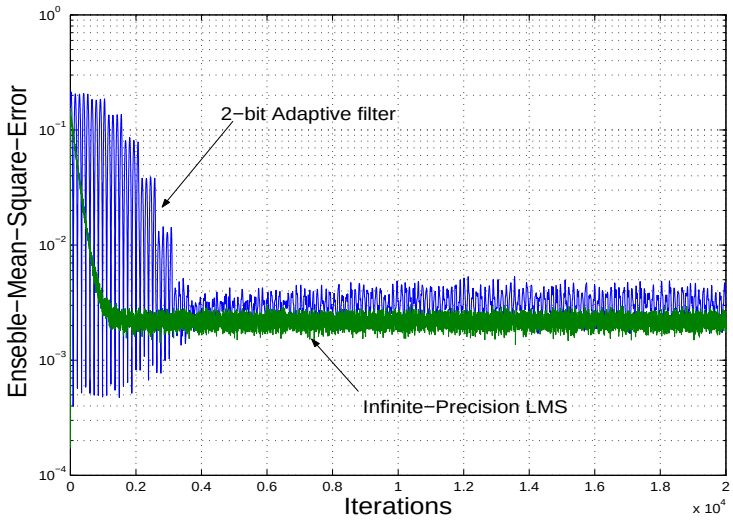

Figure 8.20: A comparison between the (undecimated) learning curves of the single-bit adaptive filter SBLL and the conventional LMS for a noisy sinusoidal input with m=20, SNR_i = 24 dB, and noise power -25.2 dB for both cases.

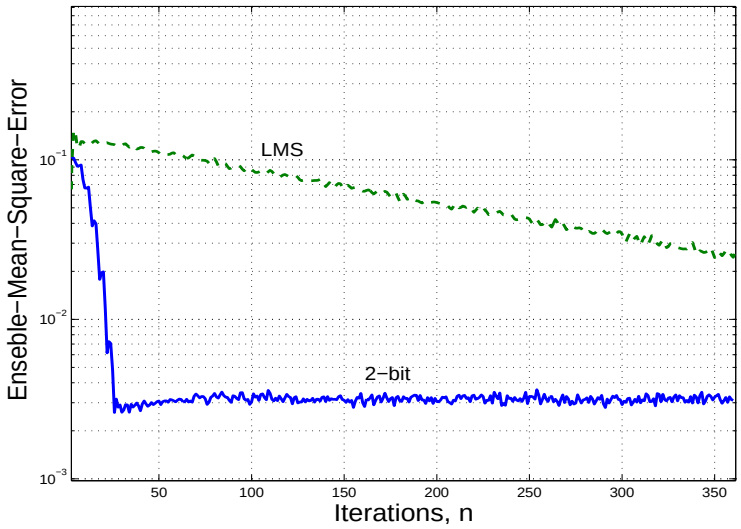

Figure 8.21: A comparison between decimated learning curves of the single-bit adaptive filter SBLL (solid) and the conventional LMS (dashed) with m=20, SNR_i = 24 dB, and noise power -25.2 dB for both cases.

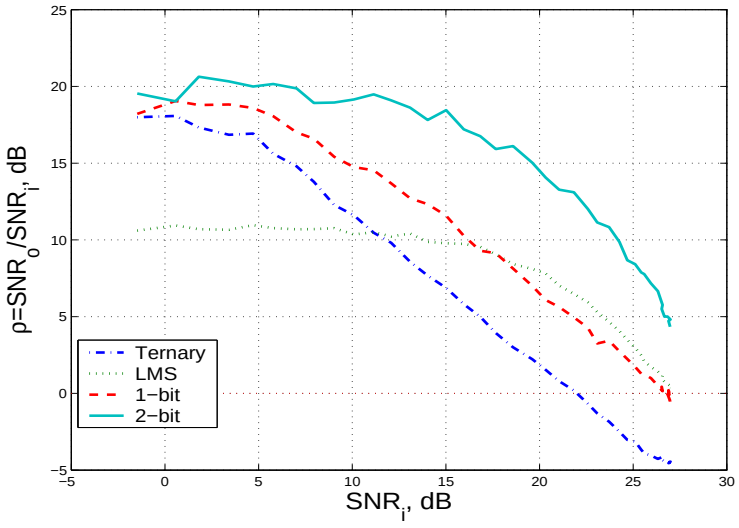

Figure 8.22: A comparison among the proposed LMS-Like adaptive filters and the conventional LMS algorithm (in terms of improvement in SNR represented by ρ).

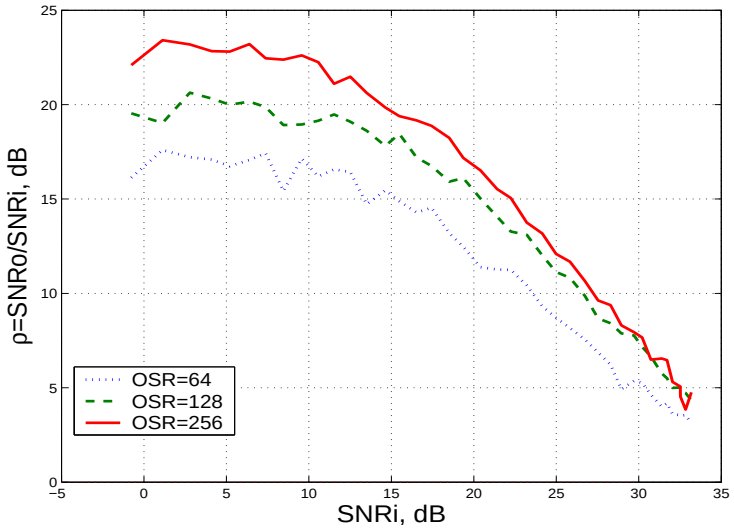

Figure 8.23: SNR improvement using the proposed 2-bit adaptive filter as a function of OSR: (solid) OSR = 256, (dashed) OSR = 128, (dotted) OSR = 64.

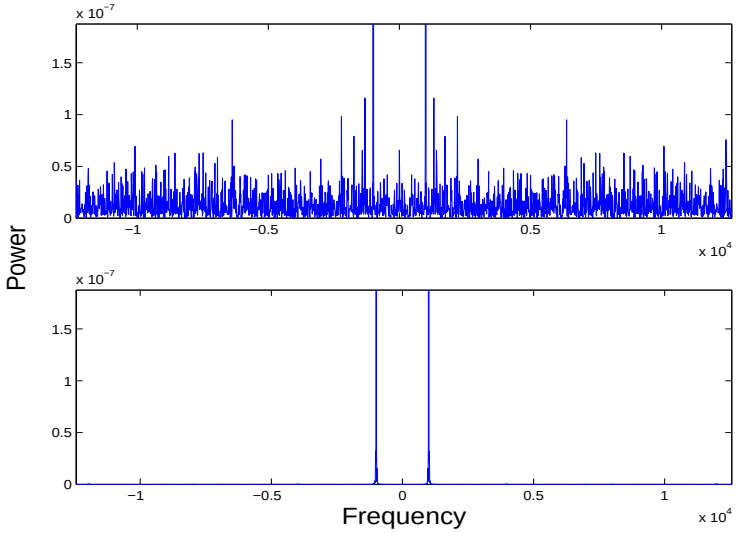

Figure 8.24: Power spectra of a sinusoid input (above, with $\text{SNR}_i = 10.4$ dB) and the corresponding output (estimation) signal (below).

mat. The proposed structure is designed and analyzed, and its performance has been evaluated (in comparison to the conventional Widrow-Hoff multi-bit LMS algorithm) in terms of convergence properties, signal-to-noise improvement, and computational complexity. Simulation results showed that the proposed adaptive structure exhibits performance that is equivalent to the infinite-precision LMS algorithm.

Finally, a 2-bit LMS-like adaptive filter has been constructed by expanding the proposed single-bit adaptive filter into 2-bit structure. This attempt is to search for the optimum word length. Simulation results showed that the 2-bit adaptive filter possesses superior performance over the other two proposed adaptive filters. This is also true for the case of infinite-precision LMS algorithm as long as noise cancelling application is considered.

Chapter 9

Conclusions and Future Work

9.1 Conclusions

In this thesis the design of single-bit, ternary, and 2-bit systems have been considered in an attempt to make short word-length digital signal processing (DSP) ready for general use as an alternative for multi-bit DSP.

Chapters 1 and 2 presented an introduction and a literature survey of single-bit techniques. Chapter 3 presented an introduction to ternary systems.

In Chapter 4, two structures for single-bit digital comb filtering are proposed and simulated. In the first structure, a comb filter is designed based on ternary filtering such that both the input signal and the target impulse response are encoded using a $\Sigma\Delta$ modulator. The second structure is based on a second-order $\Sigma\Delta$ modulator. The frequency response obtained in both cases is very near to the required response of a comb filter. The proposed filters can be built using simple hardware, and hence they are potentially suitable for VLSI implementation. They are also suitable for broadband applications such as power-line communications.

A design technique for single-bit systems using a feedback path filter to tune the response of the $\Sigma\Delta$ modulator is proposed. This may suggest utilizing ternary filters in the feedback loop in future work. A single-bit digital comb filter is designed and its performance is evaluated in terms of signal-to-quantization noise ratio (SQNR), the dynamic range (input signal level),

and stability. Moreover, we showed that the same design technique can be used for other single-bit systems, where we used it to design a multi-period resonator. It was shown that the proposed filters lend themselves very well to broadband input signals and can be utilized in emerging technologies such as the Broad-Band Power-line Communication (BPLC).

In Chapter 5, a ternary DC blocker structure is presented. This type of filtering is useful in practice to improve the stability and dynamic range of single-bit systems. The DC blocker is essentially a ternary filtering structure whose input and output are both assumed to have single-bit format. Performance has been tested for different kinds of input signals, including sinusoidal, FM, and AM-FM signals.

We also proposed a single-bit multiplierless DC-blocking structure. The input is assumed to be sigma-delta modulated bit-stream. This DC-blocker is designed using a delta modulator topology with sigma-delta modulation embedded in its feedback path. Its performance is investigated in terms of the overall signal-to-noise ratio, the effectiveness of DC removal and the stability.

In Chapter 6, a ternary-$\Sigma\Delta$ structure was analyzed mathematically without imposing any approximations. It was evident that the system exhibits a conditional limit cycle behavior. These conditions includes the initial quantization noise conditions and the constant gain parameter in addition to the dc input magnitude. The system was then simulated extensively and a random search method is utilized to discover and extract the limit cycles and identify their features. It seemed that this topology, which is similar to third order $\Sigma\Delta$ modulator possesses a highly non-linear behavior. The issue of limit cycles in higher order modulators is vital for studying the instability problem.

In Chapter 7, the stability problem was addressed using an analogy between the dynamics of the $\Sigma\Delta$ structure and the sinusoidal digital PLL system. An approximate fixed point analysis was presented and a stability criteria was derived. Simulation results were in accord with the theoretical expectations. This analysis can be extended to any higher-order Sigma-Delta topology.

The lack of adaptive LMS structures in single-bit systems has substantially limited their useability despite their major advantage of hardware simplicity over multi-bit systems.

In Chapter 8 of this thesis we introduced an approach for adaptive ternary filtering. Despite the simple structure, simulation results showed that the proposed algorithm is parallel in performance to the standard multi-bit LMS algorithm.

Moreover, a single-bit block LMS-like algorithm that seems to be quite promising is proposed. It has been shown that the proposed single-bit algorithm is comparable in performance to the multi-bit conventional LMS algorithm.

To study the optimum length of the short-word length, we introduced a 2-bit block LMS-like algorithm that seems to be quite promising. It has been shown that the proposed 2-bit algorithm has a superior performance compared to the ternary and single-bit adaptivity. It was quite impressive to find out that 2-bit adaptive scheme outperforms the conventional multi-bit LMS algorithm as long as the noise cancelling application is concerned.

A comparison among these short-word length adaptive techniques suggests a compromise between relative complexity and performance, and hence, choosing any one of them depends on the application requirements. However, according to the simulation results, we believe that 2-bit systems would be a reasonable replacement to the conventional PCM systems. We expect these adaptive approaches to open the door for short-word-length systems to be a practical alternative to multi-bit signal processing systems.

9.2 Future Work

Every effort has been made in this thesis to tackle the problematic and unresolved issues in ternary and single-bit sigma-delta modulator-based systems. However, as it is the case with all active topics, research will never stop at a

certain edge.

As the VLSI technology is reaching the physical limit both in speed and integration level, it is expected that within the next two decades, photonics will replace the conventional electronics platform. Once this occurred, and it will, the short-word length era will undoubtedly erupt and short-word length systems would replace the existing traditional PCM approach.

Apart from the new single-bit and ternary DSP applications as well as the new stability approach proposed in this Thesis, we think that the most important achievement in this work is the establishment of adaptive short-word length (single-bit, ternary and 2-bit) LMS-Like algorithms. This is so because this problem has been considered as an unresolved issue in DSP. Accordingly, we strongly recommend further research in the following directions:

1. There are some fundamental adaptive applications that need to be investigated, specifically, the short-word length *equalizer*. Also, the short-word length *matched filter* is of vital importance. These two systems should be addressed if we want to utilize single-bit ternary structures in digital communication systems. However, these tasks are nontrivial and need a lot of innovative work.
2. Studying the effect of *dithering* on the performance of the proposed short-word length adaptive structures.
3. Attempting to design and implement a complete short-word length *communication system* using FPGAs. This aim is largely dependent on the progress that may occur in the preceding future research. However, we think a communication system which is partly implemented using short-word length subsystems is feasible for the time being.

Appendix A

Recursive Equation of Third-Order ΣΔ Topology

Appendix A: Proof of Equation (6.5)

We are to prove the recursive equation (6.5)

$$u(k) = 3u(k-1) - 3u(k-2) + u(k-3) - (\alpha+2)y(k-1) + 3y(k-2) - y(k-3) + x. \tag{A.1}$$

Let $u(0) = u_o$, $u(1) = u_1$, and $u(2) = u_2$ denotes the output initial conditions of the final integrator (just before the quantizer), and y_o, y_1, y_2 denote its corresponding quantizer output values. Then,

$$u(3) = 3u_2 - 3u_1 + u_o - (\alpha+2)y_2 + 3y_1 - y_o + x \tag{A.2}$$

$$u(4) = 6u_2 - 8u_1 + 3u_o - 3(\alpha+1)y_2 + 8y_1 - 3y_o - (\alpha+2)y(3) + 4x \tag{A.3}$$

$$\begin{aligned} u(5) \;=\; & 10u_2 - 15u_1 + 6u_o - (6\alpha+4)y_2 + 15y_1 - \\ & 6y_o - 3(\alpha+1)y(3) - (\alpha+2)y(4) + 10x. \end{aligned} \tag{A.4}$$

$$\begin{aligned} u(6) &= 15u_2 - 24u_1 + 10u_o - 5(2\alpha+1)y_2 + 24y_1 - 10y_o - \\ &\quad (6\alpha+4)y(3) - 3(\alpha+1)y(4) - (\alpha+2)y(5) + 20x \end{aligned} \tag{A.5}$$

$$\begin{aligned} u(7) &= 21u_2 - 35u_1 + 15u_o - (15\alpha+6)y_2 + \\ &\quad 35y_1 - 15y_o - 5(2\alpha+1)y(3) - (6\alpha+4)y(4) - \\ &\quad 3(\alpha+1)y(5) - (\alpha+2)y(6) + 35x. \end{aligned} \tag{A.6}$$

$$\begin{aligned} u(8) &= 28u_2 - 48u_1 + 21u_o - (21\alpha+7)y_2 + \\ &\quad 48y_1 - 21y_o - 3(5\alpha+2)y(3) - 5(5\alpha+1)y(4) - \\ &\quad (6\alpha+4)y(5) - 3(\alpha+1)y(6) - (\alpha+2)y(7) + 56x. \end{aligned} \tag{A.7}$$

A general recursive formula for $u(k)$ can be found by induction using the above difference equations. We first arrange the coefficients of the initial conditions and the input x as in Table-A.1. Now we induce the coefficient formulas for each term as follows:

Table A.1: Coefficients of the initial conditions

n	u_2	u_1	u_o	y_2	y_1	y_o	x
3	3	-3	1	$-(\alpha+2)$	3	-1	1
4	6	-8	3	$-3(\alpha+1)$	8	-3	4
5	10	-15	6	$-(6\alpha+4)$	15	-6	10
6	15	-24	10	$-5(2\alpha+1)$	24	-10	20
7	21	-35	15	$-(15\alpha+6)$	35	-15	35
8	28	-48	21	$-(21\alpha+7)$	48	-21	56

u_2: $\frac{1}{2}k(k-1)$

u_1: $-k(k-2)$

u_o:$\frac{1}{2}(k-1)(k-2)$

$y_2 : -[(k-1) - \frac{1}{2}(k-1)(k-2)\alpha]$

$y_1 : k(k-2)$

$y_o : \frac{1}{2}(k-1)(k-2)$

$x : \frac{1}{6}k(k-1)(k-2)$

Then we consider the terms including $y(k)$. Table-A.2 shows a few coefficients of $y(k)$. These terms can be represented by a convolution between the output sequence $\{y(n)\}$ and the sequence $\{\eta_\alpha(n) = (\alpha n/2 + 1)(n+1)\}$ as follows:

$$-y(k) * \eta_\alpha(k) = -\sum_{n=1}^{k-3} \eta_\alpha(n) y(k-n). \tag{A.8}$$

Table A.2: Coefficients of the signum terms

k	$y(k-1)$	$y(k-2)$	$y(k-3)$	$y(k-4)$	$y(k-5)$
3	0	0	0	0	0
4	$-(\alpha+2)$	0	0	0	0
5	$-(\alpha+2)$	$-3(\alpha+1)$	0	0	0
6	$-(\alpha+2)$	$-3(\alpha+1)$	$-(6\alpha+4)$	0	0
7	$-(\alpha+2)$	$-3(\alpha+1)$	$-(6\alpha+4)$	$-5(2\alpha+1)$	0
8	$-(\alpha+2)$	$-3(\alpha+1)$	$-(6\alpha+4)$	$-5(2\alpha+1)$	$-3(5\alpha+2)$

Therefore, the overall recursive formula for $u(k)$ can now be given as:

$$\begin{aligned} u(k) &= \frac{1}{2}k(k-1)u_2 - k(k-2)u_1 + b(k)u_o \\ &\quad -b(k)y_o + k(k-2)y_1 - [(k-1) - \alpha b(k)]y_2 \\ &\quad -g(k,\alpha) + d(k)x \end{aligned}$$

where $b(k)$, $g(k)$, and $d(k)$ are given by

$$b(k) = \frac{(k-1)(k-2)}{2} \tag{A.9}$$

$$g(k,\alpha) = \sum_{n=1}^{k-3}[(\frac{\alpha}{2}n+1)(n+1)]y(k-n) \tag{A.10}$$

$$d(k) = \frac{kb(k)}{3}. \tag{A.11}$$

Appendix B

The Equivalent Function $f(k)$

Appendix B: Proof of Equation (6.12)

From (6.9) we have:

$$x = g(k,\alpha)/d(k) = \frac{6}{k(k-1)(k-2)} \sum_{n=1}^{k-3} (\frac{\alpha}{2} n + 1)(n+1)\text{sgn}(u_{k-n}).$$

The asymptotic effect of the product $(k-1)(k-2)$ in the denominator (as $k \to \infty$) is to be replaced by a certain function $f(n)$ within the above summation, that is

$$\frac{1}{(k-1)(k-2)} \sum_{n=1}^{k-3} a(n) \to \sum_{n=1}^{k-3} \frac{a(n)}{f(n)} \text{ as } k \to \infty \tag{B.1}$$

where $a(n) = (\frac{\alpha}{2} n + 1)(n+1)\text{sgn}(u_{k-n})$. Let the left- and right-hand sides of (B.1) be denoted as $S(n)$ and $H(n)$, respectively. The r^{th} discrete derivative (rate of change) of both sides of (B.1) should be

$$\frac{\Delta^r S(k)}{\Delta k^r} = \frac{\Delta^r H(k)}{\Delta k^r}. \tag{B.2}$$

At any arbitrary iteration k, the first-order discrete derivative can be found

as follows. Using Table III we get the first few expressions for $S(k)$

$$\begin{aligned} S(6) &= \frac{-1}{5\times 4}[(\alpha+2)+3(\alpha+1)+(6\alpha+4)] \\ S(7) &= \frac{-1}{6\times 5}[(\alpha+2)+3(\alpha+1)+(6\alpha+4)+5(2\alpha+1)] \\ S(8) &= \frac{-1}{7\times 6}[(\alpha+2)+3(\alpha+1)+(6\alpha+4)+5(2\alpha+1)+3(5\alpha+2)] \end{aligned}$$

form which we get the differences

$$\frac{\Delta S}{\Delta k} = S(7) - S(6) = \frac{1}{3}(\frac{\alpha}{2}) + \frac{1}{60} \tag{B.3}$$

$$\frac{\Delta S}{\Delta k} = S(8) - S(7) = \frac{1}{3}(\frac{\alpha}{2}) + \frac{1}{105}. \tag{B.4}$$

By induction, the first-order discrete derivative of $S(k)$ can be described as

$$\frac{\Delta S}{\Delta k} = S(k+1) - S(k) = \frac{1}{3}(\frac{\alpha}{2}) + \frac{2}{k(k-1)(k-2)}. \tag{B.5}$$

The second term of (B.5) represents a transient response and vanishes rapidly for large values of k, i.e.,

$$\frac{\Delta S}{\Delta k} \to \frac{1}{3}(\frac{\alpha}{2}), \quad \text{as } k \to \infty. \tag{B.6}$$

Using (B.6), it is evident that the second-order discrete derivative of $S(k)$ is equal to zero. On the other hand, from (B.1) the difference of $H(k)$ is given as:

$$\frac{\Delta H(k)}{\Delta k} = \frac{[\frac{\alpha}{2}(k-2)+1](k-1)}{f(k-2)}. \tag{B.7}$$

Now $f(k)$ can be determined by equating the asymptotic rate of change of both functions in (B.6) and (B.7)

$$f(k) \to 3(k+\frac{2}{\alpha})(k+1) \text{ as } k \to \infty. \tag{B.8}$$

Finally, substituting (B.8) into (6.9) we get

$$x \to \frac{2}{k}\sum_{n=1}^{k-3}\frac{(\frac{\alpha}{2}n+1)\mathrm{sgn}(u_{k-n})}{(n+2/\alpha)} \quad \text{as} \quad k\to\infty. \tag{B.9}$$

Now we reach equation (6.12) as follows

$$x \to \frac{\alpha}{k}\sum_{n=1}^{k}\mathrm{sgn}(u_{k-n}) + \frac{2}{k}\sum_{n=1}^{k}\frac{\mathrm{sgn}(u_{k-n})}{(n+2/\alpha)} \quad \text{as} \quad k\to\infty.$$

Appendix C

Difference Equation of M^{th}-Order $\Sigma\Delta$ System

Appendix C: Proof of Equation (6.16)

The difference equation that describes the operation of the third-order $\Sigma\Delta$ topology shown in Fig.(6.1) is given by (6.2) and will be rewritten here in its z-domain form with some rearrangement:

$$U(z)(1-z^{-1})^3 = Y(z)[-z^{-1}(\alpha+2)+3z^{-2}-z^{-3}]+z^{-1}x. \tag{C.1}$$

Now, referring to Fig.(6.2), which represents the M^{th}-order system, the difference equation of the fourth-order $\Sigma\Delta$ topology will be given by:

$$\begin{aligned} U(z)(1-z^{-1})^4 &= Y(z)[-z^{-1}(\alpha_o+\alpha_1+\alpha_2+\alpha_3)+3z^{-2}(\alpha_1+2\alpha_2+3\alpha_3) \\ &\quad -z^{-3}(\alpha_2+3\alpha_3)+z^{-4}\alpha_3]+z^{-1}x. \end{aligned} \tag{C.2}$$

Similarly, we can proceed to higher-orders.

Focusing on the left-hand side of (C.1) and (C.2), we first recall the binomial expansion

$$(a+b)^M = \sum_{n=0}^{M} \begin{pmatrix} M \\ n \end{pmatrix} a^{M-n}\, b^n, \tag{C.3}$$

then, the left-hand terms of the M^{th}-order topology will be given as

$$(1 - z^{-1})^M = \sum_{n=0}^{M} \begin{pmatrix} M \\ n \end{pmatrix} (-1)^n z^{-n}. \tag{C.4}$$

Therefore, the M^{th}-order system can be expressed in the time-domain as

$$u(k) = \sum_{n=1}^{M} \begin{pmatrix} M \\ n \end{pmatrix} (-1)^{n+1} u(k-n) + R(k) \tag{C.5}$$

where $R(k)$ denotes the remaining terms of the M^{th}-order system which involve the output $y(k)$. To find $R(k)$, we reconsider the terms constituting the right-hand sides of (C.1) and (C.2), where the i^{th} of these terms can be described in the time-domain as follows

$$i^{\text{th}} \text{ term} = \quad (-1)^i y(k-i) \sum_{n=0}^{M-1} \binom{n}{i-1} \alpha_n. \tag{C.6}$$

Thus, $R(k)$ will be expressed as

$$R(k) = \sum_{i=1}^{M} \sum_{n=0}^{M-1} (-1)^i \binom{n}{i-1} \alpha_n \, y(k-i). \tag{C.7}$$

Finally, the last integrator output (i.e., the single-bit quantizer input), $u(k)$, in the M^{th}-order $\Sigma\Delta$ topology under consideration is given as

$$u(k) \quad = \quad \sum_{n=1}^{M} (-1)^{n+1} \binom{M}{n} u(k-n) + \sum_{i=1}^{M} \sum_{n=0}^{M-1} (-1)^i \binom{n}{i-1} \alpha_n y(k-i) + x(k-1).$$

Bibliography

[1] M. Ebar, M. Rieger, and H. Schemmann, "A 1.28-GHz sigma-delta modulator for video A/D conversion," *IEEE Transactions on Consumer Electronics*, vol. 42, no. 3, pp. 357-361, Aug. 1996.

[2] J. Chiang, T. Chang, and P. Chou, "Novel noise Shaping of cascaded sigma-delta modulator for wide bandwidth applications," *IEEE Conference*, pp. 1379-1382, 2001.

[3] K. Cang, G. E. Sobelman, E. Saberinia, and A. H. Tawfik, "Performance of N-tone sigma-delta modulator for UWB-OFDM,"*IEEE Communications Society*, pp. 2483-2486, 2004.

[4] A. C. Thompson, *Techniques in Single-Bit Digital Filtering*, PhD dissertation, University of RMIT, Melbourne, Oct. 2004.

[5] B. Steele and P. O'Shea, "Design of ternary digital filters," *Proceeding of the third International Conference on Information, Communications and Signal Processing (ICICS)*, Singapore, Oct. 2001.

[6] P. W. Wong, "Fully sigma-delta modulation encoded FIR filters," *IEEE Transactions on Signal Processing*, vol. 40, no. 6, pp. 1605-1610, June 1992.

[7] M. Batman and B. Liu, "An approach to programmable CTD filters using coefficients 0, +1, and, -1," *IEEE Transaction on Circuits and Systems*, vol. CAS-27, no. 6, pp. 451-456, June 1980.

[8] N. Benvenuto, L. E. Franks, and F. S. Hill, Jr, "Dynamic programmng-methods for designing FIR filters using coefficients -1, 0, +1," *IEEE Transactions on Accostics, Speech, and Signal Processing*, vol. 34, pp. 785-792, Aug. 1986.

[9] N. Benvenuto, L. E. Franks, and F. S. Hill, Jr, "Realization of finite impulse response filters using coefficients +1, 0, -1," *IEEE Transactions on Communications*, vol. COM-33, no. 10, pp. 1117-1125, Oct. 1985.

[10] N. Benvenuto, *Realization of finite impulse response filters using coefficients +1, 0, and -1*, PhD dissertation, University of Massachusetts, Amherst, Feb. 1983.

[11] S. S. Abeysekera and K. P. Padhi, "Design of multiplier free FIR filters using a LAFD sigma-delta modulator," *ISCAS 2000, IEEE International Symposium on Circuits and Systems*, vol. II, pp. 65-68, May 2000.

[12] B. R. Steele, *Efficient signal processing through the use of sigma-delta modulation and ternary filters*, PhD dissertation, RMIT University, Melbourne, 2003.

[13] S. S. Abeysekera, X. Yao and Z. Zang, "A comparison of various low-pass filter architecture for sigma-delta modulators," *IEEE International Symposium on Circuits and Systems (ISCAS)*, vol. 2, pp. 380-383, June 1999.

[14] J. C. Candy and G. C. Temes, *Oversampling Delta-Sigma Data Converters,* New York: IEEE Press, 1992.

[15] C. S. Gunturk, J. C. Lagarias, and V. A. Vaishampayan, "On the robustness of single-loop sigma-delta modulation," *IEEE Trans. Information Theory*, vol. 47, no. 5, pp. 1735-1744, July 2001.

[16] S. R. Norsworthy, R. Schreier, G. C. Temes (Editors), "Delta-Sigma Data Converters: Theory, Design, and Simulation" *IEEE Press*, 1997.

[17] H. Fujisaka, R. Kurata, M. Sakamoto and M. Morisue, "Bit-stream signal processing and its application to communication systems," *IEE Proceedings Circuits Devices and Systems*, vol. 149, no. 3, pp. 159-166, Jun. 2002.

[18] S. M. Kershaw and M. B. Sandler, "Bit-stream signal processing on a sigma-delta bitstream," *IEE on Colloquium on Oversampling Techniques and Sigma-Delta Modulation*, pp. 9/1-9/8, 1994.

[19] A. V. Oppenheim and R. W. Schafer, "Discrete-Time Signal Processing", Second Edition, Printice-Hall, 1999.

[20] A. Tabatabaei and B. A. Wooley, "A two-path bandpass sigma-delta modulator with extended noise shaping," *IEEE journal of Solid-State Circuits*, vol. 35, no. 12,pp. 1799-1809, Dec. 2000.

[21] P. M. Aziz, H. V. Sorensen, J. van der Spiegel, "An overview of sigma-delta converters," *Signal Processing Magazine, IEEE* vol. 13, no. 1,pp. 61-84, 1996.

[22] V. Engelen and R. Van De Plassche, *Bandpass Sigma-Delta modulators stability analysis, Performance and Design aspects*, Kluwer Academic Publisher, 1999.

[23] R. Schreier and M. Snelgrove, "Bandpass sigma-delta modulation," *IEE Electronics Letters* vol. 25, pp. 1560-1561, 1989.

[24] N. Wong and Tung-Sang Ng, "DC Stability of high-order, lowpass $\Sigma\Delta$ modulators with distinct unit circle NTF zeros," *IEEE Transaction on Circuits and Systems-II: Analog and Digital Signal Processing,* vol. 50, no. 1, pp. 12-30, Jan. 2003.

[25] O. Oliaei, "Sigma-delta modulator with spectrally shaped feedback," *IEEE Transaction on Circuits and Systems-II: Analog and Digital Signal Processing,* vol. 50, no. 9, pp. 518-530, Sept. 2003.

[26] K. C. H. Chao, S. Nadeem, W. L. Lee, and C. G. sodini, "A higher-order to topology for interpolative modulators for oversampling A/D converters," *IEEE Transactions on Circuits and Systems,* vol. CAS-37, 309-318, March 1990.

[27] M. A. Al-Alaoui and R. Ferzli, "An enhanced first-order sigma-delta modulator with a controllable signal-to-noise ratio," *IEEE Transaction on Circuits and Systems-I: regular Papers,* vol. 53, no. 3, pp. 634-643, March 2006.

[28] M. A. Al-Alaoui, "Novel digital integrator and differentiator," *IEE Electronics Letters,* vol. 29, no. 10, pp. 376-378, Feb. 1993.

[29] G. R. Arce, N. A. Grabowski and N. C. Gallagher, "Weighted Median filters with sigma-delta modulation encoding," *IEEE Transaction on Signal Processing,* vol. 48, no. 2, pp. 489-498, Feb. 2000.

[30] A. J. Magrath and M. B. Sandler, "Non-linear deterministic dithering of sigma-delta modulators," IEE, pp. 2/1-2/6, 1995.

[31] S. Ouzounov, H. Hegt, and A. Van Roermund, "Sigma-delta modulators operating at a limit cycle," *IEEE Transacions on Circuits and Systems-II: Express Briefs*, vol. 53, no. 5, pp. 399-403, May 2006.

[32] F. Dachselt and S. Quitzk, "Structure and information content in sequences from the single-loop sigma-delta modulator with dc input," *Preceedings of International Symposium on Circuits and Systems 2004 (ISCAS'04)*, vol. 4, pp. IV-685-688, May 2004.

[33] D. Reefman, J. Reiss, E. Janssen, and M. Sandler, "Description of limit cycles in sigma-delta modulators," *IEEE Transaction on Circuits and Systems-I: Regular Papers,* vol. 52, no. 6, pp. 1211-1223, June 2005.

[34] P. S. V. Nataraj and J. J. Brave, "Relaiable and accurate algorithm to compute the limit cycle locus for uncertain nonlinear systems," *IEE Proceedings- Control theoy Appl.*, vol. 150, no. 5, pp. 457-466, Sept. 2003.

[35] V. Friedman, "The structure of the limit cycles in sigma-delta modulation," *IEEE Trans. COM.*, vol. 36, no. 8, pp. 972-979, 1988.

[36] O. feely and L. O. Chua, "The effect of integrator leak in modulation," *IEEE Trans. Circuits and Systems*, vol. 38, pp. 1293-1305, 1991.

[37] J. D. Reiss and M. B. Sandler, "They exist: Limit cycle in high order sigma delta modulators," *Proceedings of the 114th Convention of the Audio Engineering Society*, 2003.

[38] S. Mann and D. Tylor, "Limit cycle behaviour in the double-loop bandpass sigma-delta A/D converter," *IEEE Trans. Circuits and Systems-II, Analog Digt. Signal Process.*, vol. 46, no. 10, pp. 1086-1089, 1999.

[39] N. Bridgett and C. Lewis, "Effect of initial conditions on limit cycle performance of second order sampled data sigma-delta modulator," *Electronics Letters*, vol. 26, pp. 817-819, 1990.

[40] P. Steiner and W. Yang, "A framework for analysis of higher-order sigma-delta modulators," *IEE Electronics Letters*, vol. 44, no. 1, pp. 1-10, Jan. 1997.

[41] D. Hyun and G. Fischer, "Limit cycles and pattern noise in single-stage single-bit delta-sigma modulators," *IEEE Transactions on Circuit and Systems-I: Fundamental Theory and Applications*, vol. 49, no. 5, pp. 646-656, May 2002.

[42] W. Chou, and R. M. Gray, "Dithering and its effect on sigma-delta and multistage sigma-delta modulators," *IEEE Transaction on Information Theory*, vol. 37, no. 3, May 1991.

[43] S. Hein, "Tone suppression in general double-loop ΣΔM using chaos," *IEEE International Symposium on Circuits and Systems (ISCAS'94)*, vol. 5, pp. 449-452, June 1994.

[44] A. Ucar, "Improved stability of high order sigma-delta modulators," *IEE conference on Advanced A/d and D/A Conversion Techniques and Their Applications*, no. 466, pp. 74-78, July 1999.

[45] M. Vogels, and G. Gielen, "Efficient analysis of the sigma-delta modulators using Wavelets," *IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. III-758-761, May 2000.

[46] H. Han, H. Park, and T. Song, "A new architecture for ultrasound sigma-delta modulation beam former," *IEEE Ultrasonic Symposium*, pp. 1631-1634, 2002.

[47] S. Ardalan, and J. Paulos, "An analysis of nonlinear behaviour in sigma-delta modulators," *IEEE Transactions on Circuits and Systems*, vol. 34, pp. 593-603, 1987.

[48] S.-H. Yu and J.-S. Hu, "Sigma-delta modulators operated in optimized mode," *IEEE ISCAS*, pp. I-1080 - I-1083, 2004.

[49] H. Wang, "A geometric view of ΣΔ modulation," *IEEE Transaction on Circuits and Systems-II*, vol. 39, no. 2, pp. 402-405, 1992.

[50] R. Schreier and Y. Yang, "Stability test for single-bit sigma-delta modulators with second-order FIR noise transfer function," *IEEE Symp. Circuits and Systems*, pp. 1316-1319, 1992.

[51] J. van Engelen and R. van de Plassche, "New stability criteria for the design of lowpass sigma-delta modulators," *Proc. of Int. Symp. Low Power Electronics and Design*, pp. 114-118, 1997.

[52] P. Arpaia, F. Cennamo, P. Daponte, and H. Schumny, "Modelling and characterization of ΣΔ analog-to-digital converters," *IEEE Transactions*

on Instrumentation and Measurement, vol. 52, no. 3, pp. 978-983, June 2003.

[53] R. Farrell and O. Feely, "Application of non-linear analysis techniques to sigma-delta modulators," *IEE Colloquium on Signals Systems and Chaos*, pp. 4/1-4/6, 1997.

[54] R. M. Gray, "Over sampled sigma-delta modulation," *IEEE Trans. Commun.*, vol. COM-35, pp. 481-489, 1987.

[55] R. M. Gray, "Quantization noise spectra," *IEEE Transactions on Information Theory*, vol. 36, pp. 1220-1244, 1990.

[56] H. N. Kuhlmann and A. Buzo, "Double-loop sigma-delta modulation with dc inputs," *IEEE Transactions On Communications*, vol. COM-38, pp. 487-495, 1990.

[57] L. A. Williams and B. A. Wooley, "A third-order sigma-delta modulator with extended dynamic range," *IEEE Journal of Solid-State Circuits*, vol. 29, no. 3, pp. 193-202, March 1994.

[58] S. Hein and A. Zakhor, "On the stability of sigma-delta modulators," *IEEE Trans. Signal Processing*, vol. 41, pp. 2322-2348, 1993.

[59] S. Pinault and P. Lopresti, "On the behaviour of double-loop sigma-delta modulators," *IEEE Trans. CAS II-Analogue and Digital Processing*, vol. 40, pp.467-479, 1993.

[60] R. Farrel and O. Feely, "Bounding the integrator outputs of second order sigma-delta modulators," *IEEE Trans. CAS-II Analogue and Digital signal Processing*, 1997.

[61] P. Steiner and W. Yang, "Stability analysis of the second-order $\Sigma\Delta$ modulator," *IEEE International Symposium on Circuits and Systems (ISCAS'94)*, vol. 5, pp. 365-368, June 1994.

[62] X. Sun and K. R. Laker, "Adaptive Integrator-output bounding (AIB) for second order sigma delta ADC," *IEEE Proceedings of ICSP2000*, pp. 631-634, 2000.

[63] C. Dunn and M. Sandler, "Adaptive sigma-delta modulation for use in ADCs," *Electronics Letters*, vol. 32, no. 10, May 1996.

[64] P. Steiner and W. Yang, "Stability of high order sigma-delta modulators," *IEEE International Symposium on Circuits and Systems (ISCAS'96)*, vol. 3, pp. 52-55, May 1996.

[65] M. A. Aldajani and A. H. Sayed, "SNR performance of an adaptive sigma delta modulator," *International Symposium on Circuits and Systems (ISCAS'2001)*, vol. 1, pp. 392-394, May 2001.

[66] M. Jaggi and C. Chakavarthy, "Instantaneous adaptive delta sigma modulator," *Electrical Engineering Journal*, vol. 11, no. 1, pp. 3-6, Jan. 1986.

[67] Yu. J., M. Sandler, and R. Hwaken, "Adaptive quantisation for one-bit delta sigma modulation," *IEEE Proceedings on Circuits, Devices, and Systems*, vol. 139, no. 1, pp. 39-44, Feb. 1992.

[68] C. Dunn and M. Sandler, "Fixed and adaptive sigma delta modulator with multi-bit quantizer," *Applied Signal Processing*, vol. 3, no. 4, pp. 212-222, 1996.

[69] M. Ramesh and K. Chao, "Sigma delta analog to digital converters with adaptive quantisation," *IEEE Proceedings of Midwest Symposium on Circuits and Systems*, vol. 1.2, pp. 22-25, 1998.

[70] M. A. Aldajani and A. H. Sayed, "Stability analysis of an adaptive structure for sigma delta modulation," *IEEE international Conference on Electronics, Circuits and Systems (ICECS 2000)*, vol. 1, pp. 129-132, Dec. 2000.

[71] M. Aldajani and A. H. Sayed, 'An adaptive structure for sigma delta modulation with improved dynamic range," *Proc. 433d Midwest Symposium on Circuits and Systems*, Lansing, MI, Aug. 2000.

[72] L. O. Chua and T. Lin, "Chaos in digital filters," *IEEE Trans. Circuits and Systems*, vol. 35, pp. 648-658, 1996.

[73] C. M. Zierhofer, "Adaptive sigm-delta modulation with one-bit quantization," *IEEE Trans. Circuits and Systems-II: Analog and Digital Signal Processing*, vol. 47, pp. 408-415, May 2000.

[74] G. I. Bourdopoulos, "Adaptive order reduction scheme for high-order single-bit $\Delta\Sigma$ modulators," *IEEE Transactions on Circuits and Systems-II: Express Briefs*, vol. 51, no. 5, May 2004.

[75] R. A. Naughton, and O. Feely, "Non-linear analysis of a digital phase-locked loops," *European Conference on Circuit Theory and Design*, pp. 121-126, 1997.

[76] R. Baines, "The DSP bottelneck," *IEEE Communications magazine*, pp. 46-54, 1995.

[77] P. W. Wong and R. M. Gray, "FIR filters with sigma-delta modulation encoding," *IEEE Transactions on accostics, speech, and Signal Processing*, vol. 38, pp. 979-990, 1990.

[78] S. M. Kershaw, S. Summerfield, M. B. Sandler, and M. Anderson, "Realization and implementation of a sigma-delta bit stream FIR filter," *IEE Proceedings- Circuits, Devices, and Systems*, vol. 143, no. 5, pp. 267-273, Oct. 1996.

[79] B. Steele and P. O'Shea, "Design of Ternary low-pass filters," *Proceedings of the Fourth Australian Workshop on Signal Processing and applications (WOSPA 2002)*, Dec. 2002.

[80] C. Dick and F. Harris, "FPGA signal processing using Sigma-Delta Modulation innovative combinations of techniques and hardware for system designer," *IEEE Signal Processing Magazine*, pp. 20-35, Jan. 2000.

[81] C. Dick and F. Harris, "High-performance FPGA filters using sigma-delta modulation encoding," *IEEE International Conference on Acostics, Speech and Signal Processing, ICASSP'99*, vol. 4, pp. 2123-2126, 1999.

[82] S. Summerfield and T. Wicks, "An efficient hardware architecture implementing wave digital filters with reduced wordlength coefficients," *IEE Saraga Colloquium on digital and Analog Filters and Filtering systems*, pp. 5/1-5/4, 1992.

[83] J. C. Candy, "Decimation for sigma-delta modulation," *IEEE Transactions on Communications*, vol. COMM-34, pp. 72-76, Jan. 1986.

[84] C. Chen and A. N. Willson, "High-order sigma-delta modulation encoding for the design of efficient multiplierless FIR filters with power-of-two coefficients," *International Symposium on Circuits and Systems*, pp. 2361-2364, Jan. 1997.

[85] C. Chen and A. N. Willson, "High-order sigma-delta modulation encoding for the design of multiplierless FIR filters," *IEE Electronics Letters*, vol. 34, no. 24, pp. 2298-2300, Nov. 1998.

[86] S. R. Powell and P. M. Chau, "Efficient narrowband FIR and IFIR filters based on powers-of-two sigma-delta coefficient truncation," *IEEE Transactions on Circuits and Systems*, vol. 41, no. 8, pp. 497-505, Aug. 1994.

[87] A. C. Thompson, P. O'Shea, Z. M. Hussain, and B. R. steele, "Efficient single-bit ternary digital filtering using sigma-delta modulator," *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 162-166, Feb. 2004.

[88] B. R. Steele, *Efficient Signal Processing Through the Use of Sigma-Delta Modulation and Ternary Filters,* PhD dissertation, RMIT University, Melbourne, 2003.

[89] D. A. Johns and D. M. Lewis, "IIR filtering on sigma-delta modulated signals," *IEE Electronics Letters*, vol. 57, pp. 307-308, 1991.

[90] D. A. Johns and D. M. Lewis, "Design and analysis of delta-sigma IIR based filters," *IEEE Trans. Circuits and Systems-II: Analog and Digital Signal Processing*, vol. 40, pp. 233-240, Apr. 1993.

[91] S. M. Kershaw, S. Summerfield and M. B. Sandler, "On $\Sigma\Delta$ signal processing remodulator complexity ," *IEEE International Symposium on Circuits and Systems, (ISCAS'95)*, vol. 2, pp.881-884, 1995.

[92] D. A. Johns, W. N. Snelgrov, and A. S. Sedra, "Adaptive recursive state space using gradient based algorithm," *IEEE Trans. Circuits and Systems*, vol. 37, pp. 673-684, 1990.

[93] A. C. Thompson, *Techniques in Single-Bit Digital Filtering,* PhD dissertation, RMIT University, Melbourne, 2004.

[94] J. J. Paulos, G. T. Brauns, M. B. Steer, and S. H. Ardalan, "Improved SNR using trilevel delta-sigma modulator," *IEEE International Symposium on Circuits and Systems, Philadelphia, PA,* May 1987.

[95] S. S. Abeysekera, and K. P. Padhi, "Design of multiplier free FIR filters using a LADF sigma-delta modulator," *IEEE International Symposium on Circuits and Systems,* Geneva, Switzerland, May 2000.

[96] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing: Principles, Algorithms, and Applications,* Third Edition, Printice-Hall, 1996.

[97] K. Chuan, G. E. Sobelman, E. Saberinia, and A. H. Tewfik, "Performance of N-tone sigma-delta modulation for UWB-OFDM," *IEEE Communications Society*, pp. 1483-2486, 2004.

[98] HongMo Wang, "On the stability of third-order sigma-delta modulator," *IEE Transactions on Circuits and Systems-II: Analog and Digital Signal Processing*, vol. 47, no. 2, Oct. 2004.

[99] S. R. Norsworthy, I. G. Post, and H. Scott Fetterman, "A 14-bit 80-kHz sigma-delta A/D converter: Modeling, Design, and Performance evaluation," *IEEE Journal of Solid-State Circuits*, vol. 24, no. 2, pp. 256-2266, April 1989.

[100] D. K. Su and B. A. Wooley, "A CMOS oversampling D/A convertor with a current-mode semi-digital reconstruction filter," *IEEE J. of Solid State Circuits*, vol. 28, pp. 1224-1233, Dec. 1993.

[101] M. D. Giles, I. Kale, and R. C. S. Morling, "Effect of coefficient quantization on a general class of sigma-delta based converters- a comparative study," *IEEE Conference on Advanced A/D and D/A Conversion Techniques and Their Applications,* , No. 466, pp. 106-110, 27-28 July 1999.

[102] P. Steiner and W. Yang, "Stability analysis of the second order $\Sigma\Delta$ modulator," *IEEE International Symposium on Circuits and Systems (ISCAS'94)*,vol. 5, pp. 365 - 368, 1994.

[103] R. Schreier, "An empirical study of high-order single-bit delta-segma modulators" *IEEE Transactions on Circuits and Systems-II: Analog and Digital Signal Processing*, vol. 40, no. 8, August 1993.

[104] A. C. Thompson, Z. M. Hussain, and P. O'Shea, "A Single-bit narrowband Bandpass digital filter," *Australian Journal of Electrical and Electrnics Engineering*, in press, 2005.

[105] J. O. Smith III, *Introduction to digital filters with audio applications,* http://ccrma.stanford.edu, May 2004 draft.

[106] R. Gray, "Spectral analysis of quantization noise in a single-loop sigma-delta modulator with dc input," *IEEE Transaction on Communications*, vol. 37, no. 6, pp. 588-599, June 1989.

[107] E. Condon, and O. Feely, "Nonlinear dynamics of a nonideal sigma-delta modulator with periodic input," *Proceedings of 2004 International Symposium on Circuits ans Systems (ISCAS'04)*, vol. 4, pp. IV-804-807, May 2004.

[108] M. Martelli, *Discrete Dynamical Systems and Chaos*, Longman Scientific and Technical, 1992.

[109] R. Schreier, "Destabilizing limit cycles in delta-sigma modulators with chaos," *IEEE International Symposium on Circuits and Systems (ISCAS '93)*, vol. 2, pp. 1369-1372, May 1993.

[110] X. Yu, and Z. Galias, "Periodic behaviors in digital filter with two's complement arithmetic," *IEEE Transactions on Circuits and Systems-I: Fundamental Theory and Applications*, vol. 48, no. 10, pp. 1177-1190, Oct. 2001.

[111] N. Wong and Tung-Sang Ng, "Fast detection of instability in sigma-delta modulators based on unstable embedded limit cycles," *IEEE Transactions on Circuit and Systems-II: Express Briefs*, vol. 51, no. 8, pp. 442-449, Aug. 2004.

[112] N. Wong and Tung-Sang Ng, "State trajectory behavior in high-order, low-pass sigma-delta modulators with distinct NTF zeros," *IEEE 14th International Conference on DSP'2002*, vol. 2, pp. 1053-1055, July 2002.

[113] D. B. Ribner, "A comparison of modulator networks of high-order oversampled $\Sigma\Delta$ analog-to-digital converters," *IEEE Transactions on Circuits and Systems*, vol. 38, no. 2, pp. 145-159, Feb. 1991.

[114] N. A. Fraser, and B. Nowrouzian, "Stability analysis of muliple-feedback oversampled $\Sigma\Delta$ A/D converter configurations," *IEEE Midwest Symposium on Circuit and Systems*, vol. 2, no. 8, pp. 676-679, 2000.

[115] A. Teplinsky, E. Condon, and O. Feely, "Driven interval shift dynamics in sigma-delta modulators and phase-locked loops," *IEEE Transacions on Circuits ans Systems-I: regular papers*, vol. 52, pp. 1224-1235, June 2005.

[116] M. Martelli, *Discrete Dynamical Systems and Chaos*, Longman Scientific and Technical, 1992.

[117] A. H. Nayfeh and B. Balachandran, *Applied non-linear dynamics: Analytical, computational, and experimental methods*, Wiley Series in Non-Linear Science, 1995.

[118] H. C. Osborne, "Stability analysis of N-th power digital phase-locked loop-Part I: First-order DPLL," *IEEE Transactions on Communications*, vol. COM-28, no. 8, pp. 1343-1354, Aug. 1980.

[119] H. C. Osborne, "Stability analysis of N-th power digital phase-locked loop-part II: Second and third-order DPLLs," *IEEE Transactions on Communications*, vol. COM-28, no. 8, pp. 1355-1364, Aug. 1980.

[120] Z. M. Hussain, B. Boashash, M. Hassan-Ali, and S. R. Al-Araji, "A time-delay digital tanlock loop," *IEEE Transactions on Signal Processing*, vol. 49, no. 8, pp. 1808-1815, Aug. 2001.

[121] A. C. Thompson, Z. M. Hussain, and P. O'Shea, "Efficient digital single-bit resonator," *IEE Electronics Letters*, vol. 40, (22), pp. 1396 - 1397, 2004

[122] Proakis, J. G. *et al*: *Contemporary Communication Systems Using Matlab,* 2nd Edition, Brooks Cole : Pacific Grove, 2003

[123] B. Widrow, "Thinking about thinking: the discovery of the LMS algorithm," *IEEE Signal Processing Magazine*, pp. 100-106, Jan. 2005.

[124] B. Widrow *et al*, "Stationary and non-stationary learning characterstics of the LMS adaptive filter," *Proceedings of IEEE*, vol. 64, no. 8, pp. 1151-1162, Aug. 1976.

[125] B. Widrow and M. Kamenetsky, "On the statistical efficiency of the LMS family of adaptive algorithms," *Proceedings of IEEE Conference on Neural Networks*, vol. 4, pp. 2872-2880, July 2003.

[126] S. Haykin, *Adaptive Filter Theory,* 4th Edition, Prentice Hall, 2002.

[127] A. H. Sayed, *Fundamentals of Adaptive Filtering,* IEEE Press, 2003.

[128] B. Widrow, I. Kollar, and M. C. Liu, "Statistical theory of quantization," *IEEE Transactions on Instrumentation and Measurement*, vol. 45, no. 2, pp. 353-361, Apr. 1996.

[129] M. A. Aldajani and A. H. Sayed, "Stability and performance analysis of an adaptive sigma-delta modulator," *IEEE Transactions on Circuits and Systems-II:Analog and Digital Signal Processing*, vol. 48, no. 3, March 2001.

[130] H. J. Butterweck, "A wave theory of long adaptive filters," *IEEE Transactions on Circuits and Systems-I: Fundamental Theory and Applications*, vol. 48, no. 6, June 2001.

[131] E. Pfann and R. Stewart, "LMS adaptive filtering with $\Delta\Sigma$ modulated input signals," *IEEE Signal Processing Letters*, vol. 5, no. 4, April 1998.

[132] B. Sklar, *Digital Communications: Fundamentals and Applications,* Prentice-Hall, Upper Saddle River, NJ, 2001.

[133] G. A. Clark, S. K. Mitra, and S. R. Parker, "Block implementation of adaptive digital filter," *IEEE Transactions on Circuits and Systems*, vol. CAS-28, no. 6, June 1981.

[134] B. Widrow *et al*, "Adaptive noise cancelling: principles and applications," *Proceedings of the IEEE*, vol. 63, no. 12, Dec. 1975.

[135] A. Feuer, "Performance analysis of block least mean square algorithm," *IEEE Transactions on Circuits and Systems*, vol. CAS-32, no. 9, pp. 960-963, 1985.

[136] J. J. Shynk, "Frequency-domain and multirate adaptive filtering," *IEEE Signal Processing Magazine*, pp. 15-37, Jan. 1992.

VITA

Amin Z. Sadik received the B. Sc. (1983) and M. Sc. degrees (1988) in electrical engineering from the University of Baghdad (Iraq) and Baghdad University of Technology (Iraq), respectively. From 1989 to 1995, he was a researcher at the Scientific Research Council, Baghdad, also a lecturer in the School of Electrical Engineering, University of Technology, Baghdad (IRAQ). From 1995-2001 he was a lecturer at the University of Salahaddin, Erbil (Iraq), and from 2001-2004 he was a lecturer at the University of Al-Balqa (Jordan). He is currently pursuing his PhD degree with the School of Electrical and Computer Engineering, RMIT University, Melbourne, Australia, funded by an ARC Discovery Project. His research interests include digital signal processing and digital communications.